

# **Image Analytics and Artificial Intelligence Architectures for Medical and Forensic Applications**

A dissertation submitted by

Rahul Rajendran

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Electrical Engineering

Tufts University

May 2022

© 2022, Rahul Rajendran, All rights reserved

Advisor: Dr. Karen Panetta

# Abstract

Artificial Intelligence (AI) has emerged as a critical technology that enables machines to learn from their experiences, adapt to new inputs, and perform tasks similar to those performed by humans. With improvements in methodology and the availability of large databases, AI systems are capable of achieving exceptional performance on a variety of complex tasks. Despite these groundbreaking advances, challenges continue to exist that inhibit or even prevent the widespread adoption of AI in certain applications. For example, (1) the high complexity and energy demands of current AI models make it difficult to deploy them in resource-constrained environments, (2) the poor quality of data impacts the performance of the AI systems and leads to inaccurate predictions, and (3) the lack of understanding of the underlying process reduces the trust in and the verifiability of the decisions made by the AI system.

This dissertation addresses the present challenges and opportunities by promoting the development of high-performance, low-cost AI-inspired architectures. In particular, this dissertation focuses on addressing the following questions:

- Can the enhancement and augmentation of low-quality data improve the performance of AI systems?
- Is it possible to reduce the number of parameters in a neural network while increasing its accuracy and robustness?
- Can the integration of visual perception and cognition of the expert into current architectures demystify the “black-box” element of AI?

The contributions of this dissertation are expected to facilitate the development of efficient models that will accelerate the adoption of AI in medical and forensic applications.

*To my family*



## Acknowledgments

First and foremost, I want to thank my advisor, Dr. Karen Panetta, whose constant support and guidance, and demand for quality research were instrumental in helping me formulate a comprehensive and well documented thesis. In addition to being a brilliant researcher and professor, she is the kindest and most caring individual that I had the pleasure of working with. Her unwavering faith in me has always pushed me to become a better researcher, and I hope that someday I can be like her.

I want to thank my co-advisor, Dr. Sos Agaian, for his continued support and enthusiasm. He took me under his wings during my Master's and introduced me to the world of image processing and computer vision. He is a great mentor who guided me throughout my Ph.D. journey by offering intuitive insights into complex problems, and always encouraged me to think outside the box.

I would also like to thank my thesis committee members, Dr. Ron Lasser and Dr. Aruna Ramesh, for their valuable comments and insightful feedback. Their feedback has helped shape the main arguments in this thesis and has helped me better communicate my findings to multidisciplinary audiences. I also want to thank Dr. Aruna Ramesh, Ruby Wagmin, Iolanda Gomez, and Gina Boger from the Tufts University School of Dental Medicine for their support and contribution to developing the Tufts Dental Database.

This work would not have been accomplished without the help and support of my nearest and dearest friends. I'm extremely grateful to Shishir Rao, Srijith Rajeev, Shreyas Kamath, Varsha Krishnan, Shefali Pearson, and Prajakta Dhamorikar for their never ending support throughout my life. I want to thank them for always being there for me and offering words of encouragement

when I needed it most. From our vacations to camping adventures to all-nighters, these are memories I will cherish for the rest of my life.

I would like to extend my sincere thanks to my colleagues at the Vision and Sensing Systems Laboratory. I would also like to thank Kim Ellwood, Michael Minafo, and Miriam Santi for all their administrative support during my graduate studies.

Last but most importantly, I would like to thank my parents, Rajendran Perumal and Lalita Rajendran, and my sister Rasika Rajendran for their love, sacrifice, and tremendous support over the years. Needless to say, this thesis would not have been possible without their encouragement along the way. This thesis is dedicated to them.

I gratefully acknowledge the sponsorship granted by:

1. The National Institute of Justice (NIJ) Graduate Research Fellowship in Science, Technology, Engineering, and Mathematics, FY 2019 (Award #: 2019-R2-CX-0036). Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the NIJ.
2. The Tufts University School of Engineering, Center for Applied Brain & Cognitive Sciences (CABCS) under award numbers A452001, ARM981, AR9008. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the CABCS.

# Contents

|                                                                                   |           |
|-----------------------------------------------------------------------------------|-----------|
| <b>Abstract.....</b>                                                              | <b>I</b>  |
| <b>Acknowledgments .....</b>                                                      | <b>IV</b> |
| <b>List of Tables .....</b>                                                       | <b>IX</b> |
| <b>List of Figures.....</b>                                                       | <b>X</b>  |
| <b>List of Publications .....</b>                                                 | <b>XV</b> |
| <b>Chapter 1. Introduction .....</b>                                              | <b>1</b>  |
| 1.1. Overview .....                                                               | 1         |
| 1.2. Objectives .....                                                             | 5         |
| 1.3. Contributions .....                                                          | 6         |
| 1.4. Organization.....                                                            | 7         |
| <b>Section 1: Feature Engineering.....</b>                                        | <b>11</b> |
| <b>Chapter 2. Image Enhancement .....</b>                                         | <b>12</b> |
| 2.1. Introduction.....                                                            | 12        |
| 2.2. Proposed Work .....                                                          | 16        |
| 2.2.1. Guided Filtering .....                                                     | 17        |
| 2.2.2. Edge Enhancement .....                                                     | 19        |
| 2.2.3. Contrast Enhancement.....                                                  | 21        |
| 2.2.4. Image Fusion .....                                                         | 22        |
| 2.2.5. Contrast Limited Adaptive Histogram Equalization .....                     | 23        |
| 2.3. Experimental Results .....                                                   | 26        |
| 2.4. Summary .....                                                                | 28        |
| <b>Chapter 3. Synthetic Data Generation for Thermal Emotion Recognition .....</b> | <b>30</b> |
| 3.1. Motivation.....                                                              | 30        |
| 3.2. Introduction.....                                                            | 31        |
| 3.3. Proposed Work .....                                                          | 36        |
| 3.3.1. Architecture .....                                                         | 37        |
| 3.3.2. Tufts Face Database .....                                                  | 38        |
| 3.3.3. Dataset and synthetic data generation .....                                | 39        |
| 3.4. Results.....                                                                 | 43        |
| 3.4.1. Training Phase .....                                                       | 43        |
| 3.4.2. Comparison .....                                                           | 43        |
| 3.5. Summary .....                                                                | 45        |

|                                                                              |            |
|------------------------------------------------------------------------------|------------|
| <b>Chapter 4. Feature Detection and Matching .....</b>                       | <b>46</b>  |
| 4.1. Matching Antemortem and Postmortem Xray Radiographs.....                | 48         |
| 4.1.1. Motivation .....                                                      | 48         |
| 4.1.2. Introduction .....                                                    | 49         |
| 4.1.3. Proposed Work .....                                                   | 52         |
| 4.1.4. Experimental Results.....                                             | 60         |
| 4.2. Pixel-based Color Transfer .....                                        | 64         |
| 4.2.1. Motivation .....                                                      | 64         |
| 4.2.2. Introduction .....                                                    | 65         |
| 4.2.3. Proposed Color Transfer System.....                                   | 69         |
| 4.2.4. Experimental Results.....                                             | 76         |
| 4.2.5. Extended Application: Fusion of thermal and visible face images ..... | 79         |
| 4.3. Image Stitching .....                                                   | 81         |
| 4.3.1. Motivation .....                                                      | 81         |
| 4.3.2. Introduction .....                                                    | 81         |
| 4.3.3. Proposed Work .....                                                   | 84         |
| 4.3.4. No Reference Measure for Underwater Images .....                      | 92         |
| 4.3.5. Experimental Results.....                                             | 97         |
| 4.3.6. Extended Application: Finger image stitching.....                     | 100        |
| 4.3.7. Summary .....                                                         | 101        |
| <b>Section 2: Image Analytic.....</b>                                        | <b>103</b> |
| <b>Chapter 5. Image Segmentation .....</b>                                   | <b>104</b> |
| 5.1. Motivation.....                                                         | 104        |
| 5.2. Introduction.....                                                       | 105        |
| 5.3. Proposed Work .....                                                     | 109        |
| 5.3.1. Encoder Network.....                                                  | 110        |
| 5.3.2. Dual Decoder Network.....                                             | 113        |
| 5.3.3. Channel Attention (CA) Mechanism.....                                 | 114        |
| 5.4. Experimental Results .....                                              | 116        |
| 5.4.1. Dataset .....                                                         | 116        |
| 5.4.2. Training .....                                                        | 117        |
| 5.4.3. Ablation Study.....                                                   | 118        |
| 5.4.4. Benchmark Results.....                                                | 120        |
| 5.5. Summary .....                                                           | 125        |
| <b>Chapter 6. Image Binarization .....</b>                                   | <b>126</b> |
| 6.1. Motivation.....                                                         | 126        |
| 6.2. Introduction.....                                                       | 127        |
| 6.3. Existing Binarization Techniques .....                                  | 129        |
| 6.4. Human Visual Thresholding Method .....                                  | 133        |

|                               |     |
|-------------------------------|-----|
| 6.5. Computer Simulation..... | 139 |
| 6.6. Summary.....             | 142 |

## **Chapter 7. Initial Evaluation of Tumor Roughness to Predict Renal Cancer Subtypes 143**

|                                                                              |     |
|------------------------------------------------------------------------------|-----|
| 7.1. Introduction.....                                                       | 143 |
| 7.2. Methods .....                                                           | 144 |
| 7.2.1. Patient Population.....                                               | 144 |
| 7.2.2. Tumor Roughness Measure by Using Multi-Level Voxel Box-Counting ..... | 144 |
| 7.3. Statistical Analysis.....                                               | 146 |
| 7.4. Results.....                                                            | 147 |
| 7.5. Discussion.....                                                         | 151 |
| 7.6. Summary.....                                                            | 155 |

## **Chapter 8. Tufts Dental Database ..... 157**

|                                 |     |
|---------------------------------|-----|
| 8.1. Motivation.....            | 157 |
| 8.2. Introduction.....          | 159 |
| 8.3. Related Work .....         | 163 |
| 8.4. Tufts Dental Database..... | 168 |
| 8.5. Performance Review.....    | 178 |
| 8.5.1. Image Enhancement .....  | 179 |
| 8.5.2. Teeth Segmentation.....  | 180 |
| 8.6. Discussion.....            | 184 |
| 8.7. Summary.....               | 189 |

## **Chapter 9. Conclusion and Future Work ..... 190**

## **Reference ..... 195**

## List of Tables

|                                                                                                                                                                                                                           |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Table 3.3.1: Architecture of TERNet. Building blocks are shown in brackets with [kernel size, # of filters] x repetition.....                                                                                             | 38  |
| Table 3.4.1: Comparison of TERNet with state-of-the-art models. ....                                                                                                                                                      | 44  |
| Table 4.1.1: A comparison of the performance of the feature detectors in matching antemortem and postmortem scans .....                                                                                                   | 62  |
| Table 4.1.2: A comparison of the feature detectors in estimating the feature count range for good and bad matches.....                                                                                                    | 63  |
| Table 4.3.1: Performance analysis of the image stitching algorithms .....                                                                                                                                                 | 99  |
| Table 5.3.1: Architectural details of the encoder network.....                                                                                                                                                            | 111 |
| Table 5.4.1: Definitions of the metric utilized in this work.....                                                                                                                                                         | 118 |
| Table 5.4.2: Analysis of various normalization techniques without the features from edge enhanced images. Blue text indicates the best and Red text indicates the second-best performance. GN → Group Normalization ..... | 119 |
| Table 5.4.3: Analysis of various encoder depths with the features from edge enhanced images. Blue text indicates the best and Red text indicates the second-best performance .....                                        | 120 |
| Table 5.4.4: Segmentation results on the KITS datasets. Blue text indicates the best and Red text indicates the second-best performance .....                                                                             | 121 |
| Table 7.4.1: Demographics by tumor subtype .....                                                                                                                                                                          | 148 |
| Table 7.4.2: Multivariable linear regression model or the outcome of kidney cancer subtype. ....                                                                                                                          | 149 |
| Table 8.3.1: A comparison of panoramic radiography datasets.....                                                                                                                                                          | 164 |
| Table 8.4.1: Summary of the categorization of abnormalities based on location .....                                                                                                                                       | 173 |
| Table 8.4.2: Summary of the categorization of abnormalities based on the internal architecture .....                                                                                                                      | 173 |
| Table 8.4.3: Summary of the categorization of abnormalities based on its effects on adjacent structures.....                                                                                                              | 173 |
| Table 8.4.4: Summary of the categorization of abnormalities based on marginal characteristics .....                                                                                                                       | 174 |
| Table 8.4.5: Summary of the final categorization (fifth level) of abnormalities .....                                                                                                                                     | 174 |
| Table 8.5.1: Quality scores on a subset of images from the TDD. CLAHE algorithm performs better in comparison .....                                                                                                       | 179 |
| Table 8.5.2: Summary of the performance of the state-of-the-art algorithms in segmenting teeth from panoramic radiographs. The UNet architecture performs marginally better in comparison .                               | 183 |

## List of Figures

|                                                                                                                                                                                                                                                                                   |    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Fig. 2.1.1: A visual overview of the dissertation. The focus of the dissertation is to develop novel methods and systems for applications in medical and forensic sciences. ....                                                                                                  | 8  |
| Fig. 2.2.1: A workflow of the proposed edge-aware enhancement system .....                                                                                                                                                                                                        | 16 |
| Fig. 2.2.2: The figure displays the result of the guided filtering process where (a) original image, and (b) guided filter output .....                                                                                                                                           | 18 |
| Fig. 2.2.3: The resulting gradient map acquired after edge enhancement. (a) original image, and (b) gradient map .....                                                                                                                                                            | 21 |
| Fig. 2.2.4: Visual illustration of the Image Fusion technique [46] .....                                                                                                                                                                                                          | 23 |
| Fig. 2.2.5: The output of the CLAHE enhancement. (a) original image, and (b) CLAHE output.....                                                                                                                                                                                    | 24 |
| Fig. 2.2.6: The final enhanced image from the proposed system. (a) original image, and (b) enhanced image. ....                                                                                                                                                                   | 25 |
| Fig. 2.3.1: The figure displays a comparison of the various state-of-the-art techniques to the proposed method. The proposed method performs better in comparison. ....                                                                                                           | 26 |
| Fig. 2.3.2: The figure displays zoomed-in regions of the results in Fig. 2.3.1.....                                                                                                                                                                                               | 27 |
| Fig. 2.3.3: The figure displays a comparison of the various state-of-the-art techniques to the proposed method. The proposed method performs better in comparison. ....                                                                                                           | 28 |
| Fig. 3.2.1: Examples of visible face images (top row) and their corresponding thermal face images (bottom row) .....                                                                                                                                                              | 33 |
| Fig. 3.3.1: The proposed Thermal Face Emotion Recognition system architecture. The input to the system is thermal images. The network follows a VGG style architecture and an added custom layers. ....                                                                           | 36 |
| Fig. 3.3.2: Example images selected from the Tufts Face Database.....                                                                                                                                                                                                             | 39 |
| Fig. 3.3.3: An illustration of the different facial expressions where (a) a neutral expression, (b) smiling, (c) sleepy, (d) shocked, (e) with sunglasses.....                                                                                                                    | 40 |
| Fig. 3.3.4: Illustrative examples of the cropped thermal image, panel [a, c] are original image, and panel [b, d] are the corresponding cropped image .....                                                                                                                       | 41 |
| Fig. 3.3.5: Illustrative examples of the same expression images obtained after image enhancement- panel [a - e] are original images, guided filtered image, adaptive histogram equalization, intensity adjusted CLAHE, and Bi-Histogram equalization images, respectively.....    | 42 |
| Fig. 3.4.1: Panel [a-d] provides the confusion matrices generated by using (a)TERNet - VGG16 with pretrained VGG-Face, (b) VGG16 with pretrained ImageNet, (c) VGG19 and (d) ResNet18. The x-axis represents the predicted labels and the y-axis represents the true labels. .... | 44 |
| Fig. 4.1.1: The proposed system architecture matches panoramic antemortem and postmortem radiographs. ....                                                                                                                                                                        | 52 |
| Fig. 4.1.2: An illustration of the DoG acquired by subtracting adjacent sub-octaves and detecting the maxima and minima by comparing each point with its neighbors. [132] .....                                                                                                   | 54 |

|                                                                                                                                                                                                                                                                                                                   |    |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Fig. 4.1.3: An illustration of the image gradient and the keypoint descriptors .....                                                                                                                                                                                                                              | 57 |
| Fig. 4.1.4: The figure displays the feature detection and matching process, where (a) represent the tooth images to be matched, (b) features detected on the images using the multiscale Harris detector, (c) matched feature with erroneous matches, and (d) accurate matches after geometric verification. .... | 58 |
| Fig. 4.1.5: Example image pairs of the (a) antemortem and (b) postmortem scans.....                                                                                                                                                                                                                               | 60 |
| Fig. 4.1.6: The figure displays a panoramic radiograph and its isolated tooth. The numbers 1-31 denote the tooth number. ....                                                                                                                                                                                     | 61 |
| Fig. 4.2.1: An example of multimodal images (a) visible image, (b) infrared image, and (c) near infrared image .....                                                                                                                                                                                              | 66 |
| Fig. 4.2.2: The proposed flow diagram of the color transfer system .....                                                                                                                                                                                                                                          | 69 |
| Fig. 4.2.3: The images above depict the image matching results where (a) is the reference image, (b) is the rotated view of the near-infrared image of the same scene, and (c) displays the matched features .....                                                                                                | 70 |
| Fig. 4.2.4: Aligned images (a) reference image (b) target image .....                                                                                                                                                                                                                                             | 71 |
| Fig. 4.2.5: Results obtained after image enhancement. (a) original image (b) enhanced image.....                                                                                                                                                                                                                  | 72 |
| Fig. 4.2.6: Results obtained after image enhancement. (a) original image (b) enhanced image.....                                                                                                                                                                                                                  | 72 |
| Fig. 4.2.7: The images above demonstrates the results of the color transferred images. (a) & (b) reference image, (c) & (d) target image, and (e) & (f) recolored image .....                                                                                                                                     | 74 |
| Fig. 4.2.8: Color correction results. (a) original image, (b) target image, (c) after color correction, and (d) after re-graining. Image source [153]. ....                                                                                                                                                       | 75 |
| Fig. 4.2.9: A comparison of color transfer methods (a) reference image, (b) target image, (c) result obtained by Pitié [153], (d) result obtained by Xiao [152], and (e) proposed method.....                                                                                                                     | 76 |
| Fig. 4.2.10: A comparison of color transfer methods applied to the AMOS dataset image (a) reference image, (b) target image, (c) result obtained by Pitie [153], (d) result obtained by Xiao [152], and (e) proposed method .....                                                                                 | 77 |
| Fig. 4.2.11: A comparison of color transfer methods applied to the RGB-NIR Scene Dataset images (a) reference image, (b) target image, (c) result obtained by Pitie [153], and (d) proposed method .....                                                                                                          | 78 |
| Fig. 4.2.12: An illustration of the image fusion process. (a) visible gray level image, (b) thermal image, and (c) fused image. ....                                                                                                                                                                              | 80 |
| Fig. 4.3.1: Poor visualization of underwater images [177, 178] .....                                                                                                                                                                                                                                              | 83 |
| Fig. 4.3.2: Flow diagram of the underwater image stitching system .....                                                                                                                                                                                                                                           | 84 |
| Fig. 4.3.3: An illustration of the image pre-processing algorithm. (a) original image and (b) enhanced image .....                                                                                                                                                                                                | 86 |
| Fig. 4.3.4: Illustration of feature detection and matching. ....                                                                                                                                                                                                                                                  | 87 |
| Fig. 4.3.5: Generation of seam line using Alpha-winzorized correlation .....                                                                                                                                                                                                                                      | 90 |
| Fig. 4.3.6: Generation of seam line using weighted rank order statistics correlation. ....                                                                                                                                                                                                                        | 90 |



|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Fig. 4.3.7: Intensity differences encountered during image stitching. Image source [191] .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | 91  |
| Fig. 4.3.8: Visual analysis (a), (b) Original images, (c) ground truth, (d) alpha-winzorized result, (e) weighted rank order statistic results. ....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 97  |
| Fig. 4.3.9: Visual analysis (a), (b) Original images, (c) ground truth, (d) alpha-winzorized result, (e) weighted rank order statistic results. ....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 98  |
| Fig. 4.3.10: The results of the stitching algorithm applied to finger images. (a) left image, (b) center image, (c) right image, (d) alpha-winzorized result, and (e) weighted rank order statistics result. ..                                                                                                                                                                                                                                                                                                                                                                                                                              | 101 |
| Fig. 5.3.1: The network architecture of the proposed KTNNet - (a) provides an overall structure of KTNNet with the encoder and a dual decoder network, (b) visualizes the standard residual network proposed by He [243], (c) visualizes the residual network with a channel attention mechanism, (d) visualizes the enhance unit utilized to boost the feature extraction from the encoder network to emphasize more on significant features, and (e) is the legend for this flow diagram. ....                                                                                                                                             | 109 |
| Fig. 5.3.2: Results of the edge enhancement process. (a) original image, (b) gaussian blurred image with sigma = 2, (c) the extracted edge information, and (d) the enhanced image. As seen, the additional edges improve the resolution of the image. ....                                                                                                                                                                                                                                                                                                                                                                                  | 112 |
| Fig. 5.4.1: Example images from the KITS dataset where (a) and (d) are the input CT images, (b) and (e) are the ground truth masks of the Kidney+Tumor regions, and (c) and (f) are the ground truth masks of the Tumor region.....                                                                                                                                                                                                                                                                                                                                                                                                          | 117 |
| Fig. 5.4.2: A visual comparison of the performance of the SOTA methods in segmenting the kidney and tumor regions. (a) test image, (b) ground truth (GT) of the kidney+tumor (KT) region, (h) GT of the tumor (T) region. (c)-(f) are the results produced by the SOTA methods for the KT region and (i)-(l) are the results produced by the SOTA methods for the T region. As shown, KTNNet (g) and (m) in Blue performs better in comparison. ....                                                                                                                                                                                         | 122 |
| Fig. 5.4.3: A visual comparison of the performance of the SOTA methods in segmenting the kidney and tumor regions. (a) test image, (b) ground truth (GT) of the kidney+tumor (KT) region, (h) GT of the tumor (T) region. (c)-(f) are the results produced by the SOTA methods for the KT region and (i)-(l) are the results produced by the SOTA methods for the T region. As shown, KTNNet (g) and (m) in Blue performs better in comparison. ....                                                                                                                                                                                         | 123 |
| Fig. 5.4.4: A visual comparison of the performance of the SOTA methods in segmenting the kidney and tumor regions. (a) test image, (b) ground truth (GT) of the kidney+tumor (KT) region, (h) GT of the tumor (T) region. (c)-(f) are the results produced by the SOTA methods for the KT region and (i)-(l) are the results produced by the SOTA methods for the T region. As shown, KTNNet (g) and (m) in Blue performs better in comparison. A discrepancy in the ground truth image is noticed. The left kidney contains a tumor which is unmarked in the ground truth image. KTNNet was still able to accurately detect the tumor. .... | 124 |
| Fig. 6.3.1: Results of the existing thresholding techniques. (a) test image, (b) Otsu, (c) Niblack, (d) Sauvola, (e) Bernsen, and (f) Wolf .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | 132 |
| Fig. 6.4.1: The flow diagram of the proposed Human Visual Thresholding method .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | 133 |
| Fig. 6.4.2: H&E stained histopathology images demonstrating color variations caused by difference in slide preparation.....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | 134 |

|                                                                                                                                                                                                                                                                                                                |     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Fig. 6.4.3: Illustration of the image standardization process. (a) reference image, (b) target image, and (c) resulting standardized image .....                                                                                                                                                               | 135 |
| Fig. 6.4.4: An illustration of the <i>I<sub>mean</sub></i> image generation. (a) test image and (b) <i>I<sub>mean</sub></i> image.....                                                                                                                                                                         | 136 |
| Fig. 6.4.5: An illustration of the deviation image (a) test image and (b) deviation image.....                                                                                                                                                                                                                 | 136 |
| Fig. 6.4.6: An illustration of the <i>F<sub>2m,n</sub></i> image generation. (a) test image and (b) <i>F<sub>2m,n</sub></i> image.....                                                                                                                                                                         | 137 |
| Fig. 6.4.7: The output of the detailed enhancement. (a) deviation image and (b) enhanced deviation image .....                                                                                                                                                                                                 | 138 |
| Fig. 6.4.8: Thresholded images acquired using HVT method (a) test image (b) Result.....                                                                                                                                                                                                                        | 138 |
| Fig. 6.5.1: Visual comparison of thresholding methods: (a) Original image, (b) Otsu's method, (c) Niblack's method, (d) Sauvola's method, (e) Bernsen's method, (f) Wolf's method, and (g) HVT method.....                                                                                                     | 139 |
| Fig. 6.5.2: Zoomed view of thresholding methods: (a) Original image, (b) Otsu's method, (c) Niblack's method, (d) Sauvola's method, (e) Bernsen's method, (f) Wolf's method, and (g) proposed method. The proposed system isolates each nucleus while perfectly suppressing the background information. ....   | 140 |
| Fig. 6.5.3: Visual comparison of thresholding methods: (a) Original image, (b) Otsu's method, (c) Niblack's method, (d) Sauvola's method, (e) Bernsen's method, (f) Wolf's method, and (g) HVT method.....                                                                                                     | 140 |
| Fig. 6.5.4: Visual comparison of thresholding methods: (a) Original image, (b) Otsu's method, (c) Niblack's method, (d) Sauvola's method, (e) Bernsen's method, (f) Wolf's method, and (g) HVT method.....                                                                                                     | 141 |
| Fig. 7.4.1: Tumor roughness score image analysis. (a) Adaptive segmentation technique developed to segment the kidney and tumor for surface enhancement and investigate tumor irregularity. (b) The summary irregularity score ( $R_{avg} = 1.0413$ ) is averaged across all segments of the tumor. ....       | 150 |
| Fig. 7.5.1: Box and whisker plot of renal mass size compared to the renal tumor subtype for trend. ....                                                                                                                                                                                                        | 151 |
| Fig. 7.5.2: (a) Box and Whisker plot displaying the tumor surface roughness score compared to each renal mass subtype comparison for trend. (b) Waterfall plot for tumor roughness score for each renal cancer tumor subtype.....                                                                              | 152 |
| Fig. 7.5.3: Scatter plot noting the correlation between renal mass size and tumor roughness score (Pearson $p=0.02$ ) .....                                                                                                                                                                                    | 153 |
| Fig. 8.2.1: Three different kinds of dental radiographs. (a) periapical x-ray, (b) bitewing x-ray, (c) panoramic x-ray.....                                                                                                                                                                                    | 162 |
| Fig. 8.3.1: The acquisition system setup. The system consists of an eye-tracker and an audio recording device. ....                                                                                                                                                                                            | 168 |
| Fig. 8.4.1: Illustration of the different imaging components. (a),(g) panoramic radiograph, (b),(h) segmented mask outlining the abnormality, (c),(i) grayscale eye-tracking gaze plot, (d),(j) color quantized eye-tracking gaze plot, (e),(k) teeth mask, (f),(l) maxillomandibular region of interest. .... | 170 |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |     |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Fig. 8.4.2: Logical sequence followed while evaluating each radiograph. It depicts the five different characteristics recorded for a radiograph with abnormalities.....                                                                                                                                                                                                                                                                                                       | 171 |
| Fig. 8.4.3: An illustration of radiographs with their abnormalities labeled. (a) displays apical inflammation changes associated with teeth number 2, 3, and 30, and (b) displays a pericoronal defined corticated radiolucency associated with impacted 32, causing displacement of the tooth on the canal and possible loss of lamina dura and 31 intrauterine cyst odontogenic keratocyst of benign neoplasia there is apical inflammatory changes associated with 18..... | 172 |
| Fig. 8.4.4: Gaze plots generated from the eye-tracking data. (a) radiograph, (b) segmented mask outlining the abnormality, (c) the grayscale gaze plot displaying the expert looking at the abnormality, and (d) color quantized eye-tracking gaze plot. ....                                                                                                                                                                                                                 | 175 |
| Fig. 8.4.5: Structure of the .json file. The file contains information on abnormalities pertaining to each radiograph.....                                                                                                                                                                                                                                                                                                                                                    | 176 |
| Fig. 8.4.6: A sample of the annotated teeth and the region of interest. (a) panoramic radiograph, (b) teeth mask overlaid on the panoramic radiograph, and (c) maxillomandibular region of interest. ....                                                                                                                                                                                                                                                                     | 177 |
| Fig. 8.4.7: Folder structure of the tufts dental database .....                                                                                                                                                                                                                                                                                                                                                                                                               | 178 |
| Fig. 8.5.1: Results of the state-of-the-art enhancement algorithms where (a) input image, (b) HE, (c) CLAHE, (d) BBHE, (e) DF, and (f) Guided filtering.....                                                                                                                                                                                                                                                                                                                  | 181 |
| Fig. 8.5.2: Results of the state-of-the-art enhancement algorithms where (a) input image, (b) HE, (c) CLAHE, (d) BBHE, (e) DF, and (f) Guided filtering.....                                                                                                                                                                                                                                                                                                                  | 182 |
| Fig. 8.6.1: Illustration of the automated annotation process using eye-tracking and AI.....                                                                                                                                                                                                                                                                                                                                                                                   | 186 |
| Fig. 8.6.2: Illustration of an AI system trained to perceive radiographs as humans do .....                                                                                                                                                                                                                                                                                                                                                                                   | 187 |
| Fig. 8.6.3: A training tool that evaluates the findings of novice clinicians against the “expert AI” .....                                                                                                                                                                                                                                                                                                                                                                    | 188 |

## List of Publications

### Patents:

- 1) Panetta, Karen, Kamath K.M, Shreyas, Rao, Shishir Paramathma, Rajeev, Srijith, Rajendran, Rahul, Agaian, Sos S, “Deep perceptual image enhancement”, Utility patent application. Application #63/201,025.
- 2) Panetta, Karen, Rao, Shishir Paramathma, Rajeev, Srijith, Kamath K.M, Shreyas, Rajendran, Rahul, Agaian, Sos S, White, Jessica, “Systems, methods and program products for detecting defects using artificial intelligence analysis of multi-dimensional information data”, provisional patent submitted. Application #63/282,353.
- 3) Panetta, Karen, Rajeev, Srijith, Rajendran, Rahul, Kamath K.M, Shreyas, Rao, Shishir Paramathma, Agaian, Sos S, Ramesh, Aruna, “Methods and apparatus for dental disease detection and isolation”, provisional patent submitted. Application #63/290,315.
- 4) Panetta, Karen, Rajeev, Srijith, Rajendran, Rahul, Rao, Shishir Paramathma, Kamath K.M, Shreyas, Agaian, Sos S, Ramesh, Aruna, “Fusion-based sensing intelligence and reporting”, provisional patent submitted. Application #63/127,140.

### Journal Papers:

- 1) Panetta, Karen, Rahul Rajendran, Aruna Ramesh, Shishir Paramathma Rao, and Sos Agaian. "Tufts Dental Database: A Multimodal Panoramic X-ray Dataset for

- Benchmarking Diagnostic Systems." IEEE Journal of Biomedical and Health Informatics (2021).
- 2) Karen Panetta, Qianwen Wan, Sos Agaian, Srijith Rajeev, Shreyas Kamath K. M, Rahul Rajendran, Shishir Paramathma Rao, Aleksandra Kaszowska, Holly A. Taylor, Arash Samani, and Xin Yuan, "A Comprehensive Database for Benchmarking Imaging Systems," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 3, pp. 509-520, 1 March 2020, doi: 10.1109/TPAMI.2018.2884458.
  - 3) Rajendran, Rahul, Kevan Iffrig, Deepak K. Pruthi, Allison Wheeler, Brian Neuman, Dharam Kaushik, Ahmed M. Mansour, Karen Panetta, Sos Agaian, and Michael A. Liss. "Initial Evaluation of Computer-Assisted Radiologic Assessment for Renal Mass Edge Detection as an Indication of Tumor Roughness to Predict Renal Cancer Subtypes." Advances in Urology 2019 (2019).

#### Conference Papers:

- 1) KM, Shreyas Kamath, Rahul Rajendran, Qianwen Wan, Karen Panetta, and Sos S. Agaian. "TERNet: A deep learning approach for thermal face emotion recognition." In Mobile Multimedia/Image Processing, Security, and Applications 2019, vol. 10993, p. 1099309. International Society for Optics and Photonics, 2019.
- 2) Rajendran, Rahul, Thaweesak Trongtirakul, Karen Panetta, and Sos Agaian. "A pixel-based color transfer system to recolor nighttime imagery." In Mobile Multimedia/Image Processing, Security, and Applications 2019, vol. 10993, p. 109930N. International Society for Optics and Photonics, 2019.

- 3) KM, Shreyas Kamath, Rahul Rajendran, Karen Panetta, and Sos Agaian. "A human visual based binarization technique for histological images." In Mobile Multimedia/Image Processing, Security, and Applications 2017, vol. 10221, pp. 124-133. SPIE, 2017.
- 4) Rao, Shishir Paramathma, Rahul Rajendran, Karen Panetta, and Sos S. Agaian. "Combined transform and spatial domain based “no reference” measure for underwater images." In 2017 IEEE International Symposium on Technologies for Homeland Security (HST), pp. 1-7. IEEE, 2017.
- 5) Rajendran, Rahul, Shishir P. Rao, Karen Panetta, and Sos S. Agaian. "Adaptive alpha-trimmed correlation based underwater image stitching." In 2017 IEEE International Symposium on Technologies for Homeland Security (HST), pp. 1-7. IEEE, 2017.

# **Chapter 1. Introduction**

## **1.1. Overview**

Artificial Intelligence (AI) is one of the most prolifically growing areas spanning into every section from medicine to self-driving vehicles. AI has demonstrated a capacity to process and analyze massive amounts of data to help humans explore, explain, and expand our capabilities to understand the world around us. The significance of AI is likely to grow and more diverse sectors will use it as a foundational tool to process and analyze the massive amounts of disparate data.

Over the last few years, significant advancements in AI have been made in the field of medicine and forensic science [1-4]. In particular, deep learning (DL), a subset of AI, has been extensively used for analyzing and extracting valuable information from images. With the availability of increased computation power, DL-based algorithms are best suited for performing complex tasks and working with large amounts of unstructured data [5].

In the field of medical imaging, different imaging modalities, such as CT, MRI, X-ray, and ultrasound, are used by physicians and radiologists to detect, diagnose, and treat a wide range of disorders that occur in various sections of the human body. Imaging diagnosis involves the visual perception of images and the application of pattern recognition to differentiate normal from abnormal findings, followed by detailed characterizations of these findings that would allow for interpretation [6]. The interpretation thus derived enables appropriate patient care and management. All radiological investigation necessitates the interpretation by the radiologist to arrive at a diagnosis. Although efficient, human inspection requires specialized training and a

radiologist's time, which are increasingly expensive [7]. Human examination gives a subjective judgment, which may fluctuate from one practitioner to another, which results in a degree of uncertainty in quantitative estimation. Furthermore, inspection results could be affected by factors such as fatigue and distraction due to cognitive overload.

In forensics, identifying specific patterns in large amounts of data is critical. This could involve tasks ranging from analyzing crime scenes to recognizing patterns in texts. Traditional forensic analysis relies on forensic experts manually extracting information from crime scene evidence [8]. Not only is this time-consuming, but also complicated by subjective factors. Human error in forensic testing is one of the most common problems that forensic scientists face, and these errors frequently lead to incorrect results [9, 10]. The advent of DL systems offers the prospects of more accurate and more efficient results in forensic testing.

Despite these potentials, DL-based solutions have not yet become the norm in routine medical and forensic practices. To work as intended, DL algorithms require access to a continued supply of data for training, testing, and validating. To ensure widespread adoption, data may need to be shared across multiple institutions and potentially across nations. However, due to privacy concerns and legal restrictions, it can be difficult to access some of this data [10]. The quality of data plays a significant role in developing accurate DL models. The performance of the algorithm is highly contingent on the quality of the training data; low-quality data will result in low-quality outcomes. To parse low-quality data and generate comparable results, advanced and deeper models are required. Such models are computationally expensive to train, requiring significant memory and computational resources.



There are two well-known approaches that can be used to address the limited availability of data. The first entails applying the transfer-learning concept. Transfer learning is used to enhance the performance of a model or a system in a given domain by transferring information from a related domain [11]. While the transferred data does not directly augment the real data, it will aid in enhancing the original data input representation and the mapping function. The second approach involves synthetic data generation. Synthetic data allows you to create artificial data from scratch or use complex data manipulation techniques to create new and unique training instances. This approach is particularly useful for enriching data and eliminating bias.

In addition to the issues mentioned above, AI is also prone to cognitive bias, just like the human brain. Human cognitive biases are heuristics or mental shortcuts that influence decision-making and thinking, resulting in errors in reasoning. Stereotyping, the bandwagon effect, confirmation bias, priming, and selective perception are some examples of cognitive biases [12]. The inherent cognitive biases of the human programmer, the size, collection methodology, source of data, algorithmic designs, and manual human adjustments are some of the ways in which human cognitive bias is transferred to AI. Existing strategies for overcoming bias rely mostly on changing the dataset by adding more samples, enhancing the input data through improved pre-processing, including human-in-the-loop, and providing algorithmic confidence alongside the prediction to assist humans in determining how much emphasis should be given to the prediction. However, these methods necessitate a large workforce to program the algorithms and collect new data to be implemented. Furthermore, in the current AI framework, an AI system is trained to perform a single task, most often utilizing a single form of input data, and the entire system is activated to perform the task. As a result, the AI framework will require a large and specialized dataset,

consume more time to train each task, and consume massive amounts of computer power and energy.

Researchers are striving to create AI models that can match or even surpass human capabilities. To meet this expectation, it is crucial to have accurate data to develop such models that can mimic the behavior of a human expert. AI has demonstrated a particularly impressive ability to recognize patterns in data through correlation. But such models are fundamentally blind to the cause and the effect. For example, unlike a human doctor, AI algorithms cannot explain why a particular image may suggest a disease. Furthermore, human involvement is pivotal in the field of medicine; however, current state-of-the-art systems are limited to visual imagery alone, and humans turn into passive observers. To truly mimic an expert, the AI system must consider the visual perception and cognition of the human.

This dissertation addresses these shortcomings by promoting the development of high-performance, low-cost AI-based architectures for medical and forensic applications. In particular, this dissertation focuses on addressing the following questions:

- Can the enhancement and augmentation of low-quality data improve the performance of AI systems?
- Is it possible to reduce the number of parameters in a neural network while increasing its accuracy and robustness?
- Can the integration of visual perception and cognition of the expert into current architectures demystify the “black-box” element of AI?

In addressing these questions, this dissertation explores and demonstrates the importance of fundamental elements that aid in developing reliable AI systems. The dissertation focuses on:

- Developing algorithms and metrics to improve and analyze poor/low-quality images for better human perception and improved detection.
- Creating high-performance, lightweight neural networks for real-time applications tailored to image interpretation and deployment on edge devices.
- Enabling AI systems to combine multimodal data and generate more explainable models while maintaining a high level of performance.

## 1.2. Objectives

The overall objective of this dissertation is to develop and investigate novel methods and tools that address the aforementioned questions by demonstrating their effectiveness in medical and forensic applications, including feature enhancement and detection, segmentation, and classification.

Following are some of the tasks that were performed to accomplish the goals of this dissertation:

1. Develop an edge-aware enhancement algorithm that enhances the quality of images and augments datasets to improve the performance of DL systems.
2. Develop a scale and rotation invariant image stitching algorithm that utilizes state-of-the-art feature detection and matching algorithms to generate panoramic images.
3. Develop a lightweight, high-performance Convolutional Neural Network architecture that incorporates classical image processing techniques to improve the image segmentation accuracy and deploy such models on edge devices.

4. To implement an adaptive thresholding technique for color and grayscale images to differentiate between foreground and background segments. This technique can be utilized to annotate images and generate ground truths.
5. To introduce the first-of-its-kind comprehensive multimodal dataset for benchmarking diagnostic systems and creating novel multimodal and explainable AI systems.

### 1.3. Contributions

1. **Feature Enhancement:** The quality of data is of great importance in both medical and forensic applications. Images in these domains are often accompanied by noise and unwanted artifacts, making it hard for the observer to analyze and discern certain features. With the focus now shifting from human-based analysis to AI-based analysis, the quality of data will play a significant role. Furthermore, AI systems usually require large amounts of training data. With privacy concerns and legal restrictions, the lack of data can prove to be a bottleneck in developing efficient AI systems. This dissertation presents a novel edge-aware image enhancement algorithm to improve the quality of images. The algorithm was utilized to enhance panoramic radiograph x-ray images. In addition, state-of-the-art image enhancement techniques are employed to generate synthetic data to develop and improve a thermal emotion recognition system.
2. **Development of a scale and rotation invariant image stitching algorithm:** Two adaptive correlation techniques are developed to generate a high-resolution panoramic view of the environment. The performance of the image stitching algorithms was tested on underwater images. In addition, a new technique for underwater image quality measure is

proposed by combining the spatial and transform domain approaches based on the human visual system.

3. **Development of a dual decoder-based CNN architecture:** a new 2D CNN architecture-KTNet is developed for image segmentation, which utilizes the features obtained from edge enhanced images. The network comprises of a dilated ResNet encoder architecture and a novel dual decoder branch that was designed to individually segment the kidney and tumor regions.
4. **Multimodal Dataset:** The Tufts Dental Database, a new X-ray panoramic radiography image dataset, has been presented. This dataset consists of 1000 panoramic dental radiography images with expert labeling of abnormalities and teeth. This first-of-its-kind multimodal dataset also includes the radiologist's expertise, captured in the form of eye-tracking and audio. The comprehensive dataset will serve as a foundation for developing the next generation of multimodal and explainable AI systems.

## 1.4. Organization

Fig. 2.1.1 provides a visual overview of the dissertation. The dissertation is divided into two main sections: feature engineering and image analytics. Each section is further broken down into chapters that are tailored to specific tasks and applications.

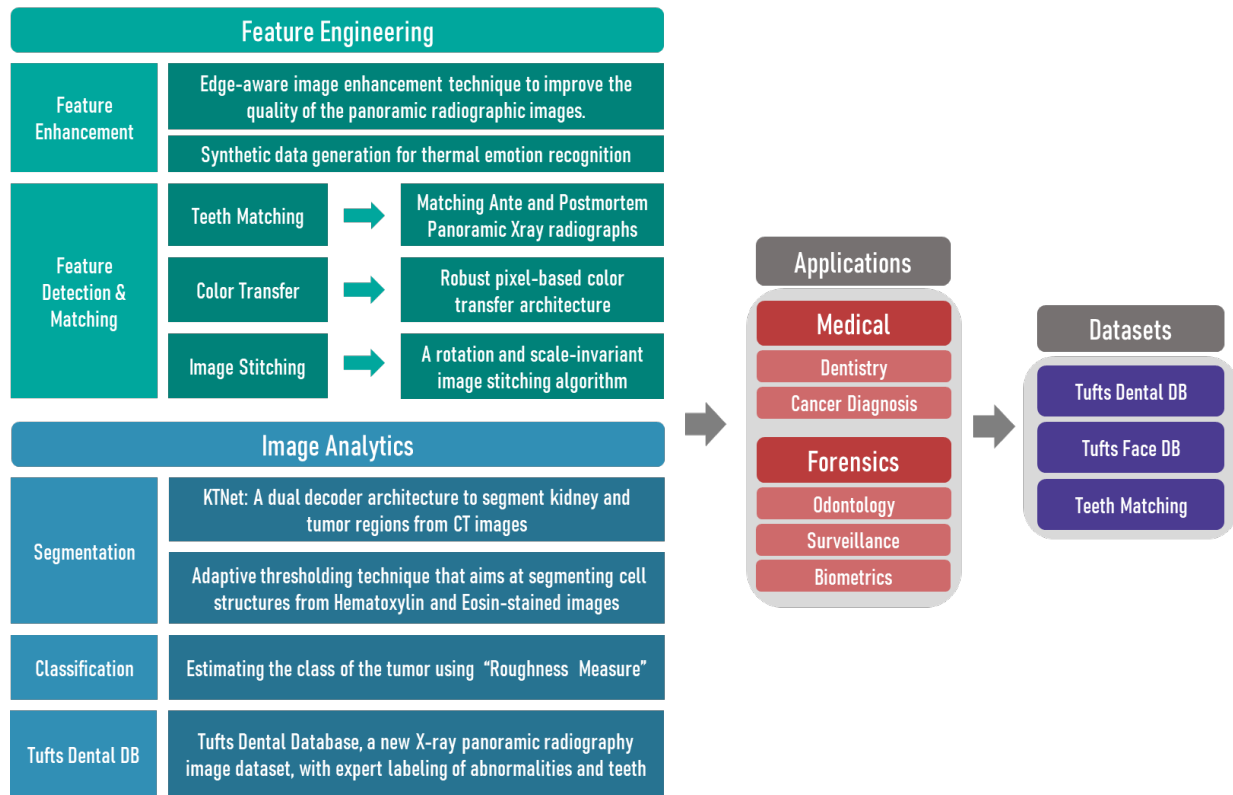


Fig. 2.1.1: A visual overview of the dissertation. The focus of the dissertation is to develop novel methods and systems for applications in medical and forensic sciences.

## **Feature Engineering**

*Feature enhancement:* Chapter 2 begins with a brief introduction to the importance of medical image enhancement and a literature survey of the current medical image enhancement techniques. The chapter presents a novel edge-aware image enhancement tool to improve the quality of panoramic radiographic X-ray images (*Objective 1*). This enhancement is designed to eliminate the noisy components of the image while simultaneously improving its contrast. Finally, the performance of the enhancement technique is quantitatively evaluated and compared with various state-of-the-art methods.

*Feature enhancement:* Chapter 3 presents a robust thermal emotion recognition system – TERNet. The customized architecture adopts features obtained via transfer learning from the VGG-Face CNN model, which is further fine-tuned with the thermal expression face data from the TUFTS face database [13]. The chapter provides further insights on the applications of feature enhancement in generating synthetic data for training and validation (*Objective 1*).

*Feature Detection and Matching:* Chapter 4 provides an overview of the state-of-the-art feature detection and matching algorithms. These algorithms were utilized for two specific tasks: (1) to match antemortem and postmortem dental radiographs and (2) to develop a color transfer algorithm. In addition, a scale and rotation invariant image stitching algorithm is proposed to generate a panoramic view of the underwater environment. The algorithm is further evaluated by testing its robustness in different applications (*Objective 2*).

### **Image Analytics**

*Segmentation:* Chapter 5 presents a novel 2D CNN architecture to segment kidney and tumor regions from Computed Tomography (CT) images. The network comprises of a Dilated ResNet encoder architecture and a novel dual decoder branch. In addition, features were extracted from edge-enhanced images to boost the performance of the system (*Objective 3*).

*Segmentation:* Chapter 6 proposes a new local human visual-based thresholding technique dedicated to nuclei segmentation from H&E stained images (*Objective 4*). H&E images are stained with specific colors to differentiate each component in the tissue. Segmentation and analysis of such images are problematic, as they present high color and cell clusters variability. The chapter also provides a visual comparison of the proposed algorithm with existing thresholding techniques.

*Classification:* Chapter 7 proposes a system to assess the aggressiveness of an incidentally detected renal mass using CT images. A tumor roughness score is employed to categorize the type of tumor. Finally, linear regression analysis noted that the roughness score is the most influential on the model, with all other demographics being equal, including tumor size.

*Tufts Dental Database:* Chapter 8 introduces a novel multimodal panoramic x-ray dataset for benchmarking diagnostic systems. This multimodal dataset consists of a tooth segmentation mask, abnormality mask, maxillomandibular region of interest mask, eye-tracking gaze map, and text description of the abnormality. The dataset opens the opportunity to seamlessly integrate the radiologist's expertise into the AI system, making it more efficient and reliable. It also paves the way for a new generation of AI systems to highlight the identified features and explain the reasons for the outputs generated (*Objective 5*).

Chapter 9 summarizes the dissertation and provides insights into future research directions and opportunities that build on the ideas presented in this thesis.



## **Section 1: Feature Engineering**

## **Chapter 2. Image Enhancement**

### **2.1. Introduction**

Medical imaging systems commonly utilize image enhancement to better visualize images of the human body and its organs. Good quality images help medical professionals in detecting and diagnosing irregularities or abnormalities. Acquiring accurate and consistent images and enabling accurate observations is challenging in medical image acquisition. Many previous efforts and ongoing research have been directed toward the problem of improving the distinguishability of significant features in medical images to simplify the diagnosis process for a given application. [14, 15]. Various factors such as restricted image dimension, pixel resolution, and the compulsive effects caused due to noise are deemed as key components for poor visualization of images. Removing or suppressing noise from medical images is an arduous and delicate task. To minimize noise while retrieving diagnostically relevant image content, a compromise must be made between preserving the image features and retrieving diagnostically relevant image content.

Image enhancement aims to improve the quality of images for subsequent analysis or image display. Image enhancement includes contrast enhancement, edge enhancement, noise filtering, sharpening, and interpolating [16, 17]. There is no known general theory of image enhancement. When an image is processed for visual interpretation, the ultimate judge of how well a method works is the expert looking at the image. The visual evaluation of image quality is highly subjective. The definition of a “good image” becomes an elusive standard used to evaluate an algorithm’s performance [18]. When the problem is processing images for machine perception, this evaluation task is somewhat easier. For instance, in a character recognition application, the

optimal image processing method would be the one that produces the best machine recognition results, disregarding other factors such as computational requirements. However, even when a clear performance criterion can be imposed on the problem, a certain amount of trial and error is usually required before selecting an image enhancement approach.[18].

Image enhancement can be broadly classified into two categories: spatial domain enhancement and frequency domain enhancement [19]. In the spatial domain, pixel values may be modified based on rules that depend on the original pixel value (local or point processes). Another class of operations combines or compares pixel values to others in their local neighborhood in various ways [20]. Spatial domain methods operate directly on the pixels of the image and are denoted by

$$g(a, b) = L[f(a, b)] \quad \text{Eq. 2.1.1}$$

where  $f(a, b)$  is the input image,  $g(a, b)$  is the processed image, and  $L$  is an operator on  $f$ , defined over some neighborhood of  $(a, b)$ . In addition,  $L$  can operate on a set of images, such as performing a pixel-by-pixel sum of  $K$  images for noise reduction.

Frequency domain techniques manipulate the image's orthogonal transform rather than the image itself. Frequency domain techniques are well suited for processing the image based on the frequency content [18]. The frequency-domain enhancement involves computing a 2-D discrete unitary transform of the image, such as the 2-D DFT, manipulating the transform coefficients with an operator  $M$ , and then performing the inverse transform. The orthogonal transform of the image has two components: magnitude and phase. The phase is used to restore the image back to the spatial domain. The transform domain enables operations on the image's frequency content; therefore, high-frequency content such as edges and other subtle information can easily be

enhanced. The edges and sharp transitions (e.g., noise) in an image contributes significantly to the high-frequency content of the Fourier transform. Whereas, the low-frequency contents in the Fourier transform are responsible for the general appearance of the image over smooth areas. The concept of filtering is easier to visualize in the frequency domain. Therefore, enhancement of image  $f(x, y)$  can be done in the frequency domain based on the Discrete Fourier Transform. This is particularly useful in convolution if the spatial extent of the point spread sequence  $h(x, y)$  is larger than the convolution theory [21].

$$G(x, y) = h(x, y) * f(x, y) \quad \text{Eq. 2.1.2}$$

where  $G(x, y)$  is the enhanced image.

Panoramic dental radiography is one of the most commonly used medical imaging techniques, both by general practitioners and dental specialists [22-24]. Various factors, such as the device used, the acquisition technique, and the subsequent processing of the acquired images, can influence the quality of the image [25, 26]. The major limitation of panoramic radiographs is their complex representation; a three-dimensional structure is represented in a two-dimensional plane, superimposing anatomical images.

Contrast enhancement is a common technique applied to enhance radiographs and assist radiologists in making better assessments. Many contrasts, brightness, detail, and edge enhancement algorithms have been proposed [27]. In particular, histogram-based algorithms are commonly used in medical image enhancement. There are many variants of histogram-based algorithms, and they have been used to improve different types of images [27-32]. Veska [33] proposed an effective X-ray image enhancement approach to detect dental caries using Contrast Limited Adaptive Histogram Equalization (CLAHE), followed by morphological operations. In

addition, homomorphic filtering was used to eliminate the nonuniformity luminance distribution in the image. Rahmi-Fajrin [32] investigated how contrast enhancement results influence the dentist's evaluation. Shreyas [34] compared the performance of the Particle Swarm Optimization, Cuckoo Search Algorithm, and Artificial Bee Colony optimization to enhance periapical dental x-rays. Chen [35] proposed a multiscale feature sensing network to enhance the quality of low-dose dental CT images under different downsampling methods. Suprijanto [36] evaluated the contrast quality of a digital image that was scanned using transmission and reflection mode based on the mean and standard deviation of the image. In addition, algorithms such as contrast stretching, histogram equalization, adaptive histogram equalization, and CLAHE were employed to enhance the panoramic radiographs. YangDai [37] proposed an enhancement method based on Single Scale Retinex. The image was divided into different regions using the knowledge of the dental CT's histogram. Finally, a weighted Single Scale Retinex method is employed with a stronger enhancement in the region of interest and a relatively weaker enhancement in the other parts. Khairunnisa [38] used a fuzzy grayscale image enhancement technique to develop an algorithm for enhancing low contrast images. Low contrast images were degraded due to insufficient lighting during image capture. By maximizing fuzzy events contained in the image, the fuzzy grayscale image enhancement method is proposed. The membership function is then tweaked using the power-law transformation and the saturation operator to improve the image quality. Uday Kumbhar [39] proposed a method to enhance the medical images by using a median filter for removing noise from the images followed by unsharp mask filters and CLAHE.

This chapter proposes an edge-preserving image enhancement system to enhance dental panoramic radiographs. The performance of the proposed system is visually evaluated and

compared to the state-of-the-art enhancement algorithms such as Histogram Equalization, CLAHE, Dynamic Fuzzy Histogram Equalization, and Bi-Histogram equalization.

## 2.2. Proposed Work

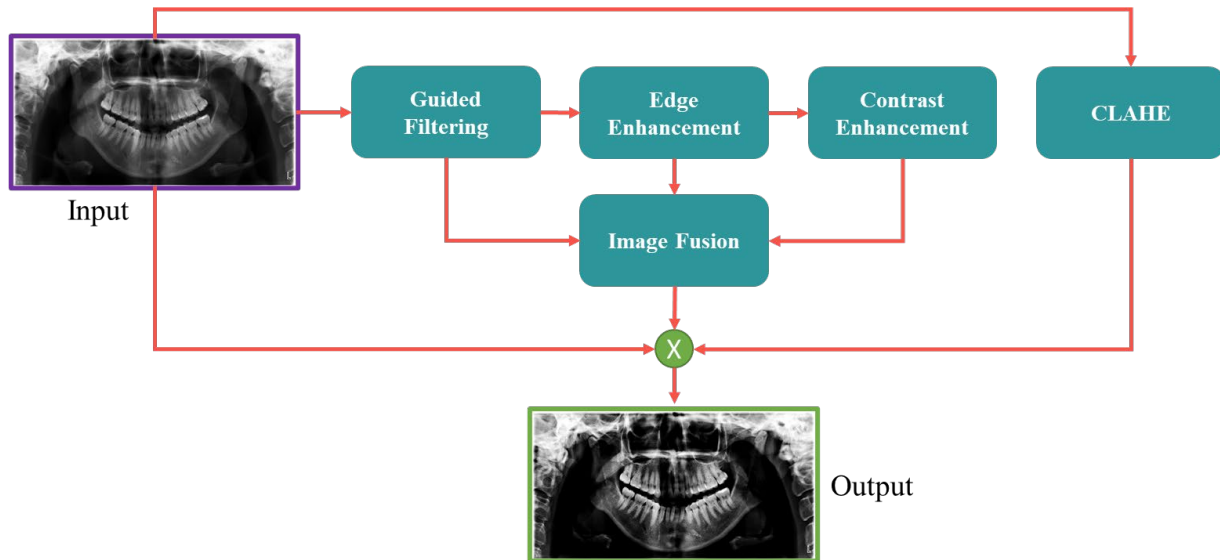


Fig. 2.2.1: A workflow of the proposed edge-aware enhancement system

This section provides a detailed description of the enhancement system. The flow diagram of the process is illustrated in Fig. 2.2.1. The system encompasses image processing techniques such as guided filtering, edge enhancement, contrast enhancement, and image fusion. The system also employs CLAHE to enhance certain radiograph features. Each component is further explained in this section.

### 2.2.1. Guided Filtering

In image processing, guided filtering [40] has become increasingly popular in eliminating noise. This filtering technique has a wide range of applications, including image fusion, colorization, image matting, and upsampling [41]. The guided filter outperforms other filters because of its non-approximate implementation, which results in superior image quality. Furthermore, this edge-preserving filter is computationally less expensive due to its invariance to the size of the filter [42]. Theoretically, a guided filter is defined as

$$G_i = \alpha_k I_i + \beta_k, \quad \forall i \in \omega_k \quad \text{Eq. 2.2.1}$$

where  $\omega_k$  is an odd-size square window. The linear coefficients  $\alpha_k$  and  $\beta_k$  are constant and are defined as

$$\alpha_k = \frac{\frac{1}{|\omega|} \sum_{i \in \omega_k} I_i P_i - \mu_k \dot{P}_k}{\delta_k + \varepsilon} \quad \text{Eq. 2.2.2}$$

$$\beta_k = \dot{P}_k - \alpha_k \mu_k \quad \text{Eq. 2.2.3}$$

where  $\delta_k$  and  $\mu_k$  represents the variance and mean of  $I$  in  $\omega_k$  respectively,  $|\omega|$  depicts the number of pixels in  $\omega_k$ , and  $\dot{P}_k$  is the mean of  $P$  in  $\omega_k$ . A pixel  $i$ , involved in all overlapping windows  $\omega_k$  that covers  $i$ , is not identical when computed in different windows. Therefore, all possible values of  $G_i$  are averaged. Hence the guided filter is defined as:

$$G_i = \alpha_i I_i + \beta_i \quad \text{Eq. 2.2.4}$$

where  $\alpha_i = \frac{1}{|\omega|} \sum_{k \in \omega_i} \alpha_k$  and  $\beta_i = \frac{1}{|\omega|} \sum_{k \in \omega_i} \beta_k$ .

For enhancing panoramic radiographs, the guided filter functions as an edge-preserving smoothing operator where the guide is similar to the filtering input image. Finally, the enhanced edge preserved output is obtained by

$$I_{enhanced} = (I_{original} - G_i) * \lambda + G_i \quad \text{Eq. 2.2.5}$$

where  $\lambda$  is a constant.  $\lambda$  was set to 2 in this case, and the results of the guided filter can be visualized in Fig. 2.2.2.



(a)



(b)

Fig. 2.2.2: The figure displays the result of the guided filtering process where (a) original image, and (b) guided filter output



### 2.2.2. Edge Enhancement

Edges carry essential structural information for object classification [43] and are a growing area of research in computer vision [43]. Additionally, edge detection has been widely used in medical imaging, object tracking, computer-aided detection systems, biometrics, remote sensing, industrial machine vision and inspection systems, and geophysical applications [43-45]. In general, edge detection kernels are separable in the sense that they smooth the image in one direction while distinguishing it in the other. As a result, they can be characterized by a low and high pass filter pair. The concept of low pass and high pass filters, which are used to acquire edge information, is examined and used to generate filters for any specification. Consider two kernels,  $K_x$  and  $K_y$ , that detect edges along the x and y directions. These kernels are defined as

$$K_x = \begin{bmatrix} L \\ P \\ F \end{bmatrix} \begin{bmatrix} H & P & F \end{bmatrix} \quad \text{Eq. 2.2.6}$$

$$K_y = \begin{bmatrix} H \\ P \\ F \end{bmatrix} \begin{bmatrix} L & P & F \end{bmatrix} \quad \text{Eq. 2.2.7}$$

where HPF and LPF are classical high and low pass filters, respectively. As these filters are separable, various kernels can be generated by simple permutation and combination.

As edge detection is an essential stride in computer vision, it is necessary to acquire the true edges to get the best results from the matching process. It is essential to choose edge detectors that fit best to the application. As shown in Eq. 2.2.8, a 7x7 truncated kernel is employed to enhance panoramic radiographs. The kernel is normalized by dividing it by the absolute value of the total sum. The kernel is applied to the image by sliding it across its entirety and the gradient

map is acquired. The kernel is further rotated by 45, 90, and 135 degrees to detect edges in those directions. Finally, the directional gradient maps are added together.

$$\begin{bmatrix} 1 & 1 & 1 & 0 & -1 & -1 & -1 \\ 1 & 2 & 2 & 0 & -2 & -2 & -1 \\ 1 & 2 & 3 & 0 & -3 & -2 & -1 \\ 1 & 2 & 3 & 0 & -3 & -2 & -1 \\ 1 & 2 & 3 & 0 & -3 & -2 & -1 \\ 1 & 2 & 2 & 0 & -2 & -2 & -1 \\ 1 & 1 & 1 & 0 & -1 & -1 & -1 \end{bmatrix} \quad \text{Eq. 2.2.8}$$

The acquired gradient map is passed through a series of morphological filters, and the final edge map is obtained. For the filtering process, top and bottom hat filters were utilized. Using a suitable structuring element, these filters can extract dark or bright features from an image [46]. The top-hat and bottom-hat filtering function is defined as

$$\begin{aligned} TopHat(I, s) &= (I \cdot s) - I \\ BottomHat(I, s) &= I - (I \circ s) \end{aligned} \quad \text{Eq. 2.2.9}$$

where  $I$  is the grayscale image and  $s$  is the structuring element. A diamond shape structuring element is employed to obtain clear and prominent edges. The final gradient image is generated as formulated in Eq. 2.2.10. The gradient map obtained is further subjected to thresholding. The thresholding value is selected based on the application. For enhancing panoramic radiographs, the threshold value was set to 0.1. The acquired edges are visualized in Fig. 2.2.3.

$$\begin{aligned} Gradient\ Image &= (Original\ Image + Tophat\ filtered\ image) - \\ &\quad Bottomhat\ filtered\ Image \end{aligned} \quad \text{Eq. 2.2.10}$$



(a)



(b)

Fig. 2.2.3: The resulting gradient map acquired after edge enhancement. (a) original image, and (b) gradient map

### 2.2.3. Contrast Enhancement

The contrast enhancement technique is employed to improve the image's contrast by altering the distribution of the intensity values. It is a vital step to image enhancement as it increases the dynamic range of the pixel's intensity, thereby producing an image that displays

enhanced contrast between the features of interest. Unlike histogram equalization, which attempts to modify the intensity values of all the pixels in the image such that the histogram is "flattened", the contrast enhancement technique linearly maps the input intensity values to the output intensity values. The contrast stretching function is defined as shown in Eq. 2.2.11, where  $\alpha = 100$ ,  $\beta = 20$ , and  $\gamma = 50$  are the centers of the transition area.

$$y(I, \alpha, \beta, \gamma) = \frac{1}{1 + e^{-0.5(I-\alpha)}} - \frac{1}{1 + e^{-0.5(I-\beta)}} + \frac{1}{1 + e^{-0.5(I-\gamma)}} \quad \text{Eq. 2.2.11}$$

#### 2.2.4. Image Fusion

The term "image fusion" refers to the process of integrating key characteristics from two or more images into a single image. This procedure can be carried out at three different levels of processing: pixel, feature, and decision levels of representation [47]. Image level fusion is the lowest preprocessing level in which the result is a gray image obtained by merging pixels. This fusion technique is more helpful for human visual or machine perception [47]. Fusion at the feature level necessitates feature extraction from the input images. These properties include edges, regions, shape, size, and length, all of which are environment dependent. Decision level fusion represents a method that combines the results from several algorithms to yield an output image. Several algorithms are executed on the input images in this fusion technique to extract information individually. Once obtained, they are merged by applying decision rules to reinforce common interpretation, resolve differences, and better understand the observed objects. A pixel-level fusion is utilized to blend the outputs at each stage for this implementation. The working of the image fusion technique is depicted in Fig. 2.2.4. A detailed description of the fusion technique can be found in [46].

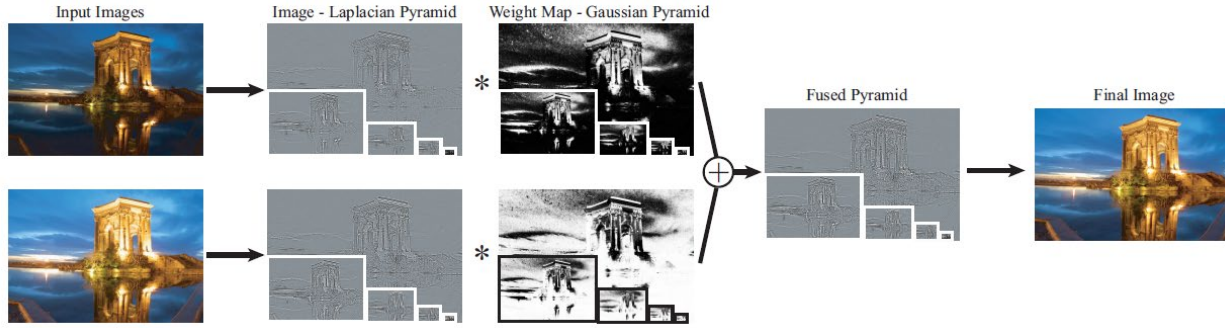


Fig. 2.2.4: Visual illustration of the Image Fusion technique [46]

### 2.2.5. Contrast Limited Adaptive Histogram Equalization

The CLAHE algorithm prevents saturation by limiting the slope associated with the gray level assignment scheme [48]. CLAHE is an improvement over the Adaptive Histogram Equalization technique, as it avoids noise amplification, is computationally fast, and controls the brightness saturation.

The enhancement is performed by dividing the image into non-overlapping regions. The histogram is estimated for each region. A clipping limit is set for the histogram that determines the degree of contrast enhancement. The clipping limit ranges from 0 to 1. The clipping limit  $\delta$  is defined as

$$\delta = \frac{M}{N} \left[ 1 + \frac{\alpha}{100} (s_{max} - 1) \right] \quad \text{Eq. 2.2.12}$$

where  $N$  is the maximum number of gray levels, and  $\alpha$  is the clip factor. The default clipping factor was set to 0.01. The clipped pixel points are evenly distributed to the histogram of each region.



(a)



(b)

Fig. 2.2.5: The output of the CLAHE enhancement. (a) original image, and (b) CLAHE output

Assume  $H(i)$  is the original histogram and  $N_{avg}$  is the average number of pixels, then the redistributed histogram is defined as

$$H(i)_{equalized} = \begin{cases} \delta & H(i) > \delta \\ \delta & H(i) + N_{avg} > \delta \\ H(i) + \delta & else \end{cases} \quad \text{Eq. 2.2.13}$$

Each histogram's redistribution may necessitate several iterations, with the number of iterations increasing as the percent clip factor decreases. The result of the CLAHE algorithm is

shown in Fig. 2.2.5. Finally, the original, fused, and CLAHE images are combined as defined in eq.  $\alpha$ ,  $\beta$ , and  $\gamma$  are set constants, and  $\alpha + \beta + \gamma = 1$ . For this application,  $\alpha$ ,  $\beta$ , and  $\gamma$  were set to 0.1, 0.2, and 0.7, respectively. The final enhanced out is shown in Fig. 2.2.6.

$$\text{Enhanced Image} = \alpha * \text{original} + \beta * \text{CLAHE} + \gamma * \text{Fused Image} \quad \text{Eq. 2.2.14}$$



(a)



(b)

Fig. 2.2.6: The final enhanced image from the proposed system. (a) original image, and (b) enhanced image.

## 2.3. Experimental Results

The performance of the proposed enhancement system is visually evaluated by comparing it to state-of-the-art image enhancement techniques that are utilized to enhance medical images, such as Histogram Equalization (HE), CLAHE, Dynamic Fuzzy Histogram Equalization (DFHE), and Bi-Histogram equalization (BBHE).



(a) original



(b) HE



(c) CLAHE



(d) DFHE



(e) BBHE



(f) Proposed Method

Fig. 2.3.1: The figure displays a comparison of the various state-of-the-art techniques to the proposed method. The proposed method performs better in comparison.



The comparison results are displayed in Fig. 2.3.1 and Fig. 2.3.3. On analyzing Fig. 2.3.1, HE over-enhanced the image, where the intensity of the entire image was amplified. CLAHE was able to improve the contrast of the image but over-enhanced the noisy components and the brighter regions of the image. BBHE and DFHE marginally enhanced the contrast but introduced additional artifacts to the image. The proposed method effectively enhanced the contrast of the image while suppressing the noise and preserving the details. Fig. 2.3.2 displays the zoomed view of a region in Fig. 2.3.1. The figure shows that the proposed method enhances the edge and texture information while preserving essential features. The red box displays the loss of texture information in the state-of-the-art techniques, where the difference in texture between the enamel and dentin layer is lost. This is well preserved in the proposed method. These improvements can also be visualized in Fig. 2.3.3.

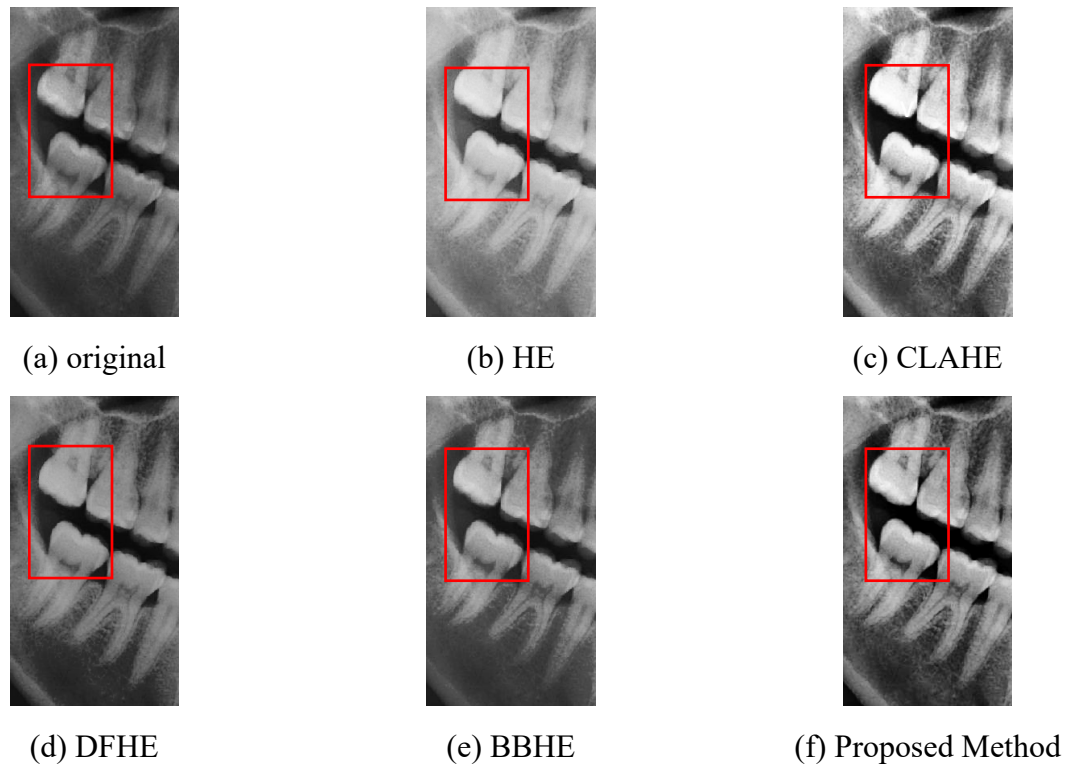
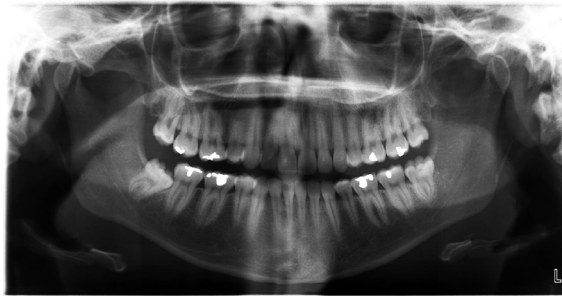


Fig. 2.3.2: The figure displays zoomed-in regions of the results in Fig. 2.3.1



(a) original



(b) HE



(c) CLAHE



(d) DFHE



(e) BBHE



(f) Proposed Method

Fig. 2.3.3: The figure displays a comparison of the various state-of-the-art techniques to the proposed method. The proposed method performs better in comparison.

## 2.4. Summary

The chapter presents an edge-aware image enhancement system that to improve the quality of panoramic radiographs. The system employs various image processing techniques such as

guided filtering, edge enhancement, contrast enhancement, and image fusion. The system employs a new directional edge filter and a contrast enhancement component that highlights the regions of interest and reduces the noise in the image. The proposed system was visually compared to the existing enhancement algorithms, generating sharper and well enhanced images.

## **Chapter 3. Synthetic Data Generation for Thermal Emotion Recognition**

### **3.1. Motivation**

Facial emotion recognition technology has a wide range of practical applications in virtual learning, cognitive psychology analysis, avatar animation, neuromarketing, human-machine interactions, and entertainment systems. Most state-of-the-art emotion recognition systems primarily focus on information in the visible spectrum. The task of identifying emotions can be challenging, as emotions vary from individual to individual. Furthermore, visible images are affected by changes in lighting conditions. Factors such as low lighting, pose variation, aging, and disguise all significantly impact the appearance of images and texture information. Despite significant progress in the field, existing systems for facial emotion recognition lack precision when compared to human performance. To overcome these challenges, thermal imaging systems are replacing visible imaging systems. Thermal images have the following advantages: (a) they are less sensitive to changes in lighting conditions, (b) they have consistent thermal signatures, and (c) they have a temperature distribution formed by the face vein branches. This chapter introduces a robust emotion recognition system using thermal images- TERNet. A customized CNN is employed to accomplish this, which possesses excellent generalization capabilities. The architecture adopts features obtained via transfer learning from the VGG-Face CNN model, which is further fine-tuned with the thermal expression face data from the Tufts Face Database.

### 3.2. Introduction

The majority of off-the-shelf deep neural networks have significantly improved performance in applications involving visible images. However, the lack of large, labeled datasets limits the use of such networks for thermal imaging applications. Synthetic data is a valuable tool for solving the data problem. Synthetic data allows you to create artificial data from scratch or use complex data manipulation techniques to create new and unique training instances. Numerous methods and tools are available to assist us in creating synthetic data that may be used in a range of applications [49-53]. The significance of synthetic data stems from its ability to generate features that are not available in real-world data. In particular, synthetic data can be used to augment/enrich existing datasets in order to create hybrid datasets that are more suitable for model training. In this approach, synthetic data is typically used to cover portions of the data distribution that are underrepresented in the real dataset, with the primary goal of mitigating dataset bias. The work presented in this chapter employs synthetic data creation to eliminate bias, simulate unencountered conditions, and improve the accuracy of the thermal emotion recognition system.

Facial emotion recognition (FER) has a wide range of practical applications in the fields of virtual learning, cognitive psychology analysis, avatar animation, neuromarketing, human-machine interactions, and entertainment systems. Emotions such as anger, disgust, fear, happiness, neutral, sadness, and surprise form the basis for emotion recognition [54]. The ability to generate representative characteristics determines the accuracy of recognizing a specific mood. On the other hand, the state of emotion often differs between individuals, making classification a difficult task [55]. FER is traditionally classified into two broad categories: static-based FER and dynamic-based FER. Static-based methods take into account only the characteristics of a single image,

whereas dynamic-based methods take into account the sequential relationship between frames [56]. In a controlled context, traditional state-of-the-art classifiers such as support vector machines, AdaBoost, and random forest have been effective at classifying posed facial expressions. However, classifying images collected in an uncontrolled, spontaneous manner or when applied to datasets for which they were not created can present its own challenges [57].

Current research focuses on visible-spectrum face images or videos. An extensive survey on existing approaches and the techniques of emotion recognition on visible images is provided in [55]. Predicting emotions from visible images is often challenging, as variations in the pose and surface of the face result in significant illumination variations in the images. Thermal face recognition systems were developed to address this issue. Thermal face recognition technology has demonstrated a high level of resistance to illumination variations, making it attractive for use in a variety of fields of research and business. Few of the applications include testing psychopathic offenders who demonstrate a lack of emotion [56], classifying cognitive workload [58], measuring emotional stress for safe driving [59], and detecting anxiety by assessing the breathing rate along with facial skin temperature [60]. Fig. 3.2.1 illustrates a set of visible face images and their corresponding thermal face images. Thermal face imaging was introduced for the first time in [61, 62] as a non-invasive, robust facial identification signature. It is primarily composed of temperature changes caused by the vein structure on the facial surface [63]. Thus, thermal face recognition is a distinguishing trait for biometric identification applications [63, 64]. In addition, thermal technology penetrates through smoke, dust, and mist, allowing for the detection of individuals in a wide variety of difficult climatic conditions [65]. On the other hand, thermal images have their own set of issues. One of the most significant issues is occlusion caused by the opaqueness of eyeglasses. Thermal images are also sensitive to the variations in environmental

temperature and physical and health conditions of the person. Moreover, given the limited availability of thermal images in datasets, training a deep neural network model would prove unfruitful.

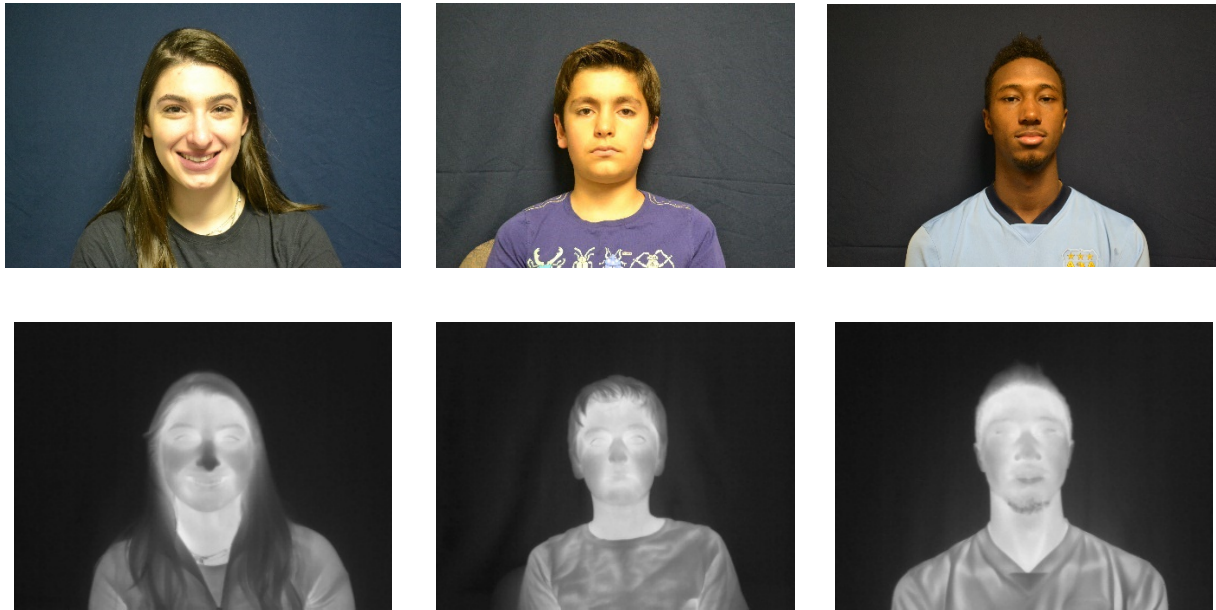


Fig. 3.2.1: Examples of visible face images (top row) and their corresponding thermal face images (bottom row)

Most state-of-the-art algorithms employ visible images to detect facial emotions [66-69]. But a major downside of visible facial emotion recognition algorithms is their inability to differentiate between genuine and fake emotions. Fake expressions can be used to hide genuine emotions. Furthermore, for visible images, the recognition rate is affected by ambient lighting. To combat this, a few researchers [70-74] have focused on emotion detection using infrared thermal images.

Khan [75] proposed a non-intrusive automated facial expression classification system by computing the variances in thermal intensities captured at thermally significant locations on human faces. Derivation of principal components, stepwise selection of optimal and most discriminating

features, and linear discriminant analysis within the reduced optimal Eigenspace were employed for classification purposes. Sophie [76] proposed a technique to categorize nine action units by employing a non-invasive technique that assesses fluctuations of facial heat patterns induced by facial muscles contractions. For quantification purposes, the leave-one-out cross-validation method with the K-nearest neighbor model was used. Trujillo [77] presented a method that used unsupervised Local and Global feature extraction techniques. Bi-modal thresholding was used for face localization, followed by extraction of Eigen features. Finally, a Support Vector Machine was used for classification purposes. Yoshitomi [74] proposed a technique that uses 2D-Discrete Cosine Transform on thermal difference images. Feature vectors were generated according to a heuristic rule. These features were then classified using the nearest neighbor criterion to detect facial expressions. Nhan [73] classified different affective states and baseline states. A genetic algorithm was employed to classify emotions. This algorithm selected the best combination among features such as various time and time-frequency features extracted from facial thermal infrared imaging data, blood volume pulse data, and respiration data. Shen [78] used infrared thermal videos to perform facial expression recognition. A subset of features was extracted according to the F-values of horizontal and vertical temperature difference sequences of different facial sub-regions. Finally, an Adaboost algorithm with the weak classifiers of k-Nearest Neighbor was employed for classifying purposes.

In machine learning, the input feature space and distribution of the training and testing data are similar. Owing to a lack of training samples, learning from a related supplementary dataset and transferring its feature representations becomes necessary. This is the conceptual foundations of transfer learning. Transfer learning emulates the human visual system by utilizing sufficient prior knowledge in other related domains when performing new tasks in the current domain. Recent



research in the field of DCNNs has demonstrated the network's ability to extract unique characteristics from input images and its superior performance across a wide variety of tasks [79]. However, when applied to FER, they have several limitations, including the following: a) a large amount of training data is required to avoid overfitting; b) expensive computational power is required to run deep layers for training purposes; c) existing thermal facial emotion databases are insufficient to train the well-known neural network with deep architecture [80], and d) significant inter-subject variation exists due to various personal characteristics, such as age, facial temperature, and gender [81]. Recently published studies have also called attention to the lack of variance in the thermal FER datasets, which fails to account for a variety of emotional states, alcohol intake or activity, and ambient temperature, leading them to call into question the rigor of the presented results, particularly in real-world scenarios[82].

To address the aforementioned issues, a DCNN based transfer learning method is proposed. The system is designed to solve the task at hand with limited data and reduce computational power by using pre-trained weights from a similar task [68]. The contribution of the work is three-fold. First, different image enhancement techniques are employed to increase the number of images in the dataset. Second, a transfer learning architecture based on DCNN is introduced for thermal facial emotion recognition. The architecture is capable of classifying emotions detected on thermal images based on the feature learned from visible images. Third, the architecture will be tested on a new and comprehensive face database, the Tufts Face Database [13], which contains different modalities of face images.

### 3.3. Proposed Work

This section presents a deep learning-based thermal emotion recognition system (TERNet). A visual block diagram of TERNet is illustrated in Fig. 3.3.1. The proposed network is initialized using the weights provided by the VGG-Face model [83], and the network is fine-tuned using thermal images from the Tufts database. VGG-Face weights were reused, as it was trained on 2.6 million face images for visible image face recognition, when compared to ImageNet, which was mainly developed for different object recognition tasks. This network was trained on  $N=2,622$  unique individuals and set up as an  $N$ -ways classifier. To perform recognition, a triplet loss was employed to verify identities by comparing the  $N$  linear predictors from the final fully connected layer with the ground truth in the Euclidean space. To perform transfer learning, the last fully connected layer is replaced with the custom layer.

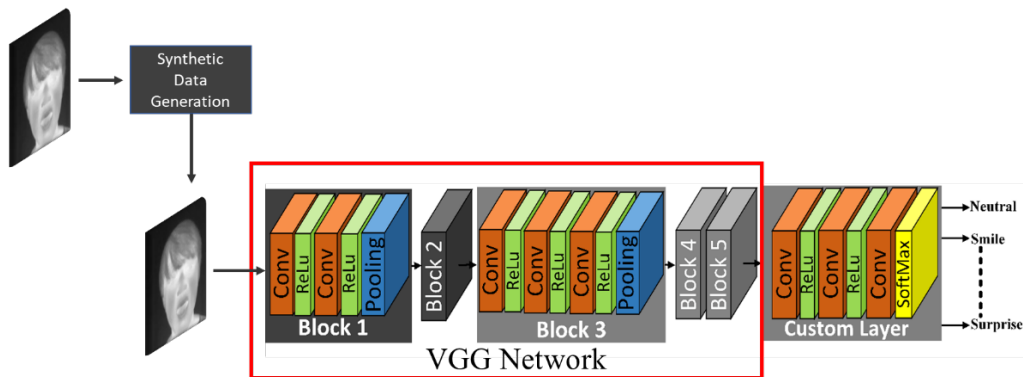


Fig. 3.3.1: The proposed Thermal Face Emotion Recognition system architecture. The input to the system is thermal images. The network follows a VGG style architecture and an added custom layers.

### 3.3.1. Architecture

The CNN architecture of TERNet is provided in Table 3.3.1. In this network, each linear operation, such as convolution layers, is followed by non-linear operations, such as a ReLU activation function [84], as defined in Eq. 3.3.1. Each block [1-5] is followed by a maxpooling layer to reduce the dimensionality of the feature maps. For the last convolution layer, a softmax function is employed and can be formulated as shown in Eq. 3.3.2.

$$\text{ReLU}(\Phi) = \max(0, \Phi) \quad \text{Eq. 3.3.1}$$

$$F(X_i) = \frac{\text{Exp}(X_i)}{\sum_{j=0}^k \text{Exp}(X_j)} \quad \text{Eq. 3.3.2}$$

The VGG-Face model used visible face images of 224 x 224, whereas TERNet uses 161 x161 resolution of the thermal face images. This image is propagated through the network, and the softmax layer provides N predictions. These N predictions are compared to the ground truth using multiclass log loss. This loss function quantifies the accuracy of a classifier by penalizing false classifications. To compute Log Loss, the classifier must assign a probability to each class rather than simply yielding the most likely class. Mathematically Log Loss is defined as

$$\text{Loss} = -\frac{1}{\lambda} \sum_{i=0}^{\lambda-1} \sum_{j=0}^{\rho-1} y_{ij} \log(p_{ij}) \quad \text{Eq. 3.3.3}$$

where  $\lambda$  is the number of samples or instances,  $\rho$  is the number of possible labels,  $y_{ij}$  is a binary indicator of whether label  $j$  is the correct classification for instance  $i$ , and  $p_{ij}$  is the model probability of assigning label  $j$  to instance  $i$ . A perfect classifier would have a Log Loss of

precisely zero. Less ideal classifiers have progressively larger values of Log Loss. The fully connected layers in the VGG-Face model are replaced by convolutional layers. These fully connected layers have fixed dimensions and require a certain number of parameters to connect the input to the output. Moreover, they eliminate spatial coordinates [85]. The replacement allows us to be more efficient as only one forward pass is needed.

| Table 3.3.1: Architecture of TERNet. Building blocks are shown in brackets with [kernel size, # of filters] x repetition |                                |                                |                                |                                |                                        |
|--------------------------------------------------------------------------------------------------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|----------------------------------------|
| Block 1_conv_x                                                                                                           | Block 2_conv_x                 | Block 3_conv_x                 | Block 4_conv_x                 | Block 5_conv_x                 | Block 6_conv_x                         |
| Conv_x<br>[3 × 3, 64] × 2                                                                                                | Conv_x<br>[3 × 3, 128] × 2     | Conv_x<br>[3 × 3, 256] × 3     | Conv_x<br>[3 × 3, 512] × 3     | Conv_x<br>[3 × 3, 512] × 3     | Conv_x<br>[5 × 5, 512]<br>[1 × 1, 512] |
| Maxpool<br>[2 × 2], stride = 2                                                                                           | Maxpool<br>[2 × 2], stride = 2 | Maxpool<br>[2 × 2], stride = 2 | Maxpool<br>[2 × 2], stride = 2 | Maxpool<br>[2 × 2], stride = 2 | [1 × 1, 4]<br>Softmax                  |

### 3.3.2. Tufts Face Database

This section introduces the Tufts Face Database [13], a comprehensive multi-modal face database introduced to benchmark imaging systems. The face image dataset was developed for multimedia, forensic, security, biometric, and entertainment applications. The database includes six different types of image modalities: photographic images, a computerized facial sketch created from a volunteer's portrait using the FACES software, thermal images, near-infrared (NIR) images, a video, and 3D images of the participant. The dataset contains more than 10,000 raw images collected from 113 volunteers of different nationalities, gender identities, ages, and ethnicity (74 females + 39 males, from more than 15 countries with an age range between 4 to 70 years old). An example of the different image modalities is shown in Fig. 3.3.2.

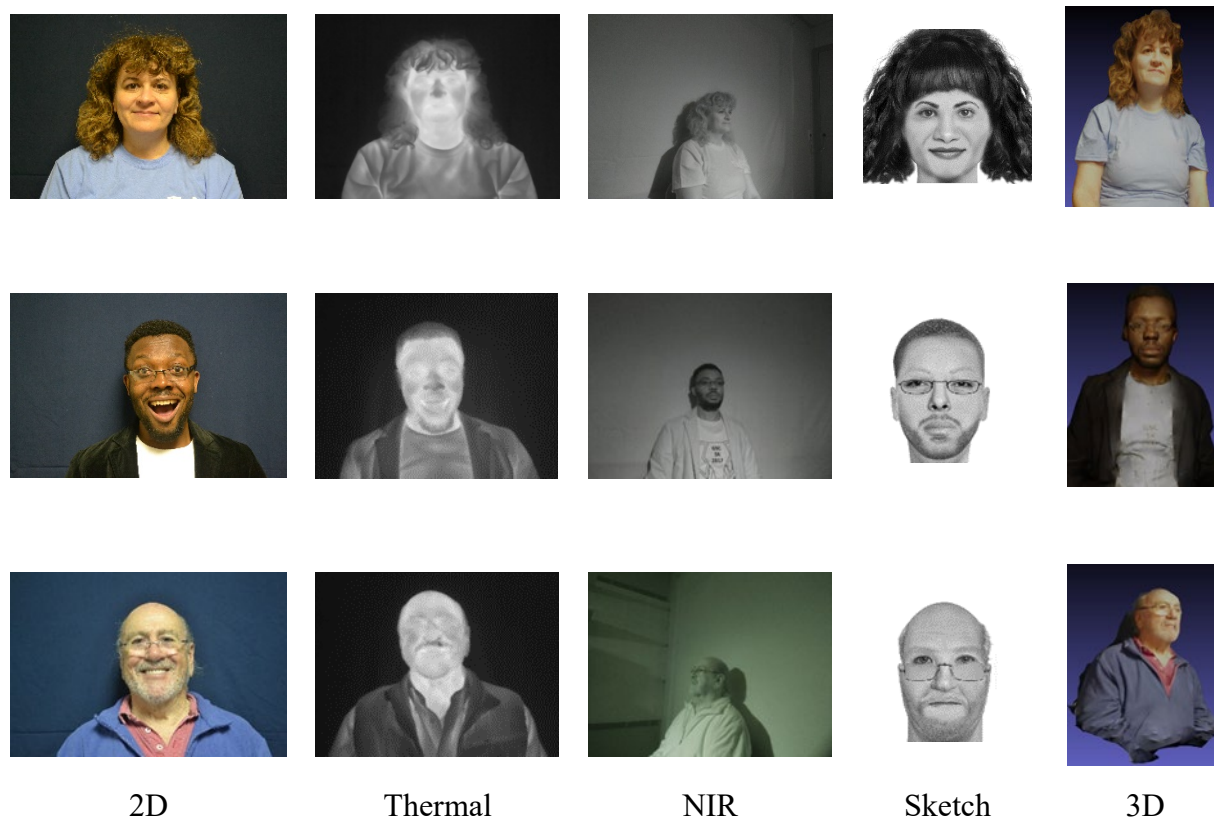


Fig. 3.3.2: Example images selected from the Tufts Face Database.

### 3.3.3. Dataset and synthetic data generation

For thermal FER, only thermal emotion images were considered from the Tufts Face Database. This dataset contains a total of 565 images from 113 participants, with five images per participant. In this work, expressions consist of a) a neutral expression, b) a smile, c) eyes closed (sleepy), d) an exaggerated shocked expression, and e) with sunglasses. However, the images with sunglasses were discarded for the emotion recognition task. An example of the different facial expressions is illustrated in Fig. 3.3.3.

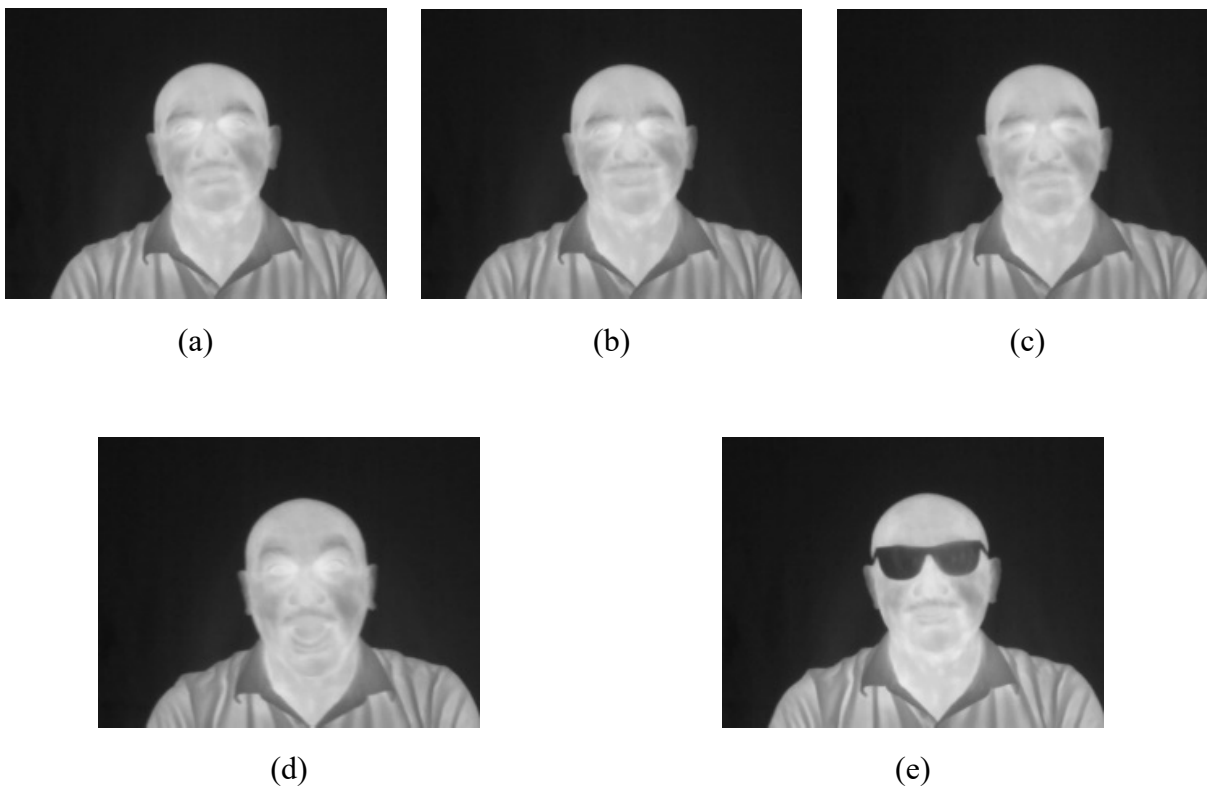
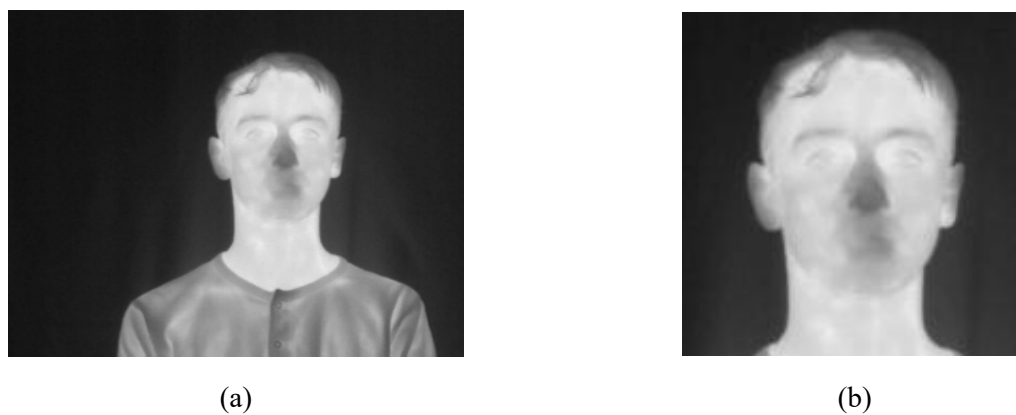


Fig. 3.3.3: An illustration of the different facial expressions where (a) a neutral expression, (b) smiling, (c) sleepy, (d) shocked, (e) with sunglasses.



*Continued*

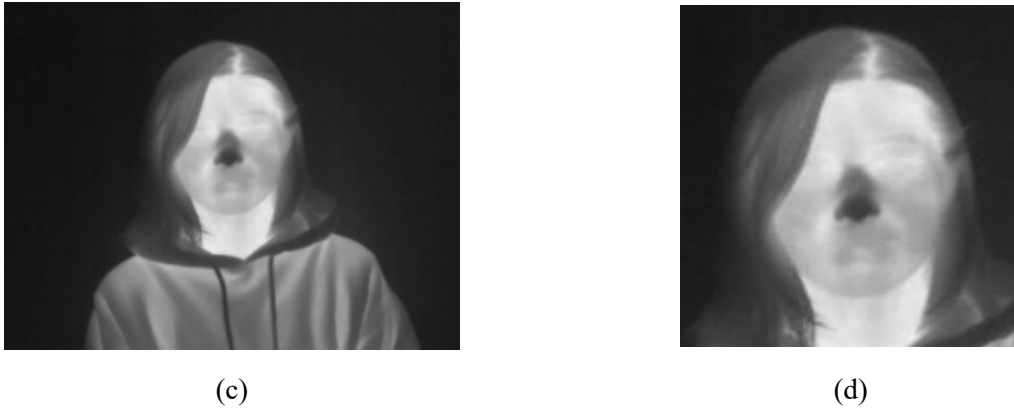


Fig. 3.3.4: Illustrative examples of the cropped thermal image, panel [a, c] are original image, and panel [b, d] are the corresponding cropped image

The images were cropped such that irrelevant information was removed. The original images were of size 336 x 256, and the cropped images were of size 161 x 161. Examples of the original and cropped image can be visualized in Fig. 3.3.4. Synthetic data was generated using classical image processing techniques to increase the amount of data. The goal was to introduce variations in the dataset, thereby increasing the number of images. The variations in the images were introduced to simulate variations in temperature conditions. In order to do so, different image enhancement techniques were employed, such as, Contrast limited adaptive histogram equalization (CLAHE) [86], guided filtering [87, 88], Bi-histogram equalization, and an intensity adjusted CLAHE technique. Examples of the enhancement results are shown in Fig. 3.3.5. As seen in the figure, the Bi-histogram equalization technique outputs introduce unwanted artifacts. The remainder of the enhancement techniques were utilized to generate the dataset for this work. For training and testing, all these images were considered and split into 70% training and 30% testing.



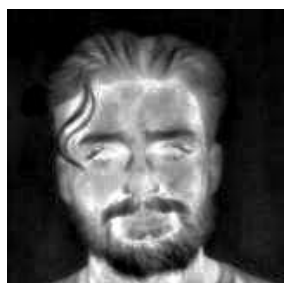
(a)



(b)



(c)



(d)



(e)



(a)



(b)



(c)



(d)



(e)

Fig. 3.3.5: Illustrative examples of the same expression images obtained after image enhancement- panel [a - e] are original images, guided filtered image, adaptive histogram equalization, intensity adjusted CLAHE, and Bi-Histogram equalization images, respectively



## 3.4. Results

### 3.4.1. Training Phase

For training, thermal images of size 161 x 161 pixels were considered. All the images were preprocessed by subtracting the mean and divided by the standard deviation of the dataset. To reduce overfitting, during training, the images were randomly rotated by +/- 15 degrees, shifted horizontally and vertically by 10 pixels, and horizontal flips were performed. Training was performed with an ADAM optimizer [89] by setting the learning rate to  $2 \times 10^{-5}$ ,  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ . The batch size was set to 32. The pretrained VGG-Face model was employed for initializing the starting layers, and the custom layers were initialized with Xavier Initialization [90]. The new model was fine-tuned till the convergence.

### 3.4.2. Comparison

Since the dataset is relatively small when compared to VGG-Face and ImageNet, pretrained network weights are employed. The performance of TERNet was computed using pretrained weights obtained by training the original VGG16 network using VGG-Face and ImageNet datasets separately. TERNet fine-tuned with pretrained weights obtained from VGG-Face performed better when compared to the model trained on ImageNet dataset. This is because the VGG-Face pretrained weights correlate better with the thermal face images and the features extracted are similar for visible and thermal face images. This can be realized by the top1 accuracy (see Table 3.4.1) obtained when the proposed model was fine-tuned using weights obtained from the aforementioned datasets.

| Table 3.4.1: Comparison of TERNet with state-of-the-art models. |                    |          |          |          |
|-----------------------------------------------------------------|--------------------|----------|----------|----------|
| Model                                                           | VGG16-custom layer |          | VGG19    | ResNet18 |
| Pretrained weights                                              | VGG-Face           | ImageNet | ImageNet | ImageNet |
| Top1 Accuracy (%)                                               | <b>96.2</b>        | 93.4     | 94.8     | 94.2     |

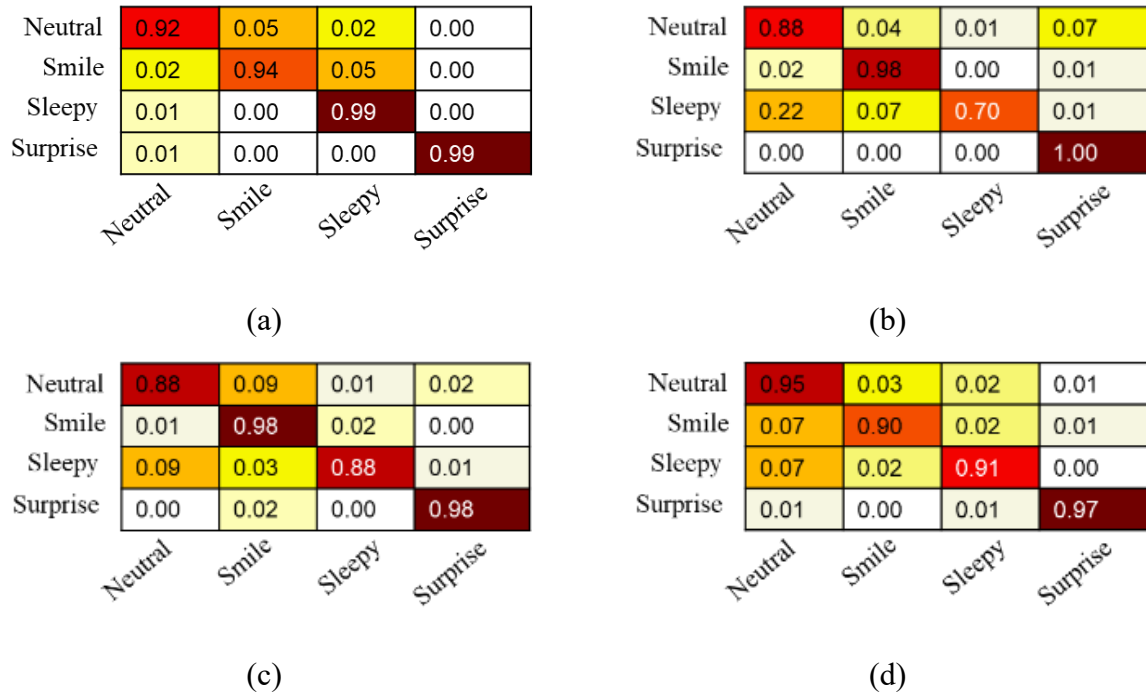


Fig. 3.4.1: Panel [a-d] provides the confusion matrices generated by using (a) TERNet - VGG16 with pretrained VGG-Face, (b) VGG16 with pretrained ImageNet, (c) VGG19 and (d) ResNet18. The x-axis represents the predicted labels and the y-axis represents the true labels.

The proposed model was compared with other state-of-the-art networks such as VGG19 and ResNet18. The feature maps obtained from the original networks were passed through average pooling layers, and finally, a fully connected layer followed by a softmax function was used to obtain the predictions. It can be inferred from Table 3.4.1 that VGG16 with custom layer and

pretrained VGG-Face weights produce the best accuracy. To observe the misclassifications among the classes, confusion matrices are provided in Fig. 3.4.1.

The images with ‘sleepy’ emotion tend to fluctuate depending on our training scheme and architecture. This might be due to very subtle changes in the face images when expressing ‘sleepy’ and ‘neutral’, especially in thermal images. This can be verified by examining the confusion matrices in Fig. 3.4.1, which indicates that the ‘sleepy’ class is often mistaken for other classes by the models. Even though the above observations highlight high accuracy, training CNNs using a small dataset with classes that are not visually distinctive is difficult.

### **3.5. Summary**

The chapter presents a robust and efficient thermal face emotion recognition system based on a custom convolution neural network, TERNet. The thermal facial recognition system utilizes pretrained weights from VGG-Face as initial facial feature maps. To extract thermal facial features and classify emotions more effectively, a custom layer is added to the VGG16 network. Extensive computer simulations show that the proposed system produces 96.2% accuracy in categorizing facial emotions. The architecture has proven to have advantageous characteristics in position-invariance, varying illumination conditions, and high noise tolerance. It was also observed that the network posed difficulty in assigning correct labels to faces depicting some of the more nuanced expressions. As a part of future work, a loss function that can distinguish between classes with subtle changes will be investigated. Furthermore, the proposed system will be prototyped and integrated into augmented reality devices to demonstrate an online assessment of human emotion.

## Chapter 4. Feature Detection and Matching

Feature detection and matching are widely used in computer vision and play a vital role in image registration [91, 92]. A feature is defined as a distinct, significant structure present in the image. Features in an image can vary from a single pixel to edges and contours, and the process of recognizing such features is known as feature detection. A feature detector's output is often a set of precise locations in an image known as feature points. These locations were selected based on their resistance to noise, transformations, and other deformations. The goal of a feature detector is to locate a set of distinct interest points or regions that are stable (invariant).

The detection of features can be automatic or manual, depending on the complexity of the application. Image features are broadly classified as local and global features [93]. Global features describe and present the features of an image as one multidimensional feature vector. In other words, the global representation approach generates a single vector containing values that measure several image properties, such as color, texture, or shape. In practice, a single vector is extracted from each image, and then two images are compared by comparing their feature vectors. Though these features are well-suited for detecting shapes and texture, they are affected by partial occlusion and clutter [94]. Moreover, global features are not suitable for the spatial localization of objects. On the other hand, local features distinctively represent an image based on relevant regions such as closed boundary regions, edges, contours, and corners, while remaining invariant to the viewpoint and illumination changes.

This chapter presents an overview of various feature detectors and their application to image matching, image registration, image stitching, and color transfer. The chapter is divided into three main sections. In particular, the feature detection and matching algorithms are utilized for

two specific tasks: (1) to match antemortem and postmortem dental radiographs and (2) to develop a color transfer algorithm. In addition, a scale and rotation invariant image stitching algorithm is proposed to generate a panoramic view of the underwater environment. The algorithm is further evaluated by testing its robustness in different applications.

## **4.1. Matching Antemortem and Postmortem Xray Radiographs**

### **4.1.1. Motivation**

Law enforcement agencies have been developing dental biometric feature-based tools in forensic identification, including identifying the victims of massive catastrophes, such as the 9/11 terrorist attack, airplane crashes, and other natural disasters such as hurricanes and tsunamis [95]. When human remains are found, the priority of the investigation is to confirm the identity of the deceased. Physical features are often destroyed in such situations, making it very difficult to identify an individual. Dental tissues have the capability of withstanding higher temperatures and pressure. The survivability of dental features makes them an ideal identification tool. Forensic dental identification methods typically use radiological documentation techniques such as periapical radiographs, bitewing films, panoramic radiographs, and CT [96, 97]. The quality of dental records scans is extremely inconsistent across the world and is accompanied by noise [98]. Hence, analyzing and visualizing different scans is tedious and time-consuming.

In some cases, features extracted from dental x-rays are insufficient to identify individuals. Also, different imaging techniques carry varying amounts of information. For example, bitewing images hold localized information about the curvature of the tooth and the roots, whereas panoramic images illustrate an end-to-end picture. Here, a solution is proposed that utilizes image-based features to match antemortem and postmortem panoramic radiographs.

### 4.1.2. Introduction

Disaster victim identification (DVI) is a demanding task that involves authorities and experts from various fields [99]. When human remains are found, the priority of the investigation is to determine the identity of the deceased. In recent years, several disaster incidents have challenged the DVI team with various difficulties [99]. Accurate identification of the deceased would aid in obsequies and issues involving criminal investigations, insurance settlements, and military proceedings [100].

Routine identification tasks are a simple one-to-one matching process; however, this is not the case in mass disasters [101]. Multiple fatality incidents (MFI) is now the common term given to the fatal results of any mass disaster. The identification of a large number of casualties in an MFI is complex. Such incidents include accidents (fires, military accidents, plane crashes), natural disasters (earthquakes, tornadoes, tsunamis, etc.), genocide, and terrorism [102]. Physical features are often destroyed in such situations, making it incredibly hard to identify an individual. This builds a perfect platform for forensic dentistry. Dental tissues have the capability of withstanding higher temperatures and pressure. Hence, they make an ideal identification tool.

In forensic dentistry, comparisons of antemortem and postmortem dental records are performed by experts to look for similarities [103]. Antemortem records are essential for developing an Automatic Dental Identification System (ADIS). The quantity and quality of antemortem dental records are highly inconsistent across the world [98]. Hence, analyzing and visualizing different scans is tedious and time-consuming. In some cases, features extracted from dental x-rays are insufficient to identify individuals [104]. The literature highlights methods and tools to aid in automated analysis.

Image segmentation methods are in practical use in radiology, keeping in mind the end goal to help or automate numerous methodologies [105-108]. The process involves subdividing the image into its constituent regions. Dental image segmentation helps identify dental abnormalities and can be utilized for applications in forensic dentistry, dental biometrics, etc., for human identification [109]. Alazab [110] proposed a fully automated segmentation algorithm that combined the region growing method with an optimized false boundary elimination method. Oprea S [111] developed an edge detection technique for dental x-ray images for segmenting decayed teeth. The major drawback of this method was that the segmentation was done manually. Since dental x-ray images are of poor quality, the proposed work did not use any enhancement technique. Thus, it failed to classify caries accurately. Rani [112] developed a fully automatic multistep segmentation method that utilizes a combination of Dual-Tree Complex Wavelet Transform and Watershed transform to segment caries from dental x-rays. Shuo [113] implemented an automatic variational level set segmentation framework to segment lesions that used a window-based feature extraction followed by principal component analysis (PCA). The images were chosen at random and segmented using hierarchical level set region detection to train the support vector machine.

Age estimation is an important characteristic used to establish the identity of any individual [114]. Kvaal [115] method is based on examining linear measurements of the pulp width, tooth width, and root length at three different root levels. The examinations were initially applied on periapical radiographs and later extended to panoramic radiographs. Cameriere [116] devised a method based on analyzing the pulp and tooth areas on dental radiographs. Methods described in [117-120] used different approaches for dental age estimation in adults based on pulp/tooth volume ratio from cone-beam computer tomography images.



Postmortem dental findings and teeth conditions are recorded in a dental chart and compared with antemortem records during DVI. Manually matching these record is physically challenging and laborious. Nomir [121] proposed a hierarchical chamfer matching method for contour matching. The segmented teeth were numbered, and matching was performed in a ranked fashion. Jain [103] suggested a computerized method for matching the ante and postmortem dental images. The tooth contour was extracted from each tooth and based on the intensity. Matching was performed on the corresponding image based on the extracted contours. Zhou [122] proposed a 3-step method to retrieve matched images. First, the images were classified into bitewing, periapical, or panoramic images. Second, the teeth on bitewing images were segmented using a top-hat filter and an active contour method. Finally, matching was performed on the corresponding image based on the boundary shape. Hosntalab [123] investigated a 3-step process for classifying and numbering teeth on multi-slice CT images. First, the tooth regions were segmented using a variation level set. Second, crucial features were extracted from these regions using the wavelet-Fourier descriptor. Lastly, the teeth were grouped into four classes using a supervised classifier. Using this technique, they achieved high classification accuracies above 94% for 804 teeth from 30 CT cases. Y Miki [124] proposed a system that investigated the utility of the DCNN for classifying tooth types using cone-beam CT. By increasing the number of training samples, the proposed work was able to achieve a high accuracy of 91%.

This work proposes an image feature-based matching system that accurately matches postmortem and antemortem panoramic radiographs. The system employs state-of-the-art feature detectors for the matching process.

### 4.1.3. Proposed Work

The general flow diagram of the proposed matching architecture is illustrated in Fig. 4.1.1. The system presented in this work is an automated tool that identifies individuals by matching postmortem scans to antemortem scans. Panoramic radiographs are often very noisy and low in contrast [125, 126]. The low image quality is a result of low dose usage. Low-dose usage is related to its effect on a patient's health [127]. Therefore, image enhancement plays a vital role in improving the quality of the radiograph, thereby improving the analysis and detection of features. The enhancement method proposed in Chapter 2 is employed to enhance the quality of the images.

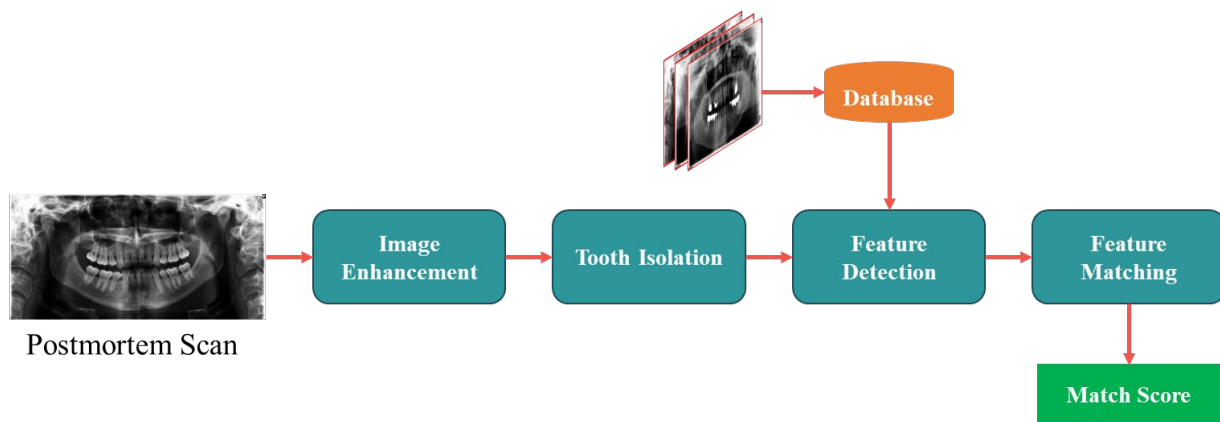


Fig. 4.1.1: The proposed system architecture matches panoramic antemortem and postmortem radiographs.

Tooth isolation or segmentation is performed to analyze each tooth individually. Several state-of-the-art algorithms have been proposed in the literature to isolate particular instances in images[37][85, 128]. The proposed system employs the Mask R-CNN [129] algorithm to isolate and number each tooth in the radiographs. Mask R-CNN is an extension of Faster R-CNN [130] that provides a class label, a bounding box coordinate, and an object mask for each tooth in the

radiograph. The model is divided into two stages, where the first stage is used for object detection, and the second stage carries out semantic segmentation. Once each tooth is isolated and numbered, the image is fed to the feature detection and matching system. This work provides an overview of the various image feature detectors and their application in matching panoramic radiographs.

***Difference of Gaussian (DoG):*** This multiscale detector improves on the Laplacian of Gaussian (LoG). Lowe [131] proposed an effective technique that detects stable features from scale-space extrema by closely approximating the LoG. It is a feature detector that is translation, rotation, and scale-invariant. The whole detection technique entails detecting extrema in the scale space, that is, across various scales and places, and then selecting keypoints based on a measure of stability. The DoG function  $D(x, y, \sigma)$  can be obtained without convolution by subtracting neighboring Gaussian pyramid scale levels separated by a factor  $k$ . It is defined as

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \quad \text{Eq. 4.1.1}$$

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma)$$

where  $G(x, y, \sigma)$  is the Gaussian kernel. The steps involved in estimating the features are: 1) Construct the scale-space of an image by convolving the image with the Gaussian kernel at various scales, 2) Divide the resulting scale-space into octaves, 3) Each generated octave is convolved again with the Gaussian to build a series of scale-space images for that octave, 4) Adjacent sub-octave scale-space are subtracted to generate the DoG, 5) To proceed to next octave, the Gaussian image is down-sampled by 2, and 6) Detect the maxima and minima of the DoG in scale-space by comparing each point with eight neighbors in the current image and with nine neighbors each in the scales above and below. This process is illustrated in Fig. 4.1.2.

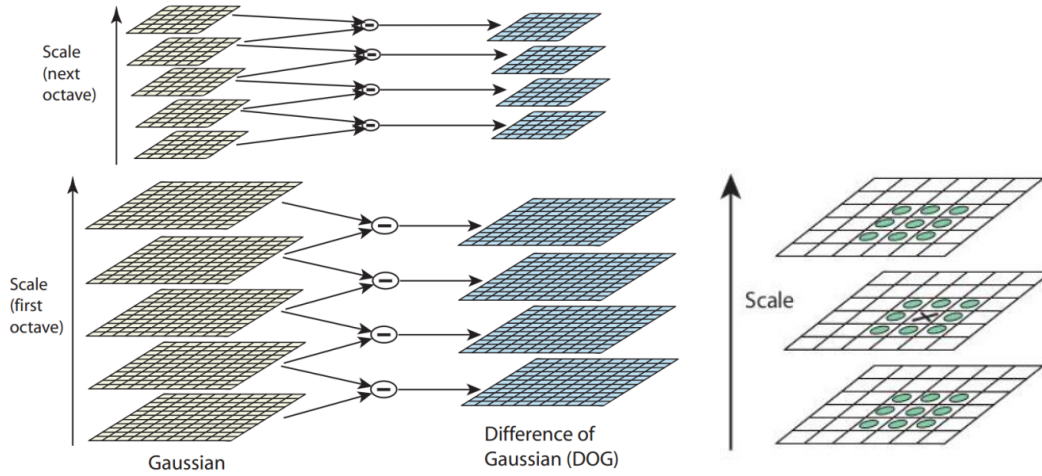


Fig. 4.1.2: An illustration of the DoG acquired by subtracting adjacent sub-octaves and detecting the maxima and minima by comparing each point with its neighbors. [132]

**Hessian Detector:** The Hessian blob detector is derived from the Hessian matrix, which is composed of second-order partial derivatives. This matrix has been used to analyze local image structures. For a given intensity image  $I(x, y)$  the Hessian matrix is defined as

$$H(x, y, \sigma) = \begin{bmatrix} I_{xx}(x, y, \sigma) & I_{xy}(x, y, \sigma) \\ I_{xy}(x, y, \sigma) & I_{yy}(x, y, \sigma) \end{bmatrix} \quad \text{Eq. 4.1.2}$$

$$\det(H) = I_{xx}I_{yy} - I_{xy}^2 \quad \text{Eq. 4.1.3}$$

where  $I_{xx}$ ,  $I_{xy}$ , and  $I_{yy}$  are the second-order image derivatives estimated using the Gaussian function with standard deviation  $\sigma$ . The steps involved are: 1) Find the Hessian matrix's determinant, which detects image structures with high signal variation in two directions, as depicted in Eq. 4.1.3. 2) Build a scale-space representation by convolving the image with Gaussians of increasing size. 3) The Gaussian kernels are chosen such that successive images differ by a particular scale ratio. The

Hessian matrix has to be made invariant to scale. Hence a factor  $\sigma = 2$  is multiplied with the Hessian matrix, where  $\sigma$  represents the scale of the image. 4) A 3 x 3 search window is swept over the entire image. If the pixel value is larger than all eight immediate neighbors and above a given threshold, a feature point is then associated with the present location.

**Harris-Laplace Detector:** is a scale-invariant corner detector proposed by Mikolajczyk [133] that integrates the Harris corner detector and a Gaussian scale-space representation. The Harris-Laplace detector utilizes the Harris detector to compute interest points at different scales. For a given 3x3 window, at each scale, a point that outputs the largest Harris response is picked as the interest point among all points within that window. After computing the interest point, the window is moved to the next point until it covers the entire image in each scale. The Harris-Laplace detector is defined as

$$M(x, y, \sigma_I, \sigma_D) = \begin{bmatrix} I_x^2(x, y, \sigma_D) & I_x I_y(x, y, \sigma_D) \\ I_x I_y(x, y, \sigma_D) & I_y^2(x, y, \sigma_D) \end{bmatrix} \quad \text{Eq. 4.1.4}$$

where  $I_x$  and  $I_y$  are the image derivatives estimated in their respective directions using a gaussian kernel  $\sigma_D$ .  $\sigma_I$  determines the current scale at which the Harris corner points are detected. The Harris-Laplace detector searches for local maxima keypoints in the current and neighboring scales by comparing the keypoint with those in the adjacent scale.

**Hessian Laplace Detector:** This is similar to Harris Laplace. The difference is that they start from the determinant of the Hessian rather than the Harris corners. The Hessian function detects points in space, and the Laplacian Function is used to determine the local maxima for all scales. The Laplacian function is defined as

$$Laplacian(x; \sigma_D) = \sigma_D^2 |I_{xx}(x; \sigma_D) + I_{yy}(x; \sigma_D)| \quad \text{Eq. 4.1.5}$$

where  $I_{xx}$  and  $I_{yy}$  are the second-order derivatives of the intensity image  $I(x, y)$ . Hessian-Laplace detector extracts a large number of features over the entire image at a marginally lower repeatability when compared to Harris-Laplace.

**Multiscale Harris Detector:** The Harris corner detector is applied at successive integration scales. For locating interest points, the eigenvalues of the matrix  $M$  are computed for each pixel, where  $M$  is defined as:

$$M(x, y) = \sum_{u,v} w(u, v) * \begin{bmatrix} I_x^2(x, y) & I_x I_y(x, y) \\ I_x I_y(x, y) & I_y^2(x, y) \end{bmatrix} \quad \text{Eq. 4.1.6}$$

where  $I_x$  and  $I_y$  are the local image derivatives in the x and y directions, and  $w(u, v)$  denotes a weighting window over the area  $(u, v)$ . This algorithm determines many points which repeat in the neighboring scales. An auto-correlation matrix is employed in the Harris detector, which aids in feature detection.

**Multiscale Hessian Detector:** As image dimensions can be different, it is necessary to introduce a measurement scale that changes within a certain range [134]. The Taylor expansion is calculated in the vicinity of the point under consideration. This expansion will approximate the structure of the image up to the second order [135]. Finally, the eigenvalues of the Hessian are used to determine the possibility of a blob feature being present. The Multiscale Hessian can extract features from images that have different dimensions.

**Scale Invariant Feature Transform (SIFT) Descriptor:** The SIFT descriptor [131, 136] combines the Difference of Gaussians (DoG) interest region detector with a feature descriptor.

This descriptor has distinct properties that make it insensitive to complications like rotation, translation, and object scaling. Many studies have been conducted on the descriptor's performance, and SIFT generally performs well compared to other local descriptors [137]. The SIFT Descriptor encodes image information in a localized set of gradient orientation histograms by providing minor shifts in the positions, achieving robustness to lighting variations. Cascade filtering was introduced, which reduced the cost of extracting features by allowing costly operations to be performed only after passing the initial test [131]. The data from the DoG is used to create the keypoint descriptors. The orientation of the data created must be changed to match the keypoint's orientation. The data is then weighted using a Gaussian with a variance of 1.5 times the keypoint scale. Using this data, a histogram centered on the keypoint is created.

The magnitude and orientation of the image gradient are sampled using a  $16 \times 16$  grid around the keypoint location, and the level of Gaussian blur is selected using the region scale. As shown in Fig. 4.1.3, the sampled region is further divided into a  $4 \times 4$  square grid with accurate scale and rotation. Every cell in this grid contains a histogram with eight orientation bins.

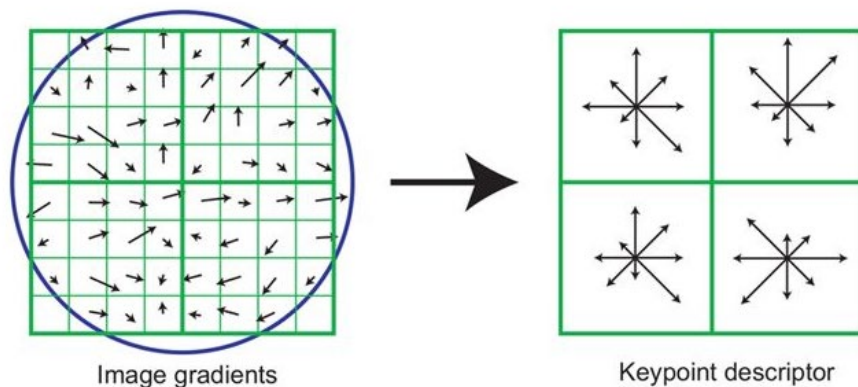


Fig. 4.1.3: An illustration of the image gradient and the keypoint descriptors

Local feature-based matching algorithms are often more efficient and are capable of handling geometrical deformation than global feature-based matching algorithms. Once the features are extracted from the postmortem and antemortem images, they are matched by estimating each keypoint's nearest neighbor. Two keypoint are matched if the distance between these points is significantly smaller than any other keypoint pair. This distance threshold is defined during the matching process.

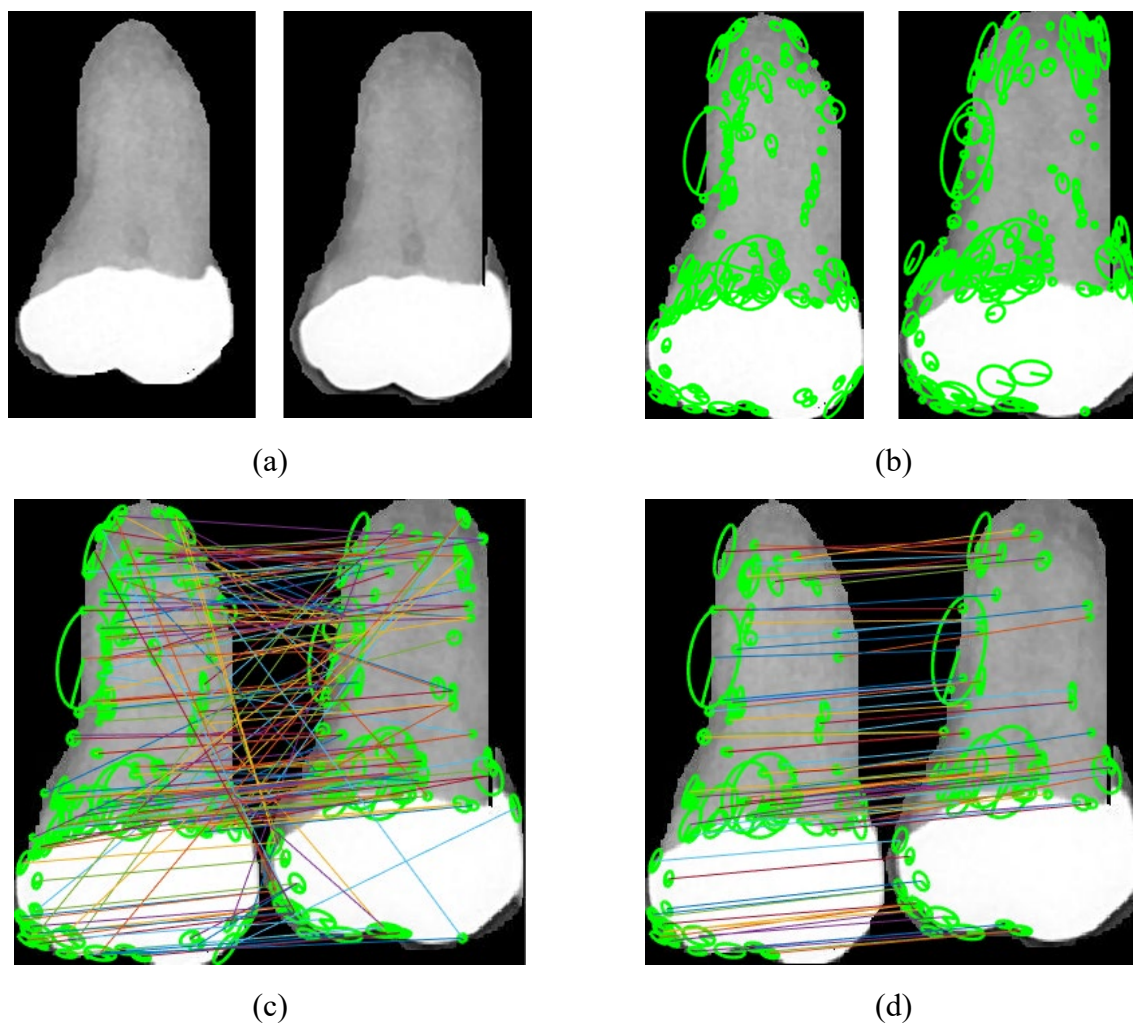


Fig. 4.1.4: The figure displays the feature detection and matching process, where (a) represent the tooth images to be matched, (b) features detected on the images using the multiscale Harris detector, (c) matched feature with erroneous matches, and (d) accurate matches after geometric verification.



An example of the matching process for a teeth pair is shown in Fig. 4.1.4. As seen in Fig. 4.1.4 (c), the matching algorithm produces erroneous matches. These matches are eliminated by performing geometrical verification. In the verification process, the similarity transform is computed using Eq. 4.1.7. This transformation is employed to map all keypoints from one image to another. The matching points are disregarded if the mapped keypoints lie beyond a defined threshold. The final matched output is shown in Fig. 4.1.4 (d). The final matched score is estimated by averaging the matched score for each tooth pair.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = s R(\theta) \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$$

Eq. 4.1.7

$\theta \rightarrow$  angular rotation;  $s \rightarrow$  scaling factor;  $(t_x, t_y) \rightarrow$  translation vector

#### 4.1.4. Experimental Results

To evaluate the performance of the proposed matching algorithm, a dataset containing antemortem and postmortem panoramic radiographs was created. The radiographs for this dataset were acquired from the Tufts University School of Dental Medicine. The radiographs were de-identified and saved with a unique identifier. These scans include images from multiple visits. To simulate antemortem and postmortem data and demonstrate the concept of matching, the first scan of the patient was considered as the antemortem data, and the following scan was considered as the postmortem data. A total of nine patients with follow-up scans were used for this study. Fig. 4.1.5 displays examples of the antemortem and postmortem scans for two patients. An additional nine dissimilar radiograph pairs were utilized to evaluate the matching performance of the proposed system.

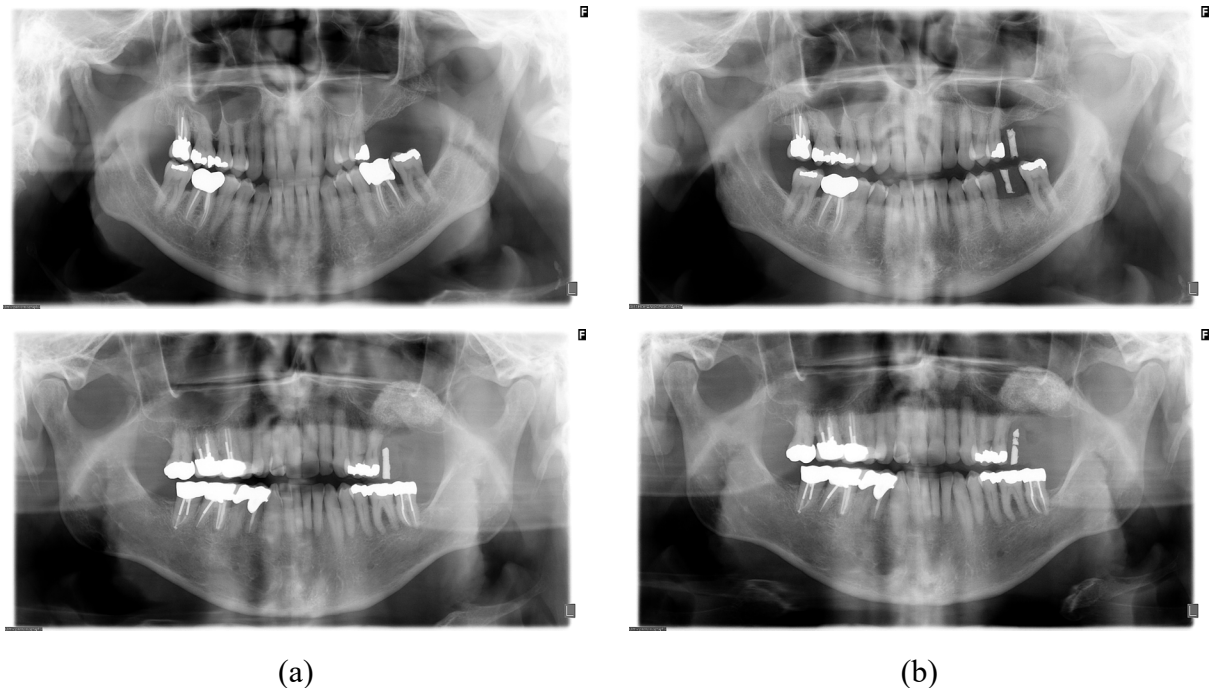


Fig. 4.1.5: Example image pairs of the (a) antemortem and (b) postmortem scans

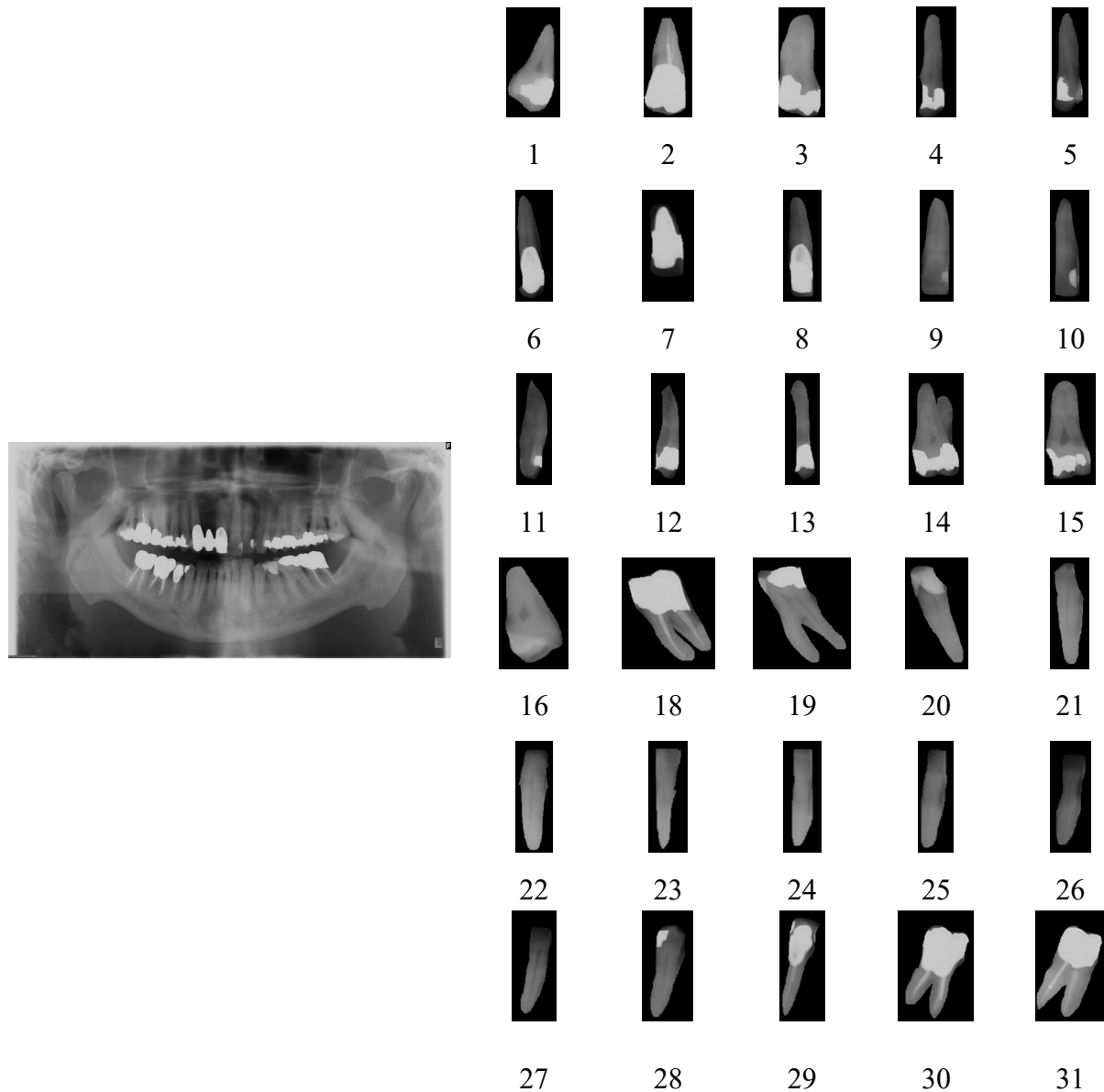


Fig. 4.1.6: The figure displays a panoramic radiograph and its isolated tooth. The numbers 1-31 denote the tooth number.

During the data acquisition process, dental images are exposed to degradations and/or deformations. Therefore, image enhancement plays a crucial role in evaluating dental images. The image enhancement process proposed in Chapter 2 is utilized to enhance the quality of the radiographs, in turn improving the feature detection accuracy. Following the enhancement process,

the teeth in the postmortem and antemortem scans are isolated/segmented using the Mask-RCNN algorithm [129]. The Mask-RCNN model was trained on the Tufts Dental Database to isolate each tooth and number each tooth in the radiograph. Fig. 4.1.6 displays a radiograph and its isolated teeth.

| Table 4.1.1: A comparison of the performance of the feature detectors in matching antemortem and postmortem scans |     |         |                 |                |                   |                    |
|-------------------------------------------------------------------------------------------------------------------|-----|---------|-----------------|----------------|-------------------|--------------------|
|                                                                                                                   | DoG | Hessian | Hessian Laplace | Harris Laplace | Multiscale Harris | Multiscale Hessian |
| Subject 1                                                                                                         | 23  | 123     | 95              | 7              | 64                | <b>703</b>         |
| Subject 2                                                                                                         | 19  | 83      | 60              | 4              | 47                | <b>480</b>         |
| Subject 3                                                                                                         | 18  | 68      | 80              | 4              | 26                | <b>372</b>         |
| Subject 4                                                                                                         | 14  | 61      | 47              | 3              | 49                | <b>433</b>         |
| Subject 5                                                                                                         | 26  | 97      | 53              | 3              | 53                | <b>510</b>         |
| Subject 6                                                                                                         | 21  | 99      | 88              | 8              | 76                | <b>922</b>         |
| Subject 7                                                                                                         | 13  | 67      | 29              | 3              | 57                | <b>489</b>         |
| Subject 8                                                                                                         | 3   | 32      | 60              | 5              | 13                | <b>483</b>         |
| Subject 9                                                                                                         | 19  | 85      | 77              | 5              | 46                | <b>578</b>         |
|                                                                                                                   |     |         |                 |                |                   |                    |
| Subject 10                                                                                                        | 15  | 42      | 51              | 4              | 26                | <b>169</b>         |
| Subject 11                                                                                                        | 14  | 42      | 50              | 3              | 25                | <b>135</b>         |
| Subject 12                                                                                                        | 11  | 38      | 51              | 3              | 20                | <b>212</b>         |
| Subject 13                                                                                                        | 10  | 38      | 42              | 2              | 12                | <b>218</b>         |
| Subject 14                                                                                                        | 13  | 45      | 70              | 1              | 21                | <b>218</b>         |
| Subject 15                                                                                                        | 12  | 33      | 42              | 3              | 24                | <b>137</b>         |
| Subject 16                                                                                                        | 4   | 14      | 59              | 2              | 12                | <b>87</b>          |
| Subject 17                                                                                                        | 12  | 36      | 33              | 8              | 35                | <b>189</b>         |
| Subject 18                                                                                                        | 14  | 46      | 84              | 3              | 16                | <b>200</b>         |

Once the teeth are isolated, the postmortem scans are matched with the antemortem scans in the database using the feature detection and matching algorithm. The feature described above were employed in the detection process, and matching scores were generated. The results for each feature detector are tabulated in Table 4.1.1, where the values depict the matching score. Subjects 1-9 have matching scans, and subjects 10-18 have dissimilar scans. These dissimilar scans simulate cases when no matches are found.

| Table 4.1.2: A comparison of the feature detectors in estimating the feature count range for good and bad matches |      |         |                 |                |                   |                    |
|-------------------------------------------------------------------------------------------------------------------|------|---------|-----------------|----------------|-------------------|--------------------|
|                                                                                                                   | DoG  | Hessian | Hessian Laplace | Harris Laplace | Multiscale Harris | Multiscale Hessian |
| Good Match                                                                                                        | 3-26 | 32-123  | 29-95           | 3-7            | 13-76             | 433-922            |
| Bad Match                                                                                                         | 4-15 | 14-46   | 33-84           | 1-8            | 12-35             | 87-218             |

As seen in Table 4.1.1, Multiscale Hessian produces the most number of feature matches. Furthermore, the multiscale hessian features can accurately measure the similarities between scans. The matching score ranges from 433 to 922 matches for similar scans, and for dissimilar scans, the matching score ranges from 87 to 218 matches. The other feature detectors generate a lower matching score. The score range for each feature detector is displayed in Table 4.1.2. As shown in the table, other feature detectors cannot discern between good and bad matches. Hence Multiscale Hessian is employed to match antemortem and postmortem scans.

## **4.2. Pixel-based Color Transfer**

### **4.2.1. Motivation**

Most state-of-the-art techniques focus primarily on daytime surveillance, and limited research has been performed on night-time monitoring. Surveillance is often more important in darker environments since many activities of interest occur at night. But, night-time imagery presents its challenges: they are mostly monochrome and very noisy. Night vision systems are also affected by a bright light or glare off a shiny surface. Color imagery has several benefits over grayscale imagery. The human eye can discriminate a broad spectrum of colors but is limited to about 100 shades of gray [138]. Also, color drives visual attention and aids in a better understanding of the scene. Moreover, the context information of the scene affects the way humans distinguish and recognize things. The essential step of the coloring process is the choice of an appropriate color image model and color mapping scheme. To enhance relevant information of night-time images, a color mapping or color transfer technique is employed. This section proposes a robust pixel-based color transfer architecture that maps the color characteristics of the daytime images to the night-time images. The architecture can also compensate for image registration issues encountered during acquisition. A visual analysis of the results demonstrates that the proposed method performs better in comparison to the state-of-the-art techniques and is robust to different imaging sensors.

### 4.2.2. Introduction

Night-time imagery plays an important role in the field of computer vision. Night vision systems are deployed in various scenarios related to surveillance, intelligence gathering, and security [139]. The two most common night-time imaging systems are low-light-level cameras (VNIR) and thermal infrared (IR) cameras. Most night-time surveillance cameras use a standard gray or green-scale representation. The footage acquired from such systems is often accompanied by noise and is hard to interpret. Moreover, night-time images of a scene contain a relatively high dynamic range in intensity, where parts of the scene are illuminated by strong light sources, while the other regions are comparatively darkened [140]. However, with the advent and increasing availability of multimodal imaging sensors, the color representation of night vision imagery is in demand. Color imagery proves advantageous over grayscale imagery as it improves visual disparity, thereby improving the understanding of the scene and driving visual attention [139]. The goal of such an application is to leverage relevant information from each modality into a single representation.

Images acquired under low light conditions are often hard to interpret due to their narrow dynamic range. Images captured in such conditions are accompanied by sensor noise. With limited information, interpreting a scene and performing recognition tasks can prove cumbersome and unreliable. Hence, preprocessing these low-light images is crucial. Most of the classical methods used for low light image enhancement are histogram-based and are primarily categorized as direct and indirect methods. Direct methods make use of a contrast measure to improve the image contrast, whereas the indirect method manipulates the dynamic range without using a contrast measure. However, such techniques usually result in false manipulation of the color space of the scene. Comprehensive research has been conducted in improving the quality of low-light images.

The most widely used image enhancement method is the Retinex theory [141]. The retinex algorithm decomposes the image into reflectance and illumination map. Most retinex algorithms utilize the reflectance map as the enhanced image [142]. Dong [143] proposed a technique to enhance low light videos, which applies an optimized dehazing technique on inverted negative images. Fu [144] proposed a weighted variational model that simultaneously enhanced the target images by estimating reflectance and illumination. Gao [145] presented a low light image enhancement method (LIME) that estimated the illumination of each pixel and predicted the total maximum value in each channel. The model, although efficient, suffered from color distortion. While the aforementioned algorithms work well, they also have some drawbacks, such as being computationally expensive or sometimes producing halo artifacts. In the presented work, enhancing the low light image resulted in better feature extraction, which in turn aided in the image registration process. To accurately perform color transfer, it was crucial to align the two images.

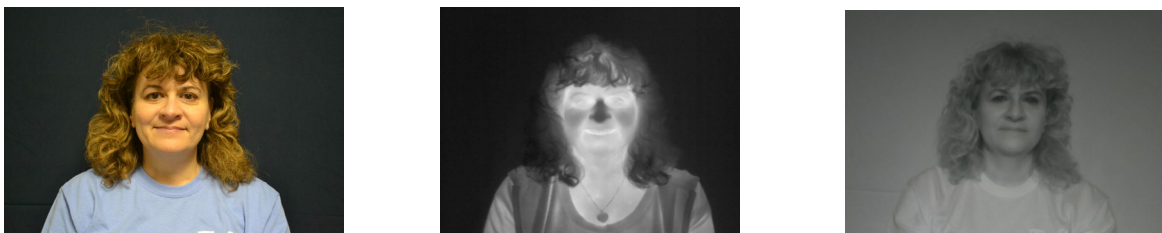


Fig. 4.2.1: An example of multimodal images (a) visible image, (b) infrared image, and (c) near infrared image

Color transfer or mapping has received a lot of attention in recent years. With the availability of multispectral imaging sensors, fusing two or more modalities improves the human's visual capabilities in discriminating objects of interest. An example of multimodal images can be seen in Fig. 4.2.1. Color transfer methods can be broadly classified into three categories:



geometric-based, statistics-based, and user-guided-based. Geometric-based color transfer methods use matching feature points in multiple images to estimate the color mapping functions. First, feature detectors such as Scale-Invariant Feature Transform (SIFT) and Speeded-Up Robust Feature (SURF) are adopted to acquire the keypoints. After obtaining the keypoints in each image, the RANdom SAmple Consensus algorithm (RANSAC) is employed to eliminate the outliers efficiently and improve the matching accuracy. Finally, a color transfer function is derived by minimizing the color difference between the corresponding keypoints and applied to the target image. Statistics-based color transfer method uses statistical correlation to generate the mapping function between the reference and target images. The mapping function transfers the color style of the reference image to the target image. Papadakis [146] proposed a histogram-based color style transfer method that utilizes the energy functional minimization technique to apply the transformation to the target image. Reinhard [147] proposed a statistics-based algorithm to perform color transfer between two images. User-guided methods generate the color transfer mapping function based on the inputs provided by the user.

Many color transfer techniques have been proposed in the literature. Wang [148] suggested a nonlinear transfer method to enhance the targets with an intense red color while the background rendered natural color appearance. Zheng [149] presented a local color night vision algorithm that enhanced the color mapping effect based on image segmentation and recognition. Toet [15] proposed a color mapping technique that maps the natural color appearance from daytime images to its respective night-time images. The method matches the statistical properties of the reference image (daytime) to the target image (night-time). One of the major drawbacks of this method was that the image registration process was performed manually. Also, the method did not provide consistent object colors and was computationally expensive. Hogervorst [150] proposed a new

color mapping technique that used the combination of a multi-band image and a corresponding natural color daytime reference image. The mapping improved the match between the multi-band image and the reference image, giving it a pleasant daytime color appearance. Pouli [151] presented a color transfer technique that can progressively reshape the histogram of a given image to match it to the histogram of another based on user-guided inputs. Xiao [152] presented a method that performed Singular Value Decomposition (SVD) on the reference and target images, following which the pixels in the target image are transformed appropriately to match the reference. Pitie [153] proposed an automated color mapping method using color iterative distribution transfer. The method uses a nonlinear and convergent algorithm to generate one-to-one color mapping using a three-dimensional probability density function. As most color transfer methods are accompanied by noise, the author proposes a technique to reduce grain noise artifacts, which adjusts the gradient field of the corrected image to match the test image. Fecker [154] proposed a color correcting technique using a basic histogram matching algorithm for the luminance component and chrominance components. Further modifying the first and last bin values of the cumulative histograms avoids possible visual artifacts. Nikolova [155] proposed a fast histogram specification algorithm for color transfer. The algorithm uses a fixed-point algorithm to minimize the functions and obtain color corrected images. With the advent of deep neural networks, a few style transfer techniques have been proposed in the literature. Gatys [156] introduced an image style transfer algorithm based on a pre-trained convolutional neural network (CNN). Extending the approach by Gatys, Luan [157] extended the approach used by Gatys by enforcing photorealism on the result image.

In this work, a pixel-based color transfer system to recolor night-time imagery is proposed. To evaluate the performance of the proposed system, we compare the results to the methods in

[153] and [152]. In addition, the robustness of the algorithm is validated by fusing visible and thermal face images.

### 4.2.3. Proposed Color Transfer System

In this section, a reference-based color transfer system is presented. A visual flow diagram of the color transfer architecture is illustrated in Fig. 4.2.2. The system comprises of four main components, namely, image registration, image preprocessing, color transfer, and color correction.

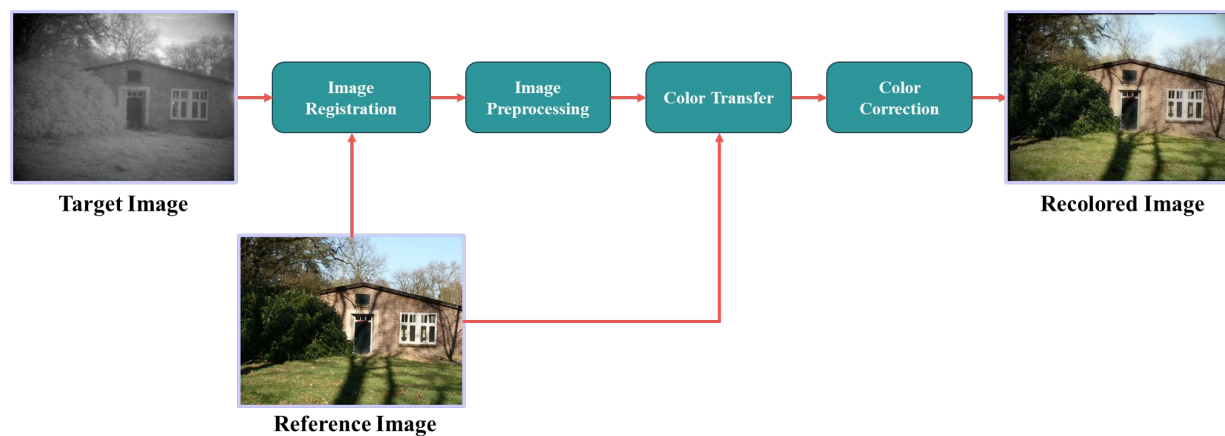


Fig. 4.2.2: The proposed flow diagram of the color transfer system

Image registration defines the process of overlaying two or more images of a scene acquired from multiple views and sensors. Image registration finds its application in medical imaging [158], military [159], traffic control [160], panoramic imaging [161, 162], agriculture [163], and remote sensing [164]. The image registration task involves two major steps: Feature matching and Homography estimation. The feature matching process performed in this step is adopted from Section 4.1.3. Homography estimation is used in applications such as image rectification, image registration, and computation of rotation and translation between two images

[165]. In order to acquire a distortion-free result, the information of how much an image must be rotated and translated with respect to the other must be known. This is achieved by estimating the homography using the Random Sample Consensus (RANSAC) model [166]. The homography estimates the scaling and rotation factor required to align the images together. An illustration of the matching and alignment process is shown in Fig. 4.2.3 and Fig. 4.2.4, respectively.

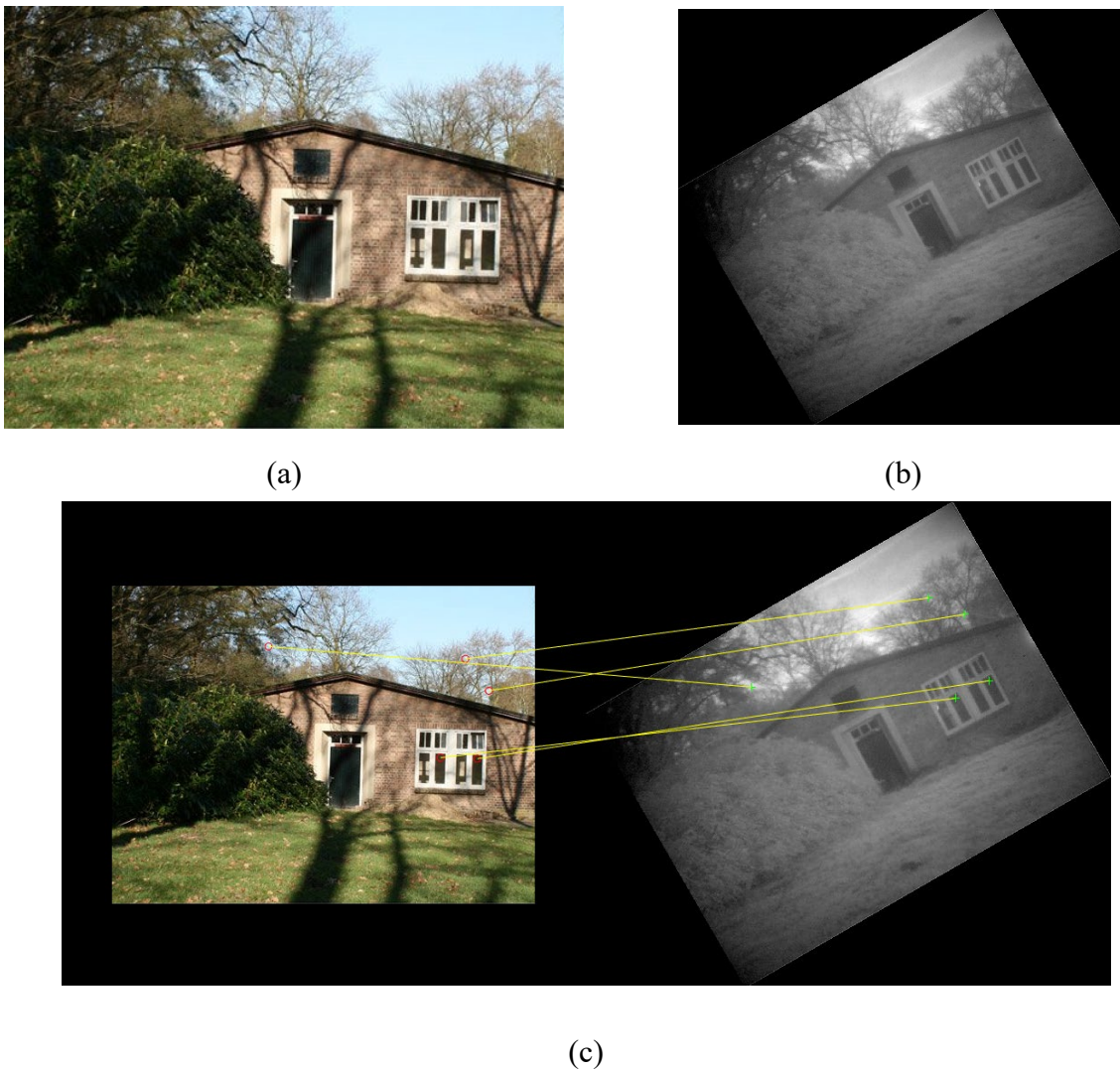


Fig. 4.2.3: The images above depict the image matching results where (a) is the reference image, (b) is the rotated view of the near-infrared image of the same scene, and (c) displays the matched features

Images are often accompanied by noise and blur during acquisition. Moreover, images captured at night are susceptible to poor visibility, hindering the image registration and color transfer process. To ensure the reliability of the systems, it is desirable to enhance the low-light images. The technique used to enhance and simultaneously denoise the images is a modified version of the method described in [88]. To eliminate the unwanted noise information, the edge

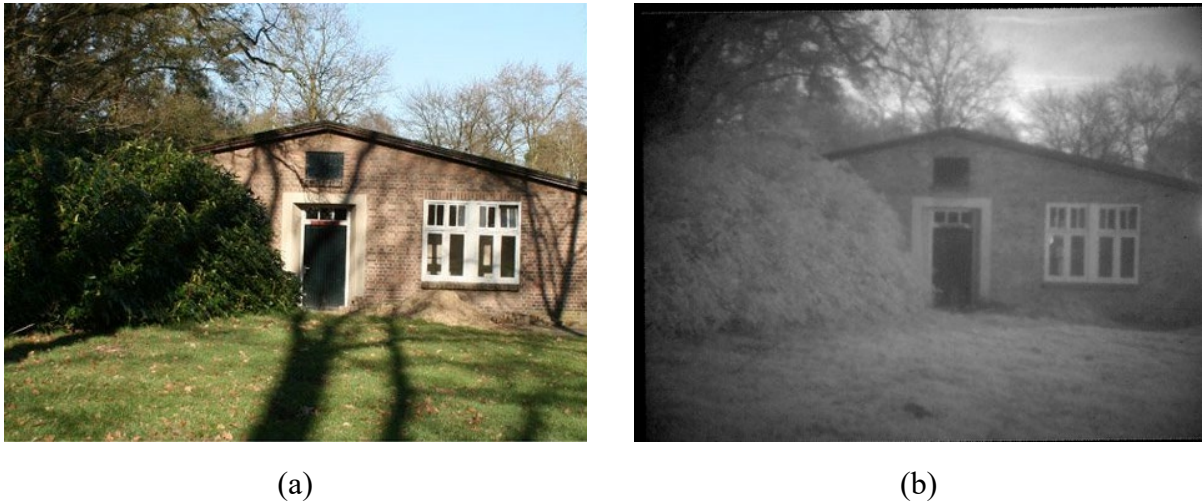


Fig. 4.2.4: Aligned images (a) reference image (b) target image

preserving method described in Chapter 2 is used. The guided filter supersedes other filters due to its non-approximate manner implementation, which results in a better quality image and reduces the run time due to its invariance with the size of the filter [42]. Results of the enhancement process can be visualized in Fig. 4.2.5 and Fig. 4.2.6.





(a)



(b)

Fig. 4.2.5: Results obtained after image enhancement. (a) original image (b) enhanced image



(a)



(b)

Fig. 4.2.6: Results obtained after image enhancement. (a) original image (b) enhanced image

The goal of the color transfer algorithm is to transfer each pixel information from the reference to the target image. For the given input image  $X$  where  $X \in [X_o, X_L]$  the color transfer function is defined by Eq. 4.2.1, where  $f_1(x)$  is the gray tone of the target image, and  $M$  is  $\max(X)$ .  $f_2(x)$  is defined as in Eq. 4.2.2.

$$\tau(X) = f_2 \left[ f_1(x) * \frac{c}{M} \right] \quad c \in [R, G, B] \quad \text{Eq. 4.2.1}$$

$$f_2 = X_o + [X_L - X_o] * k(X) \quad \text{Eq. 4.2.2}$$

where  $k(X)$  is the CDF of the image. For a given reference and target image, the result is shown in Fig. 4.2.7.



(a)



(b)



(c)



(d)

*continued*



(e)



(f)

Fig. 4.2.7: The images above demonstrates the results of the color transferred images. (a) & (b) reference image, (c) & (d) target image, and (e) & (f) recolored image

After performing the color transfer technique, a mismatch still exists between the recolored and reference image. A color correction method proposed by Pitie [16] is applied to minimize the difference between the two. The color correction technique presents the application of an iterative distribution transfer (IDT) with grain reduction. For a given dataset source  $u$ , the proposed manipulation consists of choosing an orthogonal basis  $R = [e_1, \dots, e_n]$  and applying the mapping function  $\tau$  given by:

$$\tau(u) = u + R \begin{bmatrix} t_1(e_1^T u^{(k)}) - e_1^T u^{(k)} \\ \vdots \\ t_N(e_N^T u^{(k)}) - e_N^T u^{(k)} \end{bmatrix} \quad \text{Eq. 4.2.3}$$

where  $t_i$  is the one-dimensional pdf transfer mapping for the axis  $e_i$ . Iterating this manipulation over different axes will result in a sequence of distributions that will eventually converge to the target distribution. Color mapping is often accompanied by grain artifacts. The author in [153]



proposes a grain reduction algorithm by forcing the noise to stay constant and adjusting the gradient field of the target image to match the reference image. The result of the color correction technique is shown in Fig. 4.2.8.



(a)



(b)



(c)



(d)

Fig. 4.2.8: Color correction results. (a) original image, (b) target image, (c) after color correction, and (d) after re-graining. Image source [153].

#### 4.2.4. Experimental Results

This section presents the computer simulations performed on the RGB-NIR Scene Dataset [166] and AMOS dataset [167]. The RGB-NIR Scene Dataset consists of 477 images with nine categories captured using RGB and NIR cameras. The images were captured using separate exposures from modified SLR cameras using visible and NIR filters. The images were pre-registered by extracting SIFT features and using RANSAC to find a similarity transform. The AMOS dataset consists of 17 image sequences containing 1722 images in total. These images were taken at different times of the day using fixed webcams. In order to assess the quality of the proposed method, the results were visually compared to the methods mentioned in Xiao [152] and Pitié [153].



Fig. 4.2.9: A comparison of color transfer methods (a) reference image, (b) target image, (c) result obtained by Pitié [153], (d) result obtained by Xiao [152], and (e) proposed method

The proposed algorithm was applied to the entire dataset. But, only a few images are shown for simplicity. First, the proposed color transfer technique is compared to [153] and [152] using random day-night images and the AMOS dataset. The results are displayed in Fig. 4.2.9 and Fig. 4.2.10. As seen in Fig. 4.2.9 and Fig. 4.2.10, the proposed method performs well in comparison. The result closely mimics the daytime images. The proposed system is further tested on the RGB-NIR Scene Dataset, comparing it to the method proposed by [153]. As seen in Fig. 4.2.11, the proposed system performs seamlessly compared to [153].

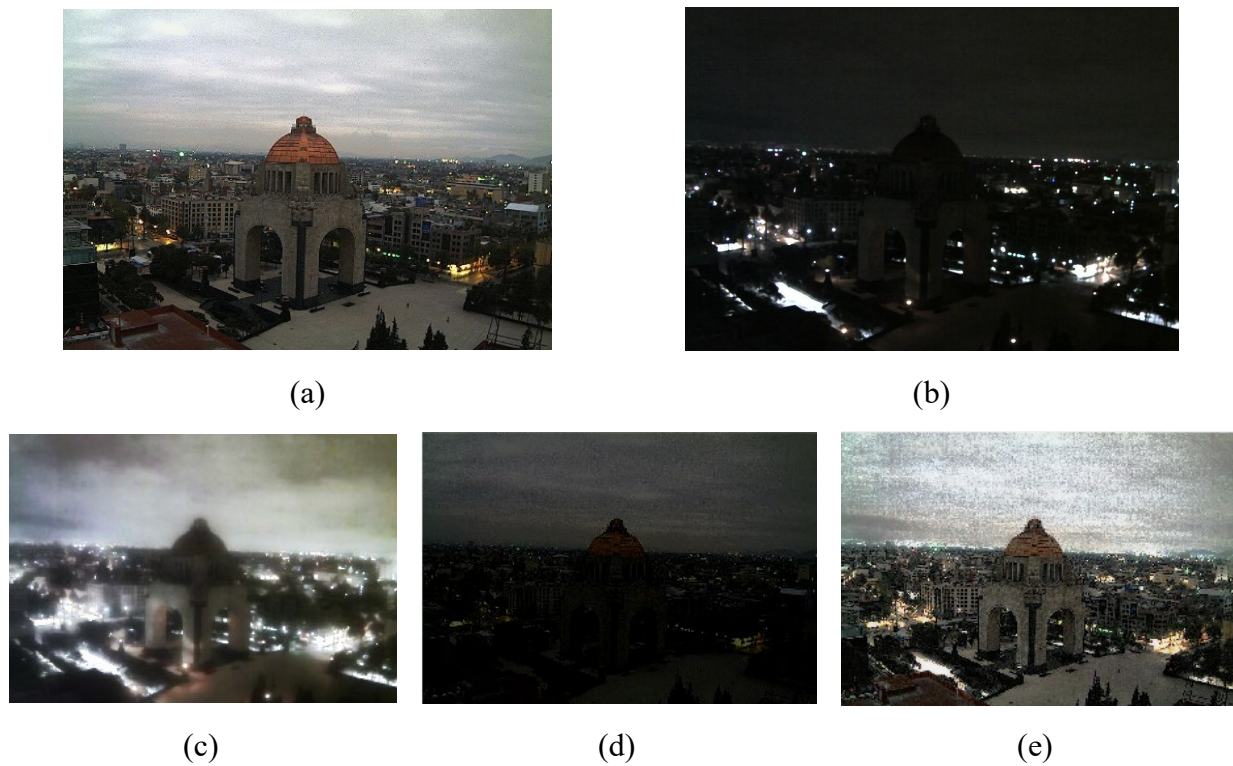


Fig. 4.2.10: A comparison of color transfer methods applied to the AMOS dataset image (a) reference image, (b) target image, (c) result obtained by Pitie [153], (d) result obtained by Xiao [152], and (e) proposed method



Moreover, it captures the minute detail present in the reference image during color transfer. Hence, the color transfer method can recolor images from different modalities. In some cases, generating a 3D of a scene from multiple sources is challenging. The images are acquired at different times during the day, resulting in a color mismatch. The proposed method can be used to recolor such images and reduce the color variation. The proposed method can also be used as a means to generate synthetic data to train and develop novel AI systems.

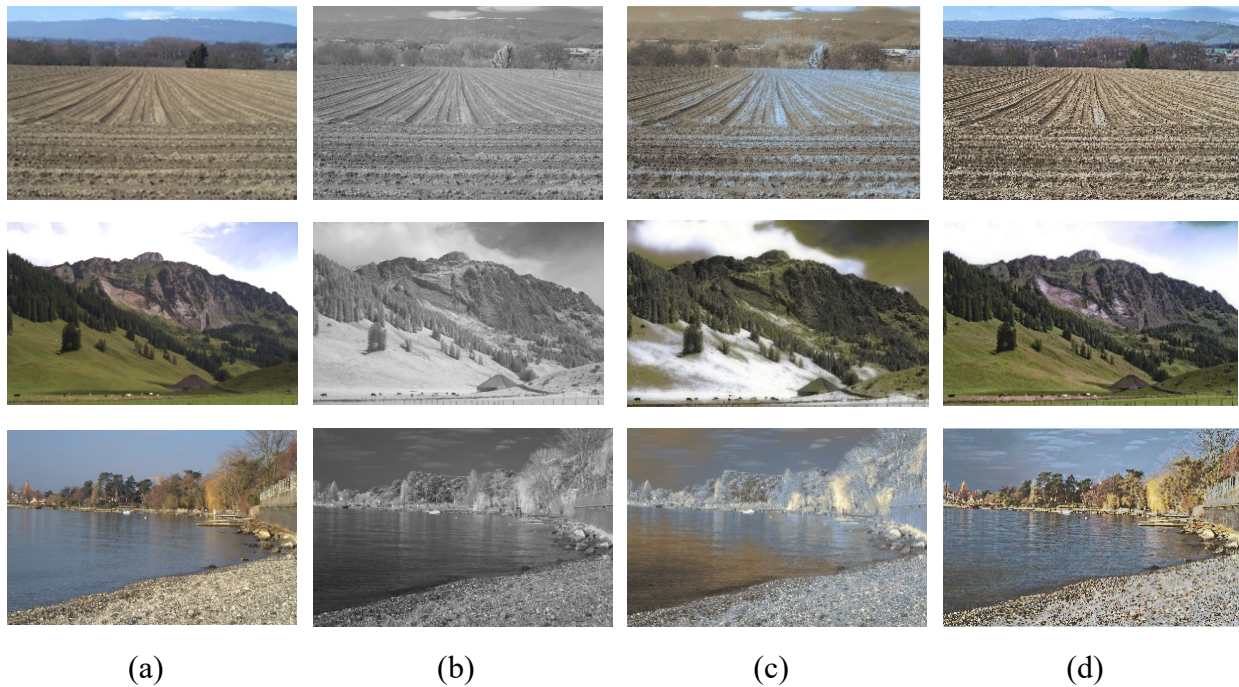


Fig. 4.2.11: A comparison of color transfer methods applied to the RGB-NIR Scene Dataset images (a) reference image, (b) target image, (c) result obtained by Pitie [153], and (d) proposed method

#### 4.2.5. Extended Application: Fusion of thermal and visible face images

Significant research has been conducted to improve the performance of automatic human authentication systems in uncontrolled environments. Despite extensive research, many issues with face authentication continue to exist due to the inherent difficulties in extracting biometric data from faces [168]. As previously stated in Chapter 3, several factors affect the performance of the face recognition system, such as pose variations, facial expression changes, occlusions, and, most importantly, illumination changes. Hence, thermal image-based face recognition systems were developed to address the issue of variation in illumination. However, thermal imagery presents its own challenges. It is sensitive to temperature changes in the surrounding environment and variations in the heat patterns of the face. Therefore, it can prove beneficial for automatic face recognition systems to fuse/combine the properties of the visible and thermal images into a single image.

$$\tau(X) = f_2 \left[ f_1(x) * \frac{I}{M} \right] \quad \text{Eq. 4.2.4}$$

where  $I$  denoted the gray level intensity.

The proposed color transfer method is modified to transfer the gray level features from the visible image to the thermal image. The fusion is performed by modifying Eq. 4.2.1 by employing the pixel transfer method to a gray level image, as defined in Eq. 4.2.4. The color correction task is excluded during the fusion process. Thermal and visible face images from the Tufts face database were used to evaluate the proposed method. Matching pairs of thermal and visible images were cropped and resized to 128 x 128 pixels. The results of the fusion process are displayed in Fig. 4.2.12.

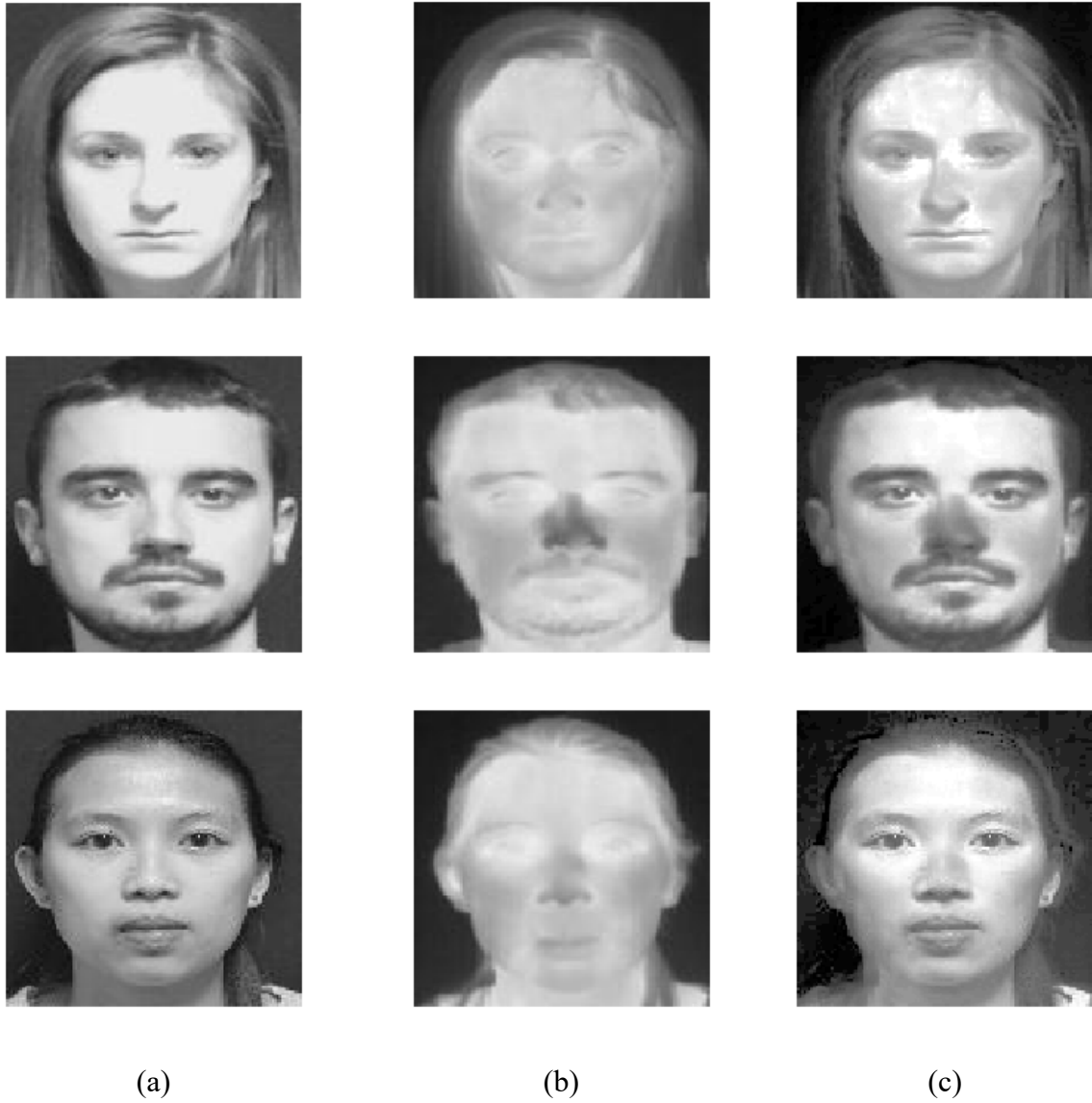


Fig. 4.2.12: An illustration of the image fusion process. (a) visible gray level image, (b) thermal image, and (c) fused image.

## **4.3. Image Stitching**

### **4.3.1. Motivation**

Image stitching is a technique of combining multiple underwater images to generate a composite image with a much larger field of view. With its advancement in various imaging applications, underwater image stitching presents its own challenges. Underwater images suffer from clouding effects and grayed-out colors, and these complexities impede the image stitching process, resulting in misalignments and uneven illumination. This section presents a system that combines multiple underwater images to generate a fused image, overcoming the impediments of a visible seamline. The method is based on two new correlation techniques: Alpha-Winzorized and Weighted Rank Order Statistics. The system is designed to adapt to the difficulties encountered in underwater image stitching. The performance of the stitching algorithm is tested against its ground truth, both subjectively and quantitatively. The quantitative analysis is performed using non-reference-based underwater quality measures such as UICM, logAMEE, and CTSUIQM. In addition, the robustness of the stitching algorithms is tested by generating end-to-end finger images.

### **4.3.2. Introduction**

Over the last few decades, automatic underwater imaging systems have become a powerful tool in oceanography study [169]. Underwater imaging has a wide range of applications, including habitat monitoring, structural monitoring [170], environmental damage assessment, and a variety of industrial uses [171]. In panoramic image stitching, advances have been made to acquire a global view of our surroundings. But its application to underwater image stitching is in its nascent

stages [172]. Underwater image stitching is used for various purposes, including damage assessment, detection of temporal changes, reconstruction of the ocean floor, 360-degree view (virtual tour of the aquatic environment), and maritime rescue [173].

Image stitching techniques are broadly categorized into two groups: direct and feature-based [174]. The direct technique is a pixel-by-pixel method, where the pixel intensities of one image are compared to another. In practice, this approach employs all of the image's pixels. The technique uses a cost function or simply considers the sum of the absolute differences between the overlapping pixels to combine the images [174]. The direct method's key advantage is that it optimally utilizes the information provided for image alignment. The technique measures the contribution of every pixel in the image. This approach is more accurate as it uses all the available information but requires human interaction to find the matching image pairs. It is also computationally heavy. The feature-based technique compares feature points of one image to another using a local descriptor, which is faster. The method establishes a correspondence between feature points. An overview of the various feature detectors is discussed in Section 4.1.3. Commonly used stitching methods are effective only when dealing with good, visible, and noiseless images with uniform illumination.

Underwater images, in general, suffer from low contrast and resolution due to poor visibility, variations in color with respect to depth, and extreme noise [175]. As seen in Fig. 4.3.1, light absorption and dispersion make underwater images seem murky and make it challenging to visualize underwater images [176]. As a result, it impacts the image stitching system's accuracy and efficiency. Aside from the above-listed issues, image acquisition is critical to image quality. While capturing images in shallow waters, the surrounding light is sufficient to acquire quality images [172]. At deeper depths, however, artificial lights are necessary for image acquisition



[172]. Artificial light sources are unreliable due to inhomogeneous illumination, resulting in a bright spot in the center surrounded by an illuminated area. As a result, such images are challenging to process.



Fig. 4.3.1: Poor visualization of underwater images [177, 178]

A few underwater image stitching algorithms have been proposed in the literature [175, 179-181]. Henhyu Li [179] proposed an algorithm that uses Speeded-Up Robust Feature (SURF) to extract feature points to stitch the images. If the feature point does not satisfy a threshold, the phase correlation method is used to stitch the images. Finally, a weighted smoothing fusion technique is used to combine the images. Chen [180] developed a method that used a combination

of Scale Invariant Feature Transform (SIFT) and wavelet fusion to generate a panoramic image. Saeed Hojjati [181] uses a color transfer method to transform images into the same color space and applies the graph cut method to blend the two images together. Pardos [175] uses pixel-level graph cut and a weighted average gradient blending operation to stitch multiple underwater images.

This chapter employs two novel adaptive correlation techniques to stitch underwater images and generate a panoramic view of the underwater environment. The correlation techniques estimate the best seam line along which the images are stitched together. The images are enhanced prior by employing a modified version of the method mentioned in [182]. The performance of the stitching algorithms is tested against its ground truth. Non-reference quality metrics such as logAMEE (image contrast) [183, 184], Underwater image colorfulness measure (UICM) [185], and Combined Transform and Spatial domain Underwater Image Quality Measure (CTSUIQM) are used in the quantitative analysis.

### 4.3.3. Proposed Work

The general flow diagram of the stitching method is illustrated in Fig. 4.3.2. The images given as input to the system must have an overlapping field of view to generate a panorama.

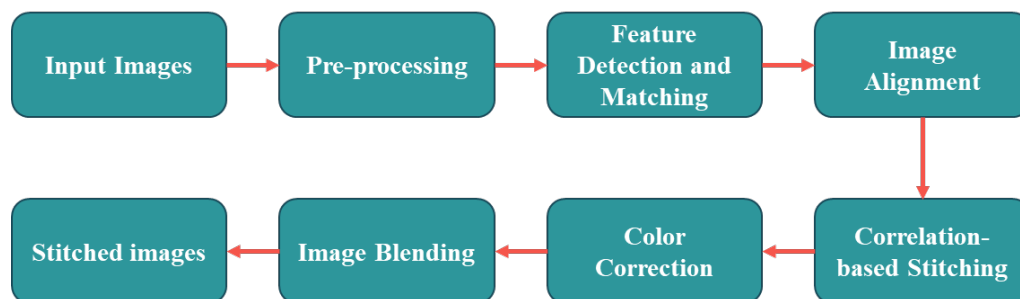


Fig. 4.3.2: Flow diagram of the underwater image stitching system

As mentioned earlier, underwater images suffer from poor contrast, which reduces the image's visual quality and makes post-processing a challenging task. Image enhancement is a vital process required for image stitching. The technique mentioned in [13] is implemented and used in this work for our system. The method follows a three-stage enhancement procedure. The first stage performs an RGB color equalization, where each color channel is balanced. The equalization keeps the dominant color channel constant, blue in this case, and enhances the other color channel by a gain factor. The second stage focuses on contrast correction of the RGB color space. The correction technique stretches the range of the intensity values to extend the desired range of values. The color correction uses the following equation

$$I_{out} = (I_{in} - c) \frac{(b - a)}{(d - c)} + a \quad \text{Eq. 4.3.1}$$

where  $I_{out}$  is the contrast corrected pixel value,  $I_{in}$  is the considered pixel value,  $a$  corresponds to the lower limit value,  $b$  is the upper limit value,  $c$  corresponds to the minimum pixel value in the image, and  $d$  is the maximum pixel value in the image. The third stage focuses on contrast correction of the HSI color space. The contrast corrected image is transformed to the HSI model, and contrast correction is performed on the Intensity and Saturation channels. A de-hazing algorithm and a guided filter are used to further improve the enhancement process. The de-hazing algorithm removes the clouding effects seen in underwater images. The guided filter [186, 187] used enhances the details of the image. The enhancement results are shown in Fig. 4.3.3.



(a)

(b)

Fig. 4.3.3: An illustration of the image pre-processing algorithm. (a) original image and (b) enhanced image

After the image pre-processing step, the images are passed to the feature detection and matching algorithm. The feature detection and matching algorithms described in [section xyz](#) are employed. For underwater stitching, the Hessian detector is used to detect features in the images, and the features of the images are matched by identifying their nearest neighbor. Fig. 4.3.4 displays the results of the detection and matching process. Finally, the images are aligned using the homography estimation technique (RANSAC) [188]. Homography plays a crucial role in estimating the scaling and rotation factor as the images can be acquired from different locations and varying distances.

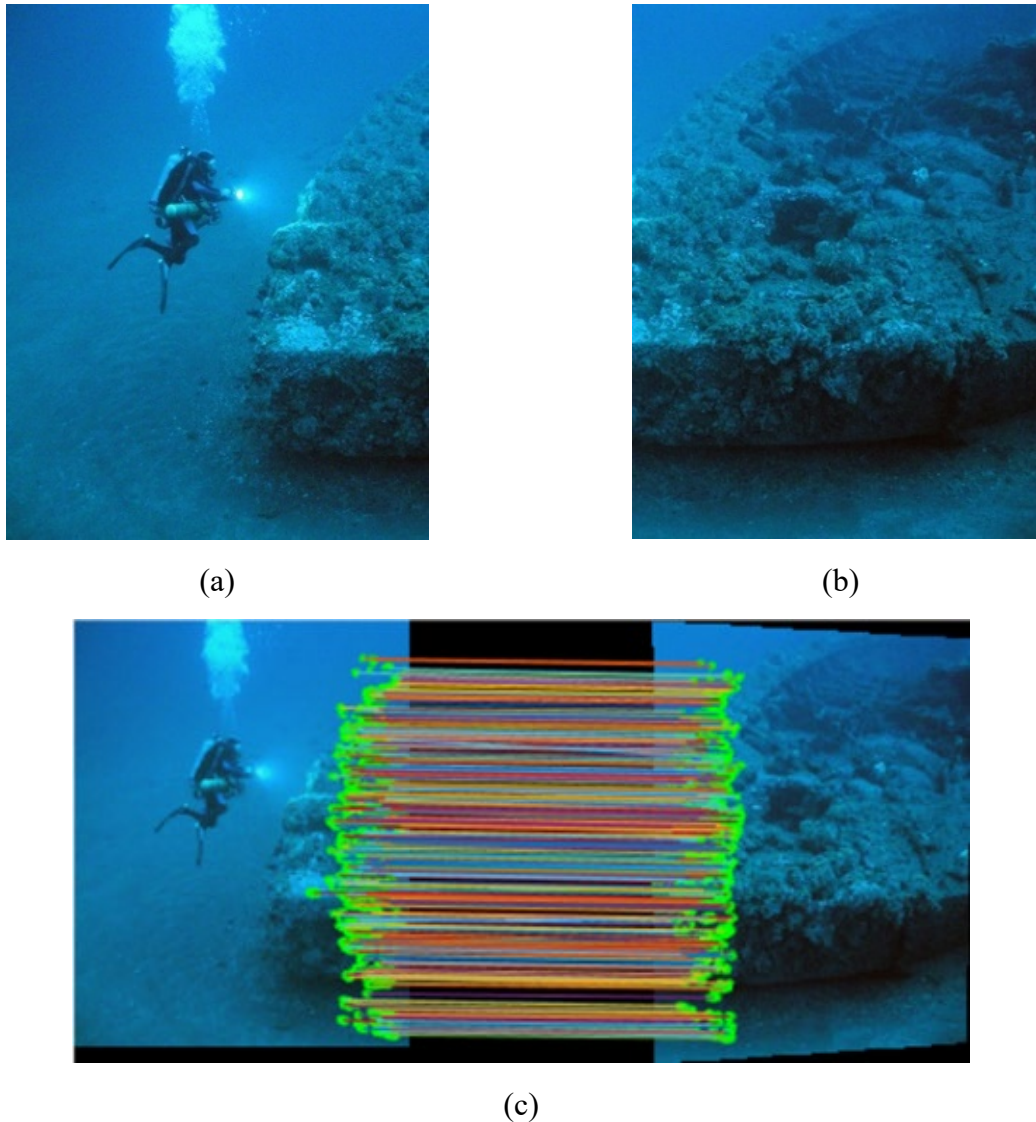


Fig. 4.3.4: Illustration of feature detection and matching.

Once the images are aligned, the overlapping regions of the two images are acquired. The maximal and minimal column coordinates define the overlapping region. Finally, the Alpha-Winzored and Weighted Rank Order Statistic based correlation algorithms are used to obtain a relationship between the two overlapping regions and to estimate the best seam line.

The correlation coefficient is a measurement of how closely two variables are related. It's also known as a statistical measure of the relationship between two or more variables. The various correlation methods are classified in terms of their strength and direction. The numerical value of a correlation determines its strength, and the direction of the correlation determines whether it is positive or negative [189]. Knowing the value of one variable will help predict the value of the other if the two variables are closely correlated. Classical correlation can be divided into three categories:

- Positive correlation: High 'X' values are often associated with high 'Y' values. As X rises, Y rises with it.
- Negative correlation: High 'X' values are related to low 'Y' values. Y lowers as X rises.
- No correlation: No consistent tendency for values on Y to increase or decrease as X increases.

Pearson's correlation coefficient, alternatively referred to as the classical correlation coefficient, measures the intensity of a linear or non-linear relationship between two variables. The Pearson's correlation coefficient is defined as

$$r(A, B) = \frac{\sum_{i=1}^n (A_i - \vec{A}) - (B_i - \vec{B})}{(\sum_{i=1}^n (A_i - \vec{A})^2 \sum_{i=1}^n (B_i - \vec{B})^2)^{0.5}} \quad \text{Eq. 4.3.2}$$

where  $\vec{A} = \frac{1}{n} \sum_{i=1}^n A_i$ , and  $\vec{B} = \frac{1}{n} \sum_{i=1}^n B_i$

The correlation coefficient is computed by individually measuring the variability of A and B. The covariance, or tendency for A and B to vary together, is then measured by the sum of their

products [21]. Depending on the direction of the relationship, the correlation can be positive or negative, and its magnitude can range from 0 to 1. The correlation coefficient is always between  $\pm 1$ . We have a perfect correlation only when all of the data points lie on a straight line. If the correlation coefficient is +1, the slope of the line is positive; if the correlation coefficient is -1, the slope of the line is negative. When there is no relationship between two variables, their correlation is 0. A value close to 0 indicates that the two variables have a random, nonlinear relationship.

The mathematical formula for computing the Alpha-winzorized and the weighted rank order statistic correlation is given by:

$$R = \frac{\sum_m \sum_n (A_{mn} - \vec{A})(B_{mn} - \vec{B})}{\sqrt{(\sum_m \sum_n (A_{mn} - \vec{A})^2)(\sum_m \sum_n (B_{mn} - \vec{B})^2)}} \quad \text{Eq. 4.3.3}$$

where,  $\vec{A}, \vec{B}$  are the  $\alpha$  – winzorized mean values given by:

$$\begin{aligned} \alpha \text{WinMean}(X_1, X_2, X_3, \dots, X_N; \alpha) \\ = \frac{1}{N} (\alpha N \cdot X_{(\alpha N+1)} + \sum_{i=\alpha N+1}^{N-\alpha N} X_{(i)} + \alpha N \cdot X_{(N-\alpha N)}); \end{aligned} \quad \text{Eq. 4.3.4}$$

The weighted rank order statistics is given by:

$$\begin{aligned} WROS(X_1, X_2, \dots, X_N; \text{weight}, \text{rank}) \\ = r^{th} \text{ order statistic of } \{a_1 \diamond X_1, a_2 \diamond X_2, \dots, a_N \diamond X_N\} \end{aligned} \quad \text{Eq. 4.3.5}$$



These correlations define the seamline along which the images are combined, as seen in Fig. 4.3.5 and Fig. 4.3.6.



Fig. 4.3.5: Generation of seam line using Alpha-winzorized correlation

To create a visually acceptable image, it is vital that the stitching procedure does not leave a trace. Factors such as exposure variance, gamma correction, and white balancing affect the image stitching process as it adds additional artifacts to the overlapping regions [190]. The issues mentioned above can be resolved by simple color correction. An example of the problem is seen in Fig. 4.3.7.



Fig. 4.3.6: Generation of seam line using weighted rank order statistics correlation.



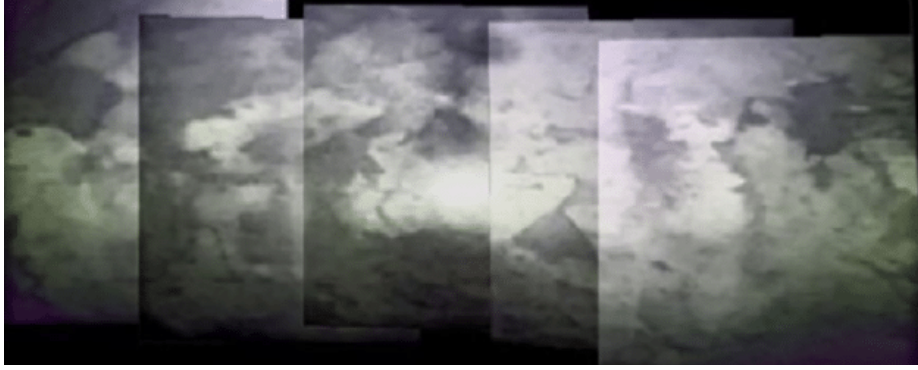


Fig. 4.3.7: Intensity differences encountered during image stitching. Image source [191]

There are two approaches to eliminating the visible seam line. The first approach finds the optimal stitch line and normalizes the energy along it. An energy function based on intensity or gradient difference can be applied. The second approach involves enhancing the pixel values of the two images, providing a smooth transition in the overlapping region. There exist several color correction algorithms [190, 192, 193]. The method described in [190] is used in the proposed system. The color correction technique eliminates the artifacts and color differences between the images by applying Poisson's equation. The equation is defined as:

$$I_{corr} = I_s + \psi \quad \text{Eq. 4.3.6}$$

where  $I_{corr}$  is the corrected image,  $\psi$  is the resulting correction function, and  $I_s$  is the overlapping image.

The proposed system uses the Laplacian pyramidal blending technique to overcome intensity differences along the seam of the overlapping regions [194]. A Difference of Gaussian pyramid is built for the source and target images. The image is downsized to different levels using a Gaussian filter. The Gaussian image is expanded to a lower level, and the difference between the

expanded image and the image at that lower level generates the Laplacian image. Once the pyramids of the two images are generated, the images are combined at different Laplacian levels.

#### **4.3.4. No Reference Measure for Underwater Images**

In the study of marine ecology, ocean engineering, and sea life monitoring, the quality of underwater images is a crucial factor [195]. Several algorithms for underwater image enhancement are based on generic quality measuring algorithms such as SNR, MSE, or visual inspection [196]. Measures like MSE and SNR require a reference/ground truth image to assess the quality of the image. Such measures cannot be used when in the absence of a ground truth image. On the other hand, visual inspection methods are biased, and the cost and resources required are high, especially when dealing with large amounts of data, as they are not automated. Hence there is a requirement for a stable, reliable, and cost and resource efficient method for evaluation of quality.

A new human visual system (HVS) based measure for underwater image quality is proposed in this work. The Combined Transform and Spatial domain Underwater Image Quality Measure (CTSUIQM) uses both transform domain and spatial domain measures and combines them using HVS. This new no-reference method differs from existing methods such as root mean square error (RMSE) and signal-to-noise ratio (SNR). Furthermore, the use of the HVS model, which is effective in dealing with human perception of underwater images [195], makes this method more suited to underwater applications.

In this work, opponent color space, as defined by Hering's opponent color theory [180], is used. Yellow – blue (YB), red-green (RG), and white-black (WB) are the color hues that are never perceived together by the human eyes and also have mutually exclusive sensory qualities [180]. YB, RG, and WB color components will be used for determining the color score.

The source image is decomposed into an n-level Gaussian and Laplacian pyramid. By performing pyramidal decomposition, the tri-stimulus intensity map is converted to transform domain coefficients, which are further used by the various spatial domain quality calculators like LogAMEE and edge ratio operator. The proposed algorithm requires the calculation of three sub-scores: color score, edge score, and contrast score. This section describes the components of each score and the procedure for calculating the final CTSUIQM score. The Gaussian pyramid calculates the color and contrast scores, while the Laplacian pyramid calculates the edge score. The Laplacian decomposition of an image is an over-complete decomposition based on the difference between successive low-pass filtered outputs. This produces the image's second-order derivative, which translates to edge information and is thus used to calculate the edge score. For all of the results presented in this work, the number of levels, n, for pyramidal decomposition is Set to 3.

As mentioned previously, RG, YB, and WB opponent color spaces will be used for color score and are formulated in Eq. 4.3.7, Eq. 4.3.8, and Eq. 4.3.9. The overall color score for underwater images is as shown in Eq. 4.3.10. The coefficients are chosen to be  $p = -0.0268$  and  $q = 0.1586$  and are obtained by the formulation given in [195].

$$RG = R - G \quad \text{Eq. 4.3.7}$$

$$YB = 0.5 * (R + G) - B \quad \text{Eq. 4.3.8}$$

$$WB = 0.33 * (R + G + B) \quad \text{Eq. 4.3.9}$$

The overall color score is formulated as:

$$Color\ Score = p * \sqrt{\mu_{\alpha, RG}^2 + \mu_{\alpha, YB}^2 + \mu_{\alpha, WB}^2} + q * \sqrt{\sigma_{\alpha, RG}^2 + \sigma_{\alpha, YB}^2 + \sigma_{\alpha, WB}^2} \quad Eq. 4.3.10$$

$$\mu_{\alpha, RG} = \frac{1}{K - R - S} \sum_{i=R+1}^{K-S} x_{RG,i} \quad Eq. 4.3.11$$

$$\sigma_{\alpha, RG}^2 = \frac{1}{K} \sum_{k=1}^K (x_{RG,k} - \mu_{\alpha, RG})^2 \quad Eq. 4.3.12$$

where  $\mu_{\alpha, RG}$  is the RS-Trimmed mean value of the RG color component. This definition of mean is used instead of the average value because it is resistant to Gaussian and impulsive noise [197]. The RS-Trimmed mean will behave like the average value when  $R = S = 0$ . Also,  $x_{RG}$  represents the intensity values of the RG color component, and  $K$  represents the total number of pixels. For all the results shown in this research,  $R$  and  $S$  values considered are  $0.2*K$ ,  $0.1*K$ , and  $0.05*K$  for each level.

The edge maps are constructed using the Laplacian pyramid. These edge maps are multiplied by the tri-stimulus input image to obtain an edge map for each color component. As a result, each color component image is composed entirely of edge pixels. These edge pixels are then fed into a window-independent Edge Score Estimator (ESE) to calculate the edge score. The newly introduced edge score estimator uses two different window sizes to map and find the sharpness of the edges accurately. It uses Edge Ratio Operator (ERO) for this purpose. The edge score is formulated as:

$$ESE = \frac{1}{\beta} \left( \frac{1}{k_1 k_2} \sum_{m=1}^{k_1} \sum_{n=1}^{k_2} \log(ERO) \right) \quad \text{Eq. 4.3.13}$$

$$ERO = \frac{|\mu_1 - \mu_2|}{\mu_1 + \mu_2} \quad \text{Eq. 4.3.14}$$

$$\mu_i = \frac{1}{w_1 w_2} \sum_{p=1}^{w_1} \sum_{q=1}^{w_2} x_{p,q} \quad \text{Eq. 4.3.15}$$

$$\beta = \max \left( \sum_{m=1}^{k_1} \sum_{n=1}^{k_2} \log \left( \frac{|\mu_1 - \mu_2|}{\mu_1 + \mu_2} \right) \right) \quad \text{Eq. 4.3.16}$$

The final edge score is formulate in Eq. 4.3.17, where  $\gamma_1 = 0.299$ ,  $\gamma_2 = 0.587$  and  $\gamma_3 = 0.114$  are selected based on the visual response of each color channel [195]

$$Edge\ Score = \sum_{c=1}^3 \gamma_c ESE(X_c) \quad \text{Eq. 4.3.17}$$

The LogAMEE score is used to estimate the contrast. To make the measurement independent of the window size, the score is calculated for three different window sizes, which are non-linearly combined. This nonlinear combination makes it more accurate than using a linear combination. The contrast score is defined as

$$Contrast\ Score = \logAMEE(Intensity\ image) \quad \text{Eq. 4.3.18}$$

$$\logAMEE = \frac{1}{k_1 k_2} \otimes \sum_{l=1}^{k_1} \sum_{k=1}^{k_2} \frac{I_{max,k,l} \ominus I_{min,k,l}}{I_{max,k,l} \oplus I_{min,k,l}} * \log \left( \frac{I_{max,k,l} \ominus I_{min,k,l}}{I_{max,k,l} \oplus I_{min,k,l}} \right) \quad \text{Eq. 4.3.19}$$

where the *Intensity image* is divided into  $k_1 k_2$  blocks, and PLIP operations are performed. This measure is the average entropy of Michelson's law for each block over the entire image [198]. PLIP operations provide an efficient nonlinear representation of relative image contrast that correlates to human perception [198].

The Underwater Quality Scores (UQS) for each  $n$  level are calculated using a linear superposition model as stated in [195]. Thus, the Color Score, the Edge Score, and the Contrast Score of each level are combined using Eq. 4.3.20. Additionally, using the advantages of nonlinear processing, the UQS of each level is combined to get the final CTSUIQM, as shown in Eq. 4.3.21. The choice of coefficients  $c_1, c_2, c_3$  is application-specific. For simulation,  $c_1 = 0.0282$ ,  $c_2 = 0.2953$  and  $c_3 = 3.5753$  are used and  $\delta_i = 1 \forall i$ .  $\alpha, \beta$ , and  $\gamma$  are parameters used to control the impact of each score on the final measure. Generally, a higher CTSUIQM score implies better quality.

$$UQS = c_1 * (Color\ Score)^\alpha + c_2 * (Edge\ Score)^\beta + c_3 * (Contrast\ Score)^\gamma \quad \text{Eq. 4.3.20}$$

$$CTSUIQM = \sqrt[n]{\prod_{i=1}^n (UQS_i)^{\delta_i}} \quad \text{Eq. 4.3.21}$$

### 4.3.5. Experimental Results

This section provides the results obtained using the Alpha Winzorized and Weighted Rank Order Statistic Correlation system. For our simulation, underwater images were acquired from [177]. Due to the unavailability of a database, the images used in this work were manually cropped and stitched back together. The advantage of doing so is the availability of a ground truth. The results acquired after stitching are compared subjectively and quantitatively to the ground truth. The final stitched images and their respective ground truths are shown in Fig. 4.3.8 and Fig. 4.3.9.

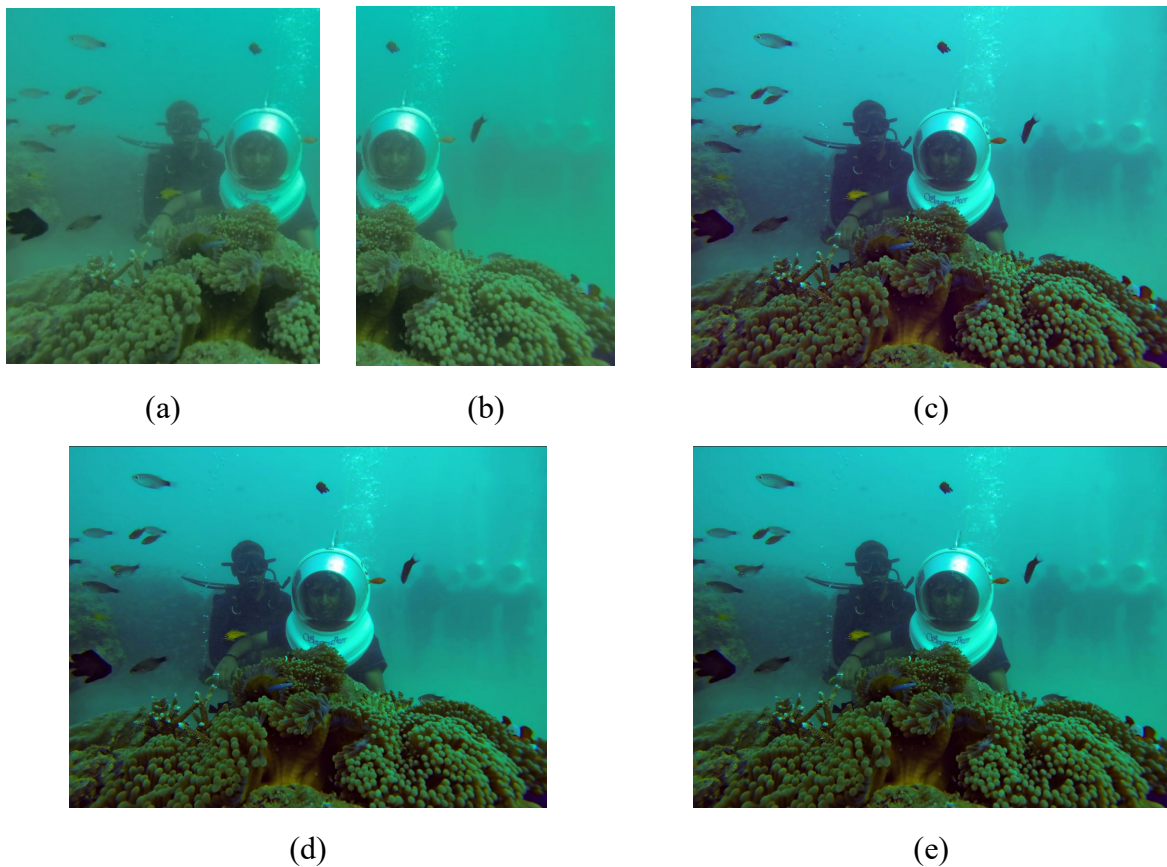


Fig. 4.3.8: Visual analysis (a), (b) Original images, (c) ground truth, (d) alpha-winzorized result, (e) weighted rank order statistic results.

As seen from Fig. 4.3.8 and Fig. 4.3.9, the final stitched image is a replica of its ground truth, indicating a perfect stitch. The color distribution is uniform and the color difference between the fusion region has been handled. To further evaluate the quality of the stitched images, non-reference quality measures such as UICM, log AMEE, and the proposed CTSUIQM have been employed. These measures help evaluate images based on colorfulness, contrast, and edge information. The stitched images are compared to their respective ground truths. The UICM measure is defined as:

$$UICM = -0.0268 \sqrt{\mu_{\alpha, RG}^2 + \mu_{\alpha, YB}^2} + 0.1586 \sqrt{\sigma_{\alpha, RG}^2 + \sigma_{\alpha, YB}^2} \quad \text{Eq. 4.3.22}$$

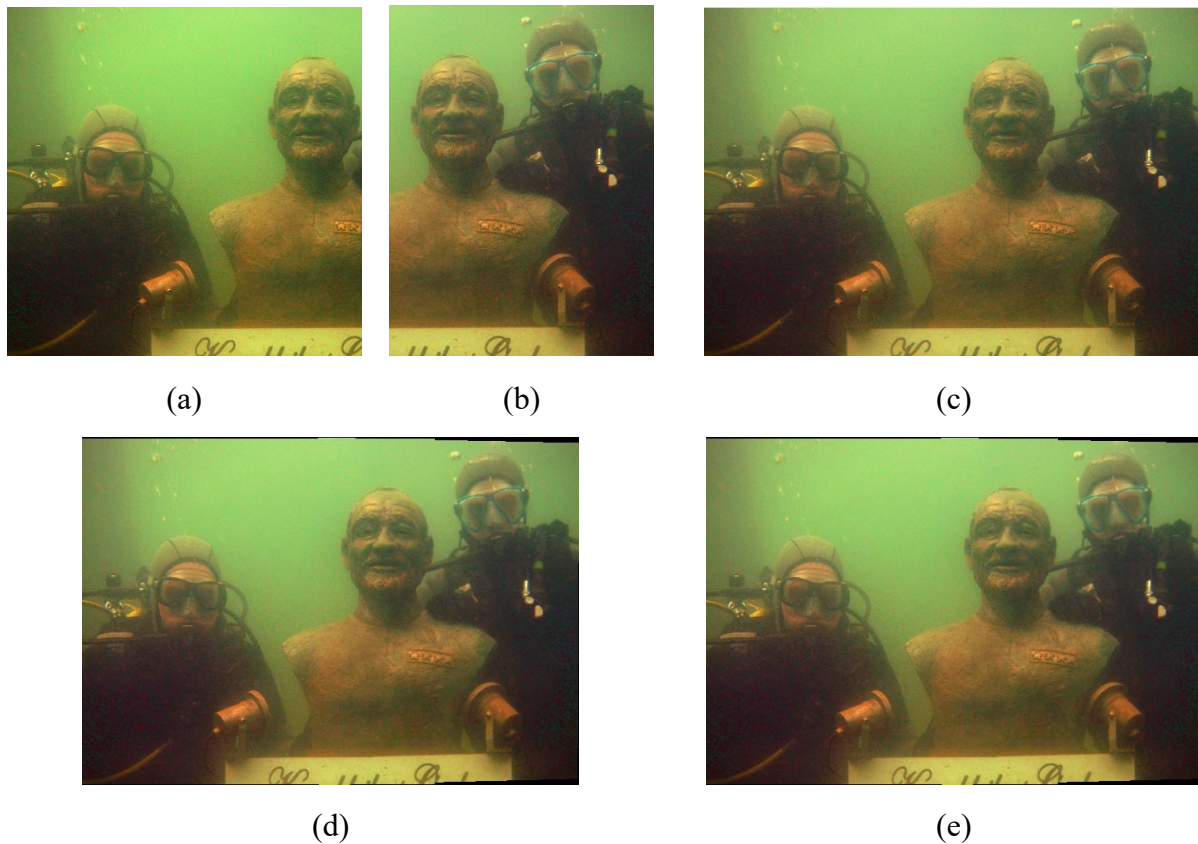


Fig. 4.3.9: Visual analysis (a), (b) Original images, (c) ground truth, (d) alpha-winzorized result, (e) weighted rank order statistic results.



where  $\mu_{\alpha, RG}$  and  $\mu_{\alpha, YB}$  are the respective alpha trimmed mean of RG and YB opponent color components.  $\sigma_{\alpha, RG}^2$  and  $\sigma_{\alpha, YB}^2$  are the respective alpha trimmed variance of RG and YB opponent color components. LogAMEE is a measure of image contrast as formulated in Eq. 4.3.19. It is similar to the AMEE measure, but PLIP arithmetic operations replace the arithmetic operations.

The measures discussed above are tested on Fig. 4.3.8 and Fig. 4.3.9. Each measure is tested on the stitched image and its resulting ground truth. The results are tabulated as shown in Table 4.3.1.

| Table 4.3.1: Performance analysis of the image stitching algorithms |                   |                   |                   |                   |                   |                   |
|---------------------------------------------------------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| Measures                                                            | Fig. 4.3.8<br>(c) | Fig. 4.3.8<br>(d) | Fig. 4.3.8<br>(e) | Fig. 4.3.9<br>(c) | Fig. 4.3.9<br>(d) | Fig. 4.3.9<br>(e) |
| UICM                                                                | 1.1195            | 1.1043            | 1.1038            | 1.2420            | 1.0860            | 1.0807            |
| logAMEE                                                             | 0.7871            | 0.8081            | 0.8062            | 1.1347            | 0.8230            | 0.8228            |
| CTSUIQM                                                             | 3.9492            | 3.3489            | 3.2883            | 2.5378            | 7.6760            | 7.6745            |

The results inferred from Table 4.3.1 imply that the quality of the stitched image is similar the ground truth. The difference in the results is due to the color correction method. Quantitative and subjective analysis shows that the final stitched image is as good as its ground truth.

#### **4.3.6. Extended Application: Finger image stitching**

Biometric authentication systems are vital in forensics and civilian applications [7]. It verifies an individual's identity by examining several characteristics such as their voice, fingerprints, face, and signature [199]. Fingerprint biometrics are the most frequently utilized in comparison to other biometric traits since they are a very reliable technique [200]. The most commonly used methods of fingerprint registration are based on the contact of a finger with a surface. It is divided into two stages: enrollment and authentication. The enrolment stage involves acquiring and processing a person's fingerprints to extract a set of features [200]. The acquired feature set is the individual's identity, which is kept in a dataset. During the authentication procedure, a person's fingerprint data is collected once more and compared to other publicly available datasets. Fingerprints are represented based on local landmarks called minutiae, which include ridge endings, ridge bifurcations, islands, etc. [201].

The traditional method for registering a fingerprint involves applying pressure to a platen surface, which acquires a set of features for authentication. Such systems are prone to nonlinear plastic distortions, accumulation of dirt on the sensors, and sensor noise [178, 202]. To overcome these challenges, touchless fingerprint sensing technologies were introduced that generated an end-to-end view of the finger. These systems were less expensive and had a faster computation time. However, one of the significant issues with finger image stitching is the visibility of a seam line. This can occur when the correlation method used is weak, or there is a color difference between the images. The proposed correlation methods overcome the issue of a visible seamline, and the results are presented in Fig. 4.3.10.

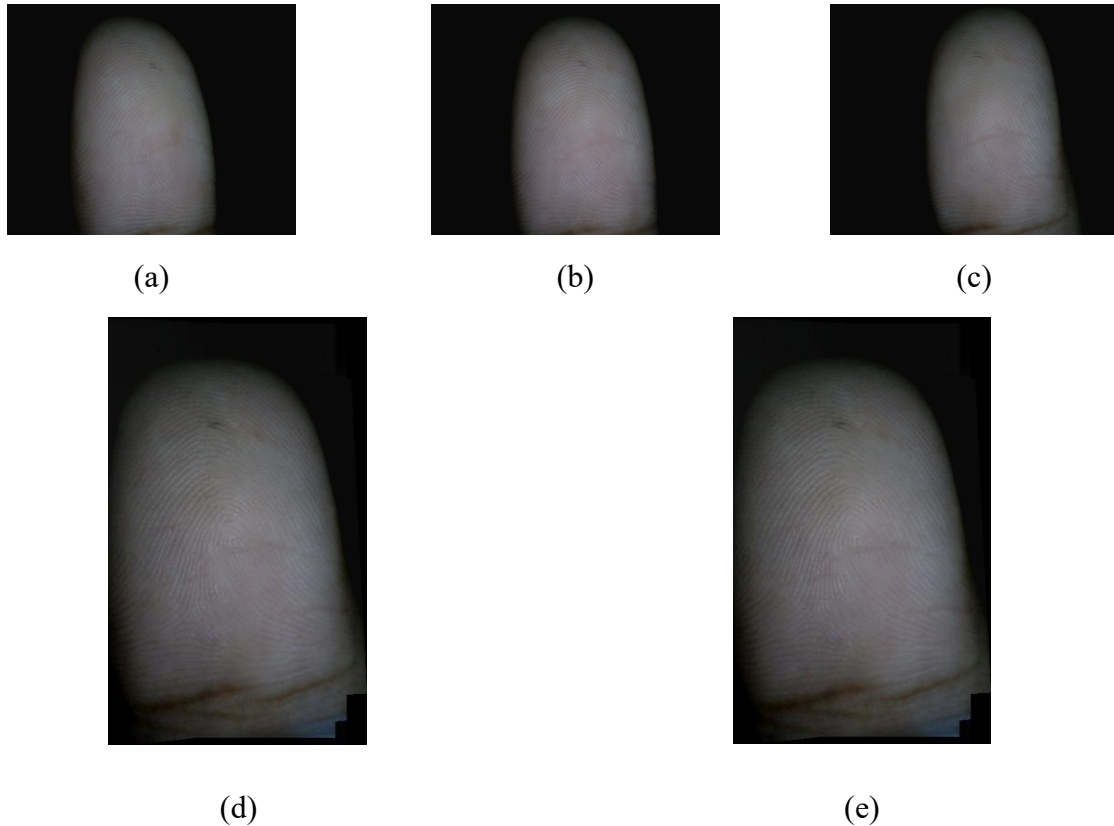


Fig. 4.3.10: The results of the stitching algorithm applied to finger images. (a) left image, (b) center image, (c) right image, (d) alpha-winzorized result, and (e) weighted rank order statistics result.

#### 4.3.7. Summary

The chapter presents an overview of the various image feature detectors and their application to panoramic radiograph matching, color transfer, and underwater image stitching. The matching algorithm produced successful results in comparing antemortem and postmortem images. The Multiscale Hessian detector performed exceptionally in extracting matching feature sets. The color transfer section introduces a new pixel-based color transfer system to recolor nighttime imagery. The system can compensate for image registration issues encountered during acquisition from different image sensors. The performance of the proposed system is evaluated

using the RGB-NIR Scene dataset and the AMOS dataset. The method can accurately transfer the color information from the reference to the target image. To further assess the method, the results are visually compared to those acquired using [153] and [152]. Our method closely mimics the daytime images compared to the state-of-the-art methods. It maps colors accurately in the visible spectrum and performs well in cross-domain color transfer. Finally, a new image stitching system that combines multiple underwater images to generate a panoramic view of the environment is introduced. The system comprises of three major sections: image enhancement, image registration, and image stitching. The image enhancement section eliminates the clouding effect and restores color to a certain extent. Feature detection and matching are implemented using SIFT. The advantage of using SIFT features is its invariance to rotation and scaling. These features were further used to estimate the homography. The proposed Alpha-Winzorized and Weighted Rank Order Statistic Correlation helped generate seamless panoramic images. The system's performance is evaluated using UICM, logAMEE, and CTSUIQM. Based on the results acquired by the measures, it can be inferred that the ground truth and stitched images are similar.

## **Section 2: Image Analytic**

## **Chapter 5. Image Segmentation**

### **5.1. Motivation**

There are more than 400,000 new cases of kidney cancer each year, and surgery is its most common treatment. Accurate segmentation and characterization of kidneys and kidney tumors is an important step in quantifying the tumor's morphological details to monitor the progression of the disease and improve treatment planning. Segmentation of kidney tumors in CT images is a challenging task due to the low contrast, irregular motion, diverse shapes, and sizes. Furthermore, manual delineation techniques are extremely time-consuming and are prone to errors due to the variability between different specialists. The literature provides the application of 3D Convolutional Neural Networks (CNNs) for the segmentation of kidneys and tumors. While effective, 3D CNNs are computationally expensive. The applications of a novel 2D CNN architecture to segment kidneys and tumors from CT images is presented. The proposed architecture uses features from enhanced images to improve the segmentation performance. Quantitative and qualitative analysis of the proposed model on the KITS19 dataset shows the improvement against recent state-of-the-art architectures.

## 5.2. Introduction

Kidney cancer represents 2.4% of the cancers worldwide, making it the 14th most common cancer [203, 204]. According to a recent study, a total of 431,288 people were diagnosed with kidney cancer in 2020 [203]. Kidney cancer is caused by malignant kidney cells that develop into tumors and appear in the kidney as tiny tubules [205]. Previous research has demonstrated that cancers can be treated more efficiently if found early, while still being contained within the renal capsule. Small renal masses are typically not aggressive and are slow-growing with low malignant potential [206]. However, if left untreated, these tumors can grow to be quite large and even spread throughout the body [206]. Renal tumors were traditionally treated by removing both the tumor and the affected kidney. But advances in surgery combined with earlier tumor detection have shifted kidney cancer treatment toward more conservative nephron-sparing procedures [207]. While imaging techniques such as CT enable precise tumor detection, the evaluation continues to rely on nephrometry grading methods [208]. Hence, there is an urgent need for computer-aided tumor delineation techniques that can assist radiologists in precisely evaluating and quantifying renal tumors.

In many screening centers and hospitals across the United States, Computer-Aided Diagnosis (CAD) has recently been part of routine clinical work and is commonly used in the identification and differential diagnosis of a variety of anomalies [209-212]. Almost all radiological investigation necessitates the interpretation by the radiologist to arrive at a diagnosis. Although efficient, human inspection requires specialized training and a radiologist's time, which are increasingly expensive [213]. Human examination gives a subjective judgment, which may fluctuate from one practitioner to another, which results in a degree of uncertainty in quantitative

estimation. Furthermore, inspection results could be affected by factors such as fatigue and distraction due to cognitive overload. In order to overcome subjective disparity and observer variability, CAD systems were developed to improve the capability of radiologists and facilitate efficient interpretation and enhanced diagnostic accuracy. It is critical to have accurate information about the kidney and tumor positions when evaluating and planning surgery. Thus, computer-aided automatic segmentation of renal tumors will assist urologists in determining the optimal treatment and minimizing any complication risks.

Automatic isolation of renal tumors in CT images is a very challenging task. These challenges include (1) low-intensity contrast between the tumor and the surrounding tissues and organs, (2) unclear and blurred boundaries of tumors, (3) variations in size, shape, and location of the tumors, and (4) presence of noise or the poor quality of images. The quality of the CT images is dependent on the X-ray dosage. Higher dosage results in a high-quality reconstruction of the image; however, this can cause serious harm to the human body [214]. Hence, lowering the dosage and ensuring CT image quality becomes crucial. Deep learning is one such area that has begun to achieve state-of-the-art (SOTA) results in many problem domains [79, 215, 216]. In the field of medical image segmentation, deep learning-based image segmentation methods have shown promising results. Deep learning has some advantages in segmentation accuracy and speed when compared to traditional machine learning and computer vision methods. Four main factors contribute to their success: To begin with, advances in neural networks are the main reason for deep learning's incredible success and its ability to decode and learn high-level features. Second, with the advent of GPUs and GPU-computing libraries, models can now be trained much faster when compared to training them on CPUs. Third, publicly available databases can be utilized for



training, allowing researchers to train and develop novel architectures. Finally, various effective optimization strategies have helped with the overall success.

Deep Convolutional Neural Networks (DCNNs) have achieved great success in the field of medical image segmentation, such as the brain [217-219], lung [220], pancreas [221, 222], and prostate [223]. Zhang [224] investigated the use of multimodal images to improve image segmentation and demonstrated its superiority over single modality inputs. Prasoon [225] applied three separate CNNs to segment tibial cartilage in low field MR images. Moeskops [226] employed a 2.5D architecture to segment multiple organs from different image modalities using a single network design. Kamnitsas [227] developed a multiscale dual-path 3D CNN to segment brain tumors in MRI images. Kleesiek [228] developed a 3D CNN architecture that used four modalities and a thresholding function for brain extraction in MR images. Cicek [229] developed a 3D U-net model to generate dense volumetric segmentation from 2D sparse slices. Zeng [230] proposed a deeply supervised 3D U-net-like fully convolutional network for segmentation of proximal femur in 3D MR images. Milletari [231] proposed a 3D image segmentation approach based on a volumetric, fully convolutional neural network trained on MRI volumes depicting prostate, and learns to predict segmentation for the whole volume at once. Yu [232] developed a fully convolutional residual network for pixel-wise melanoma recognition and segmentation. Dou [233] introduced a 3D deep supervision mechanism during the learning process to combat potential optimization difficulties and acquire a much faster convergence rate and more powerful discrimination capability. Feng [234] employed a CNN for the segmentation of lung nodules that learns discriminative regions from the activation maps of convolution units at different scales, and identifies the true nodule location with a novel candidate-screening framework.

Similarly, many kidney tumor segmentation techniques have been proposed in the literature. Andrzej [235] proposed a hybrid level set method with elliptical constraints to segment kidney tumors from CT images. Qian [236] presented a cascaded trainable segmentation model that used two orthogonal non-squared patches that trained two sub-models in a cascaded manner to segment the kidney and tumor regions from CT images. Xiaoshuai [237] proposed a deeper U-Net architecture that employed dilated convolutional blocks to replace the traditional pooling operations. Further, a hybrid loss of dice and weighted cross-entropy is used to guide the model to focus on voxels close to the boundary. Luana [238] applied the DeepLabv3+ 2.5D model to perform the kidney tumor segmentation. Zhao [239] presented a multi-scale supervised 3D U-Net, MSS U-Net, to automatically segment kidneys and kidney tumors from CT images. Yang [205] suggested a 3D fully convolutional network with a pyramid pooling module to improve segmentation in extremely large-sized kidney tumors. Yi [240] offered a three-step automated kidney tumor and cyst segmentation approach based on 3D U-Net that included segmenting the kidney and tumor from low-resolution CT, locating each kidney and fine segmenting the kidney, and finally extracting the tumor and cyst from the segmented kidney. Santini [241] proposed an automatic, multi-stage, 2.5D deep learning-based segmentation approach based on combining ResNet with Res-U-Net architectures. Li [242] proposed deep learning network architecture based on the residual and 3D UNet architecture for the kidney, renal tumor, and renal cyst segmentation.

Although 3D CNNs have a high segmentation capacity, they are complex and have a high computational cost. Furthermore, training can be challenging when dealing with volumetric data due to variance between the target and neighboring voxels. This chapter proposes a dual decoder-based 2D CNN architecture to segment kidneys and tumors from CT images. The architecture uses features from enhanced images to further improve the network's performance.

### 5.3. Proposed Work

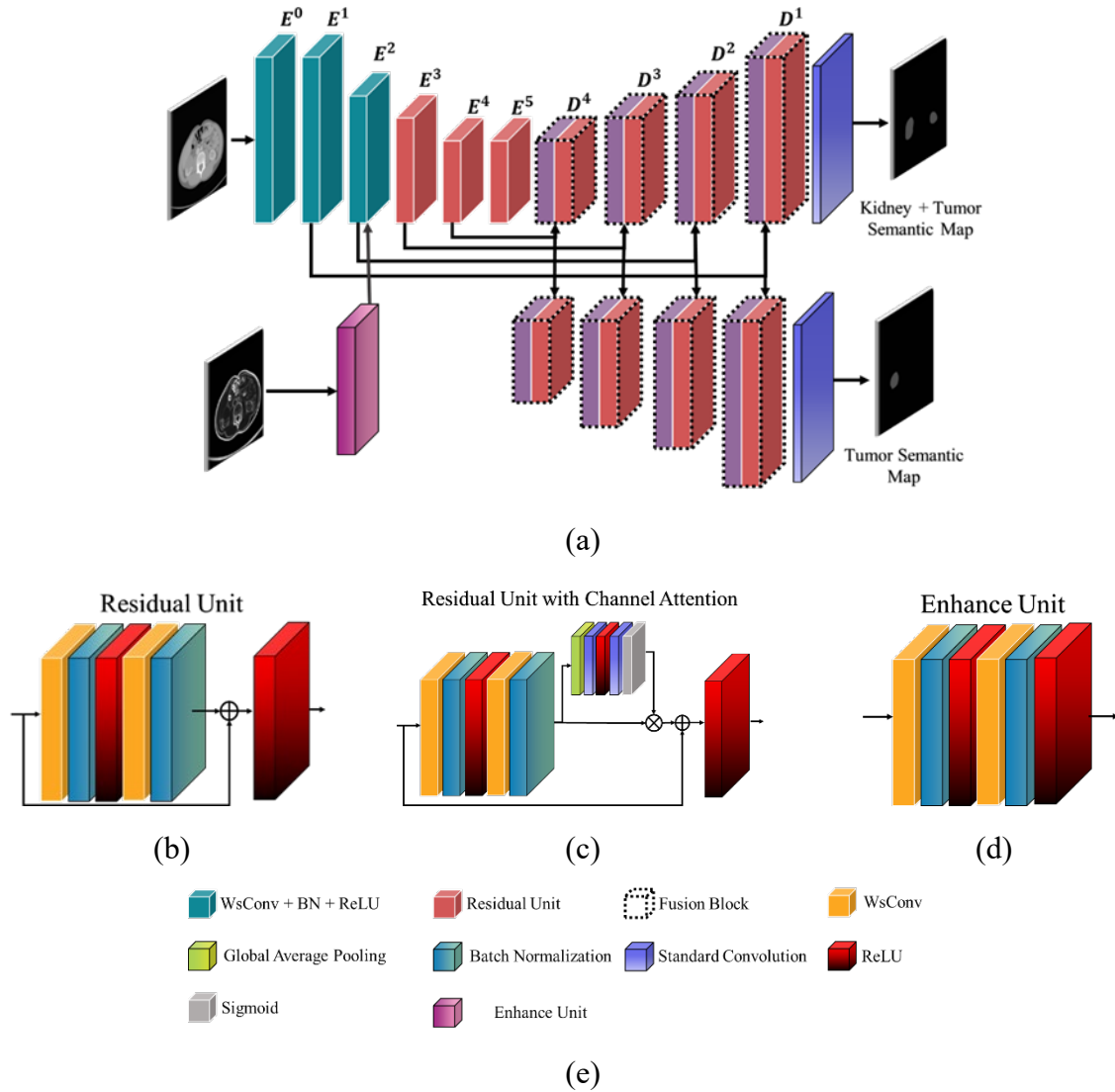


Fig. 5.3.1: The network architecture of the proposed KTNNet - (a) provides an overall structure of KTNNet with the encoder and a dual decoder network, (b) visualizes the standard residual network proposed by He [243], (c) visualizes the residual network with a channel attention mechanism, (d) visualizes the enhance unit utilized to boost the feature extraction from the encoder network to emphasize more on significant features, and (e) is the legend for this flow diagram.

This section contains an overview of the proposed dual decoder-based kidney and tumor segmentation network (KTNet). A basic flow diagram of the network is provided in Fig. 5.3.1(a). This network showcases a function  $f$  that can link each pixel from an input image  $I$  of any arbitrary size  $(m, n)$  to two semantic maps that represent kidneys and tumors with the same dimension. Following this, the novel KTNet comprises of an encoder network and two decoder networks accompanied by final pixel-wise classification layer.

This network aims at capturing, optimizing, and sharing the feature representation encompassed by the encoder with dual stage decoders, thereby producing semantic maps that are spatially consistent with sharp boundaries. Additional details of these components are provided in further subsections.

### 5.3.1. Encoder Network

The primary objective of KTNet is to develop a model that is position sensitive across two decoders and performs robust pixel-wise classification. Thus, the network builds on an encoder network that follows the design rule of LeNet-5 [244]. To capture features with varied representation capacities, this design rule adapts gradual condensation mechanism wherein the spatial size of features is reduced as they transition from high-level to low-level representations. The encoder network comprises of two feature groups which can be denoted as  $\hat{E}^\theta$  where  $\theta = 0, 1, 2$  and  $E_\lambda^\theta$  where group  $\theta = 3, 4, 5$  and  $\lambda$  indicates the number of the residual layer in that group and ranges from  $1, 2, \dots, n$ . The first feature group  $\hat{E}^\theta$  comprises of a series of weight standardized convolutional [245] (WsConv) layer followed by batch normalization (BN) and ReLU activation layer, while  $E_\lambda^\theta$  comprise of residual units. The weight standardized convolutional block was utilized to accelerate micro-batch training. For effective BN training, a large-batch training is

required, but as the batch size is decreased, BN performs poorly. Various standardization techniques have been proposed to overcome this; however, they do not rely on batch knowledge, thereby reducing the performance when compared to BN. Thus, in KNet, weight standardized convolutional layers are utilized to overcome the mini-batch problem and attain superior results. The  $E^0$  layer focuses on converting the representation from the image domain to the feature domain. The WsConv layer employs a  $3 \times 3$  kernel and produces 16 feature maps.

| Table 5.3.1: Architectural details of the encoder network |           |            |            |            |             |             |
|-----------------------------------------------------------|-----------|------------|------------|------------|-------------|-------------|
| Group name                                                |           | $E^1$      | $E^2$      | $E^3$      | $E^4$       | $E^5$       |
| Output size considering 512 x 512 input image             |           | 512 x 512  | 256 x 256  | 128 x 128  | 64 x 64     | 32 x 32     |
| [kernel size, # feature maps, stride, dilation]           | $E_0$     | [3,16,1,1] | [3,32,2,1] | [3,64,2,1] | [3,128,2,1] | [3,128,1,2] |
|                                                           | $E_{1-n}$ | -          | -          | [3,64,1,1] | [3,128,1,1] | [3,128,1,2] |

The basic structure of the residual layer used in  $E^{3-5}$  in the encoder network is formulated in Eq. 5.3.1 and can be visualized in Fig. 5.3.1 (b).

$$\Psi_{\lambda+1} = R(I(\Psi_{\lambda}) + \Omega(W_{\lambda} * \Psi_{\lambda} + \mathbb{B}_{\lambda})) \mid \{W_{\lambda} = (\varpi_{\lambda,k} : k = 1 \leq k \leq K)\} \quad \text{Eq. 5.3.1}$$

where  $\Psi_{\lambda}$  is the input feature map for the  $\lambda^{th}$  residual layer,  $W_{\lambda}$  and  $\mathbb{B}_{\lambda}$  are the associated set of weights and biases, respectively.  $\Omega$  denotes the combination of layers such as WsConv→BN→ReLU→ WsConv →BN,  $R$  denotes the ReLU activation function,  $I$  is the identity map, and  $K$  is the number of filters. In groups  $E^{2-4}$  the first layer performs downsampling by striding instead of max pooling since max pool layers lead to high amplitude, high-frequency

activations in the subsequent layers, which might increase gridding artifacts [246, 247]. As generating semantic maps rely on spatial information, it is necessary to comprehend the information and generate a semantic map with finer details.

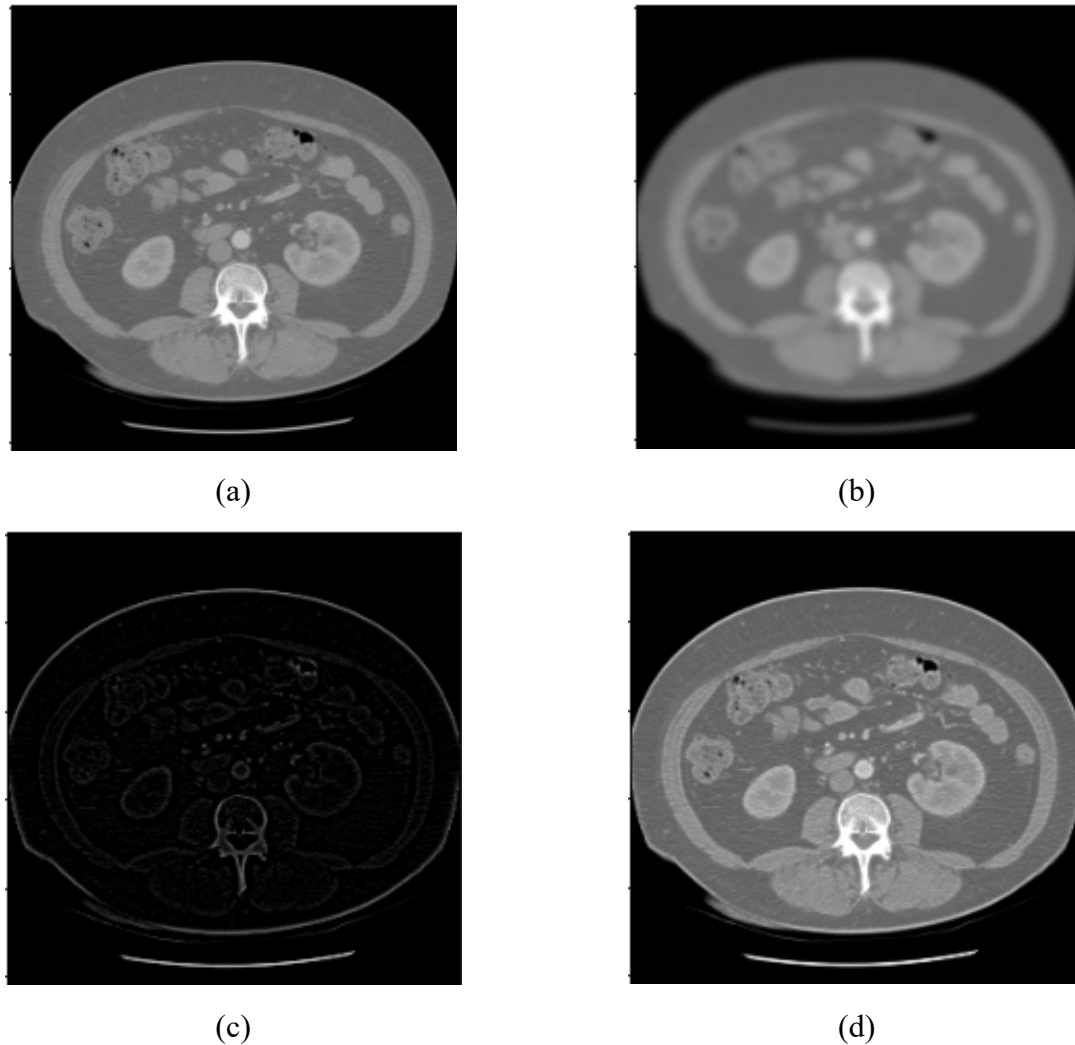


Fig. 5.3.2: Results of the edge enhancement process. (a) original image, (b) gaussian blurred image with  $\sigma = 2$ , (c) the extracted edge information, and (d) the enhanced image. As seen, the additional edges improve the resolution of the image.

However, further downsampling of features may lead to loss of information and eliminating downsampling may cause a detrimental effect on the receptive field in the following

layers. As a result, dilated convolution is used to address this limitation by adjusting the receptive fields of feature points without reducing the resolution of feature maps [246]. The architectural details of the encoder network are provided in Table 5.3.1.

To improve the representation capability of the encoder network, an edge-enhanced version of the original image is utilized. This edge-enhanced image is obtained by applying a gaussian filter and extracting the edge information from the image. These edges are added back to the original image, thereby enhancing the crucial information required to generate sharp semantic maps. This process can be formulated as Eq. 5.3.2. This image is passed through  $WsConv \rightarrow BN \rightarrow ReLU \rightarrow WsConv \rightarrow BN \rightarrow ReLU$  to generate 32 feature maps and is added to the output of  $E^2$ .

$$\hat{I} = I + (I - (G_{\sigma} \otimes I))$$

$$G_{\sigma} = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$
Eq. 5.3.2

where  $\otimes$  denotes convolution and  $\sigma$  was set to 1 empirically. A visual illustration of this mechanism is shown in Fig. 5.3.2.

### 5.3.2. Dual Decoder Network

The objective of a decoder is to upsample the encoded features to perform effective pixel-wise classification to generate sharp and spatially sound semantic maps. As KTNNet aims at generating semantic maps for kidney and tumor separately, two different decoders are employed. For simplicity of exposition, the architecture of one decoder is described as both the decoders are identical. The decoder network comprises of feature groups, which can be denoted as  $D_{\lambda}^{\theta}$ , where group  $\lambda = 1, \dots, 4$ , and  $\theta$  indicates the number of the residual layer in that particular group and is

set to 1. The feature layers of the encoder and decoder network are symmetric across the fusion block, as shown in Fig. 5.3.1 (a). Each decoder group in  $E^\theta$  mainly consists of upsampling layers, compression layers, and residual layers. The input to these groups is a combination of the feature maps from the corresponding encoder group and the previous decoder group. This aids in the propagation of contextual information to higher layers with a better resolution. The upsampling layer consists of transposed convolutions with kernel size  $2 \times 2$  and stride  $2 \times 2$ . This increases the resolution of the feature maps by a factor of 2. The compressing layer consists of  $\text{WsConv} \rightarrow \text{BN} \rightarrow \text{ReLU}$ , wherein the kernel size of WsConv is  $1 \times 1$ . This is used to compress the feature dimensions by a factor of 2. The compressed feature maps are then fed to the residual layers for further processing. Finally, the output of the group  $E^1$  is connected to a WsConv layer with kernel size  $1 \times 1$  to generate a single channel output.

### 5.3.3. Channel Attention (CA) Mechanism

Most deep learning-based semantic map generation networks treat all the feature maps equally, which is not often the case in many real-world situations. When comparing the feature maps generated by these residual layers, it is possible that a few of the features contribute much more than the rest. Additionally, the learned filters in the residual layers have a limited receptive field, and each filter output makes only marginal use of contextual information beyond the subregions. Therefore, it is necessary to implement a feature-recalibration method that provides attention to critical feature maps more than the rest. Various researchers have proposed preliminary work on applying attention to deep neural networks [248-250], however, they are not yet fully developed for low-level visual tasks such as medical semantic map generation techniques.



The primary purpose of this approach is to assign distinct values to individual channels based on their interdependencies between each convolution layer. This can be accomplished by exploring the global spatial information of each feature map. The channel attention mechanism can be formulated, as shown in Eq. 5.3.3.

$$\hat{\Theta} = \sigma \left( \mathbb{W}_{\uparrow} \left( R \left( \mathbb{W}_{\downarrow} \left( \frac{1}{(H \times W)} \sum_{m=0}^{H-1} \sum_{n=0}^{W-1} (\Theta) \right) \right) + \mathbb{B}_{\downarrow} \right) + \mathbb{B}_{\uparrow} \right) \quad \text{Eq. 5.3.3}$$

where  $\Theta = [\Theta_1, \Theta_2, \dots, \Theta_{\kappa}]$  is input feature map with  $\kappa$  number of channels/feature maps and  $H \times W$  dimensions,  $\mathbb{W}_{\downarrow}[\mathbb{B}_{\downarrow}]$  denote weight [bias] of the compression convolution, which reduces the dimension by a factor of  $\delta$ ,  $\mathbb{W}_{\uparrow}[\mathbb{B}_{\uparrow}]$  denote weight [bias] of the expansion convolution, which increases the dimension by a factor of  $\delta$ ,  $R$  denotes the ReLU activation function, and  $\sigma$  is the sigmoid activation function. The compression layer output can be realized as the fusion of local descriptors whose statistics express the entire feature map [248]. The channel attention mechanism comprises of standard convolutions with kernel size  $1 \times 1$ . This aids in learning the nonlinear interaction between the channels and ensures that emphasis is given to channels with higher distinctive features [248]. As  $\kappa$  in the encoder and decoder network varies, the gating mechanism adapts to accommodate these changes. The factor  $\delta$  is a hyperparameter, which varies the capacity of the gating mechanism. An illustration of this mechanism is illustrated in Fig. 5.3.1 (c).

## 5.4. Experimental Results

This section provides the performance evaluation of the KNet. After outlining the experimental settings, chosen datasets, and training details, the performance comparisons with SOTA methods are provided to demonstrate the effectiveness and generality of KNet.

### 5.4.1. Dataset

In this work, the KITS [251] CT database was used to train and validate our network architecture. The dataset provides a collection of multi-phase CT scans, segmentation masks, and detailed clinical outcomes for 300 patients who underwent partial or radical nephrectomy for one or more kidney tumors at the University of Minnesota Medical Center between 2010 and 2018 [251]. 210 patients were selected for the training (178), validation (16), and testing (16) set. In total, 13593 images were used for training and 1517 images were used for both testing and validation. The included volumes are characterized by different in-plane resolutions ranging from 0.437 to 1.04 mm, with a slice thickness varying among cases from a minimum of 0.5 mm up to 5.0 mm. CT images and masks were provided in a NIFTI format for each volume. The dataset was further modified to eliminate a few mislabeled volumes, as mentioned in [252]. Case IDs 15 and 17 were confirmed to be mislabeled by the challenge organization. Case IDs 23, 68, 125, and 133 in the training phase for reasons mentioned in [253] were eliminated. Fig. 5.4.1 displays a test case of the KITS dataset.

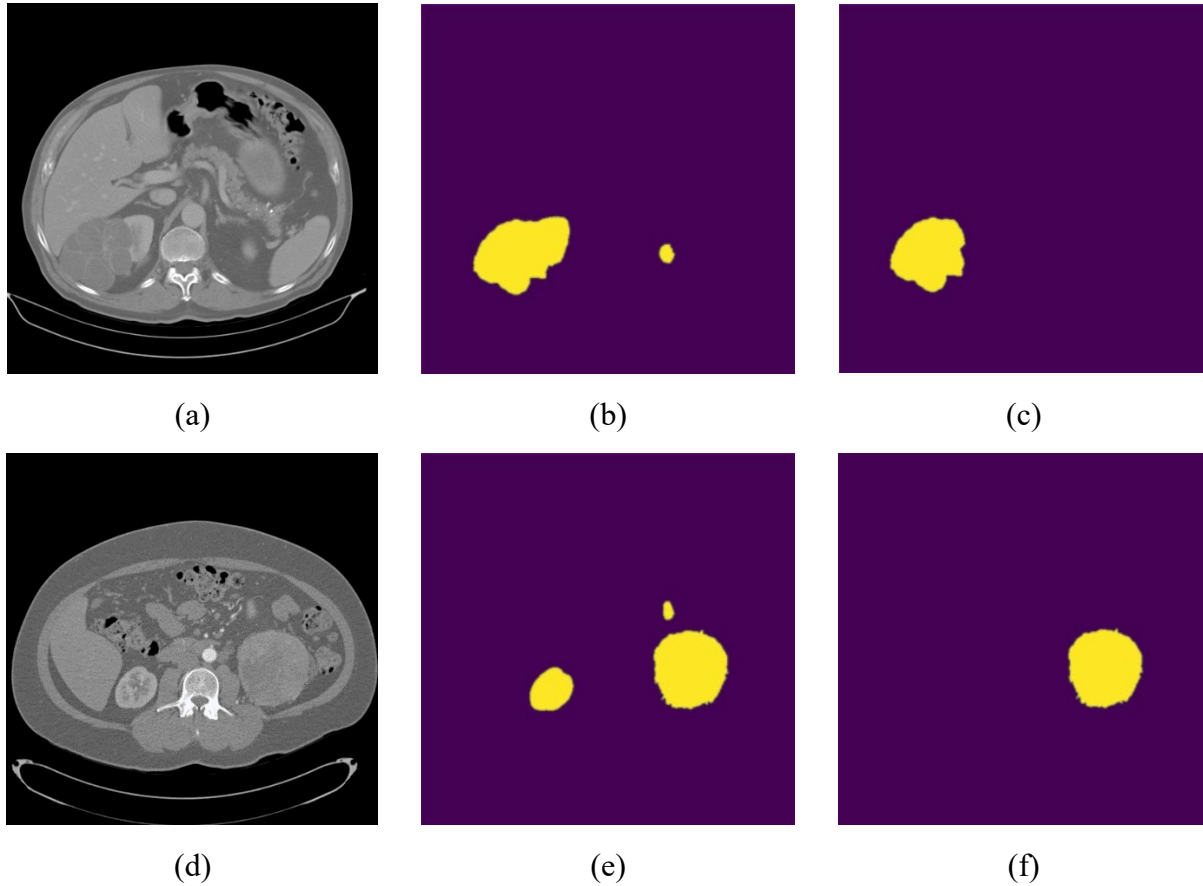


Fig. 5.4.1: Example images from the KITS dataset where (a) and (d) are the input CT images, (b) and (e) are the ground truth masks of the Kidney+Tumor regions, and (c) and (f) are the ground truth masks of the Tumor region

### 5.4.2. Training

For training, gray input images along with the corresponding semantic map was considered. The input images were augmented by random cropping to a size  $480 \times 480$ , random scaling in the range of  $(0.5, 2)$ , and random horizontal flipping along with the corresponding masks. ADAM [89] optimizer with a base learning rate of 0.001 is employed to train the model. The batch size was set to 16. The network was trained for 250 epochs, and a poly learning rate policy with the power of 0.9 was used to drop the learning rate. For training the network, Binary Cross Entropy loss was utilized and employed at the output of each decoder. For evaluation, three

metrics, namely, Pixel Accuracy, F1 metric, and Intersection over Union (IoU), were utilized. Pixel accuracy provides the percent of pixels that are accurately classified. The F1 score corresponds to an equally weighted average of the precision and recall scores. IoU, also known as the Jaccard Index, provides the ratio of the intersection of the pixel-wise classification results with the ground truth to their union. Among all these metrics, a higher percentage depicts how close the predicted class maps are to the ground truth. The formulation of these metrics is provided in Table 5.4.1.

| Table 5.4.1: Definitions of the metric utilized in this work |                                                                                                                                                |
|--------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------|
| Metric                                                       | Formulation                                                                                                                                    |
| Pixel accuracy                                               | $PA = \frac{1}{N} \sum_i^N (y_i == \hat{y}_i)$ <p>where <math>y_i</math> is a target values, and <math>\hat{y}_i</math> is the prediction.</p> |
| F1 Score                                                     | $F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$                                                                             |
| Intersection over Union                                      | $IoU = \frac{True\ Positive}{(True\ Positives + False\ Positives + False\ Negatives)}$                                                         |

### 5.4.3. Ablation Study

Extensive ablation studies on the various architectural components of KTNNet architecture were performed. Dilated residual network with 5 layers, as mentioned in Table 5.3.1 was utilized as the encoder for baseline results. In this ablation study, various parameters that affect the design of the model are presented.

Table 5.4.2: Analysis of various normalization techniques without the features from edge enhanced images. **Blue** text indicates the best and **Red** text indicates the second-best performance. GN → Group Normalization

| Models             | Pixel Accuracy | F1 Score     | IoU          | Combined Accuracy |
|--------------------|----------------|--------------|--------------|-------------------|
| CONV+BN            | 94.26          | 94.90        | 95.14        | 94.76             |
| <b>WSCONV + GN</b> | <b>94.54</b>   | <b>94.55</b> | <b>94.85</b> | <b>94.65</b>      |
| <b>WSCONV + BN</b> | <b>94.62</b>   | <b>95.01</b> | <b>95.24</b> | <b>94.96</b>      |

**Impact of varying the normalization techniques:** This section will evaluate the performance of KNet when the various normalization techniques were utilized. For a fair comparison, all the network components were fixed and only normalization layers were varied. Table 5.4.2 provides the quantitative results of this study. It can be seen that KNet with weight standardized convolution and batch normalization provided superior results compared to other normalization techniques. This indicates that weight standardization is required for accelerating micro-batch training.

**Impact of varying the depths of encoder topology:** When optimizing the network, it is critical to adjust the depth of the model to attain optimal results. Table 5.4.3 provides a set of quantitative results for varying depths of the network. In particular, only the encoder depths were varied to obtain these results. When the residual blocks were set to 1,1,2,2,2, KNet outperformed when compared to other models with lower and higher parameters. This demonstrates that the encoder depth with 1,1,2,2,2 was adequate to encapsulate all the features available in the CT image of the KITS dataset.

| Table 5.4.3: Analysis of various encoder depths with the features from edge enhanced images. <b>Blue</b> text indicates the best and <b>Red</b> text indicates the second-best performance |                 |                |              |              |                   |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------|----------------|--------------|--------------|-------------------|
| # residual blocks                                                                                                                                                                          | # of Parameters | Pixel Accuracy | F1 Score     | IoU          | Combined Accuracy |
| 1,1,1,1,1                                                                                                                                                                                  | 2.2 M           | 91.87          | 87.90        | 91.03        | 90.26             |
| 1,1,1,1,2                                                                                                                                                                                  | 2.5 M           | 95.08          | 94.5         | 94.79        | 94.79             |
| 1,1,1,2,2                                                                                                                                                                                  | 2.8 M           | 95.55          | 95.07        | 95.30        | 95.31             |
| <b>1,1,2,2,2</b>                                                                                                                                                                           | <b>2.9 M</b>    | <b>96.50</b>   | <b>95.77</b> | <b>95.94</b> | <b>96.07</b>      |
| <b>1,1,2,2,3</b>                                                                                                                                                                           | <b>3.2 M</b>    | <b>95.77</b>   | <b>95.19</b> | <b>95.41</b> | <b>95.46</b>      |

#### 5.4.4. Benchmark Results

To show the effectiveness of the proposed KNet, it is compared with other SOTA, such as UNet [128], UNet++ [254], CeNet [255], and nnUNet [256]. The base learning rate for nnUNet was set to 0.0001, as the network failed to train with a learning rate of 0.001. For all these comparisons, KNet with the configuration explained previously is utilized. The total number of parameters with this configuration is 2.87 M. The performance of this system was tested on the KITS CT database. The complete set of Pixel Accuracy, F1 score, and IoU for this dataset is provided in Table 5.4.4. These results demonstrate the performance of KNet when compared to the SOTA methods. KNet without edge enhanced feature provides a combined accuracy of 94.96%, while KNet with edge enhanced features achieves 96.07% on the KITS dataset, which is superior when compared to SOTA methods - UNet, UNet++, CeNet, and nnUNet, which produced 93.11%, 94.14%, 94.35%, and 92.24% respectively. This shows the effectiveness of the proposed network.

| Table 5.4.4: Segmentation results on the KITS datasets. <b>Blue</b> text indicates the best and <b>Red</b> text indicates the second-best performance |                        |                 |                |              |              |                   |
|-------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------|-----------------|----------------|--------------|--------------|-------------------|
| Models                                                                                                                                                | Edge Enhanced Features | # of Parameters | Pixel Accuracy | F1 Score     | IoU          | Combined Accuracy |
| UNet                                                                                                                                                  | X                      | 14.3 M          | 90.98          | 94.01        | 94.32        | 93.11             |
| UNet++                                                                                                                                                | X                      | 16 M            | 93.21          | 94.47        | 94.75        | 94.14             |
| CeNet                                                                                                                                                 | X                      | 29 M            | 94.49          | 90.86        | 94.46        | 94.35             |
| nnUNET                                                                                                                                                | X                      | 31 M            | 93.13          | 91.04        | 92.54        | 92.24             |
| KTNet                                                                                                                                                 | <b>X</b>               | <b>2.8 M</b>    | <b>94.62</b>   | <b>95.01</b> | <b>95.24</b> | <b>94.96</b>      |
| KTNet                                                                                                                                                 | ✓                      | <b>2.9 M</b>    | <b>96.50</b>   | <b>95.77</b> | <b>95.94</b> | <b>96.07</b>      |

Apart from the quantitative analysis, qualitative analysis has been illustrated in Fig. 5.4.2, Fig. 5.4.3, and Fig. 5.4.4. These examples show the complex nature of the input images and the performance of KTNet when compared to the SOTA methods. The proposed network generated sharp kidney and tumor semantic maps with a higher correlation to the ground truth despite the quality of the input images. As seen in the figures, the SOTA methods were only partially able to segment the kidney and the tumor from the CT image. In Fig. 5.4.3, UNet and UNet++ were unable to detect the tumor regions. In Fig. 5.4.4, a discrepancy in the ground truth image is noticed. The left kidney contain a tumor which is unmarked in the ground truth image. KTNet was still able to detect the tumor, while some of the other SOTA methods were unable to.

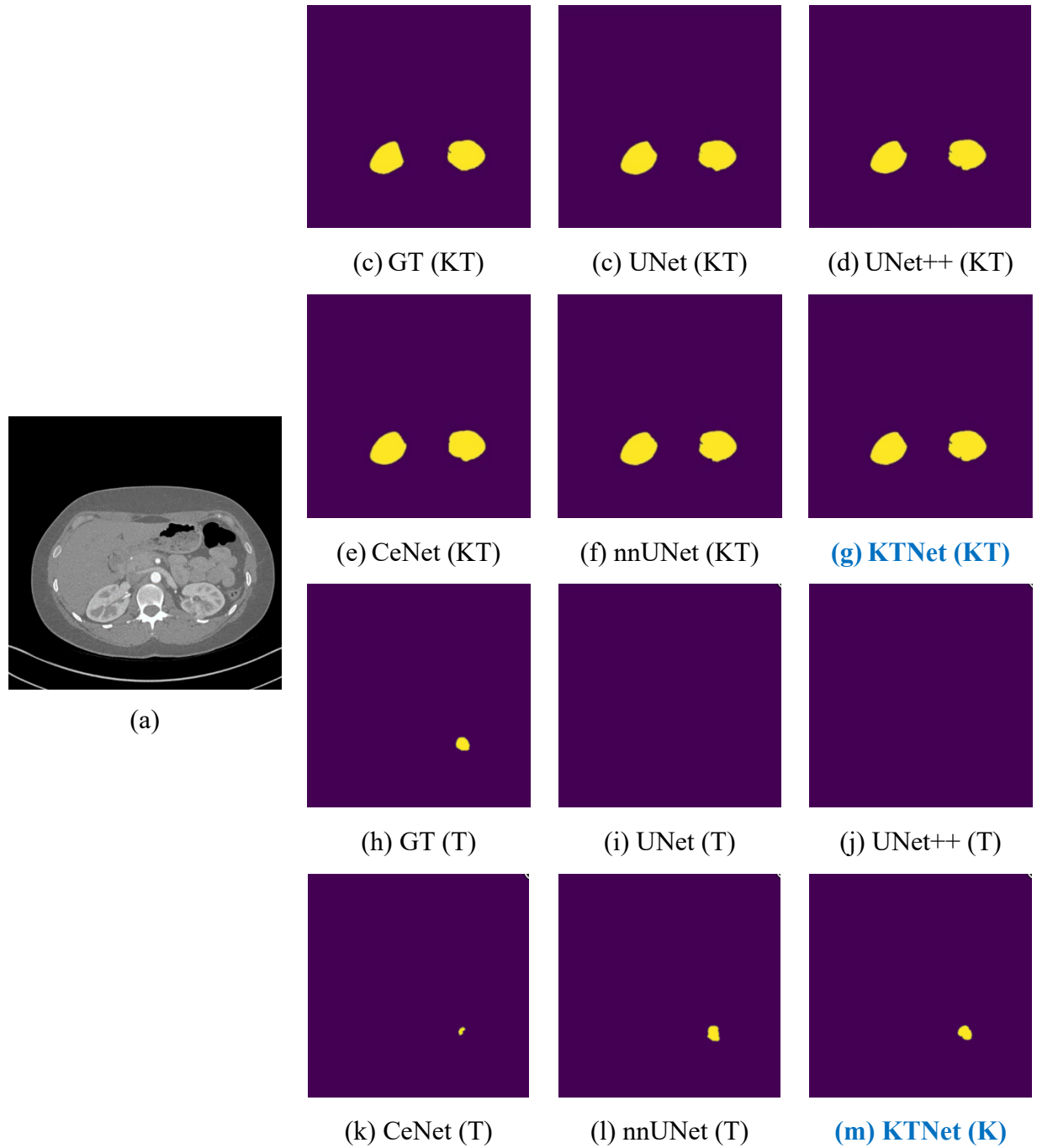


Fig. 5.4.2: A visual comparison of the performance of the SOTA methods in segmenting the kidney and tumor regions. (a) test image, (b) ground truth (GT) of the kidney+tumor (KT) region, (h) GT of the tumor (T) region. (c)-(f) are the results produced by the SOTA methods for the KT region and (i)-(l) are the results produced by the SOTA methods for the T region. As shown, KTNet (g) and (m) in **Blue** performs better in comparison.



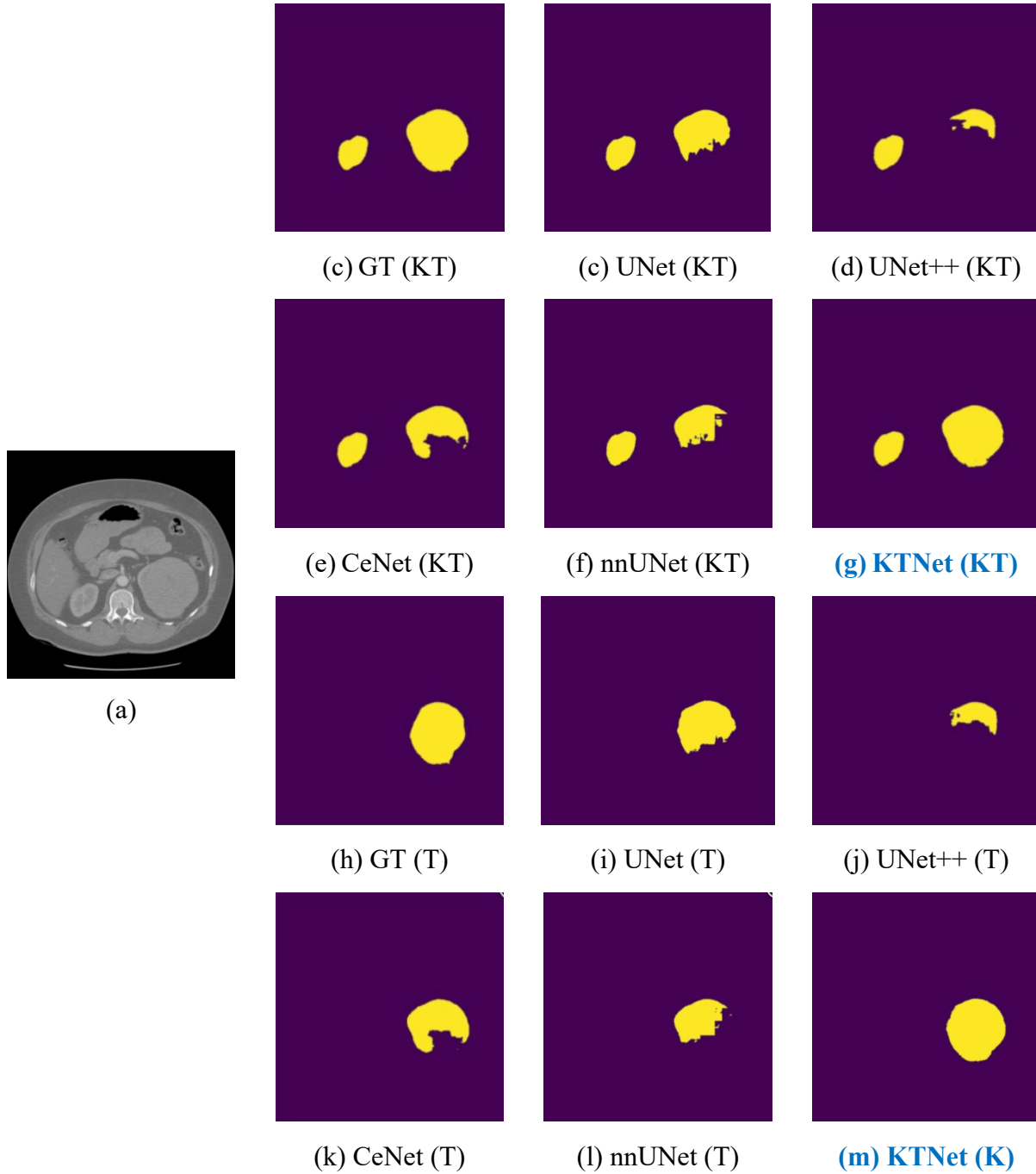


Fig. 5.4.3: A visual comparison of the performance of the SOTA methods in segmenting the kidney and tumor regions. (a) test image, (b) ground truth (GT) of the kidney+tumor (KT) region, (h) GT of the tumor (T) region. (c)-(f) are the results produced by the SOTA methods for the KT region and (i)-(l) are the results produced by the SOTA methods for the T region. As shown, KTNet (g) and (m) in **Blue** performs better in comparison.

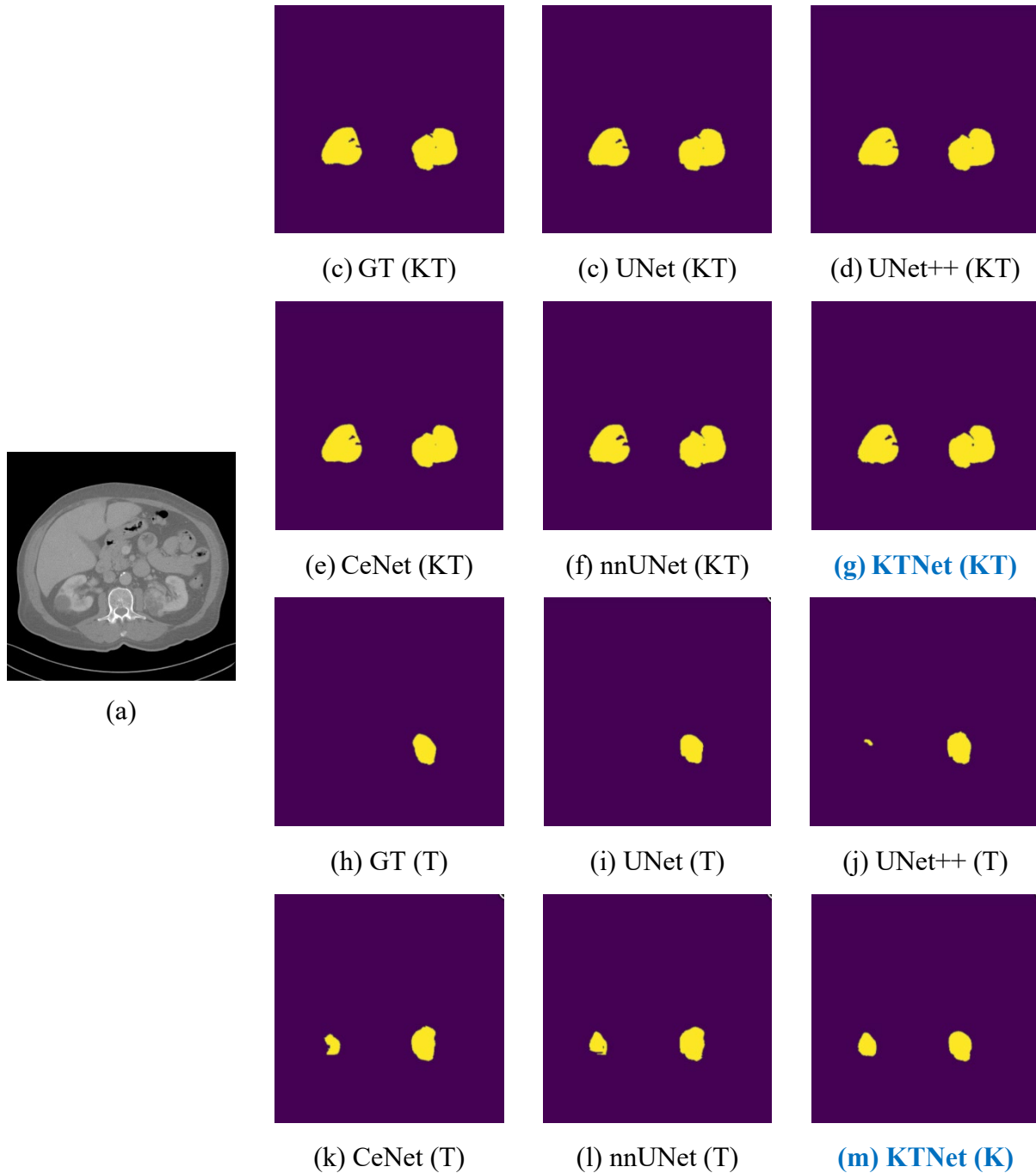


Fig. 5.4.4: A visual comparison of the performance of the SOTA methods in segmenting the kidney and tumor regions. (a) test image, (b) ground truth (GT) of the kidney+tumor (KT) region, (h) GT of the tumor (T) region. (c)-(f) are the results produced by the SOTA methods for the KT region and (i)-(l) are the results produced by the SOTA methods for the T region. As shown, KTNet (g) and (m) in **Blue** performs better in comparison. A discrepancy in the ground truth image is noticed. The left kidney contains a tumor which is unmarked in the ground truth image. KTNet was still able to accurately detect the tumor.

## 5.5. Summary

This chapter introduces a new 2D CNN architecture- KTNNet, that exploits the features obtained from edge enhanced images to segment kidneys and tumors from CT images. The network comprises of a dilated ResNet encoder architecture and a novel dual decoder branch designed to segment the kidney and tumor regions individually. In addition, features were extracted from edge-enhanced images to boost the system's performance. Extensive ablation studies and quantitative analysis conducted on the KITS CT database demonstrated the nature of design and effectiveness of the proposed KTNNet. It was able to outperform state-of-the-art techniques by achieving 96.07% combined accuracy with just **2.9 million** parameters. Quantitative analysis shows that KTNNet was able to achieve superior semantic maps for the kidney and tumor regions with crisp boundaries when compared to state-of-the-art techniques. As a part of future work, the possibilities of usage of KTNNet for other medical fields such as lung cancer segmentation and mammogram segmentation will be explored. As the parameter count is lower than the state-of-the-art systems, the possibilities of deploying KTNNet on edge devices will also be explored.

## **Chapter 6. Image Binarization**

### **6.1. Motivation**

Thresholding is a critical pre-processing step for image segmentation in the field of vision-based systems for object detection and classification. Segmentation of medical images such as Computed Axial Tomography (CAT), Magnetic Resonance Imaging (MRI), X-Ray, Phase Contrast Microscopy, and Histological images presents a number of challenges due to the high variability of human anatomy and modality variation. Recent advancements in CAD of histological images aid in disease detection and classification. Given that the majority of pathology diagnosis is dependent on the pathologist's expertise and ability, there is clearly a need for an automated assessment system. Histopathology is the microscopic examination of tissues to determine the appearance of damaged and degenerated cells. Histological images are stained with a specific color to aid in the differentiation of the tissue's components. Segmentation and analysis of such images are challenging, as they exhibit a high degree of variability in terms of color and clustering of cells. This work presented in this chapter describes an adaptive thresholding technique for segmenting cell structures from Haematoxylin and Eosin stained images. Additionally, pathologists can use these results to make an accurate diagnosis. The effectiveness of the Human Visual Thresholding (HVT) method is determined visually by comparison to state-of-the-art thresholding methods such as Otsu, Niblack, Sauvola, Bernsen, and Wolf. Computer simulations demonstrate the HVT method's effectiveness in segmenting critical data.

## 6.2. Introduction

Binarization or thresholding of images is a critical step in the field of image processing. It is an effective technique for differentiating the foreground (area of interest) from the background. Document analysis, medical image binarization, object detection in videos, biometric recognition systems, and gait recognition are just a few of the widely used applications [257-261]. Binarization is quite often the first step in image analysis and is used to locate a particular area or region of interest [262]. A binary image is a representation of a grayscale image using two gray values, namely 0 (black) and 255 (white). The benefit of acquiring a binary image is its simplicity in image processing. Binarization can be divided into two broad categories: global and local [263]. Global binarization is equivalent to a single threshold value being applied to the entire image. The thresholding value is chosen to denote the foreground and background of the image. The method is fast and produces reliable results in certain applications. Global thresholding is suitable only if the image's contrast distribution is uniform. Additional artifacts/noise may impair the performance of the thresholding technique [264]. Local thresholding techniques are used to overcome such issues. Local thresholding divides the image into overlapping blocks and assigns a threshold to each sub-block. The thresholds for each block are determined using the pixel information and its surroundings. Local thresholding generally produces superior results but is comparatively slower than global thresholding. Binarization is classified into six distinct groups, as explained in [259].

Medical imaging techniques are used to obtain information about organs or parts of the human body, physiological variations or abnormalities from the norm (for example, tumors and cysts), or structural information about bones, ligaments, and vessels [265]. Examination of medical images requires segmentation to aid pathologists in automatic detection [266]. Medical image

segmentation is important for anatomical research on normal body structure, quantifying tissues using metrics such as surface area and volume, localization of abnormalities, prognosis and treatment planning, 2D and 3D registration of imaging data, and computer-aided surgery [105-108]. For over a century, microscopy has been critical in diagnosing diseases [267, 268]. It entails the examination of a patient's surgically removed specimen [269]. Manual diagnosis of tissue cells requires considerable time and effort [268]. The accuracy of the diagnosis is entirely dependent on the pathologist's ability and experience. Early detection is critical for patient care, therefore, an automated segmentation technique is essential for perfect diagnosis and treatment planning [269-271].

The literature presents some of the techniques used in the segmentation of pathological images. Roullier [272] proposed a multiresolution algorithm that used a graph-based segmentation procedure to segment mitotic count in breast histology images. Basavanhally [273] proposed a method that used rank-based normalized cuts for detecting tubules in breast histology images. These cuts were defined by a mean shift clustering algorithm. Xu [274, 275] developed a method that used a combination of rank-based normalized cuts and active contours to segment prostate gland structures. Al-Lahham [276] proposed a method that segments out cancer cell nuclei using the K-means clustering method. Loukas [277] proposed a method that segments the cell nuclei by applying a Principal Component Analysis (PCA) technique, which performs color transformation of the image. The transformed image is then globally thresholded to segment the nuclei. Markiewicz [278] proposed a method that used color information and morphological features to train the support vector machine (SVM) classifier for cell segmentation. Kong [279] proposed a method that separated cellular regions of the image from the background. The algorithm used a controlled pixel-wise classification technique. The cells were isolated based on the radial

symmetry and shape. Karsnas [280] proposed a learning algorithm where the intensities and label dictionaries are constructed to separate the foreground from the background. The cells were isolated by merging regions with a marker-controlled watershed technique. Kim [281] developed a marker-controlled watershed algorithm based on the H-minima transform to segment the nuclei from histopathological images.

Some of the approaches mentioned above rely on color. These methods do not perform well when the image contains a larger cluster of cells touching one another [282]. Although graph-based segmentation techniques are extremely efficient, they are also very time-consuming, which limits their application to real-time cell segmentation [282]. Furthermore, learning-based methods require a sizable training dataset, and the probabilistic models must be carefully chosen. This chapter presents a thresholding technique that helps in segmenting the nuclei from Haematoxylin and Eosin (H&E) stained images. The HVT technique is compared to existing algorithms such as Otsu [283], Niblack [284], Sauvola [285], Bernsen [286], and Wolf [287]. Due to the lack of availability of ground truth images, the performance of the above-mentioned techniques is evaluated visually.

### **6.3. Existing Binarization Techniques**

Otsu's [283] is a global thresholding technique based on clustering. It is widely used as a result of its simplicity and effectiveness. The technique is predicated on the assumption that the image contains only foreground and background data. The algorithm conducts a comprehensive search for a threshold that minimizes the intra-class (foreground and background) variance. This intra-class variance is defined as

$$\sigma_x^2 = w_1(t) \sigma_1^2(t) + w_2(t) \sigma_2^2(t) \quad \text{Eq. 6.3.1}$$

where  $w_1$  and  $w_2$  are the class probabilities of the pixels set by threshold  $t$  and the class means  $\mu_1$  and  $\mu_2$  and the class variance  $\sigma_1^2$  and  $\sigma_2^2$  are defined as

$$\begin{aligned} \mu_1(t) &= \sum_{i=1}^t \frac{iP(i)}{w_1(t)} & \mu_2(t) &= \sum_{i=t+1}^I \frac{iP(i)}{w_2(t)} \\ \sigma_1^2(t) &= \sum_{i=1}^t [i - \mu_1(t)]^2 \frac{P(i)}{w_1(t)} & \sigma_2^2(t) &= \sum_{i=t+1}^I [i - \mu_2(t)]^2 \frac{P(i)}{w_2(t)} \end{aligned} \quad \text{Eq. 6.3.2}$$

where  $w_1(t) = \sum_{i=1}^t P(i)$  and  $w_2(t) = \sum_{i=t+1}^I P(i)$

Niblack [284] is a simple local thresholding method that works by sliding an overlapping window across the entire image to calculate a pixel-wise threshold. The local threshold is calculated for a given window size using the mean and standard deviation of the window's neighborhood. The Niblack threshold equation is as follows:

$$T_{local} = I_{mean} + c \sqrt{\frac{\sum I_i^2}{\text{Number of pixels}} - I_{mean}^2} \quad \text{Eq. 6.3.3}$$

where  $B$  is the binarized image,  $c$  is a constant,  $I_i$  is the pixel value, and  $I_{mean}$  is the average value of the pixels in the neighborhood.

Sauvola [285] method is a refinement of the Niblack method. The threshold is calculated using the dynamic range of the image gray values. By reducing the impact of noise on the standard deviation, it outperforms the Niblack method. The equation to estimate the threshold value is as follows:



$$T = m \left( 1 - k \left( 1 - \frac{\sigma}{R} \right) \right) \quad \text{Eq. 6.3.4}$$

where  $m$  is the mean,  $\sigma$  is the standard deviation,  $k$  is a positive value [0.1-0.5], and  $R$  is the dynamic range of the standard deviation.

Bernsen [286] proposed a local contrast-based thresholding method. There are two stages to the algorithm. The contrast threshold value is calculated as the difference between the maximum and minimum value for a given window in the first stage. The mean of the maximum and minimum gray-level values for the same window is estimated in the second stage. If the contrast value in the window falls below a certain threshold, the pixels are assigned to the background or foreground. The equation for the threshold is as follows:

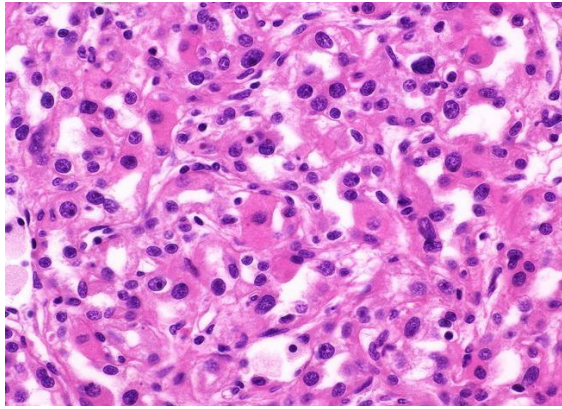
$$T(i, j) = \begin{cases} \frac{I_{max}(x, y) + I_{min}(x, y)}{2} & \text{if } C(x, y) \geq T_c \\ 0 & \text{otherwise} \end{cases} \quad \text{Eq. 6.3.5}$$

where  $I_{max}$  is the maximum gray-level value,  $I_{min}$  is the minimum gray-level value. The result is dependent on the size of the window.

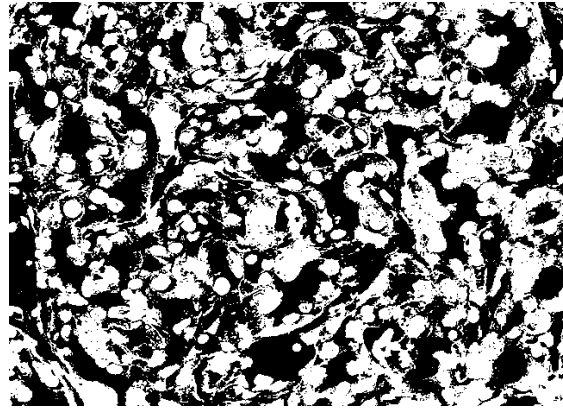
Wolf [287] proposed a method that calculates the standardized mean and contrast values of the image. The method overcomes the impediments of Sauvola's method. The threshold equation is as follows

$$T = \sqrt{(1 - k) * I_{mean} + \frac{\sigma * k}{R * (I_{mean} - M)} + k * M} \quad \text{Eq. 6.3.6}$$

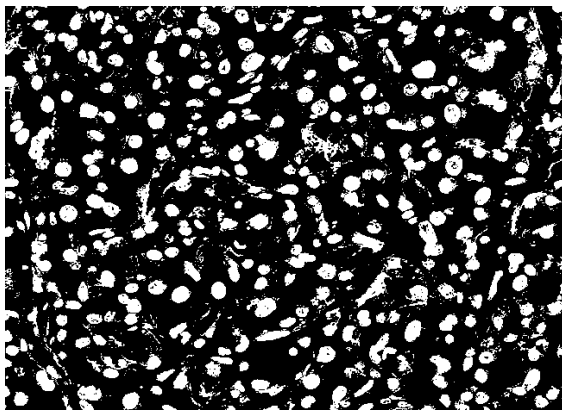
where  $k$  is the initial threshold value,  $\sigma$  is the standard deviation,  $M$  is the minimum intensity value, and  $R$  is its maximum deviation.



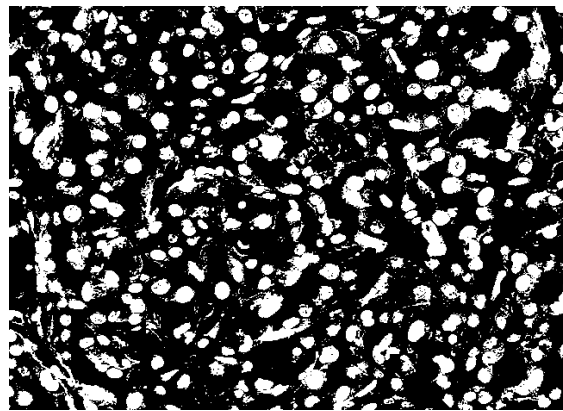
(a)



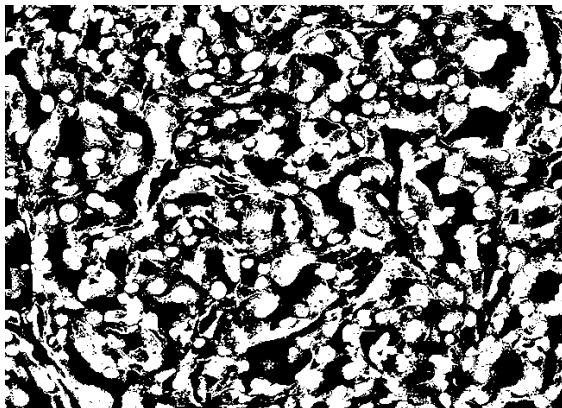
(b)



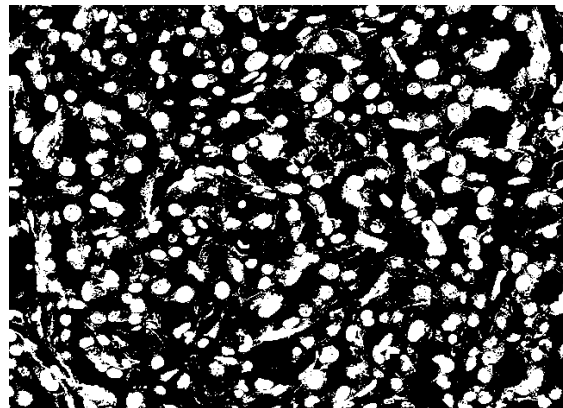
(c)



(d)



(e)



(f)

Fig. 6.3.1: Results of the existing thresholding techniques. (a) test image, (b) Otsu, (c) Niblack, (d) Sauvola, (e) Bernsen, and (f) Wolf

## 6.4. Human Visual Thresholding Method

This section describes the thresholding technique for segmenting nuclei from H&E stain images. The concept is to adjust the threshold across the image by computing the local mean and standard deviation of each pixel in the window. The methods outlined above do not work well for histological images due to the low contrast between foreground and background. Moreover, as the space between the cell is minimal, it becomes hard to separate them. Such impediments are overcome in the proposed method. The general flow diagram of the thresholding procedure is depicted in Fig. 6.4.1. The HVT technique makes use of the integral image concept as defined in [288].

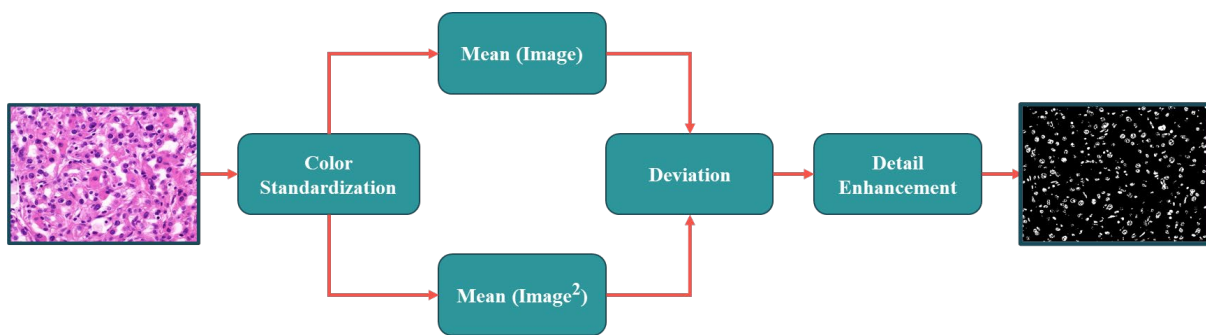


Fig. 6.4.1: The flow diagram of the proposed Human Visual Thresholding method

Color consistency remains a significant issue in whole-slide imaging, as even minor differences in the appearance of digitized slides can result in image misinterpretation by pathologists or computer-aided diagnosis systems. These differences are primarily due to variations in laboratory protocols and imaging device manufacturers. Fig. 6.4.2 illustrates the color variation in digitized whole-slide images. To overcome such variations in the data, an image standardization technique is employed. Consider a target image  $F(m,n)$  and a reference image

$K(m,n)$ . First, the images are transformed from the RGB color space to HSV color space ( $F_T(m,n)$  and  $K_T(m,n)$ ). Second, the histogram of  $K_T(m,n)$  is computed for the H, S, and V channels individually. Third, the Histogram Equalization [289] algorithm is employed to map the histogram values of  $K_T(m,n)$  to  $F_T(m,n)$ . Finally,  $F_T(m,n)$  is transformed from the HSV colorspace to RGB. The result of the image standardization technique is displayed in Fig. 6.4.3.

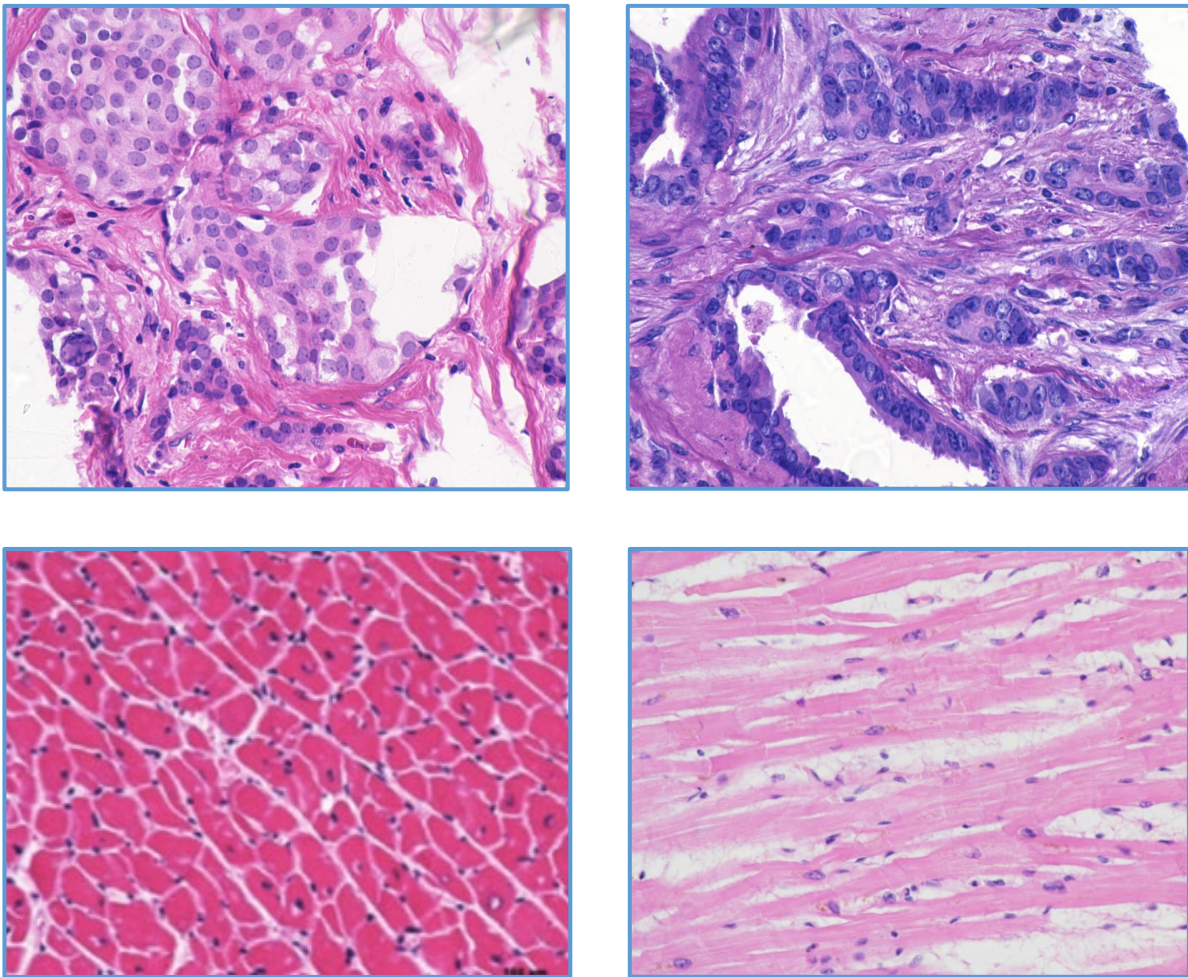


Fig. 6.4.2: H&E stained histopathology images demonstrating color variations caused by difference in slide preparation



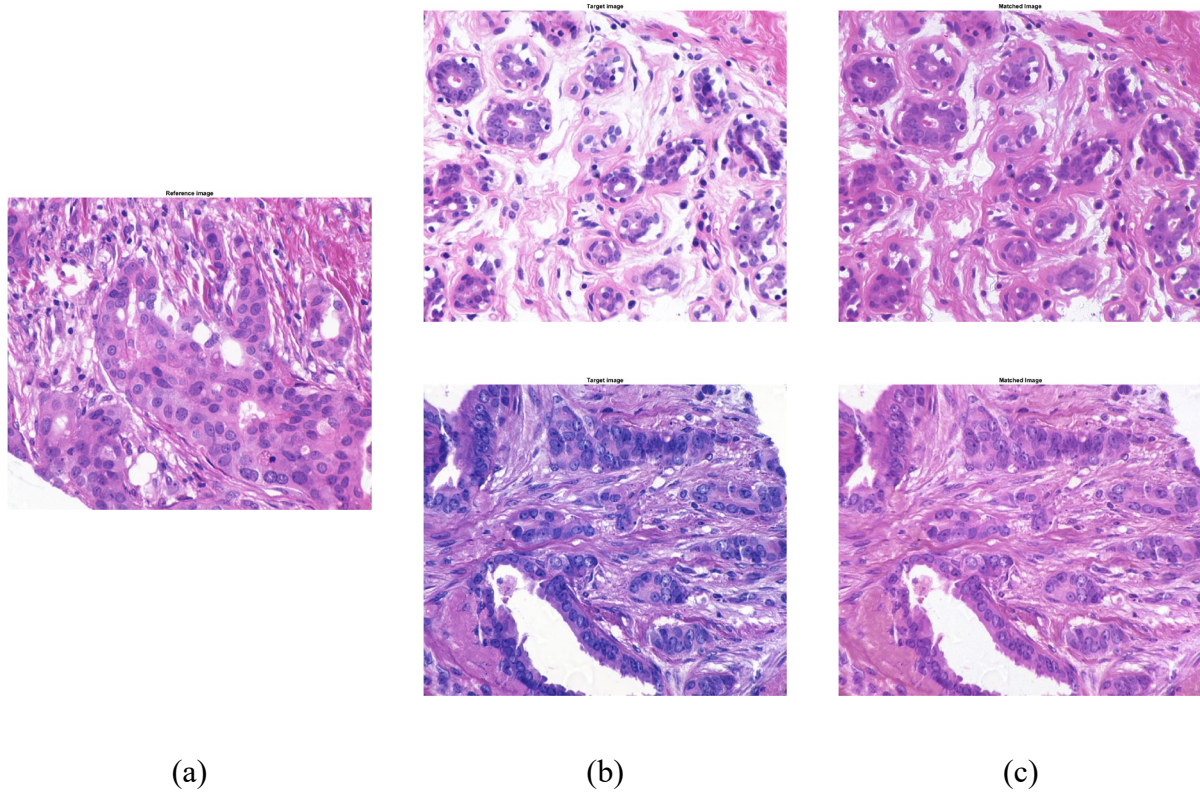


Fig. 6.4.3: Illustration of the image standardization process. (a) reference image, (b) target image, and (c) resulting standardized image

When an input image  $F(m,n)$  is considered, its integral image  $H$  at position  $(m,n)$  contains the sum of all the pixels on the top and left of that  $(m,n)$ . Thus, for a given window, the intensity of a pixel at  $(m,n)$  is defined as

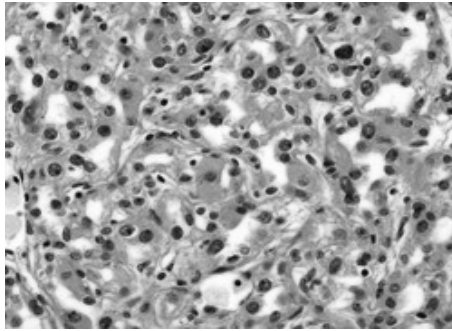
$$H(i,j) = \sum_{m=0}^i \sum_{n=0}^j F(m,n) \quad \text{Eq. 6.4.1}$$

An integral image can be generated from any grayscale image efficiently in a single iteration. Mean at a position  $(i,j)$  can be computed swiftly by:

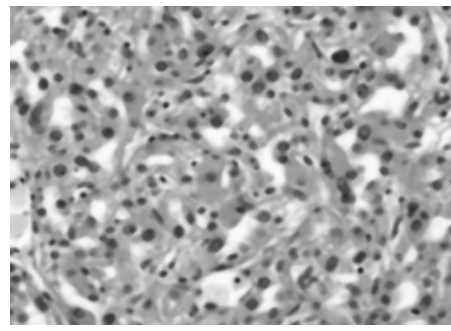
$$I_{mean}(i,j) = \frac{\left[ H\left(i + \frac{w}{2}, j + \frac{w}{2}\right) - H\left(i + \frac{w}{2}, j - \frac{w}{2}\right) + H\left(i - \frac{w}{2}, j - \frac{w}{2}\right) - H\left(i - \frac{w}{2}, j + \frac{w}{2}\right) \right]}{w^2} \quad \text{Eq. 6.4.2}$$

The output of  $I_{mean}$  is shown in Fig. 6.4.4. The standard deviation is defined in Eq. 6.4.3 and the resulting deviation image is displayed in Fig. 6.4.5.

$$\sigma(i,j) = \sqrt{\frac{1}{w^2} \sum_{m=i-\frac{w}{2}}^{i+\frac{w}{2}} \sum_{n=j-\frac{w}{2}}^{j+\frac{w}{2}} F^2(m,n) - I_{mean}^2(i,j)} \quad \text{Eq. 6.4.3}$$

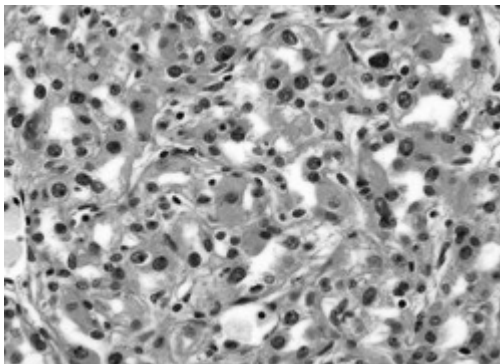


(a)

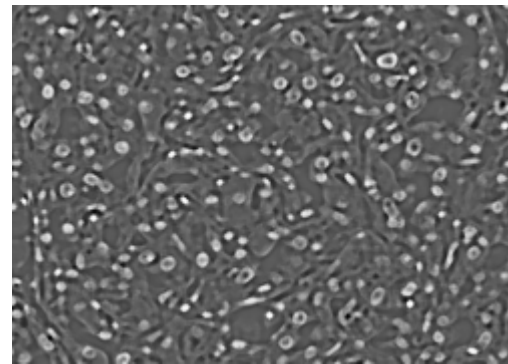


(b)

Fig. 6.4.4: An illustration of the  $I_{mean}$  image generation. (a) test image and (b)  $I_{mean}$  image.



(a)



(b)

Fig. 6.4.5: An illustration of the deviation image (a) test image and (b) deviation image.

$F^2(m, n)$  can be computed the same way as defined in Eq. 6.4.2. The output  $F^2(m, n)$  is shown in Fig. 6.4.6. The deviation image can also be combined with the original image, as shown in Eq. 6.4.6, depending on the application. As the outcome of the thresholding is dependent on the standard deviation, a guided filtering procedure is applied to the deviation image [187]. This is an edge-preserving filter and performs remarkably well due to its non-approximate means of implementation. The resulting output is shown in Fig. 6.4.7. Furthermore, it yields better quality, and the computation time is greatly reduced due to its filter size invariance. It can be defined as shown in Eq. 6.4.4.

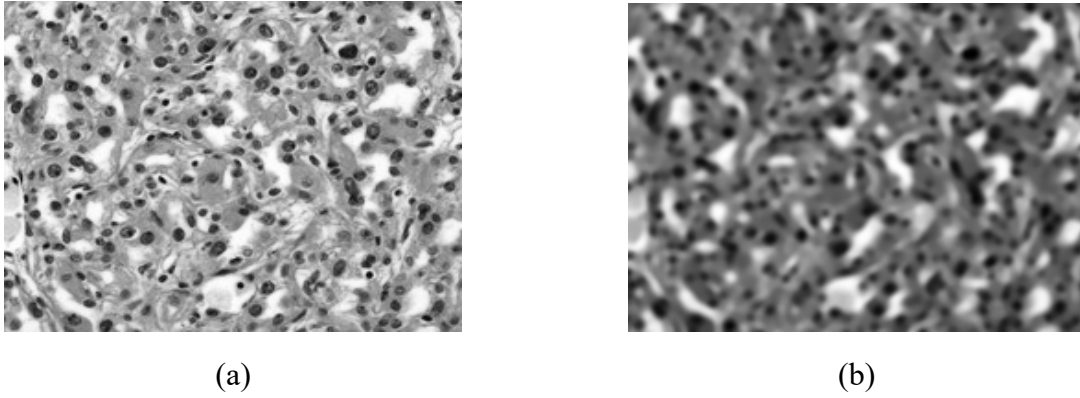


Fig. 6.4.6: An illustration of the  $F^2(m, n)$  image generation. (a) test image and (b)  $F^2(m, n)$  image.

$$F_x = \sigma_y I_x + \rho_y \in \phi_y \quad \text{Eq. 6.4.4}$$

$$\sigma_y = \frac{\frac{1}{|\phi|} \sum_{x \in \phi_y} I_x G_x - \mu_y \widehat{G}_y}{\varphi_y + \varepsilon}; \quad \rho_y = \widehat{G}_y - \sigma_y \mu_y \quad \text{Eq. 6.4.5}$$

where  $\phi_y$  is an odd-sized window,  $\varphi_y$  and  $\mu_y$  are variance and mean of  $I$ , respectively,  $|\phi|$  and  $\widehat{G}_y$  provides the count of the pixels and mean of  $G$ , respectively, in  $\phi_y$ . Finally, thresholding is performed and is given by equation Eq. 6.4.7. In this  $\gamma_1$  and  $\gamma_2$  depend on the intensity characteristics of the image into consideration.

$$\hat{\sigma} = \tau \times F + (1 - \tau) \times \sigma_y \quad \text{Eq. 6.4.6}$$

$$T_{opt} = \gamma_1 \times \left( I_{mean} \times (1 + C * (\hat{\sigma})) \right) + \gamma_2 \quad \text{Eq. 6.4.7}$$

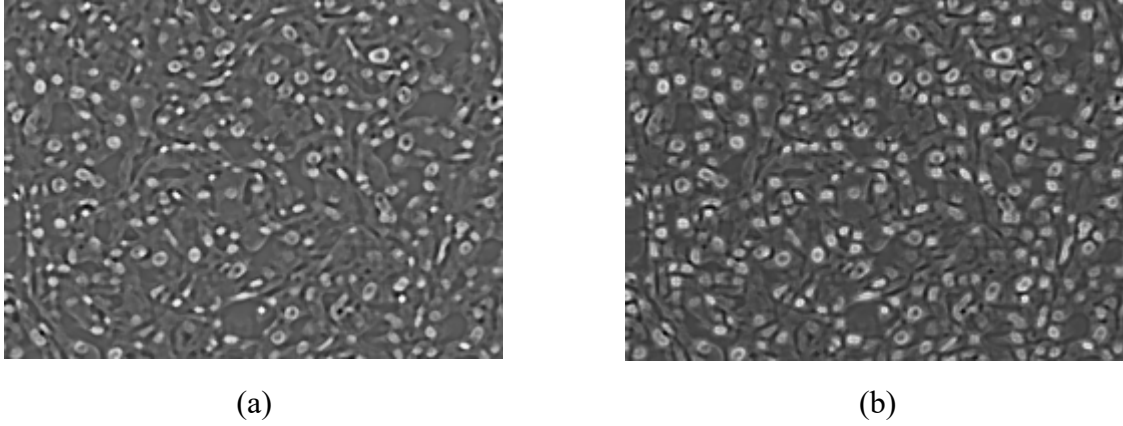


Fig. 6.4.7: The output of the detailed enhancement. (a) deviation image and (b) enhanced deviation image

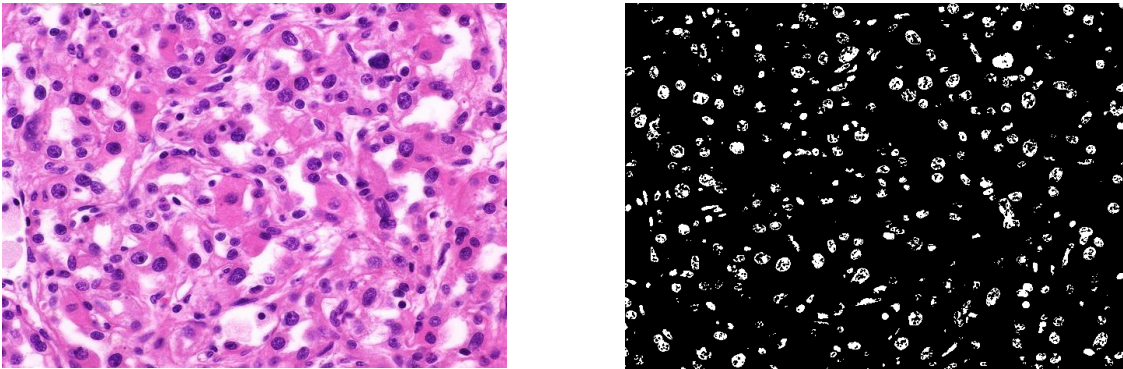


Fig. 6.4.8: Thresholded images acquired using HVT method (a) test image (b) Result

Kim [290] proposed a multi-window thresholding technique that aimed at reducing the effect of window size on documentary images. Similarly, in medical imaging, particularly, nuclei



size may vary. To make the algorithm less dependent on window sizes, three window sizes are employed to compute the thresholding. This is defined by

$$\hat{\sigma} = \kappa_1 \times T_{opt(w1)} + \kappa_2 \times T_{opt(w2)} + \kappa_3 \times T_{opt(w3)} \quad \text{Eq. 6.4.8}$$

$$\text{where } \kappa_1 + \kappa_2 + \kappa_3 = 1$$

The output acquired after applying the HVT technique is shown in Fig. 6.4.8.

## 6.5. Computer Simulation

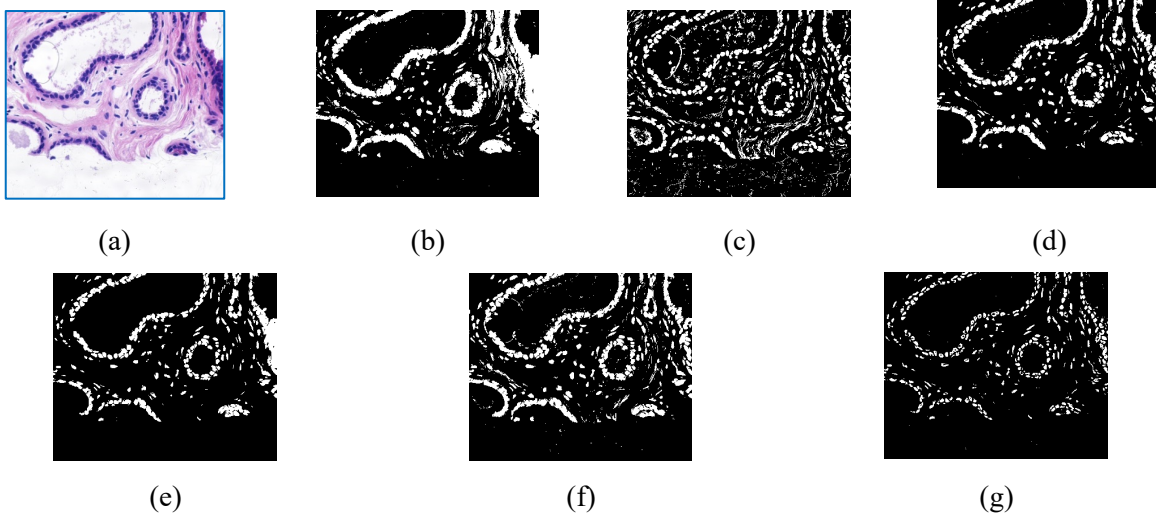


Fig. 6.5.1: Visual comparison of thresholding methods: (a) Original image, (b) Otsu's method, (c) Niblack's method, (d) Sauvola's method, (e) Bernsen's method, (f) Wolf's method, and (g) HVT method.

For computer simulations, the UCSB Bio-Segmentation Benchmark dataset is considered. This database contains 58 cellular-level breast cancer images. Each input image has a resolution of 896 x 768. The effectiveness of the HVT method is analyzed by visually comparing the results

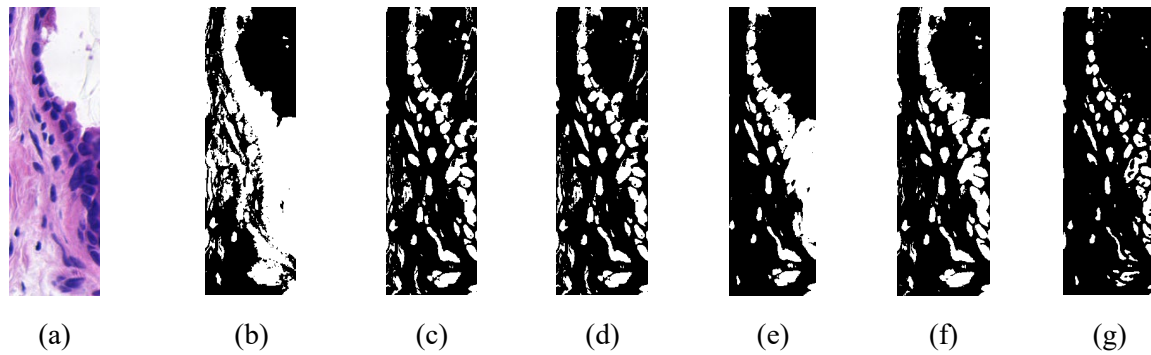


Fig. 6.5.2: Zoomed view of thresholding methods: (a) Original image, (b) Otsu's method, (c) Niblack's method, (d) Sauvola's method, (e) Bernsen's method, (f) Wolf's method, and (g) proposed method. The proposed system isolates each nucleus while perfectly suppressing the background information.

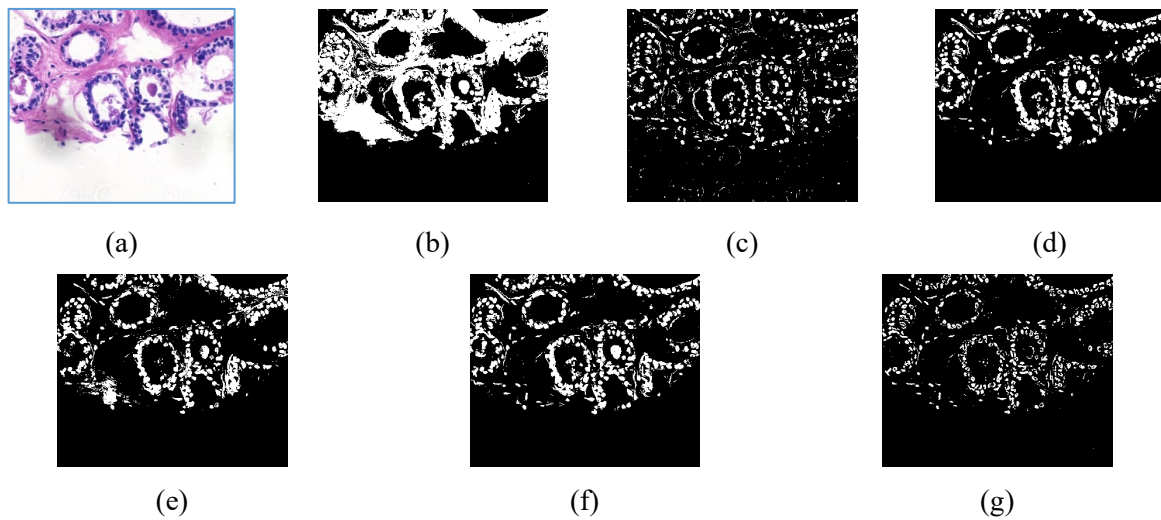


Fig. 6.5.3: Visual comparison of thresholding methods: (a) Original image, (b) Otsu's method, (c) Niblack's method, (d) Sauvola's method, (e) Bernsen's method, (f) Wolf's method, and (g) HVT method.

to the state-of-art thresholding methods such as Otsu, Niblack, Sauvola, Bernsen, and Wolf. The threshold values were set individually and the best results were chosen for each technique. The

main criteria adapted was to see if all the nuclei were detected. The post thresholding results are shown in Fig. 6.5.1, Fig. 6.5.3, and Fig. 6.5.4.

For a comparative evaluation, Fig. 6.5.1 is used. As seen in the figure, each state-of-the-art thresholding method segments the nuclei but is accompanied by a lot of background noise. Moreover, as the nuclei are close to each other, these techniques perform poorly in separating them. This is not the case in the proposed method. As seen in Fig. 6.5.1 (g), the image segments the nuclei and contains no background information. The nuclei are well defined and separated. This can be visualized clearly in Fig. 6.5.2, which displays the zoomed view of a particular section of the image.

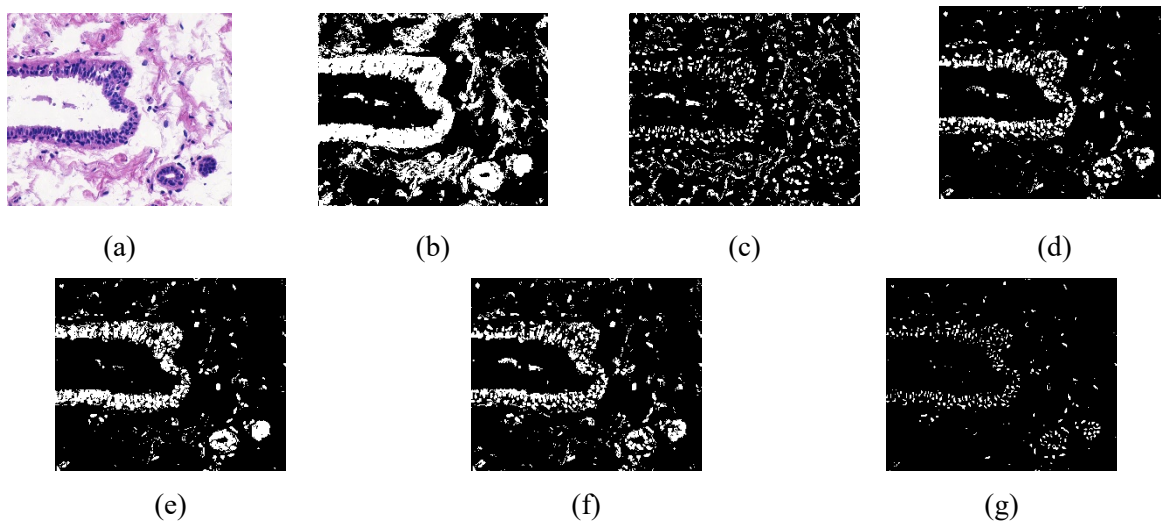


Fig. 6.5.4: Visual comparison of thresholding methods: (a) Original image, (b) Otsu's method, (c) Niblack's method, (d) Sauvola's method, (e) Bernsen's method, (f) Wolf's method, and (g) HVT method.

The method proves its robustness as it can be used to segment nuclei from colon, prostate, and neuroblastoma histology images, and is negligibly affected by the illumination variations of

the images. As the nuclei are separated, determining the size of each is easier, which assists pathologists in detecting and grading cancer types.

## 6.6. Summary

Thresholding is an important topic in vision-based systems that has received a lot of attention. However, the majority of cutting-edge techniques have been applied and tested for non-medical applications. The extraction of nuclei is a critical step in medical imaging, particularly in H&E stained histological images. Pathologists will be able to detect and diagnose diseases, such as cancer, more effectively if the nuclei are isolated from the stained images. Global thresholding techniques perform poorly in segmenting the nuclei, owing to the non-uniform illumination and poor contrast of images. Whereas, local thresholding methods handle these variations smoothly and perform better. In this work, a new local thresholding technique for nuclei segmentation from histological images is presented. In situations where the difference between the background and the foreground (nuclei) is small, the HVT method outperforms the existing techniques. Computer simulations demonstrate the HVT method's superiority over state-of-the-art methods. Further investigation will be conducted as part of future work to determine the optimal values for constants such as  $\gamma_1$ ,  $\gamma_2$  and  $C$ . In addition, the effectiveness of the thresholding method can be quantified by developing a non-reference measure due to the unavailability of ground truths.

## **Chapter 7. Initial Evaluation of Tumor Roughness to Predict Renal Cancer Subtypes**

### **7.1. Introduction**

Kidney cancer (renal cell carcinoma; RCC) is expected to account for 3.8% of all new cancers diagnosed and will be responsible for an estimated 14,400 deaths [291]. The widespread use of radiologic imaging likely accounts for the increased incidence of new cancer cases diagnosed in the United States since the 1970s. Up to 40% of these incidental masses are localized, and approximately 15-20% of these small renal masses are ultimately benign [292-297].

Small renal masses are typically not aggressive and are slow-growing with low malignant potential; therefore, urologists have adapted a monitoring program to follow these masses over time called active surveillance [298]. Active surveillance heavily relies on the growth of tumors in size and lacks uniform standardization, usually citing every 6-month imaging [299]. While size is an important factor in renal tumors, it does not solely predict the malignant potential of a small renal mass. In a study, Daugherty [300] proposed the need for different size cut-offs for the various renal cancer subtypes indicating that clear cell cancer has the most malignant potential and should be smaller than 4cm, and chromophobe has the lowest malignant potential and could wait until 7cm size. While active surveillance in appropriately selected patients is a viable alternative to surgical intervention, it still carries the burden of uncertainty, patient anxiety, risk of disease progression, and the associated financial cost to the patient [301]. As the diagnosis of small renal masses remains primarily radiographic, the development of software to assess the potential aggressiveness of a mass would be a useful aid in the clinicians' armamentarium for shared decision-making. Previous studies have attempted to discern tumor histology through the use of

current imaging techniques and automated image analysis with varying degrees of success [302-306].

In order to provide additional information on the malignant potential of small renal masses without needing a biopsy, CT scans were utilized with image enhancement techniques and edge detection to calculate the surface irregularities of renal masses scheduled for partial nephrectomy. Using this technology, an overall “Tumor Roughness Score” (TRS) is proposed to determine the distinction between renal tumor subtype and potential aggressiveness. The initial assessment of this technique is to determine its utility in active surveillance clinical follow-up and potential research trials.

## **7.2. Methods**

### **7.2.1. Patient Population**

Thirty patients who underwent nephrectomy for RCC at the University of Texas Health Science Center represent cases of clear cell RCC (n=10), papillary RCC (n=10), and chromophobe RCC (n=10). Limited demographic and follow-up data were available. Patients with known hereditary disorders and those deemed to have poor quality CT images were excluded.

### **7.2.2. Tumor Roughness Measure by Using Multi-Level Voxel Box-Counting**

The “roughness” measure of a kidney tumor requires contrast-enhanced CT scans, which can automatically detect, quantitatively measure, and distinguish lymph node involvement and local tumor invasion based on the tumor’s topography [307]. The extensive computer simulation validates the proposed roughness measure acquisition that requires a three-stage algorithm:

### **A. Improving the image quality and contrast**

In general, the quality of images is affected by factors such as restricted image dimension and pixel resolution, along with the compulsive effects caused due to noise and poor contrast. To overcome these problems, well-described enhancement methods such as guided filtering, edge enhancement, and image fusion to improve the contrast and quality of the image [308, 309] were utilized.

### **B. Multi-level image segmentation**

Multiple level segmentation aims to obtain more than one threshold for a given image and segment the image into specific regions of interest (tumor), which correspond to one background and several regions of interest [310]. The method assumes that the image contains only foreground and background information. Traditional multilevel segmentation techniques are usually susceptible to Gaussian and Impulse noise. The algorithm is a modification of the multilevel thresholding technique for image segmentation that eliminates noise and performs a comprehensive search for a threshold that minimizes the intra-class (foreground and background) alpha-trimmed variance [311]. The critical component of this method is the selection of a threshold such that it separates the pixels into multiple classes by maximizing the between-class variance. The between-class variance is defined as

$$\sigma^2 = \sum_{i=1}^M P_i (\mu_i - \mu_T)^2 \quad \text{Eq. 7.2.1}$$

where  $\sigma^2$  is the between-class variance,  $P_i$  is the probability of class  $i$ ,  $\mu_i$  is the mean value of class  $i$  and  $\mu_T$  is the total mean of the histogram.

### C. Hausdorff's measure of the regions of interest

The Hausdorff measure is a scale of the closeness of two groups of points that are subclasses of a metric space [312]. The measure assigns a scalar score to the similarity between two paths, data clouds or any group of points. It is defined by Eq. 7.2.2 and Eq. 7.2.3. Let  $(A, d)$  be a metric space. For any subset  $M \subset A$ ,  $diam(M)$  will denote the diameter of  $M$ . For any  $M \subset A$ , any  $\delta \in [0, \infty]$ , and any  $\alpha \in [0, \infty]$ , the outer measure is given by

$$H_\delta^\alpha(M) := \inf \left\{ \sum_{i=1}^{\infty} (diam(M_i))^\alpha : M \subset \bigcup_i E_i \text{ and } diam(M_i) < \delta \right\} \quad \text{Eq. 7.2.2}$$

Hence the Hausdorff  $\alpha$ -dimensional measure of  $M$  is

$$H^\alpha(M) = \lim_{\delta \rightarrow 0} H_\delta^\alpha(M) \quad \text{Eq. 7.2.3}$$

The system uses the box-counting method to generate feature points based on the presence of edges in the image. Based on these feature points, the Hausdorff dimension is estimated, and the slope of the best fit line generates a roughness score. In our case, the Hausdorff's measure is an outer measure that calculates the roughness of each level, and a combined score is generated. This score is referred to as the Tumor Roughness Score.

### 7.3. Statistical Analysis

Continuous variables were analyzed with either T-tests or ANOVA for comparisons of the five renal mass subtype groups. ANOVA post-hoc analysis used the least significant difference (LSD). All  $p$ -values are two-tailed with a significance of  $p < 0.05$ , unless otherwise specified.



Pearson correlation coefficient is used to determine correlation analysis specifically between renal mass size and tumor roughness because larger tumors are more likely to be aggressive. We utilized a linear regression model to identify the factors that are most influential in predicting tumor subtype.

## 7.4. Results

Of the thirty patients selected, three patients were excluded due to poor-quality CT images. The data from 27 patients selected for analysis comprised of the following subtypes: benign cyst (n=1), oncocytoma (n=2), clear cell RCC (n=15), papillary RCC (n=7), and chromophobe RCC (n=2). Overall demographics include median age of 56 (IQR 48 – 63), median body mass index 30.4 (27.6 – 32.0), median tumor size of 2.5 cm (IQR 1.8-4 cm), four positive margins (14.8%, 4/27) and consisted of 93% men (25/27). Only one tumor was upgraded to pT3a and a total of 3 pT1b, with all other tumors being pT1a. Race/Ethnicity includes 59% of European descent and 41% of non-European descent. The demographics of each subtype and comparison are displayed in Table 7.4.1. At a median follow-up of 3.9 years (interquartile range (IQR) 2.2-5.4 years), there were no patient deaths. None of the patients had a documented recurrence. The majority of the cases were Stage 1 disease; two cases did not have a recorded stage, and none of the subjects had documented metastasis or received neoadjuvant therapy before surgical resection (Table 7.4.1). An example of the adaptive segmentation technique for image processing for the calculation of the tumor roughness score is displayed in Fig. 7.4.1. Standard clinical CT scans were obtained and the kidney and tumor regions were segmented for surface enhancement separately. To investigate tumor irregularity, the tumor for each slice was extracted using edge detection and other technique

described in Methods to obtain a 3-dimensional surface irregularity. These slices are calculated into a composite score of tumor roughness.

| Table 7.4.1: Demographics by tumor subtype |             |                    |                    |                    |                     |                |
|--------------------------------------------|-------------|--------------------|--------------------|--------------------|---------------------|----------------|
| Renal Cancer Subtype                       | Benign Cyst | Oncocytoma         | Chromophobe        | Clear Cell         | Papillary           | <i>p</i> Value |
| Number of Patients (N)                     | 1           | 2                  | 2                  | 15                 | 7                   |                |
| Age                                        | 60          | 45 (42 - 45)       | 54 (51-54)         | 56 (42 - 66)       | 57 (51 - 63)        | 0.9            |
| Body Mass Index                            | 24.9        | 26.3 (24.1 - 26.3) | 31.8 (31.3 - 31.8) | 30.1 (27.3 - 31.6) | 32.01 (29.9 - 40.7) | 0.05           |
| Race/Ethnicity (white%)                    | 1/1 (100%)  | 1/2 (50%)          | 0/2 (0%)           | 9/15 (60%)         | 5/7 (71%)           | 0.4            |
| Biologic Gender (male %)                   | 0/1 (0%)    | 1/2 (50%)          | 2 /2(100%)         | 15/15 (100%)       | 7/7 (100%)          | 0.001          |
| Tumor Size (cm)                            | 1.8         | 5.3 (4.0-5.3)      | 4.45 (4.4-4.45)    | 2.3 (1.5 - 3.8)    | 2.5 (2.0 - 2.7)     | 0.01           |
| Tumor Stage (stage 3)                      | 0/1 (0%)    | 0/2 (0%)           | 0/2 (0%)           | 1/15 (7%)          | 0/7 (0%)            | 0.3            |
| Tumor Surface Roughness                    | 1.18        | 1.16 (1.12 - 1.16) | 1.27 (1.21 - 1.27) | 1.52 (1.38 - 1.55) | 1.56 (1.43 - 1.66)  | 0.004          |

The range of tumor size for each tumor class associated with the 27 patients is displayed in Fig. 7.5.1. The mean roughness score for each mass is 1.18, 1.16, 1.27, 1.52, and 1.56 units, respectively (Fig. 7.5.2,  $p<0.004$ ). The mean TRS Score did trend along the renal cancer subtype associated with tumor aggressiveness. The lowest quartile is 1.32 and would indicate the lowest risk based on this dataset and is visually displayed in the waterfall plot of TRS and colored by the

corresponding tumor subtype. The stage T3a upgraded tumor has a TRS of 1.46. The mean scores of positive and negative margins were identical ( $p = 0.9$ ), but both were above the lowest quartile.

| Table 7.4.2: Multivariable linear regression model of the outcome of kidney cancer subtype. |                        |         |         |
|---------------------------------------------------------------------------------------------|------------------------|---------|---------|
| Demographic                                                                                 | Coefficient (95% CI)   | t score | p Value |
| Constant                                                                                    | 2.59 (-5.8 - 0.58)     | -1.709  | 0.1     |
| Age                                                                                         | 0.004 (0.17 - 0.025)   | 0.9     | 0.69    |
| Body Mass Index                                                                             | 0.049 (0.008 - 0.107)  | 2.7     | 0.09    |
| White (yes vs. no)                                                                          | 0.114 (0.364 - 0.592)  | 0.5     | 0.62    |
| Tumor Size                                                                                  | 1.71 ( -0.351 - 0.009) | -1.2    | 0.06    |
| Roughness Score                                                                             | 2.601 (0.977 - 4.225)  | 4.2     | 0.003   |

These tumors were investigated with Furman score (n=23) and did not show significant differences between high-grade tumors vs. low-grade tumors ( $p=0.7$ ); however, 83% (19/23) were documented as Furman Grade 2 (low grade) and therefore there was not an acceptable range of tumor cellular aggressiveness based on histopathology. Renal mass size correlates with tumor roughness (Pearson,  $p=0.02$ , Fig. 7.5.3). However, tumor size itself was larger in benign tumors rather than malignant tumors ( $p=0.01$ ) giving the indication that size would not be a significant factor for aggressiveness. Linear regression analysis noted that the roughness score is the most influential on the model, with all other demographics being equal, including tumor size ( $p=0.003$ ) (Table 7.4.2).

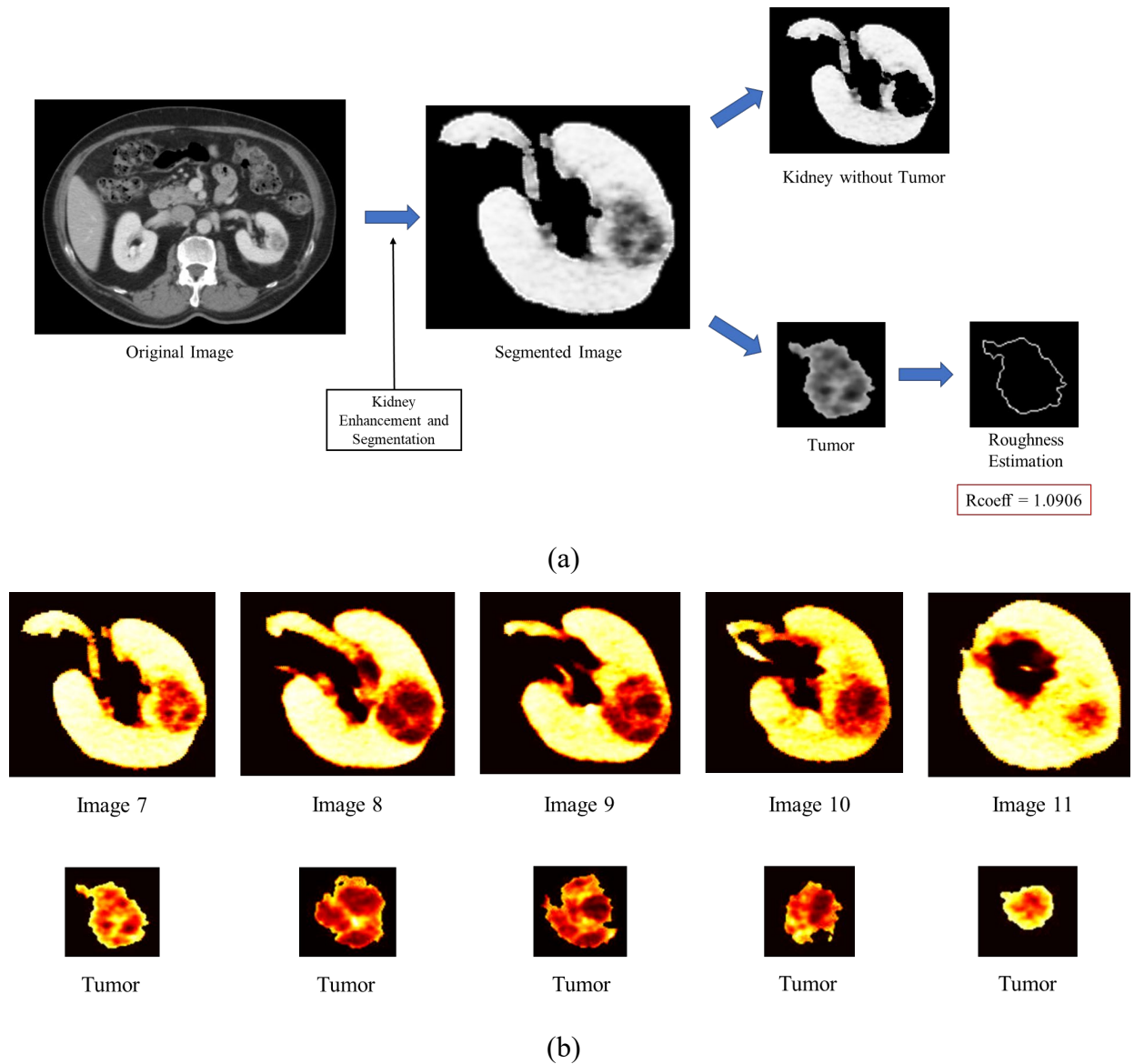


Fig. 7.4.1: Tumor roughness score image analysis. (a) Adaptive segmentation technique developed to segment the kidney and tumor for surface enhancement and investigate tumor irregularity. (b) The summary irregularity score ( $R_{avg} = 1.0413$ ) is averaged across all segments of the tumor.

## 7.5. Discussion

Herein, a data image processing technique is described to provide a tumor roughness score that can have additional information beyond tumor size. Currently, physicians use tumor size and growth rate in active monitoring strategies. However, this strategy has not been shown to be accurate regarding tumor subtype or tumor aggressiveness.

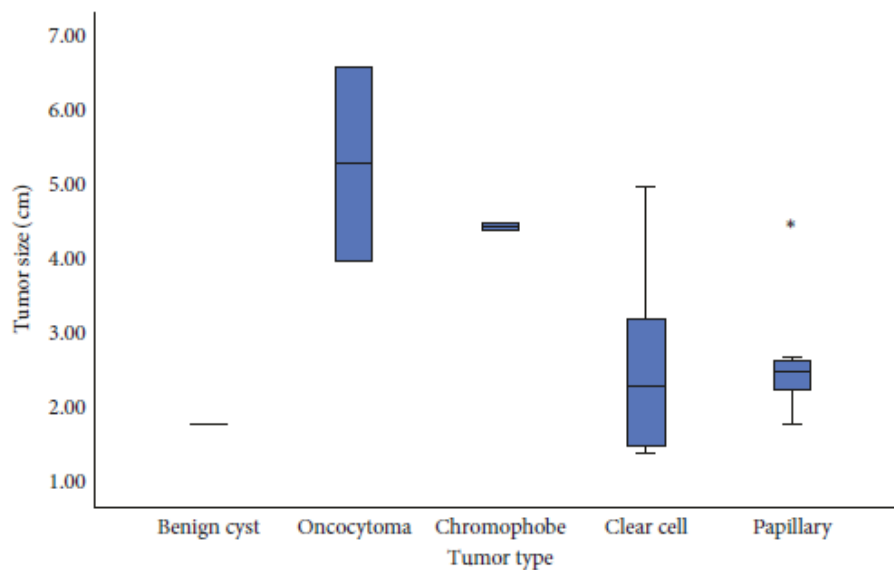


Fig. 7.5.1: Box and whisker plot of renal mass size compared to the renal tumor subtype for trend.

It is demonstrated that TRS can be a useful adjunct to image assessment and has the potential for automation using artificial intelligence and machine learning. TRS provides additional information regarding tumor subtype which correlates with the traditional RCC aggressiveness without a renal mass biopsy. This information is already available to physicians;



however, computer-assisted techniques remain limited for evaluation. The roughness scoring could provide information regarding those tumors that may be less aggressive than the clear cell and papillary cancers. These techniques could theoretically be employed in future clinical trials and to potentially implement machine learning protocols, which would enhance its predictive ability with more comprehensive datasets.

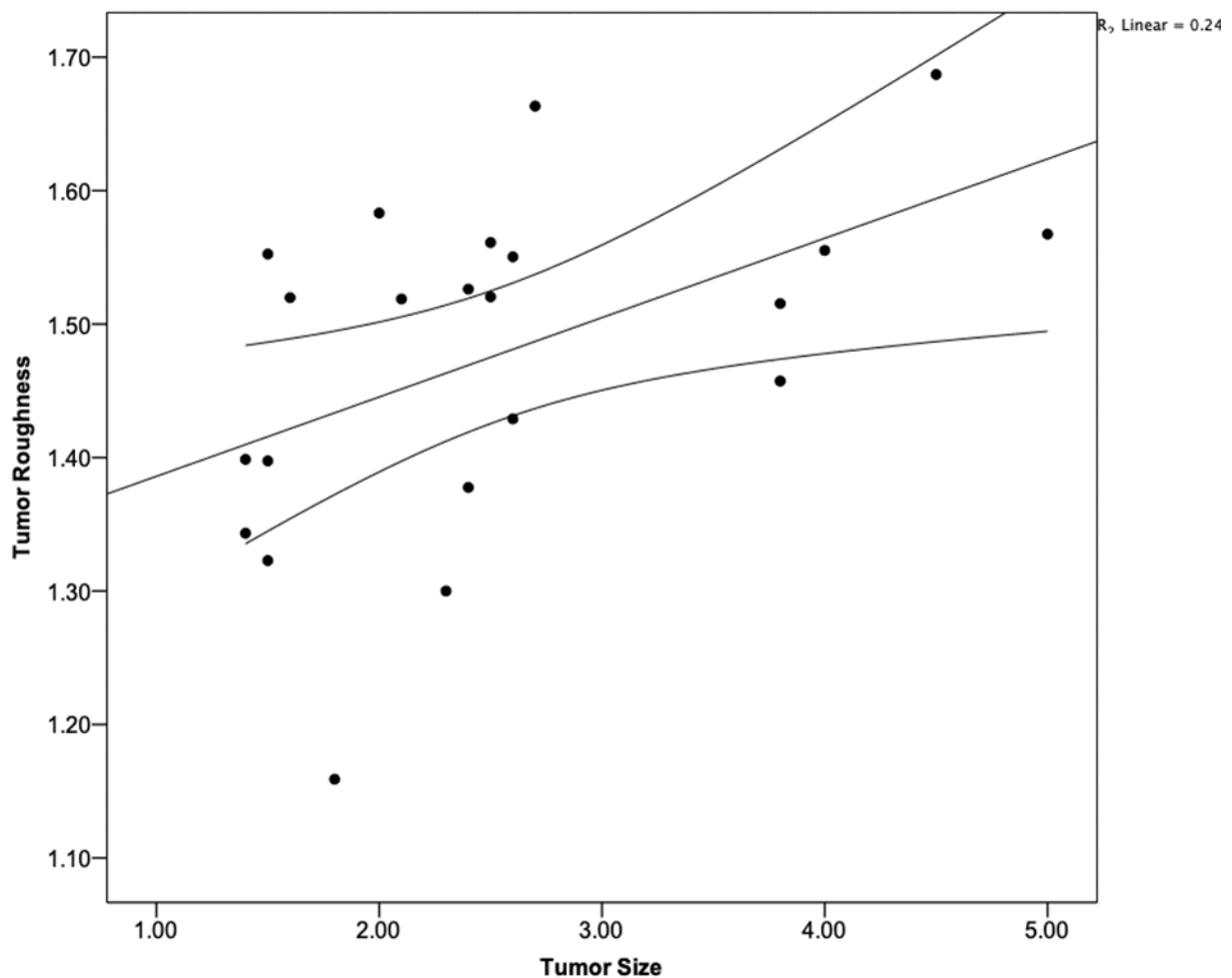


Fig. 7.5.3: Scatter plot noting the correlation between renal mass size and tumor roughness score (Pearson  $p=0.02$ )

The analysis presented in this work is distinct from previous research describing the surface contact area, which has also been shown to predict postoperative outcomes but was not investigated regarding tumor subtype [313, 314]. Tumor surface contact areas only represent how much the tumor is touching the kidney parenchyma; essentially, the test is an indicator of how exophytic a renal tumor is to its surrounding tissue.

The proposed technique enhances and outlines the entire tumor surface for irregularity not based on exophytic or endophytic, which is usually reserved for surgical resection planning. Another utilization of CT imaging is the texture analysis performed by Khene and colleagues, noting that the texture on CT could predict malignant or benign tumors as well as adherent perinephric fat during surgical intervention [315, 316]. Closure to this analysis is the work performed by Linguraru et al., who developed semi-automated quantification and classification of renal tumors to classify benign or malignant tumors, especially those patients with genetic abnormalities prone to forming multiple kidney tumors [317]. Considerable effort has been placed to detect papillary renal cancer from clear cell, including subtypes based on contrast enhancement, largely because papillary renal neoplasms have a poor or delayed enhancement that can be difficult to discern from hyperdense renal cysts [318, 319]. In addition, papillary subtype tumors are less likely to have a thick pseudo-capsule around the tumor which would coincide with having more irregular edges to the tumor [320]. The proposed technique could assist with this particular question in more extensive studies.

Data science with the uses of deep learning based automated methods is ever increasing and merging with clinical care [321]. Other authors have utilized similar techniques for image analysis. Pattern recognition is essential in radiology; however, the utilization of quantitative image analysis will provide additional objective data on tumor aggressiveness. These techniques



have been applied to lung cancer and others [322-324]. A review of other techniques for small renal mass characterization will be utilized in further evaluation and improvements [325, 326]. The goal is to add to the expanding literature using image analysis techniques to utilize data science to implement actionable data for physicians and patients.

There are some limitations to this study. First, while potentially significant associations were identified, this is a pilot study in a small cohort of patients without validated findings. Secondly, the cohort included CT scans performed at different institutions using different CT scanners, which could insert variability into the quality of the roughness score. However, in practical use, this would be the norm as other clinicians would utilize CT images from their home institutions. The assumption made here is that the tumor roughness score will have more applicability in small renal masses to have more influence on future decision-making. Unfortunately, a comparison of our method with methods of previous authors could not be made because specific algorithms have not been published [327]. We acknowledge the fast-moving field of image analysis and new techniques to separate target lesions in imaging with new techniques such as topographical assessments [328]. All renal mass types were included in this study because until surgical removal the subtype is completely unknown, providing impetus for this study. We are currently developing validation profiles of our work and the development of AI technology that can provide visual tumor characteristics in quantitative form as a biomarker.

## **7.6. Summary**

Using existing CT imaging in this proof of concept study, renal mass tumoral topography can be quantified and correlated with histology and biological aggressiveness in small renal

masses. Preliminary data is provided to suggest the use of imaging topography and genetic markers could provide additional information in the selection of patients for active surveillance or monitoring of renal masses compared to more aggressive therapy. A larger study is warranted to validate these findings and to determine if radiographic tumor surface analysis can obviate the need for conventional serum or urine biomarkers.

## **Chapter 8. Tufts Dental Database**

### **8.1. Motivation**

Imaging diagnosis involves the visual perception of an image and application of pattern recognition to differentiate normal from abnormal findings, followed by detailed characterizations of the abnormal findings that would allow for interpretation. The interpretation thus derived enables appropriate patient care and management.

The application of AI in dental healthcare has a very promising role due to the abundance of imagery and non-imagery-based clinical data. Expert analysis of dental radiographs can provide crucial information for clinical diagnosis and treatment. In recent years, DL-based systems have achieved the highest accuracy in various benchmarks, including analyzing dental X-ray images to improve clinical care quality. But, such systems are often opaque, making it challenging to comprehend their actions. One of the major reasons why successful implementation and integration of AI into routine clinical practice is rare is that most existing AI systems lack sufficient explainability and transparency. Explainable AI and Multimodal AI are emerging research topics aimed at understanding the decision-making process of an AI system. These research areas examine the metrics, data, and models used in decision-making and propose solutions to explain them explicitly. In order to develop such systems pertaining to dental healthcare, there is a dire need for data that encapsulates the cognitive process of the expert.

The Tufts Dental Database, a new X-ray panoramic radiography image dataset, has been presented in this thesis. This first-of-its-kind multimodal dataset includes the radiologist's expertise captured in the form of eye-tracking and think-aloud protocol. This dataset consists of 1000

panoramic dental radiography images with expert labeling of abnormalities and teeth. The classification of radiography images was performed based on five different levels: anatomical location, peripheral characteristics, radiodensity, effects on the surrounding structure, and the abnormality category. The contributions of this work are 1) a publicly available dataset that can help researchers to incorporate human expertise into AI and achieve more robust and accurate abnormality detection; 2) a benchmark performance analysis for various state-of-the-art systems for dental radiograph image enhancement and image segmentation using deep learning; 3) an in-depth review of various panoramic dental image datasets, along with segmentation and detection systems. The release of this dataset aims to propel the development of AI-powered automated abnormality detection and classification in dental panoramic radiographs, enhance tooth segmentation algorithms, and the ability to distill the radiologist's expertise into AI.

## 8.2. Introduction

The need for precise imaging and diagnostic tools has led to an exponential increase in computed tomography (CT) and X-ray imaging-based technologies to detect, diagnose, and treat various dental diseases in the last few decades. However, manual X-ray film interpretation and analysis may take large amounts of crucial clinical hours and may lead to misdiagnosis or underdiagnosis, owing to personal factors such as stress, fatigue, and experience levels [329]. These shortcomings can be reduced if intelligent dental X-ray film interpretation tools are developed to assist dentists in improving dental care. From this perspective, dental informatics, including automatic teeth identification, dental anomaly identification, and annotation, become an essential aspect of intelligent health care.

Utilizing classical Machine Learning (ML) techniques such as mathematical morphology [330], active contour [331], and level-set [332] have been used for teeth segmentation. Additionally, Bayesian techniques [333], linear models [334], and support vector machines [335] with hand-crafted features extracted utilizing Fourier descriptors [333], contours [336], and textures [332] have been used to perform classification. However, most of these methods require careful engineering of the feature extractor to transform the raw data into an appropriate representation for the algorithms to detect or classify the input images. The efficiency of these methods relies heavily on the image preprocessing performed and the quality of the features extracted. While some of these methods achieve satisfactory results, the test data is generally small, and they tend to perform poorly when tested on large datasets.

The limitations of the traditional approaches were overcome efficiently with the introduction of Artificial Intelligence (AI). AI-based systems' success can be mainly attributed to

the progress made in computer engineering, the massive availability of imagery data, and the development of deep learning algorithms [337]. The deep learning methods involve representation-based learning techniques that consider raw image data as input, process this data in different layers to automatically identify the optimal representations, and produce a meaningful output in the form of detection or classification of the input data. In recent years, deep learning-based methods have been employed in various medical tasks, producing promising results [338]. The main advantage of the deep learning system over the classical ML systems is its ability to learn features directly from the raw data using a general-purpose learning procedure instead of hand-crafted features.

In dentistry, AI has been applied recently to detect carious lesions on bitewing radiographs using pre-trained convolutional neural networks (CNNs) [339]. It has also been used to identify periodontal bone loss (PBL) in peri-apical radiographs and the detection of apical lesions and treatment planning in panoramic dental radiographs [340-344]. Furthermore, given the complexity and the associated diagnostic efforts, the application of CNNs on panoramic radiographs may result in a higher diagnostic yield in terms of increased accuracy, reliability, and time savings for the assessment [340].

In dentistry, imagery plays a vital role in screening and treatment. Multiple types of imagery for the same anatomical region of the same individual are taken over multiple time points, and the corresponding non-imagery-based clinical data is also present. Further, dental conditions such as caries, apical lesions, and PBL are relatively prevalent, making it easy to build the dataset to train and optimize neural networks. Despite these potentials, AI solutions have not yet become the norm in routine medical practice. Systems based on CNNs have only been recently adopted in

dental radiograph research, and the applications based on these technologies are only now entering the clinical arena [345].

As suggested in [346], some of the main reasons for AI technologies not being fully adopted in dentistry are: 1) dental data is not readily accessible due to data protection and privacy concerns, 2) datasets lack structure and are complex, multi-dimensional, and often biased with overly sick, overly healthy, or overly affluent data points, 3) datasets are relatively small compared to other image-based datasets for AI, 4) lack of "hard" gold standards and the requirement of an expert to label, and 5) lack of trust in AI, as it does not provide any feedback on how/why it arrived at the prediction.

To alleviate some of the problems mentioned above, a new X-ray dental image dataset called the "Tufts Dental Database (TDD)" is introduced in this work. TDD consists of 1000 panoramic radiographs with the abnormalities, teeth, and the maxillomandibular region of interest outlined. Among the various types of intra-oral and extra-oral radiographs, panoramic X-ray images were selected for this study due to their advantage over traditional x-ray imaging methodologies. They display (i) the entire mouth region, (ii) all the teeth in both the upper and lower jaws in a single X-ray, and (iii) the positions of fully emerged as well as emerging teeth. Furthermore, faster operation time, more diagnostic capabilities, less radiation exposure, better patient acceptance, and fewer infection control procedures [347] make it ideal for dentistry. Intraoral radiographs such as periapical and bitewing X-rays can image localized regions in the mouth but require multiple scans to capture different areas. This makes it time-consuming, causing discomfort to the patient, and can lead to an increase in radiation dose [348]. Fig. 8.2.1 shows an example for each of these three different kinds of radiographs.

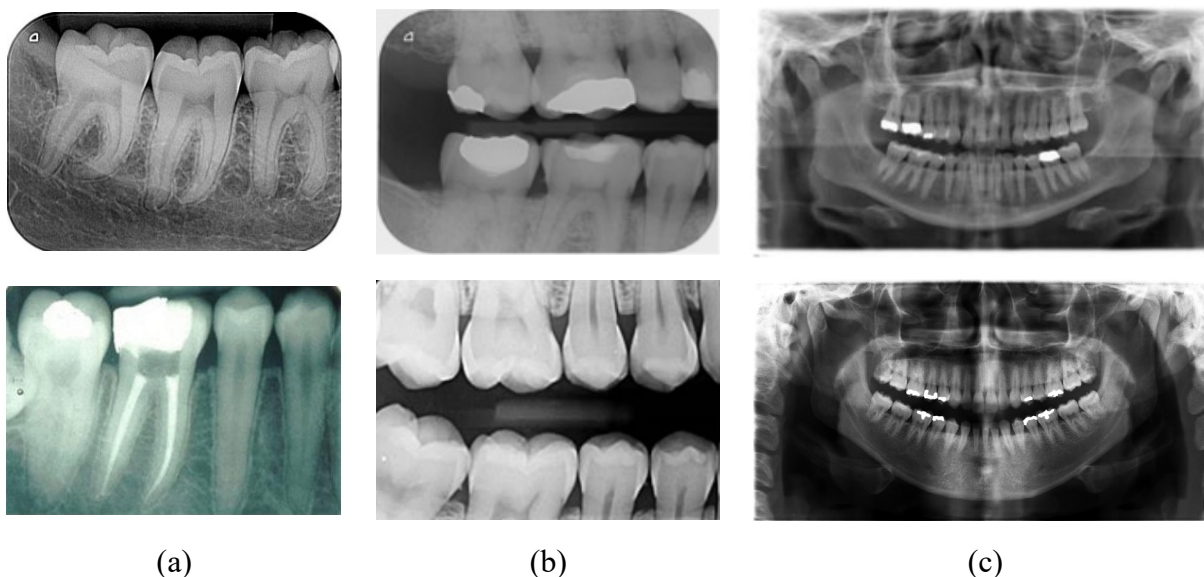


Fig. 8.2.1: Three different kinds of dental radiographs. (a) periapical x-ray, (b) bitewing x-ray, (c) panoramic x-ray

Additionally, the perception (the ability to recognize abnormal patterns) and cognition (the ability to interpret) or the "expertise" of the radiologist is captured through eye-tracking and verbal dictation or the think-aloud protocol. To distill the knowledge of the expert into an AI system, it is essential that the eye-tracking data is supplemented with the think-aloud protocol. For example, a more prolonged fixation may point to an interesting area or an area challenging to interpret. Hence, supplementing the eye-tracking data with verbal dictation allows for better interpretations. The eye-tracking details are presented in TDD as gaze maps, and the oral dictation is provided as text within the .json file.

Incorporating these aspects into TDD makes it a multimodal dataset with modalities such as X-ray images, gaze maps, and audio-text files. The multimodal nature of the dataset allows for a much more comprehensive range of applications, such as deeper investigation of visual



characteristics of medical experts and optimizing instructional design for students. For example, the Eye Movement Modeling Examples (EMME) [26] technique utilizes the experts' eye movements and think-aloud reasoning to direct the students' attention to the relevant information in the X-ray images. Furthermore, the multimodal dataset approach could be very helpful for use cases such as AI-based dental biometrics, which can utilize the X-ray images with the technicians' eye-tracking and verbal dictation for better detection. Similarly, in general-health condition investigation, oral X-ray images can be utilized along with the patients' medical records for a better prognosis.

### 8.3. Related Work

Databases: AE Rad [349] introduced a dataset containing 120 various Periapical dental X-ray images of the top and bottom jaw. An expert specified the abnormality for each image. The dataset solely focuses on dental caries and contains the ground truth for each image. Abdi [350] introduced a dataset consisting of anonymized and de-identified panoramic dental X-rays of 116 patients. The subjects cover a wide range of dental conditions, from healthy to partial and completely edentulous cases. The dataset consists of mandibles that two dentists manually segmented. Jader [351] introduced a dataset containing 1500 panoramic radiographs with each tooth labeled. To the best of our knowledge, these are the only datasets that are publicly available. The literature covers several studies that use privately owned datasets. Table 8.3.1 provides a comparison of the various datasets containing panoramic radiographs.

*Detection:* Kim [352] proposed a fully automated network using Convolution Neural Networks (CNNs) to detect periodontal bone loss (PBL) using panoramic dental radiographs. The

overall framework consists of multiple stages: the first stage was trained to extract the region of interest (the teeth region), the second stage focused on training a network segment and predicting the PBL lesion, the third stage consisted of using the pre-trained weight of the second stage encoder to create a classification network that predicts the existence of PBL in each tooth, and the final stage consists of a classification network that predicts the existence of PBL lesion specifically for the molar and premolar teeth.

| Table 8.3.1: A comparison of panoramic radiography datasets |                      |          |                    |               |                       |              |                     |
|-------------------------------------------------------------|----------------------|----------|--------------------|---------------|-----------------------|--------------|---------------------|
| Authors                                                     | Variable Detected    | # Images | Publicly Available | Labeled Teeth | Labeled abnormalities | Eye-tracking | Textual description |
| Oktaý [353]                                                 | Teeth                | 100      |                    | ✓             |                       |              |                     |
| Muramatsu [354]                                             | Teeth                | 100      |                    | ✓             |                       |              |                     |
| Jader [351]                                                 | Teeth                | 1500     | ✓                  | ✓             |                       |              |                     |
| Tuzoff [355]                                                | Teeth                | 1352     |                    | ✓             |                       |              |                     |
| Ekert [342]                                                 | Endodontic           | 85       |                    | ✓             |                       |              |                     |
| Fukuda [356]                                                | Endodontic           | 300      |                    |               | ✓                     |              |                     |
| TDD                                                         | Teeth, Abnormalities | 1000     | ✓                  | ✓             | ✓                     | ✓            | ✓                   |

The network was trained and tested on panoramic dental radiographs. Kim [357] investigated the application of CNNs to classify four types of implant fixtures. Periapical radiographs from 801 patients were used for this study and tested on state-of-the-art networks such as SqueezeNet [358], GoogLeNet [359], ResNet-18 [243], MobileNet-V2 [360], and ResNet-50 [358]. All five networks produced an accuracy of above 90% in detecting the four types of implants, with SqueezeNet and MobileNet displaying accuracy of 96% and 97%, respectively.

Chang [361] developed a hybrid framework to automatically detect and classify periodontal bone loss in dental panoramic radiographs. A total of 340 panoramic radiographs were used for detecting periodontal bone level, cementoenamel junction level, and teeth. A modified Mask R-CNN [129] based on a feature pyramid network and a ResNet101 backbone was employed to detect the abovementioned categories. Further, the percentage rate analysis of the radiographic bone loss combined the tooth long-axis with the periodontal bone and cementoenamel junction level. This percentage rate would automatically classify the periodontal bone loss. The Pearson correlation coefficient of the automatic method with the diagnosis by radiologists was 0.73 overall for the entire jaw and the intraclass correlation value of 0.91 for the entire jaw. Geetha [362] proposed a back-propagation neural network architecture to detect caries in intra-oral radiographs. The radiographs are passed through a Laplacian filter to obtain sharper images. The images are further passed through different adaptive thresholding filters to get the segmented output. These images are further fed to a feature extractor, and 16 statistical features are acquired. Finally, the features are passed into the back-propagation neural network to detect the presence of caries.

*Segmentation and Classification:* Muresan [363] proposed a deep learning-based semantic segmentation technique to classify and segment 15 different classes of specific tooth issues. The proposed solution consists of 5 main stages. The first stage focuses on image preparation, where the image is cropped to the required dimension, and the pixels are annotated for each image. A total of 1000 images were selected and annotated for semantic segmentation. The second stage employs the Efficient Residual Factorized Convolutional Network (ERFNet) [364] to perform semantic segmentation on the images. The application of the ERFNet on the test set resulted in an intersection over union score of 60.03%. In the third stage, different morphological operations were performed on different classes to separate the teeth better. The final part of this stage

consisted of a two-step labeling algorithm where each object was assigned a unique numeric label. The bounding box for each tooth or dental anomaly was computed for each instance of the label. The fourth stage focuses on inconsistencies caused by segmentation errors and utilizes a thresholding technique to eliminate pixel regions below the threshold. Finally, the classification is performed using a majority voting technique. This process is repeated for each tooth, and a number is associated with a dental abnormality. Oktay [353] proposed the application of a modified AlexNet [84] to classify teeth in panoramic X-ray radiographs. The images are further preprocessed to detect the mouth gap, which in turn gives the potential positions of the different classifications of teeth. The images are split into four quarters and fed to the AlexNet for tooth classification. Kim [365] presented automatic tooth detection and classification methods that combine the regional convolutional neural network (RCNN), Single Shot Detector, and heuristic methods. The proposed algorithm yielded precision values of 84.5%, sensitivity values of 75.5%, and specificity values of 80.4%.

*Forensics:* Fahmy [366] proposed an Automatic Dental Identification System (ADIS) architecture capable of handling large amounts of data. The architecture adopts a two-stage approach: the first stage reduces the number of candidate images from thousands to hundreds. The second stage reduces the number of candidate images to a match-list of a few cases to be presented to the forensic experts for their final decision. The experiment was tested and validated on a database consisting of antemortem and postmortem scans provided by the FBI. Experimental results demonstrated that, for 90% of the cases, the authentic images were among the top 10% of the retrieved images. Anil [103] developed a semi-automatic method to identify humans based on dental radiographs. The technique consisted of three stages: radiograph segmentation, tooth feature extraction, and tooth feature matching. Contours of the teeth were utilized as features to which a

transformation was applied to align the contours and correct the imaging geometric variations, thereby generating a matching distance. The proposed identification system was tested on 38 query images and indicated to be a feasible approach. However, the proposed method would fall short in cases where a) the images are very blurry, b) the query shape is partially occluded, c) there is a substantial change in the imaging angle between the antemortem and postmortem images, and d) some teeth have been extracted. Sathya [367] proposed a three-stage transfer learning approach to identify humans using digital radiographic images. The first stage uses a pre-trained AlexNet [84] to determine if the query tooth image belongs to the upper or lower jaw. The second stage further classifies the tooth as Molar, Premolar, Canine, or Incisor. Finally, the tooth number is identified and matched with the images in the antemortem dataset.

*Eye-tracking and radiology:* Karargyris [368] introduced a chest X-Ray dataset with 1083 images. This dataset included the eye gaze coordinates data along with the radiologist dictation audio and transcription. Khosravan [369] developed a collaborative computer-aided diagnosis (CAD) system that combined eye-tracking with CAD. By creating a novel algorithm for graph-based clustering and sparsification, this system could collaborate with radiologists and help them improve their diagnostic decisions. Stember [370] utilized eye tracking with deep learning for generating segmentation masks based on eye-tracking data and achieved results that were similar to hand annotations. Similarly, Aresta [371] and Mall [372] utilized gaze information for automatic lung nodule detection and mammogram interpretations. Hermanson [373] conducted a study to identify the visual search patterns of dentists while viewing and interpreting periapical dental radiographs. While this research showed that dentists tend to observe more significantly at regions of high contrast (radiopaque or radiolucent), it did not perform any correlation to the presence of abnormalities in the region of high contrast.

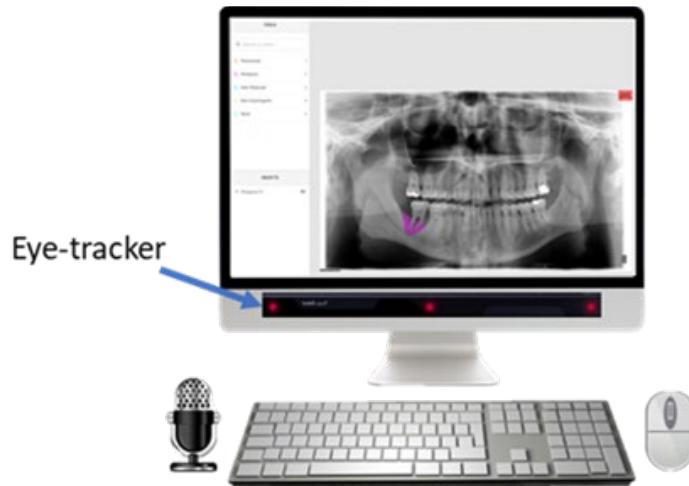


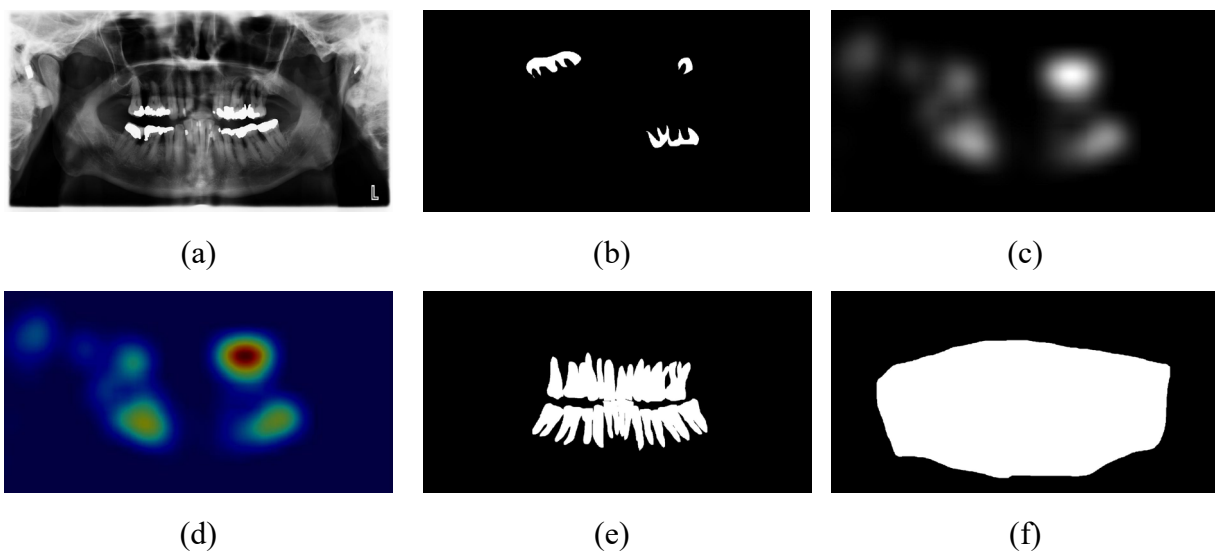
Fig. 8.3.1: The acquisition system setup. The system consists of an eye-tracker and an audio recording device.

## 8.4. Tufts Dental Database

The protocol developed to collect and publish the TDD was approved by the Tufts University Institutional Research Board (IRB ID MODCR-01-12631, approved on 7/14/2017). The protocol discussed the study's purpose and procedure, the confidentiality and public access of the database, and the protection of participants' identities. One thousand de-identified digital panoramic radiographs of Tufts University School of Dental Medicine patients have been included in this pilot study. These radiographs were randomly selected from the electronic patient database (axiUm) at the school of dental medicine. The random selection process was done through a search of the panoramic radiographic procedure codes from January 1, 2014, to December 31, 2016. Up to 1000 qualifying panoramic radiographs were selected from the generated list of chart numbers with panoramic images in the data frame. The inclusion criteria were optimum diagnostic quality

of the image with minimal or no technical errors in the image. The 1000 selected radiographs were de-identified and saved in a generic image format (TIFF/JPEG) with a unique identifier. These radiographs were annotated by an expert and a student from the Tufts University School of Dental Medicine. In addition to the annotated radiographs, eye-tracking and audio data were collected from the expert and the student. Fig. 8.3.1 displays the acquisition system setup used for this project.

The dataset consists of 6 major components: a) radiographs, b) labeled masks, c) eye tracker generated maps (gray and quantized), d) text information describing each radiograph, e) teeth mask for each radiograph with labels, and f) maxillomandibular region-of-interest mask. Fig. 8.4.1 shows an example of each imaging component. The expert and the student used Labelbox [36] to annotate and label each radiograph. To capture the annotator's eye movements, our system uses a Tobii Eye Tracker 4C, which has an operating distance of 50-95 cm [374]. The eye tracker was placed at the bottom of the screen. The audio was captured using a microphone and was synchronized with the eye tracker.



*Continued*

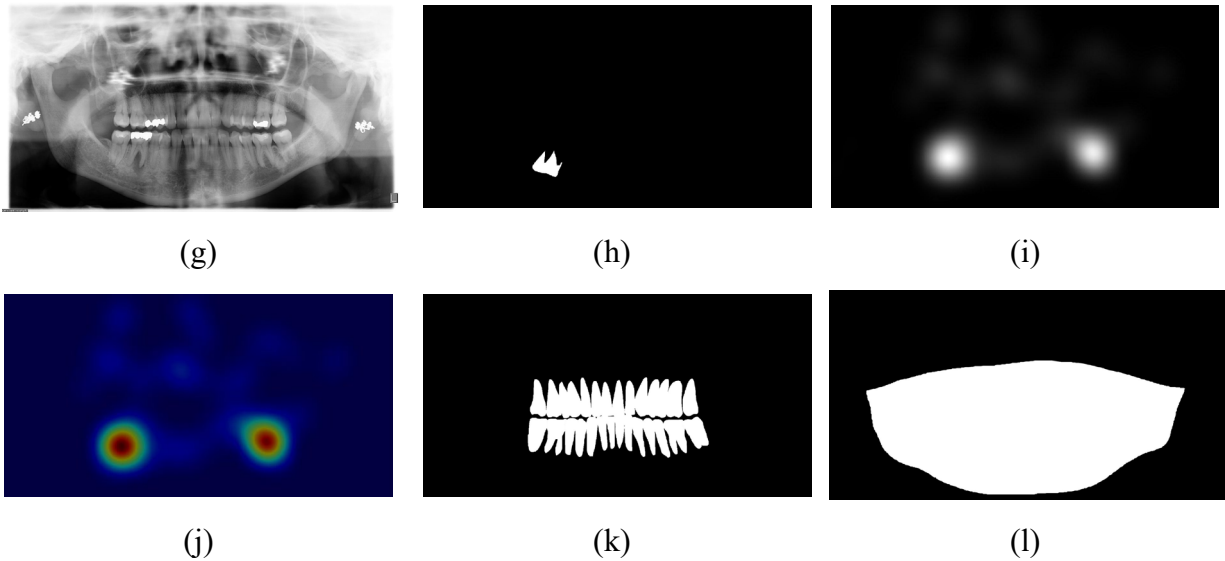


Fig. 8.4.1: Illustration of the different imaging components. (a),(g) panoramic radiograph, (b),(h) segmented mask outlining the abnormality, (c),(i) grayscale eye-tracking gaze plot, (d),(j) color quantized eye-tracking gaze plot, (e),(k) teeth mask, (f),(l) maxillomandibular region of interest.

All the digital panoramic radiographs included in this study were prescribed by clinicians based on the diagnostic and clinical needs of these patients. All panoramic images included in the dataset were obtained using OP100 Orthopantomograph (Instrumentarium Imaging/Kavo Kerr) and Plammeca Promax 2D (Henry Schein) radiographic units. Automatic exposure control was utilized to obtain clinically diagnostic images with optimal image density and contrast.

The expert utilized a systematic approach to analyze each radiograph, identify the abnormality, and describe the radiographic characteristics of the abnormality, which allowed for categorization and differential diagnosis of the entity. The student was a fourth-year dental student who had completed the Oral and Maxillofacial Radiology (OMFR) didactic course and clinical OMFR rotations with a passing grade in the radiographic interpretation case-based competency exam. The course includes instruction and assessment in the systematic analysis of radiographs.



A consistent logical sequence of steps was followed for evaluating the radiographs in this study. As each panoramic radiograph was reviewed, the expert and the student recorded their findings. Diagnostic reasoning and analytical framework have been routinely described to aid in the interpretation of diagnostic images. Therefore, the first step was recognizing the abnormality, which required a thorough knowledge of variations of normal findings. Evidence shows that the employment of a systematic search strategy by novice clinicians improves their ability to detect abnormalities in panoramic radiographs [375].

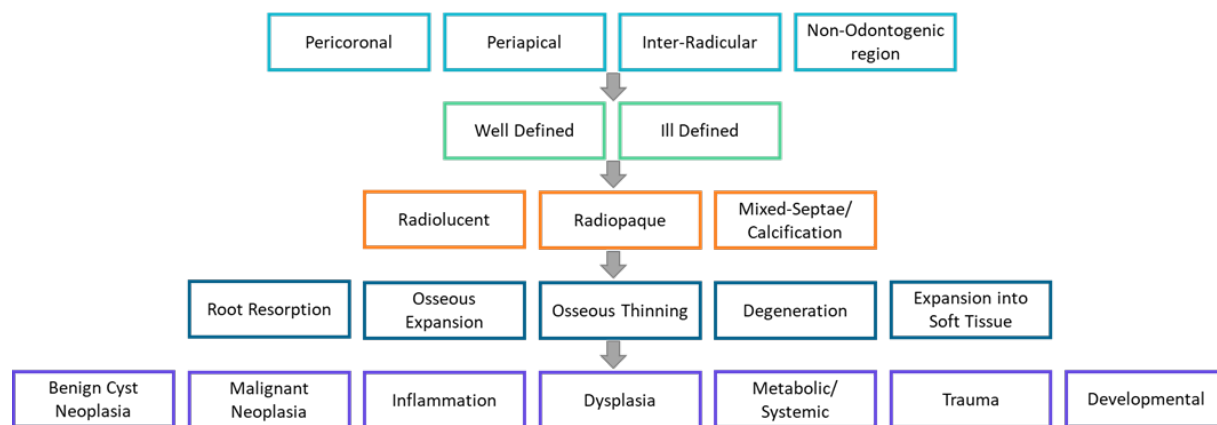


Fig. 8.4.2: Logical sequence followed while evaluating each radiograph. It depicts the five different characteristics recorded for a radiograph with abnormalities.

If an abnormality was detected, a detailed description of radiographic characteristics was recorded in five different levels, as depicted in Fig. 8.4.2. The first level described the anatomic location relative to the jaw and teeth. The second level described the periphery or marginal characteristics of the abnormality. The periphery was described primarily as well-defined versus ill-defined. The internal architecture was the third level of description and focused on the radiodensity- Radiolucent, Radiopaque, and Mixed with Septae or Calcifications. The fourth level

focused on the effects of the abnormality on adjacent structures and was recorded in terms of tooth displacement, root resorption, osseous thinning and expansion, and extension into the adjacent soft tissue or degenerative changes. Based on the overall descriptions, a category (fifth level) for the abnormality was selected from Trauma, Inflammation, Dysplasia, Developmental, Benign tumor or Cyst, Malignant Neoplasia, Systemic or Metabolic conditions. An example of a radiograph with its abnormality labeled is seen in Fig. 8.4.3.

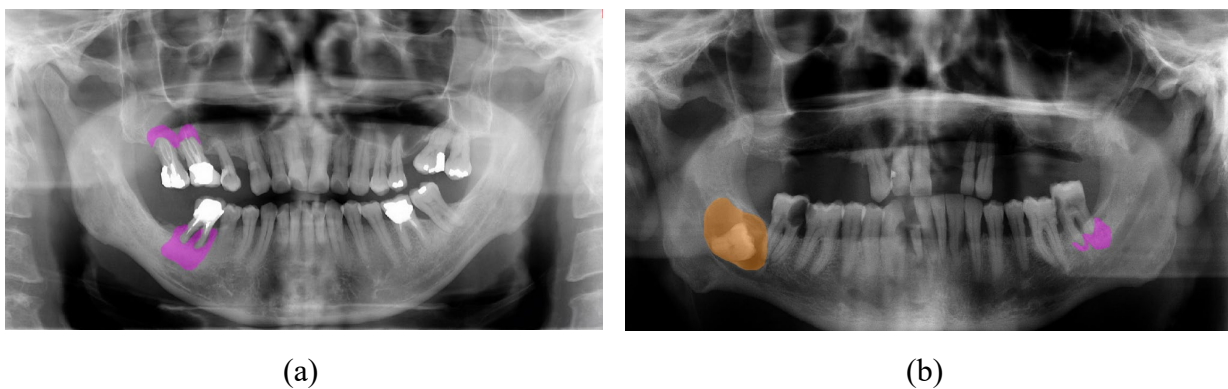


Fig. 8.4.3: An illustration of radiographs with their abnormalities labeled. (a) displays apical inflammation changes associated with teeth number 2, 3, and 30, and (b) displays a pericoronal defined corticated radiolucency associated with impacted 32, causing displacement of the tooth on the canal and possible loss of lamina dura and 31 intrauterine cyst odontogenic keratocyst of benign neoplasia there is apical inflammatory changes associated with 18.

Table 8.4.1 to Table 8.4.5 summarizes the findings of the expert and the student in describing the abnormality for all 1000 panoramic radiographs. As shown, there lie discrepancies in categorizing the abnormalities by the student. These discrepancies are a result of either missing the abnormalities in the radiograph or misclassification of the abnormalities. The five-level classification of each radiograph is recorded in the .json files for the expert and the student.

| Table 8.4.1: Summary of the categorization of abnormalities based on location |        |         |
|-------------------------------------------------------------------------------|--------|---------|
| Category                                                                      | Expert | Student |
| Periapical                                                                    | 217    | 129     |
| Non-odontogenic                                                               | 105    | 50      |
| Pericoronal                                                                   | 48     | 29      |
| Inter-radicular                                                               | 4      | 9       |
| None                                                                          | 660    | 802     |

| Table 8.4.2: Summary of the categorization of abnormalities based on the internal architecture |        |         |
|------------------------------------------------------------------------------------------------|--------|---------|
| Category                                                                                       | Expert | Student |
| Radiolucent                                                                                    | 128    | 151     |
| Radiopaque                                                                                     | 125    | 65      |
| Mixed-Septae/Calcification                                                                     | 109    | 1       |
| None                                                                                           | 11     | 0       |

| Table 8.4.3: Summary of the categorization of abnormalities based on its effects on adjacent structures |        |         |
|---------------------------------------------------------------------------------------------------------|--------|---------|
| Category                                                                                                | Expert | Student |
| Root Resorption                                                                                         | 29     | 1       |
| Osseous Expansion                                                                                       | 4      | 2       |
| Osseous Thinning                                                                                        | 18     | 0       |
| Degeneration                                                                                            | 9      | 9       |
| Expansion into Soft Tissue                                                                              | 1      | 0       |
| None                                                                                                    | 341    | 205     |

| Table 8.4.4: Summary of the categorization of abnormalities based on marginal characteristics |        |         |
|-----------------------------------------------------------------------------------------------|--------|---------|
| Category                                                                                      | Expert | Student |
| Well Defined                                                                                  | 231    | 216     |
| Ill Defined                                                                                   | 143    | 1       |
| None                                                                                          | 0      | 0       |

| Table 8.4.5: Summary of the final categorization (fifth level) of abnormalities |        |         |
|---------------------------------------------------------------------------------|--------|---------|
| Category                                                                        | Expert | Student |
| Benign Cyst Neoplasia                                                           | 39     | 25      |
| Malignant Neoplasia                                                             | 2      | 0       |
| Inflammation                                                                    | 226    | 132     |
| Dysplasia                                                                       | 8      | 49      |
| Metabolic/Systemic                                                              | 12     | 17      |
| Trauma                                                                          | 9      | 2       |
| Developmental                                                                   | 121    | 12      |
| None                                                                            | 10     | 0       |

The regions of abnormality are outlined using the Labelbox software, and the ground truth masks are generated. As the expert and the student analyzed each radiograph, their eye movements were recorded simultaneously using the Tobii eye tracker. The Tobii starter kit [374] was employed to generate a file for each radiograph containing the x and y coordinates.

These coordinates are further processed to generate the gaze plot. An example of the gaze plot can be seen in Fig. 8.4.4. The gaze plot displays the location and time spent looking at each location on the radiograph [374]. The time spent looking at a location, or commonly known as

fixation time, is shown by the diameter of the fixation circle. As the fixation time increases, the diameter of the circle also increases.

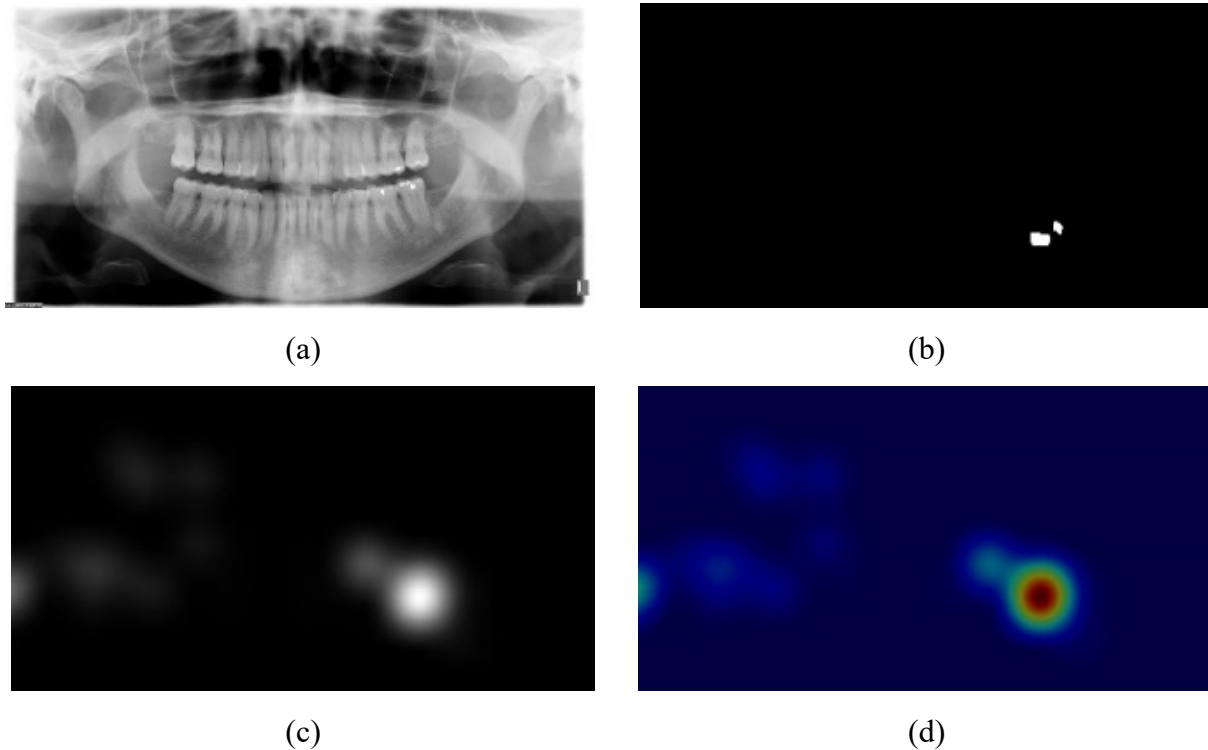


Fig. 8.4.4: Gaze plots generated from the eye-tracking data. (a) radiograph, (b) segmented mask outlining the abnormality, (c) the grayscale gaze plot displaying the expert looking at the abnormality, and (d) color quantized eye-tracking gaze plot.

The audio files acquired from the expert and the student describing each radiograph were converted to text using Google [376] and Dragon [377] automatic speech recognition platforms. The text generated for each radiograph is added to the .json file. The structure of the .json file containing the label and text information is shown in Fig. 8.4.5. In addition, our dataset has teeth labeled by the expert for all radiographs. These images also have the maxillomandibular region of interest annotated, where researchers can emphasize on the information contained within this region. Fig. 8.4.6 provides a sample of a radiograph with the teeth annotated and its

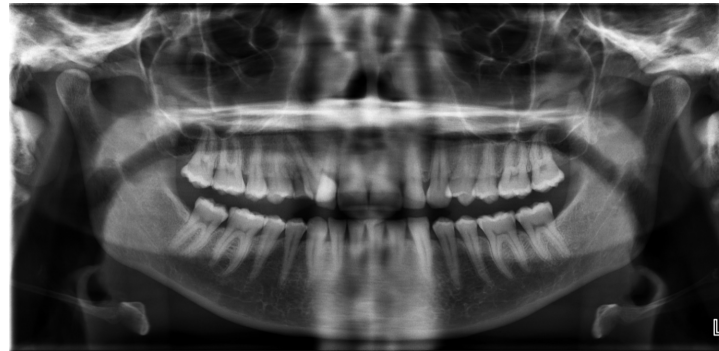
maxillomandibular region of interest. The universal tooth numbering system was used to identify and label teeth in all the panoramic images.

```
"Label": {
  "objects": [
    {
      "title": "None",
      "value": "none",
      "classifications": "none",
      "polygons": "none"
    }
  ],
  "classifications": []
},
"Description": [Within normal limits]
"External ID": "570.JPG"
},
```

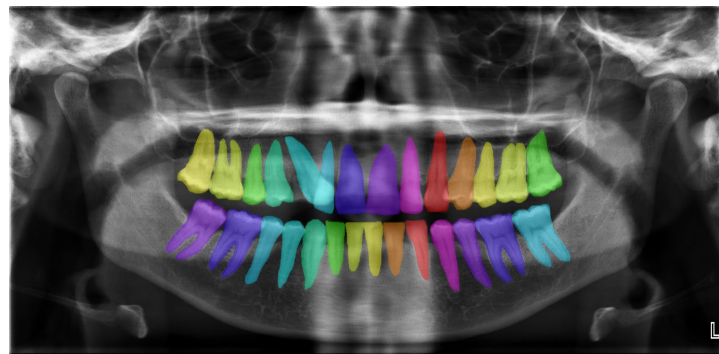
Fig. 8.4.5: Structure of the .json file. The file contains information on abnormalities pertaining to each radiograph.

The TDD contains 9000 images and 3 JSON files describing each radiograph. These images belong to 4 main subsets: panoramic radiographs, the labeled masks highlighting the abnormalities, the eye-tracking gaze plots (gray and quantized), and the labeled teeth mask. The labeled masks and gaze plots are present for both the expert and the student. The availability of this database is critical for research transparency in the field. Therefore, TDD will be publicly available for download from our research website on the Tufts ECE servers. To access the database, researchers will need to submit a request form to receive the download permission. The

information on obtaining a copy of the TDD can be found on the project website (<http://tdd.ece.tufts.edu/>). Fig. 8.4.7 shows the folder structure of the TDD.



(a)



(b)



(c)

Fig. 8.4.6: A sample of the annotated teeth and the region of interest. (a) panoramic radiograph, (b) teeth mask overlaid on the panoramic radiograph, and (c) maxillomandibular region of interest

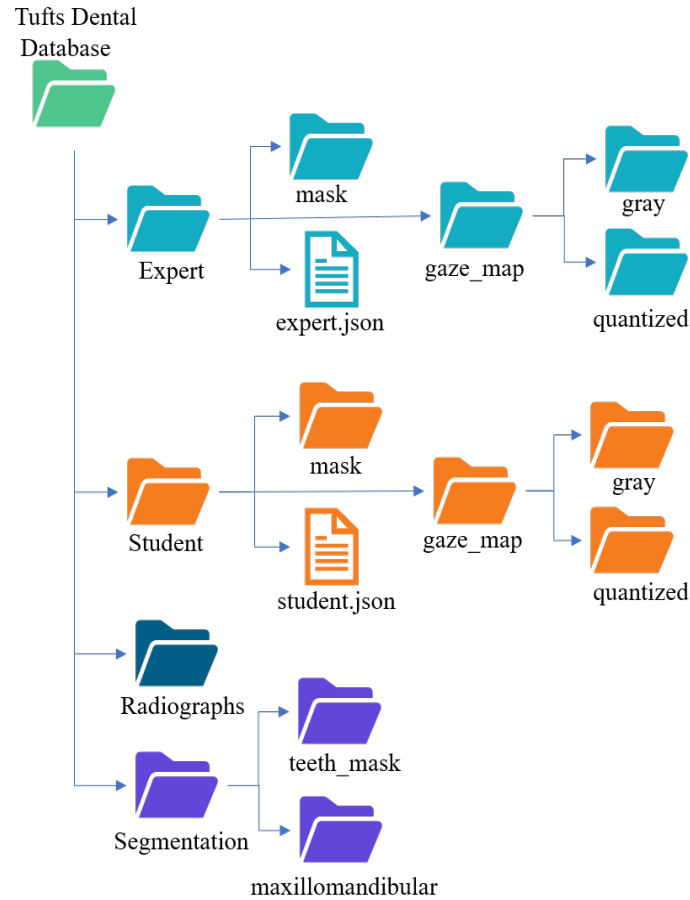


Fig. 8.4.7: Folder structure of the tufts dental database

## 8.5. Performance Review

In this section, a practical evaluation of the TDD is presented. In particular, the dataset is evaluated on recent state-of-the-art image enhancement and segmentation techniques. The performance of the segmentation and enhancement algorithms are assessed using popular metrics in computer vision and deep learning fields.



### 8.5.1. Image Enhancement

In diagnostic radiography, the goal of image enhancement is to make the image visually more appealing. This is accomplished by increasing contrast, optimizing brightness, improving sharpness, and reducing noise. Subjective image enhancement is not expected to improve the accuracy of image interpretation and is often diagnostic task-specific [378]. Images from this dataset are tested on methods previously used for medical image enhancement. These methods include Histogram Equalization (HE) [289], Contrast Limited Adaptive Histogram Equalization (CLAHE) [379], Bi-Histogram Equalization (BBHE) [380], Dynamic Fuzzy Histogram Equalization (DFHE) [381], and Guided Filtering (GF) [382].

| Table 8.5.1: Quality scores on a subset of images from the TDD. CLAHE algorithm performs better in comparison |       |        |              |               |       |        |       |        |              |        |
|---------------------------------------------------------------------------------------------------------------|-------|--------|--------------|---------------|-------|--------|-------|--------|--------------|--------|
| Image Id                                                                                                      | HE    |        | CLAHE        |               | BBHE  |        | DF    |        | GF           |        |
|                                                                                                               | AME   | LogAME | AME          | LogAME        | AME   | LogAME | AME   | LogAME | AME          | LogAME |
| 2.jpg                                                                                                         | 30.20 | 0.5011 | <b>26.69</b> | <b>0.4172</b> | 32.79 | 0.5517 | 33.19 | 0.5532 | 28.88        | 0.8585 |
| 33.jpg                                                                                                        | 28.02 | 0.4507 | <b>25.50</b> | <b>0.3853</b> | 31.95 | 0.4987 | 33.09 | 0.5384 | 28.74        | 0.8264 |
| 165.jpg                                                                                                       | 24.72 | 0.3817 | <b>21.79</b> | <b>0.3033</b> | 26.75 | 0.4026 | 28.48 | 0.4552 | 24.13        | 0.7144 |
| 278.jpg                                                                                                       | 25.04 | 0.3622 | <b>21.79</b> | <b>0.3131</b> | 25.01 | 0.3792 | 25.69 | 0.4186 | 22.09        | 0.7098 |
| 374.jpg                                                                                                       | 29.25 | 0.4754 | <b>27.31</b> | <b>0.4466</b> | 31.54 | 0.5312 | 32.27 | 0.5744 | 29.92        | 0.8989 |
| 492.jpg                                                                                                       | 27.27 | 0.4233 | <b>22.50</b> | <b>0.3299</b> | 28.88 | 0.4490 | 29.49 | 0.5084 | 24.67        | 0.7602 |
| 578.jpg                                                                                                       | 23.66 | 0.3404 | 21.68        | <b>0.3197</b> | 22.07 | 0.3447 | 23.24 | 0.4114 | <b>19.51</b> | 0.7043 |
| 665.jpg                                                                                                       | 26.46 | 0.4125 | <b>23.81</b> | <b>0.3565</b> | 27.59 | 0.4359 | 29.28 | 0.4911 | 25.74        | 0.7775 |
| 759.jpg                                                                                                       | 27.62 | 0.4420 | <b>24.60</b> | <b>0.3764</b> | 29.82 | 0.4780 | 30.70 | 0.518  | 26.13        | 0.7977 |
| 908.jpg                                                                                                       | 25.68 | 0.3815 | 21.62        | <b>0.3270</b> | 24.23 | 0.3897 | 24.44 | 0.4325 | <b>20.29</b> | 0.7344 |
| Dataset Average                                                                                               | 27.28 | 0.4310 | <b>24.32</b> | <b>0.3702</b> | 28.51 | 0.4611 | 29.14 | 0.4937 | 25.53        | 0.7970 |

In general, the performance these enhancement methods are assessed using reference-based measures such as SSIM and PSNR. However, due to the unavailability of a ground truth, non-reference-based measures are employed. AME and LogAME [15, 16, 383-386] are Human Visual System inspired quality metrics for testing algorithms and new datasets in the absence of a reference. Each image in the dataset was enhanced, and their quality scores were evaluated. An example set of the images and their respective quality scores are seen in Fig. 8.5.1 and Fig. 8.5.2. As seen in the figures, CLAHE performs better in comparison in enhancing the contrast of the images. A lower AME and LogAME measure is evidence of higher image contrast [383]. Visually comparing the results of each algorithm indicates that GF performs well in enhancing edges when compared to HE, BBHE, and DFHE. To quantify the performance of each enhancement algorithm over the entire dataset, the quality scores for each enhanced image is computed. The average across all images resulted in CLAHE performing better with an average AME score of 24.32 and a LogAME score of 0.3702. Table 8.5.1 displays the average scores and scores for a subset of images. As seen in the table, the CLAHE algorithm exhibits the best performance in enhancing the contrast of the images.

### 8.5.2. Teeth Segmentation

The performance of different deep learning-based algorithms is evaluated for segmenting teeth from panoramic radiographs. An example of the radiograph and its respective ground truth mask can be visualized in Fig. 8.4.1 (a) and Fig. 8.4.1 (e). The performance of FPN [387], UNet [128], UNet++ [388], PSPNet [389], DeeplabV3 [249], and DeeplabV3+ [390] is evaluated. Each of these networks were trained with a VGG19 [391], ResNet18 [243], and ResNet50 [243] backbone and was initialized using the ImageNet pre-trained weights.



(a)

$AME=32.3823, LogAME=0.5263$



(b)

$AME=27.5577, LogAME=0.4393$



(c)

$AME=25.0855, LogAME=0.3792$



(d)

$AME=31.6183, LogAME=0.4877$



(e)

$AME=32.0939, LogAME=0.5214$



(f)

$AME=27.9844, LogAME=0.8128$

Fig. 8.5.1: Results of the state-of-the-art enhancement algorithms where (a) input image, (b) HE, (c) CLAHE, (d) BBHE, (e) DF, and (f) Guided filtering



(a)

$AME=31.3351, LogAME=0.5364$



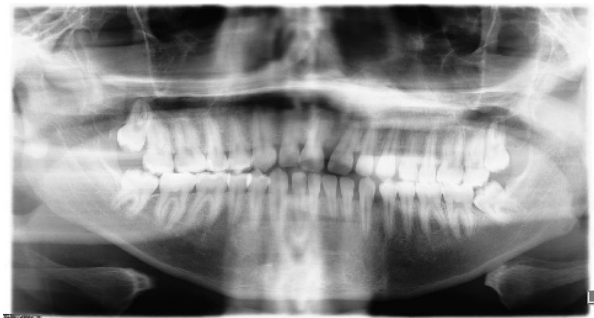
(b)

$AME=28.1996, LogAME=0.4529$



(c)

$AME=24.9148, LogAME=0.3834$



(d)

$AME=30.6813, LogAME=0.4875$



(e)

$AME=31.1098, LogAME=0.5258$



(f)

$AME=26.9720, LogAME=0.8090$

Fig. 8.5.2: Results of the state-of-the-art enhancement algorithms where (a) input image, (b) HE, (c) CLAHE, (d) BBHE, (e) DF, and (f) Guided filtering

The data comprised of 1000 radiographs and ground truth masks, which were split into training (85%) and validation (15%) sets. The models were trained on NVIDIA Tesla V100 GPUs with 32-GB memory. For training, the Adam optimizer [89] is used with an initial learning rate of  $1 \times 10^{-3}$  and weight decay of  $1 \times 10^{-4}$ . All models are trained for 150 epochs with a batch size of 16. To evaluate the performance of each model, the most popular metrics for assessing the accuracy of segmentation algorithms were employed, namely, Pixel Accuracy (PA) [392], Intersection over Union (IoU) [392], and Dice coefficient [392].

| Table 8.5.2: Summary of the performance of the state-of-the-art algorithms in segmenting teeth from panoramic radiographs. The UNet architecture performs marginally better in comparison |              |              |          |            |         |          |            |         |              |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------|--------------|----------|------------|---------|----------|------------|---------|--------------|
| Model                                                                                                                                                                                     | VGG19        |              |          | ResNet18   |         |          | ResNet50   |         |              |
|                                                                                                                                                                                           | PA (%)       | IoU (%)      | Dice (%) | PA (%)     | IoU (%) | Dice (%) | PA (%)     | IoU (%) | Dice (%)     |
| FPN                                                                                                                                                                                       | 95.11        | 86.39        | 92.30    | 95.17      | 86.37   | 92.24    | 95.08      | 86.42   | 92.29        |
| UNet                                                                                                                                                                                      | <b>95.22</b> | <b>86.67</b> | 92.46    | 95.11      | 86.42   | 92.27    | 95.13      | 86.49   | 92.36        |
| UNet++                                                                                                                                                                                    | 95.19        | 86.62        | 92.47    | 95.15      | 86.54   | 92.43    | 95.16      | 86.62   | <b>92.49</b> |
| PSPNet                                                                                                                                                                                    | 95.00        | 86.08        | 91.77    | 94.76      | 85.66   | 91.49    | 94.95      | 86.04   | 91.70        |
| DeepLabV3                                                                                                                                                                                 | -            | -            | -        | 94.91      | 86.02   | 91.87    | 95.00      | 86.21   | 92.00        |
| DeepLabV3+                                                                                                                                                                                | -            | -            | -        | 95.13      | 86.41   | 91.80    | 95.07      | 86.33   | 92.29        |
|                                                                                                                                                                                           |              |              |          |            |         |          |            |         |              |
|                                                                                                                                                                                           | PA (%)       |              |          | IoU (%)    |         |          | Dice (%)   |         |              |
| nnUNet**                                                                                                                                                                                  | 94.91        |              |          | 86.11      |         |          | 90.86      |         |              |
|                                                                                                                                                                                           |              |              |          |            |         |          |            |         |              |
|                                                                                                                                                                                           | 150 epochs   |              |          | 250 epochs |         |          | 400 epochs |         |              |
|                                                                                                                                                                                           | PA (%)       | IoU (%)      | Dice (%) | PA (%)     | IoU (%) | Dice (%) | PA (%)     | IoU (%) | Dice (%)     |
| CE-Net**                                                                                                                                                                                  | 86.52        | 71.13        | 76.92    | 91.68      | 79.75   | 84.98    | 92.67      | 81.64   | 86.62        |
| ** no backbone and pretrained weights used                                                                                                                                                |              |              |          |            |         |          |            |         |              |

Table 8.5.2 summarizes the findings of the segmentation algorithms. The table displays the scores for each model with its respective backbone architecture. The UNet and UNet++ perform marginally better in comparison. The average PA, IoU, and Dice coefficient score across all models is 95.01%, 86.19%, and 91.96%, respectively. In addition, the performance of the nnUNet [393] and CE-Net [255] architecture were tested, as these architectures were specifically designed for medical image segmentation. Both architectures were trained from scratch without the use of a backbone network. A generalized version of the nnUNet architecture is utilized and the model is trained for 150 epochs with the same hyperparameters. The nnUNet produced comparable results as seen in Table 8.5.2. In the case of CE-Net, the architecture required a lot more training iteration to achieve comparable results. After 400 epochs, the network produced a 92.67% PA score.

## **8.6. Discussion**

The applications of AI and image interpretation to generate accurate and reliable diagnosis is expanding. AI in Dentistry has a proven potential to offer assistance to clinicians in providing appropriate patient care and early diagnosis, enhance the accuracy of diagnosis, treatment planning and predicting as well as monitoring outcomes while improving efficiencies, serving as a secondary opinion, value add in forensic diagnosis, measurement of clinical/treatment outcomes and objective management of appropriate insurance coverage. Ideally, with ongoing advances in imaging and technology, computer systems could advance to an economically viable state enabling AI to serve as an adjunct to the clinical acumen of the radiologist [394].

Eye-tracking has been extensively used for research in the field of marketing, image interpretation, and psychology [395-397]. In particular, eye tracking has been evolving and expanding in understanding diagnostic interpretations in the field of medicine [368, 371-373, 398-403]. In dentistry, the interpretation of radiographs interweaves the process of perception (visually scanning the image) and cognition (decision making and diagnostic reasoning) [404]. The eye-tracking technology can be employed to specifically determine what the observer is focusing on within the image and further illustrate patterns in the scanning process. The application of eye-tracking in dentistry has provided novel opportunities to study the interpretive process and elucidate the differences in decision making, perception, misinterpretation, and misdiagnosis between novices and experts.

There is a lack in availability of annotated data and a need for a framework for incorporating eye-tracking data into AI systems. Annotating data typically involves manually tracing out regions or objects in the images. This approach is time-consuming when dealing with a large dataset and requires an expert, such as a radiologist, to perform these annotations. Moreover, most of the annotation work is performed outside clinical hours and by novice clinicians, possibly reducing the accuracy of the annotations. With the help of AI-based systems and eye-tracking technology, the process of annotating data can be automated and performed by experts during clinical hours. Advances are being made to improve disease classification by integrating eye-tracking data with deep learning techniques [369-372]. Fig. 8.6.1 demonstrates the automated annotation process. As seen in Fig. 8.6.1, the radiographs are first presented to the expert on a screen. The eye-tracking information is captured while the expert analyzes the radiograph. The captured eye-tracking data is then fed into the AI system to generate the annotations. In some cases, the findings of the radiographs are transcribed and can be used within

the AI to highlight the type of abnormality. This system can tremendously speed up the process of annotation and ensure accuracy.

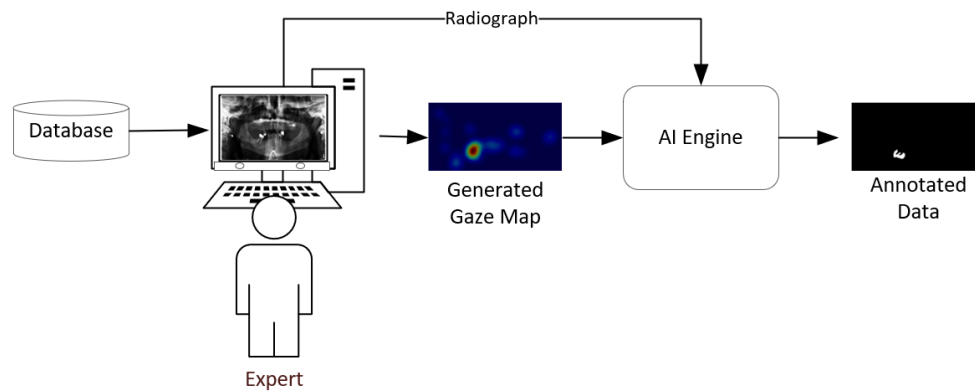


Fig. 8.6.1: Illustration of the automated annotation process using eye-tracking and AI

Researchers are striving to create AI models that can match or even surpass human capabilities. To truly mimic an expert, the AI system must consider the visual perception and cognition of the human. Fig. 8.6.2 illustrates an AI system trained to perceive radiographs as humans do. In the first pass, the radiograph is fed into the AI system and trained to predict what the expert looked at. The second pass involves passing the learned eye gaze map with the radiograph to classify the type of abnormality. Fusing different aspects to train an AI model paves the way for a new generation of AI systems to highlight the identified features and explain the reasons for the outputs generated. The trained AI system will significantly reduce errors and serve as a second opinion to the radiologist. It can further address and bridge the gaps in understanding how different radiologists perceive similar radiographs and analyze patterns or short-cuts used by the expert.

The trained AI system can further be used to teach and assist novice clinicians in interpreting radiographs. Instructors and clinicians can use the trained AI to demonstrate and assess



gaze patterns during medical training and education, in turn accelerating the transition from novice to expert. Fig. 8.6.3 illustrates a training tool example that evaluates the findings of novice clinicians. First, both the AI and the student are presented with a radiograph. The eye tracker records the movements and areas concentrated on by the novice. This acquired gaze map is then compared to the gaze map generated by the “expert AI” and the results are presented to the novice.

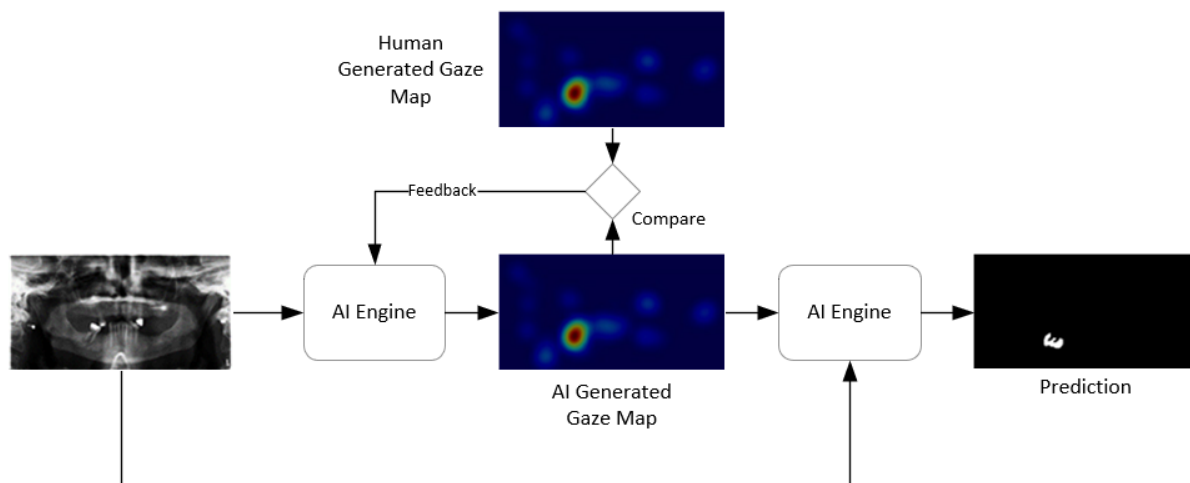


Fig. 8.6.2: Illustration of an AI system trained to perceive radiographs as humans do

Incorporating eye-tracking capabilities for the AI used in image interpretation could facilitate troubleshooting and analyzing misdiagnosis as it would provide some access to the ‘logic’ of AI-based decisions. An enhancement to education would be in the ability to allow for self-paced learning for novice clinicians who could learn specifically from expert eye-movement patterns, mimicked by AI, in the absence of the expert. Interprofessional engagement of clinicians, data scientists and computer engineers is key to the development of effective, efficient artificial intelligence tools that are framed by ethical guidelines.

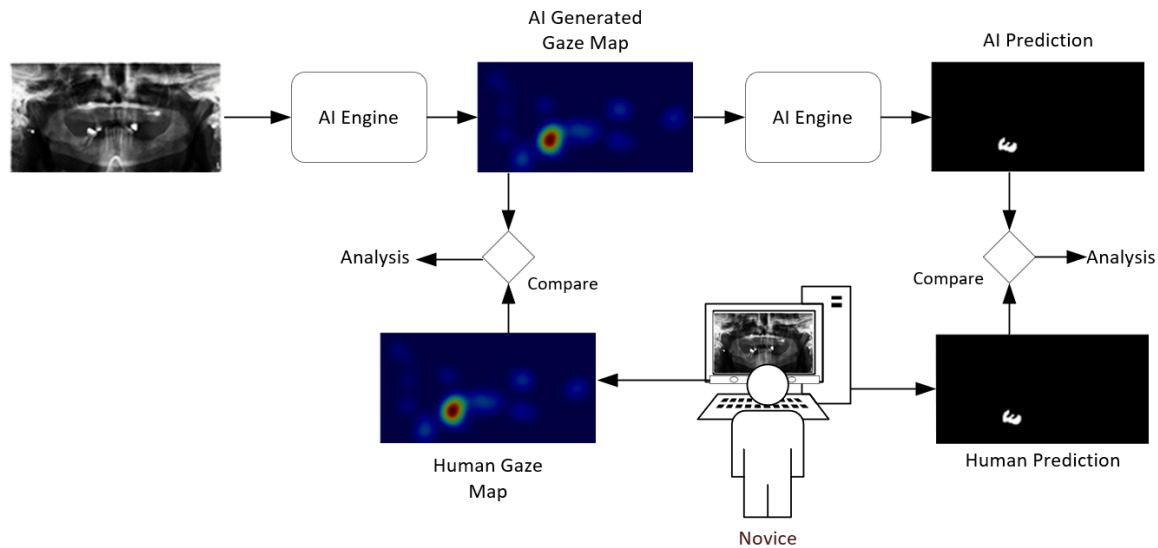


Fig. 8.6.3: A training tool that evaluates the findings of novice clinicians against the “expert AI”

The resultant artificial intelligence protocols could be employed not only for initial screening supporting clinicians by improving the efficiency but could also be employed in training future clinicians, training and improving machine learning algorithms, as well as medical claims processing. Some of the perceived limitations of intelligent AI systems that need to be addressed in the coming future are around medicolegal regulation for misdiagnoses, impact of these impersonal diagnostics on individual patients- their rights and identity, adequate planning for secure management of data accessed along with guidelines for ethical standards. Other considerations are related to consent, storage of the data, access along with the use of data currently and in the future.

## 8.7. Summary

In this work, the Tufts Dental Database, a comprehensive new dataset with the potential to revolutionize the dental AI practice, has been presented. TDD consists of a tooth segmentation mask, abnormality mask, maxillomandibular region of interest mask, eye-tracking gaze map, and text description of the abnormality. A detailed evaluation of baseline results for dental radiographic image enhancements and tooth segmentation is also provided. Furthermore, TDD opens the opportunity to seamlessly integrate the radiologist's expertise into the AI system, making it more efficient and reliable. This dataset also paves the way for a new generation of AI systems to highlight the identified features and explain the reasons for the outputs generated.

## Chapter 9. Conclusion and Future Work

This dissertation leverages the advances in computer vision and AI to develop novel algorithms and tools that analyze and interpret multimodal data. This dissertation focuses on two major components: feature engineering and image analytics, to facilitate the development of efficient models and accelerate the adoption of AI in medical and forensic applications. Feature engineering is further divided into feature enhancement and feature detection and matching. The image analytics section focuses on developing algorithms and datasets for medical image diagnosis.

In the feature enhancement aspect of this dissertation, a novel edge-aware image enhancement method is developed to improve the quality of images (Chapter 2). The method utilizes guided filtering, edge and contrast enhancement, and image fusion to suppress unwanted features and enhance the perception of information in panoramic radiographs. The method uses a directional truncated kernel that improves the edge detection process. In addition, a contrast enhancement technique based on a sigmoid function is presented. The proposed method performs better than the state-of-the-art techniques in enhancing features in the x-ray images while suppressing the noisy components. While visual analysis of the images was sufficient, we need to quantify the performance of the enhancement algorithms to promote an automated workflow. As panoramic radiographs have a complex structure, off-the-shelf quality measures do not match the visual assessment score. Hence, there is a need for a non-reference based quality metric that incorporates the Mean Opinion Score of the expert. Furthermore, CNNs can be utilized to incorporate the Mean Opinion Score and generate accurate image quality predictions.

In chapter 3, various enhancement techniques were used to create synthetic data and improve the accuracy of the thermal emotion recognition system-TERNet. The system utilized a transfer learning approach where the network is initially trained on visible images and fine-tuned to predict emotions in thermal images. The system achieved an accuracy of 96.2 % in categorizing four different expressions. While thermal face images are invariant to illumination changes, they are sensitive to temperature changes in the surrounding environment and variations in the heat patterns of the face. As a future task, the possibility of fusing thermal and visible face images to develop a robust emotion recognition system will be explored. In addition, generative adversarial networks can be employed to generate synthetic data and eliminate the bias factor in datasets.

The feature detection and matching section (Chapter 4) provides an overview of the different image feature detectors. These detectors were utilized to a) match postmortem and antemortem panoramic radiographs and b) develop a pixel-based color transfer algorithm. Different feature detectors were used to extract features from the postmortem and antemortem scans. The Multiscale Hessian detector provided the best matching score that could accurately measure the similarities between the radiographic scans. To make this system more robust, a larger dataset of corresponding antemortem and postmortem scans is required. Furthermore, features such as tooth shape, size, and location can be added to the feature measure to generate a much more accurate prediction.

The color transfer algorithm utilized the feature detectors and homography estimation to align images. Then, a pixel-to-pixel color transfer technique was developed to recolor nighttime imagery. The algorithm's performance was evaluated using the RGB-NIR Scene and the AMOS dataset. The algorithm was able to seamlessly map the information in the visible images to the NIR and nighttime images. In addition, the algorithm was extended to perform visible and thermal face

image fusion. Finally, the Alpha-Winzorized and the Weighted Rank Order Statistic, two adaptive correlation techniques, were developed to accurately stitch images and generate a panoramic view of the environment. The stitching algorithm can be used for a wide range of applications, but for the scope of this work, the algorithm was utilized to stitch underwater images. The algorithm employs the hessian detector to extract the image features and generate a holistic view of the aquatic environment. In addition, a novel underwater image quality measure, CTSUIQM, has been developed. The measure uses both transform domain and spatial domain measures and combines them using the human visual system. The experimental results display the measure's success in assessing the similarities between the ground truth and stitched images. As deep learning techniques have proven to generate superior results when compared to classical image processing techniques, future work will include CNN-based algorithms to perform color transfer and image stitching. In addition, future applications involving image stitching, such as aerial and computational photography, will be explored.

In the image analytics aspect of this dissertation, first, a novel 2D CNN architecture, KNet, to segment kidneys and tumors from CT images is presented (Chapter 5). The architecture proposes a dual decoder to generate individual semantic maps for the kidney and tumor regions. The model uses a channel attention mechanism and a feature enhancement block to improve the segmentation performance. KNet was trained on the KITS database for 250 epochs and was evaluated using the Pixel Accuracy, F1, and IoU metrics. The model outperformed the state-of-the-art techniques by achieving 96.07% combined accuracy with just 2.9 million parameters. In addition, a quantitative analysis displayed that KNet achieved superior semantic maps for the kidney and tumor regions with sharp boundaries. The work also presents an ablation study that demonstrates the process of selecting and building the architecture. As a part of future work, the

possibilities of usage of KNet for other medical fields such as lung cancer segmentation and mammogram segmentation will be explored. KNet can be modified to track and identify a disease's progression in certain patients. Furthermore, owing to the lightweight structure of KNet, the possibilities of deploying it on edge devices will also be explored.

Next, a study of renal mass tumors in CT images is presented, where tumoral topography can be quantified and correlated with histology and biological aggressiveness in small renal masses (Chapter 7). Preliminary data suggests the use of imaging topography and genetic markers could provide additional information in the selection of patients for active surveillance or monitoring of renal masses compared to more aggressive therapy. The method utilizes a multi-level box-counting approach to estimate the aggressiveness of the tumor. A linear regression analysis noted that the roughness score is the most influential on the model, with all other demographics being equal, including tumor size. As part of future work, a larger study is necessary to validate these findings and to determine if radiographic tumor surface analysis can prevent the need for conventional serum or urine biomarkers.

The next chapter of this dissertation investigates a Human Visual Thresholding (HVT) method to binarize color and grayscale images (Chapter 6). The thresholding technique was utilized to segment cell structures from Haematoxylin and Eosin stained images. In addition, an image standardization technique is used to eliminate the color variations in the staining methods. Due to the unavailability of ground truth images, the proposed method was visually analyzed and compared to various state-of-the-art thresholding techniques. The results display the method's superior performance in segmenting nuclei in situations where the difference between the background and the foreground is minimal. The HVT method can potentially automate the process

of creating ground truth data for whole-slide images, thereby reducing the clinician's time in annotating such datasets.

The final section presents the Tufts Dental Database, a multimodal dataset that captures the radiologist's expertise while annotating and interpreting panoramic dental radiographs. The expertise is captured in the form of eye-tracking gaze maps and the think-aloud protocol. The first-of-its-kind dataset presents 1000 panoramic dental radiography images with expert labeling of teeth and abnormalities. The dataset consists of 6 major components: a) radiographs, b) labeled masks, c) eye tracker generated maps (gray and quantized), d) text information describing each radiograph, e) teeth mask for each radiograph with labels, and f) maxillomandibular region-of-interest mask. In addition to the expert-generated labels, the finding of a fourth-year dental student is also presented. Furthermore, a detailed evaluation of baseline results for dental radiographic image enhancements and tooth segmentation is also provided. The Tufts Dental Database opens the opportunity to seamlessly integrate the radiologist's expertise into the AI system, making it more efficient and reliable. This dataset paves the way for the development of a new generation of AI systems capable of highlighting features and justifying the outputs generated. Furthermore, the process of data annotation can be automated by training the AI to incorporate eye-tracking and audio information, eliminating the labor-intensive process of manual annotation.



## Reference

- [1] A. Panesar, *Machine learning and AI for healthcare*. Springer, 2019.
- [2] M. Y. Shaheen, "Applications of Artificial Intelligence (AI) in healthcare: A review," *ScienceOpen Preprints*, 2021.
- [3] Z. Geradts, "The Application of Artificial Intelligence (AI) in Digital Forensic Science," in *Proceedings of the American Academy of Forensic Sciences*, 2020, vol. 26.
- [4] S. B. Khanagar *et al.*, "Application and performance of artificial intelligence technology in forensic odontology—A systematic review," *Legal Medicine*, vol. 48, p. 101826, 2021.
- [5] J. C. Ahn, A. Connell, D. A. Simonetto, C. Hughes, and V. H. Shah, "Application of artificial intelligence for the diagnosis and treatment of liver diseases," *Hepatology*, vol. 73, no. 6, pp. 2546-2563, 2021.
- [6] S. Waite *et al.*, "Analysis of perceptual expertise in radiology—Current knowledge and a new perspective," *Frontiers in human neuroscience*, vol. 13, p. 213, 2019.
- [7] T. Nawrocki, P. D. Maldjian, S. E. Slasky, and S. G. Contractor, "Artificial intelligence and radiology: have rumors of the radiologist's demise been greatly exaggerated?," ed: Elsevier, 2018.
- [8] P. K. Rao, M. Tharmavaram, and G. Pandey, "Conventional Technologies in Forensic Science," *Technology in Forensic Science: Sampling, Analysis, Data and Regulations*, pp. 17-34, 2020.
- [9] N. Sunde and I. E. Dror, "Cognitive and human factors in digital forensics: Problems, challenges, and the way forward," *Digital investigation*, vol. 29, pp. 101-108, 2019.
- [10] G. Ribeiro, J. M. Tangen, and B. M. McKimmie, "Beliefs about error rates and human judgment in forensic science," *Forensic science international*, vol. 297, pp. 138-147, 2019.
- [11] F. Zhuang *et al.*, "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43-76, 2020.
- [12] L. Azzopardi, "Cognitive biases in search: a review and reflection of cognitive biases in Information Retrieval," in *Proceedings of the 2021 conference on human information interaction and retrieval*, 2021, pp. 27-37.
- [13] K. Panetta *et al.*, "A comprehensive database for benchmarking imaging systems," *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 3, pp. 509-520, 2018.
- [14] A. Almuntashri, E. Finol, and S. Agaian, "Automatic lumen segmentation in CT and PC-MR images of abdominal aortic aneurysm," in *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2012: IEEE, pp. 2891-2896.
- [15] K. Panetta, Y. Zhou, S. Agaian, and H. Jia, "Nonlinear unsharp masking for mammogram enhancement," *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 6, pp. 918-928, 2011.
- [16] K. Panetta, C. Gao, and S. Agaian, "No reference color image contrast and quality measures," *IEEE transactions on Consumer Electronics*, vol. 59, no. 3, pp. 643-651, 2013.
- [17] S. S. Agaian, "Visual morphology," in *Electronic Imaging'99*, 1999: International Society for Optics and Photonics, pp. 139-150.
- [18] A. M. Trifas, "Medical image enhancement," University of Bucharest, 2005.
- [19] P. Janani, J. Premaladha, and K. Ravichandran, "Image enhancement techniques: A study," *Indian Journal of Science and Technology*, vol. 8, no. 22, 2015.

- [20] L. Ross and J. C. Russ, "The Image Processing Handbook," *Microscopy and Microanalysis*, vol. 17, no. 5, p. 843, 2011.
- [21] S. Bedi and R. Khandelwal, "Various image enhancement techniques-a critical review," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 2, no. 3, 2013.
- [22] S. Perschbacher, "Interpretation of panoramic radiographs," *Australian dental journal*, vol. 57, pp. 40-45, 2012.
- [23] T. S. de Oliveira Capote, M. de Almeida Gonçalves, A. Gonçalves, and M. Gonçalves, "Panoramic radiography—diagnosis of relevant structures that might compromise oral and general health of the patient," in *Emerging Trends in Oral Health Sciences and Dentistry*: IntechOpen, 2015.
- [24] C. Misch, "Dental Implant Prosthetics. China," ed: Elsevier, 2015.
- [25] O. Nomir and M. Abdel-Mottaleb, "Dental X-Ray radiographs enhancement by combining morphological operator and retinex theory," in *6th IASTED International Conference on Advances in Computer Science and Engineering, ACSE 2010*, 2010, pp. 202-204.
- [26] K. R. Jain and N. Chauhan, "Enhancement and Segmentation of Dental Radiographs Using Morphological Operations," in *Dental Image Analysis for Disease Diagnosis*: Springer, 2019, pp. 39-58.
- [27] H. Vidyasaraswathi and M. Hanumantharaju, "Review of various histogram based medical image enhancement techniques," in *Proceedings of the 2015 International Conference on Advanced Research in Computer Science Engineering & Technology (ICARCSET 2015)*, 2015, pp. 1-6.
- [28] S.-D. Chen and A. R. Ramli, "Minimum mean brightness error bi-histogram equalization in contrast enhancement," *IEEE transactions on Consumer Electronics*, vol. 49, no. 4, pp. 1310-1319, 2003.
- [29] Y. Wang, Q. Chen, and B. Zhang, "Image enhancement based on equal area dualistic sub-image histogram equalization method," *IEEE transactions on Consumer Electronics*, vol. 45, no. 1, pp. 68-75, 1999.
- [30] Y. T. Chang, J. T. Wang, W. H. Yang, and X. W. Chen, "Contrast enhancement in palm bone image using quad-histogram equalization," in *2014 International Symposium on Computer, Consumer and Control*, 2014: IEEE, pp. 1091-1094.
- [31] M. F. Khan, E. Khan, and Z. Abbasi, "Segment dependent dynamic multi-histogram equalization for image contrast enhancement," *Digital Signal Processing*, vol. 25, pp. 198-223, 2014.
- [32] H. Rahmi-Fajrin, S. Puspita, S. Riyadi, and E. Sofiani, "Dental radiography image enhancement for treatment evaluation through digital image processing," *Journal of clinical and experimental dentistry*, vol. 10, no. 7, p. e629, 2018.
- [33] V. M. Georgieva, A. D. Mihaylova, and P. P. Petrov, "An application of dental X-ray image enhancement," in *2017 13th International Conference on Advanced Technologies, Systems and Services in Telecommunications (TELSIKS)*, 2017: IEEE, pp. 447-450.
- [34] S. Simu and M. Naik, "Dental X-ray Image Enhancement using Evolutionary Algorithms," in *2020 International Conference on Computer Communication and Informatics (ICCCI)*, 2020: IEEE, pp. 1-5.
- [35] Q. Chen *et al.*, "Low-dose dental CT image enhancement using a multiscale feature sensing network," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 981, p. 164530, 2020.

- [36] E. Juliastuti and L. Epsilawati, "Image contrast enhancement for film-based dental panoramic radiography," in *2012 International Conference on System Engineering and Technology (ICSET)*, 2012: IEEE, pp. 1-5.
- [37] T.-y. YangDai and L. Zhang, "Weighted Retinex algorithm based on histogram for dental CT image enhancement," in *2014 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, 2014: IEEE, pp. 1-4.
- [38] K. Hasikin and N. A. M. Isa, "Enhancement of the low contrast image using fuzzy set theory," in *Computer Modelling and Simulation (UKSim), 2012 UKSim 14th International Conference on*, 2012: IEEE, pp. 371-376.
- [39] U. Kumbhar, V. Patil, and S. Rudrakshi, "Enhancement Of Medical Images Using Image Processing In Matlab," in *International Journal of Engineering Research and Technology*, 2013, vol. 2, no. 4 (April-2013): ESRSA Publications.
- [40] K. He, J. Sun, and X. Tang, "Guided image filtering," in *Computer Vision—ECCV 2010*: Springer, 2010, pp. 1-14.
- [41] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *Image Processing, IEEE Transactions on*, vol. 22, no. 7, pp. 2864-2875, 2013.
- [42] A. Hosni, M. Bleyer, C. Rhemann, M. Gelautz, and C. Rother, "Real-time local stereo matching using guided image filtering," in *2011 IEEE International Conference on Multimedia and Expo*, 2011: IEEE, pp. 1-6.
- [43] A. Koschan and M. Abidi, "Detection and classification of edges in color images," *IEEE Signal Processing Magazine*, vol. 22, no. 1, pp. 64-73, 2005.
- [44] P. L. Rosin, "Robust pose estimation," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 29, no. 2, pp. 297-303, 1999.
- [45] D. Nguyen, D. Halupka, P. Aarabi, and A. Sheikholeslami, "Real-time face detection and lip feature extraction using field-programmable gate arrays," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 36, no. 4, pp. 902-912, 2006.
- [46] T. Mertens, J. Kautz, and F. Van Reeth, "Exposure fusion: A simple and practical alternative to high dynamic range photography," in *Computer Graphics Forum*, 2009, vol. 28, no. 1: Wiley Online Library, pp. 161-171.
- [47] B. Yang, Z.-l. Jing, and H.-t. Zhao, "Review of pixel-level image fusion," *Journal of Shanghai Jiaotong University (Science)*, vol. 15, pp. 6-12, 2010.
- [48] S. M. Pizer, R. E. Johnston, J. P. Ericksen, B. C. Yankaskas, and K. E. Muller, "Contrast-limited adaptive histogram equalization: speed and effectiveness," in *Visualization in Biomedical Computing, 1990., Proceedings of the First Conference on*, 1990: IEEE, pp. 337-345.
- [49] D. Schraml, "Physically based synthetic image generation for machine learning: a review of pertinent literature," in *Photonics and Education in Measurement Science 2019*, 2019, vol. 11144: International Society for Optics and Photonics, p. 111440J.
- [50] P. Zdziebko and K. Holak, "Synthetic Image Generation Using the Finite Element Method and Blender Graphics Program for Modeling of Vision-Based Measurement Systems," *Sensors*, vol. 21, no. 18, p. 6046, 2021.
- [51] G. Silvano *et al.*, "Synthetic image generation for training deep learning-based automated license plate recognition systems on the Brazilian Mercosur standard," *Design Automation for Embedded Systems*, vol. 25, no. 2, pp. 113-133, 2021.
- [52] M. Turan, "Effectiveness of synthetic data generation for capsule endoscopy images," *Medicine*, vol. 10, no. 3, pp. 1015-9, 2021.

- [53] H. Achicanoy, D. Chaves, and M. Trujillo, "StyleGANs and Transfer Learning for Generating Synthetic Images in Industrial Applications," *Symmetry*, vol. 13, no. 8, p. 1497, 2021.
- [54] S. Eppel, "Setting an attention region for convolutional neural networks using region selective features, for recognition of materials within glass vessels," *arXiv preprint arXiv:1708.08711*, 2017.
- [55] D. Mehta, M. Siddiqui, and A. Javaid, "Facial emotion recognition: A survey and real-world user experiences in mixed reality," *Sensors*, vol. 18, no. 2, p. 416, 2018.
- [56] D. A. Pollina *et al.*, "Facial skin surface temperature changes during a "concealed information" test," *Annals of biomedical Engineering*, vol. 34, no. 7, pp. 1182-1189, 2006.
- [57] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in *2016 IEEE winter conference on applications of computer vision (WACV)*, 2016: IEEE, pp. 1-10.
- [58] J. Stemberger, R. S. Allison, and T. Schnell, "Thermal imaging as a way to classify cognitive workload," in *2010 Canadian Conference on Computer and Robot Vision*, 2010: IEEE, pp. 231-238.
- [59] H. Gao, A. Yüce, and J.-P. Thiran, "Detecting emotional stress from facial expressions for driving safety," in *2014 IEEE International Conference on Image Processing (ICIP)*, 2014: IEEE, pp. 5961-5965.
- [60] I. Pavlidis, J. Levine, and P. Baukol, "Thermal image analysis for anxiety detection," in *Proceedings 2001 International Conference on Image Processing (Cat. No. 01CH37205)*, 2001, vol. 2: IEEE, pp. 315-318.
- [61] F. J. Prokoski, R. B. Riedel, and J. S. Coffin, "Identification of individuals by means of facial thermography," in *Proceedings 1992 International Carnahan Conference on Security Technology: Crime Countermeasures*, 1992: IEEE, pp. 120-125.
- [62] F. J. Prokoski, J. S. Coffin, and R. B. Riedel, "Method for identifying individuals from analysis of elemental shapes derived from biosensor data," ed: Google Patents, 1992.
- [63] A. Seal, D. Bhattacharjee, M. Nasipuri, and D. K. Basu, "Minutiae based thermal face recognition using blood perfusion data," in *2011 International Conference on Image Information Processing*, 2011: IEEE, pp. 1-4.
- [64] M. K. Bhowmik *et al.*, "Thermal infrared face recognition—a biometric identification technique for robust security system," in *Reviews, refinements and new ideas in face recognition*: IntechOpen, 2011.
- [65] T. Suzuki, T. Tsuruda, K. Yamaguchi, Y. Ino, M. Honjo, and D. Miura, "Experiments on Using Thermal Imaging Camera for Fire Fighting Activity," *Fire Safety Science*, vol. 7, pp. 114-114, 2007.
- [66] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, 2005, vol. 1: IEEE, pp. 886-893.
- [67] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and vision Computing*, vol. 27, no. 6, pp. 803-816, 2009.
- [68] M. Yeasin, B. Bulot, and R. Sharma, "Recognition of facial expressions and measurement of levels of interest from video," *IEEE Transactions on Multimedia*, vol. 8, no. 3, pp. 500-508, 2006.

- [69] Y. Zhu, L. C. De Silva, and C. C. Ko, "Using moment invariants and HMM in facial expression recognition," *Pattern Recognition Letters*, vol. 23, no. 1-3, pp. 83-91, 2002.
- [70] M. M. Khan, R. D. Ward, and M. Ingleby, "Classifying pretended and evoked facial expressions of positive and negative affective states using infrared measurement of skin temperature," *ACM Transactions on Applied Perception (TAP)*, vol. 6, no. 1, p. 6, 2009.
- [71] A. Nozawa and M. Tacano, "Correlation analysis on alpha attenuation and nasal skin temperature," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2009, no. 01, p. P01007, 2009.
- [72] R. Nakanishi and K. Imai-Matsumura, "Facial skin temperature decreases in infants with joyful expression," *Infant Behavior and Development*, vol. 31, no. 1, pp. 137-144, 2008.
- [73] B. R. Nhan and T. Chau, "Classifying affective states using thermal infrared imaging of the human face," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 4, pp. 979-987, 2010.
- [74] Y. Yoshitomi, "Facial expression recognition for speaker using thermal image processing and speech recognition system," in *Proceedings of the 10th WSEAS international conference on Applied computer science*, 2010, pp. 182-186.
- [75] M. M. Khan, R. D. Ward, and M. Ingleby, "Infrared thermal sensing of positive and negative affective states," in *Robotics, Automation and Mechatronics, 2006 IEEE Conference on*, 2006: IEEE, pp. 1-6.
- [76] S. Jarlier *et al.*, "Thermal analysis of facial muscles contractions," *IEEE Trans. Affective Computing*, vol. 2, no. 1, pp. 2-9, 2011.
- [77] L. Trujillo, G. Olague, R. Hammoud, and B. Hernandez, "Automatic feature localization in thermal images for facial expression recognition," in *Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, 2005: IEEE, pp. 14-14.
- [78] P. Shen, S. Wang, and Z. Liu, "Facial expression recognition from infrared thermal videos," in *Intelligent Autonomous Systems 12*: Springer, 2013, pp. 323-333.
- [79] A. Sharif Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: an astounding baseline for recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2014, pp. 806-813.
- [80] Y. Fan, J. C. Lam, and V. O. Li, "Multi-region ensemble convolutional neural network for facial expression recognition," in *International Conference on Artificial Neural Networks*, 2018: Springer, pp. 84-94.
- [81] M. F. Valstar, B. Jiang, M. Mehu, M. Pantic, and K. Scherer, "The first facial expression recognition and analysis challenge," in *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, 2011: IEEE, pp. 921-926.
- [82] R. Shojia Ghiass, "Face recognition using infrared vision," Université Laval Quebec, QC, Canada, 2014.
- [83] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *BMVC*, 2015, vol. 1, no. 3, p. 6.
- [84] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097-1105.
- [85] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431-3440.

- [86] M. Roopaei, S. Agaian, M. Shadaram, and F. Hurtado, "Cross-entropy histogram equalization," in *2014 IEEE international conference on systems, man, and cybernetics (SMC)*, 2014: IEEE, pp. 158-163.
- [87] S. K. KM, S. Rajeev, K. Panetta, and S. S. Agaian, "Comparative study of palm print authentication system using geometric features," in *Mobile Multimedia/Image Processing, Security, and Applications 2017*, 2017, vol. 10221: International Society for Optics and Photonics, p. 102210M.
- [88] R. Rajendran, S. P. Rao, S. S. Agaian, and K. Panetta, "A versatile edge preserving image enhancement approach for medical images using guided filter," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2016: IEEE, pp. 002341-002346.
- [89] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [90] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 2010, pp. 249-256.
- [91] F. Liu, D. Zhang, C. Song, and G. Lu, "Touchless multiview fingerprint acquisition and mosaicking," *IEEE Transactions on Instrumentation and Measurement*, vol. 62, no. 9, pp. 2492-2502, 2013.
- [92] S. S. Agaian, M. M. A. Mulawka, R. Rajendran, S. P. Rao, S. K. KM, and S. Rajeev, "A comparative study of image feature detection and matching algorithms for touchless fingerprint systems," *Electronic Imaging*, vol. 2016, no. 15, pp. 1-9, 2016.
- [93] A. Oliva and A. Torralba, "Building the gist of a scene: The role of global image features in recognition," *Progress in brain research*, vol. 155, pp. 23-36, 2006.
- [94] D. Lisin, M. Mattar, M. B. Blaschko, E. G. Learned-Miller, and M. C. Benfield, "Combining local and global image features for object class recognition," in *Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, 2005: IEEE, pp. 47-47.
- [95] G. Fahmy *et al.*, "A web based tool for an automated dental identification system (ADIS)," in *Proceedings of the 2004 annual national conference on Digital government research*, 2004: Digital Government Society of North America, p. 138.
- [96] M. J. Thali, T. Markwalder, C. Jackowski, M. Sonnenschein, and R. Dirnhofer, "Dental CT imaging as a screening tool for dental profiling: advantages and limitations," *Journal of forensic sciences*, vol. 51, no. 1, pp. 113-119, 2006.
- [97] R. Kumar, A. Athota, T. Rastogi, and S. K. Karumuri, "Forensic radiology: An emerging tool in identification," *Journal of Indian Academy of Oral Medicine and Radiology*, vol. 27, no. 3, p. 416, 2015.
- [98] E. De Valck, "Major incident response: collecting ante-mortem data," *Forensic science international*, vol. 159, pp. S15-S19, 2006.
- [99] P. Pittayapat, R. Jacobs, E. De Valck, D. Vandermeulen, and G. Willems, "Forensic odontology in the disaster victim identification process," *The Journal of forensic odontology*, vol. 30, no. 1, 2012.
- [100] A. K. Verma, S. Kumar, S. Rathore, and A. Pandey, "Role of dental expert in forensic odontology," *National journal of maxillofacial surgery*, vol. 5, no. 1, p. 2, 2014.
- [101] S. H. Al Amad, "Forensic odontology," *Smile Dental Journal*, pp. 22-24, 2009.

- [102] B. Kolude, B. Adeyemi, J. Taiwo, O. Sigbeku, and U. Eze, "The role of forensic dentist following mass disaster," *Annals of Ibadan postgraduate medicine*, vol. 8, no. 2, pp. 111-117, 2010.
- [103] A. K. Jain and H. Chen, "Matching of dental X-ray images for human identification," *Pattern recognition*, vol. 37, no. 7, pp. 1519-1532, 2004.
- [104] J. Fitzpatrick *et al.*, "Optical and digital techniques for enhancing radiographic anatomy for identification of human remains," *Journal of Forensic Science*, vol. 41, no. 6, pp. 947-959, 1996.
- [105] D. L. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes, "Medical image registration," *Physics in medicine and biology*, vol. 46, no. 3, p. R1, 2001.
- [106] S. J. Dwyer *et al.*, "Medical image processing in diagnostic radiology," *IEEE Transactions on Nuclear Science*, vol. 27, no. 3, pp. 1047-1055, 1980.
- [107] W. E. L. Grimson, G. Ettinger, T. Kapur, M. E. Leventon, W. M. Wells III, and R. Kikinis, "Utilizing segmented MRI data in image-guided surgery," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 11, no. 08, pp. 1367-1397, 1997.
- [108] A. J. Worth, N. Makris, V. S. Caviness Jr, and D. N. Kennedy, "Neuroanatomical segmentation in MRI: technological objectives," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 11, no. 08, pp. 1161-1187, 1997.
- [109] G. Jader, L. Oliveira, and M. Pithon, "Automatic segmenting teeth in X-ray images: Trends, a novel data set, benchmarking and future perspectives," *arXiv preprint arXiv:1802.03086*, 2018.
- [110] M. Alazab, M. Islam, and S. Venkatraman, "Towards Automatic Image Segmentation Using Optimised Region Growing Technique," Berlin, Heidelberg, 2009: Springer Berlin Heidelberg, in *AI 2009: Advances in Artificial Intelligence*, pp. 131-139.
- [111] S. Oprea, C. Marinescu, I. Lita, M. Jurianu, D. A. Visan, and I. B. Cioc, "Image processing techniques used for dental x-ray image analysis," in *Electronics Technology, 2008. ISSE'08. 31st International Spring Seminar on*, 2008: IEEE, pp. 125-129.
- [112] R. Krithiga, C. Lakshmi, and A. Nithya, *Segmentation of dental caries from dental X-ray images using wavelet and watershed transforms*. 2014, pp. 4617-4624.
- [113] S. Li, T. Fevens, A. Krzyżak, and S. Li, "An automatic variational level set segmentation framework for computer aided dental X-rays analysis in clinical environments," *Computerized Medical Imaging and Graphics*, vol. 30, no. 2, pp. 65-74, 2006.
- [114] H. Soomer, H. Ranta, M. J. Lincoln, A. Penttila, and E. Leibur, "Reliability and validity of eight dental age estimation methods for adults," *Journal of Forensic Science*, vol. 48, no. 1, pp. 1-4, 2003.
- [115] S. I. Kvaal, K. M. Kolltveit, I. O. Thomsen, and T. Solheim, "Age estimation of adults from dental radiographs," *Forensic science international*, vol. 74, no. 3, pp. 175-185, 1995.
- [116] R. Cameriere, L. Ferrante, and M. Cingolani, "Variations in pulp/tooth area ratio as an indicator of age: a preliminary study," *Journal of Forensic Science*, vol. 49, no. 2, pp. 1-3, 2004.
- [117] F. M. Vandevoort *et al.*, "Age calculation using X-ray microfocus computed tomographical scanning of teeth: a pilot study," *Journal of Forensic Science*, vol. 49, no. 4, pp. 1-4, 2004.
- [118] F. Yang, R. Jacobs, and G. Willems, "Dental age estimation through volume matching of teeth imaged by cone-beam CT," *Forensic science international*, vol. 159, pp. S78-S83, 2006.

- [119] H. Someda *et al.*, "Age estimation based on three-dimensional measurement of mandibular central incisors in Japanese," *Forensic science international*, vol. 185, no. 1-3, pp. 110-114, 2009.
- [120] H. Aboshi, T. Takahashi, and T. Komuro, "Age estimation using microfocus X-ray computed tomography of lower premolars," *Forensic science international*, vol. 200, no. 1-3, pp. 35-40, 2010.
- [121] O. Nomir and M. Abdel-Mottaleb, "Hierarchical contour matching for dental X-ray radiographs," *Pattern Recognition*, vol. 41, no. 1, pp. 130-138, 2008.
- [122] J. Zhou and M. Abdel-Mottaleb, "A content-based system for human identification based on bitewing dental X-ray images," *Pattern Recognition*, vol. 38, no. 11, pp. 2132-2142, 2005.
- [123] M. Hosntalab, R. A. Zoroofi, A. A. Tehrani-Fard, and G. Shirani, "Classification and numbering of teeth in multi-slice CT images using wavelet-Fourier descriptor," *International journal of computer assisted radiology and surgery*, vol. 5, no. 3, pp. 237-249, 2010.
- [124] Y. Miki *et al.*, "Classification of teeth in cone-beam CT using deep convolutional neural network," *Computers in biology and medicine*, vol. 80, pp. 24-29, 2017.
- [125] N. M. Noor, N. E. A. Khalid, M. H. Ali, and A. D. A. Numpang, "Fish bone impaction using adaptive histogram equalization (ahe)," in *2010 Second International Conference on Computer Research and Development*, 2010: IEEE, pp. 163-167.
- [126] N. E. A. Khalid, M. Manaf, M. E. Aziz, and M. H. Ali, "CR images of metacarpel cortical edge detection-bone profile histogram approximation method," in *2007 International Conference on Intelligent and Advanced Systems*, 2007: IEEE, pp. 702-708.
- [127] M. Sakata and K. Ogawa, "Noise reduction and contrast enhancement for small-dose X-ray images in wavelet domain," in *2009 IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*, 2009: IEEE, pp. 2924-2929.
- [128] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, 2015: Springer, pp. 234-241.
- [129] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961-2969.
- [130] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.
- [131] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91-110, 2004.
- [132] E. Mingkhwan and W. Khawsuk, "Digital image stabilization technique for fixed camera on small size drone," in *2017 Third Asian Conference on Defence Technology (ACDT)*, 2017: IEEE, pp. 12-19.
- [133] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *International journal of computer vision*, vol. 60, no. 1, pp. 63-86, 2004.
- [134] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *International conference on medical image computing and computer-assisted intervention*, 1998: Springer, pp. 130-137.



- [135] A. Vignati *et al.*, "A fully automatic multiscale 3-dimensional Hessian-based algorithm for vessel detection in breast DCE-MRI," *Investigative radiology*, vol. 47, no. 12, pp. 705-710, 2012.
- [136] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the seventh IEEE international conference on computer vision*, 1999, vol. 2: Ieee, pp. 1150-1157.
- [137] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 10, pp. 1615-1630, 2005.
- [138] C. A. Poynton, "Smppte tutorial: "gamma" and its disguises: The nonlinear mappings of intensity in perception, crts, film, and video," *SMPTE journal*, vol. 102, no. 12, pp. 1099-1108, 1993.
- [139] M. A. Hogervorst and A. Toet, "Fast natural color mapping for night-time imagery," *Information Fusion*, vol. 11, no. 2, pp. 69-77, 2010.
- [140] Z. Zhou, M. Dong, X. Xie, and Z. Gao, "Fusion of infrared and visible images for night-vision context enhancement," *Applied optics*, vol. 55, no. 23, pp. 6480-6490, 2016.
- [141] E. H. Land, "The retinex theory of color vision," *Scientific American*, vol. 237, no. 6, pp. 108-129, 1977.
- [142] D. J. Jobson, Z.-u. Rahman, and G. A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Transactions on Image processing*, vol. 6, no. 7, pp. 965-976, 1997.
- [143] X. Dong *et al.*, "Fast efficient algorithm for enhancement of low lighting video," in *2011 IEEE International Conference on Multimedia and Expo*, 2011: IEEE, pp. 1-6.
- [144] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, "A weighted variational model for simultaneous reflectance and illumination estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2782-2790.
- [145] X. Guo, Y. Li, and H. Ling, "LIME: Low-Light Image Enhancement via Illumination Map Estimation," *IEEE Trans. Image Processing*, vol. 26, no. 2, pp. 982-993, 2017.
- [146] N. Papadakis, E. Provenzi, and V. Caselles, "A variational model for histogram transfer of color images," *IEEE Transactions on Image Processing*, vol. 20, no. 6, pp. 1682-1695, 2011.
- [147] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images," *IEEE Computer graphics and applications*, vol. 21, no. 5, pp. 34-41, 2001.
- [148] L. Wang, S. Shi, W. Jin, and Y. Zhao, "Color fusion algorithm for visible and infrared images based on color transfer in YUV color space," in *MIPPR 2007: Multispectral Image Processing*, 2007, vol. 6787: International Society for Optics and Photonics, p. 67870S.
- [149] Y. Zheng and E. A. Essock, "A local-coloring method for night-vision colorization utilizing image analysis and fusion," *Information Fusion*, vol. 9, no. 2, pp. 186-199, 2008.
- [150] M. A. Hogervorst and A. Toet, "Method for applying daytime colors to nighttime imagery in realtime," in *Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2008*, 2008, vol. 6974: International Society for Optics and Photonics, p. 697403.
- [151] T. Pouli and E. Reinhard, "Progressive color transfer for images of arbitrary dynamic range," *Computers & Graphics*, vol. 35, no. 1, pp. 67-80, 2011.

- [152] X. Xiao and L. Ma, "Color transfer in correlated color space," in *Proceedings of the 2006 ACM international conference on Virtual reality continuum and its applications*, 2006: ACM, pp. 305-309.
- [153] F. Pitié, A. C. Kokaram, and R. Dahyot, "Automated colour grading using colour distribution transfer," *Computer Vision and Image Understanding*, vol. 107, no. 1-2, pp. 123-137, 2007.
- [154] U. Fecker, M. Barkowsky, and A. Kaup, "Histogram-based prefiltering for luminance and chrominance compensation of multiview video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 9, pp. 1258-1267, 2008.
- [155] M. Nikolova and G. Steidl, "Fast ordering algorithm for exact histogram specification," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5274-5283, 2014.
- [156] L. A. Gatys, A. S. Ecker, M. Bethge, A. Hertzmann, and E. Shechtman, "Controlling perceptual factors in neural style transfer," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 3985-3993.
- [157] F. Luan, S. Paris, E. Shechtman, and K. Bala, "Deep photo style transfer," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4990-4998.
- [158] F. P. Oliveira and J. M. R. Tavares, "Medical image registration: a review," *Computer methods in biomechanics and biomedical engineering*, vol. 17, no. 2, pp. 73-93, 2014.
- [159] T. Can, A. O. Karalı, and T. Aytac, "Detection and tracking of sea-surface targets in infrared and visual band videos using the bag-of-features technique with scale-invariant feature transform," *Applied optics*, vol. 50, no. 33, pp. 6302-6312, 2011.
- [160] R. Wang, Y. Xia, G. Wang, and J. Tian, "License plate localization in complex scenes based on oriented FAST and rotated BRIEF feature," *Journal of Electronic Imaging*, vol. 24, no. 5, p. 053011, 2015.
- [161] R. Rajendran, S. P. Rao, K. Panetta, and S. S. Agaian, "Adaptive Alpha-Trimmed Correlation based underwater image stitching," in *2017 IEEE International Symposium on Technologies for Homeland Security (HST)*, 2017: IEEE, pp. 1-7.
- [162] S. P. Rao, R. Rajendran, S. S. Agaian, and M. M. A. Mulawka, "Alpha trimmed correlation for touchless finger image mosaicing," in *Mobile Multimedia/Image Processing, Security, and Applications 2016*, 2016, vol. 9869: International Society for Optics and Photonics, p. 98690U.
- [163] J. Zhu, L. Wang, R. Yang, and J. E. Davis, "Reliability fusion of time-of-flight depth and stereo geometry for high quality depth maps," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 7, pp. 1400-1414, 2011.
- [164] J. Chen, L. Luo, C. Liu, J.-G. Yu, and J. Ma, "Nonrigid registration of remote sensing images via sparse and dense feature matching," *JOSA A*, vol. 33, no. 7, pp. 1313-1322, 2016.
- [165] L. Juan and G. Oubong, "SURF applied in panorama image stitching," in *2010 2nd international conference on image processing theory, tools and applications*, 2010: IEEE, pp. 495-499.
- [166] M. Brown and S. Süssstrunk, "Multi-spectral SIFT for scene category recognition," in *CVPR 2011*, 2011: IEEE, pp. 177-184.
- [167] N. Jacobs, N. Roman, and R. Pless, "Consistent temporal variations in many outdoor scenes," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, 2007: IEEE, pp. 1-6.

- [168] Y. Xu, Z. Li, J. Yang, and D. Zhang, "A survey of dictionary learning algorithms for face recognition," *IEEE access*, vol. 5, pp. 8502-8514, 2017.
- [169] R. Garcia, J. Batlle, X. Cufi, and J. Amat, "Positioning an underwater vehicle through image mosaicking," in *Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference on*, 2001, vol. 3: IEEE, pp. 2779-2784.
- [170] D. Whang, N. Xu, S. Rangwala, K. Chintalapudi, R. Govindan, and J. W. Wallace, "Development of an embedded networked sensing system for structural health monitoring," *Center for Embedded Network Sensing*, 2004.
- [171] N. Ramanathan, M. D. Yarvis, J. Chhabra, N. Kushalnagar, L. Krishnamurthy, and D. Estrin, "A stream-oriented power management protocol for low duty cycle sensor network applications," *Center for Embedded Network Sensing*, 2005.
- [172] R. Prados, R. Garcia, and L. Neumann, *Image blending techniques and their application in underwater mosaicing*. Springer, 2014.
- [173] K. Panetta, C. Gao, and S. Agaian, "Human-visual-system-inspired underwater image quality measures," 2015.
- [174] E. Adel, M. Elmogy, and H. Elbakry, "Image stitching based on feature extraction techniques: a survey," *International Journal of Computer Applications (0975-8887) Volume*, 2014.
- [175] R. Prados, R. Garcia, N. Gracias, J. Escartin, and L. Neumann, "A novel blending technique for underwater gigamosaicing," *IEEE Journal of Oceanic Engineering*, vol. 37, no. 4, pp. 626-644, 2012.
- [176] R. Schettini and S. Corchs, "Underwater image processing: state of the art of restoration and image enhancement methods," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, no. 1, p. 746052, 2010.
- [177] C. Ancuti, C. O. Ancuti, T. Haber, and P. Bekaert, "Enhancing underwater images and videos by fusion," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012: IEEE, pp. 81-88.
- [178] Y. Song, C. Lee, and J. Kim, "A new scheme for touchless fingerprint recognition system," in *Proceedings of 2004 International Symposium on Intelligent Signal Processing and Communication Systems, 2004. ISPACS 2004.*, 2004: IEEE, pp. 524-527.
- [179] H. Li, J. Luo, C. Huang, Y. Yang, and S. Xie, "An adaptive image-stitching algorithm for an underwater monitoring system," *International Journal of Advanced Robotic Systems*, vol. 11, 2014.
- [180] M. Chen, R. Nian, B. He, S. Qiu, X. Liu, and T. Yan, "Underwater image stitching based on SIFT and wavelet fusion," in *OCEANS 2015-Genova*, 2015: IEEE, pp. 1-4.
- [181] S. Hojjati, "Underwater Image Stitching," University of Alberta, 2016.
- [182] K. Iqbal, M. Odetayo, A. James, R. A. Salam, and A. Z. H. Talib, "Enhancing the low quality images using unsupervised colour correction method," in *Systems Man and Cybernetics (SMC), 2010 IEEE international conference on*, 2010: IEEE, pp. 1703-1709.
- [183] K. Panetta, A. Samani, and S. Agaian, "Choosing the optimal spatial domain measure of enhancement for mammogram images," *Journal of Biomedical Imaging*, vol. 2014, p. 3, 2014.
- [184] S. S. Agaian, K. Panetta, and A. M. Grigoryan, "A new measure of image enhancement," in *IASTED International Conference on Signal Processing & Communication*, 2000: Citeseer, pp. 19-22.

- [185] C. Gao, K. Panetta, and S. Agaian, "Three dimensional alpha weighted quadratic filter based image color contrast enhancement," in *SPIE Defense, Security, and Sensing*, 2013: International Society for Optics and Photonics, pp. 875514-875514-12.
- [186] K. He, J. Sun, and X. Tang, "Guided image filtering," in *European conference on computer vision*, 2010: Springer, pp. 1-14.
- [187] R. Rajendran, S. P. Rao, S. S. Agaian, and K. Panetta, "A versatile edge preserving image enhancement approach for medical images using guided filter," in *Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference on*, 2016: IEEE, pp. 002341-002346.
- [188] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *International journal of computer vision*, vol. 74, no. 1, pp. 59-73, 2007.
- [189] S. F. Davis and R. A. Smith, *Introduction to Statistics and Research Methods: Pearson New International Edition PDF eBook: Becoming a Psychological Detective, An*. Pearson Higher Ed, 2013.
- [190] M. A. Sadeghi, S. M. M. Hejrati, and N. Gheissari, "Poisson Local Color Correction for Image Stitching," in *VISAPP (1)*, 2008, pp. 275-282.
- [191] M. Sheng, S. Tang, Z. Cui, W. Wu, and L. Wan, "A joint framework for underwater sequence images stitching based on deep neural network convolutional neural network," *International Journal of Advanced Robotic Systems*, vol. 17, no. 2, p. 1729881420915062, 2020.
- [192] M. Oliveira, A. D. Sappa, and V. Santos, "Unsupervised local color correction for coarsely registered images," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, 2011: IEEE, pp. 201-208.
- [193] H. S. Faridul, T. Pouli, C. Chamaret, J. Stauder, A. Trémeau, and E. Reinhard, "A Survey of Color Mapping and its Applications," in *Eurographics (State of the Art Reports)*, 2014, pp. 43-67.
- [194] L. Juan and G. Oubong, "SURF applied in panorama image stitching," in *Image Processing Theory Tools and Applications (IPTA), 2010 2nd International Conference on*, 2010: IEEE, pp. 495-499.
- [195] K. Panetta, C. Gao, and S. Agaian, "Human-visual-system-inspired underwater image quality measures," *IEEE Journal of Oceanic Engineering*, vol. 41, no. 3, pp. 541-551, 2015.
- [196] C. Ancuti, C. O. Ancuti, T. Haber, and P. Bekaert, "Enhancing underwater images and videos by fusion," in *2012 IEEE conference on computer vision and pattern recognition*, 2012: IEEE, pp. 81-88.
- [197] J. Astola and P. Kuosmanen, *Fundamentals of nonlinear digital filtering*. CRC press, 2020.
- [198] K. Panetta, S. Agaian, Y. Zhou, and E. J. Wharton, "Parameterized logarithmic framework for image enhancement," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 41, no. 2, pp. 460-473, 2010.
- [199] A. K. Jain, L. Hong, S. Pankanti, and R. Bolle, "An identity-authentication system using fingerprints," *Proceedings of the IEEE*, vol. 85, no. 9, pp. 1365-1388, 1997.
- [200] A. Jain, U. Uludag, and A. Ross, "Biometric template selection: a case study in fingerprints," in *International Conference on Audio-and Video-Based Biometric Person Authentication*, 2003: Springer, pp. 335-342.
- [201] J. Feng, "Combining minutiae descriptors for fingerprint matching," *Pattern Recognition*, vol. 41, no. 1, pp. 342-352, 2008.

- [202] H. Choi, K. Choi, and J. Kim, "Mosaicing touchless and mirror-reflected fingerprint images," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 1, pp. 52-61, 2010.
- [203] "International Agency for Research on Cancer (World Health Organization), "Kidney: Globocan 2020 - The Global Cancer Observatory"." (accessed.
- [204] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 68, no. 6, pp. 394-424, 2018.
- [205] E. Yang, C. K. Kim, Y. Guan, B.-B. Koo, and J.-H. Kim, "3D Multi-Scale Residual Fully Convolutional Neural Network for Segmentation of Extremely Large-Sized Kidney Tumor," *Computer Methods and Programs in Biomedicine*, p. 106616, 2022.
- [206] R. Rajendran *et al.*, "Initial Evaluation of Computer-Assisted Radiologic Assessment for Renal Mass Edge Detection as an Indication of Tumor Roughness to Predict Renal Cancer Subtypes," *Advances in Urology*, vol. 2019, 2019.
- [207] M. M. Nguyen, I. S. Gill, and L. M. Ellison, "The evolving presentation of renal carcinoma in the United States: trends from the Surveillance, Epidemiology, and End Results program," *The Journal of urology*, vol. 176, no. 6, pp. 2397-2400, 2006.
- [208] C. B. Lund and B. H. van der Velden, "Leveraging Clinical Characteristics for Improved Deep Learning-Based Kidney Tumor Segmentation on CT," *arXiv preprint arXiv:2109.05816*, 2021.
- [209] M. K. Santos, J. R. Ferreira, D. T. Wada, A. P. M. Tenório, M. H. N. Barbosa, and P. M. d. A. Marques, "Artificial intelligence, machine learning, computer-aided diagnosis, and radiomics: advances in imaging towards to precision medicine," *Radiologia brasileira*, vol. 52, pp. 387-396, 2019.
- [210] H. Lee and Y.-P. P. Chen, "Image based computer aided diagnosis system for cancer detection," *Expert Systems with Applications*, vol. 42, no. 12, pp. 5356-5365, 2015.
- [211] R. K. Samala, H.-P. Chan, L. M. Hadjiiski, M. A. Helvie, K. H. Cha, and C. D. Richter, "Multi-task transfer learning deep convolutional neural network: application to computer-aided diagnosis of breast cancer on mammograms," *Physics in Medicine & Biology*, vol. 62, no. 23, p. 8894, 2017.
- [212] H. Fujita, "AI-based computer-aided diagnosis (AI-CAD): the latest review to read first," *Radiological physics and technology*, vol. 13, no. 1, pp. 6-19, 2020.
- [213] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and H. J. Aerts, "Artificial intelligence in radiology," *Nature Reviews Cancer*, vol. 18, no. 8, pp. 500-510, 2018.
- [214] E. C. Lin, "Radiation risk from medical imaging," in *Mayo Clinic Proceedings*, 2010, vol. 85, no. 12: Elsevier, pp. 1142-1146.
- [215] S. K. KM, R. Rajendran, Q. Wan, K. Panetta, and S. S. Agaian, "TERNet: A deep learning approach for thermal face emotion recognition," in *Mobile Multimedia/Image Processing, Security, and Applications 2019*, 2019, vol. 10993: International Society for Optics and Photonics, p. 1099309.
- [216] K. Panetta, K. S. Kamath, S. Rajeev, and S. S. Agaian, "FTNet: Feature Transverse Network for Thermal Image Semantic Segmentation," *IEEE Access*, vol. 9, pp. 145212-145227, 2021.
- [217] M. Havaei *et al.*, "Brain tumor segmentation with deep neural networks," *Medical image analysis*, vol. 35, pp. 18-31, 2017.

- [218] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1240-1251, 2016.
- [219] B. H. Menze *et al.*, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE transactions on medical imaging*, vol. 34, no. 10, pp. 1993-2024, 2014.
- [220] A. Kalinovsky and V. Kovalev, "Lung image Ssgmentation using deep learning methods and convolutional neural networks," 2016.
- [221] M. Fu *et al.*, "Hierarchical combinatorial deep learning architecture for pancreas segmentation of medical computed tomography cancer images," *BMC systems biology*, vol. 12, no. 4, pp. 119-127, 2018.
- [222] H. R. Roth *et al.*, "Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation," in *International conference on medical image computing and computer-assisted intervention*, 2015: Springer, pp. 556-564.
- [223] L. Yu, X. Yang, H. Chen, J. Qin, and P. A. Heng, "Volumetric ConvNets with mixed residual connections for automated prostate segmentation from 3D MR images," in *Thirty-first AAAI conference on artificial intelligence*, 2017.
- [224] W. Zhang *et al.*, "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation," *NeuroImage*, vol. 108, pp. 214-224, 2015.
- [225] A. Prason, K. Petersen, C. Igel, F. Lauze, E. Dam, and M. Nielsen, "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network," in *International conference on medical image computing and computer-assisted intervention*, 2013: Springer, pp. 246-253.
- [226] P. Moeskops *et al.*, "Deep learning for multi-task medical image segmentation in multiple modalities," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2016: Springer, pp. 478-486.
- [227] K. Kamnitsas, L. Chen, C. Ledig, D. Rueckert, and B. Glocker, "Multi-scale 3D convolutional neural networks for lesion segmentation in brain MRI," *Ischemic stroke lesion segmentation*, vol. 13, p. 46, 2015.
- [228] J. Kleesiek *et al.*, "Deep MRI brain extraction: A 3D convolutional neural network for skull stripping," *NeuroImage*, vol. 129, pp. 460-469, 2016.
- [229] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation," in *International conference on medical image computing and computer-assisted intervention*, 2016: Springer, pp. 424-432.
- [230] G. Zeng, X. Yang, J. Li, L. Yu, P.-A. Heng, and G. Zheng, "3D U-net with multi-level deep supervision: fully automatic segmentation of proximal femur in 3D MR images," in *International workshop on machine learning in medical imaging*, 2017: Springer, pp. 274-282.
- [231] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*, 2016: IEEE, pp. 565-571.
- [232] L. Yu, H. Chen, Q. Dou, J. Qin, and P.-A. Heng, "Automated melanoma recognition in dermoscopy images via very deep residual networks," *IEEE transactions on medical imaging*, vol. 36, no. 4, pp. 994-1004, 2016.
- [233] Q. Dou, H. Chen, Y. Jin, L. Yu, J. Qin, and P.-A. Heng, "3D deeply supervised network for automatic liver segmentation from CT volumes," in *International conference on*

- medical image computing and computer-assisted intervention*, 2016: Springer, pp. 149-157.
- [234] X. Feng, J. Yang, A. F. Laine, and E. D. Angelini, "Discriminative localization in CNNs for weakly-supervised segmentation of pulmonary nodules," in *International conference on medical image computing and computer-assisted intervention*, 2017: Springer, pp. 568-576.
  - [235] A. Skalski, J. Jakubowski, and T. Drewniak, "Kidney tumor segmentation and detection on computed tomography data," in *2016 IEEE International Conference on Imaging Systems and Techniques (IST)*, 2016: IEEE, pp. 238-242.
  - [236] Q. Yu, Y. Shi, J. Sun, Y. Gao, J. Zhu, and Y. Dai, "Crossbar-net: A novel convolutional neural network for kidney tumor segmentation in ct images," *IEEE transactions on image processing*, vol. 28, no. 8, pp. 4060-4074, 2019.
  - [237] X. Hou *et al.*, "A triple-stage self-guided network for kidney tumor segmentation," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 2020: IEEE, pp. 341-344.
  - [238] L. B. da Cruz *et al.*, "Kidney tumor segmentation from computed tomography images using DeepLabv3+ 2.5 D model," *Expert Systems with Applications*, p. 116270, 2021.
  - [239] W. Zhao, D. Jiang, J. P. Queraltá, and T. Westerlund, "Multi-Scale Supervised 3D U-Net for Kidneys and Kidney Tumor Segmentation," *arXiv preprint arXiv:2004.08108*, 2020.
  - [240] Y. Lv and J. Wang, "Three Uses of One Neural Network: Automatic Segmentation of Kidney Tumor and Cysts Based on 3D U-Net," 2021.
  - [241] G. Santini, N. Moreau, and M. Rubeaux, "Kidney tumor segmentation using an ensembling multi-stage deep learning approach. A contribution to the KiTS19 challenge," *arXiv preprint arXiv:1909.00735*, 2019.
  - [242] D. Li, Z. Chen, H. Hassan, and B. Huang, "A Cascaded 3D Segmentation Model for Renal Enhanced CT Images," 2021.
  - [243] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.
  - [244] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.
  - [245] S. Qiao, H. Wang, C. Liu, W. Shen, and A. Yuille, "Micro-batch training with batch-channel normalization and weight standardization," *arXiv preprint arXiv:1903.10520*, 2019.
  - [246] F. Yu, V. Koltun, and T. Funkhouser, "Dilated residual networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 472-480.
  - [247] K. Panetta, S. K. KM, S. P. Rao, and S. S. Agaian, "Deep Perceptual Image Enhancement Network for Exposure Restoration," *IEEE Transactions on Cybernetics*, 2022.
  - [248] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132-7141.
  - [249] F. Wang *et al.*, "Residual attention network for image classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3156-3164.
  - [250] R. Lan, L. Sun, Z. Liu, H. Lu, C. Pang, and X. Luo, "MADNet: a fast and lightweight network for single-image super resolution," *IEEE transactions on cybernetics*, vol. 51, no. 3, pp. 1443-1453, 2020.

- [251] N. Heller *et al.*, "The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes," *arXiv preprint arXiv:1904.00445*, 2019.
- [252] <https://github.com/neheller/kits19/issues/21>.
- [253] F. Isensee and K. H. Maier-Hein, "An attempt at beating the 3D U-Net," *arXiv preprint arXiv:1908.02182*, 2019.
- [254] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep learning in medical image analysis and multimodal learning for clinical decision support*: Springer, 2018, pp. 3-11.
- [255] Z. Gu *et al.*, "Ce-net: Context encoder network for 2d medical image segmentation," *IEEE transactions on medical imaging*, vol. 38, no. 10, pp. 2281-2292, 2019.
- [256] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature methods*, vol. 18, no. 2, pp. 203-211, 2021.
- [257] B. Govindarajan, K. Panetta, and S. Agaian, "Progressive edge detection on multi-bit images using polynomial-based binarization," in *Machine Learning and Cybernetics, 2008 International Conference on*, 2008, vol. 7: IEEE, pp. 3714-3719.
- [258] S. Agaian, M. Madhukar, and A. T. Chronopoulos, "Automated screening system for acute myelogenous leukemia detection in blood microscopic images," *IEEE Systems journal*, vol. 8, no. 3, pp. 995-1004, 2014.
- [259] M. Sezgin, "Survey over image thresholding techniques and quantitative performance evaluation," *Journal of Electronic imaging*, vol. 13, no. 1, pp. 146-168, 2004.
- [260] A. G.-C. Munoz, A. de Santos Sierra, C. S. Avila, J. G. Casanova, G. B. del Pozo, and V. J. Vera, "Hand biometric segmentation by means of fuzzy multiscale aggregation for mobile devices," in *Emerging Techniques and Challenges for Hand-Based Biometrics (ETCHB), 2010 International Workshop on*, 2010: IEEE, pp. 1-6.
- [261] P. Kelly, E. Cooke, N. O'Connor, and A. Smeaton, "Pedestrian detection using stereo and biometric information," in *International Conference Image Analysis and Recognition*, 2006: Springer, pp. 802-813.
- [262] S. Roy, S. Saha, A. Dey, S. H. Shaikh, and N. Chaki, "Performance evaluation of multiple image binarization algorithms using multiple metrics on standard image databases," in *ICT and Critical Infrastructure: Proceedings of the 48th Annual Convention of Computer Society of India-Vol II*, 2014: Springer, pp. 349-360.
- [263] A. Kefali, T. Sari, and H. Bahi, "Text/Background separation in the degraded document images by combining several thresholding techniques," *WSEAS transactions on signal processing*, vol. 10, pp. 436-443, 2014.
- [264] M. Muppidi, P. Rad, S. S. Agaian, and M. Jamshidi, "Image segmentation by multi-level thresholding based on fuzzy entropy and genetic algorithm in cloud," in *System of Systems Engineering Conference (SoSE), 2015 10th*, 2015: IEEE, pp. 492-497.
- [265] L. Yi and G. Zhijun, "A review of segmentation method for MR image," in *Image Analysis and Signal Processing (IASP), 2010 International Conference on*, 2010: IEEE, pp. 351-357.
- [266] A. Almuntashri, E. Finol, and S. Agaian, "Automatic lumen segmentation in CT and PC-MR images of abdominal aortic aneurysm," in *Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference on*, 2012: IEEE, pp. 2891-2896.



- [267] A. Santamaria-Pang, J. Rittscher, M. Gerdes, and D. Padfield, "Cell segmentation and classification by hierarchical supervised shape ranking," in *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*, 2015: IEEE, pp. 1296-1299.
- [268] J. M. Haggerty, X. N. Wang, A. Dickinson, C. J. O'Malley, and E. B. Martin, "Segmentation of epidermal tissue with histopathological damage in images of haematoxylin and eosin stained human skin," *BMC medical imaging*, vol. 14, no. 1, p. 7, 2014.
- [269] C. Mosquera-Lopez, S. Agaian, A. Velez-Hoyos, and I. Thompson, "Computer-aided prostate cancer diagnosis from digitized histopathology: a review on texture-based systems," *IEEE reviews in biomedical engineering*, vol. 8, pp. 98-113, 2015.
- [270] A. Almuntashri, S. Agaian, I. Thompson, D. Rabah, O. Z. Al-Abdin, and M. Nicolas, "Gleason grade-based automatic classification of prostate cancer pathological images," in *Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on*, 2011: IEEE, pp. 2696-2701.
- [271] A. A. Nawandhar, L. Yamujala, and N. Kumar, "Image segmentation using thresholding for cell nuclei detection of colon tissue," in *Advances in Computing, Communications and Informatics (ICACCI), 2015 International Conference on*, 2015: IEEE, pp. 1199-1203.
- [272] V. Roullier, O. Lézoray, V.-T. Ta, and A. Elmoataz, "Multi-resolution graph-based analysis of histopathological whole slide images: Application to mitotic cell extraction and visualization," *Computerized Medical Imaging and Graphics*, vol. 35, no. 7, pp. 603-615, 2011.
- [273] A. Basavanahally *et al.*, "Incorporating domain knowledge for tubule detection in breast histopathology using O'Callaghan neighborhoods," in *SPIE Medical Imaging*, 2011, vol. 7963, no. 1: International Society for Optics and Photonics, p. 796310.
- [274] J. Xu, A. Janowczyk, S. Chandran, and A. Madabhushi, "A high-throughput active contour scheme for segmentation of histopathological imagery," *Medical image analysis*, vol. 15, no. 6, pp. 851-862, 2011.
- [275] J. Xu, A. Janowczyk, S. Chandran, and A. Madabhushi, "A weighted mean shift, normalized cuts initialized color gradient based geodesic active contour model: applications to histopathology image segmentation," in *SPIE Medical Imaging*, 2010: International Society for Optics and Photonics, pp. 76230Y-76230Y-12.
- [276] H. Z. Al-Lahham, R. S. Alomari, H. Hiary, and V. Chaudhary, "Automating proliferation rate estimation from Ki-67 histology images," in *SPIE Medical Imaging*, 2012: International Society for Optics and Photonics, pp. 83152A-83152A-7.
- [277] C. G. Loukas, G. D. Wilson, B. Vojnovic, and A. Linney, "An image analysis-based approach for automated counting of cancer cell nuclei in tissue sections," *Cytometry part A*, vol. 55, no. 1, pp. 30-42, 2003.
- [278] T. Markiewicz, C. Jochymski, R. Koktysz, and W. Kozłowski, "Automatic cell recognition in immunohistochemical gastritis stains using sequential thresholding and SVM network," in *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, 2008: IEEE, pp. 971-974.
- [279] H. Kong, M. Gurcan, and K. Belkacem-Boussaid, "Partitioning histopathological images: an integrated framework for supervised color-texture segmentation and cell splitting," *IEEE transactions on medical imaging*, vol. 30, no. 9, pp. 1661-1677, 2011.
- [280] A. Kårsnäs, A. L. Dahl, and R. Larsen, "Learning histopathological patterns," *Journal of pathology informatics*, vol. 2, no. 2, p. 12, 2011.

- [281] C. Jung and C. Kim, "Segmenting clustered nuclei using H-minima transform-based marker extraction and contour parameterization," *IEEE transactions on biomedical engineering*, vol. 57, no. 10, pp. 2600-2604, 2010.
- [282] A. Belsare, M. Mushrif, M. Pangarkar, and N. Meshram, "Breast histopathology image segmentation using spatio-colour-texture based graph partition method," *Journal of microscopy*, 2015.
- [283] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, no. 285-296, pp. 23-27, 1975.
- [284] W. Niblack, *An introduction to digital image processing*. Strandberg Publishing Company, 1985.
- [285] J. Sauvola and M. Pietikäinen, "Adaptive document image binarization," *Pattern recognition*, vol. 33, no. 2, pp. 225-236, 2000.
- [286] J. Bernsen, "Dynamic thresholding of grey-level images," in *International conference on pattern recognition*, 1986, vol. 2, pp. 1251-1255.
- [287] C. Wolf and J.-M. Jolion, "Extraction and recognition of artificial text in multimedia documents," *Pattern Analysis & Applications*, vol. 6, no. 4, pp. 309-326, 2004.
- [288] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137-154, 2004.
- [289] S. M. Pizer *et al.*, "Adaptive histogram equalization and its variations," *Computer vision, graphics, and image processing*, vol. 39, no. 3, pp. 355-368, 1987.
- [290] I.-J. Kim, "Multi-window binarization of camera image for document recognition," in *Frontiers in Handwriting Recognition, 2004. IWFHR-9 2004. Ninth International Workshop on*, 2004: IEEE, pp. 323-327.
- [291] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," *CA Cancer J Clin*, vol. 69, no. 1, pp. 7-34, Jan 2019, doi: 10.3322/caac.21551.
- [292] D. C. Johnson *et al.*, "Preoperatively misclassified, surgically removed benign renal masses: a systematic review of surgical series and United States population level burden estimate," *J Urol*, vol. 193, no. 1, pp. 30-5, Jan 2015, doi: 10.1016/j.juro.2014.07.102.
- [293] C. J. Kane, K. Mallin, J. Ritchey, M. R. Cooperberg, and P. R. Carroll, "Renal cell cancer stage migration: analysis of the National Cancer Data Base," *Cancer*, vol. 113, no. 1, pp. 78-83, Jul 1 2008, doi: 10.1002/cncr.23518.
- [294] A. Kutikov *et al.*, "Incidence of benign pathologic findings at partial nephrectomy for solitary renal mass presumed to be renal cell carcinoma on preoperative imaging," *Urology*, vol. 68, no. 4, pp. 737-40, Oct 2006, doi: 10.1016/j.urology.2006.04.011.
- [295] B. R. Lane *et al.*, "A preoperative prognostic nomogram for solid enhancing renal tumors 7 cm or less amenable to partial nephrectomy," *J Urol*, vol. 178, no. 2, pp. 429-34, Aug 2007, doi: 10.1016/j.juro.2007.03.106.
- [296] A. Mathew, S. S. Devesa, J. F. Fraumeni, Jr., and W. H. Chow, "Global increases in kidney cancer incidence, 1973-1992," *Eur J Cancer Prev*, vol. 11, no. 2, pp. 171-8, Apr 2002. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/11984136>.
- [297] M. M. Nguyen, I. S. Gill, and L. M. Ellison, "The evolving presentation of renal carcinoma in the United States: trends from the Surveillance, Epidemiology, and End Results program," *J Urol*, vol. 176, no. 6 Pt 1, pp. 2397-400; discussion 2400, Dec 2006, doi: 10.1016/j.juro.2006.07.144.
- [298] I. S. Gill, M. Aron, D. A. Gervais, and M. A. Jewett, "Clinical practice. Small renal mass," *N Engl J Med*, vol. 362, no. 7, pp. 624-34, Feb 18 2010, doi: 10.1056/NEJMcp0910041.

- [299] C. W. Liaw, J. S. Winoker, and R. Mehrazin, "Imaging Protocols for Active Surveillance in Renal Cell Carcinoma," *Curr Urol Rep*, vol. 19, no. 10, p. 81, Aug 13 2018, doi: 10.1007/s11934-018-0830-z.
- [300] M. Daugherty, D. Sedaghatpour, O. Shapiro, S. Vourganti, A. Kutikov, and G. Bratslavsky, "The metastatic potential of renal tumors: Influence of histologic subtypes on definition of small renal masses, risk stratification, and future active surveillance protocols," *Urol Oncol*, vol. 35, no. 4, pp. 153 e15-153 e20, Apr 2017, doi: 10.1016/j.urolonc.2016.11.009.
- [301] S. Campbell *et al.*, "Renal Mass and Localized Renal Cancer: AUA Guideline," *J Urol*, vol. 198, no. 3, pp. 520-529, Sep 2017, doi: 10.1016/j.juro.2017.04.100.
- [302] M. G. Linguraru *et al.*, "Automated noninvasive classification of renal cancer on multiphase CT," *Med Phys*, vol. 38, no. 10, pp. 5738-46, Oct 2011, doi: 10.1118/1.3633898.
- [303] R. Papalia *et al.*, "Novel kidney segmentation system to describe tumour location for nephron-sparing surgery," *World J Urol*, vol. 33, no. 6, pp. 865-71, Jun 2015, doi: 10.1007/s00345-014-1386-2.
- [304] F. Chen *et al.*, "Voxel-based whole-lesion enhancement parameters: a study of its clinical value in differentiating clear cell renal cell carcinoma from renal oncocytoma," *Abdom Radiol (NY)*, vol. 42, no. 2, pp. 552-560, Feb 2017, doi: 10.1007/s00261-016-0891-8.
- [305] H. Coy, J. R. Young, M. L. Douek, M. S. Brown, J. Sayre, and S. S. Raman, "Quantitative computer-aided diagnostic algorithm for automated detection of peak lesion attenuation in differentiating clear cell from papillary and chromophobe renal cell carcinoma, oncocytoma, and fat-poor angiomyolipoma on multiphasic multidetector computed tomography," *Abdom Radiol (NY)*, vol. 42, no. 7, pp. 1919-1928, Jul 2017, doi: 10.1007/s00261-017-1095-6.
- [306] N. J. Farber, Y. Wu, L. Zou, P. Belani, and E. A. Singer, "Challenges in RCC Imaging: Renal Insufficiency, Post-Operative Surveillance, and the Role of Radiomics," *Kidney Cancer J*, vol. 13, no. 4, pp. 84-90, Nov 2015. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/26937265>.
- [307] C. S. C. Cho J.W., Jang, J.Y, Shin, S.K, Choi, K.D., et al., "Lymph node metastases in esophageal carcinoma: an endoscopist's view. ," *Clin Endoscopy* vol. 47, no. 6, pp. 523-429, 2014.
- [308] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans Pattern Anal Mach Intell*, vol. 35, no. 6, pp. 1397-409, Jun 2013, doi: 10.1109/TPAMI.2012.213.
- [309] K. Panetta, Y. Zhou, S. Agaian, and H. Jia, "Nonlinear unsharp masking for mammogram enhancement," *IEEE Trans Inf Technol Biomed*, vol. 15, no. 6, pp. 918-28, Nov 2011, doi: 10.1109/TITB.2011.2164259.
- [310] S. A. J. Arora, Verma A., and Panigrahi P.K., "Multilevel thresholding for image segmentation through a fast statistical recursive algorithm.," *Pattern Recognition Letters*, vol. 29, pp. 119-125, 2008.
- [311] K. M. Kamath, Rajendran, R., Panetta, K., and Agaian, S., "A human visual based binarization technique for histological images.," *Proc. SPIE 10221, Mobile Multimedia/Image Processing, Security, and Applications*, 2017, Art no. 102210Q.
- [312] I. Banic, and Andrej Taranenko, "Measuring Closeness of Graphs - The Hausdorff Distance.," *Bulletin of the Malaysian Mathematical Sciences Society* vol. 40, no. 1, pp. 75-95, 2017.

- [313] V. Ficarra *et al.*, "Tumour contact surface area as a predictor of postoperative complications and renal function in patients undergoing partial nephrectomy for renal tumours," *BJU Int*, Sep 25 2018, doi: 10.1111/bju.14567.
- [314] S. Leslie *et al.*, "Renal tumor contact surface area: a novel parameter for predicting complexity and outcomes of partial nephrectomy," *Eur Urol*, vol. 66, no. 5, pp. 884-93, Nov 2014, doi: 10.1016/j.eururo.2014.03.010.
- [315] Z. E. Khene *et al.*, "Role of quantitative computed tomography texture analysis in the prediction of adherent perinephric fat," *World J Urol*, vol. 36, no. 10, pp. 1635-1642, Oct 2018, doi: 10.1007/s00345-018-2292-9.
- [316] B. A. Varghese *et al.*, "Differentiation of Predominantly Solid Enhancing Lipid-Poor Renal Cell Masses by Use of Contrast-Enhanced CT: Evaluating the Role of Texture in Tumor Subtyping," *AJR Am J Roentgenol*, vol. 211, no. 6, pp. W288-W296, Dec 2018, doi: 10.2214/AJR.18.19551.
- [317] M. G. Linguraru *et al.*, "Computer-aided renal cancer quantification and classification from contrast-enhanced CT via histograms of curvature-related features," *Conf Proc IEEE Eng Med Biol Soc*, vol. 2009, pp. 6679-82, 2009, doi: 10.1109/IEMBS.2009.5334012.
- [318] N. D. Egbert *et al.*, "Differentiation of papillary renal cell carcinoma subtypes on CT and MRI," *AJR Am J Roentgenol*, vol. 201, no. 2, pp. 347-55, Aug 2013, doi: 10.2214/AJR.12.9451.
- [319] R. P. Kopp, L. Aganovic, K. L. Palazzi, F. H. Cassidy, K. Sakamoto, and I. H. Derweesh, "Differentiation of clear from non-clear cell renal cell carcinoma using CT washout formula," *Can J Urol*, vol. 20, no. 3, pp. 6790-7, Jun 2013. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/23783049>.
- [320] J. M. Jacob *et al.*, "Characteristics of the Peritumoral Pseudocapsule Vary Predictably With Histologic Subtype of T1 Renal Neoplasms," *Urology*, vol. 86, no. 5, pp. 956-61, Nov 2015, doi: 10.1016/j.urology.2015.06.015.
- [321] L. Carin and M. J. Pencina, "On Deep Learning for Medical Image Analysis," *JAMA*, vol. 320, no. 11, pp. 1192-1193, Sep 18 2018, doi: 10.1001/jama.2018.13316.
- [322] G. Buizza *et al.*, "Early tumor response prediction for lung cancer patients using novel longitudinal pattern features from sequential PET/CT image scans," *Phys Med*, vol. 54, pp. 21-29, Oct 2018, doi: 10.1016/j.ejmp.2018.09.003.
- [323] W. Sun, M. Jiang, J. Dang, P. Chang, and F. F. Yin, "Effect of machine learning methods on predicting NSCLC overall survival time based on Radiomics analysis," *Radiat Oncol*, vol. 13, no. 1, p. 197, Oct 5 2018, doi: 10.1186/s13014-018-1140-9.
- [324] F. L. Bayisa, X. Liu, A. Garpebring, and J. Yu, "Statistical learning in computed tomography image estimation," *Med Phys*, vol. 45, no. 12, pp. 5450-5460, Dec 2018, doi: 10.1002/mp.13204.
- [325] K. Sasaguri and N. Takahashi, "CT and MR imaging for solid renal mass characterization," *Eur J Radiol*, vol. 99, pp. 40-54, Feb 2018, doi: 10.1016/j.ejrad.2017.12.008.
- [326] S. Hemachander, Verma, A., Arora, S., and Panigrahi, P.K., "Locally adaptive block thresholding method with continuity constraint," *Pattern Recognition Letters*, vol. 28, no. 1, pp. 119-124 2007.
- [327] M. Srivastava, Singh, S.K. & Panigrahi, P.K., "A Semi-Automated Statistical Algorithm for Object Separation," *Circuits Syst Signal Process*, vol. 32, no. 6, pp. 3059–3078, 2013.

- [328] H. Cui *et al.*, "A topo-graph model for indistinct target boundary definition from anatomical images," *Comput Methods Programs Biomed*, vol. 159, pp. 211-222, Jun 2018, doi: 10.1016/j.cmpb.2018.03.018.
- [329] H. Chen *et al.*, "A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films," *Scientific reports*, vol. 9, no. 1, pp. 1-11, 2019.
- [330] E. H. Said, D. E. M. Nassar, G. Fahmy, and H. H. Ammar, "Teeth segmentation in digitized dental X-ray films using mathematical morphology," *IEEE transactions on information forensics and security*, vol. 1, no. 2, pp. 178-189, 2006.
- [331] S. Shah, A. Abaza, A. Ross, and H. Ammar, "Automatic tooth segmentation using active contour without edges," in *2006 biometrics symposium: special session on research at the biometric consortium conference*, 2006: IEEE, pp. 1-6.
- [332] A. E. Rad, M. S. M. Rahim, and A. Norouzi, "Digital dental X-ray image segmentation and feature extraction," *TELKOMNIKA Indonesian Journal of Electrical Engineering*, vol. 11, no. 6, pp. 3109-3114, 2013.
- [333] M. H. Mahoor and M. Abdel-Mottaleb, "Classification and numbering of teeth in dental bitewing images," *Pattern Recognition*, vol. 38, no. 4, pp. 577-586, 2005.
- [334] F. Aeini and F. Mahmoudi, "Classification and numbering of posterior teeth in bitewing dental images," in *2010 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE)*, 2010, vol. 6: IEEE, pp. V6-66-V6-72.
- [335] A. Yuniarti, A. S. Nugroho, B. Amaliah, and A. Z. Arifin, "Classification and numbering of dental radiographs for an automated human identification system," *Telkomnika*, vol. 10, no. 1, p. 137, 2012.
- [336] P.-L. Lin, Y.-H. Lai, and P.-W. Huang, "An effective classification and numbering system for dental bitewing radiographs using teeth region and contour information," *Pattern recognition*, vol. 43, no. 4, pp. 1380-1392, 2010.
- [337] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual review of biomedical engineering*, vol. 19, pp. 221-248, 2017.
- [338] L. Liu, J. Xu, Y. Huan, Z. Zou, S.-C. Yeh, and L.-R. Zheng, "A Smart Dental Health-IoT Platform Based on Intelligent Hardware, Deep Learning, and Mobile Terminal," *IEEE journal of biomedical and health informatics*, vol. 24, no. 3, pp. 898-906, 2019.
- [339] J.-H. Lee, D.-H. Kim, S.-N. Jeong, and S.-H. Choi, "Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm," *Journal of dentistry*, vol. 77, pp. 106-111, 2018.
- [340] J. Krois *et al.*, "Deep learning for the radiographic detection of periodontal bone loss," *Scientific reports*, vol. 9, no. 1, pp. 1-6, 2019.
- [341] J.-H. Lee, D.-h. Kim, S.-N. Jeong, and S.-H. Choi, "Diagnosis and prediction of periodontally compromised teeth using a deep learning-based convolutional neural network algorithm," *Journal of periodontal & implant science*, vol. 48, no. 2, p. 114, 2018.
- [342] T. Ekert *et al.*, "Deep learning for the radiographic detection of apical lesions," *Journal of endodontics*, vol. 45, no. 7, pp. 917-922. e5, 2019.
- [343] C.-H. Wu, W.-H. Tsai, Y.-H. Chen, J.-K. Liu, and Y.-N. Sun, "Model-based orthodontic assessments for dental panoramic radiographs," *IEEE Journal of biomedical and health informatics*, vol. 22, no. 2, pp. 545-551, 2017.

- [344] Y. Gan, Z. Xia, J. Xiong, G. Li, and Q. Zhao, "Tooth and alveolar bone segmentation from dental computed tomography images," *IEEE journal of biomedical and health informatics*, vol. 22, no. 1, pp. 196-204, 2017.
- [345] F. Schwendicke, T. Golla, M. Dreher, and J. Krois, "Convolutional neural networks for dental image diagnostics: A scoping review," *Journal of dentistry*, vol. 91, p. 103226, 2019.
- [346] F. a. Schwendicke, W. Samek, and J. Krois, "Artificial intelligence in dentistry: chances and challenges," *Journal of dental research*, vol. 99, no. 7, pp. 769-774, 2020.
- [347] DentistryIQ. [Online]. Available: <https://www.dentistryiq.com/dentistry/prosthodontics-and-labatory/article/16363181/6-reasons-why-you-should-use-a-panoramic-xray-for-bitewings>.
- [348] M. Abdinian, S. M. Razavi, R. Faghihian, A. A. Samety, and E. Faghihian, "Accuracy of digital bitewing radiography versus different views of digital panoramic radiography for detection of proximal caries," *Journal of Dentistry (Tehran, Iran)*, vol. 12, no. 4, p. 290, 2015.
- [349] A. Rad, M. Rahim, A. Rehman, and T. Saba, "Digital dental X-ray database for caries screening. 3D Research, 7 (2), 1–5," ed, 2016.
- [350] A. H. Abdi, S. Kasaei, and M. Mehdizadeh, "Automatic segmentation of mandible in panoramic x-ray," *Journal of Medical Imaging*, vol. 2, no. 4, p. 044003, 2015.
- [351] G. Jader, J. Fontineli, M. Ruiz, K. Abdalla, M. Pithon, and L. Oliveira, "Deep instance segmentation of teeth in panoramic X-ray images," in *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2018: IEEE, pp. 400-407.
- [352] J. Kim, H.-S. Lee, I.-S. Song, and K.-H. Jung, "DeNTNet: Deep Neural Transfer Network for the detection of periodontal bone loss using panoramic dental radiographs," *Scientific reports*, vol. 9, no. 1, pp. 1-9, 2019.
- [353] A. B. Oktay, "Tooth detection with Convolutional Neural Networks," in *2017 Medical Technologies National Congress (TIPTEKNO)*, 2017: IEEE, pp. 1-4.
- [354] C. Muramatsu *et al.*, "Tooth detection and classification on panoramic radiographs for automatic dental chart filing: improved classification by multi-sized input data," *Oral radiology*, vol. 37, no. 1, pp. 13-19, 2021.
- [355] D. V. Tuzoff *et al.*, "Tooth detection and numbering in panoramic radiographs using convolutional neural networks," *Dentomaxillofacial Radiology*, vol. 48, no. 4, p. 20180051, 2019.
- [356] M. Fukuda *et al.*, "Evaluation of an artificial intelligence system for detecting vertical root fracture on panoramic radiography," *Oral radiology*, pp. 1-7, 2019.
- [357] J.-E. Kim, N.-E. Nam, J.-S. Shim, Y.-H. Jung, B.-H. Cho, and J. J. Hwang, "Transfer learning via deep neural networks for implant fixture system classification using periapical radiographs," *Journal of clinical medicine*, vol. 9, no. 4, p. 1117, 2020.
- [358] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size," *arXiv preprint arXiv:1602.07360*, 2016.
- [359] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1-9.
- [360] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510-4520.



- [361] H.-J. Chang *et al.*, "Deep Learning Hybrid Method to Automatically Diagnose Periodontal Bone Loss and Stage Periodontitis," *Scientific Reports*, vol. 10, no. 1, pp. 1-8, 2020.
- [362] V. Geetha, K. Aprameya, and D. M. Hinduja, "Dental caries diagnosis in digital radiographs using back-propagation neural network," *Health Information Science and Systems*, vol. 8, no. 1, pp. 1-14, 2020.
- [363] M. P. Muresan, A. R. Barbura, and S. Nedevschi, "Teeth Detection and Dental Problem Classification in Panoramic X-Ray Images using Deep Learning and Image Processing Techniques."
- [364] E. Romera, J. M. Alvarez, L. M. Bergasa, and R. Arroyo, "Erfnet: Efficient residual factorized convnet for real-time semantic segmentation," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 1, pp. 263-272, 2017.
- [365] C. Kim, D. Kim, H. Jeong, S.-J. Yoon, and S. Youm, "Automatic Tooth Detection and Numbering Using a Combination of a CNN and Heuristic Algorithm," *Applied Sciences*, vol. 10, no. 16, p. 5624, 2020.
- [366] G. F. Fahmy *et al.*, "Toward an automated dental identification system," *Journal of Electronic Imaging*, vol. 14, no. 4, p. 043018, 2005.
- [367] B. Sathya and R. Neelaveni, "Transfer learning based automatic human identification using dental traits-an aid to forensic odontology," *Journal of Forensic and Legal Medicine*, p. 102066, 2020.
- [368] A. Karagyris *et al.*, "Creation and validation of a chest X-ray dataset with eye-tracking and report dictation for AI development," *Scientific data*, vol. 8, no. 1, pp. 1-18, 2021.
- [369] N. Khosravan, H. Celik, B. Turkbey, E. C. Jones, B. Wood, and U. Bagci, "A collaborative computer aided diagnosis (C-CAD) system with eye-tracking, sparse attentional model, and deep learning," *Medical image analysis*, vol. 51, pp. 101-115, 2019.
- [370] J. N. Stemmer *et al.*, "Eye tracking for deep learning segmentation using convolutional neural networks," *Journal of digital imaging*, vol. 32, no. 4, pp. 597-604, 2019.
- [371] G. Aresta *et al.*, "Automatic lung nodule detection combined with gaze information improves radiologists' screening performance," *IEEE journal of biomedical and health informatics*, vol. 24, no. 10, pp. 2894-2901, 2020.
- [372] S. Mall, P. C. Brennan, and C. Mello-Thoms, "Modeling visual search behavior of breast radiologists using a deep convolution neural network," *Journal of Medical Imaging*, vol. 5, no. 3, p. 035502, 2018.
- [373] B. P. Hermanson, G. C. Burgdorf, J. F. Hatton, D. M. Speegle, and K. F. Woodmansey, "Visual fixation and scan patterns of dentists viewing dental periapical radiographs: an eye tracking pilot study," *Journal of endodontics*, vol. 44, no. 5, pp. 722-727, 2018.
- [374] Tobii Gaming. "Tobii Eye Tracker 4C." <https://tobiigaming.com/eye-tracker-4c/> (accessed).
- [375] M. T. Baghdady, H. Carnahan, E. W. Lam, and N. N. Woods, "Dental and dental hygiene students' diagnostic accuracy in oral radiology: effect of diagnostic strategy and instructional method," *Journal of dental education*, vol. 78, no. 9, pp. 1279-1285, 2014.
- [376] Google Cloud. "Cloud Speech-to-Text." <https://cloud.google.com/speech-to-text> (accessed).
- [377] Nuance. "Dragon Speech Recognition." <https://www.nuance.com/dragon/support/dragon-naturallyspeaking.html> (accessed).
- [378] S. C. White and M. J. Paroah, "White's and Pharoah's Oral Radiology: Principles And Interpretation," 2019.

- [379] S. M. Pizer, R. E. Johnston, J. P. Ericksen, B. C. Yankaskas, and K. E. Muller, "Contrast-limited adaptive histogram equalization: speed and effectiveness," in *[1990] Proceedings of the First Conference on Visualization in Biomedical Computing*, 1990: IEEE Computer Society, pp. 337,338,339,340,341,342,343,344,345-337,338,339,340,341,342,343,344,345.
- [380] Y.-T. Kim, "Contrast enhancement using brightness preserving bi-histogram equalization," *IEEE transactions on Consumer Electronics*, vol. 43, no. 1, pp. 1-8, 1997.
- [381] D. Sheet, H. Garud, A. Suveer, M. Mahadevappa, and J. Chatterjee, "Brightness preserving dynamic fuzzy histogram equalization," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 4, pp. 2475-2480, 2010.
- [382] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 6, pp. 1397-1409, 2012.
- [383] K. Panetta, A. Samani, and S. Agaian, "Choosing the optimal spatial domain measure of enhancement for mammogram images," *International journal of biomedical imaging*, vol. 2014, 2014.
- [384] S. S. Agaian, B. Silver, and K. A. Panetta, "Transform coefficient histogram-based image enhancement algorithms using contrast entropy," *IEEE transactions on image processing*, vol. 16, no. 3, pp. 741-758, 2007.
- [385] C. Gao, K. Panetta, and S. Agaian, "Color image attribute and quality measurements," in *Mobile Multimedia/Image Processing, Security, and Applications 2014*, 2014, vol. 9120: International Society for Optics and Photonics, p. 91200T.
- [386] E. Wharton, K. Panetta, and S. Agaian, "Human visual system based similarity metrics," in *2008 IEEE International Conference on Systems, Man and Cybernetics*, 2008: IEEE, pp. 685-690.
- [387] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117-2125.
- [388] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep learning in medical image analysis and multimodal learning for clinical decision support*: Springer, 2018, pp. 3-11.
- [389] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2881-2890.
- [390] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801-818.
- [391] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [392] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.
- [393] F. Isensee *et al.*, "nnu-net: Self-adapting framework for u-net-based medical image segmentation," *arXiv preprint arXiv:1809.10486*, 2018.
- [394] F. Amato, A. López, E. M. Peña-Méndez, P. Vañhara, A. Hampl, and J. Havel, "Artificial neural networks in medical diagnosis," ed: Elsevier, 2013.



- [395] K. Panetta *et al.*, "Iseecolor: method for advanced visual analytics of eye tracking data," *IEEE Access*, vol. 8, pp. 52278-52287, 2020.
- [396] Q. Wan, A. Kaszowska, K. Panetta, H. A. Taylor, and S. Agaian, "Enhanced head-mounted eye tracking data analysis using super-resolution," *Electronic Imaging*, vol. 2019, no. 3, pp. 647-1-647-8, 2019.
- [397] Q. Wan, A. Kaszowska, K. Panetta, H. A. Taylor, and S. Agaian, "A comprehensive head-mounted eye tracking review: software solutions, applications, and challenges," *Electronic Imaging*, vol. 2019, no. 3, pp. 654-1-654-9, 2019.
- [398] H. L. Kundel, C. F. Nodine, and D. Carmody, "Visual scanning, pattern recognition and decision-making in pulmonary nodule detection," *Investigative radiology*, vol. 13, no. 3, pp. 175-181, 1978.
- [399] H. L. Kundel, C. F. Nodine, and E. A. Krupinski, "Searching for lung nodules. Visual dwell indicates locations of false-positive and false-negative decisions," *Investigative radiology*, vol. 24, no. 6, pp. 472-478, 1989.
- [400] C. F. Nodine, H. L. Kundel, S. C. Lauver, and L. C. Toto, "Nature of expertise in searching mammograms for breast masses," *Academic radiology*, vol. 3, no. 12, pp. 1000-1006, 1996.
- [401] W. F. Auffermann, E. A. Krupinski, and S. Tridandapani, "Search pattern training for evaluation of central venous catheter positioning on chest radiographs," *Journal of Medical Imaging*, vol. 5, no. 3, p. 031407, 2018.
- [402] E. Helbren *et al.*, "Towards a framework for analysis of eye-tracking studies in the three dimensional environment: a study of visual search by experienced readers of endoluminal CT colonography," *The British journal of radiology*, vol. 87, no. 1037, p. 20130614, 2014.
- [403] L. McLaughlin, R. Bond, C. Hughes, J. McConnell, and S. McFadden, "Computing eye gaze metrics for the automatic assessment of radiographer performance during X-ray image interpretation," *International journal of medical informatics*, vol. 105, pp. 11-21, 2017.
- [404] J. K. O'Regan and A. Levy-Schoen, *Eye movements from physiology to cognition: Selected/edited proceedings of the third European conference on eye movements, Dourdan, France, September 1985*. Elsevier, 2013.