Performance Analysis of Key Facial Recognition Algorithms

Kevin Naranjo
Panetta Simulation Lab
Department of Electrical and Computer Engineering

Abstract

Over the past two decades, the field of machine learning has been quickly growing due to the increase in computational power and availability of large data sets. One of the main applications of machine learning is image classification, which can be in turn used for facial recognition. State of the art facial recognition algorithms can achieve success rates as high as the mid-90's. However, this is only based on colored and grayscale images.

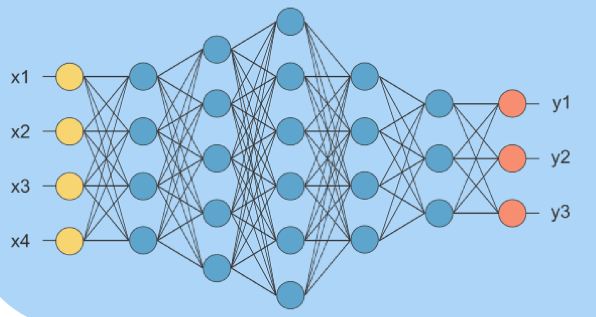
This project seeks to test modern recognition algorithms on non-conventional images, such as sketches, infrared, and night vision. This is done through the use of convolutional neural networks and transfer learning. The data set used for the training and testing of the neural networks was created at Tufts and uses a wide variety of image types.

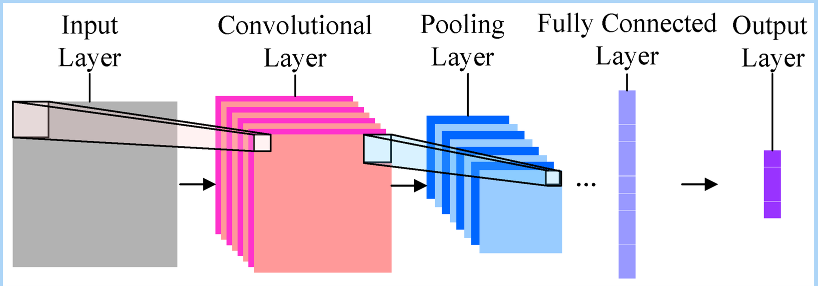
Background

The highest recognition rates for image classification are achieved through the use of data structures known as neural networks. These data structures are composed of several sequential layers, with each layer being made up of several nodes. Each node in a layer is connected to all the nodes in the layer directly in front of it, creating a fully connected structure between layers. The output of each node is the linear combination of all the inputs to that node, with each input being multiplied by it's own weight. The final layer is the output of the network and typically provides a probability for each classification ranging from 0 to 1.

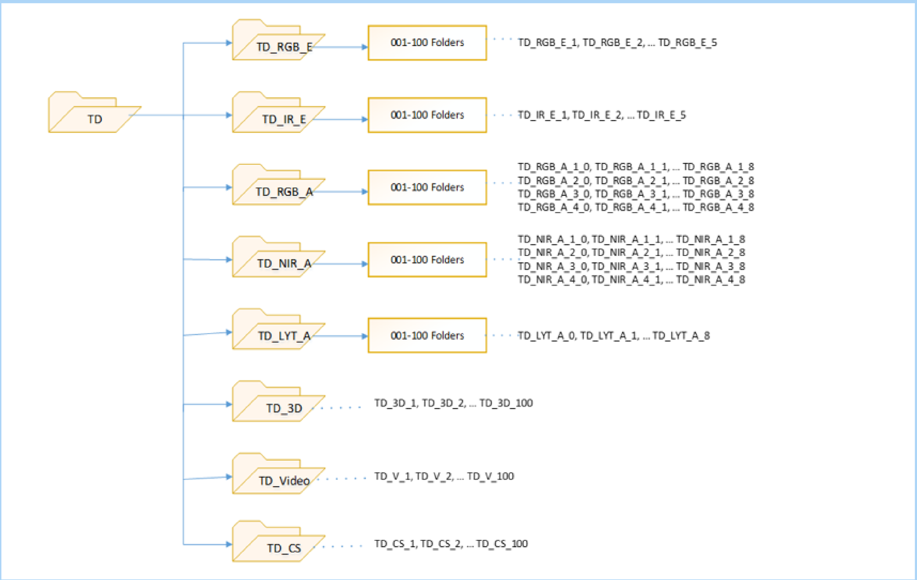
This basic structure, however, has difficulty recognizing rotated or shifted versions of the images it was trained on. This is resolved through the use of a Convolutional Neural Network (CNN). A CNN can be thought of as a basic neural network with some additional layers at the beginning. There are two types of layers that can be combined to form the front-end of the CNN, they are the convolution layer and the max pooling layer. The purpose of the convolution layers is to extract certain features, such as edges, from the image through the use of filters that scan across the data. The purpose of the max pooling layers is to reduce the size of the data while preserving the necessary information. This is done to make training faster and easier.

These networks are trained by feeding in data along with the expected output. At the beginning of the training, the output will not be correct because the initial weights in the network are randomized. However, through each iteration of the training, the network slightly adjusts each weight to get closer to the desired output for a given input.





Data Collection



Tufts Database Structure

A large portion of the project was focused on acquiring and organizing facial images to train and test the network. This was done by having a little over 100 volunteers come in and have their photos taken with several different cameras. The final result of this process became known as the Tufts Database, which can be seen on the side.

In the database, there is a directory for each of the image types, which then contain the images for each participant. The cameras used were able to take infrared, near-infrared, color, and thermal images of each participant. Additionally, a video that circled each participant was also taken. Once all the photos were taken, 3D models and computerized sketches were also created for each participant and stored in the database.

Methodology

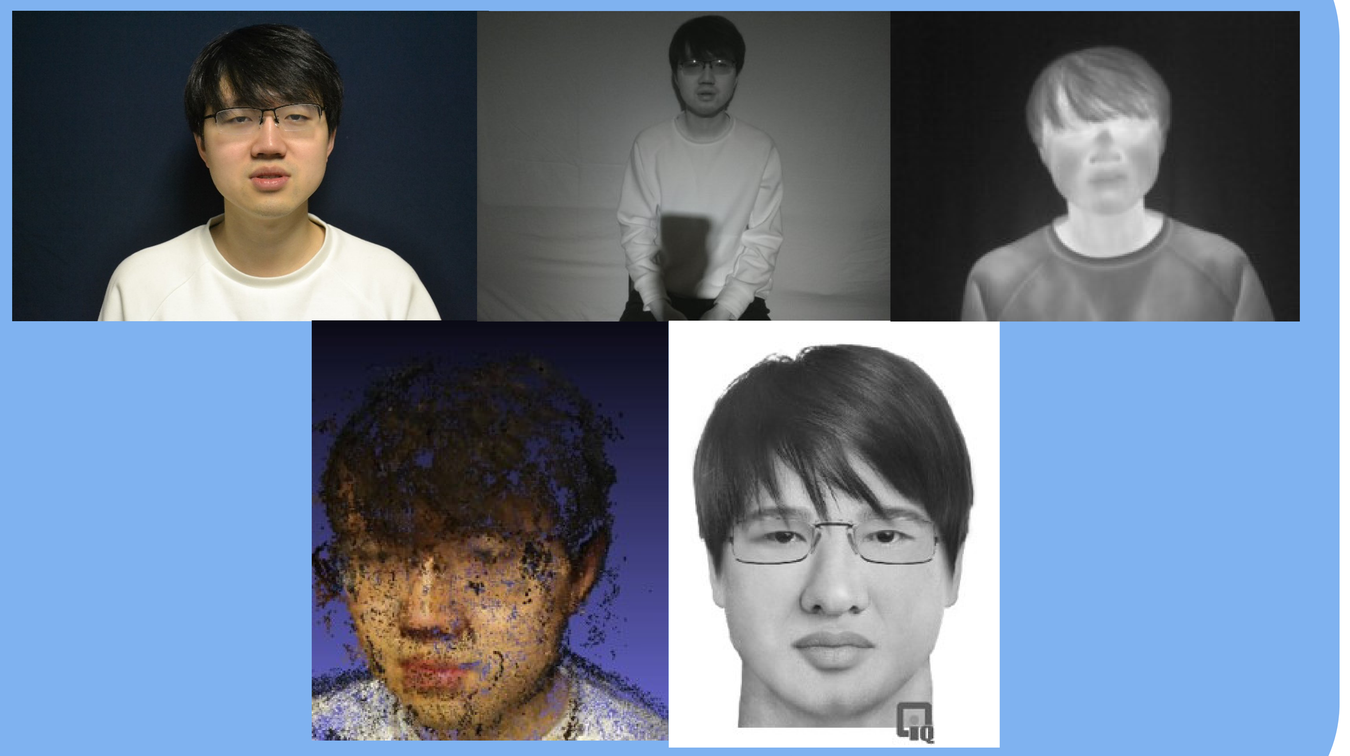
To create the neural network, the Keras library and the Tensorflow framework were used in Python. Since successful training of a CNN requires around 1 million images, we used a process known as transfer learning. The pretrained model used was VGG-Face, which can be seen on the right. Additionally, a fully connected network with a few layers was attached at the end for classification. All the weights were frozen except the lower convolution layers and the fully connected layers.

The images were categorized based on the participant's number. The network was then trained with the color images and tested on the sketches. This same process could be applied to test for thermal and near-infrared images.

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	(None, 150, 150, 3)	0
conv1_1 (Conv2D)	(None, 150, 150, 64)	1792
conv1_2 (Conv2D)	(None, 150, 150, 64)	36928
pool1 (MaxPooling2D)	(None, 75, 75, 64)	0
conv2_1 (Conv2D)	(None, 75, 75, 128)	73856
conv2_2 (Conv2D)	(None, 75, 75, 128)	147584
pool2 (MaxPooling2D)	(None, 37, 37, 128)	0
conv3_1 (Conv2D)	(None, 37, 37, 256)	295168
conv3_2 (Conv2D)	(None, 37, 37, 256)	590080
conv3_3 (Conv2D)	(None, 37, 37, 256)	590080
pool3 (MaxPooling2D)	(None, 18, 18, 256)	0
conv4_1 (Conv2D)	(None, 18, 18, 512)	1180160
conv4_2 (Conv2D)	(None, 18, 18, 512)	2359808
conv4_3 (Conv2D)	(None, 18, 18, 512)	2359808
pool4 (MaxPooling2D)	(None, 9, 9, 512)	0
conv5_1 (Conv2D)	(None, 9, 9, 512)	2359808
conv5_2 (Conv2D)	(None, 9, 9, 512)	2359808
conv5_3 (Conv2D)	(None, 9, 9, 512)	2359808
pool5 (MaxPooling2D)	(None, 4, 4, 512)	0

Total params: 14,714,688
Trainable params: 14,714,688
Non-trainable params: 0

VGG-Face Model

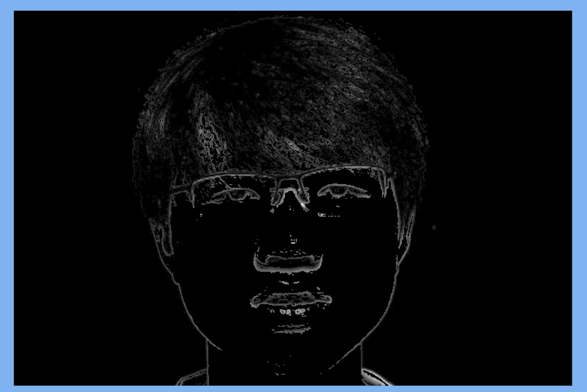


Sample Images of Participant in Tufts Database

Conclusion

After initial tests of the network, it appears that the recognition rate for the sketches is rather low compared to the colored image recognition. There are several possible paths to improve this accuracy. One of which is to adjust the layout of the fully connected layers at the end of the model as well as changing which weights are frozen in the pretrained model. Another possible improvement would be to change how the images are divided into the training set and the testing set. One final avenue that could be taken would be to modify the input images with some preprocessing in order to make the color images and the sketches appear more similar.

One possible preprocessing method is the Human Visual System developed by Panetta Simulation Lab. The next steps in the project would be to experiment with the methods mentioned above as well as working on making the Tufts database available to other researchers in the field.



HVS Image

Acknowledgements

Thanks to Tufts Summer Scholars for providing the opportunity to carry out this project. Also, thanks to Dr. Panetta as well as her head researcher Qianwen Wan for providing mentorship and advice throughout the process. Additionally, thanks to all the participants who volunteered to form the database.