Supervised Deep Learning: Real-Time Face Recognition

Amir Sotoodeh
CS5990: Deep Learning
California Polytechnic University: Pomona

Abstract

Facial recognition is a prominent topic of discussion as an application to a variety of systems pertaining to security measures. In terms of machine learning, face recognition has always been an abstract and seemingly complex task with respect to computer vision given that facial landmarks must first be accurately examined and observed. Although there have been many recent advancements in facial recognition, the greatest challenge to overcome has been the means of finding an efficient and quick ways to compute facial similarities or matches. A typical deep feed-forward neural network can be quite computationally expensive in its forward and backward propagation of data, requiring costly hardware to keep up with its capabilities. Hence, the task of real-time facial recognition for security purposes is a difficult task to complete with respects to limited hardware and efficiency. This research looks into various state-of-the-art techniques that can be utilized for facial recognition via hardware such as webcams and/or security cameras in real time. Additionally, it will discuss a history of face recognition and provide an overview of the task in how it is accomplished.

Introduction

There are many uses for facial recognition; namely, in video surveillance and aiding in search of specific criminals from available cameras in public. Additionally, it can also be applied as a safety/security measure for high authority officials to enter restricted premises. As such, it is an extremely active field in research by many data scientists and machine learning enthusiasts. Given that there exists a surplus of data from the proliferation of social media in recent years, powerful machine learning and statistical models could be built with significantly high accuracy rates. This technology is actively seen in Apple's products or Facebook's photo face recognition feature. When searching for a name, it is commonly asked to tag a specific individual after detection of a face in a photo.

The goal of this project is to research the potential usage of machine learning models such as deep convolutional neural networks to solve the task of facial recognition in real-time video feed. Additionally, research in the future will be conducted to obtain a greater or more efficient performance of this task by using state of the art statistical models while training on large public/open source dataset.

Methodology

The task of facial recognition is broken down into three steps:

1. Facial Detection
2. Facial Alignment*
3. Facial Recognition

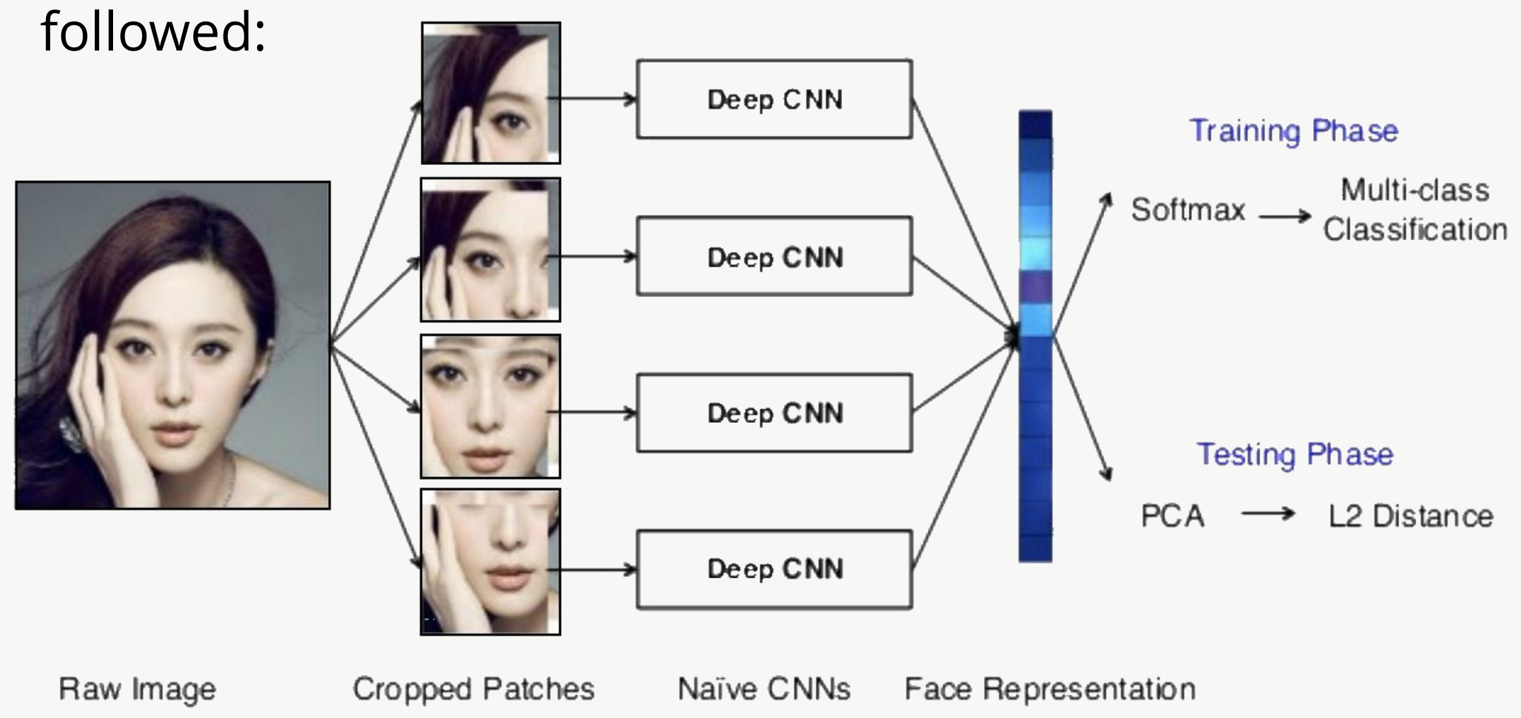
Haar features are a set of filters, much like convolution filters), that can be used to detect lines and edges. These features are applied to several regions of an image, to which the Viola-Jones algorithm is applied. The goal is to take the sum of the corresponding positions of the black pixels in the original image and subtract it from the sum of the white pixels. This becomes an order of O(n*m) computation where n is the number of rows and m is the number of columns of pixels. The Viola Jones Algorithm [3] is as follows:

$$\frac{1}{n}\sum_{black}^{n} I(X) - \frac{1}{n}\sum_{white}^{n} I(X)$$

However, to optimize this computation, we must first compute the integral image **I** for a given image. The corresponding integral image is produced when every pixel of the original image is the sum of the current pixel value and every pixel above and to the left of the image. We can now calculate the sum of pixel values in constant time using the following formula:

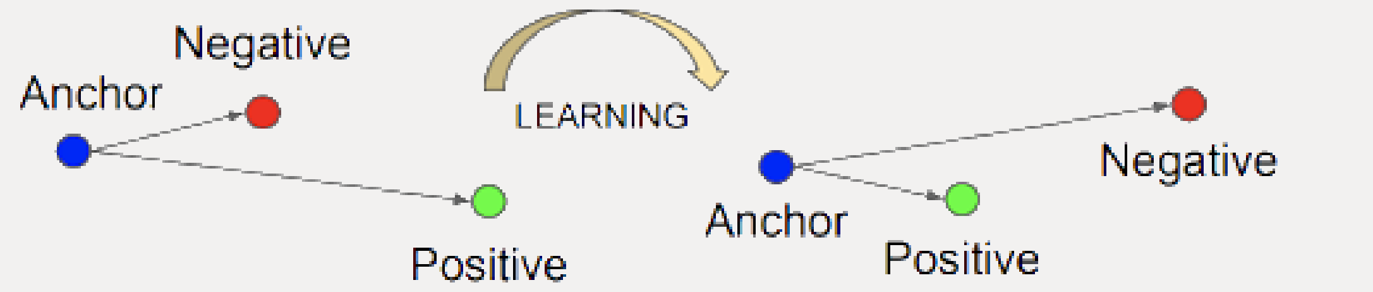
$$\sum_{(x,y)\,\epsilon\,ABCD} I(x,y) = I(D) + I(A) - I(B) - I(C)$$

Additionally, we are able to detect faces using a series of cascading convolutional networks [4]. However, although the performance of a CNN may be more accurate with respect to varying input images, real-time applications require less computational expensive algorithmic based approaches. Using classifiers such as Haar Cascade and HOG SVM are greatly preferred over deep learning approaches in terms of efficiency. The traditional face recognition approach generally followed a pattern as followed:



Recognition

The traditional approach to face recognition had several fundamental issues with respect to how robust the model was able to become. By utilizing a categorical cross entropy loss function to determine probabilities of a set amount of classes, a model will not be able to distinguish between individuals with similar facial features. Additionally, by training to identify a specific class, the accuracy for recognizing individuals significantly decreases as the number of classes increase and fewer images of the same person are used. If we wish to add another individual to the knowledge base of N people, we are forced to retrain the network to determine the appropriate weights and biases for a specific set of individuals.
Researchers from Google resolved this complication with the introduction of the **triplet loss** [1].



In contrast to the previous models utilizing a softmax cross entropy loss function, a metric loss function using triplets allows for a quantifiable way to compare and contrast facial features. A triplet consists of three components:

- Anchor - Image of an individual
- Positive Exemplar - Another image of the same individual
- Negative Exemplar - An image of an individual with similar facial features.

Computing the triplet loss:

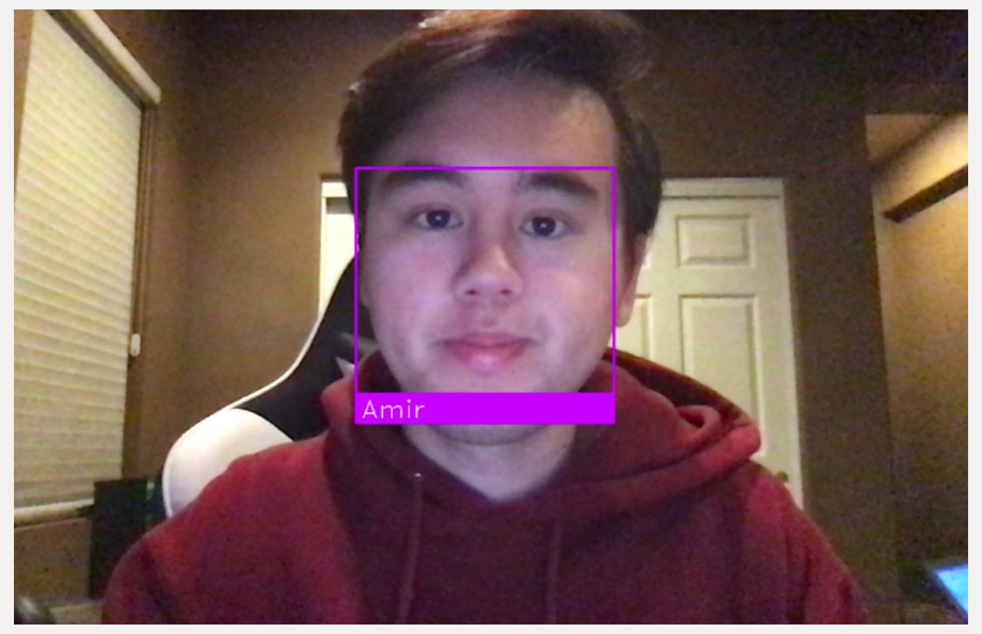
$$\mathbf{L} = \sum_{\mathbf{i}}^{\mathbf{N}} \left[\left\| \mathbf{f}(\mathbf{x_i}^{\mathbf{a}}) - \mathbf{f}(\mathbf{x_i}^{\mathbf{p}}) \right\| - \left\| \mathbf{f}(\mathbf{x_i}^{\mathbf{a}}) - \mathbf{f}(\mathbf{x_i}^{\mathbf{n}}) \right\| + \boldsymbol{\alpha} \right]$$

Where $f(x^a)$, $f(x^p)$, $f(x^n)$ are the vector embedding output of the convolutional neural network. α represents the hyper-parameter responsible for pushing the distance between L2 distances of the positive/anchor and negative/anchor embedding. By computing the L2 (Euclidean) distance between the 128-dimensional identity vector of two images, this value will represent how similar two faces are to one another. Additionally, specifying a threshold value will determine how sensitive or how different a face must be in order to be considered a match or not a match.
With respect to recognition we compare the 128-d feature vector to a set of known encodings until another vector is found within the threshold.

Results & Conclusion

With a triplet loss defined convolutional neural network, Google's FaceNet had devised the current state-of-the art method for accurate facial recognition models. Metric learning allows a network to learn the true deep features in a complex human face and compare with other faces with a single forward pass.

Using Adam Geitgey's amazing facial recognition python api, we are able to test pre-trained state of the art facial recognition models in real-time with ease. The following image is from a single frame of a webcam demonstration:



Facial recognition has a surplus of applications in the real world but still requires fine-tuning. Improvements or adjustments should be made to account for unwanted attacks such as falsified images of a face to fool a security system. Additionally, to improve the accuracy GANs are a beneficial candidate for generating effective triplets by created faces with similar, year different feature vectors with differing labels.

References & Acknowledgements

[1] Florian, S., Kalenichenko, D., Philbin, J., (March 2015) FaceNet: A Unified Embedding for Face Recognition and Clustering. Obtained From: https://arxiv.org/pdf/1503.03832.pdf
[2] Wang, M., Deng, W., (Sep 2018) Deep Face Recognition: A Survey. Obtained From: https://arxiv.org/pdf/1804.06655.pdf
[3] Cen, K, Study of Viola-Jones Real Time Face Detector. Obtained From: https://web.stanford.edu/class/cs231a/prev_projects_2016/cs231a_final_report.pdf
[4] Sun, Y., et al. Deep Convolutional Network Cascade for Facial Point Detection.
Please view the paper for a full list of references.

Thank you to Dr. Hao Ji for the inspiration and opportunity to enter the exciting and amazing field of Machine Learning. Additionally, thank you to all the authors of the preceding papers and any other references I have used to learn about this revolutionary task.