A Comparison of a Convolutional Neural Network to Eigenfaces as Facial Recognition Systems

Hanad Sharmarke (100941632)
Zayd Bille (100920315)
COMP 4107
Carleton University

Introduction

Neural networks are software systems that aim to replicate the biological structures and systems of the human brain. They are an intuitive approach to the ongoing struggle of creating intelligent computational systems.

Facial recognition began as a herculean task. Unlike handwritten digits, images of faces often have very large feature sets. This makes analyzing faces based on macro features such as mouths, noses, eyes, etc rather difficult. One of the earliest facial recognition systems was the system developed by Woodrow Wilson Bledsoe who created a program that allowed you to manually input the coordinates of certain facial features and then determined the closest related face using that information.

This field of study has come a long way since then. Throughout the past decade, facial recognition algorithms have become extremely sophisticated. Often outperforming actual human ability. In 2006, Google's facial recognition systems scored a 100% accuracy on the LFW database. Furthermore, various systems scored a similarly high accuracy in the Face Recognition Grand Challenge which is a U.S. Government sponsored evaluation of emerging facial recognition technologies. It was found that the error rate of most of these systems has decreased by an overall factor of 272 since 1993!

Our task was to implement a solution to the facial recognition task using the convolutional neural network architecture. In this report we are going to describe our implementation of that network and the results that arose from our experimentation.

Data

The dataset is a collection of JPEG pictures of famous people that are collected from the internet. Each picture is centered on a single on a single face with an RGB channel for each pixel. The data is filtered such that only people who have a minimum of 10 pictures is used which resulted in 3243 training images and 1081 test images (62 X 27 pixels) form 158 subjects.

Convolutional Neural Networks

A convolutional neural network is a category of CNN that is used for image recognition and classification. It is inspired by the visual cortex where there is a lot of connectivity between the neurons and each neuron responds to stimuli only in a restricted region. The CNN is similar to multilayer perceptron; it has an input and output layer as well as multiple hidden layers. The hidden layers of CNN consists of the convolutional layer, pooling layers, fully connected layers, and normalization layers. The convolutional layer consists of a set of filters that crosses the height and width of the input layers and computes the dot product between the entries to create a 2 dimensional activation map for each filter. The pooling layer is used to reduce the size of the amount of parameters and the computation in the network using a MAX operation. Fully connected layers connect every neuron to every activation in previous layer.

With the output volumes from the CNN architectures a softmax function cross entropy was applied with the labels (158 subjects). This gives a probability error distribution for each label. The model was then trained using a learning rate of 0.001, a decay of 0.9 and a cross-validation with batch size of 100. To evaluate the face recognition performance a random test batch was selected from the test set and the number of images the system was able to classify was taken as a percentage.

Hu et al from University of London conducted a study where they evaluated CNNs for facial recognition. They developed three CNN architectures of different sizes (small(S), medium(M) and large(L)). A variation of these architectures (Table 1) was implemented in this study.

Eigenfaces

The Eigenfaces technique for facial recognition is a technique that employs PCA on a set of face images that focuses on the overall appearance of each face as well as the variation of every face in a collection to (1) recognize whether or not an unknown image is a face and (2) determine the identity of the face in the database (or the most

similar face in the database in the case where the identity of the unknown image does not appear in the database) with a relatively high amount of accuracy. It was first introduced by Matthew Turk and Alex Pentland.
The idea behind Eigenfaces is simple - given a collection of face images, we can have a subset of images. If M equals the number of images in the collection, then let N be the number of images in the subset where N < M. This subset of images consists of images that contain only the most important and varied components of all the faces in the collection. These images are called the “eigenfaces”. It’s helpful to consider them the ‘ingredients’ for any human face as each face in the collection can be recreated with a linear combination of these eigenfaces.
The “mean” face is calculated by summing the vectors of all images divided by the number of vectors in the training set. Once the average (mean) face is obtained, the training set images are normalized by subtracting the average face from each of them. Then we calculate the eigenfaces using PCA. Afterwards we calculate the weights for each training image against each eigenface. Next, we save them in an N by M matrix, where N equals the number of training images and M equals the number of significant eigenfaces. Given a new input face we subtract the “mean” face, calculate the weight matrix for that image against each eigenface. Lastly, we calculate the distance between the weight matrix for the new image and the weight matrix for each image in the training set. If the distance is below an empirically determined threshold, the label for that training image is added to a list of candidate matches.

Results

Between the three architectures the Large CNN reached the highest number of accuracies as shown in Table 2. One hundred epochs were run and the average was taken at intervals of 10 epochs. The mean difference between the medium CNN and the small CNN is 0.032%, between the large CNN and medium CNN is 0.0878% and between the small CNN and large CNN is 0.120%. Eigenfaces implementation reached an accuracy of about 96-100%. This was after a number of trials meant to determine an appropriate threshold that would ensure a high amount of accuracy while maintaining a low number of false positives. In the original paper published by Matthew A. Turk and and Alex P. Pentland, it was reported that they reached an accuracy of 100% with a threshold that tolerated a high number of rejections (depending on the levels of preprocessing done on the images).

Discussion

The size of a CNN is very important because it impacts the representations that are used for classification. The three networks implemented had different filter sizes, strides and padding. Generally these parameters are determined through trial and error. In our study, the Large CNN (which had the highest number of accuracies) had smaller filter sizes which were used to determine the number of features that could be detected. Therefore this could suggest that the LFW data contains very fine and minute details which were used in the classification. Another reason for why the Large CNN gives higher accuracies is because it is one layer deeper. This is advantageous because it has an extra level of abstraction to learn features.

Eigenfaces system was impacted by the level of preprocessing. The dataset we used had the most significant facial features aligned. This is the standard for most basic algorithms. However this dataset takes it a step further and has the contrast in each picture normalized to a certain extent. This is very important in the eigenfaces algorithm, specifically in the step where we compute the principal components. Often times it’s the lighting in the images that account for most of the variety computed by PCA.With the lighting accounted for, we can reasonably assume that the significant eigenfaces chosen at the end of the PCA step contain the bulk of variation of the faces themselves. Therefore providing more potent analysis during the facial recognition step.

Conclusion

Going into the experimentation, we hypothesized that the medium CNN would yield the highest testing accuracy(following the Hu et al paper). We found that it was the largest CNN that yielded such results instead. Our interpretation of these results comes down to the representation of the images in the LFW dataset.

Figures

Table 1: CNN Architectures

	CNN Small	CNN Medium	CNN Large
Filter Stride Padding Max-Pooling	12x5x5 1 0 x2	16x5x5 1 0 x2	16x3x3 1 1 -
Filter Stride Padding Max-Pooling	24x4x4 1 0 x2	32x4x4 1 0 x2	16x3x3 1 1 X2
Filter Stride Padding Max-Pooling	32x3x3 2 0 x2	48x3x3 2 0 x2	32x3x3 1 1 x3(st. 2)
Filter Stride Padding Max-Pooling	-	-	48x3x3 1 1 x2
Fully Connected			

Table 2: CNN Results

	Small	Medium	Large
Epoch	Accuracy		
0-9	0.0254	0.0340	0.0426
10-19	0.0344	0.0395	0.0398
20-29	0.0234	0.0262	0.0457
30-39	0.0258	0.0313	0.0414
40-49	0.0305	0.0324	0.0488
50-59	0.0332	0.0336	0.0445
60-69	0.0313	0.0273	0.0434
70-79	0.0309	0.0406	0.0473
80-89	0.0352	0.0336	0.0406
90-99	0.0348	0.0355	0.0496

References

Original Paper on Eigenfaces
https://www.cs.ucsb.edu/~mturk/Papers/mturk-CVPR91.pdf

Basic Concepts (Wikipedia)

https://en.wikipedia.org/wiki/Artificial_neural_network
https://en.wikipedia.org/wiki/Face_Recognition_Grand_Challenge
https://en.wikipedia.org/wiki/Eigenface
https://en.wikipedia.org/wiki/Convolutional_neural_network

History of Face Recognition
https://www.facefirst.com/blog/brief-history-of-face-recognition-software/

When Face Recognition Meets with Deep Learning
https://arxiv.org/pdf/1504.02351.pdf