Facial Recognition Using a Lightweight Deep Neural Networks

Jonathan Hiebert[1], Feezan Mazhar[1], Micahl Derosa[1] and Alaa Sheta[1*]

[1]*Computer Science Department, Southern Connecticut State University, New Haven, CT 06516, USA*
[*]*Corresponding author E-mail: shetaa1@southernct.edu*

Abstract

Current facial recognition systems are still far away from the capability of the human's face perception. Facial recognition systems can continue to be improved as technology evolves. The task of face recognition is heavily explored in recent years. In this research, we provide our initial idea in developing Lightweight Deep Neural Networks for facial recognition. Although our goal was to create an optimal model that would exceed current facial recognition model performance, we could experiment and discover alternative approaches to multi-class facial recognition/classification. We tested with a dataset of 2800 images of men and women with specified image sizes. We created three CNN with various architectures, which we used to train with the chosen dataset for 20, 50, 100, and 200 classes per model. The experimental results exhibit the challenges of increasing the complexity of neural networks. From these results, we concluded that a Light CNN Model with a small number of layers had an average test accuracy of 94.19% which was the best classification performance on unseen data.

Keywords: Facial Recognition; Deep Learning; Convolutional Neural Network.

1. Introduction

The human face is a unique social identifier in our world. The distinctiveness of a human's face is a significant contributor to the concept of individuality. Thus, how we as humans perceive and register human faces has been an intriguing area of focus in psychology research for decades [1]. With the increased performance, accessibility, and affordability of computing technology in the last two decades, the study of human facial recognition has evolved immensely. Today, artificial facial recognition technologies are used worldwide by various organizations and governments.

Facial recognition software has been continuously more reliable due to the exponential increase in accuracy from these systems. For example, FaceMe [2]. Marriott has been developing a check-in software for their hotels that uses facial recognition technology to decrease the average check-in time by up to 66% [3]. Another example of everyday facial recognition utilization involves Caliburger, the California-based fast-food chain. Since 2017, Caliburger has offered a personalized experience with its facial recognition kiosks. These kiosks scan the customer's face, bringing up their user profile. They afford the customer to view past orders and even order their favorite items from the kiosk [4]. Furthermore, many retail conglomerates, such as Walmart, have been implementing systems that can track customers as they move around the store to crack down on shoplifting in their stores [5]. One of the most well-known implementations of facial recognition came in November 2017, when Apple released the iPhone X, which showcased their new "FaceID" feature. FaceID is Apple's facial recognition software that allows users to unlock their phone, make purchases in the app store, use the Apple Pay feature, and much more, all using only their face [6]. At the time of the release of FaceID on the iPhone X, 44.2% of Americans owned an iPhone, and since then, that percentage has grown to nearly 47% [7]. This implementation is responsible for exposing millions of people worldwide to the growing technology of facial recognition.

The availability of large annotated datasets and affordability of GPUs (Graphics Processing Units) are responsible for this rapid improvement in this technology [8]. As of December 2020, the top-ranked facial recognition algorithm produced an accuracy of 99.95%, while only required approximately two seconds to classify each person [9]. The dramatic decrease in the cost of GPUs, paired with the substantial increase in algorithm performance, has opened the door for various industries to utilize facial recognition technology to satisfy their own unique needs. Whether it's for security purposes or simply a method for increasing customer engagement, facial recognition is starting to become a standard in society.

In a 2020 study performed by Grand View Research [10], it was determined that, in terms of end-use, the industries with the most significant percentage of facial recognition market share in 2019 were retail/e-commerce and media/entertainment. Government and healthcare represented the 6th and 7th highest market share, respectively, as depicted in Figure 1. In the report, the focus is on facial recognition usage over the next seven years to 2027. During that forecast period, healthcare is expected to increase its market share the most, while the government is expected to be second in that regard.
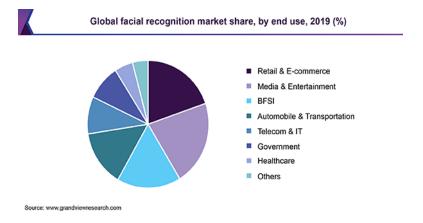

Figure 1: Global Facial Recognition Market Share, 2019 [Source: Grand View Research]

The appeal of facial recognition software for these industries is simple. The ability to identify a person accurately and efficiently is highly beneficial. However, as a wider variety of industries are exploring the benefits of this technology, there have been legitimate concerns about people's freedoms and privacy. In a 2019 survey, 26% of Americans surveyed think the use of facial recognition technology should be strictly limited by the U.S. government [11]. While a small percentage of people voice these concerns, organizations, and governments have consequently been held accountable to implement the technology only in areas of severe security needs (i.e., airport security). Furthermore, facial recognition on unsuspecting citizens has been more heavily criticized than facial recognition systems designed for customer interaction or safety.

This technology is designed to find, recognize, and distinguish between faces within a database. Facial recognition algorithms rely on biometric data — facial characteristics such as hard tissue, curves of the eye socket, nose, and chin. According to research published in April 2020 by the Center for Strategic and International Studies (CSIS), facial recognition systems have nearly absolute precision in ideal conditions, reaching a 99.95% recognition accuracy level. The National Institute of Standard Technology (NIST) states that face recognition has undergone an industrial revolution, with algorithms become increasingly more tolerant of poor-quality images [12]. Even though current machine recognition systems have reached such a high level of maturity, their success is limited by the conditions imposed by many real-life applications. However, perfect conditions are scarcely attainable in daily operations, as algorithms face various obstacles affecting their performance [13]. For example, face images acquired in an outdoor environment with illumination and/or pose changes continue to cause poor performance in even the best facial recognition algorithms.

Figure 2: Map of where facial recognition is expected to grow around the world, 2020-2025 [Source: Mordor Intelligence]

While the United States has witnessed a significant increase in the usage and deployment of facial recognition systems, an enormous change in growth has occurred in East Asia and Australia, shown in Figure 2. Facial recognition is expected to witness robust growth over the next few years due to its advantages over other biometric technologies. Facial recognition is widely preferred over other biometric technologies such as voice recognition, skin texture recognition, iris recognition, and fingerprint scanning due to its contact-less process and easy, cost-effective deployment (using cameras and existing monitoring devices).

1.1. Motivation and Goals

Our motivation for conducting this experiment stems from the fact that facial recognition systems have become increasingly apparent in our daily lives. This project aims to develop Lightweight Deep Neural Networks to detect faces from a dataset with a high level of accuracy and precision. We plan to investigate and test the neural network performance with a growing number of classes and measure the CNN performance in each case. Additionally, we want to determine any correlation between the depth of the network architecture and the level of accuracy achieved on unseen data. This paper is organized as follows. In Section 2, we provide an introduction to the CNN architecture design.

2. What is CNN?

The advancement in machine learning has been exponential in recent years, heavily influenced by the rapid progress of Convolutional Neural Networks. A Convolutional Neural Network (CNN) is a deep learning architecture that can take in an input image, assign importance (i.e., learnable weights and biases) to various objects in a picture, and be able to differentiate one from the other [14].

CNN is one of the principal models of deep learning. Traditional CNN belongs to supervised learning and is a forward pass neural network [15]. A convolutional neural network is a class of deep learning methods that will continue to dominate various computer vision tasks. Recently, CNN was used to handle many problems in face recognition [16, 17, 18].

CNN's are similar to traditional ANNs, with both having a large number of self-optimized neurons based on some learning mechanism. Each neuron receives input(s) from a previous set of neurons and performs several operations (i.e., scalar product and nonlinear function evaluation). The main difference between CNNs and traditional ANNs is that CNNs are mainly used in pattern recognition, whereas ANNs have been extensively used in various real-life problems.

CNN are complex structured models with high computational power. This architecture allows CNNs to perform several complex tasks to map an input to an output to learn complicated and nonlinear functions. Explicitly speaking, CNN consists of a massive number of interconnected nodes tuned by many learning algorithms. CNN nodes are arranged into multiple layers, each layer achieving a definite operation on its input. In Figure 3, we show a block diagram of the proposed face recognition system
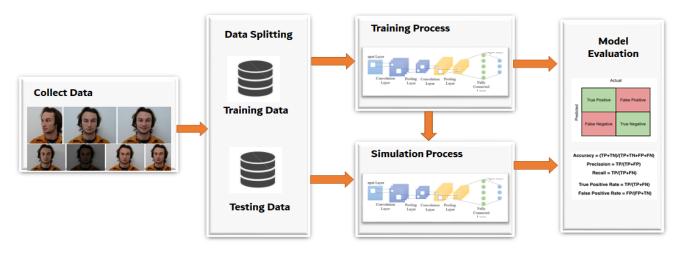

Figure 3: Block diagram of proposed face recognition system

2.1. Convolution layer

CNN's are similar to traditional ANNs, with both having a large number of self-optimized neurons based on some learning mechanism. Each neuron receives input(s) from a previous set of neurons and performs several operations (i.e., scalar product and nonlinear function evaluation). The main difference between CNNs and traditional ANNs is that CNNs are mainly used in pattern recognition, whereas ANNs have been extensively used in various real-life problems [19].

2.2. Pooling layer

Pooling layers usually are one of the layers added after the convolution layer(s). The primary purpose of the pooling layer is to reduce the amount of work that is completed in the next layer of the network. It does this by decreasing the dimensions of the provided input maps [20]—the two main types of pooling layers: max pooling and average pooling.

- Max pooling essentially takes the largest elements in a given feature map and combines them into a new and updated feature map containing the maximum elements.
- Average pooling, on the other hand, takes the average of elements in a region specified by the pooling filter and displays an updated map with the average of elements from each subsection of the original feature map.

By using these two types of layers in neural network architecture, you condense the steps used in the upcoming layers and, in turn, improve your network's performance [20] (See Figure 4).
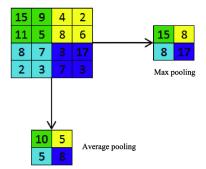
Figure 4: This figure visualizes the differences between max pooling and average pooling

2.3. Activation Functions

An activation function in a neural network is a function that takes the input from each node and connects it with its respective output. When the weighted sums are directed into the nodes, the activation function takes that input sum and generally changes it to a number between a specified upper and lower limit. This transformation that occurs is normally nonlinear. Two major types of activation functions are 1) Sigmoid and 2) ReLU (Rectified Linear Unit) [21].

2.3.1. Sigmoid

Generally speaking, in a sigmoid function, when given a negative input, the process transforms the input number close to 0. When given a positive input, the function converts the input number close to 1. Inputs given and close to 0 will be transformed into random numbers that fall between 0 and 1. The takeaway here is that if the value for the neuron is closer to 0, it becomes less activated, and when closer to 1, it becomes more activated [22].

2.3.2. ReLU

ReLU is one of the most prominent and widely used activation functions in neural networks. Unlike the sigmoid function, this function changes given inputs to a maximum value of 0 or the input value given. The takeaway here is that the more positive the value for the neuron, the more activated it becomes. An example of a famous ReLU function [22] is shown in Equation 1.

$$relu(x) = max(0, x) \tag{1}$$

2.4. Fully-connected Layer

The fully connected layer is traditionally the last layer that is added to the network. The Fully Connected (F.C.) layer takes the output that is received from the convolution, activation, or pooling layer before it and outputs a vector of N number of dimensions. N, in this case, would be the number of classes that the networks have to learn. Each value in the given vector expresses the probability of a specified class. This layer aims to take the output from the previous layer and try to classify an accurate prediction of a particular class. For this process to be possible, the output data from the layer preceding the first F.C. layer must be "flattened." The method of flattening the data before passing it as input to the first F.C. layer is essential, as the F.C. layer can only be one-dimensional [23].

2.5. Evaluation Metrics

Evaluation metrics are beneficial for evaluating a neural network model and its ability to complete its task. We will use some of the most prominent evaluation metrics to assess our model are accuracy, precision, recall, and F1 score. T.P., F.P., TN, and F.N. are the true positive, false positive, true negative, and false-negative, respectively.

- Accuracy evaluation shows the ratio made up the number of predictions the model correctly classifies to the total input samples.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

- Precision evaluates a models ability to find the correct target points that it was designed to find [24].

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

- Recall evaluates the models ability to find the specified cases in the input data set [24].

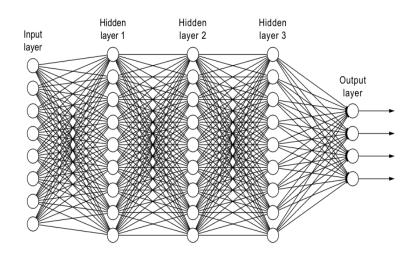$$Recall = \frac{TP}{TP + FN} \tag{4}$$

Figure 5: Fully Connected Neural Networks

- F1 score, also known as F-measure or F-score, shows the weighted average of the precision and recall values. In an F-score, the best value that can be obtained is one, and the worst being a 0 [25].

$$F1 = 2 * \frac{(precision * recall)}{(precision + recall)} \tag{5}$$

3. Dataset

For a multi-class classification model to accurately identify specific faces, many images are needed for each class. We will be using the FEI face database for our research approach. This dataset was compiled in 2006 by researchers from the University Center of FEI in Brazil. The data consists of 2800 images of Brazilian men and women, each with 14 images. There is a 1:1 ratio of men to women, within the age range of 19 to 40, featuring a wide variety of appearances and hairstyles. The images are taken against a homogeneous background in an upright frontal position with profile rotations of approximately 180 degrees. The size of the original images is 640×480 pixels.
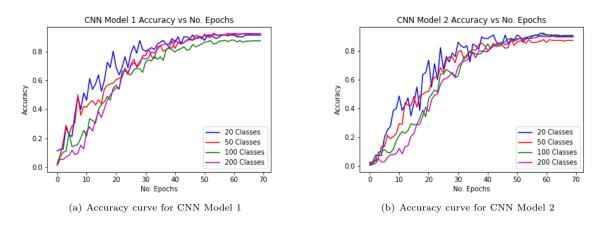

Figure 6: Some examples of image variations from the FEI face database [26]
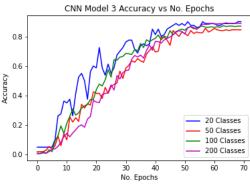
4. Experimental Results

The data set was split into 1600, 800, and 400, training, testing, and validation sets to develop the proposed set of CNN models. Our proposed approach for the facial recognition task was to create an appropriate model to classify a human face accurately. We implemented a pre-processing stage for the dataset, which includes:

- We augmented the dataset by adjusting the dimensions of the images to 480×480 pixels, and
- Applying FastAI's default data augmentation function, which randomly augments the data in various ways. Examples of such augmentations include horizontally flipping the image, warping the image, and adjusting the illumination of the image.

We then created three CNN architectures with varying complexity. The proposed models CNN consisted of one, two, and three convolutional layers, respectively. Thus, our goal is to test the CNN network's performance with the growing number of

(a) Accuracy curve for CNN Model 1

(b) Accuracy curve for CNN Model 2

(c) Accuracy curve for CNN Model 3

Figure 7: Accuracy Curves for the three proposed CNN architectures

classes. We tested each of our models on 20, 50, 100, and 200 classes (i.e., images) to achieve this. The model was trained for 70 epochs for each of these, with a fixed learning rate of 1×10^{-5}. As depicted in Table 1, our lowest test accuracy was 87.50%, which was calculated for CNN Model 2, when trained on only the first 50 classes. The highest accuracy recorded on unseen test data came from CNN Model 1 when trained on the first 100 classes.
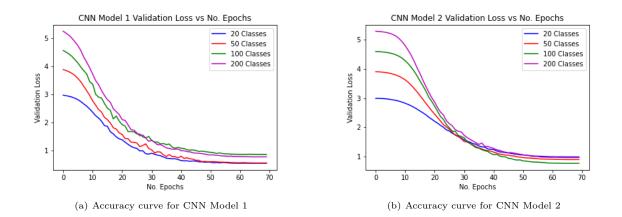
Table 1: Proposed CNN models for facial recognition

Model	No. of Classes	Training				Testing
		Accuracy%	Precision%	Recall%	Time (mm:ss)	Accuracy%
CNN Model 1	20	91.25	95.22	91.25	09:19	95.00
	50	92.00	93.87	92.00	22:10	94.00
	100	87.25	90.04	87.25	43:39	96.00
	200	92.13	94.42	92.13	89:07	91.75
CNN Model 2	20	90.00	89.08	90.00	09:19	87.50
	50	87.50	91.24	87.50	22:15	95.00
	100	91.00	92.83	91.00	51:29	93.00
	200	90.50	93.75	90.50	101:06	91.00
CNN Model 3	20	90.00	95.00	90.00	09:13	90.00
	50	84.50	89.45	84.50	22:45	91.00
	100	87.00	90.96	87.00	45:11	91.00
	200	88.88	91.58	88.88	89:40	90.25

In Figures 7(a), 7(b) and 7(c), we can see that the accuracy curves for each CNN Model were extremely similar. In each model, we can see that the model had a generally higher validation accuracy when trained on the fewest number of classes in our experiment (i.e., twenty classes). As shown in Figures 8(a), 8(b) and 8(c), the validation loss for CNN Model 1 was significantly less than the other two models, no matter the number of classes are used.

5. Conclusions

In this research, we developed several CNN architectures to handle the problem of facial recognition. Three CNN models with various complexity were explored. We determined that a Lightweight CNN with a single convolutional layer can achieve high recognition accuracy. The average test accuracy for CMM model 1 was 94.19%, which was a significant improvement over

(a) Accuracy curve for CNN Model 1

(b) Accuracy curve for CNN Model 2

(c) Accuracy curve for CNN Model 3

Figure 8: Loss Curves for the three proposed CNN architectures

CNN models 2 and 3, which had average test accuracies of 91.63% and 90.56%, respectively. We trained the deepest model on CNN Model 3, which had the worst performance on unseen data (i.e., test accuracy). Per these results, we concluded that given our experimental setup and the dataset used, there is a negative correlation between the complexity of the network architecture and the network's performance on unseen test data.

References

[1] G. H. Bower and M. B. Karlin, "Depth of processing pictures of faces and recognition memory," *Journal of Experimental Psychology*, vol. 103, pp. 751–757, 1974. [Online]. Available: https://doi.org/10.1037/h0037190

[2] H. Xu, X. Su, Y. Wang, H. Cai, K. Cui, and X. Chen, "Automatic bridge crack detection using a convolutional neural network," *Applied Sciences*, vol. 9, no. 14, p. 2867, 2019.

[3] E. Florio. (2018) At marriott, you can now check in with your face. [Online]. Available: https://www.cntraveler.com/story/marriott-alibaba-facial-recognition-hotel-check-in

[4] C. Clifford. (2018) You can pay for your burger with your face at this fast food restaurant, thanks to a.i. [Online]. Available: https://www.cnbc.com/2018/02/02/pay-with-facial-recognition-a-i-at-caliburger-in-pasadena-california.html

[5] G. Barkho. (2019) Walmart confirms use of ai-powered cameras to detect stealing. [Online]. Available: https://observer.com/2019/06/walmart-ai-cameras-detect-stealing/

[6] Apple Press Release. (2017) The future is here: iphone x. [Online]. Available: https://www.apple.com/newsroom/2017/09/the-future-is-here-iphone-x/

[7] S. O'Dea. (2021) Share of smartphone users that use an apple iphone in the united states from 2014 to 2021. [Online]. Available: https://www.statista.com/statistics/236550/percentage-of-us-population-that-own-a-iphone-smartphone/

[8] J.-C. Chen, R. Ranjan, S. Sankaranarayanan, A. Kumar, C.-H. Chen, V. M. Patel, C. D. Castillo, and R. Chellappa, "Unconstrained still/video-based face verification with deep convolutional neural networks," 2017.

[9] D. S. Abdelminaam, A. M. Almansori, M. Taha, and E. Badr, "A deep facial recognition system using computational intelligent algorithms," *PLOS ONE*, vol. 15, no. 12.

[10] Grand View Research. (2020) Facial recognition market size, share trends analysis report by technology (2d, 3d), by application (emotion recognition, attendance tracking monitoring), by end-use, and segment forecasts, 2020 - 2027. [Online]. Available: https://www.grandviewresearch.com/industry-analysis/facial-recognition-market

[11] D. Castro and M. McLaughlin, "Survey: Few americans want government to limit use of facial recognition technology, particularly for public safety or airport screening," 2019.

[12] P. J. Grother, P. J. Grother, and M. Ngan, *Face recognition vendor test (frvt).* US Department of Commerce, National Institute of Standards and Technology, 2014.

[13] W. Crumpler, "How accurate are facial recognition systems—and why does it matter," *Center for Strategic and International Studies*, vol. 14, 2020.

[14] S. Saha, "A comprehensive guide to convolutional neural networks-the eli5 way," Dec 2018. [Online]. Available: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

[15] W. You, C. Shen, X. Guo, X. Jiang, J. Shi, and Z. Zhu, "A hybrid technique based on convolutional neural network and support vector regression for intelligent diagnosis of rotating machinery," *Advances in Mechanical Engineering*, vol. 9, no. 6, p. 1687814017704146, 2017.

[16] G. Guo and N. Zhang, "A survey on deep learning based face recognition," *Computer Vision and Image Understanding*, vol. 189, p. 102805, 2019.

[17] S. Almabdy and L. Elrefaei, "Deep convolutional neural network-based approaches for face recognition," *Applied Sciences*, vol. 9, no. 20, 2019.

[18] M. Wang and W. Deng, "Deep face recognition: A survey," *Neurocomputing*, vol. 429, p. 215–244, Mar 2021. [Online]. Available: http://dx.doi.org/10.1016/j.neucom.2020.10.081

[19] V. Sharma and P. C. says:, "Role of convolutional layer in convolutional neural networks," Oct 2020. [Online]. Available: https://vinodsblog.com/2020/05/03/role-of-convolutional-layer-in-cnn/#:~:text=ConvolutionalLayer\T1\textendashAnOutlook,i.e.extractfeaturesfromit.

[20] S. Khosla. (2019) Cnn | introduction to pooling layer. [Online]. Available: https://www.geeksforgeeks.org/cnn-introduction-to-pooling-layer/

[21] Deeplizard. (2017) Accuracy, precision, recall or f1? [Online]. Available: https://deeplizard.com/learn/video/m0pIlLfpXWE

[22] S. Kansal. (2017) A quick guide to activation functions in deep learning. [Online]. Available: https://towardsdatascience.com/a-quick-guide-to-activation-functions-in-deep-learning-4042e7addd5b

[23] A. Deshpande, "A beginner's guide to understanding convolutional neural networks. 2016," *URl: adeshpande3. github. io/adeshpande3. github. io/A-Beginner's-Guide-To-Understanding-Convolutional-Neural-Networks*, 2020.

[24] K. P. Shung. (2018) Accuracy, precision, recall or f1? [Online]. Available: https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9

[25] W. Koehrsen. (2018) Beyond accuracy: Precision and recall. [Online]. Available: https://towardsdatascience.com/beyond-accuracy-precision-and-recall-3da06bea9f6c

[26] C. E. Thomaz, "Fei face database." [Online]. Available: https://fei.edu.br/~cet/facedatabase.html