Face Verification

University of Piraeus - NCSR Demokritos
Deep Learning

Stamatios Orfanos - mtn2211
stamatisorfanos99@gmail.com

Abstract

This paper presents a comprehensive study on face verification, a crucial task in biometric authentication systems. Face verification aims to determine whether two face images belong to the same individual or not, and it plays a significant role in various applications, including access control, surveillance, and identity verification. In this work, we propose a novel approach that leverages deep learning techniques for robust and accurate face verification. Our method incorporates a convolutional neural network (CNN) architecture trained on a large-scale face dataset to learn discriminative features and a similarity metric for face matching. The technique achieves exceptional performance in terms of accuracy and computational efficiency, making it suitable for real-world face verification applications.

1 Introduction

In today's digital era, secure and reliable authentication systems are of paramount importance for safeguarding sensitive information and ensuring the integrity of various applications. Biometric authentication, which relies on unique physical or behavioural characteristics of individuals, has emerged as a popular approach due to its inherent advantages of convenience and uniqueness. Among various biometric modalities, face recognition has garnered significant attention due to its non-intrusive nature and widespread availability of face imaging devices. Face verification, a critical component of face recognition systems, involves determining whether two face images belong to the same person or not.

The problem of face verification is inherently challenging due to variations in pose, illumination, occlusions, background environment and facial expressions. These factors can significantly impact the performance of traditional image processing and feature extraction techniques. Consequently, there is a growing need to develop robust and accurate face verification methods that can handle such challenges and provide reliable authentication in real-world scenarios.

In this paper, our objective is to implement the Siamese model, a powerful deep learning architecture, for face verification. The Siamese model has shown promising results in learning discriminative features and capturing similarities between face images. By leveraging this model, we aim to improve the accuracy and robustness of face verification systems.

2 Methodology

A rough outline of the methodology used to develop the face verification system includes the creation of a Siamese model. While training we have three data-sets that include the positive, anchor and negative images. Each of those data-sets is used to help the model learn the discriminative features and capturing similarities between face images.

2.1 Data

Initially the data set we are going to be using in this paper comes from the University of Massachusetts, where the data-set Labelled Faces in the Wild is found. This data set contains folders for over 5749 people and 13233 images in total. These images have a size of $250x250$ and are going to be used as the negative images to compare to our images.

For each person with multiple images we usually totally different images. The differences may be in the age of the person, the background, lighting of the space, angle of the images and many other parameters. An instance would be for the Prime Minister Tony Blair, where we can see differences in colour of hair, background, light exposure and many more.

Figure 1: Prime Minster Tony Blair

In this project we are also going to create our own data-set of images using a tool from the library OpenCv to capture and store images of ourself in order to create the positive and anchor datasets. The anchor data-set is going to include all the baseline images we are going to be compare both the positive and negative examples. In a similar manner the positive images folder is going to contain positive examples that should be verified through our model. The important part of this procedure is to create and implement an data augmentation function in order to create enough data variety to create an effective model.

The steps to achieve the data augmentation we took the following steps:

- Add random **brightness** to the image.
- Add random **contrast** to the image.
- Add random **saturation** to the image.
- **Rotate** the image random image.

For each one of the augmentation steps above we create nine images for each step. An example of this procedure is the following:

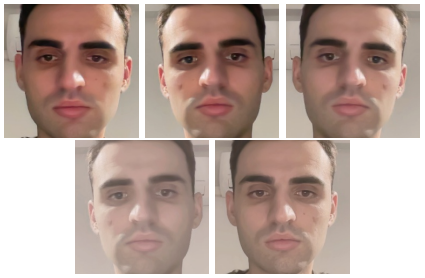

Figure 2: Data Augmentation Result

In this case we are going to use 5000 examples for each of the classes anchor, negative and positive in order to train, test and evaluate the model.

2.2 Siamese Model Overview

As we mentioned in the introduction the model we are going to be training and using for the verification image is the Siamese Model. An outline of the model would include the main layers that the model uses in order to find similarity between images.

For this project, we employ large siamese convolutional neural networks which are capable of learning generic image features useful for making predictions about unknown class distributions even when very few examples from these new distributions are available. Also it is important to be trained using standard optimisation techniques on pairs sampled from the source data.

To develop a model for one-shot image classification, we aim to first learn a neural network that can discriminate between the class-identity of image pairs, which is the standard verification task for image recognition. A simple 2 hidden layer siamese network for binary classification with logistic prediction p. The structure of the network is replicated across the top and bottom sections to form twin networks, with shared weight matrices at each layer.

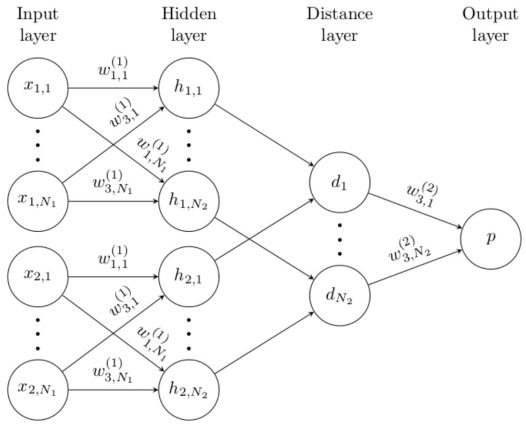

Figure 3: Main architecture of Siamese Model

3 Deep Siamese Networks

Siamese nets were first introduced in the early 1990s by Bromley and LeCun to solve signature verification as an image matching problem (Bromley et al., 1993). A Siamese neural network consists of twin networks which accept distinct inputs but are joined by an energy function at the top. This function computes some metric between the highest level feature representation on each side 3. The parameters between the twin networks are tied. Weight tying guarantees that two extremely similar images could not possibly be mapped by their respective networks to very different locations in feature space because each network computes the same function. Also, the network is symmetric, so that whenever we present two distinct images to the twin networks, the top conjoining layer will compute the same metric as if we were to we present the same two images but to the opposite twins.

The model consists of a sequence of convolutional layers, each of which uses a single channel with filters of varying size and a fixed stride of 1. The number of convolutional filters is specified as a multiple of 16 to optimise performance. The network applies a ReLU activation function to the output feature maps, optionally followed by max-pooling with a filter size and stride of 2. Thus the k_th filter map in each layer takes the following form:

$$a^k_{1,m} = maxpool(max(0, W^k_{l-1,l} * h_{1,(l-1)} + b_l), 2)$$

$$a^k_{2,m} = maxpool(max(0, W^k_{l-1,l} * h_{2,(l-1)} + b_l), 2)$$

,where W_{l-1}, l is the 3-dimensional tensor representing the feature maps for layer l and we have taken $*$ to be the valid convolutional operation corresponding to returning only those output units which were the result of complete overlap between each convolutional filter and the input feature maps. The units in the final convolutional layer are flattened into a single vector. This convolutional layer is followed by a fully-connected layer, and then one more layer computing the induced distance metric between each siamese twin, which is given to a single sigmoidal output unit.

More precisely, the prediction vector is given as $p = \sigma(\Sigma_j \alpha_j |h^j_{1,L-1} - h^j_{2,L-1}|)$, where σ is the sigmoidal activation function. This final layer induces a metric on the learned feature space of the $(L-1)th$ hidden layer and scores the similarity between the two feature vectors. The α_j are additional parameters that are learned by the model during training, weighting the importance of the component-wise distance. This defines a final Lth fully-connected layer for the network which joins the two Siamese twins. We depict one example below 4, which shows the largest version of our model that we considered. This network also gave the best result for any network on the verification task.

4 Training

In one-shot Face Verification, the most important metric in most cases, precision is considered the more crucial metric in face verification systems. Precision, also known as positive predictive value, measures the proportion of correctly identified positive instances (i.e., genuine matches) out of all instances classified as positive by the system. In the context of face verification, precision indicates how reliable the system is in correctly identifying whether two faces belong to the same individual or not. High precision implies that the system has a low false positive rate, reducing the chances of incorrectly verifying an imposter as a genuine match.

While recall, also known as sensitivity or true positive rate, is important as it measures the proportion of actual positive instances that are correctly identified by the system, it may not be the primary concern in face verification. This is because the consequences of a false positive (verifying an imposter as genuine) can be more severe than a false negative (rejecting a genuine match). In applications where security or access control is the primary objective, maintaining a high precision is crucial to prevent unauthorised access.

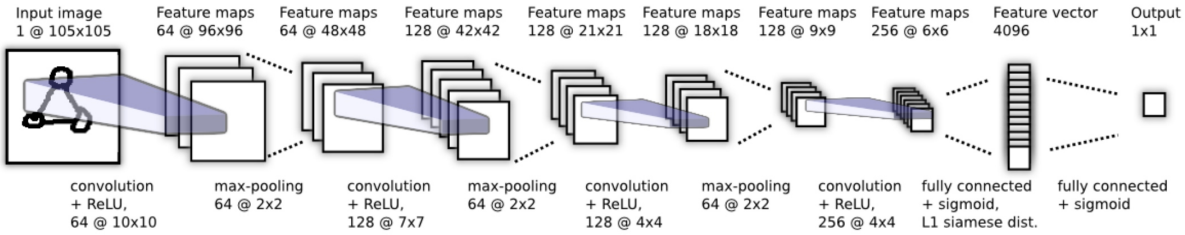

Figure 4: Siamese Model

The training process for a one-shot verification system that utilises a deep Siamese model requires us to define certain parameters. Given that the data are images we have to create a more custom training process and to be more specific we have the following:

- Epochs: 50
- Batch size: 256
- Pre-fetch size: 64
- Learning rate: 0.001

Batch size and pre-fetch size are crucial parameters to consider when training a model on image data. Choosing an appropriate batch size was quite challeging since it is important for balancing computational efficiency and model performance. A larger batch size can make better use of parallel processing capabilities, enabling faster training. However, excessively large batch sizes can lead to GPU memory limitations and hinder convergence. Conversely, smaller batch sizes may provide better generalisation and finer parameter updates, but training can become slower. On the other hand, the prefetch size determines the number of batches that are loaded into memory ahead of time, anticipating the GPU's need for the next batch. Given the amount of data and complexity of the model that requires the **training of 38,964,545 total trainable parameters** we had to include a pre-fetch process in order to avoid wasting time, while training the model

For the demo application we created a custom verification function, where for the input image we use the Siamese model and compare it to fifty positive images we created in advance. For each image we check whether the model prediction covers the model threshold and given that the majority of those images get correct predictions then we are able to verify the face for the user.

4.1 Recall-Precision in Training

The training process of a Siamese model for one-shot face verification with image data is a complex and time-consuming task due to several factors. Firstly, the Siamese architecture itself requires extensive computation and memory resources. This parallel processing, combined with the need to calculate similarity distances, results in a significant computational load. Moreover, training a Siamese model for one-shot face verification involves working with a large amount of data. Collecting and preparing a diverse data-set of face images, including positive and negative pairs, requires substantial effort and meticulous annotation. The dataset needs to cover various lighting conditions, poses, and facial expressions to ensure the model's generalisation capability.

The precision of the Siamese model for each of the fifty epochs is the following:

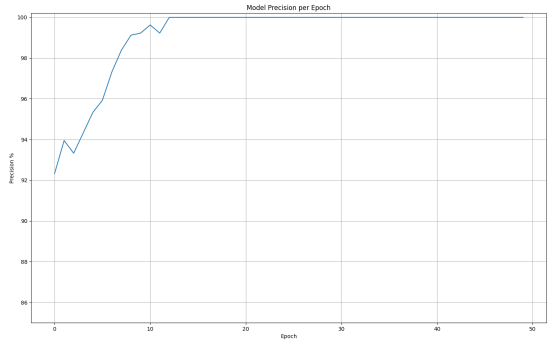

Figure 5: Siamese Model Precision

The recall of the Siamese model for each of the fifty epochs is the following:

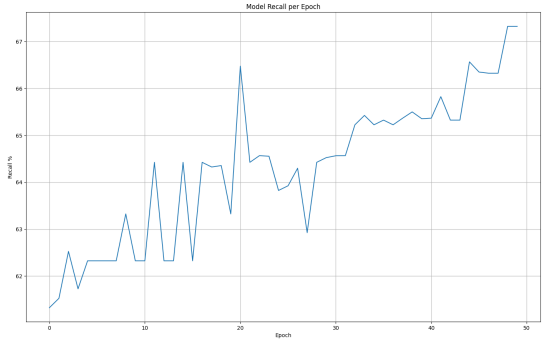

Figure 6: Siamese Model Recall

In one-shot face verification, precision is often more critical than recall. This is because the consequences of a false positive, verifying an imposter as a genuine match, can be more severe in scenarios where security or access control is the primary objective. A high precision implies that the model is effective at accurately distinguishing between genuine and imposter matches, reducing the risk of unauthorized access.

5 Conclusion

In conclusion, the project focused on developing a Siamese model for one-shot face verification, aiming to accurately identify whether two face images belong to the same individual. The training process involved a meticulous data-set preparation, incorporating positive and negative pairs of face images with diverse variations in lighting, poses, and expressions.

The Siamese architecture, with its parallel processing and similarity distance calculations, required significant computational resources. The training process was time-consuming due to the complexities involved, including the size of the data-set, multiple training iterations, hyperparameter tuning, and the incorporation of techniques such as data augmentation and regularization. The evaluation of the model's performance considered both precision and recall metrics, with precision being particularly important in minimizing false positives for security and access control. However, the optimal balance between precision and recall depended on the specific requirements of the application.

Overall, the project provided insights into the challenges and considerations involved in developing a robust one-shot face verification system, and it highlighted the importance of carefully aligning the system's objectives with the chosen evaluation metrics.

References

University of Massachusetts

Labelled Faces in the Wild

OpenCv

Siamese Model

(Bromley et al., 1993)