Siamese Network One Shot Image Recognition: My Creative Approach to Address Common Issues

LaTeX template adapted from:
European Conference on Artificial Intelligence

Fayzullo Bakhtiyorov[1]

Other group members:
Furkan Boran[2], **Jenson Allaway**[3], **Derrick Asare**[4]

Abstract. To begin with, this project evolves practical approach for face recognition, focusing on the designing and deploying of the Siamese Neural Network using TensorFlow and OpenCV tools. To exemplify this point, traditional image recognition models such as Triplet Networks require the large volume of the labelled dataset needed for training in terms of the memory usage. This might be reflected in inefficient computational performance.[8] To address this limitation, the proposed project achieves high accuracy with minimised amount of training data.

To illustrate the usage of the datasets, the negative dataset has been collected from the LFW, that stands for the Labelled Faces in the Wild. The positive and anchor images have been collected using the images of myself through the webcam of my laptop. The access to the webcam has been granted using OpenCV library.

The Siamese Neural Network, which consists of an embedding network layers and a custom Distance Layer, that performs the accurate face recognition, have been made using TensorFlow, the powerful platform for developing machine learning projects. The model is trained using binary cross-entropy loss and the Adam optimizer, with periodic checkpoints saved for reusability. [8]

To highlight the verification functionality, the real-time verification runs through comparing the anchor (input) image with pre-trained images from positive and negative datasets.[7]

The trained model has been saved for the purpose of usage in the future. It could be reloaded for the further training using advanced features or performing verification.

To sum up, the given project encounters potential minor errors in the phase of image pre-processing due to the minimum pair of positive samples, that might be considered as one of the primary disadvantages of the project deployment. [8]

1 Introduction

In the contemporary society of smart technologies, distinguishing the person authenticity in real-time has become an overwhelming challenge. The given project addresses this challenge by performing a real-time face verification through designing and deploying Siamese Neural Networks for One-shot Image Recognition. This project could be integrated as crucial feature of computer vision in security, user authentication and access control for enhancing advanced safety measures.

To begin with understanding of the problem domain, the environment with various lightning conditions, such as high and low-light settings for the image capture. [10] Furthermore, the accurate verification of image despite the change in terms of the facial expressions. Finally, the capability to store personal biometric data securely mitigating potential exposures could be considered as the main three features of the problem domain encountering in face recognition systems. [4]The primary advantage of Siamese Neural Network in contrast with its counterparts, could be considered in adaptability. For instance, the trained model performs verification with different facial expressions in different lightning conditions by tackling the various features of the problem domain.[7] In order to address these limitations comprehensively, I have opted for the appropriate algorithms and designed a custom layer to perform an accurate distinguishing the difference between the anchor and positive images. To address the problems associated with image verification models, the Binary Cross Entropy Loss Function and Adam Optimisers have been used in training phase of the project. To exemplify this point extensively, on the one hand, the Adam Optimiser, that stands for Adaptive Moment Estimation, the optimisation algorithm used in training of neural network considering its advantage to tackle the domain issues and efficient memory usage for large datasets as well as required low computational capacity.[5] On the other hand, the Binary Cross Entropy Loss function produces accurate results between the anchor image and positive labelled samples.

2 Background

Moving forward to the background of the research into neural networks for image recognition, this section consists of the analysis of existing algorithms used to deal with the problem along with the discussion about relevant papers and their summary. To begin with analysis of algorithms, two alternative options have been considered. On the one hand, the Viola-Jones algorithm, considered as traditional face detection methodology, introduced in 2001, the algorithm heavily relies on the Haar-features, rectangular patterns used for describ-

[1] School of Computing and Mathematical Sciences, University of Greenwich, London SE10 9LS, UK, email: fb2551g@gre.ac.uk
[2] School of Computing and Mathematical Sciences, University of Greenwich, London SE10 9LS, UK, email: fb7397e@gre.ac.uk
[3] School of Computing and Mathematical Sciences, University of Greenwich, London SE10 9LS, UK, email: ja3645c@gre.ac.uk
[4] School of Computing and Mathematical Sciences, University of Greenwich, London SE10 9LS, UK, email: da3210@gre.ac.uk

ing objects visually. The advantage of the given algorithm is achieving high computation accuracy. However, the sensitivity to changes of the capture environment in terms of lightning conditions and posing are the disadvantages of the algorithm. [2] On the other hand, the Histogram of Oriented Gradients, the algorithm, which encountered similar disadvantages along with its counterpart. However, the advantage of the algorithm is scalability, that might be observed in detecting objects of different sizes in the image. [1] The disadvantage, where both algorithms intersect, is the necessity of large datasets for training phase.[3] What is more, to summarise the quality of the resource materials, all of them are derived from accredited educational institutions websites. For instance, the architecture of Siamese Neural Network has been taken from the article issued by the group of academic personnel of the Computer Science Department of University of Toronto, Ontario, Canada.[7] What is more, as the potential feature of enhancing the of the project as gender detection, the following Real-Time Gender Detection in the Wild Using Deep Neural Networks has been carefully read and analysed. Furthermore, all the features proposed in the articles given above have been approved during the development and deployment of the project.[7] To sum up, in order to maintain the computational performance of the project, the alternative algorithms and the set of additional features have not been implemented.

3 Experiments and results

This section will describe the experiments conducted using 10 negative and positive pairs, along with the visual outputs. The experiment itself conducted taking 10 positive and negative labelled samples from datasets. The Keras optimiser and binary cross loss entropy used to perform the distinguishing the difference between two images.[7] Importing the Precision and Recall metrics from TensorFlow Keras Optimiser assisted for evaluating the performance of classification models. It is making predictions using the trained siamese model. The predict method of the siamese model generates predictions for pairs of input images (test input and test val). The siamese model takes two input images (anchor and positive/negative) and outputs a probability score indicating the similarity between the input images.[7] The variable y hat stores the predicted output. In the context of the siamese network, these predictions are typically binary values or probability scores representing the model's confidence in the similarity of the input pairs. Visualising two images side by side from the test dataset. The first subplot displays the image from the test input dataset, and the second subplot displays the image from the test val dataset. The figure size is set to be larger for better visibility. In order to demonstrate the results visually, two plots were set to perform this task. According to the visual representation of the outputs below, it might be clearly observed that in the first example, the images under the index 0 represent mismatch of the image pairs, which indeed inarguable. However, according to the example derived from another batch, image pairs under the index 0, supposed to be matching. The both images are the facial pictures of myself.

```
In [91]: # Post processing the results
         [1 if prediction > 0.5 else 0 for prediction in y_hat ]
Out[91]: [0, 0, 1, 0, 1, 1]

In [92]: y_true
Out[92]: array([0., 0., 1., 0., 1., 1.], dtype=float32)
```

Figure 1. The representation of results in context of arrays

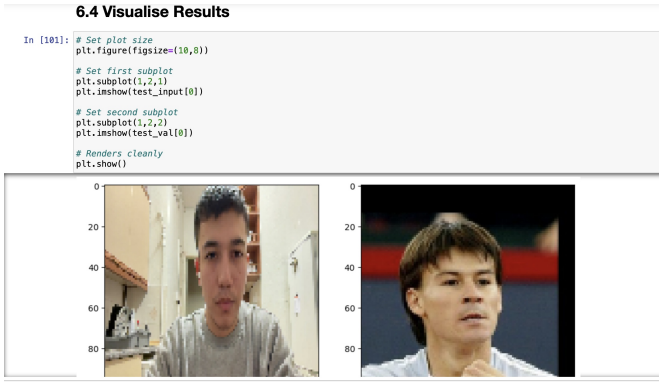

Figure 2. The visual representation of results - unverified

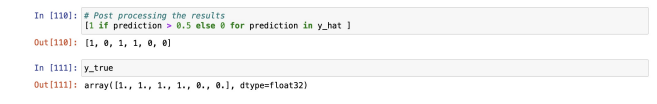

Figure 3. The representation of results in context of arrays

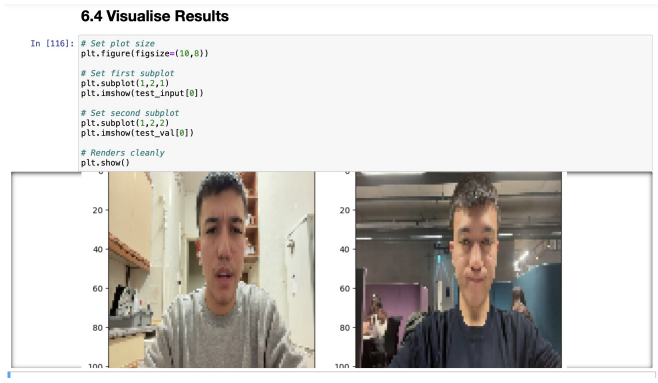

Figure 4. The visual representation of results - verified

To highlight the critical evaluation of the time taken for the model training, the chart showcasing the time taken for the training of the model with different input sizes might be found appropriate.

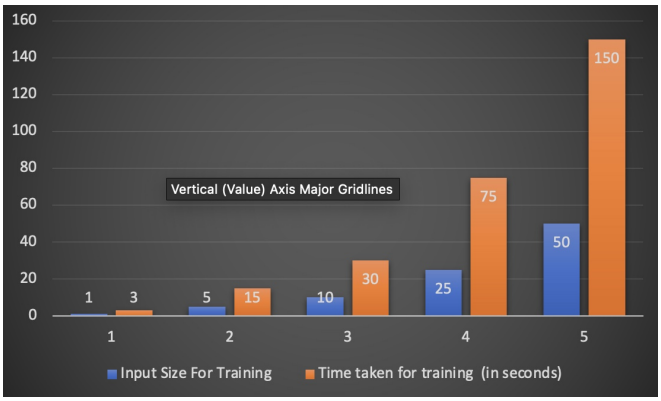

Figure 5. The analysis of time taken for the model training with different input sizes.

4 Discussion

This section includes the technical description of the architecture of Siamese Neural Network and its parameters and their usage in various scenarios. The quality and relevance of the results analysis discussed along with ethical issues of the Siamese Neural Network are considered. The crucial parameters of Siamese Neural Network involved in designing the embedding layers include input shape, the number of convolutional layers, max pooling sizes, activation functions will extensively be analysed in this section. To begin with the architecture of Siamese Neural Network, it is remarkable to mention that it consists of three main parts: the Embedding Network, Siamese L1 Distance Layer and Classification Layer.[7] The Embedding Network includes multiple convolutional layers within embedding network, max-pooling layer, and a dense layer. [7]

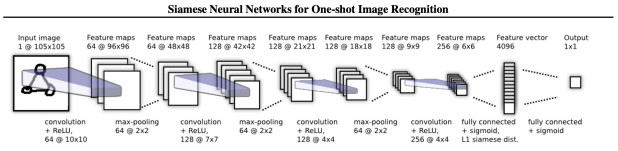

Figure 6. Siamese Neural Network Architecture

To exemplify the Embedding Network, the Input Layer is made of the shape of 105x105px and 3 channels. The channels in our case describe that anchor image will consist of red, green, and blue colour channels.[7] Furthermore, the convolution layer takes the number of filters passing through. To be more specific, it consists of 64 filters with 10x10 shape.[7] The ReLU activation function involved within the layers of Embedding Network, technically stands for rectified linear unit. It applies to hidden layers of the neural network and returns the input value if it is positive and zero otherwise.[7] The max-pooling layer keeps only the brightest pixel with high intensity and ignores the others. To be more exact, within the Embedding Network, it shrinks the picture size by focusing on the most important parts.[7] The flattening stands for arranging data from a block to a simple straight line. The main purpose behind this is optimisation in terms of the time complexity. In this phase, flattening facilitates data access to a computer before getting ready to make predictions.[7] Sigmoid is an activation function that assists in making decisions and produces an output between 0 and 1.[7]

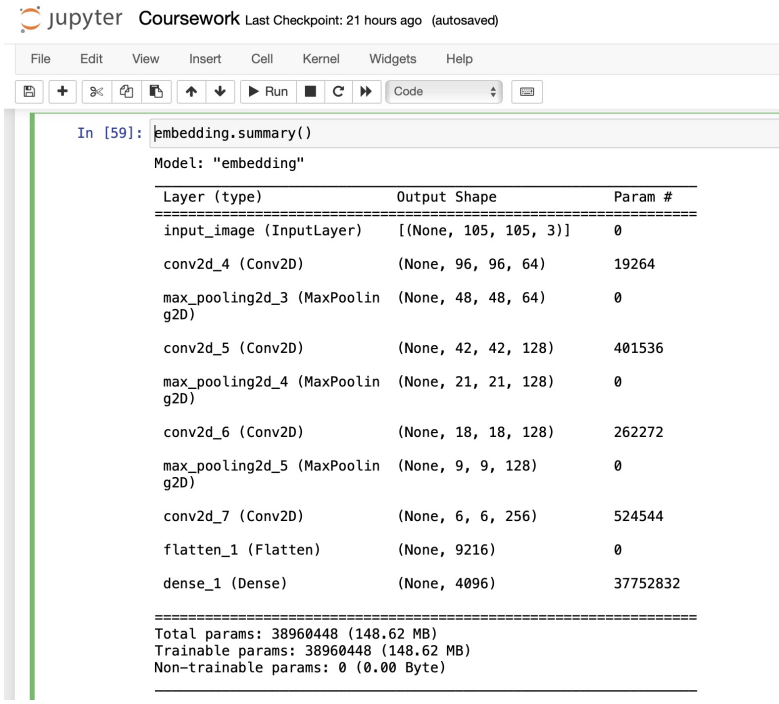

Figure 7. The practical implementation of Siamese Neural Network

The dense layer is considered to be the final layer within embedding network. There are 4096 units in the dense layer transforming the flattened data into a lower-dimensional space. [7]To maintain the efficiency of the computational performance, the shapes of layers are designed according to the article about Siamese Network Architecture.[7] During the testing under the various scenarios, the attempt failed due to the fact, that Siamese Neural Network is a set of fully connected layers, the attempt of changing the shape of one layer will be resulted in breaking the pattern between them. To exemplify the performance of the embedding layer in terms of processing data to the L1 Custom Distance layer, the following scenario could be found appropriate. The anchor and positive images will be processing through the pipelines and put together in fully connected L1 Siamese Distance Layer. [7]

Furthermore, the custom Distance layer computes the difference between two input images of Embedding Network.[7] What is more, the final dense layer, that holds sigmoid activation function, computes the value between 0 and 1, showcasing the likelihood of the images, arguing whether they belong to the same class or different classes.[7] To critically evaluate the quality of the results, the following chart can be considered as the showcase of the quality of results with different input sizes. According to the chart, given above it might be clearly observed, that per one input size, the image of myself is used in training the model. Furthermore, the achieved high accuracy of the computations credits the efficiency of Siamese Model Network in terms of being trained with minimum pairs and involved algorithms for the accurate distinguishing the difference between the input and verification images.

The following privacy concerns encountered within Siamese Neu-

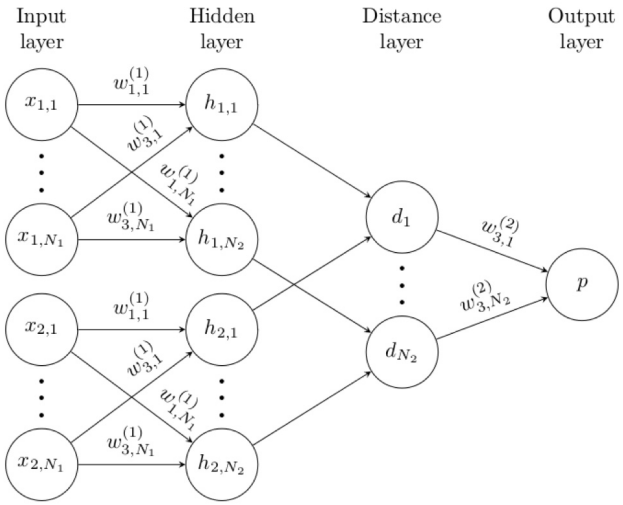

Figure 8. The scenario of processing of data through embedding layer to L1 Distance layer

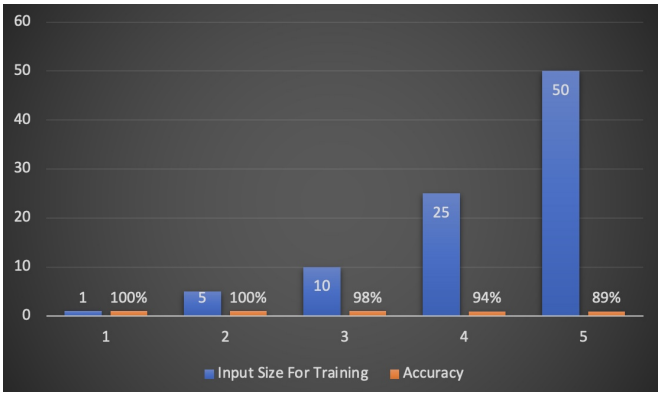

Figure 9. The critical evaluation of the quality of results with different input sizes

ral Network must be considered to address. On the one hand, considering the fact, that real-world application of this project itself arises the privacy concern as it uses images of people for training.[4] Storing and training of the model using facial images must be implemented in compliance with data protection regulations and guidelines to prevent unauthorised access. The most appropriate law in this particular case seems to be considered GDPR which stands for the General Data Protection Regulation.[9] On the other hand, the Siamese Neural Network, like other neural networks, might be vulnerable to being exposed through performing adversarial attacks, where the sensitive information, in our case the input images, may lead unauthorised access.[6] To mitigate the privacy issues and avoid the potential adversarial attacks to occur, the anonymisation of the data, through implementing end-to-end encryption technique along with other methodologies of data encryption must implemented prior to the designing the model.[6] Furthermore, to keep the system safe, regular assessing procedures must be implemented.

5 Conclusion and future work

To conclude, the remarkable achievement that could be considered as the fact of successful deployment of the project is the trained Siamese Neural Network. The given model distinguishes the similarity between the minimum pairs of input image and the pre-trained images with high accuracy and relatively low-computational capacity. To summarise the challenges I have personally met, it is the strict teaching schedule and the assignment submission deadlines. Furthermore, considering the difficulty of the project, I had to devote time beyond lab classes for debugging the project errors. To illustrate the limitation, the project uses a fixed verification threshold of 0.5 for determining whether pairs are genuine or impostor. Optimizing or adapting the threshold based on specific use cases or datasets could enhance the system's adaptability and robustness.

The critical review and evaluation of the projects from the perspective of its strengths and weaknesses, proposing including set of features clearly demonstrates the commitment for the further improvement in terms of the technical performance. To illustrate the one of the primary strengths of the project, which might be observed in the integration of the computer vision techniques from OpenCV library, and deep learning paradigms from TensorFlow. What is more, the real-time verification using the webcam of the device, avoiding involvement of other devices for image capture, makes the project applicable to a range of real-world scenarios. To exemplify this advantage, the following approach could be used for the attendance recording in our university. To illustrate the scenario, exposing the weakness of the current attendance system, the student can register the attendance of absent students using their gateway cards. However, the deployment of the project will mitigate the possible misconduct as it will capture the student's image and compare with their ID number and pre-trained positive samples. The key advantage of the algorithms used for training could be seen in the low computational capacity. To describe the technical implementation, the most up-to-date tools have been used in the project implementation. The key finding of the project is producing accurate results with the minimum amount of training pairs. The usage of the same amount of training samples in other neural networks for face detection seem to be technically impossible to implement. To highlight the future work, I am considering implementing this project using the advanced alternative algorithms and including enhanced features as age and gender detection as the project for the final academic year.

ACKNOWLEDGEMENTS

I would like to express the gratitude to my tutor, Stef Garasto, who carried me through all the crucial steps of developing my project. Furthermore, I would like to thank my team for their enthusiastic support and effective collaboration. Finally, I would like to thank my parents encouraging me in all of my pursuits and inspiring me to follow my dreams.

REFERENCES

[1] Lourdes Ramirez Cerna, Guillermo Camara-Chavez, and D Menotti, 'Face detection: Histogram of oriented gradients and bag of feature method', in *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV)*, p. 1. The Steering Committee of The World Congress in Computer Science, Computer . . . , (2013).

[2] Mehul K Dabhi and Bhavna K Pancholi, 'Face detection system based on viola-jones algorithm', *International Journal of Science and Research (IJSR)*, **5**(4), 62–64, (2016).

[3] Oscar Déniz, Gloria Bueno, Jesús Salido, and Fernando De la Torre, 'Face recognition using histograms of oriented gradients', *Pattern recognition letters*, **32**(12), 1598–1603, (2011).

[4] Zekeriya Erkin, Martin Franz, Jorge Guajardo, Stefan Katzenbeisser, Inald Lagendijk, and Tomas Toft, 'Privacy-preserving face recognition', in *Privacy Enhancing Technologies: 9th International Symposium, PETS 2009, Seattle, WA, USA, August 5-7, 2009. Proceedings 9*, pp. 235–253. Springer, (2009).

[5] Antonio Gulli and Sujit Pal, *Deep learning with Keras*, Packt Publishing Ltd, 2017.

[6] Sandy Huang, Nicolas Papernot, Ian Goodfellow, Yan Duan, and Pieter Abbeel, 'Adversarial attacks on neural network policies', *arXiv preprint arXiv:1702.02284*, (2017).

[7] Gregory Koch, Richard Zemel, Ruslan Salakhutdinov, et al., 'Siamese neural networks for one-shot image recognition', in *ICML deep learning workshop*, volume 2. Lille, (2015).

[8] Tsung-Han Tsai and Po-Ting Chi, 'A single-stage face detection and face recognition deep neural network based on feature pyramid and triplet loss', *IET Image Processing*, **16**(8), 2148–2156, (2022).

[9] Paul Voigt and Axel Von dem Bussche, 'The eu general data protection regulation (gdpr)', *A Practical Guide, 1st Ed., Cham: Springer International Publishing*, **10**(3152676), 10–5555, (2017).

[10] Wanqi Xue and Wei Wang, 'One-shot image classification by learning to restore prototypes', in *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 6558–6565, (2020).