Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/41/20230708

Application and analysis of face matching based on the Siamese model in face recognition

Zhengkai Xiang

School of Advanced Technology, Xian Jiaotong Liverpool University, Suzhou, 215123, China

Zhengkai.Xiang21@student.xjtlu.edu.cn

Abstract. In recent years, face recognition technology, and face matching particularly have broadened the application fields in various aspects of society. It is considered a combination of deep learning architecture and face recognition technology, which has been used for personal information security and safety efficiently for many years. For this, this paper aims to investigate the practical method of utilizing Siamese models to enhance the accuracy and efficiency of face matching systems. The existing challenges of low accuracy and slow recognition rates in face matching applications have been approved to be solvable by utilizing the capabilities of the Siamese model. Experimental analysis and comments from relevant practitioners demonstrate the effectiveness and potential of the Siamese model in enhancing the performance of face matching systems. To conclude, the Siamese model is introduced as a robust and efficient tool in the field of face recognition. It provides higher accuracy and efficiency compared to the traditional feature-based models. Its adaptability and advancements bring the potential to revolutionize face-matching applications and overcome current limitations. The findings from the experiments demonstrate that the utilization of the joint model can significantly enhance the performance of the matching system. The proposed model offers a potential solution to address the issue of low accuracy during the face matching phase.

Keywords: Face recognition, Face matching, Deep learning, Siamese models.

1. Introduction

Face recognition technology has gained significant attention and advancement in recent years, offering a wealth of possibilities in various fields. One practical application is the comparison of face images, a combination of deep learning architecture and face recognition technology. And its application has become a robust guarantee of personal information security. The evolution of face recognition and matching techniques has transitioned from conventional feature-based methods to more sophisticated machine learning-based approaches. Initially, face recognition systems heavily relied on the extraction and comparison of facial features, such as interocular distance, nose shape, and mouth position. Nevertheless, these methods were constrained by their reliance on precise feature extraction and were vulnerable to fluctuations in lighting conditions and facial expressions [1]. For instance, Viola-Jones [2] did significant work that used the AdaBoost algorithm for the cascade classifiers with Haar-like features.

With the advent of machine learning techniques, the field of face recognition and matching witnessed significant advancements. Machine learning models, such as deep neural networks, have demonstrated remarkable capabilities in learning discriminative features directly from raw facial images. These models [3]-[5] can capture intricate patterns and variations in facial appearance, leading to improved accuracy and robustness in face recognition and matching. One notable model in face matching is the Siamese model. The Siamese model is a type of neural network architecture that learns to compare and measure the similarity between two inputs. In the context of face matching, the Siamese model is trained on pairs of facial images, where the objective is to minimize the distance between similar faces and maximize the distance between dissimilar faces in the learned embedding space. This approach allows for more effective face matching by encoding facial features into compact and discriminative embeddings. The R-CNN-based technologies have gained significant progression in the field of object diagnosis [6]-[9]. The adaptability of the Siamese model in face-matching applications is evident in its ability to handle various challenges, such as pose variations, lighting conditions, and facial expressions. By leveraging the power of deep learning and Siamese networks, the Siamese model has shown promising results in face verification, identification, and clustering tasks. It has been successfully applied in areas such as access control systems, surveillance, and personal device authentication. The progress and advancements brought about by the Siamese model in face matching are significant. It has improved the accuracy and efficiency of face matching systems, addressing the limitations of traditional methods. This has led to enhanced security, convenience, and personalized experiences in various domains.

The objective of this paper is to investigate the utilization of the Siamese model for enhancing the performance of face matching systems. By harnessing the capabilities of the Siamese model, the aim of tackling the existing challenges of low accuracy and slow identification rates in face matching applications is met. Through experimental analysis and evaluation, the intention is to showcase the effectiveness and potential of the Siamese model in improving the performance of face matching systems. In conclusion, the Siamese model has emerged as a powerful tool in the field of face matching, offering improved accuracy and efficiency compared to traditional feature-based methods. Its adaptability and advancements have the potential to revolutionize face matching applications and overcome existing limitations. This paper aims to contribute to the ongoing research efforts in utilizing the Siamese model to enhance the performance of face matching systems, ultimately improving accuracy and speed in real-world applications. The experimental results show that utilizing the conjoined model can effectively improve the performance of the face matching system. The proposed model can solve the existing problems of low accuracy and low recognition rate in face matching applications.

2. Methodology

2.1. Dataset description

The datasets are used for training and evaluating model performance, helping the researchers and developers get insight into relevant address real-world problems [10]. Datasets are of utmost importance in the field of face matching as they offer a significant number of face images [11] for training and evaluating the efficacy of face-matching models. MongoDB, an open-source NoSQL database management system, is widely recognized for its popularity. It is specifically designed to store and manage large volumes of structured and unstructured data. In the realm of face matching experiments, MongoDB assumes a critical role. It serves as a training dataset for training deep learning models, such as Siamese networks, to acquire the feature representation of faces. Furthermore, MongoDB is utilized as an evaluation dataset to gauge the performance of face matching models. Researchers can test and compare their models on MongoDB to comprehend their effectiveness in real-world scenarios.

2.2. Proposed approach

The research objective of Siamese model experiments is to achieve accurate face matching by learning the feature representation of faces. The Siamese model is a deep learning model that maps two similar or dissimilar input samples to the same feature space for matching tasks. The researchers aim to improve the accuracy and robustness of face matching through the training and evaluation of the Siamese model.

DOI: 10.54254/2755-2721/41/20230708

The research methodology typically involves the following steps: Firstly, training the Siamese model using large-scale face recognition datasets such as MongoDB. During training, the model's parameters are optimized by minimizing a loss function to make similar face pairs closer in the feature space and dissimilar face pairs farther apart. Secondly, evaluating the trained Siamese model's performance using evaluation datasets like MongoDB. Performance metrics such as matching accuracy, recall, and data loss are computed to assess the model's performance in different scenarios and conditions. Researchers may also perform techniques such as hyperparameter tuning and data augmentation to improve the model's performance. Through Siamese model experiments, researchers gain deeper insights into the challenges and issues in face matching tasks and propose improved methods and techniques to enhance the accuracy and robustness of face matching.

2.2.1. Preprocessing. The image preprocessing pipeline typically involves several steps to enhance the quality and suitability of images for further analysis or tasks. The main purpose of image preprocessing is to reduce noise, remove irrelevant information, and standardize the images to ensure consistency and improve the performance of subsequent algorithms or models. With the advent of deep learning, features can also be learned automatically from images using deep neural networks. Convolutional neural networks (CNNs) are commonly used for extracting high-level features from images. The process is shown in Figure 1. Firstly, since the dataset contains a large number of face images, face detection, and alignment are performed to ensure consistent face positions and sizes across all images. Secondly, image cropping and scaling may be done to reduce noise and unnecessary information. Furthermore, data augmentation techniques like random rotation, translation, and flipping can be applied to improve model generalization. To achieve those steps, the pre-trained model would be introduced, and this neural network has been trained on a significantly large dataset. It generates a distinct 128-dimensional vector for each specific face when provided with images that have been cropped to only include the facial region and are properly aligned. The input image size required for this network is 160x160x3. These preprocessing steps help enhance model performance and robustness while reducing inconsistencies and noise in the dataset. The process is shown in the figure 1.

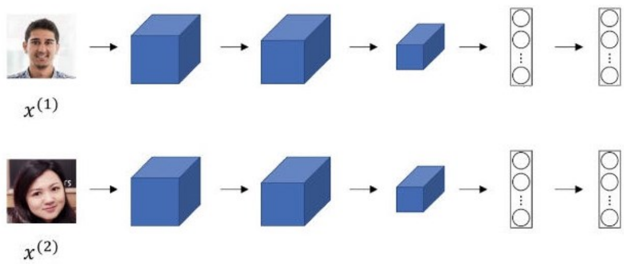

Figure 1. Preprocessing steps of the images

2.2.2. Contrastive. In this study, author proposes training a conjoined network to process pairs of images and optimize the embedding by minimizing the distance between them if they belong to the same category. Conversely, if the images represent different categories, the network is trained to increase the distance between them beyond a predefined threshold. The image source is determined by evaluating the image feature value obtained from image preprocessing. By optimizing the loss function, the degree of loss in the feature value is reduced. The contrast loss function will be minimized:

$$L_{contrastive}(x_0, x_1, y) = \frac{1}{2} y \|f(x_0) - f(x_1)\|_2^2 + \frac{1}{2}(1 - y)\{\max(0, m - \|f(x_0) - f(x_1)\|_2)\}^2 \quad (1)$$

where the coordinate x and y represent the difference between different images, with the integral of the corresponding function and corresponding maximum function for minimizing a contrastive loss function.

2.2.3. Triplet. A triple network containing anchors, positive examples and negative examples (of the same type as the anchors) is generated so that the anchors are closer to the positive examples than the negative examples. For the creation of positive and negative samples, a random forest random function is randomly generated that is closer to the positive samples overall by comparing eigenvalues. The Triplet function selects a set of positive and negative examples for each anchor, as,

$$L_{triplet}(x_a, x_p, x_n) = \max\left(0, m + \left\|f(x_a) - f(x_p)\right\|_2^2 - \left\|f(x_a) - f(x_n)\right\|_2^2\right) \tag{2}$$

where the coordinate x represents the eigenvalue of different images, with the integral of the corresponding function, and the different relationship between the integrals leads that the anchor is closer to the positive example than it is to the negative example by some margin value.

2.3. Implemented details

The research utilized Jupiter Notebook (Python 3.10.12) and the MongoDB library to implement decision tree models. Data visualization was performed using the torch vision and SoftMax libraries. The study was conducted on a Windows device equipped with an 11th Gen Intel(R) Core (TM) i5-11400H processor. The decision tree was configured with the following settings: A special Batch Sampler was used to generate n classes and n samples corresponding with each class, resulting in mini-batches of size n classes * n samples. Since MNIST is a relatively easy dataset, no additional settings were required. These settings allowed the initial sample models to accurately capture patterns and relationships in the data, enabling further optimization and improvement.

3. Result and discussion

This chapter concentrates on the optimization orientation of the Siamese model trained using a specific database. It examines the loss function and the image output of the embedded network to assess the model through metrics such as metric attributes and analysis rate. The existing models, such as CNN, are trained using a large number of images to automatically extract features, primarily local key point descriptors, from given images. The eigenvalues are withdrawn and stored as multi-dimensional coordinates. These extracted features are then compared and matched with the features already extracted in a database. The similarity or match between the image and the images in the database is determined using a threshold-based approach. The results are evaluated using metrics such as accuracy, recall, and discrimination. For instance, the triplet function generates a pair of positive and negative examples, and the features of the anchor group are further strengthened to reduce the loss in the comparison process and improve the comparison accuracy. Some experimental results are presented in the figure 2 and figure 3.

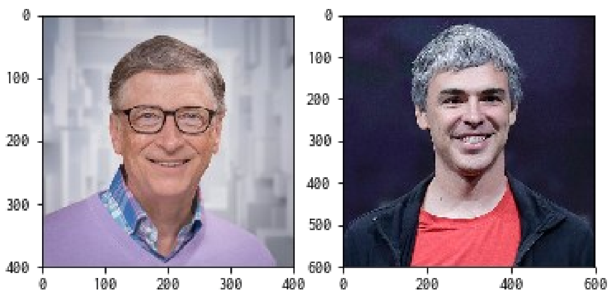

Figure 2. Bill Gates & Larrypage

DOI: 10.54254/2755-2721/41/20230708

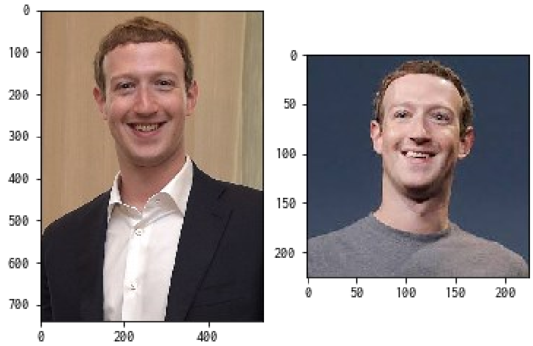

Figure 3. Mark Elliot Zuckerberg

As shown in Figures 4 and 5, the proposed model can effectively distinguish different types of data in the embedded vector space, and can reasonably calculate the differences between different types of data. And the low dimensional visualization in Figure 5 can effectively demonstrate the efficient classification performance of the proposed model applied to face matching. It should be noted that the performance and accuracy of feature extraction and matching depend on the choice of feature extraction methods and matching algorithms. Researchers typically conduct experiments and optimization to select the most suitable methods and parameters for specific tasks. The aforementioned technique of extracting features from preprocessed portraits, along with the outcomes of the comparison and training model based on this approach, collectively establish the viability of Siamese in the domain of portrait recognition.

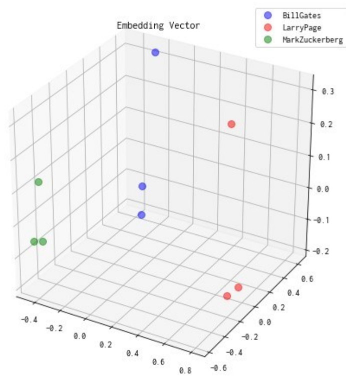

Figure 4. Embedding Vectors

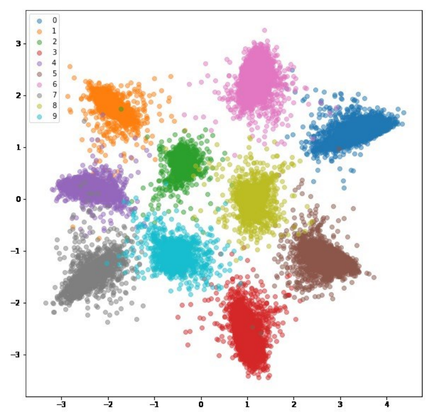

Figure 5. Embeddings for training set

DOI: 10.54254/2755-2721/41/20230708

4. Conclusion

This study introduces the Siamese model for facial image feature extraction. This model is particularly effective in tasks such as face recognition, face verification, and facial expression analysis. The Siamese architecture comprises two identical neural networks that possess shared weights and are trained to acquire a similarity metric for pairs of facial images. The adaptability of the Siamese model arises from its capacity to capture and evaluate nuanced distinctions in facial characteristics, regardless of lighting circumstances, facial expressions, and pose variations. By learning a discriminative embedding space, the Siamese model can effectively differentiate between similar and dissimilar faces. Experimental studies have further confirmed the value of the Siamese model in facial image feature extraction. These experiments involve evaluating the model's performance on benchmark datasets, such as Labeled Faces in the Wild and Mega Face, which contain a large number of face images with variations in pose, illumination, and occlusion. The results consistently demonstrate the superior performance of the Siamese model in accurately identifying and verifying faces. In the future, research will further extend to other areas such as object recognition, image retrieval, and even medical image analysis using Siamese models. The Siamese model can be utilized for its ability to learn to discriminate features and measure similarities to drive research and development in various fields that require accurate and efficient feature extraction and matching.

References

[1] Han J Lu J 2018 Multi-scale Deep Representation Learning for Face Detection 2018 IEEE Visual Communications and Image Processing (VCIP) pp 1-4

[2] Viola P JonesM 2004 Robust Real-Time Face Detection. International Journal of Computer Vision 57 pp 137–154

[3] Zhu X and Ramanan D 2012 Face detection pose estimation and landmark localization in the wild CVPR pp 2879-2886

[4] Mathias M Benenson R Pedersoli M and Van Gool L 2014 Face detection without bells and whistles ECCV pp 720-735

[5] Yan J Lei Z 2014 The fastest deformable part model for object detection CVPR pp 2497-2504

[6] Girshick R Donahue J 2014 Rich feature hierarchies for accurate object detection and semantic segmentation CVPR pp 580-587

[7] Girshick R Donahue J Darrell T and Malik J 2016 Region-based convolutional networks for accurate object detection and segmentation PAMI 38(1): pp 142-158

[8] Girshick R 2015 Fast r-cnn ICCV pp 1440-1448

[9] Ren S He K Girshick R and Sun J 2015 Faster r-cnn: Towards real-time object detection with region proposal networks NIPS pp 91-99

[10] Singh S 2019 Face recognition using Siamese network https://github.com/satinder147/Attendance-using-Face

[11] taeniae H 2018 Keras face net Github https://github.com/nyoki-mtl/keras-facenet