

A review of face recognition technologies based on deep learning

Chong Deng

Peking University, Beijing, 100091, China

dengchong1994@163.com

Abstract. Face recognition is a widely used technology in the field of security, with the goal of identity verification or recognition by analyzing and comparing facial images. In recent years, the rapid development of deep learning technology has brought significant breakthroughs to face recognition. Face recognition technology mainly involves four modules: face detection, face alignment, face representation, and face matching. Among them, face representation is the most important module in face recognition technology, which directly affects the accuracy of face recognition. This article summarizes the application of deep learning technology in the field of face recognition in the past decade and introduces the technologies as well as their development involved in the above four modules. According to analysis, it is concluded that deep learning models can learn and capture rich features in face images and thus achieve more accurate and stable face recognition. Besides, feature extraction is a critical link in deep learning face recognition, and model training and optimization are key to achieving high-performance face recognition.

Keywords: Convolutional neural network, Deep learning, Facial representation, Face recognition.

1. Introduction

Face recognition technology is an important research direction in the field of computer vision, playing a crucial role in applications such as security, human-computer interaction, and face authentication. With the rapid development of deep learning technology, face recognition technology based on deep learning has made significant breakthroughs in accuracy and robustness. The introduction of deep learning models enables researchers to learn features with strong representation capabilities from large-scale facial datasets, thereby achieving efficient face detection, feature extraction, and recognition. This article reviews the development of deep learning based on face recognition technology in recent years, introduces the key methods and algorithms involved, and focuses on exploring the application of deep learning models in face recognition, including convolutional neural networks (CNN) and commonly used loss functions and data augmentation techniques. In addition, the article also discusses the challenges faced by current face recognition technology and future development directions, in order to provide reference and inspiration for relevant researchers. Through the review of this article, it is hoped that the cutting-edge progress of deep-learning-based face recognition technology can be deeply explored, providing useful guidance and reference for practical applications and theoretical research.

2. Technical modules involved in face recognition

Face recognition technology refers to the technology that can analyze and process facial visual feature information in images or videos, and perform identity authentication. In the broad sense of the definition, face recognition refers to a series of technologies for the construction of face recognition systems, including face detection and localization, face preprocessing, and identity confirmation and matching. In the narrow sense of the definition, face recognition refers to the technology of identity confirmation and matching through the face. In the field of computer vision, face recognition is commonly discussed and researched. As seen in Figure 1, face recognition systems typically involve the following four technical modules:

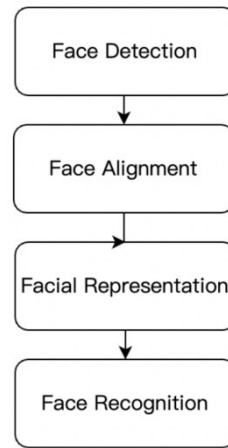


Figure 1. The process of face recognition.

The first is face detection. A face detector can locate the position of faces in the image and return the coordinates of the bounding box containing each face.

The second is face alignment, the goal of which is to scale and crop the facial image by applying a group of reference points located at fixed positions in the image. A feature point detector is typically required in the process to find a group of facial feature points in a simple two-dimensional (2D) alignment case, so as to find the best affine transformation for the most suitable reference point.

The third is face representation. In this stage, the pixel values of the facial image are converted into compact and distinguishable feature vectors, also known as templates. Ideally, all faces of the same subject are supposed to be mapped to similar feature vectors. This module is the core module of face recognition technology and will directly affect the accuracy of face recognition.

The last is face recognition. In the face recognition construction module, two templates are compared to obtain a similarity score, which gives the possibility of both belonging to the same subject. At present, the main matching algorithm in deep-learning-based face recognition technology is the cosine similarity matching algorithm.

As face representation technology is the core technology of face recognition, this article mainly discusses the development of this technology.

3. The historical development of face recognition

Face recognition technology has always been a challenging technology since its introduction. It was not until the application of convolutional neural networks in facial recognition technology that a breakthrough was made, and it was widely applied in fields and scenarios that require identity verification and recognition, such as mobile apps, train airports, and banks.

The first face recognition algorithm was proposed in the early 1970s [1,2]. Before the explosion of convolutional neural networks, people often used manually designed features combined with machine learning techniques such as principal component analysis and support vector machines for face recognition processing. Due to the influence of lighting, angle, posture, and other factors on images

containing faces, the biggest problem with traditional machine learning techniques is that under unconstrained conditions, manually designed features are difficult to represent various images.

Traditional face recognition technology mainly includes three steps: preprocessing, feature extraction, and classification. The preprocessing stage usually includes operations such as graying, smoothing, and denoising. In the feature extraction stage, some classic algorithms are used, such as LBP (Local Binary Patterns) and HOG (Histogram of Oriented Gradients), to extract local features of the image. In the classification stage, traditional machine learning algorithms such as Support Vector Machine and Decision Tree are used for classification.

Modern face recognition technology mainly includes four steps: data collection, data cleaning, feature extraction, and classification. Compared to traditional methods, modern methods place greater emphasis on data quality and diversity, while also relying more on deep learning techniques. Commonly used modern face recognition techniques include methods based on convolutional neural networks.

4. Application of convolutional neural networks in the field of face recognition

In 2012, AlexNet shined in the field of image classification. Compared with traditional machine learning algorithms, AlexNet builds an image classification model based on convolutional neural networks and uses a large amount of data to train this model. This trained model greatly improves the accuracy of image classification. Subsequently, some researchers applied convolutional neural networks to face recognition.

Convolutional neural networks completely abandon artificially designed features. People only need to build a network structure through a convolutional neural network, feed data to the network, and let the network itself learn pictures containing various scenes, extract features, and then classify them.

The accuracy of face representation technology based on convolutional neural networks is mainly affected by three factors: training data, network structure, and loss function.

In 2014, the DeepFace network [3] proposed by FaceBook first applied convolutional neural networks to face recognition technology. Researchers trained a convolutional neural network on a data set containing 4.4 million faces of 4,030 people. The accuracy rate reached 97.35%.

In 2015, Google proposed the FaceNet network [4]. The network directly maps images to Euclidean space, and then Euclidean distance can represent the similarity of faces. Moreover, the network directly trains the target and constructs a triplet training pair so that 128-dimensional features can achieve the best results so far. In the study of Schroff et al. [4], end-to-end training is used, and the training criterion is closest to the final goal. In addition, this method adopts a new loss function criterion, Triplet Loss, which considers both positive and negative samples and also takes margin into account. At the same time, the dimension of the face feature vector generated from the picture is relatively low, only 128 dimensions, and the Euclidean distance can express the similarity of the face. It can perform tasks such as face classification, face verification, and face clustering at the same time, and it is widely used.

In 2016, the Chinese Academy of Sciences designed Center Loss [5]. Most previous face classification tasks used softmax-based criteria. Center loss is proposed in the research of Wen et al. [5]. The function of this loss can make the samples of each category very close to the middle. Using Softmax + center loss joint training, it is easy to optimize and achieves good performance. This article makes the trained features more compact for the same category; it makes the training simpler and more convenient, especially the training data; it can use the information of each category, and there is very little information loss.

In the same year, Peking University (weiyang Liu) proposed L-Softmax Loss [6]. L-Softmax Loss is proposed to make features have the properties of intra-class compactness and inter-class separability, which can avoid over-fitting and achieve good results in many visual tasks. This article enables the trained features to have the characteristics of intra-class compactness and inter-class separability. In addition, it also has an intuitive geometric explanation as well as the parameter m to control the difficulty of learning or the size of the margin, thus avoiding over-fitting and directly improving softmax while no additional complexity is required for training.

In 2017, Peking University (weiyang Liu) once again proposed SphereFace (A-Softmax) [7]. Based on the idea of maximizing the intra-class distance and minimizing the inter-class distance, this article proposes the A-Softmax algorithm to make the features distinctive on the hypersphere. This enables the trained features to distinguish different types of faces. Moreover, the maximum intra-class distance should be smaller than the minimum inter-class distance, and there is theoretical guidance. This method achieves good results on many face recognition tasks.

Zheng et al. proposed Ring loss as a new loss function [8], which can significantly improve the performance of face recognition models and is widely used.

In 2019, researchers from Imperial College London proposed ArcFace [9]. Researchers have found that the main challenge in using deep learning to extract features is to design a discriminative loss function. Previous research on loss functions has achieved some results, such as center loss and SphereFace. The author of this article proposed ArcFace to obtain highly discriminative features for face recognition and has a clear geometric explanation. It has achieved good results on many face recognition tasks. The main contributions of this article are that the public source code is simple to modify and easy to reproduce; the training efficiency is high with almost no additional computational overhead; and it has an intuitive geometric explanation. The paper proposes a new face recognition algorithm called "ArcFace" that can effectively solve the "sample imbalance" problem in traditional face recognition algorithms. The algorithm uses a metric called "cosine similarity" to calculate the similarity between faces.

Liu et al. entered SphereFace and used adaptive margin and sampling methods to achieve an accuracy of 99.53% on the LFW face verification set [10]. AdaptiveFace is a method for face recognition. Its core idea is to adopt adaptive boundaries and sampling strategies. This method aims to solve challenges in face recognition, such as pose changes, illumination changes, and occlusions. AdaptiveFace proposes the concept of adaptive boundary learning, which introduces a learnable boundary to the samples of each category so that the boundary can be adjusted according to the distribution characteristics between different categories. This can improve the discriminative ability of the model and enhance the boundary discrimination between different categories. At the same time, AdaptiveFace also introduces an adaptive sampling strategy to select samples for training in a more targeted manner by adjusting the weight of samples according to the difficulty of each category. This can increase attention to difficult samples and improve the generalization ability of the model.

Shi et al. proposed a new facial feature extraction and matching method [11], which can produce Probabilistic Face Embeddings and improve the robustness of facial recognition. Probabilistic Face Embedding models face embedding as a probability distribution instead of a single vector by introducing the idea of probabilistic modeling. Doing so can better reflect the uncertainty and diversity of facial features. Specifically, Probabilistic Face Embeddings can utilize generative models such as Variational Autoencoder or Generative Adversarial Networks to treat embeddings as latent variables and approximate their posterior probability distributions. By modeling face embeddings in probabilistic space, Probabilistic Face Embeddings can provide more flexible representation capabilities and allow uncertainty to be modeled. This approach may have advantages in face-related tasks such as clustering, retrieval, and re-identification, and it is able to handle noise and variation in the data better.

Huang et al. proposed the CurricularFace stepwise learning strategy to achieve SOTA results in face recognition [12]. This literature uses a method of adjusting sample weights so that difficult and easy samples receive different attention during training, and the relationship between different categories is also adjusted through adaptive boundaries. This adaptive course learning can help the model better learn complex facial features and improve the accuracy of face recognition. By gradually adjusting the learning curve method of training sample difficulty, the best face recognition effect at the time was achieved.

Kim et al. proposed a new loss function to emphasize the importance of different difficult samples through image quality [13]. This method does this by approximating image quality with feature norms, here in the form of adaptive edge functions. Extensive experiments show that AdaFace improves the existing (SoTA) face recognition performance on 4 data sets (IJB-B, IJB-C, IJB-S, and IJB TinyFace).

5. Existing challenges and future prospects of face recognition technology based on convolutional neural networks

There are still some challenges and problems in CNN-based face recognition technology.

First, the data set quality is insufficient. Due to the diversity and complexity of face images, the quality of training data directly affects the performance of the model. Therefore, it is necessary to continuously optimize the construction method and content of the data set to improve the generalization ability and robustness of the model.

Second, there is a problem of model overfitting. During the training process, the model may be overfitted, that is, the model performs well on the training data but performs poorly on the test data. In order to avoid overfitting, techniques such as regularization and dropout can be used to constrain the complexity of the model.

Third, real-time requirements are high. In practical applications, face recognition technology needs to meet real-time requirements, that is, the analysis and comparison of face images must be completed within a short period of time. So, the algorithm needs to be optimized to increase the calculation speed.

Fourth, there are privacy protection issues. Face recognition technology involves personal privacy issues, and corresponding measures need to be taken to protect personal privacy. For example, anonymization, encrypted storage, and other technologies can be used to protect personal privacy.

Finally, in the field of face recognition, there are also problems of recognition of the same person at different ages, recognition of partial occlusion of the face, and recognition of different facial postures of the same person. They bring challenges to the accuracy and recall of face recognition algorithms.

In the future, CNN-based face recognition technology will continue to develop. Here are some possible directions for development. Face recognition can be combined with other technologies, for instance, combining CNN-based face recognition technology with other biometric recognition technologies, such as fingerprint recognition and iris recognition, can further improve the accuracy and stability of recognition.

In addition, adaptive learning technology can be introduced into the field of face recognition. Through adaptive learning, the model can automatically learn more effective feature representations from the data, thereby improving the performance of the model.

Finally, issues such as cross-domain face recognition are also worthy of in-depth study. Applying CNN-based face recognition technology to cross-domain face recognition scenarios, such as face recognition between different countries or regions, requires resolving issues such as cross-domain data differences, accents, and dialects.

6. Conclusion

Facial recognition technology based on deep learning has made tremendous progress in the past few years and has become one of the core technologies in the modern biometric field. Through large-scale data sets and powerful computing power, deep learning models are able to learn and capture rich features in face images, thereby achieving more accurate and stable face recognition.

In terms of data preprocessing, researchers have introduced techniques such as alignment, cropping, and enhancement to improve the quality and consistency of face images. At the same time, technologies such as cross-domain face recognition and robustness enhancement have also been extensively studied, making face recognition more robust under different lighting, posture, and expression conditions.

Feature extraction is a key link in deep learning face recognition. By introducing structures such as convolutional neural networks (CNN) and residual neural networks (ResNet), researchers have successfully extracted rich facial feature representations. In addition, technologies such as attention mechanism, multi-scale feature fusion, and attention map visualization have also been widely used, further improving the performance of face recognition.

Model training and optimization are key to achieving high-performance face recognition. Researchers have adopted a variety of strategies, including transfer learning, generative adversarial networks (GAN), and metric learning, to enhance the model's generalization ability and robustness. In

addition, methods such as label smoothing, sample balancing, and dynamic weight adjustment have also been proposed to help solve problems such as data imbalance and difficulty in category classification.

Future research can focus on the following three aspects: the first is the robustness improvement. It is essential to further improve the performance in the face of changes in lighting, posture, and expression, especially the recognition accuracy in complex backgrounds. Besides, in terms of cross-domain and cross-modal recognition, it is important to study how to effectively transfer knowledge to different domains and modalities to achieve model generalization capabilities and scalability. The second is the privacy and security issue. It is necessary to explore how to improve the privacy protection and resistance of face recognition systems to spoofing attacks. In terms of small-sample learning, there is a need to develop deep learning models suitable for small-sample situations to improve recognition performance in data-scarce scenarios. The third is about multi-modal fusion. Facial images, voice, posture, and other sensor data can be fused to further improve the accuracy of multi-modal face recognition systems. With the continuous improvement of computing power and the continuous innovation of deep learning methods, future deep learning face recognition technology is expected to achieve more extensive applications in fields such as security monitoring and social media.

References

- [1] Kelly, M. D. (1970). Visual identification of people by computer. tech. rep., Stanford Univ Calif Dept of Computer Science.
- [2] Kanade, T. (1973). Picture processing by computer complex and recognition of human faces. PhD Thesis. Kyoto University.
- [3] Taigman, Y., Yang, M., Ranzato, M., et al. (2014). DeepFace: Closing the gap to human-level performance in FaceVerification. IEEE Conference on Computer Vision & Pattern Recognition. IEEE Computer Society.
- [4] Schroff, F., Kalenichenko, D. and Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 815-823.
- [5] Wen, Y., Zhang, K., Li, Z. and Qiao, Y. (2016). A discriminative feature learning approach for deep face recognition. In European Conference on Computer Vision (ECCV). 499-515.
- [6] Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B. and Song, L. (2016). Large-margin softmax loss for convolutional neural networks. In Proceedings of the 33rd International Conference on Machine Learning (ICML), 48, 507-516.
- [7] Liu, W., Liu, Z. and Ma, H. (2017). SphereFace: Deep hypersphere embedding for face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 212-220.
- [8] Zheng, Y. and Pal, D. K. and Savvides, M. (2018). Ring loss: Convex feature normalization for face recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, 5089-5097.
- [9] Deng, J., Guo, J., Zhou, J., Gong, Y. and Zafeiriou, S. (2019). ArcFace: Additive angular margin loss for deep face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 4690-4699.
- [10] Liu, H., Zhu, X., Lei, Z., et al. (2019). Adaptiveface: Adaptive margin and sampling for face recognition. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 11947-11956.
- [11] Shi, Y. and Jain, A. K. (2019). Probabilistic face embeddings. Proceedings of the IEEE/CVF International Conference on Computer Vision, 6902-6911.
- [12] Huang, Y., Wang, Y., Tai, Y., et al. (2020). Curricularface: Adaptive curriculum learning loss for deep face recognition. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 5901-5910.

- [13] Kim, M., Jain, A. K. and Liu, X. (2022). AdaFace: Quality adaptive margin for face recognition. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 18750-18759.