ITM Web of Conferences **78**, 02029 (2025) https://doi.org/10.1051/itmconf/20257802029
CSEIT 2025

Analysis Loss Functions for Recognition Accuracy Improvement in Face Recognition

Zhaoyin Yang

Faculty of Data Science, City University of Macau, Macau,999078, China

Abstract. In the field of face recognition, improving recognition accuracy is an important research direction. Due to the fact that face images may be disturbed during acquisition, transmission and storage, the recognition algorithms are not robust enough to meet the requirements of practical applications. Therefore, the introduction of an effective loss function is particularly critical. In this paper, the advantages and disadvantages of different loss functions will be derived by analysing the softmax, ArcFace and AdaFace loss functions and comparing the differences between them. Through the analysis, this paper concludes that the loss functions still have different defects in the application of real scenarios and cannot effectively identify data with fewer samples. For example, Arcace loss is not able to perfectly make the samples evenly distributed on the hypersphere in real life. Therefore, in the future design and optimisation of the self-supervised pre-training framework should be strengthened to make a greater improvement in the recognition accuracy of face recognition.

1 Introduction

With the development of face recognition technology, face recognition has gradually penetrated into people's lives and become an indispensable part of life. Face unlocking of mobile phones, illegal capture by traffic cameras, face punching at work, these are all common applications of face recognition technology in our lives. These performances seem to indicate the maturity of the technology. However, there is still significant room and necessity for its development in improving the robustness of face recognition.

In terms of improving the accuracy of face recognition, people generally achieve this through the improvement of deep learning models, the development of preprocessing techniques and the evolution of loss functions. Among them, the loss function is used to achieve targeted improvement of the model's ability to handle complex samples such as occlusion and blurring, by introducing a mining strategy for difficult samples. With the traditional loss function led by softmax, face recognition often cannot handle disturbed face images with high accuracy.

Therefore, researchers try to create new loss functions to improve the robustness of face recognition. Wang Yibo et al. for the current face recognition algorithms in the lack of samples and have to anti-interference face recognition application scenarios have limitations, proposed a symmetric network-based recognition algorithm [1]. Wang Jiyong

D23909100401@cityu.edu.mo

et al. invented an occluded face recognition method based on FaceNet and attention mechanism in order to solve the problem of how to obtain the face features of the non-occluded face part so as to improve the performance and robustness of the face recognition task. [2] By innovating the loss function, the model can be made more adaptable to the recognition task under different conditions and the generalisation ability of the algorithm can be improved.

This paper will explore how to improve the robustness of face recognition by improving the existing loss function and provide suggestions and insights for future research.

2 Research in loss functions

2.1 Concept and Classification of Loss Functions

The essence of the loss function is to guide parameter adjustment by quantifying the difference between the predicted value and the true value.

In face recognition scenarios, the design of the loss function is mainly based on metric learning, focusing on intra-class compactness and inter-class differentiation. By incorporating the loss function into face recognition, the robustness of face recognition is greatly improved. The recognition accuracy of face recognition is improved by classifying images through large-scale data, grouping images of different classes into different categories, and increasing the inter-class distinction through punishment.

In regression tasks, the mean square error or mean absolute error is often used to measure the distance. However, when facing classification tasks, it relies on the cross-entropy loss to assess the probability distribution difference. Loss functions can be classified into two categories based on the type of task and mathematical properties: regression loss functions and classification loss functions.

The regression loss function formula is shown below:

$$L_2 loss: \ L\left(y, \ \hat{y}\right) = \omega(\theta)(\hat{y} - y)^2 \tag{1}$$

$$L_1 loss: \ L\left(y, \ \hat{y}\right) = \omega(\theta)|\hat{y} - y| \tag{2}$$

$\omega(\theta)$: Weighting of true values; y: True value; $\hat{y}$: The output of the model.

Regression models and artificial neural networks estimate their parameters by minimising L2 or L1 losses. InL_1loss andL_2loss ,L_2loss amplifies the distance between the estimated value and the true value by squaring it, which penalises outputs that deviate from the observed value more significantly;L_1loss takes the absolute value of the difference between the model outputs and the true value, resulting in an insensitivity to outputs that deviate from the true value, which keeps the model stable in the event of an outlier in the observation.

The classification loss function is a discontinuous segmented function which divides the estimates into 0 and 1, taking 0 for correct classification and 1 for incorrect classification. The formula is shown below:

$$L\left(\hat{y}, \ y\right) = \begin{cases} 0, \ \hat{y} = y \\ 1, \ \hat{y} \neq y \end{cases} \tag{3}$$

2.2 Loss function ArcFace

The softmax loss function essentially normalises the multiple values obtained in the neural network and the values obtained in the function are all at[0,1] . In different categories, the softmax loss function gives higher values to categories with high category probability and lower values to categories with low category probability.

In face recognition, the mathematical form of the softmax loss function is simpler, and compared to other loss functions it has less computational effort and higher efficiency. At the same time, the softmax loss function is more stable when trained on large-scale datasets, especially in low noise. However, the softmax loss function only focuses on the correctness of classification in nature, and cannot directly constrain the intra-class compactness and inter-class differentiation, which leads to its ambiguity in the decision boundary, so it is gradually eliminated [3].

For the problem that the traditional loss function softmax lacks effective constraints on intraclass compactness and interclass variability in face recognition tasks, Jiankang Deng, Jia Guo et al. proposed an additional angular edge loss (ArcFace) to improve the discriminative ability of face recognition models in 2019 [4].

They embedded features distributed around the centre of each feature on a hypersphere of radius s. After obtaining the angle θ between the feature and the true value weight by calculating $\arccos(\theta_{y_i})$, an additional angular edge penalty m is added by calculating $\theta_{y_i} + m$ and multiplying all the cos values by s. Finally, the cross-entropy loss is obtained by the softmax function and calculated.

Expressed mathematically as:

$$L = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{e^{s\cos(\theta_{y_i}+m)}}{e^{s\cos(\theta_{y_i}+m)} + \sum_{j=1,j\neq y_i}^{n} e^{s\cos\theta_j}} \tag{4}$$

θ_{y_i}: The angle between the feature vector and the corresponding category weight vector. m: a preset angular interval (e.g., 0.5 radians) to improve discriminatory properties by increasing the decision boundary. s: radius of the hypersphere.

The ArcFace loss function belongs to an improvement on the softmax function. Compared to Softmax, which produces significant ambiguities in the boundaries of decision making, the ArcFace loss can force more significant gaps between the nearest classes (Figure 1).

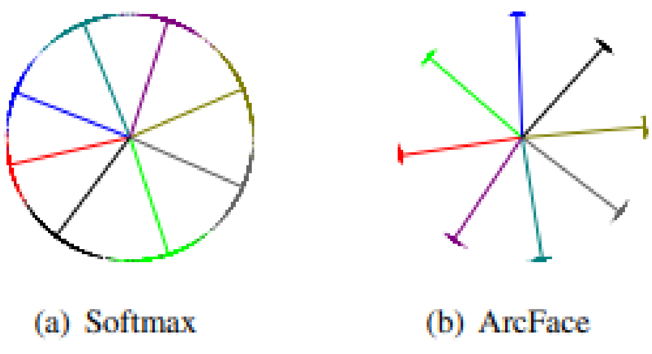

Figure 1 Additive Angular Margin Loss for Deep Face Recognition[5]

However, it also inherits the same problem that Softmax has - that is, it is too dependent on the quantity and quality of data. The classification loss function often has a huge demand for different kinds of images, when the training data is insufficient or the class samples are too small, ArcFace may not be able to adequately learn the discriminative features, resulting in overfitting or generalization ability is reduced. He also has his unique problems.

The first is that the additional angular edge penalty m in the function needs to be adjusted very finely, so it requires a lot of computation to adjust m. If the value of m is slightly off, the accuracy of the function itself is greatly reduced, and it can not complete the task properly.

The second point is its poor robustness to noisy labels; the ArcFace loss function enhances the discriminative property by forcing the interclass interval to increase, but if there are mislabels in the training data, the model may degrade in performance by overfitting to the noise.

2.3 Loss function AdaFace

In the loss function ArcFace, the fixed interval parameter m it employs, results in high and low quality samples being treated equally in the feature space and consequently affects the robustness of the model (Figure 2). Therefore, AdaFace uses an adaptive margin formula to map the feature paradigm to the interval [a_{min}, a_{max}]and uses this to adjust the value of m - low paradigms are assigned smaller m to avoid over-penalisation, and high paradigms are assigned larger m to enhance inter-class separability [5]. However, there is no strict criterion to distinguish the high quality of the image from the extreme pose of the face in the image, leading to the fact that the image quality may not be poor when the feature paradigm is at a low level and vice versa.

Figure 2 Quality Adaptive Margin for Face Recognition [5]

The maximum value of the feature paradigm mapping interval of AdaFace is not enough for some very high-quality images, so it has more significant effect when facing difficult to recognise low-quality samples, and is more widely used. The Changchun University of Science and Technology team in 2024 proposed to combine Retinaface-Resnet and adaface , Retinaface-Resnet for face detection and localisation, and adaface to solve the low-quality face recognition problem [6]. They use two loss functions to refine each other, so that the robustness of face recognition can be done more accurately.

3 The application of loss function in face recognition

Face recognition is widely used in daily life, for example, face capture of red-light runners at roadside traffic lights, face recognition to open doors in companies, and so on. With the continuous innovation of face recognition technology, the way of its application is also changing.

Tingshuo Hu and Xiqu Chen designed a face recognition attendance system using MTCNN and ArcFace [7]. The system uses MTCNN to complete face detection, then performs Gamma correction on the input face image, and finally uses ArcFace to complete face recognition thus realising a convenient and efficient attendance system that does not require active cooperation from students.

In order to verify the performance of the system, they set up a comparison experiment. It uses a real classroom scene with low light and contains 19 faces, each of which is of different sizes and shows different postures and expressions, and 5 of which are partially obscured. The feasibility of its system is verified through direct face recognition of the original image and recognition of the Gamma-corrected image, so that the result goes from the original accuracy of 73.68% to 100% later on.

Yuewen Hu, Dan Li proposed a face recognition model based on ArcFace loss function in 2020 and applied it to the field of guardianship security face recognition [8]. In the experimental results of this model and the analysis of real-time face recognition effect, its advantages and feasibility in guardianship security face recognition are verified, which provides security for the guardians of primary school students to receive delivery. Its experiments on PubFig dataset, by comparing the application of SoftMax, SphereFace, CosFace and ArcFace loss function, the result is that ArcFace has the highest accuracy of 87%, which verifies the correctness of using ArcFace loss.

Chazufushui et al. complete a pedestrian multiple loitering detection algorithm based on pedestrian detection and peak density clustering according to AdaFace, so as to monitor the most suspicious people and prevent the danger from happening [9]. The system firstly tracks the detected pedestrians by Deepsort algorithm to get the movement trajectory of pedestrians, and then analyses the trajectory of pedestrians by peak density clustering algorithm to determine whether the pedestrians are wandering or not. Finally, the AdaFace face recognition algorithm is used to match the faces of pedestrians who wander, so as to achieve the detection of multiple wanderings of pedestrians in different time periods in the region. It uses some pictures and videos from the PASCAL VOC dataset and CASIA behavioural dataset, and measures the detection accuracy of the model by the average accuracy AP, and compares it with the SSD and YOLOV4 -MobileNet Vl algorithms in terms of three aspects: AP, FPS and model size, and obtains the result of the highest accuracy value of 88.22%, which verifies its superiority.

4 Outlook

Overall, in this aspect of the loss function in face recognition, we have basically achieved inter- and intra-class discriminative optimisation and robustness to complex scenes, improving the recognition accuracy of face recognition. From softmax loss to ArcFace loss to AdaFace loss, the loss function has been iterated over and over again to address the problems of the previous loss function and to compensate for the major problems of the former.

However, for the loss function in face recognition, its applicability in real-world scenarios is still lacking. For example, ArcFace assumes that all categories are uniformly distributed on the hypersphere, but the actual data may have uneven distribution or overlap between categories, resulting in blurred decision boundaries for some categories; AdaFace assumes that the number of feature paradigms is highly correlated with the quality of the image, but the actual scenario may have low feature paradigms due to illumination or pose, which leads to the failure of the adaptive mechanism.

In the future loss function design, we should pay more attention to the practical applicability of the loss function, so that face recognition can be put into a wider range of applications. To this end, we can incorporate uncertainty modelling into the design of the loss function, allowing the model to output low confidence on unknown categories instead of forcing classification. For example, Evidential Deep Learning predicts uncertainty through Dirichlet distribution modelling to reduce the false positive rate in open-set face recognition [10].

Moreover, most of these methods rely on large-scale data, which is not done accurately and efficiently for images with fewer samples. Xinyu Zhang,Mingyu You,Jiang Zhu,Xuan Han proposed a face recognition algorithm based on joint loss function for small scale data in 2020 [11]. The algorithm was pre-trained on a large-scale public face dataset and then re-trained on a small-scale dataset to improve the feature representation of the model and reduce overfitting.

In addition, it uses traditional feature post-processing methods for comparative evaluation. The experimental results show that the method can significantly improve the retrieval accuracy of the model on the face dataset of school freshmen. Therefore, the design and optimisation of the self-supervised pre-training framework should be strengthened in the future to perform generic learning face representation with unlabelled data, together with a small number of samples for downstream tasks.

5 Conclusions

In this paper, the application of loss functions in face recognition for recognition accuracy improvement is investigated using generation-by-generation analysis and comparison of different loss functions.

The paper's concludes that the blurring of softmax loss at the decision boundary is solved by ArcFace loss through additional angular edge penalty; the problem of high-quality and low-quality samples being treated equally in the feature space brought about by the fixed interval parameter m of ArcFace loss is solved by AdaFace loss through the adaptive margin formula. This is where the loss function generation is iterated according to the main problem of the former. In the future, enhancing the utility of loss functions in real life and reducing the dependence of loss function training on large-scale data are two important issues.

Therefore, it is necessary to adapt to open world and dynamic scenarios, incorporate uncertainty modelling into the design of loss functions, and enhance the design and optimization of self-supervised pre-training frameworks.

In the future, can be based on face recognition technology, combined with anti-jamming strategies and multimodal technology, so as to improve the wide application of face recognition technology in complex scenes.

Reference

1. Yibo Wang,et al. "Research on small-sample face recognition algorithms for pairwise anti-jamming." Internet of Things Technology 15.04(2025):33-36.
2. Hangzhou University of Electronic Science and Technology,CLP Data Service Co..." An occluded face recognition method based on FaceNet and attention mechanism." CN119559681A.2025-03-04.
3. Jiankang Deng, Jia Guo,Niannan Xue,et al. ArcFace: Additive Angular Margin Loss for Deep Face Recognition[J].IEEE Xplore.2020.
4. AdaFace: Quality Adaptive Margin for Face Recognition[J].IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022.
5. D. S. Trigueros, L. Meng, and M. Hartnett, "Face Recognition: from Traditional to Deep Learning Methods," Laser & Optoelectronics Progress, vol. . 58, no. 14, p. 1415008, 2021.
6. Ren Z.W., Liu X.Y., Xu J. et al. LittleFaceNet: a small-size face recognition method based on RetinaFace and AdaFace[J]. Journal of Computing, 2023.

7. Tingshuo Hu, Xiqu Chen. Research on face recognition based on MTCNN and ArcFace[J]. Journal of Automation, 2021.
8. Yue-Wen Hu,Dan Li. Custodial security face recognition based on ArcFace loss function[J]. Modern Information Technology,2020.
9. Chazufushui, Bai Meijuan, Wei Yongyong. Pedestrian multiple wandering detection algorithm based on pedestrian detection with peak density clustering[J]. Pattern Recognition and Artificial Intelligence, 2022.
10. Ren J ,Wen J ,Zhao Z , et al. Uncertainty-Aware Deep Learning: a Promising Tool for Trustworthy Fault Diagnosis[J].IEEE/CAA Journal of Automatica Sinica,2024,11(06):1317-1330.
11. Xinyu Zhang, Mingyu You, Jiang Zhu et al. Face recognition based on joint loss function for small-scale data[J]. Journal of Beijing Institute of Technology,2020.