Impact Analysis of Different Effective Loss Functions by Using Deep Convolutional Neural Network for Face Recognition

Anh D. Nguyen[1], Dat T. Nguyen[1], Hai N. Dao[2](✉), Hai H. Le[1], and Nam Q. Tran[3]

[1] VNU University of Engineering and Technology, Hanoi, Vietnam
{19021208,19020126,hailh}@vnu.edu.vn
[2] VNU Information Technology Institute, Hanoi, Vietnam
namhai249@vnu.edu.vn
[3] Dai Nam University, Hanoi, Vietnam

Abstract. Smart Library Automation System is increasingly attractive as an effective digital library system. Adapting automation in the library helps reduce the duplication of work, time-saving, and boosts work efficiency. One such significant feature is face recognition integrated. With this application, the system can use face recognition to enter and get the details of an end user. In recent years, face recognition has achieved many prodigious accomplishments based on Deep Convolutional Neural Networks (DCNN). In addition to constructing large face datasets, designing new effective DCNN architectures and loss functions are two trends to improve the performance of face recognition systems. Therefore, many studies have been published in state-of-the-art methods, making high-accuracy face recognition systems more possible than days in the past. However, it is still difficult for all research communities to train robust face recognition models because it depends heavily on their resources. This paper investigates and analyzes the effect of several effective loss functions based on softmax. Moreover, we also evaluate how hyper-parameter settings can impact the optimization process as well as the final recognition performance of the model trained by re-implementing these methods. The results of our experiments achieve state-of-the-art figures, which show the proposed method's massive potential in improving face recognition performance.

Keywords: Face recognition · DCNN · Effective loss function

1 Introduction

Face recognition (FR) has been a prominent and long-standing topic in the research community. It is widely applied in numerous areas such as social security, health, education, banking, retail, etc. FR plays a crucial role in automated library systems with a digital library. FR applications help save time and

Y.-H. Tseng et al. (Eds.): ICADL 2022, LNCS 13636, pp. 101–111, 2022.
https://doi.org/10.1007/978-3-031-21756-2_8

avoid repetitive book issuing and returning work. Since Convolutional Neural Networks (CNN) have been proposed in recent years, the literature has witnessed the innovation of Deep Convolutional Neural Networks (DCNN) architectures that achieve remarkable results consecutively. There are many modern face datasets, such as MegaFace [10], MS1M [5], WebFace20M [26], ... which encompass a huge amount of identities and samples are released. It creates many chances for researchers to train effectively large and deep face recognition networks based DCNN such as [1,2,6,16] which present better performance than the human capability in FR. However, it is not enough to satisfy the expectation of the research community. Therefore, most recent researchers have focused on improving the classification loss function, which plays a crucial role in learning accurate FR models. Current popular loss functions for training DCNNs are mostly softmax-based classification loss. Since the learned features with original softmax loss cannot maximize inter-class variance and minimize the intra-class variance of embedding feature vectors, researchers try to design new effective loss functions based on it, which enhance discriminative power as well as still remain basic requirements of softmax. [15,21,23] have proposed some effective methods and gradually improved the accuracy by elaborating the objective of learning.

Challenges: The modern methods for augmenting training datasets and optimizing CNN architectures lead to significant achievements nowadays. However, most FR models that get State-of-the-Art (SOTA) accuracy have been trained on cost-strong computing hardware systems for a long time by experts. Unfortunately, the academic community is not able to access these resources. Many of them pragmatically own limited computational competence systems. In order to construct the FR system from scratch, they have to adopt medium-size datasets, customize DCNN architectures, and adjust the hyper-parameter settings for training. All of these changes can lead to unexpected results. Choosing appreciate loss functions is a easy way in order to improve performance of FR models. On the other hand, academic community makes the effort to design an effective loss function. Each loss function requires distinct constraints of data pre-processing, modifying CNN architectures,... It is necessary to select loss functions that are able to bring higher performance with respective lower implementation costs.

Contributions: From the aforementioned challenges, in this paper, we investigate state-of-the-art losses based on softmax, including CosFace, ArcFace, and MagFace to verify and understand the efficiency of these loss function for training deep neuron networks. Each function generally includes several hyper-parameters, which substantially impact the final performance and is usually difficult to tune. We conducted evaluation on several face benchmarks, including LFW, CFP-FP, AgeDB-30, CALFW and CPLFW.

The remainder of this paper is organized as follows. In §2, we describe our main problem and introduce some related work in FR. §3 presents our proposal. In §4, we show our experiments, evaluate our proposed work, and provide some comparisons. Finally, §5 gives the conclusions and future work.

2 Problem Formulation and Related Work

2.1 Face Recognition

Face recognition aims to match an image of someone's face to all their representations stored in the database. Typically, most FR systems have three components, primarily face detection, feature extraction, and classification. Face detection is a preprocessing stage that detects a unique face and aligns a face in the image. Next, feature extraction, which takes input as preprocessed images, adopts a DCNN to extract deep discriminative features. It returns an embedding vector as the feature representation of a face.

The classification process contains precisely two sub-processes: *(i) face verification:* calculate the one-to-one similarity between two images to determine where those pairs belong to the identical individual, and *(ii) face identification:* seek an individual by one-to-many computing similarities between a probe and all images stored in a database to identify the specific identity of a face among a set of the facial gallery. This paper mainly focuses on feature extraction and classification in face verification tasks.

2.2 Loss Function for Deep Face Recognition

Up to now, prevailing loss functions for deep FR are mainly based on variants of softmax loss. To address a defect in the pure softmax loss, which does not run well in reducing the intra-class variation (i.e., making features of the same class compact) [19], one of the most effective approaches obtained excellent performance on FR is to add the margin into the primitive loss. We can split it into two ways:

Adding Fixed Margin into Softmax. Several efforts have been proposed to enhance the discriminative power of the softmax loss by adding a fixed penalty margin for training FR models [12,19]. Margin is added to implement the constraint: the maximum inter-class distance < the minimum inter-class distance + margin. Then, learned features can be sufficiently discriminative. These methods succeed in enforcing intra-class compactness and inter-class discrepancy to improve FR performance. At the same time, they can be implemented easily with several uncomplicated code lines in the deep learning frameworks such as Pytorch or Tensorflow.

Adding Adaptive Margin into Softmax. Although the above methods have reached the SOTA performance on a number of benchmarks, they still remain a handful of drawbacks. One of them is that low-quality images may easily impact the performance of these losses. Thus, more recently, many researchers have proposed an adaptive margin strategy to automatically tune those hyper-parameters to avoid putting too much faith in the fixed margin and generate more effective supervision during training. Some of them improve inherent fixed margin-base softmax loss functions to be more flexible [9,23]. In addition to the adaptive margin, some others also proposed adding extra factors into the function as [13] for enhancing final recognition performance. Adding adaptive margin approach attempts to adopt real data with an inconsistent inter-and intra-class variation. It might limit the discriminative power and generalizability of the FR model when using fixed marginal penalty softmax losses.

3 Analysis Approach

3.1 DCNN Architecture

Training facial recognition models can be divided into two stages: training and testing. Specifically, as illustrated in Fig. 1, face images from both training set (a) and testing set (d) are processed to handle variations before feeding into feature extraction. This process serves a purpose to alleviate the effect of the environment's condition on the performance of FR. During the training stage, a DCNN (b) is utilized to extract deep discriminative features, DCNN returns an embedding vector as a feature representation of a face. Like training stage, these embedding vectors are handled in the testing stages. However, the classification layer in the DCNN is often discarded at the testing stage. Next, during the training stage, these features will be applied to a loss function (c) which optimizes deep discriminative features. That loss function is adopted to learn the FR model accurately. After the in-depth features are extracted, face matching methods (e) come to conduct the feature classification process. While face matching methods take advantage of cosine or L2 distance, they are employed to compute similarity scores among test images in the testing stage. All our experiments are conducted based on the architecture that we demonstrated above.

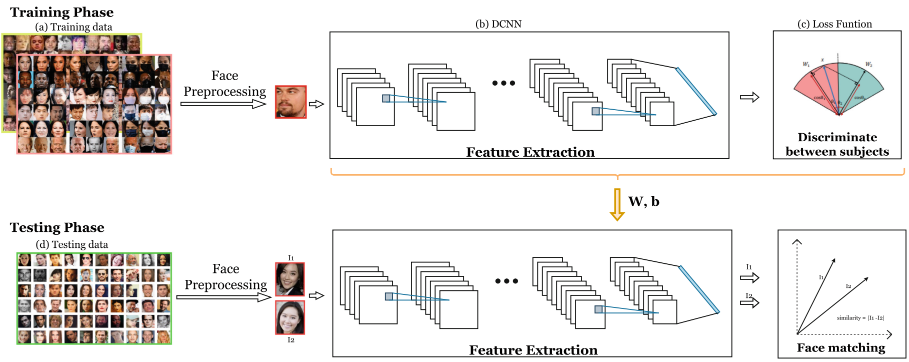

Fig. 1. Overview of training Deep Face Recognition model

Back to the problem mentioned earlier, we have to adopt the typical DCNN architecture that is ResNet50 [6] as a feature extractor and use the moderate public dataset, CASIA-WEBFACE in the training stage to be compatible with our resources. Despite being competent for these possibilities, building a facial recognition system with superior performance to humans is like [18], and there is a significant disparity between this system's accuracy and SOTA solutions. Achievements in studies of the loss function encourage us to apply these accomplishments to our situation as we search for ways to enhance our FR performance without additional hardware setup requirements. As a result, the fundamental direction we focus on in this paper is applying various loss functions to the training model in order to enhance FR performance and achieve SOTA results.

3.2 Effective Loss Function Analysis

As mentioned in 2.2, the margin-based softmax loss functions are now widely used for training FR models. Many new loss functions are published with SOTA performance for a short period. However, not all of them can easily implement or improve accuracy with our resources. Thus, we considered some functions with two criteria: (i) do not require intricately prepared samples before being fed into training; and (ii) avoid refining backbone unnecessarily and adding extra layers excessively.

Based on these criteria, we determined to use three loss functions comprise CosFace [20], ArcFace [3] and MagFace [13] for learning of DCNN model. Ideologically, to improve the discriminative competence, these loss functions implement a margin penalty on primitive softmax loss. The difference is that CosFace and ArcFace add a fixed margin while MagFace belongs to a category that adds adaptive margin into softmax.

In detail, let $\mathcal{D} = \{x_i, y_i\}_{i=1}^{N}$ where x_i denotes an input image with its corresponding label y_i and N is the number of images. In the training stage, after being fed an image, the last fully connected layer of the neuron network returns a d-dimensional embedding vector $f_i \in \mathbb{R}^d$. By defining the angle θ_j between f_i and j-th class center $w_j \in \mathbb{R}^d$ as $w_j^T f_i = \|w_j\| \, \|f_i\| \cos\theta_j$, CosFace, ArcFace can be formulated as follows:

$$\mathcal{L}_{CosFace} = -\log \frac{e^{s\left(\cos\theta_{y_i} - m\right)}}{e^{s\left(\cos\theta_{y_i} - m\right)} + \sum_{j\neq y_i} e^{s\cos\theta_j}} \tag{1}$$

$$\mathcal{L}_{ArcFace} = -\log \frac{e^{s\cos\left(\theta_{y_i} + m\right)}}{e^{s\cos\left(\theta_{y_i} + m\right)} + \sum_{j\neq y_i} e^{s\cos\theta_j}} \tag{2}$$

where m denotes the additive angular margin and s is the scaling parameter multiplied due to the norm of w and x in the tenet of all these loss functions. It is two key factors added into the Softmax to enhance performance. Specifically, we fix $||w_j|| = 1$ by L_2 normalization and presume that $||f_i|| = s$ is fixed [20]. Compared with the Softmax, margin-based variants of the Softmax loss like ArcFace and CosFace extends the decision boundary between different classes in the cosine space by a specified m and the parameter s scales up the narrow range of cosine distances, making the logits more discriminative.

In contrast, for MagFace, instead of using a specified scalar parameter as margin, it introduces two auxiliary functions related to the magnitude $a_i = ||f_i||$

without normalization of each feature f_i, the magnitude-aware angular margin $m(a_i)$ and the regularizer $g(a_i)$ following a natural institute: High norm features are easily recognizable and large margin pushes features of high norm closer to class centers. With all previous notions, we define MagFace like:

$$\mathcal{L}_{MagFace} = -\log \frac{e^{s\cos\left(\theta_{y_i}+m(a_i)\right)}}{e^{s\cos\left(\theta_{y_i}+m(a_i)\right)} + \sum_{j\neq y_i} e^{s\cos\theta_j}} + \lambda_g g\left(a_i\right) \tag{3}$$

λ_g indicates regularization losses weights. Two extra functions, $m(.)$ and $g(.)$, are convex functions that are presented in the MagFace's supplementary [13]. Due to word count constraints, we will not be able to cover everything in this paper, so we recommend reading the original paper for setting details. Based on Eqs. 1, 2 and 3, it is easy to implement three loss functions through only adjusting logits before passing it into softmax-cross entropy loss [3].

4 Evaluation

We investigate the impact analysis of several effective loss functions by performing the deep experiment of FR using DCNN models.

4.1 Dataset Preparation

For training, we choose CASIA-WEBFACE [22] containing 494,414 images of 10,575 different individuals. For evaluating the performance of DCNN-based FR, we choose several popular datasets such as LFW [7], CFP-FP [17], AgeDB-30 [14], CALFW [25] and CPLFW [24]. In there, LFW, CFP-FP, and AgeDB-30 are considered easy benchmarks and conversely, CALFW and CPLFW are higher challenges because the image quality is considered the problematic benchmark. For convenience, we fetch all datasets from the available source, InsightFace (*https://github.com/deepinsight/insightface*), in which all images are aligned to 112×112 in as the settings in ArcFace [3].

4.2 Model Training Configuration

DCNN Model Setting: All the DCNN models are trained from scratch. We implemented a widely used CNN architecture, namely ResNet50, as a backbone network. For training, we only augment training samples by random horizontal flip. Each model is trained for 32 epochs with a batch size of 128. The initial learning rate is set to 0.025 and divided by ten at 18, 28 epochs when the training loss plateaus. SGD is adopted as our optimizer with momentum is 0.9, and the weight decay parameter is set to $5e-4$. The drop ratio is fixed at 0.5 for all settings. All experiments are conducted based on Pytorch framework [8] and all the models are trained and validated for training and testing on an NVIDIA Tesla T4 (16GB) GPU.

Loss Function Hyper-parameters Setting: In ArcFace and CosFace, the margin m is choose at 0.5 and 0.35 respectively as suggested in original papers. For MagFace, there are five hyper-parameters must be selected thoroughly, as mentioned in [13]. The upper bound and lower bound of magnitude are fixed as $l_a = 10$, $u_a = 110$. Besides that, function $m(a_i)$ is described as a function defined on $[l_a, u_a]$ with $m(l_a) = l_m$, $m(u_a) = u_m$. By empirical experiment, l_m and u_m are set respectively to 0.45 and 0.8. λ_g is specified to 20 as well. The scale s has already been discussed sufficiently in several previous works [15,20], so we directly fixed its values to 64 and will not discuss its effect further.

4.3 Results

Intending to analyze the effect of the loss function in the FR DCNN model, we have built an experiment according to the method presented in Sect. 3 with the above settings. The accuracy [11] and equal error rate (ERR) metrics [4] are used to evaluate and compare our model again other SOTA. The results of specific evaluation and comparison of the accuracy of DCNN models with different loss functions and baselines are presented below:

Table 1. Accuracy metric (%) of DCNN models on five benchmarks ("*" indicates the result quoted from the original paper)

Method	LFW	CFP-FP	AgeDB-30	CALFW	CPLFW
HUMAN-Individual	97.27	–	–	82.32	81.21
HUMAN-Fusion	–	–	–	86.50	85.24
CASIA, R50, CosFace* [20]	99.33	–	–	–	–
CASIA, R50, ArcFace* [3]	99.53	95.56	95.15	–	–
MS1M-V2 , R100, MagFace* [13]	99.83	98.46	98.17	96.16	92.87
CASIA, R50, Softmax (our)	98.82	93.59	91.13	91.15	86.02
CASIA, R50, CosFace (our)	99.42	95.01	94.45	93.35	89.47
CASIA, R50, ArcFace (our)	99.43	95.12	94.88	93.88	88.98
CASIA, R50, MagFace (our)	99.43	95.11	94.55	93.82	89.27

Table 2. EER (%) of effective loss functions on all test datasets

Method	LFW	CFP-FP	AgeDB-30	CALFW	CPLFW
CosFace (our)	0.65	5.98	5.60	7.42	**12.25**
ArcFace (our)	**0.63**	6.18	**5.45**	**7.20**	12.76
MagFace (our)	0.68	**5.90**	5.68	7.23	12.70

Comparison between Effective Loss Functions and Softmax. For further understanding of the effectiveness of adding margin into softmax, we also implemented the original softmax function by the side of three proposed losses namely CosFace, ArcFace and MagFace. Results are presented in Table 1. As a baseline model, Softmax reached 98.82%, 93.59%, 91.13%, 91.15% and 86.02% accuracy on all benchmarks are LFW, CFP-FP, Age-30, CALFW and CPLFW respectively. For all other losses, they outperformed Softmax on every benchmark. Compared Softmax, CosFace archived 0.6%, 1.42%, 3.32%, 2.20%, 3.45% improvements on the respective LFW, CFP-FP, Age-30, CALFW and CPLFW. ArcFace also surpasses the baseline by 0.61%, 1.53%, 3.75%, 2.73% and 2.96% on five benchmarks respectively. As well as two previous losses, MagFace outperformed Softmax on all test datasets 0.61%, 1.52%, 3.42%, 2.67%, 3.25% consecutively. Those results clearly shown that the margin plays a key role in CosFace, ArcFace and MagFace, strengthening the discriminating power of features. Besides, our implementations for all three losses can lead to convergence without observing any difficulty similar to the advantage of Softmax.

Comparison between Effective Loss Functions. Not only compare to Softmax, we also compare among our re-implementations including CosFace, ArcFace, MagFace based EER metrics in Table 2. The result returned varied on the different benchmarks. It is interesting that CosFace showed the best result on the “hard test case”, CPLFW, 12.25%, while performed more poorly on the rest of datasets. Otherwise it not surpass ArcFace and MagFace on other benchmarks. As opposed to what we had anticipated, MagFace only performed well in the CFP-FP benchmark and is good in CALFW. On the other hand, ArcFace is archived at its best in diverse age benchmarks: on AgeDB-30, it gains 5.45% EER, which is 0.15% lower than CosFace and 0.23% lower than MagFace. Similarly, on LFW and CALFW, ArcFace also archives EER values which is lower than CosFace and MagFace approximately about (0.03%–0.2%). It illuminates that ArcFace can bring the discriminative power better than CosFace and MagFace in our scenario.

Comparison with other Baselines. Besides Softmax, we also compare our re-implementations with other published results : Human-Individual and Human-Fusion as reported in Table 1. Firstly, our models easily outperformed HUMAN-Individual on CALFW and CPLFW. They achieved accuracy better than 88% on all of these benchmarks, whereas Human-Individual only achieved performance less than 83% on both. Even on the easy benchmark LFW, our models completely defeat competence of a individual about 2%. Moreover, compared to Human-Fusion which is the third baseline, our models still did outperform it by increasing about 3.0%–7.0% on CALFW and CPLFW. Those results shown us the power of our proposed models that definitely replaces human in FR tasks.

4.4 Limitations

Apart from our achievements, we also reported the limitation of our re-implementations compare to results in original papers. Although our CosFace-based DCNN model achieves even higher than the original CosFace* 0.09% for ACC, both ArcFace and MagFace re-implementations could not reach the same results as the original papers. For the ArcFace-based DCNN model, our method has 99.43% of accuracy on LFW, which is 0.1% less than the ArcFace* as well as CFP-FP and AgeDB-30. The difference in accuracy between our models and the original ones comes mainly from the difference in computing infrastructure. For instance, the ArcFace* used 4 GPU and set the batch size to 512 for training the model. Finally, with MagFace*, they used larger datasets and a deeper backbone network, so their results ultimately outperform others in Table 1.

5 Conclusion

In this paper, we have surveyed, analyzed, and identified the main challenges of CNN-based face recognition concerning specifying the loss function. To tackle these challenges independently, we re-implement and evaluate the three novel effective loss functions, including CosFace, ArcFace, and MagFace. We demonstrated the advantages of those loss functions, which perform better than their original Softmax. Our deep experiments on recent popular datasets confirm that the face recognition accuracy is clearly improved by using a suitable loss function. Our re-implementations show significant results comparing other baselines on five widely used benchmarks: LFW, CFP-FP, AgeDB-30, CALFW, and CPLFW. Hence, it asserted the discriminative power of CosFace, ArcFace, and MagFace, respectively, and the potential for deployment in the automated library system.

In future work, we will also be interested in other types of loss functions and ensemble learning methods in order to build robust face recognition applications capable of discriminating effectively in varied environments.

References

1. Baltrusaitis, T., Zadeh, A., Lim, Y.C., Morency, L.-P.: OpenFace 2.0: facial behavior analysis toolkit. In 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), pp. 59–66 (2018)
2. Cao, Q., Shen, L., Xie, W., Parkhi, O.M., Zisserman, A.: VGGFace2: a dataset for recognising faces across pose and age. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), pp. 67–74 (2018)
3. Deng, J., Guo, J., Xue, N., Zafeiriou, S.: ArcFace: additive angular margin loss for deep face recognition. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4685–4694 (2019)
4. Du, H.P., Pham, D.H., Nguyen, H.N.: An efficient parallel method for optimizing concurrent operations on social networks. Trans. Comput. Collective Intell. 10840(XXIX), 182–199 (2018)
5. Guo, Y., Zhang, L., Hu, Y., He, X., Gao, J.: Ms-celeb-1m: a dataset and benchmark for large-scale face recognition. In: ECCV (2016)
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
7. Huang, G.B., Mattar, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Technical Report 07–49, University of Massachusetts, Amherst (2007)

8. Huang, X., Du, X., Liu, H., Zang, W.: A research on face recognition open source development framework based on PyTorch. In: 2021 International Symposium on Computer Technology and Information Science (ISCTIS), pp. 346–350 (2021)
9. Jiao, J., Liu, W., Mo, Y., Jiao, J., Deng, Z., Chen, X.: Dyn-ArcFace: dynamic additive angular margin loss for deep face recognition. Multimedia Tools Appl. **80**(17), 25741–25756 (2021)
10. Kemelmacher-Shlizerman, I., Seitz, S.M., Miller, D., Brossard, E.: The megaface benchmark: 1 million faces for recognition at scale. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4873–4882 (2016)
11. Le, H.V., Nguyen, T.N., Nguyen, H.N., Le, L.: An efficient hybrid webshell detection method for webserver of marine transportation systems. IEEE Trans. Intell. Transp. Syst. Early Access, 1–13 (2021)
12. Liu, W., Wen, Y., Yu, Z., Yang, M.: Large-margin softmax loss for convolutional neural networks. In: Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48, ICML'16, pp. 507–516 (2016)
13. Meng, Q., Zhao, S., Huang, Z., Zhou, F.: MagFace: a universal representation for face recognition and quality assessment. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 14220–14229 (2021)
14. Moschoglou, S., Papaioannou, A., Sagonas, C., Deng, J., Kotsia, I., Zafeiriou, S: AgeDB: the first manually collected, in-the-wild age database. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1997–2005 (2017)
15. Ranjan, R., Castillo, C., Chellappa, R.: L2-constrained softmax loss for discriminative face verification. CoRR, abs/1703.09507 (2017)
16. Schroff, F., Kalenichenko, D., Philbin, J.: FaceNet: a unified embedding for face recognition and clustering. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 815–823 (2015)
17. Sengupta, S., Chen, J.-C., Castillo, C., Patel, V.M., Chellappa, R., Jacobs, D.W.: Frontal to profile face verification in the wild. In: 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1–9 (2016)
18. Tao, K., He, Y., Chen, C.: Design of face recognition system based on convolutional neural network. In: 2019 Chinese Automation Congress (CAC), pp. 5403–5406 (2019)
19. Wang, F., Cheng, J., Liu, W., Liu, H.: Additive margin softmax for face verification. IEEE Signal Process. Lett. **25**(7), 926–930 (2018)
20. Wang, H., et al.: CosFace: large margin cosine loss for deep face recognition. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5265–5274 (2018)
21. Wen, Y., Zhang, K., Li, Z., Qiao, Yu.: A discriminative feature learning approach for deep face recognition. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9911, pp. 499–515. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46478-7_31
22. Yi, D., Lei, Z., Liao, S., Li, S.Z.: Learning face representation from scratch (2014)
23. Zhang, X., Zhao, R., Qiao, Y., Wang, X., Li, H.: AdaCos: adaptively scaling cosine logits for effectively learning deep face representations. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10815–10824 (2019)
24. Zheng, T., Deng, W.: Cross-pose LFW: a database for studying cross-pose face recognition in unconstrained environments. Technical Report 18–01, Beijing University of Posts and Telecommunications (2018)

25. Zheng, T., Deng, W., Hu, J.: Cross-age LFW: a database for studying cross-age face recognition in unconstrained environments. CoRR, abs/1708.08197 (2017)
26. Zhu, Z., et al.: Webface260m: a benchmark unveiling the power of million-scale deep face recognition. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10487–10497 (2021)