Proceedings of CONF-SPML 2026 Symposium: The 2nd Neural Computing and Applications Workshop 2025
DOI: 10.54254/2755-2721/2026.TJ31675

Convolutional Neural Network Based on Deep Learning Overview and Performance Optimization Research of Facial Recognition Algorithms

Yubo Lou

Department of Information Technology, Xi'ankedagaoxin University, Xi'an, China
lyguangbo@126.com

Abstract. Convolutional Neural Networks (CNNs) have been widely applied to face recognition tasks and have achieved remarkable results. However, the developmental trajectory of convolutional neural networks and the criteria for determining which algorithms can more effectively ensure recognition accuracy remain insufficiently clear. Therefore, a comparative analysis of CNN-based models is conducted. The results indicate that ElasticFace performs best on long-tailed data distributions, while MobileFaceNet is the most advantageous in edge computing scenarios.

Keywords: Convolutional Neural Networks (CNN), Face Recognition, Deep Learning

1. Introduction

Face recognition has been widely applied in various fields, such as public security, identity verification, intelligent surveillance, and human-computer interaction. Face feature extraction is a crucial step in face recognition, and effective extraction is the key to improving recognition accuracy. In recent years, Convolutional Neural Networks (CNNs) have demonstrated significant advantages in image recognition, gradually becoming the core technology in face recognition systems. Previously, most traditional face recognition methods relied on handcrafted feature extraction techniques, such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Local Binary Patterns (LBP). These methods typically yielded lower recognition accuracy when affected by variations in lighting, pose, and occlusion. In contrast, CNN-based deep learning approaches are capable of automatically learning more discriminative and hierarchical feature representations from large-scale data, significantly improving robustness and accuracy. Consequently, applying CNN-based face recognition algorithms, and optimizing their structures and training processes, has become a key path toward enhancing recognition performance.

2. Development review

In recent years, deep CNNs have achieved breakthrough progress in the field of face recognition. Many classic systems (e.g., DeepFace [1], DeepID, FaceNet [2]) have demonstrated high-performance recognition results. With the growth of large-scale training datasets and computing

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.TJ31675

power, researchers have primarily focused on designing new loss functions and lightweight network structures to further enhance discriminative ability and efficiency. Below, we introduce several mainstream models, their architectures, and key improvements.

2.1. Mainstream CNN models and architectural features

ArcFace (CVPR 2019): ArcFace [3] proposed the Additive Angular Margin Loss for face recognition. Its core lies in the normalized feature space (typically represented by a 512-dimensional vector), where a fixed angular margin is added to the target class, directly optimizing inter-class separation on the hyperspherical manifold. In practice, ArcFace generally adopts deep residual networks (e.g., ResNet-50/100) to output L2-normalized feature vectors, followed by the ArcFace loss layer during training. This method provides a clear geometric interpretation and achieved state-of-the-art performance on multiple large-scale image and video face recognition benchmarks. For instance, across more than ten public datasets (including billion-scale comparison tasks), ArcFace outperformed prior approaches.

CosFace (CVPR 2018): CosFace [4] introduced the Large Margin Cosine Loss (LMCL), which normalizes both features and class weights via L2 normalization before the softmax layer and then adds a fixed cosine margin to the target class. This maximizes inter-class decision boundaries while minimizing intra-class variations. CosFace achieved state-of-the-art results at the time on standard datasets such as LFW and YouTube Faces. The method is simple to implement, requiring only a constant term added to the logit, and does not depend on auxiliary losses for stable training.

SphereFace (2017): SphereFace [5] adopts standard CNN architectures (e.g., ResNet) as the backbone to extract image features. By imposing a multiplicative angular margin ($\cos(m\theta)$) on the normalized hyperspherical feature space, the method enforces stricter angular separation ($m > 1$) between classes. Compared with standard softmax or normalization-based losses (e.g., NormFace/Softmax Loss), SphereFace produces more discriminative angular features and decision boundaries.

MobileFaceNet (2018): MobileFaceNet [6] is a lightweight CNN model specifically designed for mobile devices, with fewer than 1M parameters. Compared to general-purpose models such as MobileNetV2, MobileFaceNet introduces optimizations tailored to facial feature extraction, such as depthwise separable convolutions and global depthwise convolution modules. These significantly reduce model size and computation. Combined with ArcFace loss during training, a 4MB MobileFaceNet model achieved 99.55% verification accuracy on LFW and 92.59% TAR (at FAR=1e-6) on MegaFace retrieval, comparable to much larger networks while offering superior inference speed.

ElasticFace (CVPRW 2022): ElasticFace [7] proposed the Elastic Margin Loss, which introduces a randomly sampled margin (drawn from a normal distribution) during each training iteration, unlike ArcFace/CosFace that use fixed margins. This flexible approach allows decision boundaries to "stretch" and "shrink" to better fit the non-uniform intra-class and inter-class variations in real data. Experimental results showed that ElasticFace surpassed fixed-margin methods (ArcFace, CosFace) across most benchmarks, achieving superior performance on seven widely used face recognition datasets. In summary, ElasticFace alleviates the limitations of fixed margins, thereby improving model generalization.

MagFace (CVPR 2021): MagFace [8] introduced a dynamic margin mechanism that adapts to sample quality. Unlike purely direction-based normalized features, MagFace incorporates feature norm (magnitude) to measure image quality (e.g., sharpness, occlusion). During training, high-quality samples are assigned larger norms and wider angular margins, while low-quality samples are

DOI: 10.54254/2755-2721/2026.TJ31675

pushed toward the sphere center to avoid overfitting noise. As a result, embeddings become not only more discriminative but also quality-aware. Experiments demonstrated that MagFace achieved leading performance on multiple face recognition benchmarks (e.g., 99.83% accuracy on LFW) and enabled the use of feature quality information for clustering and recognition tasks.(As shown in Table 1)

Table 1. Comparison of facial recognition models

Method	Core Features	Advantage	Disadvantage	Typical Applications scene
ArcFace	Based on Additive Angular Margin, optimize classification boundaries directly in the angular space.	Compact within class and significant separation effect between classes; Stable training and fast convergence; Widely applicable, SOTA benchmark model	The hyperparameter (margin value) needs to be finely tuned; Limited robustness to extreme poses/occlusions	Universal facial recognition, facial verification, 1: N retrieval
CosFace	Introducing Additive Cosine Margin to increase inter class distance in cosine space.	Efficient computation (pure cosine operation); Strong robustness to noisy data; Similar to ArcFace performance, with better performance in some scenarios	Joint adjustment of margin and temperature coefficient is required; Strong feature normalization dependency	Large scale facial recognition, low-quality image scenes
SphereFace	The first proposal was to introduce Diversified Angular Margin, which optimizes the angular space through weight normalization.	Pioneering work to drive the development of Margin based Loss; Strong explanatory power from a theoretical perspective	Unstable training (requires segmented training/Softmax assistance); Actual performance is often surpassed by ArcFace/CosFace; Hyperparameter sensitivity	Academic research, early facial recognition systems
MobileFaceNet	Lightweight network architecture (based on MobileNetV2) optimized for parameter and computational complexity on mobile devices	Extremely low computational cost (1M parameters,<0.5G FLOPs); Real time inference (mobile<10ms); Balance accuracy and efficiency	The accuracy is lower than that of large models (such as ResNet100+ArcFace); Limited ability to express features, difficult to handle complex scenes	Mobile devices, embedded systems, edge computing
ElasticFace	Propose Elastic Margin penalty to dynamically adjust the margin values of different samples.	Adaptive difficult sample mining; Enhance intra class diversity expression; More robust performance in unconstrained scenes (occlusion/pose)	High training complexity; Dependency distribution estimation strategy (such as Gaussian mixture model); Parameter tuning is more complex	Surveillance security, non cooperative facial recognition
MagFace	Jointly optimize feature amplitude and angle, and use feature module length to reflect sample quality (clarity/aspect ratio).	The feature length directly indicates the sample quality; Significant improvement in robustness to low-quality images; Support quality perception recognition (weighted matching)	Training requires additional quality supervision (implicit/explicit); During deployment, it is necessary to retain module length information; The computational cost is slightly higher than ArcFace	Low quality facial recognition, video surveillance, access control system

2.2. Performance comparison of different models

In evaluating face recognition models, commonly used public benchmark datasets include LFW, MegaFace, IJB-B, and IJB-C. Among these, LFW is primarily used for face verification and reports mean verification accuracy; MegaFace, which contains a million-level gallery, evaluates both recognition (Rank-1 accuracy) and verification (TAR@FAR=1e-6); IJB-B/C provide more challenging protocols, typically reporting True Accept Rate (TAR) at specified False Accept Rates (FAR). Representative performance comparisons of different models on these datasets are summarized below:

LFW Verification Accuracy: With the development of algorithms, LFW accuracy has approached saturation. ArcFace, trained on the refined MS-Celeb-1M dataset (MS1MV2), achieves

DOI: 10.54254/2755-2721/2026.TJ31675

approximately 99.82% [3]; CosFace reaches around 99.73% [4]; MagFace also attains about 99.8% [8]. Even lightweight models such as MobileFaceNet (4MB), when trained with ArcFace, achieve 99.55% [6. Recently proposed methods such as ElasticFace similarly surpass 99.8% [7]. This indicates that while LFW verification accuracy is no longer sufficient to distinguish state-of-the-art methods, it remains useful for reflecting baseline model effectiveness.

MegaFace Recognition/Verification Performance: In the million-scale gallery recognition task (Rank-1 accuracy), ArcFace achieves about 81.0% [3], while CosFace attains approximately 82.7% [4] (results depend on the training dataset). Improved variants of ElasticFace also achieve over 80% on MegaFace [8]. In the verification task (TAR@FAR=1e-6), ArcFace records about 96.98% [3], while CosFace achieves around 96.65% [4]. By comparison, MobileFaceNet attains 92.59% TAR at FAR=1e-6 [6], significantly outperforming earlier large CNNs, demonstrating the advantages of its efficient architecture.

IJB-B/IJB-C True Accept Rates: On the more challenging IJB benchmarks, TAR@FAR=1e-4 is commonly reported. ArcFace achieves about 94.2% TAR on IJB-B and 95.6% on IJB-C [3]. ElasticFace further improves performance through its elastic margin mechanism, reaching approximately 95.4% and 96.7% TAR on IJB-B and IJB-C, respectively [8]. CosFace yields slightly lower TAR values [4]. Overall, the new generation of models consistently surpass 90% TAR on IJB benchmarks, with ElasticFace and similar methods delivering improved recognition rates under stringent low-FAR conditions.

2.3. Public benchmark datasets and evaluation metrics

The commonly used public benchmarks and their evaluation metrics are as follows: LFW [9] and AgeDB, among other verification datasets, report face verification accuracy; MegaFace [10] reports Top-1 accuracy for gallery retrieval as well as TAR@FAR=1e–6 for verification tasks; IJB-B [11] and IJB-C [12] report TAR@FAR=1e–4 for verification tasks. Models are often compared comprehensively through ROC curves (TPR vs. FPR) and the Area Under the Curve (AUC). These metrics reflect a model's ability to distinguish between positive and negative samples. For example, ElasticFace achieves a TAR of 95.43% on IJB-B at FAR=1e–4, clearly surpassing CosFace and CenterFace, which remain around ~95.0%.(As shown in Table 2)

Table 2. Introduction to test set

Test set	Challenge type	Sample size	Feature
LFW	Unconstrained face	6,000 pairs	Benchmark testing, relatively simple
AgeDB-30	Large age difference (5-100 years old)	12,240 pairs	Cross age verification
CALFW	Cross age+cross posture	3,000 pairs	LFW's age enhanced version
CPLFW	Cross stance (deflection, occlusion)	3,000 pairs	LFW's posture enhanced version
CFP-FP	Cross stance (front vs. side)	7,000 pairs	Extreme posture difference verification

2.4. Evaluation metrics: definitions and application scenarios

Precision: The proportion of correctly predicted positive samples among all samples predicted as positive, i.e.,$Precision=\frac{TP}{TP+FP}$.

In face verification, precision measures the probability that the system is correct when outputting "same identity" (a low false acceptance rate corresponds to high precision).

Recall (or True Positive Rate, TPR): The proportion of actual positive samples correctly identified as positive, i.e.,

DOI: 10.54254/2755-2721/2026.TJ31675

$Recall=TPR=\frac{TP}{TP+FN}$.

In face verification, recall reflects the system's coverage of successfully verifying image pairs from the same identity (a low false rejection rate corresponds to high recall).

False Positive Rate (FPR): The proportion of actual negative samples incorrectly classified as positive, i.e.,

$FPR=\frac{FP}{FP+TN}$.

In practical applications, FPR determines the security threshold: the lower the FPR, the lower the probability that the system mistakenly identifies two different individuals as the same person.

ROC Curve (Receiver Operating Characteristic Curve) [13]: A curve plotting TPR against FPR at various decision thresholds. The ROC curve intuitively demonstrates the trade-off in model performance across all possible thresholds: an ideal model's curve approaches the top-left corner, while the diagonal indicates random guessing performance. The Area Under the Curve (AUC) quantifies the overall discriminative ability of the model, where AUC = 1.0 denotes perfect separation and AUC = 0.5 denotes random guessing.

AUC (Area Under ROC Curve) [13]: The numerical value of the area under the ROC curve, which can be interpreted as the probability that a randomly chosen positive sample is ranked higher than a randomly chosen negative sample. In face recognition, AUC is commonly used to comprehensively evaluate the average model performance under different error rates.

Different metrics are suitable for different application scenarios. When positive and negative samples are imbalanced or when the costs of different types of errors vary significantly, the choice of evaluation metric should be tailored to actual requirements (e.g., security vs. convenience). In such cases, additional metrics such as TAR at fixed FAR or the precision-recall curve may be more appropriate.

2.5. Comparison and analysis of test data

Table 3. Model test data for LFW, AgeDB-30, CalFW, CPLFW, CFP-FP

Method	Training Dataset	LFW	AgeDB-30	CALFW	CPLFW	CFP-FP
		Accuracy(%)	Accuracy(%)	Accuracy(%)	Accuracy(%)	Accuracy(%)
ArcFace	MS1MV2 [3,14]	99.83	98.48	95.45	92.08	98.27
CosFace	MS1MV2 [3,14]	99.73	98.11	95.12	91.24	97.91
Sphere Face	MS1MV2 [3,14]	99.42	95.15	93.18	88.35	96.13
MobileFaceNet	MS1MV2 [3,14]	99.55	96.26	93.88	88.95	96.5
ElasticFace	MS1MV2 [3,14]	99.82	98.33	95.63	92.45	98.42
MagFace	MS1MV2 [3,14]	99.85	98.17	95.78	92.51	98.36

DOI: 10.54254/2755-2721/2026.TJ31675

Table 4. Model test data in IJB-B, IJB-C, MegaFace (Rank-1)

Method	Training Dataset	IJB-B		IJB-C		MegaFace
		TAR @FAR=1e-4	TAR @FAR=1e-5	TAR @FAR=1e-4	TAR @FAR=1e-5	(R1, %)
ArcFace	MS1MV2/ MS1MV3 [3,14]	94.2%-95.6%	86.5%-90.0%	96.3%-97.6%	92.1%-95.0%	98.35%-98.5%
CosFace	MS1MV2 [3,14]	93.6%-94.8%	85.1%-88.1%	95.8%-96.9%	90.7%-93.1%	98.1%-98.3%
SphereFace	MS1MV2 [3,14]	90.3%-91.9%	78.5%-81.3%	92.8%-94.4%	84.5%-87.1%	97.8%-98.1%
MobileFaceNet	MS1MV2/ CASIA [3,14]	85.0%-89.0%	70.0%-77.0%	88.0%-92.0%	78.0%-84.0%	92.0%-95.0%
ElasticFace	MS1MV2/ MS1MV3 [3,14]	94.5%-95.8%	87.2%-90.5%	96.6%-97.7%	92.8%-95.3%	98.4%-98.6%
MagFace	MS1MV2 [3,14]	94.8%-95.9%	88.0%-90.8%	96.8%-97.8%	93.2%-95.5%	98.4%-98.6%

Experimental Observations and Interpretations(As shown in Table 3 and Table 4)

ArcFace: Exhibits highly balanced and strong performance, ranking among the top across all benchmark datasets. It has become the foundation and starting point for numerous subsequent works.

CosFace: Achieves slightly lower performance compared to ArcFace (particularly under stricter conditions such as FAR = 1e-5), yet remains a very strong baseline model.

SphereFace: As one of the earliest works introducing angular-margin losses, its performance is relatively weaker among ArcFace, CosFace, and SphereFace, especially under large-pose challenges (e.g., TAR@FAR=1e-5 on IJB-B/C).

MobileFaceNet: Designed as a lightweight model, it employs a smaller MobileNet-like architecture. Its performance is significantly lower than large-scale models (e.g., ResNet-100 based architectures), which is expected. Its main advantages lie in compact size and high inference speed, making it deployment-friendly on mobile and embedded devices.

ElasticFace: As an improvement over ArcFace by introducing elastic angular margins, it consistently demonstrates slight yet stable advantages across almost all evaluation metrics, particularly under more stringent conditions (FAR = 1e-5), indicating better discrimination for difficult samples.

MagFace: Its core innovation lies in leveraging feature magnitude as a self-supervised signal. MagFace achieves performance comparable to ElasticFace and even surpasses it under some strict evaluation protocols (e.g., IJB-B/C @ FAR=1e-5). On its proposed MagFace test set, it achieves state-of-the-art results, highlighting robustness to extreme challenges such as large pose variations and severe image blur.

MegaFace (Rank-1): All models trained with ResNet-100-level backbones and large-scale datasets (ArcFace, CosFace, SphereFace, ElasticFace, MagFace) achieve very high and comparable performance (>98%). In contrast, MobileFaceNet is constrained by model capacity, resulting in a notable performance gap.

Performance Highlights:

Maximum Accuracy: For scenarios prioritizing accuracy over model size, ElasticFace and MagFace generally represent the state-of-the-art (SOTA) or near-SOTA models at the time of their publication, slightly surpassing ArcFace on benchmarks such as IJB-B/C and MegaFace. MagFace further demonstrates advantages on its own challenging test set.

Lightweight Models: MobileFaceNet is specifically designed for mobile platforms, trading off accuracy for extremely small model size and high speed.

Strong Baselines: ArcFace remains an exceptionally strong and widely accepted baseline in terms of performance, generalization, and reproducibility. CosFace and SphereFace, while slightly weaker,

DOI: 10.54254/2755-2721/2026.TJ31675

are pioneering contributions that paved the way for subsequent improvements.

Model Selection: Choosing an appropriate model depends on trade-offs among accuracy, model size/speed, deployment platform, and specific application requirements (e.g., robustness to extreme poses or low-quality images). ElasticFace and MagFace excel in accuracy, MobileFaceNet in lightweight design, while ArcFace maintains leadership in balanced performance and widespread adoption.

3. Conclusion

Since 2020, convolutional neural networks have continued to achieve remarkable progress in face recognition. This paper provides a systematic survey of representative models, including ArcFace, CosFace, MobileFaceNet, ElasticFace, and MagFace, with a focus on their innovations in network architecture, loss function design, model compression, and feature discriminability. By comparing their performance on standard benchmarks such as LFW, MegaFace, IJB-B, and IJB-C, we observe that current face recognition systems have reached high levels of accuracy, generalization, and computational efficiency, particularly under low false acceptance rate (FAR) conditions, significantly surpassing traditional methods.

Nevertheless, despite these advances, several challenges remain. Model stability and robustness are still limited in cross-age, cross-pose, occlusion, synthetic face interference, and low-quality imaging scenarios. Moreover, deploying high-accuracy and low-latency recognition systems on edge devices demands further network compression and optimization. Security, privacy protection, and defense against adversarial attacks are also becoming increasingly critical research topics.

Looking ahead, the development of face recognition will likely rely on synergistic innovations in network architectures and supervision mechanisms. Promising directions include deep learning, self-supervised learning, generative modeling, and cross-modal fusion. Leveraging multi-source heterogeneous data, enhancing semantic feature representations, and balancing real-time efficiency with deployability are expected to be the core goals of the next phase of research.

References

[1] Taigman Y, Yang M, Ranzato M, Wolf L. DeepFace: Closing the gap to human-level performance in face verification [C]// Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), 2014: 1701–1708.

[2] Schroff F, Kalenichenko D, Philbin J. FaceNet: A unified embedding for face recognition and clustering [C]// Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), 2015: 815–823.

[3] Deng J, Guo J, Xue N, Zafeiriou S. ArcFace: Additive angular margin loss for deep face recognition [J]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 4690–4699.

[4] Wang H, Wang Y, Zhou Z, Ji X, Gong D, Zhou J, Li Z, Liu W. CosFace: Large margin cosine loss for deep face recognition [C] Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 5265–5274.

[5] Liu W, Wen Y, Yu Z, Li M, Raj B, Song L. SphereFace: Deep hypersphere embedding for face recognition. Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), 2017: 212–220.

[6] Chen S, Liu Y, Gao X, Han Z. MobileFaceNets: Efficient CNNs for accurate real-time face verification on mobile devices. Chinese Conference on Biometric Recognition. Springer, 2018: 428–438.

[7] Boutros F, Ekenel H K, Damer N. ElasticFace: Elastic margin loss for deep face recognition [J]. arXiv preprint arXiv: 2203.11190, 2022.

[8] Meng Q, Zhao S, Huang Z, Zhou F. MagFace: A universal representation for face recognition and quality assessment [J]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 14225–14234.

DOI: 10.54254/2755-2721/2026.TJ31675

[9] Learned-Miller E, Huang G B, RoyChowdhury A, Li H, Hua G. Labeled Faces in the Wild: A survey [J]. Advances in Face Detection and Facial Image Analysis, 2016: 189–248.
[10] Kemelmacher-Shlizerman I, Seitz S M, Miller D, Brossard E. The MegaFace benchmark: 1 million faces for recognition at scale. Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), 2016: 4873–4882.
[11] Whitelam C, Taborsky E, Blanton A, Maze B, Adams J, Miller D J, Kalka N, Jain A K, Duncan J A, Allen K. IARPA Janus Benchmark–B (IJB-B) face dataset [J]. CVPRW, 2017: 592–600.
[12] Maze B, Adams J, Duncan J A, Kalka N, Miller D J, Otto C, Jain A K, et al. IARPA Janus Benchmark–C (IJB-C) [J]. CVPRW, 2018: 259–259.
[13] Google, "Classification: ROC and AUC | Machine Learning, "Google for Developers. https: //developers.google.com/machine-learning/crash-course/classification/roc-and-auc.
[14] Guo Y, Zhang L, Hu Y, He X, Gao J. MS-Celeb-1M: A dataset and benchmark for large-scale face recognition. European Conference on Computer Vision (ECCV). Springer, 2016: 87–102.