



Comparative Study of Deep Learning Techniques for Automatic Face Recognition

Saffiya Zhoulei

Department of Financial Management and Analytics, Kavir Polytechnic University of Technology, Iran
saffiya.zhoulei@kput-ir.net

Peer Review Information	Abstract
<i>Submission: 13 Jan 2021</i> <i>Revision: 27 Feb 2021</i> <i>Acceptance: 14 March 2021</i> Keywords <i>Deep learning; Face recognition; Convolutional neural network; Face embeddings; Metric learning</i>	Automatic face recognition has shifted from handcrafted descriptors to deep representation learning, where convolutional neural networks learn identity-discriminative embeddings directly from facial images. This methodology paper compares major deep-learning approaches that shaped face recognition up to 2021, with emphasis on DeepFace, FaceNet, SphereFace, CosFace, and ArcFace. The comparison is organized around network representation, loss function, embedding geometry, matching strategy, reported benchmark performance, and practical robustness. A structured evaluation pipeline is proposed using face detection, alignment, normalization, embedding extraction, similarity scoring, and verification/identification metrics. Published results on Labeled Faces in the Wild (LFW) are synthesized to illustrate the historical improvement from conventional softmax-based deep models to metric-learning and angular-margin objectives. The comparison indicates that the main methodological advance is not merely greater depth, but the explicit optimization of intra-class compactness and inter-class separation in the embedding space. Remaining challenges include pose, age, occlusion, demographic imbalance, presentation attacks, privacy, and generalization to unconstrained operational conditions.

Introduction

Face recognition is a biometric task in which an individual is verified or identified from facial imagery. Traditional systems relied on engineered features and separate classifiers, but deep learning replaced this pipeline with learned hierarchical representations. DeepFace demonstrated that a deep network combined with strong facial alignment could approach human-level verification on LFW, reporting 97.35% accuracy (Taigman et al., 2014). FaceNet then formulated recognition as embedding learning and directly optimized distances among identities using triplet loss, reporting 99.63% on LFW (Schroff et al., 2015).

Later work concentrated on making the learned feature space more discriminative. SphereFace

introduced an angular margin, CosFace used a cosine-margin formulation, and ArcFace imposed an additive angular margin with a clear geometric interpretation (Liu et al., 2017; Wang et al., 2018; Deng et al., 2019). These methods progressively shifted face recognition from ordinary classification toward hyperspherical metric learning, where identity is represented by compact, well-separated feature vectors.

By 2021, surveys described deep face recognition as a mature but still challenging field because benchmark-level accuracy does not guarantee reliable operation across pose, illumination, aging, occlusion, image quality, or demographic groups (Wang & Deng, 2021). O'Toole and Castillo (2021) likewise emphasized that deep

models transformed both machine performance and scientific understanding of face processing. Therefore, a meaningful comparison should examine training objective, representation geometry, data requirements, and robustness rather than only a single headline accuracy.

Literature Review

Taigman et al. (2014) developed DeepFace, combining 3D face alignment with a nine-layer deep neural network trained on millions of images. The work established the effectiveness of large-scale supervised deep representation learning for unconstrained face verification and reported 97.35% accuracy on LFW.

Schroff et al. (2015) proposed FaceNet, which maps face images to a compact Euclidean embedding and trains with triplets consisting of anchor, positive, and negative examples. The method made recognition, verification, and clustering possible through distances in the learned space and achieved 99.63% on LFW.

Liu et al. (2017) introduced SphereFace and A-Softmax loss. Instead of optimizing only class probabilities, SphereFace enforced an angular margin between identities on a hypersphere, strengthening inter-class separability and intra-class compactness.

Wang et al. (2018) proposed CosFace, or large-margin cosine loss. Features and classifier weights are normalized, and a cosine margin is subtracted from the target class to enlarge the decision boundary in angular space. The method reported 99.73% on LFW.

Deng et al. (2019) introduced ArcFace, which adds an angular margin directly to the target angle. The formulation is geometrically interpretable, computationally lightweight, and effective across multiple large face-recognition benchmarks; reported LFW performance reaches approximately 99.8% under common ArcFace configurations.

Guo and Zhang (2019) surveyed deep-learning-based face recognition and identified persistent challenges such as pose, age, illumination, expression, and heterogeneous matching, highlighting that strong benchmark performance alone does not solve unconstrained recognition.

Wang and Deng (2021) presented a comprehensive review covering architectures, loss functions, datasets, face processing strategies, and difficult recognition scenarios. Their taxonomy shows that modern performance gains are closely tied to better loss design, larger datasets, and normalization of nuisance factors.

O'Toole and Castillo (2021) reviewed major advances brought by deep networks and noted that modern systems can reach human-level performance on selected real-world recognition tasks, while also emphasizing the importance of understanding how learned face representations generalize.

Methodology

This study uses a comparative experimental methodology. The objective is to evaluate representative deep face-recognition families under a common pipeline while controlling preprocessing, matching rule, and evaluation protocol. Because the original systems were trained on different proprietary or public datasets, the paper separates two forms of evidence: (a) a reproducible comparison framework that can be implemented with public data, and (b) literature-reported benchmark results used for historical comparison.

1. Data and Pre-processing

A public benchmark such as LFW can be used for verification. Each image is first processed by a face detector; facial landmarks are then used to align the eyes and mouth to a canonical geometry. The aligned face is cropped, resized to the input resolution required by the selected network, converted to floating-point values, and normalized. The same detection and alignment pipeline should be used across models to avoid confounding model quality with preprocessing differences.

2. Compared Deep Learning Techniques

Five representative approaches are selected: DeepFace-type softmax classification, FaceNet triplet-loss embedding, SphereFace A-Softmax, CosFace large-margin cosine loss, and ArcFace additive angular-margin loss. These approaches represent the major progression from classification-driven representations to metric and angular-margin embedding learning.

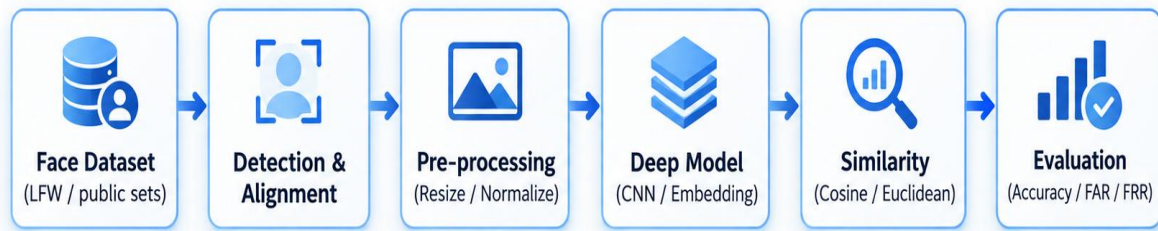


Figure 1. Proposed comparative methodology for automatic face recognition.

3. Feature Extraction and Matching

For each aligned face, the selected deep model produces an embedding vector. Verification compares two embeddings using cosine similarity or Euclidean distance. A threshold is learned on a validation split and then fixed for testing. Identification can be performed using nearest-neighbor comparison against enrolled templates. L2 normalization is applied where required by the model.

4. Evaluation Measures

Verification performance is measured using accuracy, true accept rate (TAR), false accept rate (FAR), false reject rate (FRR), and receiver operating characteristic behavior. For identification, Rank-1 recognition accuracy is used. Computational indicators such as embedding dimension, inference time, parameter count, and memory requirement should also be recorded when models are evaluated on the same hardware.

5. Experimental Procedure

1) Select a fixed face dataset and protocol; 2) detect and align every face with one preprocessing pipeline; 3) extract embeddings from each compared network; 4) normalize the embeddings as required; 5) compute similarity scores for genuine and impostor pairs; 6) determine the operating threshold on validation data; 7) evaluate the untouched test pairs; and 8) compare accuracy, FAR/FRR, robustness, and computational cost. Repeating the test over multiple folds reduces dependence on a single split.

6. Controlled Comparison Design

A controlled comparison should avoid retraining each method with unrelated data and then attributing all observed differences to the loss function. Where source code and pretrained weights are available, one option is to use the

original models only for an external historical comparison. A second, stronger option is to select a common backbone such as a residual network, train it on one fixed identity dataset, and change only the training objective. Under this design, the softmax baseline, triplet loss, A-Softmax, CosFace, and ArcFace are exposed to the same identities, image augmentations, optimizer family, number of epochs, input resolution, and evaluation splits. This isolates the contribution of the objective more effectively. Random seeds should be fixed and at least three training runs should be performed when computational resources permit. Mean performance and standard deviation can then be reported instead of a single best run.

7. Robustness and Error Analysis

A face-recognition method can appear saturated on LFW while still failing under operational variation. The proposed methodology therefore recommends an error analysis by nuisance factor. Test pairs can be grouped by large pose difference, illumination change, age difference, partial occlusion, low resolution, and expression change. Genuine pairs with low similarity are inspected as false rejects, while impostor pairs with unusually high similarity are inspected as false accepts. This analysis is important because the cost of the two error types is application dependent. A secure access-control system may prioritize a very low FAR, whereas a photo-organizing application may tolerate more false matches in exchange for higher recall. Reporting only overall accuracy hides these operating-point differences.

8. Reproducibility, Bias, and Ethical Controls

The experimental record should include dataset version, identity counts, image counts, training and validation partitions, detector and landmark model, crop size, normalization method, backbone, embedding dimension, optimizer, learning-rate schedule, loss hyperparameters,

similarity metric, and threshold-selection procedure. If demographic labels are legitimately available for a benchmark, performance should be disaggregated to check whether error rates vary strongly across groups. Such analysis must be interpreted carefully because benchmark composition may not represent deployment populations. Face data should be handled under appropriate consent, access, and privacy controls; unnecessary storage of raw facial images should be minimized. The comparative study is therefore intended to evaluate algorithms, not to infer sensitive personal traits from faces.

Results

The literature results show a clear evolution in verification accuracy as loss functions became more directly aligned with identity separation. Table 1 summarizes representative LFW results reported by the original methods or commonly cited configurations. These values should be interpreted with caution because the methods used different training datasets, architectures, and preprocessing; therefore, they are evidence of historical progress rather than a strictly controlled head-to-head experiment.

Table 1. Comparative characteristics and representative LFW verification results.

Technique	Core training objective	Embedding geometry	Reported LFW (%)	Main methodological feature
DeepFace (2014)	Softmax identity classification	Learned deep representation	97.35	3D alignment + large supervised deep network
FaceNet (2015)	Triplet loss	Euclidean embedding	99.63	Direct metric learning for verification/clustering
SphereFace (2017)	A-Softmax	Angular hypersphere	99.42	Multiplicative angular margin
CosFace (2018)	Large-margin cosine loss	Normalized cosine space	99.73	Additive cosine margin
ArcFace (2019)	Additive angular-margin loss	Normalized hypersphere	≈99.8	Geodesically interpretable angular margin

The first major gain is visible between DeepFace and FaceNet: replacing a classification-oriented representation with direct metric learning substantially improved verification accuracy while producing a compact embedding (Taigman et al., 2014; Schroff et al., 2015). SphereFace then reframed discrimination in angular space, creating a foundation for later margin-based losses (Liu et al., 2017).

CosFace simplified angular discrimination by normalizing features and weights and applying a fixed cosine margin, while ArcFace imposed the margin directly on the target angle (Wang et al., 2018; Deng et al., 2019). The resulting methods are easier to compare geometrically: a stronger model should place samples of the same identity closer together while maintaining a clear angular boundary between different identities.

Although the numerical gap between modern methods on LFW is small, the 2021 literature emphasizes that harder datasets and operational

conditions reveal larger differences. Pose, aging, occlusion, low resolution, domain shift, demographic imbalance, and attacks remain important challenges (Wang & Deng, 2021). Thus, evaluation should include difficult cross-condition datasets rather than relying exclusively on near-saturated LFW accuracy.

Conclusion and Discussion

This comparative methodology shows that the development of automatic face recognition up to 2021 is best understood as a progression from deep classification to discriminative embedding learning. DeepFace proved that large-scale deep representations and alignment could transform unconstrained recognition. FaceNet made the embedding itself the optimization target, while SphereFace, CosFace, and ArcFace increasingly imposed explicit angular separation among identities. Published LFW results indicate very high verification accuracy for all modern

approaches, but the comparison is not perfectly controlled because training data, backbone architecture, preprocessing, and compute differ. A rigorous implementation should therefore evaluate models with the same dataset split and preprocessing while reporting both recognition and computational metrics. The major research need is no longer only higher accuracy on standard datasets; it is reliable generalization across pose, age, occlusion, image quality, demographic variation, spoofing conditions, and privacy-constrained deployments. Accordingly, ArcFace-style margin learning provides a strong methodological baseline, while future comparisons should emphasize robustness, fairness, explainability, security, and efficiency in addition to accuracy.

References

- Taigman, Y., Yang, M., Ranzato, M. A., & Wolf, L. (2014). DeepFace: Closing the gap to human-level performance in face verification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1701-1708. <https://doi.org/10.1109/CVPR.2014.220>
- Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 815-823. <https://doi.org/10.1109/CVPR.2015.7298682>
- Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., & Song, L. (2017). SphereFace: Deep hypersphere embedding for face recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 212-220. <https://doi.org/10.1109/CVPR.2017.713>
- Wang, H., Wang, Y., Zhou, Z., Ji, X., Gong, D., Zhou, J., Li, Z., & Liu, W. (2018). CosFace: Large margin cosine loss for deep face recognition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5265-5274. <https://doi.org/10.1109/CVPR.2018.00552>
- Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). ArcFace: Additive angular margin loss for deep face recognition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4690-4699. <https://doi.org/10.1109/CVPR.2019.00482>
- Guo, G., & Zhang, N. (2019). A survey on deep learning-based face recognition. *Computer Vision and Image Understanding*, 189, 102805. <https://doi.org/10.1016/j.cviu.2019.102805>
- Wang, M., & Deng, W. (2021). Deep face recognition: A survey. *Neurocomputing*, 429, 215-244. <https://doi.org/10.1016/j.neucom.2020.10.081>
- O'Toole, A. J., & Castillo, C. D. (2021). Face recognition by humans and machines: Three fundamental advances from deep learning. *Annual Review of Vision Science*, 7, 543-570. <https://doi.org/10.1146/annurev-vision-093019-111701>