A Review of Deep Learning-Based Face Recognition Systems: Recent Advances, Challenges, and Future Directions

[1]Anju Kaur, [2]Puneet Devi, [3]Rachna Rajput
[1]Student, [2]Student, [3]Assistant Professor
[1]Computer Science and Engineering,
[1]Guru Kashi University, Bathinda, India
[1]anjukaur72306@gmail.com, [2]puneetdevi2004@gmail.com

Abstract

Face recognition is one of the fastest evolving research areas in computer vision due to its wide applications in security, surveillance, healthcare, banking, attendance management and smart authentication systems. Earlier face recognition systems relied on handcrafted feature extraction techniques such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Local Binary Patterns (LBP). While these approaches gave reasonable results under controlled environments, their performance degraded significantly when facial images contained variations in illumination, pose, expression, aging or partial occlusion. Deep learning has brought about a revolution in face recognition by allowing for the automatic extraction of highly discriminative facial features with the use of Convolutional Neural Networks (CNNs). State-of-the-art architectures such as DeepFace, VGG-Face, FaceNet, ArcFace, SphereFace2 and MobileFaceNet have led to remarkable improvements in recognition accuracy and computational efficiency. Each of these models involves unique learning strategies such as deep feature representation, embedding learning, angular margin optimization, binary classification and lightweight network design to overcome the limitations of previous techniques. This review paper provides a comprehensive study of the prominent deep learning-based face recognition models. The reviewed methods are analyzed based on their architectural design, learning objectives, recognition performance, advantages and limitations. Furthermore, a comparative analysis is provided to highlight the evolution of face recognition technologies. The paper also discusses current research challenges, including pose variation, illumination changes, occlusion, demographic bias, privacy preservation, and computational complexity. Finally, possible future research directions such as Vision Transformers, self-supervised learning, explainable artificial intelligence and lightweight edge-based face recognition systems are discussed. The review shows that while deep learning has greatly improved face recognition performance, further research is needed to design systems that are accurate, efficient, secure and ethically responsible for real-world deployment.

Keywords— Face Recognition, Deep Learning, Convolutional Neural Networks, DeepFace, VGG-Face, FaceNet, ArcFace, SphereFace2, MobileFaceNet, Biometrics.

I. INTRODUCTION

Face recognition is one of the most popular biometric technologies used for automatic identification or verification of individuals using their facial features. Unlike password-based authentication systems, face recognition provides a contactless, user-friendly, and efficient way of verification of identity. With the rapid progress of artificial intelligence and computer vision, face recognition has become an essential part of several real-world applications such as surveillance systems, mobile device authentication, border security, banking services, attendance monitoring, healthcare, and intelligent transportation systems. Early face recognition systems mainly used handcrafted feature extraction techniques such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Eigenfaces, Fisherfaces, and Local Binary Patterns (LBP). These methods relied on extracting handcrafted facial features and then used traditional machine learning algorithms for classification. Although these methods achieved good performance in controlled laboratory settings, they often failed when there were variations in pose, illumination, facial expression, aging, or occlusion in facial images. As a result, researchers started to explore learning-based methods that can automatically learn more robust and discriminative facial representations. Deep learning techniques, especially Convolutional Neural Networks (CNNs), have brought tremendous change to the face recognition field. CNN-based models can learn hierarchical feature representations automatically from the raw facial images without hand-crafted feature engineering, which has led to a great leap forward in recognition accuracy in the wild environment. The success of DeepFace has shown that deep neural networks can reach nearly human-level performance in face verification. Since then, many advanced models have been proposed, including VGG-Face, FaceNet, ArcFace, SphereFace2 and MobileFaceNet, which further improve the recognition accuracy by introducing novel loss functions, embedding learning strategies, angular margin optimization and lightweight network architectures for mobile devices. In addition to enhanced recognition performance, modern face recognition systems have become increasingly scalable and can handle millions of identities. However, despite these impressive advancements, several practical challenges still limit the deployment of face recognition technologies in real-world scenarios. Illumination variations, facial pose, expression, effects of aging, partial occlusion due to mask or sunglasses, demographic bias, privacy concerns, and computational requirements are still active research problems . Moreover, ethical issues regarding responsible use of biometric data have received significant attention among researchers, policy makers, and industry practitioners . The objective of this review paper is to provide a comprehensive analysis of recent deep learning based face recognition techniques. The paper reviews influential models including DeepFace , VGG-Face , FaceNet , ArcFace , SphereFace2 , and MobileFaceNet by examining their network architectures, learning mechanisms, advantages, limitations, and performance characteristics. A comparative analysis is also presented to show the evolution of these approaches. The current research challenges and future directions are discussed. The results presented in this review will help researchers and students in understanding the evolution of face recognition technologies and to identify promising areas for future research.

A. Major Contributions of this Review

The main contributions of this review paper are summarized as follows:

1. Comprehensive review of classical and modern deep learning based face recognition approaches.

2. Comparison of popular models like DeepFace, VGG-Face, FaceNet, ArcFace, SphereFace2 and MobileFaceNet.

3.Identification of strengths / limitations of each technique with respect to architecture, learning strategy, computational complexity and recognition performance.

4. Existing research challenges such as occlusion, illumination variation, pose estimation, privacy and fairness are discussed.

5. Introduction of new research areas including lightweight face recognition, Vision Transformers, explainable AI, and privacy-preserving biometric systems.

II. LITERATURE REVIEW

The research of face recognition has been revolutionized by the recent progress in deep learning. Earlier face recognition systems relied on handcrafted feature descriptors such as Eigenfaces, Fisherfaces, Local Binary Patterns (LBP), and Histogram of Oriented Gradients (HOG). Although these approaches produced satisfactory performance under controlled conditions, they were unable to generalize well to unconstrained environments where facial images exhibited variations in pose, illumination, facial expression, aging, and occlusion.

The introduction of deep convolutional neural networks (CNNs) enabled automatic learning of highly discriminative facial features directly from raw images. As a result, several deep learning-based models have been proposed to improve recognition accuracy and robustness. In this section, the most influential face recognition models that have driven forward modern biometric authentication systems are reviewed.

A. DeepFace

DeepFace proposed by Taigman et al in 2014 is one of the first deep learning based face recognition systems. Before DeepFace, the conventional approaches mainly used manually engineered features, which are usually sensitive to environmental changes. DeepFace demonstrated the power of deep neural networks in automatically learning robust facial representations and achieved human-level recognition performance.

DeepFace pipeline consists of several stages: face detection, facial landmark localization, 3D face alignment, face frontalization, feature extraction using deep convolutional neural network and face verification. One of the key innovations of DeepFace is the use of 3D alignment which normalizes the facial images before the feature extraction part, thus reducing the variations caused by different viewpoints. DeepFace's neural network architecture has several convolution layers followed by locally connected layers and fully connected layers. The model learns discriminative face representations from a large dataset of millions face images in the training phase. Then, the deep features extracted from the neural network are compared using similarity measures to decide whether two face images belong to same person. The experimental results on the Labeled Faces in the Wild (LFW) dataset show that DeepFace reached an accuracy of about 97.35%, very close to human performance. This success made deep learning the leading paradigm in face recognition research. One of the main advantages of DeepFace is its robustness to pose variations, because of its 3D face alignment module. Besides, the model introduced automatic feature learning, therefore there is no need for handcrafted descriptors. However, DeepFace needs a lot of pre-processing and high computational resources, which makes it difficult to adapt to mobile and embedded devices . Its dependence on large annotated datasets also limits its applicability in scenarios with limited training data. From a research perspective, DeepFace formed the foundation of almost all subsequent deep learning based face recognition systems. Though newer models have surpassed it, its contribution is highly important because it demonstrated the potential of deep neural networks for large scale biometric recognition.

B. VGG-Face

After the success of DeepFace which achieved a new state-of-the-art in deep face recognition by using a deeper convolutional neural network architecture, VGG-Face was introduced by Parkhi et al. The model was developed by the Visual Geometry Group (VGG) at the University of Oxford and was trained using one of the largest publicly available face datasets at that time.

The VGG-Face architecture consists of small convolutional stacks of filters with max-pooling layers and fully connected layers. The network learns robust features hierarchically from the face images without complex pre-processing techniques. The shallow convolutional layers learn simple facial features such as edges and textures whereas the deeper layers learn high discriminative semantic facial features.

One of the major contributions of VGG-Face is the release of a large pre-trained model that can be used for transfer learning. Researchers can fine-tune the network for different face recognition tasks without requiring massive computational resources for training from scratch. Consequently, VGG-Face has become one of the most widely adopted backbone networks in biometric authentication, attendance systems, surveillance, and forensic applications.

Compared with DeepFace, VGG-Face simplifies the preprocessing pipeline because it does not rely extensively on three-dimensional face alignment. Instead, it exploits deep feature learning to increase robustness to pose and illumination variations. The learned facial embeddings are highly discriminative and can be easily integrated in other recognition frameworks. However, VGG-Face has a large number of trainable parameters, which entails high computational complexity and memory requirements. This makes the model less suitable for deployment on mobile devices with limited processing capability. Additionally, inference time is relatively high compared with lightweight face recognition architectures developed in recent years.

The overall influence of VGG-Face on later research was important in showing that increasing the depth of the network and training on large-scale datasets can greatly enhance recognition accuracy . The use of VGG-Face features as a baseline for the evaluation of new face recognition algorithms is still common.

C. FaceNet

In 2015, FaceNet proposed by Schroff, Kalenichenko and Philbin introduced a fundamentally different learning strategy than previous face recognition models. Instead of directly performing multi-class classification, FaceNet learns a compact embedding representation that maps each face image into a low-dimensional Euclidean feature space. The key innovation of FaceNet is the Triplet Loss function. During training, three facial images are processed simultaneously, including an anchor image, a positive image that belongs to the same person and a negative image that belongs to an individual of different identity. The optimization goal is to minimize the distance between the anchor and positive samples and maximize the distance between the anchor and negative samples.

This embedding-based learning approach enables FaceNet to learn highly discriminative 128-dimensional representations of faces. Once trained, the similarity between two faces can be simply measured by the Euclidean distance between their embeddings. Thus, the same model can be used for face verification, face identification, face clustering and image retrieval tasks without training separate classifiers.

Experimental results show that FaceNet achieved an impressive recognition accuracy of 99.63% on Labeled Faces in the Wild (LFW) dataset, which is better than most of the existing face recognition methods at that time. The compact embedding representation also reduces storage requirements, thus making the approach very suitable for large scale face recognition systems that contain millions of identities.

In contrast to DeepFace and VGG-Face, FaceNet doesn't require the use of a traditional classification layer at inference time, but instead the learned embeddings offer a highly versatile representation that can be stored efficiently and compared across large databases, greatly improving scalability and computational efficiency.

However, the performance of FaceNet depends heavily on effective triplet selection during training. Selecting informative positive and negative samples is computationally expensive and can considerably influence convergence speed. Moreover, training requires carefully balanced datasets to avoid biased embedding learning.

FaceNet is one of the most influential face recognition frameworks despite these limitations, as it introduced metric learning to deep face recognition, and many popular algorithms such as ArcFace and SphereFace further advanced the embedding-learning idea of FaceNet.

D. ArcFace

ArcFace is one of the most influential advances in deep face recognition, proposed by Deng et al. in 2019 . The model addresses a key limitation of previous approaches by proposing the Additive Angular Margin Loss (AAM-Softmax) , which enhances the discriminative capability of facial feature embeddings. Unlike the typical softmax loss that mainly focuses on the classification accuracy, ArcFace explicitly increases the angular margin between different identities while reducing the angular margin between samples of the same identity. The architecture of ArcFace generally uses deep convolutional neural networks such as ResNet as the feature extraction backbone. After feature extraction, facial representations are normalized and projected onto a hypersphere where the angular margin is added during training. This optimization yields very discriminative embeddings that improve face verification and identification performances in unconstrained environments.

The experimental evaluations on benchmark datasets LFW, MegaFace, IJB-B and IJB-C demonstrated the state-of-the-art performance of ArcFace, which consistently outperformed previous methods such as DeepFace, VGG-Face, and FaceNet by increasing intra-class compactness and inter-class separability.

One of the major advantages of ArcFace is its simple implementation despite significant performance gains. The angular margin loss can be easily replaced with the conventional softmax loss without significant modifications of the existing CNN architectures. Moreover, ArcFace shows excellent generalization ability across various datasets. However, ArcFace needs large scale labeled datasets for proper training, and also substantial computational resources during the optimization process. Hyper-parameter tuning is also important for achieving the optimal performance. Nevertheless, ArcFace has become the de-facto baseline for many recent face recognition studies due to its superior accuracy and stable convergence.

E. SphereFace

SphereFace was introduced by Liu et al. as an effort to improve the discriminative power of deep facial representations through angular optimization. Instead of minimizing Euclidean distances, SphereFace introduces the Angular Softmax (A-Softmax) Loss, which forces facial features of the same identity to occupy compact angular regions in the embedding space.

The motivation of developing SphereFace is based on the observation that the angular distances are more proper than the Euclidean distances for characterizing the identity differences. We seek to maximize the angular margin between different classes such that the highly discriminative facial representations could be learned and the recognition accuracy could be significantly improved under the challenging conditions. The network architecture normally consists of deep residual convolutional networks with feature normalization and angular classification layers. During training, the angular margin gradually increases feature discrimination and encourages more compact intra-class representations.

Experimental studies indicate that SphereFace significantly improves recognition accuracy compared with earlier softmax-based methods. It performs particularly well in large-scale face verification tasks and has influenced the development of several subsequent margin-based loss functions.

Despite its effectiveness, SphereFace suffers from optimization difficulties because the Angular Softmax Loss introduces a highly non-linear objective function. Training becomes unstable unless careful parameter initialization and learning-rate scheduling are employed. Consequently, some simplified margin-based methods such as CosFace and ArcFace were proposed to deal with these optimization difficulties. Nevertheless, SphereFace still remains as a major milestone as it introduces angular learning to deep face recognition and demonstrates the effectiveness of hypersphere embedding spaces.

F. SphereFace2

SphereFace2 is an improved version of the SphereFace framework. Instead of formulating face recognition as a multi-class classification problem, SphereFace2 introduces a binary classification perspective that reduces the optimization complexity while improving the feature discrimination. The main goal of SphereFace2 is to reduce the optimization complexity of the Angular Softmax Loss while preserving the discriminative properties of the angular feature learning. The proposed binary classification framework results in more stable convergence and better recognition accuracy over multiple benchmark datasets. Compared with the original SphereFace, SphereFace2 demonstrates better training stability and lower computational complexity. The model also demonstrates stronger generalization capability under unconstrained environments involving pose variation, illumination changes, and facial occlusion.

Another important contribution of SphereFace2 is its flexibility. The framework can be easily combined with various CNN backbone architectures, making it suitable for different practical face recognition applications.

Although SphereFace2 significantly improves optimization compared with SphereFace, the model still depends on large annotated datasets for effective training. Computational requirements remain relatively high for large-scale recognition systems involving millions of identities.

In summary, SphereFace2 effectively integrates the benefits of angular feature learning and enhanced optimization methods, and thus makes a significant progress in the state-of-the-art face recognition research.

G. MobileFaceNet

Most deep face recognition models aim for recognition accuracy, MobileFaceNet aims for computational efficiency. MobileFaceNet is a lightweight neural network architecture designed for embedded and mobile devices, which is capable of accurate face recognition with substantially fewer parameters.

The architecture uses depth-wise separable convolutions, bottleneck residual blocks and global depth-wise convolution to reduce the computational complexity while maintaining the feature extraction capability. The proposed architectural modifications help in reducing the size of the model significantly without much effect on the recognition accuracy.

One of the most important benefits of MobileFaceNet is that it enables real-time face recognition on smartphones, IoT devices, embedded systems, and edge computing platforms. The fewer the number of parameters, the less memory is used, the faster the inference speed, and the less power is consumed, and it can be deployed on resource-constrained hardware.
Experimental evaluations demonstrate that MobileFaceNet achieves competitive recognition accuracy despite having only a fraction of the computational cost required by larger architectures such as VGG-Face and ResNet-based models.

However, lightweight architectures inevitably sacrifice a small amount of recognition accuracy compared with state-of-the-art large-scale models such as ArcFace. Nevertheless, MobileFaceNet provides an excellent trade-off between computational efficiency and recognition performance, making it highly attractive for practical biometric authentication systems.

H. Comparative Discussion of Reviewed Models

This survey of deep learning models demonstrates the progress of advanced face recognition technology in the past decade. DeepFace was the first to introduce automatic deep feature learning in convolutional neural networks, instead of hand-crafted feature engineering. VGG-Face proved the huge improvement of feature representation that could be achieved by deeper networks and training on huge datasets. FaceNet adopted metric learning with Triplet Loss to produce compact embedding representations for verification, identification and clustering. SphereFace and ArcFace aimed at improving feature discriminability by optimizing the angular margin, while SphereFace2 simplified the optimization process to a binary classification framework. MobileFaceNet addressed the challenges of practical deployment by designing lightweight architectures for real-time face recognition on mobile and embedded devices.

Finally, the development of these models is a step towards more holistic solutions that balance recognition accuracy, computational efficiency, scalability and deployment flexibility, moving away from accuracies tailored for architectures . Face recognition research continues to build on these foundational advancements through research on transformer-based architectures, self-supervised learning, explainable artificial intelligence , and privacy-preserving biometric systems.

III.COMPARATIVE ANALYSIS OF DEEP LEARNING-BASED FACE RECOGNITION MODELS

The evolution of deep learning has significantly improved the performance of face recognition systems. Each proposed model has attempted to overcome the limitations of its predecessors by introducing new learning strategies, optimization methods, and network architectures. This section presents a comparative analysis of the reviewed models based on architecture, learning strategy, computational complexity, recognition performance, and practical applications.

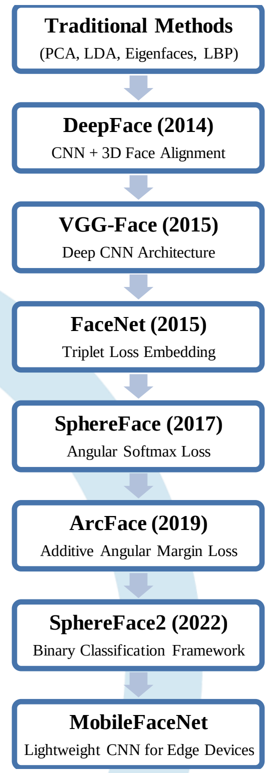

Figure 1. Evolution of deep learning-based face recognition models.

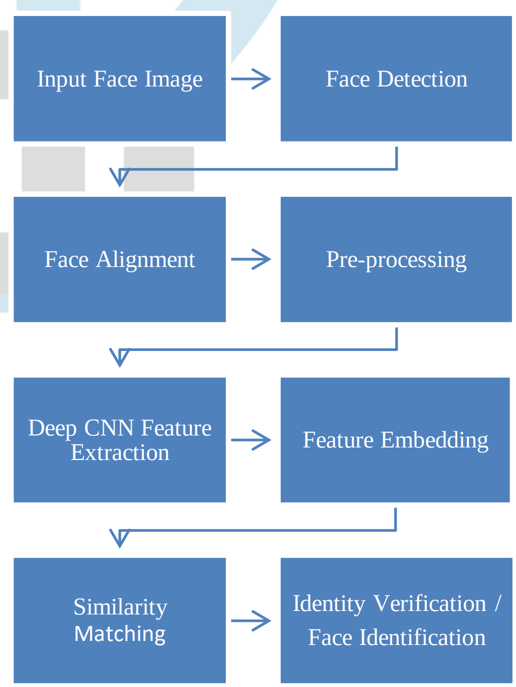

Figure 2. General workflow of a deep learning-based face recognition system.

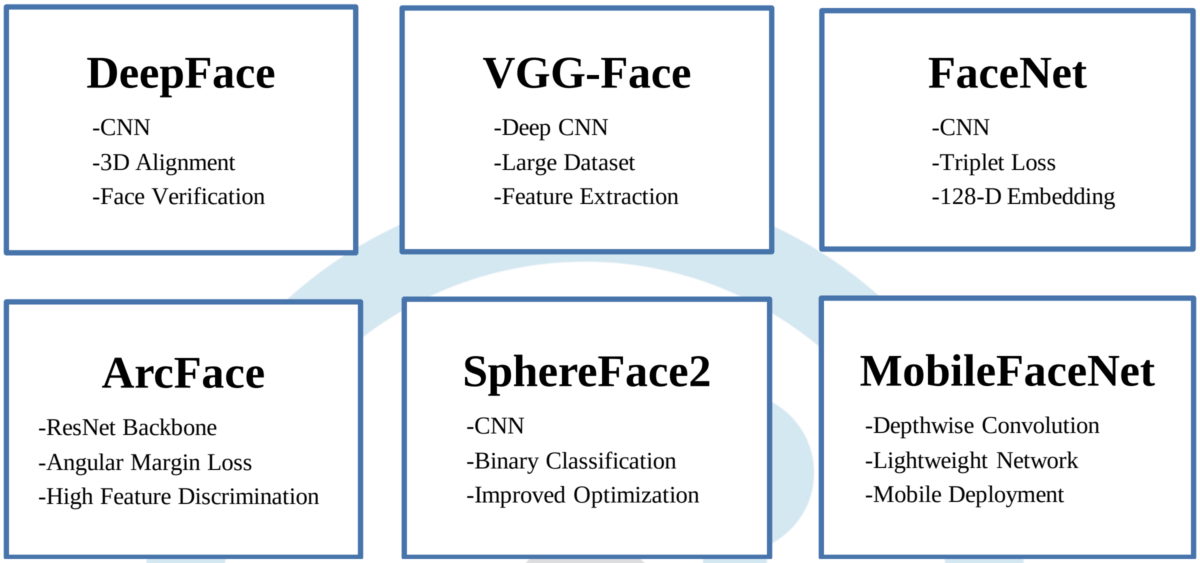

Figure 3. Comparison of major deep learning-based face recognition models.

Table I. Comparative Analysis of Reviewed Face Recognition Models

Model	Year	Learning Strategy	Advantages	Limitations	Typical Applications
DeepFace	2014	CNN + 3D Alignment	Near-human performance	High computation	Face verification
VGG-Face	2015	Deep CNN	Excellent feature extraction	Large model size	Recognition & Transfer Learning
FaceNet	2015	Triplet Loss	Compact embeddings	Difficult triplet selection	Verification & Clustering
SphereFace	2017	Angular Softmax Loss	Better feature discrimination	Training instability	Large-scale recognition
ArcFace	2019	Additive Angular Margin Loss	State-of-the-art accuracy	High computational cost	Biometric authentication
SphereFace2	2022	Binary Classification	Stable optimization	Requires large datasets	Intelligent recognition
MobileFaceNet	2018	Lightweight CNN	Fast inference	Slight reduction in accuracy	Mobile & IoT devices

Discussion

The evolution of face recognition technology from computationally expensive CNN architectures to lightweight and highly discriminative deep learning models can be seen as summarized in Table I. DeepFace demonstrated the feasibility of deep neural networks, while VGG-Face enhanced the feature representation with deeper architectures. FaceNet introduced the embedding learning with Triplet Loss, which allowed flexible similarity-based recognition.

SphereFace and ArcFace enhanced the discriminative power of the features by optimizing the angular margin, while SphereFace2 simplified the optimization process by the use of binary classification. MobileFaceNet solved the practical problems of deploying the model in mobile and embedded systems by reducing the complexity of the model. Overall, the reviewed models suggest a gradual shift towards a balance of recognition accuracy, computational efficiency, scalability and deployment flexibility.

IV. RESEARCH GAP

While much progress has been made in deep learning-based face recognition, some problems are still unsolved.

1. Most existing models require a huge amount of annotated data for effective training.

2. The recognition accuracy drops under the severe pose variations, illumination variations and facial occlusions such as masks and sunglasses.

3. Many deep learning models have millions of trainable parameters, which makes it hard to deploy them in resource-constrained devices.

4. Existing approaches are usually dedicated to improving recognition accuracy, with little attention paid to fairness, interpretability and demographic bias.

5. Privacy preservation is a major concern as the biometric information is highly sensitive.

6. Most of the current systems are supervised and require manually annotated datasets, which increase annotation cost.

Hence, future research should focus on developing lightweight, privacy-preserving, self-supervised and explainable face recognition systems which can work reliably in real-world conditions.

V. CHALLENGES IN FACE RECOGNITION

Despite the impressive progress, there are still many technical and ethical challenges for practical face recognition systems.

A. Head Pose
Variation Large variations of head pose can greatly decrease recognition accuracy as parts of the face may be occluded.

B. Illumination
Changes Variations in illumination conditions can have a significant impact on the image quality and thus influence the performance of the feature extraction.

C. Facial Occlusion
Face masks, sunglasses, scarves and hair can hide key facial features used for identification.

D. Aging
The natural process of aging can change the appearance of a face making it difficult to identify over long periods of time.

E. Demographic Bias
Recognition performance may be affected by differences in gender, ethnicity and age groups due to unbalanced datasets.

F. Privacy and Security
The unauthorized collection and abuse of facial data pose severe ethical and legal issues. Developing privacy-preserving biometric systems has therefore become an important research direction.

VI. FUTURE RESEARCH DIRECTIONS

Recent advances in artificial intelligence have opened several promising research opportunities for face recognition systems.

A. Vision Transformers
Transformer-based architectures are expected to improve long-range feature representation beyond conventional CNNs.

B. Self-Supervised learning Learning without expensive manual annotation can help to greatly reduce data labeling costs.

C. Explainable Artificial Intelligence (XAI) The future systems should provide transparent explanations for recognition decisions to build user trust.

D. Lightweight Deep Learning Models. The growing importance of efficient neural networks that can run on smartphones, IoT devices, and edge computing platforms.

E. Federated Learning Federated learning allows collaborative model training without the need to collect facial images in a centralised database, hence preserving the users' privacy.

F. Multi-modal Biometrics Biometrics systems will combine facial recognition with iris, fingerprint and voice recognition in the future to improve the accuracy and security of authentication.

Together, these emerging technologies are likely to deliver highly accurate, computationally efficient, secure, and ethically responsible face recognition systems for real-world deployment.

VII. CONCLUSION

Facial recognition has become an important research area in the fields of computer vision and artificial intelligence and one of the fastest growing areas . The transition from traditional feature-based methods to deep learning-based approaches has greatly improved the accuracy, robustness, and scalability of face recognition systems . This review paper presents a detailed analysis of several influential deep learning models including DeepFace, VGG-Face, FaceNet, SphereFace, ArcFace, SphereFace2, and MobileFaceNet . A comparative study shows that each model has made great innovations to overcome the limitations of previous methods. DeepFace showed the power of deep convolutional neural networks together with three-dimensional face alignment. VGG-Face has improved the extraction of facial features by using deeper network structures and larger-scale training datasets. FaceNet presented the metric learning based on the Triplet Loss to learn compact facial representations for verification, identification, and clustering applications. SphereFace and ArcFace further improved the recognition accuracy by proposing the angular margin-based optimization, while SphereFace2 reduced the optimization complexity with the binary classification framework. MobileFaceNet addressed the practical issues of deploying face recognition systems on mobile and embedded devices by reducing the computational complexity at the cost of acceptable recognition performance. Despite the remarkable progress, there are still several challenges to be solved. However, variations in illumination, facial pose, aging, facial occlusion, demographic bias, computational requirements, and privacy preservation still pose limitations to deploying face recognition systems in unconstrained environments. The widespread adoption of facial recognition technologies has also increased the importance of the ethical use of biometric data. Future research should be directed towards the design of lightweight explainable and privacy-preserving face recognition systems that are capable of running efficiently on resource-constrained devices. Emerging technologies such as Vision Transformers, self-supervised learning, federated learning, and multimodal biometrics are expected to be important for the next generation of intelligent biometric authentication systems . In summary, deep learning has revolutionized face recognition research by enabling automatic feature learning and highly discriminative facial representations . Further advances in artificial intelligence are expected to further improve recognition performance and to address the technical and ethical challenges associated with real-world deployment.

REFERENCES

[1] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, “DeepFace: Closing the Gap to Human-Level Performance in Face Verification,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[2] O. M. Parkhi, A. Vedaldi, and A. Zisserman, “Deep Face Recognition,” *British Machine Vision Conference (BMVC)*, 2015.

[3] F. Schroff, D. Kalenichenko, and J. Philbin, “FaceNet: A Unified Embedding for Face Recognition and Clustering,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[4] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, “SphereFace: Deep Hypersphere Embedding for Face Recognition,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[5] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, and W. Liu, “CosFace: Large Margin Cosine Loss for Deep Face Recognition,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[6] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, “ArcFace: Additive Angular Margin Loss for Deep Face Recognition,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[7] J. Meng, Y. Zhao, J. Huang, and H. Wang, “SphereFace2: Binary Classification is All You Need for Deep Face Recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[8] S. Chen, Y. Liu, X. Gao, and Z. Han, “MobileFaceNets: Efficient CNNs for Accurate Real-Time Face Verification on Mobile Devices,” *Chinese Conference on Biometric Recognition (CCBR)*, 2018.

[9] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[10] J. Guo and J. Deng, “InsightFace: 2D and 3D Face Analysis Project,” GitHub Repository, 2021.

[11] Y. Sun, X. Wang, and X. Tang, “Deep Learning Face Representation from Predicting 10,000 Classes,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[12] J. Deng, J. Guo, Y. Zhou, J. Yu, I. Kotsia, and S. Zafeiriou, “RetinaFace: Single-Shot Multi-Level Face Localisation in the Wild,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[13] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[14] R. Szeliski, *Computer Vision: Algorithms and Applications*, 2nd ed. Springer, 2022.

[15] A. Krizhevsky, I. Sutskever, and G. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” *Advances in Neural Information Processing Systems (NeurIPS)*, 2012.

[16] A. Dosovitskiy et al., “An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale,” *International Conference on Learning Representations (ICLR)*, 2021.

[17] A. Vaswani et al., “Attention Is All You Need,” *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[18] X. An, Y. Zhu, J. Xiao, F. Wu, M. Zhang, Y. Gao, and B. Qin, “Partial FC: Training 10 Million Identities on a Single Machine,” *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2021.

[19] M. Turk and A. Pentland, “Eigenfaces for Recognition,” *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[20] P. N. Belhumeur, J. Hespanha, and D. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.