Proceedings of CONF-MLA 2025 Symposium: Applied Artificial Intelligence Research
DOI: 10.54254/2755-2721/2025.BJ25394

Deep Learning Techniques for Occluded Face Recognition: A Survey and Future Directions

Simeng Zhang

Haide College, Ocean University of China, Qingdao, China
zsm5657@stu.ouc.edu.cn

Abstract: Face recognition under occlusion has become a critical research focus due to its widespread applications in surveillance, authentication, and human-computer interaction, and the frequent presence of masks, glasses, or other obstructions in real-world settings. This review systematically examines the evolution of techniques designed to handle occluded facial inputs, covering both traditional approaches and deep learning-based methods. Traditional techniques such as subspace regression, local feature analysis, and robust estimation provide interpretable and efficient solutions but are often sensitive to large or unstructured occlusions. Recent advances in deep learning, including CNNs, GANs, and self-supervised architectures, have significantly improved occlusion robustness by enabling feature reconstruction, landmark detection, and semantic completion. A comparative analysis of 22 representative studies is presented, categorized by occlusion type, methodological framework, and performance on benchmark datasets. Based on current limitations, this paper outlines future directions, including lightweight model design, multimodal fusion, and standardized occlusion-aware evaluation metrics. This review aims to provide a comprehensive reference for researchers and practitioners seeking to develop more reliable and generalizable face recognition systems under real-world occlusion challenges.

Keywords: Face recognition, Occlusion challenges, Deep learning

1. Introduction

Face recognition is an identity authentication technology based on facial biometric characteristics. In recent years, due to the rich identity-related information contained in facial images, as well as the convenience, low cost, and non-contact nature of image acquisition, face recognition technology has been widely applied in various fields such as intelligent security, identity verification, video surveillance, human-computer interaction, and social entertainment. As a technology that integrates image processing, pattern recognition, machine learning, and psychology, face recognition has significant research and application value.

Since the 1960s, when Bledsoe and others first explored face recognition methods based on pose compensation, the field has witnessed substantial progress. In 1987, Sirovich et al [1]. were the first to apply the Karhunen–Loève (K-L) transform for the optimal representation of facial images. In 1991, Turk et al [2]. proposed a method based on facial feature points, utilizing key landmarks for

© 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: 10.54254/2755-2721/2025.BJ25394

face matching. This perspective provided a novel direction for the development of face recognition technologies. Around the year 2000, research began to shift toward sparse representation and deep learning methods. To date, studies on face recognition under controlled environments have achieved a relatively mature level of performance. However, in real-world scenarios, the recognition accuracy still suffers due to various factors such as illumination, pose, expression, age, and especially occlusion [3].

Among these challenges, occlusion has emerged as a major bottleneck preventing the practical deployment of face recognition systems. Effective feature extraction becomes significantly more difficult because of the diversity of occlusion types, uncertainty in the location and size of occluded regions, and their disruptive effect on the structural integrity of facial features. In recent years, many studies have focused on face recognition under mask occlusion and have achieved considerable progress. Techniques such as reconstructing occluded regions [4] and extracting locally robust features [5] have been widely explored. However, real-life occlusions are far more complex than those caused by masks. Situations with hats, hair, glasses, hands covering faces, and multiple overlapping faces cover larger areas, create more unstable features, and often have changing and messy characteristics, making it harder for recognition systems to work.

Therefore, how to design more generalizable, robust, and efficient face recognition methods under occlusion remains a critical issue in the field. This paper focuses on the research direction of occluded face recognition, systematically reviewing and classifying existing literature, analyzing the core ideas and applicable scenarios of different technical approaches, highlighting their limitations in complex occlusion settings, and providing an outlook on future development trends. The aim is to offer theoretical support and methodological references for further studies in this domain.

2. Traditional methods for face recognition with occlusion

2.1. Subspace regression methods

The fundamental concept of subspace regression is to project facial images into a low-dimensional subspace through linear or nonlinear transformations, thereby achieving a more compact feature representation. This approach enables the classification of different facial categories into corresponding subspaces while modeling and analyzing occluded regions as independent subspaces. Operating in a low-dimensional space offers advantages such as reduced computational load and lower complexity. Representative subspace regression methods include sparse representation, collaborative representation, and occlusion dictionary-based strategies, all of which have demonstrated significant improvements in classification performance.

Sparse representation, as a transformation-based statistical modeling technique, aims to express a test facial image as a sparse linear combination of training samples. This effectively reduces intra-class variation and improves recognition robustness. Although its recognition rate on some benchmark datasets is not exceptionally high, it consistently exhibits strong resilience under occlusion conditions. For instance, Taher et al [6]. created a method that breaks down a facial image into simple parts, showing both the clear and blocked areas, which helps in recognising faces even when they are occluded or have complicated expressions. This highlights the advantage of sparse modeling in separating informative features from noise and interference [7].

Collaborative representation, on the other hand, models the entire set of facial samples from different categories and represents the test image as a collaborative linear combination of all training samples. This holistic approach enhances the robustness of face recognition. Zhang et al [8]. were the first to systematically propose this method and emphasized that, in high-dimensional feature

DOI: 10.54254/2755-2721/2025.BJ25394

spaces, the collaborative modeling itself is more critical than sparsity. That is, even without imposing explicit sparsity constraints (e.g., ℓ 1-norm), collaborative representation can still effectively distinguish between different classes. In this approach, the test image is recreated by combining samples from all classes, and it is classified by looking for the smallest difference, which makes the system better at handling blocked views, noise, and having few training samples. Compared to traditional Sparse Representation-based Classification (SRC), Collaborative Representation-based Classification (CRC) not only achieves higher computational efficiency but also outperforms in several benchmark face recognition datasets such as AR and Extended Yale B. It remains robust even under occlusion rates as high as 50% [9].

2.2. Local feature-based methods

The core idea of local feature-based methods—particularly Local Feature Analysis (LFA)—is to mitigate or eliminate the interference caused by occluded regions in face recognition by extracting localized features from facial images. This approach, often referred to as local feature weighting, is a widely adopted technique to address occlusion-related challenges. Specifically, facial images are divided into multiple subregions, and by assigning different weights to the features extracted from each region, the influence of occluded areas can be effectively reduced.

LFA constructs a set of local feature vectors using Principal Component Analysis (PCA) and applies sparsification techniques to derive a compact and correlated feature set. By focusing on localized representations rather than holistic facial patterns, LFA enhances the system's robustness to partial occlusions and localized noise.

As surveyed by Zeng et al [10]., local feature-based methods such as LFA divide the face into subregions and strategically downweight occluded parts, achieving strong resilience in partially occluded scenarios. These methods excel particularly when occlusion affects only a portion of the facial area, allowing the system to rely on the unoccluded regions for accurate recognition.

Zhang et al [11]. further advanced this line of research by proposing the Local Gabor Binary Pattern Histogram Sequence (LGBPHS), which integrates multi-scale Gabor filters with Local Binary Pattern (LBP) encoding to construct a robust and discriminative feature representation. By performing localized analysis and encoding of facial regions, their method demonstrates significant robustness against both occlusion and expression variations, thereby improving recognition performance in complex environments.

2.3. Robust estimation methods

The core concept of robust estimation methods lies in utilizing the available, albeit partially occluded, facial data to estimate the discriminative features necessary for recognition. These methods are designed to make effective use of visible information while minimizing the interference caused by occluded regions, thereby enhancing the stability and reliability of recognition systems under challenging conditions.

Robust techniques such as Robust Sparse Representation and weighted local block matching strategies have demonstrated strong adaptability, particularly in scenarios involving unpredictable occlusions. By emphasizing resilience to corruption and uncertainty, these methods enable more reliable recognition outcomes without requiring precise knowledge of the occluded areas [10].

Burgos-Artizzu et al [12]. proposed a Robust Cascaded Pose Regression (RCPR) method that effectively estimates facial landmark positions under occlusion. Their approach explicitly models the occluded regions by incorporating occlusion labels during the training phase. This technique

DOI: 10.54254/2755-2721/2025.BJ25394

enables the learning of a set of regression models that can automatically suppress the influence of occluded parts. Compared to conventional landmark detection methods, RCPR significantly improves detection accuracy under occluded conditions and does not rely on any prior knowledge about the occlusion patterns.

3. Deep learning methods for face recognition with occlusion

3.1. Cnn-based feature extraction under occlusion

CNN has become a powerful tool for face recognition since Taigman proposed DeepFace [13]. However, there are still some remaining problems to be solved. Prasad et al. examined two widely adopted architectures—VGG-Face and Lightened CNN—to assess their robustness under various occluded conditions. VGG-Face, built upon the deep VGG-16 structure, extracts high-level semantic features from large-scale face data. Its fully connected layers, particularly FC7, are often used as face embeddings for recognition tasks. In contrast, Lightened CNN employs a lightweight design with fewer parameters and incorporates Max-Feature-Map (MFM) activation to enhance feature compactness and non-linearity.

Their comparative analysis using the AR face dataset suggests that VGG-Face performs more effectively when the lower part of the face is occluded, while Lightened CNN shows greater stability under upper-face occlusion such as sunglasses. The results highlight that not only network depth but also the choice of feature layer and activation function play a critical role in maintaining recognition accuracy under partial occlusion [14].

3.2. Gan-based occlusion reconstruction

Generative Adversarial Networks (GANs) have emerged as a powerful approach for addressing the challenges of face recognition under occlusion. By learning to generate or restore missing facial regions, GAN-based frameworks can effectively recover discriminative information lost due to partial obstructions. In this section, the author reviews two representative studies that adopt GAN-based strategies for occlusion handling, each targeting different types of facial occlusion and incorporating distinct architectural innovations.

3.2.1. Masked face reconstruction with bidirectional GAN and attention

Alzubi et al [15]. proposed a comprehensive framework tailored to the problem of masked face recognition, where the lower part of the face is occluded by protective masks. Their approach utilizes a bidirectional GAN to generate synthetic unmasked faces from masked inputs and vice versa, thereby enriching the training data with complementary samples. These generated images are fed into a Dual Scale Adaptive Efficient Attention Network (DS-AEAN), which extracts both local and global facial features via parallel attention streams. The model is further optimized using an Enhanced Addax Optimization Algorithm (EAOA) to tune hyperparameters such as hidden layer size and training epochs. Experiments on benchmark masked face datasets (FMDD, RMFD) demonstrate that this framework significantly outperforms conventional models like AMaskNet and MobileNetV2 in both accuracy and computational efficiency.

DOI: 10.54254/2755-2721/2025.BJ25394

3.2.2. Transformer-aided GAN for general occlusion recovery

In a more general setting, Usama et al [16]. introduced a GAN-based deep learning model designed for arbitrary facial occlusions, including partial coverings of the eyes, nose, or side regions. The framework incorporates a Transformer encoder-decoder architecture to capture long-range dependencies and a hybrid attention mechanism to enhance feature discrimination. Additionally, Usama et al [16]. introduce a feature imputation module to estimate missing facial components based on contextual cues. Unlike the previous method which focuses on data augmentation, this model emphasizes feature-level reconstruction and semantic completion, allowing it to adapt to more diverse and irregular occlusion patterns. Testing on the Occluded Face Detection (OCFD) dataset shows that this model performs better in recognising faces with occlusions than the basic GAN-only or CNN methods.

3.3. Self-supervised and weakly-supervised occlusion-aware learning

Tu [17] proposed a self-supervised 3D facial landmark detection framework designed to handle large-pose and occluded scenarios without relying on explicitly labeled 3D data. The method utilizes unconstrained 2D facial images to guide the learning of 3D Morphable Model (3DMM) parameters and integrates three key components to improve landmark estimation under occlusion.

First, a 2D-assisted supervision strategy is adopted, in which 2D face images collected from in-the-wild conditions are used to assist 3DMM fitting. This eliminates the dependence on expensive 3D annotations while preserving the structural integrity of facial geometry. Second, a self-supervised mapping mechanism is introduced, where the model learns to optimize the projection between 3D landmarks and their corresponding 2D representations. By minimizing reprojection error, the system enhances its ability to infer 3D shape even for partially occluded faces. Third, a self-critic learning module, inspired by adversarial training, serves as a weak supervisory signal to evaluate the plausibility of the predicted 3D shapes. Unlike typical GAN frameworks, this module does not generate images but rather functions as a discriminator to refine the shape parameters through implicit feedback.

Experiments on AFLW2000-3D, AFLW-LFPA, and Florence datasets show that the method achieves robust landmark localization in both occluded and non-occluded settings. Compared to image-level reconstruction models, this approach emphasizes structural coherence and pose-invariant feature learning, offering an alternative pathway for occlusion-robust face recognition.

3.4. Siamese and metric learning-based approaches for variation robustness

Qiu et al [18]. proposed an Extended Siamese Network (ESN) to enhance face recognition performance under eyeglass occlusion and scale variation. The method extends the classic Siamese architecture by introducing a mapping layer, which aligns the features of variation images (e.g., with eyeglasses or at lower resolution) to those of their original counterparts [19]. This design allows the network to better learn variation-invariant representations. To further improve class separability, a repulsive center loss is incorporated into the network. This loss builds upon the original center loss, which minimizes intra-class feature distances, by adding a repulsive term that increases the inter-class margin [20]. Inspired by prior metric learning frameworks such as FaceNet [21], the approach combines similarity-based training with explicit class structure modeling to increase robustness under appearance variation. Empirical results on the LFW dataset demonstrate that the ESN

DOI: 10.54254/2755-2721/2025.BJ25394

outperforms standard Siamese networks and achieves recognition accuracy comparable to other state-of-the-art methods, even when trained on a relatively small dataset.

4. Conclusion

This review systematically examined face recognition under occlusion, highlighting both traditional methods and emerging deep learning-based solutions. Traditional approaches such as subspace regression, local feature analysis, and robust estimation techniques offer strong interpretability and perform well under structured occlusions. However, their performance declines significantly under unstructured or large-area occlusions, limiting their practical applicability.

In contrast, deep learning models, especially CNNs, GANs, and Transformer-based architectures have demonstrated remarkable resilience in dealing with various occlusion types. These models are capable of learning occlusion-invariant representations, reconstructing missing facial regions, and leveraging both global and local features. Nevertheless, their robustness is still challenged by real-world variability, such as dynamic occlusions, pose variations, and limited labeled datasets. Additionally, issues such as computational cost, generalizability, and the lack of standardized benchmarks remain unsolved.

Looking ahead, several promising research directions are identified. Model generalization under diverse occlusions: Improving model robustness across varying occlusion types, positions, and intensities is crucial. Approaches based on self-supervised learning, occlusion simulation, or data augmentation can mitigate the dependency on labeled occluded samples. Lightweight and efficient architectures: The practical deployment of face recognition systems, especially on mobile or embedded devices, demands lightweight models with low computational overhead. Techniques such as network pruning, knowledge distillation, and architecture search should be explored. 3D and multimodal fusion: Combining RGB with thermal, depth, or 3D facial geometry information can enhance system robustness under extreme occlusions or adverse environments. More research is needed to develop effective multimodal fusion strategies. Occlusion-aware benchmarks and evaluation metrics: Current datasets often fail to reflect real-world occlusion patterns. Constructing diverse and standardized occlusion datasets, along with specific evaluation metrics, will facilitate fair comparison and rapid iteration of occlusion-resilient models. Multi-person recognition under occlusion: Identifying multiple individuals in crowded and partially occluded scenes poses new challenges in surveillance and social applications. Developing models that can jointly handle occlusion and identity separation is an important future direction.

In summary, while face recognition under occlusion has made notable progress, developing models that are accurate, generalizable, interpretable, and deployable under real-world conditions remains an open challenge. Future research must focus on the integration of advanced learning paradigms, multimodal data, and realistic evaluation schemes to bridge the gap between academic innovation and practical application.

References

[1] Sirovich L, Kirby M.Low-dimensional procedure for the characterization of human faces [J].Journal of the Optical Society of America A, Optics and Image Science, 1987，4 (3)：519-524.
[2] Turk M.A., Pentland A.P.Face recognition using eigenface [J].journal of cognitive neuroscience, 2012, 3(1): 71-86.
[3] Zhao W, Chellappa R, Phillips PJ, Rosenfeld A (2003) Face recognition: A literature survey. ACM comput surv (CSUR) 35(4): 399-458
[4] Y. Feng, F. Wu, X. Shao, et al. Joint 3d face reconstruction and dense alignment with position map regression network [C]. ECCV, 2018, 534-551

DOI: 10.54254/2755-2721/2025.BJ25394

[5] T. Ahonen, A. Hadid, M. Pietikäinen. Face recognition with local binary patterns [C]. ECCV, 2004, 469-481
[6] S. L. Fernandes and G. J. Bala, "A Study on Face Recognition Under Facial Expression Variation and Occlusion, " in Proc. Int. Conf. Soft Computing Systems, Advances in Intelligent Systems and Computing, vol. 397, pp. 371–377, Springer, 2016. doi: 10.1007/978-81-322-2671-0_35.
[7] Taheri S, Patel VM et al (2013) Component based recognition of faces and facial expressions. Affective Comput 4(4): 360–371
[8] Zhang, L., Yang, M., Feng, X., Ma, Y., & Zhang, D. (2011). Collaborative representation based classification for face recognition. Technical Report, The Hong Kong Polytechnic University.
[9] Zhang, L., Yang, M., & Feng, X. (2011). Sparse representation or collaborative representation: Which helps face recognition? In Proceedings of the IEEE International Conference on Computer Vision (ICCV) (pp. 471–478). IEEE.
[10] Zeng, D., Veldhuis, R. N. J., & Spreeuwers, L. J. (2020).A survey of face recognition techniques under occlusion. arXiv preprint arXiv: 2006.11366.
[11] Zhang, B., Zhang, H., & Shen, L. (2010). Local Gabor Binary Pattern Histogram Sequence (LGBPHS): A novel non-statistical model for face representation and recognition. In K. W. Wong, D. Q. Miao, & Z. B. Wang (Eds.), Advances in Neural Networks – ISNN 2010 (pp. 715–723). Springer.
[12] Burgos-Artizzu, X. P., Perona, P., & Dollár, P. (2013). Robust face landmark estimation under occlusion. In Proceedings of the IEEE International Conference on Computer Vision (ICCV) (pp. 1513–1520). IEEE.
[13] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, “Deep-face: Closing the gap to human-level performance in face verification, ” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1701–1708.
[14] Prasad, P. S., Pathak, R., Gunjan, V. K., & Rao, H. V. R. (2020). Deep learning based representation for face recognition. In Proceedings of ICCCE 2019 (Lecture Notes in Electrical Engineering, Vol. 570, pp. 557–566). Springer.
[15] Alzubi, J. A., Pokkuluri, K. S., Arunachalam, R., Shukla, S. K., Venugopal, S., & Arunachalam, K. (2025). A generative adversarial network-based accurate masked face recognition model using dual scale adaptive efficient attention network. Scientific Reports, 15, 17594.
[16] Usama, M., Bourouis, S., Alassery, F., Mehmood, A., Choi, G. S., & Anwar, H. (2023). Loading… A GAN-based deep learning framework for occluded face recognition. Computers, Materials & Continua, 77(2), 1969–1988.
[17] Tu, X. (2020). Research on unconstrained face recognition (Doctoral dissertation, Beijing Jiaotong University). (In Chinese)
[18] Qiu, F., Kamata, S., & Ma, L. (2017). Deep face recognition under eyeglass and scale variation using extended Siamese network. IEEE International Conference on Image Processing, pp. 471–475.
[19] S. Chopra, R. Hadsell, and Y. LeCun, “Learning a similarity metric discriminatively, with application to face verification, ” in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol. 1. IEEE, 2005, pp. 539–546.
[20] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, “A discriminative feature learning approach for deep face recognition, ” in ECCV. Springer, 2016, pp. 499–515.
[21] F. Schroff, D. Kalenichenko, and J. Philbin, “Facenet: A unified embedding for face recognition and clustering, ” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 815–823.