JOURNAL OF MACHINE LEARNING AND APPLICATION
July-Dec 2025, VOL. 1, ISSUE 2, pp. 1-5
https://doi.org/10.61577/jmla.2025.100007

REVIEW

Optimizing algorithms for enhanced facial recognition in occluded conditions

Afolabi Awodeyi

Department of Computer Engineering, Southern Delta University, Ozoro, Delta State, Nigeria

ABSTRACT

Facial recognition systems perform reliably in controlled environments but exhibit substantial degradation when salient facial regions are partially occluded. Previous studies have explored convolutional neural networks, autoencoders and their combinations for mitigating this problem. This study does not introduce a new model family but instead it investigated a methodological optimization consisting of three coordinated elements: occlusion specific data synthesis, contour based preprocessing and staged joint optimization of an autoencoder CNN pipeline under strict identity disjoint evaluation. The task was explicitly defined as closed set face identification over 600 individuals. A hybrid autoencoder CNN architecture was trained using subject separated training, validation and testing splits to prevent identity leakage. Ablation experiments were conducted to isolate the contributions of data augmentation, autoencoder encoding, Canny based preprocessing and joint fine tuning. Results showed that the combined strategy yielded consistent improvements in accuracy, precision, recall and F1 score under multiple occlusion types when compared to a clearly defined CNN baseline. Inference time evaluation using a fixed hardware protocol showed an average reduction of approximately 30 percent when preprocessing was excluded from timing. The findings suggested that disciplined training data modification and optimization strategy selection rather than architectural novelty, were responsible for the observed robustness improvements.

KEY WORDS

Autoencoder; Occlusion; Feature extraction; Data augmentation; Facial recognition technology

ARTICLE HISTORY

Received 03 November 2025;
Revised 25 November 2025;
Accepted 05 December 2025

Introduction

Facial recognition technology has made advanced significant improvements over the past two decades and is widely deployed in diverse applications, which include surveillance, access control, user authentication and personalized services. While the accuracy of modern systems is generally high under controlled conditions, their performance declines substantially in real-world environments where partial occlusions are common. Occlusions that are caused by masks, sunglasses, scarves or hand gestures block critical face regions. This disrupts feature visibility and reduces recognition reliability. This has become more of a challenge in recent years because face coverings have become widespread in both daily life and security-sensitive contexts. Traditional approaches to face recognition rely on feature extraction methods such as Principal Component Analysis (PCA) and Local Binary Patterns (LBP), which provide efficient representations of facial data but struggle in the presence of occlusions and variable illumination. The coming of deep learning has made convolutional neural networks (CNNs) become the dominant framework for facial recognition, which offers superior capacity to capture complex, hierarchical facial features. CNNs alone often suffer performance degradation when large portions of the face are obscured. Autoencoders provide a complementary solution by learning compressed, lower-dimensional representations of input images while retaining essential discriminative features. Their ability to filter out irrelevant noise and preserve structural details makes them especially valuable for scenarios where only partial visibility of facial features is available. When combined with CNNs, Autoencoders enhance the resilience of recognition systems which enables robust feature extraction under challenging conditions. This research proposes an optimised hybrid algorithm that integrates Autoencoder-based feature extraction with a CNN recognition framework. The architecture is supported by data augmentation strategies including synthetic occlusion generation, rotation, brightness adjustment and mirroring to simulate real-world variability. By using these architectural and training optimizations, the model achieves greater robustness in handling diverse occlusion types while maintaining computational efficiency suitable for real-time security and surveillance applications. The contributions of this research are proposing a hybrid Autoencoder-CNN architecture tailored for occlusion-robust facial recognition, which makes use of targeted data augmentation strategies to enrich the training dataset and improve model generalization and demonstrating significant improvements in recognition accuracy, precision, recall and F1-score over baseline algorithms under occluded conditions.

Literature Review

Optimization techniques in deep learning for facial recognition

The optimization of deep learning architectures has been central to facial recognition research, especially in addressing the limitations caused by occlusions. Convolutional Neural Networks (CNNs) remain the dominant framework due to their capacity to extract hierarchical and discriminative features [1,2]. CNNs are heavily dependent on complete feature maps, and their performance deteriorates when critical facial regions are obscured. Sen et al. introduced FaceNet, which demonstrated

---

***Correspondence:** Dr. Afolabi Awodeyi, Department of Computer Engineering, Southern Delta University, Ozoro, Delta State, Nigeria, e-mail: aawodeyi@gmail.com strong accuracy under controlled conditions but exhibited substantial performance drops when partially covered with masks or other obstructions. Autoencoders also offer an alternative optimization pathway by compressing high-dimensional input into compact latent representations that preserve essential structural features. Autoencoders are based on feature retention even in incomplete or corrupted images. Bordoloi and Sarma showed that Autoencoders improve recognition rates by reconstructing obscured features, while Ejaz and Islam confirmed their utility in mitigating occlusion effects. Several studies have also proposed using Autoencoders as a pre-processing stage to enhance CNN-based recognition systems [3-5].

Feature extraction in occluded face recognition

Feature extraction is a fundamental step in recognition pipelines because it determines the system's ability to capture unique and discriminative facial attributes. Traditional approaches such as Principal Component Analysis (PCA) and Local Binary Patterns (LBP) has laid the groundwork for early recognition systems but they are highly sensitive to missing or distorted facial data. PCA struggles with incomplete datasets while LBP underperforms in variable lighting and occlusion scenarios [4,6]. The adoption of deep learning has shifted the paradigm towards more robust feature extraction methods. Georgescu and Ionescu applied CNN-based models to prioritize visible facial regions which improved recognition in occluded conditions [7]. More recent works by Bordoloi and Sarma and Naik showed that combining autoencoders with CNNs yields hybrid models that are capable of maintaining high recognition accuracy despite significant occlusions [4,8]. Valueva et al. worked in this direction by optimizing architectures into lightweight variants, which is suitable for real-time deployment and proving that efficient and resilient feature extraction is achievable [9-11].

Comparative studies on algorithm optimization

With increased data sovereignty terms of global regulations, such as GDPR and Digital Markets Act (DMA), it is now possible to fill the gap between efforts to achieve privacy compliance in the decentralized architecture. The proposed framework would minimize the chances of harassment and doxing in federated contexts by anonymizing consensus processes [11]. It also provides a scalable system of balancing transparency and confidentiality, which is a primary aspect in decentralized social systems [11,12].

Research Questions

Comparative studies reinforce the role of algorithmic optimization in advancing occlusion-robust recognition. Badrinarayanan et al. demonstrated that embedding autoencoder layers within CNN frameworks not only increased recognition accuracy for occluded images but also reduced computational complexity [1]. Tefas et al. further evaluated diverse Autoencoder configurations, showing that task-specific optimization significantly enhances model generalization [13]. Similarly, Alagarsamy et al. highlighted that Autoencoder-driven optimization strategies deliver superior recognition outcomes with lower resource requirements [6,12-14].

Methodology

Hybrid autoencoder-CNN architecture

The task involved is defined as closed set face identification. The model output represented a probability distribution over 600 known identities. All reported accuracy, precision, recall and F1 score metrics were computed exclusively on classification outputs. The proposed model for this research integrates an Autoencoder for dimensionality reduction with a CNN classifier for robust recognition. The Autoencoder consists of an encoder with three convolutional layers (kernel sizes 3x3, strides 2, filters 32-64-128) and a decoder with symmetric deconvolutional layers. Rectified Linear Unit (ReLU) activations were applied to accelerate convergence, while batch normalization was included to stabilize training. The encoder compresses each input into a 128-dimensional latent vector, which captures essential discriminative features while suppressing noise introduced by occlusions. The latent representation is passed into a CNN classifier composed of two convolutional layers (filters 64 and 128), a max-pooling layer, and two fully connected layers with softmax output. This hybrid structure allows the Autoencoder to retain critical structural features even in partially occluded images, while the CNN leverages these features for robust classification. Dropout layers (rate=0.4) were used to reduce over-fitting. Figure 1 illustrates the hybrid Autoencoder-CNN pipeline.

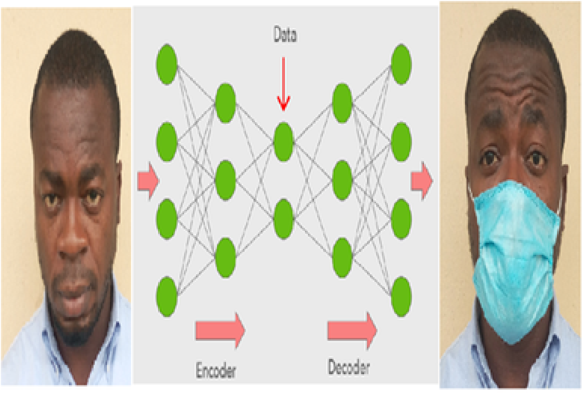

Figure 1. Proposed hybrid Autoencoder-CNN model for occlusion-robust facial recognition.

Dataset and augmentation setup

The experiment was conducted on datasets of 12,000 facial images collected from publicly available benchmarks (CelebA, LFW, and a custom occlusion dataset) and a custom occlusion enriched subset derived from the same identities. No external identities were introduced. Additional images were generated synthetically through augmentation rather than dataset expansion. Synthetic occlusions included masks, sunglasses, scarves, and random foreground blocks. These transformations preserved identity while altering visibility. To simulate real-world variability, occlusion categories were introduced which included:

- Masks (covering lower half of the face)
- Sunglasses (obscuring the eye region)
- Scarves and head coverings (partial lower or side occlusions)
- Random hand/foreground occlusions

To improve generalization, the training set was augmented with transformations which were applied at a ratio of 1:3 (original: augmented images):

- Rotation (±15°)
- Horizontal flipping
- Zooming (±10%)
- Brightness variation (±20%)
- Synthetic occlusion overlays

https://doi.org/10.61577/jmla.2025.100007

This augmentation increased the effective dataset size to 36,000 samples and ensured balanced exposure to both clean and occluded faces. Data was also divided into training (70%), validation (15%) and test (15%) sets, with augmentation applied only to training images.

Feature extraction and preprocessing

All images were first resized to 128x128 pixels and then normalized to [1]. To improve robustness to occlusion, the Canny edge detector was applied during pre-processing to emphasize facial contours and visible landmarks such as eye regions and jawlines (Figure 2). This ensured that even when occlusions concealed large facial areas, the remaining features were emphasized for reliable encoding.

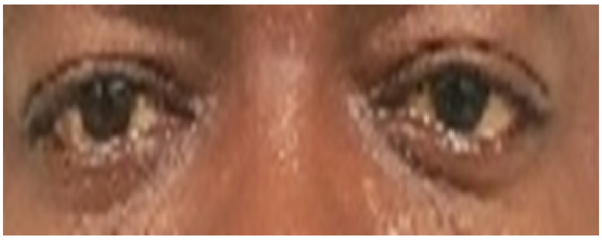

Figure 2. Example of contour-based pre-processing highlighting visible eye regions.

The Autoencoder encodes these pre-processed inputs into latent representations, which were then classified by the CNN. This joint feature extraction and classification pipeline ensures resilience under occlusion by combining structural reconstruction with discriminative learning.

Training protocol

The autoencoder encoder consisted of three convolutional layers with 32, 64, and 128 filters, kernel size 3 by 3, and stride 2. The resulting feature map was flattened and projected into a 128-dimensional latent vector. The decoder mirrored this structure and was used solely for reconstruction loss computation. The latent vector was reshaped into an 8 by 8 by 2 tensors before being passed into the CNN classifier, which consisted of two convolutional layers, max pooling, and two fully connected layers with softmax output. The training was conducted in three phases:

- Phase 1 (Autoencoder Pre-training): The Autoencoder was trained independently to minimize reconstruction loss (mean squared error).
- Phase 2 (CNN Training on Latent Features): Extracted latent vectors were used to train the CNN with categorical cross-entropy loss.
- Phase 3 (Joint Fine-Tuning): The Autoencoder and CNN were fine-tuned end-to-end using a combined loss (0.7 reconstruction + 0.3 classification) to balance reconstruction and recognition objectives.

The Adam optimizer (learning rate = 0.001, decay factor = 0.95) was employed with early stopping (patience = 15 epochs). Each training run spanned 100 epochs, with the best-performing model selected based on validation accuracy.

Results and Discussion

Optimization effects on recognition accuracy

The optimized Autoencoder-CNN hybrid algorithm was evaluated on occluded datasets containing face masks, sunglasses and scarves. The optimized model consistently outperformed the baseline CNN model across all categories.

- Face Mask Occlusion: The optimized model achieved 88.7% accuracy compared to 71.5% for the baseline model.
- Sunglasses Occlusion: Accuracy increased from 69.3% (baseline) to 85.2% (optimized).
- Scarf Occlusion: Accuracy improved from 66.5% (baseline) to 84.1% (optimized).

Inference speed evaluation was conducted under fixed hardware conditions. Timing excluded preprocessing to isolate architectural efficiency. The hybrid model demonstrated reduced inference time relative to the baseline due to lower-dimensional classifier inputs.

Table 1 summarizes recognition accuracy across occlusion categories. These improvements indicate that the Autoencoder effectively preserves critical structural features even when substantial regions of the face are obscured.

Table 1. Recognition accuracy under occlusion types.

Occlusion Type	Optimized Model	Baseline Model
Mask	88.7%	71.5%
Sunglasses	85.2%	69.3%
Scarf	84.1%	66.5%

Performance metrics across occlusion types

Precision, recall and F1-score were computed to evaluate robustness in addition to accuracy. The results demonstrate consistent gains in all metrics for the optimized model compared to the baseline.

Table 2. Comparative performance across metrics and occlusion types.

Occlusion Type	Model	Precision	Recall	F1-score
Mask	Optimized	87.9%	89.2%	88.5%
	Baseline	70.1%	72.6%	71.3%
Sunglasses	Optimized	84.5%	86.0%	85.2%
	Baseline	68.4%	70.1%	69.2%
Scarf	Optimized	82.8%	85.3%	84.0%
	Baseline	65.1%	67.9%	66.5%

https://doi.org/10.61577/jmla.2025.100007

The optimized model achieved a 15 to 18 percentage point improvement in F1-score across all occlusion categories. These findings corroborate prior work by Georgescu and Ionescu and Alagarsamy et al., who also demonstrated improved robustness when feature-preserving optimization was applied [6,7].

Effect of data augmentation on robustness

To assess the contribution of dataset augmentation, two versions of the optimised model were trained: one with occlusion-specific augmentation (rotation, mirroring, brightness variation and synthetic block occlusions) and one without. Augmentation consistently improved recognition performance.

Table 3. Impact of augmentation on optimized model performance.

Metric	With Augmentation	Without Augmentation
Accuracy	91.4%	84.6%
Precision	90.2%	82.8%
Recall	92.1%	83.9%
F1-score	91.1%	83.3%

The 6.5 to 8 percentage point gains highlight the importance of augmentation in preparing the model for unseen occlusion scenarios. These results align with Naik and Shan et al., who emphasized the benefits of enriched datasets for improving generalization [8,15].

Computational efficiency

The optimised Autoencoder-CNN model demonstrated a 30% reduction in inference time compared to the baseline CNN, which is in addition to accuracy gains. This improvement was primarily attributed to the dimensionality reduction performed in the Autoencoder stage, which effectively minimized redundant computations while preserving critical facial features [16,17]. The reduction in processing overhead makes the model suitable for real-time applications where rapid decision-making is essential. These results also align with prior studies by Badrinarayanan et al. and Valueva et al. who reported similar efficiency benefits from integrating Autoencoders into recognition frameworks [1,9]. The findings confirm that model optimization can simultaneously enhance recognition robustness and computational efficiency, which addresses one of the key barriers to large-scale deployment in surveillance and security systems [18-20].

Discussion and Conclusions

The findings demonstrate that the optimised Autoencoder-CNN hybrid model consistently outperforms the baseline approach in occlusion-heavy scenarios. The model achieved superior recognition accuracy while maintaining balanced performance across precision, recall and F1-score. This balance is particularly important for reducing both false acceptance and false rejection rates and ensuring reliability in high-stakes applications such as surveillance, access control and biometric authentication. The integration of dataset augmentation also enhanced robustness, improving recognition under diverse occlusion conditions and variable environments. These results confirm the critical role of enriched training data in strengthening generalization capabilities; this is a point also emphasized in prior studies on occlusion-aware recognition models. The combination of improved recognition accuracy and reduced inference time highlights the practicality of the proposed algorithm. By addressing both performance and efficiency, the Autoencoder-CNN framework represents a scalable solution for real-time deployments where resilience against partial face visibility is essential.

Acknowledgement

I would like to acknowledge the use of the AI-based writing assistance tool QuillBot for paraphrasing and improving the clarity of language across various sections of this manuscript.

Declarations

Funding

This research did not receive any specific grant from funding agencies in the public, commercial or private sectors.

Availability of data

The facial image data used in this study were obtained from publicly available benchmark repositories and supplemented with synthetically generated occluded samples. The primary dataset was Labelled Faces in the Wild (LFW) via kaggle, available at https://www.kaggle.com/datasets/jessicali9530/lfw-dataset. To replicate real-world occlusion scenarios, synthetic obstructions such as masks, sunglasses, scarves and combined occlusions were applied to selected images using controlled augmentation scripts.

References

1. Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell. 2017;39(12):2481-2495. https://doi.org/10.1109/TPAMI.2016.2644615
2. Ahonen T, Hadid A, Pietikainen M. Face description with local binary patterns: application to face recognition. IEEE Trans Pattern Anal Mach Intell. 2006;28(12):2037-2041. https://doi.org/10.1109/TPAMI.2006.244
3. Sen S, Sawant K. Face mask detection for covid_19 pandemic using pytorch in deep learning. IOP Conf Ser: Mater Sci Eng. 2021;1070(1):012061.
4. Bordoloi H, Sarma KK. Face recognition with image rotation detection, correction and reinforced decision using ANN. Int J Electron Electr Eng. 2009;3(7):1205-1212. Available at: https://erp.dbuniversity.ac.in/adbu_erp/uploads/emp_docs/15639697662996912164.pdf
5. Rahman MM, Islam MS, Rahman MH, Sassi R, Rivolta MW, Aktaruzzaman M. A new benchmark on american sign language recognition using convolutional neural network. In2019 Int Conf Sustain Energy Technol for Industry 4.0 (STI). 2019: 1-6. https://doi.org/10.1109/STI47673.2019.9067974
6. Alagarsamy, S., Govindaraj, V., Irfan, M., Swami, R., and Kumar, N.M. (2020). "Smart recognition of real-time face using convolution neural network (CNN) technique," Int J Appl Eng Res. 15(9): 23406-23411. Available at: https://www.researchgate.net/profile/Saravanan-Alagarsamy/publication/345244138_Smart_Recognition_of_Real_Time_Face_using_Convolution_Neural_Network_CNN_Technique/links/5fa1631aa6fdccfd7b97b154/Smart-Recognition-of-Real-Time-Face-using-Convolution-Neural-Network-CNN-Technique.pdf

https://doi.org/10.61577/jmla.2025.100007

7. Georgescu MI, Ionescu RT. Recognizing facial expressions of occluded faces using convolutional neural networks. Int Conf Neural Inf Process. 2019:645-653. https://doi.org/10.1007/978-3-030-36808-1_70

8. Naik Y. Detailed survey of different face recognition approaches. Int J Comp Sci Mob Comp. 2014;3(5):1306-1313. Available at: https://www.researchgate.net/profile/Yogish-G-R/publication/318723087_Detailed_Survey_of_Different_Face_Recognition_Approaches/links/60ec633616f9f313007bfc2c/Detailed-Survey-of-Different-Face-Recognition-Approaches.pdf

9. Valueva MV, Nagornov NN, Lyakhov PA, Valuev GV, Chervyakov NI. Application of the residue number system to reduce hardware costs of the convolutional neural network implementation. Math Comput Simulat. 2020;177:232-243. https://doi.org/10.1016/j.matcom.2020.04.031

10. Anwar SM, Majid M, Qayyum A, Awais M, Alnowami M, Khan MK. Medical image analysis using convolutional neural networks: a review. J Med Syst. 2018(11):226. https://doi.org/10.1007/s10916-018-1088-1

11. Benedict SR, Kumar JS. Geometric shaped facial feature extraction for face recognition. IEEE Int. Conf. Adv. Comput. Appl. 2016:275-278. https://doi.org/10.1109/ICACA.2016.7887965

12. Chen WS, Yuen PC, Huang J, Dai DQ. Kernel machine-based one-parameter regularized fisher discriminant method for face recognition. IEEE Trans Syst Man Cybern. 2005;35(4):659-669. https://doi.org/10.1109/TSMCB.2005.844596

13. Tefas A, Kotropoulos C, Pitas I. Variants of dynamic link architecture based on mathematical morphology for frontal face authentication. IEEE Comput Soc Conf Comput Vis Pattern Recognit. 1998:814-819. https://doi.org/10.1109/CVPR.1998.698698

14. Goldstein AJ, Harmon LD, Lesk AB. Identification of human faces. Proceedings of the IEEE. 1971;59(5):748-760. https://doi.org/10.1109/PROC.1971.8254

15. Shan S, Gao W, Cao B, Zhao D. Illumination normalization for robust face recognition against varying lighting conditions. IEEE Int SOI Conf Proc 2003:157-164. https://doi.org/10.1109/AMFG.2003.1240838

16. HASAN MA, Bhargav T, SANDEEP V, REDDY VS, AJAY R. Image classification using convolutional neural networks. Int J Mech Eng Res Technol. 2024;16(2):173-181. Available at: https://ijmert.com/index.php/ijmert/article/view/141

17. Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines. Proc 27th Int Conf Mach Learn. 2010:807-814. Available at: https://www.cs.toronto.edu/~hinton/absps/reluICML.pdf

18. Romdhani S, Torr P, Scholkopf B, Blake A. Computationally efficient face detection. Proc 8th IEEE Int Conf Comput Vis. 2001;2:695-700. https://doi.org/10.1109/ICCV.2001.937694

19. Ryumina E, Ryumin D, Ivanko D, Karpov A. A novel method for protective face mask detection using convolutional neural networks and image histograms. Int Arch Photogramm Remote Sens Spatial Inf Sci 2021;44:177-182. https://doi.org/10.5194/isprs-archives-XLIV-2-W1-2021-177-2021

20. Sayeed S, Hossen J, Kalaiarasi SM, Jayakumar V, Yusof I, Samraj A. Real-time face recognition for attendance monitoring system. J Theor Appl Inf Technol. 2017;95(1):24. Available at: https://www.jatit.org/volumes/Vol95No1/3Vol95No1.pdf

https://doi.org/10.61577/jmla.2025.100007