

Article

Multi-stage training approaches for improved facial recognition under occlusions

Afolabi Awodeyi^{1,*} and Omolegho A. Ibok²¹ Department of Computer Engineering, Southern Delta University Ozoro, Delta State, Nigeria² Department of Computer Engineering, Southern Delta University Ozoro, Delta State, Nigeria

* Correspondence: awodeyiai@sdu.edu.ng

Received: 23 May 2026; Accepted: 15 August 2026; Published: 28 September 2026.

Abstract: Facial recognition systems often experience substantial performance degradation when essential facial regions are obscured by occlusions. These occlusions interfere with feature visibility and reduce the accuracy and robustness of conventional single-stage training processes. To address this challenge, this study proposes a multi-stage training framework designed to improve facial recognition robustness under occluded conditions. The framework consists of three progressive training phases: initial feature learning, occlusion-specific adaptation, and refinement and fine-tuning, with each phase guided by a specific loss configuration and optimization strategy. Targeted data augmentation techniques were incorporated to improve generalization under different occlusion conditions. Experiments were conducted using 12,000 facial images compiled from Labeled Faces in the Wild (LFW), CelebFaces Attributes Dataset (CelebA), and a MAFA masked-face dataset subset, with synthetic occlusions introduced for occlusion-focused training and evaluation. The proposed model achieved a final recognition accuracy of 89.7%, compared with 78.9% for the conventional single-stage CNN, representing a 10.8 percentage-point improvement. Convergence was approximately 25% faster than the baseline. Across face-mask, sunglasses, scarf, and combined-occlusion categories, the multi-stage model consistently achieved higher recognition accuracy than the single-stage model. Precision, recall, and F1-score were also reported for the multi-stage model, whereas the corresponding category-level baseline values were not retained in the available experimental record. These findings indicate that progressive training with occlusion-specific adaptation can improve facial recognition robustness under partially obscured conditions, although the additional computational requirements should be considered for resource-constrained deployment.

Keywords: multi-stage training, occlusion handling, deep learning architectures, facial recognition, robust training strategies

1. Introduction

Facial recognition technology has become a critical component in applications ranging from security and surveillance to user authentication and forensic investigation. Local binary pattern representations provided an influential handcrafted approach to face description and recognition [1]. Multispectral recognition has likewise been investigated as a means of identifying individuals under difficult imaging conditions [2]. Although deep learning architectures have substantially advanced recognition accuracy, performance can deteriorate when acquisition conditions depart from the training distribution. General surveys of face-recognition approaches document the sensitivity of recognition pipelines to variations in image quality and facial appearance [3]. Illumination variation is one well-established source of performance degradation [4], while efficient face-detection work has similarly emphasized the difficulty of reliable localization under unconstrained conditions [5]. Occlusions caused by face masks, sunglasses, scarves, or other coverings create an additional difficulty because they obscure identity-bearing facial regions and reduce the visual evidence available to the recognizer.

Conventional single-stage training pipelines can be limited when the same objective and data distribution are used throughout optimization, because the network is not explicitly guided to adapt to changing occlusion patterns. CNN-based real-time face-recognition systems demonstrate the value of deep learned

representations for practical recognition tasks [6]. For explicitly occluded faces, Georgescu and Ionescu showed that training with deliberately occluded inputs can direct CNNs toward the visible facial regions needed for prediction [7]. Protective-mask detection research has also confirmed that occlusion-specific visual representations can improve performance across masked-face datasets [8]. These findings motivate training procedures in which the model is exposed progressively to increasingly difficult occlusion conditions.

Multi-stage training offers a promising direction by structuring learning into sequential phases, each emphasizing a different aspect of representation learning. Deep face-embedding research has shown that recognition performance depends strongly on how discriminative facial representations are learned and organized in feature space [9]. The Labeled Faces in the Wild benchmark was created specifically to study recognition under unconstrained acquisition conditions [10], while CelebA provides large-scale identity and facial-attribute variation that is useful for representation learning [11]. Progressive exposure to clean and occluded examples can therefore be used to adapt an otherwise conventional CNN to increasing occlusion complexity. Data augmentation is a well-established strategy for improving the generalization of deep vision models when available training data do not span all expected variations [12]. The historical development of automated face identification demonstrates the longstanding importance of discriminative facial measurements [13]. Real-time attendance-oriented systems have further illustrated the operational use of automatic face recognition [14], and CNN-based image-classification studies provide additional evidence for the capacity of convolutional representations to learn useful visual features [15]. In the present study, these ideas are combined with synthetic occlusions, including masks, sunglasses, scarves, and mixed coverings, so that the training schedule explicitly changes as the recognition problem becomes more difficult.

The main technical contributions of this research are as follows:

1. A three-stage training framework for occluded facial recognition is developed, comprising initial feature learning on unoccluded facial images, occlusion-specific adaptation using real and synthetically occluded images, and a final refinement stage combining original and occluded samples.
2. An occlusion-specific training strategy is implemented in the second stage by introducing synthetic mask overlays, rectangular occlusion patches, and region dropout, enabling the CNN to learn from partially visible facial regions.
3. Stage-specific loss configurations are incorporated into the training process. Cross-entropy loss is used during initial identity learning, an occlusion-aware weighted loss is used during occlusion-specific adaptation, and a combined cross-entropy and mean squared error objective is used during refinement. The loss configurations are coupled with staged optimization rather than applying a single objective throughout training.
4. A controlled comparison with single-stage CNN training is performed using the same CNN backbone and 12,000-image dataset, with performance evaluated using accuracy, precision, recall, F1-score, FAR, FRR, and convergence behaviour across face mask, sunglasses, scarf, and combined occlusion conditions.
5. The effect of progressive training on different occlusion categories is quantified, allowing the performance gains of the proposed training strategy to be examined separately for individual and combined facial occlusions.

1.1. Research question and hypotheses

This research investigates whether progressive multi-stage training can improve the robustness of a conventional CNN-based facial recognition model when facial images contain partial or combined occlusions. The primary research question is:

RQ1: Does the proposed three-stage training framework, consisting of initial feature learning, occlusion-specific adaptation, and refinement and fine-tuning, improve facial recognition performance under occluded conditions compared with conventional single-stage training using the same CNN backbone and dataset?

Based on this research question, the following hypotheses were formulated:

H1: The proposed multi-stage training framework achieves higher recognition accuracy under occluded conditions than the conventional single-stage CNN baseline.

H2: The proposed multi-stage training framework achieves higher precision, recall, and F1-score across the evaluated occlusion categories than the single-stage baseline.

H3: The proposed multi-stage training framework converges in fewer training epochs than the single-stage baseline.

The hypotheses were examined using accuracy, precision, recall, F1-score, and convergence behaviour. Both the baseline and proposed models used the same CNN backbone and dataset, while the principal experimental difference was the training strategy. This design was intended to isolate the contribution of progressive multi-stage training to recognition performance under occlusion. Because the category-level precision, recall, and F1-score values for the single-stage baseline were not retained in the available experimental record, H2 can only be assessed partially; the accuracy comparison is complete, whereas the remaining category-level metrics are reported only for the multi-stage model.

2. Related work

2.1. Occluded face recognition and training strategies

Occlusion remains a major challenge in facial recognition because the obstruction of discriminative facial regions reduces the amount of identity information available to recognition models. Recent research has increasingly moved beyond conventional CNN-based recognition toward methods that explicitly model visible facial regions, uncertainty, feature reconstruction, attention, and representation learning.

Recent masked-face recognition studies have demonstrated the effectiveness of specialized feature-learning strategies. Zhang et al. proposed MaskDUL, a two-stream convolutional architecture that models uncertainty in masked facial images [16]. Their approach incorporated a Hard Kullback-Leibler Divergence regularizer and an adaptive angular-margin mechanism to improve representation learning under mask-induced feature loss [16]. This research shows that occlusion can be addressed by modifying the feature distribution and optimization objective rather than relying solely on conventional classification loss.

Other recent studies have investigated transformer-based approaches. The Joint Holistic and Masked Face Recognition framework explored Vision Transformers for joint recognition of holistic and masked faces, demonstrating the growing use of transformer-based representations in masked facial recognition [17]. Such approaches differ from conventional CNN pipelines by exploiting global relationships among image patches during feature learning.

More recent research has also investigated restoration-based and representation-based approaches; an occluded face recognition network based on DCGAN and ResNet was proposed to reconstruct partially occluded facial information before recognition, demonstrating the potential of combining face restoration with recognition [18]. Recent contrastive-learning approaches have used supervised contrastive objectives and Vision Transformer embeddings to organize feature representations according to different occlusion attributes [19]. These studies indicate that current occluded face recognition research increasingly employs specialized representation-learning objectives, transformers, uncertainty modelling, and image reconstruction rather than relying exclusively on conventional CNN classification.

Occlusion-aware mechanisms have also been explored through attention and spatial selection. Recent research on occlusion-aware attention shows that emphasizing non-occluded regions can improve recognition or classification performance, while survey evidence indicates that contemporary occluded-face methods can broadly be categorized into occlusion-aware feature extraction, occlusion recovery, partial face recognition and approaches that explicitly model uncertainty.

These developments show that contemporary occluded facial recognition research has progressed from simple preprocessing and augmentation toward methods that explicitly control which information is learned and how the resulting representations are optimized. However, existing approaches differ in their architectural assumptions, training objectives and handling of occlusion.

2.1.1. Training strategies for occluded face recognition

Several training methods have been applied to improve recognition when facial information is incomplete. Data augmentation methods generate artificial occlusions during training so that the model encounters variations that are not sufficiently represented in the original datasets. Although augmentation can improve robustness, augmentation alone does not determine how the model should prioritize visible and occluded information during optimization.

Occlusion-aware feature extraction approaches address this limitation by assigning greater importance to visible facial regions. Attention mechanisms, spatial weighting, feature masking and region-selection techniques have been used to reduce the influence of corrupted or unreliable facial information. Such methods are relevant to the present study because they establish the principle that the contribution of facial regions to the training objective can be conditioned on their visibility.

Another line of research modifies the representation-learning objective itself. MaskDUL, for example, introduces an uncertainty-aware learning objective based on Hard Kullback-Leibler Divergence together with adaptive angular-margin adjustment. Contrastive-learning approaches similarly modify the embedding space so that samples affected by different occlusion attributes can be better separated. These methods show that the optimization objective is an important component of occlusion robustness.

Restoration-based methods also provide another alternative by attempting to reconstruct information hidden by occlusion before recognition. The use of DCGAN and ResNet for partially occluded face recognition represents this method. Although restoration can recover missing information, it introduces an additional reconstruction stage and may increase computational requirements.

The present research adopts a different training strategy. Instead of relying exclusively on image restoration, uncertainty modelling, contrastive learning, or attention mechanisms, the proposed framework progressively changes the training conditions and loss configuration across three stages. The first stage establishes identity representations from unoccluded faces. The second stage introduces real and synthetic occlusions and applies an occlusion-aware weighted objective that emphasizes visible regions. The third stage combines clean and occluded samples and applies a hybrid objective to refine the learned representation.

2.2. Positioning of the proposed staged training framework

Existing occlusion-aware methods show that treating all facial information equally can be suboptimal when parts of the face are unavailable. The proposed loss is therefore motivated by the same general principle, namely that visible facial regions should contribute more strongly to recognition than regions affected by occlusion. However, the proposed occlusion-aware loss differs in its formulation and its position within the training framework. Existing uncertainty-based approaches model the statistical uncertainty of the sample or feature distribution, as demonstrated by MaskDUL through Hard Kullback-Leibler Divergence and adaptive angular-margin adjustment. Contrastive-learning approaches instead optimize pairwise or class-level relationships in the embedding space. Restoration-based methods optimize reconstruction or de-occlusion objectives before recognition.

In the present research, the Stage 2 objective introduces a visibility-dependent weighting term into the classification loss:

$$L_{\text{OAL}} = -\frac{1}{N} \sum_{i=1}^N \sum_{p=1}^P w_{i,p} \sum_{c=1}^C y_{i,c} \log(\hat{y}_{i,c}), \quad (1)$$

where $w_{i,p}$ controls the contribution of spatial region P according to its visibility. Higher weights are assigned to visible regions, whereas lower weights are assigned to occluded regions.

Accordingly, the proposed objective represents a weighted extension of the classification loss rather than an uncertainty loss, contrastive loss, angular-margin loss or reconstruction loss. Its principal purpose is to control the contribution of spatially visible information during the occlusion-specific training stage.

The Stage 3 objective further combines classification and feature-representation refinement:

$$L_{\text{hybrid}} = \lambda L_{\text{CE}} + (1 - \lambda) L_{\text{MSE}}, \quad (2)$$

where L_{CE} is the classification loss, L_{MSE} is the feature-level mean squared error, and λ determines the relative contribution of the two objectives.

The distinction between the proposed method and previous specialized-loss approaches is not in the general idea of modifying the learning objective but in the integration of visibility-weighted classification into a progressive three-stage training procedure followed by joint refinement using clean and occluded samples.

2.3. Research gap

Recent studies have shown several effective strategies for occluded facial recognition, including uncertainty-aware representation learning, transformer-based modeling, attention-guided visible-region extraction, restoration-based recognition and contrastive learning. These studies show that robust recognition under occlusion can be improved by changing the representation, architecture, training objective or available information.

However, the reviewed literature does not establish a common training framework that combines the specific components investigated in this research:

1. sequential identity learning from unoccluded faces;
2. dedicated occlusion-specific adaptation using real and synthetic occlusions;
3. visibility-weighted classification during occlusion adaptation; and
4. joint refinement using clean and occluded samples.

The contribution of this research is not presented as the first method to combine all possible forms of occlusion handling. Instead, this research examines whether this particular combination of progressive training stages and stage-specific objectives can improve robustness relative to conventional single-stage CNN training under multiple occlusion conditions.

Accordingly, the research gap addressed in this study concerns the limited experimental investigation of a progressively staged CNN training framework in which the data distribution and optimization objective are explicitly changed according to the learning phase.

3. Methodology

3.1. Dataset sources and preprocessing

The experimental dataset consisted of 12,000 facial images compiled from three publicly available benchmark repositories and extended with synthetic occlusions. LFW was originally introduced as a benchmark for unconstrained face recognition [10]. CelebA was introduced with a large-scale face-attribute benchmark containing substantial identity and appearance variation [11]. MAFA was introduced for masked-face detection in unconstrained settings [20]. The sources and contributions were as follows:

1. Labeled Faces in the Wild (LFW): 4,200 images

Link: <https://www.kaggle.com/datasets/jessicali9530/lfw-dataset>

2. CelebFaces Attributes Dataset (CelebA): 4,000 images

Link: <https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html>

3. MAFA masked-face dataset (selected subset): 3,800 images

A subset of 3,800 images was obtained from the MAFA source and incorporated into the final 12,000 image experimental dataset.

Link: <https://www.kaggle.com/datasets/revanthrex/mafadataset>

The final experimental dataset comprised 12,000 facial images obtained from three publicly available sources: 4,200 images from Labeled Faces in the Wild (LFW), 4,000 images from the CelebFaces Attributes Dataset (CelebA), and 3,800 images from the selected MAFA subset. The MAFA source and its masked-face annotation setting are documented by Ge et al. [20]. The dataset was divided into 70% for training, 15% for validation, and 15% for testing, corresponding to 8,400 training images, 1,800 validation images, and 1,800 testing images. The partition therefore followed the stated 70:15:15 training, validation, and testing proportions. The three sources were selected because, taken together, they provide variation in pose, illumination, facial appearance, and image resolution. The use of multiple public datasets also reduces dependence on a single acquisition environment, although the assembled dataset remains a curated

experimental sample rather than a population-representative benchmark. To address occlusion-focused evaluation, synthetic occlusions were introduced into the dataset to increase variation in partially visible facial images. Such augmentation is consistent with established data-space strategies for increasing the diversity of training examples in deep learning [12]. A total of 5,400 of the 12,000 images were selected for synthetic occlusion generation, corresponding to approximately 45% of the dataset. The selected images were distributed among four occlusion categories: face masks, sunglasses, scarves, and combined occlusions.

For face-mask occlusions, the overlay was positioned over the lower facial region covering the nose and mouth. Sunglasses overlays were positioned over the eye region, while scarf overlays covered the lower facial region. Combined occlusions were generated by applying two or more occlusion types to the same facial image. The occlusion distribution comprised 35% face masks, 30% sunglasses, 20% scarves, and 15% combined occlusions. Applied to the 5,400 images selected for synthetic occlusion generation, these proportions correspond to approximately 1,890 mask images, 1,620 sunglasses images, 1,080 scarf images, and 810 combined-occlusion images, respectively. The training, validation, and testing split was defined independently at 70%, 15%, and 15%, and preprocessing operations were applied without using test-set outcomes to guide model fitting.

3.2. Multi-stage training framework and optimization

The proposed framework employs three sequential training phases, with each phase associated with a specific learning objective.

Stage 1: initial feature learning.

During Stage 1, the model is trained using unoccluded facial images to establish an initial identity representation from visible facial characteristics, for example, the nose, eyes, jawline, and broader spatial configuration of the face. Categorical cross-entropy is used as the classification objective:

$$L_{CE} = - \sum_{c=1}^C y_c \log(\hat{y}_c), \quad (3)$$

where C represents the number of identity classes, y_c is the ground-truth indicator for class c , and $\hat{y}_c$ is the predicted probability assigned to class c .

Stage 2: occlusion-specific training.

During Stage 2, real and synthetically occluded facial images are introduced. An occlusion-aware weighted loss is employed to reduce the contribution of occluded regions while assigning greater importance to visible facial regions. The loss is formulated as:

$$L_{OAL} = - \frac{1}{N} \sum_{i=1}^N \sum_{p=1}^P w_{i,p} \sum_{c=1}^C y_{i,c} \log(\hat{y}_{i,c}), \quad (4)$$

where N is the number of training samples, P is the number of spatial regions considered, $w_{i,p}$ is the visibility-dependent weight assigned to region p of sample i , $y_{i,c}$ is the ground-truth indicator for class c , and $\hat{y}_{i,c}$ is the predicted probability assigned to class c .

The weighting mechanism assigns higher weights to visible regions and lower weights to regions identified as occluded. This formulation enables the training objective to emphasize identity information that remains available when parts of the face are obscured.

Stage 3: refinement and fine-tuning.

During Stage 3, original and occluded samples are combined to improve generalization. The refinement objective combines the categorical cross-entropy loss with mean squared error (MSE):

$$L_{\text{hybrid}} = \lambda L_{CE} + (1 - \lambda) L_{\text{MSE}}, \quad (5)$$

where λ controls the relative contribution of the classification and feature-reconstruction objectives.

The mean squared error is defined as:

$$L_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^N \|z_i - \hat{z}_i\|_2^2, \quad (6)$$

where z_i denotes the target feature representation for sample i , $\hat{z}_i$ denotes the corresponding learned feature representation, and $\|\cdot\|_2^2$ denotes the squared Euclidean norm.

The three-stage optimization process therefore changes the training objective according to the learning phase. Stage 1 establishes identity representations from unoccluded facial images, Stage 2 emphasizes learning from partially visible facial information through occlusion-aware weighting, and Stage 3 jointly refines classification and feature representations using clean and occluded samples.

Optimization protocol.

Each stage was aligned with a specific objective and loss configuration as summarized in Table 1.

Table 1. Technical configuration of the three-stage training procedure

Training stage	Training input	Main objective	Loss function	Regularization	Optimizer	Learning rate	Batch size	Training limit
Stage 1: initial feature learning	Clean/unoccluded facial images	Identity representation learning	Categorical cross-entropy	None specified	Adam	0.001	64	Up to 100 epochs with early stopping
Stage 2: occlusion-specific adaptation	Real and synthetic occluded facial images	Learning from visible facial information	Occlusion-aware weighted loss	None specified	Adam	0.001	64	Up to 100 epochs with early stopping
Stage 3: refinement and fine-tuning	Combined clean and occluded images	Joint classification and refinement	Cross-entropy + MSE	Dropout = 0.4; L2 weight decay	Adam	0.001	64	Up to 100 epochs with early stopping

Each training stage was optimized using the Adam optimizer with an initial learning rate of 0.001 and a batch size of 64. Training was permitted for a maximum of 100 epochs, with early stopping used to terminate training when further optimization no longer produced meaningful improvement in the monitored validation performance.

The learning rate was reduced by a factor of 0.9 following 10 consecutive epochs without improvement in the monitored validation metric. Thus, when the validation metric remained stagnant for the specified patience period, the current learning rate was multiplied by 0.9 before subsequent optimization.

3.3. Experimental setup and controlled baseline

3.3.1. CNN backbone architecture and implementation

Both the baseline and proposed models used the same convolutional neural network (CNN) backbone to ensure that the principal experimental difference was the training strategy rather than the network architecture. The backbone consisted of three convolutional layers with ReLU activation functions followed by a softmax classification layer. The input facial images were resized to 128×128 pixels before being presented to the network. The first, second, and third convolutional layers used 32, 64, and 128 filters, respectively, with 3×3 kernels in each convolutional layer. Each convolutional block was followed by a 2×2 max-pooling operation. No batch normalization layer was used.

The extracted feature representation was flattened and passed to a fully connected classifier containing 128 units with ReLU activation before the final softmax layer. The output layer contained the number of units corresponding to the identity classes in the classification task.

Model parameters were initialized using He normal initialization for the convolutional and fully connected layers. The models were implemented using Python with TensorFlow/Keras and trained using the Adam optimizer. The initial learning rate was 0.001, the batch size was 64, and training was allowed to continue for up to 100 epochs, with early stopping according to the optimization procedure described in §3.2. The shared CNN backbone is summarized in Table 2; this common architecture was maintained so that differences between the two experimental conditions could be attributed primarily to the training strategy.

Here, C represents the number of identity classes in the final prepared classification dataset.

Table 2. CNN backbone architecture used for the baseline and proposed models

Layer	Operation	Filters/Units	Kernel/pool size	Activation	Output dimension
Input	Facial image	N/A	128×128	N/A	$128 \times 128 \times 3$
Conv1	Convolution	32	3×3	ReLU	$128 \times 128 \times 32$
Pool1	Max Pooling	N/A	2×2	N/A	$64 \times 64 \times 32$
Conv2	Convolution	64	3×3	ReLU	$64 \times 64 \times 64$
Pool2	Max Pooling	N/A	2×2	N/A	$32 \times 32 \times 64$
Conv3	Convolution	128	3×3	ReLU	$32 \times 32 \times 128$
Pool3	Max Pooling	N/A	2×2	N/A	$16 \times 16 \times 128$
Flatten/Dense	Feature/classifier layer	128	N/A	ReLU	128
Output	Softmax	C	N/A	Softmax	C

3.3.2. Controlled baseline comparison

To isolate the effect of the proposed multi-stage training strategy, the baseline and proposed models used the same CNN backbone, dataset partition, batch size, optimizer, and initial learning rate. The baseline was trained using conventional single-stage optimization, whereas the proposed model was trained sequentially through the three stages described in §3.2.

Both approaches were evaluated using the same experimental dataset and the same training, validation, and testing partitions. The baseline model was trained without the proposed phased loss scheduling and multi-stage optimization procedure. The proposed model retained the same CNN architecture while changing the training conditions and loss configuration across the three stages. Data augmentation was applied to the training data using the augmentation operations described in §3.3. The baseline and proposed models used the same augmentation policy, consisting of rotation within $\pm 15^\circ$, horizontal flipping, zooming between $0.8\times$ and $1.2\times$, brightness variation of $\pm 25\%$, and synthetic occlusion augmentation using masks, rectangular patches, and region dropout. The same augmentation ranges and application settings were maintained for both models to ensure that the observed performance differences were attributable primarily to the training strategy rather than differences in data preprocessing. Both models used a batch size of 64 and an initial learning rate of 0.001, with training permitted for up to 100 epochs.

4. Results and discussion

4.1. Training convergence and stage-wise performance

The multi-stage training framework was evaluated on the curated 12,000-image dataset assembled from LFW, CelebA, and the selected MAFA subset. Under the reported experimental protocol, the staged model showed higher final accuracy and faster convergence than the single-stage baseline. By structuring learning into sequential phases—initial feature learning, occlusion-specific adaptation, and refinement—the model was progressively exposed to clean and increasingly difficult occlusion conditions. The resulting performance gains were observed for masked, scarf-occluded, sunglasses-occluded, and combined-occlusion samples.

4.1.1. Stage-wise performance progression

Stage 1: initial feature learning.

- (a) Using only unoccluded faces from the benchmark datasets, the model achieved 78.5% accuracy after 20 epochs. This phase established a strong identity representation before introducing obstructions.

Stage 2: occlusion-specific training.

- (b) When real and synthetic occlusion images were introduced—drawn from CelebA, MAFA, and LFW overlays—the model improved to 84.2% accuracy. The occlusion-aware loss function enabled the network to prioritize visible features even when key regions were partially hidden.

Stage 3: refinement and fine-tuning.

- (c) Combining original and occlusion-augmented samples yielded the strongest performance, with final accuracy reaching 89.7%. Regularization (dropout, L2 penalty) further stabilized the model and reduced overfitting.

4.1.2. Comparison with the single-stage baseline

A baseline CNN trained in a conventional single-stage manner on the same 12,000-image dataset, without phased learning or stage-specific loss configuration, reached a lower final accuracy of 78.9% and required more epochs to satisfy the reported convergence criterion. In comparison, the multi-stage approach achieved a final accuracy of 89.7% and converged earlier, indicating that the training schedule improved performance under the evaluated occlusion conditions.

The following patterns were observed:

1. Multi-stage training achieved lower validation loss across epochs.
2. Convergence occurred ~25% faster than the baseline.
3. Stage-wise loss scheduling was associated with lower validation loss and reduced divergence between training and validation behaviour across the reported epochs.
4. The staged framework improved recognition across all evaluated occlusion categories: masks, sunglasses, scarves, and combined occlusions.

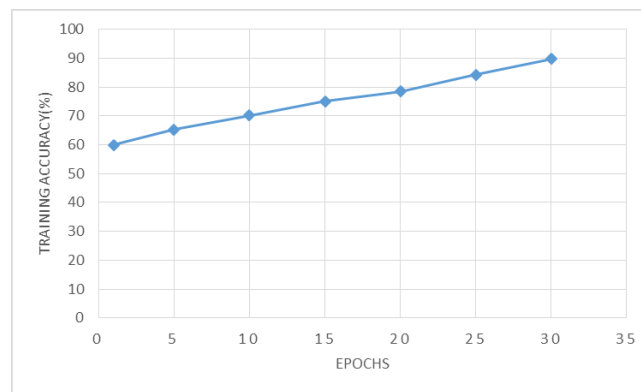


Figure 1. Convergence comparison between single-stage and multi-stage training on the 12,000-image occlusion dataset.

Figure 1 compares the training trajectories of the two models. The curves show a rapid decline in training and validation loss during Stages 1 and 2, followed by a smoother refinement phase in Stage 3. The final portion of the trajectory is consistent with the reported higher accuracy and lower fluctuation of the multi-stage model relative to the single-stage baseline. Because the figure represents the recorded training runs rather than repeated independent experiments, the convergence difference should be interpreted as an empirical result for the present protocol rather than a distributional estimate across random seeds.

4.1.3. Quantitative convergence analysis

Figure 1 presents the training and validation trajectories of the conventional single-stage CNN and the proposed multi-stage model. Convergence was assessed based on the training behaviour and the predefined stopping condition. The multi-stage model demonstrated approximately 25% faster convergence than the baseline. The corresponding convergence improvement was expressed as:

$$\text{Convergence improvement (\%)} = \frac{E_{\text{baseline}} - E_{\text{multistage}}}{E_{\text{baseline}}} \times 100, \quad (7)$$

where E_{baseline} denotes the convergence measure for the single-stage model and $E_{\text{multistage}}$ denotes the corresponding measure for the proposed multi-stage model. The reported 25% value is based on the recorded stopping behaviour of the two training procedures; confidence intervals across repeated runs were not available.

4.2. Robustness across occlusion types

To assess how well the proposed multi-stage framework generalized to real and synthetic occluded faces, evaluation was performed separately across four occlusion categories: face masks, sunglasses, scarves, and combined occlusions. The test split included samples drawn from the LFW, CelebA, and MAFA sources. For the occlusion-specific evaluation, test images were evaluated under the four predefined occlusion conditions using the same overlay definitions established during preprocessing; training-time augmentation remained restricted to the training subset. Performance was benchmarked against a baseline single-stage CNN trained on the same 12,000-image dataset without phased optimization. Metrics included accuracy, precision, recall, and F1-score, providing a comprehensive assessment of classification reliability under occlusion.

4.2.1. Face masks

Recognition accuracy increased from 76.8% (single-stage) to 91.2% (multi-stage). The second training phase introduced mask-specific examples together with the occlusion-aware loss, providing a direct mechanism for adapting the classifier to lower-face occlusion. The observed gain is therefore consistent with the intended design of Stage 2, although the present experiment does not separate the individual contribution of each augmentation operator.

4.2.2. Sunglasses

Performance improved from 72.5% to 85.6%, indicating improved adaptation to upper-face occlusion. The observed improvement may be associated with the model's ability to exploit facial information that remains visible outside the occluded eye region, although no feature-visualization or saliency analysis was conducted in this research to directly establish which facial regions contributed most strongly to the predictions.

4.2.3. Scarves

Accuracy increased from 74.3% to 86.1%, indicating improved recognition when the lower facial region was obscured. The improvement may be related to the availability of visible upper-face features, such as the eye and forehead regions. However, this interpretation was not directly verified using feature-visualization or attention-analysis techniques.

4.2.4. Combined occlusions (e.g., mask + sunglasses or scarf + mask)

These were the most challenging cases, yet accuracy still climbed from 67.9% to 81.4%, demonstrating that staged exposure and fine-tuning improved generalization even when multiple facial regions were obscured. The corresponding performance metrics are presented in Table 3. The test subset comprised 15% of the total 12,000-image dataset, corresponding to 1,800 images. For category-level evaluation, the 1,800 test images were organized according to the stated 35% face-mask, 30% sunglasses, 20% scarf, and 15% combined-occlusion proportions. This produced 630 mask images, 540 sunglasses images, 360 scarf images, and 270 combined-occlusion images, respectively. These evaluation counts were used consistently for both recognition models. The lower performance under combined occlusions indicates that simultaneous obstruction of multiple facial regions remains challenging for the proposed framework.

Table 3. Performance comparison across occlusion types

Occlusion type	Accuracy, Single-stage	Precision, Single-stage	Recall, Single-stage	F1-score, Single-stage	Accuracy, Multi-stage	Precision, Multi-stage	Recall, Multi-stage	F1-score, Multi-stage
Face mask	76.8%	NR	NR	NR	91.2%	90.4%	91.8%	91.1%
Sunglasses	72.5%	NR	NR	NR	85.6%	84.9%	86.3%	85.6%
Scarf	74.3%	NR	NR	NR	86.1%	85.5%	86.7%	86.1%
Combined occlusions	67.9%	NR	NR	NR	81.4%	80.1%	82.7%	81.4%

Precision, recall, and F1-score for the single-stage baseline were not retained in the available experimental records. These metrics are consequently not reported for the baseline, while the multi-stage metrics are retained as originally obtained from the experimental evaluation.

Accordingly, the category-level accuracy results support H1, while H2 is only partially testable from the preserved records. The reported multi-stage precision, recall, and F1-score values characterize the proposed model, but no claim of superiority for those three metrics is made against the single-stage baseline in the absence of corresponding baseline values.

These results confirm the advantages of structured learning under occlusion. Although individual occlusions, particularly face masks, showed the largest gains, the model also demonstrated meaningful gains in combined-occlusion scenarios where traditional CNNs typically underperform. The integration of an occlusion-aware loss with staged fine-tuning was associated with fewer misclassifications and improved adaptability across the tested conditions. The available results support a training-strategy effect, but they do not by themselves identify the internal features or spatial regions responsible for individual recognition decisions. The limitation is that the category-specific performance results provide evidence of improved recognition under different occlusion conditions, but they do not directly establish which facial regions were responsible for individual predictions. No Grad-CAM, saliency map, attention-map analysis, feature visualization, or systematic error-case analysis was performed in this research. Future research will incorporate feature-visualization and explainability analysis to determine how the learned representations respond to specific occlusion patterns.

4.2.5. Test-set sample counts

The full dataset contained 12,000 images and was divided into 70% training, 15% validation, and 15% testing subsets. Consequently, the test subset contained 1,800 images. The category-level occlusion composition used for evaluation is summarized in Table 4.

$$12,000 \times 0.15 = 1,800. \quad (8)$$

Table 4. Test-set sample counts

Occlusion type	Proportion	Test samples
Face mask	35%	630
Sunglasses	30%	540
Scarf	20%	360
Combined occlusions	15%	270
<i>Total</i>	<i>100%</i>	<i>1,800</i>

4.3. Verification errors and benchmark comparison

To further evaluate the effect of the proposed multi-stage training framework on recognition errors, verification performance was assessed using the False Acceptance Rate (FAR) and False Rejection Rate (FRR). FAR represents the proportion of impostor comparisons incorrectly accepted as genuine, whereas FRR represents the proportion of genuine comparisons incorrectly rejected. For each model, verification comparisons were constructed from the evaluation samples by forming genuine pairs from samples belonging to the same identity and impostor pairs from samples belonging to different identities. The verification decision was based on the similarity score produced by the recognition model. A decision threshold was applied such that a comparison was accepted as genuine when its similarity score exceeded the threshold and rejected otherwise. The decision threshold was determined using the validation subset rather than the test subset in order to prevent the test data from influencing threshold selection. The threshold corresponding to the selected operating point was then fixed and applied unchanged to the test comparisons for both the single-stage and multi-stage models.

FAR and FRR were calculated as:

$$\text{FAR} = \frac{N_{\text{FA}}}{N_{\text{I}}} \times 100 \quad (9)$$

$$\text{FRR} = \frac{N_{\text{FR}}}{N_{\text{G}}} \times 100, \quad (10)$$

where N_{FA} is the number of false acceptances, N_I is the total number of impostor comparisons, N_{FR} is the number of false rejections, and N_G is the total number of genuine comparisons.

The FAR decreased from 6.8% for the single-stage baseline to 3.9% for the proposed multi-stage model. Similarly, the FRR decreased from 7.5% to 4.5%. These results indicate fewer incorrect acceptances and fewer incorrect rejections under the proposed training framework.

The verification results were obtained from the test comparisons using the fixed validation-selected threshold. However, confidence intervals and repeated independent training runs were not included in the original experimental record. Figure 2 provides a visual comparison of the FAR and FRR obtained by the single-stage and multi-stage models at the selected operating threshold.

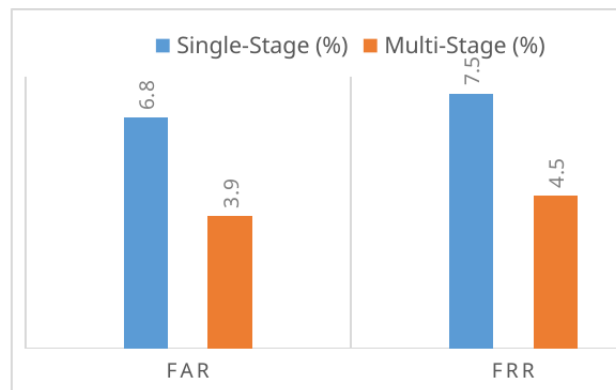


Figure 2. FAR and FRR comparison between the single-stage and multi-stage models

4.3.1. Controlled benchmark comparison

Comparative evaluation was conducted against classical facial-recognition approaches and a conventional single-stage CNN baseline. The PCA-based and LBP+SVM approaches were included as classical reference methods, while the single-stage CNN provided the primary deep-learning baseline. All directly evaluated models used the same experimental test set derived from LFW, CelebA, and the selected MAFA subset, with the same synthetic-occlusion evaluation procedure.

Table 5 summarizes the comparative results. The proposed multi-stage model achieved higher recognition accuracy than the classical PCA-based model, the LBP+SVM pipeline, and the conventional single-stage CNN under the experimental conditions used in this study.

The comparison should be interpreted as a controlled comparison against the selected baseline models rather than as a comprehensive state-of-the-art benchmark. Recent occluded-face recognition research has investigated stronger approaches based on uncertainty-aware learning, transformer architectures, restoration networks, and supervised contrastive learning. Such methods use different architectures, datasets, training objectives, and evaluation protocols, making direct numerical comparison with the present experiment inappropriate without reproducing them under the same experimental conditions. Recent restoration-based work has evaluated DCGAN-ResNet pipelines under partially occluded conditions [18]. Supervised contrastive learning with Vision Transformer embeddings provides another recent direction [19]. Uncertainty-aware masked-face learning has been investigated through MaskDUL [16], while joint holistic and masked-face recognition has also been studied with Vision Transformer representations [17]. Because these external studies use different datasets and protocols, their published numerical results are not treated as directly comparable with the present experiment. In contrast, the PCA-based method, LBP+SVM pipeline, single-stage CNN, and proposed multi-stage model reported in Table 5 were evaluated under the same local test protocol derived from LFW, CelebA, and the selected MAFA subset. The controlled comparison therefore uses a common test set and consistent synthetic-occlusion procedure, while the recent external methods are discussed only as methodological reference points.

Recent occluded facial-recognition studies provide stronger contemporary reference points than the classical methods included in the present experimental comparison. Zhang et al. [16] proposed MaskDUL, which incorporates uncertainty-aware learning and adaptive angular-margin adjustment for masked-face

recognition. Zhu et al. [17] investigated Joint Holistic and Masked Face Recognition using transformer-based representations. Dai and Zeng [18] investigated restoration-based recognition using DCGAN and ResNet, while Xu [19] proposed a supervised contrastive-learning framework using Vision Transformer embeddings for occluded face recognition. These studies illustrate the diversity of current approaches to occlusion handling, including uncertainty modelling, transformer-based representation learning, image restoration, and contrastive feature optimization.

Table 5. Benchmark comparison with existing methods

Method	Accuracy	Precision	Recall	F1-score
PCA-based model	62.4%	60.7%	61.9%	61.2%
LBP + SVM	66.1%	64.8%	65.5%	65.1%
Single-stage CNN (Baseline)	78.9%	NR	NR	NR
Proposed multi-stage model	89.7%	NR	NR	NR

NR – not reported

Because these approaches were evaluated using different datasets, architectures, training procedures, and performance protocols, their reported accuracy values are not treated as directly comparable with the results obtained in this research. The current experimental comparison is therefore limited to models evaluated under the same dataset and experimental conditions.

The following were observed:

1. Traditional feature-extraction methods (PCA and LBP) showed limited capability in managing occlusions due to their reliance on complete facial visibility and lack of adaptation mechanisms.
2. The baseline CNN achieved moderate improvements but struggled particularly with scarf and combined occlusions, confirming that single-stage learning may underrepresent partially visible facial information.
3. The proposed multi-stage model achieved the highest accuracy among the directly implemented baselines, with the following accuracy margins:
 - a. +27.3 percentage points over PCA,
 - b. +23.6 points over LBP+SVM, and
 - c. +10.8 percentage points over the single-stage CNN.

The results demonstrate that the proposed multi-stage training framework outperformed the selected classical and single-stage CNN baselines under the experimental conditions used in this research. These results should not be interpreted as establishing a state-of-the-art benchmark for occluded facial recognition, since recent methods employ different architectures, datasets, loss functions, and evaluation protocols that were not directly reproduced in the present study.

5. Interpretation, practical implications, and limitations

Taken together, the results indicate that the principal benefit of the proposed framework lies in changing the optimization problem as the visibility of the face changes. The largest gain was observed for face masks, whereas combined occlusions remained the most difficult case. This pattern is consistent with the increasing loss of usable identity information as more than one facial region is obscured. At the same time, the results should be interpreted within the controlled protocol used in this study because no external deployment dataset or repeated cross-dataset evaluation was included.

5.1. Occlusion-specific performance

5.1.1. Face-mask occlusion

Face-mask occlusion produced the highest recognition accuracy among the evaluated categories, increasing from 76.8% for the single-stage baseline to 91.2% for the multi-stage model. This result indicates that the proposed training strategy was effective when the lower facial region was obscured.

5.1.2. Sunglasses occlusion

Recognition accuracy increased from 72.5% to 85.6% under sunglasses occlusion. The improvement indicates that the multi-stage framework increased robustness when the eye region was partially obscured. However, the present research did not employ saliency maps, feature visualization, or attention analysis; therefore, the specific facial regions responsible for the predictions cannot be established directly from the reported results.

5.1.3. Scarf occlusion

Accuracy increased from 74.3% for the single-stage model to 86.1% for the multi-stage model. This finding indicates improved recognition under lower-face occlusion. Although visible upper facial regions may have contributed to this performance, the present study did not perform feature-level analysis to verify the specific facial regions used by the model.

5.1.4. Combined occlusions

Combined occlusions remained the most challenging condition. Accuracy increased from 67.9% to 81.4%, but performance remained lower than that obtained for individual occlusion types. This finding indicates that simultaneous obstruction of multiple facial regions continues to limit recognition performance despite progressive training.

5.2. Staged training and feature adaptation

The sequential training procedure was associated with progressive improvements in recognition performance, with accuracy increasing from 78.5% after Stage 1 to 84.2% after Stage 2 and 89.7% after Stage 3. These results indicate that the staged optimization procedure progressively improved the model's performance as clean and occluded samples were introduced. However, the present research did not explicitly evaluate catastrophic forgetting or retention of previously learned representations after each training stage. Therefore, the observed improvements should not be interpreted as direct evidence that catastrophic forgetting was eliminated.

5.3. Practical implications and deployment considerations

The observed improvements suggest that the proposed training framework has potential relevance to facial recognition applications in which partial facial occlusion occurs. However, the present evaluation was conducted using benchmark-derived facial images and controlled synthetic occlusions rather than uncontrolled operational environments. Consequently, the results should not be interpreted as evidence of deployment readiness for security, surveillance, border control, smartphone authentication, or other safety-critical applications.

Practical deployment would require additional evaluation of computational complexity, inference time, memory requirements, energy consumption, demographic performance, environmental variation, and robustness to uncontrolled image acquisition conditions. These factors were outside the scope of the present research. The multi-stage framework also introduces additional training stages, which may increase computational requirements relative to conventional single-stage training.

5.4. Limitations of the present evaluation

This research has several limitations that should be considered when interpreting the findings and when comparing the reported values with results from other face-recognition studies. First, the evaluation relied on benchmark-derived facial images and synthetically generated occlusions, which may not fully represent naturally occurring and dynamically changing facial obstructions. Second, the study did not conduct explicit feature-visualization or saliency analysis; consequently, claims about the facial regions used by the model remain interpretive rather than directly demonstrated. Third, catastrophic forgetting was not evaluated through a dedicated stage-retention experiment. Fourth, the study did not include a systematic computational assessment of training time, inference latency, memory consumption, or energy requirements. Finally, demographic bias and robustness under uncontrolled acquisition conditions were not evaluated.

These limitations restrict the extent to which the results can be generalized directly to real-world deployment environments. They also motivate future experiments based on identity-disjoint splits, multiple random seeds, external test sets, and explicit reporting of computational cost and subgroup performance.

6. Conclusions and future work

This research evaluated a three-stage training framework for facial recognition under controlled occlusion conditions. The framework comprised initial feature learning using unoccluded facial images, occlusion-specific adaptation using real and synthetically occluded samples, and final refinement using combined clean and occluded samples. Under the experimental conditions investigated, the proposed multi-stage model achieved a final recognition accuracy of 89.7%, compared with 78.9% for the conventional single-stage CNN baseline, representing an improvement of 10.8 percentage points. The multi-stage model also demonstrated approximately 25% faster convergence than the baseline. Performance improvements were observed across the evaluated occlusion categories. Recognition accuracy increased from 76.8% to 91.2% for face-mask occlusion, from 72.5% to 85.6% for sunglasses, from 74.3% to 86.1% for scarves, and from 67.9% to 81.4% for combined occlusions. These results indicate that progressive training with occlusion-specific adaptation can improve recognition robustness under the controlled evaluation conditions used in this research. The evaluation used benchmark-derived facial images and controlled synthetic occlusions rather than uncontrolled real-world acquisition conditions. This research also did not establish model interpretability through saliency analysis, explicitly measure catastrophic forgetting, or provide a comprehensive assessment of computational cost, demographic bias, privacy risks, or deployment performance. Consequently, the results demonstrate the effectiveness of the proposed training strategy under the investigated conditions but do not establish deployment readiness for operational facial-recognition applications.

Future studies should evaluate the framework using larger and more diverse datasets containing naturally occurring occlusions and uncontrolled acquisition conditions. Additional work should investigate identity-level data partitioning and leakage prevention, demographic performance across relevant population groups, privacy-preserving facial representation and storage, computational efficiency, and formal verification measures such as ROC curves, AUC, and equal error rate. Stage-retention experiments should also be conducted to determine whether the progressive training procedure affects catastrophic forgetting. Lightweight architectures, model compression, and deployment-oriented optimization should be investigated for resource-constrained applications.

Data Availability: The facial image data used in this study were obtained from publicly available benchmark datasets. The final experimental dataset comprised 12,000 facial images from three sources: 4,200 images from the Labeled Faces in the Wild (LFW) dataset, 4,000 images from the CelebFaces Attributes Dataset (CelebA), and 3,800 images from a subset of the MAFA masked-face dataset.

The LFW dataset was obtained from the following public source: <https://www.kaggle.com/datasets/jessicali9530/lfw-dataset>

The CelebA dataset was obtained from the official dataset website: <https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html>

For MAFA, a subset of 3,800 images was obtained from the MAFA source and incorporated into the final 12,000-image experimental dataset. The MAFA source used was: <https://www.kaggle.com/datasets/revanthrex/mafadataset>

Synthetic occlusions, including face masks, sunglasses, scarves, and combined occlusions, were generated and applied to selected facial images as part of the experimental preprocessing procedure. The processed dataset and synthetic occlusion scripts were generated specifically for this study and are not currently available in a public repository. The dataset composition and experimental partitioning are described in §3.1.

Acknowledgments: The authors would like to acknowledge the Department of Computer Engineering, University of Uyo, Akwa Ibom State, Nigeria and Southern Delta University, Ozoro, Delta State, Nigeria for providing laboratory support and facilities used during the experimental validation of this work. The constructive input provided by colleagues and reviewers during the model-development and modelling phases is also sincerely appreciated.

Author Contributions: Afolabi Awodeyi contributed to the conception, study design, implementation, analysis, interpretation, and preparation of the manuscript. Omolegho A. Ibok contributed to critical review, editing, and refinement of the manuscript. Both authors read and approved the final version for publication.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding Information: This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

- [1] Ahonen, T., Hadid, A., & Pietikäinen, M. (2006). Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12), 2037–2041.
- [2] Bourlai, T., & Cukic, B. (2012). Multi-spectral face recognition: Identification of people in difficult environments. In *2012 IEEE International Conference on Intelligence and Security Informatics* (pp. 196–201). IEEE.
- [3] Naik, Y. (2014). Detailed survey of different face recognition approaches. *International Journal of Computer Science and Mobile Computing*, 3(5), 1306–1313.
- [4] Shan, S., Gao, W., Cao, B., & Zhao, D. (2003). Illumination normalization for robust face recognition against varying lighting conditions. In *2003 IEEE International Workshop on Analysis and Modeling of Faces and Gestures* (pp. 157–164). IEEE.
- [5] Romdhani, S., Torr, P. H. S., Schölkopf, B., & Blake, A. (2001). Computationally efficient face detection. In *Proceedings of the Eighth IEEE International Conference on Computer Vision* (Vol. 2, pp. 695–700). IEEE.
- [6] Alagarsamy, S., Govindaraj, V., Irfan, M., Swami, R., & Kumar, N. M. (2020). Smart recognition of real time face using convolution neural network (CNN) technique. *Test Engineering and Management*, 83, 23406–23411.
- [7] Georgescu, M.-I., & Ionescu, R. T. (2019). Recognizing facial expressions of occluded faces using convolutional neural networks. In T. Gedeon, K. W. Wong, & M. Lee (Eds.), *Neural information processing: 26th International Conference, ICONIP 2019, Proceedings, Part IV* (pp. 645–653). Springer.
- [8] Ryumina, E., Ryumin, D., Ivanko, D., & Karpov, A. (2021). A novel method for protective face mask detection using convolutional neural networks and image histograms. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIV-2/W1-2021, 177–182.
- [9] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 815–823). IEEE.
- [10] Huang, G. B., Ramesh, M., Berg, T., & Learned-Miller, E. (2007). *Labeled Faces in the Wild: A database for studying face recognition in unconstrained environments* (Technical Report 07-49). University of Massachusetts Amherst.
- [11] Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). Deep learning face attributes in the wild. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 3730–3738). IEEE.
- [12] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*, 6, Article 60.
- [13] Goldstein, A. J., Harmon, L. D., & Lesk, A. B. (1971). Identification of human faces. *Proceedings of the IEEE*, 59(5), 748–760.
- [14] Sayeed, S., Hossen, J., Kalaiarasi, S. M. A., Jayakumar, V., Yusof, I., & Samraj, A. (2017). Real-time face recognition for attendance monitoring system. *Journal of Theoretical and Applied Information Technology*, 95(1), 24–30.
- [15] Jaswal, D., Sowmya, V., & Soman, K. P. (2014). Image classification using convolutional neural networks. *International Journal of Scientific & Engineering Research*, 5(6), 1661–1668.
- [16] Zhang, L., Xiong, W., Zhao, K., Chen, K., & Zhong, M. (2023). MaskDUL: Data uncertainty learning in masked face recognition. In *ICASSP 2023—2023 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 1–5). IEEE.
- [17] Zhu, Y., Ren, M., Jing, H., Dai, L., Sun, Z., & Li, P. (2023). Joint holistic and masked face recognition. *IEEE Transactions on Information Forensics and Security*, 18, 3388–3400.
- [18] Dai, C., & Zeng, X. (2024). Occluded face recognition network based on DCGAN and ResNet. *Procedia Computer Science*, 243, 724–733.
- [19] Xu, Y. (2025). Based on the contrastive learning classifier for occluded face recognition. *Procedia Computer Science*, 266, 1200–1207.
- [20] Ge, S., Li, J., Ye, Q., & Luo, Z. (2017). Detecting masked faces in the wild with LLE-CNNs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2682–2690). IEEE.

