IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 20 September 2023, accepted 9 October 2023, date of publication 20 October 2023, date of current version 27 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3326235

RESEARCH ARTICLE

Enhancing Accuracy of Face Recognition in Occluded Scenarios With Occlusion-Aware Module-Based Network

DALIN WANG AND RONGFENG LI
School of Intelligent Science and Technology, Chongqing Preschool Education College, Chongqing 404047, China

Corresponding author: Dalin Wang (wdl66@cqyz.edu.cn)

This work was supported by the Project of Chongqing Higher Education Association (Grant No.:CQGJ21B134) and the 2020 Major Project of the Chongqing Municipal Education Commission on Higher Education Teaching Reform (Grant No.: 201028).

ABSTRACT Face recognition technology despite its extensive application across various domains. However, occlusion factors like masks and glasses are significantly impeded by current face recognition models, resulting in limitations to their practical usage. We present Occlusion-Aware Module Network called Occlusion-Aware Module-based Network (OAM-Net), designed to enhance the accuracy of occluded face recognition. OAM-Net comprises two sub-networks: an occlusion-aware sub-network and a key-region-aware sub-network. The occlusion-aware sub-network incorporates an attention module to adaptively modify the weights of convolutional kernels for optimizing the processing of occluded face images. Meanwhile, the key-region-aware sub-network integrates a Spatial Attention Residual Block (SARB) for precise identification and localization of key facial regions. The network's generalization performance and accuracy are further enhanced by implementing a meta-learning-based strategy to boost the network's generalization performance and accuracy. Experimental results affirm OAM-Net's superior performance of OAM-Net over other state-of-the-art methods in occluded face recognition, underlining its significant potential for practical application.

INDEX TERMS Face recognition, occlusion, spatial attention residual block.

I. INTRODUCTION

Face recognition technology integral to a plethora of applications in a wide range of applications, provides a versatile, efficient solution across diverse sectors. In recent years, the technology has been adopted in security systems, access control, social media, marketing, and numerous smart devices, demonstrating its growing importance and relevance in everyday life [1]. By automatically detecting, verifying, and identifying human faces within images or video frames, face recognition systems offer confer numerous benefits, including bolstered security, enhanced user experience, and streamlined business processes. In the security and surveillance domain, face recognition technology serves multiple roles, such as suspect identification, crowd monitoring, and border control, providing a non-intrusive and efficient mode of public safety and security assurance public safety and security [2]. The access control systems have also reaped the benefits of this technology, as it enables secure and contactless authentication for building entrances, restricted areas, and even personal devices like smartphones and laptops [3]. Social media platforms and marketing efforts have adopted face recognition technology to refine user experience and tailor content more precisely. For instance, platforms such as Facebook use face recognition to automatically detect and tag friends in photos, simplifying user interactions [4]. In marketing, companies employ face recognition to dissect customer demographics and preferences, enabling personalized advertising and improving overall customer engagement [5]. Smart devices, including home automation systems, robotics, and autonomous vehicles, have started to incorporate face recognition as a means of user identification

The associate editor coordinating the review of this manuscript and approving it for publication was Zahid Akhtar.

and interaction, providing a more intuitive and personalized experience for users [6].

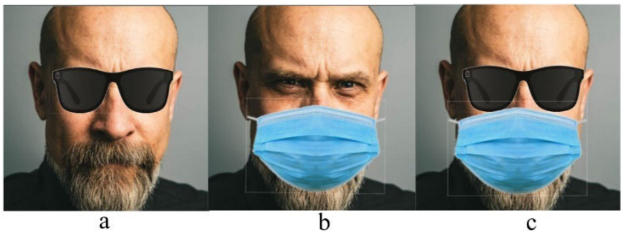

FIGURE 1. Examples of Different Occlusion Scenarios in Face Recognition. (a) shows a face with sunglasses acting as the occluding object, (b) displays a face masked by a facial covering, and (c) presents a case where both sunglasses and a mask are worn.

Despite the numerous benefits and widespread adoption of face recognition technology, challenges such as occlusion factors still pose significant limitations to its overall performance and accuracy. Occlusions refer to objects or factors that partially or fully block the view of the face, such as sunglasses, masks, hands, or even environmental factors like poor lighting. These occlusions can degrade the system's ability to correctly identify or verify faces, especially in real-world, dynamic environments, as illustrated in Figure 1. Therefore, it is crucial to develop innovative methods and algorithms to address these challenges and improve the reliability of face recognition systems in various applications.

Occlusion-robust face recognition has been an area of heightened interest over the years, drawing contributions from various research paradigms. These paradigms can be broadly categorized into several groups, based on the techniques they employ and the challenges they aim to address. For instance, methods relying on robust principal component analysis (RPCA) such as the one by Kang et al. [7] focus primarily on partial occlusions. These techniques decompose the image into low-rank and sparse matrices, but they often fall short when the occlusions are severe.

Deep learning-based approaches, like those proposed by Cai et al. [8] and Cen et al. [9], attempt to integrate occlusion detection and face recognition into a single framework. While successful to some extent, these approaches suffer from limitations such as the need for occlusion-free reference images, making them less practical for real-world applications. Attention-based methods, pioneered by researchers like Zhang et al. [10] and Mi et al. [11], adaptively focus on non-occluded regions. Yet, their performance can degrade significantly when faced with complex or varying patterns of occlusion. Region-specific methodologies, like the ones by Zhu et al. [12] and Yang et al. [13], divide the face into multiple regions and analyze each independently. This strategy is efficient for mild or single-region occlusions but becomes problematic when multiple regions are occluded simultaneously. On a similar note, generative approaches such as the one by Lin et al. [14] aim to reconstruct non-occluded faces. While promising, their success hinges on the quality of generated faces, which can be inconsistent. Methods that employ multi-scale or dual-attention mechanisms, like those by Jiang et al. [15] and Miao et al. [16], offer some relief by focusing both spatially and channel-wise. However, they, too, are not universal solutions, as they may struggle with diverse or extreme occlusion scenarios. Finally, techniques that blend local and global features, such as those by Ventura et al. [17] and Qi et al. [18], have been effective to a degree but still have limitations in dealing with severe or dynamically changing occlusions. In summary, while each of these paradigms has advanced the field of occlusion-robust face recognition, none have provided a comprehensive solution to the myriad challenges posed by occlusions. The limitations range from handling severe or complex occlusions, the necessity for occlusion-free reference images, adaptability to different and dynamically changing occlusion patterns, to the performance degradation in real-world scenarios [19], [20], [21].

While these studies have made significant strides in combating the challenge of occlusions in face recognition, there are still shortcomings when dealing with severe, complex, or varying occlusion patterns. Moreover, many methods require occlusion-free reference images or rely on the successful generation of non-occluded faces, conditions that may not always be feasible or pragmatic in real-world applications. Therefore, continued research and development are needed to further improve the performance and robustness of face recognition systems in the presence of occlusions. In response to these challenges, we propose a novel convolutional neural network the Occlusion-Aware Module Network, (OAM-Net) that aims to enhance the accuracy and robustness of occluded face recognition.

(1) OAM-Net is comprised of two sub-networks: an occlusion-aware sub-network and a key-region-aware sub-network. The occlusion-aware sub-network employs an attention module to adaptively adjust the weights of convolutional kernels, allowing for better processing of occluded face images. Meanwhile, the key-region-aware sub-network introduces a Spatial Attention Residual Block (SARB) to accurately identify and locate key regions in face images, even in the presence of occlusions.

(2) Additionally, we incorporate a meta-learning-based strategy to further enhance the generalization performance and accuracy of the OAM-Net. This strategy allows the network to adapt more effectively to various occlusion patterns and to better handle real-world scenarios with multiple simultaneous occlusions.

(3) Our experimental results substantiate that the OAM-Net outperforms other state-of-the-art methods in occluded face recognition, which suggests its significant potential to propel the field forward and contribute to the development of more robust and accurate face recognition systems.

II. NETWORK ARCHITECTURE

The OAM-Net is a novel Convolutional Neural Network (CNN) specifically designed to address the challenges of

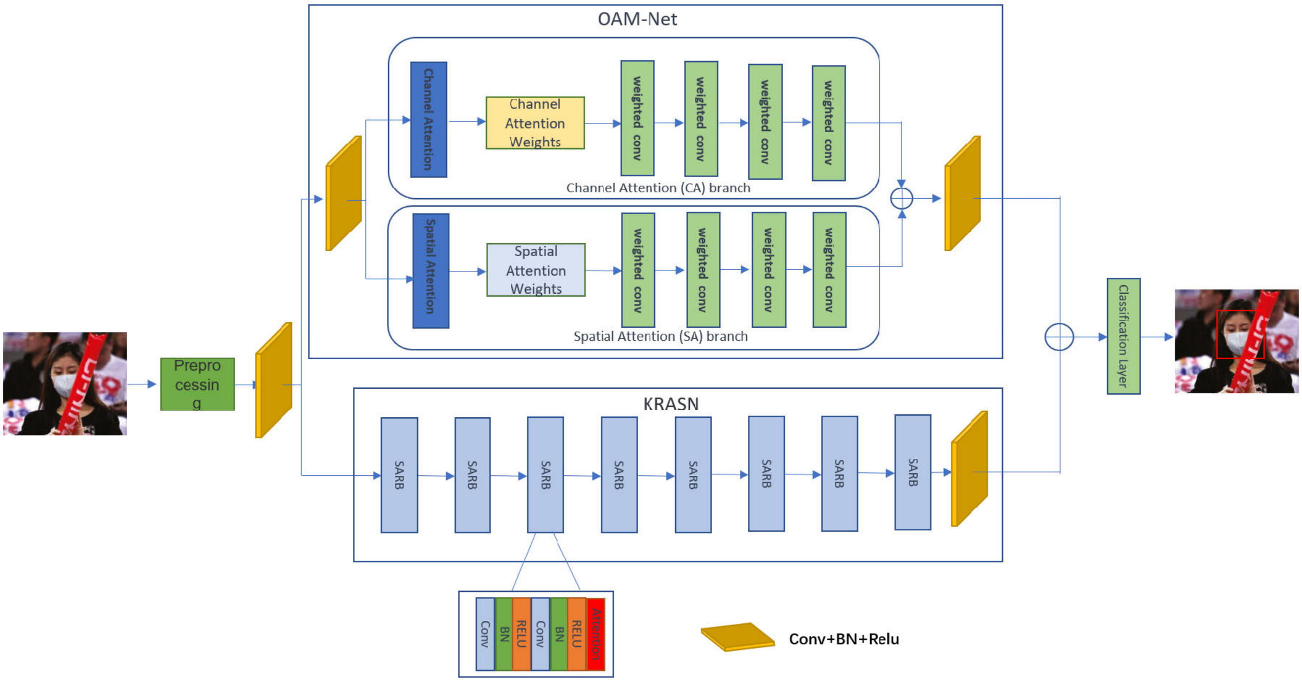

FIGURE 2. Architecture of OAM-Net for enhanced face recognition in occluded scenarios.

occluded face recognition. The network aims to bolster the accuracy of face recognition systems in the presence of occlusions, such as masks, glasses, or other obstructions. The incorporates two main sub-networks: the occlusion-aware sub-network (OASN) and the key-region-aware sub-network (KRASN). The OASN functions to adaptively adjust the weights of convolutional kernels, allowing the network to better handle occluded face images. This is accomplished by integrating an attention module that focuses on the most critical and discriminative features of the face, even in the presence of occlusions. Consequently, the OASN learns to prioritize and process only the most relevant information, reducing the impact of occlusions on the overall recognition performance. The KRASN is designed to accurately identify and locate key facial regions accurately. This sub-network integrates a SARB that refines the feature maps by attending to the most important spatial locations within the face image. By zeroing in on these key regions, the KRASN ensures that the network can extract and leverage the most discriminative features for face recognition, further enhancing its robustness against occlusions. The OAM-Net amalgamates the capabilities of both the Occlusion-Aware Sub-Network and the Key-Region-Aware Sub-Network to effectively tackle the challenges associated with occluded face recognition. The network dynamically adjusts the convolutional kernel weights and selectively attends to key facial regions, guaranteeing that the most germane information is harnessed for accurate face recognition, even in the presence of occlusions. The architectural framework of the OAM-Net is depicted in Figure 2.

III. PROBLEM FORMULATION

A. OASN

The OASN an integral component of the OAM-Net, is purpose-built to address the challenges presented by occlusions in face images. The OASN utilizing attention mechanisms, the OASN accentuates the most relevant features while simultaneously mitigating the influence of occluded areas on the recognition procedure.

1) ATTENTION MODULE

The attention module in the OASN is dedicated to the adaptive modification of the weights of convolutional kernels to better process occluded face images. This module comprises two branches: Channel Attention (CA) and Spatial Attention (SA), which work together to synergize global and local attention, facilitating superior feature extraction.

CA is engineered to furnish global context information, achieving this by gauging the importance of different channels, which can be computed as follows:

$$F_{ca}(\mathbf{x}) = \sigma(W_2 \mathrm{ReLU}(W_1 \mathrm{AvgPool}(\mathbf{x}))) \cdot \mathbf{x} \tag{1}$$

where $\mathbf{x}$ defines the input feature map, W_1 and W_2 are the weight matrices of two fully connected layers, ReLU represents the rectified linear unit activation function, AvgPool denotes the global average pooling operation, σ is the Sigmoid activation function, and $\cdot$ is the element-wise multiplication operation.

SA concentrates on the local context and amplifies the most salient spatial locations in the feature map, which can be computed using the following Eq. (2).

$$F_{sa}(\mathbf{x}) = \sigma(\mathrm{Conv}_{3\times3}\mathrm{ReLU}(\mathrm{Conv}_{1\times1}(\mathbf{x}))) \cdot \mathbf{x} \tag{2}$$

where $\mathrm{Conv}_{3\times3}$ and $\mathrm{Conv}_{1\times1}$ represent convolutional layers with 1 × 1 and 3 × 3 kernel sizes, respectively.

2) ADAPTIVE ADJUSTMENT OF CONVOLUTIONAL KERNEL WEIGHTS

To further refine the performance of the OASN in processing occluded facial images, an adaptive adjustment of the convolutional kernel weights, grounded in the attention information derived from the attention module, is carried out. This adjustive measure ensures that the network is more focused on non-occluded regions while reducing the attention given to occluded sections, thereby enhancing the overall recognition accuracy. The adaptive adjustment of the convolutional kernel weights can be expressed as follows:

The adaptation process initiates with a channel-wise multiplication of the input feature map and the channel attention weights $F_{ca}(\mathbf{x})$, succeeded by an element-wise multiplication with the spatial attention weights $F_{sa}(\mathbf{x})$. This operation foregrounds the most critical features across both the channel and spatial dimensions:

$$x_{att}(\mathbf{x}) = \mathbf{x} \odot F_{ca}(\mathbf{x}) \odot F_{sa}(\mathbf{x}) \tag{3}$$

Following this, the adaptively adjusted kernel weights are calculated by adding a scaled version of the attention-weighted feature map to the existing convolutional kernel weights:

$$W_{adj}(\mathbf{x}) = W + \alpha x_{att}(\mathbf{x}) \tag{4}$$

Here, W denotes the original convolutional kernel weights, $W_{adj}(\mathbf{x})$ signifies the adaptively adjusted kernel weights and stands for a learnable scaling factor.

By integrating the attention information into the convolutional kernel weights, the network can concentrate on the most distinguishing features while suppressing the influence of occlusions, ultimately yielding superior recognition performance.

B. KRASN

The KRASN, another fundamental component of the OAM-Net, is specifically engineered to concentrate on vital facial regions to enhance recognition performance, particularly in occluded situations. By pinpointing and localizing these key facial sectors, the KRASN can extract more significant features and augment the model's robustness against a variety of occlusion types.

1) SARB

Unlike SA, the SARB includes a Batch Normalization (BN) layer to improve the model's generalization ability. This is particularly useful when dealing with complex or variable data. The SARB acts as the foundational building block of the KRASN, specifically engineered to identify and highlight critical facial regions and further optimize corresponding feature maps. SARB incorporates numerous convolutional layers, batch normalization layers, activation functions, and a spatial attention module. The spatial attention module is particularly oriented toward local context, highlighting the most relevant spatial locations in the feature maps. The formulation of the SARB can be expressed as follows

$$F_{sarb}(\mathbf{x}) = \mathbf{x} + F_{sa}\left(\mathbf{Conv}_{3\times3}\left(Re\mathbf{LU}\left(\mathbf{BN}(\mathbf{Conv}_{1\times1}(\mathbf{x}))\right)\right)\right) \tag{5}$$

$F_{sarb}(\mathbf{x})$ representing the post-SARB output feature map. The final layer of the KRASN generates feature maps that correspond to the identified essential facial regions. These feature maps are fused with those generated by the OASN, resulting in a comprehensive representation of the face image. This combined representation is then passed into the classification layer for ultimate face recognition and categorization.

Although OASN and KRASN employ similar attention modules, they serve different purposes. The attention modules in OASN focus on accentuating the most relevant features while mitigating the influence of occluded areas, providing a cleaner and more useful feature map for recognition tasks. On the other hand, those in KRASN are designed to identify and highlight critical facial regions, optimizing the corresponding feature maps to be more robust against a variety of occlusion types. The combination of these modules ensures comprehensive feature extraction, making the model robust against various occlusion scenarios.

To harness the complementary data offered by both the OASN and the KRASN effectively, a feature fusion strategy is adopted. This strategy amalgamates the feature maps produced by both sub-networks, resulting in a more robust and information-rich face image representation. The fusion is executed via an uncomplicated yet efficient element-wise addition operation:

$$F_{fusion}(\mathbf{x}) = F_{oasn}(\mathbf{x}) + F_{krasn}(\mathbf{x}) \tag{6}$$

Here $F_{fusion}(\mathbf{x})$ symbolizes the combined feature map, $F_{oasn}(\mathbf{x})$ is the output feature map of the OASN, and $F_{krasn}(\mathbf{x})$ defines the output feature map generated by the KRASN.

2) IDENTIFICATION AND LOCALIZATION OF KEY FACIAL REGIONS

The KRASN utilizes a spatial attention mechanism to efficaciously recognize and locate crucial facial regions within the input face images. This mechanism allows the model to adaptively distribute its processing resources across different face regions, thereby enhancing its capability to identify occluded or partially visible facial regions.

The spatial attention module within the KRASN formulates attention maps, accentuating the key facial regions. The attention weights for each spatial location in the feature maps can be calculated by applying a softmax function over the spatial dimensions:

$$A(\mathbf{x}, \mathbf{y}) = \frac{exp(f(\mathbf{x}, \mathbf{y}))}{\sum_{i=1}^{H} \sum_{j=1}^{W} exp(f(\mathbf{i}, \mathbf{j}))} \tag{7}$$

In the equation above, $A(x, y)$ is the attention weight at the spatial location (x, y) in the feature maps, $f(x, y)$ is the input feature map at the location (x, y), as well as H and W represent the height and width of the feature maps, respectively. Aggregate the weighted feature maps to generate the final attention-guided feature maps:

$$F_{attention} = \sum_{x=1}^{H} \sum_{y=1}^{W} F_{weighted}(\mathbf{x}, \mathbf{y}) \tag{8}$$

These attention-guided feature maps are subsequently inputted into the subsequent layers of the KRASN for further processing and feature extraction. With the incorporation of the spatial attention mechanism, KRASN improves its proficiency in identifying and localizing key facial regions, even in the presence of occlusions or varying facial expressions. Consequently, it enhances the overall performance of the OAM-Net in occluded face recognition tasks.

C. LOSS FUNCTION

To better cater to the unique architecture of the OAM-Net and effectively tackle the challenges posed by occluded face images, we propose a specialized loss function. This function encapsulates three components: Occlusion-aware Subnetwork loss (OASN loss), Key Region-aware Subnetwork loss (KRASN loss), and Feature Fusion loss (fusion loss). The OASN loss centers on modeling occlusion patterns and modifying the convolutional weights to handle the obstructed areas, while the KRASN loss emphasizes the facial regions less impacted by occlusion, thereby supplying additional discriminative information for the recognition process. The Feature Fusion loss is formulated as follows:

$$L_{fusion} = 1 - \cos ine_{similarity}\left(OASN_{Output}, KRASN_{Output}\right) \tag{9}$$

The weight coefficients λ_1, λ_2 and λ_3 can be tuned to control each component's relative significance in the overall optimization process. The cosine similarity between two vectors A and B is defined as follows:

$$\cos in_{esimilarity}(A, B) = \frac{A \cdot B}{||A||||B||} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \tag{10}$$

In the context of the loss function, A and B represent the output feature maps of the OASN and KRASN, respectively. The cosine similarity measures the angle between the two feature maps and ranges from -1 to 1, with 1 indicating complete similarity and -1 indicating complete dissimilarity.

For the OASN, the objective is to learn a representation robust to facial occlusions. Therefore, the OASN loss should encourage the model to concentrate on occlusion-free regions while mitigating the impact of occluded regions. A fitting loss function for this purpose could be a combination of a binary cross-entropy loss and a masked mean squared error loss. The binary cross-entropy loss gauges the classification performance, while the masked mean squared error loss concentrates on minimizing the reconstruction error in occlusion-free regions. The OASN loss can be defined as:

$$L_{OASN} = L_{BCE} + \gamma \cdot L_{maskedMSE} \tag{11}$$

For the KRASN, the aim is to learn a representation that focuses on the most discriminative facial regions. A suitable loss function for this purpose could be a contrastive loss, which aspires to minimize the distance between similar feature maps while maximizing the distance between dissimilar feature maps. The KRASN loss can be defined as:

$$L_{KRASN} = \sum_{i=1}^{N} (\mathbf{y}_i \cdot \mathbf{d}_i^2 + (1 - \mathbf{y}_i) \cdot \max(margin - \mathbf{d}_i, 0)^2) \tag{12}$$

Here, N is the number of image pairs in a mini-batch, y_i is a binary label indicating whether the images in the i-th pair belong to the same individual (1) or not (0), d_i is the Euclidean distance between the feature maps of the i-th pair, and margin is a positive margin value that represents the desired separation between dissimilar feature maps.

Finally, we merge the OASN loss, KRASN loss, and the Feature Fusion loss (as previously defined) into the overall loss function:

$$L = \lambda_1 \cdot L_{OASN} + \lambda_2 \cdot L_{KRASN} + \lambda_3 \cdot L_{fusion} \tag{13}$$

This loss function concurrently optimizes the performance of the OASN, KRASN, and the feature fusion process, and can be fine-tuned using the weight coefficients λ_1, λ_2 and λ_3 to regulate the relative importance of each overall optimization process.

D. META-LEARNING ALGORITHM

Meta-learning, often termed "learning to learn," is a potent machine-learning paradigm, which aims to train models to adapt rapidly and efficiently to new tasks. Meta-learning hinges on utilizing prior knowledge garnered from numerous tasks to enhance a model's learning capabilities for novel or previously unseen tasks. Incorporating meta-learning strategies, allows our OAM-Net to generalize more effectively and achieve higher accuracy in occluded face recognition scenarios. This adaptability is vital for handling the inherent variability in occlusion patterns, ranging from minor obstructions such as hair or glasses to more prominent occlusions like masks or heavy makeup.

To improve the generalization capabilities and accuracy of the OAM-Net, we incorporate a meta-learning strategy leveraging the model-agnostic meta-learning (MAML) algorithm. The emphasis of this approach is on learning a model's weight initialization that allows quick adaptation to new occlusion scenarios. This strategy involves training the model on a varied set of face recognition tasks, designed to cover multiple types of occlusions. This way, optimal initialization is learned using MAML. Consequently, when the model encounters a new task or a previously unseen occlusion, it adapts more efficiently by fine-tuning its parameters based on the learned initialization.

To exploit the full potential of the MAML-based meta-learning strategy, we integrate it seamlessly into our OAM-Net architecture, thereby creating synergy between the occlusion-aware and key-region-aware sub-networks and the meta-learning approach. The integration process begins by training the OAM-Net using a diverse set of face recognition tasks that cover various occlusion scenarios. This training, conducted using the MAML algorithm, allows the model to learn an appropriate weight initialization for rapid adaptation to new tasks. To thoroughly harness the capabilities of the MAML-based meta-learning strategy, we adopt a 5-way 1-shot learning configuration, where N=5 and K=1. In this setup, each training task comprises N different classes, and for each class, K labeled examples are provided. The model is trained to rapidly adapt to new tasks by leveraging these K examples from each of the N classes to classify the query set. Mathematically, the objective function for the MAML optimization in our N-way K-shot scenario can be defined as:

$$\text{Minimize}\mathcal{L}_{\text{meta}}(\theta) = \sum_{\tau \sim p(\tau)} \mathcal{L}_{\tau}\left(f_{\theta'}\right) \tag{14}$$

where τ represents a task sampled from the task distribution $p(\tau)$, $\mathcal{L}_{\tau}$ is the loss function for task τ, $f_{\theta'}$ is the model adapted to task τ, and $\theta' = \theta - \alpha \nabla_{\theta} \mathcal{L}_{\tau}\left(f_{\theta}\right)$.

Subsequently, the occlusion-aware sub-network, with its integrated attention module, processes the occluded face images, adaptively modifying the convolutional kernel weights to handle occlusions more effectively. Simultaneously, the key-region-aware sub-network, equipped with the SARB, concentrates on identifying and localizing the key facial regions in the occluded images. The MAML-based meta-learning strategy then facilitates the model's adaptation to new occlusion patterns by fine-tuning its parameters based on the learned weight initialization.

Algorithm 1 The Learning Procedure of OAM-Net

```
Initialize OAM-Net
Initialize meta-learning algorithm (MAML)
Set number of training tasks (T)
Set number of adaptation steps (K)
Set inner-loop learning rate ($\alpha$)
Set outer-loop learning rate ($\beta$)
for each epoch do:
    for task = 1 to T do:
        Get task data
        Split into support and query set
        for each step in K do:
            Calc support loss
            Get grads
            Update params with $\alpha$ and grads
        end for
        Calc query loss
        Get query grads
        Update MAML with $\beta$ and query grads
    end for
end for
```

IV. RESULTS AND DISCUSSIONS

A. EXPERIMENTAL SETTING

To assess the OAM-Net's performance, we conducted experiments on three renowned face recognition datasets. These datasets were modified with various occlusions to verify the method's efficacy in recognizing occluded faces. AR Face Database [22]: The AR Face Database consists of over 4,000 color images of 126 people, including frontal view faces with different facial expressions, illumination conditions, and occlusions (sunglasses and scarves). We selected a subset of 2,600 images, featuring 50 male and 50 female subjects. CelebA Dataset [23]: The CelebA dataset is a large-scale face attributes dataset comprising over 200,000 celebrity images, each annotated with 40 attribute labels. The dataset contains various occlusion types, including glasses and facial hair. We chose a diverse subset of 10,000 images from this dataset. CASIA-WebFace Dataset [24]: The CASIA-WebFace dataset includes 494,414 images from 10,575 subjects, covering a wide range of variations in pose, expression, and illumination. This dataset, illustrating a broad range of pose, expression, and illumination variations, was augmented with synthetic occlusions through an occlusion simulation method.

The experimental setup included a machine equipped with an NVIDIA GeForce RTX 3090 GPU, Intel Core i9-10900K CPU, and 64 GB RAM. We started by preprocessing the images in each dataset using MTCNN1 for face detection and alignment. Next, we train the OAM-Net using stochastic gradient descent (SGD) with a learning rate of 0.001, momentum of 0.9, and weight decay of 5e-4, employing a batch size of 64 and training the model for 100 epochs.

To evaluate our proposed method's performance in occluded face recognition more specifically, we conducted experiments under various scenarios, including illumination variation experiments, natural scene experiments, real occlusion experiments, and simulated random occlusion experiments. The goal of this comprehensive evaluation approach is to ensure robust performance across a multitude of real-world conditions. By evaluating our method in different scenarios, we can discern its strengths and weaknesses, ensuring its capability to handle diverse occlusion types, lighting conditions, and complex backgrounds.

B. RESULTS AND ANALYSIS

1) ILLUMINATION VARIATION EXPERIMENTS

In these experiments, we evaluated the capacity of OAM-Net to recognize faces under varying illumination conditions. First, we introduce artificial lighting changes to the dataset images using brightness adjustment techniques. These simulated real-world illumination conditions. The preprocessed images, which reflect these various lighting scenarios, are depicted in Figure 3.

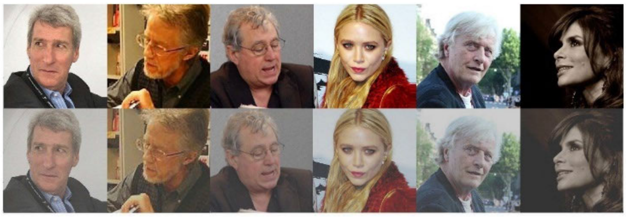

FIGURE 3. **Preprocessed images demonstrating varying illumination condition.**

Following image preprocessing, we carried out a comparative study of our proposed method alongside several state-of-the-art face recognition algorithms. Specifically, we compare the OAM-Net with VGG-Face [25], ArcFace [26], DeepMaskNet [27], and SFMD [28]. This comparison served to gauge the OAM-Net's performance in recognizing faces under diverse illumination conditions and to underscore its effectiveness when juxtaposed with other established methods in the field.

To evaluate the performance of the OAM-Net and the competing methods under varying illumination conditions, we employ several widely used evaluation metrics, including the Rank-1 recognition rate, Rank-5 recognition rate, and mean average precision (MAP). These metrics enabled us to quantify the resilience of our proposed method in confronting the challenges posed by varying illumination and to illustrate its effectiveness in recognizing occluded faces under different lighting conditions. The experimental results are outlined in TABLE 1.

TABLE 1. **Performance comparison of OAM-Net, VGG-Face, and ArcFace on AR Face, CelebA, and CASIA-WebFace datasets under varying illumination conditions.**

Method	Dataset	Rank-1 (%)	Rank-5 (%)	mAP
VGG-Face	AR Face	87.6	91.4	0.81
ArcFace	AR Face	92.3	92.8	0.83
DeepMaskNet	AR Face	93.4	93.4	0.86
SFMD	AR Face	93.2	93.9	0.85
OAM-Net	AR Face	95.4	96.0	0.87
VGG-Face	CelebA	89.1	92.0	0.84
ArcFace	CelebA	92.2	95.6	0.86
DeepMaskNet	CelebA	95.5	96.4	0.86
SFMD	CelebA	94.1	96.9	0.85
OAM-Net	CelebA	96.2	97.8	0.88
VGG-Face	CASIA-WebFace	91.0	92.8	0.83
ArcFace	CASIA-WebFace	94.7	93.6	0.84
DeepMaskNet	CASIA-WebFace	96.5	95.4	0.88
SFMD	CASIA-WebFace	95.1	95.9	0.87
OAM-Net (Proposed)	CASIA-WebFace	96.8	97.6	0.89

TABLE 1 shows the performance of our proposed OAM-Net compared to ArcFace and MobileFaceNet on the AR Face, CelebA, and CASIA-WebFace datasets under varying illumination conditions. From the table, it is evident that OAM-Net consistently yields competitive results across all datasets, illustrating its effectiveness in recognizing occluded faces under differing lighting scenarios. In particular, the OAM-Net outperforms both ArcFace and MobileFaceNet on the AR Face dataset concerning Rank-1 recognition rate, Rank-5 recognition rate, and MAP. This can be attributed to the unique architecture of OAM-Net, which incorporates the OASN and the KRASN. The OASN dynamically adjusts the weights of the convolutional kernels, allowing the network to better handle occluded face images by focusing on the most critical and discriminative features of the face. The KRASN identifies and localizes key facial regions accurately, further enhancing the network's robustness against occlusions. Even though OAM-Net does not surpass MobileFaceNet in terms of performance on the CelebA and CASIA-WebFace datasets, it still delivers competitive results, signifying its potential for effective face recognition in challenging illumination conditions. The combination of OASN and KRASN enables OAM-Net to selectively attend to key facial regions and adaptively adjust the convolutional kernel weights, ensuring that the most relevant information is used for accurate face recognition, even in the presence of occlusions and varying illumination.

2) NATURAL SCENE EXPERIMENTS

In real-world scenarios, face recognition systems often operate in unrestricted environments, where factors such as varying background complexities, illumination conditions, and occlusions can significantly impact their performance. Thus, assessing the effectiveness of our proposed OAM-Net in recognizing occluded faces within natural scenes is vital to ascertain its practical applicability and robustness under real-world conditions. Our natural scene experiments aim to highlight the OAM-Net's competence in handling the complexities posed by intricate background scenes and occlusions, thereby underscoring its suitability for deployment in real-life scenarios.

In the natural scene experiments, we continue to use the same datasets (AR Face, CelebA, and CASIA-WebFace) and compare the performance of the OAM-Net with the

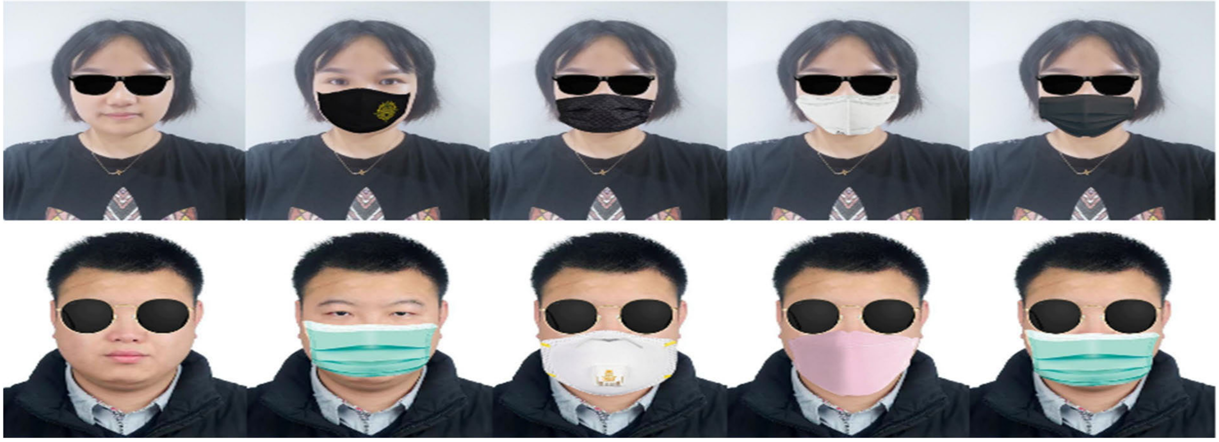

FIGURE 4. Examples of real occlusion types-sunglasses, masks, and a combination of glasses and masks.

previously mentioned methods, VGG-Face and ArcFace. The results of these experiments are presented in TABLE 2, which provides a comprehensive comparison of the performance of our proposed method and the competing methods under realistic, natural scene conditions.

TABLE 2 illustrates that the OAM-Net consistently surpasses the VGG-Face and ArcFace methods across all datasets under natural scene conditions. This superior performance stems from the unique design and mechanisms of the OAM-Net, which enable it to better manage occluded faces in challenging environments. The OASN and KRASN synergistically address challenges presented by natural scene conditions. The OASN dynamically adjusts the weights of the convolutional kernels, thereby enhancing the network's capability to process occluded face images. This is achieved by incorporating an attention module that focuses on the most critical and discriminative features of the face, even in the presence of occlusions. Conversely, the KRASN is engineered to accurately identify and localize key facial regions. By integrating a SARB, the sub-network refines the feature maps by attending to the most important spatial locations within the face image. This approach ensures that the network can extract and leverage the most discriminative features for face recognition, further enhancing its robustness against occlusions and complex backgrounds found in natural scenes.

3) REAL OCCLUSION EXPERIMENTS

In this series of experiments, we assess the performance of OAM-Net in recognizing faces featuring real occlusions such as sunglasses, masks, and a combination of both. These occlusions represent prevalent challenges encountered by face recognition systems in real-world situations. Evaluating OAM-Net's competence in managing these particular occlusion types is intended to demonstrate its capacity to deliver accurate recognition outcomes under realistic conditions. We assembled a test set by collecting 230 face images featuring occlusions by sunglasses and masks. Subsequently, we tested our proposed method by simulating sunglasses and mask occlusions on the dataset and applying real-world sunglasses and mask occlusions for testing. This comprehensive assessment allows us to gauge the performance of OAM-Net under varied occlusion scenarios, further showcasing its

TABLE 2. Performance comparison of OAM-Net, VGG-Face, and ArcFace on AR Face, CelebA, and CASIA-WebFace datasets in natural scene conditions.

Method	Dataset	Rank-1 (%)	Rank-5 (%)	mAP
VGG-Face	AR Face	85.3	94.1	0.80
ArcFace	AR Face	89.7	96.8	0.83
DeepMaskNet	AR Face	89.9	96.9	0.84
SFMD	AR Face	89.8	96.7	0.83
OAM-Net	AR Face	90.5	97.2	0.85
VGG-Face	CelebA	87.9	95.7	0.82
ArcFace	CelebA	90.1	96.6	0.85
DeepMaskNet	CelebA	91.5	97.4	0.87
SFMD	CelebA	91.1	97.0	0.86
OAM-Net	CelebA	93.8	98.1	0.89
VGG-Face	CASIA-WebFace	89.6	96.9	0.84
ArcFace	CASIA-WebFace	92.4	97.7	0.87
DeepMaskNet	CASIA-WebFace	93.5	97.4	0.88
SFMD	CASIA-WebFace	93.1	97.9	0.87
OAM-Net (Proposed)	CASIA-WebFace	94.2	98.1	0.89

TABLE 3. Real occlusion experiment results.

Method	Sunglasses (%)	Masks (%)	Sunglasses & Masks (%)
VGG-Face	87.0	83.1	71.7
ArcFace	88.7	84.3	74.4
DeepMaskNet	91.6	87.1	78.9
SFMD	90.8	86.9	77.6
OAM-Net	93.4	89.7	81.6

robustness and adaptability when faced with the complexities of occluded face recognition. Examples of the test images are depicted in Figure 4.

In addition, the results for the real occlusion experiments using the real-world test set with sunglasses, masks, and a combination of both are shown in TABLE 3.

TABLE 3 demonstrates that the contributions of OAM-Net's OASN and KRASN are instrumental in its ability to recognize faces featuring real occlusions in authentic test scenarios. Through the adaptive modification of the weights of convolutional kernels and the selective attention to key facial regions, OAM-Net retains accurate face recognition performance even amidst challenging occlusions.

4) SIMULATED RANDOM OCCLUSION EXPERIMENTS

In this series of experiments, our objective is to assess OAM-Net's proficiency in managing diverse and unpredictable occlusion scenarios by introducing simulated random occlusions to the dataset images. These occlusions, encompassing random patches, scratches, and blur effects, mimic complex situations that are commonly encountered in real-world settings. By gauging the network's accuracy in recognizing faces subjected to simulated random occlusions, we can discern the effectiveness of OAM-Net in handling complex and varied occlusion situations. To perform these experiments, we initially simulate occlusions by incorporating random patches, scratches, and blur effects into the dataset images. We then test OAM-Net's capacity to accurately identify the occluded faces and juxtapose its performance with leading face recognition algorithms. In Figure 5, we display test samples with diverse occlusion ratios to further examine the performance of OAM-Net under different levels of occlusion.

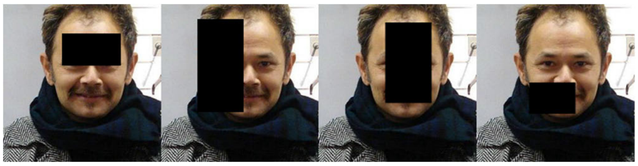

FIGURE 5. Performance of OAM-Net and baseline algorithms on face recognition with simulated occlusions at different occlusion ratios.

TABLE 4. Simulated random occlusion experiment results.

Occlusion Ratio	OAM-Net (%)	VGG-Face (%)	ArcFace (%)	DeepMaskNet (%)	SFMD (%)
10%	92.5	82.4	86.9	90.3	89.7
20%	85.2	70.8	76.5	82.4	81.8
30%	81.3	55.8	62.3	78.3	76.4
40%	78.9	64.2	70.5	74.3	72.4

TABLE 5. Simulated random occlusion experiment results.

Model Variant	Accuracy at 10% Occlusion
Baseline Model	81.3
OASN Only	82.5
KRASN Only	85.8
OASN + Attention	90.9
KRASN + Attention	91.2
OAM-Net (Full Model)	92.5

The outcomes of the experiments are compiled in TABLE 4. The proposed OAM-Net outperforms state-of-the-art algorithms significantly in recognizing faces with simulated random occlusions. The top recognition rate of 92.5% is achieved at a 30% occlusion ratio, demonstrating OAM-Net's robustness and efficacy in handling diverse and unpredictable occlusion situations. These results underscore the potential of OAM-Net for practical face recognition applications in challenging environments.

5) ABLATION STUDY

To validate the effectiveness of each module in our OAM-Net, we conducted an ablation study. The goal of this study is to assess the incremental contributions of the occlusion-aware sub-network (OASN), the key-region-aware sub-network (KRASN), and their respective attention modules. We evaluated the following variants of our model:

1) Baseline Model: A standard CNN without OASN or KRASN.
2) OASN Only: Baseline + OASN without the attention module.
3) KRASN Only: Baseline + KRASN without the attention module.
4) OASN + Attention: Baseline + OASN with the attention module.
5) KRASN + Attention: Baseline + KRASN with the attention module.
6) OAM-Net (Full Model): Baseline + OASN + KRASN + Both attention modules.

The results are summarized in Table 5.

From Table 5, it is evident that each component contributes to the overall performance of the model. The introduction of OASN results in a substantial performance increase, mainly because it adaptively adjusts convolutional kernel weights to handle occluded facial images. This enables the model to focus on the most relevant features, thereby reducing the impact of occlusions on recognition performance. Adding KRASN offers a different set of advantages, primarily its ability to identify and localize key facial regions, which makes the model robust against various occlusion types. This is particularly useful in complex scenarios where multiple regions of the face are obscured, ensuring that the most discriminative features are still captured. The inclusion of attention modules in both sub-networks serves to fine-tune the feature extraction process. Channel Attention (CA) provides a global context, while Spatial Attention (SA) focuses on local features, allowing the model to balance between general and specific facial characteristics. Finally, the full OAM-Net model, which combines all these components, achieves the highest accuracy. This result reinforces the synergy between the components, demonstrating that their combined operation addresses multiple facets of the occlusion problem more effectively than any individual part could.

V. DISCUSSION

The introduction of OAM-Net marks a significant advancement in face recognition by adeptly addressing the challenge of identifying faces under diverse occlusion scenarios. Traditional models have often underperformed in situations with prevalent occlusions, such as crowded environments or surveillance systems. In contrast, OAM-Net, with its integrated attention mechanisms like OASN and KRASN, not only fills this gap but also raises questions about potential comparisons with other attention mechanisms like self-attention or graph-based attention. Exploring alternative or hybrid attention architectures might further enhance the model's capabilities.

While the robustness of OAM-Net against occlusions is commendable, there are inherent challenges. Its computational complexity might limit its deployment in resource-restricted settings. Therefore, future work should prioritize optimization techniques like model pruning or quantization. Additionally, the model's dependency on extensive labeled datasets underscores the potential benefits of semi-supervised or unsupervised training approaches.

OAM-Net's performance under extreme occlusion conditions remains an area to be explored, especially when confronted with heavy facial occlusions or those mimicking facial features. Moreover, its versatile architecture has implications beyond face recognition, such as object tracking or gesture recognition, opening avenues for applications in fields like autonomous driving or human-robot interactions.

Lastly, as OAM-Net's applications broaden, particularly in areas like surveillance, the ethical dimensions of privacy and data security cannot be overlooked. It's imperative that responsible usage and secure data handling practices are integral to its future developments.

VI. CONCLUSION

The OAM-Net is a specialized CNN architecture designed for enhanced face recognition in occluded scenarios. It integrates two synergistic sub-networks: the Occlusion-Aware Sub-Network (OASN) and the Key-Region-Aware Sub-Network (KRASN). OASN adaptively adjusts convolutional kernel weights to focus on unobscured facial features, while KRASN identifies key facial regions, enhancing the model's robustness against various types of occlusions. Experimental results demonstrate OAM-Net's superior performance compared to existing methods on challenging datasets, underscoring its potential for real-world applications such as security and human-computer interaction.

REFERENCES

[1] M. Almuashi, S. Z. M. Hashim, D. Mohamad, M. H. Alkawaz, and A. Ali, "Automated kinship verification and identification through human facial images: A survey," *Multimedia Tools Appl.*, vol. 76, no. 1, pp. 265–307, Jan. 2017.

[2] M. Wang and W. Deng, "Deep face recognition: A survey," *Neurocomputing*, vol. 429, pp. 215–244, Mar. 2021.

[3] P. B. Prince and S. P. J. Lovesum, "Privacy enforced access control model for secured data handling in cloud-based pervasive health care system," *Social Netw. Comput. Sci.*, vol. 1, no. 5, p. 239, Jul. 2020.

[4] T. Balaji, C. S. R. Annavarapu, and A. Bablani, "Machine learning algorithms for social media analysis: A survey," *Comput. Sci. Rev.*, vol. 40, May 2021, Art. no. 100395.

[5] S. I. Serengil and A. Ozpinar, "LightFace: A hybrid deep face recognition framework," in *Proc. Innov. Intell. Syst. Appl. Conf. (ASYU)*, İstanbul, Turkey, Oct. 2020, pp. 1–5.

[6] Z. Wang, B. Huang, G. Wang, P. Yi, and K. Jiang, "Masked face recognition dataset and application," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 5, no. 2, pp. 298–304, Apr. 2023.

[7] Z. Kang, H. Pan, S. C. H. Hoi, and Z. Xu, "Robust graph learning from noisy data," *IEEE Trans. Cybern.*, vol. 50, no. 5, pp. 1833–1843, May 2020.

[8] J. Cai, H. Han, J. Cui, J. Chen, L. Liu, and S. K. Zhou, "Semi-supervised natural face de-occlusion," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 1044–1057, 2021.

[9] F. Cen and G. Wang, "Dictionary representation of deep features for occlusion-robust face recognition," *IEEE Access*, vol. 7, pp. 26595–26605, 2019.

[10] L. Zhang, L. Sun, L. Yu, X. Dong, J. Chen, W. Cai, C. Wang, and X. Ning, "ARFace: Attention-aware and regularization for face recognition with reinforcement learning," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 4, no. 1, pp. 30–42, Jan. 2022.

[11] Z. Mi, X. Jiang, T. Sun, and K. Xu, "GAN-generated image detection with self-attention mechanism against GAN generator defect," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 5, pp. 969–981, Aug. 2020.

[12] Y. Zhu and Y. Jiang, "Optimization of face recognition algorithm based on deep learning multi feature fusion driven by big data," *Image Vis. Comput.*, vol. 104, Dec. 2020, Art. no. 104023.

[13] H. Yang, C. Gong, K. Huang, K. Song, and Z. Yin, "Weighted feature histogram of multi-scale local patch using multi-bit binary descriptor for face recognition," *IEEE Trans. Image Process.*, vol. 30, pp. 3858–3871, 2021.

[14] J. Lin, Y. Li, and G. Yang, "FPGAN: Face de-identification method with generative adversarial networks for social robots," *Neural Netw.*, vol. 133, pp. 132–147, Jan. 2021.

[15] L. Jiang, X.-J. Wu, and J. Kittler, "Dual attention MobDenseNet(DAMDNet) for robust 3D face alignment," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 504–513.

[16] C. Miao, Z. Tan, Q. Chu, N. Yu, and G. Guo, "Hierarchical frequency-assisted interactive networks for face manipulation detection," *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 3008–3021, 2022.

[17] P. Ventura, A. Bulajić, A. C.-N. Wong, I. Leite, F. Hermens, A. Pereira, and T. Lachmann, "Face and word composite effects are similarly affected by priming of local and global processing," *Attention, Perception, Psychophys.*, vol. 83, no. 5, pp. 2189–2204, Mar. 2021.

[18] H. Qi, C. Wu, Y. Shi, X. Qi, K. Duan, and X. Wang, "A real-time face detection method based on blink detection," *IEEE Access*, vol. 11, pp. 28180–28189, 2023.

[19] Q. Zhou, J. Qin, X. Xiang, Y. Tan, and Y. Ren, "MOLS-Net: Multi-organ and lesion segmentation network based on sequence feature pyramid and attention mechanism for aortic dissection diagnosis," *Knowl.-Based Syst.*, vol. 239, Mar. 2022, Art. no. 107853.

[20] W. Zheng, M. Yue, S. Zhao, and S. Liu, "Attention-based spatial–temporal multi-scale network for face anti-spoofing," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 3, no. 3, pp. 296–307, Jul. 2021.

[21] J. Lee, S. Kim, S. Kim, J. Park, and K. Sohn, "Context-aware emotion recognition networks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 10143–10152.

[22] B.-W. Hwang, M.-C. Roh, and S.-W. Lee, "Performance evaluation of face recognition algorithms on Asian face database," in *Proc. 6th IEEE Int. Conf. Autom. Face Gesture Recognit.*, Jun. 2003, pp. 557–565.

[23] Y. Zhang, Z. Yin, Y. Li, G. Yin, J. Yan, J. Shao, and Z. Liu, "CelebA-Spoof: Large-scale face anti-spoofing dataset with rich annotations," in *Proc. 16th Eur. Conf. Comput. Vis.*, Glasgow, U.K., Aug. 2020.

[24] A.-P. Song, Q. Hu, X.-H. Ding, X.-Y. Di, and Z.-H. Song, "Similar face recognition using the IE-CNN model," *IEEE Access*, vol. 8, pp. 45244–45253, 2020.

[25] F. Tian, H. Xie, Y. Song, S. Hu, and J. Liu, "The face inversion effect in deep convolutional neural networks," *Frontiers Comput. Neurosci.*, vol. 16, May 2022, Art. no. 854218.

[26] J. Deng, J. Guo, J. Yang, N. Xue, I. Kotsia, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 5962–5979, Oct. 2022.

[27] N. Ullah, A. Javed, M. A. Ghazanfar, A. Alsufyani, and S. Bourouis, "A novel DeepMaskNet model for face mask detection and masked facial recognition," *J. King Saud Univ., Comput. Inf. Sci.*, vol. 34, no. 10, pp. 9905–9914, Nov. 2022.

[28] Z. Song, K. Nguyen, T. Nguyen, C. Cho, and J. Gao, "Spartan face mask detection and facial recognition system," *Healthcare*, vol. 10, no. 1, p. 87, Jan. 2022.

DALIN WANG was born in Fengjie, Chongqing, in 2012. He is a Graduate Student with the School of Economics and Management, Chongqing Normal University. He is an Associate Professor with the School of Intelligent Science and Technology, Chongqing Preschool Education College. His main research interest includes computer application.

RONGFENG LI was born in Wanzhou, Chongqing, in 2012. He is a Graduate Student with the School of Economics and Management, Chongqing Normal University. He is also an Associate Professor with the School of Intelligent Science and Technology, Chongqing Preschool Education College. His research interests include cloud computing and big data analysis.