Mobi-FaceNeXt
RESEARCH ON AN EFFICIENT FACE RECOGNITION ALGORITHM BASED ON LIGHTWEIGHT CONVOLUTIONAL NEURAL NETWORKS

by

Bin YUAN, Chang-Qing DU*, and Zi-Tian LI

School of Mathematics and Economics, School of Information Engineering,
and School of Teacher Education, Qujing Normal University, Qujing, China

Original scientific paper
https://doi.org/10.2298/TSCI2602191Y

The advent of deep learning and convolutional neural networks, in conjunction with the unremitting expansion and refinement of face recognition datasets, has precipitated a substantial advancement in face recognition technology based on convolutional neural networks. Nevertheless, in real-world implementation scenarios, numerous disadvantages persist in the deployment of facial recognition technology. The present study focuses on the research of face recognition algorithms based on lightweight convolutional neural networks. A thorough analysis of prevalent facial recognition architectures is conducted, encompassing an examination of numerous network models. The integration of diverse network models strengths is achieved to engineer a lightweight network, designated as Mobi-FaceNeXt, for the purpose of facial feature extraction. While ensuring the accuracy of face recognition, efforts are made to minimize network parameters and computational load. This makes the algorithm deployable on general embedded platforms and devices with limited computing and storage resources. This has significant practical engineering implications. In the research, a joint loss function of MagFace Loss and Center Loss is used for training, and the processed MS-Celeb-1M dataset is utilized to enhance the learning ability of the deep face recognition model for facial features. Depth-wise separable convolution is employed to reduce parameters and computations, and the algorithm is optimized to enhance the network processing speed, thereby facilitating the extraction of facial feature information. The experimental results demonstrate that the Mobi-FaceNeXt model can achieve a superior level of face recognition accuracy while maintaining a low level of network computations and parameters. The technology in question has the capacity to satisfy the requirements of embedded devices and to extract facial feature information with greater efficiency. This suggests a broad range of potential applications.

Key words: *face recognition, lightweight network, convolutional neural network*

Introduction

Since the 1970's, face recognition technology has emerged as one of the most intensively studied topics in the fields of computer vision and biometrics [1]. Initially, methods grounded in facial feature point geometric analysis were predominantly employed. These early approaches relied on measuring the distances and angles between key facial landmarks,

---

* Corresponding author, e-mail: ducq@mail.qjnu.edu.cn

such as the eyes, nose, and mouth, to identify individuals. However, they faced limitations in handling variations in facial expressions, lighting conditions, and pose [2].

Subsequently, subspace-based methods came into the spotlight. These techniques, like principal component analysis (PCA), aimed to project high dimensional facial data onto a lower dimensional subspace, reducing data complexity while retaining the most discriminative features. By representing faces as vectors in this subspace, recognition became more manageable. However, they still struggled to capture fine-grained local details [3].

Later, methods based on local features gained popularity. For instance, the local binary pattern (LBP) method focused on extracting local texture information from facial images. It described the local structure of pixels by comparing the intensity of a central pixel with its neighbors, which was effective in capturing local patterns but might lack global context [4].

Finally, a hybrid approach emerged, combining the strengths of subspace-based and feature-based methods [5]. This approach first used feature-based methods (such as LBP) to extract local features and then employed subspace methods (like PCA) to project and obtain low dimensional, discriminative features. This combination improved recognition performance to some extent by leveraging both local and global information [6].

In the 1980’s, researchers started to integrate deep learning technology into the facial recognition technology system [7]. Deep learning models, especially neural networks, have the ability to automatically learn hierarchical features from data. They can extract complex and abstract features from facial images, which helps in solving many difficulties encountered in facial recognition tasks, such as dealing with large - scale datasets, complex facial variations, and occlusion [8].

In recent years, with the rapid development of convolutional neural networks (CNN) and the availability of large scale facial datasets, significant progress has been made in the research of face recognition algorithms based on lightweight CNN. These lightweight networks are designed to have fewer parameters and lower computational complexity while maintaining high recognition accuracy [9]. They have been widely applied in various real life scenarios, including security systems, access control, and mobile device unlocking. However, challenges still remain, such as achieving a better balance between accuracy, parameter count, and computational complexity, especially when deploying these algorithms on resource constrained devices like embedded systems. This has spurred further research in this area, and this paper aims to contribute to this effort by proposing an improved lightweight face recognition algorithm [10].

Research significance

The remarkable advancements in deep learning and CNN have led to the widespread adoption of facial recognition technology based on these networks in numerous institutions worldwide. The prevailing practice entails the utilization of pre-trained models on cloud servers [11]. These systems rely on cloud computing for processing, utilize communication networks for information transmission, and execute face recognition model computations and comparisons on the servers. This approach has been instrumental in facilitating high-accuracy face recognition in numerous applications.

Nevertheless, several challenges persist. The network model parameters are frequently substantial, and the computational complexity is high. This necessitates substantial hardware resources for operation, yet concomitantly engenders issues such as elevated latency during processing and the potential for security vulnerabilities during data transmission. In scenarios necessitating real-time face recognition, such as in certain access control systems, the substantial computational demands can result in authentication delays, thereby impacting the user experience. Furthermore, in contexts characterized by limited resources, such as embedded devices, the substantial model size and elevated computational demands impede the effective implementation of face recognition technologies. Embedded devices generally possess constrained memory and computing capabilities. Consequently, executing extensive CNN models on these devices can result in performance impediments or even system failures.

The present article puts forth a series of algorithmic enhancements intended to ameliorate the limitations of embedded hardware devices. These enhancements are designed to achieve a more optimal balance among three key metrics: accuracy, parameter count, and computational complexity. By reducing the model parameter size and computational requirements without a significant loss in accuracy, the proposed algorithms can enhance the efficiency of face recognition on embedded devices. This will enable seamless integration of face recognition technology into a wider range of applications, such as smart home devices, wearable devices, and surveillance cameras with limited resources. The development of intelligent and efficient security systems will be further advanced by this technology, which will enable real-time and accurate face recognition on devices with restricted capabilities. The objective of this study is to surmount the computational challenges associated with face recognition in scenarios where resources are limited. This endeavor will facilitate the advancement and expansion of face recognition technology.

Research innovation points

The present paper expounds on several innovative aspects in the research of lightweight face recognition algorithms.

Firstly, with regard to network architecture, an exhaustive analysis of prevalent face recognition architectures is conducted. A thorough examination of various network models, including MobileNet and Mobi-FaceNeXt, has been conducted. This examination involved the integration of the strengths of multiple networks, with particular emphasis on concepts derived from advanced architectures such as Swin Transformer in the domain of computer vision. As a result of this research, a novel lightweight neural network has been developed. This network, named MobiFaceNeXt, has been designed for the extraction of facial features [12]. This novel network incorporates a bottleneck structure analogous to that observed in MobileNetV2 and Mobi-FaceNeXt, a design choice that facilitates the preservation of more information features during the extraction process. Concurrently, it implements a rapid down-sampling strategy at the inception of the network and a dimensionality reduction strategy in subsequent convolutional layers. These strategies effectively reduce the computational complexity of the network while maintaining the ability to extract discriminative facial features, rendering it more suitable for deployment on devices with limited resources.

Secondly, regarding the loss function, it innovatively uses a joint loss function composed of MagFace Loss and Center Loss [13]. ArcFace Loss, a commonly used loss function in face recognition, has certain limitations. Its fixed angle margin penalty coefficient, m, may lead to an unstable intra-class structure in practical applications. In contrast, MagFace Loss improves face recognition performance by considering the quality of face images, which is represented by the L2 norm size (modulus length) of the face feature vector [14]. MagFace Loss not only widens the inter-class spacing but also optimizes the intra-class distribution structure. By combining it with Center Loss, which calculates the distance between each sample and the center of its corresponding category to make the feature representation more compact and categories more separable, the proposed joint loss function can achieve a better training effect [15]. It can bring easily recognizable simple samples closer to the center of the class, push difficult to recognize samples away from the center, and at the same time, make the intra-class distribution more compact while maintaining a good inter-class distance [16].

Finally, in the training process, the paper uses a processed MS-Celeb-1M dataset to train the deep face recognition model. This dataset enhancement method can improve the model ability to learn diverse facial features, making the model more robust and accurate in face recognition tasks [17]. Additionally, depth-wise separable convolution is utilized to significantly reduce network parameters and computational load. At the same time, algorithm optimization techniques are applied to further improve the network processing speed, enabling more efficient extraction of facial feature information. These combined innovations in network architecture, loss function, and training process contribute to the development of a more efficient and practical lightweight face recognition algorithm [18].

Network architecture design

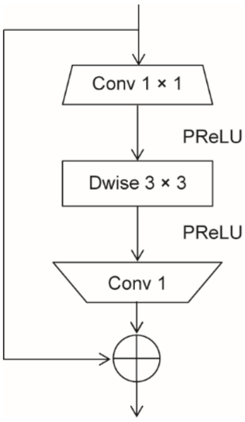

Figure 1. MobileNetV2 and Mobi-FaceNeXt inverted residual structure

In regard to the face recognition task, the present study proposes an optimization and design of lightweight neural networks, leveraging the MobileNet series and Mobi-FaceNeXt networks as a foundation. The objective is to integrate the merits of current lightweight networks in the computer vision field, such as the Swin Transformer, to create a customized lightweight neural network for facial feature extraction tasks.

The inverse residual structure, a fundamental component of both MobileNetV2 and Mobi-FaceNeXt, is subjected to rigorous analysis. As illustrated in fig. 1, this structure comprises a sequence of operations, beginning with a 1 × 1 convolution (Conv 1 × 1), followed by a 3 × 3 depth-wise separable convolution (Dwise 3 × 3), and concluding with an additional 1 × 1 convolution (Conv 1 × 1). This sequence is then followed by the PReLU activation function in both MobileNetV2 and Mobi-FaceNeXt, though it should be noted that alternative activation functions may be employed in other components of the network. The inverse residual structure is of particular importance, as it enables the preservation of additional information features. This preservation is instrumental in maintaining the model performance, ensuring its robustness and facilitating enhanced adaptation to depth-wise separable convolutions. The utilization of this configuration enables the network to concurrently fulfill two primary objectives: the reduction of both parameter and computational complexity, while concurrently achieving effective feature extraction.

In the design of the Mobi-FaceNeXt network, a similar bottleneck structure is adopted. This configuration functions as a fundamental component of the network, enabling the effective extraction of features. In the nascent stage of the Mobi-FaceNeXt network, a rapid down-sampling strategy is implemented. This strategy expeditiously diminishes the magnitude of the input feature maps. This approach has been demonstrated to markedly reduce the computational complexity of subsequent convolutional layers. For instance, if the input feature map possesses a substantial spatial size, reducing it at an early stage results in subsequent convolutions operating on reduced data volumes, thereby necessitating a reduced number of calculations.

In the subsequent convolutional layers, a dimensionality reduction strategy is employed. This strategy further reduces the computational and parameter complexity of the network. By reducing the number of channels or dimensions in the feature maps at specific stages, the network can maintain a high level of performance while utilizing fewer resources. The integration of a bottleneck structure, rapid down-sampling, and dimensionality reduction strategies within the Mobi-FaceNeXt network design renders it an effective lightweight architecture for facial feature extraction.

With regard to the activation function, the PReLU activation function is selected. The PReLU activation function can be regarded as an optimized version of the ReLU activation function [19]. A comparison of the ReLU function and the PReLU activation function reveals that the latter adds minimal parameters to the model. This approach effectively mitigates the risk of overfitting, thereby enhancing the generalizability and reliability of the model. In the context of a neural network, the phenomenon of overfitting arises when the model exhibits excessive reliance on the training data, including extraneous variations, resulting in suboptimal performance when applied to novel data. The PReLU function has been developed to address this issue by enabling the network to learn a more flexible activation pattern. It has the capacity to adaptively adjust the slope of the activation function for different neurons, thereby enhancing the network model ability to fit the data. This property renders the system highly suitable for lightweight neural networks, as it does not compromise the accuracy of model calculations [20].

It has been demonstrated that both MobileNetV2 and Mobi-FaceNeXt employ the same inverted residual structure (with different activation functions). This results in the preservation of more information features, ensuring model performance and facilitating better adaptation to depth-wise separable convolutions. The objective is to attain a reduction in parameter and computational complexity while concurrently accomplishing the extraction of features. The Mobi-FaceNeXt network, as delineated in this paper, employs a bottleneck structure that is analogous to the one described. Furthermore, Mobi-FaceNeXt implements a rapid down-sampling strategy during the initial phase of the network to expeditiously reduce the dimensionality of the input feature maps. This reduction in computational complexity is a key aspect of the subsequent convolutional layers. Concurrently, the implementation of a dimensionality reduction strategy in the subsequent convolutional layers has been demonstrated to further mitigate the computational and parameter complexity of the network [21].

Joint loss function design

In the domain of deep learning for facial feature extraction, the selection of loss function has been shown to have a substantial impact on model performance. ArcFace Loss is a prevalent loss function in the field of face recognition. The subject has been deemed to be of significant value due to its simplicity in implementation and its effectiveness in enhancing both intra-class compactness and interclass differentiation. However, a notable shortcoming of ArcFace Loss lies in its reliance on a fixed angle margin penalty coefficient, m. In practical scenarios, the quality of facial images in the training set can vary significantly, and the fixed m value may not be optimal for all image qualities. Consequently, this phenomenon can result in a chaotic intra-class distribution structure in the classification results, thereby undermining the stability of the intra-class structure [22].

To address this issue, the present paper introduces the MagFace Loss function. MagFace Loss incorporates the quality of facial images into the training process. In the context of a face recognition network, the feature vector of a face is extracted, and the L2 norm size (the modulus length of the feature vector) can be employed to represent the quality of the face image. Intuitively, a larger modulus length of the feature vector corresponds to a higher quality face image, which is typically easier to recognize. MagFace Loss has been demonstrated to achieve two primary objectives. Firstly, it has been shown to expand the interclass spacing. Secondly, it has been demonstrated to optimize the intragroup distribution structure. These effects are achieved by leveraging the concept of the modulus length of the face image feature vector [23].

Mathematically, MagFace Loss is an improvement upon ArcFace Loss. It introduces the feature modulus length, α_i, an adaptive angular margin penalty, $m(\alpha_i)$, based on the modulus length, and a penalty function, $g(\alpha_i)$, to limit larger feature modulus lengths. The formula for MagFace Loss is:

$$L_{\text{MagFace}} = -\frac{1}{N}\sum_{i=1}^{N}\left(\log\frac{\mathrm{e}^{s\{\cos[\theta_{y_i}+m(\alpha_i)]\}}}{\mathrm{e}^{s\{\cos[\theta_{y_i}+m(\alpha_i)]\}}+\sum\limits_{j\neq y_i}\mathrm{e}^{s\cos(\theta_j)}}+\lambda_g g(\alpha_i)\right) \tag{1}$$

For facial image samples, there is a direct relationship between the feature modulus length, α_i, and the angular margin penalty, $m(\alpha_i)$. A larger α_i implies a larger $m(\alpha_i)$, and *vice versa*. Due to the presence of the angular margin penalty, $m(\alpha_i)$, when α_i is smaller, $\cos[\theta_i + m(\alpha_i)]$ becomes larger. This effectively pulls low-quality facial images towards the intra-class boundary. Additionally, $g(\alpha_i)$ is defined as a monotonic decreasing function with respect to the feature module length α_i. This function pushes each sample towards the boundary of the feasible domain and guides high-quality facial images towards the center within the class.

Another important loss function in the context of face recognition is Center Loss. The working principle of Center Loss is to compute the distance between each sample and the center of its corresponding category. By adjusting the feature representation of the model based on these distances, Center Loss makes the feature representation more compact, which in turn enhances the separability of different categories.

In this paper, a new joint loss function, L, is formulated by combining MagFace Loss and Center Loss:

$$L = L_{\text{MagFace}} + \lambda L_{\text{Center}} =$$

$$= -\frac{1}{N}\sum_{i=1}^{N}\left(\log\frac{\mathrm{e}^{s\{\cos[\theta_{y_i}+m(\alpha_i)]\}}}{\mathrm{e}^{s\{\cos[\theta_{y_i}+m(\alpha_i)]\}}+\sum\limits_{j\neq y_i}\mathrm{e}^{s\cos(\theta_j)}}+\lambda_g g(\alpha_i)\right)+\frac{\lambda}{2}\sum_{i=1}^{N}x_i - C_{y_i}{}_2^2 \tag{2}$$

where λ is the weighting factor. The first part of the new joint loss function, which is the MagFace Loss, analyzes the quality of pictures based on the modulus length of their feature vectors. It brings easily recognizable simple samples closer to the center of the class and pushes difficult-to-recognize samples away from the center, thereby maintaining a favorable intra-class distribution structure. The second part, which is the Center Loss, is used for joint training. It reduces the distance between edge samples and the center of the class, making the intra-class distribution more compact. By using this joint loss function, the model can achieve a more compact intra-class distribution while maintaining a good intra-class distance, thus improving the overall performance of the face recognition algorithm [24].

Experiment and result analysis

Experimental environment

The main hardware equipment for the experiments in this paper includes the CPU of Intel I9-13900K, the graphics card of NVIDIA RTX4080S 16G. The software platforms are CUDA 11.1 and PyTorch 1.6.0. The Face recognition algorithm model is the Mobi-FaceNeXt model.

Experimental results and analysis

Comparison of loss functions

The joint loss function proposed in this article involves preset values of weighting factors, λ. Referring to the conclusion obtained from the Center Loss experiment. The preset values are 1, 0.1, 0.05, 0.01, 0.005, and 0.001. The model is Mobi-FaceNeXt. Training is carried out respectively. The accuracy rates of the three datasets are analyzed to obtain a better weighting factor. Due to the results of LFW are relatively saturated and the differences are not significant. The CFP-FP and AgeDB-30 are selected for comparison. At the same time, in the selection of the loss function. Two commonly used loss functions for face recognition as ArcFace Loss and AdaCos Loss are compared [25]. The experimental results as shown in tab. 1 are obtained.

Table 1. Comparison table of experimental results between different λ values of the joint loss function and commonly used loss functions

Loss Function	λ	LFW [%]	CFP-FP [%]	AgeDB-30 [%]
Joint loss function	1	99.69	95.42	96.69
Joint loss function	0.1	99.74	95.35	96.74
Joint loss function	0.05	99.75	95.28	96.55
Joint loss function	0.01	99.74	95.66	96.77
Joint loss function	0.005	99.74	95.33	96.64
Joint loss function	0.001	99.69	95.36	96.73
MagFace Loss	0	99.69	93.91	96.52
ArcFace Loss	–	99.64	93.33	96.19
AdaCos Loss	–	99.72	94.92	95.97

The data indicates that when using the joint loss function proposed in this paper, Mobi-FaceNeXt has the best effect. When the weight parameter $\lambda = 0.01$, the model achieves the optimal value in the CFP-FP and AgeDB-30 datasets. At the same time, it can be found that the accuracy rate of MagFace Loss is lower than that of the joint loss function. It is strongly confirming the validity and rationality of the joint loss function proposed in this article. Through comparing the experimental data, when the weight parameter $\lambda = 0.01$ the training effect is the best. Therefore, this weight parameter value will be used in all subsequent experiments [26].

Comparison of lightweight face recognition algorithms

The Mobi-FaceNeXt proposed in this article is evaluated based on three commonly used public datasets: LFW, CFP-FP, and AgeDB-30. The result data of the recognition accuracy compared with other lightweight face recognition algorithms are shown in tab. 2.

Table 2. Comparison table of accuracy rates of lightweight face recognition algorithm models

Model	LFW [%]	CFP-FP [%]	AgeDB-30 [%]	FLOP	Params
Mobi-FaceNeXt	99.54	91.66	96.05	231.3M	2.2M
LMobileNetE	99.52	88.52	96.08	–	27.7M
ShuffleFaceNet	99.44	95.09	95.15	276.8M	2.4M
MixFaceNet-S	99.62	94.63	96.65	452.7M	4.07M
Seesaw - shuffleFaceNet	99.69	93.09	96.49	155M	3.8M
Mobi-FaceNeXt	99.74	95.66	96.77	232.5M	2.43M

The experimental data demonstrate that the Mobi-FaceNeXt model proposed in this paper exhibits superior recognition accuracy on the three datasets in comparison to other models. This finding suggests that the accuracy and robustness of this model have improved in comparison to other algorithm models. A comparison of the model computational complexity with that of Mobi-FaceNeXt reveals that it remains essentially constant, exhibiting only a marginal increase of 0.23M in parameter count. Nevertheless, the accuracy rates on the LFW, CFP-FP, and AgeDB-30 datasets have shown an increase of 0.2%, 4%, and 0.72%, respectively. This outcome serves to reinforce the efficacy of the Mobi-FaceNeXt facial recognition algorithm model.

Conclusions

This research contributes to the field by building upon the extant literature of lightweight facial recognition algorithm models, including those in the MobileNet series. The development of a lightweight facial recognition algorithm neural network, designated as Mobi-FaceNeXt, was achieved through a meticulous process of optimization and design of the algorithm model. This network was meticulously engineered for the specific purpose of facial feature extraction, representing a significant advancement in the field.

The integration of the joint loss function, which incorporates MagFace Loss and Center Loss, signifies a substantial advancement in the field. This combination effectively enhances the training effect of the face recognition algorithm neural network. The model efficacy in capturing and distinguishing facial features is enhanced by the integration of two loss functions: MagFace Loss, which prioritizes the quality of face images, and Center Loss, which emphasizes minimizing the distance between samples and their class centers. This approach has been demonstrated to yield a more stable and accurate classification performance.

The lightweight neural network Mobi-FaceNeXt, in conjunction with the designed loss function, has undergone extensive training and testing on public datasets. The experimental results are highly promising. The Mobi-FaceNeXt model exemplifies the capacity to sustain a modest volume of network computation and parameters. This is a pivotal benefit, particularly when contemplating its implementation on devices characterized by limited resources, such as embedded systems. Despite its lightweight nature, the model demonstrates a noteworthy level of face recognition accuracy. The efficacy of the proposed algorithm was demonstrated through its superior performance in comparison to numerous existing lightweight face recognition algorithms on several widely used datasets, including LFW, CFP-FP, and AgeDB-30. This finding suggests that the model demonstrates a high degree of accuracy and robustness, rendering it suitable for a broad spectrum of real-world applications [27].

In practical terms, the Mobi-FaceNeXt model has the capacity to meet the specific needs of embedded devices. The technology under discussion has been demonstrated to facilitate the extraction of facial feature information with greater efficiency. This capacity is of critical importance for a variety of applications, including real-time face recognition in surveillance systems, access control in smart buildings, and user authentication in mobile devices. The model favorable balance between computational requirements and recognition accuracy suggests broad application prospects in the fields of security, smart home, and mobile computing. It is anticipated that this technology will contribute to the advancement and dissemination of face recognition technology in contexts where resources are limited, thereby facilitating more convenient and secure user experiences.

In summary, the present study offers a significant contribution to the realm of lightweight face recognition algorithms. The Mobi-FaceNeXt model provides a pragmatic solution to the challenges encountered in real-world applications. Subsequent research endeavors may concentrate on the further refinement of the model to accommodate increasingly intricate scenarios, such as the management of extreme lighting conditions, occlusions, and extensive data sets with greater efficacy.

When considering the broader context of engineering research, recent advancements in various fields have provided valuable insights for the further development of lightweight face recognition algorithms. For instance, the ResNet neural network has been demonstrated to possess a high degree of efficacy in the multi-objective prediction and optimization of vehicle acoustic packages [28]. This finding serves to illustrate the potential of neural networks in the domain of complex optimization problems. The concept of multi-objective optimization has the potential to be integrated into the refinement of the Mobi-FaceNeXt model. Subsequent research endeavors may entail the exploration of methodologies for the concurrent optimization of multiple dimensions of the model, including recognition speed, accuracy, and energy consumption. This multifaceted approach aims to enhance the model adaptability to diverse application scenarios.

The identification of structural damage [29] underscores the necessity of addressing noise and uncertainties in data. In the domain of face recognition, analogous challenges emerge, particularly when dealing with real-world data influenced by factors such as suboptimal lighting conditions, occlusions, and variations in image quality. The methodologies developed in this research for handling noise can inspire new approaches to enhance the robustness of the Mobi-FaceNeXt model in challenging environments.

As demonstrated in the following case study, the use of 3-D printed concrete [30] underscores the importance of incorporating diverse load conditions into the structural design process. In the context of face recognition, it is imperative that the model demonstrate capacity to accommodate varied *loads* or input variations, encompassing diverse facial expressions, poses, and age demographics. The notion of examining mechanical behavior and strength under various loads can be conceptualized as the study of how the Mobi-FaceNeXt model performs under different data distributions and environmental conditions.

Subsequent research endeavors may concentrate on the further refinement of the Mobi-FaceNeXt model to accommodate increasingly intricate scenarios. This includes the development of more effective strategies for managing extreme lighting conditions, such as leveraging techniques from image enhancement and normalization. To address the issue of occlusions, methods inspired by object detection and segmentation could be incorporated to identify and compensate for occluded regions during the recognition process. In the context of extensive data sets, methodologies derived from data mining and distributed computing–comparable to those employed in other large-scale engineering analyses–can be employed to enhance the model training efficiency and its capacity for generalization. By leveraging the extant research achievements in related engineering fields, the lightweight face recognition technology is expected to achieve more significant breakthroughs and have a broader range of applications in the future.

Acknowledgment

This work was supported in part by the Education Department of Yunnan Province. And in part by the Scientific Research Fund Project of Yunnan Provincial Department of Education of China under Grant No.2023J1030. And in part by the Yunnan Provincial Joint Special Project for Basic Research of Local Undergraduate Colleges and Universities under Grant No.202401BA070001-132.

References

[1] Phillips, P. J., *et al.*, The FERET Evaluation Methodology for Face - Recognition Algorithms, *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22* (2000), 10, pp. 1090-1104

[2] Zhao, W., *et al.*, Face Recognition: A Literature Survey, *ACM Computing Surveys*, *35* (2003), 4, pp. 399-459

[3] Deng, W., *et al.*, Compressive Binary Patterns: Designing a Robust Binary Face Descriptor with Random-Field Eigenfilters, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41* (2018), 3, pp. 758-767

[4] Li, L. X., *et al.*, A Review of Face Recognition Technology, *IEEE Access*, *8* (2020), 3, pp. 139110-139120

[5] Yu, L., *et al.*, Principal Patterns of Fractional-Order Differential Gradients for Face Recognition, *Journal of Electronic Imaging*, *24* (2015), 1, pp. 13-21

[6] Mei, M. Q., *et al.*, A Discriminant Subspace Learning Based Face Recognition Method, *IEEE Access*, *6* (2018), 3, pp. 13050-13056

[7] Cao, Y. J., *et al.*, Study of Friction Compensation Model for Mobile Robot's Joints, *Facta Universitatis - Series Mechanical Engineering*, *22* (2024), 4, pp. 721-740

[8] Karim, F. K., *et al.*, Innovative Mathematical Modelling Approaches to Diagnose Chronic Neurological Disorders with Deep Learning, *Thermal Science*, *28* (2024), 6B, pp. 5217-5229

[9] Krizhevsky, A, *et al.*, Imagenet Classification with Deep Convolutional Neural Networks, *Communications of the ACM*, *25* (2012), 2, pp. 84-90

[10] Wang, Z. J., Kao, H. T., Energy Consumption Prediction of Cement Calcination Process Based on Dual-Channel Temporal Convolution Neural, *Thermal Science*, *29* (2025), 1A, pp. 159-172

[11] Ojala, T., *et al.*, Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24* (2002), 7, pp. 971-987

[12] Goyal, M., *et al.*, Robust Methods for Real-Time Diabetic Foot Ulcer Detection and Localization on Mobile Devices, *IEEE journal of Biomedical and Health Informatics*, *23* (2018), 4, pp. 1730-1741

[13] Graves, A., *et al.*, A Novel Connectionist System for Unconstrained Handwriting Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31* (2008), 5, pp. 855-868

[14] Liu, W., *et al.*, Convolutional Two-Stream Network Using Multi-Facial Feature Fusion for Driver Fatigue Detection, *Future Internet*, *11* (2019), 5, pp. 115-126
[15] Li, B. Y. L., *et al.*, Face Recognition Based on Kinect, *Pattern Analysis and Applications*, *19* (2016), 4, pp. 977-987
[16] He, X., *et al.*, Face Recognition Using Laplacianfaces, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27* (2005), 3, pp. 328-340
[17] Lu, J., *et al.*, Face Recognition Using Kernel Direct Discriminant Analysis Algorithms, *IEEE Transactions on Neural Networks*, *14* (2003), 1, pp. 117-126
[18] Guo, M. H., *et al.*, Attention Mechanisms in Computer Vision: A Survey, *Computational Visual Media*, *8* (2022), 3, pp. 331-368
[19] Chang, D., *et al.*, Lightweight of ONNX Using Quantization-Based Model Compression, *The Journal of the Institute of Internet, Broadcasting and Communication*, *21* (2021), 1, pp. 93-98
[20] Hsu, G., *et al.*, RGB-D-Based Face Reconstruction and Recognition, *IEEE Transactions on Information Forensics and Security*, *9* (2014), 12, pp. 2110-2118
[21] Mehryar, E., *et al.*, Nasal Patches and Curves for Expression-Robust 3D Face Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39* (2017), 5, pp. 573-593
[22] Zhang, K, *et al.*, Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks, *IEEE Signal Processing Letters*, *23* (2016), 10, pp. 1499-1503
[23] Jiang L, *et al.*, Robust RGB-D Face Recognition Using Attribute-Aware Loss, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42* (2019), 10, pp. 2473-2481
[24] Kuang, M. H., *et al.*, A Hybrid Deep Learning Approach for Sentiment Analysis in Product Review, *Facta Universitatis - Series Mechanical Engineering*, *21* (2023), 3, pp. 479-500
[25] Gu J, *et al.*, Recent Advances in Convolutional Neural Networks, *Pattern Recognition*, *77* (2018), 12, pp. 1620-1634
[26] Zhu, X, Goldberg, A. B., Introduction to Semi-Supervised Learning, *Synthesis Lectures on Artificial Intelligence and Machine Learning*, *3* (2009), 1, pp. 120-130
[27] Song, F, *et al.*, Eyes Closeness Detection from Still Images with Multi-Scale Histograms of Principal Oriented Gradients, *Pattern Recognition*, *47* (2014), 9, pp. 2825-2838
[28] Wu, Y. R., *et al.*, Multi-Objective Prediction and Optimization of Vehicle Acoustic Package Based on ResNet Neural Network, *Sound and Vibration*, *57* (2023), 1, pp. 73-95
[29] Chen, Z., *et al.*, Multi-Role Collaborative Framework for Structural Damage Identification Considering Measurement Noise Effect, *Measurement*, *250* (2025), 117106
[30] Liu, H., *et al.*, Design of 3D printed Concrete Masonry for Wall Structures: Mechanical Behavior and Strength Calculation Methods under Various Loads, *Eng. Struct.*, *325* (2025) 119374

Paper submitted: September 19, 2024
Paper revised: May 9, 2025
Paper accepted: May 10, 2025