https://doi.org/10.31891/2307-5732-2025-355-48
УДК 004.42:004.932

MAKSYMENKO DMYTRO
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"
https://orcid.org/0009-0009-3612-173X
e-mail: dm.maksymenko@gmail.com
SHKURAT OKSANA
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"
https://orcid.org/0000-0001-7633-9121
e-mail: shkurat@pzks.fpm.kpi.ua
DYCHKA ANDRII
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"
https://orcid.org/0000-0003-0578-2788
e-mail: andriydychka@gmail.com

DEEP LEARNING METHODS BASED ON MOBILENETV2 MODEL FOR LOW-LIGHT FACE RECOGNITION

In this research paper, the problem of reduction of the recognition accuracy for face image captured in low-light environments in modern recognition methods is considered. Reviewed recognition methods are based on deep learning methods for extracting object of interest (a face), digital transformation of an image and different convolutional neural networks models (VGG-Face, ArcFace, FaceNet, DeepFace, SphereFace, CosFace, OpenFace) for feature extraction and classification. Two methods are proposed for facial recognition, based on the MobileNetV2 deep learning model architecture with the integration of additional blocks for feature extraction enhancements. One of the proposed methods includes prior classification of images depending on the lighting conditions of an image, based on the usage of a CNN with accuracy of 99.33%, after which a correspondingly-trained MobileNetV2 architecture-based model is used for recognition itself. Proposed methods were tested on the UTKFace dataset, as well as generated DistortionFace and NormalFace datasets, for recognition of face images by features such as age, gender and ethnicity. Each dataset consists of 23 708 images in total, which are then divided into parts of 16 595 images for training, 4 741 for validation and 2 371 for testing. Proposed methods for facial recognition based on MobileNetV2 showed mean absolute error of 6.38 for age, as well as accuracy of 88.09% for gender and accuracy of 72.93% for ethnicity. Proposed method for face recognition based on a lighting classification CNN model and MobileNetV2 model showed mean absolute error of 6.74 for age, as well as accuracy of 87.14% for gender and accuracy of 70.53% for ethnicity. Proposed methods demonstrate robustness in low-light face images recognition and potential for further research and improvements.

***Keywords**: Convolutional Neural Network, MobileNetV2, Deep Learning, Image Recognition, Low-Light Image, Face Recognition.*

МАКСИМЕНКО ДМИТРО
ШКУРАТ ОКСАНА
ДИЧКА АНДРІЙ
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»

МЕТОДИ ГЛИБИННОГО НАВЧАННЯ ДЛЯ РОЗПІЗНАВАННЯ ОБЛИЧ, ВІЗУАЛІЗОВАНИХ В УМОВАХ НЕДОСТАТНЬОГО ОСВІТЛЕННЯ НА ОСНОВІ МОДЕЛІ MOBILENETV2

В статті розглядається проблема зниження точності розпізнавання в сучасних методах розпізнавання облич для зображень низької якості, які візуалізовані в умовах недостатнього освітлення. Розглянуті методи розпізнавання ґрунтуються на методах глибинного навчання для виділення об'єкта інтересу (обличчя), цифрового перетворення зображення та різних моделях згорткових нейронних мереж (VGG-Face, ArcFace, FaceNet, DeepFace, SphereFace, CosFace, OpenFace) для вилучення ознак зображень та класифікації. Запропоновано два методи розпізнавання облич, що ґрунтуються на моделі глибинного навчання архітектури MobileNetV2 з інтеграцією додаткових блоків для покращення вилучення ознак. Один із запропонованих методів включає попередню класифікацію зображень відповідно до умов візуалізації зображень з використанням моделі CNN з точністю 99,33%, після чого застосовуються відповідні навчені моделі архітектури MobileNetV2. Запропоновані методи протестовано на наборі даних UTKFace та згенерованих наборах даних DistortionFace та NaturalFace для розпізнавання зображень облич за такими ознаками, як вік, стать та етнічна приналежність. Запропонований метод розпізнавання облич на основі моделі MobileNetV2 показав середню абсолютну похибку віку 6,38, точність статі – 88,09%, точність етнічної приналежності – 72,93%. Запропонований метод розпізнавання облич на основі CNN моделі та моделі MobileNetV2 показав середню абсолютну похибку віку 6,74, точність статі – 87,14%, точність етнічної приналежності – 70,53%. Запропоновані методи демонструють стійкість при розпізнаванні зображень низької якості, які візуалізовані в умовах недостатнього освітлення та потенціал для подальшого вдосконалення.

***Ключові слова**: згорткова нейронна мережка, MobileNetV2, глибинне навчання, розпізнавання зображень, зображення низької якості, розпізнавання облич.*

Стаття надійшла до редакції / Received 06.05.2025
Прийнята до друку / Accepted 26.05.2025

Introduction

Face recognition is a prominent tool to authenticate a person in many applications such as military, healthcare, public security, video surveillance, access controls for mobile devices, etc. At present, there are a considerable number of state-of-the-art artificial intelligence models specifically designed for face recognition tasks. However, their performance largely depends on the high quality of the input images. This poses a significant challenge in situations where faces are captured under low-light conditions, which substantially degrades image quality. Consequently, when the quality of input data cannot be ensured, face recognition systems may encounter serious difficulties in fulfilling their intended functions.

The primary challenge associated with low-light images lies in the physical limitations of camera hardware, both outdated and modern. This issue is particularly relevant given that a significant proportion of such images are captured using smartphones, which are typically equipped with small sensors and narrow-aperture lenses. As a result, substantial differences arise in per-pixel characteristics between images taken in high-light and low-light conditions. These differences manifest not only as reduced brightness but also through distinctive artifacts such as elevated noise levels, which often present considerable challenges for computer vision systems, sepia-like color shifts, and others. Consequently, the accuracy of recognition systems can significantly deteriorate when processing such images.

Therefore, the development of a deep learning-based method capable of effectively addressing these issues and ensuring robust performance on both high-light and low-light images represents a pertinent research objective. Advancements in this area have the potential to yield practical benefits and significantly enhance the reliability of facial recognition systems under challenging lighting conditions.

Research objective

The objective of this research is to develop an efficient and robust deep learning method for face recognition from images captured in high-light and low-light conditions. Applying a CNN-based model, the method initially classifies an image type. Applying a trained deep learning model of MobileNetV2 architecture and attentional block based on previous classification, the method extracts image features and recognizes them by age, gender, and ethnicity attributes. This enables increasing accuracy of recognition for face image not limited to different visualization conditions.

Analysis of the latest research and publications

Practical use cases of face recognition are various. Facial images can be recognized by person, emotion, race, age and other attributes. A modern face recognition is based on two general stages: face detection on an image and feature extraction from an image.

Face detection and alignment in an early stage can increase the accuracy of face recognition. There are a lot of common machine learning-based approaches to object detection, such as Haar feature-based cascade classifiers [1] and single-shot multibox detectors (SSD) [2] offered in the OpenCV library, RetinaFace model [3]. Histogram of oriented gradients (HOG) [4] and a CNN-based Max-margin object detection (MMOD) [5] and finally Multi-task cascaded convolutional networks (MTCNN) [6] offered in the Dlib library.

Deep learning models of a CNN-based architecture are an approach to feature extraction from an image and recognition. There are a lot of face-targeted models to recognize a person from an image. The VGG-Face [7], ArcFace [8], FaceNet [9], DeepFace [10], SphereFace [11], and CosFace [12] models implement person verification and recognition from a face image efficiently. Fewer deep learning models have been developed to recognize the age, gender, and race of a person [13-15]. All these models do not have architectural features specifically designed for face recognition but instead use more general convolutional neural networks, slightly tuned for the specific task. Enhancement for these models can be an attention block integrated into deep learning architecture. Attention blocks like Squeeze-and-Excitation (SE) or Convolutional Block Attention Module (CBAM) detect the most informative fragments of images for model training. The SE module introduces channel-wise attention by reweighting feature channels modules and has been plugged into many models to improve accuracy at minimal additional cost [16]. The CBAM module extends this idea by sequentially applying both channel and spatial attention. It was tested in facial emotion recognition and was found to be significantly improving performance on facial expression datasets [17]. RobFaceNet is another example of an attention module improving face-related CNN tasks [18].

Another approach to feature extraction from an image and recognition is to employ a set of deep learning models integrated in a framework. Lightface [19-20] is a hybrid face recognition framework that enables to switch face recognition models among state-of-the-art ones: VGGFace, FaceNet, OpenFace, DeepFace. Deepface [21] is a framework for recognizing a face and facial attributes – age, gender, emotion and race. Deepface wraps state-of-the-art models: VGG-Face, Google FaceNet, OpenFace, Facebook DeepFace, DeepID, ArcFace, Dlib and Sface. Deep face recognition framework in [22] consists of a feature restoration network, a feature extraction network, and an embedding matching module.

The performance of many state-of-the-art deep face recognition models and frameworks decreases significantly for images captured under low illumination. Therefore, including image enhancement methods, for example, contrast-limited adaptive histogram equalization [23], the multi-scale retinex algorithm [24], and the RetinexNet-based deep learning model [25], in state-of-the-art deep face recognition models and frameworks can improve the recognition accuracy.

Proposed method

The first proposed method for low-light face recognition is based on digital preprocessing and MobileNetV2 deep learning model. The architecture of the first proposed method is shown in Fig. 1.

The digital preprocessing block detects an object of interest (face) and aligns an image, using Haar feature-based cascade classifiers. After that, an image is resized and normalized by the brightness values in the range [0;1].

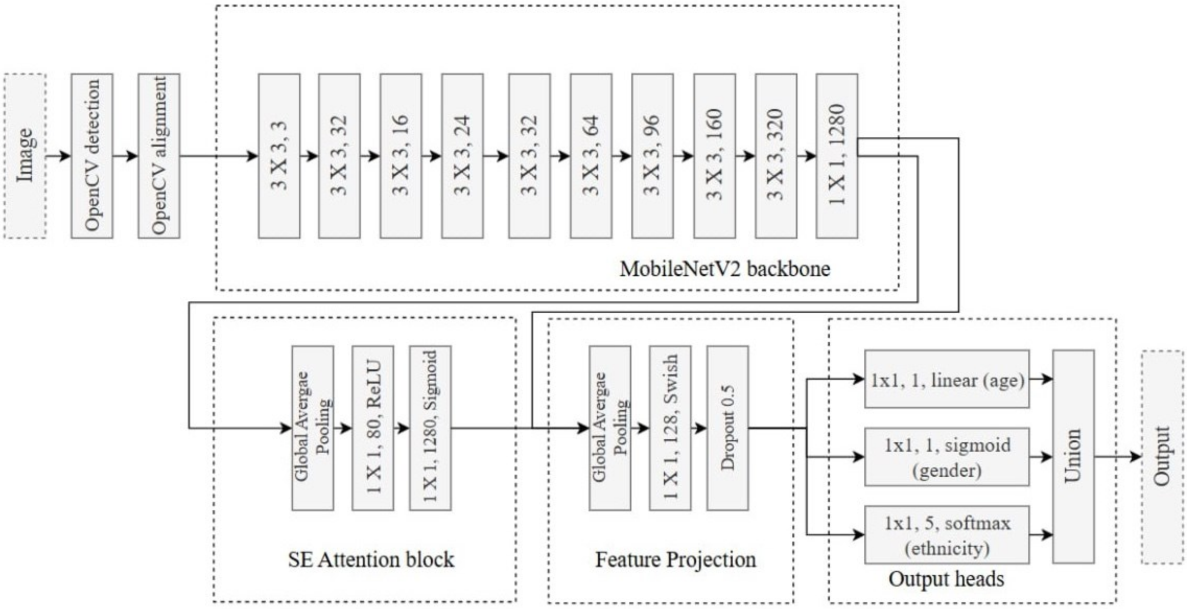

Fig.1. Architecture of the proposed method for low-light face recognition based on digital preprocessing and the MobileNetV2 model

The MobileNetV2 model recognizes a face high-light or low-light image by age, gender, and ethnicity. In the SE Attentional block, the global average pooling is performed, which averages each channel spatially and defines one value per channel. The next, ReLU activation function is applied. To expand the reduced representation back to the original number of channels, sigmoid activation function is applied. Finally, the result of the Attentional block integrates with the original feature map. In the Feature Projection block, the global average pooling is performed. The next, fully connected layer is performed. Finally, the dropout is performed, which freezes randomly half of the neurons to prevent overfitting. In the Output block, the various activation and loss functions are performed since the MobileNetV2 model recognizes age, gender, and ethnicity at the same time.

Age recognition is a prediction of a linear parameter, so the linear activation function (1) is performed, and the loss function mean squared error is estimated (2).

$$\sigma(x) = c \times x \tag{1}$$

$$Loss = \frac{1}{N} \times \sum (p - y)^2 \tag{2}$$

Gender recognition is a binary classification with male and female classes, so the sigmoid activation function (3) is performed, and the loss function binary cross entropy is estimated (4).

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{3}$$

$$Loss = -[y \cdot \log(p) + (1 - y) \cdot \log(1 - p)] \tag{4}$$

Ethnicity recognition is a multi-class classification with white, black, asian, indian and other classes, so the softmax activation function (5) is performed, and the loss function categorical cross entropy is estimated (6).

$$\sigma(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}} \tag{5}$$

$$Loss = -\sum y \times \log(p) \tag{6}$$

The second proposed method for low-light face recognition is based on digital preprocessing and three deep learning models. The architecture of the second proposed method is shown in Fig. 2.

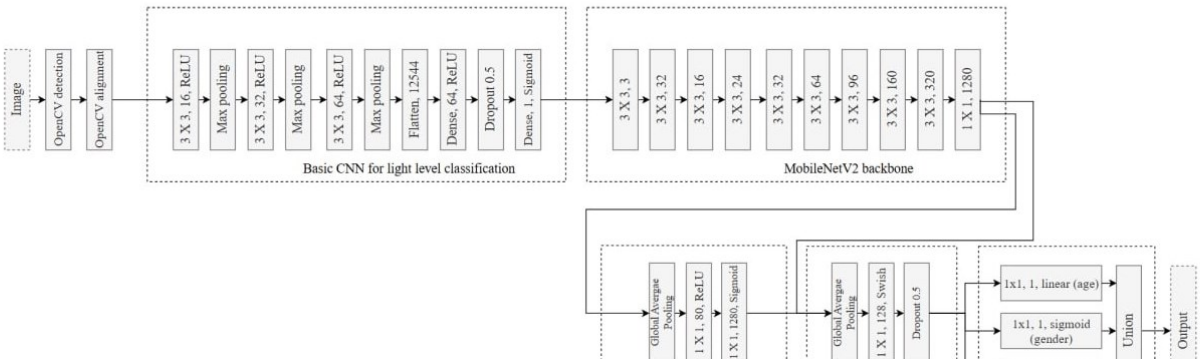

Fig.2. Architecture of the proposed method for low-light face recognition based on digital preprocessing, the MobileNetV2 and CNN-based models

The CNN-based model classifies images in high-light and low-light. Based on the result of such a previous classification, a face image is directed to MobileNetV2 model trained on low-light or high-light images. The accuracy of the proposed CNN-based model for image classification is 99.33%. The second proposed method is a hybrid method that incorporates multiple deep learning models to develop an accurate and robust solution for face recognition.

Experimental results

The proposed deep learning method based on the enhanced MobileNetV2 model for low-light face recognition has been performed by Python programming language, on TensorFlow and Keras machine learning frameworks, and Deep Face and OpenCV image processing libraries.

The proposed deep learning method has been applied to the UTKFace dataset [26], which contains 23 708 images with 5 ethnicities and numerical ages. The proposed dataset has been divided into a training set counted 16 595 images, a validation set counted 4 741 images, and a testing set counted 2 371 images. Based on images of the UTKFace dataset and the simulate module, the two novel datasets, which we call DistortionFace and NaturalFace, have been created. The processing module employs noise and sepia filters and randomized brightness reduction to images of the UTKFace dataset. All images of the DistortionFace dataset are processed by the simulate module. Half of the images of the NaturalFace dataset are processed by the simulate module. The DistortionFace and NaturalFace datasets have been divided the same as the UTKFace dataset for a training, a validation and a testing.

The MobileNetV2 model is trained on 500 epochs with an early stopping function, when the network training process is stopped if there is no improvement. Commonly, the early stopping has been between 50 and 80 epochs. The Adam optimizer with the learning rate of 0.0001 and loss weights of 0.1 for age, 1.0 for gender, and 1.0 for ethnicity has been applied. The results of training and validation of the proposed MobileNetV2 model are shown in Fig. 3.

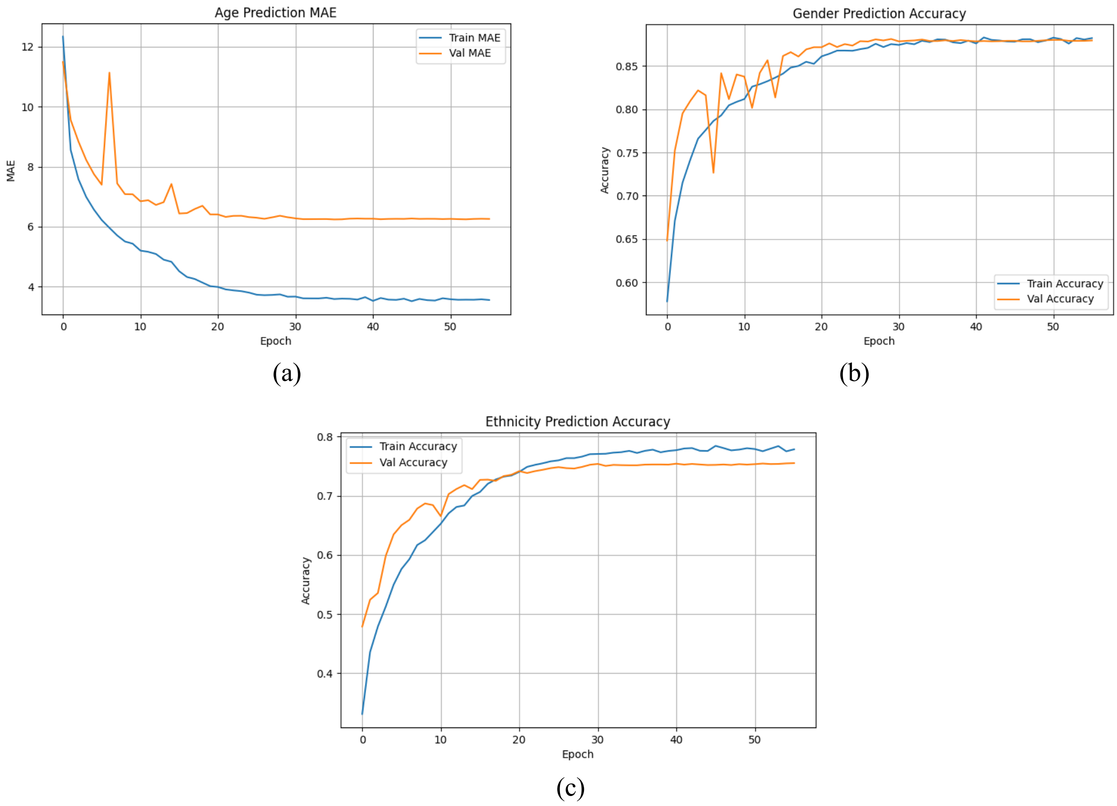

Fig. 3. The results of training and validation of the proposed MobileNetV2 model: (a) the mean absolute error (MAE) is used to predict an age, $MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$, where $\hat{y}$ – predicted age, y – true age; (b) the accuracy is used to predict a gender, $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$, where TP is true positives, FP is false positive, TN is true negatives, and FN is false negatives; (c) the accuracy is used to predict an ethnicity.

The results of face recognition by the proposed deep learning methods based on the MobileNetV2 model are shown in Fig.4.

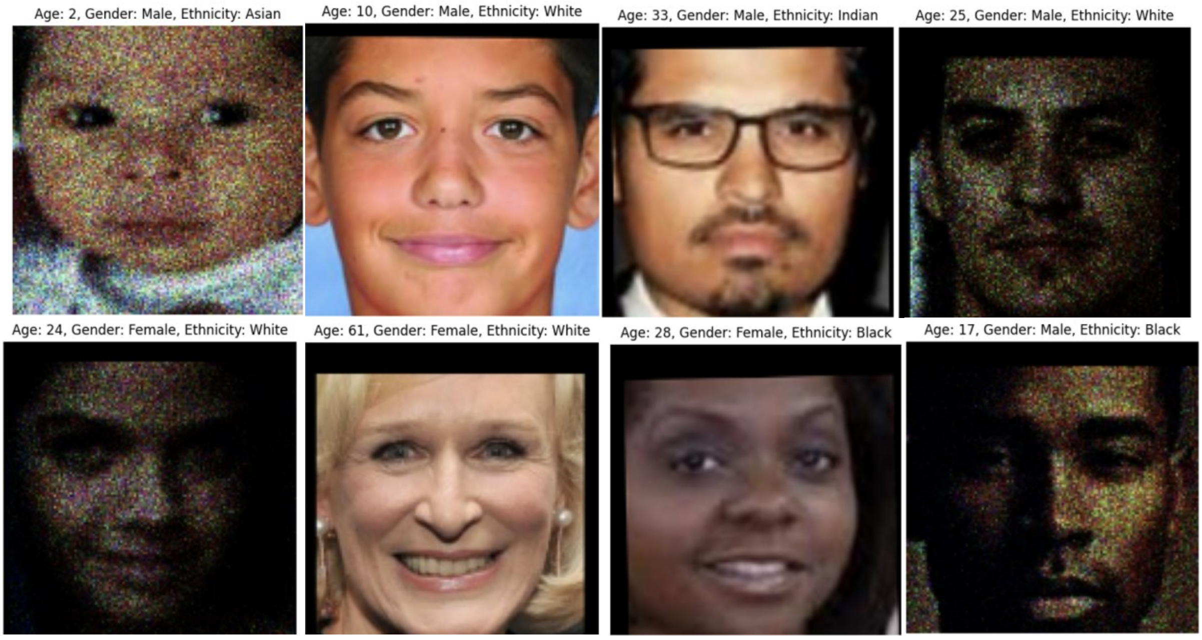

Fig.4. The results of face recognition by age, gender, and ethnicity attributes for DistortionFace and NaturalFace datasets

Table 1 and Table 2 show results of applying the proposed methods with different hyperparameters to face recognition for DistortionFace and NaturalFace datasets. Datasets for training a model, rates of learning, and an attention block are hyperparameters.

Table 1

Evaluation of the proposed deep learning-based face recognition methods for the DistortionFace dataset.

Method	Evaluation Indicators		
	MAE of Age	Gender Accuracy	Ethnicity Accuracy
MobileNetV2 model trained on the UTKFace dataset with SE attention block	22.20	53.25	47.47
MobileNetV2 model trained on the DistortionFace dataset with SE attention block	7.11	86.17	69.39
MobileNetV2 model trained on the NaturalFace dataset with SE attention block	6.80	87.06	72.09
CNN-based model and MobileNetV2 model with SE attention block	7.11	86.17	69.39

Table 2

Evaluation of the proposed deep learning-based face recognition methods for the NaturalFace dataset.

Method	Evaluation Indicators		
	MAE of Age	Gender Accuracy	Ethnicity Accuracy
MobileNetV2 model trained on the UTKFace dataset with SE attention block	14.05	72.34	62.06
MobileNetV2 model trained on the DistortionFace dataset with SE attention block	11.23	76.56	56.83
MobileNetV2 model trained on the NaturalFace dataset with SE attention block	5.96	89.12	73.78
CNN-based model and MobileNetV2 model with SE attention block	6.37	88.11	71.67
MobileNetV2 model trained on the NaturalFace dataset	6.50	86.89	72.47
MobileNetV2 model trained on the NaturalFace dataset with CBAM attention block	6.21	88.62	73.90

As is evident from the results in Tables 1 and 2, the proposed method for face recognition based on the MobileNetV2 model on average has the mean absolute error for age of 6.38, the gender accuracy is 88.09%, the ethnicity accuracy is 72.93%.

As is evident from the results in Tables 1 and 2, the proposed method for face recognition based on the CNN-based model and the MobileNetV2 model on average has the mean absolute error for age of 6.74, the gender accuracy is 87.14%, the ethnicity accuracy is 70.53%.

Accordingly, applying the proposed deep learning methods based on the MobileNetV2 model allows obtaining accurate face recognition results with both low-light and high-light images.

Conclusions

The paper introduces two deep learning methods based on the MobileNetV2 model for face recognition according to age, gender, and ethnicity attributes from face images captured in low-light and high-light conditions. The proposed face recognition methods employ a digital preprocessing block to detect and align a face image, the CNN-based model to initial classify the image in high-light and low-light, the MobileNetV2 model, Attentional, Feature Projection and Output blocks to extract features from the image and recognize ones.

Practical experiments have been shown the ability to generalize of the proposed face recognition methods. The proposed deep learning methods based on the MobileNetV2 model recognize a face image according to age, gender, and ethnicity attributes with both low-light and high-light with almost the same recognition accuracy. The proposed method for face recognition based on the MobileNetV2 model on average has the mean absolute error for age of 6.38, the gender accuracy is 88.09%, the ethnicity accuracy is 72.93%. The proposed method for face recognition based on the CNN-based model and the MobileNetV2 model on average has the mean absolute error for age of 6.74, the gender accuracy is 87.14%, the ethnicity accuracy is 70.53%.

Further research can be focused on exploring image distortion and image enhancement technologies to develop adaptive preprocessing blocks for the proposed face recognition methods. The next point of further research can be focused on exploring more models of a CNN architecture for face recognition to integrate in the proposed face recognition methods. The final point of further research can be focused on exploring a multi-class recognition of the proposed CNN-based model for previous classification of with both low-light and high-light images.

References

1. Viola, P., & Jones, M. J. (2004). Robust real-time face detection. *International Journal of Computer Vision*, 57(2), 137-154. DOI: 10.1023/B:VISI.0000013087.49260.fb
2. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A.C. (2016). SSD: Single shot multibox detector. In *European Conference on Computer Vision (ECCV 2016)*, 9905, 21-37. DOI: 10.1007/978-3-319-46448-0_2
3. Deng, J., Guo, J., Zhou, Y., Yu, J., Kotsia, I., & Zafeiriou S. (2019). RetinaFace: Single-stage dense face localisation in the wild. *arXiv:1905.00641*. DOI: 10.48550/arXiv.1905.00641
4. Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, CA, USA, 2005, 1, 886-893. DOI: 10.1109/CVPR.2005.177
5. King, D.E. (2015). Max-margin object detection. *arXiv:1502.00046*. DOI: 10.48550/arXiv.1502.00046
6. Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multi-task cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10), 1499-1503. DOI: 10.48550/arXiv.1604.02878
7. Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition. In *British Machine Vision Conference 2015 (Swansea: BMVA)*, 41.1-41.12. DOI: 10.5244/C.29.41
8. Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). ArcFace: additive angular margin loss for deep face recognition. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 4685-4694. DOI: 10.1109/CVPR.2019.00482
9. Schroff, F., Kalenichenko, D., Philbin, J. (2015). FaceNet: a unified embedding for face recognition and clustering. *arXiv:1503.03832*. DOI: 0.48550/arXiv.1503.03832
10. Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). DeepFace: closing the gap to human-level performance in face verification. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 1701-1708. DOI: 10.1109/CVPR.2014.220
11. Liu, W., Wen, Y., Yu, Z., Li, M., Raj B., & Song, L. (2017). SphereFace: Deep hypersphere embedding for face recog-nition. In *IEEE Conference Computer Vision and Pattern Recognition*, 6738-6746. DOI: 10.1109/CVPR.2017.713
12. Wang, H., Wang, Y., Zhou, Z., Ji, X., Gong, D., Zhou, J., Li, Z., & Liu, W. (2018). CosFace: Large margin cosine loss for deep face recognition. In *IEEE Conference Computer Vision and Pattern Recognition*, 5265-5274. DOI: 10.48550/arXiv.1801.09414
13. Yi, D., Lei, Z., & Li, S. Z. (2015). Age estimation by multi-scale convolutional network. *Lecture Notes in Computer Science*, 9005, 144-158. DOI: 10.1007/978-3-319-16811-1_10
14. Levi, G., & Hassncer, T. (2015). Age and gender classification using convolutional neural networks. In *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Boston, MA, USA, 34-42. DOI: 10.1109/CVPRW.2015.7301352
15. Eidinger, E., Enbar, R., & Hassner, T. (2014). Age and gender estimation of unfiltered faces. *IEEE Transactions on Information Forensics and Security*, 9(12), 2170-2179. DOI: 10.1109/TIFS.2014.2359646
16. Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-Excitation Networks. *arXiv:1709.01507*, 2018, 7132–7141. DOI: 10.48550/arXiv.1709.01507

17. Miskow, A., & Altahhan, A. (2024). Emotion Recognition with Facial Attention and Objective Activation Functions. *arXiv:2410.17740*. DOI: 10.48550/arXiv.2410.17740

18. Khalifa, A., Abdelrahman, A., Hempel, T., & Al-Hamadi, A. (2024). Towards Efficient and Robust Face Recognition through Attention-Integrated Multi-Level CNN. *Multimedia Tools and Applications*, 1–23. DOI: 10.1007/s11042-024-19521-0

19. Serengil, S.I., & Ozpinar A. (2020). LightFace: A hybrid deep face recognition framework. In *2020 Innovations in Intelligent Systems and Applications Conference (ASYU)*, Istanbul, Turkey, 1-5. DOI: 10.1109/ASYU50717.2020.9259802

20. Serengil, S.I., & Ozpinar A. (2020). HyperExtended LightFace: A facial attribute analysis framework. In *2021 International Conference on Engineering and Emerging Technologies (ICEET)*, Istanbul, Turkey, 1-4. DOI: 10.1109/ICEET53442.2021.9659697.

21. Serengil, S.I., & Ozpinar, A. (2024). A Benchmark of facial recognition pipelines and co-usability performances of modules. *Journal of Information Technologies*, 17(2), 95-107. DOI: 10.17671/gazibtd.1399077

22. Huang, Y.-H., & Chen, H. (2022). Deep face recognition for dim images. *Pattern Recognition*, 126(10). DOI: 10.1016/j.patcog.2022.108580

23. Ma, J., Fan, X., Yang, S., Zhang, X., & Zhu, X. (2017). Contrast Limited Adaptive Histogram Equalization Based Fusion in YIQ and HSI Color Spaces for Underwater Image Enhancement. Int. J. Pattern Recognit. Artif. Intell., 32. DOI: 10.1142/S0218001418540186

24. Meng, C., Zhang, J., Chu, H., Xi, K., & Zhao, B. (2022). Remote Sensing Image Enhancement Based on MSR and CLAHE. WCSE 2022 Spring. DOI: 10.18178/wcse.2022.04.059

25. Hu, Z., Shkurat, O., Przystupa, K., Kochan, O., Ivakhnenko, M. Low-light image enhancement technology based on image categorization, processing and Retinex deep network. *International Journal of Image, Graphics and Signal Processing (IJIGSP)*, 16(5), 1-13. DOI: 10.5815/ijigsp.2024.05.01

26. Zhang, Z., Song, Y., & Qi, H. (2017). Age progression/regression by conditional adversarial autoencoder. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA*, 4352-4360. DOI: 10.1109/CVPR.2017.463.