Complex & Intelligent Systems (2025) 11:139
https://doi.org/10.1007/s40747-024-01762-z

ORIGINAL ARTICLE

A joint learning method for low-light facial expression recognition

Yuanlun Xie[1,2] · Jie Ou[2] · Bihan Wen[3] · Zitong Yu[4] · Wenhong Tian[2]

Received: 20 April 2024 / Accepted: 20 December 2024 / Published online: 9 January 2025

Abstract

Existing facial expression recognition (FER) methods are mainly devoted to learning discriminative features from normal-light images. However, their performance drops sharply when they are used for low-light images. In this paper, we propose a novel low-light FER framework (termed LL-FER) that can simultaneously enhance the images and recognition tasks of low-light facial expression images. Specifically, we first meticulously design a low-light enhancement network (LLENet) to recover expressions images' rich detail information. Then, we design a joint loss to train the LLENet with FER network in a cascade manner, so that the FER network can guide the LLENet to gradually perceive and restore discriminative features which are useful for FER during the training process. Extensive experiments show that the LLENet not only achieves competitive results both quantitatively and qualitatively, but also in the LL-FER framework, which can produce results more suitable for FER tasks, further improving the performance of the FER methods.

Keywords Facial expression recognition · Low-light image enhancement · High-level vision task · Cross high- and low-level learning · Joint training

Introduction

Facial expression recognition (FER) [1] is one of the primary tasks in computer vision, and has attracted the great attention from researchers. Nowadays, FER technology has been rapidly developed, benefiting from the continuous improvement of deep learning algorithms [2–4] and increasing datasets. However, most of the datasets used in the existing FER algorithms are collected under appropriate brightness and good visual quality conditions. The performance of these algorithms can be greatly challenged when in low-light environments, as low-light facial expression images may have blurred textures and low luminance. In general, the performance degradation challenges of the FER method due to low-light environment can be mitigated by pre-training the FER model on a normal-light facial expression dataset or by improving the quality of low-light facial expression images using a trained-well low-light image enhancement (LLIE) model. However, the former does not fundamentally solve the problem of low accuracy of low-light FER, it only achieves some improvement from the perspective of training strategy, which may cause the waste of training resources to some extent. Although the latter improves the visual effect of expression images, it may be suboptimal for the FER task. Because it is trained independently of the FER task, but cannot obtain useful information (e.g., semantic information, etc.) from the FER task, only visual effects and luminance information can be recovered. Therefore, restoring certain high-level vision information during image enhancement may be one of the effective means to improve the low-light FER problem.

✉ Wenhong Tian
tian_wenhong@uestc.edu.cn

Yuanlun Xie
fengyuxiexie@163.com

Jie Ou
oujieww6@gmail.com

Bihan Wen
bihan.wen@ntu.edu.sg

Zitong Yu
yuzitong@gbu.edu.cn

1 College of Electronic Information and Electrical Engineering, Chengdu University, Chengdu, Sichuan, China

2 School of Information and Software Engineering, University of Electronic Science and Technology of China, Shahe Campus, Chengdu, China

3 The School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, Singapore

4 The Great Bay University, Dongguan, Guangdong, China

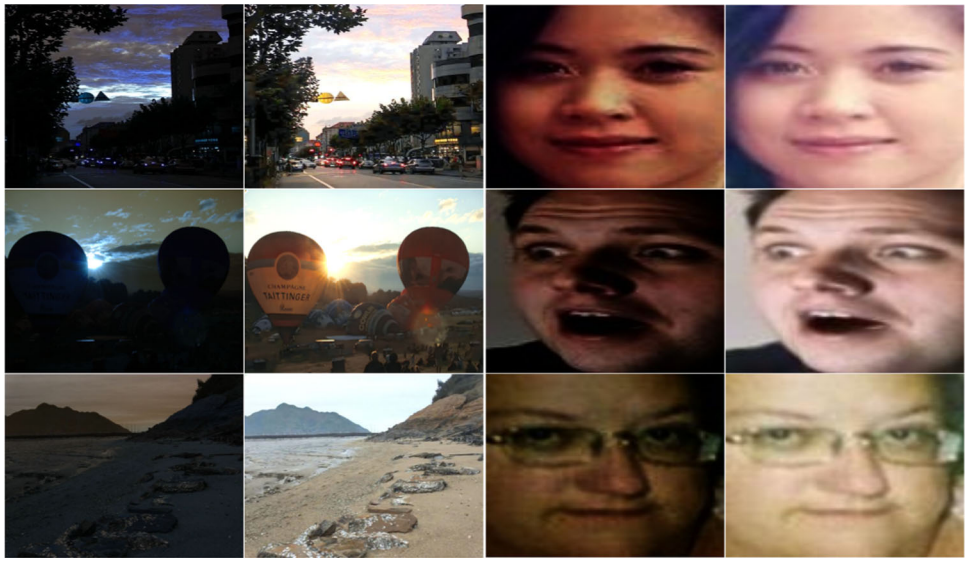

Fig. 1 The effect of the proposed LLIE method. Apply to natural low-light images and low-light facial expression images, respectively

Existing approaches have also presented a great performance on the FER tasks, such as EfficientFace [5], DMUE [6]. Unfortunately, their performance will deteriorate for low-light images due to their inability to extract discriminative features from low-light images. On the other hand, many popular LLIE methods use only the pixel-level error between the restored and the reference images to train the model. However, they ignore semantic information required for high-level vision tasks (e.g., FER), resulting in restored images which cannot effectively serve high-level vision tasks. Therefore, it is an interesting and meaningful work to explore the interaction between the low-light facial expression enhancement task and the FER task, and this has naturally becomes the main research theme of this paper. Based on this observation, we take an initiative to consider designing a suitable low-light enhancer for low-light facial expression images, that can be externally driven to simultaneously restore the color and brightness information and preserve the semantic perception information of facial expression classification tasks. To address the challenge that the semantic information required by FER task cannot be retained in the process of conventional LLIE methods, we propose a novel design a strategy to allow the LLIE model to be guided by FER during training, so as to restore facial expression images with high-level semantic information.

In this paper, we first propose a low-light image enhancement network (LLENet) by combining ResNet [7] and ConvNeXt [8] for low-light facial expression images, and secondly combine it with the FER model to form an end-to-end low-light FER framework. Further, to enable the high-level semantic information of the FER task to flow back to the enhancer, we propose to collect the loss functions of both to design a joint loss function as the overall optimization objective, which forces the enhancer to focus not only on the pixel-level information of the facial expression image during training, but also to perceive certain high-level vision information. The low-light enhancer trained by this low-light FER framework is able to recover both visual quality and high-level visual information. More meaningful, this new solution makes the improvement of the low-light FER task practically possible. Extensive experiments show that our proposed LLENet achieves remarkable enhancement for the low-light image (as shown in Fig. 1), and our cascade solution (LL-FER) for FER in low-light environment also achieves good results on two FER datasets. Overall, the main contributions of this paper are listed below:

1. We design a novel low-light FER framework (LL-FER) by combining LLIE and FER networks, which enables the LLIE network to be optimized in the right direction with the guidance of the FER network and facilitates the FER task.
2. An LLIE network (LLENet) is designed to recover the lost details and luminance of the light-degraded expression images in the framework, meanwhile, under the guidance of FER, the LLENet pays more attention to the necessary details and discriminative features required for the FER task.
3. Extensive experiments on the LLIE and FER datasets demonstrate the effectiveness of the proposed LLENet and LL-FER framework.

The rest of the paper is organized as follows: Sect. 2 presents the reviews of FER methods, LLIE methods, and some explorations of joint learning of high-level vision and low-level vision tasks. In Sect. 3, the designed LLIE method and the joint learning framework are introduced in detail. Section 4 presents the datasets used for the evaluation exper-

iments and shows the experimental results of the designed LLIE and the joint learning methods. Section 5 summarizes the proposed solution for low-light FER and suggests research directions for future work.

Related work

In this paper, we briefly review related works from two aspects: the general low-light image enhancement methods and FER models.

Low-light image enhancement

Li et al. [9, 10] propose a new single-image brightening algorithm to capture an image with small exposure time and ISO values in low-lighting condition. Three virtually differently exposed images are produced form the captured image and they are fused together to produce a brightened image. The algorithm in [9, 10] is based on the Gamma correction which introduces brightness distortion to the brightened image. A novel neural augmentation framework is proposed in [11] to overcome the problem. Three differently exposed images are first produced from a low-light image by using a physics-driven method, and are then refined by a data-driven method. They are finally fused together to produce the brightened image. Ren et al. [12] propose a trainable hybrid network, which uses an RNN-based network to describe the edge details. In [13], adversarial learning is introduced to capture visual properties beyond traditional metrics. Little work has been done in researching novel and suitable networks for low-light facial expression image enhancement. In this paper, we carefully design a LLIE network for low-light facial images enhancement. It is worth mentioning that we design the LLIE network from three perspectives: feature extraction, feature enhancement (attention mechanism and high-level feature and low-level feature fusion mechanism) and image reconstruction.

Facial expression recognition

In order to alleviate the dilemma of FER caused by occlusion situation, Xie et al. [14] present a Deep Attentive Multi-path CNN, which can locate the expression-related regions in the expression image and generate a robust image representation. To solve the class imbalance and improve the discrimination power of expression representations, Li et al. [15] introduce an adaReg loss to re-weight category importance coefficients. In addition, recent studies have shown that Vision Transformer can be applied to FER. Xue et al. [16] propse a TransFER model to learn rich relation-aware local representations. Zhang et al. [17] introduce a GAN-based model for joint facial expression recognition and image synthesis, significantly enhancing FER performance and flexibility in training. Although all of these methods achieved good results on the FER, they were all experimented on normally illuminated facial expression images, and the performance of these methods may suffer from varying degrees of attenuation when in low-light environments. In this paper, we adopt the idea of enhancing first and then recognizing for low-light FER, which effectively improves the accuracy of low-light FER, and it is worth noting that all these steps are implemented in a unified framework.

Combination of high-level vision and low-level vision

Haris et al. [18] explore the impact of image super-resolution (SR) on object detection tasks in low-resolution images. They incorporate the traditional detection loss into the training target in a new neural network training framework, and prove that this task-driven SR consistently and significantly improved the accuracy of the object detector on low-resolution images under various conditions. Tang et al. [19] propose SeAFusion to achieve real-time fusion of infrared and visible light images. A semantic loss is introduced to improve the facilitation of fusion results for high-level vision tasks. Furthermore, they propose a joint adaptive training strategy for low-level and high-level vision models. Yan et al. [20] design a semantic relation distillation loss (SRD-loss) and combined with task classification loss to jointly optimize the network for low-resolution fine-grained image classification. In summary, these methods effectively facilitate the evolution of both tasks by cascading the high-level vison task with the low-level vison task. Inspired by these works, we first attempt to combine LLIE technology with FER technology and propose the idea of using joint loss to train two cascaded networks, which effectively alleviates the dilemma of FER under low light conditions.

Our approach

We first introduce the architecture of LLENet in this section, and then demonstrate the proposed framework that cascades the LLENet model and the FER model and their specific joint training strategy in detail.

The proposed low-light image enhancement network

The architecture of the proposed LLENet is shown in Fig. 2, which mainly incorporates three module: downscaling module, upscaling module, and reconstruction module. Each module in this network will be detailed as follows.

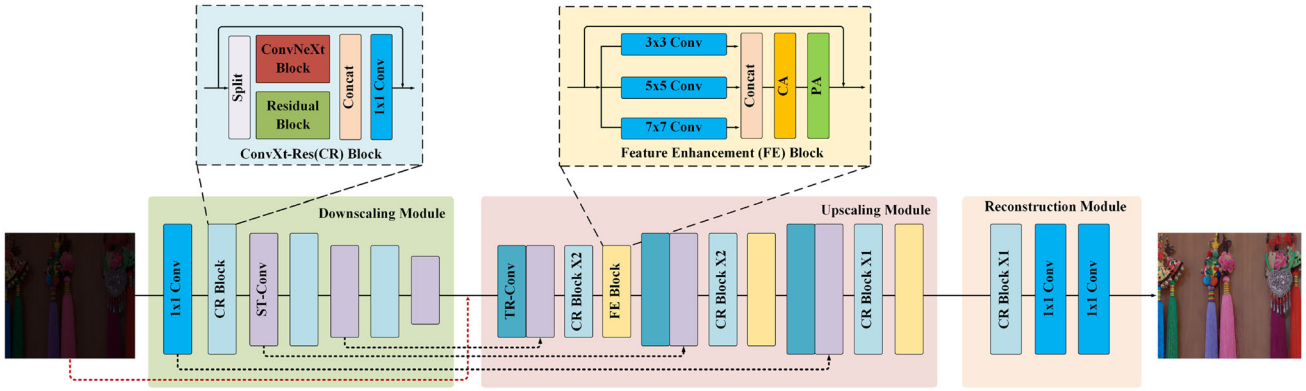

Fig. 2 The architecture of proposed LLENet. LLENet exploits the ConvNeXt-Residual(CR) block as the main building block. ST-conv and TR-Conv are 2×2 strided convolution with stride 2 and 2×2 transposed convolution with stride 2. PA and CA represent pixel attention and channel attention, respectively. The red dashed arrows indicate residual connections and the black dashed arrows refer to concatenate operations

Downscaling Module: Fig. 2 illustrates the architecture of downscaling module. The downscaling module mainly comprises ConvNeXt-Residual(CR) block, which integrates the core parts of ConvNetxt and residual network [7] to extract features.

In each CR block, the input is first split evenly into two feature map groups along the channel dimension, each of which is then fed into a ConvNeXt block and residual 3×3 convolutional block, respectively; after that, the outputs of ConvNeXt block and residual block are concatenated and then passed through a 1×1 convolution to produce the residual of the input. In a downscaling module, the input is first preprocessed a 1×1 convolution, and then is fed into the convXt-Residual (CR) block to extract features; the immediately following 2×2 strided convolution with stride 2 is used to implement feature downsampling, with the gradual increase of such operations, the original image can be deeply downsampled to obtain progressively concentrated features. We use three scales of downscaling operations for LLENet in our experiments.

As shown in the top-left subfigure of Fig. 2, a CR block fuses ConvNeXt block and conventional residual convolutional block through concatenation operation, with a residual connection to the input. Specifically, for an input feature map M, which is first divided equally into two features map M_1 and M_2. The process can be expressed as:

$$\{M_1, M_2\} = Split\,(M)\,. \tag{1}$$

Then, M_1 and M_2 are separately fed into a ConvNeXt block and Residual block to obtain Z_1 and Z_2:

$$\begin{cases} Z_1 = ConvXt\,(M_1)\,, \\ Z_2 = Residual\,(M_2)\,. \end{cases} \tag{2}$$

Finally, Z_1 and Z_2 are concatenated as the input of a 1×1 convolution, and followed by a residual connection with the input M. So, the final output of CR block is:

$$Out = Conv_{1\times1}\,(Concat\,(Z_1, Z_2)) + M. \tag{3}$$

Upscaling Module: The upscaling module is designed to restore the luminance information as well as the dimensional information from the downsampled features. In FER task, the texture information of expression image has positive effects on the classification of facial expressions. Moreover, according to our overall design, the input image of the FER task comes from the output of the LLIE network, so it is especially important to combine the underlying texture information of facial expression image in the LLIE network efficiently.

To deal with the above considerations, we carefully design the architecture of the upscaling module from the perspective of shallow feature and high-level feature fusion and feature enhancement. First, we integrate the intermediate results (shallow features) in the downscaling module into the intermediate feature map of the upscaling module by concatenation operations, so as to alleviate the problem that the texture and other information of shallow features cannot be retained during the upsampling process. Secondly, we enhance the representation ability of important features and weaken redundant features through multi-scale convolution operation and channel attention (CA) mechanism [21] and pixel attention (PA) mechanism [22]. This design can also reduce noise in low-light images to a certain extent.

Similar to the downscaling module, the upscaling module also includes the CR block, but a new feature enhancement (FE) block is introduced. In the upscaling module, the input is first upsampled by a 2×2 transposed convolution with stride 2, and then is fused with the intermediate-level features in the downsampling process. The fused features are fed into the CR block to extract features further, and the immediately fol-

lowing FE block is used to implement feature enhancement; with the gradual increase of such operations, the original downsampled features can be deeply upsampled to obtain larger-scale features. In the FE module, an input feature map F is firstly processed by three convolutions with different scales, the process can be expressed as:

$$C_i = Conv_{(2i+1)\times(2i+1)}(F), i = 1, 2, 3 \tag{4}$$

The final output of FE module is:

$$FE_{out} = PA\left(CA\left(Concat\left(C_1, C_2, C_3\right)\right)\right) + F, \tag{5}$$

where FE_{out} denote the output of the FE module, concat refer to concatenate opeartion. CA and PA are channel attention mechanism [21] and pixel attention mechanism [22], respectively.

Reconstruction Module: The reconstruction module is mainly designed to perform high quality image reconstruction work on the feature from the downscaling module. In the reconstruction module, the input is first passed through a CR block to obtain purified feature, and then the feature is fed into two 1×1 convolution operations to achieve the reconstruction of a three-channel high-light image.

Loss Function: Usually, the conventional approach is to use MAE and MSE loss functions to measure the error, however, these two loss functions only evaluate the degree of convergence at the pixel-to-pixel level, without taking into account the image structure information, the color space and other information that may affect the quality of the final generated image. In order to improve the image quality from both qualitative and quantitative aspects. We also design a new loss function by comprehensively considering the structure similarity, feature perception, color space difference, and frequency domain difference. It is expressed as:

$$\mathcal{L} = \omega_c\mathcal{L}_c + \omega_f\mathcal{L}_f + \omega_p\mathcal{L}_p + \omega_s\mathcal{L}_s, \tag{6}$$

where the $\mathcal{L}_c$, $\mathcal{L}_f$, $\mathcal{L}_p$ and $\mathcal{L}_s$ represent the loss functions of image color space difference, image frequency domain difference, feature perception and image structure similarity, and ω_c, ω_f, ω_p, ω_s are the corresponding coefficients. Details of the four loss functions are given below:

Color Space Difference Loss: We believe that the enhanced image and the normal-light (NL) image should be as consistent in the image space domain as possible. RGB image is the most familiar images in the color space, which is represented by three channels of an image, red (R), green (G) and blue (B), and can be easily stored and read by computer, but it is unfriendly to humans for color judgment, and RGB color space is a color space with poor uniformity. If the color similarity is measured directly by Euclidean distance, the result will have a large deviation from the human eye vision. There are two color spaces, HSV (Hue, Saturation, Value) and HSI (Hue, Saturation, Intensity), which are closer to people's perceptual experience of color than RGB. They intuitively to express the hue, vividness and lightness of colors, and conveniently make color comparisons. We use L1 error to measure the prediction error of an image on HSV and HSI color spaces, respectively, as shown below:

$$\mathcal{L}_c = \left\| F_E\left(LL\right)^{hsv} - NL^{hsv} \right\|^1 + \left\| F_E\left(LL\right)^{hsi} - NL^{hsi} \right\|^1 \tag{7}$$

where $F_E(\cdot)$ represent the LLIE network, LL is the input low-light image. $F_E\left(LL\right)^{hsv}$ and NL^{hsv} are the representation of enhanced and NL image in HSV color space, respectively. In addition, $F_E\left(LL\right)^{hsi}$ and NL^{hsi} denote the representation of enhanced and NL image in HSI color space, respectively.

Image Frequency Domain Difference Loss: We also propose a novel high-frequency component loss function to measure the error in the high-frequency component between the enhanced image. The wavelet transforms [23] has good localization characteristics in both time/frequency domains, which can refine the analysis of visual signals at different scales to extract the key parts from the information.

We use the difference of the high-frequency components in the horizontal and vertical direction between the enhanced image and the NL image as the image frequency domain loss. The specific expressions are as follows:

$$\mathcal{L}_f = \left\| F_E\left(I\right)^{hh} - N^{hh} \right\|^1, \tag{8}$$

where $F_E\left(I\right)^{hh}$ and N^{hh} refer to the high frequency component representation of the enhanced image and the NL image in horizontal and vertical directions, respectively.

Feature Perception Loss: To further evaluate the differences between images, we also incorporate perceptual loss to measure the feature similarity between the enhanced and NL images. Similar to [24]. The perceptual loss is defined as follows:

$$\mathcal{L}_p = \frac{1}{HWC}\sum_{i=1}^{H}\sum_{j=1}^{W}\sum_{k=1}^{C}\left\| V\left(F_E(LL)\right)_{i,j,k} - V(NL)_{i,j,k}\right\|^2, \tag{9}$$

where W, H, and C denote the three dimensions of an image respectively. $V(\cdot)$ is the pre-trained VGG network [25].

Image Structure Similarity Loss: The MAE and MSE loss functions average the differences between pixels and are not competent for structural distortion. Therefore, we adopt the SSIM [26] as the evaluation index of structural similarity,

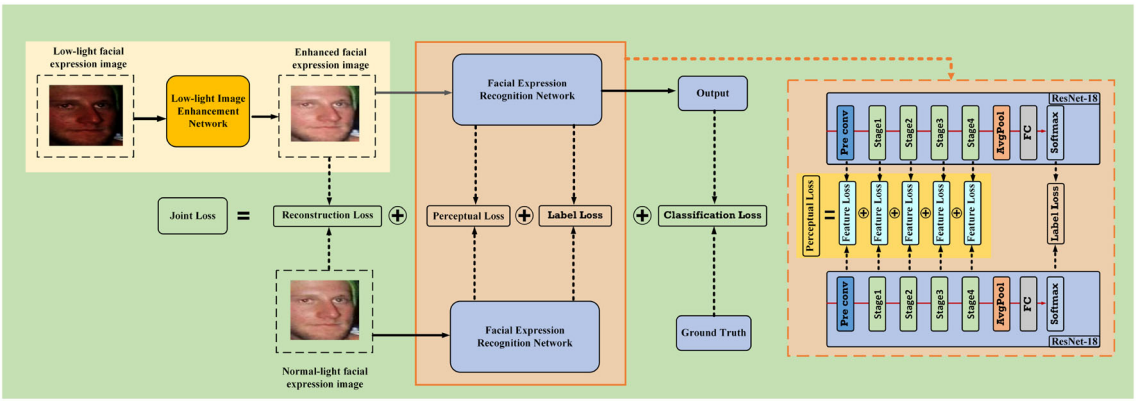

Fig. 3 The overview of our proposed cascaded architecture. The framework consists of a FER network and a LLIE network (LLENet), and the joint loss is used to train the framework

which is

$$\begin{cases} \text{SSIM}\,(e,n) = \dfrac{2\mu_e\mu_n + c_1}{\mu_e^2 + \mu_n^2 + c_1} \cdot \dfrac{2\sigma_{en} + c_2}{\sigma_e^2 + \sigma_n^2 + c_2}, \\ \mathcal{L}_s = 1 - \text{SSIM}(LL, NL), \end{cases} \tag{10}$$

where, n represent the images to be measured. The μ_e and μ_n refer to the mean values of the two images. The σ_e and σ_n denote the variances of the images. The c1 and c2 are two constants to prevent the denominator be zero ($c1 = 0.01^2$, $c2 = 0.03^2$).

The reconstruction loss L_{Rec} used to train the low-light enhancement model is defined as follows:

$$\mathcal{L}_{Rec} = \omega_1 L_1 + \omega_2 L_2 + \omega_3 \mathcal{L}, \tag{11}$$

where $\omega_1, \omega_2, \omega_3$ represent the weights for balancing the losses for L_1, L_2, $\mathcal{L}$, and the L_1 and L_2 are MAE and MSE loss functions respectively. In our experiments, we empirically set $\{\omega_1, \omega_2, \omega_3, \omega_c, \omega_f, \omega_p, \omega_s\} = \{0.1, 0.01, 1, 0.5, 0.5, 1, 1\}$.

FER guided low-light image enhancement

We propose a novel end-to-end cascade learning architecture (LL-FER) to process low-light facial expression images, via combining the LLENet and a FER network. We aim to simultaneously:

1) Reconstruct visually good facial expression images as the output of a low-light image enhancement model, guided by high-level information from the FER network.

2) Attain good accuracy in low-light FER task.

The overview of the designed LL-FER framework is displayed in Fig. 3. The LL-FER framework consists of a FER network (we use the ResNet-18 as the FER network) and a LLIE network (LLENet). The FER network can be any FER network. The cascade learning aims to allow the FER network to lead the restoration process of the LLENet, enabling it to reconstruct high-light images that are better suited to the FER network and improving the FER accuracy. Specifically, given a low-light input image, the LLIE network is first applied to enhance the input image, and the enhanced result is then fed into the following network for the FER network task, which outputs the specific facial expression category.

Joint Training Strategy: We first pre-train the FER network on normal-light dataset and freeze all trainable parameters, specifically, for the scheme, only the weights in the LLIE network are updated. In the joint training strategy, the trained-well FER network and the LLENet are just like supervisor and student, the FER network takes the role of a supervisor to guide the training of the LLENet, while the LLENet as a student utilizes the feedback information from the FER network and its efforts to optimize himself continuously and expects to achieve the goals required by the supervisor.

Joint Training Loss: The FER network is trained with a classification loss intended to improve the measure of its recognition accuracy. As shown in Fig. 3, we design a joint loss to train the LLENet in the proposed cascade framework, which integrates two main losses, one from the reconstruction loss originally used for the LLENet and the other from the high-level vision loss for the FER task: classification loss, feature perception loss, and label loss. We describe them in order as follows.

The reconstruction loss in the proposed LL-FER framework is calculated in the same way as in the LLENet, and it is obtained using Eq. 11. In addition to guide the LLIE network training by using the estimated error between the enhanced image and the NL image over the image domain, we also design FER-based perceptual loss $\mathcal{L}_{ferp}$ for the cascade framework.

$$\begin{cases} L_{fea_s} = L_1(E_s, N_s), \\ \mathcal{L}_{ferp} = \sum_{s=1}^{S} L_{fea_s}, \end{cases} \tag{12}$$

where the L_{fea_s} represents the feature loss of s-th stage in the Fig. 3, S is the number of stages in FER network. L_1 refers to the MAE loss function, E_s, N_s are features of the enhanced image and NL image passing through each stage of FER network, respectively.

In addition to the original classification loss as the final convergence target of the training framework, the label loss is introduced as an auxiliary loss to further reduce the difference in predicted label between the NL image and enhanced image. Both classification loss and label loss are cross-entropy losses, where classification loss is the error between the predicted label of the enhanced image and its ground truth label, and label loss is the error between the predicted label of the enhanced and the label generated by the NL image after passing through the FER network. The joint loss is defined as the weighted sum of the reconstruction loss, the perceptual loss and the loss for high-level vision task, which can be represented as

$$L_{Joint} = \mathcal{L}_{Res} + \mathcal{L}_{ferp} + \mathcal{L}_{Cls} + \mathcal{L}_{Label}, \tag{13}$$

where the $\mathcal{L}_{ferp}$, $\mathcal{L}_{Cls}$ and $\mathcal{L}_{Label}$ represent the FER-based perceptual loss, classification loss and label loss, respectively.

Low-light facial expression image synthesis

The low-light image synthesis method proposed by Lv et al. [27] is adopted to build a low-light facial expression dataset. Specifically, we adopt the combination of linear and gamma transformation to convert the NL facial expression image to an underexposed low-light image. The synthetic low-light facial expression image has low brightness and low contrast compared to the NL image. The low-light image synthesis method (without additional noise) can be formulated as:

$$I_{out}^{(i)} = \beta \times \left(\alpha \times I_{in}^{(i)}\right)^{\gamma}, \tag{14}$$

where i $\in \{R, G, B\}$, and $I_{in}^{(i)}$, $I_{out}^{(i)}$ denote the input image and the synthetic low-light image, respectively. α and β are linear transformations, the γ means the gamma transformation. The three parameters is sampled from uniform distribution: $\alpha \sim$ U (0.9, 1), $\beta \sim$ U (0.5, 1), $\gamma \sim$ U (1.5, 5).

Experiments

In this section, we first introduce the datasets used in our experiments, and then describe the implementation details of the experimental evaluation. Finally, we show the experiment results of our designed LLENet and joint training method on related datasets.

Dataset

1). *Low-light Image Datasets*

LOL The Low-Light (LOL) dataset [28] contains 500 images: 485 images for training and 15 images for testing. It is the first dataset for the evaluation of low-light enhancement for real scenes.

NPE The NPE dataset [29] consists of 46 images taken with the Cannon digital camera and 110 images downloaded from the websites, which is also a widely used dataset in the field of low-light image enhancement.

2). *Facial Expression Image Datasets*

RAF-DB The RAF-DB [30] is a large emotional face database with complexity and extensive annotations in real world, which includes two different categories: one for 7 classes of basic emotions, 12,271 training images, and 3,068 testing images, the other one subset for 11 classes of compound emotions. In our experiment, we use the 7 basic expressions (RAF-DB basic): Surprise, Fear, Disgust, Happiness, Sadness, Anger, Neutral.

FERPlus The FERPlus [31] provides 10 voting for each image, and also includes a way of tagging, the maximum voting method is used to remove some uncertain images. It consists of 35886 facial expression images, 28708 training images, 3589 public test images, and 3589 private test images with a total of 8 classes expressions. We use the majority voting method for experimental evaluation.

Implementation details

The proposed LLIE network takes RGB low-light images as input, we first use PSNR and SSIM [32] as the evaluation metric to measure the quality of the enhanced images, which are regularly used in other image generation tasks (e.g., image-resolution, image restoration), and then we provide visualization results for visual perception comparison. We conduct extensive experiments to compare our method with existing methods on LOL and NPE datasets. We set the size of random clipping patches to 48 × 48 and the batch size to 80. The Adam is used as the optimizer with a learning rate of 0.0003, and the model is trained for 5×10^3 iterations. As for joint training, we utilize the popular ResNet-18 network for FER tasks on two public FER datasets (RAF-DB, FERPlus), and the batch size is set to 32. The Adam is used as the optimizer with a learning rate of 0.0003, and the model is trained

Table 1 The influence of FE and HL in the proposed low-light enchancement method on LOL dataset

FE	HL	PSNR
×	×	20.91
×	✓	22.87
✓	×	21.55
✓	✓	**23.44**

Bold represents the best experimental result of the method proposed in this paper. FE is the feature enhancement block. HL represents fusing high-level features with low-level features in upscaling module network

Table 2 The influence of loss functions used in the proposed low-light enchancement method on LOL dataset

$\mathcal{L}_c$	$\mathcal{L}_f$	$\mathcal{L}_p$	$\mathcal{L}_s$	PSNR
×	×	×	×	21.37
✓	×	×	×	21.88
✓	✓	×	×	21.93
✓	✓	✓	×	22.54
✓	✓	✓	✓	**23.44**

Bold represents the best experimental result of the method proposed in this paper. The $\mathcal{L}_c, \mathcal{L}_f, \mathcal{L}_p$ and $\mathcal{L}_s$ represent the loss functions of image color space difference, image frequency domain difference, feature perception and image structure similarity

for 2×10^3 iterations. All experiments are conducted using Pytorch 1.7.1 on a Ubuntu 18.04 system with a i7-8700K CPU and a 2×32 G V100 GPU.

Low-light image enhancement

1). *Analysis of Low-light Image Enhancement Network*

To deeply analyze the proposed LLIE in this paper, several experiments will be conducted to verify the effectiveness of the method and the degree of contribution of the key parts in it. Specifically, these experiments consider the following factors: (1) the effectiveness of the combination of high-level features and low-level features in LLENet; (2) the effectiveness of the FE module in the upsampling module. As shown in Table 1, we perform evaluation experiments on the LOL dataset, it can be observed that the performance of the proposed LLIE obtains some improvement after applying the HL and FE strategies to it. Specifically, both strategies have the effect of promoting the LLENet in PSNR metrics. The effect of their combined use is more obvious than either strategy used alone, which can increase by about 2.5% relative to the case without the strategies. As a result, from the perspective of experiments, the proposed feature enhancement module and the high-level feature and low-level feature fusion strategy is beneficial to increase the performance of the proposed LLIE. This result also demonstrates that the proposed method's design intention is reasonable.

2). *Analysis of Losses used in Low-light Image Enhancement Network*

Since multiple losses ($\mathcal{L}_c$, $\mathcal{L}_f$, $\mathcal{L}_p$ and $\mathcal{L}_s$) are employed in the designed LLENet, several experiments are conducted to demonstrate their effectiveness and contribution, the model with MAE and MSE loss functions is used as the baseline model. As demonstrated in Table 2, as different loss functions are gradually added to the LLENet, the performance metric PSNR shows a trend of gradual improvement. The results suggest that using additional loss components improves the performance of the proposed LLENet. Among them, the image structure similarity loss function improves the most significantly.

Table 3 The comparison among recently advanced methods on the LOL dataset

Method	PSNR	SSIM
KinD [38]	20.87	0.8022
Zero-DCE [39]	14.86	0.5093
LIME [34]	16.76	0.5644
BIMEF [40]	13.88	0.5771
LightenNet [41]	10.30	0.3613
LLNet [42]	17.95	0.6819
Retinex-Net [43]	16.77	0.5594
EnlightenGAN [35]	17.44	0.6744
DLN [36]	21.95	0.8071
KinD++ [37]	21.30	0.8226
SNR-aware [44]	24.61	0.842
Uformer [45]	16.36	0.507
LAE-Net [46]	18.30	0.6393
RFLLE [47]	16.29	0.52
Ours	**23.44**	**0.8648**

Bold represents the best experimental result of the method proposed in this paper

3). *Comparisons with Existing Methods*

In this section, we compare both quantitative and qualitative aspects with the currently existing low-light enhancement methods, such as Retinex-Net [28], LIME [34], EnlightenGAN [35], DLN [36], Kind++ [37] and so on, these methods are highly competitive in the field of LLIE. The public code used is from the author's website and the default parameter settings provided are used for fair comparison.

The LOL and NPE are used as the evaluation datasets. Tables 3 and 4 display the numerical results of various approaches. From Table 3, it is evident that the proposed method is quantitatively most of the other competing meth-

ods with an average PSNR score of 23.443 dB and SSIM score of 0.865. The higher PSNR values indicate that the images enhanced by the proposed LLENet method have less error at pixel level and better color information recovery, the better SSIM values indicate that the proposed LLENet better retains the structural details. SNR-aware [44] focuses on optimizing PSNR/SSIM for denoising, boosting PSNR but not always aligned with perceptual quality or semantic feature preservation crucial for tasks like FER. Our LLENet adopts a balanced design, improving pixel quality (competitive PSNR) while prioritizing structural details and perceptual features for tasks like FER, integrating feature perception and structural similarity losses. This may lead to slightly lower PSNR but balances pixel accuracy with feature preservation for advanced tasks.

This suggests that the LLENet is better suited to human visual perception. We evaluated the NIQE score on NPE datasets to further validate the proposed method's generalizability, and the average value is shown in Table 4, the experimental results show that the LLENet is better than these competitors and has good generalization ability. The observation of quantitative results alone is insufficient to demonstrate the efficacy of the proposed method, it may be more convincing to evaluate it in terms of the visual effect of the enhanced images.

Figures 4 and 5 qualitatively show the visual comparison results between the LLENet and other excellent LLIE methods. For the results on LOL test set, the enhanced image frome the LIME, Zero-DCE, RUAS and EnlightenGAN are darker compared with other enhancer. Kind++ and RetinexNet perform poorly on brightness restoration and color preservation. In contrast, the proposed LLIE method achieves better perception effect. Specifically, the enhanced image generated by the LLENet not only restores the appropriate lighting but also preserves more details, and the restored color is closer to the NL image. For the results on low-light RAF-DB test set, only the facial expression image restored by DLN is acceptable in visual effect, while the rest of enhancement algorithms either have brightness distortion or produce redundant artifacts. Instead, the results recovered using the proposed enhancement algorithm are more competitive in terms of luminance recovery and detail recovery.

The proposed low-light image enhancement network (LLENet) demonstrates important progress in solving low-light image enhancement challenges, demonstrating considerable improvements over existing methods such as RetinexNet, LIME, and EnlightenGAN. By integrating high-level

Table 4 The comparison among recently advanced methods on the NPE dataset

Method	NIQE
BIMEF [40]	3.4975
CRM [48]	3.6800
LIME [34]	3.8422
NPE [29]	3.4455
KinD++ [37]	3.1466
LAE-Net [46]	3.7940
Ours	**3.0162**

Bold represents the best experimental result of the method proposed in this paper

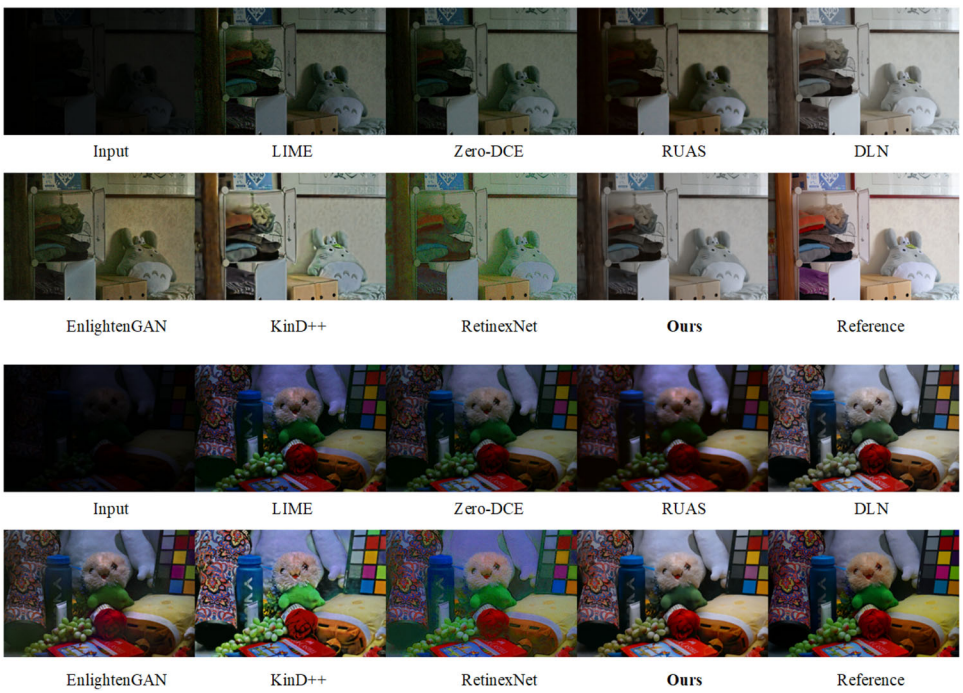

Fig. 4 The visual comparison with other famous low-light image enhancement methods on LOL dataset. The enhanced image generated by the proposed method has less noise and color distortion, and better visual effect

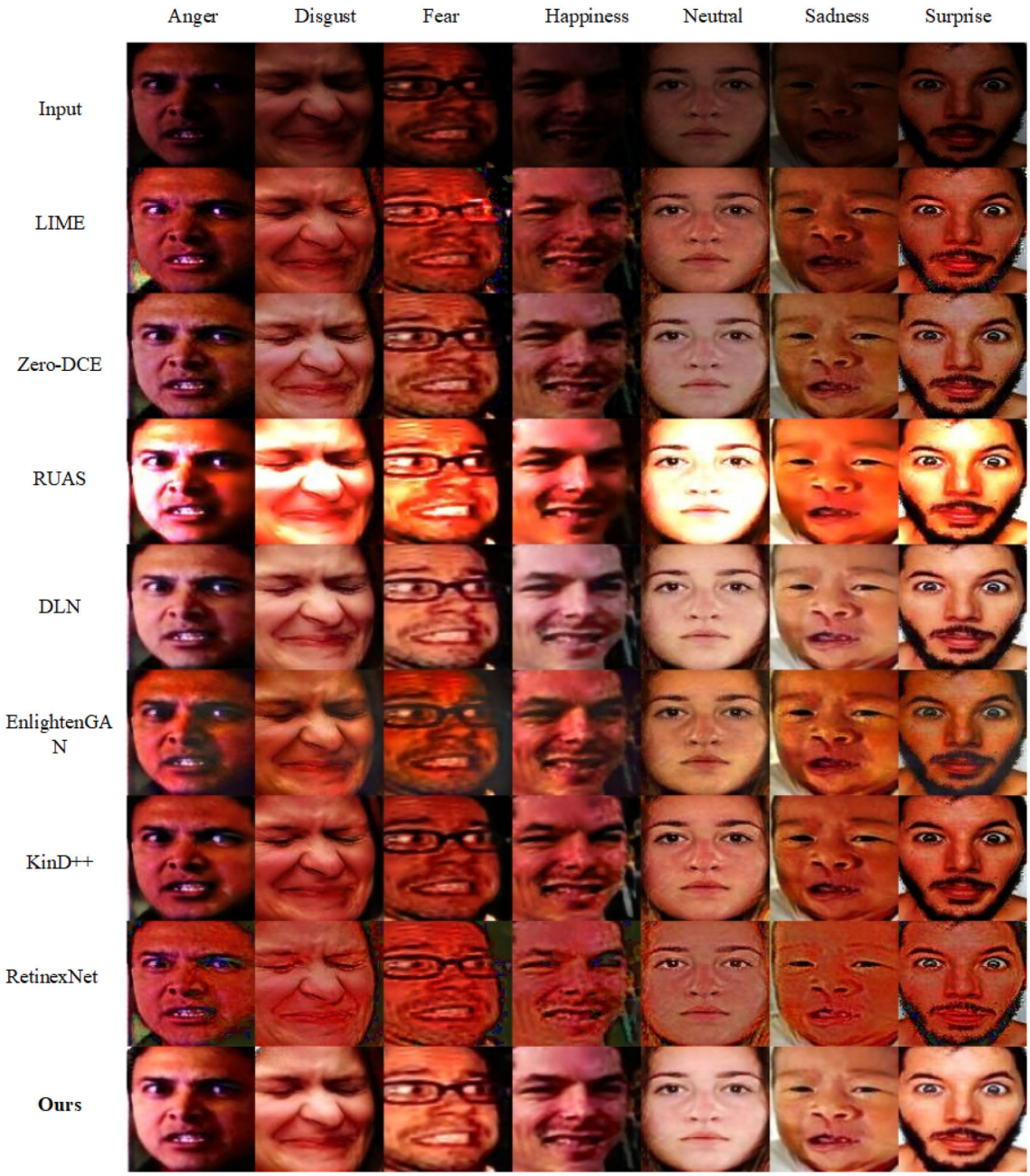

Fig. 5 The visual comparison with other excellent low-light image enhancement methods on the low-light version of RAF-DB dataset

and low-level features and introducing a novel feature enhancement (FE) module, LLENet effectively enhances image brightness while preserving key details and color fidelity. Our experimental results on the widely recognized LOL and NPE datasets show that LLENet achieves superior performance in both PSNR and SSIM metrics, confirming its efficacy in brightening pixels and structural integrity. The application of the composite loss function strategy further refines the enhancement process and ensures high-quality recovery of low-light images. Through quantitative evaluation and qualitative comparison, LLENet has been validated as an innovative approach that consistently produces better results in enhancing visual quality and realism.

When low-light image enhancement meets facial expression recognition

1). *FER Guided Low-light Image Enhancement*

Although PSNR and SSIM are widely used to evaluate the quality of recovered images, they mainly measure the

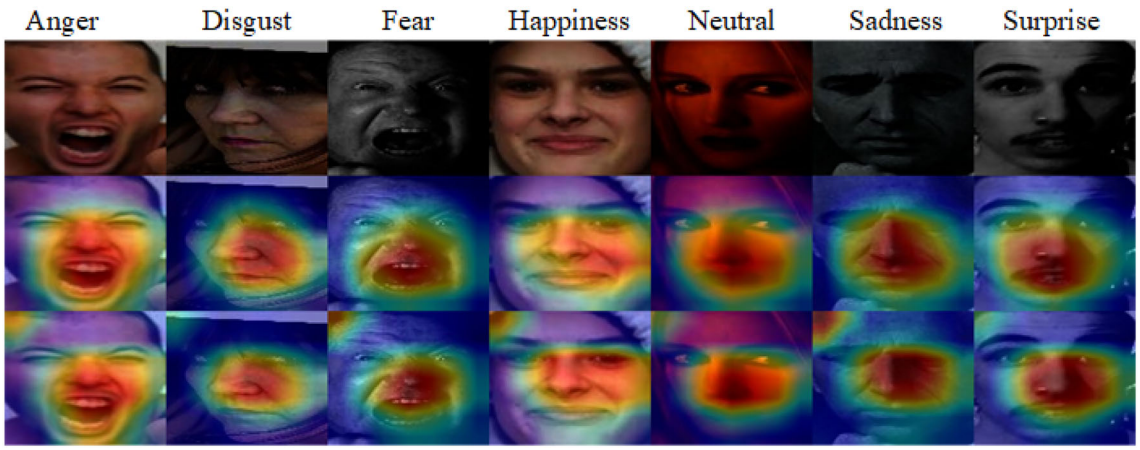

Fig. 6 The comparison of Grad CAM [33] visualization results with or without high-level vision loss. The last convolutional output is used to obtain the visualization result. The 1st row is the input low-light image, the 2nd row and 3rd row are the visualization results of applying only the reconstruction loss and the joint high-level vision loss, respectively

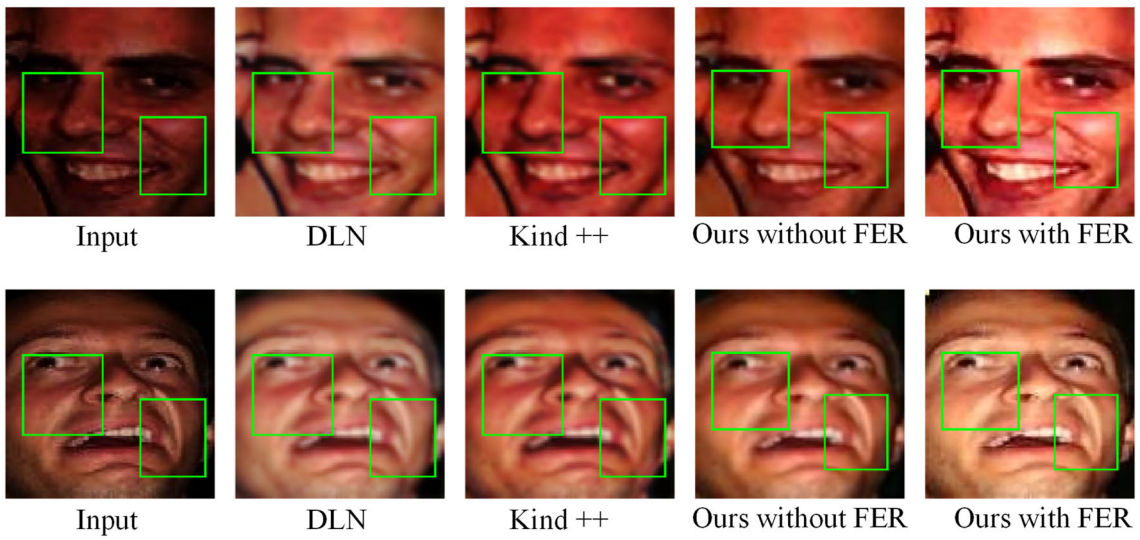

Fig. 7 Two examples of low-light facial expression image enhancement

difference between two images at the pixel level, which may always be somewhat different from the actual visual quality assessment. Therefore, we consider that future experiments will not pursue quantitative metric improvements, but rather explore how the LLIE method is affected by the FER task, and to what extent such effects can alter the performance of the FER model in low-light conditions.

In the LL-FER framework, we explore the mutual influence between LLIE and FER. The FER network is used as a high-level vision task to influence the restoration process of the designed LLENet, the loss used in the proposed framework is mainly composed of: 1) the image reconstruction losses (MAE loss, MSE loss and the newly designed loss) from the LLENet, 2) the high-level losses (perceptual loss, classification loss and label loss) from FER network. First, we investigate the impact of the loss component, by visualizing the Grad CAM [33] with relative to the input low-light expression image while utilizing (1) only the image reconstruction losses, (2) the image reconstruction losses and the high-level vision losses from the FER network. Some visualization results from the RAF-DB testset are shown in Fig. 6.

The second row in Fig. 6 is the Grad CAM visualization for only applying the reconstruction loss, and it can be seen that the network tends to respond to local regions of the facial muscles, indicating that these regions more or less contribute to FER. Instead, the third row of Fig. 6 shows the visualization when both high-level vision loss and image reconstruction loss are involved in the training. Its results are smaller and more accurate for the highlighted regions and have larger response values (darker colors) than those using only recon-

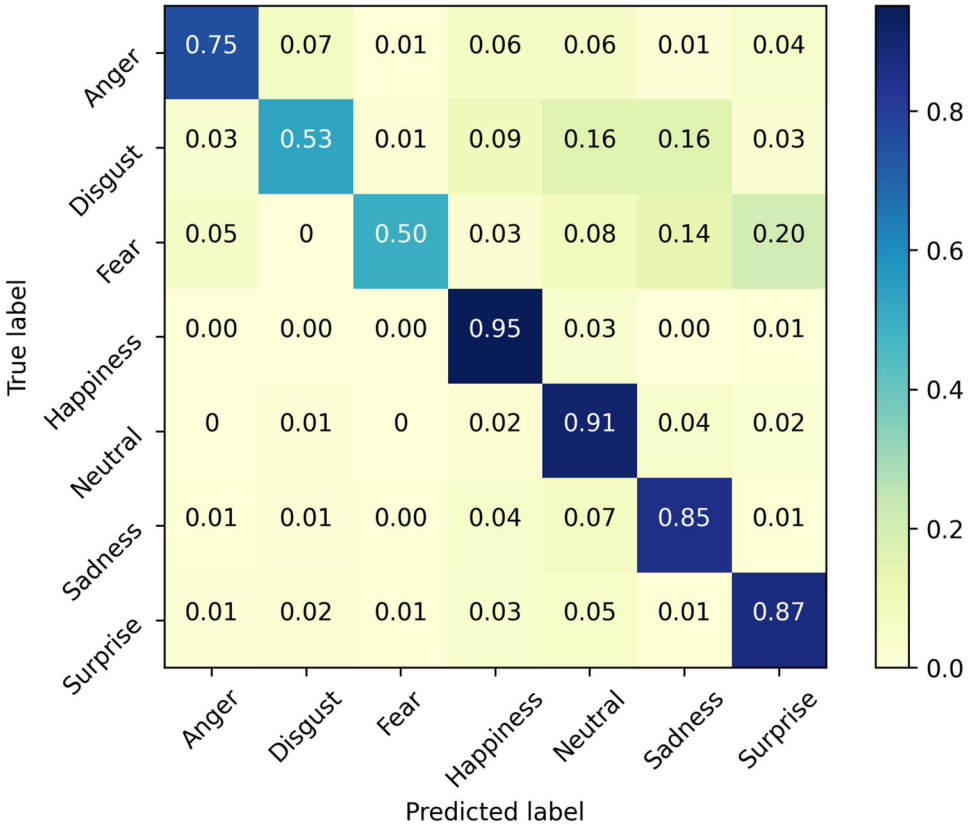

Fig. 8 The confusion matrix on the the low-light version of RAF-DB database for 7 facial expressions

Table 5 The total classification accuracy (%) comparison among recently advanced methods on the low-light versions of the RAF-DB, FERPlus datasets respectively

Method	RAF-DB	FERPlus
RAN [3]	82.43	82.31
SCN [49]	80.61	79.3
DMUE [6]	85.60	83.18
DACL [50]	83.7	82.49
KTN [51]	85.98	83.01
RUL [52]	86.18	82.92
EAC [53]	86.08	83.10
PACVT [54]	85.33	82.55
Ours	**87.48**	**85.47**

Bold represents the best experimental result of the method proposed in this paper

struction loss, and these regions often contain discriminative features useful for identifying expression categories. It can be concluded from this that the LLENet trained with FER guidance can recover certain important features that facilitate the FER task.

To further qualitatively evaluate the joint training strategy, we conduct several LLIE comparison experiments on the low-light version of the RAF-DB dataset. As shown in Fig. 7, the second and third columns are the enhanced results using DLN [36] and Kind++ [37], respectively; the fourth column is the proposed LLIE network trained separately without the guidance of FER network; the fifth column is the enhanced result using the proposed LLIE network trained jointly with a FER network. Overall, the results obtained by DLN and Kind++, and the proposed low-light enhancer are not very clear, and all suffer from blurring, which is due to the original low-light input images not being of high enough quality. However, the proposed method is slightly competitive in terms of naturalness and sharpness when observed carefully from the local area, especially the results recovered by the enhancer after FER guidance are more natural and realistic.

We present the quantitative performance comparison results in Table 5 for RAF-DB and FERPlus, respectively. In order to form a more convincing comparison, we introduce the latest research advances of FER. We utilize the code open-sourced from these studies to conduct experiments on low-light facial expression recognition. Notably, we encounter instances where the open-sourced code from certain methods is either flawed or unavailable. In these cases, we have undertaken improvements and reproductions to a extent, ensuring the robustness and validity of our comparative analysis. The proposed method achieves 87.48% and 85.47% recognition accuracy on the low-light versions of the RAF-DB and FERPlus datasets, respectively, and these

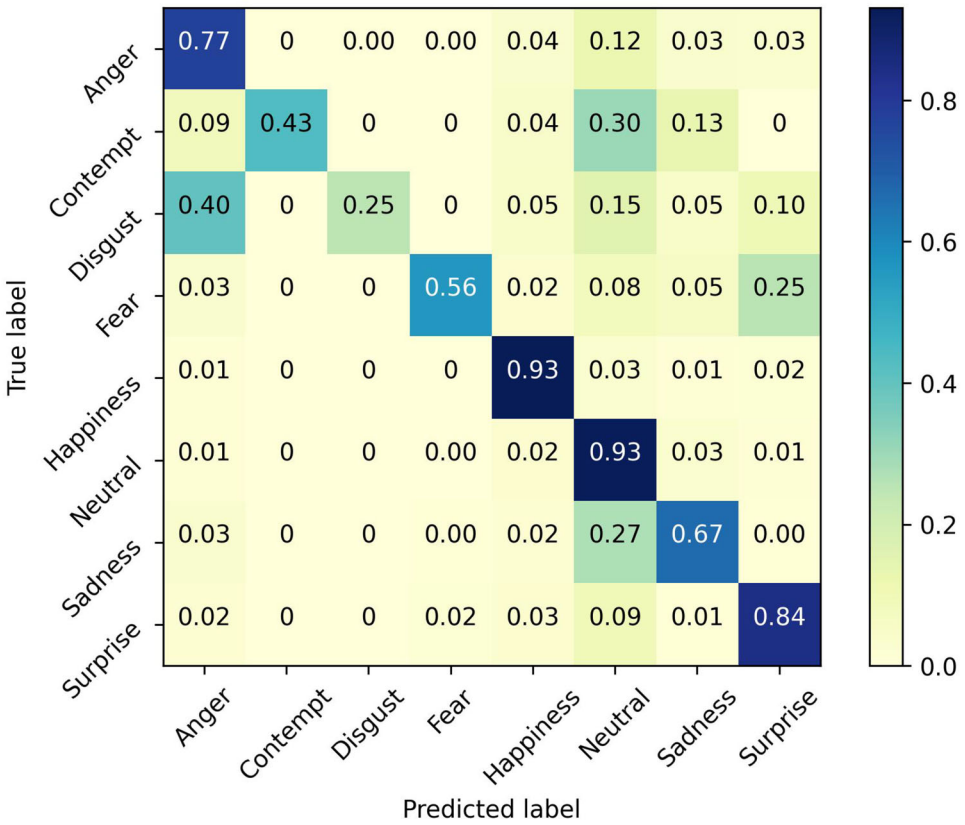

Fig. 9 The confusion matrix on the the low-light version of FERPlus database for 8 facial expressions

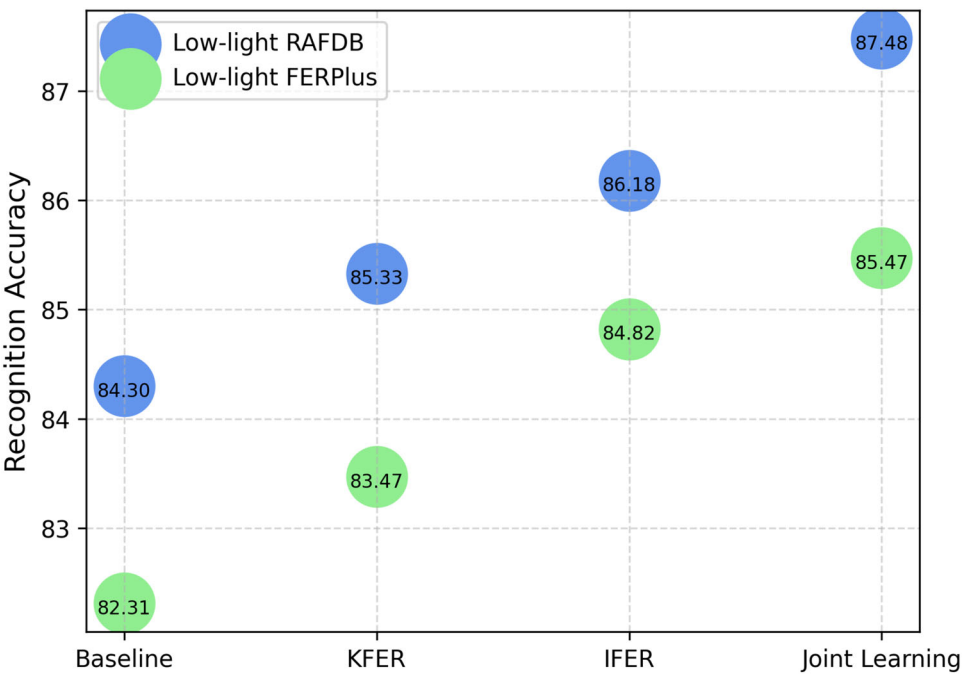

Fig. 10 The total classification accuracy (%) of different training schemes on the low-light versions of RAF-DB and FERPlus datasets, respectively

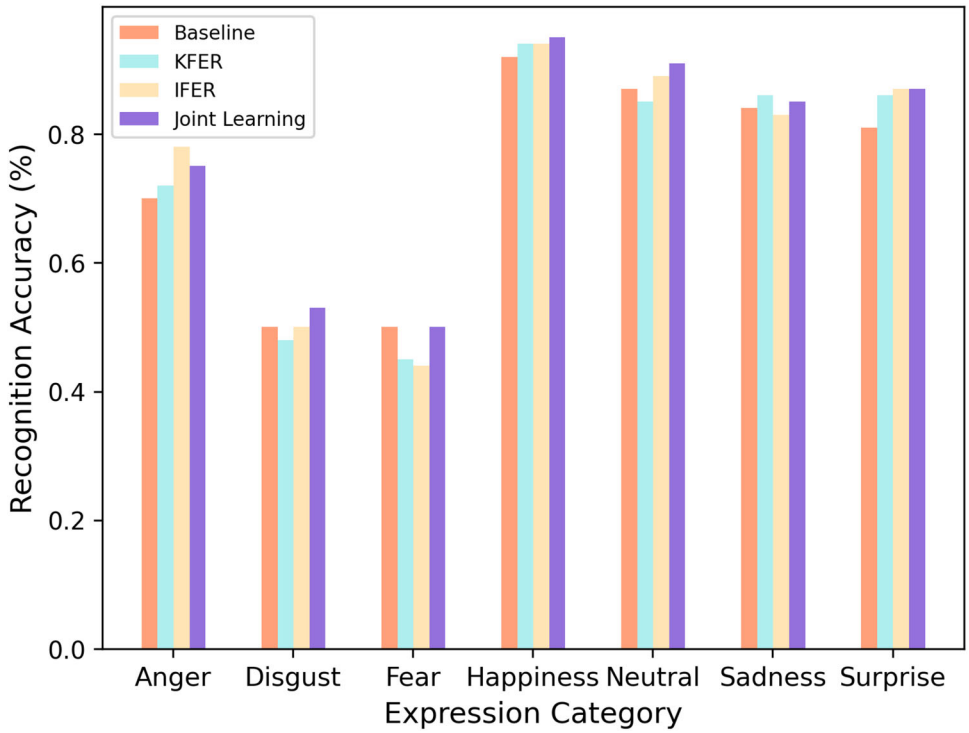

Fig. 11 The classification accuracy of various training schemes for each category of facial expressions on the low-light version of RAF-DB test set

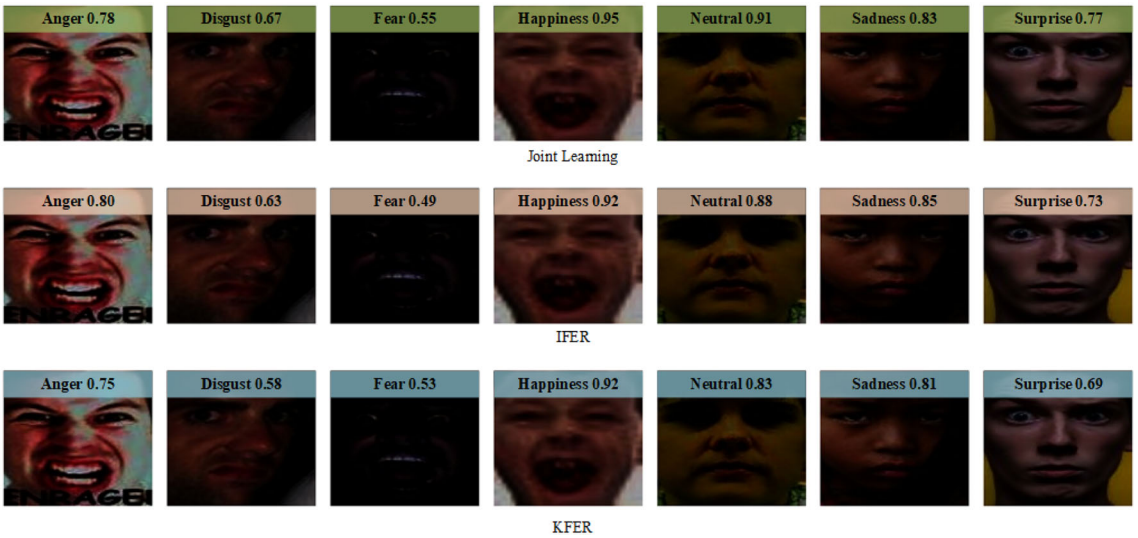

Fig. 12 The facial expression recognition scores from the proposed method, KFER and IFER (From top to bottom)

results are better than the existing methods. These results clearly demonstrate that the suggested method is competitive with state-of-the-art FER methods. It also shows that the proposed method may yield more satisfactory results when used in a practical low-light environment for FER challenges. In addition, we also present confusion matrices for the low-light versions of the RAF-DB and FERPlus datasets, as shown in Figs. 8 and 9, where it can be observed that the proposed method is very good at recognizing the facial expressions "Happiness", "Neutral" and "Surprise" in low-light environments, which are usually recognized with an accuracy higher than 84%. And the method maintains relatively acceptable performance for other facial expression categories with insufficient sample data (e.g., "Fear" and "Contempt").

2). *Comparison of Training Schemes*

We now explore how the LLIE task interacts with the FER task on FER dataet. To evaluate the impact of different low-light enhancement schemes on the performance of the FER task. We configure the following experiments:

1. Low-light expression images are directly fed into the FER network. This scheme is used as the baseline for low-light FER.
2. Low-light expression images are first enhanced by Kind++, and then fed into the FER network. This scheme is named Kind++ with FER (KFER).
3. The low-light expression images are first trained independently by the proposed LLIE network, and then the enhanced images are used as the input of the FER network to perform FER task. The scheme is named independent with FER (IFER).
4. Proposed Method: Low-light expression images are sequentially passed through the proposed LLIE network and the FER network, and the two networks are trained with joint loss. The scheme is named joint learning.

Figure 10 shows the classification accuracy of FER on various training schemes, we can clearly observe the impact of different training schemes on FER recognition performance. It is easy to find that the recognition performance of the baseline method on all datasets is at a disadvantageous position compared to the other methods. This illustrates that LLIE has potential to be employed as a pre-processing step for low-light FER, which can alleviate the problem of lower accuracy of FER caused by low-light environments to some extent. In both KFER and KFER experiments, the accuracy of FER is improved compared to the baseline model. However, their improvement is limited due to the fact that only LLIE methods are considered, but no high-level vision information is introduced; in addition, the performance of the LLIE network is not high enough, which also leads to the inability to achieve better improvement. The proposed joint learning method achieves superior accuracy to both the baseline and the independently trained methods on different FER datasets.

From Fig. 11, it can be observed that the FER network under the joint learning framework is able to achieve better performance on the low-light test dataset compared to the independently trained approach. From this, we can conclude that the joint learning approach has a positive impact on the recoverability of the LLIE network, which helps the LLIE network recover certain discriminative features required for the FER task, and this leads to the FER network being able to extract more discriminative features from the enhanced expression images more easily. On the contrary, the performance of the independently trained approach is somewhat limited.

Finally, we present some qualitative FER results in Fig. 12, which shows that the proposed joint training method performs better on some expressions and obtains larger probability scores compared to the results of other methods. Since the other methods are not trained under the guidance of the FER task, they cannot extract discriminative features directly. Instead, the proposed joint training approach guided the training of the LLIE network using the FER, which allows the FER network to learn more expression knowledge from the enhanced expression image input.

Conclusion

It is of great reference significance to explore the cross-effect between low-level image processing and high-level vision tasks for the practical application of various vision tasks. In this paper, we took an initiative to explore the relationship between the LLIE and FER task, and proposed a joint learning framework to simultaneously handle low-light expression image enhancement and FER tasks. The framework enables both tasks to achieve noticeable improvements by introducing high-level vision loss from the FER task to the LLIE network and processing both tasks in a cascade manner. Extensive qualitative and quantitative comparative experiments reveal that the proposed joint training framework has good advantages in facilitating the FER task.

Acknowledgements This research is supported by the Natural Science Foundation of Sichuan Province with Grant ID 2025ZNSFSC1498, the Sichuan Provincial Department of Science and Technology Project with Grant ID 24YFHZ0317, and the Project of Chengdu Science and Technology Bureau with Grant ID 2024-YF09-00041-SN.

Data Availability Data derived from public domain resources.

Declarations

Conflict of interest There is no competition for interests.

References

1. Li S, Deng W (2020) Deep facial expression recognition: a survey. IEEE Trans Affect Comput 1–1
2. Li Y, Zeng J, Shan S, Chen X (2018) Occlusion aware facial expression recognition using cnn with attention mechanism. IEEE Trans Image Process 28(5):2439–2450

3. Wang K, Peng X, Yang J, Meng D, Qiao Y (2020) Region attention networks for pose and occlusion robust facial expression recognition. IEEE Trans Image Process 29:4057–4069
4. Li Y, Gao Y, Chen B, Zhang Z, Lu G, Zhang D (2021) Self-supervised exclusive-inclusive interactive learning for multi-label facial expression recognition in the wild. IEEE Trans Circuits Syst Video Technol 32(5):3190–3202
5. Zhao Z, Liu Q, Zhou F (2021), Robust lightweight facial expression recognition network with label distribution training. In: Proceedings of the AAAI conference on artificial intelligence 35, 3510–3519
6. She J, Hu Y, Shi H, Wang J, Shen Q, Mei T (2021), Dive into ambiguity: latent distribution mining and pairwise uncertainty estimation for facial expression recognition. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 6248–6257
7. He K, Zhang X, Ren S, Sun J (2016), Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 770–778
8. Liu Z, Mao H, Wu C.-Y, Feichtenhofer C, Darrell T, Xie S (2022), A convnet for the 2020s. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 11976–11986
9. Li Z, Zheng J (2016), Single image brightening via exposure fusion. In: 2016 IEEE International conference on acoustics, speech and signal processing (ICASSP). IEEE, 1756–1760
10. Li Z, Wei Z, Wen C, Zheng J (2017) Detail-enhanced multi-scale exposure fusion. IEEE Trans Image Process 26(3):1243–1252
11. Zheng C, Li Z, Yang Y, Wu S (2021) Single image brightening via multi-scale exposure fusion with hybrid learning. IEEE Trans Circuits Syst Video Technol 31(4):1425–1435
12. Ren W, Liu S, Ma L, Xu Q, Xu X, Cao X, Du J, Yang M-H (2019) Low-light image enhancement via a deep hybrid network. IEEE Trans Image Process 28(9):4364–4375
13. Kim G, Kwon D, Kwon J (2019), Low-lightgan: low-light enhancement via advanced generative adversarial network with task-driven training. In: 2019 IEEE International conference on image processing (ICIP). IEEE, 2811–2815
14. Xie S, Hu H, Wu Y (2019) Deep multi-path convolutional neural network joint with salient region attention for facial expression recognition. Pattern Recogn 92:177–191
15. Li H, Wang N, Ding X, Yang X, Gao X (2021) Adaptively learning facial expression representation via cf labels and distillation. IEEE Trans Image Process 30:2016–2028
16. Xue F, Wang Q, Guo G (2021), Transfer: learning relation-aware facial expression representations with transformers. In: Proceedings of the IEEE/CVF International conference on computer vision, 3601–3610
17. Zhang X, Zhang F, Xu C (2021) Joint expression synthesis and representation learning for facial expression recognition. IEEE Trans Circuits Syst Video Technol 32(3):1681–1695
18. Haris M, Shakhnarovich G, Ukita N (2021) Task-driven super resolution: Object detection in low-resolution images. In: International conference on neural information processing. Springer, New York, 387–395
19. Tang L, Yuan J, Ma J (2022) Image fusion in the loop of high-level vision tasks: a semantic-aware real-time infrared and visible image fusion network. Inf Fusion 82:28–42
20. Yan T, Shi J, Li H, Luo Z, Wang Z (2022) Discriminative information restoration and extraction for weakly supervised low-resolution fine-grained image recognition. Pattern Recogn 127:108629
21. Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 7132–7141
22. Zhao H, Kong X, He J, Qiao Y, Dong C (2020) Efficient image super-resolution using pixel attention. In: European conference on computer vision. Springer, New York 56–72
23. Huang H, He R, Sun Z, Tan T (2017) Wavelet-srnet: a wavelet-based cnn for multi-scale face super resolution. In: Proceedings of the IEEE international conference on computer vision, 1689–1697
24. Qian R, Tan R. T, Yang W, Su J, Liu J (2018) Attentive generative adversarial network for raindrop removal from a single image. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2482–2491
25. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556
26. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612
27. Lv F, Li Y, Lu F (2021) Attention guided low-light image enhancement with a large scale low-light simulation dataset. Int J Comput Vis 129(7):2175–2193
28. Wei C, Wang W, Yang W, Liu J (2018) Deep retinex decomposition for low-light enhancement. arXiv preprint arXiv:1808.04560
29. Wang S, Zheng J, Hu H-M, Li B (2013) Naturalness preserved enhancement algorithm for non-uniform illumination images. IEEE Trans Image Process 22(9):3538–3548
30. Li S, Deng W, Du J (2017) Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2852–2861
31. Barsoum E, Zhang C, Ferrer CC, Zhang Z (2016) Training deep networks for facial expression recognition with crowd-sourced label distribution. In: Proceedings of the 18th ACM International conference on multimodal interaction, , 279–283
32. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612
33. Selvaraju R. R, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2017) Grad-cam: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International conference on computer vision (ICCV)
34. Guo X, Li Y, Ling H (2016) Lime: low-light image enhancement via illumination map estimation. IEEE Trans Image Process 26(2):982–993
35. Jiang Y, Gong X, Liu D, Cheng Y, Fang C, Shen X, Yang J, Zhou P, Wang Z (2021) Enlightengan: deep light enhancement without paired supervision. IEEE Trans Image Process 30:2340–2349
36. Wang L-W, Liu Z-S, Siu W-C, Lun DP (2020) Lightening network for low-light image enhancement. IEEE Trans Image Process 29:7984–7996
37. Zhang Y, Guo X, Ma J, Liu W, Zhang J (2021) Beyond brightening low-light images. Int J Comput Vis 129(4):1013–1037
38. Zhang Y, Zhang J, Guo X (2019), Kindling the darkness: a practical low-light image enhancer, in: Proceedings of the 27th ACM international conference on multimedia, ,1632–1640
39. Guo C, Li C, Guo J, Loy C. C, Hou J, Kwong S, Cong R (2020) Zero-reference deep curve estimation for low-light image enhancement. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 1780–1789
40. Ying Z, Li G, Gao W (2017) A bio-inspired multi-exposure fusion framework for low-light image enhancement. arXiv preprint arXiv:1711.00591
41. Li C, Guo J, Porikli F, Pang Y (2018) Lightennet: a convolutional neural network for weakly illuminated image enhancement. Pattern Recogn Lett 104:15–22
42. Lore KG, Akintayo A, Sarkar S (2017) Llnet: a deep autoencoder approach to natural low-light image enhancement. Pattern Recogn 61:650–662

43. Wei C, Wang W, Yang W, Liu J (2018) Deep retinex decomposition for low-light enhancement. arXiv preprint arXiv:1808.04560
44. Xu X, Wang R, Fu C-W, Jia J (2022) Snr-aware low-light image enhancement. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 17714–17724
45. Wang Z, Cun X, Bao J, Zhou W, Liu J, Li H (2022) Uformer: a general u-shaped transformer for image restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 17683–17693
46. Liu X, Ma W, Ma X, Wang J (2023) Lae-net: a locally-adaptive embedding network for low-light image enhancement. Pattern Recogn 133:109039
47. Yang X, Gong J, Wu L, Yang Z, Shi Y, Nie F (2023) Reference-free low-light image enhancement by associating hierarchical wavelet representations. Expert Syst Appl 213:118920
48. Ying Z, Li G, Ren Y, Wang R, Wang W (2017) A new low-light image enhancement algorithm using camera response model. In: Proceedings of the IEEE international conference on computer vision workshops, 3015–3022
49. Wang K, Peng X, Yang J, Lu S, Qiao Y (2020) Suppressing uncertainties for large-scale facial expression recognition. IEEE/CVF Conf Comput Vis Pattern Recognit (CVPR) 2020:6896–6905
50. Farzaneh A. H, Qi X (2021) Facial expression recognition in the wild via deep attentive center loss. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision, 2402–2411
51. Li H, Wang N, Ding X, Yang X, Gao X (2021) Adaptively learning facial expression representation via cf labels and distillation. IEEE Trans Image Process 30:2016–2028
52. Zhang Y, Wang C, Deng W (2021) Relative uncertainty learning for facial expression recognition. Adv Neural Inf Process Syst 34:17616–17627
53. Zhang Y, Wang C, Ling X, Deng W (2022) Learn from all: Erasing attention consistency for noisy label facial expression recognition. In: European conference on computer vision. Springer, New York, 418–434
54. Liu C, Hirota K, Dai Y (2023) Patch attention convolutional vision transformer for facial expression recognition with occlusion. Inf Sci 619:781–794