# Low-Light Face Super-resolution via Illumination, Structure, and Texture Associated Representation

**Chenyang Wang, Junjun Jiang*, Kui Jiang, Xianming Liu**

School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China
{wangchy02, jiangjunjun, jiangkui, csxm}@hit.edu.cn

## Abstract

Human face captured at night or in dimly lit environments has become a common practice, accompanied by complex low-light and low-resolution degradations. However, the existing face super-resolution (FSR) technologies and derived cascaded schemes are inadequate to recover credible textures. In this paper, we propose a novel approach that decomposes the restoration task into face structural fidelity maintaining and texture consistency learning. The former aims to enhance the quality of face images while improving the structural fidelity, while the latter focuses on eliminating perturbations and artifacts caused by low-light degradation and reconstruction. Based on this, we develop a novel low-light low-resolution face super-resolution framework. Our method consists of two steps: an illumination correction face super-resolution network (IC-FSRNet) for lighting the face and recovering the structural information, and a detail enhancement model (DENet) for improving facial details, thus making them more visually appealing and easier to analyze. As the relighted regions could provide complementary information to boost face super-resolution and vice versa, we introduce the mutual learning to harness the informative components from relighted regions and reconstruction, and achieve the iterative refinement. In addition, DENet equipped with diffusion probabilistic model is built to further improve face image visual quality. Experiments demonstrate that the proposed joint optimization framework achieves significant improvements in reconstruction quality and perceptual quality over existing two-stage sequential solutions. Code is available at https://github.com/wcy-cs/IC-FSRDENet.

## Introduction

Face super-resolution (FSR) aims to recover the fine-grained and high-resolution (HR) details from a low-resolution (LR) observed face image. It has been widely studied and applied to many outdoor computer vision systems. Although prior methods are effective in processing face images captured under normal light conditions, there is still scope for improvement when it comes to low-light conditions. In practice, due to the diverse and complex imaging environment (*i.e.*, inadequate lighting at night or the limited exposure time in the video surveillance scenario), the captured face images would suffer from both low-resolution and low-light degradations. Thus, there is the pressing need to devise technologies to recover LR face images under low-light scenarios.

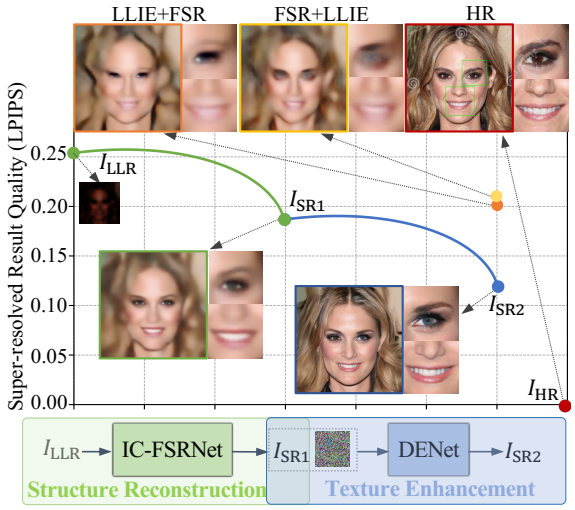


Figure 1: Super-resolved results of different methods. The bottom row presents the overall framework. IC-FSRNet is capable of restoring facial structure, but may not be able to recreate intricate details. DENet can further enhance facial details and elevate visual quality.

Prior to deep learning, conventional methods improve face resolution by designing hand-crafted priors and reconstruction operators, but struggle with the generalization on open-world scenarios. Recently, deep learning has emerged and shown impressive performance compared to conventional methods (Jiang et al. 2023) on FSR task. However, these methods are far from generating visually pleasing and content-fidelity results from the low-light low-resolution (LLR) face images since they are customized for recovering normal light LR face images. One possible approach is to perform face super-resolution and low-light image enhancement (LLIE) sequentially. This can be done either as FSR followed by LLIE (FSR+LLIE) or LLIE followed by FSR (LLIE+FSR). However, the simple cascaded solutions struggle with generating results with satisfactory facial structure and visual quality as shown in Figure 1. The reason lies in

*Corresponding author.

that the individual approach only focuses on single degradation while the input suffers from coupled degradations. That is, the former process would inevitably generate artifacts and errors, which deteriorates the latter process and causes error accumulation, ultimately arising the loss of important facial information. Thus, there is a pressing need for a method that can effectively recover low-light low-resolution face images.

A natural idea is to design a joint low-light low-resolution face super-resolution framework. As shown in Figure 1, the simple cascaded schemes (LLIE+FSR or FSR+LLIE) are deficient in capturing intricate facial features, leading to unsatisfied super-resolved results with obvious distortion and structural loss. To address this dilemma, we propose to decompose this hard task into structural fidelity maintaining and texture consistency learning. The former is tailored for enhancing the quality of LLR faces while maintaining structural fidelity, while the latter focuses on eliminating perturbations and artifacts caused by low-light degradation and reconstruction. Specifically, our method consists of two steps: an illumination correction face super-resolution network (IC-FSRNet) for capturing the face characteristic and recovering structural information, and a detail enhancement network (DENet) for improving facial details and visual quality. Given that the improved features of LLIE and FSR are complementary, whereby the intermediate results of LLIE can enhance FSR performance, and vice versa. In view of this, IC-FSRNet is built with an illumination estimation (IE) branch for estimating illumination adjustment coefficients, and a face super-resolution (FSR) branch that incorporates these coefficients to jointly perform super-resolution and illumination adjustment. More precisely, IC-FSRNet employs an iterative manner to enhance face reconstruction by fully exploring the mutual reinforcement between super-resolution and illumination adjustment. After maintaining structural fidelity, we take the merit of diffusion models in synthesizing fine image details and develop a diffusion probabilistic model-based detail enhancement network (DENet) to remove artifacts and improve the visual quality. We highlight the contributions as follows:

- We divide the dark FSR task into structural fidelity maintaining and texture consistency learning, and develop an illumination correction face super-resolution network (IC-FSRNet) and a detail enhancement model (DENet).
- We devise a joint learning scheme to boost the dark FSR, where the complementary information between the illumination correction and face super-resolution is encoded to refine each other.
- Experimental results demonstrate that the proposed method achieves state-of-the-art performance in terms of visual quality and quantitative metrics.

## Related Work

### Face Super-resolution

Prior to deep learning, researchers mainly take advantage of structure similarity (Baker and Kanade 2000), neighbour embedding (Chang, Yeung, and Xiong 2004) and other traditional techniques to improve face image quality. However, the limited representation ability makes them incapable of FSR task. More recently, deep learning techniques have become increasingly prevalent and researchers have developed various convolutional neural networks (CNN) to map LR face images to their corresponding HR counterparts (Chen et al. 2021; Lu et al. 2021; Wang et al. 2023a). Considering that the utilization of deep learning technologies in a straightforward manner may not be sufficient to accurately restore face images, researchers propose to introduce geometric prior (Chen et al. 2018; Ma et al. 2020), reference prior (Li et al. 2018, 2020a,b) and generative prior (Chan et al. 2021; Wang et al. 2021; Wang, Hu, and Zhang 2022; Wang et al. 2023c) to help recover high-quality face images. Since the receptive field of convolutional neural layers is limited while transformer can achieve a global receptive field, some FSR methods (Bao et al. 2023; Gao et al. 2023) propose to combine the transformer and CNN to capture and aggregate global and local information. Although notable performance improvement has been achieved, these methods are ineffective for recovering LR face images captured in low-light environments, as they do not consider the degradation caused by such conditions.

### Low-Light Image Enhancement

In recent years, significant progress has been made in the development of learning-based methods for low-light image enhancement. These methods can be broadly categorized into two frameworks: end-to-end framework and Retinex-based framework. The former focuses on designing networks that directly map low-light images to images with normal illumination. For example, the work of (Wang et al. 2022) proposes a flow-based network that normalizes image illumination to improve visibility. In light of the limited receptive field of CNN, LLformer (Wang et al. 2023b) develops a transformer-based LLIE methods to achieve global receptive field. Similarly, SFormer (Xu et al. 2022) not only adopts transformer structure, but also utilizes the concept of signal-to-noise ratio to achieve spatially varying enhancement and capture regional differences. Instead of using transformer, FECNet (Huang et al. 2022a) utilizes Fourier transform to capture global receptive field and explores frequency information to enhance low-light images. On the other hand, the latter relies on the Retinex theory that an image can be decomposed into reflectance and illumination. The work of (Wu et al. 2022) alternately recovers reflectance and illumination by unfolding network. Despite promising advances, they mainly enhance low-light images in high-resolution space. When applied to low-light LR face images, besides partly relighting the dark regions, they fail to recover visually pleasing faces with credible details.

### Denoising Diffusion Probabilistic Model

Denoising diffusion probabilistic model (DDPM) (Ho, Jain, and Abbeel 2020) is a popular and powerful generative model which predefines a variance schedule $\{\beta_t\}_{t=1}^{T}$ to incrementally corrupt an image, denoted as $\boldsymbol{y}_0$, into a noisy state through a forward diffusion process,

$$q\left(\boldsymbol{y}_t \mid \boldsymbol{y}_{t-1}\right)=\mathcal{N}\left(\boldsymbol{y}_t ; \sqrt{1-\beta_t} \boldsymbol{y}_{t-1}, \beta_t \boldsymbol{I}\right). \quad (1)$$

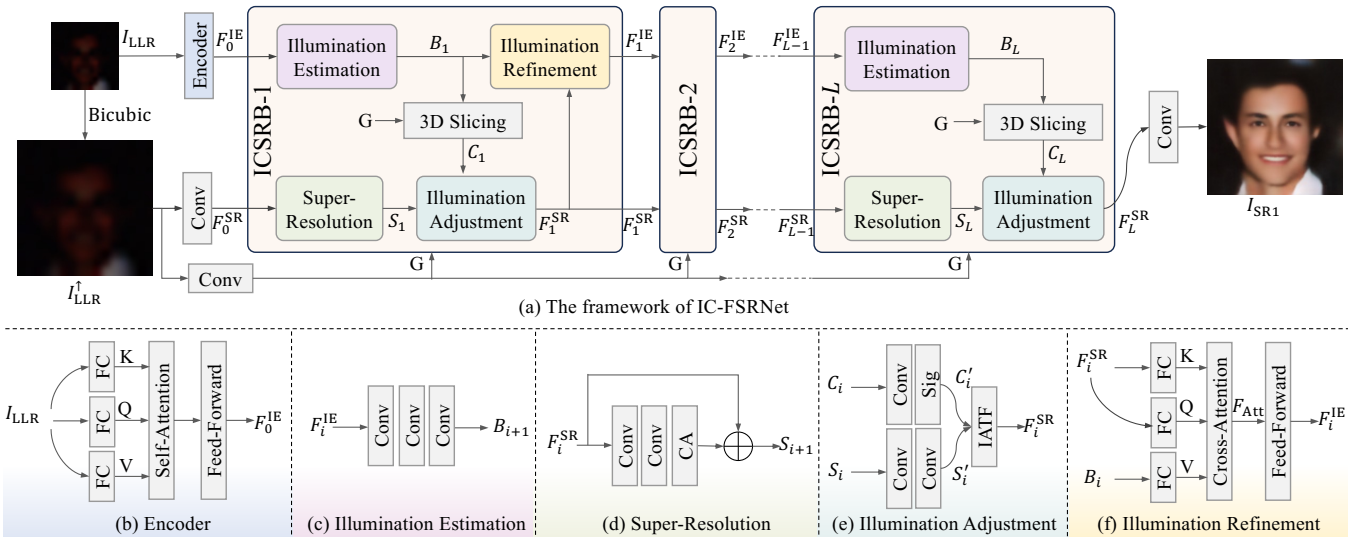


Figure 2: Overview of IC-FSRNet that consists of an illumination estimation branch (the top branch) and a super-resolution branch (the bottom branch). ICSRB is the illumination correction super-resolution block, $\boldsymbol{G}$ is guidance map extracted from $I^{\uparrow}_{\mathrm{LLR}}$, $\boldsymbol{B}$ is illumination grid, $\boldsymbol{C}$ denotes the illumination coefficient, CA is channel attention and IATF represents the illumination adjustment transform function.

By reversing the forward process, the generative process can be represented with learned mean $\boldsymbol{\mu}_\theta$ and fixed variance $\sigma_t$,

$$p_\theta\left(\boldsymbol{y}_{t-1} \mid \boldsymbol{y}_t\right)=\mathcal{N}\left(\boldsymbol{y}_{t-1} ; \boldsymbol{\mu}_\theta\left(\boldsymbol{y}_t, t\right), \sigma_t^2 \boldsymbol{I}\right). \quad (2)$$

In this study, we take the merit of the DDPM to remove artifacts and improve visual quality.

## Approach

In this paper, we aim to generate the high-quality super-resolved face images from the given low-resolution face images $\boldsymbol{I}_{\mathrm{LLR}}$ in low-light environment. Existing methods commonly focus on individually single task, resolution amplification or illumination correction, but barely consider the coupled degradation. Consequently, they are unable to handle this intricate task effectively. Although the cascaded solution (*i.e.*, FSR followed by LLIE (FSR+LLIE), or LLIE followed by FSR (LLIE+FSR)) is introduced to recover these degraded face images, limited improvement is gained by disregarding the intrinsic relations between these two tasks. As illustrated in Figure 1, face images produced by the cascaded solution lack facial structure and details, while with obvious color deviation. To mitigate this problem, we propose a novel approach that decomposes this task into structural fidelity maintaining and texture consistency learning. The former is tailored for enhancing the quality of $\boldsymbol{I}_{\mathrm{LLR}}$ faces while maintaining structural fidelity. The latter focuses on eliminating perturbations and artifacts caused by low-light degradation and reconstruction. As illustrated in Figure 1, our proposed model consists of two steps: an illumination correction face super-resolution network (IC-FSRNet) for capturing face characteristic and recovering structural information, and a detail enhancement network (DENet) for improving facial details and visual quality, dubbed as IC-FSRDENet. To be specific, we first feed the $\boldsymbol{I}_{\mathrm{LLR}}$ face image into IC-FSRNet for illumination adjustment while generating a coarse super-resolved face image. Considering that the improved features of LLIE and FSR are complementary, super-resolved face images could improve illumination adjustment and the relighted regions could provide complementary information to boost face super-resolution. In light of this, IC-FSRNet is constructed with an illumination estimation (IE) branch for estimating illumination adjustment coefficients, and a face super-resolution (FSR) branch that incorporates these coefficients to jointly perform super-resolution and illumination adjustment. In addition, IC-FSRNet iteratively performs illumination estimation and face super-resolution to explore the mutual reinforcement between them for boosting face super-resolution. Although IC-FSRNet can recover facial structure well and finish illumination adjustment, the hallucinated face images always lack high-frequency details and have noise caused by low-light degradation and reconstruction. Thus, we further develop DENet based on a diffusion model and feed the coarse results into it to improve facial details and visual quality.

### IC-FSRNet

Here we elaborate on our illumination correction face super-resolution network (IC-FSRNet). IC-FSRNet is a two-branch network as shown in Figure 2, including an illumination estimation (IE) branch to estimate illumination coefficient for illumination adjustment, and a face super-resolution (FSR) branch to reconstruct the super-resolution results. Considering that the improved features of LLIE and FSR are complementary, super-resolved face images could improve illumination adjustment and the relighted regions could provide complementary information to boost face super-resolution, the illumination coefficient estimation

and super-resolution are performed iteratively to explore the mutual relation between them for refinement.

In detail, a low-light low-resolution face image $\boldsymbol{I}_{\text{LLR}}$ is first fed into two branches. In light of that the illumination information is global in $\boldsymbol{I}_{\text{LLR}}$ face images, the encoder in IE branch adopts a transformer block (Liang et al. 2021) to encode it into $\boldsymbol{F}_0^{\text{IE}}$ to capture global response. FSR branch is responsible for increasing the resolution and learning the representation in a high-resolution space. Thus, we feed $\boldsymbol{I}_{\text{LLR}}^{\uparrow}$ upsampled by Bicubic interpolation into FSR branch and adopt a convolutional layer to extract feature $\boldsymbol{F}_0^{\text{SR}}$.

Then, an illumination correction super-resolution block (ICSRB) takes the output features as inputs to encode their mutual relation. ICSRB first performs face super-resolution and illumination estimation in parallel. The former is finished by the commonly used RCAB (Zhang et al. 2018b) to generate the enhanced feature $\boldsymbol{S}_1$. The latter is implemented by bilateral gird learning (Chen, Paris, and Durand 2007) due to its efficiency. Thanks to the bilateral gird learning, the illumination coefficient gird is estimated in low-resolution dimension and then sliced into high-resolution space with guidance map, which is more efficient. To be specific, three cascaded convolutional layers are adopted to predict bilateral illumination grid $\boldsymbol{B}_1$ from $\boldsymbol{F}_0^{\text{IE}}$. At the same time, guidance map $\boldsymbol{G}$ is extracted from $\boldsymbol{I}_{\text{LLR}}^{\uparrow}$ by a convolutional layer. With $\boldsymbol{B}_1$ and $\boldsymbol{G}$, illumination coefficient $\boldsymbol{C}_1$ can be obtained by 3D slicing, depicted as

$$\boldsymbol{C}_1 = f_{\text{Slice}}(\boldsymbol{B}_1, \boldsymbol{G}), \tag{3}$$

where $f_{\text{Slice}}$ denotes 3D slicing operation. 3D slicing first projects the illumination gird $\boldsymbol{B}_1$ with $\boldsymbol{G}$ onto a 3D grid whose first two dimensions represent 2D position in image plane and the third dimension denotes image intensity of $\boldsymbol{G}$. Then it blurs the grid with a Gaussian blur. Finally, based on the blurred bilateral grid and the guidance image $\boldsymbol{G}$, we can obtain the illumination coefficient $\boldsymbol{C}_1$ by accessing the grid value using trilinear interpolation.

Then the illumination coefficient $\boldsymbol{C}_1$ is packed into illumination adjustment block to adjust illumination of features $\boldsymbol{S}_1$ in the FSR branch, producing the adjusted result $\boldsymbol{F}_1^{\text{SR}}$. In light of that more accurate illumination coefficient can boost face super-resolution and higher resolution face images can promote illumination estimation, ICSRB is equipped with mutual reinforcement between illumination estimation and face super-resolution in an iterative manner. Specifically, in addition to utilizing illumination coefficient from IE branch to improve illumination in FSR branch, the result of illumination adjustment $\boldsymbol{F}_1^{\text{SR}}$ is fed back to the illumination refinement block of IE branch to refine the original illumination and improve the next illumination estimation. In this way, the illumination estimation is facilitated by the face super-resolution. Then, the generated super-resolution result and illumination refined result are fed into the following $L-1$ ICSRBs for more accurate mutual refinement.

After $L$ collaborations between illumination estimation and face super-resolution, a convolutional layer is applied on the result of the FSR branch to generate the reconstructed result $\boldsymbol{I}_{\text{SR1}}$. To optimize the network, $l_1$ pixel loss is adopted,

$$\mathcal{L}_{\text{IC-FSRNet}} = \left\| \boldsymbol{I}_{\text{SR1}} - \boldsymbol{I}_{\text{HR}} \right\|_1, \tag{4}$$

where $\boldsymbol{I}_{\text{HR}}$ is the ground truth face image.

**Illumination Adjustment.** Inspired by Zero-DCE (Guo et al. 2020), we resort to a transform function with the learnable illumination coefficient for illumination adjustment. To be specific, the learned coefficient $\boldsymbol{C}_i$ is first transformed into $[-1, 1]$ by a convolution layer followed by a Sigmoid function, obtaining the illumination coefficient $\boldsymbol{C}_i^{'}$. Towards FSR features, it would be fed into cascaded convolutional layers to generate $\boldsymbol{S}_i^{'}$, and then performed illumination adjustment by the following transform function:

$$\boldsymbol{F}_i^{\text{SR}} = \boldsymbol{S}_i^{'} + \boldsymbol{C}_i^{'} * \boldsymbol{S}_i^{'} * (1 - \boldsymbol{S}_i^{'}), \tag{5}$$

where $\boldsymbol{F}_i^{\text{SR}}$ is the adjusted feature via the illumination coefficient $\boldsymbol{C}_i^{'}$, and $*$ denotes the pixel-wise multiplication.

**Illumination Refinement.** In light that face super-resolution and illumination estimation can facilitate each other, we feed the super-resolved results of FSR branch back to the IE branch to refine the illumination estimation. Considering the global characteristics in illumination information, the cross-attention mechanism is introduced to explore the global informative components from the super-resolved result. Specifically, given the super-resolved result $\boldsymbol{F}_i^{\text{SR}}$ and original illumination $\boldsymbol{B}_i$, we use $\boldsymbol{B}_i$ to generate query $Q$ and use $\boldsymbol{F}_i^{\text{SR}}$ to obtain key $\boldsymbol{K}$ and value $\boldsymbol{V}$ by different full connection layers, then calculate cross-attention between them

$$\boldsymbol{F}_{\text{Att}} = f_{\text{Softmax}}(\boldsymbol{Q}\boldsymbol{K}^{\boldsymbol{T}}/\sqrt{d})\boldsymbol{V}, \tag{6}$$

where $\boldsymbol{F}_{\text{Att}}$ is attention map. Then a feed forward network is applied to generate the refined results. In this way, the super-resolved features can be utilized to refine the next illumination estimation.

## DENet

Despite IC-FSRNet can relight the dark regions and infer the facial structure, the faces hallucinated by IC-FSRNet lack facial details and with obvious artifacts, failing to provide pleasing visual experience. To this end, a detail enhancement network (DENet) is built to promote the visual quality. Inspired by the powerful generative ability of diffusion probabilistic model (DDPM) (Kawar et al. 2022), we develop DENet based on conditional DDPM with $\boldsymbol{I}_{\text{SR1}}$ as additional side information to help reverse the diffusion process. To be specific, the DENet is trained with the following objective:

$$E_{(\boldsymbol{x},\boldsymbol{y})} E_{\epsilon,\gamma} \left\| f_{\text{DENet}}(\boldsymbol{x}, \sqrt{\gamma}\boldsymbol{y}_0 + \sqrt{1-\gamma}\epsilon, \gamma) - \epsilon \right\|_1, \tag{7}$$

where $\epsilon \sim \mathcal{N}(0, I)$, $(\boldsymbol{x}, \boldsymbol{y})$ corresponds to the sample of $\boldsymbol{I}_{\text{SR1}}$ and $\boldsymbol{I}_{\text{HR}}$, $\gamma \sim p(\gamma)$, $f_{\text{DENet}}$ denotes the DENet. Thanks to the DENet, the proposed method can recover more visually pleasing face images with more detailed features.

# Experiments

In order to investigate low-light low-resolution (LLR) face super-resolution and verify the effectiveness of the proposed method, a LLR-HR face image dataset is essential. However, public synthetic dataset for LLR face super-resolution is unavailable and establishing real LLR-HR face image pairs is

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***Face Super-Resolution → Low-Light Image Enhancement*** | | | | | | | | | | | | | | | |
| | FECNet | | | | | LLformer | | | | | LEDNet | | | | |
| | PSNR | SSIM | NIQE | LPIPS | Param | PSNR | SSIM | NIQE | LPIPS | Param | PSNR | SSIM | NIQE | LPIPS | Param |
| SISN | 18.07 | 0.5532 | 11.703 | 0.2156 | 18.0M | 18.64 | 0.5674 | 10.717 | 0.2082 | 32.2M | 18.51 | 0.5651 | 12.635 | 0.2139 | 15.9M |
| SCTANet | 17.16 | 0.5218 | 12.452 | 0.2293 | 35.8M | 17.57 | 0.5240 | 11.412 | 0.2112 | 50.1M | 17.78 | 0.5409 | 14.012 | 0.2357 | 33.7M |
| SFMNet | 18.23 | 0.5650 | 11.593 | 0.2144 | 17.6M | 18.54 | 0.5688 | 10.757 | 0.2022 | 31.8M | 18.60 | 0.5749 | 12.746 | 0.2152 | 15.5M |
| ***Low-Light Image Enhancement → Face Super-Resolution*** | | | | | | | | | | | | | | | |
| | SISN | | | | | SCTANet | | | | | SFMNet | | | | |
| | PSNR | SSIM | NIQE | LPIPS | Param | PSNR | SSIM | NIQE | LPIPS | Param | PSNR | SSIM | NIQE | LPIPS | Param |
| FECNet | 17.57 | 0.5439 | 11.026 | 0.2188 | 18.0M | 17.11 | 0.4346 | 21.955 | 0.2338 | 35.8M | 17.13 | 0.5248 | 11.553 | 0.2153 | 17.6M |
| LLformer | 17.35 | 0.5337 | 10.836 | 0.2203 | 32.2M | 17.52 | 0.4543 | 18.016 | 0.2308 | 50.1M | 16.88 | 0.5128 | 10.439 | 0.2094 | 31.8M |
| LEDNet | 17.81 | 0.5431 | 10.897 | 0.2192 | 15.9M | 17.42 | 0.4638 | 19.047 | 0.2360 | 33.7M | 17.30 | 0.5261 | 11.710 | 0.2151 | 15.5M |
| IC-FSRNet (Ours) | PSNR = **21.68** | | SSIM = **0.6349** | | NIQE = 10.585 | | LPIPS = 0.1990 | | Param = 8.5M | | | | | | |
| IC-FSRDENet (Ours) | PSNR = 20.53 | | SSIM = 0.5920 | | NIQE = **6.129** | | LPIPS = **0.1267** | | Param = 18.8M | | | | | | |

Table 1: Comparisons with the state-of-the-art methods in terms of PSNR (↑), SSIM (↑), NIQE (↓), LPIPS (↓) and parameters (Param) on the CelebAMaskHQ dataset (Yu et al. 2018). The best results are in bold and the second best results are underlined.

(a) (b) (c) (d) (e) (f) (g) (h) (i) (j)

Figure 3: Visual quality comparison. (a) LLR; (b) SCTANet+LLformer; (c) LLformer+SCTANet; (d) SISN+FECNet; (e) FECNet+SISN; (f) SFMNet+LEDNet; (g) LEDNet+SFMNet; (h) IC-FSRNet; (i) IC-FSRDENet; (j) HR.

laborious and challenging. Therefore, we propose to simulate the degradation and synthesize LLR and normal light HR face image pairs based on the existing face datasets.

## Low-Light Low-Resolution Face Image Simulation

The key component of the degradation simulation includes illumination adjustment and noise addition. In the following, we elaborate on the degradation simulation.

**Illumination Adjustment**. Illumination adjustment aims to transform a normal image to the low-light image. Instead of directly using gamma correction, we utilize the combination of linear and gamma transformation to perform illumination adjustment, which can better approximate low-light images. To be specific, given a normal face image $\boldsymbol{I}_{\text{HR}}$, illumination adjustment can be formulated as:

$$\boldsymbol{I}_{\text{LL}} = \beta \times (\alpha \times \boldsymbol{I}_{\text{HR}})^{\gamma}, \tag{8}$$

where $\beta \in U(0.5, 1)$, $\alpha \in U(0.9, 1)$, $\gamma \in U(1.5, 5)$ and $I_{\text{LL}}$ is the low-light face image.

**Noise Addition**. In addition to changes in illumination, low-light environment also introduces noise. Thus, we turn to adding noise to the low-light face images. Instead of directly adding noise to face images, in-camera image signal processing (ISP) is also considered to simulate real low-light noise. ISP refers to the processing steps performed by the camera's hardware and software to convert the raw sensor data into a final image. When simulating low-light noise, it is important to consider the characteristics and effects of ISP. Thus, the noise addition process can be formulated as

$$\boldsymbol{I}_{\text{LLN}} = f_{\text{ISP}}(\boldsymbol{N}_{\text{P}}(f_{\text{ISP}}^{-1}(\boldsymbol{I}_{\text{LL}})) + \boldsymbol{N}_{\text{G}}), \tag{9}$$

where $f_{\text{ISP}}$ and $f_{\text{ISP}}^{-1}$ denote the ISP function and the inverse ISP function, $\boldsymbol{N}_{\text{P}}$ and $\boldsymbol{N}_{\text{G}}$ correspond to the Poisson noise and Gaussian noise, $\boldsymbol{I}_{\text{LLN}}$ is the generated noisy low-light face image. The $f_{\text{ISP}}$ is comprised of white balance, demosaicing, CAM-to-XYZ, XYZ-to-RGB and tone mapping.

After that, we downsample the noisy low-light face images with Bicubic interpolation by $\times 16$ to generate the final LLR face images, obtaining LLR-HR face image pairs.

## Datasets and Metrics

We conduct experiments on commonly-used high quality face dataset CelebAMask-HQ (Yu et al. 2018). From it,

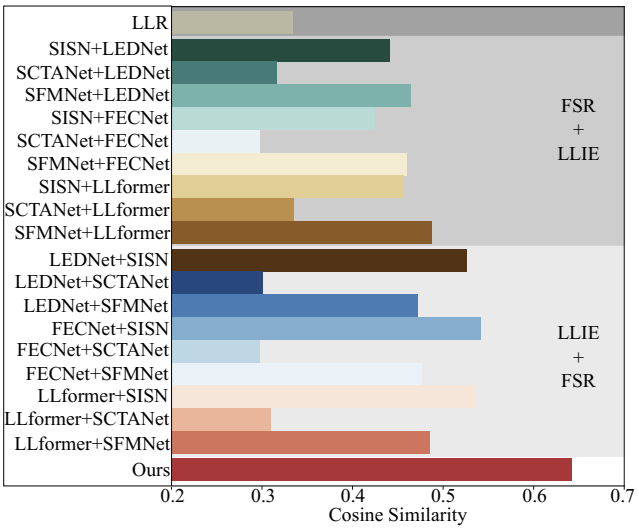


Figure 4: Cosine similarity comparison of state-of-the-art methods on CelebAMaskHQ (Yu et al. 2018) dataset.

we randomly choose 3050 face images for training, and 600 for validation, and another 300 for testing. To evaluate the model performance, four popular image quality evaluation metrics are chosen, including peak signal-to-noise ratio (PSNR), structural similarity (SSIM) (Wang et al. 2004), learned perceptual image patch similarity (LPIPS) (Zhang et al. 2018a) and natural image quality evaluator (NIQE) (Mittal, Soundararajan, and Bovik 2013).

## Implementation Details

To obtain ground truth face images, the original face images are directly resized into 256×256. Then, we apply the low-light low-resolution degradation simulation on the ground truth to generate LLR face images. In IC-FSRNet, $L$ is set as 14. IC-FSRNet and DENet are trained successively and individually. The optimizer is Adam with $\beta_1$=0.9, $\beta_2$=0.99, and $\epsilon$=1e-8. The learning rate is 1e-4 for both two networks. Towards the DENet, the backbone is UNet and other settings follow (Saharia et al. 2021). The experiments are implemented on PyTorch with one NVIDIA 3090 GPU.

## Comparison with the State-of-the-Arts

To verify the effectiveness of our proposed method, some state-of-the-art methods which are designed for FSR or LLIE, are selected and combined to form cascaded solution for low-light low-resolution face super-resolution. To be specific, the FSR methods include SISN (Lu et al. 2021), SCTANet (Bao et al. 2023) and SFMNet (Wang et al. 2023a), and the LLIE methods include FECNet (Huang et al. 2022b), LEDNet (Zhou, Li, and Change Loy 2022) and LLformer (Wang et al. 2023b). With the combination of these methods, 18 comparison methods in total can be obtained.

**Quantitative Comparison.** We have conducted a comparative analysis of our method with existing approaches, and the results are summarized in Table 1. Specifically, the upper part of Table 1 presents the results in the case that LLIE method is performed after FSR method, while the middle part lists the results in the case that LLIE method is performed before FSR method. The bottom part showcases the performance of our proposed methods. From the comparison in Table 1, it is evident that the simple cascaded solutions struggle to achieve satisfactory performance, whereas our joint method significantly outperforms the cascaded solutions. In particular, compared with the cascaded solutions, IC-FSRNet demonstrates the best performance across three evaluation metrics except NIQE, particularly excelling in PSNR and SSIM. It achieves a remarkable 3.04 dB improvement in PSNR and 0.0600 increase in SSIM compared to the second-best performance, while using fewer parameters. The incorporation of DENet further enhances LPIPS and NIQE scores but marginally affects PSNR and SSIM. Nevertheless, the PSNR and SSIM of IC-FSRDENet are still notably higher than those of the other methods. Overall, the results in Table 1 affirm the superiority of our method and its joint framework over the simple cascaded solutions, making it a highly promising and effective approach for low-light low-resolution face super-resolution.

| | PSNR | SSIM | NIQE | LPIPS |
|---|---|---|---|---|
| Model 1 | 20.39 | 0.5856 | 12.601 | 0.2466 |
| Model 2 | <u>21.48</u> | <u>0.6310</u> | <u>10.689</u> | <u>0.2018</u> |
| IC-FSRNet | **21.68** | **0.6349** | **10.585** | **0.1990** |

Table 2: Ablation study of the proposed IC-FSRNet.

**Qualitative Comparison.** The visual quality comparison of various methods are presented in Figure 3. The simple cascaded solutions exhibit artifacts and fail to restore facial structures effectively. The reason lies in that the individual approach only focuses on single degradation while the input suffers from coupled degradations. That is, the former process would inevitably generate artifacts and errors, which deteriorates the latter process and causes error accumulation, ultimately arising the loss of important facial information. In contrast, our method takes into account both low-light and low-resolution degradations and proposes a joint framework for recovering LLR face images. In terms of visual quality, IC-FSRNet successfully restores clear facial structures that are unattainable with cascaded solutions. However, faces reconstructed by IC-FSRNet still exhibit noise and lack fine details. Thanks to the DENet, our approach is capable of recovering visually pleasing high-quality face images.

## Face Recognition Results

In addition to recovering high-quality face images, it is essential for face super-resolution (FSR) methods to enhance the performance of downstream tasks, such as face recognition. Therefore, we conduct a comparative analysis between our proposed method and cascaded methods in terms of face recognition performance. To evaluate the face recognition performance, a pretrained face recognition model Deepface (Serengil and Ozpinar 2020) is employed to extract facial presentations from the hallucinated face images generated by different methods, as well as the ground truth face im-

| IC-FSRNet | GAN | DENet | PSNR | NIQE | LPIPS |
|---|---|---|---|---|---|
| ✓ | | | **21.68** | 10.585 | 0.1990 |
| ✓ | ✓ | | 20.46 | 8.837 | 0.1417 |
| | | ✓ | 18.00 | **5.453** | 0.1692 |
| ✓ | | ✓ | 20.53 | 6.129 | **0.1267** |

Table 3: Ablation study of DENet.

ages. Subsequently, the cosine similarity between these facial presentations is chosen as a measure of face recognition performance. The comparison results are presented in Figure 4. The cosine similarity achieved by our method is significantly higher than that of the comparison methods. This outcome clearly demonstrates that our method surpasses other approaches in terms of face recognition performance. This findings highlight the effectiveness and superiority of our method in improving downstream task of face recognition.

## Ablation Study

**Effectiveness of IC-FSRNet.** To analyze the effectiveness of IC-FSRNet, we delete the DENet, remove the illumination estimation branch and replace the proposed FSR branch with the commonly used RCAB (Zhang et al. 2018b), generating the baseline Model 1. Then, the illumination branch is added to the Model 1 to analyze whether the illumination estimation branch can assist FSR, obtaining Model 2. Note that the illumination branch in Model 2 only estimates illumination coefficients once. Compared with Model 1, Model 2 achieves better performance, demonstrating that the introduction of illumination information can boost the recovery of LLR face images. After that, we turn to analysis of iteration strategy. The one-step illumination estimation from LLR face in Model 2 is changed to multi-step illumination estimation from the intermediate features, and the modified model is the proposed IC-FSRNet. From Table 2, the IC-FSRNet achieves the best performance. In addition, diffusion models excel at image generation as widely recognized in academic literature. Then, is directly training DENet with original LLR image better than our two-step method? To answer it, DENet directly trained with LLR-HR face image pairs, is used for comparison. As shown in Table 3 and Figure 5, the model achieves the best NIQE but the worst PSNR, and the hallucinated faces have different facial components, and obvious color deviation and artifacts. The phenomenon demonstrates that directly training DENet without IC-FSRNet results in the model excessively focusing on generating high-quality face images and unable to guarantee structural fidelity, verifying the effectiveness of IC-FSRNet.

**Effectiveness of DENet.** DENet is tasked with improving facial details and visual quality. Thus, image visual quality and the metrics associated with human perception (*i.e.*, LPIPS and NIQE) are the basis for evaluating performance of model with and without DENet. In addition, generative adversarial network (GAN) can also improve the visual quality of face images. Thus, a GAN-based model by finetuning on pretrained IC-FSRNet, is also used for comparison. The comparison results are shown in Table 3 and

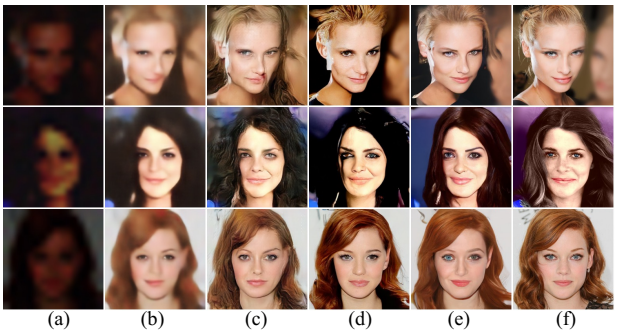


Figure 5: Ablation study. (a) LLR; (b) IC-FSRNet; (c) IC-FSRGAN; (d) DENet; (e) IC-FSRDENet; (f) HR.

Figure 5. It is obvious that the faces hallucinated by IC-FSRNet achieve the best PSNR but they are smooth and lack high-frequency details. IC-FSRGAN can improve the visual quality of the face images but bring evident artifacts. In contrast, the IC-FSRDENet can achieve better NIQE and LPIPS metrics, and the face images hallucinated by our method can provide a more visually pleasing experience.

## Discussion and Limitations

Our method may involve a lengthy inference process since DENet is built on diffusion models and inherits the characteristics of diffusion models. Furthermore, since the input suffers from complex degradation, the available information is limited, making the restoration process more challenging. it is possible that some facial attributes cannot be fully recovered and maintained. The hallucinated face images may have incorrect attributes for extreme cases. In the future, we would continue exploring and developing advanced techniques to break these limitations.

## Conclusion

We propose a novel approach for low-light low-resolution face super-resolution and decompose this hard task into structural fidelity maintaining and texture consistency learning. The former is tailored for enhancing the quality of LLR faces while maintaining structural fidelity, while the latter focuses on eliminating perturbations and artifacts caused by low-light degradation and reconstruction. Specifically, our method consists of two steps: an illumination correction face super-resolution network (IC-FSRNet) for capturing the face characteristic and recovering structural information, and a detail enhancement network (DENet) for improving facial details and visual quality. Given that LLIE and FSR are complementary, IC-FSRNet builds an illumination estimation (IE) branch for estimating illumination adjustment coefficients, and a face super-resolution (FSR) branch for face super-resolution, and iteratively performs illumination coefficient estimation and face super-resolution to let them mutually reinforce each other, promoting face reconstruction. After that, DENet based on diffusion probabilistic model is used to remove artifacts and improve the quality of coarse results. Experimental results show that our method achieves state-of-the-art performance.

## Acknowledgments

The research was supported by the National Natural Science Foundation of China (U23B2009, 92270116) and in part by the Fundamental Research Funds for the Central Universities.

## References

Baker, S.; and Kanade, T. 2000. Hallucinating Faces. *IEEE Int.conf.automatic Face I and Gesture Recognition*, 83–88.

Bao, Q.; Liu, Y.; Gang, B.; Yang, W.; and Liao, Q. 2023. SCTANet: A Spatial Attention-Guided CNN-Transformer Aggregation Network for Deep Face Image Super-Resolution. *IEEE Transactions on Multimedia*, 1–12.

Chan, K. C.; Wang, X.; Xu, X.; Gu, J.; and Loy, C. C. 2021. Glean: Generative latent bank for large-factor image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 14245–14254.

Chang, H.; Yeung, D. Y.; and Xiong, Y. 2004. Super-resolution through neighbor embedding. In *Proceeding of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, 1–275.

Chen, C.; Gong, D.; Wang, H.; Li, Z.; and Wong, K. Y. K. 2021. Learning Spatial Attention for Face Super-Resolution. *IEEE Trans. Image Processing*, 30: 1219–1231.

Chen, J.; Paris, S.; and Durand, F. 2007. Real-time edge-aware image processing with the bilateral grid. *ACM Transactions on Graphics (TOG)*, 26(3): 103–es.

Chen, Y.; Tai, Y.; Liu, X.; Shen, C.; and Yang, J. 2018. FSRNet: End-to-End Learning Face Super-Resolution with Facial Priors. In *Proceedings of The IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2492–2501.

Gao, G.; Xu, Z.; Li, J.; Yang, J.; Zeng, T.; and Qi, G.-J. 2023. CTCNet: A CNN-Transformer Cooperation Network for Face Image Super-Resolution. *IEEE Transactions on Image Processing*, 32: 1978–1991.

Guo, C. G.; Li, C.; Guo, J.; Loy, C. C.; Hou, J.; Kwong, S.; and Cong, R. 2020. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 1780–1789.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33: 6840–6851.

Huang, J.; Liu, Y.; Zhao, F.; Yan, K.; Zhang, J.; Huang, Y.; Zhou, M.; and Xiong, Z. 2022a. Deep Fourier-Based Exposure Correction Network with Spatial-Frequency Interaction. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XIX*, 163–180. Springer.

Huang, J.; Liu, Y.; Zhao, F.; Yan, K.; Zhang, J.; Huang, Y.; Zhou, M.; and Xiong, Z. 2022b. Deep Fourier-Based Exposure Correction Network with Spatial-Frequency Interaction. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XIX*, 163–180. Springer.

Jiang, J.; Wang, C.; Liu, X.; and Ma, J. 2023. Deep Learning-based Face Super-resolution: A Survey. *ACM Computing Surveys*, 55(1): 1–36.

Kawar, B.; Elad, M.; Ermon, S.; and Song, J. 2022. Denoising Diffusion Restoration Models. In *Advances in Neural Information Processing Systems*.

Li, X.; Chen, C.; Zhou, S.; Lin, X.; Zuo, W.; and Zhang, L. 2020a. Blind face restoration via deep multi-scale component dictionaries. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16*, 399–415. Springer.

Li, X.; Li, W.; Ren, D.; Zhang, H.; Wang, M.; and Zuo, W. 2020b. Enhanced blind face restoration with multi-exemplar images and adaptive spatial feature fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2706–2715.

Li, X.; Liu, M.; Ye, Y.; Zuo, W.; Lin, L.; and Yang, R. 2018. Learning warped guidance for blind face restoration. In *Proceedings of the European conference on computer vision (ECCV)*, 272–289.

Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L.; and Timofte, R. 2021. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, 1833–1844.

Lu, T.; Wang, Y.; Zhang, Y.; Wang, Y.; Wei, L.; Wang, Z.; and Jiang, J. 2021. Face Hallucination via Split-Attention in Split-Attention Network. In *Proceedings of the 29th ACM International Conference on Multimedia*, 5501–5509.

Ma, C.; Jiang, Z.; Rao, Y.; Lu, J.; and Zhou, J. 2020. Deep Face Super-Resolution With Iterative Collaboration Between Attentive Recovery and Landmark Estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5569–5578.

Mittal, A.; Soundararajan, R.; and Bovik, A. C. 2013. Making a "Completely Blind" Image Quality Analyzer. *IEEE Signal Processing Letters*, 20(3): 209–212.

Saharia, C.; Ho, J.; Chan, W.; Salimans, T.; Fleet, D. J.; and Norouzi, M. 2021. Image super-resolution via iterative refinement. *arXiv:2104.07636*.

Serengil, S. I.; and Ozpinar, A. 2020. LightFace: A Hybrid Deep Face Recognition Framework. In *Proceedings of Innovations in Intelligent Systems and Applications Conference*, 23–27.

Wang, C.; Jiang, J.; Zhong, Z.; and Liu, X. 2023a. Spatial-Frequency Mutual Learning for Face Super-Resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 22356–22366.

Wang, T.; Zhang, K.; Shen, T.; Luo, W.; Stenger, B.; and Lu, T. 2023b. Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 2654–2662.

Wang, X.; Li, Y.; Zhang, H.; and Shan, Y. 2021. Towards real-world blind face restoration with generative facial prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9168–9178.

Wang, Y.; Hu, Y.; Yu, J.; and Zhang, J. 2023c. Gan prior based null-space learning for consistent super-resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 2724–2732.

Wang, Y.; Hu, Y.; and Zhang, J. 2022. Panini-Net: GAN prior based degradation-aware feature interpolation for face restoration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 2576–2584.

Wang, Y.; Wan, R.; Yang, W.; Li, H.; Chau, L.-P.; and Kot, A. 2022. Low-light image enhancement with normalizing flow. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 2604–2612.

Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Processing*, 13(4): 600–612.

Wu, W.; Weng, J.; Zhang, P.; Wang, X.; Yang, W.; and Jiang, J. 2022. Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5901–5910.

Xu, X.; Wang, R.; Fu, C.-W.; and Jia, J. 2022. SNR-aware low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17714–17724.

Yu, C.; Wang, J.; Peng, C.; Gao, C.; Yu, G.; and Sang, N. 2018. Bisenet: Bilateral segmentation network for real-time semantic segmentation. In *Proceedings of the European Conference on Computer Vision*, 325–341.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018a. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 586–595.

Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; and Fu, Y. 2018b. Image Super-Resolution Using Very Deep Residual Channel Attention Networks. In *Proceedings of the European Conference on Computer Vision*, 286–301.

Zhou, S.; Li, C.; and Change Loy, C. 2022. Lednet: Joint low-light enhancement and deblurring in the dark. In *European Conference on Computer Vision*, 573–589. Springer.