IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 5 July 2024, accepted 24 July 2024, date of publication 29 July 2024, date of current version 7 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3435351

RESEARCH ARTICLE

A Novel Blind Restoration Method for Miner Face Images Based on Improved GFP-GAN Model

XIANMING ZHANG[1] **AND JIAOJIAO FENG**[1,2]

[1]School of Computer and Information Engineering, Heilongjiang University of Science and Technology, Harbin 150022, China
[2]Financial Technology Department, Jiangsu Changshu Rural Commercial Bank Company Ltd., Changshu 215500, China

Corresponding author: Xianming Zhang (zxm@usth.edu.cn)

This work was supported by the research projects of Basic Scientific Research Business Expenses of Provincial Colleges and Universities in Heilongjiang Province under Grant Hkdqg201911.

ABSTRACT Miner face images, as important carriers of information transmission, are an important means of digital transformation and intelligent management of mining enterprises. In order to address the issue of complex degradation factors such as noise, blurring, and low resolution, a blind restoration model for miners' face images was proposed based on improved GFP-GAN, and could make it difficult for blind image restoration to balance fidelity and authenticity. Firstly, the model introduced a UNet++ network to remove complex degradation from miners' face images using the pre-trained StyleGAN2 network as a priori knowledge. Secondly, in the channel-split spatial feature transform layer, a channel attention mechanism was introduced to better use the prior features in the pre-training network, which could make the final output of the miners' face images consider both authenticity and fidelity. Compared with other model algorithms, the experimental outcomes clearly indicate that our proposed method surpasses the current leading techniques in LPIPS (0.3827), FID (46.51), NIQE (5.206), and other indicators.

INDEX TERMS Miner face image, blind image restoration, super-resolution, attention mechanism.

I. INTRODUCTION

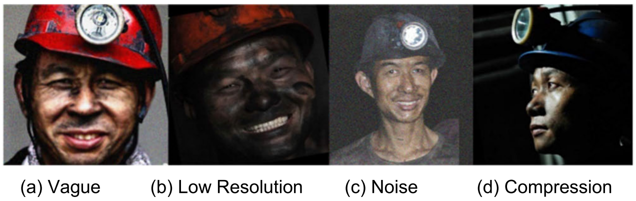
Miner face images, as important carriers of information transmission, have an important application value in mining enterprises. They can improve the safety, production, attendance efficiency and personnel management level of enterprises and are one of the important means of digital transformation and intelligent management of mining enterprises.

Due to the unique working environment of miners, their faces are often covered with a large amount of dust particles after long hours of work, which has a certain impact on the contour and facial features of the target. In addition, some collection devices, such as daily monitoring and transmission equipment, are susceptible to diverse degradative influences, including noise, compression, blur, Limited resolution and the presence of artifacts in miners' images. These are all degradation behaviors of miners' facial image information, which may result in the superposition of multiple degradation effects on miners' facial images, causing the main feature information of the face to be easily lost. In practical use, this has a dominant impact on the precision of the miners' face recognition, causing an increase in costs in mining enterprises moving towards intelligence. A variety of degraded image examples are presently exhibited in Figure 1.

Therefore, the purpose of restoring degraded images is to minimize or eliminate the degradation of image quality and to restore the original appearance of the degraded images as much as possible [1]. The effective utilization of computer and digital processing technology for facial image restoration is a focus of attention for scholars both domestically and internationally [2]. High-quality facial images of miners can also more accurately determine their identities, assisting in the intelligent management of mines.

In the early days, blind face restoration models have primarily centered their attention on statistical priors and degraded models, which are roughly compartmentalized into methods based on Bayesian inference, subspace learning,

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



(a) Vague (b) Low Resolution (c) Noise (d) Compression

FIGURE 1. **Low-quality miner face images with different degradation types.**

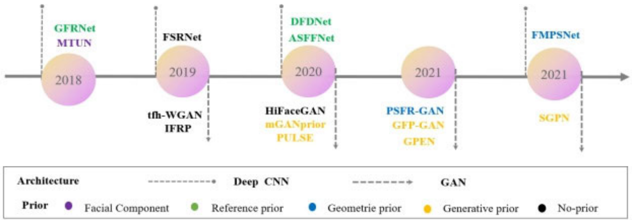


FIGURE 2. **Development of blind image restoration methods.**

sparse representation, etc. Bayesian inference-based methods struggle to repair complex degraded facial images, resulting in noisy and blurry restored images. The methods based on subspace learning and sparse representation also struggle to acquire more detailed information, which brings great difficulties to the blind restoration of facial images.

Recently, facial image blind restoration techniques have predominantly gravitated towards the realm of deep learning, as evident in Figure 2. They can be roughly compartmentalized into two categories: prior and non-prior.

A. A PRIOR-BASED DEEP RESTORATION MODEL

Among restoration methods, prior-based deep restoration models can be roughly compartmentalized into three groups: geometric prior, reference prior, and generative prior.

The deep restoration model based on geometric prior utilizes,the distinctive geometric configurations and spatial arrangement characteristics inherent to facial features, which assists the model to gradually restore high-quality facial images. In 2018, Yu et al. [3] proposed an MTUN model that utilizes the features of facial components in facial images as prior information, using four thermal maps to represent the five sense organs. In 2021, Hu first attempted to fuse 3D priors into general facial restoration networks [4]. Compared to 2D priors, feature descriptions of facial attributes can be better integrated by 3D priors, which can provide 3D morphological knowledge. In the same year, a progressive semantic perception style transformation framework, PSFR-GAN, was proposed by Chen et al., and multi-scale progressive restoration was used for blind face restoration [5]. To assess the authenticity of the generated faces, this study employed the commonly utilized metrics of Peak Signal-to-Noise Ratio (PSNR), Structure Similarity Index Measure (SSIM), and Mean Structural Similarity Index Measure (MSSIM), along with the learned perceptual image patch similarity (LPIPS) score. Furthermore, the Frechet perception distance (FID) was utilized to quantify the statistical disparity between the restored outcomes and the reference high-quality (HQ) face dataset. In 2022, Yu et al. attempted for the first time to fully utilize multiple geometric priors in the proposed MFSPSNet network, which utilizes semantic parsing maps, facial heat maps, and reference-level facial dictionaries to guide facial restoration [6].

A deep restoration method based on reference prior uses facial structures or facial component dictionaries. In 2018, a guided face recovery network (GFRNet) model was proposed by Li et al. [7] and can provide additional identity perception information to assist in the face restoration process by using fixed positive high-quality references for each identity. In 2020, utilizing multi-sample images, adaptive fusion of guidance images, and degraded image features, an improved blind face restoration technique was introduced by Zuo et al. in [8] pioneered a novel approach. Shortly after, Li et al. [9]. presented a deep face dictionary network model (DFDNet) for facial restoration; this model used the deep component dictionary generated by K-means as a reference before the restoration process.

The deep restoration method based on generative prior uses pre-trained facial GAN as the generative prior, such as StyleGAN [10] and StyleGAN2 [11], which can provide rich and diverse facial information. PULSE [12] is a representative method that optimizes the potential of pre-trained StyleGAN for self-supervised facial restoration. Inspired by PULSE, multiple potential codes were considered in the pre-trained GAN, and the program was optimized to improve image reconstruction capabilities for multi-code GAN prior (mGANprior) [13]. However, these models cannot maintain fidelity in restored facial images. In 2021, fidelity information was first extracted from low-quality facial image input bygenerative facial prior GAN (GFP-GAN) [14] and GAN prior embedded network (GPEN) [15], subsequently, a pre-trained GAN served as the decoder, effectively capturing facial priors. Specifically, the distribution of facial features was employed as a prior, enabling joint recovery and color enhancement through the utilization of the pre-trained GAN. Furthermore, GFP-GAN extensively leveraged metrics such as PSNR, SSIM, and MSSIM. It evaluated the perceptual authenticity of the generated faces using the learned perceptual image patch similarity (LPIPS) score. Additionally, the statistical distance between the restored outcomes and the reference high-quality (HQ) face dataset was quantified using the Frechet perception distance (FID). Both the advantages of GAN and DNN were effectively integrated for facial repair by GPEN. In 2022, Zhu et al. [16] proposed a SGPN network that combined shape and generation prior, established a shape recovery module, and restored facial geometry through 3D reconstruction technology.

B. NON-PRIOR-BASED DEEP RESTORATION MODEL

Compared to prior models, non-prior-based deep restoration models save computational costs and reduce computational time. Compared to CNN, GAN networks can generate more realistic images. Therefore, in 2019, Shao et al. [17] proposed a focus on designing more complex GAN models, namely tfh-WGAN, which includes GAN, WGAN, and multi-level GAN networks. An identity preservation algorithm was used to assist the GAN model effectively generates high-quality facial images, preserving accurate identity information, as demonstrated by Identity-preserving face recovery portraits (IFRP) [18]. In 2020, Yang et al. [19] constructed a HiFaceGAN model that does not include facial prior knowledge or facial degradation. The model has good robustness in blind facial restoration, and its structure includes suppression blocks and supplementary blocks. The suppression blocks are used to collect effective information, while the supplementary blocks maximize the information collected.

In the blind restoration task of miners' facial images, when faced with images affected by one or more degradation factors, non prior based restoration methods cannot utilize prior information, such as IFRP [18], which is difficult to extract identity information affected by multiple factors, resulting in poor restoration performance in blind restoration tasks. By comparison, prior based restoration methods utilize high-quality faces or facial components, including facial heatmaps and facial anatomical maps [6]. Geometric priors are constrained by geometric constraints that make it difficult to maintain the details of image information. Reference priors tend to favor reference images with the same identity and are also constrained by image integrity. Although the DFDNet [8] network constructs a face dictionary, the design of the dictionary is limited by diversity and richness. In complex scenes, whether it is geometric priors or reference priors, they cannot effectively maintain the diversity and richness of information details. Therefore, using generative priors can break through geometric constraints, not limited by dictionary size, and provide rich and diverse prior information such as texture and color to the maximum extent, so that the final recovered miner face image can balance fidelity and authenticity.

A novel blind restoration model for miners' face images based on improved GFP-GAN is proposed to tackle the intricate degradation factors encompassing noise, blurring, and low resolution, which can make it difficult for blind image restoration to balance fidelity and authenticity. There are two main innovative points in this article:

(a) By replacing the U-Net structure of the degradation removal module in the GFP-GAN model with the UNet++ structure, the model captures deeper feature information of the image, pays more attention to facial details, and obtains rich facial information from miners.

(b) To enhance the model's capability in capturing multi-scale features and delivering finer facial images, the proposed model integrated the channel-split spatial feature transform (CS-SFT) within the GFP-GAN framework, incorporating the squeeze and excitation (SE) mechanism. This approach aims to further refine the miner face images and improve their quality.

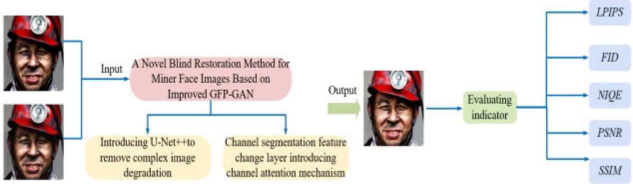

Input
A Novel Blind Restoration Method for Miner Face Images Based on Improved GFP-GAN
Introducing U-Net++to remove complex image degradation
Channel segmentation feature change layer introducing channel attention mechanism
Output
Evaluating indicator
LPIPS
FID
NIQE
PSNR
SSIM

FIGURE 3. The comprehensive framework of the model proposed in this study.

II. MATERIALS AND METHODS

As seen in Figure 3, the whole framework of this article mainly includes the following two stages. Firstly, the model introduces a UNet++ network to remove complex image degradation and uses a pre-trained StyleGAN2 network as prior knowledge. Secondly, the model introduces a channel attention mechanism on the original channel-split spatial feature transform (SFT) layer to effectively leverage the prior features embedded within the pre-trained network., which can balance the authenticity and fidelity for the final miner face output image. In comparison with other algorithms, the proposed model achieved much better experimental results for LPIPS, FID, and NIQE indicators.

A. GFP-GAN NETWORK

The network structure of the GFP-GAN is presented in Figure 4. The core architecture consists of a degradation removal module and a GAN, pre-trained as a prior, which bridge from coarse to fine through potential code mapping and multiple CS-SFT, achieving an equilibrium between fidelity and authenticity through a single forward transfer. During the training process, the intermediate restoration loss was used to eliminate complex degradation, the facial component loss of the discriminator was used to improve facial details, and the identity preservation loss was utilized to preserve facial features.

Specifically, as a prior module, pre-trained GAN is also known as facial prior. It obtains the distribution of facial details through pre-trained facial GAN in the weight of convolution and uses a StyleGAN2 network with strong generation ability for pre-training, which can provide the diversified and enriched facial details for blind facial restoration work. The typical method of generating prior knowledge in the past was to map the facial image of the network to the nearest potential code and then generate the corresponding output image through pre-trained GAN [20]. The principle of this method is to map random latent codes into facial images and then implicitly encapsulate prior information in the GAN network. In specific operational implementation, in order to achieve high fidelity facial images, it often requires a long iterative optimization process to achieve the desired effect, which to some extent limits the representativeness of semantic attributes. Therefore, GFP-GAN does not directly generate the final facial image but rather obtains intermediate convolution features generated by pre-trained GAN

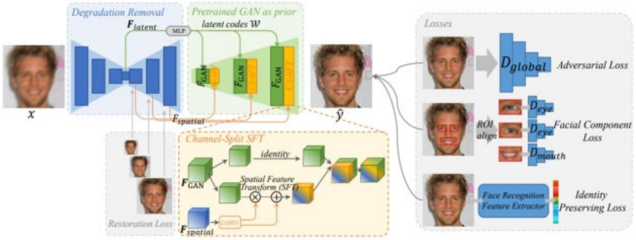


FIGURE 4. The architectural design of the GFP-GAN network.

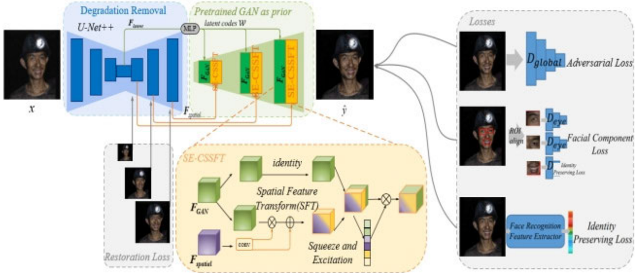


FIGURE 5. Improved GFP-GAN network architecture.

(StyleGAN2). The convolution features contain a variety of semantic information, and most importantly, the input features can be further adjusted to maintain the fidelity of the output image.

Specifically, the degradation removal module aims to promote the model's expressive power, preserve the semantic attributes for image, extract feature information, and map the information, as shown in Formulas (1) and (2).

$$W = MLP(F_{latent}) \tag{1}$$

$$F_{GAN} = StyleGAN2(W) \tag{2}$$

In the formula, given the potential coding vector F_{latent} of the face image generated by the degradation removal module, initially, the model maps the input to an intermediate potential code W via a series of multi-layer perceptron (MLP) layers, and then generates convolution features for each resolution dimension by passing the potential code W through each convolution layer in StyleGAN2

B. IMPROVED GFP-GAN NETWORK MODEL

Figure 5 illustrates the enhanced network structure of the GFP-GAN model. An unknown degraded miner face image x is input, and through blind facial image restoration, a high-quality image $\hat{y}$ is output to be as approximate to the real image y' s fidelity as possible.

Specifically, a backbone network using UNet++ as the degradation removal module was introduced first, with the main purpose of removing complex degradation from images and extracting two types of features: Firstly, the input images undergo a transformation, mapping them onto the nearest latent code within the pre-trained GAN (StyleGAN2) through latent feature. The second is the multi-resolution spatial feature $F_{spatial}$ extraction in the degradation removal module, which can be applied to modulate the pre-trained GAN (StyleGAN2) features. By multiple linear layer mapping, F_{latent} is transformed to the intermediate latent code W, and F_{GAN} represents the intermediate convolutional features generated by the pre-trained GAN (StyleGAN2) for the latent code that is similar to the input facial image. These intermediate convolution features can provide the weight of the StyleGAN2 network to capture more varied and rich priors and to restore rich details and real textures. Real miners' facial images are mostly taken in dimly lit environments, and the images contain a lot of dust. The vivid color priors generated in the facial priors enable the model to perform color enhancement, including coloring.

Secondly, through introducing the channel attention mechanism, the squeeze-and-extraction channel-split spatial feature transform (SE-CSSFT) layer gradually refines the spatial modulation of the intermediate convolution feature F_{GAN}. The specific operation was achieved using the multi-resolution spatial feature $F_{spatial}$, which can balance fidelity and authenticity.

1) DEGRADATION REMOVAL MODULE BASED ON IMPROVED U-NET

The images of miners' face in the real-world exhibit complex low resolution, artifacts, severe blurring, and noise degradation. The degradation removal module specifically removes these degradation types, reduces the burden on subsequent modules, and allows clearer facial features, F_{latent} and $F_{spatial}$, to be extracted.

The GFP-GAN network initially adopted U-Net as the backbone of the degradation removal module. U-Net was proposed in 2015 [20] and achieved some effects in the domain of medical image segmentation. U-Net adopts a left and right symmetric U-shaped structure, and the left and right parts can also be further divided into encoding and decoding parts.

On this basis, this article improves the U-Net skip connection structure by adopting the UNet++ framework, which was improved by Zhou et al. [21]. Inspired by the skip connection, Li et al. formed a UNet++ structure based on nested and dense skip connections, which can effectively improve the problem of semantic differences caused by improper feature information fusion, this is depicted in Figure 6.

Upon examination of Figure 6a, it becomes evident that UNet++ connects all four layers of U-Net together, represented by green lines in the figure, bridging a semantic gap between the encoding and decoding subnetworks. The high-resolution feature maps are gradually fused from the left encoding subnetwork to the corresponding semantic feature maps in the right decoding subnetwork and integrated through superposition to obtain features at different levels. And gradient flow can be also improved by the series of dense and nested jump paths, represented in blue in the figure.

In Figure 6b, 6c, the skip connection's structure is shown, moreover, the detailed calculation procedure is outlined in

Formula (3).

$$x^{(j,k)} = \begin{cases} H(x^{(j-1,k)}), & k=0 \\ H([[x]^{(j,i)}]_{i=0}^{k-1}, U(x^{j+1,k-1})), & k=1 \end{cases} \tag{3}$$

In the equation, $x^{(j,k)}$ is the output feature graph, where j denotes the total number of network layers, while k signifies the count of dense convolution layers featuring skip connections; $H(\cdot)$ is the convolution operation, [ ] is the cascading layer; $U(\cdot)$ is the up sampling layer. When $k = 0$, the input feature refers to the output feature from the previous layer in the encoding subnetwork. When $k = 1$, there are two different layers of encoding sub network inputs.

The improved UNet++ jump connection has a deeper range in feature capture, pays more attention to details, and further expands the range of the receptive field, which greatly improves the ability to extract features. Consequently, within the degradation removal module, both features F_{latent} and $F_{spatial}$ are utilized to transform the input image x into the nearest latent code, while simultaneously modulating the StyleGAN2 features respectively, as shown in Formula (4).

$$F_{latent}, F_{spatial} = UNet++(x) \tag{4}$$

2) CS-SFT LAYER BASED ON SQUEEZE AND EXCITATION

The latent feature F_{latent} can better preserve semantic attributes for the input miner face image. The $F_{spatial}$ generated by the degradation removal module network is used, and the input spatial feature $F_{spatial}$ is used to modulate the StyleGAN2 feature F_{GAN}. The model must efficiently combine the input spatial feature $F_{spatial}$ with the prior feature F_{GAN} of pre-trained miners' facial images as a key step in achieving experimental results, as this study wants to balance fidelity and authenticity at the same time. Given that preserving the fidelity of miners' facial images depends on local features and requires adaptive restoration at different spatial positions on miners' faces, it is crucial to preserve spatial information from input for the facial restoration of miners' facial images. Therefore, the network adopts SFT [22], which can generate affine transform parameters and can help spatial feature modulation. In addition, SFT can also combine other conditions to make it effective in image restoration.

Specifically, across each level of resolution, through several convolutional layers (α, β), a pair of affine transform parameters is generated based on the input feature $F_{spatial}$. Afterwards, subsequently, modulation is achieved by scaling and shifting the GAN feature F_{GAN}, as shown in Formulas (5) and (6).

$$\alpha, \beta = Conv(F_{spatial}) \tag{5}$$

$$F_{output} = SFT(F_{GAN}|\alpha, \beta) = \partial \odot F_{cnn} + \beta \tag{6}$$

Although *SFT* can effectively fuse the input miners' facial features in order to obtain images without losing their authenticity, the CS-SFT layer was used. The CS-SFT layer divides the prior features generated by StyleGAN2 into two branches according to the channel, and one branch directly transmits the features to save information. Direct transmission can save more prior information, which helps to improve authenticity. Another branch uses multi-resolution features to adjust the spatial features of one set of features and finally continues to concatenate the features of the two channels, as shown in Formula (7).

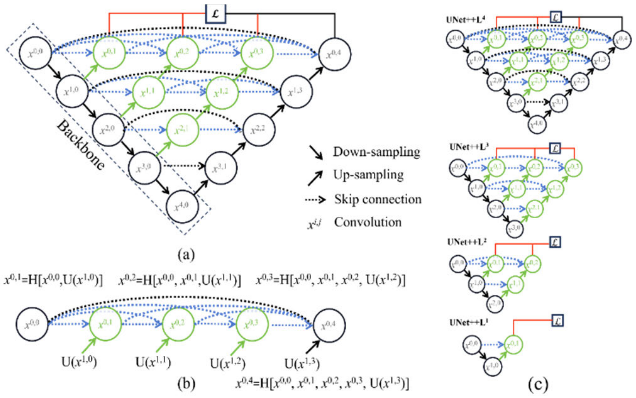


FIGURE 6. UNet++ network structure architecture:(a) Four layers network skip connection diagram (b) One layer network skip connection diagram (c) Detailed diagram of connection jump.

$$\begin{aligned} F_{output} &= CS-SFT(F_{GAN}|\alpha, \beta) \\ &= Concat[Identity(F_{GAN}^{split0}), \alpha \odot F_{GAN}^{split1} + \beta] \end{aligned} \tag{7}$$

In the formula, F_{GAN}^{split0} is the segmentation feature of F_{GAN} in the channel dimension;F_{GAN}^{split1} is the segmentation feature of F_{GAN} in the channel dimension; and $Concat[\cdot, \cdot]$ represents splicing operation.

On the basis of CS-SFT, this article further integrated Squeeze and Excitation (SE) to form SE-CSSFT. The spatial feature transformation in SE-CSSFT is divided into two branches, and the final concatenated results are input into the SE block. This module focuses on the importance of the two branch channels and can provide the learned coefficients for the two branches according to the actual situation. These values can adjust and control the proportion of facial prior information in the results, ensuring the overall balance of the final effect. The structural adjustment of this part is shown in Figure 7.

The structural change shown in Figure 7 is represented by a formula, as shown in Formula (8).

$$\begin{aligned} F_{output} &= SE-CSSFT(F_{GAN}|\alpha, \beta) \\ &= SE\{Concat[Identity(F_{GAN}^{split0}), \alpha \odot F_{GAN}^{split1} + \beta]\} \end{aligned} \tag{8}$$

Consequently, SE-CSSFT has the superiorities of incorporating prior features and modulating the input miner face image, thus achieving much better equilibrium between facial texture realism and fidelity. In addition, SE-CSSFT adopts a dual-channel approach, which can improve the expressive power of the channel-splitting spatial feature transformation layer. As it does not require too many modulation channels,

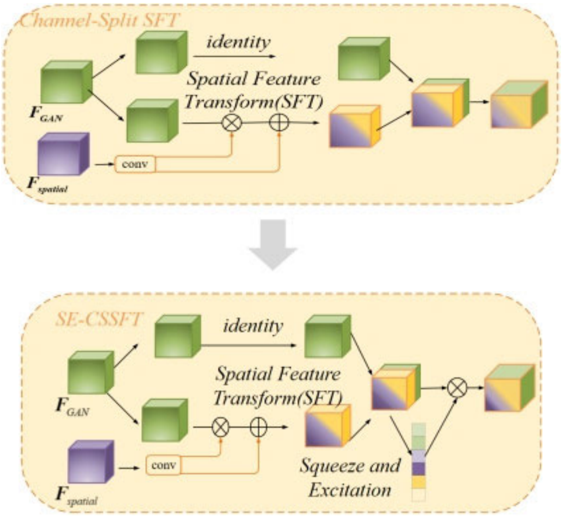


FIGURE 7. SE-CSSFT structure architecture.

each channel can also reduce complexity. The SFT layer for channel splitting at each resolution scale and the SE module can assist in more precise fusion, ultimately generating restored faces and achieving much better balance between overall authenticity and fidelity.

C. MODEL OBJECTIVES

During the training phase, the loss function of the GFP-GAN was used. The architecture of the loss function added adversarial loss for the GFP-GAN model; it also included loss of reconstruction loss, identity-preserving loss, and facial component loss. The specific losses are as follows:

1) RECONSTRUCTION LOSS

$L1$ loss and intuition loss were used as reconstruction losses, as shown in Equation (9).

$$\mathcal{L}_{rec} = \lambda_{l1}||\hat{z} - z||_1 + \lambda_{per}||\varphi(\hat{z}) - \varphi(z)||_1 \tag{9}$$

In the formula, $\varphi(\cdot)$ is the pre-trained VGG-19 network [23]; λ_{l1} signifies the L1 weight loss, whereas λ_{per} denotes the perceptual loss.

2) ADVERSARIAL LOSS

It is hoped that the network model can generate real-texture structures by adversarial loss L_{adv}, as shown in Equation (10).

$$\mathcal{L}_{adv} = -\lambda_{adv}E_{\hat{y}}softplus(D(\hat{y})) \tag{10}$$

In the formula, $D(\cdot)$ represents the StyleGAN2 discriminator; λ_{adv} denotes the weight of the adversarial loss.

3) FACIAL COMPONENT LOSS

In the GFP-GAN network, the facial component loss was introduced into local discriminators for the mouth and two eyes in order to further improve the perceived components of the miner's face.

Algorithm 1 A Novel Algorithm for Miner Face Restoration

```
1. Input of collected miners' facial images
2. Augmenting the facial image dataset of miners
3. Input of the expanded dataset of miners' facial images into the degradation mode
4. Input of paired dataset images into the algorithm
5. $U - Net + +(x) \rightarrow F_{latent}, U - Net + +(x) \rightarrow F_{spatial}, x$ iis unknown degraded facial image
6. $MLP(F_{latent}) \rightarrow W$
7. $StyGAN2(W) \rightarrow F_{GAN}$
8. $Conv(F_{spatial})$ generats a pair of affine transform parameters $\alpha$ and $\beta$
9. $SE\{Conmat[Identity(F_{GAN}^{split0}), \alpha \odot F_{GAN}^{split1} + \beta]\}$
10. Output of restored miners' facial images
```

Since the Gram matrix can represent the correlation of multi-variable characteristics and effectively obtain texture information, it can extract features from multiple local discriminator layers and match the gram statistics represented in the middle from the real and restored parts. This can generate realistic facial details and reduce unnecessary artifacts. See Formula (11).

$$\mathcal{L}_{comp} = \sum_{ROI} \lambda_{local}E_{\hat{y}ROI}[\log(1 - D_{ROI}(\hat{z}_{ROI}))] + \lambda_{fs}||Gram(\Phi(\hat{z}_{ROI})) - Gram(\Phi(z_{ROI})||_1 \tag{11}$$

where ROI refers to the region of interest encompassing both eyes and mouth components. D_{ROI} represents the local discriminator of each region. $\Phi(\cdot)$ represents the multi-resolution features for the learned discriminators. Both λ_{local} and λ_{fs} represent the weights assigned to the local discriminative loss and feature style loss, respectively.

4) IDENTITY-PRESERVING LOSS

For identity-preserving loss [24] in the model, the most prominent features are captured by using the pre-trained ArcFace model [25], which can be used for identity recognition. In the application of this model, the identity-preserving loss was defined by the feature embedding of the miner face, to ensure a minimal distance between the restored result and the ground truth in the compact deep feature space, see Formula (12), this study introduced an identity-preserving loss.

$$\mathcal{L}_{id} = \lambda_{id}||\sigma(\hat{z}) - \sigma(z)||_1 \tag{12}$$

where σ denotes the feature extractor of the miner face, that is, pre-trained face recognition ArcFace. λ_{id} signifies the importance assigned to the identity-preserving loss. Then, see Formula (13) for the overall loss of this model.

$$\mathcal{L}_{total} = \mathcal{L}_{rec} + \mathcal{L}_{adv} + \mathcal{L}_{comp} + \mathcal{L}_{id} \tag{13}$$

The proposed algorithm is outlined in Algorithm 1.

TABLE 1. Experimental environment configuration.

Equipment and Environment	Details
Processor	Intel Core i7-9750H CPU @ 2.60GHz 2.59 GHz
Graphics Card Information	NVIDIA Tesla P40 GPUs
Operating System	Linux
Programming Language	Python
Deep Learning Framework	PyTorch1.7.1
CUDA Version	11.0
Optimizer	Adam

III. EXPERIMENTAL RESULTS

A. DATASET AND IMPLEMENTATION

The dataset comprises the captured face images of the miners; there are 270 images. Further, after dataset augmentation and 6-fold augmentation, the dataset had 1620 images, and the experiments needed to be paired into training pairs, so the degradation model was used to fit the real, low-quality images. See Formula (14).

$$x = [(y \otimes k_\sigma) \downarrow_r + \eta_\delta]_{JPEG_q} \tag{14}$$

Firstly, the Gaussian blur kernel k_σ convolution was performed on the collected image y, and then the image was down-sampled by the scale factor r, and Gaussian white noise η_δ was further added. Finally, the JPEG was compressed by the quality factor q. σ was randomly selected from $\{0.2 : 10\}$, γ was randomly selected from $\{1 : 8\}$, δ was randomly selected from $\{0 : 15\}$, and q was randomly selected from $\{60 : 100\}$.Finally, there were 1620 pairs of images, eighty percent of the data were designated as training sets, while the remaining twenty percent served as test sets, and the image size was adjusted to 512 × 512.

The augmentation of the dataset is shown in Figure 8 below.

(1) Flip the original image left and right

(2) Rotate the original image by 10°

(3) For the original image, translate the length of 20-pixel units along the x and y axes respectively

(4) Enhance or weaken the contrast of the original image by 0.2

(5) Increase or decrease the brightness of the original image by 0.2

Table 1 presents the configuration details of the experimental environment for the proposed model. The pre-trained StyleGAN2 was used to generate miner face prior. In order to make the model more compact, In StyleGAN2, the channel multiplier was fixed at 1. Each SE-CSSFT layer employed two convolutional layers for training the affine parameters. During training, the minimum batch size was established as 12, the learning rate was set to 2×10^{-3}, and a total of 400 k iterations were performed.

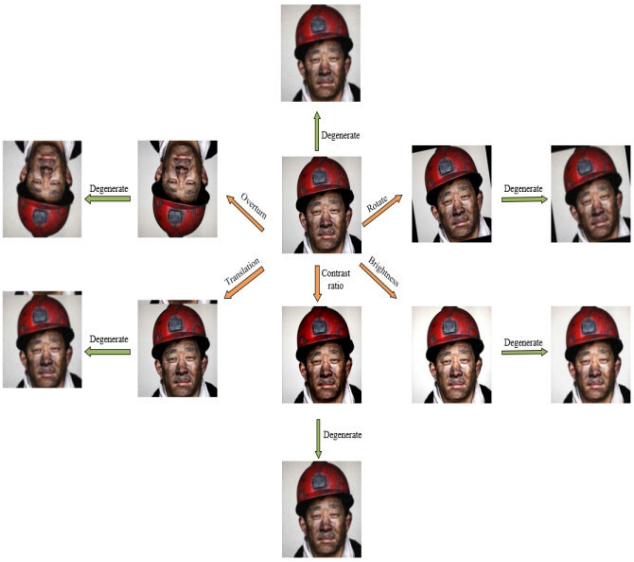


FIGURE 8. Mining face image dataset augmentation.

TABLE 2. Experimental hyperparameters.

Hyperparameters	Value
$L1$ weight λ_{l1}	0.1
Perceptual loss weight λ_{per}	1
Adversarial loss weight λ_{adv}	0.1
Loss weight for the local discriminator λ_{local}	1
Loss weight for feature style loss λ_{fs}	200
Identity-preserving loss λ_{id}	10

Table 2 details the hyperparameter configurations utilized in the training of the model for the specific experimental setup.

B. EVALUATION INDEX

1) PEAK SIGNAL-TO-NOISE RATIO (PSNR)

PSNR is a widely utilized metric for evaluating image quality. The calculation of PSNR first calculates the mean square error (MSE), and then calculates on this basis. Given a noise-free image and a noise-like image, the MSE is shown in Formula (15).

$$MSE = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} ||I(i,j) - \hat{I}(i,j)||^2 \tag{15}$$

Then the PSNR is defined in Formula (16).

$$P = 10 \log_{10}(\frac{MAX^2}{MSE}) \tag{16}$$

In the formula, MAX represents the highest pixel value of 255.PSNR is an important image evaluation index, and the higher its value, the better the image effect.

2) STRUCTURE SIMILARITY INDEX MEASURE (SSIM)

SSIM has become an important metric for measuring the similarity of image structural information, and its value range is described by [0, 1]. A higher value of PSNR indicates less distortion and greater similarity among images. Given the pixel values x and y of two images, SSIM is calculated by the brightness, contrast, and structure, as shown in Formulas (17)-(19).

$$B_r(x, y) = \frac{2m_x m_y + \theta_1}{m_x^2 + m_y^2 + \theta_1} \tag{17}$$

$$C_t(x, y) = \frac{2\delta_x \delta_y + \theta_2}{\delta_x^2 + \delta_y^2 + \theta_2} \tag{18}$$

$$S_t(x, y) = \frac{\delta_{xy} + \theta_2/2}{\delta_x + \delta_y + \theta_2/2} \tag{19}$$

In the formula, m_x and m_y denote the mean value of the pixel grayscale of images x and y, δ_x^2 and δ_y^2 represent the pixel variance of the images x and y, δ_{xy} represents the covariance of images x and y, and $\theta_1 = (\alpha_1 B_r)^2$, $\theta_2 = (\alpha_2 B_r)^2$, $\alpha_1 = 0.01$, $\alpha_2 = 0.03$.

See Formula (20) for SSIM calculation.

$$S(x, y) = B_r(x, y) * C_t(x, y) * S_t(x, y) \tag{20}$$

Compared with PSNR, SSIM is closer to the actual perception of human beings, but SSIM takes the local characteristics of the window, so the uneven distribution of the image also reflects the one-sidedness of the index to a certain extent.

3) FRECHET INCEPTION DISTANCE (FID)

FID is a description of path space similarity proposed by Maurice René Fréchet [26]. It not only considers the path space similarity but also considers the path space distance. A lower FID value corresponds to a superior image effect. and the specific calculation is shown in Formula (21).

$$F = ||m_x - m_y||^2 + Tr(\sum x + \sum y - 2(\sum x \sum y)^{1/2}) \tag{21}$$

In the formula, x and y denote the extracted feature information, obtained through the utilization of the pre-trained initial V3. m_x and m_y denote the mean value of the matrix x and y, respectively, and Tr represents the trace of a matrix.

4) LEARNED PERCEPTUAL IMAGE PATCH SIMILARITY (LPIPS)

In the field of CV, because the underlying principle is very complex, the LPIPS is typically employed to assess the similarity between various textures. The specific LPIPS is calculated by Formula (22).

$$L = d(x, x_0) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} ||w_l \odot (\hat{y}^l_{hw} - \hat{y}^l_{0hw})||_2^2 \tag{22}$$

The calculation process can be understood first by computing the depth embedding, normalizing the activations in the channel dimension, scaling each channel width, measuring the distance, and then averaging over the spatial dimension and all layers. In general, a decrease in the LPIPS value corresponds to a greater similarity among images.

TABLE 3. Comparison of experimental outcomes.

Algorithm	*LPIPS* ↓	*FID* ↓	*NIQE* ↓	*P* ↑	*S* ↑
HiFaceGAN(2020)	0.4970	64.27	6.448	25.7414	0.6176
DFDNet(2020)	0.4658	57.08	5.756	24.4203	0.6824
PSFR-GAN(2021)	0.4540	51.59	6.622	25.4336	0.6538
PULSE(2020)	0.5162	65.36	6.904	23.5285	0.6207
GFP-GAN(2021)	0.4046	48.59	5.517	25.9436	0.7142
The proposed method	**0.3827**	**46.51**	**5.206**	**26.1061**	**0.7236**

5) NATURAL IMAGE QUALITY EVALUATOR (NIQE)

The NIQE index quantifies the distance between the MVG (multivariate Gaussian) model of NSS (natural scene statistic) features. A lower NIQE value indicates a closer resemblance in quality perception features among images. It can be seen Formula (23).

$$D(\mu_1, \mu_2) = \sqrt{((\mu_1 - \mu_2)^T (\frac{cov_1 + cov_2}{2})^{-1} (\mu_1 - \mu_2))} \tag{23}$$

In the formula, μ_1 and μ_2 denote the mean vectors of the MVG models for the natural and distorted images, respectively, while cov_1 and cov_2 represent the covariance matrices corresponding to the natural and distorted image MVG models, respectively.

C. COMPARATIVE EXPERIMENT

The proposed method was compared with HiFaceGAN [19], DFDNet [9], PSFR-GAN [5], and GFP-GAN [14], which are currently popular face restoration methods, and compared them with PULSE [12], an inversion method for face blind restoration. The specific experimental results are illustrated in Table 3.

As evident from Table 3, the model introduced in this paper surpasses five other methods across various metrics. Specifically, when compared to the GFP-GAN model, the proposed model exhibits a notable improvement in PSNR value, with an increase of 0.1625dB, the SSIM has increased by 0.0094, and the LPIPS, FID, and NIQE have decreased by 0.0219, 2.08, and 0.311, respectively. Compared with the PULSE algorithm, the PSNR and SSIM values were increased by 2.5776 and 0.1029, and the LPIPS, FID, and NIQE values were reduced by 0.1335, 18.85, and 1.698, respectively. Compared with the HiFaceGAN algorithm, there was a notable increase in PSNR by 0.3647 and SSIM by 0.106, and the LPIPS, FID, and NIQE values were reduced by 0.1143, 17.76, and 1.242, respectively. Compared with the PSFR-GAN algorithm, the PSNR value rose by 0.6725, while the SSIM value increased by 0.0698, and the LPIPS, FID, and NIQE decreased by 0.0713, 5.08, and 1.416, respectively.

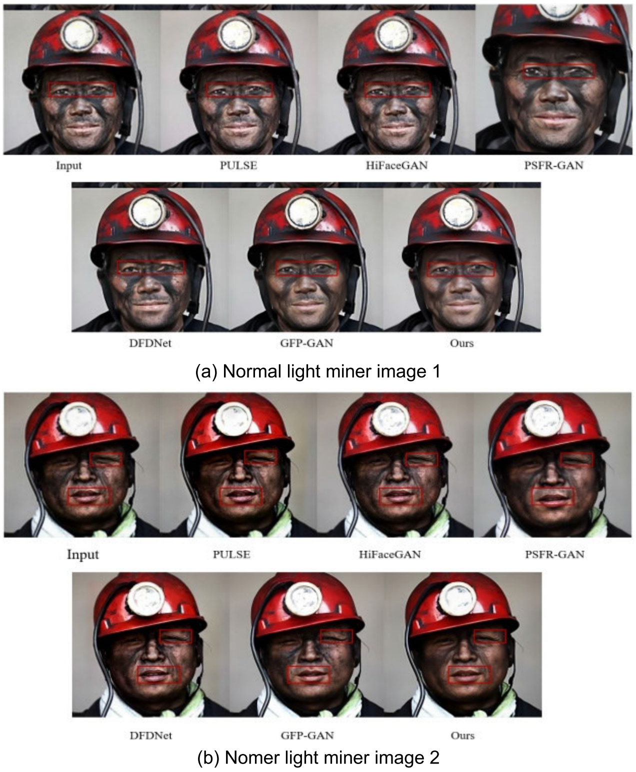
Compared with the DFDNet algorithm, the PSNR value witnessed a significant enhancement of 1.6858, whereas the SSIM value exhibited a moderate increase of 0.0412. and the LPIPS, FID, and NIQE decreased by 0.0831, 10.57, 0.55, respectively.

The PSFR-GAN, DFDNet, and GFP-GAN models that apply the face analysis graph outperformed the PULSE and HiFaceGAN models in terms of indicators, especially in the LPIPS, FID, and NIQE indicators. This also shows that in the image restoration task, the integration of prior knowledge can better improve the image quality level.

The experimental results are divided into normal light and low light, respectively, as shown in Figures 9 and 10.

As shown in Figure 9, compared with other models, the proposed model demonstrates superior sensory quality in generating output images. Notably, in the case of miners' facial images, due to the pollution of the face, there are more aspects that could be paid attention to in blind restoration compared to images in other fields. The methods of PULSE, HiFaceGAN, and PSFR-GAN have improved the blurred sense of the image, especially PSFR-GAN, but the above methods are not good in the whole effect of the miner face image. PSFR-GAN uses the method of prior knowledge to generate the miner's face image, which is relatively blunt and distorted, and there are artifacts. DFDNet adopts the restoration method of the component dictionary, which has certain advantages in the restoration of facial components, but it does not perform well in terms of facial wrinkles and authenticity of miners' facial features.

Looking at Figure 10, under low-light conditions, the images restored by the proposed model are better in terms of overall color, image fidelity, and authenticity. Although DFDNet adopts the method of using a component dictionary, the facial organs of the miners' face images under low light are seriously degraded, and the overall fidelity of the final output images is poor. The miners' face images restored by PSFR-GAN are too smooth and have a certain sense of blur. The image output by GFP-GAN is not ideal in terms of color and detail texture. In Figure 10a, the restoration of the beard is not realistic enough and appears blurred. The restoration of the beard by the proposed model is more realistic and better. In Figure 10b, the advantages of the proposed model are that the images are more realistic, there is no blurring, and more fold information is preserved in the eyes.



(a) Normal light miner image 1

(b) Nomer light miner image 2

FIGURE 9. Comparison diagram of different experiments under normal illumination.

TABLE 4. Experimental schemes of four network structures.

Program	U-Net	UNet++	SE
A (CS-SFT + U-Net)	✓		
B (SE-CSSFT + U-Net）	✓		✓
C (CS-SFT + UNet++）		✓	
D (SE-CSSFT + UNet++）		✓	✓

TABLE 5. Experimental results of four network structure schemes.

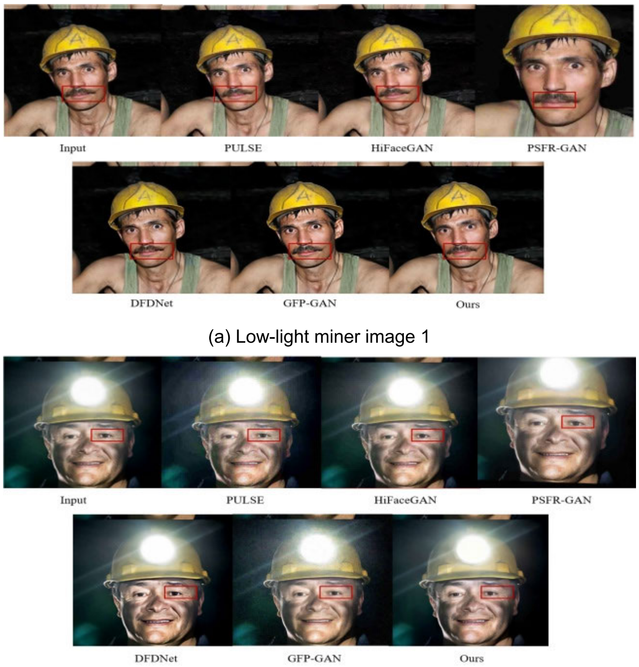
Program	LPIPS ↓	FID ↓	NIQE ↓
A (CS-SFT + U-Net)	0.4046	48.69	5.517
B (SE-CSSFT + U-Net）	0.3961	47.90	5.414
C (CS-SFT + UNet++）	0.3906	47.49	5.348
D (SE-CSSFT + UNet++）	0.3827	46.51	5.206

D. ABLATION EXPERIMENT

The ablation experiment included two parts, and the first part was mainly conducted as illustrated in Table 4.

Scheme A adopted the U-Net structure of the degradation removal module in GFP-GAN using CS-SFT; Scheme B integrated the channel attention mechanism into the channel-splitting spatial feature transformation layer based on Scheme A using SE-CSSFT; Scheme C used the UNet++ structure as the degradation removal module and used the CS-SFT layer; and Scheme D was based on Scheme C using the SE-CSSFT layer. Table 5 illustrates in the experimental results.

From Table 5, upon examination of Scheme B, it becomes evident that within the original GFP-GAN network, the use of SE-integrated SE-CSSFT resulted in a decrease of 0.0085, 0.79, and 0.103 in LPIPS, FID, and NIQE compared to Scheme A. When replacing U-Net with UNet++, LPIPS, FID, and NIQE for Scheme C were reduced by 0.014, 1.2, and 0.169, respectively, compared to Scheme A. Using SE-CSSFT with UNet++, LPIPS, FID, and NIQE for



(a) Low-light miner image 1

(b) Low-light miner image 2

FIGURE 10. Comparison diagram of different experiments under low light.

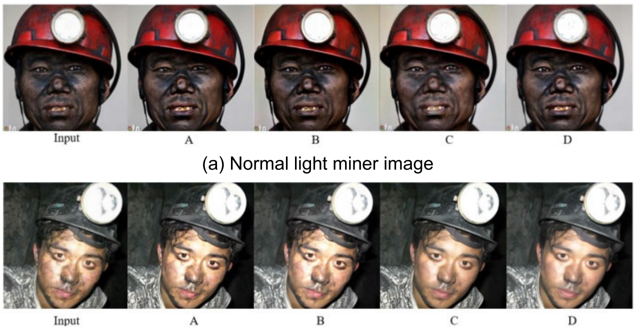
Scheme D were reduced by 0.0134, 1.39, and 0.208, respectively, compared to Scheme B. The degradation removal modules are all UNet++ structures, and Scheme D adopted SE-CSSFT, which reduced LPIPS, FID, and NIQE by 0.0079, 0.98, and 0.142, respectively, compared to Scheme C.

Two images were selected, including miners' face images under normal and low-light conditions, and the output images were visualized under different network structures, as shown in Figure 11.

Observing Figure 11, based on Scheme A, by incorporating SE into CS-SFT, Scheme B reduces artifacts visually and focuses on the details of facial features. In Figure 11a, B has a clearer contour on the teeth than A. Option C replaces U-Net with UNet++, increasing the extraction of deep features. In Figure 11b, the overall effect of the face is significantly improved. On the basis of Scheme C, SE is further integrated to improve the fidelity and authenticity of the miners' facial images generated by Scheme D.

In summary, compared to the U-Net structure, the nested dense skip connections of UNet++ can extract deep and shallow features to a greater extent, and the integration of a channel attention mechanism effectively enhances the performance of prior features within the network model, which can help better restore miners' facial images.

The second ablation experiment was based on the proposed model in this paper, using a layer-by-layer removal method to further study the influence of prior features and the channel attention mechanism on the restoration of miners' facial images. Table 6 illustrates in the experimental results.



(a) Normal light miner image

(b) Low-light miner image

FIGURE 11. Visualization of different experimental schemes.

TABLE 6. Experimental schemes of different network structures.

Program	UNet++	Channel-Split	SE
E (SE-CSSFT + UNet++)	✓	✓	✓
F (SE-SFT + UNet++）	✓		✓
G (CS-SFT + UNet++）	✓	✓	

TABLE 7. Experimental results of different network structure schemes.

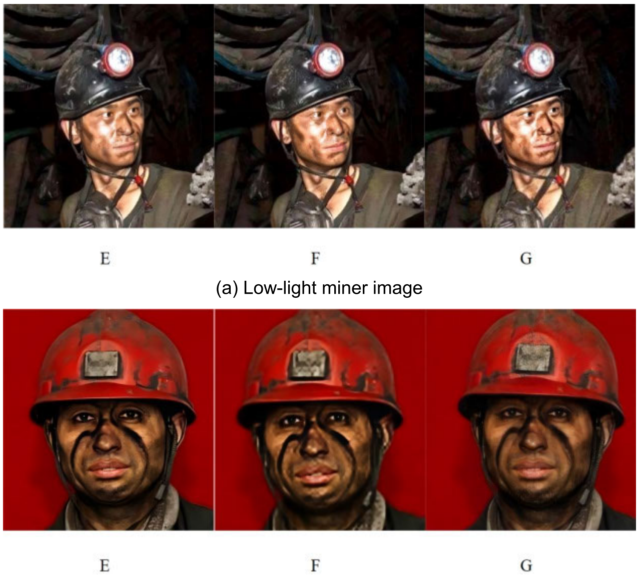
Program	LPIPS ↓	FID ↓	NIQE ↓
E (SE-CSSFT + U-Net++)	0.3827	46.51	5.206
F (SE-SFT + UNet++）	0.4029	47.49	5.348
G (CS-SFT + UNet++）	0.3906	47.08	5.319

The second ablation experiment was conducted on the basis of the degradation removal module; it was a UNet++ network. Scheme E uses SE-CSSFT, where one branch is directly transmitted to save more prior information and the other branch uses multi-resolution feature $F_{spatial}$ to modulate prior feature F_{GAN}. Finally, the features of the two channels are further concatenated and integrated into the channel attention mechanism. Scheme E is the proposed scheme in this paper.

Scheme F adopts a single-channel feature transformation layer, removing the channel split of Scheme E, and uses all the feature information from the multi-resolution spatial feature FS to modulate the features of StyleGAN2, changing the design of half modulation. Scheme G removes SE from E and adopts the channel-split spatial feature transformation layer in the original network. Table 7 illustrates in the experimental results.

Observing Table 7, when only channel split is removed, the LPIPS, FID, and NIQE of Scheme F increase by 0.0202, 0.98, and 0.142 compared to Scheme E. When removing SE again, the LPIPS, FID, and NIQE of Scheme G are improved by 0.0079, 0.57, and 0.113 compared to Scheme E, respectively, resulting in a decrease in image perception quality.

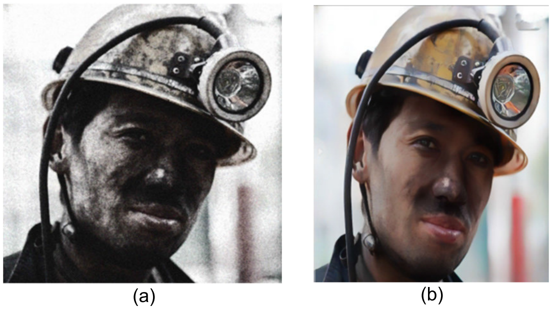
Two miners' facial images were selected, including one under normal lighting and one under low lighting, and the experimental results of E, F, and G schemes were visualized,



(a) Low-light miner image

(b) Normal light miner image

FIGURE 12. Visualization of different experimental schemes.


(a)

(b)

FIGURE 13. This proposed method restores grayscale images.

as shown in Figure 12. Scheme E showed reduced artifacts and blurring and achieved a much better balance between authenticity and fidelity in the final output image, resulting in higher image quality.

In summary, the attention channel mechanism adds coefficients to the two channels in the CS-SFT layer, enabling multi-resolution special FS to better modulate the spatial features of the prior feature FGAN. SE-CSSFT effectively elevates the perceptual quality of miners' facial images. Similarly, the use of two branches in the SE-CSSFT layer also makes prior features play a more prominent role in miner face restoration tasks.

IV. DISCUSSION

The algorithm developed in this study is based on miner face images, and has shown good adaptability and effectiveness in dealing with dark and dusty skin on human faces, ensuring high visual consistency between the output image and the original input. However, it is worth noting that the color of the dataset in this study is monotonous. For grayscale input miner images, due to the lack of color information and the fact that the miner's face often obscures dust, the algorithm may produce certain deviations when restoring facial colors, as shown in Figure 13. Secondly, it is also limited by the insufficient number of datasets, and in the future, we shall consider using plug and play sampling methods to expand the dataset [27]. Therefore, this study needs to further construct a more diverse, balanced, and data intensive training dataset.

The image examined in this paper is a miner's face image that exhibits multiple degradation factors. The pre-trained GAN and input image features are employed to modulate the image, and the output is constrained using reconstruction loss and identity preservation loss to ensure fidelity to the input. When the degradation of actual images is severe, the method may also produce minor artifacts when restoring facial details, a challenge that future research will need to address.

V. CONCLUSION

Real-world Super-Resolution (Real-SR) methods focus on dealing with diverse real-world images and have attracted increasing attention in recent years [28], [29]. In this paper introduces a novel and enhanced GFP-GAN approach specifically tailored for blind restoration of miners' facial images. Firstly, the dataset was expanded, and a degradation model was established to fit low-quality miner facial images. Secondly, to enhance the model's capacity for feature extraction, a UNet++ network with nested and dense skip connections was used as the degradation removal module in the model, helping to better capture deep and shallow features of different resolutions. The proposed model utilized latent code mapping alongside a multi-channel-split spatial segmentation feature conversion layer to seamlessly integrate the degradation removal module with the retrained facial GAN, an attention mechanism was introduced into the spatial feature conversion layer, and SE-CSSFT was used to make prior features play a greater role, ultimately achieving much better balance between fidelity and authenticity in the output image. In addition, the loss function added reconstruction loss, facial component loss, and identity-preserving loss, which guided the model to complete the blind image restoration task and achieved a better blind restoration effect.

In subsequent research, due to the need to design additional modules to utilize GAN prior information in order to generate prior information, the model may also waste computing resources to some extent. Therefore, this study shall consider using generative diffusion prior methods to reduce the complexity of the network [30].

AUTHOR CONTRIBUTION

Conceptualization: Xianming Zhang and Jiaojiao Feng; methodology: Xianming Zhang and Jiaojiao Feng; software: Xianming Zhang and Jiaojiao Feng; validation; Xianming Zhang and Jiaojiao Feng; formal analysis: Xianming Zhang; investigation: Xianming Zhang and Jiaojiao Feng; resources: Xianming Zhang; data curation: Jiaojiao Feng; writing—original draft preparation: Jiaojiao Feng and Xianming

Zhang; writing—review and editing: Xianming Zhang; visualization, Xianming Zhang; supervision: Xianming Zhang; project administration: Xianming Zhang. All authors have read and agreed to the published version of the manuscript.

ACKNOWLEDGMENT

The authors would like to thank the editor and the anonymous reviewers for their constructive comments and suggestions, which improved the quality of this article. *(Xianming Zhang and Jiaojiao Feng are co-first authors.)*

REFERENCES

[1] T. Wang, K. Zhang, X. Chen, W. Luo, J. Deng, T. Lu, X. Cao, W. Liu, H. Li, and S. Zafeiriou, "A survey of deep face restoration: Denoise, super-resolution, deblur, artifact removal," 2022, *arXiv:2211.02831*.

[2] J. Li, Z. Zhang, X. Liu, C. Feng, X. Wang, L. Lei, and W. Zuo, "Spatially adaptive self-supervised learning for real-world image denoising," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 9914–9924.

[3] X. Yu, B. Fernando, B. Ghanem, F. Porikli, and R. Hartley, "Face super-resolution guided by facial component heatmaps," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 219–235.

[4] X. Hu, W. Ren, J. Yang, X. Cao, D. Wipf, B. Menze, X. Tong, and H. Zha, "Face restoration via plug-and-play 3D facial priors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 12, pp. 8910–8926, Dec. 2022, doi: 10.1109/TPAMI.2021.3123085. https://doi.org/10.1109/tpami.2021.3123085.

[5] C. Chen, X. Li, L. Yang, X. Lin, L. Zhang, and K. K. Wong, "Progressive semantic-aware style transformation for blind face restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 11891–11900.

[6] Y. Yu, P. Zhang, K. Zhang, W. Luo, C. Li, Y. Yuan, and G. Wang, "Multi-prior learning via neural architecture search for blind face restoration," 2022, *arXiv:2206.13962*.

[7] X. Li, M. Liu, Y. Ye, W. Zuo, L. Lin, and R. Yang, "Learning warped guidance for blind face restoration," in *Proc. Eur. Conf. Comput. Vision*, 2018, pp. 278–296.

[8] X. Li, W. Li, D. Ren, H. Zhang, M. Wang, and W. Zuo, "Enhanced blind face restoration with multi-exemplar images and adaptive spatial feature fusion," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2703–2712.

[9] X. Li, C. Chen, S. Zhou, X. Lin, W. Zuo, and L. Zhang, "Blind face restoration via deep multi-scale component dictionaries," in *Proc. 16th Eur. Conf. Comput. Vis. (ECCV)*, Aug. 2020, pp. 399–415.

[10] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 12, pp. 4217–4228, Dec. 2021.

[11] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of StyleGAN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 8107–8116, doi: 10.1109/CVPR42600.2020.00813.

[12] S. Menon, A. Damian, S. Hu, N. Ravi, and C. Rudin, "PULSE: Self-supervised photo upsampling via latent space exploration of generative models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 2434–2442.

[13] J. Gu, Y. Shen, and B. Zhou, "Image processing using multi-code GAN prior," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3009–3018.

[14] X. Wang, Y. Li, H. Zhang, and Y. Shan, "Towards real-world blind face restoration with generative facial prior," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Nashville, TN, USA, Jun. 2021, pp. 9164–9174.

[15] T. Yang, P. Ren, X. Xie, and L. Zhang, "GAN prior embedded network for blind face restoration in the wild," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Nashville, TN, USA, Jun. 2021, pp. 672–681.

[16] F. Zhu, J. Zhu, W. Chu, X. Zhang, X. Ji, C. Wang, and Y. Tai, "Blind face restoration via integrating face shape and generative priors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 7652–7661.

[17] W. Z. Shao, J. J. Xu, L. Chen, Q. Ge, L. Q. Wang, B. K. Bao, and H. B. Li, "On potentials of regularized Wasserstein generative adversarial networks for realistic hallucination of tiny faces," *Neurocomputing*, vol. 364, pp. 1–15, Sep. 2019.

[18] F. Shiri, X. Yu, F. Porikli, R. Hartley, and P. Koniusz, "Identity-preserving face recovery from stylized portraits," *Int. J. Comput. Vis.*, vol. 127, pp. 863–883, Sep. 2019.

[19] L. Yang, S. Wang, S. Ma, W. Gao, C. Liu, P. Wang, and P. Ren, "HiFaceGAN: Face renovation via collaborative suppression and replenishment," in *Proc. 28th ACM Int. Conf. Multimedia*, Oct. 2020, pp. 1551–1560.

[20] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf.*, 2015, pp. 1–23.

[21] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested U-Net architecture for medical image segmentation," in *Proc. 4th Int. Workshop*, Sep. 20, 2018, pp. 1–23.

[22] X. Wang, K. Yu, C. Dong, and C. Change Loy, "Recovering realistic texture in image super-resolution by deep spatial feature transform," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 606–615.

[23] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[24] R. Huang, S. Zhang, T. Li, and R. He, "Beyond face rotation: Global and local perception GAN for photorealistic and identity preserving frontal view synthesis," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2458–2467.

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

[26] M. M. Fréchet, "Sur quelques points du calcul fonctionnel," in *Rendiconti Circolo Matematico Palermo*, vol. 22, no. 1, pp. 1–72, 1906.

[27] Y. Ma, H. Yang, W. Yang, J. Fu, and J. Liu, "Solving diffusion ODEs with optimal boundary conditions for better image super-resolution," in *Proc. Int. Conf. Learn. Represent.*, 2024, pp. 1–12.

[28] W. Zhang, X. Li, X. Chen, X. Zhang, Y. Qiao, X. M. Wu, and C. Dong, "SEAL: A framework for systematic evaluation of real-world super-resolution," in *Proc. Int. Conf. Learn. Represent.*, Jan. 2024, pp. 1–16.

[29] Z. Wan, B. Zhang, D. Chen, P. Zhang, D. Chen, F. Wen, and J. Liao, "Old photo restoration via deep latent space translation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 2, pp. 2071–2087, Feb. 2023, doi: 10.1109/TPAMI.2022.3163183.

[30] X. Chen, J. Tan, T. Wang, K. Zhang, W. Luo, and X. Cao, "Towards real-world blind face restoration with generative diffusion prior," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 1, no. 1, pp. 1–23, Apr. 2021, doi: 10.1109/TCSVT.2024.3383659.

XIANMING ZHANG received the B.S. and master's degrees in software engineering from Jilin University, Changchun, China, in 1994 and 2009, respectively. Since 2022, he has been a Senior Engineer with the School of Computer and Information Engineering, Heilongjiang University of Science and Technology. His research interests include video tracking and monitoring, image processing, and big data applications.

JIAOJIAO FENG received the B.S. and master's degrees in computer application technology from the Heilongjiang University of Science and Technology, Harbin, China, in 2020 and 2023, respectively. Since 2023, she has been a Technology Developer with the Financial Technology Department, Jiangsu Changshu Rural Commercial Bank Company Ltd. Her research interests include medical image analysis and computer vision.