Complex & Intelligent Systems (2023) 9:1439–1453
https://doi.org/10.1007/s40747-022-00865-9

ORIGINAL ARTICLE

Multi-scale progressive blind face deblurring

Hao Zhang[1] · Canghong Shi[2] · Xian Zhang[1] · Linfeng Wu[1] · Xiaojie Li[1] · Jing Peng[1] · Xi Wu[1] · Jiancheng Lv[3]

Received: 19 April 2022 / Accepted: 27 August 2022 / Published online: 13 September 2022
© The Author(s) 2022

Abstract

Blind face deblurring aims to recover a sharper face from its unknown degraded version (i.e., different motion blur, noise). However, most previous works typically rely on degradation facial priors extracted from low-quality inputs, which generally leads to unlifelike deblurring results. In this paper, we propose a multi-scale progressive face-deblurring generative adversarial network (MPFD-GAN) that requires no facial priors to generate more realistic multi-scale deblurring results by one feed-forward process. Specifically, MPFD-GAN mainly includes two core modules: the feature retention module and the texture reconstruction module (TRM). The former can capture non-local similar features by full advantage of the different receptive fields, which facilitates the network to recover the complete structure. The latter adopts a supervisory attention mechanism that fully utilizes the recovered low-scale face to refine incoming features at every scale before propagating them further. Moreover, TRM extracts the high-frequency texture information from the recovered low-scale face by the Laplace operator, which guides subsequent steps to progressively recover faithful face texture details. Experimental results on the CelebA, UTKFace and CelebA-HQ datasets demonstrate the effectiveness of the proposed network, which achieves better accuracy and visual quality against state-of-the-art methods.

Keywords Blind face deblurring · Receptive field · Supervisory attention · High-frequency texture

Introduction

Face deblurring is the task of recovering a sharp face with both edge structures and realistic details from the low-quality counterparts suffering from unknown degradation, such as different motion blur [35,45] and noise [62]. The degradation process is generally defined as:

$$I_B = K * I_S + N, \tag{1}$$

where I_B, I_S, K, and N represent the blurry image, sharp latent image, blur kernel and noise, respectively; $*$ represents the convolution. Given I_B, in face deblurring, the objective is to estimate the underlying sharp face image I_S. Moreover, the deblurring techniques can be divided into non-blind and blind deblurring methods according to whether the blur kernel K is known [31]. The latter is more challenging than the former, because it is a typical ill-posed problem with infinite feasible solution [57]. Therefore, most researchers are currently working on blind deblurring techniques [7,13,53,57,61], and this paper is no exception. In addition, accurate and fast resolution of the deblurring problem is the key to computer vision and image processing, which can produce tremendous commercial value [5].

Hao Zhang and Canghong Shi contributed equally to this work.

Xiaojie Li and Jing Peng are co corresponding authors in this work.

✉ Xiaojie Li
lixj@cuit.edu.cn

✉ Jing Peng
pengj@cuit.edu.cn

Hao Zhang
cuit_zhanghao@163.com

Canghong Shi
canghongshi@163.com

Xian Zhang
zhangxian318@163.com

Linfeng Wu
3027484833@qq.com

Xi Wu
xi.wu@cuit.edu.cn

Jiancheng Lv
lvjiancheng@scu.edu.cn

[1] School of Computer Science, Chengdu University of Information Technology, Chengdu 610225, China

[2] Xihua University, Chengdu 610039, China

[3] Sichuan University, Chengdu 610065, China

Deep learning has brought significant advances for general image deblurring tasks [7,8,31,32,59,61]. There are three main research trends [17] for this task: CNN-based methods, GAN-based methods, and prior-guided methods. SPARNet [4] introduced a facial attention unit and a spatial attention mechanism based on convolutional neural networks (CNN) to generate high-quality outputs. Kupyn et al. [31,32] proposed a deblurring method based on generative adversarial network (GAN) and demonstrated the potential of GAN for deblurring tasks. Inspired by the benefits of GAN, some progressive deblurring networks have also achieved great success for single image deblurring [7,8,61]. However, these multi-stage progressive networks lead to excessive network size and depth and difficulty maintaining a complex balance between spatial details and high-level background information [11]. Although the above methods achieved better performance for image deblurring, due to the particularity of face image itself, blind face deblurring has the following challenges [17,21]: (1) how to generate fine and realistic facial details; (2) how to achieve a good balance between visual quality and fidelity. Thus, researchers generally proposed two schemes based on Prior-guided to overcome these challenges. The first scheme considered utilizing face-specific priors on face deblurring, such as sparsity [43,44], patched similarity [47], face landmarks [2,6], face semantic labels [45,57], and face component heat maps [60], and showed the significance of these priors in face restoration. However, most of these priors inevitably suffer from degradation [49] estimated from low-quality inputs in real-world scenarios [28–30]. Although the above priors could guide facial recovery, they contain limited texture features for recovering facial detail information (e.g., hair texture, tooth contours, facial wrinkles) [56]. The second scheme is to build a reconstruction-oriented high-quality dictionary containing rich high-quality face priors for face reconstruction, leading to better reconstruction results [34,52]. However, due to the limited dictionary capacity, these methods would lose the richness and diversity of the reconstructed facial details [49]. Thus, restoring richer and faithful facial texture details with reduced reliance on priors becomes a new challenge for blind face deblurring [56].

In this paper, we propose a multi-scale progressive face-deblurring generative adversarial network (MPFD-GAN) to recover sharper faces without requiring extra inputs (i.e., face priors, facial component dictionaries). Its generator includes three parts: the encoding process, the center process, and the decoding process (pyramid reconstruction process). In the first part, we filter out noisy information and provide abundant contextual features and textural details to provide fine-grained features for subsequent steps. In the center part, we design a feature retention module (FRM). It captures broad contextual information and richer receptive field information by multiple dilated convolutions with different dilated rates. This could enhance and generate high-resolution features with rich spatial details. Moreover, FRM adding channel attention between dilated convolutions could avoid artifacts caused by the fusion of multiple receptive field information. In the last part, we propose the texture reconstruction module (TRM) and plug it into all pyramid levels to enable progressively replenishing face texture details. It consists of supervised attention and facial feature-guided reconstruction. The former computes attention maps using the previous step recovered low-scale face with the guidance of ground truth and then reweights the input features to obtain fine features by these maps. The latter extracts the high-frequency texture information from the recovered low-scale face by the Laplace operator, which guides subsequent steps to recover more detailed face texture details.

Experimental results on three publicly available datasets [37,38,65] illustrate that the proposed method can recover high-quality and detailed information-rich face images. The main contribution of this work can be summarized as follows:

- We propose an effective blind face deblurring network called MPFD-GAN, which can recover multi-scale sharp faces from blurry faces without requiring extra inputs (i.e., face priors, facial component dictionaries).
- We design a feature preservation module (FRM), which captures richer receptive field information to recover the complete structure from a blurry image, and propose the texture reconstruction module (TRM) which generates texture guidance with rich facial details to help reconstruct facial texture details.
- Extensive experiments have demonstrated that our break-method achieves better visual effect and quantitative metrics than the state-of-the-art techniques on the CelebA [38], UTKFace [65] and CelebA-HQ [37] datasets.

Related work

Image deblurring

Earlier blind deblurring methods deconvolute the estimated blur kernel with the blurry image to obtain a sharp image [41,54,55]. Although these traditional methods have a certain deblurring effect, the deblurring process of each image takes a lot of time [12].

With the development of deep learning, more excellent algorithms have been applied to image deblurring tasks [7,8,32,59,61]. Kupyn et al. [31] proposed a GAN-based end-to-end image deblurring method named DeblurGAN. It uses PatchGAN [20] as the discriminator and optimizes the network using perceptual loss [25] and adversarial loss. However, these face images recovered in this way show a checkerboard artifact and a poor deblurring effect (Fig. 1b).

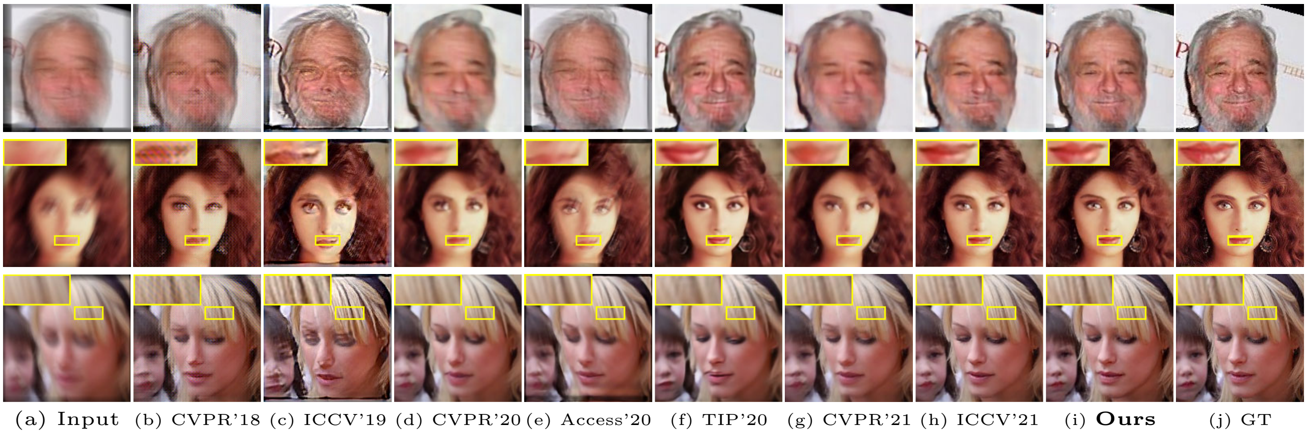

Fig. 1 Comparisons with state-of-the-art face deblurring methods: **b** DeblurGAN [31], **c** DeblurGAN-v2 [32], **d** DMPHN [8], **e** SIUN [59], **f** UMSN [57], **g** MPRNet [61] and **h** MIMO-UNet [7] on the CelebA dataset. Our method generates the complete structure and more realistic texture details, achieving the best visual quality

Subsequently, they proposed a new deblurring method named DeblurGAN-v2 [32] based on C-GAN [39]. DeblurGAN-v2 introduces the feature pyramid network as the core module to accelerate the speed of deblurring. However, their recovered face images looked quite inharmonious (Fig. 1c), and the experimentally obtained evaluation metrics also performed poorly (Table 2). Ye et al. [59] proposed a scale-iterative upscaling network to recover sharp images. However, this method failed to achieve better visual quality on face datasets (Fig. 1e), and the iterative calculation requires a large amount of memory. Moreover, Arora et al. [61] proposed a multi-stage network architecture to solve the complex image restoration problem. In this method, they first learn context-dependent features using an encoder–decoder architecture and then fuse them with local information extracted by high-resolution branching with an attention mechanism. Horizontal connections are also added into the feature processing modules at each stage to avoid information loss. However, this approach will consume a lot of computational costs and cannot be popularized well in the real world [11]. Moreover, these recovered images are relatively coarse, causing a lot of information loss in key areas such as the mouth (Fig. 1g), and they cannot recover natural and harmonious face images. Recently, Cho et al. [7] have reviewed the image restoration scheme from coarse to fine and proposed a multiple-input multiple-output deblurring structure. This structure uses asymmetric feature fusion techniques to fuse feature information at different scales. However, this method cannot restore the complete structure and faithful details (e.g., lips) when facing heavily degraded images (Fig. 1h).

Face deblurring

Existing deep learning-based blind deblurring methods can deal well with blurry images in the real world [59]. However, various existing model architectures cannot solve the deblurring problem in all settings. In face deblurring tasks, it is usually necessary to use face-specific priors to guide the recovery of face images [30,45,67]. These priors are divided into two main categories: geometric (e.g., face semantic labels) and reference priors [49]. Recently, Yasarla et al. [58] proposed to help achieve better performance in face deblurring tasks using facial semantic labels. This method is divided into two stages: (1) the first stage is to feed the blurry face images into a segmentation network to obtain different semantic labels for faces; (2) the second stage is to feed the blurry face images and the semantic labels into a multi-stream semantic network to process regions belonging to each semantic category independently and learn to combine the information from different regions to output as the final deblurring result. But these semantic labels are estimated from blurred face images that suffer from severe degradation and inevitably degrade in realistic scenes. They mainly focused on geometric constraints and did not consider how to recover critical areas of the face [49]. The recovery results obtained by this method are still blurry in critical areas such as the hair (Fig. 1f). Instead, our proposed method does not rely on estimating geometric priors from degraded blurry face images.

Methodology

Our goal is to recover high-quality faces with richer authentic details through a multi-scale progressively deblurring method (MPFD-GAN) without requiring extra inputs (i.e., face priors, facial component dictionaries). Figure 2 illustrates the overview structure of our network, which contains two key modules: FRM and TRM. In this section, we first detail the critical modules of MPFD-GAN and then describe its loss functions.

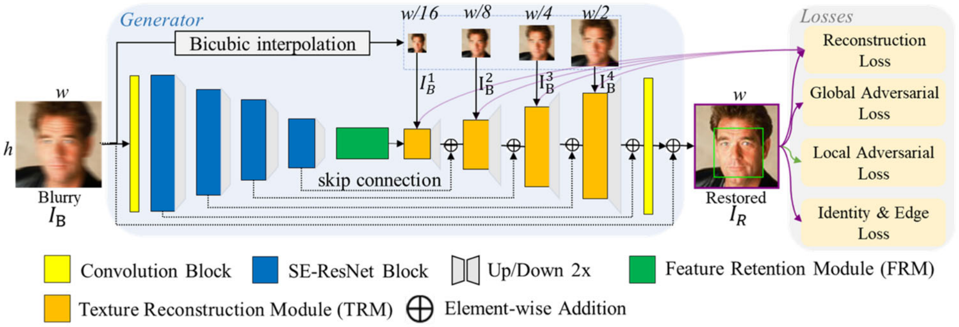

Fig. 2 The architecture of the proposed network. The generator includes three parts: the encoding process (down-sampling), the center process, and the decoding process (up-sampling) and contains two key modules: FRM and TRM. It takes I_B and $I_B^i (i = 1, 2, 3, 4)$ as input to produce the multi-scale restoration result I_R and I_R^i. Local adversarial loss is adopted across the face region (labeled in green), while the remaining losses are adopted across the entire image (labeled in purple)

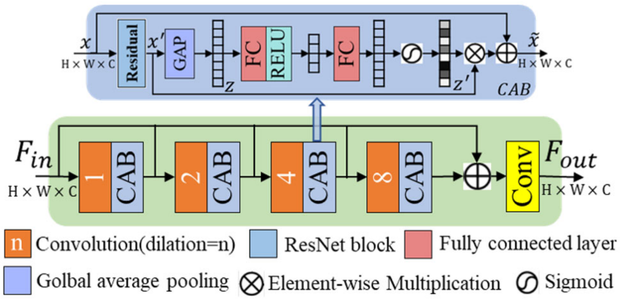

Fig. 3 The architecture of feature retention module (FRM)

Feature retention module (FRM)

To retain broad contextual information without degrading spatial information, we design a feature retention module (FRM) between the encoder and the decoder (see the green block in Fig. 2). It consists of the dilated convolution blocks and channel attention blocks (CAB) [16] with dilation rates of 1, 2, 4 and 8, respectively (see Fig. 3).

Unlike the other dilated convolution blocks [66,68], we add CAB between each two successive dilated convolution blocks to avoid artifacts caused by the fusion of multiple receptive field information. This problem can be illustrated in Fig. 4, where FRM with CAB (third row) gets sharper feature maps than FRM without CAB (second row). Moreover, we add skip connection to ensure the proposed model makes full the information from the shallow features. In this way, our network can obtain richer receptive fields without changing the feature size and generates high-resolution features with rich spatial information. Formally,

$$F_{dr\_1} = CAB\left(\text{Conv}_{dr\_1}\left(F_{\text{in}}\right)\right), \tag{2}$$

$$F_{dr\_2} = CAB\left(\text{Conv}_{dr\_2}\left(F_{dr\_1}\right)\right), \tag{3}$$

$$F_{dr\_4} = CAB\left(\text{Conv}_{dr\_4}\left(F_{dr\_2}\right)\right), \tag{4}$$

$$F_{dr\_8} = CAB\left(\text{Conv}_{dr\_8}\left(F_{dr\_4}\right)\right), \tag{5}$$

$$F_{\text{out}} = Conv\left(F_{\text{in}} + F_{dr\_1} + F_{dr\_2} + F_{dr\_4} + F_{dr\_8}\right), \tag{6}$$

where F_{in} and F_{out} denote the input and output features of the FRM. $Conv_{dr\_i}$ and F_{dr_i} represent dilated convolution operations and their corresponding output features with different dilated rates ($i = 1, 2, 4$ and 8, respectively). CAB represents channel attention block (CAB), which can be represented by a figure and the following equations:

$$z_c = GAP(x_c^{'}) = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} x_c^{'}(i, j), \tag{7}$$

$$z^{'} = \sigma\left(W_2 \delta\left(W_1 z\right)\right), \tag{8}$$

$$\tilde{x} = x^{'} \cdot z^{'} + x, \tag{9}$$

where $x^{'} \in \mathbb{R}^{H \times W \times C}$ represents the features obtained from the input $x \in \mathbb{R}^{H \times W \times C}$ after a ResNet block [14], $H \times W$ denotes the spatial dimension and C is the number of channels. $x_c^{'} \in \mathbb{R}^{H \times W}$ represents the features of the cth channel of $x^{'}$. $z^{'} \in \mathbb{R}^{1 \times 1 \times C}$ represents a one-dimensional vector composed of the feature weights of each channel. δ and σ are ReLU and sigmoid activation functions, respectively. W_1 and W_2 represent the parameters of the first and second fully connected layers, respectively. $\tilde{x} \in \mathbb{R}^{H \times W \times C}$ represents the output of CAB.

Texture reconstruction module (TRM)

To enable the model to progressively replenish face texture details, we propose a texture reconstruction module (TRM) and plug it into every pyramid level (different scales in the reconstruction process, see Fig. 2). The structural details of the TRM are shown in Fig. 5, where we restrict the reconstructed image I_R^i at each pyramid level to be infinitely close

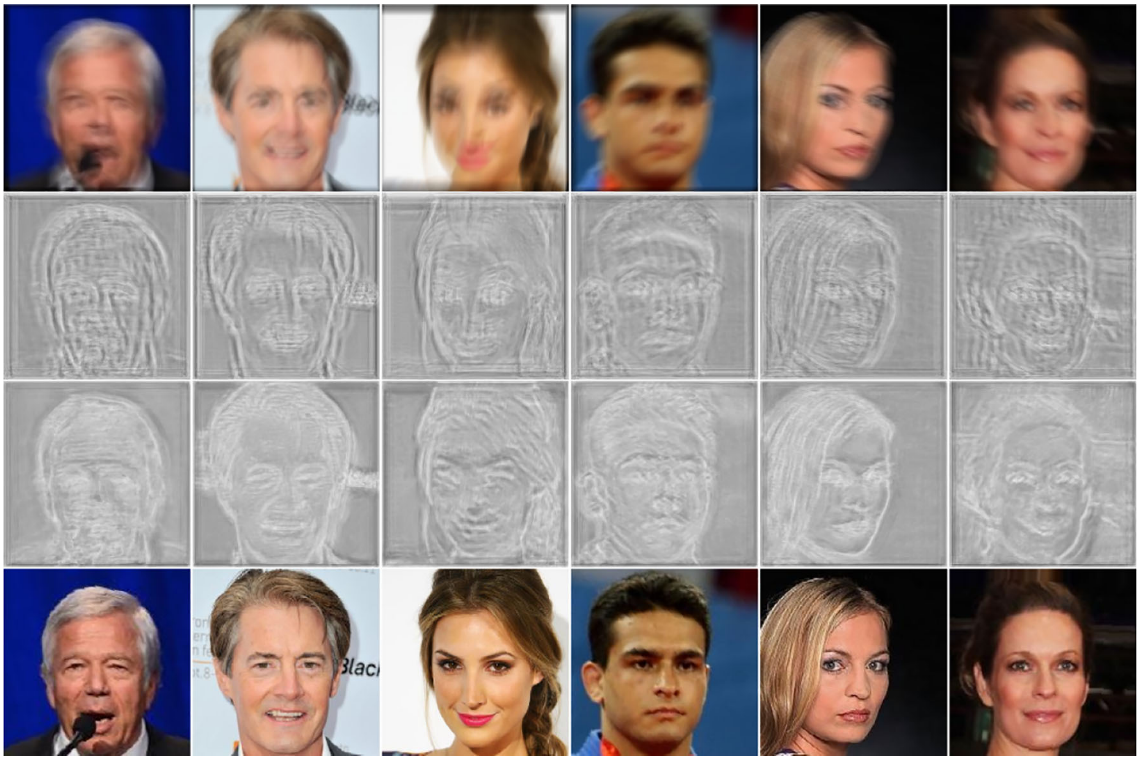

Fig. 4 FRM's output feature map. The first row is the blurry image of the model input, the second and third rows show the FRM's output feature map without and with CAB, respectively, and the last row is the ground truth image

to the pyramid of the ground truth image I_G^i $(i = 1, 2, 3, 4)$ in terms of realism and fidelity. The contribution of TRM is twofold. First, we compute attention maps by the reconstructed low-resolution image supervised by ground truth and use these maps to reweight the input features F_{in} to generate F_{SA} containing more beneficial features. Second, we extract the high-frequency texture information from the reconstructed low-resolution sharp face I_R^i by the Laplace operator as facial texture guidance $I_{t.g}^i$. This facial texture guidance can provide sufficient texture information for the subsequent image reconstruction process and help recover the reality texture details of the face. As illustrated in Fig. 6, we can visually observe that the texture guide contains more and more facial texture information when the scale gradually increases. Finally, we fuse the features $(F_{SA}, I_R^i, I_{t.g.}^i)$ following the channel attention block (CAB) [16] to suppress the less informative channels at the current pyramid level and only allow the beneficial ones to pass to the next step.

Formally, as shown in Fig. 5, TRM first estimates the residual image $I_{R-B}^i \in \mathbb{R}^{H\times W\times 3}$ by convolving the input features $F_{\text{in}} \in \mathbb{R}^{H\times W\times C}$ with a 3×3 convolution kernel, where $H \times W$ represents the size of the features and C represents the number of channels. The element-wise summation of the residual image I_{R-B}^i with the input low-resolution blurry image $I_B^i \in \mathbb{R}^{H\times W\times 3}$ can reconstruct the low-resolution sharp image $I_R^i \in \mathbb{R}^{H\times W\times 3}$. We provide the ground truth image of the corresponding size to constrain the reconstructed $I_R^i \in \mathbb{R}^{H\times W\times 3}$, which improves the fidelity of the reconstructed results. We further perform the convolution operation and sigmoid activation on I_R^i to obtain the corresponding per-pixel attention maps $\in \mathbb{R}^{H\times W\times C}$. It can help TRM to re-calibrate transformed features F_{in} (after 3×3 convolution) and then get attention-augmented fea-

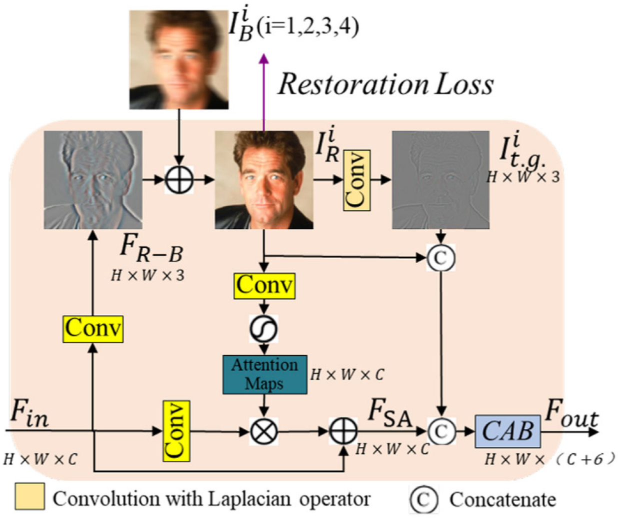

Fig. 5 The architecture of texture reconstruction module (TRM)

tures $F_{SA} \in \mathbb{R}^{H \times W \times C}$. This process can be expressed by the following equation:

$$I_R^i = Conv\left(F_{\text{in}}\right) \oplus \left(I_B^i\right), \tag{10}$$

$$F_{SA} = F_{\text{in}} + Conv(F_{\text{in}}) \otimes Sigmoid(Conv(I_R^i)). \tag{11}$$

Subsequently, the high-frequency face texture guidance $I_{t.g.}^i \in \mathbb{R}^{H \times W \times 3}$ is extracted from I_R^i by the Laplace operator (see Eq. 12). It can provide adequate texture information for the subsequent step to gradually restore higher-resolution sharp image I_R^{i+1}.

$$\begin{aligned} I_{t.g.}^i\left(x, y\right) &= I_R^i\left(x+1, y\right) + I_R^i\left(x-1, y\right) \\ &+ I_R^i\left(x, y+1\right) + I_R^i\left(x, y-1\right) - 4I_R^i\left(x, y\right), \end{aligned} \tag{12}$$

where $I_R^i\left(x, y\right)$ denotes the pixel value on the location (x, y) in the image I_R^i. We set the padding parameter in the Conv2d function [42] to 1 to ensure that $x+1$, $x-1$, $y+1$, and $y-1$ do not cross the boundary.

Finally, we splice the various efficient features obtained earlier, suppress channels with less information in the current scale by the channel attention block (CAB), and pass the beneficial channel information into the next step as the output $F_{\text{out}} \in \mathbb{R}^{H \times W \times (C+6)}$. Formally:

$$F_{\text{out}} = CAB(Concat[F_{SA}, I_R^i, I_{t.g.}^i]). \tag{13}$$

Dual discriminators

Inspired by the work of Zhang et al. [64], we design dual discriminators based on relativistic GAN [26]. It consists of a global discriminator and a local discriminator for the facial region. The global discriminator constrains the overall spatial consistency, while the local discriminator provides fine-grained facial feature distribution to restore photorealistic and harmonious faces.

Specifically, both the recovered sharp image I_R and the ground truth image I_G are passed into the global discriminator to judge the realistic of I_R (see Eq. 16). Then we feed the face regions extracted from I_R and I_G to the local discriminator to judge the authenticity of the recovered face regions (see Eq. 17). At the beginning of training, the global discriminator can ensure that I_R and I_G are consistent in the overall structure. When I_R approaches I_G infinitely in spatial, the local discriminator conduces to recover edges and detailed textures of the facial region of I_R.

Loss function

The training objective of our method is achieved by minimizing the total loss that consists of: (1) reconstruction loss constraints in the restoration results to the pyramid of the ground truth image; (2) adversarial loss for restoring image details and facial realistic textures; (3) edge loss further enhancing the quality and visual realism of facial details; and (4) identity preservation loss to protect the original identity information of the input image.

Reconstruction loss

To obtain recovery results on different scales and strengthen the deblurring ability, both pyramid reconstruction loss and perceptual loss are used. We found that using the robust Charbonnier loss [3] form better handles outliers and improves performance. The reconstruction loss is defined as follows:

$$\begin{aligned} \mathcal{L}_{\text{rec}} = \lambda_{\text{pyramid}} &\sum_{i=1}^{4} \sqrt{\left\|I_R^i - I_G^i\right\|^2 + \varepsilon^2} \\ &+ \lambda_{\text{char}} \sqrt{\|I_R - I_G\|^2 + \varepsilon^2} \\ &+ \lambda_{\text{per}} \sqrt{\left\|\emptyset\left(I_R\right) - \emptyset\left(I_G\right)\right\|^2 + \epsilon^2}, \end{aligned} \tag{14}$$

where I_R and I_G represent the original size recovery result and the corresponding ground truth, respectively. I_R^i and I_G^i represent the low-resolution outputs and their corresponding ground truth (with $\frac{1}{2^{5-i}}$ times the original size). $\emptyset(\cdot)$ denotes the pretrained VGG19 network [46] with ImageNet [9] and we use the first 5 feature maps of maxpooling layers (after activation) [31]. λ_{pyramid} and λ_{per} represent the loss weights of the pyramid reconstruction loss and perceptual loss, respectively. We empirically set $\epsilon = 1e-3$ in all experiments [61].

Adversarial loss

We employ relative adversarial loss [50] to recover sharper contours and detailed texture. Meanwhile, for facial regions, we introduce a local relative discriminator to enhance the model's perception of the facial area. The adversarial loss for the generator is defined as follows:

$$\mathcal{L}_{\text{adv}}^G = \mathcal{L}_{\text{global\_adv}}^G + \mathcal{L}_{\text{local\_adv}}^G, \tag{15}$$

where $L_{\text{global\_adv}}^G$ and $L_{\text{local\_adv}}^G$ represent the global adversarial loss and the local adversarial loss for the generator, respectively. They are defined as follows:

$$\begin{aligned} \mathcal{L}_{\text{global\_adv}}^G = &-\mathbb{E}_{I_G}\left[\log\left(1 - D_{\text{global}}^{Ra}\left(I_G, I_R\right)\right)\right] \\ &- \mathbb{E}_{I_R}\left[\log(D_{\text{global}}^{Ra}(I_R, I_G))\right], \end{aligned} \tag{16}$$

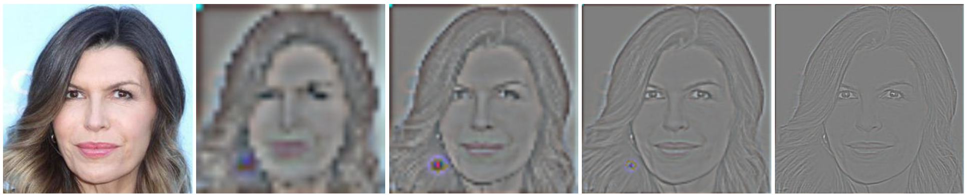
ground truth (a) $I^1_{t.g.}$ (b) $I^2_{t.g.}$ (c) $I^3_{t.g.}$ (d) $I^4_{t.g.}$

Fig. 6 Face texture guidance generated by reconstruction in TRM at different pyramid levels

$$\begin{aligned}\mathcal{L}^{G}_{\text{local\_adv}} = &-\mathbb{E}_{\Delta(I_G)}\left[\log\left(1 - D^{Ra}_{\text{local}}\left(\Delta(I_G), \Delta(I_R)\right)\right)\right] \\ &- \mathbb{E}_{\Delta(I_R)}\left[\log(D^{Ra}_{\text{local}}(\Delta(I_R), \Delta(I_G)))\right],\end{aligned} \tag{17}$$

where $D^{Ra}(I_G, I_R) = \sigma(D(I_G) - \mathbb{E}_{I_R}[D(I_R)])$ is the relativistic average discriminator [50], $\sigma(\cdot)$ is the sigmoid function and $D(\cdot)$ represents the output non-transformed of the discriminator. $\mathbb{E}_{I_G}[\cdot]$ represents the operation of averaging all ground truth in the small batch. $\Delta(\cdot)$ represents face area extractor, i.e., the Dlib C++ library in our implementation.

The adversarial loss for the discriminator is in a symmetrical form:

$$\begin{aligned}\mathcal{L}^{D}_{\text{global\_adv}} = &-\mathbb{E}_{I_G}\left[\log\left(D^{Ra}_{\text{global}}\left(I_G, I_R\right)\right)\right] \\ &- \mathbb{E}_{I_R}\left[\log(1 - D^{Ra}_{\text{global}}(I_R, I_G))\right],\end{aligned} \tag{18}$$

$$\begin{aligned}\mathcal{L}^{D}_{\text{local\_adv}} = &-\mathbb{E}_{\Delta(I_G)}\left[\log\left(D^{Ra}_{\text{local}}\left(\Delta(I_G), \Delta(I_R)\right)\right)\right] \\ &- \mathbb{E}_{\Delta(I_R)}\left[\log(1 - D^{Ra}_{\text{local}}(\Delta(I_R), \Delta(I_G)))\right].\end{aligned} \tag{19}$$

Edge loss

Recent studies have found that adding auxiliary functions in addition to the reconstruction loss would get better deblurring performance [23]. Therefore, we consider edge loss to constrain the differences between frequency spaces so that the final output recovers more realistic high-frequency details. Formally:

$$\mathcal{L}_{\text{edge}} = \sqrt{||\Omega(I_R) - \Omega(I_G)||^2 + \epsilon^2}, \tag{20}$$

where $\Omega(\cdot)$ represents the edge map extracted from the image by the Laplacian operator [22].

Identity preservation loss

To improve the fidelity and authenticity of identity characteristics while deblurring, we refer [18,33] to apply identity preservation loss to our method. We first extract facial features from recovered faces and corresponding ground truth using the pre-trained ArcFace model [10], and calculate the cosine distance between these features as identity preservation loss L_{id}. Formally:

$$\mathcal{L}_{\text{id}} = 1 - \cos(\varphi(I_R), \varphi(I_G)), \tag{21}$$

where $\cos(\cdot, \cdot)$ is the cosine similarity of two vectors, and $\varphi(\cdot)$ represents the face feature extractor, i.e., ArcFace [10].

Algorithm 1 *MPFD-GAN Algorithm.*

```
Require: $\alpha$: learning rate, $m$: batch size, $W_G$: generator parameters,
    $W_{\text{global}}$: global discriminator parameters,$G$: generator, $W_{\text{local}}$: local
    discriminator parameters, $D_{\text{global}}$: global discriminator, $D_{\text{local}}$: local
    discriminator, $I_G$: ground truth, $\Delta$: face extractor.
    while $G$ and $D$ have not converged do
        Sample a batch from the blurry face images $\{I^i_B\}^m_i$
        Restoration multi-scale result by generator:
        $\{I^i_R\}^m_i \sim G\{I^i_B\}^m_i$
        Compute the losses:
        $\mathcal{L}_{\text{rec}}, \mathcal{L}_{\text{edge}}, \mathcal{L}_{\text{id}} \leftarrow \{I^i_G, I^i_R\}^m_{i=1}$ (see Eqs. 14, 20 and 21 )
        Discriminate the distribution of $I^i_R$ real or fake:
        $\mathcal{L}^G_{\text{global\_adv}}, \mathcal{L}^D_{\text{global\_adv}} \leftarrow D_{\text{global}}\{I^i_G, I^i_R\}^m_{i=1}$ (see Eqs. 16 and 18
    )
        Discriminate the distribution of face in $I^i_R$ real or fake:
        $\mathcal{L}^G_{\text{local\_adv}}, \mathcal{L}^D_{\text{local\_adv}} \leftarrow D_{\text{local}}\{\Delta(I^i_G), \Delta(I^i_R)\}^m_{i=1}$ (see Eqs. 17
    and 19 )
        Update $W_G$ by $\mathcal{L}_{\text{rec}}, \mathcal{L}_{\text{edge}}, \mathcal{L}_{\text{id}}, \mathcal{L}^G_{\text{global\_adv}}, \mathcal{L}^G_{\text{local\_adv}}$
        Update $W_{\text{global}}$ and $W_{\text{local}} \sim \{\mathcal{L}^D_{\text{global\_adv}}, \mathcal{L}^D_{\text{local\_adv}}\}$
    end while
```

Overall loss

The whole loss is summarized as follows:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{rec}}\mathcal{L}_{\text{rec}} + \lambda_{\text{adv}}\mathcal{L}^G_{\text{adv}} + \lambda_{\text{edge}}\mathcal{L}_{\text{edge}} + \lambda_{\text{id}}\mathcal{L}_{\text{id}}, \tag{22}$$

where λ_{rec}, λ_{adv}, λ_{edge} and λ_{id} are the tradeoff parameters. Inspired by previous work, [7,36,49,50,61], we empirically set $\lambda_{\text{pyramid}} = 10$, $\lambda_{\text{char}} = 100$, $\lambda_{\text{per}} = 0.5$, $\lambda_{\text{rec}} = 1$, $\lambda_{\text{adv}} = 0.01$, $\lambda_{\text{edge}} = 30$, and $\lambda_{\text{id}} = 2$.

The whole algorithm is summarized in Algorithm 1.

Experiments

Datasets and implementation

Datasets

We separately conduct experiments on three publicly available face datasets to demonstrate the effectiveness of our method: (1) high-resolution face dataset (image size: 256 × 256), which consists of 29,996 blurry face images generated using the CelebA-HQ dataset [37]; (2) middle-resolution face dataset (image size: 192 × 192), which consists of 23,708 blurry face images generated using the UTKFace dataset [65]; (3) low-resolution face dataset (image size: 160 × 160), which consists of 196,973 blurry face images generated using the CelebA dataset [38]. As shown in Table 1, we respectively split the training, validation, and testing datasets according to previous work [57,58,61].

To simulate a real blurring scene, we use a blur kernel of size 25 × 25 to blur the sharp image (CelebA-HQ) with a motion angle of 45 degrees and then add random white Gaussian noise [19]. In addition, we generate 22,127 pairs of clean blurry data based on the UTKFace dataset following the approach of Yasarla et al. [58]. To further demonstrate the performance of our network in realistic scenarios, we applied a more complex blur to the CelebA dataset by referring to the ideas of Boracchi et al. [1]. Specifically, we use the Markov process to generate random motion trajectory vectors [24]. Next, we perform sub-pixel interpolation on the trajectory vector to gain motion blur kernels of size from 13 × 13 to 29 × 29 [31]. Finally, we randomly adopt one to three blur kernels to blur the original images and add random white Gaussian noise to obtain a blurred face dataset that is infinitely close to the real world. This approach can simulate the sudden movements that occur when people press the camera button or try to compensate for camera shake [1].

Implementation

MPFD-GAN is an end-to-end learned method for blind face deblurring. It does not require any pre-trained model to generate facial prior. We train separate models for two different datasets with the following settings. The training batch-size is set as 8 and 14 on the CelebA-HQ and CelebA datasets, respectively. We augment the training data with horizontal flip and use AdamW [27] as the optimizer for a total of 500 epochs. Furthermore, the initial learning rate is set as 1×10^{-3} and gradually decreased to 1×10^{-4} using the Multi-Step LR strategy [49]. With the PyTorch framework [42], all experiments implement on a GeForce RTX 2080Ti GPU.

Table 1 Training, validation and test splitting results for three datasets

DataSet	Train	Valid	Test
CelebA-HQ [37]	27,996	1000	1000
UTKFace [65]	22,127	790	791
CelebA [38]	158,109	19,380	19,484

Comparison with state-of-the-art work

To verify the effectiveness of our method on blind face deblurring tasks, we compare MPFD-GAN with several state-of-the-art methods: DeblurGAN [31], DeblurGAN-v2 [32], DMPHN [8], SIUM [59], UMSN [57], MPRNet [61] and MIMO-UNet [7]. For a fair comparison, we used their published official codes and completely followed their experimental setup. We performed a quantitative and qualitative comparison of the test results of all methods on the CelebA, UTKFace and CelebA HQ datasets, respectively.

Quantitative comparison

Like similar tasks [57,61], we adopted non-reference perceptual metrics FID [15] and NIQE [40] to measure the realness. As for the facial fidelity, we adopt perceptual metrics (LPIPS [63]) and pixel metrics (PSNR and SSIM [51]). LPIPS compares the difference between image patches, PSNR measures the distance between pixels, and SSIM assesses similarity between structure, contrast and luminance. To verify the ability of different methods to recover identity features, we also calculated the face similarity between the recovered results and the corresponding ground truth using cosine similarity as in Deng et al [10]. Furthermore, we first introduce the mean normalized error (MNE) [48] into face deblurring tasks to evaluate the recovery performance of facial contours (according face key point offset distance). Formally:

$$MNE = \frac{\sum_{i}^{N} ||I_{(i)}^{R} - I_{(i)}^{G}||_2}{N \times d_{io}} \times 100\%, \tag{23}$$

where $I_{(i)}^{R}$ is the coordinates of the ith face key point and $I_{(i)}^{G}$ is its corresponding ground truth. d_{io} and N denote the distance between the eyes and the number of key points of the face, respectively. N is set to 68 in this article. The smaller value means that the face contour is closer to the ground truth.

Tables 2, 3 and 4 show the quantitative results of our method and the state-of-the-art deblurring methods on the CelebA, UTKFace and CelebA-HQ datasets, respectively. We can see that our method obtains the lowest LPIPS on the CelebA and CelebA-HQ datasets, which indicates that our output is perceptually closest to the ground truth. Our method achieves the lowest FID and NIQE, which indicates that the output of our method has a small distance from the

Table 2 Quantitative comparison of the state-of-the-art methods on **CelebA** dataset

Methods	LPIPS (%)↓	FID↓	NIQE↓	FS. (%)↑	MNE (%)↓	PSNR↑	SSIM↑
Input	30.19	99.17	9.12	71.3	5.39	21.78	0.77
DeblurGAN [31]	7.77	40.23	17.29	87.3	2.96	23.48	0.75
DeblurGAN-v2 [32]	9.42	74.82	6.55	82.7	3.58	21.10	0.75
DMPHN [8]	10.29	22.30	7.54	93.0	2.47	28.33	0.91
SIUN [59]	11.73	22.16	7.36	89.2	3.41	22.16	0.84
UMSN [58]	15.04	34.85	7.94	89.5	2.50	27.38	0.89
MPRNet [61]	8.76	13.38	6.48	92.7	2.25	29.27	0.92
MIMO-UNet [7]	3.31	8.54	5.69	96.8	1.91	31.61	0.95
MPFG-GAN (Ours)	**1.57**	**5.98**	**5.44**	**98.3**	**1.77**	**33.13**	**0.96**
GT	0	0	4.97	100	0	∞	1

Bold and underline indicate the best and the second best performance, respectively. FS. represents the face similarity. "↑" denotes higher is better, and "↓" denotes lower is better

Table 3 Quantitative comparison of the state-of-the-art methods on **UTKFace dataset**

Methods	LPIPS (%)↓	FID↓	NIQE↓	FS. (%)↑	MNE (%)↓	PSNR↑	SSIM↑
Input	44.16	217.70	29.80	54.6	5.19	21.95	0.795
DeblurGAN [31]	11.28	58.72	106.22	84.2	2.66	23.83	0.763
DeblurGAN-v2 [32]	21.17	176.53	16.74	64.2	4.49	21.27	0.764
DMPHN [8]	15.67	53.88	5.55	76.4	2.82	28.12	0.904
SIUN [59]	20.55	158.33	6.18	76.0	3.81	22.93	0.855
UMSN [58]	20.38	73.96	5.92	73.8	3.20	26.83	0.884
MPRNet [61]	20.67	66.69	6.00	70.8	3.65	26.32	0.881
MIMO-UNet [7]	10.31	47.03	5.40	85.9	2.24	29.97	0.927
MPFG-GAN (Ours)	**8.33**	**34.55**	**4.86**	**90.2**	**1.87**	**31.47**	**0.939**
GT	0	0	2.91	100	0	∞	1

Bold and underline indicate the best and the second best performance, respectively. FS. represents the face similarity. "↑" denotes higher is better, and "↓" denotes lower is better

Table 4 Quantitative comparison of the state-of-the-art methods on **CelebA-HQ** dataset

Methods	LPIPS (%)↓	FID↓	NIQE↓	FS. (%)↑	MNE (%)↓	PSNR↑	SSIM↑
Input	44.10	95.88	10.30	55.4	4.88	21.77	0.623
DeblurGAN [31]	6.87	15.22	8.97	93.2	1.78	26.29	0.817
DeblurGAN-v2 [32]	8.26	23.81	5.43	93.1	1.82	25.42	0.797
DMPHN [8]	7.89	16.49	4.78	95.8	2.34	25.94	0.811
SIUN [59]	5.98	19.11	6.27	98.5	**1.43**	29.88	0.901
UMSN [58]	9.90	21.67	5.56	94.4	1.64	27.29	0.842
MPRNet [61]	10.88	31.27	7.26	94.6	1.75	27.39	0.845
MIMO-UNet [7]	5.28	13.03	4.47	98.4	1.46	30.07	0.898
MPFG-GAN (Ours)	**4.17**	**9.93**	**3.16**	**98.7**	1.43	**30.15**	**0.902**
GT	0	0	2.55	100	0	∞	1

Bold and underline indicate the best and the second best performance, respectively. FS represents the face similarity. "↑" denotes higher is better, and "↓" denotes lower is better

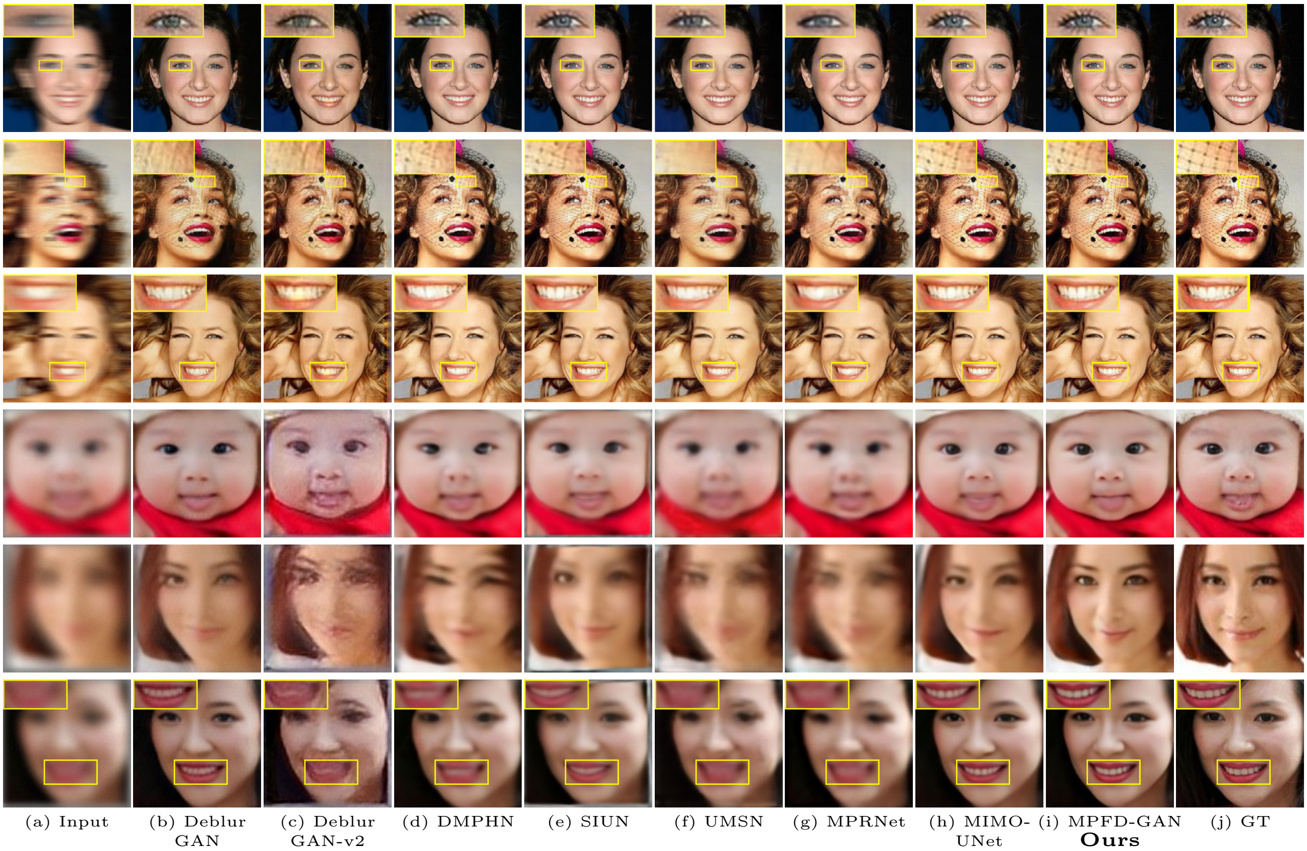

Fig. 7 Qualitative comparison with state-of-the-art face deblurring methods: **b** DeblurGAN [31], **c** DeblurGAN-v2 [32], **d** DMPHN [8], **e** SIUN [59], **f** UMSN [57], **g** MPRNet [61] and **h** MIMO-UNet [7] on the **CelebA-HQ (first three lines)** and **UTKFace (last three lines)** dataset for face blind deblurring. Our MPFD-GAN produces faithful texture details in eyes, veils and teeth

Table 5 Ablation on the network architecture

Configuration	LPIPS (%)↓	FID↓	NIQE↓	FS. (%)↑	MNE (%)↓	PSNR↑	SSIM↑
Our MPFD-GAN	**1.57**	**5.98**	**5.44**	**98.3**	**1.77**	**33.13**	**0.96**
(a) w/o SE-ResNet Block	3.71	12.54	5.88	97.2	1.89	30.92	0.94
(b) w/o FRM	3.70	13.05	5.93	97.1	1.92	30.58	0.94
(c) w/o TRM	6.77	16.54	6.35	95.4	2.13	28.98	0.92
(d) -Facial Region Discriminator	3.62	12.93	5.95	97.2	1.90	30.81	0.94

FS. represents the face similarity. "↑" denotes higher is better, and "↓" denotes lower is better

Table 6 Ablation on the loss functions

L_{pyramid}	L_{edge}	L_{id}	LPIPS (%)↓	FID↓	NIQE↓	FS. (%)↑	MNE (%)↓	PSNR↑	SSIM↑
✓	✓	✓	**1.57**	**5.98**	**5.44**	**98.3**	**1.77**	**33.13**	**0.96**
✓	✓	×	3.52	12.12	5.96	97.0	1.93	30.67	0.94
✓	×	✓	3.27	11.63	5.9	97.3	1.92	30.8	0.94
×	✓	✓	3.45	12.39	5.83	97.2	1.91	30.89	0.94

L_{pyramid}, L_{edge}, L_{id} denote pyramid restoration loss, edge loss, and identity preservation loss, respectively. FS represents the face similarity. "↑" denotes higher is better, and "↓" denotes lower is better

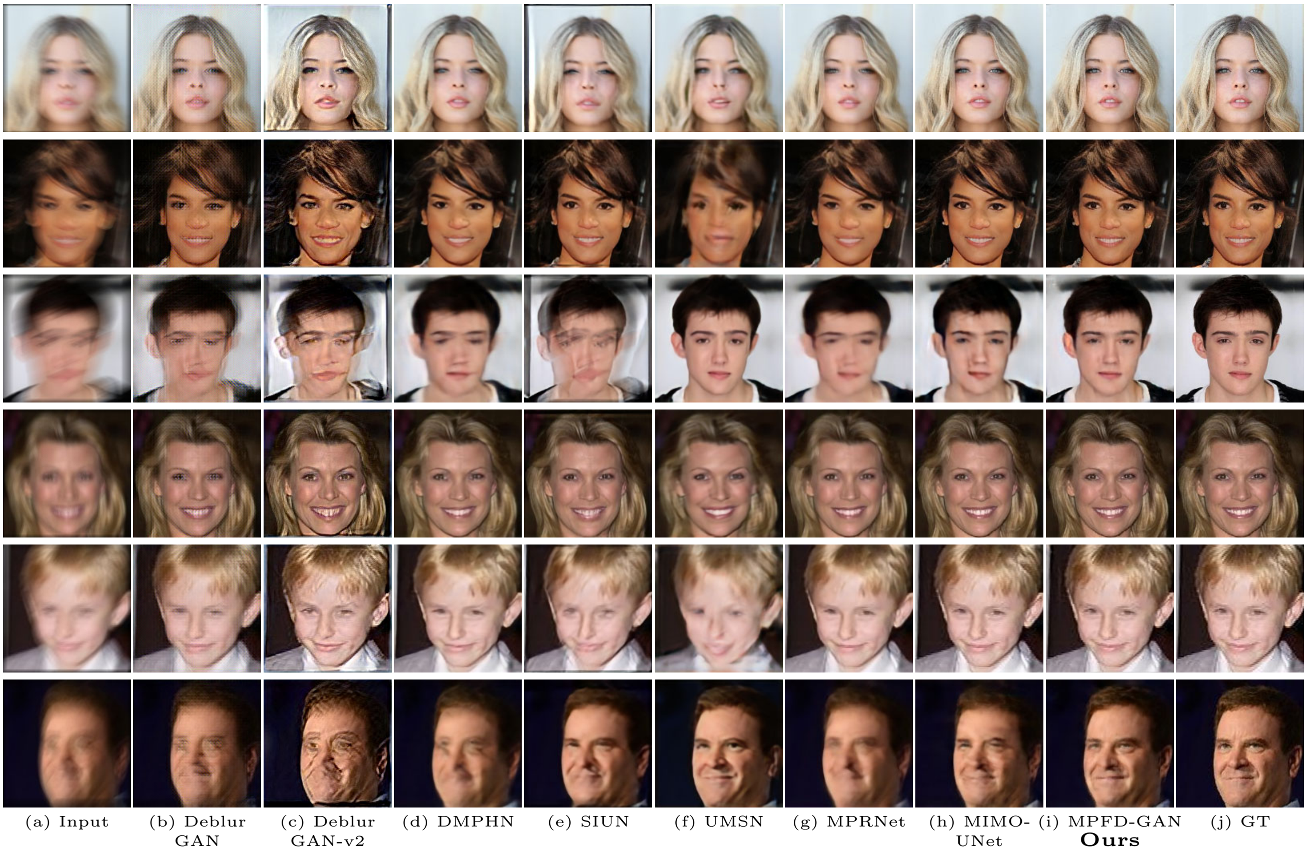

Fig. 8 Qualitative comparison of our method (i) and the current state-of-the-art network: **b** DeblurGAN [31], **c** DeblurGAN-v2 [32], **d** DMPHN [8], **e** SIUN [59], **f** UMSN [57], **g** MPRNet [61] and **h** MIMO-UNet [7] on the **CelebA** dataset. Our MPFD-GAN effectively removes blur and generates faces that are natural, authentic and visually closer to the ground truth

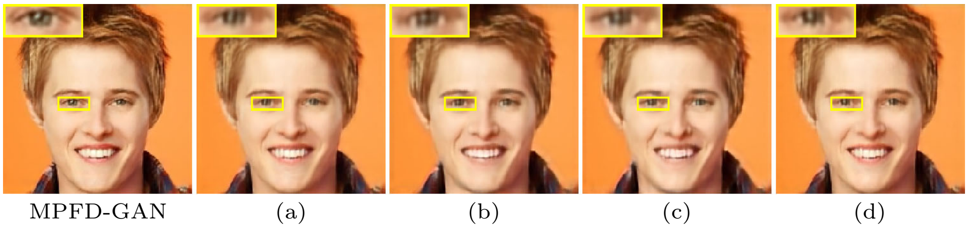

Fig. 9 Ablation studies on **a** SE-ResNet Block, **b** FRM, **c** TRM and **d** facial discriminator. The first column shows the recovery results of the complete MPFD-GAN, and the remaining columns show the recovery results of the corresponding configuration in Table 5

natural image distribution and the realistic face distribution, respectively. Our method also obtains the highest PSNR and SSIM, which indicates that our output results are closest to ground truth at the structure and pixel level. Furthermore, the face in our output results has the highest similarity to the ground truth (shown in Tables 2, 3 and 4), demonstrating that our method recovers better identity information. Our MPFD-GAN also obtains a lower MNE (slightly higher than the SIUN method on CelebA-HQ), which indicates that our approach recovers the facial contour closer to the ground truth.

Qualitative comparison

Figures 7 and 8 show the qualitative results of the various methods on CelebA-HQ, UTKFace and CelebA datasets. We can observe from Fig. 7 that our MPFD-GAN recovers realistic details such as eyes (eyelashes, etc.), facial decorations and teeth. That is due to TRM fusing detailed high-frequency texture information and filtering the feature information in space and dimension, letting the useful features pass to the next scale. As shown in Fig. 8, our MPFD-GAN recovers the complete structure and more realistic texture details to achieve the best visual quality on the CelebA dataset. Among them, the recovery results obtained by DeblurGAN have significant checkerboard artifacts (see Fig. 8b). The deblurred output obtained by DeblurGAN-v2 and SIUM also has significant ghosting and no good visual performance (see Fig. 8c, e). The face images recovered by DMPHN and UMSN suffer from facial disharmony (see Fig. 8d, f). In addition, we found that MPRNet and MIMO-UNet are not effective at deblurring face images that suffer from high blurring influence (see Fig. 8, rows 3 and 5). They also fail to remove blur in local details (e.g., eyes, teeth, etc.) and structures in enlarged image regions(see Fig. 7g, h).

Ablation studies

To better understand the roles of different components of MPFD-GAN and the training strategy, we conduct an ablation study by introducing some variants of the proposed method and comparing their blind face deblurring performance in this subsection. All ablation experiments perform on the CelebA dataset.

Ablation on network architecture

As shown in Table 5 (configuration a) and Fig. 9a, we can observe that the recovered face images lose facial details and get poor evaluation metrics when replacing the SE-ResNet block in MPFD-GAN with the ResNet block. The results indicate that the SE-ResNet block is essential for MPFD-GAN to extract effective features from blurry images. When FRM and TRM are removed separately (configuration b and c), we can observe that (1) the face images in the recovered results lose a lot of details in the eye region, causing the eyes to remain blurry (see Fig. 9b, c); (2) the performance of both perceptual metrics and pixel-wise metrics in Table 5 (configuration b and c) is degraded. These comparison results demonstrate that FRM and TRM make very significant contributions in the deblurring process of MPFD-GAN. Finally, we compare the experimental results with and without the facial region discriminator (see Table 5; Fig. 9d). The results show that the local discriminator can prompt the model to better restore the distribution of facial regions and effectively recover realistic details.

Loss function for ablation studies

At the same time, we further investigate the contribution of different loss terms by adjusting the weight of each loss in Eq. (22), and the results are shown in Table 6. When we remove the pyramid restoration loss, the performance of each metric obtained through the experiment decreases. This experimental result indicates that pyramid restoration loss enhances the recovery ability of MPFD-GAN to form blurry images. This intermediate supervision, which is helpful to recover sharp face images (multi-scale) in each scale of upsampling, also enhances the overall deblurring performance of the model. When MPFD-GAN removes the supervision on realistic texture details (edge loss) and face features (identity preservation loss), it will result in the final recovered images not giving the best performance (see Table 6).

All the above ablation experiments again demonstrate that the design scheme of MPFD-GAN and our proposed individual modules are very effective for the blind face deblurring task.

Conclusion

In this paper, we propose a multi-scale progressive deblurring network named MPFD-GAN to deblur the face image unknown degradation without requiring extra inputs (i.e., face priors and facial component dictionaries). This approach mainly includes two core modules: FRM and TRM. The former can explore multi-scale receptive field information to help MPFD-GAN recover the complete image structure. The latter prompts MPFD-GAN progressively reconstruct facial texture details by fusing high-frequency texture information. Comparative experiments on the CelebA, UTKFace, and CelebA-HQ datasets demonstrate the superiority and robustness of our MPFD-GAN. Ablation experiments further verify the effect of core modules (namely FRM and TRM) of MPFD-GAN for the above tasks. In conclusion, MPFD-GAN provides a robust and easy-to-use solution for face deblurring task.

Future work

In future work, we consider employing MPFD-GAN to solve the challenging task of blind face restoration. The task demands recovering high-quality faces from the low-quality counterparts suffering more complex unknown degradation [5], such as low-resolution, compression artifacts, color fading, and lossy compression. To this end, we performed some simple experiments to test the potential of our MPFD-GAN for this task. Following the setting of the experimental data by Li et al. [34], we retrain the proposed model and UMSN [57] on the FFHQ dataset [28] and test both methods on three real-world datasets: CelebChild-Test [49], LFW-Test [28] and WebPhoto-Test [49]. The experimental results are shown in Fig. 10 and we can see that MPFD-GAN achieves better visual performance than UMSN [57] on real-world datasets. The above experiments are just a simple attempt of MPFD-GAN for blind face restoration. We will do more experiments in the future to solve the task well.

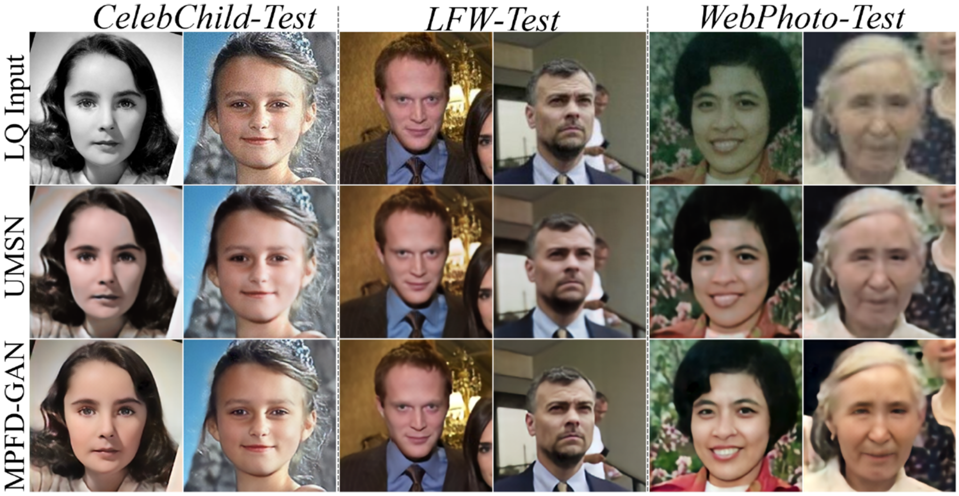

Fig. 10 Visual comparison between UMSN [57] and MPFD-GAN (Ours) on real-world low-quality images

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant No. 42130608) and the Sichuan Science and Technology program (Grant No. 2021YFQ0053, 22ZDYF3709, 2020JDTD0020, 23NSFSC0999, 2022YFG0026, 2021YFG0018).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Boracchi G, Foi A (2012) Modeling the performance of image restoration from motion blur. IEEE Trans Image Process 21(8):3502–3517
2. Bulat A, Tzimiropoulos G (2018) Super-fan: integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with gans. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 109–117
3. Charbonnier P, Blanc-Feraud L, Aubert G, Barlaud M (1994) Two deterministic half-quadratic regularization algorithms for computed imaging. In: Proceedings of 1st international conference on image processing, vol 2, pp 168–172. IEEE
4. Chen C, Gong D, Wang H, Li Z, Wong K (2021) Learning spatial attention for face super-resolution. IEEE Trans Image Process 30:1219–1231
5. Chen C, Li X, Yang L, Lin X, Zhang L, Wong K-YK (2021) Progressive semantic-aware style transformation for blind face restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 11896–11905
6. Chen Y, Tai Y, Liu X, Shen C, Yang J (2018) Fsrnet: End-to-end learning face super-resolution with facial priors. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2492–2501
7. Cho S-J, Ji S-W, Hong J-P, Jung S-W, Ko S-J (2021) Rethinking coarse-to-fine approach in single image deblurring. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 4641–4650
8. Das SD, Dutta S (2020) Fast deep multi-patch hierarchical network for nonhomogeneous image dehazing. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, pp 482–483
9. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L (2009) Imagenet: a large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition, pp 248–255. IEEE
10. Deng J, Guo J, Xue N, Zafeiriou S (2019) Arcface: additive angular margin loss for deep face recognition. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 4690–4699
11. Fan Y, Hong C, Wang X, Zeng Z, Guo Z (2021) Multi-input-output fusion attention module for deblurring networks. In: 2021 IEEE international conference on big data (Big Data), pp 3176–3182
12. Feng H, Guo J, Ge SS (2020) Sharpgan: Receptive field block net for dynamic scene deblurring. arXiv preprint arXiv:2012.15432
13. Gupta A, Joshi N, Zitnick CL, Cohen MF, Curless B (2010) Single image deblurring using motion density functions. In: European conference on computer vision
14. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
15. Heusel M, Ramsauer H, Unterthiner T, Nessler B, Hochreiter S (2017) Gans trained by a two time-scale update rule converge to a local Nash equilibrium. In: Advances in neural information processing systems, vol 30
16. Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7132–7141
17. Hu K, Liu Y, Liu R, Lu W, Yu G, Fu B (2022) Enhancing quality of pose-varied face restoration with local weak feature sensing and gan prior. arXiv preprint arXiv:2205.14377

18. Huang R, Zhang S, Li T, He R (2017) Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis. In: Proceedings of the IEEE international conference on computer vision, pp 2439–2448
19. Hui J, Liu C (2008) Motion blur identification from image gradients. In: 2008 IEEE computer society conference on computer vision and pattern recognition (CVPR 2008), 24–26 June 2008, Anchorage, Alaska, USA
20. Isola P, Zhu J-Y, Zhou T, Efros AA (2017) Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1125–1134
21. Jiang J, Wang C, Liu X, Ma J (2021) Deep learning-based face super-resolution: a survey. ACM Comput Surv (CSUR) 55(1):1–36
22. Jiang K, Wang Z, Yi P, Chen C, Huang B, Luo Y, Ma J, Jiang J (2020) Multi-scale progressive fusion network for single image deraining
23. Jiao J, Cao Y, Song Y, Lau R (2018) Look deeper into depth: monocular depth estimation with semantic booster and attention-driven loss. In: Proceedings of the European conference on computer vision (ECCV), pp 53–69
24. John and Odentrantz (2000) Markov chains: Gibbs fields, Monte Carlo simulation, and queues. Technometrics 42(4):438–439
25. Johnson J, Alahi A, Fei-Fei L (2016) Perceptual losses for real-time style transfer and super-resolution. In: European conference on computer vision, pp 694–711. Springer
26. Jolicoeur-Martineau A (2018) The relativistic discriminator: a key element missing from standard gan. arXiv preprint arXiv:1807.00734
27. Karras T, Aila T, Laine S, Lehtinen J (2017) Progressive growing of gans for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196
28. Karras T, Laine S, Aila T (2019) A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 4401–4410
29. Karras T, Laine S, Aittala M, Hellsten J, Lehtinen J, Aila T (2020) Analyzing and improving the image quality of stylegan. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 8110–8119
30. Kim D, Kim M, Kwon G, Kim (D-S) Progressive face super-resolution via attention to facial landmark. arXiv preprint arXiv:1908.08239
31. Kupyn O, Budzan V, Mykhailych M, Mishkin D, Matas J (2018) Deblurgan: Blind motion deblurring using conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 8183–8192
32. Kupyn O, Martyniuk T, Wu J, Wang Z (2019) Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 8878–8887
33. Li L, Bao J, Yang H, Chen D, Wen F (2019) Faceshifter: towards high fidelity and occlusion aware face swapping. arXiv preprint arXiv:1912.13457
34. Li X, Chen C, Zhou S, Lin X, Zuo W, Zhang L (2020) Blind face restoration via deep multi-scale component dictionaries. In: European conference on computer vision, pp 399–415. Springer
35. Lin S, Zhang J, Pan J, Liu Y, Ren J (2020) Learning to deblur face images via sketch synthesis. Proc AAAI Conf Artif Intell 34(7):11523–11530
36. Liu H, Jiang B, Song Y, Huang W, Yang C (2020) Rethinking image inpainting via a mutual encoder–decoder with feature equalizations. In: European conference on computer vision, pp 725–741. Springer
37. Liu Z, Luo P, Wang X, Tang X (2015) Deep learning face attributes in the wild. In: Proceedings of the IEEE international conference on computer vision, pp 3730–3738
38. Loshchilov I, Hutter F (2017) Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101
39. Mirza M, Osindero S (2014). Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784
40. Mittal A, Soundararajan R, Bovik AC (2012) Making a "completely blind" image quality analyzer. IEEE Signal Process Lett 20(3):209–212
41. Noroozi M, Chandramouli P, Favaro P (2017) Motion deblurring in the wild. In: German conference on pattern recognition, pp 65–77. Springer
42. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, et al (2019) Pytorch: an imperative style, high-performance deep learning library. In: Advances in neural information processing systems, vol 32
43. Patel VM, Easley GR, Healy DM (2009) Shearlet-based deconvolution. IEEE Trans Image Process 18(12):2673–2685
44. Schuler CJ, Christopher Burger H, Harmeling S, Scholkopf B (2013) A machine learning approach for non-blind image deconvolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1067–1074
45. Shen Z, Lai W-S, Xu T, Kautz J, Yang M-H (2018) Deep semantic face deblurring. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 8260–8269
46. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556
47. Sun L, Cho S, Wang J, Hays J (2013) Edge-based blur kernel estimation using patch priors. In: IEEE international conference on computational photography (ICCP), pp 1–8. IEEE
48. Wang N, Gao X, Tao D, Yang H, Li X (2017) Facial feature point detection: a comprehensive survey. Neurocomputing 275(1):50–65
49. Wang X, Li Y, Zhang H, 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, pp 9168–9178
50. Wang X, Yu K, Wu S, Gu J, Liu Y, Dong C, Qiao Y, Change Loy C (2018) Esrgan: enhanced super-resolution generative adversarial networks. In: Proceedings of the European conference on computer vision (ECCV) workshops
51. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612
52. Wang Z, Zhang J, Chen R, Wang W, Luo P (2022) Restoreformer: High-quality blind face restoration from undegraded key-value pairs. arXiv preprint arXiv:2201.06374
53. Whyte O, Sivic J, Zisserman A, Ponce J (2010) Non-uniform deblurring for shaken images. In: Computer vision and pattern recognition
54. Xu L, Jia J (2010) Two-phase kernel estimation for robust motion deblurring. In: European conference on computer vision, pp 157–170. Springer
55. Xu L, Zheng S, Jia J (2013) Unnatural l0 sparse representation for natural image deblurring. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1107–1114
56. Yang L, Wang S, Ma S, Gao W, Liu C, Wang P, Ren P (2020) Hifacegan: Face renovation via collaborative suppression and replenishment. In: Proceedings of the 28th ACM international conference on multimedia, pp 1551–1560
57. Yasarla R, Perazzi F, Patel VM (2020) Deblurring face images using uncertainty guided multi-stream semantic networks. IEEE Trans Image Process 99:1–1

58. Yasarla R, Perazzi F, Patel VM (2020) Deblurring face images using uncertainty guided multi-stream semantic networks. IEEE Trans Image Process 29:6251–6263
59. Ye M, Lyu D, Chen G (2020) Scale-iterative upscaling network for image deblurring. IEEE Access 8:18316–18325
60. Yu X, Fernando B, Ghanem B, Porikli F, Hartley R (2018) Face super-resolution guided by facial component heatmaps. In: Proceedings of the European conference on computer vision (ECCV), pp 217–233 (2018)
61. Zamir SW, Arora A, Khan S, Hayat M, Khan FS, Yang M-H, Shao L (2021) Multi-stage progressive image restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 14821–14831
62. Zhang K, Zuo W, Chen Y, Meng D, Zhang L (2017) Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising. IEEE Trans Image Process 26(7):3142–3155
63. Zhang R, Isola P, Efros AA, Shechtman E, Wang O (2018) The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 586–595
64. Zhang X, Wang X, Shi C, Yan Z, Li X, Kong B, Lyu S, Zhu B, Lv J, Yin Y, et al (2021) De-gan: Domain embedded gan for high quality face image inpainting. Pattern Recognition, p 108415
65. Zhang Z, Song Y, Qi H (2017) Age progression/regression by conditional adversarial autoencoder. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR)
66. Zhou X-Y, Zheng J-Q, Yang G-Z (2018) Atrous convolutional neural network (acnn) for biomedical semantic segmentation with dimensionally lossless feature maps. arXiv preprint arXiv:1901.09203, p 68
67. Zhu S, Liu S, Loy CC, Tang X (2016) Deep cascaded bi-network for face hallucination. In: European conference on computer vision, pp 614–630. Springer
68. Zou W, Jiang M, Zhang Y, Chen L, Lu Z, Wu Y (2021) Sdwnet: A straight dilated network with wavelet transformation for image deblurring. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 1895–1904

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.