GFNet: A Gradient Information Compensation-Based Face Super-Resolution Network

SHENGYANG LUO AND JINBO LU
School of Electronics and Information Engineering, Southwest Petroleum University, Chengdu 637001, China
Corresponding author: Jinbo Lu (lujb@swpu.edu.cn)

This work was supported in part by the National Natural Science Foundation for Young Scientists of China under Grant 61603319 and Grant 61601385.

ABSTRACT Face super-resolution (FSR) is defined as the generation of high-resolution face images from low-resolution face images. Existing FSR approaches usually improve the performance by combining deep learning with additional tasks such as face parsing and landmark prediction. However, the additional data requires manual labeling, and facial landmark heatmaps and parsing maps cannot represent the intrinsic geometric structure of facial components. In this paper, we introduce a FSR network based on gradient information compensation named GFNet, which consists of feature residual blocks (FRBs) and gradient extraction blocks (GEBs). Specifically, the GEB constructs pixel-level gradient maps directly from the feature maps without requiring data labels and extracts gradient features to compensate for the missing high-frequency components in the face features; the FRB extracts the face features in the network. Furthermore, we introduced a feature fusion mechanism between the GEB and the FRB, which fuses the face features with the gradient features. We evaluate the performance of proposed network on the two public datasets: CelebA-HQ dataset and Helen dataset. Experimental results show that the proposed method is able to reconstruct fine face images, which outperforms the other state-of-the-art methods such as SRResnet, FSRNet, and MSFSR.

INDEX TERMS Super-resolution, face hallucination, gradient information, feature fusion.

I. INTRODUCTION

Face super-resolution (FSR), also known as face hallucination, generates high-resolution (HR) face images from low-resolution (LR) face images. With the development of video technology, face information detection has become a recent research hotspot [1]–[3]. However, due to the empirical factors such as devices and distances, the face images are usually low-resolution, which may hinder face information detection.

Unlike the single image super-resolution (SISR) tasks [4]–[8], the difficulty of FSR tasks lies in the recovery of facial components, i.e., eye and nose. These facial components have large pixel variations in face images, and therefore, it is difficult for deep neural networks trained using mean square error loss to recover these facial components. In addition, FSR tasks generally require reconstructing face images at high upscale factors (e.g., 8×), while SISR tasks generally focus on reconstructing images at low upscale factors (e.g., 2×, 4×). Therefore, the FSR task is more complex since there are more mapping functions.

Some methods [9]–[13] assist FSR networks in face image recovery by adding additional tasks such as face parsing, landmark prediction, etc. In [9], the authors used a priori estimation networks to predict landmark heatmaps and parsing maps. Although the network extracts more facial structure information by introducing these additional tasks, there exist the following drawbacks: 1) it requires manual labeling of data for the additional tasks; 2) estimating face priors from low-resolution inputs is itself a difficult task; and 3) facial landmark heatmaps and parsing maps cannot represent the intrinsic geometric structure of facial components, such as the nose bridge.

In this paper, we propose a FSR network for faces based on gradient information compensation. First, to reconstruct
fine facial details of faces, we propose a gradient extraction block (GEB), which constructs pixel-level gradient maps directly from face features; second, we employ a feature residual block (FRB) to extract face features in the network. By introducing a channel attention mechanism, the FRB can focus on important features; finally, we introduce a feature fusion mechanism between the GEB and the FRB, which allows the network to utilize features at different levels. The primary contributions of this paper are summarized as follows:

1) We propose the GFNet, a face super-resolution framework capable of processing very low-resolution face images (e.g., 16 × 16). Without requiring additional training task, GFNet is able to generate fine-grained facial components of faces.

2) We develop a GEB and a FRB to effectively extract the features. The GEB can directly extract gradient features from the original LR image to compensate for the missing high-frequency components of face features, which solves the problem that most FSR tasks require additional face priors; the feature fusion mechanism is applied between the GEB and the FRB, which makes the FRB attaches channel attention to fusing face feature maps and gradient feature maps, and makes the GEB extract gradient information guided by the face features.

3) Experimental results on CelebA-HQ dataset and Helen dataset show that the proposed method outperforms the state-of-the-art algorithms both qualitatively and quantitatively.

II. RELATED WORK

A. FACE SUPER-RESOLUTION

FSR was first proposed by Baker and Kanade [14]. Following this work, various FSR methods have been proposed. Traditional FSR methods can be roughly classified into two categories: global model-based methods [15]–[20] and part-based methods [21]–[24]. Global model-based methods reconstruct super-resolution images by learning the overall appearance mapping, such as principle component analysis. This method requires the LR input to be precisely aligned with the HR image. Part-based methods extract facial components, which requires the detection of facial landmarks in the LR image. The performance of both methods degrades significantly in FSR tasks with a high upscale factor.

Recently, convolutional neural networks have been successfully used to face super-resolution. Zhu et al. [25] proposed a Cascade Bi-Network (CBN) to recover facial details by estimating the dense correspondence field. Due to the difficulty of accurate dense correspondence field estimation at low resolution, it performs poorly on human faces in different poses. Huang et al. [26] used wavelet coefficients predicted from LR images to reconstruct HR images. Zhang et al. [27] used enhanced facial boundaries and coarse-to-fine supervision to recover fine faces. Compared to landmark points, the enhanced facial boundaries can represent richer contours of facial components, however, the representation of the intrinsic geometric structure is yet to be explored.

Some works utilizes additional face priors to train the network. Chen et al. [9] fused face priors with face features and train the network in an end-to-end manner. Ma et al. [28] introduced a loop iteration for landmark estimation with face recovery in a recurrent iterative network. Dogan et al. [13] proposed a GWAI net that takes the LR image and a high-resolution guide face as inputs. They convert the posture of the guide face to be the same as the LR image. Although introducing additional face priors helps to improve the performance, there are two drawbacks: 1) additional manually labeled data is required, and 2) estimating face priors from LR images is itself a difficult task.

Unlike the above methods, the proposed GFNet uses gradient maps with pixel-level accuracy to compensate for the missing high-frequency components in face features, which means that we can obtain more semantic information even in very low-resolution images. The gradient extraction block constructs the gradient image directly from the face features, so no additional supervised labeling is required.

B. GRADIENT CORRELATION METHOD

In previous SISR studies, researchers have used gradient information to reshape the edges of images [29], [30]. They used a cascade framework consisting of two identical networks to process gradient images and LR images separately; however, the structure of this single cascade framework cannot cope with the complex FSR tasks. Yang et al. [31] divided the face image into three parts: facial component, contour, and smoothed region. They matched the gradient maps of the three parts with the reconstructed HR image to maintain the image structure. The gradient image of the facial component is determined by a set of landmark points and two label sets. They then further improved their algorithm to process the compressed face images [32]. Pei et al. [33] proposed the iterative gradient constrained weighted sparse representation method, where they exploit the gradient information of the images during the patch representation. They combined both the $l_1$ reweighted constraint and the gradient information of the images, and fused them into the sparse representation to improve the performance. Although the above method improves the performance, there are still several shortcomings: 1) the deep learning-based FSR task usually requires a large amount of data. The above method divides the face images into different regions, meaning that it requires more supervised labels. 2) When processing very low resolution face images, the errors caused by landmark detection can lead to artifacts in the reconstructed images. In contrast, the gradient extraction block can extract gradient information without landmark detection. 3) To generating a gradient map for the whole face, the network should get a more comprehensive view of the facial structure.
III. PROPOSED METHOD

A. OVERVIEW

Fig. 1 shows the structure of the proposed network, which consists of three main parts: downscale, feature extraction, and upscale parts. Each part consists of several FRBs and GEBs. We introduce a feature fusion mechanism between the two blocks. To enable the network to focus on features at different scales, we designed the downscale part as well as the upscale part. In the Downscale part, we change the structure of the latter FEBs and GEBs to achieve downsampling (see Sections III-B and III-C for details). In the upscale part, we change the structure of the former FEBs and GEBs to achieve upsampling. Let $I_{LR}$, $I_{SR}$, and $I_{HR}$ denote the low-resolution image, the super-resolved image and the ground truth HR image, respectively.

First, we upsample $I_{LR}$ to the same size as $I_{HR}$ as follows.

$$I_{UP} = F_{bic}(I_{LR}),$$  
(1)

where $F_{bic}(\cdot)$ denotes the bicubic interpolation function, and $I_{UP}$ denotes the upsampled image. Then, $I_{UP}$ is fed into the GFNet.

$$I_{SR} = F_{GFNet}(I_{UP}),$$  
(2)

where $F_{GFNet}$ denotes the function of our GFNet. Given a set of training set $\{I_{LR}^{(i)}, I_{HR}^{(i)}\}_{i=1}^{N}$ containing $N$ samples, we optimize GFNet by minimizing the pixel-level $L_2$ loss given by

$$L_{pix}(\Theta) = \frac{1}{N} \sum_{i=1}^{N} \left\| I_{SR}^{(i)} - I_{HR}^{(i)} \right\|_2^2,$$  
(3)

where $\Theta$ denotes the parameter set of GFNet.

B. FEATURE RESIDUAL BLOCK

The GFNet needs to process both face features and gradient features simultaneously. We add the residual channel attention block (RCAB) [34] to the FRB, which makes the network focus its channel attention on fusing face feature maps and gradient feature maps. As shown in Fig. 1, the FRB consists of two $3 \times 3$ convolutional layers and a RCAB. Each convolutional layer is followed by a batch normalization (BN) and a rectified linear unit (ReLU) activity function. Fig. 2(a) is the structure of RCAB, channel attention is integrated into residual block. The residual component is mainly obtained by two $3 \times 3$ convolution layers. In channel attention, the weights on feature map channels are calculated by two $1 \times 1$ convolution layers. They adjust weights automatically. In this process, the feature map channels are rescaled with ratio $r$ by two $1 \times 1$ convolution layers. In this paper, we set $r$ to 16. Since residual blocks have achieved great success in SR tasks [35] and FSR tasks [6], [36], so we add residual connections to the feature residual module and determine the number of channels by a $1 \times 1$ convolutional layer. We assume that the input of the $i$-th feature extraction module is $x_{i-1}$ of 128 channels, then the process of feeding it into the $i$-th feature extraction module can be expressed as follows:

$$f_i = Conv_{64}^{128} \left( R \left( Conv_{64}^{128} (x_{i-1}) \right) \right),$$  
(4)

$$x_i = Conv_{64}^{128} (x_{i-1})_1 + f_i,$$  
(5)

where $f_i$ denotes the output of the last convolutional layer in the feature extraction module, $Conv_{scal}^{scal}(\cdot)$ denotes an $n \times n$ convolutional layer with in input channels and out output channels, followed by BN and ReLU. $R(\cdot)$ denotes the RCAB function.

As shown in Fig. 3(a), in the downscale part, we use a convolutional layer with a step size of 2 to achieve downsampling. In the upscale section, we use a nearest-neighbor upsampling layer with a convolutional layer to slightly modify the feature extraction module to achieve upsampling. Nearest-neighbor upsampling layer with a convolutional layer helps to avoid checkerboard artifacts [37]. Thus equation (5) becomes:

$$x_i = F_{scal} (x_{i-1}) + f_i,$$  
(6)

where $F_{scal}(\cdot)$ denotes the scale transformation function.

C. GRADIENT EXTRACTION BLOCK

Gradient information has been widely used in various image tasks to improve system performance, such as image translation [38], SISR [29], [30], and image restoration [39]. Gradient information can compensate the missing parts in high-frequency components such as edges and structures. So it can improve the problem of easy generation of smooth images using $L_2$ loss in SR tasks. In this paper, we design the GEB to extract the gradient information from the features. As shown in Fig. 1, the GEB consists of three convolutional layers and one gradient layer. Since we need to extract the gradient information from the features of three channels, we set the kernel size of the second convolutional layer to $1 \times 1$. In the gradient layer, we obtain the gradient map by calculating the difference between adjacent pixels as follows:

$$I_{a}(A) = I(a+1, b) - I(a-1, b),$$

$$I_{b}(A) = I(a, b+1) - I(a, b-1),$$

$$\nabla I(A) = (I_{a}(A), I_{b}(A))$$

$$T(I) = \|\nabla I\|_2,$$  
(7)

where $T(\cdot)$ denotes the gradient map extraction, and the element of gradient map is the gradient length of the pixel of coordinate $A = (a, b)$, where $I_{a}(A)$ is the pixel gradient of a pixel A along direction $a$ on the image I. It is the same as the pixel value difference between the previous pixel and the latter pixel in direction $a$. It is also the same case in direction $b$. $\nabla I(A)$ is a two-dimensional vector representing the pixel gradient of A. By calculating the two-dimensional norms of these two-dimensional vectors of all pixels on the image I, the gradient map $T(I)$ can be obtained.

We assume that the input of the $i$-th gradient extraction module is $y_{i-1}$ with 128 channels, then the process of feeding
it to the $i$-th feature extraction module can be expressed as
\[ y_i = \text{Conv}^3_{64} \left( T \left( \text{Conv}^6_{64} \left( \text{Conv}^{128}_{64} (y_{i-1})_3 \right)_3 \right)_3 \right)_3 \]  
(8)

where $y_i$ denotes the output of the $i$-th gradient extraction module. As shown in Fig. 3(b), the scale scaling function of the gradient extraction module is achieved by using the same strategy as described in Section III-B.

D. FEATURE FUSION MECHANISM

We found that the FRB was rich in structural information, which is crucial for the GEB to extract gradient features. Since the gradient features carrying high frequency information in the GEB, this facilitates the feature residual module to obtain richer face features. Therefore, we introduce the feature fusion mechanism, which fuses the features between the two modules and input the fused features into the next FRB and GEB respectively, as shown in Fig. 1. This process can be expressed as
\[ z_i = x_i \odot y_i \]  
(9)

where $\odot$ denotes the concatenate of feature maps, and $z_i$ denotes the feature after fusing the output $x_i$ of the $i$-th feature residual module and the output $y_i$ of the $i$-th gradient extraction module. Equation. (4), (5), (6) and (8) can be rewritten as:
\[ f_i = \text{Conv}^6_{64} \left( R \left( \text{Conv}^{128}_{64} (z_{i-1})_3 \right)_3 \right)_3, \]
(10)
\[ x_i = \text{Conv}^{128}_{64} (z_{i-1})_3 + f_i, \]
(11)
\[ x_i = f_i + F_{scal} (z_{i-1}), \]
(12)
\[ y_i = \text{Conv}^3_{64} \left( T \left( \text{Conv}^6_{64} \left( \text{Conv}^{128}_{64} (z_{i-1})_3 \right)_3 \right)_3 \right)_3, \]
(13)

IV. EXPERIMENT

A. DATASET

We used two commonly used face datasets, CelebA-HQ [40] and Helen [41], to evaluate our model. CelebA-HQ is an updated version of CelebA [42], with a total of 30k images, each with a resolution of 1024 x 1024. We randomly selected 17k images as the training set and 13k images as the test set.

For the Helen dataset, we randomly selected 2000 images as the training set and 330 images as the test set.

B. IMPLEMENTATION DETAILS

We resized the images of the CelebA-HQ dataset to 128 x 128 pixels as the ground truth HR images. For the Helen dataset, we cropped the 128 x 128 pixel face images based on the annotations of the facial components of the images. We then downsampled the HR images to 16 x 16 pixels by bicubic interpolation as the LR input and enhanced the training images by randomly rotating them by 90°, 180°, 270°, and horizontal flip. We used PSNR and SSIM as performance metrics to evaluate the model.

We use the Adam optimizer to train the model, where $\beta_1 = 0.9$, $\beta_2 = 0.999$. The initial learning rate is $1 \times 10^{-4}$ and halved at 15k, 30k, 45k, 90k, 180k iterations. The batch size is set to 16. We implemented the code in Pytorch.

C. ABLATION STUDY

We conducted a series of ablation experiments on the CelebA-HQ dataset to validate the effectiveness of our proposed method.
To evaluate the effect of gradient information, considering that the gradient layer is an artificial design without any learnable parameters, and its computational complexity is negligible, we keep the parameters of GFNet constant and remove the gradient layer in the BEM to observe the performance of the network. The new model is defined as GFNet-WG. The visualization results are shown in Fig. 4. As can be seen from the local zoomed-in view, with the help of the gradient information, our network is able to generate a more three-dimensional face structure (e.g., nose) with more accurate contours than GFNet-WG. Table 1 shows the results of GFNet-WG in both PNSR and SSIM tests. By comparing the PSNR and SSIM scores in results of GFNet-WG with GFNet, it can be concluded that the proposed gradient information compensation strategy into the face features is effective.

2) GFNet-FM
We evaluate the effectiveness of the feature fusion mechanism by eliminating the concatenate operation of the feature map between the FRB and the GEB, as shown in Fig. 5. We define the model with the feature fusion mechanism at $M$ cancelled as GFNet-FM, where $M \in \{0, 8, 15\}$. When $M$ is set to 15, the feature exchange mechanism is located between the last feature residual module and the gradient extraction module. As can be seen from Table 2, the smaller $M$ is, the better the model performs in the test. This means that the feature exchange mechanism can improve the performance of the network.

D. COMPARISONS WITH STATE-OF-THE-ART METHODS
We compare our model with state-of-the-art methods on the CelebA-HQ and Helen datasets. In generic SR methods, we chose SRCNN [4], SRResNet [6], and VDSR [5]; in FSR methods, we chose PFSRNet [10], FSRNet [9], and MSFSR [27]. Since only the test code of FSRNet was released, we re-implement FSRNet with Pytorch. All of the above models were trained using the same dataset as our model. The experimental results on the test set show that our models are able to generate finer facial components compared to state-of-the-art methods.

Table 3 shows the quantitative comparisons on the test sets of CelebA-HQ and Helen datasets. Compared with other SISR and FSR methods, GFNet achieved the highest scores in both PSNR and SSIM tests. Table 4 shows the computational complexity and parameter amount of each method. We calculate the FLOPs on the assumption that the resolution of generated SR images is $128 \times 128$. Compared to other FSR methods, GFNet outperforms existing methods in parameter while the result of FLOPs is higher than the others. As can be seen from Fig. 6, most methods generate facial components with blurred shapes, in contrast, GFNet can generate results closer to the real images.

We found that SRCNN and VDSR in the SISR method could not achieve satisfactory results in the $8 \times$ FSR task. As shown in Table 3, their scores on PSNR and SSIM are far behind those of the FSR method. The visualization results
in Fig. 7 show that their recovered faces are blurred and some facial components are missing, e.g., eye and nose. This is due to the fact that the face images are much more complex compared to the images processed by the SISR task. In addition, we evaluated the model with 8x upscaling factors, and additional face information was needed for face image reconstruction at very low resolution. Note that the images recovered by SRResNet are perceptually closer to HR images than SRCNN and VDSR, due to the deeper network layers and the use of a large number of residual blocks in SRResNet [42]. This proves from the side that the proposed feature residual blocks inspired by ResNet are beneficial for the FSR task. GIDN proposed a network of distilling gradient information to improve performance. Compared with other SISR methods, GIDN can generate clearer outlines.

We compared the results of GFNet with those of PFSRNet, FSRNet and MSFSR. From the results in Table 3, we found that GFNet achieves the best PSNR and SSIM scores. The visualization results in Fig. 8 show that FSRNet cannot recover the correct facial components, so the images it generated are not realistic. This is because that although parsing maps can represent the contours of facial components, they cannot represent the intrinsic structures. Specifically, as shown in Fig. 9, a parsing map can represent the general shape of the eye, but it ignores the structure of the eye. PFSRNet generates facial components with artifacts, which is caused by the difficulty in detecting landmarks in very low-resolution images. In contrast, the proposed GFNet is able to generate more realistic face images and does not produce artifacts. The images generated by MSFSR using the enhanced

---

**TABLE 3.** Average of PSNR(dB) and SSIM values on CelebA-HQ and Helen at 8x SR. Note for FSRNet*: We re-implement FSRNet with Pytorch.

| Dataset    | SRCNN | VDSR | SRResNet | GIDN | PFSR | FSRNet* | MSFSR | Ours       |
|------------|-------|------|----------|------|------|---------|-------|------------|
| CelebA-HQ  | 20.70/0.5344 | 21.28/0.6140 | 23.10/0.6524 | 23.48/0.6752 | 24.07/0.6869 | 23.59/0.6704 | 24.49/0.7035 | 24.89/0.7124 |
| Helen      | 20.64/0.5735 | 21.97/0.6409 | 23.64/0.6785 | 23.89/0.6873 | 23.42/0.6718 | 24.01/0.6808 | 24.32/0.6983 | 24.53/0.7016 |

**TABLE 4.** Average of PSNR(dB) and SSIM values on CelebA-HQ and Helen at 8x SR.

| Parameter | SRCNN | VDSR | SRResNet | GIDN | PFSR | FSRNet* | MSFSR | Ours       |
|-----------|-------|------|----------|------|------|---------|-------|------------|
| Parameter | 57K   | 665K | 1090K    | 1640K| 8962K| 3179K   | 6305K | 1173k      |
| PLOPs     | 1.13G | 10.90G | 1.36G    | 1.18G| 10.07G| 15.45G  | 8.45G | 11.65G     |

**FIGURE 6.** Qualitative comparisons with the state-of-the-art methods on CelebA-HQ and Helen datasets.

**FIGURE 7.** Qualitative comparisons with the state-of-the-art SISR method on Helen dataset.
facial boundaries are similar to GFNet, but the facial structure is still blurred. As shown in Fig. 10, the gradient map is richer and more detailed for the facial structure compared to the enhanced facial boundary. The effectiveness of our use of gradient information to compensate for face features is demonstrated by qualitative comparisons with state-of-the-art methods.

E. LIMITATIONS

Generally, there remains a challenge in reconstructing images at 16× upscale factor for FSR methods. In Fig. 11, we show the visual results of GFNet on upscale factor of 16×. Although the GFNet compensates gradient information to face features to improve performance, the reconstructed images are not satisfactory and some facial components are missing, e.g., nose and lip. This is mainly because at 16× upscale factor, and it is challenging to obtain enough gradient information from LR images. The network cannot obtain enough additional face information, leading to the difficulty of reconstructing face images.

V. CONCLUSION

In this paper, we proposed a GFNet for FSR task, which consists of FRBs and GEBs. The FRB extracts the face features and the GEB extracts the gradient information. We also introduced a feature exchange mechanism between the two modules, which allows the network to utilize features at different levels. Extensive experimental results on the CelebA-HQ and Helen datasets demonstrate that GFNet can generate faces with finer facial components compared to other SR models at 8× upscale factor. Following the idea of this work, future work can be extended to develop better GEB and to investigate better combination method of gradient information with face features.

REFERENCES

[1] J. Yang, L. Luo, J. Qian, Y. Tai, F. Zhang, and Y. Xu, “Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 1, pp. 156–171, Jan. 2017.
[2] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, “DeepFace: Closing the gap to human-level performance in face verification,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Columbus, OH, USA, Jun. 2014, pp. 1701–1708.

[3] X. Li, S. Liu, J. Yang, and M.-H. Yang, “Generative face completion,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Honolulu, HI, USA, Jul. 2017, pp. 3911–3919.

[4] C. Dong, C. C. Loy, K. He, and X. Tang, “Learning a deep convolutional network for image super-resolution,” in Proc. Eur. Conf. Comput. Vis. Zurich, Switzerland: Springer, 2014, pp. 184–199.

[5] J. Kim, J. K. Lee, and K. M. Lee, “Accurate image super-resolution using very deep convolutional networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 1646–1654.

[6] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, “Photo-realistic single image super-resolution using a generative adversarial network,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Honolulu, HI, USA, Jul. 2017, pp. 4681–4690.

[7] M. S. Rad, B. Bozorgtabar, U.-V. Marti, M. Basler, H. K. Ekenel, and J.-P. Thiran, “SROBB: Targeted perceptual loss for single image super-resolution,” in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Seoul, South Korea, Oct. 2019, pp. 2710–2719.

[8] J. Kim, J. K. Lee, and K. M. Lee, “Deeply-recursive convolutional network for image super-resolution,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 1637–1645.

[9] Y. Zhang, Y. Wu, and L. Chen, “MSFSR: A multi-stage face super-resolution with iterative collaboration between attentive recovery and landmark estimation,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Seattle, WA, USA, Jun. 2020, pp. 5769–5778.

[10] C. Ma, Y. Yao, Y. Cheng, C. Chen, J. Lu, and J. Zhou, “Structure-preserving super resolution with gradient guidance,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Seattle, WA, USA, Jun. 2020, pp. 7769–7778.

[11] B. Meng, L. Wang, Z. He, G. Jeon, Q. Dou, and X. Yang, “Gradient information distillation network for real-time single-image super-resolution,” J. Real-Time Image Process., vol. 18, no. 2, pp. 333–344, Apr. 2021.

[12] C.-Y. Yang, S. Liu, and M.-H. Yang, “Structured face hallucination,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Portland, OR, USA, Jun. 2013, pp. 1099–1106.

[13] S. Liu and M.-H. Yang, “Compressed face hallucination,” in Proc. IEEE Int. Conf. Image Process. (ICIP), Paris, France, Oct. 2014, pp. 4032–4036.

[14] X. Pei, Y. Guan, P. Cai, and T. Dong, “Face hallucination via gradient constrained spare representation,” IEEE Access, vol. 6, pp. 4577–4586, 2018.

[15] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, “Image super-resolution using very deep residual channel attention networks,” in Proc. Eur. Conf. Comput. Vis. Pattern Recognit. (ECCV), Munich, Germany, 2018, pp. 286–301.

[16] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[17] Y. Tai, J. Yang, and X. Liu, “Image super-resolution via deep recursive residual network,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Honolulu, HI, USA, Jul. 2017, pp. 3147–3155.

[18] A. Odena, V. Dumoulin, and C. Olah, “Deconvolution and checkerboard artifacts,” Distill, vol. 1, no. 10, p. e5, Oct. 2016.

[19] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, “Photo-realistic single image super-resolution using a generative adversarial network,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, Jun. 2016, pp. 1870–1878.

[20] T. Karras, T. Aila, S. Laine, and J. Lehtinen, “Progressive growing of GANs for improved quality, stability, and variation,” 2017, arXiv:1710.10196.

[21] V. Le, J. Brandt, Z. Lin, L. Bourdev, and T. S. Huang, “Interactive facial feature localization,” in Proc. Eur. Conf. Comput. Vis. Florence, Italy: Springer, 2012, pp. 679–692.

[22] Z. Liu, P. Luo, X. Wang, and X. Tang, “Deep learning face attributes in the wild,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Santiago, Chile, Dec. 2015, pp. 3730–3738.