Proceedings of the 6th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/95/2024CH0052

# Personalized Super Resolution with Face Prior

**Hanzhang Lu**[1,3,*], **Ji Qi**[2,4]
[1]Chengdu University of Technology
[2]Beijing University of Technology

[3]lu.hanzhang@student.zy.cdut.edu.cn
[4]qiji21029210@emails.bjut.edu.cn
*corresponding author

**Abstract.** Face Super-Resolution (FSR) is a critical technology in computer vision that aims to reconstruct high-resolution facial images from low-resolution inputs. Despite recent advancements, current FSR methods struggle to accurately reconstruct personalized and detailed features. This paper proposes a novel FSR approach that addresses these challenges through a personalized feature extraction and fusion framework. Our method integrates a U-Net based downsampling mechanism to extract individual-specific features from high-resolution reference images, which are then fused with a pre-trained Generative Adversarial Network (GAN) for enhanced reconstruction. We introduce a comprehensive loss function that combines reconstruction, adversarial, facial component, and identity preservation losses to guide the learning process. Extensive experiments on the augmented FFHQ dataset demonstrate that our approach significantly improves the reconstruction of rich facial features, particularly for older individuals, outperforming existing state-of-the-art methods in both quantitative metrics and qualitative visual assessments.

**Keywords:** Image processing, Facial prior, Super-Resolution

## 1. Introduction

Face Super-Resolution (FSR), a specialized branch of single image super-resolution (SISR), aims to bridge this gap by algorithmically enhancing low-resolution facial images to high-resolution counterparts. The significance of FSR extends beyond mere visual enhancement, playing a pivotal role in facial recognition systems, forensic image analysis, and various multimedia applications.

Even though, in recent years, this field has seen major development, significant difficulties still remain in the processing of reconstructing human faces that are detailed and real. Problems surrounding human-face SR are often addressed by using image priors. The explicit way of introducing these priors is through reference images while the implicit method uses pre-trained generative models.

However, utilizing the auxiliary priors extracted from low-resolution images or pre-trained generative modes has several limitations: (1) priors derived from low-resolution images like geometry information inherently lack detailed features and texture, which results in a loss of finer details necessary for high-quality reconstruction. (2) general priors from generative models, $eg$. GANs, are unable to reconstruct personalized features like wrinkles well, which are crucial for maintaining the individuality and authenticity of the reconstructed image.

There are sufficient reasons to believe that using personalized super-resolution capabilities to enhance the super-resolution (SR) model can solve these issues. In this pursuit, we propose a novel FSR

DOI: 10.54254/2755-2721/95/2024CH0052

framework that incorporates personalized feature extraction and a fusion mechanism. Our approach leverages one or several high-resolution reference images of the target individual, processed through a U-Net architecture to extract personalized facial characteristics. Such kind of features extracted from high-resolution images contain rich and detailed facial features. These features are then integrated as prior knowledge into a pre-trained Generative Adversarial Network (GAN) to guide the reconstruction process. Also, we develop a comprehensive loss function to guide the training process and ensure high-quality super-resolution. we introduce reconstruction loss on U-net when doing feature extraction, identity preserving loss, adversarial loss for restoring realistic textures, and facial component loss to further enhance details. Moreover, we employ a boosting method during the training stage, which improves the model's performance when reconstructing complicated and rare features.

The key contributions of our work are as follows:

- We introduce a novel FSR architecture that synergistically combines personalized feature extraction with the generative power of pre-trained GANs, effectively reducing the search space while enhancing reconstruction quality, especially on personalized features.
- We develop a comprehensive loss function that integrates reconstruction, adversarial, facial component, and identity preservation objectives, ensuring both visual fidelity and identity consistency in the super-resolved images.
- We propose a data augmentation strategy for the FFHQ dataset, incorporating a higher proportion of images featuring older individuals, to improve the model's capability in reconstructing rich facial features across diverse age groups.
- We conduct extensive experiments demonstrating significant improvements over current methods, particularly in handling complex facial features and preserving personal identity.

## 2. Related work

Our work builds upon and extends several key areas in computer vision and image processing. This section provides a comprehensive overview of relevant research in image restoration, face super-resolution, and the application of prior knowledge in these domains.

**Image Restoration**. Image restoration is a process that can improve the quality of degraded images by means of various techniques available. Traditional methods often rely on analytical models and optimization techniques [1]. Deep learning has highlighted the dominance of convolutional neural networks (CNNs), offering superior performance in various restoration tasks [2]. Recent advancements include the use of generative adversarial networks (GANs) [3] and attention mechanisms [4], which have pushed the boundaries of image restoration quality.

**Face Super-Resolution.** Face super-resolution, a specialized subset of image restoration, focuses on the unique challenges posed by facial images. Early works in this domain adapted general super-resolution techniques to facial images [5]. However, the complex structure and high-level semantics of faces necessitated more specialized approaches.

Yu Chen et al. [6] proposed FSRNet, a pioneering work that incorporated facial landmark heatmaps and parsing maps as prior knowledge to guide the super-resolution process. This approach demonstrated the importance of facial structural information in achieving high-quality results. Building on this concept, Xintao Wang et al. [7] introduced the use of pre-trained GANs as a generative facial prior, significantly enhancing the realism of restored faces.

Despite these advancements, current methods often struggle with reconstructing personalized features, particularly for faces with rich textures such as those of older individuals. Our work addresses this limitation by introducing a personalized feature extraction mechanism.

**Prior Knowledge in Image Reconstruction.** The incorporation of prior knowledge has been a key factor in improving the performance of image reconstruction tasks. In the context of face super-resolution, priors can take various forms:

DOI: 10.54254/2755-2721/95/2024CH0052

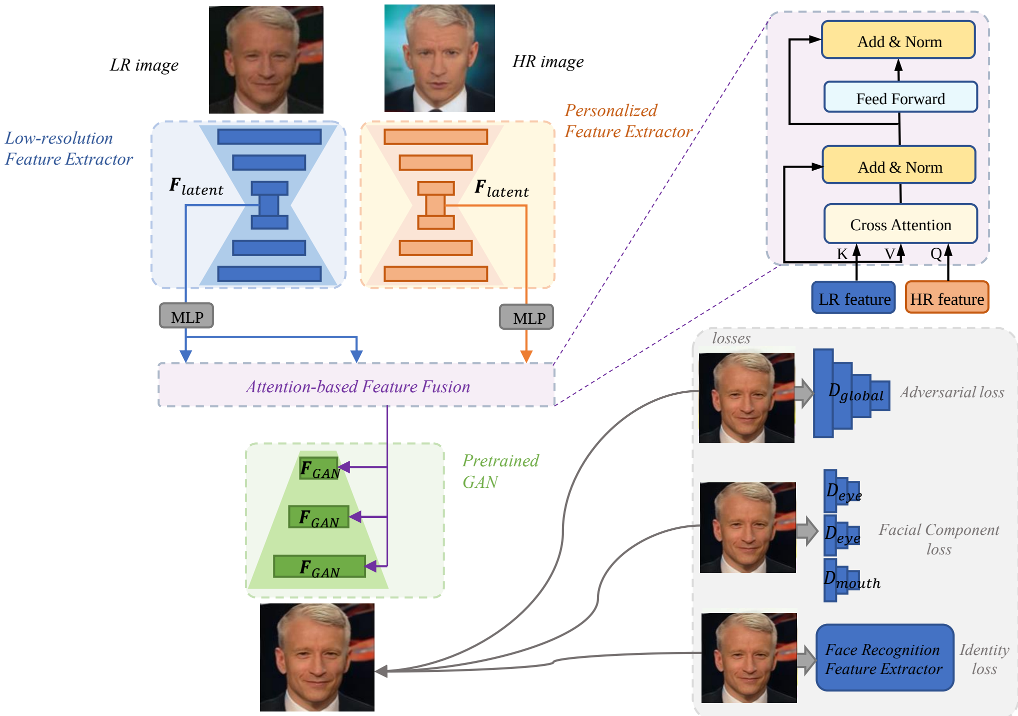



**Figure 1. Overview of PSP-GAN framework.** It consists of a low-resolution feature extractor, a high-resolution feature extractor, and a pretrained face GAN as facial prior. The features extracted from LR image and SR reference image are fused by a attention-based feature fusion module, and finally the fused feature is used to guide the reconstruction process of the pretrained GAN.

- **Geometric Priors:** Facial landmarks and parsing maps provide structural guidance, ensuring that reconstructed faces maintain proper proportions and feature locations [6].
- **Statistical Priors:** Learned from large datasets, these priors capture the statistical properties of facial images, guiding the reconstruction towards more plausible results [7].
- **Identity Priors:** Methods that incorporate identity-preserving constraints ensure that the reconstructed high-resolution face maintains the identity of the input low-resolution face [8].

**Feature Fusion.** Feature fusion is a crucial step in many machine learning and deep learning applications, particularly in scenarios where multiple sources of information need to be combined. Various methods, such as early fusion, attention-based fusion and late fusion, have been proposed to achieve effective feature fusion.

Early fusion concatenates features collected from various sources at the input level before feeding them into the model; and late fusion integrates the features at the decision level by merging the outputs of multiple models. However, these methods may not capture the complex interactions between features effectively.

Attention-based fusion methods have become especially popular because they are able to combine and weigh features dynamically according to how relevant they are to the current task. Attention mechanisms, initially introduced in the context of machine translation [9], have been widely adopted in various applications.

**Local Component Discriminators.** Recent work has shown the effectiveness of using local component discriminators to enhance the quality of facial details in restoration tasks. These discriminators focus on specific regions of the face, such as eyes, nose, and mouth, ensuring that these critical areas are reconstructed with high fidelity [10]. Our approach builds upon this concept by incorporating a facial component loss that specifically targets the enhancement of perceptually significant facial features.

In summary, while existing methods have made significant strides in face super-resolution, they often fall short in reconstructing personalized features, particularly for faces with complex textures. Our work addresses this gap by introducing a novel framework that combines personalized feature extraction with the generative power of GANs, guided by a comprehensive loss function that ensures both overall quality and preservation of individual characteristics.

## 3. Methodology

We describe a GAN-Based framework in this section. The aim of face super-resolution is giving an input image $\boldsymbol{x}$from degradation and model trying to estimate a high-quality image $\hat{y}$ which is visually closer to the ground truth $\boldsymbol{y}$ while maintaining facial details(wrinkles, freckles, etc.) as possible.

Our proposed approach to face super-resolution integrates personalized feature extraction with a pre-trained GAN framework, addressing the challenges of reconstructing unique facial characteristics while maintaining overall image quality. This section details model architecture, and the comprehensive loss function guiding our training process.

### *3.1. Model Architecture*

Our model architecture, illustrated in Figure 1, consists of three main components: a Personalized Feature Extractor, a Feature Fusion Module, and a Pre-trained GAN Network.

*3.1.1. Low-resolution Feature Extractor* In practical blind face restoration scenarios, images often suffer complex degradation such as noise, low-resolution and blur. Thus, we need The low-resolution feature extractor module to remove the degradation and extract feature$\boldsymbol{F_{latent}}$. In the framework, we employ a U-Net[11] structure as the core component for mitigating degradation. The U-Net design an encoder pathway which can efficiently extract multi-level features while the decoder pathway can reconstruct these features while preserving vital spatial information for locating the features. By doing so, we aim to address the complexities inherent in degradation processes effectively.

Formally, given a low-resolution image $I_{LR}$, $F_{latent}$ can be extracted from the Low-resolution Feature Extractor $E_{LR}$:

$$F_{latent} = E_{LR}(I_{LR}) \tag{1}$$

where $F_{latent}$ is the representation of low-resolution image.

*3.1.2. Personalized Feature Extractor* Similar with the low-resolution Feature Extractor, the Personalized Feature Extractor also utilizes a U-Net architecture to process a high-resolution reference image of the target individual. This component is designed to capture person-specific facial characteristics across multiple scales.

Formally, given a high-resolution reference image $I_{HR}$, the Personalized Feature Extractor $E_p$ produces a set of multi-scale features:

$$F_p = E_p(I_{HR}) \tag{2}$$

where $F_p$ represents the extracted personalized features at different scales.

DOI: 10.54254/2755-2721/95/2024CH0052

*3.1.3. Feature Fusion Module* The Feature Fusion Module combines the personalized features extracted from the reference image with the latent representation of the low-resolution input. This module employs an attention mechanism to dynamically weight the importance of different feature channels, allowing for adaptive integration of personalized and general facial features.

Let $F_{LR}$ be the features extracted from the low-resolution input $I_{LR}$ and $F_p$ as the features extracted from the high-resolution reference image. The fusion process can be described as:

$$\begin{aligned} F_{fused} &= Fusion(F_p, F_{LR}) \\ &= Attention(Q = F_p, K = F_{LR}, V = F_{LR}) \end{aligned} \tag{3}$$

where $Fusion(\cdot)$ represents the attention-based fusion operation. More specifically, $F_{LR}$ acts as both the key and the value, $F_p$ acts as the query. The cross-attention mechanism computes the attention weights, enabling the model to incorporate personalized features effectively.

*3.1.4. Pre-trained GAN Network* We leverage a pre-trained GAN as the backbone of our super-resolution network. The GAN's generator $G$ is fine-tuned using the fused features to produce the final high-resolution output:

$$I_{SR} = G(F_{fused}) \tag{4}$$

where $I_{SR}$ is the super-resolved output image.

*3.2. Model Objectives*

In the real-world super-resolution task, to improve the realism and accuracy of image generation as much as possible under the premise of guaranteeing the training cost and efficiency, we mainly adopt the GFP-GAN[12] architecture, using multiple loss functions as the approximation function of the training model.

*Ensure the authenticity of the output.* The basic task of super-resolution task is generating images which close to the ground truth. Using loss functions, constraints the outputs $\hat{y}$ close to the ground-truth $y$. In the pixel level, the reconstruction loss $\mathcal{L}_{rec}$ ensures the super-resolved output similar to real high-resolution image perceptually.Considering the L1 loss and perceptual loss[13, 14], the reconstruction loss function $\mathcal{L}_{rec}$ described as follow:

$$\mathcal{L}_{rec} = \lambda_{l1}\|I_{SR} - I_{HR}\|_1 + \lambda_{per}\|\phi(I_{SR}) - \phi(I_{HR})\|_1 \tag{5}$$

where $\phi$ represents the VGG-19 network and $\lambda_{l1}$ and $\lambda_{per}$ represents the loss weight of L1 and perceptual loss.

Moreover, in the deep feature space, the architecture introduces an feature preserving loss $\mathcal{L}_{pre}$. Inspired by perceptual loss, identity discrimination can get the most prominent features from the feature embedding of the input image. Adopt with the pretrained face recognition ArcFace[15]model, the feature preserving loss is defined as follow:

$$\mathcal{L}_{pre} = \lambda_{pre}\|\eta(\hat{I_{SR}}) - \eta(I_{HR})\|_1 \tag{6}$$

where $\lambda_{pre}$ represent the weight of preserving loss and $\eta$ denotes face feature extractor.


**Figure 2.** Qualitative comparison for human face restoration on both young and old people. Our model produces more faithful personalized details than GFP-GAN.

*Maintain the details of the output.* In order to better generate the details of the output images, we train the different discriminators separately. By dividing facial space into different regions which we are interested, the discriminators can perceptually improve the significant facial features close to the natural distributions of different facial areas.

In this model, there are three different discriminators which could cut facial space into three regions: mouth, right eyes and left eyes. We describe the loss of those regions of interest as followed: $\mathcal{L}_{ROI}$:

$$\mathcal{L}_{ROI} = \sum_{i \in \{eyes, nose, mouth\}} \lambda_{local} \mathbb{E}_{\hat{\mathbb{I}_{\mathbb{SR}}}} [\log(1 - D(I_{SR}))] \tag{7}$$

where $\lambda_{local}$ represent the loss weights of local discriminative loss and D is the local discriminator for the region from the component collection.

To further encourage model to favor the solutions in the natural image, we employ adversarial loss $\mathcal{L}_{adv}$. Similar to the StyleGAN2[16], the logistic loss[17] is adopted:

$$\mathcal{L}_{adv} = -\lambda_{adv} \mathbb{E}[\log D(I_{SR})] \tag{8}$$

where $D$ is the discriminator network and $\lambda_{adv}$ represents the adversarial loss weight .

The overall loss is the the combination of the above losses:

$$\mathcal{L}_{total} = \mathcal{L}_{rec} + \mathcal{L}_{adv} + \mathcal{L}_{ROI} + \mathcal{L}_{pre} \tag{9}$$

### *3.3. Training Strategy*

We adopt a progressive training strategy, starting with datasets such as FFHQ [18] that exhibit significant diversity in gender, race, and age. In the final epochs of training, we enhance the model with more challenging samples, focusing on older faces with rich textures.

This comprehensive approach enables our model to effectively tackle the challenges of personalized face super-resolution, especially for faces with rich and complex features.

## 4. Experiment results

### *4.1. Datasets and Implementation*

*Dataset* We base our work on the FFHQ [18] dataset (Flickr-Faces-HQ), a high-quality collection of 70,000 facial images at 1024x1024 resolution. To address the limitations in existing datasets and models regarding the representation of diverse age groups and facial features, we augment the FFHQ dataset as follows:

- 70% of the training images are retained from the original FFHQ dataset, maintaining a broad representation of facial variations.

- 30% of the training images are specifically selected to represent older individuals, focusing on features such as pronounced wrinkles, age spots, and other complex textures typically underrepresented in standard datasets.

This augmentation strategy ensures that our model is exposed to a wider range of facial features during training, improving its ability to reconstruct diverse faces accurately.

*Data Process.* To create training pairs, we follow the practice in GFP-GAN[7]:

$$x = [(y * k_\sigma) \downarrow r + n_\delta]_{\mathrm{JPEG}_q} \tag{10}$$

First, we use the Gaussian blur kernel $k_\sigma$ to convolve the high-resolution image(GT) $y$. Then, we perform a down-sampling operation using a scale factor $r$. Next, we add the additive white Gaussian noise $n_\delta$ to the image. Lastly, JPEG compresses the image using quality factor $q$.

*Implementation.* For the general facial prior, we choose the pretrained StyleGAN2 [18] with $512^2$ outputs. In order to achieve compact model size, we set StyleGAN2's channel multiplier to one. The UNet for low-resolution and high-resolution feature extractor include seven upsamples and seven downsamples, with every sample having a residual block [19].

The size of the mini-batch for training is 3. We use color jittering and horizontal flip to increase the training data. The three components used to compute face component loss are mouth, right eye, and left eye because of their perceptual significance. We use Adam optimizer Adam optimizer [20] to train our model for 800k iterations in total. We set the learning rate to $2 \times 10^{-3}$ and then it decayed at the 700k-th, and 750k-th iterations by a factor of 2. The implementation of the models takes place using the PyTorch framework, and their training is supported by a NVIDIA GeForce RTX 3090 GPU.

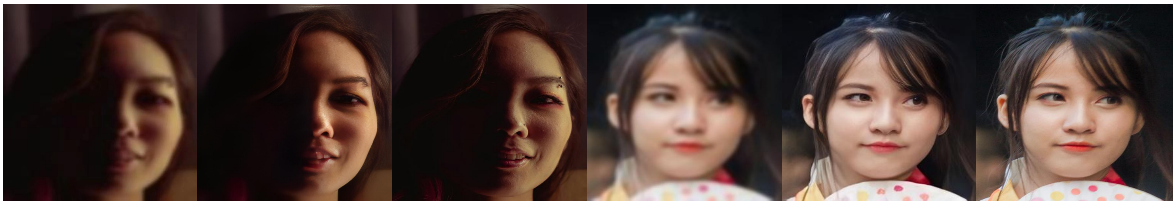
### *4.2. Experimental Results*

We first target images of young people, comparing in our study a self-developed facial restoration model with a GFP-Gan model. Both techniques are mainly used to process facial images of young people and are evaluated comparatively in terms of detail recovery, colour reproduction and visual effects. Analysed in terms of detail recovery, the GFP model tends to generate smoother and more embellished facial features, sometimes at the expense of a certain level of realism in this processing. For example, facial wrinkles and skin texture are over-smoothed in images processed by the GFP model, which improves visual appeal but may also result in the loss of real details of facial features. On the contrary, our model focuses more on restoration while maintaining the quality of the original image and is able to more realistically reproduce the details of the eyes, hair and skin. In terms of colour reproduction, our model shows a clear advantage in that it is more accurate to reproduce the original tones and brightness of the image, maintaining the naturalness and realism of the colours. This is especially important for applications that require a highly realistic reflection of the subject's own features. The GFP model, on the other hand, is more aggressive in its colour processing, enhancing the saturation and contrast of colours, which makes the image more vibrant and striking, but at the same time may cause unnatural and over-embellished colours. In terms of visual effects, although the images from the GFP model are more appealing at first glance, our model does a better job of maintaining the natural expressions and features of the subjects. This treatment not only preserves the character's personality and authenticity, but also avoids the distortion of the character's features caused by over-embellishment. It seems that our model has a clear advantage in maintaining colour authenticity and realistic reproduction of facial details, and is more suitable for application scenarios that require a high degree of realism. The GFP model, on the other hand, may be more suitable for the commercial advertising or entertainment media industries, where more vibrant colours and smooth facial features may be preferred.

In image processing for the elderly, we are concerned with the effectiveness of the two models in processing images of the elderly, especially in terms of detail and realism preservation. Our model is

DOI: 10.54254/2755-2721/95/2024CH0052

optimised specifically for wrinkles in the elderly and its ability to capture and reproduce wrinkle details is strengthened by data augmentation and the inclusion of a specifically elderly dataset. In contrast, the GFP model underwent excessive smoothing in some cases, compromising the realistic representation of facial features. It performed well in preserving the facial features of the elderly, especially the wrinkles near the corners of the eyes, mouth and nose. This processing maintains the age-specific features of the subjects and enhances the naturalness and expressiveness of the facial expressions. This is especially evident in the comparison images, where our model is able to clearly show each wrinkle without making the face look too smooth or distorted. For the image of the older woman, our model more accurately reproduces the natural colours and details of the hair, especially in the darker areas of the hair and the changes in light and shadow. This is essential to enhance the visual realism of the overall image. On the contrary, the GFP model, when processing the same image, although the overall colours are more vivid, there are some unnatural dark spots in the dark processing of the hair, which may be due to the fact that the model has over-enhanced the contrast of the image during smoothing.



Input Our Model GT Input Our Model GT

**Figure 3.** Limitations for Asian People's Facial Features

*4.3. Limitations*

Since the dataset is mainly optimised for facial features of Western populations, when dealing with low-resolution images of Asians, especially when the eyes are particularly blurred, our model tends to recover the eyes according to Western eye features. This may result in the recovered image eyes showing light brown eyes that do not match Asian features. Meanwhile the position of the eyeballs may be shifted. Such positional shifts may even cause visual disharmony, such as eyes appearing unnaturally large or small, further reducing the natural authenticity of the face and possibly affecting the accuracy of the individual's identification. In addition, the model also performs poorly in preserving details of some fine facial decorations, such as eye make-up and eyebrow studs, which are important elements in cultural expression and personal style, but are often overlooked or lost in the processing.

While our model performs well in maintaining wrinkles and hair details in older adults, it still leaves much to be desired in terms of overall image clarity enhancement, especially at very low resolutions. The current model's enhanced ability to handle details is not sufficient to fully overcome the effects of low input image quality. These limitations highlight the need for future research.

## 5. Discussion and future works

In this project, we tackled the challenge of enhancing low-resolution human facial images using Face Super-Resolution (FSR) technology. Our approach mainly focuses on tackling the challenge of handling personalized and detailed features in facial images. To meet these challenges, we proposed a novel method incorporating personalized feature extraction and fusion modules and develop comprehensive loss functions. Additionally, we used a boosting training method to improve the reconstruction of specific features. Our results showed significant improvements in the reconstruction of high-quality, detailed facial images, particularly in handling rich features.

However, there are still several areas that need improvement. Firstly, we did not propose or utilize specific metrics to measure the quality of the super-resolution images. Secondly, the computational efficiency of our approach needs to be optimized to ensure faster training and inference, making the model more practical for real-world applications.

Given more time and resources, we would pursue the following avenues to extend and improve our work:

- Quantity Metric: Use quantity metrics like FID [21] and NIQE [22] to evaluate the quality of human faces super resolution rather than just providing visualization.
- Cross-Modality Feature Fusion: Explore the integration of features from different imaging modalities (e.g., thermal, depth) to enhance robustness in challenging environmental conditions.
- Adaptive Reference Selection: Develop intelligent algorithms for automatically selecting or synthesizing optimal reference images for each input, potentially improving efficiency and accuracy in real-world applications.

## 6. Acknowledgement

Ji Qi and Hanzhang Lu contributed equally to this work and should be considered co-first authors.

## References

[1] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," *European conference on computer vision*, 2014.

[2] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, 2017.

[3] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017.

[4] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," *Proceedings of the European conference on computer vision (ECCV)*, 2018.

[5] S. Baker and T. Kanade, "Hallucinating faces," *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 2000.

[6] Y. Chen, Y. Tai, X. Liu, C. Shen, and J. Yang, "Fsrnet: End-to-end learning face super-resolution with facial priors," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[7] X. Wang, Y. Li, H. Zhang, and Y. Shan, "Towards real-world blind face restoration with generative facial prior," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.

[8] K. Grm, F. Pernuš, L. Ibrahimi, and V. Štruc, "Face hallucination using cascaded super-resolution and identity priors," *IEEE Transactions on Image Processing*, 2020.

[9] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[10] J. Li, F. Fang, K. Mei, and G. Zhang, "Enhanced super-resolution generative adversarial networks," in *European Conference on Computer Vision*, Springer, 2020.

[11] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, Springer, 2015.

[12] X. Wang, Y. Li, H. Zhang, and Y. Shan, "Towards real-world blind face restoration with generative facial prior," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

[13] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European Conference on Computer Vision (ECCV)*, pp. 694–711, Springer, 2016.

[14] C. Ledig, L. Theis, F. Huszár, 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 *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4681–4690, 2017.

[15] J. Deng, J. Guo, X. Niannan, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4690–4699, 2019.

[16] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of stylegan," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8110–8119, 2020.

[17] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[18] T. Karras, S. Laine, and T. Aila, "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, 2019.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *arXiv preprint arXiv:1512.03385*, 2015.

[20] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.

[21] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," in *Advances in Neural Information Processing Systems*, pp. 6626–6637, 2017.

[22] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2012.