APSIPA Transactions on Signal and Information Processing, 2024, 13, e36

Overview Paper

Deep Learning for Face Super-Resolution: A Techniques Review

Bolin Zhu[1], Kanghui Zhao[1#], Tao Lu[1*], Junjun Jiang[2], Zhongyuan Wang[3], Kui Jiang[2] and Zixiang Xiong[4]

[1] *Hubei Key Laboratory of Intelligent Robot, Wuhan Institute of Technology, Wuhan, China*
[2] *School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China*
[3] *School of Computer Science, Wuhan University, Wuhan, China*
[4] *Dept. Electrical and Computer Engineering, Texas A&M University, College Station, Texas, USA*

ABSTRACT

Face Super-Resolution (FSR) represents a significant branch of image super-resolution, aiming to reconstruct low-resolution face images into high-resolution counterparts. Recently, driven by rapid advancements in deep learning technology, FSR methods using deep learning have achieved notable subjective and objective reconstruction quality, attracting extensive industrial attention. However, detailed classifications of FSR methods remain limited. Therefore, this survey systematically and comprehensively reviews deep learning-based FSR methods. Initially, we introduce the background and technical framework of FSR. Subsequently, we detail the FSR problem definition, alongside commonly used datasets, evaluation metrics, and loss functions. We conduct comprehensive researches in deep learning FSR methods and classify them according to their solution strategies. Within each category, we begin with a general method description, and subsequently introduce

*Corresponding author: lutxyl@gmail.com
#Kanghui, Zhao is co-first author.

Received 11 July 2024; revised 08 September 2024; accepted 07 October 2024
ISSN 2048-7703; DOI 10.1561/116.20240045

representative approaches and discuss their respective pros and cons. Finally, we address current challenges in FSR methods and propose future research directions.

Keywords: Face super-resolution, deep learning, survey, face characteristics

1 Introduction

With the surge of the digital era, video calls, live streaming, and surveillance cameras are constantly generating massive amounts of face images. However, limited by device performance, transmission bandwidth, and other factors, these face images often have low resolution, which restricts the further development of related applications. Face Super-Resolution (FSR) technology can enhance the resolution and detail level of low-quality (LQ) and low-resolution (LR) face images. Therefore, its extensive application value and research content have always been a hot topic in the fields of image processing and computer vision.

In 2000, Baker and Kanade [2] first proposed the concept of FSR (Face Super-Resolution), using mathematical theoretical models to restore low-resolution face images. Subsequently, studies by Liu *et al.* [73], Gunturk *et al.* [35], Wang *et al.* [114], Chakrabarti *et al.* [7], Park and Lee [90], and Yang *et al.* [129] mainly focused on reconstructing low-resolution face images through global methods. On the other hand, studies by Chang *et al.* [8] Kim and Kwon [58], Ma *et al.* [86], and Yang *et al.* [130] primarily used local methods such as sparse coding, neighbor embedding, and local patch-based representation. Since then, FSR has become a mainstream research direction, with increasing innovative methods being proposed. For instance, Huang *et al.* [45] used canonical correlation analysis (CCA) to reconstruct details, while Wang *et al.* [122] used Gaussian and Laplace norms to solve the problem within a Bayesian framework. With the development of deep learning, studies by Zhou *et al.* [149], Huang *et al.* [46], and Cai *et al.* [6] combined deep learning with FSR and applied them to images or videos, achieving good performance. As FSR continues to develop, organizing the existing different deep learning-based FSR methods becomes much important. In this paper, we conduct a comparative study of various deep learning-based FSR methods.

The main contributions of this survey are as follows:

- The survey provides a comprehensive review of the latest deep learning-based FSR technologies, including problem definitions, commonly used evaluation metrics and loss functions, face datasets, and various types of deep learning-based FSR methods.

- The survey highlights the various architectural designs and techniques used in existing deep learning-based methods, demonstrating how they achieve good performance on both subjective and objective metrics.
- The survey investigates and discusses the existing issues in the FSR field and provides insights into future developments.

In the following, Figure 1 shows the basic framework of the survey. In the second section, we introduce the definition of the FSR problem, as well as some face datasets, evaluation metrics, and loss functions. In the third section, we discuss existing deep learning-based FSR methods. Based on the different ways of solving the blur kernel in FSR, these methods are categorized into five types: General FSR, Prior-guided FSR, Reference FSR, Multi-task FSR, and Blind FSR. Finally, in the fourth section, we summarize the issues presented in existing FSR methods, further discuss their limitations, and look forward to further technological advancements.

2 Problem Settings and Terminology

2.1 Problem Definition

FSR focuses on recovering the corresponding HR face image from an observed LR face image. The general mathematical degradation model can be written as:

$$I_{HR} = \psi^{-1}(I_{LR}, \delta), \tag{1}$$

where ψ^{-1} represents the inverse operation of the face image degradation process, including blur kernel, downsampling, and noise, etc. I_{HR} denotes the original HR face image, and δ represents reconstruction function parameters. FSR aims to approximate the inverse operation of the degradation model, but can only achieve results close to it. This process can be represented as:

$$I_{SR} = F(I_{LR}, \delta), \tag{2}$$

where F is the FSR model (inverse degradation model), δ represents the parameters of F, and I_{SR} represents the super-resolved results. The optimization of δ can be defined as:

$$\hat{\delta} = \underset{\delta}{argmax}\, L(I_{SR}, I_{HR}), \tag{3}$$

where L represents the loss function, and $\hat{\delta}$ denotes the optimal parameters of the trained model.

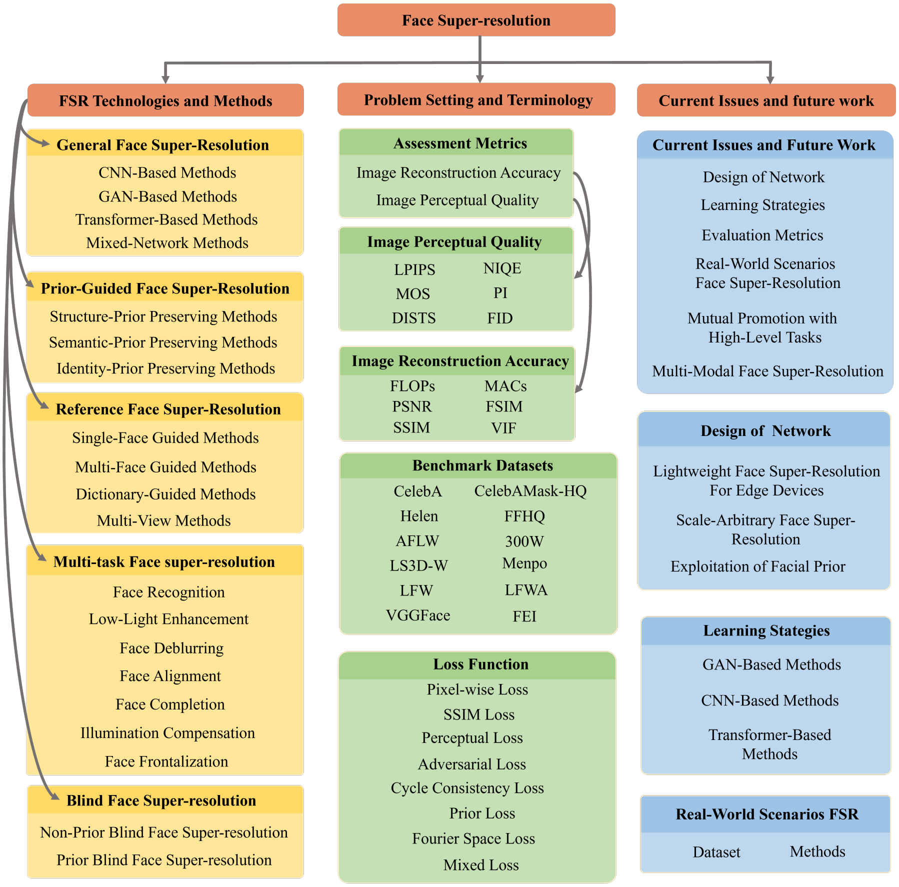
Face Super-resolution
FSR Technologies and Methods
Problem Setting and Terminology
Current Issues and future work
General Face Super-Resolution
CNN-Based Methods
GAN-Based Methods
Transformer-Based Methods
Mixed-Network Methods
Prior-Guided Face Super-Resolution
Structure-Prior Preserving Methods
Semantic-Prior Preserving Methods
Identity-Prior Preserving Methods
Reference Face Super-Resolution
Single-Face Guided Methods
Multi-Face Guided Methods
Dictionary-Guided Methods
Multi-View Methods
Multi-task Face super-resolution
Face Recognition
Low-Light Enhancement
Face Deblurring
Face Alignment
Face Completion
Illumination Compensation
Face Frontalization
Blind Face Super-resolution
Non-Prior Blind Face Super-resolution
Prior Blind Face Super-resolution
Assessment Metrics
Image Reconstruction Accuracy
Image Perceptual Quality
Image Perceptual Quality
LPIPS
NIQE
MOS
PI
DISTS
FID
Image Reconstruction Accuracy
FLOPs
MACs
PSNR
FSIM
SSIM
VIF
Benchmark Datasets
CelebA
CelebAMask-HQ
Helen
FFHQ
AFLW
300W
LS3D-W
Menpo
LFW
LFWA
VGGFace
FEI
Loss Function
Pixel-wise Loss
SSIM Loss
Perceptual Loss
Adversarial Loss
Cycle Consistency Loss
Prior Loss
Fourier Space Loss
Mixed Loss
Current Issues and Future Work
Design of Network
Learning Strategies
Evaluation Metrics
Real-World Scenarios Face Super-Resolution
Mutual Promotion with High-Level Tasks
Multi-Modal Face Super-Resolution
Design of Network
Lightweight Face Super-Resolution For Edge Devices
Scale-Arbitrary Face Super-Resolution
Exploitation of Facial Prior
Learning Stageties
GAN-Based Methods
CNN-Based Methods
Transformer-Based Methods
Real-World Scenarios FSR
Dataset
Methods

Figure 1: The basic framework of the survey

In real-world situations, it is usually impossible to obtain precise details about the degradation model and its related parameters. Therefore, researchers often use mathematical models to simulate the degradation process as accurately as possible. This approach helps generate LR and HR image pairs that are crucial for training purposes. The simplest mathematical model is

$$I_{LR} = (I_{HR}) \downarrow_s, \tag{4}$$

where ↓ denotes the downsampling operation, and s is the scaling factor. However, this basic pattern falls short of accurately replicating real-world degradation processes. To better emulate these real-life scenarios, researchers have developed degradation processes that combine multiple operations as below:

$$I_{LR} = J((I_{HR} \otimes k) \downarrow_s + n), \tag{5}$$

where k is the blurring kernel, $\otimes$ represents the convolutional operation, n denotes the noise, and J denotes the image compression.

2.2 Benchmark Datasets

With the introduction of FSR concept, an increasing number of face datasets have emerged for use. These datasets vary in several aspects, such as the number of images released and the number of annotated features included. As shown in Table 1, we have listed some face image datasets and provided their publication dates, the number of images included, and the number of annotated features they contain.

Table 1: Summary of public face image datasets for FSR.

Dataset	Release Time	Number	Feature-Points
CelebA [80]	2015	202,599	5
CelebAMask-HQ [65]	2020	30,000	19
Helen [64]	2012	2330	68
FFHQ [53]	2019	70,000	×
AFLW [61]	2011	25,993	21
300W [99]	2013	3,827	68
LS3D-W [5]	2017	230,000	×
Menpo [72]	2017	×	68
LFW [44]	2019	13,233	×
LFWA [80]	2016	13,143	40
VGGFace [91]	2015	2.6 million	×
FEI [19]	2006	2,800	×
LFPW [3]	2011	1432	29
AFW [152]	2012	205	6
WiderFace [133]	2015	32,203	×
WFLW [125]	2018	10,000	98

2.3 Assessment Metrics and Loss Function

In the realm of deep learning-based FSR methods, the selection of a loss function can gauges the disparity between I_{HR} and I_{SR}, significantly influences the training guidance. Once a network is effectively trained, the reconstruction efficacy of these methods can be assessed using evaluation metrics. In practical applications, the choice of an appropriate loss function can be tailored to suit specific needs. Given the interplay between loss functions and evaluation metrics, we will explore them collectively in this section.

2.3.1 Assessment Metrics

Generally, there are two main methods for quality evaluation: subjective evaluation and objective evaluation. Subjective evaluation relies on human judgment, typically involving interviewers who watch and assess the quality of the generated images. This method always produces results consistent with human perception, but it is time-consuming, inconvenient, and expensive. In contrast, objective evaluation primarily uses statistical data to reflect the quality of the generated images. Objective evaluation methods often yield results that differ from subjective evaluation metrics because they are based on mathematical calculations rather than human visual perception, which can lead to controversies in assessing image quality. Here, we introduce the evaluation metrics.

Floating Point Operations(FLOPs): FLOPs refers to the number of floating-point calculations required by an algorithm or model. It is commonly used to measure the computational complexity of a model. The larger the FLOPs, the more computational resources are needed, indicating a higher model complexity.

Multiply–Accumulate Operations(MACs): Represent multiply-add operations, where 1 MACs includes one multiplication and one addition, roughly equivalent to 2 FLOPs. Therefore, there is typically a 2x relationship between MACs and FLOPs.

Peak Signal-to-Noise Ratio(PSNR) [48]: Given I_{HR} and I_{SR}, the mean square error(MSE) between them is firstly calculated, then the PSNR is obtained:

$$MSE = \frac{\|I_{SR} - I_{HR}\|_2^2}{HWC}, \tag{6}$$

$$PSNR = 10 \log_{10} \frac{M^2}{MSE} \tag{7}$$

where H, W and C denote the height, width, and channel of the image, respectively. M is the maximum possible pixel value (i.e., 255 for 8-bit images). The smaller the pixel-wise difference between the two images, the higher the PSNR.

Structural Similarity (SSIM): SSIM [123] measures the structural similarity between two images. To be specific, SSIM measures similarity from three aspects: luminance, contrast, and structure. Given I_{HR} and I_{SR}, SSIM is obtained by

$$SSIM = l(I_{HR}, I_{SR}) * C(I_{HR}, I_{SR}) * S(I_{HR}, I_{SR}), \tag{8}$$

where $l(I_{HR}, I_{SR})$, $C(I_{HR}, I_{SR})$ and $S(I_{HR}, I_{SR})$ represent the similarity of the luminance, contrast and structure. SSIM ranges from 0 to 1. The higher the structural similarity of the two images, the larger the SSIM.

Feature Similarity (FSIM): FSIM [142] is a variant of SSIM that accounts for the non-uniform importance of pixels within an image. For instance, pixels along the edges of objects are deemed more critical for delineating the object's structure compared to those in background regions. Thus, the enhancement over SSIM lies in FSIM's ability to differentiate important areas and assign suitable weights accordingly. For grayscale images, the calculation method of FSIM is as follows:

$$FSIM = \frac{\sum_{x \in \Omega} S_L(x) * PC_m(x)}{\sum_{x \in \Omega} PC_m(x)}, \tag{9}$$

where $x \in \Omega$ denote the I_{HR} and I_{SR} images that need to be calculated, $S_L(x)$ is the product of the similarity between the extracted gradient features and phase features controlled by some parameters to determine their proportion. $PC_m(x)$ represents the phase consistency between I_{HR} and I_{SR} images. When calculating color images, it is necessary to consider chromaticity consistency:

$$FSIM_c = \frac{\sum_{\Omega} S_{PC}(x) * S_G(x) * [S_I(x) * S_Q(x)]^{\lambda} * PC_m(x)}{\sum_{\Omega} PC_m(x)}, \tag{10}$$

where $S_{PC}(x)$ and $S_G(x)$ representing the phase consistency and gradient feature similarity of I_{HR} and I_{SR} images, respectively. $[S_I(x) * S_Q(x)]^{\lambda}$ denote the color similarity.

Visual Information Fidelity (VIF): VIF [102] measures the preservation of visual information between distorted and original images. VIF evaluates image quality by comparing structural and textural similarities between two images. The I_{SR} and I_{HR} are divided into non-overlapping blocks, and structural (luminance, contrast, structural similarity) and textural (color, orientation, frequency) information are computed separately. VIF values range from 0 to 1, with values closer to 1 indicating higher visual fidelity and better quality.

Learned Perceptual Image Patch Similarity (LPIPS): LPIPS [145] is used to measure the distance between two images in deep feature space, and compared to PSNR and SSIM, LPIPS is more in line with human perception. The more similar the two images are, the smaller the LPIPS value.

Mean Opinion Score (MOS): Contrasting with the objective quantitative metrics mentioned earlier. MOS [104] is obtained by soliciting perceptual quality scores from human raters for the tested images. The final MOS value is calculated as the arithmetic mean of these ratings. The reliability of MOS can vary depending on the number of human raters involved. With a small number of raters, MOS may be biased, whereas it tends to be more faithful with a larger number of raters, providing a better representation of perceived image quality.

Differentiable Image Saliency Transform for Improved Scalability and Portability of Image Quality Assessment (DISTS): DISTS [20] is a differentiable image transformation method that converts any input image into a representation based on the saliency relationships between pixels. DISTS first computes a saliency score for each pixel and then reorders the pixels accordingly to construct a new image representation. As DISTS is differentiable, it can be directly used to train neural networks, enhancing the scalability and portability of image quality assessment. Traditional image quality assessment methods often rely on metrics based on human visual perception, such as PSNR and SSIM. However, these methods may not adapt well to different application scenarios and hardware platforms, and they have high computational complexity, making them unsuitable for resource-constrained environments like mobile devices.

Natural Image Quality Evaluator (NIQE): NIQE [87] measures the distance between two multivariate Gaussian models: one fitted to natural images and the other to the evaluated image, without requiring ground truth images. Specifically, NIQE fits a multivariate Gaussian model using quality-aware features derived from natural scene statistics. These features capture characteristics common in natural images. The lower the NIQE score, the better the visual quality of the evaluated image, indicating closer similarity to natural image statistics and thus higher perceived quality.

Perceptual Index (PI): PI is a metric based on NIQE. Generally, the lower the PI value, the more pleasant and better the image quality appears.

Fréchet Inception Distance (FID): FID focuses on the difference between I_{HR} and I_{HR} in a distribution-wise manner, and it is always applied to assess the visual quality of face images. The better the visual quality, the smaller the FID.

2.3.2 Loss Function

Loss functions provide the target metric that the network aims to minimize during the training process. The most common ones are pixel-based L_1 loss and pixel-based L_2 loss (also known as MSE loss), but these can respectively lead to slow convergence and smooth images. Subsequently, more loss functions have been proposed by researchers, such as Pixel-wise Loss and Perceptual Loss.

Pixel-wise Loss: Pixel-wise loss is used to compute the loss between the predicted image and the target image at the pixel level. Common examples include L_1 loss, L_2 loss, Huber loss [52] and Carbonnier penalty function [62]. Pixel-wise losses can enhance the PSNR of the generated images, but they often result in images that are overly smooth and lack fine details.

SSIM Loss: SSIM loss builds on MSE loss by focusing more on the structural similarity between super-resolved image and the original HR one rather than calculating pixel-by-pixel differences:

$$\mathcal{L}(I_{HR}, I_{SR}) = \frac{1}{2}(1 - F_{SSIM}(I_{HR}, I_{SR})), \tag{11}$$

where F_{SSIM} denotes the function of SSIM. Except for SSIM loss, multi-scale SSIM loss can calculate SSIM loss at different scales.

Perceptual Loss: Perceptual loss compares the high-level perceptual and semantic differences between images. It extracts the lower-level features of an image from the output of the early layers of a pre-trained network, then uses a simple pixel-level loss to compare the differences between the feature tensors of the target and the output values:

$$\mathcal{L}(I_{HR}, I_{SR}, \Psi, l) = \|\Psi^l(I_{HR}) - \Psi^l(I_{SR})\|_2, \tag{12}$$

where Ψ is the pretrained network and l is the l-th layer. The I_{SR} produced using perceptual loss often appears more visually appealing but tends to have lower PSNR compared to methods based on pixel-wise losses.

Cycle Consistency Loss: Cycle consistency loss is proposed by CycleGAN [151], involves two collaborative models: a super-resolution model that enhances I_{LR} to I_{SR}, and a degradation model that downsamples I_{SR} back to I'_{LR}. Additionally, the degradation model downsamples high-resolution face images to I_{HLR} which are then restored to I'_{HR} by the FSR model. The purpose of the cycle consistency loss is to ensure that the generated low-resolution image remains consistent with the original input image,

$$\mathcal{L}(I_{LR}, I'_{LR}, I_{HR}, I'_{HR}) = \|I_{LR} - I'_{LR}\|_2 + \|I_{HR} - I'_{HR}\|_2. \tag{13}$$

Prior Loss: Apart from the above loss functions, some prior knowledge can also be introduced into FSR models to participate in high-quality image reconstruction, such as sparse prior, gradient prior, and edge prior. Among them, gradient prior loss and edge prior loss are the most widely used prior loss functions, which are defined as follows:

$$\mathcal{L}_{TV}(I_{SR}) = \frac{1}{HWC}\sum_{i,j,k}\sqrt{(I_{SR}^{i,j+1,k} - I_y^{i,j,k})^2 + (I_{SR}^{i+1,j,k} - I_y^{i,j,k})^2}, \tag{14}$$

$$\mathcal{L}_{Edge}(I_{SR}, I_y, E) = \frac{1}{HWC}\sum_{i,j,k}\|E(I_{SR}^{i,j+1,k}) - E(I_y^{i,j+1,k})\|_1, \tag{15}$$

where E is the image edge detector, and $E(I_{SR}^{i,j+1,k})$ and $E(I_y^{i,j+1,k})$ are the image edges extracted by the detector. The purpose of the prior loss is to optimize some specific information of the image toward the expected target so that the model can converge faster and the reconstructed image will contain more texture details.

Fourier Space Loss: The design of perceptual losses predominantly focuses on the spatial domain. However, SR is inherently connected to the frequency domain, as downsampling primarily removes high-frequency components. To address this issue, Fuoli *et al.* [24] propose a novel Fourier Space Loss, which emphasizes frequency content by calculating the frequency components using the Fast Fourier Transform (FFT). Firstly, the image is transformed into Fourier space using the FFT. The method then calculates the amplitude difference F_f, $|.|$ and phase difference, $\angle$ of all frequency components between output image and ground truth image. The averaged differences are computed as the total frequency loss as follows:

$$L_f, |.| = \frac{2}{UV} \sum_{u=0}^{U/2-1} \sum_{v=0}^{V-1} \left| |\hat{Y}|_{u,v} - |Y|_{u,v} \right|, \tag{16}$$

$$L_f, \angle = \frac{2}{UV} \sum_{u=0}^{U/2-1} \sum_{v=0}^{V-1} \left| \angle\hat{Y}_{u,v} - \angle Y_{u,v} \right|, \tag{17}$$

$$L_f = \frac{1}{2} L_f, |.| + \frac{1}{2} L_f, \angle. \tag{18}$$

3 FSR Technologies and Methods

FSR typically involves two primary steps:

1. Preprocessing low-resolution images.
2. Generating high-resolution images through predictive models.

Various deep learning-based FSR methods have been developed, leveraging different types of prior information extracted from face images or high-quality face references to improve the reconstruction process. Recently, new generative models and priors have been introduced to advance FSR techniques. We categorizes FSR Methods into five types based on how they address face mapping:

1. General FSR: These methods primarily rely on network architectures such as CNNs, GANs, and Transformers to learn HR image reconstruction in a data-driven manner, often assuming fixed or implicitly learned blur kernels without explicit prior information or kernel estimation.

2. Prior-guided FSR: These methods enhance reconstruction by incorporating prior information such as facial structure or edge details. Compared to General FSR, Prior-Guided FSR excels in precision, particularly in restoring fine facial details and improving image fidelity, making it more suitable for tasks requiring detailed facial restoration.

3. Reference FSR: These methods leverage structural, semantic, or identity priors in combination with external reference images (e.g., different angles of facial images or face dictionaries) to improve reconstruction accuracy. However, challenges like misalignment between reference and target images due to variations in viewpoint, lighting, or expressions, and the availability of high-quality reference data, can limit its practicality.

4. Multi-task FSR: These methods combine FSR with other tasks such as face recognition or low-light enhancement, using joint learning to share information across tasks. While this enhances performance, it requires significantly more computational resources due to the complexity of managing multiple objectives, which poses challenges in resource-constrained environments.

5. Blind FSR: Unlike standard FSR methods, blind FSR simultaneously estimates the unknown blur kernel and reconstructs the high-resolution image, making it well-suited for handling more complex real-world conditions. However, balancing the accuracy of kernel estimation and image reconstruction can be challenging, potentially leading to suboptimal results if one aspect dominates the process.

3.1 General FSR

In General FSR, without utilizing face characteristics, a tailored network is designed to optimize the potential of exploring effective network structures. In the early stages of deep learning development, the initial methods employed CNN networks, while subsequent advancements utilized various sophisticated architectures (such as projection networks, residual networks, channel attention, etc.) to enhance the network's fitting ability. Subsequently, various FSR methods using advanced network structures were proposed. As shown in Figure 2, we categorize general FSR into four types: CNN-based methods, GAN-based methods, transformer-based methods, and mixed-network methods.

3.1.1 CNN-based Methods

Inspired by the pioneering work in applying deep learning to single-image super-resolution, BCCNN [149] became the first method to apply CNNs to FSR tasks, they directly extract facial representation information through a dual-channel

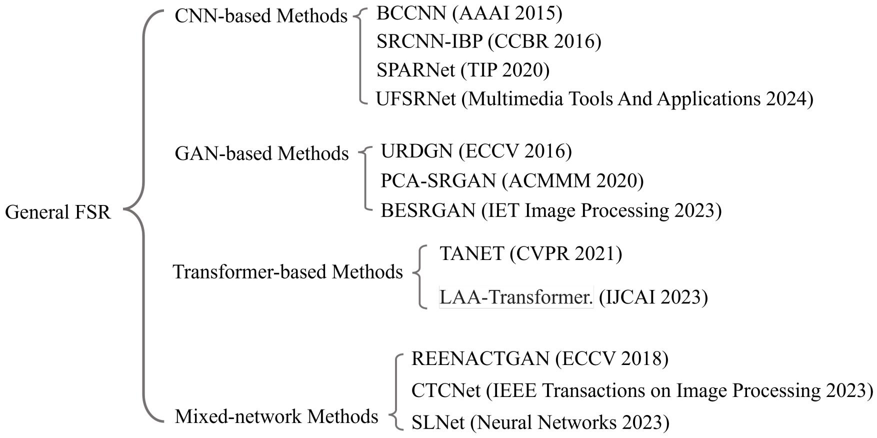
General FSR
CNN-based Methods
BCCNN (AAAI 2015)
SRCNN-IBP (CCBR 2016)
SPARNet (TIP 2020)
UFSRNet (Multimedia Tools And Applications 2024)
GAN-based Methods
URDGN (ECCV 2016)
PCA-SRGAN (ACMMM 2020)
BESRGAN (IET Image Processing 2023)
Transformer-based Methods
TANET (CVPR 2021)
LAA-Transformer. (IJCAI 2023)
Mixed-network Methods
REENACTGAN (ECCV 2018)
CTCNet (IEEE Transactions on Image Processing 2023)
SLNet (Neural Networks 2023)

Figure 2: The typical Methods of General FSR

CNN. Subsequently, with the enhancement of Iterative back projection (IBP) in general image super-resolution performance, Huang *et al.* [43] introduced IBP as an independent post-processing module in the SRCNN-IBP task. Following this, the introduction of channel attention and spatial attention garnered widespread attention. Attention mechanism-based methods, such as those in [78, 9], further improved objective metrics. Representative methods include E-ComSupResNet [17], which integrates attention channel mechanisms, and SPARNet [54], which focuses on spatial attention. In addition to these methods, [74, 88] design cascaded models and utilize multi-scale information to enhance performance.

Global methods can capture global information but fail to recover face details well. To address this, methods based on local recovery of face images have emerged. SRDSI [40] transforms spatial domain face images to the frequency domain, using VDSR [56] and sparse representation to recover low and high-frequency information respectively, and finally fuses the frequency domain information to obtain high-resolution face images. There are also patch-based FSR methods, such as those in [60, 23]. Following the above works, methods considering global-local approaches have been proposed to capture global structures and recover local details simultaneously. Methods like in [107, 81] designed a global upsampling network to simulate global constraints and a local enhancement network to learn face detail features. To capture global cues and restore local details, DPDFN [50] constructs two separate branches to learn global face contours and local face component details, then fuses the results from both branches to generate the final SR results. Wang *et al.* [113] proposes a U-shaped face network based on wavelet transform. First, the downsampling unit uses two depth-separable convolution blocks as the main branch, extracting features through a feature calibration branch and a residual branch. Then, it utilizes Discrete Wavelet Transform (DWT) and Inverse Discrete Wavelet Transform (IDWT) to extract high-frequency details.

3.1.2 GAN-based Methods

GAN was proposed by Goodfellow *et al.* [29] in 2014, which can solve the problem of excessive smoothing caused by CNN methods. Due to its ability to generate more detailed and realistic facial images, GAN has gradually become popular in the field of FSR. In the early days, paired data was commonly used to train discriminators and generators. Recently, pre-trained models were mainly used to generate prior auxiliary model training.

URDGN [136] used a discriminator to distinguish between real HR face images and reconstructed images In the early stages. The generator was used to create SR face images to deceive the discriminator and match the distribution of HR face images. Many subsequent works continued this idea, with MLGE [59] improving it by focusing more on the edge regions of face images. Luo and Huang [84], Indradi *et al.* [49], Chen and Tong [13], and Bin *et al.* [4] also employed generative models to face images. PCA-SRGAN [22] does not directly feed the entire face image into the discriminator. Instead, it decomposes the face image into components via PCA and progressively feeds more components of the face image into the discriminator to reduce its learning difficulty. However, SPGAN [143] argued that a single probability value is too fragile to represent the entire image. Instead, the discriminator outputs a discrimination matrix with the same resolution as the input image and employs a supervised pixel-level adversarial loss to recover more realistic face images. BESRGAN [97] is a high-fidelity boundary equilibrium network, which effectively reduces artifacts and distortions. Specifically, they introduced a fidelity ratio to control the adversarial influence of the discriminator on the generator and then used a balanced perceptual discriminator to match the perceptual loss distribution.

3.1.3 Transformer-based Methods

The Transformer [108] believes that traditional recurrent neural network (RNN) (such as LSTM [31], GRU [16]) compute sequentially, limiting their parallelization capabilities. This sequential nature forces each variable t to wait for $t-1'$s result, which restricts efficiency, especially over long spans where information loss can occur. The Transformer addresses these issues by

1. Using an attention mechanism to reduce the distance between any two positions in a sequence to a constant.

2. Eliminating sequential processing constraints, making it highly suitable for parallelization and efficient information integration across sequences.

Wang *et al.* [118] employ a self-attention mechanism to enhance face structure representation. They model global and local features separately, enhancing both global face structure consistency and local face detail fidelity. Li *et al.* [66] proposed a novel self-refinement mechanism based on the Transformer, Their approach adaptively reconstructs coarse-to-fine texture perception using a wavelet fusion module that integrates shallow structural and deep detailed features in the frequency domain.

3.1.4 Mixed-network Methods

Despite the significant advancements made with GAN, issues like mode collapse and training instability persist. Researchers have continuously introduced various techniques, such as loss functions and training methodologies. Since the introduction of deep convolutional generative adversarial networks (DC-GAN) [95] in 2015, which extended GANs with CNN architectures, successful GAN models have relied on CNN-based generators and discriminators. Traditional GAN-based methods rely on convolutional neural networks have limited receptive fields, leading to a loss of details at deeper levels. Wu *et al.* [126] introduced a novel approach where they map the initial face to a boundary latent space instead of pixel space to avoid structural artifacts. They then use a transformer to adapt this boundary to the target boundary and finally reconstruct face features using a decoder based on target features.

While CNN-based FSR methods have achieved good results, they have limitations. Multi-task joint learning requires additional dataset labeling, and prior networks significantly increase computational costs. Additionally, the limited receptive fields of CNN reduce the fidelity and naturalness of reconstructed face images, leading to poor reconstruction results. Local methods (CNN-based methods) focus primarily on local face details, whereas global methods (Transformer-based methods) typically capture global face structures. FaceFormer [119] combines the global facial information modeling capability of Transformers, which excel at capturing long-range visual dependencies, with the local modeling ability of CNNs to restore fine-grained facial details. CTCNet [25] use a multi-scale encoder-decoder architecture as the backbone network and designe a Global-Local Feature Cooperation Module, including face structure attention units and Transformer modules, to enhance consistency in restoring both local face details and global face structures. Zeng *et al.* [140] employ a hierarchical feature learning framework to obtain shallow information from lower spatial layers. They then refine this shallow information, which had accumulate errors due to deep convolutional networks, resulting in intermediate reconstruction results. Finally, advanced spatial feature representation was improved through a multi-scale context-aware encoder-decoder for face reconstruction.

Discussion: We discuss the advantages and disadvantages of these subcategories within general FSR methods. Overall, the distinction between CNN-based and GAN-based methods lies in their adversarial training approach. CNN-based methods typically use pixel-wise loss, achieving higher PSNR values but often resulting in smoother images. GAN-based methods, on the other hand, may recover more detailed features but tend to have lower PSNR, while the reconstructed face images appear visually pleasing. Transformer-based methods excel in capturing global details due to their self-attention mechanisms, but they require substantial computational resources and memory. Additionally, they are often weaker in extracting pixel-level features of the images. Hybrid methods that combine elements of both approaches can often leverage the strengths of each. For example, networks incorporating both CNN and Transformer components can achieve excellent results in terms of both PSNR and LPIPS.

3.2 Prior-guided FSR

Prior-guided FSR methods typically utilize facial features, such as edge details, gradient changes, landmarks, parsing maps, and heatmaps, to constrain model training. Based on how prior information is used in FSR methods, prior-guided FSR can be categorized into four parts: structural prior preservation, semantic prior preservation, identity prior preservation, and multi-prior preservation FSR methods.

3.2.1 Structure prior preserving FSR

Structure-preserving FSR methods initially utilize face structure information, such as edges, gradients, face heatmaps, face landmarks, face parsing maps, and mixed priors, to guide the reconstruction and training processes. Some of these methods extract prior information from LR face images using a pre-trained model or a sub-network related to the main network.

Edge: Face edge information typically represents the contours of the face, eyes, nose, mouth, and other prominent features, providing additional auxiliary information to help reconstruct corresponding details and clarity. Ko *et al.* [59] observed the edge information at different scales of the face, enhancing edge information to reconstruct high-resolution face images. Yu *et al.* [137] aggregated edge information and attention by parallel connecting channel-wise and spatial-wise components. They integrated attention fusion strategies into the residual module and used edge blocks to extract edge information, allowing adaptive interpolation at multiple scales in the reconstruction part. Shahbakhsh *et al.* [101] proposed an edge-attention architecture that targets edge textures. By reducing the differences between the generated image and the actual image features and enhancing the edges of low-resolution images using Unsharp Masking (UM) and Local Binary Pattern (LBP).

Gradient: In FSR tasks, gradients provide crucial information about image edges and textures. For example, they can help identify edges where pixel values change sharply, aiding in the restoration of fine details like skin textures and hair strands. Luo *et al.* [83] designed an FSR network based on gradient information compensation, consisting of residual blocks and feature extraction blocks. Specifically, it constructs pixel-level gradient images directly from feature maps without requiring additional data labels, compensating for missing high-frequency components in face features.

Face Heatmaps: Face heatmaps represents the importance or attention levels of different areas in a face image, typically using color coding to indicate the significance of each region. Yu *et al.* [135] proposed a method combining multi-task CNNs with face heatmaps to address the impact of local information on mapping low-resolution to high-resolution face images. The CNN reconstructs LQ images and predicts salient regions with face heatmaps, guiding the upsampling process to generate high-quality details. They also utilize both low-level intensity similarity and mid-level face structure information to explore spatial constraints in LR input images. Xiu *et al.* [128] propose the double discriminative FSR network (DDFSRNet) that enhances the reconstruction of key facial components through the perceptual similarity loss, facial heatmap loss, and dual adversarial loss.

Face Landmarks: Face landmarks represent key points on a face image, typically used to locate and describe face features, including key positions like the eyes (inner and outer corners, pupil), nose (tip, wings), etc. Dogan *et al.* [21] guided subsequent training with another unconstrained HR face image of the same person. Due to variations in age, expression, pose, and size, they used adversarial training. Kim *et al.* [55] progressively restored face details, proposing a face alignment network for landmarks extraction. They introduced facial attention loss to enhance facial attributes by combining pixel differences with heatmap values and compress the face alignment network (FAN) for efficient landmark heatmap extraction.

Face Parsing Maps: Face parsing maps segment a face image into different semantic regions, assigning each pixel to a category representing different face parts or features, providing more granular structural information beyond simple face detection or keypoint localization. Wang *et al.* [109] first used an attention module to extract parsing map priors from LR face images. Given that high-resolution features contain more precise spatial information while low-resolution features provide robust contextual information, they designed a multi-scale refinement block to maintain spatial and contextual information, refining feature representations using multi-scale features. Liu *et al.* [75] considered the highly structured nature of faces, utilizing parsing maps to fully leverage LR image information. They fused parsing maps and network features at different dimensions, using parsing maps as masks to assign different weights and loss functions to key face regions.

Mixed Prior: By combining different face priors, FSRNet [12] integrated geometric priors from face landmark heatmaps and parsing maps. Using GAN to reconstruct high-quality SR face images without requiring precise alignment. Hu *et al.* [41] combined 3D face priors to capture sharp face structures explicitly. By incorporating face attribute parameters into 3D deformation knowledge, this method leverages face structure and identity information to effectively handle images with large pose variations. Specifically, it includes a 3D face rendering branch to obtain 3D prior information. CHNet [82] fuses HR facial components with LR background to generate new LR images using facial parsing maps, learning the LR to LRmix mapping to ensure LRmix can be obtained from LR images in tests and real-world datasets. HFNet [141] fused face texture and structure information through an implicit learning dual-branch network with four key components: a deep feature extractor, two interaction modules, and a supervised attention-based fusion network. The extractor uses two-stage cross-dimension attention for texture enhancement and structure reconstruction. HFNet employs information exchange blocks for feature fusion and adaptively aggregates the enhanced maps.

3.2.2 Semantic prior preserving FSR

Semantic prior preserving FSR leverage semantic information to maintain the structure, features, and contextual meaning of face images, ensuring that the super-resolved images are more accurate, realistic, and consistent with the original image semantics. These semantic cues often include face contours, eyes, and attributes such as age, gender, and expressions.

Estimated Attribute Methods: Estimated attribute methods utilize estimated face attributes or features, such as the positions of facial components or other semantic information, to guide FSR training process. CSPGAN [76] utilized semantic probability maps of facial components are utilized to adjust features in the CSPGAN via affine transformations. To address the overly smooth performance of the generative network, a gradient loss is introduced to recover high-frequency details. Li *et al.* [67] proposed an end-to-end gradient enhancement branch and semantic guidance mechanism. The gradient enhancement branch reconstructs high-resolution gradient maps under the constraints of two proposed gradient losses. Then, combining features in both image and gradient spaces, they achieve super-resolution of facial images while preserving geometric structure. Additionally, the proposed semantic guidance mechanism adaptively reconstructs sharp edges and enhances local details in different facial regions under the guidance of semantic parsing maps. Due to the absence of LR face attribute information and errors in training data, the key facial attributes (such as age and gender) of the restored face may differ from the initial LR face images. DebiasFR [71] explicitly models facial attributes, allowing adjustments of facial attributes in the output HR facial images.

Given Attribute Methods: Given attribute methods utilize provided attribute information to assist in face images reconstruction. These attributes can encompass various facial features such as the spatial positions of facial components, facial expressions, lighting conditions, etc. For instance, some methods may employ the spatial positions of facial components adjust features, thereby enhancing the preservation of facial details and structure. Li *et al.* [68] proposed an open-source face SR framework based on facial semantic attribute transformation and self-attention structure enhancement. This framework introduces facial semantic information (i.e., face attributes) and facial structure information (i.e., facial boundaries) in two consecutive stages. In the first stage, face attributes are combined with facial features to generate intermediate HR results with plausible attributes. In the second stage, face boundary heatmaps are estimated from the input, and these are then fused to produce the final HR face image.

3.2.3 Identity prior preserving FSR

FSR methods that preserve identity priors focus on utilizing specific face identity information to enhance the restoration process, ensuring that crucial individual features in face images are retained. This approach is essential for maintaining identity-related facial attributes.

Face Recognition-based Methods: Using face recognition technology to preserve facial identity features and improve the accuracy and quality of FSR. Chen *et al.* [10] effectively learn identity-aware features by decomposing them into two orthogonal components: magnitude and angle. They project identity features into hyperspherical space, where magnitude and angle respectively represent feature quality and identity information. By decomposing these features, the model focuses more on restoring textures related to identity.

Pairwise Data-based Methods: Many existing FSR methods primarily focus on generating pleasing texture details while overlooking the important high-level face attribute of identity. Addressing this gap, DIDNet [14] proposed a dual-loop network framework that utilizes identity information constraints. DIDNet consists of two closed-loop networks: one for generating HR images to preserve identity in the HR feature space, and another for learning the degradation process to leverage low-resolution identity information. By combining dual identity constraints, the method renders face images with distinct features.

3.2.4 Multiple-prior preserving FSR

In response to the challenges posed by deep neural networks (DNNs) that integrate face priors requiring additional labels, longer training times, and larger computational resources, EIPNet [57] proposed a lightweight network using edge blocks to extract and integrate perceptual edge information with feature maps across scales, progressively enhancing local and global structural details. An identity loss function preserves identity information by comparing feature distributions between super-resolved and real images. The Luminance-Chrominance Error (LCE) separates luminance and color components, reducing color reliance and reflecting differences in RGB and YUV color spaces, enabling high-quality 8x super-resolved images. Most traditional methods rely on paired data, which is hard to obtain and do not fully utilize face prior knowledge. To address this, Li *et al.* [67] proposed an end-to-end unsupervised network with a gradient enhancement branch and a semantic guidance mechanism. The gradient enhancement branch reconstructs high-resolution gradient maps under two gradient losses, integrating features in image and gradient spaces to preserve geometric structures. The semantic guidance mechanism includes a semantic adaptive sharpening module and a semantic-guided discriminator, using parsing maps to adaptively reconstruct sharp edges and local details of different face regions.

Disscussion: Prior-guided FSR methods leverage inherent face priors like edge maps, gradient information, face landmarks, heatmaps, and parsing maps to enhance high-resolution face image reconstruction. By integrating these priors into specialized network architectures, these methods effectively preserve face details and attributes, resulting in superior image quality with enhanced structure and texture fidelity. However, they face challenges such as dependency on annotated datasets, increased computational requirements, and difficulties handling variations in face attributes and poses. Future advancements may focus on optimizing computational efficiency and seamlessly integrating multiple priors to improve the performance and applicability of prior-guided FSR methods in real-world scenarios.

3.3 Reference FSR

Most of the aforementioned methods do not consider using HQ face images with the same identity as the LR face images, relying solely on the LR face images as reference information. This limitation prevents the extraction of more detailed face priors. Therefore, reference-based FSR methods use high-quality face images as references to enhance the face restoration process. These reference face images can be single or multiple images, or even dictionary-based guides. When reference images come from different angles and there are multiple images, it allows for a multi-view analysis of face information, referred to as multi-view reference methods. We categorize reference-based FSR methods into four types: single-face guided methods, multi-face guided methods, dictionary-guided methods, and multi-view methods. The comparison of reference FSR methods is shown in Table 2.

Table 2: Comprison of reference FSR methods.

Reference FSR	**Methods**	**Same identity**	**Alignment**
Single-face Guided	GFRNet [69]	✓	Landmark
	GWAInet [21]	✓	Flow field
Multi-face Guided	ASFFNet [77]	✓	Moving least-square
Dictionary-Guided	JSRFC [70]	×	-
Multi-View	Wang *et al.* [115]	✓	Texture
	Wang *et al.* [116]	✓	Texture

Single-face guided methods: When only a single face serves as the reference image, it typically corresponds to a high-quality image with the same identity as the LR face image, often a frontal view. Examples include GFRNet [69] and GWAInet [21]. GFRNet takes degraded observations and high-quality guidance images of the same identity as inputs. It employs a warping sub-network (WarpNet) and a reconstruction sub-network (RecNet) to predict the flow field and warp the guided image to correct pose and expression. The degraded observation and warped guidance image are then used to produce the restoration results. GWAInet utilizes another unconstrained HR face images of the same person to guide the process of applying 8x super-resolution to face images. It is trained in an adversarial manner, using a warp sub-network to align the content of the guiding image to the input image and a feature fusion chain to merge features extracted from the warped guiding image and the input image.

Multi-face guided methods: When a LR face image has only one HQ reference face image available, single-face guided methods typically achieve satisfactory results. However, in some applications, utilizing multiple high-quality face images can provide additional supplementary information. ASFFNet [77] is a novel data-driven pyramidal feature fusion strategy. It suppresses inconsistencies by learning how to spatially filter conflicting information, thereby improving the scale-invariance of features, with almost no additional inference overhead. Therefore, ASFFNet addresses these challenges using weighted least squares alignment [100] and AdaIN [47]. Finally, they designed an Adaptive Feature Fusion Block to generate an attention mask for supplementing LR face images information with that of the reference images.

Dictionary-guided methods: Dictionary-guided FSR Methods utilizes a collection of reference images without requiring identity consistency. These methods create a dictionary of face components from various reference images, selecting those with similar components to the LR face. The selected components are aligned and extracted to form a comprehensive dictionary that guides the face restoration process, enhancing the preservation of facial details and structures. JSRFC Based on observations that different individuals may share similar face components, dictionary-guided methods have been proposed, including Joint Super-Resolution and face Composite (JSRFC) [70] do not require identity consistency between the reference and LR face images. Instead, they construct a component dictionary to enhance face restoration. For example, JSRFC selects reference images with similar components to the LR face image (each reference face image is annotated with vectors indicating which components are similar). Then, the LR face image is aligned with the reference face images, and corresponding components are extracted to form the component dictionary.

Multi-view methods: While single-view FSR methods have shown promising performance, extending these methods to handle multi-view FSR presents greater challenges due to variations in face pose and the need to integrate texture information from multiple low-resolution viewpoints. Wang *et al.* [115] utilized multi-view texture compensation and leveraged texture attention mechanisms to transfer high-precision texture compensation information from multiple viewpoints to a fixed viewpoint. Wang *et al.* [116] extract rich texture information from different viewpoints, which can serve as effective priors for reconstructing frontal face images. They focused on extracting more texture information from multi-view face images and propagated high-precision texture compensation information to frontal face images.

Disscussion: Single-face and multi-face guided FSR rely on HQ reference face images with the same identity as the LR face image. However, these methods are limited by the availability of such reference images. Furthermore, aligning the LR face image with the HQ reference face image remains a challenge even when such references exist. In contrast, dictionary-guided methods expand their applicability by breaking the identity constraint, but at the cost of increased complexity in face reconstruction. Conversely, multi-view methods utilize rich texture information from different viewpoints, leading to superior reconstruction results.

3.4 Multi-task FSR

Multi-task FSR refers to the approach where a single model is designed to simultaneously handle multiple related tasks associated with enhancing the resolution and quality of face images. This methodology is particularly advantageous in scenarios where different aspects of image enhancement need to be addressed concurrently, such as improving resolution, enhancing details, reducing noise, and correcting artifacts. Based on the nature of tasks, as shown in Figure 3, we categorizes multi-task FSR into seven types: Face recognition

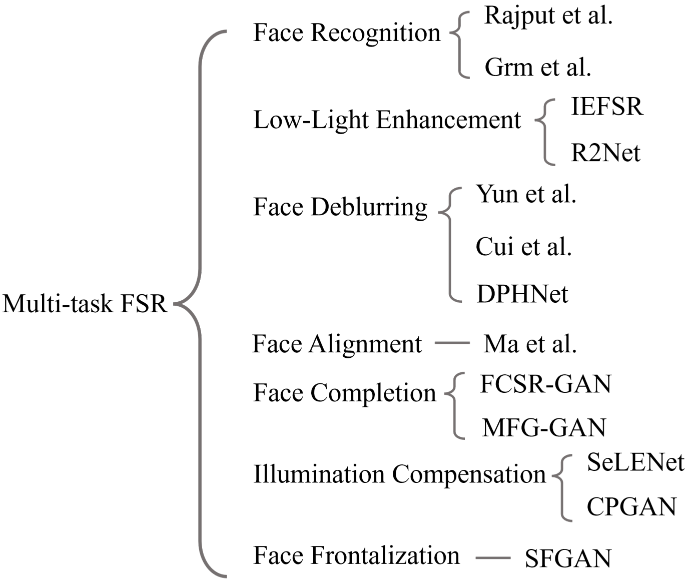

Figure 3: The Classification of Multi-task FSR

and FSR, Low-light enhancement and FSR, Face deblurring and FSR, Face alignment and FSR, Face Completion and FSR, Illumination Compensation and FSR, and Face Frontalization and FSR.

3.4.1 Face Recognition and FSR

Face Recognition focuses on enhancing the resolution and quality of face images to assist in accurately identifying individuals in high-resolution images. In real-world surveillance scenarios, Face Recognition (FR) systems face challenges due to captured LR and noisy probe images. To address this, Rajput *et al.* [96] inherited the advantages of function interpolation and dictionary-based SR techniques. Function interpolation aids in generating more discriminative outputs, while dictionary-based methods help in mitigating noise effects during the reconstruction process. Grm *et al.* [32] established identity matching across different sources and proposed an approach combining FSR, resolution matching, and multi-scale template accumulation to reliably identify faces in remote surveillance videos, including images from low-quality sources.

3.4.2 Low-light Enhancement and FSR

To address the challenge of enhancing face images under low-light conditions and produce clearer and more detailed images, IEFSR [36] enabled the restoration of low-light face images of 32 × 32 pixels to high-resolution faces through an 8x magnification process. A coarse LR recovery network reconstructs these faces, revealing hidden details. The generator uses noisy style blocks for visual realism, and spectral normalization in the discriminator enhances training stability. Hai *et al.* [37] proposed R2RNet to address degradation in low-light images using the Retinex theory. It includes three sub-networks: Decom-Net for decomposition, Denoise-Net for denoising, and Relight-Net for contrast enhancement and detail preservation. R2RNet leverages spatial and frequency information to improve contrast and retain details. Wang *et al.* [117] propose a novel duplex fusing-embedding learning approach to tackle low-light environments challenge. In the fusion phase, shallow features from both tasks are bidirectionally fused into a consistent feature space. In the embedding phase, fused features from previous iterations are fed back and embedded into the deep features of both tasks, enhancing the learning of feature representations.

3.4.3 Face Deblurring and FSR

Face deblurring aims to enhance the visual clarity and fidelity of face images degraded by motion blur, out-of-focus blur, or other factors. Yun *et al.* [139] proposed an adversarial framework to reconstruct high perceptual quality and deblurred HR face images. They utilized a simple five-layer CNN to extract LR images' feature maps, which were then fed into two branches of an encoder-decoder network to generate HR face images with and without blur. Cui *et al.* [18] directly obtained the HR and clear face images from LR and blurred face images. By parallelly connecting super-resolution feature extraction branches and deblurring feature extraction branches, they effectively avoided error accumulation. They introduced a hybrid attention mechanism to enhance feature selection in both channel and spatial dimensions. Additionally, a new multi-scale feature fusion module was proposed to effectively integrate features from these two tasks. DPHNet [94] proposed a dual-branch hybrid network where the Transformer branch captured global features, while the CNN branch focused on extracting local features. The convolutional block attention module aggregated features extracted by the two branches, enabled interaction between channel and spatial domain information.

3.4.4 Face Alignment and FSR

Face alignment involves the precise localization and identification of key facial landmarks in an image or video frame, aimed at determining their spatial positions relative to the face. However, challenges arise in accurately aligning face features across different images or poses, particularly under severe occlusions or extreme poses. These challenges stem from two main reasons:

1. Difficulty in modeling long-range dependencies and constructing effective face shape constraints.
2. Limitations in the scale and diversity of annotated face datasets available for training.

To address these, based on Transformer for data distillation, Ma *et al.* [85] incorporated Transformer-based heatmap detection to model more efficient face shape constraint relationships. They designed a quality-aware pseudo-label sample distillation network to assess the quality of pseudo-label data generated by the Transformer heatmap detection network and aids the Transformer in mitigating inherent biases of CNNs.

3.4.5 Face Completion and FSR

Face completion aims to restore or fill in missing parts of a face image to create a complete and visually plausible representation. When combined with FSR techniques, it enhances the resolution of face images while completing missing parts or features, thereby improving the visual integrity and quality of the reconstructed images. FCSR-GAN [6] employed multi-task learning to simultaneously perform face completion and FSR in an end-to-end manner. The generator of FCSR-GAN aims to recover unoccluded HR face images from input LR face images that may contain occlusions. The discriminator in FCSR-GAN utilizes a series of carefully designed loss functions (adversarial loss, perceptual loss, pixel loss, smoothness loss, style loss, and face prior loss) to ensure high-quality reconstruction of the high-resolution face images. MFG-GAN [79] integrates graph convolution and feature pyramids, employing a multi-scale feature map GAN to restore occluded LR face images to unoccluded HR face images.

3.4.6 Illumination Compensation and FSR

The color information of images is often influenced by factors such as light sources and the color biases of capturing devices, leading to overall color shifts like cooler or yellowish tones. To counteract these color deviations and facilitate subsequent image processing, illumination compensation is necessary. SeLENet [63] decomposes face images into face normals and albedo maps, then they used spherical harmonic lighting coefficients of ambient white light to enhance and reconstruct the illumination conditions of the input image, resulting in neutral light face images. CPGAN [146] restores realistic high-resolution face images while compensating for low and uneven illumination by enhancing face details using information from the input image and additional high-resolution face images for illumination compensation.

3.4.7 Face Frontalization and FSR

Face frontalization refers to the process of generating frontal-facing images of faces from non-frontal or multi-view images, which finds extensive applications in face recognition, video surveillance, and identity verification. Ning *et al.* [89] combined FSR technology with face frontalization methods involves first transforming non-frontal or arbitrary pose face images into frontal views, followed by enhancing the resolution and quality of the images using super-resolution techniques. SF-GAN [134] generates high-resolution frontal faces from LR inputs while preserving identity. It uses intra-class and inter-class constraints: orthogonal loss encourages diverse subject representations, while triplet loss enhances identity preservation. SF-GAN includes an SR side-view module to maintain fine details in side-view faces, improving realism in synthesizing frontal images from non-frontal inputs.

3.4.8 Discussion

Multi-task FSR methods integrate multiple enhancement tasks into a single model, enhancing face image resolution, detail, and quality for applications like face recognition, surveillance, and image analysis. They address tasks such as deblurring, alignment, and illumination compensation, offering versatile image enhancement capabilities. However, designing and training these methods can be complex due to integrating diverse tasks and managing multiple loss functions. Computational demands and task prioritization trade-offs also impact performance. Recent advances have improved handling of these challenges, promising robust and efficient face image enhancement in practical applications.

3.5 Blind FSR

Blind Face Super-Resolution (BFR) tasks involve enhancing face images' resolution or quality without direct access to corresponding HR images during training. In classic FSR tasks, LR images are assumed to be degraded versions of HR images using a predefined degradation kernel, typically a bicubic downsampling blur kernel. However, real-world degradations are much more complex and often involve various intricate factors in unpredictable combinations, making their exact formulation unknown. This complexity leads to domain gaps between training samples with bicubic downsampling and actual images, resulting in poor performance when deploying networks trained solely on bicubic kernels in practical applications. BFR tasks hold significant potential in diverse domains such as digital arts, computer graphics, social media, and mobile applications. We categorized BFR into two types based on the use of priors: none-prior and prior. None-prior methods do not rely on explicit prior knowledge about the degradation process or HR images during training, relying instead on learning directly from LR images. In contrast, prior-based methods incorporate prior knowledge about potential degradation types or high-resolution image statistics to guide the restoration process, leveraging external information to enhance performance in challenging scenarios. Table 3 shows the typical BFR methods.

Table 3: Summary of BFR

BFR	**Methods**	**Prior**	**Network**
None-Prior BFR	HiFaceGAN [131]	-	U-Net
	STUNet [144]	-	U-Net, Transformer
	GCFSR [38]	-	Transformer
	DAEFR [106]	-	Transformer
	BFRFormer [28]	-	Transformer
	Meta-USR [42]	-	Meta Model
	MRDA [127]	-	Meta Model
Prior BFR	RWSR [1]	LR/HR prior	GAN
	Goswami *et al.* [30]	Gererative prior	GAN
	zheng *et al.* [148]	LR/HR prior	GAN
	CodeFormer [150]	Codebook combination prior	Transformer
	SCGAN [39]	Gererative prior	Cycle-GAN
	RSenFace [124]	Semantic prior	-
	Difface [138]	Gererative prior	Diffusion Model
	BPSR3 [26]	Gererative prior	Diffusion Model
	BFRffusion [11]	Generative prior	Diffusion Model
	PGDiff [132]	Generative prior	Diffusion Model
	DR2 [121]	Generative prior	Diffusion Model
	DiffMAC [27]	Generative prior	Diffusion Model

3.5.1 None-Prior BFR

Current face restoration research typically relies on image degradation priors or explicit guiding labels for training, limiting their ability to handle heterogeneous degradations and complex background content in real-world scenarios. HiFaceGAN [131] addresses a more challenging and practical “double-blind” problem, known as Face Renovation (FR), which eliminates the need for both of these priors. HiFaceGAN’s multi-stage framework includes multiple nested CSR units that progressively replenish face details using hierarchical semantic guidance extracted from a frontend content-adaptive suppression module.

STUNet [144] designs a scalable and transferable series of U-Net [98] models for blind face restoration. Leveraging attention mechanisms and a shifted windowing scheme, STUNet captures interactions among distant pixels to focus more on critical features while ensuring efficient training. GCFSR [38] proposes a generative and controllable FSR framework designed for multi-factor super-resolution tasks using an encoder-generator architecture. It incorporates two modules: style modulation aims to generate realistic face details, while feature modulation dynamically blends multi-layer encoded features and generated features based on the magnification factor.

DAEFR [106] minimize the domain gap and information loss when restoring HQ images from LQ ones by promoting effective collaboration between the LQ and HQ branches through joint training, thereby improving code prediction and recovery quality. BFRFormer [28] introduce wavelet discriminators and aggregate attention modules to remove blocking artifacts, adaptively normalize spectra, and balance consistency to mitigate training instability and overfitting inherent in CNN-based methods. To quickly learn new tasks and overcome inconsistencies between training and testing face image scenes, meta-learning methods are trained to construct adaptive models that adjust parameters at test time based on input image characteristics, enabling the model to acquire a "learning ability.". Hu *et al.* [42] design a meta-restoration module to handle various degradation factors, promising practical application prospects. Xia *et al.* [127] use meta-learning to generalize extensive external data, rapidly adapt to specific complex degradations, and extract implicit degradation information.

3.5.2 Prior BFR

BFR methods typically rely on priors or assumptions about image degradation processes during training. These priors include models for blur kernels, noise characteristics, compression artifacts, and other forms of degradation that occur in real-world scenarios. By incorporating these priors into the training process, the models aim to enhance the resolution of LQ to HQ ones without explicit guidance from HR counterparts during training.

RWSR [1] proposes an LR/HR training-to-generation framework. Initially, it estimates parameters such as blur kernel, noise, compression, etc., from real LR faces. Subsequently, it generates LR and HR face image pairs with these estimated parameters for GAN training. To achieve better perceptual quality, they interchange commonly used losses like VGG-Loss [51] and LPIPS-Loss [145] to attain more detailed reconstructions with lower noise. Goswami *et al.* [30] employ GANs to generate synthetically degraded LR images paired with their corresponding HR counterparts, training with a combination of pixel-level and adversarial losses. Finally, they enforce the similarity between encoded features learned from clean and degraded images using Entropy

Regularized Wasserstein Divergence to enhance model robustness. Zheng *et al.* [148] employed semi-dual optimal transport in their research to steer the learning process, resulting in the development of a semi-dual optimal transport CycleGAN. They address the discrepancy between generated LR faces during training and actual LR face images during testing. To mitigate this, researchers introduced characteristic regularization [15]. Built upon CycleGAN, CR facilitates the learning of mappings between real LR face images and synthesized LR face images.

CodeFormer [150] model the global composition and context of LQ faces using the codebook combination prior for latent code prediction. Additionally, they improve adaptability to various degradations and introduce a controllable feature transformation module, allowing for flexible trade-offs between fidelity and quality. SCGAN [39] established two independent degradation branches in the forward and backward cycle consistency reconstruction processes. Both processes share a common restoration branch. They mitigated the domain gap between real-world LR face images and generated LR face images and achieves accurate and robust FSR performance by regularizing the shared restoration branch through forward and backward cycle consistency learning processes. RSemFace [124] first design a degradation stage to synthesize low-resolution face images degraded by various interpolations, noise levels, blur kernels, and even real-world interferences. In the generation stage, it generates coarse super-resolution face images and extracts semantic features as prior information, which support the Semantic Feature Attention Blocks and fine super-resolution face image reconstruction under semantic loss.

In order to more effectively simulate complex degradation processes and reduce the frequency of adjusting fidelity, perceptual loss, and other hyper-parameters. Therefore, Yue *et al.* [138] utilize a diffusion model to improve the ability to restore face shapes and details by establishing the posterior distribution from low-quality to high-quality images. Gao *et al.* [26] design a multi-scale deep back-projection network based on diffusion models to enhance the quality of recovered images at different scales. Chen *et al.* [11] use pre-trained stable diffusion-generated priors to guide training of self-attention networks. To enhance the realism of reconstruction results and the generalization ability of diffusion models, Yang *et al.* [132] individually model degradation attributes to guide the reverse diffusion process. DR2 [121] first converts degraded images into rough but degradation-invariant predictions, then uses an enhancement module to restore these rough predictions to high-quality images. Gao *et al.* [27] propose a diffusion-information-diffusion framework that highly generalizes face features across different degradation scenarios and heterogeneous domains.

3.5.3 Discussion

BFR reconstructs low-quality face images with unknown degradation. In real-world scenarios, degradation is complex and can't be precisely modeled by a single function. Diffusion model-based methods simulate this process by sequentially adding blur, noise, and other factors, effectively handling various degradations to produce clearer images, but they are slow and have weak generalization for non-image or structured data. Meta-learning methods, on the other hand, adapt quickly to new tasks with high generalization but are heavily dependent on task distribution and data quality. BFR holds significant practical importance in real-world applications.

4 Current Issues and Future Work

First, we categorize methods based on deep learning into five types according to how face mappings are solved: General FSR, Prior-guided FSR, Reference FSR, Multi-task FSR, and Blind FSR. Then, further subdivisions are made based on various network architectures or whether face priors are used. Of course, researchers' exploration of FSR methods extends beyond what we have introduced. To gain a comprehensive understanding of this rapidly evolving field is quite challenging, and omissions may occur. Therefore, this review serves as an educational tool, providing insights into deep learning-based FSR methods for researchers.

Despite significant breakthroughs in FSR problems with the development of deep learning technologies, FSR still faces many challenges and will become more numerous with new issues and field explorations. Below, we illustrate current FSR issues and future trends.

4.1 Design of Network

The backbone network has a crucial impact on performance, especially when designing an efficient training network that can effectively improve objective evaluation metrics such as PSNR and SSIM. Therefore, we can draw inspiration from novel network designs such as DiTs [92], RWKV [93] and Mamba [34]. We can also refer to general image super-resolution tasks, where many well-designed network structures have been proposed, like IPG [105] and SinSR [120], to design efficient deep networks more suitable for FSR tasks.

In addition to network efficacy, corresponding hardware has seen rapid development in recent years, enabling the deployment of large models in practical applications. Using large models can better adapt to specific image characteristics in certain scenarios, such as improving the demand for high-precision face images in nighttime surveillance or low-light conditions. This is because large models possess powerful feature extraction and representation capabilities, enabling them to fit a wider range of data distributions and enhance generalization performance.

4.1.1 Lightweight FSR for Edge Devices

Implementing lightweight FSR on edge devices requires a combination of model optimization techniques and efficient architectures. Techniques such as quantization, pruning, and knowledge distillation significantly reduce computational demands and storage requirements. Leveraging lightweight convolutional neural networks enhances performance while maintaining efficiency. Additionally, utilizing attention mechanisms to focus on key facial regions helps improve the effectiveness of super-resolution without compromising efficiency. For instance, integrating FSR frameworks and deploying them on edge devices [110] can achieve high-quality face image super-resolution while meeting the resource constraints of edge computing.

4.1.2 Scale-Arbitrary FSR

Currently, arbitrary-scale FSR technology faces challenges such as strong data dependency, high model complexity, and insufficient real-time processing capability. Training high-quality models requires diverse datasets covering different ages, genders, and ethnicities, which increases model complexity and computational resource demands for practical applications [103]. Despite these challenges, the future is promising. With algorithm optimization, improved computational efficiency, and richer datasets, arbitrary-scale FSR technology is expected to play a more significant role.

4.1.3 Exploitation of face Prior

Utilizing face prior knowledge in FSR encounters challenges in data acquisition and processing, particularly requiring large-scale and diverse training data that encompass variations in age, race, and facial expressions. Additionally, the complexity of algorithms and the demand for computational resources are critical factors limiting the practical applications of FSR technology. However, advancements in technology and research present opportunities to address these challenges through algorithm optimization, improved computational efficiency, and the availability of more diverse datasets. For instance, the work of Zhao *et al.* [147] proposes a hyper-Laplacian prior-based SR method, which could be applied to FSR. Specifically, their model consists of three components: a rough reconstruction subnetwork(RRS), a hyper-Laplacian prior subnetwork(HPS), and an image refinement enhancement subnetwork(RES). In the RRS, LR images are reconstructed into rough SR images using a set of residual blocks. In the HPS, the hyper-Laplacian prior is introduced for LR images to provide additional texture information, followed by the implementation of a prior loss that imposes second-order supervision on the SR image. Finally, the outputs of the RRS and HPS are fused and fed into the RES for HQ image reconstruction. These approaches offer new directions for FSR, and with further optimization and research, they are expected to enhance the performance and expand the application potential of FSR technology.

4.2 Learning Strategies

The learning strategy of FSR includes the selection and optimization of deep learning models, effective data preprocessing, design and balancing of loss functions, integration of multiple technologies, and consideration of real-time processing. These strategies work together to improve the accurate reconstruction of face details and quality in the model.

4.2.1 Loss Functions

Pixel-wise L1 or L2 loss tends to produce super-resolution results with high PSNR and SSIM values, while perceptual loss and adversarial loss tend to generate visually pleasing results, performing well in terms of LPIPS and FID. There is no universal evaluation metric that can perfectly balance these aspects. For instance, Cycle Consistency Loss proposed by Zhu *et al.* [151] maintains network training through loss coordination across paired images, while prior loss aims to control the influence of priors on the network.

4.2.2 Unsupervised, few-shot and Meta learning FSR

Currently, unsupervised, few-shot, and meta-learning face significant challenges in FSR. Data scarcity and variability in quality limit the generalization ability in few-shot learning, while meta-learning requires a diverse set of meta-tasks for effective training. Moreover, the complexity of model structures and the demand for substantial computational resources are crucial constraints for practical applications. However, with technological advancements and deeper research, future prospects involve addressing these challenges through algorithm optimization, enhancing dataset diversity, and improving computational efficiency. These efforts will drive FSR technology towards higher performance and broader application domains, opening new possibilities for visual processing technologies.

4.3 Evaluation Metrics

Evaluation metrics are crucial in the field of FSR; however, there currently exists a lack of standardization and consistency among these metrics. Different studies employ diverse and often non-aligned evaluation metrics, which limits the comparability and reproducibility of results. In the future, with deeper research and the establishment of standards, we can expect to see a more comprehensive and objective evaluation metrics system. This system will accurately reflect the performance of super-resolution models in fidelity preservation, detail retention, computational efficiency, and real-world applications. Such advancements will provide a more robust foundation for technology assessment and propel FSR technology towards higher levels of development and widespread application.

4.4 Real-world Scenarios FSR

In real-world scenarios, FSR technology plays a crucial role in enhancing the clarity and details of face images captured under various conditions. However, it faces constraints in processing time for real-time applications such as video conferencing or surveillance, necessitating efficient processing. Furthermore, achieving consistent high-quality results across diverse face features, expressions, and lighting conditions remains a challenge. Despite these obstacles, advancements in deep learning and computational efficiency provide hope for future developments.

4.4.1 Dataset

The datasets for FSR technology currently suffer from issues such as lack of diversity, varying scales, and a lack of standardized evaluation criteria. However, with advancements in data collection and annotation technologies, anticipated future directions include enhancing dataset diversity and scale. This enhancement aims to improve model generalization and adaptability, enabling more precise and reliable face reconstruction in complex scenarios.

4.4.2 Method

FSR methods currently face several significant challenges: algorithm complexity and computational demands, adaptability to diverse face features and complex lighting conditions, and the specificity of evaluation metrics. These challenges impact the transparency and credibility of technological advancements. However, with continuous progress in deep learning algorithms and hardware technologies, optimizing algorithm structures and utilizing emerging hardware accelerators can significantly enhance the computational efficiency and real-time processing capabilities of FSR techniques. Moreover, introducing more sophisticated deep learning models and richer datasets can improve the model's understanding of various face features and environmental conditions. Lastly, establishing unified evaluation standards and more accurate metric systems will aid in comprehensive assessments of technological performance and application effectiveness.

4.5 Mutual Promotion with High-level Tasks

In the mutual promotion of FSR technology with advanced tasks such as face recognition, current challenges include computational complexity when dealing with large-scale data and maintaining consistent high-quality outputs under varying conditions like different lighting and face expressions. Looking forward, the focus will be on advancements in deep learning algorithms and computational capabilities to optimize real-time processing. This advancement aims to provide more accurate and efficient solutions in domains like security monitoring and medical diagnostics, thereby promoting widespread application and further innovation.

4.6 Multi-modal FSR

Multi-modal information, such as depth, texture, and lighting, provides additional contextual cues. Integrating these cues can help models better understand the three-dimensional structure and surface characteristics of faces, such as skin color, pores, and wrinkles [33]. However, most existing methods primarily utilize computer vision techniques, neglecting multi-modal algorithms that involve text, sound, and images. Wang *et al.* [112] propose a pan-sharpening framework, intrinsic decomposition knowledge distillation. The teacher network decomposes HR images, while the student network uses enhanced illumination from LR to reconstruct the image. By combining depth information to guide texture restoration and using lighting information to enhance facial shadows and highlights, better results can be achieved. Subsequently, they propoes PSCINN [111], which, under the guidance of the PAN image, uses resampled latent variables from a prior distribution and low-resolution images to predict the pan-sharpened image in an information-preserving manner.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (62072350, 62171328, 62401410); by Central Government Guides Local Science and Technology Development Special Projects (ZYYD2022000021); by the Hubei Science and technology innovation team (T2023009); by the

National Natural Science Foundation of Hubei (2023AFB158); by Enterprise Technology Innovation Project (2022012202015060); by the Youths Science Foundation of Wuhan institute of technology (XJ2023000103).

References

[1] A. Aakerberg, K. Nasrollahi, and T. B. Moeslund, "Real-world super-resolution of face-images from surveillance cameras", *IET Image Processing*, 16(2), 2022, 442–52.

[2] S. Baker and T. Kanade, "Hallucinating faces", in *Proceedings Fourth IEEE international conference on automatic face and gesture recognition (Cat. No. PR00580)*, IEEE, 2000, 83–8.

[3] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, and N. Kumar, "Localizing parts of faces using a consensus of exemplars", *IEEE transactions on pattern analysis and machine intelligence*, 35(12), 2013, 2930–40.

[4] H. Bin, C. Weihai, W. Xingming, and L. Chun-Liang, "High-quality face image SR using conditional generative adversarial networks", *arXiv preprint arXiv:1707.00737*, 2017.

[5] A. Bulat and G. Tzimiropoulos, "How far are we from solving the 2d & 3d face alignment problem?(and a dataset of 230,000 3d facial landmarks)", in *Proceedings of the IEEE international conference on computer vision*, 2017, 1021–30.

[6] J. Cai, H. Han, S. Shan, and X. Chen, "FCSR-GAN: Joint face completion and super-resolution via multi-task learning", *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2(2), 2019, 109–21.

[7] A. Chakrabarti, A. Rajagopalan, and R. Chellappa, "Super-resolution of face images using kernel PCA-based prior", *IEEE Transactions on Multimedia*, 9(4), 2007, 888–92.

[8] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding", in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.* Vol. 1, IEEE, 2004, I–I.

[9] C. Chen, D. Gong, H. Wang, Z. Li, and K.-Y. K. Wong, "Learning spatial attention for face super-resolution", *IEEE Transactions on Image Processing*, 30, 2020, 1219–31.

[10] J. Chen, J. Chen, Z. Wang, C. Liang, and C.-W. Lin, "Identity-aware face super-resolution for low-resolution face recognition", *IEEE Signal Processing Letters*, 27, 2020, 645–9.

[11] X. Chen, J. Tan, T. Wang, K. Zhang, W. Luo, and X. Cao, "Towards real-world blind face restoration with generative diffusion prior", *IEEE Transactions on Circuits and Systems for Video Technology*, 2024.

[12] 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, 2492–501.

[13] Z. Chen and Y. Tong, “Face super-resolution through wasserstein gans”, *arXiv preprint arXiv:1705.02438*, 2017.

[14] F. Cheng, T. Lu, Y. Wang, and Y. Zhang, “Face super-resolution through dual-identity constraint”, in *2021 IEEE international conference on multimedia and expo (ICME)*, IEEE, 2021, 1–6.

[15] Z. Cheng, X. Zhu, and S. Gong, “Characteristic regularisation for super-resolving face images”, in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, 2435–44.

[16] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning phrase representations using RNN encoder-decoder for statistical machine translation”, *arXiv preprint arXiv:1406.1078*, 2014.

[17] V. Chudasama, K. Nighania, K. Upla, K. Raja, R. Ramachandra, and C. Busch, “E-comsupresnet: Enhanced face super-resolution through compact network”, *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 3(2), 2021, 166–79.

[18] Y. Cui, C. Tang, and Q. Huang, “Joint face super-resolution and deblurring using multi-task feature fusion network”, in *7th International Conference on Vision, Image and Signal Processing (ICVISP 2023)*, Vol. 2023, IET, 2023, 57–61.

[19] N. Di Domenico, G. Borghi, A. Franco, and D. Maltoni, “Combining Identity Features and Artifact Analysis for Differential Morphing Attack Detection”, in *International Conference on Image Analysis and Processing*, Springer, 2023, 100–11.

[20] K. Ding, K. Ma, S. Wang, and E. P. Simoncelli, “Image quality assessment: Unifying structure and texture similarity”, *IEEE transactions on pattern analysis and machine intelligence*, 44(5), 2020, 2567–81.

[21] B. Dogan, S. Gu, and R. Timofte, “Exemplar guided face image super-resolution without facial landmarks”, in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2019, 0–0.

[22] H. Dou, C. Chen, X. Hu, Z. Xuan, Z. Hu, and S. Peng, “PCA-SRGAN: Incremental orthogonal projection discrimination for face super-resolution”, in *Proceedings of the 28th ACM international conference on multimedia*, 2020, 1891–9.

[23] Z. Feng, J. Lai, X. Xie, D. Yang, and L. Mei, “Face hallucination by deep traversal network”, in *2016 23rd International Conference on Pattern Recognition (ICPR)*, IEEE, 2016, 3276–81.

[24] D. Fuoli, L. Van Gool, and R. Timofte, “Fourier space losses for efficient perceptual image super-resolution”, in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, 2360–9.

[25] G. Gao, Z. Xu, J. Li, J. Yang, T. Zeng, and G.-J. Qi, “CTCNet: A CNN-transformer cooperation network for face image super-resolution”, *IEEE Transactions on Image Processing*, 32, 2023, 1978–91.

[26] J. Gao, N. Tang, and D. Zhang, “A Multi-Scale Deep Back-Projection Backbone for Face Super-Resolution with Diffusion Models”, *Applied Sciences*, 13(14), 2023, 8110.

[27] N. Gao, J. Li, H. Huang, Z. Zeng, K. Shang, S. Zhang, and R. He, “DiffMAC: Diffusion Manifold Hallucination Correction for High Generalization Blind Face Restoration”, *arXiv preprint arXiv:2403.10098*, 2024.

[28] G. Ge, Q. Song, G. Zhu, Y. Zhang, J. Chen, M. Xin, M. Tang, and J. Wang, “BFRFormer: Transformer-based generator for Real-World Blind Face Restoration”, *arXiv preprint arXiv:2402.18811*, 2024.

[29] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets”, *Advances in neural information processing systems*, 27, 2014.

[30] S. Goswami, Aakanksha, and A. Rajagopalan, “Robust super-resolution of real faces using smooth features”, in *European Conference on Computer Vision*, Springer, 2020, 169–85.

[31] A. Graves and A. Graves, “Long short-term memory”, *Supervised sequence labelling with recurrent neural networks*, 2012, 37–45.

[32] K. Grm, B. K. Özata, V. Štruc, and H. K. Ekenel, “Meet-in-the-middle: Multi-scale upsampling and matching for cross-resolution face recognition”, in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, 120–9.

[33] A. Gruber, E. Collins, A. Meka, F. Mueller, K. Sarkar, S. Orts-Escolano, L. Prasso, J. Busch, M. Gross, and T. Beeler, “GANtlitz: Ultra High Resolution Generative Model for Multi-Modal Face Textures”, in *Computer Graphics Forum*, Vol. 43, No. 2, Wiley Online Library, 2024, e15039.

[34] A. Gu and T. Dao, “Mamba: Linear-time sequence modeling with selective state spaces”, *arXiv preprint arXiv:2312.00752*, 2023.

[35] B. K. Gunturk, A. U. Batur, Y. Altunbasak, M. H. Hayes, and R. M. Mersereau, “Eigenface-domain super-resolution for face recognition”, *IEEE transactions on image processing*, 12(5), 2003, 597–606.

[36] K. Guo, M. Hu, S. Ren, F. Li, J. Zhang, H. Guo, and X. Kui, “Deep illumination-enhanced face super-resolution network for low-light images”, *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 18(3), 2022, 1–19.

[37] J. Hai, Z. Xuan, R. Yang, Y. Hao, F. Zou, F. Lin, and S. Han, "R2rnet: Low-light image enhancement via real-low to real-normal network", *Journal of Visual Communication and Image Representation*, 90, 2023, 103712.

[38] J. He, W. Shi, K. Chen, L. Fu, and C. Dong, "Gcfsr: a generative and controllable face super resolution method without facial and gan priors", in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, 1889–98.

[39] H. Hou, J. Xu, Y. Hou, X. Hu, B. Wei, and D. Shen, "Semi-cycled generative adversarial networks for real-world face super-resolution", *IEEE Transactions on Image Processing*, 32, 2023, 1184–99.

[40] X. Hu, P. Ma, Z. Mai, S. Peng, Z. Yang, and L. Wang, "Face hallucination from low quality images using definition-scalable inference", *Pattern Recognition*, 94, 2019, 110–21.

[41] X. Hu, W. Ren, J. LaMaster, X. Cao, X. Li, Z. Li, B. Menze, and W. Liu, "Face super-resolution guided by 3d facial priors", in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IV 16*, Springer, 2020, 763–80.

[42] X. Hu, Z. Zhang, C. Shan, Z. Wang, L. Wang, and T. Tan, "Meta-USR: A unified super-resolution network for multiple degradation parameters", *IEEE Transactions on Neural Networks and Learning Systems*, 32(9), 2020, 4151–65.

[43] D. Huang and H. Liu, "Face hallucination using convolutional neural network with iterative back projection", in *Biometric Recognition: 11th Chinese Conference, CCBR 2016, Chengdu, China, October 14-16, 2016, Proceedings 11*, Springer, 2016, 167–75.

[44] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database forstudying face recognition in unconstrained environments", in *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*, 2008.

[45] H. Huang, H. He, X. Fan, and J. Zhang, "Super-resolution of human face image using canonical correlation analysis", *Pattern Recognition*, 43(7), 2010, 2532–43.

[46] H. Huang, R. He, Z. Sun, and T. Tan, "Wavelet-srnet: A wavelet-based cnn for multi-scale face super resolution", in *Proceedings of the IEEE international conference on computer vision*, 2017, 1689–97.

[47] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization", in *Proceedings of the IEEE international conference on computer vision*, 2017, 1501–10.

[48] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment", *Electronics letters*, 44(13), 2008, 800–1.

[49] S. D. Indradi, A. Arifianto, and K. N. Ramadhani, “Face image super-resolution using inception residual network and gan framework”, in *2019 7th International Conference on Information and Communication Technology (ICoICT)*, IEEE, 2019, 1–6.

[50] K. Jiang, Z. Wang, P. Yi, T. Lu, J. Jiang, and Z. Xiong, “Dual-path deep fusion network for face image hallucination”, *IEEE Transactions on Neural Networks and Learning Systems*, 33(1), 2020, 378–91.

[51] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution”, in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, Springer, 2016, 694–711.

[52] R. Kalarot, T. Li, and F. Porikli, “Component attention guided face super-resolution network: Cagface”, in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2020, 370–80.

[53] 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*, 2019, 4401–10.

[54] M. A. Khoei, A. Yousefzadeh, A. Pourtaherian, O. Moreira, and J. Tapson, “Sparnet: Sparse asynchronous neural network execution for energy efficient inference”, in *2020 2nd IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS)*, IEEE, 2020, 256–60.

[55] D. Kim, M. Kim, G. Kwon, and D.-S. Kim, “Progressive face super-resolution via attention to facial landmark”, *arXiv preprint arXiv:1908.08239*, 2019.

[56] J. Kim, J. K. Lee, and K. M. Lee, “Accurate image super-resolution using very deep convolutional networks”, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, 1646–54.

[57] J. Kim, G. Li, I. Yun, C. Jung, and J. Kim, “Edge and identity preserving network for face super-resolution”, *Neurocomputing*, 446, 2021, 11–22.

[58] K. I. Kim and Y. Kwon, “Single-image super-resolution using sparse regression and natural image prior”, *IEEE transactions on pattern analysis and machine intelligence*, 32(6), 2010, 1127–33.

[59] S. Ko and B.-R. Dai, “Multi-laplacian GAN with edge enhancement for face super resolution”, in *2020 25th International Conference on Pattern Recognition (ICPR)*, IEEE, 2021, 3505–12.

[60] W.-J. Ko and S.-Y. Chien, “Patch-based face hallucination with multitask deep neural network”, in *2016 IEEE International Conference on Multimedia and Expo (ICME)*, IEEE, 2016, 1–6.

[61] M. Koestinger, P. Wohlhart, P. M. Roth, and H. Bischof, “Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization”, in *2011 IEEE international conference on computer vision workshops (ICCV workshops)*, IEEE, 2011, 2144–51.

[62] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep laplacian pyramid networks for fast and accurate super-resolution", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, 624–32.

[63] H. A. Le and I. A. Kakadiaris, "SeLENet: A semi-supervised low light face enhancement method for mobile face unlock", in *2019 International Conference on Biometrics (ICB)*, IEEE, 2019, 1–8.

[64] V. Le, J. Brandt, Z. Lin, L. Bourdev, and T. S. Huang, "Interactive facial feature localization", in *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part III 12*, Springer, 2012, 679–92.

[65] C.-H. Lee, Z. Liu, L. Wu, and P. Luo, "Maskgan: Towards diverse and interactive facial image manipulation", in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, 5549–58.

[66] G. Li, J. Shi, Y. Zong, F. Wang, T. Wang, and Y. Gong, "Learning attention from attention: Efficient self-refinement transformer for face super-resolution", in *Proceedings of the International Joint Conference on Artificial Intelligence*, Vol. 2, 2023.

[67] L. Li, J. Tang, Z. Ye, B. Sheng, L. Mao, and L. Ma, "Unsupervised face super-resolution via gradient enhancement and semantic guidance", *The Visual Computer*, 37, 2021, 2855–67.

[68] M. Li, Z. Zhang, J. Yu, and C. W. Chen, "Learning face image super-resolution through facial semantic attribute transformation and self-attentive structure enhancement", *IEEE Transactions on Multimedia*, 23, 2020, 468–83.

[69] X. Li, M. Liu, Y. Ye, W. Zuo, L. Lin, and R. Yang, "Learning warped guidance for blind face restoration", in *Proceedings of the European conference on computer vision (ECCV)*, 2018, 272–89.

[70] X. Li, G. Duan, Z. Wang, J. Ren, Y. Zhang, J. Zhang, and K. Song, "Recovering extremely degraded faces by joint super-resolution and facial composite", in *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*, IEEE, 2019, 524–30.

[71] Z. Li, D. Zeng, X. Yan, Q. Shen, and B. Tang, "Analyzing and Combating Attribute Bias for Face Restoration.", in *IJCAI*, 2023, 1151–9.

[72] Y. Lin, J. Shen, Y. Wang, and M. Pantic, "Roi tanh-polar transformer network for face parsing in the wild", *Image and Vision Computing*, 112, 2021, 104190.

[73] C. Liu, H.-Y. Shum, and C.-S. Zhang, "A two-step approach to hallucinating faces: global parametric model and local nonparametric model", in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, Vol. 1, IEEE, 2001, I–I.

[74] H. Liu, Z. Han, J. Guo, and X. Ding, "A noise robust face hallucination framework via cascaded model of deep convolutional networks and manifold learning", in *2018 IEEE International Conference on Multimedia and Expo (ICME)*, IEEE, 2018, 1–6.

[75] H. Liu, Y. Yang, and Y. Liu, "W-Net: A Facial Feature-Guided Face Super-Resolution Network", *arXiv preprint arXiv:2406.00676*, 2024.

[76] L. Liu, S. Wang, and L. Wan, "Component semantic prior guided generative adversarial network for face super-resolution", *IEEE Access*, 7, 2019, 77027–36.

[77] S. Liu, D. Huang, and Y. Wang, "Learning spatial fusion for single-shot object detection", *arXiv preprint arXiv:1911.09516*, 2019.

[78] Y. Liu, Z. Dong, K. P. Lim, and N. Ling, "A densely connected face super-resolution network based on attention mechanism", in *2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, IEEE, 2020, 148–52.

[79] Z. Liu, C. Zhang, Y. Wu, and C. Zhang, "Joint face completion and super-resolution using multi-scale feature relation learning", *Journal of Visual Communication and Image Representation*, 93, 2023, 103806.

[80] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild", in *Proceedings of the IEEE international conference on computer vision*, 2015, 3730–8.

[81] T. Lu, J. Wang, J. Jiang, and Y. Zhang, "Global-local fusion network for face super-resolution", *Neurocomputing*, 387, 2020, 309–20.

[82] T. Lu, Y. Wang, Y. Zhang, J. Jiang, Z. Wang, and Z. Xiong, "Rethinking prior-guided face super-resolution: A new paradigm with facial component prior", *IEEE Transactions on Neural Networks and Learning Systems*, 35(3), 2022, 3938–52.

[83] S. Luo and J. Lu, "GFNet: a gradient information compensation-based face super-resolution network", *IEEE Access*, 10, 2022, 8073–80.

[84] Y. Luo and K. Huang, "Super-resolving tiny faces with face feature vectors", in *2020 10th International Conference on Information Science and Technology (ICIST)*, IEEE, 2020, 145–52.

[85] J. Ma, X. Li, J. Li, J. Wan, T. Liu, and G. Li, "Quality-aware face alignment using high-resolution spatial dependencies", *Multimedia Tools and Applications*, 83(14), 2024, 42165–87.

[86] X. Ma, J. Zhang, and C. Qi, "Hallucinating face by position-patch", *Pattern Recognition*, 43(6), 2010, 2224–36.

[87] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer", *IEEE Signal processing letters*, 20(3), 2012, 209–12.

[88] H. Nie, Y. Lu, and J. Ikram, "Face hallucination via convolution neural network", in *2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI)*, IEEE, 2016, 485–9.

[89] X. Ning, F. Nan, S. Xu, L. Yu, and L. Zhang, “Multi-view frontal face image generation: a survey”, *Concurrency and Computation: Practice and Experience*, 35(18), 2023, e6147.

[90] J.-S. Park and S.-W. Lee, “An example-based face hallucination method for single-frame, low-resolution facial images”, *IEEE Transactions on Image Processing*, 17(10), 2008, 1806–16.

[91] O. Parkhi, A. Vedaldi, and A. Zisserman, “Deep face recognition”, in *BMVC 2015-Proceedings of the British Machine Vision Conference 2015*, British Machine Vision Association, 2015.

[92] W. Peebles and S. Xie, “Scalable diffusion models with transformers”, in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, 4195–205.

[93] B. Peng, E. Alcaide, Q. Anthony, A. Albalak, S. Arcadinho, S. Biderman, H. Cao, X. Cheng, M. Chung, M. Grella, *et al.*, “Rwkv: Reinventing rnns for the transformer era”, *arXiv preprint arXiv:2305.13048*, 2023.

[94] T. Qiu and Y. Yan, “DPHNet: Dual-Path Hybrid Network for Blurry Face Image Super-Resolution”, *IEEE Access*, 2024.

[95] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks”, *arXiv preprint arXiv:1511.06434*, 2015.

[96] S. S. Rajput and K. Arya, “A robust face super-resolution algorithm and its application in low-resolution face recognition system”, *Multimedia Tools and Applications*, 79(33), 2020, 23909–34.

[97] X. Ren, Q. Hui, X. Zhao, J. Xiong, and J. Yin, “BESRGAN: Boundary equilibrium face super-resolution generative adversarial networks”, *IET Image Processing*, 17(6), 2023, 1784–96.

[98] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation”, in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, Springer, 2015, 234–41.

[99] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, “300 faces in-the-wild challenge: The first facial landmark localization challenge”, in *Proceedings of the IEEE international conference on computer vision workshops*, 2013, 397–403.

[100] S. Schaefer, T. McPhail, and J. Warren, “Image deformation using moving least squares”, in *ACM SIGGRAPH 2006 Papers*, 2006, 533–40.

[101] M. B. Shahbakhsh and H. Hassanpour, “Edge-attention network for preserving structure in face super-resolution”, *Multimedia Tools and Applications*, 2024, 1–21.

[102] H. R. Sheikh and A. C. Bovik, “Image information and visual quality”, *IEEE Transactions on image processing*, 15(2), 2006, 430–44.

[103] J. Shi, Y. Wang, Z. Yu, G. Li, X. Hong, F. Wang, and Y. Gong, "Exploiting multi-scale parallel self-attention and local variation via dual-branch transformer-CNN structure for face super-resolution", *IEEE Transactions on Multimedia*, 2023.

[104] R. C. Streijl, S. Winkler, and D. S. Hands, "Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives", *Multimedia Systems*, 22(2), 2016, 213–27.

[105] Y. Tian, H. Chen, C. Xu, and Y. Wang, "Image Processing GNN: Breaking Rigidity in Super-Resolution", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, 24108–17.

[106] Y.-J. Tsai, Y.-L. Liu, L. Qi, K. C. Chan, and M.-H. Yang, "Dual associated encoder for face restoration", *arXiv preprint arXiv:2308.07314*, 2023.

[107] O. Tuzel, Y. Taguchi, and J. R. Hershey, "Global-local face upsampling network", *arXiv preprint arXiv:1603.07235*, 2016.

[108] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need", *Advances in neural information processing systems*, 30, 2017.

[109] C. Wang, J. Jiang, Z. Zhong, D. Zhai, and X. Liu, "Super-resolving face image by facial parsing information", *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2023.

[110] G. Wang, J. Li, J. Xie, J. Xu, and B. Yang, "EfficientSRFace: An Efficient Network with Super-Resolution Enhancement for Accurate Face Detection", in *Asian Conference on Pattern Recognition*, Springer, 2023, 74–87.

[111] J. Wang, T. Lu, X. Huang, R. Zhang, and X. Feng, "Pan-sharpening via conditional invertible neural network", *Information Fusion*, 101, 2024, 101980.

[112] J. Wang, Q. Zhou, X. Huang, R. Zhang, X. Chen, and T. Lu, "Pan-sharpening via intrinsic decomposition knowledge distillation", *Pattern Recognition*, 149, 2024, 110247.

[113] T. Wang, Y. Xiao, Y. Cai, G. Gao, X. Jin, L. Wang, and H. Lai, "UFSRNet: U-shaped face super-resolution reconstruction network based on wavelet transform", *Multimedia Tools and Applications*, 2024, 1–19.

[114] X. Wang and X. Tang, "Hallucinating face by eigentransformation", *IEEE transactions on systems, man, and cybernetics, part C (applications and reviews)*, 35(3), 2005, 425–34.

[115] Y. Wang, T. Lu, R. Xu, and Y. Zhang, "Face super-resolution by learning multi-view texture compensation", in *MultiMedia Modeling: 26th International Conference, MMM 2020, Daejeon, South Korea, January 5–8, 2020, Proceedings, Part II 26*, Springer, 2020, 350–60.

[116] Y. Wang, T. Lu, F. Yao, Y. Wu, and Y. Zhang, "Multi-view texture learning for face super-resolution", *IEICE TRANSACTIONS on Information and Systems*, 104(7), 2021, 1028–38.

[117] Y. Wang, T. Lu, Y. Yao, Y. Zhang, and Z. Xiong, "Learning to hallucinate face in the dark", *IEEE Transactions on Multimedia*, 2023.

[118] Y. Wang, T. Lu, Y. Zhang, J. Jiang, J. Wang, Z. Wang, and J. Ma, "Tanet: a new paradigm for global face super-resolution via transformer-cnn aggregation network", *arXiv preprint arXiv:2109.08174*, 2021.

[119] Y. Wang, T. Lu, Y. Zhang, Z. Wang, J. Jiang, and Z. Xiong, "FaceFormer: Aggregating global and local representation for face hallucination", *IEEE Transactions on Circuits and Systems for Video Technology*, 33(6), 2022, 2533–45.

[120] Y. Wang, W. Yang, X. Chen, Y. Wang, L. Guo, L.-P. Chau, Z. Liu, Y. Qiao, A. C. Kot, and B. Wen, "SinSR: diffusion-based image super-resolution in a single step", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, 25796–805.

[121] Z. Wang, Z. Zhang, X. Zhang, H. Zheng, M. Zhou, Y. Zhang, and Y. Wang, "Dr2: Diffusion-based robust degradation remover for blind face restoration", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, 1704–13.

[122] Z.-Y. Wang, Z. Han, R.-M. Hu, and J.-J. Jiang, "Noise robust face hallucination employing Gaussian–Laplacian mixture model", *Neurocomputing*, 133, 2014, 153–60.

[123] Z. Wang and A. C. Bovik, "A universal image quality index", *IEEE signal processing letters*, 9(3), 2002, 81–4.

[124] H. Wu, H. Qi, H. Zhang, Z. Jin, D. Salihu, and J.-F. Hu, "Reconstruction with robustness: A semantic prior guided face super-resolution framework for multiple degradations", *Image and Vision Computing*, 140, 2023, 104857.

[125] W. Wu, C. Qian, S. Yang, Q. Wang, Y. Cai, and Q. Zhou, "Look at boundary: A boundary-aware face alignment algorithm", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, 2129–38.

[126] W. Wu, Y. Zhang, C. Li, C. Qian, and C. C. Loy, "Reenactgan: Learning to reenact faces via boundary transfer", in *Proceedings of the European conference on computer vision (ECCV)*, 2018, 603–19.

[127] B. Xia, Y. Tian, Y. Zhang, Y. Hang, W. Yang, and Q. Liao, "Meta-learning based degradation representation for blind super-resolution", *IEEE Transactions on Image Processing*, 2023.

[128] J. Xiu, X. Qu, and H. Yu, "Double discriminative face super-resolution network with facial landmark heatmaps", *The Visual Computer*, 39(11), 2023, 5883–95.

[129] C.-Y. Yang, S. Liu, and M.-H. Yang, “Structured face hallucination”, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, 1099–106.

[130] J. Yang, J. Wright, T. S. Huang, and Y. Ma, “Image super-resolution via sparse representation”, *IEEE transactions on image processing*, 19(11), 2010, 2861–73.

[131] L. Yang, S. Wang, S. Ma, W. Gao, C. Liu, P. Wang, and P. Ren, “Hifacegan: Face renovation via collaborative suppression and replenishment”, in *Proceedings of the 28th ACM international conference on multimedia*, 2020, 1551–60.

[132] P. Yang, S. Zhou, Q. Tao, and C. C. Loy, “PGDiff: Guiding Diffusion Models for Versatile Face Restoration via Partial Guidance”, *Advances in Neural Information Processing Systems*, 36, 2024.

[133] S. Yang, P. Luo, C.-C. Loy, and X. Tang, “Wider face: A face detection benchmark”, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, 5525–33.

[134] Y. Yin, J. P. Robinson, S. Jiang, Y. Bai, C. Qin, and Y. Fu, “Superfront: From low-resolution to high-resolution frontal face synthesis”, in *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, 1609–17.

[135] X. Yu, B. Fernando, B. Ghanem, F. Porikli, and R. Hartley, “Face super-resolution guided by facial component heatmaps”, in *Proceedings of the European conference on computer vision (ECCV)*, 2018, 217–33.

[136] X. Yu and F. Porikli, “Ultra-resolving face images by discriminative generative networks”, in *European conference on computer vision*, Springer, 2016, 318–33.

[137] Y. Yu, J. Jiang, H. Chang, H. Zheng, and S. Wang, “Face Super-Resolution via Joint Edge Information and Attention Aggregation Network”, *Computers and Electrical Engineering*, 111, 2023, 108931.

[138] Z. Yue and C. C. Loy, “Difface: Blind face restoration with diffused error contraction”, *arXiv preprint arXiv:2212.06512*, 2022.

[139] J. U. Yun, B. Jo, and I. K. Park, “Joint face super-resolution and deblurring using generative adversarial network”, *IEEE Access*, 8, 2020, 159661–71.

[140] K. Zeng, Z. Wang, T. Lu, J. Chen, J. Wang, and Z. Xiong, “Self-attention learning network for face super-resolution”, *Neural Networks*, 160, 2023, 164–74.

[141] K. Zengy, Z. Wang, T. Luz, J. Chen, Z. He, and Z. Han, “Implicit Mutual Learning With Dual-Branch Networks for Face Super-Resolution”, *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2024.

[142] L. Zhang, L. Zhang, X. Mou, and D. Zhang, “FSIM: A feature similarity index for image quality assessment”, *IEEE transactions on Image Processing*, 20(8), 2011, 2378–86.

[143] M. Zhang and Q. Ling, "Supervised pixel-wise GAN for face super-resolution", *IEEE Transactions on Multimedia*, 23, 2020, 1938–50.

[144] P. Zhang, K. Zhang, W. Luo, C. Li, and G. Wang, "Blind face restoration: Benchmark datasets and a baseline model", *Neurocomputing*, 574, 2024, 127271.

[145] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, 586–95.

[146] Y. Zhang, I. W. Tsang, Y. Luo, C.-H. Hu, X. Lu, and X. Yu, "Copy and paste GAN: Face hallucination from shaded thumbnails", in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, 7355–64.

[147] K. Zhao, T. Lu, J. Wang, Y. Zhang, J. Jiang, and Z. Xiong, "Hyper-Laplacian Prior for Remote Sensing Image Super-Resolution", *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[148] W. Zheng, L. Yan, W. Zhang, C. Gou, and F.-Y. Wang, "Guided cyclegan via semi-dual optimal transport for photo-realistic face super-resolution", in *2019 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2019, 2851–5.

[149] E. Zhou, H. Fan, Z. Cao, Y. Jiang, and Q. Yin, "Learning face hallucination in the wild", in *Proceedings of the AAAI conference on artificial intelligence*, Vol. 29, No. 1, 2015.

[150] S. Zhou, K. Chan, C. Li, and C. C. Loy, "Towards robust blind face restoration with codebook lookup transformer", *Advances in Neural Information Processing Systems*, 35, 2022, 30599–611.

[151] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks", in *Proceedings of the IEEE international conference on computer vision*, 2017, 2223–32.

[152] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild", in *2012 IEEE conference on computer vision and pattern recognition*, IEEE, 2012, 2879–86.