Survey on Deep Face Restoration: From Non-blind to Blind and Beyond

Wenjie Li, Mei Wang, Kai Zhang, Juncheng Li, Xiaoming Li, Yuhang Zhang, Guangwei Gao*, *Senior Member, IEEE,* Weihong Deng*, *Member, IEEE* and Chia-Wen Lin, *Fellow, IEEE*

Abstract—Face restoration (FR) is a specialized field within image restoration that aims to recover low-quality (LQ) face images into high-quality (HQ) face images. Recent advances in deep learning technology have led to significant progress in FR methods. In this paper, we begin by examining the prevalent factors responsible for real-world LQ images and introduce degradation techniques used to synthesize LQ images. We also discuss notable benchmarks commonly utilized in the field. Next, we categorize FR methods based on different tasks and explain their evolution over time. Furthermore, we explore the various facial priors commonly utilized in the restoration process and discuss strategies to enhance their effectiveness. In the experimental section, we thoroughly evaluate the performance of state-of-the-art FR methods across various tasks using a unified benchmark. We analyze their performance from different perspectives. Finally, we discuss the challenges faced in the field of FR and propose potential directions for future advancements. The open-source repository corresponding to this work can be found at *https://github.com/24wenjie-li/Awesome-Face-Restoration*.

Index Terms—Face restoration, Survey, Deep learning, Non-blind/Blind, Joint restoration tasks, Facial priors.

1 Introduction

Face restoration (FR) aims to improve the quality of degraded face images and recover accurate and high-quality (HQ) face images from low-quality (LQ) face images. This process is crucial for various downstream tasks such as face detection [1], face recognition [2], and 3D face reconstruction [3]. The concept of face restoration was first introduced by Baker *et al.* [4] in 2000. They developed a pioneering prediction model to enhance the resolution of low-resolution face images. Since then, numerous FR methods have been developed, gaining increasing attention from researchers in the field. Traditional FR methods primarily involve deep analysis of facial priors and degradation approaches. However, these methods often struggle to meet engineering requirements. With breakthroughs in deep learning technology, a multitude of deep learning-based methods specifically designed for FR tasks have emerged. Deep learning networks, utilizing large-scale datasets, are capable of effectively capturing diverse mapping relationships between degraded face images and real face images. Consequently, deep learning-based FR methods [5], [6] have demonstrated significant advantages over traditional methods, offering more robust solutions.

- *: Corresponding author.
- Wenjie Li, Mei Wang, Yuhang Zhang and Weihong Deng are with the Pattern Recognition and Intelligent System Laboratory, School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing, China. (e-mail: {cswjli, wangmei1, zyhzyh, whdeng}@bupt.edu.cn).
- Kai Zhang is with the Computer Vision Lab, ETH Zürich, Zürich, Switzerland (e-mail: kai.zhang@vision.ee.ethz.ch).
- Juncheng Li is with the School of Communication and Information Engineering, Shanghai University, Shanghai, China. (e-mail: cvjunchengli@gmail.com).
- Xiaoming Li is with the Nanyang Technological University, Singapore. (e-mail: csxmli@gmail.com).
- Guangwei Gao is with the Intelligent Visual Information Perception Laboratory, Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing, China. (e-mail: csggao@gmail.com).
- Chia-Wen Lin is with the Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan. (e-mail: cwlin@ee.nthu.edu.tw).

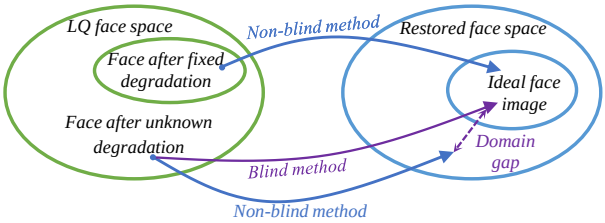


Figure 1: Domain interpretation of differences between non-blind and blind method. If the degradation factors affecting the stochastic LQ face differ from those assumed by the non-blind method (*e.g.*, bicubic downsampling or fixed blur kernel), it can result in a significant domain gap between the restored face image and the ideal HQ face image.

Most deep learning-based face restoration methods are trained using a fully supervised approach, where HQ face images are artificially degraded to synthesize paired LQ face images for training. In earlier non-blind methods [5], [7], [8], HQ face images were degraded using fixed degradation techniques, typically bicubic downsampling. However, as shown in Fig. 1, when the model is trained on LQ facial images synthesized in this specific manner, there can be a notable domain gap between the restored facial images and ideal HQ facial images. To address this issue, blind methods [6], [9], [10] have been developed. These methods simulate the realistic degradation process by incorporating an array of unknown degradation factors such as blur, noise, low resolution, and lossy compression. By considering more complex and diverse degradation scenarios and accounting for variations in poses and expressions, blind restoration methods have proven to be more applicable to real-world scenarios. Furthermore, a series of joint face restoration tasks have emerged to tackle specific challenges in face restoration [11], [12], [13], [14], [15]. These tasks include joint face alignment and restoration [11], joint face recognition and restoration [12], joint illumination compensation and

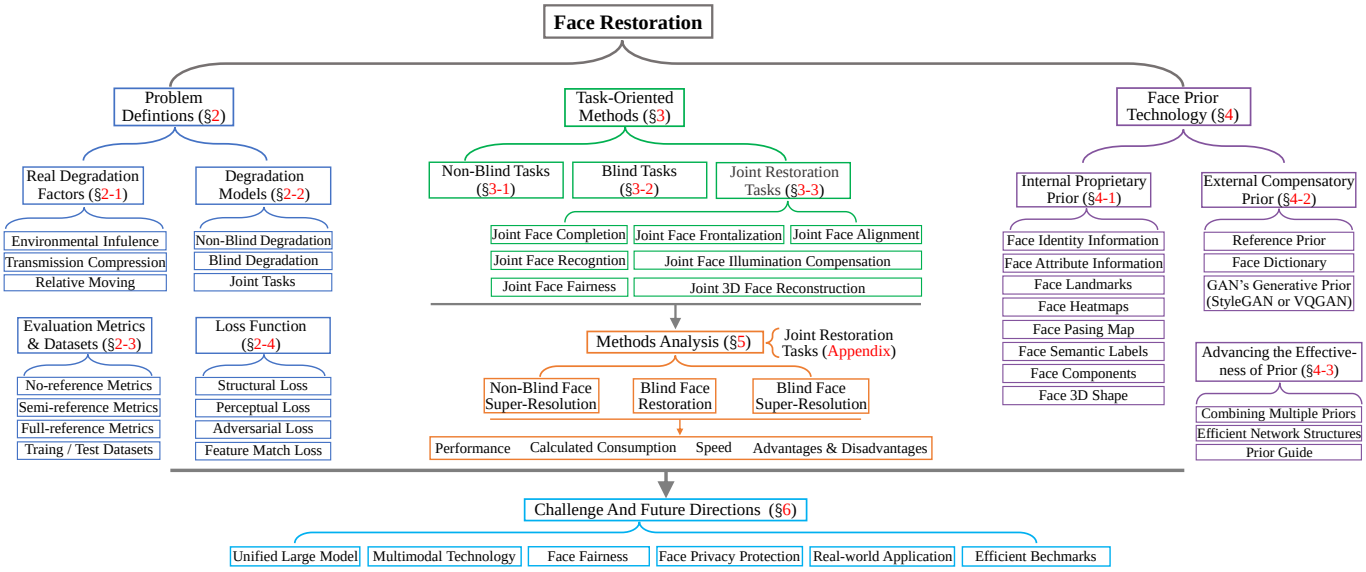


Figure 2: Outline of our deep learning-based face restoration survey.

Table 1: A summary of other deep learning based FR reviews.

Year	Surveys	Related Topic	Venue
2019	Liu *et al.* [16]	GAN-based face super-resolution	IET
2021	Jiang *et al.* [17]	Deep learning-based face super-resolution	CSUR
2023	Wang *et al.* [18]	Deep learning-based face restoration	Arxiv

restoration [13], joint 3D face reconstruction and restoration [14], and joint face fairness and restoration [15]. Building upon these advancements, our paper aims to provide a comprehensive survey of deep learning-based non-blind/blind face restoration methods and their joint tasks. By presenting this overview, we aim to shed light on the current state of development in the field, the technical approaches employed, the existing challenges, and potential directions.

Despite the rapid growth in the field of FR, there is a relative scarcity of reviews specifically focusing on deep learning-based FR methods. As depicted in TABLE 1, Liu *et al.* [16] provided a review of face super-resolution methods based on generative adversarial networks, but it solely focused on a specific technique within FR. Jiang *et al.* [17] presented an overview of deep learning-based face super-resolution, covering FR tasks beyond super-resolution, but the emphasis remained on summarizing face super-resolution. Wang *et al.* [18] conducted a survey on FR, however, it adopted a classification pattern of sub-tasks in the image restoration domain, such as denoising, deblurring, super-resolution, and artifact removal. These patterns might not effectively generalize to existing FR methods, which could result in the omission of joint tasks related to FR. In contrast, our review provides a comprehensive summary of current FR methods from three distinct classification perspectives: blind, non-blind, and joint restoration tasks. By considering these perspectives, we not only encompass a broader range of methods related to FR but also clarify the characteristics of methods under different tasks. In the experimental section, while Wang's work [18] primarily focused on blind methods, we conduct a comprehensive analysis of both blind and non-blind methods across various aspects. Furthermore, we provide a comparison of the methods within the joint tasks. As a result, our work provides an accurate perspective on non-blind/blind tasks and joint tasks, aiming to inspire new research within the community through insightful analysis.

The main contributions of our survey are as follows: **(I)** We compile the factors responsible for the degradation of real-world images and explain the degradation models used to synthesize diverse LQ face images. **(II)** We classify the field of FR based on blind, non-blind tasks and joint tasks criteria, providing a comprehensive overview of technological advancements within these domains. **(III)** Addressing the uncertainties stemming from the absence of consistent benchmarks in the field, we conduct a fair comparison of popular FR methods using standardized benchmarks. Additionally, we discuss the challenges and opportunities based on the experimental results.

Fig. 2 provides an overview of the structure of this survey. In Section 2, we summarize the real-world factors contributing to the appearance of LQ face images and present corresponding artificial synthesis methods. We also discuss notable benchmarks used in the field. In Section 3, we introduce existing methods for different subtasks within FR. Section 4 covers various popular priors and methods for enhancing prior validity in the restoration process. In Section 5, We conduct extensive experiments to compare state-of-the-art FR methods. Section 6 addresses the challenges faced in FR and presents potential future directions. Finally, we conclude this survey in Section 7.

2 Problem Definitions

In this section, we will discuss the presence of degradation factors in real-world scenarios, followed by an introduction to artificial degradation models. Additionally, we will cover commonly used loss functions, evaluation metrics, and datasets that are frequently employed in this field.

2.1 Real Degradation Factors

In real-world scenarios, face images are susceptible to degradation during the imaging and transmission process due to the complex environment. The degradation of facial images is primarily caused by the limitations of the physical imaging equipment and external imaging conditions. We

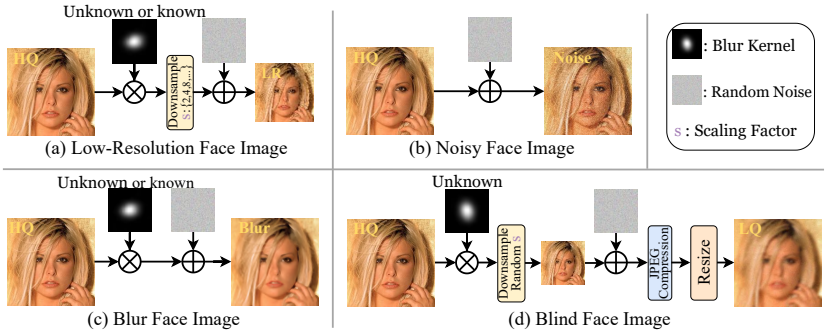


Figure 3: Methods for generating various types of degradation facial images.

Table 2: Summary of key evaluation metrics.

Metrics	Highlight
PSNR [19]	Full reference, pixel-by-pixel comparison of the differences between both.
SSIM [20]	Full reference, focus on differences in brightness, contrast, structure, *etc.*
MS-SSIM [21]	Full reference, average SSIM for windows.
LPIPS [22]	Full reference, focus on the visual perceptual similarity between both.
IDD [23]	Full reference, assess identity consistency.
FID [24]	Semi-reference, measure the difference in distribution between both.
NIQE [25]	No reference, evaluate image naturalness.
MOS [26]	Subjective scoring by groups.

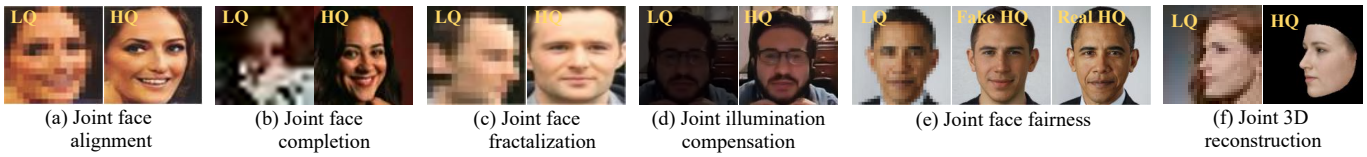


Figure 4: Demonstration of LQ and HQ face images for some joint face restoration tasks.

can summarize the main factors contributing to image degradation as follows: (1) Environmental influence: Particularly the low or high light conditions; (2) Camera shooting process: Internal factors related to the camera itself, such as optical imaging conditions, noise, and lens distortion, as well as external factors like relative displacement between the subject and the camera, such as camera shake or capturing moving face; (3) Compression during transmission: Lossy compression during image transmission and surveillance storage. To replicate realistic degradation, researchers have made various attempts. Initially, they utilized fixed blur kernels, such as Gaussian blur or downsampling, to simulate realistic blurring or low resolution. Later, randomized blur kernels were experimented with to improve robustness by introducing a wider range of degradation patterns. Additionally, considering the diversity of face-related tasks, extensive research has been conducted on joint FR tasks to recover LQ faces in specific scenes.

2.2 Degradation Models

Due to the challenge of acquiring real HQ and LQ face image pairs, researchers often resort to using degradation models to generate synthetic LQ images I_{lq} from HQ images I_{hq}. Generally, The I_{lq} is the output of the I_{hq} after degradation:

$$I_{lq} = D(I_{hq}; \delta), \tag{1}$$

where D represents the degradation function and δ represents the parameter involved in the degradation process (*e.g.*, the downsampling or noise or blur kernel). As shown in Fig. 3, different δ can result in various types of degradation. Existing FR tasks can be categorized into four subtasks based on the type of degradation: face denoising, face deblurring, face super-resolution, and blind face restoration. The distinction between non-blind and blind lies in whether the degradation factors are known. TA subtask in face restoration is considered non-blind when the degradation factors are known and can be explicitly modeled. Conversely, if the degradation factors are unknown and cannot be precisely modeled, the FR task is classified as blind.

• **Non-blind Degradation Models.** (I) The non-blind task primarily focuses on face super-resolution (FSR) [41], also known as face hallucination [27]. As shown in Fig. 3 (a), its degradation model involves degrading a high-resolution (HR) face image into a low-resolution (LR) face image. When the blur kernel is pre-determined and remains constant, such as a Gaussian blur kernel or any other well-defined blur kernel, FSR can be categorized as a non-blind task. The degradation model can be described as follows:

$$I_{lr} = (I_{hq} \otimes k_f)\downarrow_s + n, \tag{2}$$

where I_{lr} represents the LR face image, I_{hq} represents the HR face image, $\otimes$ represents the convolutional operation, k_f represents the fixed blur kernel, $\downarrow_s$ denotes the downsampling operation with scale factor s, typically set to 4, 8, 16 and 32, and n represents the additive Gaussian noise. Additionally, most researchers directly employ this degradation model to simplify the FSR's degradation process as:

$$I_{lr} = (I_{hq})\downarrow_s. \tag{3}$$

(II) Face denoising [68] and face deblurring [69] primarily focus on removing additive noise from face images or simulating the removal of motion blur in a realistic face captured by a camera. Similarly, as shown in Fig. 3 (b) and (c), when the blur kernel remains constant, they can be classified as non-blind tasks. Their degradation model can be described separately as:

$$I_n = I_{hq} + n, \tag{4}$$

$$I_b = I_{hq} \otimes k_f + n, \tag{5}$$

where I_n represents the face image containing noise, I_b represents the blurred image, I_{hq} represents the clean HQ face image, k_f represents the fixed blur kernel and n represents the additive Gaussian noise.

• **Blind Degradation Models.** (I) When the blur kernel in degradation models is randomly generated or composed of multiple unknown blur kernels, the nature of the blur kernel becomes essentially unknown. In such cases, both face super-resolution [52] and face deblurring [70] can be classified as blind tasks. As shown in Fig. 3 (a) and (c), their degradation processes can be described separately as follows:

$$I_{lr} = (I_{hq} \otimes k_u)\downarrow_s + n, \tag{6}$$

Table 3: Summary of benchmark datasets used in existing face restoration methods.

Year	Dataset	Size	Attributes	Landmarks	Parsing maps	Identity	HQ-LQ	Methods
2008	**LFW** [2]	13K	73	✗	✗	✓	HQ	C-SRIP [27], LRFR [28], DPDFN [29], *etc.*
2010	**Multi-PIE** [30]	755.4K	✗	✗	✗	✓	HQ	FSGN [31], CPGAN [32], MDCN [33], *etc.*
2011	**AFLW** [34]	26K	✗	21	✗	✗	HQ	FAN [35], JASRNet [36], *etc.*
2011	**SCFace** [37]	4.2K	✗	4	✗	✗	HQ	MNCE [8], SISN [38], CTCNet [39], *etc.*
2012	**Helen** [40]	2.3K	✗	194	✓	✗	HQ	DIC [41], SAAN [42], SCTANet [43], *etc.*
2014	**CASIA-WebFace** [44]	494.4K	✗	2	✗	✓	HQ	MDFR [45], C-SRIP [27], GFRNet [9], *etc.*
2015	**CelebA** [46]	202.6K	40	5	✗	✓	HQ	FSRNet [5], SPARNet [47], SFMNet [48], *etc.*
2016	**Widerface** [1]	32.2K	✗	✗	✗	✗	HQ	Se-RNet [49], SCGAN [50], *etc.*
2017	**LS3D-W** [51]	230K	✗	68	✗	✗	HQ	Super-FAN [52], SCGAN [50], *etc.*
2017	**Menpo** [53]	9K	✗	68/39	✗	✗	HQ	SAM3D [54], [55], *etc.*
2018	**VGGFace2** [56]	3310K	✗	✗	✗	✓	HQ	GFRNet [9], ASFFNet [57], GWAInet [58], *etc.*
2019	**FFHQ** [59]	70K	✗	68	✗	✗	HQ	mGANprior [60], GFPGAN [6], VQFR [61], *etc.*
2020	**CelebAMask-HQ** [62]	30K	✗	✗	✓	✗	HQ	MSGGAN [63], GPEN [10], Pro-UIGAN [64], *etc.*
2022	**EDFace-Celeb-1M** [65]	1700K	✗	✗	✗	✗	HQ-LQ	STUNet [66], *etc.*
2022	**CelebRef-HQ** [67]	10.6K	✗	✗	✗	✓	HQ	DMDNet [67], *etc.*

$$I_b = I_{hq} \otimes k_u + n, \tag{7}$$

where k_u is the unknown blur kernel, and the remaining variables have the same meanings as described above for non-blind face super-resolution and face deblurring.

(II) Since the above tasks focus on a single type of degradation, they face challenges in handling severely degraded face images encountered in real-world scenarios. Blind face restoration [23], [71], [72] aims to address this limitation by considering more complex degradations, making it the most prominent task in the field currently. GFRNet [9] is a pioneering work in blind face restoration by introducing a more intricate degradation model aimed at simulating realistic deterioration for the first time. As shown in Fig. 3 (d), the degradation model in blind face restoration encompasses random noise, unknown blur, arbitrary scale downsampling, and random JPEG compression artifacts. This degradation process can be formulated as follows:

$$I_{lq} = \{JPEG_q((I_{hq} \otimes k_u)\downarrow_{s_r} + n_r)\}\uparrow_{s_r}, \tag{8}$$

where I_{lq} and I_{hq} represent the low-quality and high-quality face images, respectively. $JPEG_q$ represents JPEG compression operation with arbitrary quality factor, k_u represents an unknown blur kernel. $\downarrow_{s_r}$ and $\uparrow_{s_r}$ represent down-sampling and up-sampling operations with arbitrary scale factors s_r, respectively. n_r represents random noise.

• **Joint Tasks.** Due to the multitude of joint tasks, we do not introduce the degradation models for each of them individually. Fig. 4 showcases several examples of joint tasks, depicted from left to right: (a) Joint face alignment and restoration [11]: This task addresses the challenge of misaligned faces by aligning and restoring them. (b) Joint face completion and restoration [73]: The objective is to handle face occlusions and restore the missing regions in the face image. (c) Joint face frontalization and restoration [74]: This task focuses on recovering frontal faces from side faces, enhancing their appearance and quality. (d) Joint face illumination compensation and restoration [13]: This task aims to restore faces captured in low-light conditions, compensating for the lack of illumination. (e) Joint face fairness and restoration [15]: This task aims to improve the accuracy of face restoration across different human races, promoting fairness and inclusivity. (f) Joint 3D face reconstruction [3]: This task aims to improve the accuracy of 3D reconstruction of low-quality faces. In each case, the HQ face images are represented on the right, while the degraded LQ face images corresponding to each specific task are shown on the left. These joint tasks are designed to address face restoration challenges in specific scenarios and hold practical significance in their respective domains.

2.3 Evaluation Metrics And Datasets

We have compiled a selection of the most widely used evaluation metrics in the field of FR, as presented in TABLE 2. We classify these metrics into three groups: full-reference metrics, which necessitate paired HQ face images; semi-reference metrics, which only require unpaired HQ face images; and no-reference metrics, which don't involve any face images for measurement. Additionally, more metrics can be found at *https://github.com/chaofengc/Awesome-Image-Quality-Assessment*. Furthermore, we summarize commonly used benchmark datasets for FR in TABLE 3, including the number of face images, the facial features included, the availability of HQ-LQ pairs, and previous methods that have utilized these datasets. For datasets that only provide HQ images, we need to synthesize the corresponding LQ images using the degradation model introduced in Sections 2.2.

2.4 Loss Function

The researchers aim to estimate the approximation of the HQ face image I_{hq}, denoted as $\hat{I}_{hq}$, from the LQ face image I_{lq}, following:

$$\hat{I}_{hq} = D^{-1}(I_{lq}, \delta) = F(I_{lq}, \theta), \tag{9}$$

where F represents the face restoration method and θ represents the parameters of the method. During the training, the optimization process can be formulated as follows:

$$\hat{\theta} = \arg\min L(\hat{I}_{hq}, I_{hq}), \tag{10}$$

where $\hat{\theta}$ represents the optimization parameter in the training process, L represents the loss between $\hat{I}_{hq}$ and I_{hq}. Different loss functions can yield varying results in face restoration. Initially, researchers commonly used structural losses; however, these losses have limitations, such as over-smoothing the output images. To overcome these limitations, perceptual losses and adversarial losses were developed. Furthermore,

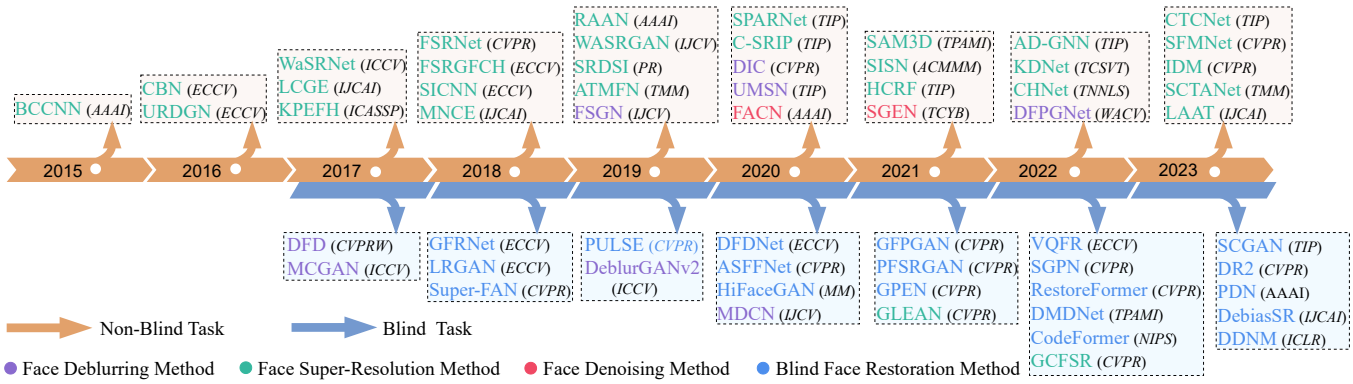


Figure 5: Milestones of deep learning-based non-blind/blind task methods, including its name and venues.

because of the structured nature of faces, a large number of face-specific losses have also been proposed.

• **Structural loss.** Structural losses are employed to minimize the structural differences between two face images. The most commonly used structural losses are pixel-wise losses, which include $L1$ loss [6], [10], [23] and the $L2$ loss [5], [57], [71]. They can be formulated as

$$L_i = \left\| I_{hq}(h,w,c) - \hat{I}_{hq}(h,w,c) \right\|_i, i \in \{1,2\}, \quad (11)$$

where h, w, and c represent the height, width, and number of channels, respectively. The pixel-level loss also encompasses the Huber loss [75] and Carbonnier penalty function. Furthermore, in addition to the pixel-level losses, textural losses have been developed. These include the SSIM loss [27], which promotes image textural similarity, and the cyclic consistency loss [50], which facilitates cooperation between recovery and degradation processes. While minimizing these structural losses encourages the restored image to closely match the ground truth image in terms of pixel values, resulting in a similar structure between the two face images and a higher PSNR value, there is a disadvantage. The recovered face image, however, tends to be too smooth and lacks fine details.

• **Perceptual loss.** The perceptual loss is intended to enhance the visual quality of the recovered images by comparing them to the ground truth images in the perceptual domain using a pre-trained network, such as VGG, Inception *etc.*. The prevalent approach is to calculate the loss based on features extracted from specific intermediate or higher layers of the pre-trained network, as these features represent high-level semantic information within the image. Denoting the l-th layer involved in the computation of the pre-trained network as φ_l, its perceptual loss L_{per}^l can be expressed as follows:

$$L_{per}^l = \left\| \varphi_l(I_{hq}(h,w,c)) - \varphi_l(\hat{I}_{hq}(h,w,c)) \right\|_2, \quad (12)$$

• **Adversarial loss.** The adversarial loss is a common type of loss used in GAN-based face restoration methods [23], [71], [72]. In this setup, the generator G aims to generate an HQ face image to deceive the discriminator D, while the discriminator D strives to distinguish between the generated image and the ground-truth image. The generator and discriminator are trained alternately to generate visually more realistic images. The loss can be expressed as follows:

$$L_{adv,D} = E_{I_{hq}}[\log(1 - D(G(I_{hq}))) + \log(D(I_{hq}))], \quad (13)$$

$$L_{adv,G} = E_{I_{hq}}[\log(1 - D(G(I_{hq})))], \quad (14)$$

where $L_{adv,G}$ and $L_{adv,D}$ are the adversarial losses of the generator and discriminator, respectively. It is worth noting that the use of adversarial loss can sometimes result in training instabilities, so careful parameter tuning is necessary. Furthermore, although models trained with adversarial loss can generate visually appealing results, they may also introduce artifacts, resulting in less faithful face images.

• **Feature match loss.** The structured nature of the human face allows for the integration of specific structural features into the supervised process, leading to improved accuracy in restoration. These features include face landmarks [5], face heatmaps [52], 3D face shape [54], semantic-aware style [76], face parsing [71], facial attention [35], face identity [27], and facial components [6]. Among these, the face landmarks loss is widely utilized and can be described as

$$L_{landmarks} = \frac{1}{N} \sum_{n=1}^{N} \left\| l_{x,y}^n - \hat{l}_{x,y}^n \right\|_2, \quad (15)$$

where N is the number of facial landmarks, and $l_{x,y}^n$ and $\hat{l}_{x,y}^n$ represent the coordinates of the n-th landmark point in the HQ face and the recovered face, respectively. Face-specific losses take into account the specific characteristics and details of facial images. By incorporating these losses, the model can better preserve facial attributes, improve facial details, and enhance the overall visual quality of the restored face.

3 Task-Oriented Methods

In this section, we will summarize and discuss the methodology for each of the three types of face restoration tasks: non-blind tasks, blind tasks, and joint restoration tasks. Fig. 5 illustrates several notable methods in recent years that focus on non-blind and blind tasks. Fig. 6 showcases several landmark methods in recent years that specialize in joint face restoration tasks.

3.1 Non-blind Tasks

The initial attempts in the field of FR primarily focused on non-blind methods. Earlier non-blind methods did not consider facial priors and directly mapped LQ images to HQ images, as depicted in Fig. 7 (a). One pioneering work is the bi-channel convolutional neural network (BCCNN) proposed by Zhou *et al.* [77], which significantly surpasses previous conventional approaches. This network combines the extracted face features with the input face features and

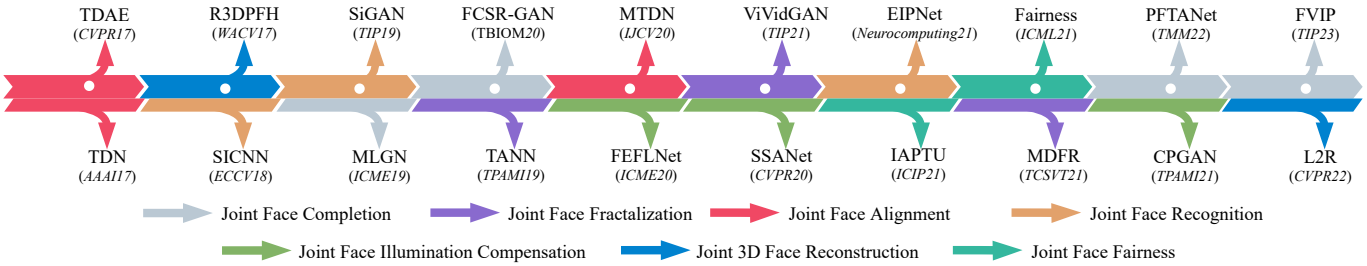


Figure 6: Milestones of the Joint Face Restoration methods, including its name and venues.

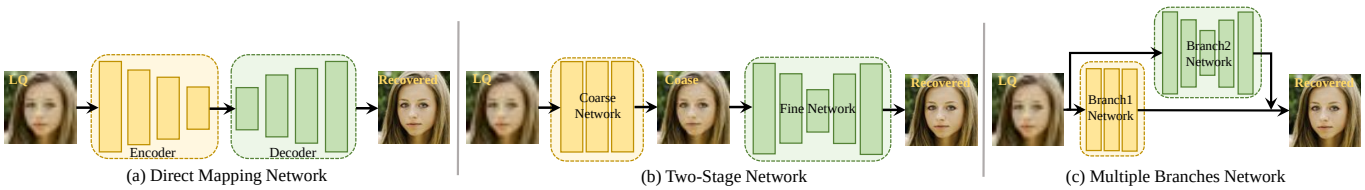
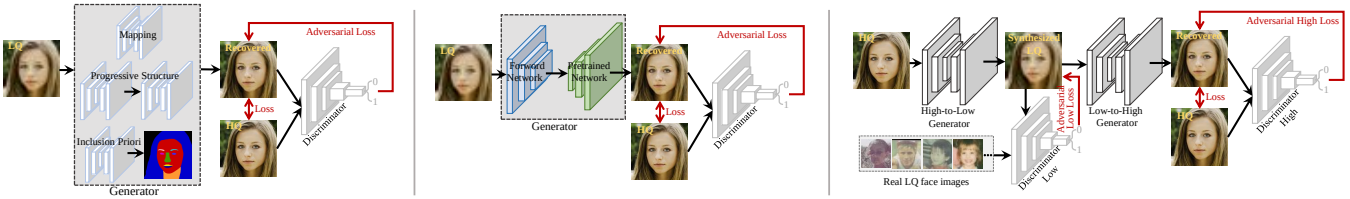


Figure 7: Summary of the architecture of general methods for non-blind face restoration.

utilizes a decoder to reconstruct HQ face images, leveraging its strong fitting capability. Similarly, other methods [78], [79], [80] also adopt direct LQ to HQ mapping networks. Subsequently, non-blind methods incorporated novel techniques, such as learning strategies and prior constraints, into the mapping network to achieve more robust and accurate face restoration. Specifically, as shown in Fig. 7 (b), one class of methods adopts a two-stage approach for face restoration, consisting of roughing and refining stages. For example, CBN [81] employs a cascaded framework to address the performance limitations observed in previous methods when dealing with misaligned facial images. LCGE [82], MNCE [8], and FSGN [31] generate facial components that approximate real landmarks and enhance them by recovering details. FSRNet [5] obtains a rough face image through a network and then refines it using a heatmap and a resolving map of facial landmarks. DIDnet [83] and ATSENet [84] utilize facial identity or attributes to enhance the features extracted by the initial network and recover face images with higher confidence. FAN [35] employs a facial attention prior loss to constrain each incremental stage and gradually increase the resolution. Another class of methods adopts a multi-branch structure for facial restoration, as depicted in Fig. 7 (c). For example, KPEFH [85] utilizes multiple branches in the network to predict key components of the face separately. FSRGFCH [86] enhances the quality of facial details by predicting the face component heatmap with an additional branching in the network. UMSN [87] employs multiple branches to predict regions of different semantic categories of the face separately and then combines them.

Attention mechanisms have demonstrated their effectiveness in image restoration methods [88], [89]. Subsequently, there has been a significant focus on integrating attention mechanisms [90] to enhance the handling of important facial regions. Various networks based on attention mechanisms have been developed, as illustrated in Fig. 9. Attention can be categorized into four types: channel attention, spatial attention, self-attention, and hybrid attention. Channel attention-based approaches [42], [91], [92], [93] emphasize the relative weights between different feature channels in the model, enabling selective emphasis on important channels. Spatial attention-based approaches [47], [55], [94] focus on capturing spatial contextual information about features, enabling the model to prioritize features relevant to key face structures. Self-attention-based approaches [95], [96], [97] mainly capture global facial information, yielding excellent performance. Some approaches [35], [47] also enhance individual attention mechanisms to better suit the specific requirements of FR tasks. Hybrid attention-based approaches [39], [43], [98] combine the aforementioned three main types of attention, aiming to leverage the advantages of different attention types to improve the overall performance of restoration models. Furthermore, some approaches leverage specific types of prior to guide the network. For instance, SAAN [42] incorporates the face parsing map, FAN [35] incorporates the face landmark, SAM3D [54], [55] incorporates the 3D face information, HaPSR [99] incorporates the face heatmap, and CHNet [94] incorporates the face components. To direct attention more precisely, some methods have started to artificially delineate and recover different regions of the face image. WaSRNet [100] employs wavelet transform to convert various regions of the image into coefficients and then performs restoration processing at different levels in the wavelet coefficient domain. SRDSI [101] uses PCA to decompose faces into low-frequency and high-frequency components and then employs deep and sparse networks to recover these two parts, respectively. SFMNet [48] integrates information extracted from its spatial and frequency branches, enhancing the texture of the contour.

The Generative Adversarial Network (GAN) has gained significant popularity due to its ability to generate visually appealing images. It consists of a generator and a discriminator. The generator's role is to produce realistic samples to deceive the discriminator, while the discriminator's task is to distinguish between the generator's output and real data. GAN architectures used in FR can be classified into three types: general GAN, pre-trained embedded GAN, and cyclic GAN. Non-blind methods primarily employ the general GAN structure depicted in Fig. 8 (a). In 2016, Yu *et al.* [7] introduced the first GAN-based face super-resolution network (URDGN). This network utilizes a discriminative network to learn fundamental facial features, and a generative network leverages adversarial learning to combine these features with the input face. Since then, many different GAN-based face



(a) General GAN architecture (b) Pre-trained embedding GAN architecture (c) The cyclic GAN architecture

Figure 8: Summary of architecture of GAN-based methods for face restoration.

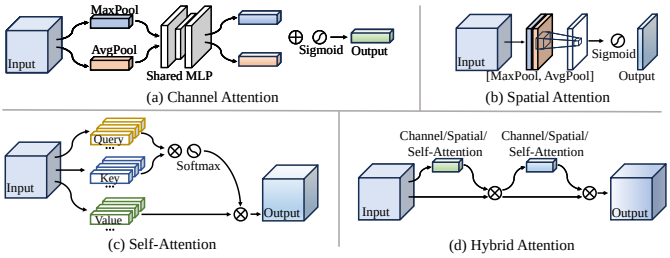


Figure 9: The architecture of Attention-based methods.

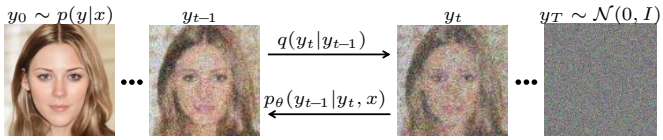


Figure 10: The diffusion denoising principle involves two main processes: forward diffusion process q and reverse inference process p. In the q, Gaussian noise is gradually added to the target image from left to right. In contrast, p iteratively denoises the target image, proceeding from right to left.

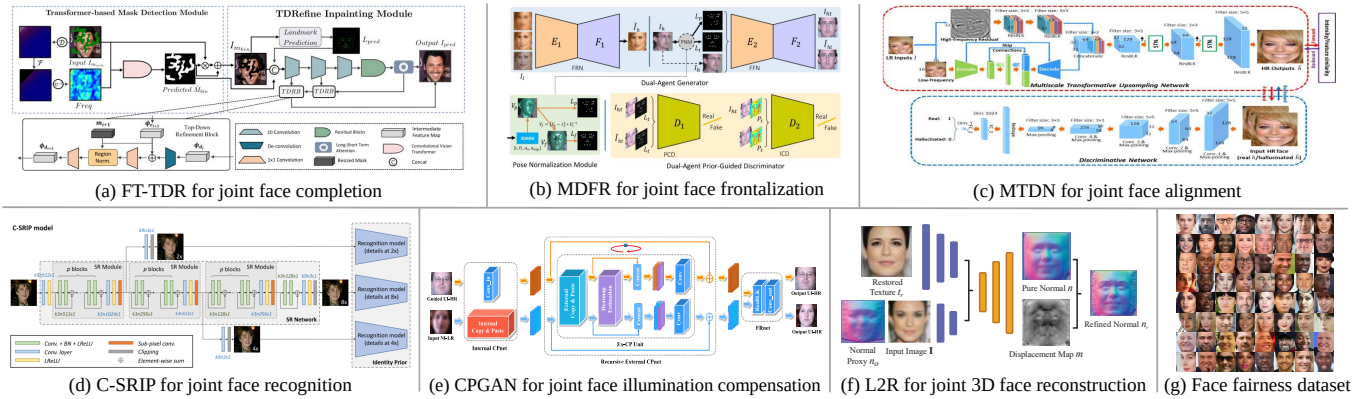
restoration methods have been extended in the non-blind task, showing promising recovery results. Some methods focus on designing progressive GANs, including two- or multi-stage approaches [63], [102], [103]. Others concentrate on embedding face-specific prior information, such as facial geometry [49], [104], [105], facial attributes [106], or identity information [107] into the GAN framework. It is worth noting that given the excellent performance of GAN, many non-GAN-based methods [5], [39], [41], [43], [47], [98] also provide a GAN version of their approach for reference.

However, GAN-driven methods often suffer from pattern collapse, resulting in a lack of diversity in the generated images. The diffusion probabilistic model (DDPM) have been proposed as an alternative approach. As shown in Fig. 10, given samples drawn from an unknown conditional distribution $p(y|x)$, the input-output image pair is denoted as $D = \{x_i, y_i\}$. DDPM learns the parameter approximations of $p(y|x)$ through a stochastic iterative refinement process that maps the source image x to the target image y. Specifically, DDPM starts with a purely noisy image $y_T \sim \mathcal{N}(0, I)$, and the model refines the image through successive iterations $(y_{T-1}, y_{T-2}, ..., y_0)$ based on the learned conditional transformation distribution $p_\theta(y_{t-1}|y_t, x)$, refining the image until $y_0 \sim p(y|x)$. In 2022, SRDiff [108] introduced a diffusion-based model for face super-resolution. It incorporated residual prediction throughout the framework to accelerate convergence. Then, SR3 [109] achieved super-resolution by iterative denoising the conditional images generated by the denoising diffusion probabilistic model, resulting in more realistic outputs at various magnification factors. IDM [110] combined an implicit neural representation with a denoising diffusion model. This allowed the model to continuous-resolution requirements and provide HQ face restoration with improved scalability across different scales.

3.2 Blind Tasks

In practical applications, researchers have observed that methods originally designed for non-blind tasks often struggle to effectively handle real-world LQ face images. Consequently, the focus of face restoration is gradually shifting towards blind tasks to address a broader range of application scenarios and challenges associated with LQ images. One of the earliest blind methods is DFD [111], introduced by G. Chrysos *et al.*, which employs a modified ResNet architecture for blind face deblurring. Then, MCGAN [112] leveraged GAN techniques to significantly improve the model's robustness in tackling blind deblurring tasks. However, this approach exhibits limited efficacy when encountering more complex forms of degradation. As a result, subsequent endeavors in the realm of blind tasks have predominantly employed GAN-driven methodologies. Some methods adopt the general GAN structure depicted in Fig. 8 (a). For example, DeblurGAN-v2 [113], HiFaceGAN [114], STUNet [66], GCFSR [72], and FaceFormer [115] all design novel and intricate network architectures for blind face restoration. Additionally, many methods use more complex GAN networks with prior information. GFRNet [9], ASFFNet [57] and DMDNet [67] utilize a bootstrap network with reference to prior to guide the recovery network, employing a two-stage strategy for better face restoration. MDCN [33] and PFSRGAN [71] employ a two-stage network consisting of a face semantic label prediction network and a recovery parsing network for reconstruction. Furthermore, Super-FAN [52], DFDNet [116], and RestoreFormer [23] integrate face structure information or face component dictionary into GAN-based algorithms to enhance the quality of blind LQ facial images.

Pre-trained GAN-based models have become the most popular approach in the field of blind face restoration since generative models [59], [119] can produce realistic and HQ face images. As shown in Fig. 8 (b), the pre-trained GAN embedding architecture involves adding an additional pre-trained generative GAN [59], [120] into the generator network. For example, GPEN [10] incorporates a pre-trained StyleGAN as a decoder within a U-network. It utilizes features extracted from the input by the decoder to refine the decoder's output, significantly improving restoration results compared to the general GAN structure. GFPGAN [6] goes a step further by integrating features from various scales within the encoder through spatial transformations into a pre-trained GAN employed as a decoder. Other networks, such as GLEAN [76], Panini-Net [121], SGPN [122], DEAR-



(a) FT-TDR for joint face completion (b) MDFR for joint face frontalization (c) MTDN for joint face alignment

(d) C-SRIP for joint face recognition (e) CPGAN for joint face illumination compensation (f) L2R for joint 3D face reconstruction (g) Face fairness dataset

Figure 11: Examples of methods for joint tasks. (a): FT-TDR [117] for joint face complementation; (b): MDFR [45] for joint face frontalization; (c): MTDN [118] for joint face alignment; (d): C-SRIP [27] for joint face recognition; (e): CPGAN [13] for joint face illumination complementation; (f): L2R [3] for joint 3D face reconstruction; (g): EDFace-Celeb-1M [65] dataset for joint face fairness.

GAN [123], DebiasSR [124], PDN [125], and others, also embrace this architecture. They incorporate a pre-trained StyleGAN or its variations into a GAN generator, complementing it with their individually crafted network architectures to cater to their specific application requirements. To further enhance the fidelity of the generated images, methods like VQFR [61], CodeFormer [126], and others employ pre-train VQGAN to enhance facial details. They achieve this by employing discrete feature codesets extracted from HQ face images as prior. The discrete codebook prior, acquired within a smaller agent space, significantly reduces uncertainty and ambiguity compared to the continuous StyleGAN prior.

Another category of blind methods focuses on addressing the challenge of obtaining paired LQ and HQ images in real-world scenarios. Inspired by CycleGAN [127], as shown in Fig. 8 (c), LRGAN [128] employs an cyclic GAN architecture consisting of two GAN networks. The initial high-to-low GAN generates LQ images that mimic real-world conditions and pairs them with corresponding HQ images. Subsequently, the second low-to-high GAN network is used to restore and enhance the quality of the generated LQ face images for restoration purposes. SCGAN [50] takes a step further by guiding the generation of paired LQ images through the creation of degenerate branches from HQ images. This approach further reduces the domain gap between the generated LQ and the authentic LQ images. Additionally, diffusion-denoising techniques for blind tasks aim to improve robustness in severely degraded scenarios when compared to non-blind tasks. DR2 [129] employs this technique to enhance the robustness of the blind restoration process and reduce artifacts often observed in the output face images. DDPM [130] refines the spatial content during backpropagation to improve the robustness and realism of the restoration in challenging scenarios. DIFFBFR [131] takes a different approach by initially restoring the LQ image and subsequently employing an LQ-independent unconditional diffusion model to refine the texture, rather than directly restoring the HQ image from a noisy input.

3.3 Joint Restoration Tasks

In this section, we will discuss some essential components of FR, which include joint face completion and restoration, joint face frontalization, joint face alignment, joint face recognition, joint face illumination compensation, joint 3D face reconstruction, and joint face fairness. And we have shown representative methods for each of them in Fig. 11.

• **Joint Face Completion.** It is an important branch of FR, as real-world captured face images may suffer from both blurring and occlusion. One class of methods focuses on normal-resolution complements. MLGN [132] and Swin-CasUNet [133] directly employ general networks for completion, but their fidelity is unsatisfactory. Given that accurately estimating occluded facial features is the key challenge in face completion, integrating prior information empowers models to infer critical details such as facial contours under occlusion. As a result, facial priors are extensively integrated into the majority of methods. For example, ID-GAN [134] uses facial identity, SwapInpaint [135] uses reference face, PFTANet [136] employs face semantic labels, FT-TDR [117] utilizes face landmarks, and others ([137]) uses face components. Another class of methods focuses on low-resolution face completion, where the initial methods [73], [138], [139] address occluded parts through patching first before performing restoration work. However, this type of method can result in a significant accumulation of errors in the final results. In contrast, MFG-GAN [140] utilizes graph convolution and customized loss functions to achieve end-to-end restoration. UR-GAN [64] utilizes landmarks guidance to progressively fix occluded and LQ faces.

• **Joint Face Frontalization.** Existing FR methods are primarily designed for frontal faces, and when applied to non-frontal faces, artifacts in the reconstructed results become evident. The first attempt to address this issue was made by TANN [74]. It utilized a discriminative network to enforce that the side-face generated face image should be close to the front-face image, aligning the faces in the same plane. Subsequently, VividGAN [141] employed a fractalization network combined with a fine feature network to further optimize the face details under fractalization. MDFR [45] introduced a 3D pose-based module to estimate the degree of face fractalization. It proposed a training strategy that integrates the recovery network with face fractalization end-to-end. Furthermore, inspired by the aforementioned methods, some approaches [142], [143] also combine the tasks of face completion and frontalization to address them jointly.

• **Joint Face Alignment.** Most FR methods require the use of aligned face training samples for optimal performance. Therefore, researchers have developed various methods for

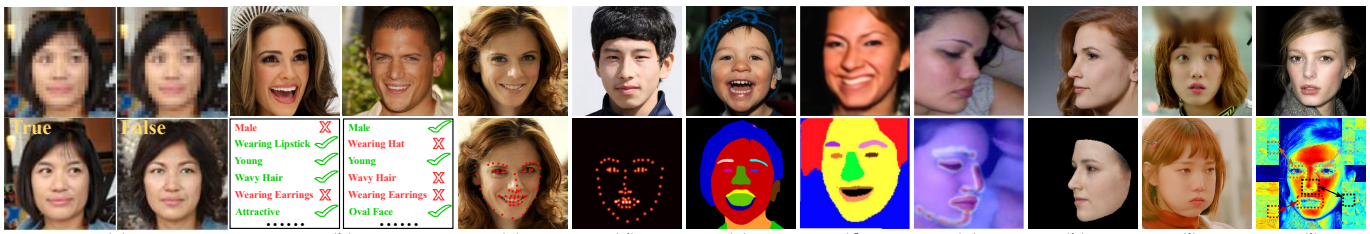


Figure 12: Visualization of popular Facial priors. (a) Facial identity; (b) Facial attributes; (c) Facial landmarks; (d) Facial heatmaps; (e) Facial parsing map; (f) Facial semantic labels; (g) Facial components; (h) 3D Face; (i) Facial reference priors; (j) Facial dictionary.

joint face alignment. Yu *et al.* were among the first to attempt embedding a spatial transformation layer as a generator and utilizing a discriminator to improve the alignment and upsampling. They developed TDN [11] and MDTN [118] using this approach. To handle possible noise in unaligned faces, they also developed a method [144] that incorporates downsampling and upsampling within the TDN framework to minimize the noise's impact. JASRNet [36] achieves quality alignment in parallel by supervising facial landmarks and HQ face images. Another approach [145] utilizes a face 3D dictionary alignment scheme to accomplish alignment.

• **Joint Face Recognition.** Some restoration methods [60], [146] may result in recovered face images that diverge from their original identities, making them unsuitable for downstream face recognition tasks. Since face recognition heavily relies on local features such as the eyes, many priors struggle to accurately emphasize these specific areas. One swift solution involves applying a pre-trained face recognition model after the restoration. This helps determine whether the restored face image aligns with the ground truth in terms of identity, enhancing restoration accuracy by incorporating identity-related prior knowledge. Some examples of these methods include SICNN [147], LRFR [28], and others [148], [149], [150]. C-SRIP [27] improves upon this approach by recovering multiple scales of face images through different branches and supervising the recovered face images at different scales using a pre-trained face recognition network. Furthermore, some methods, including SiGAN [151], FH-GAN [152], WaSRGAN [153], and others, further enhance performance by incorporating discriminators into the restoration process.

• **Joint Face Illumination Compensation.** Due to the unsatisfactory restoration performance of current algorithms on low-light LQ faces, this task has garnered significant attention. The main challenge in this task is detecting facial contours under low light conditions. As the first work, SeLENet [154] segments the input low-light face into human face normals and light coefficients. It then augments the existing lighting coefficients to complete the lighting compensation process. CPGAN [32] employs an internal CPNet to accomplish detail restoration from the input facial image. Additionally, it utilizes an external CPNet to compensate for background lighting using externally guided images. Furthermore, Zhang *et al.* [13] further improves CPGAN by introducing landmark constraints and recursive strategies. Ding *et al.* [155] employs a face localization network to detect facial landmarks, and then utilize these landmarks to better restore face contours and key features. Later, NASFE [156] introduces an automatic search strategy to discover an optimized network architecture specifically designed for the given task.

• **Joint 3D Face Reconstruction.** With the advancement in 3D technology, there has been growing interest in achieving 3D face reconstruction from LQ face images or recovering reconstructed LR 3D faces. R3DPFH [14] focused on predicting corresponding HQ 3D face meshes from LR faces containing noise. Utilizing the Lucas-Kanade algorithm, Qu *et al.* [157] aimed to improve the accuracy of 3D model fitting. Furthermore, Li *et al.* [158] and Uddin *et al.* [159] utilized techniques for 3D point clouds to infer HR mesh data from LQ or incomplete 3D face point clouds. In contrast to the aforementioned methods, L2R [3] directly reconstructed HQ faces from LQ faces by learning to recover fine-grained 3D details on the proxy image.

• **Joint Face Fairness.** Existing datasets often fail to adequately represent the distribution of human races, which can introduce biases towards specific racial groups in trained methods. One class of approaches focuses on algorithmic fairness by employing suitable algorithms to mitigate racial bias. Ajil *et al.* [15] define theoretical concepts of race fairness and implement their defined notion of conditional proportional representation through a posteriori sampling, which helps achieve fairer face restoration. Noam *et al.* [160] enhance the feature extractor to better capture facial features, attributes, and racial information, by incorporating multi-faceted constraints to reduce racial bias. Another class of approaches tackles the problem by building more ethnically balanced and comprehensive datasets. Zhang *et al.* [65] developed the EDFace-Celeb-1M dataset, which covers 1.7 million photographs from different countries with relatively balanced ethnicity. Subsequently, Zhang *et al.* [66] synthesized datasets for FR, namely EDFace-Celeb-1M and EDFace-Celeb-150K, which have made significant contributions to the progress of face fairness by providing more diverse and representative data.

4 Face Prior Technology

Considering the inherent structured attributes of faces, many methods in the aforementioned tasks have chosen to incorporate facial priors to enhance restoration outcomes. To provide a better understanding of the diverse roles played by these priors in face restoration, this section focuses on exploring the technology of facial priors. We present these priors in Fig. 12 for reference. Based on whether they additionally utilize the structural information of the external face, we categorize these priors into two classes: internal proprietary prior-based methods and external compensatory prior-based methods. A summary of representative methods can be found in TABLE 4. In the following sections, we will discuss these two classes of methods and their network structures in detail. It is worth noting that a few methods [122], [161] utilize both priors.

Table 4: An overview of representative blind / non-blind and joint face restoration methods based on deep learning techniques.

Methods	Publication	Prior	Task	Improved Technology	Highlight
BCCNN [77]	AAAI 2015	Plain	Non-Blind Task	Plain	Bi-channel CNN
ATMFN [91]	TMM 2019			GAN+Attention	Channel Attention Mechanism
SPARNet [47]	TIP 2020			Attention	Spatial Attention Mechanism
MSG-GAN [63]	CVPR 2020			GAN	Multi-Scale GAN
Faceformer [95]	TCSVT 2022			Attention	Self-Attention Mechanism
SCTANet [43]	TMM 2022			Attention	Self-attention / Spatial Attention
IDM [110]	CVPR 2023			Diffusion Model	Diffusion Probabilistic Models
LCGE [82]	IJCAI 2017	Internal Proprietary Prior		Prior	Facial Components Prior
AEDN [162]	CVPR 2018			GAN+Prior	Given Facial Attribute Prior
FSRNet [5]	CVPR 2018			Prior	Facial Landmarks / Parsing Maps Prior
FACN [163]	AAAI 2020			Prior	Estimated Facial Attribute Prior
DIC [41]	CVPR 2020			Attention+Prior	Facial Landmarks / Components Prior
PAP3D [55]	TPAMI 2021			Attention+Prior	3D Facial Prior / Spaital Attention
HCRF [164]	TIP 2021			Prior	Facial Semantic Labels Prior
GWAInet [58]	CVPR 2019	External Compensatory Prior		GAN+Prior	High Quality Images As Reference Prior
KDFSRNet [105]	TCSVT 2022			Prior	Pre-trained Teacher's Knowledge As Generative Prior
LRGAN [128]	ECCV 2018	Plain	Blind Task	GAN	Unsupervised / Two-Stage GAN
HiFaceGAN [114]	MM 2020			GAN	Semantic-Guided Generation
GCFSR [72]	CVPR 2022			GAN	Generative And Controllable Framework
SCGAN [50]	TIP 2023			GAN	Unsupervised / Semi-Cycled GAN
DR2 [129]	CVPR 2023			Diffusion Model	Diffusion-based Robust Degradation Remover
IDDM [165]	ICCV 2023			Diffusion Model	Iteratively Learned System
Super-FAN [52]	CVPR 2018	Internal Proprietary Prior		GAN+Prior	Facial Heatmap Prior
MDCN [33]	IJCV 2020			GAN+Prior	Semantic labels Prior
UMSN [87]	TIP 2020			GAN+Prior	Facial Components Prior
PSFRGAN [71]	CVPR 2021			GAN+Prior	Facial Parsing Maps Prior
SGPN [122]	CVPR 2022			GAN+Prior	3D Faical / Pre-trained Generative Prior
GFRNet [9]	ECCV 2018	External Proprietary Prior		GAN+Prior	High Quality Images As Reference Prior
PULSE [146]	CVPR 2020			GAN+Prior	Pre-trained StyleGAN's Generative Prior
ASFFNet [57]	CVPR 2020			GAN+Attention+Prior	Reference / Landmark Prior
DFDNet [116]	ECCV 2020			GAN+Prior	Faical Component Dictionaries Prior
GFPGAN [6]	CVPR 2021			GAN+Prior	Pre-trained StyleGAN's Generative Prior
DMDNet [67]	TPAMI 2022			GAN+Prior	Faical Component Dictionaries Prior / Reference Prior
RestoreFormer [23]	CVPR 2022			GAN+Attention+Prior	Faical Component Dictionaries
VQFR [61]	ECCV 2022			GAN+Prior	Pre-trained VQGAN's Codebook Prior
CodeFormer [126]	NIPS 2022			GAN+Prior	Pre-trained VQGAN's Codebook Prior
DebiasFR [124]	IJCAI 2023			GAN+Prior	Pre-trained StyleGAN's Generative Prior
TDN [11]	AAAI 2017	Plain	Joint Task	GAN	CNN / Joint Face Alignment
TANN [74]	TPAMI 2019			GAN	CNN / Joint Face Frontalization
MTDN [118]	IJCV 2020			GAN	CNN / Joint Face Alignment
EDFace-Celeb-1M [65]	TPAMI 2022			Plain	Dataset / Joint Face Fairness
SICNN [147]	ECCV 2018	Internal Proprietary Prior		Prior	Identity Prior / Joint Face Recognition
FCSR-GAN [73]	TBIOM 2020			GAN+Prior	Landmark/ Semantic labels / Joint Face Compensation
JASRNet [36]	AAAI 2020			Prior	Landmark Prior / Joint Face Alignment
ID-GAN [36]	TCSVT 2020			GAN+Prior	Semantic labels / Identity Prior / Joint Face Recognition
SiGAN [151]	TIP 2019			GAN+Prior	Identity Prior / Joint Face Recognition
MDFR [45]	TCSVT 2021			GAN+Prior	Landmark/ 3D Facial Prior / Joint Face Frontalization
FT-TDR [117]	TMM 2022			GAN+Attention+Prior	Landmark Prior/ Self-Attention / Joint Face Completion
L2R [3]	CVPR 2022			GAN+Prior	Generative / 3D Prior / Joint 3D Face Reconstruction
FVIP [166]	TIP 2023			GAN+Prior	3D Face Prior / Joint Face Completion
CPGAN [32]	CVPR 2020	External Proprietary Prior		GAN+Attention+Prior	Reference Prior / Joint Illumination Compensation
ViVidGAN [141]	TIP 2021			GAN+Attention+Prior	Reference Prior / Joint Illumination Compensation
IAPTU [141]	ICIP 2021			GAN+Prior	Pre-trained Generative Prior / Joint Face Fairness

4.1 Internal Proprietary Prior

This type of method primarily utilizes knowledge about the attributes and structural features inherent to the face itself. It incorporates information such as identity, facial features, and contours to guide the face restoration process. Common techniques employed in this approach include identity recognition, facial landmarks creation, semantic labeling maps, and more.

The first type of information used is the face's own 1D information, such as identity prior and attribute prior. Identity prior refers to information related to an individual's identity, indicating whether the restored face corresponds to the same person as the ground truth. Integrating identity prior to the restoration process enhances the model's ability to faithfully recover facial features. Methods based on identity prior, such as SICNN [147], FH-GAN [152], IPFH [12], C-SRIP [27], and others, aim to maintain identity consistency between the restored image and the HQ face image. During training, these frameworks typically include a restoration network and a pre-trained face recognition network. The face recognition network serves as an identity prior, determining whether the restored face belongs to the same identity as the HQ face, thereby improving the identity accuracy of the restored face. The face attribute prior provides 1D semantic information about the face for face restoration, such as attributes like long hair, age, and more. This prior aids the model in understanding and preserving specific facial characteristics during the restoration process. For instance, incorporating age attributes into the restoration process assists models in accurately preserving natural textures such as skin wrinkles. Earlier methods, such as EFSRSA [162], ATNet [167], ATSENet [84], AACNN [168], and others, directly connect the attribute information to the LQ image or its extracted features. Other methods, like AGCycleGAN [106] and FSRSA [162], use a discriminator to encourage the network to pay more attention to attribute features during restoration. However, these methods may experience significant performance degradation when attributes are missing. To address this

issue, attribute estimation methods [92], [163] have been proposed. These approaches design appropriate attribute-based losses that enable the network to adaptively predict attribute information. RAAN [92] utilizes three branches to separately predict face shape, texture, and attribute information. It emphasizes either face shape or texture based on the attribute channel. FACN [163] introduces the concept of capsules to enhance the recovered face. This is achieved by performing multiplication or addition operations between the face attribute mask estimated by the network and the semantic or probabilistic capsule obtained from the input.

Another class of methods emphasizes the use of the face's unique 2D geometric or 3D spatial information as priors. Facial landmarks [35], [41], [117] and facial heatmaps [99], [117], [161] are examples of these priors, representing coordinate points or probability density maps that indicate key facial components such as the eyes, nose, mouth, and chin. They provide accurate and detailed facial location information. Methods like DIC [41] utilize the predicted coordinates of facial landmarks from the prior estimation network to guide the restoration network. However, using a large number of facial landmarks may lead to error accumulation in coordinate estimation, particularly for severely degraded face images, resulting in distortion of the restored facial structure. In contrast, facial parsing maps [5], [49], [71] and facial semantic labels [33], [69], [87] are more robust to severe degradation as they segment the face into regions. Even if some regions are severely degraded, intact regions can still guide the restoration process. Moreover, these priors contain more comprehensive facial information, enabling the restoration model to better understand the overall facial structure and proportions, leading to more coherent restorations. However, these priors may involve multiple semantic labels for different facial regions, requiring more complex networks [71], [87] to address semantic ambiguity. On the other hand, facial components [86], [94], [137] provide a straightforward representation of critical facial features, reducing the need for complex models while effectively guiding the restoration process. In addition to the aforementioned 2D facial priors, Hu *et al.* [54] introduced the use of a 3D face prior to handle faces with large pose variations. Subsequent 3D prior-based methods [122], [166], [169] demonstrated their robustness in handling complex facial structures and significant pose changes. There are also methods [84], [136] that strive to achieve more comprehensive restoration by synergistically combining multiple internal proprietary priors.

4.2 External Compensatory Prior

Methods that leverage external priors primarily rely on externally guided faces or information sources derived from external HQ face datasets to facilitate the face restoration process. These external priors can take various forms, including reference priors, face dictionary priors, and pre-trained generative priors.

Reference prior-based methods [9], [57], [58] utilize HQ face images of the same individual as a reference to enhance the restoration of a target face image. The challenge lies in effectively handling reference faces with varying poses and lighting conditions. GFRNet [9] is the pioneering work in this field. It employs a sub-network called WarpNet, coupled with a landmark loss, to rectify pose and expression disparities present in the reference face. This enables the model to effectively utilize reference faces that exhibit differences compared to the face undergoing restoration. GWAInet [58] utilizes the structure of the generative network of the GAN and achieves favorable results without relying on facial landmarks. Subsequently, ASFFNet [57] further enhances performance by refining the selection of the guide face and improving the efficiency of feature fusion between the guide face and the image to be recovered.

However, the above methods require reference images for both training and inference, which limits their applicability in various scenarios. To address this limitation, DFDNet [116] employs a strategy that creates a facial component dictionary. Initially, a dictionary comprising facial elements such as eyes, nose, and mouth is categorized from an HQ face dataset. During the training phase, the network dynamically selects the most analogous features from the component dictionary to guide the reconstruction of corresponding facial parts. RestoreFormer [23] integrates Transformer architecture and leverages the face component loss to more effectively utilize the potential of the facial component dictionary. DMDNet [67] leverages external facial images as well as other images of the same individual to construct two distinct facial dictionaries. This process enables a gradual refinement from the external dictionary to the personalized dictionary, resulting in a coarse-to-fine bootstrapping approach.

Unlike face dictionary that requires manual separation of facial features, pre-trained face GAN models [59], [119], [120] can automatically extract information beyond facial features, including texture, hair details, and more. This makes approaches based on pre-trained generative priors simpler and more efficient. PULSE [146] is a pioneering breakthrough in FR that utilizes generative prior. It identifies the most relevant potential vectors in the pre-trained GAN feature domain for the input LQ face. Subsequently, mGANprior [60] enhances the PULSE method by incorporating multiple potential spatial vectors derived from the pre-trained GAN. However, these methods are complex and may struggle to ensure fidelity in restoration while effectively leveraging the input facial features. Approaches like GLEAN [170], GPEN [10], and GFPGAN [6] integrate a pre-trained GAN into their customized networks. They employ GAN's generative prior to guide the forward process of the network, effectively leveraging the input facial features and leading to improved fidelity in restoration. Subsequent techniques [121], [123], [124], [125], [171] aim to enhance the efficacy of pre-trained GAN priors by investigating optimal strategies for integrating pre-trained GANs with forward networks or exploring more efficient forward networks. SGPN [122] incorporates a 3D shape prior along with the generative prior to enhance restoration, combining both spatial and structural information. Apart from approaches based on pre-trained StyleGAN [59], [119], there is another category of methods built upon pre-trained VQGAN [120]. The key advantage of VQGAN lies in its utilization of a vector quantization mechanism, enabling accurate manipulation of specific features within the generated face images. Additionally, the training of it is more stable compared to some variants of StyleGAN. VQFR [61] leverages discrete codebook vectors from VQGAN, using optimally sized compression patches and a parallel de-

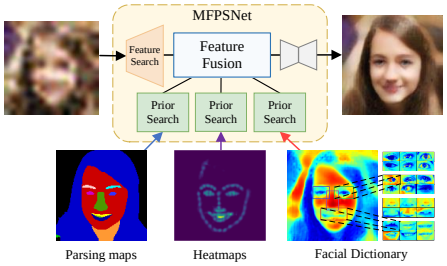


Figure 14: The representative method MFPSNet [161] for enhancing restoration process using multiple priors.

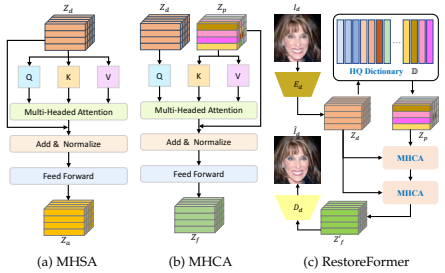


Figure 15: The representative method RestoreFormer [23] for designing networks that more efficiently utilize priors.

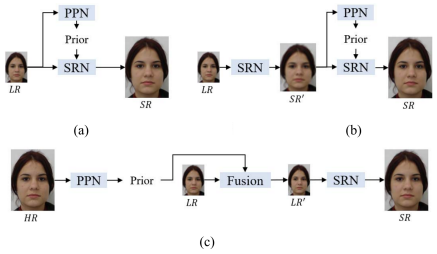


Figure 16: The method CHNet [94] for enhancing the effectiveness of priors by changing the prior guidance approach.

coder to improve detail and fidelity in the restored outcomes. Codeformer [126] integrates Transformer technology into its network architecture, achieving a favorable trade-off between quality and fidelity with a controlled feature conversion module. Zhao *et al.* [172] explores the utilization of pre-training priors, aiming to strike a harmonious equilibrium between generation and restoration aspects.

4.3 Advancing the Effectiveness of Prior

In this section, we will delve into approaches aimed at enhancing the effectiveness of prior knowledge for facial restoration. These approaches include combining multiple priors, developing efficient network structures, and adopting the prior guide approach.

• **Combining Multiple Priors.** Since different prior are suitable to different scenarios, the effectiveness of prior utilization diminishes significantly when inappropriate priors are used. To address this issue, some methods enhance the effectiveness of individual prior in facial restoration by incorporating multiple priors during the restoration process, leveraging the flexible complementarity of various prior information. Fig. 14 illustrates MFPSNet [161], which utilizes multiple priors including face parsing maps, face landmarks, and face dictionary to assist in restoration. Compared to approaches relying on a single prior, MFPSNet exhibits better robustness in highly blurry scenes. In general, some methods [5], [67], [84], [87] make use of either multiple internal proprietary priors or multiple external compensating priors. For example, UMSN [87] employs both face semantic labels and facial components as priors. DMDNet [67] utilizes both facial dictionaries and external reference faces. Additionally, some methods [6], [23], [122] combine internal proprietary priors with external compensating priors. For instance, SGPN [122] leverages a 3D face shape prior alongside a pre-trained GAN prior. However, employing an approach that utilizes multiple priors requires increased computational resources for prior estimation and often demands a larger dataset for modeling.

• **Efficient Network Structures.** Initial methods [5], [41] primarily focused on utilizing simple residual block structures for prior fusion, although these structures were not always optimal solutions. Subsequently, some methods [6], [57], [61], [126] aimed to design more efficient networks for prior fusion or estimation to enhance restoration performance. As depicted in Fig. 15, RestoreFormer [23] designs a custom multi-head cross-attention mechanism (MHCA) to comprehensively integrate facial dictionary information with facial features, showcasing significantly superior performance compared to multi-head self-attention (MHSA) alone. Similarly, ASFFNet [57] enhances the fusion of prior information with facial semantic features through a specially crafted adaptive spatial feature fusion block. VQFR [61] employs a parallel decoder structure to blend the generated prior information with low-level features, ensuring enhanced fidelity without compromising the quality of the prior guidance.

• **Prior Guide.** The way the prior is bootstrapped plays a crucial role in determining its effectiveness, as different bootstrapping methods yield varying restoration outcomes. For example, PFSRGAN [71] aims to enable the model to more effectively leverage the raw input information by directly estimating the prior knowledge from the LQ facial images to guide the restoration. In contrast, FSRNet [5] partially restores the LQ faces before estimating the prior to address inaccuracies in prior knowledge estimation. JASRNet [36] adopts a bootstrapping structure with parallel communication to fully leverage the interaction between prior estimation and restoration. Furthermore, as illustrated in Fig. 16, CHNet [94] modifies the process of estimating priors by opting to estimate them from HQ faces instead of directly or indirectly from LQ faces. For more comprehensive generalizations regarding the prior guide approach, please refer to the provided *supplementary material*.

5 Methods Analysis

In this section, we conducted a comprehensive evaluation of the key non-blind and blind face restoration methods. And due to the extensive range of joint tasks, their comparisons are in the *supplementary material*.

5.1 Experimental Setting

• **Non-blind Tasks.** We utilized the initial 18,000 images from CelebA dataset [46] for training purpuse. For testing, we randomly selected 1,000 images from CelebA dataset and 50 random images from Helen dataset [40]. All images were cropped and resized to a size of 128×128. The LQ images were derived by downsampling the HQ images using bicubic interpolation, as described in Eq. 3.

• **Blind Tasks.** We followed the degradation model used in GFPGAN [6] and conducted training and testing on the FFHQ and CelebA-HQ datasets, respectively. The degradation process is defined by Eq. 8 and Eq. 6, which represent blind restoration and blind super-resolution respectively. In these equations, the parameters σ, δ, r, and q of the degradation model are randomly drawn from the ranges $\{0.2 : 10\}$,

Methods (×8)	CelebA				Helen			
	PSNR↑	SSIM↑	LPIPS↓	FID↓	PSNR↑	SSIM↑	LPIPS↓	FID↓
RCAN [88]	27.45	0.7824	0.2205	174.0	25.50	0.7383	0.3437	219.3
FSRNet [5]	27.05	0.7714	0.2127	170.4	25.45	0.7364	0.3090	228.8
FACN [163]	27.22	0.7802	0.1828	167.7	25.06	0.7189	0.3113	218.0
SPARNet [47]	27.73	0.7949	0.1995	161.2	26.43	0.7839	0.2674	211.5
DIC [41]	-	-	-	-	26.15	0.7717	0.2158	214.1
SISN [38]	27.91	0.7971	0.2005	162.3	26.64	0.7908	0.2571	210.7
SwinIR [89]	27.88	0.7967	0.2001	163.2	26.53	0.7856	0.2644	213.2
CTCNet [39]	**28.37**	**0.8115**	**0.1702**	156.9	**27.08**	0.8007	0.2094	205.8
SCTANet [43]	28.26	0.8100	0.1710	156.8	27.01	**0.8068**	0.1901	203.3
SFMNet [48]	27.85	0.7967	0.1837	**156.5**	26.98	0.8049	**0.1865**	**199.5**
Input	23.61	0.6779	0.4899	362.2	22.95	0.6762	0.4912	289.1

Table 5: Performance comparison of key non-blind methods on CelebA and Helen Test Sets. In this paper, the best and the second best values are highlighted and underlined respectively.

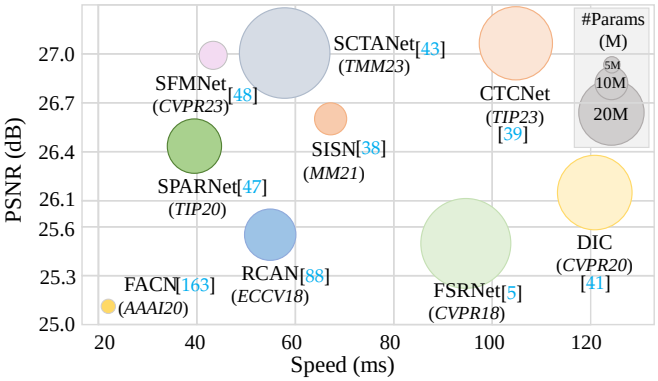


Figure 17: Complexity analysis of non-blind methods on Helen.

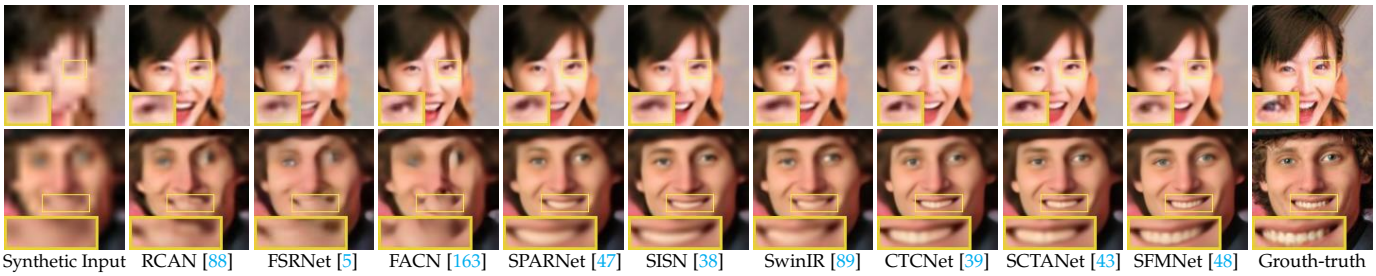


Figure 18: Visual comparison of different non-blind methods on the CelebA (first row) test set and Helen (second row) test set.

Method	RCAN [88]	FSRNet [5]	FACN [163]	SPARNet [47]
Params	15.7M	27.5M	4.4M	16.6M
MACs	4.7G	40.7G	12.5M	7.1G
Speed	56ms	89ms	22ms	40ms
DIC [41]	SISN [38]	CTCNet [39]	SCTANet [43]	SFMNet [48]
22.8M	9.8M	22.4M	27.7M	8.6M
35.5G	2.3G	47.2G	10.4G	30.6G
122ms	68ms	106ms	58ms	48ms

Table 6: Speed and overhead comparison of typical non-blind methods that measured on 128 × 128 images. We test all models using an NVIDIA RTX 3090 GPU.

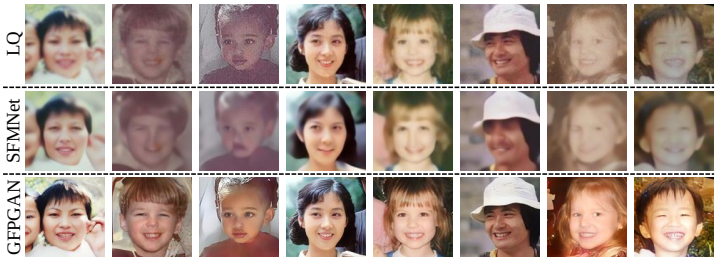


Figure 19: Comparison of non-blind/blind methods in reality.

$\{1:8\}$, $\{0:20\}$, and $\{60:100\}$, respectively. Furthermore, to ensure a more comprehensive evaluation, we incorporated real-world datasets such as LFW-Test, WebPhoto-Test, CelebChild, and CelebAdult. All images were aligned and resized to a size of 512×512.

• **Evaluation Metric.** We employed fully reference metrics, such as PSNR, SSIM, LPIPS, and IDD. These metrics assess various aspects including pixel structure similarity, visual fidelity, and identity preservation. In addition, we also utilized non-reference or semi-reference metrics like NIQE and FID. These metrics allow us to evaluate image fidelity and visual quality without the need for actual landmarks or reference images.

5.2 Quantitative Evaluation

Regarding the non-blind task, we chose to focus on evaluating non-blind super-resolution methods due to their predominant emphasis in the field. TABLE 5 presents a compilation of ten state-of-the-art non-blind methods, including fine-tuned image restoration methods [88], [89], methods based on attention mechanisms [38], [39], [43], [48], and methods relying on various priors [5], [41], [163]. Among these, methods employing hybrid attention mechanisms, namely CTCNet [39], SCTANet [43], and SFMNet [48], achieve either the best or second-best performance across all metrics on both test sets. TABLE 6 provides detailed information about the model characteristics of these methods, including parameters, computation, and inference duration. Furthermore, Fig. 17 visually illustrates the efficiency of these techniques through three perspectives: performance, inference speed, and model size. Notably, attention-based methods, particularly SFMNet [48], stand out as they achieve superior performance while maintaining smaller computational loads. Finally, Fig. 18 provides a visual comparison of these methods.

A range of state-of-the-art methods were selected for the blind task, including approaches that do not rely on prior such as network architecture design [72], [114] and diffusion modeling techniques [109]). Additionally, techniques utilizing internally-specific priors such as parsing maps [71] and 3D face shapes [122] were considered. Furthermore, methods employing external compensatory prior like pre-trained StyleGAN prior [6], [10], [122], pre-trained VQGAN prior [61], [126], face dictionary [23], [116], and reference prior [67]) were also included. To evaluate the application scope of non-blind and blind methods, we randomly selected several real face photos and restored them using the fine-tuned non-blind method SFMNet [48] and the blind method GFPGAN [6] respectively. As shown in Fig. 19, it is evident that SFMNet struggles to effectively handle real-world photos, while GFPGAN, despite showing some racial bias

Table 7: Comparison of primary blind method performance on synthetic test set CelebA-HQ and real datasets LFW-Test, WebPhoto, Celeb-Child, and Celeb-Adult. Speed and overhead comparison of typical non-blind methods that measured on 512×512 images.

Methods	Param	MACs	Speed	CelebA-HQ						LFW-Test		WebPhoto		CelebChild		CelebAdult	
				PSNR↑	SSIM↑	LPIPS↓	IDD↓	FID↓	NIQE↓	FID↓	NIQE↓	FID↓	NIQE↓	FID↓	NIQE↓	FID↓	NIQE↓
PSFRGAN [71]	60.2M	464.9G	53ms	24.65	.6443	.4199	.6664	43.33	4.099	49.53	4.095	84.98	4.151	106.6	4.670	104.1	4.246
HiFaceGAN [114]	79.9M	40.7G	90ms	24.92	.6195	.4770	.7310	66.09	5.002	64.50	4.520	116.1	4.943	113.0	4.871	104.0	4.340
DFDNet [116]	133.3M	599.8G	2.1s	24.26	.6042	.4421	.6884	54.34	5.921	59.69	4.776	93.28	5.812	107.1	4.452	105.6	3.782
GPEN [10]	71.1M	138.1G	235ms	25.59	.6894	.4009	.6019	**36.46**	5.364	57.00	5.071	101.3	6.326	112.1	4.945	110.8	4.362
GFPGAN [6]	48.7M	51.6G	**46ms**	25.08	.6777	.3646	.5709	42.59	4.158	50.04	3.965	87.13	4.228	111.4	4.447	105.0	4.033
VQFR [61]	71.8M	1.07T	495ms	24.14	.6360	.3515	.5959	41.29	**3.693**	50.65	**3.590**	**75.41**	**3.608**	105.2	**3.938**	105.0	**3.756**
GCFSR [72]	88.7M	119.8G	145ms	**26.31**	**.7085**	**.3400**	**.5122**	50.10	4.943	52.23	4.998	93.27	5.640	115.1	5.326	107.1	4.824
SGPN [122]	**15.2M**	**18.3G**	134ms	24.93	.6583	.3702	.6028	39.44	4.095	44.95	3.863	75.61	4.269	109.4	4.234	104.9	4.402
RestoreFormer [23]	72.4M	340.8G	172ms	24.64	.6060	.3655	.5339	41.82	4.405	48.38	4.169	77.33	4.459	**101.2**	4.580	**103.5**	4.321
CodeFormer [126]	73.6M	292.4G	98ms	25.15	.6699	.3432	.6171	52.43	4.650	52.36	4.484	83.19	4.705	116.2	4.983	111.1	4.541
DMDNet [67]	40.4M	187.2G	219ms	25.62	.6933	.3670	.6179	39.94	4.786	**43.38**	4.617	88.55	5.154	114.2	4.884	114.2	4.884
Input	-	-	-	23.35	.6848	.4866	.8577	144.0	13.2	137.6	11.0	170.1	12.7	144.4	9.03	118.3	7.56

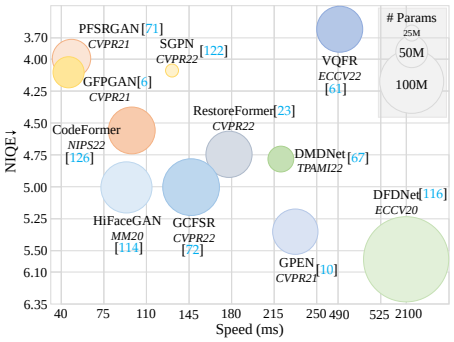


Figure 21: Complexity analysis of blind methods on synthetic test set CelebA-HQ.

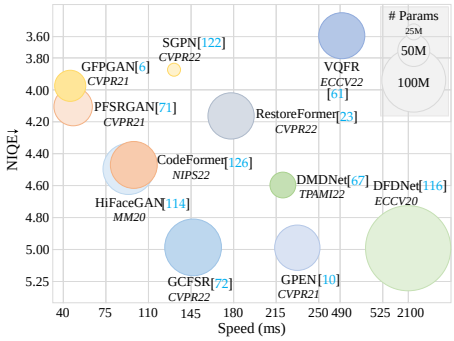


Figure 22: Complexity analysis of blind methods on the real test set LFW-Test.

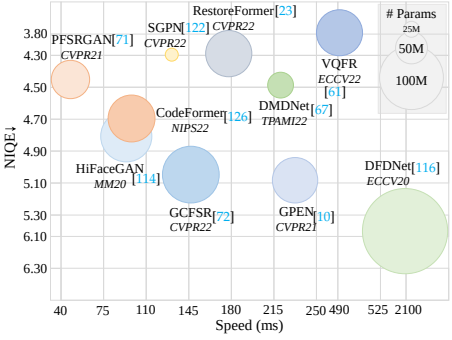


Figure 23: Complexity analysis of blind FSR methods on CelebA-HQ ($\times 8$).

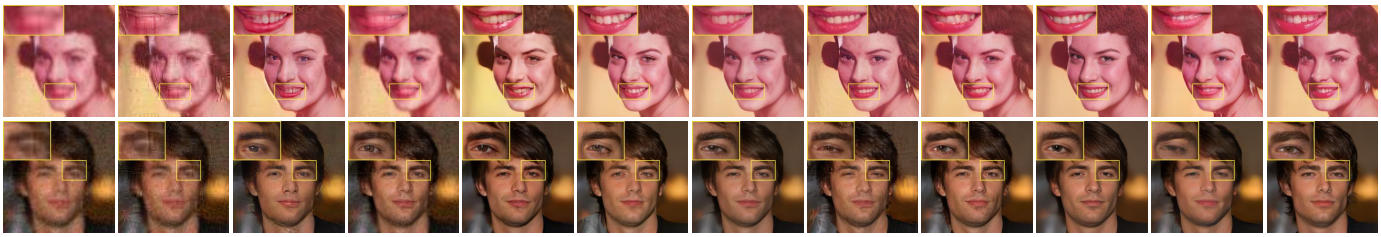


Figure 24: Visual comparison of different blind methods on the CelebA-HQ test set for blind face restoration.

in certain images, generally offers superior visual quality. Consequently, the blind method holds greater promise for real-world applications.

In the context of the blind task, our evaluation primarily focuses on blind face restoration, as blind methods primarily emphasize this specific direction. We also complement the evaluation with blind super-resolution. TABLE 7 presents a comprehensive quantitative assessment of these techniques across three dimensions: model size, inference speed, and performance on synthetic and real datasets. It can be observed that GCFSR achieves the best performance in several metrics that measure the structural similarity of restored face images. In terms of image fidelity and perceptual quality, pre-trained GAN-based methods, with VQFR [61] being a notable representative, exhibit superior performance. Methods such as DMDNet [23], SGPN [122], and GPEN [10] strike a better balance between structural similarity and perceptual quality. Furthermore, to handle more complex degradation, blind methods tend to employ larger models compared to non-blind approaches, resulting in slower inference times. Fig. 21 and Fig. 22 illustrate the efficiency trade-offs of these methods on the synthetic and real test sets, respectively. In these figures, methods closer to the upper-left corner with smaller circles are considered more efficient. The figure demonstrate that both PFSRGAN [71] and pre-trained GAN-based methods [6], [122], [126] are more efficient. Comparative visualization of their visual effects can be observed in Fig. 24 and Fig. 25. It's noticeable that methods [6], [61], [126] relying on pre-trained GAN priors tend to achieve superior performance when dealing with severely degraded facial images. Finally, as depicted in Fig. 26, we have selected four metrics - SSIM for face similarity, IDD for identity consistency, LPIPS for sensory quality assessment, and FID for image fidelity - to highlight the strengths and weaknesses of each method in terms of face image quality. It is evident that some methods [23], [122], while exhibiting better sensory quality, show subpar performance in two metrics, such as SSIM and IDD. On the other hand, methods [72], [126] with higher structural and identity similarity often display inferior sensory quality. Therefore, there is a clear need for the development of more balanced approaches to address these disparities.

The second part focuses on blind super-resolution, and TABLE 8 provides a comprehensive quantitative performance comparison of these methods at three scales: $\times 4$, $\times 8$, and $\times 16$. It is apparent that priori-free methods like GCFSR [72]

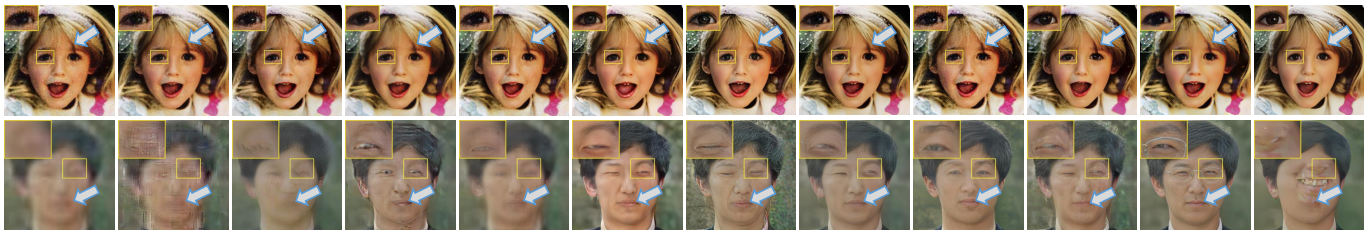


Figure 25: Qualitative comparison of restoration for the real test sets, including Celeb-Child (first), and Celeb-Adult (second row).

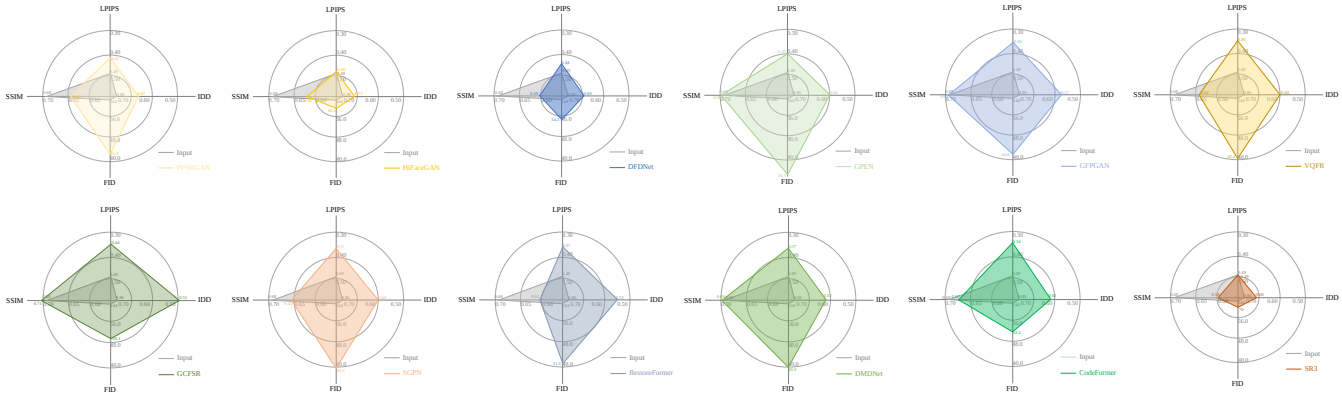


Figure 26: Balanced analysis of various blind methods across four major metrics: SSIM, LPIPS, FID, and IDD.

and HiFaceGAN [114] excel in face structure restoration. However, they exhibit shortcomings in FID and NIQE metrics, suggesting that their restored images might lack realism and may contain artifacts. On the other hand, pre-trained GAN-based approaches such as GPEN [10], VQFR [61], and SGPN [122] perform better in these two metrics, indicating more realistic and artifact-free results. Moving forward, Fig. 23 illustrates the efficiency of methods at the $\times 8$ scale, with PFSRGAN [71] and SGPN [122] emerging as the more efficient choices. Lastly, in Fig. 27, we present a visual comparison of methods at three scales. Notably, SGPN [122] and CodeFormer [126], leveraging pre-trained GAN priors, perform favorably without introducing artifacts when dealing with substantial downsampling factors.

6 Challenge And Future Directions

After reviewing various tasks and techniques and evaluating some prominent methods, it is clear that significant progress has been made. However, several challenges still persist in this domain. Additionally, there are numerous promising research opportunities to tackle these challenges and further advance the field of facial restoration.

• **Unified Large Model.** Prominent advancements in macro-modeling, exemplified by techniques such as Generative Pre-Training (GPT) and the Segment Anything Model (SAM), have had a significant impact on the field of computer vision. However, existing face restoration techniques often have a limited scope. Most models are designed to address specific challenges such as super-resolution or deblurring, or they focus on a single joint task. Consequently, there is a pressing demand in the industry for comprehensive large-scale models that are capable of restoring a wide spectrum of degraded facial images.

• **Multimodal Technology.** The successful utilization of GPT-4 in integrating images and text opens up new possibilities. For example, linguistic commands can be input to achieve selective restoration of features such as hair, eyes, and skin. Language-based instructions can also be employed to achieve specific restoration effects, such as emphasizing high resolution or maintaining identity resemblance. However, current models face challenges in precisely controlling these factors due to a lack of interpretability or handling intersectionality across different domains. As a result, the interpretability of FR models and their application in multimodal tasks could emerge as significant research areas.

• **Face Fairness.** The majority of FR datasets, such as CelebA and FFHQ, primarily collect facial images from specific geographical regions. It leads to the current trained models focusing on recovering facial features that are typical of those specific regions, while potentially disregarding distinctions in facial characteristics across various areas, such as variations in skin color. As a result, restoration results for individuals with black or yellow skin tones may inadvertently exhibit features characteristic of white individuals. Addressing this challenge requires the development of algorithms that mitigate racial bias in FR or the creation of datasets that prioritize racial balance.

• **Face Privacy Protection.** With the widespread use of facial recognition technology, improving recognition accuracy in specific scenarios (such as low light or blur) is closely linked to face restoration techniques. However, during the process of repairing and recognizing faces, there is a risk of facial information leakage. This highly sensitive data is closely associated with financial transactions and access permissions. Unfortunately, current FR methods often ignore this aspect. Therefore, ensuring the protection of facial privacy during restoration remains an ongoing challenge and opportunity.

• **Real-world Applications.** The challenges faced by facial restoration applications are two-fold: the disparity between synthetic and real data domains, and the significant com-

Table 8: Quantitative comparisons with primary blind methods on CelebA-HQ for $\times 4$, $\times 8$, $\times 16$ super-resolution.

Methods	CelebA-HQ ($\times 4$)						CelebA-HQ ($\times 8$)						CelebA-HQ ($\times 16$)					
	PSNR↑	SSIM↑	LPIPS↓	IDD↓	FID↓	NIQE↓	PSNR↑	SSIM↑	LPIPS↓	IDD↓	FID↓	NIQE↓	PSNR↑	SSIM↑	LPIPS↓	IDD↓	FID↓	NIQE↓
PSFRGAN [71]	27.99	.7777	.3055	.2924	42.35	4.623	25.50	.6921	.3639	.4445	47.56	4.446	23.20	.6250	.4216	.8603	49.31	4.197
HiFaceGAN [114]	29.49	.8030	.2736	**.2065**	39.72	4.535	**26.76**	.7156	.3496	**.3704**	51.32	4.830	23.68	.6179	.4746	1.012	92.31	5.836
DFDNet [116]	27.47	.7542	.3108	.2888	41.26	4.710	25.26	.6336	.3982	.4097	45.58	6.054	23.24	.5768	.4713	.9003	60.06	7.070
GPEN [10]	28.35	.7974	.2600	.2972	47.83	4.603	26.60	**.7359**	.3193	.4052	54.17	5.086	**24.12**	**.6772**	.3950	.8329	68.35	5.896
VQFR [61]	26.29	.7201	.2989	.3654	43.98	**3.884**	24.84	.6657	.3287	.4600	45.72	**3.826**	22.17	.5853	.3761	.8128	**38.42**	**3.431**
GCFSR [72]	**30.73**	**.8383**	**.2369**	.2132	52.02	4.915	26.66	.7299	.3073	.4146	54.74	5.059	22.90	.6371	.3799	.8564	46.99	4.622
SGPN [122]	28.64	.8040	.2456	.2581	41.06	4.425	26.18	.7127	**.3033**	.3846	44.69	4.330	23.65	.6487	**.3602**	**.7506**	46.66	4.444
RestoreFormer [23]	24.99	.6669	.3353	.4146	41.38	4.392	24.65	.6560	.3495	.4525	41.66	4.340	22.60	.5944	.3974	.8068	38.45	4.216
CodeFormer [126]	27.10	.7465	.3020	.4462	51.30	4.739	25.75	.6947	.3229	.5115	51.42	4.698	23.26	.6260	.3666	.7776	48.69	4.496
DMDNet [67]	28.43	.8081	.2724	.3080	**39.06**	4.652	26.31	.7208	.3292	.3967	**41.49**	4.576	22.91	.6327	.3890	.8318	39.61	4.358
SR3 [109]	25.02	.4797	.4998	.3968	52.78	9.366	23.14	.4781	.4826	.4735	61.68	7.614	22.11	.4057	.6377	1.139	111.2	13.78
Input	31.05	.8488	.2425	.2216	107.0	8.155	27.51	.7585	.3748	.5898	195.7	11.24	24.21	.6815	.5007	1.076	163.7	13.47

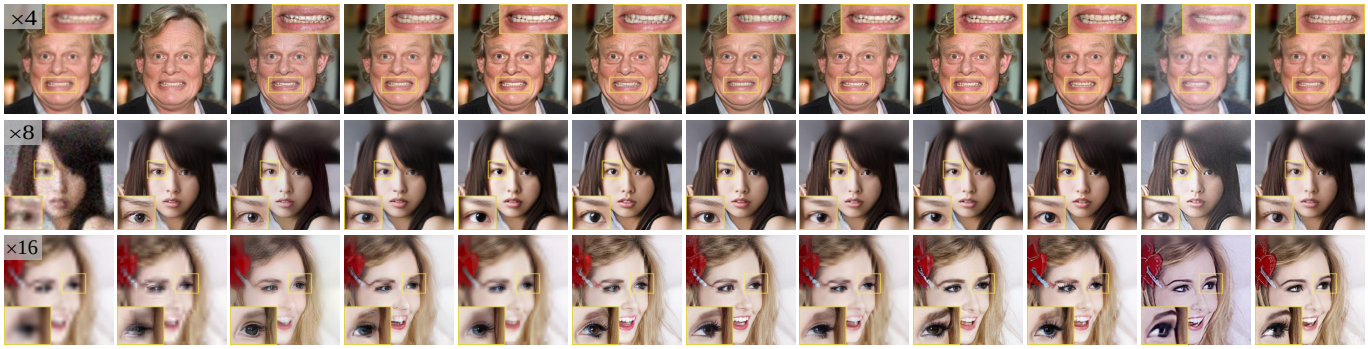


Figure 27: Qualitative comparisons on CelebA-HQ for $\times 4$ (first row), $\times 8$ (second row), $\times 16$ (third row) super-resolution.

putational costs. The domain difference is evident in the fact that real-world images undergo more complex forms of degradation compared to synthetic counterparts, resulting in persistent artifacts after applying existing restoration techniques. Additionally, the computational overhead of current methods is excessive for deployment on mobile devices, limiting their scalability. To address these challenges, research efforts should focus on developing realistic image degradation models to capture the complexities of real-world degradation, exploring unsupervised restoration approaches to alleviate the reliance on large annotated datasets, and investigating model compression and acceleration techniques to reduce computational costs. These endeavors will contribute to the advancement of applications related to video face restoration and the restoration of aged photographs, ultimately enhancing their practicality and usability.

• **Effective Benchmarks.** Several commonly used benchmarks in current face restoration, including datasets, loss functions, baseline network architectures, and evaluation metrics, may not provide optimal solutions. For example, some datasets may lack comprehensive coverage, leading to limited generalization of the models. Flawed loss functions may result in undesired artifacts in the restored faces. Existing network architectures may not be suitable for all restoration tasks, limiting their applicability. Additionally, evaluating restoration results solely based on quantitative metrics may overlook important aspects of human perceptual quality. Ongoing research efforts are actively addressing these issues, leading to improvements in various areas of face restoration. However, these concerns remain focal points for future investigations.

7 Conclusion

In this review, we provide a systematic exploration of deep learning-based approaches for face restoration. We begin by discussing the factors that contribute to the degradation of facial images and artificial degradation processes. Subsequently, we categorize the field into three distinct task categories: non-blind, blind, and joint tasks, and discuss their evolution and technical characteristics. Furthermore, we shed light on prevailing methodologies that utilize facial priors, including both internal proprietary and external compensatory priors. And we summarize the prevalent strategies for enhancing the effectiveness of priors in face restoration. Then, We conduct a thorough comparison of cutting-edge methods, highlighting their respective strengths and weaknesses. Finally, we dissect the prevailing challenges within the existing paradigms and provide insights into potential directions for advancing the field. Overall, This comprehensive review aims to serve as a valuable reference for researchers who are starting their journey in developing techniques aligned with their research aspirations.

8 Acknowledgements

This work was supported by China Postdoctoral Science Foundation under Grant 2022M720517 and National Natural Science Foundation of China under Grant 62306043.

References

[1] 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, pp. 5525–5533.

[2] 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 *Proceedings of Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, 2008.

[3] Z. Zhang, Y. Ge, Y. Tai, X. Huang, C. Wang, H. Tang, D. Huang, and Z. Xie, "Learning to restore 3d face from in-the-wild degraded images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2022, pp. 4237–4247.

[4] S. Baker and T. Kanade, "Hallucinating faces," in *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition*. IEEE, 2000, pp. 83–88.

[5] 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, pp. 2492–2501.

[6] X. Wang, Y. Li, H. Zhang, and Y. Shan, "Towards real-world blind face restoration with generative facial prior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021, pp. 9168–9178.

[7] X. Yu and F. Porikli, "Ultra-resolving face images by discriminative generative networks," in *Proceedings of the European Conference on Computer vVision*. Springer, 2016, pp. 318–333.

[8] J. Jiang, Y. Yu, J. Hu, S. Tang, and J. Ma, "Deep cnn denoiser and multi-layer neighbor component embedding for face hallucination," *Proceedings of the International Joint Conference on Artificial Intelligence*, 2018.

[9] 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 vVision*, 2018, pp. 272–289.

[10] T. Yang, P. Ren, X. Xie, and L. Zhang, "Gan prior embedded network for blind face restoration in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021, pp. 672–681.

[11] X. Yu and F. Porikli, "Face hallucination with tiny unaligned images by transformative discriminative neural networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, 2017.

[12] X. Cheng, J. Lu, B. Yuan, and J. Zhou, "Identity-preserving face hallucination via deep reinforcement learning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 12, pp. 4796–4809, 2019.

[13] Y. Zhang, I. W. Tsang, Y. Luo, C. Hu, X. Lu, and X. Yu, "Recursive copy and paste gan: Face hallucination from shaded thumbnails," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 8, pp. 4321–4338, 2021.

[14] Y. Zhong, Y. Pei, P. Li, Y. Guo, G. Ma, M. Liu, W. Bai, W. Wu, and H. Zha, "Face denoising and 3d reconstruction from a single depth image," in *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*. IEEE, 2020, pp. 117–124.

[15] A. Jalal, S. Karmalkar, J. Hoffmann, A. Dimakis, and E. Price, "Fairness for image generation with uncertain sensitive attributes," in *Proceedings of the International Conference on Machine Learning*, 2021, pp. 4721–4732.

[16] H. Liu, X. Zheng, J. Han, Y. Chu, and T. Tao, "Survey on gan-based face hallucination with its model development," *IET Image Processing*, vol. 13, no. 14, pp. 2662–2672, 2019.

[17] J. Jiang, C. Wang, X. Liu, and J. Ma, "Deep learning-based face super-resolution: A survey," *ACM Computing Surveys*, vol. 55, no. 1, pp. 1–36, 2021.

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

[19] A. Hore and D. Ziou, "Image quality metrics: Psnr vs. ssim," in *Proceedings of the International Conference on Pattern Recognition*. IEEE, 2010, pp. 2366–2369.

[20] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[21] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proceedings of the Asilomar Conference on Signals, Systems & Computers*, vol. 2. Ieee, 2003, pp. 1398–1402.

[22] 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, pp. 586–595.

[23] Z. Wang, J. Zhang, R. Chen, W. Wang, and P. Luo, "Restoreformer: High-quality blind face restoration from undegraded key-value pairs," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17 512–17 521.

[24] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

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

[26] T. Hoßfeld, P. E. Heegaard, M. Varela, and S. Möller, "Qoe beyond the mos: an in-depth look at qoe via better metrics and their relation to mos," *Quality and User Experience*, vol. 1, pp. 1–23, 2016.

[27] K. Grm, W. J. Scheirer, and V. Štruc, "Face hallucination using cascaded super-resolution and identity priors," *IEEE Transactions on Image Processing*, vol. 29, pp. 2150–2165, 2019.

[28] S.-C. Lai, C.-H. He, and K.-M. Lam, "Low-resolution face recognition based on identity-preserved face hallucination," in *Proceedings of the IEEE International Conference on Image Processing*. IEEE, 2019, pp. 1173–1177.

[29] 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*, vol. 33, no. 1, pp. 378–391, 2020.

[30] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-pie," *Image and Vision Computing*, vol. 28, no. 5, pp. 807–813, 2010.

[31] Y. Song, J. Zhang, L. Gong, S. He, L. Bao, J. Pan, Q. Yang, and M.-H. Yang, "Joint face hallucination and deblurring via structure generation and detail enhancement," *International journal of computer vision*, vol. 127, pp. 785–800, 2019.

[32] 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 Conference on Computer Vision and Pattern Recognition*, 2020, pp. 7355–7364.

[33] Z. Shen, W.-S. Lai, T. Xu, J. Kautz, and M.-H. Yang, "Exploiting semantics for face image deblurring," *International Journal of Computer Vision*, vol. 128, pp. 1829–1846, 2020.

[34] 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 *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2011, pp. 2144–2151.

[35] 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.

[36] Y. Yin, J. Robinson, Y. Zhang, and Y. Fu, "Joint super-resolution and alignment of tiny faces," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, 2020, pp. 12 693–12 700.

[37] M. Grgic, K. Delac, and S. Grgic, "Scface–surveillance cameras face database," *Multimedia Tools and Applications*, vol. 51, pp. 863–879, 2011.

[38] T. Lu, Y. Wang, Y. Zhang, Y. Wang, L. Wei, Z. Wang, and J. Jiang, "Face hallucination via split-attention in split-attention network," in *Proceedings of the ACM International Conference on Multimedia*, 2021, pp. 5501–5509.

[39] 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*, vol. 32, pp. 1978–1991, 2023.

[40] V. Le, J. Brandt, Z. Lin, L. Bourdev, and T. S. Huang, "Interactive facial feature localization," in *Proceedings of the European Conference on Computer Vision*, 2012, pp. 679–692.

[41] C. Ma, Z. Jiang, Y. Rao, J. Lu, and J. Zhou, "Deep face super-resolution with iterative collaboration between attentive recovery and landmark estimation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2020, pp. 5569–5578.

[42] T. Zhao and C. Zhang, "Saan: Semantic attention adaptation network for face super-resolution," in *Proceedings of the IEEE International Conference on Multimedia and Expo*. IEEE, 2020, pp. 1–6.

[43] Q. Bao, Y. Liu, B. Gang, W. Yang, and Q. Liao, "Sctanet: A spatial attention-guided cnn-transformer aggregation network for deep face image super-resolution," *IEEE Transactions on Multimedia*, 2023.

[44] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," *arXiv preprint arXiv:1411.7923*, 2014.

[45] X. Tu, J. Zhao, Q. Liu, W. Ai, G. Guo, Z. Li, W. Liu, and J. Feng, "Joint face image restoration and frontalization for recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 3, pp. 1285–1298, 2021.

[46] 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, pp. 3730–3738.

[47] 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*, vol. 30, pp. 1219–1231, 2020.

[48] C. Wang, J. Jiang, Z. Zhong, and X. Liu, "Spatial-frequency mutual learning for face super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2023, pp. 22 356–22 366.

[49] X. Yu, L. Zhang, and W. Xie, "Semantic-driven face hallucination based on residual network," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 3, no. 2, pp. 214–228, 2021.

[50] 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*, vol. 32, pp. 1184–1199, 2023.

[51] 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, pp. 1021–1030.

[52] ——, "Super-fan: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with gans," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 109–117.

[53] S. Zafeiriou, G. Trigeorgis, G. Chrysos, J. Deng, and J. Shen, "The menpo facial landmark localisation challenge: A step towards the solution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 170–179.

[54] 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 *Proceedings of the European Conference on Computer Vision*. Springer, 2020, pp. 763–780.

[55] X. Hu, W. Ren, J. Yang, X. Cao, D. Wipf, B. Menze, X. Tong, and H. Zha, "Face restoration via plug-and-play 3d facial priors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 12, pp. 8910–8926, 2021.

[56] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," in *Proceedings of the IEEE International Conference on Automatic Face & Gesture Recognition*. IEEE, 2018, pp. 67–74.

[57] X. Li, W. Li, D. Ren, H. Zhang, M. Wang, and W. Zuo, "Enhanced blind face restoration with multi-exemplar images and adaptive spatial feature fusion," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2706–2715.

[58] B. Dogan, S. Gu, and R. Timofte, "Exemplar guided face image super-resolution without facial landmarks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0–0.

[59] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4401–4410.

[60] J. Gu, Y. Shen, and B. Zhou, "Image processing using multi-code gan prior," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2020, pp. 3012–3021.

[61] Y. Gu, X. Wang, L. Xie, C. Dong, G. Li, Y. Shan, and M.-M. Cheng, "Vqfr: Blind face restoration with vector-quantized dictionary and parallel decoder," in *Proceedings of the European Conference on Computer Vision*. Springer, 2022, pp. 126–143.

[62] C.-H. Lee, Z. Liu, L. Wu, and P. Luo, "Maskgan: Towards diverse and interactive facial image manipulation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 5549–5558.

[63] A. Karnewar and O. Wang, "Msg-gan: Multi-scale gradients for generative adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 7799–7808.

[64] Y. Zhang, X. Yu, X. Lu, and P. Liu, "Pro-uigan: Progressive face hallucination from occluded thumbnails," *IEEE Transactions on Image Processing*, vol. 31, pp. 3236–3250, 2022.

[65] K. Zhang, D. Li, W. Luo, J. Liu, J. Deng, W. Liu, and S. Zafeiriou, "Edface-celeb-1 m: Benchmarking face hallucination with a million-scale dataset," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[66] P. Zhang, K. Zhang, W. Luo, C. Li, and G. Wang, "Blind face restoration: Benchmark datasets and a baseline model," *arXiv preprint arXiv:2206.03697*, 2022.

[67] X. Li, S. Zhang, S. Zhou, L. Zhang, and W. Zuo, "Learning dual memory dictionaries for blind face restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[68] S. Cheng, Y. Wang, H. Huang, D. Liu, H. Fan, and S. Liu, "Nbnet: Noise basis learning for image denoising with subspace projection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2021, pp. 4896–4906.

[69] Z. Shen, W.-S. Lai, T. Xu, J. Kautz, and M.-H. Yang, "Deep semantic face deblurring," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8260–8269.

[70] W.-S. Lai, Y. Shih, L.-C. Chu, X. Wu, S.-F. Tsai, M. Krainin, D. Sun, and C.-K. Liang, "Face deblurring using dual camera fusion on mobile phones," *ACM Transactions on Graphics*, vol. 41, no. 4, pp. 1–16, 2022.

[71] C. Chen, X. Li, L. Yang, X. Lin, L. Zhang, and K.-Y. K. Wong, "Progressive semantic-aware style transformation for blind face restoration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021, pp. 11 896–11 905.

[72] 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 Conference on Computer Vision and Pattern Recognition*, 2022, pp. 1889–1898.

[73] 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*, vol. 2, no. 2, pp. 109–121, 2019.

[74] X. Yu, F. Shiri, B. Ghanem, and F. Porikli, "Can we see more? joint frontalization and hallucination of unaligned tiny faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 9, pp. 2148–2164, 2019.

[75] R. Kalarot, T. Li, and F. Porikli, "Component attention guided face super-resolution network: Cagface," in *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*, 2020, pp. 370–380.

[76] K. C. Chan, X. Wang, X. Xu, J. Gu, and C. C. Loy, "Glean: Generative latent bank for large-factor image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021, pp. 14 245–14 254.

[77] 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.

[78] Z. Feng, J. Lai, X. Xie, D. Yang, and L. Mei, "Face hallucination by deep traversal network," in *Proceedings of the International Conference on Pattern Recognition*. IEEE, 2016, pp. 3276–3281.

[79] T. Lu, H. Wang, Z. Xiong, J. Jiang, Y. Zhang, H. Zhou, and Z. Wang, "Face hallucination using region-based deep convolutional networks," in *Proceedings of the IEEE International Conference on Image Processing*. IEEE, 2017, pp. 1657–1661.

[80] Z. Chen, J. Lin, T. Zhou, and F. Wu, "Sequential gating ensemble network for noise robust multiscale face restoration," *IEEE Transactions on Cybernetics*, vol. 51, no. 1, pp. 451–461, 2019.

[81] S. Zhu, S. Liu, C. C. Loy, and X. Tang, "Deep cascaded bi-network for face hallucination," in *Proceedings of the European Conference on Computer Vision*. Springer, 2016, pp. 614–630.

[82] Y. Song, J. Zhang, S. He, L. Bao, and Q. Yang, "Learning to hallucinate face images via component generation and enhancement," *arXiv preprint arXiv:1708.00223*, 2017.

[83] F. Cheng, T. Lu, Y. Wang, and Y. Zhang, "Face super-resolution through dual-identity constraint," in *Proceedings of the IEEE International Conference on Multimedia and Expo*. IEEE, 2021, pp. 1–6.

[84] 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*, vol. 23, pp. 468–483, 2020.

[85] K. Li, B. Bare, B. Yan, B. Feng, and C. Yao, "Face hallucination based on key parts enhancement," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2018, pp. 1378–1382.

[86] 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*, 2018, pp. 217–233.

[87] R. Yasarla, F. Perazzi, and V. M. Patel, "Deblurring face images using uncertainty guided multi-stream semantic networks," *IEEE Transactions on Image Processing*, vol. 29, pp. 6251–6263, 2020.

[88] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks,"

in *Proceedings of the European Conference on Computer Vision*, 2018, pp. 286–301.

[89] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "Swinir: Image restoration using swin transformer," in *Proceedings of the IEEE International Conference on Computer Vision*, 2021, pp. 1833–1844.

[90] Q. Cao, L. Lin, Y. Shi, X. Liang, and G. Li, "Attention-aware face hallucination via deep reinforcement learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 690–698.

[91] K. Jiang, Z. Wang, P. Yi, G. Wang, K. Gu, and J. Jiang, "Atmfn: Adaptive-threshold-based multi-model fusion network for compressed face hallucination," *IEEE Transactions on Multimedia*, vol. 22, no. 10, pp. 2734–2747, 2019.

[92] J. Xin, N. Wang, X. Gao, and J. Li, "Residual attribute attention network for face image super-resolution," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 9054–9061.

[93] 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*, vol. 3, no. 2, pp. 166–179, 2021.

[94] 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*, 2022.

[95] 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*, 2022.

[96] H. Qi, Y. Qiu, X. Luo, and Z. Jin, "An efficient latent style guided transformer-cnn framework for face super-resolution," *IEEE Transactions on Multimedia*, 2023.

[97] 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," *Proceedings of the International Joint Conference on Artificial Intelligence*, 2023.

[98] Q. Bao, B. Gang, W. Yang, J. Zhou, and Q. Liao, "Attention-driven graph neural network for deep face super-resolution," *IEEE Transactions on Image Processing*, vol. 31, pp. 6455–6470, 2022.

[99] C. Wang, J. Jiang, and X. Liu, "Heatmap-aware pyramid face hallucination," in *Proceedings of the IEEE International Conference on Multimedia and Expo*. IEEE, 2021, pp. 1–6.

[100] 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, pp. 1689–1697.

[101] 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*, vol. 94, pp. 110–121, 2019.

[102] 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 ACM International Conference on Multimedia*, 2020, pp. 1891–1899.

[103] S. Lin, J. Zhang, J. Pan, Y. Liu, Y. Wang, J. Chen, and J. Ren, "Learning to deblur face images via sketch synthesis," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, 2020, pp. 11 523–11 530.

[104] J. Li, B. Bare, S. Zhou, B. Yan, and K. Li, "Organ-branched cnn for robust face super-resolution," in *Proceedings of the IEEE International Conference on Multimedia and Expo*. IEEE, 2021, pp. 1–6.

[105] C. Wang, J. Jiang, Z. Zhong, and X. Liu, "Propagating facial prior knowledge for multitask learning in face super-resolution," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 11, pp. 7317–7331, 2022.

[106] Y. Lu, Y.-W. Tai, and C.-K. Tang, "Attribute-guided face generation using conditional cyclegan," in *Proceedings of the European Conference on Computer Vision*, 2018, pp. 282–297.

[107] M. Zhang and Q. Ling, "Supervised pixel-wise gan for face super-resolution," *IEEE Transactions on Multimedia*, vol. 23, pp. 1938–1950, 2020.

[108] H. Li, Y. Yang, M. Chang, S. Chen, H. Feng, Z. Xu, Q. Li, and Y. Chen, "Srdiff: Single image super-resolution with diffusion probabilistic models," *Neurocomputing*, vol. 479, pp. 47–59, 2022.

[109] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, "Image super-resolution via iterative refinement," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 4, pp. 4713–4726, 2022.

[110] S. Gao, X. Liu, B. Zeng, S. Xu, Y. Li, X. Luo, J. Liu, X. Zhen, and B. Zhang, "Implicit diffusion models for continuous super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2023, pp. 10 021–10 030.

[111] G. G. Chrysos and S. Zafeiriou, "Deep face deblurring," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 69–78.

[112] X. Xu, D. Sun, J. Pan, Y. Zhang, H. Pfister, and M.-H. Yang, "Learning to super-resolve blurry face and text images," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 251–260.

[113] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang, "Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 8878–8887.

[114] 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 ACM International Conference on Multimedia*, 2020, pp. 1551–1560.

[115] A. Li, G. Li, L. Sun, and X. Wang, "Faceformer: Scale-aware blind face restoration with transformers," *arXiv preprint arXiv:2207.09790*, 2022.

[116] X. Li, C. Chen, S. Zhou, X. Lin, W. Zuo, and L. Zhang, "Blind face restoration via deep multi-scale component dictionaries," in *Proceedings of the European Conference on Computer Vision*. Springer, 2020, pp. 399–415.

[117] J. Wang, S. Chen, Z. Wu, and Y.-G. Jiang, "Ft-tdr: Frequency-guided transformer and top-down refinement network for blind face inpainting," *IEEE Transactions on Multimedia*, 2022.

[118] X. Yu, F. Porikli, B. Fernando, and R. Hartley, "Hallucinating unaligned face images by multiscale transformative discriminative networks," *International Journal of Computer Vision*, vol. 128, no. 2, pp. 500–526, 2020.

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

[120] P. Esser, R. Rombach, and B. Ommer, "Taming transformers for high-resolution image synthesis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021, pp. 12 873–12 883.

[121] Y. Wang, Y. Hu, and J. Zhang, "Panini-net: Gan prior based degradation-aware feature interpolation for face restoration," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 3, 2022, pp. 2576–2584.

[122] F. Zhu, J. Zhu, W. Chu, X. Zhang, X. Ji, C. Wang, and Y. Tai, "Blind face restoration via integrating face shape and generative priors," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2022, pp. 7662–7671.

[123] Y. Hu, Y. Wang, and J. Zhang, "Dear-gan: Degradation-aware face restoration with gan prior," *IEEE Transactions on Circuits and Systems for Video Technology*, 2023.

[124] Z. Li, D. Zeng, X. Yan, Q. Shen, and B. Tang, "Analyzing and combating attribute bias for face restoration," *Proceedings of the International Joint Conference on Artificial Intelligence*, 2023.

[125] Y. Wang, Y. Hu, J. Yu, and J. Zhang, "Gan prior based null-space learning for consistent super-resolution," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 3, 2023, pp. 2724–2732.

[126] 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*, vol. 35, pp. 30 599–30 611, 2022.

[127] 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, pp. 2223–2232.

[128] A. Bulat, J. Yang, and G. Tzimiropoulos, "To learn image super-resolution, use a gan to learn how to do image degradation first," in *Proceedings of the European Conference on Computer Vision*, 2018, pp. 185–200.

[129] 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 Conference on Computer Vision and Pattern Recognition*, 2023, pp. 1704–1713.

[130] Y. Wang, J. Yu, and J. Zhang, "Zero-shot image restoration using denoising diffusion null-space model," *The Eleventh International Conference on Learning Representations*, 2023.

[131] X. Qiu, C. Han, Z. Zhang, B. Li, T. Guo, and X. Nie, "Diffbfr: Bootstrapping diffusion model towards blind face restoration," in *Proceedings of the ACM International Conference on Multimedia*, 2023.

[132] J. Liu and C. Jung, "Facial image inpainting using multi-level generative network," in *Proceedings of the IEEE International Conference on Multimedia and Expo*. IEEE, 2019, pp. 1168–1173.

[133] C. Zeng, Y. Liu, and C. Song, "Swin-casunet: Cascaded u-net with swin transformer for masked face restoration," in *Proceedings of the International Conference on Pattern Recognition*. IEEE, 2022, pp. 386–392.

[134] S. Ge, C. Li, S. Zhao, and D. Zeng, "Occluded face recognition in the wild by identity-diversity inpainting," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 10, pp. 3387–3397, 2020.

[135] H. Li, W. Wang, C. Yu, and S. Zhang, "Swapinpaint: Identity-specific face inpainting with identity swapping," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 7, pp. 4271–4281, 2021.

[136] Y. Zhang, X. Zhang, C. Shi, X. Wu, X. Li, J. Peng, K. Cao, J. Lv, and J. Zhou, "Pluralistic face inpainting with transformation of attribute information," *IEEE Transactions on Multimedia*, 2022.

[137] Y. Bai, R. He, W. Tan, B. Yan, and Y. Lin, "Fine-grained blind face inpainting with 3d face component disentanglement," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2023, pp. 1–5.

[138] L. Yang, B. Shao, T. Sun, S. Ding, and X. Zhang, "Hallucinating very low-resolution and obscured face images," *arXiv preprint arXiv:1811.04645*, 2018.

[139] G. Gao, L. Tang, F. Wu, H. Lu, and J. Yang, "Jdsr-gan: Constructing an efficient joint learning network for masked face super-resolution," *IEEE Transactions on Multimedia*, vol. 25, pp. 1505–1512, 2023.

[140] 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*, vol. 93, p. 103806, 2023.

[141] Y. Zhang, I. W. Tsang, J. Li, P. Liu, X. Lu, and X. Yu, "Face hallucination with finishing touches," *IEEE Transactions on Image Processing*, vol. 30, pp. 1728–1743, 2021.

[142] K. Li and Q. Zhao, "If-gan: Generative adversarial network for identity preserving facial image inpainting and frontalization," in *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*. IEEE, 2020, pp. 45–52.

[143] Q. Duan, L. Zhang, and X. Gao, "Simultaneous face completion and frontalization via mask guided two-stage gan," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 6, pp. 3761–3773, 2021.

[144] X. Yu and F. Porikli, "Hallucinating very low-resolution unaligned and noisy face images by transformative discriminative autoencoders," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 3760–3768.

[145] A. Abbasi and M. Rahmati, "Identity-preserving pose-robust face hallucination through face subspace prior," *arXiv preprint arXiv:2111.10634*, 2021.

[146] S. Menon, A. Damian, S. Hu, N. Ravi, and C. Rudin, "Pulse: Self-supervised photo upsampling via latent space exploration of generative models," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2437–2445.

[147] K. Zhang, Z. Zhang, C.-W. Cheng, W. H. Hsu, Y. Qiao, W. Liu, and T. Zhang, "Super-identity convolutional neural network for face hallucination," in *Proceedings of the European Conference on Computer Vision*, 2018, pp. 183–198.

[148] E. Ataer-Cansizoglu, M. Jones, Z. Zhang, and A. Sullivan, "Verification of very low-resolution faces using an identity-preserving deep face super-resolution network," *arXiv preprint arXiv:1903.10974*, 2019.

[149] J. Mathai, I. Masi, and W. AbdAlmageed, "Does generative face completion help face recognition?" in *Proceedings of the International Conference on Biometrics*. IEEE, 2019, pp. 1–8.

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

[151] C.-C. Hsu, C.-W. Lin, W.-T. Su, and G. Cheung, "Sigan: Siamese generative adversarial network for identity-preserving face hallucination," *IEEE Transactions on Image Processing*, vol. 28, no. 12, pp. 6225–6236, 2019.

[152] B. Bayramli, U. Ali, T. Qi, and H. Lu, "Fh-gan: Face hallucination and recognition using generative adversarial network," in *Proceedings of the Neural Information Processing International Conference*. Springer, 2019, pp. 3–15.

[153] H. Huang, R. He, Z. Sun, and T. Tan, "Wavelet domain generative adversarial network for multi-scale face hallucination," *International Journal of Computer Vision*, vol. 127, no. 6-7, pp. 763–784, 2019.

[154] H. A. Le and I. A. Kakadiaris, "Selenet: A semi-supervised low light face enhancement method for mobile face unlock," in *Proceedings of the International Conference on Biometrics*. IEEE, 2019, pp. 1–8.

[155] X. Ding and R. Hu, "Learning to see faces in the dark," in *Proceedings of the IEEE International Conference on Multimedia and Expo*. IEEE, 2020, pp. 1–6.

[156] R. Yasarla, H. R. V. Joze, and V. M. Patel, "Network architecture search for face enhancement," *arXiv preprint arXiv:2105.06528*, 2021.

[157] C. Qu, C. Herrmann, E. Monari, T. Schuchert, and J. Beyerer, "Robust 3d patch-based face hallucination," in *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*. IEEE, 2017, pp. 1105–1114.

[158] J. Li, F. Zhu, X. Yang, and Q. Zhao, "3d face point cloud super-resolution network," in *Proceedings of the IEEE International Joint Conference on Biometrics*. IEEE, 2021, pp. 1–8.

[159] K. Uddin, T. H. Jeong, and B. T. Oh, "Incomplete region estimation and restoration of 3d point cloud human face datasets," *Sensors*, vol. 22, no. 3, p. 723, 2022.

[160] N. Gat, S. Benaim, and L. Wolf, "Identity and attribute preserving thumbnail upscaling," in *Proceedings of the IEEE International Conference on Image Processing*. IEEE, 2021, pp. 2708–2712.

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

[162] X. Yu, B. Fernando, R. Hartley, and F. Porikli, "Super-resolving very low-resolution face images with supplementary attributes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 908–917.

[163] J. Xin, N. Wang, X. Jiang, J. Li, X. Gao, and Z. Li, "Facial attribute capsules for noise face super resolution," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, 2020, pp. 12 476–12 483.

[164] Z.-S. Liu, W.-C. Siu, and Y.-L. Chan, "Features guided face super-resolution via hybrid model of deep learning and random forests," *IEEE Transactions on Image Processing*, vol. 30, pp. 4157–4170, 2021.

[165] Y. Zhao, T. Hou, Y.-C. Su, X. J. Li, M. Grundmann *et al.*, "Towards authentic face restoration with iterative diffusion models and beyond," in *Proceedings of the IEEE International Conference on Computer Vision*, 2023.

[166] W. Yang, Z. Chen, C. Chen, G. Chen, and K.-Y. K. Wong, "Deep face video inpainting via uv mapping," *IEEE Transactions on Image Processing*, vol. 32, pp. 1145–1157, 2023.

[167] M. Li, Y. Sun, Z. Zhang, H. Xie, and J. Yu, "Deep learning face hallucination via attributes transfer and enhancement," in *Proceedings of the IEEE International Conference on Multimedia and Expo*. IEEE, 2019, pp. 604–609.

[168] C.-H. Lee, K. Zhang, H.-C. Lee, C.-W. Cheng, and W. Hsu, "Attribute augmented convolutional neural network for face hallucination," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 721–729.

[169] R. Dey and V. N. Boddeti, "3dfacefill: An analysis-by-synthesis approach to face completion," in *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*, 2022, pp. 1586–1595.

[170] K. C. Chan, X. Xu, X. Wang, J. Gu, and C. C. Loy, "Glean: Generative latent bank for image super-resolution and beyond," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 3, pp. 3154–3168, 2022.

[171] Z. Hou, L. Li, and X. Guo, "Feature-guided blind face restoration with gan prior," in *Proceedings of the IEEE International Conference on Multimedia and Expo*. IEEE, 2022, pp. 1–6.

[172] Y. Zhao, Y.-C. Su, C.-T. Chu, Y. Li, M. Renn, Y. Zhu, C. Chen, and X. Jia, "Rethinking deep face restoration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2022, pp. 7652–7661.