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

WENJIE LI, Beijing University of Posts and Telecommunications, China
MEI WANG, Beijing Normal University, China
KAI ZHANG, Nanjing University, China
JUNCHENG LI, East China Normal University, China
XIAOMING LI, Nanyang Technological University, Singapore
YUHANG ZHANG, Beijing University of Posts and Telecommunications, China
GUANGWEI GAO, Nanjing University of Science and Technology, China
ZHANYU MA, Beijing University of Posts and Telecommunications, China

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. Furthermore, we explore the various facial priors commonly utilized in restoration 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 real-world applications and challenges faced in the field of FR, 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.*

CCS Concepts: • **General and reference → Surveys and overviews**; • **Computing methodologies → Reconstruction**; **Neural networks**.

Additional Key Words and Phrases: Face restoration, Survey, Deep learning, Non-blind/Blind, Joint restoration.

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 [163], face recognition [63], and 3D face reconstruction [195]. 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

Corresponding Author: Guangwei Gao.
Authors' addresses: Wenjie Li, Beijing University of Posts and Telecommunications, Beijing, China, lewj2408@gmail.com; Mei Wang, Beijing Normal University, Beijing, China, wangmei1@bnu.edu.cn; Kai Zhang, Nanjing University, Suzhou, China, cskaizhang@gmail.com; Juncheng Li, East China Normal University, Shanghai, China, cvjunchengli@gmail.com; Xiaoming Li, Nanyang Technological University, Singapore, csxmli@gmail.com; Yuhang Zhang, Beijing University of Posts and Telecommunications, Beijing, China, zyhzyh@bupt.edu.cn; Guangwei Gao, Nanjing University of Science and Technology, Nanjing, China, csggao@gmail.com; Zhanyu Ma, Beijing University of Posts and Telecommunications, Beijing, China, mazhanyu@bupt.edu.cn.

0360-0300/2024/2-ART $15.00
https://doi.org/

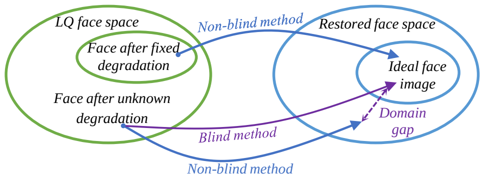


Fig. 1. Domain interpretation of differences between non-blind and blind method. If the degradation factors affecting the LQ face differ from those assumed by the non-blind method, it can result in a significant domain gap between the restored face image and the ideal HQ face image.

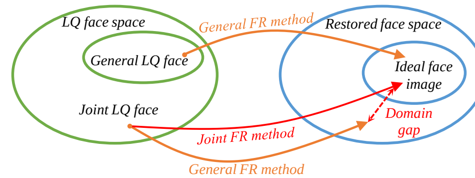


Fig. 2. Domain interpretation of differences between general and joint FR method. If a general FR method is used to accomplish a joint FR task (*e.g.*, joint face completion), it can result in a significant domain gap between the restored face image and the ideal HQ face image.

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

Year	Survey Title	Venue
2018	Super-resolution for biometrics: A comprehensive survey [117]	PR
2019	Survey on GAN-based face hallucination with its model development[104]	IET
2021	Deep Learning-based Face Super-Resolution: A Survey[66]	ACM CSUR
2023	A Survey of Deep Face Restoration: Denoise, Super-Resolution, Deblur, Artifact Removal[146]	Arxiv

face images. Since then, numerous FR methods have been developed, gaining increasing attention from researchers. Traditional FR methods [18, 39, 103, 120] 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, many deep learning-based methods specifically designed for FR tasks have emerged. Deep learning networks, utilizing large-scale datasets, can effectively capture diverse mapping relationships between degraded face images and real face images. Consequently, deep learning-based FR methods [20, 93, 94, 147, 156] have demonstrated significant advantages over traditional methods, offering more robust solutions.

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 [20, 67, 171], 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 and ideal HQ facial images. To address this issue, blind methods [97, 147, 164] 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, as shown in Fig. 2, to address the challenges faced by general methods such as blind/non-blind FR methods dealing with the joint FR tasks, a series of joint FR methods have emerged to tackle specific challenges in face restoration [25, 172, 191, 199]. These joint FR tasks include joint face alignment and restoration [172], joint face recognition and restoration [25], joint illumination compensation and restoration [191], and joint 3D face reconstruction and restoration [199]. Building upon these advancements, our paper aims to comprehensively survey 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.

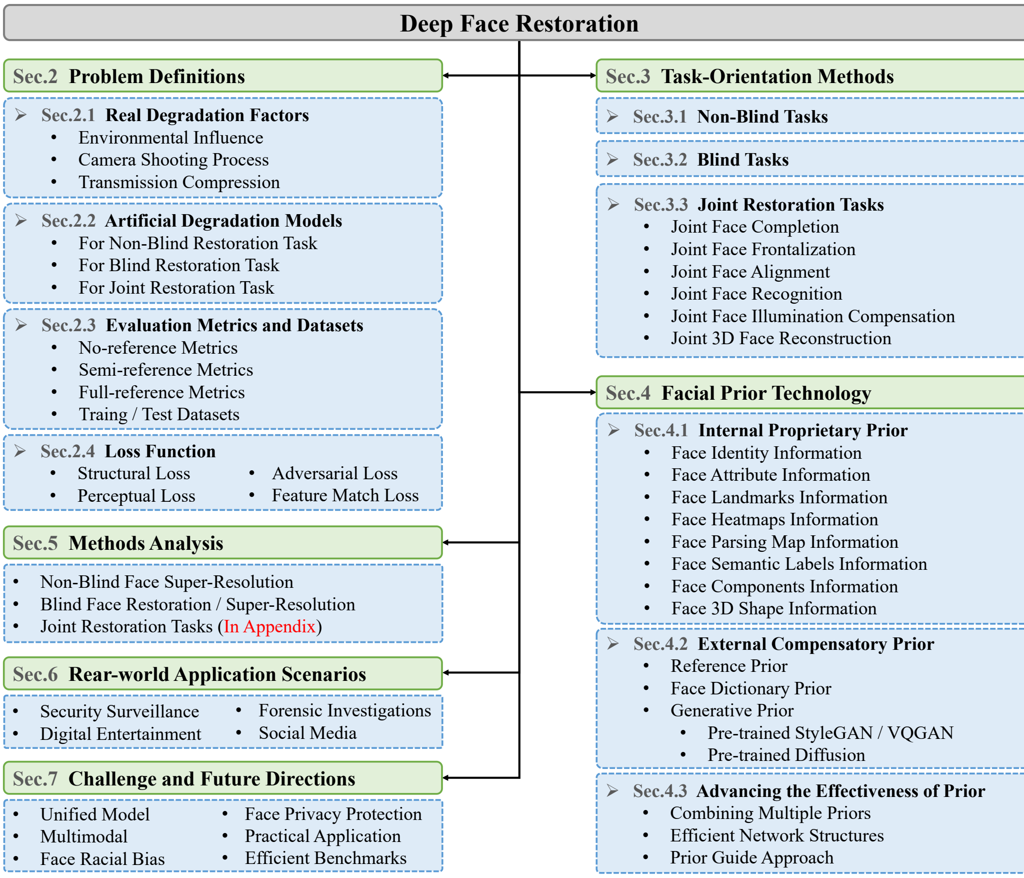


Fig. 3. Outline of our deep learning-based face restoration survey.

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.* [104] provide a review of face super-resolution methods based on generative adversarial networks, but it solely focuses on a specific technique within FR. Kien *et al.* [117] summarize some deep learning-based face super-resolution methods, but it is not comprehensive. Jiang *et al.* [66] present 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.* [146] conduct a survey on FR, however, it adopts 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 FR-related joint tasks. In contrast, our review comprehensively summarizes 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 [146] primarily focuses 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 in our ***Appendix***. 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 task 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. 3 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. Section 3 introduces 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 discusses the FR techniques used in real-world applications. Section 7 addresses the challenges faced in FR and presents potential future directions. Finally, we conclude this survey in Section 8.

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 frequently employed datasets in this field.

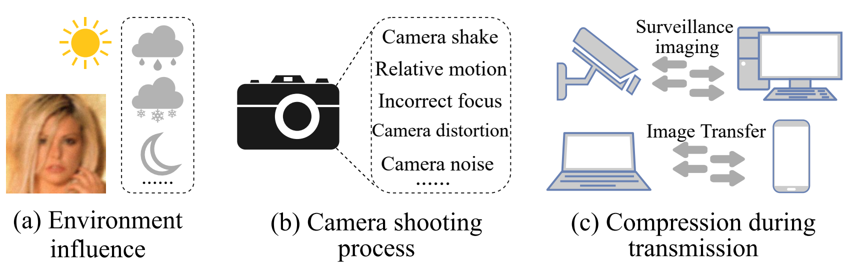


Fig. 4. Several mainstream factors affect the quality of face images. (a) Environment influences, including rain, snow, haze, and low light; (b) Interference in the camera imaging process; (c) Image compression during information transmission.

2.1 Real Degradation Factors

In real-world scenarios, face images are susceptible to degradation during imaging and transmission due to the complex environment. The limitations of the physical imaging equipment and external imaging conditions primarily cause the degradation of facial images. As shown in Fig. 4, we 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

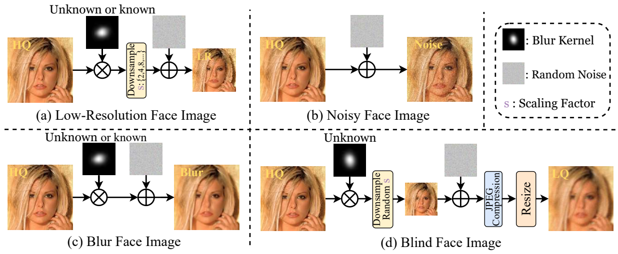


Fig. 5. Methods for generating various types of degraded facial images.

Table 2. Summary of key metrics.

Metrics	Highlight
PSNR [54]	Full reference, pixel-by-pixel comparison of the differences between both.
SSIM [152]	Full reference, focus on differences in brightness, contrast, structure, *etc.*
MS-SSIM [153]	Full reference, average SSIM for windows.
LPIPS [186]	Full reference, focus on the visual perceptual similarity between both.
IDD [154]	Full reference, assess identity consistency.
FID [53]	Semi-reference, measure the difference in distribution between both.
NIQE [116]	No reference, evaluate image naturalness.
MOS [55]	Subjective scoring by groups.

were experimented with to improve robustness by introducing a more comprehensive 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. 5, 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. The FR 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 subtask is classified as blind.

• **Non-blind Degradation Models.** (I) The non-blind task primarily focuses on face super-resolution (FSR) [112], also known as face hallucination [48]. As shown in Fig. 5 (a), its degradation model involves degrading a high-resolution (HR) face image into a low-resolution (LR) face image. FSR can be categorized as a non-blind task when the blur kernel is pre-determined and remains constant, such as a Gaussian blur kernel or any other well-defined blur kernel. The degradation model can be described as follows:

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

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

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

(II) Face denoising [24] and face deblurring [128] 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. 5 (b) and (c), when the blur kernel remains constant, they can be classified

(a) Joint face alignment (b) Joint face completion (c) Joint face fractalization (d) Joint illumination compensation (e) Joint 3D reconstruction

Fig. 6. Demonstration of LQ and HQ face images for joint face restoration tasks.

as non-blind tasks. Their degradation model can be described separately as:

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

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

where I_n represents the noisy face image, I_b represents the blurred image, I_{hq} represents the clean HQ face image, k_f represents the fixed blur kernel and n_δ 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 [9] and face deblurring [80] can be classified as blind tasks. As shown in Fig. 5 (a) and (c), their degradation processes can be described separately as follows:

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

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

where k_u is a parameters unknown blur kernel, the remaining variables have the same meanings as described above for non-blind face super-resolution and 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 [17, 52, 154] aims to address this limitation by considering more complex degradations, making it the most prominent task in the field currently. GFRNet [97] 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. 5 (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:

$$I_{lq} = \{JPEG_q((I_{hq} \otimes k_u)\downarrow_s + n_\delta)\}\uparrow_s, \tag{8}$$

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. $\downarrow_s$ and $\uparrow_s$ represent down-sampling and up-sampling operations with arbitrary scale factors s, respectively. n_δ 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. 6 showcases several examples of joint tasks, depicted from left to right: (a) Joint face alignment and restoration [172]: This task addresses the challenge of misaligned faces by aligning and restoring them. (b) Joint face completion and restoration [11]: The objective is to handle face occlusions and restore the facial image's missing regions. (c) Joint face frontalization and restoration [175]: This task focuses on recovering frontal faces from side faces, enhancing their appearance and quality. (d) Joint face illumination compensation and restoration [191]: This task aims to restore faces captured in low-light conditions, compensating for the lack of illumination. (e) Joint 3D face reconstruction [195]: 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

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** [63]	13K	73	✗	✗	✔	HQ	C-SRIP [48], LRFR [79], DPDFN [68], *etc.*
2010	**Multi-PIE** [49]	755.4K	✗	✗	✗	✔	HQ	FSGN [130], CPGAN [192], MDCN [129], *etc.*
2011	**AFLW** [77]	26K	✗	21	✗	✗	HQ	FAN [74], JASRNet [168], *etc.*
2011	**SCFace** [47]	4.2K	✗	4	✗	✗	HQ	MNCE [67], SISN [110], CTCNet [44], *etc.*
2012	**Helen** [81]	2.3K	✗	194	✔	✗	HQ	DIC [112], SAAN [196], SCTANet [6], *etc.*
2013	**300W** [125]	3.8K	✗	68	✗	✗	HQ	JASRNet [168], *etc.*
2014	**CASIA-WebFace** [167]	494.4K	✗	2	✗	✔	HQ	MDFR [135], C-SRIP [48], GFRNet [97], *etc.*
2015	**CelebA** [106]	202.6K	40	5	✗	✔	HQ	FSRNet [20], SPARNet [16], SFMNet [141], *etc.*
2016	**Widerface** [163]	32.2K	✗	✗	✗	✗	HQ	Se-RNet [176], SCGAN [56], *etc.*
2017	**LS3D-W** [8]	230K	✗	68	✗	✗	HQ	Super-FAN [9], SCGAN [56], *etc.*
2017	**Menpo** [180]	9K	✗	68/39	✗	✗	HQ	SAM3D [60, 61], *etc.*
2018	**VGGFace2** [13]	3310K	✗	✗	✗	✔	HQ	GFRNet [97], ASFFNet [96], GWAInet [32], *etc.*
2019	**FFHQ** [72]	70K	✗	68	✗	✗	HQ	mGANprior [50], GFPGAN [147], VQFR [51], *etc.*
2020	**CelebAMask-HQ** [82]	30K	✗	✗	✔	✗	HQ	MSGGAN [71], GPEN [164], Pro-UIGAN [193], *etc.*
2022	**EDFace-Celeb-1M** [182]	1700K	✗	✗	✗	✗	HQ-LQ	STUNet [185], *etc.*
2022	**CelebRef-HQ** [98]	10.6K	✗	✗	✗	✔	HQ	DMDNet [98], *etc.*

designed to address face restoration challenges in specific scenarios and hold practical significance in their respective domains.

2.3 Evaluation Metrics And Datasets

We compile 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, facial features included, the availability of HQ-LQ pairs, and previous methods that utilize these datasets. We need to synthesize corresponding LQ face images for datasets that only provide HQ face images using degradation models 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:

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

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, 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 [93, 112, 154] and the $L2$ loss [17, 20, 96]. They can be formulated as

$$L_1 = \left\| I_{hq}(h, w, c) - \hat{I}_{hq}(h, w, c) \right\|_1, \quad L_2 = \left\| I_{hq}(h, w, c) - \hat{I}_{hq}(h, w, c) \right\|_2, \tag{11}$$

where h, w, and c represent the height, width, and number of channels, respectively. The pixel-level loss also encompasses the Huber loss [70] and the Carbonnier penalty function. Furthermore, textural losses have been developed in addition to the pixel-level losses. These include the SSIM loss [48], which promotes image textural similarity, and the cyclic consistency loss [56], which facilitates cooperation between recovery and degradation processes. While structural loss preserves the overall facial details and structure, ensuring a close match to the ground truth and improving PSNR, this can result in a face that appears overly smooth, lacking fine details, and less natural in terms of texture and lighting, often leading to a perceptually stiff reconstruction that may not align with human aesthetic expectations.

• **Perceptual loss.** The perceptual loss is intended to enhance the visual quality of recovered images by comparing them to 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 pre-trained networks, as these features represent high-level semantic information. Denoting the l-th layer involved in the computation of pre-trained networks as φ_l, its perceptual loss L^l_{per} can be expressed as follows:

$$L^l_{per} = \left\| \varphi_l(I_{hq}(h, w, c)) - \varphi_l(\hat{I}_{hq}(h, w, c)) \right\|_2. \tag{12}$$

This loss enhances the perceptual quality, making restored details (*e.g.*, expression, eye shape, skin texture) appear more realistic to human vision. However, it might prioritize global visual appeal over local accuracy, which can lead to a subtle loss of sharpness or precision in certain areas.

• **Adversarial loss.** The adversarial loss is a common type of loss used in GAN-based face restoration methods [17, 52, 154]. In this setup, the generator G aims to generate an HQ face image to deceive the discriminator D. In contrast, 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:

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

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

where $L_{adv,G}$ and $L_{adv,D}$ are the adversarial losses of the generator and discriminator, respectively. The use of adversarial loss improves the visual realism of facial texture and natural appearance. However, it may introduce training instabilities and artifacts, requiring careful parameter tuning. While it generates appealing results, it may also lead to a loss of detail in less prominent areas and cause over-optimization of specific features, resulting in inconsistencies with surrounding regions.

• **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 [20], face heatmaps [9], 3D face shape [60], semantic-aware style [14], face parsing [17], facial attention [74], face identity [48], and facial components [147]. Among these, the face landmarks loss is widely utilized and can be described as

$$L_{landmarks} = \frac{1}{N} \sum_{n=1}^{N} \left\| l^n_{x,y} - \hat{l}^n_{x,y} \right\|_2, \tag{15}$$

where N is the number of facial landmarks, and $l^n_{x,y}$ and $\hat{l}^n_{x,y}$ represent the coordinates of the n-th landmark point in the HQ face and the recovered face, respectively. Feature matching loss focuses on preserving facial attributes by aligning key facial components (*e.g.*, eyes, nose) with their counterparts in the target image, improving both facial details and overall visual quality. However,

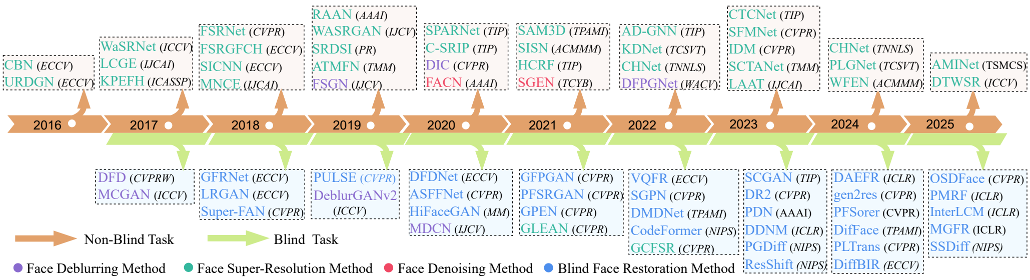


Fig. 7. Milestones of deep learning-based non-blind/blind task methods, including their names and venues.

excessive emphasis on these components may result in a loss of detail in other regions, leading to inconsistencies between local facial features and the overall facial structure.

• **Discuss.** Early methods [20, 112] primarily relied on structural loss, which preserved facial details but resulted in overly smooth images, lacking fine texture. To address this, feature matching loss is introduced [9, 60] to improve the alignment of facial components, enhancing attribute preservation. However, it still causes inconsistencies across different regions. The addition of perceptual loss [16, 44] improves visual quality by aligning high-level features, enhancing realism, but still leaving some lacking fine details. Influenced by the generative task, adversarial loss is introduced in GAN-based methods [147, 201] to improve visual realism. While it enhances the natural appearance, it also causes artifacts and over-optimization in specific regions. By combining these loss functions, modern methods aim to balance fine details and perceptual quality.

3 TASK-ORIENTATION 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. 7 illustrates several notable methods in recent years that focus on non-blind and blind tasks. Fig. 8 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. 9 (a). One pioneering work is the bi-channel convolutional neural network (BCCNN) proposed by Zhou *et al.* [200], which significantly surpasses previous conventional approaches. This network combines the extracted face features with the input face features and utilizes a decoder to reconstruct HQ face images, leveraging its strong fitting capability. Similarly, other methods [21, 41, 108] 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. 9 (b), one class of methods adopts a two-stage approach for face restoration, consisting of roughing and refining stages. For example, CBN [204] employs a cascaded framework to address the performance limitations observed in previous methods when dealing with misaligned facial images. LCGE [131], MNCE [67], and FSGN [130] generate facial components that approximate real landmarks and enhance them by recovering details. FSRNet [20] obtains a rough face image through a network and then refines it using a heatmap and a resolving map of facial landmarks. DIDnet [23] and ATSENet [91] utilize facial identity or attributes to enhance the features extracted by the initial network and recover face images with higher confidence. JASRNet [168] achieves high-quality restoration in parallel

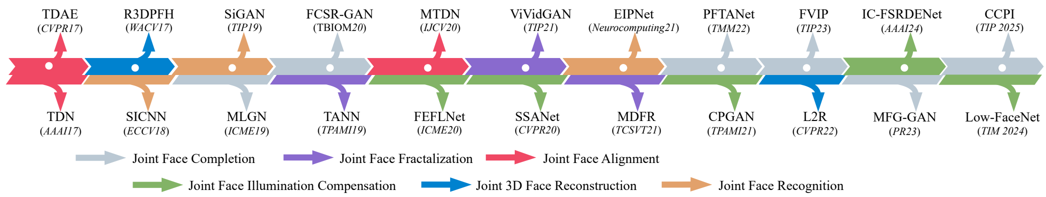


Fig. 8. Milestones of the Joint Face Restoration methods, including their names and venues. Over time, joint tasks have transitioned from early tasks like joint alignment, detection, and frontalization to joint 3D face reconstruction, completion, and illumination compensation.

by supervising facial landmarks and HQ face images. FAN [74] employs a facial attention prior loss function 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. 9 (c). For example, KPEFH [88] utilizes multiple branches in the network to predict key components of the face separately. FSRGFCH [169] enhances the quality of facial details by predicting the face component heatmap with an additional branching in the network. HCRF [107] utilizes random forests to recover different regions of the face semantics predicted by the network separately.

Attention mechanisms have demonstrated their effectiveness in image restoration methods [100, 187]. Subsequently, there has been a significant focus on integrating attention mechanisms [12] to enhance the handling of important facial regions. Various networks based on attention mechanisms have been developed, as illustrated in Fig. 10. Attention can be categorized into four types: channel attention, spatial attention, self-attention, and hybrid attention. Channel attention-based approaches [27, 69, 158, 196] emphasize the relative weights between different feature channels in the model, enabling selective emphasis on important channels. Spatial attention-based approaches [16, 61, 109] focus on capturing spatial contextual information about features, enabling the model to prioritize features relevant to key face structures. Self-attention-based approaches [84, 121, 150] mainly capture global facial information, yielding excellent performance. Some approaches [16, 74] also enhance individual attention mechanisms to suit the specific requirements of FR tasks better. Hybrid attention-based approaches [5, 6, 44, 93**?**] 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 guiding the network. For instance, SAAN [196] incorporates the face parsing map, FAN [74] incorporates the face landmark, SAM3D [60, 61] incorporates the 3D face information, HaPSR [139] incorporates the face heatmap, and CHNet [109] incorporates the face components. To direct attention more precisely, some methods have started to delineate and recover different regions of the face image artificially. WaSRNet [64] employs the 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 [59] 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 [141] integrates information extracted from its spatial and frequency branches, enhancing the texture of the contour. WFEN [93] leverages the wavelet transform to minimize the loss of facial features during downsampling.

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

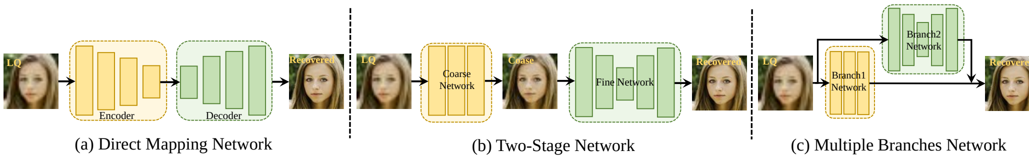


Fig. 9. Summary of the architecture of general methods for non-blind face restoration.

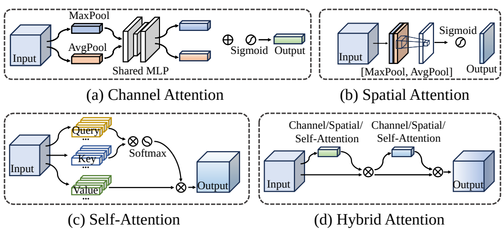


Fig. 10. The network architecture of Attention-based face restoration methods.

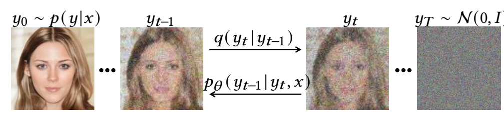
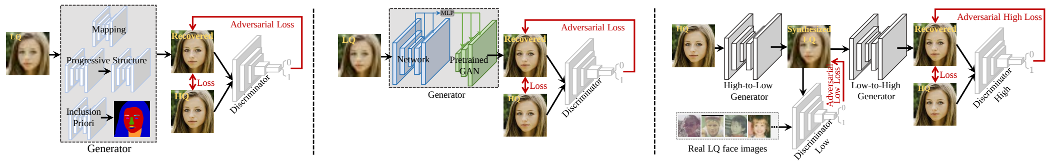


Fig. 11. The diffusion denoising principle involves two main processes: forward diffusion process q and reverse inference process p. In the q, Gaussian noise is gradually added to the target image from left to right. In contrast, p iteratively denoises the target image, proceeding from right to left.

methods primarily employ the general GAN structure depicted in Fig. 12 (a). In 2016, Yu *et al.* [171] 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 restoration methods have been extended to the non-blind task, showing promising recovery results. Some methods focus on designing progressive GANs, including two- or multi-stage approaches [33, 71, 101]. Others concentrate on embedding face-specific prior information, such as facial geometry [176], facial attributes [111], or identity information [184] into the GAN framework. It is worth noting that, given the excellent performance of GAN, many non-GAN-based methods [5, 6, 16, 20, 44, 112] 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 denoising diffusion probabilistic model (DDPM) has been proposed as an alternative approach. As shown in Fig. 11, 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 [86] introduced a diffusion-based model for face super-resolution. It incorporated residual prediction throughout the framework to accelerate convergence. Then, SR3 [126] achieves super-resolution by iteratively denoising the conditional images generated by the denoising diffusion probabilistic model, resulting in more realistic outputs at various magnification factors. IDM [45] combines an implicit neural representation with a denoising diffusion model. This allows the model to meet continuous-resolution requirements and provide HQ face restoration with improved scalability across different scales. ResDiff [127] utilizes a CNN to recover the low-frequency portion of LQ face images and diffusion models to focus on recovering the high-frequency portion of LQ face images. UCDIR [189] uses a lightweight UNet for initial guidance and spatially adaptive conditioning to improve perceptual quality in diffusion models. DTWSR [34] enhances the



(a) General GAN architecture (b) Pre-trained GAN embedding architecture (c) Using unpaired images GAN architecture

Fig. 12. Summary of the architecture of GAN-based methods for face restoration.

detail of the reconstruction by improving the correlation of multi-scale frequency-domain features in diffusion models through the discrete wavelet transform.

• **Discuss.** Non-blind tasks primarily rely on simple downsampling models, such as bicubic interpolation. Current approaches heavily depend on structural loss and feature matching loss to improve overall facial consistency and component alignment. When combined with specific architectural designs—such as multi-branch architectures and various attention mechanisms—these methods prioritize processing critical facial regions to achieve more natural results. Despite these advancements, non-blind methods are still constrained by limitations in efficient architecture innovations. In practical applications, they also remain hindered by domain gaps and often produce visible artifacts when the locations of damage are unknown.

3.2 Blind Tasks

In practical applications, researchers have observed that methods originally designed for non-blind tasks often struggle to handle real-world LQ face images effectively. 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 [26], introduced by G. Chrysos *et al.*, which employs a modified ResNet architecture for blind face deblurring. Then, MCGAN [160] leverages 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. 12 (a). For example, DeblurGAN-v2 [78], HiFaceGAN [161], STUNet [185], and GCFSR [52] all design novel and intricate network architectures for blind face restoration. Additionally, many methods use more complex GAN networks with prior information. GFRNet [97], ASFFNet [96], and DMDNet [98] utilize a bootstrap network with reference to prior to guiding the recovery network, employing a two-stage strategy for better face restoration. MDCN [129] and PFSRGAN [17] employ a two-stage network consisting of a face semantic label prediction network and a recovery parsing network for reconstruction. Furthermore, Super-FAN [9], DFDNet [95], and RestoreFormer [154] 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 gradually developed in the field of blind face restoration since generative models [72, 73] can produce realistic and HQ face images. As shown in Fig. 12 (b), the pre-trained GAN embedding architecture involves adding an additional pre-trained generative GAN [36, 72] into the generator network. For example, GPEN [164] 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 [147] 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 [14], Panini-Net [149], SGPN [202], DEAR-GAN [62], DebiasSR [99], PDN [148], 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 [51], CodeFormer [201], and others employ pre-trained VQGAN to enhance facial details. They achieve this by employing discrete feature codesets extracted from HQ face images as prior. The discrete 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 [203], as shown in Fig. 12 (c), LRGAN [10] employs a cyclic GAN architecture consisting of two GAN networks. The initial high-to-low GAN generates LQs that mimic real-world conditions and pairs them with corresponding HQs. Subsequently, the second low-to-high GAN network is used to restore and enhance the quality of generated LQ face images for restoration purposes. SCGAN [56] takes a step further by guiding the generation of paired LQs through the creation of degenerate branches from HQs. This approach further reduces the domain gap between generated and authentic LQs. Additionally, diffusion-denoising techniques for blind tasks aim to improve robustness in severely degraded scenarios compared to non-blind tasks. DR2 [156] employs this technique to enhance the robustness of blind restoration and reduce artifacts often observed in results. DDPM [151] refines the spatial content during backpropagation to improve the robustness and realism of the restoration in challenging scenarios. DiffBFR [122] takes a different approach by initially restoring LQ images and employing an LQ-independent unconditional diffusion to refine textures rather than directly restoring HQ images from noisy inputs. PGDiff [162] introduces partial bootstrapping to make the utilization of pre-trained diffusion modeling methods more adaptable to real degradations. WaveFace [115] uses a wavelet transform to decompose face images into low and high-frequency components, with a diffusion model recovering the low-frequency part and a unified network restoring high-frequency details to preserve identity and enhance efficiency. PLTrans [157] learns a degradation-unaware representation using latent diffusion-based regularization and refines features with a latent dictionary of high-quality face priors. DifFace [178] gradually transfers LQ from intermediate states to HQ by recursively applying a pre-trained diffusion model. ResShift [179] reduces sampling steps by using a Markov chain for faster transitions to improve efficiency in diffusion. Gen2res [31] improve the fidelity of existing diffusion-based methods by constraining the generation space. PFStorer [137] presents a personalized face restoration using diffusion models, fine-tuning with a few identity images to preserve identity and details without requiring alignment. T2I [42] utilizes texts to guide the denoising to improve the consistency of results with inputs. Suin *et al.* [132] use a conditional diffusion framework with an identity-preserving conditioner to improve perceptual quality. SSDiff [94] employs pseudo-label construction and staged guidance to leverage diffusion priors for old photo face restoration. BFRffusion [19], DiffBIR [102], StableSR [144], and InterLCM [92] further improve the fidelity and quality of restoration by fine-tuning priors in pre-trained text-to-image diffusion models (*e.g.*, Stable diffusion [124]). OSDFace [143] performs single-step inference by generating prompts through enhanced understanding of facial features. PMRF [118] further enhances face restoration through the concept of flow matching.

• **Discuss.** Blind face restoration addresses real-world degradation with unknown factors. Early methods simulate a variety of degradation patterns, but they are computationally expensive. Pre-trained GAN and diffusion priors have improved realism, but challenges such as the introduction of artifacts and a trade-off between fidelity and over-optimization. Blind methods benefit from incorporating various face priors, while they use flexible prior embedding architectures to handle complex distortions. Future improvements in blind restoration will depend on better priors embedded in excellent network architecture to balance fidelity and restoration quality.

3.3 Joint Restoration Tasks

In this section, we will discuss beyond components of FR, which include joint face completion and restoration, joint face frontalization, joint face alignment, joint face recognition, joint face illumination compensation, and joint 3D face reconstruction.

• **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 [105] and Swin-CasUNet [181] use general networks for completion, but their fidelity is limited. Since accurately estimating occluded features is a key challenge, most methods integrate facial priors to infer critical details. Examples include ID-GAN [46] (facial identity), SwapInpaint [85] (reference face), PFTANet [194] (semantic labels), and FT-TDR [142] (landmarks). Another class of methods focuses on low-resolution face completion. Early methods [11, 43] address occlusion through patching first before performing restoration work, but this may lead to error accumulation. In contrast, UR-GAN [193] uses landmarks to progressively restore occluded faces. Pang *et al.* [119] enhance inversion with progressive sampling for better restoration. CCPI [188] employs cross-perception collaboration schemes to enhance completion. The challenge is recovering features closely related to unobscured areas, requiring generation strategies to solve this problem when occlusion is high.

• **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 [175]. It utilizes a discriminative network to enforce the side-face generated face image to be close to the front-face image, aligning the faces in the same plane. Subsequently, VividGAN [190] employs a fractalization network combined with a fine feature network to optimize the face details under fractalization further. MDFR [135] introduces 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 [35, 89] also combine the tasks of face completion and frontalization to address them jointly. With the widespread use of generative priors, it may be possible to estimate more realistic frontal faces by taking advantage of the technique.

• **Joint Face Alignment.** Most FR methods require the use of aligned face training samples for optimal performance. Therefore, researchers have developed various methods for joint face alignment. Yu *et al.* are among the first to attempt embedding a spatial transformation layer as a generator and utilizing a discriminator to improve the alignment and upsampling. They develop TDN [172] and MDTN [174] using this approach. To handle possible noise in unaligned faces, they also develop a method [173] that incorporates downsampling and upsampling within the TDN framework to minimize the noise's impact. Another approach [1] utilizes a face 3D dictionary alignment scheme to accomplish alignment. The latest methods [141] tend to use well-aligned face datasets for training and achieve better performance compared to the above methods.

• **Joint Face Recognition.** Some methods [50, 114] may result in recovered face images that diverge from their original identities, making them unsuitable for downstream face recognition tasks. This task aims to solve the identity consistency problem of recovered faces. Since face recognition heavily relies on local features such as the eyes, many priors struggle to emphasize these specific areas accurately. One swift solution involves applying a pre-trained face recognition model after restoration. This helps determine whether restored face images align with the ground truth in terms of identity, enhancing restoration accuracy by incorporating identity-related prior knowledge.

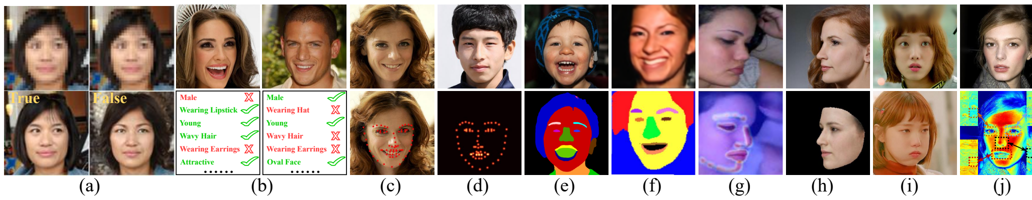


Fig. 13. Visualization of popular facial priors. (a) Facial identity; (b) Facial attributes; (c) Facial landmarks; (d) Facial heatmaps; (e) Facial parsing map; (f) Facial semantic labels; (g) Facial components; (h) 3D Face shape; (i) Facial reference priors; (j) Facial dictionary.

Some examples of these methods include SICNN [183], LRFR [79], and others [75, 113]. C-SRIP [48] 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 [58], FH-GAN [7], WaSRGAN [65], and others, further enhance performance by incorporating discriminators into restoration.

• **Joint Face Illumination Compensation.** This joint task aims to recover faces degraded by both low light and low quality. CPGAN [192] uses internal and external networks for detail restoration and background relighting. Zhang *et al.*[191] improve CPGAN with landmark constraints and recursive strategies. Ding *et al.*[30] use face landmarks to enhance contour restoration. Low-FaceNet [37] enhances the brightness of real-world face images through unsupervised contrastive learning. IC-FSRDENet [138] integrates mutual learning and diffusion models to iteratively improve restoration under low-light conditions.

• **Joint 3D Face Reconstruction.** With the advancement in 3D technology, there has been a growing interest in achieving 3D face reconstruction from LQ face images or recovering reconstructed LR 3D faces. R3DPFH [199] focuses on predicting corresponding HQ 3D face meshes from LR faces containing noise. Utilizing the Lucas-Kanade algorithm, Qu *et al.* [123] aim to improve the accuracy of 3D model fitting. Furthermore, Li *et al.* [87] and Uddin *et al.* [136] utilize 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 [195] directly reconstructs HQ faces from LQ faces by learning to recover fine-grained 3D details on the proxy image.

4 FACE PRIOR TECHNOLOGY

Unlike image restoration, face restoration requires maintaining the coherence and naturalness of facial features, especially in restoring key areas such as the eyes and mouth. In contrast, image restoration typically focuses more on overall consistency and detail recovery, without necessarily emphasizing the fine restoration of specific parts. Based on this, using appropriate face priors becomes particularly important, as it helps the model better understand the structure of face images and effectively restore damaged areas. Therefore, 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 existing priors in Fig. 13 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 [177, 202] utilize both priors. These methods are categorized based on the specific type of prior they primarily focus on.

Table 4. An overview of deep learning-based representative blind/non-blind and joint FR methods.

Methods	Publication	Prior	Task	Improved Technology	Highlight
BCCNN [200]	AAAI 2015	Plain	Non-Blind Task	Plain	Bi-channel CNN
ATMFN [69]	TMM 2019			GAN+Attention	Channel Attention Mechanism
SPARNet [16]	TIP 2020			Attention	Spatial Attention Mechanism
MSG-GAN [71]	CVPR 2020			GAN	Multi-Scale GAN
SFMNet [141]	CVPR 2023			Attention	Self-attention / Fourier
IDM [45]	CVPR 2023			Diffusion Model	Diffusion Probabilistic Models
CTCNet [6]	TIP 2023			Attention	Self-attention / Spatial Attention
WFEN [93]	MM 2024			Attention	Self-Attention Mechanism
DTWSR [34]	ICCV 2025			Diffusion Model	Wavelet-based Diffusion
LCGE [131]	IJCAI 2017	Internal Proprietary Prior		Prior	Facial Components Prior
FSRGFCH [169]	ECCV 2018			GAN+Prior	Facial Heatmap / Components Prior
AEDN [170]	CVPR 2018			GAN+Prior	Given Facial Attribute Prior
FSRNet [20]	CVPR 2018			Prior	Facial Landmarks / Parsing Maps Prior
FACN [159]	AAAI 2020			Prior	Estimated Facial Attribute Prior
DIC [112]	CVPR 2020			Attention+Prior	Facial Landmarks / Components Prior
PAP3D [61]	TPAMI 2021			Attention+Prior	3D Facial Prior / Spaital Attention
HCRF [107]	TIP 2021			Prior	Facial Semantic Labels Prior
GWAInet [32]	CVPR 2019	External Compensatory Prior		GAN+Prior	High Quality Images As Reference Prior
KDFSRNet [140]	TCSVT 2022			Prior	Pre-trained Teacher's Knowledge As Generative Prior
LRGAN [10]	ECCV 2018	Plain	Blind Task	GAN	Unsupervised / Two-Stage GAN
HiFaceGAN [161]	MM 2020			GAN	Semantic-Guided Generation
GCFSR [52]	CVPR 2022			GAN	Generative And Controllable Framework
SCGAN [56]	TIP 2023			GAN	Unsupervised / Semi-Cycled GAN
DR2 [156]	CVPR 2023			Diffusion Model	Diffusion-based Robust Degradation Remover
IDDM [197]	ICCV 2023			Diffusion Model	Iteratively Learned System
Super-FAN [9]	CVPR 2018	Internal Proprietary Prior		GAN+Prior	Facial Heatmap Prior
MDCN [129]	IJCV 2020			GAN+Prior	Semantic labels Prior
UMSN [166]	TIP 2020			GAN+Prior	Facial Components Prior
PSFRGAN [17]	CVPR 2021			GAN+Prior	Facial Parsing Maps Prior
SGPN [202]	CVPR 2022			GAN+Prior	3D Faical / Pre-trained Generative Prior
GFRNet [97]	ECCV 2018	External Proprietary Prior		GAN+Prior	High Quality Images As Reference Prior
PULSE [114]	CVPR 2020			GAN+Prior	Pre-trained StyleGAN's Generative Prior
ASFFNet [96]	CVPR 2020			GAN+Attention+Prior	Reference / Landmark Prior
GFPGAN [147]	CVPR 2021			GAN+Prior	Pre-trained StyleGAN's Generative Prior
VQFR [51]	ECCV 2022			GAN+Prior	Pre-trained VQGAN's Codebook Prior
CodeFormer [201]	NIPS 2022			GAN+Prior	Pre-trained VQGAN's Codebook Prior
DMDNet [98]	TPAMI 2023			GAN+Prior	Faical Component Dictionaries Prior / Reference Prior
PGDiff [162]	NIPS 2023			Diffusion Model	Pre-trained Diffusion Prior
DAEFR [134]	ICLR 2024			GAN+Prior	Pre-trained VQGAN's Codebook Prior
gen2res [31]	CVPR 2024			Diffusion+Prior	Pre-trained Diffusion Prior
DifFace [178]	TPAMI 2024			Diffusion+Prior	Pre-trained Diffusion Prior
PMRF [118]	ICLR 2025			Diffusion+Prior	Pre-trained Diffusion Prior / Flow Matching
MGFR [133]	ICLR 2025			Diffusion+Prior	Pre-trained Diffusion Prior / HQ Reference
SSDiff [94]	NIPS 2025			Diffusion+Prior	Pre-trained Diffusion Prior
TDN [172]	AAAI 2017	Plain	Joint Task	GAN	CNN / Joint Face Alignment
TANN [175]	TPAMI 2019			GAN	CNN / Joint Face Frontalization
FSRDENet [138]	AAAI 2024			Diffusion Model	Diffusion / Joint Illumination Compensation
CCPI [188]	TIP 2025			GAN	CNN / Joint Face Completion
SICNN [183]	ECCV 2018	Internal Proprietary Prior		Prior	Identity Prior / Joint Face Recognition
FCSR-GAN [11]	TBIOM 2020			GAN+Prior	Landmark/ Semantic labels / Joint Face Compensation
ID-GAN [168]	TCSVT 2020			GAN+Prior	Semantic labels / Identity Prior / Joint Face Recognition
SiGAN [58]	TIP 2019			GAN+Prior	Identity Prior / Joint Face Recognition
MDFR [135]	TCSVT 2021			GAN+Prior	Landmark/ 3D Facial Prior / Joint Face Frontalization
FT-TDR [142]	TMM 2022			GAN+Attention+Prior	Landmark Prior/ Self-Attention / Joint Face Completion
L2R [195]	CVPR 2022			GAN+Prior	Generative / 3D Prior / Joint 3D Face Reconstruction
FVIP [165]	TIP 2023			GAN+Prior	3D Face Prior / Joint Face Completion
CPGAN [192]	CVPR 2020	External Proprietary Prior		GAN+Attention+Prior	Reference Prior / Joint Illumination Compensation
ViVidGAN [190]	TIP 2021			GAN+Attention+Prior	Reference Prior / Joint Illumination Compensation

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 restoration. Standard techniques employed in this approach include identity recognition, facial attribute description, 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 recover facial features faithfully. Methods based on identity prior, such as SICNN [183], FH-GAN [7], IPFH [25], C-SRIP [48], 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 restoration. 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 [170], ATNet [90], ATSENet [91], AACNN [83], and others, directly connect the attribute information to the LQ image or its extracted features. Other methods, like AGCycleGAN [111] and FSRSA [170], 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 [158, 159] have been proposed. These approaches design appropriate attribute-based losses that enable the network to predict attribute information adaptively. RAAN [158] utilizes three branches to predict face shape, texture, and attribute information separately. It emphasizes either face shape or texture based on the attribute channel. FACN [159] 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 using the face's unique 2D geometric or 3D spatial information as priors. Facial landmarks [74, 112, 142] and facial heatmaps [139, 142, 177] 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 [112] 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 [17, 20, 176] and facial semantic labels [128, 129, 166] are more robust to severe degradation as they segment the face into regions. Even if some areas are severely degraded, intact regions can still guide restoration. Moreover, these priors contain more comprehensive facial information, enabling the restoration model to understand the overall facial structure and proportions better, leading to more coherent restorations. However, these priors may involve multiple semantic labels for different facial regions, requiring more complex networks [17, 166] to address semantic ambiguity. On the other hand, facial components [3, 109, 169] provide a straightforward representation of critical facial features, reducing the need for complex models while effectively guiding the restoration process. In addition to those above 2D facial priors, Hu *et al.* [60] introduce using a 3D face prior to handling faces with significant pose variations. Subsequent 3D prior-based methods [29, 165, 202] demonstrate their robustness in handling complex facial structures and significant pose changes. There are also methods [91, 194] that strive to achieve more comprehensive restoration by synergistically combining multiple internal proprietary priors.

• **Discuss.** Internal proprietary priors primarily rely on geometric or attribute properties intrinsic to the face itself. Consequently, they can be employed to recreate face images more authentically. Nevertheless, accurately estimating and effectively utilizing such priors remains a challenge, particularly in scenarios characterized by extremely blurred facial features.

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 restoration. These external priors can take various forms, including reference priors, dictionary priors, and pre-trained generative priors.

Reference prior-based methods [32, 96, 97] utilize HQ face images of the same individual as a reference to enhance the restoration of a target face image. The challenge lies in handling reference faces with varying poses and lighting conditions. GFRNet [97] is the pioneering work, which employs a WarpNet, coupled with a landmark loss, to rectify pose and expression disparities present in reference faces. This enables models to effectively utilize reference faces that exhibit differences compared to those undergoing restoration. GWAInet [32] utilizes the structure of generative networks of GAN and achieves favorable results without relying on facial landmarks. Subsequently, ASFFNet [96] further enhances performance by refining the selection of guide faces and improving the efficiency of feature fusion between guide faces and inputs. MGFR [133] enhances fidelity by integrating attribute text prompts, HQ reference, and identity information.

However, the above methods require reference images, limiting their applicability in various scenarios. To address this limitation, DFDNet [95] employs a strategy that creates a facial component dictionary. Initially, an HQ face dataset categorizes a dictionary comprising facial elements such as eyes, nose, and mouth. 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 [154] and RestoreFormer++ [155] integrate the Transformer and leverage the face component loss to more effectively utilize the potential of the facial component dictionary. DMDNet [98] leverages external facial reference 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 [36, 72, 73] can automatically extract information beyond facial features, including texture, hair details, and more. This makes methods based on pre-trained generative priors simpler. PULSE [114] is a pioneering breakthrough in FR that utilizes a generative prior. It identifies the most relevant potential vectors in the pre-trained GAN feature domain for input LQ faces. Subsequently, mGANprior [50] enhances the PULSE method by incorporating multiple potential spatial vectors derived from the pre-trained GAN. However, these methods are complex and may need help to ensure fidelity in restoration while effectively leveraging input facial features. Approaches like GLEAN [15], GPEN [164], and GFPGAN [147] integrate a pre-trained GAN into their customized networks. They employ GANs' generative prior to guide the forward process of networks, effectively leveraging input facial features and leading to improved fidelity in restoration. Subsequent techniques [57, 62, 99, 148, 149] 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 [202] incorporates a 3D shape prior along with the generative prior to enhancing restoration, combining both spatial and structural information.

Apart from approaches based on pre-trained StyleGAN [72, 73], there is another category of methods built upon pre-trained VQGAN [36]. The critical advantage of VQGAN lies in its utilization of a vector quantization mechanism, enabling accurate manipulation of specific features within the generated face images. Additionally, its training is more stable than some of StyleGAN's variants. VQFR [51] leverages discrete codebook vectors from VQGAN, using optimally sized compression patches and a parallel decoder to improve detail and fidelity in the restored outcomes. Codeformer [201] 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.* [198] explore the utilization of pre-training priors, aiming to strike a harmonious equilibrium between generation and restoration aspects. DAEFR [134] additionally introduces an auxiliary LQ codebook branch with information extracted from LQ inputs to collaborate with the restoration branch. In addition, DifFace [178] utilizes a pre-trained diffusion model for the first time to assist in blind face restoration. Subsequently, PGDiff [162] and SSDiff [94] further improve the performance of such methods by bootstrapping the diffusion prior. BFRffusion [19], DiffBIR [102], InterLCM [92] and OSDFace [143] improve performance by fine-tuning the prior of the large diffusion model.

• **Discuss.** External compensatory prior is well-suited for offering a robust solution in scenarios with more pronounced face image degradation, successfully restoring clear faces in the majority of cases. However, it does not guarantee the fidelity of the restored face images. Consequently, when employing this type of prior, it is essential to fully leverage the input face features to guide the external prior to enhancing the fidelity of the results.

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 priors are suitable for 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 by incorporating multiple priors during face restoration, leveraging the flexible complementarity of various prior information. MFPSNet [177] utilizes multiple priors, including face parsing maps, face landmarks, and a 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 [20, 91, 98, 166] use either multiple internal proprietary priors or external compensating priors. For example, UMSN [166] employs face semantic labels and facial components as priors. DMDNet [98] utilizes both facial dictionaries and external reference faces. Additionally, some methods [147, 154, 202] combine internal proprietary priors with external compensating priors. For instance, SGPN [202] leverages a 3D face shape prior alongside a pre-trained GAN prior. FREx [22] combines structure-accurate 3D priors and texture-rich 2D priors, enabling the restoration of natural-looking faces even under severe degradation and extreme poses. It is worth noting that individual prior combinations also affect the restoration results. Combining internal proprietary prior with each other, *e.g.*, the combination of face landmarks and parsing maps facilitates the continuous adjustment of facial landmarks' positions according to the facial semantic regions, which can more accurately localize damaged facial features and faithfully deal with facial changes. The combination of an internal proprietary prior and an external compensatory prior allows the estimated face geometry prior to being utilized to adjust for unfaithful face features introduced by the externally compensated 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 [20, 112] primarily focused on utilizing simple residual block structures for prior fusion, although these structures were not always optimal solutions. Subsequently, some methods [51, 96, 147, 201] aimed to design more efficient networks for prior fusion or estimation to enhance restoration performance. RestoreFormer [154] 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 [96] enhances the fusion of prior information with facial semantic features through a specially crafted adaptive spatial feature fusion block. VQFR [51] employs a parallel decoder structure to blend the generated prior information

Methods	CelebA				Helen			
	PSNR↑	SSIM↑	LPIPS↓	FID↓	PSNR↑	SSIM↑	LPIPS↓	FID↓
FSRNet [20]	27.05	0.7714	0.2127	170.4	25.45	0.7364	0.3090	228.8
FACN [159]	27.22	0.7802	0.1828	167.7	25.06	0.7189	0.3113	218.0
SPARNet [16]	27.73	0.7949	0.1995	161.2	26.43	0.7839	0.2674	211.5
DIC [112]	27.42	0.7840	0.2129	166.5	26.15	0.7717	0.2158	214.1
SISN [110]	27.91	0.7971	0.2005	162.3	26.64	0.7908	0.2571	210.7
SwinIR [100]	27.88	0.7967	0.2001	163.2	26.53	0.7856	0.2644	213.2
CTCNet [44]	**28.37**	**0.8115**	**0.1702**	156.9	**27.08**	0.8007	0.2094	205.8
SCTANet [6]	28.26	0.8100	0.1710	156.8	27.01	**0.8068**	0.1901	203.3
SFMNet [141]	27.85	0.7967	0.1837	**156.5**	26.98	0.8049	**0.1865**	**199.5**
WFEN [93]	28.04	0.8032	0.1803	**156.5**	27.01	0.8051	0.2148	203.6
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 at the scale of ×8. In this paper, the best and the second-best values are highlighted and underlined respectively.

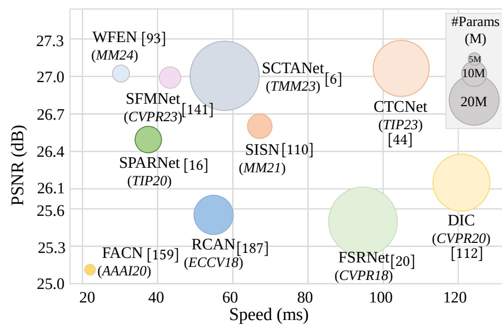
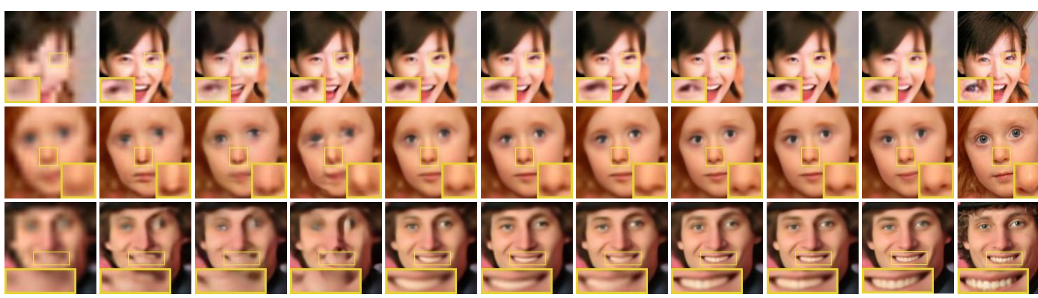


Fig. 14. Complexity analysis of non-blind methods on Helen test set.



Synthetic Input RCAN [187] FSRNet [20] FACN [159] SPARNet [16] SISN [110] SwinIR [100] CTCNet [44] SCTANet [6] SFMNet [141] Grouth-truth

Fig. 15. Visual comparison of different non-blind methods on the CelebA and Helen test sets.

with low-level features, ensuring enhanced fidelity without compromising the quality of the prior guidance. Overall, exploring the model structure based on the model improves the model performance to a certain extent, but brings limited performance improvement for scenarios such as non-positive faces and highly blurred.

• **Prior Guide Approach.** 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 [17] aims to enable the model to leverage the raw input information more effectively by directly estimating the prior knowledge from the LQ facial images to guide the restoration. In contrast, FSRNet [20] partially restores the LQ faces before estimating the prior to address inaccuracies in prior knowledge estimation. JASRNet [168] adopts a bootstrapping structure with parallel communication to leverage the interaction between prior estimation and restoration fully. Furthermore, CHNet [109] modifies the process of estimating priors by estimating 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 ***Appendix***.

• **Discuss.** Each strategy above for augmenting the prior validity of faces comes with its own advantages and disadvantages, requiring consideration based on the specific scenarios. For instance, combining multiple priors is beneficial for performance but may not be optimal for the practical deployment of the model. The efficient network structure needs to be carefully designed, and different degrees of blurred face images may be suitable for distinct prior guide approaches.

5 METHODS ANALYSIS

In recent years, face restoration methods have focused on two subtasks: face super-resolution and blind face restoration. In this section, we evaluate existing non-blind and blind methods for

Method	FSRNet [20]	FACN [159]	SPARNet [16]	DIC [112]
Params	27.5M	4.4M	10.6M	22.8M
MACs	40.7G	12.5G	7.1G	35.5M
Speed	89ms	22ms	40ms	122ms
SISN [110]	CTCNet [44]	SCTANet [6]	SFMNet [141]	WFEN [93]
9.8M	22.4M	27.7M	8.6M	6.8M
2.3G	47.2G	10.4G	30.6G	7.5M
68ms	106ms	58ms	48ms	31ms

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

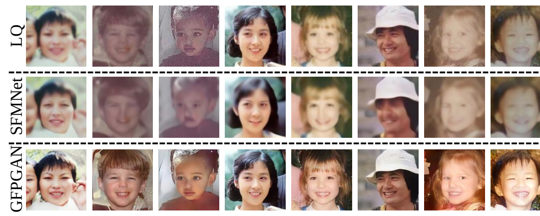


Fig. 16. Comparison of non-blind/blind face restoration methods in real-world scenarios.

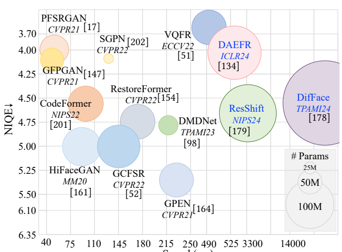


Fig. 17. Complexity analysis of blind FR methods on the synthetic face test set CelebA-HQ.

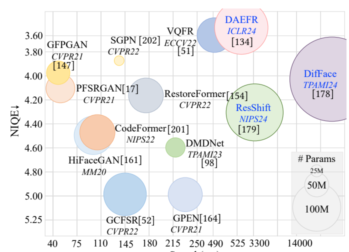


Fig. 18. Complexity analysis of blind FR methods on the real face test set LFW-Test.

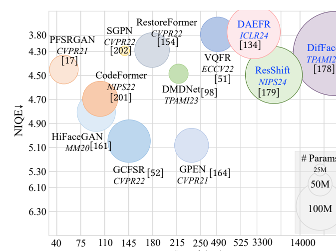


Fig. 19. Complexity analysis of blind FSR methods on the synthetic face test set CelebA-HQ (×8).

these tasks, with additional comparisons provided in the ***Appendix***. Given the broad scope of joint tasks, their comparisons are also included there. Methods presented in all performance comparison Tables of this survey are sorted chronologically from oldest to newest.

5.1 Experimental Setting

5.1.1 Non-blind Task. We utilize the initial 18,000 face images from the CelebA dataset for training. For testing, we randomly select 1,000 faces from the CelebA dataset and 50 random faces from the Helen dataset. All face images are cropped and resized to a size of 128×128. LQs are derived by downsampling HQs with a factor of ×8 using bicubic interpolation, as described in Eq.3. Models are optimized using the Adam optimizer for 100 epochs, with an initial learning rate of 2e-4.

5.1.2 Blind Task. We conduct training on the FFHQ dataset, which consists of 70,000 high-quality face images. We follow the degradation model used in GFPGAN [147] to get synthesized LQ face images. The degradation process is defined by Eq.8 and Eq.6, which represent blind restoration and blind super-resolution, respectively. In Eq.6, and 8, the standard deviation σ in Gaussian blur kernels, noise intensity δ, downsample scale s, and JPEG quality scale q in degradation models are randomly drawn from ranges $\{0.2 : 10\}$, $\{1 : 8\}$, $\{0 : 20\}$, and $\{60 : 100\}$, respectively. Models are trained with the Adam optimizer for 800k iterations, starting with a learning rate of 2e-3, which was then decayed by a factor of 2 at the 700k-th and 750k-th iterations.

We conduct testing on the CelebA-HQ dataset, which is a synthetic dataset with 3,000 face images from its testing partition, and the LQ images were generated in the same way as during training. Furthermore, we incorporate real-world face test sets such as LFW-Test with 1,711 LQ face images, WebPhoto-Test with 407 LQ face images, CelebChild with 180 LQ face images, and CelebAdult with 180 LQ face images to ensure a more comprehensive evaluation. All face images in training and testing are aligned and resized to a size of 512×512.

Table 7. Comparison of primary blind method performance on the synthetic test set CelebA-HQ and real datasets LFW-Test, WebPhoto, Celeb-Child, and Celeb-Adult. Speed and overhead are measured on 512×512 images. Methods presented in Table 5, Table 7, and Table 8 are sorted chronologically from oldest to newest.

Methods	Param	MACs	Speed	CelebA-HQ					LFW-Test		WebPhoto		CelebChild		CelebAdult	
				PSNR↑	LPIPS↓	IDD↓	FID↓	NIQE↓	FID↓	NIQE↓	FID↓	NIQE↓	FID↓	NIQE↓	FID↓	NIQE↓
HiFaceGAN [161]	79.9M	40.7G	90ms	24.92	.4770	.7310	66.09	5.002	64.50	4.520	116.1	4.943	113.0	4.871	104.0	4.340
DFDNet [95]	133.3M	599.8G	2.1s	24.26	.4421	.6884	54.34	5.921	59.69	4.776	93.28	5.812	107.1	4.452	105.6	3.782
PSFRGAN [17]	60.2M	464.9G	53ms	24.65	.4199	.6664	43.33	4.099	49.53	4.095	84.98	4.151	106.6	4.670	104.1	4.246
GPEN [164]	71.1M	138.1G	235ms	25.59	.4009	.6019	**36.46**	5.364	57.00	5.071	101.3	6.326	112.1	4.945	110.8	4.362
GFPGAN [147]	48.7M	51.6G	**46ms**	25.08	.3646	.5709	42.59	4.158	50.04	3.965	87.13	4.228	111.4	4.447	105.0	4.033
VQFR [51]	71.8M	1.07T	495ms	24.14	.3515	.5959	41.29	**3.693**	50.65	3.590	75.41	**3.608**	105.2	**3.938**	105.0	3.756
GCFSR [52]	88.7M	119.8G	145ms	**26.31**	**.3400**	**.5122**	50.10	4.943	52.23	4.998	93.27	5.640	115.1	5.326	107.1	4.824
SGPN [202]	**15.2M**	**18.3G**	134ms	24.93	.3702	.6028	39.44	4.095	44.95	3.863	75.61	4.269	109.4	4.234	104.9	4.402
RestoreFor [154]	72.4M	340.8G	172ms	24.64	.3655	.5339	41.82	4.405	48.38	4.169	77.33	4.459	**101.2**	4.580	103.5	4.321
CodeFormer [201]	73.6M	292.4G	98ms	25.15	.3432	.6171	52.43	4.650	52.36	4.484	83.19	4.705	116.2	4.983	111.1	4.541
DMDNet [98]	40.4M	187.2G	219ms	25.62	.3670	.6179	39.94	4.786	43.38	4.617	88.55	5.154	114.2	4.884	114.2	4.884
PGDiff [162]	47.7M	127.5G	16.3s	22.69	.4134	.9606	47.79	4.859	48.39	3.894	96.06	5.117	121.0	5.070	103.9	4.774
DAEFR [134]	111.2M	449.5G	627ms	23.07	.3697	.7602	40.73	3.816	47.53	**3.552**	75.38	4.041	105.7	4.072	101.7	**3.752**
ResShift [179]	118.6M	5491G	3.3s	25.82	.3435	.5545	43.76	4.375	52.38	4.300	**74.74**	4.451	108.1	4.635	106.4	4.337
DifFace [178]	175.4M	268.8G	14.4s	24.76	.3994	.7656	37.88	4.286	46.33	4.019	80.49	4.361	104.7	4.198	**97.6**	3.921
DiffBIR [102]	1717M	24234G	13.2s	25.38	.3878	.5367	58.16	6.083	**40.32**	5.737	91.83	6.069	118.9	5.549	108.8	5.651
Input	-	-	-	23.35	.4866	.8577	144.0	13.23	137.6	11.02	170.1	12.7	144.4	9.03	118.3	7.56

5.1.3 Evaluation Metric. We employ full 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 utilize non-reference or semi-reference metrics like NIQE and FID. These metrics allow us to evaluate image fidelity and visual quality without ground truth.

5.2 Experimental Results

5.2.1 Non-blind Task. Regarding the non-blind task, we choose 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 [100, 187], methods based on attention mechanisms [6, 44, 93, 110, 141], and methods relying on various priors [20, 112, 159]. Among these, methods employing hybrid attention mechanisms, namely CTCNet [44], SCTANet [6], and SFMNet [141], 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. 14 visually illustrates the efficiency of these techniques through three perspectives: performance, inference speed, and model size. Notably, attention-based methods, particularly SFMNet [141] and WFEN [93], stand out as they achieve superior performance while maintaining smaller computations. Finally, Fig. 15 provides a visual comparison of these methods. Attention-based methods also recovered the clearest results compared to other types of methods.

5.2.2 Blind Task. A range of state-of-the-art methods are selected for the blind task, including approaches that do not rely on prior knowledge, such as network architecture design [52, 161] and diffusion modeling techniques [126]). Additionally, techniques utilizing internally-specific priors such as parsing maps [17] and 3D face shapes [202] are considered. Furthermore, methods employing external compensatory prior like pre-trained StyleGAN prior [147, 164, 202], pre-trained VQGAN prior [51, 134, 201], pre-trained diffusion prior [102, 162, 178, 179], face dictionary [98, 154], and reference prior [98]) are also included.

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

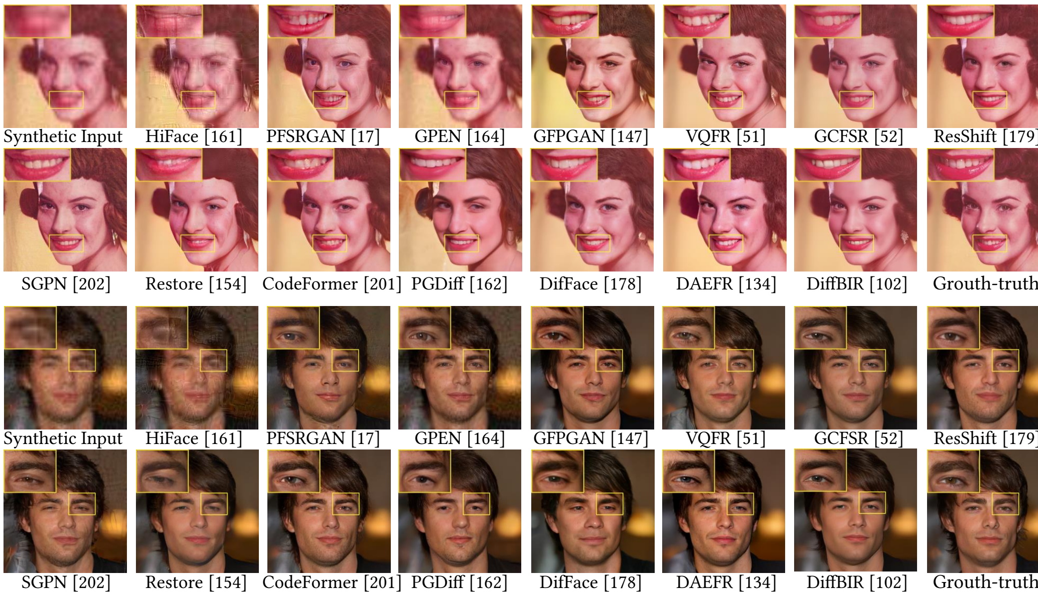


Fig. 20. Visual comparison of different blind methods on the CelebA-HQ test set for blind face restoration.

techniques across three dimensions: model size, inference speed, and performance on the synthetic dataset CelebA-HQ. It can be observed that GCFSR [52] achieves the best performance in structural similarity metrics of restored face images. Regarding visual and perceptual quality, pre-trained GAN-based [51, 164] and diffusion-based methods [178] exhibit superior performance. Methods such as DMDNet [154], SGPN [202], and GPEN [164] 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. 17 illustrates the efficiency trade-offs of these methods on the synthetic. This figure considers methods closer to the upper-left corner with smaller circles as more efficient. The figure demonstrates that both PFSRGAN [17] and pre-trained GAN-based methods [147, 201, 202] are more efficient, whereas diffusion-based approaches [102, 178, 179] generally exhibit a poor efficiency. Comparative visualization in Fig. 20 shows that methods relying on pre-trained GAN [147, 201] or diffusion priors [102] tend to achieve superior performance when dealing with severely degraded facial images. Finally, as depicted in Fig. 22, we select 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 image quality. It is evident that some methods [154, 178, 202], while exhibiting better sensory quality, show subpar performance in two metrics, such as SSIM and IDD. On the other hand, methods [52, 201] with higher structural and identity similarity often display inferior perceptual quality. Therefore, the community needs to develop more balanced approaches to address these disparities.

TABLE 7 also provides the performance of the blind method on four real-world datasets: LFW-Test, WebPhoto, CelebChild, and CelebAdult. As shown in this table, the methods VQFR [51] and DAEFR [28], which are based on the VQGAN-generated prior, achieved most of the best and second-best results. In addition, as shown in Fig. 21, in the face of more complex real-world scenarios, many of the blind methods have serious distortions. The ones that perform better are CodeFormer [201] and DAEFR [28], which are based on the VQGAN generating prior, and ResShift [179], which is based on the diffusion prior. It is important to note that, although the latest

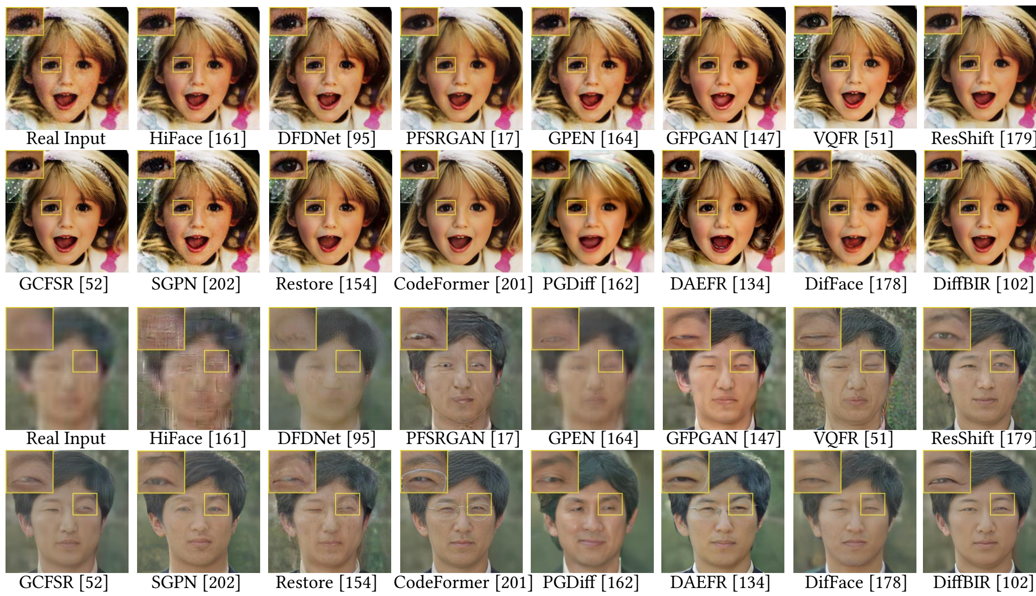


Fig. 21. Qualitative comparison of restoration for the real test sets, including Celeb-Child and WebPhoto-Test.

diffusion-based methods [102, 178, 179] don't achieve the highest metrics on all benchmarks, they are primarily designed to handle extremely blurred scenarios. As shown in the second image of Fig. 21, diffusion-based methods exhibit greater robustness in the extreme conditions, demonstrating a lower susceptibility to distortion. However, existing benchmarks encompass a broad range of degradations without prioritizing any particular type, meaning the metrics may struggle to reflect the advantages of the latest diffusion methods. Furthermore, we give the efficiency trade-off in Fig. 18, this efficiency comparison show GFPGAN [147] and PFSRGAN [17] are relatively efficient methods. Therefore, we need to choose the reasonably appropriate methods for actual needs.

The second part focuses on blind super-resolution, and TABLE 8 provides a quantitative performance comparison of these methods at three scales: ×4, ×8, and ×16. Prior-free methods like GCFSR [52] excel in face structure. However, they exhibit shortcomings in FID and NIQE metrics, suggesting that their restored faces might lack realism and may contain artifacts. On the other hand, pre-trained GAN-based and diffusion-based approaches such as VQFR [51], DAEFR [134], and DifFace [178] perform better in these two metrics, indicating more realistic and artifact-free results. Moving forward, Fig. 19 illustrates the efficiency of methods at the ×8 scale, with PFSRGAN [17] and SGPN [202] emerging as the more efficient choices. Lastly, in Fig. 23, we visually compare methods at three scales. Methods leveraging pre-trained GAN or diffusion priors perform favorably without introducing artifacts when dealing with substantial downsampling factors.

• **Performance Analysis.** Due to varying priorities in prior knowledge, no single method excels across all evaluation metrics. Approaches that achieve the highest fidelity scores tend to avoid generative priors, which can hinder the faithful reproduction of facial details. In contrast, methods utilizing generative priors, such as GAN or diffusion, perform better on visual metrics, with diffusion models being effective in handling severe blurring or profile-view scenarios. However, due to the diverse degradations in our data, the latest methods do not achieve optimal performance on all metrics. In summary, performance variations arise from trade-offs between priors and the balance between model size and restoration quality, with each method excelling in specific scenarios.

Table 8. Quantitative comparisons with blind methods on CelebA-HQ for ×4, ×8, ×16 super-resolution.

Methods	CelebA-HQ (×4)					CelebA-HQ (×8)					CelebA-HQ (×16)				
	PSNR↑	LPIPS↓	IDD↓	FID↓	NIQE↓	PSNR↑	LPIPS↓	IDD↓	FID↓	NIQE↓	PSNR↑	LPIPS↓	IDD↓	FID↓	NIQE↓
PSFRGAN [17]	27.99	.3055	.2924	42.35	4.623	25.50	.3639	.4445	47.56	4.446	23.20	.4216	.8603	49.31	4.197
HiFaceGAN [161]	29.49	.2736	**.2065**	39.72	4.535	**26.76**	.3496	**.3704**	51.32	4.830	23.68	.4746	1.012	92.31	5.836
DFDNet [95]	27.47	.3108	.2888	41.26	4.710	25.26	.3982	.4097	45.58	6.054	23.24	.4713	.9003	60.06	7.070
GPEN [164]	28.35	.2600	.2972	47.83	4.603	26.60	.3193	.4052	54.17	5.086	**24.12**	.3950	.8329	68.35	5.896
VQFR [51]	26.29	.2989	.3654	43.98	3.884	24.84	.3287	.4600	45.72	3.826	22.17	.3761	.8128	**38.42**	**3.431**
GCFSR [52]	**30.73**	**.2369**	.2132	52.02	4.915	26.66	.3073	.4146	54.74	5.059	22.90	.3799	.8564	46.99	4.622
SGPN [202]	28.64	.2456	.2581	41.06	4.425	26.18	**.3033**	.3846	44.69	4.330	23.65	**.3602**	.7506	46.66	4.444
RestoreFor [154]	24.99	.3353	.4146	41.38	4.392	24.65	.3495	.4525	41.66	4.340	22.60	.3974	.8068	38.45	4.216
CodeFormer [201]	27.10	.3020	.4462	51.30	4.739	25.75	.3229	.5115	51.42	4.698	23.26	.3666	.7776	48.69	4.496
DMDNet [98]	28.43	.2724	.3080	39.06	4.652	26.31	.3292	.3967	41.49	4.576	22.91	.3890	.8318	39.61	4.358
DAEFR [134]	23.11	.3525	.7109	39.86	**3.792**	23.02	.3614	.7283	41.09	**3.764**	22.20	.3864	.8331	42.55	3.739
ResShift [179]	28.60	.2741	.3195	43.79	4.698	26.72	.3081	.3908	44.89	4.532	23.29	.3736	.8023	42.46	4.166
DifFace [178]	25.26	.3838	.6882	**36.02**	3.995	24.86	.3917	.7031	**36.85**	4.163	23.57	.4200	.8181	42.01	4.487
DiffBIR [102]	27.21	.3296	.2854	49.44	5.769	25.89	.3628	.3895	49.64	5.829	23.28	.4218	**.7038**	61.11	5.939
Input	31.05	.2425	.2216	107.0	8.155	27.51	.3748	.5898	195.7	11.24	24.21	.5007	1.076	163.7	13.47

• **Discuss.** To evaluate the application scope of non-blind and blind methods, we randomly selected several real face photos. We restored them using a fine-tuned non-blind method, SFMNet [141], and a blind method, GFPGAN [147], respectively. As shown in Fig. 16, it is evident that SFMNet struggles to handle real-world photos effectively. In contrast, GFPGAN, despite showing some racial bias in certain images, generally offers superior visual quality. Consequently, the blind method holds greater promise for real-world applications. However, blind methods require more computational load than non-blind methods, which subsequent researchers will need to consider optimizing.

6 REAL-WORLD APPLICATION SCENARIOS

Face restoration methods have significant potential in enhancing visual quality across various fields, and their real-world applications continue to grow as the technology improves, including:

• **Security and Forensic Applications:** Face restoration plays a crucial role in both security surveillance and forensic investigations by enhancing the quality of facial images captured in low-resolution or degraded conditions. In security, it improves facial recognition and identification, boosting the effectiveness of surveillance systems. Similarly, in forensic science, face restoration aids in recovering facial features from unclear or obscured images, assisting in criminal investigations, identifying suspects or victims, and supporting cold cases.

• **Digital Entertainment:** In areas like gaming, animation, and movie remastering, face restoration is used for character generation and to restore facial features in old or low-quality video footage. This process enhances visual quality and realism, contributing to a more immersive and lifelike user experience. By improving the clarity of facial details, it elevates content quality, making it more engaging for audiences.

• **Social Media:** With the growing popularity of social media and Communication, face restoration techniques are also applied in applications that improve the quality of user-uploaded photos or videos, especially in situations where face images are taken under poor lighting conditions, low resolution, or with motion blur. However, it's important to note that facial restoration may be misidentified as a forgery by deepfake detection systems, so its use in these contexts should be approached with caution [2, 38, 145].

7 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 persist in this domain.

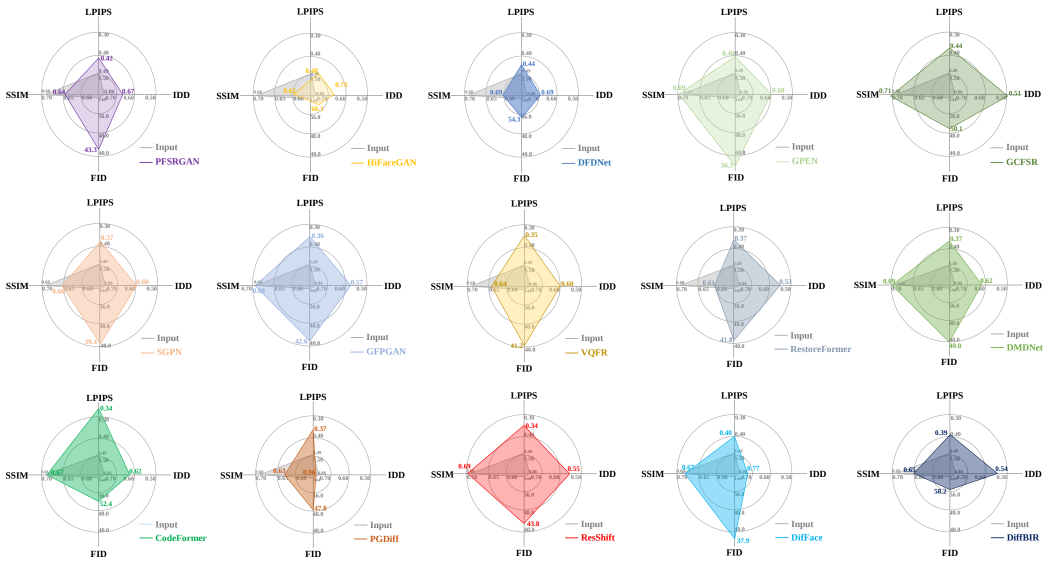


Fig. 22. A balanced analysis of multiple blind methods including PFSRGAN [17], HiFaceGAN [161], DFDNet [95], GPEN [164], GCFSR [52], SGPN [202], GFPGAN [147], VQFR [51], RestorFormer [154], DMDNet [98], CodeFormer [201], PGDiff [162], ResShift [179], DifFace [178], and DiffBIR [102] is performed through four main indicators: SSIM, LPIPS, FID, and IDD.

Additionally, numerous promising research opportunities exist 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 [76] (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. Consequently, there is a pressing demand in the industry for comprehensive, large-scale models capable of restoring a broad 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 selectively restore features such as hair, eyes, and skin. Language-based instructions can also 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.

• **Model Bias.** Most FR datasets, such as CelebA and FFHQ, collect facial images from specific geographical regions. It leads to the current trained models focusing on recovering facial features 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 creating 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

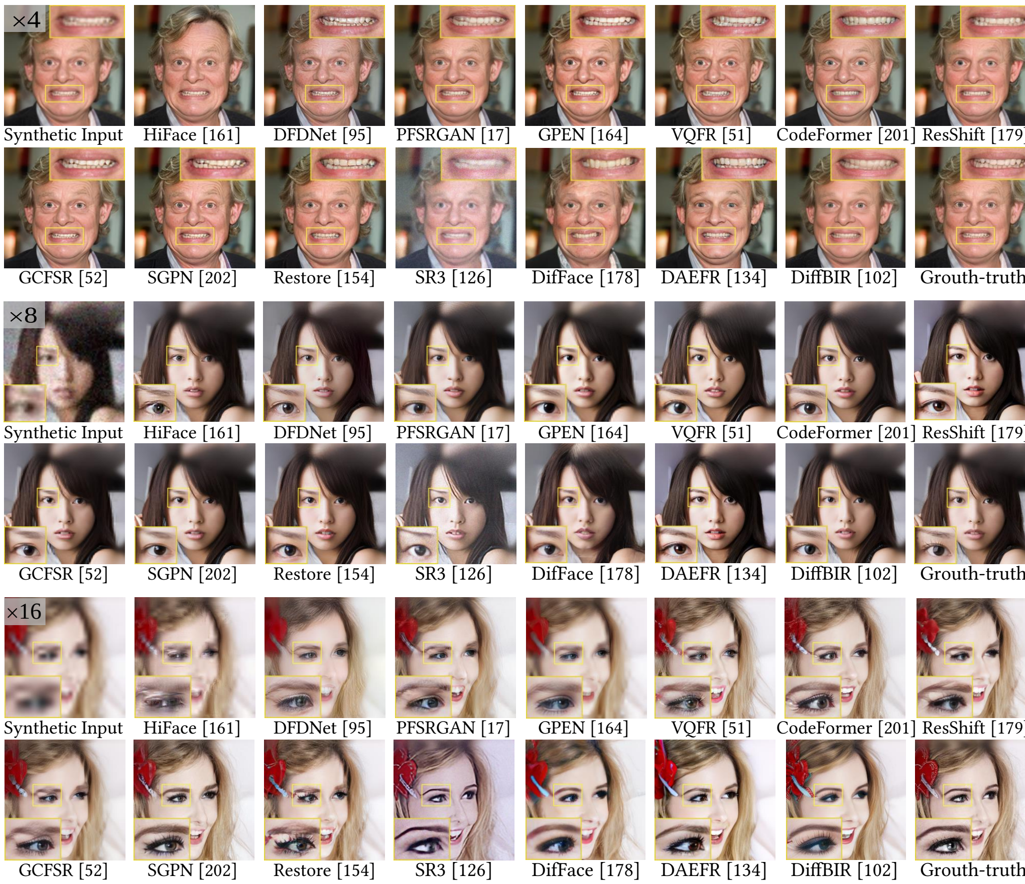


Fig. 23. Qualitative comparisons on CelebA-HQ for ×4, ×8, ×16 face super-resolution.

a risk of facial information leakage. This compassionate data is closely associated with financial transactions and access permissions. Unfortunately, current face restoration methods often ignore this aspect. Therefore, ensuring the protection of facial privacy during restoration remains an ongoing challenge and opportunity.

• **Practical Applications.** The challenges faced by facial restoration applications are two-fold: the disparity between synthetic and real data domains and significant computational costs. Compared to synthetic counterparts, real-world images undergo more complex forms of degradation. For example, most methods primarily focus on Gaussian blur kernels, while motion blur and spatially-varying blur caused by camera displacement a potential factors that could negatively impact these methods in real scenarios. Additionally, the computational overhead of existing methods is excessive for deployment on mobile devices. To address these challenges, research efforts should focus on developing realistic degradation models to model real-world degradation or add general blur kernels in training, exploring unsupervised restoration approaches to alleviate the reliance on large annotated datasets, and investigating model compression techniques to reduce computational costs. These endeavors will contribute to advancing practical applications [40].

• **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 [72, 106] may lack comprehensive coverage, limiting the models' generalization. Flawed loss functions may result in undesired artifacts in the restored faces. Existing network architectures [51, 147, 154] may not be suitable for all restoration tasks, limiting their applicability. Additionally, evaluating restoration results solely based on quantitative metrics may overlook essential 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.

8 CONCLUSION

In this review, we provide a systematic exploration of deep learning-based face restoration methods. We begin by discussing factors that contribute to the degradation of facial images and artificial degradations. 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 methods utilizing facial priors, including internal proprietary and external compensatory priors. We summarize the prevalent strategies for enhancing the effectiveness of priors in face restoration. Then, we thoroughly compare cutting-edge methods, highlighting their respective strengths and weaknesses, and giving application scenarios of face restoration. Finally, we dissect the prevailing challenges within existing paradigms and provide insights into potential directions for advancing the field. Overall, we aim to serve as a valuable reference for researchers who are starting their journey in developing techniques aligned with their research aspirations.

9 ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Grant 62301306, 62225601, U23B2052), the Science and Technology Commission of Shanghai Municipality under Grant 23YF1412800, Beijing Natural Science Foundation Project No. L242025, the Fundamental Research Funds for the Beijing University of Posts and Telecommunications under Grant 2025AI4S15.

REFERENCES

[1] Ali Abbasi and Mohammad Rahmati. 2021. Identity-Preserving Pose-Robust Face Hallucination Through Face Subspace Prior. *arXiv preprint arXiv:2111.10634* (2021).

[2] Wasim Ahmad, Yan-Tsung Peng, Yuan-Hao Chang, Gaddisa Olani Ganfure, and Sarwar Khan. 2025. Capst: Leveraging capsule networks and temporal attention for accurate model attribution in deep-fake videos. *ACM Transactions on Multimedia Computing, Communications and Applications* 21, 4 (2025), 1–23.

[3] Yu Bai, Ruian He, Weimin Tan, Bo Yan, and Yangle Lin. 2023. Fine-Grained Blind Face Inpainting with 3D Face Component Disentanglement. In *ICASSP*. IEEE, 1–5.

[4] Simon Baker and Takeo Kanade. 2000. Hallucinating faces. In *FG*. IEEE, 83–88.

[5] Qiqi Bao, Bowen Gang, Wenming Yang, Jie Zhou, and Qingmin Liao. 2022. Attention-driven graph neural network for deep face super-resolution. *IEEE Transactions on Image Processing* 31 (2022), 6455–6470.

[6] Qiqi Bao, Yunmeng Liu, Bowen Gang, Wenming Yang, and Qingmin Liao. 2023. SCTANet: A Spatial Attention-Guided CNN-Transformer Aggregation Network for Deep Face Image Super-Resolution. *IEEE Transactions on Multimedia* (2023).

[7] Bayram Bayramli, Usman Ali, Te Qi, and Hongtao Lu. 2019. FH-GAN: Face hallucination and recognition using generative adversarial network. In *ICONIP*. Springer, 3–15.

[8] Adrian Bulat and Georgios Tzimiropoulos. 2017. How far are we from solving the 2d & 3d face alignment problem?(and a dataset of 230,000 3d facial landmarks). In *ICCV*. 1021–1030.

[9] Adrian Bulat and Georgios Tzimiropoulos. 2018. Super-fan: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with gans. In *CVPR*. 109–117.

[10] Adrian Bulat, Jing Yang, and Georgios Tzimiropoulos. 2018. To learn image super-resolution, use a gan to learn how to do image degradation first. In *ECCV*. 185–200.

[11] Jiancheng Cai, Hu Han, Shiguang Shan, and Xilin Chen. 2019. FCSR-GAN: Joint face completion and super-resolution via multi-task learning. *IEEE Transactions on Biometrics, Behavior, and Identity Science* 2, 2 (2019), 109–121.

[12] Qingxing Cao, Liang Lin, Yukai Shi, Xiaodan Liang, and Guanbin Li. 2017. Attention-aware face hallucination via deep reinforcement learning. In *CVPR*. 690–698.

[13] Qiong Cao, Li Shen, Weidi Xie, Omkar M Parkhi, and Andrew Zisserman. 2018. Vggface2: A dataset for recognising faces across pose and age. In *FG*. IEEE, 67–74.

[14] Kelvin CK Chan, Xintao Wang, Xiangyu Xu, Jinwei Gu, and Chen Change Loy. 2021. Glean: Generative latent bank for large-factor image super-resolution. In *CVPR*. 14245–14254.

[15] Kelvin CK Chan, Xiangyu Xu, Xintao Wang, Jinwei Gu, and Chen Change Loy. 2022. GLEAN: Generative latent bank for image super-resolution and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 3 (2022), 3154–3168.

[16] Chaofeng Chen, Dihong Gong, Hao Wang, Zhifeng Li, and Kwan-Yee K Wong. 2020. Learning spatial attention for face super-resolution. *IEEE Transactions on Image Processing* 30 (2020), 1219–1231.

[17] Chaofeng Chen, Xiaoming Li, Lingbo Yang, Xianhui Lin, Lei Zhang, and Kwan-Yee K Wong. 2021. Progressive semantic-aware style transformation for blind face restoration. In *CVPR*. 11896–11905.

[18] Liang Chen, Jinshan Pan, Junjun Jiang, Jiawei Zhang, Zhen Han, and Linchao Bao. 2021. Multi-stage degradation homogenization for super-resolution of face images with extreme degradations. *IEEE Transactions on Image Processing* 30 (2021), 5600–5612.

[19] Xiaoxu Chen, Jingfan Tan, Tao Wang, Kaihao Zhang, Wenhan Luo, and Xiaochun Cao. 2024. Towards real-world blind face restoration with generative diffusion prior. *IEEE Transactions on Circuits and Systems for Video Technology* (2024).

[20] Yu Chen, Ying Tai, Xiaoming Liu, Chunhua Shen, and Jian Yang. 2018. Fsrnet: End-to-end learning face super-resolution with facial priors. In *CVPR*. 2492–2501.

[21] Zhibo Chen, Jianxin Lin, Tiankuang Zhou, and Feng Wu. 2019. Sequential gating ensemble network for noise robust multiscale face restoration. *IEEE Transactions on Cybernetics* 51, 1 (2019), 451–461.

[22] Zhengrui Chen, Liying Lu, Ziyang Yuan, Yiming Zhu, Yu Li, Chun Yuan, and Weihong Deng. 2024. Blind Face Restoration under Extreme Conditions: Leveraging 3D-2D Prior Fusion for Superior Structural and Texture Recovery. In *AAAI*, Vol. 38. 1263–1271.

[23] Fangfang Cheng, Tao Lu, Yu Wang, and Yanduo Zhang. 2021. Face super-resolution through dual-identity constraint. In *ICME*. 1–6.

[24] Shen Cheng, Yuzhi Wang, Haibin Huang, Donghao Liu, Haoqiang Fan, and Shuaicheng Liu. 2021. Nbnet: Noise basis learning for image denoising with subspace projection. In *CVPR*. 4896–4906.

[25] Xiaojuan Cheng, Jiwen Lu, Bo Yuan, and Jie Zhou. 2019. Identity-preserving face hallucination via deep reinforcement learning. *IEEE Transactions on Circuits and Systems for Video Technology* 30, 12 (2019), 4796–4809.

[26] Grigorios G Chrysos and Stefanos Zafeiriou. 2017. Deep face deblurring. In *CVPRW*. 69–78.

[27] Vishal Chudasama, Kartik Nighania, Kishor Upla, Kiran Raja, Raghavendra Ramachandra, and Christoph Busch. 2021. E-comsupresnet: Enhanced face super-resolution through compact network. *IEEE Transactions on Biometrics, Behavior, and Identity Science* 3, 2 (2021), 166–179.

[28] Trinetra Devkatte, Shiv Ram Dubey, Satish Kumar Singh, and Abdenour Hadid. 2025. Face to Cartoon Incremental Super-Resolution Using Knowledge Distillation. In *ICPR*. Springer, 99–114.

[29] Rahul Dey and Vishnu Naresh Boddeti. 2022. 3DFaceFill: An analysis-by-synthesis approach to face completion. In *WACV*. 1586–1595.

[30] Xin Ding and Ruimin Hu. 2020. Learning to see faces in the dark. In *ICME*. 1–6.

[31] Zheng Ding, Xuaner Zhang, Zhuowen Tu, and Zhihao Xia. 2024. Restoration by Generation with Constrained Priors. In *CVPR*. 2567–2577.

[32] Berk Dogan, Shuhang Gu, and Radu Timofte. 2019. Exemplar guided face image super-resolution without facial landmarks. In *CVPRW*. 0–0.

[33] Hao Dou, Chen Chen, Xiyuan Hu, Zuxing Xuan, Zhisen Hu, and Silong Peng. 2020. PCA-SRGAN: Incremental orthogonal projection discrimination for face super-resolution. In *ACMMM*. 1891–1899.

[34] Peng Du, Hui Li, Han Xu, Paul Barom Jeon, Dongwook Lee, Daehyun Ji, Ran Yang, and Feng Zhu. 2025. Diffusion Transformer meets Multi-level Wavelet Spectrum for Single Image Super-Resolution. In *ICCV*. 19700–19710.

[35] Qingyan Duan, Lei Zhang, and Xinbo Gao. 2021. Simultaneous face completion and frontalization via mask guided two-stage GAN. *IEEE Transactions on Circuits and Systems for Video Technology* 32, 6 (2021), 3761–3773.

[36] Patrick Esser, Robin Rombach, and Bjorn Ommer. 2021. Taming transformers for high-resolution image synthesis. In *CVPR*. 12873–12883.

[37] Yihua Fan, Yongzhen Wang, Dong Liang, Yiping Chen, Haoran Xie, Fu Lee Wang, Jonathan Li, and Mingqiang Wei. 2024. Low-facenet: face recognition-driven low-light image enhancement. *IEEE Transactions on Instrumentation and Measurement* 73 (2024), 1–13.

[38] Shuaijv Fang, Zhiyong Zhang, and Bin Song. 2025. Deepfake detection model combining texture differences and frequency domain information. *ACM Transactions on Privacy and Security* 28, 2 (2025), 1–16.

[39] Reuben A Farrugia and Christine Guillemot. 2017. Face hallucination using linear models of coupled sparse support. *IEEE Transactions on Image Processing* 26, 9 (2017), 4562–4577.

[40] Ruicheng Feng, Chongyi Li, and Chen Change Loy. 2024. Kalman-inspired feature propagation for video face super-resolution. In *ECCV*. Springer, 202–218.

[41] Zhanxiang Feng, Jianhuang Lai, Xiaohua Xie, Dakun Yang, and Ling Mei. 2016. Face hallucination by deep traversal network. In *ICPR*. IEEE, 3276–3281.

[42] Kanchana Vaishnavi Gandikota and Paramanand Chandramouli. 2024. Text-guided Explorable Image Super-resolution. In *CVPR*. 25900–25911.

[43] Guangwei Gao, Lei Tang, Fei Wu, Huimin Lu, and Jian Yang. 2023. JDSR-GAN: Constructing an Efficient Joint Learning Network for Masked Face Super-Resolution. *IEEE Transactions on Multimedia* 25 (2023), 1505–1512.

[44] Guangwei Gao, Zixiang Xu, Juncheng Li, Jian Yang, Tieyong Zeng, and Guo-Jun Qi. 2023. Ctcnet: A cnn-transformer cooperation network for face image super-resolution. *IEEE Transactions on Image Processing* 32 (2023), 1978–1991.

[45] Sicheng Gao, Xuhui Liu, Bohan Zeng, Sheng Xu, Yanjing Li, Xiaoyan Luo, Jianzhuang Liu, Xiantong Zhen, and Baochang Zhang. 2023. Implicit diffusion models for continuous super-resolution. In *CVPR*. 10021–10030.

[46] Shiming Ge, Chenyu Li, Shengwei Zhao, and Dan Zeng. 2020. Occluded face recognition in the wild by identity-diversity inpainting. *IEEE Transactions on Circuits and Systems for Video Technology* 30, 10 (2020), 3387–3397.

[47] Mislav Grgic, Kresimir Delac, and Sonja Grgic. 2011. SCface–surveillance cameras face database. *Multimedia Tools and Applications* 51 (2011), 863–879.

[48] Klemen Grm, Walter J Scheirer, and Vitomir Štruc. 2019. Face hallucination using cascaded super-resolution and identity priors. *IEEE Transactions on Image Processing* 29 (2019), 2150–2165.

[49] Ralph Gross, Iain Matthews, Jeffrey Cohn, Takeo Kanade, and Simon Baker. 2010. Multi-pie. *Image and Vision Computing* 28, 5 (2010), 807–813.

[50] Jinjin Gu, Yujun Shen, and Bolei Zhou. 2020. Image processing using multi-code gan prior. In *CVPR*. 3012–3021.

[51] Yuchao Gu, Xintao Wang, Liangbin Xie, Chao Dong, Gen Li, Ying Shan, and Ming-Ming Cheng. 2022. VQFR: Blind face restoration with vector-quantized dictionary and parallel decoder. In *ECCV*. Springer, 126–143.

[52] Jingwen He, Wu Shi, Kai Chen, Lean Fu, and Chao Dong. 2022. Gcfsr: a generative and controllable face super resolution method without facial and gan priors. In *CVPR*. 1889–1898.

[53] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. 2017. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *NeurIPS*, Vol. 30.

[54] Alain Hore and Djemel Ziou. 2010. Image quality metrics: PSNR vs. SSIM. In *ICPR*. IEEE, 2366–2369.

[55] Tobias Hoßfeld, Poul E Heegaard, Martín Varela, and Sebastian Möller. 2016. QoE beyond the MOS: an in-depth look at QoE via better metrics and their relation to MOS. *Quality and User Experience* 1 (2016), 1–23.

[56] Hao Hou, Jun Xu, Yingkun Hou, Xiaotao Hu, Benzheng Wei, and Dinggang Shen. 2023. Semi-cycled generative adversarial networks for real-world face super-resolution. *IEEE Transactions on Image Processing* 32 (2023), 1184–1199.

[57] Zhengzhang Hou, Liang Li, and Xiaojie Guo. 2022. Feature-Guided Blind Face Restoration with GAN Prior. In *ICME*. 1–6.

[58] Chih-Chung Hsu, Chia-Wen Lin, Weng-Tai Su, and Gene Cheung. 2019. Sigan: Siamese generative adversarial network for identity-preserving face hallucination. *IEEE Transactions on Image Processing* 28, 12 (2019), 6225–6236.

[59] Xiao Hu, Peirong Ma, Zhuohao Mai, Shaohu Peng, Zhao Yang, and Li Wang. 2019. Face hallucination from low quality images using definition-scalable inference. *Pattern Recognition* 94 (2019), 110–121.

[60] Xiaobin Hu, Wenqi Ren, John LaMaster, Xiaochun Cao, Xiaoming Li, Zechao Li, Bjoern Menze, and Wei Liu. 2020. Face super-resolution guided by 3d facial priors. In *ECCV*. Springer, 763–780.

[61] Xiaobin Hu, Wenqi Ren, Jiaolong Yang, Xiaochun Cao, David Wipf, Bjoern Menze, Xin Tong, and Hongbin Zha. 2021. Face restoration via plug-and-play 3d facial priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 12 (2021), 8910–8926.

[62] Yujie Hu, Yinhuai Wang, and Jian Zhang. 2023. Dear-gan: Degradation-aware face restoration with gan prior. *IEEE Transactions on Circuits and Systems for Video Technology* (2023).

[63] Gary B Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller. 2008. 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*.

[64] Huaibo Huang, Ran He, Zhenan Sun, and Tieniu Tan. 2017. Wavelet-srnet: A wavelet-based cnn for multi-scale face super resolution. In *ICCV*. 1689–1697.

[65] Huaibo Huang, Ran He, Zhenan Sun, and Tieniu Tan. 2019. Wavelet domain generative adversarial network for multi-scale face hallucination. *International Journal of Computer Vision* 127, 6-7 (2019), 763–784.

[66] Junjun Jiang, Chenyang Wang, Xianming Liu, and Jiayi Ma. 2021. Deep learning-based face super-resolution: A survey. *Comput. Surveys* 55, 1 (2021), 1–36.

[67] Junjun Jiang, Yi Yu, Jinhui Hu, Suhua Tang, and Jiayi Ma. 2018. Deep CNN denoiser and multi-layer neighbor component embedding for face hallucination. *IJCAI* (2018).
[68] Kui Jiang, Zhongyuan Wang, Peng Yi, Tao Lu, Junjun Jiang, and Zixiang Xiong. 2020. Dual-path deep fusion network for face image hallucination. *IEEE Transactions on Neural Networks and Learning Systems* 33, 1 (2020), 378–391.
[69] Kui Jiang, Zhongyuan Wang, Peng Yi, Guangcheng Wang, Ke Gu, and Junjun Jiang. 2019. ATMFN: Adaptive-threshold-based multi-model fusion network for compressed face hallucination. *IEEE Transactions on Multimedia* 22, 10 (2019), 2734–2747.
[70] Ratheesh Kalarot, Tao Li, and Fatih Porikli. 2020. Component attention guided face super-resolution network: Cagface. In *WACV*. 370–380.
[71] Animesh Karnewar and Oliver Wang. 2020. Msg-gan: Multi-scale gradients for generative adversarial networks. In *CVPR*. 7799–7808.
[72] Tero Karras, Samuli Laine, and Timo Aila. 2019. A style-based generator architecture for generative adversarial networks. In *CVPR*. 4401–4410.
[73] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2020. Analyzing and improving the image quality of stylegan. In *CVPR*. 8110–8119.
[74] Deokyun Kim, Minseon Kim, Gihyun Kwon, and Dae-Shik Kim. 2019. Progressive face super-resolution via attention to facial landmark. *arXiv preprint arXiv:1908.08239* (2019).
[75] Jonghyun Kim, Gen Li, Inyong Yun, Cheolkon Jung, and Joongkyu Kim. 2021. Edge and identity preserving network for face super-resolution. *Neurocomputing* 446 (2021), 11–22.
[76] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. 2023. Segment anything. In *ICCV*. 4015–4026.
[77] Martin Koestinger, Paul Wohlhart, Peter M Roth, and Horst Bischof. 2011. Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization. In *ICCVW*. 2144–2151.
[78] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. 2019. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *ICCV*. 8878–8887.
[79] Shun-Cheung Lai, Chen-Hang He, and Kin-Man Lam. 2019. Low-resolution face recognition based on identity-preserved face hallucination. In *ICIP*. IEEE, 1173–1177.
[80] Wei-Sheng Lai, Yichang Shih, Lun-Cheng Chu, Xiaotong Wu, Sung-Fang Tsai, Michael Krainin, Deqing Sun, and Chia-Kai Liang. 2022. Face deblurring using dual camera fusion on mobile phones. *ACM Transactions on Graphics* 41, 4 (2022), 1–16.
[81] Vuong Le, Jonathan Brandt, Zhe Lin, Lubomir Bourdev, and Thomas S Huang. 2012. Interactive facial feature localization. In *ECCV*. 679–692.
[82] Cheng-Han Lee, Ziwei Liu, Lingyun Wu, and Ping Luo. 2020. Maskgan: Towards diverse and interactive facial image manipulation. In *CVPR*. 5549–5558.
[83] Cheng-Han Lee, Kaipeng Zhang, Hu-Cheng Lee, Chia-Wen Cheng, and Winston Hsu. 2018. Attribute augmented convolutional neural network for face hallucination. In *CVPRW*. 721–729.
[84] Guanxin Li, Jingang Shi, Yuan Zong, Fei Wang, Tian Wang, and Yihong Gong. 2023. Learning Attention from Attention: Efficient Self-Refinement Transformer for Face Super-Resolution. *IJCAI* (2023).
[85] Honglei Li, Wenmin Wang, Cheng Yu, and Shixiong Zhang. 2021. SwapInpaint: Identity-specific face inpainting with identity swapping. *IEEE Transactions on Circuits and Systems for Video Technology* 32, 7 (2021), 4271–4281.
[86] Haoying Li, Yifan Yang, Meng Chang, Shiqi Chen, Huajun Feng, Zhihai Xu, Qi Li, and Yueting Chen. 2022. Srdiff: Single image super-resolution with diffusion probabilistic models. *Neurocomputing* 479 (2022), 47–59.
[87] Jiaxin Li, Feiyu Zhu, Xiao Yang, and Qijun Zhao. 2021. 3D face point cloud super-resolution network. In *IJCB*. IEEE, 1–8.
[88] Ke Li, Bahetiyaer Bare, Bo Yan, Bailan Feng, and Chunfeng Yao. 2018. Face hallucination based on key parts enhancement. In *ICASSP*. IEEE, 1378–1382.
[89] Kunjian Li and Qijun Zhao. 2020. If-gan: Generative adversarial network for identity preserving facial image inpainting and frontalization. In *FG*. IEEE, 45–52.
[90] Mengyan Li, Yuechuan Sun, Zhaoyu Zhang, Haonian Xie, and Jun Yu. 2019. Deep learning face hallucination via attributes transfer and enhancement. In *ICME*. 604–609.
[91] Mengyan Li, Zhaoyu Zhang, Jun Yu, and Chang Wen Chen. 2020. Learning face image super-resolution through facial semantic attribute transformation and self-attentive structure enhancement. *IEEE Transactions on Multimedia* 23 (2020), 468–483.
[92] Senmao Li, Kai Wang, Joost van de Weijer, Fahad Shahbaz Khan, Chun-Le Guo, Shiqi Yang, Yaxing Wang, Jian Yang, and Ming-Ming Cheng. 2025. InterLCM: Low-Quality Images as Intermediate States of Latent Consistency Models for Effective Blind Face Restoration. In *ICLR*.

[93] Wenjie Li, Heng Guo, Xuannan Liu, Kongming Liang, Jiani Hu, Zhanyu Ma, and Jun Guo. 2024. Efficient face super-resolution via wavelet-based feature enhancement network. In *ACM MM*. 4515–4523.
[94] Wenjie Li, Xiangyi Wang, Heng Guo, Guangwei Gao, and Zhanyu Ma. 2025. Self-Supervised Selective-Guided Diffusion Model for Old-Photo Face Restoration. In *NeurIPS*.
[95] Xiaoming Li, Chaofeng Chen, Shangchen Zhou, Xianhui Lin, Wangmeng Zuo, and Lei Zhang. 2020. Blind face restoration via deep multi-scale component dictionaries. In *ECCV*. Springer, 399–415.
[96] Xiaoming Li, Wenyu Li, Dongwei Ren, Hongzhi Zhang, Meng Wang, and Wangmeng Zuo. 2020. Enhanced blind face restoration with multi-exemplar images and adaptive spatial feature fusion. In *CVPR*. 2706–2715.
[97] Xiaoming Li, Ming Liu, Yuting Ye, Wangmeng Zuo, Liang Lin, and Ruigang Yang. 2018. Learning warped guidance for blind face restoration. In *ECCV*. 272–289.
[98] Xiaoming Li, Shiguang Zhang, Shangchen Zhou, Lei Zhang, and Wangmeng Zuo. 2023. Learning Dual Memory Dictionaries for Blind Face Restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2023).
[99] Zelin Li, Dan Zeng, Xiao Yan, Qiaomu Shen, and Bo Tang. 2023. Analyzing and Combating Attribute Bias for Face Restoration. *IJCAI* (2023).
[100] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. 2021. Swinir: Image restoration using swin transformer. In *ICCV*. 1833–1844.
[101] Songnan Lin, Jiawei Zhang, Jinshan Pan, Yicun Liu, Yongtian Wang, Jing Chen, and Jimmy Ren. 2020. Learning to deblur face images via sketch synthesis. In *AAAI*, Vol. 34. 11523–11530.
[102] Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Bo Dai, Fanghua Yu, Yu Qiao, Wanli Ouyang, and Chao Dong. 2024. Diffbir: Toward blind image restoration with generative diffusion prior. In *ECCV*. Springer, 430–448.
[103] Ce Liu, Heung-Yeung Shum, and Chang-Shui Zhang. 2001. A two-step approach to hallucinating faces: global parametric model and local nonparametric model. In *CVPR*, Vol. 1. IEEE, I–I.
[104] Heng Liu, Xiaoyu Zheng, Jungong Han, Yuezhong Chu, and Tao Tao. 2019. Survey on GAN-based face hallucination with its model development. *IET Image Processing* 13, 14 (2019), 2662–2672.
[105] Jie Liu and Cheolkon Jung. 2019. Facial image inpainting using multi-level generative network. In *ICME*. 1168–1173.
[106] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. 2015. Deep learning face attributes in the wild. In *ICCV*. 3730–3738.
[107] Zhi-Song Liu, Wan-Chi Siu, and Yui-Lam Chan. 2021. Features guided face super-resolution via hybrid model of deep learning and random forests. *IEEE Transactions on Image Processing* 30 (2021), 4157–4170.
[108] Tao Lu, Hao Wang, Zixiang Xiong, Junjun Jiang, Yanduo Zhang, Huabing Zhou, and Zhongyuan Wang. 2017. Face hallucination using region-based deep convolutional networks. In *ICIP*. IEEE, 1657–1661.
[109] Tao Lu, Yuanzhi Wang, Yanduo Zhang, Junjun Jiang, Zhongyuan Wang, and Zixiang Xiong. 2022. Rethinking prior-guided face super-resolution: a new paradigm with facial component prior. *IEEE Transactions on Neural Networks and Learning Systems* (2022).
[110] Tao Lu, Yuanzhi Wang, Yanduo Zhang, Yu Wang, Liu Wei, Zhongyuan Wang, and Junjun Jiang. 2021. Face hallucination via split-attention in split-attention network. In *ACM MM*. 5501–5509.
[111] Yongyi Lu, Yu-Wing Tai, and Chi-Keung Tang. 2018. Attribute-guided face generation using conditional cyclegan. In *ECCV*. 282–297.
[112] Cheng Ma, Zhenyu Jiang, Yongming Rao, Jiwen Lu, and Jie Zhou. 2020. Deep face super-resolution with iterative collaboration between attentive recovery and landmark estimation. In *CVPR*. 5569–5578.
[113] Joe Mathai, Iacopo Masi, and Wael AbdAlmageed. 2019. Does generative face completion help face recognition?. In *ICB*. IEEE, 1–8.
[114] Sachit Menon, Alexandru Damian, Shijia Hu, Nikhil Ravi, and Cynthia Rudin. 2020. Pulse: Self-supervised photo upsampling via latent space exploration of generative models. In *CVPR*. 2437–2445.
[115] Yunqi Miao, Jiankang Deng, and Jungong Han. 2024. WaveFace: Authentic Face Restoration with Efficient Frequency Recovery. In *CVPR*. 6583–6592.
[116] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. 2012. Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters* 20, 3 (2012), 209–212.
[117] Kien Nguyen, Clinton Fookes, Sridha Sridharan, Massimo Tistarelli, and Mark Nixon. 2018. Super-resolution for biometrics: A comprehensive survey. *Pattern Recognition* 78 (2018), 23–42.
[118] Guy Ohayon, Tomer Michaeli, and Michael Elad. 2024. Posterior-mean rectified flow: Towards minimum mse photo-realistic image restoration. In *ICLR*.
[119] Yun Pang, Jiawei Mao, Libo He, Hong Lin, and Zhenping Qiang. 2024. An Improved Face Image Restoration Method Based on Denoising Diffusion Probabilistic Models. *IEEE Access* (2024).
[120] Jeong-Seon Park and Seong-Whan Lee. 2008. An example-based face hallucination method for single-frame, low-resolution facial images. *IEEE Transactions on Image Processing* 17, 10 (2008), 1806–1816.

[121] Haoran Qi, Yuwei Qiu, Xing Luo, and Zhi Jin. 2023. An Efficient Latent Style Guided Transformer-CNN Framework for Face Super-Resolution. *IEEE Transactions on Multimedia* (2023).

[122] Xinmin Qiu, Congying Han, ZiCheng Zhang, Bonan Li, Tiande Guo, and Xuecheng Nie. 2023. DiffBFR: Bootstrapping Diffusion Model Towards Blind Face Restoration. In *ACMMM*.

[123] Chengchao Qu, Christian Herrmann, Eduardo Monari, Tobias Schuchert, and Jürgen Beyerer. 2017. Robust 3D patch-based face hallucination. In *WACV*. IEEE, 1105–1114.

[124] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-resolution image synthesis with latent diffusion models. In *CVPR*.

[125] Christos Sagonas, Georgios Tzimiropoulos, Stefanos Zafeiriou, and Maja Pantic. 2013. 300 faces in-the-wild challenge: The first facial landmark localization challenge. In *ICCVW*. 397–403.

[126] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. 2022. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 4 (2022), 4713–4726.

[127] Shuyao Shang, Zhengyang Shan, Guangxing Liu, LunQian Wang, XingHua Wang, Zekai Zhang, and Jinglin Zhang. 2024. Resdiff: Combining cnn and diffusion model for image super-resolution. In *AAAI*, Vol. 38. 8975–8983.

[128] Ziyi Shen, Wei-Sheng Lai, Tingfa Xu, Jan Kautz, and Ming-Hsuan Yang. 2018. Deep semantic face deblurring. In *CVPR*. 8260–8269.

[129] Ziyi Shen, Wei-Sheng Lai, Tingfa Xu, Jan Kautz, and Ming-Hsuan Yang. 2020. Exploiting semantics for face image deblurring. *International Journal of Computer Vision* 128 (2020), 1829–1846.

[130] Yibing Song, Jiawei Zhang, Lijun Gong, Shengfeng He, Linchao Bao, Jinshan Pan, Qingxiong Yang, and Ming-Hsuan Yang. 2019. Joint face hallucination and deblurring via structure generation and detail enhancement. *International journal of computer vision* 127 (2019), 785–800.

[131] Yibing Song, Jiawei Zhang, Shengfeng He, Linchao Bao, and Qingxiong Yang. 2017. Learning to hallucinate face images via component generation and enhancement. In *IJCAI*.

[132] Maitreya Suin, Nithin Gopalakrishnan Nair, Chun Pong Lau, Vishal M Patel, and Rama Chellappa. 2024. Diffuse and Restore: A Region-Adaptive Diffusion Model for Identity-Preserving Blind Face Restoration. In *WACV*. 6343–6352.

[133] Keda Tao, Jinjin Gu, Yulun Zhang, Xiucheng Wang, and Nan Cheng. 2024. Overcoming false illusions in real-world face restoration with multi-modal guided diffusion model. In *ICLR*.

[134] Yu-Ju Tsai, Yu-Lun Liu, Lu Qi, Kelvin CK Chan, and Ming-Hsuan Yang. 2024. Dual Associated Encoder for Face Restoration. *ICLR* (2024).

[135] Xiaoguang Tu, Jian Zhao, Qiankun Liu, Wenjie Ai, Guodong Guo, Zhifeng Li, Wei Liu, and Jiashi Feng. 2021. Joint face image restoration and frontalization for recognition. *IEEE Transactions on Circuits and Systems for Video Technology* 32, 3 (2021), 1285–1298.

[136] Kutub Uddin, Tae Hyun Jeong, and Byung Tae Oh. 2022. Incomplete region estimation and restoration of 3D point cloud human face datasets. *Sensors* 22, 3 (2022), 723.

[137] Tuomas Varanka, Tapani Toivonen, Soumya Tripathy, Guoying Zhao, and Erman Acar. 2024. PFStorer: Personalized Face Restoration and Super-Resolution. In *CVPR*. 2372–2381.

[138] Chenyang Wang, Junjun Jiang, Kui Jiang, and Xianming Liu. 2024. Low-Light Face Super-resolution via Illumination, Structure, and Texture Associated Representation. In *AAAI*, Vol. 38. 5318–5326.

[139] Chenyang Wang, Junjun Jiang, and Xianming Liu. 2021. Heatmap-aware pyramid face hallucination. In *ICME*. 1–6.

[140] Chenyang Wang, Junjun Jiang, Zhiwei Zhong, and Xianming Liu. 2022. Propagating facial prior knowledge for multitask learning in face super-resolution. *IEEE Transactions on Circuits and Systems for Video Technology* 32, 11 (2022), 7317–7331.

[141] Chenyang Wang, Junjun Jiang, Zhiwei Zhong, and Xianming Liu. 2023. Spatial-Frequency Mutual Learning for Face Super-Resolution. In *CVPR*. 22356–22366.

[142] Junke Wang, Shaoxiang Chen, Zuxuan Wu, and Yu-Gang Jiang. 2022. Ft-tdr: Frequency-guided transformer and top-down refinement network for blind face inpainting. *IEEE Transactions on Multimedia* (2022).

[143] Jingkai Wang, Jue Gong, Lin Zhang, Zheng Chen, Xing Liu, Hong Gu, Yutong Liu, Yulun Zhang, and Xiaokang Yang. 2025. Osdface: One-step diffusion model for face restoration. In *CVPR*. 12626–12636.

[144] Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin CK Chan, and Chen Change Loy. 2024. Exploiting diffusion prior for real-world image super-resolution. *International Journal of Computer Vision* (2024), 1–21.

[145] Li Wang, Xiangtao Meng, Dan Li, Xuhong Zhang, Shouling Ji, and Shanqing Guo. 2024. DEEPFAKER: a unified evaluation platform for facial deepfake and detection models. *ACM Transactions on Privacy and Security* 27, 1 (2024), 1–34.

[146] Tao Wang, Kaihao Zhang, Xuanxi Chen, Wenhan Luo, Jiankang Deng, Tong Lu, Xiaochun Cao, Wei Liu, Hongdong Li, and Stefanos Zafeiriou. 2022. A survey of deep face restoration: Denoise, super-resolution, deblur, artifact removal. *arXiv preprint arXiv:2211.02831* (2022).

[147] Xintao Wang, Yu Li, Honglun Zhang, and Ying Shan. 2021. Towards real-world blind face restoration with generative facial prior. In *CVPR*. 9168–9178.
[148] Yinhuai Wang, Yujie Hu, Jiwen Yu, and Jian Zhang. 2023. Gan prior based null-space learning for consistent super-resolution. In *AAAI*, Vol. 37. 2724–2732.
[149] Yinhuai Wang, Yujie Hu, and Jian Zhang. 2022. Panini-Net: GAN prior based degradation-aware feature interpolation for face restoration. In *AAAI*, Vol. 36. 2576–2584.
[150] Yuanzhi Wang, Tao Lu, Yanduo Zhang, Zhongyuan Wang, Junjun Jiang, and Zixiang Xiong. 2022. FaceFormer: aggregating global and local representation for face hallucination. *IEEE Transactions on Circuits and Systems for Video Technology* (2022).
[151] Yinhuai Wang, Jiwen Yu, and Jian Zhang. 2023. Zero-Shot Image Restoration Using Denoising Diffusion Null-Space Model. *ICLR* (2023).
[152] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* 13, 4 (2004), 600–612.
[153] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. 2003. Multiscale structural similarity for image quality assessment. In *ACSSC*, Vol. 2. Ieee, 1398–1402.
[154] Zhouxia Wang, Jiawei Zhang, Runjian Chen, Wenping Wang, and Ping Luo. 2022. Restoreformer: High-quality blind face restoration from undegraded key-value pairs. In *CVPR*. 17512–17521.
[155] Zhouxia Wang, Jiawei Zhang, Tianshui Chen, Wenping Wang, and Ping Luo. 2023. Restoreformer++: Towards real-world blind face restoration from undegraded key-value pairs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2023).
[156] Zhixin Wang, Ziying Zhang, Xiaoyun Zhang, Huangjie Zheng, Mingyuan Zhou, Ya Zhang, and Yanfeng Wang. 2023. DR2: Diffusion-based Robust Degradation Remover for Blind Face Restoration. In *CVPR*. 1704–1713.
[157] Lianxin Xie, Csbingbing Zheng, Wen Xue, Le Jiang, Cheng Liu, Si Wu, and Hau San Wong. 2024. Learning Degradation-unaware Representation with Prior-based Latent Transformations for Blind Face Restoration. In *CVPR*. 9120–9129.
[158] Jingwei Xin, Nannan Wang, Xinbo Gao, and Jie Li. 2019. Residual attribute attention network for face image super-resolution. In *AAAI*, Vol. 33. 9054–9061.
[159] Jingwei Xin, Nannan Wang, Xinrui Jiang, Jie Li, Xinbo Gao, and Zhifeng Li. 2020. Facial attribute capsules for noise face super resolution. In *AAAI*, Vol. 34. 12476–12483.
[160] Xiangyu Xu, Deqing Sun, Jinshan Pan, Yujin Zhang, Hanspeter Pfister, and Ming-Hsuan Yang. 2017. Learning to super-resolve blurry face and text images. In *ICCV*. 251–260.
[161] Lingbo Yang, Shanshe Wang, Siwei Ma, Wen Gao, Chang Liu, Pan Wang, and Peiran Ren. 2020. Hifacegan: Face renovation via collaborative suppression and replenishment. In *ACMMM*. 1551–1560.
[162] Peiqing Yang, Shangchen Zhou, Qingyi Tao, and Chen Change Loy. 2023. PGDiff: Guiding Diffusion Models for Versatile Face Restoration via Partial Guidance. In *NeurIPS*.
[163] Shuo Yang, Ping Luo, Chen-Change Loy, and Xiaoou Tang. 2016. Wider face: A face detection benchmark. In *CVPR*. 5525–5533.
[164] Tao Yang, Peiran Ren, Xuansong Xie, and Lei Zhang. 2021. Gan prior embedded network for blind face restoration in the wild. In *CVPR*. 672–681.
[165] Wenqi Yang, Zhenfang Chen, Chaofeng Chen, Guanying Chen, and Kwan-Yee K Wong. 2023. Deep Face Video Inpainting via UV Mapping. *IEEE Transactions on Image Processing* 32 (2023), 1145–1157.
[166] Rajeev Yasarla, Federico Perazzi, and Vishal M Patel. 2020. Deblurring face images using uncertainty guided multi-stream semantic networks. *IEEE Transactions on Image Processing* 29 (2020), 6251–6263.
[167] Dong Yi, Zhen Lei, Shengcai Liao, and Stan Z Li. 2014. Learning face representation from scratch. *arXiv preprint arXiv:1411.7923* (2014).
[168] Yu Yin, Joseph Robinson, Yulun Zhang, and Yun Fu. 2020. Joint super-resolution and alignment of tiny faces. In *AAAI*, Vol. 34. 12693–12700.
[169] Xin Yu, Basura Fernando, Bernard Ghanem, Fatih Porikli, and Richard Hartley. 2018. Face super-resolution guided by facial component heatmaps. In *ECCV*. 217–233.
[170] Xin Yu, Basura Fernando, Richard Hartley, and Fatih Porikli. 2018. Super-resolving very low-resolution face images with supplementary attributes. In *CVPR*. 908–917.
[171] Xin Yu and Fatih Porikli. 2016. Ultra-resolving face images by discriminative generative networks. In *ECCV*. Springer, 318–333.
[172] Xin Yu and Fatih Porikli. 2017. Face hallucination with tiny unaligned images by transformative discriminative neural networks. In *AAAI*, Vol. 31.
[173] Xin Yu and Fatih Porikli. 2017. Hallucinating very low-resolution unaligned and noisy face images by transformative discriminative autoencoders. In *CVPR*. 3760–3768.

[174] Xin Yu, Fatih Porikli, Basura Fernando, and Richard Hartley. 2020. Hallucinating unaligned face images by multiscale transformative discriminative networks. *International Journal of Computer Vision* 128, 2 (2020), 500–526.

[175] Xin Yu, Fatemeh Shiri, Bernard Ghanem, and Fatih Porikli. 2019. Can we see more? joint frontalization and hallucination of unaligned tiny faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42, 9 (2019), 2148–2164.

[176] Xiaoyuan Yu, Langwen Zhang, and Wei Xie. 2021. Semantic-driven face hallucination based on residual network. *IEEE Transactions on Biometrics, Behavior, and Identity Science* 3, 2 (2021), 214–228.

[177] Yanjiang Yu, Puyang Zhang, Kaihao Zhang, Wenhan Luo, and Changsheng Li. 2023. Multiprior learning via neural architecture search for blind face restoration. *IEEE Transactions on Neural Networks and Learning Systems* (2023).

[178] Zongsheng Yue and Chen Change Loy. 2024. Difface: Blind face restoration with diffused error contraction. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2024).

[179] Zongsheng Yue, Jianyi Wang, and Chen Change Loy. 2024. Resshift: Efficient diffusion model for image super-resolution by residual shifting. In *NeurIPS*, Vol. 36.

[180] Stefanos Zafeiriou, George Trigeorgis, Grigorios Chrysos, Jiankang Deng, and Jie Shen. 2017. The menpo facial landmark localisation challenge: A step towards the solution. In *CVPRW*. 170–179.

[181] Chengbin Zeng, Yi Liu, and Chunli Song. 2022. Swin-CasUNet: Cascaded U-Net with Swin Transformer for Masked Face Restoration. In *ICPR*. IEEE, 386–392.

[182] Kaihao Zhang, Dongxu Li, Wenhan Luo, Jingyu Liu, Jiankang Deng, Wei Liu, and Stefanos Zafeiriou. 2022. EDFace-Celeb-1 M: Benchmarking Face Hallucination with a Million-scale Dataset. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2022).

[183] Kaipeng Zhang, Zhanpeng Zhang, Chia-Wen Cheng, Winston H Hsu, Yu Qiao, Wei Liu, and Tong Zhang. 2018. Super-identity convolutional neural network for face hallucination. In *ECCV*. 183–198.

[184] Menglei Zhang and Qiang Ling. 2020. Supervised pixel-wise GAN for face super-resolution. *IEEE Transactions on Multimedia* 23 (2020), 1938–1950.

[185] Puyang Zhang, Kaihao Zhang, Wenhan Luo, Changsheng Li, and Guoren Wang. 2024. Blind face restoration: Benchmark datasets and a baseline model. *Neurocomputing* 574 (2024), 127271.

[186] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*. 586–595.

[187] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. 2018. Image super-resolution using very deep residual channel attention networks. In *ECCV*. 286–301.

[188] Yongle Zhang, Yimin Liu, Hao Fan, Ruotong Hu, Jian Zhang, and Qiang Wu. 2025. Consistent Image Inpainting with Pre-Perception and Cross-Perception Collaborative Processes. *IEEE Transactions on Image Processing* (2025).

[189] Yi Zhang, Xiaoyu Shi, Dasong Li, Xiaogang Wang, Jian Wang, and Hongsheng Li. 2024. A unified conditional framework for diffusion-based image restoration. In *NeurIPS*, Vol. 36.

[190] Yang Zhang, Ivor W Tsang, Jun Li, Ping Liu, Xiaobo Lu, and Xin Yu. 2021. Face hallucination with finishing touches. *IEEE Transactions on Image Processing* 30 (2021), 1728–1743.

[191] Yang Zhang, Ivor W Tsang, Yawei Luo, Changhui Hu, Xiaobo Lu, and Xin Yu. 2021. Recursive copy and paste gan: Face hallucination from shaded thumbnails. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 8 (2021), 4321–4338.

[192] Yang Zhang, Ivor W Tsang, Yawei Luo, Chang-Hui Hu, Xiaobo Lu, and Xin Yu. 2020. Copy and paste GAN: Face hallucination from shaded thumbnails. In *CVPR*. 7355–7364.

[193] Yang Zhang, Xin Yu, Xiaobo Lu, and Ping Liu. 2022. Pro-uigan: Progressive face hallucination from occluded thumbnails. *IEEE Transactions on Image Processing* 31 (2022), 3236–3250.

[194] Yang Zhang, Xian Zhang, Canghong Shi, Xi Wu, Xiaojie Li, Jing Peng, Kunlin Cao, Jiancheng Lv, and Jiliu Zhou. 2022. Pluralistic Face Inpainting with Transformation of Attribute Information. *IEEE Transactions on Multimedia* (2022).

[195] Zhenyu Zhang, Yanhao Ge, Ying Tai, Xiaoming Huang, Chengjie Wang, Hao Tang, Dongjin Huang, and Zhifeng Xie. 2022. Learning to Restore 3D Face from In-the-Wild Degraded Images. In *CVPR*. 4237–4247.

[196] Tianyu Zhao and Changqing Zhang. 2020. Saan: Semantic attention adaptation network for face super-resolution. In *ICME*. 1–6.

[197] Yang Zhao, Tingbo Hou, Yu-Chuan Su, Xuhui Jia Li, Matthias Grundmann, et al. 2023. Towards Authentic Face Restoration with Iterative Diffusion Models and Beyond. In *ICCV*.

[198] Yang Zhao, Yu-Chuan Su, Chun-Te Chu, Yandong Li, Marius Renn, Yukun Zhu, Changyou Chen, and Xuhui Jia. 2022. Rethinking deep face restoration. In *CVPR*. 7652–7661.

[199] Yicheng Zhong, Yuru Pei, Peixin Li, Yuke Guo, Gengyu Ma, Meng Liu, Wei Bai, WenHai Wu, and Hongbin Zha. 2020. Face denoising and 3D reconstruction from a single depth image. In *FG*. IEEE, 117–124.

[200] Erjin Zhou, Haoqiang Fan, Zhimin Cao, Yuning Jiang, and Qi Yin. 2015. Learning face hallucination in the wild. In *AAAI*, Vol. 29.

[201] Shangchen Zhou, Kelvin Chan, Chongyi Li, and Chen Change Loy. 2022. Towards robust blind face restoration with codebook lookup transformer. In *NeurIPS*, Vol. 35. 30599–30611.

[202] Feida Zhu, Junwei Zhu, Wenqing Chu, Xinyi Zhang, Xiaozhong Ji, Chengjie Wang, and Ying Tai. 2022. Blind face restoration via integrating face shape and generative priors. In *CVPR*. 7662–7671.

[203] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*. 2223–2232.

[204] Shizhan Zhu, Sifei Liu, Chen Change Loy, and Xiaoou Tang. 2016. Deep cascaded bi-network for face hallucination. In *ECCV*. Springer, 614–630.