Restoration by Generation with Constrained Priors

Zheng Ding[1†], Xuaner Zhang[2], Zhuowen Tu[1], Zhihao Xia[2]
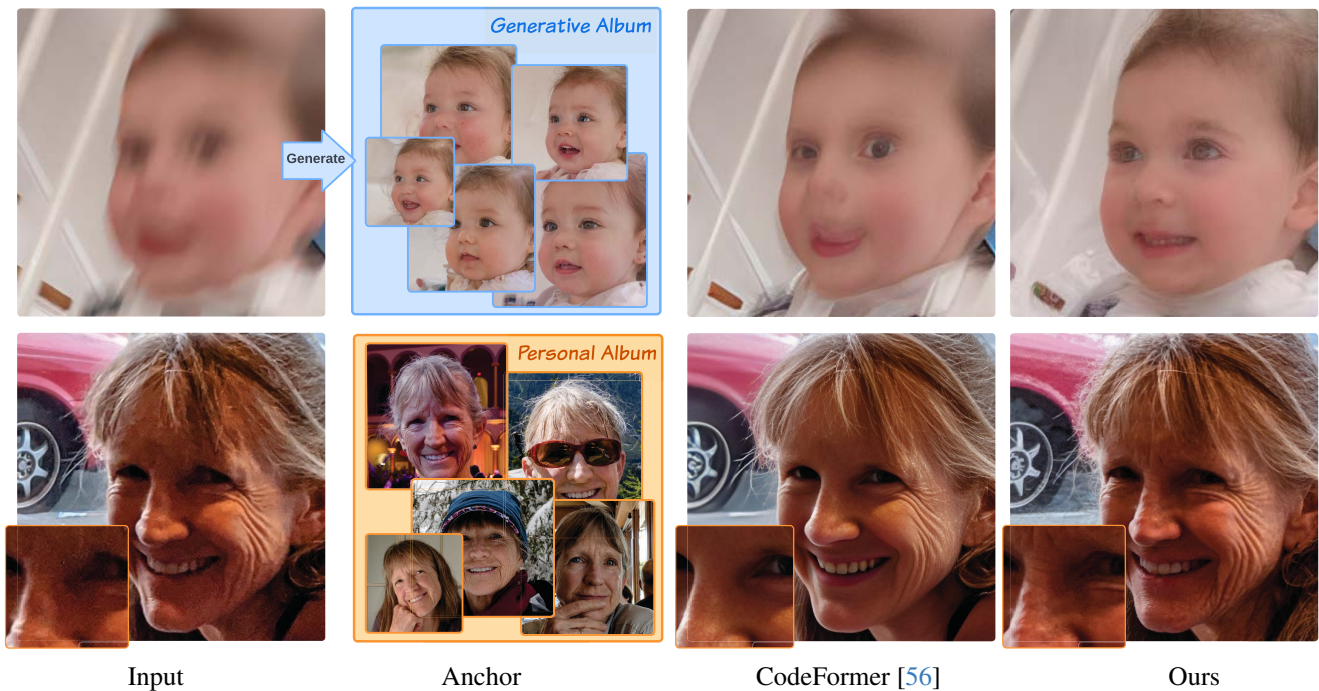
[1]UC San Diego   [2]Adobe

Figure 1. We harness the generative capacity of a diffusion model for image restoration. By constraining the generative space with a generative or personal album, we can directly use a pre-trained diffusion model to produce a high-quality and realistic image that is also faithful to the input identity. Without any assumption on the degradation type, we are able to generalize to real-world images that exhibit complicated degradation. We compare our restoration result with CodeFormer, a state-of-the-art baseline [56]. Our method generalizes better to different types of degradation while more faithfully preserving the input identity. Images are best viewed zoomed in on a big screen.

Abstract

The inherent generative power of denoising diffusion models makes them well-suited for image restoration tasks where the objective is to find the optimal high-quality image within the generative space that closely resembles the input image. We propose a method to adapt a pretrained diffusion model for image restoration by simply adding noise to the input image to be restored and then denoise. Our method is based on the observation that the space of a generative model needs to be constrained. We impose this constraint by finetuning the generative model with a set of anchor images that capture the characteristics of the input image. With the constrained space, we can then leverage the sampling strategy used for generation to do image restoration. We evaluate

against previous methods and show superior performances on multiple real-world restoration datasets in preserving identity and image quality. We also demonstrate an important and practical application on personalized restoration, where we use a personal album as the anchor images to constrain the generative space. This approach allows us to produce results that accurately preserve high-frequency details, which previous works are unable to do. Project webpage: https://gen2res.github.io.

1. Introduction

Image restoration involves recovering a high-quality natural image x from its degraded observation $y = H(x)$ is a fundamental task in low-level vision. The challenge lies in finding a solution that 1) matches the observation through a set of degradation steps; and 2) aligns with the distribution of

---

† Work done during an internship at Adobe.

x. In scenarios where the degradation process H is unknown, the problem becomes a blind image restoration problem.

Discriminative learning approaches [11, 40, 48, 56] aim to solve this inverse problem directly by training an inverse model $F(y)$, typically a neural network, using datasets of low- and high-quality image pairs (x, y). However, the trained model is limited to restoring images with degradations H present in the training set. This limitation places the burden of generalization on the construction of the training set. The effectiveness of these methods also heavily depends on the capacity of the inversion model and the characteristics of the loss function. Model-based optimization methods [5, 20, 31, 32, 53], on the other hand, assume that the degradation model is only known at inference time. They focus on learning the image prior $p(x)$, which can be represented as regularization terms [32], denoising networks [31, 51], or more recently pre-trained diffusion models [5, 20]. However, these methods generally assume that the degradation process is known at inference time, limiting their practicality and often relegating them to synthetic evaluations.

In this paper, we adopt a markedly different approach to the image restoration problem. We observe that humans are able to recognize a degraded image (i.e., a 'bad photo') and envision a fix without knowing the imperfections in the image formation process. Such insights rely on our inherent understanding of what constitutes a high-quality image. Building on this observation, we propose to approach image restoration using the recent success of large generative models, which possess the capacity of forming high-quality imagery. Unlike prior works, we do not make any assumption on the degradation process. Our method solely relies on a well-trained denoising diffusion model.

The challenges then arise in how to project the input image into the generative process given the models are trained on mostly clean images. And once projected, how to constrain the generation to preserve the useful features in the input, e.g., the identity. We address the input projection by adding Gaussian noise to the low-quality image to be restored, matching the distribution of clean images added with noise. Once projected, we can then denoise the image as is normally done in the generation process of a diffusion model. To handle the second challenge of preserving useful signals in the input, we propose to constrain the generative space by finetuning the model with anchor images that share characteristic features with the low-quality input. When the anchor is given, such as from an album of other photos of the same identity, we can simply finetune the model with the provided images. When the anchor is missing, as in most single-image restoration scenarios, we propose to use a generative album as the anchor. The generative album is a set of clean images generated from the diffusion model with the low-quality input image imposing soft guidance, and thus closely resembles the input image.
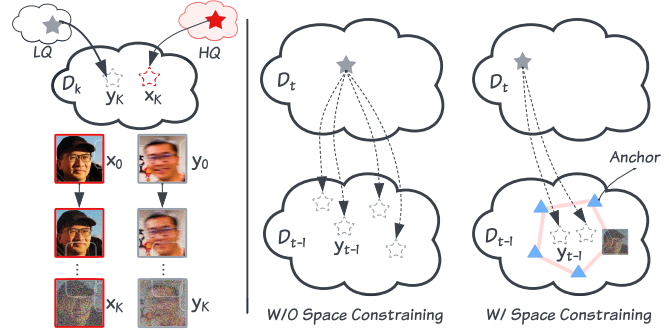

Figure 2. **Left: Image projection.** When sufficient Gaussian noise is added to the low- and high-quality image, we can bring them to the same distribution. The low-quality image can thus be denoised with a pre-trained diffusion model. **Right: With and without space constraining.** A regular diffusion step lands y_t in an arbitrary position in the generative space; with space constraining, the path of generation becomes more constrained towards the space defined by the anchor images.

Surprisingly, we find that our straightforward approach yields high-quality results on blind image restoration. Unlike previous methods, our approach does not rely on paired training data or assumptions about the degradation process. It thus generalizes well to real-world images with unknown degradation types, such as noise, motion blur, and low resolution. By effectively harnessing the generative capacity of a pre-trained diffusion model, our generation-based restoration approach produces high-quality and realistic images that are faithful to the input identity.

2. Related Works

Supervised Learning for Image Restoration. The trend of leveraging advanced neural network architectures for image restoration has spanned from CNNs [2, 39, 50, 52, 54] to GANs [23, 24, 26], and more recently, to transformers [29, 43, 49] and diffusion models [34, 35, 45]. One aspect remains unchanged: these methods are trained on datasets comprising pairs of high-quality and low-quality images. Typically, these image pairs are synthetically generated, depicting a single type of degradation, leading to task-specific models for denoising [12, 39, 46, 50, 52], deblurring [2, 23, 24, 45], or super-resolution [26, 35, 41]. However, they fall short when applied to real-world low-quality images, which often suffer from diverse, unknown degradations.

In specific domains, particularly with facial images, numerous works have focused on training blind restoration models that simulate various degradation types during training. For instance, GFPGAN [40] and GPEN [48] enhance pretrained GAN networks with modules to leverage generative priors for blind face restoration. Recent approaches like CodeFormer [56], VQFR [11] and RestoreFormer[49] exploit the low-dimensional space of facial images to achieve impressive results. Emerging works have also started build-

ing upon the success of diffusion models [7, 14, 37]. For example, IDM [55] trains a conditional diffusion model for face image restoration by injecting low-quality images at different layers of the model. Conversely, DR2 [44] combines the generative capabilities of pre-trained diffusion models with existing face restoration networks. Another line of works [27, 28] seeks to enhance the results by incorporating additional information present in a guide image or photo album, which is often available in practice. Nevertheless, these methods rely on a synthetic data pipeline for training, which limits their generalizability. Diverging from these methodologies, our approach does not use paired data, synthetic or real, allowing it to generalize naturally to real data without succumbing to artifacts.

Model-based Image Restoration. Unlike supervised learning methods, model-based methods often form a posterior of the underlying clean image given the degraded image, with a likelihood term from the degradation process and an image prior. [31, 53] proposed using denoising networks as the image prior. These priors are integrated with the known degradation process during inference, and the Maximum A Posteriori (MAP) problem is addressed through approximate iterative optimization methods. DGP [30] proposes image restoration through GAN inversion, searching for a latent code that generates an image closely matching the input image after processing it through the known degradation. The recent success of pre-trained foundational diffusion models has inspired works [3, 16, 18, 19] to utilize diffusion models as such priors. Kawar *et al*. [20] and Wang *et al*. [42] proposed an unsupervised posterior sampling method using a pre-trained denoising diffusion model to solve linear inverse problems. Chung *et al*. [5] extends diffusion solvers to general noise inverse problems. Despite these advancements, these methods generally assume that the degradation process is known at inference, limiting their practicality to synthetic evaluations. In contrast, our method does not assume any knowledge of the degradation model at training or inference.

Personalized Diffusion Models. Personalization methods aim to adapt pre-trained diffusion models to specific subjects or concepts by leveraging data unique to the target case. In text-to-image synthesis, many works opt for customization by fine-tuning with personalized data, adapting token embeddings of visual concepts [9, 10], the entire denoising network [33], or a subset of the network [22]. Recent studies [15, 36, 47] propose bypassing per-object optimization by training an encoder to extract embeddings of subject identity and injecting them into the diffusion model's sampling process. In other domains, DiffusionRig [8] learns personalized facial editing by fine-tuning a 3D-aware diffusion model on a personal album. In this work, we demonstrate that a personalized diffusion model represents a constrained generative space, directly usable for sampling high-quality images

to restore images of a specific subject, without additional complexities. For *single-image restoration*, unlike previous instance-based personalization methods [15, 36, 47], we generate an album of images close to the input and then constrain the diffusion model using this generative album. This approach enables restoration by directly sampling from the fine-tuned model, eliminating the need for guidance.

3. Method

3.1. Preliminaries

A diffusion model approximates its training image distribution $p_\theta(x_0)$ by learning a model θ that effectively reverses the process of adding noise. The commonly used Denoising Diffusion Probabilistic Models (DDPM) gradually introduce Gaussian noise into a clean image x_0:

$$x_t = \sqrt{\alpha_t}x_0 + \sqrt{1-\alpha_t}\epsilon, \quad \text{where} \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \quad (1)$$

The reverse generative process aims to progressively denoise x_t until it is free from noise. Once a diffusion model is trained, for any given time t and the corresponding noisy image x_t, it can iteratively denoise by sampling from $p(x_0|x_t)$ using the trained model.

The objective of image restoration, on the other hand, is to recover the latent high-quality image x_0 from a low-quality, partially observed image y_0. Contrary to previous methods that decompose the posterior distribution into the likelihood $p(y_0|x_0)$ and the prior $p(x_0)$ to solve a MAP problem, we propose to recover the complete observation by directly sampling from the posterior:

$$\hat{x} \sim p(x_0|y_0) \quad (2)$$

3.2. Restoration by Generation

We aim to maximally leverage the generative capacity of the diffusion model by using its iterative sampling process for restoration. A critical observation underlies this approach: when sufficient Gaussian noise is added to the degraded observation y_0, the resultant image y_t:

$$y_t = \sqrt{\alpha_t}y_0 + \sqrt{1-\alpha_t}\epsilon, \quad \text{where} \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \quad (3)$$

becomes indistinguishable from the underlying clean image x_0 with the same noise. That is, there exists a large enough K such that

$$y_K \approx x_K \quad (4)$$

This phenomenon becomes apparent from Eq 1 and 3 as α decreases and when the same noise ϵ is sampled. It is also demonstrated in Fig 2, where adding noise to high-quality and low-quality images can progressively align their distributions, making them more similar over time, this suggests:

$$p(x_0|y_K) \approx p(x_0|x_K) \quad (5)$$

Based on this observation, we can sample a clean image x_0 from $p(x_0|y_K)$ using the same sampling process as from
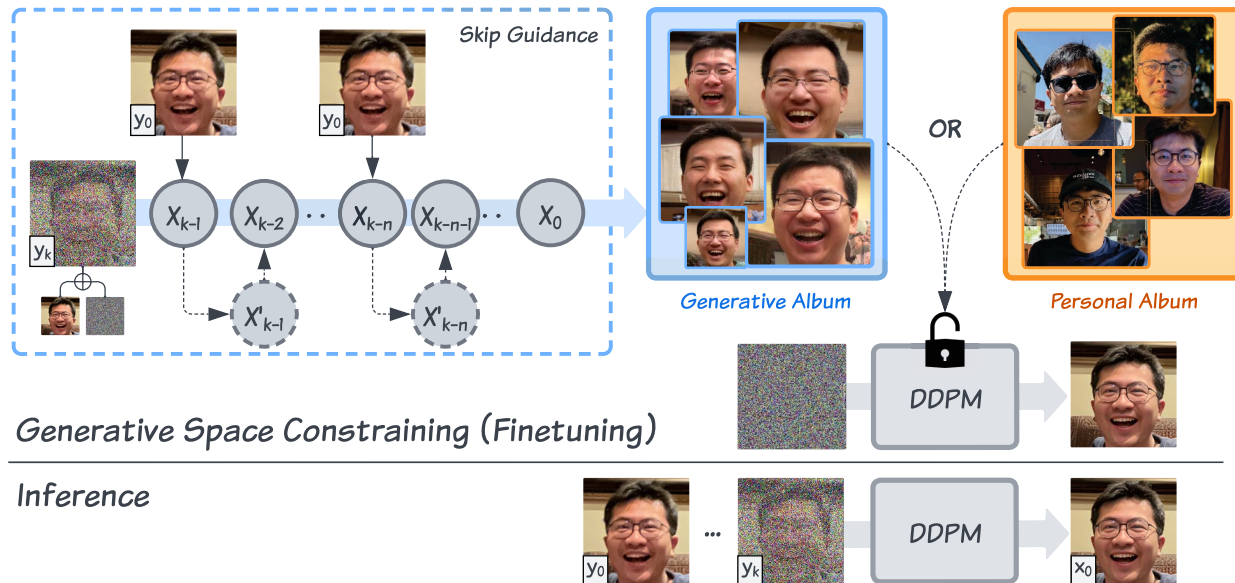
Figure 3. **An illustration of our finetuning and inference stage.** The core of our method is to constrain the generative space by fine-tuning a pre-trained diffusion model with either a generative album or a personal album. The generative album is generated from the input low-quality image with skip guidance to loosely follow the characteristics of the input. Once the generative space is constrained, at inference time, we can simply add noise to the input low-quality image and pass it through the diffusion model to do restoration.

$p(x_0|x_K)$; in other words, we can denoise y_K iteratively directly with the pre-trained diffusion model. Since the sampling process remains unchanged, the resultant image should match the quality of the images generated from the original diffusion model.

We find it critical to select the optimal time K, which determines the amount of noise added to the low-quality input image to start the sampling process. If too little noise is added, the discrepancy between x_K and y_K becomes large, yielding low-quality samples as y_K does not align with the training distribution $p(x_K)$ of the diffusion model. On the other hand, with too excessive noise added, the original contents in the input y_K are hardly discernible. The generated sample, though with high quality, will not be faithful to the input. We aim to produce high-quality samples, while mitigating the information loss, and achieve so by constraining the generative space of the pre-trained diffusion model.

3.3. Generative Space Constraining

The loss of information is inherent in the diffusion process. Due to the stochasticity of the forward Markov chain, the clean image generated using the reverse process from x_t may not match the original x_0. The larger t is, the larger the generative space $p(x_0|x_t)$ spans. The learned score functions guide x_t to the clean image space without constraining its content. This property is desirable for a generative model where the diversity of generation is valued. However, this is not ideal for image restoration where the input contents also need to be preserved. The goal is thus to constrain the generative space to a small subspace that tightly surrounds the underlying clean image.

We propose to use a set of anchor images to fine-tune the diffusion model, thus imposing the generative space. These anchor images can be given in the form of a *personal album*, or be generated as a *generative album* in the common scenario of single image restoration.

Personal Album as Additional Information. In many real-world scenarios, additional information about the underlying clean image beyond a single degraded observation is available, such as an album of different clean images of the same subject. We personalize the pre-trained diffusion model in this case — fine-tuning it with the personal album. This approach naturally addresses the ill-posed nature of single-image restoration, producing results containing authentic high-frequency details absent in the degraded observation. This is demonstrated in identity preservation in face restoration tasks (Sec 4.2).

Generative Album from a Single Degraded Observation. For single-image restoration, due to its ill-posed nature, we can only constrain the generative space to a subspace of high-quality realistic images close to the degraded observation. To generate this album of high-quality images, we follow approaches similar to previous works on guided image generation [1, 5, 38]. Specifically, given a degraded image y_0, we first add noise ϵ_K to obtain y_K, then denoise it progressively with the pre-trained diffusion model. For the denoised image x_t, we apply a simple L_1 guidance that computes the distance between the input degraded image and the generated

	Wider-Test		WebPhoto-Test		LFW-Test		Deblur-Test	
	FID ↓	MUSIQ ↑	FID ↓	MUSIQ ↑	FID ↓	MUSIQ ↑	FID ↓	MUSIQ ↑
Input	183.03	15.68	161.82	20.26	131.68	27.51	169.43	27.53
GFPGAN[40]	59.38	56.48	114.15	55.13	64.10	60.46	178.40	58.03
CodeFormer[56]	48.57	55.70	98.55	55.20	66.31	58.72	163.47	57.09
VQFR[11]	52.64	54.23	105.94	52.44	63.73	57.52	168.36	54.45
DR2(+VQFR)[44]	69.40	53.62	143.96	51.92	67.70	57.42	173.33	55.34
Ours	**46.38**	**58.73**	**96.44**	**57.71**	**56.32**	**60.68**	**135.33**	**60.20**

Table 1. **Quantitative comparison on real-world single-image blind face restoration on four datasets.**

image:

$$x'_t = x_t - \lambda \nabla_{x_t} ||y_0 - \hat{x}_{0,t}||_2^2 \qquad (6)$$

Unlike previous methods where the guidance needs to be strongly followed, our guidance, the low-quality input, is an approximation. Instead of applying the guidance at every step [5, 38], we propose to apply this approximated guidance periodically at every n steps. The proposed *Skip Guidance* enforces the generated image to loosely follow the information in the degraded input while retaining the quality of images in the generative steps. We repeat this process multiple times to generate a set of images that form a generative album, which is used to fine-tune the diffusion model.

Once the diffusion model is fine-tuned with a personal or generative album, we restore a degraded image y_0 by adding noise ϵ_K. Then, we iteratively denoise y_K using the fine-tuned model for K steps, *without* further guidance. Notably, our approach does not rely on paired data for training and makes no assumptions about the degradation process at training or inference.

4. Experiments

With the core observation that generation can be directly applied for restoration, our method requires only a pre-trained unconditional diffusion model and is applicable to any image domain for which the diffusion model has been trained. We first show results of our restoration-by-generation approach on the standard task of single-image blind face restoration in Sec 4.1. In Sec 4.2, we extend our approach to personalized face restoration. Here, the objective is to restore a degraded image of a subject using other clean images of the same identity. Sec 4.3 presents the adaptation of our method to different image categories, such as dogs and cats, by simply swapping the pre-trained diffusion model. Notably, as our method does not presume any specific form of degradation, all our evaluations are conducted on real images with unknown degradation.

4.1. Blind Face Restoration with Generative Album

For the task of single-image blind face restoration, we utilize an unconditional diffusion model pretrained on the FFHQ dataset [17]. We first assess our approach on three widely-used real-world face benchmarks with degradation

levels ranging from heavy to mild: Wider-Test (970 images) [56], LFW-Test (1771 images) [40], and Webphoto-Test (407 images) [40]. These datasets are collections of in-the-wild images aligned using the method employed in FFHQ [17].

Our approach uses a *generative album* as the anchor for restoring these in-the-wild images. For each input low-quality image, we generate 16 images with skip guidance to form the album. We then fine-tune the diffusion model using this album to constrain the generative space. The process involves adding noise to the input low-quality image and denoising it for K steps with the fine-tuned model, where $K = 200$. The model is fine-tuned for 3,000 iterations with a batch size of 4 and a learning rate of 1e-5.

We benchmark our method against state-of-the-art supervised alternatives for blind face restoration, including the GAN-based GFPGAN [40], two codebook-based approaches (Codeformer [56] and VQFR [11]), and a diffusion-based approach DR2 [44]. Except for DR2, which combines a diffusion model with the pretrained supervised face restoration model VQFR [11], all methods utilize supervised training with synthetic low-quality images from FFHQ.

Quantitative and qualitative results are provided. For the former, we use FID [13] and MUSIQ(Koniq) [21] as metrics following CodeFormer [56]. The quantitative scores are in Table 1. Previous methods, except for DR2 [44], are trained on FFHQ-512×512 for restoration. For a fair comparison, we downsize the outputs of these methods to 256×256 for metric calculation. Our results surpass all previous methods in terms of FID and MUSIQ across all datasets, despite not undergoing a supervised training approach for image restoration. Qualitative comparisons in Figure 4 illustrate that our method produces high-quality restoration results akin to those from an unconditional diffusion model, even with severely degraded input images.

Our method's agnosticism to the degradation process leads to superior generalization capabilities. To further demonstrate this, we constructed a motion blur dataset (Deblur-Test) by selecting 67 images from [25] featuring moderate to severe real motion blur. The synthetic data pipeline in other supervised approaches does not model motion blur, resulting in poor performance on this out-of-distribution dataset. In contrast, our method consistently restores clean images from complex non-uniform motion

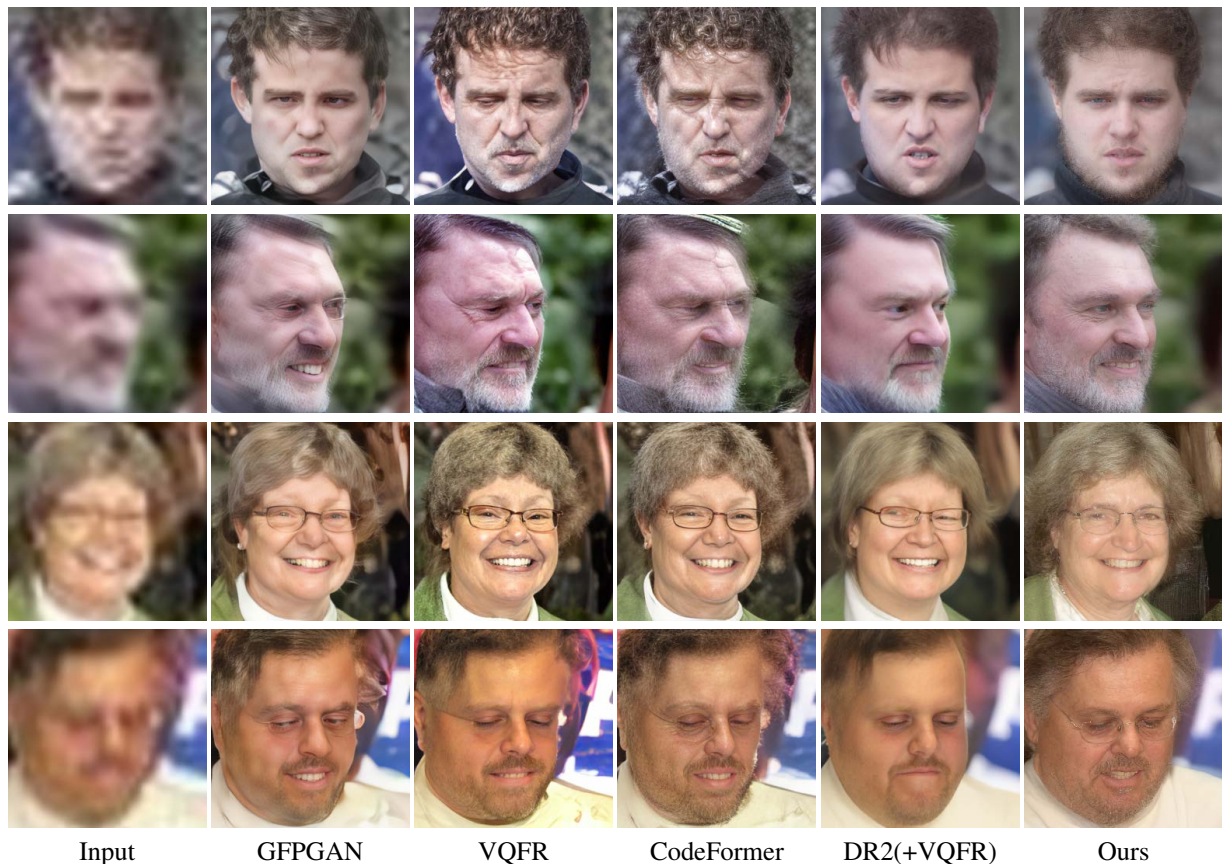| Input | GFPGAN | VQFR | CodeFormer | DR2(+VQFR) | Ours |

Figure 4. **Qualitative comparison with baselines on Wider-Test.** With strong generative capacity of the diffusion model, our method performs well on severely degraded images. We are able to produce high-quality and realistic images while prior works suffer from unrealistic artifacts.

blur, as seen in Figure 5, outperforming previous methods significantly, as shown in Table 1.
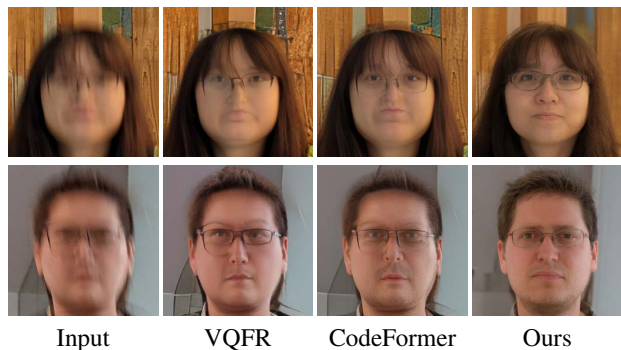

| Input | VQFR | CodeFormer | Ours |

Figure 5. **Comparison with previous methods on Deblur-Test.** Previous methods do not include motion blur as part of the degradation simulation for training, and thus fail to restore the images. In contrast, our method does not make assumptions on the degradation types and generalizes more robustly.

4.2. Personalized Face Restoration

We now evaluate our method on personalized restoration. Given a set of clean images of a subject, the goal is to restore any degraded image of the same subject using personalized features to preserve identity and recover high-frequency details that may have been lost in the degraded image. Our method naturally incorporates the personal album as the anchor. We use a personal album that contains around 20 images with diversity in pose, hairstyle, accessories, lighting, etc. We fine-tune on the personal album for 5,000 iterations. The model can then be used to restore any low-quality images of the same subject through direct sampling.

We compare our method against three single-image-based works: Codeformer [56], VQFR [11], DR2 [44], as well as an exemplar-based approach ASFFNet [28] which also incorporates a personal album for additional information. We evaluate our approach on three subjects: an elderly woman (Subject A), Obama and Hermione. We present the qualitative comparison in Figure 6. Single-image-based methods struggle to preserve identity – for example, wrinkles and other facial structures are often missing in the results of CodeFormer or DR2 for the elderly subject, altering their age and identity. By using a photo album as reference, ASFFNet preserves identity better, but fails to produce high-quality results. Our method, on the other hand, directly samples
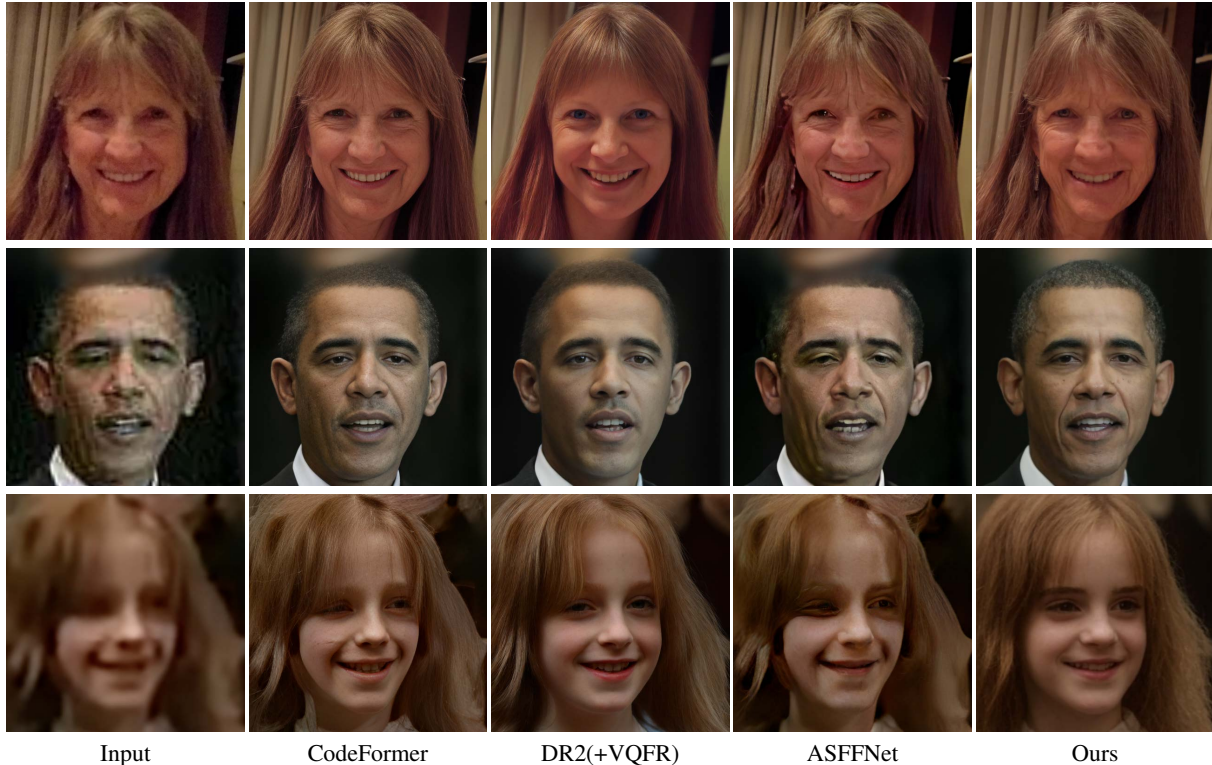
Figure 6. **Qualitative Comparison on personalized face restoration.** From top to bottom: Subject A, Obama and Hermione. With a personal album as anchor, we are able to restore images with faithful preservation of the input identity. Previous single-image methods alter the identity with lost details; previous reference-based methods fail to produce high-quality images and are prone to artifacts.

from the personalized generative space to do restoration, and thus produces faithful and high-quality results.

We also provide quantitative evaluation in Table 2 where we focus on the identity preservation. We use the identity score which uses the cosine similarity of the features given by a face recognition network ArcFace[6]. For each subjects, we collect around 20 test images and compute their average identity scores. Table 2 shows that our method preserves the identity of the subject much better than both single-image-based methods and the exemplar-based approach ASFFNet.

	Subject A	Obama	Hermione
Input	0.721	0.502	0.483
CodeFormer[56]	0.633	0.558	0.518
VQFR[11]	0.560	0.527	0.483
DR2(+VQFR)[44]	0.384	0.400	0.392
ASFFNet[28]	0.694	0.574	0.522
Ours	**0.731**	**0.716**	**0.664**

Table 2. **IDS comparison on three subjects.** We use the cosine similarity of the features given by ArcFace[6] to compute identity score.

4.3. Beyond Face Restoration

Our model does not make any assumptions about the type of degradation or image contents, allowing it to be easily extended to other categories of data where a generative

model is available. Specifically, we evaluate our approach's ability to generalize to restoring dog and cat images. We pre-train two diffusion models with the same architecture, one for dogs and one for cats, on the AFHQ Dog and Cat datasets [4]. Our testing involves three subjects: a gray cat, an English golden retriever, and an Australian shepherd. For each subject, we fine-tune the pre-trained diffusion model using an album of around 20 images. Once fine-tuned, given a low-quality image, we add noise to it and then denoise it using the fine-tuned model. Qualitative results in Figure 7 demonstrate that our method can effectively reconstruct high-frequency details such as fur, while preserving the identity.

5. Ablation Studies

Noise Step K. Our restoration-by-generation approach is predicated on the observation that sufficient noise added to a degraded image y_0 and subsequent denoising of the noisy image y_K with a pre-trained diffusion model yields a high-quality, realistic image. Here, we demonstrate this observation and analyze the effect of the choice of K, which determines the noise level added to initiate the sampling process. Figure 8 displays sampled images from y_K for varying K values. A smaller K leads to a y_K that falls outside the typical diffusion process's training trajectory, resulting in lower-quality sampled output. Conversely, while
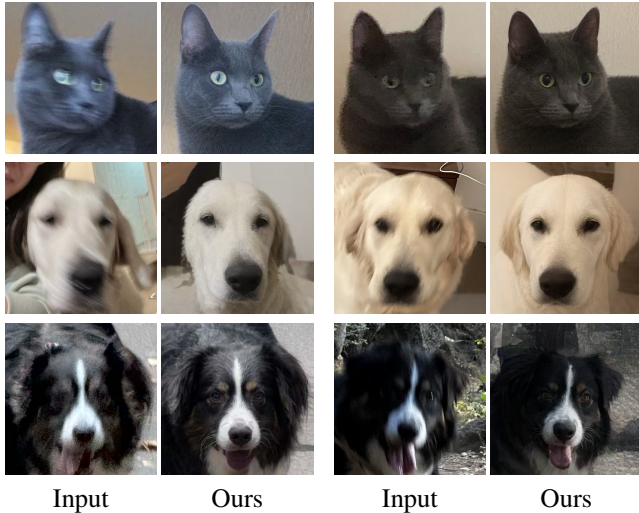
Figure 7. **Results on real-world cat/dog restoration.** Our method easily extends to other categories with corresponding pre-trained diffusion models. We show results on cats and dogs where we can reconstruct high-frequency details while preserving the identity.

a larger K enhances sample quality as hypothesized, it may also produce outputs less faithful to the input.

Constraining Prior with Generative Album. In the same Figure 8, we illustrate the significance of prior constraining and the effectiveness of using a generative album. As shown, a generative space that is too diverse increases the difficulty of sampling high-quality images from a given input, especially when K is small. Conversely, for large K values, the sampled image can deviate significantly from the input. Constraining the generative space with an album close to the input ensures preservation of input information in the output for large K, while still allowing high-quality sampling from small K. Ablation on Skip Guidance is included in the supplementary.

Constraining Prior with Personal Album. When a personal album is available, we directly constrain the generative space with this album. This not only improves output quality and faithfulness, as with the generative album, but also aids in recovering information absent in the input. As demonstrated in Figure 9, compared to an unconstrained model (i.e., the pre-trained diffusion model), the personalized model produces higher-quality images that better preserve identity.

6. Conclusion

We propose a method for image restoration that involves simply adding noise to a degraded input and then denoising it with a diffusion model. The key to our approach is constraining the generative space with a set of anchor images. We demonstrate in single-image restoration tasks that this method yields high-quality restoration results, surpassing previous supervised approaches. Furthermore, we show
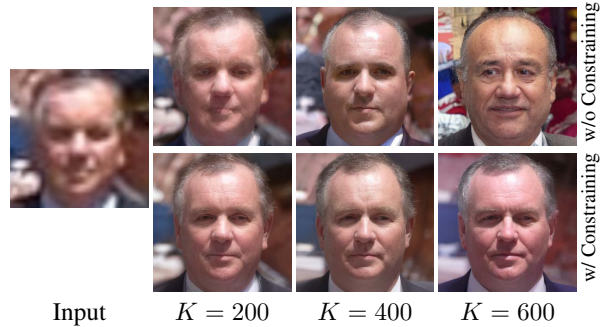

Figure 8. **Ablation on Noise Step K and Constraining with Generative Album.** As K increases, quality of images sampled from y_K improves, but alignment with the input reduces. Fine-tuning with a generative album notably enhances both image quality and input fidelity.
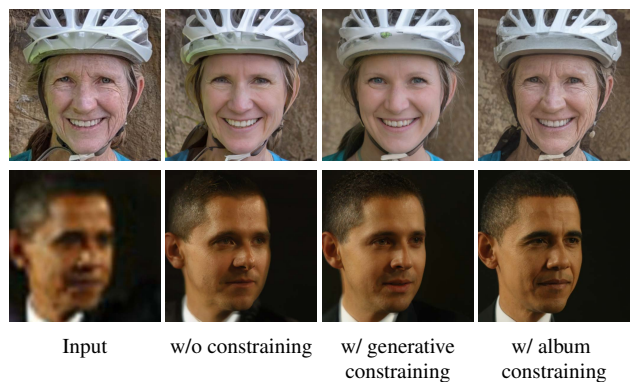

Figure 9. **Constraining with personal album.** Personalized model produces higher-quality images that better preserve identity compared to the model without constraining.

that constraining the generative space with a personal album leads to a personalized restoration-by-generation model that is effective for any image of the same subject, producing results with high quality and faithful details.

Limitations and Future Work. Unlike the personalization case, for single-image restoration, our approach requires fine-tuning for each input image. This is relatively slow compared to feed-forward approaches. Investigating methods to constrain the generative space without fine-tuning could be interesting. Furthermore, we have primarily validated our approach on class-specific image restoration tasks, largely due to the absence of a high-quality pre-trained diffusion model for natural images. Exploring whether our approach remains effective within a more diverse generative space would be intriguing. Such exploration could potentially address the challenge of blind restoration for general images.

References

[1] Arpit Bansal, Hong-Min Chu, Avi Schwarzschild, Soumyadip Sengupta, Micah Goldblum, Jonas Geiping, and Tom Goldstein. Universal guidance for diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 843–852, 2023. 4

[2] Sung-Jin Cho, Seo-Won Ji, Jun-Pyo Hong, Seung-Won Jung, and Sung-Jea Ko. Rethinking coarse-to-fine approach in single image deblurring. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4641–4650, 2021. 2

[3] Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon. Ilvr: Conditioning method for denoising diffusion probabilistic models. *arXiv preprint arXiv:2108.02938*, 2021. 3

[4] Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. Stargan v2: Diverse image synthesis for multiple domains. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8188–8197, 2020. 7

[5] Hyungjin Chung, Jeongsol Kim, Michael T Mccann, Marc L Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. *arXiv preprint arXiv:2209.14687*, 2022. 2, 3, 4, 5

[6] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4690–4699, 2019. 7

[7] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021. 3

[8] Zheng Ding, Xuaner Zhang, Zhihao Xia, Lars Jebe, Zhuowen Tu, and Xiuming Zhang. Diffusionrig: Learning personalized priors for facial appearance editing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12736–12746, 2023. 3

[9] Rinon Gal, Yuval Alaluf, Yuval Atzmon, Or Patashnik, Amit H Bermano, Gal Chechik, and Daniel Cohen-Or. An image is worth one word: Personalizing text-to-image generation using textual inversion. *arXiv preprint arXiv:2208.01618*, 2022. 3

[10] Rinon Gal, Moab Arar, Yuval Atzmon, Amit H Bermano, Gal Chechik, and Daniel Cohen-Or. Encoder-based domain tuning for fast personalization of text-to-image models. *ACM Transactions on Graphics (TOG)*, 42(4):1–13, 2023. 3

[11] Yuchao Gu, Xintao Wang, Liangbin Xie, Chao Dong, Gen Li, Ying Shan, and Ming-Ming Cheng. Vqfr: Blind face restoration with vector-quantized dictionary and parallel decoder. In *European Conference on Computer Vision*, pages 126–143. Springer, 2022. 2, 5, 6, 7

[12] Shi Guo, Zifei Yan, Kai Zhang, Wangmeng Zuo, and Lei Zhang. Toward convolutional blind denoising of real photographs. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1712–1722, 2019. 2

[13] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017. 5

[14] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 3

[15] Xuhui Jia, Yang Zhao, Kelvin CK Chan, Yandong Li, Han Zhang, Boqing Gong, Tingbo Hou, Huisheng Wang, and Yu-Chuan Su. Taming encoder for zero fine-tuning image customization with text-to-image diffusion models. *arXiv preprint arXiv:2304.02642*, 2023. 3

[16] Zahra Kadkhodaie and Eero Simoncelli. Stochastic solutions for linear inverse problems using the prior implicit in a denoiser. *Advances in Neural Information Processing Systems*, 34:13242–13254, 2021. 3

[17] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019. 5

[18] Bahjat Kawar, Gregory Vaksman, and Michael Elad. Snips: Solving noisy inverse problems stochastically. *Advances in Neural Information Processing Systems*, 34:21757–21769, 2021. 3

[19] Bahjat Kawar, Gregory Vaksman, and Michael Elad. Stochastic image denoising by sampling from the posterior distribution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1866–1875, 2021. 3

[20] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *Advances in Neural Information Processing Systems*, 35:23593–23606, 2022. 2, 3

[21] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5148–5157, 2021. 5

[22] Nupur Kumari, Bingliang Zhang, Richard Zhang, Eli Shechtman, and Jun-Yan Zhu. Multi-concept customization of text-to-image diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1931–1941, 2023. 3

[23] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8183–8192, 2018. 2

[24] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8878–8887, 2019. 2

[25] Wei-Sheng Lai, Yichang Shih, Lun-Cheng Chu, Xiaotong Wu, Sung-Fang Tsai, Michael Krainin, Deqing Sun, and Chia-Kai Liang. Face deblurring using dual camera fusion on mobile phones. *ACM Transactions on Graphics (TOG)*, 41(4):1–16, 2022. 5

[26] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017. 2

[27] Xiaoming Li, Ming Liu, Yuting Ye, Wangmeng Zuo, Liang Lin, and Ruigang Yang. Learning warped guidance for blind face restoration. In *Proceedings of the European conference on computer vision (ECCV)*, pages 272–289, 2018. 3

[28] Xiaoming Li, Wenyu Li, Dongwei Ren, Hongzhi Zhang, Meng Wang, and Wangmeng Zuo. Enhanced blind face restoration with multi-exemplar images and adaptive spatial feature fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2706–2715, 2020. 3, 6, 7

[29] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1833–1844, 2021. 2

[30] Xingang Pan, Xiaohang Zhan, Bo Dai, Dahua Lin, Chen Change Loy, and Ping Luo. Exploiting deep generative prior for versatile image restoration and manipulation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):7474–7489, 2021. 3

[31] Yaniv Romano, Michael Elad, and Peyman Milanfar. The little engine that could: Regularization by denoising (red). *SIAM Journal on Imaging Sciences*, 10(4):1804–1844, 2017. 2, 3

[32] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992. 2

[33] Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22500–22510, 2023. 3

[34] Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 Conference Proceedings*, pages 1–10, 2022. 2

[35] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4713–4726, 2022. 2

[36] Jing Shi, Wei Xiong, Zhe Lin, and Hyun Joon Jung. Instantbooth: Personalized text-to-image generation without test-time finetuning. *arXiv preprint arXiv:2304.03411*, 2023. 3

[37] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020. 3

[38] Jiaming Song, Qinsheng Zhang, Hongxu Yin, Morteza Mardani, Ming-Yu Liu, Jan Kautz, Yongxin Chen, and Arash Vahdat. Loss-guided diffusion models for plug-and-play controllable generation. In *Proceedings of the 40th International Conference on Machine Learning*, pages 32483–32498. PMLR, 2023. 4, 5

[39] Chunwei Tian, Yong Xu, Zuoyong Li, Wangmeng Zuo, Lunke Fei, and Hong Liu. Attention-guided cnn for image denoising. *Neural Networks*, 124:117–129, 2020. 2

[40] Xintao Wang, Yu Li, Honglun Zhang, and Ying Shan. Towards real-world blind face restoration with generative facial prior. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9168–9178, 2021. 2, 5

[41] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1905–1914, 2021. 2

[42] Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. In *The Eleventh International Conference on Learning Representations*, 2022. 3

[43] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17683–17693, 2022. 2

[44] Zhixin Wang, Ziying Zhang, Xiaoyun Zhang, Huangjie Zheng, Mingyuan Zhou, Ya Zhang, and Yanfeng Wang. Dr2: Diffusion-based robust degradation remover for blind face restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1704–1713, 2023. 3, 5, 6, 7

[45] Jay Whang, Mauricio Delbracio, Hossein Talebi, Chitwan Saharia, Alexandros G Dimakis, and Peyman Milanfar. Deblurring via stochastic refinement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16293–16303, 2022. 2

[46] Zhihao Xia and Ayan Chakrabarti. Identifying recurring patterns with deep neural networks for natural image denoising. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2426–2434, 2020. 2

[47] Guangxuan Xiao, Tianwei Yin, William T Freeman, Frédo Durand, and Song Han. Fastcomposer: Tuning-free multi-subject image generation with localized attention. *arXiv preprint arXiv:2305.10431*, 2023. 3

[48] Tao Yang, Peiran Ren, Xuansong Xie, and Lei Zhang. Gan prior embedded network for blind face restoration in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 672–681, 2021. 2

[49] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5728–5739, 2022. 2

[50] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017. 2

[51] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. Learning deep cnn denoiser prior for image restoration. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3929–3938, 2017. 2

[52] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. *IEEE Transactions on Image Processing*, 27(9):4608–4622, 2018. 2

[53] Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte. Plug-and-play image restoration with deep denoiser prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):6360–6376, 2021. 2, 3

[54] Yulun Zhang, Kunpeng Li, Kai Li, Bineng Zhong, and Yun Fu. Residual non-local attention networks for image restoration. *arXiv preprint arXiv:1903.10082*, 2019. 2

[55] Yang Zhao, Tingbo Hou, Yu-Chuan Su, Xuhui Jia, Yandong Li, and Matthias Grundmann. Towards authentic face restoration with iterative diffusion models and beyond. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7312–7322, 2023. 3

[56] Shangchen Zhou, Kelvin Chan, Chongyi Li, and Chen Change Loy. Towards robust blind face restoration with codebook lookup transformer. *Advances in Neural Information Processing Systems*, 35:30599–30611, 2022. 1, 2, 5, 6, 7