Residual Denoising Diffusion Models

Jiawei Liu[1,2,3], Qiang Wang[1,4], Huijie Fan[1,2]*, Yinong Wang[5], Yandong Tang[1,2], Liangqiong Qu[5]*
[1]State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences
[2]Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Science
[3]University of Chinese Academy of Sciences [4]Shenyang University [5]The University of Hong Kong
{liujiawei,wangqiang,fanhuijie,ytang}@sia.cn, liangqqu@hku.hk

Abstract

We propose residual denoising diffusion models (RDDM), a novel dual diffusion process that decouples the traditional single denoising diffusion process into residual diffusion and noise diffusion. This dual diffusion framework expands the denoising-based diffusion models, initially uninterpretable for image restoration, into a unified and interpretable model for both image generation and restoration by introducing residuals. Specifically, our residual diffusion represents directional diffusion from the target image to the degraded input image and explicitly guides the reverse generation process for image restoration, while noise diffusion represents random perturbations in the diffusion process. The residual prioritizes certainty, while the noise emphasizes diversity, enabling RDDM to effectively unify tasks with varying certainty or diversity requirements, such as image generation and restoration. We demonstrate that our sampling process is consistent with that of DDPM and DDIM through coefficient transformation, and propose a partially path-independent generation process to better understand the reverse process. Notably, our RDDM enables a generic UNet, trained with only an L1 loss and a batch size of 1, to compete with state-of-the-art image restoration methods. We provide code and pre-trained models to encourage further exploration, application, and development of our innovative framework (https://github.com/nachifur/RDDM).

1. Introduction

In real-life scenarios, diffusion often occurs in complex forms involving multiple, concurrent processes, such as the dispersion of multiple gases or the propagation of different types of waves or fields. This leads us to ponder whether the denoising-based diffusion models [17, 51] have limitations in focusing solely on denoising. Current diffusion-based image restoration methods [22, 39, 48, 49, 82] extend the diffusion model to image restoration tasks by using degraded images as a condition input to implicitly guide the reverse generation process, without modifying the original denoising diffusion process [17, 51]. However, the reverse process starting from noise seems to be unnecessary, as the degraded image is already known. The forward process is non-interpretability for image restoration, as the diffusion process does not contain any information about the degraded image, as shown in Fig. 1(a).

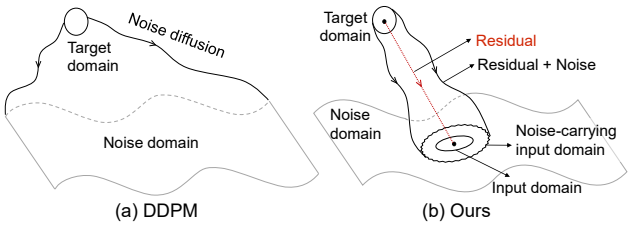

Figure 1. Denoising diffusion process - DDPM [17] (a) and our residual denoising diffusion process (b). For image restoration, we introduce residual diffusion to represent the diffusion direction from the target image to the input image.

In this paper, we explore a novel dual diffusion process and propose Residual Denoising Diffusion Models (RDDM), which can tackle the non-interpretability of a single denoising process for image restoration. In RDDM, we decouple the previous diffusion process into residual diffusion and noise diffusion. Residual diffusion prioritizes certainty and represents a directional diffusion from the target image to the conditional input image, and noise diffusion emphasizes diversity and represents random perturbations in the diffusion process. Thus, our RDDM can unify different tasks that require different certainty or diversity, e.g., image generation and restoration. Compared to denoising-based diffusion models for image restoration, the residuals in RDDM clearly indicate the forward diffusion direction and explicitly guide the reverse generation process for image restoration, as shown in Fig. 1(b).

Specifically, we redefine a new forward process that al-

*Corresponding author.

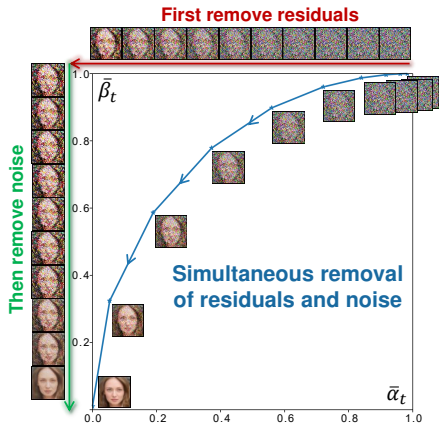

Figure 2. Decoupled dual diffusion framework. The previous forward diffusion process is decoupled into residual diffusion and noise diffusion, while in the reverse process, the simultaneous sampling can be decoupled into first removing the residuals and then removing noise.

lows simultaneous diffusion of residuals and noise, wherein the target image progressively diffuses into a purely noisy image for image generation or a noise-carrying input image for image restoration. Unlike the previous denoising diffusion model [17, 51], which uses one coefficient schedule to control the mixing ratio of noise and images, our RDDM employs two independent coefficient schedules to control the diffusion speed of residuals and noise. We found that this independent diffusion property is also evident in the reverse generation process, e.g., readjusting the coefficient schedule within a certain range during testing does not affect the image generation results, and removing the residuals firstly, followed by denoising (see Fig. 2), can also produce semantically consistent images. Our RDDM is compatible with widely used denoising diffusion models, i.e., our sampling process is consistent with that of DDPM [17] and DDIM [51] by transforming coefficient schedules. In addition, our RDDM natively supports conditional inputs, enabling networks trained with only an ℓ_1 loss and a batch size of 1 to compete with state-of-the-art image restoration methods. We envision that our models can facilitate a unified and interpretable image-to-image distribution transformation methodology, highlighting that residuals and noise are equally important for diffusion models, e.g., the residual prioritizes certainty while the noise emphasizes diversity. The contributions of this paper are summarized as follows:

- We propose a novel dual diffusion framework to tackle the non-interpretability of a single denoising process for image restoration by introducing residuals. Our residual diffusion represents a directional diffusion from the target image to the conditional input image.
- We introduce a partially path-independent generation process that decouples residuals and noise, highlighting their roles in controlling directional residual shift (certainty) and random perturbation (diversity), respectively.
- We design an automatic objective selection algorithm to choose whether to predict residuals or noise for unknown new tasks.
- Extensive experiments demonstrate that our method can be adapted to different tasks, e.g., image generation, restoration, inpainting and translation, focusing certainty or diversity, and involving paired or unpaired data.

2. Related Work

Denoising diffusion models (e.g., DDPM [17], SGM [52, 53], and DDIM [51]) were initially developed for image generation. Subsequent image restoration methods [14, 39, 48] based on DDPM and DDIM feed a degraded image as a conditional input to a denoising network, e.g., DvSR [62], SR3 [49], and WeatherDiffusion [82], which typically require large sampling steps and batch sizes. Additionally, the reverse process starting from noise in these methods seems unnecessary and inefficient for image restoration tasks. Thus, SDEdit [41], ColdDiffusion [2], InDI [11], and I2SB [29] propose generating a clear image directly from a degraded image or noise-carrying degraded image. InDI [11] and I2SB [29], which also present unified image generation and restoration frameworks, are the most closely related to our proposed RDDM. Specifically, the forward diffusion of InDI, I2SB, and our RDDM consistently employs a mixture of three terms (i.e., input images I_{in}, target images I_0, and noise ϵ), extending beyond the denoising-based diffusion model [17, 51] which incorporates a mixture of two terms (i.e., I_0 and ϵ). However, InDI [11] and I2SB [29] opt for estimating the target image or its linear transformation term to replace the noise estimation, akin to a special case of our RDDM (SM-Res). In contrast, we introduce residual estimation while also embracing noise for both generation and restoration tasks. Our RDDM can further extend DDPM [17], DDIM [51], InDI [11], and I2SB [29] to independent double diffusion processes, and pave the way for the multi-dimensional diffusion process. We highlight that residuals and noise are equally important, e.g., the residual prioritizes certainty while the noise emphasizes diversity. In addition, our work is related to coefficient schedule design [44, 48], variance strategy optimization [3, 4, 24, 44], superimposed image decomposition [12, 81], curve integration [47], stochastic differential equations [53], and residual learning [15] for image restoration [1, 32, 56, 70, 72, 75]. See Appendix A.5 for detailed comparison.

3. Background

Denoising diffusion models [17, 50] aim to learn a distribution $p_\theta(I_0) := \int p_\theta(I_{0:T})dI_{1:T}$[1] to approximate a tar-

[1]To understand diffusion from an image perspective, we use I instead of x in DDPM [17].

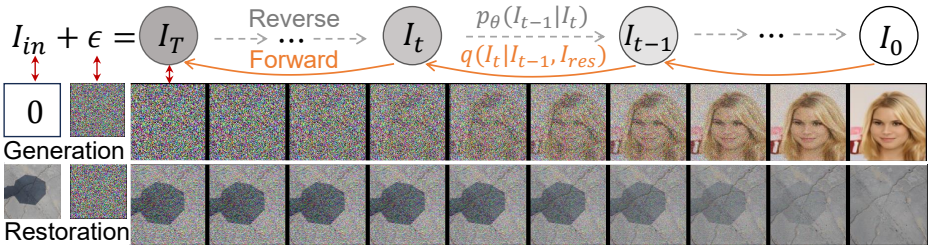

Figure 3. The proposed residual denoising diffusion model (RDDM) is a unified framework for image generation and restoration (a shadow removal task is shown here). We introduce residuals (I_{res}) in RDDM, redefining the forward diffusion process to involve simultaneous diffusion of residuals and noise. The residuals ($I_{res} = I_{in} - I_0$) diffusion represents the directional diffusion from the target image I_0 to the degraded input image I_{in}, while the noise (ϵ) diffusion represents the random perturbations in the diffusion process. In RDDM, I_0 gradually diffuses into $I_T = I_{in} + \epsilon$, $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. In the third columns, I_T is a purely noisy image for image generation since $I_{in} = 0$, and a noise-carrying degraded image for image restoration as I_{in} is the degraded image.

get data distribution $q(I_0)$, where I_0 are target images and $I_1, \ldots, I_T$ ($T = 1000$) are latent images of the same dimension as I_0. In the forward process, $q(I_0)$ is diffused into a Gaussian noise distribution using a fixed Markov chain,

$$q(I_{1:T}|I_0) := \prod_{t=1}^{T} q(I_t|I_{t-1}), \tag{1}$$
$$q(I_t|I_{t-1}) := \mathcal{N}(I_t; \sqrt{\alpha_t} I_{t-1}, (1-\alpha_t)\mathbf{I}), \tag{2}$$

where $\alpha_{1:T} \in (0, 1]^T$. $q(I_t|I_{t-1})$ can also be written as $I_t = \sqrt{\alpha_t} I_{t-1} + \sqrt{1-\alpha_t}\epsilon_{t-1}$. In fact, it is simpler to sampling I_t from I_0 by reparameterization [25, 26],

$$I_t = \sqrt{\bar{\alpha}_t} I_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, \tag{3}$$

where $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $\bar{\alpha}_t := \prod_{s=1}^{t} \alpha_s$. The reverse process is also a Markov chain starting at $p_\theta(I_T) \sim \mathcal{N}(I_T; \mathbf{0}, \mathbf{I})$,

$$p_\theta(I_{0:T}) := p_\theta(I_T) \prod_{t=1}^{T} p_\theta(I_{t-1}|I_t), \tag{4}$$
$$p_\theta(I_{t-1}|I_t) := \mathcal{N}(I_{t-1}; \mu_\theta(I_t, t), \Sigma_t \mathbf{I}), \tag{5}$$

where $p_\theta(I_{t-1}|I_t)$ is a learnable transfer probability (the variance schedule Σ_t is fixed). A simplified loss function [17] is derived from the maximum likelihood of $p_\theta(I_0)$, i.e., $L(\theta) := \mathbb{E}_{I_0 \sim q(I_0), \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})} \left[\|\epsilon - \epsilon_\theta(I_t, t)\|^2 \right]$. The estimated noise ϵ_θ can be used to represent μ_θ in $p_\theta(I_{t-1}|I_t)$, thus I_{t-1} can be sampled from $p_\theta(I_{t-1}|I_t)$ step by step.

4. Residual Denoising Diffusion Models

Our goal is to develop a dual diffusion process to unify and interpret image generation and restoration. We modify the representation of $I_T = \epsilon$ in traditional DDPM to $I_T = I_{in} + \epsilon$ in our RDDM, where I_{in} is a degraded image (e.g., a shadow, low-light, or blurred image) for image restoration and is set to 0 for image generation. This modification is compatible with the widely used denoising diffusion model, e.g., $I_T = 0 + \epsilon$ is the pure noise (ϵ) for generation. For image restoration, I_T is a noisy-carrying degraded image ($I_{in} + \epsilon$), as shown in the third column in Fig. 3. The modified forward process from I_0 to $I_T = I_{in} + \epsilon$ involves progressively degrading I_0 to I_{in}, and injecting noise ϵ. This naturally results in a dual diffusion process, a residual diffusion to model the transition from I_0 to I_{in} and a noise diffusion. For example, the forward diffusion process from the shadow-free image I_0 to the noisy carrying shadow image I_T involves progressively adding shadows and noise, as shown in the second row in Fig. 3.

In the following subsections, we detail the underlying theory and the methodology behind our RDDM. Inspired by residual learning [15, 31, 32], we redefine each forward diffusion process step in Section 4.1. For the reverse process, we present a training objective to predict the residuals and noise injected in the forward process in Section 4.2. In Section 4.3, we propose three sampling methods, i.e., residual prediction (SM-Res), noise prediction (SM-N), and "residual and noise prediction" (SM-Res-N).

4.1. Directional Residual Diffusion Process with Perturbation

To model the gradual degradation of image quality and the increment of noise, we define the single forward process step in our RDDM as follows:

$$I_t = I_{t-1} + I_{res}^t, \qquad I_{res}^t \sim \mathcal{N}(\alpha_t I_{res}, \beta_t^2 \mathbf{I}), \tag{6}$$

where I_{res}^t represents a directional mean shift (residual diffusion) with random perturbation (noise diffusion) from state I_{t-1} to state I_t, the residuals I_{res} in I_{res}^t is the difference between I_{in} and I_0 (i.e., $I_{res} = I_{in} - I_0$), and two independent coefficient schedules α_t and β_t control the residual and noise diffusion, respectively. In fact, it is simpler to sample I_t from I_0 (like Eq. 3),

$$\begin{aligned} I_t =& I_{t-1} + \alpha_t I_{res} + \beta_t \epsilon_{t-1}, \\ =& I_{t-2} + (\alpha_{t-1} + \alpha_t) I_{res} + (\sqrt{\beta_{t-1}^2 + \beta_t^2})\epsilon_{t-2} \\ =& \ldots \\ =& I_0 + \bar{\alpha}_t I_{res} + \bar{\beta}_t \epsilon, \end{aligned} \tag{7}$$

where $\epsilon_{t-1}, \dots \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $\bar{\alpha}_t = \sum_{i=1}^{t} \alpha_i$ and $\bar{\beta}_t = \sqrt{\sum_{i=1}^{t} \beta_i^2}$. If $t = T$, $\bar{\alpha}_T = 1$ and $I_T = I_{in} + \bar{\beta}_T \epsilon$. $\bar{\beta}_T$ can control the intensity of noise perturbation for image restoration (e.g., $\bar{\beta}_T^2 = 0.01$ for shadow removal), while $\bar{\beta}_T^2 = 1$ for image generation. From Eq. 6, the joint probability distributions in the forward process can be defined as:

$$q(I_{1:T}|I_0, I_{res}) := \prod_{t=1}^{T} q(I_t|I_{t-1}, I_{res}), \tag{8}$$

$$q(I_t|I_{t-1}, I_{res}) := \mathcal{N}(I_t; I_{t-1} + \alpha_t I_{res}, \beta_t^2 \mathbf{I}). \tag{9}$$

Eq. 7 defines the marginal probability distribution $q(I_t|I_0, I_{res}) = \mathcal{N}(I_t; I_0 + \bar{\alpha}_t I_{res}, \bar{\beta}_t^2 \mathbf{I})$. In fact, the forward diffusion of our RDDM is a mixture of three terms (i.e., I_0, I_{res}, and ϵ), extending beyond the widely used denoising diffusion model that is a mixture of two terms, i.e, I_0 and ϵ. A similar mixture form of three terms can be seen in several concurrent works, e.g., InDI [11], I2SB [29], IR-SDE [40], and ResShift [67].

4.2. Generation Process and Training Objective

In the forward process (Eq. 7), residuals (I_{res}) and noise (ϵ) are gradually added to I_0, and then synthesized into I_t, while the reverse process from I_T to I_0 involves the estimation of the residuals and noise injected in the forward process. We can train a residual network $I_{res}^{\theta}(I_t, t, I_{in})$ to predict I_{res} and a noise network $\epsilon_\theta(I_t, t, I_{in})$ to estimate ϵ. Using Eq. 7, we obtain the estimated target images $I_0^{\theta} = I_t - \bar{\alpha}_t I_{res}^{\theta} - \bar{\beta}_t \epsilon_\theta$. If I_0^{θ} and I_{res}^{θ} are given, the generation process is defined as,

$$p_\theta(I_{t-1}|I_t) := q_\sigma(I_{t-1}|I_t, I_0^{\theta}, I_{res}^{\theta}), \tag{10}$$

where the transfer probability $q_\sigma(I_{t-1}|I_t, I_0, I_{res})$[2] from I_t to I_{t-1} is,

$$\begin{aligned} q_\sigma(I_{t-1}|I_t, I_0, I_{res}) &= \mathcal{N}(I_{t-1}; I_0 + \bar{\alpha}_{t-1} I_{res} \\ &+ \sqrt{\bar{\beta}_{t-1}^2 - \sigma_t^2} \frac{I_t - (I_0 + \bar{\alpha}_t I_{res})}{\bar{\beta}_t}, \sigma_t^2 \mathbf{I}), \end{aligned} \tag{11}$$

where $\sigma_t^2 = \eta \beta_t^2 \bar{\beta}_{t-1}^2 / \bar{\beta}_t^2$ and η controls whether the generation process is random ($\eta = 1$) or deterministic ($\eta = 0$). Using Eq. 10 and Eq. 11, I_{t-1} can be sampled from I_t via:

$$\begin{aligned} I_{t-1} = I_t &- (\bar{\alpha}_t - \bar{\alpha}_{t-1}) I_{res}^{\theta} \\ &- (\bar{\beta}_t - \sqrt{\bar{\beta}_{t-1}^2 - \sigma_t^2}) \epsilon_\theta + \sigma_t \epsilon_t, \end{aligned} \tag{12}$$

where $\epsilon_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. When $\eta = 1$, our RDDM has the sum-constrained variance, while DDPM has preserving variance (see Appendix A.4). When $\eta = 0$ (i.e., $\sigma_t = 0$), the sampling process is deterministic,

$$I_{t-1} = I_t - (\bar{\alpha}_t - \bar{\alpha}_{t-1}) I_{res}^{\theta} - (\bar{\beta}_t - \bar{\beta}_{t-1}) \epsilon_\theta. \tag{13}$$

[2]Eq. 11 does not change $q(I_t|I_0, I_{res})$ in Appendix A.2.

We derive the following simplified loss function for training (Appendix A.1):

$$L_{res}(\theta) := \mathbb{E}\left[\lambda_{res} \left\| I_{res} - I_{res}^{\theta}(I_t, t, I_{in}) \right\|^2\right], \tag{14}$$

$$L_{\epsilon}(\theta) := \mathbb{E}\left[\lambda_{\epsilon} \left\| \epsilon - \epsilon_\theta(I_t, t, I_{in}) \right\|^2\right], \tag{15}$$

where the hyperparameters $\lambda_{res}, \lambda_\epsilon \in \{0, 1\}$, and the training input image I_t is synthesized using I_0, I_{res}, and ϵ by Eq. 7. I_t can also be synthesized using I_{in} (replace I_0 in Eq. 7 by $I_0 = I_{in} - I_{res}$),

$$I_t = I_{in} + (\bar{\alpha}_t - 1) I_{res} + \bar{\beta}_t \epsilon. \tag{16}$$

4.3. Sampling Method Selection Strategies

For the generation process (from I_t to I_{t-1}), I_t and I_{in} are known, and thus I_{res} and ϵ can represent each other by Eq. 16. From Eq. 14, 15, 16, we propose three sampling methods as follows.

SM-Res. When $\lambda_{res} = 1$ and $\lambda_\epsilon = 0$, the residuals I_{res}^{θ} are predicted by a network, while the noise ϵ_θ is represented as a transformation of I_{res}^{θ} using Eq. 16.

SM-N. When $\lambda_{res} = 0$ and $\lambda_\epsilon = 1$, the noise ϵ_θ is predicted by a network, while the residuals I_{res}^{θ} are represented as a transformation of ϵ_θ using Eq. 16.

SM-Res-N. When $\lambda_{res} = 1$ and $\lambda_\epsilon = 1$, both the residuals and the noise are predicted by networks.

To determine the optimal sampling method for real-world applications, we give empirical strategies and automatic selection algorithms in the following.

Empirical Research. Table 1 presents that the SM-Res shows better results for image restoration but offers a poorer FID for generation. On the other hand, the SM-N yields better frechet inception distance (FID in [16]) and inception scores (IS), but is ineffective in image restoration (e.g., PSNR 11.34 for shadow and 16.30 for low-light). This may be due to the inadequacy of using ϵ_θ to represent I_{res}^{θ} in Eq. 16 for restoration tasks. We attribute these inconsistent results to the fact that **residual predictions prioritize certainty, whereas noise predictions emphasize diversity**. In our experiments, we use SM-N for image generation, SM-Res for low-light (LOL [61]), and SM-Res-N for other image restoration tasks. For an unknown new task, we empirically recommend using SM-N for those requiring greater diversity and SM-Res for tasks that demand higher certainty.

Automatic Objective Selection Algorithm (AOSA). To automatically choose between SM-Res or SM-N for an unknown task, we develop an automatic sampling selection algorithm in Appendix B.2. This algorithm requires only a single network and learns the hyperparameter in Eq. 15, enabling a gradual transition from combined residual and noise training (akin to SM-Res-N) to individual prediction (SM-Res or SM-N). This plug-and-play training strategy requires less than 1000 additional training iterations and is

Sampling Method	Generation (CelebA)		Shadow removal (ISTD)			Low-light (LOL)		Deraining (RainDrop)	
	FID (↓)	IS (↑)	MAE(↓)	PSNR(↑)	SSIM(↑)	PSNR(↑)	SSIM(↑)	PSNR(↑)	SSIM(↑)
SM-Res	31.47	1.73	4.76	30.72	0.959	**25.39**	**0.937**	31.96	0.9509
SM-N	**23.25**	**2.05**	81.01	11.34	0.175	16.30	0.649	19.15	0.7179
SM-Res-N	28.90	1.78	**4.67**	**30.91**	**0.962**	23.90	0.931	**32.51**	**0.9563**

Table 1. Sampling method analysis. The sampling steps are 10 on the CelebA 64 × 64 [36] dataset, 5 on the ISTD [57] dataset, 2 on the LOL [61] dataset, and 5 on the RainDrop [45] dataset.

fully compatible with the current denoising-based diffusion methods [17]. Our RDDM using AOSA has the potential to provide a unified and interpretable methodology for modeling, training, and inference pipelines for unknown target tasks.

Comparison with Other Prediction Methods. Our SM-N is similar to DDIM [51] (or DDPM [17]), which only estimates the noise, and is consistent with DDPM and DDIM by transforming the coefficient/variance schedules in Eq. 12 (the proof in Appendix A.3),

$$\bar{\alpha}_t = 1 - \sqrt{\bar{\alpha}^t_{DDIM}}^{\,3}, \bar{\beta}_t = \sqrt{1-\bar{\alpha}^t_{DDIM}}, \\ \sigma_t^2 = \sigma_t^2(DDIM). \tag{17}$$

In fact, current research has delved into numerous diffusion forms that extend beyond noise estimation. For example, IDDPM [44] proposes that it is feasible to estimate noise (ϵ_θ), clean target images (I_0^θ), or the mean term (μ_θ) to represent the transfer probabilities (i.e., $p_\theta(I_{t-1}|I_t)$ in Eq. 5). The score-based generative model (SGM) [52] and Schrödinger Bridge (I2SB [29]) estimate the score of noisy data (i.e., the sum of residuals and noise $\sum_{i=1}^{t} I^t_{res}$). Cold-Diffusion [2] and InDI [11] estimate the clean target images (I_0). Rectified Flow [35] predicts the residuals (I_{res}) to align with the image linear interpolation process without noise diffusion (i.e., $I_T = I_{in}$). A detailed comparison can be found in Appendix A.5.

These previous/concurrent works choose to estimate the noise, the residual, the target image, or its linear transformation term. In contrast, we introduce residual estimation while also embracing noise for both generation and restoration. Residuals and noise have equal and independent status, which is reflected in the forward process (Eq. 7), the reverse process (Eq. 13), and the loss function (Eq. 15). This independence means that the noise diffusion can even be removed and only the residual diffusion retained to model the image interpolation process (when $\bar{\beta}_T = 0$ in Eq. 7, RDDM degenerates to Rectified Flow [35]). In addition, this property derives a decoupled dual diffusion framework in Section 5.

5. Decoupled Dual Diffusion Framework

Upon examining DDPM from the perspective of RDDM, we discover that DDPM indeed involves the simultaneous diffusion of residuals and noise, which is evident as Eq. 48 becomes equivalent to Eq. 44 in Appendix A.3. We find that it is possible to decouple these two types of diffusion. Section 5.1 presents a decoupled forward diffusion process. In Section 5.2, we propose a partially path-independent generation process and decouple the simultaneous sampling into first removing the residuals and then removing noise (see Fig. 6(d) and Fig. 17). This decoupled dual diffusion framework sheds light on the roles of deresidual and denoising in the DDPM generation process.

[3] $\bar{\alpha}^t_{DDIM}$ here is α_t of DDIM [51].

Schedules	FID (↓)	IS (↑)
Linear (DDIM [51])	28.39[4]	2.05
Scaled linear [48]	28.15	2.00
Squared cosine [44]	47.21	**2.64**
α_t (mean), β_t^2 (mean)	38.35	2.22
α_t (linearly increasing), β_t^2 (linearly increasing)	40.03	2.45
α_t (linearly decreasing), β_t^2 (linearly decreasing)	27.82	2.26
α_t (linearly decreasing), β_t^2 (linearly increasing)	**23.25**	2.05

Table 2. Coefficient schedules analysis on CelebA (64 × 64) [36]. In our RDDM, the residual diffusion and noise diffusion are decoupled, so one may design a better schedule in the decoupled coefficient space, e.g., α_t (linearly decreasing), β_t^2 (linearly increasing). To be fair, all coefficient schedules were retrained using the same network structure, training, and evaluation. The sampling method is SM-N with 10 sampling steps using Eq. 13.

5.1. Decoupled Forward Diffusion Process

Our defined coefficients (α_t, β_t^2) offer a distinct physical interpretation. In the forward diffusion process (Eq. 7), α_t controls the speed of residual diffusion and β_t^2 regulates the speed of noise diffusion. In the reverse generation process (Eq. 13), $\bar{\alpha}_t$ and $\bar{\beta}_t$ are associated with the speed of removing residual and noise, respectively. In fact, there are no constraints on α_t and β_t^2 in Eq. 7, meaning that the residual diffusion and noise diffusion are independent of each other. Utilizing this decoupled property and the difference between these two diffusion processes, we should be able to design a better coefficient schedule, e.g., α_t (linearly decreasing) and β_t^2 (linearly increasing) in Table 2. This aligns with the intuition that, during the reverse generation pro-

[4] Our RDDM is implemented based on the popular diffusion repository github.com/lucidrains/denoising-diffusion-pytorch. Differences in network structure and training details may lead to poorer FID. We have verified sampling consistency with DDIM [51] in Table 3(a) and Appendix A.3.

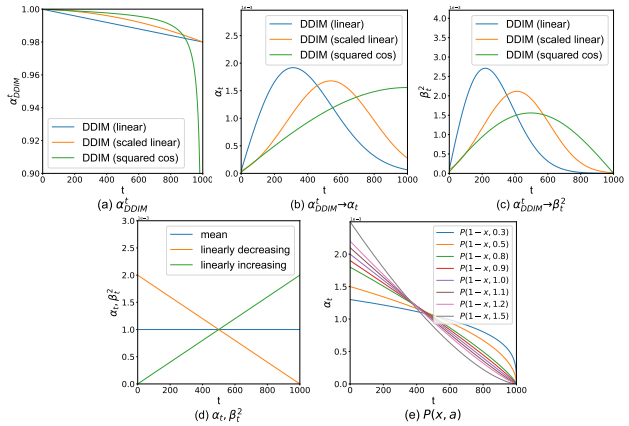

Figure 4. Coefficient transformation from DDIM [51] to RDDM using Eq. 17. (a) We show several schedules for α^t_{DDIM}, e.g., linear [51], scaled linear [48], and squared cosine [44]. (b) We transform α^t_{DDIM} into α_t in our RDDM. (c) We transform α^t_{DDIM} into β^2_t in our RDDM. (d) A few simple schedules. (e) $P(x, a)$ is a normalized power function (see Eq. 18). "mean", "linearly increasing", and "linearly decreasing" in (d) can be denoted as $P(x, 0)$, $P(x, 1)$ and $P(1 - x, 1)$, respectively. See Algorithm 1 in Appendix A.3 for more details of (b) and (c).

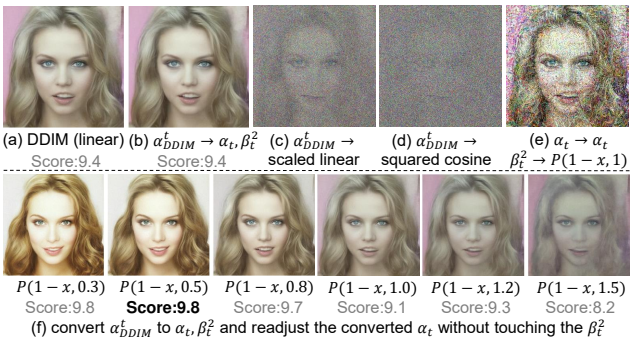

Figure 5. Analysis of readjusting coefficient schedules. We find that changing the α_t schedule barely affects the denoising process in (f) and edited faces may have higher face scores when assessed using AI face scoring software[5]. These images were generated using a pre-trained UNet on the CelebA (256×256) dataset [36] with 10 sampling steps.

cess (from T to 0), the estimated residuals become increasingly accurate while the estimated noise should also weaken progressively. Therefore, when t is close to 0, the deresidual pace should be faster and the denoising pace should be slower. Since our α_t and β^2_t represent the speed of diffusion, we name the curve in Fig. 4 (b-d) the *diffusion speed curve*.

5.2. Partially Path-independent Generation Process

In the original DDPM [17] or DDIM [51], when the α^t_{DDIM} schedule changes, it is necessary to retrain the denoising network because this alters the diffusion process [44, 48]. As shown in Fig. 5(c)(d), directly changing the α^t_{DDIM} schedule causes denoising to fail. Here, we propose a path-independent generation process, i.e., modifying the diffusion speed curve does not cause the image generation process to fail. We try to readjust the diffusion speed curve in the generation process. First, we convert the α^t_{DDIM} schedule of a pre-trained DDIM into the α_t and β^2_t schedules of our RDDM using Eq. 17 (from Fig. 5(a) to Fig. 5(b). We then readjust the converted α_t schedules using the normalized power function ($P(x, a)$ in Fig. 5(f)), without touching the β^2_t schedule that controls noise diffusion, as shown in Fig. 5(f). $P(x, a)$ is defined as (a is a parameter of the power function),

$$P(x, a) := x^a / \int_0^1 x^a dx, \text{where } x = t/T. \tag{18}$$

These schedule modifications shown in Fig. 5 lead to the following key findings.

1. Fig. 5(f) shows that modifying the residual diffusion speed curve (α_t) leads to a drastic change in the generation results, probably due to I^θ_{res} being represented as a transformation of ϵ_θ using Eq. 16.

2. As the time condition t represents the current noise intensity in the denoising network ($\epsilon_\theta(I_t, t, 0)$), modifying the noise diffusion speed curve (β^2_t) causes t to deviate from accurately indicating the current noise intensity, leading to denoising failure, as shown in Fig. 5(e).

Nonetheless, we believe that, corresponding to the decoupled forward diffusion process, there should also be a path-independent reverse generation process. To develop a path-independent generation process, we improve the generation process based on the above two key findings:

1. Two networks are used to estimate I^θ_{res} and ϵ_θ separately, i.e., SM-Res-N-2Net in Appendix B.2.

2. $\bar{\alpha}_t$ and $\bar{\beta}_t$ are used for the time conditions embedded in the network, i.e., $I^\theta_{res}(I_t, t, 0) \to I^\theta_{res}(I_t, \bar{\alpha}_t \cdot T, 0)$, $\epsilon_\theta(I_t, t, 0) \to \epsilon_\theta(I_t, \bar{\beta}_t \cdot T, 0)$.

These improvements lead to a partially path-independent generation process, as evidenced by the results shown in Fig. 6(c).

Analysis of Partially Path-independence via Green's Theorem. "Path-independence" reminds us of Green's theorem in curve integration [47]. From Eq. 13, we have:

$$I_t - I_{t-1} = (\bar{\alpha}_t - \bar{\alpha}_{t-1}) I^\theta_{res} + (\bar{\beta}_t - \bar{\beta}_{t-1}) \epsilon_\theta, \tag{19}$$

$$\begin{aligned} \mathrm{d}I(t) =& I^\theta_{res}(I(t), \bar{\alpha}(t) \cdot T, 0) \mathrm{d}\bar{\alpha}(t) \\ &+ \epsilon_\theta(I(t), \bar{\beta}(t) \cdot T, 0) \mathrm{d}\bar{\beta}(t), \end{aligned} \tag{20}$$

where $I(t) = I(0) + \bar{\alpha}(t) I_{res} + \bar{\beta}(t)\epsilon$. Given inputs $I(t)$ and $\bar{\alpha}(t)$, the denoising network learns to approximate the noise ϵ in $I(t)$ by estimating ϵ_θ. If this network is trained well and robust enough, it should be able to avoid the interference of

[5] https://ux.xiaoice.com/beautyv3

Figure 6. Partially path-independent generation process. (a1) We trained a denoising network using the DDIM linear schedule [51]. (a2-a3) We modified the α_t and β_t^2 schedules during testing. (b) We trained two networks to remove noise and residuals. In contrast to the sharply varying images in (a2-a3) and the noisy images in (b2-b3), (c) shows that we constructed a path independent generation process where modifications to the diffusion speed curve can generate a noise-free image with little variation in image semantics. (d) The simultaneous sampling in (d1) or (c) can be decomposed into first removing residuals and then noise (d2), or removing noise and then residuals (d3). In (d3), diversity is significantly reduced because noise is removed first.

the residual terms $\bar{\alpha}(t)I_{res}$ in $I(t)$. This also applies to a robust residual estimation network. Thus, we have

$$\frac{\partial I_{res}^\theta(I(t),\bar{\alpha}(t)\cdot T)}{\partial \bar{\beta}(t)} \approx 0, \frac{\partial \epsilon_\theta(I(t),\bar{\beta}(t)\cdot T)}{\partial \bar{\alpha}(t)} \approx 0. \quad (21)$$

If the equation in Formula 21 holds true, it serves as a necessary and sufficient condition for path independence in curve integration, which provides an explanation for why Fig. 6(c) achieves a partially path-independent generation process. The path-independent property is related to the network's resilience to disturbances and applies to disturbances that vary within a certain range. However, excessive disturbances can lead to visual inconsistencies, e.g., readjusting α_t and β_t^2 to $P(x,5)$. Thus, we refer to this generative property as partially path-independent. We also investigated two reverse paths to gain insight into the implications of the proposed partial path independence. In the first case, the residuals are removed first, followed by the noise: $I(T) \overset{-I_{res}}{\rightarrow} I(0) + \bar{\beta}_T\epsilon \overset{-\bar{\beta}_T\epsilon}{\rightarrow} I(0)$. The second case involves removing the noise first and then the residuals: $I(T) \overset{-\bar{\beta}_T\epsilon}{\rightarrow} I_{in} \overset{-I_{res}}{\rightarrow} I(0)$. The first case (Fig. 6(d2)) shows

(a) CelebA (FID)	5 steps	10 steps	15 steps	20 steps	100 steps
DDIM	69.60	40.45	32.67	30.61	23.66
DDIM→RDDM	69.60	40.41	32.71	30.77	24.92

(b) Shadow Removal	MAE(↓)			SSIM(↑)			PSNR(↑)		
	S	NS	ALL	S	NS	ALL	S	NS	ALL
DSC [19] ¶	9.48	6.14	6.67	0.967	-	-	33.45	-	-
FusionNet [13]	7.77	5.56	5.92	0.975	0.880	0.945	34.71	28.61	27.19
BMNet [79]	7.60	4.59	5.02	0.988	0.976	0.959	35.61	32.80	30.28
DMTN [31]	7.00	4.28	4.72	**0.990**	**0.979**	**0.965**	35.83	33.01	30.42
Ours (RDDM)	**6.67**	**4.27**	**4.67**	0.988	**0.979**	0.962	**36.74**	**33.18**	**30.91**

(c) Low-light	PSNR(↑)	SSIM(↑)	LPIPS (↓)
KinD++ [76]	17.752	0.760	0.198
KinD++-SKF [68]	20.363	0.805	0.201
DCC-Net [77]	22.72	0.81	-
SNR-Aware [66]	24.608	0.840	0.151
LLFlow [59]	25.19	0.93	**0.11**
LLFormer [58]	23.649	0.816	0.169
Ours (RDDM)	**25.392**	**0.937**	0.116

(d) Deraining	PSNR(↑)	SSIM(↑)
AttnGAN [45]	31.59	0.9170
DuRN [34]	31.24	0.9259
RainAttn [46]	31.44	0.9263
IDT [64]	31.87	0.9313
RainDiff64 [82]	32.29	0.9422
RainDiff128 [82]	32.43	0.9334
Ours (RDDM)	**32.51**	**0.9563**

Table 3. Quantitative comparison results of image generation on the CelebA (256×256) dataset [36], shadow removal on the ISTD dataset [57], low-light enhancement on the LOL [61] dataset, and deraining on the RainDrop [45] dataset. "S, NS, ALL" in (b) denote shadow area (S), non-shadow area (NS) and whole image (ALL). The sampling steps are 5 for shadow removal and deraining, 2 for low-light.

that removing residuals controls semantic transitions, while the second case (Fig. 6(d3)) shows that diversity is significantly reduced because noise is removed first. Fig. 6(d) validates our argument that residuals control directional semantic drift (certainty) and noise controls random perturbation (diversity). See Appendix B.4 for more details.

6. Experiments

Image Generation. We can convert a pre-trained[6] DDIM [51] to RDDM by coefficient transformation using Eq. 17, and generate images by Eq. 12. Table 3(a) verifies that the quality of the generated images before and after the conversion is nearly the same[7]. We show the generated face images with 10 sampling steps in Fig. 7(a).

Image Restoration. We extensively evaluate our method on several image restoration tasks, including shadow removal, low-light enhancement, deraining, and deblurring on 5 datasets. Notably, our RDDM uses an identical UNet and is trained with a batch size of 1 for all these tasks. In contrast, SOAT methods often involve elaborate network architectures, such as multi-stage [13, 59, 80], multi-branch [10], Transformer [58], and GAN [27], or sophisticated loss functions like the chromaticity [20], texture similarity [74], and edge loss [70]. Table 3 and Fig. 7(b-c) show that our RDDM is competitive with the SOTA restora-

[6]https://huggingface.co/google/ddpm-celebahq-256

[7]The subtle differences in larger sampling steps may stem from errors introduced by numerical representation limitations during coefficient transformation, which may accumulate and amplify in larger sampling steps.

Figure 7. Application of our RDDM. (a) Image generation on the CelebA dataset [36]. (b) Shadow removal on the ISTD dataset [57]. (c) Low-light enhancement on the LOL dataset [61]. (d) Image inpainting (center and irregular mask). (e) The image translation process can be regarded as first translating the semantics and then generating the details. These images in (b) are magnified using MulimgViewer [30].

tion methods. See Appendix B for more training details and comparison results.

We extend DDPM [17]/DDIM [51], initially uninterpretable for image restoration, into a unified and interpretable diffusion model for both image generation and restoration by introducing residuals. However, the residual diffusion process represents the directional diffusion from target images to conditional input images, which does not involve a priori information about the image restoration task, and therefore is not limited to it. Beyond image generation and restoration, we show examples of image inpainting and image translation to verify that our RDDM has the potential to be a unified and interpretable methodology for image-to-image distribution transformation. **We do not intend to achieve optimal performance on all tasks by tuning all hyperparameters.** The current experimental results show that RDDM 1) achieves consistent **image generation** performance with DDIM after coefficient transformation, 2) competes with state-of-the-art **image restoration** methods using a generic UNet with only an ℓ_1 loss, a batch size of 1, and fewer than 5 sampling steps, and 3) has satisfactory visual results of **image inpainting** and **image translation** (see Fig. 7(d-e), Fig. 14, or Fig. 15 in Appendix B.3), which validates our RDDM.

7. Conclusions

We present a unified dual diffusion model called Residual Denoising Diffusion Models (RDDM) for image restoration and image generation. This is a three-term mixture framework beyond the previous denoising diffusion framework with two-term mixture. We demonstrate that our sampling process is consistent with that of DDPM and DDIM through coefficient schedule transformation, and propose a partially path-independent generation process. Our experimental results on four different image restoration tasks show that RDDM achieves SOTA performance in no more than five sampling steps. We believe that our model and framework hold the potential to provide a unified methodology for image-to-image distribution transformation and pave the way for the multi-dimensional diffusion process.

8. Acknowledgments

This work was supported by National Natural Science Foundation of China under Grants 61991413, 62273339, 62073205, 62306253.

References

[1] Saeed Anwar and Nick Barnes. Densely residual laplacian super-resolution. *IEEE TPAMI*, 44(3):1192–1204, 2020. 2

[2] Arpit Bansal, Eitan Borgnia, Hong-Min Chu, Jie S Li, Hamid Kazemi, Furong Huang, Micah Goldblum, Jonas Geiping, and Tom Goldstein. Cold diffusion: Inverting arbitrary image transforms without noise. *arXiv preprint arXiv:2208.09392*, 2022. 2, 5, 16

[3] Fan Bao, Chongxuan Li, Jiacheng Sun, Jun Zhu, and Bo Zhang. Estimating the optimal covariance with imperfect mean in diffusion probabilistic models. In *Proc. ICML*, 2022. 2

[4] Fan Bao, Chongxuan Li, Jun Zhu, and Bo Zhang. Analytic-dpm: an analytic estimate of the optimal reverse variance in diffusion probabilistic models. In *Proc. ICLR*, 2022. 2

[5] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*. Springer, 2006. 13

[6] Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand. Learning photographic global tonal adjustment with a database of input / output image pairs. In *Proc. CVPR*, 2011. 21

[7] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proc. CVPR*, pages 3291–3300, 2018. 21

[8] Zipei Chen, Chengjiang Long, Ling Zhang, and Chunxia Xiao. CANet: A Context-Aware Network for Shadow Removal. In *Proc. ICCV*, pages 4723–4732, 2021. 19, 20

[9] Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. Stargan v2: Diverse image synthesis for multiple domains. In *Proc. CVPR*, pages 8188–8197, 2020. 23, 28

[10] Xiaodong Cun, Chi Man Pun, and Cheng Shi. Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting GAN. In *Proc. AAAI*, pages 10680–10687, 2020. 7, 17, 19, 20

[11] Mauricio Delbracio and Peyman Milanfar. Inversion by direct iteration: An alternative to denoising diffusion for image restoration. *Transactions on Machine Learning Research*, 2023. 2, 4, 5, 16, 23, 26

[12] Huiyu Duan, Wei Shen, Xiongkuo Min, Yuan Tian, Jae-Hyun Jung, Xiaokang Yang, and Guangtao Zhai. Develop then rival: A human vision-inspired framework for superimposed image decomposition. *IEEE TMM*, 2022. 2

[13] Lan Fu, Changqing Zhou, Qing Guo, Felix Juefei-Xu, Hongkai Yu, Wei Feng, Yang Liu, and Song Wang. Auto-Exposure Fusion for Single-Image Shadow Removal. In *Proc. CVPR*, pages 10566–10575, 2021. 7, 17, 19, 20

[14] Lanqing Guo, Chong Wang, Wenhan Yang, Siyu Huang, Yufei Wang, Hanspeter Pfister, and Bihan Wen. Shadowdiffusion: When degradation prior meets diffusion model for shadow removal. In *Proc. CVPR*, 2023. 2, 16, 28

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proc. CVPR*, pages 770–778, 2016. 2, 3

[16] 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. In *Proc. NeurIPS*, 2017. 4

[17] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Proc. NeurIPS*, pages 6840–6851, 2020. 1, 2, 3, 5, 6, 8, 12, 13, 14, 15, 16, 19

[18] Weipeng Hu and Haifeng Hu. Domain discrepancy elimination and mean face representation learning for nir-vis face recognition. *IEEE Signal Processing Letters*, 28:2068–2072, 2021. 26

[19] Xiaowei Hu, Chi Wing Fu, Lei Zhu, Jing Qin, and Pheng Ann Heng. Direction-aware spatial context features for shadow detection and removal. *IEEE TPAMI*, 42(11): 2795–2808, 2020. 7, 19, 20

[20] Yeying Jin, Aashish Sharma, and Robby T. Tan. DC-ShadowNet: Single-Image Hard and Soft Shadow Removal Using Unsupervised Domain-Classifier Guided Network. In *Proc. ICCV*, pages 5007–5016, 2021. 7, 17

[21] Yeying Jin, Wenhan Yang, Wei Ye, Yuan Yuan, and Robby T Tan. Des3: Attention-driven self and soft shadow removal using vit similarity and color convergence. *arXiv preprint arXiv:2211.08089*, 2022. 25

[22] Yeying Jin, Wenhan Yang, Wei Ye, Yuan Yuan, and Robby T Tan. Shadowdiffusion: Diffusion-based shadow removal using classifier-driven attention and structure preservation. *arXiv preprint arXiv:2211.08089*, 2022. 1

[23] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. In *Proc. ICLR*, 2018. 23, 26, 28

[24] Diederik Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. In *Proc. NeurIPS*, pages 21696–21707, 2021. 2

[25] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *Proc. ICLR*, 2014. 3

[26] Diederik P. Kingma and Max Welling. An introduction to variational autoencoders. *Found. Trends Mach. Learn.*, 12 (4):307–392, 2019. 3

[27] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *Proc. ICCV*, pages 8878–8887, 2019. 7, 17, 23

[28] Hieu Le and Dimitris Samaras. Shadow removal via shadow image decomposition. In *Proc. ICCV*, pages 8578–8587, 2019. 17

[29] Guan-Horng Liu, Arash Vahdat, De-An Huang, Evangelos A Theodorou, Weili Nie, and Anima Anandkumar. I^2sb: Image-to-image schrödinger bridge. In *Proc. ICML*, 2023. 2, 4, 5, 16, 23

[30] Jiawei Liu. MulimgViewer: A multi-image viewer for image comparison and image stitching. 8

[31] Jiawei Liu, Qiang Wang, Huijie Fan, Wentao Li, Liangqiong Qu, and Yandong Tang. A decoupled multi-task network for shadow removal. *IEEE TMM*, 2023. 3, 7, 19, 20

[32] Jiawei Liu, Qiang Wang, Huijie Fan, Jiandong Tian, and Yandong Tang. A shadow imaging bilinear model and three-branch residual network for shadow removal. *IEEE TNNLS*, pages 1–15, 2023. 2, 3

[33] Risheng Liu, Long Ma, Jiaao Zhang, Xin Fan, and Zhongxuan Luo. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *Proc. CVPR*, pages 10561–10570, 2021. 21

[34] Xing Liu, Masanori Suganuma, Zhun Sun, and Takayuki Okatani. Dual residual networks leveraging the potential of paired operations for image restoration. In *Proc. CVPR*, pages 7007–7016, 2019. 7

[35] Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *Proc. ICLR*, 2023. 5, 16

[36] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proc. ICCV*, 2015. 5, 6, 7, 8, 15

[37] Zhihao Liu, Hui Yin, Yang Mi, Mengyang Pu, and Song Wang. Shadow removal by a lightness-guided network with training on unpaired data. *IEEE TIP*, 30:1853–1865, 2021. 19, 20

[38] Gunter Loffler, Grigori Yourganov, Frances Wilkinson, and Hugh R Wilson. fmri evidence for the neural representation of faces. *Nature neuroscience*, 8(10):1386–1391, 2005. 26

[39] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *Proc. CVPR*, pages 11461–11471, 2022. 1, 2

[40] Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. Image restoration with mean-reverting stochastic differential equations. In *Proc. ICML*, 2023. 4

[41] Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. Sdedit: Guided image synthesis and editing with stochastic differential equations. In *Proc. ICLR*, 2021. 2

[42] Qiang Meng, Shichao Zhao, Zhida Huang, and Feng Zhou. Magface: A universal representation for face recognition and quality assessment. In *Proc. CVPR*, pages 14225–14234, 2021. 26

[43] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proc. CVPR*, pages 3883–3891, 2017. 22, 23, 28

[44] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *Proc. ICML*, pages 8162–8171, 2021. 2, 5, 6, 16

[45] Rui Qian, Robby T Tan, Wenhan Yang, Jiajun Su, and Jiaying Liu. Attentive generative adversarial network for raindrop removal from a single image. In *Proc. CVPR*, pages 2482–2491, 2018. 5, 7, 17, 22, 23, 28

[46] Yuhui Quan, Shijie Deng, Yixin Chen, and Hui Ji. Deep learning for seeing through window with raindrops. In *Proc. ICCV*, pages 2463–2471, 2019. 7

[47] K. F. Riley, M. P. Hobson, and S. J. Bence. *Mathematical Methods for Physics and Engineering: A Comprehensive Guide*. Cambridge University Press, 3 edition, 2006. 2, 6

[48] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proc. CVPR*, pages 10684–10695, 2022. 1, 2, 5, 6

[49] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE TPAMI*, 45(4): 4713–4726, 2022. 1, 2, 16, 27, 28

[50] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *Proc. ICML*, pages 2256–2265, 2015. 2

[51] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *Proc. ICLR*, 2021. 1, 2, 5, 6, 7, 8, 13, 14, 15, 16, 17, 24

[52] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. 32, 2019. 2, 5

[53] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *Proc. ICLR*, 2021. 2, 15, 16

[54] Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. In *Proc. ICML*, 2023. 29

[55] Maitreya Suin, Kuldeep Purohit, and AN Rajagopalan. Spatially-attentive patch-hierarchical network for adaptive motion deblurring. In *Proc. CVPR*, pages 3606–3615, 2020. 23

[56] Zhengzhong Tu, Hossein Talebi, Han Zhang, Feng Yang, Peyman Milanfar, Alan Bovik, and Yinxiao Li. Maxim: Multi-axis mlp for image processing. In *Proc. CVPR*, pages 5769–5780, 2022. 2

[57] Jifeng Wang, Xiang Li, and Jian Yang. Stacked Conditional Generative Adversarial Networks for Jointly Learning Shadow Detection and Shadow Removal. In *Proc. CVPR*, pages 1788–1797, 2018. 5, 7, 8, 17, 18, 19, 20, 26, 27, 28

[58] Tao Wang, Kaihao Zhang, Tianrun Shen, Wenhan Luo, Bjorn Stenger, and Tong Lu. Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. In *Proc. AAAI*, pages 2654–2662, 2023. 7, 27

[59] Yufei Wang, Renjie Wan, Wenhan Yang, Haoliang Li, Lap-Pui Chau, and Alex Kot. Low-light image enhancement with normalizing flow. In *Proc. AAAI*, pages 2604–2612, 2022. 7, 27

[60] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proc. CVPR*, pages 17683–17693, 2022. 23, 24, 28

[61] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *Proc. BMVC*, 2018. 4, 5, 7, 8, 20, 21, 28

[62] Jay Whang, Mauricio Delbracio, Hossein Talebi, Chitwan Saharia, Alexandros G Dimakis, and Peyman Milanfar. Deblurring via stochastic refinement. In *Proc. CVPR*, pages 16293–16303, 2022. 2, 16, 23

[63] Hugh R Wilson, Gunter Loffler, and Frances Wilkinson. Synthetic faces, face cubes, and the geometry of face space. *Vision research*, 42(27):2909–2923, 2002. 26

[64] Jie Xiao, Xueyang Fu, Aiping Liu, Feng Wu, and Zheng-Jun Zha. Image de-raining transformer. *IEEE TPAMI*, 2022. 7

[65] Ke Xu, Xin Yang, Baocai Yin, and Rynson WH Lau. Learning to restore low-light images via decomposition-and-enhancement. In *Proc. CVPR*, pages 2281–2290, 2020. 17, 21, 23, 28

[66] Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. Snr-aware low-light image enhancement. In *Proc. CVPR*, pages 17714–17724, 2022. 7, 21, 22

[67] Zongsheng Yue, Jianyi Wang, and Chen Change Loy. Resshift: Efficient diffusion model for image super-resolution by residual shifting. In *Proc. NeurIPS*, 2023. 4

[68] Wu Yuhui, Pan Chen, Wang Guoqing, Yang Yang, Wei Jiwei, Li Chongyi, and Heng Tao Shen. Learning semantic-aware knowledge guidance for low-light image enhancement. In *Proc. CVPR*, 2023. 7, 21

[69] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for real image restoration and enhancement. In *Proc. ECCV*, pages 492–511, 2020. 21

[70] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proc. CVPR*, pages 14821–14831, 2021. 2, 7, 17, 23

[71] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for fast image restoration and enhancement. *IEEE TPAMI*, 45(2):1934–1948, 2022. 21

[72] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE TIP*, 26(7):3142–3155, 2017. 2

[73] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proc. CVPR*, pages 586–595, 2018. 17, 21

[74] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *Proc. ACMMM*, pages 1632–1640, 2019. 7, 17

[75] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image restoration. *IEEE TPAMI*, 43(7):2480–2495, 2020. 2

[76] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond brightening low-light images. *IJCV*, 129: 1013–1037, 2021. 7, 21

[77] Zhao Zhang, Huan Zheng, Richang Hong, Mingliang Xu, Shuicheng Yan, and Meng Wang. Deep color consistent network for low-light image enhancement. In *Proc. CVPR*, pages 1899–1908, 2022. 7

[78] Chuanjun Zheng, Daming Shi, and Wentian Shi. Adaptive unfolding total variation network for low-light image enhancement. In *Proc. ICCV*, pages 4439–4448, 2021. 21

[79] Yurui Zhu, Jie Huang, Xueyang Fu, Feng Zhao, Qibin Sun, and Zheng-Jun Zha. Bijective mapping network for shadow removal. In *Proc. CVPR*, pages 5627–5636, 2022. 7, 17, 19, 20

[80] Yurui Zhu, Zeyu Xiao, Yanchi Fang, Xueyang Fu, Zhiwei Xiong, and Zheng-Jun Zha. Efficient model-driven network for shadow removal. In *Proc. AAAI*, 2022. 7, 17, 19, 21

[81] Zhengxia Zou, Sen Lei, Tianyang Shi, Zhenwei Shi, and Jieping Ye. Deep adversarial decomposition: A unified framework for separating superimposed images. In *Proc. CVPR*, pages 12806–12816, 2020. 2

[82] Ozan Özdenizci and Robert Legenstein. Restoring vision in adverse weather conditions with patch-based denoising diffusion models. *IEEE TPAMI*, pages 1–12, 2023. 1, 2, 7, 16, 23