# Noisy-Residual Continuous Diffusion Models for Real Image Denoising

1st Xuan Long
*Institute of Information Science, Beijing Jiaotong University*
Beijing, China
xuanlong@bjtu.edu.cn

2nd Meiqin Liu*
*Institute of Information Science, Beijing Jiaotong University*
Beijing, China
mqliu@bjtu.edu.cn

3rd Qi Tang
*Institute of Information Science, Beijing Jiaotong University*
Beijing, China
qitang@bjtu.edu.cn

4th Chao Yao
*School of Computer and Communication Engineering University of Science and Technology Beijing*
Beijing, China
yaochao@ustb.edu.cn

5th Jian Jin
*Alibaba-NTU Singapore Joint Research Institute Nanyang Technological University*
Singapore
jian.jin@ntu.edu.sg

6th Yao Zhao
*Institute of Information Science, Beijing Jiaotong University*
Beijing, China
yzhao@bjtu.edu.cn

***Abstract*—The generation paradigm of diffusion model (DM) inspires numerous works to approach the image denoising problem iteratively. However, DM-based image denoising methods typically require long serial sampling chains, resulting in substantial sampling time and computation. To address this issue, we propose a Noisy-Residual Continuous Diffusion Model (RCDM). It constructs a path between clean and noisy images by shifting their noisy residual during forward process, which significantly shortens diffusion distance. To approximate the path, Noisy Residual Tracer Network (NRTNet) is adopted to estimate the derivative of each point along the path. For further acceleration, clean images are iteratively sampled from noisy images in the reverse process, where the sampling intervals are learnable and skippable. Moreover, we devise a two-stage training strategy to minimize the curvature of the learned path. Experimental results demonstrate that the proposed method achieves superior performance with fewer sampling steps in real image denoising.**

***Index Terms*—image denoising, diffusion model, sampling acceleration**

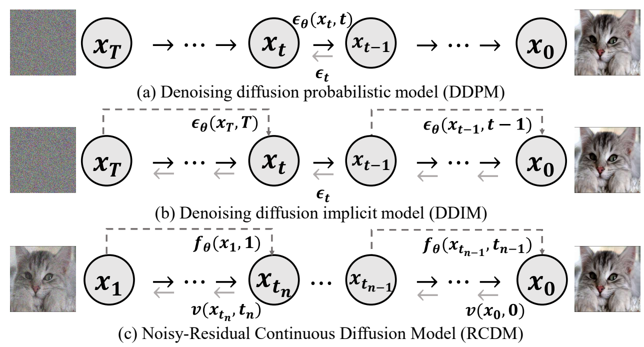


Fig. 1: (a) DDPM defines discrete Markov chain between clean images and Gaussian noises. (b) DDIM designs discrete non-Markov chain in the reverse process between the clean images and the Gaussian noises. (c) RCDM constructs continuous forward process and skippable reverse process between the clean image and the corresponding noisy.

## I. Introduction

Image denoising aims to recover clean images from distorted ones corrupted by different types of noise, which is ill-posed and a challenging task. Image denoising holds practical significance in various real-life scenarios, spanning from medical imaging to surveillance systems, which is attributed to the enhancement of image representation and the improved accuracy of visual analysis [1]. Consequently, image denoising attracts considerable attention due to its diverse applications and the demand for high-quality images.

The deep learning methods based on convolutional neural networks (CNNs) [2]–[4] have been the prevailing streaming in the field of image denoising. These approaches leverage the powerful representation learning capabilities of CNNs, allowing them to capture complex patterns and structures. Nevertheless, the limited receptive field prevents the network from extracting global information, which is crucial in image denoising. To address this limitation, various researchers make efforts to broaden the receptive field, employing techniques such as pooling [5] and deformable convolution [6], etc. Furthermore, the convolutional kernel, utilizing a mechanism of parameter sharing, lacks the adaptability to fuse spatial features in an optimal manner. Therefore, some studies also introduce kinds of attention mechanisms [7], [8] to focus on important regions and learnable kernel bases to model different local structures [9].

Compared to CNNs, Transformer [10], [11] possesses the capability to capture global information via the self-attention (SA) mechanism, rendering them successful in various visual tasks, including image denoising. However, the advantages in

* Corresponding author.
This work is supported in part by the National Key Research and Development Program of China under Grant 2022ZD0118001; and in part by the National Natural Science Foundation of China under Grant 62120106009, Grant 62372036, and Grant 62332017.

long-range dependency modeling brought by SA come at the cost of increased computational overhead. Hence, it's desirable to explore a trade-off between performance and computational efficiency. Some solutions, such as sparse attention [12], window-based attention [13], and applying SA across the channel dimension [14], etc., are proposed to reduce the computational complexity and enhance the feasibility of Transformer in image denoising. Additionally, amount of methods combine CNNs and Transformer to aggregate local and global features, resulting in significant improvements [15], [16].

Building on the advancements achieved by CNNs and Transformers in the realm of image denoising, another noteworthy approach that recently captured significant attention is diffusion models (DMs). The inherent ability of DMs to iteratively generate high-quality and realistic samples from Gaussian noises presents a distinctive path for image restoration [17]. It consists of two processes, i.e., the forward process and the reverse process. In the reverse process of most DMs, Gaussian noises gradually undergo denoising to sample the photo-realistic images. Given that the objective of image denoising is to recover clean images from noisy ones, it coincides with the reverse process. Consequently, it's natural and rational to apply the diffusion model to image denoising.

Despite the considerable success of DMs in image generation, there exists a necessity for several modifications to tailor DMs to image denoising better. As illustrated in Fig.1, we propose a noisy-residual continuous diffusion model (RCDM) based on Ordinary Differential Equation (ODE) rather than Stochastic Differential Equations (SDE). The intuition behind this is to eliminate randomness, establishing a straightforward and deterministic path between the clean and noisy images. Meanwhile, the continuous process enables the sampling time intervals to be learnable and skippable, thus accelerating the sampling process. Specifically, the forward process initiates from the clean images and incrementally shifts to the noisy images by incorporating noisy residuals rather than Gaussian noises. In the process of sampling a clean image, we employ the Noisy Residual Tracer Network (NRTNet) to predict the derivative, differing from the noise or clean image typically predicted by previous DMs (such as DDPM [18] and DDIM [19] in Fig.1). Furthermore, we devise a two-stage training strategy comprising the prediction training stage and the fine training stage. Among them, the former guides the network to approximate the deterministic denoising path at a coarse granularity, while the latter traverses the entire generative process to refine the sampled image directly, minimizing the curvature of the learned path. Experimental results show that our method achieves superior performance with fewer sampling time steps. The main contributions are listed as follows:

- We devise a noisy-residual continuous diffusion model (RCDM) for image denoising. It consists of a forward process that shifts in the noisy residual space and a non-Markov reverse process, achieving efficient denoising.
- We design a learnable continuous sampling schema, enabling adaptive removal of noise. It eliminates most noise in the early steps and concentrates on the hard-to-remove noise in the later sampling steps.
- We propose a two-stage training strategy involving the derivative prediction and the rectification of the cumulative error to make the curved path to be straighter, gaining the superior denoising performance.

## II. METHODOLOGY

In this section, we present a noisy-residual continuous diffusion model (RCDM) tailored for image denoising. We will provide a detailed description of the proposed RCDM, as well as the training and sampling strategies employed.

### A. Overview

RCDM consists of a forward process and a reverse process, along with a transformer-based denoising network named Noisy Residual Tracer Network (NRTNet). In the forward process, the noisy residual between clean image $x_0$ and noisy image $x_1$ is incrementally added to the clean one, yielding intermediate noisy images $x_t$. It serves as the input of NRTNet, as shown in Fig. 2. To be specific, NRTNet first utilizes a convolution with the kernel size of $3 \times 3$ to extract shallow features $F_s \in \mathbb{R}^{H \times W \times C}$, where $H \times W$ denotes the spatial dimension and $C$ is the number of channels. Subsequently, $F_s$ is hierarchically refined by walking past the stacked noisy residual differentiators (NRDs), arranged akin U-shape. NRD is designed to capture long-range correlations within the input $F_{in}$ along the channel dimension and generate output $F_{out}$. The time embedding module is incorporated into NRD, aiming to encode time step $t$ to provide the prior of noise intensity. Besides, the skip connection bridges the features at the same level in the encoder and decoder to assist the recovery. Finally, fine details are further enriched by a $3\times3$ convolution layer for prediction of the derivative $f_\theta$ located at $x_t$. In the reverse process, the residual-related noises are iteratively removed from the noisy image with specified time intervals and predicted derivative by NRTNet, gaining denoised samples.

### B. Forward and Reverse Process

**Forward process**. Residual-related noise is gradually introduced into the clean image, finally obtaining the noisy counterpart. In the noisy-residual continuous diffusion model, the intermediate noisy image $x_t$ indexed by a continuous time $t$ [21] is defined as follows:

$$x_t = x_{t-dt} + v(x_{t-dt}, t-dt)dt, t \in (0, 1) \tag{1}$$

where $v(x_{t-dt}, t-dt) = x_1 - x_0 = \frac{x_{t-dt}-x_0}{t-dt}$ represents the speed of $x_{t-dt}$ as it travels to $x_t$, possessing the same value as the residual. Mathematically, it can be viewed as the derivative at $x_{t-dt}$. It is worth mentioning that the forward process terminates with the corresponding noisy image rather than a pure Gaussian noise, which is unnecessary for image denoising. The transition distribution is formulated based on the derivative as follows:

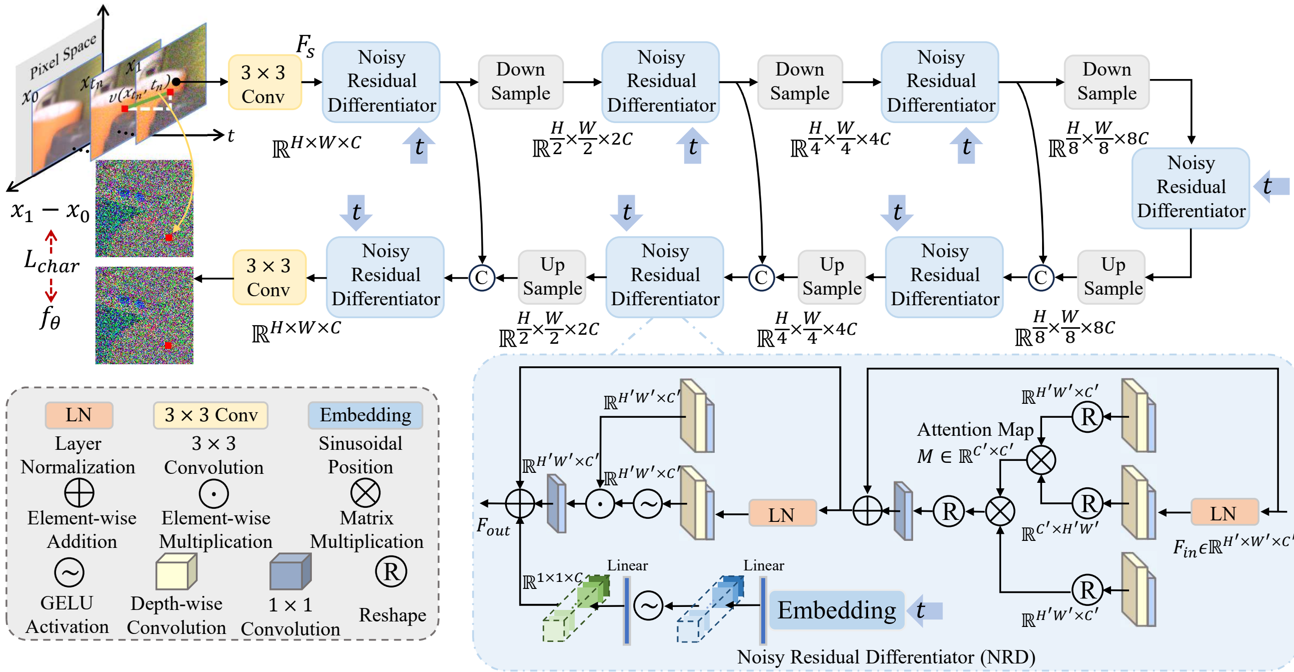


Fig. 2: The architecture of the Noisy Residual Tracer Network (NRTNet) is composed of stacked noise residual differentiators (NRD) akin U-shape [20]. The time embedding module is incorporated into the NRD block to provide prior of noise intensity. The whole network is optimized by predicting the residual-related derivative to approximate the straightforward path from $x_1$ to $x_0$.

$$q(x_t|x_{t-dt}) = \mathcal{N}(x_t; x_{t-dt} + v(x_{t-dt}, t-dt)dt, \beta(t)\boldsymbol{I}), \quad t \in [0,1] \tag{2}$$

where $\mathcal{N}(\cdot)$ denotes the Gaussian distribution. To establish a straightforward and deterministic path from $x_{t-dt}$ to $x_t$, the variance of the random noise in $q(x_t|x_{t-dt})$ is set to 0, i.e., $\beta(t) = 0$. The marginal distribution at arbitrary time $t$ is formulated as:

$$q(x_t|x_0) = \mathcal{N}(x_t; x_0 + \int_0^t v(x_t, t)dt, \int_0^t \beta(t)dt\boldsymbol{I}), \quad t \in [0,1] \tag{3}$$

where the coefficient of random noise $\int_0^t \beta(t)dt = 0$ in the proposed method. By constructing such forward process, we establish a shorter path between $x_0$ and $x_1$ compared to previous diffusion based methods [18], [19], thereby effectively decreasing the cost of diffusing and sampling. Given the definite forward diffusion process of noisy residual, the noise in distorted image can be iteratively removed in reverse sampling.

**Reverse process.** We define a learnable and continuous generative process to estimate the posterior distribution $p_\theta(x_{0:1})$. The inverse transition kernel, represented by $p_\theta(x_{t-dt}|x_t)$, leverages knowledge from $q(x_{t-dt}|x_t, x_0, x_1)$. Intuitively, given a noisy image observation $x_t$, together with the derivative (or residual) at each step and time interval, RCDM can sample high-quality denoised images iteratively. The target distribution $q(x_{t-dt}|x_t, x_0, x_1)$ can be formally represented in terms of the conditional probability of $x_t$:

$$q(x_{t-dt}|x_t, x_0, x_1) = \mathcal{N}(x_{t-dt}; x_t - v(x_t, t)dt, \beta(t)\boldsymbol{I}), \quad t \in [0,1] \tag{4}$$

where $\beta(t) = 0$. Nevertheless, the value of $v(x_t, t)$ is unavailable during sampling. Therefore, the predicted derivative $f_\theta(x_t, t)$ by NRTNet is opted as a rationable alternative. It differs from most existing DMs where the neural network predicts noise or clean images at each time step.

Same as DDIM [19], the reverse process of RCDM is defined based on non-Markov chain to further accelerate sampling, and the predicted target distribution is as follows:

$$p_\theta(x_{t-n\times dt}|x_t, x_1) = \mathcal{N}(x_{t-dt}; x_t - nf_\theta(x_t, t)dt, \widetilde{\beta}(t)\boldsymbol{I}), \quad t \in [0,1], n = 1, 2, 3, ... \tag{5}$$

where $f_\theta(\cdot)$ denotes the predicted derivative of NRTNet, $\widetilde{\beta}(t) = 0$. The parameters of model $\theta$ are optimized by :

$$\mathcal{J}(f_\theta) = \mathbb{E}_{x_{0:1} \sim p_\theta(x_{0:1})} L_{char}(v(x_t, t), f_\theta(x_t, t)) \tag{6}$$

where $L_{char}(\cdot)$ denotes the Charbonnier loss.

### C. Sampling and Training Strategies

**Sampling schema.** To accelerate the sampling process, we devise a non-Markov and time-step-learnable sampling schema.

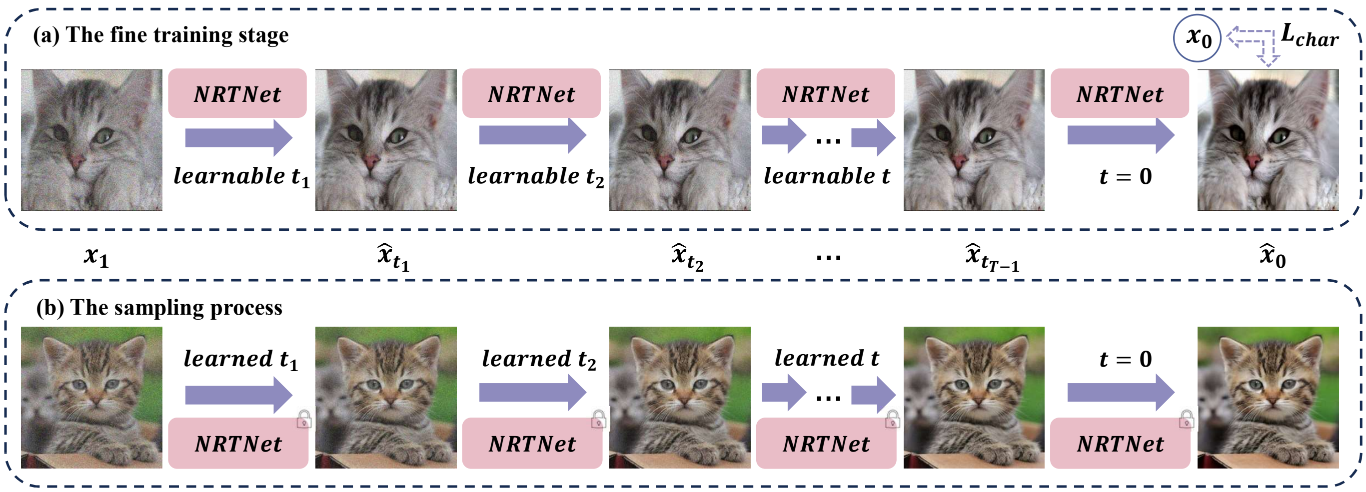


Fig. 3: (a) **The fine training stage**. The parameters of NRTNet are further optimized to mitigate the cumulative error and minimize the curvature of the learned path by traversing the entire sampling process and directly refining the sampled clean image. (b) **The sampling process**. The choice of time step during sampling process is adaptive to noise intensity in the image.

---

**Algorithm 1** The Prediction Training Stage

---

**Input:** Noisy image $x_1$, original clean image $x_0$, noisy residual tracer $f_\theta$.

**repeat**
- $t \sim U(0,1)$
- $v(x_t, t) = x_1 - x_0 = \frac{x_t - x_0}{t}$
- $x_t = x_0 + t \times v(x_t, t)$
- $\nabla_\theta L_{char}(v(x_t, t), f_\theta(x_t, t))$

**until** converged

---

Specifically, the sampling process is defined based on a non-Markov chain, allowing for the selection of arbitrary sampling step. Typically, only $3-4$ steps are sufficient to yield photo-realistic samples. Moreover, the continuity of RCDM is determined by a continuous parameter $t$, enabling the sampling steps to be learnable and skippable for more effective and efficient noise removal, which will be explained in detail in the following training strategy.

**Two-stage training strategy.** The path between clean images $x_0$ and noisy images $x_1$ is assumed to be deterministic and straightforward. Charbonnier loss is employed to constrain the training of NRTNet. However, following the initial training of NRTNet, the unsatisfactory estimation of derivative results in cumulative error during sampling process. To address this issue, we propose a two-stage training strategy. In the prediction training stage, NRTNet is optimized to approximate the deterministic denoising path by predicting the derivative, as presented in Algorithm 1. And the second one, named as fine training stage, aims to rectify the cumulative error and minimize the curvature of the learned path. Simultaneously, it focuses on learning the sampling time step, enabling an adaptive sampling process. In this phase, RCDM directly refines the final output of the sampling process under the supervision of the ground truth $x_0$ to further optimize the NRTNet, as described in Algorithm 2.

---

**Algorithm 2** The Fine Training Stage

---

**Input:** Noisy image $x_1$, pretrained noisy residual tracer $f_\theta$, time steps $T$.

$t_0 = 0, t_T = 1$

**repeat**
- **for** $n = T, ..., 1$ **do**
  - $t_{n-1}$ is a learnable time variable, $t_{n-1} \in [0,1]$ **if** $n > 1$ **else** $0$
  - $\hat{x}_{t_{n-1}} = \hat{x}_{t_n} - (t_n - t_{n-1}) \times f_\theta(\hat{x}_{t_n}, t_n)$
- **end for**
- $\nabla_\theta L_{char}(x_0, \hat{x}_0)$

**until** converged

---

It is noteworthy that the time variable $t$ follows a uniform distribution over $[0, 1]$, which controls the intensity of noise added to the clean image. In contrast to most of the previous discrete diffusion models, NRTNet is trained to predict the derivatives of intermediate images $x_t$ at arbitrary continuous time step. What's more, in the fine training stage, RCDM is capable to dynamically determine the skipped time intervals based on the specific noise intensity and characteristics of the image. Ideally, the larger the value of $t$, the more the network is expected to efficiently remove the majority of noise, thus skipping long time intervals and saving more computational cost. When $t$ tends to be small, it will force the network to concentrate on removing the imperceptible and tiny noise.

## III. Performance Evaluation

### A. Experiment Settings

The proposed noisy-residual continuous diffusion model (RCDM) is implemented using PyTorch. During the prediction training stage, the cosine annealing is employed to steadily decrease the learning rate from $3 \times 10^{-4}$ to $1 \times 10^{-6}$. Subsequently, NRTNet undergoes further training in the fine training stage with a learning rate ranging from $1 \times 10^{-6}$ to $1 \times 10^{-7}$.

TABLE I: Quantitative results on real image denoising. The best performance is displayed in **bold**.

| Dataset | Method | Convolutional Neural Network | | | | | | | | Transformer | | | Diffusion Model | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | DnCNN [2] | RIDNet [3] | VDN [22] | DANet+ [23] | CyscleISP [24] | MIRNet [25] | MPRNet [4] | NBNet [26] | Uformer [13] | SwinIR [27] | Restormer [14] | C2F-DFT [28] | NRTNet (ours) |
| ***SIDD*** | PSNR ↑ | 23.66 | 38.71 | 39.28 | 39.47 | 39.52 | 39.72 | 39.71 | 39.75 | 39.77 | 39.77 | **40.02** | 39.84 | <u>39.89</u> |
| | SSIM ↑ | 0.583 | 0.951 | 0.956 | 0.957 | 0.957 | <u>0.959</u> | 0.958 | <u>0.959</u> | <u>0.959</u> | 0.958 | **0.960** | **0.960** | **0.960** |
| ***DND*** | PSNR ↑ | 32.43 | 39.26 | 39.38 | 39.58 | 39.56 | 39.88 | 39.80 | 39.89 | 39.96 | 40.01 | <u>40.03</u> | 39.95 | **40.10** |
| | SSIM ↑ | 0.790 | 0.953 | 0.952 | 0.955 | <u>0.956</u> | <u>0.956</u> | 0.954 | 0.955 | <u>0.956</u> | 0.958 | <u>0.956</u> | 0.955 | **0.957** |

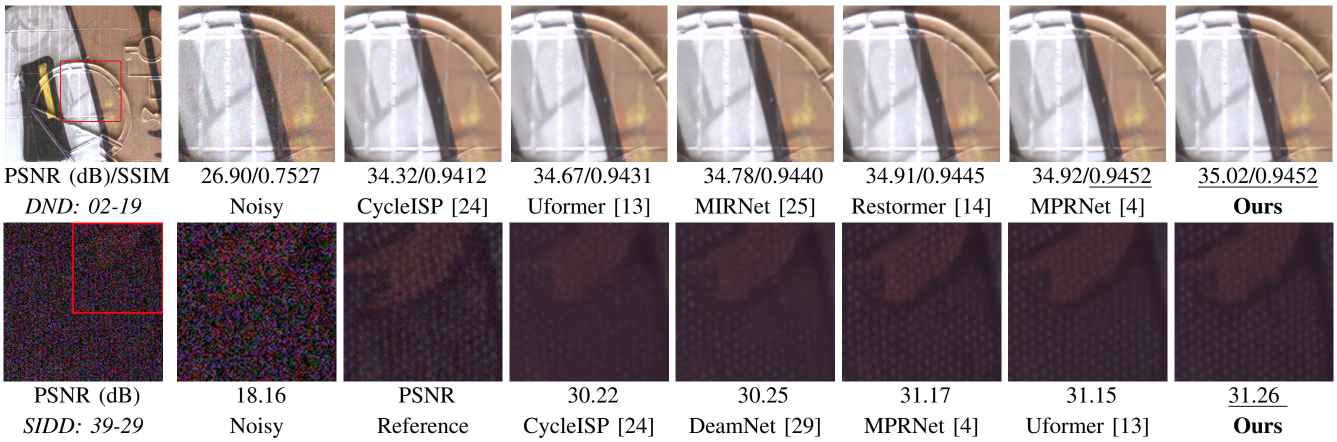


Fig. 4: Visual comparisons of real image denoising. **Top**: Visual results on *DND* testset. **Bottom**: Visual results on *SIDD* testset.

The (patch size, batch size) is maintained at $(128^2, 16)$ through the entire training process. The two-stage training is conducted on 2 NVIDIA 3090 GPUs, while sampling is performed on 1 NVIDIA 3090 GPU.

### B. Quantitative and Qualitative Evaluation

For real image denoising, RCDM is trained on the *SIDD* dataset [30] and evaluated on the *SIDD* validation set and *DND* test set [31]. In comparison to some image denoising methods based on CNN, Transformer and Diffusion Model, our approach demonstrates superior performance. As shown in Table I, NRTNet with RCDM surpasses all CNN-based methods and most of the Transformer-based algorithms on *SIDD* validation set. Notably, the proposed method achieves a PSNR gain of 0.05 dB over the DM-based denoising method C2F-DFT [28] on the *SIDD* validation set. On the *DND* test set, NRTNet also attains the highest values for PSNR and SSIM compared to other methods. For visual results, the most faithful and realistic denoising results are evident in both two test sets, as illustrated in Fig. 4.

Furthermore, we conduct experiments on the sampling time of NRTNet with different sampling schemes. As shown in Table III, the proposed RCDM has a similar sampling time as DDIM, while reducing substantial amount of time compared to DDPM on single image and the whole *SIDD* validation set, respectively. Moreover, RCDM achieves the best results in both PSNR and SSIM.

### C. Ablation Study

**Effect of two-stage training scheme.** We analyze the improvements on the proposed two-stage training scheme. In Table II, the denoising results on *DND* test set are significantly improved by adding the fine training stage following the prediction training stage. Fig. 5 shows the comparisons on *SIDD* validation set, it can be observed that promising results are yielded after the fine training stage, which rectifies the cumulative error.

TABLE II: Effect of the two-stage training strategy.

| Training Scheme | PSNR (dB) | SSIM |
|---|---|---|
| (a) Prediction training (3 sampling steps) | 39.63 | 0.950 |
| (b) Prediction training (1 sampling step) | 39.97 | 0.955 |
| (c) Fine training (3 sampling steps) | **40.07** | **0.956** |

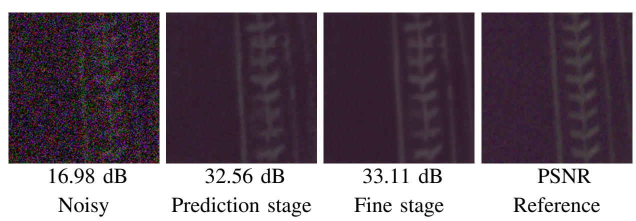


Fig. 5: Visual results of two training stages.

**Effect of learnable time intervals.** We further investigate the impact of the learnable time intervals in the fine training stage, and show the results in Table III. Learnable $t$ can improve the denoising quality by 0.58 dB gain of PSNR on *SIDD* test set.
**Effect of sampling steps.** We study the impact of timesteps in the fine training stage. Specifically, we change the number of timesteps from 1 to 5. The results of PSNR and the perception measurement LPIPS on *SIDD* test set are shown in Fig. 6. The consumption of computing resources increases linearly as the steps grow. Considering the trade-off between the sampling steps of the diffusion reverse process and computational cost, we adopt 4 sampling steps in practice.

TABLE III: Sampling time of different methods. *-T: T steps

| Sampling Scheme | PSNR | SSIM | Time (s) |
|---|---|---|---|
| DDPM-50 | 38.28 | 0.945 | 3.48/4451.12 |
| DDIM-3 | 39.30 | 0.954 | 0.24 / 309.58 |
| RCDM-3 | **39.88** | **0.956** | **0.23 / 298.75** |

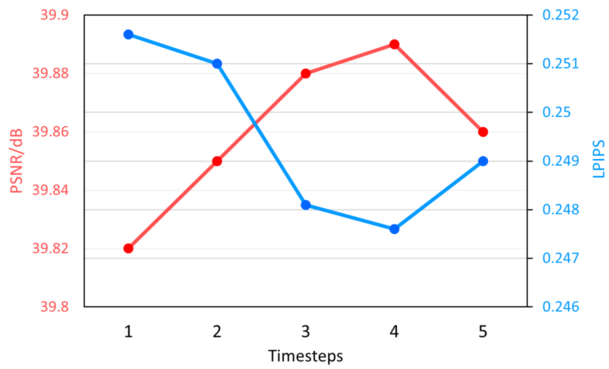


Fig. 6: Effect of time steps in the fine training stage.

## IV. Conclusions

In this paper, we propose a noisy-residual continuous diffusion model (RCDM) for image denoising. It defines a forward process that establishes a straightforward and deterministic path shifted towards noisy images by incrementally incorporating noisy residuals into clean images. And the Noisy Residual Tracer Network (NRTNet) is adopted to predict the derivative (or residual) of each point along the path. In this way, the clean images can be iteratively sampled from the noisy images in the reverse process, where the sampling intervals are learnable and skippable. Furthermore, we devise a two-stage training strategy to minimize the curvature of the learned path in order to denoise more effectively. Under configuration of the shorter diffusing process and non-Markov chain-based sampling process, it gains superior performance with fewer sampling time steps.

## References

[1] Jian Jin, Xingxing Zhang, Xin Fu, Huan Zhang, Weisi Lin, Jian Lou, and Yao Zhao, "Just noticeable difference for deep machine vision," *TCSVT*, vol. 32, no. 6, pp. 3452–3461, 2021.

[2] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang, "Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising," *TIP*, vol. 26, no. 7, pp. 3142–3155, 2017.

[3] Saeed Anwar and Nick Barnes, "Real image denoising with feature attention," in *CVPR*, 2019.

[4] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao, "Multi-stage progressive image restoration," in *CVPR*, 2021.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *TPAMI*, vol. 37, no. 9, pp. 1904–1916, 2015.

[6] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei, "Deformable convolutional networks," in *ICCV*, 2017.

[7] Meiqin Liu, Shuo Jin, Chao Yao, Chunyu Lin, and Yao Zhao, "Temporal consistency learning of inter-frames for video super-resolution," *TCSVT*, vol. 33, no. 4, pp. 1507–1520, 2022.

[8] Hongguang Zhu, Yunchao Wei, Xiaodan Liang, Chunjie Zhang, and Yao Zhao, "CTP: Towards vision-language continual pretraining via compatible momentum contrast and topology preservation," in *ICCV*, 2023.

[9] Yi Zhang, Dasong Li, Xiaoyu Shi, Dailan He, Kangning Song, Xiaogang Wang, Hongwei Qin, and Hongsheng Li, "KBNet: Kernel basis network for image restoration," *arXiv:2303.02881*, 2023.

[10] Chao Yao, Shuaiyong Zhang, Mengyao Yang, Meiqin Liu, and Junpeng Qi, "Depth super-resolution by texture-depth transformer," in *ICME*, 2021.

[11] Qi Tang, Yao Zhao, Meiqin Liu, Jian Jin, and Chao Yao, "Semantic lens: Instance-centric semantic alignment for video super-resolution," in *AAAI*, 2024.

[12] Xiang Chen, Hao Li, Mingqiang Li, and Jinshan Pan, "Learning a sparse transformer network for effective image deraining," in *CVPR*, 2023.

[13] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li, "Uformer: A general u-shaped transformer for image restoration," in *CVPR*, 2022.

[14] 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 *CVPR*, 2022.

[15] Mo Zhao, Gang Cao, Xianglin Huang, and Lifang Yang, "Hybrid transformer-CNN for real image denoising," *ISPL*, vol. 29, pp. 1252–1256, 2022.

[16] CM Fan, TJ Liu, and KH Liu, "SUNet: Swin transformer UNet for image denoising.," *arXiv:2202.14009*, 2022.

[17] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi, "Image super-resolution via iterative refinement," *TPAMI*, vol. 45, no. 4, pp. 4713–4726, 2022.

[18] Jonathan Ho, Ajay Jain, and Pieter Abbeel, "Denoising diffusion probabilistic models," in *NeurIPS*, 2020.

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

[20] Meiqin Liu, Chenming Xu, Chao Yao, Chunyu Lin, and Yao Zhao, "JNMR: Joint non-Linear motion regression for video frame interpolation," *TIP*, vol. 32, pp. 5283–5295, 2023.

[21] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole, "Score-based generative modeling through stochastic differential equations," *arXiv:2011.13456*, 2020.

[22] Zongsheng Yue, Hongwei Yong, Qian Zhao, Deyu Meng, and Lei Zhang, "Variational denoising network: Toward blind noise modeling and removal," in *NeurIPS*, 2019.

[23] Zongsheng Yue, Qian Zhao, Lei Zhang, and Deyu Meng, "Dual adversarial network: Toward real-world noise removal and noise generation," in *ECCV*, 2020.

[24] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao, "CycleISP: Real image restoration via improved data synthesis," in *CVPR*, 2020.

[25] 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 *ECCV*, 2020.

[26] Shen Cheng, Yuzhi Wang, Haibin Huang, Donghao Liu, Haoqiang Fan, and Shuaicheng Liu, "NBNet: Noise basis learning for image denoising with subspace projection," in *CVPR*, 2021.

[27] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte, "SwinIR: Image restoration using swin transformer," in *ICCV*, 2021.

[28] Liyan Wang, Qinyu Yang, Cong Wang, Wei Wang, Jinshan Pan, and Zhixun Su, "Learning a coarse-to-fine diffusion transformer for image restoration," *arXiv:2308.08730*, 2023.

[29] Chao Ren, Xiaohai He, Chuncheng Wang, and Zhibo Zhao, "Adaptive consistency prior based deep network for image denoising," in *CVPR*, 2021.

[30] Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown, "A high-quality denoising dataset for smartphone cameras," in *CVPR*, 2018.

[31] Tobias Plotz and Stefan Roth, "Benchmarking denoising algorithms with real photographs," in *CVPR*, 2017.