# CDPMSR: CONDITIONAL DIFFUSION PROBABILISTIC MODELS FOR SINGLE IMAGE SUPER-RESOLUTION

*Axi Niu[1], Kang Zhang[2], Trung X. Pham[2], Jinqiu Sun[1]\*, Yu Zhu[1], In So Kweon[2], Yanning Zhang[1]*

[1]Northwestern Polytechnical University
[2] Korea Advanced Institute of Science and Technology (KAIST)

## ABSTRACT

Diffusion probabilistic models (DPM) have been widely adopted in image-to-image translation to generate high-quality images. Prior attempts at applying the DPM to image super-resolution (SR) have shown that iteratively refining a pure Gaussian noise with a conditional image using a U-Net trained on denoising at various-level noises can help obtain a satisfied high-resolution image for the low-resolution one. To further improve the performance and simplify current DPM-based super-resolution methods, we propose a simple but non-trivial DPM-based super-resolution post-process framework,*i.e.*, cDPMSR. After applying a pre-trained SR model on the to-be-test LR image to provide the conditional input, we adapt the standard DPM to conduct conditional image generation and perform super-resolution through a deterministic iterative denoising process. Our method surpasses prior attempts on both qualitative and quantitative results and can generate more photo-realistic counterparts for the low-resolution images with various benchmark datasets including Set5, Set14, Urban100, BSD100, and Manga109. *Code will be published after accepted.*

***Index Terms***— Diffusion Probabilistic Models, Image-to-Image Translation, Conditional Image Generation, Image Super-resolution.

## 1. INTRODUCTION

Over the years, single image super-resolution (SISR) has drawn active attention due to its wide applications in computer vision, such as object recognition, remote sensing, and so on. SISR aims to obtain a high-resolution (HR) image containing great details and textures from a low-resolution (LR) image by an SR method, which is a classic ill-posed inverse problem [1]. To establish the mapping between HR and LR images, various CNN-based methods had been proposed. Among them, methods based on the deep generative model have become one of the mainstream, mainly including

**Fig. 1**. Illustration of our method. The model contains a stochastic forward diffusion process which gradually adds noise to an $I^{HR}$ image. And a deterministic denoise process is applied to recover high-resolution and realistic images $I^{SR}$ corresponding to $I^{LR}$ images.

GAN-based [2, 3, 4] and flow-based methods [5, 6, 7], which have shown convincing image generation ability.

GAN-based SISR methods [2, 3, 4] used a generator and a discriminator in an adversarial way to encourage the generator to generate realistic images. Specifically, the generator generates an SR result for the input, and the discriminator is used to distinguish if the generated SR is true. It combines content losses (*e.g.*, $L_1$ or $L_2$) and adversarial losses to optimize the whole training process. Due to their strong learning abilities, GAN-based methods become popular for image SR tasks [4, 8, 9]. However, these methods are easy to meet mode collapse and the training process is hard to converge with complex optimization [10, 11] and adversarial losses often introduce artifacts in SR results, leading to large distortion [12, 13]. Another line of methods based on deep generative models is flow-based methods, which directly account for the ill-posed problem with an invertible encoder [14, 15, 16, 17]. It transforms a Gaussian distribution into an HR image space instead of modeling one single output and inherently resolves the pathology of the original "one-to-many" SR problem. Optimized by a negative loglikelihood loss, these methods avoid training instability but suffer from extremely large footprints and high training costs due to the strong architectural constraints to keep the bijection between latents and data [16].

Lately, the adoptions of diffusion probabilistic models (DPM) have shown promising results in image generative

tasks [18]. Furthermore, the prior attempts [19, 16] at applying the DPM to image SR have also proved their effectiveness for obtaining satisfied SR images. In [19], the authors propose a two-stage SR framework. First, they design an SR structure and pre-train it to obtain a conditional input for the DPM process. Then they redesign the U-net in DPM. The training process of this method is relatively complicated, and it does not consider combining existing pre-trained SR models, such as EDSR [20], RCAN [21], and SwinIR [22]. Similarly, [16] apply the bicubic up-sampled LR image as the conditional input directly. Different from them, our method leverage the development of the current SISR methods to provide more plausible conditional inputs. And we adapt a deterministic sampling way in the inference phase to make the restoration of the SR image better and faster.

Our work is most similar to [19] which is the first to apply DPM to the SR task. In this work, we propose a simple but non-trivial SR post-process framework for image SR based on the conditional diffusion model,$i.e.$, cDPMSR. Unlike the existing techniques, our cDPMSR adopts the pre-trained SR methods to provide the conditional input, which is more plausible than the one used in [16, 19]. And it brings significant improvement in perceptual quality over existing SOTA methods across multiple standard benchmarks. By simply concatenating a Gaussian noise and the conditional input and using MAE loss (*i.e.*, $L_1$ ) to optimize the diffusion model, our method makes the training process more concise compared with [16, 19]. Our contributions are summarized as follows:

- To the best of our knowledge, we are the first to propose an SR post-process framework based on the existing SR models and diffusion probabilistic model.

- Compared to existing SOTA SR methods, our cDPMSR achieves superior perceptive results and can generate more photo-realistic SR results.

- Compared with existing DPM-based SR methods, our cDPMSR adopts a deterministic sampling way in the inference phase, which helps to obtain a better balance between distortion and perceptual quality.

## 2. PROPOSED METHOD

Fig.1 illustrates the whole process of our cDPMSR. The following section introduces the method in detail.

### 2.1. Stocatic Diffusion Process

Given a SISR dataset $(\boldsymbol{I}^{HR}, \boldsymbol{I}^{LR}) \sim D$, we adopt the diffusion probabilistic model (DPM) [18, 23] to map a normal Gaussian noise $\boldsymbol{x}_T \sim \mathcal{N}(0,\mathbf{1})$ to a high-resolution image $I^{HR}$ with a corresponding conditional image $\boldsymbol{I}^C$. We will talk about the choice of the conditional image later. The DPM contains latent variables $\boldsymbol{x} = \{\boldsymbol{x}_t | t = 0, 1, ..., T\}$, where $\boldsymbol{x}_0 = I^{HR}$, $\boldsymbol{x}_T = \mathcal{N}(0,\mathbf{1})$. And it is defined on a noise schedule $\alpha_t$ and $\sigma_t$ such that $\lambda_t = \alpha_t^2/\sigma_t^2$, the signal-to-noise-ratio, is monotonically decreasing with $t$.

**Forward stochastic diffusion process.** We define the forward process $q(\boldsymbol{x}_t | \boldsymbol{I}^{HR}) := q(\boldsymbol{x}_t | \boldsymbol{x}_0)$ of DPM with a Gaussian process by the Markovian structure:

$$q(\boldsymbol{x}_t | \boldsymbol{x}_0) = \mathcal{N}(\boldsymbol{x}_0; \alpha_t \boldsymbol{x}_0, \sigma_t^2 \mathbf{1})$$

$$q(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}) = \mathcal{N}(\boldsymbol{x}_t; (\alpha_t/\alpha_{t-1})\boldsymbol{x}_{t-1}, (1 - \frac{\lambda_t}{\lambda_{t-1}})\sigma_t^2 \mathbf{1}). \tag{1}$$

The forward process gradually adds noise into an image $\boldsymbol{x}_0$ to generate latent variables $\boldsymbol{x}_1, ..., \boldsymbol{x}_T$ for the original image $\boldsymbol{x}_0$. With the Gaussian distribution reparameterization trick, we can write the latent variable $x_t$ as:

$$\boldsymbol{x}_t = \alpha_t \boldsymbol{x}_0 + \sigma_t \boldsymbol{\epsilon}, \boldsymbol{\epsilon} \sim \mathcal{N}(0,\mathbf{1}). \tag{2}$$

Following [18] we set $\alpha_t = \cos 0.5\pi \frac{t}{T}$ and $\sigma_t = \sqrt{1 - \alpha_t^2}$.
**Model training.** The optimization target of DPM is denoising $\boldsymbol{x}_t \sim p(\boldsymbol{x}_t | \boldsymbol{x}_0)$ to get estimated $\hat{\boldsymbol{x}}_0$ with a U-Net $f_\theta(x_t, t, I^C) := \hat{\boldsymbol{x}_0} \approx \boldsymbol{x}_0$. Same with [18], we use the following loss function to train the model:

$$L := \mathbb{E}_{t,(\boldsymbol{x}_0, I^C), \boldsymbol{\epsilon}}[\|\boldsymbol{x}_0 - f_\theta(\alpha_t \boldsymbol{x}_0 + \sigma_t \boldsymbol{\epsilon}, t, I^C)\|], \tag{3}$$

where $t$ is uniformly sampled between 1 and $T$. In [18, 23] the above loss function is justified as optimizing the usual variational bound on negative log-likelihood with discarding the loss weighting. Here, different with [18] we add an additional input $I^C$ as the conditional image to guide the model to keep the same content with $I^C$ during the denoising process. And different from current DPM-based super-resolution methods [16, 19] which reverses the diffusion process by estimating noise, we directly let the model predict the image.

---

**Algorithm 1** cDPMSR

**Training :** train predictor $f_\theta$
Input: Dataset $D$, schedule $\alpha_t, \sigma_t$, timesteps $T$, pretrained super resolution model $\phi_\theta$

1: **repeat**
2:     $(\boldsymbol{I}^{HR}, \boldsymbol{I}^{LR}) \sim D, t \sim \texttt{Uniform}(\{1, ..., T\})$,
    $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
3:     $\boldsymbol{I}^C = \phi_\theta(I^{LR})$
4:     $\boldsymbol{x}_t = \alpha_t \boldsymbol{I}^{HR} + \sigma_t \boldsymbol{\epsilon}$
5:     Take a gradient descent step on
       $\nabla_\theta \| \boldsymbol{I}^{HR} - f_\theta(\boldsymbol{x}_t, t, \boldsymbol{I}^C) \|$
6: **until** converged

**Inference:** super resolve $\boldsymbol{I}^{LR}$
Input: trained predictor $f_\theta$, pretrained super resolution model $\phi_\theta$

1:   $\boldsymbol{I}^C = \phi_\theta(I^{LR})$
2:   $\boldsymbol{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{1})$
3:   **for** $t = T, ..., 1$ **do**
4:     $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\boldsymbol{\epsilon} = \mathbf{0}$
5:     $\boldsymbol{x}_{t-1} = \alpha_{t-1} f_\theta(x_t, t, I^C) + \sigma_{t-1}(\boldsymbol{x}_t - \alpha_t \hat{\boldsymbol{x}_0})/\sigma_t$
6:   **end for**

---

### 2.2. Deterministic Denoise Process

**Trained model sampling.** Different from [19, 16] sampling via a stochastic way, we adopt a deterministic manner to conduct the reverse process $p_\theta(\boldsymbol{x}_{t-1} | \boldsymbol{x}_t)$, which is an implicit

**Fig. 2**. Qualitative comparison with SOTAs performed on 'head' from Set5 ($\times 4$ scale, best view in zoomed-in.)

**Fig. 3**. Qualitative comparison with SOTAs performed on 'img003 ' from Urban100 ($\times 4$ scale. Cropped and zoomed in for a better view.)

probabilistic model [24]. Compared with the sampling strategy used in [19, 16], it can achieve higher quality images with less inference time which is critical for this task. Given the image $\boldsymbol{x}_t$ at step $t$, we can write the generation process of $\boldsymbol{x}_{t-1}$ via a forward posterior $q_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \hat{x}_0)$ as follows:

$$
\begin{aligned}
p_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t) &= q_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \hat{x}_0) \\
&= \mathcal{N}(\alpha_{t-1}\hat{\boldsymbol{x}}_\theta(\boldsymbol{x}_t, t) + \sigma_{t-1}\hat{\boldsymbol{z}}, \mathbf{0}) \quad (4) \\
\boldsymbol{x}_{t-1} &= \alpha_{t-1}\hat{\boldsymbol{x}}_\theta(\boldsymbol{x}_t, t) + \sigma_{t-1}\hat{\boldsymbol{z}},
\end{aligned}
$$

where $\hat{\boldsymbol{x}}_0$ is predicted with $f_\theta(x_t, t, I^C)$ and $\hat{\boldsymbol{z}}$ is the estimated noise which can be calculated with Equation 2, $\hat{\boldsymbol{z}} = (\boldsymbol{x}_t - \alpha_t\hat{\boldsymbol{x}}_0)/\sigma_t$. Integrating the above equation we have:

$$
\boldsymbol{x}_{t-1} = \alpha_{t-1}f_\theta(x_t, t, I^C) + \sigma_{t-1}(\boldsymbol{x}_t - \alpha_t\hat{\boldsymbol{x}}_0)/\sigma_t. \quad (5)
$$

**Conditional image choice.** To get realistic super-resolution images, [19, 16] also introduced DPM with conditional denoising on a pre-trained feature extractor or a bicubic upsampled image on low-resolution image. In this work, we leverage the power of the current development of SISR to provide a better conditional image. Specifically, given a low-resolution image $I^{LR}$ and a pre-trained super-resolution model $\phi_\theta$, we generate our conditional image by $I^C = \phi_\theta(I^{LR})$, which has been proved to be more plausible for obtaining results with better perceptual quality in ablation study 3.3. Algorithm 1 shows the whole process of our cDPMSR.

## 3. EXPERIMENTS

### 3.1. Experimental Settings

**Implementation Details.** We use 800 image pairs in DIV2K as the training set. We take public benchmark datasets, *i.e.*, Set5, Set14, Urnban100, BSD100, and Manga109 as the test set to compare with other methods. For the diffusion model, we set $T = 1000$ for training and $T = 100$ during inference time. We take the pre-trained super-resolution models (SwinIR [22], EDSR [20], and RCAN [21]) to provide the initial super-resolution image,*i.e.*the conditional input image. The conditional diffusion model is trained with Adam optimizer and batch size 16, learning rate $1 \times 10^{-4}$ for 300k steps. The architecture of the model is the same as the one in [16].

**A note on metrics.** The previous study has shown the distortion and perceptual quality are at odds with each other, and there is a trade-off between them [25]. Since our work focuses on the perceptual quality, except the distortion metrics: PSNR and SSIM, we also provide perceptual metrics: LPIPS [26] to show that our method can generate better perceptual results than other methods. LPIPS is recently introduced as a reference-based image quality evaluation metric, which computes the perceptual similarity between the ground truth and the SR image.

**Table 1**. Results for $4\times$SR on Set5, Set14, BSD100, Urban100, and Manga109. The best three perceptual results are highlighted in red, green, and blue colors, respectively. The **bold** represents the best distortion result among generative-based methods.

| Method | | EDSR | RCAN | ESRGAN | SwinIR | SRFlow | SRDiff | SR3 | Ours | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | EDSR+ | RCAN+ | SwinIR+ |
| Set5 | LPIPS↓ | 0.0922 | 0.1096 | 0.0596 | 0.0899 | 0.0767 | 0.0770 | 0.1265 | 0.0619 | 0.0606 | 0.0564 |
| | PSNR↑ | 32.43 | 32.64 | 30.46 | 32.69 | 28.35 | 30.94 | 27.31 | 30.78 | 30.84 | **31.03** |
| | SSIM↑ | 0.8985 | 0.9002 | 0.8516 | 0.9018 | 0.8138 | **0.8738** | 0.7844 | 0.8684 | 0.8645 | 0.8676 |
| Set14 | LPIPS↓ | 0.1445 | 0.1387 | 0.0867 | 0.1416 | 0.1318 | 0.1009 | 0.1531 | 0.0883 | 0.0865 | 0.0827 |
| | PSNR↑ | 28.68 | 28.85 | 26.28 | 28.82 | 24.97 | **27.23** | 25.48 | 27.11 | 27.02 | 27.14 |
| | SSIM↑ | 0.7883 | 0.7885 | 0.6980 | 0.7918 | 0.6908 | **0.7432** | 0.6889 | 0.7316 | 0.7257 | 0.7341 |
| BSD100 | LPIPS↓ | 0.1517 | 0.1536 | 0.0834 | 0.1497 | 0.1831 | 0.1041 | 0.1392 | 0.0935 | 0.0953 | 0.0834 |
| | PSNR↑ | 27.73 | 27.74 | 25.29 | 27.86 | 24.65 | 25.95 | 25.21 | 25.74 | 25.87 | **25.96** |
| | SSIM↑ | 0.7425 | 0.7430 | 0.6495 | 0.7466 | 0.6573 | **0.6833** | 0.6498 | 0.6681 | 0.6706 | 0.6743 |
| Urban100 | LPIPS↓ | 0.1251 | 0.1220 | 0.0944 | 0.1185 | 0.1279 | 0.1077 | 0.1993 | 0.0997 | 0.0997 | 0.0934 |
| | PSNR↑ | 26.65 | 26.75 | 24.35 | 27.08 | 23.65 | 25.34 | 22.49 | 25.45 | 25.59 | **25.86** |
| | SSIM↑ | 0.8036 | 0.8066 | 0.7327 | 0.8165 | 0.7312 | 0.7661 | 0.6336 | 0.7649 | 0.7681 | **0.7796** |
| Manga109 | LPIPS↓ | 0.0628 | 0.0544 | 0.0420 | 0.0592 | 0.0660 | 0.0473 | 0.1427 | 0.0409 | 0.0396 | 0.0374 |
| | PSNR↑ | 31.06 | 31.20 | 28.48 | 31.67 | 27.14 | 28.67 | 24.69 | 29.07 | 29.39 | **29.60** |
| | SSIM↑ | 0.9160 | 0.9170 | 0.8595 | 0.9227 | 0.8244 | 0.8851 | 0.7568 | 0.8791 | 0.8816 | **0.8874** |

## 3.2. Quantitative and qualitative results

To verify the effectiveness of our cDPMSR, we select some SOTA generative methods to conduct the comparative experiments, including ESRGAN [2], SRFlow [5], SRDiff [19], SR3 [16]. We selected EDSR [20], RCAN [21], and SwinIR [22] to provide conditional input, respectively. Therefore, we report three cases for our cDPMASR, *i.e.*, **EDSR+**, **RCAN+**, and **SwinIR+**. All the results are obtained by the provided codes or from the publicized papers. As shown in Fig.2 and 3, the results of EDSR, RCAN, and SwinIR are so smooth that some details are missed. Because these methods are PSNR-directed and they all focus on obtaining results with good distortion [25] and they certainly got good PSNR results in Tab.1. It seems the results generated by ESRGAN in Fig. 2 and 3 include more details than other methods, but it introduces too many false artifacts compared to the ground truth. And the results of SRflow seem a little noisy. With better conditional input, our method exhibits superior performance on both quantitative and qualitative results than SR3 [16]. Though SRdiff obtains comparable numeric results in Tab.1, the visual results of our cDPMSR are closer to ground truths (Especially, the forehead in Fig.2 and the plants in Fig.3).

**Table 2**. Results of ablation study for different conditional inputs on BSD100. ($4 \times$ SR)

| Method | LR | SR-RRDB | ours | | |
|---|---|---|---|---|---|
| | | | EDSR+ | RCAN+ | SwinIR+ |
| LPIPS↓ | 0.1412 | 0.1096 | 0.0935 | 0.0953 | **0.0834** |
| PSNR↑ | 24.35 | 25.01 | 25.74 | 25.87 | **25.96** |
| SSIM↑ | 0.6402 | 0.6589 | 0.6681 | 0.6706 | **0.6743** |

## 3.3. Ablation Study

**Different conditional inputs.** Here, we conduct experiments to verify how different conditional inputs influence performance. We adopt LR, SRs generated by EDSR, RCAN,



**Fig. 4**. Visual results on '69020' from BSD100 under different conditional inputs ($\times$4 scale. Best view with zoomed-in.)

SwinIR, and the RRDB trained in [19] as the conditional inputs to perform experiments, respectively. As shown in Tab.2 and Fig.4, without any pre-processing, the result under LR conditional performs worst in both quantitative and qualitative. After being pre-trained by RRDB, EDSR, RCAN, SwinIR, the conditional inputs can restore more details, which pushes our cDPMSR model gets better performance.

## 4. CONCLUSION

Our work revisits DPM in SR and reveals taking a pre-super-resolved version for the given LR image as the conditional input can help to achieve a better high-resolution image. Based on this, we propose a simple but non-trivial DPM-based super-resolution post-process framework, *i.e.*, cDPMSR. By taking a pre-super-resolved version of the given LR image and adapting the standard DPM to perform super-resolution, our cDPMSR improves both qualitative and quantitative results and can generate more photo-realistic counterparts for the low-resolution images on benchmark datasets (Set5, Set14, Urban100, BSD100, Manga109). In the future, we will extend our cDPMSR to images with more complex degradation.

# 5. REFERENCES

[1] Axi Niu, Yu Zhu, Chaoning Zhang, Jinqiu Sun, Pei Wang, In So Kweon, and Yanning Zhang, "Ms2net: Multi-scale and multi-stage feature fusion for blurred image super-resolution," *TCSVT*, 2022.

[2] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy, "Esrgan: Enhanced super-resolution generative adversarial networks," in *ECCV workshops*, 2018.

[3] Jae Woong Soh, Gu Yong Park, Junho Jo, and Nam Ik Cho, "Natural and realistic single image super-resolution with explicit natural manifold discrimination," in *CVPR*, 2019.

[4] Chunwei Tian, Xuanyu Zhang, Jerry Chun-Wei Lin, Wangmeng Zuo, Yanning Zhang, and Chia-Wen Lin, "Generative adversarial networks for image super-resolution: A survey," *arXiv preprint arXiv:2204.13620*, 2022.

[5] Andreas Lugmayr, Martin Danelljan, Luc Van Gool, and Radu Timofte, "Srflow: Learning the super-resolution space with normalizing flow," in *ECCV*, 2020.

[6] Jingyun Liang, Andreas Lugmayr, Kai Zhang, Martin Danelljan, Luc Van Gool, and Radu Timofte, "Hierarchical conditional flow: A unified framework for image super-resolution and image rescaling," in *ICCV*, 2021.

[7] Valentin Wolf, Andreas Lugmayr, Martin Danelljan, Luc Van Gool, and Radu Timofte, "Deflow: Learning complex image degradations from unpaired data with conditional flows," in *CVPR*, 2021.

[8] Sefi Bell-Kligler, Assaf Shocher, and Michal Irani, "Blind super-resolution kernel estimation using an internal-gan," *NeurIPS*, 2019.

[9] Mohammad Emad, Maurice Peemen, and Henk Corporaal, "Dualsr: Zero-shot dual learning for real-world super-resolution," in *WACV*, 2021.

[10] Luke Metz, Ben Poole, David Pfau, and Jascha Sohl-Dickstein, "Unrolled generative adversarial networks," *arXiv preprint arXiv:1611.02163*, 2016.

[11] Suman Ravuri and Oriol Vinyals, "Classification accuracy score for conditional generative models," *NeurIPS*, 2019.

[12] Andreas Lugmayr, Martin Danelljan, and Radu Timofte, "Ntire 2021 learning the super-resolution space challenge," in *CVPR*, 2021.

[13] Jay Whang, Mauricio Delbracio, Hossein Talebi, Chitwan Saharia, Alexandros G Dimakis, and Peyman Milanfar, "Deblurring via stochastic refinement," in *CVPR*, 2022.

[14] Younggeun Kim and Donghee Son, "Noise conditional flow model for learning the super-resolution space," in *CVPR*, 2021.

[15] Jingyun Liang, Kai Zhang, Shuhang Gu, Luc Van Gool, and Radu Timofte, "Flow-based kernel prior with application to blind super-resolution," in *CVPR*, 2021.

[16] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi, "Image super-resolution via iterative refinement," *TPAMI*, 2022.

[17] Charles Laroche and Matias Tassano, "Bridging the domain gap in real world super-resolution," in *ICIP*, 2022.

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

[19] Haoying Li, Yifan Yang, Meng Chang, Shiqi Chen, Huajun Feng, Zhihai Xu, Qi Li, and Yueting Chen, "Srdiff: Single image super-resolution with diffusion probabilistic models," *Neurocomputing*, 2022.

[20] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *CVPR workshops*, 2017.

[21] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu, "Image super-resolution using very deep residual channel attention networks," in *ECCV*, 2018.

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

[23] 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.

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

[25] Yochai Blau and Tomer Michaeli, "The perception-distortion tradeoff," in *CVPR*, 2018.

[26] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *CVPR*, 2018.