# CDFormer: When Degradation Prediction Embraces Diffusion Model for Blind Image Super-Resolution

Qingguo Liu, Chenyi Zhuang, Pan Gao*, Jie Qin*
College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics
{liuqingguo, chenyi.zhuang, pan.gao, jie.qin}@nuaa.edu.cn

## Abstract

*Existing Blind image Super-Resolution (BSR) methods focus on estimating either kernel or degradation information, but have long overlooked the essential content details. In this paper, we propose a novel BSR approach, Content-aware Degradation-driven Transformer (CDFormer), to capture both degradation and content representations. However, low-resolution images cannot provide enough content details, and thus we introduce a diffusion-based module $CDFormer_{diff}$ to first learn Content Degradation Prior (CDP) in both low- and high-resolution images, and then approximate the real distribution given only low-resolution information. Moreover, we apply an adaptive SR network $CDFormer_{SR}$ that effectively utilizes CDP to refine features. Compared to previous diffusion-based SR methods, we treat the diffusion model as an estimator that can overcome the limitations of expensive sampling time and excessive diversity. Experiments show that CDFormer can outperform existing methods, establishing a new state-of-the-art performance on various benchmarks under blind settings. Codes and models will be available at https://github.com/I2-Multimedia-Lab/CDFormer.*

## 1. Introduction

Image Super-Resolution (SR), which aims to reconstruct high-resolution (HR) images from their low-resolution (LR) counterparts, is a long-standing low-level vision problem in the research community. Advanced methods based on deep learning networks start from the assumption that the LR image is degraded from the HR image by a specific process, which can be formulated as follows:

$$I_{LR} = (I_{HR} * k_h)\downarrow_s + n. \tag{1}$$

where $k_h$ is a blur kernel, $\downarrow_s$ is a downsampling operation with scale factor $s$, $n$ is additive Gaussian noise. In practice,

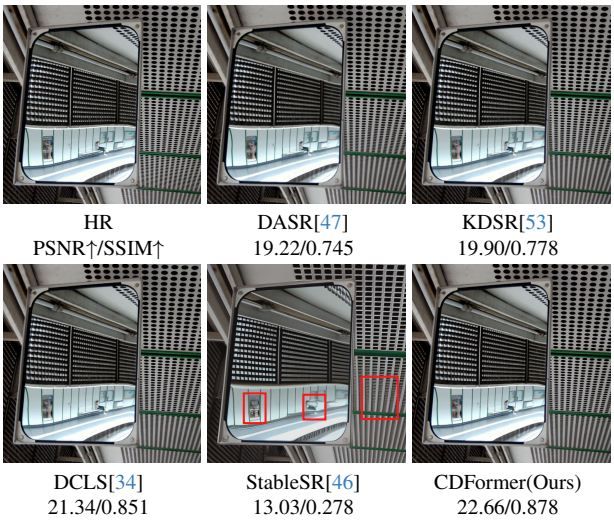



Figure 1. Blind image Super-Resolution for scale 4 on kernel width 1.2. Our proposed CDFormer with CDP is capable of producing sharp and clean textures and outperforms previous state-of-the-art approaches DASR, KDSR, DCLS, and StableSR.

however, the transformation from HR image to LR image is complicated and usually difficult to formulate simply as Eq. (1). As a result, previous works [6, 8–10, 22, 25–27, 29] explore non-blind image Super-Resolution, where the kernel and downsampling are assumed to be known as the prior, have faced great challenges in real-world scenarios, thus extending to Blind image Super-Resolution (BSR).

Existing BSR methods can be divided into two categories: Kernel Prediction (KP) [1, 12, 28, 34] and Degradation Prediction (DP) [31, 47, 53]. KP methods produce desirable SR results in most cases, but still face severe performance drops when it has to deal with complex degradations. This inevitable limitation is subject to the unavailability of real-word kernels as well as training distribution mismatch. Furthermore, KP methods have to be limited to blur kernel-based degradation and cannot address other types of degradation (*e.g.*, noise). In contrast, DP methods that DASR [47] exemplify, explore degradation representa-

*Corresponding authors.

tions in an unsupervised manner. However, although DASR can outperform some KP methods [1, 12] for simple degradation, a gap remains between DP and state-of-the-art KP methods [28, 34] for more complex cases.

More recently, the advances in Diffusion Models (DMs) [13, 38, 41] have revolutionized the community of image synthesis [7, 14, 42] as well as image restoration tasks, such as inpainting [5, 32] and Super-Resolution [17, 39]. Several works [24, 33, 39, 46] have explored the powerful generative ability of DMs for Super-Resolution and attracted extensive attention. We argue that although these works provide novel viewpoints and solutions for image SR, they suffer from two main drawbacks: (1) treating DMs as the main SR net requires numerous inference steps ($\sim$ 50 to 1000 steps), which will be computationally expensive and not suitable for real-time applications. Although the former issue can be alleviated by reducing iteration number [49], it will lead to quality deterioration of the SR results. (2) undesirable artifacts such as joint misalignment or texture distortion have been introduced as a result of the error propagation inherent in single-step noise prediction models.

As shown in Fig. 1, DCLS [34], a state-of-the-art KP method, exhibits significant improvements compared to DP methods, DASR [47] and KDSR [53], in terms of PSNR and SSIM. Despite being on the leading edge of DM-based SR approaches, StableSR [46] suffers from incorrect textures (an ellipse is reconstructed as a square) and loss of detail (the person in the mirror disappears). We suspect that the erroneous texture results via StableSR may be blamed on the diversity in the pre-trained DMs which is excessive for SR task. Due to the low quality of given image, the priors in pre-trained DMs may misinterpret the given LR images and reconstruct them in a wrong context.

In this paper, we rethink the existing DP methods that concentrate on estimating degradation representations from LR images while neglecting the essential content information. We propose CDFormer, a Content-aware Degradation-driven Transformer network for BSR. Instead of employing the DM as the whole SR network, we design a Content Degradation Prior (CDP) generation module $CDFormer_{diff}$, where the DM is treated as an estimator to recreate CDP from LR images. The Content-aware Degradation-driven Transformer SR module $CDFormer_{SR}$ further utilizes the estimated CDP via several injection modules to adaptively refine features and benefit reconstruction. Our main contributions are as follows:

- We introduce a Content Degradation Prior (CDP) generation module. The CDP is learned from the pairs of HR and LR images in the first stage, while recreated from the LR images solely via a diffusion-based estimator in the second stage.
- We propose a CDP-guided SR network where CDP is injected via learnable affine transformations as well as interflow mechanisms to improve the representation of both high- and low-frequency details.
- Experiments demonstrate the superiority of CDFormer, leading to a new state-of-the-art performance. With content estimation, CDFormer achieves unprecedented SR results even for severely degraded images.

## 2. Related Work

### 2.1. Image Super-Resolution

**Non-blind Image Super-Resolution** algorithms, pioneered by SRCNN [8], mostly start with an assumption in the degradation process (*e.g.* bicubic downsampling and specific blur kernels) and can produce appealing SR results for the synthetic cases, employing either recurrent neural networks (RNNs) [18, 43], adversarial learning [23, 48, 50] or attention mechanisms [6, 30, 36, 37]. To meet the real-world challenges with multiple degrading effects, SRMD [8] proposes to incorporate an LR image with a stretched blur kernel, and UDVD [54] utilizes consecutive dynamic convolutions and multi-stage loss to refine image features. Recently, Transformer-based networks [3, 4, 27] have emerged and achieved state-of-the-art results. Nevertheless, most non-blind image Super-Resolution methods fail to cope with complicated degradation cases.

**Blind Image Super-Resolution** then emerged to deal with real-world scenarios where the degradation kernels are complicated. Previous methods based on Kernel Prediction (KP) utilize explicit [12, 16, 19, 40, 56] or implicit [1, 34] techniques to predict blur kernels, and then guide non-blind SR networks through kernel stretching strategies. DCLS [34] presents a least squares deconvolution operator to produce clean features using estimated kernels. While KP methods require the ground truth labels of the degradation kernels, Degradation Prediction (DP) methods further encourage learning degradation representation. DASR [47] extracts degradation representation in an unsupervised manner. KDSR [53] takes full advantage of knowledge distillation to distinguish different degradations.

Although the above DP methods can handle real-world degradations to some extent, we do emphasize that only degradation representation has been utilized in previous works, which may render the network agnostic to the texture information. In contrast, our proposed CDFormer explicitly takes both degradation and content representations into consideration and can achieve remarkable improvement.

### 2.2. Diffusion Models for Super-Resolution

Diffusion models (DMs) [7, 13, 20] have been proving to be a powerful generative engine. Compared to other generative models like Generative Adversarial Networks (GANs), DMs define a parameterized Markov chain to optimize the lower variational bound on the likelihood function, allowing

it to fit all data distributions. DMs for the super-resolution task have been investigated in several papers. SR3 [39] and SRDiff [24] adopt DDPM [13] for SR to reconstruct the LR image via iterative denoising, StableSR [46] further explore the Latent Diffusion Model (LDM) [38] to enhance efficiency. However, as discussed above, the use of DMs as SR net can induce error propagation and consequently incorrect textures in SR results. Moreover, DMs are known for being quite expensive in terms of time and computation, which is crucial for real-time applications.

The solution of generating particular information by diffusion models has provided a new perspective to enhance accuracy and stability. DiffIR [52] that trains the DM to estimate prior representation achieves state-of-the-art performance in image restoration. More than DiffIR, we encourage the diffusion process as an estimator to recover both degradation and content representations from the LR images, therefore improving reconstruction in textural details.

# 3. Methodology

## 3.1. Motivation

Previous BSR methods based on KP have been observed to be ineffective when dealing with complex degradations. DP approaches instead estimate degradation representation, while still in a content-independent manner. To address these limitations, we propose to predict both degradation and content representations, thus reconstructing images with more harmonious textures.

The content detail is supposed to be rich in high-resolution images. However, SR task given only low-resolution images naturally lacks such information. In this case, we adopt the diffusion model to retrace content and degradation representations from LR images. The diffusion module in CDFormer is to estimate high-level information rather than reconstruct low-level images, thus we can overcome the limitations of low efficiency and excessive diversity that existed in previous diffusion-based SR approaches.

Moreover, recent research has shown the robust modeling capabilities of Transformers compared with pure CNN architectures, however, still lacks an inductive bias for modeling low-frequency information [4]. We are motivated to redesign an adaptive SR network that takes the full advantage of the estimated CDP to model both high- and low-frequency information for accurate reconstruction.

## 3.2. Our Method

As illustrated in Fig. 2, our proposed method is composed of a Content Degradation Prior generation module $CDFormer_{diff}$ and a Content-aware Degradation-driven Transformer SR module $CDFormer_{SR}$. We apply a two-stage training strategy to steer the diffusion model. Sec. 3.2.1 describes the training of ground-truth encoder $E_{GT}$ that learns CDP from both LR and HR images. Sec. 3.2.2 presents the pipeline to generate CDP from LR images by encoder $E_{LR}$ with the diffusion process. In both stages, $CDFormer_{SR}$ is trainable, which is stacked by some residual groups and a reconstruction net. Each residual group introduces several Content-aware Degradation-driven Refinement Blocks (CDRBs), where Content Degradation Injection Module (CDIM) is applied to infuse the estimated CDP. For simplicity, we provide a brief background of the diffusion model in the supplementary material and the following will focus on our approach.

### 3.2.1 STAGE 1: Learn CDP from $I_{LR}$ and $I_{HR}$

In this stage, $E_{GT}$ will be trained to construct the real data distribution under the supervision of both HR and LR images, and $CDFormer_{SR}$ will also be trained to ensure that estimated representations have been efficiently used.

The estimator $E_{GT}$ is designed to extract the degradation information from LR and HR images pair, as well as the content details from the HR images, which is the so-called CDP, denoted as:

$$Z_0 = E_{GT}(Concat((I_{HR}) \downarrow_s, I_{LR}), I_{HR}), \quad (2)$$

where the CDP $Z_0 \in \mathbb{R}^{C_z}$, $\downarrow_s$ denotes a Pixel-Unshuffle operation with scale $s$. Specifically, $E_{GT}$ is implemented by several residual blocks and an MLP layer, as depicted in the right of Fig. 2. The fusion of CDP with features in $CDFormer_{SR}$ is accomplished in the injection module CDIM through channel-wise affine transformations which are learnable during training, formulated as:

$$F^{'} = Linear(Z_0) \odot Norm(F) + Linear(Z_0), \quad (3)$$

where $F, F^{'} \in \mathbb{R}^{H \times W \times C}$ are input and output feature maps respectively, $\odot$ is a element-wise multiplication, $Norm(\cdot)$ denotes Layer Normalization.

The goal of SR network $CDFormer_{SR}$ is to reconstruct high-resolution images with the guidance of CDP. To enhance the representation ability, CDRB combines spatial attention and channel attention mechanisms, and further incorporates CNN and Transformer features via interflow mechanism, enabling the CDRB module to adaptively refine both high and low-frequency information.

As depicted in Fig. 3, $i$-th CDRB involves four CDIMs to inject CDP into the feature maps $F_i^j \in \mathbb{R}^{H \times W \times C}$ by:

$$\hat{F}_i^j = CDIM_{i,j}(F_i^j, Z_0), i = 1, ..., N_2, j = 1, 2, 3, 4 \quad (4)$$

where $Z_0$ is the CDP predicted by $E_{GT}$, $CDIM_{i,j}$ is the $j$-th CDIM in $i$-th CDRB.

To ensure effective learning of representations, we apply two kinds of self-attention and deep convolution operations. To be specific, we first utilize Spatial Window Self-Attention (SW-SA) that calculates attention scores within

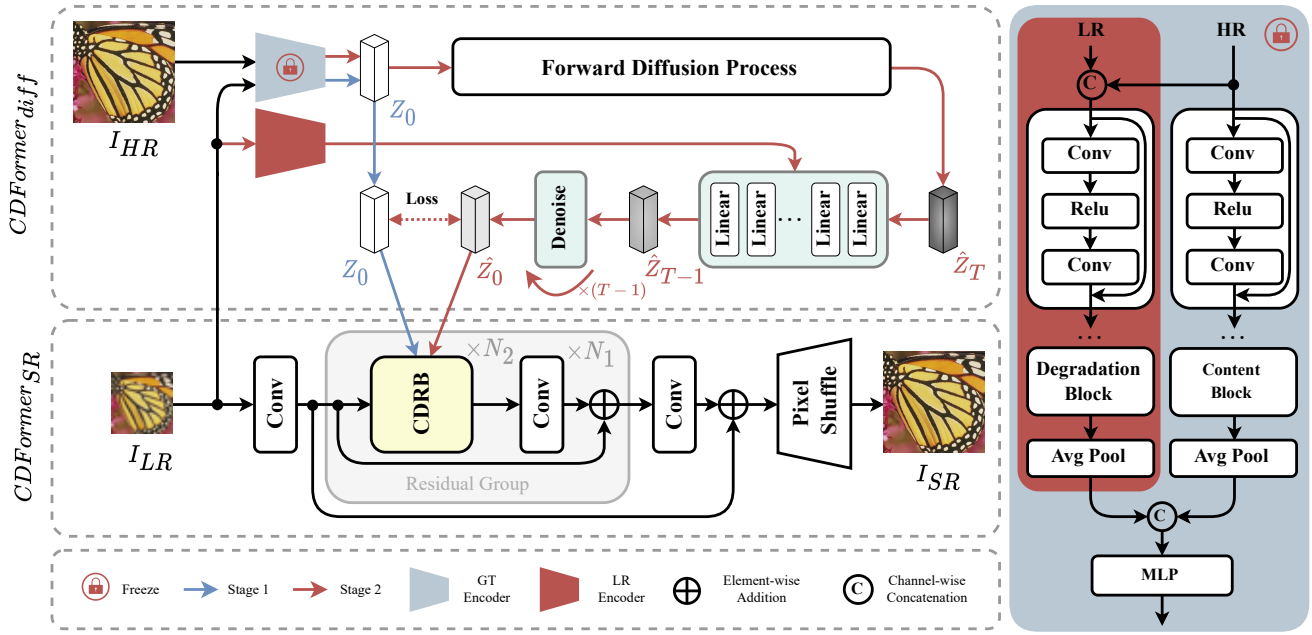



Figure 2. Overall architecture of our proposed CDFormer. In the first stage (blue line), we train the GT encoder to learn Content Degradation Prior (CDP) from both HR and LR images to guide the SR network $CDFormer_{SR}$. In the second stage (red line), only LR images are input into LR encoder to produce conditional vectors, which helps the diffusion model to recreate CDP.

non-overlapping windows. Given the input features $\hat{F}_i^1 = CDIM_{i,1}(F_i^1, Z_0)$, we first obtain query, key, and value elements through linear projection:

$$Q = \hat{F}_i^1 W_Q, K = \hat{F}_i^1 W_K, V = \hat{F}_i^1 W_V, \tag{5}$$

where $W_Q, W_K, W_V \in \mathbb{R}^{C\times C}$ are learnable parameter matrices, with no bias appended. Window partition is then applied to divide $Q, K, V$ into $\frac{HW}{N_w}$ non-overlapping windows, each window with $N_w$ length. These flattened and reshaped elements are denoted as $Q_s, K_s$, and $V_s$. Finally, the output features are obtained as follows:

$$F_i^S = Softmax(Q_s K_s^T / \sqrt{d_k}) V_s, \tag{6}$$

where $d_k$ is the variance of the attention score.

Moreover, we introduce an interflow mechanism that utilizes two kinds of distiller between two branches to adaptively modulate the CNN and Transformer features. Specifically, for feature maps in $\mathbb{R}^{H\times W\times C}$, Channel Distiller $\tilde{C}$ transforms to $\mathbb{R}^{1\times 1\times C}$, while Spatial Distiller $\tilde{S}$ changes to $\mathbb{R}^{H\times W\times 1}$. We formulate the above process as:

$$\begin{aligned} F_i^S &= SW\text{-}SA(\hat{F}_i^1), F_i^W = DConv(\hat{F}_i^1), \\ F_i^2 &= F_i^S \odot \tilde{S}(F_i^W) + F_i^W \odot \tilde{C}(F_i^S) + F_i^1, \\ F_i^3 &= FFN(\hat{F}_i^2) + F_i^2 \end{aligned} \tag{7}$$

where $DConv(\cdot)$ denotes a depth-wise convolution layer, $\hat{F}_i^2 = CDIM_{i,2}(F_i^2, Z_0)$, $FFN(\cdot)$ is a feed-forward network with GELU activation.

Different from SW-SA which learns pixel-wise relations, Channel-wise Self-Attention (CW-SA) focuses on understanding relationships between channels. Given feature maps after injection operation as $\hat{F}_i^3 = CDIM_{i,3}(F_i^3, Z_0)$, query, key, and value elements $Q_c, K_c, V_c$ are projected from $\hat{F}_i^3$. Channel-wise relationship is computed by:

$$F_i^C = Softmax(Q_c^T K_c / \alpha) V_c, \tag{8}$$

where $\alpha$ is a learnable temperature parameter.

Again, feature aggregation is achieved by introducing the interflow between channel-wise self-attention and deep convolution operations. This technique allows the network to capture both global and local dependencies. This process is formulated as follows:

$$\begin{aligned} F_i^C &= CW\text{-}SA(\hat{F}_i^3), \dot{F}_i^W = DConv(\hat{F}_i^3), \\ F_i^4 &= F_i^C \odot \tilde{C}(\dot{F}_i^W) + \dot{F}_i^W \odot \tilde{S}(F_i^C) + F_i^3, \\ F_{i+1} &= FFN(\hat{F}_i^4) + F_i^4 \end{aligned} \tag{9}$$

where $\tilde{C}$ and $\tilde{S}$ denote the channel distiller and spatial distiller, respectively. $\hat{F}_i^4 = CDIM_{i,4}(F_i^4, Z_0)$.

Finally, we jointly optimize $E_{GT}$ and $CDFormer_{SR}$ by reconstruction loss between HR images and reconstructed SR images as the training objective for the first stage:

$$\mathcal{L}_{rec} = \|I_{HR} - I_{SR}\|_1, \tag{10}$$

#### 3.2.2 STAGE 2: Generate CDP from $I_{LR}$

Theoretically, $I_{HR}$ is full of content details. However, for the SR task, such high-resolution images are unknown dur-

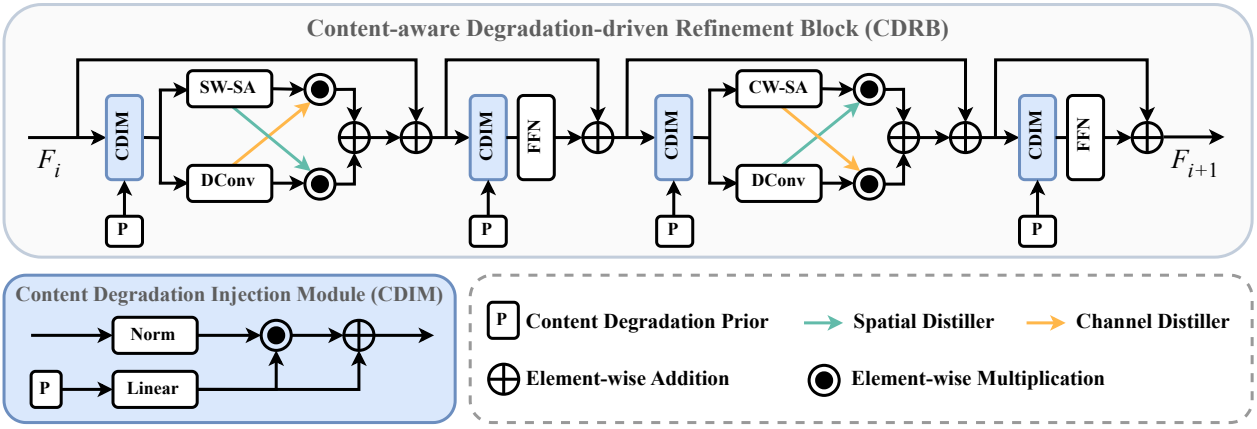



Figure 3. Details of Content-aware Degradation-driven Refinement Block (CDRB).

ing inference. In the second stage, we therefore propose to treat the diffusion model as an estimator and generate CDP from LR images, exploiting the capability of diffusion models to approximate the real data distribution.

Specifically, we reuse the pre-trained encoder $E_{GT}$ in the first stage to produce the initial CDP $Z_0 \in \mathbb{R}^{C_z}$, which is supposed to be the ground truth representation of content and degradation. Following the normal routine to train a diffusion model, we add Gaussian noise to $Z_0$ in the forward diffusion process to produce a noisy representation:

$$Z_T = \sqrt{\bar{\alpha}_T} Z_0 + \sqrt{1-\bar{\alpha}_T}\epsilon \tag{11}$$

where $T$ is the total number of time steps, $\bar{\alpha}_T = \prod_{i=1}^{T} \alpha_i$ is predefined schedule, and added noise $\epsilon \sim \mathcal{N}(0,1)$.

During the reverse process, $E_{LR}$ that is a replicate of $E_{GT}$ with the degradation branch only, will be trained to produce a one-dimensional condition vector $c \in \mathbb{R}^{C_z}$ from the LR images. This conditional vector will then be fused into each reverse step, guiding the diffusion model to generate a great representation of CDP based on the LR images. Notice that each training step $CDFormer_{diff}$ performs the whole sampling process with $T$ iterations, which is different from the traditional DMs that minimize $||\epsilon - \epsilon_\theta(x_t, t, c)||$ for a single diffusion step. Instead, we compute $\hat{Z}_0$ in every training step as follows:

$$\hat{Z}_{t-1} = \frac{1}{\sqrt{\alpha_t}}(\hat{Z}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(\hat{Z}_t, t, c))) + \sigma_t \epsilon, \quad t = T, ..., 1 \tag{12}$$

where the one-dimensional condition vector $c = E_{LR}(I_{LR})$ is concatenated with $\hat{Z}_t$, variance $\sigma_t = \sqrt{1-\alpha_t}$.

The generated representation $\hat{Z}_0$ is expected to recreate the lost information in LR images, thus we compute the $L_1$ distance between the CDP from $E_{GT}$ and from diffusion as:

$$\mathcal{L}_{diff} = \|Z_0 - \hat{Z}_0\|_1 \tag{13}$$

To ensure efficient learning, $CDFormer_{diff}$ and $CDFormer_{SR}$ will be trained jointly, by minimizing the loss function $\mathcal{L}_{loss} = \alpha_{diff}\mathcal{L}_{diff} + \mathcal{L}_{rec}$. Notice that in the second stage, we inject CDP $\hat{Z}_0$ predicted by diffusion model into $CDFormer_{SR}$, instead of $Z_0$ in the first stage. The algorithm is detailed in the supplementary material.

During training, $\hat{Z}_T$ is actually given from $I_{HR}$, while for inference we exclusively perform the reverse diffusion process from a Gaussian noise, i.e., $\hat{Z}_T \sim \mathcal{N}(0,1)$. The conditional vector $c$ obtained from LR images will also participate in. Utilizing the denoising ability of the diffusion model, CDFormer is able to generate CDP from LR images, with abundant content and degradation representations to guide $CDFormer_{SR}$ to reconstruct both high- and low-frequency information. Moreover, $CDFormer_{diff}$ achieves plausible results with fewer sampling iterations ($T = 4$ in our experiments) and parameters ($\sim$ 3M).

## 4. Experiments

### 4.1. Experiment Settings

**Implementation Details.** CDFormer in baseline is stacked by 6 residual groups, each containing 3 CDRBs with window size $8 \times 32$ and 180 channels, and the channel for CDP is $C_z = 256$. For a fair comparison, we develop a lightweight variant of CDFormer, denoted as CDFormer-S, where CDRB window size is reduced to $4 \times 16$ and channel is reduced to 96. We use Adam [21] optimizer with $\beta_1 = 0.9, \beta_2 = 0.99$ to update parameters, and $\alpha_{diff} = 0.01$ for the second stage training. Each stage is trained with 300 epochs and batch size 4. We set the initial learning rate as $1 \times 10^{-4}$, and half it every 125 epochs.

**Datasets and Metrics.** Our models are trained on two widely used datasets: DIV2K [44] and Flickr2K [29]. For testing, we adopt four benchmark datasets: Set5 [2], Set14 [55], B100 [35], Urban100 [15]. All methods will be evaluated by PSNR and SSIM [51] unless otherwise specified.

Table 1. Comparison with diffusion-based SR models test on 192 × 192 resolution for ×4 scale. KP method DCLS for reference.

| Method | Params(M) | FLOPs(T) | Time(s) | Set14 / kernel width=0 | | | | Urban100 / kernel width=0 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | PSNR ↑ | SSIM ↑ | FID ↓ | LPIPS ↓ | PSNR ↑ | SSIM ↑ | FID ↓ | LPIPS ↓ |
| SR3[39] | 187.65 | 6601 | 532 | 25.41 | 0.7462 | 83.63 | **0.246** | - | - | - | - |
| StableSR [46] | 1409.11 | - | 71 | 19.66 | 0.5406 | 97.84 | 0.282 | 20.56 | 0.6753 | 54.21 | **0.195** |
| DCLS[34] | 13.63 | 0.279 | 0.55 | 28.61 | 0.7816 | 78.21 | 0.293 | 26.50 | 0.7973 | 43.25 | 0.220 |
| CDFormer-S | 11.09 | 0.355 | 0.14 | 28.80 | 0.7876 | 78.69 | 0.287 | 26.56 | 0.8023 | 26.40 | 0.214 |
| CDFormer | 24.46 | 0.725 | 0.39 | **29.00** | **0.7918** | **75.43** | 0.280 | **27.21** | **0.8189** | **23.94** | 0.197 |

Table 2. Quantitive evaluation by PSNR (↑) on noise-free degradations with isotropic Gaussian kernels. Best in **red** and second in blue.

| Method | Scale | Set5 | | | | Set14 | | | | B100 | | | | Urban100 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Kernel Width | | 0 | 0.6 | 1.2 | 1.8 | 0 | 0.6 | 1.2 | 1.8 | 0 | 0.6 | 1.2 | 1.8 | 0 | 0.6 | 1.2 | 1.8 |
| Bicubic | ×2 | 33.66 | 32.30 | 29.28 | 27.07 | 30.24 | 29.21 | 27.13 | 25.47 | 29.56 | 28.76 | 26.93 | 25.51 | 26.88 | 26.13 | 24.46 | 23.06 |
| RCAN [58] | | **38.27** | 35.91 | 31.20 | 28.50 | **34.12** | 32.31 | 28.48 | 26.33 | **32.41** | 31.16 | 28.04 | 26.26 | **33.34** | 29.80 | 25.38 | 23.44 |
| DASR [47] | | 37.87 | 37.47 | 37.19 | 35.43 | 33.34 | 32.96 | 32.78 | 31.60 | 32.03 | 31.78 | 31.71 | 30.54 | 31.49 | 30.71 | 30.36 | 28.95 |
| DCLS [34] | | 38.06 | 38.04 | 37.66 | 36.06 | 33.62 | 33.52 | 33.52 | 32.27 | 32.23 | 32.25 | 32.12 | 30.93 | 32.36 | 32.17 | 31.81 | 30.23 |
| CDFormer (Ours) | | 38.25 | **38.25** | **37.88** | **36.32** | 34.10 | **34.01** | **33.88** | **32.57** | 32.40 | **32.39** | **32.25** | **31.03** | 33.11 | **32.62** | **32.08** | **30.47** |
| Kernel Width | | 0 | 0.8 | 1.6 | 2.4 | 0 | 0.8 | 1.6 | 2.4 | 0 | 0.8 | 1.6 | 2.4 | 0 | 0.8 | 1.6 | 2.4 |
| Bicubic | ×3 | 30.39 | 29.42 | 27.24 | 25.37 | 27.55 | 26.84 | 25.42 | 24.09 | 27.21 | 26.72 | 25.52 | 24.41 | 24.46 | 24.02 | 22.95 | 21.89 |
| RCAN [58] | | 34.74 | 32.90 | 29.12 | 26.75 | 30.65 | 29.49 | 26.75 | 24.99 | 29.32 | 28.56 | 26.55 | 25.18 | 29.09 | 26.89 | 26.89 | 22.30 |
| DASR [47] | | 34.06 | 34.08 | 33.57 | 32.15 | 30.13 | 29.99 | 28.66 | 28.42 | 28.96 | 28.90 | 28.62 | 28.13 | 27.65 | 27.36 | 26.86 | 25.95 |
| DCLS [34] | | 34.62 | 34.68 | 34.53 | 33.55 | 30.33 | 30.39 | 30.42 | 29.76 | 29.16 | 29.21 | 29.20 | 28.68 | 28.53 | 28.50 | 28.29 | 27.47 |
| CDFormer (Ours) | | **34.79** | **34.85** | **34.61** | **33.73** | **30.73** | **30.70** | **30.60** | **29.94** | **29.34** | **29.36** | **29.30** | **28.79** | **29.20** | **29.01** | **28.68** | **27.86** |
| Kernel Width | | 0 | 1.2 | 2.4 | 3.6 | 0 | 1.2 | 2.4 | 3.6 | 0 | 1.2 | 2.4 | 3.6 | 0 | 1.2 | 2.4 | 3.6 |
| Bicubic | ×4 | 28.42 | 27.30 | 25.12 | 23.40 | 26.00 | 25.24 | 23.83 | 22.57 | 25.96 | 25.42 | 24.20 | 23.15 | 23.14 | 22.68 | 21.62 | 20.65 |
| RCAN [58] | | 32.63 | 30.26 | 26.72 | 24.66 | 28.87 | 27.48 | 24.93 | 23.41 | 27.72 | 26.89 | 25.09 | 23.93 | 26.61 | 24.71 | 22.25 | 20.99 |
| DASR [47] | | 31.99 | 31.92 | 31.75 | 30.59 | 28.50 | 28.45 | 28.28 | 27.45 | 27.51 | 27.52 | 27.43 | 26.83 | 25.82 | 25.69 | 25.44 | 24.66 |
| KDSR [53] | | 32.42 | 32.34 | 32.13 | 31.02 | 28.67 | 28.66 | 28.55 | 27.80 | 27.64 | 27.67 | 27.60 | 26.98 | 26.36 | 26.29 | 26.06 | 25.21 |
| DCLS [34] | | 32.36 | 32.35 | 32.19 | 31.14 | 28.61 | 28.66 | 28.57 | 27.78 | 27.69 | 27.73 | 27.65 | 27.02 | 26.50 | 26.50 | 26.24 | 25.34 |
| CDFormer (Ours) | | **32.69** | **32.65** | **32.24** | **31.33** | **29.00** | **28.99** | **28.75** | **27.93** | **27.86** | **27.86** | **27.76** | **27.12** | **27.21** | **27.06** | **26.66** | **25.72** |


Figure 4. Visual results of Imgs in Urban100, for scale factor 4 and kernel width 1.2. Best marked in **red**.

Note that, some comparison results in the following are blank, since some prior models are extremely large and they cannot run on our GPU NVIDIA RTX 4090 in some scenarios that require higher memory.

### 4.2. Comparison with State-of-the-art Methods

**Isotropic Gaussian Kernels Noise-free Degradation**. We first conduct a quantitative comparison with two diffusion-based SR methods: SR3 [39] that employs DDPM [13] and StableSR [46] that employs LDM [38]. For a fair comparison, we retrain SR3 on blind settings. We also list the start-of-the-art KP method DCLS [34] for reference. As shown in Tab. 1, CDFormer achieves significant improvement in terms of both objective (PSNR and SSIM) and subjective (FID and LPIPS) metrics compared to diffusion-based methods. SR3 and StableSR that adopt pre-trained diffusion models both show expensive costs in time and model size. Our method instead treats the diffusion model as an estimator, therefore CDFormer can address their limitation and is comparable to DCLS.

We next compare our method with state-of-the-art BSR approaches on noise-free degradations with only isotropic

Table 3. Quantitive evaluation by PSNR (↑) on Set14 for ×4 SR with anisotropic Gaussian kernels and noises. Best marked in **bold**.

| method | noise | Blur Kernel | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| DnCNN[57]+RCAN[58] | 0 | 26.44 | 26.22 | 24.48 | 24.23 | 24.29 | 24.19 | 23.9 | 23.42 | 23.01 |
| | 5 | 26.10 | 25.90 | 24.29 | 24.07 | 24.14 | 24.02 | 23.74 | 23.31 | 22.92 |
| | 10 | 25.65 | 25.47 | 24.05 | 23.84 | 23.92 | 23.8 | 23.54 | 23.14 | 22.77 |
| DnCNN[57]+DCLS[34] | 0 | 27.56 | 27.49 | 26.32 | 25.99 | 25.88 | 26.03 | 25.70 | 24.65 | 23.95 |
| | 5 | 26.20 | 26.02 | 24.44 | 24.21 | 24.28 | 24.14 | 23.88 | 23.40 | 22.98 |
| | 10 | 25.47 | 25.33 | 24.06 | 23.87 | 23.91 | 23.79 | 23.58 | 23.16 | 22.78 |
| DASR[47] | 0 | 27.99 | 27.97 | 27.53 | 27.45 | 27.43 | 27.22 | 27.19 | 26.83 | 26.21 |
| | 5 | 27.25 | 27.18 | 26.37 | 26.16 | 26.09 | 25.96 | 25.85 | 25.52 | 25.04 |
| | 10 | 26.57 | 26.51 | 25.64 | 25.47 | 25.43 | 25.31 | 25.16 | 24.80 | 24.43 |
| KDSR[53] | 0 | 28.26 | 28.38 | 27.98 | 27.98 | 27.94 | 27.75 | 27.69 | 27.35 | 26.52 |
| | 5 | 27.56 | 27.55 | 26.67 | 26.49 | 26.44 | 26.35 | 26.19 | 25.78 | 25.25 |
| | 10 | 26.85 | 26.79 | 25.86 | 25.70 | 25.68 | 25.59 | 25.42 | 25.06 | 24.64 |
| CDFormer(Ours) | 0 | **28.63** | **28.68** | **28.29** | **28.31** | **28.21** | **28.07** | **28.04** | **27.57** | **26.97** |
| | 5 | **27.76** | **27.72** | **26.79** | **26.61** | **26.59** | **26.50** | **26.35** | **25.97** | **25.51** |
| | 10 | **27.00** | **26.92** | **25.98** | **25.80** | **25.80** | **25.70** | **25.55** | **25.22** | **24.83** |

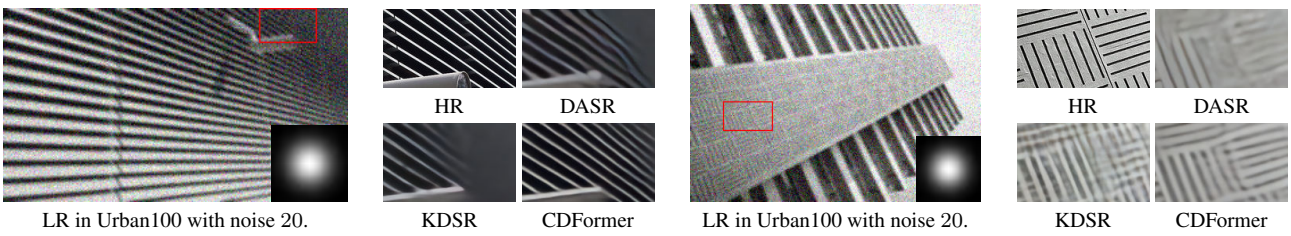



Figure 5. Visualization of SR results via different DP methods on anisotropic Gaussian kernels and noises.

Gaussian kernels under blind settings, including the DP methods DASR [47] and KDSR [53], KP method DCLS [34], and a non-blind SR method RCAN [58] for reference. As shown in Tab. 2, the quantitative results indicate that CDFormer can surpass existing BSR methods. Specifically, CDFormer outperforms the state-of-the-art method DCLS by up to 0.71 dB on Urban100 at a scale of $\times 4$ in all blind settings. However, DCLS relies heavily on accurate kernel estimation, and thus risks significant performance drops in the case of inaccurate estimation or complex degradations. Among DP methods, either DASR or KDSR may lack stability in degradation extraction. In contrast, CDFormer which estimates both content and degradation representations, can reconstruct SR results with sharper and more harmonious textures, evident from the qualitative results in Fig. 4, demonstrating the efficiency of the CDP in properly guiding CDFormer and enabling robustness.

**General Degradation with Anisotropic Gaussian Kernels and Noise**. To better evaluate our method in the setting of complex degradations, we follow DASR [47] that uses 9 blur kernels and different noise levels. For a fair comparison, we use DnCNN [57] as a denoising module to preprocess images for some methods that inherently lack the ability to perform denoising (RCAN [58] and DCLS [34]). The quantitative results are listed in Tab. 3. Compared to KDSR, CDFormer achieves a PSNR improvement ranging from 0.1 dB to 0.5 dB over all settings. Notice that our method achieves the most significant PSNR improvement when the noise level is set to 0. However, improvement is limited as the noise level increases. We attribute this phenomenon to the limitations of SR networks, when given severely damaged LR images, having too minimal information to reconstruct ideal SR images. Nevertheless, CDFormer incorporating CDP performs better texture reconstruction in most settings. The qualitative comparison in Fig. 5 further indicates that CDFormer with the guidance of both content and degradation representations can generate more accurate and clear textures. More results are provided in the supplementary material.

Table 4. Ablation study of CDP on Set5. Best in **blod**.

| Method | Degradation | Content | kernel width=1.2 | | kernel width=2.4 | |
|---|---|---|---|---|---|---|
| | | | PSNR ↑ | SSIM ↑ | PSNR ↑ | SSIM ↑ |
| model1 | ✘ | ✘ | 32.074 | 0.8922 | 32.006 | 0.8894 |
| model2 | ✔ | ✘ | 32.372 | 0.8958 | 32.167 | 0.8902 |
| model3 | ✘ | ✔ | 32.451 | 0.8959 | 32.287 | 0.8909 |
| model4 (Ours) | ✔ | ✔ | **32.564** | **0.8981** | **32.393** | **0.8926** |

## 4.3. Ablation Study

In this section, we will investigate several key components of CDFormer. All the following experiments are conducted on Set5 [2] dataset with scale factor 4. We also provide ablation experiments for iterations number $T$ and CDRB in the supplementary material.

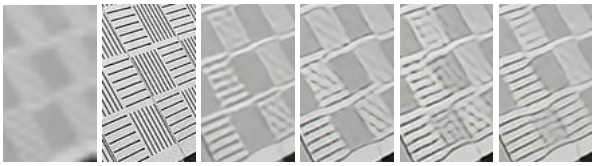
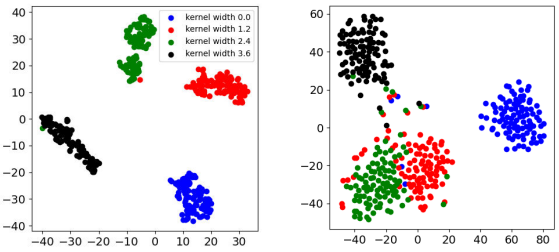



Figure 6. Visualization of ablation study for CDP.




(a) CDFormer (b) DASR[47]

Figure 7. T-SNE results of CDFormer and DASR.

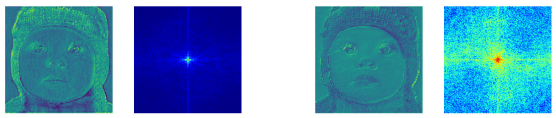
**Effects of Content Degradation Prior (CDP).** We first perform an ablation study on the proposed CDP under the lightweight settings. As listed in Tab. 4, compared to model1 with no prior information, model2 with degradation prior, and model3 with content prior, our method (model4) with the guidance of both content and degradation representations can achieve the best performance. Fig. 6 demonstrates the improvement of CDP visually, further proving the effectiveness of CDP. Our baseline model is able to reconstruct SR images with more accurate textures and fewer distortions, highlighting the effects of CDP on preserving image content and regulating structural textures.

To compare the learned representations between previous DP methods and our proposed CDFormer, we utilize T-SNE [45] to visualize representations in Fig. 7. Specifically, we use 100 images with 4 different degradation kernels to generate LR images. The visualization comparison indicates that CDFormer exhibits superior clustering results. It is notable that DASR employs the contrastive learning mechanism to push away the different degradations and pull close the same degradations. In contrast, CDFormer without any explicit estimation is able to distinguish different degradations, which is also verified by LAM [11] visualization provided in the supplementary material. Furthermore, we utilize the Fourier transform to visualize and analyze the effect of CDP. The results in Fig. 8 explain that CDP can strengthen low-frequency components. This further confirms that features with CDP contain a greater abundance of low-frequency information, which can be assumed to be local content details, resulting in reconstructed images with more harmonious textures and clear edges.

**Effects of Diffusion Model.** We next conduct an ablation study on the diffusion model. We employ a two-stage strategy that $E_{GT}$ trained in the first stage will supervise the training of $E_{LR}$ and the diffusion model in the second stage. Notice that the content details in $E_{GT}$ are extracted from HR images, while CDFormer in the second stage utilizes the LR images only and recreates CDP via the diffusion process. As demonstrated in Tab. 5, diffusion model with the conditional vector from LR images is capable of generating reasonable content and degradation representations, even surpassing the ground truth model. This observation indicates the powerful ability of the diffusion model to capture the data distribution and also confirms the superiority of our method.

Table 5. Ablation study of diffusion model on Urban100.

| Method | kernel width=0 | | kernel width=1.2 | |
|---|---|---|---|---|
| | PSNR ↑ | SSIM ↑ | PSNR ↑ | SSIM ↑ |
| CDFormer (Stage1) | 26.945 | 0.8121 | 26.923 | 0.8096 |
| CDFormer (Stage2) | **27.210** | **0.8189** | **27.058** | **0.8116** |



(a) w/o CDP (b) w/ CDP

Figure 8. Visualization for feature maps and Fourier plots of CDP.

## 5. Conclusion

In this paper, we propose a novel Blind image Super-Resolution method dubbed CDFormer. The distinctiveness of our proposed method lies in the idea of introducing the concept of Content Degradation Prior (CDP) to provide rich low- and high-frequency information for reconstruction. The diffusion model is cleverly used as an estimator to recreate the CDP from the LR images. Since the vector to be learned is one-dimensional, our denoising process is more efficient in sampling and computation. We redesign an adaptive SR network to take full advantage of CDP via injection modules as well as the interflow mechanism to enhance feature representation. Various experiments show that CDFormer not only achieves improved performance with more accurate texture reconstruction in complex degradation scenarios, but also confirms the diffusion model with infinite possibilities in super-resolution.

## Acknowledgement

This work was partially supported by the National Natural Science Foundation of China (No. 62272227 & No. 62276129), and the Natural Science Foundation of Jiangsu Province (No. BK20220890).

## References


[1] Sefi Bell-Kligler, Assaf Shocher, and Michal Irani. Blind super-resolution kernel estimation using an internal-gan. In *NeurIPS 2019*. 1, 2

[2] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie-Line Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In *BMVC 2012,*. 5, 7

[3] Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao. Pre-trained image processing transformer. In *CVPR 2021*, . 2

[4] Zheng Chen, Yulun Zhang, Jinjin Gu, Linghe Kong, Xiaokang Yang, and Fisher Yu. Dual aggregation transformer for image super-resolution. In *CVPR 2023*, . 2, 3

[5] Hyungjin Chung, Byeongsu Sim, and Jong Chul Ye. Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction. In *CVPR 2022*. 2

[6] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. Second-order attention network for single image super-resolution. In *CVPR 2019*. 1, 2

[7] Prafulla Dhariwal and Alexander Quinn Nichol. Diffusion models beat gans on image synthesis. In *NeurIPS 2021*. 2

[8] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *ECCV 2014*, . 1, 2

[9] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In *ECCV 2016*, .

[10] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, 38(2):295–307, 2016. 1

[11] Jinjin Gu and Chao Dong. Interpreting super-resolution networks with local attribution maps. In *CVPR 2021*. 8

[12] Jinjin Gu, Hannan Lu, Wangmeng Zuo, and Chao Dong. Blind super-resolution with iterative kernel correction. In *CVPR 2019*. 1, 2

[13] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *NeurIPS 2020*. 2, 3, 6

[14] Jonathan Ho, Chitwan Saharia, William Chan, David J. Fleet, Mohammad Norouzi, and Tim Salimans. Cascaded diffusion models for high fidelity image generation. *J. Mach. Learn. Res.*, 23:47:1–47:33, 2022. 2

[15] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *CVPR 2015*. 5

[16] Shady Abu Hussein, Tom Tirer, and Raja Giryes. Correction filter for single image super-resolution: Robustifying off-the-shelf deep super-resolvers. In *CVPR 2020*. 2

[17] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *NeurIPS 2022*. 2

[18] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Deeply-recursive convolutional network for image super-resolution. In *2016 IEEE Conf. Comput. Vis. Pattern Recog., CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 1637–1645. IEEE Computer Society, 2016. 2

[19] Soo Ye Kim, Hyeonjun Sim, and Munchurl Kim. Koalanet: Blind super-resolution using kernel-oriented adaptive local adjustment. In *CVPR 2021*. 2

[20] Diederik Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. *NeurIPS 2021*. 2

[21] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR 2015*. 5

[22] Xiangtao Kong, Hengyuan Zhao, Yu Qiao, and Chao Dong. Classsr: A general framework to accelerate super-resolution networks by data characteristic. In *CVPR 2021*. 1

[23] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew P. Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *CVPR 2017*. 2

[24] 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*, 479:47–59, 2022. 2, 3

[25] Wenbo Li, Xin Lu, Shengju Qian, and Jiangbo Lu. On efficient transformer-based image pre-training for low-level vision. In *IJCAI 2023*, . 1

[26] Zheyuan Li, Yingqi Liu, Xiangyu Chen, Haoming Cai, Jinjin Gu, Yu Qiao, and Chao Dong. Blueprint separable residual network for efficient image super-resolution. In *CVPR Workshops 2022*, .

[27] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *ICCV Workshops 2021*, . 1, 2

[28] Jingyun Liang, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Mutual affine network for spatially variant kernel estimation in blind image super-resolution. In *ICCV 2021*, . 1, 2

[29] 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*. 1, 5

[30] Ding Liu, Bihan Wen, Yuchen Fan, Chen Change Loy, and Thomas S. Huang. Non-local recurrent network for image restoration. In *NeurIPS 2018*. 2

[31] Qingguo Liu, Pan Gao, Kang Han, Ningzhong Liu, and Wei Xiang. Degradation-aware self-attention based transformer for blind image super-resolution. *CoRR*, abs/2310.04180, 2023. 1

[32] Andreas Lugmayr, Martin Danelljan, Andrés Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *CVPR 2022*. 2

[33] Ziwei Luo, Fredrik K. Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B. Schön. Refusion: Enabling large-size realistic image restoration with latent-space diffusion models. In *CVPR 2023 Workshops*, . 2

[34] Ziwei Luo, Haibin Huang, Lei Yu, Youwei Li, Haoqiang Fan, and Shuaicheng Liu. Deep constrained least squares for blind image super-resolution. In *CVPR 2022*, . 1, 2, 6, 7

[35] David R. Martin, Charless C. Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *ICCV 2001*. 5

[36] Yiqun Mei, Yuchen Fan, and Yuqian Zhou. Image super-resolution with non-local sparse attention. In *CVPR 2021*. 2

[37] Ben Niu, Weilei Wen, Wenqi Ren, Xiangde Zhang, Lianping Yang, Shuzhen Wang, Kaihao Zhang, Xiaochun Cao, and Haifeng Shen. Single image super-resolution via a holistic attention network. In *ECCV 2020*. 2

[38] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *CVPR 2022*. 2, 3, 6

[39] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J. Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(4):4713–4726, 2023. 2, 3, 6

[40] Jae Woong Soh, Sunwoo Cho, and Nam Ik Cho. Meta-transfer learning for zero-shot super-resolution. In *CVPR 2020*. 2

[41] Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *ICML 2015*. 2

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

[43] Ying Tai, Jian Yang, and Xiaoming Liu. Image super-resolution via deep recursive residual network. In *CVPR 2017*. 2

[44] Radu Timofte, Eirikur Agustsson, and son on. NTIRE 2017 challenge on single image super-resolution: Methods and results. In *CVPR Workshops 2017*. 5

[45] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9 (11), 2008. 8

[46] Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin C. K. Chan, and Chen Change Loy. Exploiting diffusion prior for real-world image super-resolution. *CoRR*, abs/2305.07015, 2023. 1, 2, 3, 6

[47] Longguang Wang, Yingqian Wang, Xiaoyu Dong, Qingyu Xu, Jungang Yang, Wei An, and Yulan Guo. Unsupervised degradation representation learning for blind super-resolution. In *CVPR 2021*, . 1, 2, 6, 7, 8

[48] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *ICCV Workshops 2021*, . 2

[49] Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. In *ICLR 2023*, . 2

[50] 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*, . 2

[51] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4): 600–612, 2004. 5

[52] Bin Xia, Yulun Zhang, Shiyin Wang, Yitong Wang, Xinglong Wu, Yapeng Tian, Wenming Yang, and Luc Van Gool. Diffir: Efficient diffusion model for image restoration. In *ICCV 2023*, . 3

[53] Bin Xia, Yulun Zhang, Yitong Wang, Yapeng Tian, Wenming Yang, Radu Timofte, and Luc Van Gool. Knowledge distillation based degradation estimation for blind super-resolution. In *ICLR 2023*, . 1, 2, 6, 7

[54] Yu-Syuan Xu, Shou-Yao Roy Tseng, Yu Tseng, Hsien-Kai Kuo, and Yi-Min Tsai. Unified dynamic convolutional network for super-resolution with variational degradations. In *CVPR 2020*. 2

[55] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *Curves and Surfaces - 7th International Conference, Avignon, France, June 24-30, 2010, Revised Selected Papers*, pages 711–730. Springer, 2010. 5

[56] Kai Zhang, Luc Van Gool, and Radu Timofte. Deep unfolding network for image super-resolution. In *CVPR 2020*, . 2

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

[58] 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*, . 6, 7