Zero-Shot Inference-Time Rectification for Real-World Arbitrary-Scale Super-Resolution

Yifan Zuo[1], Tianlin Zhu[1], Zhenlong Xia[1], Jiebin Yan[1⋆], Xiaoshui Huang[2], Sanqian Li[1], Yuming Fang[1], and Qiang Wu[3]

[1] School of Computing and Artificial Intelligence, Jiangxi University of Finance and Economics, Nanchang, China kenny0410@126.com, jiebinyan@foxmail.com
[2] School of Public Health, Shanghai Jiao Tong University, Shanghai, China
[3] Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, Australia

Abstract. Arbitrary-scale single image super-resolution (ASSR) has achieved tremendous success under synthetic degradations, but still faces performance deterioration in real-world scenarios where complex distortions are highly entangled with continuous zooming scales. While constructing paired real-world datasets with fractional scales attempts to alleviate this gap, such physical data is fundamentally constrained by discrete scale coverage and hardware-related biases. To overcome these limitations, we propose the Continuous Degradation Rectifier (CDR), a plug-and-play generative proxy that achieves zero-shot domain adaptation at inference time without requiring any real-world paired training data. Designed for any off-the-shelf ASSR model pre-trained on synthetic data, CDR leverages a conditional diffusion process to dynamically project scale-entangled real-world inputs into a canonical bicubic latent space, rectifying the continuous degradation manifold. Specifically, CDR achieves this through a progressive orthogonal decoupling mechanism: it first purifies scale-invariant content from real-world low-resolution images via a discrete codebook bottleneck, and subsequently extracts scale-independent degradation and continuous scale modulation from unpaired bicubically downsampled reference image, which are orthogonally trained by scale-equivariant contrastive learning. Finally, these orthogonal representations are seamlessly integrated via scale-adaptive bandwidth modulation to guide the reverse diffusion process. Extensive experiments on real-world benchmarks demonstrate that equipping frozen, off-the-shelf ASSR models with CDR not only bypasses the need for domain-specific retraining but also establishes a new state-of-the-art.

Keywords: Arbitrary-Scale Super-Resolution · Zero-Shot Adaptation · Real-World Super-Resolution · Conditional Diffusion

1 Introduction

Single image super-resolution (SISR) aims to reconstruct high-resolution (HR) images from low-resolution (LR) observations [8, 21, 22, 46]. Recently, Arbitrary-

⋆ Corresponding author: jiebinyan@foxmail.com

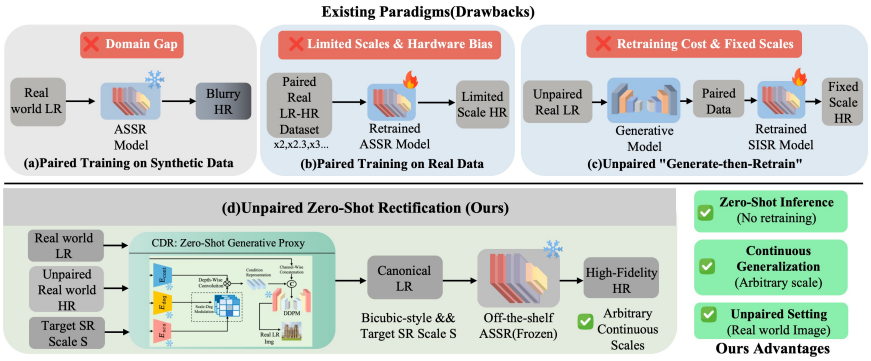

Fig. 1: Comparison of different real-world ASSR/SISR paradigms. (a) suffers from a domain gap when applied to real-world images, (b) is inherently restricted by limited scales and hardware biases, (c) incurs prohibitive computational costs and remains confined to fixed scales, (d) Operating without any real-world paired training data, our Continuous Degradation Rectifier (CDR) acts as a generative proxy for any ASSR model pre-trained on synthetic data. It dynamically projects real-world LR images into a canonical bicubic space, expanding the model scope to real-world distributions without any fine-tuning.

Scale Super-Resolution (ASSR) methods utilizing implicit neural representations (INRs) have achieved tremendous success by modeling images as continuous functions, allowing pixel values to be queried at any continuous spatial coordinates [4, 6, 7, 14, 16, 17, 19]. However, these advanced ASSR models are primarily trained under the assumption of ideal synthetic degradations (*e.g.*, bicubic downsampling). When deployed in real-world scenarios, their performance deteriorates drastically due to the severe domain gap induced by complex distortions that are highly entangled with continuous zooming scales [12, 37, 38, 44]. To bridge this gap, several real-world datasets have been constructed, ranging from discrete-scale collections (*e.g.*, City100 [5], RealSR [3], DRealSR [40]) to recent fractional-scale ones (*e.g.*, RealArbiSR [20], COZ [10]). However, these physical data are inherently constrained by limited scale coverage and significant hardware-related biases. Alternatively, unpaired SISR paradigms have been extensively investigated for real-world SISR [2, 25, 43, 45]. By leveraging generative models such as GANs [11, 18, 39, 47] and Diffusion Models [13, 23, 28, 31, 35], these methods synthesize realistic LR images from unpaired HR images, effectively converting the unpaired setting into a paired one for subsequent supervised training [42]. Nevertheless, existing generative models are inherently restricted to *discrete, predefined scales* (*e.g.*, $\times 4$). Naively extending this "generate-then-retrain" paradigm to arbitrary-scale upsampling is highly impractical. Because the target scales are continuously distributed, generating paired data to retrain the ASSR model for infinite fractional scales would incur prohibitively high computational costs.

In this paper, to tackle the highly ill-posed nature of real-world ASSR **without requiring any real-world paired training data**, we propose a paradigm shift from computationally expensive discrete retraining to zero-shot inference-

time adaptation (as illustrated in Fig. 1). We introduce the **Continuous Degradation Rectifier (CDR)**, a plug-and-play generative proxy designed to rectify continuous degradation manifolds. Rather than retraining ASSR models with offline-generated paired data, CDR dynamically projects scale-entangled real-world LR inputs into a canonical bicubic latent space during the inference phase. Specifically, by purifying scale-invariant content from real-world LR observations and recombining it with orthogonalized scale-independent degradation and continuous scale modulation from *unpaired* bicubically downsampled reference images [15], CDR robustly converts real-world inputs into purified, synthetic-style counterparts. These rectified images can then be directly fed into any ASSR model pre-trained on synthetic data to achieve high-fidelity, arbitrary-scale reconstruction without fine-tuning.

Achieving this zero-shot continuous domain alignment requires disentangling highly coupled physical variables [32]. To this end, CDR employs a progressive orthogonal decoupling mechanism. First, unlike existing methods that merely focus on separating content and degradation representations on fixed scales, we repurpose a discrete codebook as a strict domain-conversion bottleneck. This physically filters out high-frequency real-world artifacts and purifies the content representation. By supervising the codebook with clean, bicubic-style targets across random fractional scales, we enforce the codebook to function as a dedicated scale-normalizing anchor that normalizes representations across arbitrary continuous zoom levels. Second, to address the physical reality that degradations evolve dynamically as a camera lens zooms, we elevate standard "Content-Degradation" 2D feature decoupling to a "Content-Degradation-Scale" 3D orthogonal space via a novel scale-equivariant contrastive learning scheme. By geometrically anchoring contrastive triplets on continuous scaling factors, rather than content or specific noise types, we force the network to explicitly track how degradations mutate across different magnifications. This conceptually elegant design successfully orthogonalizes scale-independent degradation and continuous scale modulation. Finally, these orthogonal representations are seamlessly fused via scale-adaptive bandwidth modulation to guide the reverse diffusion process [33]. This mechanism dynamically adjusts the degradation kernel's frequency bandwidth to adapt to the target continuous scale, elegantly mirroring the physical principles of optical zooming.

In summary, our main contributions are threefold. (a) We propose a novel zero-shot real-world ASSR framework that successfully adapts to complex physical degradations at inference time without requiring any real-world paired data. By introducing the Continuous Degradation Rectifier, a universal plug-and-play generative proxy, we seamlessly equip any synthetic-trained ASSR model with real-world continuous-scale generalization capabilities. (b) We introduce a progressive orthogonal decoupling mechanism alongside a scale-adaptive bandwidth modulation strategy. By synergizing a discrete codebook bottleneck with scale-equivariant contrastive learning, we successfully purify scale-invariant content and orthogonalize scale-independent degradation against continuous scale modulation, subsequently integrating them via bandwidth modulation to guide

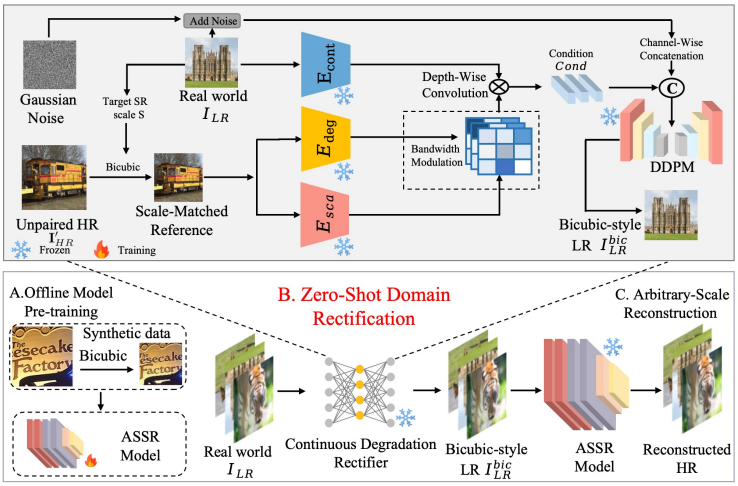

Fig. 2: Inference Phase of CDR. CDR acts as a plug-and-play zero-shot proxy. For a target scale s, it purifies content from the real-world input $\mathbf{I}_{LR}$ and extracts orthogonalized scale-independent degradation and scale modulation from an unpaired HR reference $\mathbf{I}'_{HR}$ bicubically downsampled by s. By fusing these representations, CDR dynamically projects the real-world image into a canonical bicubic latent space $\mathbf{I}_{LR}^{bic}$, ready for the frozen ASSR model.

the reverse diffusion process. (c) Extensive experiments demonstrate that our method establishes a new state-of-the-art (SOTA) across multiple real-world benchmarks, effectively bridging the gap between synthetic degradation modeling and continuous-scale real-world inference while completely bypassing the prohibitive cost of domain-specific ASSR retraining.

2 Related Work

2.1 Implicit Neural Representations for ASSR

Traditional SISR models are fundamentally constrained to predefined integer scaling factors (*e.g.*, $\times 4$) [22, 46]. To break this discrete barrier, ASSR was introduced to infer continuous upsampling scales within a unified framework. Following the pioneering LIIF [7], numerous Implicit Neural Representation (INR) architectures [4, 6, 16, 19] and recent operator-based methods [17, 24] have advanced continuous-scale upsampling. However, these ASSR models exhibit severe performance degradation in real-world scenarios due to their strict reliance on ideal synthetic training degradations (*e.g.*, bicubic downsampling). When confronted with complex physical distortions, their continuous query mechanisms suffer from massive domain shifts, producing pronounced artifacts. While constructing continuous-scale real-world datasets like COZ [10] mitigates this gap, such physical data collection is prohibitively expensive, range-limited, and prone to hardware biases. To bypass these constraints, our work proposes a data-efficient alternative: rectifying degraded inputs directly at inference time.

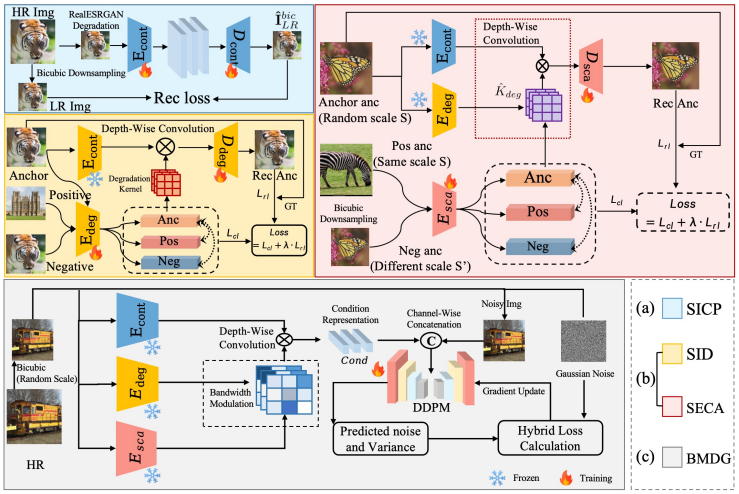

Fig. 3: Training Phase via Progressive Orthogonal Decoupling. The training is divided into sequential stages with a freezing mechanism to prevent feature entanglement. **(a) SICP:** Purifying scale-invariant content via a discrete codebook bottleneck. **(b) SID & SECA:** Orthogonalizing scale-independent degradation and continuous scale modulation via sequential contrastive learning. **(c) BMDG:** Guiding the conditional diffusion process via scale-adaptive bandwidth modulation.

2.2 Real-World SISR and Unpaired Data Generation

To bridge the "Synthetic-Real" domain gap, early methods relied on manual degradation pipelines [9, 27, 38, 41, 44], which inherently lack the capacity to encompass the vast diversity of physical distortions. Consequently, unpaired degradation modeling emerged as a robust alternative, learning degradation distributions directly from unaligned real-world LR and HR images. Methods utilizing GANs [25, 43] and explicit distribution modeling via coupled dictionaries [36] or conditional Diffusion Models [28] effectively synthesize realistic LR-HR pairs, converting the unpaired setting into a paired one for supervised training. However, existing unpaired data generators are structurally confined to *predefined, discrete scaling factors* (*e.g.*, $\times 4$). Because physical optical zooming is continuous, iteratively synthesizing paired data and retraining ASSR models for infinite fractional scales is computationally prohibitive. To bypass this costly "generate-then-retrain" paradigm, our Continuous Degradation Rectifier (CDR) serves as a universal zero-shot inference-time proxy. By **purifying** scale-invariant content and **orthogonalizing** scale-independent degradation against continuous scale modulation, CDR dynamically projects complex physical inputs into an idealized bicubic space. This elegantly unlocks continuous-scale generalization for off-the-shelf, synthetic-trained ASSR models without requiring a single step of fine-tuning.

3 Methodology

3.1 Overall Framework

Given a real-world LR image $\mathbf{I}_{LR}$ and a continuous target scale s, we aim to reconstruct the HR image $\mathbf{I}_{SR}$ without requiring any real-world paired training data. As illustrated in Fig. 2, our approach consists of three sequential stages: **(a) Offline Model Pre-training**, **(b) Zero-Shot Domain Rectification**, and **(c) Arbitrary-Scale Reconstruction**. To bypass the prohibitive cost of continuous domain-specific retraining, we propose the Continuous Degradation Rectifier (CDR), a zero-shot inference-time generative proxy.

Inference Phase: Zero-Shot Domain Rectification. CDR acts as a plug-and-play module for any synthetic-trained ASSR network (*e.g.*, LIIF [7]). Given a target scale s, an arbitrary, unpaired HR reference image $\mathbf{I}'_{HR}$ is bicubically downsampled by s to serve as a scale-matched reference. CDR extracts purified scale-invariant content from $\mathbf{I}_{LR}$ via encoder E_{cont}, and recombines it with orthogonalized scale-independent degradation and continuous scale modulation extracted from the reference via E_{deg} and E_{sca}, respectively. This fusion conditions a reverse diffusion process to synthesize a canonical bicubic-style variant $\mathbf{I}^{bic}_{LR}$, which is seamlessly fed into the frozen ASSR model.

Training Phase: Progressive Orthogonal Decoupling. During pre-training on synthesized complex degradations, empowering CDR with robust domain-conversion capabilities strictly requires disentangling highly coupled physical variables: content, scale-independent degradation, and continuous scale modulation. To prevent conditional posterior collapse where the diffusion model might over-rely on dominant content signals or fixed downsampling scales, we introduce a *progressive orthogonal decoupling mechanism* (Fig. 3). Representations are learned sequentially, with each sub-module strictly frozen upon convergence to guarantee feature orthogonality:

Scale-Invariant Content Purification (SICP): A discrete codebook bottleneck (E_{cont}) is trained to structurally filter out real-world artifacts and extract a pure, scale-invariant content representation.

Contrastive Degradation-Scale Orthogonalization: With E_{cont} frozen, we sequentially train a degradation encoder (E_{deg}) and a scale modulation encoder (E_{sca}) via a novel scale-equivariant contrastive scheme to strictly orthogonalize Scale-Independent Degradation (SID) and continuous scale modulation via Scale-Equivariant Contrastive Anchoring (SECA).

Bandwidth-Modulated Diffusion Guidance (BMDG): With all encoders frozen, we fuse the orthogonal features via scale-adaptive bandwidth modulation to guide the reverse diffusion process, targeting the canonical bicubic distribution.

3.2 Scale-Invariant Content Purification via Discrete Bottleneck

The first step in our progressive orthogonal decoupling is to purify a pristine, scale-invariant content representation $\mathbf{F}_{cont}$ from the scale-entangled input. Although recent degradation modeling methods [28] have explored discrete bottlenecks to separate content from degradations at fixed scales, extending this

paradigm to arbitrary scales introduces a critical challenge: complex distortions are highly entangled with continuous zooming scales. To address this, we tailor a Vector Quantized Variational Autoencoder (VQ-VAE) [30] to act as a strict domain-conversion bottleneck. This physically filters out high-frequency artifacts, actively purifying the content representation to remain strictly invariant to continuous scale modulations. The content extractor comprises an encoder E, a learnable discrete codebook $\mathcal{Z} = \{\mathbf{z}_k\}_{k=1}^{K} \in \mathbb{R}^{K \times d}$ containing K latent embeddings, and a decoder D_{cont}. During pre-training, to simulate scale-entangled real-world complexities, we apply high-order synthetic degradations (*e.g.*, Real-ESRGAN [38]) to HR images across continuous random scales $s \in [1, 4]$, yielding complex degraded LR images $\tilde{\mathbf{I}}_{LR}$. The encoder E first maps $\tilde{\mathbf{I}}_{LR}$ to continuous latent variables, which are subsequently quantized to the nearest codebook entries via a nearest-neighbor lookup:

$$\mathbf{z}_{opt} = \arg\min_{\mathbf{z}_k \in \mathcal{Z}} \|E(\tilde{\mathbf{I}}_{LR}) - \mathbf{z}_k\|_2. \tag{1}$$

For simplicity, we define the composite operation of continuous encoding and discrete quantization as our unified content encoder E_{cont}, denoted as $\mathbf{F}_{cont} = E_{cont}(\tilde{\mathbf{I}}_{LR}) = \mathbf{z}_{opt}$. To enforce the scale-invariant and degradation-free properties of $\mathbf{F}_{cont}$, we supervise the decoder D_{cont} using a canonical bicubic-style variant $\mathbf{I}_{LR}^{bic}$, which is not related to random degradations, to infer the prediction $\hat{\mathbf{I}}_{LR}^{bic}$. Crucially, to provide perfectly aligned spatial supervision, this target $\mathbf{I}_{LR}^{bic}$ is generated by applying clean bicubic downsampling to the corresponding HR image at the exact same random scale s. The reconstruction process from the quantized representation is formulated as:

$$\hat{\mathbf{I}}_{LR}^{bic} = D_{cont}(\mathbf{F}_{cont}). \tag{2}$$

Scale-Invariant Purification Mechanism: While the discrete bottleneck inherently lacks the representational vocabulary to express high-frequency variations introduced by continuous scales [28], our insight lies in explicitly harnessing this property to establish a robust scale-normalizing anchor. Specifically, by supervising the codebook $\mathcal{Z}$ with clean, bicubic-style targets across random continuous scales, we force the codebook to function as a dedicated anchor that normalizes representations across arbitrary zoom levels. The module is optimized using the standard VQ-VAE objective [30]. Once converged, E_{cont} is strictly frozen for all subsequent stages to prevent gradient leakage from scale- and degradation-adaptive modules.

3.3 Contrastive Degradation-Scale Orthogonalization

With the content successfully purified, the subsequent task is to stably project complex degradation manifolds into the canonical bicubic space across arbitrary scales. To strictly orthogonalize scale-independent degradations from continuous scale modulations, we formulate a sequential two-stage training process.

Stage I: Scale-Independent Degradation Representation. With the purified content encoder E_{cont} strictly frozen, we first extract the scale-independent degradation representation. Following the paradigm of RealDGen [28], we employ a degradation encoder E_{deg} and a decoder, training them exclusively anchored at a predefined baseline scale (*e.g.*, $\times 2$). By restricting the scale, E_{deg} is forced to learn scale-independent degradation properties ($\mathbf{F}_{deg}$), effectively isolating fundamental degradation characteristics from continuous scale variations.

Stage II: Scale-Equivariant Contrastive Anchoring. Once E_{deg} converges, we freeze both E_{cont} and E_{deg} to train the scale modulation encoder E_{sca}. To force E_{sca} to perceive the dynamic effect of arbitrary scales on degradation representations, we introduce a novel scale-based contrastive learning strategy. Specifically, let $\mathbf{I}_{HR}^{A}$ and $\mathbf{I}_{HR}^{B}$ denote two distinctly different HR images randomly sampled from the training set. We define our contrastive triplets by applying clean bicubic downsampling as follows:

- **Anchor (anc):** Image $\mathbf{I}_{HR}^{A}$ downsampled at a random continuous scale $s \in [1, 4]$.
- **Positive (pos):** The *different* image $\mathbf{I}_{HR}^{B}$ downsampled at the *same* scale s.
- **Negative (neg):** The *same* image $\mathbf{I}_{HR}^{A}$ downsampled at a *different* random scale $s' \neq s$.

This triplet construction enforces a strict geometric constraint: it compels the network to group samples exclusively by their continuous downsampling scales while explicitly disregarding semantic content. To implement this, E_{sca} takes both the degraded image and the scale factor s as inputs. Inspired by GIDF [48], we employ SwinIR [21] as the backbone. The image is directly fed into the network, while the scale factor is parameterized using the cell representation $[\frac{2}{S_h}, \frac{2}{S_w}]$ from LIIF [7]. The scale representation is then fused with the positional bias in the Transformer blocks, enabling scale-adaptive feature encoding. The module is supervised via the triplet margin loss $\mathcal{L}_{cl}$:

$$\begin{aligned}\mathcal{L}_{cl} = \sum_{i=1}^{n} \max \Big(0, &\|E_{sca}(\mathbf{anc}) - E_{sca}(\mathbf{pos}_i)\|_2^2 \\ &- \|E_{sca}(\mathbf{anc}) - E_{sca}(\mathbf{neg}_i)\|_2^2 + \text{margin}\Big),\end{aligned} \tag{3}$$

where n denotes the number of triplet samples in a mini-batch. Optimized via Eq. 3, the resulting continuous scale modulation feature $\mathbf{F}_{sca}$ becomes highly scale-equivariant while remaining strictly invariant to spatial content.

Representational Completeness and Integration. To guarantee that these decoupled representations collectively capture the complete degradation manifold, we employ a scale-adaptive decoder D_{sca}. It takes the purified content $\mathbf{F}_{cont}$, the scale-independent degradation $\mathbf{F}_{deg}$, and the continuous scale modulation $\mathbf{F}_{sca}$ from the anchor image as inputs; the fusion details are identical to the condition generation described in Sec. 3.4. The entire module is jointly

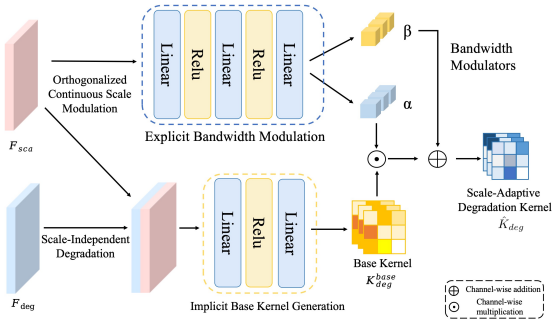

Fig. 4: Architecture of the Bandwidth-Modulated Diffusion Guidance. To simulate the physical effect of optical zooming, the continuous scale modulation $\mathbf{F}_{sca}$ and scale-independent degradation $\mathbf{F}_{deg}$ are jointly utilized to infer a base structural kernel. Concurrently, an affine modulation subnetwork dynamically predicts channel-wise parameters ($\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$) exclusively from $\mathbf{F}_{sca}$ to explicitly modulate the base kernel's frequency bandwidth, yielding the final scale-adaptive degradation kernel.

trained by minimizing the L_2 reconstruction loss on the anchor image:

$$\mathcal{L}_{rl} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\mathbf{I}}_{LR}^{i} - \mathbf{anc}_i\right)^2, \tag{4}$$

where $\hat{\mathbf{I}}_{LR}^{i}$ denotes the predicted reconstruction. Through this sequential learning strategy, we successfully orthogonalize the scale-independent degradation and the continuous scale modulation representations.

3.4 Bandwidth-Modulated Diffusion Guidance

With the latent representations of purified content $\mathbf{F}_{cont}$, scale-independent degradation $\mathbf{F}_{deg}$, and continuous scale modulation $\mathbf{F}_{sca}$ successfully orthogonalized, the final step is to integrate them to guide the conditional diffusion model. While recent degradation modeling methods (*e.g.*, RealDGen [28]) utilize kernel-based feature modulation to fuse content and degradation at fixed scales, they inherently lack the capacity to account for how optical blur dynamically scales across continuous focal lengths. To bridge this gap, as illustrated in Fig. 4, we build upon the fixed-scale modulation paradigm and propose a novel scale-adaptive bandwidth modulation. By explicitly modulating the spatial filter's frequency bandwidth using our orthogonalized scale representation, we elegantly adapt the base degradation to arbitrary continuous scales.

Implicit Base Kernel Generation. First, to capture the joint manifold of the base degradation and the target scale, we concatenate the scale-independent degradation $\mathbf{F}_{deg}$ and the continuous scale modulation $\mathbf{F}_{sca}$. This concatenated feature is fed into a subnetwork consisting of two linear layers to infer a base degradation kernel $\mathbf{K}_{deg}^{base} \in \mathbb{R}^{C\times 3\times 3}$, where C denotes the channel dimension and 3×3 represents the spatial filter size.

Table 1: Performance comparison on RealSR and DRealSR datasets. "+ours" denotes the results of combining the ASSR model pre-trained on synthetic data with our proxy. The best and second-best results are highlighted in **bold** and underlined, respectively. Values in blue represent the **gain** compared to the base ASSR method.

Method	Scale	RealSR			DRealSR		
		PSNR ↑	SSIM ↑	LPIPS ↓	PSNR↑	SSIM ↑	LPIPS ↓
Fixed-Scale Methods							
Real-ESRGAN [38]	×4	24.3974	0.6798	0.3881	26.3510	0.6935	0.3842
BSRGAN [44]	×4	23.2665	0.6043	0.5093	26.1970	0.6869	0.4319
SynDiff [42]	×4	25.2461	0.7588	0.2371	28.0125	0.8136	0.1739
ICF [26]	×4	25.9200	0.7420	0.2845	28.0583	0.8230	0.2295
PiSA [34]	×4	25.5000	0.7417	0.2672	28.3100	0.7804	0.2960
AdaDiffSR [9]	×4	24.1900	0.7485	0.2595	25.6700	**0.8415**	0.2627
OSEDiff [41]	×4	25.1500	0.7341	0.2921	27.9200	0.7835	0.2968
RealDGen [28]	×4	26.1615	**0.7940**	0.2226	28.6629	0.8360	**0.1497**
Arbitrary-Scale Methods (Ours)							
LIIF+ours	×2	34.7984 (+4.5211)	**0.9214** (+0.0496)	0.0523 (-0.1047)	-	-	-
	×3	30.0942 (+2.7988)	0.8614 (+0.0769)	0.1018 (-0.1925)	-	-	-
	×4	26.8051 (+0.8238)	0.7822 (+0.0470)	0.1943 (-0.1810)	28.7324 (+0.6737)	0.7743 (-0.0296)	0.2367 (-0.1036)
CiaoSR+ours	×2	**34.9864** (+4.6858)	0.9194 (+0.0475)	**0.0510** (-0.1061)	-	-	-
	×3	**30.5238** (+3.2225)	**0.8676** (+0.0831)	**0.0946** (-0.2002)	-	-	-
	×4	**26.9161** (+0.9250)	0.7842 (+0.0488)	**0.1910** (-0.1852)	**28.8057** (+0.7349)	0.7728 (-0.0313)	0.2339 (-0.0936)
HIIF+ours	×2	34.7425 (+4.4664)	0.9191 (+0.0477)	0.0520 (-0.1052)	-	-	-
	×3	30.2250 (+2.9467)	0.8633 (+0.0795)	0.1005 (-0.1944)	-	-	-
	×4	26.7693 (+0.7965)	0.7822 (+0.0474)	0.1964 (-0.1798)	28.6953 (+0.6502)	0.7704 (-0.0321)	0.2382 (-0.0972)

Explicit Bandwidth Modulation. While $\mathbf{K}_{deg}^{base}$ establishes the structural prototype of the degradation, it requires dynamic bandwidth adjustment to precisely match the continuous fractional scale. To achieve this explicit modulation, the scale feature $\mathbf{F}_{sca}$ is independently passed through an affine modulation subnetwork composed of three linear layers. This module predicts channel-wise affine parameters: scaling coefficients $\boldsymbol{\alpha} \in \mathbb{R}^C$ and biases $\boldsymbol{\beta} \in \mathbb{R}^C$. These parameters act as bandwidth modulators, applying an affine transformation to the base kernel:

$$\hat{\mathbf{K}}_{deg} = \boldsymbol{\alpha} \odot \mathbf{K}_{deg}^{base} + \boldsymbol{\beta}, \tag{5}$$

where $\odot$ denotes channel-wise multiplication. This operation yields the final scale-adaptive degradation kernel $\hat{\mathbf{K}}_{deg}$, physically mimicking how focal length adjustments stretch or shrink optical blur kernels.

Conditional Diffusion Guidance. Having synthesized the dynamic scale-adaptive degradation kernel, we explicitly embed these degradation characteristics into the purified content feature. Specifically, we perform a depth-wise convolution ($\otimes$) using $\hat{\mathbf{K}}_{deg}$ over the content representation $\mathbf{F}_{cont}$:

$$\mathbf{Cond} = \hat{\mathbf{K}}_{deg} \otimes \mathbf{F}_{cont}, \tag{6}$$

where **Cond** serves as the condition representation for the Denoising Diffusion Probabilistic Model (DDPM). During pre-training, inputs derived from synthesized complex degradations supervise the diffusion model in reconstructing the canonical bicubic distribution. During the zero-shot inference phase, the proposed CDR is directly applied to real-world inputs. Specifically, it fuses the purified content from the real-world $\mathbf{I}_{LR}$ with orthogonalized degradation and scale representations extracted from a scale-matched reference (*i.e.*, an arbitrary,

Table 2: Performance comparison of ASSR on COZ dataset. "+ours" denotes the results of combining the ASSR model pre-trained on synthetic data with our proxy. The best and second-best results are highlighted in **bold** and underlined, respectively. (PSNR↑/SSIM↑/LPIPS↓) in blue represent the **gain** compared to the ASSR model.

Scale	LIIF+ours	CiaoSR+ours	HIIF+ours
×2	**31.7725**/**0.8262**/**0.1292** (+3.9240/+0.0780/-0.0663)	31.5785/0.8211/0.1307 (+3.4634/+0.0684/-0.0623)	31.7361/0.8241/0.1317 (+3.9061/+0.0761/-0.0626)
×2.3	30.6637/0.8279/0.1295 (+3.7804/+0.0640/-0.0686)	30.5092/0.8238/0.1278 (+3.6336/+0.0612/-0.0732)	**30.6929**/**0.8284**/**0.1270** (+3.8539/+0.0658/-0.0753)
×2.7	29.9360/0.8115/0.1614 (+3.7620/+0.0694/-0.1006)	29.8588/0.8093/**0.1570** (+3.6770/+0.0677/-0.1085)	**29.9795**/**0.8128**/0.1586 (+3.8189/+0.0703/-0.1086)
×3	29.9801/0.7878/0.2180 (+3.8876/+0.0928/-0.1132)	29.9378/0.7858/**0.2144** (+3.8413/+0.0905/-0.1197)	**30.0020**/**0.7887**/0.2186 (+3.9084/+0.0935/-0.1137)
×3.5	28.5565/0.7797/0.2239 (+3.6895/+0.0857/-0.1673)	28.5668/0.7797/**0.2177** (+3.7002/+0.0861/-0.1770)	**28.6156**/**0.7823**/0.2217 (+3.7466/+0.0882/-0.1705)
×4	28.6130/0.7615/0.2862 (+3.8943/+0.0950/-0.1637)	28.6329/0.7608/**0.2824** (+3.9104/+0.0942/-0.1703)	**28.6824**/**0.7640**/0.2860 (+3.9622/+0.0974/-0.1646)
×5	27.6456/0.7431/0.3270 (+3.7069/+0.0940/-0.1912)	27.6803/0.7424/**0.3265** (+3.7328/+0.0931/-0.1950)	**27.7463**/**0.7471**/0.3300 (+3.8034/+0.0979/-0.1894)
×6	26.4225/0.7244/**0.3706** (+3.6274/+0.0942/-0.1967)	26.4327/0.7226/0.3753 (+3.6260/+0.0922/-0.1962)	**26.5105**/**0.7280**/0.3782 (+3.7107/+0.0977/-0.1898)

unpaired HR image $\mathbf{I}'_{HR}$ bicubically downsampled by s). This equips the DDPM to dynamically rectify complex physical inputs into the idealized bicubic space.

4 Experiments

4.1 Experimental Setup

Datasets and Metrics. We train our progressive decoupling modules on DIV2-K [1] and pre-train the conditional diffusion model on Flickr2K [29]. Depending on the progressive decoupling stages, training patches are synthesized via high-order degradations across continuous random scales $s \in [1, 4]$ (*e.g.*, for E_{cont} and DDPM), or anchored at a predefined fixed scale to enforce scale-independence (*e.g.*, for E_{deg}). To evaluate generalization across real-world degradation manifolds, we measure PSNR, SSIM, and LPIPS under two settings: fixed-scale (×4) evaluation on RealSR [3] and DRealSR [40], and continuous arbitrary-scale (×2 to ×6) evaluation on the COZ dataset [10].

Implementation Details. We evaluate CDR using three frozen, off-the-shelf ASSR backbones (LIIF [7], CiaoSR [4], HIIF [16]) pre-trained on synthetic data. In our proxy, E_{cont} employs a VQ-VAE bottleneck, while E_{deg} and E_{sca} utilize residual and Swin Transformer blocks, respectively. During zero-shot inference, scale-matched references are generated by bicubically downsampling random DF2K images. The framework is implemented in PyTorch and trained on RTX 3090 GPUs.

4.2 Quantitative Evaluation on Fixed Scales

We evaluate our framework against SOTA fixed-scale real-world SR methods. Unlike baselines requiring domain-specific retraining, our CDR serves as a uni-

versal *zero-shot inference-time proxy*. Despite tackling continuous-scale generalization without ASSR fine-tuning, our method achieves superior performance. As shown in Table 1 (RealSR, $\times 4$), **CiaoSR [4]+ours** records 26.9161 dB (PSNR) and 0.1910 (LPIPS), significantly surpassing classical GANs [38] and the diffusion SOTA, **RealDGen [28]** (26.1615 dB / 0.2226). Similarly, on DRealSR, we establish new PSNR benchmarks (28.8057 dB *vs.* 28.6629 dB for RealDGen). Furthermore, compared to unrectified ASSR baselines, CDR yields substantial gains (blue Δ): consistently reducing LPIPS by 0.10–0.20 and improving PSNR by up to +0.92 dB. This confirms CDR successfully bridges the real-to-bicubic gap, empowering off-the-shelf models to outperform specialized fixed-scale models.

4.3 Continuous-Scale Performance on COZ Dataset

We evaluate continuous-scale real-world generalization on the COZ dataset ($\times 2$ to $\times 6$, Table 2). While off-the-shelf ASSR models (*e.g.*, LIIF [7], CiaoSR [4], HIIF [16]) theoretically support arbitrary scales, they suffer severe degradation on physical inputs due to the synthetic-to-real domain gap. Integrating CDR as a zero-shot inference-time proxy consistently yields massive PSNR improvements of **+3.46 to +4.0 dB** across all backbones. For instance, LIIF+ours achieves 28.6130 dB at $\times 4$ (an absolute gain of **+3.89 dB** over the baseline), demonstrating that our scale-adaptive bandwidth modulation effectively shields base models from domain collapse.

Crucially, CDR avoids the traditional distortion-perception trade-off, unlocking simultaneous improvements across all metrics. The robust PSNR gains are consistently accompanied by enhanced structural fidelity (SSIM up to +0.0979) and reduced perceptual artifacts (LPIPS up to -0.2002). Even at extreme out-of-distribution magnifications (*e.g.*, $\times 5$ and $\times 6$), these enhancements remain highly stable. Overall, CDR effectively bridges the domain gap and establishes a new state-of-the-art benchmark for real-world ASSR.

4.4 Qualitative Analysis

Qualitative comparisons in Fig. 5 (RealSR, $\times 4$) and Fig. 6 (COZ, $\times 6$) demonstrate the superior perceptual fidelity of our method. On RealSR, while specialized baselines (*e.g.*, Real-ESRGAN [38], SynDiff [42], RealDGen [28]) struggle with over-smoothing, granular noise, or stroke fractures, our CDR+LIIF cleanly recovers continuous fabric textures and sharp text boundaries without ringing artifacts. Furthermore, under extreme $\times 6$ magnification on COZ, original ASSR backbones suffer severe degradation, producing heavily smeared architectural textures and thickened stroke fusions. Equipping these backbones with our CDR proxy explicitly recovers high-frequency details, reconstructing distinct, aliasing-free building facades and crisp text boundaries. These consistent improvements confirm that CDR unlocks exceptional structural accuracy and visual cleanliness across diverse real-world scenes.

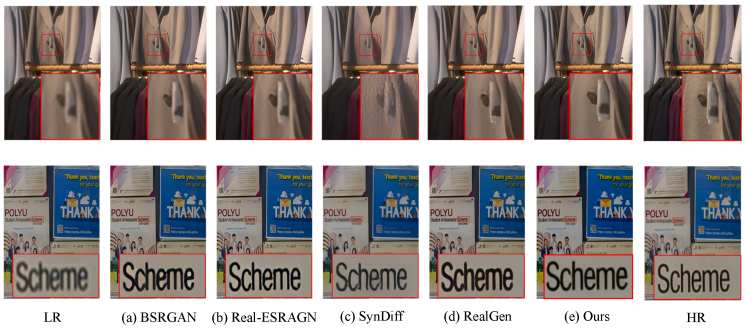

Fig. 5: Visual comparison on fabric textures and printed text (RealSR, $\times 4$). CDR-enhanced models recover clearer weaving patterns and sharper character boundaries compared to specialized real-world SR methods.

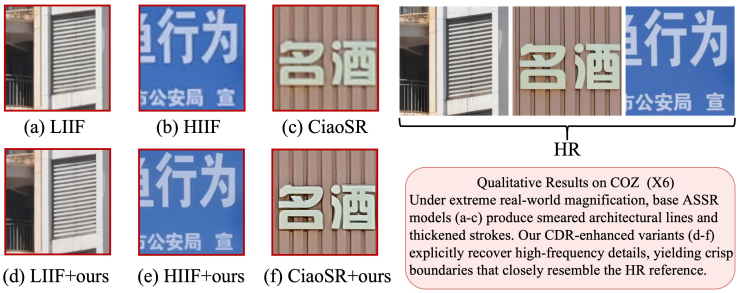

Fig. 6: Qualitative comparison on COZ dataset at an extreme $\times 6$ scale. Our method (d-f) demonstrates consistent improvements in edge sharpness and structural fidelity over ASSR baselines (a-c).

4.5 Ablation Study

Effectiveness of Scale Incorporation The core innovation of CDR lies in disentangling complex real-world degradations from continuous scaling factors. To validate this, we compare our continuous-scale learning scheme against baselines trained specifically on fixed scales (*i.e.*, $\times 2$, $\times 3$, or $\times 4$) without scale modulation (Table 3). We define the "Origin" baseline as directly feeding real-world images into a bicubic-pretrained ASSR model (LIIF). As shown, models trained on fixed scales suffer from severe scale-degradation entanglement. They yield only marginal improvements over the "Origin" baseline even on their targeted in-distribution scales, and fail to generalize to out-of-distribution continuous factors (*e.g.*, $\times 5$, $\times 6$). In contrast, by explicitly injecting continuous scale representations ($\mathbf{F}_{sca}$), our framework robustly models the joint degradation-scale manifold. This architectural design translates into absolute PSNR gains of $+$**3.5 to $+$4.0 dB** across all integer and fractional scales, confirming the necessity of dynamic scale modulation for real-world ASSR.

Design of Bandwidth Modulation Strategy To fuse scale ($\mathbf{F}_{sca}$) and degradation ($\mathbf{F}_{deg}$) representations, we evaluate eight strategies (Table 4). Naive arith-

Table 3: Ablation study on scale conditioning evaluated on COZ dataset (PSNR). "Origin" denotes direct inference using LIIF without our CDR proxy.

Training Setting	In-distribution Scale						Out-of-distribution Scale	
	×2	×2.3	×2.7	×3	×3.5	×4	×5	×6
Origin	27.8485	26.8833	26.1740	26.0925	24.8670	24.7187	23.9387	22.7951
Train on fixed ×2	28.2575	27.0387	26.2721	26.0842	24.9619	24.8292	23.9307	22.8876
Train on fixed ×3	28.2483	27.0151	26.3334	26.0892	25.0535	24.8364	23.9349	22.8996
Train on fixed ×4	28.2358	26.9972	26.3320	26.3711	25.0572	24.8576	23.9399	22.9156
Ours (Continuous Scale)	**31.7725**	**30.6637**	**29.9360**	**29.9801**	**28.5565**	**28.6130**	**27.6456**	**26.4225**

Table 4: Ablation study on the Scale-adaptive Bandwidth Modulation strategy evaluated on COZ dataset (PSNR).

Fusion Strategy	×2	×3	×4
Feature Addition	28.0059	26.3641	26.3572
Feature Concatenation	27.7121	26.8999	26.1973
Global Mod. on $\mathbf{F}_{deg}$	28.0163	26.5875	26.3726
Global Mod. on $\mathbf{K}_{d}$	27.9845	26.8901	26.4143
Global Mod. on $\mathbf{K}_{dc}$	28.3928	26.9515	26.4270
Channel-wise Mod. on $\mathbf{F}_{deg}$	28.1277	26.9104	26.4425
Channel-wise Mod. on $\mathbf{K}_{d}$	28.2760	27.0157	26.5597
Channel-wise Mod. on $\mathbf{K}_{dc}$ (Ours)	**31.7725**	**29.9801**	**28.6130**

metic fusions restrict expressiveness, yielding sub-optimal performance. Conversely, modulating spatial kernels rather than latent features yields superior results by aligning directly with the physical nature of optical blur. Furthermore, explicitly injecting $\mathbf{F}_{sca}$ into kernel generation ($\mathbf{K}_{dc}$) achieves stronger scale-awareness than relying solely on $\mathbf{F}_{deg}$ ($\mathbf{K}_{d}$). However, global modulation shares parameters across channels, ignoring diverse degradation distributions. Thus, our Scale-adaptive Bandwidth Modulation applies *channel-wise* affine modulation to $\mathbf{K}_{dc}$. This strategy achieves 31.7725 dB at ×2, delivering a substantial **+3.38 dB** PSNR gain over the best global variant (28.3928 dB), with consistent improvements at ×3 and ×4. These results confirm that channel-specific bandwidth control over scale-aware kernels is essential to accurately simulate continuous magnification.

5 Conclusion

This paper addresses the highly ill-posed challenge of real-world ASSR without requiring any real-world paired training data. We propose the Continuous Degradation Rectifier (CDR), a plug-and-play zero-shot inference-time proxy that completely bypasses the prohibitive computational costs of domain-specific model retraining. Designed for any ASSR model pre-trained on synthetic data, CDR dynamically projects complex, scale-entangled physical inputs into a canonical bicubic space during inference. By synergizing a discrete codebook bottleneck with a novel scale-equivariant contrastive learning scheme, CDR effectively purifies scale-invariant content and strictly orthogonalizes scale-independent degradation against continuous scale modulation. These decoupled representations

are subsequently integrated via scale-adaptive bandwidth modulation to guide the reverse diffusion process. Extensive experiments confirm that CDR thoroughly bridges the synthetic-to-real domain gap, establishing a formidable new benchmark for real-world arbitrary-scale super-resolution.

Acknowledgements

This work was partially supported by the National Natural Science Foundation of China under Grant Nos. U24A20220, 62132006, 62271237, and 62461028; the Natural Science Foundation of Jiangxi Province under Grant Nos. 20252BAC230003 and 20242BAB26014; and the Hunan Natural Science Foundation Project under Grant No. 2025JJ50338.

References

1. Agustsson, E., Timofte, R.: Ntire 2017 challenge on single image super-resolution: Dataset and study. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition Workshops (2017)
2. Bulat, A., Yang, J., Tzimiropoulos, G.: To learn image super-resolution, use a gan to learn how to do image degradation first. In: Proceedings of the European Conference on Computer Vision. pp. 185–200 (2018)
3. Cai, J., Zeng, H., Yong, H., Cao, Z., Zhang, L.: Toward real-world single image super-resolution: A new benchmark and a new model. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 3086–3095 (2019)
4. Cao, J., Wang, Q., Xian, Y., Li, Y., Ni, B., Pi, Z., Zhang, K., Zhang, Y., Timofte, R., Van Gool, L.: Ciaosr: Continuous implicit attention-in-attention network for arbitrary-scale image super-resolution. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. pp. 1796–1807 (2023)
5. Chen, C., Xiong, Z., Tian, X., Zha, Z.J., Wu, F.: Camera lens super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (June 2019)
6. Chen, H.W., Xu, Y.S., Hong, M.F., Tsai, Y.M., Kuo, H.K., Lee, C.Y.: Cascaded local implicit transformer for arbitrary-scale super-resolution. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. pp. 18257–18267 (2023)
7. Chen, Y., Liu, S., Wang, X.: Learning continuous image representation with local implicit image function. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. pp. 8628–8638 (2021)
8. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence **38**(2), 295–307 (2016)
9. Fan, Y., Liu, C., Yin, N., Gao, C., Qian, X.: Adadiffsr: Adaptive region-aware dynamic acceleration diffusion model for real-world image super-resolution. In: Proceedings of the European Conference on Computer Vision. pp. 396–413 (2024)
10. Fu, H., Peng, F., Li, X., Li, Y., Wang, X., Ma, H.: Continuous optical zooming: A benchmark for arbitrary-scale image super-resolution in real world. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. pp. 3035–3044 (2024)

11. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks. Communications of the ACM **63**(11), 139–144 (2020)
12. Gu, J., Lu, H., Zuo, W., Dong, C.: Blind super-resolution with iterative kernel correction. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. pp. 1604–1613 (2019)
13. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
14. Hu, X., Mu, H., Zhang, X., Wang, Z., Tan, T., Sun, J.: Meta-sr: A magnification-arbitrary network for super-resolution. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. pp. 1575–1584 (2019)
15. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. pp. 1125–1134 (2017)
16. Jiang, Y., Kwan, H.M., Peng, T., Gao, G., Zhang, F., Zhu, X., Sole, J., Bull, D.: Hiif: Hierarchical encoding based implicit image function for continuous super-resolution. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. pp. 2289–2299 (2025)
17. Kim, J., Kim, T.K.: Arbitrary-scale image generation and upsampling using latent diffusion model and implicit neural decoder. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. pp. 9202–9211 (2024)
18. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. pp. 4681–4690 (2017)
19. Lee, J., Jin, K.H.: Local texture estimator for implicit representation function. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. pp. 1929–1938 (2022)
20. Li, Z., Li, M., Fan, J., Chen, L., Tang, Y., Lu, J., Zhou, J.: Learning dual-level deformable implicit representation for real-world scale arbitrary super-resolution. In: Proceedings of the European Conference on Computer Vision. pp. 352–368 (2024)
21. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: Swinir: Image restoration using swin transformer. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1833–1844 (2021)
22. Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M.: Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition Workshops. pp. 136–144 (2017)
23. Lin, X., He, J., Chen, Z., Lyu, Z., Dai, B., Yu, F., Qiao, Y., Ouyang, W., Dong, C.: Diffbir: Toward blind image restoration with generative diffusion prior. In: Proceedings of the European Conference on Computer Vision. pp. 430–448 (2024)
24. Liu, X., Tang, H.: Difffno: Diffusion fourier neural operator. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. pp. 150–160 (2025)
25. Maeda, S.: Unpaired image super-resolution using pseudo-supervision. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. pp. 291–300 (2020)

26. Neshatavar, R., Yavartanoo, M., Son, S., Lee, K.M.: Icf-srsr: Invertible scale-conditional function for self-supervised real-world single image super-resolution. In: Proceedings of the IEEE Winter Conference on Applications of Computer Vision. pp. 1557–1567 (2024)
27. Peng, J., Luo, X., Fu, J., Liu, D.: Confidence-based iterative generation for real-world image super-resolution. In: Proceedings of the European Conference on Computer Vision. pp. 323–341 (2024)
28. Peng, L., Li, W., Pei, R., Ren, J., Xu, J., Wang, Y., Cao, Y., Zha, Z.J.: Towards realistic data generation for real-world super-resolution. In: International Conference on Learning Representations. vol. 2025, pp. 55077–55101 (2025)
29. Plummer, A.B., Wang, L., Cervantes, M.C., Caicedo, C.J., Hockenmaier, J., Lazebnik, S.: Flickr30k entities: Collecting region-to-phrase correspondences for richer image-to-sentence models. International Journal of Computer Vision pp. 74–93 (2017)
30. Razavi, A., Van den Oord, A., Vinyals, O.: Generating diverse high-fidelity images with vq-vae-2. In: Proceedings of Advances in Neural Information Processing Systems. vol. 32, pp. 1–11 (2019)
31. Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D.J., Norouzi, M.: Image super-resolution via iterative refinement. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(4), 4713–4726 (2023)
32. Shocher, A., Cohen, N., Irani, M.: "zero-shot" super-resolution using deep internal learning. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. pp. 3118–3126 (2018)
33. Su, X., Song, J., Meng, C., Ermon, S.: Dual diffusion implicit bridges for image-to-image translation. In: Proceedings of the International Conference on Learning Representations. pp. 1–12 (2023)
34. Sun, L., Wu, R., Ma, Z., Liu, S., Yi, Q., Zhang, L.: Pixel-level and semantic-level adjustable super-resolution: A dual-lora approach. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2333–2343 (2025)
35. Wang, J., Yue, Z., Zhou, S., Chan, K.C., Loy, C.C.: Exploiting diffusion prior for real-world image super-resolution. International Journal of Computer Vision **132**(12), 5929–5949 (2024)
36. Wang, L., Li, J., Wang, Y., Hu, Q., Guo, Y.: Learning coupled dictionaries from unpaired data for image super-resolution. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. pp. 25712–25721 (2024)
37. Wang, L., Wang, Y., Dong, X., Xu, Q., Yang, J., An, W., Guo, Y.: Unsupervised degradation representation learning for blind super-resolution. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. pp. 10581–10590 (2021)
38. Wang, X., Xie, L., Dong, C., Shan, Y.: Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1905–1914 (2021)
39. Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., Change Loy, C.: Esrgan: Enhanced super-resolution generative adversarial networks. In: Proceedings of the European Conference on Computer Vision Workshops. pp. 1–16 (2018)
40. Wei, P., Xie, Z., Lu, H., Zhan, Z., Ye, Q., Zuo, W., Lin, L.: Component divide-and-conquer for real-world image super-resolution. In: Proceedings of the European Conference on Computer Vision. pp. 101–117 (2020)
41. Wu, R., Sun, L., Ma, Z., Zhang, L.: One-step effective diffusion network for real-world image super-resolution. In: Proceedings of Advances in Neural Information Processing Systems. pp. 92529–92553 (2024)

42. Yang, T., Ren, P., Zhang, L., et al.: Synthesizing realistic image restoration training pairs: A diffusion approach. arXiv preprint arXiv:2303.06994 (2023)
43. Yuan, Y., Liu, S., Zhang, J., Zhang, Y., Dong, C., Lin, L.: Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition Workshops. pp. 701–710 (2018)
44. Zhang, K., Liang, J., Van Gool, L., Timofte, R.: Designing a practical degradation model for deep blind image super-resolution. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 4791–4800 (2021)
45. Zhang, Y., Liu, S., Dong, C., Zhang, X., Yuan, Y.: Multiple cycle-in-cycle generative adversarial networks for unsupervised image super-resolution. IEEE transactions on Image Processing **29**, 1101–1112 (2019)
46. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. In: Proceedings of the European conference on computer vision. pp. 286–301 (2018)
47. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2223–2232 (2017)
48. Zuo, Y., Hu, Y., Xu, Y., Wang, Z., Fang, Y., Yan, J., Jiang, W., Peng, Y., Huang, Y.: Learning guided implicit depth function with scale-aware feature fusion. IEEE Transactions on Image Processing **34**, 3309–3322 (2025)