This ICCV paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Not All Degradations Are Equal: A Targeted Feature Denoising Framework for Generalizable Image Super-Resolution

Hongjun Wang[1], Jiyuan Chen[2], Zhengwei Yin[1], Xuan Song[3]*, Yinqiang Zheng[1]*
[1]The University of Tokyo
[2]The Hong Kong Polytechnic University
[3]Jilin University
songxuan@jlu.edu.cn, yqzheng@ai.u-tokyo.ac.jp

Abstract

Generalizable Image Super-Resolution aims to enhance model generalization capabilities under unknown degradations. To achieve such goal, the models are expected to focus only on image content-related features instead of overfitting degradations. Recently, numerous approaches such as Dropout [19] and Feature Alignment [31] have been proposed to suppress models' natural tendency to overfitting degradations and yields promising results. Nevertheless, these works have assumed that models overfit to all degradation types (e.g., blur, noise, JPEG), while through careful investigations in this paper, we discover that models predominantly overfit to noise, largely attributable to its distinct degradation pattern compared to other degradation types. In this paper, we propose a targeted feature denoising framework, comprising noise detection and denoising modules. Our approach presents a general solution that can be seamlessly integrated with existing super-resolution models without requiring architectural modifications. Our framework demonstrates superior performance compared to previous regularization-based methods across five traditional benchmark and datasets, encompassing both synthetic and real-world scenarios.

1. Introduction

Image Super-Resolution (SR) has achieved remarkable progress in recent years, largely driven by the rapid evolution of deep learning techniques [13, 18, 25, 32]. However, despite these advancements, deploying SR models in practical real-world applications remains highly challenging, which stems primarily from a persistent domain gap between synthetic training pipelines and the complex, diverse degradations encountered in real-world imagery.

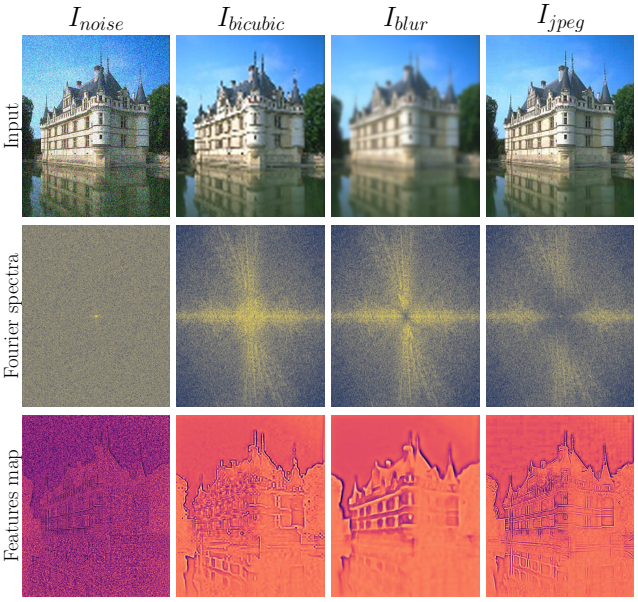
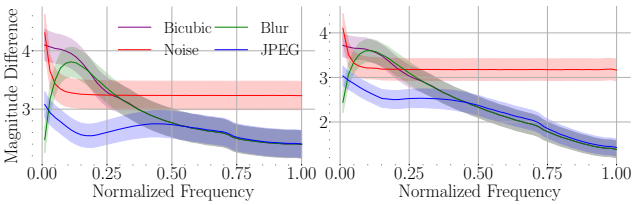


Figure 1. **Visualization of different degradation types, including input images (top), their Fourier spectra of residual images (middle), and SRResNet feature maps (bottom).** Noise (first columns) shows distinct characteristics compared to others.

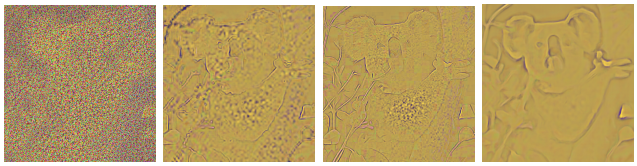
Conventional SR models are typically trained on synthetic low-resolution (LR) images, which are generated using predefined degradations (e.g., bicubic downsampling) [30, 44]. This simplified degradation process, while computationally convenient, is fundamentally misaligned with the intricate degradation patterns observed in real photographs. This misalignment arises from three key factors: (1) Real-world paired LR-HR data collection is non-trivial, requiring precisely aligned captures across different camera systems, lighting conditions, and imaging pipelines [7, 47]; (2) Real-world degradations involve complex interactions between sensor noise, optical aberrations, compression artifacts, and environmental factors, which are difficult to fully character-

*Corresponding Author



(a) DIV2K (b) BSD100

Figure 2. **Frequency-domain analysis of image degradation showing average magnitude differences across normalized frequencies for DIV2K training set (a) and BSD100 test set (b).** The distinctive spectral characteristics of noise degradation contribute to model overfitting issues.



(a) SRResNet (b) +Dropout (c) +Alignment (d) +TFD (ours)

Figure 3. **Visualization of feature representations across different methods. The baseline model (a) shows significant noise corruption.** While dropout [19] and feature alignment [31] reduce noise to some extent, they both retain noticeable artifacts. Our denoising approach effectively preserves structural details while thoroughly suppressing noise artifacts.

ize using simple analytical models [26, 45]; and (3) Existing real-world SR datasets often have limited coverage, capturing only a narrow subset of real degradations, thus lacking the diversity needed for robust generalization [6, 37].

To bridge this gap, recent works have introduced more realistic, high-order degradation modeling pipelines [34, 45, 46], designed to synthesize complex degradation mixtures that better mimic real-world scenarios by mixing several basic degradations, like blur, noise, JPEG. While these efforts represent significant progress, Kong et al. [19] observed that even with improved degradation modeling, existing SR models still tend to overfit to specific degradation patterns, ultimately limiting their generalization potential.

In response, regularization-based strategies such as dropout [19] and feature alignment [31] have been proposed to mitigate degradation overfitting. However, these methods often implicitly assume uniform overfitting across all degradation types, applying generic regularization regardless of the degradation source (e.g., blur, noise, or compression artifacts). Such uniform treatment overlooks the degradation-specific characteristics that drive overfitting. In particular, through detailed analysis, we uncover a critical observation: *noise degradations exhibit uniquely disruptive behaviors that make them the dominant factor contributing to overfitting*.

To systematically analyze degradation impacts, we decompose high-order degradations [34] into individual components and examine their isolated effects on model performance. As shown in Figure 1, noise exhibits random, unstructured spatial patterns, unlike the localized or texture-aware degradations (e.g., blur, JPEG). This irregularity propagates into the feature space, corrupting structural consistency far more severely than other degradations. In Figure 2, noise further amplifies mid-to-high frequency magnitudes [12], while other degradations follow smoother spectral decay. But, in low-frequency regions, due to the contamination of various degradations, it more challenging to overfit to specific degradation. This discrepancy makes noise prone to overfitting, severely impairing the model's ability to learn reliable content features.

Analysis of SRResNet's internal feature maps (Figure 3) reveals that conventional regularization techniques inadequately suppress noise, resulting in persistent artifacts. While the baseline SRResNet exhibits significant noise amplification, established strategies like dropout and feature alignment offer only marginal improvement. This highlights a critical limitation: conventional regularization fails to effectively disentangle noise from content features, ultimately compromising both image fidelity and robustness.

To explicitly address this issue, we propose a targeted feature denoising (TFD) framework, which directly tackles noise-induced overfitting in SR models. Instead of applying uniform constraints across all features, it dynamically detects noise-corrupted features and applies adaptive denoising, preserving content-related structures while suppressing noise artifacts. Our design is: **1)** *noise-aware*: unlike generic regularization, TFD on noise—the primary source of overfitting—and applies degradation-specific treatment. **2)** *model-agnostic*: feature denoising framework is a plug-and-play module compatible with diverse SR architectures [9, 20]. **3)** *lightweight*: The selective denoising mechanism introduces minimal computational overhead. Comprehensive experiments on ten synthetic and real-world benchmarks demonstrate that TFD consistently enhances generalization across diverse degradation scenarios, significantly outperforming state-of-the-art regularization techniques.

2. Related Work

Single Image Super-Resolution. Deep learning has revolutionized single image super-resolution, beginning with SRCNN's pioneering CNN application [13]. Subsequent architectural innovations—including residual learning [18], dense connectivity [49], and attention mechanisms [48]—have progressively improved reconstruction quality. However, these methods primarily excel on synthetic benchmarks with well-defined degradation models, limiting their real-world applicability. Blind super-resolution has emerged to address this limitation by enhancing im-

ages without prior knowledge of degradation processes. Recent approaches pursue two complementary strategies: (1) *Practical degradation modeling*, where Zhang et al. [45] proposed comprehensive degradation pipelines, Wang et al. [33] introduced Real-ESRGAN with carefully designed training data synthesis, and Liang et al. [24] developed degradation-adaptive networks with improved computational efficiency. (2) *Explicit degradation estimation*, exemplified by the IKC framework's iterative kernel refinement [16], Zheng et al.'s unfolded deep learning approach with domain adaptation [50], and Bell-Kligler et al.'s GAN-based kernel estimation for unknown degradations [3].

Generalizable Image Super-Resolution. Achieving robust generalization across diverse degradation conditions remains a fundamental challenge in SISR. Recent advances have explored complementary strategies to address this issue. Component-based approaches include Wei et al.'s degradation-specific subcomponent decomposition [36] and Li et al.'s knowledge transfer from face-specific models [22]. Data-centric methods encompass Chen et al.'s human-guided ground-truth generation [8] and Sahak et al.'s denoising diffusion probabilistic models for severe degradations [29]. Regularization strategies have also proven effective, with Kong et al. [19] introducing dropout as stochastic regularization to prevent degradation-specific overfitting, and Wang et al. [31] developing feature alignment techniques for domain-invariant representations. Recent complementary work has explored random noise injection on feature statistics, data efficiency considerations, and flexible manipulable restoration approaches [40–42]. Our proposed method differs from existing general regularization strategies by explicitly addressing noise overfitting, which we identify as a primary bottleneck in model generalization.

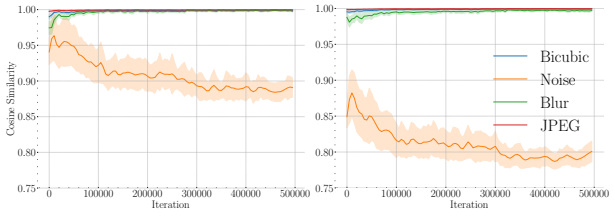
3. Analysis on Feature Representation

We investigate noise overfitting from both empirical and theoretical perspectives, revealing its disruptive effect on content preservation.

Empirical Analysis on Noise Overfitting. To quantify the effect of different degradations on feature retention, we analyze feature similarity throughout training. Specifically, we train SwinIR and SRResNet using RealESRGAN [34] on DIV2K [1] and evaluate on BSD100 [27] under isolated degradations. Given intermediate feature representations of ground-truth and degraded images at training step t, denoted as $h_{gt}^{(t)}$ and $h_{deg}^{(t)}$, we compute cosine similarity:

$$\text{CosSim}(h_{gt}^{(t)}, h_{deg}^{(t)}) = \frac{h_{gt}^{t} \cdot h_{deg}^{(t)}}{\|h_{gt}^{(t)}\| \cdot \|h_{deg}^{(t)}\|}. \tag{1}$$

As shown in Figure 4, noise-induced degradation leads to a sharper and more severe drop in feature similarity compared to blur and JPEG compression. This suggests that noise significantly disrupts content preservation, as noise randomness fundamentally interferes with coherent feature learning. These findings motivate our targeted noise-aware strategy to explicitly mitigate noise-related overfitting.



(a) SwinIR (b) SRResNet

Figure 4. **Feature similarity between degraded and clean images across training on BSD100.** Noise leads to sharper content drift than other degradations, indicating overfitting.

Theoretical Analysis on Noise Overfitting. To explain noise overfitting, we conduct a frequency analysis via 2D Fast Fourier Transform (FFT) denoted as $\mathcal{F}$. By leveraging the linearity property of the Fourier transform [15] and noise additivity [21], we can decompose degraded images as: $\mathcal{F}(I_{deg}) = \mathcal{F}(I_{\text{content}}) + \mathcal{F}(I_{\text{noise}})$. Unlike other degradations that concentrate in low frequencies, noise spans all frequencies, as shown in Figure 2. This spectral spread amplifies overfitting, as explained by the Frequency Principle [38]:

Theorem 3.1 (Frequency Principle) *Let $I_{SR}(t)$ be the network output at training step t and I_{HR} the target high-resolution image. Define $S^k(t) = \mathcal{F}(I_{SR}(t))_{\mathbf{k}}$ and $H^k = \mathcal{F}(I_{HR})_{\mathbf{k}}$ as their respective Fourier coefficients at frequency $\mathbf{k}$. For the relative error $\Delta F(\mathbf{k}, t) = |S^k(t) - H^k|/|H^k|$ and frequencies $\mathbf{k}_1, \mathbf{k}_2$ with $|\mathbf{k}_1| < |\mathbf{k}_2|$*

$$\begin{aligned} &\Delta F(\mathbf{k}_1, t) < \Delta F(\mathbf{k}_2, t), \\ &\left|\frac{d}{dt}\Delta F(\mathbf{k}_1, t)\right| > \left|\frac{d}{dt}\Delta F(\mathbf{k}_2, t)\right|. \end{aligned} \tag{2}$$

This theorem provides crucial insights into the behavior of neural networks during training, which shows that:

- *Networks learn low-frequency components first*: The error for low frequencies ($\Delta F(\mathbf{k}_1, t)$) is consistently smaller than for high frequencies ($\Delta F(\mathbf{k}_2, t)$).
- *Low-frequency learning occurs faster*: The rate of error reduction $|\frac{d}{dt}\Delta F(\mathbf{k}_1, t)|$ is greater for low frequencies than high frequencies.

Theorem 3.1 behavior is reflected in the gradient update: $\frac{d\mathcal{L}}{d\theta} \propto \sum_\omega w(\omega, t)\mathcal{F}(I_{deg}, \omega)$, where $w(\omega, t)$ encodes the frequency-dependent learning dynamics over training time. As training progresses, the effective signal-to-noise ratio (SNR) [5] in gradient updates monotonically decreases:

$$\text{SNR}(t) = \frac{\sum_\omega w(\omega,t)|\mathcal{F}(I_{content},\omega)|^2}{\sum_\omega w(\omega,t)|\mathcal{F}(I_{noise},\omega)|^2}. \quad (3)$$

Since $w(\omega,t)$ shifts emphasis from low to high frequencies with increasing t, and content energy diminishes at higher frequencies while noise remains uniform across the spectrum, we obtain: $\text{SNR}(t+1) < \text{SNR}(t),\ \forall t \geq 0$. This temporal degradation of SNR explains the increasing vulnerability to noise overfitting in later training stages.

However, unlike degradations such as blur or JPEG that selectively impair certain frequency bands, noise spans the full frequency spectrum (see Figure 2). Since noise is semantically meaningless and inherently random, this overfitting directly compromises content fidelity, explaining the severe drop in feature similarity observed in Figure 4.

This spectral entanglement between high-frequency image content and noise poses a unique challenge for conventional learning pipelines. It highlights the necessity for noise-aware training strategies that explicitly disentangle meaningful content from noise across all frequency bands.

4. Targeted Feature Denoising Framework

In this section, we introduce our novel targeted feature denoising framework, designed to explicitly identify and suppress noise-contaminated features while preserving semantically-relevant content representations. The overall pipeline, illustrated in Figure 5, consists of two essential components: 1) a noise detection module, which predicts the likelihood of noise corruption at the feature level, and 2) a frequency-spatial denoising module, which performs selective feature refinement by jointly leveraging complementary frequency and spatial clues.

Noise Detection Module. Our noise detection module builds upon the key observation from Dosselmann and Yang [14] that *noise contamination significantly amplifies high-frequency spectral components, caused by abrupt pixel intensity variations.* To leverage this property, we design a lightweight and adaptable noise detection module f_θ that explicitly captures these spectral signatures. This module is designed to be *plug-and-play*, seamlessly integrating with spatial-domain backbones, while maintaining strong generalizability to unseen noise types due to its spectral grounding. Given an intermediate feature map $h \in \mathbb{R}^{C\times H\times W}$ extracted from the backbone encoder, we first transform h into the frequency domain via the Fourier transform: $\mathcal{F}(h) = F_r + iF_i$ where F_r and F_i represent the real and imaginary components, respectively. To emphasize noise-specific spectral responses, we apply two independent learnable filters $\mathcal{W}_r$ and $\mathcal{W}_i$ to these frequency components:

$$F_r' = \phi(\mathcal{W}_r \circledast F_r), \quad F_i' = \phi(\mathcal{W}_i \circledast F_i) \quad (4)$$

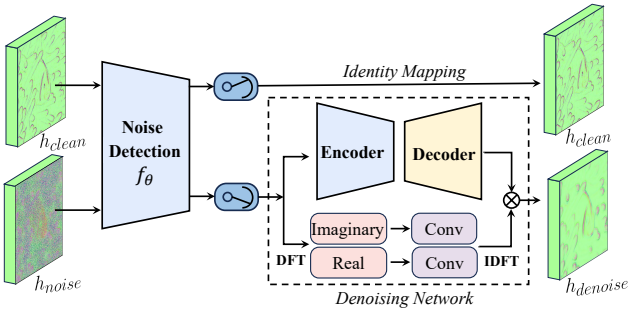


Figure 5. **Overview of the proposed denoising framework.** The noise estimation module (f_θ) first decides whether denoising is required. If not, the features are directly passed through an identity shortcut. Otherwise, the features are refined by a denoising network (dashed box), which integrates spatial processing (encoder-decoder) and frequency processing (DFT-IDFT). $\otimes$ denotes feature fusion.

where ϕ denotes the ReLU activation and $\circledast$ represents convolution, which adaptively amplifies noise-specific spectral signatures while preserving essential structural information.

The frequency domain contains both amplitude and phase information, while direct classification in the frequency domain may focus too much on amplitude and ignore phase. Through Inverse FFT $\mathcal{F}^{-1}$, we allow these two types of information to be reintegrated to provide a more comprehensive feature representation. The feature map is then processed by Inverse FFT, global average pooling Ψ, flattened into a feature vector, and passed through a lightweight convolutional classifier to predict whether the feature is clean or noisy:

$$y = \text{ConvClassifier}(\text{Flatten}(\Psi(\mathcal{F}^{-1}(F_r' + iF_i')))) \in \mathbb{R}^2,$$

where y denotes binary logits indicating the noise state.

Frequency-Spatial Denoising Module. To effectively exploit the complementary strengths of frequency-domain analysis and spatial-domain processing, we propose a dual-path denoising architecture. As illustrated in Figure 1, this design is motivated by the observation that noise exhibits distinct characteristics from other degradations across frequency and spatial domains. By jointly leveraging these two perspectives, our module achieves enhanced noise suppression while preserving semantic content.

Frequency Domain Branch. To improve computational efficiency, this branch reuses the Fourier transform results $\mathcal{F}(h_n)$ already obtained from the noise detection module f_θ. Given these pre-computed frequency components, we introduce two specialized learnable filters $\mathcal{W}_r^n$ and $\mathcal{W}_i^n$, which are applied independently to the real and imaginary parts of the spectral representation. The denoised frequency representation is then transformed back to the spatial do-

main, generating a noise attention mask:

$$h_{\text{freq}} = \sigma\left(\mathcal{F}^{-1}\left(\mathcal{W}_r^n \circledast \mathcal{F}_r + \mathcal{W}_i^n \circledast \mathcal{F}_i\right)\right), \quad (5)$$

where $\circledast$ denotes Hadamard product, and σ is the sigmoid function. The result $h_{\text{freq}} \in [0,1]^{C\times H\times W}$ encodes per-pixel noise likelihood, guiding spatial refinement.

2) Spatial Domain Branch. This branch performs hierarchical reconstruction via an encoder-bottleneck-decoder pipeline:

$$h_{\text{up}} = \mathcal{D}(\Gamma(\mathcal{E}(h_n))), \quad (6)$$

where $\mathcal{E}$ and $\mathcal{D}$ denote the encoder and decoder, and Γ is a non-linear bottleneck operator. Each block contains a residual attention unit defined as:

$$h_{\text{res}} = \mathcal{W}_{1\times1}\left(\text{CA}\left(\mathcal{W}_{3\times3}^{\text{dw}}\left(\mathcal{W}_{1\times1}(\text{LN}(h_n))\right)\right)\right) + h_n, \quad (7)$$

where LN is layer normalization, $\mathcal{W}_{3\times3}^{\text{dw}}$ is a depthwise convolution, and CA is a channel attention module. Skip connections between $\mathcal{E}$ and $\mathcal{D}$ enable signal preservation. The output is fused with input features via:

$$h_{\text{spatial}} = \Lambda([h_{\text{up}}, h_n]), \quad (8)$$

where Λ is a lightweight 1×1 convolution that consolidates scale-aware residuals.

3) Cross-Domain Fusion. We apply the frequency-guided mask to spatial features:

$$h_{\text{denoised}} = h_{\text{freq}} \odot h_{\text{spatial}}, \quad (9)$$

where $\odot$ denotes element-wise multiplication. This operation adaptively scales each channel and pixel according to its noise probability, enhancing restoration accuracy in contaminated regions.

Training Strategy. We use a tri-objective loss to balance fidelity, noise detection, and feature consistency.

1) Reconstruction Loss. The main objective supervises pixel-wise recovery:

$$\mathcal{L}_{\text{rec}} = \|I_{\text{SR}} - I_{\text{HR}}\|_1, \quad (10)$$

where I_{SR} is the super-resolved output and I_{HR} is the clean reference.

2) Noise Classification Loss. The noise classifier f_θ is supervised using cross-entropy:

$$\mathcal{L}_{\text{cls}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=0}^{1} y_{i,c}\log(\hat{y}_{i,c}), \quad (11)$$

with $y_{i,c} \in \{0,1\}$ the ground-truth label and $\hat{y}_{i,c}$ the predicted softmax probability.

3) Feature Consistency Loss. To preserve high-level semantics, we enforce that denoised features remain close to clean references:

$$\mathcal{L}_{\text{feat}} = \|h_{\text{denoised}} - h_{\text{ref}}\|_1, \quad (12)$$

where h_{ref} is extracted from the noise-free image via the same encoder.

Total Loss. The final objective is:

$$\mathcal{L} = \mathcal{L}_{\text{rec}} + \lambda_{\text{cls}}\mathcal{L}_{\text{cls}} + \lambda_{\text{feat}}\mathcal{L}_{\text{feat}}. \quad (13)$$

To prevent early overfitting to uncertain noise estimates, we activate denoising only when the predicted noise confidence surpasses a 0.75 threshold. This dynamic scheduling allows structure-preserving training in early stages, with aggressive suppression deferred to later epochs.

5. Experiments

Datasets. For training, we use the DIV2K dataset [2], which contains 800 high-quality images with multi-degradations settings used in Blind SR [34]. For evaluation, we utilize five standard SR benchmark datasets: Set5 [4], Set14 [39], BSD100 [27], Manga109 [28], and Urban100 [17]. We also evaluate on real-world datasets including DIV2K validation tracks (Difficult, Wild, and Mild) and the real-world DSLR dataset [6] with Canon and Nikon subsets to demonstrate generalization capabilities.

Degradation Modeling and Experimental Protocol. Following previous works [19, 31, 33], we implemented a comprehensive degradation pipeline simulating real-world scenarios. We adopted "second-order" degradation settings with eight configurations: (1) Clean (bicubic downsampling), (2) Blur (Gaussian blur + downsampling), (3) Noise (additive Gaussian noise, $\sigma = 20$, post-downsampling), (4) JPEG (compression quality=30, post-downsampling), and (5)-(8) combinations thereof. This diverse set enables thorough evaluation of model robustness and generalization across realistic distortions.

Quantitative Results. We evaluate TFD through rigorous benchmarking and real-world generalization testing.

Performance on Traditional Benchmark Datasets. Table 1 highlights the performance of TFD when applied to multiple SR models. Notably, on SRResNet [20], TFD achieves significant PSNR gains of 1.86dB, 0.85dB, 0.41dB, and 0.77dB on Set5, Set14, BSD100, and Urban100, respectively. This trend becomes even more pronounced under noise-corrupted settings, where TFD consistently improves performance across all tested architectures, averaging a 0.89dB gain. Although our method specifically targets noise, these results suggest that by preventing the model from overfitting to noise patterns, we free up model capacity that can be redirected toward better handling all degradation types. This confirms our hypothesis that *noise overfitting is a critical bottleneck for generalization in Blind SR - even when noise isn't present in the test image. The consistent improvements across different architectures demonstrate that addressing noise overfitting benefits general SR performance regardless of the underlying model structure.*

Table 1. **Average PSNR and SSIM of different methods in $\times$4 blind SR on five benchmarks with eight types of degradations.**

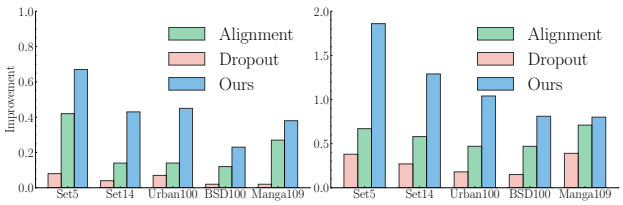
Data	Method	Clean		Blur		Noise		JPEG		Blur+Noise		Blur+JPEG		Noise+JPEG		Blur+Noise+JPEG		Average	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Set5	SRResNet [20]	24.85	0.7295	24.73	0.7113	23.69	0.6422	23.69	0.6714	23.25	0.6142	23.41	0.6534	23.10	0.6508	22.68	0.6249	23.68	0.6622
	+TFD	**26.71**	**0.7295**	**26.20**	**0.7113**	**24.31**	**0.6422**	**24.49**	**0.6714**	**23.39**	**0.6142**	**23.92**	**0.6534**	**23.67**	**0.6508**	**22.99**	**0.6249**	**24.46**	**0.6622**
	RRDB [32]	25.18	0.7344	25.12	0.7176	22.92	0.6317	23.82	0.6739	23.44	0.6183	23.45	0.6542	23.32	0.6548	22.81	0.6279	23.76	0.6641
	+TFD	**26.83**	**0.7773**	**26.59**	**0.7632**	**24.96**	**0.7068**	**24.56**	**0.6974**	**24.07**	**0.6681**	**24.04**	**0.6741**	**23.81**	**0.6753**	**23.14**	**0.6432**	**24.75**	**0.7700**
	HAT [11]	26.40	0.7608	25.84	0.7367	24.86	0.7071	24.54	0.6843	23.93	0.6676	23.95	0.6590	23.84	0.6728	23.15	0.6417	24.56	0.6913
	+TFD	**26.84**	**0.7923**	**26.38**	**0.7753**	**24.98**	**0.7040**	**24.64**	**0.7034**	**24.03**	**0.6680**	**24.05**	**0.6820**	**23.94**	**0.6728**	**23.25**	**0.6417**	**24.76**	**0.7049**
	SwinIR [23]	26.25	0.7498	26.03	0.7317	24.15	0.6484	24.37	0.6816	23.80	0.6249	23.84	0.6601	23.67	0.6609	22.99	0.6315	24.39	0.6736
	+TFD	**26.92**	**0.7989**	**26.43**	**0.7863**	**24.76**	**0.7037**	**24.56**	**0.7095**	**23.90**	**0.6712**	**24.03**	**0.6894**	**23.77**	**0.6744**	**23.09**	**0.6443**	**24.68**	**0.7097**
Set14	SRResNet [20]	23.25	0.6286	23.05	0.6085	22.50	0.5563	22.36	0.5817	22.23	0.5304	22.10	0.5637	22.06	0.5637	21.77	0.5441	22.41	0.5721
	+TFD	**24.54**	**0.6286**	**24.15**	**0.6085**	**23.02**	**0.5563**	**23.11**	**0.5817**	**22.43**	**0.5304**	**22.79**	**0.5637**	**22.60**	**0.5637**	**22.13**	**0.5441**	**23.10**	**0.5721**
	RRDB [32]	23.74	0.6352	23.36	0.6129	22.33	0.5542	22.59	0.5853	22.47	0.5341	22.17	0.5648	22.29	0.5675	21.95	0.5467	22.61	0.5751
	+TFD	**24.72**	**0.6625**	**24.37**	**0.6436**	**23.49**	**0.6044**	**23.27**	**0.5974**	**22.85**	**0.5720**	**22.89**	**0.5765**	**22.75**	**0.5756**	**22.25**	**0.5506**	**23.32**	**0.5978**
	HAT [11]	24.30	0.6390	23.85	0.6145	23.33	0.5959	23.15	0.5885	22.71	0.5640	22.70	0.5662	22.67	0.5697	22.17	0.5450	23.11	0.5854
	+TFD	**24.83**	**0.6777**	**24.44**	**0.6558**	**23.60**	**0.6102**	**23.30**	**0.6014**	**22.83**	**0.5766**	**22.84**	**0.5787**	**22.75**	**0.5789**	**22.23**	**0.5529**	**23.35**	**0.6040**
	SwinIR [23]	24.53	0.6453	24.25	0.6241	23.46	0.5680	23.14	0.5935	22.53	0.5350	22.73	0.5738	22.59	0.5732	22.20	0.5507	23.18	0.5830
	+TFD	**24.96**	**0.6936**	**24.60**	**0.6735**	**23.56**	**0.6044**	**23.23**	**0.6063**	**22.70**	**0.5738**	**22.83**	**0.5833**	**22.69**	**0.5775**	**22.30**	**0.5533**	**23.36**	**0.6082**
BSD100	SRResNet [20]	23.06	0.5943	22.99	0.5755	22.45	0.5160	22.48	0.5526	22.26	0.4905	22.34	0.5365	22.22	0.5343	22.05	0.5158	22.48	0.5394
	+TFD	**23.87**	**0.5943**	**23.71**	**0.5755**	**22.67**	**0.5160**	**22.96**	**0.5526**	**22.36**	**0.4905**	**22.78**	**0.5365**	**22.52**	**0.5343**	**22.29**	**0.5158**	**22.89**	**0.5394**
	RRDB [32]	23.38	0.5994	23.32	0.5803	22.09	0.5119	22.73	0.5563	22.39	0.4926	22.47	0.5382	22.42	0.5371	22.15	0.5175	22.62	0.5417
	+TFD	**24.11**	**0.6191**	**24.05**	**0.6058**	**23.13**	**0.5628**	**23.19**	**0.5637**	**22.74**	**0.5356**	**22.95**	**0.5458**	**22.69**	**0.5436**	**22.41**	**0.5234**	**23.15**	**0.5625**
	HAT [11]	23.78	0.5976	23.55	0.5769	23.03	0.5586	23.07	0.5571	22.65	0.5306	22.82	0.5385	22.65	0.5406	22.35	0.5196	22.99	0.5524
	+TFD	**24.11**	**0.6287**	**23.97**	**0.6137**	**23.15**	**0.5640**	**23.17**	**0.5631**	**22.85**	**0.5368**	**22.91**	**0.5449**	**22.75**	**0.5430**	**22.45**	**0.5218**	**23.17**	**0.5645**
	SwinIR [23]	23.91	0.6062	23.83	0.5870	23.27	0.5253	23.04	0.5610	22.61	0.4950	22.82	0.5432	22.61	0.5397	22.34	0.5207	23.05	0.5473
	+TFD	**24.14**	**0.6415**	**24.06**	**0.6263**	**23.50**	**0.5595**	**23.27**	**0.5684**	**22.84**	**0.5331**	**23.05**	**0.5505**	**22.84**	**0.5420**	**22.57**	**0.5218**	**23.28**	**0.5679**
Urban100	SRResNet [20]	21.24	0.6351	21.06	0.6090	20.82	0.5656	20.60	0.5949	20.46	0.5277	20.30	0.5652	20.43	0.5761	20.10	0.5436	20.63	0.5771
	+TFD	**22.28**	**0.6351**	**21.89**	**0.6090**	**21.20**	**0.5656**	**21.30**	**0.5949**	**20.52**	**0.5277**	**20.84**	**0.5652**	**20.87**	**0.5761**	**20.33**	**0.5436**	**21.15**	**0.5771**
	RRDB [32]	21.57	0.6404	21.18	0.6106	19.61	0.5487	20.93	0.5996	20.57	0.5297	20.40	0.5667	20.74	0.5807	20.24	0.5458	20.66	0.5778
	+TFD	**22.44**	**0.6654**	**22.13**	**0.6441**	**21.66**	**0.6166**	**21.45**	**0.6108**	**20.99**	**0.5755**	**20.93**	**0.5764**	**21.09**	**0.5887**	**20.53**	**0.5521**	**21.40**	**0.6037**
	HAT [11]	22.05	0.6412	21.56	0.6094	21.39	0.6028	21.28	0.5999	20.70	0.5596	20.72	0.5642	20.93	0.5799	20.34	0.5427	21.12	0.5875
	+TFD	**22.58**	**0.6782**	**22.23**	**0.6568**	**21.79**	**0.6284**	**21.55**	**0.6174**	**21.10**	**0.5874**	**20.98**	**0.5837**	**21.18**	**0.5986**	**20.58**	**0.5610**	**21.50**	**0.6139**
	SwinIR [23]	22.18	0.6489	21.90	0.6204	20.56	0.5614	21.32	0.6050	20.89	0.5350	20.79	0.5724	20.98	0.5841	20.45	0.5498	21.13	0.5846
	+TFD	**22.63**	**0.6923**	**22.31**	**0.6725**	**21.61**	**0.6243**	**21.47**	**0.6228**	**20.95**	**0.5829**	**20.93**	**0.5894**	**21.08**	**0.5982**	**20.55**	**0.5618**	**21.44**	**0.6180**
Manga109	SRResNet [20]	18.42	0.6467	18.75	0.6453	18.32	0.5903	18.30	0.6266	18.60	0.5851	18.53	0.6226	18.25	0.6142	18.43	0.6091	18.45	0.6175
	+TFD	**19.22**	**0.6467**	**19.52**	**0.6453**	**18.98**	**0.5903**	**18.96**	**0.6266**	**19.14**	**0.5851**	**19.11**	**0.6226**	**18.83**	**0.6142**	**18.92**	**0.6091**	**19.09**	**0.6175**
	RRDB [32]	18.59	0.6498	18.64	0.6437	18.30	0.5900	18.41	0.6285	18.83	0.5886	18.43	0.6208	18.38	0.6167	18.41	0.6088	18.50	0.6183
	+TFD	**19.28**	**0.6632**	**18.64**	**0.6437**	**19.09**	**0.6352**	**19.05**	**0.6375**	**19.21**	**0.6270**	**19.09**	**0.6322**	**18.84**	**0.6197**	**18.91**	**0.6133**	**19.01**	**0.6340**
	HAT [11]	19.49	0.6666	19.66	0.6608	19.35	0.6481	19.26	0.6444	19.40	0.6367	19.29	0.6348	19.12	0.6337	19.10	0.6237	19.33	0.6436
	+TFD	**19.59**	**0.6778**	**19.76**	**0.6759**	**19.34**	**0.6579**	**19.26**	**0.6524**	**19.50**	**0.6481**	**19.31**	**0.6444**	**19.12**	**0.6399**	**19.10**	**0.6300**	**19.37**	**0.6533**
	SwinIR [23]	19.10	0.6583	19.27	0.6523	18.71	0.5964	18.95	0.6372	19.07	0.5924	19.02	0.6308	18.79	0.6242	18.80	0.6153	18.96	0.6259
	+TFD	**19.20**	**0.6739**	**19.37**	**0.6709**	**19.34**	**0.6690**	**18.91**	**0.6489**	**19.17**	**0.6417**	**19.12**	**0.6414**	**18.89**	**0.6358**	**18.90**	**0.6264**	**19.34**	**0.6690**

Figure 6 further reveals that TFD consistently outperforms existing regularization techniques. *The detailed data of Figure 6 is in supplementary material.* On SwinIR, TFD improves PSNR by 0.67dB on Set5 and 0.45dB on Urban100, with even larger gains observed for SRResNet (up to 1.86dB). Compared to general-purpose regularization such as Dropout or feature alignment, which apply uniform constraints across all features, TFD explicitly disentangles noise from content features. This targeted noise suppression not only preserves fine details but also enhances robustness under compound degradations. Importantly, TFD's architecture-agnostic design consistently benefits both transformer-based models (SwinIR, HAT) and CNN-based models (SRResNet, RRDB), demonstrating its broad applicability.

Generalization to Real-world Degradations. We further evaluate robustness on real-world datasets (Table 2), where TFD consistently outperforms existing methods. On DIV2K variants, TFD improves PSNR by 0.62dB, 0.60dB, and 0.52dB. On Canon and Nikon datasets, TFD achieves remarkable gains of 1.50dB and 1.46dB. Notably, TFD reaches 25.72dB PSNR on Canon, surpassing Dropout (24.86dB) and feature alignment (25.13dB). TFD also improves LPIPS by up to 0.020, demonstrating superior perceptual quality. These results validate TFD's ability to selectively suppress noise, addressing a key bottleneck in generalizable super-resolution.

Ablation Study. We conduct systematic ablations to evaluate the effectiveness, efficiency, and robustness of our framework across three key dimensions.

Component Analysis. Table 3 presents our architectural investigation across diverse network backbones. The full TFD model delivers impressive gains across all degradation scenarios, improving PSNR by 1.86dB (SRResNet) and 0.67dB (SwinIR) on Set5, with particularly strong performance on noise-corrupted Urban100 images (+1.04dB).



(a) SwinIR (b) SRResNet

Figure 6. **PSNR improvement comparison of different enhancement methods across benchmark datasets.**

Table 2. **Comparison of different strategies in real setting.**

Dataset	Method	PSNR ↑	SSIM ↑	LPIPS ↓	MAD ↓	NLPD ↓
DIV2K Difficult [2]	SRResNet	17.83	0.584	0.565	0.954	0.914
	+Dropout	18.02	0.591	0.576	0.933	0.920
	+Alignment	18.07	0.597	0.570	0.933	0.923
	+TFD	**18.45**	**0.615**	**0.560**	**0.909**	**0.910**
DIV2K Wild [2]	SRResNet	17.60	0.582	0.531	0.954	0.903
	+Dropout	17.77	0.590	0.545	0.938	0.908
	+Alignment	17.83	0.595	0.532	0.936	0.911
	+TFD	**18.20**	**0.605**	**0.525**	**0.927**	**0.900**
DIV2K Mild [2]	SRResNet	16.98	0.567	0.522	0.965	0.895
	+Dropout	17.14	0.574	0.536	0.953	0.900
	+Alignment	17.21	0.580	0.522	0.952	0.903
	+TFD	**17.50**	**0.590**	**0.515**	**0.942**	**0.890**
Canon [6]	SRResNet	24.22	0.787	0.327	0.629	0.702
	+Dropout	24.86	0.792	0.332	0.580	0.711
	+Alignment	25.13	0.795	0.328	0.566	0.709
	+TFD	**25.72**	**0.802**	**0.315**	**0.559**	**0.700**
Nikon [6]	SRResNet	23.85	0.754	0.368	0.684	0.712
	+Dropout	24.37	0.759	0.375	0.650	0.715
	+Alignment	24.74	0.762	0.370	0.629	0.720
	+TFD	**25.31**	**0.773**	**0.348**	**0.622**	**0.700**

When analyzing individual components, we observe that both spectral-domain (FD) and spatial-domain (SD) denoising pathways provide complementary benefits, supporting our dual-path design rationale. The noise detection module proves crucial—its removal causes significant performance drops (1.30dB for SRResNet, 0.50dB for SwinIR on Set5). *This substantial decline occurs because the ND module enables adaptive processing through selective feature routing—clean features bypass denoising via identity mapping while corrupted features undergo restoration. Without this discriminative capability, the model applies unnecessary transformations to clean features, compromising generalization to out-of-distribution degradation patterns.*

Model Complexity vs. Performance. Our frequency analysis revealed noise's unique spectral characteristics, but implementing this insight requires efficient design. Table 4 examines this trade-off using SwinIR. The optimal configuration (Full model) adds just 51.60G MACs (+9.0%) while delivering substantial quality improvements (+0.41dB on BSD100, +0.52dB on Urban100, +0.64dB on Manga109). Further increasing complexity (Heavy configuration) yields diminishing returns—a classic sign of overfitting to noise patterns rather than content features. This confirms that our dual-path denoising framework strikes an ideal balance, effectively targeting noise corruption with minimal computational overhead.

Table 3. **Ablation study on Noise Detection (ND), Spatial Denoising (SD), and Frequency Domain (FD) processing.**

Model	TFD Components			Set5		Urban100	
	ND	SD	FD	Clean	Noise	Clean	Noise
SRResNet	✓	✓	✓	**26.71**	**24.31**	**22.28**	**21.20**
	✗	✓	✓	26.32	24.15	21.89	20.96
	✓	✗	✓	26.28	24.08	21.85	20.92
	✓	✓	✗	25.41	23.92	21.43	20.87
	✗	✗	✗	24.85	23.69	21.24	20.82
SwinIR	✓	✓	✓	**26.92**	**24.76**	**22.63**	**21.61**
	✗	✓	✓	26.58	24.52	22.35	21.38
	✓	✗	✓	26.54	24.48	22.31	21.35
	✓	✓	✗	26.42	24.32	22.25	21.15
	✗	✗	✗	26.25	24.15	22.18	20.56

Table 4. **Model complexity vs. performance.**

Model	Parameters & Complexity			PSNR on Test Sets		
	Params (M)	MACs	Increase	BSD100	Urban100	Manga109
SwinIR	10.72	571.77	-	23.05	22.18	19.10
+ Light	11.12	588.91	+3.00%	23.21	22.38	19.23
+ Medium	11.68	606.08	+6.00%	23.24	22.51	19.28
+ Full	12.21	623.37	+9.02%	**23.28**	**22.63**	**19.34**
+ Heavy	12.83	640.71	+12.06%	23.26	22.60	19.31
HAT	12.01	599.78	-	22.99	22.05	19.49
+ Light	12.42	616.92	+2.86%	23.10	22.25	19.53
+ Medium	12.98	634.08	+5.72%	23.14	22.41	19.55
+ Full	13.52	651.38	+8.60%	**23.17**	**22.58**	**19.59**
+ Heavy	14.14	668.71	+11.49%	23.16	22.56	19.57

Hyperparameter Sensitivity. The effectiveness of our noise-content disentanglement depends on balancing classification accuracy and denoising strength. Table 5 shows that $\lambda_{cls} = 0.10$ and $\lambda_{denoise} = 0.01$ consistently achieves optimal results across datasets and architectures, with improvements up to 1.86dB (SRResNet) on Set5. This sweet spot reflects the inherent trade-off between noise suppression and content preservation identified in our theoretical analysis. Higher classification weights overemphasize degradation identification at the expense of reconstruction quality, while excessive denoising weights risk removing content features along with noise. The consistent optimal configuration across network paradigms demonstrates that our approach addresses a fundamental limitation in super-resolution rather than architecture-specific issues.

Qualitative Results. Figure 7 presents qualitative comparisons on challenging textures from BSD100 under the bicubic+noise20 setting. We highlight two representative examples: a pyramid with fine geometric patterns and a landscape with dense foliage—both exemplifying high-frequency details that are particularly vulnerable to noise corruption. As shown, conventional super-resolution models (SRResNet, RRDBNet) visibly suffer from noise overfitting, leading to texture degradation and structural blurring—consistent with our hypothesis on corrupted feature representations. Existing regularization methods offer lim-

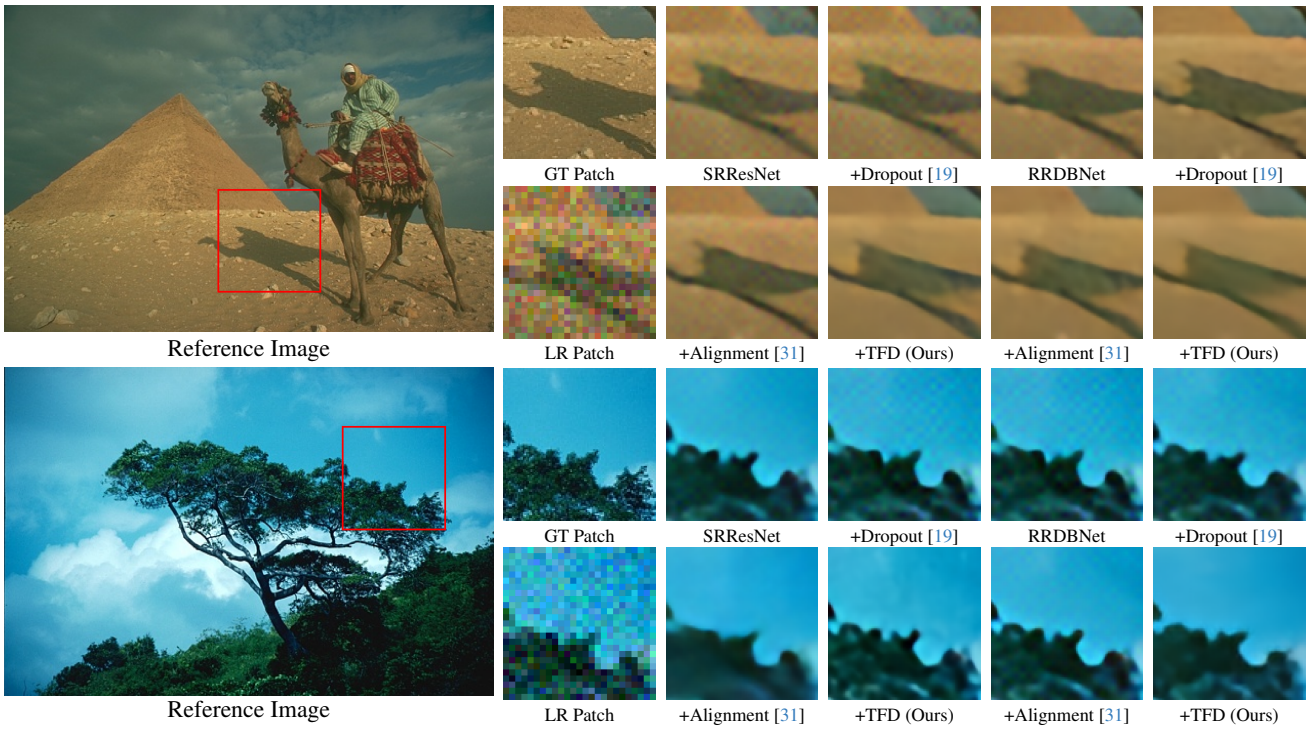


Figure 7. **Visual comparison of different super-resolution methods on BSD100 dataset with bicubic_noise20 degradation.**

Table 5. **Analysis of loss weight hyperparameters on model performance across benchmark datasets.**

Model	Loss Weights		Test Sets (PSNR)				
	λ_{cls}	$\lambda_{denoise}$	Set5	Set14	BSD100	Urban100	Manga109
SRResNet	0.05	0.005	26.43	23.89	22.62	20.92	18.87
	0.10	0.010	**26.71**	**24.10**	**22.89**	**21.15**	**19.09**
	0.20	0.020	26.63	24.06	22.84	21.08	19.01
	0.15	0.005	26.54	23.96	22.76	21.02	18.96
	0.05	0.015	26.48	23.92	22.73	20.95	18.90
SwinIR	0.05	0.005	26.71	24.78	23.08	21.22	19.12
	0.10	0.010	**26.92**	**24.96**	**23.28**	**21.44**	**19.34**
	0.20	0.020	26.84	24.90	23.20	21.39	19.28
	0.15	0.005	26.77	24.85	23.15	21.36	19.23
	0.05	0.015	26.74	24.81	23.11	21.32	19.19

ited improvements: Dropout reduces noise at the expense of oversmoothing, while feature alignment preserves structures but retains noise residues. These observations align with our frequency analysis, confirming the spectral overlap between noise and high-frequency content as a core challenge. In contrast, our proposed TFD framework achieves significantly cleaner and sharper reconstructions. The pyramid's edges are crisp and well-defined, while the foliage retains natural texture without residual noise. This highlights the effectiveness of our dual-path design: the frequency branch isolates noise-corrupted components, while the spatial branch preserves semantic content, enabling perceptual quality that aligns closely with our quantitative gains.

6. Conclusion

This paper addresses generalizable image super-resolution by identifying noise as the primary source of feature corruption limiting model generalization. Through frequency analysis, we demonstrated noise's distinctive spectral characteristics compared to other degradations. We proposed a lightweight feature denoising framework comprising noise detection and dual-path denoising modules that selectively suppress noise-related features while preserving content details. Our model-agnostic framework integrates seamlessly with various SR architectures without structural modifications and demonstrates effectiveness on real-world images with complex degradations.

Acknowledgment

This work was partially supported by the Research Institute of Trustworthy Autonomous Systems at Southern University of Science and Technology, grants from the Jilin Provincial International Cooperation Key Laboratory for Super Smart City and the Jilin Provincial Key Laboratory of Intelligent Policing, the SPRING GX project of the University of Tokyo (grant number JPMJSP2108), JST-Mirai Program JPMJMI23G1, and JSPS KAKENHI Grant Numbers 24KK0209, 24K22318, and 22H00529.

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2017. 3

[2] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 126–135, 2017. 5, 7, 1

[3] Sefi Bell-Kligler, Assaf Shocher, and Michal Irani. Blind super-resolution kernel estimation using an internal-gan. *Advances in Neural Information Processing Systems*, 32, 2019. 3

[4] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie Line Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. 2012. 5, 1

[5] George Box. Signal-to-noise ratios, performance criteria, and transformations. *Technometrics*, 30(1):1–17, 1988. 3

[6] Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3086–3095, 2019. 2, 5, 7, 1

[7] Chang Chen, Zhiwei Xiong, Xinmei Tian, Zheng-Jun Zha, and Feng Wu. Camera lens super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1652–1660, 2019. 1

[8] Du Chen, Jie Liang, Xindong Zhang, Ming Liu, Hui Zeng, and Lei Zhang. Human guided ground-truth generation for realistic image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14082–14091, 2023. 3

[9] Haoyu Chen, Jinjin Gu, and Zhi Zhang. Attention in attention network for image super-resolution. *arXiv preprint arXiv:2104.09497*, 2021. 2

[10] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European conference on computer vision*, pages 17–33. Springer, 2022.

[11] Xiangyu Chen, Xintao Wang, Jiantao Zhou, Yu Qiao, and Chao Dong. Activating more pixels in image super-resolution transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 22367–22377, 2023. 6

[12] Xiangyu Chen, Zheyuan Li, Yuandong Pu, Yihao Liu, Jiantao Zhou, Yu Qiao, and Chao Dong. A comparative study of image restoration networks for general backbone network design. In *European Conference on Computer Vision*, pages 74–91. Springer, 2024. 2

[13] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015. 1, 2

[14] Richard W Dosselmann and Xue Dong Yang. No-reference noise and blur detection via the fourier transform. *Dept. of Computer Science, Univ. of Regina. Regina, SK (Saskatchewan), Canada*, 2012. 4

[15] Jack D Gaskill. *Linear systems, Fourier transforms, and optics*. John Wiley & Sons, 1978. 3

[16] Jinjin Gu, Hannan Lu, Wangmeng Zuo, and Chao Dong. Blind super-resolution with iterative kernel correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1604–1613, 2019. 3

[17] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5197–5206, 2015. 5, 1

[18] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016. 1, 2

[19] Xiangtao Kong, Xina Liu, Jinjin Gu, Yu Qiao, and Chao Dong. Reflash dropout in image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6002–6012, 2022. 1, 2, 3, 5, 8

[20] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017. 2, 5, 6, 4

[21] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. *arXiv preprint arXiv:1803.04189*, 2018. 3

[22] Xiaoming Li, Chaofeng Chen, Xianhui Lin, Wangmeng Zuo, and Lei Zhang. From face to natural image: Learning real degradation for blind image super-resolution. In *European Conference on Computer Vision*, pages 376–392. Springer, 2022. 3

[23] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *IEEE International Conference on Computer Vision Workshops*, 2021. 6, 1, 4

[24] Jie Liang, Hui Zeng, and Lei Zhang. Efficient and degradation-adaptive network for real-world image super-resolution. In *European Conference on Computer Vision*, pages 574–591. Springer, 2022. 3

[25] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2017. 1

[26] Anran Liu, Yihao Liu, Jinjin Gu, Yu Qiao, and Chao Dong. Blind image super-resolution: A survey and beyond. *IEEE transactions on pattern analysis and machine intelligence*, 45(5):5461–5480, 2022. 2

[27] David Martin, Charless 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 *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, pages 416–423. IEEE, 2001. 3, 5, 1

[28] Yusuke Matsui, Kota Ito, Yuji Aramaki, Azuma Fujimoto, Toru Ogawa, Toshihiko Yamasaki, and Kiyoharu Aizawa. Sketch-based manga retrieval using manga109 dataset. *Multimedia Tools and Applications*, 76(20):21811–21838, 2017. 5, 1

[29] Hshmat Sahak, Daniel Watson, Chitwan Saharia, and David Fleet. Denoising diffusion probabilistic models for robust image super-resolution in the wild. *arXiv preprint arXiv:2302.07864*, 2023. 3

[30] Radu Timofte, Eirikur Agustsson, Luc Van Gool, Ming-Hsuan Yang, and Lei Zhang. Ntire 2017 challenge on single image super-resolution: Methods and results. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 114–125, 2017. 1

[31] Hongjun Wang, Jiyuan Chen, Yinqiang Zheng, and Tieyong Zeng. Navigating beyond dropout: An intriguing solution towards generalizable image super resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25532–25543, 2024. 1, 2, 3, 5, 8

[32] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 0–0, 2018. 1, 6, 4

[33] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1905–1914, 2021. 3, 5, 1

[34] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *International Conference on Computer Vision Workshops (ICCVW)*, 2021. 2, 3, 5

[35] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17683–17693, 2022.

[36] Pengxu Wei, Ziwei Xie, Hannan Lu, Zongyuan Zhan, Qixiang Ye, Wangmeng Zuo, and Liang Lin. Component divide-and-conquer for real-world image super-resolution. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VIII 16*, pages 101–117. Springer, 2020. 3

[37] Yunxuan Wei, Shuhang Gu, Yawei Li, Radu Timofte, Longcun Jin, and Hengjie Song. Unsupervised real-world image super resolution via domain-distance aware training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13385–13394, 2021. 2

[38] Zhi-Qin John Xu, Yaoyu Zhang, Tao Luo, Yanyang Xiao, and Zheng Ma. Frequency principle: Fourier analysis sheds light on deep neural networks. *arXiv preprint arXiv:1901.06523*, 2019. 3

[39] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma. Image super-resolution via sparse representation. *IEEE transactions on image processing*, 19(11):2861–2873, 2010. 5, 1

[40] Zhengwei Yin, Hongjun Wang, Guixu Lin, Weihang Ran, and Yinqiang Zheng. Random is all you need: Random noise injection on feature statistics for generalizable deep image denoising. In *The Thirteenth International Conference on Learning Representations*. 3

[41] Zhengwei Yin, Guixu Lin, Mengshun Hu, Hao Zhang, and Yinqiang Zheng. Flexir: Towards flexible and manipulable image restoration. In *Proc. ACM Int. Conf. Multimedia*, pages 6143–6152, 2024.

[42] Zhengwei Yin, Mingze Ma, Guixu Lin, and Yinqiang Zheng. Exploring data efficiency in image restoration: A gaussian denoising case study. In *Proc. ACM Int. Conf. Multimedia*, pages 2564–2573, 2024. 3

[43] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5728–5739, 2022.

[44] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Learning a single convolutional super-resolution network for multiple degradations. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3262–3271, 2018. 1

[45] Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4791–4800, 2021. 2, 3

[46] Wenlong Zhang, Guangyuan Shi, Yihao Liu, Chao Dong, and Xiao-Ming Wu. A closer look at blind super-resolution: Degradation models, baselines, and performance upper bounds. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 527–536, 2022. 2

[47] Xuaner Zhang, Qifeng Chen, Ren Ng, and Vladlen Koltun. Zoom to learn, learn to zoom. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3762–3770, 2019. 1

[48] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 286–301, 2018. 2

[49] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2472–2481, 2018. 2

[50] Hongyi Zheng, Hongwei Yong, and Lei Zhang. Unfolded deep kernel estimation for blind image super-resolution. In *European Conference on Computer Vision*, pages 502–518. Springer, 2022. 3