# A New Dataset and Framework for Real-World Blurred Images Super-Resolution

Rui Qin<sup>1,2</sup>, Ming Sun<sup>2</sup>, Chao Zhou<sup>2</sup>, and Bin Wang<sup>1,†</sup>

<sup>1</sup> School of Software, Tsinghua University, Beijing, China qr20@mails.tsinghua.edu.cn, wangbins@tsinghua.edu.cn <sup>2</sup> Kuaishou Technology, Beijing, China {sunming03, zhouchao}@kuaishou.com

Abstract. Recent Blind Image Super-Resolution (BSR) methods have shown proficiency in general images. However, we find that the efficacy of recent methods obviously diminishes when employed on image data with blur, while image data with intentional blur constitute a substantial proportion of general data. To further investigate and address this issue, we developed a new super-resolution dataset specifically tailored for blur images, named the Real-world Blur-kept Super-Resolution (ReBlurSR) dataset, which consists of nearly 3000 defocus and motion blur image samples with diverse blur sizes and varying blur intensities. Furthermore, we propose a new BSR framework for blur images called Perceptual-Bluradaptive Super-Resolution (PBaSR), which comprises two main modules: the Cross Disentanglement Module (CDM) and the Cross Fusion Module (CFM). The CDM utilizes a dual-branch parallelism to isolate conflicting blur and general data during optimization. The CFM fuses the well-optimized prior from these distinct domains cost-effectively and efficiently based on model interpolation. By integrating these two modules, PBaSR achieves commendable performance on both general and blur data without any additional inference and deployment cost and is generalizable across multiple model architectures. Rich experiments show that PBaSR achieves state-of-the-art performance across various metrics without incurring extra inference costs. Within the widely adopted LPIPS metrics, PBaSR achieves an improvement range of approximately 0.02-0.10 with diverse anchor methods and blur types, across both the ReBlurSR and multiple common general BSR benchmarks. Code here.

Keywords: Blind Super-resolution  $\cdot$  Image Blur  $\cdot$  Dataset

# 1 Introduction

Blind Image Super-Resolution (BSR) [6, 7, 9, 10, 37, 48, 52, 65, 75, 84] aims to realistically reconstruct High-Resolution (HR) images from Low-Resolution (LR) images with unknown degradation. Benefiting from the construction of various benchmarks [2, 3, 21, 38, 44, 68, 74], the BSR technique has achieved considerable

<sup>&</sup>lt;sup>†</sup> Corresponding author.

#### 2 R. Qin et al.



Fig. 1: Recent BSR methods' performance on blur data. Left: Examples of real images with blur. Right: LPIPS on blur and non-blur images in DIV2K-Val.

success in the field of general images [20, 27, 28, 37, 50, 64] as well as in various specific image categories, such as face images [4, 34, 35], cartoon images [68], text images [8, 53, 79], and remote sensing images [18, 69].

Blurring, a prevalent form of image degradation, is influenced by factors such as aperture size and exposure time, manifesting in actual images with different visual perceptions. Previous research on blur often treated it negatively, focusing on its detection [29,60] and removal [1,11]. However, upon investigation [19,62], we found that blurring is also a commonly used visual perception enhancement photography technique. Take the real sample from DIV2K [2] in Fig. 1a as an example, it accentuates the foreground squirrel by defocusing the background. In addition, this type of technique is broadly utilized in numerous commercial multimedia editing platforms, evidenced by the 'Blur' features within the basic 'Filter' menus of Adobe Photoshop, Krita, and LunaPic. Therefore, we believe that the intentionally set blur should be preserved rather than entirely removed during the BSR process, which has not been paid much attention to yet. As shown in Fig. 1a, recent methods over-enhance the blur region despite the overall high PSNR/SSIM, which impairs the actual perception quality. To quantitatively verify our surmise, we also used a more reasonable perceptual metric, LPIPS [77], for a brief evaluation on the DIV2K-Val [2] benchmark (100 samples) and split it into samples with blur vs. those without blur. As shown in Fig. 1b, the LPIPS of blur samples generally decreases by 0.01~0.06 across different methods compared to the fully-focused samples without blur. To further investigate this neglected problem, we develop a new super-resolution benchmark specifically tailored for the restoration of images with blur, named the Real-world Blur-kept Super-Resolution (ReBlurSR) dataset. It contains 2,931 high-quality (HQ) images with diverse blur sourced from the existing super-resolution (e.g., DIV2K, DIV8K [21], Flickr2K [38]), blur-related datasets (e.g., CUHK [57], EBD [30]) and web images. We meticulously labeled and categorized the blur data among various characteristics, such as types, sizes, and intensities.

Except for the ReBlurSR dataset, we argue that an optimal blur BSR framework should adaptively handle blur images and enhance their restoration performance without affecting general image processing quality. Furthermore, it should integrate easily with existing SOTA models, with minimal additional inference costs. In response to these criteria, we propose the Perceptual-Blur-adaptive Super-Resolution (PBaSR), the first blur-adaptive BSR framework. The PBaSR framework consists of two main modules: the Cross Disentanglement Module (CDM) and the Cross Fusion Module (CFM). On the one hand, the CDM, with its branch tailored for blur, separates blur from non-blur data for focused learning and reduces conflicts in uniform training. On the other hand, the CFM utilizes adaptive cross-branch weight interpolation, fostering the exchange of information between branches while keeping feature spaces aligned. This setup allows for the effective fusion of optimal priors from each branch in its respective domains through cost-efficient weight averaging, thereby handling both blur and non-blur data well without adding any inference or deployment costs. By synergizing these modules, PBaSR substantially elevates the performance of various recent mainstream BSR methods on blur data. Comprehensive testing on Re-BlurSR and several general BSR benchmarks with various quantitative metrics shows that PBaSR achieves state-of-the-art performance in blur image blind super-resolution and matches recent top-performing methods for general data processing. PBaSR consistently enhances performance across different perceptual metrics, suitable for various methods and blur types, showing a 0.02 to 0.10 improvement in LPIPS without any additional inference costs. In summary, our main contributions are as follows:

- We observed that recent BSR methods ignore the treatment of images with intentional blur, and have limitations on the preservation of blur regions.
- In response to this phenomenon, we propose the Real-world Blur-kept Super-Resolution (ReBlurSR) dataset for blur image blind super-resolution. It contains 2,931 high-quality images with diverse blur regions, nearly three times the size of the commonly used SR benchmark DIV2K.
- We validate the limitations of current methods on blur BSR, and propose the new Perceptual-Blur-adaptive Super-Resolution (PBaSR) framework for blur BSR. Extensive experiments on multiple benchmarks show that PBaSR effectively improves BSR performance on blur data while maintaining commendable performance on general data, with no additional inference cost.

# 2 Related Work

### 2.1 Single Image Super-Resolution

Since the advent of SRCNN [15], many CNN-based super-resolution frameworks [14,32,58] have been developed. EDSR [38] and RDN [78] improved SISR using Residual Dense Blocks [25]. Recently, frameworks using VQVAE [6,52,83] and Transformer [37, 39, 54, 84] have shown better results. BSR is particularly challenging due to diverse degradation, leading many to use GANs [45, 66, 85] for enhancing texture and incorporating well-designed degradation synthesis processes [65,75] with high-quality datasets. As noted in Sec. 1, intentional blur can improve visual perception [19,62]. Fig. 1 shows that over 20% of DIV2K-Val samples are blurred, and current methods struggle with such data. Thus, researching blur BSR and constructing datasets is beneficial for BSR techniques. 4 R. Qin et al.

#### 2.2 Image Super-Resolution Benchmark Datasets

Recent image super-resolution (SR) has been significantly driven by the development of comprehensive benchmarks. The DIV2K [2], including 1000 high-quality (HQ) 2K resolution images, has become a widely used standard for evaluating SR models due to its diverse and detailed content. Similarly, the Flickr2K [38], with its 2650 HQ images from various scenarios, provides sufficient data for training robust SR models. For evaluation, the BSDS100 [43] is widely used, with 100 challenging images with complex natural scenes. The Set5 [3] and Set14 [74] datasets, despite their smaller sizes, remain popular for their historical significance. Besides, the Urban100 [26] and Manga109 [44] datasets cater to specific needs, with the former focusing on urban scenes and the latter on comic images. Though these datasets provide robust platforms for the evaluation of general SR tasks, there is still a lack of specialized benchmarks for blur image SR.

### 3 Methodology

#### 3.1 Blur in Real-World Blind Super-Resolution

Though recent BSR methods excel in general image enhancement, they do not exhibit optimal performance when facing blur images, while blur is widespread in high-quality multimedia content and is an important photographic technique.

To illustrate this, we selected blur samples from the DIV2K-val dataset, revealing recent SR methods' limitations, such as over-texturization and vignetting destruction, as shown in Fig. 1a. These problems are overlooked by PSNR/SSIM metrics due to their insensitivity to smooth areas. To further assess the inadequacy quantitatively, we conducted statistical experiments using the DIV2K-Val dataset. Specifically, we used the blur detection method D-DFFNet [29] to differentiate samples into those with and without blur. We found that over 20% of samples had noticeable blur, which is a non-negligible proportion. Besides, as Fig. 1b indicates, the SOTA BSR methods showed a marked decrease in effectiveness on the blur samples, with a 0.01-0.06 drop in LPIPS.

These findings highlight blur's significance in high-quality media, a gap in current datasets and BSR methods. Therefore, this work aims to develop a focused subset for this gap and improve BSR methods' performance on blur data.

#### 3.2 Real-World Blur-Kept Super-Resolution Dataset

To support the training and evaluation of BSR methods on blur images, we created the high-quality **Re**al-world **Blur**-kept **S**uper-**R**esolution dataset, namely, ReBlurSR dataset, comprising of two subsets. We selected 2210 real-world blur images from sources like DIV2K, Flickr2K, DIV8K, CUHK, EBD, and the web. Additionally, we synthesized 601 blur images using popular diffusion-based models. These 2811 images form the ReBlurSR-Train set. For realistic evaluation, the ReBlurSR-Test set includes 120 real blur images from existing validation benchmarks. Each sample has a High-Resolution (HR) image and a blur region map, indicating blur pixels with 0 and non-blur pixels with 1.



**Fig. 2:** Data collection and partition of ReBlurSR. (a) Three methods of collecting the ReBlurSR dataset. (b) The samples from different data partitions.

**Data Collection** For diversity and realism, we gathered data from recent public datasets, the web and synthesized realistic blur images using multi-modal models and image-generating models. As shown in Fig. 2a, our data collection involved:

1) From blur-specific datasets EBD [30] and CUHK [57], we chose large samples (over  $512 \times 512$  pixels) with less than 80% blur region and paired the provided manually labeled blur maps with images, resulting in 849 samples.

2) Using D-DFFNet [29] and method in [31] for blur estimation on universal SR datasets, including Flickr2K [38], DIV2K [2], and DIV8K [21], we filtered out samples with less than 5% blur region and manually corrected detection errors of estimated blur maps. Besides, we collected 315 HQ samples from the web using the same processing, resulting in 1,362 valid samples in total.

3) For synthetic imagery, we developed a pipeline to create synthetic blur images using Large Language Models (LLMs) and diffusion-based methods. As shown in Fig. 3b, the process begins with extracting concise image descriptions from real datasets using the GIT [63] method. GPT-3.5 [49] diversifies these descriptions, specifying blur types. The Stable Diffusion [55] model transforms these descriptions into  $512 \times 512$  images, which are then upscaled to  $2048 \times 2048$ due to computational limitations. After blur detection and manual verification to remove errors, we produced 601 synthetic samples (more details in the Supp.).

4) Furthermore, we applied the filtering process to real-world samples from the validation sets of DIV2K [2], CUHK [57] and EBD [30], resulting in the collection of 120 samples for the ReBlurSR-Test set.

In total, ReBlurSR comprises 2,210 real blur images and 601 synthetic blur images for training purposes, alongside 120 real blur images designated for testing. This collection is nearly *threefold* larger than the widely used DIV2K and **1.14 times** larger than Flickr2K, which is now accessible online for public use. **Data Partition** Following prior blur research [1, 60], our ReBlurSR dataset encompasses the two prevalent types of blur found in natural images: **1**) Defocus Blur primarily arises from aperture settings. Smaller apertures reduce it,



**Fig. 3:** Data analysis of the ReBlurSR dataset. (a) The sample distribution of image resolution and blur type. (b) The local gradient in the blur region of different blur intensity subsets. (c) The sample distribution and the HR images' quality assessment based on NIQE [47] of different blur area sizes and blur intensities in ReBlurSR.

but larger ones cause defocus blur for objects outside the focal plane, intensifying with distance from the focal plane. Modern photography often uses it to highlight the foreground by blurring less important areas. 2) Motion Blur results from relative movement during exposure, such as fast-moving subjects or camera shake, and is more pronounced in low-light conditions requiring longer exposures. While previous studies focused on correcting unintended blur, our work aims to preserve professionally intended motion blur. For fine-grained assessment, we conducted a detailed dataset partition based on the following criteria:

**Blur Area Size:** We categorized each sample based on the blur area's proportion of the total image size into three groups: small (<45%), medium ( $45\%\sim55\%$ ), and large (>55%). In defocus blur samples, the blur area size predominantly falls in the small to medium range, whereas in motion blur samples, it is mostly medium to small and large. Real samples are shown in Fig. 2b.

**Blur Intensity:** Given the challenges in quantitatively predicting blur intensity, we relied on human visual assessment to classify images into three intensity levels: little (minimal edge and texture loss), middle (noticeable edge overlap and significant texture loss), and heavy (extensive edge loss and almost complete texture elimination), as shown in Fig. 2b. To intuitively assess the quality of our classification, we calculated the pixel-level gradient in blur regions of different intensity subsets. As shown in Fig. 3b, the gradient decreases with increasing intensity levels, indicating the degradation of textures and edges as expected.

**Data Analysis** In Fig. 3, we show the statistical distribution of resolution, blur type, size, and intensity in the ReBlurSR dataset's high-resolution (HR) images. The dataset mainly consists of high-quality images larger than  $1920 \times 1080$ , with a higher prevalence of defocus blur compared to motion blur, likely due to the former's wider application. Fig. 3b and Fig. 3c indicate that samples with lower blur intensity have better NIQE metrics. Additionally, there is a correlation between blur intensity and blur area size: images with extensive blurring tend to have larger blur areas, while those with smaller blur areas exhibit less blurring.

#### 3.3 Perceptual-Blur-Adaptive Super-Resolution

Given the current methods' disparate results on blur and general data, we posit that an ideal blur-adaptive BSR framework should meet the following criteria: - **Robustness.** The framework must enhance blur image processing quality without compromising the performance of general data.

- Efficiency. The framework should integrate with various methods requiring minimal modifications, and incur negligible additional costs in inference.

**Revisiting Recent Methods' Performance** To investigate the relationship between general and blur data, and to evaluate the performance of current models, we conducted three demo experiments using FeMaSR, as detailed below:

(1) We explored the impact of data diversity by maintaining constant total training data amount and altering the blur data (ReBlurSR-Train) to general data ratio for FeMaSR fine-tuning. Findings in Fig. 4a illustrate a performance trade-off, with LPIPS shifts around 0.05 for general data and 0.08 for blur data, respectively, indicating that increasing specific data types boosts performance in its respective category but fails to simultaneously improve both.

(2) To investigate convergence dynamics over time, we unified both data types for an extended fine-tuning (370k iterations) and tracked FeMaSR's performance from 240k to 370k iterations. Fig. 4b shows LPIPS scores for general and blur data oscillating between 0.415~0.385 and 0.395~0.355, respectively, revealing a negative correlation. This further highlights the conflict between the two data types, which cannot be resolved simply by unified training.

(3) To explore why models excel with single-class data but falter postmerging, we analyzed adversarial loss on blur data. By comparing the average discriminator hinge loss between blur and non-blur pixels on HR Blur data, Fig. 4c reveals that loss in blur regions escalates from  $25\% \sim 30\%$  to  $150\% \sim 300\%$ relative to non-blur areas as blur intensity and blur area size grow, signifying diminished confidence and discriminative capacity of the classifier in blur regions.

In summary, the existing framework struggles with robustness and efficiency when handling both blur and general data simultaneously. It cannot improve performance on both data types at once, indicating a robustness issue, while splitting inference parameters by data type contradicts efficiency principles. To address these challenges, we propose the Perceptual-Blur-adaptive Super-Resolution (PBaSR) framework, which effectively merges blur and general data into a unified model without extra inference or deployment costs. The PBaSR



**Fig. 4:** (a) The performance of FeMaSR [6] fine-tuned on different proportions of general (G) and blur (B) data (keeping the total training data amount the same). (b) FeMaSR's performance during late iterations which were unified trained on the combination of all general and blur images (darker colors denote later iterations). (c) The adversarial loss of general discriminator on the HR images of ReblurSR-Test.



Fig. 5: The structure of the Cross Disentanglement Module (CDM).

framework incorporates the Cross Disentanglement Module (CDM) and the Cross Fusion Module (CFM), which will be detailed in the sections below.

**Cross Disentanglement Module** In this part, we introduce the Cross Disentanglement Module (CDM). Insights from Fig. 4a and 4b show that separate training on distinct data types enhances performance in their respective datasets. Inspired by this, we created the CDM with two branches for independently processing each data type, enabling disentanglement at the sample level for general and blur data. The structure of CDM is detailed in Fig. 5.

First, the general branch processes general data using existing training strategies, while the distinct blur branch is tailored for training on blur data. Fig. 4c indicates that training blur data within the conventional unconditional adversarial learning framework leads to suboptimal results, for which the reduced confidence of the discriminator in the blur regions is one of the primary factors. Given the  $0.3 \sim 0.5$  GAN loss disparity between blur and non-blur areas, we advocate for customized loss adjustments in the blur branch for better optimization in blur regions, diverging from general data strategies. We posit that incorporating a priori information as an additional loss calculation guide is beneficial. Thus, blur maps from the ReBlurSR dataset, derived from human annotation and auto-detection via D-DFFNet with a 0.5 threshold for binary conversion. have been integrated. These maps, with a manual-to-auto annotation ratio of about 20%:80%, inform the redefined GAN loss for the blur branch. As shown in Eq. (1), the blur map M is concatenated with the input image and then input into the discriminator  $D_B$  for conditional discrimination. By incorporating the blur map as a conditional factor, the adversarial loss  $L_{D_B}$  is computed in a manner that respects the distinct characteristics of different regions,

$$L_{D_B} = \frac{1}{WHC} \times \sum_{c=1}^{C} \sum_{w=1}^{W} \sum_{h=1}^{H} \left\{ (1 - D_B(I_{HR_B}|M)_{w,h,c}) + (1 + D_B(I_{SR_B}|M)_{w,h,c}) \right\}$$
(1)

Thus, with these adjustments, we believe that the blur branch will more adeptly and efficiently process blur data. To support this, we executed a simple exper-



**Fig. 6:** Map visualization of the discriminator loss in the HR blur samples. More blue for larger residual and red for the reverse.

Table 1: Discriminator loss on HR de-
focus data in ReblurSR-Test. $CDM_0$
denotes CDM trained without the blue
conditional adversarial loss. 'Blur', 'Fo-
cus', and 'All' in $L_D$ on blur data de-
note the corresponding regions' loss.

	Mothod	LP	IPS	$L_D$ on blur data					
Method		Blur	General	Blur	Focus	All			
	Baseline	0.4178	0.3848	0.7047	0.8588	1.5636			
	$CDM_0$	0.3786	0.3809	0.5677	0.5926	1.1603			
	CDM	0.3564	0.3826	0.3974	0.6051	1.0026			

iment to assess the impact of integrating the a priori blur map condition. As indicated in Tab. 1, employing the blur condition led to a reduction in the loss of blur pixels by approximately 0.17 and yielded a 0.022 improvement in LPIPS for blur samples when compared to the baseline classifier's performance on the ReBlurSR-Test HR data, without markedly impacting the loss in non-blur regions. Furthermore, visual analysis of the discriminator loss residual map on specific samples, depicted in Fig. 6, further corroborates that the inclusion of the blur condition results in a notable decrease in the loss within blur regions. These observations affirm that the blur map serves as a valuable auxiliary for more accurate discrimination in blur regions.

**Cross Fusion Module** The CDM enables us to attain commendable results on both blur and general data. Yet, as outlined in Sec. 3.3, we aim to avoid adding complexity and computational demands during inference, a challenge given CDM's requirement for doubled parameters and extra data type distinction. An initial thought might be to average the branch weights after training, but this method falls short of expectations. Tab. 2 shows that while each branch excels on its target data type, yielding about 0.01 average LPIPS improvement over the unified training baseline, the performance after weight averaging (Tab. 2 col.4) is still similar to that of the unified baseline (Tab. 2 col.1), indicating a persistent sub-optimal trade-off. This issue stems from inadequate communication between the branches during training. To overcome this and improve performance without additional inference costs, we propose the Cross Fusion Module (CFM), with its structure and operational phases detailed below and in Fig. 7.

**Table 2:** LPIPS comparison of FeMaSR trained unified on both kinds of data and the averaged model of two branches of  $CDM_{FeMaSR}$  after being trained separately.

Test Data	Unified Trained FeMeSD	CDM <sub>FeMaSR</sub> only						
Test Data	Unnied Trained Femask	General Branch	Blur Branch	Poster-Average				
General Data	0.3976	0.3855	0.4234	0.3890				
Blur Data	0.3714	0.4100	0.3649	0.3711				



**Fig. 7:** The structure of the Cross Fusion Module (CFM). 't' denotes the iteration number. 'k' denotes the interval of CFM operation.

Training Stage During the training phase, our goal is for the weights from both the general and blur branches to communicate effectively. Given the need for efficiency in a multi-GPU training setup, we aim to optimize communication costs while enhancing quality. Therefore, we argue that an optimal communication strategy should meet two key criteria: **first**, to achieve optimal results on both kinds of data, the weights of the general  $W_G$  and blur  $W_B$  branches should exhibit a certain level of difference; **second**, to ensure effective fusion and inference, communication intervals should not be excessively long. Therefore, we adopt weight interpolation as a cost-effective method of cross-branch communication, balancing weight distance dynamically with infrequent updates. Specifically, we interpolate the weights of the general branch  $W_G$  and the blur branch  $W_B$  at every k iteration adaptive to the cross-branch distance and apply these interpolated weights  $(W'_{G,B})$  to the subsequent iteration,

$$W_{G}^{'} = \lambda W_{G} + (1-\lambda)W_{B}, W_{B}^{'} = \lambda W_{B} + (1-\lambda)W_{G}, \lambda = \lambda_{0} + (1-\lambda_{0})\frac{W_{G} \cdot W_{B}}{2\|W_{G}\|\|W_{B}\|}.$$
(2)

Inference Stage After training, we combine the weights from both branches to create a unified model. This involves equal interpolation of both branches to establish the final model parameters, expressed as  $W_{PBaSR} = W_G/2 + W_B/2$ .

The CFM and CDM do not modify the model's architecture, ensuring the inference process aligns with the original model and avoids extra inference or



Fig. 8: Comparison of CDM trained with and without CFM. (a) The cosine distance between two branches' weights at different iterations. (b) The LPIPS on general and blur data at different iterations. 'Baseline' denotes the unified training as in Fig. 4b.

11

deployment costs. CFM's role in maintaining weight distance between different branches during training allows for the effective fusion of priors for different data types via simple interpolation averaging during inference. For clarity, we conducted an ablation analysis comparing the CDM trained with and without CFM. As shown in Fig. 8a, the cosine distance between the two branches' weights is effectively constrained within a certain range by CFM, while the distance without CFM fluctuates obviously. Besides, Fig. 8b shows that the LPIPS of CDM without CFM on general and blur data are entangled in the same region as the unified trained baseline, whereas CFM achieves a 0.01-0.015 LPIPS improvement.

# 4 Experiment

#### 4.1 Experimental Setup

#### Dataset

Training Datasets: For general training data, we employed widely recognized super-resolution datasets, including DIV2K [2], Flickr2K [38], and DIV8K [21], covering a broad spectrum of high-resolution, real-world images. Regarding blur data, we utilized the ReBlurSR dataset as the primary source. We applied the degradation model from BSRGAN [75] for low-resolution (LR) image synthesis. *Testing Datasets:* We assessed performance on general data using benchmark datasets including DIV2K [2], Urban100 [26], BSDS100 [43], Manga109 [44], Set14 [74], and Set5 [3]. For blur data, we evaluated methods using the ReBlurSR-Test dataset. LR inputs are generated from a mixed degradation of BSRGAN [75] and RealESRGAN [66]. Considering experimental efficiency, DIV2K [2] is used as the default general benchmark in all ablation analyses.

**Evaluation** We use perceptual metrics as the main performance evaluation metrics. Specifically, we employ six widely recognized metrics to evaluate the performance of our proposed method, namely Learned Perceptual Image Patch Similarity (LPIPS) [77], Visual Information Fidelity (VIF) [56], Gradient Magnitude Similarity Deviation (GMSD) [71], Visual Saliency Index (VSI) [76], Deep Image Structure and Texture Similarity (DISTS) [12], and Attention-based Hybrid Image Quality (AHIQ) assessment [33]. For reference, the PSNR/SSIM and more perceptual metrics [5, 17, 82] results are also provided in the supplement.

Implementation Details In the implementation, branch initial weights are based on those from the general BSR task, using officially released weights. The optimization process for each branch is conducted independently, utilizing an Adam optimizer with a learning rate set at  $1 \times 10^{-4}$ . We set the batch size of each GPU and the patch size of HR to 16 and 256, respectively. Each batch is composed of an equal amount of general and blur data. The  $\lambda_0$  and k in CFM are set to 0.99 and 20, respectively. The training spanned a total of 200,000 iterations, executed on two NVIDIA V100 GPUs, with PyTorch serving as the programming framework. All experiments are conducted using x4 upscaling. 12 R. Qin et al.

**Table 3:** Quantitative comparison of PBaSR with the SOTA methods on blur data and general data. The blur validation set is ReBlurSR-Test. The general validation set consists of six benchmarks, including DIV2K, Urban100, BSDS100, Manga109, Set14, and Set5. The **best** and **second-best** results are bolded in black and red, respectively.

Data	Metric	SwinIR	Real-ESRGAN	MM-RealSR	FeMaSR	CAL-GAN	HAT	SRFormer	PBaSRESPCAN	PBaSB	PBaSR <sub>SREormer</sub>	
		(2021)	(2021)	(2022)	(2022)	(2023)	(2023)	(2023)	-ESIGAN -FEMAS		SAFormer	
	LPIPS $\downarrow$	0.4048	0.4199	0.4270	0.4037	0.4511	0.3924	0.3974	0.3986	0.3564	0.3740	
	AHIQ ↑	0.2396	0.2215	0.2204	0.2323	0.1857	0.2282	0.2394	0.2386	0.2670	0.2620	
Defocus	DISTS $\downarrow$	0.2074	0.2313	0.2342	0.1861	0.2607	0.2502	0.2139	0.1952	0.1733	0.1734	
Blur	VIF ↑	0.0873	0.0842	0.0833	0.0862	0.0821	0.0859	0.0896	0.0907	0.0938	0.0931	
	$GMSD \downarrow$	0.1843	0.1900	0.1851	0.1781	0.1854	0.1888	0.1796	0.1773	0.1737	0.1709	
	VSI ↑	0.9565	0.9547	0.9552	0.9602	0.9573	0.9528	0.9577	0.9609	0.9621	0.9625	
	LPIPS	0.4252	0.4104	0.4238	0.4659	0.4605	0.3856	0.4145	0.3791	0.3624	0.3887	
	AHIQ ↑	0.2852	0.2920	0.2771	0.2330	0.2343	0.2964	0.2846	0.2986	0.2884	0.2931	
Motion	DISTS $\downarrow$	0.2105	0.2212	0.2324	0.2097	0.2415	0.2250	0.2127	0.1907	0.1771	0.1844	
Blur	VIF ↑	0.1192	0.1135	0.1125	0.1225	0.1121	0.1171	0.1238	0.1244	0.1320	0.1292	
	$GMSD \downarrow$	0.1713	0.1731	0.1706	0.1775	0.1770	0.1679	0.1670	0.1648	0.1629	0.1649	
	VSI ↑	0.9744	0.9733	0.9721	0.9760	0.9749	0.9723	0.9748	0.9771	0.9782	0.9781	
	$LPIPS \downarrow$	0.4202	0.4703	0.4952	0.3986	0.4775	0.4850	0.4284	0.4463	0.3912	0.3937	
	$AHIQ \uparrow$	0.1944	0.1890	0.1707	0.1932	0.1557	0.1828	0.1990	0.1860	0.2190	0.2223	
C	DISTS $\downarrow$	0.2475	0.2839	0.3060	0.2303	0.2928	0.3115	0.2615	0.2587	0.2331	0.2229	
General	VIF ↑	0.0646	0.0568	0.0563	0.0619	0.0560	0.0598	0.0674	0.0617	0.0668	0.0673	
	$GMSD \downarrow$	0.2387	0.2515	0.2536	0.2225	0.2418	0.2523	0.2402	0.2350	0.2261	0.2243	
	VSI ↑	0.9176	0.9080	0.9059	0.9257	0.9139	0.9060	0.9174	0.9192	0.9257	0.9260	

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

For comprehensive and reasonable comparative analysis, we conducted our PBaSR on three widely recognized BSR methods, namely Real-ESRGAN [65], FeMaSR [6], and SRFormer [84]. We compared them with four other state-of-the-art (SOTA) BSR methods (including SwinIR [37], MM-RealSR [48], CAL-GAN [51], and HAT [9]), evaluating performance on both general and blur data. To ensure fairness in the comparison, the results for these SOTA methods were derived using the code and weights available from their official repositories. Additionally, we included comparative experiments to assess the impact of fine-tuning with the ReBlurSR dataset in our ablation studies, thereby providing a more nuanced understanding of our framework's performance under varied training conditions. The quantitative and qualitative results are shown in Tab. 3 and Fig. 9.

As shown in Tab. 3, our PBaSR<sub>FeMaSR</sub> (Tab. 3 col.9) outperforms the leading method for blur data, HAT [9] (Tab. 3 col.6), with an improvement of  $0.04 \sim 0.09$ in LPIPS [77], while maintaining comparable results on general data (Tab. 3 last 6 rows) with the best-performing FeMaSR [6] (Tab. 3 col.4) and SRFormer [84] (Tab. 3 col.7) across all metrics. Fig. 9 presents several examples showcasing different types of blur. Taking the first sample in Fig. 9, most general BSR methods (Fig. 9 cols. 1, 2, and 4.) generate artifacts and over-texturization in blur regions. In contrast, our PBaSR (Fig. 9 col.5) exhibits superior blur preservation both visually and quantitatively. While CAL-GAN [51] (Fig. 9 col.3 row.1) preserves blur relatively well, it tends to over-smooth and lose texture in focused areas (Fig. 9 col.3, row.2). PBaSR, however, not only enhances restoration in blur regions (Fig. 9 row.1 and 3) but also achieves the best perception and LPIPS metrics in focused areas (Fig. 9 Row.2 and 4). In summary, both quantitative and qualitative analyses reveal that PBaSR effectively alleviates recent methods' limitations in blur image blind super-resolution and also maintains comparative texture restoration performance in general blind super-resolution tasks.



Fig. 9: Comparison of  $PBaSR_{FeMaSR}$  with SOTA methods on blur regions (up) and non-blur regions (down). More results are provided in the supplement.

#### **4.3** Ablation Study

Scalability for Blur Data To assess PBaSR's flexibility with different blur types, we categorized the ReBlurSR-Train set by blur types and progressively included various blur types during training, beginning with defocus and then motion blur data. We conducted detailed evaluations on subsets of distinct blurs in the ReBlurSR-Test. Tab. 4 shows that adding defocus data (from the 1st to 2nd and the 4th to 5th rows) enhances performance on the defocus blur test for a 0.04 LPIPS improvement without a heavy negative impact on general data performance. Likewise, incorporating all blur data (from the 2nd to 3rd and the 5th to 6th rows) leads to enhancements across all blur data. These findings reveal PBaSR's robust scalability across different blur types, effectively adjusting to new blur categories while maintaining performance on previously trained types. Effect of Training Data To elucidate PBaSR's effectiveness, we undertook an extensive evaluation based on FeMaSR, examining how different types and amounts of training data affect performance. Fig. 10a reveals that without PBaSR (gray points), training solely on general or blur data improves results for that specific category but leads to reduced performance for the other. Merging both data types for training fails to yield simultaneous gains in both areas (sky blue point). Conversely, PBaSR (blue point) shows notable improvements for

Table 4: Scalability of  $PBaSR_{FeMaSR}$  on dif- Table 5: Generalizability across various ferent blur subsets. "+G", "+D", and "+M" model structures. "Base" denotes the ofdenote general, defocus blur, and motion blur ficial weights. "Flop" denotes the flop opdata, respectively. Top: Blur Intensity; Bot- eration number when processing a  $256 \times$ tom: Blur area size. 256 patch.

LPIPS

Train Data		LPIPS							Mathad	M 117	Б I	E1	LPIPS						
+6	+I	)+	М		Defocu	is Blur		Motion Blur			a .	Method	nodel 1 ype	Framework	r iops	Defocus	Motion	General	
				Little	Middle	Heavy	All	Little	Middle	Heavy	All	General 0.3848 0.3856 4 0.3826	ESRGAN	CNN	Base	11.7T	0.4199	0.4104	0.4739
1	×		<	0.4237	0.4014	0.3689	0.4037	0.4871	0.4706	0.4461	0.4659				Unified Finetune		0.4148	0.3910	0.4419
				0.3800	0.3717	0.3315	0.3657	0.4250	0.3833	0.3430	0.3800				PBaSR(ours)		0.3986	0.3791	0.4390
۰,	•	. 1	2	0.3505	0.9505	0.0010	0.3501	0.4200	0.0000	0.9905	0.0000		FeMaSR	VQVAE	Base	15.0T	0.4037	0.4586	0.3848
~	V	`	<u></u>	0.3704	0.5595	0.3190	0.3304	0.3995	0.3782	0.3203	0.3024				Unified Finetune		0.3714	0.3905	0.3976
				Small	Medium	Large	All	Small	Medium	Large	All				PBaSR(ours)		0.3564	0.3663	0.3826
√	$\times$	3	<	0.4178	0.4180	0.3368	0.4037	0.4352	0.3898	0.4765	0.4659	0.3848			Base		0.3974	0.4145	0.4170
$\checkmark$	~		<	0.3719	0.3848	0.3043	0.3657	0.3691	0.3323	0.3850	0.3800	0.3856 0.3826	SRFormer	Transformer	Unified Finetune	2.0T	0.4139	0.4144	0.4150
1	1	,	$\langle  $	0.3660	0.3774	0.2901	0.3564	0.3559	0.3239	0.3660	0.3624				PBaSR(ours)		0.3740	0.3887	0.3892



Fig. 10: Ablations on ReBlurSR-Test (Defocus) and DIV2K-Val. (a) FeMaSR trained with different training data settings. "×N" denotes the used ratio of the data type. " $1\times B + 1\times G$ " denotes the unified training with all training data. (b)/(c) Effect of the communication frequency 'k' and ratio interpolation ' $\lambda_0$ ' of CFM in Eq. 2.

both data types, evidencing the successful disentanglement and integration of these categories by the CDM and CFM.

Generalizability across Anchor Method Structures To evaluate PBaSR's versatility across different SR model architectures, we implemented the PBaSR framework on three widely recognized architectures: CNN, VQVAE, and Transformers. Tab. 5 shows that PBaSR yielded noticeable enhancements across these mainstream SR architectures. Additionally, our framework proved effective across models of varying sizes and types, from SRFormer to FeMaSR, showing consistent performance improvements regardless of model scale. Notably, SRFormer and FeMaSR saw higher gains, possibly due to their superior fitting capabilities. Effect of CFM Communication Strategy To evaluate the impact of the communication strategy in CFM, we performed ablation studies on its frequency and ratio of cross-branch interpolation with PBaSR<sub>FeMaSR</sub>. Fig.10b shows that either too frequent or infrequent communication (k = 1 or k = 100) decreases LPIPS by  $0.005 \sim 0.01$  due to inappropriate distances between branch weights, leading to poor interpolation results. Fig. 10c indicates that the absence of communication ( $\lambda_0 = 1$ ) or overly aggressive interpolation ( $\lambda_0 = 0$ ) negatively affects performance while a  $\lambda_0$  value between 0.9 ~ 0.99 ensures performance stability. Additionally, we compared CFM against other feature communication methods like feature distillation and teacher-student learning in supplement, which also indicates the effectiveness and efficiency of CFM.

# 5 Conclusion

In this work, we explored blind super-resolution for real-world blur images and created the ReBlurSR dataset containing 2931 diverse blur images. We proposed a novel Perceptual-Blur-adaptive Super-Resolution (PBaSR) framework to address the limitations of current methods in processing blur data. Extensive evaluations on various benchmarks reveal that PBaSR significantly improves performance on real-world blur images while maintaining strong results on general data, without incurring extra inference or deployment costs.

# Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant 62072271.

## References

- Abuolaim, A., Brown, M.S.: Defocus deblurring using dual-pixel data. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part X 16. pp. 111–126. Springer (2020)
- Agustsson, E., Timofte, R.: Ntire 2017 challenge on single image super-resolution: Dataset and study. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (July 2017)
- Bevilacqua, M., Roumy, A., Guillemot, C.M., Alberi-Morel, M.L.: Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In: British Machine Vision Conference (2012), https://api.semanticscholar.org/ CorpusID:5250573
- Chen, C., Li, X., Yang, L., Lin, X., Zhang, L., Wong, K.Y.K.: Progressive semanticaware style transformation for blind face restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11896– 11905 (2021)
- Chen, C., Mo, J., Hou, J., Wu, H., Liao, L., Sun, W., Yan, Q., Lin, W.: Topiq: A top-down approach from semantics to distortions for image quality assessment. arXiv preprint arXiv:2308.03060 (2023)
- Chen, C., Shi, X., Qin, Y., Li, X., Han, X., Yang, T., Guo, S.: Real-world blind super-resolution via feature matching with implicit high-resolution priors. In: Proceedings of the 30th ACM International Conference on Multimedia. pp. 1329–1338 (2022)
- Chen, H., Hao, J., Zhao, K., Yuan, K., Sun, M., Zhou, C., Hu, W.: Cassr: Activating image power for real-world image super-resolution. arXiv preprint arXiv:2403.11451 (2024)
- Chen, J., Li, B., Xue, X.: Scene text telescope: Text-focused scene image superresolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 12026–12035 (June 2021)
- Chen, X., Wang, X., Zhou, J., Qiao, Y., Dong, C.: Activating more pixels in image super-resolution transformer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 22367–22377 (June 2023)
- Chen, Z., Wu, Z., Zamfir, E., Zhang, K., Zhang, Y., Timofte, R., Yang, X., Yu, H., Wan, C., Hong, Y., et al.: Ntire 2024 challenge on image super-resolution (x4): Methods and results. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6108–6132 (2024)
- Cho, S.J., Ji, S.W., Hong, J.P., Jung, S.W., Ko, S.J.: Rethinking coarse-to-fine approach in single image deblurring. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4641–4650 (2021)
- Ding, K., Ma, K., Wang, S., Simoncelli, E.P.: Image quality assessment: Unifying structure and texture similarity. CoRR abs/2004.07728 (2020), https://arxiv. org/abs/2004.07728
- Ding, K., Ma, K., Wang, S., Simoncelli, E.P.: Comparison of full-reference image quality models for optimization of image processing systems. International Journal of Computer Vision 129, 1258–1281 (2021)

- 16 R. Qin et al.
- Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13. pp. 184–199. Springer (2014)
- 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 (2015)
- Fritsche, M., Gu, S., Timofte, R.: Frequency separation for real-world superresolution. In: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). pp. 3599–3608. IEEE (2019)
- 17. Ghildyal, A., Liu, F.: Shift-tolerant perceptual similarity metric. In: European Conference on Computer Vision (2022)
- Gong, Y., Liao, P., Zhang, X., Zhang, L., Chen, G., Zhu, K., Tan, X., Lv, Z.: Enlighten-gan for super resolution reconstruction in mid-resolution remote sensing images. Remote Sensing 13(6), 1104 (2021)
- 19. Google Chrome team, F.B.: Blur camera background (2023), https://developer.chrome.com/blog/background-blur?hl=zh-cn
- Gu, J., Lu, H., Zuo, W., Dong, C.: Blind super-resolution with iterative kernel correction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1604–1613 (2019)
- Gu, S., Lugmayr, A., Danelljan, M., Fritsche, M., Lamour, J., Timofte, R.: Div8k: Diverse 8k resolution image dataset. In: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). pp. 3512–3516. IEEE (2019)
- 22. Hinton, G., Vinyals, O., Dean, J.: Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531 (2015)
- Hore, A., Ziou, D.: Image quality metrics: Psnr vs. ssim. In: 2010 20th international conference on pattern recognition. pp. 2366–2369. IEEE (2010)
- Horé, A., Ziou, D.: Image quality metrics: Psnr vs. ssim. In: 2010 20th International Conference on Pattern Recognition. pp. 2366–2369 (2010). https://doi.org/10. 1109/ICPR.2010.579
- Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4700–4708 (2017)
- Huang, J.B., Singh, A., Ahuja, N.: Single image super-resolution from transformed self-exemplars. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2015)
- Huang, Y., Li, S., Wang, L., Tan, T., et al.: Unfolding the alternating optimization for blind super resolution. Advances in Neural Information Processing Systems 33, 5632–5643 (2020)
- Ji, X., Cao, Y., Tai, Y., Wang, C., Li, J., Huang, F.: Real-world super-resolution via kernel estimation and noise injection. In: proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. pp. 466–467 (2020)
- Jin, Y., Qian, M., Xiong, J., Xue, N., Xia, G.S.: Depth and dof cues make a better defocus blur detector. arXiv preprint arXiv:2306.11334 (2023)
- Jin, Y., Qian, M., Xiong, J., Xue, N., Xia, G.S.: Depth and dof cues make a better defocus blur detector. arXiv preprint arXiv:2306.11334 (2023)
- Kim, B., Son, H., Park, S.J., Cho, S., Lee, S.: Defocus and motion blur detection with deep contextual features. In: Computer Graphics Forum. vol. 37, pp. 277–288. Wiley Online Library (2018)

17

- Kim, J., Lee, J.K., Lee, K.M.: Deeply-recursive convolutional network for image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1637–1645 (2016)
- 33. Lao, S., Gong, Y., Shi, S., Yang, S., Wu, T., Wang, J., Xia, W., Yang, Y.: Attentions help cnns see better: Attention-based hybrid image quality assessment network. arXiv preprint arXiv:2204.10485 (2022)
- 34. Li, X., Chen, C., Zhou, S., Lin, X., Zuo, W., Zhang, L.: Blind face restoration via deep multi-scale component dictionaries. In: European conference on computer vision. pp. 399–415. Springer (2020)
- Li, X., Li, W., Ren, D., Zhang, H., Wang, M., Zuo, W.: Enhanced blind face restoration with multi-exemplar images and adaptive spatial feature fusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2706–2715 (2020)
- 36. Liang, J., Zeng, H., Zhang, L.: Details or artifacts: A locally discriminative learning approach to realistic image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2022)
- Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: Swinir: Image restoration using swin transformer. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 1833–1844 (2021)
- Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K.: Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 136–144 (2017)
- Liu, G., Ding, Y., Li, M., Sun, M., Wen, X., Wang, B.: Reconstructed convolution module based look-up tables for efficient image super-resolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12217–12226 (2023)
- 40. Liu, H., Wu, M., Yuan, K., Sun, M., Tang, Y., Zheng, C., Wen, X., Li, X.: Adadqa: Adaptive diverse quality-aware feature acquisition for video quality assessment. In: Proceedings of the 31st ACM International Conference on Multimedia. p. 6695–6704. Association for Computing Machinery (2023)
- 41. Lu, Y., Li, X., Pei, Y., Yuan, K., Xie, Q., Qu, Y., Sun, M., Zhou, C., Chen, Z.: Kvq: Kaleidoscope video quality assessment for short-form videos. arXiv preprint arXiv:2402.07220 (2024)
- Ma, C., Yang, C.Y., Yang, X., Yang, M.H.: Learning a no-reference quality metric for single-image super-resolution. Computer Vision and Image Understanding 158, 1–16 (2017)
- 43. Martin, D., Fowlkes, C., Tal, D., Malik, J.: 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. vol. 2, pp. 416–423. IEEE (2001)
- Matsui, Y., Ito, K., Aramaki, Y., Fujimoto, A., Ogawa, T., Yamasaki, T., Aizawa, K.: Sketch-based manga retrieval using manga109 dataset. Multimedia Tools and Applications 76, 21811–21838 (2017)
- Menon, S., Damian, A., Hu, S., Ravi, N., Rudin, C.: Pulse: Self-supervised photo upsampling via latent space exploration of generative models. In: Proceedings of the ieee/cvf conference on computer vision and pattern recognition. pp. 2437–2445 (2020)
- Mittal, A., Soundararajan, R., Bovik, A.C.: Making a "completely blind" image quality analyzer. IEEE Signal processing letters 20(3), 209–212 (2012)

- 18 R. Qin et al.
- Mittal, A., Soundararajan, R., Bovik, A.C.: Making a "completely blind" image quality analyzer. IEEE Signal Processing Letters 20(3), 209-212 (2013). https: //doi.org/10.1109/LSP.2012.2227726
- Mou, C., Wu, Y., Wang, X., Dong, C., Zhang, J., Shan, Y.: Metric learning based interactive modulation for real-world super-resolution. In: European Conference on Computer Vision (ECCV) (2022)
- 49. OpenAI: Chatgpt: Optimizing language models for dialogue. https://openai.com/chatgpt (2023), accessed: 2024-01-18
- Pan, X., Zhan, X., Dai, B., Lin, D., Loy, C.C., Luo, P.: Exploiting deep generative prior for versatile image restoration and manipulation. IEEE Transactions on Pattern Analysis and Machine Intelligence 44(11), 7474–7489 (2021)
- Park, J., Son, S., Lee, K.M.: Content-aware local gan for photo-realistic superresolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10585–10594 (2023)
- Qin, R., Sun, M., Zhang, F., Wen, X., Wang, B.: Blind image super-resolution with rich texture-aware codebook. In: Proceedings of the 31st ACM International Conference on Multimedia. pp. 676–687 (2023)
- Qin, R., Wang, B., Tai, Y.W.: Scene text image super-resolution via content perceptual loss and criss-cross transformer blocks. arXiv preprint arXiv:2210.06924 (2022)
- Qu, Y., Yuan, K., Zhao, K., Xie, Q., Hao, J., Sun, M., Zhou, C.: Xpsr: Cross-modal priors for diffusion-based image super-resolution. arXiv preprint arXiv:2403.05049 (2024)
- 55. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models (2021)
- Sheikh, H.R., Bovik, A.C.: Image information and visual quality. IEEE Transactions on image processing 15(2), 430–444 (2006)
- Shi, J., Xu, L., Jia, J.: Discriminative blur detection features. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2965–2972 (2014)
- 58. Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., Wang, Z.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1874–1883 (2016)
- Soundararajan, R., Bovik, A.C.: Rred indices: Reduced reference entropic differencing for image quality assessment. IEEE Transactions on Image Processing 21(2), 517–526 (2011)
- 60. Tang, C., Liu, X., Zheng, X., Li, W., Xiong, J., Wang, L., Zomaya, A.Y., Longo, A.: Defusionnet: Defocus blur detection via recurrently fusing and refining discriminative multi-scale deep features. IEEE Transactions on Pattern Analysis and Machine Intelligence 44(2), 955–968 (2020)
- 61. Tang, C., Liu, X., Zhu, X., Zhu, E., Sun, K., Wang, P., Wang, L., Zomaya, A.: R<sup>2</sup>mrf: Defocus blur detection via recurrently refining multi-scale residual features. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 12063–12070 (2020)
- 62. team, A.P.L.: Photography tips: when to use a large aperture. (2024), https://www.adobe.com/creativecloud/photography/hub /guides/when-touse-large-aperture.html
- Wang, J., Yang, Z., Hu, X., Li, L., Lin, K., Gan, Z., Liu, Z., Liu, C., Wang, L.: Git: A generative image-to-text transformer for vision and language. arXiv preprint arXiv:2205.14100 (2022)

- 64. 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/CVF Conference on Computer Vision and Pattern Recognition. pp. 10581–10590 (2021)
- 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/CVF international conference on computer vision. pp. 1905–1914 (2021)
- 66. 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 (ECCV) workshops. pp. 0–0 (2018)
- Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing 13(4), 600–612 (2004)
- Wu, Y., Wang, X., Li, G., Shan, Y.: Animesr: learning real-world super-resolution models for animation videos. Advances in Neural Information Processing Systems 35, 11241–11252 (2022)
- Xiao, Y., Yuan, Q., Jiang, K., He, J., Wang, Y., Zhang, L.: From degrade to upgrade: Learning a self-supervised degradation guided adaptive network for blind remote sensing image super-resolution. Information Fusion 96, 297–311 (2023)
- 70. Xie, L., Wang, X., Chen, X., Li, G., Shan, Y., Zhou, J., Dong, C.: Desra: Detect and delete the artifacts of gan-based real-world super-resolution models. In: Proceedings of the 40th International Conference on Machine Learning. ICML'23 (2023)
- Xue, W., Zhang, L., Mou, X., Bovik, A.C.: Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. IEEE transactions on image processing 23(2), 684–695 (2013)
- Yuan, K., Kong, Z., Zheng, C., Sun, M., Wen, X.: Capturing co-existing distortions in user-generated content for no-reference video quality assessment. In: Proceedings of the 31st ACM International Conference on Multimedia. p. 1098–1107. Association for Computing Machinery (2023)
- 73. Yuan, K., Liu, H., Li, M., Sun, M., Sun, M., Gong, J., Hao, J., Zhou, C., Tang, Y.: Ptm-vqa: Efficient video quality assessment leveraging diverse pretrained models from the wild. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2835–2845 (June 2024)
- Zeyde, R., Elad, M., Protter, M.: On single image scale-up using sparserepresentations. In: Curves and Surfaces: 7th International Conference, Avignon, France, June 24-30, 2010, Revised Selected Papers 7. pp. 711–730. Springer (2012)
- 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/CVF International Conference on Computer Vision. pp. 4791–4800 (2021)
- Zhang, L., Shen, Y., Li, H.: Vsi: A visual saliency-induced index for perceptual image quality assessment. IEEE Transactions on Image Processing 23(10), 4270– 4281 (2014). https://doi.org/10.1109/TIP.2014.2346028
- 77. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 586–595 (2018)
- Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y.: Residual dense network for image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2472–2481 (2018)

- 20 R. Qin et al.
- 79. Zhao, C., Feng, S., Zhao, B.N., Ding, Z., Wu, J., Shen, F., Shen, H.T.: Scene text image super-resolution via parallelly contextual attention network. In: Proceedings of the 29th ACM International Conference on Multimedia. p. 2908–2917 (2021)
- Zhao, K., Yuan, K., Sun, M., Li, M., Wen, X.: Quality-aware pre-trained models for blind image quality assessment. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 22302–22313 (June 2023)
- Zhao, K., Yuan, K., Sun, M., Wen, X.: Zoom-vqa: Patches, frames and clips integration for video quality assessment. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 1302–1310 (June 2023)
- Zheng, H., Yang, H., Fu, J., Zha, Z.J., Luo, J.: Learning conditional knowledge distillation for degraded-reference image quality assessment. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 10222–10231 (2021). https://doi.org/10.1109/ICCV48922.2021.01008
- Zhou, S., Chan, K., Li, C., Loy, C.C.: Towards robust blind face restoration with codebook lookup transformer. Advances in Neural Information Processing Systems 35, 30599–30611 (2022)
- Zhou, Y., Li, Z., Guo, C.L., Bai, S., Cheng, M.M., Hou, Q.: Srformer: Permuted self-attention for single image super-resolution. arXiv preprint arXiv:2303.09735 (2023)
- 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)
- Zhu, Q., Hao, J., Ding, Y., Liu, Y., Mo, Q., Sun, M., Zhou, C., Zhu, S.: Cpga: Coding priors-guided aggregation network for compressed video quality enhancement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2964–2974 (June 2024)