

Blind Super Resolution with Reference Images and Implicit Degradation Representation

Huu-Phu Do^{1*}, Po-Chih Hu^{1*}, Hao-Chien Hsueh¹, Che-Kai Liu¹,
Vu-Hoang Tran², and Ching-Chun Huang^{1**}

¹ National Yang Ming Chiao Tung University, Taiwan

² Ho Chi Minh City University of Technology and Education, Vietnam

Abstract. Previous studies in blind super-resolution (BSR) have primarily concentrated on estimating degradation kernels directly from low-resolution (LR) inputs to enhance super-resolution. However, these degradation kernels, which model the transition from a high-resolution (HR) image to its LR version, should account for not only the degradation process but also the downscaling factor. Applying the same degradation kernel across varying super-resolution scales may be impractical. Our research acknowledges degradation kernels and scaling factors as pivotal elements for the BSR task and introduces a novel strategy that utilizes HR images as references to establish scale-aware degradation kernels. By employing content-irrelevant HR reference images alongside the target LR image, our model adaptively discerns the degradation process. It is then applied to generate additional LR-HR pairs through down-sampling the HR reference images, which are keys to improving the SR performance. Our reference-based training procedure is applicable to proficiently trained blind SR models and zero-shot blind SR methods, consistently outperforming previous methods in both scenarios. This dual consideration of blur kernels and scaling factors, coupled with the use of a reference image, contributes to the effectiveness of our approach in blind super-resolution tasks.

1 Introduction

Image super-resolution (SR) is a widely studied task in the field of image processing, aiming to enhance the resolution and quality of low-resolution (LR) images. Extracting detailed information from an LR image for super-resolution poses significant challenges. Previous researchers have employed deep learning techniques to improve the quality of LR images [3, 8, 12, 17]. This task typically requires a substantial amount of high-resolution (HR) images and their corresponding LR counterparts for training an SR network through a supervised approach. Prior studies often utilized a bicubic kernel to synthesize low-resolution images from high-resolution ones [12]. This process can be formulated as:

$$I^{LR} = (I^{HR} \otimes k) \downarrow_s + n, \quad (1)$$

* Equal contributions

** Corresponding Author

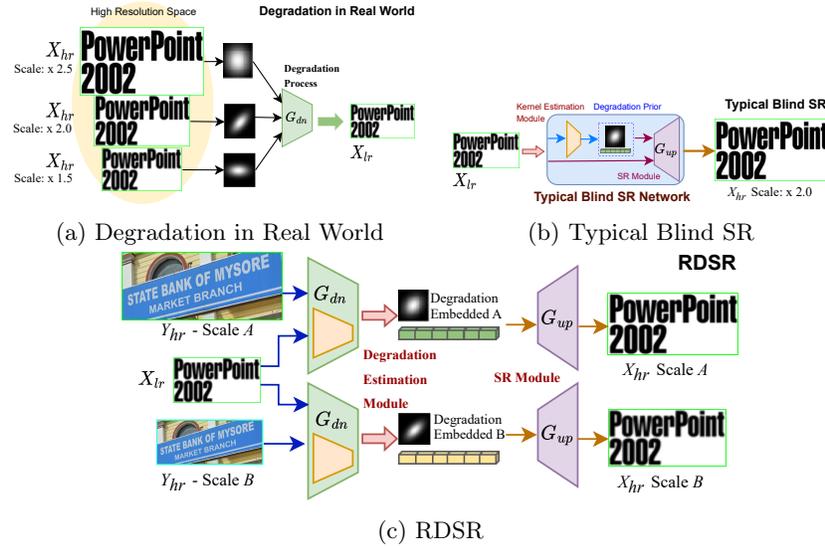


Fig. 1: (a) Degradation with multi scales in real world (b) Typical Blind Super-Resolution (c) Proposed Method - RDSR.

where I^{LR} denotes the low-resolution image, I^{HR} is the high-resolution image, k represents the degradation kernel, \downarrow_s represents downsampling operation with scale factor s and n is the added noise.

In practical applications, blur kernels are often much more complex than the typically bicubic ones. This domain gap can result in significant performance degradation when applying these networks to real-world scenarios. Therefore, there is a critical need to focus on super-resolution (SR) techniques that can handle unknown blur kernels, known as blind SR. Currently, existing blind SR techniques [2, 5] primarily focus on deriving a single degradation kernel that is assumed to be universally applicable across different SR scales. As depicted in Fig. 1(b), predominantly involve estimating kernels for a specific scaling factor and incorporating them within the super-resolution (SR) model. Consequently, thereby overlook the nuances introduced by varying scales. Effective blind super-resolution, however, requires accounting for the variability in blur kernels and the ability to adapt to different scaling factors. As illustrated in Fig. 1(a), the degradation kernel associated with a specific scaling factor significantly differs from kernels tailored for other scaling factors, emphasizing the need for a more flexible and scale-aware approach for kernel estimation.

Therefore, in this paper, we propose an approach for a degradation kernel that is inherently scale-aware. This is crucial for accurately super-resolving a low-resolution (LR) image to its high-resolution (HR) counterparts across various scales, necessitating a degradation kernel that is precisely tailored for each specific scenario. To achieve this goal, we introduce a reference-based training

procedure that integrates content-irrelevant HR reference images into the super-resolution (SR) model as conditional information. This innovative approach allows the SR model to leverage diverse high-resolution data, enhancing its ability to reconstruct high-quality images from low-resolution inputs. These reference images serve dual purposes: (a) to introduce a scaling factor as a condition for the SR network, and (b) to provide additional HR information, thereby facilitating image enhancement at the desired resolution. Fig. 1(c) depicts the overall concept of our method.

Diverging from existing reference-based SR approaches [6,13,20], our methodology does not directly utilize detailed information from the reference content; instead, it adaptively estimates a scale-aware kernel to model the degradation process accurately from the resolution of the reference images to that of the LR input. Our proposed procedure, termed “**Reference Degradation Super Resolution (RDSR)**” consists of upsampling and downsampling networks and is adaptable to various blind SR networks and zero-shot super-resolution networks. We also evaluate our instance implementation, the RDSR model, which utilizes KernelGAN [2] and a well-trained DASR [19] as the backbones for the upsampling and downsampling networks respectively, and achieve the superior performance compared to traditional SR methods. The main findings and contributions of this paper are summarized as follows:

1. The degradation processes for an LR image derived from HR images with different resolutions should be distinct. This suggests that the corresponding blur kernel from HR images downsampled to varying resolutions should also differ.
2. Our proposed approach integrates reference images into the architecture to facilitate scaling information for refining the degradation kernel, thereby optimizing the performance of the super-resolution network.
3. In contrast to other reference-based super-resolution methods that rely on learning high-resolution (HR) content from closely related reference images, our approach permits the use of content-irrelevant HR reference images. This strategy effectively mitigates the risk of content inconsistency issues, which are prevalent in traditional reference-based SR methods.
4. We introduce adversarial and regularization losses to optimize our proposed network. The adversarial loss encourages the upsampling network to align with the domain distribution of the reference images. Meanwhile, the regularization loss ensures consistency in degradation priors between the reference and target images.
5. Our reference-based training procedure is specifically designed for the target LR image and operates independently of other LR images. This approach is applicable not only to well-trained blind SR models but also to zero-shot super-resolution models.

2 Related Work

Numerous studies [22, 23] focus on single image super-resolution (SISR) tasks with known degradation. However, these methods may experience significant

performance drops when the degradation process is misaligned with the training samples. In addressing blind SR, various approaches have been proposed, categorized into four main groups: degradation estimation methods, unsupervised methods, zero-shot methods, and reference-based methods.

2.1 Degradation Estimation of Blind SR

To tackle image super-resolution with unknown degradation, numerous studies concentrate on degradation estimation. The objective is to uncover clues within the LR images and utilize the estimated degradation to guide the super-resolution process. These methods fall into two categories: explicit kernel estimation and implicit kernel estimation.

Explicit Kernel Estimation The explicit kernel estimation methods involve directly evaluating a degradation kernel, as demonstrated in previous works such as IKC [5] and KernelGAN [2]. IKC introduced an iterative method for correcting the kernel based on intermediate SR images during the training stage. However, this correction method requires multiple degradation kernels, along with corresponding LR and HR images during training. KernelGAN proposed a kernel estimation approach that learns the degradation kernel by establishing internal connections between patches in a single LR image. This sets it apart from other SR methods that rely on external training materials. This attribute not only offers flexibility but is also precisely customized for the target images.

Implicit Kernel Estimation Different from explicit kernel estimation which may suffer a performance drop if the kernel estimation error increases, DASR [19] and CDSR [24] proposed implicit kernel estimation for learning the degradation representations using contrastive learning. The kernel estimator produces a degradation representation, serving as an implicit degradation prior, and this representation is seamlessly integrated into the SR network to guide the model through various degradation scenarios.

2.2 Unsupervised Super-Resolution Methods

Unsupervised SR methods [14, 15] distinguish themselves from supervised approaches that rely on paired LR and HR data with synthetic degradation kernels. Instead, these methods focused on learning the distribution within the domain encompassing bicubic and real-world degradation processes. Following the acquisition of domain distribution knowledge, a domain corrector is employed to enhance SR performance.

2.3 Zero-shot Super-Resolution Methods

While the majority of SR methods necessitate paired data comprising LR, HR, and degradation kernel information for training, this approach poses challenges

in terms of adaptability for individual images. In contrast, zero-shot super-resolution methods, exemplified by [4, 18, 21] train models using only a single target LR image. These approaches yield an SR model tailored to the specific characteristics of the target images, deviating from the conventional generalized SR models. In the context of this, DualSR proposes a dual-path architecture optimized for blind super-resolution using only the input LR image. RZSR, on the other hand, advocates the creation of a reference dataset through cross-scale matching using the input target image. Our RDSR also targeted at this task.

2.4 Reference-based Super-Resolution Methods

Reference-based super-resolution is a subfield within the field of super-resolution. In the prior studies, exemplified by [6, 13, 20], the input comprises a target LR image along with a corresponding reference HR image. Notably, the content of the reference HR image is typically closely associated with the content of the target LR image. To leverage this relationship, most methods in this category utilize a feature-matching procedure to identify or evaluate related patches between the reference and target images. Once corresponding patches are identified, the SR procedure learns high-resolution information directly from the matched patches in the reference image. These methods assume a strong correlation between the reference and target images. However, inaccuracies in the matching process may lead to unexpected results. Additionally, the creation of paired target and reference images imposes additional costs.

3 Proposed Method

We seek to harness the advantages of employing reference HR images in addressing zero-shot blind SR tasks. The following section provides an overview of our design and showcases instances of implementation. It illustrates the process of integrating reference images into our network based on the target image. Furthermore, our training process only uses the target LR image and the provided irrelevant reference HR images. Specifically, our method yields an SR model tailored to the target image, making it a zero-shot super-resolution method.

3.1 Overview and Pipeline

We propose a novel procedure to optimize the super-resolution task. As illustrated in Fig. 2, our method involves four key steps. The first step (Fig. 2(a)) involves identifying a preliminary degradation process from the target LR image X_{lr-tar} . This degradation process is embedded and can be represented either implicitly or explicitly in downsampling network G_{dn} and upsampling network G_{up} , denoted by $K_{H \rightarrow L}$ and $K_{L \rightarrow H}$, respectively. Specifically, $K_{H \rightarrow L}$ performs the mapping from HR to LR image, while $K_{L \rightarrow H}$ performs the mapping from LR to HR image.

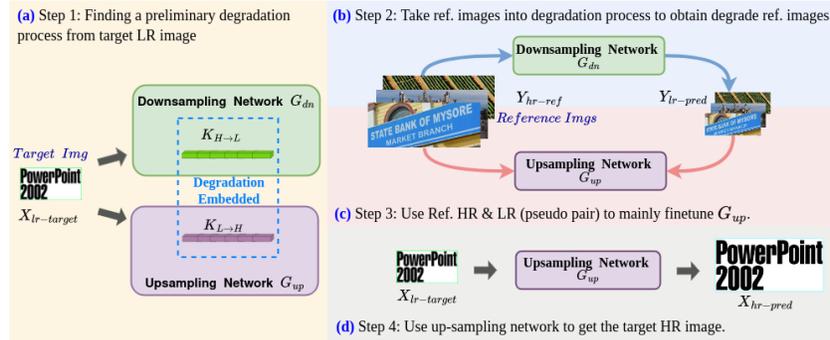


Fig. 2: Overview of our Reference Degradation Super-Resolution Architecture. The diagram illustrates the RDSR framework, showcasing the sequential steps. (a) The degradation estimator first identifies the preliminary degradation process. (b) Reference images are introduced to capture the scale-aware degradation process. (c)(d) The reference image pairs, generated by the downsampling network, serve as a supervised signal to train the SR network, culminating in the production of high-quality super-resolution results.

Next, given the HR reference images, we aim to generate pseudo data through the downsampling network G_{dn} , which is initialized from the previous step, as depicted in Fig. 2(b). Unlike existing methods that utilize HR images with content similar to the LR images, our approach leverages irrelevant HR images. These irrelevant HR images Y_{hr-ref} are downsampled to generate corresponding LR images $Y_{lr-pred}$, which are then employed as a data augmentation technique during the training phase.

Subsequently, as illustrated in Fig. 2(c), these pseudo paired data Y_{hr-ref} and $Y_{lr-pred}$ are primarily employed for fine-tuning the upsampling network G_{up} . This process allows G_{up} to progressively adapt to the reference image scale by refining the degradation representation $K_{L \rightarrow H}$ through constraints derived from the degradation of the target LR image and the downsampled reference HR images. This fine-tuning results in substantial improvements in super-resolution (SR) performance. Additionally, since the downsampling network G_{dn} is specifically initialized for the LR target images, it is also simultaneously finely adjusted to align the degradation representation $K_{H \rightarrow L}$ with that of the upsampling network $K_{L \rightarrow H}$.

Finally, following comprehensive fine-tuning, the upsampling network G_{up} is utilized for downstream tasks that take LR target image (X_{lr-tar}) as input and generate corresponding HR image ($X_{hr-pred}$), as illustrated in Fig. 2(d). Importantly, our approach offers the flexibility to interchange the downsampling network G_{dn} and the upsampling network G_{up} .

3.2 Implementation

In this section, we describe an instance of our procedure in term RDSR model, which utilizes KernelGAN [2] and DASR [19] architecture as the backbone for downsampling network G_{dn} and upsampling network G_{up} respectively. Based on our procedure described at section 3.1, to simply, we divide the implementation into two parts: Initial Phase and Fine-tune Phase, as illustrated in Fig. 3.

In the initial phase as depicted in Fig. 3(a), corresponding to Step 1, we aim to initialize the weights of G_{up} and G_{dn} to tailor them for the degradation process of the target LR image. Specifically, we employ a pre-trained DASR model [19] G_{up} , which includes a kernel estimator $E_k_{L \rightarrow H}$ to estimate the implicit degradation representation $K_{L \rightarrow H}$ and an upscaler network to generate the upscaled image based on the LR input and the obtained $K_{L \rightarrow H}$. In contrast, KernelGAN [2] G_{dn} , which explicitly extracts the kernel to compute the degradation representation $K_{H \rightarrow L}$ for downsampling the image, is randomly initialized and roughly optimized by using target LR and target HR pairs. These paired data are generated by utilizing the pre-trained G_{up} and the target LR image X_{lr-tar} to produce the target HR image $X_{hr-pred}$.

In the finetune phase, corresponding to Steps 2, 3, and 4, our objective is to generate additional LR-HR pairs by downsampling content-irrelevant reference HR images and utilizing them as references to establish scale-aware degradation kernels. As illustrated in Figure 3(b), G_{up} and G_{dn} form two branches: the target branch and the reference branch, creating a dual-path pipeline inspired by CycleGAN [25], with shared weights.

In the reference branch, the HR input is initially downsampled by G_{dn} and then upsampled back to the original scale by G_{up} . In contrast, the target branch first upsamples the target LR image using G_{up} and then downscales it back to the original LR size using G_{dn} . This process is designed to integrate HR image details into the model through the reference branch and ensure that the model maintains optimal parameters for processing the target LR images through the target branch. These objectives can be addressed by incorporating consistency loss along with constraints on the degradation representation, helping the model establish scale-aware degradation based on the HR images. Additionally, the discriminator D_{up} is incorporated to distinguish between real HR images and upscaled target LR images produced by G_{up} . Ideally, the G_{up} and G_{dn} should act as inverses of each other in the target and reference branches. Therefore, leveraging reference HR images at a specific scale as additional data enhances the super-resolution performance for downstream tasks on target LR images.

3.3 Training Objective

In this section, we outline our training objectives, which encompass implementing cycle consistency training and integrating both the target and reference branches. Additionally, we utilize the adversarial network D_{up} to align the output of the upsampling network with the HR distribution. To maintain consistency in the degradation process across both branches, we introduce a regularization

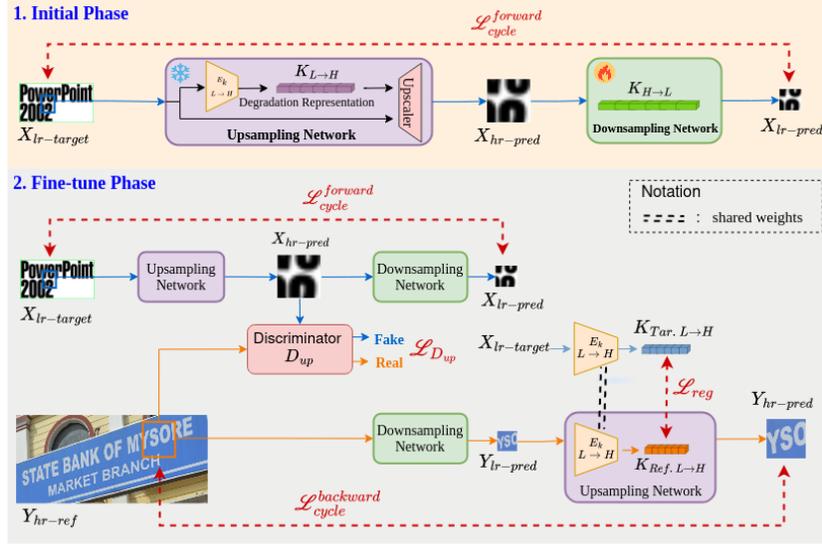


Fig. 3: Overview of RDSR Implementation.

loss for the degradation representation. Detailed descriptions of these specific training objectives will be provided in the following paragraphs. Note that both LR and HR images are processed as patches throughout the entire process.

Cycle consistency loss. The cycle consistency loss \mathcal{L}_{cycle} consists of forward and backward cycle consistency losses during the training process of RDSR. It comprises the Charbonnier loss [10] and the perceptual loss [7], utilizing an ImageNet-pretrained VGG-19 network denoted as $\phi(\cdot)$. The forward cycle is performed on the target branch to facilitate minimizing the disparity between the original target LR image x and the cycle-generated target LR image. Ideally, this implies that $G_{dn}(G_{up}(x)) = x$, ensuring that the model maintains optimal parameters for the target images while injecting HR information. The forward cycle consistency loss can be formulated as follows:

$$\mathcal{L}_{cycle}^{forward} = \mathbb{E}_x[\sqrt{(G_{dn}(G_{up}(x)) - x)^2 + \epsilon^2} + \|\phi_n(G_{dn}(G_{up}(x))) - \phi_n(x)\|_1], \quad (2)$$

To establish scale-awareness effectively from HR images y into the model, applying the backward cycle consistency loss is crucial. Ideally this aims for $G_{up}(G_{dn}(y)) = y$. This loss is performed on the reference branch and can be written as:

$$\mathcal{L}_{cycle}^{backward} = \mathbb{E}_y[\sqrt{(G_{up}(G_{dn}(y)) - y)^2 + \epsilon^2} + \|\phi_n(G_{up}(G_{dn}(y))) - \phi_n(y)\|_1], \quad (3)$$

where ϵ is set empirically to 10^{-6} during the experiments and $\phi_n(\cdot)$ is the feature at layer n of loss network ϕ . The total cycle consistency loss is then the sum of the forward and backward cycle consistency losses :

$$\mathcal{L}_{cycle} = \lambda_{cycle}^{forward} \mathcal{L}_{cycle}^{forward} + \lambda_{cycle}^{backward} \mathcal{L}_{cycle}^{backward}. \quad (4)$$

Generative and Discriminator loss. Taking inspiration from SRGAN [11], we design an adversarial loss for the upsampling network G_{up} to encourage it to generate more detailed images similar to the reference image. The adversarial loss for the generator is defined as follows:

$$\mathcal{L}_{gan} = \min_{G_{up}} \mathbb{E}_x [(D_{up}(G_{up}(x)) - 1)^2]. \quad (5)$$

On the other hand, the discriminator D_{up} aims to distinguish fake generated high-resolution from real reference high-resolution images. Its objective is defined as follows:

$$\mathcal{L}_{D_{up}} = \min_{D_{up}} (\mathbb{E}_y [(D_{up}(y) - 1)^2] + \mathbb{E}_x [D_{up}(G_{up}(x))^2]). \quad (6)$$

Regularization loss. To ensure consistency in the degradation process between target image x and reference image y , we assume that the degradation representation $K_{L \rightarrow H}$ of the target LR image from the target branch should ideally be similar to that of the downsampled reference image in the reference branch, implying $E_{k \ L \rightarrow H}(x) = E_{k \ L \rightarrow H}(G_{dn}(y))$. Therefore, we introduce a regularization loss \mathcal{L}_{reg} , computed with L1 distance in our framework.

$$\mathcal{L}_{reg} = \mathbb{E}_{x,y} [||E_{k \ L \rightarrow H}(x) - E_{k \ L \rightarrow H}(G_{dn}(y))||_1]. \quad (7)$$

Finally, the total loss is defined as the sum of all components:

$$\mathcal{L}_{total} = \mathcal{L}_{cycle} + \lambda_{reg} \mathcal{L}_{reg} + \mathcal{L}_{gan}, \quad (8)$$

where λ values control the importance of the different loss function terms.

4 Experiments and Results

4.1 Datasets and Implementation Details

For evaluation of SR performance, we use the datasets DIV2K [2] as target images and DIV2K [1] as reference images for our experiments. DIV2K contains 100 images, LR images blurred with randomly generated kernels, which are 11x11 anisotropic Gaussians. DIV2K contains 800 images and serves as a reference collection in our experiments.

Although our method implies recovering to multiple scaling factors, for better comparison with other methods, we adopt scales commonly used $\times 2$ and $\times 4$ to

evaluate the results. To evaluate our method’s performance, we adopted PSNR (Peak Signal-to-Noise Ratio) on Y channel of the images after being converted to the YCbCr color space and SSIM (Structural Similarity Index Measure).

In our experiments, we train 3000 iterations with a frozen G_{up} in the first stage to obtain the preliminary downsampling network G_{dn} . Subsequently, G_{up} is unfrozen to train 500 iterations per reference image on $E_k_{L \rightarrow H}$ and the up-scaling network, respectively. Concurrently, G_{dn} is fine-tuned with minor adjustments. The initial learning rates are set 2×10^{-3} and 1×10^{-5} for G_{dn} and G_{up} respectively. We utilize the Adam [9] optimizer with $\beta_1 = 0.25$ and $\beta_2 = 0.99$. Additionally, λ_{cycle}^{target} is set to 5, λ_{cycle}^{ref} to 1, and λ_{reg} to 20 based on our experiments for the whole training. The process for a LR image requires approximately 30min on a single RTX 2080Ti GPU.

Our RDSR method can generate multiple results for a target image utilizing various input reference images. To improve the quality of our training, we monitor specific criteria throughout the training process. Specifically, we observe the cycle consistency loss of the target image, deciding when to preserve the network parameters. Furthermore, we integrate a no-reference quality assessment index BRISQUE [16] to guide the decision of whether to save the output of the network.

4.2 Experimental Results and Comparison

Our model utilizes well-trained DASR [19] models as the backbone. These DASR models consist of $\times 2$ and $\times 4$ settings, which are trained using isotropic and anisotropic kernels, respectively. As shown in Table 1, we have achieved superior results in terms of PSNR and SSIM compared to the baseline DASR in both isotropic and anisotropic kernel types at scales 2 and 4. Furthermore, our method surpasses Bicubic and typical zero-shot super-resolution methods such as DualSR [4], RZSR [21] in PSNR. Additionally, as demonstrated in Fig. 4, while bicubic interpolation results in blurry and oversmoothed images, DASR [19] yields more detailed results. Our method produces the sharpest and clearest results among them. It excels in restoring fine details and achieves visually superior results compared to both the baseline DASR [19] and zero-shot methods, owing to its effective scale awareness.

4.3 Ablation Study

In this section, we study our RDSR model with different loss settings to evaluate their impact. Additionally, we investigate the importance of finding the preliminary degradation process from target LR images during the model initialization phase through experiments with ground truth (GT) kernels. Furthermore, one of our key contributions is the integration of HR images, where the characteristics of HR images—such as their types and quantity—play a critical role in influencing the final outcomes. Therefore, we analyze the impact of HR image selection methods and the number of reference images used in our study.

Table 1: Quantitative comparison on DIV2KRRK, we evaluate both blind super-resolution methods and zero-shot super-resolution methods.

Scale	Method	PSNR-Y	SSIM
×2	Bicubic	28.829	0.8139
	DualSR [4]	30.92	0.8728
	RZSR [21]	30.55	0.8704
	DASR [19] <i>isotropic</i> + RDSR (Ours)	30.743	0.8586
		30.973	0.8638
	DASR [19] <i>anisotropic</i> + RDSR (Ours)	31.051	0.8668
	31.161	0.8683	
×4	Bicubic	25.471	0.7088
	DASR [19] <i>isotropic</i> + RDSR (Ours)	27.441	0.7635
		27.560	0.7663
	DASR [19] <i>anisotropic</i> + RDSR (Ours)	28.142	0.7769
	28.233	0.7794	

Table 2: Ablation Study with RDSR ×2 with isotropic kernel on DIV2KRRK.

Method	\mathcal{L}_{ref}	\mathcal{L}_{gan}	\mathcal{L}_{reg}	PSNR	SSIM
DASR <i>iso</i>	×	×	×	30.743	0.8586
RDSR <i>iso</i> (<i>baseline</i>)	✓	×	×	30.876	0.8609
RDSR <i>iso</i> + \mathcal{L}_{gan}	✓	✓	×	30.927	0.8621
RDSR <i>iso</i> + \mathcal{L}_{gan} + \mathcal{L}_{reg}	✓	✓	✓	30.973	0.8638

The Designed Loss Functions. Table 2 demonstrates the effectiveness of our designed loss functions. Experimental results on the DIV2K×2 dataset under isotropic settings show that integrating the adversarial loss \mathcal{L}_{gan} improved performance by 0.05 dB in PSNR and 0.0012 in SSIM compared to the baseline. Furthermore, the inclusion of the regularization loss \mathcal{L}_{reg} resulted in an additional gain of 0.046 dB in PSNR and 0.0017 in SSIM.

Ground-truth Kernel. In our architecture, the performance is minimally impacted by the content of the reference images. However, the degradation process has been identified as a critical factor. Table 3 illustrates the substantial performance improvements achieved by using the ground truth kernel in place of the downsampling network G_{dn} . If a more effective kernel estimation method is discovered, our approach can readily integrate it by substituting the new method into the downsampling network.

Comparison of Reference Image Selection Methods. As mentioned earlier, the reference images employed in our methods may not inherently contain content relevant to the target images. Although the specific content of the reference images has minimal impact on our method, the selection of a reference image introduces subtle variations in the outcome of the target image. In response to

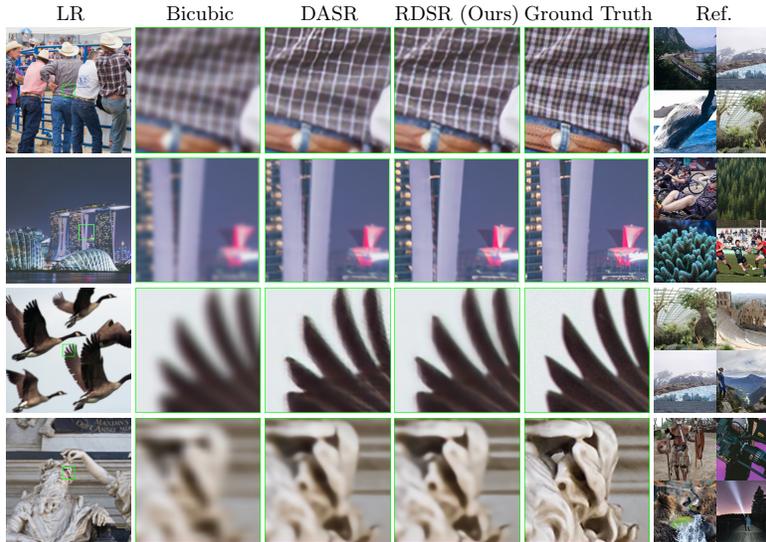


Fig. 4: Qualitative comparison of the RDSR based on "Blind Super Resolution" method. The order from top to bottom sequentially present the results of isotropic and anisotropic kernels at $\times 2$ and $\times 4$ scaling factors, respectively.

Table 3: Comparison of ground truth kernel on RDSR $\times 2$ with anisotropic kernel.

Method	scale	PSNR	SSIM
DASR <i>aniso</i>	$\times 2$	31.051	0.8668
RDSR <i>aniso</i>	$\times 2$	31.161	0.8683
RDSR <i>aniso</i> + GT kernel	$\times 2$	33.144	0.8994

Table 4: Comparison of reference image selection policies on DIV2K $\times 2$ datasets.

Method & Datasets	Ref. Selection Method	PSNR-Y	SSIM
DualSR + Ref. add-on DIV2K $\times 2$	auto-selection	26.228	0.7353
	random selection	26.190	0.7340
	reverse auto-selection	26.164	0.7338

this, we have devised an automatic selection method to choose the reference image from the high-resolution collection. This method involves calculating the mean values of the reference images in the RGB channels, respectively, and selecting the closest match to the target image. In Table 4, we compare three distinct selection policies: auto-selection, random selection, and reverse auto-selection, wherein the farthest images are chosen as the reference images—an inverse approach to auto-selection. The results indicate that selecting HR images with similar color distribution (automatic selection method) leads to better SR outcomes.

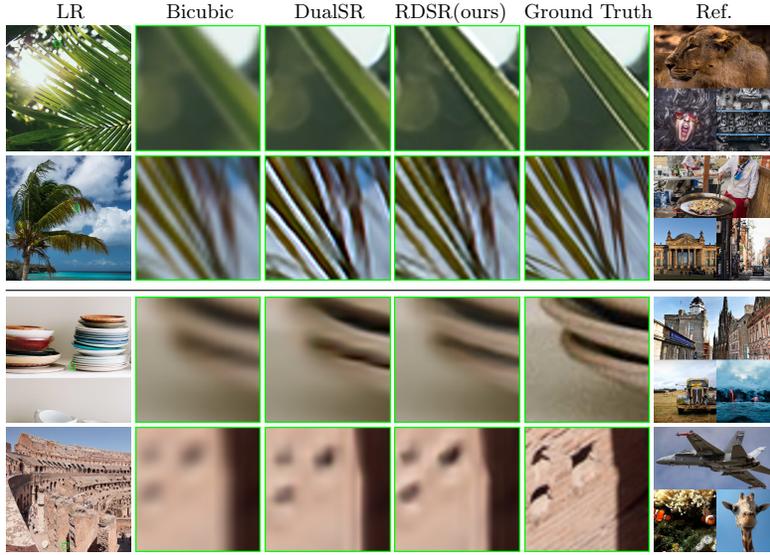


Fig. 5: Qualitative comparison of DualSR with the reference add-on module. The first two rows display results for a scale of $\times 2$, while the subsequent two rows show results for a scale of $\times 4$.

Comparison of the Number of Reference Images. Our method acquires domain knowledge from the provided reference HR images. To investigate the optimal number of reference images (denoted as N) required to impart sufficient domain knowledge, we conducted experiments using 1, 3, and 5 reference images. As depicted in Table 5, we observed a consistent trend where increasing the number of reference images correlates positively with improved SR performance. By increasing the number of reference images, the model gains deeper insights into a broader range of image characteristics, thereby leading to enhanced super-resolution (SR) results.

4.4 Study of Reference-based Module on ZSSR Method

With our proposed procedure for integrating HR images into the model, we extend this approach to zero-shot super-resolution (ZSSR) methods. In our experiments, we integrate the reference add-on module into the DualSR [4] training network. The DIV2K dataset is used as our target image dataset, and we select reference images from DIV2KRRK. The PSNR and SSIM values are reported in Table 6, which demonstrates the benefits of applying the reference add-on module. Fig. 5 provides a visual comparison between our results and those obtained from DualSR, highlighting the effective reduction of artifacts achieved by our methods.

Table 5: Quantitative analysis of the impact of the number (N) of reference images.

Method & Datasets	N	PSNR-Y	SSIM
DualSR + Ref. add-on	1	30.594	0.8551
DIV2KRRK $\times 2$	3	30.880	0.8622
	5	30.918	0.8636

Table 6: Quantitative comparison of reference add-on module based on DualSR [4] on Track 2 of DIV2K.

Method	Scale	PSNR-Y	SSIM
DualSR	$\times 2$	26.011	0.7310
DualSR + Ref. add-on module	$\times 2$	26.228	0.7353
DualSR	$\times 4$	23.079	0.6466
DualSR + Ref. add-on module	$\times 4$	23.230	0.6388

5 Conclusion

We present a novel and versatile approach to significantly enhance super-resolution (SR) performance by leveraging additional high-resolution images. Our method utilizes these supplementary images to facilitate learning a scale-specific degradation process. This approach distinguishes itself from conventional reference-based SR methods, which typically rely on reference images closely resembling the target image. Moreover, our proposed method is applicable in both well-trained blind SR and zero-shot SR scenarios. Additionally, our procedure offers flexibility in selecting the network backbone, enabling the substitution of kernel estimators and SR networks with more advanced techniques tailored specifically to the target image. Experimental results demonstrate that our specific instance, RDSR, surpasses baseline blind super-resolution algorithms in terms of both PSNR and SSIM, consistently delivering visually superior results across various scenarios.

Acknowledgments This work was financially supported in part (project number: 112UA10019) by the Co-creation Platform of the Industry Academia Innovation School, NYCU, under the framework of the National Key Fields Industry-University Cooperation and Skilled Personnel Training Act, from the Ministry of Education (MOE) and industry partners in Taiwan. It also supported in part by the National Science and Technology Council, Taiwan, under Grant NSTC-112-2221-E-A49-089-MY3, Grant NSTC-110-2221-E-A49-066-MY3, Grant NSTC-111-2634-F-A49-010, Grant NSTC-112-2425-H-A49-001, and in part by the Higher Education Sprout Project of the National Yang Ming Chiao Tung University and the Ministry of Education (MOE), Taiwan. And we would like to express our gratitude for the support from MediaTek Inc.

References

1. Agustsson, E., Timofte, R.: Ntire 2017 challenge on single image super-resolution: Dataset and study. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 126–135 (2017)
2. Bell-Kligler, S., Shocher, A., Irani, M.: Blind super-resolution kernel estimation using an internal-gan. *Advances in Neural Information Processing Systems* **32** (2019)
3. 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)
4. Emad, M., Peemen, M., Corporaal, H.: Dualsr: Zero-shot dual learning for real-world super-resolution. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 1630–1639 (2021)
5. 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)
6. Jiang, Y., Chan, K.C., Wang, X., Loy, C.C., Liu, Z.: Robust reference-based super-resolution via c2-matching. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2103–2112 (2021)
7. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part II* 14. pp. 694–711. Springer (2016)
8. 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)
9. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014)
10. Lai, W.S., Huang, J.B., Ahuja, N., Yang, M.H.: Fast and accurate image super-resolution with deep laplacian pyramid networks. *IEEE transactions on pattern analysis and machine intelligence* **41**(11), 2599–2613 (2018)
11. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4681–4690 (2017)
12. 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)
13. Lu, L., Li, W., Tao, X., Lu, J., Jia, J.: Masa-sr: Matching acceleration and spatial adaptation for reference-based image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6368–6377 (2021)
14. Lugmayr, A., Danelljan, M., Timofte, R.: Unsupervised learning for real-world super-resolution. In: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). pp. 3408–3416. IEEE (2019)
15. Maeda, S.: Unpaired image super-resolution using pseudo-supervision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 291–300 (2020)

16. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. *IEEE Transactions on image processing* **21**(12), 4695–4708 (2012)
17. Qiu, Y., Wang, R., Tao, D., Cheng, J.: Embedded block residual network: A recursive restoration model for single-image super-resolution. In: *Proceedings of the IEEE/CVF international conference on computer vision*. pp. 4180–4189 (2019)
18. Shocher, A., Cohen, N., Irani, M.: “zero-shot” super-resolution using deep internal learning. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 3118–3126 (2018)
19. 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)
20. Yang, F., Yang, H., Fu, J., Lu, H., Guo, B.: Learning texture transformer network for image super-resolution. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 5791–5800 (2020)
21. Yoo, J.S., Kim, D.W., Lu, Y., Jung, S.W.: Rzsrr: Reference-based zero-shot super-resolution with depth guided self-exemplars. *IEEE Transactions on Multimedia* (2022)
22. Zhang, K., Gool, L.V., Timofte, R.: Deep unfolding network for image super-resolution. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 3217–3226 (2020)
23. Zhang, K., Zuo, W., Zhang, L.: Learning a single convolutional super-resolution network for multiple degradations. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 3262–3271 (2018)
24. Zhou, Y., Lin, C., Luo, D., Liu, Y., Tai, Y., Wang, C., Chen, M.: Joint learning content and degradation aware feature for blind super-resolution. In: *Proceedings of the 30th ACM International Conference on Multimedia*. pp. 2606–2616 (2022)
25. 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)