

# Pairwise Distance Distillation for Unsupervised Real-World Image Super-Resolution

Yuehan Zhang<sup>1</sup>, Seungjun Lee<sup>2</sup>, and Angela Yao<sup>1</sup>

<sup>1</sup> National University of Singapore  
{zyuehan, ayao}@comp.nus.edu.sg

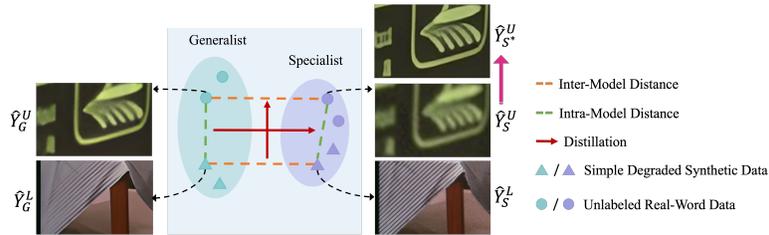
<sup>2</sup> Korea University  
9penguin9@korea.ac.kr

**Abstract.** Standard single-image super-resolution creates paired training data from high-resolution images through fixed downsampling kernels. However, real-world super-resolution (RWSR) faces unknown degradations in the low-resolution inputs, all the while lacking paired training data. Existing methods approach this problem by learning blind general models through complex synthetic augmentations on training inputs; they sacrifice the performance on specific degradation for broader generalization to many possible ones. We address the unsupervised RWSR for a targeted real-world degradation. We study from a distillation perspective and introduce a novel pairwise distance distillation framework. Through our framework, a model specialized in synthetic degradation adapts to target real-world degradations by distilling intra- and inter-model distances across the specialized model and an auxiliary generalized model. Experiments on diverse datasets demonstrate that our method significantly enhances fidelity and perceptual quality, surpassing state-of-the-art approaches in RWSR. The source code is available at <https://github.com/Yuehan717/PDD>.

## 1 Introduction

Single-image super-resolution (SISR) predicts high-resolution (HR) images from low-resolution (LR) counterparts. The standard SISR model addresses predefined downsampling kernels, *e.g.* bicubic interpolation. However, real-world scenarios of SISR (RWSR) encompass unknown degradations in LR images, including blur, noise, and JPEG compression artifacts [35, 43, 45] with diverse combinations. The various and unknown degradations pose additional challenges when learning an RWSR model.

Existing RWSR methods focus on blind generalization. Such methods synthesize paired training data with extensive and harsh degradations, involving multiple rounds of random blurring, added noise, resizing, and compression [35, 43]. The complex degradation pipeline allows the model to generalize to diverse unknown conditions. We refer to such a model addressing various degradations as a *generalist*. However, as much of the model’s capacity is devoted to handling multiple conditions, these “jack-of-all-trades” models come with performance trade-



**Fig. 1:**  $\hat{Y}_G^U$  and  $\hat{Y}_G^L$  are reconstructions for real-world and bicubic-interpolated (BI) inputs using a blind generalized model (generalist);  $\hat{Y}_S^U$  and  $\hat{Y}_S^L$  are counterparts using a standard SR model (bicubic specialist). The specialist enhances the BI inputs with clearer details, while the generalist does better for the real-world one. We distill the intra- and inter-model distances for an improved real-world reconstruction  $\hat{Y}_S^U$ .

offs [45], *i.e.*, they have inferior performance compared to the model only optimized for the tested degradation ( $\hat{Y}_G^L$  and  $\hat{Y}_S^L$  in Fig. 1).

SR models can be optimized for specific and known input degradations through supervised learning [9, 11, 18, 36]. The LR-HR pairs for training are created by degrading HR images with fixed kernels, *e.g.*, bicubic interpolation, and Gaussian blur. We refer to models adept at the specific input degradation as a *specialist*. Degradations in RWSR datasets [2, 39] always belong to a specific domain [45]. For example, images captured by a specific camera model tend to feature consistent sensor noise [2]. However, the formulation of a real-world degradation is always unknown, posing difficulties in creating paired training data. As such, learning specialist models for such real-world domains is challenging, and strategies that do not require direct supervision are crucial.

This paper tackles unsupervised RWSR from a knowledge distillation perspective. Distillation has been successful for domain adaptation in high-level tasks [6, 22, 38], but it is under-explored in low-level vision. Our distillation framework adapts a specialist model of known synthetic degradation to target real-world degradations. We use an auxiliary generalist model to provide generalization knowledge of real-world degradations that the specialist lacks. However, naive distillation only transfers knowledge from the generalist model; we aim to improve beyond the generalist’s moderate performance. To that end, we also consider the low-level characteristics of the specialist’s predictions on synthetic samples as a reference for the distillation.

Our method uses the strength of both the specialist and the generalist by exploring the relationships between their predictions. Given a generalist and a specialist for synthetic degradation, we consider predictions of synthetic samples and real-world samples from both models (see Fig. 1). The generalist model has smaller gaps in qualities of real-world and synthetic predictions, featuring close low-level characteristics between domains [20, 21]. In contrast, the specialist outputs distanced low-level characteristics but high-quality predictions for synthetic samples ( $\hat{Y}_S^L$  in Fig. 1), providing valuable low-level characteristics for reference.

We propose to let the specialist imitate prediction *relationships* from the generalist - that is, push the specialist’s real-world predictions ( $\hat{Y}_S^U$  in Fig. 1) to have low-level characteristics similar to its high-quality synthetic predictions. If such similarity is satisfied by the specialist, it would exhibit two consistencies. First, for two input samples, the difference between their predictions from the same model is intra-model distance; such distances would be consistent for the specialist and the generalist. Secondly, for a given input sample, the difference between generalist and specialist predictions is inter-model distance; such distances would be consistent whether the sample is synthetic or real-world. These intra- and inter-model distance consistencies form the basis of our pairwise distance distillation framework. Together with designs on configurations of specialist and generalist models, our adapted specialist model experimentally shows significant improvements over the generalist model.

**Contributions.** To summarize, the contributions of our work are three-fold:

- To the best of our knowledge, we are the first to propose a distillation perspective for unsupervised RWSR to combine generalist and specialist models.
- We propose a novel pairwise distance distillation framework, emphasizing transferring intra-model and inter-model distances to enhance the specialist’s performance in real-world scenarios.
- Experiments on three benchmarks demonstrate that our approach improves the off-the-shelf models regarding fidelity and perceptual performances.

## 2 Related Works

**Single-Image Super-Resolution (SISR)** methods use deep neural networks of various architectures, such as residual networks [19, 36] and transformers [18, 33]. Despite improving the standard SISR performance with dedicated architectures, these methods struggle to generalize to real-world scenarios [34, 40].

**Real-World Image Super-Resolution (RWSR)** aims to address unknown input degradations. The infeasibility of creating paired training data without known kernels presents a significant challenge in learning specialized models for RWSR. There are two primary strategies in existing RWSR. The first assumes that networks trained on diverse and challenging synthetic degradations will effectively generalize to real-world data [35, 43]. Even though real-world degradations are complex, within a specific setting or deployment, they tend to be limited in the domain and less diverse than the training pipelines. Consequently, [45] proposes customizing the synthetic pipeline to better match specific real-world data. Nonetheless, an unresolved gap exists between synthesized and real-world degradations. Instead, our method operates directly on real-world data and avoids the intrinsic gap from synthetic degradation.

The second strategy connects real-world and synthetic degradation with transfer learning. Most works use image-to-image translation [5, 30, 40] to adapt the synthetic degradation to real-world ones. Such methods resort to intricate designs like CycleGAN [50], which struggle to replicate real-world degradations [34]

reliably. Our proposed method explores feature distances between model outputs and bypasses the need to mimic real-world degradations.

**Knowledge Distillation** is initially introduced for model compression [10]. In low-level vision, several works [7, 28, 47] distill large models into more efficient ones by enforcing similarity between their internal features or predictions. Recently, knowledge distillation has served diverse purposes in high-level tasks like segmentation, including domain adaptation [6, 15, 27] and transfer learning [38, 41]. However, such applications are less discussed in low-level vision.

### 3 Method

Our method adapts the specialist in synthetic degradation to an unlabeled real-world domain. We use the low-level characteristics of synthetic predictions from the specialists and the knowledge of the generalist model. Fig. 2 is the overview of our approach and the novel Pairwise Distance Distillation (PDD) method, which enforces the consistency of the intra- and inter-model distances to improve the specialist ( $M_S$ )’s real-world predictions. Sec. 3.1 gives notions used in this section and Sec. 3.2 illustrates the unsupervised formulation. Sec. 3.3 explains the details of PDD and Sec. 3.4 provides the full optimization aims and discusses the static and Exponential Moving Average (EMA) versions of our method.

#### 3.1 Definitions & Setup

When provided with an LR input  $X \in \mathbb{R}^{H \times W \times C}$  - where  $H \times W$  is the spatial resolution and  $C$  is the number of color channels - a SISR model aims to reconstruct an HR image  $\hat{Y} \in \mathbb{R}^{rH \times rW \times C}$ , scaling by a factor  $r$ . Due to the difficulty in obtaining paired training data, the conventional approach in SISR involves generating an LR image  $X$  by downsampling the HR ground-truth  $Y$ , *i.e.*,  $X = D_{\downarrow}(Y)$ . This downsampling operation  $D_{\downarrow}$  typically refers to bicubic interpolation, although other alternatives exist [1].

In a more extensive setting, it is typical for the LR image to be affected by various factors, such as blurring, noise, compression artifacts, *etc.* To exhibit these degradations, the LR image  $X$  can be formed by applying  $D_{\downarrow}$  along with one or multiple degradations  $D_i$ , in any order, *i.e.*,

$$X = D_1 \circ \dots \circ D_n(Y). \quad (1)$$

Here, we include  $D_{\downarrow}$  as one of the degradations  $\{D_i\}_{i \in [1, n]}$  for simplicity. Now, let’s consider two categories of degradation sets consisting of known degradations:  $\mathcal{D}_S$ , which comprises simple degradations with a few factors, and  $\mathcal{D}_G$  characterized by complex degradations with a wide range of factors:

$$\mathcal{D}_S = \{D_i\}_{i \in \mathcal{S}}, \quad \mathcal{D}_G = \{D_i\}_{i \in \mathcal{G}}, \quad \text{where } |\mathcal{G}| \gg |\mathcal{S}|. \quad (2)$$

In the simplest case for  $\mathcal{D}_S$ , the set  $\mathcal{D}_S = \{D_{\downarrow}\}$ , which represents the standard SISR setup. For the broader case,  $\mathcal{D}_G$  should encompass diverse and challenging

degradations beyond  $\{D_{\downarrow}\}$  to ensure that models trained on this data generalize well to blind settings [35, 43].

Consider two models,  $M_S$  and  $M_G$ , trained on datasets constructed with  $\mathcal{D}_S$  and  $\mathcal{D}_G$  respectively. Both models are optimized with common supervised SR learning losses [13, 36] including MSE-based loss, VGG loss, and adversarial training (see Eq. (9)). Since  $\mathcal{D}_S$  comprises a narrower range of degradations compared to  $\mathcal{D}_G$ ,  $M_S$  is a “*specialist*” model for the limited domain defined by  $\mathcal{D}_S$ . On the other hand,  $M_G$  is a “*generalist*” model as it achieves moderate performance across the broader domain derived by  $\mathcal{D}_G$ . Comparing  $M_S$  and  $M_G$ , we assume the specialist’s superiority to the generalist in  $\mathcal{D}_S$  and the generalist’s better handling of unknown degradations. These assumptions are empirically validated in Tab. 1.

### 3.2 Unsupervised Learning Through Distillation

Our approach aligns with the unsupervised setup [5, 30, 40], wherein we work with a collection of real-world LR images alongside unpaired HR clean images. We denote the set of LR images as  $\{X^U\}$  with the same unknown degradations  $\{D_i\}_{i \in \mathcal{U}}$ . Additionally, we generate LR counterparts  $\{X^L\}$  from HR images using specific degradations from  $\mathcal{D}_S$ , *e.g.* bicubic interpolation. The specialist  $M_S$  is adept in restoring  $X^L$ , while  $M_G$  performs moderately in restoring  $X^L$  and  $X^U$ . We aim to adapt  $M_S$  to  $\{X^U\}$  with the aid of  $M_G$ . As in [10], a naive distillation approach is having  $M_S$  imitate the predictions from  $M_G$  for unlabeled inputs:

$$\begin{aligned} \hat{Y}_S^L &= M_S(X^L), \hat{Y}_S^U = M_S(X^U), \hat{Y}_G^U = M_G(X^U), \\ \mathcal{L}_{ND} &= \mathcal{L}_L(\hat{Y}_S^U, \hat{Y}_G^U) + \lambda \mathcal{L}_L(\hat{Y}_S^L, Y^L), \end{aligned} \quad (3)$$

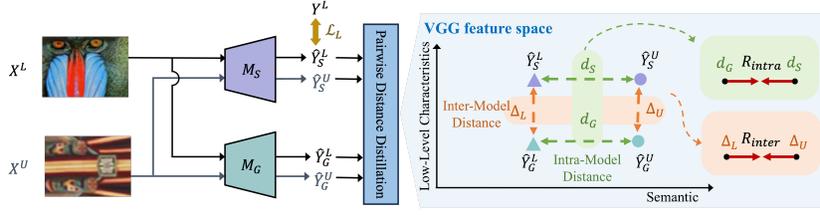
where  $Y^L$  represents the ground-truth image in the labeled domain,  $\mathcal{L}_L$  denotes the supervised loss formulated in Eq. (9), and  $\lambda$  is a scale factor to balance between the distillation and primary objective of  $M_S$ . However, this simple imitation approach relies solely on information from the generalist  $M_G$  and fails to harness the strengths of both models.

### 3.3 Pairwise Distance Distillation (PDD)

Consider any pair of LR images - a real-world  $X^U$  and a synthesized  $X^L$ . As shown in Fig. 2, we obtain four predictions by applying  $M_S$  and  $M_G$ :

$$\begin{aligned} \hat{Y}_S^U &= M_S(X^U), \hat{Y}_G^U = M_G(X^U), \\ \text{and } \hat{Y}_S^L &= M_S(X^L), \hat{Y}_G^L = M_G(X^L). \end{aligned} \quad (4)$$

Our distillation explores the relationship between predictions in Eq. (4) to seek for knowledge combination. We use VGG features [31] of the predictions as our basis for exploration. VGG, as a classification model, inherently captures *semantic* information. Yet it also captures *low-level characteristics* related to



**Fig. 2:** Schematic of Pairwise Distance Distillation (PDD).  $X^U$  and  $X^L$  are unlabeled (real-world) and labeled (synthetic) inputs. Model  $M_G$  is a generalist trained with extensive synthetic pipeline, while  $M_S$  specializes in  $X^L$ . The prediction  $\hat{Y}_S^L$  is supervised by its ground truth throughout training. PDD enforces the consistency between  $\{\Delta_U, \Delta_L\}$  and between  $\{d_S, d_G\}$  to improve  $M_S$ 's real-world performance.

image quality [37], *e.g.* blur and sharpness, and is widely used in the perceptual loss for super-resolution tasks [13, 35, 36].

As a generalist,  $M_G$  outputs synthetic and real-world predictions with small quality gaps, which is characterized by the overlapping distributions of low-level characteristics of  $\{\hat{Y}_G^L\}$  and  $\{\hat{Y}_G^U\}$  [20, 21]. Following [20], Fig. 5a projects the low-level features of  $\{\hat{Y}_G^L\}$  (DIV2K) and  $\{\hat{Y}_G^U\}$  (NTIRE20). The generalist's coverage strongly overlaps while the specialist's predictions are well-separated (see Fig. 5b). The within-model relationship of the generalist exhibits the generalizability to real-world degradation.

On the other side, the high image quality of  $\{\hat{Y}_S^L\}$ , due to  $M_S$ 's specialization, makes its low-level characteristics a valuable reference. To integrate the knowledge for real-world performance, we propose to make the specialist imitate the generalist's within-model relationship, that is, pushing the low-level characteristics of real-world predictions from  $M_S$ ,  $\{\hat{Y}_S^U\}$ , to be similar to that of its synthetic predictions  $\{\hat{Y}_S^L\}$ . To achieve this aim, we first *assume* such similarity is satisfied for  $M_S$  and conclude the following consistencies about VGG feature distances between predictions:

1. Intra-model distances for predictions of the same input pair should be consistent across  $M_S$  and  $M_G$  (green shade in Fig. 2).
2. Inter-model distances for predictions of a single input should be consistent across synthetic and real-world domains (orange shade in Fig. 2).

Our rationales for these consistencies are as follows. First, we recall the low-level characteristics of predictions from the same model are similar. We discussed it for  $M_G$  with Fig. 5a and made such an assumption for  $M_S$ . Thus, the distance between the same model's predictions  $\{\hat{Y}_G^L, \hat{Y}_G^U\}$  (or  $\{\hat{Y}_S^L, \hat{Y}_S^U\}$ ), *i.e.* intra-model distance, reflects mainly semantic differences due to the close low-level characteristics. As it naturally follows that the applied SR models should not change semantics, intra-model distances should be consistent for  $M_S$  and  $M_G$ .

Second, we note the low-level characteristics of predictions from the two models differ, *e.g.*  $\hat{Y}_S^L$  has higher quality than the counterpart  $\hat{Y}_G^L$ . The distance

between the two predictions  $\{\hat{Y}_G^L, \hat{Y}_S^L\}$  (or  $\{\hat{Y}_G^U, \hat{Y}_S^U\}$ ) of a given sample, *i.e.* inter-model distance, captures such differences in low-level characteristics, as semantics stay constant for the predictions of the same input. Given the similarity of low-level characteristics within a single model, the inter-model distances for synthetic and real-world samples should be consistent.

We structure the two consistencies as distillations on the intra- and inter-model distances. Our method encourages  $M_S$ 's real-world predictions  $\{\hat{Y}_S^U\}$  to have similar low-level characteristics as its synthetic predictions  $\{\hat{Y}_S^L\}$ , taking it as a reference from high-quality images.

**Intra-model Distance Distillation** enforces consistency on the distances between predictions from the same model (green shading in Fig. 2). Consider the  $\ell_1$  distance  $d_G^{ij}$  between predictions from the generalist  $M_G$ , *i.e.*  $\hat{Y}_G^L$  and  $\hat{Y}_G^U$ , in VGG feature space, and the distance  $d_S^{ij}$  for the specialist  $M_S$  is similarly defined:

$$\begin{aligned} d_G^{ij} &= \|\Phi_{ij}(\hat{Y}_G^L) - \Phi_{ij}(\hat{Y}_G^U)\|_1, \\ d_S^{ij} &= \|\Phi_{ij}(\hat{Y}_S^L) - \Phi_{ij}(\hat{Y}_S^U)\|_1, \end{aligned} \quad (5)$$

where  $\Phi_{ij}$  refers to the  $j$ -th layer in the  $i$ -th residual block of VGG19. For  $d_G^{ij}, d_S^{ij} \in \mathbb{R}^{c \times h \times w}$ ,  $h \times w$  is the spatial resolution and  $c$  is the number of channels. We enforce the consistency between  $d_G^{ij}$  and  $d_S^{ij}$  by minimizing their difference measured with the Cross-Entropy (CE)  $R_{intra}$  :

$$R_{intra} = -\frac{1}{hw} \sum_{i,j} \sum_{m,n} \mathbf{S}(d_G^{ij}[m,n]) \log \mathbf{S}(d_S^{ij}[m,n]), \quad (6)$$

where  $[m,n]$  is a spatial index of the feature map, and  $\mathbf{S}(\cdot)$  denotes the SoftMax function with the input size being  $\mathbb{R}^c$ . The element-wise CE reflects the negative log-likelihood of distances locally. While other measures are feasible, CE is empirically a good choice in favoring the overall sharpness (see Sec. 4.4).

**Inter-model Distance Distillation** enforces the consistency of changes in low-level characteristics between predictions from different models (orange shading in Fig. 2). Given the two predictions for a single input, *i.e.*  $\{\hat{Y}_G^L, \hat{Y}_S^L\}$  (or  $\{\hat{Y}_G^U, \hat{Y}_S^U\}$ ), we first calculate their feature distance  $\Delta_L^{ij}$  (or  $\Delta_U^{ij}$ ) to represent the differences in low-level characteristics. However, enforcing consistency on  $\Delta_L^{ij}$  and  $\Delta_U^{ij}$  is less straightforward, as the spatial layout of corresponding inputs can differ. Thus, we compute the inter-model distances with the following:

$$\begin{aligned} \Delta_L^{ij} &= \text{Gram}(\Phi_{ij}(\hat{Y}_S^L) - \Phi_{ij}(\hat{Y}_G^L)), \\ \Delta_U^{ij} &= \text{Gram}(\Phi_{ij}(\hat{Y}_S^U) - \Phi_{ij}(\hat{Y}_G^U)). \end{aligned} \quad (7)$$

$\text{Gram}(\cdot)$  refers to the Gram matrix, which calculates the correlations among vectorized feature maps along the channel dimension. The Gram matrix captures the statistics while collapsing the spatial layout [8].

For distillation, we ensure the consistency between inter-model distances by minimizing  $R_{inter}$ , the Frobenius norm between  $\Delta_U^{ij}$  and  $\Delta_L^{ij}$ :

$$R_{inter} = \sum_{i,j} \|\Delta_U^{ij} - \Delta_L^{ij}\|_{\mathcal{F}}, \quad (8)$$

where  $\|\cdot\|_{\mathcal{F}}$  computes the Frobenius norm. The Frobenius norm of Gram Matrix difference is interpreted in [8, 16] as measuring distribution discrepancy.

### 3.4 Full Method

The network  $M_S$  is fully optimized with supervised losses for labeled synthetic data  $X^L$  and unsupervised losses for real-world data  $X^U$ .

**Supervised Losses** maintain the specialization in synthetic degradation and reduce overfitting to the distillation. With the ground-truth image for prediction  $\hat{Y}_S^L$ , denoted as  $Y^L$ , we implement a combination of supervised loss functions. It includes a wavelet-based loss  $L_{\text{wv}}$ , a perceptual loss  $L_{\text{vgg}}$ , and generative loss  $L_{\text{gan}}$  for adversarial training:

$$\mathcal{L}_L(\hat{Y}_S^L, Y^L) = \alpha_1 L_{\text{wv}}(\hat{Y}_S^L, Y^L) + \alpha_2 L_{\text{vgg}}(\hat{Y}_S^L, Y^L) + \alpha_3 L_{\text{gan}}(\hat{Y}_S^L), \quad (9)$$

where  $L_{\text{vgg}}$  and  $L_{\text{gan}}$  are adopted from RealESRGAN [35], with balancing weights  $\alpha_1$ ,  $\alpha_2$ , and  $\alpha_3$ . We use but omit writing the discriminator loss for brevity.  $L_{\text{wv}}$  refers to an  $\ell_1$ -based wavelet loss [4, 49]:

$$L_{\text{wv}}(\hat{Y}_S^L, Y^L) = \sum_i \omega_i \|W(\hat{Y}_S^L)_i - W(Y^L)_i\|_1, \quad (10)$$

where  $W(\cdot)_i$  extracts the  $i$ -th wavelet channel, and  $\omega_i$  weights its importance. For the wavelet transform, we apply the Haar wavelet [14] for all output channels.

**Unsupervised Losses.** In the absence of ground truth for  $\hat{Y}_S^U$ , we opt to optimize for the consistency outlined in Eq. (6) and Eq. (8) as regularizers. To fully utilize the discriminator’s knowledge of realness, we also incorporate the generative loss in Eq. (9):

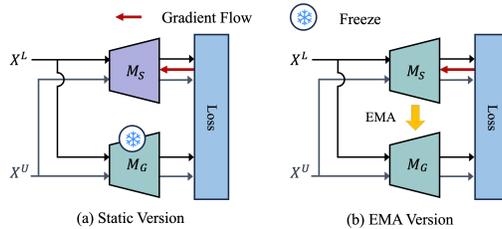
$$\mathcal{L}_U = \lambda_1 R_{\text{intra}} + \lambda_2 R_{\text{inter}} + \lambda_3 L_{\text{gan}}(\hat{Y}_S^U), \quad (11)$$

where  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  balance the optimization aims.  $L_{\text{gan}}$  here uses the same discriminator in Eq. (9). The supervised loss in Eq. (9) and the unsupervised distillation loss in Eq. (11) are combined as our final optimization objective, expressed as  $\mathcal{L} = \mathcal{L}_L + \mathcal{L}_U$ .

**Color Correction.** In practice, our approach often leads to color shifts due to the regularization of distances within the feature space. To address this, we rectify the output by normalizing the mean and variance of each color channel with those of the corresponding input channels [35]. Further details are provided in Supplementary Sec. E.

**Learning Configurations.** Our proposed PDD is a learning framework that can utilize pre-trained networks. It is recommended that  $M_G$  is pre-trained with extensive degradations  $\mathcal{D}_G$ , as in [35, 43]. The *static* configuration (refer to Fig. 3 (a)) initializes  $M_S$  with the model pre-trained on specific synthetic degradation ( $\mathcal{D}_S$ ), and  $M_G$  is frozen throughout the training process.

Alternative configuration initializes  $M_G$  and  $M_S$  with the same model pre-trained with  $\mathcal{D}_G$  and maintain  $M_G$  as an Exponential Moving Average (EMA)



**Fig. 3:** Two initialization options for the generalist and specialist models. (a) The static configuration initializes  $M_G$  with a model pre-trained by the complex synthetic pipeline and  $M_S$  with a model pre-trained by simple degradation in  $X^L$ . Weights of  $M_G$  are frozen during distillation. (b) Both  $M_S$  and  $M_G$  are initialized with a pre-trained generalized model. Weights of  $M_G$  are the EMA version of  $M_S$ .

version of  $M_S$  (see Fig. 3 (b)). Despite  $M_S$  not being initialized as a specialist, it can readily specialize to the synthetic degradation  $\mathcal{D}_S$  through supervised learning in Eq. (9). Empirically, the EMA version tends to be more effective, likely because the updated knowledge of real-world data in  $M_S$  also benefits  $M_G$  through EMA, enabling  $M_G$  to yield better references for optimization.

## 4 Experiments

### 4.1 Settings

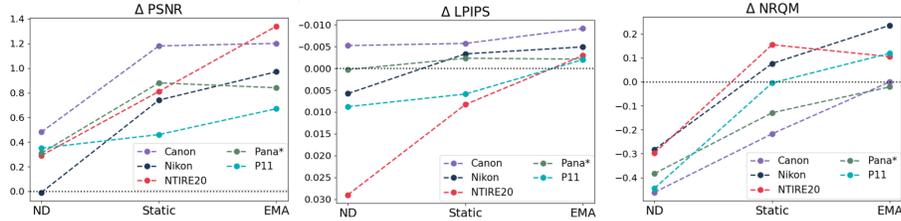
**Datasets.** We experiment on real-world datasets, RealSR [2] and DRealSR [39], and a synthetic dataset without released degradation settings, NTIRE20 [23]. For experiments on RealSR and DRealSR, we synthesize labeled training data  $\{X^L\}$  from high-resolution images in the DF2K dataset [1,32]. For the NTIRE20 dataset, we use the provided target clean images to synthesize low-resolution counterparts. For synthesizing  $\{X^L\}$ , we use only bicubic interpolation as the default and investigate other options for  $\mathcal{D}_S$  in the Supplementary Sec. D.b. RealSR or DRealSR has multiple data sources, and we optimize our methods for each source separately. For DRealSR, we report results for Panasonic and Olympus sources, as the others do not have sufficient training samples.

**Training Details.** We crop patches of size  $48 \times 48$  as inputs and train the model with the ADAM optimizer [12] using the settings  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ , and  $\epsilon = 10^{-8}$ . The batch size is 16, including equal numbers of labeled and unlabeled inputs. The initial learning rate is  $1e - 4$  and halved after 25K iterations. The static version needs 50K iterations, and the EMA version is trained for 100K iterations due to the initial specialist underfitting to the synthetic domain. All models are for  $\times 4$  super-resolution. Other details are in Supplementary Sec. A.

**Evaluation.** We evaluate models on testing sets of RealSR, DRealSR, and the validation set of NTIRE20. We crop non-overlapped  $512 \times 512$  patches from LR images in DRealSR to avoid memory exhaustion. For other datasets, we use the full images. To fully use the provided high-resolution counterparts, we report

**Table 1:**  $M_G$  and  $M_S(\text{static})$  columns show *initial* performances on the bicubic-interpolated data  $\{X^L\}$  from Set14 [42] and  $\{X^U\}$  with unknown degradation from NTIRE20 [23].  $M_S(\text{EMA})$  column shows the performance of  $M_S$  in the EMA version on  $\{X^L\}$  after training process.

Domain	Metric	$M_G$	$M_S(\text{static})$	$M_S(\text{EMA})$	Domain	Metric	$M_G$	$M_S(\text{static})$
$\{X^L\}$	PSNR $\uparrow$	23.98	24.81	26.26	$\{X^U\}$	PSNR $\uparrow$	25.08	19.82
	LPIPS $\downarrow$	0.2349	0.1337	0.1471		LPIPS $\downarrow$	0.2504	0.7552



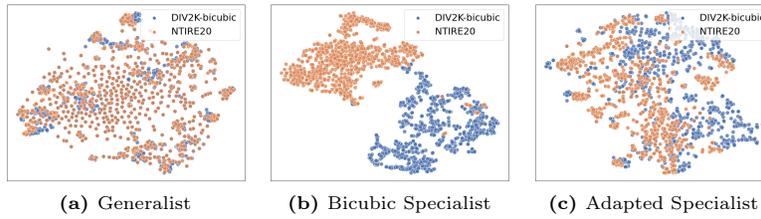
**Fig. 4:** Improvements over the Generalist on five unlabeled data domains, where a lower LPIPS score is better. ND improves fidelity scores (PSNR) but dramatically drops perceptual scores (LPIPS and NRQM). Both versions of our method achieve better improvements than ND for all reported metrics. For each domain, the Static version improves at least one of the perceptual metrics; the EMA version improves all.

scores of full-reference metrics, PSNR, SSIM, and LPIPS [44]. The PSNR and SSIM scores are calculated on the Y-channel of YCbCr format images. Additionally, we follow BSRGAN [43] and supplement with the no-reference metric NRQM [24] to evaluate the perceptual clearness, as the ground truth in real-world captured datasets [2, 39] also feature some blur. We use the implementation in IQA-PyTorch<sup>3</sup> for all metrics. It is important to consider both fidelity and perceptual scores; fidelity metrics are worse at detecting blurriness while perceptual metrics are not sensitive to high-frequency artifacts [25, 26, 29, 48].

## 4.2 Effectiveness of PDD

**Comparisons.** We compare the two versions of our method in Fig. 3 with the naive distillation defined in Eq. (3) regarding their improvements over the generalist network  $M_G$ . We choose RealESRGAN [35] as the pre-trained generalist model and ESRGAN [36], a standard SISR model, as the pre-trained specialist for the naive distillation and the static version in Fig. 3 (a). Tab. 1 shows the initial performances of the generalist and specialist for  $\{X^L\}$  and  $\{X^U\}$ . As we assumed in Sec. 3.1, the specialist performs better than the generalist for  $\{X^L\}$  while  $M_G$  generalizes better for  $\{X^U\}$ . For the EMA version in Fig. 3 (b),  $M_S$  as well as  $M_G$  is initialized with RealESRGAN. Through the supervised learning in Eq. (9),  $M_S(\text{EMA})$  shows specialized performance for labeled data  $\{X^L\}$ .

<sup>3</sup> <https://github.com/chaofengc/IQA-PyTorch>



**Fig. 5:** Visualization of low-level features for predictions of DIV2K [1] (bicubic) and NTIRE20 (unknown) following [20] (a) Generalist’s predictions for the two domains has overlapped distribution. (b) The predictions from the specialist that only adept in bicubic interpolation have separated distributions. (c) After applying our method (static version), predictions for two domains are pushed close.

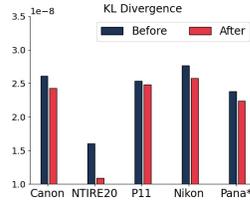
Fig. 4 plots the improvements over the pretrained generalist for Naive Distillation (ND), Static, and EMA versions of our method. Compared to the generalist, ND dramatically drops in perceptual scores (LPIPS, NRQM) despite maintaining or improving the PSNR. Both versions of our method perform better than ND; compared to the generalist, our static version improves PSNR significantly and at least one perceptual score for each data domain. The EMA version improves all metrics with the highest magnitudes.

**Change in Low-Level Characteristics.** We extract low-level features for visualization following [20]. Fig. 5b visualize predictions generated by the bicubic specialist (ESRGAN), which has separate distributions. Fig. 5c verifies the predictions from the specialist after applying our method (static version), where the separated distributions are pushed closer. Visualizing low-level differences between images with arbitrary content is challenging [20]. As such, we also quantitatively compare the KL divergence between the low-level features. As shown in Fig. 6, applying our method decreases the KL divergence between labeled (DIV2K-bicubic) and unlabeled predictions. Details of the visualization and KL-divergence calculations are in Supplementary Sec. B.

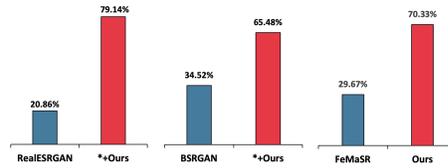
### 4.3 Comparisons with Other Methods

**Quantitative Comparison.** We consider two variants of our method by applying the EMA version on RealESRGAN [35] and BSRGAN [43]. In Tab. 2, we compare their performances with other state-of-the-art RWSR methods, including FeMaSR [3], EDAN [17], DASR [40], DAN [11], RealESRGAN [35], BSRGAN [43], and TG [46], which released code and model weights that enable reproduction. We regenerate results to evaluate these methods, while Supplementary Sec. C discusses related but not directly comparable methods.

As shown in Tab. 2, EDAN, DAN, and DASR perform strongly on the fidelity metrics (PSNR, SSIM) but are poor perceptually, especially on the NRQM score. The qualitative comparison highlights the blurry effects. Our method enhances RealESRGAN and BSRGAN across all metrics and demonstrates superior performance compared to TG when both are applied to RealESRGAN. For FeMaSR,



**Fig. 6:** KL divergence between specialist’s unlabeled and labeled predictions. The KL divergence between labeled and unlabeled predictions is lower after applying our method (red) than that in the specialist before adaption (blue).



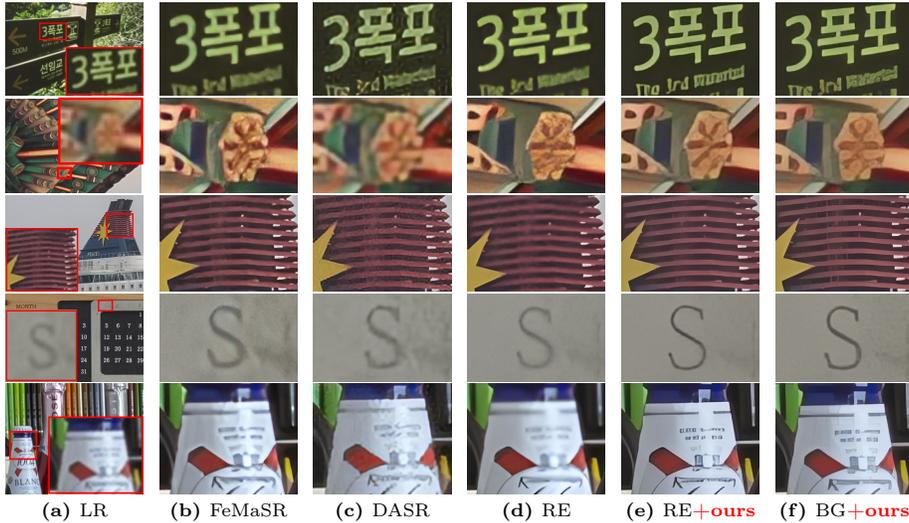
**Fig. 7:** User study on outputs of 30 real-world images. We recruit 30 evaluators for each image. Each vote is for a better quality within a pair of options. We compared our method to RealESRGAN and BSRGAN by using them as pre-trained models; FeMaSR is compared to RealESRGAN+ours. Our method gains at least 65% votes.

**Table 2:** Quantitative comparison with state-of-the-art methods. DAN, EDAN, and DASR are separately listed as they favor fidelity metrics but yield poor perceptual scores. For other methods, **bold** and underline mark the best and second-best scores. Subscripts in the last two columns show their difference from the corresponding pre-trained generalist, where improvement is colored red and drop is blue.

Dataset	Metric	DAN	EDAN	DASR	FeMaSR	RealESRGAN	BSRGAN	RealESRGAN	RealESRGAN	BSRGAN
		[11]	[17]	[40]	[3]	[43]	[35]	+TG [46]	+Ours	+Ours
NTIRE20 -Valid	PSNR $\uparrow$	26.83	26.59	-	23.54	25.08	25.44	25.56	<b>26.42</b> <sub>(+1.340)</sub>	<u>26.37</u> <sub>(+0.930)</sub>
	SSIM $\uparrow$	0.7171	0.7400	-	0.6665	0.7061	0.6984	0.7193	<b>0.7297</b> <sub>(+0.024)</sub>	<u>0.7238</u> <sub>(+0.025)</sub>
	LPIPS $\downarrow$	0.5747	0.2678	-	<b>0.2360</b>	0.2504	0.2645	<u>0.2420</u>	0.2475 <sub>(-0.003)</sub>	0.2595 <sub>(-0.005)</sub>
	NRQM $\uparrow$	4.8784	5.8481	-	<b>6.6245</b>	6.1213	6.2779	5.7926	6.2263 <sub>(+0.105)</sub>	<u>6.2781</u> <sub>(+0.0002)</sub>
RealSR -Canon	PSNR $\uparrow$	26.67	26.36	26.71	24.29	24.74	25.57	25.03	<u>25.94</u> <sub>(+1.200)</sub>	<b>26.13</b> <sub>(+0.560)</sub>
	SSIM $\uparrow$	0.7740	0.7774	0.7782	0.7460	0.7634	0.7683	<u>0.7720</u>	<b>0.7733</b> <sub>(+0.010)</sub>	0.7626 <sub>(-0.006)</sub>
	LPIPS $\downarrow$	0.4095	0.3026	0.2507	0.2809	0.2607	0.2573	0.2615	<b>0.2516</b> <sub>(-0.009)</sub>	<u>0.2517</u> <sub>(-0.006)</sub>
	NRQM $\uparrow$	3.0407	3.6360	3.9383	6.0151	<u>6.0649</u>	6.0293	5.4250	<u>6.0647</u> <sub>(-0.0002)</sub>	<b>6.1323</b> <sub>(+0.103)</sub>
RealSR -Nikon	PSNR $\uparrow$	26.29	25.37	25.63	23.94	24.31	24.79	24.56	<b>25.28</b> <sub>(+0.970)</sub>	<u>25.07</u> <sub>(+0.280)</sub>
	SSIM $\uparrow$	0.7533	0.7460	0.7473	0.7097	0.7406	0.7334	<u>0.7406</u>	<b>0.7470</b> <sub>(+0.014)</sub>	0.7294 <sub>(-0.001)</sub>
	LPIPS $\downarrow$	0.4133	0.3200	0.2785	0.3046	0.2851	<u>0.2797</u>	0.2949	0.2802 <sub>(-0.005)</sub>	<b>0.2713</b> <sub>(-0.008)</sub>
	NRQM $\uparrow$	3.2056	4.4872	4.4952	<u>5.9488</u>	5.6685	5.9387	5.1942	5.9044 <sub>(+0.236)</sub>	<b>6.1276</b> <sub>(+0.189)</sub>
DRealSR -Panasonic	PSNR $\uparrow$	28.99	28.12	-	25.02	27.03	26.93	27.51	<b>27.87</b> <sub>(+0.840)</sub>	<u>27.69</u> <sub>(+0.760)</sub>
	SSIM $\uparrow$	0.8123	0.8006	-	0.7118	0.7782	0.7621	<u>0.7843</u>	<b>0.7939</b> <sub>(+0.016)</sub>	0.7752 <sub>(+0.013)</sub>
	LPIPS $\downarrow$	0.4235	0.2747	-	0.3088	0.2703	0.2820	0.2836	<b>0.2682</b> <sub>(-0.002)</sub>	<u>0.2687</u> <sub>(-0.013)</sub>
	NRQM $\uparrow$	2.8131	4.5915	-	<b>6.0342</b>	5.3751	<u>5.5865</u>	4.6483	5.3540 <sub>(-0.021)</sub>	<u>5.5608</u> <sub>(-0.026)</sub>
DRealSR -Olympus	PSNR $\uparrow$	28.74	27.85	-	24.38	26.73	26.49	<b>27.43</b>	<u>27.40</u> <sub>(+0.670)</sub>	27.25 <sub>(+0.764)</sub>
	SSIM $\uparrow$	0.8063	0.7979	-	0.6691	0.7667	0.7568	<u>0.7843</u>	<b>0.7846</b> <sub>(+0.018)</sub>	0.7635 <sub>(+0.007)</sub>
	LPIPS $\downarrow$	0.4595	0.3555	-	0.4007	<u>0.3159</u>	0.3283	0.3416	<b>0.3139</b> <sub>(-0.002)</sub>	0.3268 <sub>(-0.001)</sub>
	NRQM $\uparrow$	2.7961	3.7989	-	<b>6.2590</b>	5.2530	5.5376	4.1459	5.3722 <sub>(+0.120)</sub>	<u>5.5957</u> <sub>(+0.058)</sub>

we achieve competitive perceptual scores and significantly surpass their fidelity scores.

**User Study** is conducted by presenting two reconstructions of the same LR image and asking for the better one. We compare the EMA version of our method with methods RealESRGAN, BSRGAN, and FeMaSR. Each comparison is conducted on 30 different real-world results among 30 evaluators. As shown in Fig. 7, at least 65% evaluators vote for our method in each comparison. Although quan-



**Fig. 8:** Qualitative comparisons. LRs are from the RealSR-Canon dataset. Columns (e) and (f) are the results of applying our methods on RealESRGAN (RE) and BSRGAN (BG). FeMaSR and DASR can not predict reasonable details.

titatively competing, FeMaSR yields unpleasant artifacts to human eyes, further shown in qualitative comparison.

**Qualitative Comparison** is in Fig. 8. FeMaSR [3] produces high-frequency artifacts, while outputs from DASR [40] suffer blur or artifacts. Compared to RealESRGAN, our method predicts sharper patterns without introducing distortion. More comparisons are in Supplementary Sec. F.

#### 4.4 Ablations

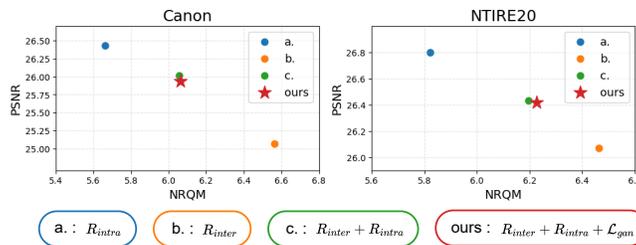
**Ablation on loss terms** reveals the effectiveness of each term in Eq. (11). All experiments take RealESRGAN as the pre-trained model and follow the EMA implementation of our method in Fig. 3 (b). Fig. 9 shows that  $R_{intra}$  and  $R_{inter}$  relate differently to fidelity and perception. Their combination achieves balanced improvements in both directions, and the  $L_{gan}$  increases pattern sharpness with a slight trade-off in PSNR. Controlling the ratio between weights of  $R_{intra}$  and  $R_{inter}$  results in fidelity and perceptual quality trade-offs. We choose their balance in Sec. 4.3 and refer to Supplementary Sec. D.a for more discussion.

**Effectiveness of EMA** is further validated by comparing to a *SingleFixed* version, which also uses RealESRGAN as the pre-trained model for both  $M_S$  and  $M_G$  but fixes the  $M_G$  during training. As shown in Tab. 3, updating the weights of the generalist network through EMA enables higher performance in both fidelity and perception.

**Choice of  $R_{intra}$  measurement.** Eq. (6) use Cross-Entropy (CE) for measuring the difference between  $d_S^{ij}$  and  $d_G^{ij}$ . Here, we alternate the CE to  $\ell_1$  distance

**Table 3:** Collection of ablations on EMA and choice of measurement in Eq. (6). **Bold** and underline mark the best and second best scores. SingleFixed uses the same pre-trained model to initialize  $M_S$  and  $M_G$  as the EMA version but fixes the weights of  $M_G$  during training, resulting in inferior performance. Substituting CE in Eq. (6) to  $\ell_1$  biases the fidelity scores.

Method	Canon		NTIRE20		Olympus	
	PSNR	NRQM	PSNR	NRQM	PSNR	NRQM
Generalist	24.74	<b>6.0649</b>	25.08	6.1213	26.73	5.2530
SingleFixed	25.71	5.9803	25.75	6.0938	<u>27.47</u>	5.2071
$\ell_1$	<b>26.17</b>	5.8061	<b>26.44</b>	<u>6.1997</u>	<b>27.72</b>	4.9602
Ours	<u>25.94</u>	<u>6.0647</u>	<u>26.42</u>	<b>6.2263</b>	27.40	<b>5.3722</b>



**Fig. 9:** Ablations of loss terms for distillation in Eq. (11). Using one of  $R_{intra}$  (blue) and  $R_{inter}$  (orange) bias fidelity or perceptual quality (a.-b.), while the combination of them (c.-d.) achieves a balanced improvement (see green and red dots).

without changing the ratio between weights of  $R_{intra}$  and  $R_{inter}$ . As shown in Tab. 3, CE favors the overall sharpness while  $\ell_1$  focuses more on fidelity. Although  $\ell_1$  is a feasible choice, we use CE in our method due to the importance of sharpness for human eyes [13].

## 5 Conclusion

This paper studies unsupervised real-world image super-resolution from the distillation perspective. The approach considers both the real-world knowledge of a generalized SR model and the low-level characteristics of synthetic predictions of the specialist model, adapting a synthetic domain specialist to target real-world degradation. The core of our method is the Pairwise Distance Distillation. By enforcing the consistencies for intra- and inter-model distances, our method aims at pushing the low-level characteristics of the specialist’s real-world predictions towards its synthetic predictions. As a learning scheme, our method can improve off-the-shelf models regarding fidelity and perception on multiple real-world datasets. Additionally, we emphasize our approach to offering a new perspective toward addressing real-world super-resolution challenges.

**Acknowledgement** This research/project is supported by the Ministry of Education, Singapore, under the Academic Research Fund Tier 1 (FY2022).

## References

1. 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)
2. Cai, J., Zeng, H., Yong, H., Cao, Z., Zhang, L.: Toward real-world single image super-resolution: A new benchmark and a new model. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3086–3095 (2019)
3. 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)
4. Deng, X., Yang, R., Xu, M., Dragotti, P.L.: Wavelet domain style transfer for an effective perception-distortion tradeoff in single image super-resolution. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 3076–3085 (2019)
5. Fritsche, M., Gu, S., Timofte, R.: Frequency separation for real-world super-resolution. In: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). pp. 3599–3608. IEEE (2019)
6. Gao, H., Guo, J., Wang, G., Zhang, Q.: Cross-domain correlation distillation for unsupervised domain adaptation in nighttime semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9913–9923 (2022)
7. Gao, Q., Zhao, Y., Li, G., Tong, T.: Image super-resolution using knowledge distillation. In: Asian Conference on Computer Vision. pp. 527–541. Springer (2018)
8. Gatys, L.A., Ecker, A.S., Bethge, M.: Image style transfer using convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2414–2423 (2016)
9. 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)
10. Hinton, G., Vinyals, O., Dean, J.: Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531 (2015)
11. 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)
12. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
13. 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)
14. Lepik, Ü., Hein, H.: Haar wavelets. In: *Haar Wavelets: With Applications*, pp. 7–20. Springer (2014)
15. Li, W., Li, L., Yang, H.: Progressive cross-domain knowledge distillation for efficient unsupervised domain adaptive object detection. *Engineering Applications of Artificial Intelligence* **119**, 105774 (2023)

16. Li, Y., Wang, N., Liu, J., Hou, X.: Demystifying neural style transfer. arXiv preprint arXiv:1701.01036 (2017)
17. Liang, J., Zeng, H., Zhang, L.: Efficient and degradation-adaptive network for real-world image super-resolution. In: European Conference on Computer Vision. pp. 574–591. Springer (2022)
18. 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)
19. 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)
20. Liu, Y., Liu, A., Gu, J., Zhang, Z., Wu, W., Qiao, Y., Dong, C.: Discovering distinctive " semantics" in super-resolution networks. arXiv preprint arXiv:2108.00406 (2021)
21. Liu, Y., Zhao, H., Gu, J., Qiao, Y., Dong, C.: Evaluating the generalization ability of super-resolution networks. *IEEE Transactions on pattern analysis and machine intelligence* (2023)
22. Liu, Y., Zhang, W., Wang, J.: Adaptive multi-teacher multi-level knowledge distillation. *Neurocomputing* **415**, 106–113 (2020)
23. Lugmayr, A., Danelljan, M., Timofte, R.: Ntire 2020 challenge on real-world image super-resolution: Methods and results. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 494–495 (2020)
24. 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)
25. Ma, C., Rao, Y., Cheng, Y., Chen, C., Lu, J., Zhou, J.: Structure-preserving super resolution with gradient guidance. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 7769–7778 (2020)
26. Maggioni, M., Tanay, T., Babiloni, F., McDonagh, S., Leonardis, A.: Tunable convolutions with parametric multi-loss optimization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20226–20236 (2023)
27. Nguyen-Meidine, L.T., Belal, A., Kiran, M., Dolz, J., Blais-Morin, L.A., Granger, E.: Unsupervised multi-target domain adaptation through knowledge distillation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 1339–1347 (2021)
28. Park, S., Kwak, N.: Local-selective feature distillation for single image super-resolution. arXiv preprint arXiv:2111.10988 (2021)
29. Park, S.H., Moon, Y.S., Cho, N.I.: Perception-oriented single image super-resolution using optimal objective estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1725–1735 (2023)
30. Romero, A., Van Gool, L., Timofte, R.: Unpaired real-world super-resolution with pseudo controllable restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 798–807 (2022)
31. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
32. Timofte, R., Agustsson, E., Van Gool, L., Yang, M.H., Zhang, L.: Ntire 2017 challenge on single image super-resolution: Methods and results. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 114–125 (2017)

33. Wang, H., Chen, X., Ni, B., Liu, Y., Liu, J.: Omni aggregation networks for lightweight image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 22378–22387 (2023)
34. Wang, W., Zhang, H., Yuan, Z., Wang, C.: Unsupervised real-world super-resolution: A domain adaptation perspective. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4318–4327 (2021)
35. 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)
36. 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)
37. Wang, Y., Cao, Y., Zha, Z.J., Zhang, J., Xiong, Z.: Deep degradation prior for low-quality image classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11049–11058 (2020)
38. Wang, Y., Wang, L., Shi, S., Li, V.O., Tu, Z.: Go from the general to the particular: Multi-domain translation with domain transformation networks. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 9233–9241 (2020)
39. Wei, P., Xie, Z., Lu, H., Zhan, Z., Ye, Q., Zuo, W., Lin, L.: Component divide-and-conquer for real-world image super-resolution. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VIII 16. pp. 101–117. Springer (2020)
40. Wei, Y., Gu, S., Li, Y., Timofte, R., Jin, L., Song, H.: Unsupervised real-world image super resolution via domain-distance aware training. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13385–13394 (2021)
41. Yang, C., Pan, J., Gao, X., Jiang, T., Liu, D., Chen, G.: Cross-task knowledge distillation in multi-task recommendation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36, pp. 4318–4326 (2022)
42. Zeyde, R., Elad, M., Protter, M.: On single image scale-up using sparse-representations. In: Curves and Surfaces: 7th International Conference, Avignon, France, June 24–30, 2010, Revised Selected Papers 7. pp. 711–730. Springer (2012)
43. 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)
44. 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)
45. Zhang, R., Gu, J., Chen, H., Dong, C., Zhang, Y., Yang, W.: Crafting training degradation distribution for the accuracy-generalization trade-off. arXiv preprint arXiv:2305.18107 (2023)
46. Zhang, W., Li, X., Shi, G., Chen, X., Qiao, Y., Zhang, X., Wu, X.M., Dong, C.: Real-world image super-resolution as multi-task learning. *Advances in Neural Information Processing Systems* **36** (2024)
47. Zhang, Y., Chen, H., Chen, X., Deng, Y., Xu, C., Wang, Y.: Data-free knowledge distillation for image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7852–7861 (2021)
48. Zhang, Y., Ji, B., Hao, J., Yao, A.: Perception-distortion balanced admm optimization for single-image super-resolution. In: European Conference on Computer Vision. pp. 108–125. Springer (2022)

49. Zhong, Z., Shen, T., Yang, Y., Lin, Z., Zhang, C.: Joint sub-bands learning with clique structures for wavelet domain super-resolution. *Advances in neural information processing systems* **31** (2018)
50. 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)