ICF-SRSR: Invertible scale-Conditional Function *for* Self-Supervised Real-world Single Image Super-Resolution

Reyhaneh Neshatavar[1*] Mohsen Yavartanoo[1*] Sanghyun Son[1] Kyoung Mu Lee[1,2]
[1]Dept. of ECE & ASRI, [2]IPAI, Seoul National University, Seoul, Korea
{reyhanehneshat,myavartanoo,thstkdgus35,kyoungmu}@snu.ac.kr

Abstract

Single image super-resolution (SISR) is a challenging ill-posed problem that aims to up-sample a given low-resolution (LR) image to a high-resolution (HR) counterpart. Due to the difficulty in obtaining real LR-HR training pairs, recent approaches are trained on simulated LR images degraded by simplified down-sampling operators, e.g.*, bicubic. Such an approach can be problematic in practice due to the large gap between the synthesized and real-world LR images. To alleviate the issue, we propose a novel Invertible scale-Conditional Function (ICF), which can scale an input image and then restore the original input with different scale conditions. Using the proposed ICF, we construct a novel self-supervised SISR framework (ICF-SRSR) to handle the real-world SR task without using any paired/unpaired training data. Furthermore, our ICF-SRSR can generate realistic and feasible LR-HR pairs, which can make existing supervised SISR networks more robust. Extensive experiments demonstrate the effectiveness of our method in handling SISR in a fully self-supervised manner. Our ICF-SRSR demonstrates superior performance compared to the existing methods trained on synthetic paired images in real-world scenarios and exhibits comparable performance compared to state-of-the-art supervised/unsupervised methods on public benchmark datasets. The code is available from this link.*

1. Introduction

Single image super-resolution (SISR) as a fundamental vision problem is a procedure to reconstruct a super-resolution (SR) image from a single low-resolution (LR) image. SISR is an active research topic and has attracted increasing attention in low-level computer vision. It has many applications in various fields such as medical imaging [17, 43], face recognition [19, 60], satellite image processing [32, 51] and security video surveillance [35, 67].

*equal contribution

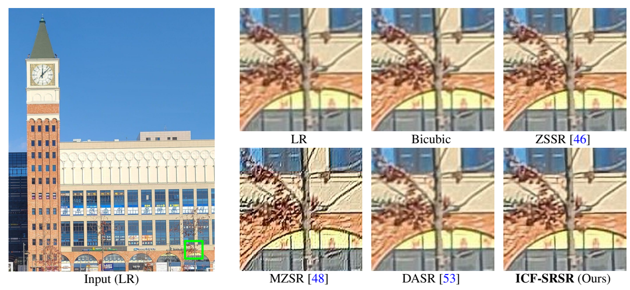


Figure 1. **Real-world image super-resolution**. We train our ICF-SRSR on a single real-world smartphone photo in a self-supervised manner to get the result for scale $\times 2$. The other listed methods are also zero-shot [46, 48] or unsupervised [53] methods.

Recent state-of-the-art (SOTA) SR methods have achieved remarkable progress due to the development of deep convolutional neural networks (CNNs). They are usually trained on synthetic inputs in a fully supervised fashion, where LR images are generated by bicubic down-sampling from their HR counterparts. Nevertheless, models trained on the synthetic datasets cannot generalize well when applied to real-world inputs [6, 7]. Another problem is that acquiring well-constructed LR-HR pairs from the real world is very challenging due to cost problems or hardware limitations [6, 7, 68]. Therefore, it is a common scenario that we have LR images only rather than having LR-HR training pairs. Several approaches adopt unsupervised adversarial training [16] and leverage unpaired LR-HR images to alleviate the situation. By jointly training down-sampling and up-sampling networks [5, 36, 37, 62, 72], those methods aim to generate synthetic LR images that have similar characteristics of given unpaired LR examples. Then, the synthesized training pairs can be leveraged to optimize the up-sampling network. However, such unsupervised strategies require appropriate HR images, even though those images are not paired with the given LR images. Also, Son *et al.* [49] have identified that those methods are biased toward some hand-crafted functions *e.g.*, nearest or bicubic interpolation, which limits the generalization.

In this paper, we present a novel self-supervised real-world SR framework, ICF-SRSR, to overcome the aforementioned challenges. To this end, we first propose a concept of Invertible scale-Conditional Function (ICF). It is designed to perform up-sampling and down-sampling within a single model, conditioned by the scale arguments s and $1/s$, respectively. Therefore, we can resize an input by a given scale s and restore the initial input by taking the inverse scale $1/s$. Without utilizing paired/unpaired training images nor any specific down-sampling operator *e.g.*, bicubic, ICF-SRSR containing a learnable ICF can be trained in a fully self-supervised manner. Moreover, our method can generate realistic LR-HR image pairs from a set of given images useful for training the other off-the-shelf methods. In the experiments, we demonstrate the ability of our ICF-SRSR to learn from real-world datasets, restore high-/lower-resolution images, and evaluate our method on other datasets in a self-supervised manner. Our main contributions are threefold:

- Our ICF-SRSR is a self-supervised framework for the SISR task that performs simultaneous SR and down-sampling based on the proposed ICF.
- Our ICF-SRSR can learn a feasible resizing function directly from real-world LR images. Our self-supervised approach performs better on real-world SR than existing methods trained on synthetic datasets, even with training on a single image, as evident in Fig. 1.
- Our ICF-SRSR can also down-sample given natural images, which enables us to construct realistic training pairs. Therefore, we can train off-the-shelf SR methods using the generated pairs by our ICF-SRSR in the absence of real paired training samples.

2. Related Works

In this section, we review recent SR methods from the perspective of training supervision.

2.1. Supervised image super-resolution

Starting from Dong *et al.* [12], CNNs [13, 45] have become a standard for SISR. Following VDSR [28], several methods such as LapSRN [30], EDSR [34], and SR-GAN [31] have taken advantage of residual learning. Advanced approaches utilize dense connections [56, 71], channel attention [11, 42, 70], and back-projection [21, 22], and even transformers [8, 14, 33, 40, 58, 63] for high-performance SR architectures. Furthermore, recent attempts extend the task to continuous scaling factors [9, 23, 47, 54] and even to arbitrary shapes [50].

Nevertheless, supervised methods are still vulnerable when a given LR image is degraded by an unknown down-sampling function [49] that is not seen during training. Therefore, several methods [10, 18, 25] jointly estimate latent kernel parameters and SR images to alleviate the issue. Rather than up-sampling LR images directly, Correction filter [26] first converts a given input to resemble a bicubic down-sampled image and applies off-the-shelf SR methods. Still, they require supervision from synthetic LR-HR pairs for training, which prevents their real-world applications.

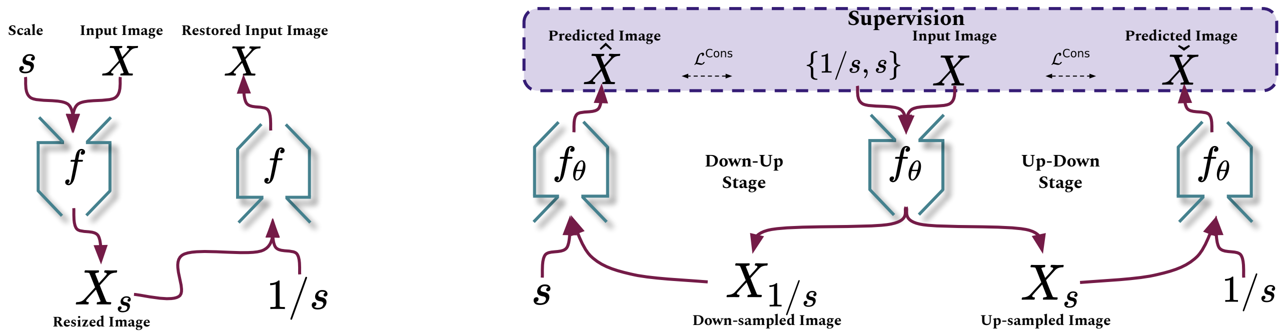
2.2. Unsupervised super-resolution

To reduce biases from synthetic training data, zero-shot methods are trained on a given LR input only, without relying on supervision from large-scale data. Ulyanov *et al.* [52] has shown that the structure of CNNs can be prior for natural image representation which can be utilized for the SR task. Based on internal patch recurrence [41], ZSSR [46] is trained on numerous sub-patches of the given image to construct an input-specific SR model. Later, there has been an attempt to integrate external and internal learning using model-agnostic meta-learning [15]. MZSR [48] is firstly trained on a large-scale paired dataset with multiple degradation parameters and then adopted to a given image during the inference time.

However, the zero-shot methods assume that the degradation pipeline for a given image is known, which is less practical. To implement fully-blind SR methods, internal patch recurrence properties have played a critical role [41]. Based on such a background, KernelGAN [3] predicts a kernel that matches the distribution of the down-sampled image and the original input in an unsupervised manner. The estimated kernel can also be utilized for several SR models [46, 66] for more accurate reconstruction. Rather than explicitly utilize the concept of image distribution, we construct self-supervised chains to learn the SR model without assuming a specific degradation model.

2.3. Cyclic architectures for super-resolution

On the other hand, a class of methods interprets SR as a domain transfer problem between LR and HR distributions. They introduce cyclic architectures [27] with adversarial loss [16, 44, 73] to train consecutive down-sampling and SR networks. CinCGAN [62] utilizes the concept of cycle consistency to train the model on unpaired LR-HR images. Under the cyclic framework [5, 36, 37, 72], down-sampling models are trained to simulate the distribution of training LR images. Then, the following SR network can learn to generalize on given LR images even if the corresponding HR pairs do not exist. However, they are still biased toward handcrafted down-sampling functions [49] and lack generalization. Without using adversarial loss, Guo *et al.* [20] combine paired and unpaired data to train a dual regression network with a loop. In this paper, we further propose a self-supervised approach without requiring either paired/unpaired training data or a specific down-sampling operator.



(a) ICF (b) ICF-SRSR

Figure 2. **Overview of our proposed method.** (a) We introduce an invertible scale-conditional function (ICF), which receives an input image and an arbitrary scale condition and generates a resized image. It outputs the same input image for the resized image and the inverse scale condition. (b) We propose a self-supervised SISR framework ICF-SRSR, in which a learnable ICF up-samples and down-samples a given image with different scale conditions and can reproduce the same input from the generated images by the inverse scales using the defined loss functions between the predicted images and the original input.

2.4. Real-world super-resolution

To overcome the limitations of existing methods when handling real-world data, several approaches have captured paired LR-HR images in the wild. While they are still limited due to scene diversity [7], accurate alignment [6, 59], real-world datasets help generalization of existing SR models with more practical training data. Zhang *et al*. [68] and Xu *et al*. [61] leverage RAW and RGB images together to deliver better reconstruction quality. Nevertheless, those pairs require careful alignment and complicated hardware setup, which are not scalable. Recently, Real-ESRGAN [55] and BSRGAN [65] aim to synthesize more realistic and diverse LR images to improve the generalization ability of existing SR models. Still, they cannot leverage information from real-world images and heavily depend on such a synthesis process. On the other hand, our fully self-supervised framework does not require synthetic or real-world pairs and can be trained on arbitrary LR images.

3. Method

We first introduce an Invertible scale-Conditional Function (ICF) to design our self-supervised real-world single image super-resolution framework (ICF-SRSR); then, we discuss our defined loss functions and the network architecture. For convenience, we denote $X \in \mathbb{R}^{H\times W\times 3}$ as the input LR image with arbitrary sizes of H and W.

3.1. Invertible scale-Conditional Function

For a given input X, a conditional function $f(X|s)$ returns different outputs for different conditions s. In this paper, we design an Invertible scale-Conditional Function (ICF) as a specific conditional function, which can act as an operation and the inverse operation for different scale conditions. Without losing generality, we consider f as an image-to-image mapping and s as an arbitrary scaling factor, respectively. Then, we can resize an arbitrary image X as follows:

$$X_s = f\left(X|s\right), \tag{1}$$

where $X_s \in \mathbb{R}^{sH\times sW\times 3}$ is a resized image. Furthermore, for the same function f, we can get the original input X again using the inverse scaling factor $1/s$ as follows:

$$X = f\left(X_s|1/s\right). \tag{2}$$

Therefore, f as an ICF can project an image to its arbitrary scale representation and back-project it to the original input for the scale conditions s and $1/s$, respectively. Fig. 2a illustrates the concept of our ICF. We note that if $s = 1/s = 1$ the function is identity which implies $f(X|1) = X$.

3.2. Self-supervised SISR using ICF

One of the challenges in real-world SR is that we cannot acquire the ground-truth HR image for an arbitrary LR image. To overcome this limitation, we develop a novel self-supervised SR framework, ICF-SRSR, based on the concept of ICF. As shown in Fig. 2b, our method can simultaneously super-resolve and down-sample the given LR image X with different scale conditions s and $1/s$, without requiring any paired/unpaired LR-HR training samples. Specifically, we first parameterize an ICF f_θ with CNNs and utilize its property to optimize the model. Then, we repeatedly apply f_θ to an LR image X with different scale conditions to acquire two outputs $\check{X}, \hat{X} \in \mathbb{R}^{H\times W\times 3}$ as follows:

$$\begin{aligned} f_\theta(f_\theta(X|s)|1/s) &= f_\theta(X_s|1/s) = \check{X}, \\ f_\theta(f_\theta(X|1/s)|s) &= f_\theta(X_{1/s}|s) = \hat{X}, \end{aligned} \tag{3}$$

where for $s > 1$, $X_s \in \mathbb{R}^{sH\times sW\times 3}$ and $X_{1/s} \in \mathbb{R}^{H/s\times W/s\times 3}$ are generated super-resolution (SR) and low-low-resolution (LLR) images, respectively. For simplicity, we assume that both H/s and W/s are integers.

For an ideal ICF f_θ, both $\check{X}$ and $\hat{X}$ in Eq. (3) should be the same as the original LR image X. Therefore, we train f_θ in a self-supervised manner by reducing the distance between X and the generated images $\check{X}$ and $\hat{X}$ in two stages simultaneously, as shown in Fig. 2b. In the up-down stage, we minimize the distance between $\check{X}$ and X. By doing so, the network can learn to down-sample the generated SR image X_s by restoring the output $\check{X}$ as the approximation of the original input X. On the other hand, in the down-up stage, we aim to approximate the original input X by reducing the distance between $\hat{X}$ and X. Then, the network can learn to up-sample the generated LLR image $X_{1/_s}$. Therefore, by leveraging the learned up-sampler and down-sampler applied on the generated images $X_{1/_s}$ and X_s, respectively, we can generate favorable SR and LLR images X_s and $X_{1/_s}$ by employing the learned model f_θ on the input X with the scale conditions s and $1/s$, respectively.

We note that our method is different from CycleGAN [73], which utilizes unpaired LR-HR images and performs two independent cycles on LR and HR images separately. Rather, our model is trained in a self-supervised manner by optimizing the f_θ jointly with two stages on LR images only, without requiring the adversarial loss. In other words, f_θ can perform simultaneous up-sampling and down-sampling without requiring prior information or paired/unpaired data.

3.3. Training loss functions

To train the proposed ICF f_θ, we design a set of self-supervised loss functions. First, we formulate the consistency loss $\mathcal{L}^{\text{Cons}}$, which preserves information during the simultaneous up-down and down-up stages. The proposed consistency loss $\mathcal{L}^{\text{Cons}}$ on the approximated LR images $\hat{X}$ and $\check{X}$, and the original input X is defined as follows:

$$\mathcal{L}^{\text{Cons}} = \|\hat{X} - X\| + \|\check{X} - X\|. \tag{4}$$

For simplicity, we use $\|\cdot\|$ to represent the L1 norm. The proposed consistency term $\mathcal{L}^{\text{Cons}}$ guarantees to generate reliable up-sampled and down-sampled images simultaneously. Furthermore, to stabilize the training and preserve colors between the input and intermediate images X_s and $X_{1/_s}$, we utilize the low-frequency loss [49]. We implement the low-pass filter with a spatial pooling operator $\mathbf{P}(\cdot, w, s)$, where w and s are window size and stride, respectively. Our color-preserving loss $\mathcal{L}^{\text{Color}}$ is defined as follows:

$$\begin{aligned}\mathcal{L}^{\text{Color}} = &\|\mathbf{P}(X_s, 4s, 4s) - \mathbf{P}(X, 4, 4)\| \\ &+ \|\mathbf{P}\left(X_{1/_s}, 4, 4\right) - \mathbf{P}(X, 4s, 4s)\|,\end{aligned} \tag{5}$$

where the window size and stride are adjusted to match dimensions between each of (X_s, X) and $\left(X_{1/_s}, X\right)$. The total training objective $\mathcal{L}^{\text{Total}}$ is the combination of the aforementioned two loss terms, which is defined as follows:

$$\mathcal{L}^{\text{Total}} = \mathcal{L}^{\text{Cons}} + \lambda_{\text{Color}}\mathcal{L}^{\text{Color}}. \tag{6}$$

3.4. Network architecture

Our ICF-SRSR architecture leverages a single model to handle different scale conditions. To implement the proposed method, we modify the existing SISR model, *e.g.*, EDSR [34] as our baseline backbone architecture. Since the body part is invariant to the scale image (*i.e.*, the input and output have the same resolution), we introduce multiple tail parts for different scale conditions. Employing a single network with the shared body part is more efficient and can improve performance by observing more augmented data, *i.e.*, images with different scales, during the training. In the supplementary material, we provide the details of the network architecture and illustrate that our method is model-agnostic and can leverage different SOTA baselines.

4. Experiments

We first introduce training and evaluation configurations of the proposed ICF-SRSR framework. Then we conduct comprehensive experiments, extensive quantitative and qualitative comparisons with the other methods, and an in-depth analysis of our proposed method.

4.1. Training details

Dataset. We train and evaluate our method on two scenarios. 1) Synthetic datasets, where the training and testing LR images are synthesized by a uniform degradation process (*e.g.*, bicubic down-sampling) from HR images. 2) Real-world datasets, which provide paired LR-HR images from the real-world captured by adjusting the focal length of a camera.

To train our ICF-SRSR, we use 800 bicubic LR images from the DIV2K [1] dataset. For evaluation, we adopt five standard benchmarks: Set5 [4], Set14 [64], BSD100 [38], Urban100 [24], and Manga109 [39]. We also use the high-quality DIV2K validation set for evaluation.

To train and evaluate our ICF-SRSR under real-world scenarios, we utilize real-world datasets [6, 59] for the SISR task. RealSR-V3 [6] includes paired LR-HR images captured by two different cameras, Canon and Nikon. For each camera, about 200 training images are captured from different scenes for each scaling factor $\times 2$, $\times 3$, and $\times 4$. We used only LR images with scaling factors $\times 2$ and $\times 4$ for training and evaluated our method on the pairs of tests 50 for each scale. DRealSR [59] also contains images captured by five DSLR cameras. We conduct our experiments using images for $\times 2$ and $\times 4$ SR, containing 884 and 840 LR images, respectively. For evaluation, we use 83 and 93 test pairs in DRealSR for $\times 2$ and $\times 4$, respectively.

Hyperparameters. During the training, we extract random patches of size 48×48 from LR images of both synthetic and real-world datasets. For all our experiments, we set the batch size to 16, and $\lambda_{\text{Color}} = 0.2$. Random flip and

Supervision	**Method**	**Set5** ×2/×4	**Set14** ×2/×4	**BSD100** ×2/×4	**Urban100** ×2/×4	**Manga109** ×2/×4	**DIV2K** ×2/×4
	Bicubic	33.66/28.42	30.24/26.00	29.56/25.96	26.88/23.14	30.80/24.89	31.01/26.66
Supervised	VDSR [28]	37.53/31.35	33.03/28.01	31.90/27.29	30.76/25.18	37.22/28.83	33.66/28.17
	EDSR [34]	38.11/32.46	33.92/28.80	32.32/27.71	32.93/26.64	39.10/31.02	**36.22**/30.52
	CARN [2]	37.76/32.13	33.52/28.60	32.09/27.58	31.92/26.07	38.36/30.47	- /30.10
	RCAN [70]	38.27/32.63	34.12/28.87	32.41/27.77	33.34/26.82	39.44/31.19	36.13/30.52
	RDN [71]	38.24/32.47	34.01/28.81	32.34/27.72	32.89/26.61	39.18/31.00	- / -
	DRN-S [20]	37.80/32.68	33.30/28.93	31.97/27.78	31.40/26.84	38.11/31.52	35.77/**30.79**
	LIIF [9]	38.17/32.50	33.97/28.80	32.32/27.74	32.87/26.68	- / -	34.99/29.27
	ELAN [69]	**38.36/32.75**	**34.20/28.96**	**32.45/27.83**	**33.44/27.13**	**39.62/31.68**	- / -
Unsupervised	SelfExSR [24]	36.49/30.31	32.22/27.40	31.18/26.84	29.54/24.82	35.78/27.82	- / -
	ZSSR [46]	37.37/31.13	33.00/28.01	31.65/27.12	29.34/24.12	35.57/27.04	**34.45/29.08**
	MZSR [48]	37.25/31.59	33.16/27.90	31.64/ -	30.41/25.52	**36.70/29.58**	- / -
	DASR [53]	**37.87/31.99**	**33.34/28.50**	**32.03/27.52**	**31.49/25.82**	- / -	- / -
Self-supervised	**ICF-SRSR** (Ours)	37.01/30.81	32.86/27.76	31.54/26.99	30.39/24.72	36.45/28.01	35.19/29.48
	EDSR (LLR,LR) (Ours)	**37.09/31.06**	**32.91/27.97**	**31.63/27.10**	**30.51/24.92**	**36.68/28.29**	**35.26/29.64**

Table 1. **Quantitative comparisons on synthetic datasets.** We compare ICF-SRSR with several supervised/unsupervised methods on the benchmarks [4,24,38,39,64] and DIV2K [1] validation set for scales ×2 and ×4 with PSNR metric. ICF-SRSR refers to our self-supervised method, while EDSR (LLR,LR) is the model EDSR trained on our generated pairs (LLR,LR) of the DIV2K.

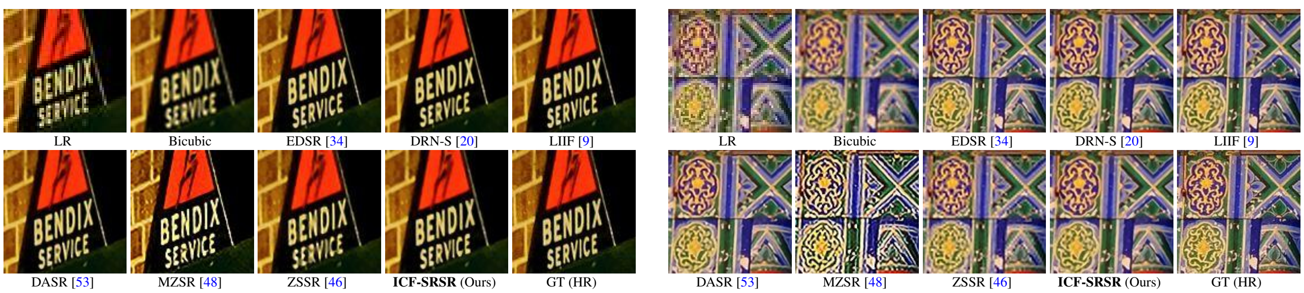


Figure 3. **Qualitative comparisons on a synthetic dataset.** We compare our ICF-SRSR method with bicubic up-scaling, supervised methods EDSR [34], DRN-S [20], and LIIF [9] and also unsupervised methods DASR [53], MZSR [48], and ZSSR [46] trained on the DIV2K [1] training set and evaluated on the DIV2K validation set for scale ×2.

rotation augmentations are applied to the input images to increase the number of effective training samples. We train our model using ADAM [29] optimizer with initial learning rate 1×10^{-4}, which decays by a factor 0.5 after every 200 epoch. For quantitative comparisons, we adopt structural similarity (SSIM) [57] and peak signal-to-noise ratio (PSNR) on the luminance channel for experiments on synthetic datasets and real-world dataset DRealSR [59] and also on RGB channels for RealSR-V3 dataset [6]. All experiments are done using PyTorch 1.8.1 and Quadro RTX 8000 GPUs.

4.2. Evaluation on synthetic datasets

We train our ICF-SRSR on the DIV2K [1] dataset with EDSR-baseline [34] and test it on the public benchmark datasets [4, 24, 38, 39, 64] and also the validation set of DIV2K. We note that the proposed method is trained in a self-supervised manner by targeting a certain scale s. Specifically, we train $(\times 2, \times 1/2)$ ICF and $(\times 4, \times 1/4)$ ICF independently. Tab. 1 shows extensive comparisons between the proposed self-supervised approach and the other representative supervised/unsupervised SR methods with the PSNR metric. We demonstrate that our ICF-SRSR approach achieves superior performance compared to the SelfExSR [24] model and comparable performance to the other unsupervised and supervised methods. We note that the ground truth HR images in Set5 and Set14 are relatively noisier than the other datasets, preventing our self-supervised framework from learning accurate scaling functions. We will discuss more details about the noisy cases in our supplementary material. Notably, ICF-SRSR outperforms unsupervised method ZSSR [46] by 1.05dB on scale ×2 of Urban100 dataset and supervised methods [9, 28] on both scales of DIV2K validation set.

Moreover, we apply the trained ICF-SRSR to LR images from the DIV2K training dataset and get LLR-LR paired images. Then, we train off-the-shelf EDSR on the synthesized paired data from scratch and evaluate it on the test datasets as shown in Tab. 1. The results demonstrate that EDSR (LLR, LR) trained on our generated pairs (LLR, LR) achieves supe-

Training Set	Supervision	Method	RealSR (Canon) ×2 (PSNR/SSIM)	RealSR (Canon) ×4 (PSNR/SSIM)	RealSR (Nikon) ×2 (PSNR/SSIM)	RealSR (Nikon) ×4 (PSNR/SSIM)	DRealSR ×2 (PSNR/SSIM)	DRealSR ×4 (PSNR/SSIM)
		Bicubic	30.35/0.876	25.80/0.744	29.66/0.854	25.50/0.718	32.67/0.877	30.56/0.820
Synthetic	Supervised	EDSR [34]	**30.58/0.880**	26.05/0.754	**30.00/0.861**	25.89/0.735	**32.82**/0.869	**30.64/0.821**
		RRDB [56]	- / -	26.05/ -	- / -	25.91/ -	- / -	30.55/ -
		IKC [18]	- / -	25.71/0.751	- / -	25.27/**0.740**	- / -	- / -
		BilndSR [10]	27.99/0.822	- / -	26.68/0.794	- / -	- / -	- / -
		DRN-S [20]	30.57/0.879	**26.07/0.755**	29.99/0.860	**25.92**/0.736	32.81/**0.879**	30.63/**0.821**
Real-world	Supervised	EDSR [34]	32.45/0.913	27.59/0.792	31.59/**0.888**	27.14/0.771	34.24/**0.908**	**32.03**/0.855
		RRDB [56]	- / -	**27.90**/ -	- / -	**27.39**/ -	33.89/0.906	31.92/0.856
		RCAN [70]	**32.69/0.919**	27.66/0.793	**31.61/0.888**	27.09/0.771	**34.34/0.908**	31.85/**0.857**
		LP-KPN [6]	- / -	27.76/**0.807**	- / -	26.34/**0.774**	33.88/ -	31.58/ -
		DRN-S [20]	32.50/0.912	27.79/0.805	31.43/0.884	- / -	33.91/0.898	- / -
	Unsupervised	ZSSR [46]+ [3]	28.79/0.826	23.68/0.673	27.54/0.799	22.46/0.645	- / -	- / -
	Self-supervised	**ICF-SRSR** (Ours)	30.98/0.885	26.27/0.763	30.31/0.864	25.89/**0.742**	32.87/0.880	30.65/0.821
		EDSR (LLR,LR) (Ours)	**31.13/0.888**	**26.32/0.764**	**30.33/0.865**	**25.92/0.742**	**32.91/0.881**	**30.68/0.823**

Table 2. **Quantitative comparison on real-world datasets.** We compare our self-supervised ICF-SRSR and EDSR (LLR,LR), *i.e.*, the model EDSR [34] trained on our generated paired dataset (LLR,LR), to several supervised/unsupervised methods trained on synthetic DIV2K [1], real-world RealSR-V3 [6] and DRealSR [59] datasets for scales ×2 and ×4 with PSNR and SSIM metrics.

rior performance than ICF-SRSR, which illustrates the merit of our method to generate useful training image pairs.

Fig. 3 further visualizes the qualitative results of ICF-SRSR on two validation images from the DIV2K dataset [1]. Our method achieves results comparable to those of the supervised methods [9, 34] while restoring more details compared to unsupervised methods [46, 48]. We note that the results in ZSSR [46] show lost information and scratched texts, and on MZSR [48] include severe artifacts and color shifting. For an in-depth comparison, we also provide quantitative results with SSIM metric in our supplementary material.

4.3. Evaluation on real-world datasets

We train and evaluate ICF-SRSR for each scale ×2 and ×4 independently on the LR images of each Canon and Nikon camera from the real-world dataset RealSR-V3 [6] separately and also on the LR images of the real-world dataset DRealSR [59] in a self-supervised manner. We further train the model EDSR [34] on our generated (LLR, LR) image pairs. We compare our method with the supervised methods [6, 20, 34, 56, 70] trained on real paired images, which serve as the upper bounds for the SR problem.

On the other hand, we employ the pre-trained supervised models EDSR [34], RRDB [56], IKC [18], BlindSR [10] and DRN-S [20] on the synthetic DIV2K [1] dataset to super-resolve the LR images in the testing sets of RealSR-V3 [6] and DRealSR [59]. Moreover, we utilize Kernel-GAN [3] to approximate the down-sampling kernel from a single LR image and use ZSSR [46] as a zero-shot SR to apply to real LR images. Our extensive comparisons with the various methods trained on real and synthetic datasets are summarized in Tab. 2. We illustrate that our self-supervised method can achieve superior performance compared to the methods pre-trained on the synthetic datasets and unsupervised method ZSSR [46]+Kernel-GAN [3] in terms of both PSNR and SSIM metrics, which emphasizes the fact that the trained models on synthetic datasets with known degradations cannot perform well on real-world scenarios. We qualitatively compare our method with the various existing methods on the RealSR-V3 dataset and visualize the SR results and their corresponding error maps with respect to the GT (HR) in Fig. 4. We demonstrate that our self-supervised method can achieve comparable and sometimes better performance to the supervised method LP-KPN [6] trained on real paired images. We note that our method is generally more suitable for restoring the texture and preserving color compared to supervised method IKC [18] and unsupervised method ZSSR [46]+Kernel-GAN [3] as evident in appearance and PSNR, SSIM, and mean absolute error (MAE) metrics. We show more qualitative results in the supplementary material.

4.4. Ablation study

We conduct various ablation studies on the model design, down-sampling operators, few-shot learning, augmentation, and the effect of loss functions to better analyze our method.

Model design. We conduct an experiment to show the superiority of a developed baseline as a single conditional model compared to two independent models and also the effect of training our two-stage framework compared to training each Up-Down and Down-Up stage separately. Our results on synthetic dataset DIV2K [1] and Canon and Nikon images from real-world dataset RealSR-V3 [6] for scale ×2

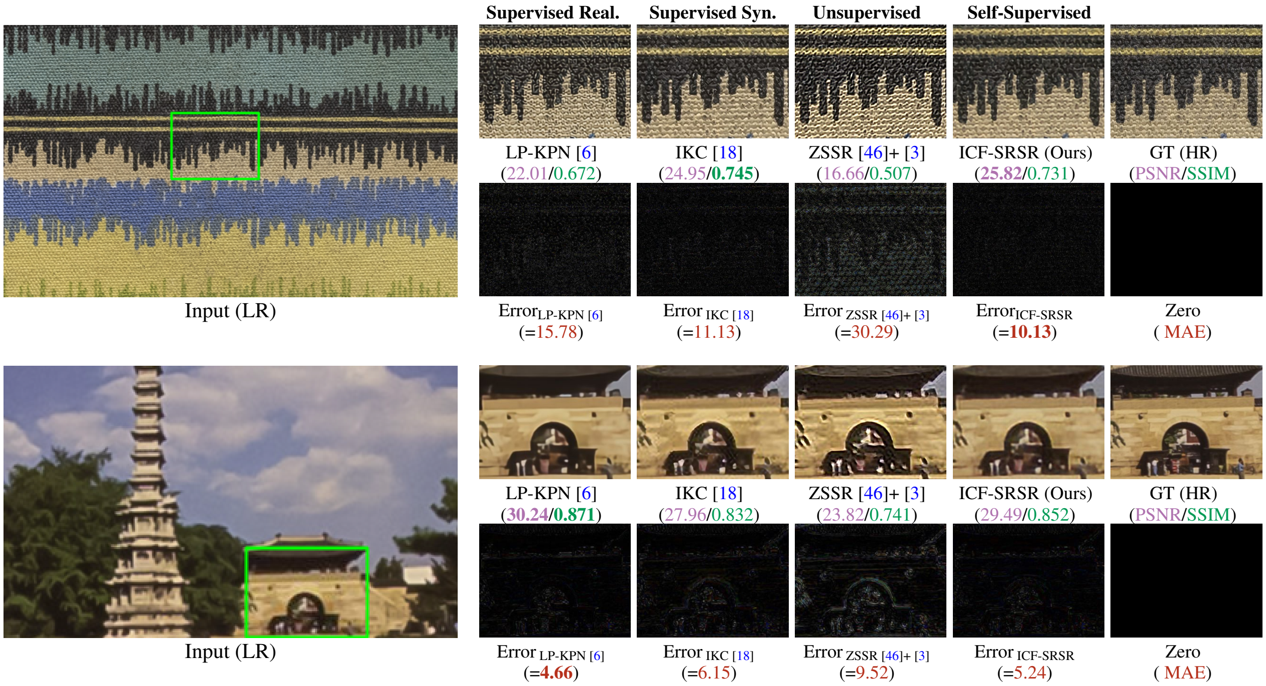


Figure 4. **Qualitative comparisons on a real-world dataset.** We visualize the super-resolution results (first row) and their corresponding error maps with respect to the GT (second row) for an image captured by each Nikon and Canon camera. We compare our self-supervised method ICF-SRSR with the supervised method LP-KPN [6] and the unsupervised method ZSSR [46]+ [3] trained on the RealSR-V3 [6] dataset and the supervised method IKC [18] trained on synthetic dataset DIV2K [1] for scale $\times 4$ with PSNR, SSIM, and MAE metrics.

show that training with two independent models or using only one stage (half) results in unsatisfactory performance, demonstrating the uniqueness of our method in using a single invertible scale-conditional model as shown in Tab. 3.

Method	**DIV2K** ($\times 2$)	**Canon** ($\times 2$)	**Nikon** ($\times 2$)
Two Models	34.81	30.61	30.01
Up-Down	29.92	28.56	27.52
Down-Up	34.59	30.58	30.00
ICF-SRSR	**35.19**	**30.98**	**30.31**

Table 3. **Ablation on model design.**

Evaluation of down-sampling. Due to the invertibility attribute of ICF, our method can be interpreted as a learnable down-sampler. Therefore, we analyze our model f_θ as a down-sampling operator in three aspects.

First. We train ICF-SRSR on HR images from RealSR-V3 [6] and evaluate the model on HR images of the test dataset to gather the generated down-sampled images. Then, we compare ground-truth LR images with our generated LR images, as well as LR images obtained by down-sampling functions *e.g.*, Nearest, Bicubic, Gaussian+Nearest, and Gaussian+bicubic ($\sigma = 0.4$). Tab. 4 provides a comparison of LR images for different down-sampling models based on PSNR. The values show the superiority of our learnable down-sampling method in generating more realistic LR images compared to ones with other down-sampling operators.

Down-sampling	**Canon**		**Nikon**	
	$\times 2$	$\times 4$	$\times 2$	$\times 4$
Nearest	29.35	24.51	28.54	23.91
Bicubic	30.27	25.76	29.71	25.56
Gaussian+Nearest	29.62	24.65	28.87	24.09
Gaussian+Bicubic	30.61	25.95	30.12	25.81
ICF-SRSR	**32.46**	**28.93**	**32.12**	**29.15**

Table 4. **Ablation on down-sampling performance.**

Second. We further analyze our learnable down-sampling operator f_θ compared to non-learnable down-sampling approaches. We use our learnable down-sampling operator f_θ, bicubic down-sampling, and Gaussian ($\sigma = 0.4$) filtering followed by different nearest and bicubic down-sampling operators to generate the LLR images from given input LR images on the training sets. Then, we train the model EDSR on the generated paired images (LLR, LR) to learn generating SR images given LR counterparts. We summarize the results for scale $\times 2$ of the benchmarks Set5 [4] and Set14 [64], and Canon and Nikon sets of RealSR-V3 [6] dataset for both non-learnable and our learnable down-sampling operators in Tab. 5. The results indicate the effect of our learnable down-sampling operator to generate appropriate image pairs for training, which results in a significant improvement compared to known down-sampling operators.

Third. By using different down-sampling methods, we first generate LR samples from the real training HR images and

Down-sampling	**Set5**	**Set14**	**Canon**	**Nikon**
Bicubic	35.30	31.53	30.41	29.80
Gaussian+Nearest	30.79	28.39	29.41	28.60
Gaussian+Bicubic	35.43	31.84	30.47	29.86
ICF-SRSR	**37.09**	**32.91**	**31.13**	**30.33**

Table 5. **Comparison with non-learnable down-sampling operators to generate paired training data for SR task.**

then train a vanilla EDSR model using the generated pairs, *i.e.*, (LR, HR). As shown in Tab. 6, our synthesized pairs can provide more suitable training data compared to ones by previous learnable down-sampling methods ADL [49] and DRN-S [20] as the EDSR performs much better for the $\times 2$ SR tasks on real dataset RealSR-V3 [6].

Downsampling	**Canon** ($\times 2$)	**Nikon** ($\times 2$)
ADL [49]	30.76	30.44
DRN-S [20]	30.82	30.24
ICF-SRSR	**31.94**	**31.24**

Table 6. **Comparison with learnable down-sampling operators to generate paired training data for SR task.**

Few-shot learning. We train and evaluate our method on small datasets to show the advantage of our method to learning from only a few images without requiring a large-scale training dataset. Therefore, we train the model ICF-SRSR (Small) on the test sets of synthetic datasets Set14 [64], BSD100 [38] and Urban100 [24] and also real-world datasets RealSR-V3 [6] and DRealSR [59] and show their results on the corresponding test datasets in Tab. 7. We demonstrate that our method can achieve slightly lower performance even when trained on very small datasets compared to our model ICF-SRSR (Large) trained on large-scale training datasets.

Training set	**Set14**		**BSD100**		**Urban100**	
	$\times 2$	$\times 4$	$\times 2$	$\times 4$	$\times 2$	$\times 4$
Large	32.86	27.76	31.54	26.99	30.39	24.72
Small	32.44	27.19	31.34	26.82	30.26	24.66
Training set	**Canon**		**Nikon**		**DRealSR**	
	$\times 2$	$\times 4$	$\times 2$	$\times 4$	$\times 2$	$\times 4$
Large	30.98	26.26	30.31	25.89	32.87	30.65
Small	30.67	26.08	29.99	25.76	32.83	30.62

Table 7. **Few-shot learning.**

Multi-scale augmentation. As we mention in Sec. 3.4, augmented data with different scales can lead to performance improvement. Therefore, when we train ICF-SRSR directly on the test samples, we adopt diverse scaling factors as well as their reciprocals to compensate for the limited number of training data. In Tab. 8, we show that increasing the number of inputs induced by various scaling factors, *e.g.*, $\times 2$, $\times 4$, and $\times 8$, and their inverses can lead to obtaining superior performance on the RealSR-V3 [6] dataset. More details about our multi-scale augmentation strategy are described in our supplementary material.

Scale	**Canon** ($\times 2$)	**Nikon** ($\times 2$)
2	30.67	29.99
2,4	30.75	30.09
2,4,8	**30.78**	**30.11**

Table 8. **Multi-scale augmentation.**

Effects of loss functions. We analyze the effects of losses discussed in Sec. 3.3. As shown in Tab. 9, our novel self-supervised consistency loss $\mathcal{L}^{\text{Cons}}$ can drastically improve the performance when it is added to the loss $\mathcal{L}^{\text{Color}}$ on both synthetic and real-world datasets. In our supplementary material, we further discuss the effect of the weight λ_{Color}.

Loss	**DIV2K** ($\times 2$)	**Canon** ($\times 2$)	**Nikon** ($\times 2$)
$\mathcal{L}^{\text{Color}}$ only	30.31	29.12	28.38
$\mathcal{L}^{\text{Color}}$,$\mathcal{L}^{\text{Cons}}$	**35.19**	**30.98**	**30.31**

Table 9. **Effect of loss functions.**

5. Conclusion

We propose ICF, a novel invertible scale-conditional function that receives an image and an arbitrary scaling factor and generates the resized image, and can reconstruct the same input image by the given resized image and the inverse scaling factor. Then, we utilize ICF to design a self-supervised real-world single-image super-resolution framework ICF-SRSR. Accordingly, our framework is able to generate up-sampled and down-sampled images simultaneously, where the generated down-sampled images can be used to construct paired images appropriate for training existing models. Extensive experiments demonstrate the strengths of our self-supervised method on both synthetic and real-world datasets and superior performance on the real-world dataset compared to supervised models trained on the synthetic datasets.

Limitations and future works. One remaining limitation is that we only apply our method to a few real-world datasets due to the lack of aligned LR-HR image pairs for evaluation in other real-world datasets. Therefore, we aim to provide a large-scale real-world dataset from various scenes for better evaluation in our future work. Moreover, we will investigate the applications of our defined ICF to self-supervised image warping and other image restoration tasks.

Acknowledgements. This work was supported in part by the IITP grants [No.2021-0-01343, Artificial Intelligence Graduate School Program (Seoul National University), No.2021-0-02068, and No.2023-0-00156], the NRF grant [No.2021M3A9E4080782] funded by Korea government (MSIT).

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *CVPR Workshops*, 2017. 4, 5, 6, 7

[2] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. Fast, accurate, and lightweight super-resolution with cascading residual network. In *ECCV*, 2018. 5

[3] Sefi Bell-Kligler, Assaf Shocher, and Michal Irani. Blind super-resolution kernel estimation using an internal-gan. In *NeurIPS*, 2019. 2, 6, 7

[4] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie Line Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In *BMVC*, 2012. 4, 5, 7

[5] Adrian Bulat, Jing Yang, and Georgios Tzimiropoulos. To learn image super-resolution, use a GAN to learn how to do image degradation first. In *ECCV*, 2018. 1, 2

[6] Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. In *ICCV*, 2019. 1, 3, 4, 5, 6, 7, 8

[7] Chang Chen, Zhiwei Xiong, Xinmei Tian, Zheng-Jun Zha, and Feng Wu. Camera lens super-resolution. In *CVPR*, 2019. 1, 3

[8] Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao. Pre-trained image processing transformer. In *CVPR*, 2021. 2

[9] Yinbo Chen, Sifei Liu, and Xiaolong Wang. Learning continuous image representation with local implicit image function. In *CVPR*, 2021. 2, 5, 6

[10] Victor Cornillère, Abdelaziz Djelouah, Wang Yifan, Olga Sorkine-Hornung, and Christopher Schroers. Blind image super-resolution with spatially variant degradations. *ACM TOG*, 2019. 2, 6

[11] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. Second-order attention network for single image super-resolution. In *CVPR*, 2019. 2

[12] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *ECCV*, 2014. 2

[13] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In *ECCV*, 2016. 2

[14] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *ICLR*, 2021. 2

[15] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *ICML*, 2017. 2

[16] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *NeurIPS*, 2014. 1, 2

[17] Hayit Greenspan, Sharon Peled, Gal Oz, and Nahum Kiryati. Mri inter-slice reconstruction using super-resolution. In *MICCAI*, 2001. 1

[18] Jinjin Gu, Hannan Lu, Wangmeng Zuo, and Chao Dong. Blind super-resolution with iterative kernel correction. In *CVPR*, 2019. 2, 6, 7

[19] Bahadir K Gunturk, Aziz Umit Batur, Yucel Altunbasak, Monson H Hayes, and Russell M Mersereau. Eigenface-domain super-resolution for face recognition. *IEEE TIP*, 2003. 1

[20] Yong Guo, Jian Chen, Jingdong Wang, Qi Chen, Jiezhang Cao, Zeshuai Deng, Yanwu Xu, and Mingkui Tan. Closed-loop matters: Dual regression networks for single image super-resolution. In *CVPR*, 2020. 2, 5, 6, 8

[21] Muhammad Haris, Gregory Shakhnarovich, and Norimichi Ukita. Deep back-projection networks for super-resolution. In *CVPR*, 2018. 2

[22] Muhammad Haris, Gregory Shakhnarovich, and Norimichi Ukita. Recurrent back-projection network for video super-resolution. In *CVPR*, 2019. 2

[23] Xuecai Hu, Haoyuan Mu, Xiangyu Zhang, Zilei Wang, Tieniu Tan, and Jian Sun. Meta-SR: A magnification-arbitrary network for super-resolution. In *CVPR*, 2019. 2

[24] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *CVPR*, 2015. 4, 5, 8

[25] Yan Huang, Shang Li, Liang Wang, Tieniu Tan, et al. Unfolding the alternating optimization for blind super resolution. In *NeurIPS*, 2020. 2

[26] Shady Abu Hussein, Tom Tirer, and Raja Giryes. Correction filter for single image super-resolution: Robustifying off-the-shelf deep super-resolvers. In *CVPR*, 2020. 2

[27] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *CVPR*, 2017. 2

[28] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *CVPR*, 2016. 2, 5

[29] Diederik P Kingma and J Ba. Adam: Amethod for stochastic optimization. In *ICLR*, 2015. 5

[30] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *CVPR*, 2017. 2

[31] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *CVPR*, 2017. 2

[32] Xiaodong Li, Yun Du, and Feng Ling. Sub-pixel-scale land cover map updating by integrating change detection and sub-pixel mapping. *PE&RS*, 2015. 1

[33] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *ICCV*, 2021. 2

[34] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *CVPR Workshops*, 2017. 2, 4, 5, 6

[35] Frank Lin, Clinton Fookes, Vinod Chandran, and Subramanian Sridharan. Investigation into optical flow super-resolution for surveillance applications. In *WDIC APRS*, 2005. 1

[36] Andreas Lugmayr, Martin Danelljan, and Radu Timofte. Unsupervised learning for real-world super-resolution. In *ICCV Workshops*, 2019. 1, 2

[37] Shunta Maeda. Unpaired image super-resolution using pseudo-supervision. In *CVPR*, 2020. 1, 2

[38] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *ICCV*, 2001. 4, 5, 8

[39] Yusuke Matsui, Kota Ito, Yuji Aramaki, Azuma Fujimoto, Toru Ogawa, Toshihiko Yamasaki, and Kiyoharu Aizawa. Sketch-based manga retrieval using manga109 dataset. *Multimedia Tools and Applications*, 2017. 4, 5

[40] Yiqun Mei, Yuchen Fan, and Yuqian Zhou. Image super-resolution with non-local sparse attention. In *CVPR*, 2021. 2

[41] Tomer Michaeli and Michal Irani. Nonparametric blind super-resolution. In *CVPR*, 2013. 2

[42] Ben Niu, Weilei Wen, Wenqi Ren, Xiangde Zhang, Lianping Yang, Shuzhen Wang, Kaihao Zhang, Xiaochun Cao, and Haifeng Shen. Single image super-resolution via a holistic attention network. In *ECCV*, 2020. 2

[43] Sharon Peled and Yehezkel Yeshurun. Superresolution in mri: application to human white matter fiber tract visualization by diffusion tensor imaging. *MRM*, 2001. 1

[44] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv*, 2015. 2

[45] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *CVPR*, 2016. 2

[46] Assaf Shocher, Nadav Cohen, and Michal Irani. "zero-shot" super-resolution using deep internal learning. In *CVPR*, 2018. 1, 2, 5, 6, 7

[47] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In *NeurIPS*, 2020. 2

[48] Jae Woong Soh, Sunwoo Cho, and Nam Ik Cho. Meta-transfer learning for zero-shot super-resolution. In *CVPR*, 2020. 1, 2, 5, 6

[49] Sanghyun Son, Jaeha Kim, Wei-Sheng Lai, Ming-Hsuan Yang, and Kyoung Mu Lee. Toward real-world super-resolution via adaptive downsampling models. *IEEE TPAMI*, 2021. 1, 2, 4, 8

[50] Sanghyun Son and Kyoung Mu Lee. SRWarp: Generalized image super-resolution under arbitrary transformation. In *CVPR*, 2021. 2

[51] Andrew J Tatem, Hugh G Lewis, Peter M Atkinson, and Mark S Nixon. Super-resolution land cover pattern prediction using a hopfield neural network. *RSE*, 2002. 1

[52] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *CVPR*, 2018. 2

[53] Longguang Wang, Yingqian Wang, Xiaoyu Dong, Qingyu Xu, Jungang Yang, Wei An, and Yulan Guo. Unsupervised degradation representation learning for blind super-resolution. In *CVPR*, 2021. 1, 5

[54] Longguang Wang, Yingqian Wang, Zaiping Lin Lin, Jungang Yang, Wei An, and Yulan Guo. Learning for scale-arbitrary super-resolution from scale-specific networks. *arXiv*, 2020. 2

[55] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data. In *ICCV Workshops*, 2021. 3

[56] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. ESRGAN: Enhanced super-resolution generative adversarial networks. In *ECCV Workshops*, 2018. 2, 6

[57] Zhou Wang, Alan C Bovik, Hamid R Sheikh, Eero P Simoncelli, et al. Image quality assessment: From error visibility to structural similarity. *IEEE TIP*, 2004. 5

[58] Zhendong Wang, Xiaodong Cun, Jianmin Bao, and Jianzhuang Liu. Uformer: A general u-shaped transformer for image restoration. *arXiv*, 2021. 2

[59] Pengxu Wei, Ziwei Xie, Hannan Lu, Zongyuan Zhan, Qixiang Ye, Wangmeng Zuo, and Liang Lin. Component divide-and-conquer for real-world image super-resolution. In *ECCV*, 2020. 3, 4, 5, 6, 8

[60] Frederick W Wheeler, Xiaoming Liu, and Peter H Tu. Multi-frame super-resolution for face recognition. In *BTAS*, 2007. 1

[61] Xiangyu Xu, Yongrui Ma, and Wenxiu Sun. Towards real scene super-resolution with raw images. In *CVPR*, 2019. 3

[62] Yuan Yuan, Siyuan Liu, Jiawei Zhang, Yongbing Zhang, Chao Dong, and Liang Lin. Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. In *CVPR Workshops*, 2018. 1, 2

[63] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. *arXiv*, 2021. 2

[64] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *ICCS*, 2010. 4, 5, 7, 8

[65] Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In *ICCV*, 2021. 3

[66] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Learning a single convolutional super-resolution network for multiple degradations. In *CVPR*, 2018. 2

[67] Liangpei Zhang, Hongyan Zhang, Huanfeng Shen, and Pingxiang Li. A super-resolution reconstruction algorithm for surveillance images. *Signal Processing*, 2010. 1

[68] Xuaner Zhang, Qifeng Chen, Ren Ng, and Vladlen Koltun. Zoom to learn, learn to zoom. In *CVPR*, 2019. 1, 3

[69] Xindong Zhang, Hui Zeng, Shi Guo, and Lei Zhang. Efficient long-range attention network for image super-resolution. In *ECCV*, 2022. 5

[70] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *ECCV*, 2018. 2, 5, 6

[71] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *CVPR*, 2018. 2, 5

[72] Tianyu Zhao, Wenqi Ren, Changqing Zhang, Dongwei Ren, and Qinghua Hu. Unsupervised degradation learning for single image super-resolution. *arxiv*, 2018. 1, 2

[73] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, 2017. 2, 4