Dual Adversarial Adaptation for
Cross-Device Real-World Image Super-Resolution

Xiaoqian Xu[1] Pengxu Wei[*1] Weikai Chen[2] Yang Liu[1] Mingzhi Mao[1] Liang Lin[1] Guanbin Li[1]

[1]Sun Yat-sen University [2]Tencent America

Abstract

Due to the sophisticated imaging process, an identical scene captured by different cameras could exhibit distinct imaging patterns, introducing distinct proficiency among the super-resolution (SR) models trained on images from different devices. In this paper, we investigate a novel and practical task coded cross-device SR, which strives to adapt a real-world SR model trained on the paired images captured by one camera to low-resolution (LR) images captured by arbitrary target devices. The proposed task is highly challenging due to the absence of paired data from various imaging devices. To address this issue, we propose an unsupervised domain adaptation mechanism for real-world SR, named Dual ADversarial Adaptation *(DADA), which only requires LR images in the target domain with available real paired data from a source camera. DADA employs the Domain-Invariant Attention (DIA) module to establish the basis of target model training even without HR supervision. Furthermore, the dual framework of DADA facilitates an Inter-domain Adversarial Adaptation (InterAA) in one branch* for *two LR input images* from two domains, *and an Intra-domain Adversarial Adaptation (IntraAA) in* two branches *for* an LR input image. *InterAA and IntraAA together improve the model transferability from the source domain to the target. We empirically conduct experiments under six* Real→Real *adaptation settings among three different cameras, and achieve superior performance compared with existing state-of-the-art approaches. We also evaluate the proposed DADA to address the adaptation to the video camera, which presents a promising research topic to promote the wide applications of real-world super-resolution. Our source code is publicly available at* https://github.com/lonelyhope/DADA.

1. Introduction

Single image super-resolution (SISR), which super-resolves low-resolution images (LRs) and reconstructs their

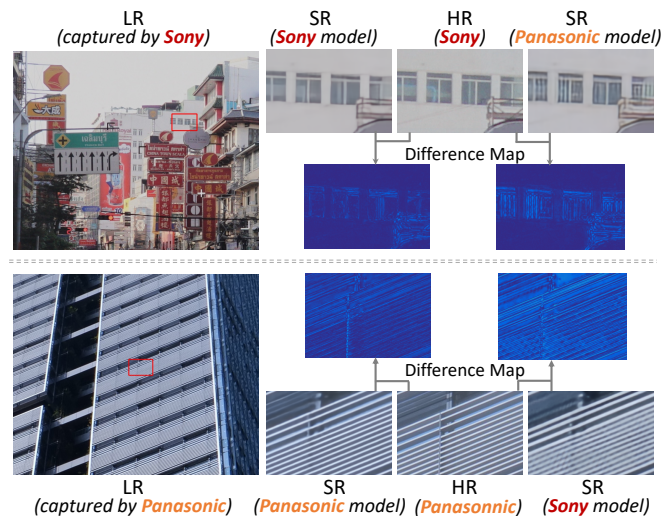*Corresponding author: weipx3@mail.sysu.edu.cn

Figure 1. Comparison of real SR results across different trained models from different camera data in the DRealSR dataset [18]. The Difference Map denotes the absolute difference between ground-truth HR and SR image (The brightness in the map reflects the magnitude of the difference).
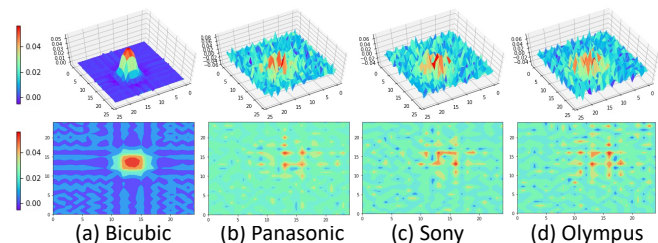

Figure 2. Degradation kernels for bicubic down-sampling and three different cameras in DRealSR [18]. USRNet [22] is employed to estimate those kernels by minimizing $\|(HR \otimes k) \downarrow_s -LR\|$, where $\otimes$ means convolution, and $\downarrow_s$ means down-sampling for scaling factor s by choosing the upper left item for every $s * s$ grid. The kernel size is set to $25 * 25$ and the scaling factor s is 4.

high-resolution counterparts (HRs), is a fundamental task in low-level computer vision. The emergence of deep learning significantly contributes to the SR progress and SISR is usually cast as a supervised learning task with paired LR-HR
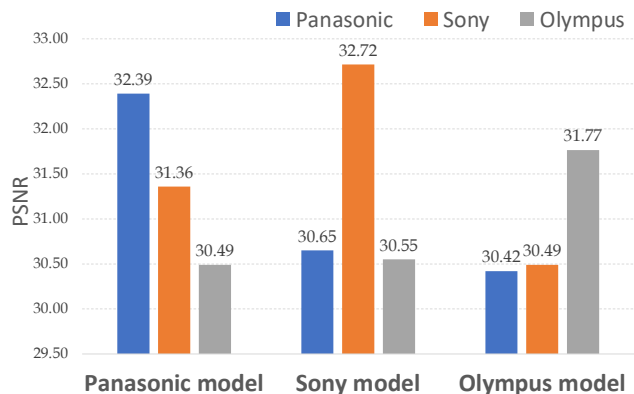
Figure 3. Cross-device evaluation of real SR models trained from images of individual cameras. We use models trained on the data collected by different cameras to test the images taken by a specific camera.
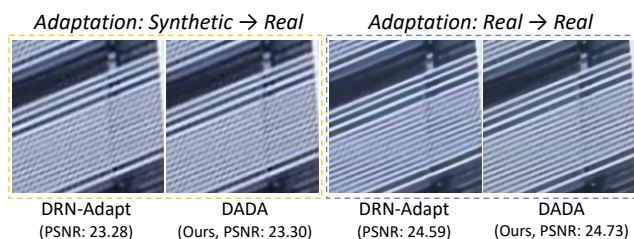

Figure 4. SR result comparison between synthetic to real adaptation and real to real adaptation (*Sony → Panasonic*).

images [3, 7, 8, 16, 24]. However, due to the difficulty of collecting LR-HR image pairs, typical learning-based methods learn to map synthetic low-resolution images to the original counterpart to achieve super-resolution reconstruction, which is constantly criticized with poor model generalization in practical scenarios. Accordingly, real SR has come up to explore the real image degradation and several real-world SR datasets collected by capturing paired LR-HR data via optical zoom of DSLR cameras emerged immediately, *e.g.*, *RealSR* [1] and *DRealSR* [18]. Considering the vulnerability of deep networks for quasi-imperceptible noises, robust real-world SR is further investigated to defense the adversarial attacks for practical applications [21].

In comparison with synthetic image degradation, real-world SR exhibits a crucial challenge, *i.e.*, *diverse degradation processes among devices*, due to different imaging processes across devices, especially across different camera types. The *cross-device domain gap* is demonstrated in Fig. 1. It is observed that in cross-device real SR model evaluation (Sony→Panasonic, or Panasonic→Sony), it presents a distinct device-specific degradation gap: compared with that trained on Panasonic (Sony) images, an SR model trained only on Sony (Panasonic) images (*i.e.*, *Sony* (*Panasonic*) model) super-resolve an LR image, captured by Panasonic (Sony) camera, into an SR image that has more blurry details or even distorted artifacts and larger differ-

ences from ground truth HR image. To explain this phenomenon, degradation kernels for different (camera) degradations are analyzed in Fig.2. Real image degradation kernels are different across different cameras. We regard device-specific degradation gap as *domain gap* in this work and also empirically demonstrate its consequence of the performance degradation in cross-device/domain setting in Fig. 3. However, this domain gap is ubiquitous for many realistic applications, *e.g.*, image/video enhancement for all kinds of phones or GoPro cameras, and classic old movie restoration. Usually, it is extremely labor-intensive and difficult to collect paired data for each camera, and even is impossible to obtain paired data, *e.g.*, classic old movies.

To mitigate this issue, we are the first to explore Unsupervised Domain Adaptation (UDA) for Real-World Image Super-Resolution across Devices. Under this setting, given paired real LR-HR images captured by one camera (*source camera/domain*), the goal is to adapt the model to the *target domain* that has only LR images captured by another camera (*Real → Real* adaptation). This is more rational than conventional UDA SR from source domain with paired synthesized LR-HR images to target domain with real images (*Synthetic → Real* adaptation). As shown in Fig. 2, synthetic degradation (*e.g.*, widely-used bicubic downsampling) is engaged with simple kernels; realistic degradation is heterogeneous and more complex. The complexity of real kernels brings challenges to the *Real → Real* adaptation task. Due to the significant distinction between synthetic and real degradation, it is certainly difficult to coordinate a source model of synthetic degradation to a target domain with realistic data, and has an inferior performance in the target domain. This is evidenced in Fig. 4. Overall, our UDA SR across devices, namely *Real → Real* adaptation, is more practical for realistic applications.

In this paper, we propose a Dual ADversarial Adaptation model (DADA) to explore unsupervised domain adaptation for real-world image super-resolution across devices. Rooted in the Component Divide-and-Conquer model (CDC) [18], DADA has a source branch and a target branch in a symmetric architecture and each branch is essentially a cycle image reconstruction with an up-sampling module and a down-sampling module. DADA employs the Domain-Invariant Attention (DIA) module to provide component guidance masks for up-sampling modules in two branches for each input. Additionally, the dual framework of DADA facilitates an Inter-domain Adversarial Adaptation (InterAA) in *one branch* for *two LR input images* from two domains, and an Intra-domain Adversarial Adaptation (IntraAA) in *two branches* for *an LR input image*. InterAA and IntraAA together improve the model transferability from the source domain to the target.

In summary, our main contributions are three-fold:

- We are devoted to the early attempt to explore the

cross-device domain gap in real-world image super-resolution. To mitigate this issue, a Dual ADversarial Adaptation model (DADA) is proposed for unsupervised domain adaptation from the source domain with paired real data to the target with only real LR images.

- We propose the Inter-domain adversarial Adaptation (InterAA) and the Intra-domain adversarial Adaptation (IntraAA) to train the model in a dual architecture for unsupervised *Real* to *Real* SR adaptation.

- Extensive experiments on six *Real* to *Real* adaptation settings among three different camera domains demonstrate the superiority of our DADA over traditional SR methods on real-world image super-resolution when adapting the model across different cameras.

2. Related Work

2.1. Image Super-Resolution

With a remarkable feature learning ability, Convolutional Neural Network (CNN) based methods have brought considerable improvements in the field of single image super-resolution compared to traditional methods. SRCNN [3] is the first to employ the CNN network in the SR task, which is an end-to-end three-layer CNN network to learn the feature mapping from input LR images to HR images. Subsequently, much deeper and more complex networks have been proposed, *e.g.*, ESPCN [12], SRResNet [7], EDSR [8], SRDenseNet [14], RCAN [24] and ESRGAN [16], introducing structures such as dense connections, attention modules and non-local modules, and the SR performance is constantly improved. However, due to the difficulty of collecting real LR-HR image pairs, those deep learning-based approaches synthetically down-sample HR images into their LR counterparts. With synthetic LR-HR pairs, it essentially casts SISR as a supervised learning problem. However, simple downsampling methods cannot simulate the realistic degradation, causing performance degradation when directly applying the trained model in practice.

2.2. Real-World Image Super-Resolution

To break the data bottleneck of synthetic degradation, real-world image super-resolution attracts increasing interest. A well-prepared real-world SR dataset is RealSR [1], which has paired LR-HR images captured by zooming lens of DSLR cameras. Subsequently, DRealSR, as a more challenging large-scale real-world SR dataset, has been built with five DSLR cameras. Due to the vulnerability of deep neural networks, how adversarial perturbations affect SR models is explored for robust real-world super-resolution in the low-level vision [21]. However, existing real SR research treats images from different cameras equally without discrimination and the cross-device domain gap, brought by different realistic degradations derived from different cameras. This is ignored in existing works, let alone conventional SR with synthetic image degradation (it is limited to uniform and simple down-sampling degradation kernels, *e.g.*, bicubic down-sampling, for images no matter which camera they are captured by).

2.3. Unsupervised Domain Adaptation in SR

Most UDA researches are on high-level vision tasks, *e.g.*, image classification, object detection and semantic segmentation [2, 4, 6, 9, 11, 15, 26]. Few are devoted to low-level vision, which is more challenging for UDA since it concentrates more on pixel-wise adaptation and is not easy just by simple feature alignment or sample distribution alignment like UDA in high-level vision. Conventional UDA SR generally aims to leverage abundant paired synthetic LR-HR images (source domain) to transfer the model to super-resolve real images (target domain). Motivated by Cycle-GAN [25], Yuan et al. [20] propose the CinCGAN model, to transfer real-world LR images to synthetic LR images. Guo et al. [5] propose a method that trains the synthetic paired data and unpaired real data together with cyclic constraints. Wei et al. [19] proposed a model named DASR, which uses unpaired real images to handle the real-world SR problem. They use domain-gap aware training and domain-distance weighted supervision strategies to narrow the domain gap between source data and target data. However, these methods are not aware of the divergence of realistic image degradation between different cameras. We term this divergence as cross-device domain gap and propose a DADA model to address this *Real* to *Real* unsupervised domain adaptation.

3. Methodology

We consider the real-world SR problem of unsupervised domain adaptation across devices, where real LR-HR image pairs from a camera are provided in the source domain, while only real LR images from another camera are accessible in the target domain. Noted that the device domain gap mentioned here stems from the imaging differences of two different models of cameras for the same scene[1]. With source LR-HR real image pairs and target real LR images, we aim to train a UDA model for real SR in the target domain. In this section, we will elaborate on the proposed DADA to explore the unsupervised domain adaptation problem for real-world image super-resolution across devices.

---

[1]Different cameras might have different Image Signal Processors (ISPs) for image degradation. Existing real SR datasets are collected by different cameras with different types/brands and do not include different cameras with the same ones. In this work, different cameras/devices specifically indicate different camera brands, namely different domains.
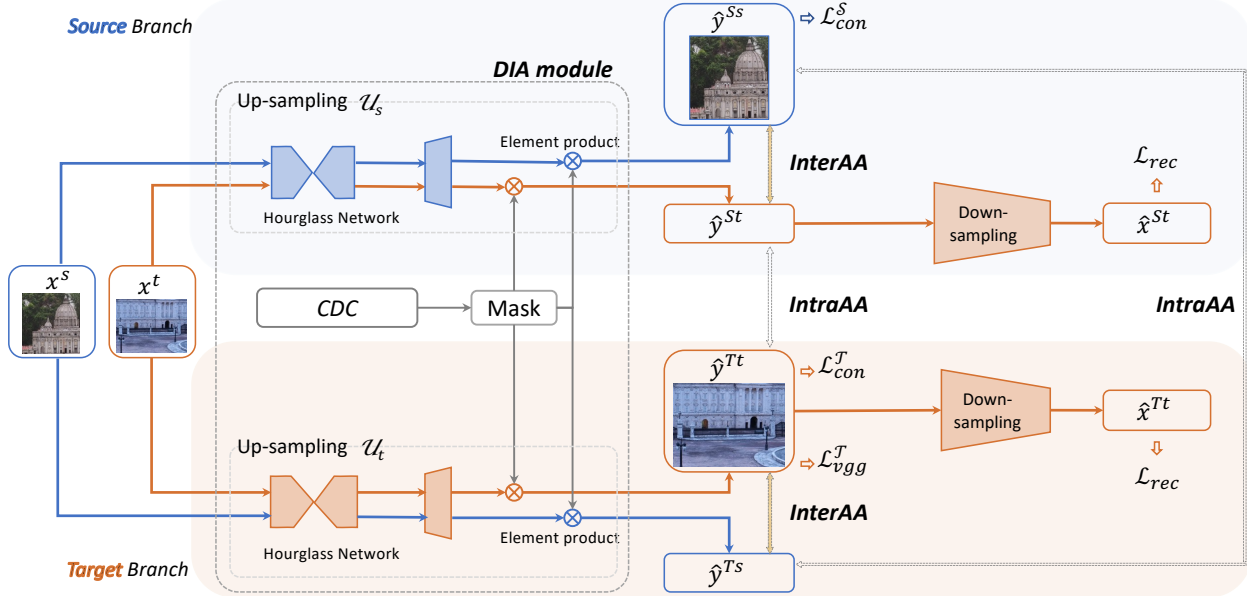
Figure 5. The proposed framework of our DADA. It includes a source branch and a target branch, both of which have an LR-HR-LR reconstruction structure. In their up-sampling networks, a Domain-Invariant Attention (DIA) module provides component attentive masks by a pre-trained CDC model. For InterAA, two LR images from different domains are fed into the up-sampling network for adversarial adaptation in one branch. For IntraAA, one LR image is respectively fed into up-sampling networks in source and target branches for the adversarial adaptation. For testing, the trained up-sampling network in target branch is utilized for inference.

3.1. Overview

(Close-set) UDA in high-level vision has an underlying prerequisite that the source domain and the target domain share the same semantic categories. This is convenient to learn domain-invariant semantic features as the basis for model adaptation. However, for UDA in real SR, it is intractable to figure out what is essentially the basis for adaptation. To address this issue, on one hand, our DADA leverages the stability of extracting mid-level image components from low-level image pixels, to build a domain-invariant attention module, rather than learning domain-invariant features. On the other hand, the proposed DADA model inherits the cycle-consistent reconstruction structure (*LR→HR→LR*) for the target domain. Considering the inaccessibility of HR images in the target domain, DADA employs a training strategy with inter-domain and intra-domain adversarial adaptation in a dual architecture. In Fig. 5, our DADA consists of two symmetrical branches. Each branch is a *LR→HR→LR* reconstruction network, including an up-sampling module and a down-sampling module. One branch is dominated by source data with paired LR-HR supervisions, named *source branch*; the other is mainly responsible for the target domain, named *target branch*. Taking a source LR image x^s and a target LR image x^t as inputs, they are sent to the up-sampling modules in these two branches respectively, resulting in four SR outputs. We use the *Inter-domain Adversarial Adaptation* (InterAA) scheme to process SR outputs from different inputs of the same branch. In contrast, the *Intra-domain Ad-versarial Adaptation* (IntraAA) scheme processes SR outputs from the same inputs in different branches. As stated in CDC [18], image components (*i.e.*, the flat, edges and corner points) are content-dependent, and thus they are domain-invariant and loosely influenced by different cameras. This motivates us to establish a domain-invariant attention module based on a pre-trained CDC. The two branches are trained collaboratively with inter-domain and intra-domain adversarial adaptation.

3.2. Dual Adversarial Adaptation Model

For source domain, paired data $\{x_i^s, y_i^s\}_{i=1,...,M}$ are accessible, where x_i^s is the LR image and y_i^s is the corresponding HR image. For the target domain, only LR images $\{x_j^t\}_{j=1,...,N}$ can be accessed and their HR counterparts $\{y_j^t\}$ are unknown.

DADA takes a source LR image x_i^s and a target LR image x_j^t as inputs. Both LR images are fed into the source and target branches together, respectively. In one branch, besides the *LR→HR→LR* reconstruction process, x_i^s and x_j^t are re-solved into SR images by the same up-sampling module (a generator), which are further sent to a discriminator for source/target domain discrimination. We name this process *Inter-domain Adversarial Adaptation* (InterAA). Between two branches, a target (source) LR image goes through the *LR→HR→LR* reconstruction in each branch, and the SR images generated by the upsampling modules of the two branches will be further sent to a branch discriminator to distinguish their branch source. We name this
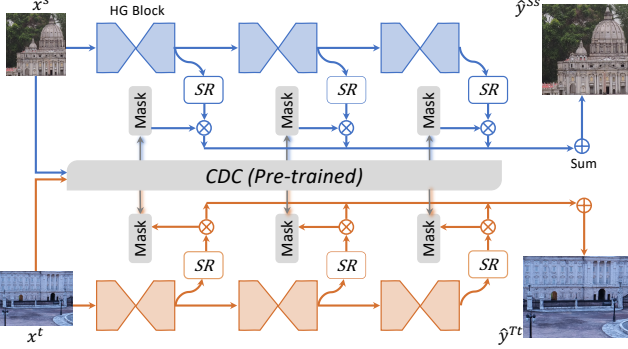
Figure 6. Domain-invariant attention module. For simplicity, we show its detailed network structure by taking two LR images as inputs in source and target branches, respectively.

process *Intra-domain Adversarial Adaptation* (IntraAA).

Domain-Invariant Attention Module. We denote the up-sampling modules as $\mathcal{U}_s$ for the source branch and $\mathcal{U}_t$ for the target branch. They follow the same neural network architecture as CDC [18]. With the hourglass network as the backbone, CDC builds three component-attentive blocks to address the issue of model overfitting to the simple image regions or contents for reconstruction. The three image components are the flat, edges and corner points, respectively. Considering those components are relatively invariant to the camera hardwares, it is believed that they are invariant for cross-device SR. This is regarded as the root for real to real adaptation. Accordingly, DADA utilizes a CDC model pre-trained on source pairs to provide domain-invariant attention masks (*i.e.*, component masks in CDC) for LR input images. Namely, the two up-sampling modules in each branch share the same parameter weights to generate component masks, as shown in Fig. 6. Like CDC, those masks weigh three intermediate SR results, respectively, and then sum weighted results to generate the final SR result.

Inter-domain Adversarial Adaptation. In each branch, LR images from source domain and target domain are handled by the same up-sampling module. There are GT HR images for source inputs, thus we can impose content supervision on the source SR images. The InterAA process is to align the target domain with the source domain.

In the source branch, the up-sampling module $\mathcal{U}_s$ forwards a source LR image x_i^s and a target LR image x_j^t, respectively, and produces their corresponding super-resolution results $\hat{y}_i^{Ss}$ and $\hat{y}_j^{St}$. For those two SR images, *i.e.*, the source SR result and the target SR result in the source branch, the discriminator $\mathcal{D}_s^{\text{inter}}$ is employed to distinguish which domain they come from. By imposing an adversarial loss between $\hat{y}_i^{Ss}$ and $\hat{y}_j^{St}$, we force the network to produce results close to the source domain for the target input. In this way, domain alignment is achieved in an adversarial manner, thereby improving the model's ability to

process target data. It is worth noting that since the two LR inputs from the two domains are very different, *e.g.*, content and color, which increases the difficulty of adversarial training between them, we let the discriminator distinguish their SR images on the Y channel. This can avoid the bias to the image content or style.

Symmetrically, in the target branch, both the source LR image and the target LR image are up-sampled by $\mathcal{U}_t$, generating two SR results $\hat{y}^{Ts}$ and $\hat{y}^{Tt}$, which are then discriminated by the discriminator $\mathcal{D}_t^{\text{inter}}$. There are two differences here from the source branch: 1) to avoid strong and dominant supervision from the source domain for $\mathcal{U}_t$ in the target branch, there is no supervision for the source SR $\hat{y}_i^{Ts}$. 2) Since source supervision is not used in the target branch, to stabilize the training of $\mathcal{U}_t$, we impose supervision on target SR images. Due to the lack of target HR images, we use a pseudo-label for the target SR $\hat{y}_j^{Tt}$. Specifically, we use the SR result $\hat{y}_j^{St}$ of the target input in the source branch as the label for $\hat{y}_j^{Tt}$.

Thus, in the source branch, the InterAA adversarial loss for the generator $\mathcal{U}_s$ and the discriminator $\mathcal{D}_s^{\text{inter}}$ is defined respectively as follows,

$$\mathcal{L}_{\text{inter}}^{\mathcal{S},\mathcal{G}} = \mathbb{E}_{x_i^s}[log(1 - \mathcal{D}_s^{\text{inter}}(\mathcal{U}_s(x_i^s)))], \qquad (1)$$

$$\begin{aligned}\mathcal{L}_{\text{inter}}^{\mathcal{S},\mathcal{D}} = &\mathbb{E}_{x_j^t}[log(1 - \mathcal{D}_s^{\text{inter}}(\mathcal{U}_s(x_j^t)))]+ \\ &\mathbb{E}_{x_i^s}[log(\mathcal{D}_s^{\text{inter}}(\mathcal{U}_s(x_i^s)))].\end{aligned} \qquad (2)$$

Symmetrically, in the target branch, the adversarial loss of InterAA for the generator $\mathcal{U}_t$ and the discriminator $\mathcal{D}_t^{\text{inter}}$ is respectively defined as follows,

$$\mathcal{L}_{\text{inter}}^{\mathcal{T},\mathcal{G}} = \mathbb{E}_{x_j^t}[log(1 - \mathcal{D}_t^{\text{inter}}(\mathcal{U}_t(x_j^t)))], \qquad (3)$$

$$\begin{aligned}\mathcal{L}_{\text{inter}}^{\mathcal{T},\mathcal{D}} = &\mathbb{E}_{x_i^s}[log(1 - \mathcal{D}_t^{\text{inter}}(\mathcal{U}_t(x_i^s)))]+ \\ &\mathbb{E}_{x_j^t}[log(\mathcal{D}_t^{\text{inter}}(\mathcal{U}_t(x_j^t)))].\end{aligned} \qquad (4)$$

Intra-domain Adversarial Adaptation. InterAA adapts the model in the source or target branch with different LR images from two domains as inputs. Oppositely, DADA utilizes intra-domain adversarial adaptation between source and target branches by taking the same LR images as inputs. Specifically, taking the source LR image x_i^s as an input of two branches, i.e., the two branches employ $\mathcal{U}_s$ and $\mathcal{U}_t$ to super-resolve it into SR images, $\hat{y}_i^{Ss}$ and $\hat{y}_i^{Ts}$, respectively. The discriminator $\mathcal{D}_s^{\text{intra}}$ is used to identify which branch $\hat{y}_i^{Ss}$ and $\hat{y}_i^{Ts}$ are generated from. Similarly, the target LR image will be sent to the two up-sampling modules in the two branches to obtain two SR counterparts, and then be distinguished by the discriminator $\mathcal{D}_t^{\text{intra}}$. Namely, the two up-sampling modules, $\mathcal{U}_s$ and $\mathcal{U}_t$ cooperatively fool $\mathcal{D}_t^{\text{intra}}$. By adopting IntraAA, we force the two up-sampling modules in the two branches to produce results that are close to each other, even though they are under different supervision, namely the source GT HR supervision and the target

pseudo SR supervision. In this way, we impose indirect and adversarial supervision on the target branch.

Accordingly, for a source LR image, the adversarial loss of IntraAA for the generator $\mathcal{U}_s$ and the discriminator $\mathcal{D}_s^{\text{intra}}$ is respectively defined as

$$\mathcal{L}_{\text{intra}}^{\mathcal{S},\mathcal{G}} = \mathbb{E}_{x_i^s}[log(1 - \mathcal{D}_s^{\text{intra}}(\mathcal{U}_s(x_i^s)))], \tag{5}$$

$$\mathcal{L}_{\text{intra}}^{\mathcal{S},\mathcal{D}} = \mathbb{E}_{x_i^s}[log(1 - \mathcal{D}_s^{\text{intra}}(\mathcal{U}_t(x_i^s)))] + \\ \mathbb{E}_{x_i^s}[log(\mathcal{D}_s^{\text{intra}}(\mathcal{U}_s(x_i^s)))]. \tag{6}$$

Similarly, for a target LR image, the adversarial loss of IntraAA for the generator $\mathcal{U}_t$ and the discriminator $\mathcal{D}_t^{\text{intra}}$ is respectively defined as

$$\mathcal{L}_{\text{intra}}^{\mathcal{T},\mathcal{G}} = \mathbb{E}_{x_j^t}[log(1 - \mathcal{D}_t^{\text{intra}}(\mathcal{U}_t(x_j^t)))], \tag{7}$$

$$\mathcal{L}_{\text{intra}}^{\mathcal{T},\mathcal{D}} = \mathbb{E}_{x_j^t}[log(1 - \mathcal{D}_t^{\text{intra}}(\mathcal{U}_s(x_j^t)))] + \\ \mathbb{E}_{x_j^t}[log(\mathcal{D}_t^{\text{intra}}(\mathcal{U}_t(x_j^t)))]. \tag{8}$$

3.3. Training Objective

Our training objective loss function $\mathcal{L}$ includes the reconstruction loss $\mathcal{L}_{\text{rec}}$, source content loss $\mathcal{L}_{con}^{\mathcal{S}}$, target content loss $\mathcal{L}_{con}^{\mathcal{T}}$, target VGG loss $\mathcal{L}_{\text{vgg}}^{\mathcal{T}}$ and the aforementioned adversarial loss of InterAA and IntraAA.

$$\mathcal{L} = \mathcal{L}_{con}^{\mathcal{S}} + \mathcal{L}_{con}^{\mathcal{T}} + \alpha\mathcal{L}_{\text{rec}} + \beta\mathcal{L}_{\text{vgg}}^{\mathcal{T}} + \lambda_1\mathcal{L}_{\text{inter}}^{\mathcal{S},\mathcal{G}} + \\ \lambda_2\mathcal{L}_{\text{inter}}^{\mathcal{T},\mathcal{G}} + \lambda_3\mathcal{L}_{\text{intra}}^{\mathcal{S},\mathcal{G}} + \lambda_4\mathcal{L}_{\text{intra}}^{\mathcal{T},\mathcal{G}}, \tag{9}$$

where hyper-parameters α, β and $\lambda_1 \sim \lambda_4$ are weight scalars. **Reconstruction loss $\mathcal{L}_{rec}$:** Each branch follows a cycle reconstruction framework ($LR{\to}HR{\to}LR$). For each input, we compute the L_1 reconstruction loss in each branch. **Content loss $\mathcal{L}_{con}$:** The Gradient Weighted (GW) loss [18] is utilized to compute the pixel-wise content loss. In source branch, $\mathcal{L}_{con}^{\mathcal{S}} = \mathcal{L}_{GW}(\hat{y}_i^{Ss}, y_i^s)$. In target branch, $\mathcal{L}_{con}^{\mathcal{T}} = \mathcal{L}_{GW}(\hat{y}_j^{Tt}, \hat{y}_j^{St})$. **VGG loss $\mathcal{L}_{\text{vgg}}^{\mathcal{T}}$:** In the target branch, to alleviate the negative effects brought by pseudo HRs, VGG loss is used to constrain the target up-sampling module $\mathcal{U}_t$. VGG-19 [13] is the feature extractor ϕ whose $conv5\_3$ features are used. $\mathcal{L}_{\text{vgg}}^{\mathcal{T}} = \mathbb{E}_{x_j^t}[\|\phi(\hat{y}_j^{Tt}) - \phi(\bar{y}_j^t)\|_1]$, where $\bar{y}_j^t = \mathcal{U}_{s_0}(x_j^t)$, where $\mathcal{U}_{s_0}$ is the CDC model pretrained on source data.

4. Experiments

4.1. Experimental Settings

Dataset. Our experiments have been conducted on DRealSR dataset [18] for UDA of real-world SR across cameras. DRealSR is the only real-world SR dataset that involves multiple cameras and has distinct indications about which camera each image was captured by. It is collected by five DSLR cameras, *i.e.*, Panasonic, Sony, Olympus, Nikon and Canon. In our experiments, image pairs from three cameras (Panasonic, Sony and Olympus) are chosen for training and testing. For each camera, they are split into a training set and a testing set, respectively. There are 197 image pairs of Panasonic, 145 pairs of Sony and 190 pairs of Olympus for training, and 20, 17 and 19 corresponding pairs for testing. Similar to [18], in the training stage, LR images are cropped into patches with the size of $48 * 48$. Our experiments are conducted for $\times 4$ scaling factor.

Implementation Details. We apply the Adam optimizer to train our model. The learning rate is $1e\text{-}4$. DADA takes a pair of a source image and a target image as an input, where source and target images are selected randomly. In each iterative step during training, four pairs are selected randomly as a batch. Data augmentations includes random crop, random rotation and flip. All the discriminators are PatchGAN [25]. The down-sampling module includes eight Residual blocks and two strided convolution layers [19]. For inference, the $\mathcal{U}_t$ network is employed to produce SR results. $\lambda_1 \sim \lambda_4$ are all set to 0.005 and $\alpha = 0.1, \beta = 0.01$. Three commonly used metrics are adopted for evaluation, *i.e.*, Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM) [17] and LIPIS [23]. Following the similar setting to CDC [18], PSNR is computed on the Y channel and SSIM is on RGB images.

4.2. Comparison with State-of-the-Art Methods

We are the first to explore the unsupervised domain adaptation problem between different devices w.r.t the super-resolution task. We compare our method with existing *Synthetic* to *Real* UDA methods for SR, including Cycle-in-Cycle Generative Adversarial Networks (CinCGAN) [20], Domain-distance Aware Super-Resolution (DASR) [19] and Dual Regression Adaptation Network (DRN-Adapt) [5]. For fair experimental comparisons, we implement them under the *Real* to *Real* adaptation setting by replacing the bicubic images with real LR images in the source domain. The super-resolution network of CinCGAN is also the CDC model [18]. Their comparison results are provided in Table 1. In this table, *Source Only* is the model trained with paired source data without model adaptation. *Target Only* means the model trained with real paired data in the target domain.

Real→Real Adaptation. We conduct experimental comparisons under six settings of *Real* to *Real* adaptation among three cameras. Compared with the baseline *Source Only* model, our DADA achieves significant performance gains. For example, in the *Panasonic→Sony* task, it improves the *Source Only* model from 31.36dB to 32.13dB (PSNR). In comparison with the state-of-the-art UDA methods, our method presents a superior performance and achieves best PSNR and SSIM on most of the six adaptation tasks. For instance, in the *Sony→Olympus* adaptation task, our method outperforms CinCGAN by 0.91dB (PSNR), DASR by 1.22dB, and DRN-Adapt by 0.42dB, respectively. Notably, in terms of the LPIPS metric, DASR achieves a

Method	Panasonic → Sony			Sony → Panasonic			Olympus → Panasonic		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
Real→Real									
Target Only	32.72	0.854	0.302	32.39	0.846	0.316	32.39	0.846	0.316
Source Only	31.36	0.838	0.319	30.65	0.820	0.383	30.42	0.818	0.372
CinCGAN [20]	27.76	0.821	0.391	28.33	0.792	0.410	29.37	0.799	0.381
DASR [19]	30.08	0.777	**0.269**	30.45	0.772	**0.316**	30.06	0.785	**0.272**
DRN-Adapt [5]	31.85	0.845	0.321	30.96	0.821	0.380	30.80	0.822	0.356
DADA (Ours)	**32.13**	**0.849**	0.327	**31.25**	**0.825**	0.363	**31.27**	**0.824**	0.348
Synthetic → Real									
Source Only	31.39	0.829	0.369	30.43	0.807	0.433	30.42	0.808	0.437
CinCGAN [20]	27.59	0.788	0.405	27.19	0.743	0.414	28.38	0.739	0.422
DASR [19]	29.95	0.764	**0.298**	29.79	0.749	**0.339**	30.02	0.777	**0.293**
DRN-Adapt [5]	31.42	0.829	0.359	30.47	0.808	0.429	30.45	0.808	0.433
DADA (Ours)	**31.50**	**0.830**	0.369	**30.72**	**0.809**	0.376	**30.74**	**0.808**	0.362

Method	Panasonic → Olympus			Sony → Olympus			Olympus → Sony		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
Real→Real									
Target Only	31.77	0.833	0.375	31.77	0.833	0.375	32.72	0.854	0.302
Source Only	30.49	0.816	0.439	30.55	0.810	0.457	30.49	0.814	0.330
CinCGAN [20]	28.85	0.791	0.461	30.17	0.814	0.443	30.05	0.823	0.365
DASR [19]	29.32	0.768	**0.306**	29.86	0.762	**0.372**	30.29	0.787	**0.270**
DRN-Adapt [5]	30.73	0.816	0.431	30.66	0.810	0.459	31.47	0.833	0.312
DADA (Ours)	**31.08**	**0.820**	0.433	**31.08**	**0.817**	0.438	**32.05**	**0.843**	0.343
Synthetic → Real									
Source Only	30.08	0.799	0.479	30.08	0.799	0.472	31.41	0.828	0.371
CinCGAN [20]	28.43	0.766	0.407	29.34	0.767	0.451	29.50	0.792	0.392
DASR [19]	28.30	0.752	**0.375**	29.51	0.755	**0.402**	29.40	0.737	**0.327**
DRN-Adapt [5]	30.11	0.799	0.475	30.11	0.799	0.473	31.45	0.829	0.362
DADA (Ours)	**30.40**	**0.800**	0.403	**30.62**	**0.803**	0.411	**31.52**	**0.829**	0.355

Table 1. Performance evaluation for UDA real SR across devices. (*Target Only* model are trained with GTs. Except *Target Only*, the highest performance among UDA SR methods are highlighted in bold.)

Figure 7. Comparison of UDA real SR results with state-of-the-art methods for *Real* (Sony) to *Real* (Panasonic) adaptation.

Method	InterAA	IntraAA	DIA	PSNR	SSIM	LIPIS
DADA		✓	✓	30.76	0.812	0.472
	✓		✓	30.97	0.815	0.438
	✓	✓		31.03	0.816	0.445
	✓	✓	✓	**31.08**	**0.817**	**0.438**

Table 2. Ablation study.

rather high performance among all the methods including our DADA, however, its evaluations in terms of PSNR and SSIM are inferior in comparison with DRN-Adapt and our DADA. Though it indeed produces perceptually-clear SR images, it is of obvious noises and artifacts. Particularly, Our DADA significantly outperforms DASR with PSNR gains of 0.80 dB (at least) and 2.05 dB (at most), and SSIM gains of 0.04 (at least) and 0.07 (at most). In essence, the main reason to explain this phenomenon is that DASR employs adversarial training of the predicted SR images against the real-world HR images while DADA utilizes adversarial training between two predicted SR images. Thus, the former enforces DASR to produce SR images as similar (to HRs) as possible, but inevitably invites obvious noises and artifacts. Besides, qualitatively, Fig.7 shows the visual comparison of different methods in the *Sony → Panasonic* task. It is observed that our DADA yields the visually better SR image than other comparison methods. For instance,

Figure 8. SR results of video images on the REDS dataset. For *Real→Real* adaptation across devices, only qualitative visualizations of SR results are provided. Because only HR video images are realistic data, we take those HR video images in REDS as the target LR images to verify the proposed model.

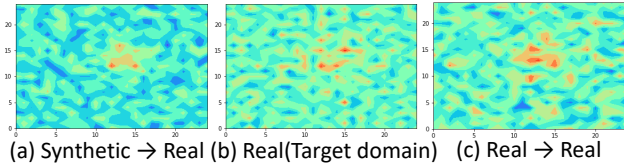(a) Synthetic → Real (b) Real(Target domain) (c) Real → Real

Figure 9. Comparison of degradation kernels for generated SR images under different adaptations from *Panasonic* to *Sony*.

the predicted SR image by *Source Only* is blurred due to the device gap. The result of CinCGAN is sharp, but the color is distinctly different from the ground-truth HR image. Instead, our DADA produces the clear result more close to the *Target Only*.

Synthetic→Real Adaptation. We also provide the evaluation results of *Synthetic* to *Real* adaptation in Table 1. We use the synthetic pairs in the source dataset, where LR images are obtained by bicubic downsampling HR images. It is observed that our DADA still achieves superior performance under six adaptation settings. Moreover, in general, all the models have significantly high performance under *Real* to *Real* adaptation than those of under *Synthetic* to *Real* adaptation. This demonstrates that the adaptation from *Real* to *Real* can achieve higher image quality than that from *Synthetic* to *Real*.

4.3. Model Evaluation and Analysis

Ablation study. We conduct the model ablation studies in the *Sony→Olympus* task, as shown in Table 2. (1) *DIA:* In the DIA module, for an input image, we share the attention mask on the source branch and the target branch. Without DIA, we train the mask generator separately in each branch, and the result drops from 31.08dB to 31.03dB (PSNR). (2) *InterAA:* InterAA employs an SR model as the generator for source LR image and target LR image. It improves the model from 30.76dB to 31.08dB (PSNR). (3) *IntraAA:* IntraAA brings 0.11dB improvements.

Adaptation kernel analysis. In Fig. 9, we employ USR-Net to show degradation kernels for generated SR images under different adaptation. Fig.9(b) is the GT real kernels for the target domain. It is demonstrated that *Synthetic* to *Real* adaptation (Fig. 9(a)) is limited by simple image degradation in the source domain and fails to fill the large domain gap cross devices. Instead, *Real* to *Real* adaptation (Fig. 9(c)) presents a favorable transferring.

Evaluation for video camera. To fully verify our method, we additionally conduct an adaptation experiment to a video

camera on REDS [10]. REDS is a video dataset proposed for video deblurring and super-resolution tasks, where each video is captured by the GoPro HERO6 Black camera. Considering its LR frames are synthetic, we only use its original video frames as LR images for training. We randomly select 22 videos for training and the rest are for testing. In our work, temporal relationships of videos are not considered and *Sony* in DRealSR is considered as the source domain. In Fig. 8, we provide qualitative comparison results of *Sony→REDS*. It is observed that CinCGAN tends to produce SR results with wrong color and serious artifacts, and DASR also has chromatic aberration and distortion problems. Our DADA has a clear and more natural SR result. This presents a promising research interest, which would promote the practical applications of real-world SR to broad scenarios, *e.g.*, video enhancement for handheld devices.

5. Conclusion

In this paper, we propose exploring the cross-device domain gap in the real-world super-resolution, facilitating an adaptation from a source domain with paired real LR-HR data to a target domain with only LR images. To mitigate this issue, a Dual ADversarial Adaptation (DADA) model is proposed. It leverages the stability of extracting mid-level image components from images, to build a domain-invariant attention module, rather than learning domain-invariant features. Besides, considering the inaccessibility of HR images in target domains, our DADA employs a training strategy with inter-domain and intra-domain adversarial adaptation in a dual architecture. Extensive experiments are conducted under six *Real→Real* adaptation settings among three different cameras in the DRealSR dataset, demonstrating that the proposed DADA achieves superior performance compared with existing state-of-the-art approaches. Additionally, we also evaluate the proposed DADA to address the adaptation to a video camera, which presents a promising research topic to promote the wide applications of real-world super-resolution.

Broader impacts and limitations. All the experiments are conducted on DRealSR, the only one dataset including multiple cameras with detailed device information, for cross-device real SR. The task might be more complex due to certain unknown issues when considering more different devices for image degradation. This task and the proposed method might need more analyses and evaluations.

References

[1] Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3086–3095, 2019. 2, 3

[2] Yuhua Chen, Wen Li, Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Domain adaptive faster r-cnn for object detection in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3339–3348, 2018. 3

[3] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *European conference on computer vision*, pages 184–199. Springer, 2014. 2, 3

[4] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *The journal of machine learning research*, 17(1):2096–2030, 2016. 3

[5] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5407–5416, 2020. 3, 6, 7

[6] Jiaxing Huang, Shijian Lu, Dayan Guan, and Xiaobing Zhang. Contextual-relation consistent domain adaptation for semantic segmentation. In *European conference on computer vision*, pages 705–722. Springer, 2020. 3

[7] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017. 2, 3

[8] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017. 2, 3

[9] Yawei Luo, Liang Zheng, Tao Guan, Junqing Yu, and Yi Yang. Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2507–2516, 2019. 3

[10] Seungjun Nah, Sungyong Baik, Seokil Hong, Gyeongsik Moon, Sanghyun Son, Radu Timofte, and Kyoung Mu Lee. Ntire 2019 challenge on video deblurring and super-resolution: Dataset and study. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019. 8

[11] Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, and Tatsuya Harada. Maximum classifier discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3723–3732, 2018. 3

[12] 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 *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883, 2016. 3

[13] K. Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2015. 6

[14] Tong Tong, Gen Li, Xiejie Liu, and Qinquan Gao. Image super-resolution using dense skip connections. In *Proceedings of the IEEE international conference on computer vision*, pages 4799–4807, 2017. 3

[15] Yi-Hsuan Tsai, Wei-Chih Hung, Samuel Schulter, Kihyuk Sohn, Ming-Hsuan Yang, and Manmohan Chandraker. Learning to adapt structured output space for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7472–7481, 2018. 3

[16] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision workshops*, 2018. 2, 3

[17] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 6

[18] 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 *European Conference on Computer Vision*, pages 101–117. Springer, 2020. 1, 2, 4, 5, 6

[19] Yunxuan Wei, Shuhang Gu, Yawei Li, Radu Timofte, Longcun Jin, and Hengjie Song. Unsupervised real-world image super resolution via domain-distance aware training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13385–13394, 2021. 3, 6, 7

[20] 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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 701–710, 2018. 3, 6, 7

[21] Jiutao Yue, Haofeng Li, Pengxu Wei, Guanbin Li, and Liang Lin. Robust real-world image super-resolution against adversarial attacks. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 5148–5157, 2021. 2, 3

[22] Kai Zhang, Luc Van Gool, and Radu Timofte. Deep unfolding network for image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3217–3226, 2020. 1

[23] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the*

IEEE conference on computer vision and pattern recognition, pages 586–595, 2018. 6

[24] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision*, pages 286–301, 2018. 2, 3

[25] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017. 3, 6

[26] Xinge Zhu, Jiangmiao Pang, Ceyuan Yang, Jianping Shi, and Dahua Lin. Adapting object detectors via selective cross-domain alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 687–696, 2019. 3