Real-World Super-Resolution using Generative Adversarial Networks

Haoyu Ren*, Amin Kheradmand*, Mostafa El-Khamy, Shuangquan Wang, Dongwoon Bai, Jungwon Lee
SOC R&D, Samsung Semiconductor, Inc.
9868 Scranton Road, San Diego, CA, USA
{haoyu.ren, a.kheradmand, mostafa.e, shuangquan.w, dongwoon.bai, jungwon2.lee}@samsung.com

Abstract

Robust real-world super-resolution (SR) aims to generate perception-oriented high-resolution (HR) images from the corresponding low-resolution (LR) ones, without access to the paired LR-HR ground-truth. In this paper, we investigate how to advance the state of the art in real-world SR. Our method involves deploying an ensemble of generative adversarial networks (GANs) for robust real-world SR. The ensemble deploys different GANs trained with different adversarial objectives. Due to the lack of knowledge about the ground-truth blur and noise models, we design a generic training set with the LR images generated by various degradation models from a set of HR images. We achieve good perceptual quality by super resolving the LR images whose degradation was caused by unknown image processing artifacts. For real-world SR on images captured by mobile devices, the GANs are trained by weak supervision of a mobile SR training set having LR-HR image pairs, which we construct from the DPED dataset which provides registered mobile-DSLR images at the same scale. Our ensemble of GANs uses cues from the image luminance and adjusts to generate better HR images at low-illumination. Experiments on the NTIRE 2020 real-world super-resolution dataset show that our proposed SR approach achieves good perceptual quality.

1. Introduction

Image super-resolution (SR) generates a high-resolution (HR) image from a given low-resolution (LR) image by attempting to recover the missing information. SR methods have already been deployed in many computer vision applications such as surveillance, face and iris recognition, and medical image processing. Recently, deep convolutional neural networks (CNNs) have been deployed to solve the image super-resolution problem, as they demonstrate significant accuracy improvements.

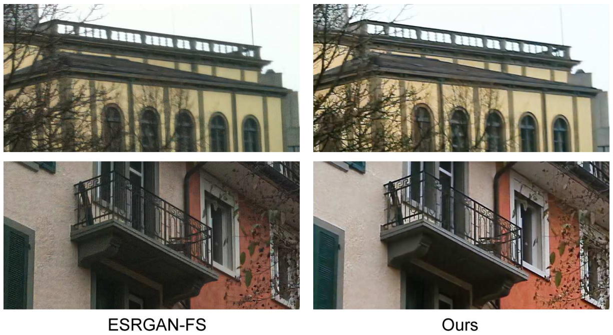

Figure 1. Example x4 SR outputs on the NTIRE real-world SR challenge track 2 testing images.

Due to the lack of real-world LR-HR patches, in most of the prior arts, images are bicubic-downsampled to create the LR-HR training pairs. This results in clean and noise free LR images. Unfortunately, in real-world scenario where the images come straight from the camera, there will always be additional noise or unknown degradations [18]. As a result, the state-of-the-art CNN methods trained only to reconstruct images artificially downsampled using bicubic downsampling may lead to dramatic artifacts when applied to real-world images. To solve this problem, some researchers capture images by using different focal lengths of digital single-lens reflex (DSLR) cameras [3], and further align them by some registration algorithms. However, the DSLR imaging system is still different from the mobile imaging system which are commonly-used in real-world. In addition, the alignment of images captured between different scales is relatively difficult.

In this paper, we study the real-world super-resolution problem from the following aspects. First, we create a generic training set by adopting multiple degradations on the high-quality HR images based on reasonable assumptions of the image processing artifacts, such as different downsampling methods, different blur kernels, and different noises. We show that Generative Adversarial Network (GAN) previously shown to work on paired SR datasets, can perform well and achieve good generalization ability

*The first two authors have equal contributions.

on the images without any knowledge of the exact degradation model after training them on our new dataset. Second, we propose a weakly supervised method to train the GANs for super resolving real-world mobile images. Without any knowledge about the mobile HR-LR degradation model, we create a mobile SR dataset by generating paired LR-HR images from the registered mobile-DSLR images at the same scale provided by the DPED dataset [11]. We use the mobile images as LR, and apply our SR model trained by the generic training set on the paired DSLR images to create super resolved HR images with good perceptual quality. After fine-tuning our SR-GAN model on these LR-HR pairs, we observe clear performance improvement when testing on mobile images, as given in Fig. 1. When testing on the NTIRE 2020 real-world SR challenge track 2 images, our SR method achieves clearly better perceptual quality compared to the state-of-the-art real-world SR method ESRGAN-FS [8] trained on the DPED dataset as well. Third, we investigate the fusion based on different GAN networks to improve the overall perceptual quality of the resulting SR images.

The contributions of this paper are highlighted as follows:

- Our generic SR model trained on the SR dataset generated by multiple degradations generalizes well on the images with unknown degradation caused by image processing artifacts.
- We design a mobile SR dataset based on registered mobile-DSLR image pairs at same scale, where the DSLR images are super resolved with our generic SR model. Fine-tuning our SR model on this dataset improves the perceptual quality on mobile images.
- Our GAN-based fusion is capable of improving the perceptual quality and reducing the artifacts of the estimated HR images.

2. Related Work

2.1. PSNR-oriented super-resolution

During the past years, lots of research addressed the image super-resolution problem. Most of them are still based on synthetic bicubic downsampling degradation and benchmarked in terms of PSNR. Some early SR algorithms refer to filtering approaches, such as bilinear, bicubic, and Lanczos filtering [7]. These filtering algorithms may generate smooth outputs without recovering any high-frequency information. They are computationally efficient, but the accuracy is limited because they oversimplify the SR problem. Other approaches assume a kind of mapping between the LR space and the HR space. Such mapping could be learned from a large number of LR-HR pairs by the sparse coding-based image representation [30][28], where a sparse coefficient vector is shared between both the LR space and the HR space. Recently, CNNs have been widely used for image super-resolution. In [6], Dong et al. designed a 3-layer CNN learning a mapping between the bicubic upsampled LR space and the corresponding HR space. To further improve the accuracy, more complicated networks are proposed. Kim et al. [12] increased the layer number to 20 and used small filters and a high learning rate with adjustable gradient clipping. Kim et al. [13] proposed to use a deep recursive network with a skip connection, where the same weights are shared by multiple convolutional layers. Tai. et al. [26] further integrated the recursive structure into residual network. Dahl et al. [5] combined the ResNet with a pixel recursive super-resolution, which showed promising results on face and bed SR. Lim et al. [16] designed an enhanced deep super-resolution network (EDSR) by removing unnecessary modules in conventional residual networks and expanding the model size. Tai et al. [27] proposed a very deep persistent memory network (MemNet) that introduces a memory block, consisting of a recursive unit and a gate unit, to explicitly mine persistent memory through an adaptive learning process. Haris et al. [10] proposed deep backprojection networks that exploit iterative up- and downsampling layers, providing an error feedback mechanism for projection errors at each stage. Zhang et al. [34] proposed very deep residual channel attention networks (RCANs), which have a residual in residual (RIR) structure and a channel attention mechanism to adaptively rescale channel-wise features.

2.2. Perceptual-oriented super-resolution

PSNR is not consistent to human vision, which implies that a SR network with better PSNR can lead to poor perceptual quality [2]. To solve this problem, Ledig et al. [15] employed a very deep residual network and further presented the super-Resolution generative adversarial network (SR-GAN) to obtain HR images with textures similar to natural textures. Wang et al. [29] improved the SRGAN by introducing the Residual-in-Residual Dense Block (RRDB) without batch normalization as the basic network building unit and using relativistic GAN to let the discriminator predict relative realness. But in real-world super-resolution, due to the unknown degradation and lack of training data, the SR methods with the assumption of bicubic degradation still don't work well. Recently, several approaches have been introduced for super resolving real LR images. Zhang et al. [31] fed extra information related to the blur and noise in addition to the input LR images at both training time and test time to improve the performance of the SR model on different degradations. However, it explicitly assumes the knowledge of the degradation. In order to extend this multiple degradation approach to blind SR where the degradation in the input images is unknown, Gu et al. [9] used a predictor network for blur kernel estimation as well as a corrector network to reduce the artifacts from the output of the main SR network. Zhou et al. [35] tried to estimate the camera blur in real LR images by exploiting an existing blur kernel estimation algorithm [20]. The estimated blur kernels are then used to generate synthetic LR-HR image paris for training the SR network. Lugmayr et al. [17] started from an unpaired dataset of LR-HR image pairs and trained a CycleGAN network to transform the bicubically downsampled HR images into the LR images which represent the domain defined by real image characteristics. They used these LR-HR image pairs generated using this approach to train a GAN-based SR network. In [8], the transformation from the bicubically downsampled domain to the source LR domain was modeled using a GAN network by separating the low and high frequencies in the bicubic-downsampled image and training the discriminator of the GAN just using the high-frequency component for improving the GAN training. Bell et al. [1] proposed the blind SR algorithm using an unsupervised approach for estimating the SR kernel from the input LR image. The estimated SR kernel can then be plugged into existing SR algorithms for high-quality results for blind SR.

3. Proposed Method for Real-World SR

Our method for real-world super-resolution consists of 3 steps. The first step is dataset generation for real-world SR. The second step is training different Generative Adversarial Networks (GANs) for generation of the super resolved images. More specifically, we follow the Enhance Super-resolution Generative Adverserial Networks (ESRGAN) [29] framework, and use the Residual Channel Attention Network (RCAN) [34] as the generator instead of Residual-in-Residual Dense Block (RRDB). During the testing, the generator (RCAN) is directly used to estimate the HR image from a given LR image. To further improve the perceptual quality, we train two RCANs by using different discriminators and hyper-parameters in ESRGAN. This grants us two SR networks with complementary characteristics. The third step is to design an ensemble strategy for the fusion of the results from the trained SR-GANs. Our final SR prediction is a pixel-wise ensemble of these two RCANs, as given in Fig. 2.

3.1. Dataset construction for real-world SR

We generate two different datasets, the first dataset is for generic real-world SR with no assumption on the source of the LR images. The second dataset is for real-world SR when it is known that the LR images are captured by a mobile phone camera.

3.1.1 Synthetic dataset generation for robust real-world SR

To train an SR network that has robust performance without prior knowledge about the HR-LR degradation model, we construct an SR dataset based on multiple degradations from the image processing artifacts. More specifically, we generate an LR image x from an HR image y following the generic degradation model formulated as

$$x = \mathcal{N}\left(D(y * k)\right), \tag{1}$$

where D is the downsampling operation, k is a blur kernel, $*$ denotes the convolution, and $\mathcal{N}(t)$ applies noise to the input t, where the noise model is not necessarily additive.

Downsampling We consider multiple downsampling methods, including nearest neighbor, bilinear, bicubic, and Lanczos. The downsampling method is randomly selected when generating the LR patches.

Blur kernel Different from image deblurring, the blur kernel setting of SR is usually simple. We use the most commonly-used isotropic Gaussian blur kernel parameterized by the standard deviation. In our implementation, we randomly sample the standard deviation of Gaussian kernel in the range of [0.2, 3], and fix the kernel size as 15×15.

Noise Most of the real-world LR images are noisy due to some image processing artifacts. It is shown that some real-world noises consist of Gaussian, Poisson, or Poisson-Gaussian components in [22]. So we randomly pick the Gaussian, Poisson, and Poisson-Gaussian noises when generating the LR images. The parameters are based on reasonable assumption of real-world image processing artifacts, where the sigma of Gaussian noise is randomly selected from the range [0, 25], the peak of Poisson noise is uniformly sampled from the range [50, 150]. When generating the Poisson-Gaussian noise, we follow the procedure in [22]. More specifically, we generate the Poisson-Gaussian noise as

$$N(t) = \alpha p + n, \tag{2}$$

where n is i.i.d Gaussian noise, $p \sim \mathcal{P}(t)$ is a Poisson random variable with mean t. We use a similar Poisson peak range as [50, 150], but reduce the Gaussian sigma range to [0, 5]. The constant α is set to 1.

Some examples of the LR images generated by the above approach are given in Fig. 3, which demonstrates the diversity of the degradations in the constructed SR training set.

3.1.2 Synthetic dataset generation for mobile real-world SR

In real-world super-resolution, a high-quality dataset of LR-HR images which matches the domain of target imaging device is crucial for the performance of the SR net-

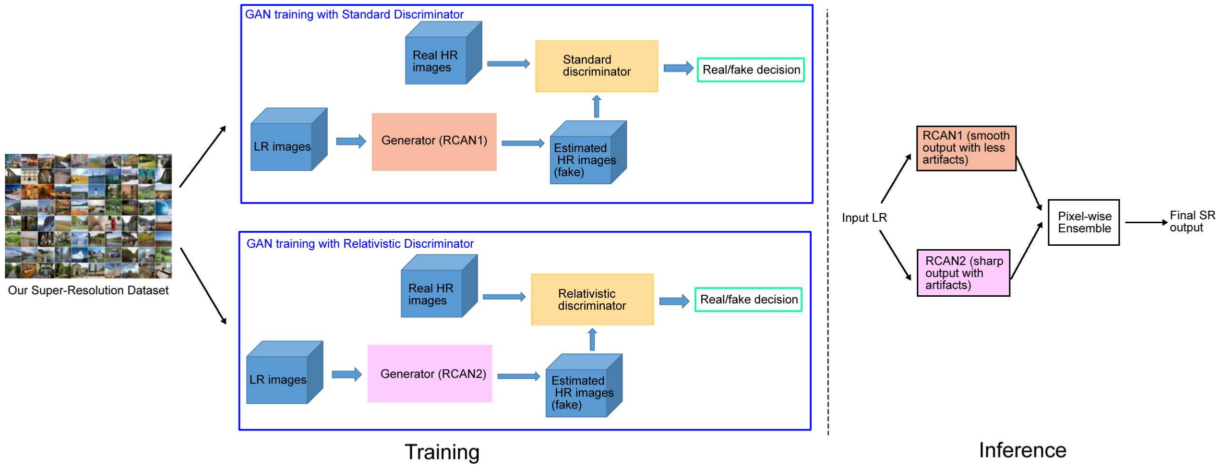
Our Super-Resolution Dataset
GAN training with Standard Discriminator
Real HR images
Standard discriminator
Real/fake decision
LR images
Generator (RCAN1)
Estimated HR images (fake)
GAN training with Relativistic Discriminator
Real HR images
Relativistic discriminator
Real/fake decision
LR images
Generator (RCAN2)
Estimated HR images (fake)
Training
Input LR
RCAN1 (smooth output with less artifacts)
RCAN2 (sharp output with artifacts)
Pixel-wise Ensemble
Final SR output
Inference

Figure 2. Overview of our SR approach.

works. There have been several efforts focusing on creating datasets for real-world SR. In [3], the data is captured using two DSLR camera models from Nikon and Canon. The authors captured images with different focal lengths to derive the HR image and the corresponding LR images with different zoom factors. The resulting LR-HR images are then registered to create the final paired SR dataset [3]. However, these DSLR images don't generalize well on mobile images commonly-used in the senerio of real-world super-resolution. In [4], the SR datasets are captured by mobile phone and DSLR separately. But due to the difficulty of the registration where the images come from both different scales and different domains, these mobile images are not aligned with DSLR images.

In order to alleviate these shortcomings of the existing methods, we propose an effective methodology for creating high-quality mobile SR dataset. It is motivated by the fact that it is much easier to register images from different domains captured at the same scale, instead of different scales. More specifically, we propose to create a new mobile SR dataset using registered image pairs at same scale, without the knowledge of the exact HR-LR degradation model. For this purpose, we use the DSLR Photo Enhancement Dataset (DPED) [11] which provides registered mobile-DSLR patch pairs at same scale for image enhancement. We create our mobile SR dataset by super resolving the DSLR patches to create the HR images. The LR images in our proposed dataset are the corresponding mobile patches. We observe that even applying a simple bicubic upsampling algorithm to super resolve the DSLR patches can generate high-quality HR patches for training the mobile SR networks. We further use our SR model trained on the generic dataset described in Section 3.1.1 to super resolve these DSLR patches. This gives us a clear improvement in visual quality after training the SR network as compared to the data generation based on the bicubic upsampling. More details can be found in Section 4.3.2.

Figure 3. Example LR images of our generic SR training set.

We show that fine-tuning a GAN network using the proposed mobile SR dataset results in high-quality outputs when applied to super resolve real mobile LR images. Our method actually follows a kind of weakly supervised way, so that the perceptual quality of the SR output is much better than prior art [8] without the usage of paired data. Our proposed technique also makes the practical application of super resolving mobile images for digital zoom more feasible.

3.2. Super-resolution GANs

3.2.1 RCAN

The Residual Channel Attention Network (RCAN) is proposed in [34]. RCAN is based on a residual in residual (RIR) structure, which consists of several residual groups with long skip connections. Each residual group contains some residual blocks (ResBlock) with short skip connections. In each ResBlock, the channel attention mechanism is utilized to adaptively rescale channel-wise features by considering inter-dependencies among channels. We use RCAN as the generator when training our GAN networks.

3.2.2 ESRGAN

It is known that the pixel-wise PSNR-oriented SR methods usually give over-smoothed results and are not able to properly recover the high-frequency details [14]. SRGAN [14] takes advantage of the strength of the GAN networks to model the space of natural images, and uses perceptual and adversarial losses to guide the SR network to favor output images that reside on the manifold of natural images. After that, several modifications related to the perceptual-driven, GAN-based approach in SRGAN have been introduced [29, 33, 25]. We follow the ESRGAN [29] framework due to the usage of relativistic discriminator, which is able to create sharper edges and more realistic texture details.

We use ESRGAN with RCAN generator to train our SR networks. During the training, our generator loss function L_G^R consists of the L_1 image loss, the perceptual loss L_p, and the adversarial loss L_a^R, which is similar to [29], as described in Eq. 3

$$L_G^R = L_p + \lambda L_a^R + \eta L_1, \tag{3}$$

$L_1 = \mathbb{E}_x[\|G(x) - y\|_1]$ calculates the L_1 distance between the super resolved image $G(x)$ from the RCAN generator $G(.)$ and the ground-truth HR image y. $\mathbb{E}_x[.]$ denotes the operation of taking average for all the images in the mini-batch. The perceptual loss L_p calculates a feature map distance between $G(x)$ and y with the usage of a pre-trained 19-layer VGG network. We denote the output image from the RCAN generator network G as $x_f = G(x)$ which means the fake image, and the corresponding real image as x_r. The adversarial loss L_a^R is based on the relativistic GAN discriminator [29] and is defined as

$$\begin{aligned} L_a^R = &-\mathbb{E}_{x_r}[\log(1 - D_R(x_r, x_f))] \\ &-\mathbb{E}_{x_f}[\log(D_R(x_f, x_r))], \end{aligned} \tag{4}$$

in which the discriminator D_R predicts if the real image x_r is more realistic than the fake image x_f rather than deciding if an input image is absolutely real or fake.

The hyper-parameters λ and η in Eq. 3 determine the contribution of different loss components in the final loss function. One can increase the parameter η to reduce the quantitative error of the estimate while increasing the adversarial loss weight will result in an improvement in perceptual quality of the results. Furthermore, we train another GAN with RCAN generator but a different generator loss L_G^S based on standard GAN [14]

$$L_G^S = L_p + \lambda L_a^S + \eta L_1, \tag{5}$$

where L_a^S is the adversarial loss based on standard GAN. We observe that the resulting SR estimates trained by the loss functions in Eq. 3 and Eq. 5 exhibit some complementary characteristics. We employ these different characteristics via a simple yet effective fusion strategy to improve the overall visual quality of the images, as described in the following section.

3.3. Ensemble fusion of SR-GANs

We observe that the SR outputs generated by the RCAN trained by the relativistic GAN loss in Eq. 3 show good perceptual quality in high-frequency regions. In contrast, the SR outputs generated by the RCAN trained by the standard GAN loss in Eq. 5 generate less artifacts in the smooth regions of some low-illumination images. Motivated by existing pixel-wise fusion work [24, 21, 23], we propose a fusion method using the SR estimates of two RCAN generators. We opt for a selective average technique based on the median brightness of all pixels in the image in order to improve the visual quality on low-illumination images. We denote the HR output from the RCAN model trained using relativistic GAN loss as y_{SR}^R, and the HR output from the RCAN model trained using standard GAN loss function as y_{SR}^S. The fused output image y_{SR}^{fused} is derived as

$$y_{SR}^{fused} = \begin{cases} \alpha y_{SR}^R + \beta y_{SR}^S & if \quad Y_{med} < \gamma \\ y_{SR}^R & otherwise \end{cases} \tag{6}$$

where Y_{med} is the median of the pixel intensity values of all the pixels in the Y (luminance) component of the YCbCr color space representation of y_{SR}^R. The difference between our fusion framework and the GAN-based image interpolation in [29] is that we are fusing two GAN models trained based on different adversarial losses with different complementary effects, so that we make sure that the perceptual quality of the two images used for fusion are close. This ensures that we may reduce artifacts in some regions of low-illumination images while not sacrificing the overall perceptual quality. In contrast, the image interpolation in [29] uses the PSNR-oriented and GAN-based SR estimates which can lead to a reduction in overall perceptual quality of the fused images.

4. Experimental Results

4.1. Training on NTIRE 2020 datasets

We use the datasets from NTIRE 2020 real-world super-resolution [19] (RealSR) challenge to evaluate our approaches. The NTIRE 2020 RealSR challenge aims to stimulate the research on real-world super-resolution, where the degradation is unknown and unpaired data is given. In this setting, there exists no ground-truth reference images that can be directly employed for training. Instead, the model is required to be learned from only a set of source domain images, originating from for instance a particular camera sensor. NTIRE 2020 RealSR challenge contains two tracks, which investigate different types of source domains. Both of these two tracks provide unpaired images from the source domain and the target domain. The target domain of these two tracks are a same set of high-quality clean images. In track 1, the source domain consists of images containing artifacts produced by some denoising algorithms. In contrast, the source domain of track 2 consists of images with artifacts originating from image enhancement operations of the smartphone. The degradations of these two tracks are different. The scaling factors of both tracks are x4.

We apply our SR approach on the NTIRE RealSR datasets. For track 1, we train our RCANs with ESRGAN on the generic SR training set described in Section 3.1.1. When generating our generic SR dataset, we use the 800 target domain images provided by track 1 training data, as well as additional 2,650 images from Flickr2K dataset as HR images to generate the LR images with random degradation. The degradation parameters are randomly selected for the downsample method, blur kernel, and noise. When training the RCAN with relativistic GAN, the weights of the loss functions in Eq. 3 are set as $\lambda = 0.005, \eta = 0.01$. When training the RCAN with standard GAN, the weights of the loss functions in Eq. 5 are set as $\lambda = 0.005, \eta = 0.005$. The final SR output is the ensemble of two RCANs as described in Section 3.3, where the fusion parameters are set as $\alpha = 0.6, \beta = 0.4, \gamma = 64$. These fusion parameters are tuned by having a good trade-off between the artifacts in smooth regions and the sharpness in textured regions on track 1 validation images.

For track 2, we generate our mobile SR dataset following the instructions in Section 3.1.2. We use 160K mobile patches provided by the DPED training set [11] with 'iphone' labels as the LR images, and apply our track 1 SR model on the corresponding aligned DSLR patches to generate the x4 super resolved HR patches. Our RCANs used in track 1 are fine-tuned on this mobile SR dataset as our track 2 SR model using same weights of the loss functions. The final output is still an ensemble of the fine-tuned RCANs, with same fusion parameters.

4.2. Results

For track 1, we employ the commonly-used PSNR and SSIM metrics along with the perceptual metric LPIPS [32] to quantitatively evaluate the performance of our proposed algorithm on the validation data. Table 1 summarizes the numerical comparisons of different SR approaches on the track 1 validation data. It can be seen that the networks trained on our generic SR dataset have clearly better accuracy compared to the network trained on bicubic degradation. Our SR solution achieves the lowest LPIPS, which implies that we have the best perceptual quality, as illustrated in Fig. 4. The standard ESRGAN [29] based on bicubic downsample still has a lot of noises when testing on the track 1 images with unknown degradation. Comparing to training a PSNR-oriented RCAN on same generic SR training set, we also achieve better perceptual quality although the PSNR and SSIM are lower. This is consistent to the fact that perceptual-oriented GAN is more suitable for real-world super-resolution.

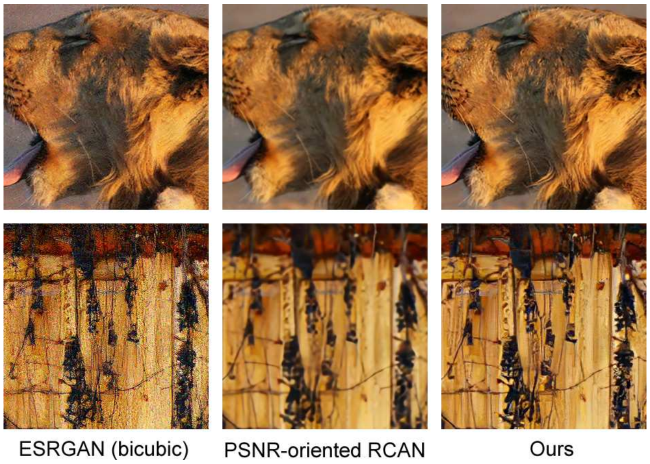

Figure 4. Qualitative comparisons of different SR algorithms on the testing images of NTIRE 2020 RealSR challenge track 1, with scaling factor x4.

Table 1. PSNR (dB)/SSIM/LPIPS evaluation of different SR methods on the validation data of NTIRE 2020 RealSR challenge track 1, with scaling factor x4. The lower LPIPS, the better.

Method	Training	PSNR(dB)/SSIM/LPIPS
ESRGAN [29]	Bicubic SR	19.06/0.2424/0.7552
PSNR-oriented RCAN	Generic SR	**27.36/0.7620**/0.3680
Ours	Generic SR	25.78/0.7119/ **0.2482**

For track 2, since we do not have access to the ground-truth images, we give qualitative analysis to verify the effectiveness of our approach. Fig. 5 compares our proposed algorithm with some of the existing SR approaches when applying on the testing images of track 2. We find that our proposed algorithm is capable of producing much higher quality images as compared to the standard ESRGAN [29] trained on bicubically downsampled images, as well as the state-of-the-art real-world SR algorithm ESRGAN-FS [8] which is also trained on DPED dataset. ESRGAN-FS [8] considers the DPED mobile images as HR, and uses a downsampling GAN to generate the LR images. An ESRGAN with frequency separation (ESRGAN-FS) is further trained using these LR images. In contrast, we use the mobile-DSLR image pairs at the same scale, without the knowledge of the exact HR-LR degradation. So our solution is a kind of weakly supervised approach. The perfor-

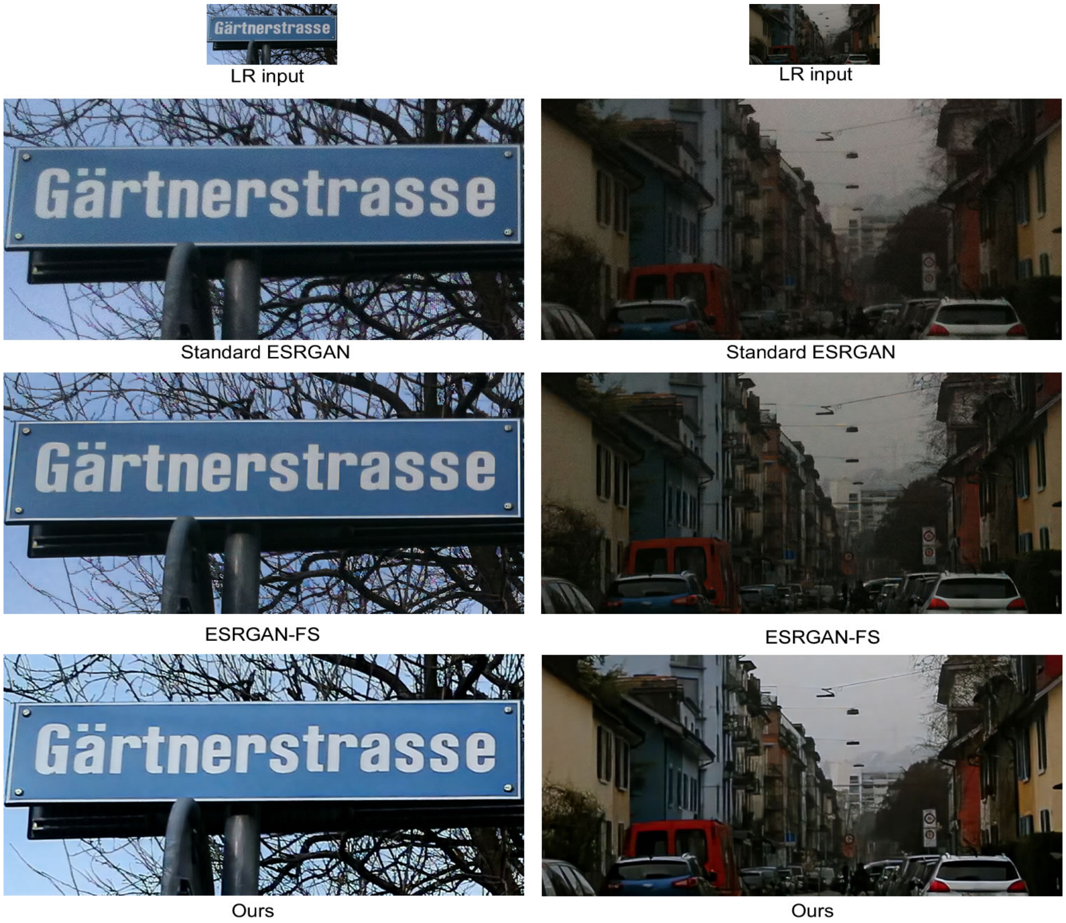

Figure 5. Qualitative comparisons of different SR algorithms on the testing images of NTIRE 2020 RealSR challenge track 2, with scaling factor x4.

mance is clearly better than that of the ESRGAN-FS algorithm which follows an unsupervised way of data generation. These results verify the effectiveness of the usage of our mobile SR dataset generated by super resolving registered DSLR images.

4.3. Ablation study

4.3.1 Using fusion on the SR outputs

In this section, we give the ablation study of our fusion method. Fig. 6 shows some SR outputs on the track 1 testing images with or without fusion (if no fusion applied, only the SR output of the RCAN trained by relativistic GAN is visualized). It can be seen that we are able to reduce the unpleasant artifacts in the low-illumination images by using the proposed fusion algorithm while avoiding losing sharpness. It demonstrates the effectiveness of our proposed fusion algorithm for enhancing the overall visual quality of the images.

Table 2. Different ways to use DPED images to generate our mobile SR dataset.

Label	LR	HR
T1	mobile	DSLR + bicubic upsample
T2	mobile	DSLR + perceptual-oriented SR
T3	mobile	DSLR + PSNR-oriented SR
T4	mobile + downsample	DSLR

4.3.2 Different ways to generate mobile SR dataset

There are multiple ways to create our mobile SR dataset based on registered mobile-DSLR image pairs, as given in Table 2. We can either keep the same scale of mobile images and super resolve the DSLR images (T1, T2, T3), or downsample the mobile images while keeping same scale of DSLR images. The differences between T1, T2, and T3 are how to super resolve the HR images, which can be implemented by simple bicubic upsampling (T1), applying perceptual-oriented SR model such as our track 1 RCAN (T2), or a PSNR-oriented SR model (T3, same as the PSNR-oriented RCAN in Table 1).

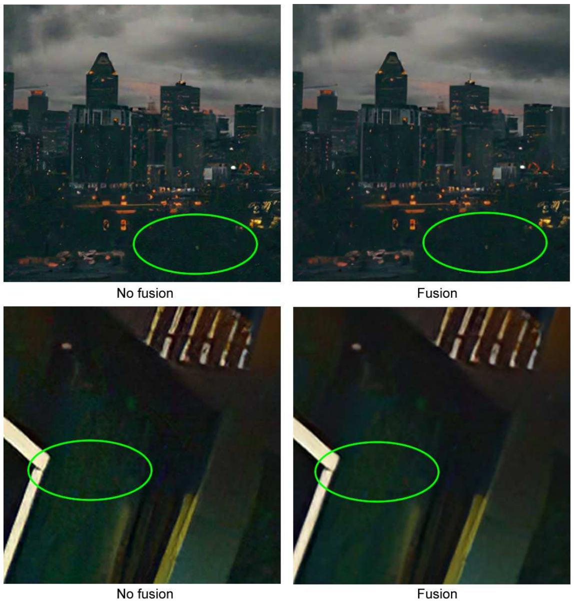

Figure 6. Comparison of the SR outputs of track 1 testing images with and without fusion. We can see that the artifacts in the green circles are improved by the fusion.

In Fig. 7, we give the comparison of the SR outputs of track 2 test set generated by different SR networks trained by applying T1, T2, T3, and T4 on DPED dataset. It can be seen that T2 gives the best perceptual quality. The reason is that the HR images used in training T2 have better perceptual quality than T1, T3, and T4. T3 looks slightly better than T1 due to the reason that the PSNR-oriented SR may generate better HR patches than simple bicubic upsampling. T4 still shows some artifacts because the LR images in T4 are actually blurred twice, one is by the downsampling degradation and the other is by the mobile camera. In contrast, the testing images of NTIRE RealSR track 2 are only blurred by the mobile camera. The artifacts of T4's output come from this model-mismatch problem.

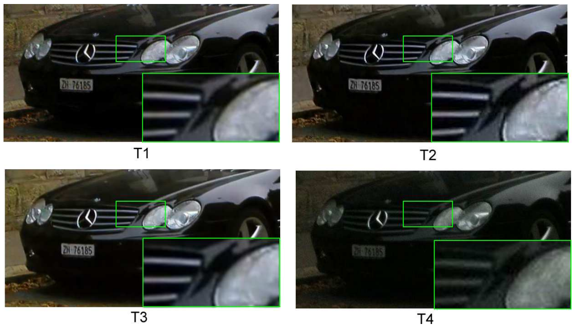

Figure 7. Visualizations of the SR outputs on track 2 testing images, using networks trained on our mobile SR datasets generated by T1, T2, T3, and T4 in Table 2. T2 shows the best perceptual quality.

5. Conclusion

In this paper, we discussed three issues useful for real-world super-resolution. First, we constructed a generic SR dataset using multiple HR-LR degradation models. We also showed that a generic SR model based on GAN can be successfully trained for robust real-world SR using our generic SR dataset. Second, we have proposed a novel weakly supervised SR scheme to generate higher resolution images from those captured by real mobile cameras. A mobile SR dataset is created using registered mobile-DSLR images, where the mobile images are used as LR and the super resolved DSLR images are used as HR. The perceptual quality on mobile images can be improved by fine-tuning our generic SR model on this mobile SR dataset. Third, we demonstrate that the ensemble of different GANs can reduce the artifacts of low-illumination images, which contributes to the overall perceptual quality. Experimental results on NTIRE 2020 RealSR challenge dataset demonstrate the effectiveness of our method.

References

[1] Sefi Bell-Kligler, Assaf Shocher, and Michal Irani. Blind super-resolution kernel estimation using an internal-GAN. In *Advances in Neural Information Processing Systems*, pages 284–293, 2019.

[2] Yochai Blau and Tomer Michaeli. The perception-distortion tradeoff. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 6228–6237, 2018.

[3] 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 *IEEE International Conference on Computer Vision*, pages 3086–3095, 2019.

[4] Chang Chen, Zhiwei Xiong, Xinmei Tian, Zheng-Jun Zha, and Feng Wu. Camera lens super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1652–1660, 2019.

[5] Ryan Dahl, Mohammad Norouzi, and Jonathon Shlens. Pixel recursive super resolution. *arXiv preprint arXiv:1702.00783*, 2017.

[6] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE Transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2016.

[7] Claude E Duchon. Lanczos filtering in one and two dimensions. *Journal of applied meteorology*, 18(8):1016–1022, 1979.

[8] Manuel Fritsche, Shuhang Gu, and Radu Timofte. Frequency separation for real-world super-resolution. *arXiv preprint arXiv:1911.07850*, 2019.

[9] Jinjin Gu, Hannan Lu, Wangmeng Zuo, and Chao Dong. Blind super-resolution with iterative kernel correction. In *IEEE conference on computer vision and pattern recognition*, pages 1604–1613, 2019.

[10] Muhammad Haris, Gregory Shakhnarovich, and Norimichi Ukita. Deep back-projection networks for super-resolution.

In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1664–1673, 2018.

[11] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. DSLR-quality photos on mobile devices with deep convolutional networks. In *IEEE International Conference on Computer Vision*, pages 3277–3285, 2017.

[12] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1646–1654, 2016.

[13] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Deeply-recursive convolutional network for image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1637–1645, 2016.

[14] 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 *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4681–4690, 2017.

[15] Christian Ledig, Lucas Theis, Ferenc Huszár, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[16] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 136–144, 2017.

[17] Andreas Lugmayr, Martin Danelljan, and Radu Timofte. Unsupervised learning for real-world super-resolution. *arXiv preprint arXiv:1909.09629*, 2019.

[18] Andreas Lugmayr, Martin Danelljan, Radu Timofte, et al. Aim 2019 challenge on real-world image super-resolution: Methods and results. In *IEEE International Conference on Computer Vision Workshops*, 2019.

[19] Andreas Lugmayr, Martin Danelljan, Radu Timofte, et al. Ntire 2020 challenge on real-world image super-resolution: Methods and results. *IEEE Computer Vision and Pattern Recognition Workshops*, 2020.

[20] Jinshan Pan, Deqing Sun, Hanspeter Pfister, and Ming-Hsuan Yang. Blind image deblurring using dark channel prior. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1628–1636, 2016.

[21] Haoyu Ren, Mostafa El-Khamy, and Jungwon Lee. Image super resolution based on fusing multiple convolution neural networks. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1050–1057, 2017.

[22] Haoyu Ren, Mostafa El-Khamy, and Jungwon Lee. Dn-resnet: Efficient deep residual network for image denoising. In *Asian Conference on Computer Vision*, pages 215–230. Springer, 2018.

[23] Haoyu Ren, Mostafa El-Khamy, and Jungwon Lee. Cnf+ct: Context network fusion of cascade trained convolutional neural networks for image super-resolution. *IEEE Transactions on Computational Imaging*, 2019.

[24] Yaniv Romano, John Isidoro, and Peyman Milanfar. RAISR: Rapid and accurate image super resolution. *IEEE Transactions on Computational Imaging*, 3(1):110–125, 2016.

[25] Jae Woong Soh, Gu Yong Park, Junho Jo, and Nam Ik Cho. Natural and realistic single image super-resolution with explicit natural manifold discrimination. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 8122–8131, 2019.

[26] Ying Tai, Jian Yang, and Xiaoming Liu. Image super-resolution via deep recursive residual network. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[27] Ying Tai, Jian Yang, Xiaoming Liu, and Chunyan Xu. Memnet: A persistent memory network for image restoration. In *IEEE International Cconference on Computer Vision*, pages 4539–4547, 2017.

[28] Radu Timofte, Vincent De Smet, and Luc Van Gool. A+: Adjusted anchored neighborhood regression for fast super-resolution. In *Asian Conference on Computer Vision*, pages 111–126. Springer, 2014.

[29] 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 *European Conference on Computer Vision*, 2018.

[30] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma. Image super-resolution via sparse representation. *IEEE Transactions on image processing*, 19(11):2861–2873, 2010.

[31] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Learning a single convolutional super-resolution network for multiple degradations. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3262–3271, 2018.

[32] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 586–595, 2018.

[33] Wenlong Zhang, Yihao Liu, Chao Dong, and Yu Qiao. Ranksrgan: Generative adversarial networks with ranker for image super-resolution. In *IEEE International Conference on Computer Vision*, pages 3096–3105, 2019.

[34] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *European Conference on Computer Vision*, pages 286–301, 2018.

[35] Ruofan Zhou and Sabine Susstrunk. Kernel modeling super-resolution on real low-resolution images. In *IEEE International Conference on Computer Vision*, pages 2433–2443, 2019.