

Unsupervised Real-World Super Resolution with Cycle Generative Adversarial Network and Domain Discriminator

Gwantae Kim¹ Jaihyun Park¹ Kanghyu Lee¹ Junyeop Lee¹ Jeongki Min¹ Bokyeung Lee¹
David K. Han² Hanseok Ko¹

¹Korea University, Seoul, Korea ²Army Research Laboratory, Adelphi, MD, USA

Abstract

This paper proposes an unsupervised single-image Super-Resolution(SR) model using cycleGAN and domain discriminator to solve the problem of SR with unknown degradation using unpaired dataset. In previous approaches, paired dataset is required for training with assumed levels of image degradation. In real world SR applications, however, training sets are typically not of low and high resolution image pairs, but only low resolution images with unknown degradation are provided as inputs. To address the problem, we introduce a cycle-in-cycle GAN based unsupervised learning model using an unpaired dataset. In addition, we combine several losses attributed to image contents, such as pixel-wise loss, VGG feature loss and SSIM loss, for stable learning and performance improvement. We also propose a domain discriminator, which consists of noise discriminator, texture discriminator and color discriminator, to guide generated images to follow target domain distribution rather than source domain. We validate effectiveness of our model in quantitative and qualitative experiments using NTIRE2020 real-world SR challenge dataset.

1. Introduction

Many deep learning based Single-Image Super-Resolution(SISR) methods have recently achieved outstanding performance improvement not only in Peak Signal to Noise Ratio(PSNR)-based evaluation[3, 10, 9, 12, 26, 13, 2] but also perceptual image quality evaluation[12, 22]. These methods require high-quality High-Resolution(HR) image and its down-sampled Low-Resolution(LR) image pairs for training in supervised learning manner. However, in real-world scenarios, only unpaired LR images are given. Since supervised training is not possible, an alternative approach would be desirable. In addition, real world images are often degraded by noise, requiring further processing for



Figure 1. $\times 4$ super resolution results for "0878" in the validation set of the NTIRE 2020 Real World Super-Resolution Challenge Track 1 dataset. Our proposed method successfully reconstructs a clean result with texture details compared to state-of-the-art models.

enhanced images. AIM2019 Real-image SR Challenge[16] has been designed specifically to tackle these difficult issues in SR tasks.

Along the same motivation, NTIRE2020 Real World Super-Resolution(SR) Challenge[17] provided a dataset, which consists of degraded HR images and un-paired high quality HR images. The dataset also included two smaller sets of LR images grouped as validation and test. The cause of degradation on these images is unknown. Since the challenge is focused on perceptual image quality, the aim of the challenge is to reconstruct visually pleasing images and not necessarily focused on quantitative metrics.

Since Dong et al.[3] proposed a SR method using Convolutional Neural Networks(CNN), numerous supervised learning based SR approaches[3, 10, 9, 12, 26, 13, 2] have been proposed. Super Resolution Convolution Neural Networks(SRCNN)[3] introduced an end-to-end deep learning approach using CNN networks. In SRCNN, it's been shown that performance improvement can be achieved with deeper layers. However, these improvements are incremental as deeper layers would result vanishing/exploding gradients or overfitting. To avoid these limitations, Very Deep Super Resolution(VDSR)[9] was proposed as a residual-learning method. With the residual connection proposed by He et al.[6], deeper networks can be trained accurately. Additionally, other approaches[26, 22] presented improved feature extraction blocks, such as Residual Dense Block(RDB) or Residual-in-Residual Dense Block(RRDB), which exploits skip-connection structures to allow larger receptive fields and to avoid overfitting.

Despite some successes among these supervised learning based methods, particularly in terms of high PSNR or other quantitative metrics, a major drawback is that the SR images constructed are often not realistic in fine details and may leave blurred impressions perceptually. Moreover, these methods do not tackle the problem of artifacts created in the SR process from the noise in input images. As SR process inevitably involves upsampling, learning based upsampling methods often produce checkerboard patterns[23].

Therefore, we propose an unsupervised single image super-resolution model using a cycleGAN[28] and a domain discriminator to tackle these three issues and to solve the tasks of NTIRE2020 Real-world Image SR Challenge. To produce more realistic and detailed images, we first develop a cycleGAN style multi-stage learning model based on the structure of the Cycle-in-cycle Generative Adversarial Networks(CinCGAN)[24] and Enhanced Super Resolution Generative Adversarial Networks(ESRGAN)[22]. Additionally, we define a content loss by combining a pixel-wise loss, a VGG features loss and a Structural Similarity(SSIM) loss. For the checkerboard pattern issue, we modify ESRGAN's up-sampling structure to take advantages of bilinear interpolation and transposed convolution. For mitigating the input noise created artifacts, we propose a novel discriminator, named domain discriminator, to remove artifacts and preserve color information and texture details. The proposed model is validated using NTIRE2020 Real World Image SR Challenge Track 1 and Track 2 datasets. Our model is able to generate visually more realistic images compared to state-of-the-art approaches.

The main contributions of this paper are:

- We propose a novel and effective cycle-in-cycle GAN based structure to remove image processing artifacts coming from noise in LR images.

- We introduce a new content loss, which consists of L1 loss, VGG feature loss and SSIM loss, to produce more realistic and detailed SR images.
- We propose a novel upsampling method for mitigating checkerboard effects with improved learning stability and enhanced perceptual quality.
- We present a domain discriminator effective in transforming images from a source domain to a target domain.

These proposed elements have been shown efficacious in removing artifacts, reconstructing texture details, and preserving color of the source domain image. We demonstrate the visual improvement of the proposed method by an empirical study.

2. Related work

First we investigated some approaches that adopt Generative Adversarial Networks(GAN)[5] to improve perceptual image quality[12, 22, 21]. Super Resolution Generative Adversarial Networks(SRGAN)[12] introduced a domain transfer model between low and high resolution domains based on a GAN model. ESRGAN[22], which is based on SRGAN, introduced a novel loss function for improving the discriminator to generate more realistic images. Although quantitative scores of these GAN-based models are not better than non-GAN models based on CNN, they generated more perceptually realistic images. Since these approaches require a paired dataset to train generator and discriminator, they are not suitable for the task here.

To handle the un-paired dataset for the given super resolution task, we explored unsupervised learning approaches[28, 7, 24, 19, 4]. CycleGAN[28] proposed un-paired image-to-image translation with cycle consistency loss. Although it achieved successful results in style transfer of images, such as horse to zebra, the transferred images sometimes contain artifacts. Park et al.[19] introduced multi-discriminator learning with CycleGAN for image enhancement task in underwater images. WESPE[7] introduced adversarial color and texture discriminators for image enhancement using a Gaussian blur kernel and grayscale images to focus on color distribution and texture of the image separately. While it is not an SR method, we adopt these discriminators as our input images contain noise and require image enhancement in the SR task. CinCGAN[24] proposed an unsupervised cycle-in-cycle network to solve a blind SR problem and a denoising problem at the same time. However, they only used pixel-wise loss for content loss, hence, the performance was limited to texture recovery.

SR and in several other applications of GAN recently reported some impressive performances of their GAN mod-

els by incorporating additional roles of discriminators [25, 1, 8, 22]. RankSRGAN[25] defined ranker module using Siamese-like architecture to generate visually pleasing images. Self-supervised GAN[1] proposed a discriminator with rotation-based self-supervision. Their discriminator classifies not only the image is real or fake but also the rotation degree of the image. RaGAN[8] and ESRGAN[22] introduced a relativistic discriminator to enhance effects of the adversarial loss. Similarly, we define a new role of a discriminator by having it determine if an image is closer to the target domain or to the source domain. A loss function computed from this new role of the discriminator has been shown effective in the SR task.

3. Proposed Method

In this challenge, a set of LR images with unknown degradation and a set of clean HR images are given. We define the noisy LR images as x and the clean HR images as z . The problem is to train a function F to transform x as $\hat{x} = F(x)$ such that \hat{x} belongs to the same target domain as z . Training F in a single step is too hard and it's been known that such an attempt may result noisy patterns. We, therefore, separate the problem into two parts: a domain transfer for noise removal and another domain transfer from LR to SR. In the domain transfer noise removal, G_1 learns the mapping from the noisy source domain to the target clean domain of LR images. In the super-resolution part, SR learns the mapping from the LR space to the HR space. Therefore, the function F can be defined as $F(x) = SR(G_1(x))$.

The overall architecture is shown in Fig.2. The model consists of three generators G_1, G_2 and G_3 , one SR model and three discriminators D_N, D_Y and D_C .

We first train the generators G_1, G_2 and the discriminators D_N, D_Y, D_C in a cycleGAN setting in order to train the domain transfer mapping function G_1 for noise removal.

Secondly, we construct an upsampling module that combines a bilinear interpolation upsampling and a transposed convolution upsampling to solve the checkerboard effects and to enforce learning stability. We use an HR image z in target domain and its bicubically downsampled image $z' = B(z)$ for pre-training. The SR model is trained by a pixel-wise L1 loss. Finally, the generators G_1, G_2, G_3 , the discriminators D_N, D_Y, D_C and SR are trained with in a modified cycleGAN architecture for generating more realistic and detailed SR images.

3.1. Domain Transfer from noisy LR to clean LR images

We first develop a domain transfer model G_1 for denoising and effective image enhancement. We employ the CinCGAN[24] architecture and WESPE[7] discriminators to solve the problem. The detailed network architec-

ture of generators and discriminators, which are based on CinCGAN[24], is presented in Fig.2. We essentially modified the number of blocks and normalization strategy, aimed at effective domain transfer learning. We employ the network structure in G_1 and G_2 with skip-connections and residual blocks for denoising. The input and the output of G_1 are an LR image in source domain and an LR image in target domain, respectively. The input and the output of G_2 are an LR image in target domain and an LR image in source domain, respectively. We do not use batch normalization because our batch size is small. We also do not use instance normalization because it tends to follow style information of the target domain excessively. Instead, we use a batch-instance normalization[18] for discriminator because it has an advantage of capturing style and content information from the images.

In [27], combining several loss functions, such as L1 loss and SSIM loss, enhances training stability and helps the network achieve performance improvement. We employ pixel-wise content loss, VGG19 feature loss[20] and SSIM loss for perceptual quality improvement and preserving edge presence. The final content loss is defined as follows:

$$\begin{aligned} Content(a, b) &= \frac{1}{N} \sum_i ||a_i - b_i||_1 \\ &+ \frac{1}{N} \sum_i ||\phi(a_i) - \phi(b_i)||_2^2 \\ &+ 2 - SSIM(a, b) \end{aligned} \quad (1)$$

where ϕ denotes the features before activation function extracted from the VGG19 networks. Using Equation (1), the final cycle consistency loss for LR images is defined as:

$$\mathcal{L}_{cyc}^{LR_x} = Content(G_2(G_1(x)), x) \quad (2)$$

$$\mathcal{L}_{cyc}^{LR_{z'}} = Content(G_1(G_2(z')), z') \quad (3)$$

$$\mathcal{L}_{cyc}^{LR} = \mathcal{L}_{cyc}^{LR_x} + \mathcal{L}_{cyc}^{LR_{z'}} \quad (4)$$

We also use an identity loss to improve color presence performance.

$$\mathcal{L}_{idt}^{LR} = Content(G_1(z'), z) \quad (5)$$

The three discriminators D_N, D_Y, D_C help the generators and the SR model achieve learning the accurate mapping function from the source domain to the target domain. We propose a special discriminator, named "domain discriminator", which is different from the standard discriminator in SRGAN. The domain discriminator predicts with a probability that a generated image is in source domain or target domain instead of playing the usual role of classifying whether the image is real or fake. Therefore, the source

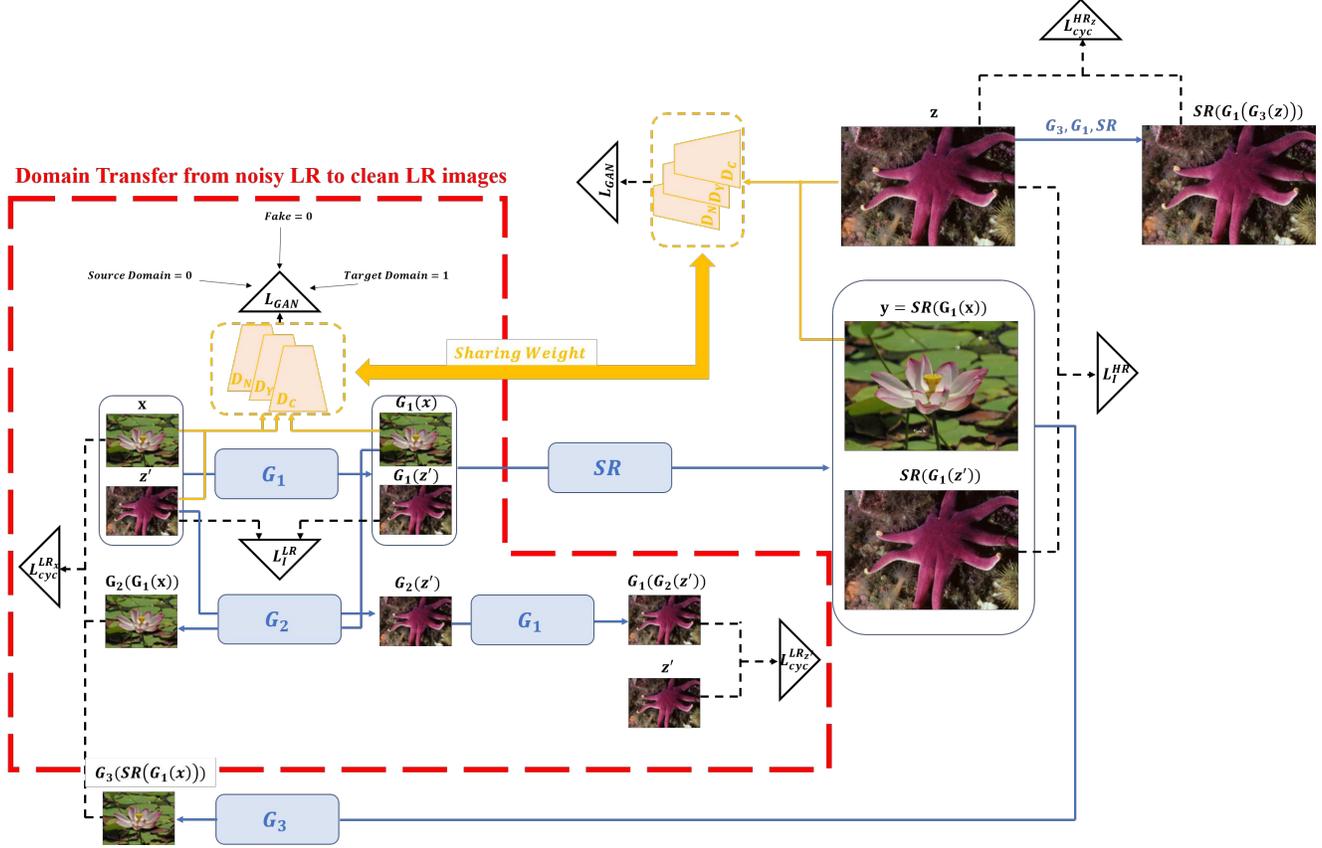


Figure 2. The proposed training model architecture, where G_1 , G_2 and G_3 are generators and SR is super resolution networks. D_N , D_Y and D_C are discriminator for noise, texture and color, respectively. Initially, G_1 , G_2 and D_N , D_Y , D_C are trained by domain transfer learning for LR image. Secondly, SR model is trained by supervised learning. Finally, The overall network architecture is tuned by the domain transfer learning.

domain image can be used as the prior information for the discriminator. The basic concept of the domain discriminator is depicted in Fig. 3. Following Goodfellow et al.[5]

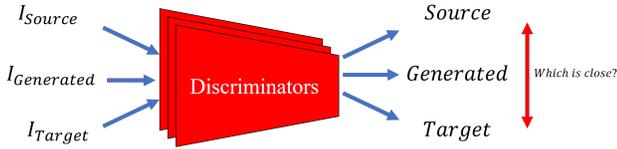


Figure 3. Concept of the proposed domain discriminator.

we apply the adversarial min-max framework to the domain discriminator concept as follows:

$$\min_{\theta_G} \max_{\theta_D} \mathbb{E}_{x \sim p_{train}(x)} [\log(1 - D(x))] + \mathbb{E}_{x \sim p_G(x)} [\log(1 - D(G_1(x)))] + \mathbb{E}_{z \sim p_{train}(z)} [\log(D(z))] \quad (6)$$

where x is source domain image, z is target domain image. When a source domain image is passed through the discrim-

inator, it is classified as "0". Likewise, target domain image is classified to "1". With this arrangement, the generator can learn to create solutions that are classified to 1, which subsequently follows the target domain images.

D_N essentially assesses the level of noise present and general image quality. The adversarial loss of D_N is defined by:

$$\mathcal{L}_{GAN_N}^{LR} = - \sum_i \log(1 - D_N(G_1(x))) \quad (7)$$

where x is the source domain image and $G_1(x)$ is the generated image. We also use an adversarial texture loss and an adversarial color loss, which were proposed by WESPE[7]. In WESPE, the author asserted that image texture quality and image color quality can be measured by D_Y and D_C , respectively. The adversarial loss of D_Y and D_C are defined

as:

$$\mathcal{L}_{GAN_Y}^{LR} = - \sum_i \log(1 - D_Y(G_1(x)_g)) \quad (8)$$

$$\mathcal{L}_{GAN_C}^{LR} = - \sum_i \log(1 - D_C(G_1(x)_b)) \quad (9)$$

where a subscript *g* is grayscale image and a subscript *b* is Gaussian blurred image. The shape of Gaussian kernel is the same as WESPE. Our full objective function for LR image is:

$$\mathcal{L}_{GAN}^{LR} = \omega_1 \mathcal{L}_{GAN_N}^{LR} + \omega_2 \mathcal{L}_{GAN_Y}^{LR} + \omega_3 \mathcal{L}_{GAN_C}^{LR} \quad (10)$$

$$\mathcal{L}^{LR} = \mathcal{L}_{idt}^{LR} + \lambda_1 \mathcal{L}_{cyc}^{LR} + \lambda_2 \mathcal{L}_{GAN}^{LR} \quad (11)$$

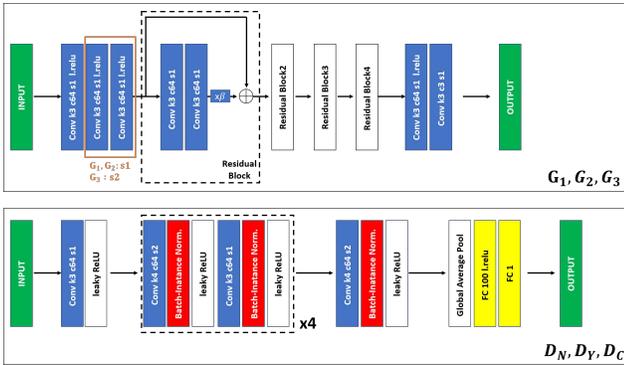


Figure 4. The model structure of generator and discriminator. β is residual scaling. The stride of the second and third layer of G_1 and G_2 is 1, but the stride of G_3 is 2 because of downsampling.

3.2. Super-Resolution Learning

In this stage, the SR model is pre-trained in a supervised manner. We employ ESRGAN[22] structure to the SR model because ESRGAN achieves good qualitative results in perceptual-driven approaches. However, a transposed convolution upsampling module of the ESRGAN often results checkerboard effects. Therefore, we modify the up-sampling module that combines a bilinear interpolation module and a transposed convolution module. Since the interpolation with normal convolution has benefits of learning stability and removing checkerboard pattern artifacts[15], we add an interpolation based method and transposed convolution with residual scaling to take advantage of both. We found this to be beneficial for enhanced perceptual quality. Furthermore, following [15]’s results, our final output forms SR images upon going through the final convolution layer with an activation process. The architecture of the SR model is shown in Fig. 5.

The SR model is pre-trained with target domain data for training stability. The loss function computed from L1 loss

between SR images $SR(z')$ and HR images z is defined as:

$$\mathcal{L}^{SR} = \frac{1}{N} \sum_i ||SR(z'_i) - z_i||_1 \quad (12)$$

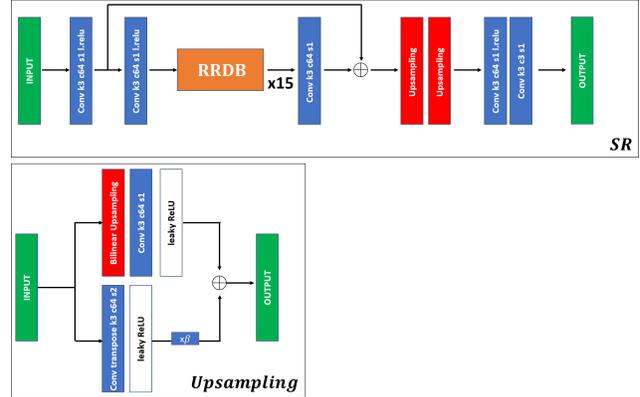


Figure 5. The model structure of SR model and the proposed upsampling module. β is residual scaling. 15 RRDB modules are used for residual connection module. We introduce an upsampling module that contains bilinear upsampling and transposed convolution.

3.3. Domain Transfer Learning for the Overall Network

In sections 3.1 and 3.2, the mapping function G_1 and SR are pre-trained because an end-to-end training strategy is too hard. However, G_1 is trained with only LR images and SR is trained with bicubic downsampled images and paired HR images. Since our final goal is estimating the target domain HR image from the source domain LR image, a fine-tuning is needed for an optimized solution. Therefore, the pre-trained models G_1 , G_2 , SR , D_N , D_Y and D_C are loaded and fine-tuned with the loss from LR and HR images. Compared to section 3.1, we add one additional cycle, which consists of G_1 , SR and G_3 modules, to address domain transfer learning between source domain LR images and target domain HR images. The cycle-consistency loss and identity of second cycle are defined as:

$$\mathcal{L}_{cyc}^{HR_x} = Content(SR(G_1(x)), x) \quad (13)$$

$$\mathcal{L}_{cyc}^{HR_z} = Content(SR(G_1(G_3(z))), z) \quad (14)$$

$$\mathcal{L}_{idt}^{HR} = Content(SR(G_1(z')), z) \quad (15)$$

The GAN loss of D_N , D_Y and D_C are added because the discriminators deal with information from the HR images.

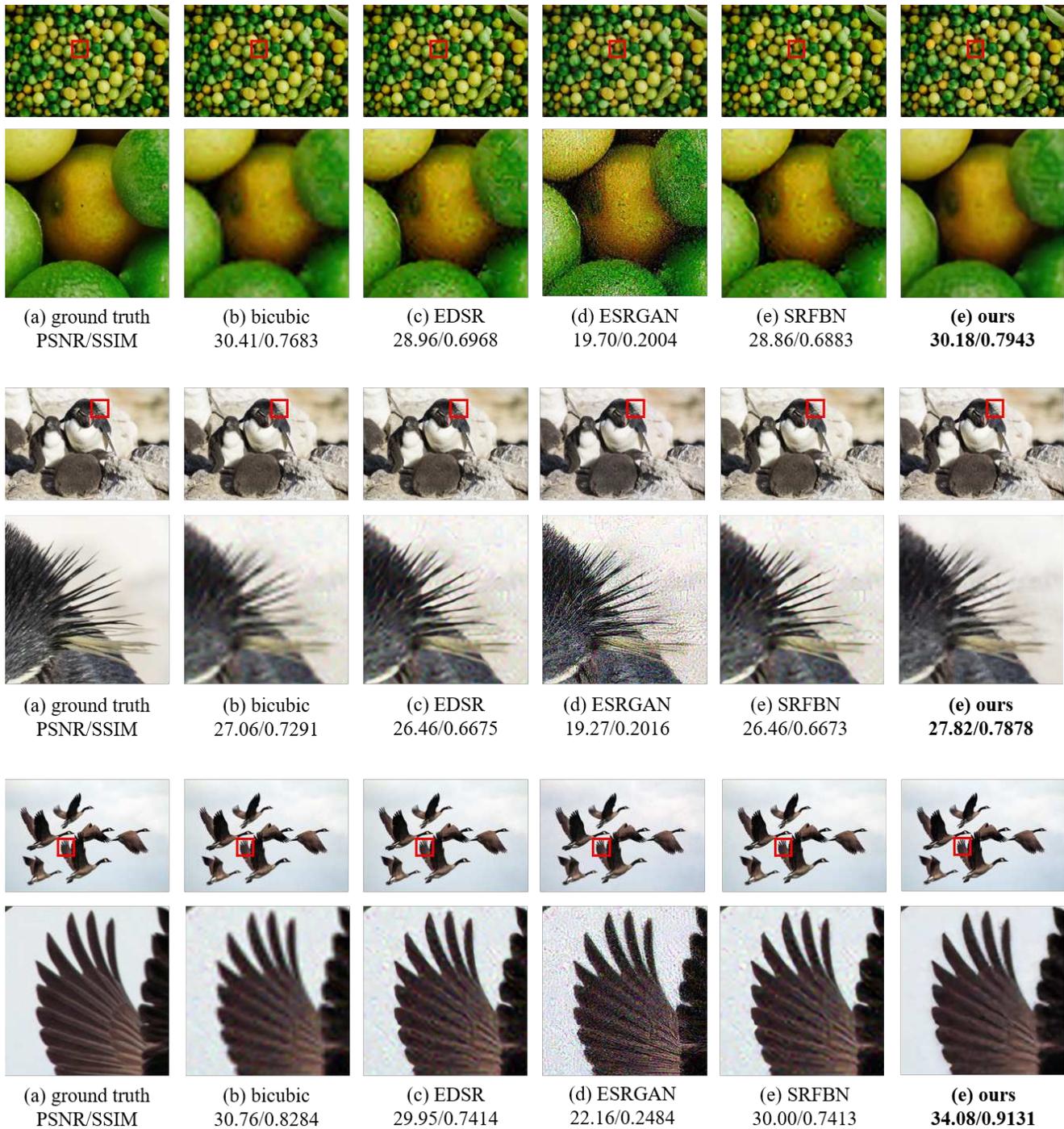


Figure 6. Super-resolution results of "0802", "0842" and "0896" in the validation set of the NTIRE2020 Real World SR challenge Track 1 with scale factor $\times 4$. Our method is able to remove unknown noise and reconstruct detailed textures successfully.

Table 1. Quantitative results on NTIRE 2020 Real World SR challenge Track 1 validation dataset of the proposed model compared to state-of-the-art methods.

Methods	bicubic	EDSR	ESRGAN	SRFBN	ours
PSNR/SSIM	25.49/0.6721	25.36/0.6404	19.04/0.2423	25.37/0.6417	26.39/0.7329

The new GAN loss is:

$$\mathcal{L}_{GAN_N}^{HR} = \mathcal{L}_{GAN_N}^{LR} - \sum_i \log(1 - D_N(SR(G_1(x)))) \quad (16)$$

$$\mathcal{L}_{GAN_Y}^{HR} = \mathcal{L}_{GAN_N}^{LR} - \sum_i \log(1 - D_Y(SR(G_1(x))_g)) \quad (17)$$

$$\mathcal{L}_{GAN_C}^{HR} = \mathcal{L}_{GAN_N}^{LR} - \sum_i \log(1 - D_C(SR(G_1(x))_b)) \quad (18)$$

The final objective function is defined as :

$$\mathcal{L}_{GAN} = \omega_1 \mathcal{L}_{GAN_N}^{HR} + \omega_2 \mathcal{L}_{GAN_Y}^{HR} + \omega_3 \mathcal{L}_{GAN_C}^{HR} \quad (19)$$

$$\mathcal{L}_{cyc} = \mathcal{L}_{cyc}^{LR} + \mathcal{L}_{cyc}^{HR_x} + \mathcal{L}_{cyc}^{HR_z} \quad (20)$$

$$\mathcal{L}_{idt} = \mathcal{L}_{idt}^{LR} + \mathcal{L}_{idt}^{HR} \quad (21)$$

$$\mathcal{L}_{total} = \mathcal{L}_{idt} + \lambda_1 \mathcal{L}_{cyc} + \lambda_2 \mathcal{L}_{GAN} \quad (22)$$

4. Experimental Results

In this section, we first introduce the dataset and the training details. Next, the performance of the proposed model is evaluated by comparing with several state of the art SR methods. Finally, we perform an ablation study to validate the proposed discriminators and the training model structure.

4.1. Training Data

The dataset from NTIRE 2020 Real World SR challenge is composed of 2650 training images in the source domain, 800 training images in the target domain, 100 validation images, and 100 testing images in Track 1. As described earlier, the source domain HR images of the dataset are not paired with the target domain HR images. Furthermore, there are some unknown degradation in the source domain HR images. The LR images are not provided for training data. In Track 2, 5901 source domain training images in the source domain, 800 training images in the target domain, 112 validation images, and 100 testing images are provided. The source domain images are captured by a smartphone containing color degradation and sensor noise.

We use all the training data in source domain and target domain during training phase. We augment the dataset with flipping and 90 degree rotation of images which are already in the dataset. Our experiments are performed with a scaling factor of $\times 4$. We use a 56×56 size patch for LR images and a 224×224 size patch for HR images. We conduct testing on the provided 100 validation images in Track 1 and 100 testing images in Track 2. Note that Track 2 does not have ground truth image.

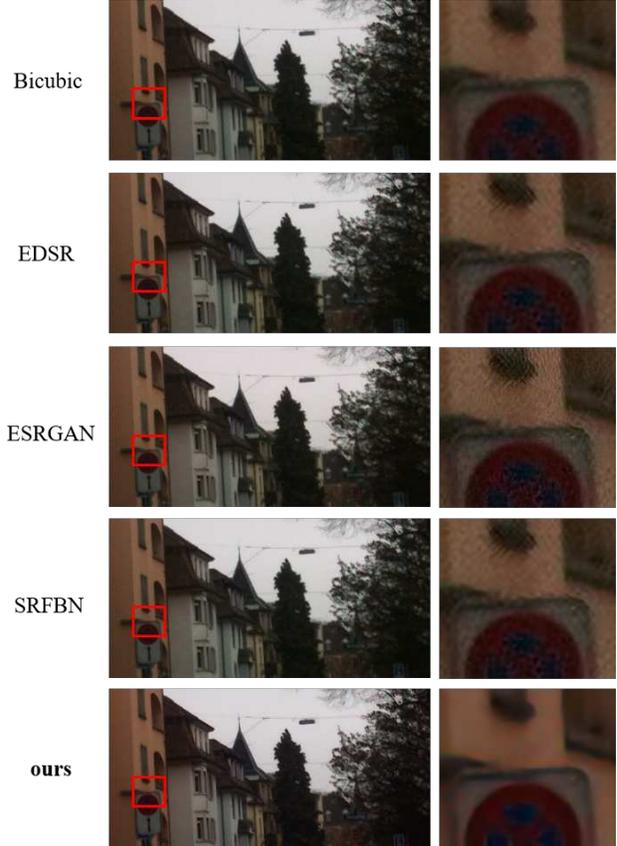


Figure 7. Super-resolution results of "00028" in the test set of the NITRE2020 Real World SR challenge Track 2 with scale factor $\times 4$. Since other state-of-the-art methods are not trained with smartphone image dataset, they produce large artifacts on the smartphone images. In contrast, our method alleviates the artifacts and improves perceptual quality.

4.2. Training Details

We divide our training process into three stages. Initially, we train generators G_1, G_2 and discriminators D_N, D_Y, D_C for mapping LR images in a noisy domain to LR images in a noise-free domain. We set the parameters as $\omega_1 = \omega_2 = \omega_3 = 1, \lambda_1 = 1, \lambda_2 = 0.01$, respectively. We train our model with an Adam optimizer[11] with $\beta_1 = 0.5, \beta_2 = 0.999$, batch size 16 and a learning rate of 1×10^{-4} . Since this process is pre-training, we train our model 100000 iterations, which is not fully optimized. Next, we train the SR with a supervised learning strategy. We also train our model with an Adam optimizer with a batch size 8 and a learning rate of 1×10^{-4} . At this point, we train the model over 800000 iterations, until it converges.

Finally, we train the overall structure for mapping LR images in the source domain to HR images in the target domain. We set the parameters as $\omega_1 = \omega_2 = 2, \omega_3 = 1, \lambda_1 = 0.5, \lambda_2 = 5 \times 10^{-4}$ for track 1 and $\omega_1 = 1, \omega_2 = \omega_3 =$

$2, \lambda_1 = 2, \lambda_2 = 1 \times 10^{-3}$ for track 2. We also use an Adam optimizer with batch size 2. The learning rate is initialized as 1×10^{-4} and then it was decreased by a factor of 2 every 100000 iterations. We train the model over 300000 iterations until it converges. The network is trained on three NVIDIA Titan Xp GPUs and the model is implemented using tensorflow.

4.3. Quantitative Evaluation Measures

In order to quantitatively compare the different approaches we use metrics such as PSNR and SSIM. PSNR is calculated over all pixel values in RGB channels and SSIM is measured by taking a mean of SSIM index over the RGB channels.

4.4. Results

We demonstrate effectiveness of the proposed method against state-of-the-art approaches for SISR: EDSR[14], ESRGAN[22], SRFBN[13]. We use the publicly available code and optimized models provided by the authors. Table 1 shows the average PSNR and SSIM values of the NTIRE2020 validation set with the different methods. Our method outperforms the previous methods in denoising performance. The EDSR and SRFBN, which are PSNR-oriented method, has almost the same PSNR and SSIM because the resultant images remain noisy. Especially, the performance of ESRGAN, which is a perceptual quality-oriented approach, is not better than the others. We analyze that the noise is enhanced because the ESRGAN model recognized noise as a pattern of the image. Therefore, our proposed model has a quantitative performance advantage when an input image contains noise. The results of perceptual image quality in Track 1 and Track 2 are shown in Fig.6 and Fig.7, respectively. Our approach successfully overcomes the unknown degradation problem and reconstructs more realistic images.

4.5. Ablation Study

To validate the advantages of the discriminator, we implemented a vanilla GAN discriminator, which classifies an image to be real or fake, for comparison. The network structure is the same, but the minmax problem and GAN loss are modified respectively as:

$$\min_{\theta_G} \max_{\theta_D} \mathbb{E}_{z' \sim p_{train}(z)} [\log(D_N(z))] + \mathbb{E}_{x \sim p_G(x)} [\log(1 - D_N(G_1(x)))] \quad (23)$$

$$\mathcal{L}_{GAN} = - \sum_i \log(1 - D(G_1(x))) - \sum_i \log(1 - D(SR(G_1(x)))) \quad (24)$$

The quantitative results of the vanilla GAN and the proposed method are summarized in Table 2. As shown by the results, the proposed GAN loss model outperforms the model with vanilla GAN loss in PSNR and SSIM both.

Table 2. Quantitative results of the ablation study on NTIRE 2020 Real World SR challenge Track 1 validation dataset, comparing the proposed GAN loss model vs. vanilla GAN loss model

Methods	proposed loss	vanillaGAN loss
PSNR/SSIM	26.39/0.7329	25.73/0.6979

Moreover, the qualitative results are illustrated in Fig. 8. The proposed method is shown to restore edge details of the image clearly compared to the adversarial loss of the vanilla GAN.

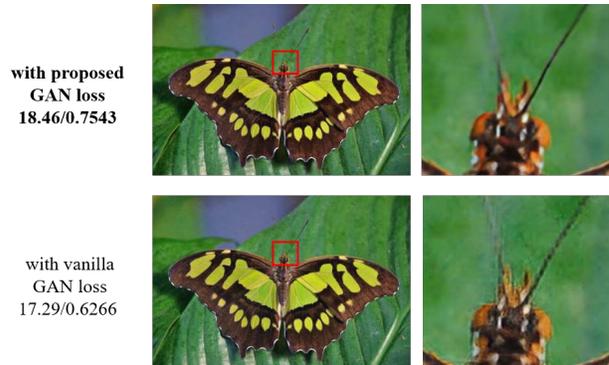


Figure 8. Super-resolution results of "0828" in the validation set of the NTIRE2020 Real World SR challenge Track 1 with scale factor $\times 4$. Our method is able to reconstruct texture details better than the vanilla GAN based model.

5. Conclusion

Inspired by the recent successful domain transfer learning approaches, we proposed an unsupervised learning method to solve the SISR problem with unknown degradation using an unpaired dataset. To avoid checkerboard artifacts while preserving details, we proposed an upsampling module that combines bilinear interpolation and transposed convolution. In addition, we developed a content loss using several types of loss and introduced a domain discriminator to improve quantitative and qualitative results. Compared to the state-of-the-art approaches, our method generated more realistic and visually pleasing images when input image was corrupted by unknown degradation.

6. Acknowledgement

This material is based upon work supported by the Air Force Office of Scientific Research under award number FA2386-19-1-4001.

References

- [1] Ting Chen, Xiaohua Zhai, Marvin Ritter, Mario Lucic, and Neil Houlsby. Self-supervised gans via auxiliary rotation loss. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 12154–12163, 2019.
- [2] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. Second-order attention network for single image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 11065–11074, 2019.
- [3] 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, 2015.
- [4] Manuel Fritsche, Shuhang Gu, and Radu Timofte. Frequency separation for real-world super-resolution. In *ICCV Workshops*, 2019.
- [5] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.
- [6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [7] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. Wespe: weakly supervised photo enhancer for digital cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 691–700, 2018.
- [8] Alexia Jolicoeur-Martineau. The relativistic discriminator: a key element missing from standard gan. *arXiv preprint arXiv:1807.00734*, 2018.
- [9] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016.
- [10] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Deeply-recursive convolutional network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1637–1645, 2016.
- [11] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [12] 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.
- [13] Zhen Li, Jinglei Yang, Zheng Liu, Xiaomin Yang, Gwanggil Jeon, and Wei Wu. Feedback network for image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3867–3876, 2019.
- [14] 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.
- [15] Andreas Lugmayr, Martin Danelljan, and Radu Timofte. Un-supervised learning for real-world super-resolution. In *ICCV Workshops*, 2019.
- [16] Andreas Lugmayr, Martin Danelljan, Radu Timofte, et al. Aim 2019 challenge on real-world image super-resolution: Methods and results. In *ICCV Workshops*, 2019.
- [17] Andreas Lugmayr, Martin Danelljan, Radu Timofte, et al. Ntire 2020 challenge on real-world image super-resolution: Methods and results. *CVPR Workshops*, 2020.
- [18] Hyeonseob Nam and Hyo-Eun Kim. Batch-instance normalization for adaptively style-invariant neural networks. In *Advances in Neural Information Processing Systems*, pages 2558–2567, 2018.
- [19] Jaihyun Park, David K Han, and Hanseok Ko. Adaptive weighted multi-discriminator cyclegan for underwater image enhancement. *Journal of Marine Science and Engineering*, 7(7):200, 2019.
- [20] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [21] 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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8122–8131, 2019.
- [22] 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 (ECCV)*, pages 0–0, 2018.
- [23] Zhihao Wang, Jian Chen, and Steven CH Hoi. Deep learning for image super-resolution: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.
- [24] 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.
- [25] Wenlong Zhang, Yihao Liu, Chao Dong, and Yu Qiao. Ranksrgan: Generative adversarial networks with ranker for image super-resolution. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3096–3105, 2019.
- [26] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2472–2481, 2018.
- [27] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz. Loss functions for image restoration with neural networks. *IEEE Transactions on computational imaging*, 3(1):47–57, 2016.

- [28] 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.