Guided Frequency Separation Network for Real-World Super-Resolution

Yuanbo Zhou, Wei Deng, Tong Tong and Qinquan Gao

College of Physics and Information Engineering, Fuzhou University
{webbozhou,weideng.chn,ttraveltong}@gmail.com, gqinquan@fzu.edu.cn

Abstract

Training image pairs are unavailable generally in real-world super-resolution. Although the LR images can be down-scaled from HR images, some real-world characteristics (such as artifacts or sensor noise) have been removed from the degraded images. Therefore, most of state-of-the-art super-resolved methods often fail in real-world scenes.

In order to address aforementioned problem, we proposed an unsupervised super-resolved solution. The method can be divided into two stages: domain transformation and super-resolution. A color-guided domain mapping network was proposed to alleviate the color shift in domain transformation process. In particular, we proposed the Color Attention Residual Block (CARB) as the basic unit of the domain mapping network. The CARB which can dynamically regulate the parameters is driven by input data. Therefore, the domain mapping network can result in the powerful generalization performance. Moreover, we modified the discriminator of the super-resolution stage so that the network not only keeps the high frequency features, but also maintains the low frequency features. Finally, we constructed an EdgeLoss to improve the texture details. Experimental results show that our solution can achieve a competitive performance on NTIRE 2020 real-world super-resolution challenge.

1. Introduction

Recovering high resolution (HR) images from low resolution (LR) images is called super-resolution (SR), which is a basic problem in computer vision. We have witnessed the remarkable development of SR in last years. The emergence of Convolutional Neural Network (CNN) brings a lot of state-of-the-art methods such as [4, 16, 17, 29, 28]. However, these methods often fail to generate high quality images in real-world scenes.

The reason for the above problem is that most models were trained on artificial image pairs. As the LR images are resulted from a known degradation (*e.g.* bicubic down-scaling), the process of SR only recovers the losing details of the degradation operation rather than the nature images. Moreover, the bicubic down-scaling removed the characteristics of real-world images such as artifacts, sensor noise and other nature characteristics, which makes the training data so clean [19]. If we directly utilize these degradation data to train SR model, it cannot work well as the difference between distributions of training data and those of testing data.

Faced this challenging problem, while we are not able to collect the real-world image pairs, the similar domain images with real-world can be generated by Generative Adversarial Networks (GAN) [10] indirectly. Thereby, many state-of-the-art algorithms were proposed recently, such as [6, 19]. However, these methods break the consistency of color. Although the clean domain images have been transferred to the real-world domain images, the color shifts will hamper the SR process. Specifically, the color shifts allow for partial optimization of the SR network towards color consistency, rather than just recovering the losing of texture detail of LR images. As a result, the performance of the SR network was greatly compromised. The result of SR not only changes the color of the original image, but also causes the over-smoothed image.

In order to tackle aforementioned problem, we proposed a CARB as the basic unit of the domain mapping network. The CARB which can dynamically regulate the parameters is driven by another color feature guided network. The color feature guided network dynamically adjust the parameters of CABA by extracting the mean and variance of the input image, so that each image has a color independent distribution. Therefore, the domain mapping network can maintain the color consistency and obtain the powerful generalization performance.

Inspired by the work in [6], we utilized the strategy of frequency separation. We designed a SR discriminator that treats the low and high frequency features separately. The discriminator can ensure the realness and completeness of

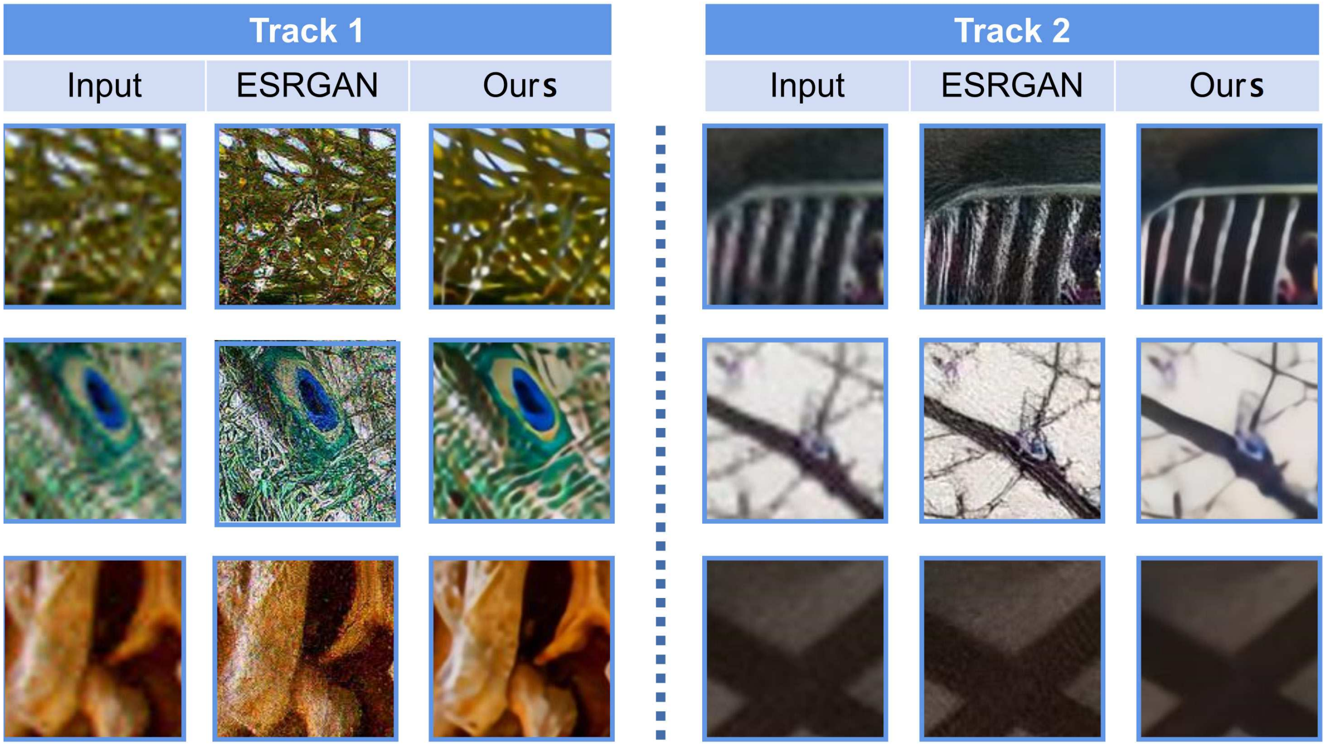

Figure 1. The visual results of SR with an upscaling factor of 4, which demonstrates the effectiveness of our proposed method in real-world SR.

original image, which is critical for many applications such as medical imaging [9] and surveillance [23].

To further enhance the detail presentation of the SR network, we constructed an EdgeLoss with Canny operator [2] by detecting the edge of HR and SR. This loss can effectively make the training of SR more edge-oriented, to the point of getting more edge details that are not normally available by the usual losses.

All in all, the whole solution can achieve a competitive performance for real-world images. Some visual results can be seen in Fig 1. The code is available at `https://github.com/fzuzyb/2020NTIRE-Guided-Frequency-Separation-Network-for-RWSR`. Our contribution can be summarized as follows:

1. We proposed a domain mapping network consists of CARB
2. We designed a SR discriminator that can treat the low and high frequency features separately
3. We constructed an Edgeloss with Canny operator

2. Related work

As we know, the task of SR has always been an ill-posed problem. It's received a lot of enthusiastic attention for researchers over years. While there are many classic methods to solve this problem such as [26, 3, 7, 13], the learning-based approaches grow explosively with the advent of deep-learning. Thereby, more and more state-of-the-art methods continue to emerge. Typically, SRCNN was proposed by Dong *et al.* [4, 5] to address the problem, which is an end-to-end convolutional neural network with supervised leaning and the training data comes from bicubic down-scaling LR and corresponding HR data. Based on this idea, many excellent algorithms have been proposed like [12, 15, 1, 28] to improve the quantitative results. In particular, the presentation of EDSR [18] make the PSNR metric achieve the peak.

Nevertheless, the training process usually use L1 or L2 loss, resulting in a lot of high-frequency details being lost. Thus, the over-smoothed result was outputted, and the visual quality of SR usually is poor. In order to tackle this problem, a pioneer work SRGAN [16] was proposed to produce the significant human visual quality. It introduced the GAN and perceptual loss in SR model, which yields more photo-realistic results comparing with prior works. Along this work, [24, 29] was proposed to further increase the subjective visual perception of the results. Especially ESRGAN [29] was produced, which brings the subjective state-of-the-art algorithm to SR. However, aforementioned methods were limited when faces the real-world dataset like DPED dataset [11]. If we directly use the LR image that

Figure 2. Degradation of real-world data by bicubic downscaling, which removed the sensor noise. The cropped area can easily to distinguish.

was got from the down-scaling HR image to train the SR model, which will outputs the poor results. Therefore, many methods in 2019 AIM challenge [20] was proposed to tackle this problem. Especially, [19] was proposed by Lugmayr *et al*. by utilizing CycleGAN [32]. It can produce the training data which is similar with the distribution of real-world. Furthermore, the frequency separation network was proposed by Fritsche [6], which only uses a GAN to produce the state-of-the-art performance and proposed frequency separation idea for SR.

Although recent approaches have made significant success, the color shifts are inevitable during GAN training. Therefore, in this paper, we proposed a systematic solution to alleviate this phenomenon so that the performance of real-world SR can be further improved.

3. Proposed method

3.1. Overview

As previously stated, the learning-base method is to estimate the mapping from LR to HR with pair data. Unfortunately, it relies on the artificial image pairs. General methods to make training data is to down-scale from the HR images, however there is a gap between training data and testing data. Many real-world characteristics are removed such as sensor noise and artifacts, which can be seen in Fig. 2. Therefore, we proposed a real-world SR solution in an unsupervised manner to perform this challenge. Our solution can be divided two stages: unsupervised SR data generation stage and supervised SR stage. The overall architecture can be seen in Fig. 3.

3.2. Problem formulation

In unsupervised SR data generation stage, we let $\mathcal{Y}$ denotes the domain of real-world HR images. The $\mathcal{X}$ denotes the domain of LR images which down-scales from $\mathcal{Y}$. $\mathcal{Z}$ is the true domain of real-world LR images. We focus on finding a mapping f_1 from $\mathcal{X}$ to $\mathcal{Z}$ to make $f_1(x)$ as similar to z as possible in characteristic while maintains the content, where $x \in \mathcal{X}$, $z \in \mathcal{Z}$.

In supervised SR stage, the image pairs are generated by first stage can be utilized to train SR model. Thereby, we let the $\hat{\mathcal{Z}}$ denotes the generated domain from $\mathcal{X}$. We need to find another mapping f_2 from $\hat{\mathcal{Z}}$ to $\mathcal{Y}$ to make $f_2(\hat{z})$ as similar to y as possible, where $\hat{z} \in \hat{\mathcal{Z}}$, y $\in \mathcal{Y}$.

To sum up, as long as we can make $\hat{\mathcal{Z}}$ and $\mathcal{Z}$ similar enough. Ideally $\hat{\mathcal{Z}}$ = $\mathcal{Z}$, the problem can be simplified into the supervised SR problem.

3.3. Unsupervised SR data generation

Network architecture. In order to implement the domain transformation, we adopt the idea of GAN [10]. Especially, we transfer the LR images by official DSGAN [6] model. However, there is color shift in the degenerate results, which can be seen Fig. 6. If we utilize these pair data to train SR model directly, the results of SR have the phenomenon of over-smoothed image. After our analysis, we believe that the reason for this is Instance Normalization layer lacks a priori about color independence. Therefore, to address this problem, we add a color-guided network to dynamically output the image color features which can be performed with mean and variance so that can be provided to AdaIN [8]. As the HSV space: S indicates saturation and it is related to image variance. V indicates value and it is related to image mean. Therefore, we proposed a generator $G_{x\rightarrow z}$ consisting of two parts, a main network of CARB units and its corresponding parameter network.

The details of generator can be seen Fig. 4. The top half of the network is a guided parameter network, to yield the bias (mean) and weight (variance) of CARB. The bias is the global information, so we utilize several convolutions with kernel size of 3 and three global pooling layers with kernel size of 5 to extract it. After than, the original image subtracts this global information will be fed into the sigmoid layer. The global information is used as bias, and the final output value is used as weight for CARB. For the

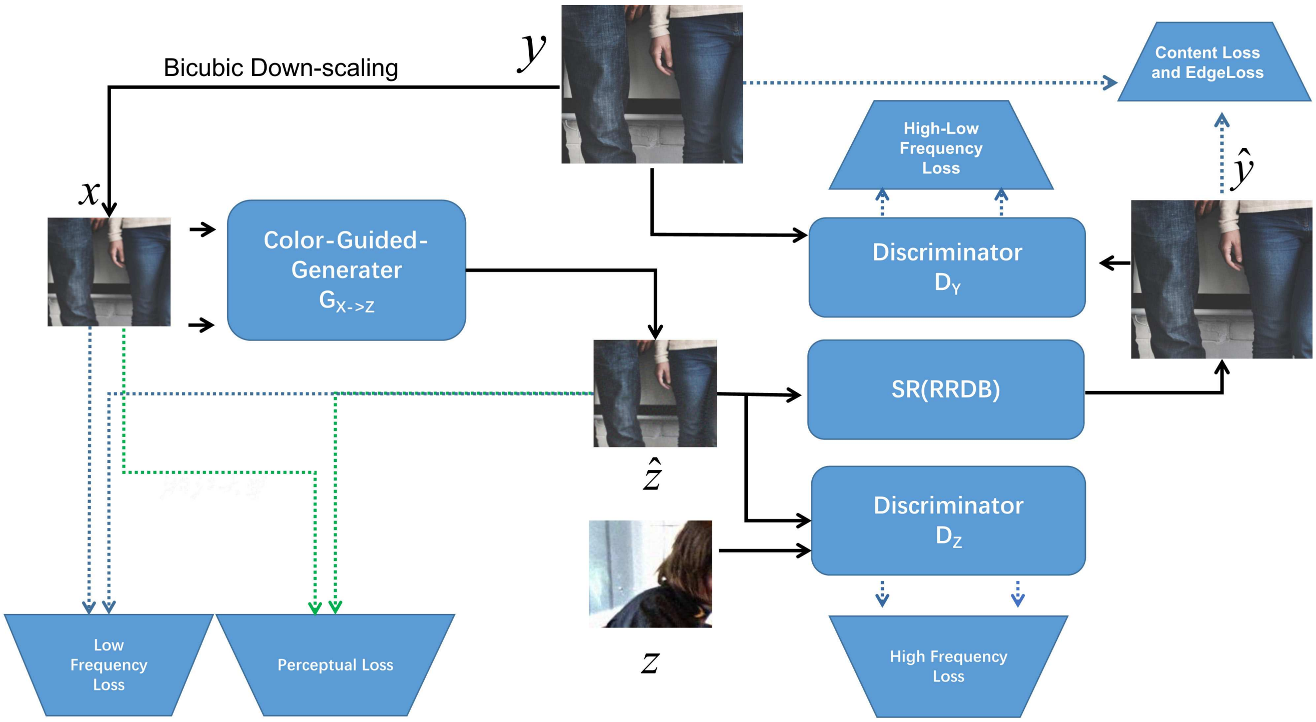

Figure 3. The overall architecture. the details of generator and discriminator can be seen Fig. 4 and Fig. 5.

CARB, this is a residual block. We combine spatial attention [30] and AdaIN [8] idea to enhance spatial perception. Therefore, the content and color of the original image can be maintained.

The details of discriminator can be seen Fig. 5. We follow the idea of frequency separation [6]. There is a Guassian high-pass filter before several convolution which kernel size of 3, to extract the high frequency information. This design allows the discriminator $G_z(\cdot)$ to treat only the high-frequency part for real and fake image, making the training of the whole GAN more stable and fast convergent.

For each real-world HR image y, the result of bicubic down-scaling is x where $y \in \mathcal{Y}$, $x \in \mathcal{X}$. The $G_{x \to z}(x)$=$\hat{z}$ denotes real-world LR image where $\hat{z} \in \hat{\mathcal{Z}}$. The $\hat{z}$ and z will be fed into $D_z(\cdot)$ to distinguish fake or real.

Loss funcition. In order for the generator to do domain transfer effectively, we combine three losses, low frequency loss $\mathcal{L}_{low}$, perceptual loss $\mathcal{L}_{per}$ and high frequency loss $\mathcal{L}_{high}$. The low frequency loss is defined as Eq. 1.

$$\mathcal{L}_{low} = \frac{1}{n} \sum_{i=1}^{n} \|F_L(G_{x \to z}(x_i)) - F_L(x_i)\|_1, \quad (1)$$

where $F_L(\cdot)$ is a Guassian low-pass filter, n is the batchsize, $x_i \in \mathcal{X}$

For perceptual loss we use the pre-trained VGG16 network [25], it can be defined as Eq. 2

$$\mathcal{L}_{per} = \frac{1}{n} \sum_{i=1}^{n} \|F_{Vgg}(G_{x \to z}(x_i)) - F_{Vgg}(x_i)\|_2, \quad (2)$$

In order enchance the realistic of image, we use the LS-GAN [22] strategy. Thereby the high frequency loss $\mathcal{L}_{high}$ can be defined as Eq. 3.

$$\mathcal{L}_{high} = \frac{1}{n} \sum_{i=1}^{n} \|D_z(G_{x \to z}(x_i)) - 1\|_2, \quad (3)$$

Thus, the total loss of generator can be represented in Eq. 4.

$$\mathcal{L}_{TGtotal} = \lambda_{t1} * \mathcal{L}_{low} + \lambda_{t2} * \mathcal{L}_{per} + \lambda_{t3} * \mathcal{L}_{high}, \quad (4)$$

Finally, the loss of discrimnator is defined as Eq. 5.

$$\mathcal{L}_{TDtotal} = \frac{1}{2} \sum_{i=1}^{n} \|D_z(G_{x \to z}(x_i)) - 0\|_2 + \frac{1}{2} \sum_{i=1}^{n} \|D_z(z) - 1\|_2, \quad (5)$$

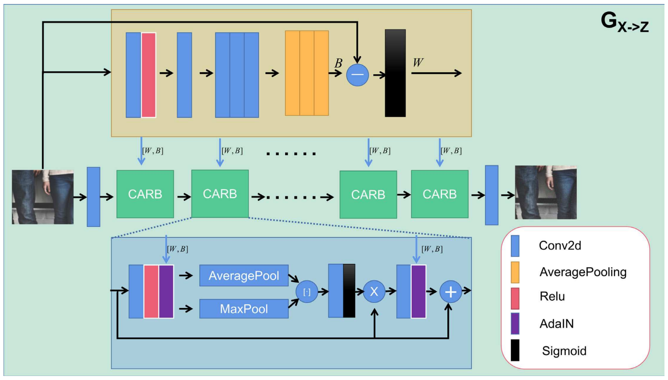

Figure 4. The details of generator

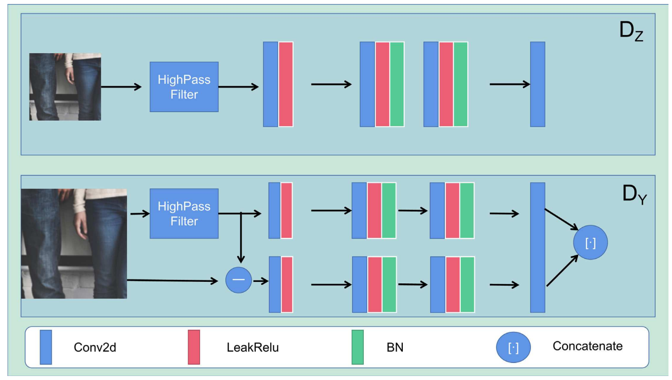

Figure 5. The details of discriminator

3.4. Supervised SR

Network architecture. After domain transformation, the generated image $\hat{z}$ and the y make up the image pairs for the training of supervised SR. As stated in section 3.2, we just to solve a mapping f_2 from $\hat{\mathcal{Z}}$ to $\mathcal{Y}$. In order to improve the subjective visual quality, we also use LSGAN [22]. On the one hand, the generator $G_{\hat{z}\to y}$ only consists of nine RRDBs [29] which is a network with less computation. On the other hand, as the generator usually yields fake texture, we let the discrimnator D_y contains of two networks, high frequency network and low frequency network so that not only keeps the high frequency features, but also maintains the low frequency features. The network can be seen Fig. 5.

Loss funcition. In training process of SR, we find the perceptual loss also bring slight color shift. Unlike [29], we remove the perception loss and add an Edgeloss which makes the color consistency to keep well. Therefore, the total loss of generator consists of content loss $\mathcal{L}_c$, EdgeLoss $\mathcal{L}_e$ and adversial loss $\mathcal{L}_{adv}$. For the content loss, it aims to maintain the content of the original image, which can be defined as Eq. 6.

$$\mathcal{L}_c = \frac{1}{n}\sum_{i=1}^{n}\|(G_{\hat{z}\to y}(\hat{z}_i)) - y_i)\|_1, \tag{6}$$

where n is the batchsize, $z_i \in \hat{\mathcal{Z}}$ generated by $G_{x\to z}(\cdot)$, $y_i \in \mathcal{Y}$.

For the EdgeLoss, we want to the training process focus on image edge details, which can enhance the visual quality effectively . In our solution, we utilize the Canny operator to extract the edge of y_i and $G_{x\to z}(\cdot)$. Thereby, the EdgeLoss $\mathcal{L}_e$ can be written as Eq. 7.

$$\mathcal{L}_e = \frac{1}{n}\sum_{i=1}^{n}\|F_E(G_{\hat{z}\to y}(\hat{z}_i)) - F_E(y_i)\|_2, \tag{7}$$

where F_E denotes Canny operator, n is the batchsize, $z_i \in \hat{\mathcal{Z}}$ is generated by $G_{x\to z}(\cdot)$, $y_i \in \mathcal{Y}$.

For the adversial loss, since we use LSGAN strategy, the adversial loss can be written as Eq. 8.

$$\mathcal{L}_{adv} = \frac{1}{n}\sum_{i=1}^{n}\|D_y(G_{\hat{z}\to y}(\hat{z}_i)) - 1\|_2, \tag{8}$$

Thus, the total loss of generator can be written as Eq. 9.

$$\mathcal{L}_{SGtotal} = \lambda_{s1} * \mathcal{L}_c + \lambda_{s2} * \mathcal{L}_e + \lambda_{s3} * \mathcal{L}_{adv}, \tag{9}$$

Likewise, the total loss of discrimnator can be written as Eq. 10.

$$\mathcal{L}_{SDtotal} = \frac{1}{2}\sum_{i=1}^{n}\|D_y(G_{\hat{z}\to y}(\hat{z}_i)) - 0\|_2 + \frac{1}{2}\sum_{i=1}^{n}\|D_z(y) - 1\|_2, \tag{10}$$

4. Results

4.1. Implements detail

The whole network including the unsupervised domain transformation network and the SR network is implemented using PyTorch 1.1. Firstly, we need to generate the unpair data to train domain transformation network, therefore the real-world HR image $y \in \mathcal{Y}$ is bicubic down-scaled with factor of 1/4 so that we can get x. Furthermore, we crop the x and z to 128×128 patches. The domain transformation network is trained with 300,000 iterations and the batch_size is 8. The optimizer is Adam [14] with $\beta_1 = 0.5, \beta_2 = 0.999$ and the initial learning rate is set 1e-4. Especially, we set $\lambda_{t1} = 1$, $\lambda_{t2} = 0.05$, $\lambda_{t3} = 0.05$ on a Nvidia RTX 2080 TI.

After training the domain transformation network, the image pairs can be generated. We crop the $\hat{z}$ and y to 120×120 and 480×480 patches so that can improve the speed of IO. In training process, we randomly crop the LR patched to 64 × 64. The SR network is also trained 300,000

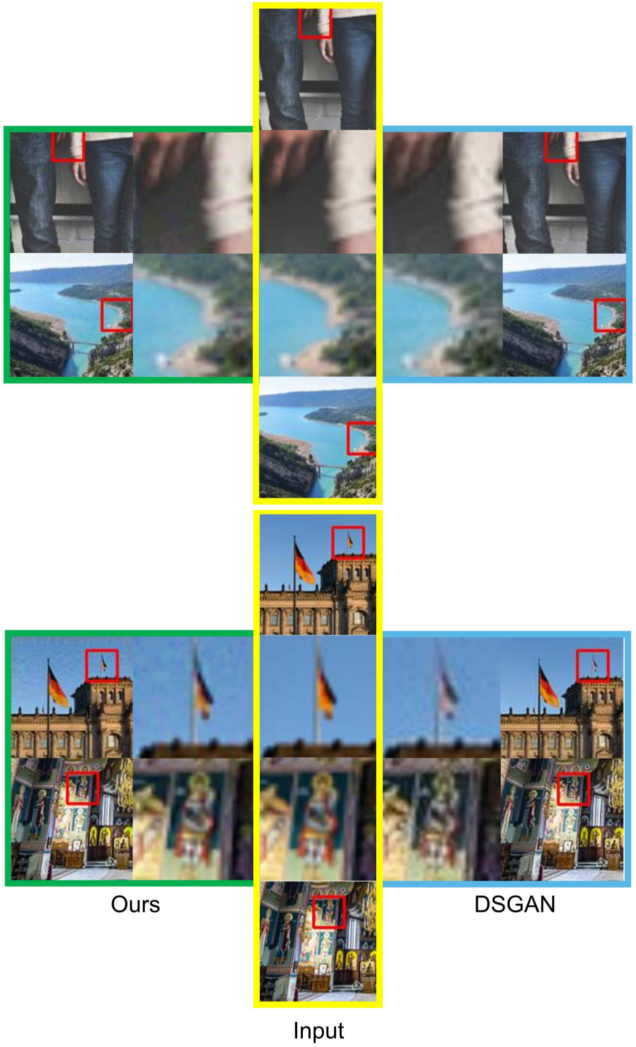

Figure 6. The qualitative results of domain transformation. The yellow box is the input data that will be transferred. The green box is generated by our method. The blue box is generated by DSGAN [6]. We can see there is color shift in DSGAN [6]. Zoom in for best view.

iterations and the batch_size is 12. The optimizer is Adam [14] with $\beta_1 = 0.9, \beta_2 = 0.999$ and the initial learning rate is set 2e-4. Especially, we set $\lambda_{s1} = 1$, $\lambda_{s2} = 0.1$, $\lambda_{s3} = 0.05$ on two Nvidia RTX 2080 TI.

4.2. Domain transformation results

To validate the proposed method effectively, we evaluated the performance of domain transformation network. We map from x to $\hat{z}$. Especially, we randomly sample 10 images from clean DF2K dataset [29] which is a mergence of DIV2K [27] and Flickr2K [27]. We down-scale ten images with factor of 1/4. After that, they will be fed into the generator $G_{x\rightarrow z}(\cdot)$. The final qualitative results can be seen in

Method	PSNR ↑	SSIM ↑	LPIPS ↓
ESRGAN	18.64	0.225	0.8174
SDSR	22.73	0.456	0.4384
TDSR	21.59	0.4083	0.4609
Our	29.76	0.8233	0.2764

Table 1. The quantitative results on the DF2K dataset. ↑ denotes the higher the more similar. Conversely, ↓ represents a lower and more similar.

Fig. 6. The results of DSGAN [6] is generated by official model. We can see that DSGAN has color shift for some images.

4.3. The comparison of super-resolved results

In this section, in order to compare the quantitative results with the state-of-the-art methods, we randomly sample 10 images from the DF2K dataset as the validation dataset. The degradation comes from artifacts. Table 1 shows the quantitative results of PSNR and SSIM metrics. Meanwhile, the LPIPS [31] also was reported to describe the perceptual quality. The visual qualitative results can be seen Fig. 7. All models for comparison are from the official pre-trained models.

As Fig. 7, many state-of-the-art algorithms have heavily artifacts, which also makes their quantitative indicators and our algorithms have a large gap.

4.4. NTIRE 2020 real-world super-resolution challenge

NTIRE 2020 real-world super-resolution challenge is divided into two tracks: one is to process the images with artifacts and the other is to process the smartphone images [21]. Those images should be super-resolved with an factor of 4 is the final goal. Our team participated two tracks. Both Track 1 and Track 2 will utilize Mean Opinion Score (MOS) as an evaluation metric.

Track 1. For Track 1, since the GT is unknown, we only provide the comparison of qualitative results with the state-of-the-art algorithms, which can be seen in Fig. 8.

Track 2. For Track 2, we also provide the comparison of qualitative results with the state-of-the-art algorithms, which can be seen in Fig. 9. The SR results using ESRGAN [29], SDSR [6] and TDSR [6] have noise or other unnecessary distortions for real-world images.

4.5. Ablation study

In this section, to evaluate the contribution of different modules in the systematic solution, we do corresponding experiments. All results are the same as the validation dataset used in the Section 4.3. We evaluate the method proposed by ourself on following different setting. The qualitative results can be seen Fig. 10. The quantitative results

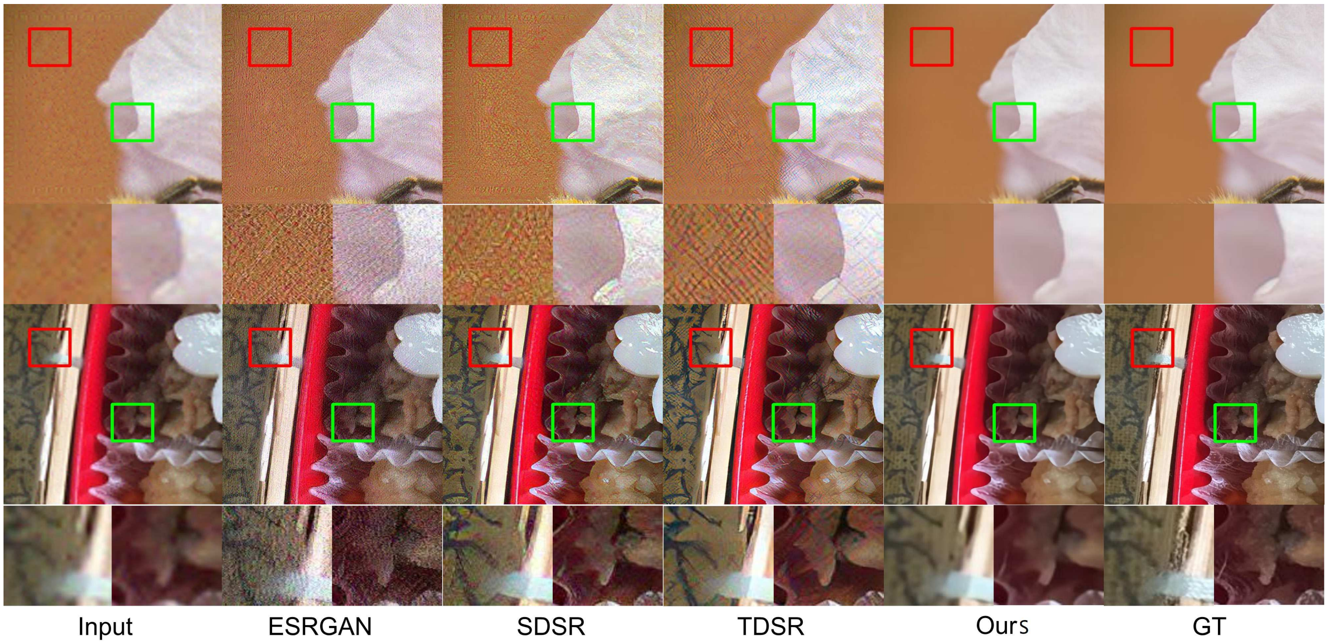

Figure 7. The qualitative results of different methods on DF2K dataset. We compared the state-of-the-art methods, ESRGAN [29], SDSR [6] and TDSR[6].

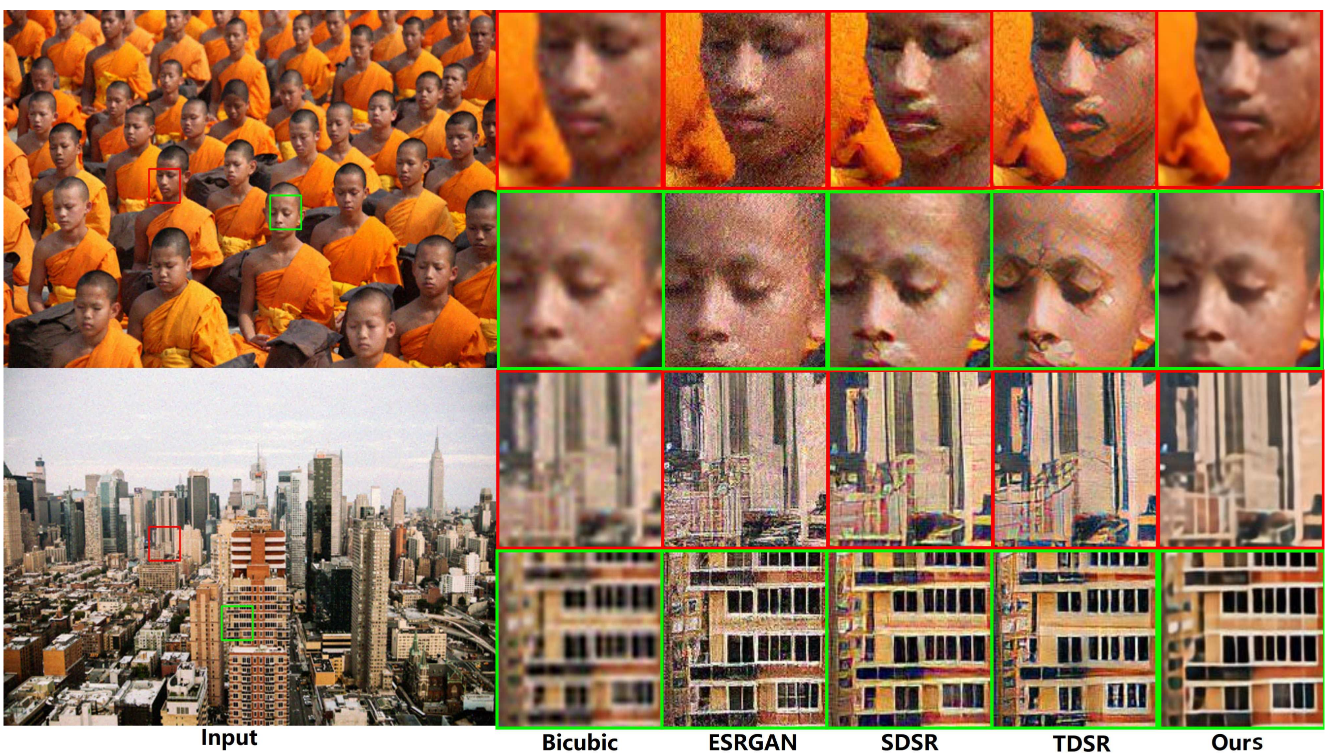

Figure 8. The qualitative results of Track 1. We compared the state-of-the-art methods, ESRGAN [29], SDSR [6] and TDSR[6]

can be seen Table. 10.

Bicubic and GAN: This is an standard method which LR comes from HR using bicubic down-scaling directly, meanwhile the SR process is trained by GAN. We can see

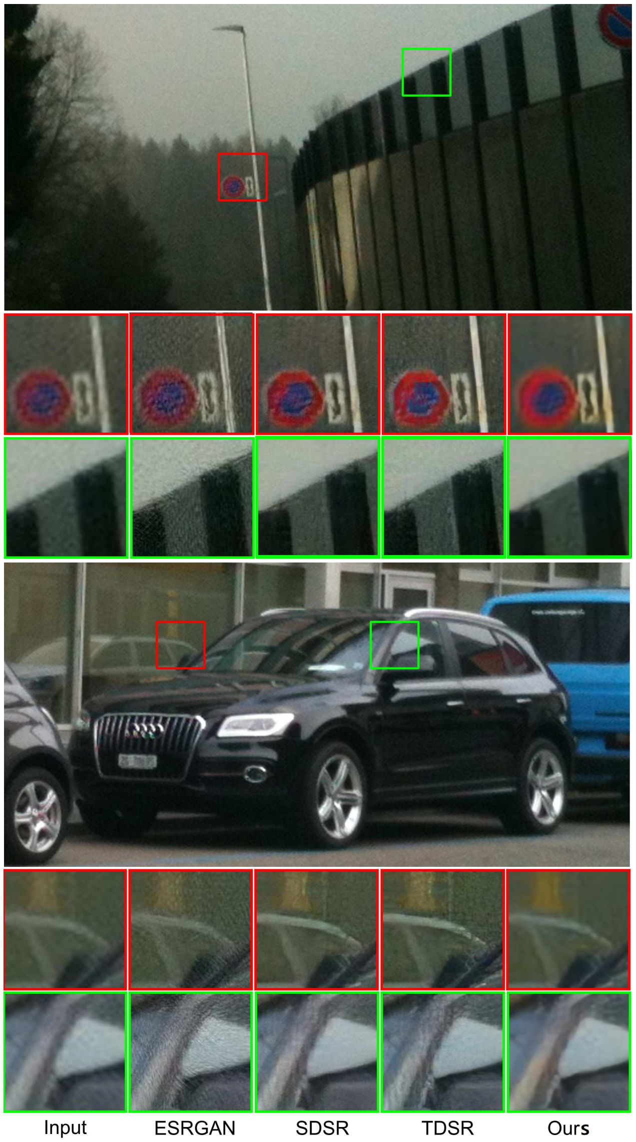

Figure 9. The qualitative results of Track 2. We compared the state-of-the-art methods, ESRGAN [29], SDSR [6] and TDSR[6].

the results are heavier distortion.

CARB and GAN: The SR training data is generated by our domain transformation network. It also uses the standard GAN. We find there are fake texture.

CARB and GAN-FS: The SR training data is generated by our domain transformation network, however the discrimnator uses the frequency separation idea. We can see that the fake texture phenomenon is alleviated, but some details is lost.

CARB, GAN-FS and EdgeLoss: The SR training data is generated by our domain transformation network, meanwhile the frequency separation idea and EdgeLoss is combined to enhance the details of edge. We find the results are

Method	PSNR ↑	SSIM ↑	LPIPS ↓
Bicubic_GAN	18.64	0.225	0.8174
CARB_GAN	20.69	0.444	0.3928
CARB_GAN-FS	26.44	0.5069	0.4024
CARB_GAN-FS_EdgeLoss	29.76	0.8233	0.2764

Table 2. This Table reports the quantitative results of SR for different setting. ↑ denotes the higher the more similar. Conversely, ↓ represents a lower and more similar. The specific settings can be seen in Section 4.5.

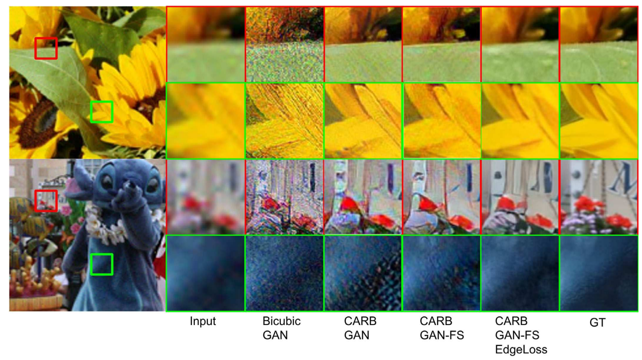

Figure 10. This figure reports the qualitative results of SR for different setting. The specific settings can be seen in Section 4.5.

more similar with GT.

5. Conclusion

In this paper, we proposed a CARB as the unit of domain transformation network, which can effectively to map images from a domain to another domain in characteristic and the content and color will be maintained. Furthermore, we modified the discrimnator of ESRGAN to distinguish high frequency and low frequency separation which aims to accelerate the convergence of training model and maintain both high frequency and low frequency features. Finally, the EdgeLoss was constructed to enhance the edge details. Our systematic solution achieved a significant improvement. We will investigate how to further improve the realness of the images in future work.

Acknowledgments. This work was supported by National Natural Science Foundation of China under Grant 61901120 and Grant 61802065, in part by the Science and Technology Program of Fujian Province of China under Grant 2019YZ016006.

References

[1] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. Fast, accurate, and lightweight super-resolution with cascading residual network. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 252–268, 2018.

[2] John Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6):679–698, 1986.

[3] Hong Chang, Dit-Yan Yeung, and Yimin Xiong. Super-resolution through neighbor embedding. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 1, pages I–I. IEEE, 2004.

[4] 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.

[5] 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.

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

[7] Daniel Glasner, Shai Bagon, and Michal Irani. Super-resolution from a single image. In *2009 IEEE 12th international conference on computer vision*, pages 349–356. IEEE, 2009.

[8] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1501–1510, 2017.

[9] Yawen Huang, Ling Shao, and Alejandro F Frangi. Simultaneous super-resolution and cross-modality synthesis of 3d medical images using weakly-supervised joint convolutional sparse coding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6070–6079, 2017.

[10] Mehdi Mirza Bing Xu David Warde-Farley Sherjil Ozair Aaron Courville Ian Goodfellow, Jean Pouget-Abadie and Yoshua Bengio. Generative adversarial nets. In *In Advances in neural information processing systems*, page 26722680, 2014.

[11] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. Dslr-quality photos on mobile devices with deep convolutional networks. In *Proceedings of the 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 *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016.

[13] Kwang In Kim and Younghee Kwon. Single-image super-resolution using sparse regression and natural image prior. *IEEE transactions on pattern analysis and machine intelligence*, 32(6):1127–1133, 2010.

[14] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[15] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 624–632, 2017.

[16] 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.

[17] 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.

[18] 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.

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

[20] Andreas Lugmayr, Martin Danelljan, Radu Timofte, et al. Aim 2019 challenge on real-world image super-resolution: Methods and results. In *ICCV Workshops*, 2019.

[21] Andreas Lugmayr, Martin Danelljan, Radu Timofte, et al. Ntire 2020 challenge on real-world image super-resolution: Methods and results. *CVPR Workshops*, 2020.

[22] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2794–2802, 2017.

[23] Pejman Rasti, Tonis Uiboupin, Sergio Escalera, and Gholamreza Anbarjafari. Convolutional neural network super resolution for face recognition in surveillance monitoring. In *International conference on articulated motion and deformable objects*, pages 175–184. Springer, 2016.

[24] Mehdi SM Sajjadi, Bernhard Scholkopf, and Michael Hirsch. Enhancenet: Single image super-resolution through automated texture synthesis. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4491–4500, 2017.

[25] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[26] Jian Sun, Zongben Xu, and Heung-Yeung Shum. Image super-resolution using gradient profile prior. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.

[27] Radu Timofte, Eirikur Agustsson, Luc Van Gool, Ming-Hsuan Yang, and Lei Zhang. Ntire 2017 challenge on single image super-resolution: Methods and results. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 114–125, 2017.

[28] Tong Tong, Gen Li, Xiejie Liu, and Qinquan Gao. Image super-resolution using dense skip connections. In *Proceed-*

ings of the IEEE International Conference on Computer Vision, pages 4799–4807, 2017.

[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 *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 0–0, 2018.

[30] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 3–19, 2018.

[31] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018.

[32] 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.