symmetry

MDPI

Article

Generative Adversarial Network-Based Super-Resolution Considering Quantitative and Perceptual Quality

Can Li [1], Liejun Wang [1,*], Shuli Cheng [1] and Naixiang Ao [2]

[1] College of Information Science and Engineering, Xinjiang University, Urumqi 830046, China; lican_0301@163.com (C.L.); cslxjuedu@126.com (S.C.)
[2] Xinjiang Lianhai INA-INT Information Technology Ltd., Urumqi 830000, China; aonaixiang@xjlhcz.com
* Correspondence: Correspondence: wljxju@xju.edu.cn; Tel.: +86-139-9981-6618

Received: 21 February 2020; Accepted: 4 March 2020; Published: 11 March 2020

Abstract: In recent years, the common algorithms for image super-resolution based on deep learning have been increasingly successful, but there is still a large gap between the results generated by each algorithm and the ground-truth. Even some algorithms that are dedicated to image perception produce more textures that do not exist in the original image, and these artefacts also affect the visual perceptual quality of the image. We believe that in the existing perceptual-based image super-resolution algorithm, it is necessary to consider Super-Resolution (SR) image quality, which can restore the important structural parts of the original picture. This paper mainly improves the Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) algorithm in the following aspects: adding a shallow network structure, adding the dual attention mechanism in the generator and the discriminator, including the second-order channel mechanism and spatial attention mechanism and optimizing perceptual loss by adding second-order covariance normalization at the end of feature extractor. The results of this paper ensure image perceptual quality while reducing image distortion and artefacts, improving the perceived similarity of images and making the images more in line with human visual perception.

Keywords: super-resolution; generative adversarial networks; attention mechanism; shallow network

1. Introduction

Image super-resolution reconstruction converts low-resolution images into high-resolution images to obtain images as close as possible to real images. After Dong et al. pioneered SRCNN [1], image super-resolution reconstruction algorithms based on neural networks emerged in an endless stream and achieved remarkable results. These algorithms include the image distortion-driven algorithms (i.e., Peak-Signal-to-Noise Ratio (PSNR) value) [1–6], and perception-driven image super-resolution algorithms [7–12], The distortion-based network structure makes the restored image too smooth, losing considerable high-frequency information and texture information, which does not correspond with human visual perception [9]. Subsequent image super-resolution reconstruction algorithms based on the generative adversarial network (GAN) [9,10,13–15], improve the visual perception of the image while producing some unpleasant artefacts which reduce the image quality. To some extent, these artefacts also affect the visual perceptual quality of the image. Figure 1 shows some of the experimental results. It can be seen that the experimental results in this paper are closer to the structure of the original picture, reduce the generation of other textures, and correspond better with human visual perception.

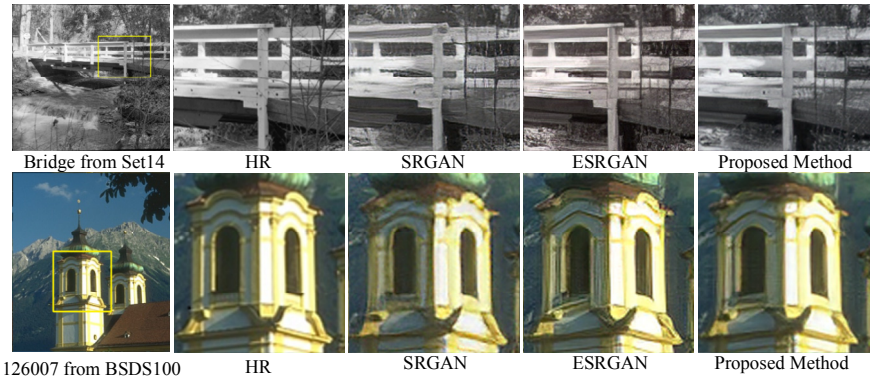


Figure 1. Super-resolution reconstruction is magnified four times. We compare our results with Super-Resolution Generative Adversarial Networks (SRGAN) and Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN). Our results are closer to the original High-Resolution (HR) image, reducing artefacts.

Recently, [16] proposed that the perceptual quality and the degree of image distortion do not completely correspond. There is an unrealizable region between distortion and perception, which makes the perception and distortion reach the optimal value, indicating that the perception and distortion may always be contradictory. There are a number of algorithms dedicated to weighing perceptual quality and distortion [8,13,17]. We summarize some super-resolution reconstruction algorithms and find that researchers mainly improve these through three aspects: network structure, loss function and algorithm for trade-off and conversion between perception and distortion. In terms of network structure, researchers initially considered widening and deepening the network structure based on the residual network or the dense block [4] and later introduced the GAN [18]. Zhang et al. designed the residual dense block [19] considering the advantages and disadvantages of residual block and dense block. In addition, Wang et al. proposed a priori algorithm for image super-resolution. The author proposed adding semantic prior conditions to improve the perceptual quality of SR images [12]. Pan et al. proposed the dual CNN [20] to establish a dual path CNN network to optimize the structure and detail in the image. Pan's proposed physical GAN [21] believes that when the SR picture is close to the HR picture, the statistical characteristics of the corresponding low-resolution images should be the same. EUSR [7] proposed a new enhanced upscale module. In the process of training, considering both image quality and perceptual quality, the weight of perception and distortion in the image is changed by changing the weight between each loss function. Recently, researchers have found that the dependence of the middle feature layer should be considered instead of simply widening or deepening the network. Therefore, the attention mechanism is introduced into image super-resolution reconstruction [22–24], which makes the network focus more on the informative part of the feature map and improves the expressive ability of the network. SOCA [23] proposed a second-order attention mechanism. The author believed that higher-order statistical features should be considered. The results are excellent in both quantitative matrix and visual quality. For the loss function, perceptual loss [25] is proposed, which enables the network to minimize the gap between SR and HR in feature space, rather than in pixel space, which can improve the perceptual quality of generated pictures. Merchez et al. proposed contextual loss to maintain natural image statistics [11,26]. In addition, the proposed discriminative losses [9,18] can also improve the perceptual quality of SR images. A relative discriminator [27] has also been used in image super-resolution [14,28]. Vu et al. [28] enhanced RaGAN by wrapping the focal loss. EUSR [7] proposed minimizing the discrete cosine transform (DCT) coefficient between SR and HR images so that the two pictures are consistent in the frequency domain.

However, there is still a large gap between real images and predicted images. Distortion-based images result in too smooth images, while perception-based super-resolution algorithms tend to result in the over-distortion of SR images for parts of complex textures. Looking back on the previous works,

we find that the network always tends to extract the deep features, regardless of the generator or the discriminator, while ignoring the low-level features, which means that only the high-level features are approximated, and there is still a large gap between SR and HR in the low-level features. In addition, this paper also considers that dependence between intermediate feature layers should be considered in both the generator and discriminator to improve the expressive ability of the network, and more focus should be on the informative part of the feature map. In addition, the previous perceptual loss knowledge is optimized in the first-order feature statistics, and the optimization of the feature statistics above the first order is not considered. To solve these problems, this paper proposes a shallow generator and a shallow discriminator. A shallow generative network generates low-level features. The shallow discriminative network minimizes the statistical gap between SR and HR images in the low-level features. In addition, dual attention (DUA) is added to both the generator and discriminator, including a spatial attention mechanism and channel attention mechanism, to improve the expressive ability of the network. Specifically, by referring to [23], second-order covariance pooling is used in channel attention mechanisms to extract higher-order statistical features. Two attention mechanisms are added to the discriminator to extract higher-order statistical features. The discriminator can more accurately impose more complex geometric features on the global image structure and constrain the generation of SR images generated by the network. At the same time, this paper improves the perceptual loss and calculates the loss function on the higher level of image features. The covariance normalization [29,30] is introduced into the feature extraction layer to minimize the difference between SR and HR in higher-order statistical features. In summary, the main contributions of this article are as follows:

1. To make full use of the original low-resolution images, we should not only narrow the gap between SR and HR at high-levels but also narrow the gap between low-levels. A shallow generator and a shallow discriminator are added to obtain a closer picture of the original real image.
2. Considering the dependencies between feature maps, we introduce a second-order channel attention mechanism and self-attention mechanism on the generator and the discriminator, so that the network focuses on more informative parts and improves the network's expressive ability and discriminative ability, which more accurately restrain pictures generated by the generation network.
3. For perceptual loss, we introduce covariance normalization in the feature extraction layer so that the perceptual loss can improve the perceptual quality of SR pictures from higher-order statistical features for more discriminative representations.
4. We improve the perceptual quality of the image while considering the distortion of the image, making the generated SR image more suitable for human visual perception.

The latter part of this paper consists of the following aspects. The second part introduces the related work of this paper. The third part introduces the algorithms mentioned in this paper, including the specific network structure, attention mechanism, and second-order covariance pooling. In the fourth part, the experimental results are presented and compared with other popular perception-based image super-resolution algorithms. Finally, the research results are summarized.

2. Related Work

In the last decade, with the rapid development of deep learning, deep learning has been widely used in all aspects of visual images [31–33]. Image super-resolution reconstruction based on deep learning has developed rapidly. Most of the algorithms based on deep learning mainly improve network structure [8,10,12,19] and loss function [9,11,25,26,28,34]. In addition, to explore the dependencies between feature maps, many algorithms introduce attention mechanisms [22–24]. The work related to this article will be introduced from the above aspects.

2.1. Network Structure

Because of the powerful expression ability of CNNs, image super-resolution reconstruction algorithms based on CNNs have rapidly increased. Various kinds of deformation based on the CNN network have also emerged. These include distortion-based EDSR [35], which laid the foundation for many subsequent network structures based on distortion algorithms. SRGAN [9] was the first to introduce the generative adversarial network into image super-resolution reconstruction. ESRGAN [14] replaced the residual block in SRGAN with a residual in residual dense block (RRDB), which enables the network to accommodate more residual networks and dense connections. Zhang et al. proposed RankSRGAN [15] by ranking perceptual index (NIQE or PI). In addition, Pan's dual CNN [20] optimizes the image from two aspects: details and structure. Two convolutional networks are established for the details and structure of the image. Detailed loss and structured loss are used to optimize the details and structure of the SR image, respectively. The Pan's proposed dual CNN [20] network provided great inspiration for this article. A double-layer CNN network [36] is also proposed, which comprises a deep network and a shallow network. The author believes that the training of a deep network and shallow network can accelerate the convergence speed and improve the quality of the generated SR image.

In this paper, a shallow generator is added to the generator to extract the low-level features of the image so that the whole network can not only approximate the original image in the high-level feature but also approximate the original image in the low-level feature. The shallow discriminator is also added to the original discriminator so that the discriminator can minimize the statistical characteristic difference in the image from the high-level and low-level features and make the final SR image closer to the original image. In addition, this paper does not add the results produced by the deep network and the results produced by the shallow network directly, as in [36], but adds the feature maps of the shallow network and the deep network after upsampling, obtains the final feature maps, and enters the subsequent feature mapping network (which is explained in detail in the following sections).

2.2. Loss Function

The loss function, including perceptual loss and discriminative loss, is a very important factor in the neural network. Perceptual loss [25] is proposed to minimize the loss of feature space between HR and SR to improve image perceptual quality. SRGAN [9] uses discriminative loss to learn the statistical properties of natural images in discriminative training to produce more realistic SR images. The relativistic discriminator proposed by the relativistic GAN [27,28] shows that the effect of the relativistic discriminator can be improved by wrapping a focal loss [34]. NatSR [37] remodeled super-resolution, improved the network's discriminator, and improved the perceptual quality of SR images. The proposed second-order statistical feature can improve the discriminative representation of the network. Mechrez proposed that contextual loss [11,26] can make the SR image have the same statistical properties as the HR image.

In this article, a dual discriminative network is used, including the deep discriminator and shallow discriminator. The relativistic GAN is used in both discriminators. In the perceptual loss, we introduce covariance normalization in the feature extraction layer, which enables the network to minimize the feature difference between the HR image and the SR image from higher-order statistical features to achieve better discriminative representation and perceptual quality.

2.3. Attention Mechanism

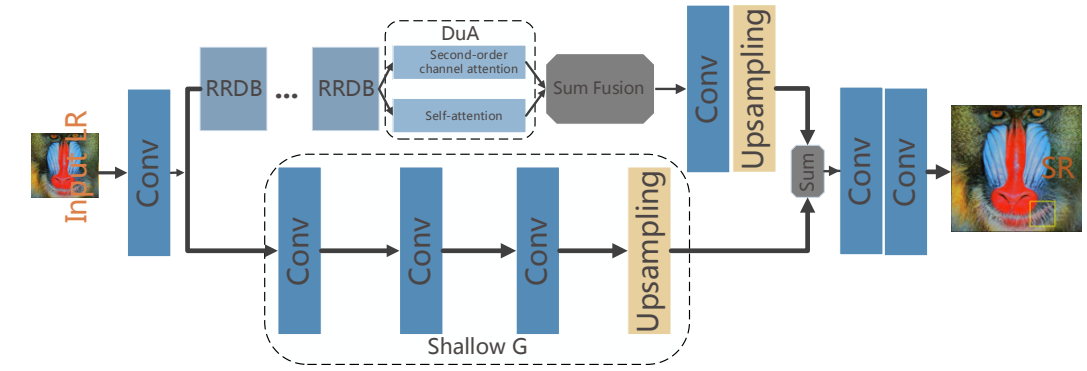
Attention mechanism: The attention mechanism is derived from the study of human vision. In essence, the attention mechanism in deep learning is similar to the human selective attention mechanism. The image attention mechanism model can avoid or ignore the noise portion in the images and improve the signal-to-noise ratio of the restored image. The attention mechanism proposed in SENet [38] is used in image super-resolution reconstruction by RCAN [22]. CBAM [24] inputs the spatial attention mechanism [39] into the network structure in series with the channel attention

mechanism. The channel attention mechanism and the spatial attention mechanism are adopted in SOCA [23]. The channel attention mechanism is a second-order channel attention mechanism, which can adaptively adjust the channel inter-dependencies in higher-order statistical features. However, in most articles, the attention mechanism is mostly added to the generator, ignoring the need for the expressive quality of the discriminator.

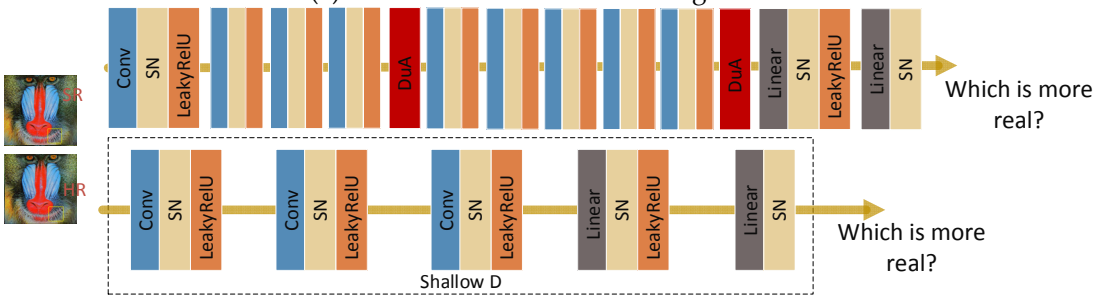
In this paper, the second-order channel attention mechanism and spatial attention mechanism are added to the generator and the discriminator to make the generator focus more on the informative part and improve the image signal-to-noise ratio, in the discriminator, more complex geometric features can be applied to the global image structure more accurately through the constraints of two attention mechanisms.

3. Methods

The main purpose of this paper is to make the SR image closer to the original image, ensure the image quality, reduce the artifacts generated by the GAN network, and thus closer to the human visual perception. This chapter introduces the network structure, loss function and attention mechanism of this paper. In the network structure, this paper refers to the network framework in ESRGAN. A shallow generator and a shallow discriminator are added to ESRGAN. DUA blocks are designed to be placed in the upper network of the generator, and two DUA blocks are placed in the discriminator (as shown in Figure 2) to restore a more realistic image texture. In addition, the perceptual loss is improved by adding covariance normalization at the end of the feature extractor.



(**a**) The network structure of the generator



(**b**) The network structure of the discriminator

Figure 2. Network structure: the upper part represents the generator, and the lower part represents the discriminator. We added a shallow generator and a shallow discriminator to the generator and the discriminator, respectively. At the same time, the spatial attention mechanism and the second-order channel attention mechanism (the position indicated by dual attention (DUA) in the figure) are added to the generator and the discriminator. The two attention mechanisms are merged in parallel. In the discriminator, we used a red rectangle to represent the DUA block.

3.1. Generator

After passing through the first convolutional layer, LR images enter the high-level feature extraction network and the low-level feature extraction network to extract high-level and low-level features, respectively. This article uses a convolutional layer as a shallow feature extractor

$$F_0 = H_{SF}(I_{LR}) \tag{1}$$

where H_{SF} denotes the first convolutional layer, I_{LR} denotes the low-resolution picture as input, and F_0 as the shallow feature layer enters the subsequent high-level feature extraction network and the low-level feature extraction network

$$F_{HF} = H_{HL}(F_0) \tag{2}$$

$$F_{SF} = H_{LL}(F_0) \tag{3}$$

H_{HL} denotes a deep feature extractor that extracts high-level features, and H_{LL} denotes a shallow feature extractor that extracts shallow-level features. H_{HL} includes RRDBs, DUA attention mechanism and upsampling layer. The DUA will be elaborated in the following sections. H_{LL} includes three convolution layers and an upsampling layer. The output of the high-level network structure and the shallow-level network structure to undergo feature fusion.

$$F_{TF} = F_{HF} + F_{SF} \tag{4}$$

F_{TF} denotes a total feature after fusion, F_{HF} and F_{SF} are feature-fused by element-by-element addition, the fused feature enters the final feature mapping layer.

$$I_{SR} = H_{MF}(F_{TF}) \tag{5}$$

H_{MF} denotes the last feature mapping layer and can also be regarded as a feature reconstruction layer.

3.2. Discriminator

In this paper, the relative GAN (RaGAN) [27] is used. The author of RaGAN believes that the probability of real data being true should be reduced, while increasing the probability that false data seem to be true. In the adversarial loss of the generator, not only the fake data but also the real data are involved, which can explain the a priori condition that half of the data in the dataset of the incoming discriminator are false. This article uses the relativistic average discriminator to replace the standard GAN [18]. The expression for the relative average discriminator is

$$D_{Ra}(x_r, x_f) = \sigma(C(x_r) - E_{x_f}[C(x_f)]) \tag{6}$$

x_r and x_f represent real data (HR images) and false data (SR images), respectively, σ is a sigmoid function, and $C()$ is the output of the non-transformation discriminator. E_{x_f} is the average of all false data in a mini batch. The final adversarial loss is defined as

$$l_D^{Ra} = -E_{x_r}[\log(D_{Ra}(x_r, x_f))] - E_{x_f}[\log(1 - D_{Ra}(x_f, x_r))] \tag{7}$$

The adversarial loss for the generator is defined as

$$l_G^{Ra} = -E_{x_r}[\log(1 - D_{Ra}(x_r, x_f))] - E_{x_f}[\log(D_{Ra}(x_r, x_f))] \tag{8}$$

The above is the relative discriminator we use, and in this paper, we use the deep discriminator D_{Ra}^D, and the shallow discriminator D_{Ra}^S. The deep discriminator uses nine convolutional layers, each

followed by SN spectral normalization [40], followed by the leakyReLU activation function, and finally the fully connected layer.

Three convolutional layers are used in the shallow discriminator to obtain low-level features of the SR and HR pictures

$$D_{Ra}^{D} = \sigma(C_{DD}(x_r) - E_{x_f}[C_{DD}(x_f)]) \tag{9}$$

$$D_{Ra}^{S} = \sigma(C_{SD}(x_r) - E_{x_f}[C_{SD}(x_f)]) \tag{10}$$

where C_{DD} and C_{SD} represent the output of the deep discriminative network and the shallow discriminative network, respectively. D_{Ra}^{D} and D_{Ra}^{S} represent the deep discriminator and the shallow discriminator, respectively.

The adversarial loss of the deep and shallow discriminator is defined respectively as

$$l_{D\_D}^{Ra} = -E_{x_r}[\log(D_{Ra}^{D}(x_r, x_f))] - E_{x_f}[\log(1 - D_{Ra}^{D}(x_f, x_r))] \tag{11}$$

$$l_{D\_S}^{Ra} = -E_{x_r}[\log(D_{Ra}^{S}(x_r, x_f))] - E_{x_f}[\log(1 - D_{Ra}^{S}(x_f, x_r))] \tag{12}$$

The final loss function of the discriminator is defined as

$$l_{D}^{Ra} = (l_{D\_D}^{Ra} + l_{D\_S}^{Ra})/2 \tag{13}$$

Similarly, the generator's loss function is also composed of deep and shallow adversarial loss:

$$l_{G\_D}^{Ra} = -E_{x_r}[\log(1 - D_{Ra}(x_r, x_f))] - E_{x_f}[\log(D_{Ra}(x_r, x_f))] \tag{14}$$

$$l_{G\_S}^{Ra} = -E_{x_r}[\log(1 - D_{Ra}(x_r, x_f))] - E_{x_f}[\log(D_{Ra}(x_r, x_f))] \tag{15}$$

The ultimate generator loss function is still

$$l_{G}^{Ra} = (l_{G\_D}^{Ra} + l_{G\_S}^{Ra})/2 \tag{16}$$

3.3. Perceptual Loss

This paper optimizes the perceptual loss by adding a covariance normalization to the last layer of the feature extractor and constrains SR images from higher-order statistical features. The papers [29,30] shows that the second-order statistical features can increase the expressive ability and discriminative ability of the network and can specify the shape of the feature distribution. For the input image, the covariance normalization generates a normalized covariance matrix as a representation, which characterizes the correlation of the feature channels and specifies the shape of the feature distribution. Considering these advantages, this paper applies covariance normalization to the high-order feature extractors.

Covariance normalization:

For a feature map with dimensions $H \times W \times C$, $F = [f_1, \ldots, f_c]$, there are a total of C channels, each of which has a scale of $H \times W$. We reshape the feature map F into a feature matrix of $C \times S$, where $S = H \times W$. Thus, the feature matrix covariance matrix is

$$\Sigma = X\bar{I}X^T \tag{17}$$

where $\bar{I} = \frac{1}{S}(I - \frac{1}{S}1)$, I and 1 represent the unit matrix of the S×S dimension and the matrix of total 1, respectively, and the T denotes the matrix transpose.

After calculating the covariance matrix, we normalize the covariance. The covariance matrix is a symmetric semi-definite matrix, which allows eigenvalue decomposition

$$\Sigma = U\Lambda U^T \tag{18}$$

where U is an orthogonal matrix and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_C)$ is a symmetric matrix of non-incremental eigenvalues. The energy of the covariance matrix can be converted into the energy of the eigenvalue

$$\hat{Y} = \Sigma^{\alpha} = U\Lambda^{\alpha}U^{T} \tag{19}$$

where α is a positive real value, $\Lambda^{\alpha} = \mathrm{diag}({\lambda_1}^{\alpha}, \ldots, {\lambda_C}^{\alpha})$

When $\alpha = 1$, there is no normalization; when $\alpha < 1$, the nonlinear shrinkage is greater than the eigenvalue of 1.0, and the stretching is less than the eigenvalue of 1.0. Referring to the paper [29], we set $\alpha = 0.5$ to obtain more discriminative representations.

In this paper, a VGG network is used as a feature extractor, and covariance normalization is added at the end of the feature extractor to minimize the perceptual loss of SR images from a higher-order feature level. In summary, the perceptual loss of the paper is

$$L_{\text{Perceptual}} = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} \left(Cov(I^{HR}) - Cov(G(I^{LR})) \right) \tag{20}$$

where I^{HR} and I^{LR} represent the high-resolution images and low-resolution images, respectively. *Cov()* indicates the feature maps through the covariance normalization after feature extract layer. In order to highlight the role of covariance normalization, we use the abbreviation of "covariance" as the symbol. *G()* represents the generator operator.

Finally, the loss function for the generator is

$$L_G = L_{\text{Perceptual}} + \beta L_G^{\text{Ra}} + \gamma L_1 \tag{21}$$

where $L_1 = E_{x_i} \left\| G(x_i) - y \right\|_1$ is the 1-norm distance between the SR images $G(x_i)$ and the HR images y. β and γ represent the weights between the loss functions.

3.4. Attention Mechanism

3.4.1. Channel Attention Mechanism

In order to obtain better expression ability, we introduce the second-order channel attention mechanism in both generator and discriminator, for the feature map of H×W×C, after covariance normalization mentioned in the previous section, the feature map of C×C is obtained, denoted as $\hat{f}$. let $\hat{f} = [f_1, \ldots, f_C]$, where $f_i \in R^{C \times 1}$. By compressing each element of $\hat{f}$, We can get the c-dimensional vector

$$\hat{f_C} = [z_1, \ldots, z_C] \in R^{C \times 1} \tag{22}$$

where

$$z_i = H_{CVP}(\hat{f}) = \frac{1}{C} \sum_{j=1}^{C} f_i(j) \tag{23}$$

where $H_{CVP}()$ represents the covariance pooling; i denotes the i^{th} channel. We use covariance pooling instead of the original global average pooling. Compared to the traditional first-order global average pooling, the covariance pooling can explore the feature distribution to capture higher-order statistics than the first-order global average pooling, so as to obtain more discriminative representations. Similar to SENet [38], we introduce a gate mechanism to send the resulting $\hat{f_{cop}} \in R^{C \times 1}$ to the subsequent activation function, as shown in Figure 3.

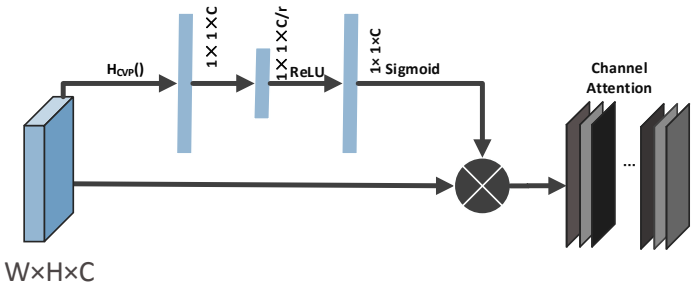


Figure 3. Of the channel attention mechanism.

After obtaining $\hat{f_{cop}}$, the two connected dimensions are the fully connected layers of C/r and C, respectively, and the two fully connected layers are followed by the ReLU activation function and the sigmoid activation function, respectively. Finally, the weight of C is $w_c \in R^{C\times 1}$.

Then, the final feature is

$$f_{sa} = f \cdot w_c \tag{24}$$

where f represents the feature map before the second-order channel attention mechanism, and f_{sa} represents the feature map weighted by the channel attention mechanism, which shows that the channel attention mechanism can adaptively adjust the dependency between the feature maps.

3.4.2. Self-Attention

In SAGAN [41], the self-attention mechanism[42] is a complement to convolution, which helps to model the long-term, multi-level dependence between image regions. With the self-attention mechanism, when the generator generates an image, each position in the image is carefully coordinated with the distant details of the image. In addition, the discriminator can more accurately apply complex geometric features to the global image structure. This paper adopts the same attention mechanism framework as SAGAN.

In this article, the feature maps of the second-order attention mechanism and spatial attention mechanism are fused into generators and discriminators in an element-by-element way. We did not choose cascading to save GPU memory. In addition, considering that feature maps depend on low-level features, the details extracted from the underlying network are more from the bottom, while the high-level network can extract the global semantic information. Therefore, attention is applied to high-level features, because high-level features can provide a large enough perceptual field, and the data in a channel are sufficient to represent the global features. Therefore, as shown in Figure 2, the DUA block is placed in the upper part of the network structure.

4. Experience

Like other articles [7,9], this article magnifies a low-resolution image four times, where the low-resolution image is obtained by bicubic downsampling four times. The high-resolution images are cropped to a 128×128 image with the batch size set to 16. In the loss function, α is equal to 0.01 and β is equal to 5×10^{-3}. There are 23 RRDBs in the generator and one DUA is added. The discriminator uses the VGG19 network and uses two linear operations. This paper trains a total of 50×10^4 times. This paper uses the parameters of ESRGAN as the pretraining model, and we believe that the algorithm currently achieves the highest perceptual quality.

4.1. Data

During the training, the training dataset is the DIV2K dataset, which has 800 high-resolution pictures. In this paper, the training set is expanded by horizontal flipping and 90-degree rotation. During testing, set5, set14, PIRM verification set and BSD100 are used as test sets.

4.2. Evaluation Methods

This paper uses PSNR and Structure Similarity Index Measurement (SSIM) values to estimate image distortion and perceptual index (PI) and Root Mean Square Error (RMSE) to estimate perception-distortion trade-off. PI was evaluated by the non-reference of Ma's score [43] and NIQE [44], PI = 1/2 ((10-Ma) + NIQE). The lower the perceptual index, the better the perceptual quality of the image. The higher the PSNR and SSIM, the less noise the image has. In addition, this article also refers to learned perceptual image patch similarity (LPIPS)[45,46], which we describe as Perceptual Similarity. In this paper, CNN features are used to represent the visual perception of images. The author introduces a large-scale perceptual similarity dataset determined by 484 k people and uses the dataset to train existing VGG networks and Alex networks. LPIPS perceptual similarity is trained according to the person's score; the lower the perceived similarity value, the higher the perceptual quality of the image.

4.3. Experimental Results

This paper compares the experimental results with the current popular perceptual-based image super-resolution reconstruction algorithms: CX [26], EnhanceNet [47], SRGAN [9], EPSR3 [13], EUSR_pirm [7], ESRGAN [14], RankSRGAN [15]. SRGAN first introduces the GAN network into image resolution reconstruction and proposes that the traditional algorithm improves the image distortion too much and reduces the image perceptual quality. RankSRGAN uses indicators such as PI, NIQE, and Ma's score, that are more consistent with human visual perception as optimization goals. ESRGAN improves SRGAN in three aspects: network structure, perceptual loss and adversarial loss. CX adds contextual loss to SRGAN. Among these algorithms, we adopt algorithms with higher perceptual quality, such as EPSR3 and EUSR_pirm. These algorithms are evaluated on SET5, SET14, BSDS100 and PIRM_val datasets. The results are shown in Figure 4. The image restored by the perception-based super-resolution reconstruction algorithm has more or less CNN-based distortion, noise and some artefacts, which also affect human visual perception. Therefore, even though the optimal PI algorithm still does not correspond with human visual perception, people do not want to see artefacts and noise. For example, in the beard part of SET14 Baboon, the CX algorithm has many artefacts. EnhanceNet and SRGAN, EPSR3, etc., are not clear in the recovery of the beard part. The beard part of ESRGAN seems to have recovered the texture details of the beard in general, but a closer look reveals that there are some beard textures in the recovered image that are not found in the original image; our algorithm can reduce these artefacts. On the coastguard of Set14 in Figure 4, the SR images in EnhanceNet, SRGAN and CX are deformed. In ESRGAN, noise and artefacts are generated at the texture of the water waves and at the junction of people and water waves. In EPSR3, the images are too smooth, and some high-frequency details are lost. The same conclusion can also be obtained by observing the algorithm comparison of other pictures. In BSDS's 126007, EnhanceNet, SRGAN, EUSR, EPSR3, CX, RankSRGAN produce deformation and distortion in the wall part of the building, and ESRGAN produced distortion in the window part of the building. The algorithm of this paper reduces the deformation and distortion of the building part of the figure and improves the image structure. In addition, according to the objective evaluation index below each image, the results of this paper achieve the best perceptual similarity. While guaranteeing high image perceptual quality, our algorithm also reduces image distortion, improves PSNR, and reduces the presence of SR image artefacts, which is equivalent to improving image visual perceptual quality from another aspect. In BSDS100's 304034, as shown in Figure 4, the grass and branches have a relatively complex texture. In the lower PSNR results of SRGAN and CX, ESRGAN and SRGAN, the image produces distortion and noise, while in the higher PSNR algorithm, EUSR can still see deformation. In EPSR2 and RankSRGAN, the SR picture is too smooth and loses a lot of detail. The results of this paper ensure that high PSNR can still clearly see the texture in the image and achieve optimal perceptual similarity. In conclusion, the algorithm for obtaining image perceptual quality at the expense of image quality does not accord with human visual perception and vice versa. By observing the LPIPS perceptual similarity of these

results, the algorithm in this paper reached the optimal value. This proves that our algorithm also has the advantage of perception quality at CNN feature level. Referring to the objective evaluation indicators in the table below, Table 1 shows the PI/RMSE evaluation of image perceptual quality by each algorithm in SET5, SET14 and BSDS100. Table 2 shows the PSNR/SSIM of various algorithms to evaluate image distortion. Table 3 shows the perceptual similarity of each algorithm's results.

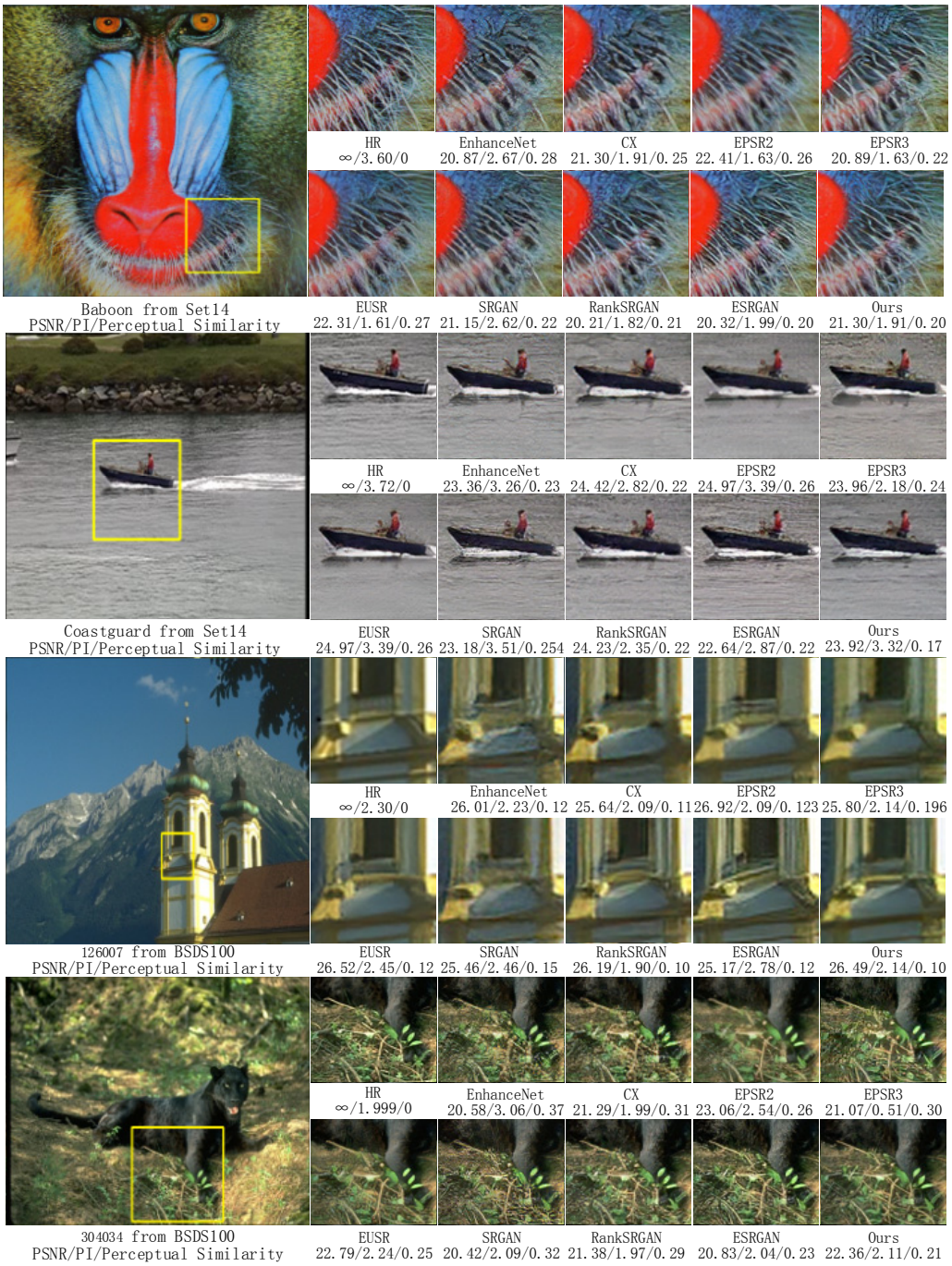


Figure 4. We compare our algorithm We compare our algorithm with the latest recent popular perception-based algorithm. From these pictures, we can see that the experimental results in this paper can guarantee optimal perception similarity and good image quality.

Table 1. PI/RMSE of various algorithms in each dataset; red represents the optimal value, blue represents the suboptimal value.

PI/RMSE	Set5	Set14	BSDS100	PIRM_val
EnhanceNet	2.926/10.088	3.018/18.068	2.908/17.515	2.688/15.985
SRGAN	3.355/9.313	2.882/17.432	2.531/17.138	–
EUSR	4.904/7.596	3.094/15.834	2.683/15.049	2.353/12.579
EPSR2	4.112/7.446	3.025/15.626	2.746/14.569	2.388/12.409
EPSR3	3.257/8.930	2.698/17.074	2.199/16.782	2.069/15.359
CX	3.295/9.583	2.759/17.441	2.250/18.781	2.131/15.248
RankSRGAN	3.083/8.702	2.615/17.143	2.131/16.500	2.021/14.993
ESRGAN	3.320/8.219	2.926/18.161	2.337/17.093	2.299/15.569
ours	3.176/7.883	2.813/16.728	2.326/16.375	2.211/14.115

Table 2. PSNR/SSIM of various algorithms in each dataset; red represents the optimal value, blue represents the suboptimal value.

PSNR/SSIM	SET5	SET14	BSDS100	PIRM_val
EnhanceNet	28.573/0.81	24.967/0.651	24.368/0.614	25.069/0.646
SRGAN	29.426/0.836	25.186/0.665	24.569/0.625	–
EUSR	31.045/0.863	26.416/0.705	25.651/0.669	27.265/0.728
EPSR2	31.240/0.865	26.552/0.709	25.896/0.667	27.350/0.728
EPSR3	29.586/0.841	25.452/0.681	24.726/0.636	25.459/0.666
CX	29.116/0.832	25.148/0.671	24.039/0.629	25.410/0.675
RankSRGAN	29.796/0.839	26.484/0.703	25.505/0.649	25.622/0.659
ESRGAN	30.318/0.871	26.406/0.722	24.479/0.677	25.577/0.696
ours	30.586/0.862	27.024/0.742	25.896/0.693	26.224/0.712

Table 3. LPIPS perception similarity of various algorithms in each dataset; red represents the optimal value, blue represents the suboptimal value.

LPIPS	SET5	SET14	BSDS100	PIRM_val
EnhanceNet	0.102	0.168	0.209	0.167
SRGAN	0.084	0.154	0.189	–
EUSR	0.081	0.155	0.194	0.146
EPSR2	0.078	0.161	0.198	0.143
EPSR3	0.089	0.163	0.200	0.187
CX	0.081	0.152	0.190	0.145
RankSRGAN	0.072	0.143	0.176	0.139
ESRGAN	0.067	0.151	0.166	0.132
ours	0.066	0.134	0.163	0.126

The results in the table represent the average values of all image results in the dataset. According to the results in Table 1, some algorithms are dedicated to optimizing the RMSE (EUSR, EPSR2) of the image, and some algorithms are dedicated to optimizing the PI value of the image (CX, SRGAN, ESRGAN, OURS). The algorithm in this paper is superior to ESRGAN in both PI and RMSE. Compared with RankSRGAN, the results of this paper are not as good as RankSRGAN in PI, but better than RankSRGAN in RMSE. As for PSNR/SSIM, according to the results in Table 2, the results of this paper reached the optimal in the SET14 and BSDS100 dataset, and the optimal SSIM value in the SET5 dataset. The PIRM_val dataset is only lower than the RMSE-based EUSR and EPSR2. According to the results in Table 3, our algorithm achieves the best performance in the LPIPS perception similarity index and shows that the proposed algorithm has the best-perceived similarity in terms of CNN features, which indicates that the algorithm in this paper is more suitable for human visual perception at the CNN feature level. To prove that the proposed algorithm does indeed correspond with human visual perception, we evaluated the mean opinion score of the results of each algorithm.

Mean Opinion Score (MOS): To evaluate whether the results of each algorithm are close to human visual perception, this paper tests the mean opinion score. We asked 40 students to perform MOS measurements in this experiment. For each classmate, we randomly display 20 pictures from the PIRM_test dataset in each algorithm. Each student ranked the algorithms according to how close they are to the original image, the closest being Rank1, followed by Rank2, and so on. In this experiment, we have chosen the most popular algorithms ESRGAN and RankSRGAN, EPSR algorithm that can adjust PI and RMSE, and the classic algorithm EnhanceNet. The results are shown in Figure 5. Because our results produce fewer artifacts, the ratio of our algorithm to Rank1 is the largest, and the overall results are better than ESRGAN and RankSRGAN, far more than EPSR2, EPSR3 and EnhanceNet, indicating that our algorithm is closer to the original image in visual perception.

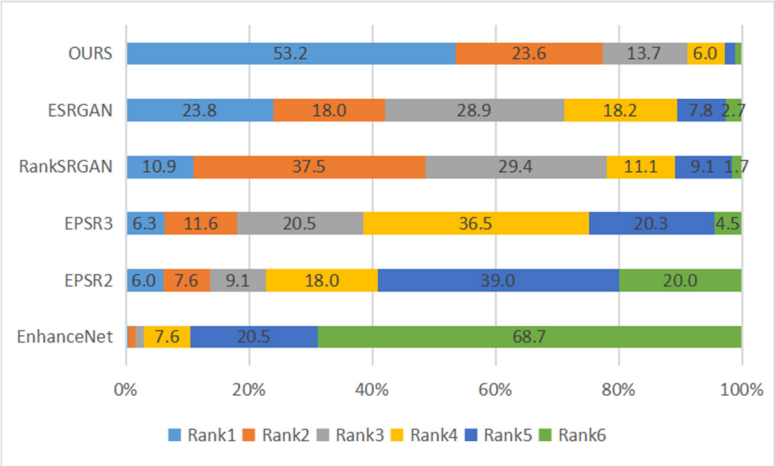


Figure 5. The ranking results of user studies and the current popular algorithm.

4.4. Ablation Eeperiences

To prove that the proposed scheme is effective, we performed an ablation experiment to prove that each module of the experiment contributes to the image super-resolution algorithm. As shown in Figure 6, the first column and the second column show the real image. The third column shows the experimental results of the original ESRGAN, and then the corresponding results when adding each algorithm. We elaborate on each improvement separately.

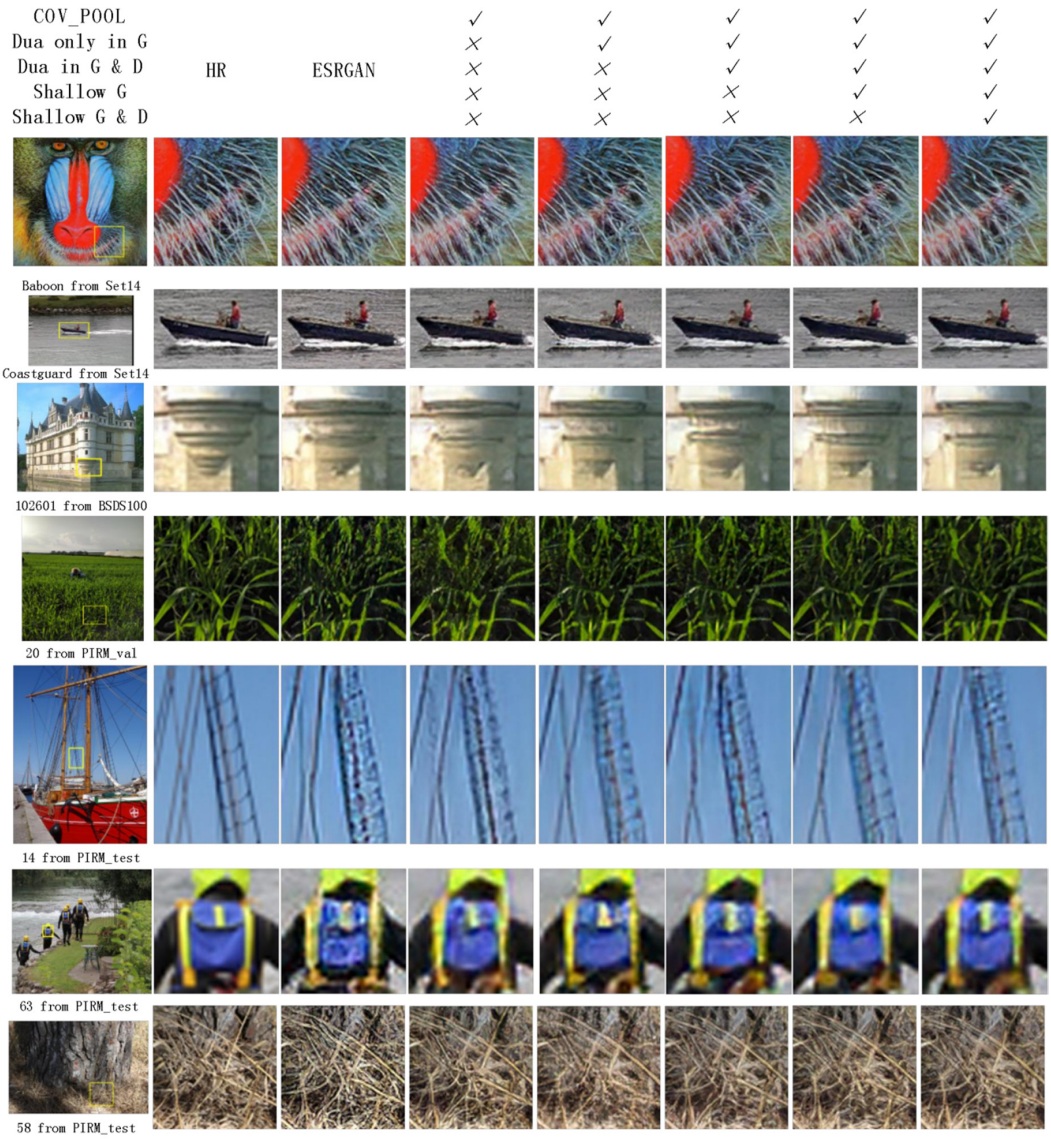


Figure 6. An overview of the impact of the addition of each module in this article. Each column represents the effect of adding the corresponding module; the corresponding module is displayed at the top of the picture; √ indicates to add this module, × indicates not added. The yellow rectangle indicates the area we mainly display.

4.4.1. Covariance Normalization (COVNORM)

In this paper, a second-order covariance normalization is added at the end of the feature extractor for the perceptual loss of the image so that the perceptual loss can minimize the high-order features between the HR image and the SR picture. As shown in the image results of the third and fourth columns in Figure 6, adding the covariance normalized image reduces the artefacts of the ESRGAN image, such as the baboon in SET14. In the original algorithm, the excess texture of the beard is significantly reduced and restored. ESRGAN algorithm shows more serious noise and distortion in the parts with complex texture, such as the 20 from PIRM_val. With CovNorm added to the perceptual loss, there is much less noise in the picture. In addition, adding covariance normalization to the perceptual feature extractor can preserve more texture details (referring to the third and the fourth column of 102,601 from BSDS100 of Figure 6), such as animal hair, as shown in Figure 7.

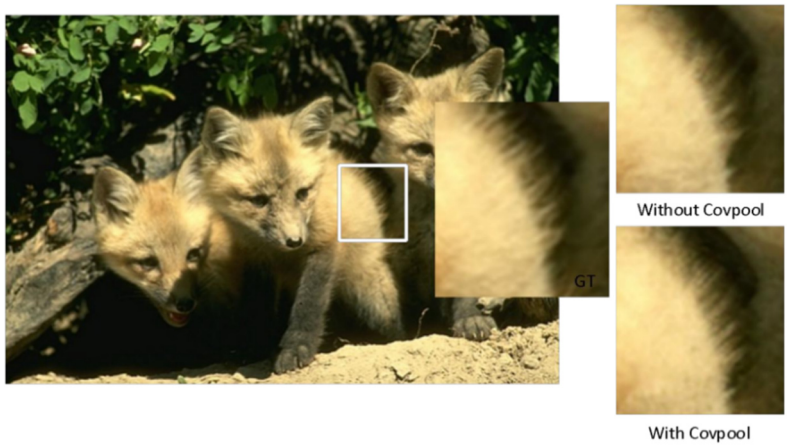


Figure 7. CovNorm enables SR images to produce sharp textures; the GT represents the Ground truth image.

Figure 8 shows that, after adding CovNorm, the SR picture is closer to the original image in brightness. As can be seen from the brightness map on the right, after adding CovNorm in the feature extractor, the brightness curve of SR almost coincides with the brightness curve of the real image, which improves the brightness of the generated image.

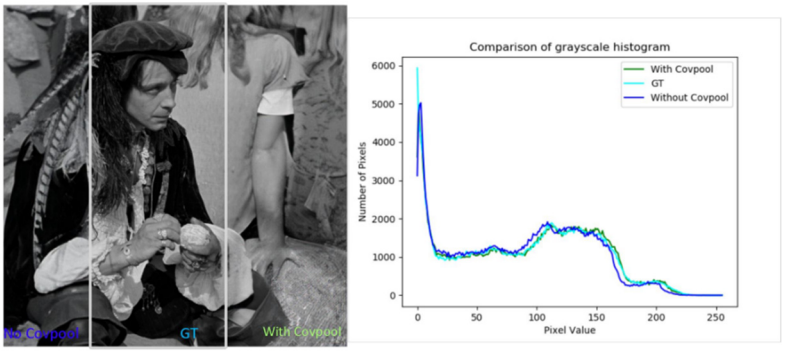


Figure 8. CovNorm ensures that the brightness of the image is closer to the Ground-truth image.

4.4.2. DUA Only in THE Generator

In this paper, DUA is added to the upper layer of the generator, which enhances the expressive ability of the network, and carefully coordinates the dependence between the channel features of the image and the correlation of each position in the feature map, making the texture of the generated image closer to the texture of the HR image. According to the comparison of the fourth column and the fifth column of Figure 6, for example, after the DUA attention mechanism is added to the beard portion of the baboon in SET14, the beard portion reduces the non-existent texture and noise on the basis of the upper portion. Including the results of 63 from PIRM_test, a network with a dual attention mechanism can reduce the noise present in the original image without affecting the perceptual quality of the image (we can still clearly see the texture portion of the image).

4.4.3. DUA in G and D

This article not only adds DUA to the generator of the network but also adds DUA to the discriminator. The discriminator can more accurately apply complex geometric feature constraints to the global image structure, which can more accurately constrain the geometric features of the image. From the comparison in the fifth and sixth columns in Figure 6, it can be seen that the 20 pictures of PIRM_VAL can recover a more real structure after adding DUA to the discriminator, unlike the unclear or even messy structure in the original picture.

4.4.4. Shallow G

To make full use of low-resolution images and fully exploit the features of each level, the shallow features of the image cannot be ignored when considering the deep features of the image. The effect of adding a shallow network can be seen from the sixth and seventh columns in Figure 6. After considering the shallow information of the LR image, the network can restore a smoother image and ensure the perceptual quality of the image.

In order to observe the effect of the shallow network, we extract the output of the network and the shallow network (after upsampling) and show the results after feature fusion. As shown in the Figure 9, adding a shallow network can obtain richer texture and image information. In addition, according to Table 4 and Figure 6, adding a shallow network can improve the image quality of the generated images and reduce artefacts.

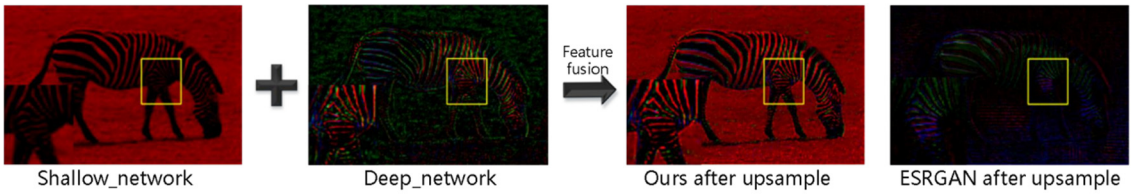


Figure 9. Comparing the features of the added shallow network with the original feature. The third image represents the result of the fusion of deep and shallow features, and the fourth image represents the original feature map.

Table 4. PSNR/SSIM and PI/RMSE of various algorithms in BSDS100 dataset.

BSDS100	ESRGAN	CovNorm	DUA in G	DUA in G and D	Shallow G	Shallow G and D
PSNR/SSIM	25.402/0.683	25.943/0.703	26.147/0.708	26.176/0.711	26.184/0.711	26.224/0.712
PI/RMSE	2.184/15.484	2.147/14.566	2.078/14.211	2.086/14.110	2.103/14.099	2.116/14.023

4.4.5. Shallow D

In the discriminator for GAN networks, if only the statistical information of high-level feature maps is considered and the statistical information of low-level feature maps is ignored, it may cause SR images to not completely correspond to ground-truth images, so a shallow discriminator is added in this paper. As shown in the last column of Figure 6, after adding shallow D, the entire structure information of the SR picture is more accurate. For example, in 1,002,601 in BSDS100, the details of the building can be displayed more accurately and, in the coastguard in SET14, the reduced artefacts and false information in the figure show that the method in this paper is effective.

Finally, Table 4 shows the PSNR/SSIM and PI/RMSE results of the PIRM_test dataset after adding each algorithm.

According to this result, various algorithms added in this paper can improve the quality of SR images. In particular, adding CovNorm to the loss function and adding DUA to the generator and discriminator can both improve the PSNR of the SR image and improve the perceptual quality of the image (based on PI/RMSE). Adding a shallow network later can further improve the PSNR and RMSE of the SR image.

5. Conclusions

In this exposition, we improve the ESRGAN algorithm which achieved a significant breakthrough in SR image perception quality. However, in the process of enhancing image perception, the algorithm sacrifices the quality of the image and produces many non-existent textures in SR image. This paper puts forward an improvement to this problem.

We improve the network structure by adding shallow generators and shallow discriminator, so that we can pay attention to the shallow information in the image. In the perceptual loss, we add covariance normalization to optimize the perceptual loss from higher-order image statistical characteristics. Considering the higher-order dependence between image features, the second-order channel attention mechanism and spatial attention mechanism are added to both the generator and the discriminator. Compared with other experiments, our algorithm can reduce the generation of artefacts and improve the image quality (PSNR,SSIM) without sacrificing the image quality seriously.

Author Contributions: Conceptualization, L.W. and C.L.; methodology, C.L.; software, C.L.; validation, S.C. and L.W.; formal analysis, N.A.; writing—original draft preparation, C.L.; writing—review and editing, L.W. and S.C.; visualization, S.C. and N.A.; All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of Xinjiang Uygur Autonomous Region, grant number 2019D01C033, in part by the National Science Foundation of China under Grant 61771416 and U1903213, in part by the CERNET Innovation Project under Grant NGII20180201, and in part by the Creative Research Groups of Higher Education of Xinjiang Uygur Autonomous Region under Grant XJEDU2017T002.

Acknowledgments: Theoretical analysis and discussion, Wang Liejun and Cheng Shuli; conceptualization, Can Li and Liejun Wang; Formal analysis, Naixiang Ao; Methodology, Can Li; Software, Can Li; Validation, Liejun Wang and Shuli Cheng; Visualization, Shuli Cheng and Naixiang Ao; Writing – original draft, Can Li; Writing – review & editing, Liejun Wang and Shuli Cheng.

Conflicts of Interest: The authors have declared that no competing interests exist.

References

1. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Trans.* **2016**, *38*, 295–307. [CrossRef] [PubMed]
2. Lai, W.; Huang, J.; Ahuja, N.; Yang, M. Fast and Accurate Image Super-Resolution with Deep Laplacian Pyramid Networks. *IEEE Trans.* **2019**, *41*, 2599–2613. [CrossRef] [PubMed]
3. Kim, J.; Lee, J.K.; Lee, K.M. Deeply-Recursive Convolutional Network for Image Super-Resolution. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 1637–1645.
4. Tong, T.; Li, G.; Liu, X.; Gao, Q. Image Super-Resolution Using Dense Skip Connections. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 4809–4817.
5. Tai, Y.; Yang, J.; Liu, X.; Xu, C. MemNet: A Persistent Memory Network for Image Restoration. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 4549–4557.
6. Chao, D.; Chen, C.L.; He, K.; Tang, X. Learning a Deep Convolutional Network for Image Super-Resolution. In Proceedings of the European Conference on Computer Vision (ECCV), Zurich, Switzerland, 6–12 September 2014; Springer: Cham, Switzerland, 2014; pp. 184–199.
7. Kim, J.; Lee, J. Deep Residual Network with Enhanced Upscaling Module for Super-Resolution. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018; pp. 913–9138.
8. Choi, J.H.; Kim, J.H.; Cheon, M.; Lee, J.S. Deep learning-based image super-resolution considering quantitative and perceptual quality. *arXiv* **2018**, arXiv:1809.04789. [CrossRef]
9. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Shi, W. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 105–114.
10. Luo, X.; Chen, R.; Xie, Y.; Qu, Y.; Li, C. Bi-GANs-ST for Perceptual Image Super-Resolution. In Proceedings of the European Conference on Computer Vision (ECCV); Springer: Cham, Switzerland, 2019; pp. 20–34.
11. Mechrez, R.; Talmi, I.; Shama, F.; Zelnik-Manor, L. Learning to Maintain Natural Image Statistics. *arXiv* **2018**, arXiv:1803.04626.
12. Wang, X.; Yu, K.; Dong, C.; Loy, C.C. Recovering Realistic Texture in Image Super-Resolution by Deep Spatial Feature Transform. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 606–615.

13. Vasu, S.; Madam, N.T.; Rajagopalan, A.N. Analyzing Perception-Distortion Tradeoff Using Enhanced Perceptual Super-Resolution Network. In Proceedings of the European Conference on Computer Vision (ECCV); Springer: Cham, Switzerland, 2019; pp. 114–131.
14. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Change Loy, C. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. In Proceedings of the European Conference on Computer Vision (ECCV); Springer: Cham, Switzerland, 2019; pp. 63–79.
15. Zhang, W.; Liu, Y.; Dong, C.; Qiao, Y. RankSRGAN: Generative Adversarial Networks with Ranker for Image Super-Resolution. In Proceedings of the IEEE International Conference on Computer Vision; IEEE: Piscataway, NJ, USA, 2019; pp. 3096–3105.
16. Blau, Y.; Michaeli, T. The Perception-Distortion Tradeoff. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 6228–6237.
17. Michelini, P.N.; Zhu, D.; Liu, H. Multi–scale Recursive and Perception–Distortion Controllable Image Super–Resolution. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2019; pp. 3–19.
18. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2014.
19. Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; Fu, Y. Residual Dense Network for Image Super-Resolution. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2472–2481.
20. Pan, J.; Liu, S.; Sun, D.; Zhang, J.; Liu, Y.; Ren, J.; Yang, M.H. Learning Dual Convolutional Neural Networks for Low-Level Vision. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3070–3079.
21. Pan, J.; Liu, Y.; Dong, J.; Zhang, J.; Ren, J.; Tang, J.; Yang, M.H. Physics-Based Generative Adversarial Models for Image Restoration and Beyond. *arXiv* **2018**, arXiv:1808.00605v1. [CrossRef] [PubMed]
22. Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. Image Super-Resolution Using Very Deep Residual Channel Attention Networks. In Proceedings of the European Conference on Computer Vision (ECCV); Springer: Cham, Switzerland, 2018; pp. 294–310.
23. Dai, T.; Cai, J.; Zhang, Y.; Xia, S.; Zhang, L. Second-Order Attention Network for Single Image Super-Resolution. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019; pp. 11057–11066.
24. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision (ECCV); Springer: Cham, Switzerland, 2018; pp. 3–19.
25. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2016; pp. 694–711.
26. Mechrez, R.; Talmi, I.; Zelnik-Manor, L. The Contextual Loss for Image Transformation with Non-Aligned Data. In Proceedings of the European Conference on Computer Vision (ECCV); Springer: Cham, Switzerland, 2018; pp. 768–783.
27. Jolicoeur-Martineau, A. The relativistic discriminator: A key element missing from standard GAN. *arXiv* **2018**, arXiv:1807.00734.
28. Vu, T.; Luu, T.M.; Yoo, C.D. Perception-Enhanced Image Super-Resolution via Relativistic Generative Adversarial Networks. In Proceedings of the European Conference on Computer Vision (ECCV); Springer: Cham, Switzerland, 2019; pp. 98–113.
29. Li, P.; Xie, J.; Wang, Q.; Zuo, W. Is Second-Order Information Helpful for Large-Scale Visual Recognition? In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2089–2097.
30. Li, P.; Xie, J.; Wang, Q.; Gao, Z. Towards Faster Training of Global Covariance Pooling Networks by Iterative Matrix Square Root Normalization. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 947–955.
31. Blau, Y.; Mechrez, R.; Timofte, R.; Michaeli, T.; Zelnik-Manor, L. The 2018 PIRM Challenge on Perceptual Image Super-Resolution. In Proceedings of the European Conference on Computer Vision (ECCV); Springer: Cham, Switzerland, 2018; pp. 334–355.
32. Wang, D.; Wang, L. On OCT Image Classification via Deep Learning. *IEEE Photonics J.* **2019**, *11*, 1–14. [CrossRef]

33. Cheng, S.; Wang, L.; Du, A. Histopathological Image Retrieval Based on Asymmetric Residual Hash and DNA Coding. *IEEE Access* **2019**, *7*, 101388–101400. [CrossRef]
34. Lin, T.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2999–3007.
35. Lim, B.; Son, S.; Kim, H.; Nah, S.; Mu Lee, K. Enhanced Deep Residual Networks for Single Image Super-Resolution. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 1132–1140.
36. Wang, Y.; Wang, L.; Wang, H.; Li, P. End-to-End Image Super-Resolution via Deep and Shallow Convolutional Networks. *IEEE Access* **2019**, *7*, 31959–31970. [CrossRef]
37. Soh, J.W.; Park, G.Y.; Jo, J.; Cho, N.I. Natural and Realistic Single Image Super-Resolution with Explicit Natural Manifold Discrimination. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019; pp. 8114–8123.
38. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
39. Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-local Neural Networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7794–7803.
40. Miyato, T.; Kataoka, T.; Koyama, M.; Yoshida, Y. Spectral Normalization for Generative Adversarial Networks. *arXiv* **2018**, arXiv:1802.05957v1.
41. Zhang, H.; Goodfellow, I. Self-Attention Generative Adversarial Networks. *arXiv* **2018**, arXiv:1805.08318v2.
42. Liu, D.; Wen, B.; Fan, Y.; Loy, C.C.; Huang, T.S. Non-local recurrent network for image Restoration. In *Neural Information Processing Systems (NIPS)*; MIT Press: Cambridge, MA, USA, 2018.
43. Mittal, A.; Soundararajan, R.; Bovik, A.C. Making a "Completely Blind" Image Quality Analyzer. *IEEE Signal Process. Lett.* **2013**, *20*, 209–212. [CrossRef]
44. Mittal, A.; Moorthy, A.K.; Bovik, A.C. No-Reference Image Quality Assessment in the Spatial Domain. *IEEE Trans.* **2012**, *21*, 4695–4708. [CrossRef] [PubMed]
45. Gondal, M.W.; Schölkopf, B.; Hirsch, M. The Unreasonable Effectiveness of Texture Transfer for Single Image Super-resolution. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 80–97.
46. Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 586–595.
47. Sajjadi, M.S.M.; Schölkopf, B.; Hirsch, M. EnhanceNet: Single Image Super-Resolution through Automated Texture Synthesis. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 4501–4510.