Multi-Perspective Discriminators Based Generative Adversarial Network for Image Super Resolution

Oh-Young Lee¹, Yoon-Ho Shin¹, and Jong-Ok Kim¹, (Member, IEEE)

¹School of Electrical Engineering, Korea University, Seoul, 02841, Rep. of Korea (e-mail: {triple0, yoon731, jokim}@korea.ac.kr)

Corresponding author: Jong-Ok Kim (e-mail: jokim@korea.ac.kr).

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2019R1A2C1005834).

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(1ITP-2019-2018-0-01421) supervised by the IITP(Institute of Information & communications Technology Planning & Evaluation).

ABSTRACT Recently, generative adversarial network-based image super resolution has been investigated, and it has been shown to lead to overwhelming improvements in subjective quality. However, it also leads to checkerboard artifacts and the unpleasing high-frequency (HF) components. In this paper, we propose a multi-discriminators-based image super resolution method that distinguishes those artifacts from various perspectives. First, the DCT perspective discriminator is proposed because the checkerboard artifacts are easily separated on the frequency domain. Second, the gradient perspective discriminator is proposed, because the unpleasing HF components can be discriminated on the gradient magnitude distribution. These proposed multi-perspective discriminators can easily identify artifacts, and they can help the generator reproduce artifact-less SR images. The experimental results show that the proposed SR-GAN with multi-perspective discriminators achieves objective and subjective quality improvements in terms of PSNR, SSIM, PI and MOS, as compared to the conventional SR-GAN by reducing the aforementioned artifacts.

INDEX TERMS Image super-resolution, Deep learning for super resolution, SR GAN, Multi-discriminators.

I.INTRODUCTION In recent years, spatial resolution enhancement has become a very important research topic in the field of image processing [1-3]. Moreover, with the arrival of the era of 8K displays, there has been growing interest in increasing the spatial resolution of images.

Although image super-resolution (SR) has been extensively studied for years [4-8][38-41], it remains challenging due to the fact that it is an inherently ill-posed problem. Numerous non-deep-learning-based methods such as self-similarity [4-7] and sparse representation [8] have been proposed and shown to lead to significant quality improvements. However, a couple of limitations still remain in estimating the correlation between low-resolution (LR) and high resolution (HR) images in edge and texture regions.

As interest in deep learning has recently increased, a few deep learning-based SR methods have been proposed [9-16][42-53]. Deep networks allow us to easily learn the correlation between LR and HR and they typically achieve better performance than non-deep-learning approaches. However, the common MSE (mean square error) used for cost calculation has a fundamental difficulty in accurately measuring complex signal structures, such as texture. Due to the average effect of MSE, it is difficult to learn for those complex texture regions. The MSE loss is not always proportional to human perception quality, and it is likely to be insufficient for accurate measurement of perceptual quality. Thus, perceptual loss such as MSE of VGG has been increasingly studied for visual quality enhancement using deep learning. Although VGG features contribute to increasing the perceptual quality, they are not sufficient to fully represent the actual characteristics of the features due to the MSE effect (MSE of VGG is used for loss). Moreover, the VGG features lead to HF noise.

The discovery of an adversarial network is very helpful for image SR. In contrast to using a generative network
by itself, the use of generative and adversarial networks (GAN) together can lead to substantially improved subjective quality.

Specifically, it can enhance the sharpness of the texture region in an image. While the adversarial network attempts to distinguish between real and fake images, the generative network attempts to fool the adversarial network. In this optimization process, the generative network obtains an enhanced ability to reconstruct high-frequency (HF). However, some artifact and noise are generated, which should be overcome for GAN-based SR. First, the checkerboard artifacts, as shown in Figure 2, are partially attributed to the upsampling of input features to a target resolution within the SR-GAN network and they can be reduced by using interpolated input or MSE loss [26]. However, both solutions lead to blurring in the output image. The second concern is unpleasing HF noise. Since the aim of the generator is to fool the discriminator, the performance of the generator depends on that of the discriminator. As mentioned in [21], the conventional discriminator distinguishes between real and fake images based on the amount of HF. As a result, if even arbitrary HF noise is added to the fake image, the conventional discriminator may classify it as real [21]. Because of this, GAN-based SR methods tend to produce less meaningful and unpleasing HF noise in the resulting images, as shown in Figure 2. This leads to low objective quality, such as PSNR.

In this paper, we propose a new multi-discriminators-based SR-GAN framework that distinguishes between real and fake images from various perspectives, as shown in Figure 1. It was experimentally confirmed that the SR results differ depending on the discriminator. Based on this, we attempt to overcome the existing shortcomings by using the multi-perspective discriminators that are newly proposed in this paper. In addition to a conventional spatial domain discriminator, we propose two novel discriminators for accurately distinguishing SR images in other aspects, which are discriminators on both the DCT and gradient domains. While the spatial discriminator in SR-GAN only relies on the amount of HF, in this paper, we focus on discrimination from the perspectives of gradient and DCT, as well as the amount of HF that should be used to reduce unpleasing HF noise and checkerboard artifacts. The DCT domain discriminator compares real and fake images on the frequency domain. The checkerboard artifacts are repeated periodically on the spatial domain. When an image with the periodical artifact is transformed by DCT, the artifacts are locally concentrated in specific HF region, and can thus be easily addressed. The gradient magnitude distribution is a method that can be used to statistically measure the structural complexity of image signals. In addition, it is closely related to the spectral characteristics. The larger the gradient magnitude is, the higher the frequency is as well. Therefore, the gradient magnitude distribution discriminator can be used to evaluate whether or not the HF components of an SR image are well restored.

The main contributions can be summarized as follows:

- Multi-perspective discriminators for SR-GAN are newly proposed, and they work on distinct domains such as frequency and gradient as well as spatial.
- The frequency and gradient discriminators are proposed to both remove artifacts and refine the restored HF, while the conventional spatial discriminator contributes to the restoration of HF signals. In other words, the usages of the proposed discriminators are different from those of conventional discriminators.
- The proposed method achieves higher visual quality with artifact reduction than the existing GAN-based SR methods by adding frequency and gradient discriminators.

The rest of this paper is organized as follows. In Section II, related works are discussed. In Section III, we begin by describing the reference SR-GAN, then, present the proposed multi-perspective discriminators in detail. In Section IV, performance evaluations of the proposed...
evaluators are presented. Finally, Section V concludes the paper.

II. RELATED WORK

SR algorithms have been extensively studied for years [4-8]. Recently, since [9] proposed Convolution Neural Networks (CNN)-based SR methods, various deep learning-based SR algorithms have been investigated. These algorithms have used CNN to estimate the correlation between LR and HR images, as opposed to a specific model based on special characteristics such as sparse representation and, self-similarity [4-8]. [10-14] propose novel network structures suitable for SR that involve using methods such as skip connections, residual blocks, and Laplacian pyramid. They extract and learn features that are difficult for humans to understand and use these to reconstruct HR images. They show better performances than conventional non-deep-learning-based techniques. However, since they only consider Mean Square Error (MSE) and not perceptual quality, they have the limitations of all MSE-based methods. MSE is known to lead to blurry results, since it has an average effect. The importance of perceptual quality has already been emphasized in style transfer [15]. [16], which was inspired by [15], proposes VGG feature-based perceptual losses. They show less blurred results than MSE-based deep learning SR, but they also have side effects such as the occurrence of artifacts and lower PSNRs. Although novel network structures for deep learning SR have been proposed [42-53], they all still have certain limitations.

One study has shown that the subjective quality of the SR results can be significantly increased by using the Generative Adversarial Network [17]. Following the introduction of GAN, various algorithms using GAN have been proposed in SR [18-23]. In [18], rather than MSE loss, VGG loss is used to improve the visual quality with adversarial loss. As a result, the subjective quality is greatly improved. However, PSNR is low and there are checkerboard artifacts. [26] presents a detailed description of the checkerboard artifacts. If interpolated input or MSE-based loss is used, the checkerboard artifacts can be easily removed. However, the results will be blurred. In [19], the authors add texture matching loss to the generator loss function to produce more realistic textures and further reduce the occurrence of visually implausible artifacts. Although [19] shows a more natural HF than [18], it still has a lower PSNR. [20] also proposes a novel generator loss function with the EUSR network [24]. However, instead of MSE loss and VGG loss, it uses content loss and differential content loss, which both use L1-norm. Moreover, DCT loss is proposed to measure the similarity on the frequency domain. Although it reduces unpleasing HF and succeeds in achieving high objective quality in terms of PSNR and SSIM, it has a higher perceptual index (PI) value than SR-GAN (a high PI value indicates a low subjective quality).

Along with research into improving the generator, research into improving the discriminator has also begun to be conducted. [21] proposes a pre-trained feature domain discriminator using a generator network with short-and-long range skip connections. The authors expect that the feature discriminator distinguishes SR and HR based not simply on HF but on structural features. They justify their use of multi-discriminators by obtaining high PSNR and SSIM values along with clarity close to the ground-truth. However, the use of VGG based discriminator has a limitation in terms of signal accuracy. Since VGG has been proposed for image classification, pooling layers are included and the VGG features are low-resolution. Thus, VGG feature maps tend to be global rather than local, leading to the occurrence of new artifacts around the edge region. In [22], the authors propose an improved network architecture and relativistic discriminator. In order to improve the structure of the generator, they remove the batch normalization and use the Residual-in-Residual Dense Block (RRDB), which combines a multi-level residual network and dense connections. Inspired by [25], the authors apply a relativistic discriminator to SR. The discriminator learns to judge whether one image is more realistic than another, which guides the generator to synthesize more detailed textures. Moreover, instead of MSE, they use L1-norm-based content loss and modified perceptual loss for the generator loss function. As a result, they achieve consistently better performance than conventional SR methods in terms of subjective and objective qualities. However, jagging and hallucination artifacts often occur near the strong edge. In [54], the authors propose deep dense skip connection based generator to reduce over-smoothing and use the Wasserstein distance based discriminator to improve the stability of learning. They achieve higher PSNR, SSIM and MOS with less blur. However, although the MSE loss is used, there is a lot of noise near the edge region due to the limitation of the Wasserstein distance based discriminator. This has an adverse effect on subjective quality.

In summary, the studies that propose a novel discriminator [21, 22, 54] have successfully reduced the blur artifact, but they have suffered from noise generation, which degrades the image quality.

DCT transform-based approaches have been widely used for image denoising [27-28]. Since noise is a high frequency component, it is better separated in the frequency domain than in the spatial domain. [27-28] show that the separated noise components by DCT transform can be easily removed. The gradient magnitude distribution (GMD) shows how many components each gradient of an image has. This allows for a quantitative evaluation of the image structure quality. [29] uses the GMD as a measure of the texture quality to show the superiority of SR performance. [30] proposes the gradient histogram preservation (GHP)-
based denoising framework for enhancing the texture. It evaluates the degree of the enhanced texture using the GMD and attempts to make the GMD of the denoised image close to the original.

The key idea of this paper is to distinguish between real and fake images using various perspective discriminators to reduce SR artifacts and noise. In a case study, we select three discriminators (spatial, DCT and GMD) to reduce the checkerboard artifacts and unpleasing HF. The standard discriminator only distinguishes whether or not the amount of HF is large. The rest of the proposed discriminators attempt to distinguish not only whether or not the amount of HF is large, but also whether or not the shape and magnitude of HF are correct by analyzing them from a different perspective.

III. THE PROPOSED METHOD

Our goal is to generate a clean and natural HR image that is similar to the original by reducing the checkerboard artifacts and unpleasing HF. For this GAN-based SR study, we straightforwardly extend the SR-GAN framework [18] by adding two different discriminators as shown in Figure 1. In this section, we first describe the reference SR-GAN framework, then introduce the three proposed discriminators, which run on a spatial domain, a DCT domain, and a gradient magnitude domain.

A. NETWORK ARCHITECTURE

As proposed in [18], the SR-GAN consists of two networks, called the generator and discriminator. The generator network produces the HR image given an LR image, while the discriminator network distinguishes between the original HR and its SR version. As the two networks compete against each other, the generator improves its SR performance.

The generator network $G_{\theta_G}$ is trained as a feed-forward CNN parameterized by $\theta_G$. Here $\theta_G$ indicates the weights and biases of the deep network and it is obtained by optimizing a generator loss function $l^S_R$ for SR as follows.

$$\hat{\theta}_G = \arg\min_{\theta_G} \frac{1}{N} \sum_{n=1}^{N} l^S_R(G_{\theta_G}(I_{n}^{LR}), I_{n}^{HR})$$

(1)

where, $I_{n}^{LR}$ and $I_{n}^{HR}$ are the $n$ th training LR and HR images, respectively. The loss function $l^S_R$ is described in detail in Section III.B.

The discriminator, $D_{\theta_D}$ and parameter $\theta_D$ are similar to $G_{\theta_G}$ and $\theta_G$ for the generator, respectively. $\theta_D$ is obtained as follows.

$$\hat{\theta}_D = \arg\min_{\theta_D} \frac{1}{N} \sum_{n=1}^{N} l^D_{S_R} (D_{\theta_D}(I_{n}^{LR}), D_{\theta_D}(I_{n}^{HR}))$$

(2)

Regarding the adversarial min-max problem, it can be arranged as follows.

$$\min_{\theta_G} \max_{\theta_D} E[\log(D_{\theta_D}(G_{\theta_G}(I^{SR}))) + E[\log(1-D_{\theta_D}(I^{SR})))$$

(3)

In order to distinguish real images from fake images, $D_{\theta_D}$ attempts to maximize equation (3), while in an attempt to fool the discriminator $D_{\theta_D}$, $G_{\theta_G}$ aims to minimize equation (3). $D_{\theta_D}$ is optimized in an alternating manner along with $G_{\theta_G}$ to solve the adversarial min-max problem in (3).

In this paper, our network is a single generator with three discriminators, and it is basically identical to that in [18].
As shown in Figure 4, our generator consists of a few convolution layers, 16 identical residual blocks and a sub-pixel convolution layer. On the other hand, discriminators contain eight convolution layers with increasing channels, which use $3 \times 3$ filter kernels. In addition, they are followed by two dense layers and sigmoid activation function. The proposed three discriminators have the same architecture aside from one aspect. The inputs of the Spatial and DCT discriminators are 2-dimensional, whereas the input of the GMD discriminator is 1-dimensional. Therefore, the GMD discriminator uses 1-dimensional filters as opposed to 2-dimensional filters.

**B. PRE-TRAINING OF THE GENERATOR NETWORK**

In adversarial training, the initial value has a significant effect on the results. In order to prevent the adversarial training from failing to learn, the generative network is first trained prior to adversarial training. An MSE loss is used as the generator loss function, and it is written as follows:

$$ l_{MSE} = \|I^{HR} - I^{SR}\|_2 $$

(4)

Pre-training the generator with MSE loss is an important process in the SR-GAN. This can contribute to the stable output results in terms of accurate signal reconstruction. This is complementary for adversarial training, the primary focus of which is generating a more realistic signal, as opposed to signal accuracy. Although HF has not been sufficiently restored with MSE loss during this pre-training step, it will be further reconstructed in the next main training step.

**C. MULTI-PERSPECTIVE дискриминаторы**

As mentioned above, the conventional GAN framework consists of a generator and a discriminator. In order to enhance the performance, the discriminator network is extended with three different kinds of discriminators, as shown in Figure 3. The first discriminator $D_{spa}$ is the same as the one used in the existing SR-GAN. It distinguishes real images from fake images on the spatial domain. The second one $D_{dct}$ discriminates between real and fake images by inspecting their DCT coefficients. It aims to remove the checkerboard artifacts. Finally, the third one $D_{gmd}$ inspects the gradient magnitude distribution of images. It attempts to restore the natural HF by comparing between the GMDs of the generator output and the real ones.

For the three discriminators, the generator loss function $l_G$ is defined as follows:

$$ l_G = l^c_G + \lambda_1 \cdot l^{spa}_A + \lambda_2 \cdot l^{dct}_A + \lambda_3 \cdot l^{gmd}_A $$

(5)

where $l^c_G$ is the content loss and $\lambda_1, \lambda_2, \lambda_3$ are the weights for the adversarial loss terms, which are $l^{spa}_A, l^{dct}_A$ and $l^{gmd}_A$ from the three discriminators. $D_{spa}, D_{dct}$ and $D_{gmd}$
are trained by minimizing the discriminator loss functions $l_{D}^\text{spa}$, $l_{D}^\text{dct}$ and $l_{D}^\text{gmd}$, respectively. Increasing $\lambda_1, \lambda_2, \lambda_3$ results in excessive enhancement that can lead to artificial visual quality, while decreasing $\lambda_1, \lambda_2, \lambda_3$ reduces the effect of the discriminators, leading to less enhancement of HF. In this paper, optimal $\lambda_1, \lambda_2, \lambda_3$ are determined through diverse experiments. Each of the three discriminators is trained separately, and in the same way as SR-GAN, the generator and discriminators are trained alternatingly. Next, we will now explain in detail the content loss, each discriminator, and the corresponding loss.

**Content Loss $l_{C}^G$** Content loss is a term that makes the generator output to be restored, perceptually close to the ground-truth, regardless of the discriminator. In general, pixel-wise MSE is widely used. However, due to its blurred side-effect, the MSE of VGG features is used as follows.

$$l_{C}^G = \|\phi(I_{HR}) - \phi(I_{SR})\|_2$$

where $\phi$ is the feature map obtained by the VGG network.

**Spatial Adversarial Loss $l_{A}^\text{spa}$** The spatial adversarial loss is the same as the adversarial loss in standard SR-GAN. The spatial discriminator distinguishes between real and fake images in the spatial domain. The easiest way to tell the difference between real and fake images in the spatial domain is to check the amount of HF component. Therefore, this term makes the results have a rich HF. The spatial adversarial loss $l_{A}^\text{spa}$ from the spatial discriminator $D_{\text{spa}}$ is defined as follows.

$$l_{A}^\text{spa} = -\log(D_{\text{spa}}(I_{SR}))$$

**DCT Adversarial Loss $l_{A}^\text{dct}$** In order to reduce the checkerboard artifacts, we propose a DCT domain discriminator. The checkerboard artifacts are a pattern that is periodically repeated as shown in Figure 2. Based on this characteristic, we find that the artifact is well separated on the DCT domain. In general, when natural images are transformed onto the DCT domain, their ultra-high frequency components such as top-right, bottom-left and bottom-right typically have small values, almost close to zero. However, the checkerboard artifacts are outstandingly separated on the DCT domain, and they are transformed into high DCT coefficient values in the ultra-high frequency regions as shown in Figure 5. The DCT discriminator identifies whether or not the generated image has the checkerboard artifacts. Thus, this term makes it so that the generator is trained not to produce the checkerboard artifacts. The DCT adversarial loss $l_{A}^\text{dct}$ from the DCT discriminator $D_{\text{dct}}$ is defined as follows.

$$l_{A}^\text{dct} = -\log(D_{\text{dct}}(C_{\text{th},r}(I_{SR})))$$

where $C_{\text{th},r}$ is the thresholded binary map of the DCT coefficient and, $W_r$ and $B_a$ are the truncation and thresholding functions, respectively; these are expressed as follows.
In this paper, spatial signals are transformed by DCT in order to separate the artifacts. However, for typical natural images, the DCT coefficient values in LF (low frequency) are substantially larger than those in both MF and HF. If the DCT coefficients are compared across all of the frequency ranges, LF has a dominant effect on the comparison. However, LF is not a critical factor for the discrimination between real and fake images. Recall that the ultimate goal of SR is to restore HF such as image details. In particular, in order to distinguish the artifact well, HF is more important than LF in this case. Therefore, two constraints are applied to reduce the importance of LF component. At first, the DCT coefficients are converted into a binary image by the thresholding in (11), as shown in Figure 5 (b). This binarization ensures that all DCT coefficients have uniform importance. The threshold, ‘th’ in equation (11) should be carefully determined such that DCT coefficients corresponding to the checkerboard artifacts not be zero-padded. The checkerboard artifacts should be monitored for on the DCT domain by the discriminator in order to be removed. This is determined experimentally in this paper. Secondly, the LF region is truncated on the DCT domain, and it is defined by a region which is within a distance \( r \) as shown in (10) and Figure 5 (c). In other words, LF is excluded from the DCT adversarial loss \( l_{dct}^A \) and the importance of HF increases relatively. The amplitudes of HF coefficients are relatively smaller than those of LF. Thus, as \( r \) in equation (10) becomes smaller, LF has a dominant effect on the overall loss, leading to less enhancement of HF. By contrast, as \( r \) increases, HF including the checkerboard artifacts is enhanced, leading to sharper visual quality, but the checkerboard artifacts are not reduced sufficiently. Thus, it would be better to only discriminate HF coefficients (including the checkerboard artifacts). Under this trade-off relationship, \( r \) is determined experimentally.

**GMD Adversarial Loss** \( l_{gmd}^A \) The gradient magnitude distribution is an indicator that statistically represents the overall signal structure of an image. In particular, it has been used to measure texture characteristics [29-30]. In this paper, we propose a novel adversarial loss from the GMD perspective. It is expected that this could facilitate the discrimination between real and fake images in terms of HF quality. The GMD adversarial loss is expected to be effective for determining whether or not the HF structures are realistically restored. Note that the spatial discriminator focuses solely on the discrepancy of image signals. This leads to the restoration of less meaningful and incongruous HF. By contrast, the GMD discriminator focuses on the overall signal structure statistically. This means that the GMD discriminator considers the harmonization of the restored image with the original. The GMD discriminator causes the result to have an HF with gradient magnitude and shape that are congruous with those of the original image. As a result, the GMD discriminator makes the HF of the result more natural without hallucination artifact and unpleasing HF noise. The GMD adversarial loss \( l_{gmd}^A \) from the GMD discriminator \( D_{gmd} \) is defined as follows.

\[
I_{gmd}^A = - \log(D_{gmd}(H(I^{SR})))
\]  (12)

where \( H \) is the gradient magnitude distribution function, which is defined as follows.
As shown in Figure 6 (a), the gradient magnitude is obtained by equation (14). Then, a histogram of the gradient magnitude is regarded as the gradient magnitude distribution as shown in equation (13) and Figure 6 (b).

### IV. EXPERIMENTAL RESULTS

#### A. EXPERIMENT

All of the networks are trained on a NVIDIA GTX 1080ti graphic processor and i7-8700 CPU using 800 training images from div2k, which are cropped for patch-wise training. These training images are obtained from github [31] and are distinct from the test images. LR images are obtained by down-sampling HR images with a factor of 2. In the above environment, it takes two days to train SR-GAN, while the proposed method takes four days. The time taken for the test is not changed because the proposed discriminators only affect the training time.

For evaluation, we use public datasets such as Set5 [32] and Set14 [33] as well as 100 test images from div2k. All of the experiments are performed with a scale factor of 2 between LR and HR. For quantitative evaluation, we use the measures of PSNR and SSIM [34]. Although they are a renowned visual quality metric, they sometimes fail to capture the visual perception quality of an image. For this reason, Perceptual index [35], the linear combination NIQE [36] and MA [37], is also used to evaluate the perceptual quality. Unlike PSNR and SSIM, a lower PI indicates better quality. Moreover, the mean opinion score (MOS) test is conducted on seven techniques and 12 test images by 20 subjects. Finally, we use Kullback-Leibler divergence (KLD) for GMD to compare the similarity between the SR output and ground-
DCT coefficient comparisons. (a) ground-truth, (c) SG (SR-GAN) [18], (d) SG with DCT loss for generator (SG-D) [20], (f) SG with the DCT discriminator (MPDSD), (b) enlargements of yellow boxes in (a) and (c), and (e) enlargements of yellow boxes in (d) and (f).

Figure 8. DCT coefficient comparisons. (a) ground-truth, (c) SG (SR-GAN) [18], (d) SG with DCT loss for generator (SG-D) [20], (f) SG with the DCT discriminator (MPDSD), (b) enlargements of yellow boxes in (a) and (c), and (e) enlargements of yellow boxes in (d) and (f).

The performance is compared with non-deep-learning-based self-similarity SR as well as the state-of-the-art deep learning-based SR methods of, SR-CNN and SR-GAN.

In this paper, we propose multi-perspective discriminators for SR-GAN. Including the VGG-based discriminator described in [21], we evaluate the various combinations of discriminators in order to confirm the effect of each discriminator as listed in Table 1. The DCT loss for the generator in [20] is also compared with the same network.

B. ABLATION STUDY

Before presenting the comparisons with conventional methods, we first verify the effects of the DCT and GMD discriminators.

DCT discriminator In this sub-section, we explain why the DCT discriminator is proposed and how it works. As shown in Figure 7 (b), the SR-GAN generates the checkerboard artifacts which can be clearly observed as separate coefficients in the ultra-HF region (left-bottom, right-top-and-bottom) on the DCT domain, as shown in Figure 8 (b) and (c). In other words, the checkerboard artifacts are easily separated on the frequency domain compared to on the spatial domain. The DCT discriminator distinguishes between real and fake according to the presence of the separated checkerboard artifacts on the frequency domain.

That’s why we select the DCT for reducing the checkerboard artifacts. However, one may wonder which network (generator or discriminator) is more suitable for DCT. To answer this question, the standard SR-GAN is compared with the SR-GANs using DCT in two different ways. One is to use the DCT as a generator loss (SG-D), and the other is to use the DCT domain discriminator.

Figure 9. Gradient magnitude distribution comparisons. (a) GMD of ground-truth, SR-GAN and MPDSD. (b) enlargement of yellow box in (a).

Figure 10. Visual comparisons. (a) GM discriminator, (b) GMD discriminator, and (c) ground-truth.
HF, discriminator can reconstruct more natural HF by its accuracy, since the discriminator focuses on the amount of HF. This may differ from the ground-truth, tends to be reconstructed. In order to overcome this problem, we adopt the GMD discriminator.n{\textbf{GMD discriminator}} Although using the DCT discriminator reduces the occurrence of the checkerboard artifacts, HF components are recovered insufficiently and unpleasing HF noise still occurs. In order to overcome this problem, we adopt the GMD discriminator. In this sub-section, we describe why we propose the gradient domain discriminator, why the GMD discriminator is selected rather than the GM discriminator and how it works. SR-GAN increases the amount of HF through adversarial loss. However, due to the limitation of the spatial discriminator, unpleasing HF, which differs from the ground-truth, tends to be reconstructed. In the discriminator, the method of measuring the reconstructed HF quality, not the amount of HF, is needed for reducing the unpleasing HF and we select the GMD rather than the GM as a discriminator. The GM discriminator focuses on the amount of gradient magnitude, similar to the spatial discriminator which focuses on the amount of HF. This may increase the GM component, but it also leads to unpleasing HF noise. By contrast, since the GMD shows how many components each gradient of an image has, it can be used as a measure of the SR performance. The GMD discriminator distinguishes between the ground-truth and SR, based on whether the probability of specific GM goes well with neighboring GM. This prevents unpleasing HF noise which

Table 2. Average PSNR, SSIM and PI comparisons.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR</td>
<td>27.99</td>
<td>32.08</td>
<td>28.96</td>
<td>29.35</td>
<td>29.17</td>
<td>29.62</td>
<td>29.54</td>
<td>29.50</td>
<td>30.66</td>
<td></td>
</tr>
<tr>
<td>SSIM</td>
<td>0.8972</td>
<td>0.9473</td>
<td>0.9035</td>
<td>0.9117</td>
<td>0.9120</td>
<td>0.9129</td>
<td>0.9140</td>
<td>0.9134</td>
<td>0.9175</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR</td>
<td>25.52</td>
<td>29.70</td>
<td>26.14</td>
<td>26.68</td>
<td>26.31</td>
<td>26.73</td>
<td>27.08</td>
<td>26.83</td>
<td>27.56</td>
<td></td>
</tr>
<tr>
<td>SSIM</td>
<td>0.8291</td>
<td>0.8975</td>
<td>0.8406</td>
<td>0.8471</td>
<td>0.8421</td>
<td>0.8455</td>
<td>0.8480</td>
<td>0.8398</td>
<td>0.8538</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR</td>
<td>28.68</td>
<td>33.35</td>
<td>27.11</td>
<td>27.64</td>
<td>29.57</td>
<td>29.83</td>
<td>30.05</td>
<td>27.48</td>
<td>29.89</td>
<td></td>
</tr>
<tr>
<td>SSIM</td>
<td>0.8716</td>
<td>0.9276</td>
<td>0.8804</td>
<td>0.8863</td>
<td>0.8896</td>
<td>0.8917</td>
<td>0.8906</td>
<td>0.8761</td>
<td>0.8916</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. MOS test from three perspectives.

<table>
<thead>
<tr>
<th>Network</th>
<th>The checkerboard artifacts</th>
<th>Unpleasing HF noise</th>
<th>Overall Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>SG [18]</td>
<td>3.73</td>
<td>3.83</td>
<td>5.53</td>
</tr>
<tr>
<td>SGw [54]</td>
<td>6.57</td>
<td>2.87</td>
<td>5.22</td>
</tr>
<tr>
<td>SG-D [20]</td>
<td>5.82</td>
<td>4.71</td>
<td>5.84</td>
</tr>
<tr>
<td>MPDSD</td>
<td>6.96</td>
<td>4.94</td>
<td>6.07</td>
</tr>
<tr>
<td>MPDw</td>
<td>4.35</td>
<td>5.98</td>
<td>5.92</td>
</tr>
<tr>
<td>MPDx [21]</td>
<td>4.23</td>
<td>3.86</td>
<td>6.26</td>
</tr>
<tr>
<td>MPDx [21]</td>
<td>7.11</td>
<td>6.15</td>
<td>7.23</td>
</tr>
</tbody>
</table>

Table 4. KL-Divergence comparisons.

<table>
<thead>
<tr>
<th>Network</th>
<th>HF reconstruction (D_{KL})</th>
</tr>
</thead>
<tbody>
<tr>
<td>SG [18]</td>
<td>0.05873</td>
</tr>
<tr>
<td>SGw [54]</td>
<td>0.05381</td>
</tr>
<tr>
<td>SG-D [20]</td>
<td>0.04634</td>
</tr>
<tr>
<td>MPDSD</td>
<td>0.04528</td>
</tr>
<tr>
<td>MPDw</td>
<td>0.04270</td>
</tr>
<tr>
<td>MPDx [21]</td>
<td>0.04635</td>
</tr>
<tr>
<td>MPDx [21]</td>
<td>0.03506</td>
</tr>
</tbody>
</table>

(MPDSD). In Figure 7, MPDSD and SG-D show less the checkerboard artifacts than SR-GAN. However, there still remain some artifacts for SG-D as shown in Figure 7 (c) and Figure 8 (d). The proposed SR-GAN with the DCT discriminator significantly reduces the artifacts on both the spatial and the frequency domains as confirmed in Figure 7 (d) and Figure 8 (f). The SG-D cannot sufficiently reduce the artifacts due to the use of MSE based pixel-wise loss. In SR-GAN, unlike the generator which just focuses on pixel accuracy, since the discriminator focuses on the amount of HF, discriminator can reconstruct more natural HF by its feedback on HF. Likewise, the DCT domain discriminator focuses on ultra-HF region through training and reduces the artifacts significantly. Meanwhile, since the DCT loss for generator just reflects pixel errors, it has performance limitation like spatial domain generator of SR-GAN. Moreover, since the DCT loss is the MSE of DCT coefficients, SG-D shows more blurred results than MPDSD. This means that the DCT discriminator is more effective in reducing the occurrence of the checkerboard artifacts than generator with DCT loss.
Figure 11. Subjective quality comparisons of 807 in Div2k.

Figure 12. Subjective quality comparisons of 861 in Div2k.

Figure 13. Subjective quality comparisons of 865 in Div2k.
Figure 14. DCT coefficients comparisons of 865 in Div2k.
occurs intensively at a specific frequency, and also makes the restored HF look natural.

Figure 9 shows the comparison of GMDs. Under small gradient magnitudes, the GMDs of SR-GAN and MPD$_{SD}$ are very similar. However, for the rest of it, we can see that the MPD$_{SG}$ is closer to the ground-truth. We can confirm from Figure 9 that the GMD discriminator has restored HFs more similar to the ground-truth successfully. The second column in Table 4, HF reconstruction ($D_{G}$), represents the results of quantitative measurement with KLD, which is an indicator of how much the HF is restored. In Table 4, the KLD value is the average of the div2k test images, with a smaller KLD indicating that the GMD distribution is closer to the original.

Comparing MPD$_{SG}$ with SG and MPD$_{SD}$ in Table 4 show that, its KLD is the smallest. This means that MPD$_{SG}$ can reconstruct more of the HF component making it closer to the original than SG and MPD$_{SD}$. The increased amount of HF can also be confirmed on the DCT domain, as shown in Figure 14. MPD$_{SG}$ restores more HF than MPD$_{SD}$ in the lower-right region of the DCT domain (the right region of the red line). These results ensure that the GMD discriminator is effective in sufficiently restoring HF. Figure 10 shows why we propose the GMD discriminator, not the GM discriminator. The design goal of the discriminator is to remove the unpleasing noise and refine the restored HF. The GM discriminator focuses on the amount of GM like the spatial discriminator. Thus, as shown in Figure 10 (a), although the amount of GM increases, unpleasing GM still occurs in the check-skirt, stair railing and building window. On the other hand, although the amount of increased GM in Figure 10 (b) is less than that in Figure 10 (a), its results have less unpleasing GM and look more natural. It is thought that the distribution of gradient magnitudes is more suitable for HF quality discrimination rather than gradient magnitude.

### C. VISUAL COMPARISON

Table 2 shows the quantitative evaluation comparisons, where the best values are marked in bold. Three notable points can be derived from Table 2. The first is the superiority of the proposed method in terms of quantitative evaluation. We can see that for all test images, SR-CNN has the highest PSNR and SSIM because of the use of the MSE loss. However, its perceptual quality is very low (i.e., it has the highest perceptual index (PI)), which is attributed to the severe blur degradation. The blurry phenomenon of SR-CNN can be observed in Figures 11, 12, and 13. On the other hand, aside from SR-CNN, the proposed method (MPD$_{SDG}$) has not only the highest PSNR and SSIM but also the lowest PI for most test images. The second is the use of DCT loss for both the generator and discriminator. In MPD$_{SD}$, DCT loss is adopted in the discriminator, while in existing SG-D, it is adopted in the generator. As confirmed in Table 2, MPD$_{SD}$ achieves higher visual quality than SG-D. Moreover, its capability to recover HF is also superior to that of SG-D, as shown in Table 4. Thus, it is experimentally confirmed that using the DCT discriminator is more effective than using the DCT loss for the generator. The last point that should be noted is the combination of each discriminator on the proposed multi-perspective GAN. Although the DCT discriminator reduces the occurrence of the checkerboard artifacts, it cannot sufficiently reconstruct the HF. Thus, MPD$_{SD}$ has higher PI than MPD$_{SG}$ and MPD$_{SV}$. The VGG feature-based discriminator can reproduce more structural HF than the spatial discriminator. However, using this discriminator also leads to hallucination artifacts and unpleasing HF noise near the edge as illustrated in Figures 11, 12, and 13. As a result, MPD$_{SV}$ has lower PI than MPD$_{SD}$ while having lower PSNR than MPD$_{SD}$. On the other hand, since MPD$_{SG}$ reconstructs more natural HF without artifacts, it has higher PSNR and SSIM with lower PI than MPD$_{SV}$.

In Table 2, MPD$_{SDG}$ achieves better PI than SG in Set5 and Set14, but achieves worse PI in div2k. This is related to the ultra-HF components that the images have. MPD$_{SD}$ reduces the checkerboard artifacts which are easily separated on the DCT domain. This is very effective for images which have less ultra-high frequency components such as Set5, Set14. As a result, MPD$_{SD}$ achieves better PI in both Set5 and Set14. However, when some images in div2k have ultra-high frequency components that overlap with those of the checkerboard artifacts, the ultra-HFs are also removed with the checkerboard artifacts together. This is why MPD$_{SD}$ obtains worse PI than SG in div2k. Nonetheless, MPD$_{SD}$ has better HF reconstruction than SG due to the reduction of the checkerboard artifacts as confirmed in Table 4. The reason that MPD$_{SDG}$ achieves worse PI and PSNR than MPD$_{SD}$ in div2k is similar to the reason that MPD$_{SD}$ achieves worse PI in div2k. Despite the use of the GMD discriminator for restoring HF, due to the use of the DCT discriminator, MPD$_{SDG}$ cannot sufficiently reconstruct the HF in the div2k data-set. Nonetheless, we can confirm that the GMD shape of MPD$_{SDG}$ is closer to ground-truth than that of MPD$_{SG}$ in Table 4, since both checkerboard and unpleasing HF artifacts are reduced well. As PSNR is not always directly proportional to SSIM, PI is not always proportional to $D_{KL}$ of GMD. For experimental performance verification, we conduct the MOS test, which is a subjective experiment by a group of human observers.

Table 3 shows MOS test comparisons from the three perspectives: the checkerboard artifacts, unpleasing HF noise and overall performance. The MOS test ranges from 0 (the lowest quality) to 10 (the best quality). The score regarding the checkerboard artifacts shows that the use of DCT is effective for reducing the checkerboard artifacts. Moreover, we can confirm once again that the DCT discriminator is more effective in reducing the checkerboard artifacts than the generator with DCT loss. The score regarding unpleasing HF shows that only the GMD discriminator is suitable for preventing the occurrence of unpleasing HF noise. For the overall performance score, the proposed method obtains an overwhelming score. Since SG has both artifacts, it receives
a low score. SG_W shows that the checkerboard artifacts are reduced by MSE. As a result, it obtains a high score on the checkerboard artifacts. However, due to severe artifact near the edge, SG_W achieves the lowest score in unpleasing HF. Despite the high score in the checkerboard artifacts, SG_W gets the lowest score due to severe unpleasing HF noise. MPD_{SG} reduces the checkerboard artifacts, but has unpleasing HF noise. By contrast, MPD_{SG} reduces the unpleasing noise, but has the checkerboard artifacts. Thus, both MPD_{SG} and MPD_{SG} cannot receive a high score. Despite having both checkerboard and unpleasing HF artifacts, MPD_{SV} receives the highest score among any methods except for the proposed method due to the largest amount of the reconstructed HF. And finally, the proposed method receives score about one point higher than MPD_{SV}, because it reduces both artifacts, even though the amount of reconstructed HF is slightly less than MPD_{SV}.

In summary, using the DCT feature reduces the occurrence of the checkerboard artifacts. However, using it for the discriminator is more effective than using it for the generator in terms of HF reconstruction. The GMD discriminator prevents the occurrence of unpleasing HF noise. And the simultaneous use of both GMD and DCT discriminators reduces both artifacts and makes the reconstructed HF closer to ground-truth.

Figures 11-13 show comparisons of the proposed method with the existing methods described in [7, 9, 18, 21, 54]. In Figures 11-13, the SS-SR results [7] show the limits of non-deep-learning-based approaches. They have severe blurry artifacts and fail to restore the edge structure. Although the SRCNN [9] shows better results than non-deep-learning-based approaches, they still suffer from blurry artifacts caused by MSE loss. SG shows less blurred results than SRCNN, but generates the checkerboard artifacts, unpleasing HF and hallucination artifact. WGAN leads to more enhanced HF than GAN. This means that WGAN increases the amount of HF, but results in severe artifacts near the edge. MPD_{SV} achieves results that are substantially less blurred, because using the VGG discriminator improves the perceptual quality. However, it has no effect on the checkerboard artifacts. Moreover, unpleasing HF and hallucination artifacts are severe. On the other hand, the proposed MPD_{SGD} reduces the occurrence of the checkerboard artifacts by adopting the DCT discriminator, and suppresses incongruous and unpleasing HF by adopting the GMD discriminator. As a result, MPD_{SGD} shows clearer and more natural visual quality.

As shown in Figure 11, SG, MPD_{SG}, and MPD_{SV} generate the checkerboard artifacts. By contrast, the methods using DCT feature loss such as MPD_{SG} and MPD_{SG} show artifact-less results. MPD_{SV} looks less blurred and clearer, but has hallucination artifact near the branch (it can be observed more easily by zooming in on the monitor). Unlike the spatial discriminator, which induces the generator to increase the amount of HF, the VGG discriminator induces the generator to increase the amount of structural HF until the results look similar to the original. In this process, the sharpness of the results is improved by increasing the structural HF. However, the VGG discriminator has a problem similar to that of the spatial discriminator. In the same way as the spatial discriminator was fooled by arbitrary HF noise, the VGG discriminator was fooled by arbitrary structural HF noise. This led to the occurrence of hallucination artifacts. On the other hand, in the proposed method, the DCT discriminator distinguishes the SR results according to the existence of artifacts and the GMD discriminator focuses on the naturalness of the gradient from a probabilistic perspective. Therefore, it is difficult to fool them by arbitrary noise unlike the spatial and VGG discriminator. This leads to an artifact-less result. The proposed method significantly alleviates the occurrence of the checkerboard artifacts in the lake region and the occurrence of hallucination artifacts in the branch shown in Figure 11.

Figure 12 confirms the superiority of the proposed methods in terms of HF restoration. SG and MPD_{SV} fail to reconstruct the HF in a manner similar to the ground-truth in the vertical edge of the building. SG_W reconstructs rich amount of HF, but there are a lot of noises near the edge of building than other methods. MPD_{SG} reconstructs the HF closer to ground-truth than SG and MPD_{SV}, but there are noises in the vertical edge of the building and the horizontal edge of the top roof. It should be noted that the proposed method reconstructs the edge of the building better without artifacts in the wall and roof of the building.

Figure 13 shows the unpleasing HF noise as well as the checkerboard artifacts. SG, MPD_{SG}, and MPD_{SV} have the checkerboard artifacts in the sky region. The tower of SG looks blurred and artifacts are shown near the edge. Unlike the SG, MPD_{SV} shows less blurry artifacts, but it has remarkable unpleasing noise due to the disadvantage of the VGG discriminator. By contrast, the proposed method achieves enhanced clarity without artifacts, unlike the other results.

Finally, Figure 14 compares the DCT coefficients of the SR outputs. For convenience, enlarged images of the yellow boxes are added to the center. As described earlier, the yellow box of ground-truth is clean, but that of SG is not. We can see in the yellow boxes that the methods with DCT features are cleaner than those without DCT features. Regarding of SG, although it does not use the DCT features, the checkerboard artifacts are reduced well due to MSE loss. However, we can see from Figure 14 that the DCT coefficients are newly created at a specific region unlike the ground-truth. This means that unpleasing HF noise such as artifact near the edge occurs intensively. This results in a serious degradation of the subject quality.

In Figure 14, a red line is drawn diagonally for the purpose of easy comparison. The lower-right region below the red line represents HF components, and its high
intensities indicate more reconstruction of HF. As illustrated in Figure 14, the HF components below the red line are not only almost close to zero for SRCNN, but they also even above the red line. This indicates that the HF component has been reconstructed very poorly. By checking that MPD$_{SG}$ has more HFs below the red line than MPD$_{SD}$, we can see that the GMD discriminator is more appropriate for restoring HF than DCT. It is particularly worth noting that the yellow box of the proposed is the cleanest while achieving abundant HFs (as confirmed by HFs being below the red line).

V. DISCUSSION AND CONCLUSION

We propose a novel multi-perspective discriminators based SR-GAN. It is very difficult to simply evaluate SR performance using a single measure. The visual quality of the SR image should be measured from multi-perspectives such as the extent of HF restoration, blur, and artifact. These quality factors cannot be discriminated with a single measure. This motivates us to propose multi-perspective discriminators, which measure the visual quality from a variety of aspects. The conventional SR-GAN exhibits certain side-effects, such as occurrence of the checkerboard artifacts, and unpleasing HF. In order to overcome these drawbacks, we propose the two discriminators for SR-GAN in addition to the existing spatial one: DCT and GMD discriminators. The DCT discriminator is useful for reducing the checkerboard artifacts because it is easily separated on the frequency domain. The GMD discriminator employs gradient distribution, which is effective for the quality measurement of HF. Thus, it could reduce the unpleasing HF noise and improve the restoration of HF that is congruous with the image. The experimental results demonstrate that the proposed framework achieves improved SR image quality in terms of both objective and subjective quality such as PSNR, SSIM, PL, KL of GMD and MOS.

In this paper, we experimentally showed that multiple discriminators are effective in removing artifacts caused by SR-GAN. Future research should include further study of a novel perspective discriminator, which is motivated by existing generator loss such as texture and consistence loss. Texture loss is also known as style loss. Each image has its own style such as sky, forest, sea and so on. The discriminator on the style domain may help the generator produce the HF beyond the structural HF, which looks similar to the style of the ground-truth. SR-GAN tends to produce results similar to the original but, not equal to the original due to the characteristic of adversarial loss. The use of a consistence domain discriminator may be the solution for this problem. The multiple discriminators based GAN method can be expanded upon for image restoration, such as image denoising. Such methods can be effective from various perspectives such as noise reduction, the generation of realistic textures and so on.

REFERENCES


This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2019.2942779, IEEE Access