electronics

Article

PSRGAN: Perception-Design-Oriented Image Super Resolution Generative Adversarial Network

Tao Wu [1,†], Shuo Xiong [2,*,†], Hui Liu [1], Yangyang Zhao [3,*], Haoran Tuo [1], Yi Li [1], Jiaxin Zhang [1] and Huaizheng Liu [1]

[1] National Model Software Institute, Huazhong University of Science and Technology, Wuhan 430074, China; wutao1972@hust.edu.cn (T.W.); liuh_@hust.edu.cn (H.L.); thr@hust.edu.cn (H.T.); liyi99@hust.edu.cn (Y.L.); jiaxinzhang@hust.edu.cn (J.Z.); liuhuaizheng@hust.edu.cn (H.L.)
[2] PSS Lab of Big Data and National Communication Strategy, MOE, Huazhong University of Science and Technology, Wuhan 430074, China
[3] Hytera Communication Co., Ltd., Shenzhen 518057, China
* Correspondence: xiongshuo@hust.edu.cn (S.X.); yangyangzhao0803@gmail.com (Y.Z.)
† These authors contributed equally to this work.

Abstract: Among recent state-of-the-art realistic image super-resolution (SR) intelligent algorithms, generative adversarial networks (GANs) have achieved impressive visual performance. However, there has been the problem of unsatisfactory perception of super-scored pictures with unpleasant artifacts. To address this issue and further improve visual quality, we proposed a perception-design-oriented PSRGAN with double perception turbos for real-world SR. The first-perception turbo in the generator network has a three-level perception structure with different convolution kernel sizes, which can extract multi-scale features from four $\frac{1}{4}$ size sub-images sliced by original LR image. The slice operation expands adversarial samples to four and could alleviate artifacts during GAN training. The extracted features will be eventually concatenated in later 3×2 upsampling processes through pixel shuffle to restore SR image with diversified delicate textures. The second-perception turbo in discriminators has cascaded perception turbo blocks (PTBs), which could further perceive multi-scale features at various spatial relationships and promote the generator to restore subtle textures driven by GAN. Compared with recent SR methods (BSRGAN, real-ESRGAN, PDM_SR, SwinIR, LDL, etc.), we conducted an extensive test with a $\times 4$ upscaling factor on various datasets (OST300, 2020track1, RealSR-Canon, RealSR-Nikon, etc.). We conducted a series of experiments that show that our proposed PSRGAN based on generative adversarial networks outperforms current state-of-the-art intelligent algorithms on several evaluation metrics, including NIQE, NRQM and PI. In terms of visualization, PSRGAN generates finer and more natural textures while suppressing unpleasant artifacts and achieves significant improvements in perceptual quality.

Keywords: perception design; image super resolution; generative adversarial network; artifact suppression; intelligent computing

check for updates

Citation: Wu, T.; Xiong, S.; Liu, H.; Zhao, Y.; Tuo, H.; Li, Y.; Zhang, J.; Liu, H. PSRGAN: Perception-Design-Oriented Image Super Resolution Generative Adversarial Network. *Electronics* **2023**, *12*, 4420. https://doi.org/10.3390/electronics12214420

Academic Editor: Silvia Liberata Ullo

Received: 14 September 2023
Revised: 13 October 2023
Accepted: 17 October 2023
Published: 27 October 2023

1. Introduction

Single-image super-resolution (SISR) aims to reconstruct a high-resolution (HR) image from a low-resolution (LR) one. The traditional methods for solving the SR problems are mainly interpolation-based methods [1–4] and reconstruction-based methods [5–7]. Intelligent computing has also been applied in the field of image super-resolution. Super-resolution methods based on genetic algorithms, guided by imaging models, utilize optimization techniques to seek the optimal estimation of the original image. At its core, this approach transforms the problem of reconstructing multiple super-resolved images into a linear system of equations. The convolutional neural network (CNN) has greatly promoted the vigorous development of SR field and demonstrates vast superiority over traditional methods. The main reason it achieves good results is due to its strong capability of learning

rich features from big data in an end-to-end manner [8]. CNN-based SR methods often use PSNR as the evaluation metric; although some SR methods achieve good results for PSNR, it is still not completely satisfactory in terms of perception.

The generative adversarial network (GAN) [9] has achieved impressive visual performance in the field of super-resolution (SR) since the pioneering work of SRGAN [10]. GANs have proven their capability to generate more realistic images with high perceptual quality. In pursuit of further enhancing visual quality, Wang et al. proposed ESRGAN [11]. Given the challenges of collecting well-paired datasets in real-world scenarios, unsupervised GANs have been introduced [12,13]. BSRGAN [14] and real-ESRGAN [15] are dedicated to simulating the practical degradation process to obtain better visual results on real datasets.

However, perceptual dissatisfaction accompanied by unpleasant artifacts still exists in GAN-based SR models because of insufficient design in either generators or discriminators. In GAN-based SR methods, it is obvious that the decisive capability to recover naturally finer textures in generators is dependent largely on the guidance of discriminators through GAN training, but discriminators are usually cloned from well-known networks (U-net [16], VGG [17], etc.) suitable for image segmentation or classification, which might not fully lead generators to restore subtle textures in SR. Moreover, the design of generators should be perceptive enough to extract multi-scale image features from low-resolution (LR) images and mitigate artifacts.

Research hypotheses and questions: Perceived quality improvement: How can we design a network structure of PSRGAN to suppress artifact generation in images, and how can we achieve the effect of suppressing artifacts? Generative adversarial network image quality assessment: Which evaluation metrics are used to assess the generated images to ensure their perceived quality is enhanced? Adversarial training stability: How can we ensure the stability and convergence of our PSRGAN training? To address these issues and further improve the visual quality of the restored SR images, we redesigned both generators and discriminators; the contributions of this paper are mainly in four aspects:

- We present a novel perception-design-oriented PSRGAN with double perception turbos, which can generate real-world SR images with naturally finer textures while suppressing unpleasant artifacts by $\times 4$ upscaling factors (see Figure 1).
- We design the first-perception turbo in the generator network, characterized by slice operation and a three-level perception structure, which can extract multi-scale features from sliced sub-images and mitigate artifacts.
- We propose the second-perception turbo in the discriminator network with cascaded perception turbo blocks, which can further promote the generator to restore subtle textures.
- We demonstrate that the proposed PSRGAN has achieved state-of-the-art perceptual capabilities calculated by NIQE, NRQM, and PI.

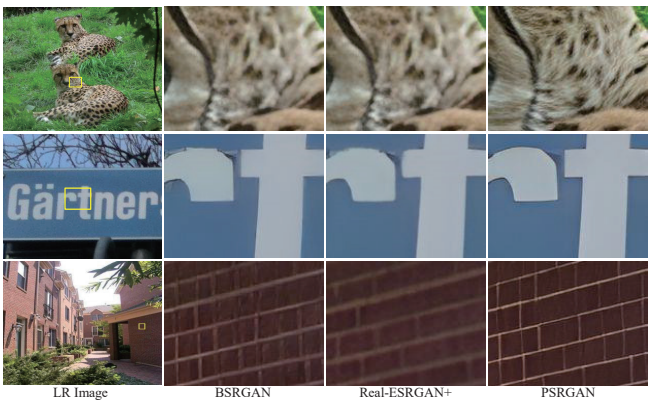

Figure 1. Comparisons of visual quality among BSRGAN [14], real-ESRGAN+ [1], and PSRGAN on real-life images by $\times 4$ upscaling. The PSRGAN can generate naturally finer textures and remove or alleviate annoying artifacts for real-world images. Zoom in for best view.

2. Related Work

Single-image super-resolution: SRCNN [18] is the first method to apply deep learning to SR reconstruction, and a series of learning-based works are subsequently proposed [19–23]. ESPCN [24] introduces an efficient sub-pixel convolution layer to perform the feature extraction stages in the LR space instead of HR space. VDSR [19] uses a very deep convolutional network. EDSR [25] removes the batch normalization layers from the network. SRGAN [10] first uses the GAN network for the SR problem and proposes perceptual loss, including adversarial loss and content loss. Based on human perceptual characteristics, the residual in the residual dense block strategy (RRDB) is exploited to implement various depths in network architectures [11,26]. ESRGAN [11] introduces the residual-in-residual dense block (RRDB) into the generator. RealSR [27] estimates various blur kernels and real noise distributions to synthesize different LR images. CDC [28] proposes a divide-and-conquer SR network. Luo et al., in [29], propose a probabilistic degradation model (PDM). Shao et al., in [30], propose a sub-pixel convolutional neural network (SPCNN) for image SR reconstruction.

Perceptual-driven approaches: The PSNR-oriented approaches lead to overly smooth results and a lack of high-frequency details, and the results sometimes do not agree with the subjective human perception. In order to improve the perceptual quality of SR results, the perceptual-driven approach is proposed. Based on the idea of perceptual similarity [31], Li Feifei et al. propose perceptual loss in [32]. Then, textures matching loss [33] and contextual loss [34] are introduced. ESRGAN [11] improves the perceptual loss by using the features before activation and wins the PIRM perceptual super-resolution challenge [35]. Christian Szegedy et al. propose inception [36], which can extract more features with the same amount of computation, thus improving the training results. For the purpose of extracting multi-scale information and enhance the feature discriminability, RFB-ESRGAN [8] applies the receptive field block (RFB) [37] to super resolution and wins the NTIRE 2020 perceptual extreme super-resolution challenge. There is still plenty of room for perceptual quality improvement [38].

The design of discriminator networks: The discriminator in SRGAN is VGG-style, which is trained to distinguish between SR images and GT images [10]. ESRGAN borrows ideas from relativistic GAN to improve the discriminator in SRGAN [11]. Real-ESRGAN improves the VGG-style discriminator in ESRGAN to an U-Net design [15]. In [39], Alejandro et al. propose a novel convolutional network architecture named "stacked hourglass", which captures and consolidates information across all scales of the image. Inspired by [39], we propose a new discriminator structure, which can guide the generator to recover finer textures. All the related work as Table 1 shows.

Table 1. Related work on design of discriminator networks.

Different Methods	Design of Discriminator Networks
SRGAN	VGG-style, which is trained to distinguish between SR images
ESRGAN	borrows ideas from relativistic GAN to improve the discriminator in SRGAN
Real-ESRGAN	proposed an U-Net design
RFB-ESRGAN	proposed stacked hourglass network which captures and consolidates information across all scales of the image

Artifact suppression: The instability of the training of GANs often leads to the introduction of many perceptually unpleasant artifacts while generating details in the GAN-based SR networks [40]. There have been several SR models focusing on solving the problem. Zhang et al. propose a supervised pixel-wise generative adversarial network (SPGAN) to obtain higher-quality face images [41]. Gong et al., in [42], overcome the effect of artifacts in the super-resolution of remote sensing images using self-supervised hierarchical perceptual loss. Real-ESRGAN uses spectral normalization (SN) regularization to stabilize

the training dynamics [15]. We propose a algorithm named "image slice and multi-scale feature extraction", which can generate more delicate textures and suppress artifacts.

The evaluation metrics: The DCNN-based SR approaches have two main optimization objectives: the distortion metric (e.g., PSNR, SSIM, IFC, and VIF [43–45]) and perceptual quality (e.g., the human opinion score; no-reference quality measures such as Ma's score [46], NIQE [47], BRISQUE [48], and PI [49]) [50]. Yochai et al. in [49] have revealed that distortion and perceptual quality are contradictory and there is always a trade-off between the two. Algorithms that are superior in terms of perceptual quality tend to be poorer in terms of, e.g., PSNR and SSIM. However, sometimes there is also inconsistency between the results observed by human eyes and these perceptual quality metrics. Because the no-reference metrics do not always match perceptual visual quality [51], some SR models such as SRGAN perform mean-opinion-score (MOS) tests to quantify the perceptual ability of different methods [10]. We use NIQE, NRQM, and PI as our image quality metrics, which do not depend on the GT image to measure the perceptual quality of the reconstructed image [52]. The related work on evaluation metrics as Table 2 shows.

Table 2. Related work on evaluation metrics.

Evaluation Metrics	Advantage	Disadvantage
Distortion metrics	Simple calculation	Greater inconsistency with perceived quality
Human opinion score	Consistent with visual perception	High labor costs
No-reference quality measures	Balancing consistency with perceived quality and computational cost	There is some inconsistency with visual perception

The transformer: Vaswani et al. in [36] propose a new simple network architecture, transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Transformer continues to show amazing capabilities in the NLP domain. Many researches have started to try to apply the powerful modeling ability of transformer to the field of computer vision [53]. In [54], Yang et al. propose TTSR, in which LR and HR images are formulated as queries and keys in transformer, respectively, to encourage joint feature learning across LR and HR images. Swin transformer [55] combines the advantages of convolution and transformer. Liang et al. in [56] propose SwinIR based on Swin transformer. Vision transformer is computationally expensive and consumes high GPU memory, so Lu et al. in [57] propose ESRT, which uses efficient transformers (ET), a lightweight version of the transformer structure.

3. Proposed Methods

To further improve perceptual quality as well as mitigate artifacts in SISR, we proposed a novel perception-design-oriented super resolution generative adversarial network (PSRGAN) with double perception turbos. In this section, we first introduce the generator network-containing first-perception turbo (GPT) and then describe the construction of the discriminator network with the second-perception turbo (DPT). At last, we discuss the perceptual loss function used.

3.1. Generator Network

The generator network consists of two components: first-perception turbo, and the feature blending and upsampling component (FBUC) as shown in Figure 2.

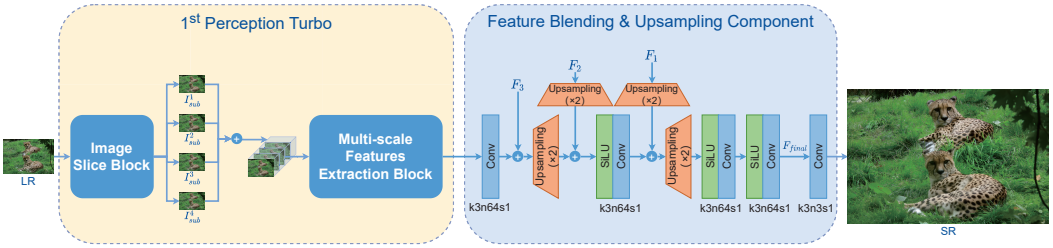

Figure 2. Architecture of generator network with corresponding kernel size (k), number of feature maps (n), and stride (s) indicated for each convolutional layer, where F_1, F_2, and F_3 are multi-scale features extracted by MFEB described in Figure 3.

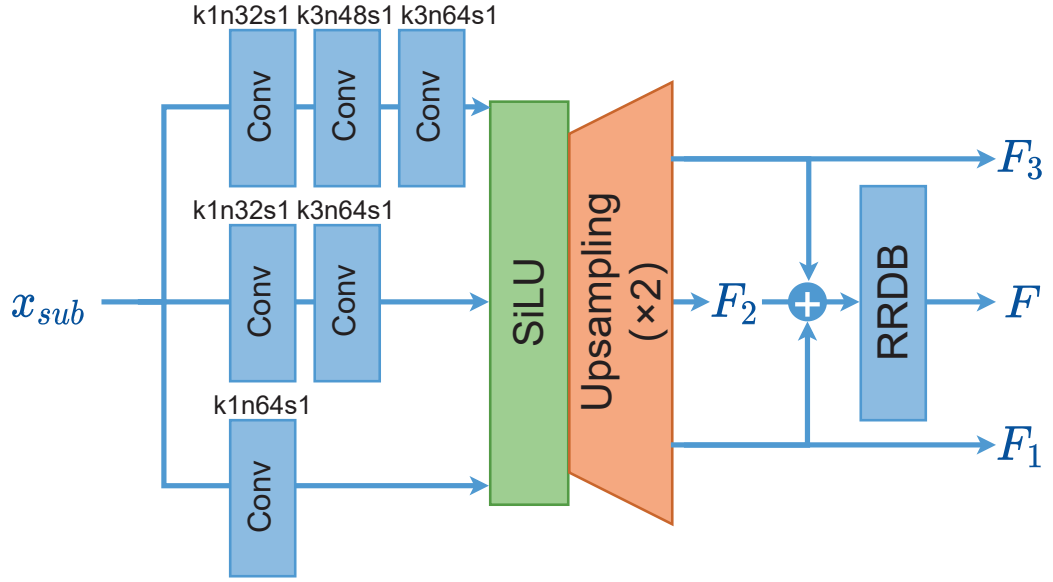

Figure 3. Design of MFEB in first-perception turbo.

The first perception turbo has two major blocks: the image slice block (ISB) and the multi-scale feature-extraction block (MFEB). The image slice block (ISB) produces four $\frac{1}{4}$ size sub-images (I^1_{sub}, I^2_{sub}, I^3_{sub}, and I^4_{sub}) from the low-resolution image I^{LR} via pixel reassembly. Specifically, suppose I^{LR} has the resolution of $2m \cdot 2n$ pixels or padding to $2m \cdot 2n$ pixels; the sliced sub-images are $m \cdot n$ pixels. If the upper left pixel is denoted as $(0,\ 0)$, and the lower right pixel is denoted as $(2m-1, 2n-1)$, the relationship of the pixels between I^{LR} and the sub-images can be formulated as below.

$$\begin{cases} I^{LR} = \{(k,\ t) \mid 0 \le k < 2m,\ 0 \le t < 2n,\ k,\ t\ \in\ N \cup 0\} \\ I^1_{sub} = \{(2p, 2q) \mid 0 \le p < m,\ 0 \le q < n,\ p,\ q\ \in\ N \cup 0\} \\ I^2_{sub} = \{(2p+1, 2q) \mid 0 \le p < m,\ 0 \le q < n\} \\ I^3_{sub} = \{(2p, 2q+1) \mid 0 \le p < m,\ 0 \le q < n\} \\ I^4_{sub} = \{(2p+1, 2q+1) | 0 \le p < m,\ 0 \le q < n\} \end{cases} \tag{1}$$

The slice method above has the following characteristics:

- The slice splits the LR image to multiple detail adversarial sub-images while preserving the pixel integrity of the LR image.
- The subsequent MFEB could extract multi-scale features from smaller adversarial samples; thus, the generator is capable of generating diverse and delicate textures.
- The slice weakens the correlations among noisy pixels in I^{LR}, which can effectively reduce noises and further alleviate artifacts in the restored SR image. Although the correlations among adjacent pixels might be also impaired, the meaningful semantic features will be eventually recovered in the SR image through GAN training.

The multi-scale feature extraction block (MFEB, Figure 3): It has been proven that each learned filter has its specific functionality and that a reasonably larger filter size could grasp richer structural information, which in turn could lead to better results [18]. The MFEB is perceptually designed to extract diverse image features from the LR image by three groups of convolutional layers inspired by inception networks [36], as depicted in Figure 3. Please refer to Appendix B for more detail.

The first convolution group has a tiny receptive field, used to retain micro subtle features, denoted as k1-n64-s1.

The second convolution group has a medium receptive field, used to capture moderate features, denoted as k1-n32-s1, k3-n64-s1.

The third convolution group has a large receptive field, used to seize macro features, denoted as k1-n32-s1, k3-n48-s1, k3-n64-s1.

The outputs of the three convolution groups are activated using the Sigmoid weighted liner unit (SiLU) and then $\times 2$ upsampled via pixel-shuffle to obtain multi-scale features F_1, F_2, F_3. The process can be formulated as:

$$F_i = [SiLU(Convs_i(x_{sub}))] \uparrow_s, i \in \{1, 2, 3\}. \tag{2}$$

where $Convs_i(x_{sub}), i \in \{1, 2, 3\}$ denote the three convolution groups, SiLU is the activation function, $\uparrow$ denotes upsampling, s denotes the scale factor and $s = 2$ in this block, and $F_i, i \in \{1, 2, 3\}$ indicate the 3-scale feature maps extracted. Subsequently, the obtained feature maps F_1, F_2, F_3 are added in the channel dimension as input, residual in residual dense block (RRDB) [11] is adopted to further capture semantic information and improve the recovered textures, and the output is denoted as F. The formal processing in the first-perception turbo is described in Algorithm 1.

Algorithm 1 Image slice and multi-scale feature extraction

Input: LR images set $\mathcal{X}$.
Output: Multi-scale features F_1, F_2, F_3, deeper features F.
1: **for all** I_{LR} such that $I_{LR} \in \mathcal{X}$ **do**
2: generate I^1_{sub},I^2_{sub},I^3_{sub},I^4_{sub} through slice operation from I_{LR}.
3: Get x_{sub} by merging the four sub-images I^1_{sub},I^2_{sub},I^3_{sub},I^4_{sub} in color channel dimension.
4: **for all** i such that $1 \leq i \leq 3$ **do**
5: input x_{sub} to $Convs_i$,$SiLU$,$2UP$ obtain F_i,
6: **end for**
7: generate $F = RRDB(F_1 + F_2 + F_3)$,
8: **end forreturn** F_1, F_2, F_3, F.

Feature blending and upsampling component (FBUC, Figure 2): The FBUC reassembles the obtained multi-scale features to generate the corresponding I^{SR} counterpart of I^{LR}. In the upsampling phase, the FBUC upsamples I^{LR} with diversfied features F as the input via pixel shuffle and gradually blends the features extracted by the MFEB. The upsampling process can be formulated as follows:

$$F_{final} = f_{Conv-SiLU}(f_{Conv-SiLU}(f_{Conv-SiLU}((F + F_3) \uparrow_s + (F_2) \uparrow_s) + (F_1) \uparrow_s) \uparrow_s) \tag{3}$$

where '+' denotes concatenation operation, $\uparrow$ denotes upsampling, s denotes the scale factor, and $s = 2$. $f_{Conv-SiLU}$ denotes one convolutional kernel, SiLU is the activation function, and F_{final} denotes the final features obtained from the FBUC. F_{final} is passed through a triple convolutional layer with the kernel size of 3×3 and finally outputs I^{SR}, which is $\times 4$ upscaling according to the original I^{LR}.

3.2. Discriminator Network

We proposed a novel discriminator containing the pre-processing block, cascaded perception turbo blocks (PTBs), and the post-processing block. The structure of the discriminator is depicted in Figure 4.

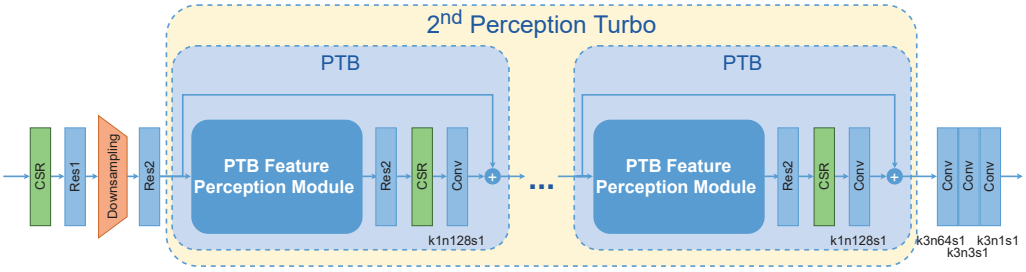

Figure 4. Discriminator network structure with second-perception turbo. The structure of CSR, Res1, and Res2 are shown in Figure 5.

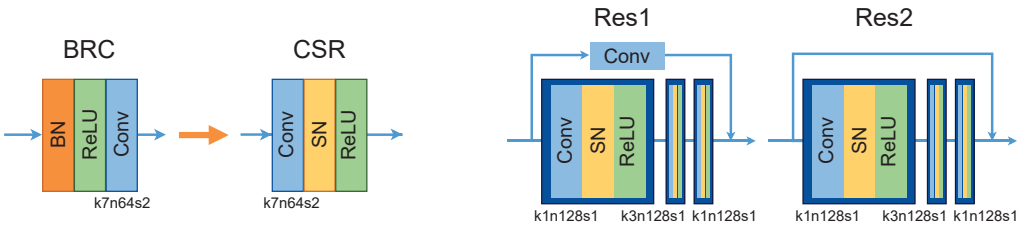

Figure 5. (**Left**): Differences between BRC and CSR; (**right**): structure of Res1 and Res2 in PTBs.

The pre-processing block is utilized for the initial feature perception of I^{SR} and I^{HR}. As shown in Figure 4, it includes a CSR block, two residual blocks, and a downsampling layer. The CSR block consists of a convolution layer, an SN layer, and a ReLU activation function. The specific structure of the two residual blocks Res1 and Res2 is shown in the Figure 5.

The second-perception turbo is the core structure of this discriminator, which consists of cascaded PTBs. In order to further promote the generator to restore subtle textures, we proposed the PTB structure and made the following four improvements on the basis of hour-glass module [39]:

- As shown in the Figure 5, we adopt the CSR structure instead of BRC, which consists of the BN layer, the ReLU activation function, and the convolutional layer. It has been proven that removing the BN layers can prevent BN artifacts of SR images, improve the performance, and reduce the computational complexity in the SR task [25]. In addition, we improve the perceptual loss by using the features before activation, which could provide stronger supervision for brightness consistency and texture recovery [11].
- In the upsampling procedure, we use pixel-shuffle instead of nearest neighbor interpolation, which may lose pixel information.
- In the downsampling layer, we use convolution instead of Maxpool2d operation, which may lose the integrity of feature map.
- We enlarge the input channels of PTB to 128, which improves the perceptive capabilities of the discriminator.

The post-processing block consists of three convolutional layers to further learn features and output a feature map that benefits the computation of adversarial loss.

Based on the above improvements, the discriminator could further perceive multi-scale features at various spatial relationships and promote the generator to restore subtle textures driven by GAN.

3.3. Perception Loss

We introduced the loss function similar to ESRGAN, which is a hybrid weighted loss function that takes into account pixel-level recovery and visual perception effects and is

able to achieve better super-resolution quality. Therefore, the total loss function of the generator L_G is a weighted combination of several losses: the adversarial loss L_{GAN}, pixel loss L_{Pixel}, and perceptual loss L_{Percep}. The loss function of the discriminator L_D is the adversarial loss L_{GAN}. The L_G is described as follows:

$$L_G = \alpha L_{GAN} + \beta L_{Pixel} + \gamma L_{Percep} \tag{4}$$

where $L_{Pixel} = E_{x_i} \parallel G(x_i) - y \parallel_1$ is the one-norm distance between the recovered image $G(x_i)$ and HR image y; it thus evaluates the average degree of approximation of I^{SR} and I^{HR} over pixels. α, β, γ are coefficients to balance different loss terms. Moreover, L_{Percep} is gained by introducing a fine-tuned VGG19 network to calculate the one-norm distance between the recovered image $G(x_i)$ and high-level features of y. It is used to evaluate the approximation of I^{SR} and I^{HR} in human perception. The perceptual loss is calculated as follows:

$$L_{Percep} = \mathbb{E}_{x_i} \parallel VGG(G(x_i)) - VGG(y) \parallel_1 \tag{5}$$

L_{GAN} aims to distinguish the SR image from the HR image by the superior perceptive capability of the discriminator, which could help to learn sharper edges and more detailed textures; it can be formulated as follows:

$$L_{GAN} = -\mathbb{E}_{x_{hr}}[log(1 - D(x_{hr}, x_{sr}))] - \mathbb{E}_{x_{sr}}[log(D(x_{sr}, x_{hr}))] \tag{6}$$

4. Experiments

In this section, we will discuss our PSRGAN model trained in RGB three channels.

4.1. Training Details

The experiments are performed with a scaling factor of $\times 4$ between LR and HR images; we obtain corresponding four-times smaller LR images by degrading the HR pictures, which are cropped to size 400×400 using the high-order [15] algorithm. Meanwhile, the patch size of cropped HR is 256×256, and the patch size of LR is 64×64. When training, the batch size is set to 12×2, which means that we use two GPUs and the batch size per GPU is 12.

The training process is divided into two stages. One is the pre-training generator, and the other is conducting GAN training combined with the generator and discriminator. First, in the pre-training process, we purely train the generator with the L1 loss. The learning rate is 2×10^{-4}, and the sum of the iteration is 0.4 million. Then, we employ the pre-training generator model as an initialization for the generator. The GAN is trained with a combination of L1 loss, perception loss, and GAN loss, with weights of 1, 1, and 0.1, respectively. The learning rate is set to 1×10^{-4} for both the generator and discriminator, and the sum of iteration is 0.28 million. Pre-training with L1 loss is beneficial to obtain more visually pleasing results by avoiding undesired local optima for the generator. Moreover, it can help the discriminator to distinguish more on the textures part so that the discriminator can receive relatively better super-resolved images during GAN training.

For optimization, we use Adam [58] with $\beta 1 = 0.9$, $\beta 2 = 0.99$. We alternately update the generator and discriminator network until the model converges. We implement our models with the PyTorch framework and train them using NVIDIA GeForce RTX 3090 GPUs.

4.2. Data

For training, we use the DIV2K dataset [59], the Flickr2K dataset [21], and the OutdoorSceneTraining(OST) dataset [60] as training datasets. We employ these large datasets with rich textures, which help to generate SR pictures with more natural and subtle textures [11].

We evaluate our models on widely used benchmark datasets, including OST300 [60], PIRM_Self_val [35], 2020track1 [51], RealSR-Canon [61], DRealSR_Test_x4 [28], and RealSR-

Nikon [61]. In particular, the images from RealSR-Canon and RealSR-Nikon are the center subimages of original images, and those larger than $1K \times 1K$ are cropped to $1K \times 1K$.

4.3. Qualitative Results

Due to the accessibility of SR methods, we compare our PSRGAN with several state-of-the-art methods, including BSRGAN, PDM_SR, SwinIR [56], LDL [40], ESRGAN, and real-ESRGAN+. We have shown some representative qualitative results with NIQE in Figure 6 and Table 3. More detailed results calculated by NRQM and PI are presented in Tables 4–6. It can be observed from the figure that the results of our proposed PSRGAN outperforms previous approaches in both details and clearness, with fewer artifacts. For instance, PSRGAN can produce clearer, more natural lion fur (see 0901) and more detailed wall structures (see OST_278) than BSRGAN and LDL, whose textures are unnatural, skewed, and contain unpleasing noise. Compared with PSRGAN, ESRGAN and real-ESRGAN+ fail to produce enough details. Moreover, PSRGAN is capable of boosting visual sharpness (see DSC_1454_x1), while other methods either produce blurry structures (ESRGAN, PDM_SR, and SwinIR) or do not generate enough details (BSRGAN). In addition, previous GAN-based methods sometimes introduced unpleasant artifacts such as BSRGAN and real-ESRGAN+. Our PSRGAN eliminates these artifacts and obtains cleaner results (see Canon_40_x1).

Table 3. NIQE scores on diverse testing datasets—the lower, the better. Colors R, G, and B indicate the best first, second, and third NIQE results among models on each dataset row. The calculation method of NIQE is derived from the basic SR package of PyTorch 1.11.0 + cu113.

	Bicubic	BSRGAN	PDM_SR	SwinIR	LDL	ESRGAN	real-ESRGAN+	PSRGAN
OST300	7.600	3.309	4.319	2.921	**2.817**	3.501	2.806	**2.735**
DRealSR_Test_x4	9.772	**4.803**	7.667	4.698	5.250	8.644	4.846	**4.533**
RealSR-Canon	13.480	5.998	10.015	4.956	5.637	13.096	**5.352**	**4.499**
RealSR-Nikon	13.017	6.377	9.544	**4.819**	5.712	12.443	**5.180**	5.164
PIRM_Self_val	7.747	3.808	5.132	3.683	3.539	**3.516**	3.350	**3.330**
2020track1	7.596	**3.783**	4.101	3.618	3.958	7.440	3.820	**3.411**

Table 4. NIQE scores on diverse testing datasets—the lower, the better. Colors R, G, and B indicate the best first, second, and third NIQE results among models on each dataset row. The calculation method of NIQE is in PIRM2018 derived from https://github.com/roimehrez/PIRM2018 (accessed on 1 June 2023).

	Bicubic	BSRGAN	PDM_SR	SwinIR	LDL	ESRGAN	real-ESRGAN+	PSRGAN
OST300	7.612	3.414	4.308	**3.034**	4.56	3.551	2.929	**2.826**
DRealSR_Test_x4	9.766	4.818	7.635	9.765	8.372	8.632	**4.848**	**4.543**
RealSR-Canon	13.442	6.046	10.008	4.985	13.187	13.101	**5.346**	**4.512**
RealSR-Nikon	13.006	6.435	9.537	**4.834**	12.39	12.446	**5.176**	5.169
PIRM_Self_val	7.746	3.838	5.195	3.716	**2.986**	3.511	**3.363**	3.311
2020track1	7.606	**3.813**	4.096	7.606	**3.249**	7.217	3.835	3.423

Table 5. NRQM scores on diverse testing datasets—the higher, the better. Colors R, G, and B indicate the best first, second, and third NRQM results among models on each dataset row. The calculation method of NRQM is in PIRM2018 derived from https://github.com/roimehrez/PIRM2018 (accessed on 1 June 2023).

	Bicubic	BSRGAN	PDM_SR	SwinIR	LDL	ESRGAN	real-ESRGAN+	PSRGAN
OST300	3.266	6.319	5.737	6.58	5.683	6.236	**6.576**	**6.714**
DRealSR_Test_x4	2.576	**5.264**	3.536	2.576	3.317	3.244	5.295	**5.551**
RealSR-Canon	2.337	4.571	2.484	**4.861**	2.548	2.476	5.743	**6.131**
RealSR-Nikon	2.366	4.635	2.597	**5.249**	2.866	2.681	5.69	**5.839**
PIRM_Self_val	3.76	8.091	6.096	8.191	**8.393**	8.401	8.347	**8.524**
2020track1	3.307	6.219	5.99	3.307	**6.493**	**6.591**	6.133	6.504

Table 6. PI scores on diverse testing datasets—the lower, the better. Colors R, G, and B indicate the best first, second, and third PI results among models on each dataset row. The calculation method of PI is in PIRM2018 derived from https://github.com/roimehrez/PIRM2018 (accessed on 1 June 2023).

	Bicubic	BSRGAN	PDM_SR	SwinIR	LDL	ESRGAN	real-ESRGAN+	PSRGAN
OST300	7.173	3.548	4.286	**3.227**	4.438	3.658	3.176	**3.056**
DRealSR_Test_x4	8.595	4.777	7.05	8.595	7.527	7.694	**4.777**	**4.495**
RealSR-Canon	10.552	5.738	8.762	**5.062**	10.319	10.313	4.802	**4.191**
RealSR-Nikon	10.32	5.9	8.47	**4.793**	9.762	9.883	4.743	**4.665**
PIRM_Self_val	6.994	2.874	4.549	2.763	**2.297**	2.555	**2.509**	2.394
2020track1	7.15	**3.797**	4.053	7.15	**3.378**	5.313	3.851	3.459

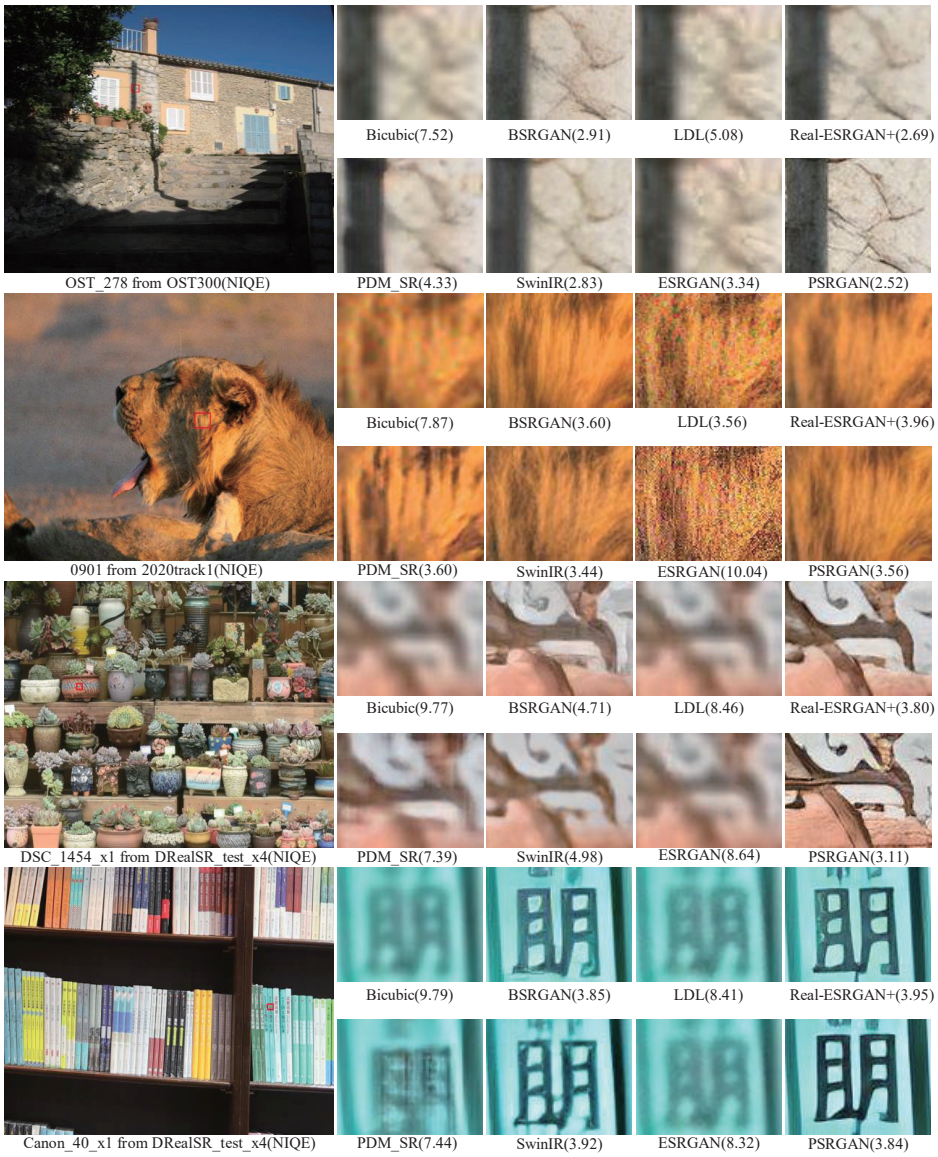

Figure 6. Qualitative results of PSRGAN. PSRGAN produces more subtle textures and clearer structures, e.g., animal texture and building structure, as well as fewer unpleasant artifacts, e.g., artifacts in fonts. Zoom in for best view.

Although the NIQE score of PSRGAN is not always best, we still believe that exploring the effect of focusing on the human visual perception of real pictures is crucial for SR; after all, the existing perception indexes do not reflect all the problems. Please refer to Appendix C for more qualitative results.

4.4. Ablation Study

In order to study the effects of each component in the proposed PSRGAN, we gradually modify the discriminators of PSRGAN and compare their differences. The overall visual comparison is illustrated in Figure 7. Each column represents a model with its configurations shown at the top. The red sign indicates the best performance. A detailed discussion is provided as Table 7 follows.

Table 7. Model with different configurations.

	Second	Third	Fourth	Fifth
PTBs	3	5	5	7
Channels	128	128	256	128

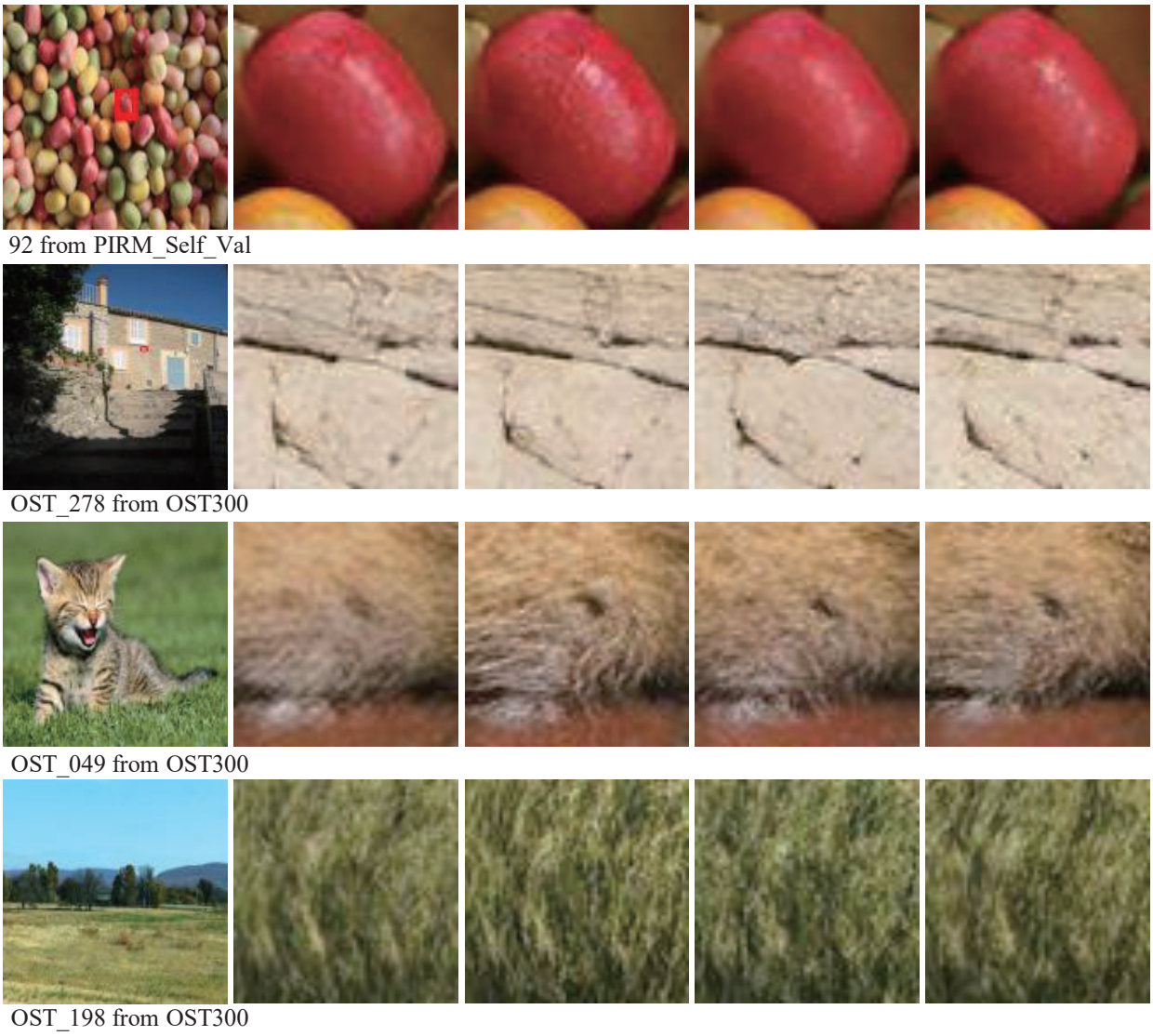

Figure 7. Visual comparisons of different configurations in PSRGAN. The red sign indicates the best performance.

Number of PTBs: The discriminator with the optimal number of cascaded PTBs has a strong representation capacity to capture semantic information, which can further improve the recovered textures, especially for regular structures like the wall of image OST_278 in Figure 6. We set the order of the number to 2, 3, 4, 5, 6, and 7 for experimentation, respectively. For simplisity, we only demonstrate the results of 3, 5, and 7 numbers; the experimental results are depicted in Figure 7. As shown, when the number is 5, the results are relatively sharper with richer textures than others. For some cases, a prominent difference can be observed from the second, third and fifth column in Figure 7.

Channel size of PTB: The different channel sizes of PTB influence the perceptive capabilities of the discriminator. We have tested on 3, 128, and 256 channels. For simplisity, we only demonstrate the results of 128 and 256 channels, as shown in Figure 7. When the channel size is 128, the results are clearer and have fewer artifacts.

Cross verification between PTBs and U-net: Please refer to Appendix A for details.

4.5. Running Times

Our method achieves moderate GPU run times for both training and testing, thanks to its design characteristics. Our model achieves outstanding super-resolution performance, reaching a superior level of quality after a rigorous training regimen of 490 k iterations. Our model exhibits test times on multiple datasets that are comparable to existing state-of-the-art models. Notably, when compared to SwinIR and LDL, our model demonstrates a significant advantage in test time efficiency. The algorithms were trained and tested on a server with NVIDIA GeForce RTX 3090 GPUs. Tables 8 and 9 compare the running times of different state-of-the-art models.

Table 8. The GPU run times for training of different networks. The unit is the number of iterators, and k represents thousands. Since Bicubic is not an adversarial neural network, there is no number of iterators.

	Bicubic	**BSRGAN**	**PDM_SR**	**SwinIR**	**LDL**	**ESRGAN**	**real-ESRGAN+**	**PSRGAN**
GAN Training Times (iters)	None	1000 k	200 k	500 k	400 k	400 k	400 k	490 k

Table 9. The GPU run times of different networks on diverse datasets. The unit is time, where m stands for minutes and s stands for seconds.

	Bicubic	**BSRGAN**	**PDM_SR**	**SwinIR**	**LDL**	**ESRGAN**	**real-ESRGAN+**	**PSRGAN**
OST300	54 s	5 m 4 s	5 m 13 s	23 m 54 s	7 m 15 s	5 m 36 s	5 m 16 s	5 m 29 s
DRealSR_Test_x4	1 m 2 s	5 m 47 s	5 m 46 s	21 m 56 s	8 m 34 s	6 m 21 s	5 m 58 s	6 m 20 s
RealSR-Canon	17 s	2 m	2 m 1 s	9 m 6 s	2 m 20 s	2 m	2 m 3 s	2 m 9 s
RealSR-Nikon	22 s	2 m 31 s	2 m 30 s	9 m 26 s	3 m 12 s	2 m 31 s	2 m 34 s	2 m 42 s
PIRM_Self_val	1 s	6 s	6 s	17 s	19 s	5 s	7 s	7 s
2020track1	10 s	53 s	54 s	2 m 56 s	1 m 18 s	55 s	56 s	58 s

5. Discussion

In this study, we present the perception-design-oriented image super resolution generative adversarial network (PSRGAN), an innovative approach that fuses generative adversarial networks (GANs) and human perceptual insights. Through extensive experiments and analysis of the model, we have achieved the following major achievement.

Perceptually guided super-resolution enhancement: We successfully combined human perceptual insights and used them to guide super-resolution processes. This resulted in sharper, more realistic, and more human-perceivable high-resolution image generation, as illustrated by Figure 6, where our PSRGAN generates more detailed textures of animal hairs, fewer artifacts, and a sharper edge in text-related images.

The experimental results: Our extensive experiments show that PSRGAN achieves significant performance gains on multiple datasets and tasks. Quantitative evaluations show that PSRGAN outperforms traditional super-resolution methods (real-ESRGAN+, ESRGAN, and BSRGAN) on multiple standard image quality metrics such as NIQE, NRQM,

and PI. More encouragingly, the images generated by PSRGAN are closer to the high-resolution original images in terms of human perception.

Limitations: Despite our satisfactory achievements, we have to recognize some limitations of PSRGAN. Computational requirements: the training and inference of PSRGAN requires a large number of computational resources, which may be a challenge for some applications. Data diversity: while our model performs well on multiple datasets, performance may be degraded in specific domains or with uneven data distribution.

In my opinion, the SR network will definitely develop in the direction of breaking through its current limitations in the future, and the trend of super-resolution application is to reduce the computational burden and to apply it to diversified datasets.

6. Conclusions

We have presented a PSRGAN model that achieves superior perceptual quality both in terms of evaluation metrics and visual effects. According to the experimental results, our proposed PSRGAN based on generative adversarial networks outperforms current state-of-the-art intelligent algorithms (BSRGAN, real-ESRGAN, PDM_SR, SwinIR, LDL, etc.) on several evaluation metrics (NIQE, NRQM and PI), with a $\times 4$ upscaling factor on various datasets (OST300, DRealSR_Test_x4, RealSR-Canon, etc.). The PSRGAN model mainly consists of two kinds of perception turbo (PT), GPT in the generator network, and DPT in the discriminator network. In terms of visual effects, the proposed image slice block mitigates the artifacts and noise in the reconstructed image, the three-level perception structure in GPT which could extract diversified textures. The cascaded PTBs in DPT could further promote the generator to restore subtle textures.

Author Contributions: All authors contributed to the study conception and design. T.W. and Y.Z. participated in the design of the network structure, S.X. is responsible for the budget and paper quality and revision, and the network implementation was carried out by Y.L. and H.T. Material preparation, data collection, and analysis were performed by H.T., H.L. (Huaizheng Liu) and J.Z. The model training and experimental results collection was carried out by Y.L., H.T. and H.L. (Huaizheng Liu). Y.Z. participated in the system integration and testing. The first draft of the manuscript was written by H.L. (Hui Liu) and J.Z. All of the authors commented on previous versions of the manuscript. All authors have read and agreed to the published version of the manuscript

Funding: This work was supported by Tencent Technology (Shenzhen) Co., Ltd.

Data Availability Statement: The calculation method of NIQE is derived from the following resources available in the public domain: https://github.com/roimehrez/PIRM2018 (accessed on 1 June 2023). The verification program is available at https://drive.google.com/file/d/1AZmkgvgfcBicTP9tM5X-gahnirskm6-1/view?usp=share_link (accessed on 1 June 2023). The datasets analyzed during the current study are derived from the following resources available in the public domain: https://github.com/XPixelGroup/BasicSR/blob/master/docs/DatasetPreparation.md (accessed on 1 June 2023).

Acknowledgments: The authors would like to thank Tencent Company for helpful budget and resource on topics related to this work, special the Qilong Kou, the department of Tencent Institute of Games.

Conflicts of Interest: The authors have no competing interests to declare that are relevant to the content of this article. This work did not involve human participants or animal research.

Appendix A. Cross Verification between PTBs and U-Net

In PSRGAN, we now call its generator network PSRNet and its discriminator network PTBs. In this section, we further compare the differences between two kinds of GANs, PSRNet+PTBs, and PSRNet+U-net on diverse testing datasets. From the results in Table A1, we can conclude that using PTBs can promote the generator to restore more perceptive SR images driven by GAN; more qualitative comparisons are shown in Figure A1.

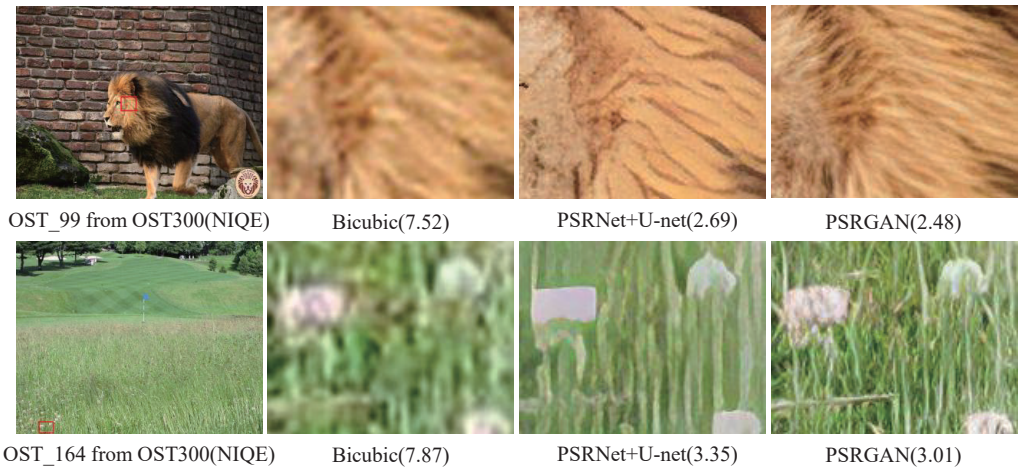

Figure A1. Qualitative comparisons on representative real-world samples with $\times 4$ upscaling factors. PSRNet+PTBs outperforms PSRNet+U-net in terms of both restoring texture details (See OST_99) and producing clearer results (See OST_164).

Table A1. NIQE scores on several diverse testing datasets. The lower, the better.

	PIRM_self_val	OST300	RealSR-Nikon	RealSR-Canon
PSRNet+U-net	3.527	2.830	5.673	5.896
PSRGAN (PSRNet+PTBs)	3.330	2.735	5.164	4.499

Appendix B. Structure of Multi-Scale Feature Extraction Block

We conducted experiments on the number of convolutional groups for multi-scale feature-extraction block (MFEB) in the generator network. As the experimental results in Table A2 show, the SR results show better performance when the number of convolutional groups is three.

Table A2. NIQE scores of feature extraction block at different scales on diverse testing datasets; the lower, the better. The calculation method of NIQE is derived from the basic SR package of PyTorch 1.11.0+cu113.

Groups	1	2	3	4	5
OST300	6.650362	6.635382	**6.520303**	6.626211	6.61622
DRealSR_Test_x4	7.892549	7.932709	**7.818715**	7.993731	7.821493
RealSR-Canon	**10.179904**	10.313314	10.208351	10.32831	10.3463
RealSR-Nikon	10.445691	10.491598	10.409443	10.582925	**10.378877**
PIRM_Self_val	6.717716	6.717983	**6.653905**	6.699813	6.694974
2020track1	6.596391	6.583975	**6.46227**	6.60103	6.526587

Appendix C. More Qualitative Results

Figure A2. More qualitative results of PSRGAN and NIQE are provided for reference. [×4 upscaling].

References

1. Duchon, C.E. Lanczos filtering in one and two dimensions. *J. Appl. Meteorol. Climatol.* **1979**, *18*, 1016–1022. [CrossRef]
2. Zhang, L.; Wu, X. An edge-guided image interpolation algorithm via directional filtering and data fusion. *IEEE Trans. Image Process.* **2006**, *15*, 2226–2238. [CrossRef] [PubMed]
3. Wu, Y.; Ding, H.; Gong, M.; Qin, A.; Ma, W.; Miao, Q.; Tan, K.C. Evolutionary multiform optimization with two-stage bidirectional knowledge transfer strategy for point cloud registration. *IEEE Trans. Evol. Comput.* **2022**. [CrossRef]
4. Wu, Y.; Zhang, Y.; Ma, W.; Gong, M.; Fan, X.; Zhang, M.; Qin, A.; Miao, Q. Rornet: Partial-to-partial registration network with reliable overlapping representations. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**. [CrossRef] [PubMed]
5. Dai, S.; Han, M.; Xu, W.; Wu, Y.; Gong, Y.; Katsaggelos, A.K. Softcuts: A soft edge smoothness prior for color image super-resolution. *IEEE Trans. Image Process.* **2009**, *18*, 969–981. [PubMed]
6. Sun, J.; Xu, Z.; Shum, H.Y. Image super-resolution using gradient profile prior. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, Alaskapp, 23–28 June 2008; pp. 1–8.
7. Yan, Q.; Xu, Y.; Yang, X.; Nguyen, T.Q. Single image superresolution based on gradient profile sharpness. *IEEE Trans. Image Process.* **2015**, *24*, 3187–3202. [PubMed]
8. Shang, T.; Dai, Q.; Zhu, S.; Yang, T.; Guo, Y. Perceptual extreme super-resolution network with receptive field block. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 13–19 June 2020; pp. 440–441.
9. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Adv. Neural Inf. Process. Syst.* **2020**, *63*, 139–144. [CrossRef]
10. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-realistic single-image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4681–4690.

11. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Change Loy, C. Esrgan: Enhanced super-resolution generative adversarial networks. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, 8–14 September 2018.
12. Yuan, Y.; Liu, S.; Zhang, J.; Zhang, Y.; Dong, C.; Lin, L. Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–23 June 2018; pp. 701–710.
13. Zhang, Y.; Liu, S.; Dong, C.; Zhang, X.; Yuan, Y. Multiple cycle-in-cycle generative adversarial networks for unsupervised image super-resolution. *IEEE Trans. Image Process.* **2019**, *29*, 1101–1112. [CrossRef]
14. Zhang, K.; Liang, J.; Van Gool, L.; Timofte, R. Designing a practical degradation model for deep blind image super-resolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 4791–4800.
15. Wang, X.; Xie, L.; Dong, C.; Shan, Y. real-ESRGAN: Training real-world blind super-resolution with pure synthetic data. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 1905–1914.
16. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; Proceedings, Part III 18; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
17. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
18. Dong, C.; Loy, C.C.; He, K.; Tang, X. Learning a deep convolutional network for image super-resolution. In Proceedings of the Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, 6–12 September 2014; Proceedings, Part IV 13; Springer: Berlin/Heidelberg, Germany, 2014; pp. 184–199.
19. Kim, J.; Lee, J.K.; Lee, K.M. Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1646–1654.
20. Lai, W.S.; Huang, J.B.; Ahuja, N.; Yang, M.H. Deep laplacian pyramid networks for fast and accurate super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 624–632.
21. Timofte, R.; Agustsson, E.; Van Gool, L.; Yang, M.H.; Zhang, L. Ntire 2017 challenge on single-image super-resolution: Methods and results. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 114–125.
22. Haris, M.; Shakhnarovich, G.; Ukita, N. Deep back-projection networks for super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 1664–1673.
23. Dong, C.; Loy, C.C.; Tang, X. Accelerating the super-resolution convolutional neural network. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Proceedings, Part II 14; Springer: Berlin/Heidelberg, Germany, 2016; pp. 391–407.
24. Shi, W.; Caballero, J.; Huszár, F.; Totz, J.; Aitken, A.P.; Bishop, R.; Rueckert, D.; Wang, Z. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1874–1883.
25. Lim, B.; Son, S.; Kim, H.; Nah, S.; Mu Lee, K. Enhanced deep residual networks for single-image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 136–144.
26. Musunuri, Y.R.; Kwon, O.S. Deep residual dense network for single-image super-resolution. *Electronics* **2021**, *10*, 555. [CrossRef]
27. Ji, X.; Cao, Y.; Tai, Y.; Wang, C.; Li, J.; Huang, F. Real-world super-resolution via kernel estimation and noise injection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 13–19 June 2020; pp. 466–467.
28. Wei, P.; Xie, Z.; Lu, H.; Zhan, Z.; Ye, Q.; Zuo, W.; Lin, L. Component divide-and-conquer for real-world image super-resolution. In Proceedings of the Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020; Proceedings, Part VIII 16; Springer: Berlin/Heidelberg, Germany, 2020; pp. 101–117.
29. Luo, Z.; Huang, Y.; Li, S.; Wang, L.; Tan, T. Learning the degradation distribution for blind image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 6063–6072.
30. 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 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 606–615.
31. Shao, G.; Sun, Q.; Gao, Y.; Zhu, Q.; Gao, F.; Zhang, J. Sub-Pixel Convolutional Neural Network for Image super-resolution Reconstruction. *Electronics* **2023**, *12*, 3572. [CrossRef]
32. Bruna, J.; Sprechmann, P.; LeCun, Y. super-resolution with deep convolutional sufficient statistics. *arXiv* **2015**, arXiv:1511.05666.
33. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Proceedings, Part II 14; Springer: Berlin/Heidelberg, Germany, 2016; pp. 694–711.
34. Sajjadi, M.S.; Scholkopf, B.; Hirsch, M. Enhancenet: Single-image super-resolution through automated texture synthesis. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 4491–4500.

35. 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), Munich, Germany, 8–14 September 2018; pp. 768–783.
36. 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) Workshops, Munich, Germany, 8–14 September 2018.
37. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
38. Liu, S.; Huang, D. Receptive field block net for accurate and fast object detection. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 385–400.
39. Zhang, K.; Gu, S.; Timofte, R. Ntire 2020 challenge on perceptual extreme super-resolution: Methods and results. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 13–19 June 2020; pp. 492–493.
40. Newell, A.; Yang, K.; Deng, J. Stacked hourglass networks for human pose estimation. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Proceedings, Part VIII 14; Springer: Berlin/Heidelberg, Germany, 2016; pp. 483–499.
41. Liang, J.; Zeng, H.; Zhang, L. Details or artifacts: A locally discriminative learning approach to realistic image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5657–5666.
42. Zhang, M.; Ling, Q. Supervised pixel-wise GAN for face super-resolution. *IEEE Trans. Multimed.* **2020**, *23*, 1938–1950. [CrossRef]
43. Gong, Y.; Liao, P.; Zhang, X.; Zhang, L.; Chen, G.; Zhu, K.; Tan, X.; Lv, Z. Enlighten-GAN for super resolution reconstruction in mid-resolution remote sensing images. *Remote Sens.* **2021**, *13*, 1104. [CrossRef]
44. Sheikh, H.R.; Bovik, A.C.; De Veciana, G. An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Trans. Image Process.* **2005**, *14*, 2117–2128. [CrossRef]
45. Sheikh, H.R.; Bovik, A.C. Image information and visual quality. *IEEE Trans. Image Process.* **2006**, *15*, 430–444. [CrossRef]
46. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]
47. Ma, C.; Yang, C.Y.; Yang, X.; Yang, M.H. Learning a no-reference quality metric for single-image super-resolution. *Comput. Vis. Image Underst.* **2017**, *158*, 1–16. [CrossRef]
48. Mittal, A.; Soundararajan, R.; Bovik, A.C. Making a "completely blind" image quality analyzer. *IEEE Signal Process. Lett.* **2012**, *20*, 209–212. [CrossRef]
49. Mittal, A.; Moorthy, A.K.; Bovik, A.C. No-reference image quality assessment in the spatial domain. *IEEE Trans. Image Process.* **2012**, *21*, 4695–4708. [CrossRef]
50. Blau, Y.; Michaeli, T. The perception-distortion tradeoff. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6228–6237.
51. Vasu, S.; Thekke Madam, N.; Rajagopalan, A. Analyzing perception-distortion tradeoff using enhanced perceptual super-resolution network. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, 8–14 September 2018.
52. Lugmayr, A.; Danelljan, M.; Timofte, R. Ntire 2020 challenge on real-world image super-resolution: Methods and results. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 13–19 June 2020; pp. 494–495.
53. Cai, J.; Zeng, H.; Yong, H.; Cao, Z.; Zhang, L. Toward real-world single-image super-resolution: A new benchmark and a new model. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 3086–3095.
54. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*.
55. He, E.; Chen, Q.; Zhong, Q. SL-Swin: A transformer-Based Deep Learning Approach for Macro-and Micro-Expression Spotting on Small-Size Expression Datasets. *Electronics* **2023**, *12*, 2656. [CrossRef]
56. Yang, F.; Yang, H.; Fu, J.; Lu, H.; Guo, B. Learning texture transformer network for image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 5791–5800.
57. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 10012–10022.
58. Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L.; Timofte, R. Swinir: Image restoration using swin transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 1833–1844.
59. Lu, Z.; Li, J.; Liu, H.; Huang, C.; Zhang, L.; Zeng, T. transformer for single-image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 457–466.

60. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
61. Agustsson, E.; Timofte, R. Ntire 2017 challenge on single-image super-resolution: Dataset and study. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 126–135.