ESRGAN - Enhanced Super-Resolution Generative Adversarial Networks论文翻译——中英文对照

ESRGAN - Enhanced Super-Resolution Generative Adversarial Networks论文翻译——中英文对照

文章作者：Tyan

博客：noahsnail.com | CSDN | 简书

声明：作者翻译论文仅为学习，如有侵权请联系作者删除博文，谢谢！

翻译论文汇总：https://github.com/SnailTyan/deep-learning-papers-translation

ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks

Abstract

The Super-Resolution Generative Adversarial Network (SR-GAN) [1] is a seminal work that is capable of generating realistic textures during single image super-resolution. However, the hallucinated details are often accompanied with unpleasant artifacts. To further enhance the visual quality, we thoroughly study three key components of SRGAN – network architecture, adversarial loss and perceptual loss, and improve each of them to derive an Enhanced SRGAN (ESRGAN). In particular, we introduce the Residual-in-Residual Dense Block (RRDB) without batch normalization as the basic network building unit. Moreover, we borrow the idea from relativistic GAN [2] to let the discriminator predict relative realness instead of the absolute value. Finally, we improve the perceptual loss by using the features before activation, which could provide stronger supervision for brightness consistency and texture recovery. Benefiting from these improvements, the proposed ESRGAN achieves consistently better visual quality with more realistic and natural textures than SRGAN and won the first place in the PIRM2018-SR Challenge [3]. The code is available at https://github.com/xinntao/ESRGAN.

摘要

超分辨率生成对抗网络(SR-GAN)[1]是一项开创性的工作，其能够在单图像超分辨率期间生成逼真的纹理。然而，虚幻的细节常常伴随讨厌的伪像。为了进一步增强视觉质量，我们充分研究了SRGAN的三个关键组成部分

——网络架构、对抗损失和感知损失，并对每一个都进行了改进以取得增强的SRGAN(ESRGAN)。特别的是，我们引入了没有批归一化的Residual-in-Residual Dense Block(RRDB)作为基本的网络构架单元。此外，我们借鉴了相对GAN[2]中的思想，让判别器预测相对真实性而不是绝对值。最后，我们通过使用激活前的特征改进感知损失，这可以对亮度一致性和纹理复原提供更强的监督。得益于这些改进，相比于SRGAN，提出的ESRGAN一致地取得了更好的视觉质量、更多真实自然的纹理，并在PIRM2018-SR Challenge[3]中获得了第一名。源码地址：https://github.com/xinntao/ESRGAN。

1 Introduction

Single image super-resolution (SISR), as a fundamental low-level vision problem, has attracted increasing attention in the research community and AI companies. SISR aims at recovering a high-resolution (HR) image from a single low-resolution (LR) one. Since the pioneer work of SRCNN proposed by Dong et al. [4], deep convolution neural network (CNN) approaches have brought prosperous development. Various network architecture designs and training strategies have continuously improved the SR performance, especially the Peak Signal-toNoise Ratio (PSNR) value [5,6,7,1,8,9,10,11,12]. However, these PSNR-oriented approaches tend to output over-smoothed results without sufficient high-frequency details, since the PSNR metric fundamentally disagrees with the subjective evaluation of human observers [1].

1 引言

作为一个基本的低级视觉问题，单图像超分辨率(SISR)在研究领域和AI公司中引起了越来越多的关注。SISR目标是从一张低分辨率(LR)图像复原出一张高分辨率(HR)图像。从Dong等[4]提出SRCNN的开创性工作以来，深度卷积神经网络(CNN)方法带来了繁荣的发展。各种网络架构设计和训练策略持续地改善SR性能，尤其是峰值信噪比(PSNR)的值[5,6,7,1,8,9,10,11,12]。然而，这些面向PSNR的方法趋向于输出过于平滑的结果，缺少足够的高频细节，因为PSNR度量从根本上与人类观察者的主观评价[1]不符。

Several perceptual-driven methods have been proposed to improve the visual quality of SR results. For instance, perceptual loss [13,14] is proposed to optimize super-resolution model in a feature space instead of pixel space. Generative adversarial network [15] is introduced to SR by [1,16] to encourage the network to favor solutions that look more like natural images. The semantic image prior is further incorporated to improve recovered texture details [17]. One of the milestones in the way pursuing visually pleasing results is SRGAN [1]. The basic model is built with residual blocks [18] and optimized using perceptual loss in a GAN framework. With all these techniques, SRGAN significantly improves the overall visual quality of reconstruction over PSNR-oriented methods.

已经提出了一些感知驱动的方法来改进SR结果的视觉质量。例如，提出感知损失[13,14]来优化在特征空间而不是像素空间中的超分辨率模型。[1,16]引入生成对抗网络[15]到SR中以鼓励网络支持看起来更像自然图像的解。语义图像先验被进一步合并以改善恢复的纹理细节[17]。追寻视觉愉悦效果的方法中的里程碑之一是

SRGAN[1]。基本模型是用残差块构建的[18]，并在GAN框架中使用感知损失来进行优化。通过所有这些技术，与面向PSNR的方法相比，SRGAN显著改善了重建的整体视觉质量。

However, there still exists a clear gap between SRGAN results and the ground-truth (GT) images, as shown in Fig. 1. In this study, we revisit the key components of SRGAN and improve the model in three aspects. First, we improve the network structure by introducing the Residual-in-Residual Dense Block (RDDB), which is of higher capacity and easier to train. We also remove Batch Normalization (BN) [19] layers as in [20] and use residual scaling [21,20] and smaller initialization to facilitate training a very deep network. Second, we improve the discriminator using Relativistic average GAN (RaGAN) [2], which learns to judge "whether one image is more realistic than the other" rather than "whether one image is real or fake". Our experiments show that this improvement helps the generator recover more realistic texture details. Third, we propose an improved perceptual loss by using the VGG features *before activation* instead of after activation as in SRGAN. We empirically find that the adjusted perceptual loss provides sharper edges and more visually pleasing results, as will be shown in Sec. 4.4. Extensive experiments show that the enhanced SRGAN, termed ESRGAN, consistently outperforms state-of-the-art methods in both sharpness and details (see Fig. 1 and Fig. 7).

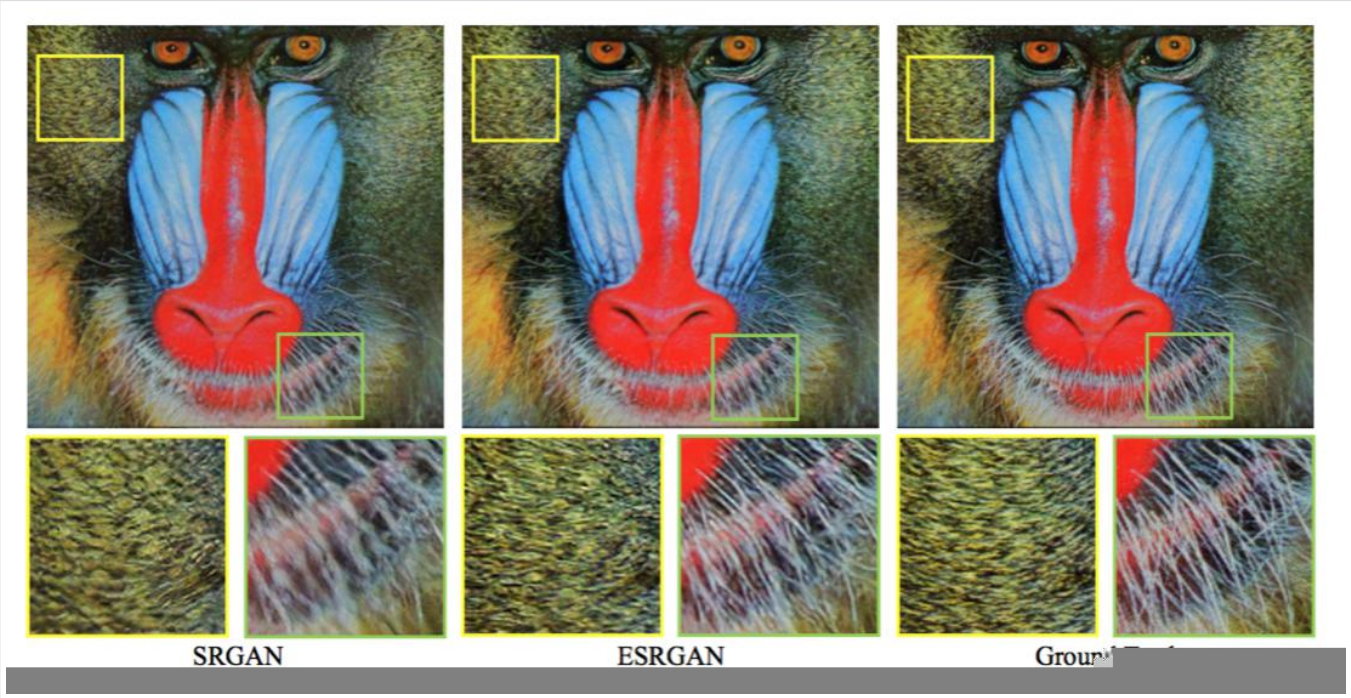


Fig.1: The super-resolution results of ×4 for SRGAN, the proposed ESRGAN and the ground-truth. ESRGAN outperforms SRGAN in sharpness and details.

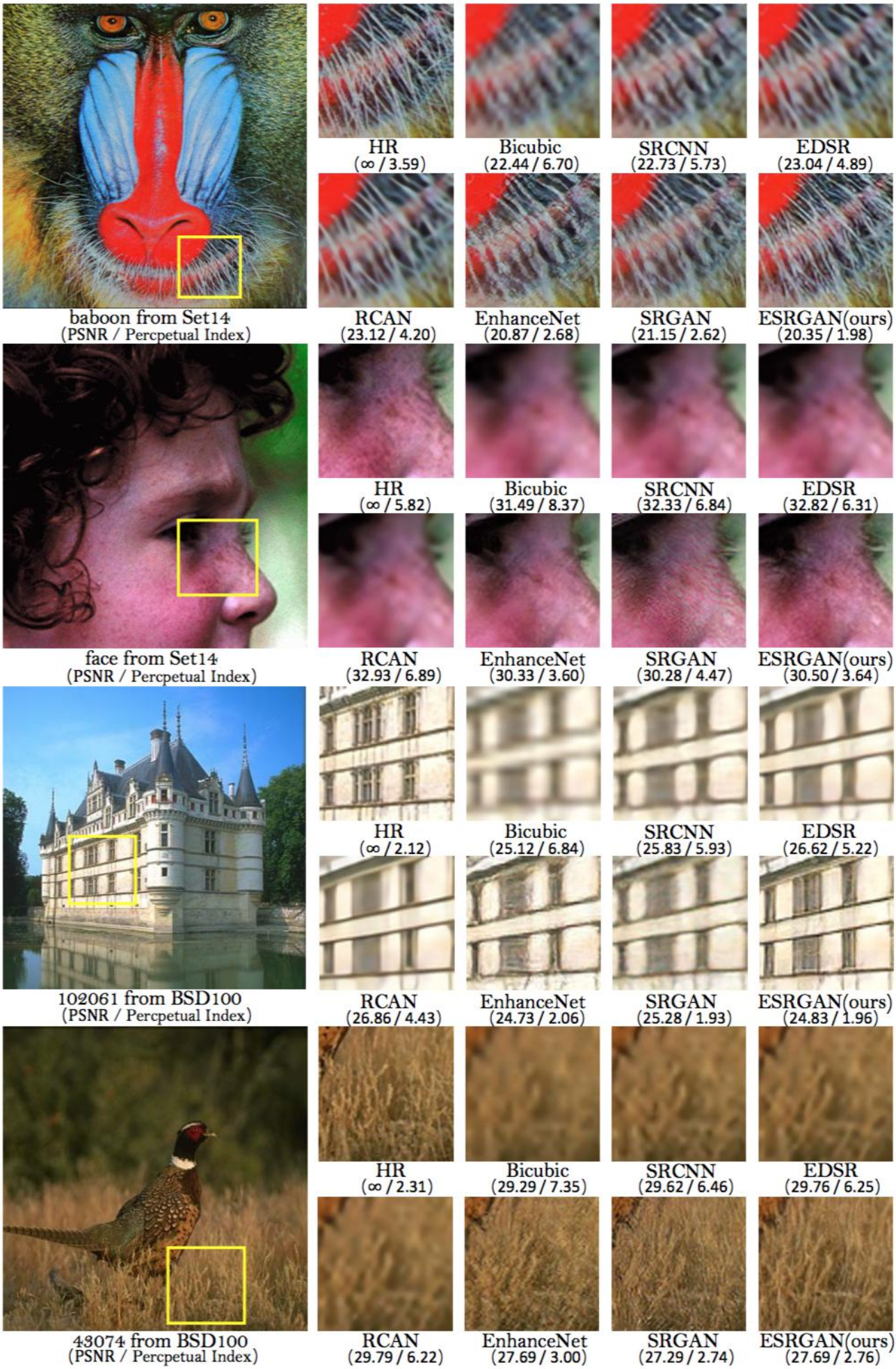

HR
(∞ / 3.59)
Bicubic
(22.44 / 6.70)
SRCNN
(22.73 / 5.73)
EDSR
(23.04 / 4.89)
baboon from Set14
(PSNR / Percpetual Index)
RCAN
(23.12 / 4.20)
EnhanceNet
(20.87 / 2.68)
SRGAN
(21.15 / 2.62)
ESRGAN(ours)
(20.35 / 1.98)
HR
(∞ / 5.82)
Bicubic
(31.49 / 8.37)
SRCNN
(32.33 / 6.84)
EDSR
(32.82 / 6.31)
face from Set14
(PSNR / Percpetual Index)
RCAN
(32.93 / 6.89)
EnhanceNet
(30.33 / 3.60)
SRGAN
(30.28 / 4.47)
ESRGAN(ours)
(30.50 / 3.64)
HR
(∞ / 2.12)
Bicubic
(25.12 / 6.84)
SRCNN
(25.83 / 5.93)
EDSR
(26.62 / 5.22)
102061 from BSD100
(PSNR / Percpetual Index)
RCAN
(26.86 / 4.43)
EnhanceNet
(24.73 / 2.06)
SRGAN
(25.28 / 1.93)
ESRGAN(ours)
(24.83 / 1.96)
HR
(∞ / 2.31)
Bicubic
(29.29 / 7.35)
SRCNN
(29.62 / 6.46)
EDSR
(29.76 / 6.25)
43074 from BSD100
(PSNR / Percpetual Index)
RCAN
(29.79 / 6.22)
EnhanceNet
(27.69 / 3.00)
SRGAN
(27.29 / 2.74)
ESRGAN(ours)
(27.69 / 2.76)

Fig.7: Qualitative results of ESRGAN. ESRGAN produces more natural textures, e.g., animal fur, building structure and grass texture, and also less unpleasant artifacts, e.g., artifacts in the face by SRGAN.

然而，如图1所示，SRGAN结果与真实(GT)图像之间仍然存在明显的差距。在本研究中，我们重新审视SRGAN的关键组件，并在三个方面改进模型。首先，我们通过引入Residual-in-Residual Dense Block(RDDB)改进网络架构，该结构具有较高的能力且更容易训练。我们像[20]中一样也移除了批归一化(BN)[19]层，使用残差缩放[21,20]和更小的初始化来促进训练一个非常深的网络。其次，我们使用相对平均GAN(RaGAN)[2]来改进判别器，RaGAN学习判断“一张图像是否比另一张更真实”而不是“一张图像时真的还是假的”。我们的实验表明这个改进有助于生成器恢复更多的真实纹理细节。第三，我们提出了一种改进的感知损失，使用激活之前的VGG特征来代替SRGAN中激活之后的VGG特征。从经验上我们发现调整之后的感知损失提供了更清晰的边缘和视觉上更令人满意的结果，如4.4节所示。大量的实验表明增强SRGAN(称为ESRGAN)在清晰度和细节方面都始终优于最新的方法（见图1和图7）。

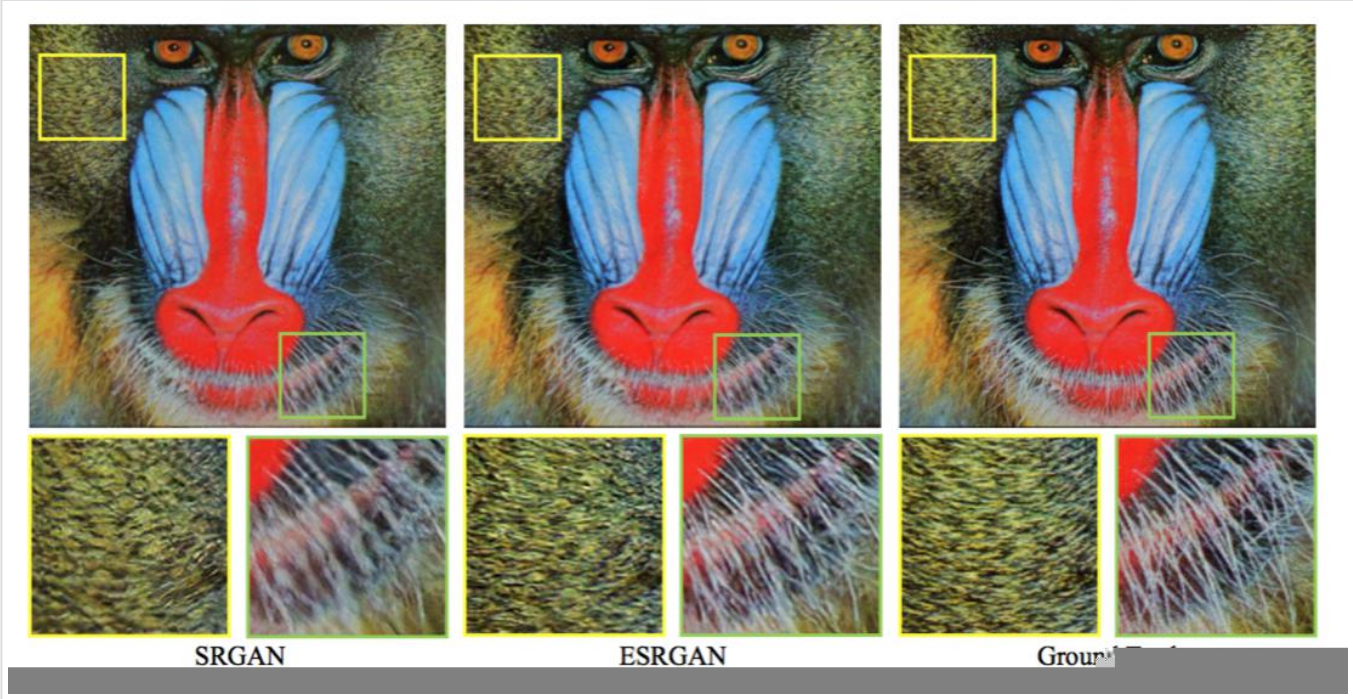


图1：SRGAN、提出的ESRGAN和实际的4倍超分辨率结果。ESRGAN在清晰度和细节方面优于SRGAN。

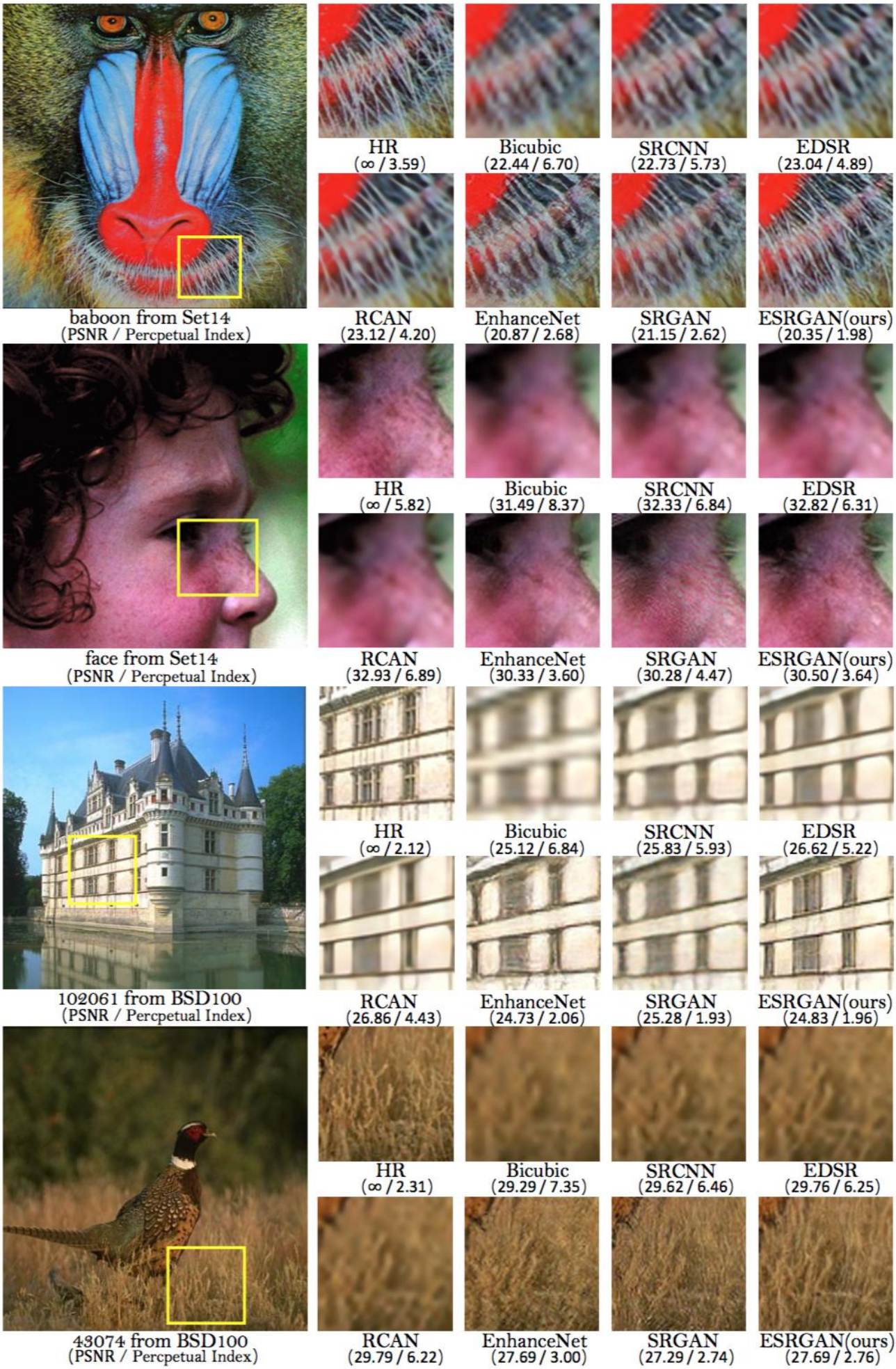
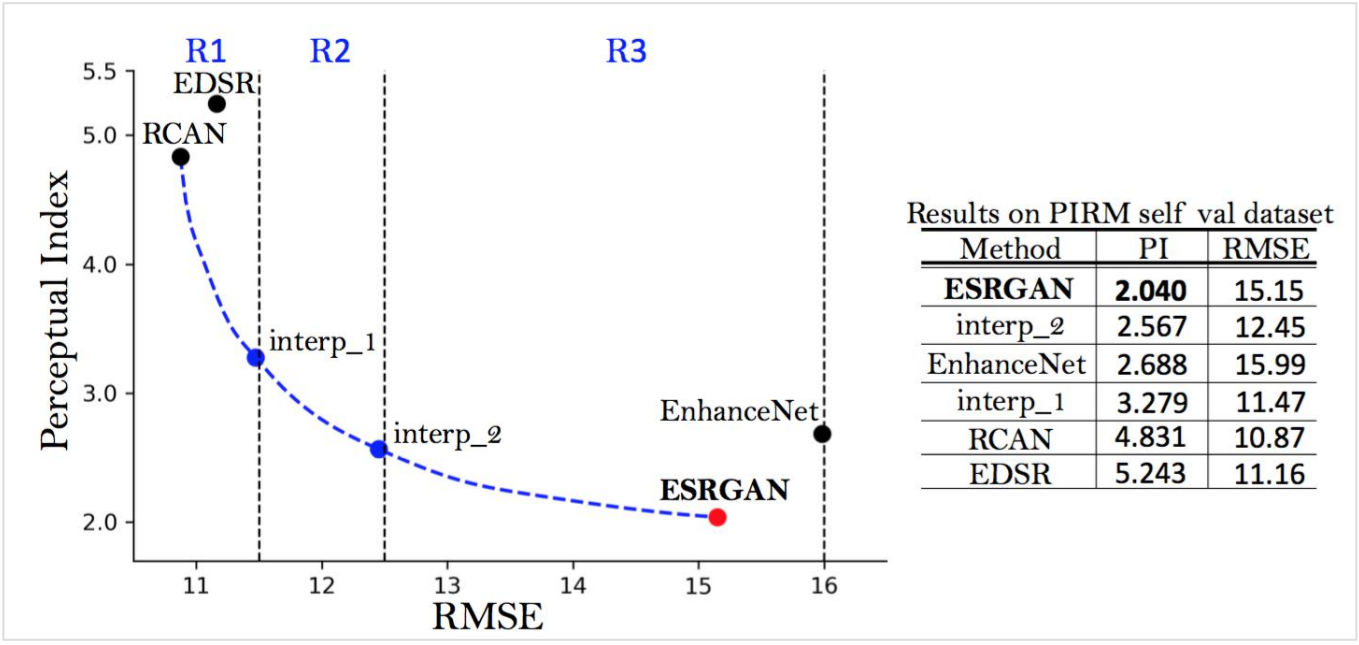

HR
(∞ / 3.59)
Bicubic
(22.44 / 6.70)
SRCNN
(22.73 / 5.73)
EDSR
(23.04 / 4.89)
baboon from Set14
(PSNR / Percpetual Index)
RCAN
(23.12 / 4.20)
EnhanceNet
(20.87 / 2.68)
SRGAN
(21.15 / 2.62)
ESRGAN(ours)
(20.35 / 1.98)
HR
(∞ / 5.82)
Bicubic
(31.49 / 8.37)
SRCNN
(32.33 / 6.84)
EDSR
(32.82 / 6.31)
face from Set14
(PSNR / Percpetual Index)
RCAN
(32.93 / 6.89)
EnhanceNet
(30.33 / 3.60)
SRGAN
(30.28 / 4.47)
ESRGAN(ours)
(30.50 / 3.64)
HR
(∞ / 2.12)
Bicubic
(25.12 / 6.84)
SRCNN
(25.83 / 5.93)
EDSR
(26.62 / 5.22)
102061 from BSD100
(PSNR / Percpetual Index)
RCAN
(26.86 / 4.43)
EnhanceNet
(24.73 / 2.06)
SRGAN
(25.28 / 1.93)
ESRGAN(ours)
(24.83 / 1.96)
HR
(∞ / 2.31)
Bicubic
(29.29 / 7.35)
SRCNN
(29.62 / 6.46)
EDSR
(29.76 / 6.25)
43074 from BSD100
(PSNR / Percpetual Index)
RCAN
(29.79 / 6.22)
EnhanceNet
(27.69 / 3.00)
SRGAN
(27.29 / 2.74)
ESRGAN(ours)
(27.69 / 2.76)

图7：ESRGAN的定性结果。ESRGAN生成了更自然的纹理，例如，动物皮毛，建筑物结构和草坪纹理，以及更少的令人不快的伪影，例如SRGAN中脸上的伪影。

We take a variant of ESRGAN to participate in the PIRM-SR Challenge [3]. This challenge is the first SR competition that evaluates the performance in a perceptual-quality aware manner based on [22], where the authors claim that distortion and perceptual quality are at odds with each other. The perceptual quality is judged by the non-reference measures of Ma's score [23] and NIQE [24], i.e., perceptual index $= \frac{1}{2}((10 - Ma) + NIQE)$. A lower perceptual index represents a better perceptual quality.

我们采用ESRGAN的一个变种来参加PIRM-SR挑战赛[3]。这个挑战是第一个在[22]的基础上以察觉感知质量的方式评估性能的SR竞赛，[22]中作者声称失真和感知质量相互矛盾。感知质量是通过Ma分数[23]和NIQE[24]的非参考度量来判断的，即感知指数$= \frac{1}{2}((10 - Ma) + NIQE)$。更低的感知指数表示更好的感知质量。

As shown in Fig. 2, the perception-distortion plane is divided into three regions defined by thresholds on the Root-Mean-Square Error (RMSE), and the algorithm that achieves the lowest perceptual index in each region becomes the regional champion. We mainly focus on region 3 as we aim to bring the perceptual quality to a new high. Thanks to the aforementioned improvements and some other adjustments as discussed in Sec. 4.6, our proposed ESRGAN won the first place in the PIRM-SR Challenge (region 3) with the best perceptual index.



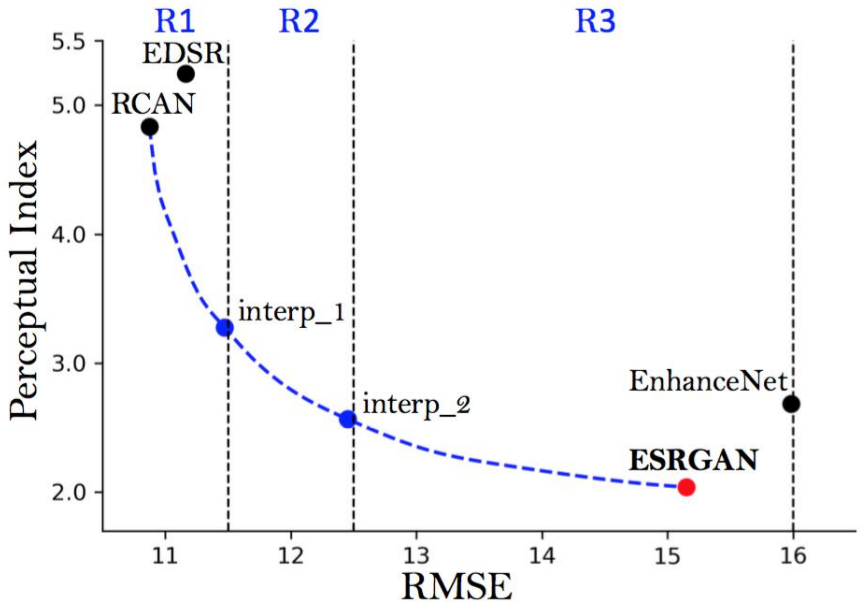
Results on PIRM self val dataset

Method	PI	RMSE
ESRGAN	**2.040**	15.15
interp_2	2.567	12.45
EnhanceNet	2.688	15.99
interp_1	3.279	11.47
RCAN	4.831	10.87
EDSR	5.243	11.16

Fig.2: Perception-distortion plane on PIRM self validation dataset. We show the baselines of EDSR [20], RCAN [12] and EnhanceNet [16], and the submitted ESRGAN model. The blue dots are produced by image interpolation.

如图2所示，通过均方根误差(RMSE)的阈值，将感知失真平面分成三个区域，每个区域中取得最低感知指数的算法为区域冠军。我们主要关注区域3，因为我们旨在将感知质量提升到新的高度。由于上述的改进和4.6节中

讨论的一些其它调整，我们提出的ESRGAN在PIRM-SR挑战赛（区域3）中以最好的感知指数赢得了第一名。



Results on PIRM self val dataset

Method	PI	RMSE
ESRGAN	**2.040**	15.15
interp_2	2.567	12.45
EnhanceNet	2.688	15.99
interp_1	3.279	11.47
RCAN	4.831	10.87
EDSR	5.243	11.16

图2：PIRM自验证集上的感知失真平面。我们展示了EDSR[20]，RCAN[12]，EnhanceNet[16]以及提交的ESRGAN模型的基准线。蓝色的点通过图像插值生成。

In order to balance the visual quality and RMSE/PSNR, we further propose the network interpolation strategy, which could continuously adjust the reconstruction style and smoothness. Another alternative is image interpolation, which directly interpolates images pixel by pixel. We employ this strategy to participate in region 1 and region 2. The network interpolation and image interpolation strategies and their differences are discussed in Sec. 3.4.

为了平衡视觉质量和RMSE/PSNR，我们进一步提出了网络插值策略，其可以持续地调整重建风格和平滑度。另一种替代方案是图像插值，其直接逐像素地插值图像。我们采用这个策略来参加区域1和区域2。网络插值和图像插值策略以及它们的差异在3.4节中讨论。

2 Related Work

We focus on deep neural network approaches to solve the SR problem. As a pioneer work, Dong et al. [4,25] propose SRCNN to learn the mapping from LR to HR images in an end-to-end manner, achieving superior performance against previous works. Later on, the field has witnessed a variety of network architectures, such as a deeper network with residual learning [5], Laplacian pyramid structure [6], residual blocks [1], recursive learning [7,8], densely connected network [9], deep back projection [10] and residual dense network [11]. Specifically, Lim et al. [20] propose EDSR model by removing unnecessary BN layers in the residual block and expanding the model size, which achieves significant improvement. Zhang et al. [11] propose to use effective residual dense block in SR, and they further explore a deeper network with channel attention [12], achieving the

state-of-the-art PSNR performance. Besides supervised learning, other methods like reinforcement learning [26] and unsupervised learning [27] are also introduced to solve general image restoration problems.

2 相关工作

我们专注于解决SR问题的深度神经网络方法。作为开创性工作，Dong等[4,25]提出了SRCNN以端到端的方式来学习从LR到SR图像的映射，取得了优于之前工作的性能。后来，这个领域见证了各种网络架构，例如具有残差学习的神经网络[5]，拉普拉斯金字塔结构[6]，残差块[1]，递归学习[7,8]，密集连接网络[9]，深度反向投影[10]和残差密集网络[11]。具体来说，Lim等[20]通过移除残差块中不必要的BN层以及扩展模型尺寸提出了EDSR模型，取得了显著的改善。Zhang等[11]在SR中提出了使用有效的残差密集块，并且他们进一步开发了一个使用通道注意力[12]的更深网络，取得了最佳的PSNR性能。除了监督学习之外，也引入了其它的方法像强化学习[26]以及无监督学习[27]来解决一般的图像复原问题。

Several methods have been proposed to stabilize training a very deep model. For instance, residual path is developed to stabilize the training and improve the performance [18,5,12]. Residual scaling is first employed by Szegedy et al. [21] and also used in EDSR. For general deep networks, He et al. [28] propose a robust initialization method for VGG-style networks without BN. To facilitate training a deeper network, we develop a compact and effective residual-in-residual dense block, which also helps to improve the perceptual quality.

已经提出了一些方法来稳定训练非常深的模型。例如，开发残差路径来稳定训练并改善性能[18,5,12]。Szegedy等[21]首次采用残差缩放，也在EDSR中使用。对于一般的深度网络，He等[28]为没有BN的VGG风格的网络提出了一个鲁棒的初始化方法。为了便于训练更深的网络，我们也开发了一个简洁有效的残差套残差密集块，这有助于改善感知质量。

Perceptual-driven approaches have also been proposed to improve the visual quality of SR results. Based on the idea of being closer to perceptual similarity [29,14] perceptual loss [13] is proposed to enhance the visual quality by minimizing the error in a feature space instead of pixel space. Contextual loss [30] is developed to generate images with natural image statistics by using an objective that focuses on the feature distribution rather than merely comparing the appearance. Ledig et al. [1] propose SRGAN model that uses perceptual loss and adversarial loss to favor outputs residing on the manifold of natural images. Sajjadi et al. [16] develop a similar approach and further explored the local texture matching loss. Based on these works, Wang et al. [17] propose spatial feature transform to effectively incorporate semantic prior in an image and improve the recovered textures.

感知驱动的方法已经被提出用来改善SR结果的视觉质量。基于更接近于感知相似度[29,14]的想法提出感知损失[13]，通过最小化特征空间而不是像素空间的误差来增强视觉质量。通过使用专注于特征分布而不是只比较外观的目标函数，开发上下文损失[30]来生成具有自然图像统计的图像。Ledig等[1]提出SRGAN模型，使用感知

损失和对抗损失来支持位于自然图像流形的输出。Sajjadi等[16]开发了类似的方法并进一步探索了局部纹理匹配损失。基于这些工作，Wang等[17]提出空间特征变换来有效地将语义先验合并到图像中并改进恢复的纹理。

Throughout the literature, photo-realism is usually attained by adversarial training with GAN [15]. Recently there are a bunch of works that focus on developing more effective GAN frameworks. WGAN [31] proposes to minimize a reasonable and efficient approximation of Wasserstein distance and regularizes discriminator by weight clipping. Other improved regularization for discriminator includes gradient clipping [32] and spectral normalization [33]. Relativistic discriminator [2] is developed not only to increase the probability that generated data are real, but also to simultaneously decrease the probability that real data are real. In this work, we enhance SRGAN by employing a more effective relativistic average GAN.

在整个文献中，通常通过与GAN[15]的对抗训练来获得写实主义照片。最近有很多工作致力于开发更有效的GAN框架。WGAN[31]提出最小化Wasserstein距离的合理和有效近似，并通过权重修剪来正则化判别器。其它对判别器的正则化包括梯度修剪[32]和谱归一化[33]。开发的相对判别器[2]不仅提高了生成数据真实性的概率，而且同时降低了真实数据真实性的概率。在这项工作中，我们通过采用更有效的相对平均GAN来增强SRGAN。

SR algorithms are typically evaluated by several widely used distortion measures, e.g., PSNR and SSIM. However, these metrics fundamentally disagree with the subjective evaluation of human observers [1]. Non-reference measures are used for perceptual quality evaluation, including Ma's score [23] and NIQE [24], both of which are used to calculate the perceptual index in the PIRM-SR Challenge [3]. In a recent study, Blau et al. [22] find that the distortion and perceptual quality are at odds with each other.

SR通常通过几种广泛使用的失真测量方式来进行评估，例如PSNR和SSIM。然而，这些度量从根本上与人类观察者的主观评估不一致[1]。非参考度量通常用于感知质量评估，包括Ma的分数[23]和NIQE[24]，两者都用于PIRM-SR挑战赛中[3]计算感知指数。在最近的一项研究中，Blau等[22]发现失真和感知质量相互矛盾。

3 Proposed Methods

Our main aim is to improve the overall perceptual quality for SR. In this section, we first describe our proposed network architecture and then discuss the improvements from the discriminator and perceptual loss. At last, we describe the network interpolation strategy for balancing perceptual quality and PSNR.

3 提出的方法

我们的主要目标是提高SR的整体感知质量。在本节中，我们首先描述我们提出的网络架构，然后讨论判别器和感知损失的改进。最后，我们描述用于平衡感知质量和PSNR的网络插值策略。

3.1 Network Architecture

In order to further improve the recovered image quality of SRGAN, we mainly make two modifications to the structure of generator G: 1) remove all BN layers; 2) replace the original basic block with the proposed Residual-in-Residual Dense Block (RRDB), which combines multi-level residual network and dense connections as depicted in Fig. 4.

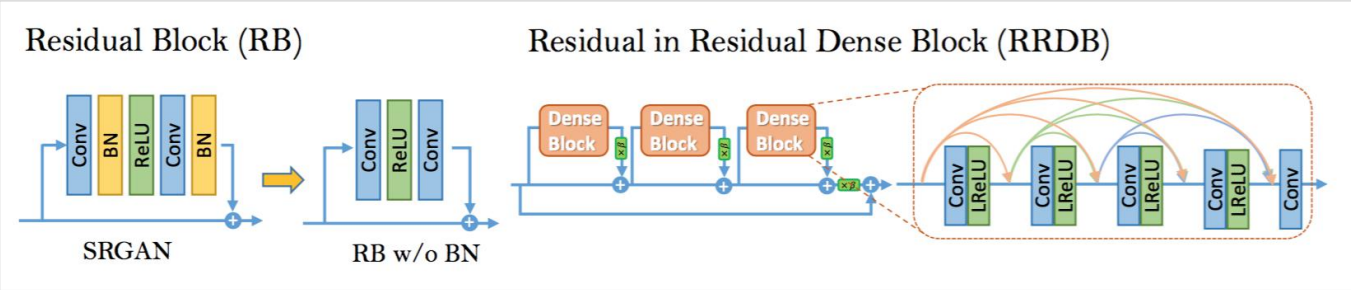


Fig.4: Left: We remove the BN layers in residual block in SRGAN. Right: RRDB block is used in our deeper model and β is the residual scaling parameter.

3.1 网络架构

为了进一步改进SRGAN复原的图像质量，我们主要对生成器G的架构进行了两个修改：1）移除所有的BN层；2）用提出的残差套残差密集块(RRDB)替换原始的基本块，它结合了多层残差网络和密集连接，如图4所示。

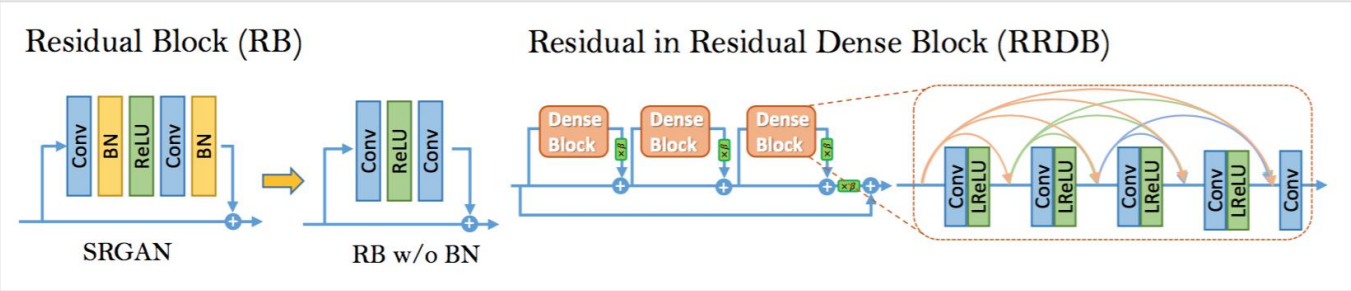


图4：左：我们移除了SRGAN残差块中的BN层。右：RRDB块用在我们的更深模型中，β是残差尺度参数。

Removing BN layers has proven to increase performance and reduce computational complexity in different PSNR-oriented tasks including SR [20] and deblurring [35]. BN layers normalize the features using mean and variance in a batch during training and use estimated mean and variance of the whole training dataset during testing. When the statistics of training and testing datasets differ a lot, BN layers tend to introduce unpleasant artifacts and limit the generalization ability. We empirically observe that BN layers are more likely to bring artifacts when the network is deeper and trained under a GAN framework. These artifacts occasionally appear among iterations and different settings, violating the needs for a stable performance over training. We therefore remove BN layers for stable training and consistent performance. Furthermore, removing BN layers helps to improve generalization ability and to reduce computational complexity and memory usage.

在不同的面向PSNR的任务（包括SR[20]和去模糊[35]）中，已经证实了移除BN层可以提高性能并降低计算复杂度。BN层在训练中使用一批数据的均值和方差对特征进行归一化，并在测试中使用整个训练集估计的均值和方差。当训练集和测试集的统计差别很大时，BN层趋向于引入令人不快的伪影并限制泛化能力。我们凭经验观察到，当网络较深且在GAN架构下训练时，BN层更可能带来伪影。这些伪影有时会在迭代中间和不同的设置下出现，违背了训练过程中对于稳定性能的需求。因此，我们为了稳定的训练和一致的性能移除了BN层。此外，移除BN层有助于提高泛化能力并降低计算复杂度及内存使用。

We keep the high-level architecture design of SRGAN (see Fig. 3), and use a novel basic block namely RRDB as depicted in Fig. 4. Based on the observation that more layers and connections could always boost performance [20,11,12], the proposed RRDB employs a deeper and more complex structure than the original residual block in SRGAN. Specifically, as shown in Fig. 4, the proposed RRDB has a residual-in-residual structure, where residual learning is used in different levels. A similar network structure is proposed in [36] that also applies a multilevel residual network. However, our RRDB differs from [36] in that we use dense block [34] in the main path as [11], where the network capacity becomes higher benefiting from the dense connections.

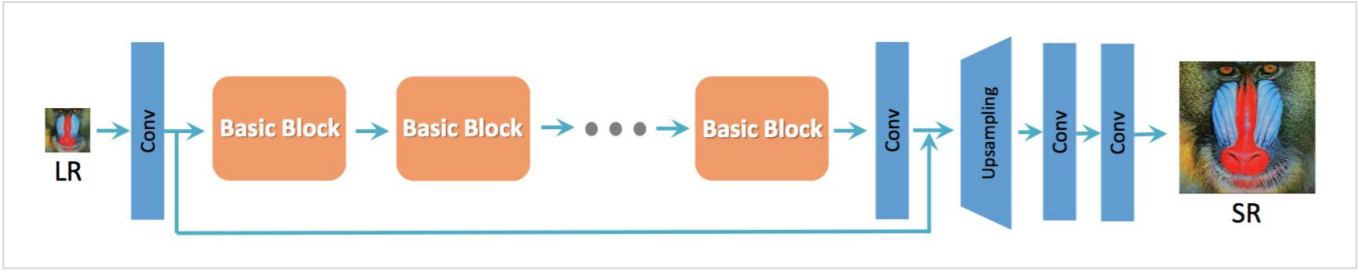


Fig. 3: We employ the basic architecture of SRResNet [1], where most computation is done in the LR feature space. We could select or design “basic blocks” (e.g., residual block [18], dense block [34], RRDB) for better performance.

我们保留了SRGAN的高级架构设计（见图3），并使用了一个新颖的名为RRDB的基本块，如图4所示。基于观测，更多的层和连接总是可以提升性能[20,11,12]，与SRGAN中的原始残差块相比，提出的RRDB采用了更深更复杂的架构。具体地说，如图4所示，提出了的RRDB有残差套残差的结构，其中残差学习用在不同的级别中。[36]中提出的类似结构也适用于多级残差网络。然而，我们的RRDB与[36]的不同在于我们在主路径中使用了如[11]的密集块[34]，受益于密集连接其网络容量变得更高。

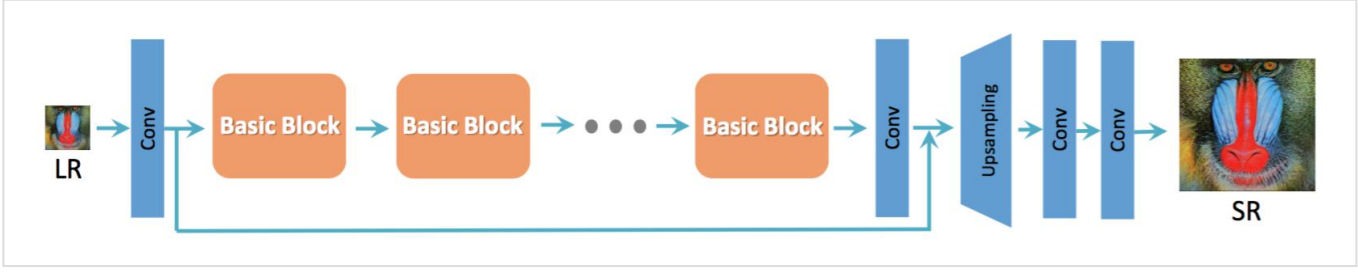


图3：我们采用SRResNet[1]的基本架构，大多数计算都在LR特征空间进行。我们可以为了更佳的性能选择或设计“基础块”（例如，残差块[18]，密集块[34]，RRDB）。

In addition to the improved architecture, we also exploit several techniques to facilitate training a very deep network: 1) residual scaling [21,20], i.e., scaling down the residuals by multiplying a constant between 0 and 1 before adding them to the main path to prevent instability; 2) smaller initialization, as we empirically find residual architecture is easier to train when the initial parameter variance becomes smaller. More discussion can be found in the *supplementary material*.

除了改进架构之外，我们也利用几种技术来促进训练非常深的网络：1）残差缩放[21,20]，例如在将残差加到主路径上之前，通过将其乘以一个0-1之间的常量来缩小残差以防止不稳定性；2）更小的初始化，因为我们凭经验发现当初始参数方差变得更小时，残差结构更容易训练。更多讨论可在*补充材料*中找到。

The training details and the effectiveness of the proposed network will be presented in Sec. 4.

训练细节和提出网络的有效性将在第4节中介绍。

3.2 Relativistic Discriminator

Besides the improved structure of generator, we also enhance the discriminator based on the Relativistic GAN [2]. Different from the standard discriminator D
in SRGAN, which estimates the probability that one input image x is real and natural, a relativistic discriminator tries to predict the probability that a real
image x_r is relatively more realistic than a fake one x_f, as shown in Fig. 5.

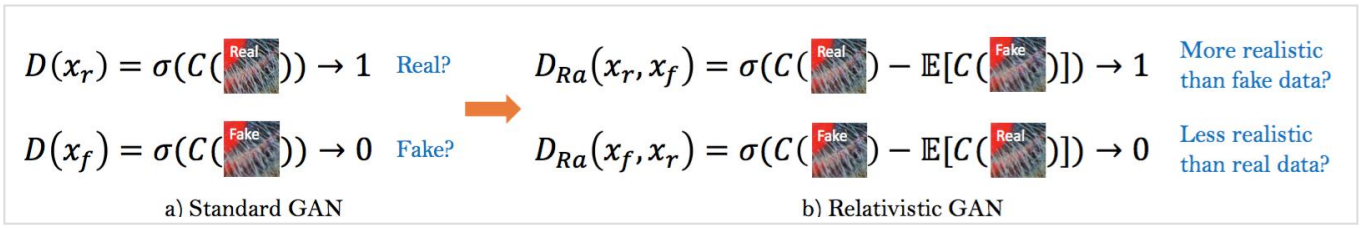


Fig. 5: Difference between standard discriminator and relativistic discriminator.

3.2 相对判别器

除了改进生成器架构之外，我们还在相对GAN[2]的基础上增强了判断器。不同于SRGAN中的标注判别器D，D估算输入图像x是真实自然的概率，相对判别器尝试预测真实图像x_r比假图像x_f相对更真实的概率，如图5所示。

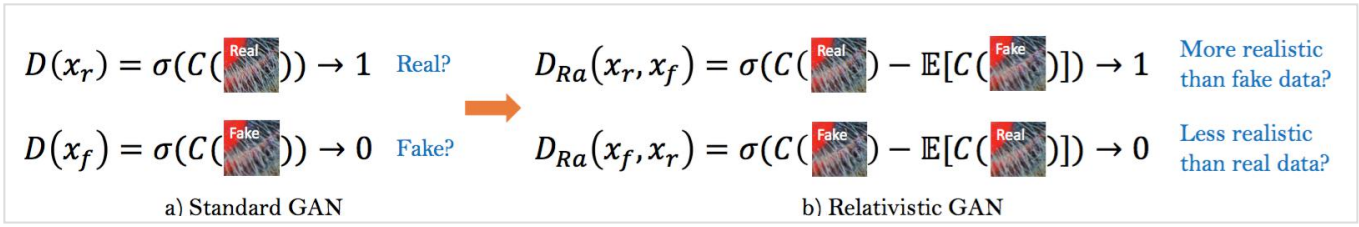


图5：标准判别器和相对判别器的差异。

Specifically, we replace the standard discriminator with the Relativistic average Discriminator RaD [2], denoted as D_{Ra}. The standard discriminator in SRGAN can be expressed as $D(x) = \sigma(C(x))$, where σ is the sigmoid function and $C(x)$ is the non-transformed discriminator output. Then the RaD is formulated as $D_{Ra}(x_r, x_f) = \sigma(C(x_r) - \mathbb{E}_{x_f}[C(x_f)])$, where $\mathbb{E}_{x_f}[\bullet]$ represents the operation of taking average for all fake data in the mini-batch. The discriminator loss is then defined as:

$$L_D^{Ra} = -\mathbb{E}_{x_r}[log(D_{Ra}(x_r, x_f))] - \mathbb{E}_{x_f}[1 - log(D_{Ra}(x_f, x_r))]. \quad (1)$$

The adversarial loss for generator is in a symmetrical form:

$$L_G^{Ra} = -\mathbb{E}_{x_r}[1 - log(D_{Ra}(x_r, x_f))] - \mathbb{E}_{x_f}[log(D_{Ra}(x_f, x_r))], \quad (2)$$

where $x_f = G(x_i)$ and x_i stands for the input LR image. It is observed that the adversarial loss for generator contains both x_r and x_f. Therefore, our generator benefits from the gradients from both generated data and real data in adversarial training, while in SRGAN only generated part takes effect. In Sec. 4.4, we will show that this modification of discriminator helps to learn sharper edges and more detailed textures.

具体来说，我们用相对平均判别器RaD[2]代替标准判别器，记为D_{Ra}。SRGAN中的标准判别器可表示为$D(x) = \sigma(C(x))$，其中σ是sigmoid函数，$C(x)$是非变换判别器输出。然后RaD用公式表示为$D_{Ra}(x_r, x_f) = \sigma(C(x_r) - \mathbb{E}_{x_f}[C(x_f)])$，其中$\mathbb{E}_{x_f}[\bullet]$表示对小批次中所有假数据取平均值的操作。然后判别器损失定义为：

$$L_D^{Ra} = -\mathbb{E}_{x_r}[log(D_{Ra}(x_r, x_f))] - \mathbb{E}_{x_f}[1 - log(D_{Ra}(x_f, x_r))]. \quad (1)$$

生成器的对抗损失呈对称形式：

$$L_G^{Ra} = -\mathbb{E}_{x_r}[1 - log(D_{Ra}(x_r, x_f))] - \mathbb{E}_{x_f}[log(D_{Ra}(x_f, x_r))], \quad (2)$$

其中$x_f = G(x_i)$和x_i代表输入LR图像。可以看出，生成器的对抗损失包含x_r和x_f。因此，在对抗训练中，我们的生成器受益于生成数据和真实数据的梯度，而在SRGAN中仅生成部分起作用。在4.4节中，我们将展示判别器的这种修改有助于学习更清晰的边缘和更细致的纹理。

3.3 Perceptual Loss

We also develop a more effective perceptual loss L_{percep} by constraining on features before activation rather than after activation as practiced in SRGAN.

3.3 感知损失

通过约束激活之前的特征而不是SRGAN中实践的激活之后的特征，我们还开发了一种更有效的感知损失L_{percep}。

Based on the idea of being closer to perceptual similarity [29,14], Johnson et al. [13] propose perceptual loss and it is extended in SRGAN [1]. Perceptual loss is previously defined on the activation layers of a pre-trained deep network, where the distance between two activated features is minimized. Contrary to the convention, we propose to use features before the activation layers, which will overcome two drawbacks of the original design. First, the activated features are very sparse, especially after a very deep network, as depicted in Fig. 6. For example, the average percentage of activated neurons for image 'baboon' after VGG19-54 layer is merely 11.17%. The sparse activation provides weak supervision and thus leads to inferior performance. Second, using features after activation also causes inconsistent reconstructed brightness compared with the ground-truth image, which we will show in Sec. 4.4.

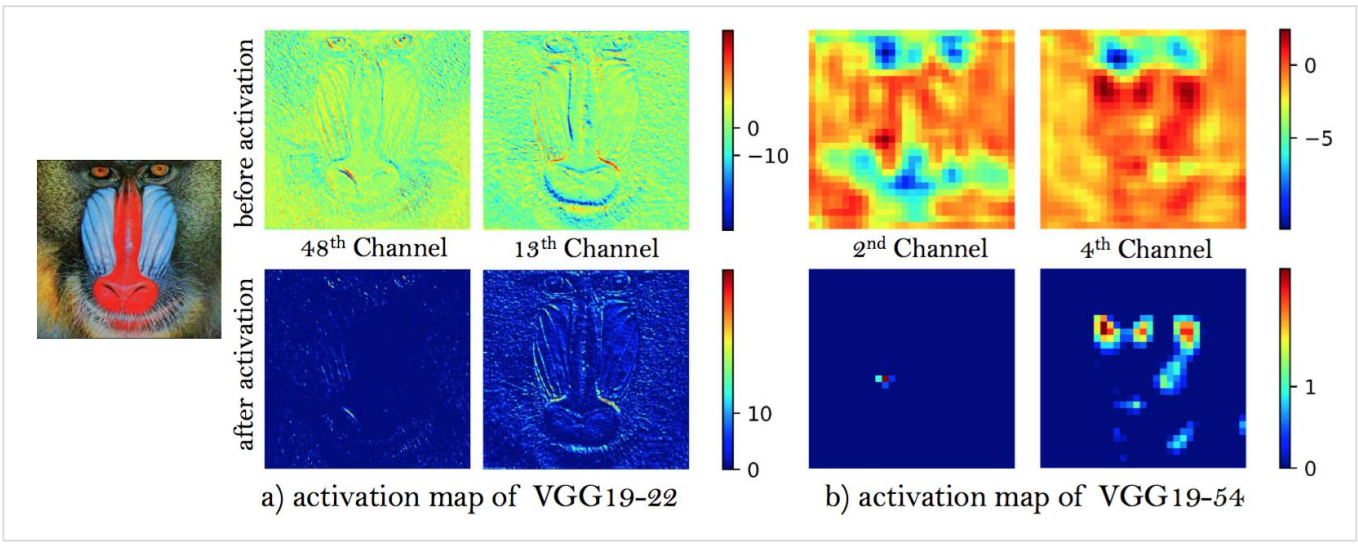


Fig.6: Representative feature maps before and after activation for image 'baboon'. With the network going deeper, most of the features after activation become inactive while features before activation contains more information.

基于更接近感知相似[29,14]的想法，Johnson等[13]提出了感知损失并在SRGAN[1]中得到了扩展。之前的感知损失定义在预训练深度网络的激活层上，最小化两个激活特征之间的距离。与常规用法相反，我们提出使用激活层之前的特征，这将克服原始设计的两个缺点。首先，激活特征非常稀疏，尤其是在非常深的网络之后，如图6所示。例如，图像"狒狒"在VGG19-54层之后激活神经元的平均百分比只有11.17%。稀疏的激活提供了弱

监督，因此导致性能较差。其次，与真实图像相比，使用激活之后的特征也会导致重建亮度不一致，这将在4.4节中展示。

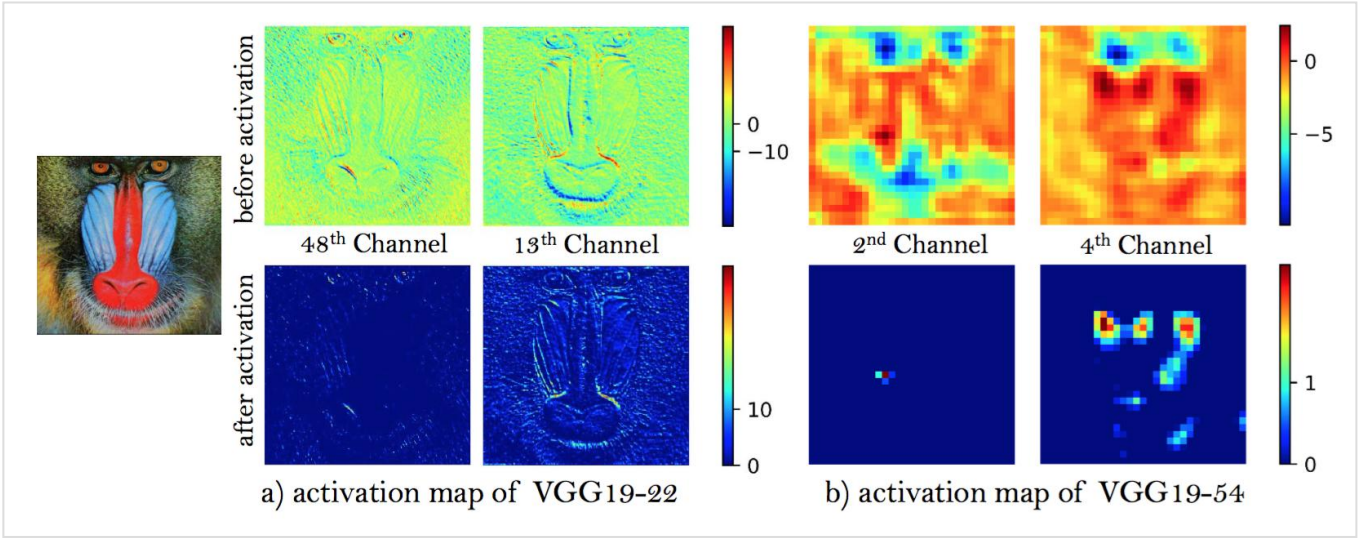


图6：图像“狒狒”激活之前和激活之后代表性的特征映射。随着网络加深，大多数激活之后的特征变得不活跃而激活之前的特征包含更多的信息。

Therefore, the total loss for the generator is:

$$L_G = L_{percep} + \lambda L_G^{Ra} + \eta L_1 \tag{3}$$

where $L_1 = \mathbb{E}_{x_i}||G(x_i) - y||_1$ is the content loss that evaluate the 1-norm distance between recovered image $G(x_i)$ and the ground-truth y, and λ, η are the coefficients to balance different loss terms.

因此，生成器的全部损失为：

$$L_G = L_{percep} + \lambda L_G^{Ra} + \eta L_1 \tag{3}$$

，其中$L_1 = \mathbb{E}_{x_i}||G(x_i) - y||_1$是内容损失，用来评估恢复图像$G(x_i)$和真实图像$y$之间的1范数距离，$\lambda, \eta$是平衡不同损失项的系数。

We also explore a variant of perceptual loss in the PIRM-SR Challenge. In contrast to the commonly used perceptual loss that adopts a VGG network trained for image classification, we develop a more suitable perceptual loss for SR–MINC loss. It is based on a fine-tuned VGG network for material recognition [38], which focuses on textures rather than object. Although the gain of perceptual index brought by MINC loss is marginal, we still believe that exploring perceptual loss that focuses on texture is critical for SR.

我们在PIRM-SR挑战赛中也探索了感知损失的变种。与采用图像分类训练的VGG网络的常用感知损失相比，我们为SR–MINC损失开发了一种更合适的感知损失。它是基于材料识别[38]的微调VGG网络，该网络注重于纹理

而不是目标。尽管MINC损失带来的感知指数收益是微不足道的，但我们仍然认为，采用注重纹理的感知损失对于SR至关重要。

3.4 Network Interpolation

To remove unpleasant noise in GAN-based methods while maintain a good perceptual quality, we propose a flexible and effective strategy – network interpolation. Specifically, we first train a PSNR-oriented network G_{PSNR} and then obtain a GAN-based network G_{GAN} by fine-tuning. We interpolate all the corresponding parameters of these two networks to derive an interpolated model G_{INTERP}, whose parameters are:

$$\theta_G^{INTERP} = (1-\alpha)\theta_G^{PSNR} + \alpha\theta_G^{GAN} \tag{4}$$

where G_{INTERP}, G_{PSNR} and G_{GAN} are the parameters of θ_G^{INTERP}, θ_G^{PSNR} and θ_G^{GAN}, respectively, and $\alpha \in [0, 1]$ is the interpolation parameter.

3.4 网络插值

为了去除基于GAN方法中讨厌的噪声同时保持好的感知质量，我们提出了一种弹性有效的策略——网络插值。具体来说，我们首先训练一个面向PSNR的网络G_{PSNR}，然后通过微调获得一个基于GAN的网络G_{GAN}。我们插值这两个网络的所有对应参数来取得插值模型G_{INTERP}，其参数为：

$$\theta_G^{INTERP} = (1-\alpha)\theta_G^{PSNR} + \alpha\theta_G^{GAN} \tag{4}$$

其中G_{INTERP}, G_{PSNR}和G_{GAN}分别是θ_G^{INTERP}, θ_G^{PSNR}和θ_G^{GAN}的参数，$\alpha \in [0, 1]$为插值参数。

The proposed network interpolation enjoys two merits. First, the interpolated model is able to produce meaningful results for any feasible α without introducing artifacts. Second, we can continuously balance perceptual quality and fidelity without re-training the model.

提出的网络插值有两个优点。首先，插值模型对于任何合理的α都能产生有意义的结果而不会产生伪影。其次，我们可以持续平衡感知质量和保真度都不必重新训练模型。

We also explore alternative methods to balance the effects of PSNR-oriented and GAN-based methods. For instance, one can directly interpolate their output images (pixel by pixel) rather than the network parameters. However, such an approach fails to achieve a good trade-off between noise and blur, i.e., the interpolated image is either too blurry or noisy with artifacts (see Sec. 4.5). Another method is to tune the weights of content loss and adversarial loss, i.e., the parameter λ and η in Eq. (3). But this approach requires tuning loss weights and fine-tuning the network, and thus it is too costly to achieve continuous control of the image style.

我们也探索了替代方法来平衡面向PSNR方法和基于GAN方法的影响。例如，可以直接插值它们的输出图像（逐像素）而不是网络参数。然而，这种方法不会在噪声和模糊之间取得良好的权衡，即插值图像或太模糊或

带有伪影的噪声太大（见4.5节）。另一种方法是调整内容损失和对抗损失的权重，即方程3中的参数λ和η。但这种方法要求调整损失权重并微调网络，因此实现图像风格的连续控制代价很高。

4 Experiments

4.1 Training Details

Following SRGAN [1], all experiments are performed with a scaling factor of ×4 between LR and HR images. We obtain LR images by down-sampling HR images using the MATLAB bicubic kernel function. The mini-batch size is set to 16. The spatial size of cropped HR patch is 128 × 128. We observe that training a deeper network benefits from a larger patch size, since an enlarged receptive field helps to capture more semantic information. However, it costs more training time and consumes more computing resources. This phenomenon is also observed in PSNR-oriented methods (see supplementary material).

4 实验

4.1 训练细节

按照SRGAN[1]，所有实验在LR和HR图像间均以4倍的尺度系数进行。我们通过使用MATLAB双三次核函数对HR图像进行下采样来获得LR图像。最小批次大小设置为16。裁剪的HR图像块的空间大小为128×128。我们观察到，训练更深的网络可以从更大的批次大小中获益，因为扩大的感受野有助于捕获更多的语义信息。但是，这会花费更多的训练时间并消耗更多的计算资源。这种现象也可以在面向PSNR的方法中观察到（见补充材料）。

The training process is divided into two stages. First, we train a PSNR-oriented model with the L1 loss. The learning rate is initialized as 2×10^{-4} and decayed by a factor of 2 every 2×10^5 of mini-batch updates. We then employ the trained PSNR-oriented model as an initialization for the generator. The generator is trained using the loss function in Eq. (3) with $\lambda = 5 \times 10^{-3}$ and $\eta = 1 \times 10^{-2}$. The learning rate is set to 1×10^{-4} and halved at [50k, 100k, 200k, 300k] iterations. Pre-training with pixel-wise loss helps GAN-based methods to obtain more visually pleasing results. The reasons are that 1) it can avoid undesired local optima for the generator; 2) after pre-training, the discriminator receives relatively good super-resolved images instead of extreme fake ones (black or noisy images) at the very beginning, which helps it to focus more on texture discrimination.

训练过程分为两个阶段。首先，我们训练一个具有L1损失的面向PSNR的模型。学习率初始化为2×10^{-4}，每2×10^5个小批次更新的衰减因子为2。然后，我们采用训练的面向PSNR的模型作为生成器的初始化。生成器训练使用等式3中的损失函数，$\lambda = 5 \times 10^{-3}$，$\eta = 1 \times 10^{-2}$。学习率设置为$1 \times 10^{-4}$，在[50k, 100k, 200k, 300k]次迭代之后减半。使用逐像素损失进行预训练有助于基于GAN的方法获得视觉上更好的结果。原

因是：1）它可以避免生成器不希望的局部最优；2）在预训练之后，最初判别器可以收到相对好的超分辨率图像而不是极端假的图像（黑色或噪声图像），这有助于其更关注纹理判别。

For optimization, we use Adam [39] with $\beta_1 = 0.9, \beta_2 = 0.999$. We alternately update the generator and discriminator network until the model converges. We use two settings for our generator – one of them contains 16 residual blocks, with a capacity similar to that of SRGAN and the other is a deeper model with 23 RRDB blocks. We implement our models with the PyTorch framework and train them using NVIDIA Titan Xp GPUs.

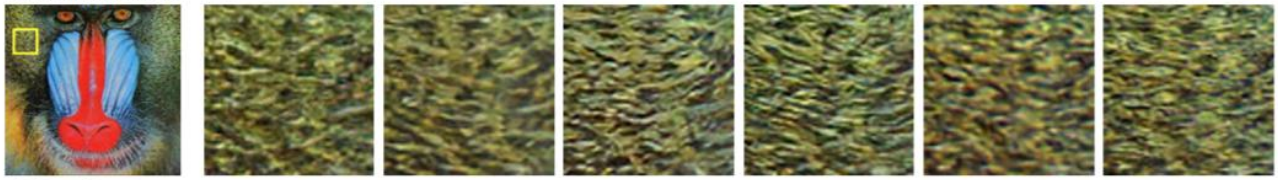
为了优化，我们使用Adam[39]，其中$\beta_1 = 0.9$，$\beta_2 = 0.999$。我们交替更新生成器和判别器网络，直到模型收敛。我们为生成器使用了两种设置——其中一种包含16个残差块，能力类似于SRGAN，另一种是具有23个RRDB块的更深的模型。我们使用PyTorch框架实现我们的模型，并使用NVIDIA Titan Xp GPU对其进行训练。

4.2 Data

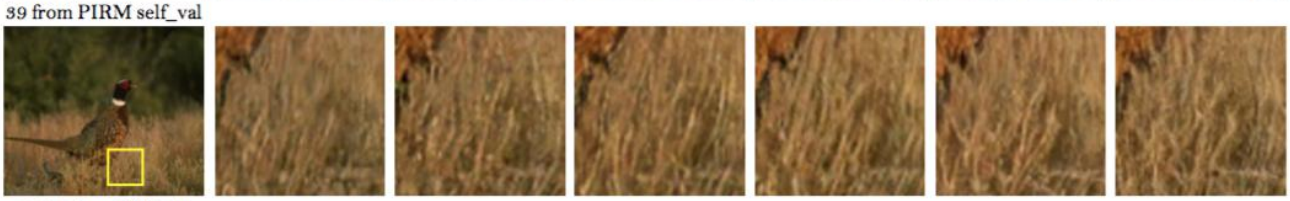
For training, we mainly use the DIV2K dataset [40], which is a high-quality (2K resolution) dataset for image restoration tasks. Beyond the training set of DIV2K that contains 800 images, we also seek for other datasets with rich and diverse textures for our training. To this end, we further use the Flickr2K dataset [41] consisting of 2650 2K high-resolution images collected on the Flickr website, and the OutdoorSceneTraining (OST) [17] dataset to enrich our training set. We empirically find that using this large dataset with richer textures helps the generator to produce more natural results, as shown in Fig. 8.

1st	2nd	3rd	4th	5th	6th	7th
BN?	✓	✗	✗	✗	✗	✗
Activation?	After	After	Before	Before	Before	Before
GAN?	Standard GAN	Standard GAN	Standard GAN	RaGAN	RaGAN	RaGAN
Deeper with RRDB?	✗	✗	✗	✗	✓	✓
More data?	✗	✗	✗	✗	✗	✓


baboon from Set14


baboon from Set14


39 from PIRM self_val


43074 from BSD100


69015 from BSD100


6 from PIRM self_val


20 from PIRM self_val


208001 from BSD100

Fig. 8: Overall visual comparisons for showing the effects of each component in ESRGAN. Each column represents a model with its configurations in the top. The red sign indicates the main improvement compared with the previous model.

4.2 数据

对于训练，我们主要使用DIV2K数据集[40]，它是用于图像复原任务的高质量（2K分辨率）数据集。除了包含800张图像的DIV2K训练集外，我们也搜寻了其它具有丰富多样纹理的数据集进行训练。为此，我们进一步使用Flickr2K数据集[41]，包含Flickr网站上收集的2650张2K高分辨率图像，OutdoorSceneTraining(OST)[17]数据集来丰富我们的训练集。我们凭经验发现，使用具有丰富纹理的大型数据集有助于生成器产生更自然的结果，如图8所示。

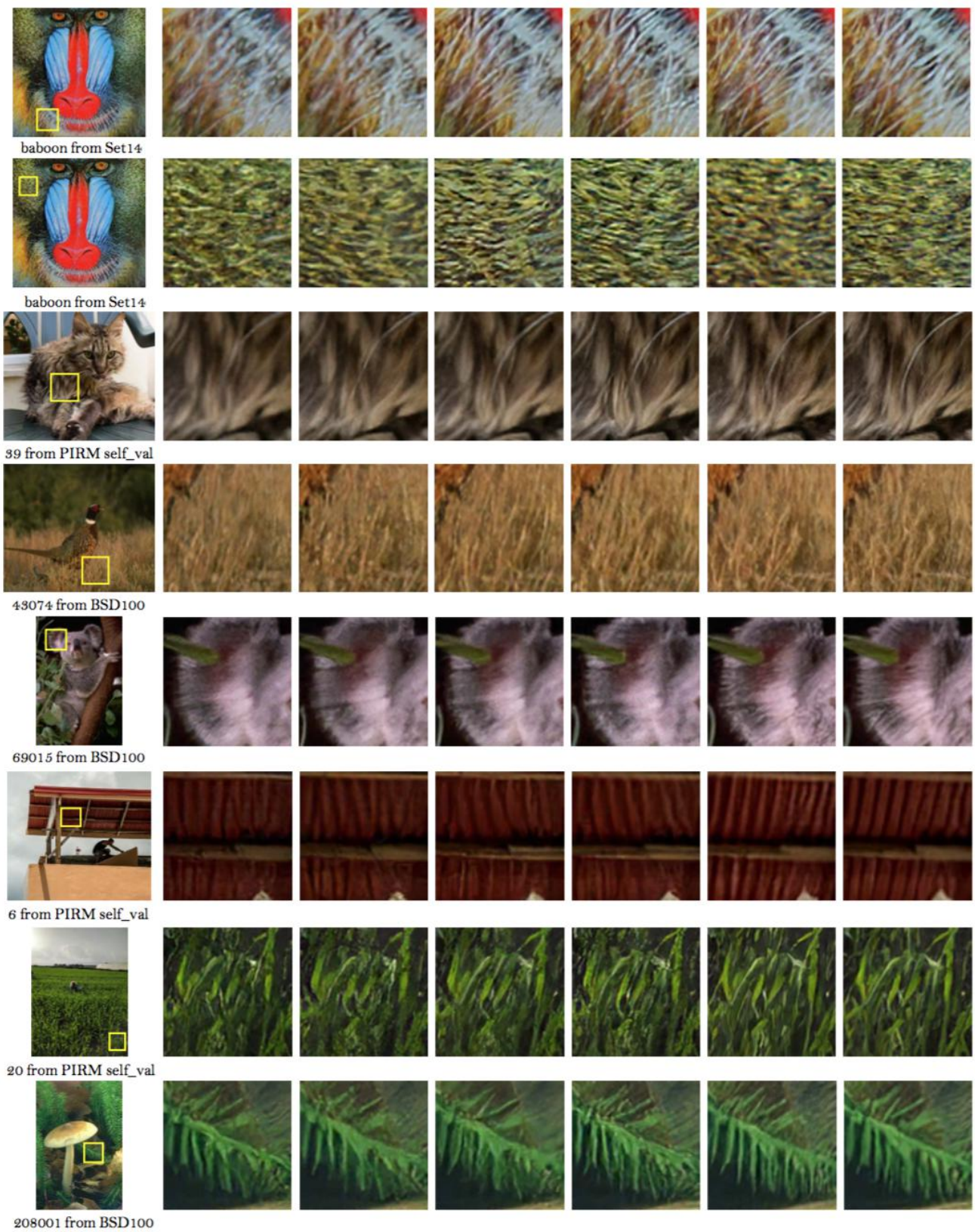

1st
2nd
3rd
4th
5th
6th
7th
BN?
Activation?
After
After
Before
Before
Before
Before
GAN?
Standard GAN
Standard GAN
Standard GAN
RaGAN
RaGAN
RaGAN
Deeper with RRDB?
More data?
baboon from Set14
baboon from Set14
39 from PIRM self_val
43074 from BSD100
69015 from BSD100
6 from PIRM self_val
20 from PIRM self_val
208001 from BSD100

图8：展示ESRGAN中每个组件效果的整体视觉比较。每一列表示一个模型，其配置在顶部。红色符号表示与前面模型相比的主要改进。

We train our models in RGB channels and augment the training dataset with random horizontal flips and 90 degree rotations. We evaluate our models on widely used benchmark datasets – Set5 [42], Set14 [43], BSD100 [44], Urban100 [45], and the PIRM self-validation dataset that is provided in the PIRM-SR Challenge.

我们在RGB通道训练模型，并通过随机水平翻转和90度旋转来增强训练集。我们在广泛使用的基准数据集——Set5[42]，Set14[43]，BSD100[44]，Urban100[45]以及PIRM-SR挑战赛提供的PIRM自验证数据上评估我们的模型。

4.3 Qualitative Results

We compare our final models on several public benchmark datasets with state-ofthe-art PSNR-oriented methods including SRCNN [4], EDSR [20] and RCAN [12], and also with perceptual-driven approaches including SRGAN [1] and EnhanceNet [16]. Since there is no effective and standard metric for perceptual quality, we present some representative qualitative results in Fig. 7. PSNR (evaluated on the luminance channel in YCbCr color space) and the perceptual index used in the PIRM-SR Challenge are also provided for reference.

4.3 定性结果

我们将最终的模型与最新的面向PSNR的方法包括SRCNN[4]，EDSR[20]和RCAN[12]，以及感知驱动的方法包括在SRGAN[1]和EnhanceNet[16]在一些公开基准数据集上进行了比较。由于对于感知质量没有有效标准的度量标准，我们在图7中展示了一些具有代表性的结果，也提供了PSNR（在YCbCr颜色空间的亮度通道上评估）和PIRM-SR挑战赛中的感知指数供参考。

It can be observed from Fig. 7 that our proposed ESRGAN outperforms previous approaches in both sharpness and details. For instance, ESRGAN can produce sharper and more natural baboon's whiskers and grass textures (see image 43074) than PSNR-oriented methods, which tend to generate blurry results, and than previous GAN-based methods, whose textures are unnatural and contain unpleasing noise. ESRGAN is capable of generating more detailed structures in building (see image 102061) while other methods either fail to produce enough details (SRGAN) or add undesired textures (EnhanceNet). Moreover, previous GAN-based methods sometimes introduce unpleasant artifacts, e.g., SRGAN adds wrinkles to the face. Our ESRGAN gets rid of these artifacts and produces natural results.

从图7可以看出，我们提出的ESRGAN在清晰度和细节方面都优于之前的方法。例如，与面向PSNR的方法（更趋向于产生模糊的结果）和以前的基于GAN的方法（纹理不自然并包含令人不快的噪声）相比，ESRGAN可以产生更清晰更自然的狒狒胡须和草的纹理（见图43074）。在建筑物中（见图102061），ESRGAN能够产生更

详细的结构而其它的方法要么不能产生足够的细节(SRGAN)，要么添加不必要的纹理(EnhanceNet)。此外，以前基于GAN的方法有时会引入令人不快的伪影，例如SRGAN会在脸上添加皱纹。我们的ESRGAN除去了这些伪影并产生了自然的结果。

4.4 Ablation Study

In order to study the effects of each component in the proposed ESRGAN, we gradually modify the baseline SRGAN model and compare their differences. The overall visual comparison is illustrated in Fig. 8. Each column represents a model with its configurations shown in the top. The red sign indicates the main improvement compared with the previous model. A detailed discussion is provided as follows.

4.4 消融研究

为了研究提出的ESRGAN中每个组件的效果，我们逐渐修改基准的SRGAN模型并比较它们的差异。完整的视觉比较如图8所示。每一列表示一个模型，其配置在顶部。红色符号表明与前面模型相比的主要改进。详细讨论提供如下。

BN removal. We first remove all BN layers for stable and consistent performance without artifacts. It does not decrease the performance but saves the computational resources and memory usage. For some cases, a slight improvement can be observed from the 2nd and 3rd columns in Fig. 8 (e.g., image 39). Furthermore, we observe that when a network is deeper and more complicated, the model with BN layers is more likely to introduce unpleasant artifacts. The examples can be found in the supplementary material.

移除BN。为了稳定和没有伪影的一致性能，我们首先移除了所有的BN层。它不会降低性能但会节省计算资源和内存使用。在某些情况下，从图8中的第2列和第3列可以观察到轻微的改进（例如，图39）。此外，我们观察到当网络更深更复杂时，具有BN层的模型更可能引入令人不快的伪影。可以在补充材料中找到示例。

Before activation in perceptual loss. We first demonstrate that using features before activation can result in more accurate brightness of reconstructed images. To eliminate the influences of textures and color, we filter the image with a Gaussian kernel and plot the histogram of its gray-scale counterpart. Fig. 9a shows the distribution of each brightness value. Using activated features skews the distribution to the left, resulting in a dimmer output while using features before activation leads to a more accurate brightness distribution closer to that of the ground-truth.

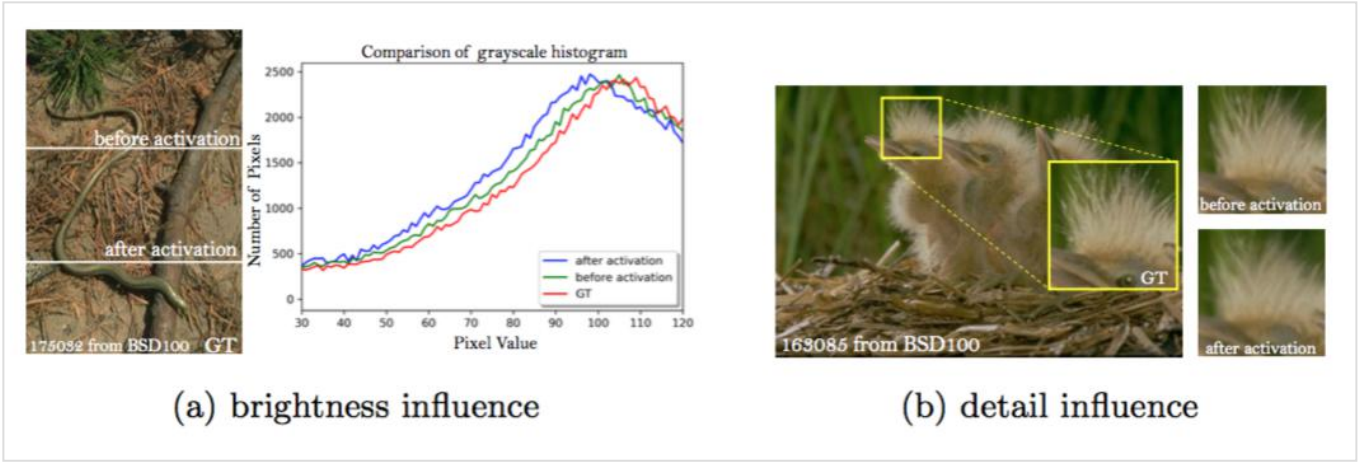


Fig. 9: Comparison between before activation and after activation.

感知损失在激活之前。我们首先证实了使用激活之前的特征可以使重建图像的亮度更准确。为了消除纹理和颜色的影响，我们使用高斯核对图像进行了滤波并绘制了其对应灰度图像的直方图。图9a展示了每一个亮度值的分布。使用激活的特征会使分布偏向左，导致了较暗的输出，而使用激活之前的特征会得到更精确的亮度分布，更接近于真实图像的亮度分布。

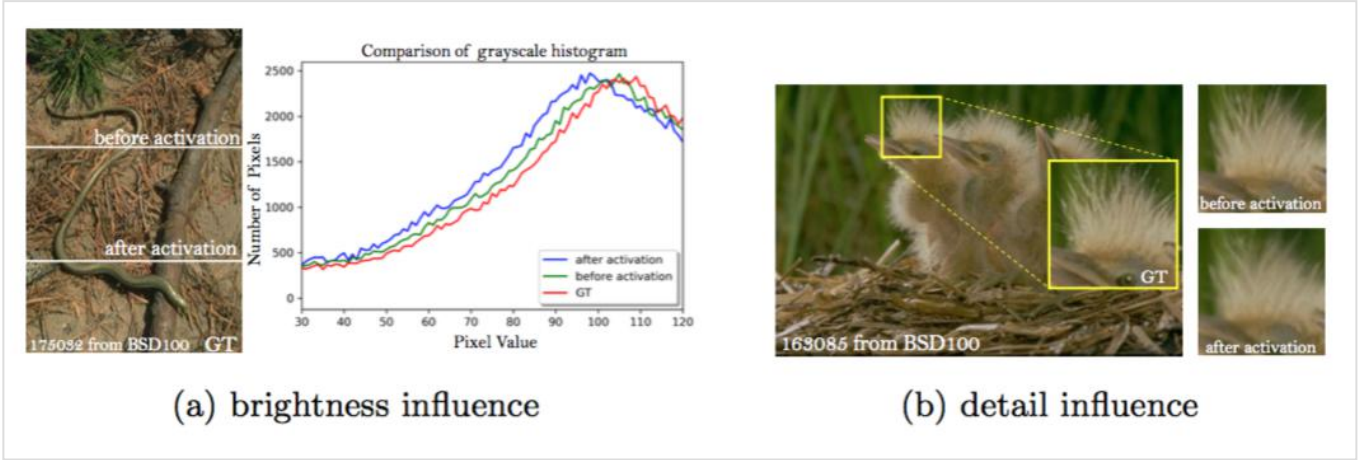


图9：激活之前和激活之后的比较。

We can further observe that using features before activation helps to produce sharper edges and richer textures as shown in Fig. 9b (see bird feather) and Fig. 8 (see the 3rd and 4th columns), since the dense features before activation offer a stronger supervision than that a sparse activation could provide.

我们可以进一步观察到，使用激活之前的特征有助于产生更清晰的边缘和更丰富的纹理，如图9b（见鸟羽）和图8（见第三列和第四列）所示，因为与稀疏激活提供的特征相比，激活之前的密集特征能提供更强的监督。

RaGAN. RaGAN uses an improved relativistic discriminator, which is shown to benefit learning sharper edges and more detailed textures. For example, in the 5th column of Fig. 8, the generated images are sharper with richer textures than those on their left (see the baboon, image 39 and image 43074).

RaGAN。RaGAN使用改进的相对判别器，证明了其有利于学习更清晰的边缘和更细致的纹理。例如，在图8的第5列中，生成的图像比其左侧的图像更清晰，具有更丰富的纹理（见狒狒，图39和图43074）。

Deeper network with RRDB. Deeper model with the proposed RRDB can further improve the recovered textures, especially for the regular structures like the roof of image 6 in Fig. 8, since the deep model has a strong representation capacity to capture semantic information. Also, we find that a deeper model can reduce unpleasing noises like image 20 in Fig. 8.

具有RRDB的更深网络。具有提出的RRDB的更深模型可以进一步改善恢复的纹理，尤其是像图8中图像6的屋顶这样的常规结构，因为深度模型具有强大的表示能力来捕获语义信息。 我们也发现更深的模型可以减少像图8中图像20这样的令人不快的噪声。

In contrast to SRGAN, which claimed that deeper models are increasingly difficult to train, our deeper model shows its superior performance with easy training, thanks to the improvements mentioned above especially the proposed RRDB without BN layers.

与SRGAN声称的更深的模型越来越难训练相比，由于上述提供的改进尤其是提出的没有BN层的RRDB，我们更深的模型展示了它容易训练且优越性能。

4.5 Network Interpolation

We compare the effects of network interpolation and image interpolation strategies in balancing the results of a PSNR-oriented model and GAN-based method. We apply simple linear interpolation on both the schemes. The interpolation parameter α is chosen from 0 to 1 with an interval of 0.2.

4.5 网络插值

我们比较了网络插值和图像插值策略在平衡面向PSNR模型与基于GAN方法的结果方面的作用。我们在这个两个方案中应用了简单的线性插值。插值参数α从间隔为0.2的0-1之间选取。

As depicted in Fig. 10, the pure GAN-based method produces sharp edges and richer textures but with some unpleasant artifacts, while the pure PSNRoriented method outputs cartoon-style blurry images. By employing network interpolation, unpleasing artifacts are reduced while the textures are maintained. By contrast, image interpolation fails to remove these artifacts effectively.

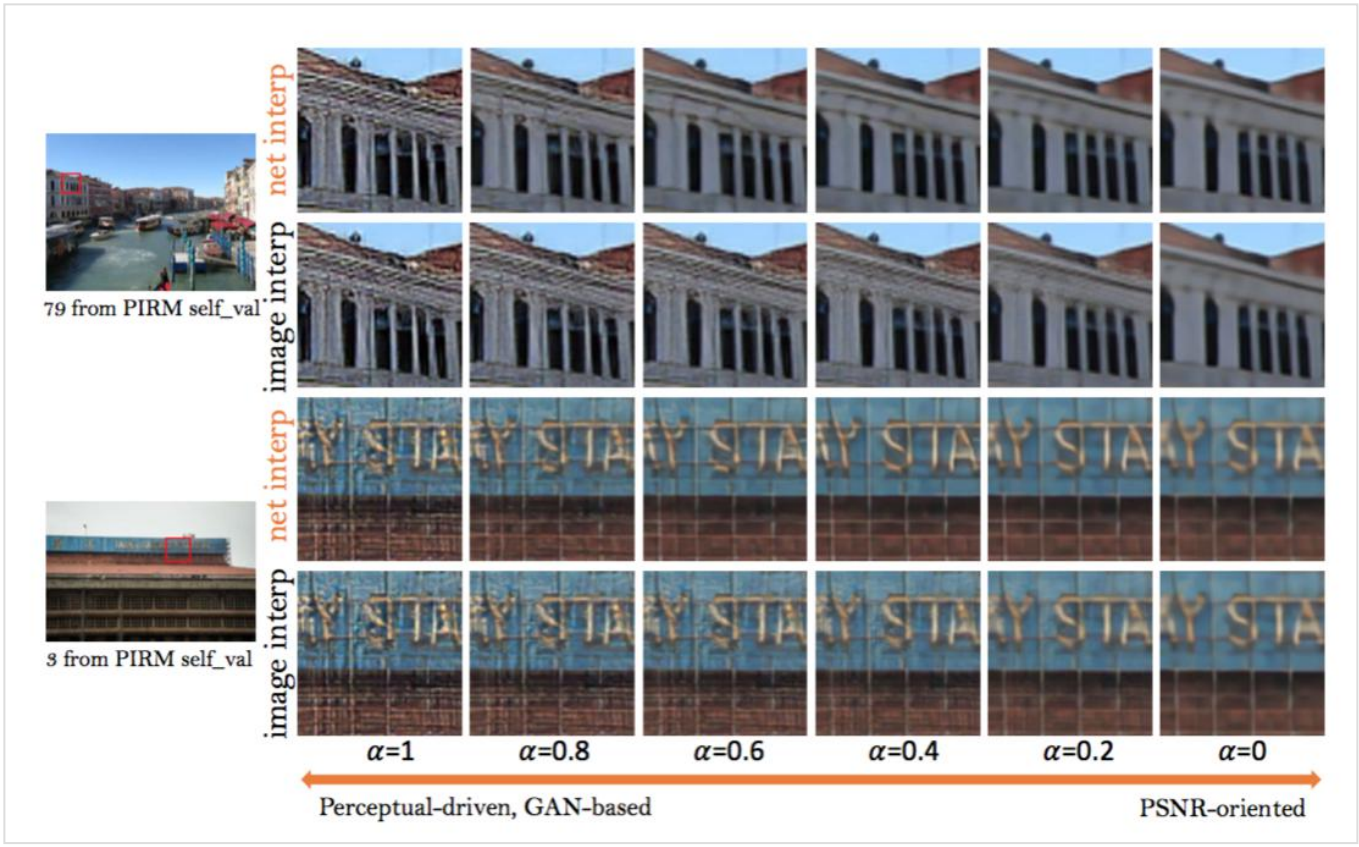


Fig. 10: The comparison between network interpolation and image interpolation.

如图10所示，单纯的基于GAN的方法会产生清晰的边缘和更丰富的纹理，但带有一些令人不快的伪影，而单纯的面向PSNR方法会输出卡通风格的模糊图像。通过采用网络插值，在减少令人不快的伪影的同时保持了纹理。相比之下，图像插值不能有效消除这些伪影。

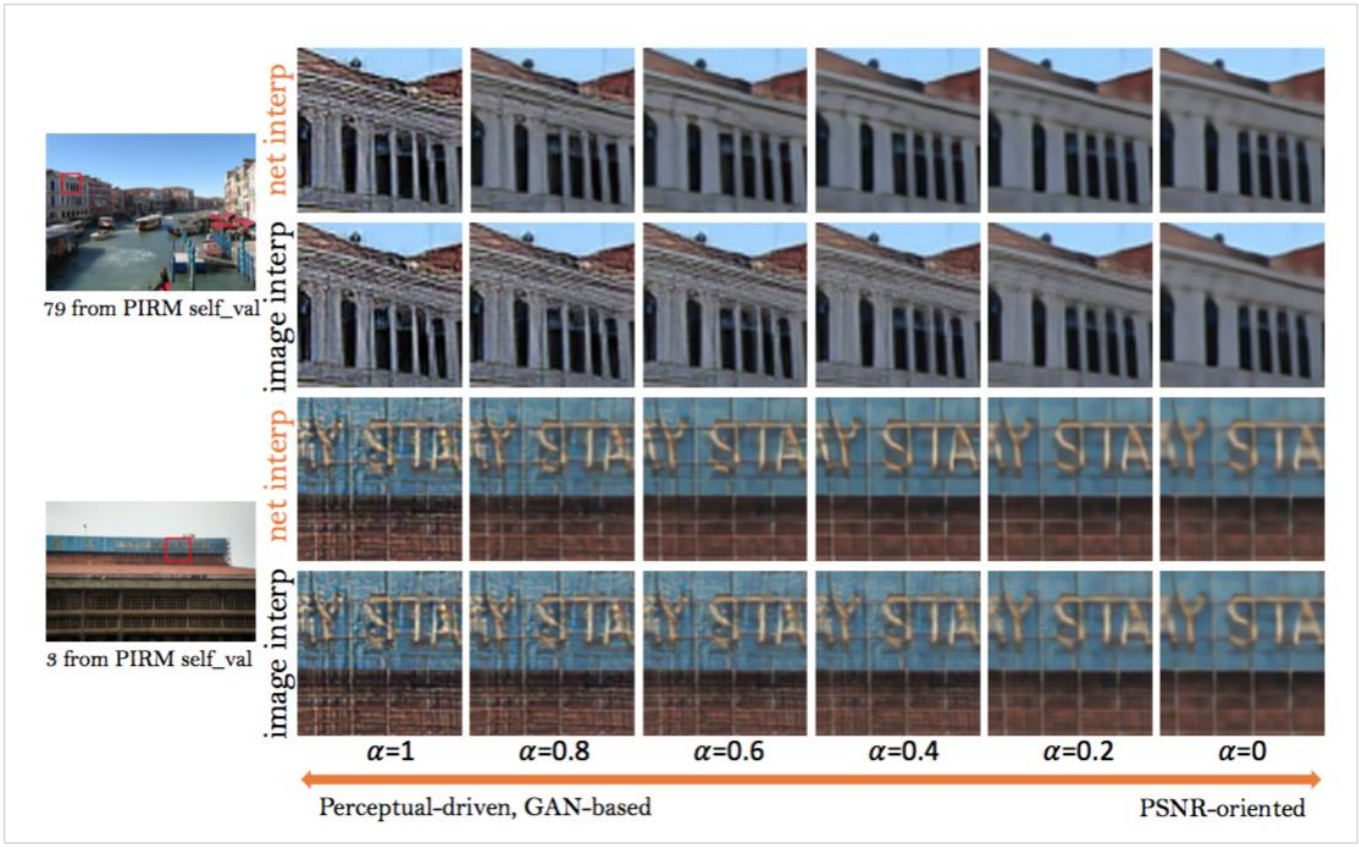


图10：网络插值和图像插值的比较。

Interestingly, it is observed that the network interpolation strategy provides a smooth control of balancing perceptual quality and fidelity in Fig. 10.

有趣的是，在图10中观察到网络插值策略提供了对平衡感知质量和保真度的平滑控制。

4.6 The PIRM-SR Challenge

We take a variant of ESRGAN to participate in the PIRM-SR Challenge [3]. Specifically, we use the proposed ESRGAN with 16 residual blocks and also empirically make some modifications to cater to the perceptual index. 1) The MINC loss is used as a variant of perceptual loss, as discussed in Sec. 3.3. Despite the marginal gain on the perceptual index, we still believe that exploring perceptual loss that focuses on texture is crucial for SR. 2) Pristine dataset [24], which is used for learning the perceptual index, is also employed in our training; 3) a high weight of loss L_1 up to $\eta = 10$ is used due to the PSNR constraints; 4) we also use back projection [46] as post-processing, which can improve PSNR and sometimes lower the perceptual index.

4.6 PIRM-SR挑战赛

我们采用ESRGAN的一个变种来参加PIRM-SR挑战赛[3]。具体来说，我们使用提出的具有16个残差块的ESRGAN，并根据经验进行了一些修改来迎合感知指数。1）使用MINC损失作为感知损失的一个变种，如3.3节所述。尽管在感知指数上有边际收益，但我们仍认为采用专注于纹理的感知损失对于SR至关重要；2）我们的训练中也使用了Pristine数据集[24]来学习感知指数；3）由于PSNR约束，L_1损失的权重高达$\eta = 10$；4）我们也使用反向投射[46]作为后处理，其可以改善PSNR，有时会降低感知指数。

For other regions 1 and 2 that require a higher PSNR, we use image interpolation between the results of our ESRGAN and those of a PSNR-oriented method RCAN [12]. The image interpolation scheme achieves a lower perceptual index (lower is better) although we observed more visually pleasing results by using the network interpolation scheme. Our proposed ESRGAN model won the first place in the PIRM-SR Challenge (region 3) with the best perceptual index.

对于其它需要较高PSNR的区域1和2，我们在ESRGAN的结果和面向PSNR方法RCAN[12]的结果之间使用图像插值。尽管通过使用网络插值方案我们观察到了视觉上更令人满意的效果，但图像插值方案取得了较低的感知指数（越低越好）。我们提出的ESRGAN模型以最好的感知指数赢得了PIRM-SR挑战赛（区域3）的第一名。

5 Conclusion

We have presented an ESRGAN model that achieves consistently better perceptual quality than previous SR methods. The method won the first place in the PIRM-SR Challenge in terms of the perceptual index. We have

formulated a novel architecture containing several RDDB blocks without BN layers. In addition, useful techniques including residual scaling and smaller initialization are employed to facilitate the training of the proposed deep model. We have also introduced the use of relativistic GAN as the discriminator, which learns to judge whether one image is more realistic than another, guiding the generator to recover more detailed textures. Moreover, we have enhanced the perceptual loss by using the features before activation, which offer stronger supervision and thus restore more accurate brightness and realistic textures.

5 结论

我们提出了一种ESRGAN模型，它比以前的SR方法始终取得更好的感知质量。就感知指数而言，该方法在PIRM-SR挑战赛中获得了第一名。我们构建了一种包含一些没有BN层的RDDB块的新颖架构。此外，采用了包括残差缩放和较小初始化的有用技术，以促进提出的深度模型的训练。我们还介绍了使用相对GAN作为判别器，其学习判断一张图像是否比另一张更真实，引导生成器恢复更详细的纹理。此外，我们通过使用激活之前的特征增强了感知损失，它提供了更强的监督，从而恢复了更精确的亮度和真实纹理。

Acknowledgement. This work is supported by SenseTime Group Limited, the General Research Fund sponsored by the Research Grants Council of the Hong Kong SAR (CUHK 14241716, 14224316. 14209217), National Natural Science Foundation of China (U1613211) and Shenzhen Research Program (JCYJ20170818164704758, JCYJ20150925163005055).

致谢。这项工作由商汤科技支持，香港特别行政区研究资助局（CUHK 14241716、14224316、14209217），中国国家自然科学基金（U1613211）和深圳研究计划（JCYJ20170818164704758，JCYJ20150925163005055）赞助。

References

1. Ledig,C.,Theis,L.,Huszá r,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: CVPR. (2017)

2. Jolicoeur-Martineau, A.: The relativistic discriminator: a key element missing from standard gan. arXiv preprint arXiv:1807.00734 (2018)

3. Blau, Y., Mechrez, R., Timofte, R., Michaeli, T., Zelnik-Manor, L.: The pirm challenge on perceptual super resolution. https://www.pirm2018.org/PIRM-SR. html (2018)

4. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: ECCV. (2014)

5. Kim, J., Kwon Lee, J., Mu Lee, K.: Accurate image super-resolution using very deep convolutional networks. In: CVPR. (2016)

6. Lai, W.S., Huang, J.B., Ahuja, N., Yang, M.H.: Deep laplacian pyramid networks for fast and accurate super-resolution. In: CVPR. (2017)

7. Kim, J., Kwon Lee, J., Mu Lee, K.: Deeply-recursive convolutional network for image super-resolution. In: CVPR. (2016)

8. Tai, Y., Yang, J., Liu, X.: Image super-resolution via deep recursive residual network. In: CVPR. (2017)

9. Tai, Y., Yang, J., Liu, X., Xu, C.: Memnet: A persistent memory network for image restoration. In: ICCV. (2017)

10. Haris, M., Shakhnarovich, G., Ukita, N.: Deep backprojection networks for super- resolution. In: CVPR. (2018)

11. Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y.: Residual dense network for image super-resolution. In: CVPR. (2018)

12. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. In: ECCV. (2018)

13. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: ECCV. (2016)

14. Bruna, J., Sprechmann, P., LeCun, Y.: Super-resolution with deep convolutional sufficient statistics. In: ICLR. (2015)

15. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: NIPS. (2014)

16. Sajjadi, M.S., Schö lkopf, B., Hirsch, M.: Enhancenet: Single image super-resolution through automated texture synthesis. In: ICCV. (2017)

17. Wang, X., Yu, K., Dong, C., Loy, C.C.: Recovering realistic texture in image super-resolution by deep spatial feature transform. In: CVPR. (2018)

18. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR. (2016)

19. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: ICMR. (2015)

20. Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M.: Enhanced deep residual networks for single image super-resolution. In: CVPRW. (2017)

21. Szegedy, C., Ioffe, S., Vanhoucke, V.: Inception-v4, inception-resnet and the impact of residual connections on learning. arXiv preprint arXiv:1602.07261 (2016)

22. Blau, Y., Michaeli, T.: The perception-distortion tradeoff. In: CVPR. (2017)

23. Ma, C., Yang, C.Y., Yang, X., Yang, M.H.: Learning a no-reference quality metric for single-image super-resolution. CVIU 158 (2017) 1–16

24. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a completely blind image quality analyzer. IEEE Signal Process. Lett. 20(3) (2013) 209–212

25. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. TPAMI 38(2) (2016) 295–307

26. Yu, K., Dong, C., Lin, L., Loy, C.C.: Crafting a toolchain for image restoration by deep reinforcement learning. In: CVPR. (2018)

27. Yuan, Y., Liu, S., Zhang, J., Zhang, Y., Dong, C., Lin, L.: Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. In: CVPRW. (2018)

28. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: ICCV. (2015)

29. Gatys, L., Ecker, A.S., Bethge, M.: Texture synthesis using convolutional neural networks. In: NIPS. (2015)

30. Mechrez, R., Talmi, I., Shama, F., Zelnik-Manor, L.: Maintaining natural image statistics with the contextual loss. arXiv preprint arXiv:1803.04626 (2018)

31. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein gan. arXiv preprint arXiv:1701.07875 (2017)

32. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C.: Improved training of wasserstein gans. In: NIPS. (2017)

33. Miyato, T., Kataoka, T., Koyama, M., Yoshida, Y.: Spectral normalization for generative adversarial networks. arXiv preprint arXiv:1802.05957 (2018)

34. Huang, G., Liu, Z., Weinberger, K.Q., van der Maaten, L.: Densely connected convolutional networks. In: CVPR. (2017)

35. Nah, S., Kim, T.H., Lee, K.M.: Deep multi-scale convolutional neural network for dynamic scene deblurring. In: CVPR. (2017)

36. Zhang, K., Sun, M., Han, X., Yuan, X., Guo, L., Liu, T.: Residual networks of residual networks: Multilevel residual networks. IEEE Transactions on Circuits and Systems for Video Technology (2017)

37. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)

38. Bell, S., Upchurch, P., Snavely, N., Bala, K.: Material recognition in the wild with the materials in context database. In: CVPR. (2015)

39. Kingma, D., Ba, J.: Adam: A method for stochastic optimization. In: ICLR. (2015)

40. Agustsson, E., Timofte, R.: Ntire 2017 challenge on single image super-resolution: Dataset and study. In: CVPRW. (2017)

41. Timofte, R., Agustsson, E., Van Gool, L., Yang, M.H., Zhang, L., Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M., et al.: Ntire 2017 challenge on single image super-resolution: Methods and results. In: CVPRW. (2017)

42. Bevilacqua, M., Roumy, A., Guillemot, C., Alberi-Morel, M.L.: Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In: BMVC, BMVA press (2012)

43. Zeyde, R., Elad, M., Protter, M.: On single image scale-up using sparse-representations. In: International Conference on Curves and Surfaces, Springer (2010)

44. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: ICCV. (2001)

45. Huang, J.B., Singh, A., Ahuja, N.: Single image super-resolution from transformed self-exemplars. In: CVPR. (2015)

46. Timofte, R., Rothe, R., Van Gool, L.: Seven ways to improve example-based single image super resolution. In: CVPR. (2016)

47. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: International Conference on Artificial Intelligence and Statistics. (2010)

Deep Learning

ESRGAN - Enhanced Super-Resolution Generative Adversarial Networks论文翻译——中文版

Python的"is None" vs "==None"