Single Image Super-Resolution Based On GAN

--From SRGAN to ESRGAN

Huapeng Wu

5/18/2019

Contents

Image super-resolution

图像超分辨率（Super Resolution, SR）就是将单幅或多幅低分辨率（Low Resolution, LR）的图像通过一定的算法重建到具有更高的像素密度，更多的细节信息，更细腻的画质的高分辨率图像(High Resolution, HR)。

超分辨是一种典型的不适定问题（ill-posed），因为对于任何一个输入的LR像素都存在多个HR像素的解。即两个不同的高分辨率图像块经过退化后可能得到相同的低分辨率图像块。

Image super-resolution

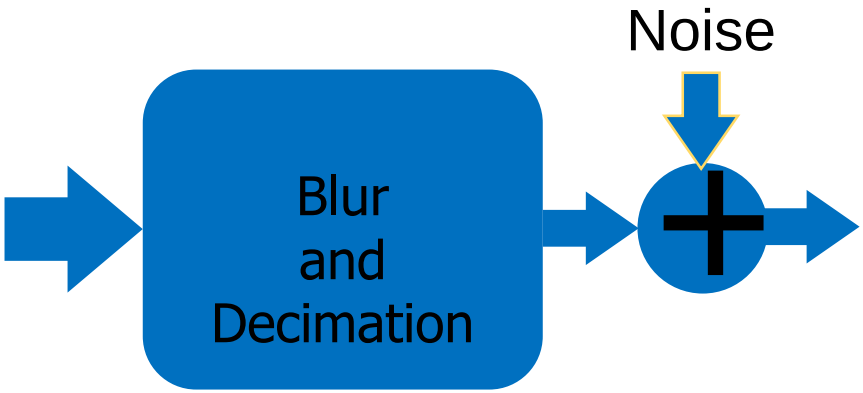
Problems:

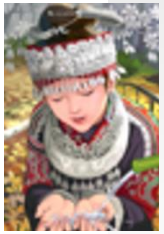
- **Multi-frame super-resolution**
- **Single image super-resolution**

Methods:

- **Interpolation based**
- **Reconstruction based**
- **Learning based**

Single Image Super-Resolution (SISR)

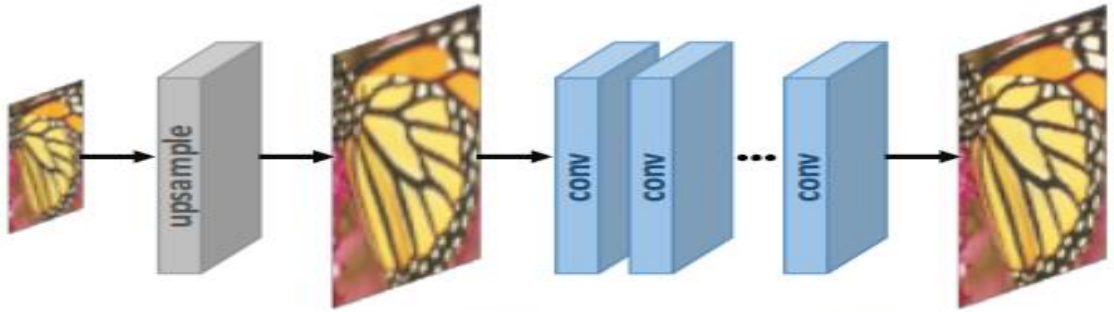




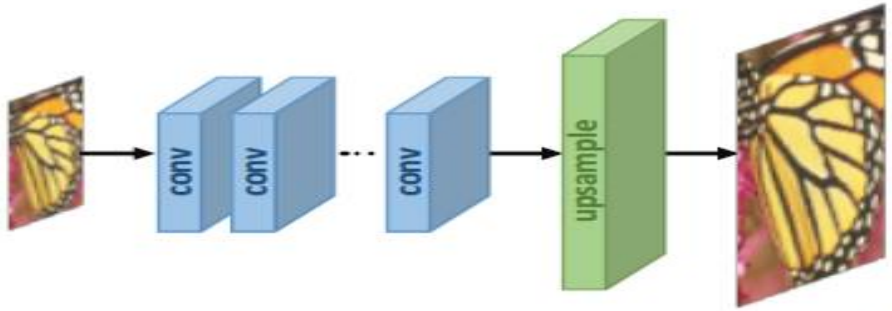
Recover the HR image from the LR one

SISR

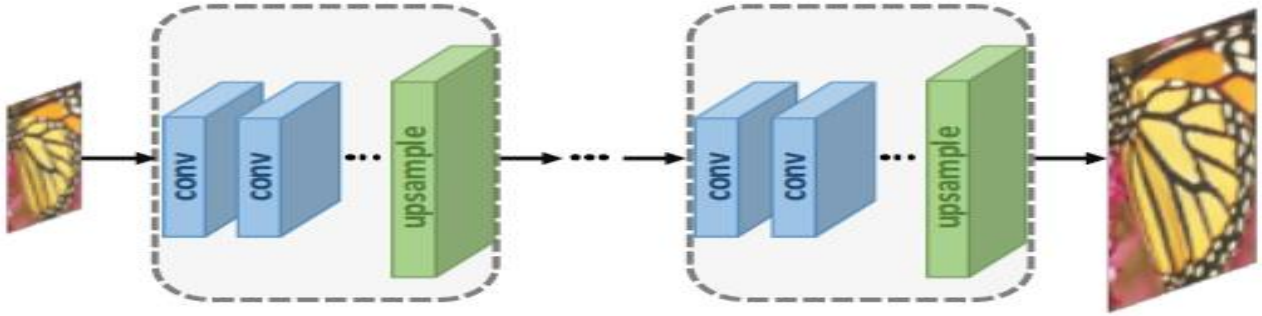
SISR (based on deep learning)



(a) Pre-upsampling SR



(b) Post-upsampling SR



(c) Progressive upsampling SR

SISR

The choice of the objective functioin is crucial. Recent work has largely focused on minimizing the mean squared reconstruction error (MSE).

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(I(i) - \hat{I}(i))^2,$$

$$\mathrm{PSNR} = 10 \cdot \log_{10}(\frac{L^2}{\mathrm{MSE}}).$$

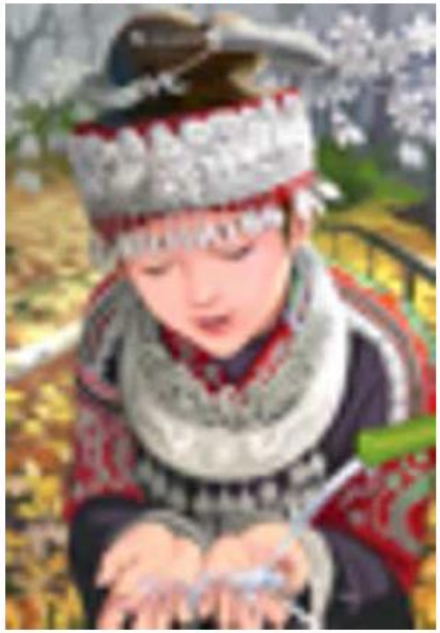
where I and $\hat{I}$ are the ground truth image and reconstructed image, both of which are with N pixels, L is fixed (general is 255). The PSNR is only related to the pixel-level MSE between images.

Reference: Wang, Zhihao, Jian Chen, and Steven CH Hoi. "Deep learning for image super-resolution: A survey." *arXiv preprint arXiv:1902.06068* (2019).

Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial (SRGAN)

The resulting estimates have high peak signal-to-noise ratios (PSNR), but they are often lacking high-frequency details and are perceptually unsatisfying in the sense.

bicubic (21.59dB/0.6423)



SRResNet (23.53dB/0.7832)

SRGAN (21.15dB/0.6868)

original

References: Wang, Zhihao, Jian Chen, and Steven CH Hoi. "Deep learning for image super-resolution: A survey." *arXiv preprint arXiv:1902.06068* (2019).

Ledig, Christian, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken 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*, pp. 4681-4690. 2017.

SRGAN

In SRGAN, they propose a perceptual loss function which consists of an adversarial loss and a content loss.

Perceptual loss function l^{SR} :

$$l^{SR} = \underbrace{\underbrace{l_{\mathrm{X}}^{SR}}_{\text{content loss}} + \underbrace{10^{-3} l_{Gen}^{SR}}_{\text{adversarial loss}}}_{\text{perceptual loss (for VGG based content losses)}}$$

Content loss:

1). Previous pixel-wise MSE loss:

SRGAN

$$l_{MSE}^{SR} = \frac{1}{r^2 WH} \sum_{x=1}^{rW} \sum_{y=1}^{rH} (I_{x,y}^{HR} - G_{\theta_G}(I^{LR})_{x,y})^2$$

2). To get closer to perceptual similarity, author define the VGG loss based on the ReLU activation layers of the pre-trained 19 layer VGG network.

$$l_{VGG/i.j}^{SR} = \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}(G_{\theta_G}(I^{LR}))_{x,y})^2$$

where I^{LR} and I^{HR} are LR image and HR image, r is downsampling factor. W, H, C are the size of LR image. G_{θ_G} is a generator network with parameter θ_G, With $\emptyset_{i,j}$ we indicate the feature map obtained by the j-th convolution (after activation) before the i-th maxpooling layer within the VGG19 network.

SRGAN

Adversarial loss:

GAN:
$$\min_{\theta_G}\max_{\theta_D}\ \mathbb{E}_{I^{HR}\sim p_{\text{train}}(I^{HR})}[\log D_{\theta_D}(I^{HR})]+ \mathbb{E}_{I^{LR}\sim pG(I^{LR})}[\log(1-D_{\theta_D}(G_{\theta_G}(I^{LR}))]$$

Our ultimate goal is to train a generating function G.

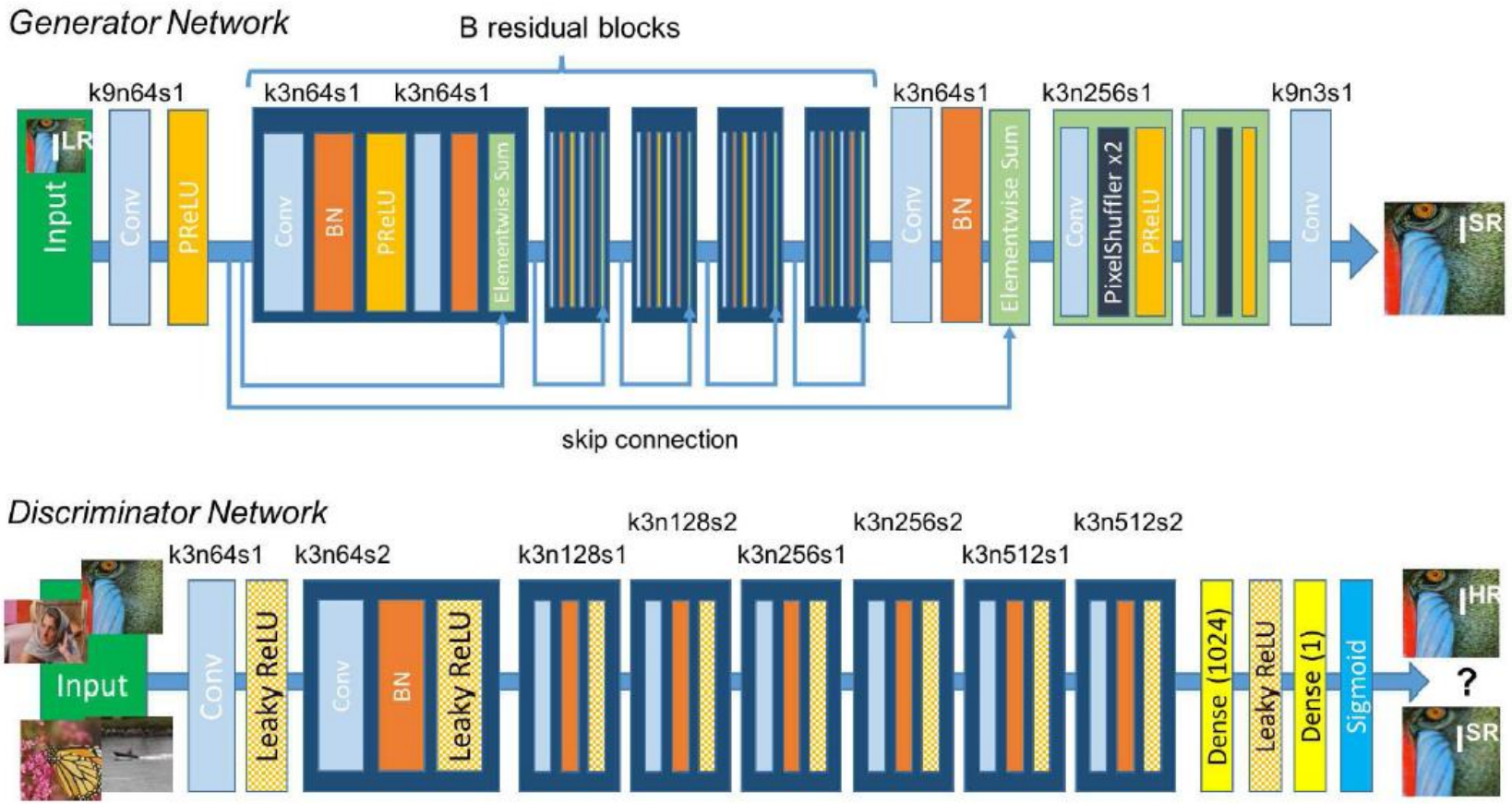
$$l^{SR}_{Gen}=\sum_{n=1}^{N}-\log D_{\theta_D}(G_{\theta_G}(I^{LR}))$$

The generative loss l^{SR}_{Gen} is defined based on the probabilities of the discriminator $D_{\theta_D}(G_{\theta_G}(I^{LR}))$ over all training samples.

Final loss : $$\hat{\theta}_G=\arg\min_{\theta_G}\frac{1}{N}\sum_{n=1}^{N}l^{SR}(G_{\theta_G}(I^{LR}_n),I^{HR}_n)$$

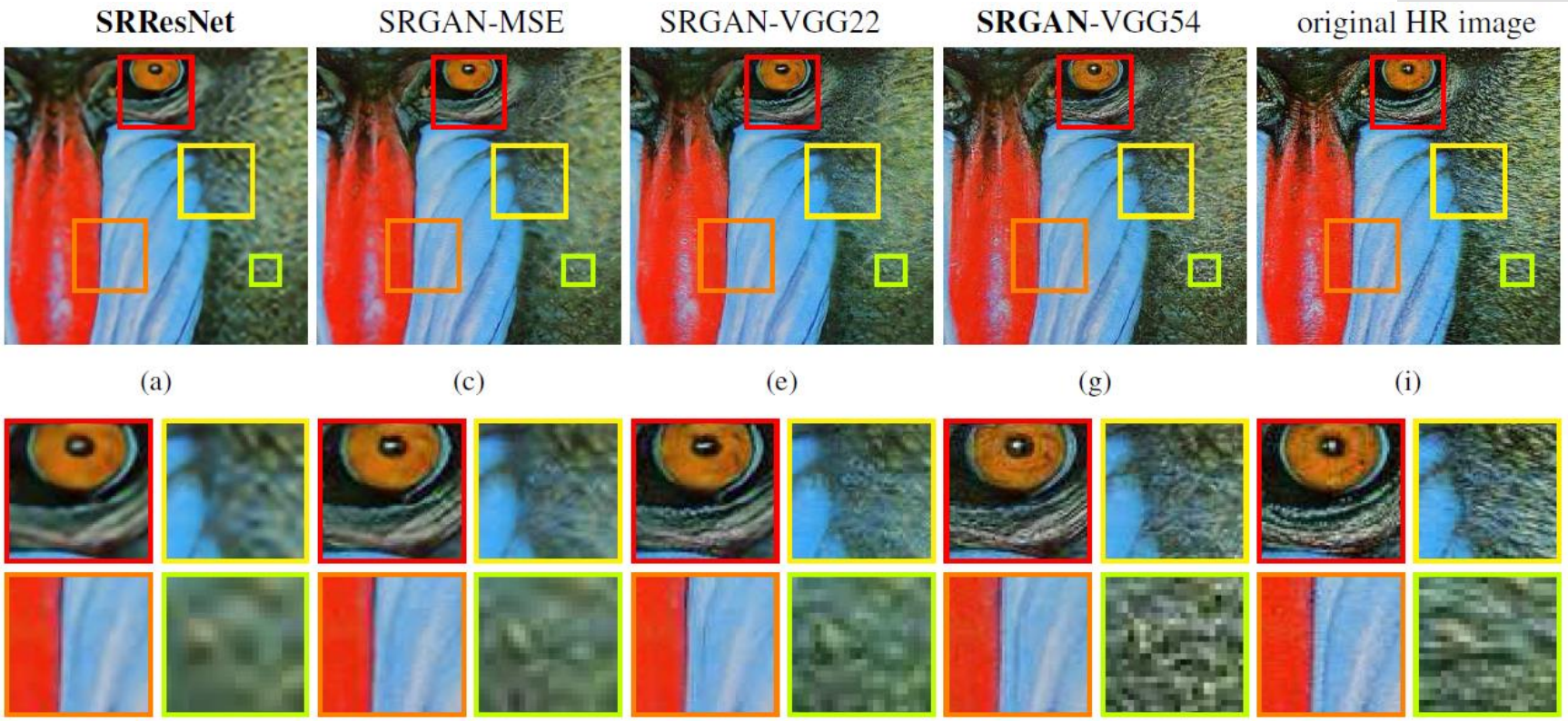
N is the number of images.

SRGAN



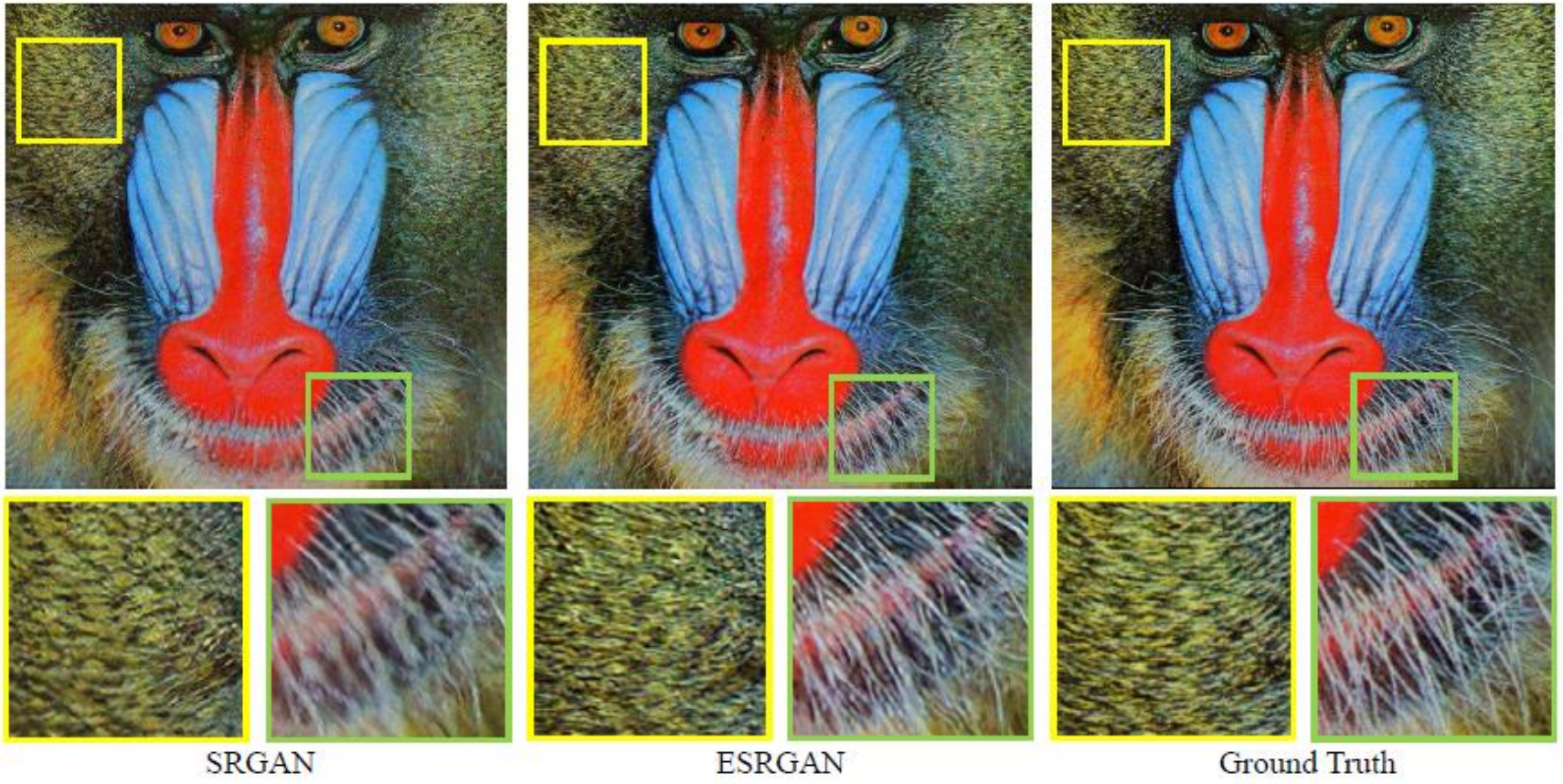
Architecture of Generator and Discriminator Network with corresponding kernel size (k), number of feature maps (n) and stride (s) indicated for each convolutional layer.

SRGAN


SRResNet
SRGAN-MSE
SRGAN-VGG22
SRGAN-VGG54
original HR image
(a)
(c)
(e)
(g)
(i)

ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks

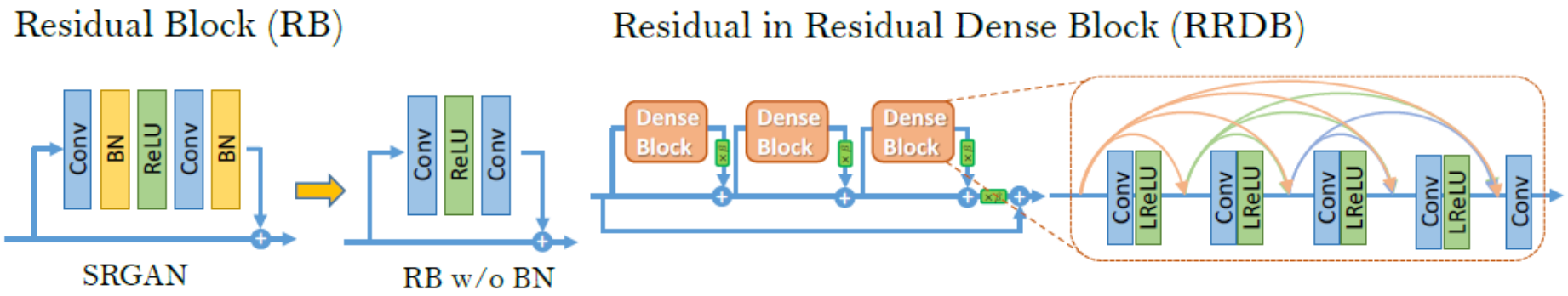
ESRGAN thoroughly study three key components of SRGAN – network architecture (remove BN), adversarial loss (relative realness) and perceptual loss (features before activation).



Reference: Wang, Xintao, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. "Esrgan: Enhanced super-resolution generative adversarial networks." In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 0-0. 2018.

ESRGAN

Network Architecture:



Left: 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 ([0,1], prevent instability).

ESRGAN

Remove BN: 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.

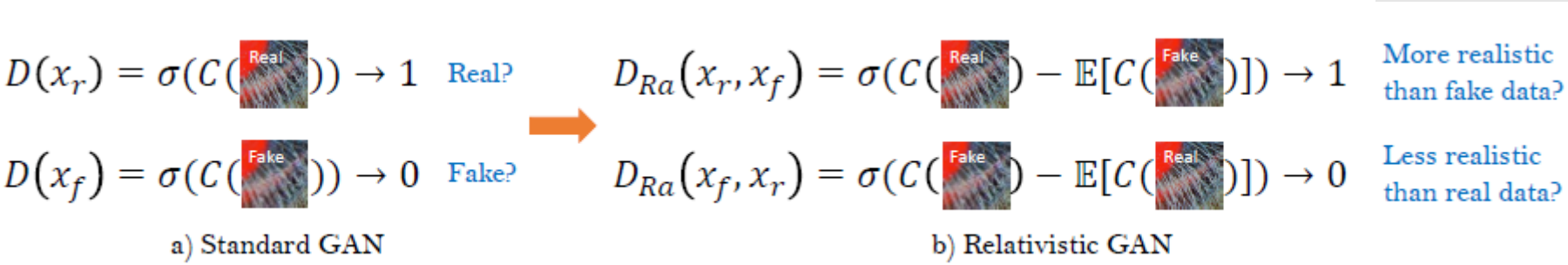
Removing BN layers helps to improve generalization ability and to reduce computational complexity and memory usage.

BN

Without BN

ESRGAN

Relativistic Discriminator: Relativistic average GAN (RaGAN), which learns to judge "whether one image is more realistic than the other" rather than "whether one image is real or fake".



where σ is the sigmoid function and C(x) is the non-transformed discriminator output, $E_{x_f}(\cdot)$ represents the operation of taking average for all fake data in the mini-batch.

The discriminator loss:

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

Reference: Jolicoeur-Martineau, Alexia. "The relativistic discriminator: a key element missing from standard GAN." arXiv preprint arXiv:1807.00734 (2018).

ESGAN

The adversarial loss for generator:

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

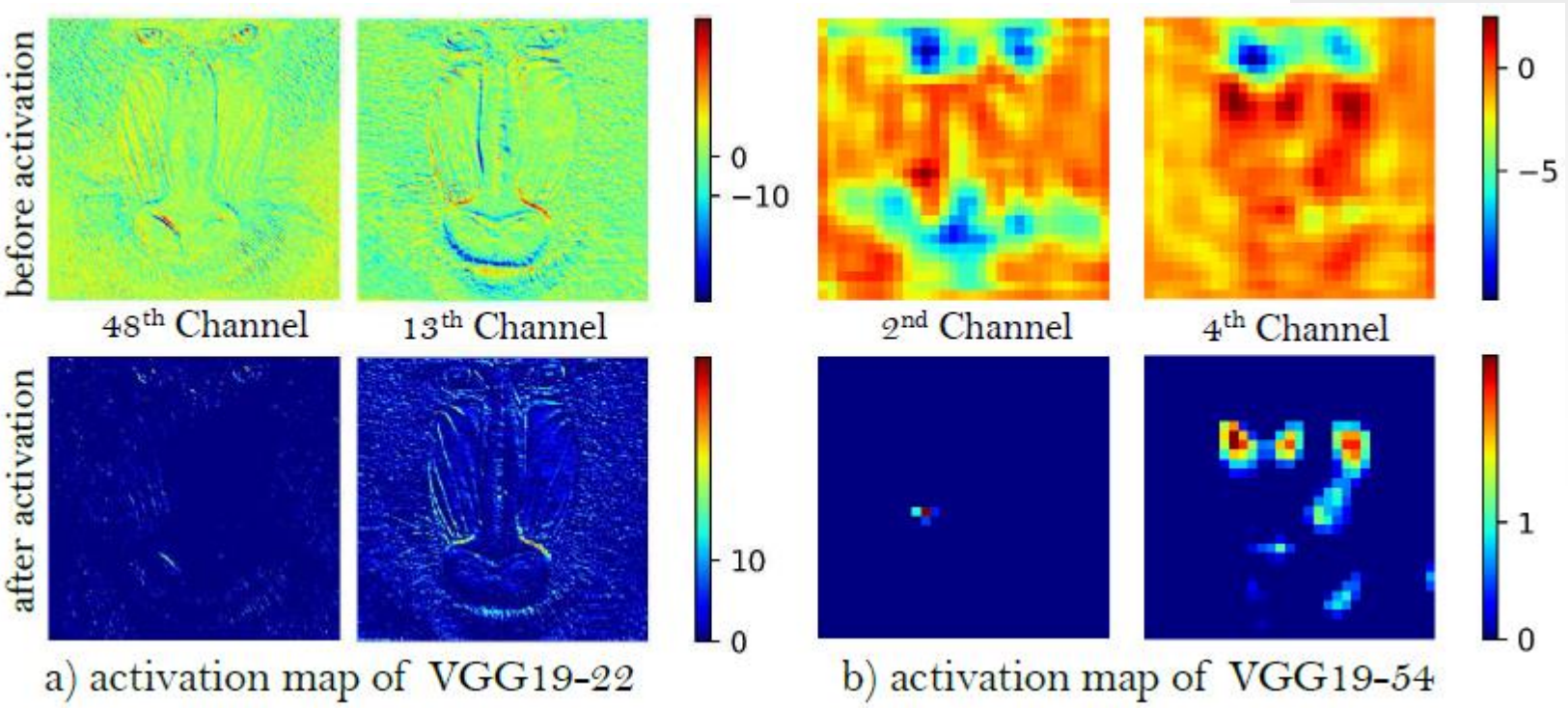
ESGAN benefits from the gradients from both generated data and real data in adversarial training, while in SRGAN only generated part takes effect.

SRGAN RaGAN

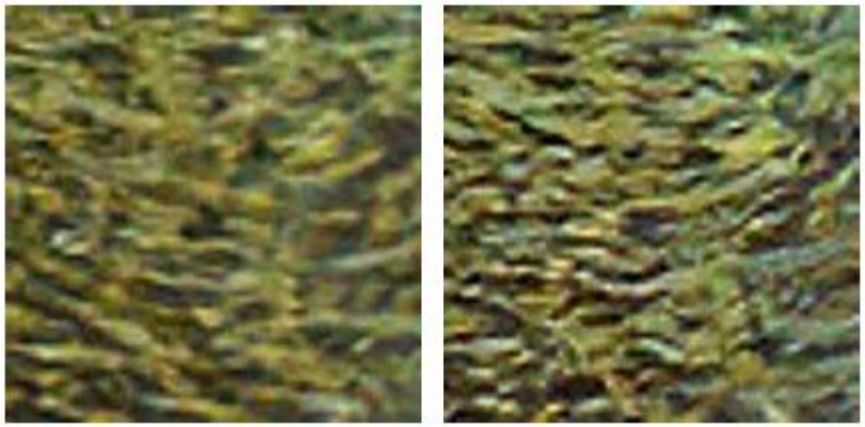
ESGAN

Perceptual loss: constrain on features before activation rather than after activation as practiced in SRGAN.

The activated features are very sparse, especially after a very deep network:



ESRGAN



after before

The total loss:

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

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.

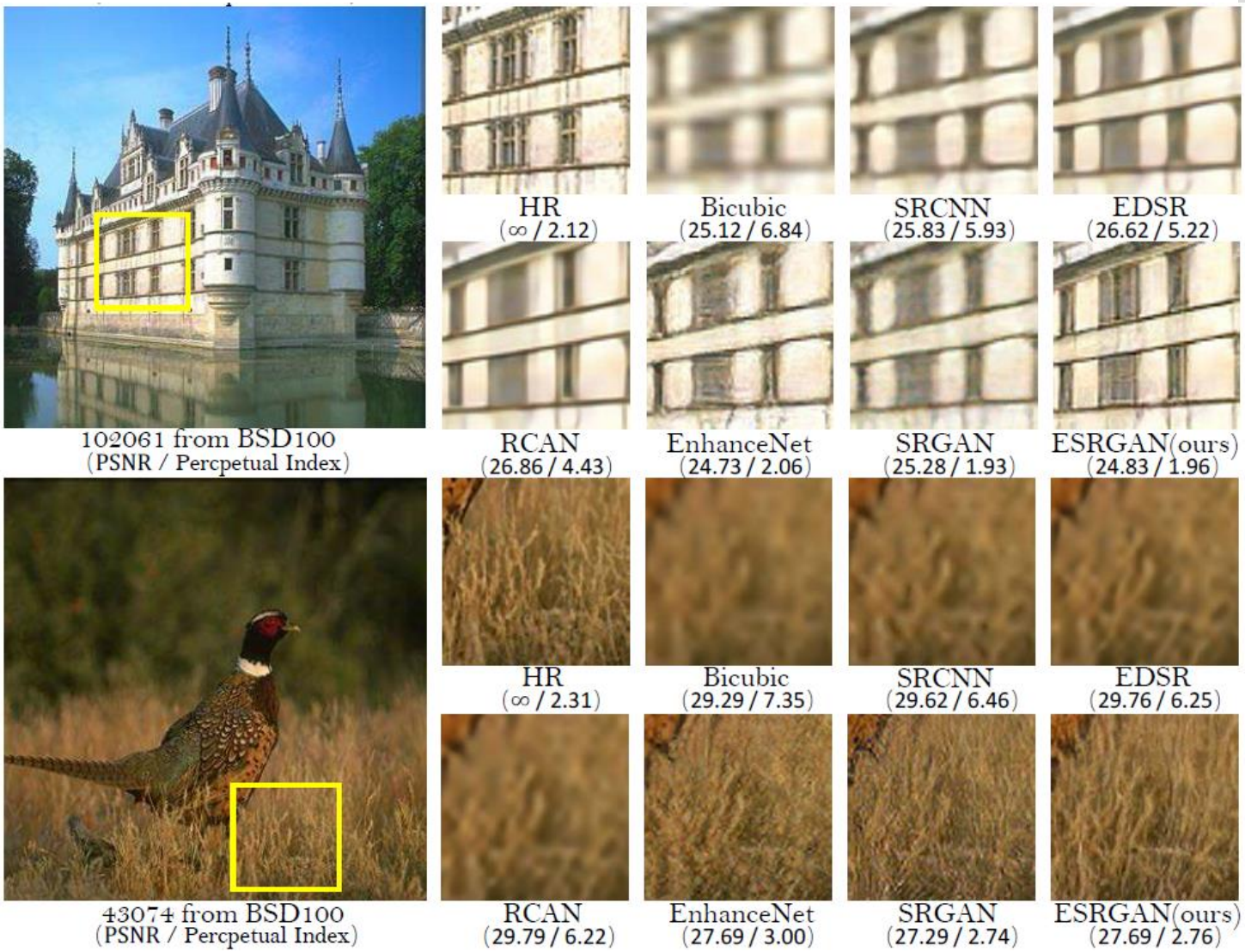
ESRGAN

Network Interpolation: To remove unpleasant noise in GAN-based methods. First train a PSNR-oriented network G_{PSNR} and then obtain a GAN-based network G_{GAN} by fine-tuning, and interpolate all the corresponding parameters of these two networks to derive an interpolated model.

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

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

ESRGAN


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)

Thank You !