

SRGAN

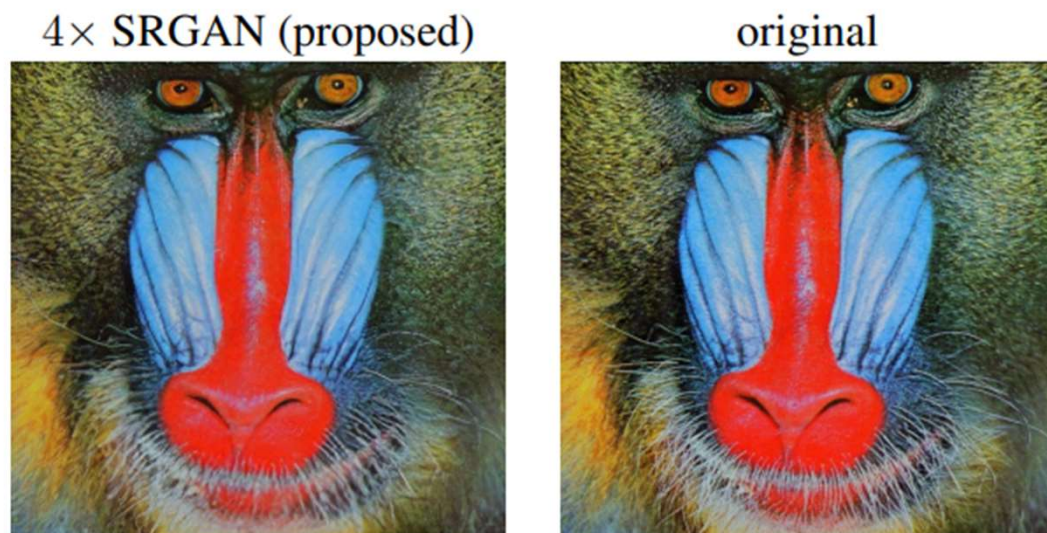
**Photo-Realistic Single Image Super-Resolution
Using a Generative Adversarial Network**

2021210088 허지혜

Abstract

- CNN을 사용하면서 SR(Super Resolution) 정확도와 속도가 빨라졌지만 Up-scale 과정에서 이미지의 디테일을 어떻게 복구할지에 관한 문제가 남아있다.
- SRGAN 모델을 제시하여 이미지의 해상도를 높이자.

Introduction



- SR(Super Resolution)은 저화질에서 고화질로 바꾸는 것을 의미하는데 고화질 이미지의 디테일이 부족하다는 문제가 있다.
- 이유로 고화질 이미지와 실제 이미지 간의 차이를 mse로 많이 최소화하는데 mse가 최소화 될 수록 psnr(최대 신호잡음비)은 최대화되기 때문이다.

Figure 1: Super-resolved image (left) is almost indistinguishable from original (right). [4× upscaling]

Introduction

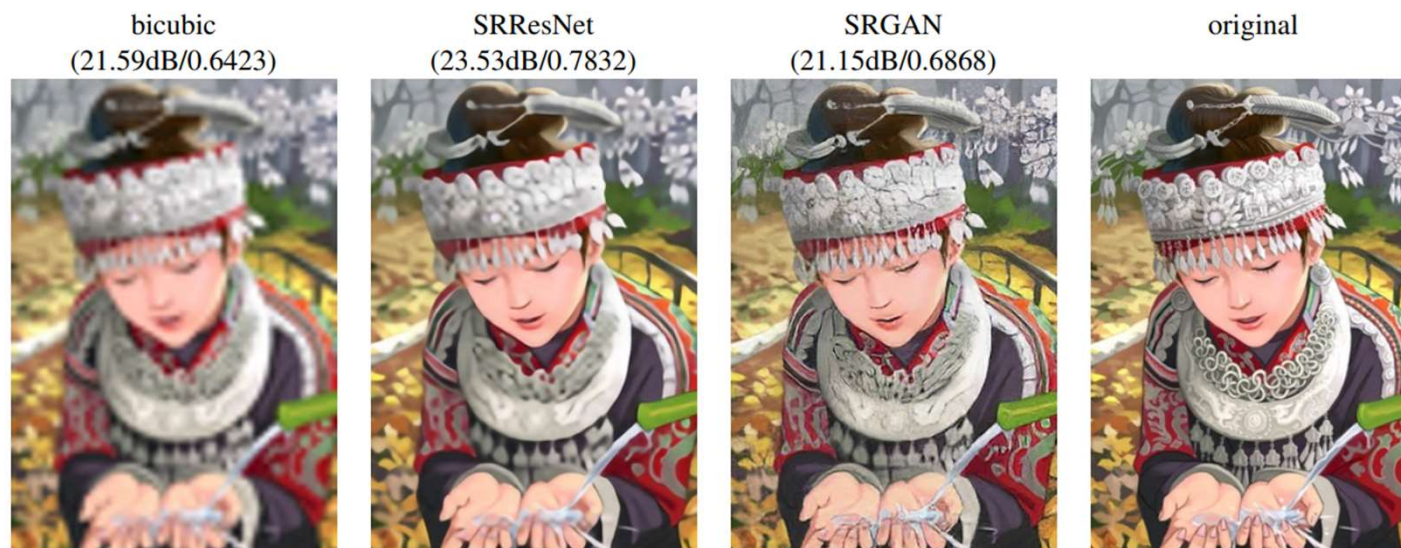


Figure 2: From left to right: bicubic interpolation, deep residual network optimized for MSE, deep residual generative adversarial network optimized for a loss more sensitive to human perception, original HR image. Corresponding PSNR and SSIM are shown in brackets. [4× upscaling]

- 위 그림처럼 psnr이 높다고 더 나은 결과를 보장할 순 없다.
- 위 논문에서는 mse 대신 지각적인 유사성에 중점을 둔 loss 함수를 제시한다.

Method - SRGAN

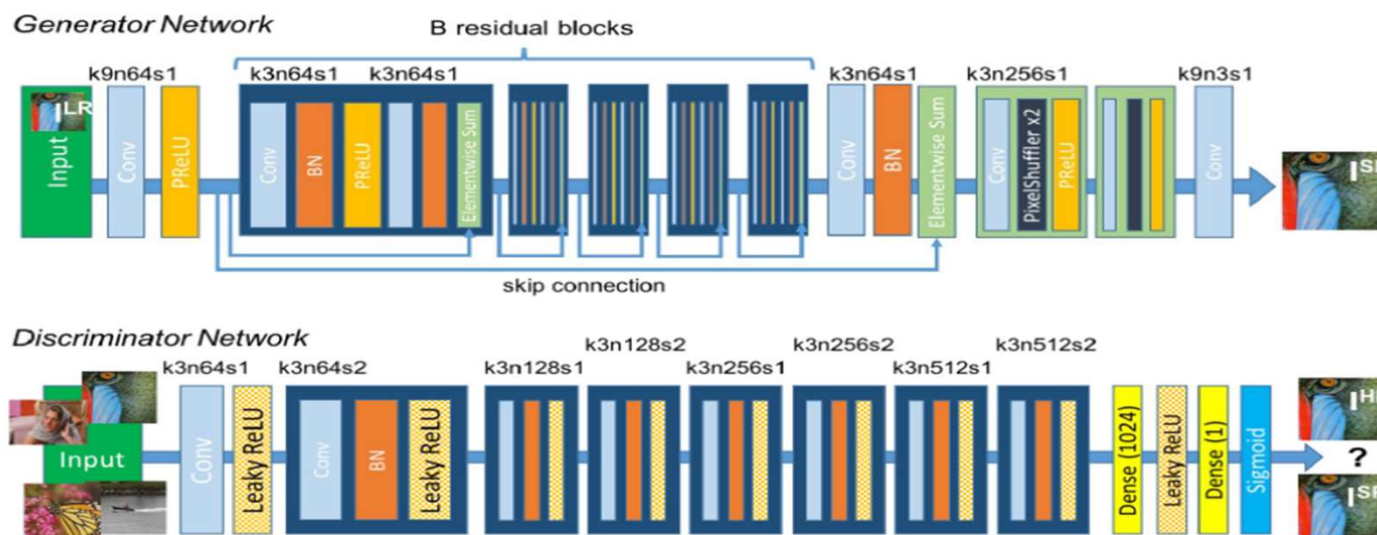


Figure 4: Architecture of Generator and Discriminator Network with corresponding kernel size (k), number of feature maps (n) and stride (s) indicated for each convolutional layer.

- Generator의 Activation 함수로 PReLU 함수를 사용한다.
- SRGAN에서는 SRCNN과 다르게 Generator 후반에 PixelShuffler를 사용한다고 한다.
- Discriminator에서는 maxpooling 말고 stride를 이용하여 크기를 줄인다.

Method – SRGAN Loss

Perceptual loss = content loss + adversarial loss
content loss

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

adversarial loss

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

- 일반적으로 MSE에 기반하여 설계되지만 앞서 얘기했듯 그런 경우에는 PSNR이 낮게 나오거나 사람 눈에 이상한 경우로 출력이 될 수 있다.
- 논문에서는 이를 방지하기 위해 두 가지 loss를 섞은 perceptual loss를 이용하여 계산하였다.

Experiments

- 모든 실험은 저화질과 고화질 사이의 4배의 scale-factor로 진행된다. 이는 pixel의 크기를 16배 감소한다는 의미이다.
- 이를 정량화 하려고 MOS TEST를 진행하였다.

- SRGAN-MSE : 표준 MSE를 손실로 지정
- SRGAN-VGG22 : 하위 레벨의 특징을 나타내는 Feature map에 정의된 손실
- SRGAN-VGG54 : 네트워크의 깊은 layer에서 더 높은 수준의 feature map에 정의된 손실

SRGAN-MSE를 쓸 때 가장 좋은 PSNR을 얻는다.

Table 1: Performance of different loss functions for SR-ResNet and the adversarial networks on Set5 and Set14 benchmark data. MOS score significantly higher ($p < 0.05$) than with other losses in that category*. [4× upscaling]

Set5	SRResNet-		SRGAN-		
	MSE	VGG22	MSE	VGG22	VGG54
PSNR	32.05	30.51	30.64	29.84	29.40
SSIM	0.9019	0.8803	0.8701	0.8468	0.8472
MOS	3.37	3.46	3.77	3.78	3.58
Set14					
PSNR	28.49	27.19	26.92	26.44	26.02
SSIM	0.8184	0.7807	0.7611	0.7518	0.7397
MOS	2.98	3.15*	3.43	3.57	3.72*

Experiments

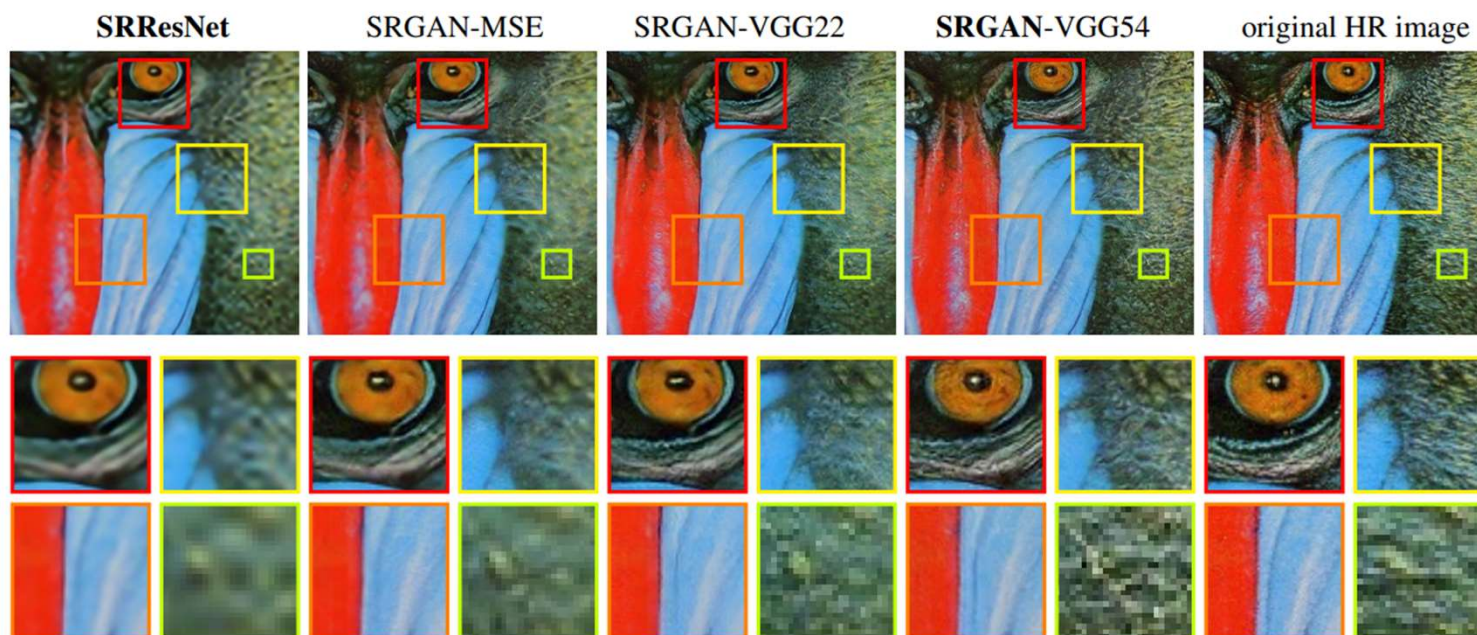
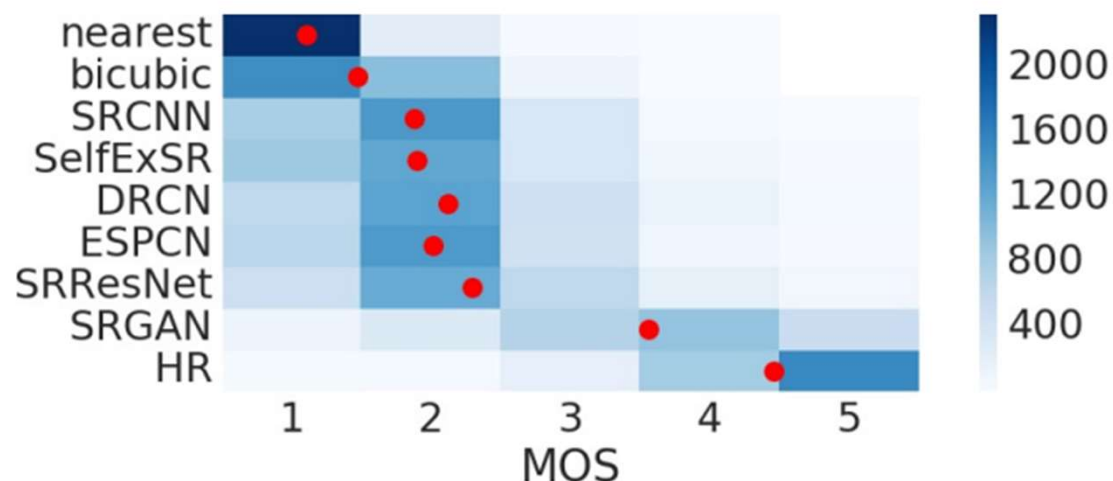


Figure 6: **SRResNet** (left: a,b), **SRGAN-MSE** (middle left: c,d), **SRGAN-VGG2.2** (middle: e,f) and **SRGAN-VGG54** (middle right: g,h) reconstruction results and corresponding reference HR image (right: i,j). [4× upscaling]

- 이미지에서 디테일 부분은 SRGAN-VGG54가 잘 표현하는 것을 확인할 수 있다.

Discussion and future work



- SRGAN이 MOS TEST에서 우수한 성능을 보임을 확인할 수 있다.
- 네트워크 구조가 깊을수록 좋은 성능을 보이지만 High-frequency artifact(고주파 인공물) 문제로 학습하기 힘들다.

Reference

- [1] Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network : <https://arxiv.org/abs/1609.04802>
- [2] <https://velog.io/@kanghoon12/%EB%85%BC%EB%AC%B8%EB%A6%AC%EB%B7%B0-SRGANPhoto-Realistic-Single-Image-Super-Resolution-Using-a-Generative-Adversarial-Network>

감사합니다.