

---

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

[CVPR, 2017]

---

2023. 04. 03.

김성수

Data Mining and Quality Analytics

# 연구배경

- Task명

## ❖ Task: Super Resolution (SR)

- 저화질 이미지가 입력으로 돌아오면 고화질 이미지로 개선하는 것
- Input: 저화질 이미지(LR) // Output: 고화질 이미지(HR)
  - ✓ 이때, 저화질 이미지는 고화질 이미지를 Down sampling하여 구축



[Input: 저화질 이미지]

  
**Super Resolution**



[Output: 고화질 이미지]

# 연구배경

- Task명

## ❖ Task: Super Resolution (SR)

- 저화질 이미지가 입력으로 돌아오면 고화질 이미지로 개선하는 것
- Input: 저화질 이미지(LR) // Output: 고화질 이미지(HR)
  - ✓ 이때, 저화질 이미지는 고화질 이미지를 Down sampling하여 구축
  - ✓ 크기를 작게 하는 것이 왜 저화질? → 이미지를 다시 크게하면 화질이 안 좋아진 것을 볼 수 있음



[Input: 저화질 이미지]

➡  
**Super Resolution**



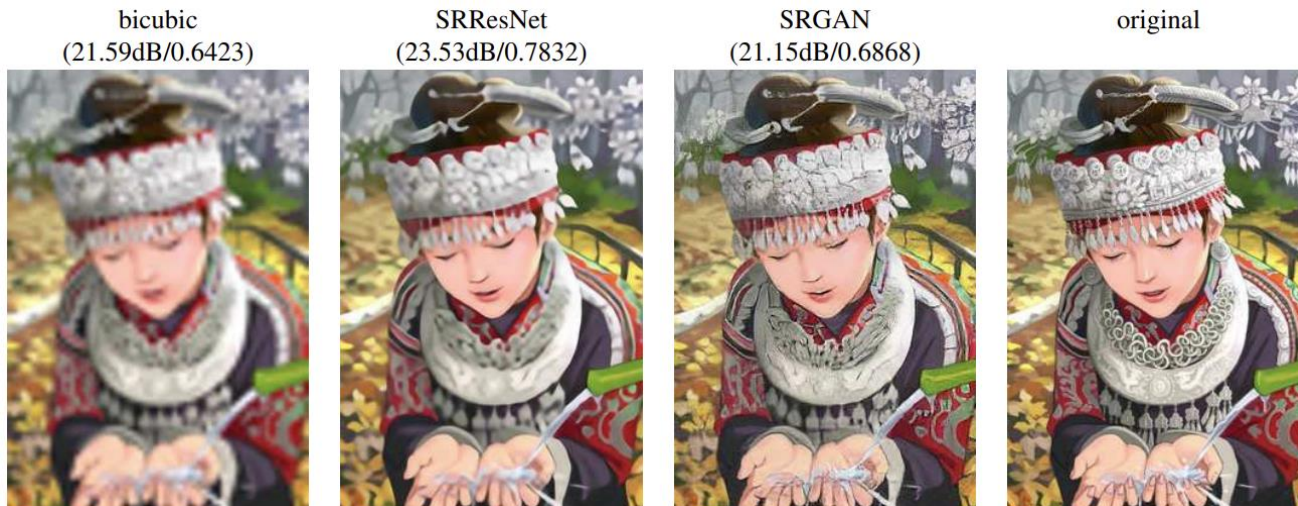
[Output: 고화질 이미지]

$$\text{PSNR} = 10 \log \frac{\text{Max(pixel values)} \cdot S^2}{\text{MSE}}$$

## ❖ Limitation of Previous Research

- 선행연구의 SR은 미세한 특징까지는 잘 복원하지 못함
- MSE를 최소화로 학습하는 것은 이미지의 세부적인 부분을 잘 복원하지 못함 (Over-smooth)
  - ✓ MSE로 학습 시 SR의 척도 중 하나인 PSNR 수치를 좋게 만들 수 있음 (두 식의 관련성 존재)
  - ✓ 그러나 실제 정성적 평가에서는 무조건 PSNR이 높다고 이미지가 고화질은 아님

\* MSE를 최소화: 복원된 이미지와 입력 값인 LR이미지의 픽셀 차이를 최소화



[Input: 저화질 이미지]

[Output: 고화질 이미지]

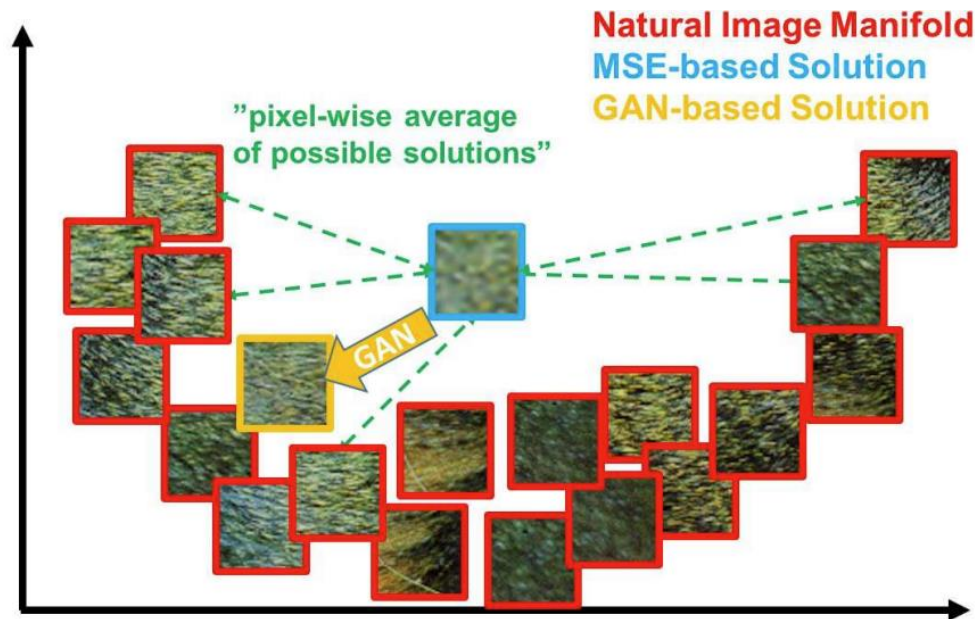
# 연구배경

- 선행연구의 한계

## ❖ Limitation of Previous Research

- 선행연구의 SR은 미세한 특징까지는 잘 복원하지 못함
- MSE를 최소화로 학습하는 것은 이미지의 세부적인 부분을 잘 복원하지 못함 (Over-smooth)
  - ✓ MSE로 학습 시 SR의 척도 중 하나인 PSNR 수치를 좋게 만들 수 있음 (두 식의 관련성 존재)
  - ✓ 그러나 실제 정성적 평가에서는 무조건 PSNR이 높다고 이미지가 고화질은 아님

\* MSE를 최소화: 복원된 이미지와 입력 값인 LR이미지의 픽셀 차이를 최소화



## ❖ Overcome the Limitation

- MSE 기반이 아닌, GAN Loss를 기반으로 학습
- Adversarial Loss와 Content Loss의 가중합 Loss를 새롭게 제안
  - ✓ 이때, Content Loss에는 Pretrained VGG를 활용

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

**[기존 연구의 MSE-based Loss]**

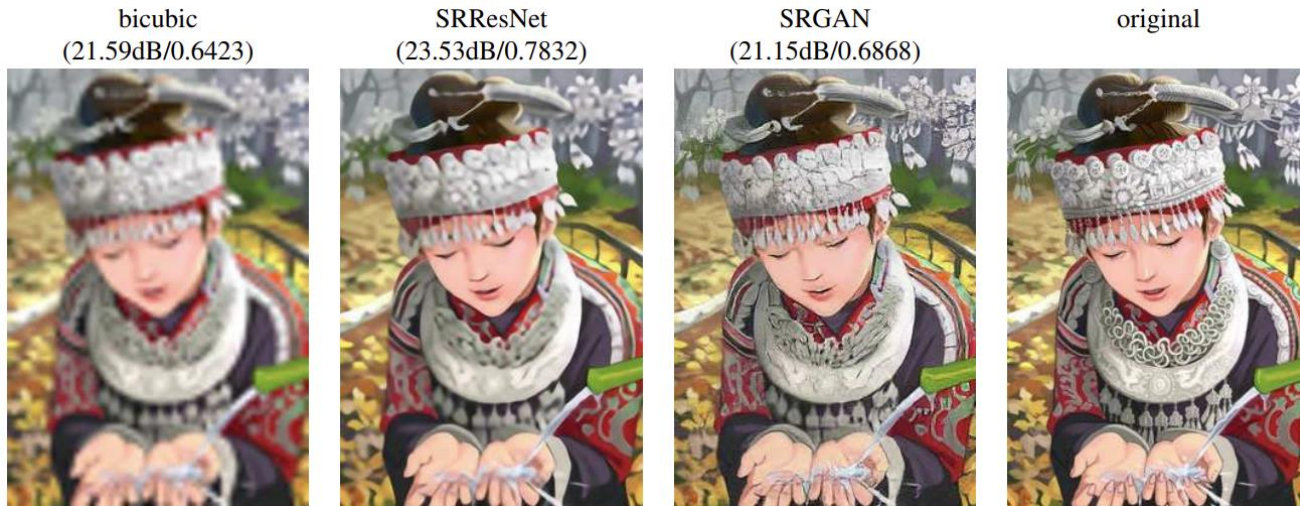
$$l^{SR} = \underbrace{l_X^{SR}}_{\text{content loss}} + \underbrace{10^{-3} l_{Gen}^{SR}}_{\text{adversarial loss}}$$

**[본 연구의 Perceptual Loss]**



## ❖ Contribution

- 해상도를 4배 수준으로 크게 했을 때도 좋은 성능을 보임
  - ✓ 이는 높이/넓이를 각각 4배 씩 증가하는 것으로, 픽셀수가 16배 많아지는 수치
- PSNR의 한계 지적: **“무조건 PSNR이 높다고 좋은 이미지는 아니다.”**
  - ✓ 대안으로 MOS Test를 제안했으며, 본 방법론은 MOS Test에서 가장 좋은 성능을 보임
  - ✓ MOS Test(Mean Opinion Test): 인간이 직접 이미지의 해상도에 대해 점수를 책정

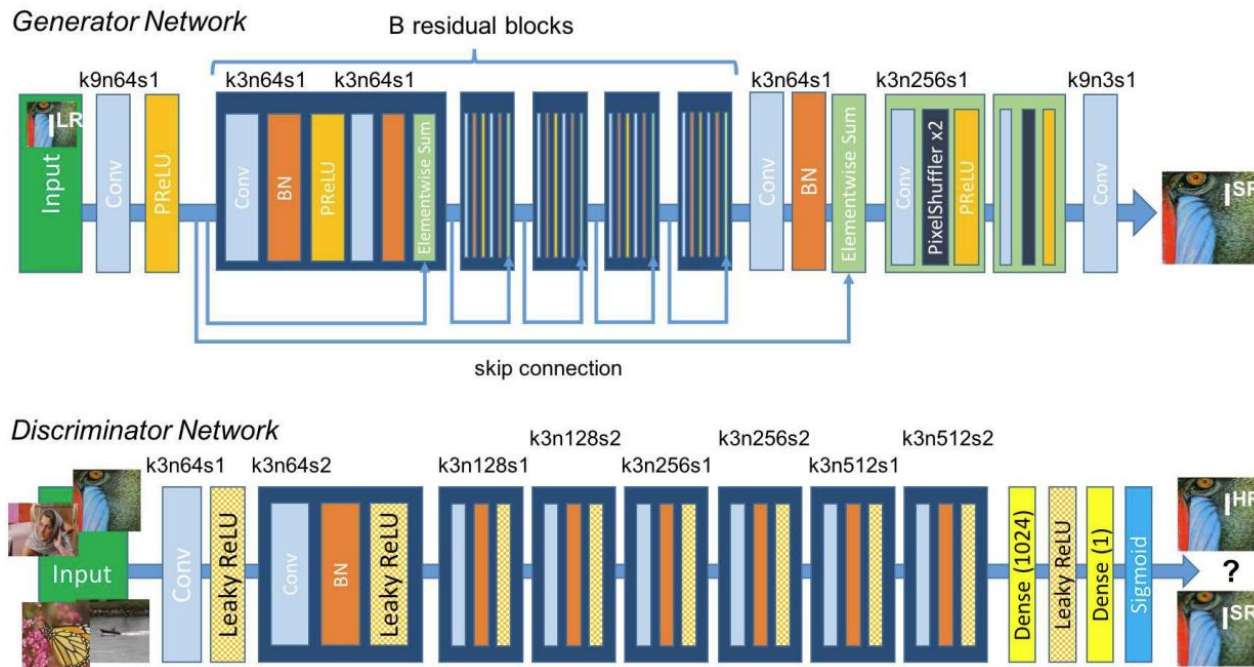


[Input: 저화질 이미지]

[Output: 고화질 이미지]

## ❖ 방법론 개요

- Adversarial Loss와 Content Loss의 가중합 Loss로 학습



[Model Architecture]



## ❖ 방법론 개요

- Adversarial Loss와 Content Loss의 가중합 Loss로 학습

$$\hat{\theta}_G = \arg \min_{\theta_G} \frac{1}{N} \sum_{n=1}^N \underbrace{l^{SR}}_{\text{복원된 이미지}}(\underbrace{G_{\theta_G}(I_n^{LR})}_{\text{기존 고해상도 이미지}}, \underbrace{I_n^{HR}}_{\text{기존 고해상도 이미지}})$$

## [최종 파라미터 Selection]

$$l^{SR} = \underbrace{l_X^{SR}}_{\text{content loss}} + \underbrace{10^{-3} l_{Gen}^{SR}}_{\text{adversarial loss}}$$

perceptual loss (for VGG based content losses)

## [최종 Loss]

$$l^{SR} = \underbrace{l_X^{SR}}_{\text{content loss}} + \underbrace{10^{-3} l_{Gen}^{SR}}_{\text{adversarial loss}}$$

perceptual loss (for VGG based content losses)

## ❖ Adversarial Loss

- Generator에서 복원된 이미지와 실제 고해상도 이미지를 구분하지 못하도록 학습

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

[GAN의 아이디어]

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

[본 연구의 Adversarial Loss]

\*\*\*궁금: 왜 기존 고해상도 이미지는 Input으로 안 넣지? 안 넣으면 GAN학습에 의미가 있나?\*\*\*

$$l^{SR} = \underbrace{l_X^{SR}}_{\text{content loss}} + \underbrace{10^{-3}l_{Gen}^{SR}}_{\text{adversarial loss}}$$

perceptual loss (for VGG based content losses)

## ❖ Content Loss

- 일반적인 MSE Loss는 Over-smooth되는 경향이 존재
- Content Loss: Pretrained-VGG19의 중간 Feature Map에 대해 MSE 비교

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

[MSE Loss]

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

[Content Loss]

$\phi_{i,j}$

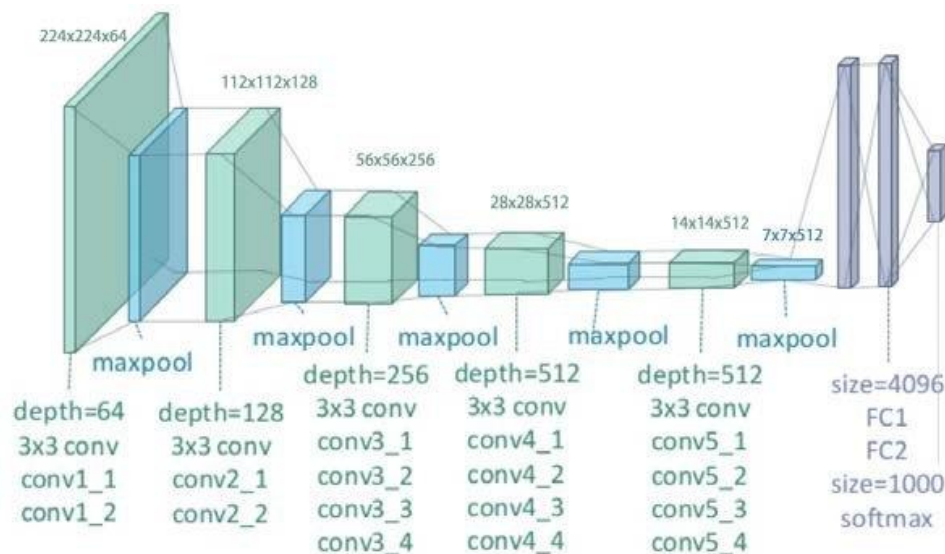
i번째 max pooling 이전의  
j번째 convolution layer

$$l^{SR} = \underbrace{l_X^{SR}}_{\text{content loss}} + \underbrace{10^{-3} l_{Gen}^{SR}}_{\text{adversarial loss}}$$

perceptual loss (for VGG based content losses)

## ❖ Content Loss

- 일반적인 MSE Loss는 Over-smooth되는 경향이 존재
- Content Loss: Pretrained-VGG19의 중간 Feature Map에 대해 MSE 비교



[VGG19 Architecture]

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

[Content Loss]

$\phi_{i,j}$

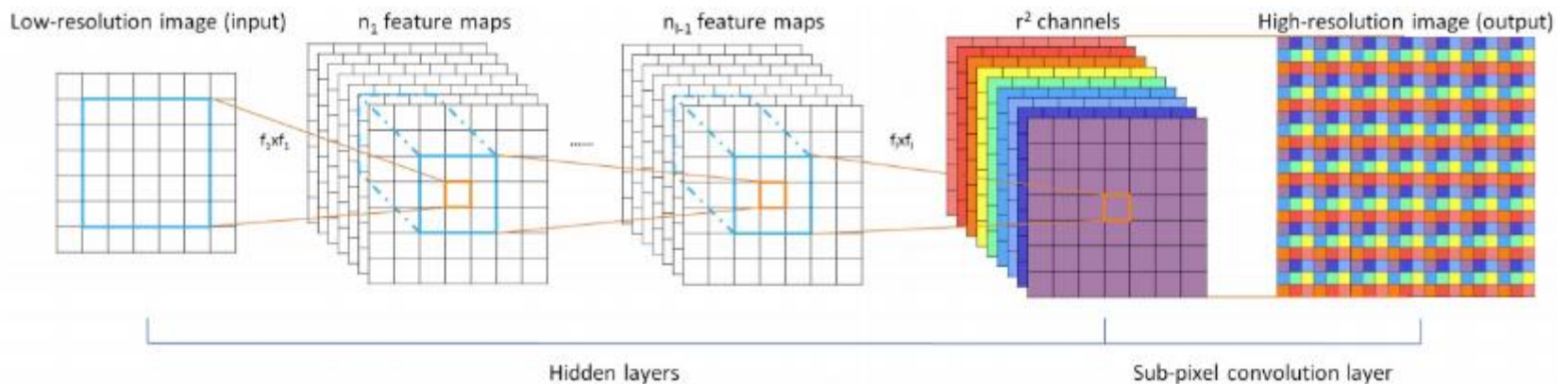
i번째 max pooling 이전의  
j번째 convolution layer

# APPENDIX

- Pixel Shuffle (Sub Pixel Convolution)

## ❖ 어떻게 이미지의 크기가 커지지?

- 궁금증: CNN은 분명 이미지를 작게 만들거나, 최소한 같은 크기로 만드는데, 어떻게 해당 네트워크에서는 이미지가 커질까?
  - ✓ Pixel Shuffle (Sub Pixel Convolution)을 활용
- $R^2$  Channels,  $7 \times 7$  Image  $\rightarrow$  1 Channel  $7R \times 7R$  Image
  - ✓ 무조건 Output Channel이 1개? No.  $(16, 4, 4) \rightarrow (4, 8, 8)$ 도 가능



[Pixel Shuffle (Sub Pixel Convolution)]