Real-World Super Resolution
CVPR 2020 – Ntire 2020 challenge

2020 VDSL 연구실 하계 세미나

서 유 림

Vision & Display Systems Lab.

Dept. of Electronic Engineering, Sogang University

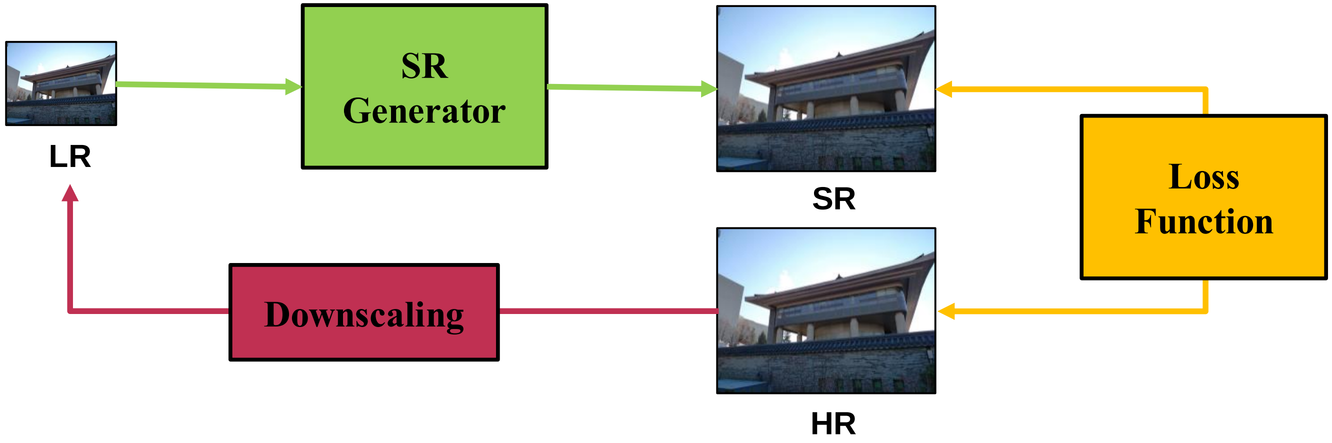
Outline

Motivation

- Super Resolution (SR) 이란?
  - Low Resolution (LR) 영상을 High Resolution (HR) 영상으로 변환하는 기술
  - ill-posed problem
    - Multiple solutions for LR input
  - 딥러닝 기반의 학습 방법을 통해 비약적인 성능 향상
    - SRCNN[1], VDSR[2], SRResNet 과 SRGAN[3], EDSR [4], RCAN [5], ESRGAN [6]…

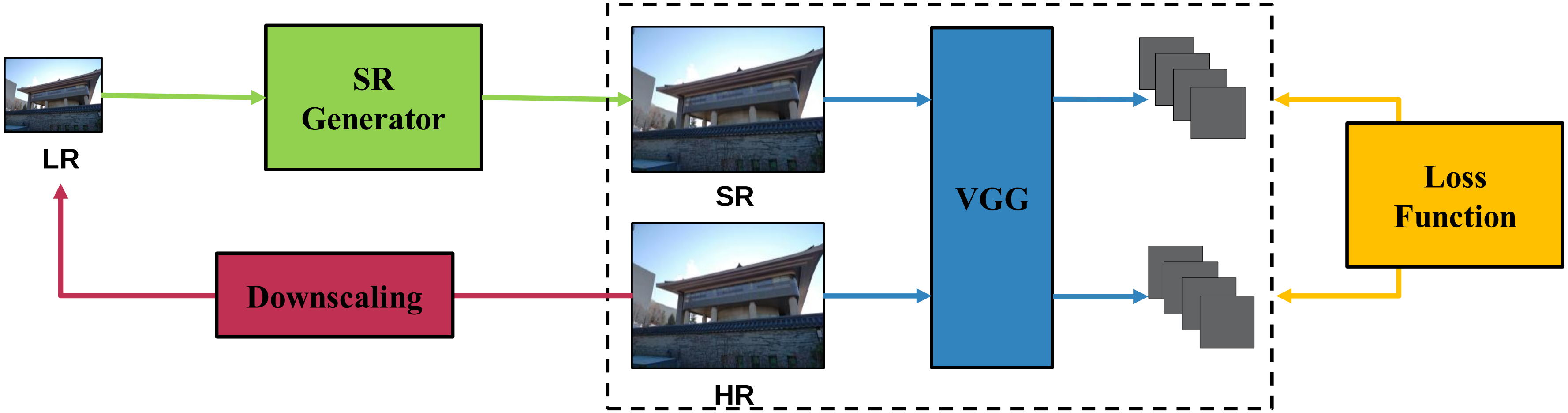
Motivation

- 학습 방법에 따른 분류
  - PSNR-oriented SR
    - 대표적인 영상 평가 지표인 PSNR (Peak Signal-to-Noise Ratio) 수치를 높이기 위한 학습
    - Pixel-wise Loss Function (ex) L1, L2 function) 를 사용해 SR 출력 영상과 HR 영상 사이의 차이를 줄이도록 네트워크 학습
    - SRCNN, VDSR, EDSR, RCAN…
    - 장점 : 높은 PSNR 수치, 낮은 구조적 왜곡
    - 단점 : blur artifact 및 고주파수 성분 복원 어려움



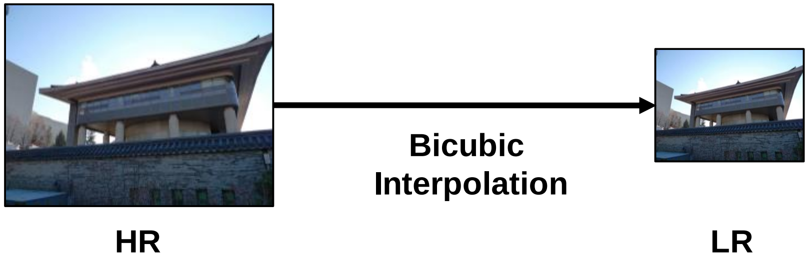
Motivation

- 학습 방법에 따른 분류
  - Perceptual-Oriented SR
    - PSNR-Oriented SR의 단점을 해결하기 위해 Peceptual loss [7] 과 GAN 사용
      - Perceptual loss : SR 영상과 HR 영상을 pre-trained 된 VGG 네트워크에 입력 후 출력 사이의 차이를 줄이도록 네트워크 학습
    - SRGAN, ESRGAN, EnhanceNet…
    - 장점 : 높은 인지적 성능, 낮은 blur artifact
    - 단점 : 구조적 왜곡 및 noise 발생

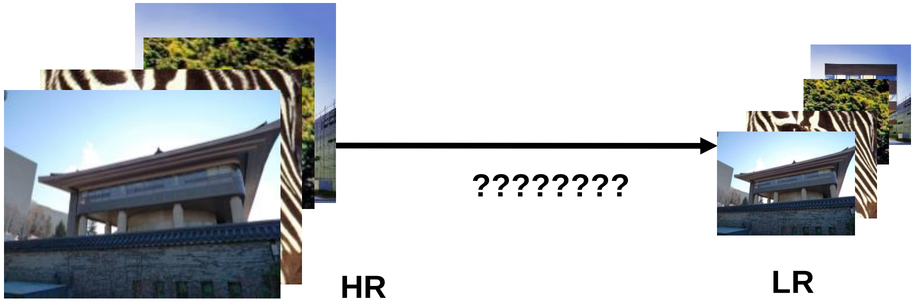


Motivation

- Real-World에서의 적용
  - 일반적으로 학습 데이터 셋을 만드는 방법
    - Bicubic interpolation 을 downscaling 시 사용



    - 그러나 실제 Real-World에서의 영상은 HR 영상에서 Bicubic interpolation 으로 downscaling 된 영상이 아님



Motivation

- Real-World에서의 적용 (scaling factor x4)
  - 일반적인 학습데이터셋으로 Pre-Trained RCAN에 적용
    - HR 영상을 downscaling하는 방법에 따른 출력 영상의 quality 차이
    - 즉 기존의 학습 방법으로는 임의의 Real-World 영상에 높은 성능 불가능

HR

Nearest

Bilinear

Bicubic

Lanczos

PNSR : 20.2726

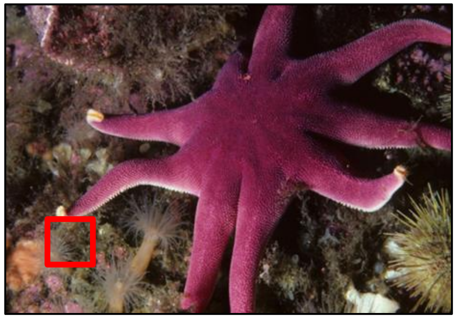
PNSR : 24.1343

PNSR : 24.2644

PNSR : 24.2668

Motivation

- Real-World에서의 적용 (scaling factor x4)
  - 일반적인 학습데이터셋으로 Pre-Trained RCAN에 적용
    - HR 영상을 downscaling하는 방법에 따른 출력 영상의 quality 차이
    - 즉 기존의 학습 방법으로는 임의의 Real-World 영상에 높은 성능 불가능



HR

Nearest

Bilinear

Bicubic

Lanczos

PNSR : 25.6983

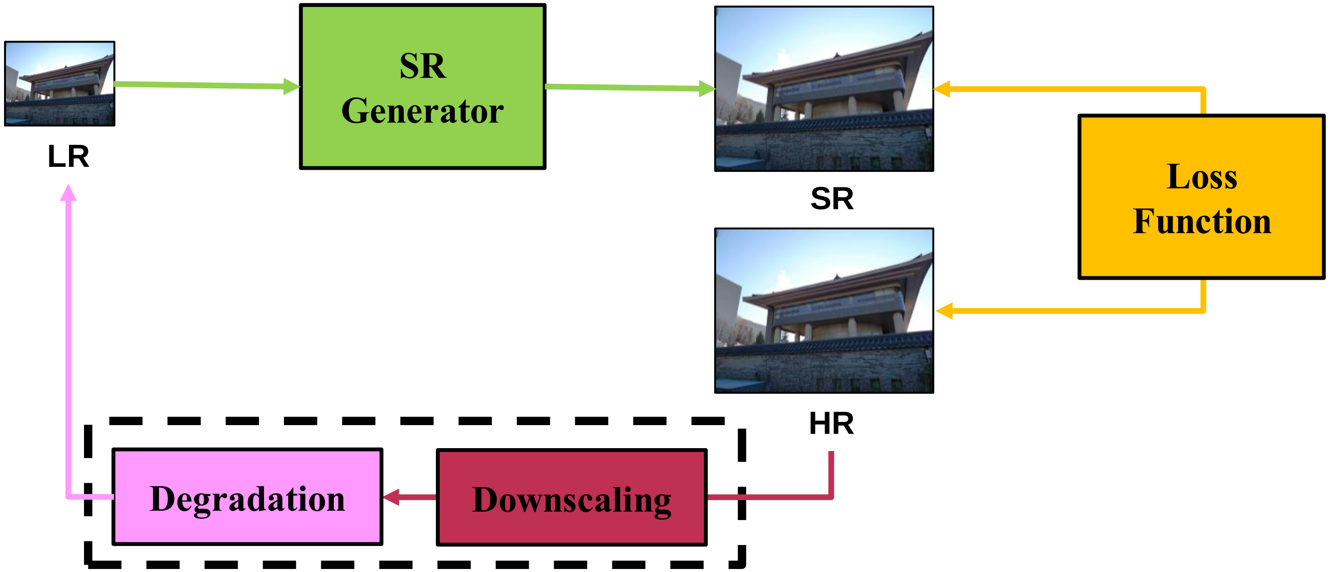
PNSR : 28.7693

PNSR : 29.5592

PNSR : 29.5267

Motivation

- Real-World에서의 적용이 잘되려면?
  - Real-World 영상에 최대한 대응되는 학습데이터셋이 필요 (Blind-SR)
    - HR 영상에서 Bicubic interpolation 기반의 downsampling을 통해 LR영상을 생성할 경우 임의의 입력 LR 영상에 대해 네트워크가 적절히 이를 반영하지 못하는 문제가 발생
    - LR-HR pair가 없을 때 (weakly or unsupervised learning) 대응할 수 있는 학습 방법
    - 다양한 Downsamping & Degradation을 반영할 수 있는 네트워크



Motivation

- Ntire 2020 image challenges
  - Real World Super-Resolution
    - Samsung-SLSL-MSL → Proposed Method 1
      - Track 1 → 2등 Track 2 → 평가 제외
    - Impressionism → Proposed Method 2
      - Track 1, 2 → 1등

Team	PSNR↑	SSIM↑	LPIPS↓	MOS↓
Impressionism	$24.67_{(16)}$	$0.683_{(13)}$	$0.232_{(1)}$	$2.195_{(1)}$
Samsung-SLSI-MSL	$25.59_{(12)}$	$0.727_{(9)}$	$0.252_{(2)}$	$2.425_{(2)}$
BOE-IOT-AIBD	$26.71_{(4)}$	$0.761_{(4)}$	$0.280_{(4)}$	$2.495_{(3)}$
MSMers	$23.20_{(18)}$	$0.651_{(17)}$	$0.272_{(3)}$	$2.530_{(4)}$
KU-ISPL	$26.23_{(6)}$	$0.747_{(7)}$	$0.327_{(8)}$	$2.695_{(5)}$
InnoPeak-SR	$26.54_{(5)}$	$0.746_{(8)}$	$0.302_{(5)}$	$2.740_{(6)}$
ITS425	$27.08_{(2)}$	$0.779_{(1)}$	$0.325_{(6)}$	$2.770_{(7)}$

Team	NIQE↓	BRISQUE↓	PIQE↓	NRQM↑	PI↓	IQA-Rank↓	MOR↓
Impressionism	$5.00_{(1)}$	$24.4_{(1)}$	$17.6_{(2)}$	$6.50_{(1)}$	$4.25_{(1)}$	3.958	$1.54_{(1)}$
AITA-Noah-A	$5.63_{(4)}$	$33.8_{(5)}$	$29.7_{(8)}$	$4.23_{(8)}$	$5.70_{(6)}$	7.720	$3.04_{(2)}$
ITS425	$8.95_{(18)}$	$52.5_{(18)}$	$88.6_{(18)}$	$3.08_{(18)}$	$7.94_{(18)}$	14.984	$3.30_{(3)}$
AITA-Noah-B	$8.18_{(17)}$	$50.1_{(12)}$	$88.0_{(17)}$	$3.23_{(15)}$	$7.47_{(17)}$	13.386	$3.57_{(4)}$
Webbzhou	$7.88_{(15)}$	$51.1_{(15)}$	$87.8_{(16)}$	$3.27_{(14)}$	$7.30_{(15)}$	12.612	$4.44_{(5)}$
Samsung-SLSI-MSL	$6.25_{(7)}$	$37.3_{(6)}$	$26.0_{(5)}$	$4.31_{(7)}$	$5.97_{(7)}$	6.662	-

NTIRE 2020 image challenges

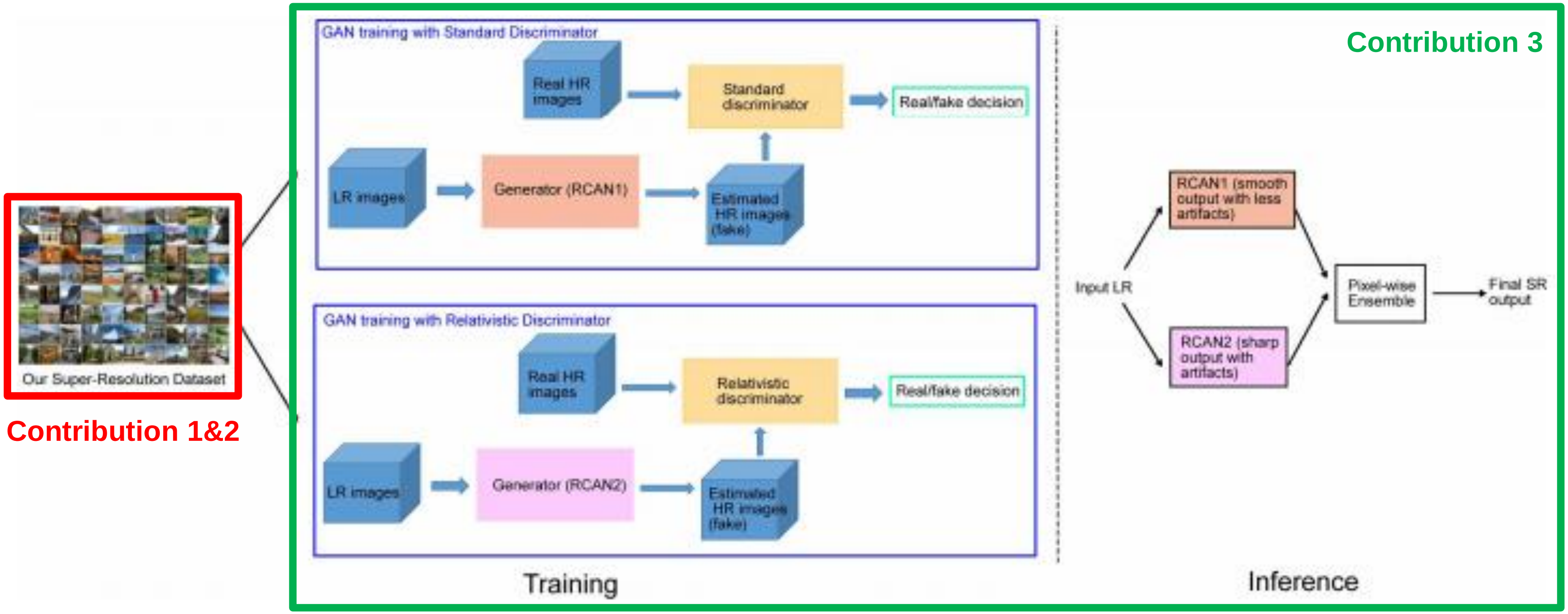
- Perceptual Extreme Super-Resolution
- Real World Super-Resolution - Track 1 Image Processing Artifacts
- Real World Super-Resolution - Track 2 Smartphone Images
- Real Image Denoising - Track 1 rawRGB
- Real Image Denoising - Track 2 sRGB
- Image Deblurring - Track 1: on desktop
- Image Deblurring - Track 2: on smartphone
- Demoireing - Track 1 Single image
- Demoireing - Track 2 Burst
- Spectral Reconstruction from an RGB Image - Track 1 Clean
- Spectral Reconstruction from an RGB Image - Track 2 Real World
- NonHomogeneous Dehazing

Proposed Method 1

- Real-World Super-Resolution using Generative Adversarial Networks
  - Contribution 1
    - Our generic SR model trained on the SR dataset generated by multiple degradations generalizes well on the images with unknown degradation caused by image processing artifacts
  - Contribution 2
    - We design a mobile SR dataset based on registered mobile-DSLR image pairs at same scale, where the DSLR images are super resolved with our generic SR model. Fine-tuning our SR model on this dataset improves the perceptual quality on mobile images.
  - Contribution 3
    - Our GAN-based fusion is capable of improving the perceptual quality and reducing the artifacts of the estimated HR images.

Proposed Method 1

- Network Architecture
  - Contribution 1, 2 → 학습 데이터셋 만드는 방법
  - Contribution 3 → 네트워크 설계
    - SRGAN, ESRGAN의 Generator 가 아닌 RCAN을 SR generator 로 사용



Proposed Method 1

- Contribution 1
  - 다양한 degradation이 랜덤하게 적용된 학습 데이터셋을 생성

$$x = N(D(y * k))$$

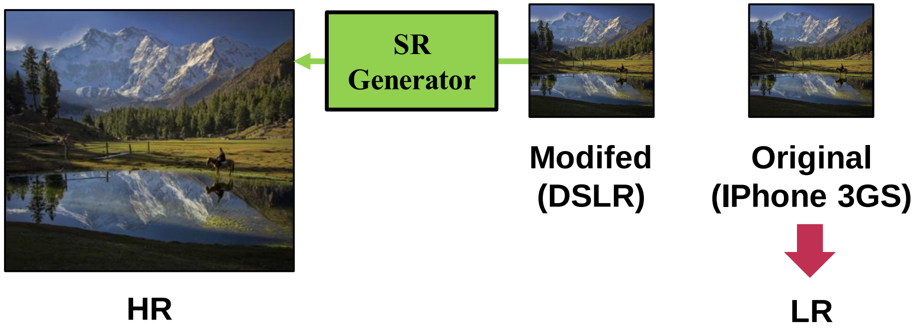
LR 영상 **Noise** **Down scaling** **HR 영상** **Blur Kernel**

- Challenge 에서 제공하는 Source Domain 영상 (image processing artifacts) 및 Target Domain 영상 (Clean High Quality) 을 HR 영상으로 설정하고 위의 식 적용하여 LR 영상 생성
- Downscaling : Nearest Neighbor, Bilinear, Bicubic and Lanczos
- Blur Kernel : Gaussian blur kernel in range [0.2, 3] and kernel size 15 x 15
- Noise : Gaussian noise in range [0,5] or Poisson in range [50, 150] or Gaussian-Poisson

Proposed Method 1

- Contribution 2
  - Reference 영상이 없는 real-world 영상의 학습을 위한 pair 데이터셋을 만들기 위해 pre-trained 된 SR generator를 사용 후 fine-tuning

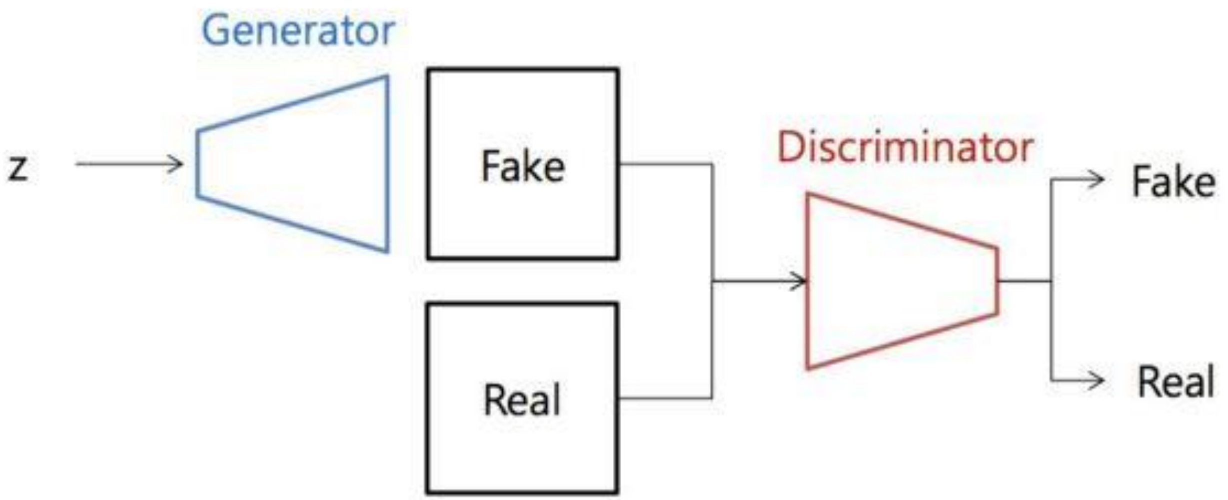
Track 2 제공



- 저자는 Track 2 에서 제공한 IPhone 3GS로 촬영된 Original data를 LR 영상으로 사용
- 좌측의 Modified (DSLR Enhancement [8] 적용) 영상을 pre-trained 된 SR generator를 통과하여 upscaling 된 영상을 생성 후 해당 영상을 HR 영상으로 학습에 사용

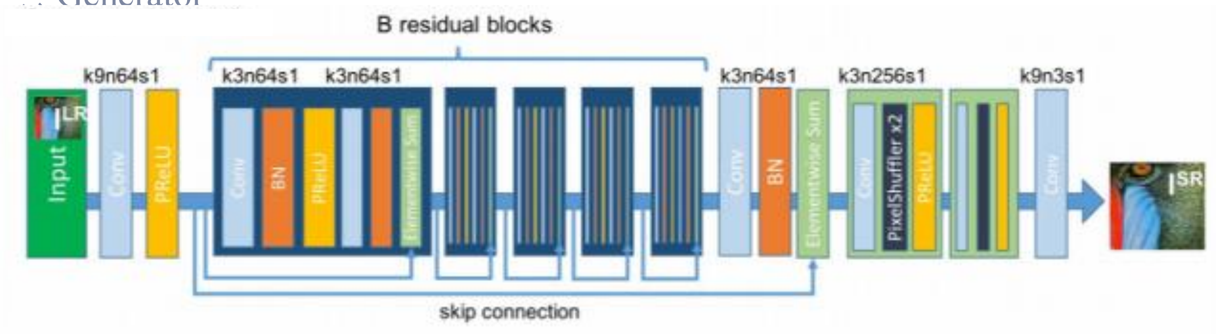
Proposed Method 1

- Contribution 3
  - Related Work
    - GAN (Super Resolution Generative Adversarial Networks)
      - SR 영상을 출력하는 Generator 와 real인지 fake인지 구분하는 Discriminator로 구성
      - Real = HR 영상, z = LR 영상
      - Generator 출력 (Fake) = SR 출력 영상

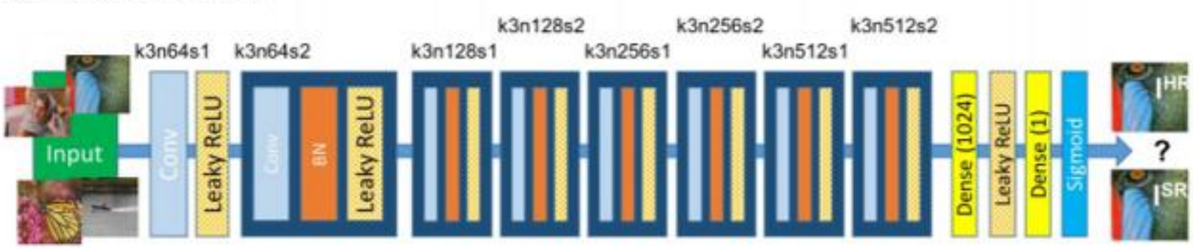


Proposed Method 1

- Contribution 3
  - Related Work
    - SRGAN (Super Resolution using Generative Adversarial Networks)
      - Generator

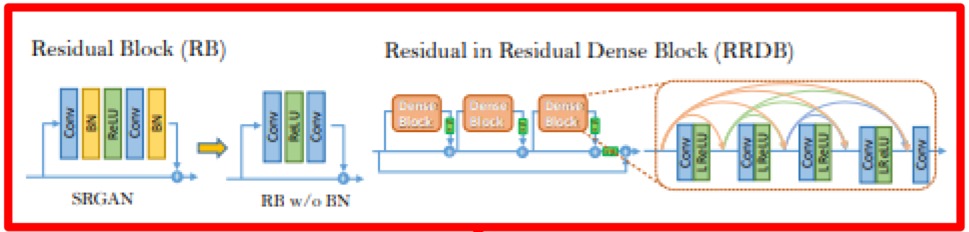


      - Discriminator

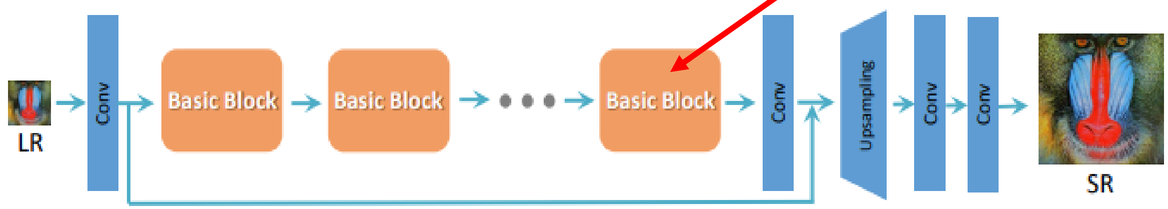


Proposed Method 1

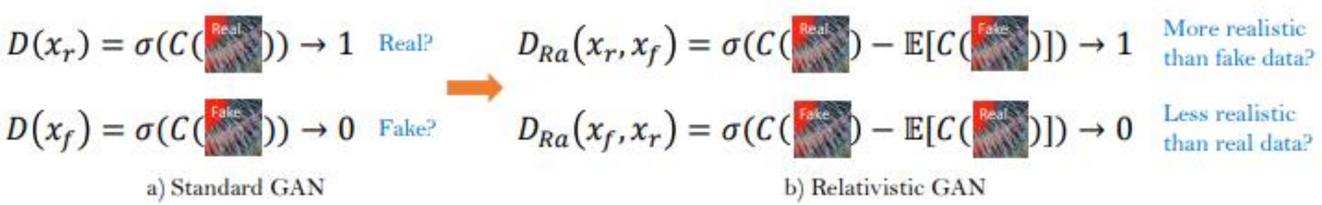
- Contribution 3
  - Related Work
    - ESRGAN (Enhanced Super Resolution using Generative Adversarial Networks)



- Generator

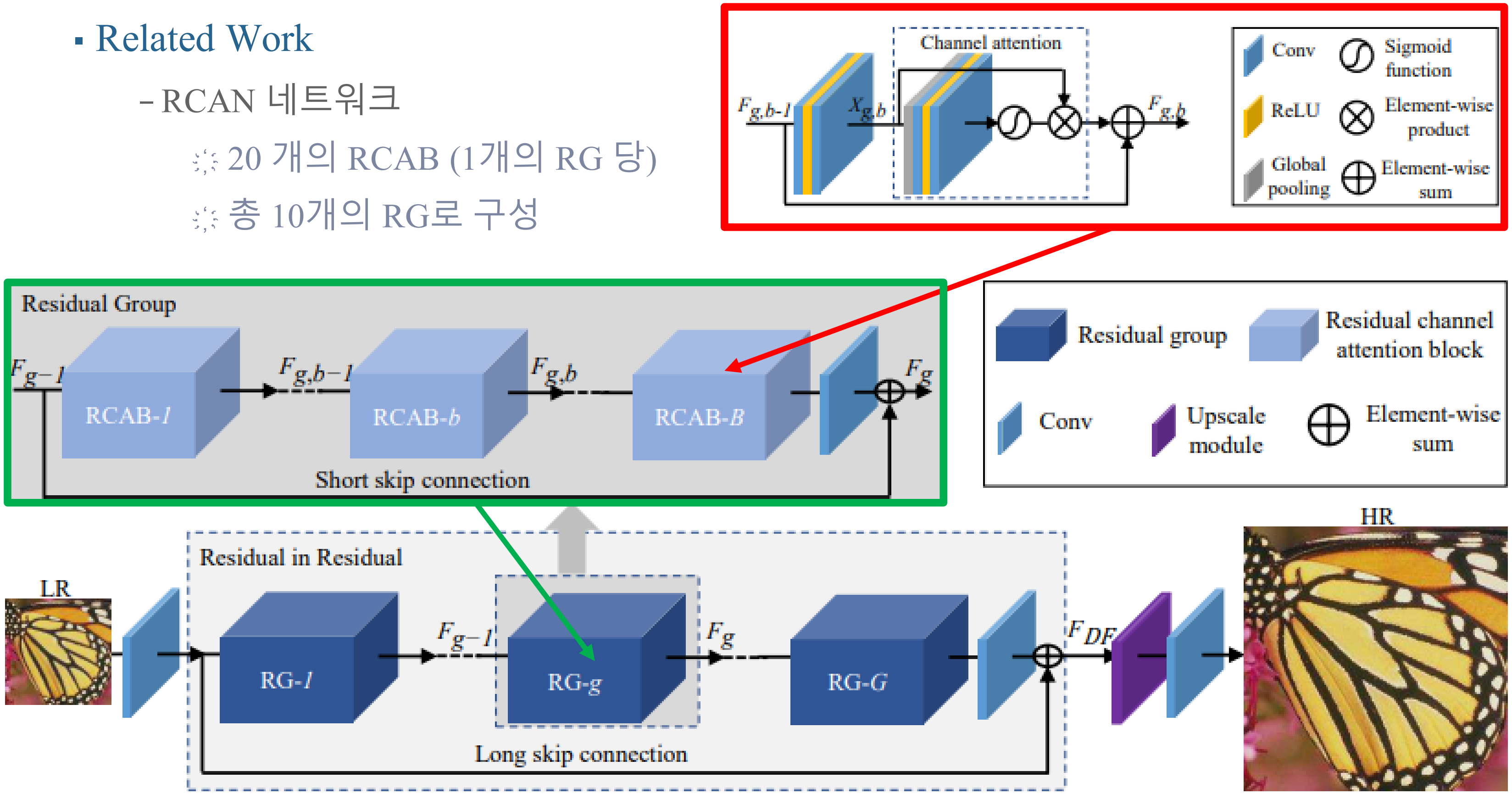


- Discriminator
  - ✓ With RaD, the discriminator estimates the probability that the given real data is more realistic than fake data, on average. [9]



Proposed Method 1

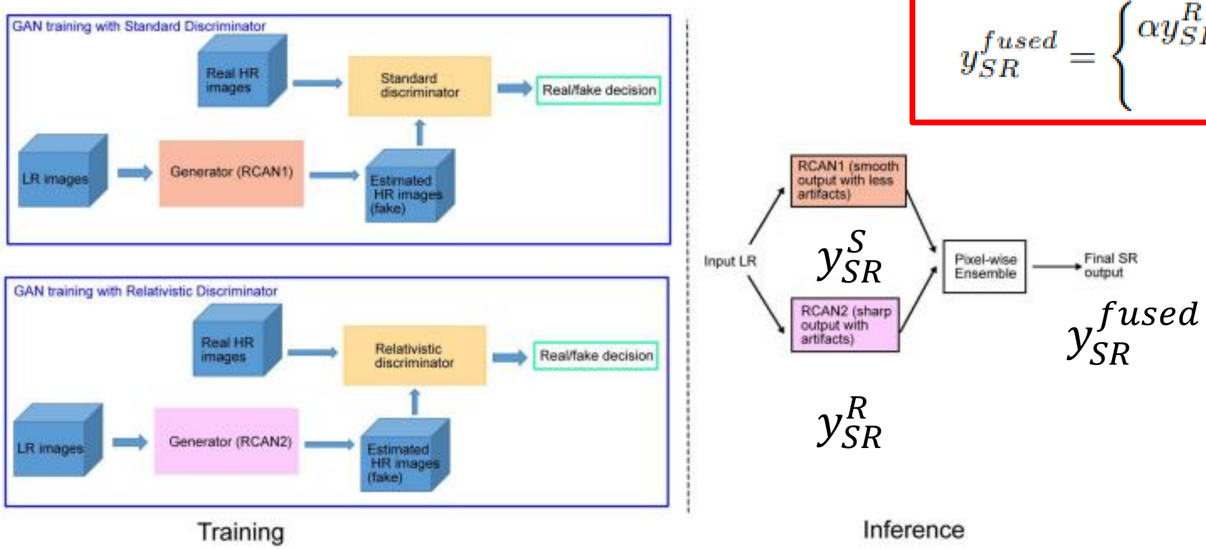
- Contribution 3
  - Related Work
    - RCAN 네트워크
      - 20 개의 RCAB (1개의 RG 당)
      - 총 10개의 RG로 구성



Proposed Method 1

- Contribution 3
  - 두개의 GAN을 사용하여 네트워크 fusion
    - Standard discriminator = 임의의 영상이 real 혹은 fake 인지 판별 → SRGAN 사용
    - Relativistic discriminator = real 영상이 fake 영상 보다 realistic 한지를 판별 → ESRGAN 사용
    - Y_{med} : median of the pixel intensity values in Y channel
      - ∴ Y value 가 낮은 부분의 image quality가 높아진다

$$y_{SR}^{fused} = \begin{cases} \alpha y_{SR}^{R} + \beta y_{SR}^{S} & if \quad Y_{med} < \gamma \\ y_{SR}^{R} & otherwise \end{cases}$$



Experimental Results

- Track 1
  - Bicubic SR : 기존의 dataset
  - Generic SR : 제안하는 dataset
  - LPIPS [10]
    - 영상의인지적 성능을 판단

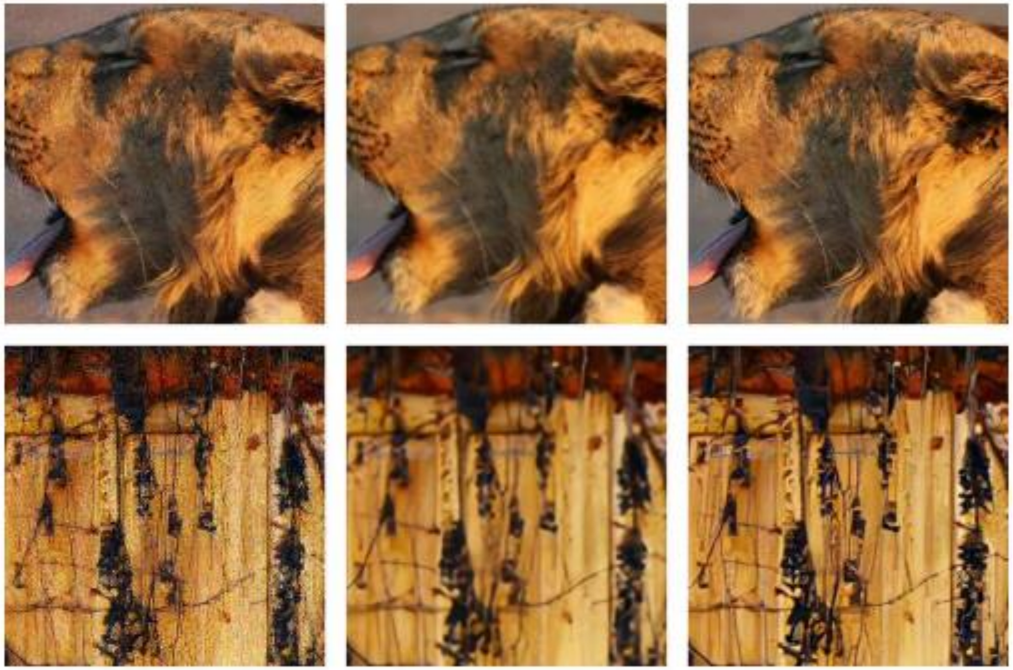


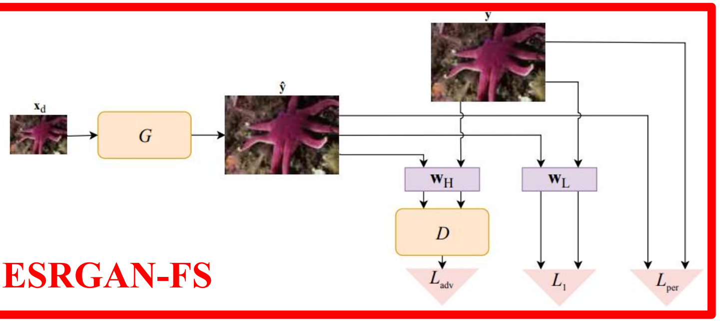
Figure 4. Qualitative comparisons of different SR algorithms on the testing images of NTIRE 2020 RealSR challenge track 1, with scaling factor x4.

Table 1. PSNR (dB)/SSIM/LPIPS evaluation of different SR methods on the validation data of NTIRE 2020 RealSR challenge track 1, with scaling factor x4. The lower LPIPS, the better.

Method	Training	PSNR(dB)/SSIM/LPIPS
ESRGAN [29]	Bicubic SR	19.06/0.2424/0.7552
PSNR-oriented RCAN	Generic SR	**27.36/0.7620**/0.3680
Ours	Generic SR	25.78/0.7119/ **0.2482**

Experimental Results

- Track 2
  - Standard ESRGAN
    - 기존의 dataset으로 학습한 ESRGAN
  - ESRGAN-FS [11]
    - Frequency Seperation을 적용한 ESRGAN
    - DPED dataset 사용
      - Mobile image → HR
      - downsampling GAN으로 LR 이미지 생성



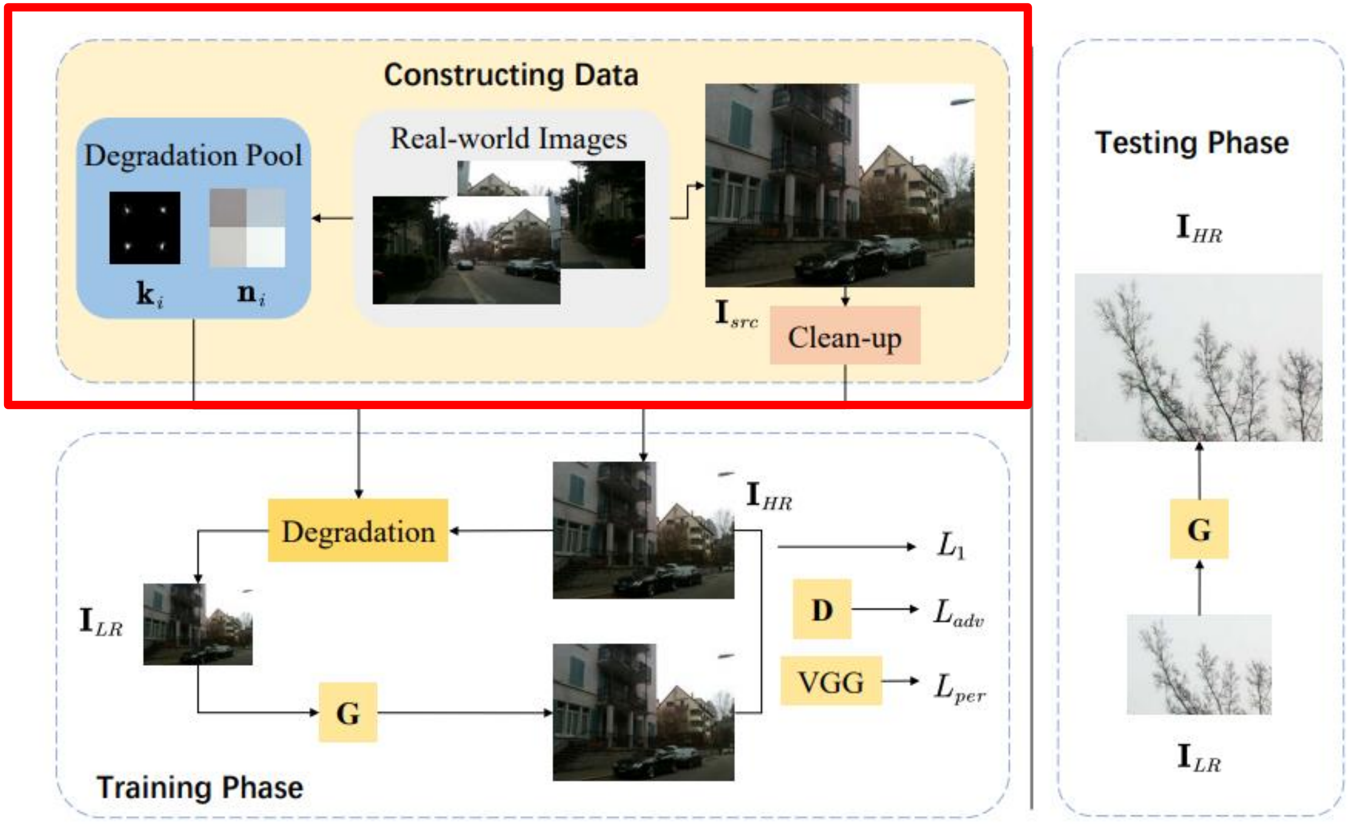
ESRGAN-FS

Proposed Method 2

- Real-World Super-Resolution via Kernel Estimation and Noise Injection
  - Contribution 1
    - We propose a novel degradation framework RealSR under real-world setting, which provides realistic images for super-resolution learning.
  - Contribution 2
    - By estimating the kernel and noise, we explore the specific degradation of blurry and noisy images.
  - Contribution 3
    - We demonstrate that the proposed RealSR achieves state-of-the-art results in terms of visual quality

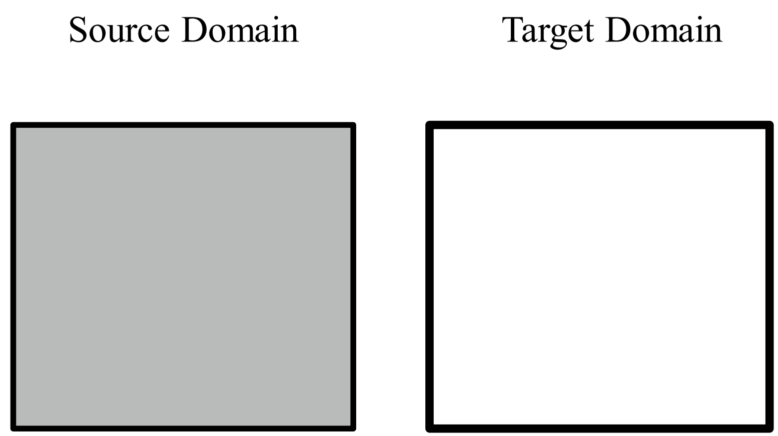
Proposed Method 2

- Network Archtecture
  - Contribution 1, 2 → 학습 데이터셋 만드는 방법



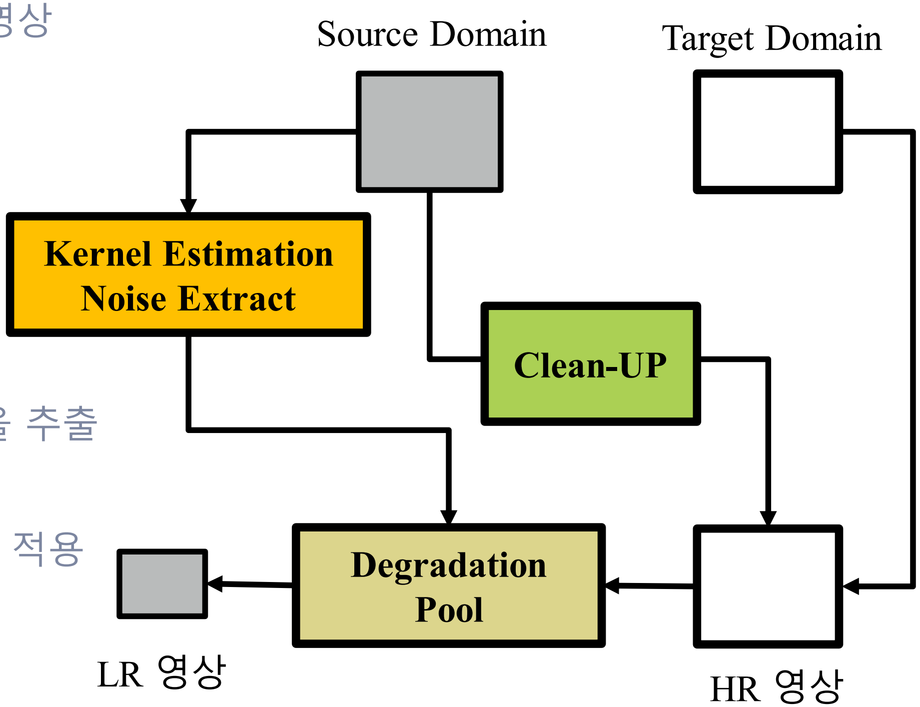
Proposed Method 2

- Contribution
  - Ntire Challenge의 데이터셋
    - 주어진 데이터셋은 LR-HR pair 데이터셋이 아니기 때문에 주어진 영상에서 pair를 형성하여야 한다.



Proposed Method 2

- Contribution
  - 학습데이터 얻는 과정
    - 원하는 HR-LR pair
      - LR : Source domain에서의 degradation 및 noise가 적용된 LR
      - HR : Target domain의 clean한 HR 영상
    - Clean-up
      - HR 영상의 개수를 늘리기 위해 이상적인 bicubic downsampling 적용
      - noise 및 artifact 제거 효과
    - Kernel Estimation / Noise Extract
      - Source domain 영상의 degradation을 추출
    - Degradation pool
      - 위에서 추출한 degradation을 HR에 적용
        - ✓ Noise Injection



Proposed Method 2

- Contribution
  - Related Work

정확한 K 를 알수 있다면 LR 영상에서 HR 영상으로 변환하는데 유용

– ZSSR (Zero-shot Super Resolution) [12] (CVPR 2018)

$$\mathrm{LR} = \mathrm{HR} * \boxed{K_{\downarrow}}$$

- Blind Super Resolution (SR)
- 입력 LR 영상으로 일반적으로 bicubic downsamping된 영상이 아닌 임의의 downsampling 기법으로 생성된 영상의 해상도를 변화하는 SR 알고리즘
- 다량의 학습데이터를 필요로 하지 않고 하나의 입력 영상을 기반으로 수행

(a) SR under aliasing:

Ground truth (PSNR /SSIM) | EDSR+ [12] (21.64 / 0.6641) | ZSSR (ours) (25.02 / 0.7658)

(b) SR under unknown *non-ideal* downscaling kernel:

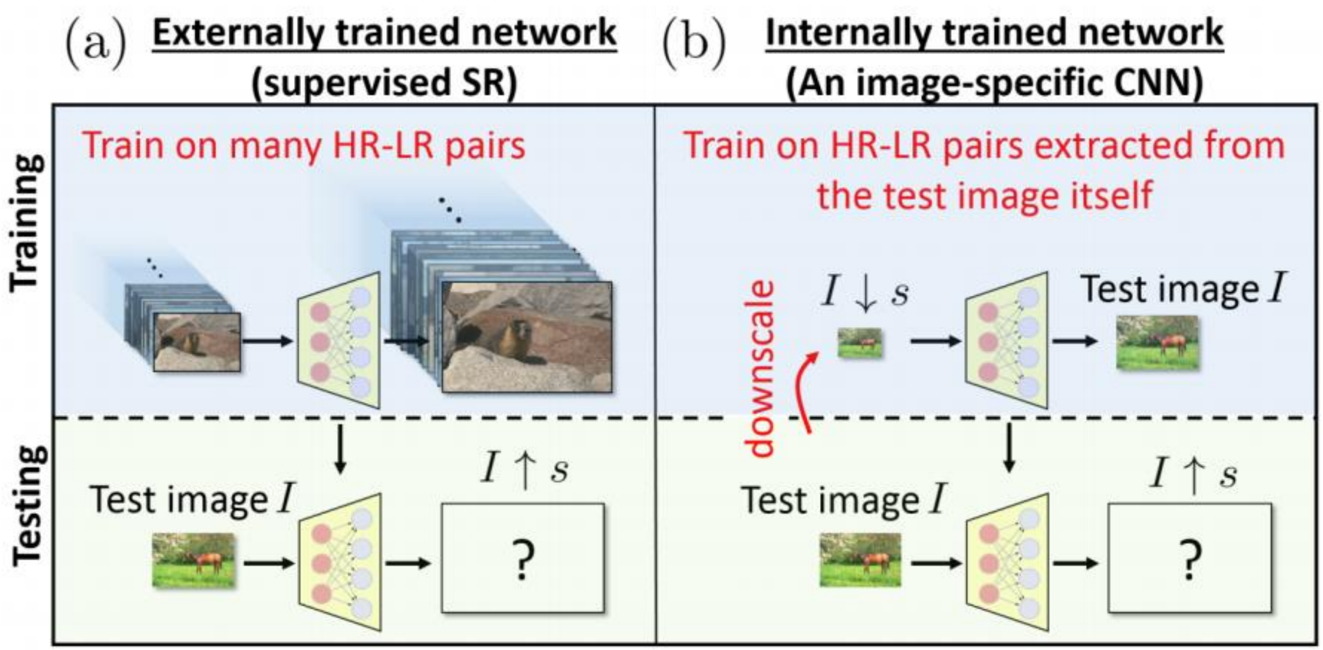
Ground truth (PSNR /SSIM) | EDSR+ [12] (24.44 / 0.7006) | ZSSR (ours) (27.62 / 0.8367)

Proposed Method 2

- Contribution
  - Related Work
    - ZSSR (Zero-shot Super Resolution)

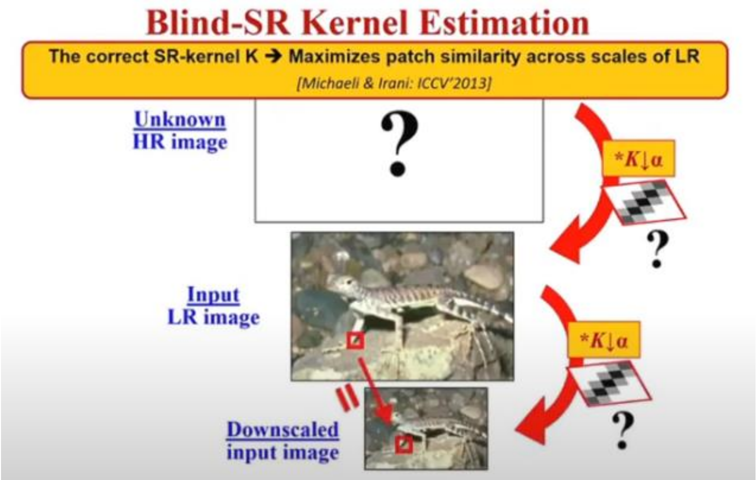
Michaeli & Irani [13] 의 Kernel Estimation 사용 (ICCV 2013)

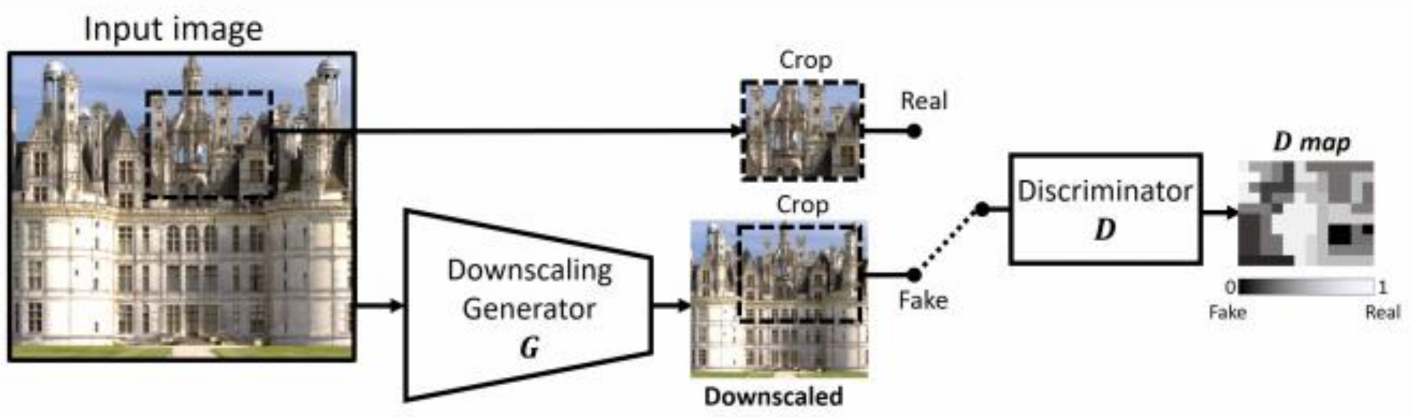
큰 학습데이터셋을 필요로 하지 않고 downsampling 방법을 구해 입력 LR 영상 하나로 학습



Proposed Method 2

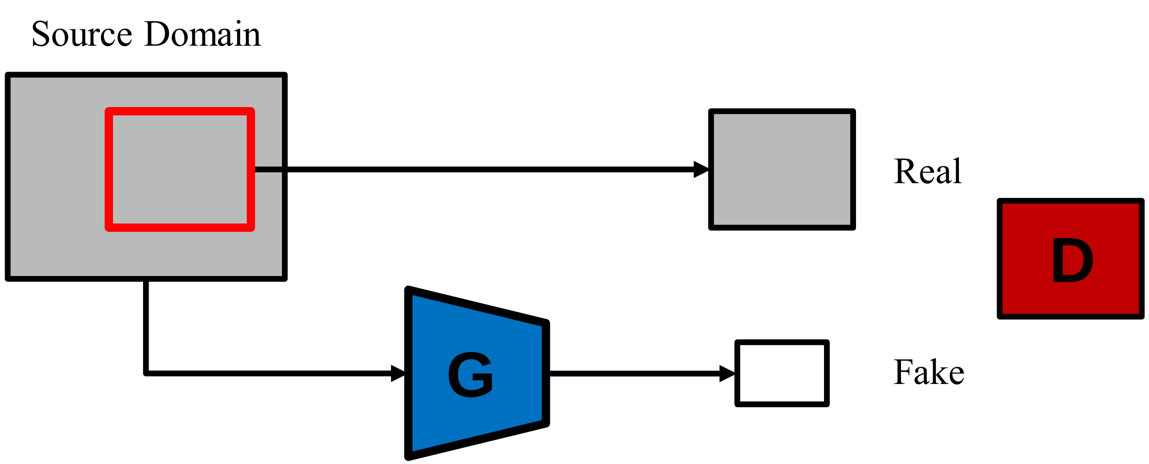
- Contribution
  - Related Work
    - KernelGAN (NIPS 2019) [14]
      - Downsamping 기법을 알아내기 위해 KernelGAN 사용
      - Downscaling Generator → Linear network
        - ✓여러 개의 global minima 후보를 생성하여 학습
      - Real 과 Fake 의 patch similarity 가 유사해지도록 학습





Proposed Method 2

- Contribution
  - Kernel Estimation
    - source domain의 영상의 저주파수 정보를 downsampled 영상이 유지하는데 도움
    - 임의의 영상을 입력으로 하였을 때 Source Domain의 patch distribution을 가지는 downsampled 영상을 출력하는 Generator (kernel) 학습
    - 각각의 영상 별로 kernel 을 출력하여 Degradation pool에 추가



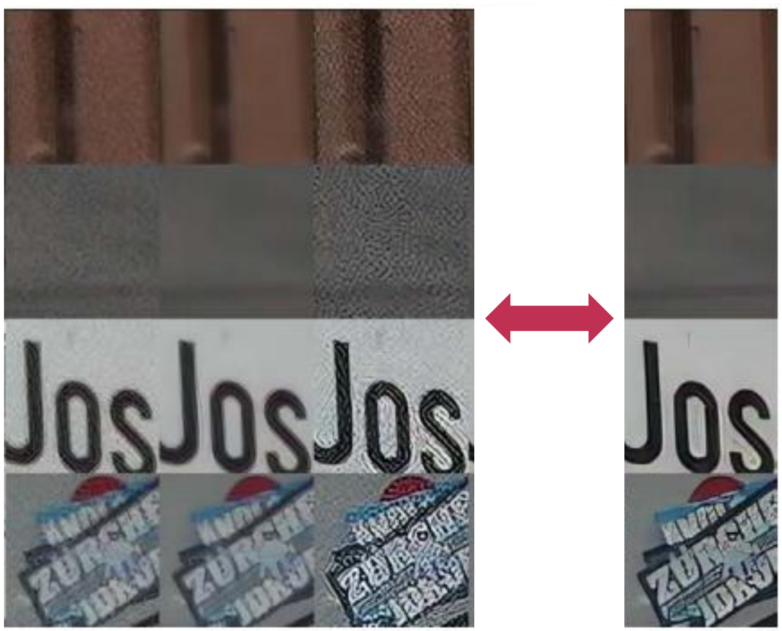
Algorithm 1 Realistic Degradation of our RealSR

Input: Real images set $\mathcal{X}$, HR images set $\mathcal{Y}$, downsampling scale factor s
Output: Realistic paired images $\{\mathbf{I}_{LR}, \mathbf{I}_{HR}\}$

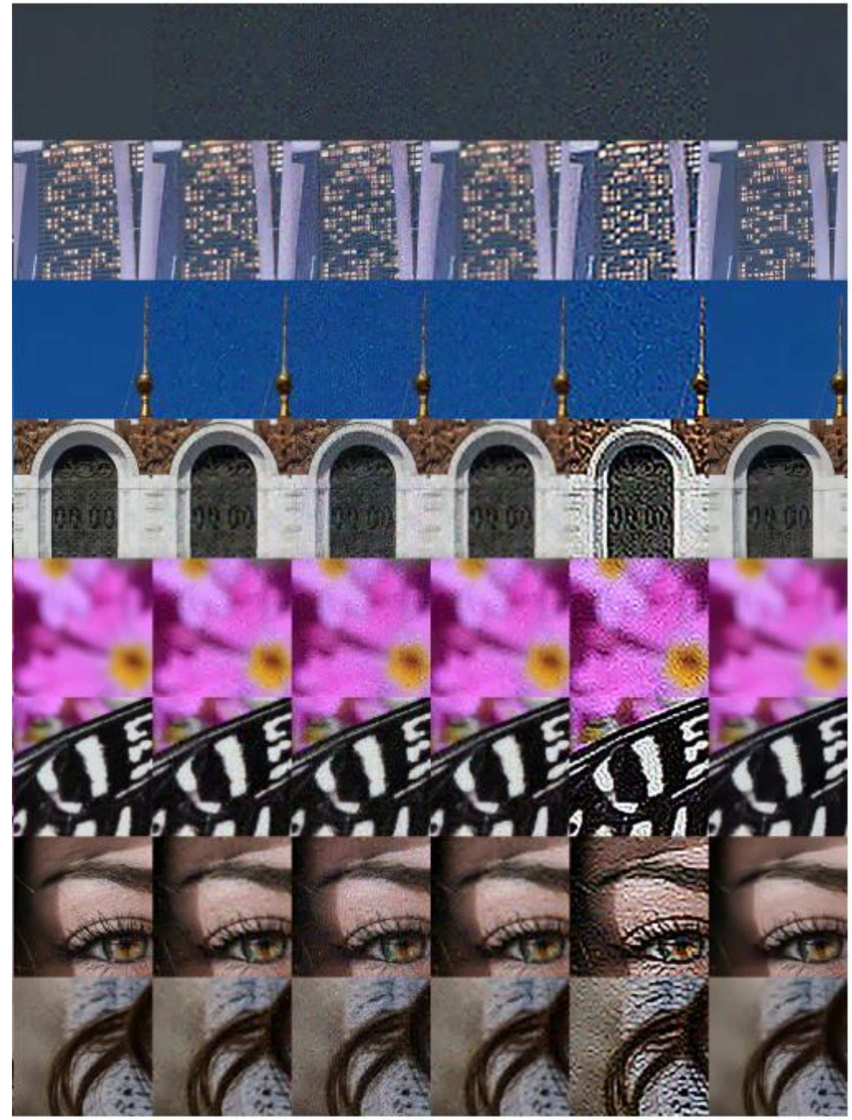
Initialize kernel pool $\mathcal{K} = \emptyset$
Initialize noise pool $\mathcal{N} = \emptyset$
for all $\mathbf{I}_{src}$ such that $\mathbf{I}_{src} \in \mathcal{X}$ **do**
  Estimate $\mathbf{k}$ from $\mathbf{I}_{src}$ by solving Eqn. 4
  Add $\mathbf{k}$ to $\mathcal{K}$
  Crop $\mathbf{n}$ from $\mathbf{I}_{src}$
  if $\mathbf{n}$ meet Eqn. 7 **then**
    Add $\mathbf{n}$ to $\mathcal{N}$
  end if
end for
for all $\mathbf{I}_{HR}$ such that $\mathbf{I}_{HR} \in \mathcal{Y}$ **do**
  Randomly select $\mathbf{k}_i \in \mathcal{K}$, $\mathbf{n}_j \in \mathcal{N}$
  Generate $\mathbf{I}_{LR}$ with $\mathbf{k}_i$ and $\mathbf{n}_j$
end for
return $\{\mathbf{I}_{LR}, \mathbf{I}_{HR}\}$

Experimental Results

- Track 1
  - Ablation Study
    - Noise → Kernel Estimation 제거
    - Kernel → Noise Injection 제거



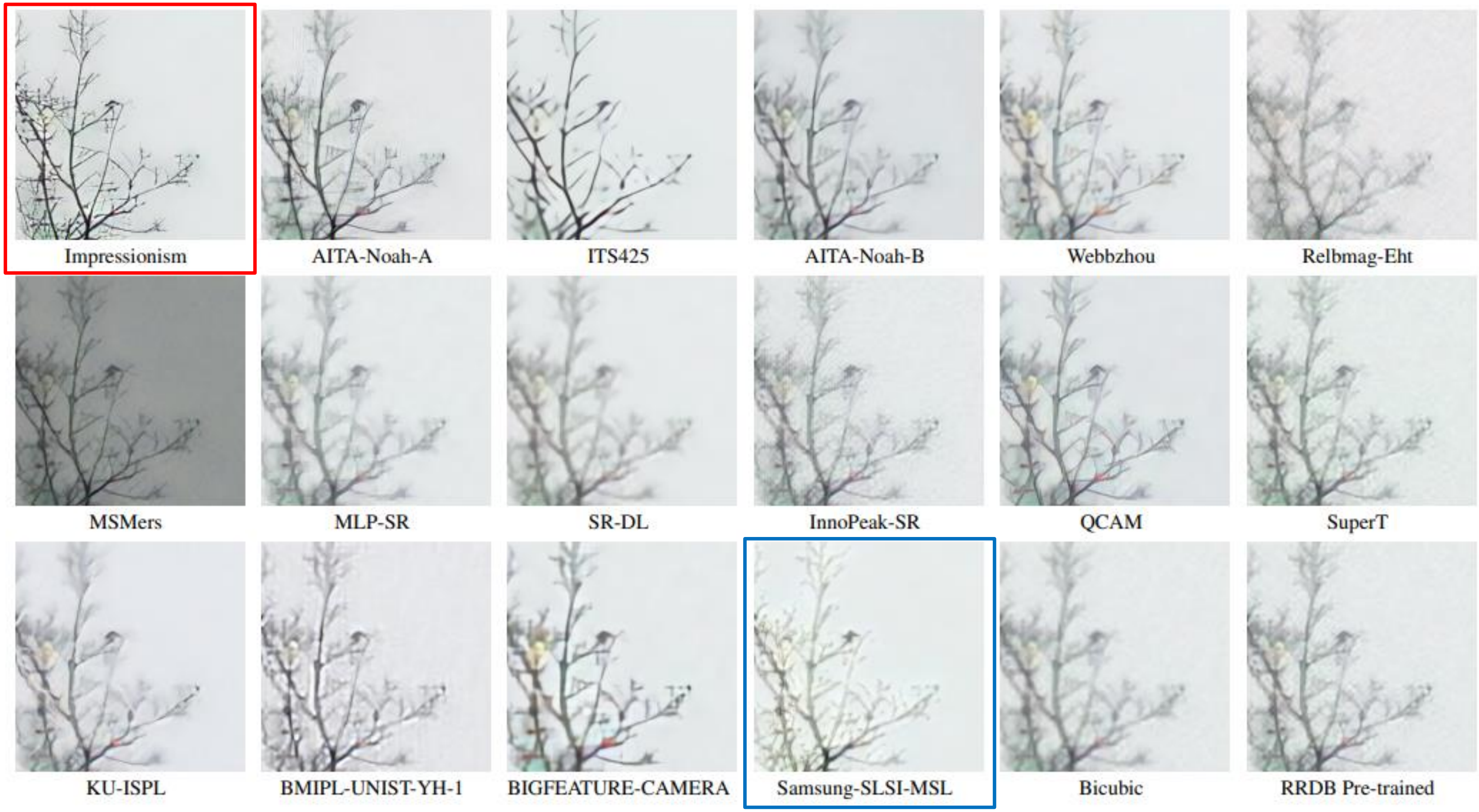
Bicubic Noise Kernel Patch



GT EDSR ESRGAN ZSSR K-ZSSR Ours

Experimental Results

- Track 2



References

- 1. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. TPAMI 38(2), 295–307 (2016)
- 2. Kim, J., Kwon Lee, J., Mu Lee, K.: Accurate image super-resolution using very deep convolutional networks. In: CVPR (2016)
- 3. Ledig, C., Theis, L., Husz´ar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., Shi, W.: Photo-realistic single image superresolution using a generative adversarial network. In: CVPR (2017)
- 4. Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M.: Enhanced deep residual networks for single image super-resolution. In: CVPRW (2017)
- 5. Zhang, Yulun, et al. "Image super-resolution using very deep residual channel attention networks." Proceedings of the European Conference on Computer Vision (ECCV). 2018.
- 6. Wang, Xintao, et al. "Esrgan: Enhanced super-resolution generative adversarial networks." Proceedings of the European Conference on Computer Vision (ECCV). 2018.
- 7. Johnson, Justin, Alexandre Alahi, and Li Fei-Fei. "Perceptual losses for real-time style transfer and super-resolution." European conference on computer vision. Springer, Cham, 2016.
- 8. Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. DSLR-quality photos on mobile devices with deep convolutional networks. In IEEE International Conference on Computer Vision, pages 3277– 3285, 2017.

References

- 9. Jolicoeur-Martineau, A.: The relativistic discriminator: a key element missing from standard gan. arXiv preprint arXiv:1807.00734 (2018)
- 10. Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 586–595, 2018.
- 11. Manuel Fritsche, Shuhang Gu, and Radu Timofte. Frequency separation for real-world super-resolution. arXiv preprint arXiv:1911.07850, 2019.
- 12. Assaf Shocher, Nadav Cohen, and Michal Irani. “zero-shot” super-resolution using deep internal learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3118–3126, 2018.
- 13. T. Michaeli and M. Irani. Nonparametric blind super-resolution. In International Conference on Computer Vision (ICCV), 2013.
- 14. Sefi Bell-Kligler, Assaf Shocher, and Michal Irani. Blind super-resolution kernel estimation using an internal-gan. In Advances in Neural Information Processing Systems, pages 284–293, 2019.