Image Super Resolution

7기 최명헌
8기 장준혁
조보경
최윤서
황진우

목차

A table of Contents

Part 1,

왜 Image Super Resolution인가

Why Image Super Resolution?

사진 화질 개선

압축된 사진 복원

과거의 사진 화질 개선

CCTV 범인 추적

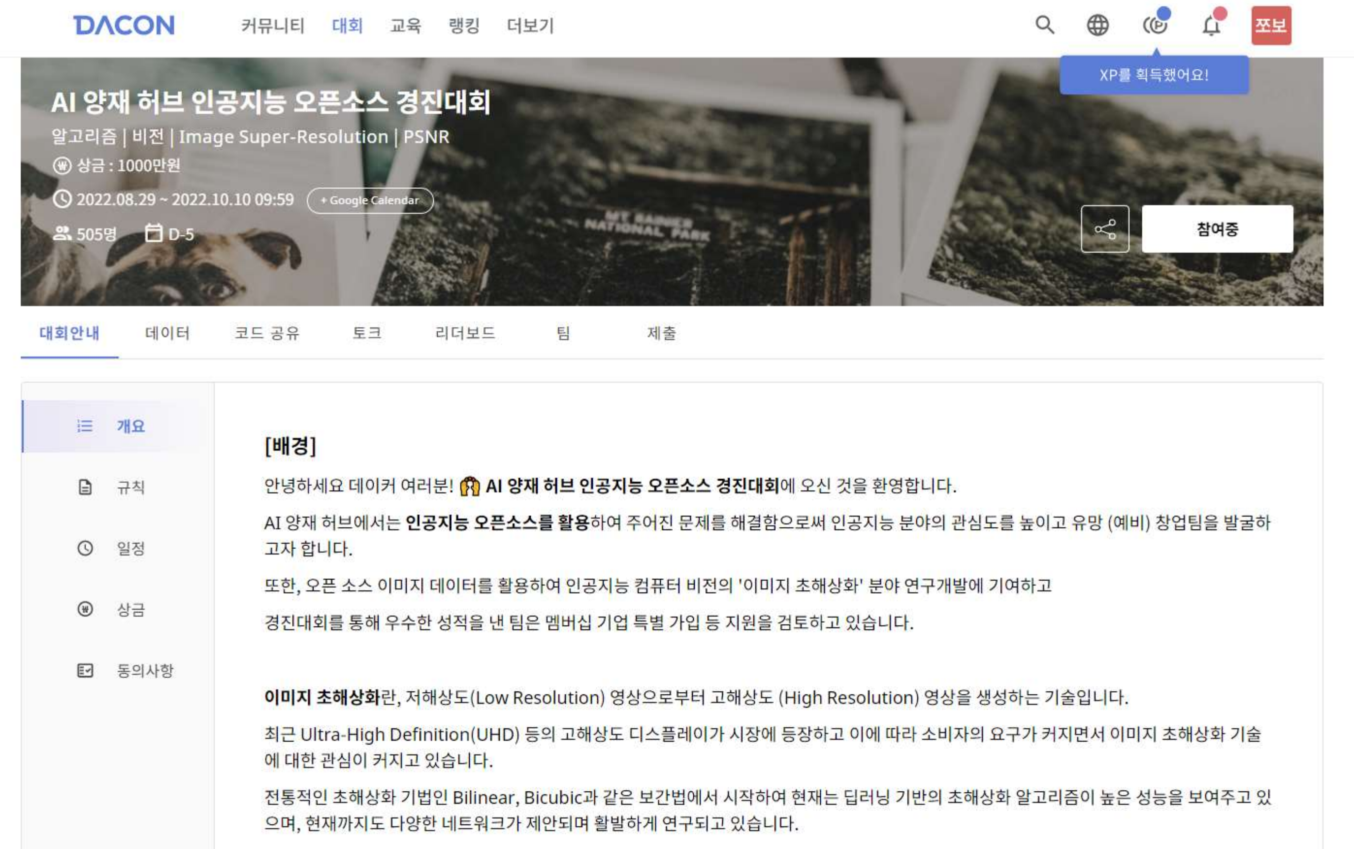
차량 번호판

범인 얼굴 색출

Part 2,

Data 소개

Part 2 # Data 소개



"LR to HR"

저해상도, 고해상도

촬영 이미지 사용

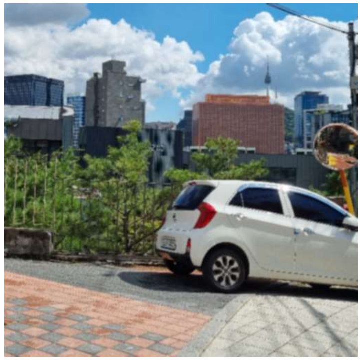
Part 2

Data 소개

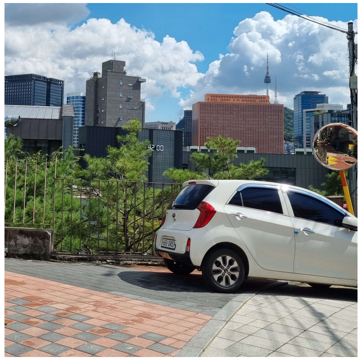
Train data(1640개)

Low-Resolution Image (input)와 High-Resolution Image (target) pair로 존재

ex)



512*512 LR



2048*2048 HR

Test(18개)

Low-Resolution Image

512*512 LR

Data 소개

Train(1640개)

LR Image (input)
(512*512)

HR Image (target)
(2048*2048)

Train	Valid	Test
0.6	0.2	0.2
984개	328개	328개

평가 metric for SR

PSNR

Peak-Signal-to-Noise-Ratio

영상 압축했을 때 화질이 얼마나 손실되었는지 평가

"PSNR이 높을수록 원본 영상에 비해 손실이 적다는 의미"

$$PSNR = 10\log\frac{s^2}{MSE}, \text{(where, } s^2 = \max(pixel - value))$$

실제 HR 이미지와의 유사성

SSIM

Structural Similarity Index Map

수치적 에러가 아닌 인간의 시각적 화질 차이를 평가

"SSIM이 높을수록 원본 영상의 품질에 가깝다는 의미"

$$SSIM(x,y) = [l(x,y)]^{\alpha} \cdot [c(x,y)]^{\beta} \cdot [s(x,y)]^{\gamma}$$

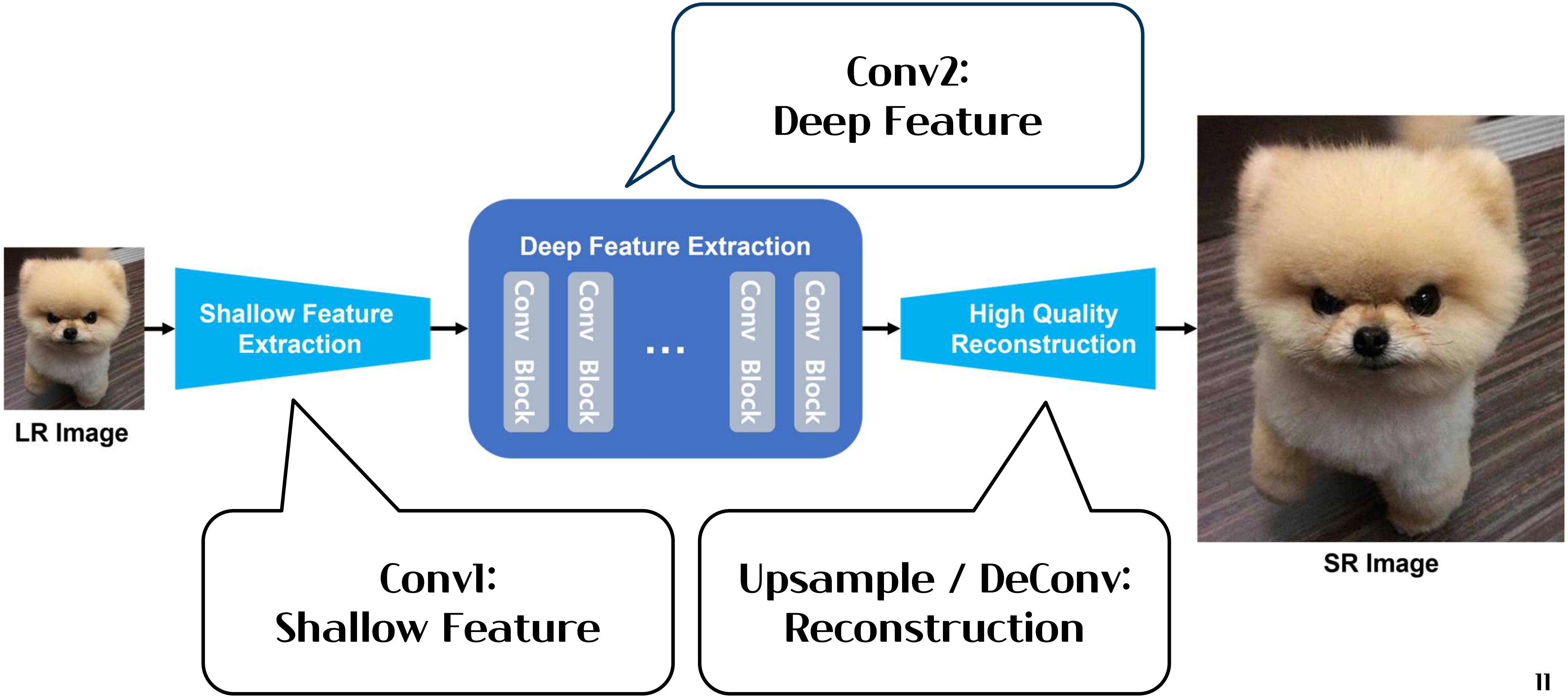
$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}$$

인간이 느끼는 품질

Part 3,

적용 가능한 모델

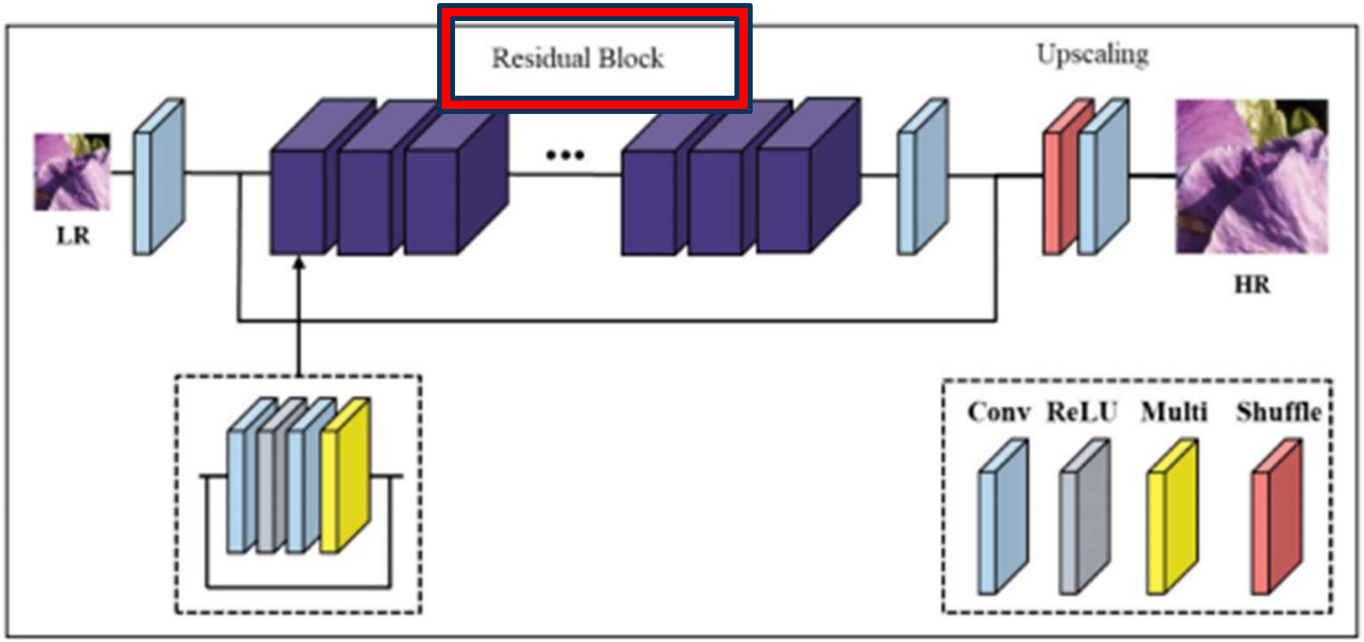
CNN based


Conv2:
Deep Feature
LR Image
Shallow Feature Extraction
Deep Feature Extraction
Conv Block
Conv Block
...
Conv Block
Conv Block
High Quality Reconstruction
SR Image
Conv1:
Shallow Feature
Upsample / DeConv:
Reconstruction

Part 3

CNN based

EDSR



-BN Layer 제거
Network의 range flexibility 감소 방지

-Residual Scaling
Feature map 개수 증가

Part 3

GAN based

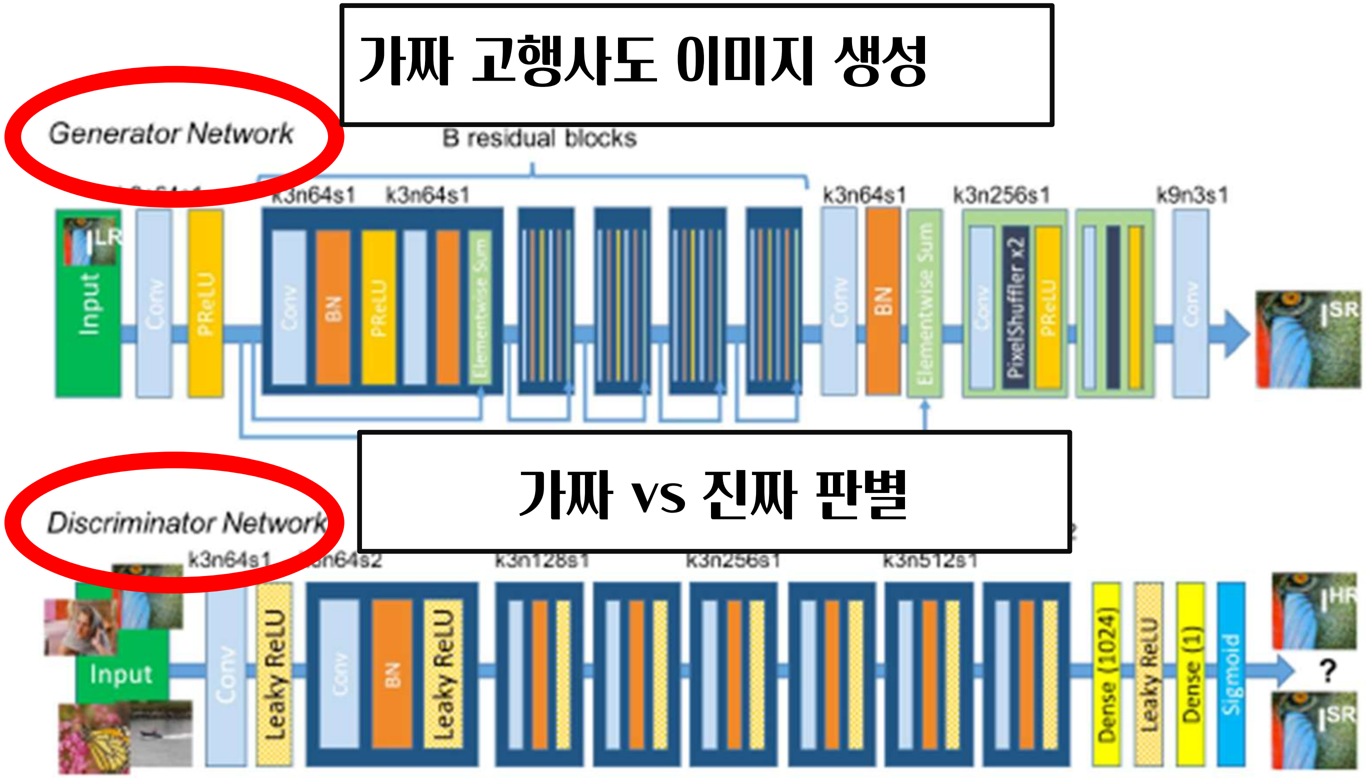
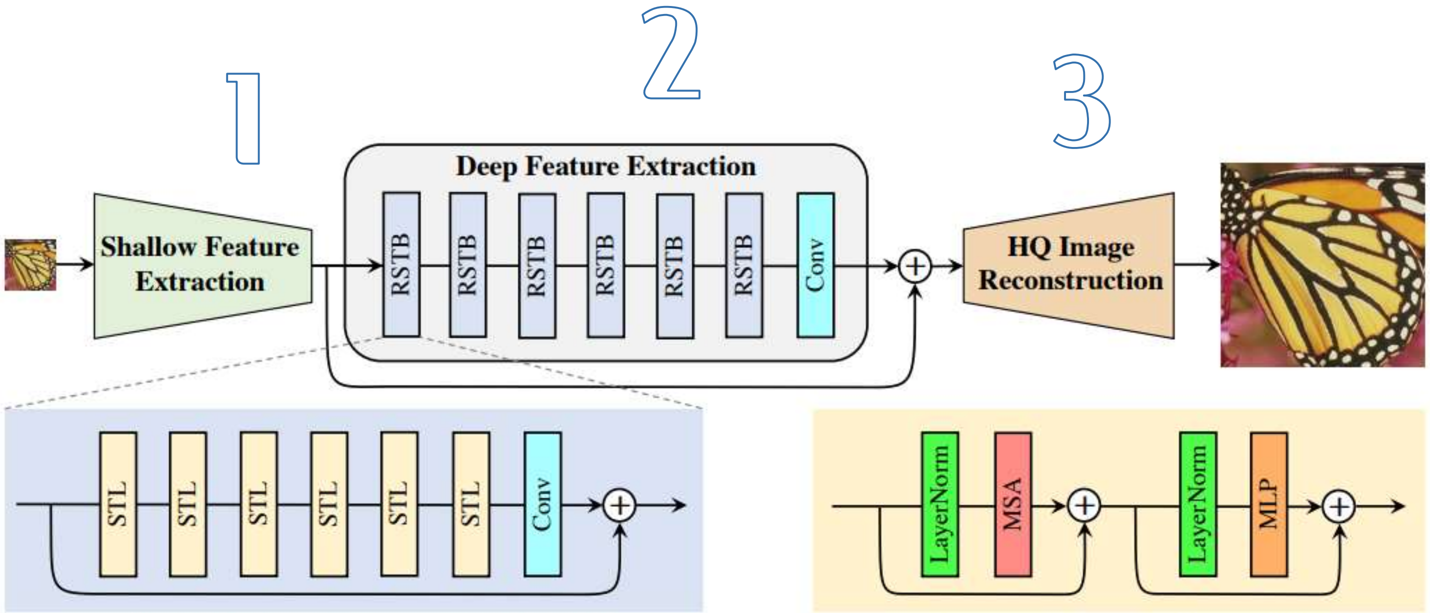


Figure 4: Architecture of Generator and Discriminator Network with corresponding kernel size (k), number of feature maps (n) and stride (s) indicated for each convolutional layer.

Transformer based

SWINIR



(a) Residual Swin Transformer Block (RSTB)

(b) Swin Transformer Layer (STL)

Figure 2: The architecture of the proposed SwinIR for image restoration.

1 Convolution Layer -> Shallow Feature

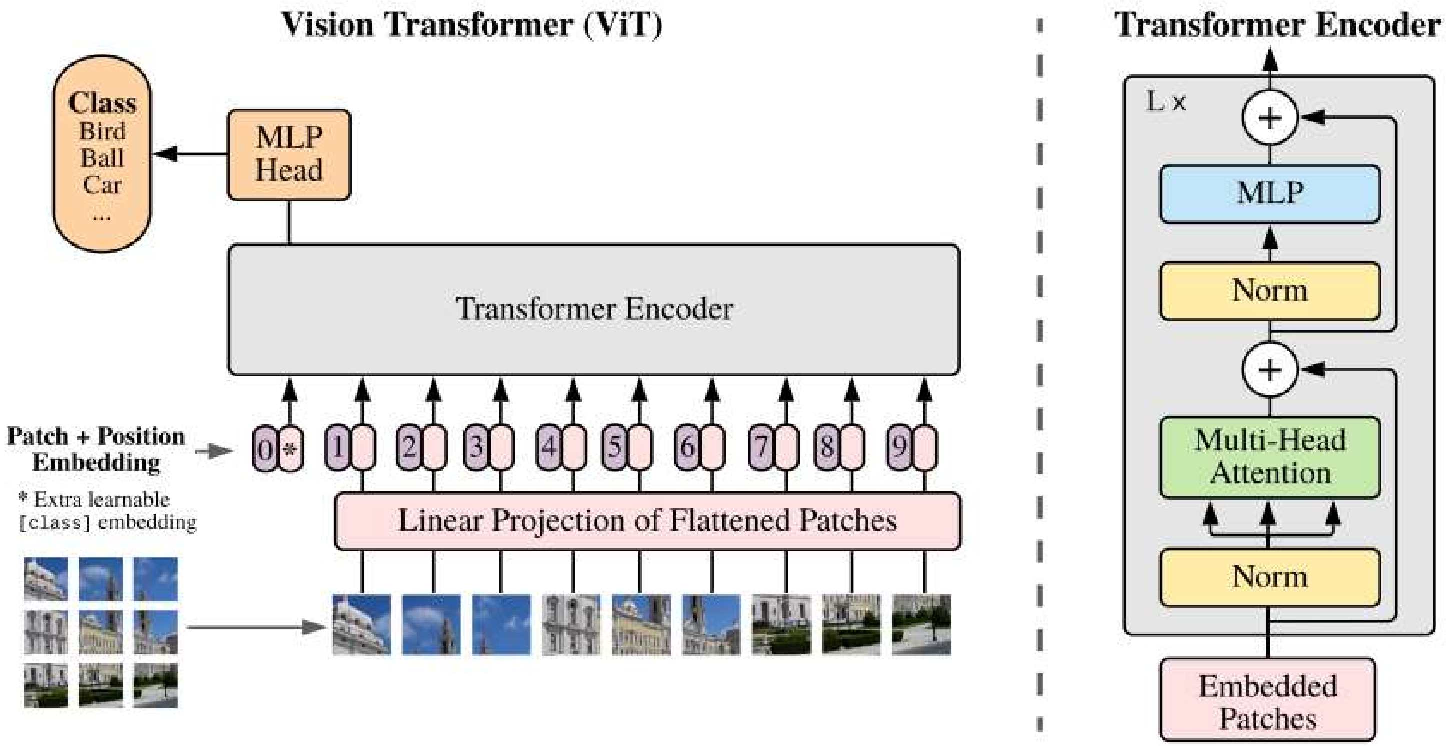
reconstruction module 에 직접 연결

2 Deep Feature 추출

3 Shallow Feature +Deep Feature

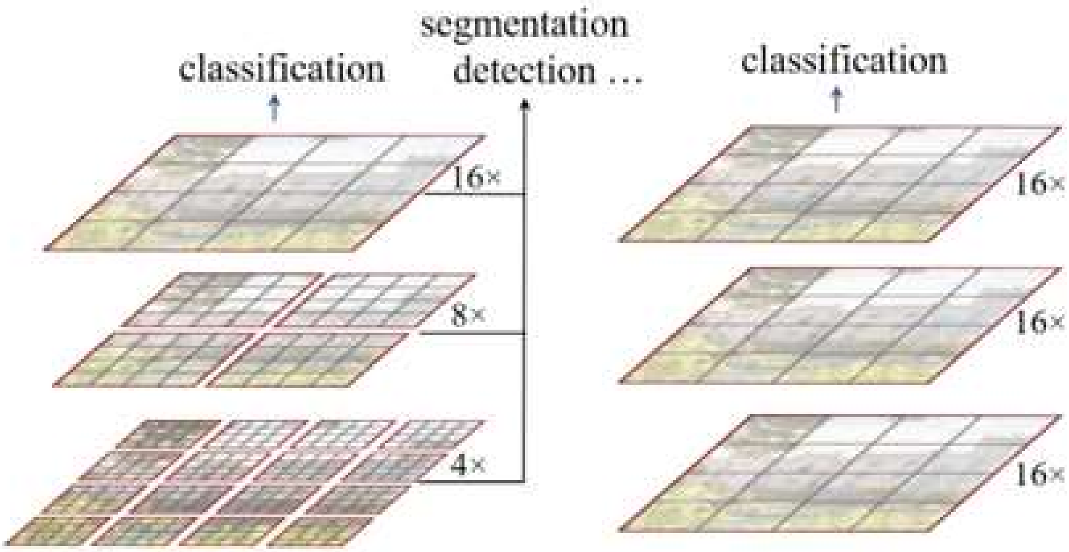
Transformer based

Vision Transformer



Transformer based

Hierarchical Feature Map



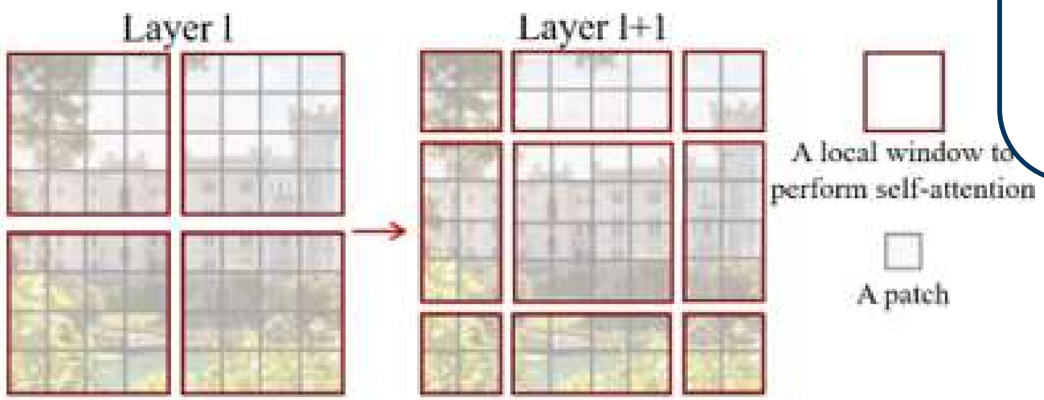
(a) Swin Transformer (ours)
(b) ViT

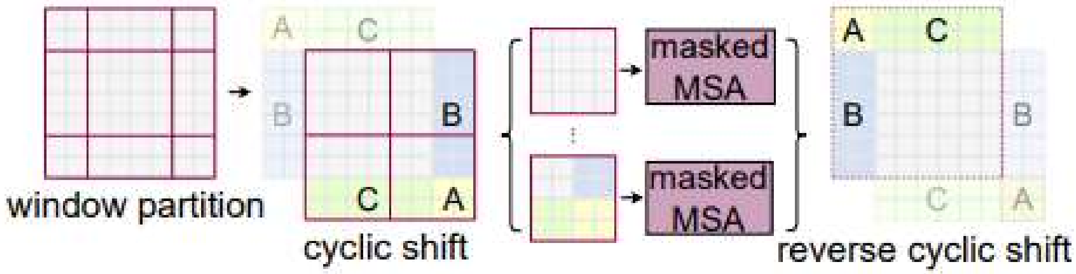
이미지의 특징을 더 잘 보기 위해
Segmentation

Transformer based

Shifted Window

가장자리의 학습률 증가





Window Slide
패치들이 안 겹치게

Shifted Window
가장자리 패치를 이동

Transformer based

SWINIR

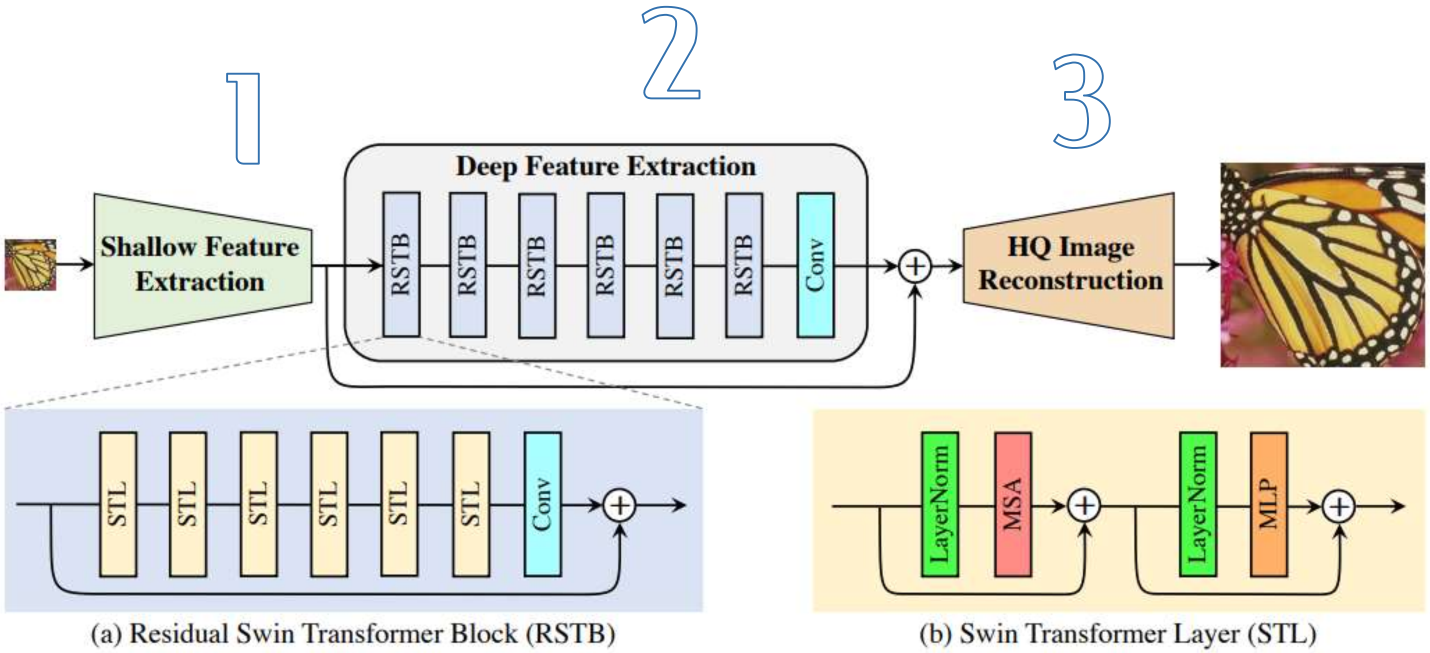


Figure 2: The architecture of the proposed SwinIR for image restoration.

1 Convolution Layer -> Shallow Feature

reconstruction module 에 직접 연결

2 Deep Feature 추출

3 Shallow Feature +Deep Feature

Part 4,

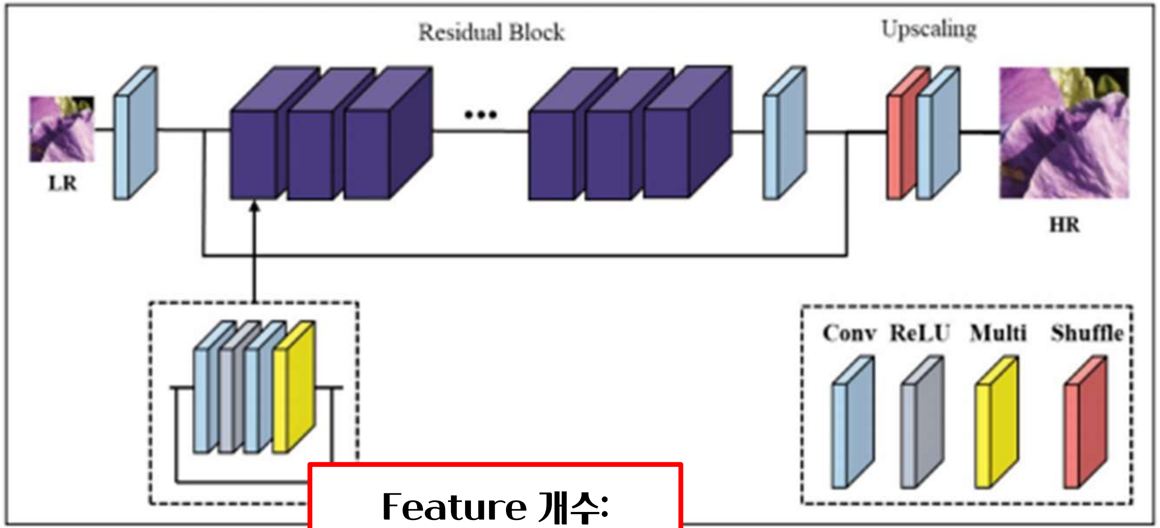
Toy Model

Part 4

Toy Model; EDSR

Structure

Residual Block:
16개



Feature 개수:
64개

Hyperparameter

epoch : 10
loss function : L1 Loss
optimizer : Adam
batch_size = 1
learning_rate = 0.0001
input size = 512,512,3
output size = 2048, 2048, 3

After 10 epoch

train loss : 0.00851
train PSNR : 20.2574
valid loss : 0.00987
valid PSNR : 20.1106

Part 4

Toy Model; SWINIR

Structure

shallow feature extraction

deep feature extraction

RSTB
Attention head 6개
STL * 6개

RSTB
Attention head 6개
STL * 6개

RSTB
Attention head 6개
STL * 6개

Conv2d

HQ Image Reconstruction

Hyperparameter

epoch : 10
loss function : L1 Loss
optimizer : Adam
batch_size = 1
learning_rate = 0.0001
input size = 512,512,3
output size = 2048, 2048, 3

After 10 epoch

train loss : 0.0548
train PSNR : 21.9173
valid loss : 0.0550
valid PSNR : 21.9173

Toy Model

성능

	PSNR
EDSR	22.7684
SWINIR	21.1658

Dacon Leader board 기준

Toy Model

Toy Model Result

Issue!

Input Size의 증가 + 큰 모델 사이즈
=
OOM(Out Of Memory)

→ 모델의 일부 Layer 삭제

→ 성능 저하

Solution!

512*512 사이즈의 Image를
64*64 사이즈로 자르자!

		논문 Architecture	Toymodel Architecture
INPUT Size		64*64 or 44*44	512 * 512
EDSR	Residual Block	64개	16개
	Feature Dimension	256개	64개
SwinIR	RSTB	4개	3개
	STL in 1 RSTB	6개	6개
	Attention Heads in 1 RSTB	6개	6개

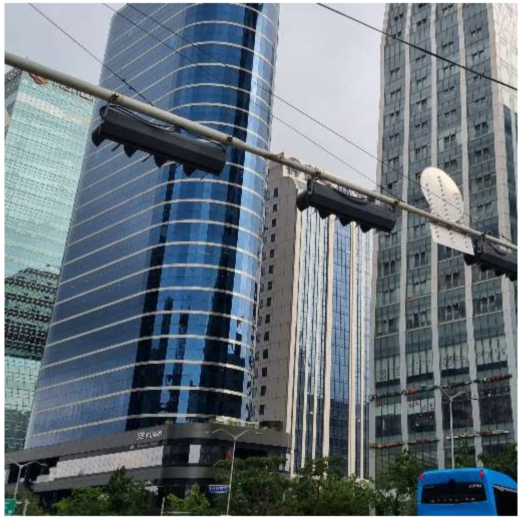
Part5,

Improvement

Part 5

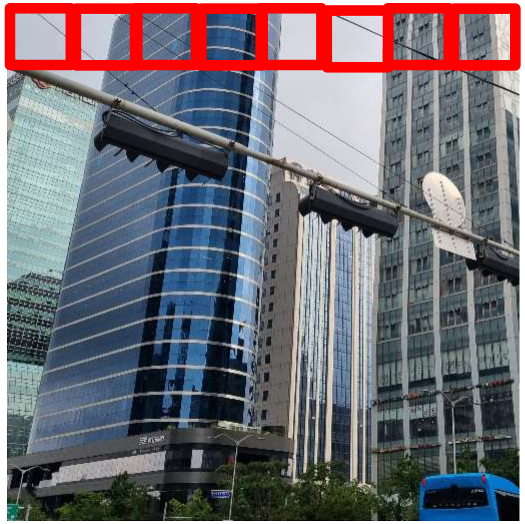
Data processing

해결방안



512*512

기존의 LR Image (Input)



64*64

1장의 input image(512*512)를

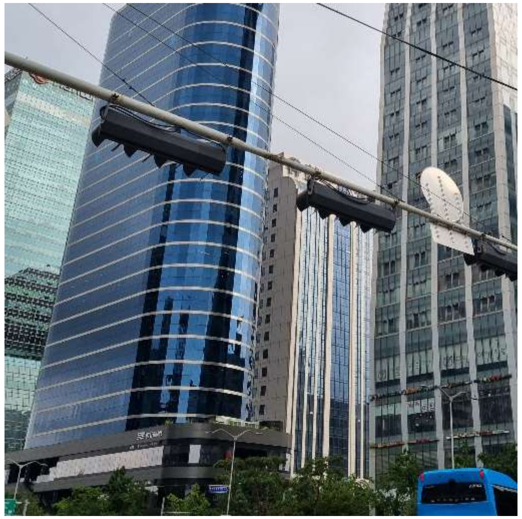
64장의 patch(64*64)로 잘라서

총 1640장*64개 = 104,960개의 이미지로 학습

Part 5

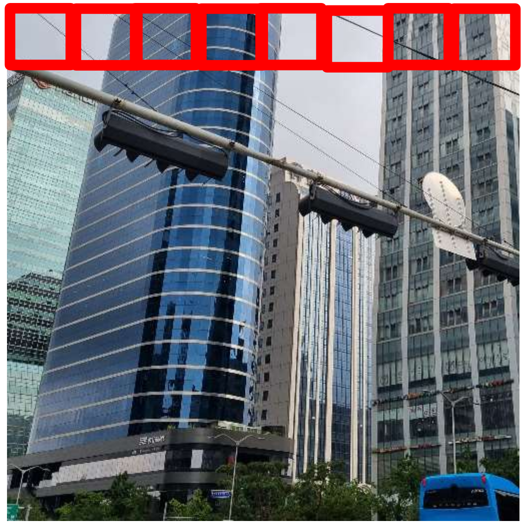
Data processing

해결방안



2048*2048

기존의 HR Image (Target)



256*256

LR, HR 모두 64개 패치로 자름

~~512*512 -> 2048*2048~~

~~Upscaling 모델 만들자!~~

64*64 -> 256*256

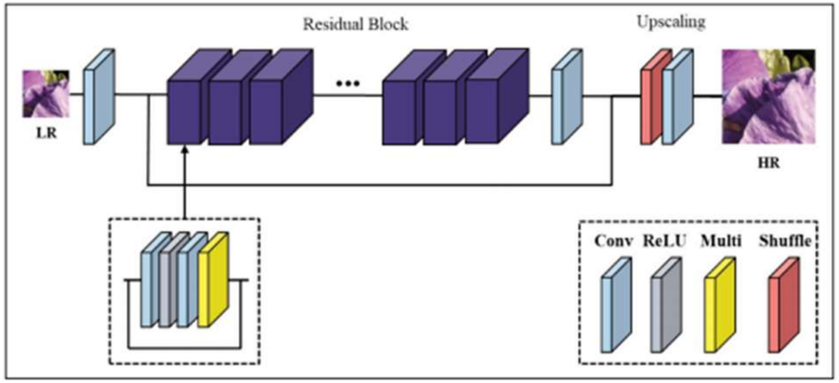
Upscaling 모델 만들자!

Custom Model

EDSR + SwinIR

EDSR

CNN:
국소적 특징



SWINIR

Transformer:
전역적 특징

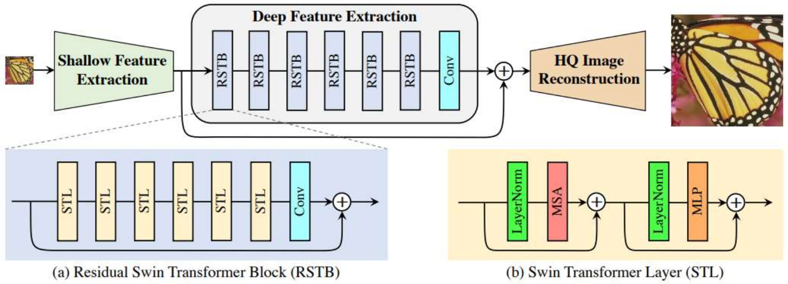
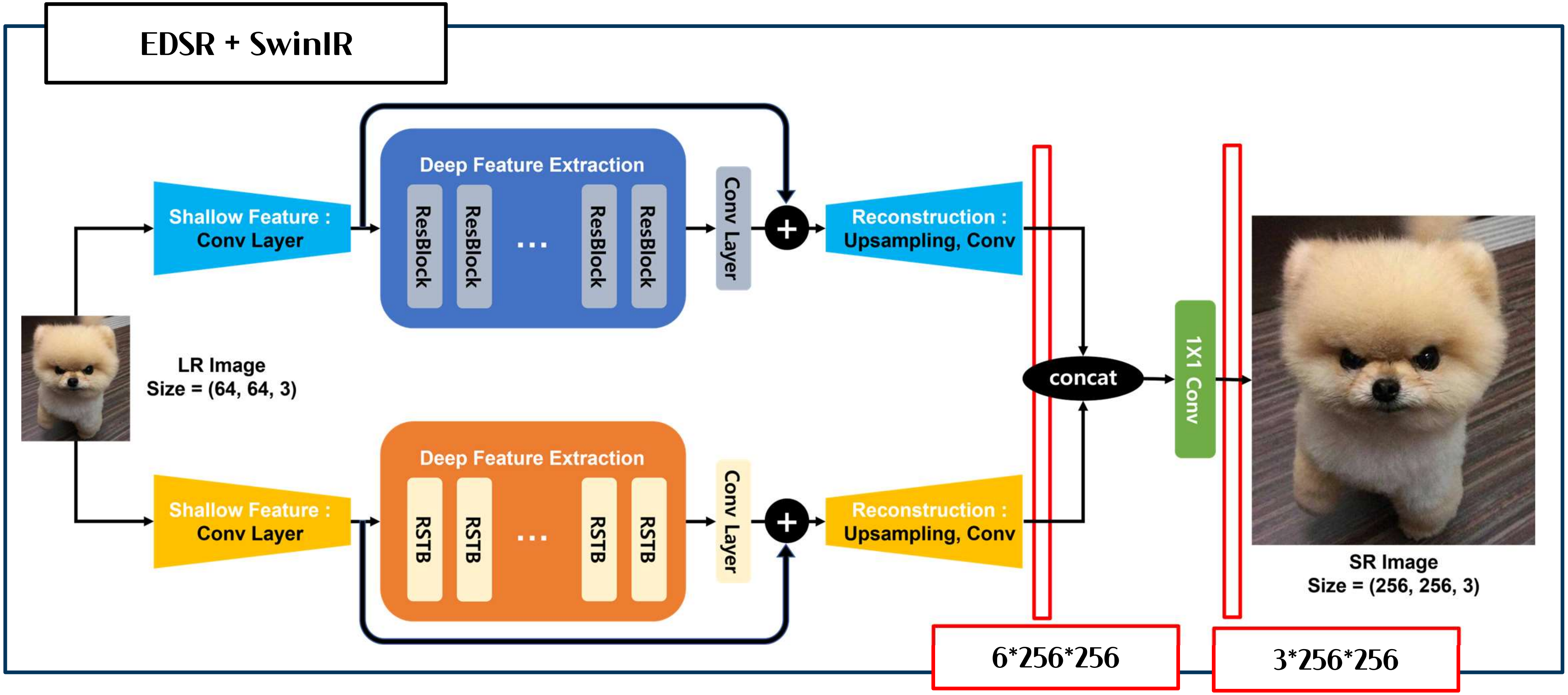


Figure 2: The architecture of the proposed SwinIR for image restoration.

Part 5

Custom Model



Custom Model Training

학습과정

15 epoch

Train : Valid = 8 : 2

15 epoch

Train : Valid = 1 : 0
+ Horizontal / Vertical Flip

Model Structure

SwinIR :
- RSTB : 4개
- STL in 1 RSTB : 6개
- Attention Heads : 6개
- Window Size : 8

EDSR :
- Featrue Dim : 256
- ResBlock : 64개

Hyperparameter

epoch : 10
loss function : L1 Loss
optimizer : Adam
scheduler : Cosine Annealing
batch_size = 16
learning_rate = 0.0001
input size = 64, 64, 3
output size = 512, 512, 3

Custom Model Training

Result

<u>15 epoch</u>

	Train	Valid
비율	0.8	0.2
loss	0.0529	0.0521
PSNR	21.1272	21.2176

<u>15 epoch</u>

+ Horizontal / Vertical Flip

	Train	Valid	Dacon 제출
비율	1.0	0.0	.
loss	0.05486	X	.
PSNR	20.7179	X	23.1245

Part6,

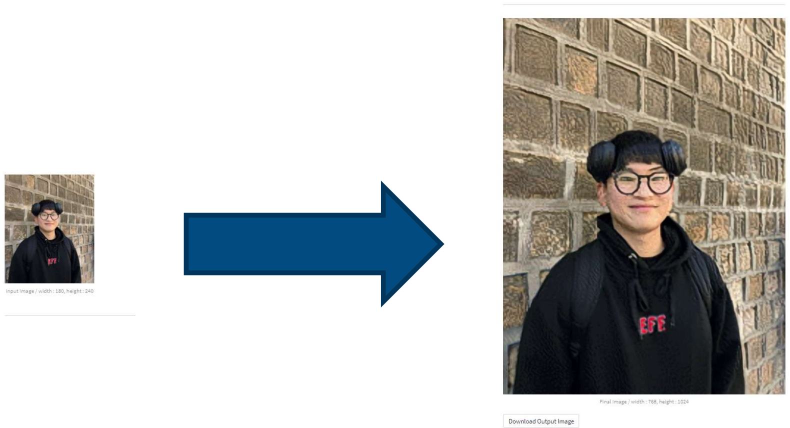
Streamlit

Limitation

1. Resource- 학습환경

2. Patch로 인한 가장자리 비율 증가

3. Detail한 요소가 많은 이미지



Reference

강석주, 서유림 (2020). 딥러닝 기반 Super Resolution 기술의 현황 및 최신 동향 방송과 미디어 = Broadcasting and media magazine v.25 no.2 , pp.7 – 16.

Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee,(2017) "Enhanced Deep Residual Networks for Single Image Super-Resolution," 2nd NTIRE: New Trends in Image Restoration and Enhancement workshop and challenge on image super-resolution in conjunction with CVPR

Ledig, Christian, et al.(2017) "Photo-realistic single image super-resolution using a generative adversarial network." Proceedings of the IEEE conference on computer vision and pattern recognition.

Liang, Jingyun, et al.(2021) "Swinir: Image restoration using swintransformer." Proceedings of the IEEE/CVF International Conference on Computer Vision.

Lim, Bee, et al (2017). "Enhanced deep residual networks for single image super-resolution." Proceedings of the IEEE conference on computer vision and pattern recognition workshops.

Wang, Xintao, et al.(2018) "Esrgan: Enhanced super-resolution generative adversarial networks." Proceedings of the European conference on computer vision (ECCV) workshops.

Thank You for Listening