
Super Resolution with Diffusion models

DMQA Open Seminar

2024.07.19

Data Mining & Quality Analytics Lab.

정구진

발표자 소개



정구진

- 고려대학교 산업경영공학과 대학원 재학
- Data Mining & Quality Analytics Lab. (김성범 교수님)
- Ph.D. Student (2023.03 ~ Present)

Research Interest

- Computer vision
- Generative models

Contact

- kujhin@korea.ac.kr
-

TABLE OF CONTENTS

01

- Overview

- Super Resolution
- Ill-posed problem

02

- SR3

- 논문 소개
- DDPM
- 방법론
- 결과

03

- SR3+

- 논문 소개
- Degradation
- 결과

04

- StableSR

- 논문 소개
- 방법론
- Time aware encoder
- CFW
- 결과

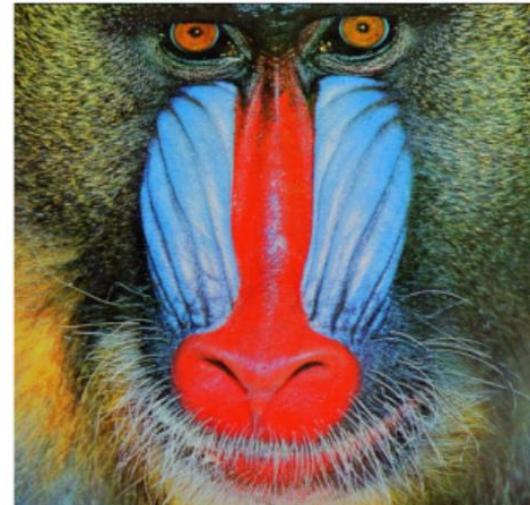
01 Super resolution

Super Resolution (SR) 이란?

- 저화질 이미지를 고화질로 해상도를 높여주는 task
- 입력값 : 저화질 이미지 (Low Resolution, LR)
- 출력값 : 고화질 이미지 (High Resolution, HR)



Low-resolution image



High-resolution image

01 Super resolution

Ill-Posed Problem

- 저해상도를 고해상도로 복원할 때, 타겟이 되는 고해상도 이미지가 정답이 여러 개 존재할 수 있음

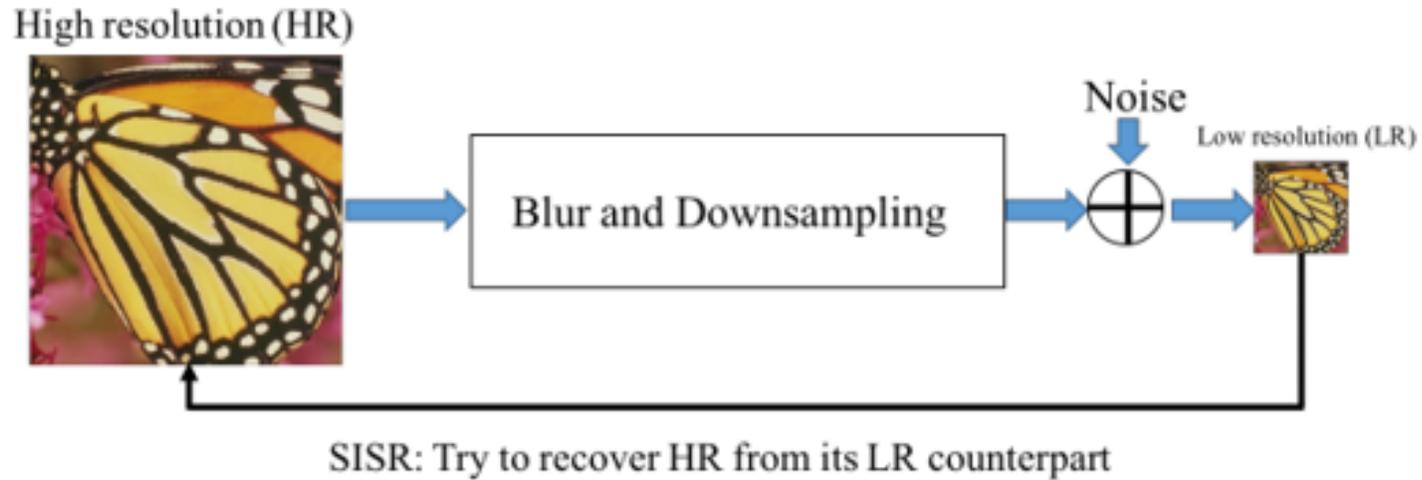
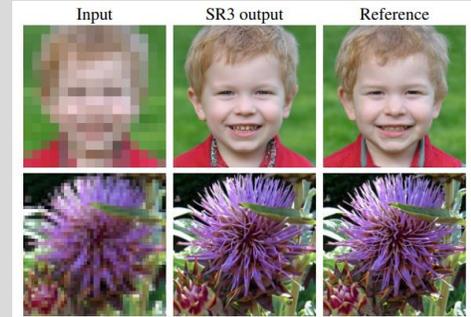




Image Super-Resolution via Iterative Refinement

Chitwan Saharia,[†] Jonathan Ho, William Chan, Tim Salimans, David J. Fleet, Mohammad Norouzi
 {sahariac, jonathanho, williamchan, salimans, davidfleet, mnorouzi}@google.com
 Google Research, Brain Team



연구 배경(CVPR, 2021)

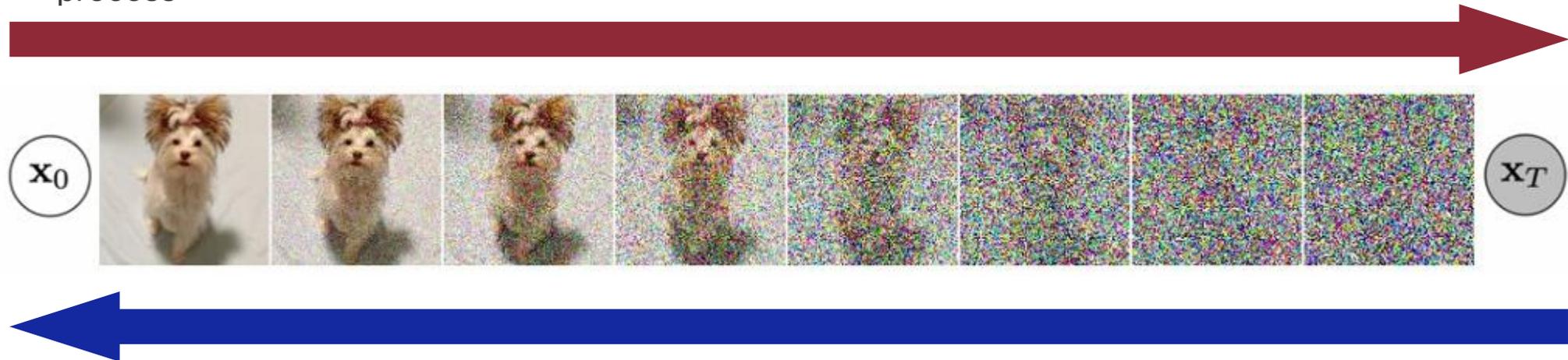
Deep generative model들은 이미지의 복잡한 경험적 분포를 학습하는데 성공하였으나, 다음과 같은 문제가 발생

1. Autoregressive model : 고해상도 이미지 생성에 **많은 비용**이 필요
2. VAE(Variational AutoEncoder), NF(Normalizing Flow) : 종종 **좋지 못한 샘플**이 생성
3. GAN : **최적화 불안정성**과 mode collapse 해결이 필요

02 Diffusion model

Diffusion model의 기본 개념

- Diffusion은 '확산'을 의미
- 어떤 분자의 움직임이 있다고 가정 할 때, 그 움직임은 **Gaussian distribution**을 따르며 **Markov process** 형태로 움직임
- **Forward Process** : 이미지 x_0 가 완전한 gaussian noise x_T 가 될 때까지 gaussian noise를 점진적으로 추가하는 markov process
- **Reverse Process** : Gaussian noise x_T 에서 점진적으로 gaussian noise를 제거하여 이미지 x_0 를 복원하는 markov process





02 Diffusion model

Diffusion model의 기본 개념

- Diffusion은 '확산'을 의미
- 어떤 분자의 움직임이 있다고 가정 할 때, 그 움직임은 **Gaussian dis**
- **Forward Process** : 이미지 X_0 가 완전한 gaussian noise X_T 가 될 process
- **Reverse Process** : Gaussian noise X_T 에서 점진적으로 gau process

종료 Improving Sampling Speed of Diffusion Models
Open DMQA Seminar 2023.02.10
조한성

Improving Sampling Speed of Diffusion Models

발표자: 조한성

2023년 2월 10일
오후 1시 ~
온라인 비디오 시청 (YouTube)

세미나 정보 보기 →

종료 Conditional Diffusion Models
KOREA UNIVERSITY
Jong Hyun Lee 2023.06.09

Conditional Diffusion Models

발표자: 이종현

2023년 6월 16일
오전 12시 ~
온라인 비디오 시청 (YouTube)

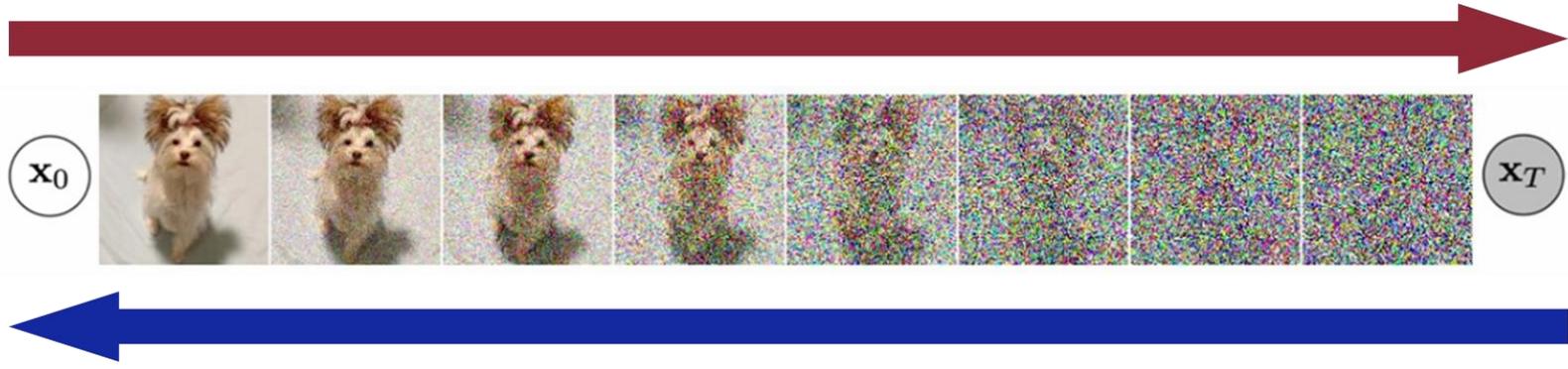
세미나 정보 보기 →



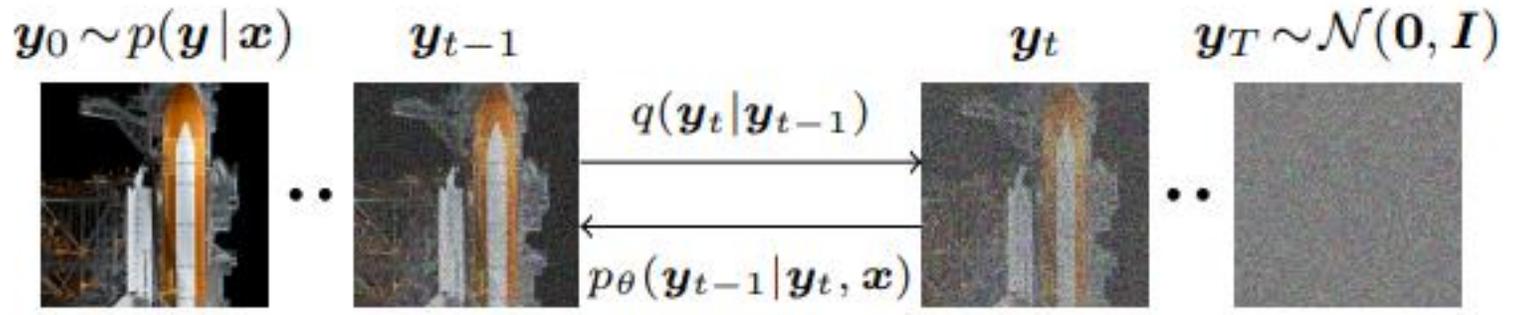


02 방법론

[DDPM]



[SR3]

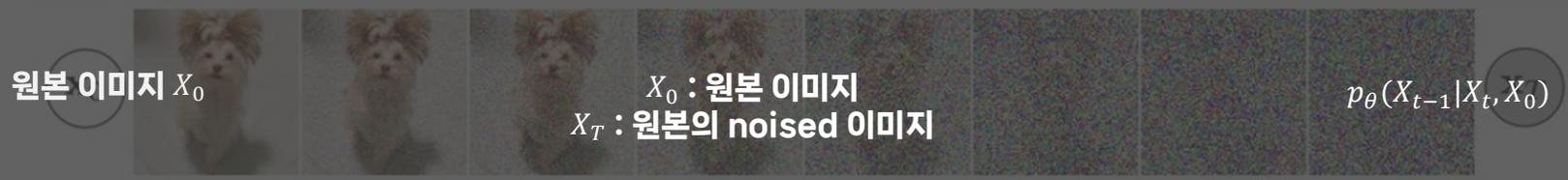




02 방법론

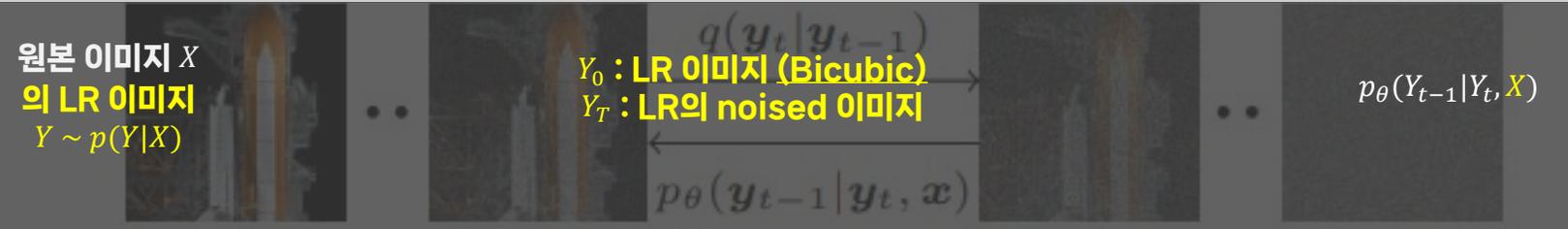
[DDPM]

Input → Diffusion process의 대상 → Inference



[SR3]

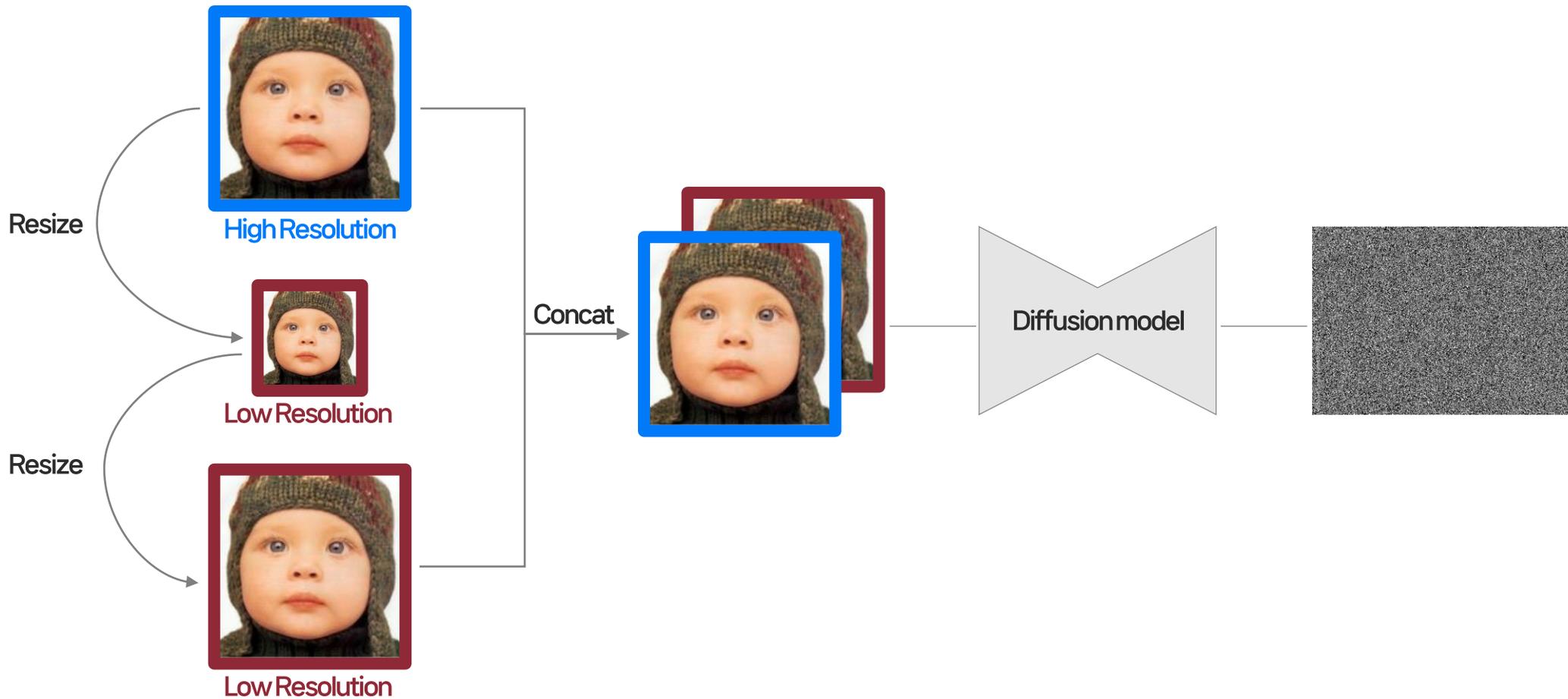
Input $y_t \sim p(y|x)$ → Diffusion process의 대상 y_t → Inference $y_T \sim \mathcal{N}(0, I)$





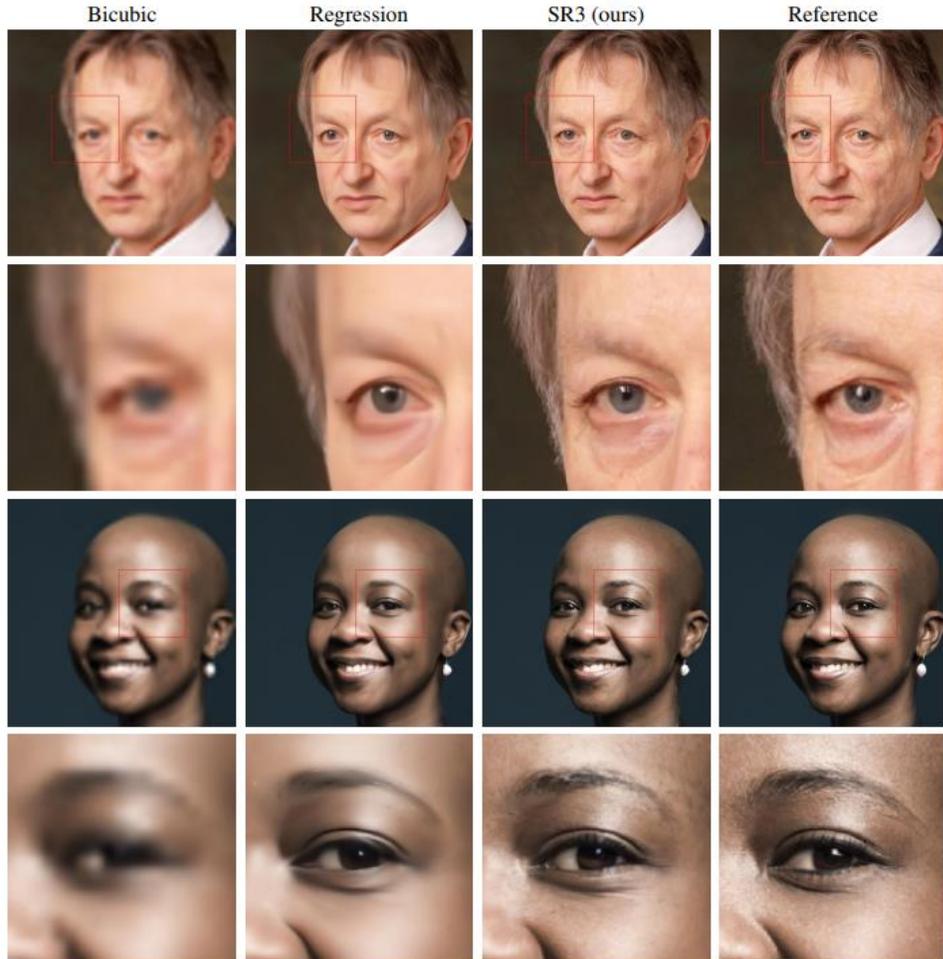
SR3

02 Framework





SR 이미지 생성 결과 – Automated metrics



PSNR / SSIM 결과

Metric	PULSE [28]	FSRGAN [7]	Regression	SR3
PSNR ↑	16.88	23.01	23.96	23.04
SSIM ↑	0.44	0.62	0.69	0.65
Consistency ↓	161.1	33.8	2.71	2.68

FID / IS 결과

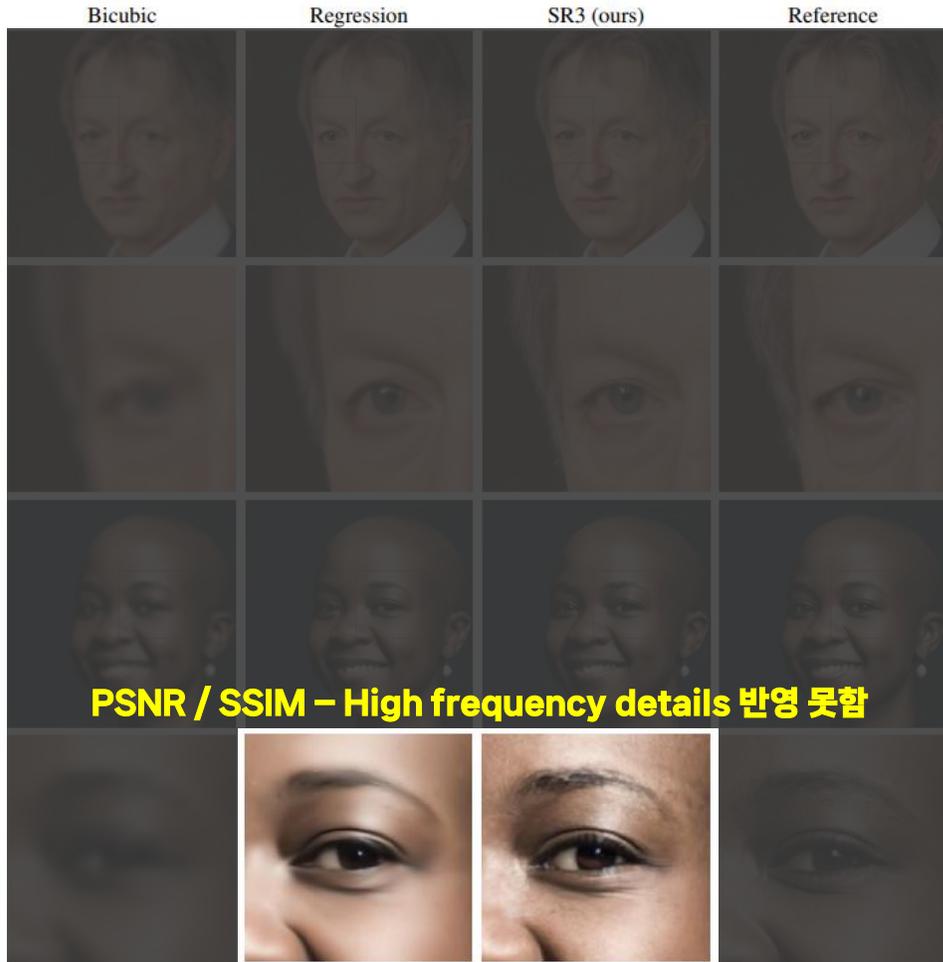
Model	FID ↓	IS ↑	PSNR ↑	SSIM ↑
Reference	1.9	240.8	-	-
Regression	15.2	121.1	27.9	0.801
SR3	5.2	180.1	26.4	0.762

Top-1/5 error

Method	Top-1 Error	Top-5 Error
Baseline	0.252	0.080
DRCN [22]	0.477	0.242
FSRCNN [13]	0.437	0.196
PsyCo [35]	0.454	0.224
ENet-E [44]	0.449	0.214
RCAN [64]	0.393	0.167
Regression	0.383	0.173
SR3	0.317	0.120



SR 이미지 생성 결과 – Automated metrics



• PSNR / SSIM 결과

Metric	PULSE [28]	FSRGAN [7]	Regression	SR3
PSNR ↑	16.88	23.01	23.96	23.04
SSIM ↑	0.44	0.62	0.69	0.65
Consistency ↓	161.1	33.8	2.71	2.68

• FID / IS 결과

Model	FID ↓	IS ↑	PSNR ↑	SSIM ↑
Reference	1.9	240.8	-	-
Regression	15.2	121.1	27.9	0.801
SR3	5.2	180.1	26.4	0.762

• Top-1/5 error

Method	Top-1 Error	Top-5 Error
Baseline	0.252	0.080
DRCN [22]	0.477	0.242
FSRCNN [13]	0.437	0.196
PsyCo [35]	0.454	0.224
ENet-E [44]	0.449	0.214
RCAN [64]	0.393	0.167
Regression	0.383	0.173
SR3	0.317	0.120

SR 이미지 생성 결과 – Human evaluation

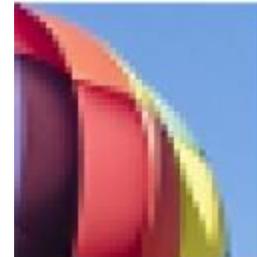
Low Resolution



Output 1
(Ground truth)



Output 2
(Generated SR)



Output 1
(Ground truth)



Output 2
(Generated SR)

Q1. **Output1**과 **Output2** 중에 어떤 이미지가 LR이미지를 더 잘 표현했나요?

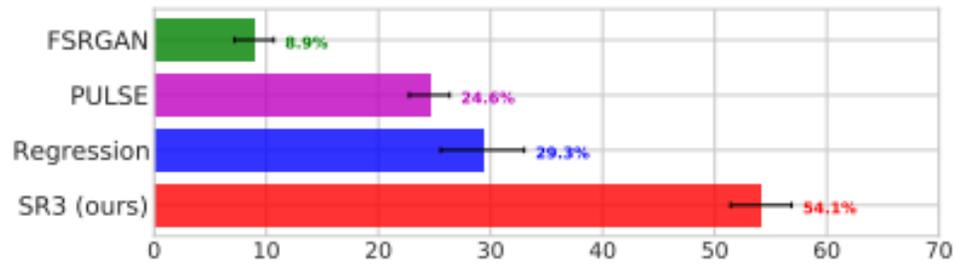
Q2. **Output1**과 **Output2** 중에 어떤 이미지가 진짜 카메라로 찍은 이미지 같은가요?



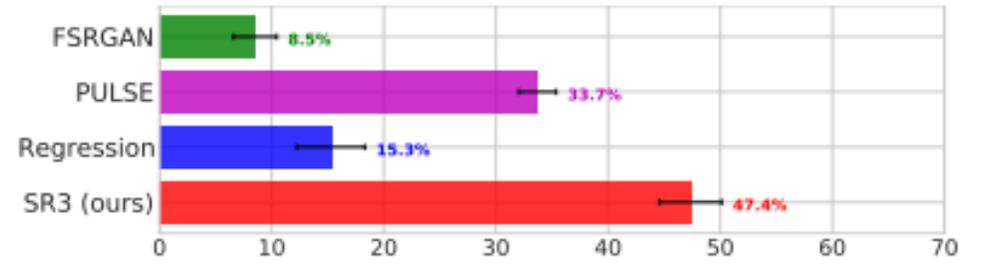
02 SR 이미지 생성 결과 – Human evaluation

Fool rate : Ground truth(정답)가 아닌 model이 생성한 이미지를 선택한 확률

Fool rates (3 sec display w/ inputs, 16 × 16 → 128 × 128)



Fool rates (3 sec display w/o inputs, 16 × 16 → 128 × 128)



Q1. Output1과 Output2 중에 어떤 이미지가 LR이미지를 더 잘 표현했나요?

Q2. Output1과 Output2 중에 어떤 이미지가 진짜 카메라로 찍은 이미지 같은가요?



SR3+

03

논문 소개

Denoising Diffusion Probabilistic Models for Robust Image Super-Resolution in the Wild

Hshmat Sahak^{1,2} Daniel Watson¹ Chitwan Saharia¹ David Fleet¹



연구 배경(CVPR, 2023)

- SR3는 OOD의 경우(예를 들어 알려지지 않은 degradation을 가진 이미지) 잘 작동하지 않음
- Robust한 모델 구축을 위해 downsampling시 SR3보다 더 큰 손상을 줌
- SOTA 달성



SR3+

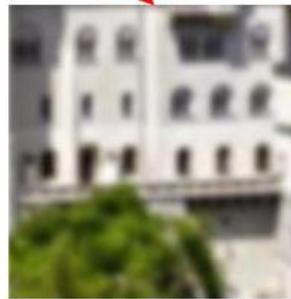
03 DDPM for Robust Image Super-Resolution in the Wild

Blind Super Resolution

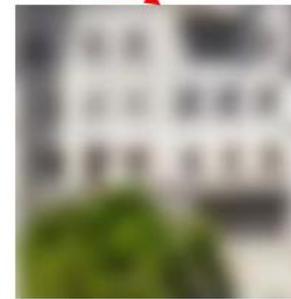
다양한 degradation 기법들로 현실에 있을 법한 LR 이미지를 생성



HR



LR①
Gaussian Blur



LR②
Gaussian Blur
JPEG Compression



LR③
Gaussian Noise



SR3+

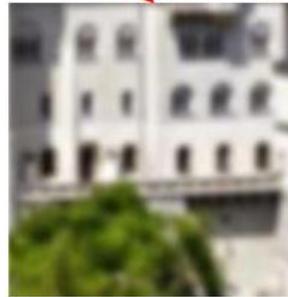
03 DDPM for Robust Image Super-Resolution in the Wild

Blind Super Resolution

다양한 degradation 기법들로 현실에 있을 법한 LR 이미지를 생성



HR



LR①
Gaussian Blur



LR②
Gaussian Bl
JPEG Compre

종료

Blind Super Resolution

2023. 09. 01.
김성수
Data Mining and Quality Analytics

Blind Super Resolution

발표자:  김성수

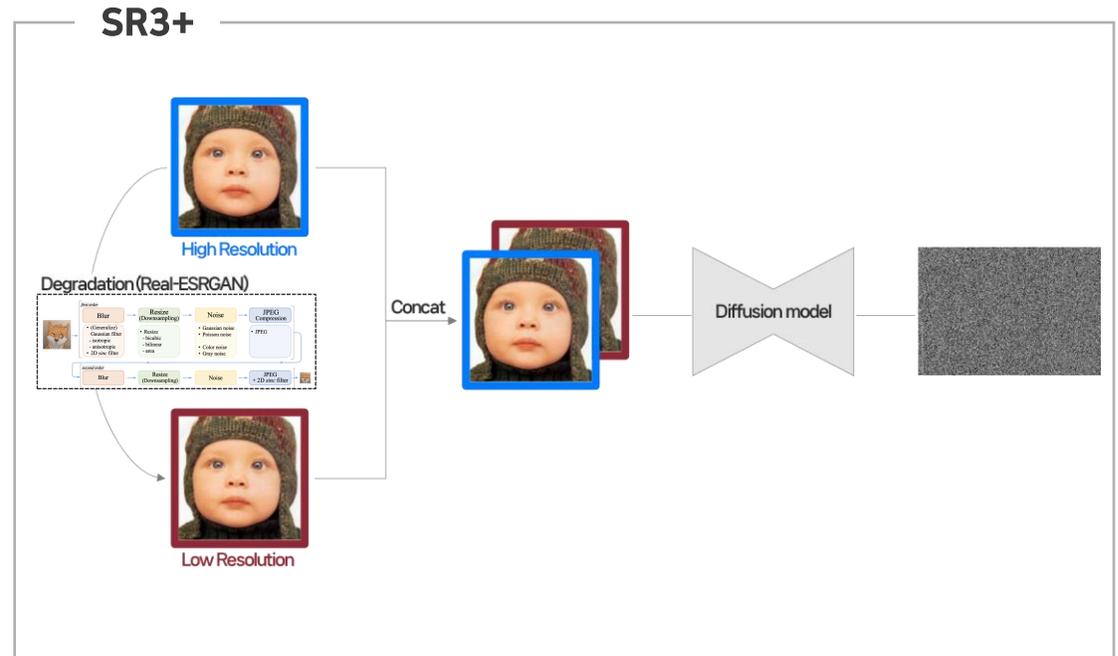
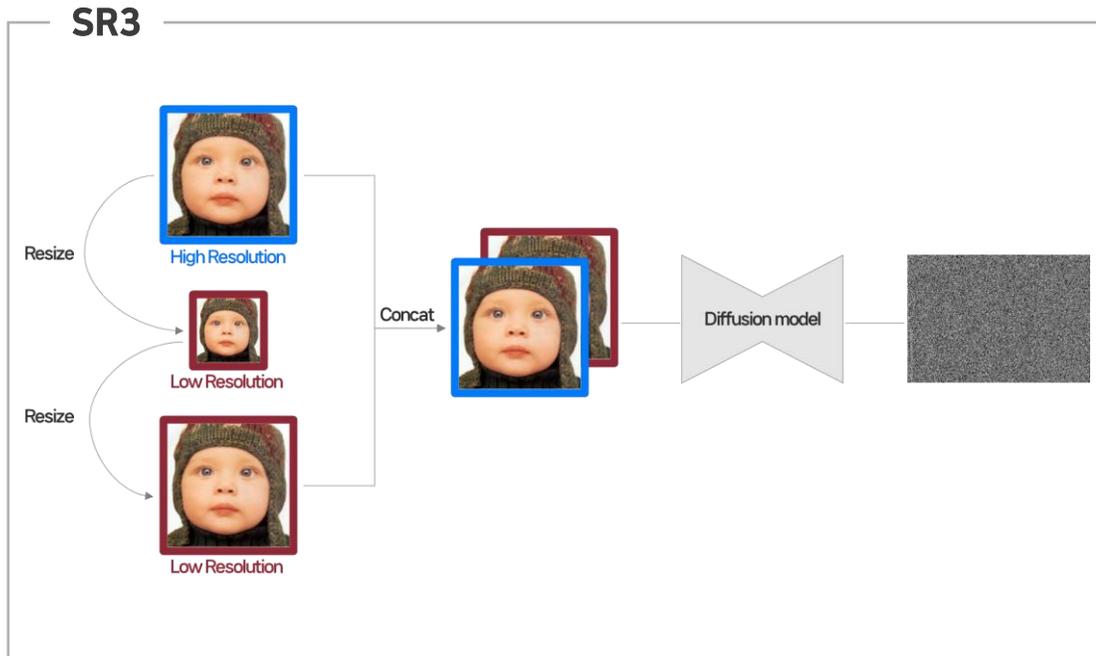
📅 2023년 9월 1일
🕒 오후 1시 ~
▶ 온라인 비디오 시청 (YouTube)

[세미나 정보 보기 →](#)



SR3+

03 SR3+ : Degradation 방법

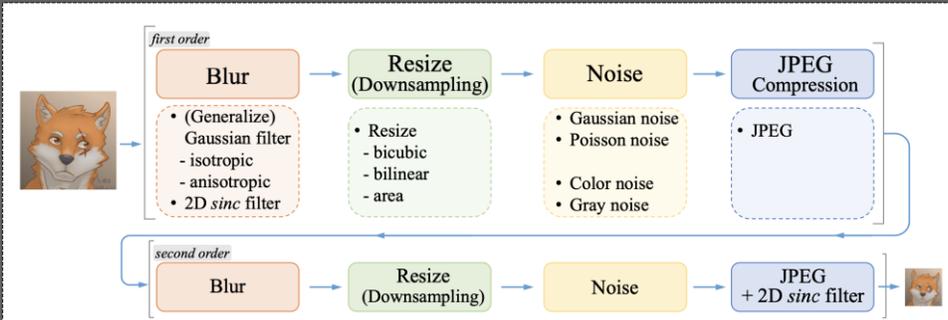




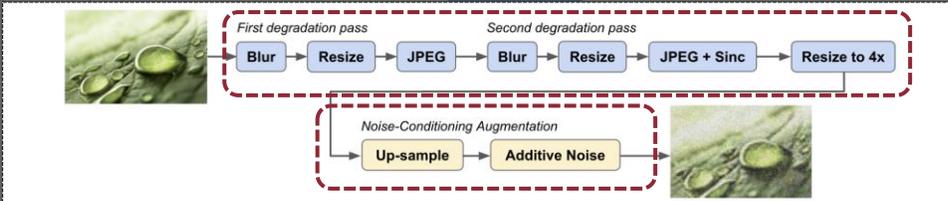
SR3+

03 Degradation 방법

Real-ESRGAN



SR3+



```
def apply_degradation(img, step):
    # 1. 블러
    radiuses = [3, 5, 7, 9, 11]
    kernel_rad = random.choice(radiuses)
    if step == 1:
        img = blur(img, step=step, kernel_rad=kernel_rad)
    else:
        img = blur(img, step=step, kernel_rad=kernel_rad)

    # 2. resizing
    if step == 1:
        scale_range = (0.15, 1.5)
    else:
        scale_range = (0.3, 1.2)
    resize_scale_factor = np.random.uniform(scale_range[0], scale_range[1])
    new_size = (int(img.width * resize_scale_factor), int(img.height * resize_scale_factor))
    # resizing 방법 선택
    methods = [Image.BOX, Image.BICUBIC, Image.BILINEAR]
    method = random.choice(methods)
    # random 하게 resize
```

Real-ESRGAN의 degradation과 흐름은 유사하지만, 중간 과정에서의 세부 조건 및 수치에 차이가 있음

```
# 3. JPEG 아티팩트
if step == 1:
    img = jpeg_compression(img)
else:
    img = sinc_filter(img, kernel_rad)
    img = jpeg_compression(img)
    img = sinc_filter(img, kernel_rad)

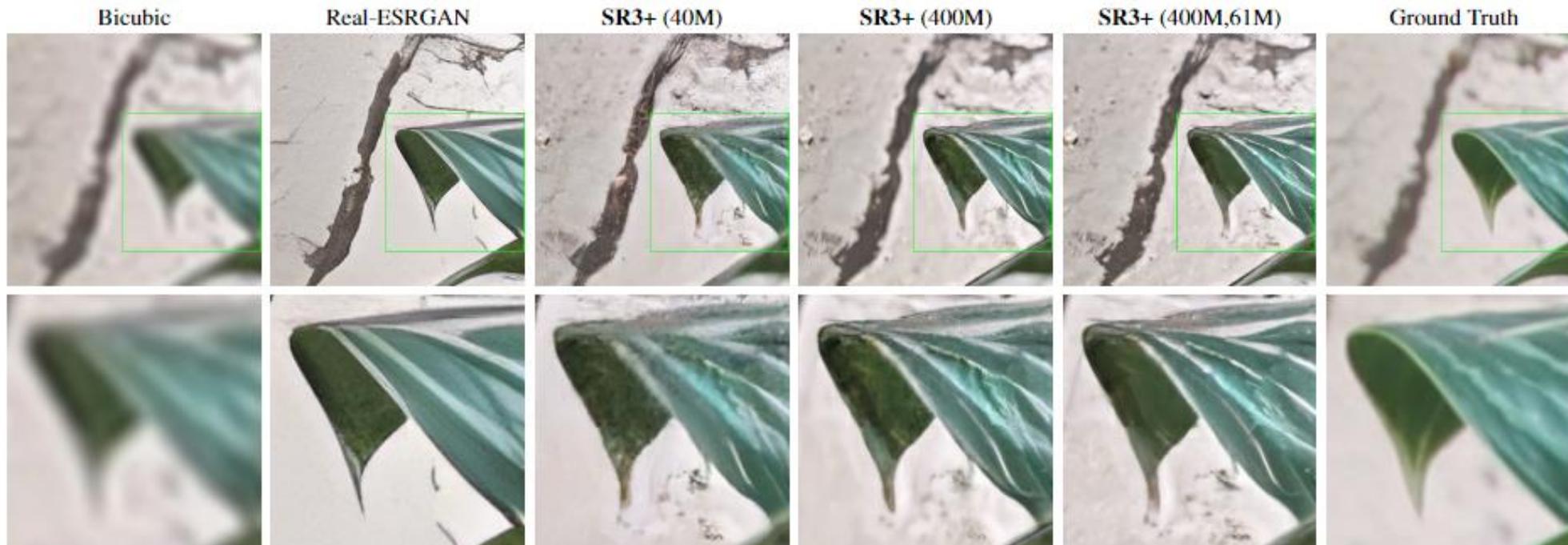
return img
```



SR3+

03

SR 이미지 생성 결과



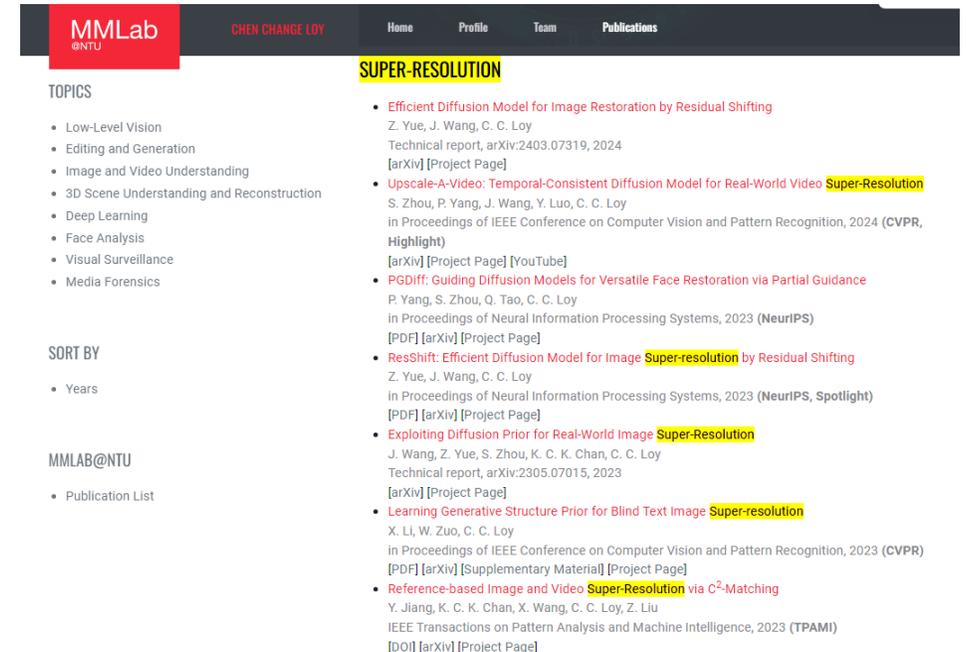
SR Model (Parameter Count, Dataset)	FID(10k) ↓		PSNR ↑		SSIM ↑	
	RealSR	DRealSR	RealSR	DRealSR	RealSR	DRealSR
Real-ESRGAN	34.21	37.22	25.14	25.85	0.7279	0.7808
SR3+ (40M, DF2K + OST)	<u>31.97</u>	40.26	24.84	25.18	0.6827	0.7201
SR3+ (400M, DF2K + OST)	<u>27.34</u>	<u>36.28</u>	23.84	24.36	0.662	0.719
SR3+ (400M, 61M Dataset)	24.32	32.37	24.89	25.74	0.6922	0.7547

Exploiting Diffusion Prior for Real-World Image Super-Resolution

Jianyi Wang · Zongsheng Yue · Shangchen Zhou · Kelvin C.K. Chan ·
Chen Change Loy

연구 배경(arXiv, 2023)

- 사전 학습된 text-to-image diffusion model에 포함된 prior knowledge를 이용한 새로운 super resolution 접근 방식
- Time-aware encode라는 fine-tune과 몇 개의 feature를 조절하여 SR task 수행
- SOTA 달성



MMLab@NTU

CHEN CHANGE LOY

Home Profile Team Publications

SUPER-RESOLUTION

TOPICS

- Low-Level Vision
- Editing and Generation
- Image and Video Understanding
- 3D Scene Understanding and Reconstruction
- Deep Learning
- Face Analysis
- Visual Surveillance
- Media Forensics

SORT BY

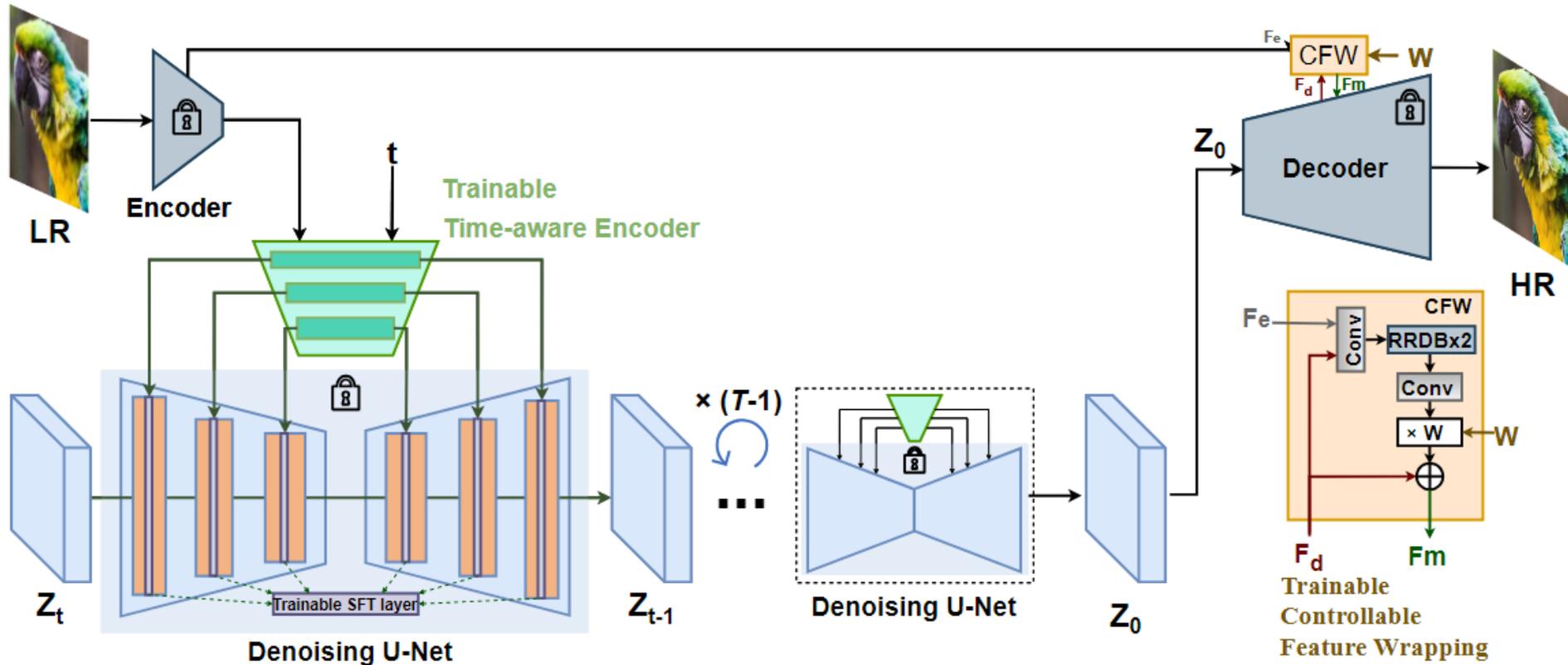
- Years

MMLAB@NTU

- Publication List

Publications

- **Efficient Diffusion Model for Image Restoration by Residual Shifting**
Z. Yue, J. Wang, C. C. Loy
Technical report, arXiv:2403.07319, 2024
[arXiv] [Project Page]
- **Upscale-A-Video: Temporal-Consistent Diffusion Model for Real-World Video Super-Resolution**
S. Zhou, P. Yang, J. Wang, Y. Luo, C. C. Loy
in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2024 (CVPR, Highlight)
[arXiv] [Project Page] [YouTube]
- **PGDiff: Guiding Diffusion Models for Versatile Face Restoration via Partial Guidance**
P. Yang, S. Zhou, Q. Tao, C. C. Loy
in Proceedings of Neural Information Processing Systems, 2023 (NeurIPS)
[PDF] [arXiv] [Project Page]
- **ResShift: Efficient Diffusion Model for Image Super-resolution by Residual Shifting**
Z. Yue, J. Wang, C. C. Loy
in Proceedings of Neural Information Processing Systems, 2023 (NeurIPS, Spotlight)
[PDF] [arXiv] [Project Page]
- **Exploiting Diffusion Prior for Real-World Image Super-Resolution**
J. Wang, Z. Yue, S. Zhou, K. C. K. Chan, C. C. Loy
Technical report, arXiv:2305.07015, 2023
[arXiv] [Project Page]
- **Learning Generative Structure Prior for Blind Text Image Super-resolution**
X. Li, W. Zuo, C. C. Loy
in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2023 (CVPR)
[PDF] [arXiv] [Supplementary Material] [Project Page]
- **Reference-based Image and Video Super-Resolution via C^2 -Matching**
Y. Jiang, K. C. K. Chan, X. Wang, C. C. Loy, Z. Liu
IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023 (TPAMI)
[DOI] [arXiv] [Project Page]





04 Time aware encoder

제약조건

1. 결과 모델은 LR input을 조건으로 받아 그럴듯한 HR 이미지를 생성함 (LR은 유일한 구조적 정보의 원천)
2. Stable diffusion의 prior를 보존하기 위해 최소한의 tuning

Feature Modulation

- SFT(Spatial Feature Transformation)을 통해 feature map 조절
- LR 이미지로부터 추출한 구조적 정보를 삽입하게 해줌

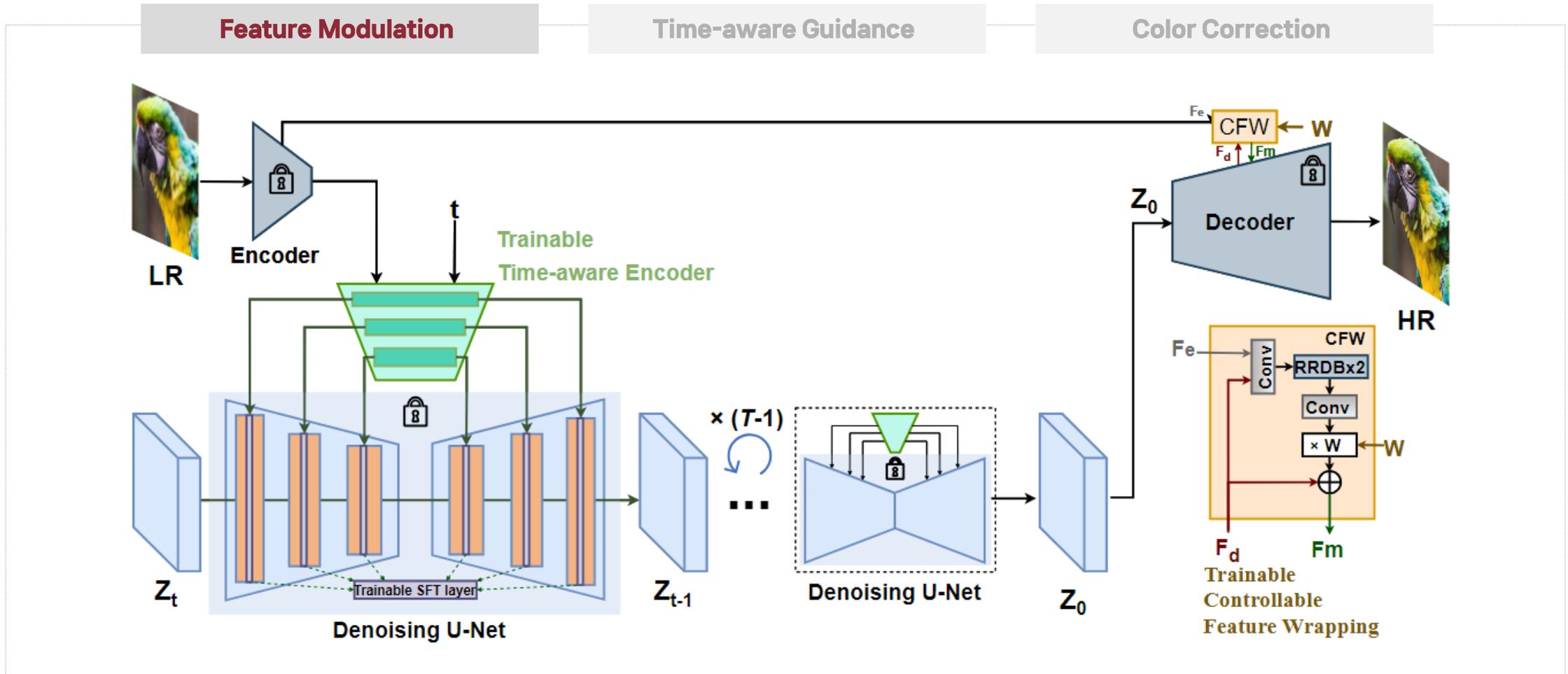
Time-aware Guidance

- Time-embedding 레이어 덕분에 LR feature로부터 추출한 condition을 잘 조절할 수 있음
- 이미지 품질과 fidelity 향상에 도움을 주며, SNR로 평가

Color Correction

- Diffusion model은 종종 color shift 문제가 발생
- Color normalization과 Wavelet 분해를 통해 해결

04 Time aware encoder



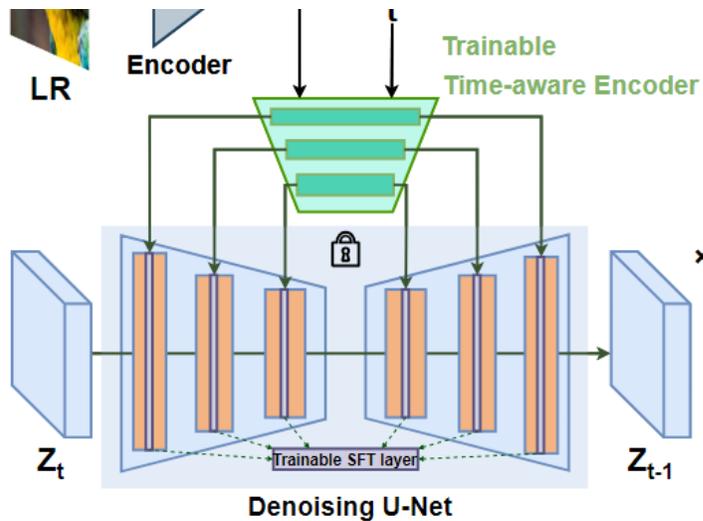
04 Time aware encoder

Feature Modulation

Time-aware Guidance

Color Correction

Wang, X. et al., (CVPR, 2018) Recovering realistic texture in image super-resolution by deep spatial feature transform



- LR 이미지로부터 multi scale feature들을 추출
- 해당 Feature들을 SFT(Spatial Feature Transformation)을 통해 stable diffusion의 residual block feature map을 조절하는데 활용

$$\hat{F}_{dif}^n = (1 + \alpha^n) \odot F_{dif}^n + \beta^n; \quad \alpha^n, \beta^n = \mathcal{M}_\theta(F^n),$$

- α, β 는 affine 파라미터이고, \mathcal{M}_θ 는 Conv layer
 - \hat{F}_{dif}^n 를 통해서 feature map 조절
- Feature map 튜닝이 이루어지는 동안 stable diffusion의 weight는 고정, 인코더와 SFT layer만 학습
- LR 이미지로부터 추출한 구조적 정보를 삽입하게 해줌

Time aware encoder

Feature Modulation

Time-aware Guidance

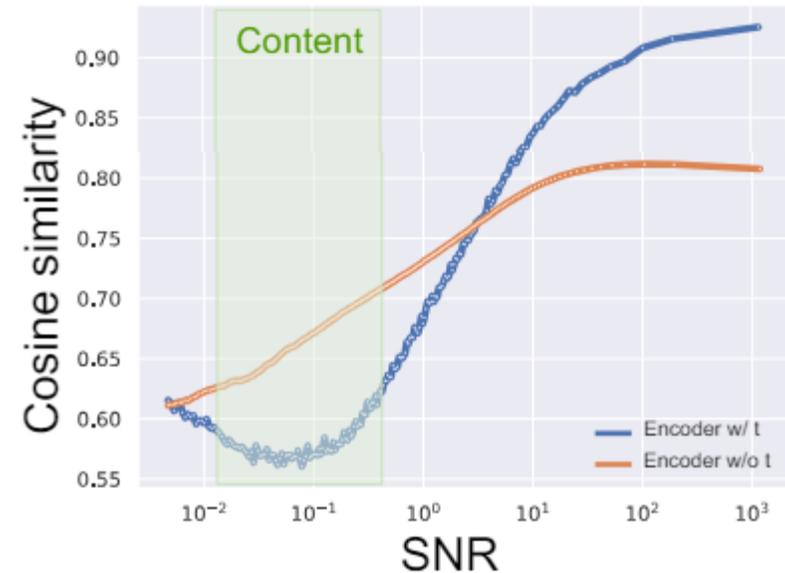
Color Correction

Choi, J. et al., (CVPR, 2022) Perception prioritized training of diffusion models

- 인코더의 time embedding layer 덕분에 LR feature로부터 추출한 condition을 잘 조절하는 것은 이미지 품질과 fidelity 향상에 도움 → SNR로 평가

$$SNR = \frac{P_s}{P_n} : \text{신호(signal)의 전력과 노이즈(noise)의 전력}$$

- 기존 연구에 따라 SNR이 0.05일 때, SR performance가 좋음
- 이를 입증하기 위해 stable diffusion에서 SFT 이전의 feature들과 SFT 이후의 feature들 간의 condition strength를 측정하여 코사인 유사도 계산
 - 코사인 유사도가 낮을수록 인코더에서 제공하는 특성이 diffusion model의 원래 특성과 다르다는 것을 의미
 - 인코더 특성이 모델에 강력하고 명확한 조건을 제공





04 Time aware encoder

Feature Modulation

Time-aware Guidance

Color Correction

Choi, J. et al., (CVPR, 2022) Perception prioritized training of diffusion models

< Color shift 문제 발생 >



Color normalization

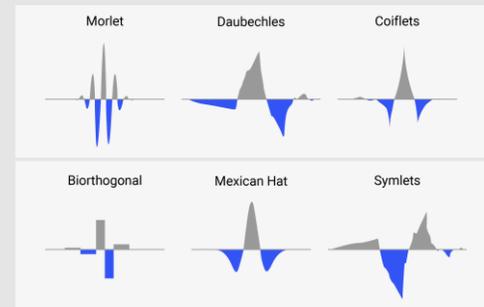
- LR에서 추출한 평균과 표준편차를 이용하여 생성된 HR값의 색상을 보정하는 작업
- x 는 LR의 input이고 \hat{y} 는 생성된 HR 이미지, c 는 RGB채널

$$y^c = \frac{\hat{y}^c - \mu_{\hat{y}}^c}{\sigma_{\hat{y}}^c} \cdot \sigma_x^c + \mu_x^c$$

Wavelet-based color correction

- Wavelet 분해를 통해 고주파 성분(H)과 저주파 성분(L)을 따로 추출하여 색상 보정

$$L^i = C_i(L^{i-1}, k), H^i = L^{i-1} - L^i$$



<https://tech.onepredict.ai/94d98ece-06be-4215-b5ef-87a58ab8d2e3>



04 Time aware encoder

Feature Modulation

Time-aware Guidance

Color Correction

Choi, J. et al., (CVPR, 2022) Perception prioritized training of diffusion models

< Color shift 문제 발생 >



Color normalization

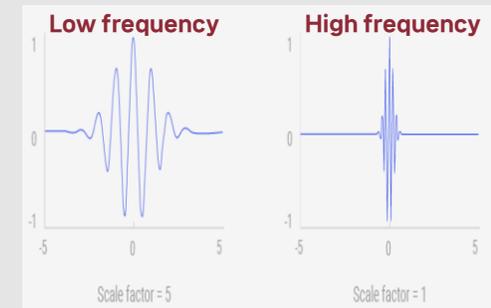
- LR에서 추출한 평균과 표준편차를 이용하여 생성된 HR값의 색상을 보정하는 작업
- x 는 LR의 input이고 \hat{y} 은 생성된 HR 이미지, c 는 RGB채널

$$y^c = \frac{\hat{y}^c - \mu_{\hat{y}}^c}{\sigma_{\hat{y}}^c} \cdot \sigma_x^c + \mu_x^c$$

Wavelet-based color correction

$$L^i = C_i(L^{i-1}, k), H^i = L^{i-1} - L^i$$

$$k = \begin{bmatrix} 1/16 & 1/8 & 1/16 \\ 1/8 & 1/4 & 1/8 \\ 1/16 & 1/8 & 1/16 \end{bmatrix}$$



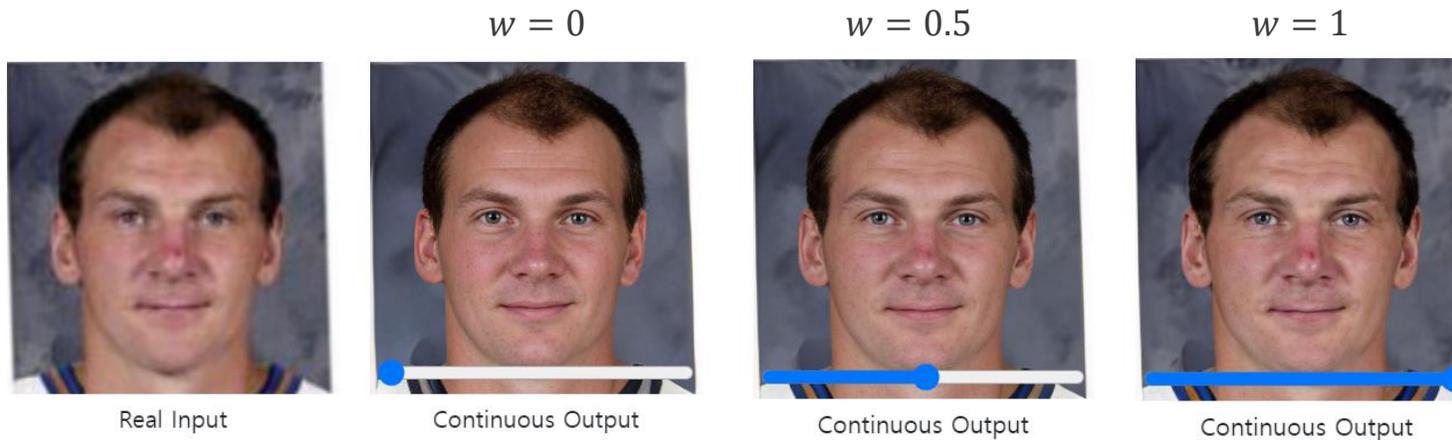
<https://tech.onepredict.ai/94d98ece-06be-4215-b5ef-87a58ab8d2e3>

Controllable Feature Wrapping(CFW)

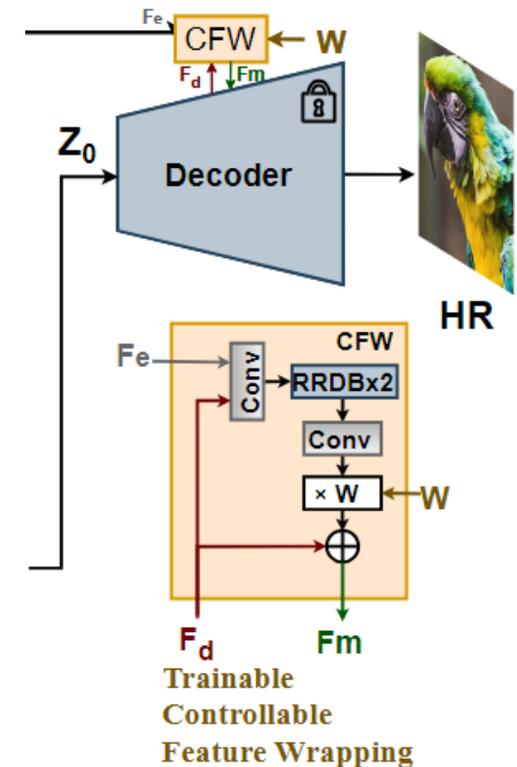
Zhou, S. et al., (NeurIPS, 2022) Towards robust blind face restoration with codebook lookup transformer

- Fidelity와 Realism 사이 균형을 위해 CFW 모듈 도입
 - Fidelity : 원본이나 실제 대상을 얼마나 정확하게 재현하는가, 즉 ground truth와 얼마나 일치하는가
 - Realism : 현실감을 얼마나 반영하고 있는가
- Stable diffusion이 latent space상에서 구현되므로, 인코더 feature를 통해 디코더 feature를 조절

$$F_m = F_d + C(F_e, F_d; \theta) \times w$$

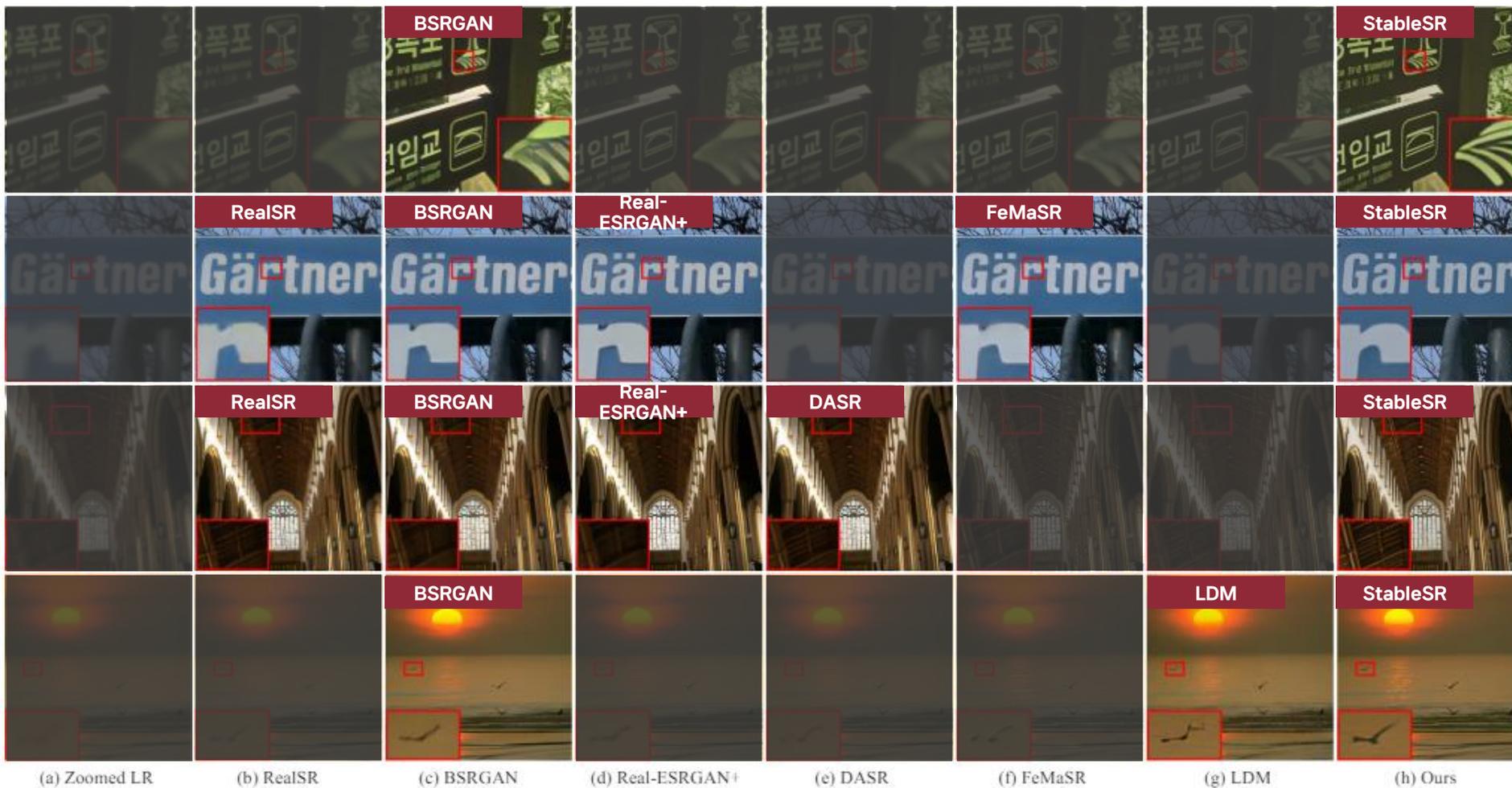


가장 좋은 성능 = 모델 채택





SR 이미지 생성 결과 - 정량평가



SR 이미지 생성 결과 - 정량평가

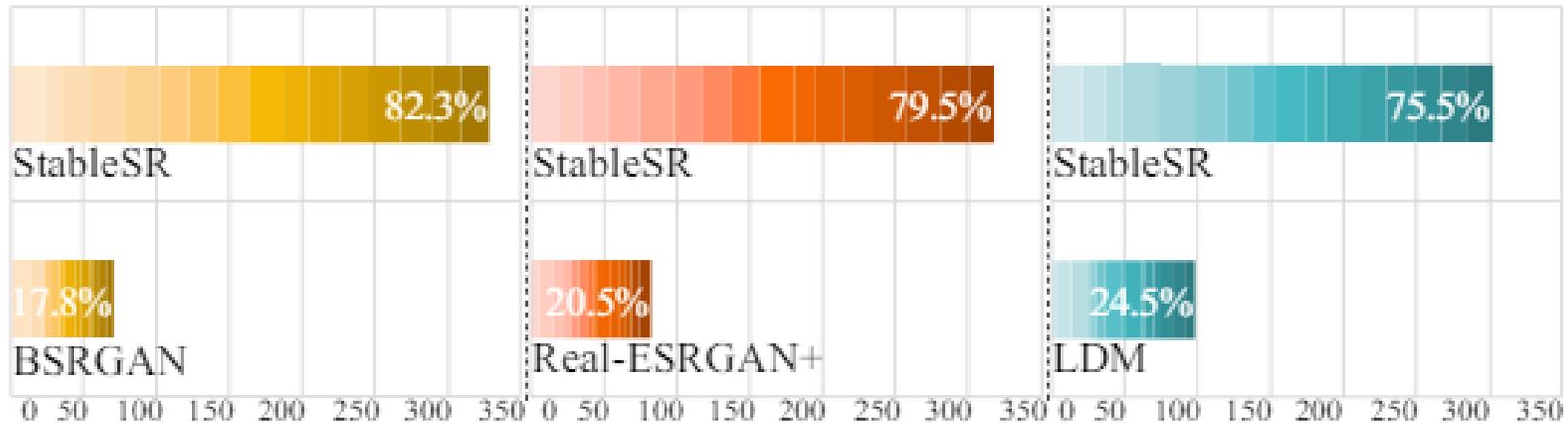


Red : Best performance, **Blue** : Second best performance

Datasets	Metrics	RealSR	BSRGAN	DASR	Real-ESRGAN+	FeMaSR	LDM	SwinIR-GAN	IF_III	StableSR
DIV2K Valid	PSNR ↑	24.62	<u>24.58</u>	24.47	24.29	23.06	23.32	23.93	23.36	23.26
	SSIM ↑	0.5970	0.6269	0.6304	0.6372	0.5887	0.5762	<u>0.6285</u>	0.5636	0.5726
	LPIPS ↓	0.5276	0.3351	0.3543	0.3112	0.3126	0.3199	0.3160	0.4641	<u>0.3114</u>
	FID ↓	49.49	44.22	49.16	37.64	35.87	<u>26.47</u>	36.34	37.54	24.44
	CLIP-IQA ↑	0.3534	0.5246	0.5036	0.5276	0.5998	<u>0.6245</u>	0.5338	0.3980	0.6771
	MUSIQ ↑	28.57	61.19	55.19	61.05	60.83	<u>62.27</u>	60.22	43.71	65.92
RealSR	PSNR ↑	27.30	26.38	<u>27.02</u>	25.69	25.06	25.46	26.31	25.47	24.65
	SSIM ↑	0.7579	0.7651	<u>0.7707</u>	0.7614	0.7356	0.7145	0.7729	0.7067	0.7080
	LPIPS ↓	0.3570	<u>0.2656</u>	0.3134	0.2709	0.2937	0.3159	0.2539	0.3462	0.3002
	CLIP-IQA ↑	0.3687	0.5114	0.3198	0.4495	0.5406	<u>0.5688</u>	0.4360	0.3482	0.6234
	MUSIQ ↑	38.26	<u>63.28</u>	41.21	60.36	59.06	58.90	58.70	41.71	65.88
DRealSR	PSNR ↑	30.19	28.70	<u>29.75</u>	28.62	26.87	27.88	28.50	28.66	28.03
	SSIM ↑	<u>0.8148</u>	0.8028	0.8262	0.8052	0.7569	0.7448	0.8043	0.7860	0.7536
	LPIPS ↓	0.3938	0.2858	0.3099	<u>0.2818</u>	0.3157	0.3379	0.2743	0.3853	0.3284
	CLIP-IQA ↑	0.3744	0.5091	0.3813	0.4515	0.5634	<u>0.5756</u>	0.4447	0.2925	0.6357
	MUSIQ ↑	26.93	<u>57.16</u>	42.41	54.26	53.71	53.72	52.74	30.71	58.51
DPED-iphone	CLIP-IQA ↑	0.4496	0.4021	0.2826	0.3389	0.5306	0.4482	0.3373	0.2962	<u>0.4799</u>
	MUSIQ ↑	45.60	45.89	32.68	42.42	<u>49.95</u>	44.23	43.30	37.49	50.48



SR 이미지 생성 결과 - User study



Summary

- SR3는 Diffusion model을 통해 SR 문제를 해결한 최초의 연구
- SR3+는 Blind SR을 적용하여 SR3보다 robust한 모델을 구축
- SR3와 SR3+는 시각적으로는 훌륭해 보이나, 전통적인 SR 평가 지표인 PSNR과 SSIM에서는 아쉬운 결과를 보임
- StableSR은 사전 학습된 stable diffusion 모델을 사용하여 최고 성능을 달성
- Diffusion model의 초기 형태인 DDPM에서 벗어나 stable diffusion 기반 모델이 생성되었으며, 평가 관점에서는 여전히 정성적 평가를 참조

고맙습니다