



# DiffBIR

Towards Blind Image Restoration with Generative Diffusion Prior

## 출처

<https://arxiv.org/pdf/2308.15070v1.pdf>

## Abstract

text-to-image diffusion model을 이용한 이미지 복구 문제 해결 모델인 DiffBIR을 소개합니다. 이 모델은 two-stage로 이루어져 있으며 첫 번째 stage는 다양한 낮은 품질의 데이터를 개선시키는 복구 모듈을 미리 학습시킵니다. 그리고 두 번째 stage는 diffusion 모델을 이용해 현실적인 이미지를 생성해냅니다. 특히, 우리는 LAControlNet을 전이학습에 모듈로 사용하au stable diffusion은 생성 능력을 유지합니다. 마침내, 우리는 퀄리티와 재현성을 균형있게 조정가능한 모듈을 소개합니다. 많은 super-resolution 실험에서 우리는 좋은 성능을 나타냄을 입증했습니다.

## Introduction

이미지 복구는 낮은 품질에서 높은 품질의 이미지를 재구축하는데 목적이 있습니다. 전형적으로 denoising, deblurring, super-resolution과 같은 작업으로 이루어집니다. 이전에 많은 복구 알고리즘이 있었지만, 생성에는 제약이 많았습니다. 현실 세계에서의 저품질의 이미지 문제로 인해 blind image restoration(BIR)의 경우 많은 주목을 받았습니다. BIR은 저품질을 자연스러운 고품질 이미지로 변경하는 것이 목표입니다. 그리고 특정한 이미지가 아닌 많은 산업에 적용되어집니다.

BIR의 연구는 크게 3가지로 나누어집니다. 각각의 연구는 많은 성과가 있었지만 제약도 존재합니다. BSR은 현실 이미지 super-resolution문제를 다루며 알 수 없는 낮은 품질의 이미지를 고해상도로 바꿉니다. 가장 유명한 모델은 BSRGAN과 REAL-ESRGAN이 존재합니다. 여러 저품질 문제를 해결하기 위해서는 degradation 섞기 방식과 높은 수준의 degradation 모델링이 개별적으로 요구되며 이들은 gan loss를 통해 end to end 방식을 가집니다. 이들은 이미지에서 대부분의 degradation을 없애지만 자연스러운 디테일을 생성하지 못합니다. 게다가 degradation세팅이 X4, X8의 super-resolution수준으로 제약을 가집니다.

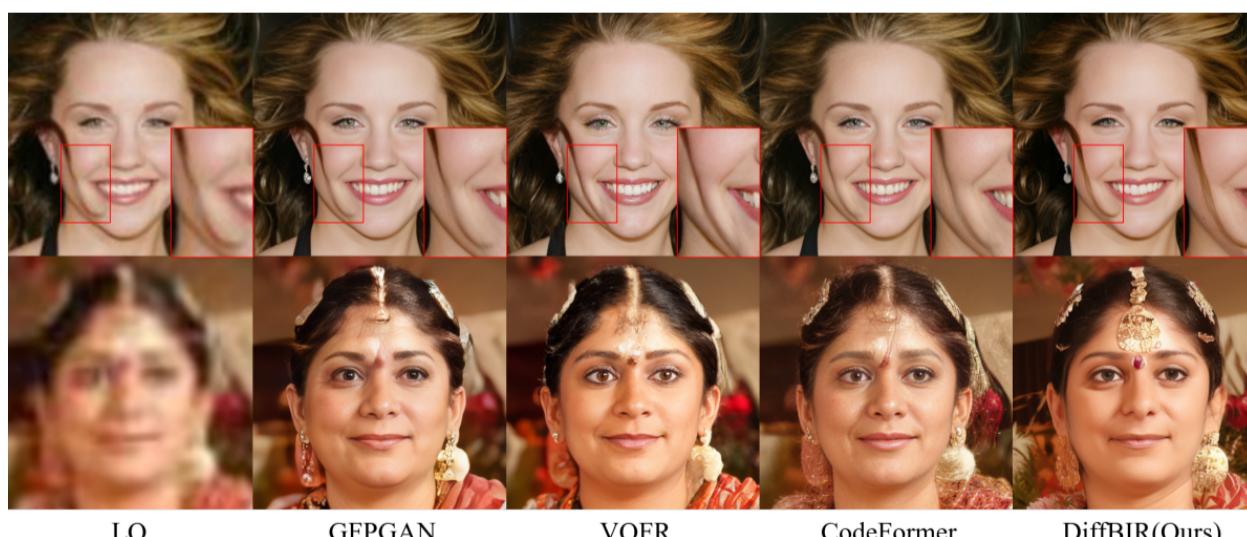
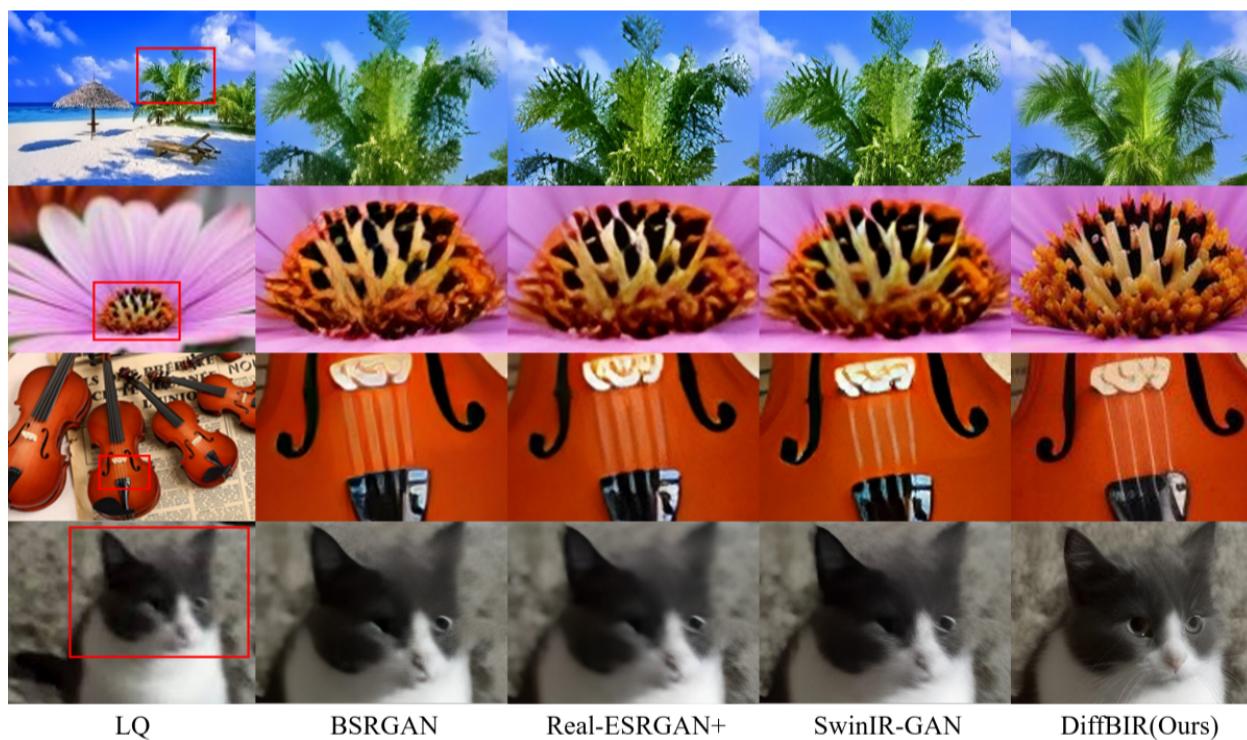
두 번째 방식은 ZIR은 새롭게 등장한 연구로 DDRM, DDNM 그리고 GDP로 이뤄집니다. 사전에 강력한 diffusion 모델을 수반하며 GAN 기반 모델보다 강력한 생성 능력을 지닙니다. degradation 사전 가정을 가지고 인상적인 zero-shot 복구를 할 수 있습니다. 그러나 알고 있는 degradation만을 다룰 수 있다는 단점이 존재합니다. 다시 말하자면, 일반적인 degradation image는 복구 가능하나 알 수 없는 degradation이 있을 경우 잘 해내지 못합니다.

세 번째 방식인 BFR은 사람 얼굴 복구에 초점을 맞췄습니다. 최고의 방식은 CodeFormer와 VQFR입니다. BSR 방식은 유사한 파이프라인을 가지지만, degradation 모델과 생성 네트워크에서 다릅니다. 이미지 공간이 작기 때문에 VQGAN과 TRANSFORMER를 활용할 수 있어 좋은 성능을 냅니다. 그러나 이는 사람 얼굴이라는 도메인 한정적입니다. BFR은 대개 이미지 사이즈에서 제약이 존재하며 일반적인 이미지에 적용이 어렵습니다. 위와 같은 전체적인 분석을 통해 BIR 방식은 전반적인 degradation에서의 일반적인 이미지의 자연스러운 복구가 불가능한 것을 알 수 있습니다. 그러므로 우리는 새로운 BIR 방식을 제시하여 이러한 제약을 극복해냅니다.

이 논문에서는 우리는 이전 연구들의 이점을 살려 합친 프레임 워크를 사용합니다. 특히 real-word degradation을 생성하는 확장된 degradation 모델을 사용합니다. 그리고 잘 학습되어진 stable diffusion을 생성 능력을 키우기 위하여 사용합니다. 마지막으로 two-stage 방식을 사용해 현실성과 재현성을 보장합니다.

우리는 일반화 능력을 키우기 위해서 BSR에서의 다양한 degradation과 BFR의 넓은 degradation 범위를 결합했습니다. 두 번째로 Stable Diffusion을 사용하기 위해 sub-network로 LAControlNet을 사용해 우리의 테스크에 최적화시키도록 하였습니다. ZIR과 유사하게, 사전 학습된 Stable diffusion의 경우 생성 능력을 유지하기 위해 전이학습간 고정되어집니다. 세 번째로 이미지 복구의 현실성과 신뢰성을 주기 위해서 우리는 Restoration Module을 첫 번째로 적용시켜 대부분의 degradation을 제거합니다. 그리고 Generation Module을 사용해 추가 튜닝을 진행시켜 새로운 텍스쳐를 생성합니다. 이러한 파이프라인 없으면, 모델은 너무나도 스무스한(비현실적인) 결과를 내보내게 될 것입니다. 게다가, 사람들의 다양한 요구를 만족시키기 위해 우리는 1stage와 2stage의 영향력 controllable module을 제시합니다. 이는 재학습 과정 없이 latent image guidance를 제시하여 가능했습니다. 기울기 영향은 latent 이미지 거리가 현실성과 재현성 사이에서의 trade off 관계로 구현되어집니다.

위와 같은 요소로 인해, DiffBIR은 BSR과 BFR에서 좋은 성능을 보였습니다. BSR, BFR은 DiffBIR과는 몇몇 측면에서 다른 양상을 보입니다. 복잡한 이미지에서 BSR의 경우 비현실적인 부분을 생성해내지만, DiffBIR은 시각적으로 뛰어난 결과물을 생성해냅니다. DiffBIR은 디테일적 부분을 잘해내는데 그 이유는 BSR에서는 디테일적인 것을 지우게 되는데 DiffBIR은 오히려 그러한 구조를 살리고자 합니다. 게다가, 극심한 degradation을 복구해낼 수 있으며 현실적인 이미지를 생성해냅니다. 얼굴 복원에 관해서는 DiffBIR은 다른 부분에 의해 흐릿한 얼굴 영역에 대해서 좋은 재현 결과를 보여줍니다. 결론적으로 DiffBIR은 BSR과 BFR에서 처음으로 동시에 좋은 프레임워크를 지닙니다.



## Related Work

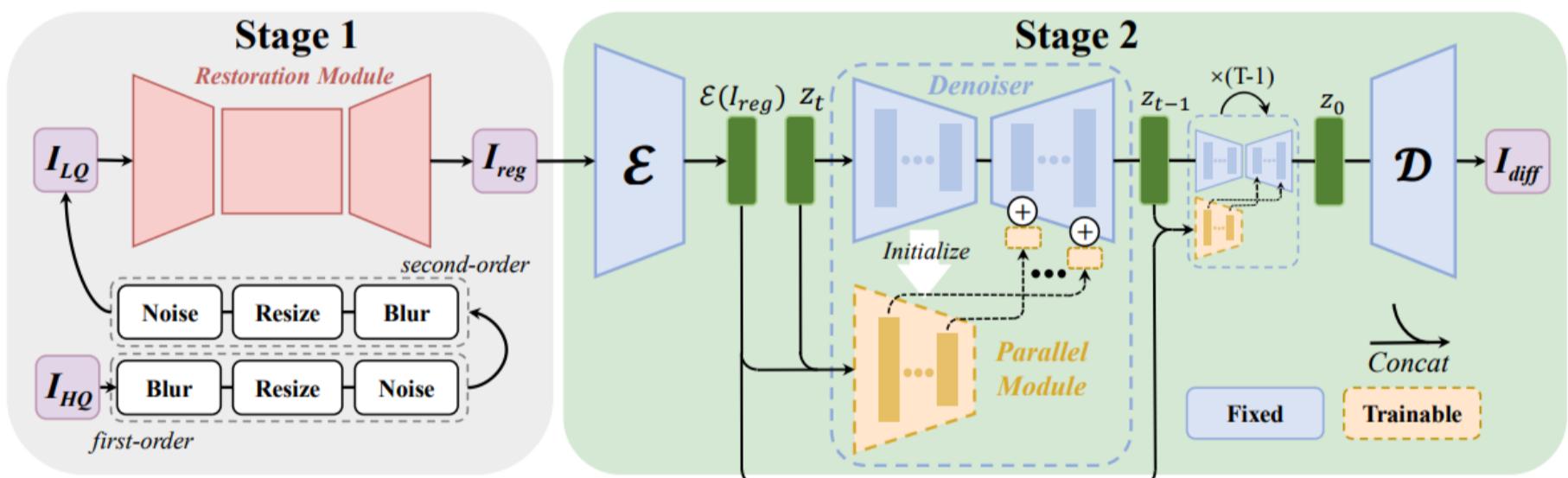
**Blind Image Super-Resolution.** BSR에서의 정교한 현실에서 degradation을 해결하고자 합니다. 특히 BSRGAN은 랜덤 샘플 전략을 기반으로 실용적인 degradation 합성 방식을 하고자 했으며, Real-ESRGAN은 고차원의 degradation model을 사용했습니다. 둘다 GANs를 활용했습니다. SwinIR-GAN은 Swin Transformer를 사용했습니다. FeMaSR은 사전 학습된 VQ-GAN을 기반으로 SR을 피쳐 매칭 문제로 사용했습니다. BSR방식이 degradation을 제거하는데 유용하지만 현실적인 디테일한 부분을 생성해내지 못합니다.

**Zero-shot Image Restoration.** ZIR은 사전 학습된 네트워크를 이용합니다. 초기에는 GAN's latent space에서 latent code를 탐색하는 데 집중했습니다. 최근에는 DDPM을 활용하는 방식으로 추세고 이어집니다. DDRM은 SVD기반 접근으로 이미지 복구를 효율적으로 수행하며 DDNM은 이론적으로 벡터의 range-null space 구성을 분석하고 sampling schedule을 구성합니다. 분류기 방식의 영향을 받아 GDP는 더욱 편리하고 효율적인 인퍼런스 방식을 보여줍니다. 이러한 연구들은 zero-shot에서 좋은 결과를 보입니다. 그러나 만족스럽지 못한 복구 결과를 내보입니다.

**Bline Face Restoration.** 이미지 생성 특정 분야로, 얼굴 이미지는 구조적인 정보를 수반합니다. 초기 방식은 기하학적 사전 얼굴 파싱 맵, 얼굴 랜드마크, 얼굴 요소 히트맵 또는 참고 사전 정보를 보조 정보로써 활용했습니다. 빠른 발전으로 많은 방식은 얼굴 복구에서 뛰어난 성과를 보였습니다. 대표적인 GAN 사전 방식은 고품질, 높은 재현성의 이미지를 보여줍니다.

## Methodology

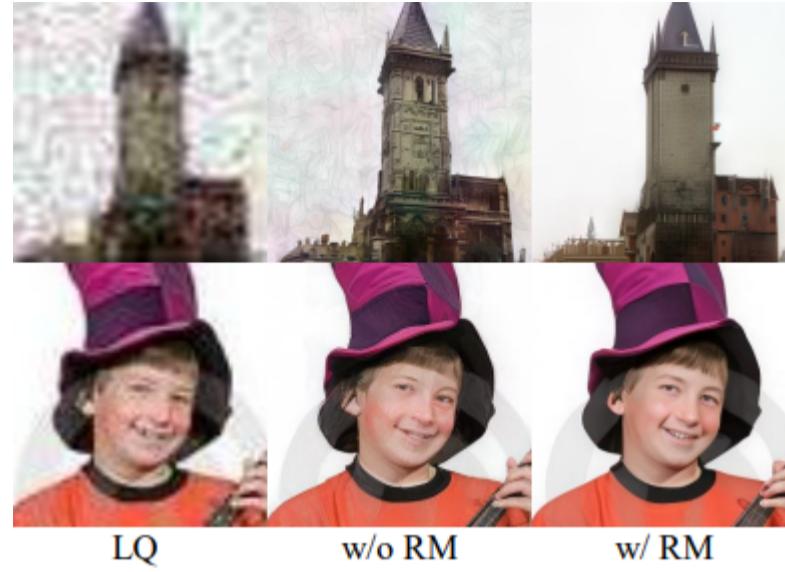
연구에서 우리는 강력한 생성 모델인 Stable Diffusion을 blind restoration을 해결하기 위해 사용하고자 합니다. 제안한 프레임워크는 two-stage 방식으로 효율적이고, 강건하고 유연합니다. 먼저, 우리는 Restoration Module으로 노이즈나 왜곡과 같은 이미지에서의 오염을 회귀로 스를 사용하여 제거합니다. 지역적인 텍스쳐와 디테일들은 여전히 부족하기 때문에 우리는 Stable Diffusion으로 정보 손실을 복구하도록 사용합니다. 특히 우리는 사전 학습된 SwinIR을 다양한 degradation을 제거하기 위해 사용합니다. 그리고 생성 쪽에서 자연스러운 복구 결과를 생성합니다. 게다가 controllable module을 통해 현실성과 재현성을 조절할 수 있습니다.



### 3.1 Pretraining for Degradation Removal

**Degradation Model.** BIR은 알 수 없는 복잡한 degradation이 있는 저품질(LQ) 이미지에서 복구된 깨끗한 이미지를 만드는 것을 목표로 합니다. 전형적으로, 흐릿함, 노이즈, 압축, 저해상도가 수반되어집니다. degradation 공간을 더 좋게 채우기 위해서 다양하고 고차원적인 포괄적인 degradation 모델을 사용합니다. degradation 중에서 blur, resize 그리고 noise는 실제 이미지에서 중요한 요인입니다. blur에서는 isotropic Gaussian 그리고 anisotropic Gaussian kernels를 사용합니다. resize에서는 영역 resize, 선형 resize, bicubic resize를 사용하고, noise의 경우 Gaussian noise, Poisson noise 그리고 JPEG 압축 노이즈를 포함합니다. 고차원 degradation에 관해서는 blur-resize-noise process 2차원 degradation을 사용합니다. 이러한 방식은 image restoration 목적으로 사용되었기 때문에 원래의 이미지 사이즈로 복구되어집니다.

**Restoration Module.** 강건한 생성 모델을 만들기 위해, 먼저 대부분의 degradation을 제거하는 보수적으로 가능한 방식을 사용합니다. 그러면 생성 모듈을 사용해 잃은 정보를 재생산합니다. 이러한 설계는 diffusion model이 질감과 디테일에 집중하여 생성하도록 도와주어 좋은 품질의 이미지를 만들어냅니다.



우리는 SwinIR을 수정하여 사용합니다, 특히, 우리는 8분의 1로 줄이는 다운 샘플하기 위한 픽셀 unshuffle방식을 활용합니다. 그리고 3\*3 convolution layer가 얕은 피쳐 추출기로 사용되어집니다. 모든 연속적인 트랜스포머 방식은 latent diffusion model 처럼 저해상도 공간에서 수행되어집니다. 깊은 피쳐 추출기는 Residual Swin Transformer Blocks(RSTB)를 선택하며 각각의 RSTB는 몇개의 Swin Transformer Layers(STL)를 가집니다. 얕고 깊은 피쳐는 저해상도와 고해상도 정보를 유지하기 위해 더해집니다. 깊은 피쳐를 다시 원래 이미지 공간으로 업샘플링 하기 위해 우리는 최근접 보간을 3번 시도하고 각각의 보간은 하나의 conv layer와 Leaky ReLU 활성화 함수로 이뤄집니다. Restoration module을 L2 pixel loss를 최소화하기 위해 최적화 되어집니다.

$$I_{reg} = \text{SwinIR}(I_{LQ}), \quad \mathcal{L}_{reg} = \|I_{reg} - I_{HQ}\|_2^2,$$

$I_{LQ}$ 는 low quality를 뜻하고,  $I_{HQ}$ 는 high quality를 의미합니다.  $I_{reg}$ 는 regression learning에서 얻어지며 이후 latent diffusion model을 전이학습하는데 사용됩니다.

### 3.2 Leverage Generative Prior for Image Reconstruction

**Preliminary: Stable Diffusion.** Stable diffusion기반으로 우리의 방식을 사용합니다. Diffusion 모델의 경우는 데이터의 분포를 추정하며 디노이징 과정을 거치며 데이터 샘플을 생성하는 것을 학습합니다. 더 효율적이고 안정적인 학습을 위해서 Stable Diffusion은 image  $x$ 를 latent  $z$ 로 이를 다시 구성하는 인코더와 디코더를 가지는 사전 학습 autoencoder를 사용합니다. 이는 hybrid 목적함수인 VAE, Patch-GAN 그리고 LPIPS를 사용합니다. diffusion과 denoising 과정은 latent space에서 수행되어집니다. Gaussian noise variance(0,1)을 time  $t$ 마다 더해주며 noisy latent를 생성합니다.

$$z_t = \sqrt{\bar{\alpha}_t} z + \sqrt{1 - \bar{\alpha}_t} \epsilon$$

$\epsilon \sim N(0, I)$ , 에서  $t$ 가 충분히 크다면, latent Ztsms rjdml 표준 가우시안 분포를 뛸 것입니다. 네트워크는 노이즈  $c$  조건에서 noise 앱실론을 예측하며 학습되어집니다. 목적 함수는 다음과 같습니다.

$$\mathcal{L}_{ldm} = \mathbb{E}_{z,c,t,\epsilon} [\|\epsilon - \epsilon_\theta(z_t = \sqrt{\bar{\alpha}_t} z + \sqrt{1 - \bar{\alpha}_t} \epsilon, c, t)\|_2^2]$$

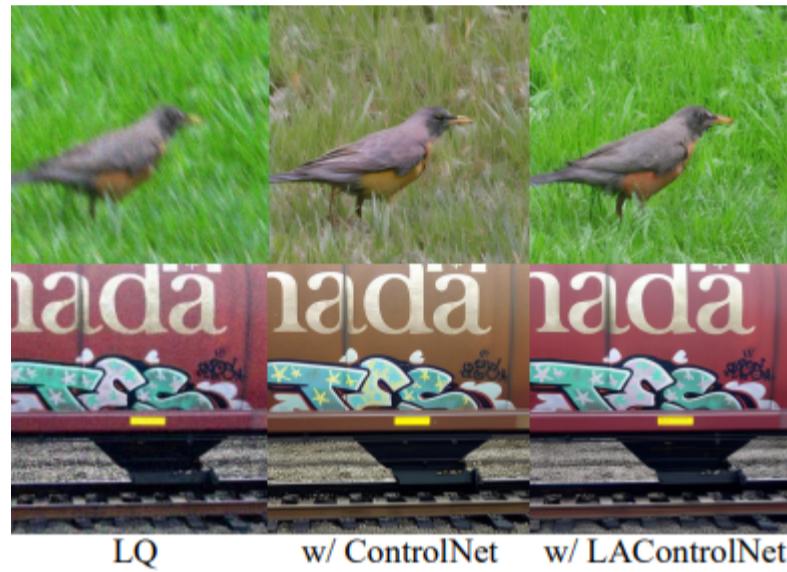
$x, c$ 는 데이터셋과  $z$ 에서부터 샘플링 되어집니다.  $t$ 는 균일하게 샘플링되고 앱실론은 스탠다드 가우시안 분포에서 샘플링 되어집니다.

**LAControlNet.** stage-one에서 대부분의 degradation이 제거 될지라도,  $I_{reg}$ 는 종종 너무 지나치게 매끄러워지며 자연스러운 고품질 이미지에서 멀어지게 됩니다. 그래서 우리는 사전 학습된 Stable Diffusion을  $I_{reg} - I_{HQ}$  쌍으로 이미지를 재구축하게 됩니다. 먼저 우리는 Stable Diffusion의 사전 학습 VAE Encoder를 활용해  $I_{reg}$ 를 latent space로 보냅니다. 조건 latent  $I_{reg}$ 를 얻게 됩니다. UNet denoiser는 encoder, middle block, decoder를 포함한 latent diffusion을 수행합니다. 특히, decoder는 encoder에서 피쳐를 받고 이들을 다른 스케일로 결합시킵니다. 여기서 우리는 같은 encoder와 middle block을 포함하는 UNet denoiser 병렬 모델을 제작했습니다. 그러면, 병렬 모델의 인풋으로 조건 latent  $I_{reg}$ 를 샘플링된 노이즈와 결합시킵니다. 이러한 결합 방식은 첫 번째 convolution layer의 채널을 증가시켜주며 새로운 더해진 파라미터를 0으로 시작하게끔 합니다. 다른 파라미터는 사전 학습된 UNet denoiser의 체크포인트에서 시작되어집니다. 병렬 모듈의 아웃풋은 원래의 Unet decoder에 더해집니다. 게다가 1\*1 convolution layer는 각 스케일에서 addition operation을 적용하기 전에 사용됩니다. 전이학습

을 하며 병렬 모듈과  $1 \times 1$  convolution layers는 상호보완적으로 최적화되어집니다. 우리는 다음과 같은 latent diffusion 목적함수를 최적화시켜 최소화하고자 합니다.

$$\mathcal{L}_{Diff} = \mathbb{E}_{z_t, c, t, \epsilon, \mathcal{E}(I_{reg})} [\|\epsilon - \epsilon_\theta(z_t, c, t, \mathcal{E}(I_{reg}))\|_2^2].$$

이 스테이지에서의 얻어진 결과는  $I_{diff}$ 입니다. 요약하자면, 오직 UNet denoiser에서 skip-connected된 feature만이 우리 테스크에서 튜닝되어집니다. 이러한 전략은 작은 학습 데이터에서의 오버피팅을 완화시켜줍니다. 그리고 Stable Diffusion에서의 높은 품질 생성 능력을 유지할 수 있습니다. 더 중요한 건, 조건 메커니즘은 전체를 scratch부터 학습하는 ControlNet과 비교했을 때 이미지 복구 테스크에서 간단하며 효율적입니다. 잘 학습된 VAE의 Encoder인 LAControlNet은 latent 변수로 조건 이미지를 같은 표현 공간으로 주입시킬 수 있습니다. 이러한 전략은 latent diffusion의 내재적 지식과 외부 정보와의 격차를 효과적으로 완화시킬 수 있습니다. 실제로, 직접적인 ControlNet의 활용은 극심한 색깔 변환을 일으킬 수 있습니다.



### 3.3 Latent Image Guidance for Fidelity-Realness Trade-off

two-stage 접근 방식으로 이미 좋은 결과를 얻게 되었지만, 현실성과 재현성 사이의 trade-off 관계는 다양한 유저의 선호에 따라 변경되어집니다. 그래서 우리는 stage-one에서의  $I_{reg}$ 를 통해 얻는 denoising process를 안내해줄 수 있는 controllable module을 제시합니다. 노이지 이미지에서 타겟 클래스를 예측하도록 학습된 classifier를 이용한 guidance는 고안됩니다. 대부분의 경우 사전 학습 모델은 깨끗한 이미지에서 학습되어집니다. 이러한 상황에서 디퓨전 과정 생성을 조절하는 중간 변수  $x_0$ 로 억제합니다. 특히, sampling 과정에서 노이즈를 예측하여 노이즈 이미지에서 깨끗한 이미지를 추정합니다. 이를 보면 diffusion과 denoising processes는 latent space를 기반으로 함을 볼 수 있어 우리는 다음과 같은 식을 이용하여 깨끗한 latent  $z$ 를 얻고자 합니다.

$$\tilde{z}_0 = \frac{z_t}{\sqrt{\bar{\alpha}_t}} - \frac{\sqrt{1 - \bar{\alpha}_t} \epsilon_\theta(z_t, c, t, \mathcal{E}(I_{reg}))}{\sqrt{\bar{\alpha}_t}}$$

전반적인 방식은 반복적으로 latent 피쳐 사이에서 공간과 컬러의 관계를 학습시킵니다. 그리고 생성된 latent를 이전 latent의 요소를 보존하도록 학습합니다. 그러므로, 하나는 얼마나 정보를 유지할지를 다루며 아웃풋 이미지를 얼마나 스무스하게 만들지 결정합니다. 전체적인 latent 가이던스는 아래와 같습니다.

---

**Algorithm 1** Latent-guided diffusion, given a diffusion model  $\epsilon_\theta$ , and the VAE’s encoder  $\mathcal{E}$  and decoder  $\mathcal{D}$

---

**Input:** Guidance image  $I_{reg}$ , text description  $c$  (set to empty), diffusion steps  $T$ , gradient scale  $s$   
**Output:** Output image  $\mathcal{D}(z_0)$   
 Sample  $z_T$  from  $\mathcal{N}(0, \mathbf{I})$   
**for**  $t$  from  $T$  to 1 **do**  
 $\tilde{z}_0 \leftarrow \frac{z_t}{\sqrt{\bar{\alpha}_t}} - \frac{\sqrt{1 - \bar{\alpha}_t} \epsilon_\theta(z_t, c, t, \mathcal{E}(I_{reg}))}{\sqrt{\bar{\alpha}_t}}$   
 $\mathcal{L} = \mathcal{L}(\tilde{z}_0, \mathcal{E}(I_{reg}))$   
 Sample  $z_{t-1}$  by  $\mathcal{N}(\mu_\theta(z_t) - s \nabla_{\tilde{z}_0} \mathcal{L}, \sigma_t^2)$   
**end for**  
**return**  $\mathcal{D}(z_0)$

---

## Experiments

### 4.1 Datasets, Implementation, Metric

**Datasets.** ImageNet dataset와 FFGQ dataset 기타 등을 이용하여 학습을 진행했습니다.

**Implementation.** restoration module은 8 residual Swin Transformer blocks (RSTB)를 채택했고, RSTB는 6개의 Swin Transformer Layers를 지닙니다. head는 6개 그리고 window size는 8입니다. DiffBIR은 512X512 사이즈보다 큰 이미지를 다룰 수 있으며 512보다 작은 사이드가 있다면 512까지 사이즈를 키우고 나중에 이를 줄입니다.

**Metrics.** PSNR, SSIM and LPIPS를 채택합니다. 더 좋은 평가를 얻기 위해 IQA metrics를 사용합니다. 우리는 IDSL을 평가하고 FID를 이용합니다.

### 4.2 Comparisons with State-of-the-Art Methods

BSR에서 우리는 다른 sota model과 비교를 진행합니다.

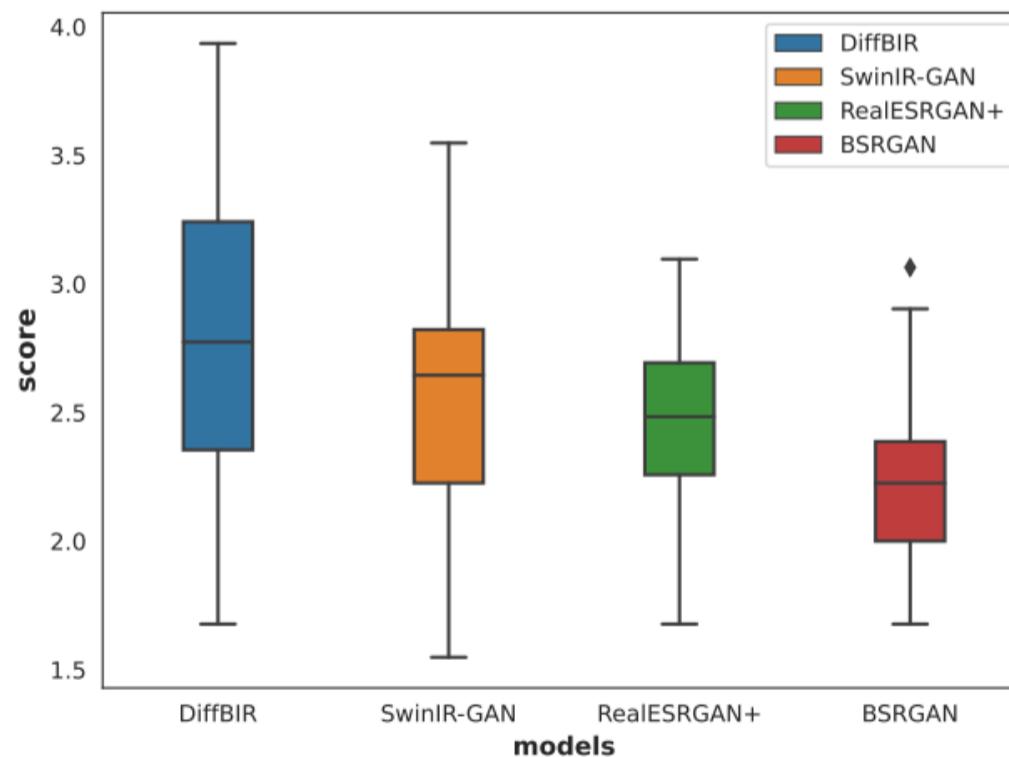
Table 1: Comparison with state-of-the-art BSR and ZIR methods on real-world datasets with a  $4 \times$  upsampling scale. **Red** and **blue** indicate the best and second best performance. The top 3 results are marked as **gray**.

Dataset	Metric	DDNM [57]	GDP [16]	Real-ESRGAN+[55]	BSRGAN [64]	SwinIR-GAN [36]	FeMaSR [6]	DiffBIR(Ours)
RealSRSet	MANIQA↑	0.4535	0.4581	0.5376	0.5640	0.5295	0.5247	0.5906
	NIQE↓	6.8415	5.0626	5.7401	5.6074	5.6093	5.2353	6.0738
Real47	MANIQA↑	0.4813	0.5237	0.5900	0.5889	0.5721	0.5718	0.6293
	NIQE↓	6.4768	3.9866	3.9103	4.0338	3.9910	4.1731	3.9240



다른 모델에 비해 텍스트에서 더욱 더 자연스럽고 정교한 결과를 출력해냈습니다. 그리고 다른 모델의 결과는 지나치게 스무스한 반면 우리는 디테일한 부분들을 살려냈습니다.

또한, user study를 통해 평가를 진행하였습니다.



BFR task에서도 평가를 진행합니다.

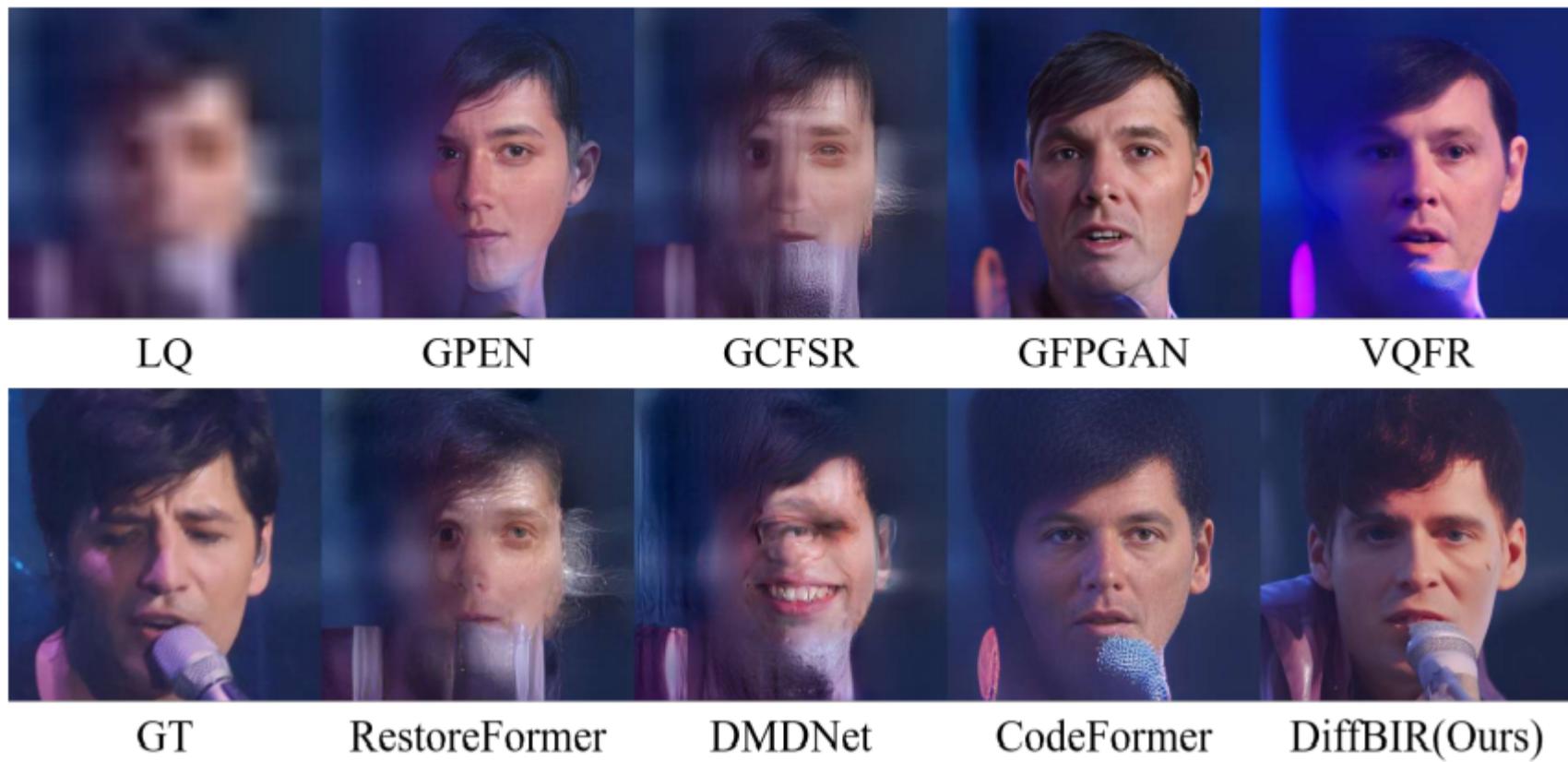


Table 2: Comparison with state-of-the-art methods for BFR on both synthetic and real-world face datasets. **Red** and **blue** indicate the best and second best performance. The top 3 results are marked as gray.

Dataset	Synthetic CelebA-Test					Wild		
	PSNR↑	SSIM↑	LPIPS↓	FID↓	IDS↑	LFW-Test	WIDER-Test	CelebChild-Test
Method								
GPEN [61]	21.3995	0.5742	0.4687	23.92	0.48	51.97	46.35	76.58
GCFSR [19]	<b>21.8791</b>	<b>0.6072</b>	0.4577	35.49	0.44	52.20	40.86	76.29
GFPGAN [54]	21.6953	0.6060	0.4304	21.69	0.49	52.11	41.70	80.69
VQFR [18]	21.3014	<b>0.6132</b>	<b>0.4116</b>	<b>20.30</b>	0.48	49.88	<b>37.87</b>	<b>74.76</b>
RestoreFormer [59]	21.0025	0.5283	0.4789	43.77	<b>0.56</b>	48.43	49.79	<b>70.54</b>
DMDNet [35]	21.6617	0.6000	0.4828	64.79	<b>0.67</b>	<b>43.36</b>	40.51	79.38
CodeFormer [68]	<b>22.1519</b>	0.5948	<b>0.4055</b>	22.19	0.47	52.37	38.78	79.54
DiffBIR(Ours)	21.7509	0.5971	0.4573	<b>20.02</b>	0.51	<b>39.58</b>	<b>32.35</b>	75.94

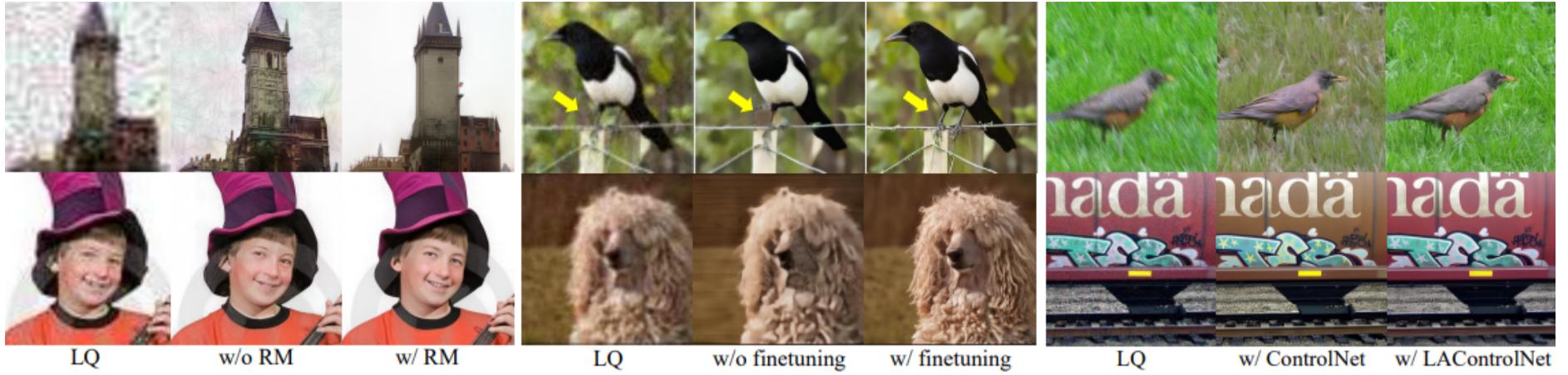
### 4.3 Ablation Studies

**The Importance of Restoration Module.** two-stage 파이프라인의 효과를 알아봅니다. 여기서 RM을 제거하고 바로 디퓨전 모델로 주입을 하였습니다. Restoration module 제거시 많은 성능 하락이 있었습니다.

Table 3: The effectiveness of restoration module. The best results are denoted as Red.

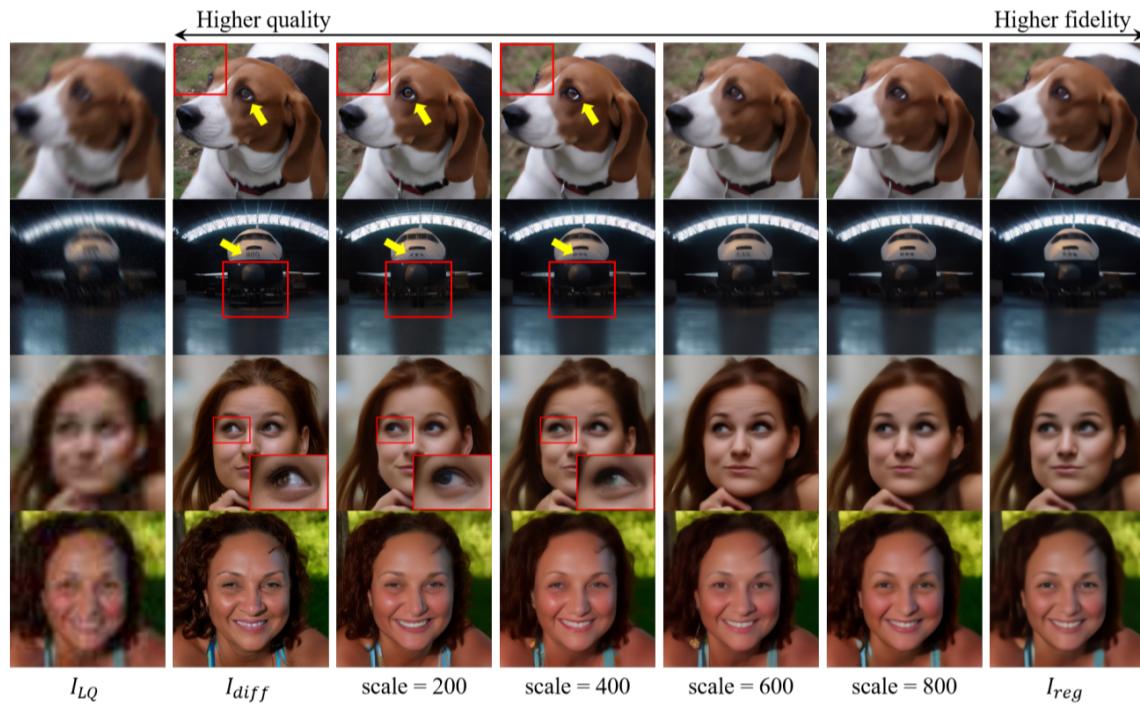
Dataset	Face			General	
	LFW-Test	WIDER-Test	CelebChild-Test	RealSRSet	Real47
Method	FID↓	FID↓	FID↓	MANIQA↑	MANIQA↑
DiffBIR(w/o RM)	40.78	33.22	75.98	0.582	0.624
DiffBIR(w/ RM)	<b>39.58</b>	<b>32.35</b>	<b>75.94</b>	<b>0.591</b>	<b>0.629</b>

**The Necessity of Finetuning Stable Diffusion.** 또한, stable diffusion을 fine tuning하지 않을 경우 비현실적인 이미지를 생성하는 경우가 생깁니다. 사진에서 보다시피 새의 다리가 하나 밖에 없는 경우가 생깁니다. (가운데)



**The Effectiveness of LAControlNet.** 우리는  $I_{reg}$ 를 latent space로 encoding하는 LAControlNet을 강조합니다. scratch부터 학습을 하는 ControlNet과 비교를 진행합니다. ControlNet의 경우 색깔의 변환 결과를 초래합니다. 누군가는 불균형 샘플링은 최적화의 가능성을 증진 시켜 더 좋은 결과를 낼 수 있다고 할 수 있지만 그럼에도 불구하고 우리의 방식이 더 좋은 결과를 나타냈습니다.

**The Flexibility of Controllable Module.** 생성 복구 모델은 예상치 않은 디테일을 생성할 수 있습니다. 사용자의 선호에 따른 이미지 결과를 보여줍니다. 스무스한 높은 재현율을 원할 경우  $I_{reg}$ , 그게 아니라면 현실적인 이미지를 원한다면  $I_{diff}$ 를 선택할 수 있습니다.



## Conclusion and Limitations

Stable Diffusion기반으로 현실적인 이미지 복구를 하는 **DiffBIR** 이미지 복구 통합 프레임워크를 제안합니다. 이는 현실성과 재현성을 동시에 만족하는 two-stage restoration과 generation stage를 가지고 있습니다. BSR과 BFR task에서 동시에 SOTA를 달성합니다. 또한 diffusion model은 restoration task로 fine tuning 진행이 되어 보다 적합해졌습니다. 그러나 DiffBIR은 50번의 sampling steps이 존재하여 높은 컴퓨팅 자원이 요구됨과 인퍼런스 시간이 발생합니다.