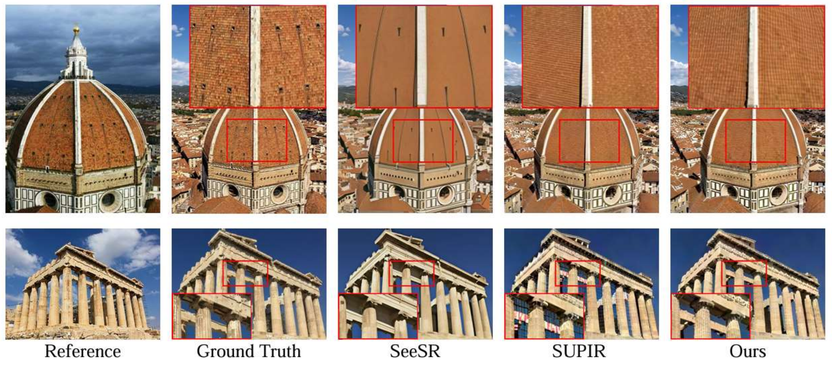
ReFIR: Grounding Large Restoration Models with Retrieval Augmentation

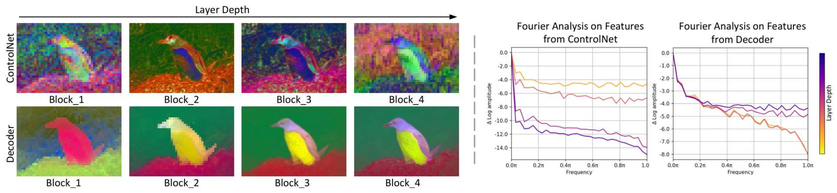
Hang Guo[1], Tao Dai[2,*], Zhihao Ouyang[3], Taolin Zhang[1], Yaohua Zha[1], Bin Chen[4], Shu-tao Xia[1,5]
[1]Tsinghua University; [2]Shenzhen University; [3]Aitist.ai; [4]Harbin Institute of Technology; [5]Peng Cheng Laboratory
*Corresponding author

Motivation



- Diffusion-based restoration models have gained success
- But sometimes, inaccurate results appear on hard cases
- Similar to LLMs, we define it as the 'hallucination ' dilemma of diffusion restoration models
- Train a larger model for hard samples seems impractical
- Our Idea: Inject 'external knowledge' from extra reference images!

Model Probing



We can divide the working mechanism of diffusion restoration models into two stages

1. Denoising Structure Reconstruction

The self-attention in the ControlNet reconstructs a clear overall structure from the noised representation

2. Detail Texture Restoration

The self-attention in the UNet decoder fills scene-specific textures based on the denoised structure map.

💡 We can transfer high-quality, scene-specific textures from the reference images during the detail texture restoration stage!

Method

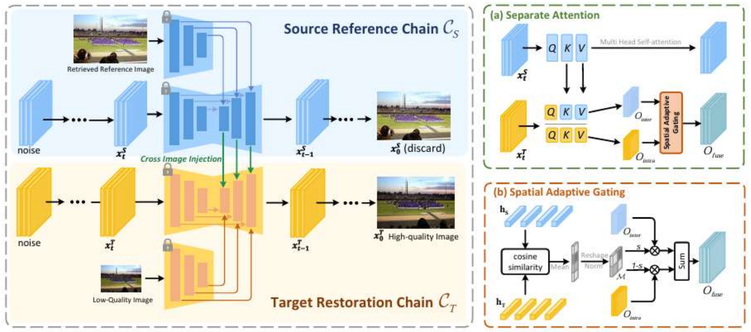
Overview

- First Stage: Retrieve reference images
- Second Stage: Use the retrieved images

Reference Image Retrieval

Simple solution: Use the Similarity matching of CLIP embedding to find the Top1 related images as the reference

Reference Image Retrieval



- Separate Attention

$$O_{intra} = \texttt{Attention}(Q_T, K_T, V_T), \quad O_{inter} = \texttt{Attention}(Q_T, K_S, V_S).$$

- Spatial Adaptive Gating

$$O_{fuse} = (\mathbf{1} - s\mathcal{M}) \otimes O_{intra} + s\mathcal{M} \otimes O_{inter}$$

- Distribution alignment

$$O'_{fuse} = \texttt{AdaIN}(O_{fuse}, O_{intra})$$

Advantages of ReFIR

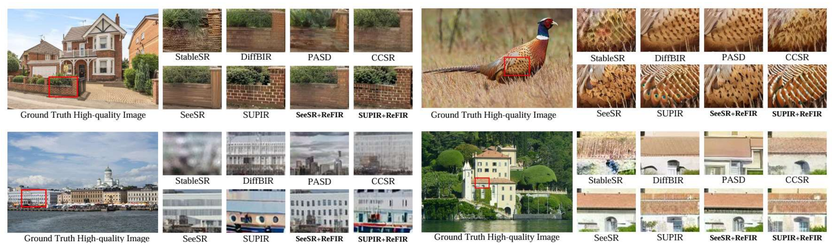
⭐ Training-free, only need to modify the inference!

⭐ Can be applied to multiple existing methods!

Comparison to SoTA

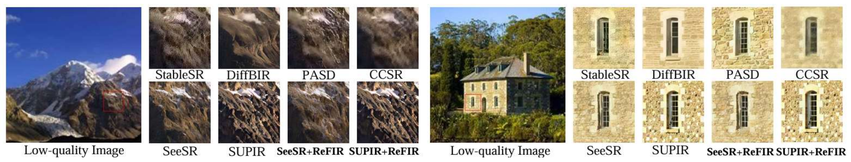
Restoration with Ideal Reference

Method	CUFED5 PSNR↑	SSIM↑	LPIPS↓	NIQE↓	FID↓	WR-SR PSNR↑	SSIM↑	LPIPS↓	NIQE↓	FID↓
C2-Matching [45]	20.77	0.5169	0.7282	8.4438	282.43	22.63	0.5627	0.7177	8.3238	157.61
DATSR [46]	20.75	0.5130	0.7301	8.6765	282.19	22.62	0.5620	0.7210	8.4329	157.54
MrefSR [47]	20.84	0.5218	0.7853	9.6524	286.44	22.68	0.5703	0.7748	9.7742	156.57
BSRGAN [9]	20.22	0.5256	0.4135	4.2204	203.17	22.07	0.5735	0.4073	3.8703	133.50
Real-ESRGAN [8]	20.31	0.5543	0.3698	3.8832	175.91	22.14	0.5974	0.3631	3.7001	97.88
StableSR [16]	20.46	0.4480	0.6532	6.3433	292.69	21.22	0.4421	0.5899	5.2040	145.07
DiffBIR [15]	19.76	0.4886	0.3820	3.5629	154.75	21.30	0.5284	0.3938	3.8736	76.05
PASD [17]	20.22	0.4959	0.5252	5.4828	208.64	21.12	0.5254	0.4292	4.2505	98.16
SeeSR [19]	19.94	0.5195	0.3660	3.7912	142.92	21.73	0.5658	0.3501	4.0155	65.78
SeeSR+ReFIR	20.32	0.5289	0.3338	3.7831	134.62	21.86	0.5664	0.3460	3.9089	61.22
Δimprovement	**+0.38**	**+0.0094**	**+0.0322**	**+0.0081**	**+8.30**	**+0.13**	**+0.0006**	**+0.0041**	**+0.1066**	**+4.56**
SUPIR	18.97	0.4665	0.4807	4.5624	168.26	20.91	0.5426	0.3791	3.7587	75.85
SUPIR+ReFIR	19.00	0.4729	0.4341	4.2085	148.69	21.02	0.5497	0.3785	3.7478	71.82
Δimprovement	**+0.03**	**+0.0064**	**+0.0466**	**+0.3539**	**+19.57**	**+0.11**	**+0.0071**	**+0.0006**	**+0.0109**	**+4.03**

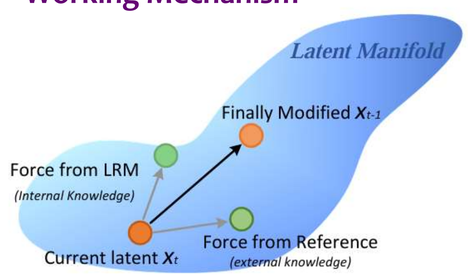


Restoration in the Wild

Metrics	StableSR [16]	DiffBIR [15]	PASD [17]	CCSR [18]	SeeSR [19]	SUPIR [20]	SeeSR+ReFIR (Ours)	SUPIR+ReFIR (Ours)
NIQE↓	3.7695	2.8458	5.1603	5.5082	4.7432	3.5076	4.4566(+0.2866)	3.4593(+0.0483)
MUSIQ↑	51.95	65.20	49.01	32.26	55.54	59.84	57.13(+1.59)	60.49(+0.65)
CLIPIQA↑	0.6852	0.7845	0.5863	0.4568	0.6575	0.5692	0.6732(+0.0157)	0.5722(+0.003)



Working Mechanism



Additional Resources

paper code