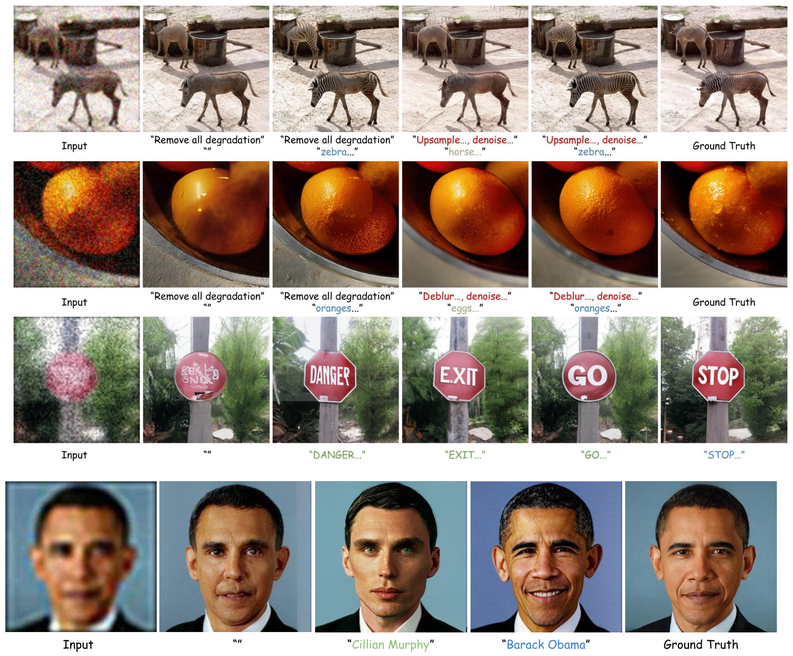
SPIRE: Semantic Prompt-Driven Image Restoration

Chenyang Qi, Zhengzhong Tu, Keren Ye, Mauricio Delbracio, Peyman Milanfar, Qifeng Chen, Hossein Talebi

Goal and Motivation



Other degradations

Major Contributions

- We introduce the first unified text-driven image restoration model that supports both semantic prompts and restoration instructions. Our experiments demonstrate that incorporating semantic prompts and restoration instructions significantly enhances the restoration quality.
- Our proposed approach provides a mechanism for users to adjust the category and strength of the restoration effect based on their subjective preferences.
- We demonstrate that text can serve as universal guidance control for low-level image restoration, eliminating the need for task-specific model design.

Framework

Parameterized			Real-ESRGAN		
Degradation Process	p(choose)	Restoration Prompt	Degradation Process	p(choose)	Restoration Prompt
Gaussian Blur	0.5	Deblur with {sigma$\in$ [0.2, 3.0]} or Deblur	Blur	1.0	Remove all degradation
Downsample	0.5	Upsample to {resizing factor$\in$ [1.0, 7.0]} or Upsample	Resize	1.0	
Gaussian Noise	0.5	Denoise with {sigma$\in$ [0.0, 0.12]} or Denoise	Noise	1.0	
JPEG	0.5	Dejpeg with quality {quality factor$\in$ [30, 92]} or Dejpeg	JPEG	1.0	

In the training phase, we synthesize a degraded version y, of a clean image x. Our degradation synthesis pipeline also creates a restoration prompt cr , which contains numeric parameters that reflect the intensity of the degradation introduced.

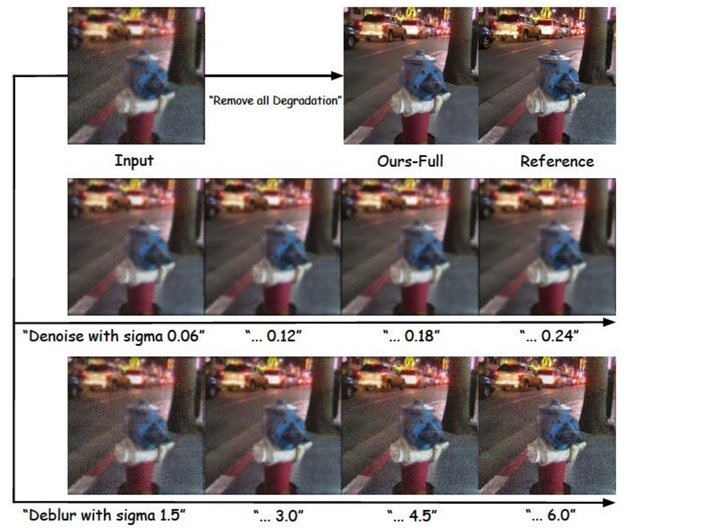
During test time, the users can employ the SPIRE framework as either a blind restoration model with restoration prompt "Remove all degradation" and empty semantic prompt ∅, or manually adjust the restoration cr and semantic prompts cs to obtain what they want.

Quantitative Comparison

Method	Prompts Sem	Prompts Res	Parameterized Degradation with synthesized c_r FID↓	LPIPS↓	PSNR↑	SSIM↑	CLIP-I↑	CLIP-T↑	Real-ESRGAN Degradation without c_r FID↓	LPIPS↓	PSNR↑	SSIM↑	CLIP-I↑	CLIP-T↑
SwinIR [30]	✗	✗	43.22	0.423	**24.40**	**0.717**	0.856	0.285	48.37	0.449	**23.45**	**0.699**	0.842	0.284
StableSR [67]	✗	✗	20.55	0.313	21.03	0.613	0.886	0.298	25.75	0.364	20.42	0.581	0.864	0.298
DiffBIR [34]	✗	✗	17.26	0.302	22.16	0.604	0.912	0.297	19.17	0.330	21.48	0.587	0.898	0.298
ControlNet-SR [84]	✗	✗	13.65	0.222	23.75	0.669	0.938	0.300	16.99	0.269	22.95	0.628	0.924	0.299
Ours w/o text	✗	✗	12.70	0.221	23.84	0.671	0.939	0.299	16.25	**0.262**	23.15	0.636	0.929	0.300
DiffBIR [34] + SDEdit [37]	✓	✗	19.36	0.362	19.39	0.527	0.891	0.305	17.51	0.375	19.15	0.521	0.887	**0.308**
DiffBIR [34] + CLIP [46]	✓	✗	18.46	0.365	20.50	0.526	0.896	**0.308**	20.31	0.374	20.45	0.539	0.885	0.307
ControlNet-SR + CLIP [46]	✓	✗	13.00	0.241	23.18	0.648	0.937	0.307	15.16	0.286	22.45	0.610	0.926	**0.308**
Ours	✓	✓	**11.34**	**0.219**	23.61	0.665	**0.943**	0.306	**14.42**	**0.262**	23.14	0.633	**0.935**	**0.308**

Qualitative Comparison

Restoration Prompting



Prompt space walking visualization for the restoration prompt. Given the same degraded input (upper left) and empty semantic prompt ∅, our method can decouple the restoration direction and strength via only prompting the quantitative number in natural language.

Result on Real-world Data with Synthetic Caption