Face2Scene: Facial Degradation as an Oracle for Scene Restoration

Amirhossein Kazerouni, Maitreya Suin, Tristan Aumentado-Armstrong, Sina Honari, Amanpreet Walia, Iqbal Mohomed, Konstantinos G. Derpanis, Babak Taati, Alex Levinshtein

Samsung Research AI Center-Toronto

University of Toronto · Vector Institute · UHN · York U

CVPR JUNE 3-7, 2026 DENVER COLORADO

Project Page

Motivation

Facial Oracle | **Diffusion Priors** | **Scene Restoration**

Existing ref-based face restoration methods excel at enhancing faces **but** neglect the surrounding scene, leaving backgrounds and bodies degraded.

- Face-only models are not applicable to full scenes
- Scene restoration is critical for human-centric images
- A face can act as an oracle for the degradation of the whole image

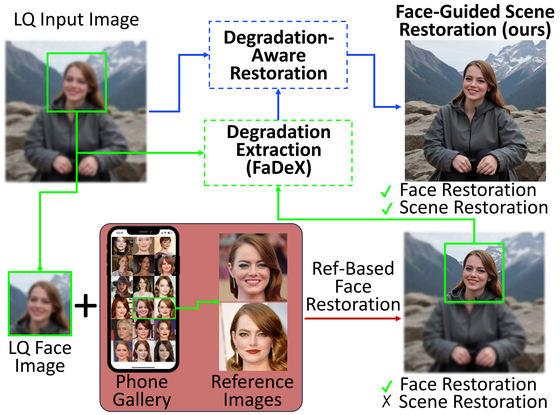
How can we leverage reference-based face restoration models for full scenes?

The Idea

Treat the face as a degradation probe; its stable geometry and available identity references make it the perfect oracle.

- Restore the face first using identity references.
- Extract a degradation code from the LQ ↔ HQ face pair.
- Use that code to condition full-scene restoration.

Overview



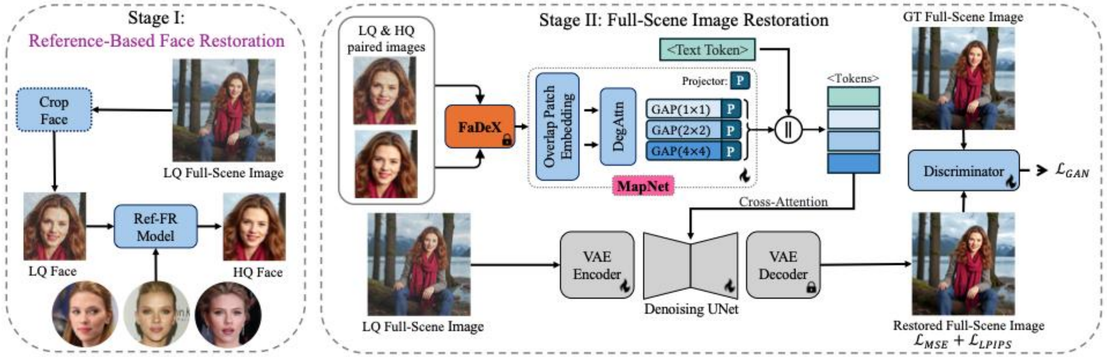
Restore the face. Then use it to restore the scene.

Components

Ref-FR: Reference-based face restoration transfers person-specific priors to a degraded face.

FaDeX: Face-derived Degradation eXtractor estimates a degradation code from the LQ ↔ HQ face pair, ignoring content.

MapNet: Multi-scale token mapper turns the code into degradation-aware tokens to condition a diffusion model.



End-to-end pipeline. The restored face yields a degradation code that conditions the diffusion model to restore the whole scene.

InScene Synthetic Dataset

Prompt: **First**, full-body photography of woman in a montane cloud forest. **Second**, vertical orientation, direct perspective, wide-angle view 1.2, distant background full detailed, everything in frame clearly visible, ultra high detail, balanced exposure, textured bark and moss rendered clearly. **Third**, rusted pipes, grates, and riveted beams in the background, with silver overcast light, distant objects clearly rendered, edge-to-edge clarity. **Fourth**, plaster texture and wood grain remain defined, consistent detail across near and far surfaces, cinematic realism, ultra high detail.

CelebRef-HQ + InfiniteYou → Facial Reference + Generated Scene Image

InScene Real Dataset

Real-World Gallery Images

Results

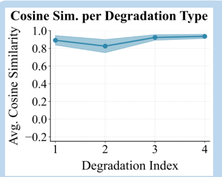
Best Fidelity | **Highest Quality** | **Fast**

Table 1: InScene Synthetic Validation Dataset (↓ = lower better, ↑ = higher better) [C-IQA=CLIP-IQA, M-IQA=MANIQA]

Methods	DISTS ↓	LPIPS ↓	PSNR ↑	SSIM ↑	FID ↓	MUSIQ ↑	C-IQA ↑	M-IQA ↑	LIQE ↑	TOPIQ ↑
SUPIR	0.136	0.312	**24.08**	**0.629**	24.85	70.20	0.602	0.363	4.02	0.601
InvSR	0.132	0.321	23.86	0.590	22.92	71.12	0.707	0.405	4.09	0.620
S3Diff	0.113	0.256	23.60	0.592	18.06	72.18	0.698	0.386	4.42	0.623
Face2Scene	**0.101**	**0.242**	22.90	0.557	**15.26**	**74.76**	**0.764**	**0.435**	**4.72**	**0.652**

Table 2: InScene Synthetic-Degradation Real-Image Validation Dataset (↓ = lower better, ↑ = higher better)

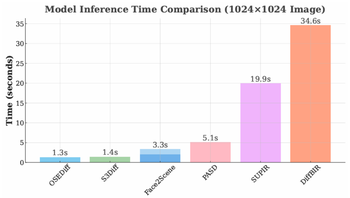
Methods	DISTS ↓	LPIPS ↓	PSNR ↑	SSIM ↑	FID ↓	MUSIQ ↑	C-IQA ↑	M-IQA ↑	LIQE ↑	TOPIQ ↑
SUPIR	0.236	0.556	17.64	0.478	45.22	70.43	0.547	0.401	4.13	0.607
InvSR	0.231	0.547	17.41	0.492	49.35	72.91	0.691	0.457	4.43	0.644
S3Diff	0.223	0.515	17.14	0.489	**38.64**	73.82	0.673	0.448	4.71	0.663
Face2Scene	**0.118**	**0.250**	**22.90**	**0.620**	42.21	**75.37**	**0.702**	**0.471**	**4.80**	**0.678**



Separating Content vs. Degradation in FaDeX Code

Table 3: InScene Real-Degradation Real-Image Test Dataset

Methods	MUSIQ ↑	CLIP-IQA ↑	MANIQA ↑	LIQE ↑	TOPIQ ↑
SUPIR	60.4119	0.3506	0.3108	2.8347	0.4210
InvSR	72.7069	0.4559	0.4145	4.3888	0.6665
S3Diff	69.4941	0.5423	0.4014	4.1888	0.5805
Face2Scene	**73.3047**	**0.6407**	**0.4859**	**4.6973**	**0.6846**



Conclusion: A New Paradigm for Restoration

- ✓ **Facial Oracle Guidance:** faces enable precise degradation extraction.
- ✓ **FaDeX + MapNet:** disentangle content from degradation, then map to multi-scale tokens.
- ✓ **InScene Datasets:** curated synthetic + real-world benchmarks for human-centric restoration.
- ✓ **SotA performance:** best fidelity and lowest artifact rate across all metrics.