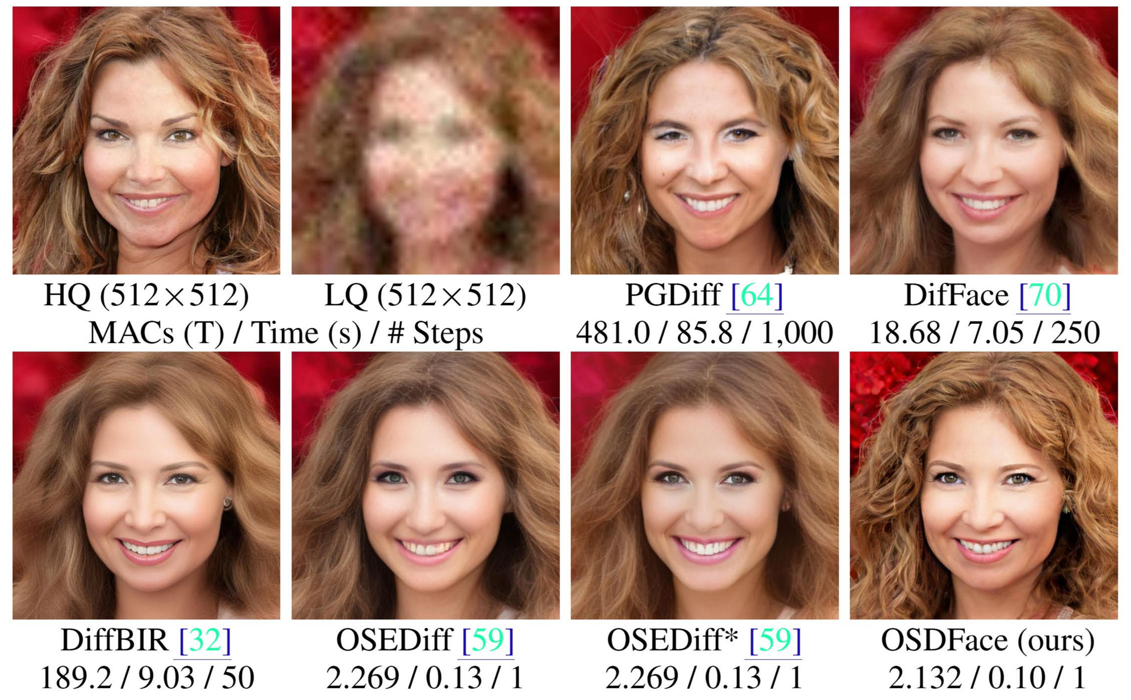
OSDFace: One-Step Diffusion Model for Face Restoration

Jingkai Wang[1†], Jue Gong[1†], Lin Zhang[1], Zheng Chen[1]
Xing Liu[2], Hong Gu[2], Yutong Liu[1*], Yulun Zhang[1*], Xiaokang Yang[1]
[1]Shanghai Jiao Tong University, [2]vivo Mobile Communication Co., Ltd

CVPR Nashville JUNE 11-15, 2025

SHANGHAI JIAO TONG UNIVERSITY



Contribution

- **OSDFace: One-Step Diffusion model for FACE restoration**. The **First** attempt to utilize one-step diffusion for restoring faces.
- **VRE: Visual Representation Embedder**. Using low-quality **VQ dictionary**, VRE captures rich prior from LQ images for a deeper understanding of visual content.
- **Realistic Face Alignment**. Facial identity loss for identity consistency and GAN loss for distribution alignment.
- **SOTA** performance on **Face Restoration** Lower complexity, Smaller model size, 80ms for 512×512 face.

Methods

❖ Architecture

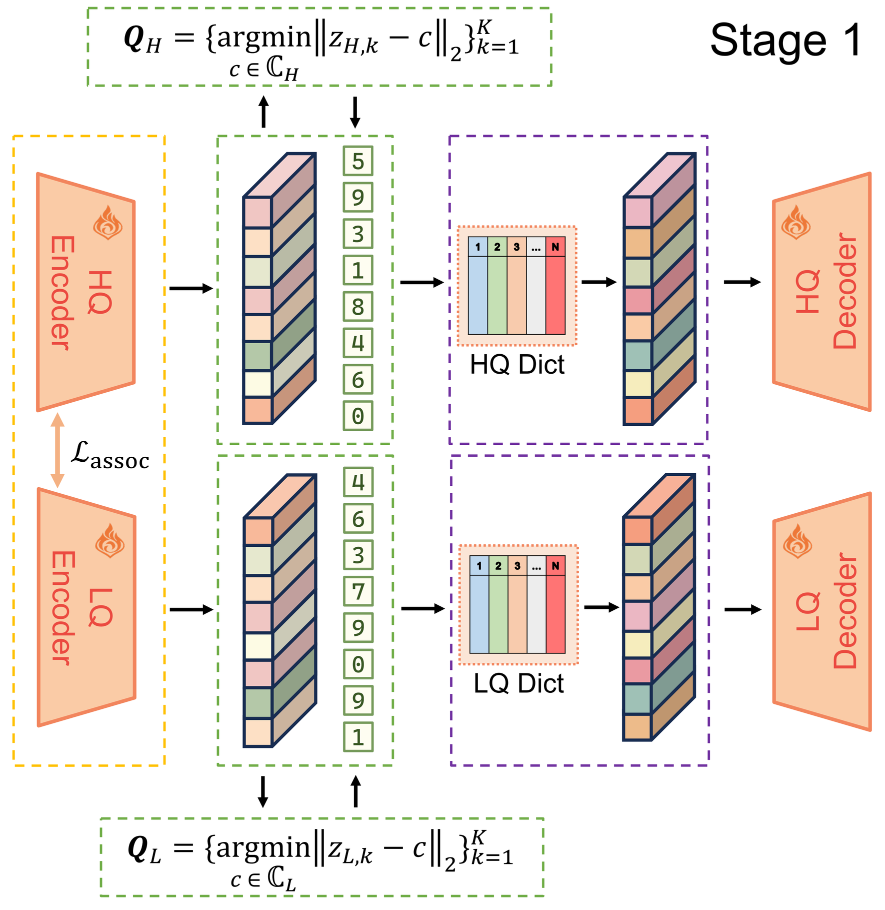


Fig.1. Illustration of Stage 1 Arch.

❖ Visual Representation Embedder (VRE)

- The architecture consists of the VRE and VAE decoder;
- Two VQ dictionaries corresponding to the HQ and LQ image categories and train VQVAE using self-reconstruction.
- Align the categories between LQ and HQ faces.
- Enhancing the diagonal correlation within VQ dictionaries

❖ Architecture

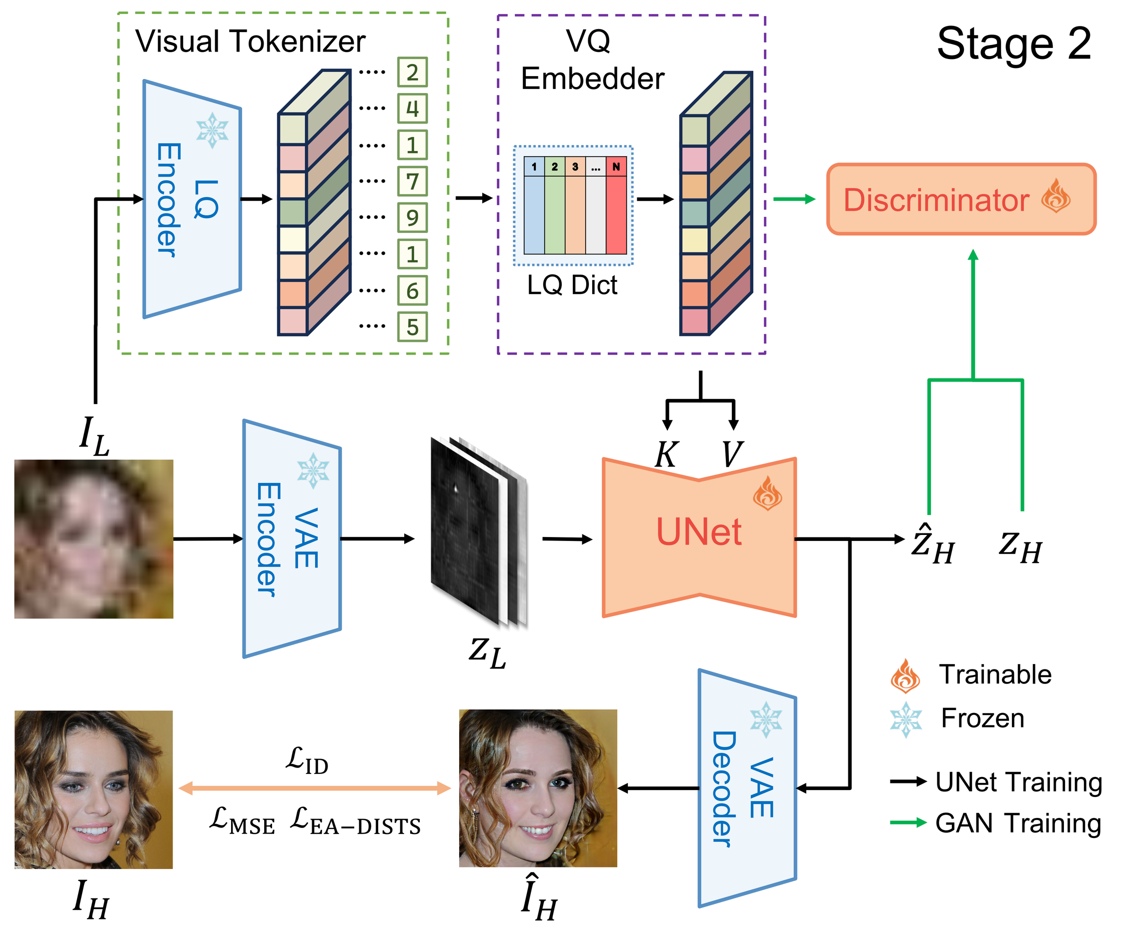


Fig.2. Illustration of Stage 2 Arch.

❖ Realistic Face Alignment

Loss Functions:

- Pixel Reconstruction
- Facial Identity
- Perceptual
- GAN

Total Arch:

- Stable Diffusion
- LoRA

Advantages:

- More harmonious
- More realistic
- More ID consistent

Experiments

Methods	C-IQA↑	M-IQA↑	MUSIQ↑	NIQE↓	FID↓
Learnable	0.6714	0.5312	75.0339	4.0855	45.4041
DAPE [58]	0.6863	0.5622	74.8107	4.3801	45.9380
VRE (w/o assoc)	0.6791	0.5180	75.2328	3.8228	44.3360
VRE (w/ assoc)	0.6946	0.5356	75.2911	3.8793	41.9502

Tab.1. Ablation Study of VRE

$\mathcal{L}_{ID}$	$\mathcal{L}_{EA\text{-}DISTS}$	$\mathcal{L}_{\mathcal{G}}$	$\mathcal{L}_{EA\text{-}LPIPS}$	C-IQA↑	M-IQA↑	MUSIQ↑	NIQE↓	FID↓
	✓	✓		0.6724	0.5243	74.5986	4.0190	41.7842
✓		✓		0.6710	0.5387	74.1060	4.3223	55.1016
✓		✓	✓	0.6674	0.5470	75.2021	4.1484	46.1684
✓	✓			0.6590	0.5081	75.2336	3.9857	45.7834
✓	✓	✓		0.6946	0.5356	75.2911	3.8793	41.9502

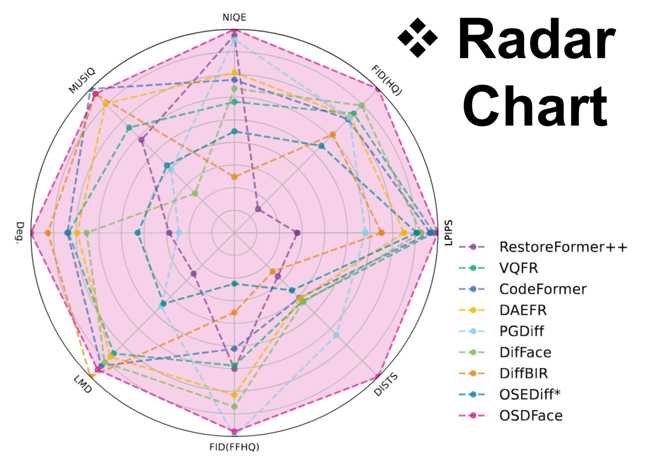
Tab.2. Ablation Study of Losses

Methods	PGDiff [61]	DifFace [67]	DiffBIR [30]	OSEDiff [57]	OSDFace
Step	1,000	250	50	1	1
Time (s)	85.81	7.05	9.03	0.13	0.10
Param (M)	176.4	175.4	3,042	1,302	978.4
MACs (G)	480,997	18,682	189,208	2,269	2,132

Tab.3. Complexity comparison

Comparisons

❖ Radar Chart



❖ Quantitative Comparisons ❖ Synthetic Dataset

Type	Methods	LPIPS↓	DISTS↓	MUSIQ↑	NIQE↓	Deg.↓	LMD↓	FID(FFHQ)	FID(HQ)↓
Non-Diffusion	RestoreFormer++ [56]	0.4535	0.2301	72.3612	3.9524	70.5083	8.8019	57.3723	72.7880
	VQFR [14]	0.3390	0.2131	73.2969	4.8374	62.2595	5.7297	58.1044	24.9093
	CodeFormer [71]	0.3412	0.2151	75.9432	4.5157	62.1972	5.3819	62.0280	26.8595
	DAEFR [49]	0.3580	0.2146	74.9853	4.4161	62.8184	5.6332	52.0341	26.0458
Multi-Step Diffusion	PGDiff [61] (s=1,000)	0.3866	0.1949	69.5676	4.0010	71.5996	7.3109	44.6258	26.3694
	DifFace [67] (s=250)	0.3469	0.2126	66.7451	4.6381	63.4511	5.4759	49.8075	22.2370
	DiffBIR [30] (s=50)	0.3740	0.2340	75.6360	6.2801	61.0238	5.1042	71.7767	32.5109
One-Step Diffusion	OSEDiff [57] (s=1)	0.4708	0.3165	51.2566	6.6968	72.7112	10.9948	89.8160	68.3579
	OSEDiff* [57] (s=1)	0.3496	0.2200	69.9807	5.3280	67.4026	7.4082	81.3624	37.1309
	OSDFace (ours, s=1)	0.3365	0.1773	75.6398	3.8840	60.0708	5.2867	45.4150	17.0617

Table 1. Synthetic Dataset – CelebA-Test from DAEFR degradation pipeline.

❖ Real-World Datasets

Method	Wider-Test					LFW-Test					WebPhoto-Test				
	C-IQA↑	M-IQA↑	MUSIQ↑	NIQE↓	FID↓	C-IQA↑	M-IQA↑	MUSIQ↑	NIQE↓	FID↓	C-IQA↑	M-IQA↑	MUSIQ↑	NIQE↓	FID↓
RestoreFormer++ [56]	0.7159	0.4767	71.332	3.7231	45.398	0.7024	0.5108	72.250	3.8434	50.253	0.6950	0.4902	71.484	4.0202	75.071
VQFR [14]	0.7069	0.5044	71.417	4.0357	37.866	0.7098	0.5339	74.385	3.8356	49.800	0.6769	0.4909	70.906	4.6095	84.776
CodeFormer [71]	0.6986	0.4958	73.406	4.1188	38.765	0.6890	0.5266	75.484	4.4377	52.341	0.6918	0.5034	74.001	4.6273	83.197
DAEFR [49]	0.6975	0.5205	74.143	3.5701	36.701	0.6964	0.5420	75.838	3.4788	47.527	0.6696	0.4934	72.698	3.9333	75.474
PGDiff [61] (s=1,000)	0.5824	0.4531	68.135	3.9315	35.862	0.5975	0.4858	71.244	4.0112	41.209	0.5653	0.4460	68.599	3.9930	86.954
DifFace [67] (s=250)	0.5924	0.4299	64.907	4.2380	37.099	0.6075	0.4577	69.617	3.9016	46.127	0.5737	0.4189	65.116	4.2474	79.556
DiffBIR [30] (s=50)	0.8084	0.6625	75.321	5.5903	35.343	0.7948	0.6735	76.421	5.6782	40.320	0.7441	0.5839	72.272	6.0093	91.834
OSEDiff [57] (s=1)	0.6235	0.4616	66.538	5.1921	42.014	0.6428	0.5022	72.577	4.7994	49.054	0.6321	0.4713	69.322	5.4122	111.121
OSEDiff* [57] (s=1)	0.6193	0.4752	69.101	5.0869	47.883	0.6186	0.4879	71.707	4.8002	51.048	0.6254	0.4823	69.816	5.3253	109.236
OSDFace (ours, s=1)	0.7284	0.5229	74.601	3.7741	34.648	0.7203	0.5493	75.354	3.8710	44.629	0.7106	0.5162	73.935	3.9864	84.597

❖ Visual Comparisons