Supplemental Material for 3D GAN Inversion with Facial Symmetry Prior

Fei Yin[1], Yong Zhang[2]†, Xuan Wang[3], Tengfei Wang[4], Xiaoyu Li[2], Yuan Gong[1], Yanbo Fan[2], Xiaodong Cun[2], Ying Shan[2], Cengiz Öztireli[5], Yujiu Yang[1]†
[1] Shenzhen International Graduate School, Tsinghua University
[2]Tencent AI Lab [3]Ant Group [4] HKUST [5]University of Cambridge

Appendix

A. Ablation Study on MEAD

To further verify the designed modules and strategy of our method, we conduct ablation study experiments on a multi-view dataset, MEAD [4]. The quantitative results are shown in Tab. 1. '$\mathcal{W}^+$ Inversion' denotes optimizing latent in $\mathcal{W}^+$ space with 500 iterations using only the ground truth image. '+ Symmetry Prior' denotes optimizing latent in $\mathcal{W}^+$ space with 500 iterations employing both original and symmetric view. '+ Joint Optimization' would further optimize generator parameters with $1,000$ iterations. '+ Geometry Regularization' would regularize the shape correctness during the joint optimization process. '+ Warping Pseudo' would introduce depth-guided 3D warping pseudos for supervision. It can be seen that the symmetry prior can strongly boost the vanilla inversion method, especially when inputting a side face (*e.g.*, L60, R60). The other designed modules enhance the baselines to a different extent from the rest columns. The results demonstrate that reasonable geometry of our method can help the model synthesize robust and consistent texture, which aligns with the conclusion of the manuscript.

B. Comparison with 2D GAN Inversion.

We apply rotation editing [3] directions to the latent code of StyleGAN-2 [1] to mimic the camera rotation. The comparison is shown in Fig. 1. The rotation of [3] makes changes to the identity. While our method can generate consistent and high-fidelity portraits in different views.

C. Additional Results on In-the-wild Images

Following the baseline comparison in the manuscript, we provide additional inversion results on in-the-wild images shown in Fig. 2 and Fig. 3, which demonstrate the effectiveness of our 3D GAN inversion method.

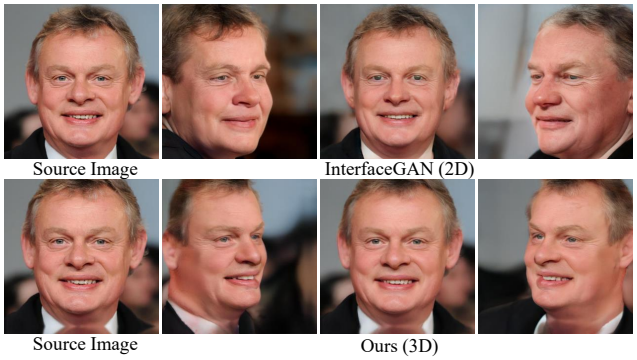

Figure 1. Comparison of 2D and 3D GAN inversion along with viewpoint change.

References

[1] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *CVPR*, 2020. 1

[2] Daniel Roich, Ron Mokady, Amit H Bermano, and Daniel Cohen-Or. Pivotal tuning for latent-based editing of real images. *arXiv preprint arXiv:2106.05744*, 2021. 3, 4

[3] Yujun Shen, Jinjin Gu, Xiaoou Tang, and Bolei Zhou. Interpreting the latent space of gans for semantic face editing. In *CVPR*, 2020. 1

[4] Kaisiyuan Wang, Qianyi Wu, Linsen Song, Zhuoqian Yang, Wayne Wu, Chen Qian, Ran He, Yu Qiao, and Chen Change Loy. Mead: A large-scale audio-visual dataset for emotional talking-face generation. In *ECCV*, 2020. 1, 2

Work done during an internship at Tencent AI Lab.
†Corresponding Author.

Method	View	MSE ↓	LPIPS ↓	MS-SSIM ↓	ID ↑
$\mathcal{W}^+$ Inversion	F	0.04853	0.3358	0.2681	0.8124
+ Symmetry Prior		0.04119	0.3253	0.2531	0.8148
+ Joint Optimization		0.03474	0.3161	0.2210	0.8364
+ Geometry Regularization		0.03315	0.3123	0.2158	0.8363
+ Warping Pseudo (Ours)		**0.03296**	**0.3088**	**0.2135**	**0.8388**
$\mathcal{W}^+$ Inversion	L30	0.05158	0.3286	0.2659	0.8111
+ Symmetry Prior		0.04278	0.3002	0.2375	0.8245
+ Joint Optimization		0.03321	0.2827	0.2054	0.8457
+ Geometry Regularization		**0.03303**	0.2828	0.2053	**0.8481**
+ Warping Pseudo (Ours)		0.03399	**0.2796**	**0.2025**	0.8469
$\mathcal{W}^+$ Inversion	L60	0.08951	0.4200	0.3485	0.7421
+ Symmetry Prior		0.04824	0.3251	0.2633	0.8202
+ Joint Optimization		0.04087	0.3144	0.2424	0.8270
+ Geometry Regularization		**0.04032**	0.3134	0.2416	**0.8281**
+ Warping Pseudo (Ours)		0.04069	**0.3113**	**0.2379**	0.8272
$\mathcal{W}^+$ Inversion	R30	0.05888	0.3478	0.2938	0.7987
+ Symmetry Prior		0.03825	0.3013	0.2421	0.8244
+ Joint Optimization		0.03133	0.2820	0.2083	0.8455
+ Geometry Regularization		**0.03134**	0.2817	0.2081	0.8471
+ Warping Pseudo (Ours)		0.03203	**0.2807**	**0.2057**	**0.8529**
$\mathcal{W}^+$ Inversion	R60	0.09239	0.4229	0.3587	0.7461
+ Symmetry Prior		0.05352	0.3361	0.2744	0.8140
+ Joint Optimization		0.04565	0.3166	0.2465	0.8329
+ Geometry Regularization		**0.04488**	0.3161	0.2448	0.8307
+ Warping Pseudo (Ours)		0.04541	**0.3160**	**0.2400**	**0.8335**

Table 1. Quantitative comparison on MEAD [4]. View denotes the yaw angle of the input image. F is frontal, L is left side, and R is right side. 30 and 60 are the rotation degrees. The metrics are calculated between the ground truth and the synthetic images in different views.

Figure 2. Qualitative comparisons with PTI [2] on in-the-wild images.

Figure 3. Qualitative comparisons with PTI [2] on in-the-wild images.