.1 SUPPLEMENTARY

.1.1 GAN INVERSION PERFORMANCE

We use an optimization-based GAN inversion method (Abdal et al., 2019) to find optimal latent vectors that can best reconstruct the real images via the generator. To examine the effect of GAN inversion performance on our approach, we terminated the GAN inversion approach at different training steps, i.e., 500 and 4000 iterations. We show averaged reconstruction MSE loss in Tab. 3. In this case, we reconstructed 20 real face images and edited their "Smile" and the "Blond hair" attribute with 10 different degrees ε, i.e., 200 images in total for each attribute editing. Quantitative evaluation on image identity preservation and numerical changes on the other semantically independent attributes for the reconstructed face images is in Tab. 4. The results in Tab. 4 suggest that the performance of the GAN inversion method affects our method to some degree. Visualized inversion and editing results are shown in Fig. 18. The qualitative results suggest that our method still works remarkably well on the worse inversion image.

Table 3: **Averaged MSE loss of reconstructing 20 real face images.** The GAN inversion method (Abdal et al., 2019) was trained and terminated at 4k and 500 iterations with averaged MSE loss in the right column.

Training iterations	MSE
4k iters	1657.40
500 iters	3096.75

Table 4: **Quantitative evaluation of identity preservation (ID) and attribute changes (Attr)**. We edited two attributes for the real face images, i.e., "Smile" (col.2-7) and "Blond hair" (col.8-13).

	Smile						Blond hair					
	ID (↑)			Attr (↓)			ID (↑)			Attr (↓)		
$\lvert\hat{\varepsilon}\rvert$	(0, .3]	(.3, .6]	(.6, .9]	(0, .3]	(.3, .6]	(.6, .9]	(0, .3]	(.3, .6]	(.6, .9]	(0, .3]	(.3, .6]	(.6, .9]
4k iters	**.997**	**.994**	**.986**	**.077**	**.108**	**.147**	**.988**	**.979**	**.94**	**.1**	**.112**	**.16**
500 iters	.996	.99	.982	.087	.138	.165	.978	.936	.914	.122	.158	.185

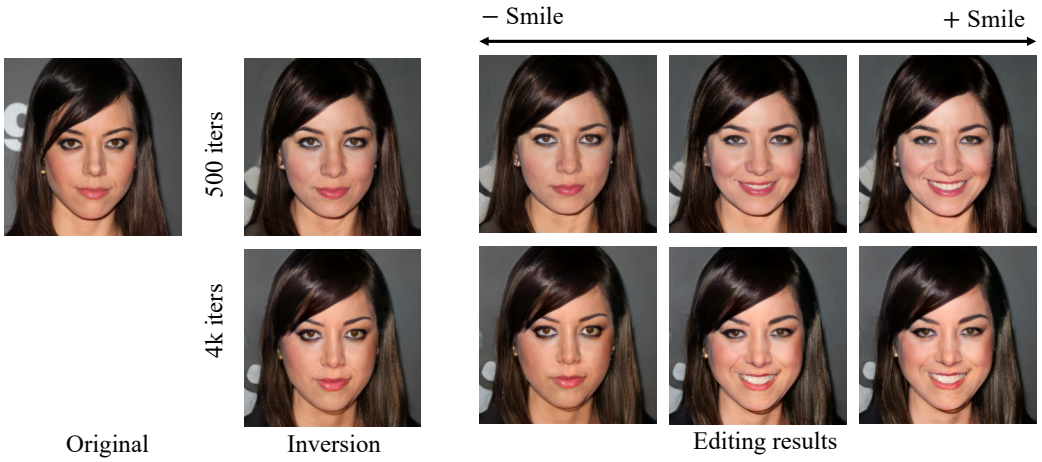
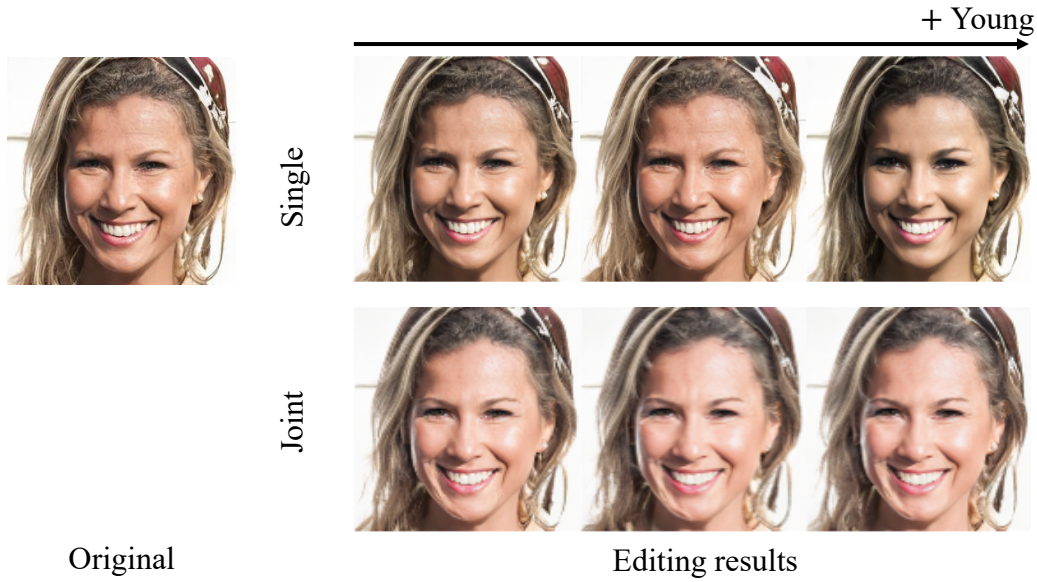


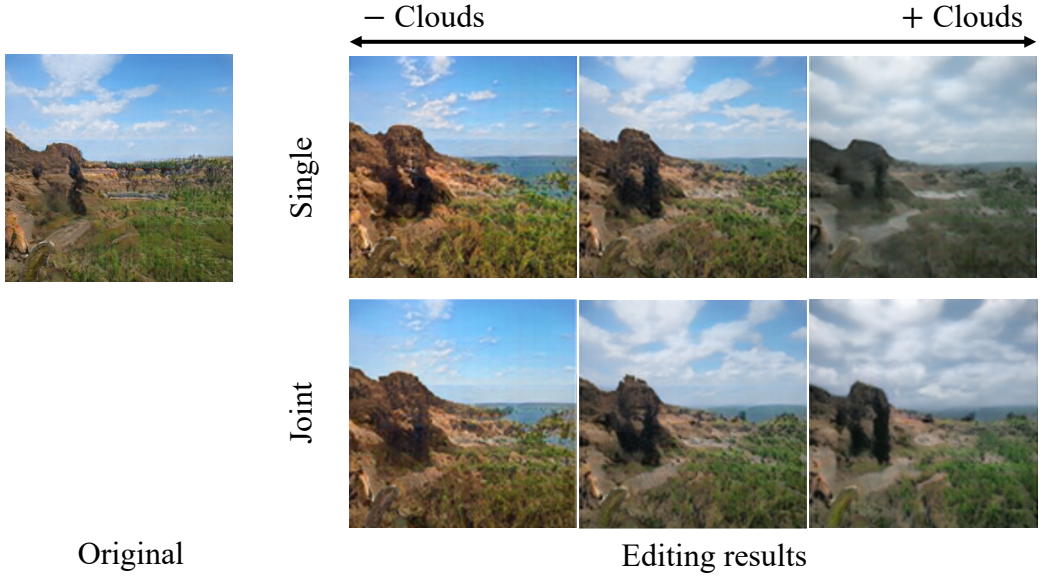
Figure 18: **Visual comparisons of "Smile" editing with different reconstructed images**: (*col.1*) Original image; (*col.2*) reconstructed images with 500 (top) and 4k (bottom) inversion optimization steps; (*col.3-6*) results of editing the "Smile" attribute.

.1.2 ABLATION STUDY

We conduct an ablation study with single- and joint-distribution training strategies in our method. Specifically, single-distribution sampling refers to learn one attribute direction at a time. In contrast, joint-distribution training means to train multiple attribute directions simultaneously. Fig. 19 shows the visualized results with the two training strategies on the face and the scene dataset. We observe that the model with the single-distribution training strategy may generate more unexpected changes as the manipulation degree is getting large, e.g., darker scene colors by the model trained for a single attribute, shown in Fig. 19 (b).



(a) Visual comparisons of single- and joint-distribution sampling on the face dataset



(b) Visual comparisons of single- and joint-distribution sampling on the scene dataset

Figure 19: **Visual comparisons of single- and joint-distribution training.** Single-distribution training refers to train one attribute direction at a time, while joint-distribution training means to train multiple attribute directions simultaneously.