DynamicFace: High-Quality and Consistent Face Swapping for Image and Video using Composable 3D Facial Priors

Supplementary Materials

1. Detailed Ablation Study on Facial Composable Conditions

To rigorously validate the necessity of each facial condition, we conduct additional ablation experiments exclusively on the FaceForensics++ (FF++) dataset. We systematically evaluated the impact of each composable facial condition by training the model while removing individual components on purely public dataset, ensuring reproducibility. Crucially, Table 1 validated that no single condition could be removed without significant performance degradation, confirming their complementary yet disentangled contributions. Figure 2 shows the role of disentangled conditions and their collective effect.

Methods	ID Simi. ↑	Pose↓	Expr.↓	Mouth↓	Eye↓
w/o. landmark	55.45	1.66	3.46	2.59	0.47
w/o. normal	56.32	1.93	3.13	2.96	0.43
w/o. lighting	56.30	**1.34**	2.89	2.27	0.46
w/o. bg	**73.05**	2.43	4.01	2.66	0.48
Ours	56.22	1.46	**2.60**	**2.06**	**0.33**

Table 1. Ablation study on different facial conditions.

2. Why Disentanglement Matters

The core strength of DynamicFace lies in its disentangled facial conditions, which enable independent control over identity, motion, and environmental attributes. We explore more possible applications to validate this:

Enhancing shape similarity of ID Preserving Text-to-Image Generation. Despite advancements in identity-preserving methods, existing approaches may still exhibit suboptimal ID similarity and inadequate face shape control. Shape-aware normal maps of DynamicFace anchor facial geometry to the source identity, enabling localized repairs (e.g. jaw realignment) without distorting the target's expression or pose. We first generate portraits in different styles using ID Preserving methods and then enhance the generated results. Our method resolves identity leakage in the jaw and eye shape as shown in Figure 1.

Motion-Consistent Artifact Restoring. By decoupling shape, expression, background, and illumination, DynamicFace could selectively restore corrupted regions without altering unrelated attributes (e.g., fixing a torn texture while preserving eye blinks). Our disentangled facial conditions enable precise control over motion attributes (e.g., expression, pose) while preserving the identity of the source face. This allows us to address facial artifacts in the results from other portrait generation works . Expression-related landmarks ensure repaired regions align with the original motion dynamics as shown in Figure 3, 4.

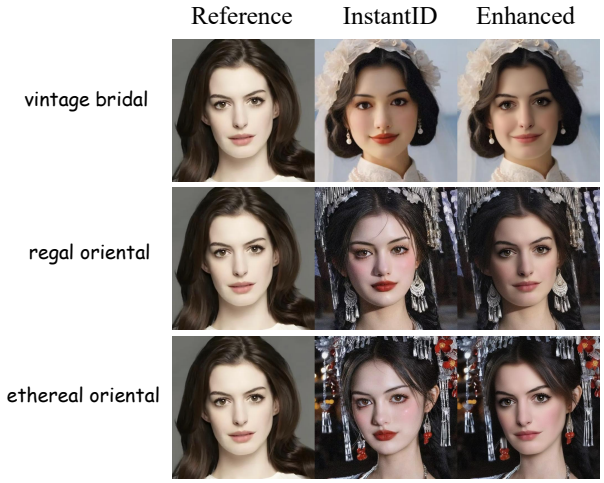

Figure 1. **Enhance results genearted by ID-Preserving works.**

3. Extended Real World Visualizations

We provide additional visual results on real-world videos to demonstrate robustness under challenging scenarios: extreme expressions and occlusions. The videos are sorted out together in the supplementary materials, including visual comparison with latest video face swapping methods.

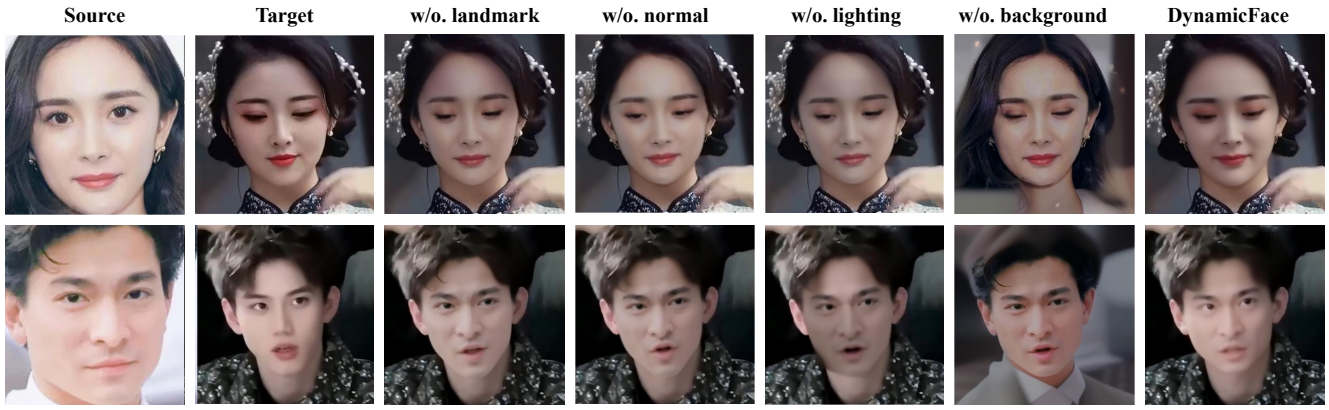

Figure 2. **Impact of disentangled facial conditions on face swapping.**

Figure 3. **Disentangled priors enable motion-consistent facial artifact correction in generated videos**

Figure 4. **Disentangled priors enable motion-consistent facial artifact correction in generated videos**

Figure 5. **Celebrity face swapping results under varied illumination and pose conditions**

Figure 6. **Celebrity face swapping results under varied illumination and pose conditions**