DreamActor-M1: Holistic, Expressive and Robust Human Image Animation with Hybrid Guidance

Supplementary Material

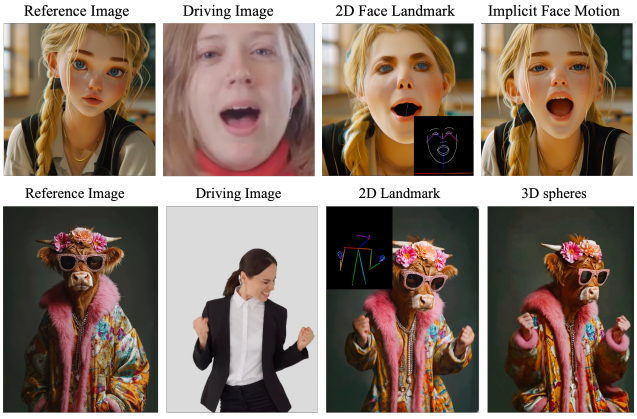

Figure 1. Qualitative comparisons of baseline (skeleton maps) and baseline + control signals.

A. Additional Ablation Details for Hybrid Control Signals

For diffusion-based image animation, the conditioning strategy proves pivotal. Our key contribution lies in extracting information-rich, decoupled control signals from driving videos. Here, we supplement additional visualizations and numerical results to demonstrate the effectiveness of our hybrid control signals. We provide quantitative comparison results of the baseline (Seaweed model + skeleton maps) and baseline plus various control signals in Tab. 1, and qualitative comparison results in Fig. 1. The results clearly show that the combination of implicit face motion and head sphere controls achieves superior performance in both identity preservation and motion fidelity. To further verify that our superior results stem from the hybrid control signal design rather than the seaweed backbone, we also compare our method with UniAnimate-DiT in Fig. 2, whose backbone is the superior Wan2.1-14B-I2V model.

B. Cross-dataset Generalization

We conduct quantitative comparisons with both pose transfer and portrait animation methods, as reported in Tab. 1 and Tab. 2 (with 20 randomly selected samples per test dataset). Notably, our method achieves state-of-the-art performance on public benchmarks, demonstrating its consistent superiority in multiple evaluation settings.

Figure 2. Our method vs UniAnimate-DiT.

Methods	FID↓	SSIM↑	PSNR↑	LPIPS↓	FVD↓
Baseline	33.13	0.757	21.70	0.222	131.6
+ Facial Reps	27.66	0.818	23.16	0.209	123.4
+ Head Sphs	**27.27**	**0.821**	**23.93**	**0.206**	**122.0**

Table 1. Ablation study on our collected dataset.

Methods	FID↓	SSIM↑	PSNR↑	LPIPS↓	FVD↓
AA	53.34	0.707	26.72	0.244	161.3
Champ	52.61	0.710	26.87	0.246	159.8
MimicMotion	50.83	0.739	27.82	0.231	155.2
DisPose	42.78	0.805	28.91	0.228	150.9
Ours	**31.63**	**0.820**	**29.22**	**0.211**	**125.7**

Table 2. Quantitative comparisons with pose transfer methods on Fashion-Video dataset.

Methods	FID↓	SSIM↑	PSNR↑	LPIPS↓	FVD↓
LivePortrait	30.34	0.791	22.66	0.269	151.1
X-Portrait	30.11	0.784	23.71	0.272	148.5
SkyReels-A1	29.30	0.809	24.02	0.263	144.3
Act-One	28.81	0.807	24.91	0.252	135.9
Ours	**25.24**	**0.818**	**27.36**	**0.224**	**112.0**

Table 3. Quantitative comparisons with portrait animation methods on VoxCeleb dataset.