

MikuDance: Animating Character Art with Mixed Motion Dynamics

Supplementary Material

A. Supplementary on Experiments

Quantitative comparisons on foreground-only results.

The quantitative results of previous methods are primarily evaluated on videos with static backgrounds and simple human motions. In contrast, our test dataset includes high-dynamic motions in both the foreground and background. To provide a clearer evaluation, we report additional comparisons on foreground-only results in Table 1, which remove complex backgrounds using a segmentation mask generated by BiRefNet (CAAI-AIR'24) and focus solely on the quality of character animation. The results show that MagicPose outperforms AniAny in character quality, as it better preserves the character's appearance. Among all methods, our MikuDance consistently achieves superior results across all metrics.

Scene motion visualization. Figure 1 presents a visualized example of scene motion, depicted using directed line segments (a) and dense flow (b).

Animating video frames. Since quantitative evaluation requires ground-truth video to calculate metrics, we collected an unseen test video dataset and used the first frames as reference images. Additionally, most results reported in prior works, such as Animate Anyone and DISCO, involve the animation of video frames where the reference image is naturally aligned with the motion guidance. Examples of video frame animations generated by MikuDance are illustrated in Figure 2. These results show that, when handling the simpler task of animating video frames, MikuDance exhibits stable, high-quality performance even with complex and high-dynamic motion guidance. Notably, the primary motivation behind MikuDance is to animate in-the-wild character art, offering broader application potential compared to prior works focused on animating video frames.

Additional results on high-dynamic motions. In Figure 3, we present additional animation results generated by MikuDance under high-dynamic motion guidance, which in-

cludes large pose variations, extensive camera movements, turning around, and incomplete body postures. These results further demonstrate MikuDance's robust ability to handle complex character and scene motion.

Additional results on various shapes and styles. MikuDance demonstrates strong generalizability across a wide range of characters and art styles. Additional results are illustrated in Figure 4. Remarkably, even in extremely challenging cases, such as the one shown in the third row of Figure 4, MikuDance consistently generates vivid animations that preserve both the character's appearance and the scene structure.

B. Demo Videos and User Study Details

We provide two demo videos. MikuDance-Results.mp4 showcases the animation results presented in the paper. Notably, MikuDance-Queencard.mp4 is a complete Music Video with animations fully generated by MikuDance, using reference character art created by SDXL-Neta.

Figure 5 shows an example of the user study webpage. In one case, 50 volunteers were invited to rank four anonymous results based on three carefully defined evaluation criteria. We provided 20 videos for the volunteers to evaluate, and the average time required to complete the entire questionnaire was over 30 minutes.

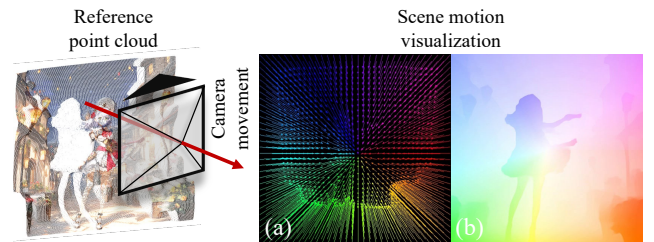


Figure 1. Visualization of the scene motion.

Methods	Image					Video	
	FID \downarrow	SSIM \uparrow	PSNR \uparrow	LISPIS \downarrow	L1 \downarrow	FID-VID \downarrow	FVD \downarrow
AniAny	27.927	0.625	14.831	0.431	3.307E-05	15.197	326.842
AniAny*	17.037	0.801	17.194	0.202	2.284E-05	7.546	211.384
DISCO	31.221	0.533	13.155	0.518	4.824E-05	22.828	564.892
MagicPose	25.248	0.695	15.240	0.362	3.014E-05	13.248	299.912
UniAnimate	29.818	0.601	14.147	0.448	3.902E-05	18.416	381.485
MikuDance (Ours)	14.835	0.846	17.809	0.137	2.060E-05	6.765	194.124

Table 1. Quantitative comparisons on foreground-only results.

Reference

MikuDance (Ours)



Figure 2. Supplementary experiments on animating video frames.

Reference



MikuDance (Ours)



Figure 3. Additional results on high-dynamic motions.

Reference



MikuDance (Ours)



Figure 4. Additional results on various shapes and styles.

User Study for Character Art Animation


This project aims to **generate animations for anime character art**. Specifically, we use 2D human pose sequences as guidance and any anime character art as a reference to generate motion videos of the character, with the entire process completed by an AI model.

We can evaluate the quality of the generated animation through the following three aspects:


1. Overall Quality: Is the animation coherent, aesthetically pleasing, and does it match the characteristics of the character?
2. Frame Quality: Are individual frames clear and reasonable? Do the character's movements follow the pose guidance conditions?
3. Temporal Quality: Is the video temporally smooth, consistent, and do the transitions between movements flow naturally?

The following video contains **20 sets** of character art and their corresponding generated animations. Please watch the video carefully and rank the quality of Animations A, B, and C. The three animations will play sequentially. **Rank the best quality option first.**


* 01 Overall Quality of Video 1



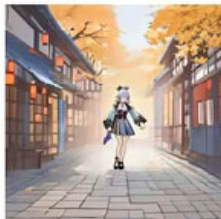
Reference image




Pose Guidance




A



B



C



D

A	⋮
B	⋮
C	⋮
D	⋮

Figure 5. An example of our user study webpage.