

MultiAnimate: Pose-Guided Image Animation Made Extensible

Supplementary Material

001 1. t-SNE Analysis of Latent Tokens

002 To illustrate the discriminatory power of our method, we
003 visualize the feature distribution within the latent space. We
004 extract the patchified latent tokens corresponding to each
005 individual by leveraging their respective tracking masks.
006 These person-specific latent features are then subjected to
007 dimensionality reduction using t-SNE for a clear 2D repre-
008 sentation.

009 As depicted in Fig. 1, the left panel showcases the
010 results achieved with our proposed method, while the
011 right panel presents the corresponding visualization for
012 UniAnimate-DiT. The distinct, well-clustered groups in our
013 result demonstrate that our framework effectively disentangles
014 the identities in the latent space, thereby preventing
015 person confusion during multi-person animation.

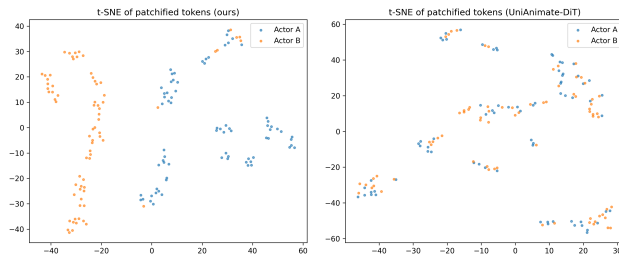


Figure 1. t-SNE visualization of patchified latent tokens. Left: Our method. Right: UniAnimate-DiT. Our approach clearly distinguishes latent features for different individuals, indicating successful identity disentanglement.

016 2. Extended Multi-Actor Comparative Results

017 We present additional qualitative comparisons focusing on
018 multi-actor image animation.

019 Given that previous evaluations demonstrated consis-
020 tently weaker performance for Dispose and MimicMotion,
021 we primarily focus our extended comparative results on
022 VACE and UniAnimate-DiT. Our method consistently ex-
023 hibits superior performance.

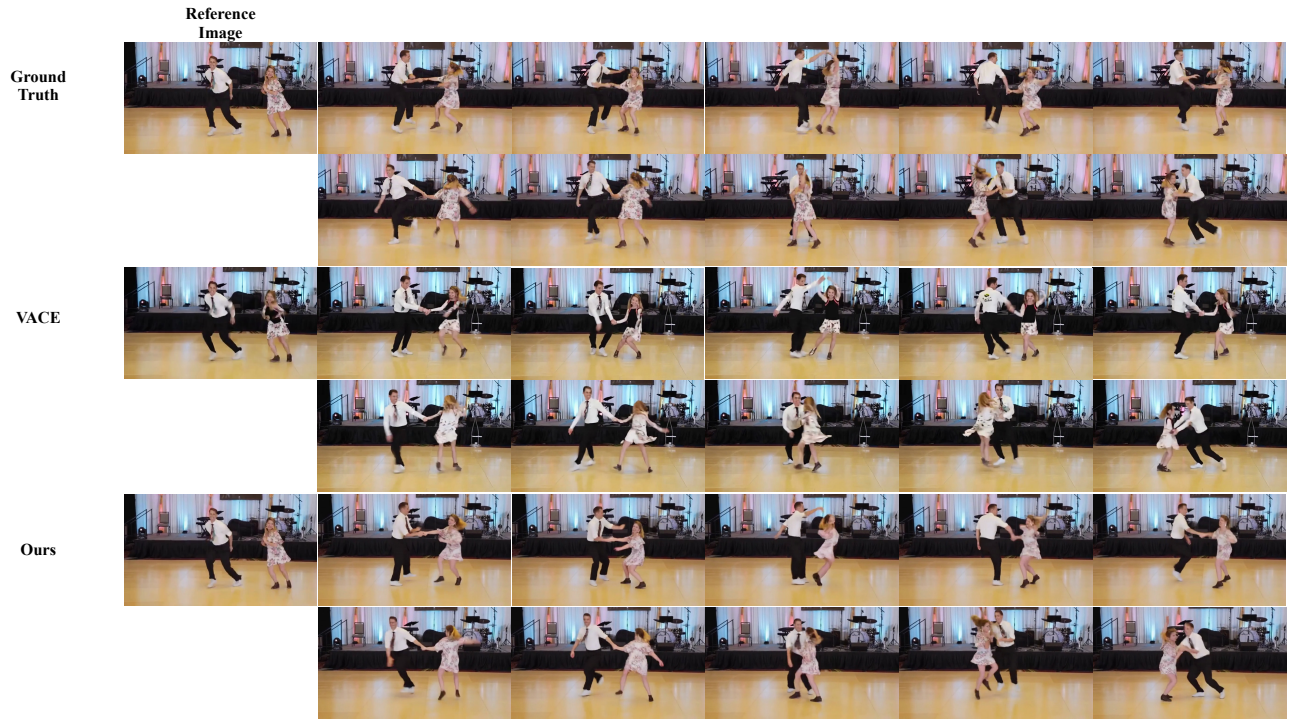


Figure 2. More comparing examples.



Figure 3. More comparing examples.

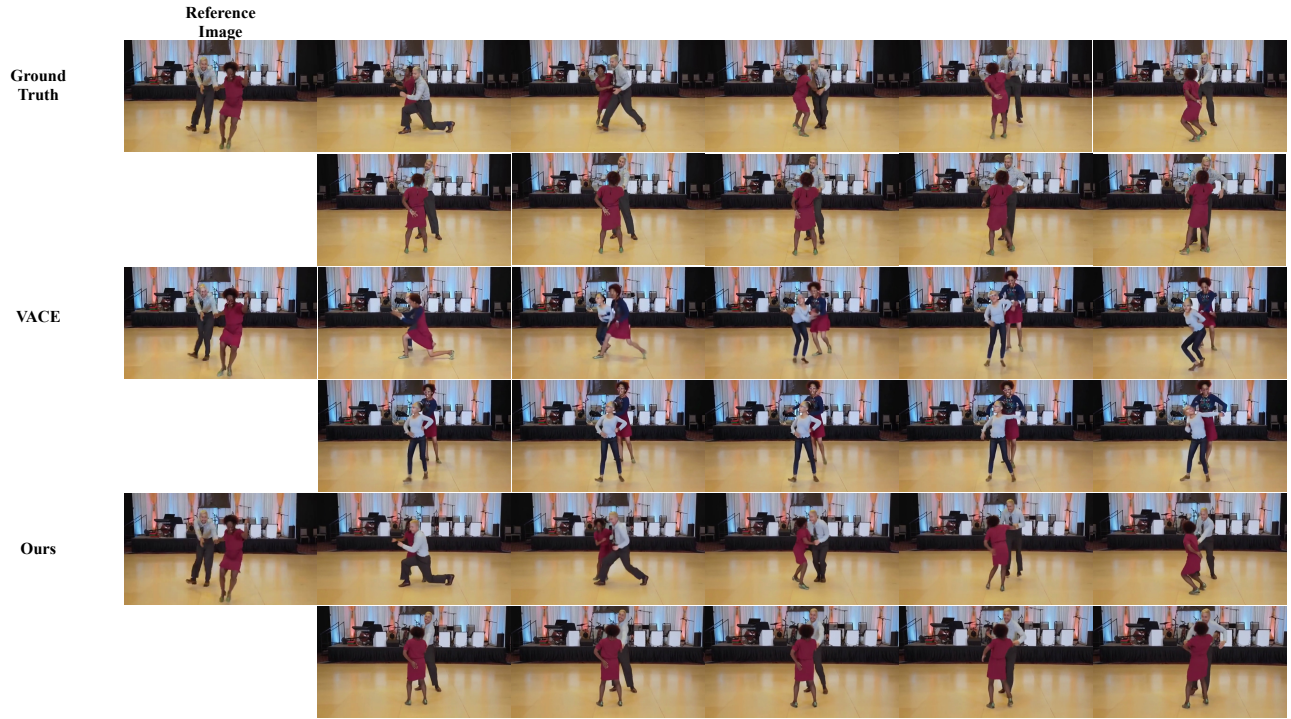


Figure 4. More comparing examples.

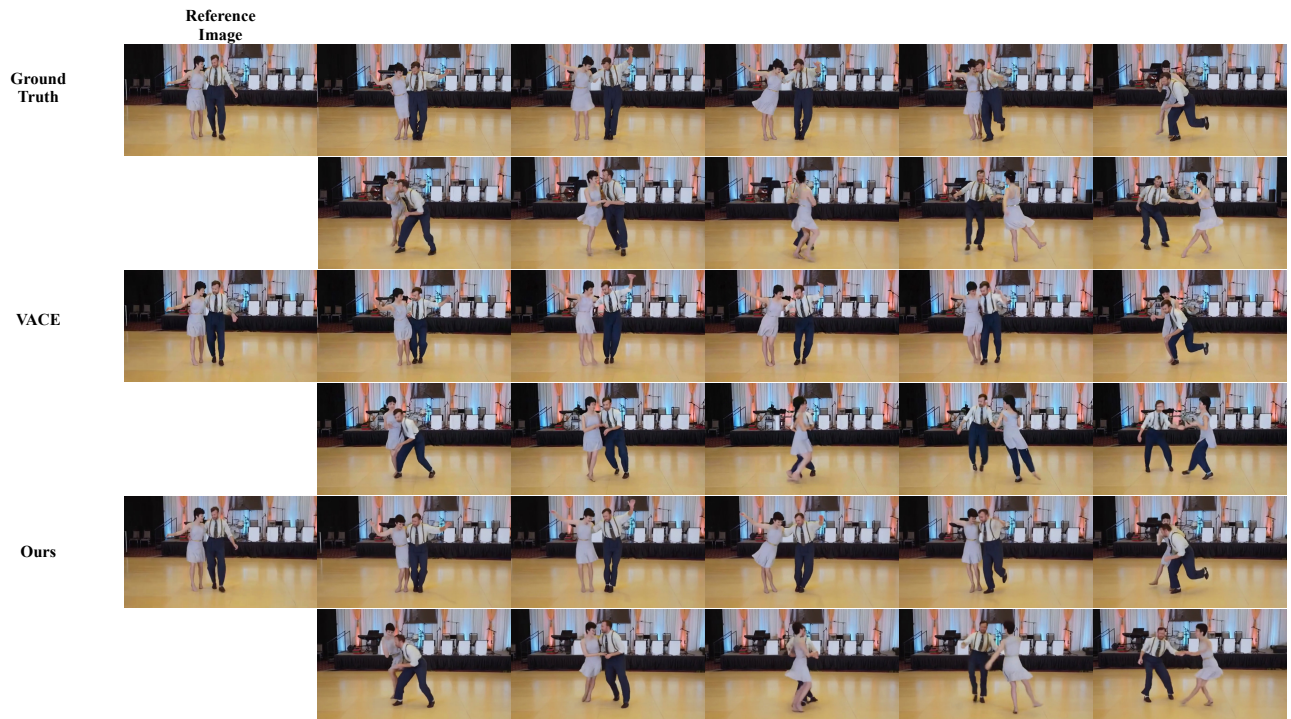


Figure 5. More comparing examples.



Figure 6. More comparing examples.



Figure 7. More comparing examples.

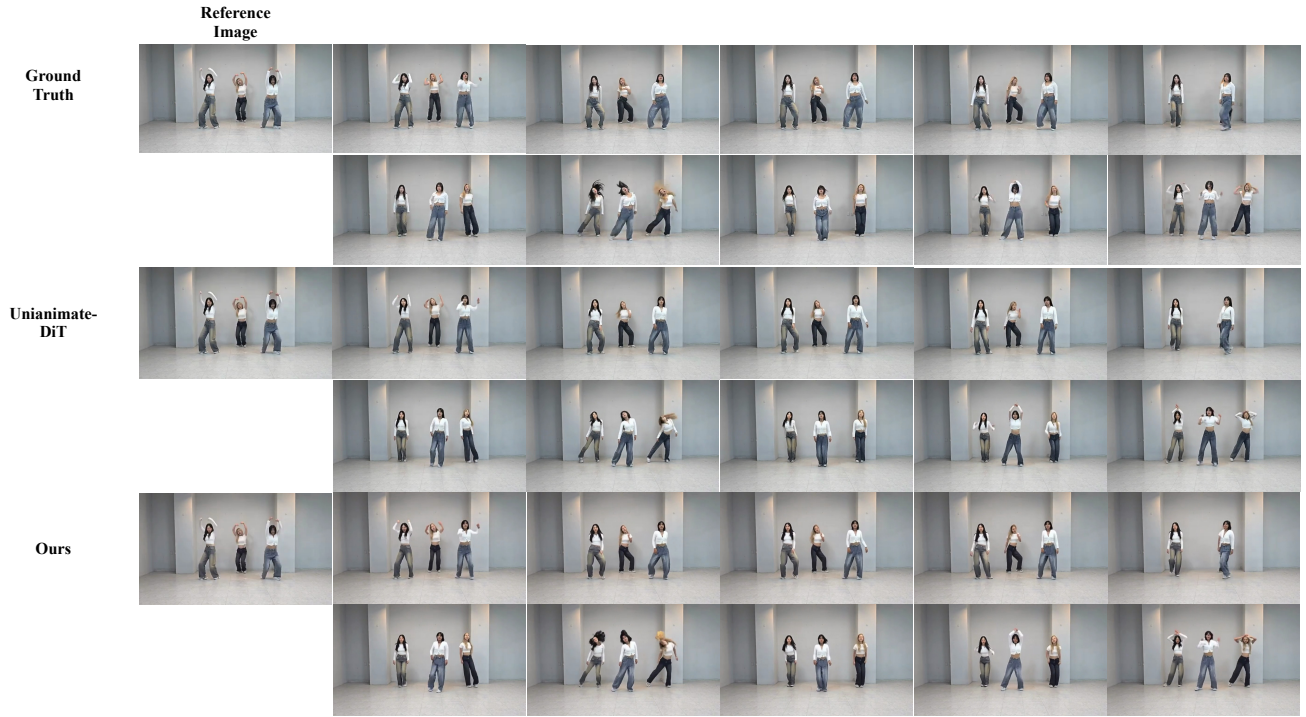


Figure 8. More comparing examples.

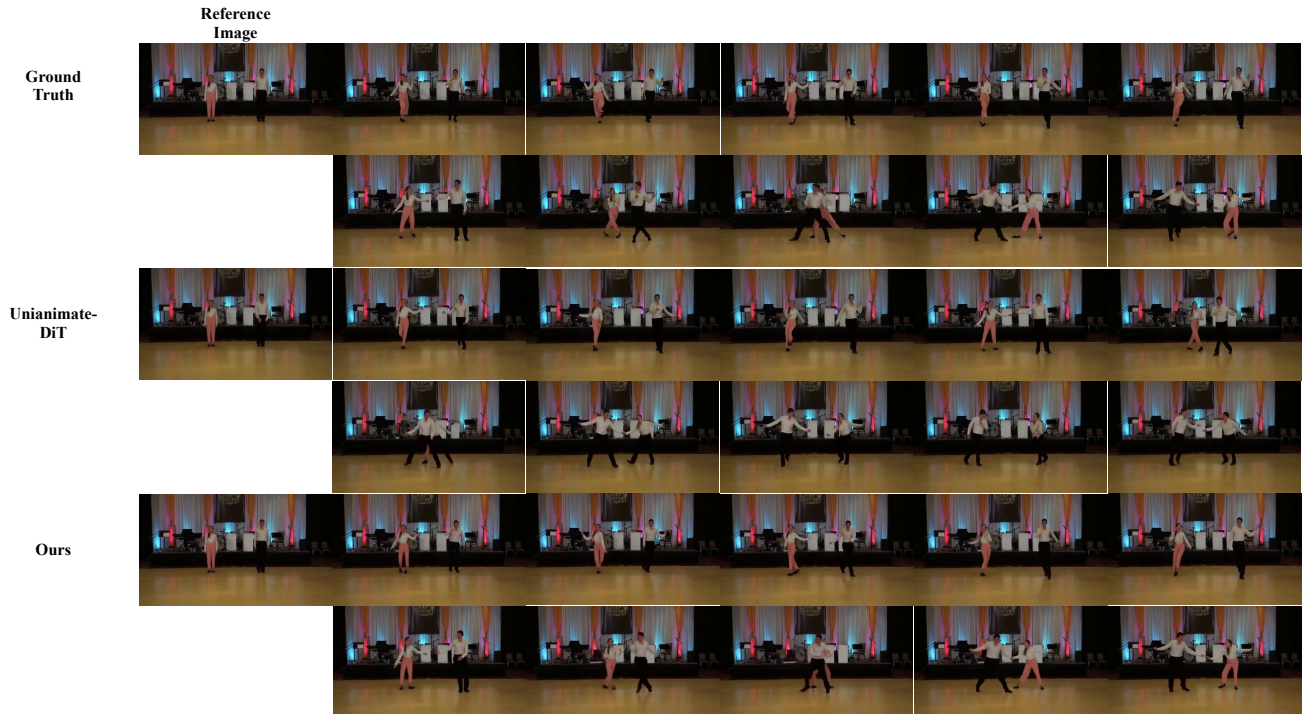


Figure 9. More comparing examples.