

MimicMotion: High-Quality Human Motion Video Generation with Confidence-aware Pose Guidance

Yuang Zhang^{†‡}, Jiayi Gu[†], Li-Wen Wang[†], Han Wang^{†‡},
Junqi Cheng[†], Yuefeng Zhu[†], Fangyuan Zou[†]

[†]*Tencent* [‡]*Shanghai Jiao Tong University*

Tencent 腾讯



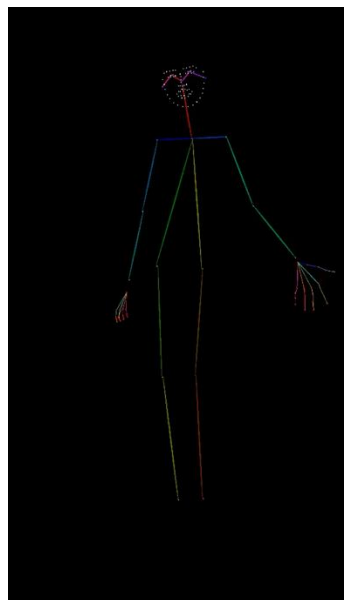
上海交通大学
SHANGHAI JIAO TONG UNIVERSITY

Introduction

Reference image



Pose sequences



MimicMotion

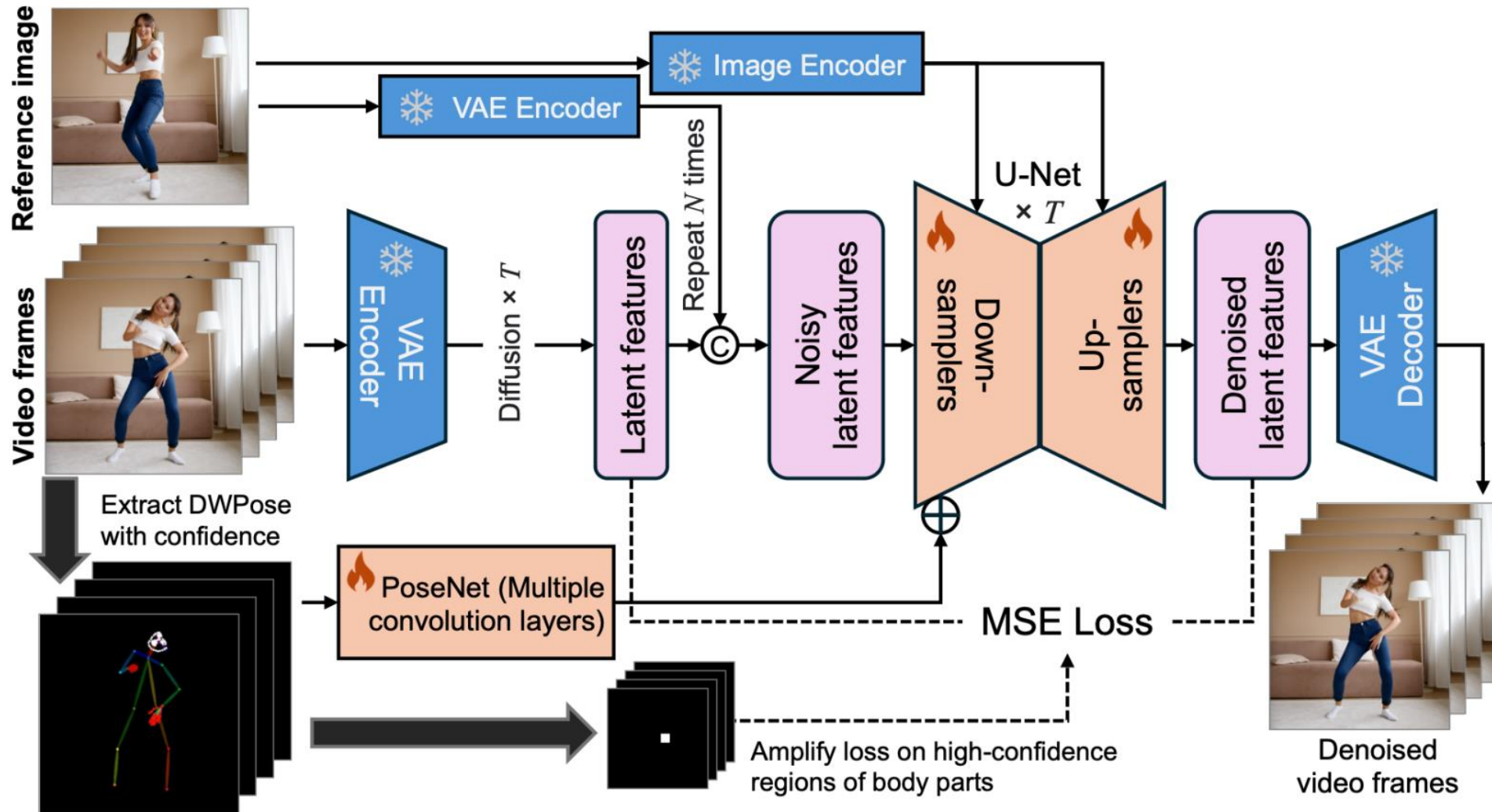
Generated video



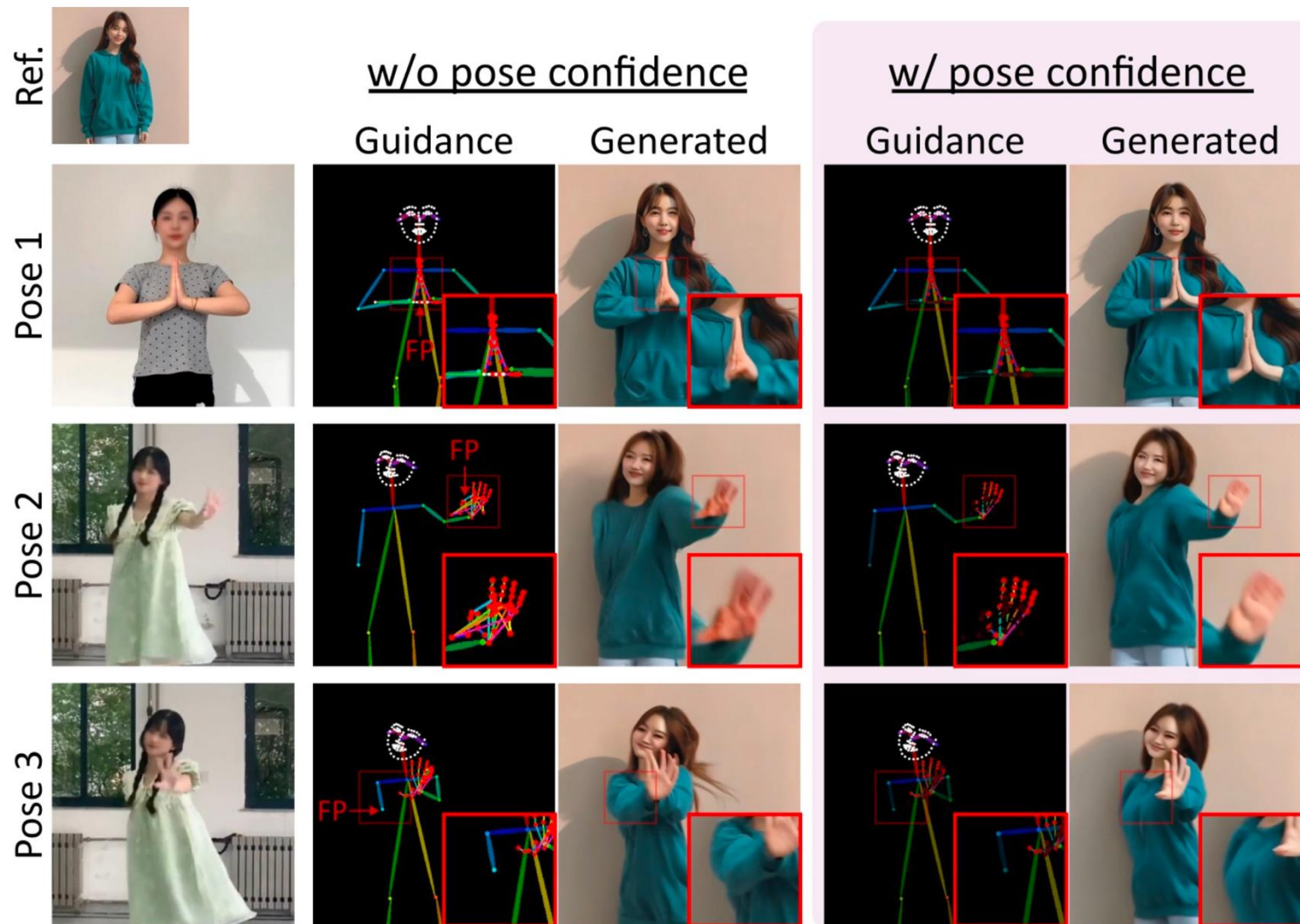
Highlights of MimicMotion:

1. Confidence-aware pose guidance: high frame quality and temporal smoothness.
2. Regional loss amplification: reduces image distortion in key regions.
3. Progressive latent fusion: long and smooth videos.

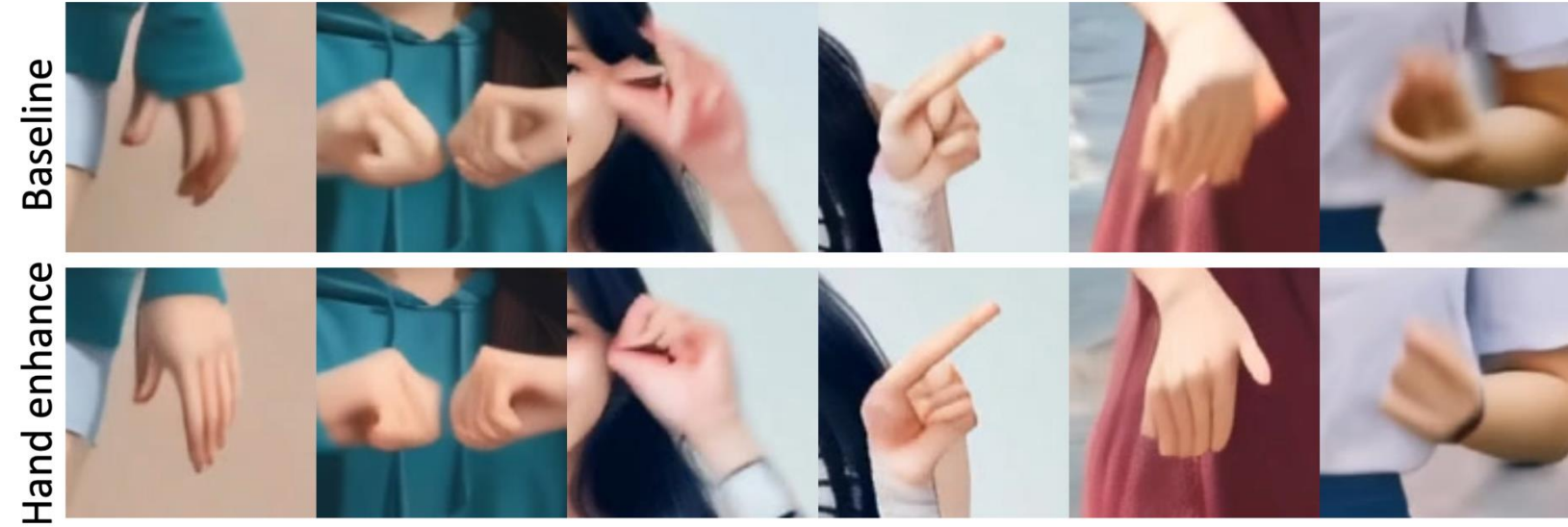
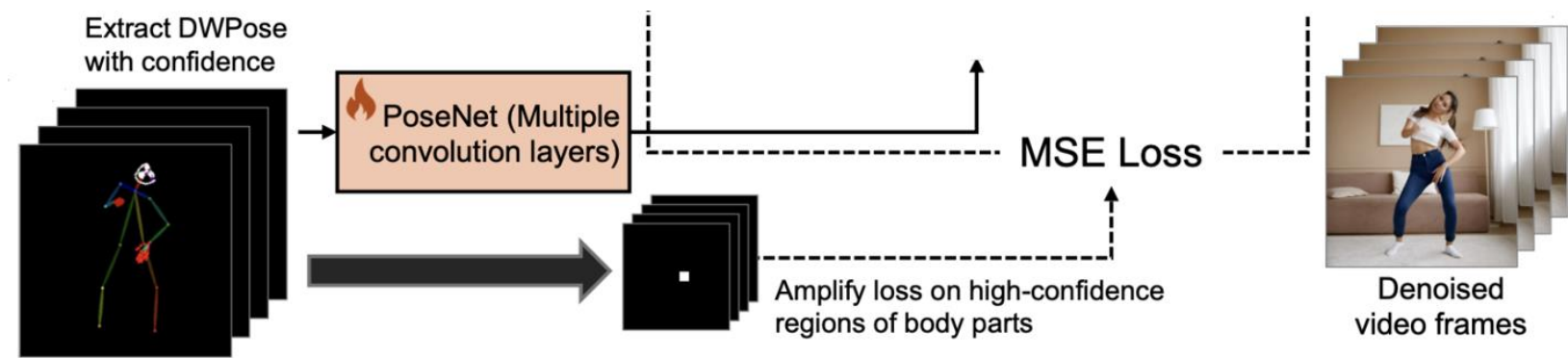
Pose-Guided Video Diffusion Model



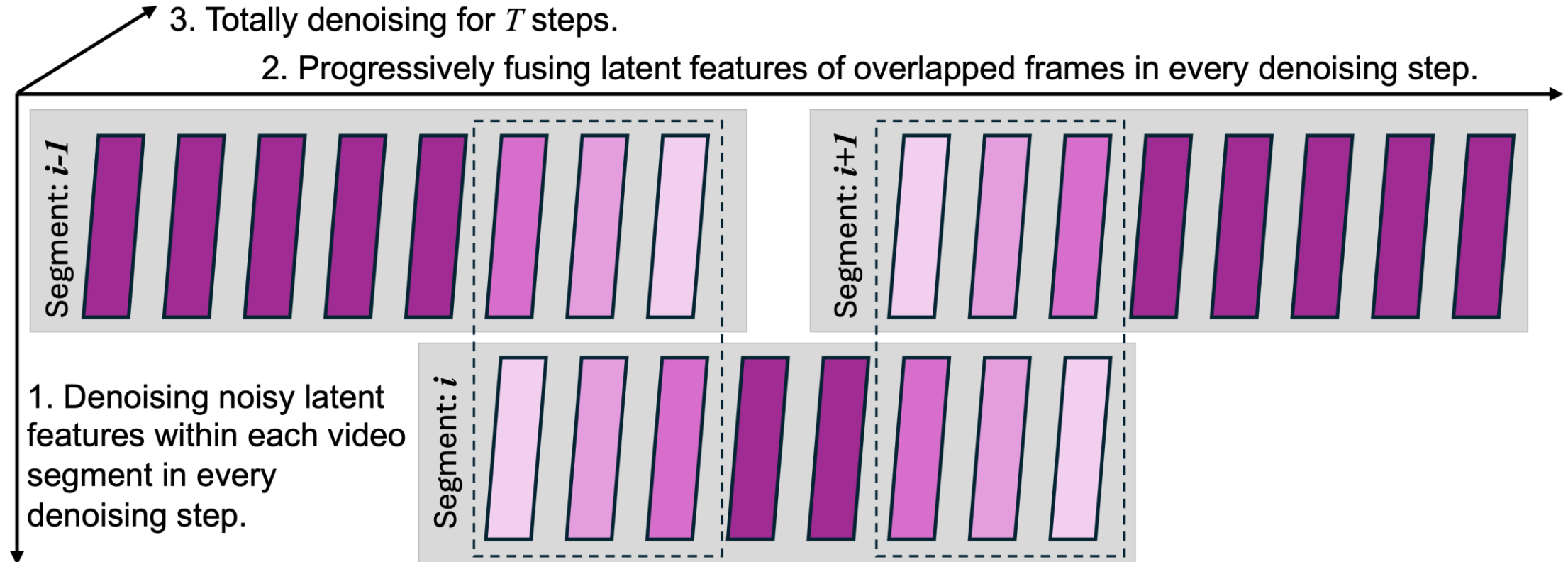
Confidence-aware Pose Guidance



Region-specific hands refiner



Progressive Latent Fusion



Progressive Latent Fusion

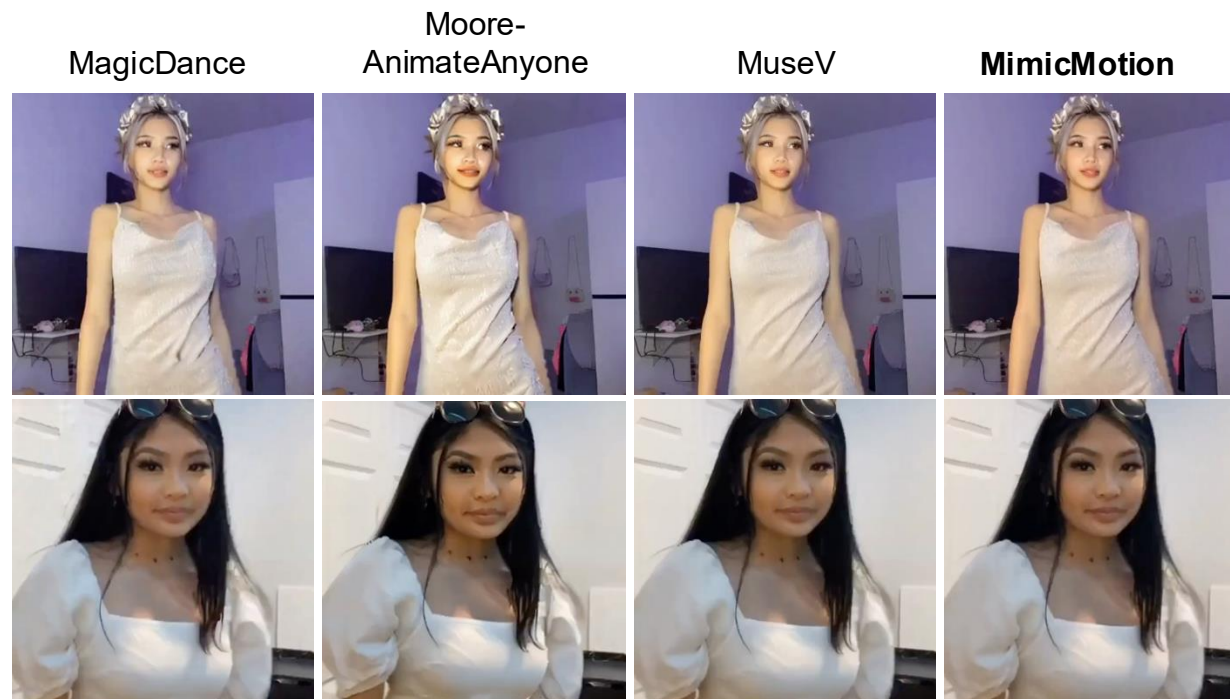
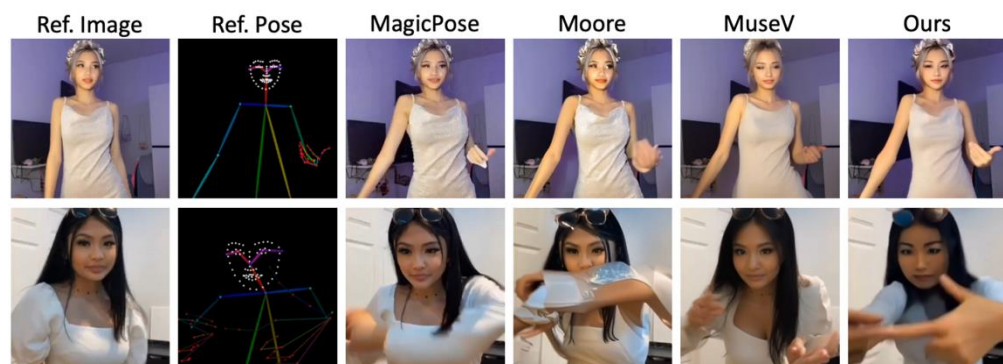


(a) Without progressive latent fusion.

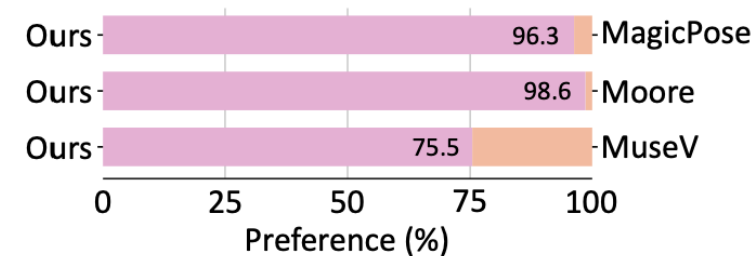


(b) With progressive latent fusion.

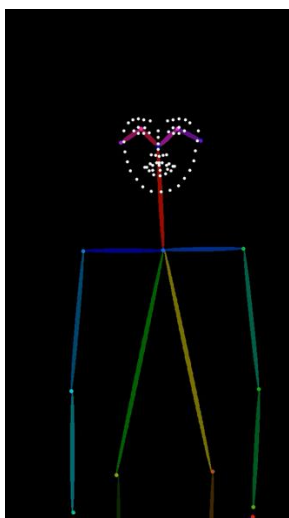
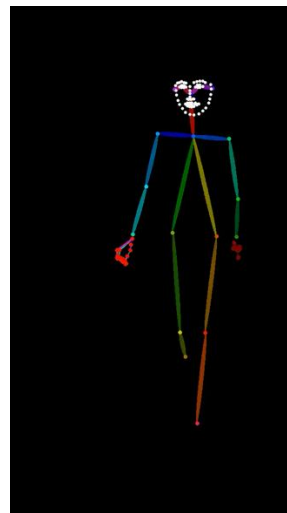
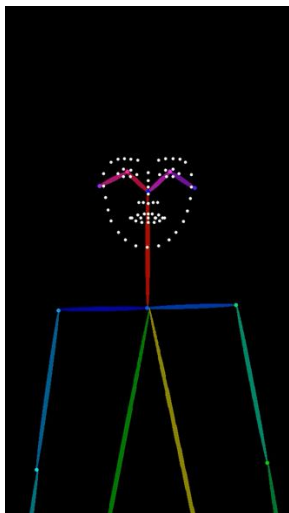
Experiments



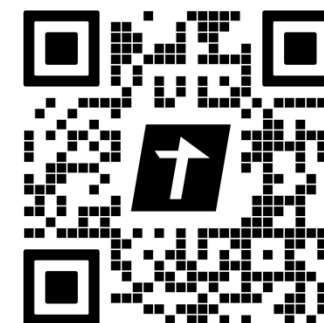
Method	FID-VID↓	FVD↓
MagicPose [4]	13.3	916
Moore [38]	12.4	728
MuseV [39]	14.6	754
AnyMotion (ours)	9.3	594



Showcases



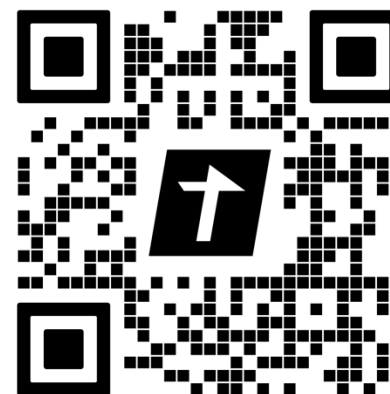
*More showcases can be
found on our project page:*



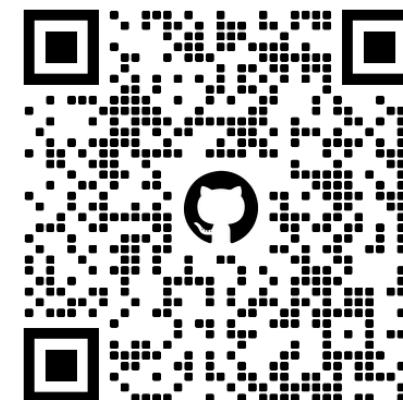
The End

Thank You!

- Project page: <https://tencent.github.io/MimicMotion>
- Code page: <https://github.com/Tencent/MimicMotion>
- Contact: imjiaxi@gmail.com



Project page



Code page