MDMP: Multi-modal Diffusion for supervised Motion Predictions with uncertainty

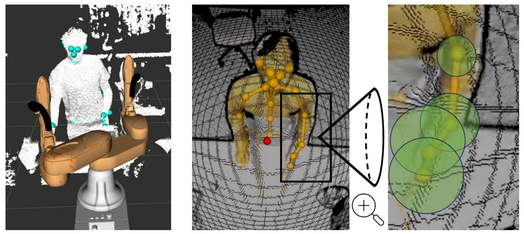
Leo Bringer, Joey Wilson, Kira Barton, Maani Ghaffari
University of Michigan

CVPR Nashville JUNE 11-15, 2025

Motivation:

Contributions:

- Diffusion model guided by **text & skeletal data** for accurate long-term motion prediction.
- Associated **spatial uncertainty estimation** per joint and timestep.
- **Graph-based transformer** capturing spatial-temporal motion dynamics efficiently.

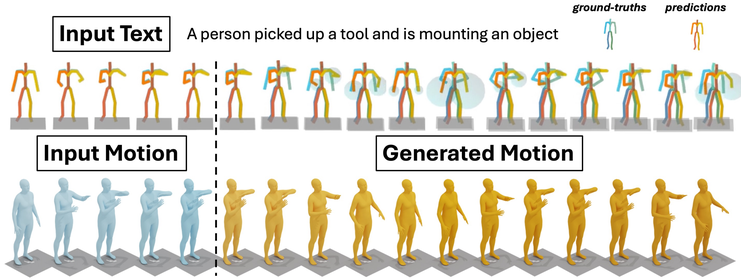


In many human-robot collaborative settings such as industrial robotics or self-driving cars **estimating uncertainty could significantly improve spatial awareness and safety** by incorporating zones of presence with varying confidence levels.

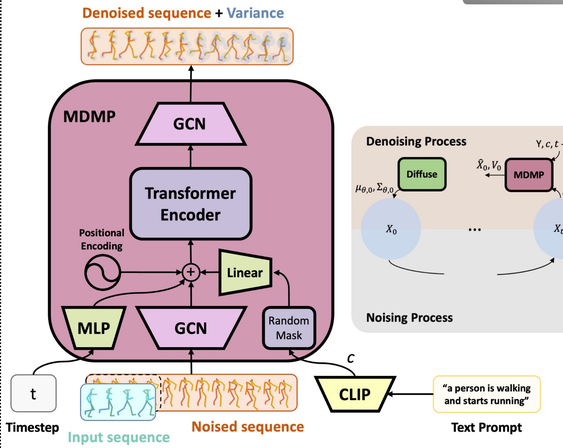
Method:

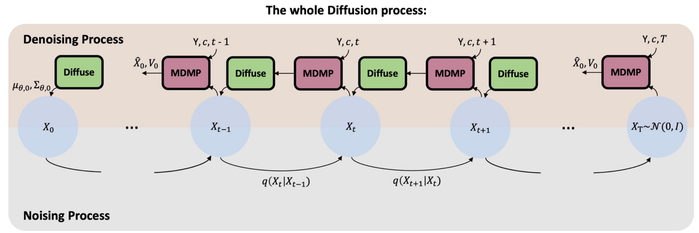
Summary

Our method integrates **both skeletal motion & text** to generate **long-term** motion predictions **with uncertainty:**

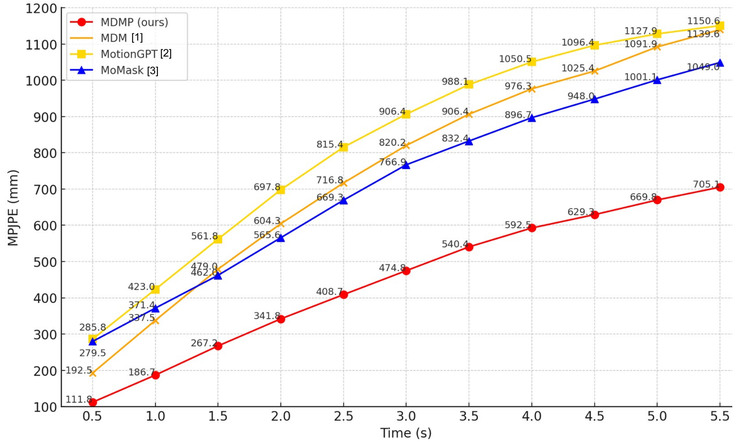


Architecture Overview:



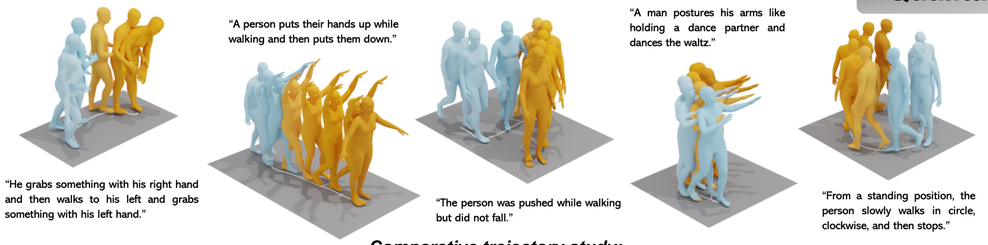


Quantitative Results:



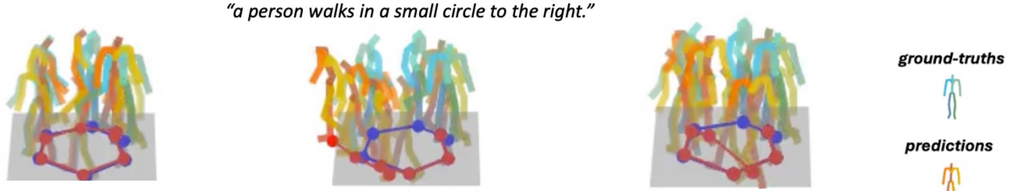
Qualitative Results:

Some visual results:

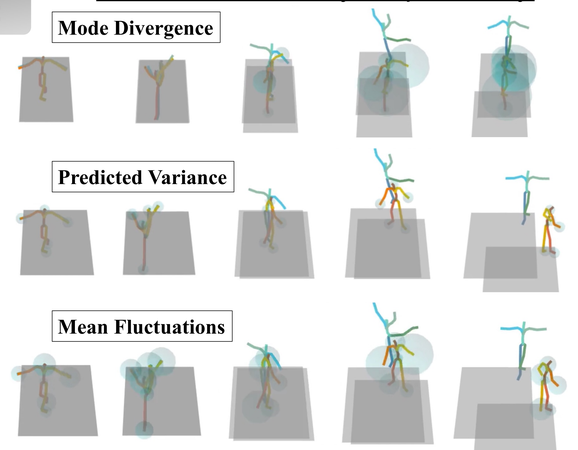


Comparative trajectory study:

"a person walks in a small circle to the right."



Different index for uncertainty - Comparative study:



[1] Guy Tevet, et al. Human motion diffusion model. *ICLR*, 2023
[2] Biao Jiang et al. Motiongpt: Human motion as a foreign language. *NeurIPS*, 2023
[3] Chuan Guo, et al. Momask: Generative masked modeling of 3d human motions. *CVPR*, 2024