UniCon3R: Unified Contact-aware 4D Human-Scene Reconstruction from Monocular Video

Tanuj Sur[1], Shashank Tripathi[2], Nikos Athanasiou[2], Ha Linh Nguyen[1], Kai Xu[1], Michael J. Black[2], Angela Yao[1]

[1] National University of Singapore [2]Max Planck Institute for Intelligent Systems, Tübingen, Germany

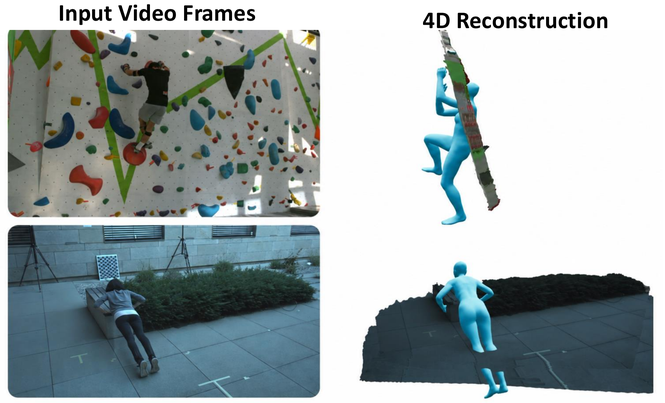
NEURAL INFORMATION PROCESSING SYSTEMS

Goal

- Recover physically grounded 4D humans and scenes from streaming monocular video through contact-aware, online reconstruction.

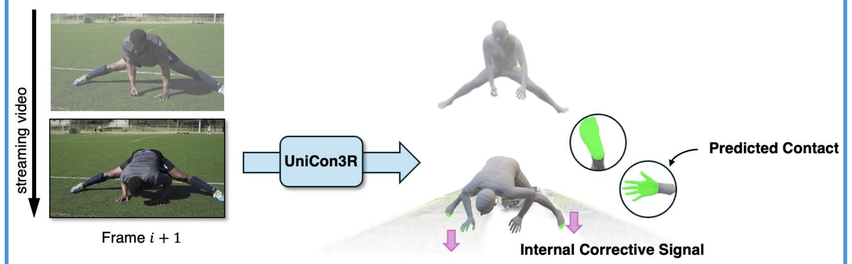
Challenges

- **Physical Plausibility:** Existing methods[1, 2, 3] often produce floating, or penetrating human meshes.
- **Ambiguous Contact:** Human-scene contact is difficult to infer from monocular.



Key Insight

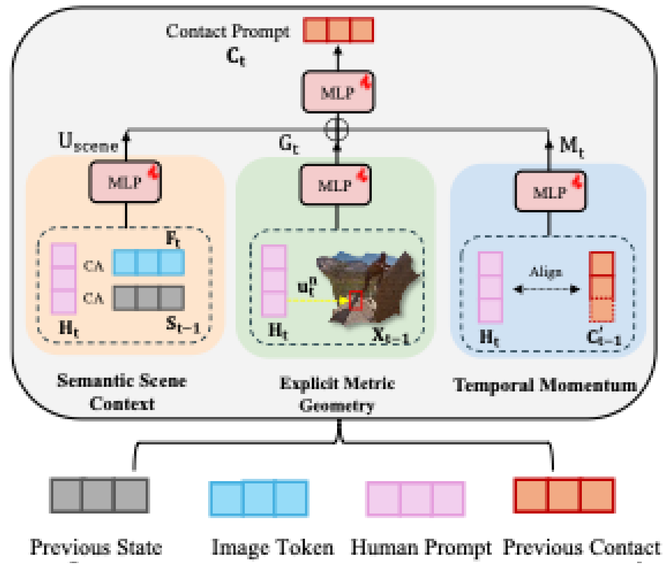
- Explicitly modelling human-scene interactions through contact provides physical constraints.



Scene-Aware Contact Prompt

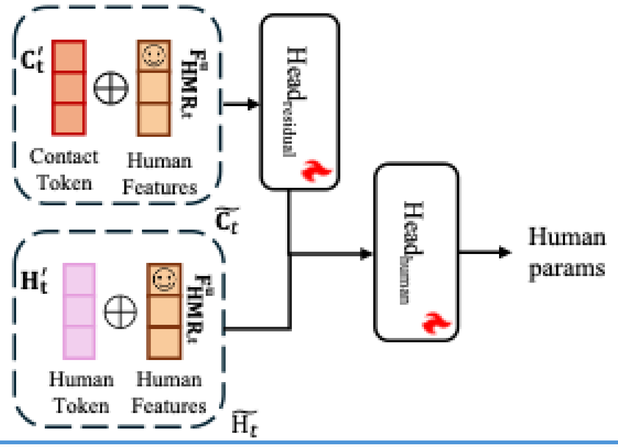
Designing the contact prompt is **not trivial**:

- Contact estimation is ill-posed from one view.
- Image features lack explicit metric geometry.
- Inferred contact can be temporally unstable.

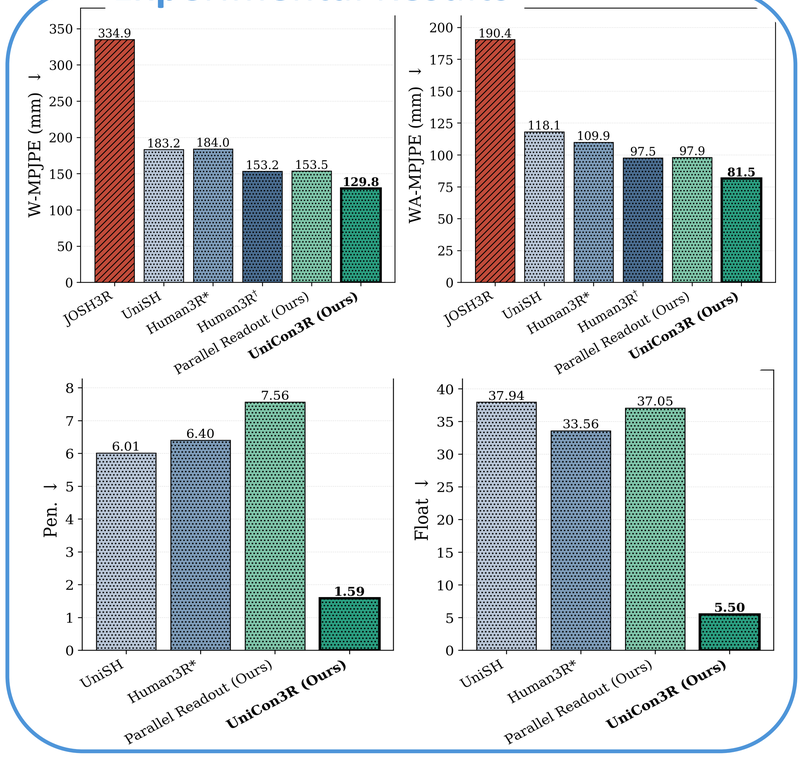


Contact-Guided Latent Refinement

- Contact is not treated as an auxiliary prediction.
- Contact directly refines the human representation.

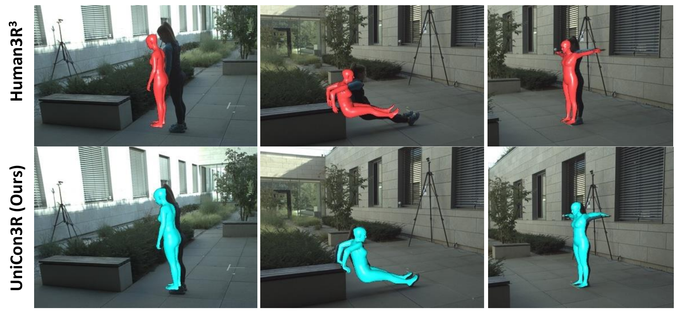


Experimental Results

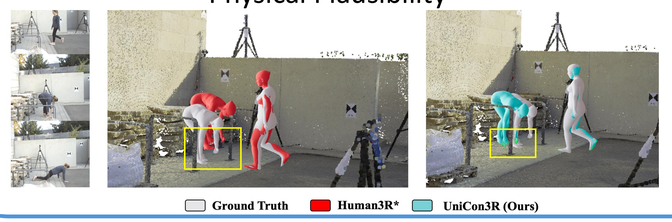


Qualitative Results

Local Human Mesh Reconstruction



Physical Plausibility



References:
1. JOSH3R, Liu et al., ICLR' 26
2. UniSH, Li et al., CVPR' 26
3. Human3R, Chen et al., ICLR' 26

Webpage: https://surtantheta.github.io/UniCon3R

NUS National University of Singapore

MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS