

A APPENDIX

A.1 MORE VISUALIZATIONS OF HUMAN INTERACTIONS

We show visualization results on human interactions. Fig. 5 demonstrates that our pipeline can generate well-aligned image-annotation pairs where people are with close interactions. The generated data pairs are of great value in enhancing existing human interaction datasets collected in the studio environment. (e.g., Hi4D [Yin et al. \(2023\)](#) and CHI3D [Fieraru et al. \(2020\)](#))

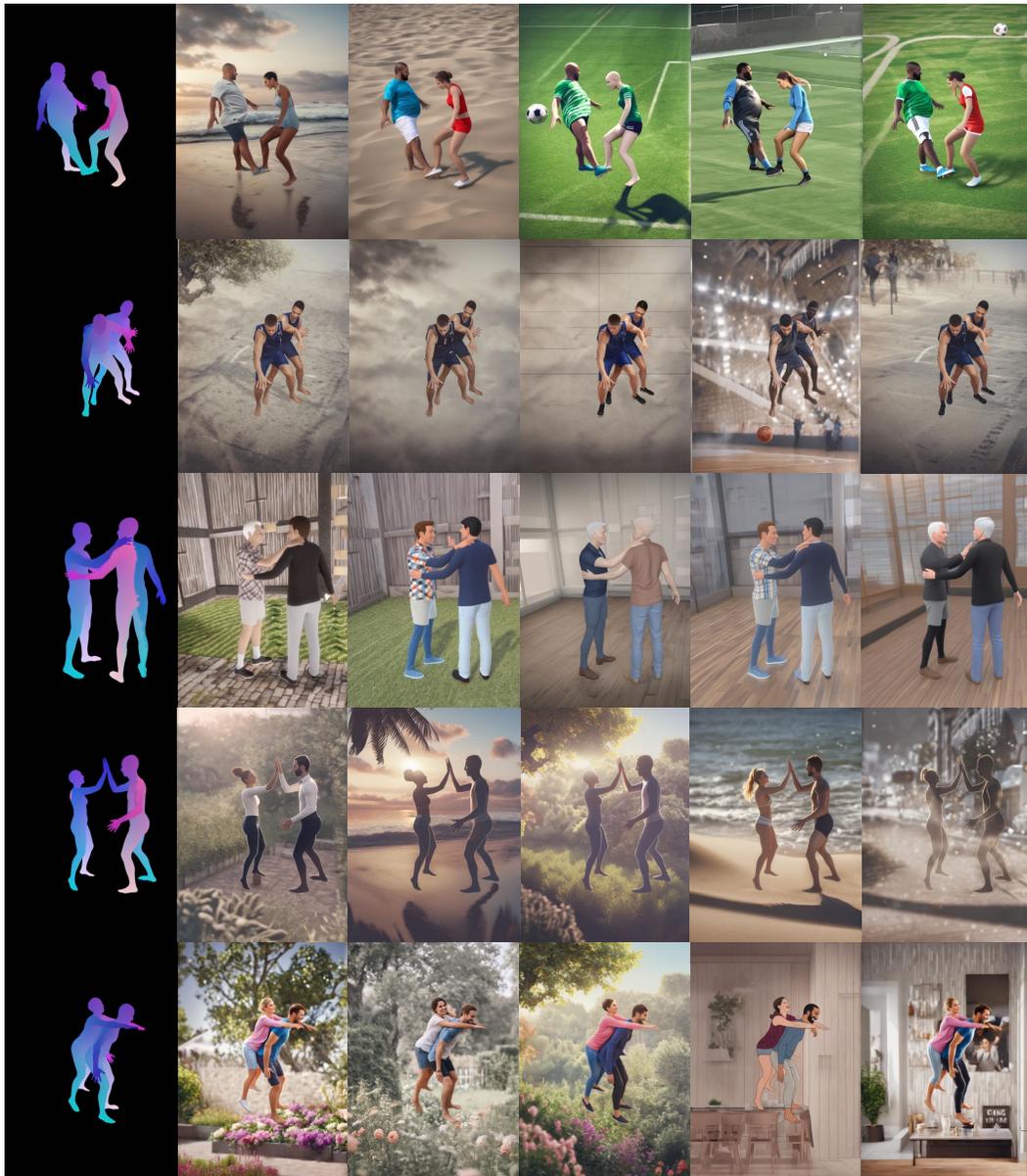


Figure 5: Visualization of Human interaction. The SMPL interaction annotations are sampled from the Hi4D [Yin et al. \(2023\)](#) dataset.

A.2 FAILURE CASES OF THE INITIAL TRAINING PAIRS

In Fig. 6 we show some failure cases of the initial data pairs generated by ControlNet. The inconsistency of the image and 3D mesh would affect the performance of the 3D human pose estimation. Thus, it is necessary to conduct the label refinement proposed in this work.



Figure 6: Failure cases of the initial data pairs generated by ControlNet. Highlight with red circles.

A.3 TEXT PROMPT EXAMPLES GENERATED BY LLM

In Appendix A.3 we show some text prompt examples, which are generated by ChatGPT (OpenAI (2023)) with diverse human actions and scenes.

gender	action	environment
a man	playing soccer	at the park
a woman	reading a book	on the beach
a woman	dancing	at a nightclub
a man	eating dinner	at a restaurant
a woman	walking	in the city
a woman	swimming	in the pool
a man	shopping	at the mall
a woman	running	in the park
a man	studying	at the library
a man	working	at the office
a man	chatting	at a cafe

Table 6: Text prompt examples.