Table A: More baselines on AMASS, we fed imputed tracking signals to their provided pre-trained model.

Methods	x	Imputation	y	MPJPE	MPJVE	MPJRE	Jitter
AvatarPoser [10]	🎧 & 🤚	Interpolation	🎧 & 🤚	40.42	64.07	16.37	27.89
Bodiffusion [3]	🎧 & 🤚	Interpolation	🎧 & 🤚	46.45	75.33	17.99	2793.32
AvatarJLM [34]	🎧 & 🤚	Interpolation	🎧 & 🤚	25.02	68.42	14.14	32.18
AvatarPoser [10]	🎧 & 🤚	MAE	🎧 & 🤚	9.88	62.31	5.98	37.89
Bodiffusion [3]	🎧 & 🤚	MAE	🎧 & 🤚	7.35	31.33	5.47	1254.84
AvatarJLM [34]	🎧 & 🤚	MAE	🎧 & 🤚	7.12	37.60	5.24	16.95
DSPoser (Ours)	🎧 & 🤚	MAE	🎧 & 🤚	**5.51**$^{\pm 0.02}$	**24.19**$^{\pm 0.10}$	**4.09**$^{\pm 0.02}$	**4.27**$^{\pm 0.03}$

Table B: Metrics Related to Computational Complexity

Module	# of Params	MACs	Time
VQ-VAE	17.9 M	3.6 G	3 ms
MAE	51.3 M	23.3 G	4 ms
VQ-Diffusion	74.2 M	1190.2 G	958 ms

Table C: Ablation study on training steps and inference steps. MPJPE is reported.

infer. steps	training steps 25	33	50	100	infer. time (ms)
25	5.83	5.92	5.69	8.72	278
33	-	5.67	5.63	5.58	348
50	-	-	5.61	5.53	522
100	-	-	-	5.51	1013

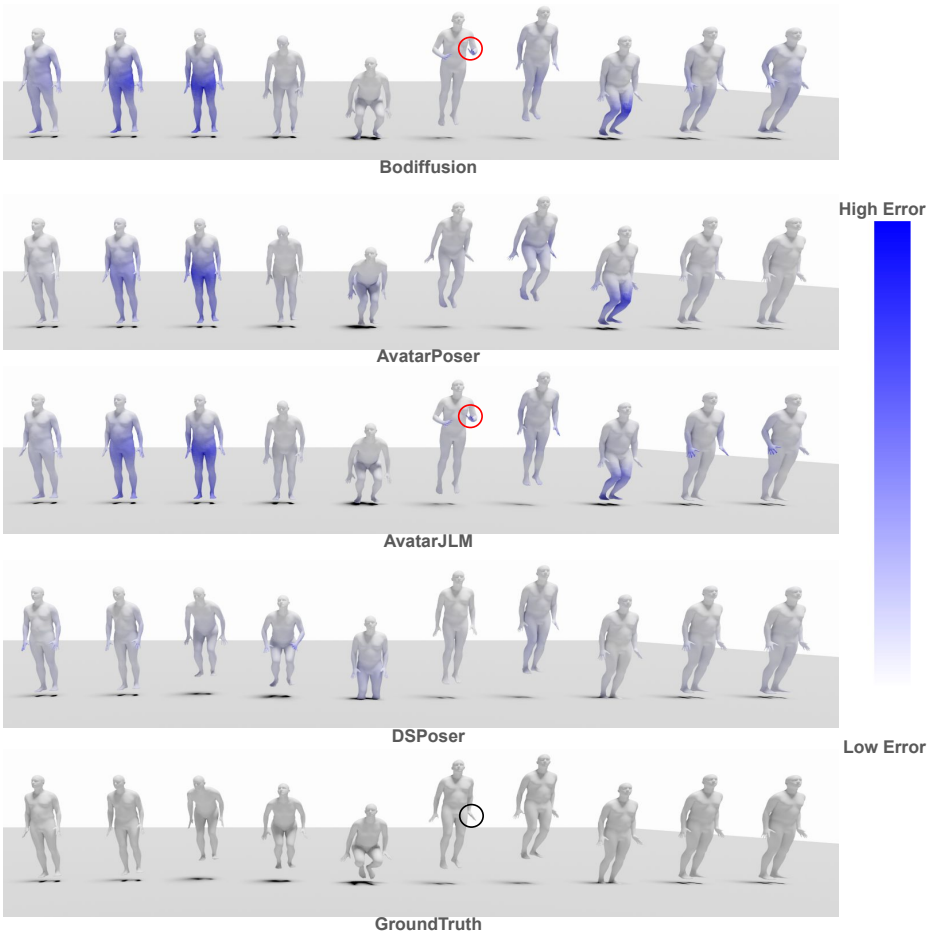

Figure A: Qualitative results comparing DSPoser (Ours) against the baselines. Color gradient indicates an absolute positional error, with a higher error corresponding to higher blue intensity. Results demonstrate that motions generated by DSPoser exhibit greater similarity to the ground truth. Furthermore, it highlights higher errors (indicated with red circles) for baselines when the hand is occluded in the ground truth pose (indicated with a black circle).