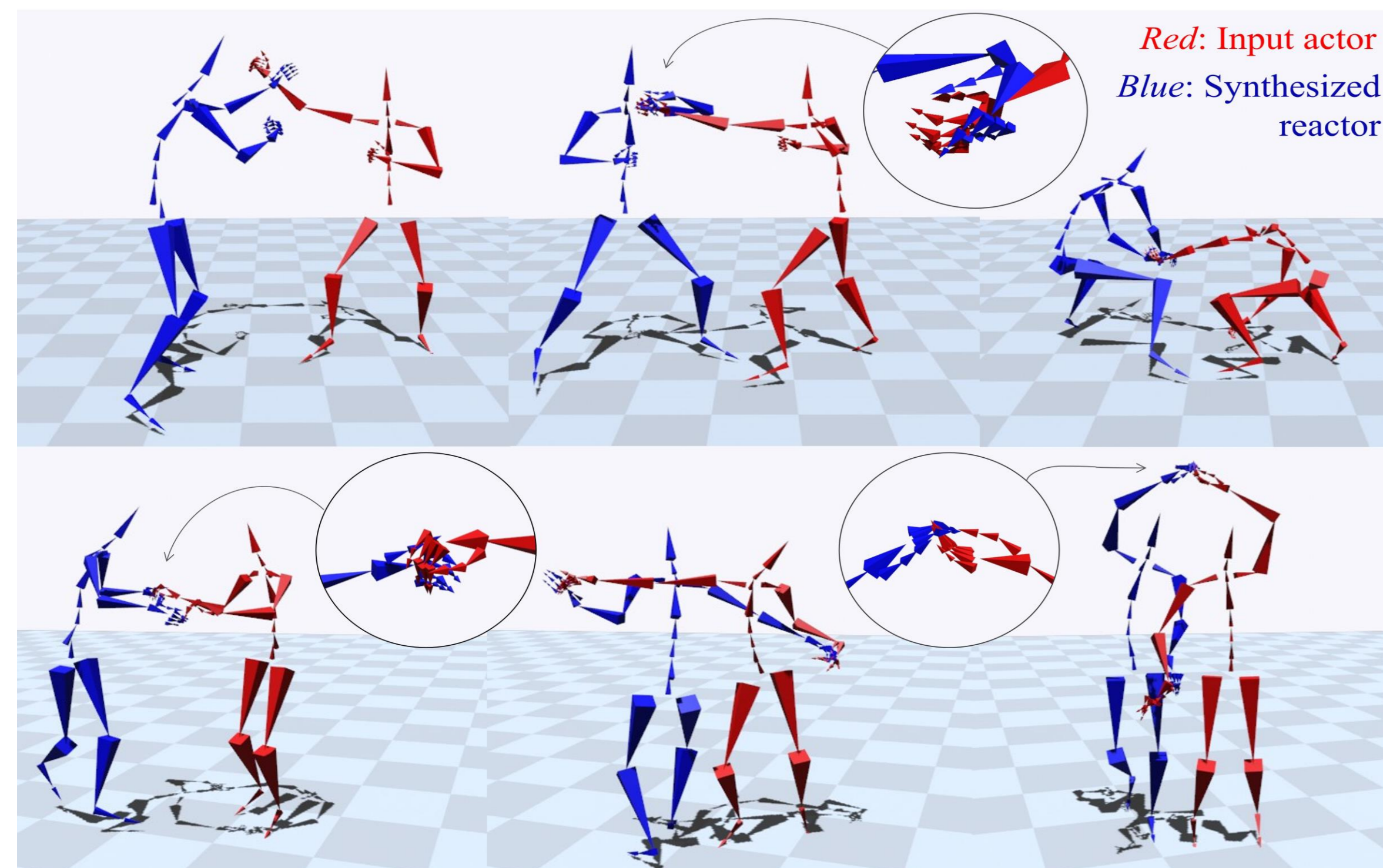




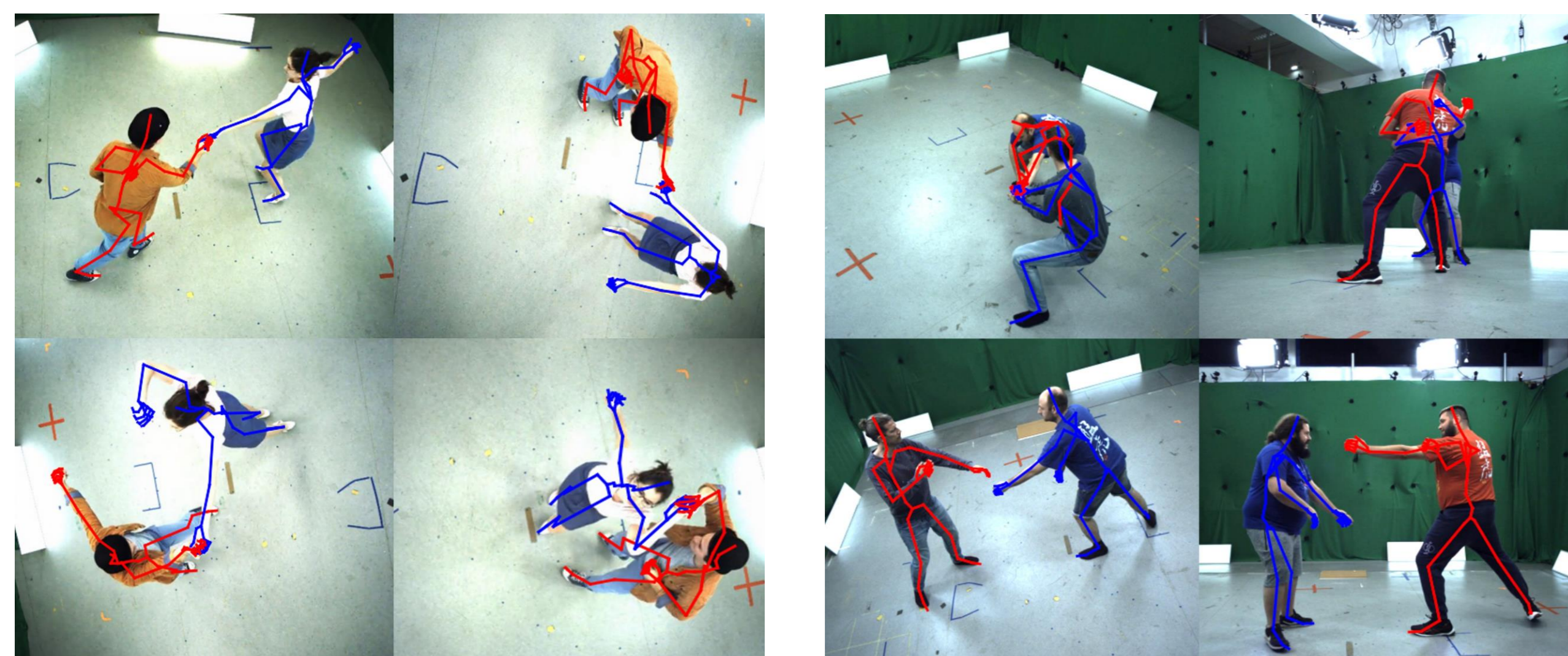
Goal

- Generate 3D full-body and hand motions of a **reactor** given the 3D motion of an **input actor** in a two-person interaction scenario.



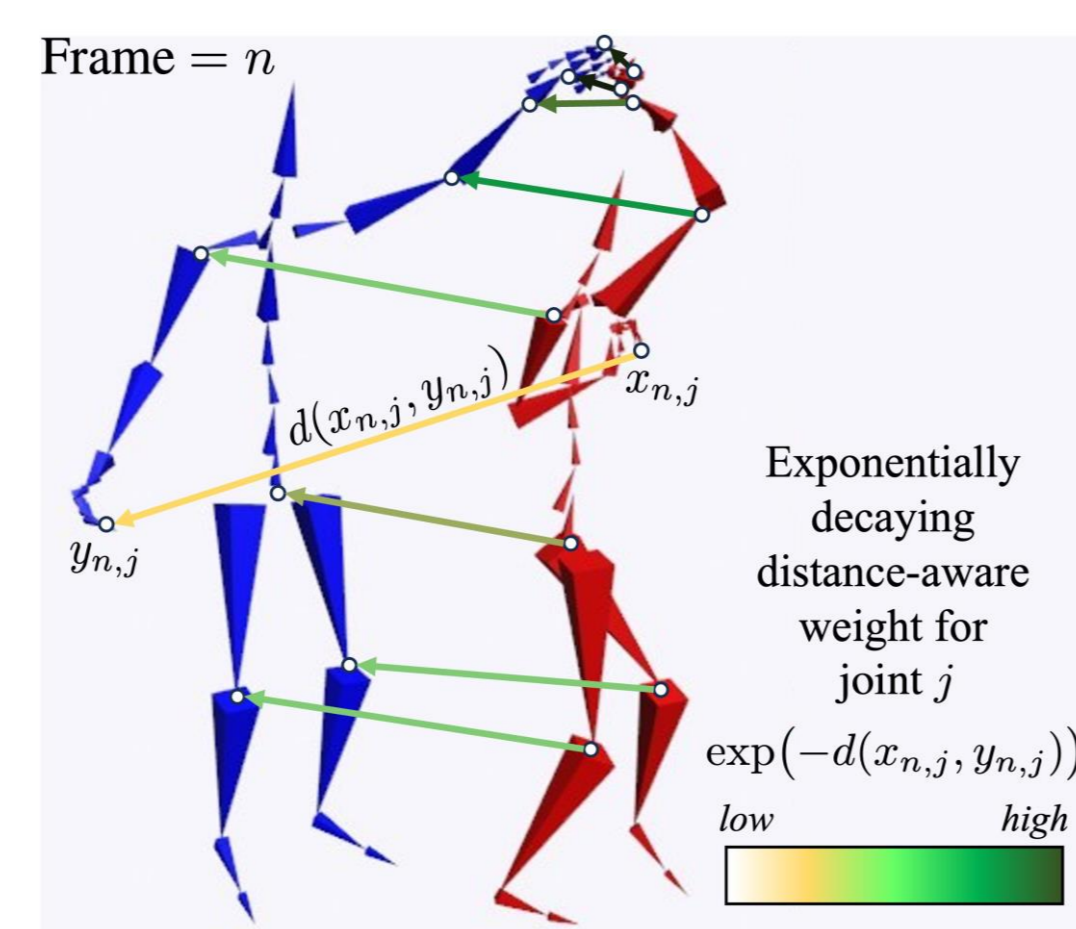
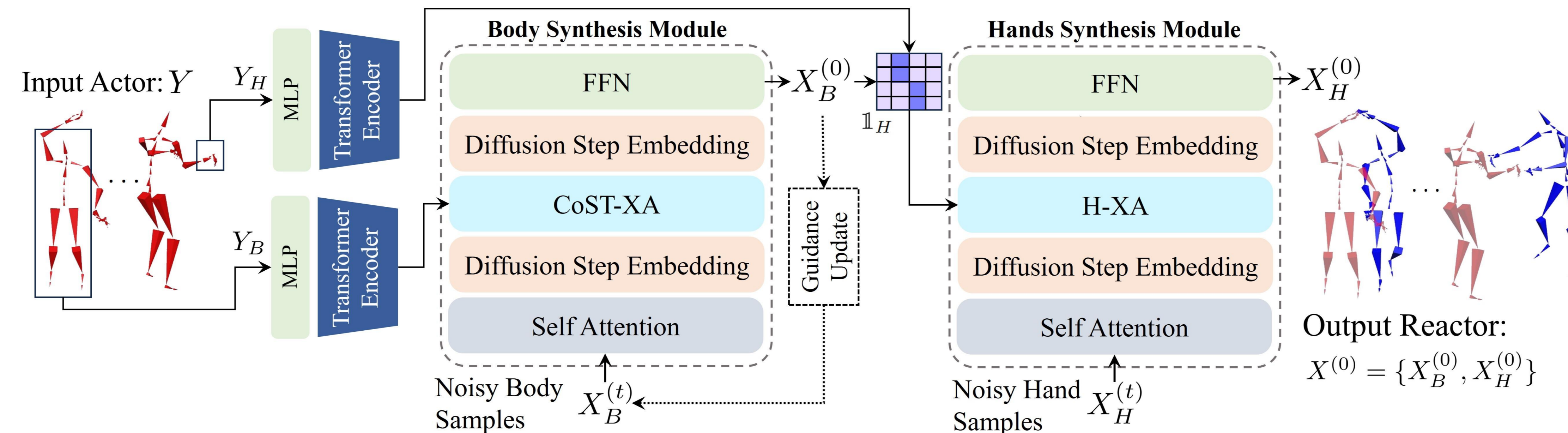
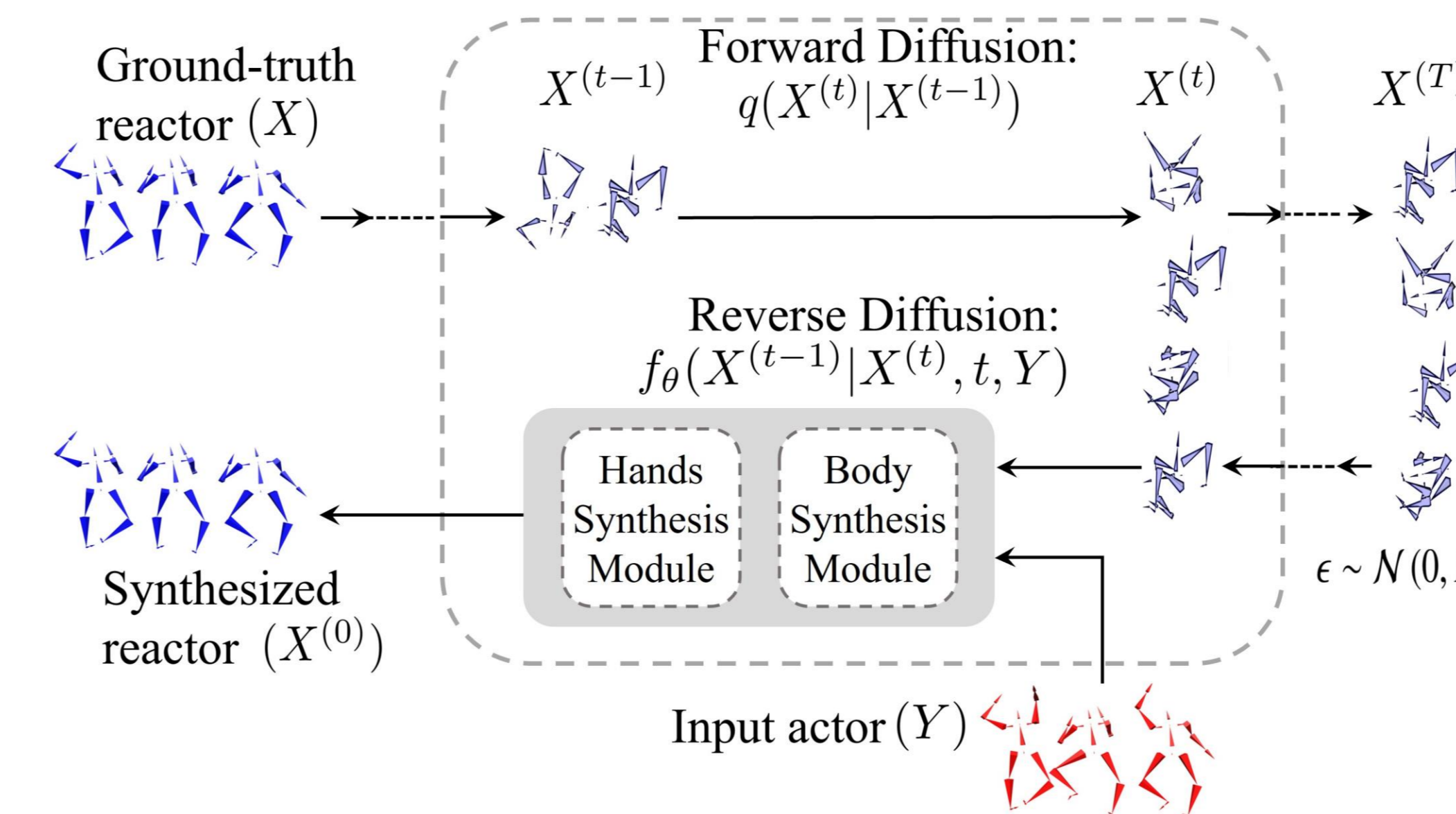
ReMoCap Dataset

- Novel dataset for two-person interactions with full-body and hand motions.
- Consists of two challenging, interactive scenarios: **Lindy Hop dancing** and **Ninjutsu**



DDPM for Reactive Motion Synthesis

- Train a **denoising diffusion-based probabilistic model** conditioned on the 3D motion sequence of the actor.
- Learn the **fine-grained synchronization** between the actor and the reactor through a **two-stage cascaded framework**
- We first synthesize the denoised body samples and then use them for **hand-interaction-aware attention masking** to synthesize the denoised hand samples.



Exponentially decaying reaction loss focus more on the **reactor's** joints that are closer to the **actor's** joints.

Inference time spatial guidance

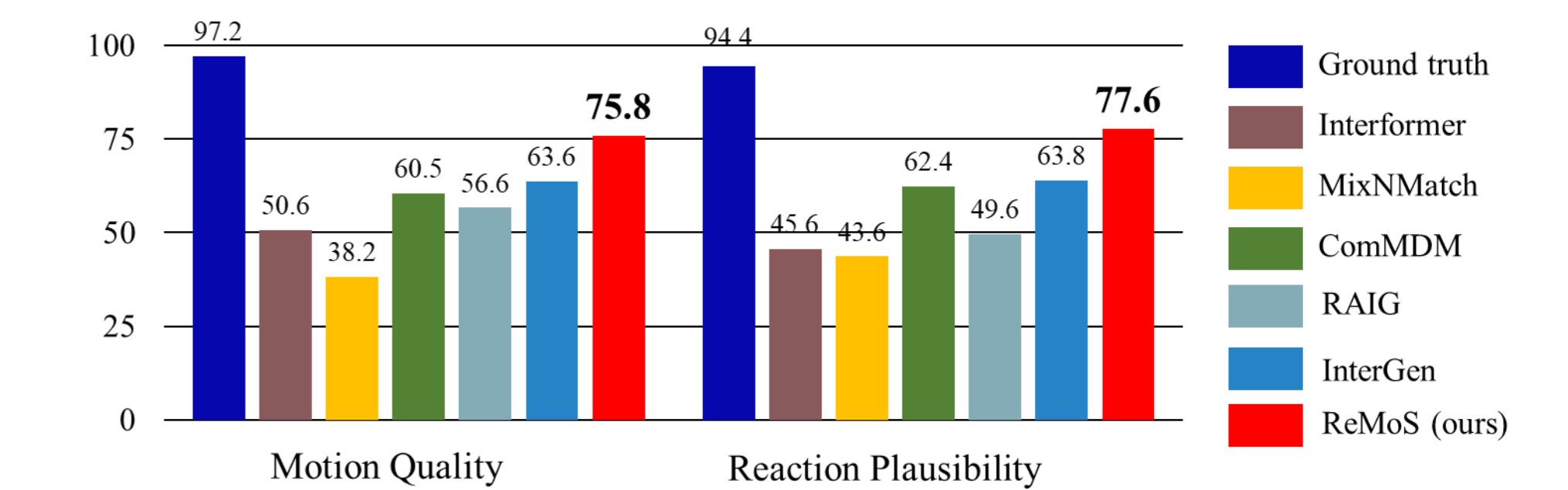
$$G = \arg \min_{\hat{\phi}} \left(\left\| \mathbb{1}_{H_A} \odot \phi - \mathbb{1}_{H_R} \odot \hat{\phi} \right\| \right)$$

$$X_B^{(0)} = X_B^{(0)} - \gamma \nabla_{X_B^{(0)}} G(\phi, \hat{\phi})$$

Results

Method	Lindy Hop					Ninjutsu				
	MPJPE (mm) ↓	MPJVE (mm) ↓	FID ↓ (body)	FID ↓ (hands)	Div →	MPJPE (mm) ↓	MPJVE (mm) ↓	FID ↓ (body)	FID ↓ (hands)	Div →
GT	-	-	-	-	7.57	-	-	-	-	10.51
InterFormer	66.6	8.26	0.53	0.65	4.54	270.2	3.4	0.57	0.68	6.48
MixNMatch	70.2	10.3	0.77	0.78	2.48	257.2	5.2	0.74	0.72	4.83
ComMDM	59.4	4.41	0.32	0.53	7.48	201.2	4.1	0.34	0.58	9.98
RAIG	71.2	4.32	0.47	0.63	8.45	199.1	5.1	0.21	0.63	10.11
InterGen	62.6	3.92	0.30	0.61	7.21	172.6	3.9	0.32	0.57	9.98
ReMoS	40.7	2.26	0.12	0.26	7.62	139.2	3.3	0.16	0.35	10.26

Quantitative evaluation on the ReMoCap dataset



Perceptual study to compare motion quality and reaction plausibility with baseline methods.

