



Paper ID
3827

Dynamic Motion Blending for Versatile Motion Editing

Nan Jiang^{1,2*} Hongjie Li^{1*} Ziyue Yuan^{1*} Zimo He^{1,2,3}
Yixin Chen² Tengyu Liu² Yixin Zhu¹✉ Siyuan Huang²✉

¹ Institute for AI, Peking University ² State Key Laboratory of General Artificial Intelligence, BIGAI

³ Yuanpei College, Peking University * Equal contribution ✉ yixin.zhu@pku.edu.cn, syhuang@bigai.ai

<https://awfuact.github.io/motionrefit/>

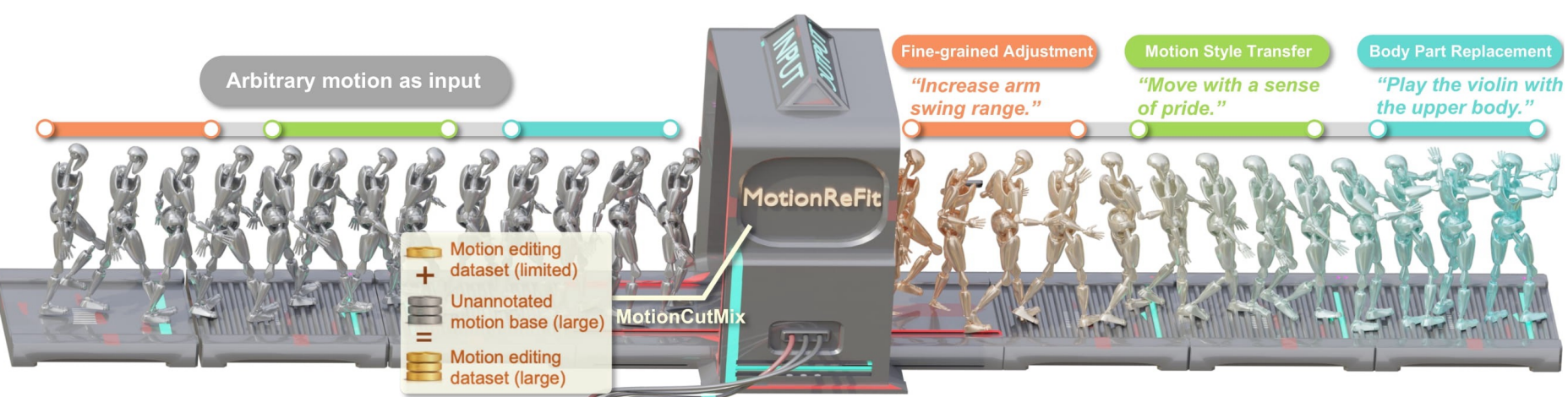


Project Page

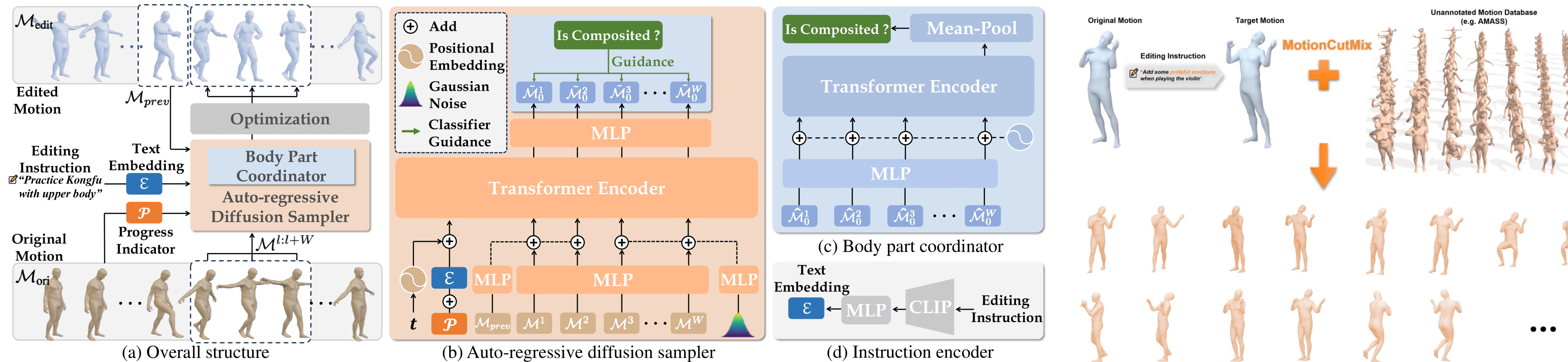


Contributions

- ✓ A **method** for text-guided human motion editing that requires **only** original motion and editing instruction as input.
- ✓ A **training technique** for online motion editing data augmentation.
- ✓ A **dataset** for diverse motion editing tasks.



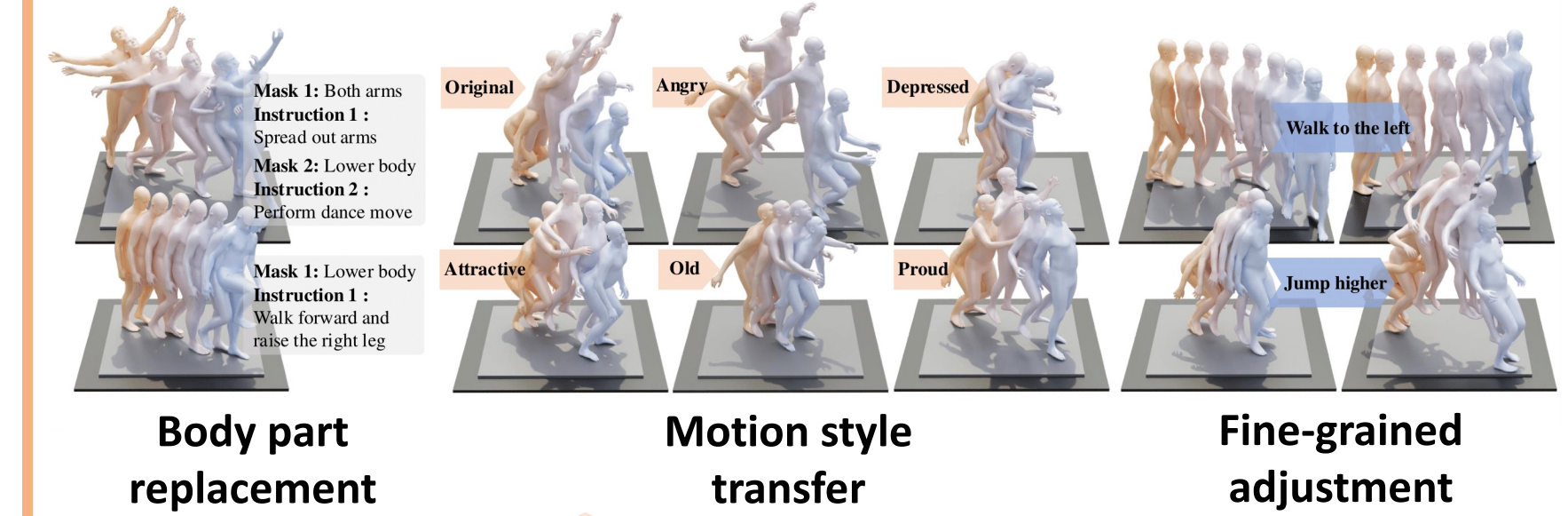
Motion Editing Method



Diffusion-based motion editing model

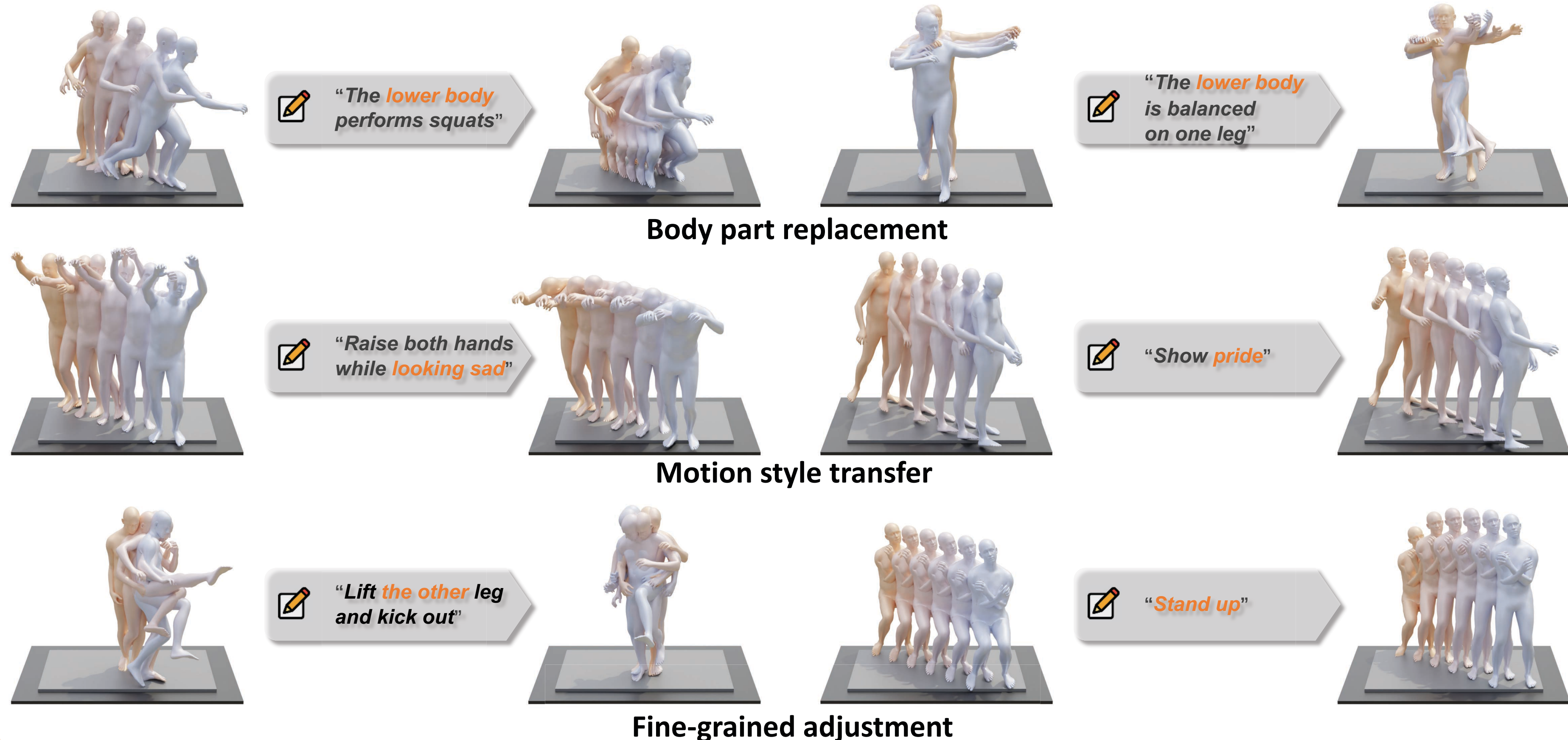
Online data augmentation

Proposed Dataset



- ✓ **STANCE**, an extensive **motion-captured** and **manually annotated** dataset for **diverse motion editing tasks**.
- ✓ Comprises **30,000+** annotated motion editing sequences.

Motion Editing Results



Experimental Results

Comparison with baselines

Table 1. Quantitative comparison across body part replacement (upper) and style transfer (lower) tasks. Each metric reports mean over 10 evaluations with 95% confidence intervals ($\pm$). Arrows ($\rightarrow$) indicate metrics where values closer to real data are better. **Bold** denotes best performance.

Method	FID $\downarrow$	Diversity $\rightarrow$	FS $\downarrow$	Edited-to-Source Retrieval				Edited-to-Target Retrieval			
				R@1 $\rightarrow$	R@2 $\rightarrow$	R@3 $\rightarrow$	AvgR $\rightarrow$	R@1 $\uparrow$	R@2 $\uparrow$	R@3 $\uparrow$	AvgR $\downarrow$
Real Data	0.01	36.06	0.98	52.08	54.32	56.00	8.28	100.0	100.0	100.0	1.00
MDM-BP [56]	0.44	36.71	0.91	69.11	79.75	85.14	2.20	39.05	46.39	50.57	8.92
TMED [7]	0.52	35.37	0.90	38.59	44.10	48.67	9.31	42.70	52.89	58.32	6.47
TMED w/ MCM	0.54	35.67	0.90	41.29	46.13	49.80	9.38	50.62	61.95	68.52	4.48
Ours w/o MCM	0.23	36.34	0.96	93.17	96.30	97.33	1.27	51.18	53.71	55.30	8.51
Ours w/o BC	0.23	36.18	0.97	52.51	56.03	58.19	7.54	60.78	67.17	71.11	4.74
Ours full	0.20	36.01	0.97	52.48	56.13	58.59	7.46	61.37	68.35	72.20	4.65
Real Data	0.01	33.98	0.98	50.94	62.88	67.40	6.28	100.0	100.0	100.0	1.00
MDM-BP [56]	0.39	33.64	0.89	62.40	82.78	89.62	1.96	38.89	53.51	60.24	7.14
TMED [7]	1.54	34.37	0.90	28.44	40.03	46.53	8.48	24.76	38.33	45.62	8.12
TMED w/ MCM	0.84	34.35	0.92	39.83	55.00	62.92	5.37	33.02	47.60	56.94	6.15
Ours w/o MCM	0.23	34.05	0.93	87.05	98.33	99.41	1.16	51.39	63.58	67.88	7.15
Ours w/o BC	0.16	34.51	0.95	45.52	57.05	62.29	6.57	62.26	74.69	79.90	3.51
Ours full	0.14	34.19	0.94	47.67	57.71	62.50	6.46	63.82	76.35	80.69	3.48

Ablation study of online data augmentation

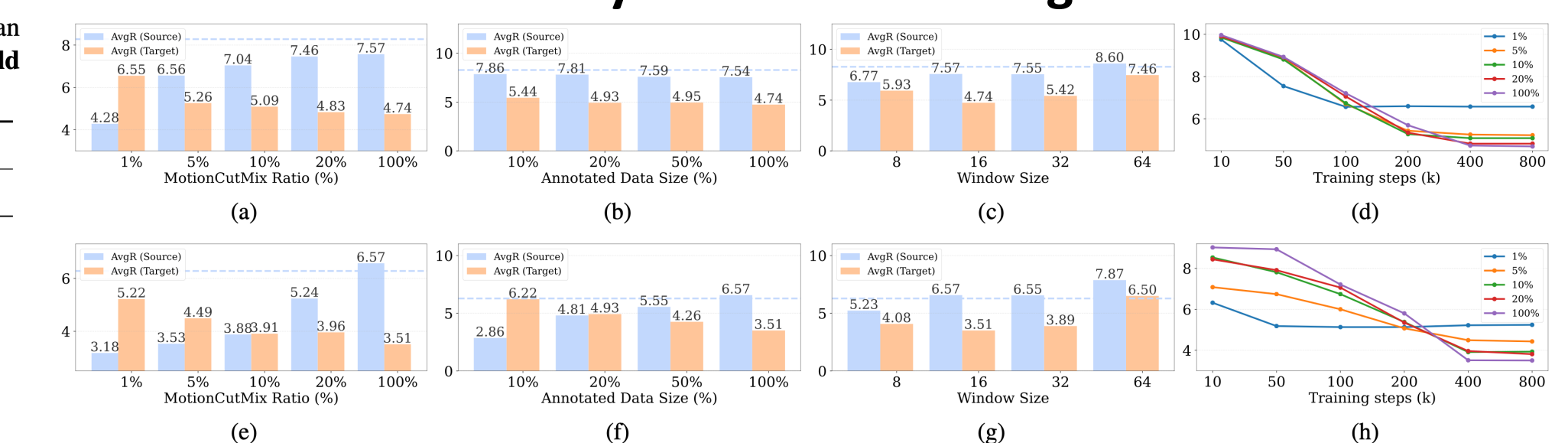


Figure 6. Ablation analyses for body part replacement (a-d) and style transfer (e-h), reporting AvgR metrics. Edited-to-Target AvgR shown only for (d) and (h), with blue dotted lines indicating real data Edited-to-Source AvgR. Parameters studied: (a,e) MotionCutMix ratio, (b,f) annotated data volume, (c,g) temporal window size, and (d,h) convergence patterns at varying MotionCutMix ratios. All training converges within 800k steps.

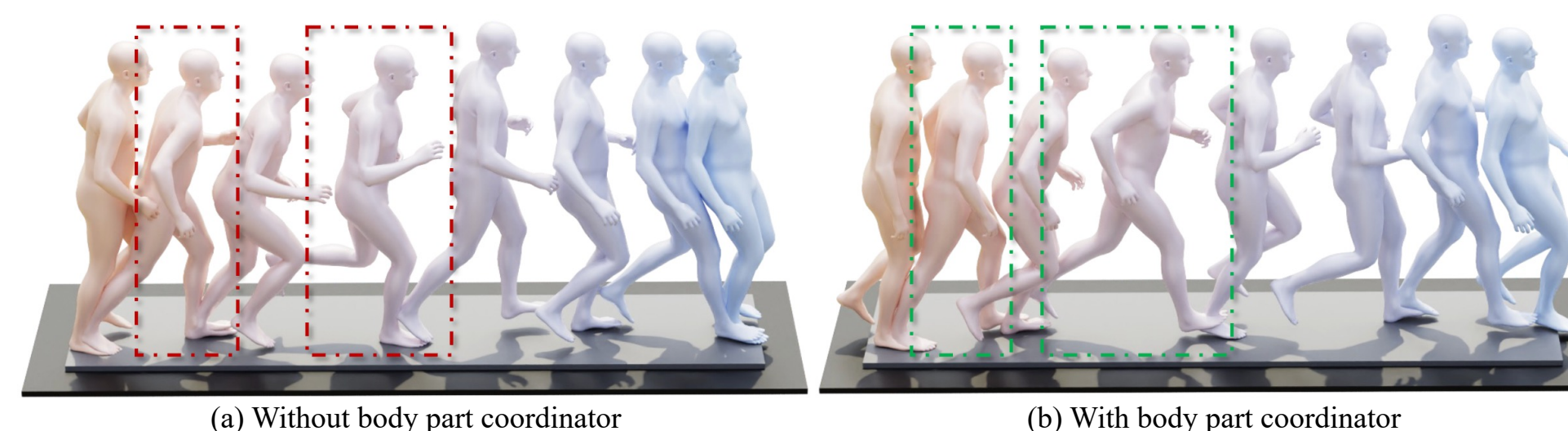


Figure 7. Impact of body part coordinator on motion quality. Examples show paired results using identical random seeds, highlighting how coordinator prevents unnatural synchronous movements of same-side limbs (arm and leg moving forward together).

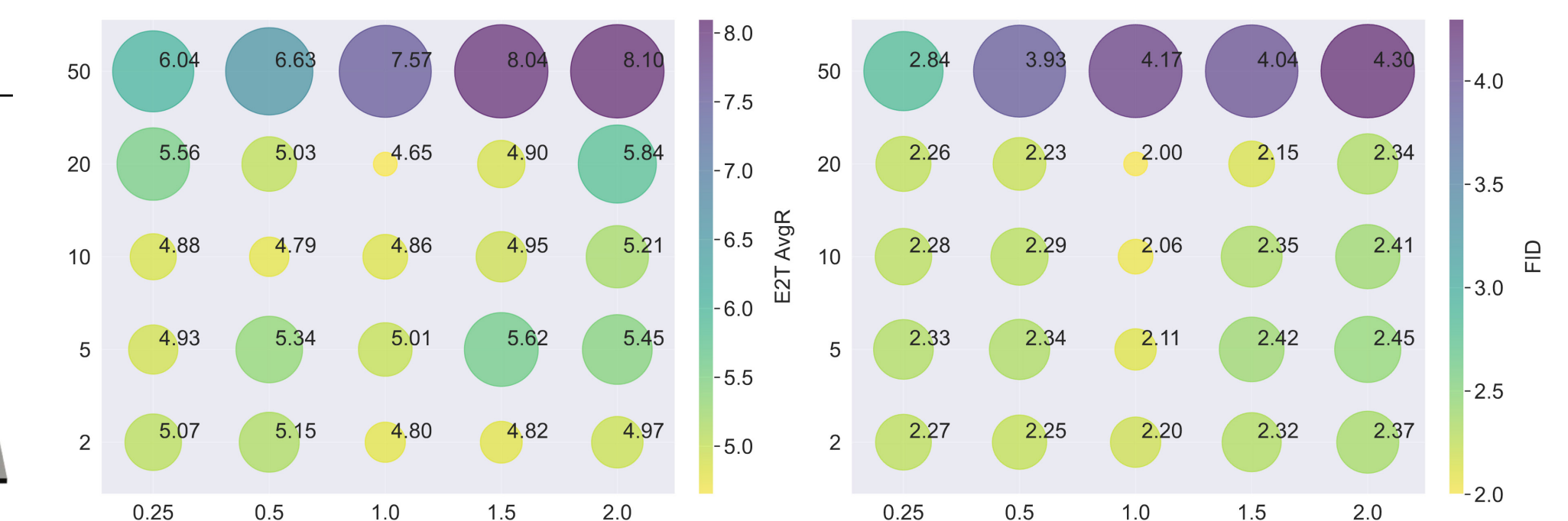


Figure A1. Ablation results on classifier guidance. We illustrate the E2T AvgR (upper) and FID (lower) performance of MotionRefit for the body part replacement task. The x-axis represents guidance strength, whereas the y-axis depicts guidance steps count.

Ablation study of guidance terms