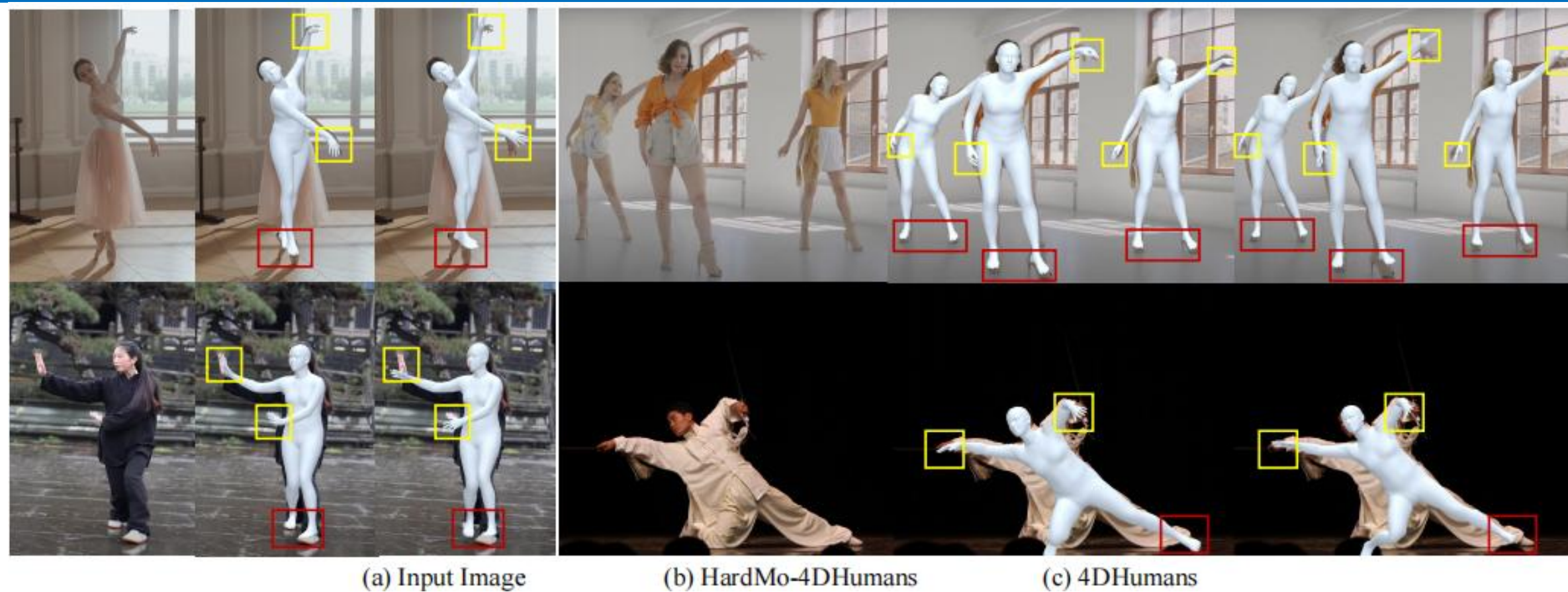
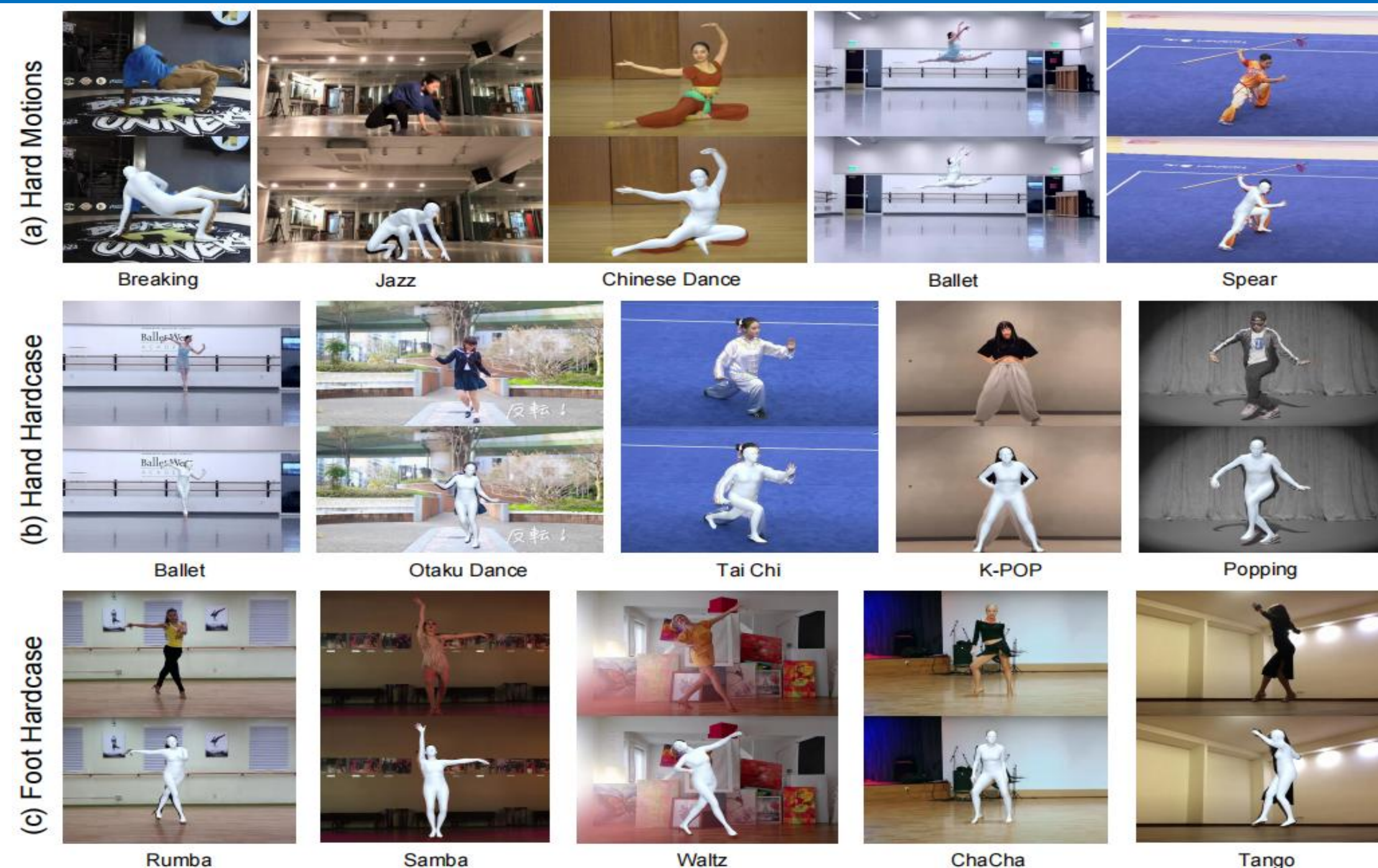


Motivation



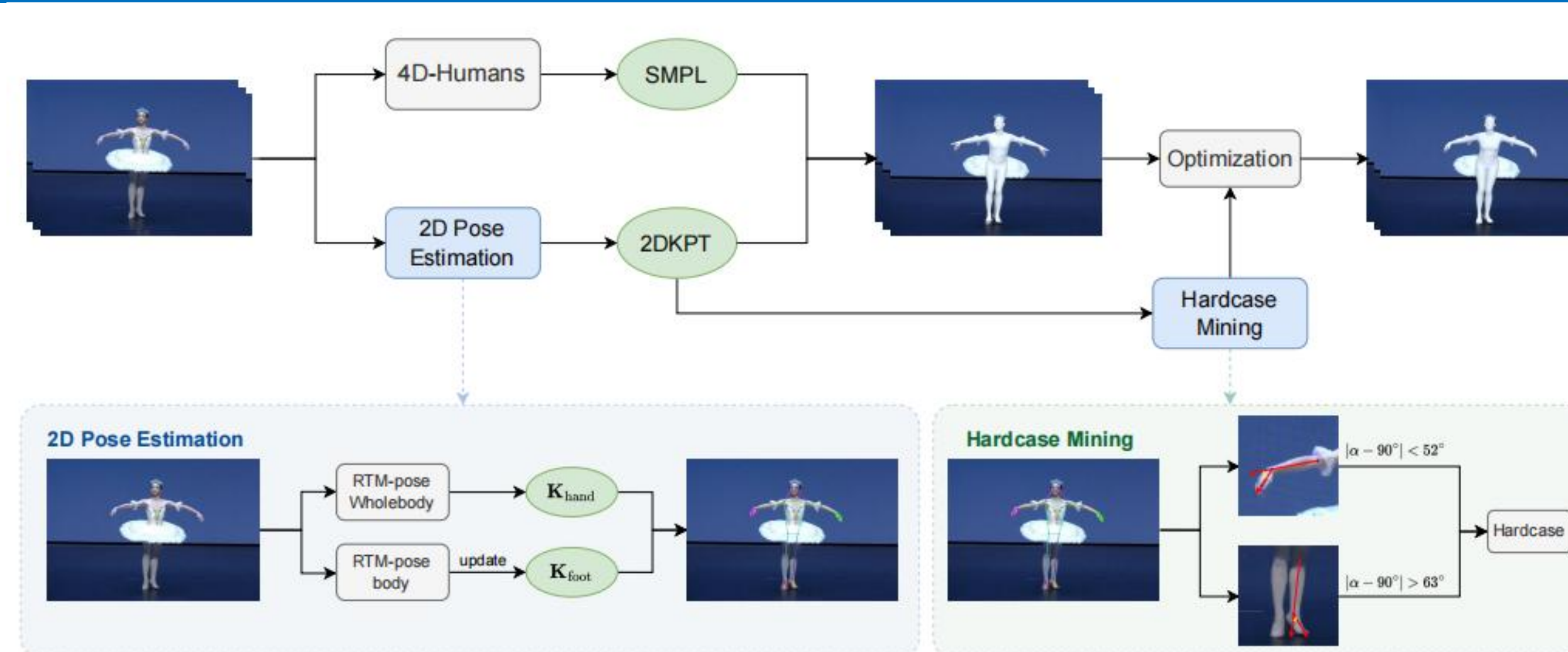
Monocular human mesh recovery has made rapid progress. Despite impressive performance on public benchmarks, existing methods struggle with unusual poses, limiting their deployment in challenging scenarios like dance and martial arts.

Dataset



- We propose a large-scale dataset, named HardMo, contains 7M images along with precise annotations covering 15 categories of dance and 14 categories of martial arts.
- We also construct two subsets named HardMo-Hand and HardMo-Foot to specifically handle the two hardcases.

Method



- First, we label the HardMo dataset with pseudo-labels through the Normal Annotation stage.
- Second, we refine these labels to obtain the HardMo-Foot and HardMo-Hand datasets with accurate pseudo-labels through the Hardcase Annotation stage.

Experiments

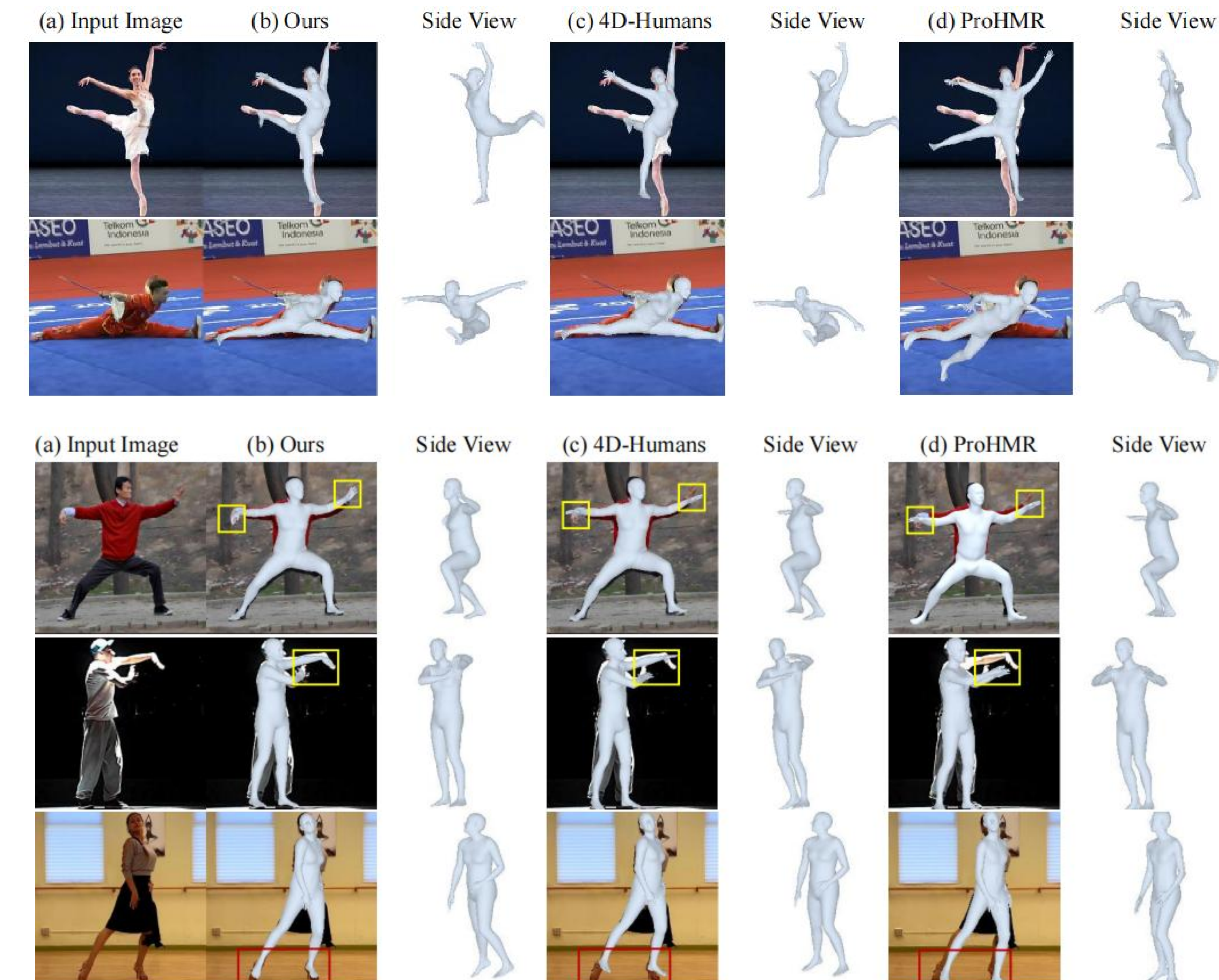
- Benchmark on HardMo, HardMo-Foot and HardMo-Hand

Method	MPJPE↓		PA-MPJPE↓		PCK@0.01↑		PCK@0.05↑	
	Body	Hand	Body	Hand	Body	Hand	Body	Hand
ProHMR [22]	113.8	74.2	0.10	0.73	0.12	0.03	0.75	0.47
CLIFF [24]	93.3	56.2	0.29	0.90	0.32	0.08	0.93	0.64
HMR [18]†	61.1	42.1	0.40	0.96	0.54	0.36	0.98	0.98
HardMo-HMR†	46.0	31.5	0.47	0.97	0.61	0.46	0.99	0.99
HardMo-HMR	36.0	25.0	0.56	0.98	0.58	0.54	0.99	0.99

Method	MPJPE↓		PA-MPJPE↓		PCK@0.01↑		PCK@0.05↑	
	Body	Foot	Body	Foot	Body	Foot	Body	Foot
4DHumans ^a [10]	83.1	52.7	0.16	0.92	0.15	0.04	0.93	0.49
4DHumans ^b [10]	36.6	23.0	0.48	0.98	0.56	0.11	0.99	0.63
HardMo-4DHumans†	29.9	20.2	0.55	0.98	0.61	0.46	0.99	0.99
HardMo-4DHumans	26.0	18.0	0.60	0.98	0.58	0.54	0.99	0.99

Method	HardMo-foot P1								HardMo-foot P2							
	MPJPE↓		PA-MPJPE↓		PCK@0.01↑		PCK@0.05↑		MPJPE↓		PA-MPJPE↓		PCK@0.01↑		PCK@0.05↑	
	Body	foot	Body	foot	Body	foot	Body	foot	Body	foot	Body	foot	Body	foot	Body	foot
ProHMR [22]	117.0	213.6	73.2	53.1	0.13	0.02	0.77	0.50	90.0	131.9	54.4	41.5	0.14	0.03	0.82	0.64
CLIFF [24]	93.5	169.4	49.5	42.0	0.31	0.14	0.91	0.82	73.1	120.4	39.5	32.0	0.42	0.22	0.97	0.92
HardMo-HMR (w/o OPT)	38.9	58.3	23.4	16.2	0.57	0.43	0.99	0.98	42.1	58.1	25.6	21.1	0.57	0.51	0.98	0.98
HardMo-HMR (w/ OPT)	23.6	34.9	15.4	10.8	0.71	0.58	0.99	0.99	27.5	39.5	17.7	13.2	0.67	0.60	0.99	0.99
4DHumans ^a [10]	86.7	143.7	47.2	41.8	0.16	0.08	0.95	0.79	77.1	100.0	40.1	33.3	0.22	0.14	0.98	0.92
4DHumans ^b [10]	40.0	88.1	21.1	26.5	0.59	0.14	0.99	0.86	34.9	77.0	19.0	25.4	0.64	0.16	0.99	0.92
HardMo-4DHumans† (w/ OPT)	20.9	29.4	13.5	9.0	0.68	0.50	0.99	0.99	24.4	34.5	15.5	11.0	0.68	0.59	0.99	0.99
HardMo-4DHumans (w/ OPT)	19.8	29.3	13.0	9.0	0.70	0.53	0.99	0.99	23.3	34.8	14.6	10.8	0.71	0.60	0.99	0.99

Quantitative Visible Results



HardMo-4DHumans can resolve these hardcases perfectly.

Conclusion

- We develop an efficient and scalable pipeline for automatic annotation and hardcase mining. This system offers a potent solution to the data scarcity issue in the motion domain.
- HardMo bridges the domain gap, containing 7 million images across over 300 different scenarios. As subsets of HardMo, HardMo-Hand and HardMo-Foot, the first of their kind, focus on solving the inherent hardcase issues.