



Action-guided 3D Human Motion Prediction

Jiangxin Sun

Sun Yat-sen University

`sunjx5@mail2.sysu.edu.cn`

Zhang Lin

Sun Yat-sen University

`linzh59@mail2.sysu.edu.cn`

Xintong Han

Huya Inc

`hanxintong@huya.com`

Jian-Fang Hu*

Sun Yat-sen University

`hujf5@mail.sysu.edu.cn`

Jia Xu

Huya Inc

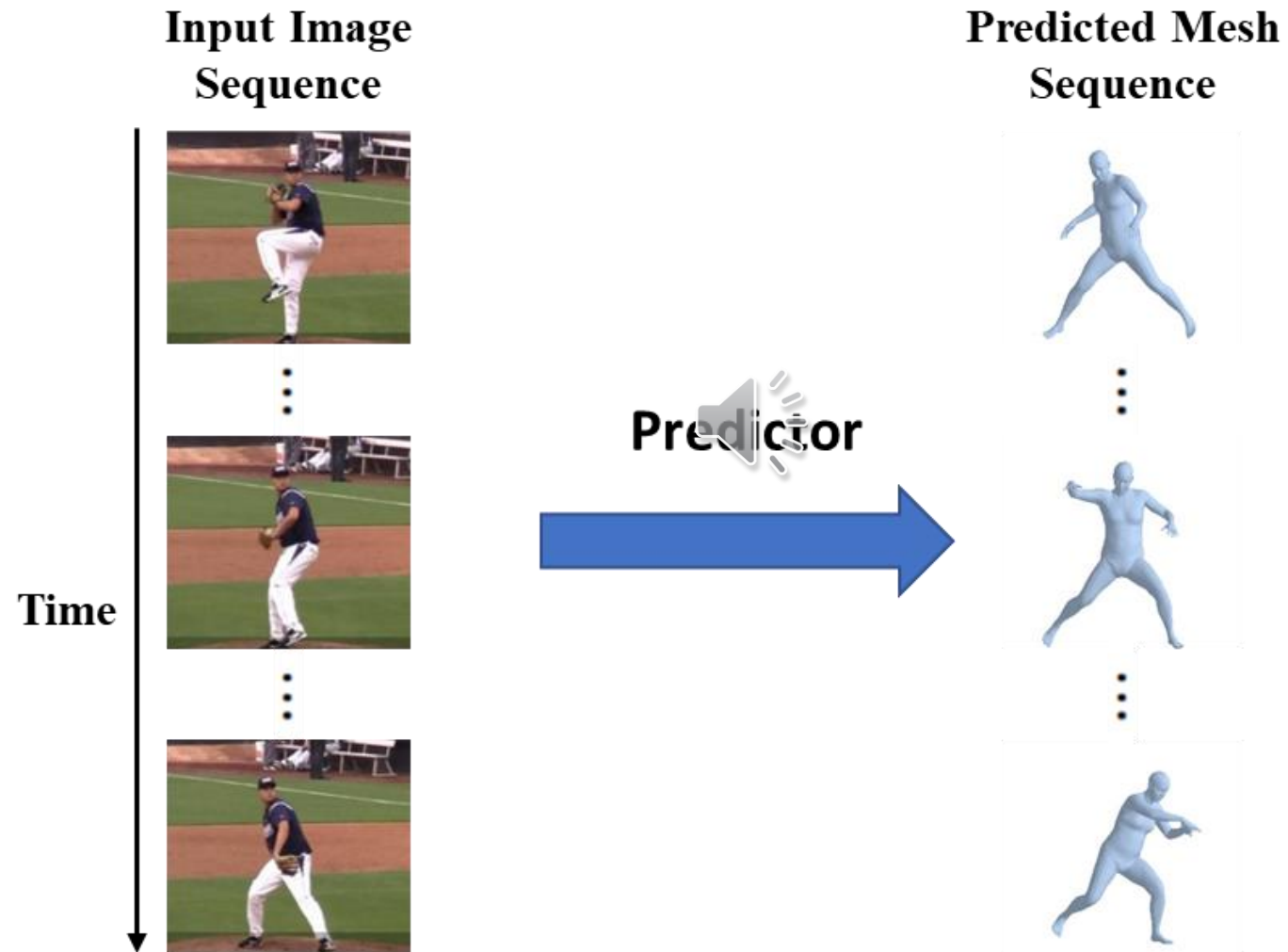
`xujia@huya.com`

Wei-Shi Zheng

Sun Yat-sen University

`wszheng@ieee.org`

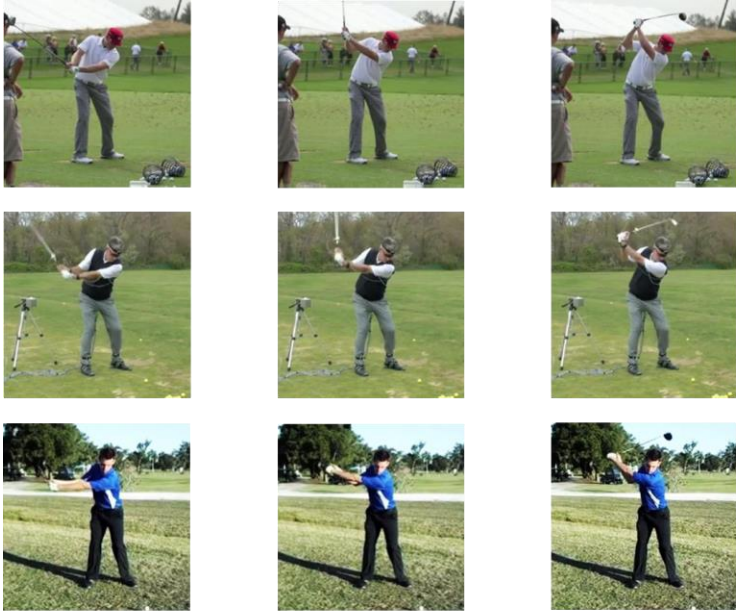
Background



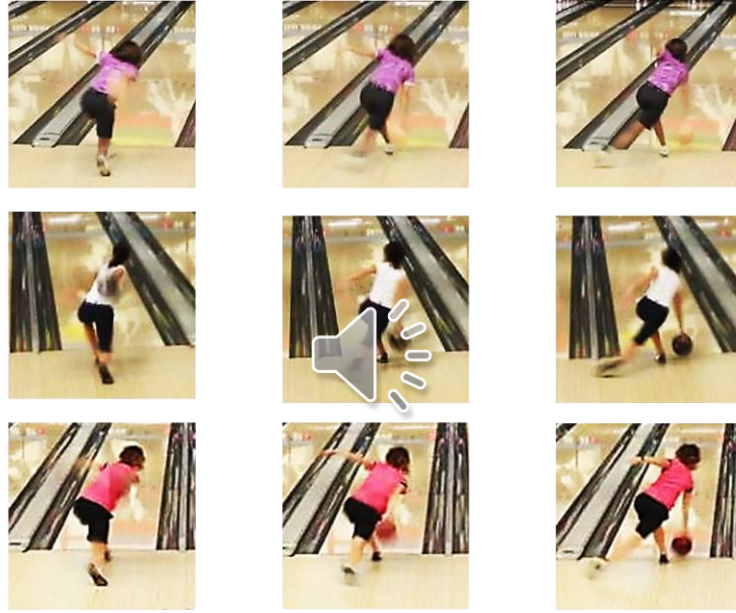
Predicted Mesh Sequence



Motivation



Golf Swing



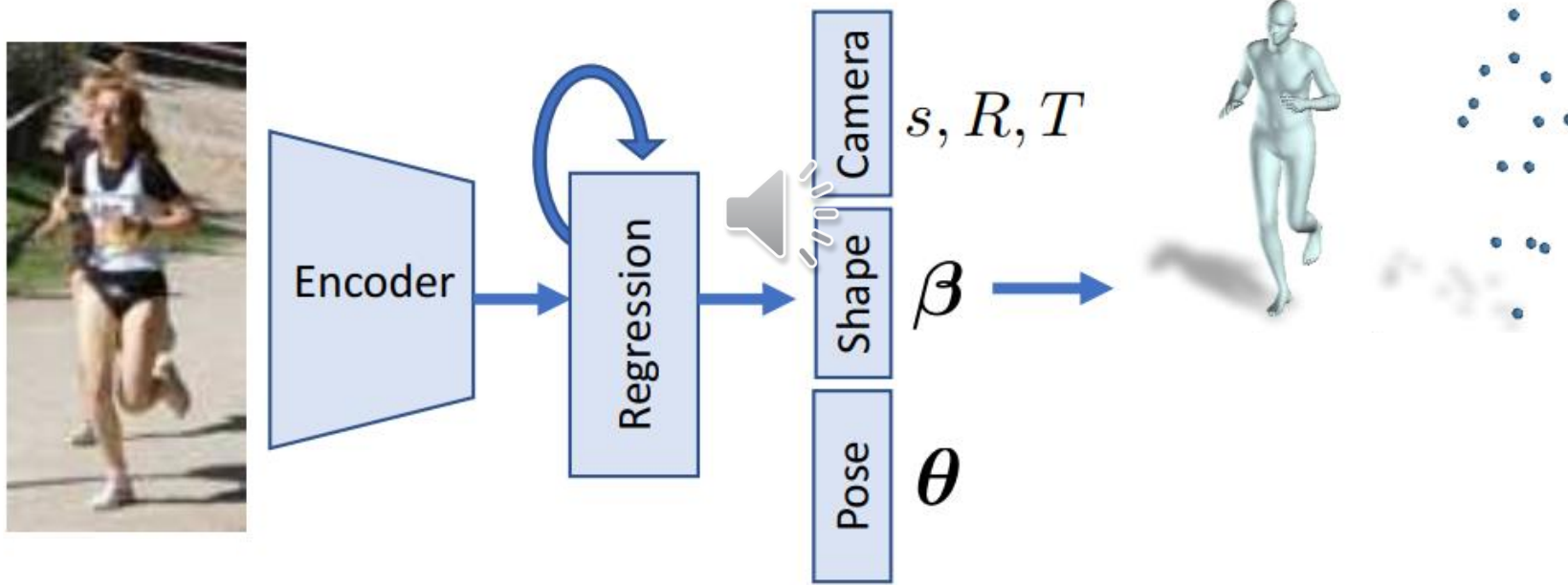
Bowling



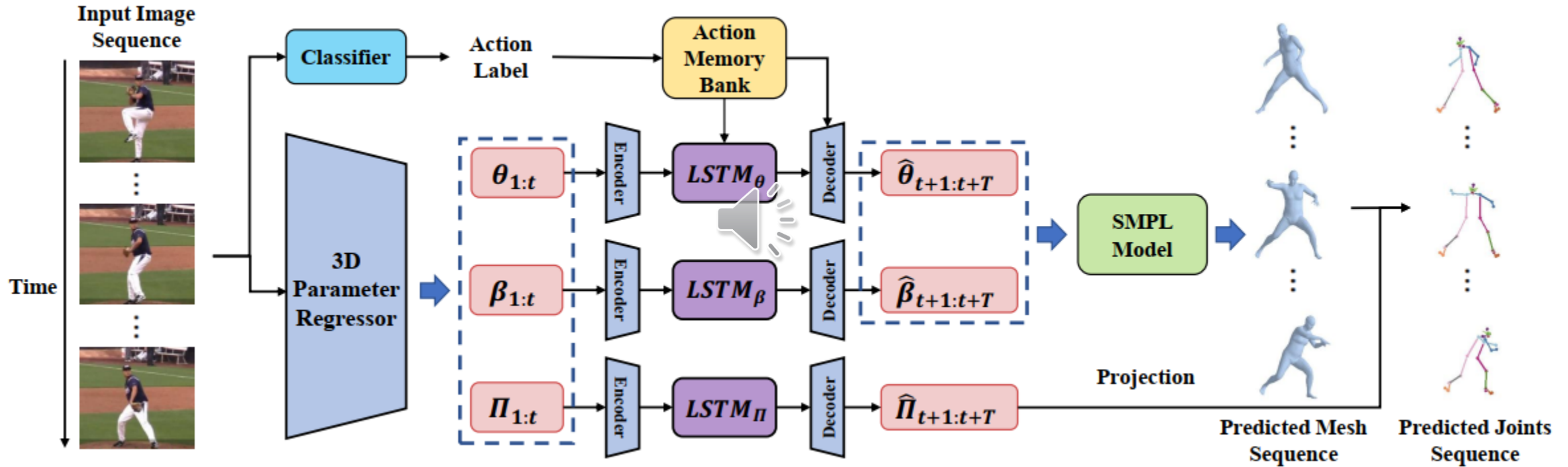
Baseball Pitch

representative motion dynamics

3D Human Body



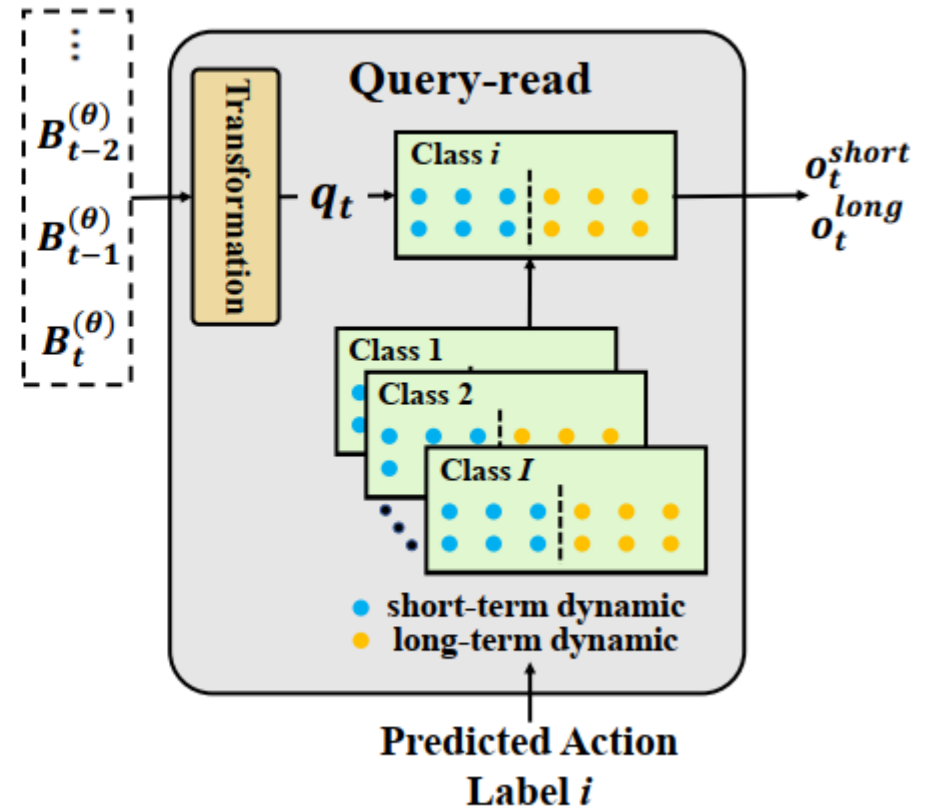
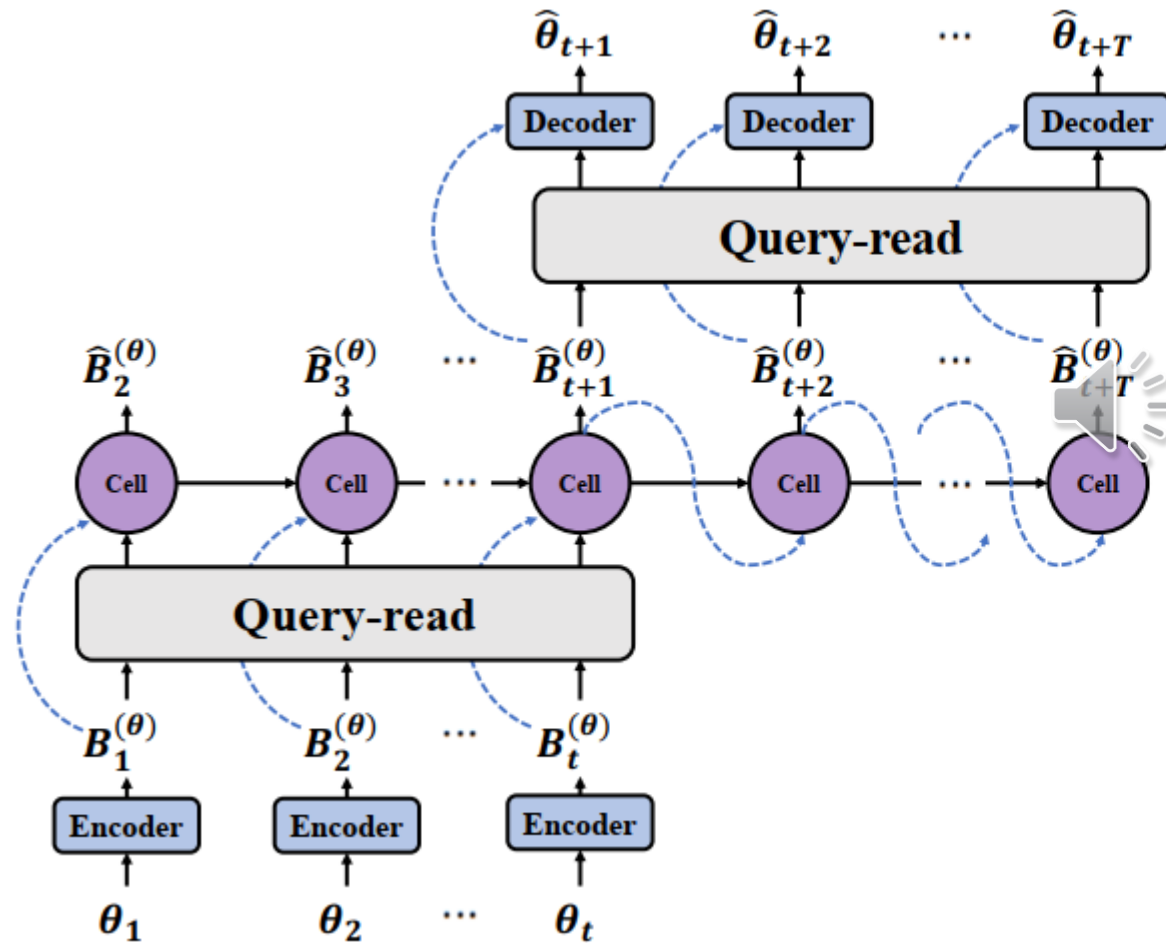
Action-guided Feature Prediction



Action Memory Bank



Action-guided Predictor





$$\mathcal{L} = \mathcal{L}_{pred} + \mathcal{L}_{prior} + \mathcal{L}_{action}$$

Experiments



Dataset & Metric

- Penn Action
in-the-wild sports videos with 2D annotations
- Human 3.6M
indoor human activity videos with 3D annotations
- Evaluation
for 2D annotation: PCK (percentage of correct keypoints)
for 3D annotation: PA-MPJPE
(mean error of per joint position after applying Procrustes Alignment)
- DTW: Dynamic Time Warping (unreasonable in this task)



Penn Action



Human 3.6M

Experiments



Comparison Results

Table 1: Comparison results with recent approaches for human motion prediction.

	Method	Human 3.6M (PA-MPJPE ↓)					Penn Action (PCK ↑)				
		1	5	10	20	30	1	5	10	20	30
	Oracle	56.9	56.9	56.9	56.9	56.9	81.6	81.6	81.6	81.6	81.6
With DTW	Nearest Neighbor	90.3	95.1	100.6	108.6	114.2	63.2	61.5	60.6	58.7	57.8
	Constant	59.7	65.3	72.8	84.3	90.4	78.3	71.7	64.9	56.2	49.7
	PHD [42]	57.7	59.5	61.1	62.1	65.1	81.2	80.0	79.0	78.2	77.2
	Ours	57.3	58.9	60.3	61.4	62.8	81.4	80.6	79.7	79.1	78.5
Without DTW	Nearest Neighbor	90.3	99.8	110.3	124.7	133.3	62.5	57.6	53.7	44.6	41.1
	Constant	59.7	71.4	85.9	101.4	102.8	78.3	65.5	54.6	42.3	32.7
	PHD [42]	57.7	61.2	64.4	67.1	81.1	81.2	77.2	72.4	67.9	60.1
	Ours	57.3	59.6	61.7	62.5	75.9	81.4	79.1	76.7	72.8	66.5

Experiments



Ablation Study

Table 2: Evaluation of our action context modeling on Human3.6M dataset.

Method	Reconstruction error ↓			
	5	10	20	30
Baseline	63.1	66.5	68.9	83.3
+ Prediction with Bank	61.2	63.8	65.7	79.6
+ Decoding with Bank	59.6	61.9	63.4	77.1
+ Action Constraint	59.6	61.7	62.5	75.9



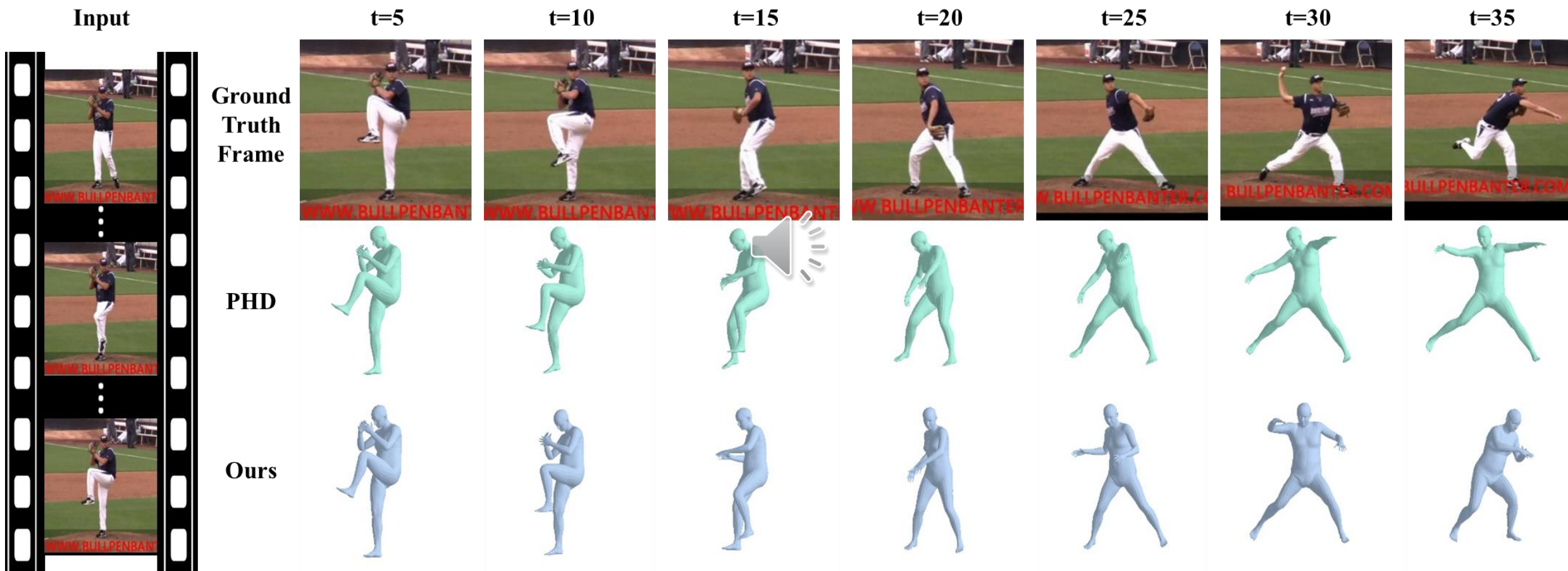
Table 3: Evaluation of our action-specific memory bank on Human3.6M dataset.

	Reconstruction error ↓			
	5	10	20	30
Baseline	63.1	66.5	68.9	83.3
Action-agnostic Bank	61.4	64.6	66.3	80.7
Action-specific Bank	59.6	61.9	63.4	77.1

Experiments



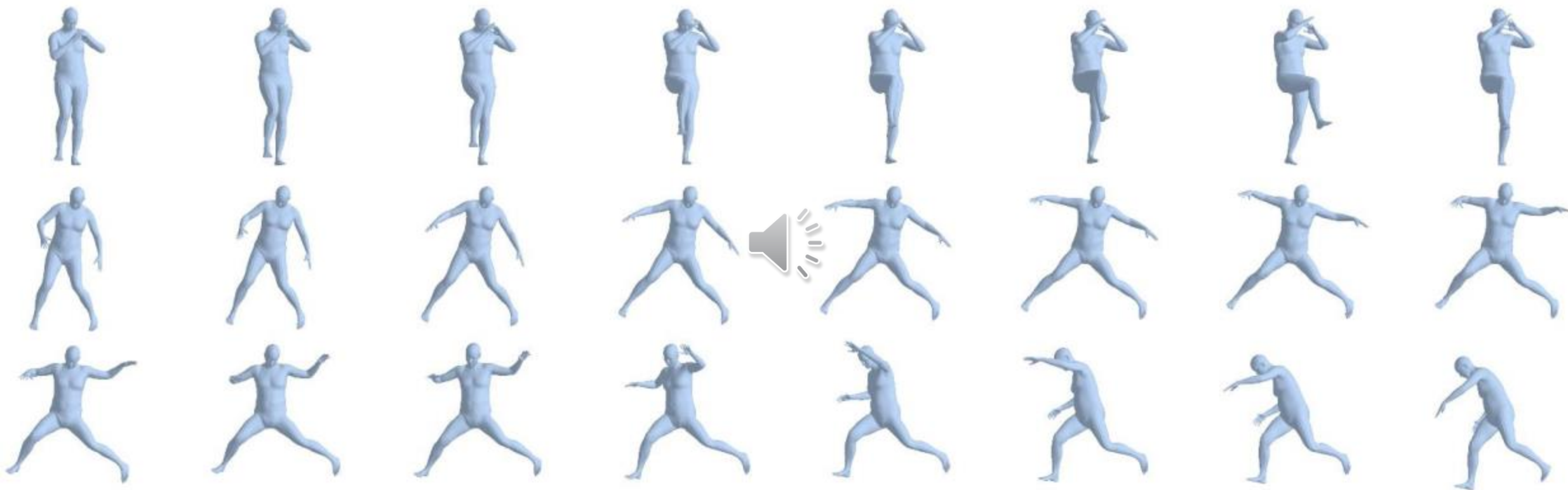
Visualization



Experiments



Visualization



Each row visualizes an item in our action memory bank for the baseball pitch class.



Thanks !