Imitation Learning for Human Pose Prediction

Borui Wang, Ehsan Adeli, Hsu-kuang Chiu, De-An Huang, Juan Carlos Niebles
Stanford University

Introduction

The Human Pose Prediction Problem: predict a sequence of future human poses given a sequence of human poses observed in the past.

Challenges: previous methods for this prediction problem were mainly based on purely supervised training of various RNN architectures, but they face two severe challenges:

- **Generalize poorly** over unseen domains of the human motion space.
- Hard to keep a **balance between long-term and short-term** prediction accuracies.

Our solution: we propose a new **reinforcement learning** formulation for the problem of human pose prediction, and use a combination of two **imitation learning** algorithms to train our pose prediction agent under this RL formulation: one is based on **behavioral cloning** and the other one is based on **generative adversarial imitation learning**.

RL Formulation of Human Pose Prediction

Overview: we define a **Markov Decision Process** to model the generation of pose prediction sequences. We divide the whole prediction sequence evenly into *K* steps, where each step contains a window of *m* pose vectors to predict.

Definition: at each step *i* during the prediction:

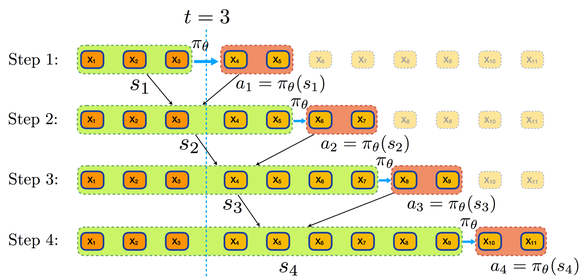
- **State** - the state of the MDP is defined as the list of all previous pose vectors:

$$s_i = \{x_1, x_2, \ldots, x_{t+(i-1)\times m}\}$$

- **Action** - the action of the MDP is defined as the next length-*m* window of pose vectors:

$$a_i = \{x_{t+(i-1)\times m+1}, \ldots, x_{t+i\times m}\}$$

- **Transition Dynamic** - deterministic: $s_{i+1} = \{s_i, a_i\}$



Imitation Learning for Human Pose Prediction

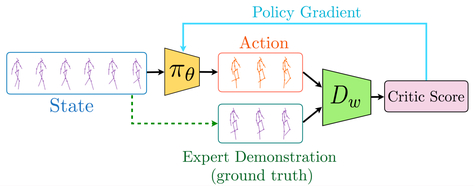
We use a combination of two different imitation learning approaches to train our pose prediction agent under this RL formulation:

- **Behavioral Cloning (BC)**: trains the prediction agent for the optimal policy:

$$\pi_{\theta^*} = \arg\min_{\pi_\theta \in \Pi} \mathbb{E}_{s\sim d_{\pi_E}}\left[l\left(\pi_\theta(s), \pi_E(s)\right)\right]$$

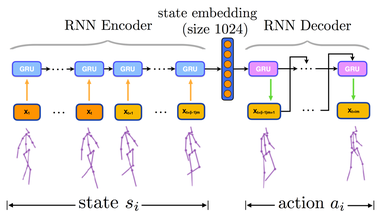
- **WGAIL-div**: alternatively continues training the prediction agent and a critic to learn an optimal policy using the objective function:

$$\min_{\pi_\theta} \max_{D_w \in \mathcal{S}\times\mathcal{A}\to\mathbb{R}} \left(\mathbb{E}_{\pi_\theta}[D_w(s,a)] - \mathbb{E}_{\pi_E}[D_w(s,a)] - k \mathbb{E}_{(\hat{s},\hat{a})\sim\mathbb{P}_u}\left[\|\nabla_{(\hat{s},\hat{a})} D_w((\hat{s},\hat{a}))\|^p\right]\right)$$

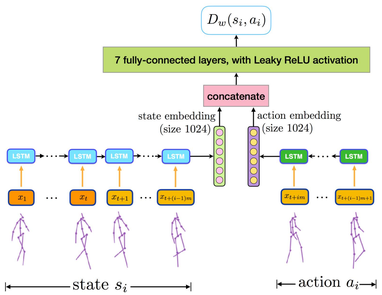


Model Architectures

- **Policy Generator Network**: we adopt a sequence-to-sequence architecture for our prediction agent policy generator network. It generates future pose sequences:



- **Critic Network**: we adopt a seven-layer fully connected neural architecture for our critic network in *WGAIL-div* training:



Experiment Results

- **Results on the Human 3.6M Dataset:** with the Mean Angle Error (MAE) performance metric

	Walking					Eating					Smoking					Discussion				
milliseconds	80	160	320	560	1000	80	160	320	560	1000	80	160	320	560	1000	80	160	320	560	1000
ERD [10]	1.30	1.56	1.84	2.00	2.38	1.66	1.93	2.88	2.36	2.41	2.34	2.74	3.73	3.68	3.82	2.67	2.97	3.23	3.47	2.92
LSTM-3LR [10]	1.18	1.50	1.67	1.81	2.20	1.36	1.79	2.29	2.49	2.82	2.05	2.34	3.10	3.24	3.42	2.25	2.33	2.45	2.48	2.93
SRNN [20]	1.08	1.34	1.60	1.90	2.13	1.35	1.71	2.12	2.28	2.58	1.90	2.30	2.90	3.21	3.23	1.67	2.03	2.20	2.39	2.43
Dropout-AE [12]	1.00	1.11	1.39	1.55	1.39	1.31	1.49	1.86	1.76	2.01	0.92	1.03	1.15	1.38	1.77	1.11	1.20	1.38	1.53	**1.73**
Residual [29]	0.32	0.54	0.72	0.86	0.96	0.25	0.42	0.64	0.94	1.30	0.33	0.60	1.01	1.23	1.83	0.34	0.74	1.04	1.43	1.75
Zero-velocity [29]	0.39	0.68	0.99	1.35	1.32	0.27	0.48	0.73	1.04	1.38	0.26	0.48	0.97	1.02	1.69	0.31	0.67	0.94	1.41	1.96
TP-RNN [8]	0.25	0.41	0.58	0.74	0.77	0.20	0.33	0.53	0.84	1.14	0.26	0.48	0.88	0.98	1.66	0.30	0.66	0.98	1.39	1.74
Ours	**0.21**	**0.34**	**0.53**	**0.67**	**0.69**	**0.17**	**0.30**	**0.52**	**0.79**	**1.13**	**0.23**	**0.44**	**0.86**	**0.95**	**1.63**	**0.27**	**0.56**	**0.82**	**1.34**	1.81

	Purchases					Sitting					Sitting down					Taking photo				
millisec	80	160	320	400	560	1000	80	160	320	400	560	1000	80	160	320	400	560	1000	80	160	320	400	560	1000
Residual	0.60	0.86	1.24	1.30	1.58	2.26	0.44	0.74	1.19	1.40	1.57	2.03	0.51	0.93	1.44	1.65	1.94	2.55	0.33	0.65	0.97	1.09	1.19	1.47
TP-RNN	0.59	0.82	1.12	1.18	1.52	2.28	0.41	0.66	1.07	1.22	1.35	1.74	0.41	0.79	1.13	1.27	1.47	1.93	0.26	0.51	0.80	0.95	1.08	1.35
Ours	**0.54**	**0.78**	**1.07**	**1.14**	**1.46**	**2.23**	**0.29**	**0.48**	**0.87**	**1.04**	**1.21**	**1.58**	**0.34**	**0.68**	**1.01**	**1.14**	**1.34**	**1.78**	**0.17**	**0.37**	**0.60**	**0.72**	**0.84**	**1.06**

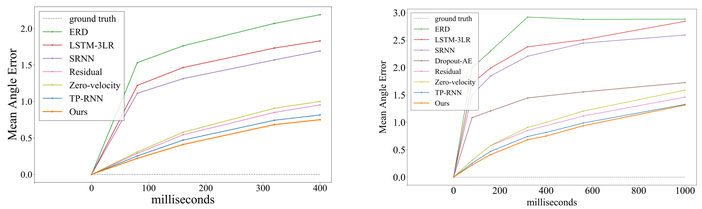
	Directions						Greeting						Talking on the phone						Posing					
millisec	80	160	320	400	560	1000	80	160	320	400	560	1000	80	160	320	400	560	1000	80	160	320	400	560	1000
Residual	0.44	0.69	0.83	0.94	1.03	1.49	0.53	0.88	1.29	1.45	1.72	1.89	0.61	1.12	1.57	1.74	1.59	1.92	0.47	0.87	1.49	1.76	1.96	**2.35**
TP-RNN	0.38	0.59	**0.75**	**0.83**	**0.95**	**1.38**	0.51	0.86	1.27	1.44	1.72	1.81	0.57	1.08	1.44	1.59	**1.47**	1.68	0.42	0.76	1.29	1.54	**1.75**	2.47
Ours	**0.27**	**0.46**	0.81	0.89	**0.95**	1.41	**0.43**	**0.75**	**1.17**	**1.33**	**1.62**	**1.72**	**0.54**	**1.05**	**1.40**	**1.56**	1.52	**1.67**	**0.27**	**0.55**	**1.16**	**1.41**	1.83	2.69

	Waiting						Walking dog						Walking together						Average over all 15 actions					
millisec	80	160	320	400	560	1000	80	160	320	400	560	1000	80	160	320	400	560	1000	80	160	320	400	560	1000
Residual	0.34	0.65	1.09	1.28	1.61	**2.27**	0.56	0.95	1.28	1.39	1.68	1.92	0.31	0.61	0.84	0.89	1.00	1.43	0.43	0.75	1.11	1.24	1.42	1.83
TP-RNN	0.30	0.60	1.09	1.31	1.71	2.46	0.53	0.93	1.24	1.38	1.73	1.98	0.23	0.47	0.67	0.71	0.78	1.28	0.37	0.66	0.99	1.11	1.30	1.71
Ours	**0.25**	**0.53**	**0.96**	**1.19**	**1.59**	2.39	**0.46**	**0.80**	**1.12**	**1.31**	**1.65**	**1.86**	**0.19**	**0.41**	**0.61**	**0.64**	**0.67**	**1.15**	**0.31**	**0.57**	**0.90**	**1.02**	**1.23**	**1.65**

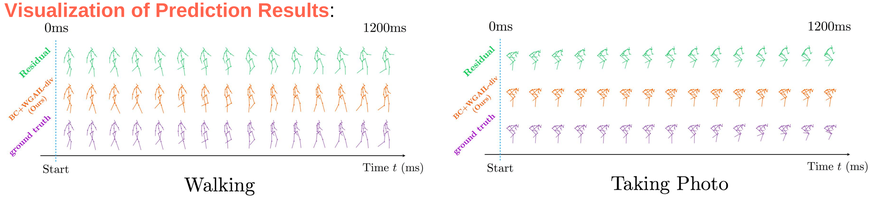
- **Ablation Study:**

millisec	80	160	320	400	560	1000
BC	0.23	0.43	0.72	0.78	0.96	1.36
WGAIL-div	0.23	0.44	0.74	0.80	0.98	1.33
BC+WGAIL-div	**0.21**	**0.41**	**0.69**	**0.75**	**0.94**	**1.32**

- Comparison of average MAE over different actions in short-term (left) and long-term (right) prediction



- **Visualization of Prediction Results**:



Acknowledgement: We would like to thank Mindtree and Panasonic for their support.