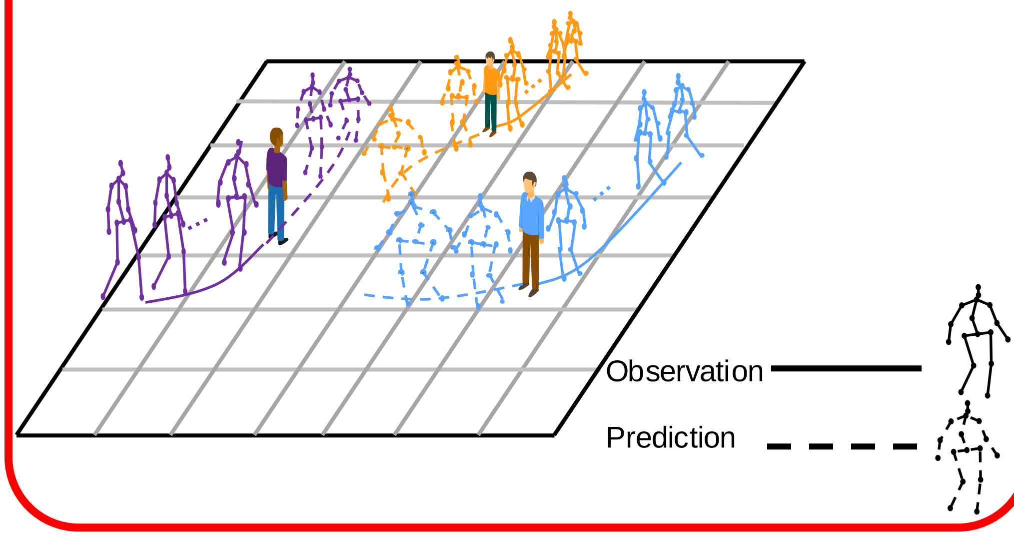
Multi-Transmotion: Pre-trained Model for Human Motion Prediction

Yang Gao, Po-Chien Luan, Alexandre Alahi

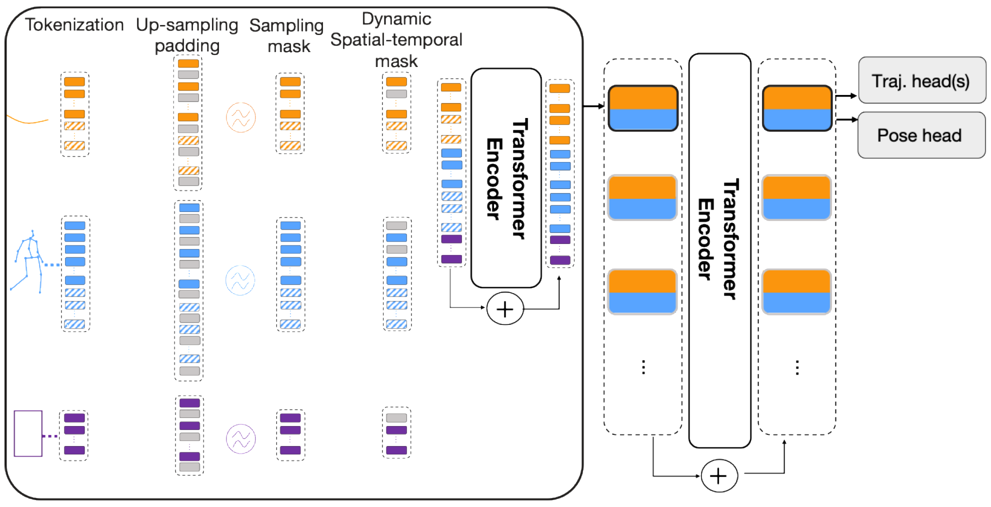
Overview

Task: Predicting future **human trajectory** and **3D pose** from observational data.
Motivation: Missing of foundation model to predict human motion with different fps, horizon, keypoints settings.
Approach:
* Create a Unified Human Motion Data Framework
* Propose a pre-trained model that can adapt to varying frame settings and keypoints settings.

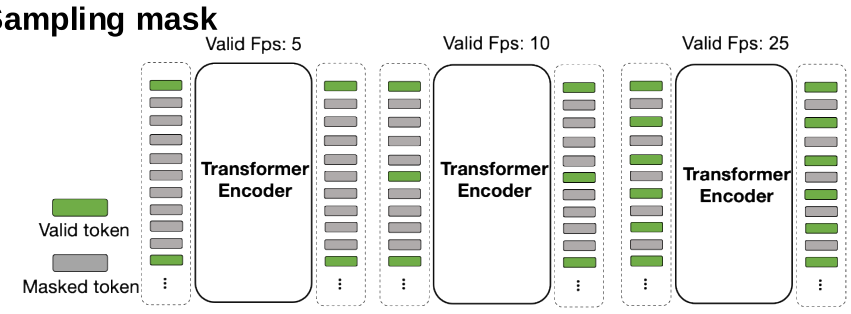


Method

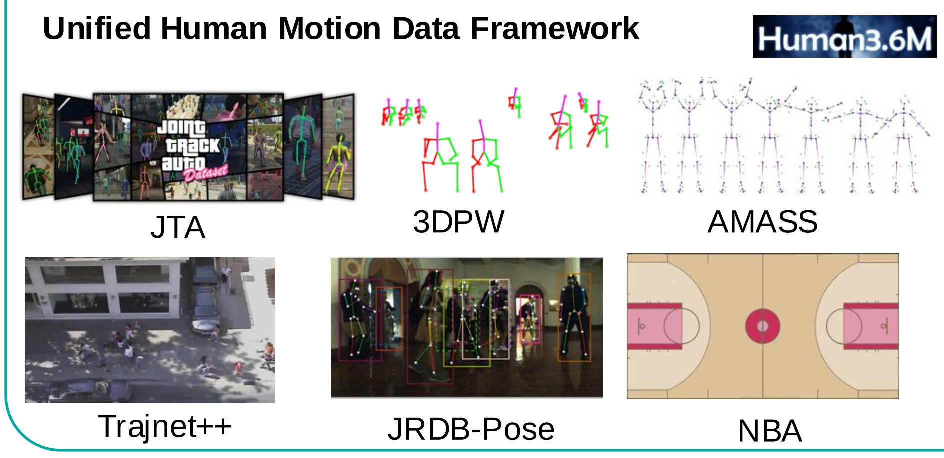
Model architecture



Sampling mask

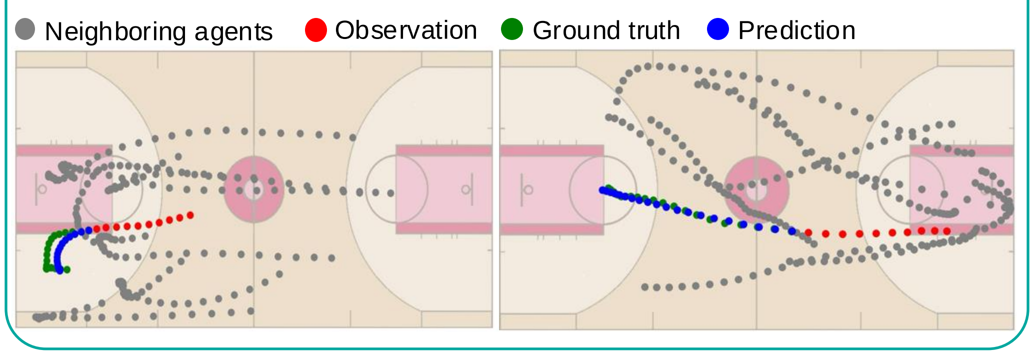


Unified Human Motion Data Framework

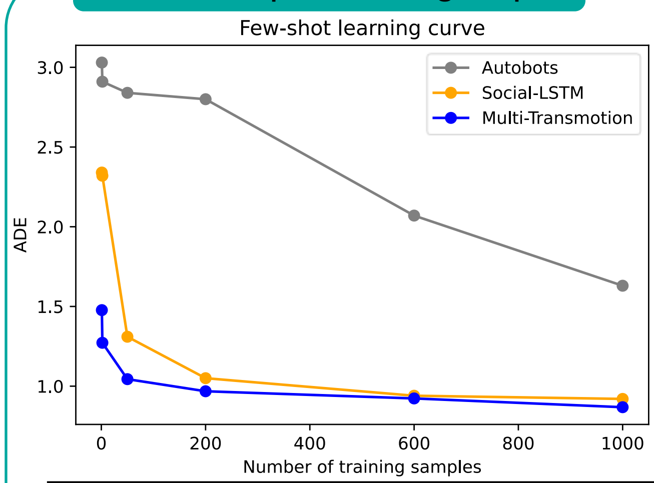


Trajectory pred. on NBA

Models	Venue	$MinADE_{20}$	$MinFDE_{20}$
Social-GAN	CVPR 18	1.59	2.41
Trajectron++	ECCV 20	1.15	1.57
GroupNet	CVPR 22	0.96	1.30
Leapfrog	CVPR 23	0.81	1.10
Social-Transmotion	ICLR 24	<u>0.78</u>	<u>1.01</u>
Multi-Transmotion	(ours)	**0.75**	**0.97**



How does pre-training help?

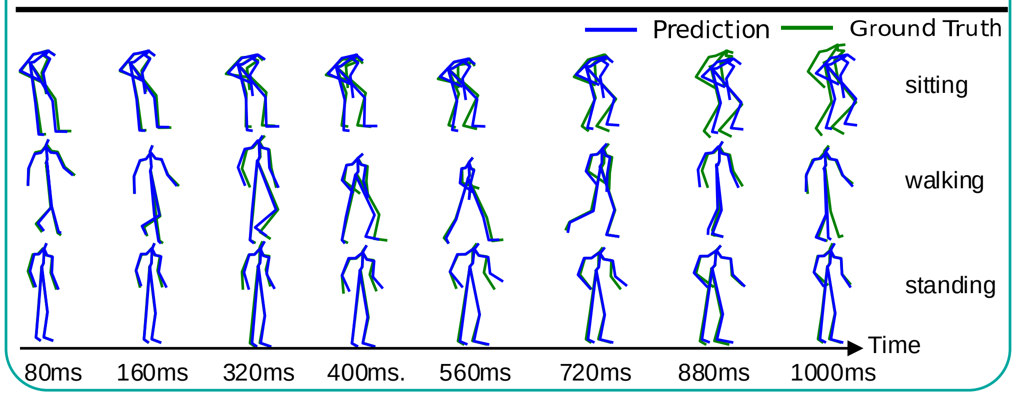


Key takeaways:
* More data efficient
* Better generalization

Datasets	Pre-trained model	Specific model
NBA	**0.75/0.97**	0.77/0.98
Trajnet++	**0.54/1.13**	0.57/1.22
AMASS	**66.91**	69.58
3DPW	**73.74**	76.77

Pose pred. on AMASS

Models	160 ms	400 ms	720 ms	1000 ms
ConvSeq2Seq	36.9	67.6	87	93.5
LTD-10-10	**19.3**	44.6	75.9	91.2
LTD-10-25	20.7	45.3	65.7	75.2
HRI	20.7	<u>42</u>	<u>58.6</u>	67.2
ST-Trans	21.3	42.5	**58.3**	**66.6**
Multi-Transmotion	**19.3**	**41.4**	<u>58.6</u>	<u>66.9</u>



Application in robot navigation

	Completion time in seconds (gain)	Collision rate (gain)
Robot navigation w/o our predictor	16.46	1.93%
Robot navigation w/ our predictor	**16.20 (+2%)**	0.39% (+80%)