What and How? Jointly Forecasting Human Action and Pose

Yanjun Zhu, Yanxia Zhang, Qiong Liu, Andreas Girgensohn, David Doermann

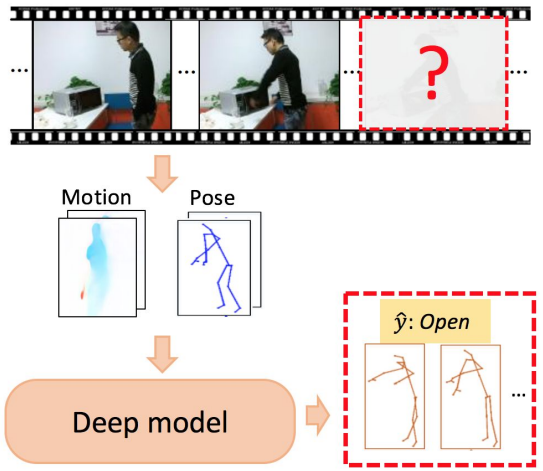
Motivation

- Human can **anticipate** actions.
- Forecasting actions can prevent harmful actions.

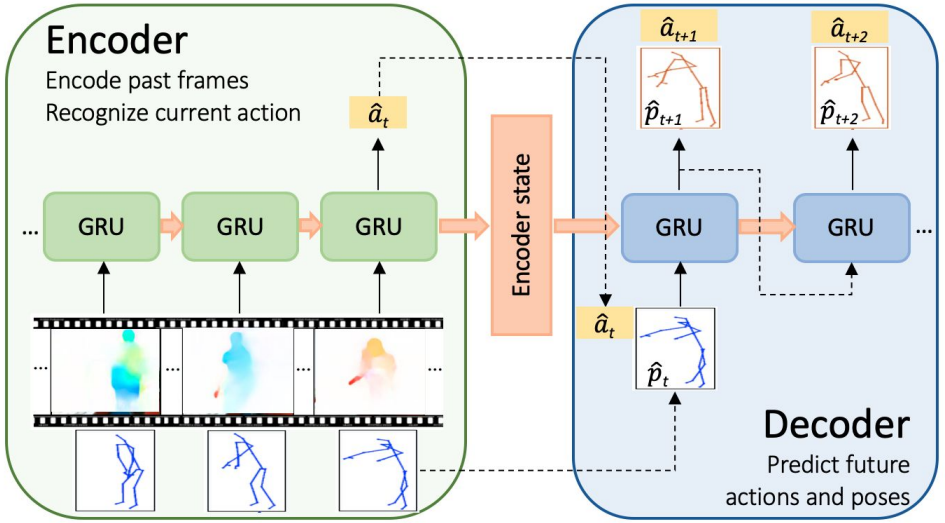
- **What?** Future action categories
- **How?** Future body positions

Challenges and Goals

- Forecasting future action labels and human pose jointly.
- The two tasks help each other.



Model Architecture



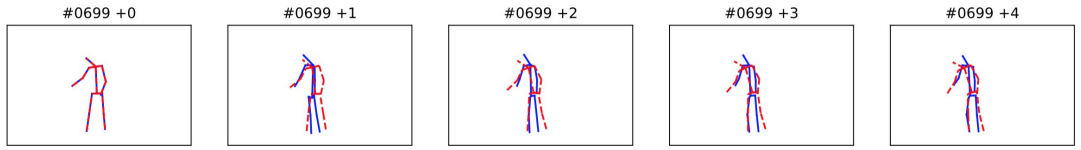
❖ Contributions

- A novel seq2seq recurrent model to forecast actions and poses.
- Skeleton pose and global motion are combined via image coordinates.
- A new evaluation metric, Scaled Changing Distance (SCD), to measure the continuity of predicted action sequences.

❖ Future Work

- Improve the smoothness of predicted pose sequence

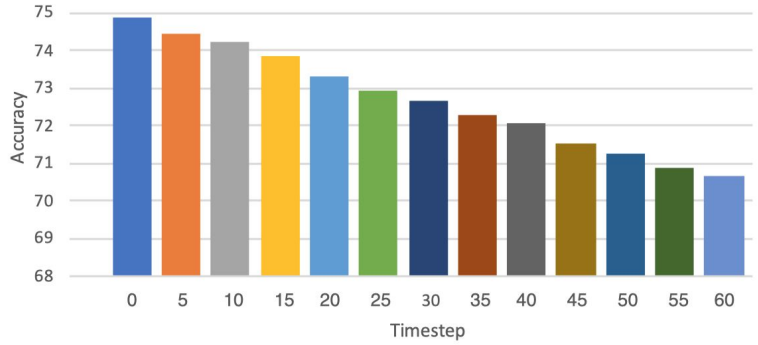
Results: Pose Forecasting



Results: Action Prediction

Forecasting Methods	Acc_0	Acc_1	Acc_4	avg.
Two-stream + FC	-	22.37	-	-
Two-stream + LSTM	-	55.60	-	-
iDT + LSTM	-	65.20	-	-
Multi-label LSTM (ours)	74.95	74.74	**74.56**	74.77
Multi-task Seq2seq (ours)	**76.36**	**76.12**	74.34	**75.21**

- Action prediction accuracy on IkeaDB dataset.



- Action prediction accuracies of different lengths, from 0 to 60 timesteps.

Results: Joint Learning

Methods	Acc_{avg} (%)	$F1$	E_{pose} (e-2)
RNN-SW	-	0.60	-
JCR-RNN	-	0.65	-
Zero-velocity	87.72	0.67	4.48
Ours (joint learning)	**88.11**	**0.68**	**3.84**
Ours (action only)	87.94	0.66	-
Ours (pose only)	-	-	4.29

- Action and pose forecasting evaluation on OAD dataset.