# What and How? Jointly Forecasting Human Action and Pose

<u>Yanjun Zhu</u>, Yanxia Zhang, Qiong Liu, Andreas Girgensohn, David Doermann





## Why Forecasting?

- Human daily life
- Quick response
- Imagination



Lan, T., Chen, T. C., & Savarese, S. (2014). A hierarchical representation for future action prediction. In ECCV (pp. 689–704).

### Why Action and Pose Together?

- Healthcare robot
- WHAT is the future action and HOW will people perform it?
- Pose keypoints







https://spectrum.ieee.org/automaton/robotics/medical-robots/riba-ii-healthcare-robot-gets-bigger-muscles-cuter-ears

# Why Joint Learning?

- Multi-task learning
- Category classification and pose keypoints regression
- The two tasks would help each other.



#### **Model Architecture**



#### **Results:** Action

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.

### **Results: Forecasting Length**



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

#### **Results:** Pose

Forecasting results (blue) and ground truth (red)



### **Results: Joint Learning**

Methods	$Acc_{avg}$ (%)	F1	$E_{\rm pose}$ (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.

