

Learning Robot Activities from First-Person Human Videos Using Convolutional Future Regression

Wenjie Niu

May 21, 2018

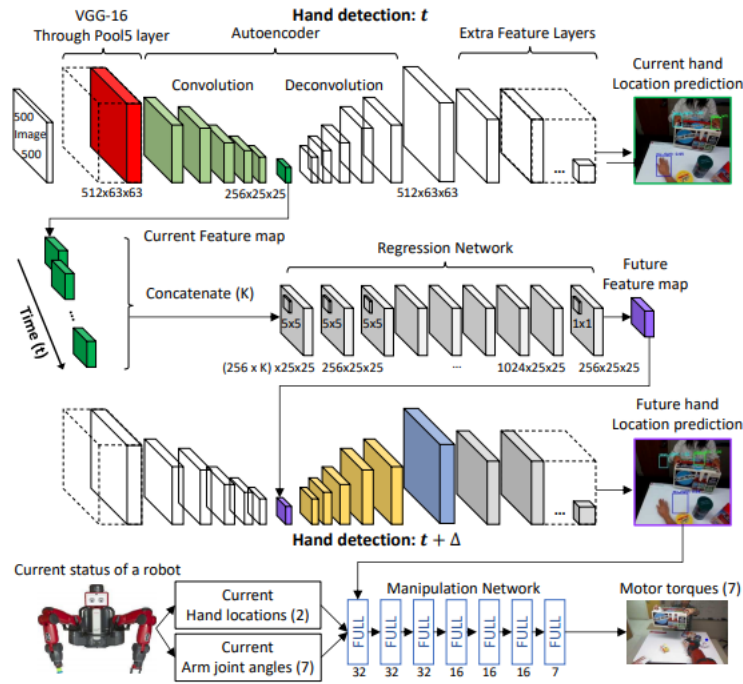


Figure 1: Overview of our approach

As shown in Fig 1, this regression network does not require activity labels or hand/object labels in videos for its training. The manipulation network (the last row) generates robot control commands given current robot joint state, current robot hand locations, and predicted future robot hand locations.

[1] We employ two components for achieving the goal. The first component is the perception component that consists of two fully convolutional neural networks:

1. an extended version of the Single Shot MultiBox Detector (SSD) for human hand detection and

Method	Evaluation					
	Precision		Recall		F-measure	
Hand-crafted features	0.30	0.37	0.15	0.19	0.20	0.25
Hands only	4.78	3.70	5.06	4.06	4.87	3.81
SSD w/ future Annot.	27.53	23.36	9.09	8.96	13.23	12.62
Deep Regressor (ours): K=1	27.04	16.50	21.71	14.71	23.45	14.99
Deep Regressor (ours): K=5	29.97	15.37	23.89	16.45	25.40	15.51
Deep Regressor (ours): K=10	36.58	16.91	28.78	17.96	30.90	17.02

Table 1: Evaluation of predicting 1-sec future hand locations

2. a future regression network to predict the intermediate scene representation corresponding to the future frame. The second component is the manipulation component that maps 2-D hand locations in the image coordinate to the actual motor control using fully connected layers.

Table 1 shows quantitative results of our future hand prediction. Here, K represents number of frames we used as an input for our regression network. We can clearly observe that our approach significantly outperforms all baselines, including the state-of-the-art object detector SSD modified for the hand prediction.

References

- [1] Lee, Jangwon and Ryoo, Michael S. *Learning Robot Activities From First-Person Human Videos Using Convolutional Future Regression*. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. 2017. **July**: 1,2