

Learning Human Pose from Multiple Noisy Kinect Data

Yiying Li

Input: A Single Image

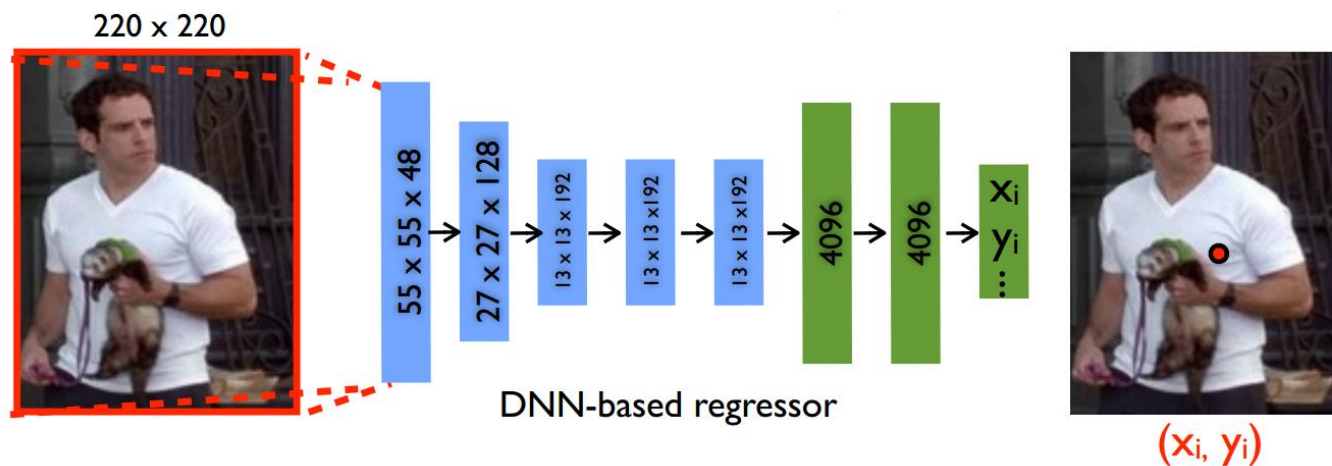


One Kinect

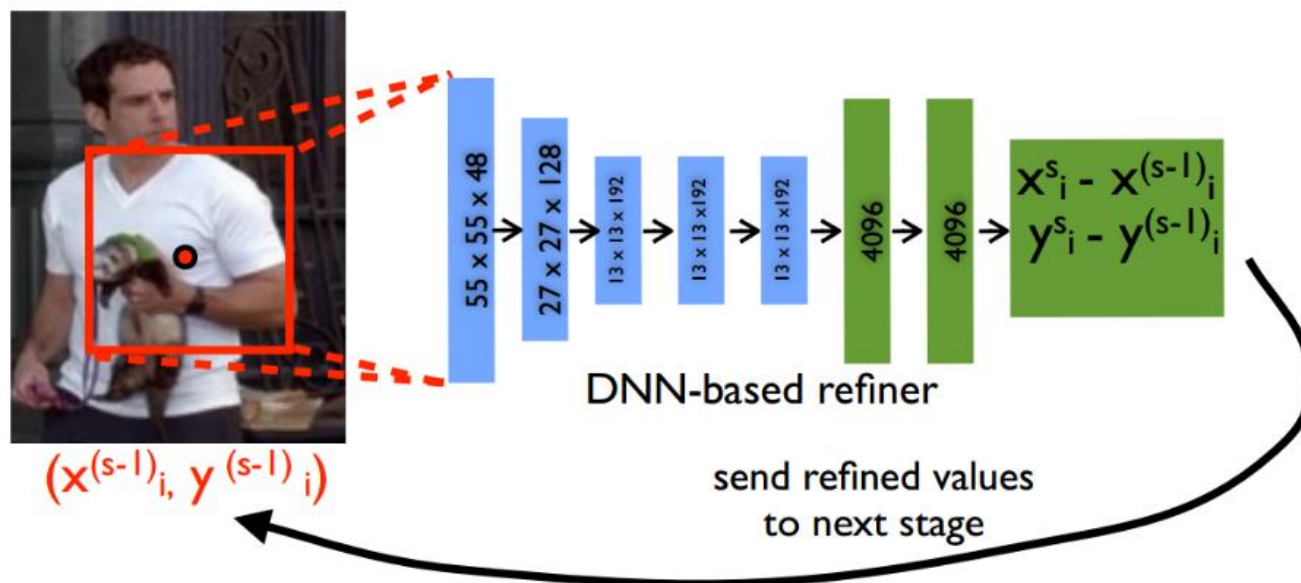
- Use one Kinect as ground truth for skeleton data
- Person Detector
 - Normalized positions in bounding box
- Use a cascade of Deep Neural Networks
- Loss is a linear regressor
 - Sum of L2 norm

DNN Architecture

Whole Person
Stage 1



Bounding Box per Joint
Stage 2+



Multiple Kinects



Multiple Kinects

- Kinects Skeletons are noisy
- Use multiple Kinect to detect how each Kinect
 - Basically characterize noise
- Increase the loss of by the noise
 - $L = \sum_{i=1}^n ||x_i - f_i(x; \theta)||^2 + \beta < \text{some term} >$
- Working on seeing if this error can be modeled implicitly through another hidden layer

Problems

- Overfitting to the environment
 - Add images from other data sets
 - Mix in a motion capture data set as non noisy training data
- No ground truth dataset
 - Manual labeling of test set