Multi-person Pose Estimation using Depth and Thermal Data Modalities

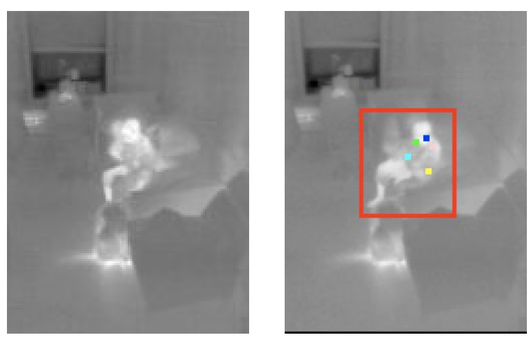
Rishab Mehra [1,2] Meena Chetty [1] John Kamalu [1]
[1] Stanford University, School of Engineering [2] Stanford AI Lab

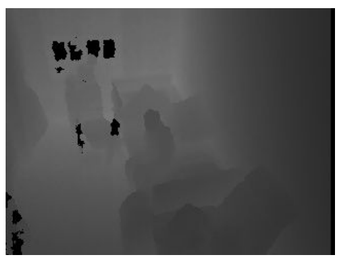
Introduction

Pose Estimation helps us perform tasks such as activity recognition and risk detection. In hospitals and other sensitive environments, we often do not have access to RGB data due to privacy concerns.

In this work, we create a novel human pose dataset with only thermal and depth modalities, which have been approved by the Senior Home authorities, and then transfer RGB pose models[1] to use only these modalities.

Data





Upper left: thermal image. Lower left: Depth Image. Upper right: labeled image

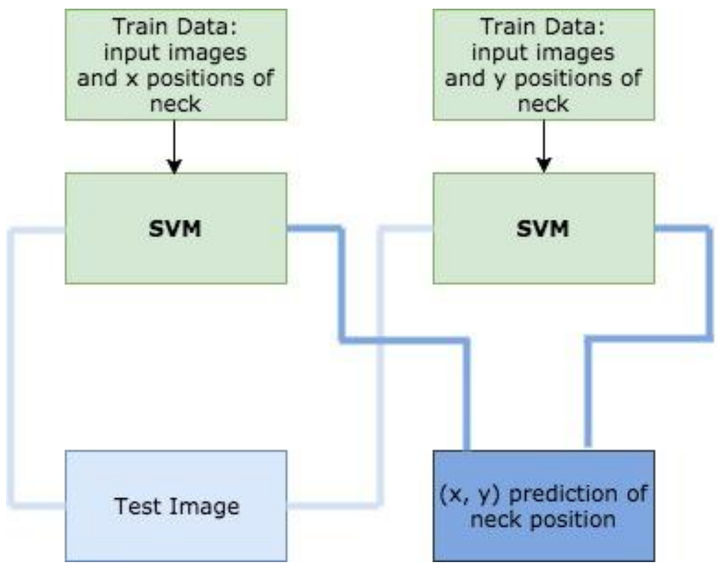
Data was collected over 2 days in a San Francisco senior home:

- Thermal Data
  - 8 frames/sec
- Depth Data
  - 24 frames/sec

We manually labeled 5 upper body joints for training: base of neck, left elbow, left shoulder, right elbow, right shoulder

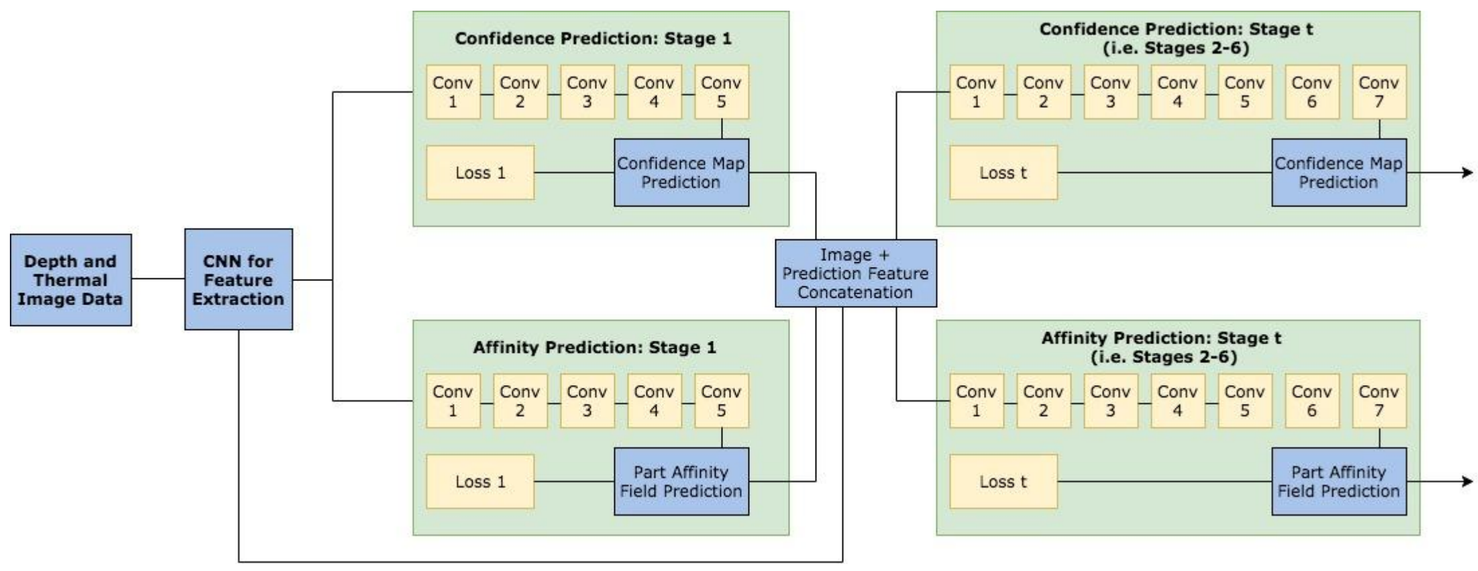
Methods

Baseline: SVM for Joint Prediction



Baseline predicts location of only the neck base joint using 2 SVM's: 1 to predict x position and 1 to predict y position.

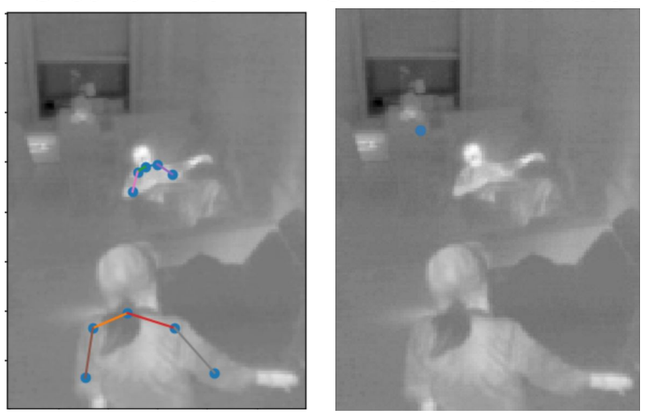
Adapted RGB Confidence Map[1] and Part Affinity Field Model for Depth and Thermal Data

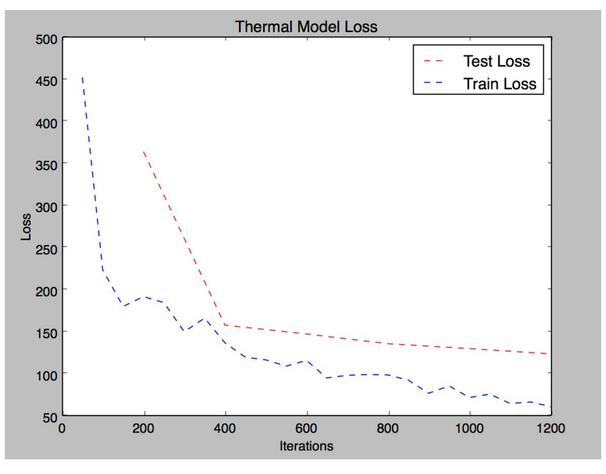


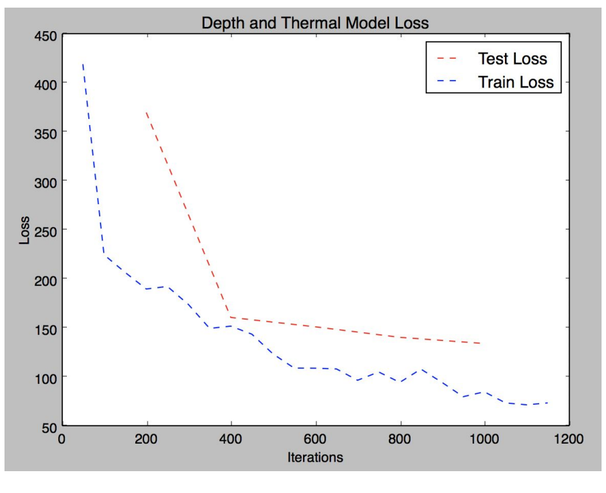
2-branch 7-layer convolutional model to predict Confidence Maps of limbs and Part Affinity Fields for the relationships between body parts using all 5 labeled joints.

Results

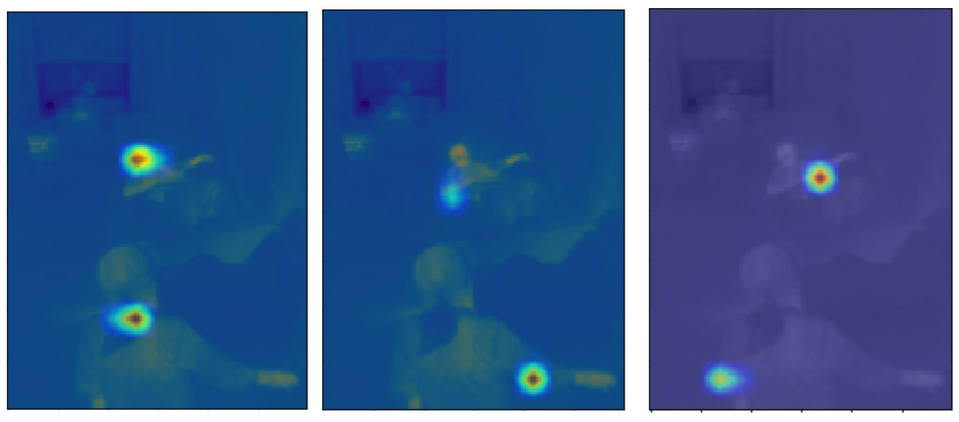
Ground Truth **SVM Neck Prediction**



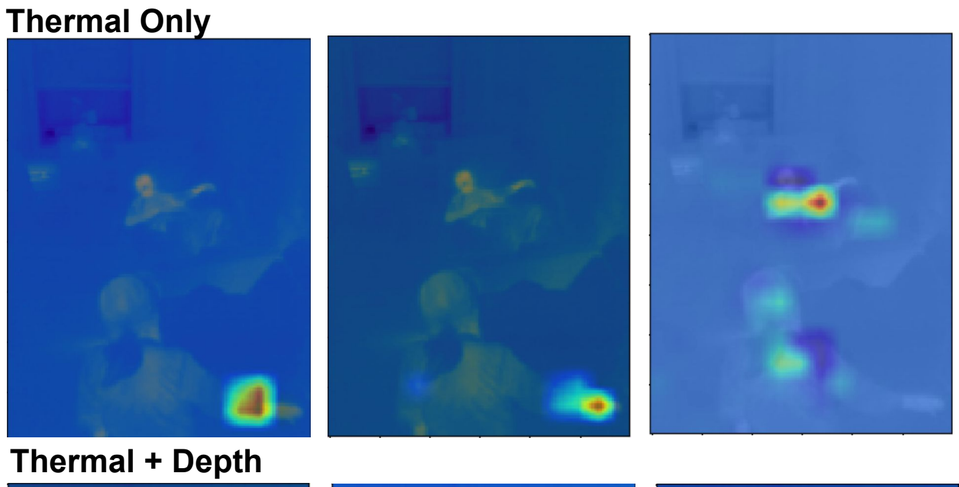




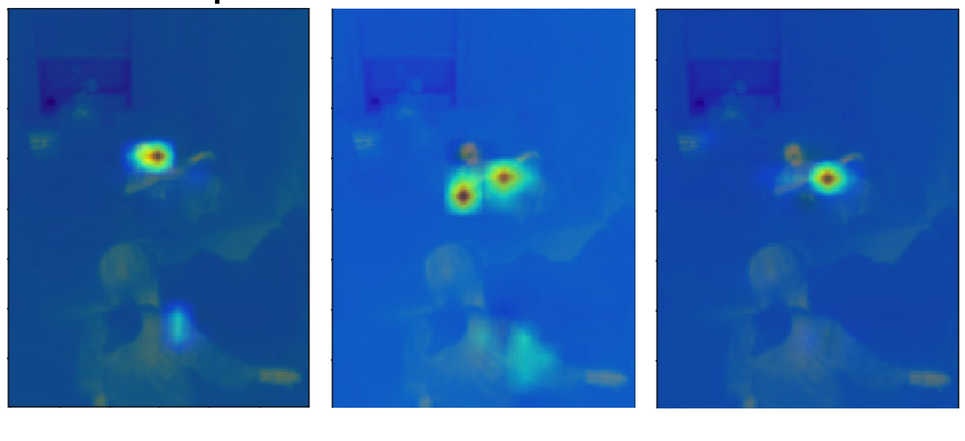
Neck Ground Truth **Right Elbow** **Left Elbow**



Thermal Only



Thermal + Depth



Acknowledgements

We thank the members of the Stanford Vision and Learning Lab, particularly Serena Yeung, for allowing us to use the Senior Home Data for the project, and providing us with GPU resources to train the models. We also thank Onlock for allowing us to collect data at their Senior homes. Finally, we thank Albert Haque and Sanyam Mehra for their useful comments throughout the quarter.

Citations

1. Cao, T. Simon, S.-E. Wei, and Y. Sheikh. Realtime multiperson 2d pose estimation using part affinity fields. In CVPR, 2017.
2. Abdolrahim, A. Kadkhodamohammadi, M. Gangi, N. Mathelin, and Padoy. Articulated clinician detection using 3d pictorial structures on rgb-d data. In CVPR, 2016.
3. S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh. Convolutional pose machines. In CVPR, 2016.