# Egocentric Height Estimation

Jessie Finocchiaro
Aisha Urooj Khan
Dr. Ali Borji

# Compact 3D CNN

- Used Shmuel Poleg's CNN and made it regression based*
- Made last layer have one output and ReLU activation

Poleg, Yair, Ariel Ephrat, Shmuel Peleg, and Chetan Arora. "Compact CNN for Indexing Egocentric Videos." 2016 IEEE Winter Conference on Applications of Computer Vision (WACV) (2016): n. pag. Web.

# Network input and output

- Input: Sparse Optical Flow Vectors
  - 60 frames concatenated for x and y directions
  - 32x32x120
  - Typically 10-20 of these inputs per video
- Output
  - A continuous measurement for each 60 frame segment of the video.
  - Take the average of results to get estimate for entire video

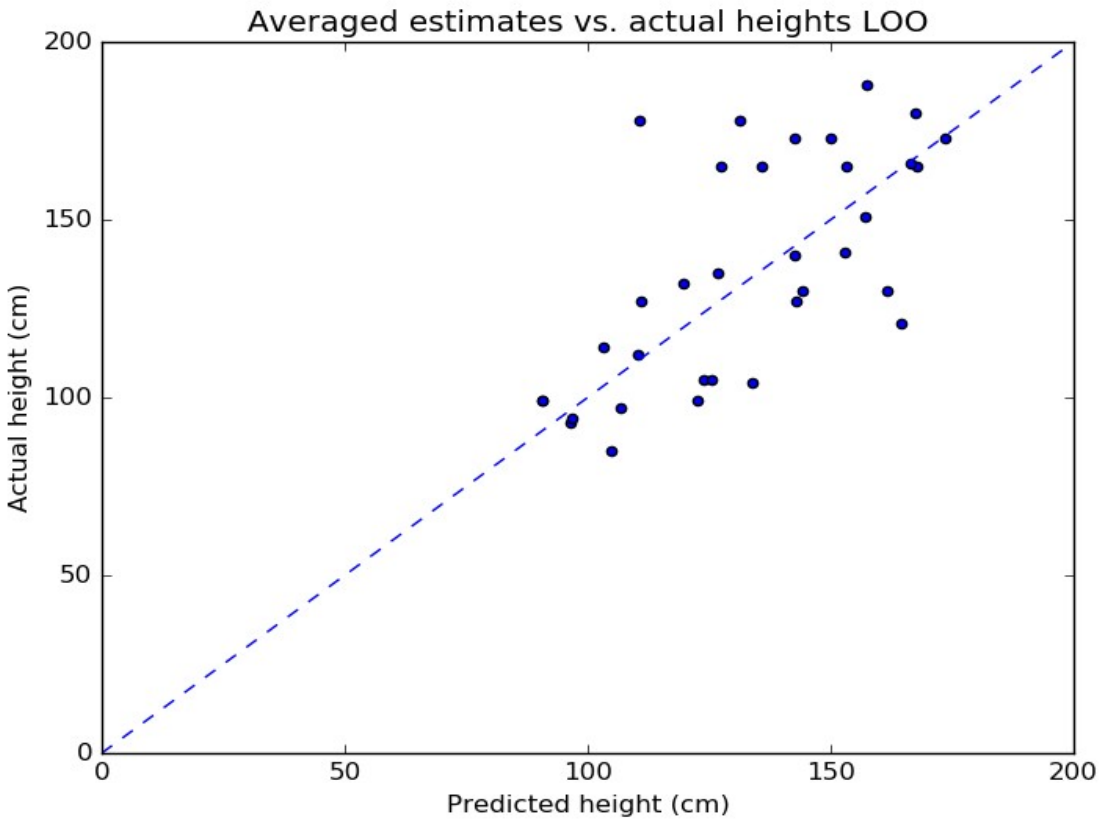
# Compact 3D CNN

- Normalized data between 0 and 1
- Tried training network with learning rates of 0.01, 0.001, 0.0001

| | Actual | Network | Retrained; decreased LR |
|---|---|---|---|
| Chest | 121 | 121.087 | 164.48 |
| Waist | 99 | 92.96 | 90.7 |
| Head | 166 | 174.56 | 166.457 |
| Loss | ---- | ---- | .1058 |

# Leave One Out training

- 10 subjects
- Train on 9, test on 1
- R^2 = .4557
- MAE = 18.03 cm

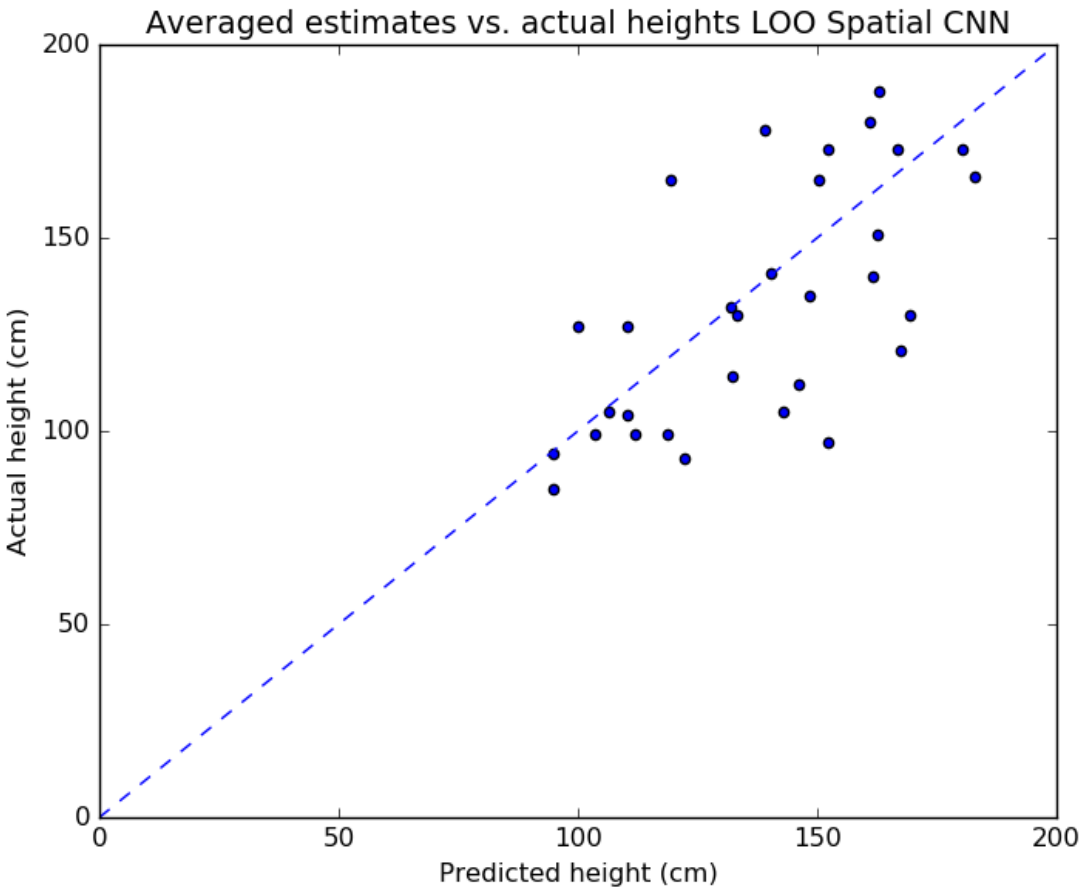


# Spatial CNN

- Use same CNN, but instead of inputting optical flow vectors, input is pixel intensities
- Condense input from 120 frames to 60 frames

# Spatial CNN Results- LOO

- R^2 = .4267
- MAE = 19.51 cm



# Next steps

- Will combine temporal and spatial CNNs to make a two-stream CNN to see if results improve
- Person identification from a small dataset
  - Try to improve the results of Peleg's paper *An Egocentric Look at Video Photographer Identity*