

---

# Personalized Pose Estimation for Body Language Understanding

Zhengyuan Yang, Jiebo Luo

Department of Computer Science  
University of Rochester



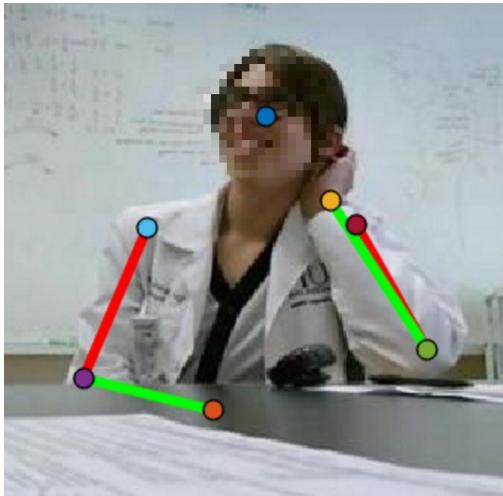
HAJIM  
SCHOOL OF ENGINEERING  
& APPLIED SCIENCES  
UNIVERSITY OF ROCHESTER

DEPARTMENT OF  
COMPUTER SCIENCE



# Pose Estimation

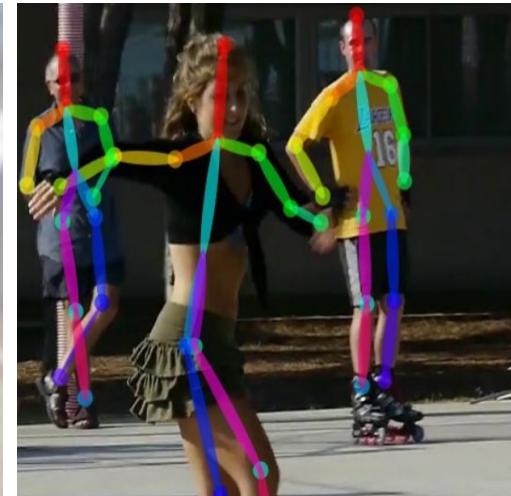
- Locating the keypoints on human body.



Upper body  
pose estimation



Full body  
pose estimation



Multi-person  
pose estimation



# Motivation

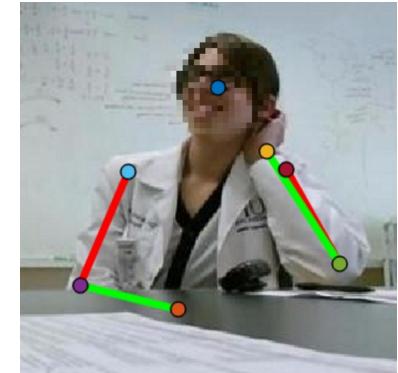
- Public Security



- Human-Computer Interaction



- Video Analysis



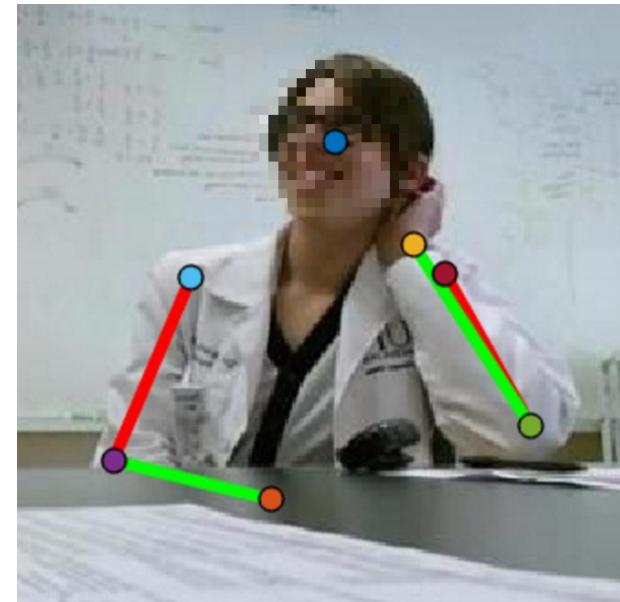
# Literature Review

- Human motion studies with wearable sensors  
Using wearable sensors to record the motion of body parts.
- Direct Mapping methods  
Regressing image features directly into a vector of joints coordinates.
- Parts-based methods  
Decomposing appearances into body part image patches.
- Heatmap methods  
A set of confidence maps as the regression output.

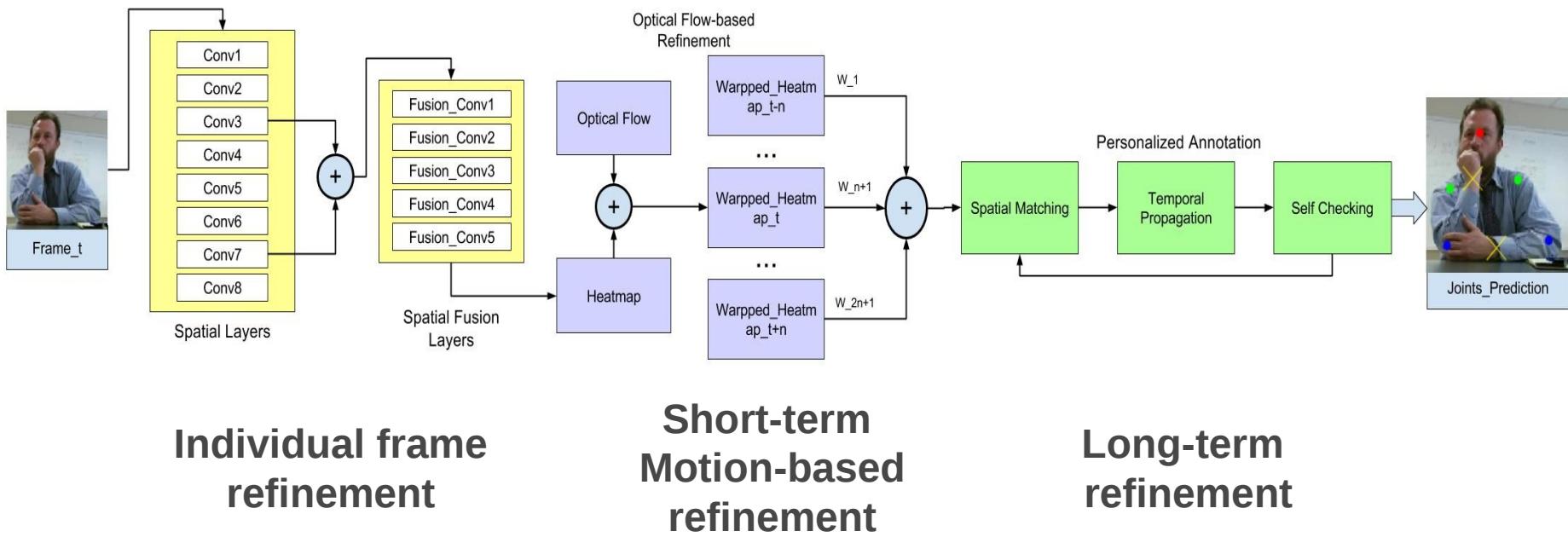


# Motivation beyond Typical Solutions

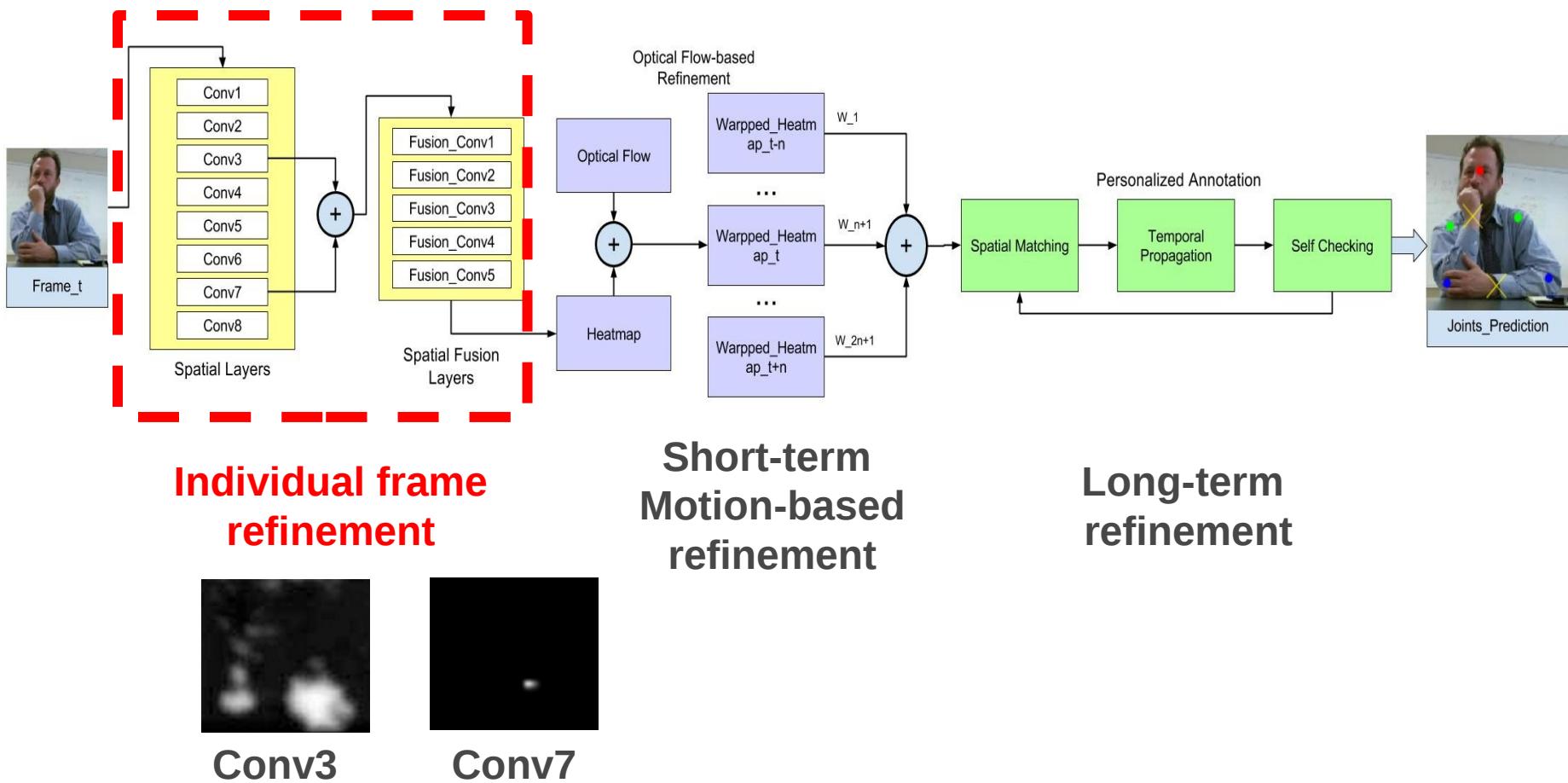
- Seeking perfect results on stable videos.
- Application in conversation monitoring:
  - Less motion
  - Accuracy and stability.



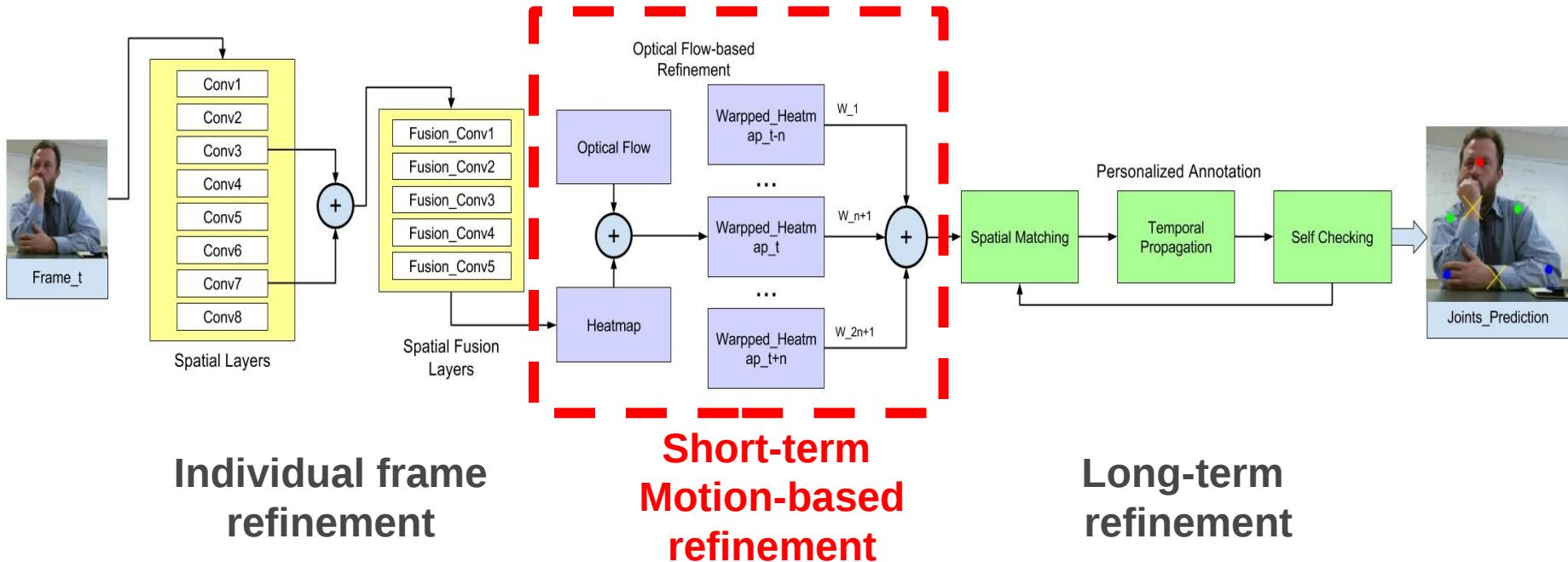
# Our Framework



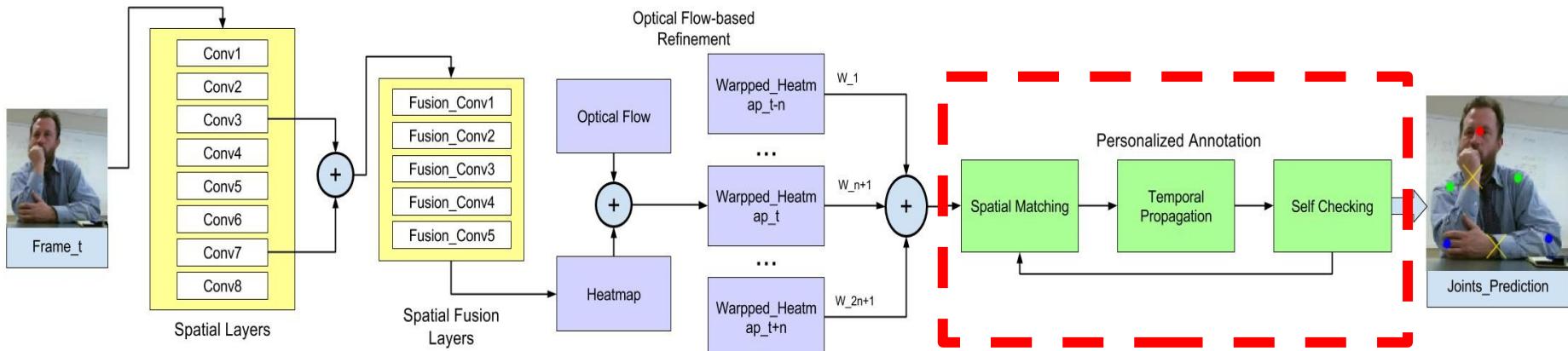
# Individual Frame Refinement



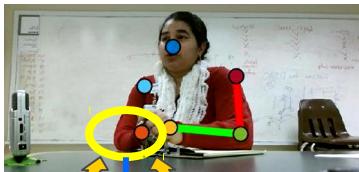
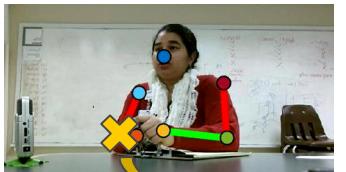
# Short-term Refinement



# Long-term Refinement



Individual frame  
refinement



Short-term  
Motion-based  
refinement



Long-term  
refinement



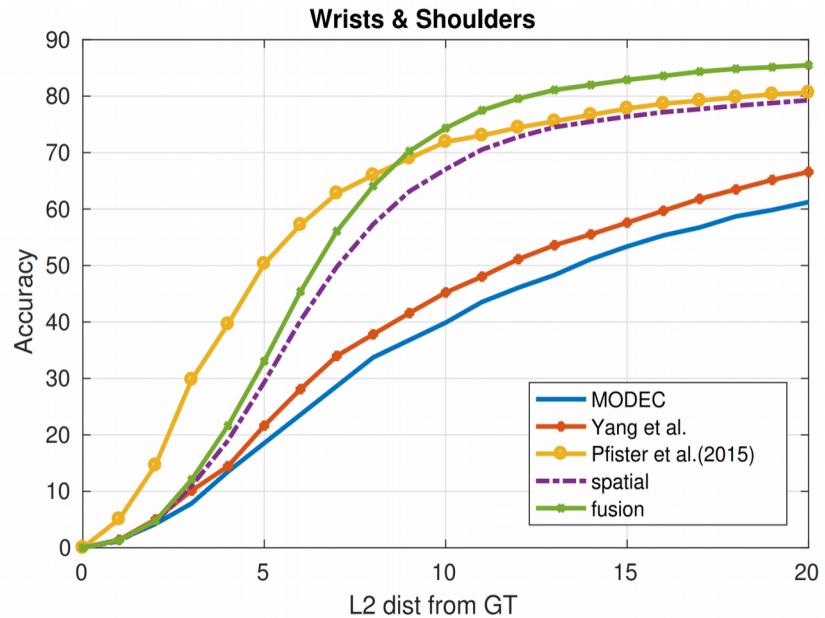
# URMC Dataset

- Eight videos lasting from 15 to 30 minutes. Labeled with joints positions.
- A scene depicts the conversation between a patient and a psychiatrist. To infer symptoms with body languages.
  - Sitting behind table: only upper body.
  - Less body part motion.



# Experimental Results

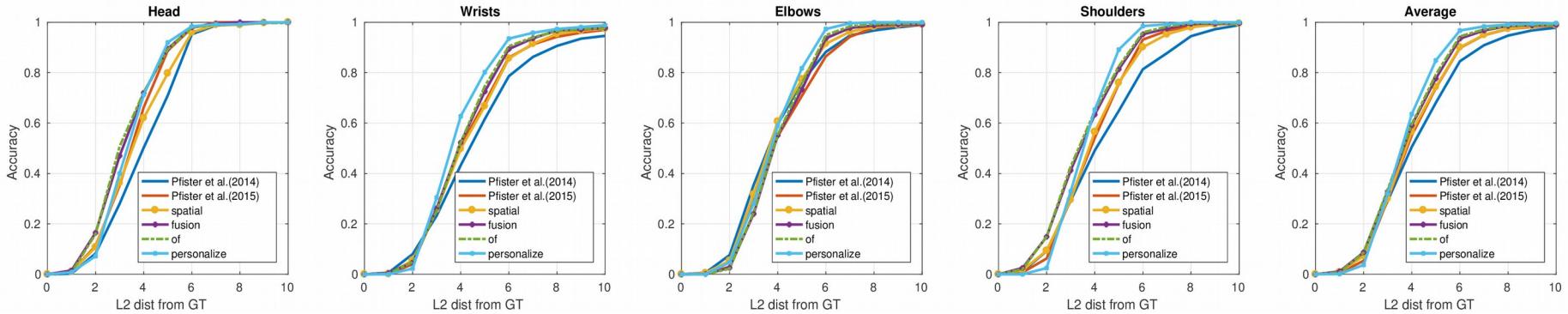
- Quantitative Evaluation
- Measurements on public datasets: FLIC, ChaLearn.
- Radius distance threshold for accuracy measurement.



# Experimental Results

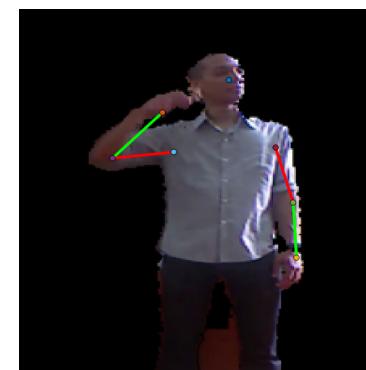
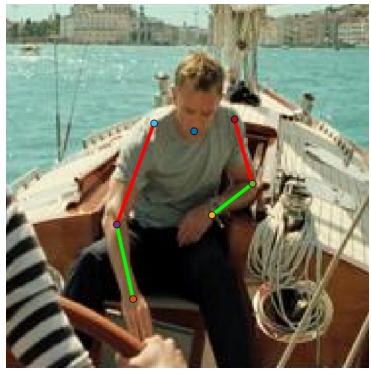
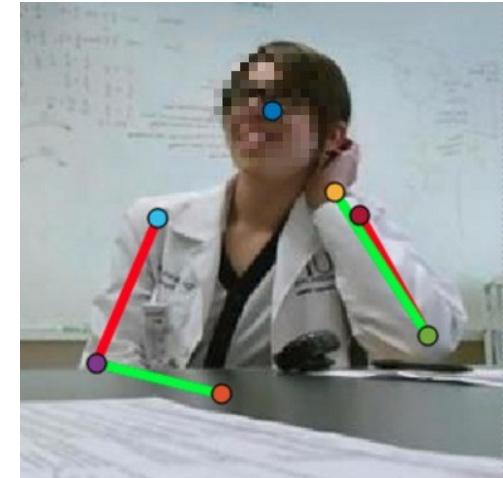
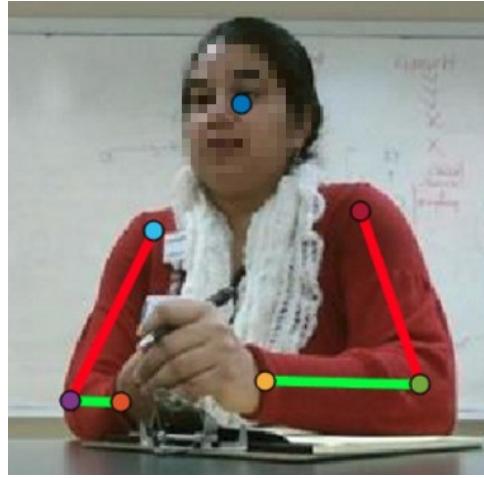
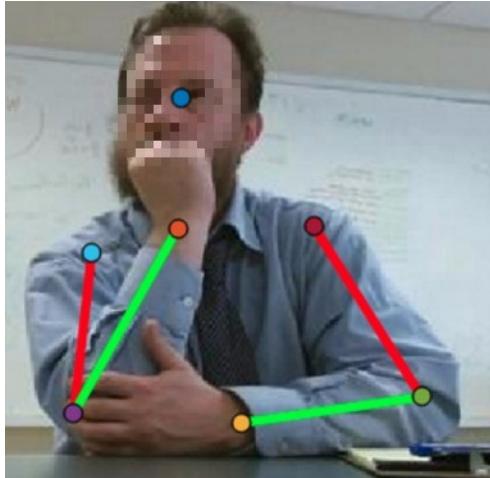
- Quantitative Evaluation
- Best Performance on all joints.
- Prove the effectiveness of the three refinements.

Method	Head	Wrsts	Elbws	Shldrs	Avg.
Pfister et al. [16]	76.8	65.4	75.6	70.9	71.5
Pfister et al. [12]	88.8	71.3	69.6	84.8	77.1
Spatial	79.8	66.8	77.4	76.0	74.2
Fusion	89.5	72.5	73.3	81.4	77.7
Fusion + OF	89.3	74.6	75.6	82.5	79.3
Fusion + OF + P	<b>92.0</b>	<b>80.2</b>	<b>81.8</b>	<b>89.2</b>	<b>84.9</b>



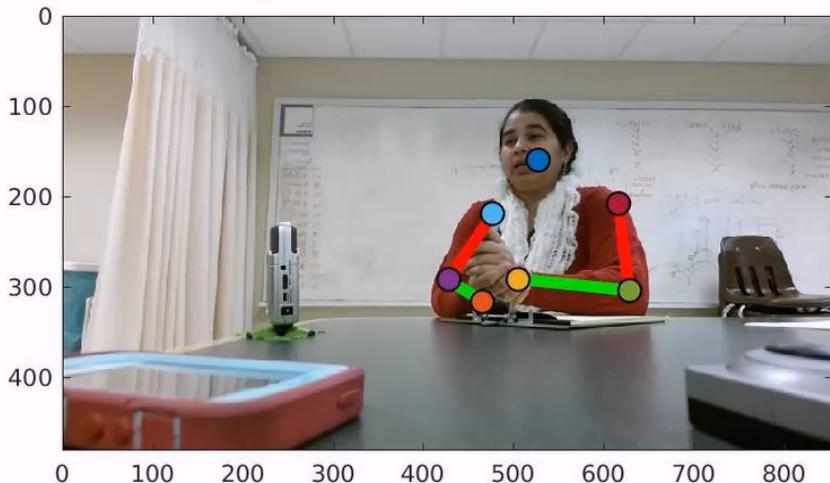
# Experimental Results

- Qualitative Evaluation



# Experimental Results

- Qualitative Evaluation



# Summary

- We propose a framework for human pose estimation under stable videos with the three refinement steps. These three steps are:
  1. A fully convolutional spatial fusion architecture for frame level refinement.
  2. Optical flow based short-term refinement.
  3. Personalized annotation step for refinement with long-term information.
- We also establish a psychiatric conversation dataset (URMC dataset) for subtle body language extraction and mental disorder detection.



**Thank you for your attention!  
Questions?**

Contact: [jiebo.luo@gmail.com](mailto:jiebo.luo@gmail.com)

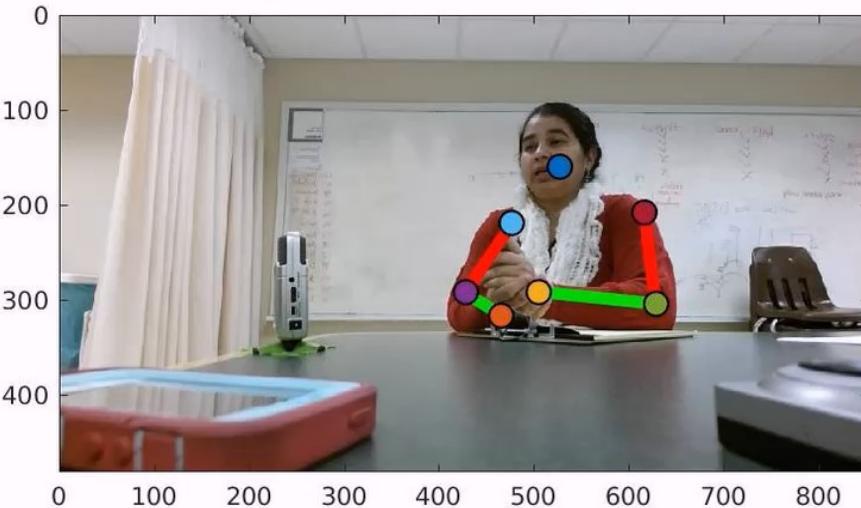


# Back Up Slides

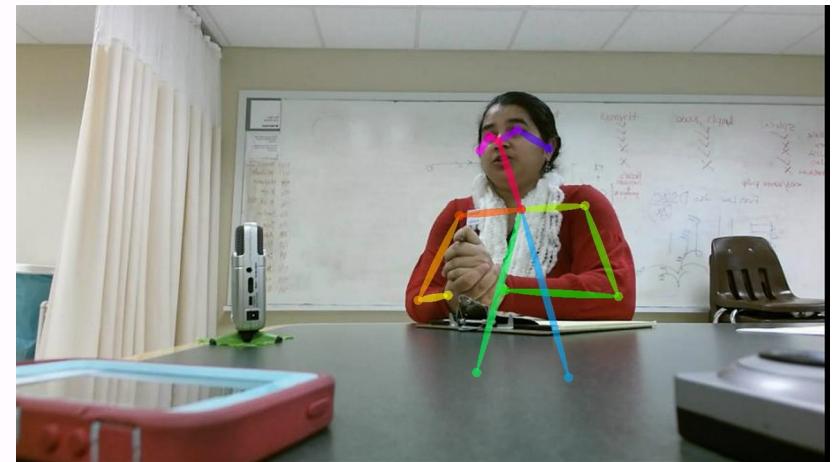


# Experimental Results

- Comparing to OpenPose



Proposed Method

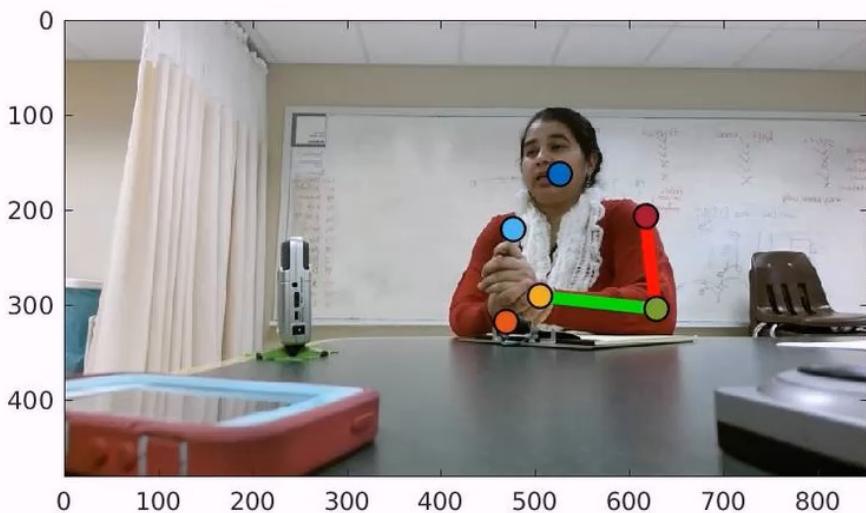


OpenPose

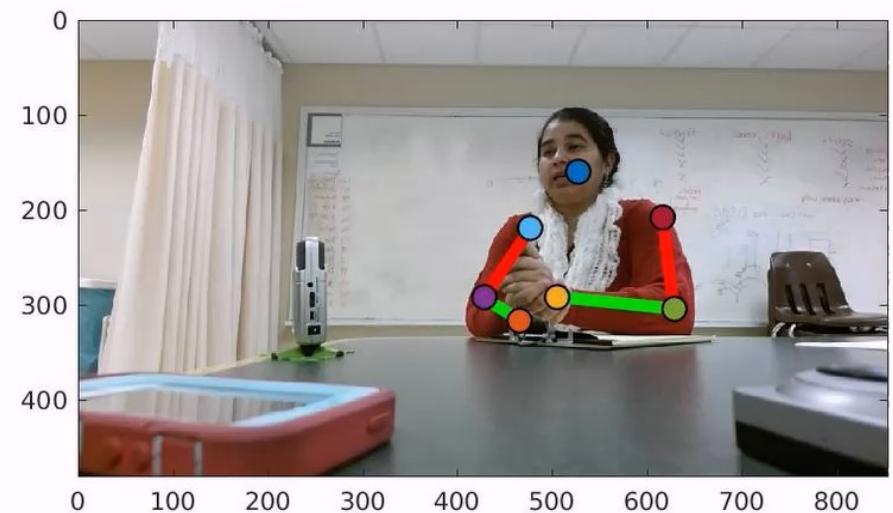


# Experimental Results

- Demo of personalize step



Before personalized step



After personalized step



# Experimental Results

- Further example on occluded case

