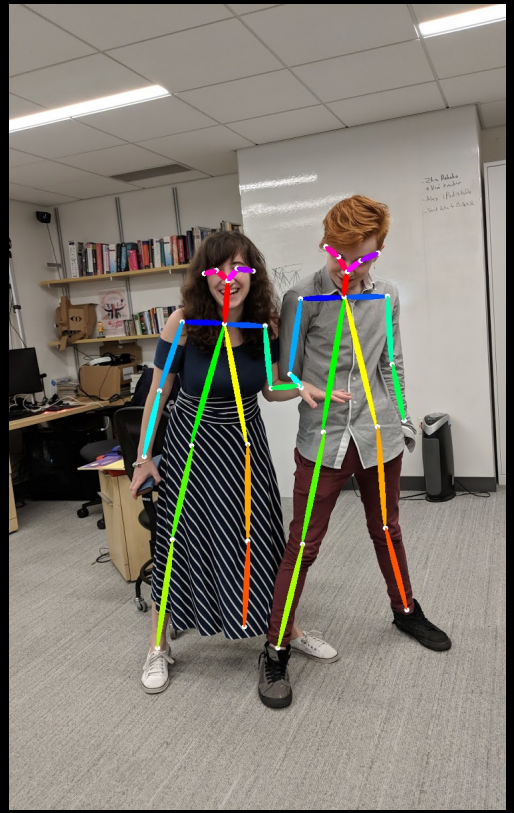
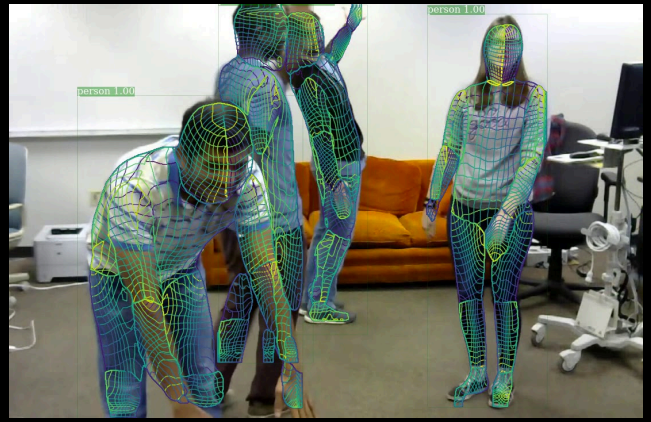
Through-Wall Human Sensing with RF Signals

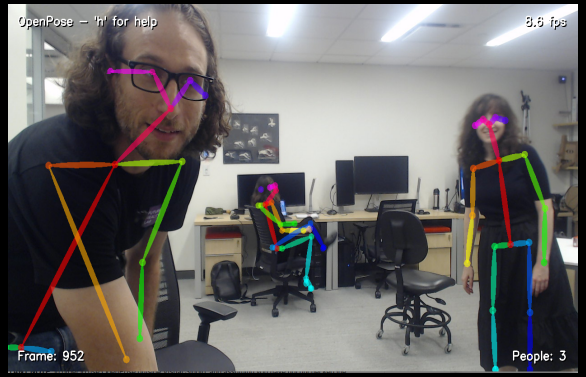
Mingmin Zhao

MIT CSAIL

Sensing Humans in the Environment

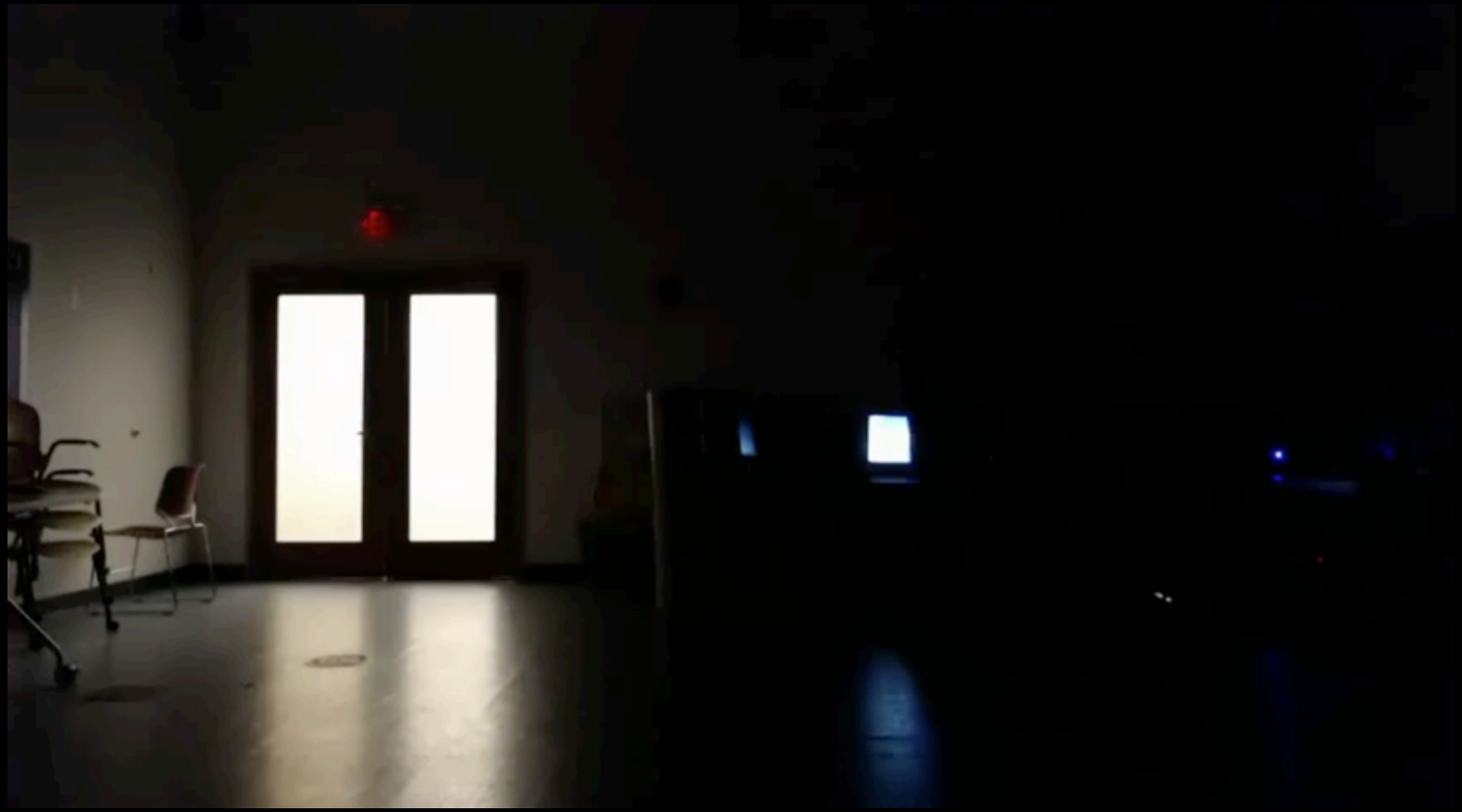




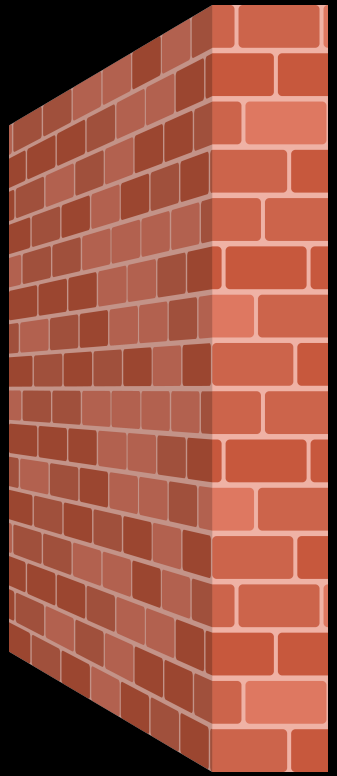


Occlusion is a fundamental challenge for vision

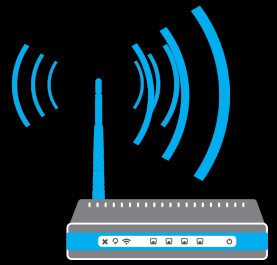
Vision also fails in bad lighting conditions

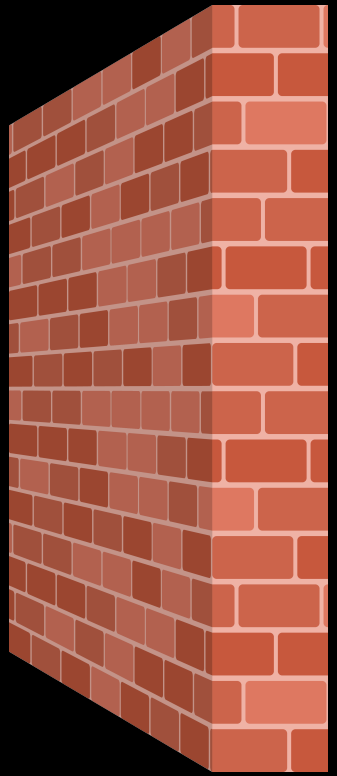


Want to see the human through walls & in the dark



Want to see the human through walls & in the dark





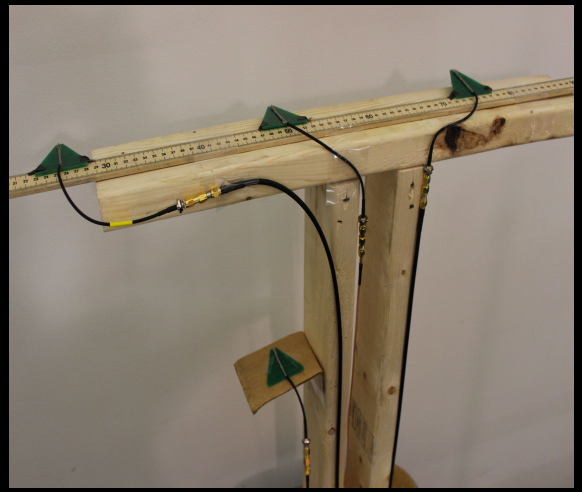
RF-based approach

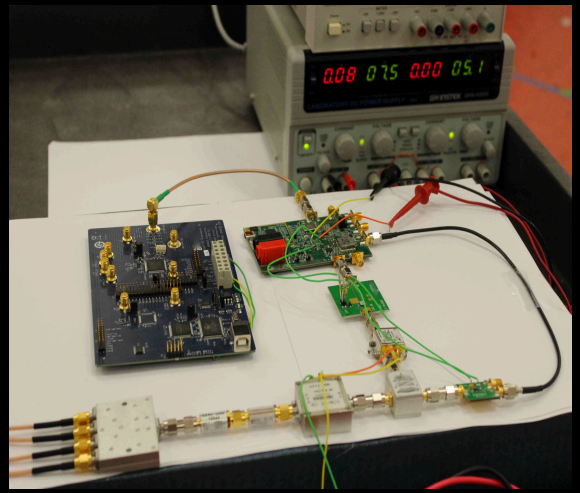
DARPA see-through-wall (mid 2000)

MIT Lincoln Lab (2011)

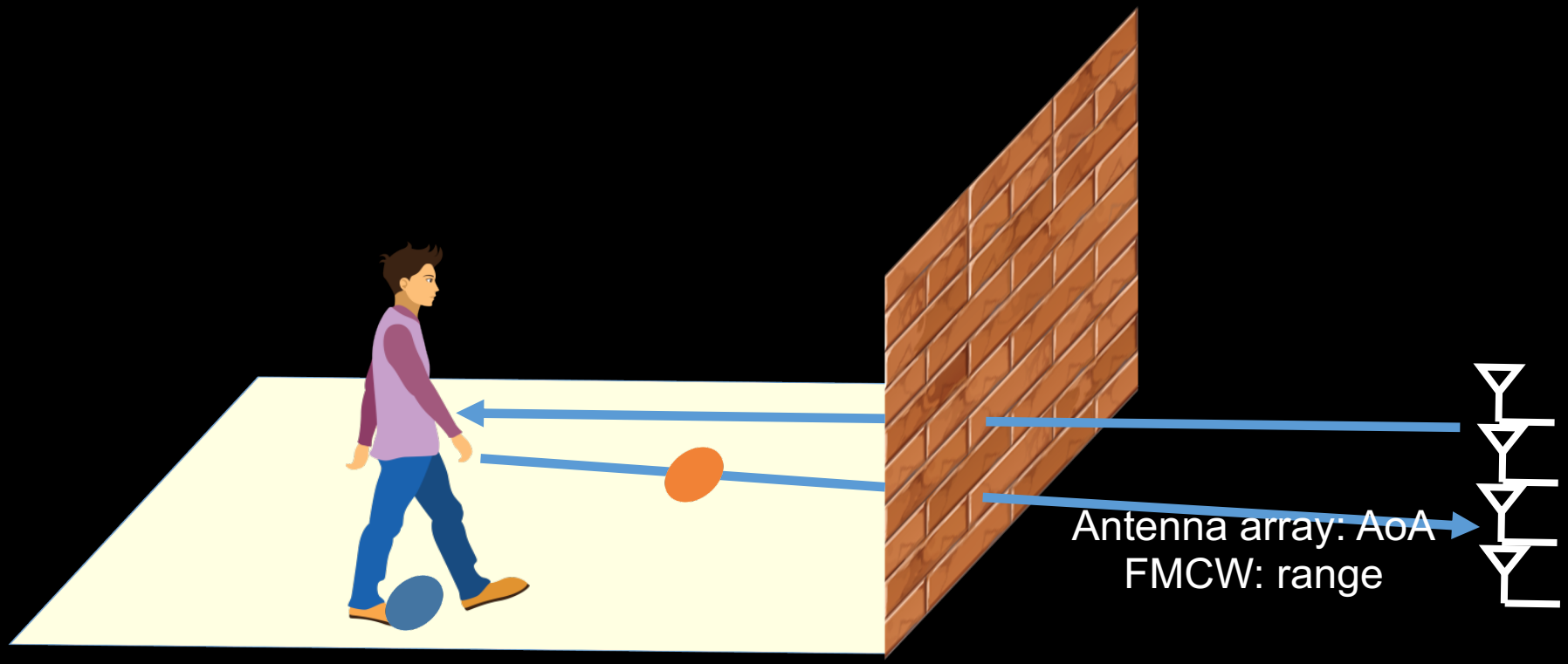
RF-based Approach

Wi-Vi and WiTrack from MIT
(2013)





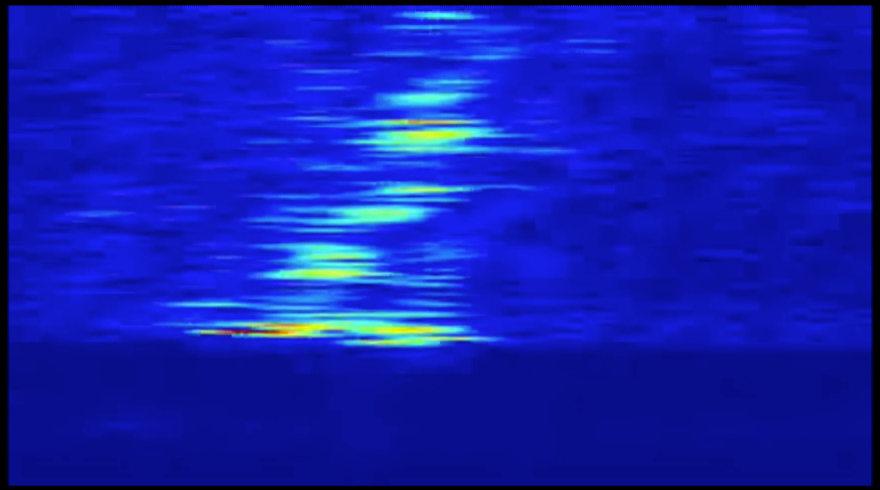
RF-based Approach



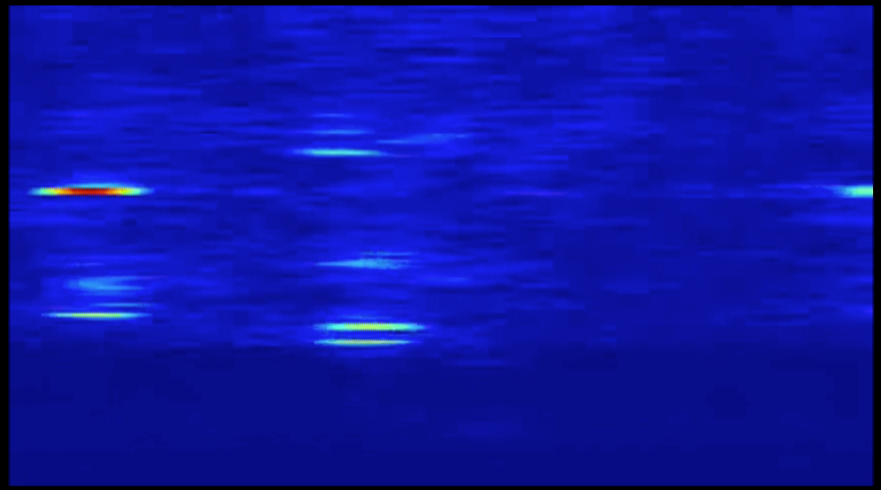
Outline

1. Challenges in RF signals for human sensing
2. Through-Wall human pose estimation (2D & 3D)
3. A framework for human sensing with RF signals
    a. Human Mesh Recovery
    b. Human Action Recognition
    c. Human Identification

How to train a model to estimate pose from RF?



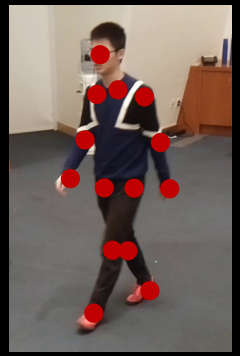
Vertical RF heatmaps

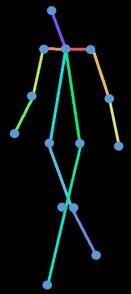


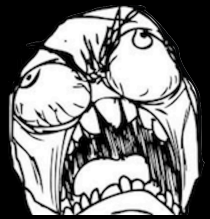
Horizontal RF heatmaps

Challenge: How to obtain labeled data?

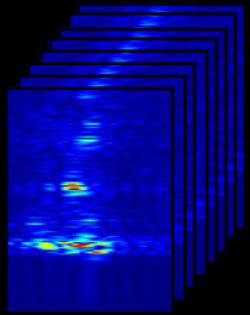
annotate skeleton?







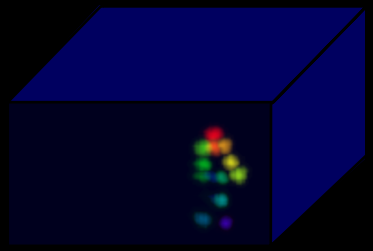
annotate skeleton?



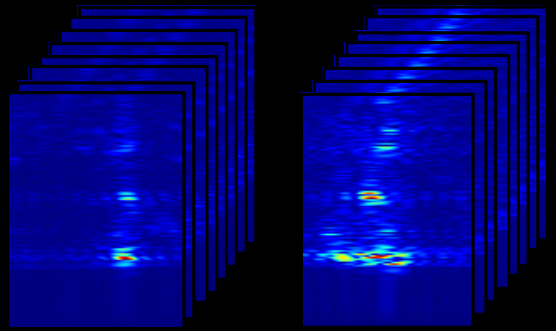
Idea: Cross-Modal Supervision

RGB images

Teacher Network
(Vision-based)

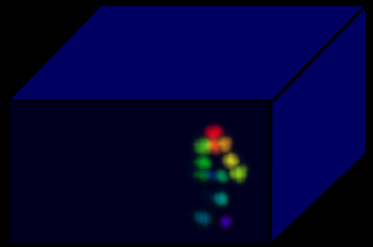


supervisory
signals

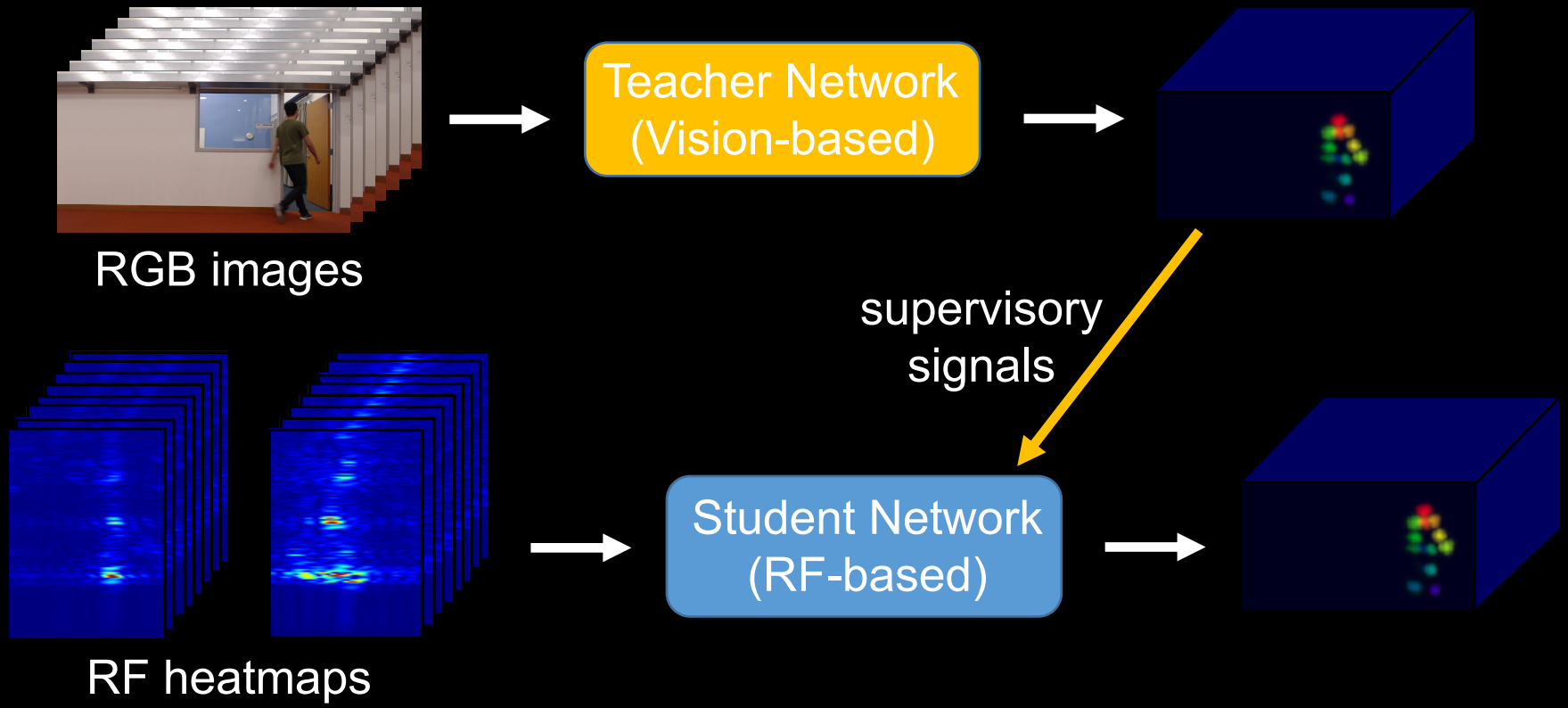


RF heatmaps

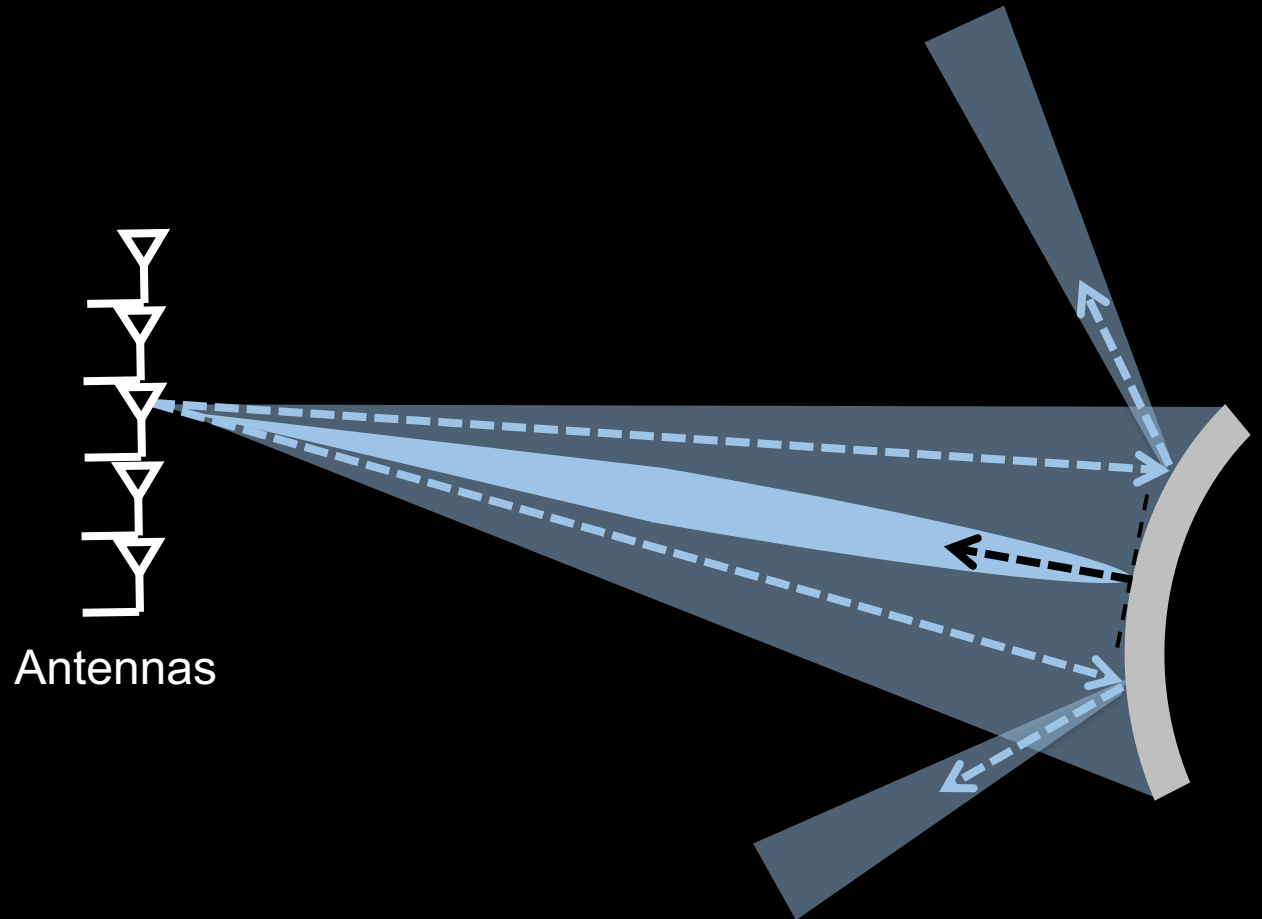
Student Network
(RF-Based)



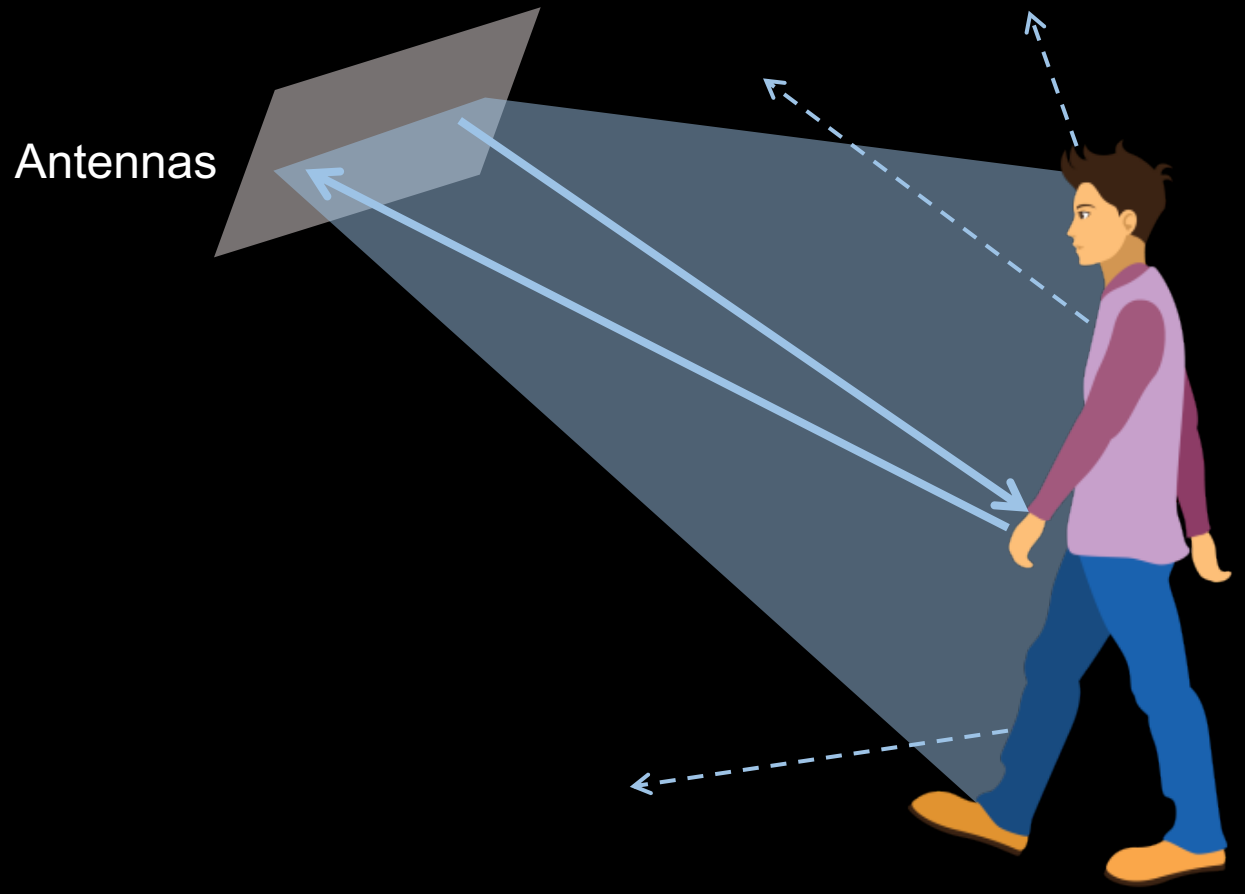
During inference



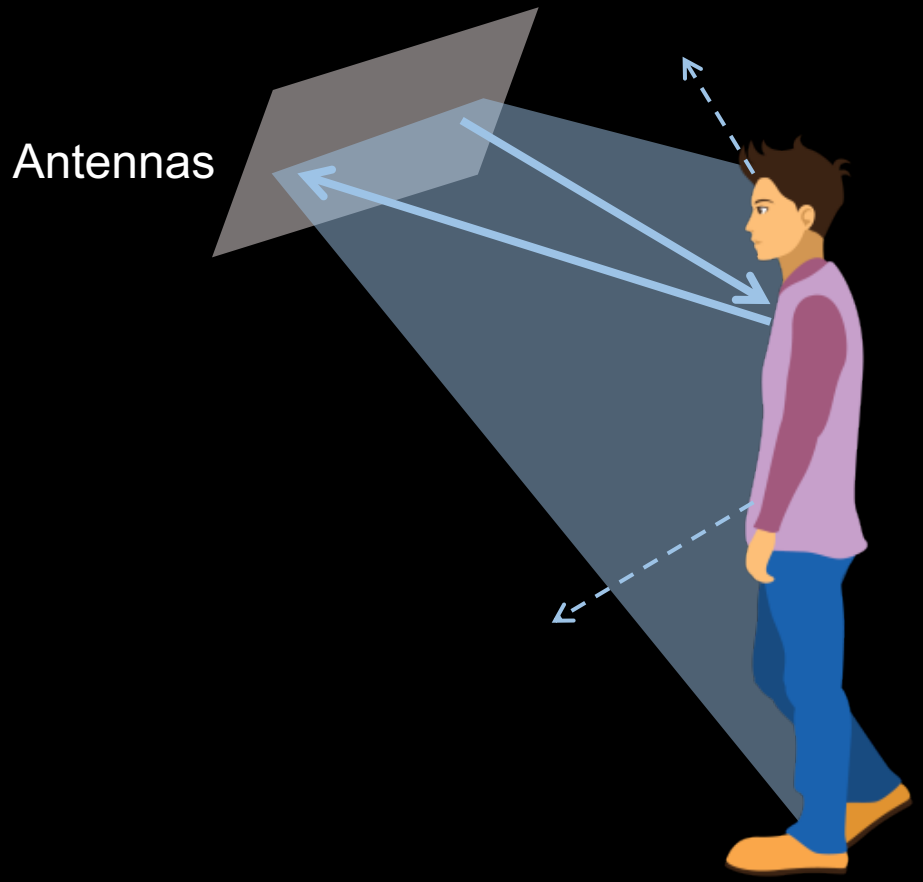
Challenge: Specularity of Human Body



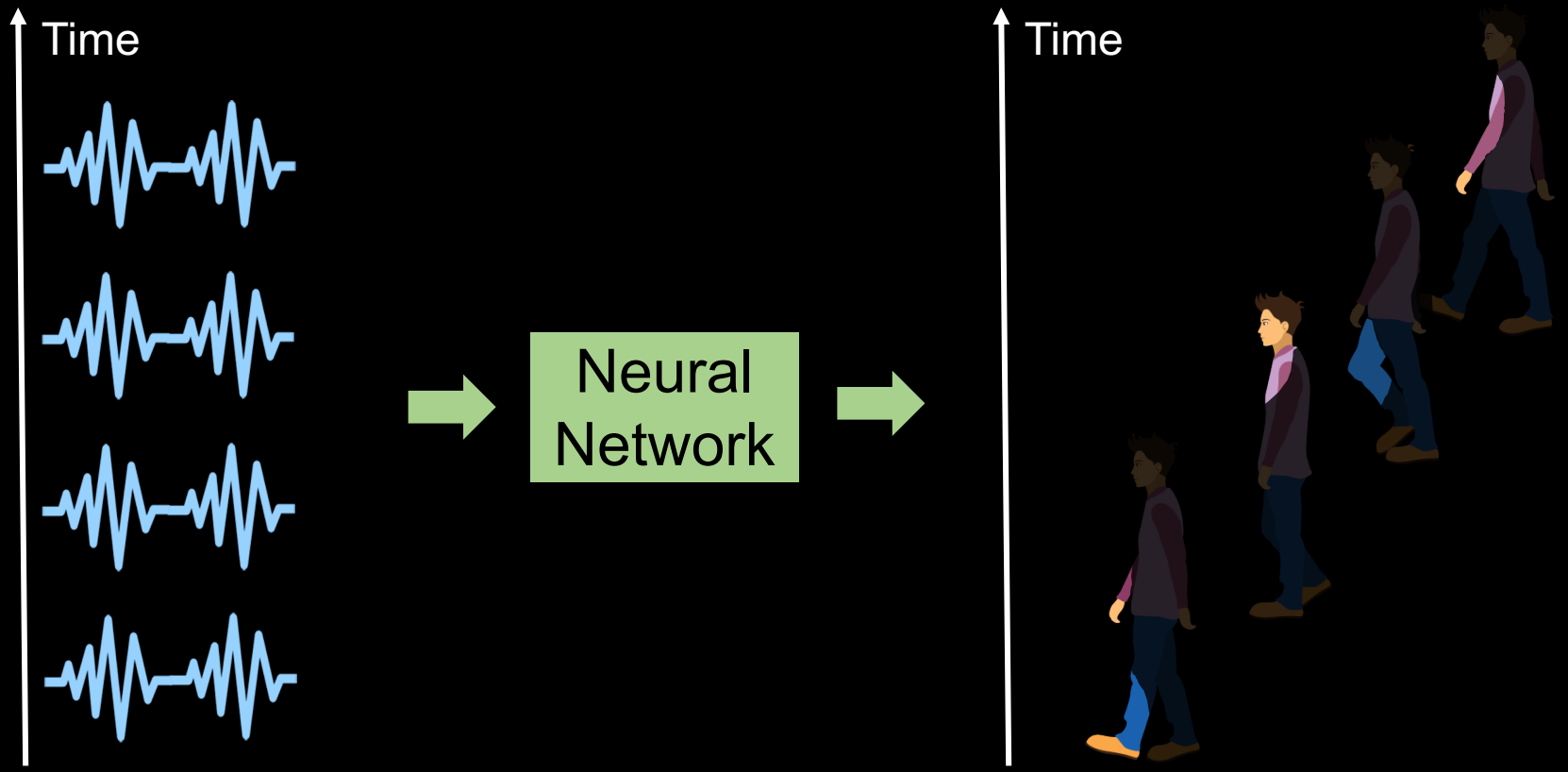
A Snapshot Doesn’t Have Skeleton



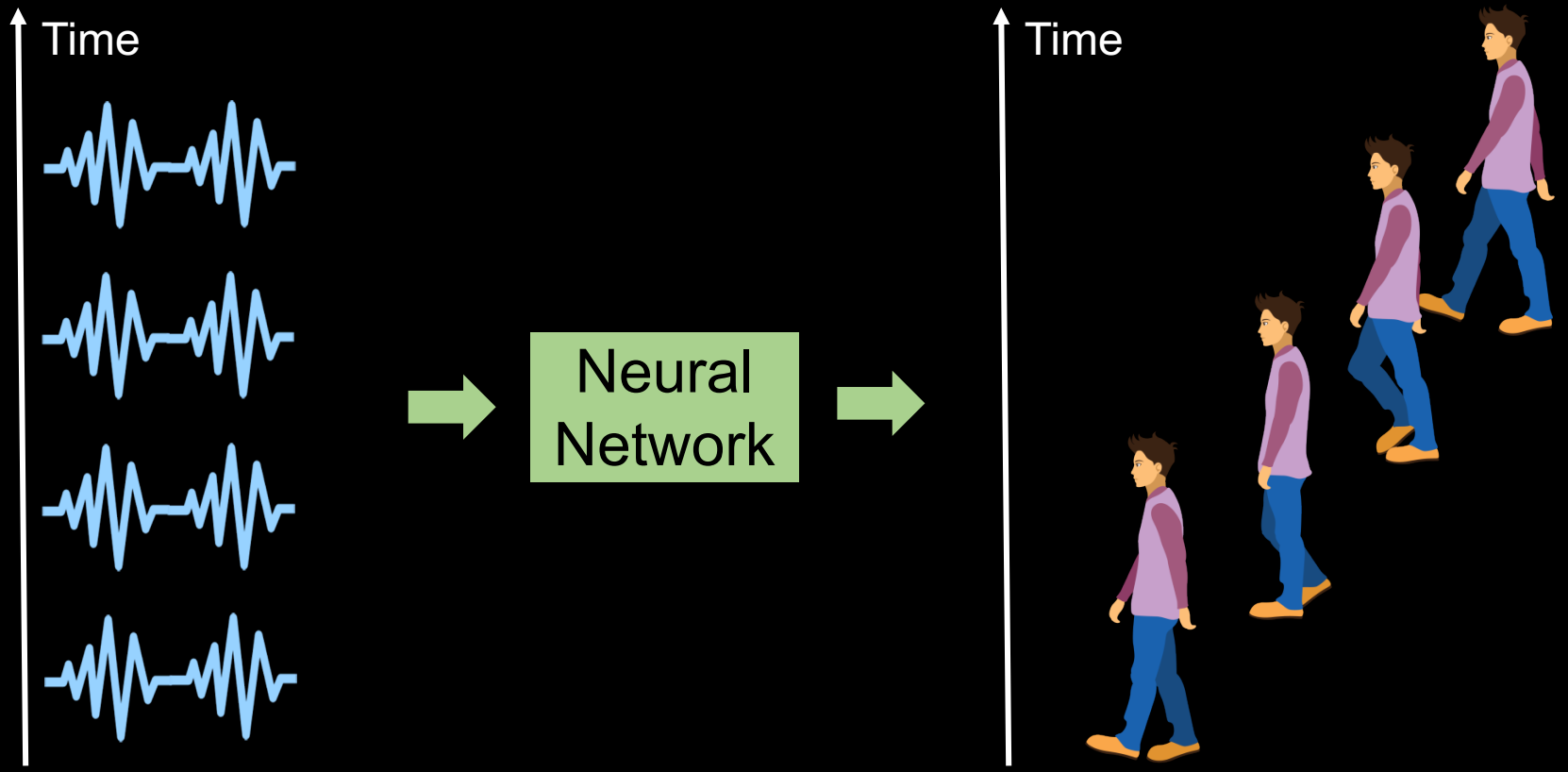
Solution: Use Human Motion Across Time



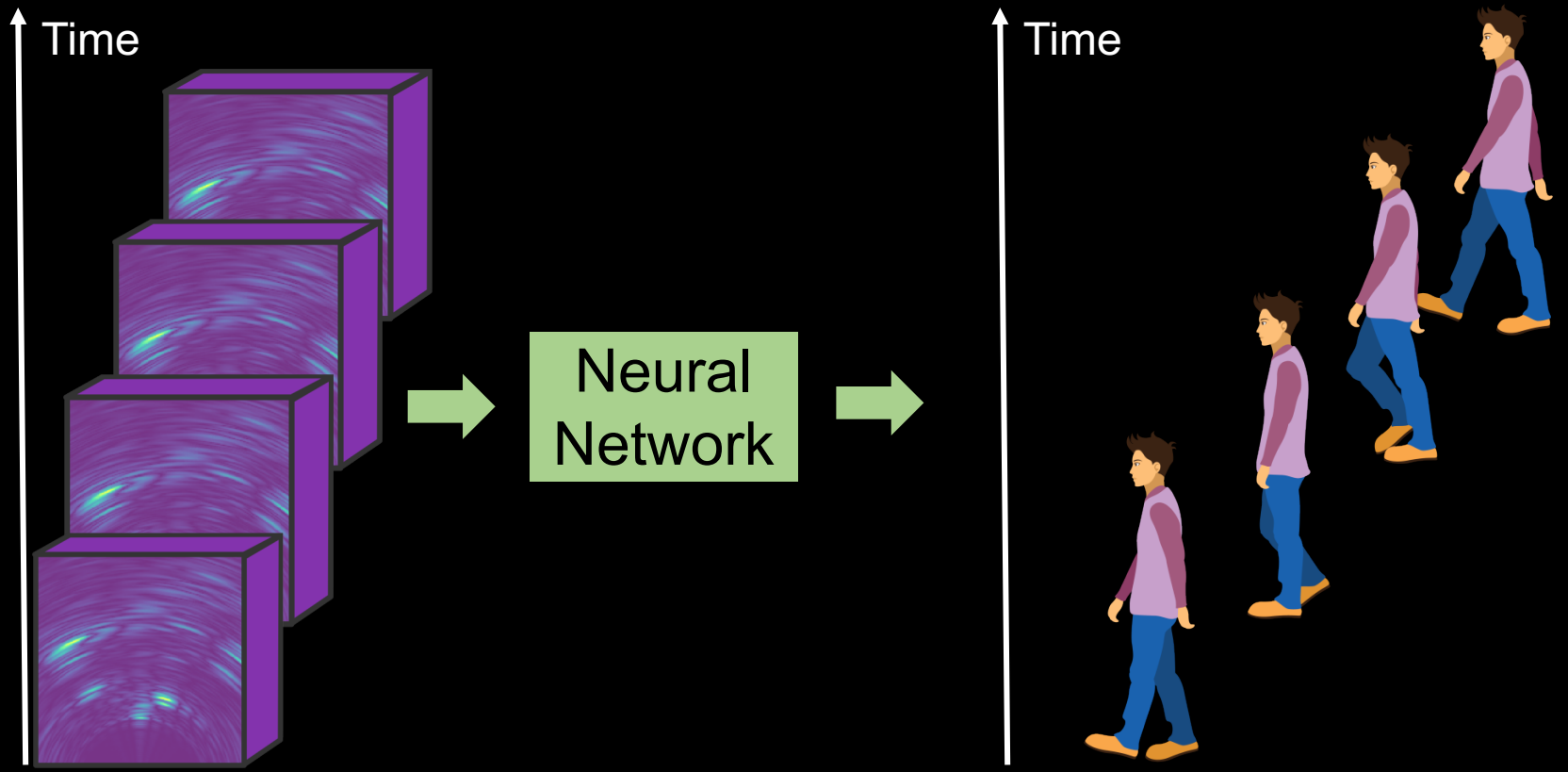
Solution: Use a series of RF snapshots



Solution: Extracts limbs and fills in missing parts

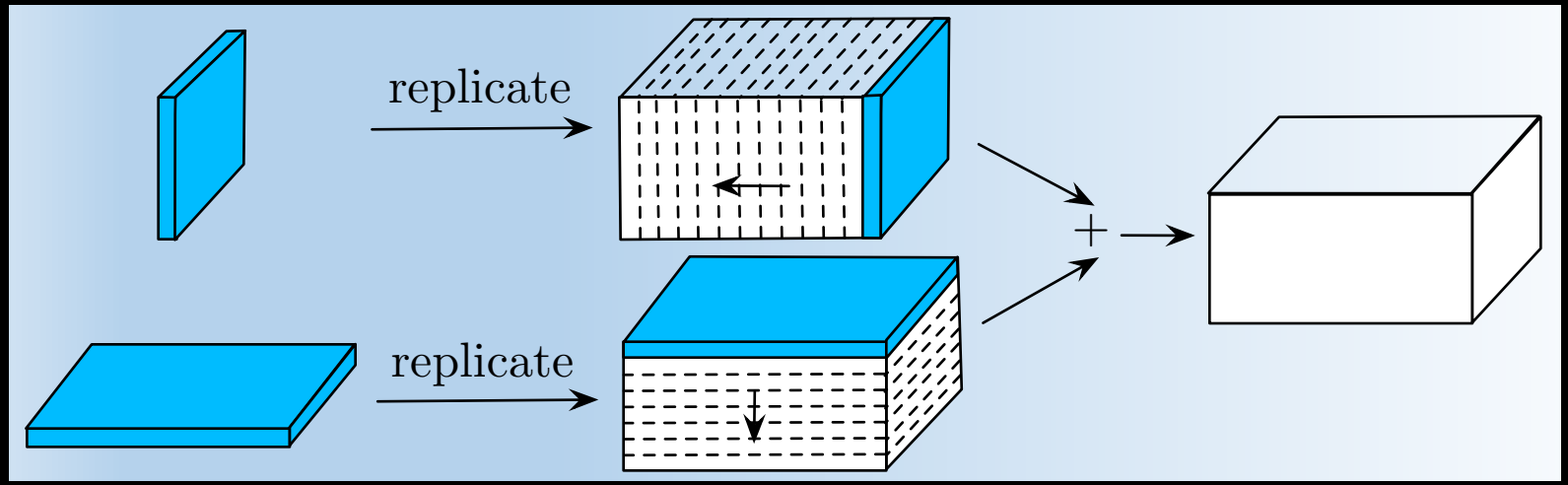


Challenge: 4D signals are too large for NN!

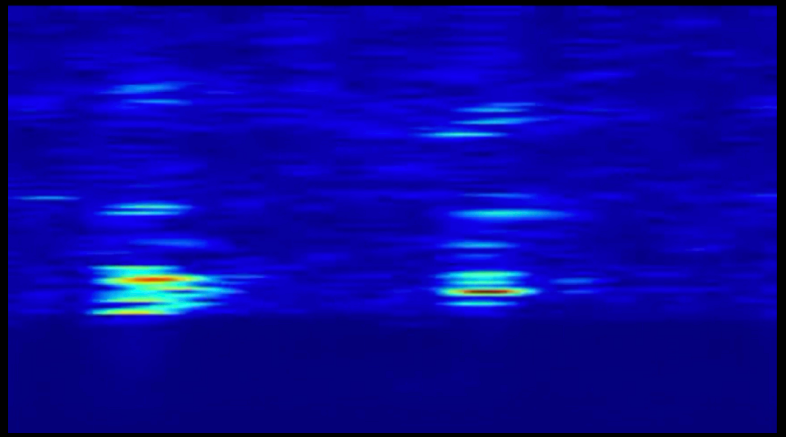


Solution: Neural Network Decomposition

Theorem (informal): An RF-based 4D Neural Networks is equivalent to a combination of two 3D Neural Networks

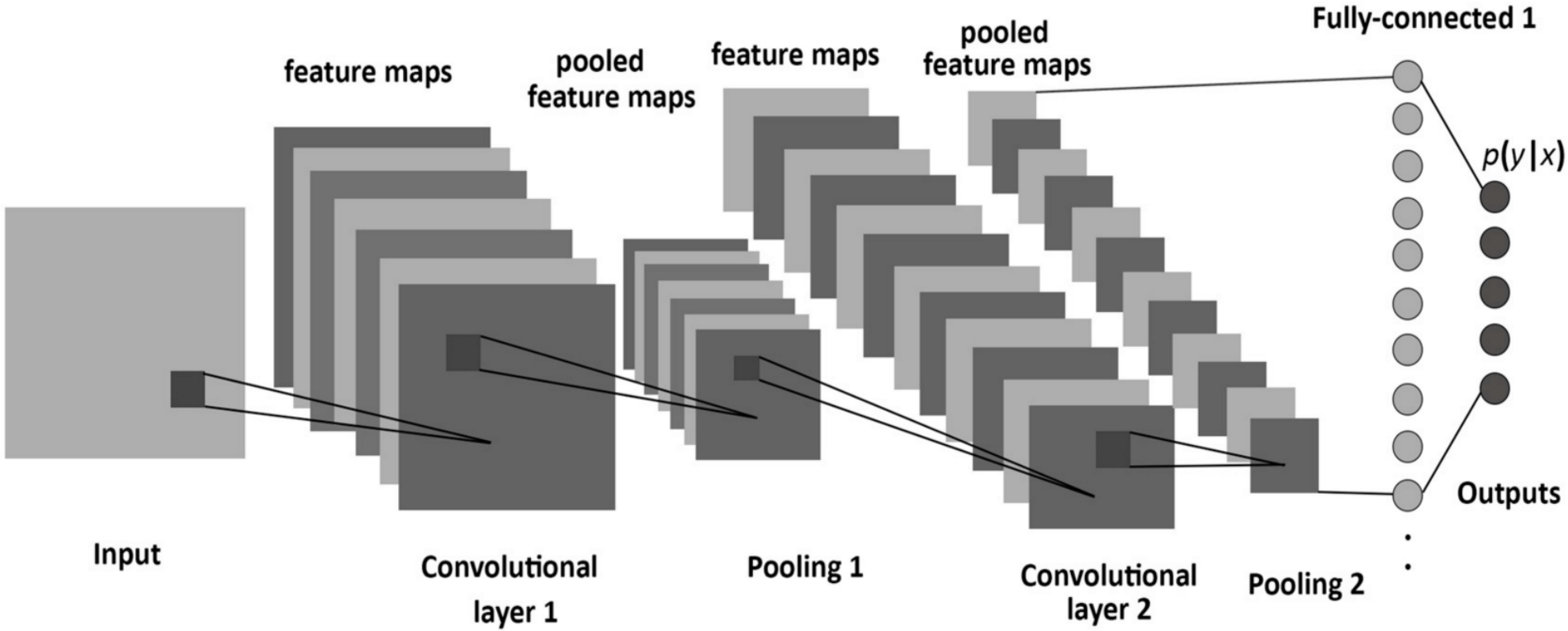


How about Multipath?

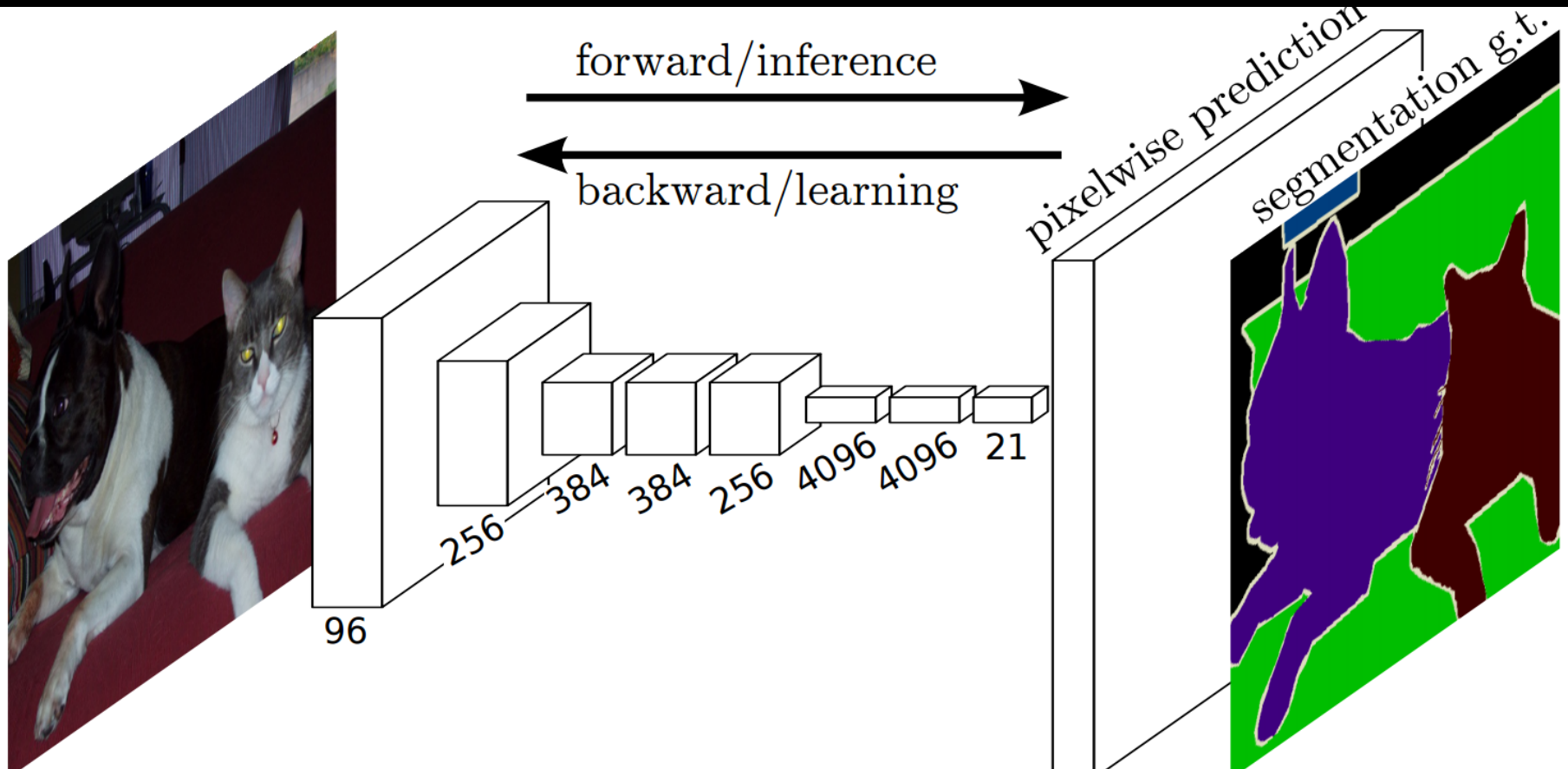


RF-Pose

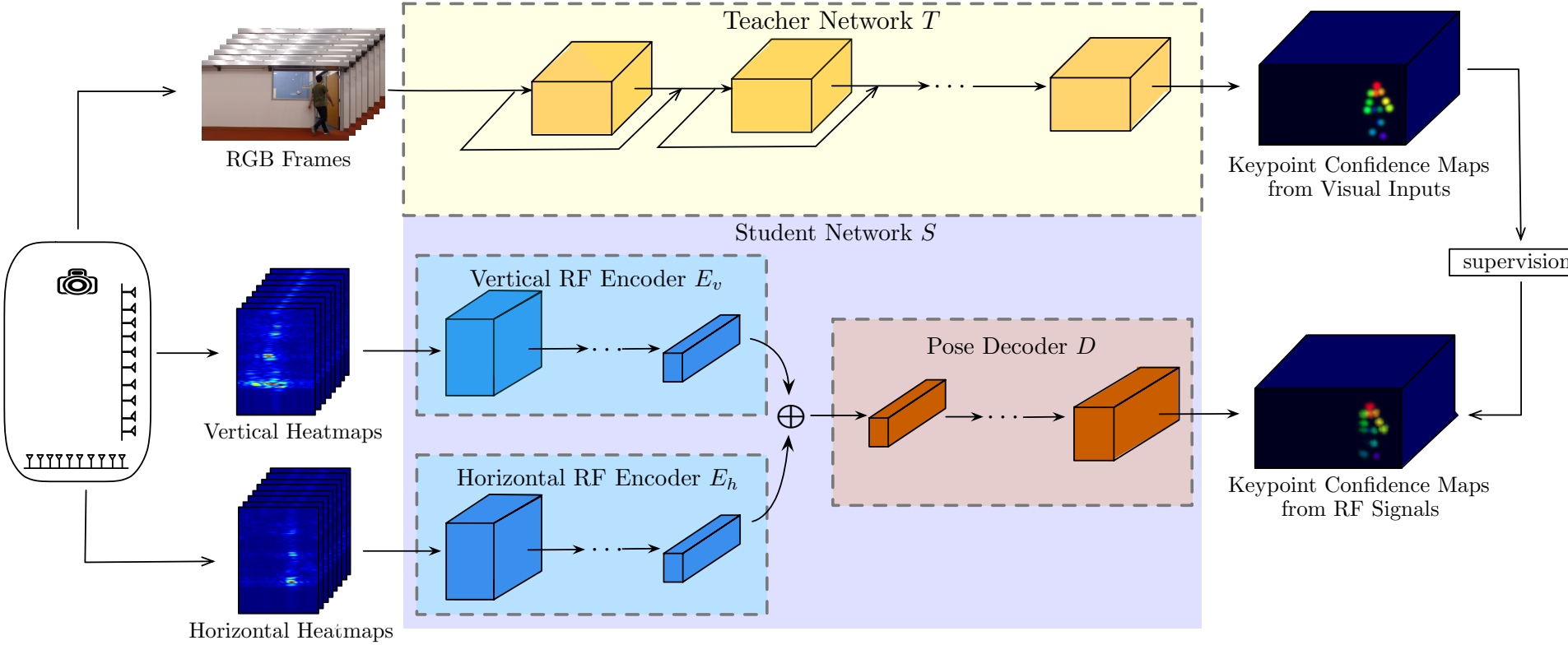
Convolution Neural Network (CNN):



Fully Convolution Network (FCN):



Model Architecture of RF-Pose:

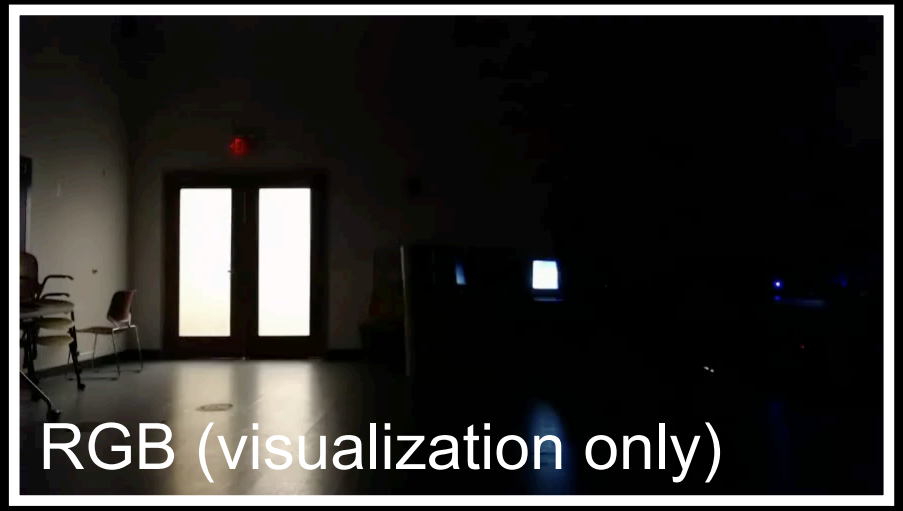


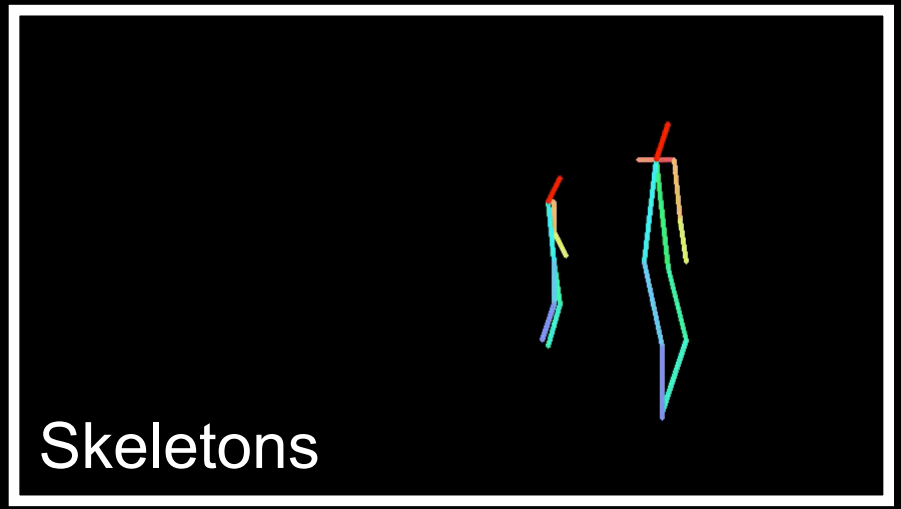
Through-wall poses using **only** RF

Skeletons

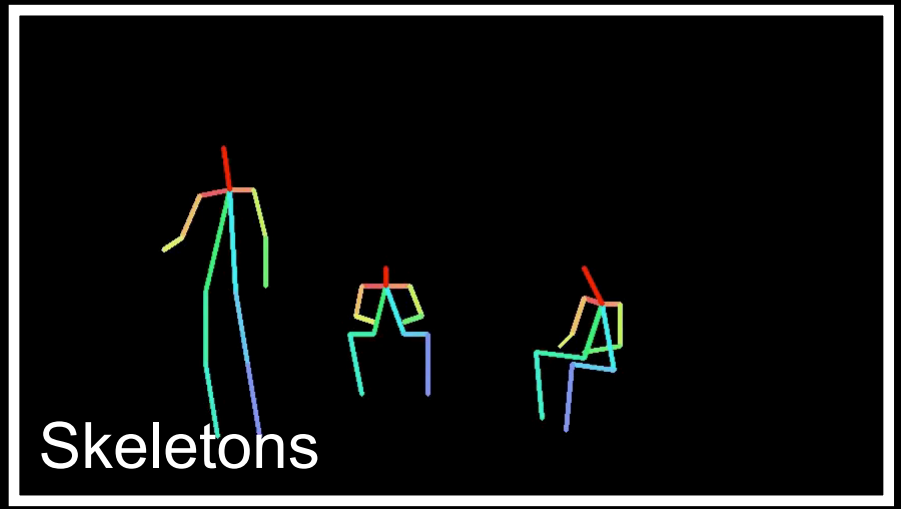
Confidence Maps

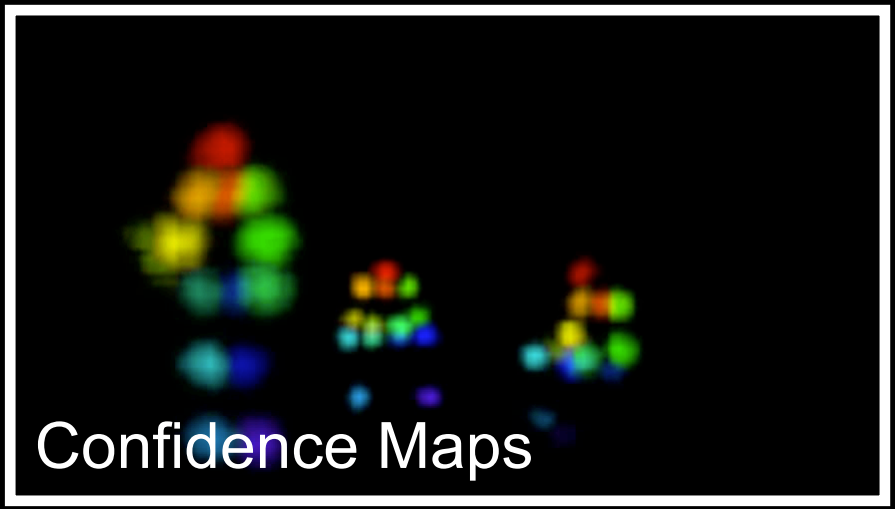
RF-Pose also works in bad lighting





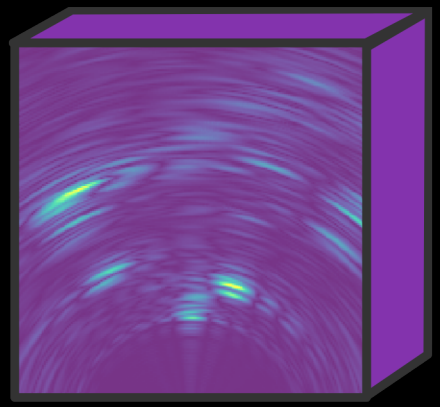
RF-Pose works with different environment and daily activities





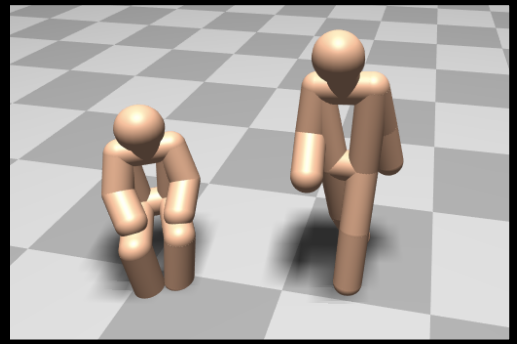
RF-Pose3D

Model Design: Complexity



3D RF tensor

Neural Network

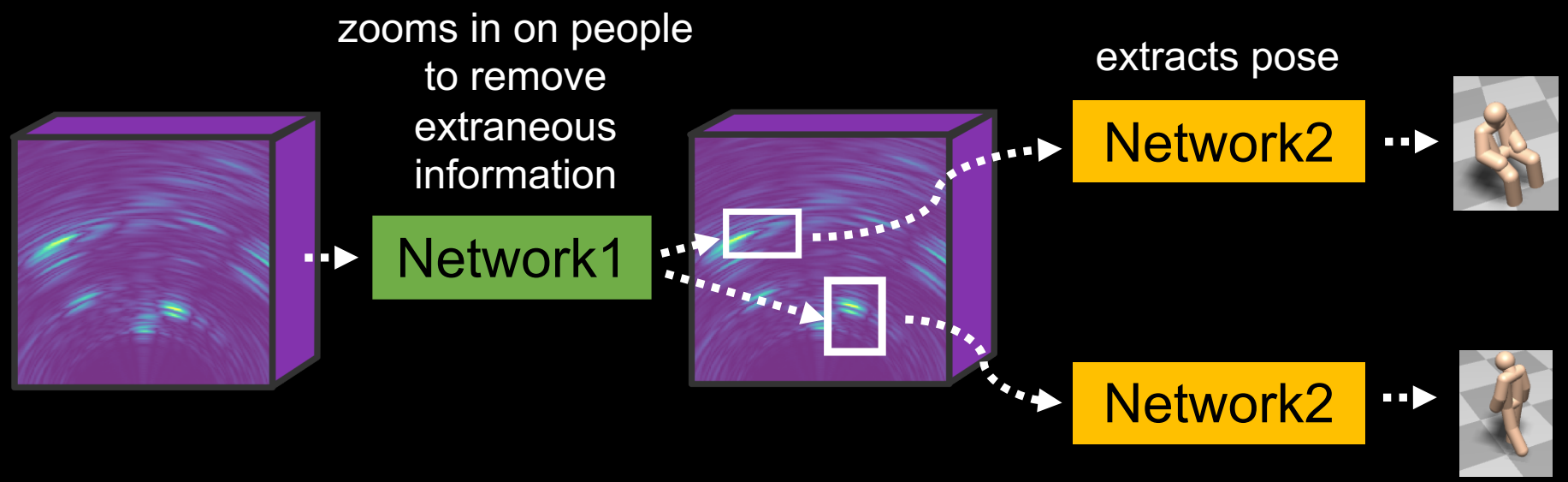


Model Design: Complexity

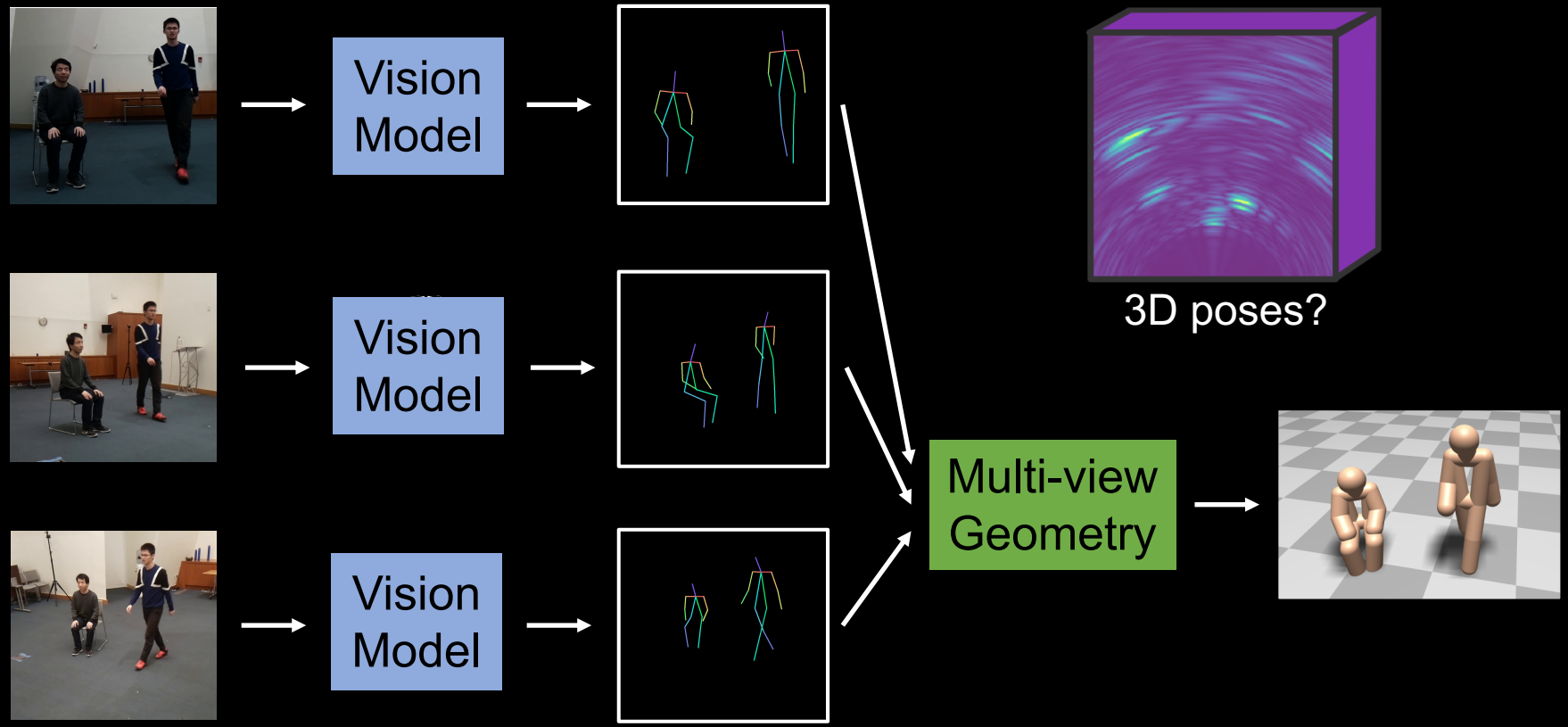
Implications:

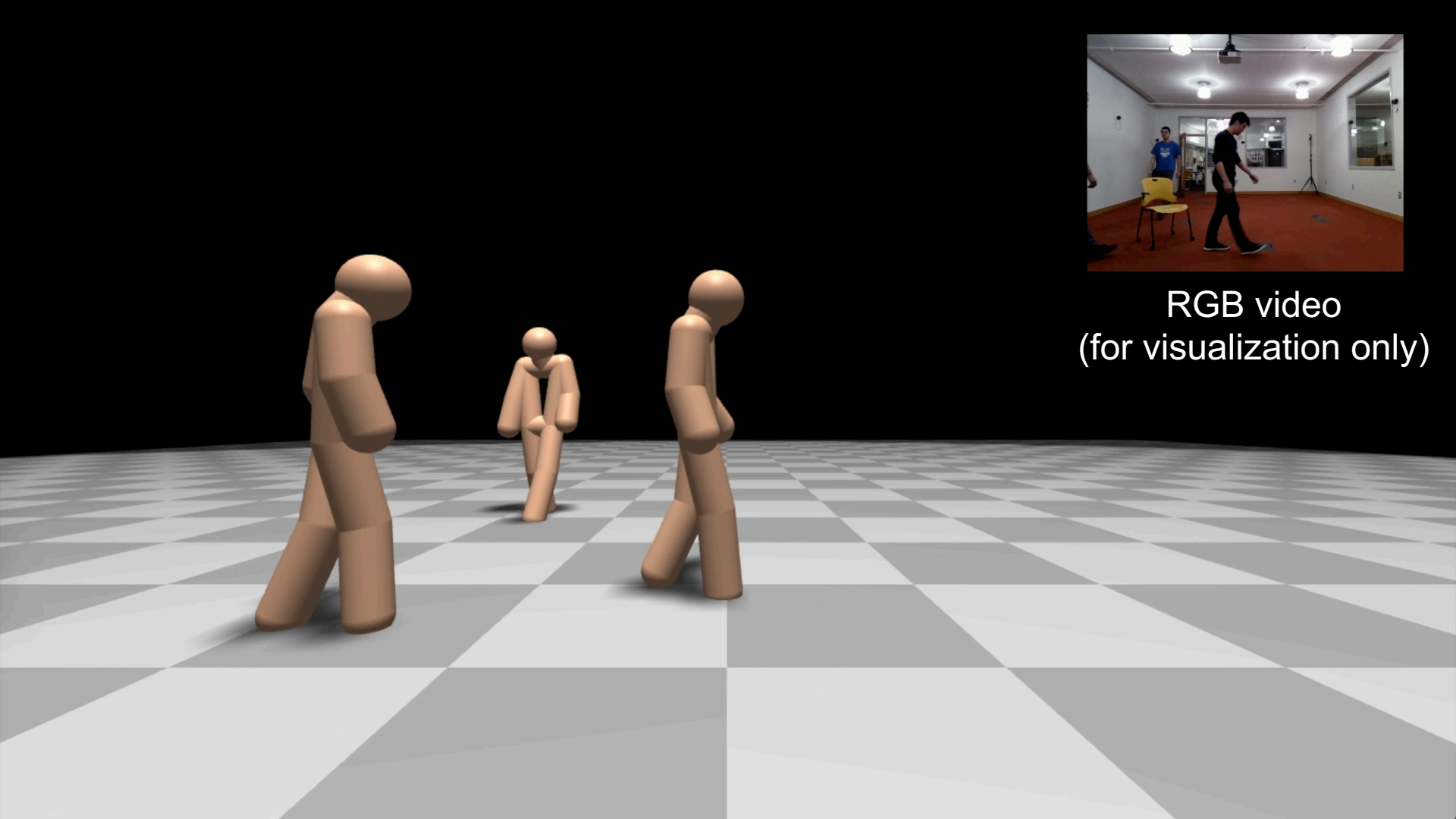
- Large neural networks
- Huge amount of labeled examples
- Very hard to train

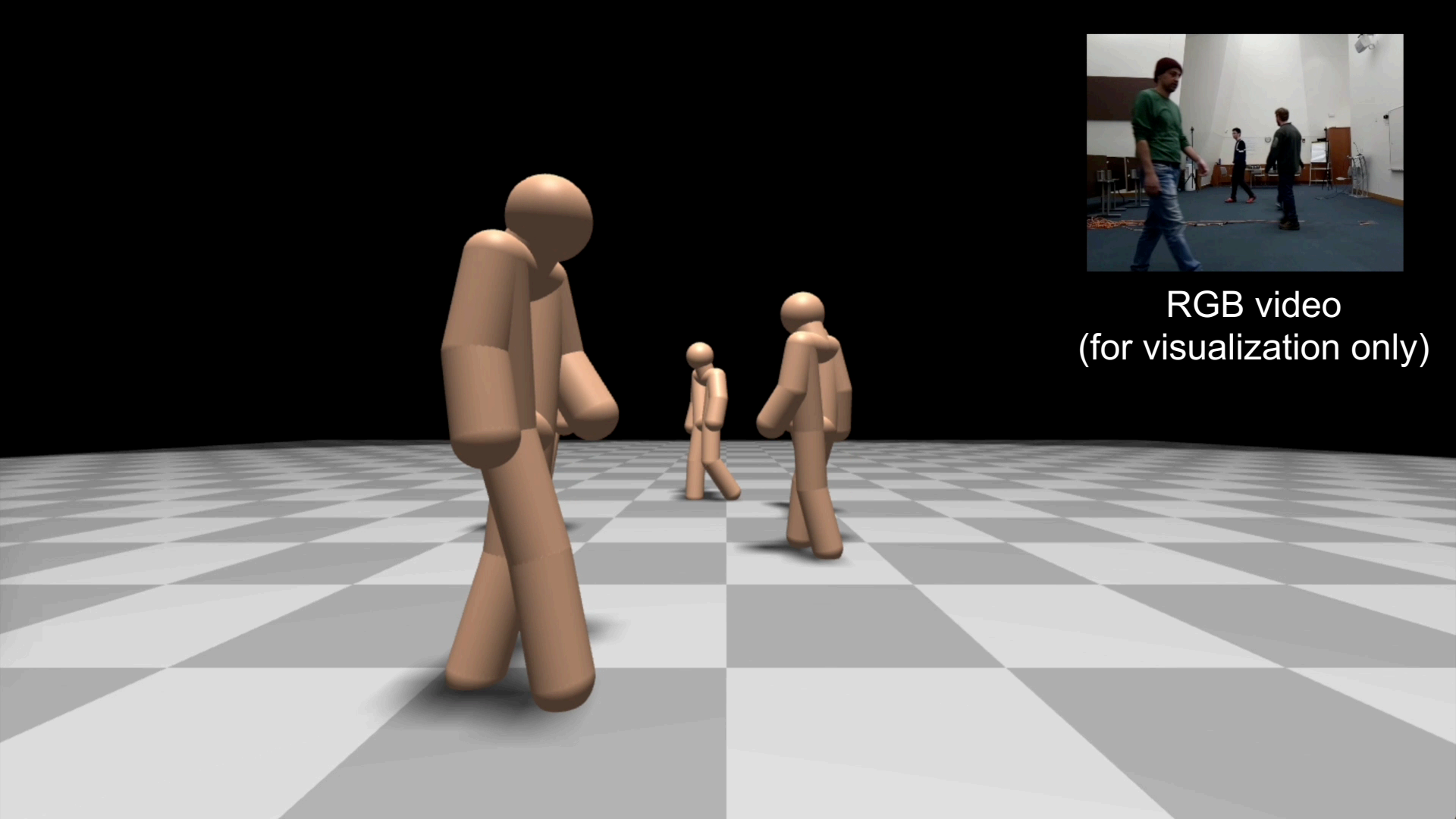
Solution: Two-Stage Model for Task Separation



Automatic labeling of 3D poses with cameras

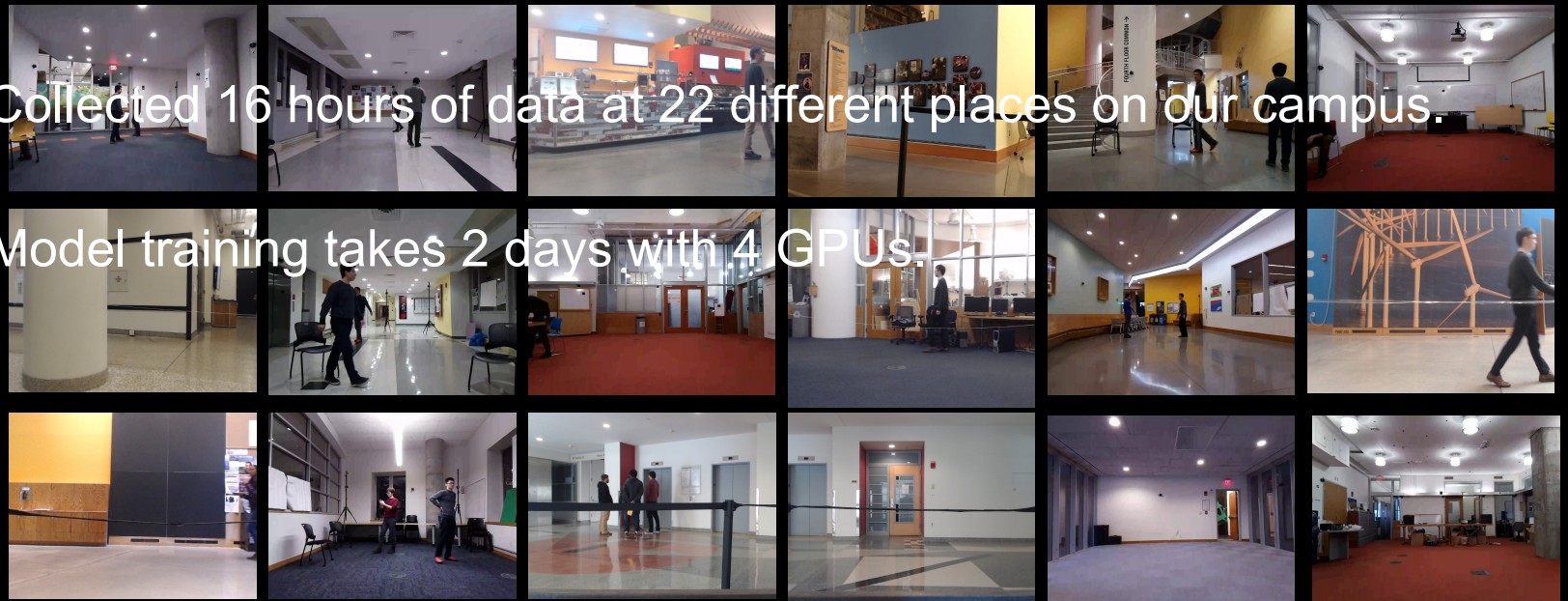



RGB video
(for visualization only)


RGB video
(for visualization only)

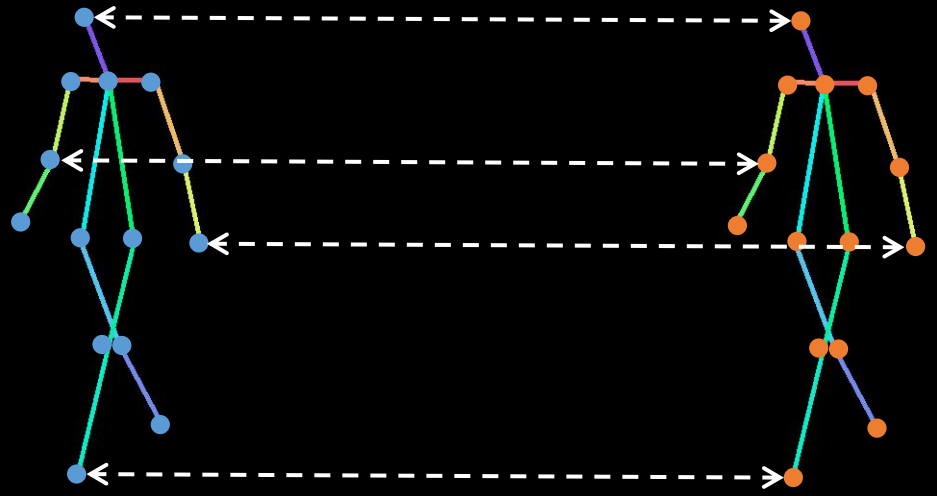
Implementation

- HW is similar to past work; uses an FMCW radio with antenna arrays
- Model implemented with decomposition in PyTorch.
- Collected 16 hours of data at 22 different places on our campus.
- Model training takes 2 days with 4 GPUs.



How Accurate is the Skeleton?

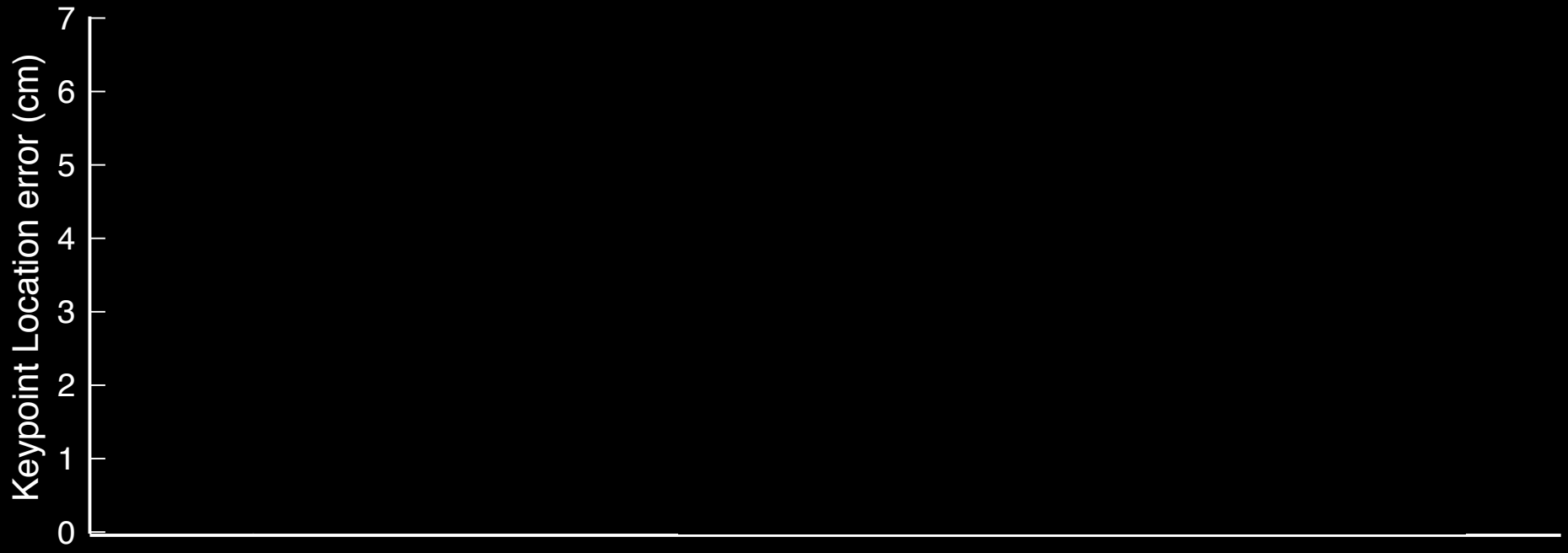
Metric: Keypoint localization error



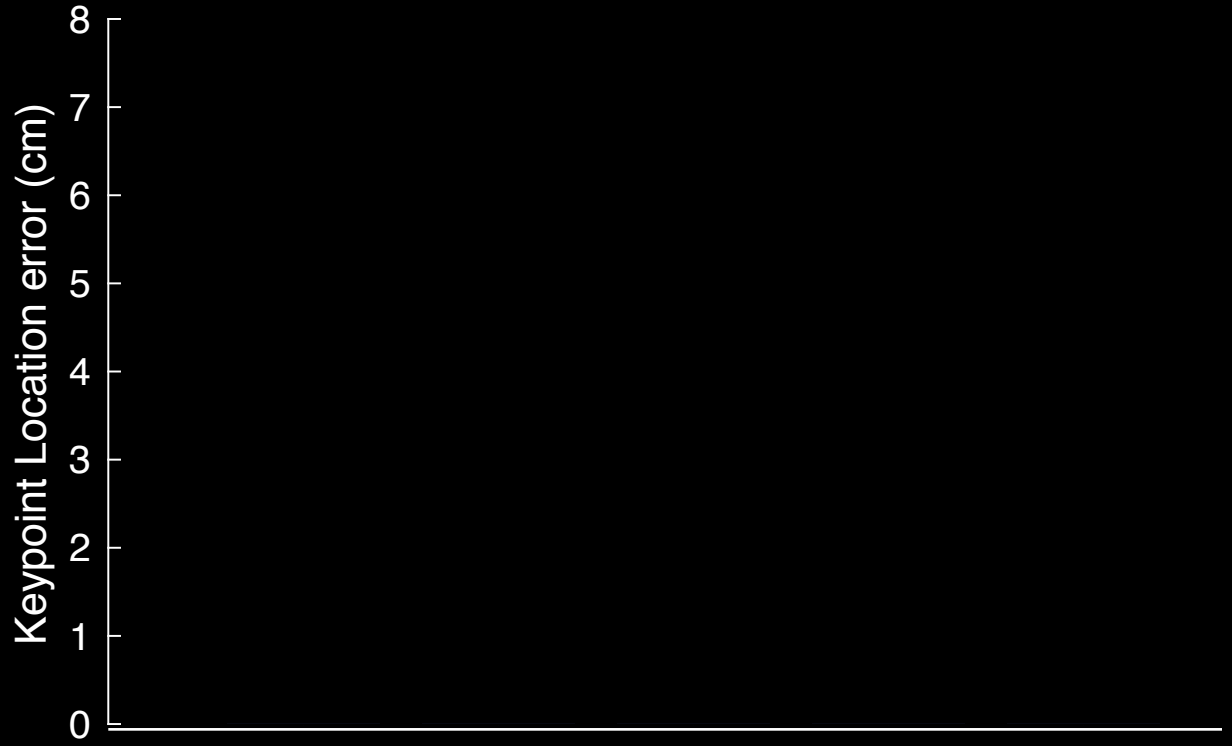
Predicted Skeleton

Ground Truth Skeleton

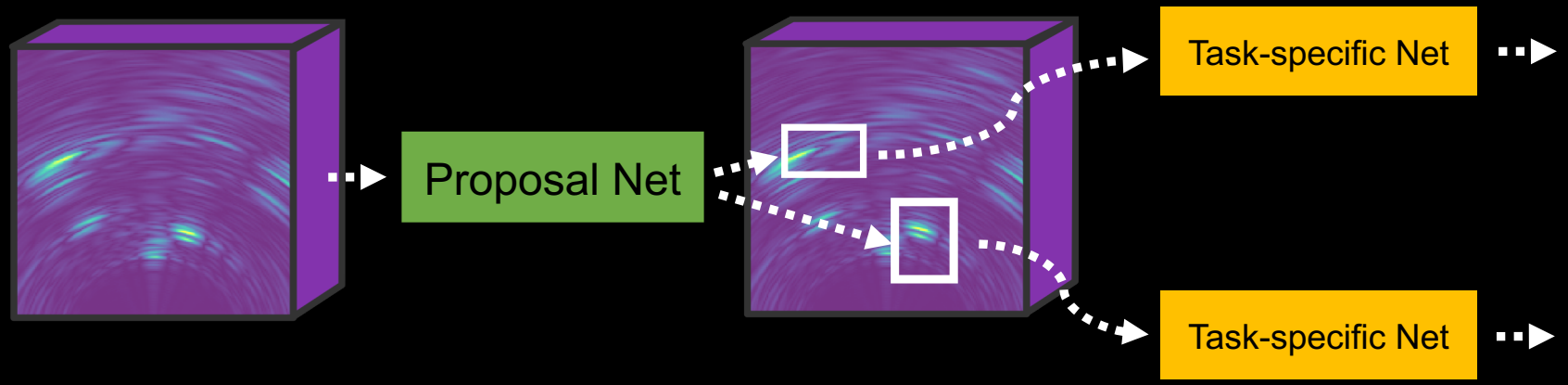
Skeleton Accuracy for Different Keypoints



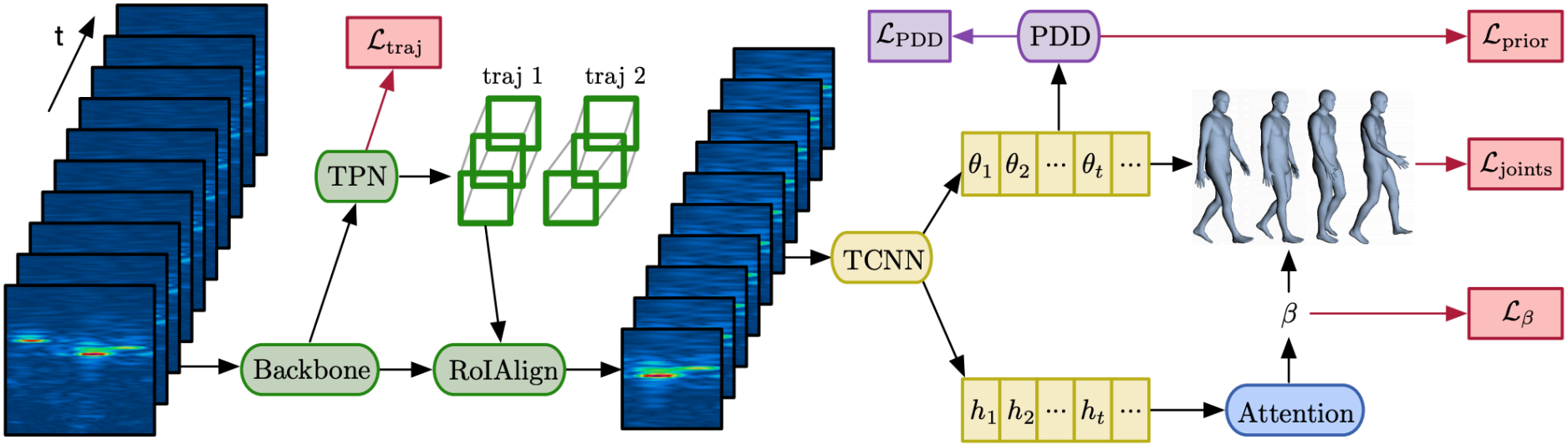
Skeleton Accuracy vs. Number of People

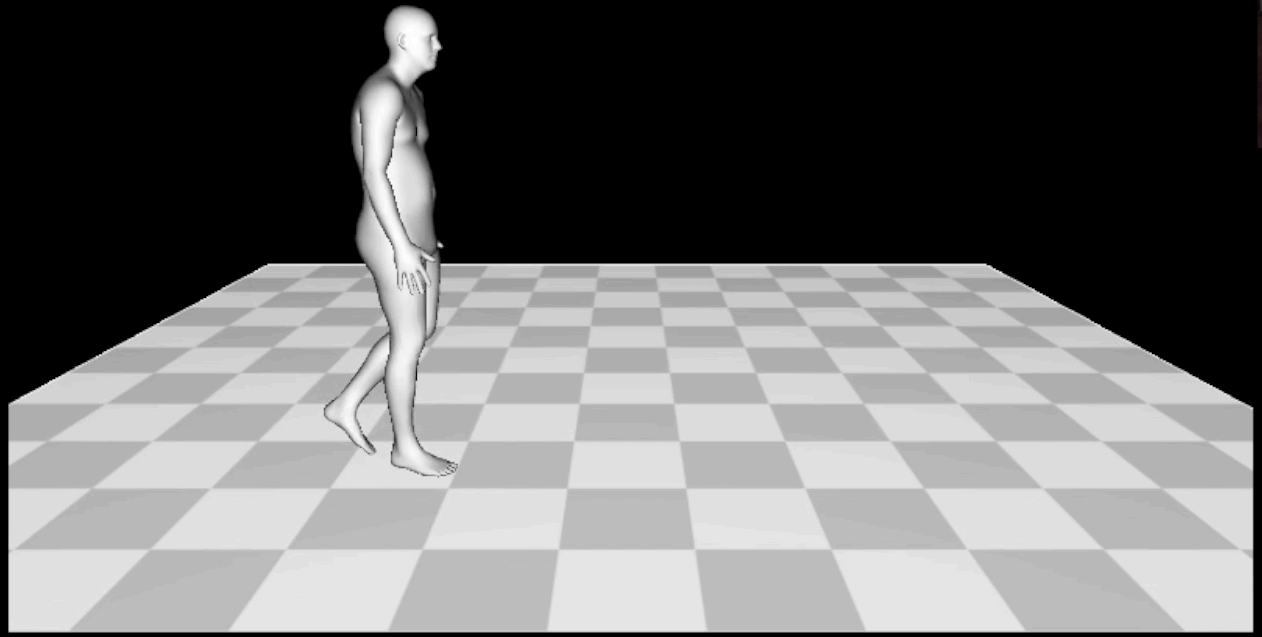


A Top-down Framework for Human Sensing with RF



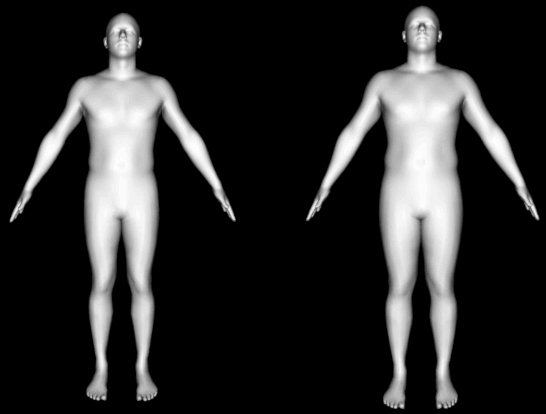
Human Mesh Recovery





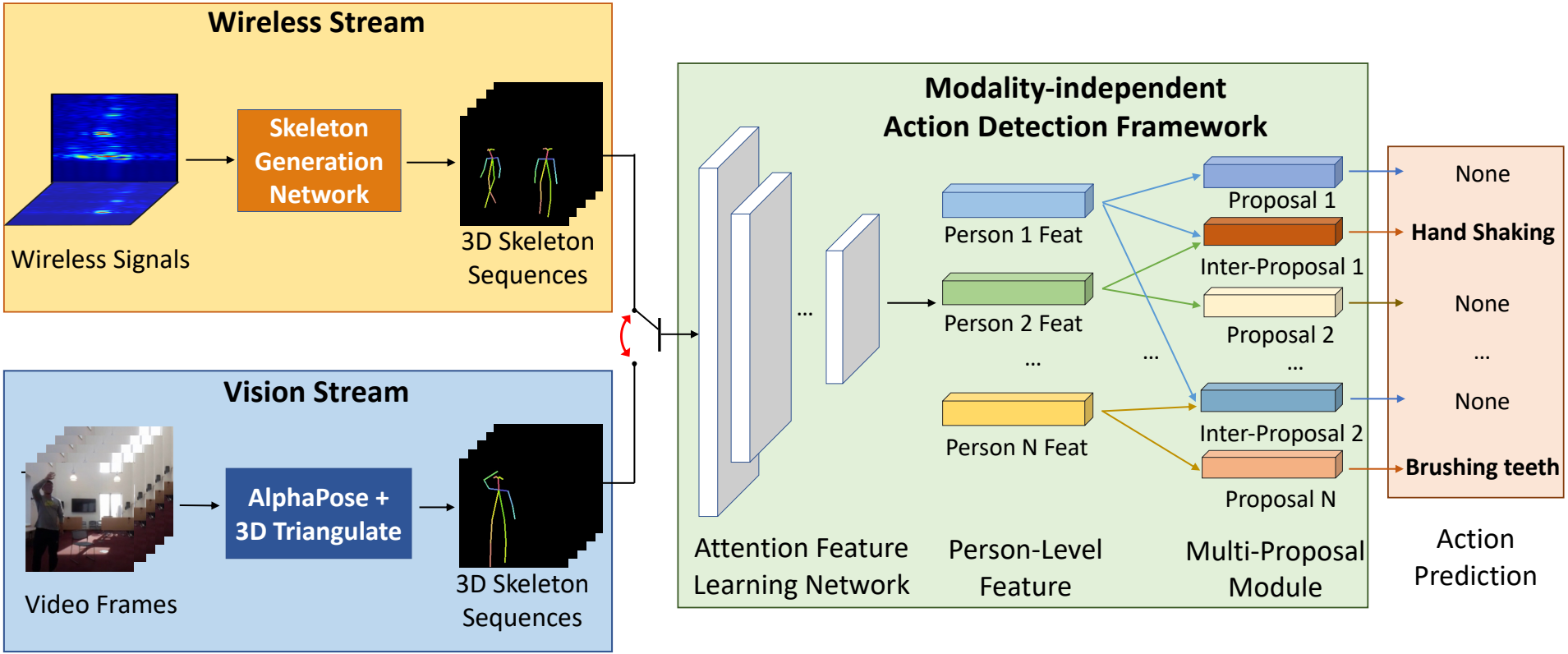
Dynamic human meshes from RF

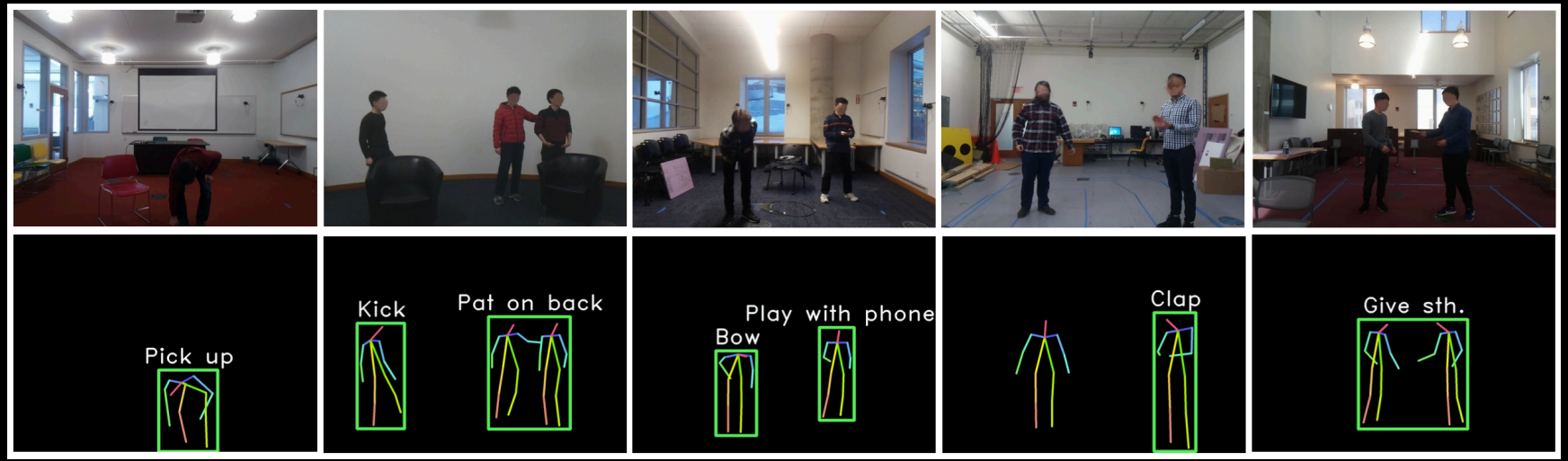
RGB view

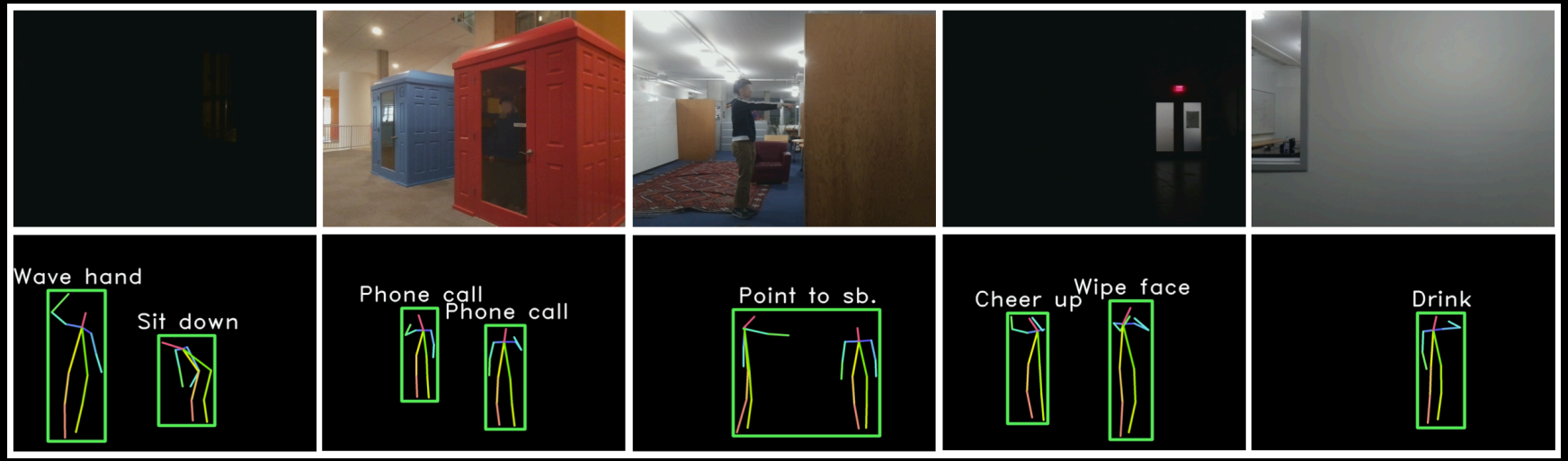


Body shapes from RF

Human Action Recognition




Pick up
Kick
Pat on back
Bow
Play with phone
Clap
Give sth.


Wave hand
Sit down
Phone call
Phone call
Point to sb.
Cheer up
Wipe face
Drink

Through-Wall Human Sensing with RF Signals

For more information, please visit our websites:

RF-Pose: **RFPose.csail.mit.edu**

RF-Pose3D: **RFPose3D.csail.mit.edu**

RF-Avatar: **RFAvatar.csail.mit.edu**

RF-Action: **RF-Action.csail.mit.edu**

MIT CSAIL