Lecture 22: Machine Learning for Tracking and Imaging

RF-Pose

<u>How can we track human movements?</u>

- There are two types of detection systems:
  - <u>Keypoint detection</u>: detect and track specific points on human body and follow these points' movements.
  - <u>Mesh detection</u>: detect changes in movement by outlining a more exact contour of the human body.
- Occlusion is a challenge for vision-based systems (i.e. cameras)
  - When there are objects in the way or when it dark, it's difficult for vision-based systems to accurately track movement.
  - This is not a problem with RF because RF can move through objects and low-light settings while still reflecting off human bodies.

<u>How does RF sensing address human tracking despite occlusions?</u>

- There were many past systems which performed imaging behind walls, but they required complex set ups and only worked in limited settings.
- RF sensing works by tracking reflected RF signals off the human body.
  - WiTrack reminder: tracking done by combining distance (measuring time of flight) and angle (calculated using multiple antennas).
- Can use RF signal to create a heat map that identifies where humans are on a given 2D plane.
  - RF-Pose leverages RF to create two heat maps: one mapping horizontal angle vs. distance, and another mapping vertical angle vs. distance.
  - These two heat maps are then plotted to the cartesian plane.
    - Intuitively, the conversion is somewhat similar to converting polar coordinates to cartesian, for 3D space we go from R/theta/pi to X/Y/Z.

<u>What are some challenges in ML for RF?</u>

<u>Challenge 1</u>: It is difficult to label data just based off an RF signal

- Solution: Cross-Modal Supervisions
- Teacher network is pretrained and it provides the supervisory signals to the student network.
  - In RF-Pose, the teacher network has a vision-based input and software to label/track human poses, while the student network has an RF-based input. The student network's inputs are paired with the corresponding teacher network's labels/tracking.
- RF-Pose's network never trained in environment with walls or low-lighting, but the model was still able to generalize to these scenarios (performed well on out-of-distribution points; RF signal input should not be affected by these occlusions anyway).

Challenge 2: Specularity of human body (human body acts as mirror, i.e. reflector and scatterer), meaning only some of the reflections will go back to the device, making it difficult to track humans at a level more granular than just entire-body movement.

- Solution: track keypoint movement over time. As a person moves, the device collects different samples of different parts of the body, and then they are put together in the neural network.
- This results in a 4D input to the neural network (3D space + time), which is computationally expensive.
    - Solution: neural network decompression
        - Have the neural network use two 3D input streams (2D horizontal map + time) + (2D vertical map + time).

How does the model display these keypoints visually?

- Different keypoints (body parts) are identified by their reflections at certain positions relative to other keypoints. For example, the head is specified by being higher than all other keypoints on the vertical map, the hands are lower on the vertical map but also the two outermost keypoints on the horizontal map, etc.
- Displayed keypoints are each a unique color, and each keypoint is also displayed with a confidence map (density of color is confidence).

How can the 2D display made by RF-Pose be mapped to 3D?

- Key idea: first identify where people are in heat maps, then focus attention on those areas and use another neural network to extract the pose.
- Automatic labelling by using the Cross-Modal supervision model (i.e. feed label of video/picture of pose, feed it through the network with the RF signal, then focus on the pose using multi-view geometry).

HawkEye

What is HawkEye?

- Primary goal: Capture and examine RF point cloud of stationary cars and map that to vision-based ground truth (like cross-modal model)
    - Collected RF data by moving 1 antenna (emulated 2D antenna array)
    - Capturing the exact shape and position of the car is difficult due to specularity
- Solution: Since there is a prior (aka we already know shapes of cars) we can be specific and fit these RF point clouds to a car.
- Limitations: only works with single cars, must be stationary, uses SAR not real time.