People Counting from mmWave Radar Point Clouds with Graph Neural Networks

Author: Bernadett Bakos b.a.bakos-1@student.tudelft.nl
Supervisors: Girish Vaidya, Marco Zuñiga Zamalloa

1. Background

- Cameras are a widely used approach for crowd sensing. They however come at the cost of invading people's privacy. To overcome this issue, mmWave (short-wavelength electromagnetic waves) sensors can be used as replacement for cameras, as they only collect non-identifiable data.
- *New challenge*: mmWave radars record points in space, how to tell the number of people sensed by the radar?

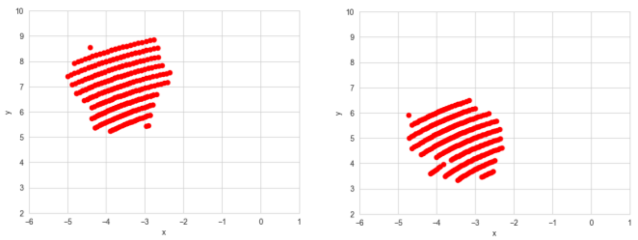

Figure 1: 5 frames of 3 people (left) and 4 people (right) walking in front of the radar

- The PointNet architecture for people counting, focusing only on the spatial properties, achieved 83% accuracy [1].
- *Idea*: include temporal properties via forming a graph
- Graph Neural Networks (GNNs) have been used for gesture recognition from mmWave radar point clouds with an accuracy of 90.53% [2]

2. Research Question

How can the accuracy of people counting from mmWave radar point clouds be improved by using a Graph Neural Network architecture?

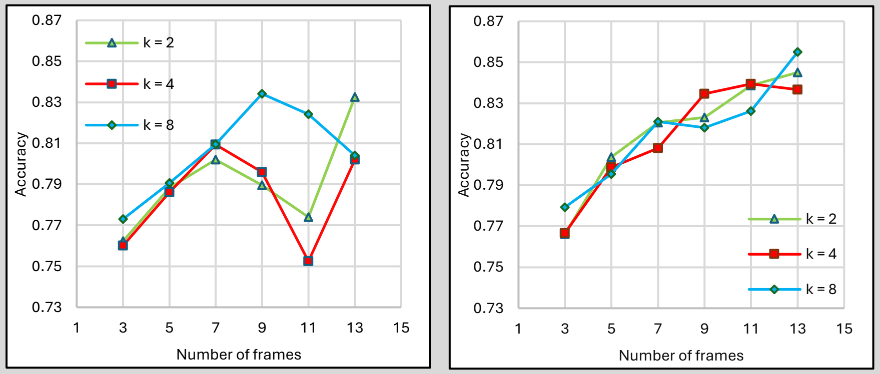

Figure 5: Accuracy of farthest (left) and random (right) neighbors with different number of frames and neighbors

3. Methodology

- **Temporal Graph formation**: The edges in the graph connects successive frames by connecting each point in a frame to k other points in the next frame based on some rule. We considered 3 rules: *nearest*, *farthest*, and *random* neighbors. Fig. 3 demonstrates graph generation with farthest neighbors. The nodes in the graph correspond to the points from the frame with a matching color.
- **Graph processing**: The graphs get processed through a Message-passing Neural Network (MPNN), as shown in Fig. 2. Message passing collects information from neighboring nodes, and therefore includes the temporal depth in the learning process. Self-attention determines the importance of the neighboring nodes.
- The model classifies into 6 classes namely: 1,2,3,4,5 people and bikes.

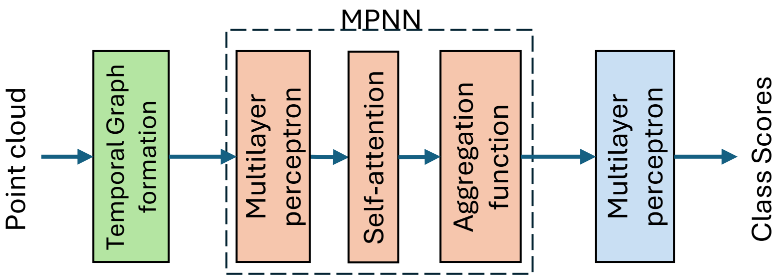

Figure 2: Architecture of the proposed model, inspired by the Tesla model [2]

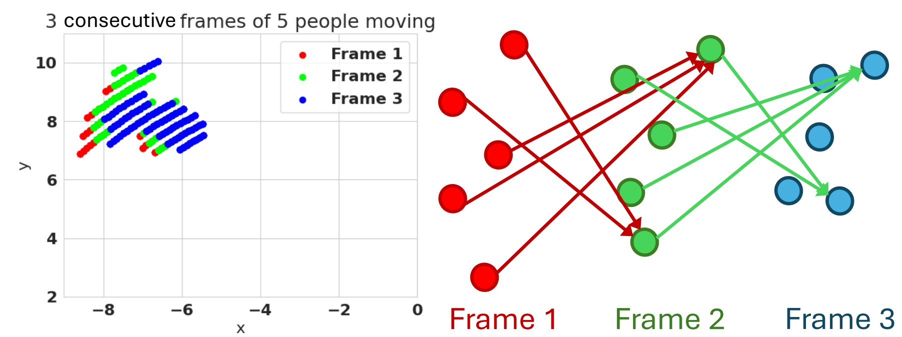

Figure 3: Graph generation from 3 frames with $k = 1$ farthest neighbor

4. Results

- The comparison of the 3 different edge formation rules is presented in Fig. 4. It shows that farthest and random neighbors outperform nearest neighbors.
- As shown in Fig. 5, the performance of random neighbors increases with the number of frames. The accuracy of farthest neighbors, turns inconsistent after 7 frames. No trend can be observed in terms of number of neighbors.
- The achieved accuracy is **80.47%** for 11 frames and 2 random neighbors. The confusion matrix is shown in Fig. 6.
- The proposed model performs worse than the PointNet architecture by 2.53%.
- While the PointNet model uses all the available data, we had to exclude almost 30% of it due to the graph formation requirements.

Nearest neighbors	67.72%
Farthest neighbors	79.11%
Random neighbors	81.41%

Figure 4: Accuracy of different edge formations

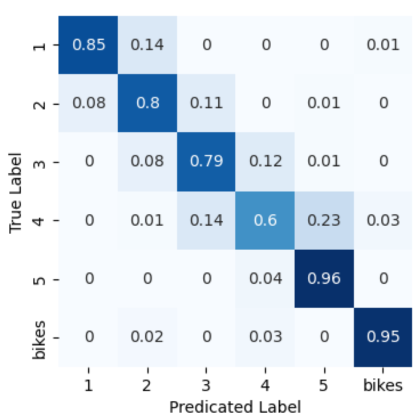

Figure 6: Confusion matrix

5. Limitations and Conclusion

- Only 3 edge formation functions were explored.
- Due to the small number of consecutive frames, at most 13 frames were considered at graph formation.
- The proposed model has achieved a lower accuracy than the PointNet architecture. It however shows potential towards a larger amount of temporal data.

References

[1] G. Vaidya and M. Zuniga, "Exploiting mmWave and Deep-Learning Models to Estimate People Count in Urban Scenarios".
[2] D. Salami, R. Hasibi, S. Palipana, P. Popovski, T. Michoel, and S. Sigg, "Tesla-Rapture: A Lightweight Gesture Recognition System From mmWave Radar Sparse Point Clouds," IEEE Transactions on Mobile Computing, vol. 22, no. 8, pp. 4946–4960, Aug. 2023, doi: 10.1109/TMC.2022.3153717.