Temporal Dynamics Modelling for People Counting in Point Clouds

An Extension on PointNet and MARS through LSTM Integration

Author
Marina Escribano Esteban
m.escribanoesteban@student.tudelft.nl

Supervisors
Marco Zuñiga Zamalloa
Girish Vaidya

1 Background

- **People counting** is needed across domains from high-traffic zone optimisation to forensic investigation.
- Camera surveillance raises ethical concerns about people's privacy, needs strict lighting and is expensive.
- We can use a **millimetre wave (mmWave) radar** which reconstructs individuals as point clouds [1].

Figure 1: Surveillance camera generates an image or video of people [2].

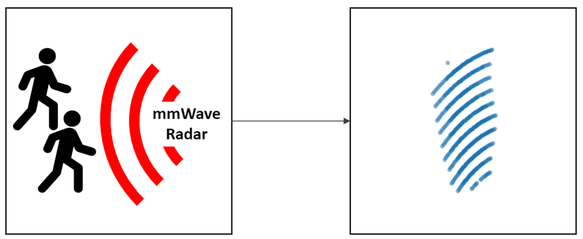

Figure 2: mmWave radar creates point cloud reconstruction.

- **PointNet**: Takes point cloud data as input to segment and classify these 3D sets into their true object class [3].
- **MARS**: Reconstructs human joints and skeletons from mmWave point clouds (estimated joint position) [4].
- **LSTM**: Long Short-Term Memory is a deep learning, sequential neural network that allows information to persist.

2 Objective

How do PointNet and MARS perform when extended by an LSTM to count the people in a point cloud?

3 Methodology

Data set: 19,346 sequences of people in groups of one to five and bikes.

PointNet
- Data pre-processing: Standardise the clusters and add random noise.
- The classification network is adapted to classify into six classes instead of sixteen.

MARS
- Data pre-processing: Standardise the clusters and sort by x, and y coordinates.
- Adapt the linear function used for regression to the Logarithmic Softmax function used in classification tasks.

LSTM
- Build an LSTM of a single layer for the extended models.

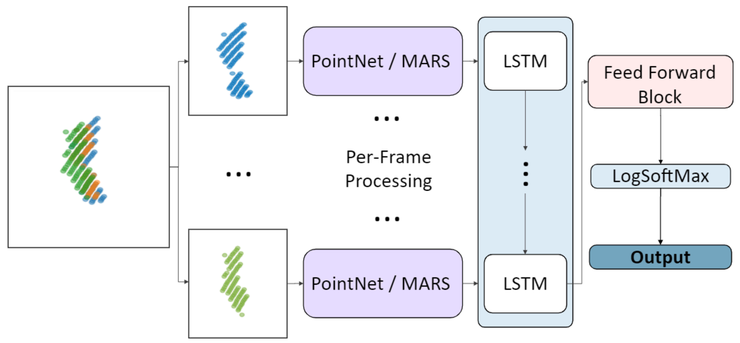

Figure 3: This diagram illustrates the complete architecture of the models including the Feed Forward layer.

4 Results

The four models were trained on the following settings:

- Split of 60% **training**, 20% **validation** and 20% **test data**
- **Number of frames per sequence**: 5
- **Features**: X-coordinate, Y-coordinate, SNR (Singal-To-Noise Ratio), Velocity
- **Epochs**: 30

Model	No. Parameters	Accuracy
PointNet	142,790	62.5%
MARS	4,200,758	60.1%
PointNet +LSTM	2,243,974	80%
MARS +LSTM	5,318,966	89.1%

Figure 4: A summary of each model's total number of parameters and accuracy.

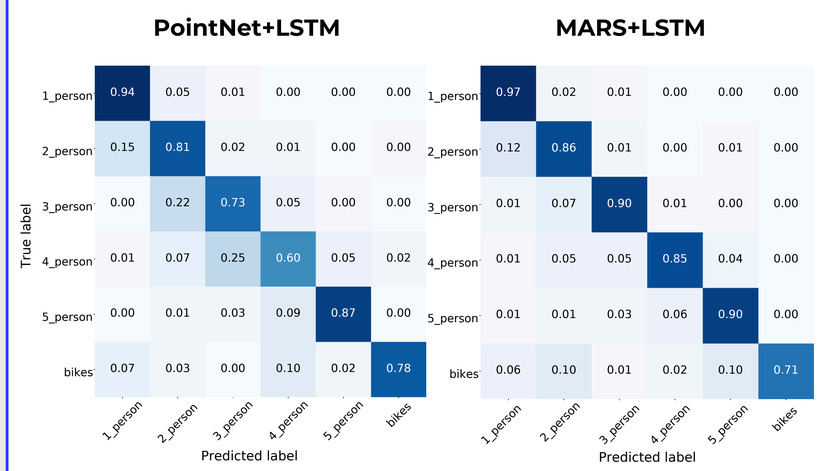

Figure 5: Confusion Matrix of the LSTM-extended models on the test data set.

References

[1] C. Iovescu, 'The fundamentals of millimeter wave sensors', 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:30625501.
[2] Miami TV Mounting. (n.d.). Store with security camera installation. Retrieved June 23, 2024, from https://www.miamitvmounting.com/cameras.html
[3] Charles Ruizhongtai Qi et al. 'PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation'. In: (Dec. 2016)
[4] S. An and U. Y. Ogras, 'MARS: mmWave-based Assistive Rehabilitation System for Smart Healthcare', ACM Transactions on Embedded Computing Syste,s, vol. 20 no. 5s, pp. 1-22, Sep. 2021, doi:10.1145/3477003

5 Discussion and Conclusion

Insights
- LSTMs process frames separately and capture the small changes in local structure. This local information is lost in PointNet from the use of max pooling, however, the CNN in MARS can capture these small changes in cloud structure and therefore perform better when extended by an LSTM.

Limitations
- The number of frames per sequence is limited to a maximum of five.
- All the data was captured from one radar position and the models have not been trained to generalise to more angles.

Conclusion
In conclusion, leveraging the temporal data through an LSTM improves the accuracy of counting people from point clouds in both PointNet and MARS.

TU Delft