Temporal Dynamics in Human Pose Estimation models

Monitoring People without cameras: Privacy is important!

Author
Dan Teodor Savastre

Supervisor
Girish Vaidya

Responsible Professor
Marco Zuñiga Zamalloa

1 Background

- **What is Human Pose Estimation (HPE)?**
  The task of locating key points in the human body and recreating a digital skeleton of the person being analyzed
- **Why do we need HPE?**
  Applications in patient monitoring, sports analysis, robotics, traffic monitoring, and entertainment [1-3]
- **How do we do HPE?**
  - **Cameras:** Require good lighting and poses privacy concerns
  - **Wearables:** Require regular charging, prone to noise [2]
  - **mmWave Radars:**
    - Sense a person as a group of points, as seen in Figure 1.
    - Addresses the privacy concerns of cameras
    - Less strict requirements on deployment conditions
- **Temporal Dynamics:**
  - Patterns and correlations observed over time within a sequence of data. Figure 2 shows the captured pose of three consecutive frames, from which temporal dependencies can be extracted

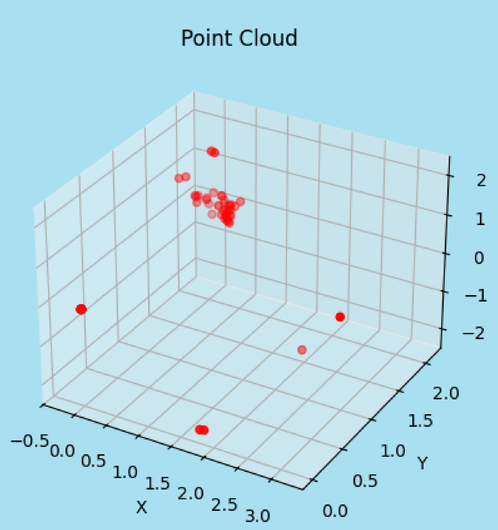

Figure 1. Point cloud of one frame

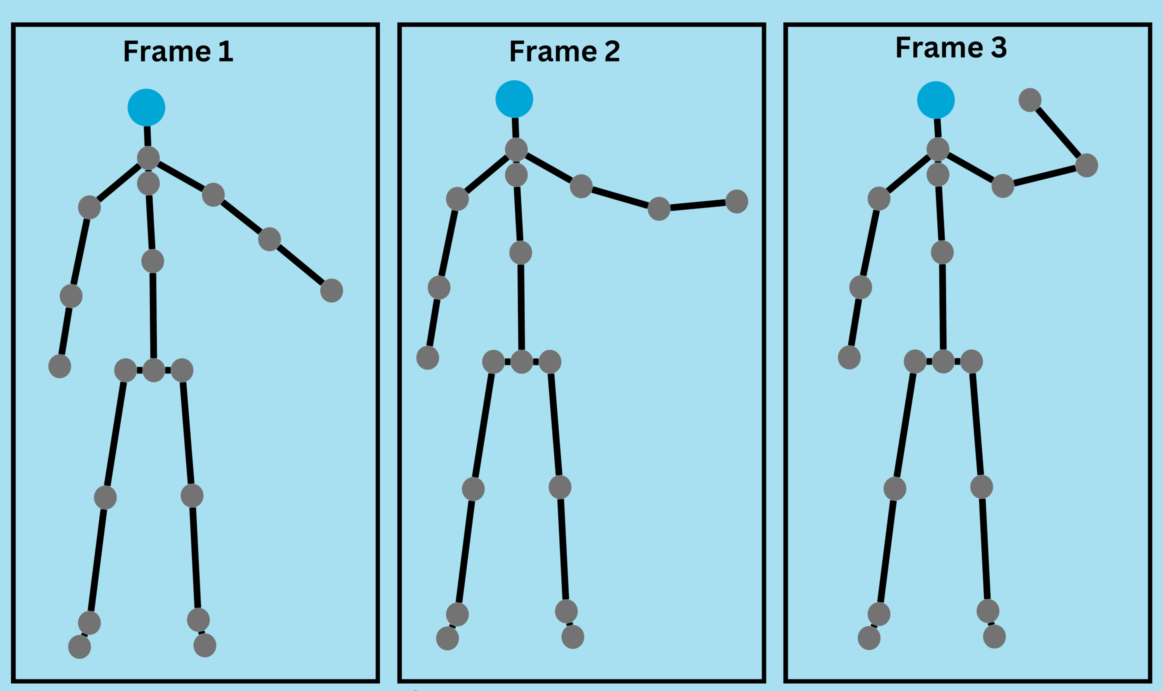

Figure 2. Illustration of human pose across three consecutive frames.

2 Research Question

- **Main Research Question:**
  - What is the impact Temporal Dynamics have on HPE models?
- **Sub-questions:**
  - How is the precision affected?
  - How is the model complexity affected?
- **Hypothesis:**
  - Accounting for temporal dynamics in HPE models will improve the precision without significantly incresing the model complexity

3 Methodology

State-of-the-art Spatial Model:

- MARS: mmWave-based Assistive Rehabilitation System [1] was proposed for assistive rehabilitation in patients with motor disorders.
- Uses point cloud data to estimate 3D coordinates of 19 key points.

Temporal Model:

- This research focuses on enhancing the MARS model by integrating Long Short-Term Memory (LSTM)
- The modified model architecutre can be seen in Figure 3.

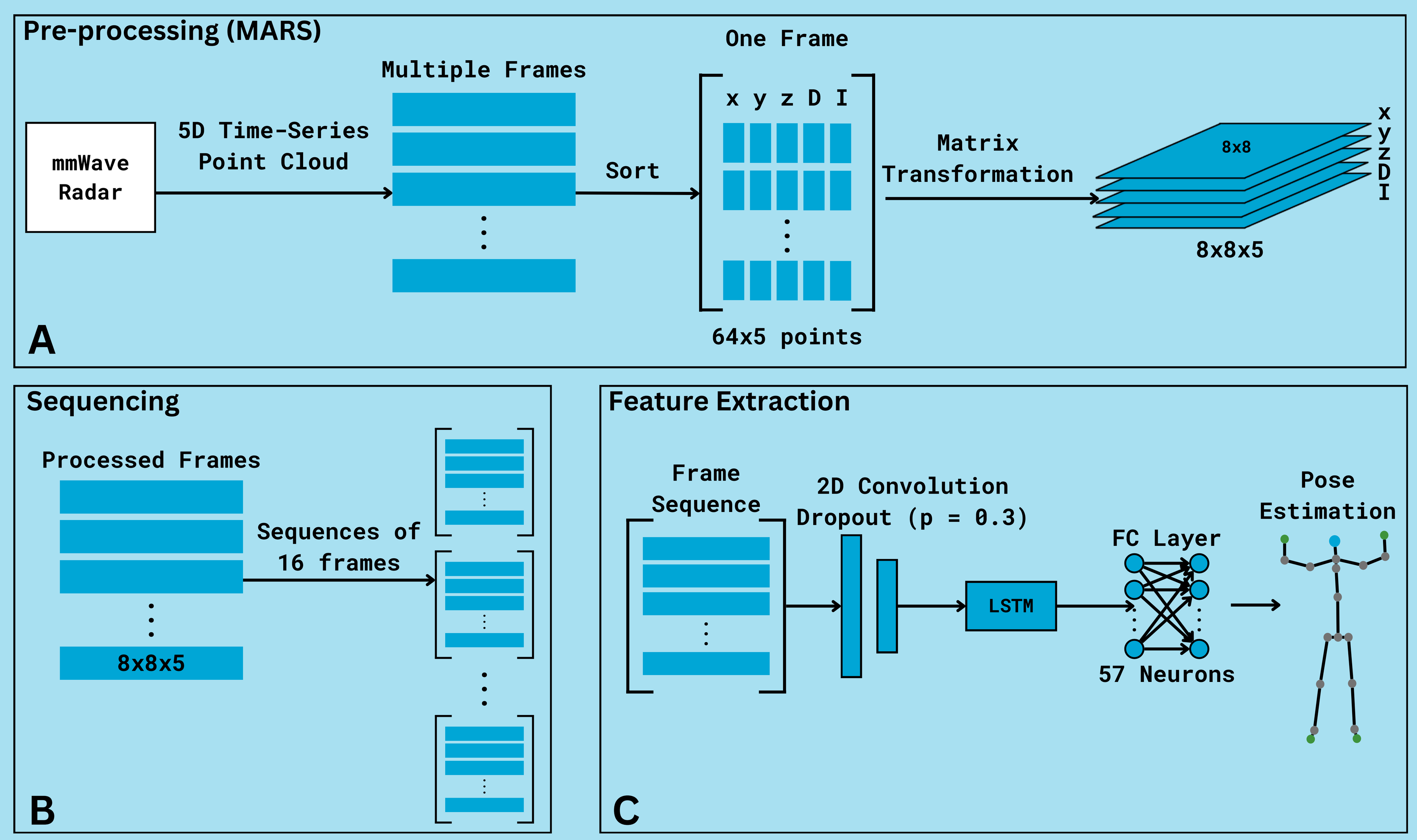

Figure 3. Overview of temporal model architecture. **(A)** Pre-processing strategy from MARS; **(B)** Dividing the time-series into sequences of frames for LSTM; **(C)** Feature extraction using CNN layers of MARS, an LSTM layer and a FC layer as output.

4 Model Optimization

To accurately evaluate the Temporal model's performance, the following parameters need to be optimized:

- **Number of LSTM Units:** Defines the dimensionality of the LSTM layer. Figure 4 shows that having too few units results in a loss of information, while too many results in increased training times.
- **LSTM Memory size:** From Figure 5, it can be seen that small memory sizes give little accuracy improvement and after 16 frames of memory, the minimal improvements come with significant complexity increase.

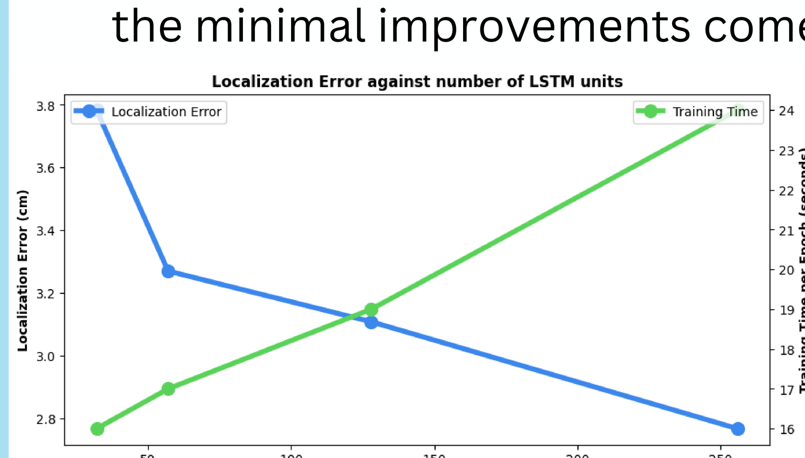

Figure 4. Localization error for optimizing **Number of LSTM units**

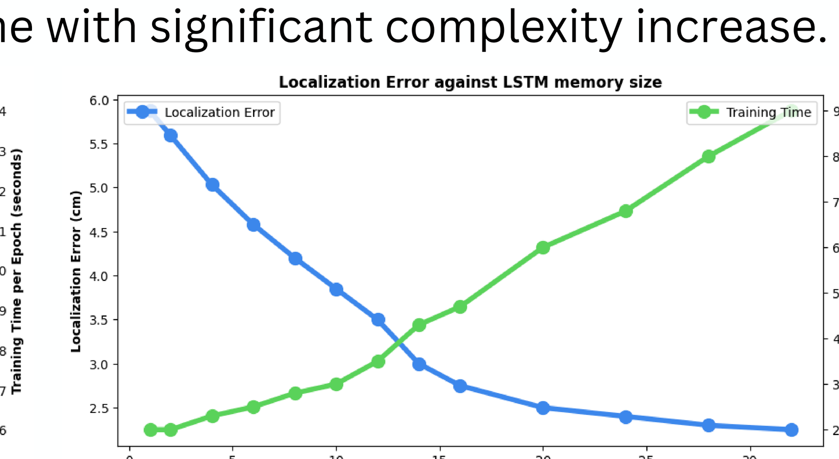

Figure 5. Localization error for optimizing **LSTM Memory Size**

5 Experimental Results

MARS Dataset

The boxplots in Figure 6. show that the Temporal Model provides a significant improvement over MARS. Experimental resluts show that incorporating temporal dynamcs in HPE reduce:

- **MAE by 53%** - Overall accuracy increased
- **RMSE by 45%** - Improvement in worst case scenarios

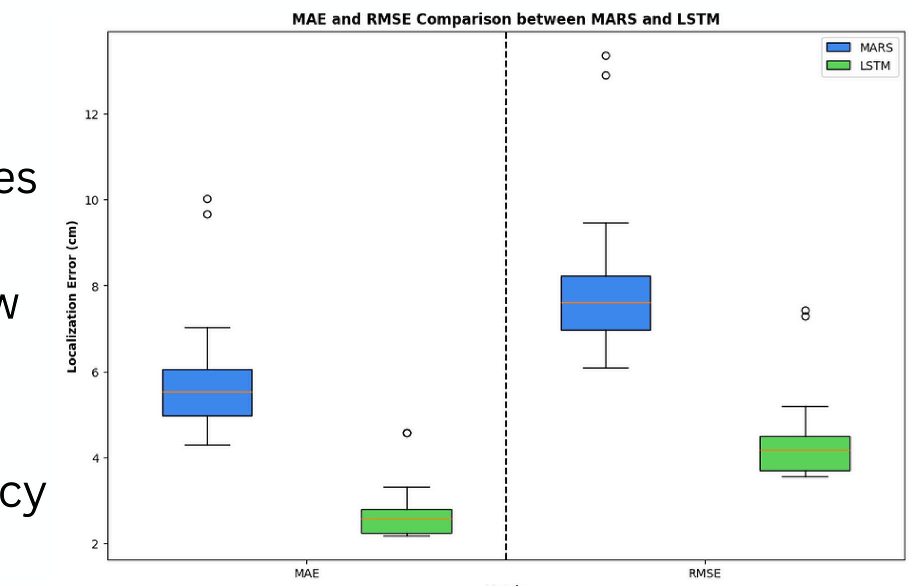

Figure 6. Box plot comparing the localization error of MARS with the Temporal Model

Moving Target Dataset

Evaluating the performance on a dataset featuring more complex movements , the temporal model reduced:

- **MAE by 34%** - Better accuracy for complex movements
- **RMSE by 82%** - Temporal model is effective at reducing outliers and large errros

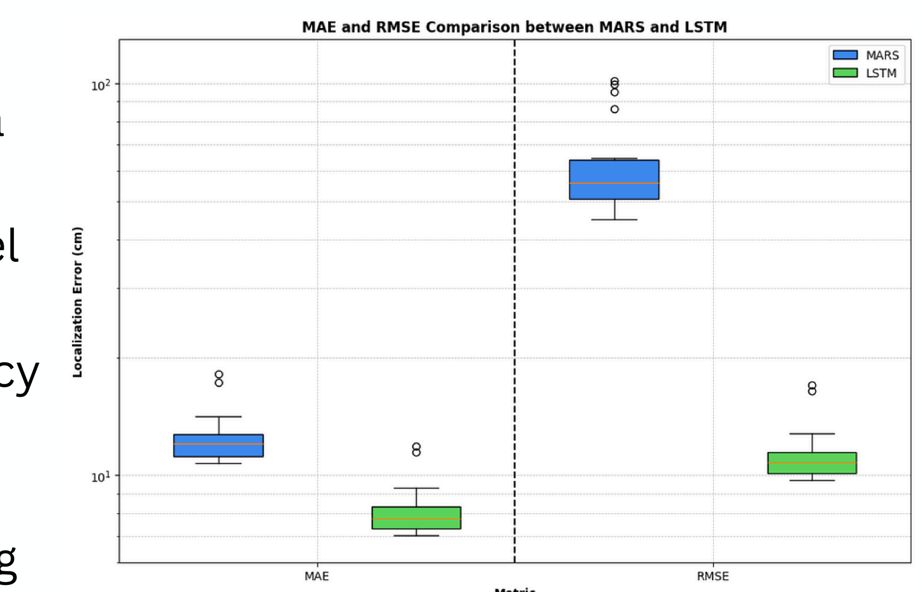

Figure 7. Box plot comparing the localization error of MARS with the Temporal Model

6 Conclusions & Future Work

- Accounting for Temporal Dynamics can significantly improve the performance of HPE models, with reduced impact on model complexity
- Temporal models perform better on more complex movements, such as the Moving Target Dataset.
- Future research could investigate different temporal architectures and compare their performance.
- More comprehensive open-source mmWave datasets are needed to establish benchmarks for future models and for easier comparisons.

7 References

[1] Sizhe An and Umit Y. Ogras. Mars: mmwave-based assistive rehabilitation system for smart healthcare. ACM Transactions on Embedded Computing Systems, 20(5s):1–22, 2021.

[2] G. Bhat, N. Tran, H. Shill, and U. Y. Ogras. w-har: An activity recognition dataset and framework using low-power wearable devices. Sensors, 20(18):26, 2020.

[3] B. Solongontuya, K. J. Cheoi, and M. H. Kim. Novel side pose classification model of stretching gestures using three-layer lstm. Journal of Supercomputing, 77(9):10424– 10440, 2021.

PAPER

CODE

TU Delft