A TinyML system for gesture detection using 3D pre-processed data

Author: Sem van den Broek
s-broek@outlook.com

Supervisors: Mingkun Yang & Ran Zhu
Responsible professor: Qing Wang

1. Background

- Demand for **touchless interfaces** has increased
- Cost, power & availability play a role in system design
- **Machine learning** has been proposed as possible solution for recognition of gesture patterns [1] [2]
- **Accuracy improvement** & **dataset collection** are key research points

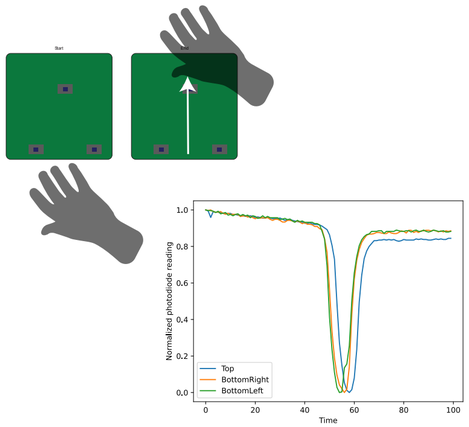

Figure 1: Example of swipe-up gesture sample

2. Research Question

"How to perform gesture detection on a testbed with one Arduino Nano 33 BLE and three OPT101 photodiodes using machine learning based on 3-D pre-processed data?"

Subquestions for:

- Collecting **training data**
- **Type** and **parameters** of model
- **Deployment** of model
- **Metrics** of model

3. Methodology

- Collect data using **specialized tooling** and **research participants**
- Convert input data from **3 photodiodes** to pre-processed 3-D input stream
- Split data **between-candidate** for more representative accuracy metrics

Input	Dataset															
Split	Train												Test			
Shuffled random	C3				C2				C4				C1			
Shuffled random	S10	S9	S11	S12	S6	S7	S5	S8	S13	S16	S15	S14	S3	S2	S1	S4

Figure 2: Visualisation of train/test split

- Use a **Recurrent Neural Network (RNN)** like LSTM
- **Optimize** model to run on microcontroller
- Investigate inference problems using **a different microcontroller**

4. Results

- Right & left hands **(65/35 ratio)**
- **±230** measurements per gesture
- Best 3D model: **ConvLSTM-128** [1]
- K-Fold accuracy (10 folds): **70.7%**

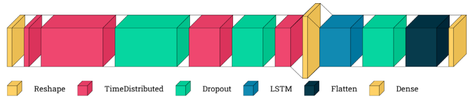

Figure 3: Visualisation of ConvLSTM-128 model [3]

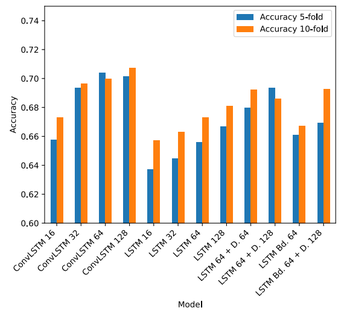

Figure 4: Fold accuracy of various models

- Confusion is apparent in **zoom in** and **zoom out** gestures
- Swipe gestures perform better

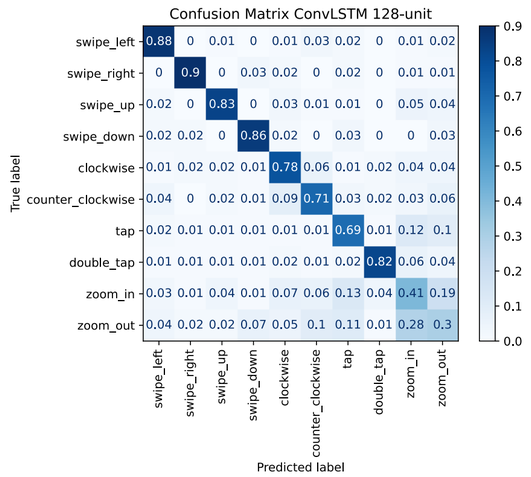

Figure 5: Confusion Matrix of ConvLSTM-128 model

- Inference output **not correct**
- Inference times on Arduino is sufficient for real-time usage
- Inference times on different microcontroller architectures differ

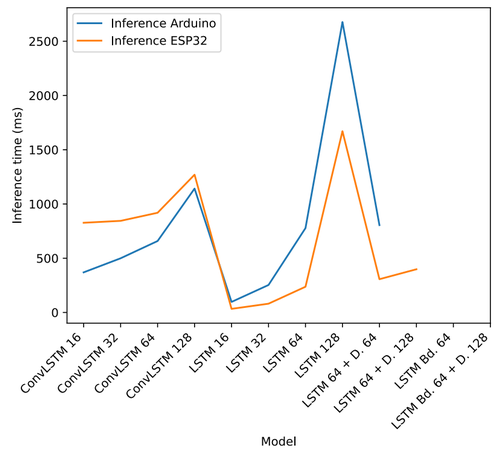

Figure 6: Inference times of various models

5. Conclusion

- **Convolutional model** seems to have an advantage
- **Accuracy** not high enough for real-world usage
- **Inference** fast enough for normal applications
- **10-classification** feasible
- Gestures like **zoom in** and **zoom out** could be replaced

Future work:

- **More data** for better accuracy
- Possible hardware change for more intricate gestures
- Correct inference output on embedded devices

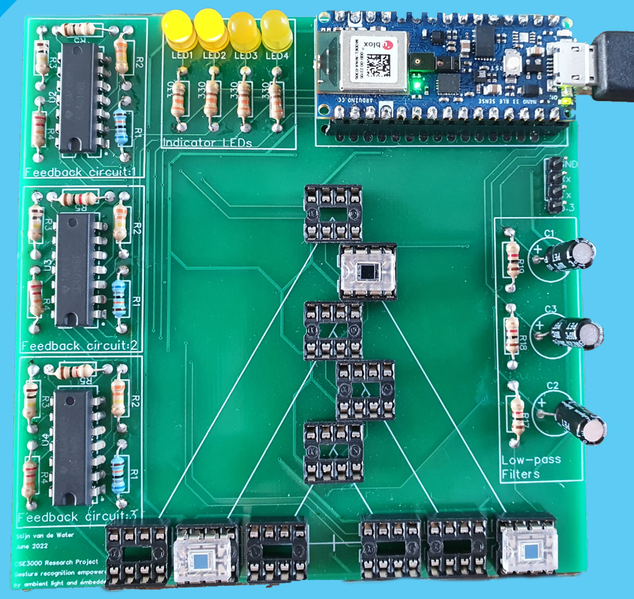

Figure 7: Photo of the custom-built testbed

[1] M. Lipski, 'Hand Gesture Recognition on Arduino Using Recurrent Neural Networks and Ambient Light', Delft University of Technology, 6 2022.
[2] W. Narchi, 'Recognising Gestures Using Ambient Light and Convolutional Neural Networks: Adapting Convolutional Neural Networks for Gesture Recognition on Resource-constrained Microcontrollers', Delft University of Technology, 6 2022.
[3] P. Gavrikov, 'visualkeras', GitHub repository. GitHub, 2020.

TU Delft