Recognising Gestures Using Ambient Light and Convolutional Neural Networks

CSE3000 Research Project
William Narchi (w.narchi-1@student.tudelft.nl)
Responsible Professor: Qing Wang | Supervisors: Mingkun Yang - Ran Zhu

TU Delft Delft University of Technology

1. Introduction

- Goal of overall research project is to enable **affordable and non-intrusive gesture recognition**
- Usage of **ambient lighting** coupled with low-cost **photodiodes** (light sensors) and an **embedded platform**
- Hand movement creates unique shadow patterns for each gesture
- Utilising **machine learning** for recognition
- Resource-constrained, highly economic environment
  - Arduino Nano 33 BLE
  - 3 photodiodes in total

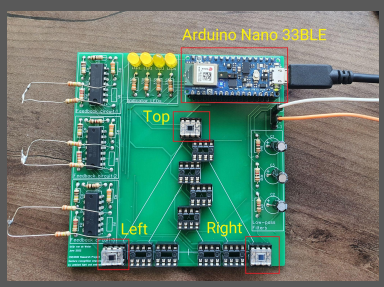

Figure 1. Gesture recognition assembly

2. Research Questions

1. *"Which model could be used for gesture recognition, based on 2-D pre-processed data (like picture recognition)?"*
2. *"How to compress the used deep learning to make it real-time on Arduino Nano 33 BLE?"*

3. Methodology

Machine Learning Architecture

- Structure readings from 3 photodiodes as an **(n X 3) image** with a single 'channel'
- Use (relatively shallow) **Convolutional Neural Network (CNN)**

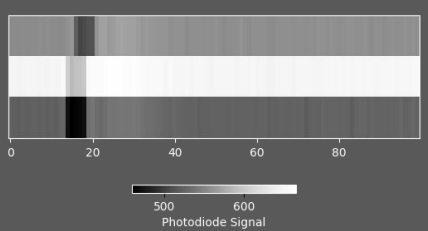

Figure 2. Visualisation of CNN input corresponding to an upwards swipe (rendered as (3 x n) for visibility)

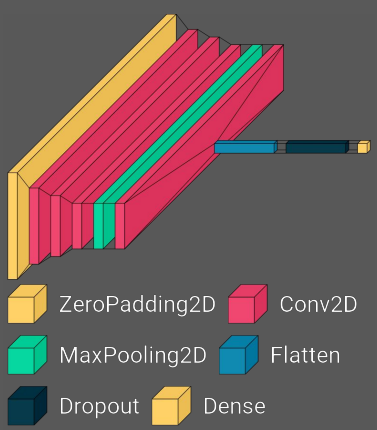

Figure 3. Structure of final CNN [1]

Embedded Optimisations

- Utilise **TensorFlow Lite** for running on the Arduino platform
  - Allows for various optimizations, including **quantization** and **operator omission**
  - Significant improvement to final model size
    - Unoptimised: **105KiB**
    - Optimised: **32KiB**

4. Results and Evaluation

- **6 models** evaluated in total
- Dataset consists of **10 gestures** from **~50 participants** in light conditions from 0 lux to 100,000 lux
- **K-fold cross-validation** utilised with entire **raw** dataset
- Accuracy ranges from **70%** to **87%** depending on test scenario
- Optimisations drop accuracy by **2%** to **10%** depending on test scenario
- All models suitable for real-time operation
  - Optimised model sizes range from **16KiB** to **37KiB**
  - Inference latency ranges from **44ms** to **128ms**
- Results of the **final model**
  - Accuracy (5-fold): **79.220% (± 6.516%)**
  - Accuracy (10-fold): **86.798% (± 5.722%)**
  - Optimised accuracy (5-fold): **75.388% (± 6.376%)**
  - Optimised accuracy (10-fold): **80.954% (± 4.753%)**
  - Size: **32,849 bytes**
  - Inference latency: **78ms**
- Overall system memory usage
  - RAM: **69,536 bytes**
  - Flash memory: **162,224 bytes**

5. Conclusion

- Amicable results, real-time gesture recognition feasible with specified setup
- Future work
  - Experimentation with additional CNN architectures
  - Inclusion of processing pipeline

6. References

[1] P. Gavrikov, visualkeras. GitHub, 2020. [Online]. Available: https://github.com/paulgavrikov/visualkeras