Digit recognition with visible light using three photodiodes and 3D-preprocessed data

A research project for CSE3000 Research Project by Gijs van de Linde (**G.vandeLinde@student.tudelft.nl**)
With assistance from responsible professor Qing Wang and supervisors Ran Zhu and Mingkun Yang.

TU Delft

1. Introduction

The COVID pandemic made humans mentally aware of the bacterial and viral traces they leave behind on objects. Though these traces on shared surfaces are rarely the cause of infection [1], they did cause "touch anxiety" among people. This was reported by therapists [2] and found in a survey by London South Bank University [3]. This caused a necessity for public devices allowing for touch-free interaction, which has inspired researchers to look for non-contact means of taking input from a user, as an alternative to touch screens. A solution is researched in this paper that uses visible light to detect digit gestures using 3D-formatted data.

2. Background

It's important to define what is meant by "3D-formatted data". This refers to the data being formatted in frames, which makes the data more like a video instead of an image.
Figure 1 shows data points of photodiodes over time, divided into frames of size 3.

	Frame 1			Frame 2			Frame 3			Frame 4		
PD 1	0.4	0.5	0.5	0.6	0.6	0.6	0.7	0.1	0.2	0.2	0.1	0.1
PD 2	0.2	0.1	0.6	0.5	0.5	0.4	0.2	0.2	0.2	0.1	0.1	0.1
PD 3	0.9	0.8	0.9	0.9	0.9	0.5	0.1	0.1	0.1	0.2	0.2	0.1

t→

Figure 1: visualization of 3D data

The hardware for this project was designed last year by Stijn van de Water [4]. An important conclusion that was taken from this research was that in a triangular setup, a distance of 5 cm between each photodiode was found the best balance between the precision and robustness of the system. The algorithm for the automatic adaption of photodiodes to the light environment was also used (as described in section 5.1). The setup is shown in Figure 2. It uses an Arduino 33 BLE and three photodiodes.

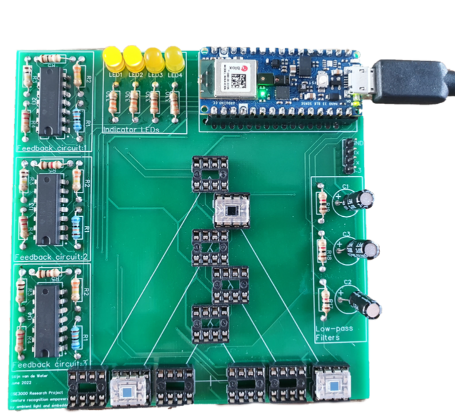

Figure 2: hardware setup of the project

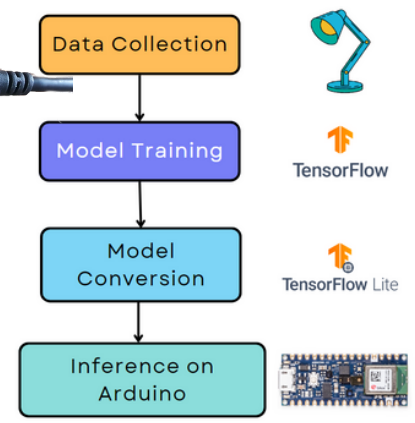

Figure 3: Methodology visually summarized.

3. Method

The methodology for this research is briefly summarized in figure 3. It consisted of data collection, training models, and converting these models to TensorFlow Lite models, such that they could be run on the Arduino Nano 33 BLE. A data set was created in a controlled lighting environment, using. This data set contains approximately twenty participants. Each collected gesture consisted of a 2-second sample at 1000 Hz. A bright lamp was placed approximately 30 cm above the PCB during the experiment, which caused the photodiode collaboration algorithm calibrate to 122000 Ohm.

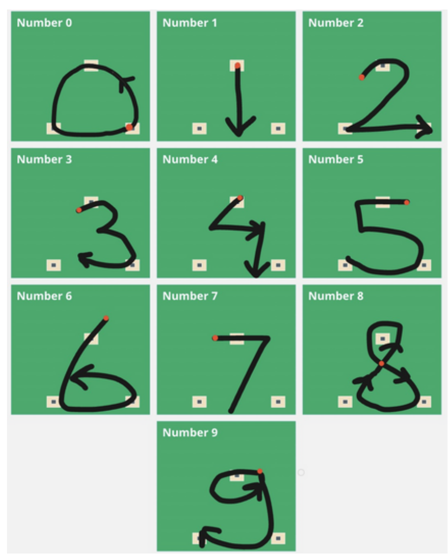

Figure 4: Way participants were instructed to draw digits

We chose multiple deep learning models to test. To evaluate the deep learning models, we did 5-Fold cross-validation using a split based on candidates. This resulted in the average accuracy and standard deviation metrics. The second metric that is considered is the size of the model converted to a TensorFlow Lite version. During evaluation the frame size was fixed to 5, and the learning rate to 0.0005.

During data collection a gesture detection algorithm was used to capture the data samples. This makes the dataset more representative, since this would be the way a sample would be presented to the model during final inference on the Arduino also. Furthermore an effort was made to make the data representative by using candidates from different ages and genders. It should however be noted that it is exclusively made up of right-handed people.

4. Results and conclusions

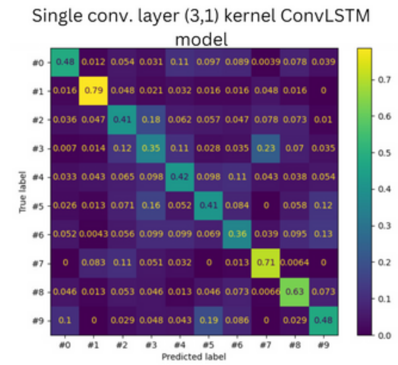

Figure 5: confusion matrix for first model

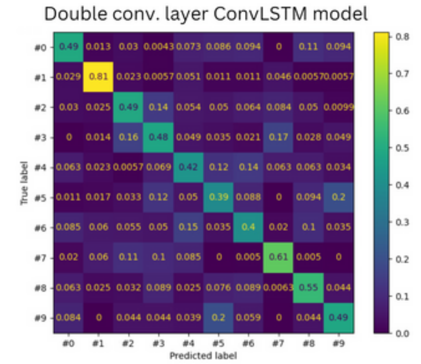

Figure 6: confusion matrix for second model

There are two models that performed the best during our experiments. First off the ConvLSTM with a single convolutional layer that uses a (3,1) shaped kernel with a KFold cross-validation accuracy of 0.491 ± 0.073, for which the confusion matrix is shown in Figure 5. Secondly, the ConvLSTM model with two convolutional layers of which one uses a (3,1) shaped kernel and one uses a (2, 2) shape with a KFold cross-validation accuracy of 0.500 ± 0.091, for which the confusion matrix is shown in Figure 6.

5. Future work

Future work that should be considered is summarized here:

- Research different frame sizes.
- Research overfitting when using different learning rates.
- Creating a model using augmented data.
- Research quantization and run evaluation on a fully quantized model.
- Do an analysis of the inference latencies for our models.
- Do an analysis on the threshold used for gesture detection.

[1] E. Goldman. "Exaggerated risk of transmission of COVID-19 by fomites". in The Lancet Infectious Diseases: 20.8 (2020), pages 892–893. DOI: https://doi.org/10.1016/S1473-3099(20)30561-2.
[2] Ellie Violet Bramley. Therapists report huge rise in cases of anxiety as England ends Covid rules. 2021. URL: https://www.theguardian.com/world/2021/jul/11/therapists-report-huge-rise-in-cases-of-anxiety-as-england-ends-covid-rules.
London South Bank University. The pandemic's mental toll: new survey finds one in five suffer from Covid-19 Anxiety Syndrome. 2021. URL: https://www.lsbu.ac.uk/about-us/news/the-pandemics-mental-toll-new-survey-finds-one-in-five-suffer-from-covid-19-anxiety-syndrome.
[4] S. van de Water. Designing an adaptable and low-cost system for gesture recognition using visible light.2022.