EESTEC Challenge Milano 2022 Final Round

Trivial Group (Janez Justin, Frenk Dragar), Slovenia

# Scope and purpose

This document provides a description of our submission, including data collection, pre-processing, model architecture and detailed specification.

# Intended audience

Infineon representatives in Milan.

Table of contents

[Table of contents 1](#_Toc104480006)

[1 Data 2](#_Toc104480007)

[1.1 Collection 2](#_Toc104480008)

[1.2 Pre-processing 2](#_Toc104480009)

[2 Algorithm 4](#_Toc104480010)

[2.1 Architecture 4](#_Toc104480011)

[2.2 Performance 4](#_Toc104480012)

[2.3 Specification 4](#_Toc104480013)

[3 Evaluation and Conclusion 5](#_Toc104480014)

[Revision history 6](#_Toc104480015)

# Data

## Collection

The data was collected in three separate environments – one outside and two inside. All places had a wall at least 3-4 meters in front of the radar and no obstacles placed in the area.

The first dataset – “Bolognese” was created using 4 capture sessions, each lasting about a minute, totaling 1200 frames of footage with the classes (0, 1, 2, 3 people) balanced, with each representing 25% of the total dataset. “Bolognese” is based inside of Polytechnic Milano and features moving people in the frame (exaugurated movements).

The second dataset – “Pesto” was created using 4 capture sessions, each lasting about two minutes, totaling about 3000 frames of footage with the classes (0, 1, 2, 3 people) balanced, with each representing about 25% of the total dataset. “Pesto” was recorded outside of Polytechnic Milano and features moving people in the frame (exaugurated movements).

The third dataset – “Tono” was created one capture session, each lasting about two minutes, totaling about 3000 frames of footage with the classes (0, 1, 2, 3 people) balanced, with each representing about 25% of the total dataset. The video was then manually labeled by timestamps. “Tono” was recorded inside of Polytechnic Milano and features moving people in the frame (exaugurated movements). The background is about 3 meters away, shielded by a makeshift table wall.

In our training, we made use of all these datasets and combined them freely in order to provide some robustness regarding people detection in different environments.

## Pre-processing

Pre-processing was first done using Infineon’s provided script, which applies the Fast Fourier Transform algorithm on raw sensor data, in order to get back a Doppler map.

After the Doppler map is acquired, we apply some more pre-processing:

* Removing the middle rows (horizontal position 32:34 in the array), as they were practically static in the sensor data input and might interfere with the results,
* Subtracting the mean of the array from the data,
* Removing any data from more than 4.5 meters in distance (as the specification requests only 3 meters to be detected),
* Normalization (0.0 – 1.0 floating point numbers as inputs).

We also introduced some Gaussian noise upon the training data (0.05 \* rand(0,1) on about 20% of the training dataset) in order to improve model robustness.

Shape

Description automatically generatedA picture containing application

Description automatically generatedA picture containing text, white

Description automatically generatedA picture containing chart

Description automatically generated

*Figure 1: Images (from left to right) show: plotted raw sensor data, transformed sensor data (FFT), transformed data with the middle rows removed, and training data with added Gaussian noise.*

# Algorithm

## Architecture

We iterated through a couple of Neural network models, using the Pytorch library. The models “Penne”, “Fettucine”, “Rotini” and “Spaghetti” were all implemented using a varying amount of 2d convolutional layers, fully linked layers, and ReLu activation functions.

In the end, our best performing model was “Spaghetti” model, contains 1 max pooling layer, 5 convolution layers, 7 linear layers, and “LeakyReLu” activation functions.

The model was trained using a Cross-Entropy loss function and Adam optimizer.

## Performance

We tested our model on the split dataset (80-20 split), where we achieved a 98% classification accuracy, which is unfortunately optimistic, as the model overfit onto the training set, which was very similar to the test set. This means the classification accuracy is not a valid metric in this case.

Graphical user interface

Description automatically generated

Figure 2: "Spaghetti al Pesto e Tono" model confusion matrix on the dataset test split

In our internal and “intuition base” testing, we found that using the model was correct most of the time when predicting the “0” and “1” people classes, but had some trouble differentiating between the “2” and “3” people classes. It is however highly accurate as a binary classificatory (moving objects vs. non-moving objects).

## Specification

The model is 881 kb in size, which makes it not quite small, but it should be able to run on some embedded devices with more memory.

# Evaluation and Conclusion

The “Spaghetti al Pesto e Tono” model performs reasonably well using its dataset and CNN architecture, but there needs to be some more time put into dataset crafting, preprocessing, and model architecture.

The model size could probably be greatly reduced as well.

Revision history

| Document version | Date of release | Description of changes |
| --- | --- | --- |
| v1.0 | 26. 5. 2022 | Initial commit. |
|  |  |  |
|  |  |  |