Project II - Transformers
Project Plan

Kinga Frańczak, 313335
Grzegorz Zakrzewski, 313555

1 What we are going to test

1.1 Network architectures

We are going to test and compare the following network architectures:

- a simple, feed-forward neural network with three (or more) dense layers utilizing the ReLU activation function and an additional dropout layer;
- the architecture based on the best neural network architecture developed during Project I, which consists of a convolutional neural network with three convolutional layers, pooling layers interspersed between them, a normalization layer, and dense layers at the end, one employing the ReLU activation function and the other utilizing the softmax activation function for classification purposes;
- Transformer architecture implemented using the `MultiHeadAttention` layer from Keras along with other necessary dense, normalization, and dropout layers.

1.2 Hyper-parameters

Given that the focus of this project is on Transformers, we intend to explore the influence of parameters such as the number of attention heads and the size of each attention head for query and key. After reviewing the literature, we will propose several different values for these parameters and assess their impact on the results.

1.3 Predictions & evaluation

Ideally, we aim to train our models on the full dataset, which includes all ten base classes as well as the *silence* and *unknown* classes. We will consider three scenarios:

- utilizing one Transformer to recognize all ten base classes, as well as the two special cases;
- employing a Transformer for the base classes and a separate Transformer for the special cases, distinguishing whether the observation belongs to the *silence* class, the *unknown* class, or neither of these special cases;
- using a Transformer for the base classes and two separate Transformers for each of the special cases.

2 Short schedule

The final four weeks will be allocated as follows: the first week for reviewing Transformer resources and defining network architectures and hyper-parameters, the second and third weeks for coding and conducting experiments, and the last week for report writing.