

# Hands-On Assignment

## Deep Learning

### 880008-M-6

Spring, 2021

## Description

This is a group assignment. Groups should be of 5 people. The assignment is a shared task consisting of a re-identification problem.

You will need to implement a model and evaluate it on a given test set. You have to submit your predictions for the test set, as well as a report describing your solution(s) and the code you used. Your predictions will be compared to the ones of a baseline model.

This assignment is worth 30% of your course grade. The assignment grade will be based on your work's quality as judged by the instructors based on your **report** and **code**.

Additionally, you can get a bonus based on your ranking on a shared task leader board. Specifically:

- if your rank first, you will receive bonus 2 points;
- if your score is no better than a provided baseline you will receive no bonus;
- for intermediate ranks the bonus points will be linearly interpolated. The performance of the baseline solution will be shown on Codalab.

All the information about the Codalab submission will be specified on a separated document later on.

Passing the assignment is not mandatory to pass the course, but it is highly advisable. There will be no resit for this assignment as passing the it is not compulsory. The exam may include questions that might be easier to answer if you have worked on the shared task.

## 1 Shared Task – Re-identification from voice

Analyzing sequential data or time series is very relevant and extensively explored task in the field of Deep Learning. This kind of data appears in many domains and formats; for example, stock prices, videos, and electrophysiological signals.

In this task, you will be working with voice recordings. In this particular case, the data set consists of short audio fragments of different people reading aloud. Everyone's voice has specific

characteristics displayed in, for example, pitch timbre or tone, which allow us to recognize anyone's voice as unique.

The human voice normally spans a frequency range from about 100Hz to 8kHz. The lowest frequency of any voice signal is called the Fundamental Frequency. The average fundamental frequency for a male voice is 125Hz, and for a female voice it is 200Hz. When speaking, consonants take up space between 2kHz and 5kHz. These sounds pass quickly and can help make speech more intelligible. Similarly, vowel sounds are most prominent between 500Hz and 2kHz.

In this task, the goal is to build a Deep Learning model that is capable of identifying whether two recordings belong to the same person or not. More specifically, given a set of recordings, your final model should be capable of determining which of the recordings are most closely related to each other, and so creating a measure of how likely two recordings belong to the same person.

Each sequence in the data set contains only one person's voice and lasts a second. There are several recordings for each person. Every sequence has an id (label) assigned to it. All of the recordings that belong to the same person have the same id.

Your final model should be evaluated on a given unlabeled Test set that will be provided later. None of the audios in the test set belong to any of the persons in the training set.

## 2 Data format

The data is given in a .pkl file that should be loaded using the pickle library. The file contains the sequences and ids (labels) in two arrays: [data, labels]. Each sequence has 11025 points, corresponding to 1 second. There is more than one sequence with the same id for every id in the set.

## 3 Important dates and deliverables

### 3.1 Report

A one-page report should be submitted by March 21st, 2021. The report should include the following:

- Title including names and student numbers
- A diagram of the architecture used.
- Brief description of your experiments, including training, hyperparameters, and optimization.
- A table with the results and, optionally, a graph, e.g. a line chart showing the accuracy progression over epochs/iterations; or an AUC-ROC graph.
- Short discussion and conclusions on the performance of your solution, preferably in bullet points.
- The name of the account under which you submit your results to Codalab.

**Your report should be a PDF document, with a single page of content**, and optionally additional pages for references and appendices. Note that the content page needs to be self-contained, and the appendices should only contain auxiliary material. If any of the points above are in the appendix the report will be considered as not presented.

### 3.1.1 Code

Your code should be a plain Python script that can be run to generate your predictions. You do not need to include the training data or the weights of your trained model.

### 3.1.2 Submission format

For the submission, your report and code should be in a single zip file named with your group ID, e.g. group\_1.zip, and submitted through the assignment on Canvas.

## 3.2 Performance

In addition to the report, you will need to submit a file with the predictions to the competition's server. The competition will be hosted on <https://competitions.codalab.org>. You will need a Codalab account for the group; make sure to indicate this account's name in your report. The account should be created with a TiU email account. There will be a separate document on the submission to Codalab with additional details.