

# Lab Homeworks – Fall 2024

July 2024

## Homework L2 (10p):

Build and train a multi-input **PyTorch** network that will accept both the raw signal and the Mel Spectrogram for classifying **GTZAN** dataset. For the head extracting features from Mel Spectrograms, you should experiment with **at least 3** different pre-trained 2D architectures from **torchvision**<sup>1</sup>, using transfer learning and fine-tuning.

**Note:** you should not use the final output of a pre-trained network (i.e. logits), but rather some feature vector computed in a previous layer.

## Homework L3 (10p):

Given the **ObjectDetectionModel** defined in Lab 3, your task is to measure the influence **n\_cells** has on recognizing small objects. That is, you should experiment with training under different values **nc<sub>i</sub>** for **n\_cells**, measure the detection accuracy for traffic signs occupying less than different fractions  $f_j$ ,  $f_i \in (0, 1)$ , of the input image, and construct a final *heatmap* containing accuracy scores for each combination (**nc<sub>i</sub>**,  $f_j$ ).

**Note:** You should experiment with at least 5 values for **nc<sub>i</sub>**, and at least 5 values for  $f_i$  (i.e., the final heatmap should be at least  $5 \times 5$  in size). Since we encoded all bounding boxes as squares, you'll use  $f_i = \frac{h^2}{HW}$ , where  $H, W$  are the full image dimensions, and  $h$  is the bounding box height (or width).

## Homework L4 (10p):

Compare the influence of different pre-trained word embeddings from **torchtext** (or from anywhere else) before and after fine-tuning on the task of movie review classification from Lab 4.

---

<sup>1</sup><https://pytorch.org/vision/stable/models.html#classification>