Classification of vehicles based on the audio signal

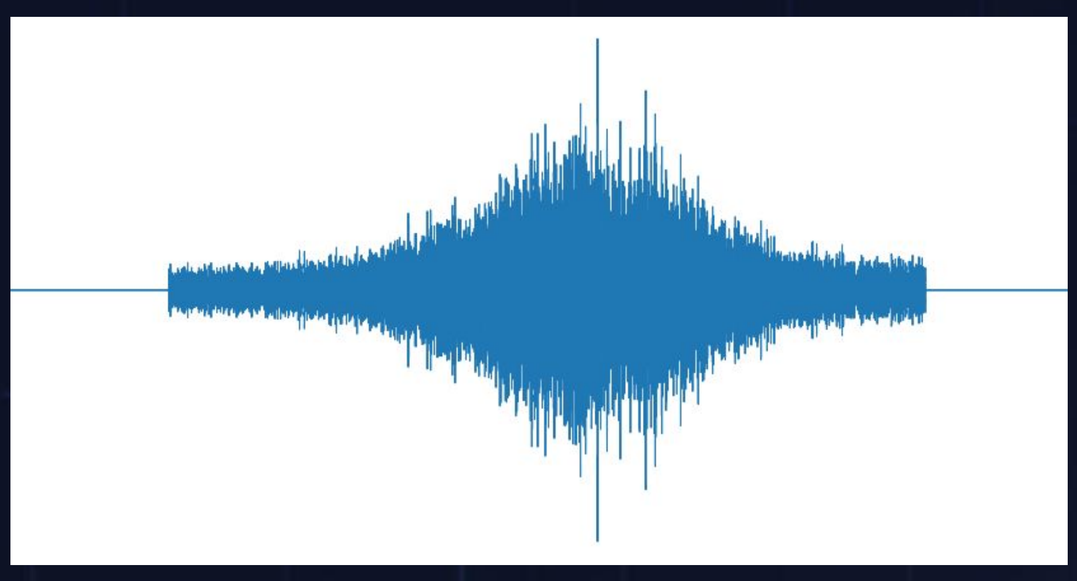
Contractors: Mateusz Erezman, Michał Hajdasz, Dariusz Kobiela
Uczenie Maszynowe, Wydział Elektroniki, Telekomunikacji i Informatyki, Politechnika Gdańska
Project supervisor: mgr inż. Szymon Zaporowski

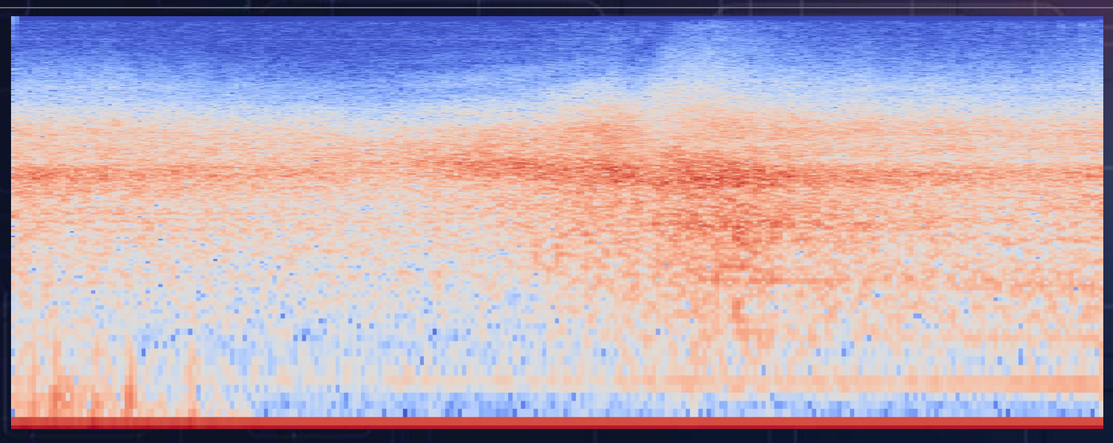
Project goal

The goal of the project was to create a neural network model, which basing on the input data in the form of an audio files would be able to recognize the type of the vehicle passing. The project was a continuation of the research carried out by Czyżewski, Kurowski and Zaporowski (doi.org/10.1121/1.5137275).

Input data

Input data for the project was the raw audio file with the sound of the approaching vehicle. After data preprocessing, all the source videos were saved in the WAV format. Additionally, the video in AVI format of the vehicle was provided. The video data was used in order to tag the audio data into one of the 3 classes: car, motorcycle or truck. According to the pilot study, the "truck" class contained was the tag for vehicles such as truck, bus and van. Most of the data was manually tagged by the authors of this article.





Preprocessing

Then, the labeled and converted to the wave format data is transformed using MFCC (Mel-frequency cepstral coefficients) and spectrograms - visual representation of the spectrum of frequencies of a signal as it varies with time. It was decided to use log-spectrograms because of the fact that they were bringing more useful information.

	precision	recall	f1-score	support
car	0.94	0.89	0.91	185
truck	0.75	0.86	0.80	70
motorcycle	0.70	0.78	0.74	9
macro avg	0.80	0.84	0.82	264
weighted avg	0.88	0.88	**0.88**	264

Model and results

According to the SLR (Systematic Literature Review), the most popular model to be used was CNN (Convolutional Neural Network). The obtained results confirms, that the best **CNN model,** implemented in pytorch with the usage of spectrograms and with the augmented data, achieves the metrics values on the level **of accuracy = 0.875** and **f1-score = 0.88**. It was decided to use time shift and masking out with horizontal and vertical black bars (time and frequency). Therefore, comparing to the solution shown in Czyżewski et al., where SVC (Support Vector Classifier) model achieved the weighted accuracy at the level of 0.93 and weighted f1-score = 0.54, the proposed CNN model significantly beats the SVM in the terms of f1-score. The accuracy of CNN model is slightly lower, however, the approach showed in Czyżewski et al. takes into consideration only 2 classes (car and truck), while this approach takes 3 classes (car, truck and motorcycle).

Projekt współfinansowany ze środków Unii Europejskiej w ramach Europejskiego Funduszu Rozwoju Regionalnego.
Program Operacyjny Polska Cyfrowa na lata 2014-2020.
Oś priorytetowa nr 3 „Cyfrowe kompetencje społeczeństwa”, działanie nr 3.2 „Innowacyjne rozwiązania na rzecz aktywizacji cyfrowej”.
Tytuł projektu: „Akademia Innowacyjnych Zastosowań Technologii Cyfrowych (AI Tech)”.