SZAKDOLGOZAT

Nyilas Péter
Mérnökinformatikus hallgató részére

Közlekedési objektumok detekcióját megvalósító neurális hálós modell tanítása és magyarázata

A közlekedési objektumok, mint például autók, gyalogosok, kamionok, motorkerékpárok, biciklik és egyéb dinamikus közlekedési szereplők felismerése kritikus fontosságú feladat az önvezető, illetve a vezetést támogató rendszerek, valamint közlekedési monitoring rendszerek fejlesztése során. A neurális hálókon alapuló objektumdetekciós módszerek lehetővé teszik, hogy ezek az objektumok valós időben és nagy pontossággal detektálhatók legyenek.

A szakdolgozat célja egy olyan mély neurális hálózat kialakítása és tanítása, amely képes a közlekedési objektumok automatikus felismerésére és azonosítására képek alapján. Emellett az alkalmazás biztonságkritikus jellege miatt a hallgatónak meg kell vizsgálnia a modell magyarázhatósági aspektusait is, különös tekintettel a modell interpretálhatóságára.

A hallgató feladatának a következőkre kell kiterjednie:

- Végezzen irodalomkutatást a konvolúciós mély neurális hálók és a magyarázó technikák témájában!
- Végezzen adatgyűjtést, és készítse elő az adatokat a háló tanításához és validálásához!
- Végezze el a választott objektumdetekcióra használható neurális háló tanítását és értékelje ki annak teljesítményét!
- Alkalmazza a megismert modelfüggő és modelfüggetlen magyarázó módszereket a modell döntéseinek elemzésére!
- Alkalmazzon legalább két magyarázó módszert, és hasonlítsa össze azok hatékonyságát és eredményeit!

Tanszéki konzulens: Dr. Hullám Gábor, docens

Budapest, 2024.09.26.

……..
Dr. Dabóczi Tamás
tanszékvezető, egyetemi tanár

Budapesti Műszaki és Gazdaságtudományi Egyetem
Villamosmérnöki és Informatikai Kar
Mesterséges Intelligencia és Rendszertervezés Tanszék

1117 Budapest, Magyar Tudósok krt. 2. I. ép. I.E.444.
Telefon: 463-2057, Fax: 463-4112
http://www.mit.bme.hu • e-mail: mitadm@mit.bme.hu

Budapest University of Technology and Economics
Faculty of Electrical Engineering and Informatics
Department of Artificial Intelligence and Systems Engineering

Training and Interpretation of a Neural Network Model for Traffic Object Detection

Bachelor's Thesis

Author
Péter Nyilas

Advisor
dr. Gábor Hullám

October 9, 2024

Contents

HALLGATÓI NYILATKOZAT

Alulírott *Nyilas Péter*, szigorló hallgató kijelentem, hogy ezt a szakdolgozatot meg nem engedett segítség nélkül, saját magam készítettem, csak a megadott forrásokat (szakirodalom, eszközök stb.) használtam fel. Minden olyan részt, melyet szó szerint, vagy azonos értelemben, de átfogalmazva más forrásból átvettem, egyértelműen, a forrás megadásával megjelöltem.

Hozzájárulok, hogy a jelen munkám alapadatait (szerző(k), cím, angol és magyar nyelvű tartalmi kivonat, készítés éve, konzulens(ek) neve) a BME VIK nyilvánosan hozzáférhető elektronikus formában, a munka teljes szövegét pedig az egyetem belső hálózatán keresztül (vagy autentikált felhasználók számára) közzétegye. Kijelentem, hogy a benyújtott munka és annak elektronikus verziója megegyezik. Dékáni engedéllyel titkosított diplomatervek esetén a dolgozat szövege csak 3 év eltelte után válik hozzáférhetővé.

Budapest, 2024. október 9.

Nyilas Péter
hallgató

Kivonat

Jelen dokumentum egy diplomaterv sablon, amely formai keretet ad a BME Villamosmérnöki és Informatikai Karán végző hallgatók által elkészítendő szakdolgozatnak és diplomatervnek. A sablon használata opcionális. Ez a sablon LaTeX alapú, a *TeXLive* TeX-implementációval és a PDF-LaTeX fordítóval működőképes.

Abstract

This document is a LaTeX-based skeleton for BSc/MSc theses of students at the Electrical Engineering and Informatics Faculty, Budapest University of Technology and Economics. The usage of this skeleton is optional. It has been tested with the *TeXLive* TeX implementation, and it requires the PDF-LaTeX compiler.

1 Introduction and Motivations

Accurate detection of traffic relevant objects is a critical task in the field of automotive and transportation systems. There can be multiple scenarios where different types of object detection is needed. In order to develop a highly sophisticated automated driving system a very crucial part of it is, to efficiently gather data from sensors located in the chassis and processing them to create a virtual model of the world surrounding our self-driving vehicle. Among the various solutions to this problem, the use of camera sensors as the primary source of information is prevalent. These sensors provide rich visual data, which can be effectively processed using different types of processing algorithms.

The biggest challenge in this field is to develop an algorithm that can accurately detect and classify objects in real-time, while being robust to various environmental conditions. This is where deep learning models come into play. These models have shown remarkable performance in object detection tasks, because of their ability to learn complex patterns , and their capability to generalize well.

However, the biggest drawback of these models is that they are often considered as black boxes, we cannot simply predict their output from the input we gave them. Tha fact that we cannot understand how they came to a certain conclusions, motivated the development(like [**?**]) of various model interpretation techniques. These techniques differ in their approach, but they all aim to provide insights into the decision-making process of the model. The complexity of interpreting models in these domains arises from the intricate nature of the data and the sophisticated algorithms used. Image processing models, for instance, must analyze vast amounts of visual data to detect and classify objects accurately. This complexity makes it challenging to trace the decision-making process, yet it is essential for identifying potential biases, improving model performance, and gaining the trust of users and regulatory bodies. Therefore, enhancing the interpretability of models in these areas is vital for advancing technology while maintaining safety and ethical standards.

2 Objectives

The main objectives of my thesis are as follows:

- To train a neural network for traffic object detection using the YOLOv8 architecture.
- To interpret the model's predictions using external solutions such as LIME, SHAP and EigenCAM.
- To evaluate the performance and interpretability of the model on a real-world dataset.
- To analyze the results and draw conclusions on the effectiveness of the model for traffic object detection.
- To propose future research directions and improvements for the model.

3 The Model, its training and data management

3.1 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are a class of deep learning models that are mostly used for computer vision tasks. Given their ability to learn spatial hierarchies of features and their various types of data representations, they are perfect for the task of traffic object detection.

Characteristic layer types

Convolutional layers

Pooling layers

Fully connected layers

Sampling layers

Activation Functions Activation functions are used in CNNs to introduce non-linearity into the model. They help the model learn complex patterns and make accurate predictions. Some common activation functions include ReLU, Sigmoid, and Tanh.

ReLU

Sigmoid

Tanh

3.2 Yolov8

Introduction to YOLO The YOLO (You Only Look Once) model is a cutting-edge object detection system that has gained a reputation for its speed and accuracy. It processes images in real time and identifies objects within them. One of the latest iteration, YOLOv8, has seen improvements in both performance and efficiency.

Model architecture

Backbone The YOLOv8 model is built on the bases of convolutional neural networks (CNN) that aims to extract the most essential features from the input images. It has been designed with both depth and efficiency in mind, with the intention of capturing intricate details while maintaining high processing speeds.

Neck The neck component of the YOLOv8 architecture serves as a bridge between the backbone and the head. It aggregates features from different layers of the backbone, enhancing the model's ability to detect objects at various scales.

Head The head of the YOLOv8 model is responsible for making the final predictions. It processes the aggregated features from the neck and outputs bounding boxes and class probabilities for detected objects.

3.3 Training

The training process for the YOLOv8 model involves feeding it a large dataset of labeled images. The model learns to identify objects by minimizing the difference between its predictions and the actual labels. Techniques such as data augmentation and regularization are used to improve the model's generalization capabilities.

4 Dataset and formats

4.1 Cityscapes

The Cityscapes dataset is a large-scale dataset used for training and evaluating object detection models. It contains high-resolution images of urban scenes, with detailed annotations for various objects such as cars, pedestrians, and traffic signs. The dataset is widely used in the field of computer vision for tasks such as semantic segmentation and object detection.

4.2 Format conversion and datatypes

Format conversion is a crucial step in preparing the dataset for training. The images and annotations are converted into a format that the YOLOv8 model can process This involves resizing images, normalizing pixel values, and converting annotations into a suitable format.

5 Model interpretation using external solutions

5.1 Importance of Interpretation

Interpreting machine learning models is essential for understanding their decision-making processes. It helps in identifying biases, improving model performance, and building trust with users. Interpretation techniques provide insights into how models make predictions and highlight the most influential features.

5.2 Model agnostic methods

Local Interpretable Model-agnostic Explanations LIME is a popular model-agnostic interpretation method that explains individual predictions by approximating the model locally with an interpretable model. It perturbs the input data and observes the changes in the model's predictions to identify the most important features.

Shapley Additive explanations SHAP is another model-agnostic method that provides consistent and accurate explanations for model predictions. It is based on cooperative game theory and assigns a Shapley value to each feature, representing its contribution to the prediction.

5.3 Model specific methods

EigenCAM EigenCAM is a model-specific interpretation method that visualizes the regions of an image that are most important for a model's prediction. It computes the principal components of the feature maps and highlights the areas that contribute the most to the final decision.

Summary

This document presents a comprehensive study on the training and interpretation of a neural network model for traffic object detection. The key points discussed in the chapters are summarized as follows:

- **Model, Training, and Data**: The YOLOv8 model architecture, including its backbone, neck, and head components, is detailed. The training process and the Cityscapes dataset used for training are also discussed.
- **Model Interpretation Using External Solutions**: The importance of model interpretation is highlighted. Various model-agnostic methods such as LIME and SHAP, as well as model-specific methods like EigenCAM and EigenGradCAM, are explained.

The study concludes that effective training and interpretation of neural network models are crucial for accurate and reliable traffic object detection.

Bibliography

[] Yu Liang, Siguang Li, Chungang Yan, Maozhen Li, and Changjun Jiang. Explaining the black-box model: A survey of local interpretation methods for deep neural networks. *Neurocomputing*, 419:168–182, 2021. ISSN 0925-2312. DOI: https://doi.org/10.1016/j.neucom.2020.08.011. URL https://www.sciencedirect.com/science/article/pii/S0925231220312716.