FACULTY OF APPLIED SCIENCES
UNIVERSITY
OF WEST BOHEMIA

DEPARTMENT OF
COMPUTER SCIENCE
AND ENGINEERING

Master's Thesis

Computer Vision Applications in Video Recordings for Traffic Signal Detection and Classification on Czech Railways

Daniel Schnurpfeil

PILSEN, CZECH REPUBLIC

2025

FACULTY OF APPLIED SCIENCES
UNIVERSITY
OF WEST BOHEMIA

DEPARTMENT OF
COMPUTER SCIENCE
AND ENGINEERING

Master's Thesis

Computer Vision Applications in Video Recordings for Traffic Signal Detection and Classification on Czech Railways

Bc. Daniel Schnurpfeil

Thesis advisor
Ing. Pavel Mautner, Ph.D.

PILSEN, CZECH REPUBLIC **2025**

Citation in the bibliography/reference list:
SCHNURPFEIL, Daniel. *Computer Vision Applications in Video Recordings for Traffic Signal Detection and Classification on Czech Railways*. Pilsen, Czech Republic, 2025. Master's Thesis. University of West Bohemia, Faculty of Applied Sciences, Department of Computer Science and Engineering. Thesis advisor Ing. Pavel Mautner, Ph.D.

ZÁPADOČESKÁ UNIVERZITA V PLZNI
Fakulta aplikovaných věd
Akademický rok: 2024/2025

Studijní program: Informatika a její specializace
Forma studia: Prezenční
Specializace/kombinace: Zpracování přirozeného jazyka (ZPJ18np)

Podklad pro zadání DIPLOMOVÉ práce studenta

Jméno a příjmení:	**Bc. Daniel SCHNURPFEIL**
Osobní číslo:	**A22N0074P**
Adresa:	Luženická 677, Domažlice – Týnské Předměstí, 34401 Domažlice 1, Česká republika
Téma práce:	Využití metod počítačového vidění ve videozáznamech pro detekci a klasifikaci návěstidel na českých železnicích
Téma práce anglicky:	Computer Vision Applications in Video Recordings for Traffic Signal Detection and Classification on Czech Railways
Jazyk práce:	Angličtina
Vedoucí práce:	Ing. Pavel Mautner, Ph.D. Katedra informatiky a výpočetní techniky

Zásady pro vypracování:

- Seznamte se s problematikou návěstidel a návěstních znaků, zejména se zaměřením na jejich vizuální charakteristiky a odlišnosti.
- Prostudujte videa z veřejně dostupných zdrojů (např. YouTube kanál parnici.cz) obsahující železniční návěstidla.
- Navrhněte a implementujte metody pro získání snímků popřípadě sérií snímků návěstidel/návěstních znaků z dostupných videozáznamů.
- Navrhněte metody a implementujte řešení pro detekci a klasifikaci světelných návěstidel, případně návěstních znaků.
- Na dostatečně velké množině dat ověřte funkčnost implementovaných řešení.
- Zhodnoťte a popište dosažené výsledky.

Seznam doporučené literatury:

Dodá vedoucí diplomové práce.

Podpis studenta: Datum:

Podpis vedoucího práce: Datum:

Declaration

I hereby declare that this Master's Thesis is completely my own work and that I used only the cited sources, literature, and other resources. This thesis has not been used to obtain another or the same academic degree.

I acknowledge that my thesis is subject to the rights and obligations arising from Act No. 121/2000 Coll., the Copyright Act as amended, in particular the fact that the University of West Bohemia has the right to conclude a licence agreement for the use of this thesis as a school work pursuant to Section 60(1) of the Copyright Act.

V Plzni, on 14 March 2025

. . . .

Daniel Schnurpfeil

Abstract

English abstract

Abstrakt

Czech abstract

Keywords

computer vision • czech railways

Contents

1 Introduction

Czech Railways 2

todo - tady krátké intro

todo - zmínit evropský zabezpečovací systém, siemens mobility ...

2.1 Situation in Recent Years

todo - tady popsat situaci v čechách a na moravě (slezku)
todo - zmínit - Dopravní a návěstní předpis pro tratě nevybavené evropským vlakovým zabezpečovačem a že to je hlavní zaměření

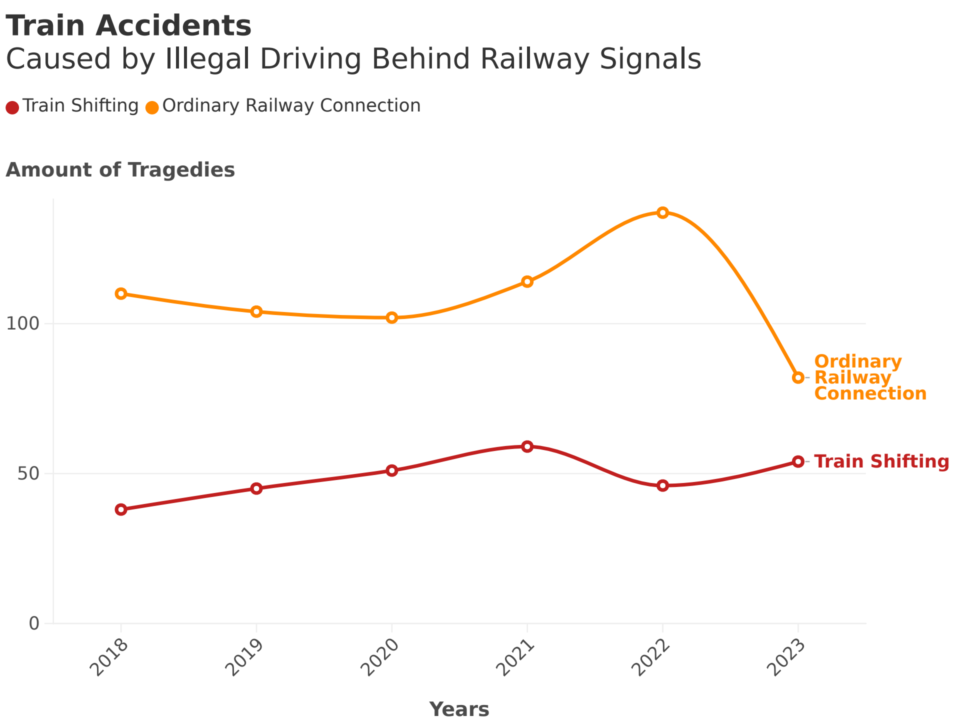

Figure 2.1: Train Accidents Caused by Illegal Driving Behind Railway Signals, source [25]

2.2 Railway Signals

Railway signals represent a visual communication tool for train drivers. Their main purpose is to show important safety information for train driver. These signals contain specific combinations of lights, shapes, and colors to transmit clear instructions about speed limits, track availability, and required actions.

Light signals on Czech railways operate through a system of colored lights mounted on standardized signal posts. The most frequent signal colors are red, green, yellow, and white, with each color that carry distinct meaning. Red lights typically indicate stop request, while green lights could allow unlimited movement. The yellow light serves as a warning sign, preparing drivers for the following limitations. White lights are often in shunting[1] signals or as additional indicators.

The signals combine these colors in various patterns to communicate more complex messages. For example, two vertically positioned yellow lights (Figure 2.2) inform the driver to reduce speed and expect a stop signal ahead. The position and blinking of light(s) adds another piece of information to the basic colors.

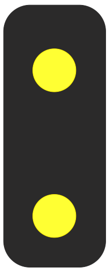

Figure 2.2: Limit 40 km/h and warning

Fixed railway signs complement the light-based system. These include physical signs and markers that display speed limits, distance warnings, and track identification. Their design emphasizes visibility in various weather conditions through reflective materials and high-contrast color schemes. Signal boards often use standardized shapes such as circles or white triangles that serve as warning signs. They are not part of the thesis.

[1]Shunting in railways is the process of moving trains, wagons within a station to assemble, disassemble, or relocate them for operational purposes.

2.2.1 Fore Signal (Předvěst)

In SŽ D1 Část První.[Svo24] regulation, they describe fore signal in czech called ("předvěst"). This signal is dependent on main signal. Typically, this signal adds another notification for the train driver before the driver could see the main signal.

2.2.2 Single-Light Signals

In this section we will describe single-light signals and their characteristics described in the railway signaling regulation SŽ D1 Část První.[Svo24] It is important to mention that described signaling systems on railway tracks are not equipped with the European Train Control System.

There are several distinct single-light signals, each characterized by a specific color and blinking pattern. For example, the "Volno" signal, represented by a steady green light, allows the train to run.

On dependent signals at upstream points fore signals (předvěsti), this signal indicates there is a similar signal on the following main signal. In contrast, signals such as

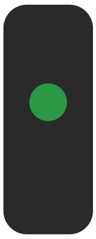

Figure 2.3: Návěst Volno (steady green light)

"Očekávejte rychlost 40 km/h" (slowly flashing yellow light) or "Očekávejte rychlost 100 km/h" (rapidly flashing green light) permit train movement while preparing the driver for a speed restriction at the subsequent signal, typically positioned at least at braking distance. These speed-related signals are from 40 km/h to 120 km/h. they are realized by patterns slow or rapid flashing and colors (yellow or green). **[Space reserved for Figure: Obrázek 103 - Návěst Očekávejte rychlost 100 km/h (rapidly flashing green light)]**

2.2.3 Stop Signal (návěst Stůj)

In the regulation SŽ D1 Část První[Svo24] is described also Stop Signal. It is named "návěst Stůj" in the Czech railway signaling system, and it is a single red light on the main signal devices. This signal is the most significant safety mechanism in the railway infrastructure. Based on this signal, the train driver has to stop the locomotive approximately 10 meters in front of the signal device when displaying the

Figure 2.4: Návěst Očekávejte rychlost 40 km/h (slowly flashing yellow light)

Stop Signal. This signal is also used for shunting operations or special maintenance vehicles. In situations where the main signal is not positioned directly adjacent to the track, the train must stop before reaching the end of the train path indicator. There are two types of stop signal. The first is absolute and the second is permissive

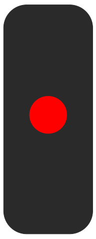

Figure 2.5: Návěst Stůj (singe red light)

type. The absolute signal means that when the red changes to something different that allows the train to move, the train driver can continue. The second permissive type allows the train to continue in certain cases that are described in SŽ D1 Část První[Svo24] with more details.

2.2.4 Multiple-Light Signals

The multiple-light signaling architecture uses a vertical light elements where each light combination mean specific state. This system follows a structured logic in which:

1. The lower light element adjust speed restrictions
2. The upper light element has a similar predictive function as is described in section 2.2.1 Fore Signal.

The semantics of this system are described through color coding and flashing patterns. The signal states are:

- Static yellow lights - indicating speed restrictions
- Flashing yellow lights - communicating anticipatory speed reductions
- Green lights (static and flashing) - permitting higher speed
- Numerical indicators shows precise speed thresholds

There are also horizontal illuminated bars that correspond to specific speed thresholds. For example, a horizontal yellow bar indicates a 60 km/h restriction, while a horizontal green bar corresponds to an 80 km/h threshold.

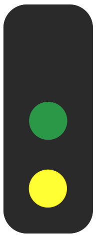

Figure 2.6: Limit 60 km/h and go

The SŽ D1 Část První [Svo24] further describes the relationship between the current and anticipated signal states. That is particularly useful when:

- The predicted speed restriction is more important than the current one (the need to slow down)
- The predicted speed restriction is higher than the current restriction (the need to speed up)

2.2.4.1 Repeater Signals

There are also special states that repeat signals within the Czech railway infrastructure. These repeater signals perform a function inform the locomotive driver about speed limits imposed by subsequent (aditional white) signals located at not very good stopping distances.

todo dopsat

2.2.5 Convolutional Neural Network

The convolutional neural network is designed to process multidimensional data [YBH15]. Such data are, for example, color images, which can be represented by, for example, three two-dimensional arrays containing pixel intensities in three color channels (red, green and blue).

In the simplest terms, convolution is a mathematical operation that in our case is used to modify an image to, for example, an image where the edges are highlighted, which is important for the objects we are looking for there.[Sch22]

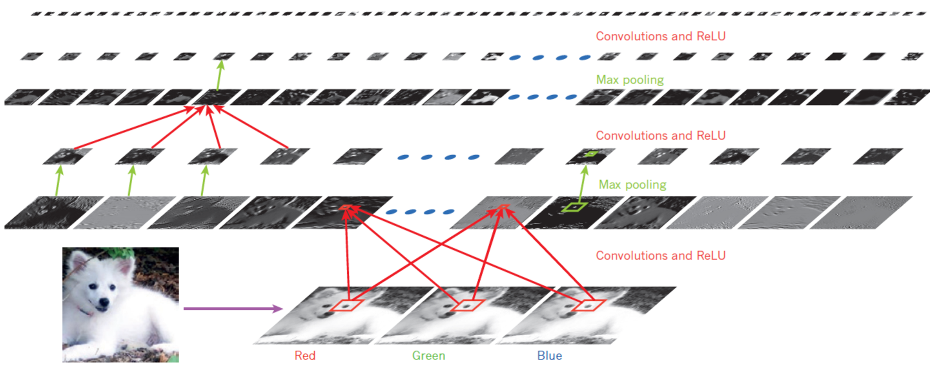

Figure 2.7: CNN Layers (picture is taken from [YBH15])

2.2.5.1 Convolution

Now, let us look at the convolution in more detail. As an example, consider a black and white image represented by the matrix shown on the left in the image 2.8. The values in the matrix represent the brightness intensities of the pixels. Next, we have the so-called kernel matrix (the convolution mask in Figure 2.8). Both matrices are then processed in the following function:

$$V_{i,j} = (M, N)_{i,j} = \sum_{a=-k}^{k} \sum_{b=-k}^{k} M(i-a, j-b) \cdot N(a, b), \tag{2.1}$$

where $V_{i,j}$ is the resulting pixel value at the position of indices i and j, M is the area in the V matrix, and N is the kernel matrix of k rows and k columns. The convolution can be seen in the following figure 2.8. It is clear in the 2.8 figure that for a kernel

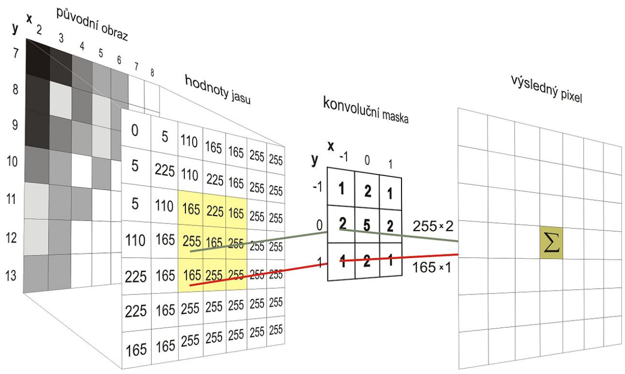

Figure 2.8: Convolution Example (picture is taken from [06])

matrix of three columns and three rows, that each index value is multiplied by the index value of a given region of the matrix with the same dimensions as the kernel matrix.

2.2.5.2 **Pooling**

This method is similar to the convolution described in the previous paragraphs of the 2.2.5.1 section. The method traverses the image matrix by regions and calculates just one pixel using a defined function, for example, by averaging the values in the given regions (average pooling) or by calculating the maximum value in each subregion (max pooling).[Sch22]

2.2.5.3 **CNN Architecture**

The structure of a convolutional neural network is composed of two parts. The first is the part that processes the input image. This part consists of convolutional and pooling layers. The result of this part is a vector of features that is used as input to the layered neural network.

2.2.5.4 **You Only Look Once**

YOLO (You Only Look Once) represents a great improvement in object detection approaches, reformulating the detection problem as regression rather than classification. Unlike traditional methods that employ sliding windows or region proposals, YOLO processes the entire image in a single evaluation, predicting bounding boxes and class probabilities simultaneously through a unified neural network architecture [**yolo**]. The system divides the input images into an S×S grid, where each grid cell predicts N bounding boxes with associated confidence scores and class probabilities. These predictions are represented as an S×S×(N*5+C) tensor, where C is the number of classes. The network architecture consists of 24 convolutional layers described in previous sections. They are followed by 2 fully connected layers that take inspiration from the GoogLeNet model but 1×1 reduction layers followed by 3×3 convolutional layers. During inference, YOLO processes images at 448×448 resolution, generating approximately 98 bounding boxes per image. This unified approach is able end-to-end optimization directly on detection performance, the result is extremely fast processing that speeds up base YOLO model operates at 45 frames per second while a faster variant reaches 155 frames per second. This real-time performance, combined with YOLO's ability to detect objects in images, makes it particularly suitable for applications that require robust object detection. So, that is the reason why YOLO is chosen as the main tool in this thesis.

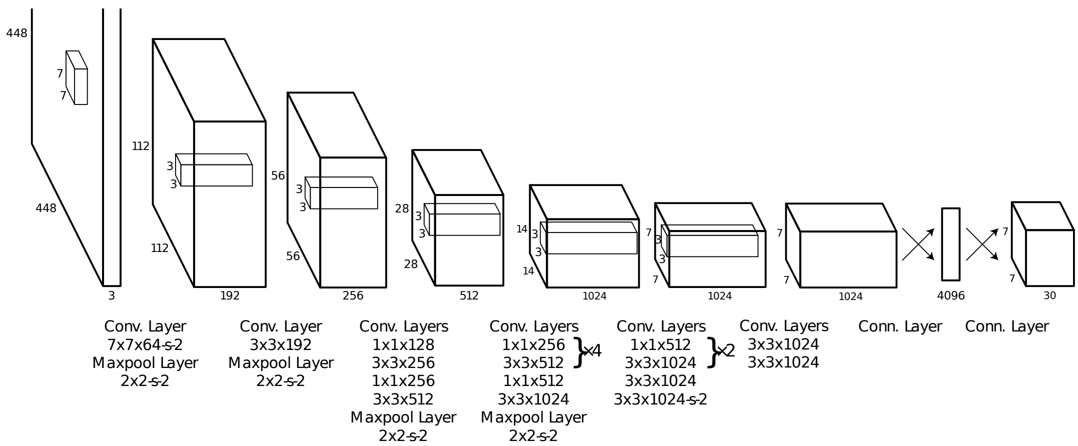

Figure 2.9: One of the first YOLO architectures (Diagram is taken from [Red+16])

2.2.5.5 Real-Time DEtection TRansformer

todo The architectural design of RT-DETR strategically balances computational efficiency and detection accuracy through two principal innovations. First, the model implements an efficient hybrid encoder that decouples intra-scale interaction and cross-scale fusion processes, significantly reducing computational overhead compared to traditional Transformer encoders. This hybrid design incorporates an Attention-based Intra-scale Feature Interaction (AIFI) module that selectively applies self-attention operations to high-level features with rich semantic content, while eliminating redundant processing of lower-level features.

Complementing this encoder architecture, RT-DETR introduces an uncertainty-minimal query selection mechanism that optimizes object query initialization. Unlike previous approaches that rely solely on classification confidence, this novel selection process explicitly models the joint latent variable of classification and localization confidence, minimizing uncertainty in the selected features and providing higher-quality initial queries to the decoder.

The empirical evaluation demonstrates RT-DETR's superior performance characteristics. The RT-DETR-R50 variant achieves 53.1% Average Precision (AP) on the COCO validation dataset while processing 108 frames per second (FPS) on an NVIDIA T4 GPU. The RT-DETR-R101 configuration reaches 54.3% AP at 74 FPS. These metrics surpass all comparable YOLO detectors in both accuracy and computational efficiency. Furthermore, RT-DETR offers practical advantages through flexible speed tuning capabilities, allowing adaptation to various deployment scenarios without retraining by adjusting decoder layer configuration.

Perhaps most significantly, RT-DETR eliminates the need for Non-Maximum Suppression (NMS) post-processing, which introduces computational bottlenecks and hyperparameter dependencies in traditional detection frameworks. This end-to-end approach provides more stable and consistent performance while simultaneously improving inference speed.

3 State of The Art

this is related[Sta+22]

Data Analysis & Methodology

4

4.1 Data Resources

4.2 ETL

Study of publicly available sources (e.g., YouTube channel parnici.cz) Methods for extracting individual frames and image sequences … [Lin+15]

4.2.1 Data Annotation

4.2.1.1 YOLO

`Limitations`

4.2.1.2 Heuristics

4.2.1.3 Data Transformation

4.2.1.4 Datat Load

4.3 Region of Interest

Proposed methods for identifying light signals in images
- enlarge bounding box (ROI) from YOLO detections

Figure 4.1: Original detection example (figure is from [Svo24])

Techniques for recognizing specific signal aspects

Implementation 5

Details of the implemented solution

5.1 Dataset Storage

5.2 Experiment Playground

5.3 Training Scripts

Technologies and libraries used

5.4 Applied Technologies

5.4.1 Ultralytics Yolo

5.4.2 Open CV

Challenges encountered and solutions applied

5.4.3 Czech Metacenter

Results 6

Description of the testing process

6.1 Train Dataset

6.2 Eval Dataset

6.3 Test Dataset

parnici cz a strojvedouci .com

Process of compiling a comprehensive dataset for testing
Presentation of results
Analysis of system performance

6.4 Signal Detection

6.4.0.1 Baseline

6.5 Signal Classification

6.5.0.1 Baseline

6.6 Signal Recognition

Signal Detection

+

Signal Classification

6.6.0.1 Baseline

Discussion

7

Interpretation of results Comparison with existing methods Limitations of the current approach

Conclusion 8

Summary of achievements Contributions to the field Suggestions for future work

Bibliography

[Sch22] SCHNURPFEIL, Daniel. *Architectural Style Classification on Snapshots of Buildings*. ZCU, 2022. Available also from: `https://naos-be.zcu.cz/server/api/core/bitstreams/b699dd9d-fb36-436b-a9d4-1bdb7d6e13c3/content`.

[YBH15] YANN; BENGIO, Yoshua; HINTON, Geoffrey. *Deep learning*. cs.toronto.edu, 2015. Available also from: `https://www.cs.toronto.edu/~hinton/absps/NatureDeepReview.pdf`.

[06] wikimedia.org, 2006. Available also from: `https://upload.wikimedia.org/wikipedia/commons/thumb/c/c5/Konvoluce_2rozm_diskretni.jpg/1280px-Konvoluce_2rozm_diskretni.jpg`.

[Lin+15] LIN, Tsung-Yi et al. *Microsoft COCO: Common Objects in Context*. 2015. Available from arXiv: `1405.0312 [cs.CV]`.

[Red+16] REDMON, Joseph; DIVVALA, Santosh; GIRSHICK, Ross; FARHADI, Ali. *You Only Look Once: Unified, Real-Time Object Detection*. 2016. Available from arXiv: `1506.02640 [cs.CV]`.

[Sta+22] STAINO, Andrea; SUWALKA, Akshat; MITRA, Pabitra; BASU, Biswajit. Real-Time Detection and Recognition of Railway Traffic Signals Using Deep Learning. *Journal of Big Data Analytics in Transportation*. 2022, vol. 4, pp. 57–71. Available from DOI: `10.1007/s42421-022-00054-7`.

[25] *Statistiky mimořádných událostí | Drážní inspekce* [Online]. 2025. Available also from: `https://di.gov.cz/mimoradne-udalosti/statistiky-mimoradnych-udalosti`.

[Svo24] SVOBODA, Jiří. *Dopravní a návěstní předpis pro tratě nevybavené evropským vlakovým zabezpečovačem* [Online]. 2024. Available also from: `https://provoz.spravazeleznic.cz/Portal/Show.aspx?oid=2179979`.

List of Figures

List of Tables

List of Listings