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 21 March 2025

. . . .

Daniel Schnurpfeil

Abstract

English abstract

Abstrakt

Czech abstract

Keywords

computer vision • czech railways

Contents

1 Introduction **3**

2 Czech Railways **4**
2.1 Railway Signals 5
2.1.1 Fore Signal (Předvěst) 6
2.1.2 Single-Light Signals 6
2.1.3 Stop Signal (návěst Stůj) 6
2.1.4 Multiple-Light Signals 7
2.1.5 Convolutional Neural Network 9

3 Data Analysis & Methodology **14**
3.1 Data Resources 14
3.1.1 Dataset Characteristics and Technical Parameters 14
3.1.2 Study of Publicly Available Sources 16
3.1.3 Integrated Dataset Analysis 17
3.2 ETL 18
3.2.1 Data Annotation 18
3.3 Region of Interest 19

4 State of The Art **20**

5 Implementation **21**
5.1 Dataset Storage 21
5.2 Experiment Playground 21
5.3 Training Scripts 21
5.4 Applied Technologies 21
5.4.1 Ultralytics Yolo 21
5.4.2 Open CV 21
5.4.3 Czech Metacenter 21

6 Results **22**
6.1 Signal Detection 22
6.2 Signal Classification 22
6.3 Signal Recognition 22

7 Discussion **23**

8 Conclusion **24**

Bibliography **25**

List of Figures **27**

List of Tables **28**

List of Listings **29**

Introduction

1

Background on Railway Signaling Systems
Thesis objectives and scope

Czech Railways 2

Railway transport systems through Czechia have been improved in recent years with respect to the modernization of the safety infrastructure. The Czech railway network is approximately 9,500 kilometers[19] long and has many signaling technologies, from legacy mechanical systems to the modern European Train Control System (ETCS).

Statistical analyses of railway incidents within the Czech railway system reveal a lot of signal violations. In visualization, train accidents caused by illegal driving behind railway signals in 2018-2024 have demonstrated a high number, with annual incidents between 136 and 183 cases[25]. The temporal distribution analysis shows

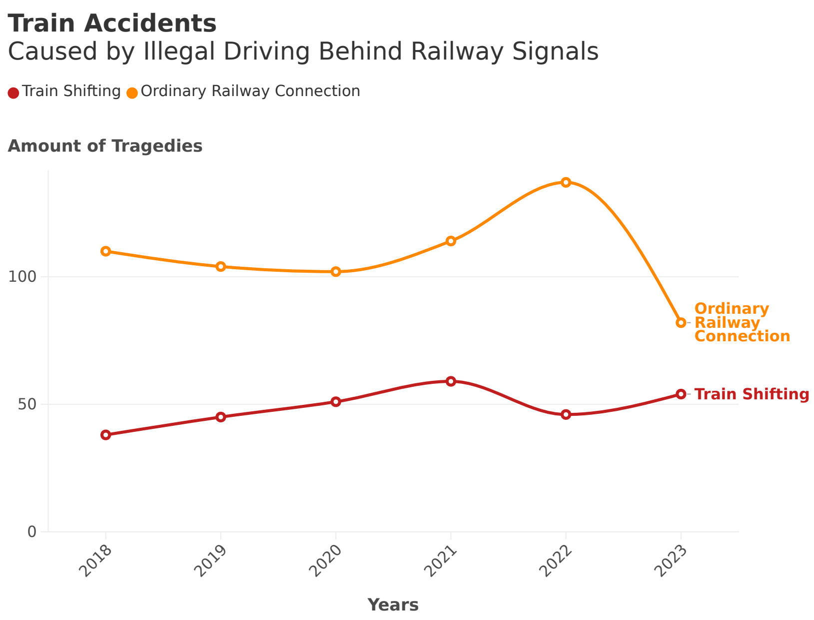

Figure 2.1: Train Accidents Caused by Illegal Driving Behind Railway Signals, data are taken from [25]

that a significant number of incidents frequently occur during the summer months and transitional seasonal periods. The data indicates a bifurcation between violations occurring during regular train operations (vlak) versus shunting operations (posun), with train operations consistently representing approximately 68-75% of total incidents during the measured period.

A critical factor that influences railway safety in Czechia is the heterogeneous nature of signaling infrastructure. Although the main railways are modernized with contemporary European standards, regional and secondary lines continue to operate without advanced train protection systems. This can lead to difficulties for driver adaptation across a lot of infrastructure environments.

The Czech national *Traffic and Signaling Regulations for Lines Not Equipped with European Train Protection*[Svo24] sets up the operational characteristics of rail transport. This regulatory framework defines both visual signal characteristics and operational responses, creating context for computer vision implementation.

2.1 Railway Signals

Railway signals represent a visual communication tool for train drivers. These signals contain specific combinations of lights, shapes, and colors to transmit instructions about speed limits, track availability, and required actions.

The 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. Each of the color or color combinations has a 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 2.3) inform the driver to reduce speed and expect a stop signal ahead. The position and blinking of the light(s) add another piece of information to the basic colors.

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, which 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.

2.1.1 Fore Signal (Předvěst)

In the *Traffic and Signaling Regulations for Lines Not Equipped with European Train Protection*[Svo24], they describe the fore signal in Czech called (”předvěst”). This signal depends on the main signal. Typically, this signal adds another notification for the train driver before the driver could see the main signal. They are typically smaller than regular single-light signals and multiple-light signals.

2.1.2 Single-Light Signals

In this section we will describe single-light signals and their characteristics described in the *Traffic and Signaling Regulations for Lines Not Equipped with European Train Protection*.[Svo24]

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.

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

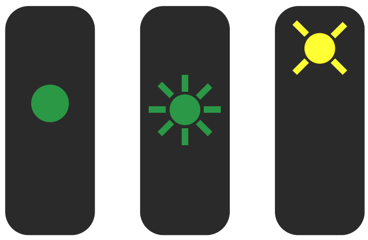

Figure 2.2: "Návěst Volno" (steady green light),
"Návěst Očekávejte rychlost 100 km/h" (rapidly flashing green light),
"Návěst Očekávejte rychlost 40 km/h" (slowly flashing yellow light)

In contrast, signals such as ”Očekávejte rychlost 40 km/h” (**slowly flashing yellow light**) or ”Očekávejte rychlost 100 km/h” (**rapidly flashing green light**) allow the movement of the train 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)**.

2.1.3 Stop Signal (návěst Stůj)

In the regulation *Traffic and Signaling Regulations for Lines Not Equipped with European Train Protection*[Svo24] are also described Stop Signal. It is named "návěst Stůj" in the Czech railway signaling system and 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 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 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 *Traffic and Signaling Regulations for Lines Not Equipped with European Train Protection*[Svo24] with more details.

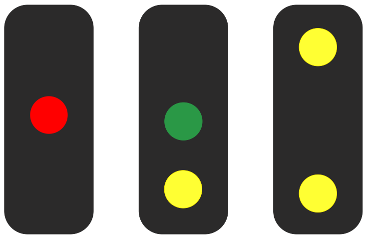

Figure 2.3: "Návěst Stůj" (singe red light), Limit 60 km/h and go ,
Limit 40 km/h and warning

2.1.4 Multiple-Light Signals

The multiple-light signaling architecture uses vertical light elements where each light combination means a 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 2.1.1 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** - indicating predict next 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.

The *Traffic and Signaling Regulations for Lines Not Equipped with European Train Protection* [Svo24] further describe 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)

2.1.4.1 Repeater Signals

There are also special states that repeat signals within the Czech railway infrastructure. These repeater signals perform a function to inform the locomotive driver about speed limits imposed by subsequent (additional white) signals located at not very good stopping distances. The architectural realization of these repeater signals is based on a multi-illumination system. The repeater signals are categorized into a few types. The first of them is **White light with yellow light above** that permits the movement of the train and the maximum track speed until the next main signal, but also indicates the next "Stop" signal at an insufficient braking distance. The second type is family of **Speed-Specific Repeater Signals**:

- **Expect Speed 40 km/h** - the white light with the slowly flashing yellow light above indicates next speed restriction of 40 km/h, 30 km/h, or 50 km/h
- **Expect Speed 60 km/h** - White light with the rapidly flashing yellow light above indicates next speed restriction of 60 km/h or 70 km/h
- **Expect Speed 80 km/h** - White light with the slowly flashing green light above indicates next speed restriction of 80 km/h
- **Expect Speed 100 km/h** - White light with the rapidly flashing green light above indicates next speed restriction of 100 km/h
- **Expect Speed 120 km/h** - White light with rapidly flashing green light and illuminated yellow "12" numeral above indicates next speed restriction of 120 km/h

- **Speed 40 km/h with Various Expected Speeds** - Yellow light, white light above, with varying signal patterns at top indicates immediate 40 km/h restriction and next speed restrictions depending on top light configuration which was described in 2.1.4 section.

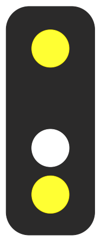

Figure 2.4: "Rychlost 40 km/h a opakování návěsti Výstraha" - Speed 40 km/h and repeating the signal Warning

2.1.5 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]

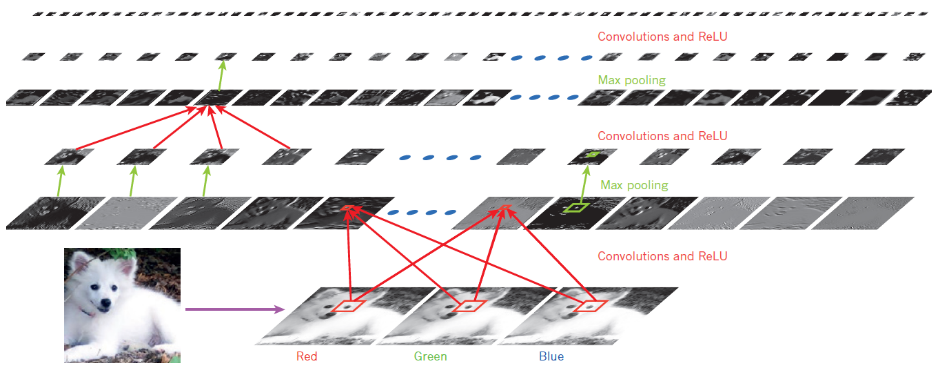

Figure 2.5: CNN Layers (picture is taken from [YBH15])

2.1.5.1 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 2.6. 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 2.6). 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 2.6. It is clear in the figure 2.6 that for a kernel

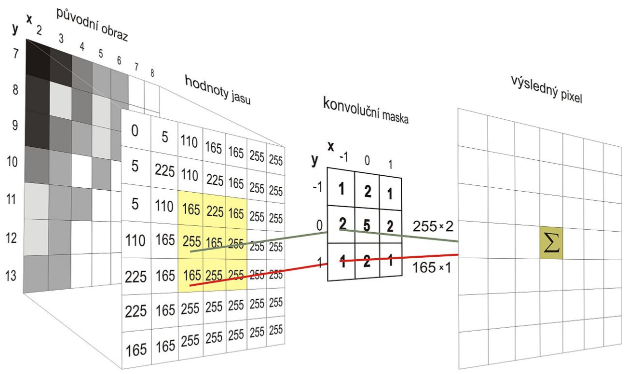

Figure 2.6: Convolution Example (picture is taken from [06])

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

2.1.5.2 Pooling

This method is similar to the convolution described in the previous paragraphs of the 2.1.5.1 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]

2.1.5.3 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.

2.1.5.4 You Only Look Once

YOLO (You Only Look Once) introduced in [**yolo**] is the advanced Convolutional Neural Network variant. Unlike traditional methods that use sliding windows or region suggestions, YOLO processes the entire image in a single evaluation and predicts bounding boxes and class probabilities simultaneously through a unified neural network architecture [**yolo**]. The system partitions 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

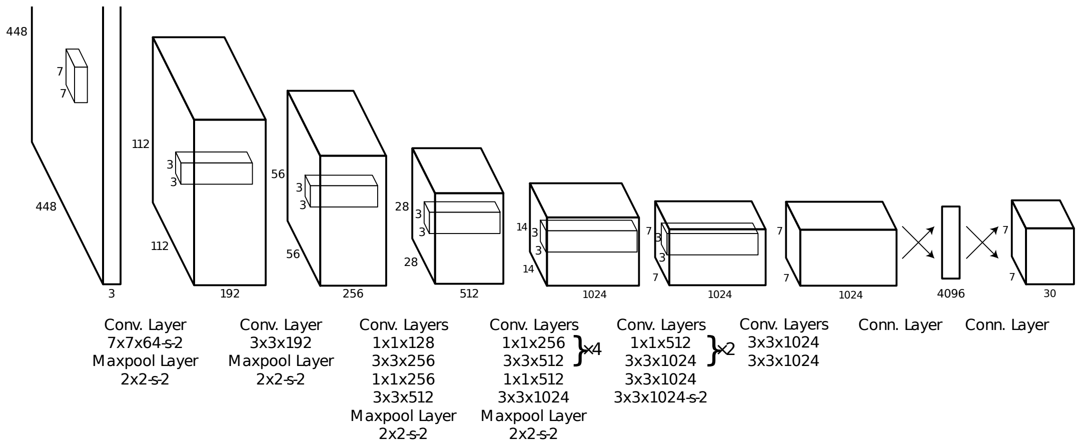

Figure 2.7: One of the first YOLO architectures (Diagram is taken from [Red+16])

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.

2.1.5.5 Real-Time DEtection TRansformer

The architecture of the Real-Time DEtection TRansformer (RT-DETR) comes from the End-to-End Object Detection with Transformers paper[Car+20] publicized by Facebook AI research.

The Detection Transformer uses an encoder-decoder architecture, which was originally made for natural language processing tasks. In the diagram 2.8 there is its architecture. The Detection Transformer has a convolutional backbone to extract visual features from the image, which are then passed to the transformer encoder. The encoder processes these features and creates a representation of the image that is then passed to the decoder. The decoder then uses a set of learned object queries to predict the classes and positions (bounding boxes) of objects in the image. This architecture is not quite optimal for real-time use cases due to its computational complexity and relatively high inference latency. The transformer architecture, while powerful for modeling global interactions between image elements, involves multiple attention mechanisms that require significant computational resources. The self attention operations scale quadratically with input size, which becomes problematic for high-resolution images or video frames.

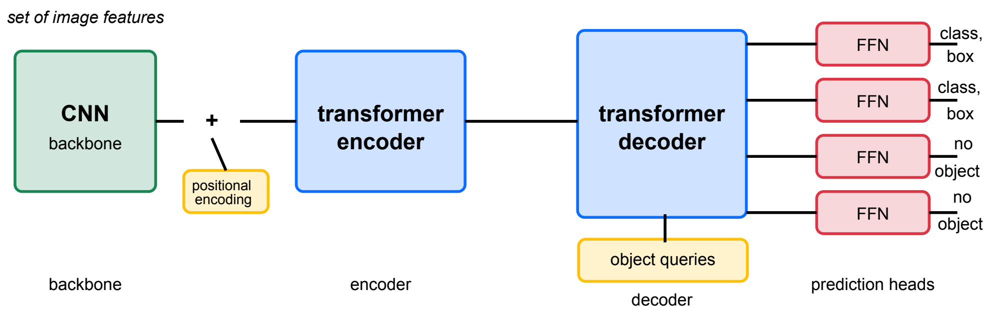

Figure 2.8: Simple diagram of the Detection Transformer architecture

The researchers from China[Lv+23] did a few things for optimization and created the Real-Time DEtection TRansformer that includes an uncertainty-minimal query selection mechanism that optimizes object query initialization. Unlike previous approaches that rely on classification confidence, this process explicitly performs joint variable classification and localization confidence, which minimizes the uncertainty of the selected features and makes the initial queries of higher quality to the decoder.

The researchers from China[Lv+23] also compared their models with the detectors YOLOv5, YOLOv6, YOLOv7 and YOLOv8. Their RT-DETR-R50 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 outperform some of the comparable YOLO detectors in both accuracy and computational efficiency. In addition, RT-DETR offers practical advantages with flexible rate tuning options that allow adaptation to different deployment scenarios without the need for retraining by adjusting the decoding layer configuration.

Most significantly, RT-DETR eliminates the need for Non-Maximum Suppression (NMS) post-processing[2] that has some issues for hyperparameter dependencies in traditional detection frameworks.

[2] Non-Maximum Suppression (NMS) post-processing is technuque of deleting duplicities in detected regions of interest.

Data Analysis & Methodology

3

This chapter examines the exploitative data analysis and methodological approaches used in the development of computer vision systems for the detection and classification of railway signals in the Czech railway infrastructure. The methodology introduces a robust pipeline for data extraction, pre-processing, feature extraction, and model development, with particular emphasis on addressing domain-specific challenges inherent to railway visual signal interpretation.

3.1 Data Resources

This section presents an exploitative data analysis of publicly available video resources to train machine learning models in the detection of railway traffic signals. The resources are from YouTube channels:

- https://www.youtube.com/@parnicicz4773
- http://www.youtube.com/@strojvedouci_com

3.1.1 Dataset Characteristics and Technical Parameters

The corpus consists of visual data captured at 60 frames per second with full HD resolution (1920×1080 pixels). The corpus has an optimal spatial-temporal density for feature extraction in signal detection algorithms. This technical configuration uses edge detection and color segmentation processes that are useful for railway signal classification tasks, while at the same time enabling optical flow computation for motion analysis. The corpus is suitable for both traditional computer vision pipelines and deep learning architectures, such as convolutional neural network training through standardized input dimensionality or Real-Time DEtection TRansformer.

Following list consists of notes and explanations for the following tables:

1. Video title (simplified for readability)
2. <-> – Duration in `hours:minutes` format
3. **Avg** – Average brightness
4. **Var** – Temporal variation (how much the brightness changes over time)

Table 3.1: Train Video Analytics Data

Video Title	<->	**Avg**	**Var**
Cabview z vlaku 3 Rosice n L-Vojtěchov	0:44	117.91	9.17
Cabview z vlaku 2 Žďárec u Skutče-Slatiňany	0:20	117.39	25.31
Cabview z vlaku 6 Pardubice-Hlinsko	1:05	114.53	18.39
Cabview z vlaku 7 Holice-Chrudim	0:46	108.69	18.92
Cabview z vlaku 8 Čáslav-Třemošnice	1:02	108.18	17.20
Cabview z vlaku 14 Pardubice-Holice-Týniště	1:52	96.24	14.69
Cabview 16 Pardubice-Týniště n O-Hradec Králové	1:30	84.46	12.89
Cabview 21 RegioFox Havlíčkův Brod-Pardubice	1:50	94.77	21.60
Cabview 15 Hradec Králové-Choceň-Moravany	1:31	94.26	19.43
Cabview 19 Pardubice-Polička duben 2023	1:45	121.70	12.74
Cabview 17 Chrudim-Havlíčkův Brod-Chrudim	3:06	107.52	23.19
Cabview 18 Pardubice-Polička trať 238 261	1:46	94.59	16.34

Table 3.2: Parnici

Video Title	<->	**Avg**	**Var**
4K Nýřany - Dioss 3100 Kafemlejnek	0:11	96.60	10.61
4K Týn nad Vltavou - Číčenice 810 RegioMouse	0:32	97.62	13.15
4K Radnice - Plzeň 842 Kvatro	0:44	97.60	20.60
4K Plzeň - Plasy Regioshark 844	0:42	104.65	12.23
4K Bezdružice - Plzeň 842 Kvatro	1:03	86.76	14.75
4K Týn n Vltavou - Protivín 814 RegioNova	0:47	130.20	10.27
4K Broumov - Meziměstí 854 Hydra + 954	0:15	109.53	4.75
4K Kolín - Rataje n Sázavou 814 RegioNova	0:53	104.89	15.14
4K Ledečko - Zruč nad Sázavou 814 RegioNova	0:48	94.43	18.68
4K Zruč nad Sázavou - Kutná Hora 814 RegioNova	1:10	111.75	12.20
4K Čáslav - Třemošnice 810 Šukafon	0:28	120.79	7.45
4K Chornice - Velké Opatovice - Skalice	1:57	99.19	9.51
4K Příkosice - Rokycany 814 RegioNova	0:21	121.70	3.74
4K Rokycany - Příkosice 814 RegioNova	0:23	125.55	6.83

Table 3.3: Outliers

Video Title	<->	Avg	Var	FPS
Cabview z vlaku 5 Havlíčkův Brod-Chotěboř	0:20	113.15	8.29	≈ 50
Cabview z vlaku 13 první sníh Pardubice-Hlinsko	2:13	91.72	41.28	30
Cabview 20 Pardubice-Slatiňany-Pardubice	0:51	96.96	25.04	≈ 24

3.1.2 Study of Publicly Available Sources

3.1.2.1 Cabview Dataset

The Cabview dataset constitutes a substantial corpus of railway footage from the Czech transportation network. Statistical analysis reveals the following technical characteristics pertinent to computer vision applications:

- **Temporal Extent:** The collection comprises 12 videos with a cumulative duration of 17 hours 18 minutes 29 seconds, with individual videos ranging from 20 minutes 24 seconds to 3 hours 6 minutes 15 seconds (mean duration: 1 hour 26 minutes 32 seconds).
- **Luminance Distribution:** Videos exhibit mean brightness values ranging from 84.46 to 121.70 (scale 0-255), with an average brightness of 105.02 (σ = 11.35) across the dataset, providing diverse illumination conditions for model training.
- **Temporal Variation:** The dataset demonstrates considerable variation in frame-to-frame brightness fluctuation ($\mu = 17.49$, $\sigma = 4.47$), with values ranging from 9.17 to 25.31, necessitating robust image stabilization preprocessing for consistent feature detection.
- **Information Entropy:** Histogram entropy measurements indicate moderate to high information content ($\mu = 7.63$, $\sigma = 0.12$), with values ranging from 7.40 to 7.78, suggesting substantial visual complexity beneficial for feature extraction tasks.

3.1.2.2 Parnici Dataset

The Parnici dataset offers complementary content with distinct technical properties:

- **Temporal Characteristics:** The collection consists of 14 videos with a total duration of 10 hours 10 minutes 55 seconds, ranging from 10 minutes 37 seconds to 1 hour 56 minutes 46 seconds (mean duration: 43 minutes 38 seconds).

- **Resolution Enhancement:** Videos are predominantly captured in 4K resolution, providing superior spatial detail for signal detection at greater distances, particularly beneficial for early detection systems.
- **Luminance Metrics:** The dataset exhibits mean brightness values from 86.76 to 130.20, with an average of 107.23 ($\sigma = 12.65$), offering slightly higher overall luminance compared to the Cabview corpus.
- **Stability Characteristics:** Temporal variation measurements ($\mu = 11.42$, $\sigma = 4.72$) indicate generally more stable recording conditions than the Cabview dataset, with values ranging from 3.74 to 20.60.
- **Information Content:** Histogram entropy values ($\mu = 7.66$, $\sigma = 0.10$) demonstrate consistent visual complexity across the dataset, with measurements ranging from 7.41 to 7.80.

3.1.3 Integrated Dataset Analysis

The combined datasets present complementary characteristics for machine learning model development:

- **Temporal Coverage:** The aggregated corpus comprises approximately 27 hours 29 minutes of railway operation footage, providing substantial data volume for model training and validation procedures.
- **Environmental Diversity:** The integrated dataset spans diverse illumination conditions (84.46-130.20 brightness range) and temporal dynamics (3.74-25.31 variation range, excluding outliers), facilitating robust model generalization across varying operational scenarios.
- **Frame Extraction Potential:** At the standard 60fps acquisition rate, the combined dataset yields approximately 5.94 million individual frames, presenting a substantial pool for strategic sampling and annotation protocols.
- **Anomalous Cases:** Three identified outlier videos exhibit atypical characteristics, including extreme brightness variation (41.28) and non-standard frame rates (24-50fps), providing valuable edge cases for model robustness evaluation.

Both data sources are suitable for machine learning. The Cabview videos has an extended duration and higher temporal variation, while the Parnici dataset provides superior resolution and more consistent recording conditions.

3.2 ETL

leave blank

3.2.1 Data Annotation

leave blank

3.2.1.1 YOLO

`Limitations`

3.2.1.2 Heuristics

3.2.1.3 Data Transformation

3.2.1.4 Datat Load

3.3 Region of Interest

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

Figure 3.1: Original detection example (figure is from [Svo24])

Techniques for recognizing specific signal aspects

4 State of The Art

this is related[Sta+22]

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
- parnici cz a strojvedouci .com
- Process of compiling a comprehensive dataset for testing
- Presentation of results
- Analysis of system performance

6.1 Signal Detection

6.1.0.1 Baseline

6.2 Signal Classification

6.2.0.1 Baseline

6.3 Signal Recognition

Signal Detection
- +
- Signal Classification

6.3.0.1 Baseline

7 Discussion

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.

[Car+20] CARION, Nicolas et al. *End-to-End Object Detection with Transformers*. 2020. Available from arXiv: 2005.12872 [cs.CV].

[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.

[19] *Europe: rail network density by country size 2019*. 2019. Available also from: https://www.statista.com/statistics/1243209/europe-rail-network-density-per-country-size/.

[Lv+23] LV, Wenyu et al. *DETRs Beat YOLOs on Real-time Object Detection*. 2023. Available from arXiv: 2304.08069 [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

2.1 Train Accidents Caused by Illegal Driving Behind Railway Signals, data are taken from [25] 4
2.2 "Návěst Volno" (steady green light),
"Návěst Očekávejte rychlost 100 km/h" (rapidly flashing green light),
"Návěst Očekávejte rychlost 40 km/h" (slowly flashing yellow light) . . 6
2.3 "Návěst Stůj" (singe red light), Limit 60 km/h and go ,
Limit 40 km/h and warning 7
2.4 "Rychlost 40 km/h a opakování návěsti Výstraha" - Speed 40 km/h and repeating the signal Warning 9
2.5 CNN Layers (picture is taken from [YBH15]) 9
2.6 Convolution Example (picture is taken from [06]) 10
2.7 One of the first YOLO architectures (Diagram is taken from [Red+16]) 11
2.8 Simple diagram of the Detection Transformer architecture 12

3.1 Original detection example (figure is from [Svo24]) 19

List of Tables

3.1 Train Video Analytics Data 15
3.2 Parnici 15
3.3 Outliers 16

List of Listings