MASTER THESIS

Thesis submitted in fulfillment of the requirements for the degree of Master of Science in Engineering at the University of Applied Sciences Technikum Wien - Degree Program Data Science

Multi-sensor rail track detection in automatic train operations

By: Attila Kovacs

Student Number: 2110854031

Supervisors: Lukas Rohatsch
Daniele Capriotti

Wien, January 28, 2024

Declaration

“As author and creator of this work to hand, I confirm with my signature knowledge of the relevant copyright regulations governed by higher education acts (see Urheberrechtsgesetz /Austrian copyright law as amended as well as the Statute on Studies Act Provisions / Examination Regulations of the UAS Technikum Wien as amended).

I hereby declare that I completed the present work independently and that any ideas, whether written by others or by myself, have been fully sourced and referenced. I am aware of any consequences I may face on the part of the degree program director if there should be evidence of missing autonomy and independence or evidence of any intent to fraudulently achieve a pass mark for this work (see Statute on Studies Act Provisions / Examination Regulations of the UAS Technikum Wien as amended).

I further declare that up to this date I have not published the work to hand nor have I presented it to another examination board in the same or similar form. I affirm that the version submitted matches the version in the upload tool.“

Wien, January 28, 2024 Signature

Kurzfassung

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Schlagworte: Deep Learning, Computer Vision, Segmentation, Automatic Train Operations

Abstract

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Keywords: Deep Learning, Computer Vision, Segmentation, Automatic Train Operations

Contents

1 Introduction

According to the International Energy Agency (IEA), the global demand for passenger and freight transportation will more than double by 2050 compared to 2019 (International Energy Agency, 2019). However, a greater demand entails higher energy consumption as well as increased CO2 emissions and atmospheric pollutants. Given the fact that railway is one of the most efficient and reliable modes of transportation there seems to be consensus between politicians and researchers that a greater reliance of rail has the potential to counterbalance the negative impacts of transportation (Islam et al., 2016; Pagand et al., 2020). The IEA lists minimizing costs per passenger-kilometer or ton-kilometer moved as one of three pillars that are essential to increase the market share of rail transportation[1].

Automatic Train Operations (ATO) which refers to a system that automates different aspects of train operations is expected be one of the key drivers of a more efficient and competitive railway system (ALSTOM Transport SA, 2021; Europe's Rail Joint Undertaking, 2019). ATO is estimated to reduce energy consumption by up to 45%, increase the level of punctuality, increase operational flexibility, and allow for a 50% better utilization of the infrastructure when combined with other technologies.

ATO relies on advanced technologies that are used to perceive and interpret the railway environment in order to allow autonomous operations with minimal or no human intervention (Deutsche Bahn AG, 2022). One aspect of ATO is the precise identification and localization of railway tracks. The ability to detect and isolate tracks based on video images is essential for ensuring the safe navigation of trains through the railway network or in shunting yards. Accurate track detection ensures that the train can make informed decisions, such as adjusting speed, navigating turns, and responding to potential obstacles.

Traditional methods of track detection often rely on rule-based algorithms and image processing techniques, but these approaches may face challenges in diverse environmental conditions such as bad weather, complex background, lighting variations (e.g., day and night), and dirty cameras. This master's thesis addresses the task of multi-sensor rail track detection in the context of ATO. We explore deep learning techniques, particularly convolutional neural networks (CNNs), that have demonstrated great success in computer vision tasks, including image segmentation. The application of deep learning to track detection is expected to outperform conventional non-AI-based techniques and thereby improving the accuracy and robustness of the system. Our analysis is based on a multi-sensors dataset, including images of normal RGB

[1] The other pillars are maximizing revenues from rail systems, and ensuring that all forms of transport (especially road transportation) pay not only for the use of the infrastructure they need, but also for the adverse impacts they generate.

cameras, high-resolution cameras, and infrared cameras, with different orientations, respectively. This multi-sensor approach allows to compare the the effectiveness of different cameras and informs the deployment of those in order to improve the robustness of track detection in diverse conditions.

In the context of rail track detection, researchers have explored various areas. Yet, applying deep learning techniques to detect rail tracks is a relatively raw field. In particular, there is no research that is focusing on comparing different input images such as RGB and infrared cameras and images that are oriented to the left, center, and right of the locomotive. The contribution of this thesis to the literature is three-fold: First, we select and train a deep learning model capable of accurately detecting and segmenting railway tracks using data from RGB cameras, high-resolution cameras, and infrared cameras. In contrast to approaches that have been specifically tailored to the task, we apply a general framework that is easier to use by practitioners without elaborate software engineering skills. The results of the deep learning model are compared to a non-AI based method specialized in identifying lines in images. Second, we conduct a comprehensive performance evaluation to assess the accuracy and computational efficiency of the proposed track detection system on images generated by different cameras. Third, we explore the integration of the developed model into real-world applications by applying to identify tracks in video streams.

By achieving these objectives, this research provides valuable insights and advancements to the field of railway automation, with implications for improving the safety and efficiency of automatic train operations.

2 Literature review

Traditionally, rail track detection has been performed by first extracting features of an image (e.g., gradient-based thresholds) and then detecting rails. These approaches achieve good results in certain conditions. However, deep-learning based approaches are often more robust in real-world environments (Giben et al., 2015; Li and Peng, 2022; Wang et al., 2019). Deep learning techniques, particularly CNNs, have emerged as powerful tools for image segmentation tasks, demonstrating success in various computer vision applications. Recent surveys on image segmentation and object detection using deep-learning techniques is provided by Cheng et al. (2023) and Zaidi et al. (2022), respectively.

The following sections examine related research in track detection, considering both deep learning-based segmentation and traditional non-AI segmentation methods.

2.1 Traditional rail track detection

While deep learning has shown remarkable success in track detection, non-AI segmentation techniques continue to play a role in this field as they allow the integration of domain-specific knowledge and rules into the algorithm and require less data for training. These methods are often referred to as line segment detectors and involve traditional computer vision techniques such as thresholding, edge/contour detection, template matching, and region growing (Almazàn et al., 2017; Grompone von Gioi et al., 2010, 2012; Sahoo et al., 1988).

Kaleli and Akgul (2009) present a dynamic programming algorithm to extract the rail tracks in front of the train. The idea is to first identify the vanishing point which refers to the imaginary intersection of the tracks as the distance between the tracks decreases from the bottom of the image to the top. This step is based on computing the gradient and applying Hough transform to detect the straight lines that indicate the tracks. Next, dynamic programming is used to extract the space between the two tracks. Qi et al. (2013) apply a method based on histogram of oriented gradients (HOG) to identify tracks and switches. First, HOG features are computed; railway tracks are then identified by a region-growing algorithm. The proposed method is able to predict the patch the train will travel by detecting the setting of the switches. Nassu and Ukai (2011) introduce an approach that performs rail extraction by matching edge features to candidate templates.

While the previously mentioned approaches focus on images by on-board cameras, Purica et al. (2017) examines the detection of tracks in aerial images taken by drones. The solution approach is based on Hough transform.

(Arastounia, 2015) and (Yang and Fang, 2014) develop methods to recognize railroad infrastructure from 3D LIDAR data. In (Arastounia, 2015), railway components such as rail tracks, contact cables, catenary cables, masts, and cantilevers are classified based on local neighborhood structure, shape of objects, and topological relationships among objects. (Yang and Fang, 2014) focus on the detection of tracks. The authors utilize the geometry and reflection intensity of the tracks to extract features and identify tracks.

2.2 Deep-learning based rail track detection

Deep-learning based techniques such as semantic segmentation incorporate convolutional neural networks (CNNs) and other deep architectures to automatically learn features from raw image data. Semantic segmentation aims to assign a label to each pixel in the image, distinguishing between the pixels that belong to the rail tracks and those that represent the background and is therefore particularly well suited for rail track detection.

Giben et al. (2015) and Le Saux et al. (2018) were among the first authors who evaluated the performance of deep learning-based segmentation against traditional segmentation techniques in rail track detection. In Giben et al. (2015), the authors propose a CNN for localizing and inspecting the condition of railway component based on gray-scale images. The authors report that the CNN model is better suited to capture complex patters compared to approaches that rely on traditional texture features (e.g., discrete Fourier transforms of local binary pattern histograms). Le Saux et al. (2018) detect rail tracks in aerial images by devising a CNN based approach and different traditional approaches such as thresholding.

Wang et al. (2019) propose the RailNet – a deep-learning based rail track segmentation algorithm that combines the ResNet50 backbone with a fully convolutional network. In order to train the model, the authors compile a non-public dataset consisting of 3000 images from forward-facing on-board cameras. Experiments show that RailNet is able to outperform general purpose models for segmentation.

A machine-learning based approach is proposed by (Teng et al., 2016) where features are extracted from super-pixels (i.e., a group of adjacent pixels with similar characteristics) and classified by applying a previously trained support vector machine.

2.3 Lane detection

Lane detection for road vehicles is similar to rail track detection for locomotives in the sense that both tasks aim to identify and segment elongated shapes in complex environments that vary in lighting conditions, shadows, and occlusions. The field of lane detection appears to have a richer body of literature which might be attributed to the existence of well-established benchmark datasets such as TuSimple (2017) and CULane (Pan et al., 2018).

Early work on lane detection is based on traditional approaches such as Hough transform and clustering (Duda and Hart, 1972; Ma and Xie, 2010). However, recently the focus of researchers has shifted to deep-learning based approaches such as CNN and recurrent neural networks (RNN) (Meyer et al., 2021; Wang et al., 2022; Zheng et al., 2022). Tang et al. (2021) and Yang (2023) provide comprehensive surveys on lane detection approaches. In (Yang, 2023), the authors propose a combined approach in which the advantages of traditional and deep-learning based methods are mixed.

3 Datasets

Deep learning algorithms require a labeled dataset with a large number of images in order to train models that are capable of detecting objects in real-time with a high level of accuracy. Compared to the domain of autonomous vehicles on the road, there is only a very limited number of datasets for railway applications. The rail semantics dataset 2019 (RailSem19) is the first publicly available dataset for detecting objects (including rails) in the railway domain (Zendel et al., 2020). The French railway signaling dataset (FRSign) is a dataset focusing only on traffic lights (Harb et al., 2020), whereas the Railway Pedestrian Dataset (RAWPED) is focusing on pedestrian detection methods (Toprak et al., 2020). The dataset proposed by Wang et al., 2019 - railroad segmentation dataset – is not available to the public. The Rail-DB dataset (Xinpeng and Xiaojiang, 2022) comprises 7.432 annotated images, featuring different scenarios (e.g., weather conditions). This thesis is based on the first freely available multi-sensor data set for the development of fully automated driving in the railway sector (Deutschland Deutsche Bahn AG, 2023; Tagiew, 2023). The setup on the locomotive comprises infrared/RGB cameras, lidar, radar, positioning, and acceleration sensors. The images may contain up to 20 different object classes that are annotated using polylines, polygons, bounding boxes and cuboids. For our purposes (i.e., detecting tracks), we will focus on images of infrared and RGB cameras. Three examples of annotated images are given in Figure 1, one for each sensor respectively. The dataset features 18.543 annotations of tracks.

3.1 Mutli-sensor dataset by Deutsch Bahn

3.2 RailSem dataset

4 Experiments

4.1 Modeling and performance evaluation

4.2 Non-AI based segmentation

4.3 Deep-learning based segmentation

5 Results

6 Conclusion

Etwas Text... Hier kommen noch einige Abkürzunge vor zum Beispiel Alphabet (ABC),world wide web (WWW) und Rolling on floor laughing (ROFL).

Literaturverweise sollten automatisch verwaltet werden, vor allem, wenn es viele Quellenverweise gibt. Beispiele sind Kopka (2005a), Kopka (2005b), Goossens et al. (2002), Teschl et al. (2014), Humenberger et al. (2007), Humenberger et al. (2010), Zinner et al. (2007), Hemetsberger (2007), Siemens Automation Technology (2011), Siemens Automation Technology (2014), International Standards Office (1998), Atmel Corporation (2011), Humenberger (2011), Pohn (2010). Das verwendete Zitierformat (bzw. das Format des Literaturverzeichnisses) ist entspechend der Vorgaben der Studiengänge zu wählen.

6.1 Algorithms

Use a defined environment for algorithms.

Algorithm 1 is an example from the gallery (https://www.overleaf.com/latex/examples/euclids-algorithm-an-example-of-how-to-write-algorithms-in-latex/mbysznrmktqf) .

Algorithm 1 Euclid's algorithm

procedure EUCLID(a, b) ▷ The g.c.d. of a and b
    $r \leftarrow a \bmod b$
    while $r \neq 0$ **do** ▷ We have the answer if r is 0
        $a \leftarrow b$
        $b \leftarrow r$
        $r \leftarrow a \bmod b$
    return b ▷ The gcd is b

Bibliography

Emilio J. Almazàn, Ron Tal, Yiming Qian, and James H. Elder. Mcmlsd: A dynamic programming approach to line segment detection. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5854–5862, 2017. doi: 10.1109/CVPR.2017.620.

ALSTOM Transport SA. Autonomous mobility: The future of rail is automated, 2021. URL https://www.alstom.com/autonomous-mobility-future-rail-automated.

Mostafa Arastounia. Automated recognition of railroad infrastructure in rural areas from lidar data. *Remote Sensing*, 7(11):14916–14938, 2015. ISSN 2072-4292. doi: 10.3390/rs71114916. URL https://www.mdpi.com/2072-4292/7/11/14916.

Atmel Corporation. Atmel atmega16 – 8-bit microcontroller with 16k bytes in-system programmable flash, 2011. URL http://www.atmel.com/dyn/resources/prod_documents/doc2466.pdf.

Jieren Cheng, Hua Li, Dengbo Li, Shuai Hua, and Victor S. Sheng. A survey on image semantic segmentation using deep learning techniques. *Computers, Materials & Continua*, 74(1):1941–1957, 2023. ISSN 1546-2226. doi: 10.32604/cmc.2023.032757. URL http://www.techscience.com/cmc/v74n1/49879.

Deutsche Bahn AG. Automatic train operation (ato), 2022. URL https://digitale-schiene-deutschland.de/en/Automatic-Train-Operation.

Richard O. Duda and Peter E. Hart. Use of the hough transformation to detect lines and curves in pictures. *Commun. ACM*, 15(1):11–15, jan 1972. ISSN 0001-0782. doi: 10.1145/361237.361242. URL https://doi.org/10.1145/361237.361242.

Europe's Rail Joint Undertaking. Innovation in the spotlight: Towards unattended mainline train operations (ato goa 4), 2019. URL https://rail-research.europa.eu/highlight/innovation-in-the-spotlight-towards-unattended-mainline-train-operations-ato-goa4/.

Xavier Giben, Vishal M. Patel, and Rama Chellappa. Material classification and semantic segmentation of railway track images with deep convolutional neural networks. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 621–625, 2015. doi: 10.1109/ICIP.2015.7350873.

Michel Goossens, Frank Mittelbach, and Alexander Samarin. *Der LaTeX Begleiter*. Addison-Wesley Deutschland, Bonn, 2002.

Rafael Grompone von Gioi, Jeremie Jakubowicz, Jean-Michel Morel, and Gregory Randall. Lsd: A fast line segment detector with a false detection control. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(4):722–732, 2010. doi: 10.1109/TPAMI.2008.300.

Rafael Grompone von Gioi, Jérémie Jakubowicz, Jean-Michel Morel, and Gregory Randall. LSD: a Line Segment Detector. *Image Processing On Line*, 2:35–55, 2012. https://doi.org/10.5201/ipol.2012.gjmr-lsd.

H. Hemetsberger. Ait stereo sensor im einsatz während der darpa urban challenge 2007, 2007. AIT Austrian Institute of Technology.

M. Humenberger, D. Hartermann, and W. Kubinger. Evaluation of stereo matching systems for real world applications using structured light for ground truth estimation. In *Proceedings of the Tenth IAPR Conference on Machine Vision Applications (MVA2007)*, pages 433–436. MVA Conference Committee, 2007.

Martin Humenberger. Real-time stereo matching for embedded systems in robotic applications, 2011.

Martin Humenberger, Christian Zinner, Michael Weber, Wilfried Kubinger, and Markus Vincze. A fast stereo matching algorithm suitable for embedded real-time systems. *Computer Vision and Image Understanding*, 114(11):1180–1202, 2010.

International Energy Agency. The future of rail, 2019. URL https://www.iea.org/reports/the-future-of-rail.

International Standards Office. Iso 690 – information and documentation: Bibliographical references: Electronic documents, 1998.

Dewan Md Zahurul Islam, Stefano Ricci, and Bo-Lennart Nelldal. How to make modal shift from road to rail possible in the european transport market, as aspired to in the eu transport white paper 2011. *European transport research review*, 8(3):1–14, 2016.

Fatih Kaleli and Yusuf Sinan Akgul. Vision-based railroad track extraction using dynamic programming. In *2009 12th International IEEE Conference on Intelligent Transportation Systems*, pages 1–6, 2009. doi: 10.1109/ITSC.2009.5309526.

Helmut Kopka. *LaTeX, Band 1: Einführung*. Pearson Studium, München, 3 edition, 2005a.

Helmut Kopka. *LaTeX, Band 1: Einführung*. Pearson Studium, München, 3 edition, 2005b. URL http://www.pearson-studium.de.

B. Le Saux, A. Beaupère, A. Boulch, J. Brossard, A. Manier, and G. Villemin. Railway detection: From filtering to segmentation networks. In *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 4819–4822, 2018. doi: 10.1109/IGARSS.2018.8517865.

Xinpeng Li and Xiaojiang Peng. Rail detection: An efficient row-based network and a new benchmark. In *Proceedings of the 30th ACM International Conference on Multimedia*, MM '22, page 6455–6463, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450392037. doi: 10.1145/3503161.3548050. URL https://doi.org/10.1145/3503161.3548050.

Chao Ma and Mei Xie. A method for lane detection based on color clustering. In *2010 Third International Conference on Knowledge Discovery and Data Mining*, pages 200–203, 2010. doi: 10.1109/WKDD.2010.118.

Annika Meyer, Philipp Skudlik, Jan-Hendrik Pauls, and Christoph Stiller. Yolino: Generic single shot polyline detection in real time. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2916–2925, 2021.

Bogdan Tomoyuki Nassu and Masato Ukai. Rail extraction for driver support in railways. In *2011 IEEE Intelligent Vehicles Symposium (IV)*, pages 83–88, 2011. doi: 10.1109/IVS.2011.5940410.

I Pagand, C Carr, and J Doppelbauer. Fostering the railway sector through the european green deal. *European Union Agency For Railways*, 2020.

Xingang Pan, Jianping Shi, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Spatial as deep: Spatial cnn for traffic scene understanding. In *AAAI Conference on Artificial Intelligence (AAAI)*, February 2018.

Johannes Pohn. Condition monitoring systeme für die zustandorientierte instandhaltung von windkraftanlagen, 2010.

Andrei I. Purica, Beatrice Pesquet-Popescu, and Frederic Dufaux. A railroad detection algorithm for infrastructure surveillance using enduring airborne systems. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2187–2191, 2017. doi: 10.1109/ICASSP.2017.7952544.

Zhiquan Qi, Yingjie Tian, and Yong Shi. Efficient railway tracks detection and turnouts recognition method using hog features. *Neural Computing and Applications*, 23:245–254, 2013.

P.K Sahoo, S Soltani, and A.K.C Wong. A survey of thresholding techniques. *Computer Vision, Graphics, and Image Processing*, 41(2):233–260, 1988. ISSN 0734-189X. doi: https://doi.org/10.1016/0734-189X(88)90022-9. URL https://www.sciencedirect.com/science/article/pii/0734189X88900229.

Siemens Automation Technology. Simatic, 2011. URL http://www.automation.siemens.com/mcms/topics/de/simatic/Seiten/Default.aspx.

Siemens Automation Technology. Simatic, 2014. [Online] Verfügbar unter: <http://www.automation.siemens.com/mcms/topics/de/simatic/Seiten/Default.aspx> [Zugang am 17.10.2014].

Jigang Tang, Songbin Li, and Peng Liu. A review of lane detection methods based on deep learning. *Pattern Recognition*, 111:107623, 2021.

Zhu Teng, Feng Liu, and Baopeng Zhang. Visual railway detection by superpixel based intracellular decisions. *Multimedia Tools and Applications*, 75:2473–2486, 2016.

Susanne Teschl, Karl Michael Göschka, and Günter Essl. Leitfaden zur Verfassung einer Bachelorarbeit oder Master Thesis, 2014. URL www.technikum-wien.at.

TuSimple. Tusimple, 2017. URL https://github.com/TuSimple/tusimple-benchmark/tree/master/doc/lane_detection.

Jinsheng Wang, Yinchao Ma, Shaofei Huang, Tianrui Hui, Fei Wang, Chen Qian, and Tianzhu Zhang. A keypoint-based global association network for lane detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1392–1401, 2022.

Yin Wang, Lide Wang, Yu Hen Hu, and Ji Qiu. Railnet: A segmentation network for railroad detection. *IEEE Access*, 7:143772–143779, 2019. doi: 10.1109/ACCESS.2019.2945633.

Bisheng Yang and Lina Fang. Automated extraction of 3-d railway tracks from mobile laser scanning point clouds. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(12):4750–4761, 2014. doi: 10.1109/JSTARS.2014.2312378.

Mingyue Yang. Lane detection methods survey for automatic driving. In *Journal of Physics: Conference Series*, volume 2547, page 012015. IOP Publishing, 2023.

Syed Sahil Abbas Zaidi, Mohammad Samar Ansari, Asra Aslam, Nadia Kanwal, Mamoona Asghar, and Brian Lee. A survey of modern deep learning based object detection models. *Digital Signal Processing*, 126:103514, 2022. ISSN 1051-2004. doi: https://doi.org/10.1016/j.dsp.2022.103514. URL https://www.sciencedirect.com/science/article/pii/S1051200422001312.

Tu Zheng, Yifei Huang, Yang Liu, Wenjian Tang, Zheng Yang, Deng Cai, and Xiaofei He. Clrnet: Cross layer refinement network for lane detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 898–907, 2022.

Christian Zinner, Wilfried Kubinger, and Richard Isaacs. Pfelib: a performance primitives library for embedded vision. *EURASIP Journal on Embedded Systems*, 2007:1–14, 2007. URL http://downloads.hindawi.com/journals/es/2007/049051.pdf.

List of Figures

List of Tables

List of source codes

List of Abbreviations

ABC Alphabet

WWW world wide web

ROFL Rolling on floor laughing

ATO Automatic Train Operations

CNN Convolutional Neural Network

A Anhang A

B Anhang B