2nd Conference on Production Systems and Logistics

A Deep Learning Framework for Automated Collection and Analysis of Traffic Data based on Identifying and Classifying Delivery Vehicles in Logistics

Meng Jin[1], Lars Andreas Mauch[2], Bernd Bienzeisler[1]
[1]*Fraunhofer IAO | Forschungs- und Innovationszentrum Kognitive Dienstleistungssystem , Heilbronn, Germany*
[2]*Fraunhofer IAO | Energy Innovation, Stuttgart, Germany*

Abstract

In urban logistics, analyzing urban traffic data plays an important role in achieving higher schedule reliability and delivery time efficiency. To increase the diversity of urban traffic data, we developed a solution for the automated collection and analysis of different types of traffic data. These are needed to optimize and control the flow of traffic.

The use of traditional on-road sensors (e.g., inductive loops) for collecting data is necessary but not currently sufficient because it cannot draw any conclusions about the type of goods being transported. In this paper, we propose a framework in which different classes of delivery vehicles and types of goods being shipped are identified in road videos by deep-learning-based image recognition method. Video sequences are automatically evaluated according to the following criteria: (i) distinguish between individual and commercial vehicles, (ii) identify the category of commercial vehicles, for example, van, box trucks, small trucks, etc. (iii) identify the special features of the vehicle body (such as the name of the carrier) to classify commercial transportation of food, general goods or package services, etc. Using this method, logistics throughput of a designated city or region and the peak time of goods transportation can be obtained. This provides the carrier with better pre-advice and potential actions to improve transportation efficiency. For the evaluation of our framework, we collected real street videos at different time points in the main traffic arteries of Heilbronn, Germany. In particular, the difference between traffic flow of logistics services before and during the COVID-19 epidemic was compared. The results of implementation and testing demonstrated a high-precision, low-latency performance of the framework for obtaining urban logistics data.

Keywords

Urban Freight Transportation; Identification and Classification Delivery Vehicles; Logistics; Deep Learning; Image Recognition

1. Introduction

The complexity of urban traffic and the diversity of its impacts require very thorough data collection methods since the collection and analysis of traffic data play a significant role in traffic safety, traffic pollution, and urban road planning. In this paper, we focus on the collection method of commercial vehicle traffic data, because they account for a large part of the traffic data and are particularly useful for urban logistics service providers, who compete with other road users for scarce traffic space in the inner city, the continuous traffic

data are extremely important for them to optimize transportation operations and make reliable decisions through data analysis.

The technology of collecting traffic data has been studied for decades. It can be easily divided into two categories, the intrusive and non-intrusive methods [1]. The intrusive methods consist of a data recorder or a sensor placing on or in the road, such as traditional road sensors (induction loops). Such technologies have been employed for many years. However, due to various factors, including limited coverage and possible inability to obtain continuous traffic count data, such as pavement repair, construction, and maintenance. In this case, non-intrusive data can be used as an alternative for a short period or to supplement traffic data diversity, which is based on remote observation, including manual counting, passive and active infrared, microwave radar, video image detection, etc. [1]. These different types of data collectors have their advantages and disadvantages, and their combination provides more complete traffic data. In this paper, we mainly focus on video image detection technology because other data collection methods have made outstanding contributions to vehicle counting and geographic information. Obtaining more vehicle feature information through images, such as vehicle classification, vehicle size, weight, or the type of goods being transported, to supplement the defects of current traffic data is the focus of our work. According to literature surveys, most existing video-based detectors can only provide a few macro traffic parameters, such as traffic count and average speed. Several algorithms based on computer vision have been developed to classify vehicles, but most of them only use traditional image processing and mathematical methods, such as calculating the vehicle's length in pixels for classification [5]. The result is that trucks can only be divided into long vehicles and short vehicles, and the accuracy of the results cannot achieve very high since the vehicle's length is susceptible to road reflections, lighting changes, shadows, and other interference factors. In addition, manual input of the exact road area to be observed and the threshold used to determine the length of the truck are also required. In this paper, a new video-based traffic data collection framework is developed to track each passing vehicle, the types of vehicles and the features of vehicle body will be recognized using deep learning methods. It no longer relies on manual input of thresholds and precise road positions in advance. By identifying features on the vehicle, such as letters, logistics service providers and the items being transported can also be deduced. In this way, high-precision and detailed traffic data can be collected, and the process of data collection is simplified compared with the complex algorithm that required manual input before.

The paper starts with a short overview of the methods for collecting traffic data in previous works. In section 3, the emphasis is put on the framework for identifying the features of commercial vehicles on the road based on the deep-learning method, the detailed methods used to distinguish between individuals and commercial vehicles, classify commercial vehicles, and identify transported items will also be described in detail. The purpose of section 4 is to evaluate our proposed method and show the results of applying it to roads to collect traffic data. The final section concludes this research effort and proposes further research topics.

2. Related Work

Before presenting the details of our framework related studies are briefly introduced. The methods of traffic data collection have been an area of interest in intelligent transportation system for the past few decades. [1] introduced all data collection methods on the market and analyzed their advantages and disadvantages, such as capabilities and restrictions. [2] in 2010 provided a framework for optimizing logistics time based on traffic data collection and proved that traffic data is beneficial for estimating and optimizing logistics delivery time. [3] designed a traffic information system, as a test, traffic flow and vehicle speed were collected in Berlin, Germany through fixed measuring equipment and specially equipped vehicles. [4] analyzed the advantages and disadvantages of traffic data based on video detection technology and Automatic Traffic Recorder. The results showed that although the data based on video detection technology

are not very accurate, they can provide useful traffic data information. The acceptability of the data should depend on the accuracy requirements of the use cases. [6] developed a new video-based traffic data collection system. Each passing vehicle can be tracked and classified in mixed traffic situations. The average speed of each passing vehicle is identified as the main contribution of this paper. [7] proposed a system based on video detection to classify vehicles and estimate the target color. [8] showed an algorithm for classifying vehicles based on five-dimensional feature vectors, and each vehicle can be tracked by a temporal tracking methodology. Through this process, both the macro and micro parameters of the vehicle will be available. However, due to the high computational complexity, real-time collection of traffic data is difficult to be applied. [9] proposed a method to detect and classify vehicles based on a three-dimensional (3D) model in aerial imagery, but the previous condition of this method is to build a 3D matching vehicle model for training, that is complicated and time-consuming, the accuracy is obviously affected by the constructed 3D model. Generally, these previous studies still have the following shortcomings: (i) many methods can only detect single-lane vehicles (ii) Manual input of the detection area or setting of detection parameters is required in advance (iii) Insufficient types of vehicles, such as only divided into two classes by length (iv) lacks an identification method to identify the transported object. The framework described in this paper aims to overcome these shortcomings.

3. Methodology

In this section, the main steps of our deep learning-based traffic data collection and analysis framework will be introduced, which can be briefly called DeepTraffic.

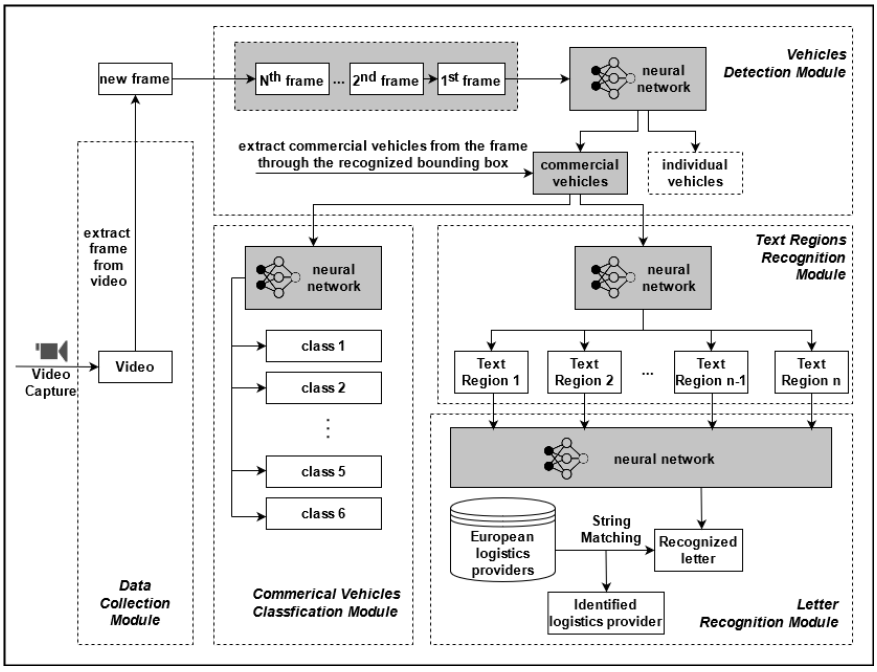

Figure 1: Flow Chart of the DeepTraffic Framework

3.1 Framework Overview

DeepTraffic framework has five modules: data collection, vehicle detection, commercial vehicle classification, text regions recognition, and letter recognition. Figure 1 shows the flow chart of our framework. First, a digital video is taken by a camera placed on the road to capture multiple lanes in the same direction. Our framework takes road video as input and extracts every frame in real-time as the new input for the next step. In the next vehicle detection module, we trained a neural network with various labeled

real-world pictures to detect vehicles and distinguish between private and commercial vehicles. Through this step, the commercial vehicle can be marked with a bounding box in the video in real-time and extracted from the background as a new image, which is used as a new input and enters the following two modules at the same time: in the commercial vehicle classification module, we employed a simplified VGG network to divide commercial vehicles into six categories, in the text regions recognition module, an open-source deep learning model is utilized to obtain the features of the vehicle body. The recognized text area will be transferred to the letter recognition module, where all English or German letters can be identified from the picture. To ensure the correct rate of recognition, a string-matching algorithm is used to match the recognized letters with the name of the courier service provider in our pre-established database Finally, we tracked the vehicle and counted the number of passing vehicles by analyzing the vehicle's type and possible logistics providers in consecutive frames. The following sections will explain the methods involved in the entire process in more detail.

3.2 Collection of Image Samples

In a deep-learning-based framework, many sample pictures are needed to train the neural network. To make our system adapt to each road section, various video images on highways, country roads and urban roads in different periods and weather conditions are taken by *YI* cameras with a frequency of 30 frames per second (FPS). Figure 2 shows some examples of our images. We shot 6 hours of video on the bridge above highway A8 and B14 in Stuttgart, Germany, highway A6 in Heilbronn, Germany, and 9 hours of video along the main road in the city Heilbronn center. We collect data for research purposes only, according to data protection regulations we will not collect any personal information from the video. All video data will not be disclosed and will be deleted after analysis.

Figure 2: Examples of Training Images

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Figure 3: YOLO V3 network structure [11]

3.3 Vehicle Detection

Deep Neural Networks (DNN) based vehicle detection methods have been applied to many image/video applications and have achieved state-of-the-art on various data set. A large data set is required for transfer learning if we apply them directly, which is a laborious and expensive construction. However, if we train them on a small data set to solve a complex task, it usually leads to overfitting problems. We intend to use DNN with our small data set to accomplish the following two simple goals: (i) detecting vehicles, (ii) distinguishing between private and commercial vehicles. The detailed vehicle classification network will be applied separately. Because if commercial vehicles are classified in detail on this complex DNN model,

overfitting often occurs. After searching, we found that a DNN architecture called YOLO V3 [11] [12] has been widely used in object detection tasks and achieved excellent results. To perform vehicle detection more effectively, we choose the YOLO V3 framework as the basic structure and fine-tuned it based on our data set. More than 5,000 pictures from 25 hours of video are selected as a training image and labeled by an open-source software called LabelImg [10]. Figure 3 shows the basic structure of YOLO V3 network, which is trained on COCO dataset with 80 classes [11]. Therefore, we reduced the number of categories to 2 to suit our tasks, namely private and commercial vehicles. Figure 4 shows the detection results of the YOLO V3 model fine-tuned by our data set**: s**ome vehicles in the distance have been identified, but this is not what we expected because we need clear pictures to prepare for future vehicle classification or body feature recognition. We used the following criterion to remove these outliers: (i) we only select objects with a confidence value higher than 0.6 as valid objects. (ii) we set a general road section (the green box in Figure 4) as a valid region, and the detected object is only valid when the center of the object is within the selected valid region. The results show that the YOLO V3 model can identify all private and commercial vehicles in the video with an accuracy rate of 97%. Since we are analyzing consecutive frames, even if certain vehicles are not recognized in some individual frames, these vehicles will appear in at least three frames to ensure that no passing vehicles are missed.

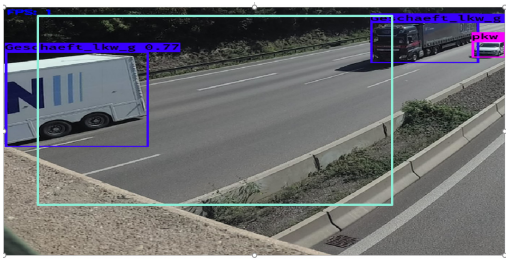

Figure 4: Examples of detected Vehicles

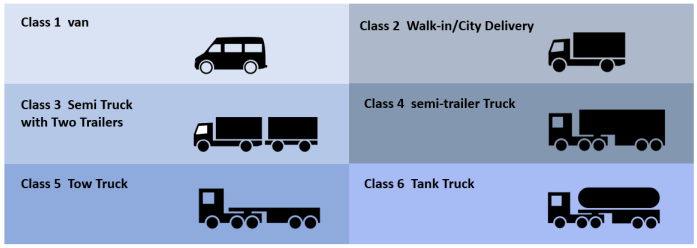

Figure 5: Examples of detected Vehicles

3.4 Commercial Vehicles Classification

In the previous step, the recognized commercial vehicle will be extracted from the video automatically and saved locally used as the new input for the vehicle classification module. The corresponding frame number is also stored in the image name and used later for vehicle counting. As shown in Figure 5, we classify vehicles into six classes: van, walk-in/city delivery, semi-truck with two trailers, semi-trailer truck, tow truck, and tank truck. These types of vehicles are commonly used in German logistics transportation. The CNN architecture we used for classification is SmallerVGGNet, a simplified version of the original VGGNet, which was first introduced in [13] in 2014. We do not need such a vast network structure of the original version for our task with only six categories, so we simplified the network with only four repeated Conv blocks with different kernel sizes (see Figure 6). The input image of the network is a fixed-size RGB image. We resized all the extracted images to 224*244 and sent them to the network. First, four conv blocks are used to extract features. Each conv block contains one 2D convolutional layer with different kernel sizes, a Relu activation function to scale the output, one batch normalization layer to standardize the inputs, and one max pooling layer for aggregating the results of convolutional layers by only passing on the strongest feature. The output should be rolled out through Flatten layer and connected to a dense layer, which has exactly the number of neurons that corresponds to the number of different classes to be recognized. In this case, we used 600 classified images as the data set, of which 480 are used as training images and 120 are used as test images. The network was trained on a Nvidia 2080Ti GPU, the accuracy of the training set reached 98.2%, and the accuracy of the test set fluctuated between 90% and 95% after 30 Epochs.

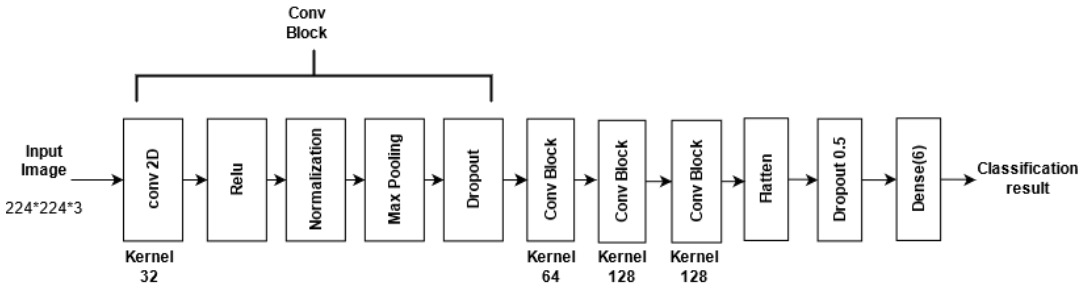

Figure 6: Structure of SmallerVGGNet

Figure 7: Example of text recognition on the vehicle body

3.5 Text Regions Recognition

To detect the text on the vehicle body, we directly used an open-source deep learning model called EAST proposed by Zhou et al. in 2017 [14]. It can run at near real-time at 13 FPS on 720p images and obtain state-of-the-art text detection accuracy. The vehicle image extracted in the vehicle detection module is also adopted as input in this module. During the recognition process, we found that the text on the car body is not parallel in the picture because of the shooting angle, as shown in the leftmost picture in Figure 7, which makes the text harder to recognize. To overcome this difficulty, the input picture is preprocessed: according to the angle at which we took the video, the picture can be rotated 5 degrees clockwise to make the text in the image parallel, as shown in the middle picture in Figure 7, this angle of rotation can be applied to almost all the pictures we took. The rightmost picture in Figure 7 shows the final recognition result. The text area on the vehicle body will be recognized and cut out.

3.6 Letter Recognition and Matching

To identify the logistics provider and assume the items being shipped, we need to convert the recognized text into letters. In this module, we used the open-source scene text recognition (STR) benchmark [15], which was proposed by Baek et al. in 2019 and integrated and analyzed the current advanced models for character recognition. It pointed that the accuracy of the best model on different data sets is between 74% and 94.9%, which means that using this model to recognize the characters in our cases cannot be completely accurate. Figure 8 shows a part of the recognition result, that is stored in a text file and contains the following information: the ID of the video, the number of frames in the video, the position of the recognized vehicle in the frame, the recognized vehicle type, the probability of the recognized vehicle type, and the characters recognized by the vehicle body. From the results, we concluded that some recognized characters are slightly different, even in the same vehicle in consecutive frames. To obtain the unique logistics provider, we collected information on 5188 logistics companies in Germany and Austria in advance and calculated the similarity between the recognized characters and them through the Levenshtein algorithm and matched it to the unique logistics provider in our database. In consideration of data protection, we do not public our logistics company database here. After matching the logistics provider in our database, we count the number of vehicles according to the following criteria: (i) if there are vehicles of the same logistics provider in consecutive frames, this vehicle can be regarded as one vehicle, and one vehicle is counted. (ii) if there are no characters recognized in consecutive frames, such as a white vehicle, then we use the position of the

identified vehicle in the picture to determine whether it is the same vehicle, because as the car moves forward, the position of the car must be closer to our camera. To determine the type of vehicle, we perform the following calculation: In consecutive frames, if the identified vehicle type is unique, the vehicle type can be simply derived. If more than two different types are identified, the type with the highest probability is selected as the final result. The item being transported can be inferred from the name of the logical operator. For example, from the result of "DACHSER Intelligent logistics", it can be known that the item being transported is food, if “DHL or DPD” are recognized, it can be inferred to be a package service. An example of result of our framework is shown in Figure 9.

```
video11_frame0139_000_347_Semi Truck with two trailers_099_
video11_frame0188_061_320_Tow Truck_099_
video11_frame0196_296_462_Semi Truck with two trailers_099_Intelligent|DACHSER
video11_frame0197_260_450_Semi Truck with two trailers_099_Intelligent|DACHSER
video11_frame0198_213_431_Semi Truck with two trailers_099_DACHSER|Intelligent
video11_frame0199_158_424_Semi Truck with two trailers_099_Intelligent|DACHSER
video11_frame0200_105_392_Semi Truck with two trailers_099_Intelligent|Logistics|DACHSER
video11_frame0201_022_364_Semi Truck with two trailers_099_Intelligent|Antoristically|DACHSER
video11_frame0202_000_344_Semi Truck with two trailers_099_Logistics|STACHS|DACHSER|Intelligent
video11_frame0203_006_307_Semi Truck with two trailers_099_Logistics|Intelligent|DACHSER
```

Figure 8: Example of text recognition

```
['video10', 'frame0002', 'Walk-in', 'Concerverting']
['video10', 'frame0029', 'Tank Truck', 'ABE']
['video10', 'frame0062', 'Semi-trailer Truck', 'Wiedemeyer']
['video10', 'frame0069', 'Tow Truck', '']
['video10', 'frame0101', 'Semi-trailer Truck', 'EUROTIRE']
['video10', 'frame0131', 'Semi-trailer Truck', 'BOR']
['video10', 'frame0145', 'Semi-trailer Truck', 'KING']
['video11', 'frame0010', 'Semi-trailer Truck', 'burkhardt']
['video11', 'frame0022', 'Semi-trailer Truck', 'Simon SPEDITION Darmstadt']
['video11', 'frame0071', 'Semi-trailer Truck', 'ACTION'],
['video11', 'frame0081', 'Tank Truck','Alloborthors Contranter transter']
['video11', 'frame0115', 'Semi-trailer Truck', 'ACTION']
['video11', 'frame0203', 'Semi Truck with two trailers', 'DACHSER Inteligent Logistics']
```

Figure 9: Example of text recognition on the vehicle

4. Results

Table 1: Experimental results for vehicle detection

	total	True Positive	True Negative	False Negative	False Positive	correct (%)
result	630	612	10	0	9	97.1

Experimental results and discussion on the performance of the DeepTraffic framework are described in the section. To demonstrate the effectiveness of our framework, the experiments are conducted with the following procedures: we chose 9 hours of video data as the test set, which were shot on three main roads in Heilbronn, Germany. Considering the intensity of the sun and the weather conditions, videos are shot between 9:00-10:00, 13:00-14:00, and 16:00-17:00 on sunny, cloudy, and rainy days. No input images are included in the sample images for training. Table 1 shows the results of vehicle detection modules. We compared the results from our framework with the results observed manually (ground-truth data). Among the 630 commercial vehicles, 612 were correctly identified and the recognition accuracy rate reached 97.1%. 10 commercial vehicles were not identified because of occlusion, 8 large private vehicles, such as SUVs, were incorrectly identified as commercial vehicles.

Table 2: Experimental results for vehicle classification

	Class1	Class2	Class3	Class4	Class5	Class6	total	correct	%
Class1	65	18	0	1	0	0	84	65	77.3

Class2	0	218	0	16	0	1	234	218	93.2
Class3	0	0	28	4	0	0	32	28	87.5
Class4	0	3	2	231	0	0	236	231	97.9
Class5	0	0	0	2	24	1	27	24	88.9
Class6	0	0	0	3	0	14	17	14	82.3
total							630	580	92.1

*Class1=Van, Class2=Walk-in, Class3= Semi-Truck with two trailers, Class4= Semi-trailer truck, Class5=Two truck, Class6=Tank truck

Table 2 shows the results of our evaluation of one-hour video classification, which classifies the commercial vehicles in each frame. The correct average rate is close to 92.1%. In this experiment, only when the classification result matches the human visual judgment results, the system classification results are regarded as the correct answer. The later steps of using vehicle counting also bring more accurate classification results since the same vehicle may be divided into different categories due to different distances in consecutive frames. The final classification results will be analyzed through the subsequent steps of vehicle counting, and the category with a higher probability will be selected. It can also be concluded from the experimental that the results on cloudy and rainy days are better than those for sunny days because the images are more stable without the influence of sunlight on cloudy or rainy days.

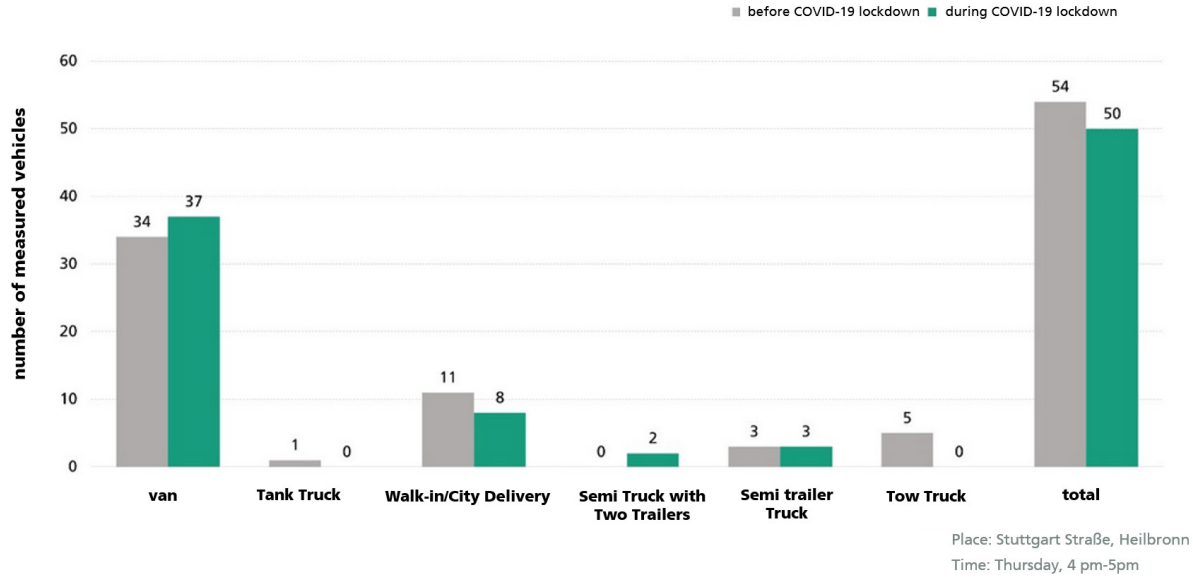

Figure 10: Traffic flows before and during the COVID-19 lockdown

In consideration of data protection, the result of the next text regions recognition module and letter recognition module will not be shown here. From the results, we can conclude that, although the text recognition algorithms have a high accuracy rate, only one-third of passing vehicles are clearly identified which logistics provider belongs to. There are three reasons why the logistic provider has not been identified: (i) there are no recognizable characters on the vehicle, (ii) there are only some characters used for advertising on the vehicle, which cannot match the providers in our database, (iii) the identified provider is not in our database. After analyzing the identified logistics providers, we found that the vehicles of German logistics providers appeared the most frequently. To determine the items they deliver, the service areas of them should be researched one by one. The second place is parcel services such as DHL, PDP, and GLS, which appear 1

to 5 times per hour. The third place is the delivery service of some supermarkets, which is used to deliver food and drinks.

To put our system into practice, we used our system to collect and process traffic data in the city of Heilbronn to understand whether lockdown has an impact on commercial traffic. We took two one-hour videos from 4 pm to 5 pm on Thursday at the same place in Heilbronn before (27.02.2020) and during lockdown (09.04.2020). Comparing the results, we counted that the amount of commercial traffic on Stuttgarter Street was almost the same as the traffic collected before the lockdown (see Figure 10). the results indicate that city delivery traffic was less affected by lockdown than private transport. It is noticeable that no tow truck was detected during the Corona period, obviously, there was less needed to tow away incorrectly parked or defective cars.

5. Conclusion and Outlook

This paper proposed a framework that combines different deep learning algorithms to collect and analyze each commercial vehicle on the road by using video images. The evaluation results from the three test locations in Heilbronn are encouraging. In all the test data, the accuracy of vehicle detection is higher than 97.1%, and the accuracy of vehicle classification and vehicle counting is higher than 92.1%. Letter recognition can identify the providers of one-third of the vehicles on the road. Of course, it depends to a large extent on the data in our pre-set database and whether the characteristics of the vehicle body are apparent. The test results show that the proposed DeepTraffic framework works stably and effectively under tested traffic conditions. However, the following problems still need us to improve further: (i) the current prototype framework will be affected by the environment: for example, it cannot work at night because our camera does not support taking night photos. Or during a rainy day, the text on the car body will be deformed due to rain, resulting in a decrease in the accuracy of recognition. (ii) vehicle occlusion significantly affects the accuracy of classification. In addition, more robust algorithms should be researched and explored to solve the light reflection problem to improve the reliability of the system.

References

[1] Leduc, G. (2008) 'Road traffic data: Collection methods and applications', Working Papers on Energy, Transport and Climate Change, 1(55), 1-55.

[2] Ehmke, J. F., Mattfled, D. (2010) 'Data allocation and application for time-dependent vehicle routing in city logistics', European Transport

[3] Fleischmann, B., Gietz, M. and Gnutzmann, S. (2004) 'Time-Varying Travel Times in Vehicle Routing', Transportation Science, 2: 160-173.

[4] Suh, W., Anderson, J., Guin, A., Hunter, M. (2015) 'Evaluation of Video Detection System as a Traffic Data Collection Method', Scientia Iranica 22 (6), S. 2092-2102

[5] Zhang G, Avery RP, Wang Y. (2007) 'Video-Based Vehicle Detection and Classification System for Real-Time Traffic Data Collection Using Uncalibrated Video Cameras', Transportation Research Record.;1993(1):138-147. doi:10.3141/1993-19

[6] Li, S., Yu. H., Zhang, J., Yang, K., Bin, R. (2014) 'Video-based traffic data collection system for multiple vehicle types', IET Intelligent Transport Systems 8(2):164-174

[7] Hasegawa, O., Kanade, T. (2005) 'Type classification, color estimation, and specific target detection of moving targets on public streets', Mach. Vis. Appl.,16, pp. 116–121

[8] Mallikarjuna, C., Phanindra, A., Ramachandra Rao, K. (2009) ‘Traffic data collection under mixed traffic conditions using video image processing’,Transp. Res. Rec. J. Transp. Eng.,135, (4), pp. 174–182

[9] Khan, S.M., Cheng, H., Matthies, D., Sawhney, H. (2010) ‘3D model-based vehicle classification in aerial imagery’, Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 1681–1687

[10] Tzutalin (2015) LabelImg Free Software: MIT License.

[11] Redmon, J., Farhadi, A (2015) ‘You only look once: Unified, real-time object detection’, arXiv preprintarXiv:1506.02640.

[12] Redmon, J., Farhadi, A. (2018) ‘YOLOv3: An Incremental Improvement’, cite arxiv:1804.02767Comment: Tech Report.

[13] Simonyan, K., Zisserman, A. (2015) ‘Very Deep Convolutional Networks for Large-Scale Image Recognition’, CoRR, abs/1409.1556.

[14] Zhou, X., Yao, C., Wen, H., Wang, Y., Zhou, S., He, W., Liang, J. (2017) ‘EAST: An Efficient and Accurate Scene Text Detector’

[15] Baek, J., Kim, G., Lee, J., Park, S., Han, D., Yun, S., Oh, S. Lee, H., (2019) ‘What Is Wrong with Scene Text Recognition Model Comparisons? Dataset and Model Analysis’, International Conference on Computer Vision (ICCV)

Biography

Meng Jin (*1994) holds a Master`s Degree (M.Sc.) in Software Engineering from the University of Stuttgart with majors in computer vision. Since 2018 she has been working as research associate in Research and Innovation Center for Cognitive Service System (KODIS) of the Fraunhofer Institute for Industrial Engineering IAO.

Lars Andreas Mauch (*1987) has been a research associate at the Fraunhofer Institute for Industrial Engineering IAO in Stuttgart since 2017. As a member of the team "Energy Innovation" he focuses on city logistics and the electrification of urban commercial transport. Part of his research is the integration of decentralized and electrified fleets in transport systems. Before joining Fraunhofer IAO, he studied civil engineering at Karlsruhe University of Applied Sciences, where he worked for two years as a research assistant in the field of automated driving and machine-readable infrastructure.

Bernd Bienzeisler (*1964) holds a doctoral degree (Dr. rer. oec.) in business administration and is the head of the Research and Innovation Center for Cognitive Service System (KODIS) of the Fraunhofer Institute for Industrial Engineering IAO. His research focusses on data driven services and business models in different branches and industries.