Commercial Vehicle Traffic Detection from Satellite Imagery with Deep Learning

Moritz Blattner [1] **Michael Mommert** [1] **Damian Borth** [1]

Abstract

Road freight traffic is a major greenhouse gas emitter: commercial vehicles (CVs) contribute ~7% to the global CO_2 emission budget, a fraction that is likely to increase in the future. The quantitative monitoring of CV traffic rates, while essential for the implementation of targeted road emission regulations, is costly and as such only available in developed regions. In this work, we investigate the feasibility of estimating hourly CV traffic rates from freely available Sentinel-2 satellite imagery. We train a modified Faster R-CNN object detection model to detect individual CVs in satellite images and feed the resulting counts into a regression model to predict hourly CV traffic rates. This architecture, when trained on ground-truth data for Switzerland, is able to estimate hourly CV traffic rates for any freeway section within 58% (MAPE) of the actual value; for freeway sections with historic information on CV traffic rates, we can predict hourly CV traffic rates up to within 4% (MAPE). We successfully apply our model to freeway sections in other countries and show-case its utility by quantifying the change in traffic patterns as a result of the first COVID-19 lockdown in Switzerland. Our results show that it is possible to estimate hourly CV traffic rates from satellite images, which can guide civil engineers and policy makers, especially in developing countries, in monitoring and reducing greenhouse gas emissions from CV traffic.

1. Introduction

Driven by an increasing demand in international parcel shipping through e-commerce and other effects, the annual parcel shipping volume passed 100 billion parcels in 2019 and is expected to double within the next five years (Statista, 2020). Road freight, which currently accounts for more than 7% of the annual global CO_2 emissions (IEA, 2021; 2019), will remain the dominant mode of surface freight transportation (Eurostat, 2018) in the near future, likely leading to even higher emissions, as carbon-neutral solutions for heavy modes of transportion have not yet matured (ITF, 2021). Ground-based traffic measurement systems provide quantitative information on the rate and composition of traffic for a given location. Such information is required by civil engineers, policy-makers and other stakeholders for road planning and the implementation of road emission regulations. Switzerland, with its dense road network, has over 500 of such systems in operation, which monitor traffic rates 24/7, while simultaneously categorizing all passing vehicles into ten different vehicle size classes (Bundesamt für Strassen ASTRA, 2009). However, such a dense ground-based detection system is costly in installation and maintenance and thus often unavailable in developing countries, where much of the growth in CV traffic is occurring (Kaack et al., 2018).

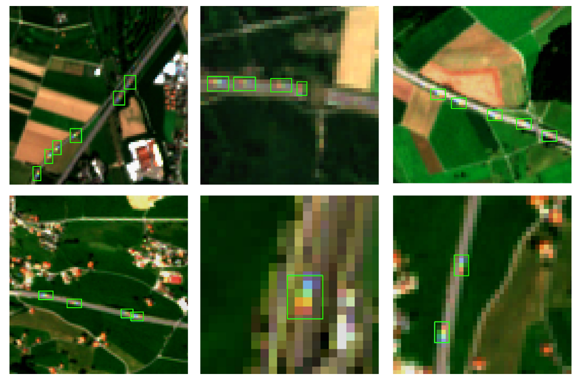


Figure 1. Observations of moving CVs (green boxes) in Sentinel-2 satellite imagery. The "rainbow"-effect is caused by a temporal delay between acquiring the individual bands.

The use of remote imaging data, as provided by the European Copernicus Programme and its two Sentinel-2 satellites, may be utilized to estimate hourly CV traffic rates. CVs and other vehicles are typically too small to be spatially resolved in Sentinel-2 images. However, CVs of a sufficient size and velocity ($\geq$ 70km/h) generate a charac-

[1]Institute of Computer Science, University of St. Gallen, St. Gallen, Switzerland. Correspondence to: Moritz Blattner <moritz.blattner@student.unisg.ch>.

Tackling Climate Change with Machine Learning workshop at ICML 2021.

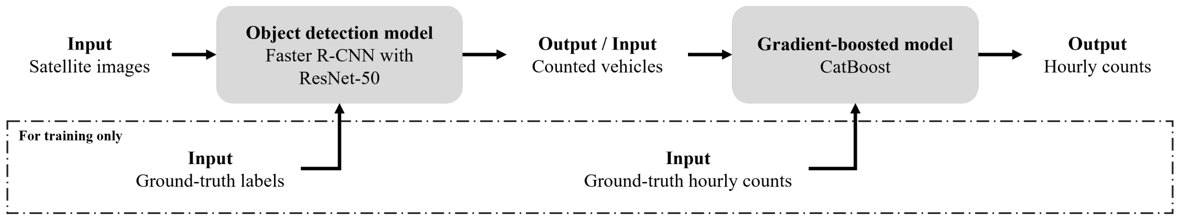


Figure 2. Schematic overview of our pipeline.

teristic pattern due to the temporal offset between different bands during imaging, which facilitates their detection (see Figure 1). This work tries to tackle the missing traffic census in many parts of the world by combining satellite images with deep learning. The goal is to facilitate a new automated monitoring technique to quantify CV traffic rates anywhere in the world.

CV traffic monitoring enables stakeholders to improve road planning and enforce environmental protection regulations to monitor and minimize greenhouse gas emissions from CV traffic, and to identify illegal industrial operations (logging, mining, waste dumps...). Additionally, the data can be utilized for economic analyses, such as the tracking of transportion routes.

2. Related Work

Deep learning techniques are compelling in the task of identifying, localizing and quantifying objects in images. Such methods have been successfully applied on remote sensing data to identify various buildings and construction sites (Dandabathula, 2019; Stankov & He, 2014; Mommert et al., 2021), military facilities (Nur Ömeroğlu et al., 2019), oil tanks (Zhang et al., 2015), tree crowns (Hung et al., 2012), smoke plumes (Mommert et al., 2020) and more. The task of vehicle detection has been approached with high-resolution imaging data and based on single-stage and two-stage object detection methods (Zhou et al., 2020; Kaack et al., 2019). Kaack et al. (2019) went one step further and cross-validated their model with ground-based traffic count stations to make a probabilistic prediction of annual average daily CV traffic.

This work combines the use of freely available Sentinel-2 data with a calibration-based approach on actual CV traffic rate data, making this model globally applicable. The approach to detect commercial vehicles on Sentinel-2 satellite images was first pursued by Fisser (2020), who utilized a multi-thresholding model that compares individual pixel values with predefined thresholds to detect individual CVs on images.

3. Data and Methodology

3.1. Freeway Location Selection

We identify a total of 33 freeway sections in Switzerland that are suitable for our approach by meeting the following criteria: (1) the distance between two successive entry or exit points is sufficiently long and (2) the freeway section in question must be equipped with a traffic measurement system (Bundesamt für Strassen ASTRA, 2009) to provide traffic rate ground-truth data. The selection of freeway sections mirrors the different geographical characteristics within the country. The resulting freeway sections have visible lengths varying between one to more than 10 km; short lengths are sometimes caused by tunnel elements.

3.2. Satellite Image Data

For each freeway section, we download satellite images from ESA's Sentinel-2 satellites taken over the years 2019 and 2020, resulting in 3,670 images. Constrained by the satellites' orbits, observations typically are obtained around noon local time with spacings of at least 5 days. We only utilize those Sentinel-2 bands with 10 m resolution to be able to identify CVs against the background (bands 2, 3, 4 and 8). We split each multi-band image into tiles of 100 px $\times$ 100 px roughly centered on the freeway utilizing a road mask. We annotate all visible CVs on the freeway with rectangular bounding boxes (Tkachenko et al., 2020-2021). The created dataset consists of 5,683 labeled images with 4,686 annotations (individual CVs). Around one third of the images contain CVs, while two thirds show empty roads or cloudy images.

3.3. Traffic Count Data

We obtain ground-truth traffic data from the federal road office in Switzerland (Bundesamt für Strassen ASTRA, 2020). The data consist of hourly traffic rates and distinguish between ten different vehicle categories. Across the selected 33 freeway sections, we have strongly frequented ones with well above 5,000 CVs and buses a day, whereas smaller sections only count few hundreds. We utilize ground-data with the goal to convert CV counts in a single image to an estimate for the hourly CV traffic rate.

3.4. CV Detection

As object detection model, we use a Faster R-CNN architecture (Ren et al., 2015) with a ResNet-50 (He et al., 2016) backbone for feature extraction, similar to Kaack et al. (2019). The network is implemented using PyTorch (Paszke et al., 2019). We adjust image scaling parameters and adapt the anchor box sizes to match the size of the CVs in the images, which only cover a few pixels. We tested different image upscaling values to counteract the down-sampling ratio of the ResNet-50. We ran 49 randomized hyperparameter searches to define the optimal batch size, learning rate, weight decay and momentum. Random image mirroring, flipping, and rotations by integer multiples of 90° are utilized as data augmentations. As evaluation metrics, we utilize precision, recall and mean average precision (mAP, for Intersection-over-Union, IoU, thresholds of 0.1, 0.3 and 0.5). We train the network for 400 epochs on one NVIDIA Tesla V100-SXM2. The main result of the object detection model is a count of CV instances in each image, serving as a snapshot count from which traffic rates are extrapolated.

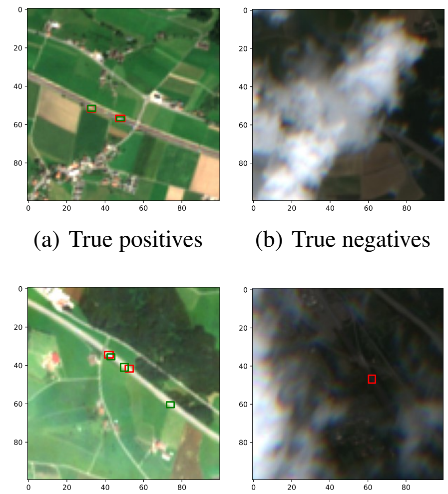
3.5. CV Traffic Rate Regression

For the traffic rate estimation from CV counts as provided by the object detection model, we utilize the gradient-boosted tree-based implementation CatBoost (Prokhorenkova et al., 2018). We enrich the detected CV count with observation specific features, i.e. the weekday and the percentage of the freeway area that is covered by clouds at sensing time. We also include section-specific features, such as the number of lanes, distance to next largest city and more, however, these features did not improve the results and were discarded.

We train the model in a global and a local setting. In the **global** setting we used 70% of the freeway sections as training data and 30% for testing; data from each freeway section is only used in either of the samples, but not both. In the **local** setting we train one model per freeway section and used the same split for the observations per station. The intuition behind these settings is that the global setting applies to unknown freeway sections, whereas the local setting simulates a freeway section for which historic traffic rate information is available, which can be learned by the model. We train the CatBoost model with the objective to minimize the root mean square error (RMSE); we utilize a maximum number of leaves of 64 and a maximum tree depth of 6. We evaluate the results with the mean absolute error (MAE), RMSE and the mean absolute percentage error (MAPE).

Table 1. Object detection model test set results.

Model	**AP @IoU Levels**		
	AP@0.1	AP@0.3	AP@0.5
Faster RCNN	0.711	0.697	0.641
Faster RCNN + data augmentation	**0.723**	**0.703**	**0.650**



(a) True positives (b) True negatives

(c) False negative (d) False positive

Figure 3. Example predictions from our object detection model. Green boxes correspond to ground-truth labels, red boxes correspond to model predictions.

4. Results

4.1. CV Detection

Figure 3 shows example images and corresponding model predictions for bounding boxes around CVs from our trained model. Table 1 lists the AP for the model trained with and without data augmentations for different IoU thresholds. We see that the model trained with data augmentation slightly outperforms the model without those augmentations. Furthermore, AP peaks for an IoU threshold of 0.1, resulting in recall and precision metrics of 0.7 each, as evaluated on our test data set.

4.2. CV Traffic Rate Regression

We test the regression model performance using both the global and local settings on our test data set and receive the results labeled as "baseline" in Table 2. Upon inspection of the results, we found systematic prediction outliers that were due to (1) extremely short freeway sections or (2) non-negligible cloud coverage over the freeway, or both. By filtering freeway sections with lengths $\geq$1 km and cloud coverage $\leq$20%, we are able to achieve significantly better performances (labeled "global" and "local" in Table 2), which we adopt as our final regression models.

Figure 4 illustrates the residual errors between the actual and predicted values for both models. We can see in the global model, that observations with low traffic density tend to be overestimated, while observations with high frequencies are underestimated. Furthermore, we can see that clusters are formed. In the case of the local model, which has seen

Table 2. Traffic rate prediction results.

Setting	R^2	MAE	RMSE	MAPE
Global baseline	0.46	123	180	0.84
Global	**0.65**	**102**	**162**	**0.58**
Local baseline	0.93	34	66	0.09
Local	**0.95**	**33**	**60**	**0.04**

Table 3. Benchmarking of our model compared to other work.

	Fisser (2020)	Our model
# of predictions	329	435
TP / FP / FN	162 / 167 / 219	344 / 91 / 37
Precision	0.49	0.79
Recall	0.43	0.90

("historic") CV traffic rates for the specific freeway section, the model utilizes learned traffic rate means to arrange these clusters on the unity line, which drastically reduces the variance of the predictions.

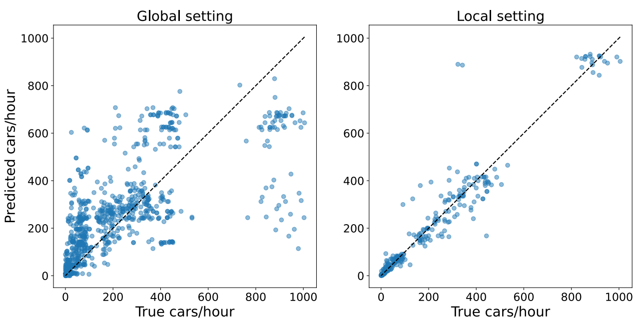


Figure 4. Actual vs. predicted regression results in global and local setting.

5. Discussion

Our **object detection** approach to identifying CVs in Sentinel-2 image data leads to robust detections, resulting in precision and recall scores of 0.7 each. To put these results in context, we compare our results to results obtained by the multi-thresholding model developed by Fisser (2020) on the same data set. Results are summarized in Table 3, revealing that our model performs significantly better. Specifically, we found the Fisser (2020) model to be susceptible to confusing clouds as CVs and to be less robust with respect to different geographical characteristics.

The **regression model** in the global setting achieves reasonable results, while the model in the local setting performs much better. For the local model to be applicable, historic ground-truth traffic rates are required to guide the model to learn average traffic rates for the corresponding freeway section. This behavior is represented well in Figure 4.

5.1. CV Detection: Domain Shift

We test our CV detection approach on different geographical areas, featuring different vegetation and soil types, by randomly selecting freeway sections across five countries (Brazil, US, China, Germany and Brazil) on five different continents. For the former two, we select dense forest areas and for the latter three, we select arid environments. Precision and recall metrics based on manual annotations for these images, are almost identical to those for our original test data set (Section 4.1) and consistent throughout this sample.

5.2. CV Traffic Rates: Impact of COVID-19 Lockdown

We analyze the influence of COVID-19 lockdown measures on CV traffic rates during the first infection wave in Switzerland with a focus on freeway sections in border proximity, as they were predominantly affected. Figure 5 shows the reduction seen over all border areas. Following columns break down the effect for individual stations. We see that the border to Italy (Moleno) was strongly affected, followed by the border to France (Nyon). The border to Germany (Basel) was barely affected by the lockdown measures. In general, our CV traffic rate predictions agree well with the ground-truth measurements.

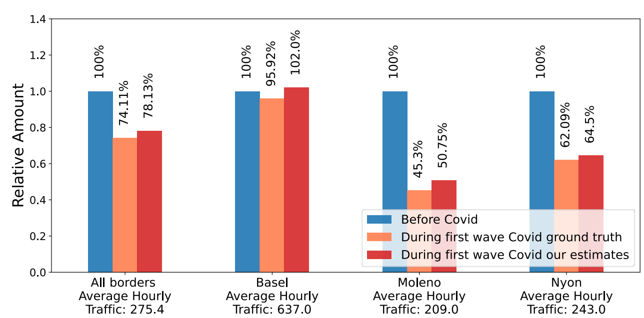


Figure 5. Relative change in CV traffic rates in Swiss border regions during the first COVID-19 lockdown.

6. Conclusions

This work presents an approach to detect CVs and estimate CV traffic rates from Sentinel-2 satellite images. We show that it is possible to measure hourly CV traffic rates for any freeway section with a MAPE of 58% of the true value or with a RMSE of $\sim$160 vehicles per hour. For freeway sections with historic CV traffic rate data, we can predict CV traffic volumes with an RMSE of $\sim$60 vehicles per hour or within a MAPE of 4%. Our model pipeline is suitable to estimate CV traffic rates to support civil engineers and policymakers globally to monitor greenhouse gas emissions.

Acknowledgements

The authors would like to acknowledge the Copernicus Programme and the Swiss *Bundesamt für Strassen* for providing the data utilized in this study.

References

Bundesamt für Strassen ASTRA. Richtlinie Verkehrszähler. *ASTRA 13012*, 1.06:1–24, 2009.

Bundesamt für Strassen ASTRA. Schweizerische automatische Strassenverkehrszählung (SASVZ), 2020. URL https://www.astra.admin.ch/astra/de/home/dokumentation/daten-informationsprodukte/verkehrsdaten/daten-publikationen/automatische-strassenverkehrszaehlung.html. Accessed on 2020-11-10.

Dandabathula, G. Automatic Detection of Overhead Water Tanks From Satellite Images Using Faster-Rcnn. *International Journal of Advanced Research in Computer Science*, 10:34–37, 2019.

Eurostat. Energy, transport and environment indicators. Technical report, European Union, 2018.

Fisser, H. Truck Detection Sentinel2 Upscaling, 2020. URL https://github.com/hfisser/Truck_Detection_Sentinel2_COVID19. Accessed on 2020-10-30.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.

Hung, C., Bryson, M., and Sukkarieh, S. Multi-class predictive template for tree crown detection. *Isprs Journal of Photogrammetry and Remote Sensing*, 68:170–183, 2012.

IEA. Transport sector CO2 emissions by mode in the Sustainable Development Scenario, 2000-2030, 2019. URL https://www.iea.org/data-and-statistics/charts/transport-sector-co2-emissions-by-mode-in-the-sustainable-development-scenario-2000-2030. Accessed on 2021-05-20.

IEA. Global Energy Review: CO2 Emissions in 2020, 2021. URL https://www.iea.org/articles/global-energy-review-co2-emissions-in-2020. Accessed on 2021-05-20.

ITF. ITF Transport Outlook 2021. *OECD Publishing*, pp. 167–168, Chapter 5, 2021.

Kaack, L., Chen, G. H., and Morgan, M. G. Truck traffic monitoring with satellite images. *Proceedings of the 2nd ACM SIGCAS Conference on Computing and Sustainable Societies*, 2019.

Kaack, L. H., Vaishnav, P., Morgan, M. G., Azevedo, I., and Rai, S. Decarbonizing intraregional freight systems with a focus on modal shift. *Environmental Research Letters*, 13(8), 2018.

Mommert, M., Sigel, M., Neuhausler, M., Scheibenreif, L., and Borth, D. Characterization of Industrial Smoke Plumes from Remote Sensing Data. *ArXiv*, 2020. ISSN 23318422.

Mommert, M., Scheibenreif, L., Hanna, J., and Borth, D. IEEE International Geoscience and Remote Sensing Symposium 2021, accepted for publication, 2021.

Nur Ömeroğlu, A., Kumbasar, N., Argun Oral, E., and Ozbek, I. Y. Hangar Detection from Satellite Images with Mask-RCNN and YOLOv2 Algorithms. *27th Signal Processing and Communications Applications Conference (SIU)*, 2019.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S. Pytorch: An imperative style, high-performance deep learning library. In *NeurIPS*, 2019.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., and Gulin, A. CatBoost: unbiased boosting with categorical features. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2018.

Ren, S., He, K., Girshick, R., and Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39:1137–1149, 2015.

Stankov, K. and He, D. Detection of Buildings in Multispectral Very High Spatial Resolution Images Using the Percentage Occupancy Hit-or-Miss Transform. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(10):4069–4080, 2014.

Statista. Global parcel shipping volume between 2013 and 2026 (in billion parcels), 2020. URL https://www.statista.com/statistics/1139910/parcel-shipping-volume-worldwide/. Accessed on 2021-05-20.

Tkachenko, M., Malyuk, M., Shevchenko, N., Holmanyuk, A., and Liubimov, N. Label Studio: Data labeling software, 2020-2021. URL https://github.com/h eartexlabs/label-studio. Open source software available from https://github.com/heartexlabs/label-studio.

Zhang, L., Shi, Z., and Wu, J. A Hierarchical Oil Tank Detector With Deep Surrounding Features for High-Resolution Optical Satellite Imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(10):4895–4909, 2015.

Zhou, L., Liu, J., and Chen, L. Vehicle detection based on remote sensing image of Yolov3. *2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, 1:468–472, 2020.