

Traffic Pattern Analysis at Urban Intersections through Vehicle Detection in Aerial Imagery

Reza Bahmanyar, Jens Hellekes, Manuel Mühlhaus, Veronika Gstaiger, Franz Kurz

Remote Sensing Technology Institute, German Aerospace Center (DLR), Wessling, Germany
{reza.bahmanyar; jens.hellekes; manuel.muehlhaus; veronika.gstaiger; franz.kurz}@dlr.de

Keywords: Aerial imagery, Deep learning, Object detection, Traffic monitoring, Urban mobility.

Abstract

This paper explores the application of aerial image sequences to analyze vehicle flows at urban intersections, with the goal of generating data that can inform adaptive traffic signal timing and improve urban mobility in complex, interconnected road networks. Using deep learning techniques, we detect vehicles in aerial images taken at different times of the day and week at a given urban intersection. This approach allows us to infer vehicle density, identify queuing patterns, and analyze traffic light cycles. Such analysis can assess the robustness of signal timing under different traffic flows and the factors that influence intersection performance. In addition, the use of aerial imagery allows for the derivation of often inaccessible signal timing plans while providing extensive coverage to evaluate surrounding traffic conditions that may impact the target intersections. Ultimately, this research contributes to improved traffic management strategies and supports the development of smart city infrastructure.

1. Introduction

As urbanization continues, efficient traffic management has become increasingly important in promoting mobility and improving the quality of urban life. Traditional approaches to traffic signal timing, defined as the process of selecting appropriate values for timing parameters implemented in traffic signal controllers and related systems, often rely on fixed schedules that may not adapt well to fluctuating traffic conditions. This paper explores the use of aerial image sequences to analyze vehicle flows at urban intersections, generating actionable data to inform adaptive traffic signal timing and improve urban mobility in complex, interconnected road networks.

Aerial imagery offers a promising alternative to traditional data collection methods, which typically include tube counters or induction loops (in-ground installation) or video cameras and radar/microwave/thermal detectors installed at traffic light poles (above-ground). These conventional technologies, often installed near traffic lights, primarily measure actual traffic volumes and can overlook broader traffic dynamics, particularly in saturated conditions where demand exceeds traffic volume, resulting in queuing. In practice, ongoing monitoring and maintenance of signal timing plans requires the collection of periodic flow data to identify changes in traffic characteristics and determine the need for further investigation. Ideally, data collection occurs on a 24-hour per week basis or through peak hour movement counts, while additional traffic characteristics such as travel times and queue lengths are beneficial for developing signal timing strategies (National Academies of Sciences, Engineering, and Medicine, 2015).

Some previous works suggest the use of GPS data to track vehicle locations and speeds, as well as cell phone data for more aggregated insights into travel patterns and traffic conditions (Tzika-Kostopoulou et al., 2024). However, these data sources often suffer from limited availability and accuracy for lane-based analysis. Remote sensing data, particularly from Unmanned Aerial Vehicle (UAV), are increasingly being used in urban traffic monitoring applications (Bouguettaya et al., 2022). However, to the best of our knowledge, no previous



Figure 1. Our target intersection with the overlaid segmentation. Our analysis will focus on the highlighted entry legs.

studies have specifically focused on using aerial imagery data to extract information relevant to signalized intersections. This research aims to fill this gap by determining whether the total cycle length and phase lengths per leg of an intersection can be reliably inferred from aerial image sequences. Such insights could improve the coordination of traffic flow and help urban planners adapt to the demands of modern urban environments. In addition, the availability of large archives of aerial imagery enables new use cases for traffic analysis, making it possible to conduct large-scale urban monitoring and traffic pattern analysis for better urban planning and management.

For our study, we selected an intersection in Brunswick, Germany, that has been frequently imaged in our aerial image

archive and has a high aerial coverage over time. Figure 1 shows an overview of the intersection of interest. We focus our analysis on the four entry legs of the intersection, segmenting the first 150 m of each leg into 30 m tiles. The study legs and segmentations are also shown in Figure 1. We then filter out all existing images from the intersection based on Ground Sampling Distance (GSD), time intervals, and acquisition times, resulting in five image sequences with a total of 2,229 images taken on different days and at different times within a week. Next, we apply a state-of-the-art deep learning method to detect vehicles in the aerial images, and count the number of vehicles in each segment to obtain vehicle density. This density information, analyzed over time, forms the basis for our further investigations.

2. Dataset

To create our dataset, we searched our unique database of 888,346 aerial images collected between 2011 and 2024. This database contains images acquired over Central Europe using the proprietary 3K and 4K camera systems mounted on airplanes and helicopters (Kurz et al., 2012, Kurz et al., 2014). The images vary in GSD from 2 to 30 cm/pixel and cover a wide range of acquisition conditions, viewing angles, times of day, seasons, and camera configurations.

The 3K and 4K camera systems developed by the German Aerospace Center (DLR) are advanced airborne imaging systems currently deployed on specially modified DLR research aircraft. The systems are equipped with three Canon EOS 1Ds Mark II cameras, allowing for optimal ground coverage by aligning each camera according to specific imaging requirements. They support various applications, including traffic monitoring, by capturing high-resolution data. The compact processing unit and precise on-board Inertial Measurement Unit (IMU) enable instant geo-referencing of recorded data with high accuracy. In addition, the data can be transmitted directly to a ground station via a C-band data link, facilitating real-time applications such as rapid situational awareness in urban surveillance and disaster management (Merkle et al., 2020, Wieland et al., 2023).

2.1 Dataset Generation

The study area, a frequently imaged intersection in Brunswick, Germany, was selected for its high aerial coverage over time. Using an intersection operator, 9,003 aerial images covering this location at different times of day and year were identified in the database. These images span several years and were taken sporadically, not at fixed intervals. Figure 2 illustrates the footprint of these aerial images overlaid on a map of Brunswick, with the study area highlighted in a red rectangle. As can be seen, the study area has a high concentration of images, providing extensive temporal insight into this urban intersection.

To refine the dataset for further analysis, images with a GSD greater than 10 cm/pixel and those forming very short sequences (less than approximately 4 minutes) were filtered out. From these, images taken within a single week were selected for analysis, resulting in a total of 2,229 images. These selected images were ortho-projected using measured image positions and orientations. However, due to variations in georeferencing quality, offsets of 1–2 m may exist between images. While these offsets are not further corrected, they are managed by allowing tolerance ranges in subsequent analysis steps. For consistency,



Figure 2. Footprint of 9,003 aerial images (black) overlaid with the study intersection (red rectangle) in Brunswick, Germany.

ID	Date	Day	Start Time	# Images	Duration [s]
1	26.09.2022	Monday	11:49	450	450
2	27.09.2022	Tuesday	13:59	199	199
3	27.09.2022	Tuesday	14:12	248	248
4	29.09.2022	Thursday	12:58	1,140	3,954
5	29.09.2022	Thursday	13:57	192	192

Table 1. Details of the five image sequences used in our analysis.

all ortho-projected images were cropped to the same coordinates to create chronological image stacks for the study area. The selected dataset consists of five image sequences, each recorded at 1 frame per second, spanning three days in September 2022, as summarized in Table 1. Since each sequence may consist of multiple recordings, its total duration may exceed the number of images in the sequence.

Figure 1 provides an overview of the target intersection, focusing on the four entry legs. For detailed analysis, we segment the first 150 m of each leg into five tiles of 30 m each. However, for the WE leg, only four segments are considered, as most images cover only 120 m. The 30 m segment length is chosen based on the average vehicle speed on such roads and the image frame rate, ensuring that each vehicle is visible at least once in each segment during the sequence.

3. Vehicle Detection

In the next step, we apply a Deep Learning (DL) method to detect vehicles in our aerial image dataset. Specifically, we use an adaptation of the state-of-the-art transformer model DINO (Zhang et al., 2022), modified for object detection in aerial imagery using oriented bounding boxes, referred to as DINO_{OBB} (Mühlhaus et al., 2023). We trained the model on the EAGLE dataset (Azimi et al., 2021). To improve its performance and adaptability, we fine-tuned the model on the VETRA dataset (Hellekes et al., 2024), which contains images that are more similar to ours in terms of scenes and GSD. We perform inference on the dataset by applying three different scaling factors (0.5, 1.0, and 1.5) to each image. We use a sliding window of size 1024 × 1024 pixels that moves over each image with a step of 824 pixels. After detecting vehicles, we merge the resulting detections using a Non-Maximum Suppression (NMS) approach to eliminate redundant predictions. Although the training datasets include multiple classes of vehicles, for the purposes of our analysis, we combine all classes into a single vehicle class, focusing on overall vehicle presence rather than class differentiation.

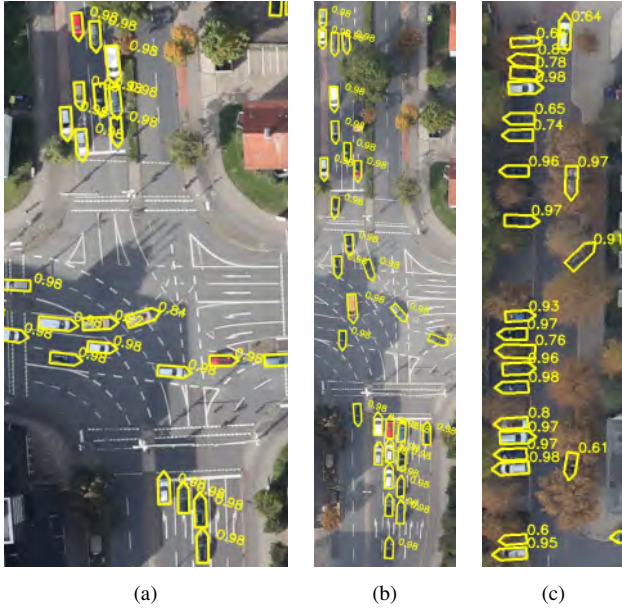


Figure 3. Qualitative analysis of detections: (a) and (b) show the EW and NS directions of the target intersection area, (c) shows a narrow residential street. Detection confidence scores for each bounding box are displayed alongside.

Since our dataset lacks annotations, a quantitative evaluation of vehicle detection is not possible. Therefore, we perform a qualitative evaluation of the detection results. The model trained on the EAGLE dataset achieved a **Mean Average Precision (mAP)** of 78.5% on the EAGLE test set. The reduced complexity of our dataset compared to EAGLE facilitates accurate detections, as shown in [Figure 3a](#) and [Figure 3b](#). However, we observed that the model struggles with strong occlusions, truncated objects, or poor lighting, which are rare in these images, especially in the areas of primary focus. [Figure 3c](#) illustrates the challenges on a narrower residential street, where tree occlusions can lead to wrong detection, such as curbstones being mistaken for vehicles, resulting in lower confidence scores and box localisation accuracy. However, because these problematic regions are outside the intended study area, they do not affect subsequent analysis steps.

4. Results and Discussion

Our primary research question is whether the total cycle length and phase lengths per leg of an intersection can be reliably inferred from aerial image sequences, and whether such information could support the coordination of traffic flow through the intersection. Ideally, frequent monitoring over extended periods should be available to accurately analyse traffic patterns over multiple signal cycles. However, given the sporadic nature of our dataset, our focus is on determining whether intermittent data can still provide meaningful insights into our research question.

In this section we outline our approach to traffic pattern analysis and present our findings. Our analysis focuses on the variation in vehicle density across the segmented areas of the four entry legs at the target intersection, as described in [Section 2.1](#) and shown in [Figure 1](#). To calculate the density, we first determine the centre point of each detected bounding box obtained in [Section 3](#), and then count the points within each segment. By using only the centre points, we compensate for small positional

offsets in the images caused by georeferencing, as explained in [Section 2.1](#). The calculated densities are visualised using colour-coded overlays on the intersection image, together with a density plot over time, as shown in the figures in this section. These visualizations form the basis of our analysis and discussion. Our findings are based on an analysis of all sequences in the dataset, although only selected examples are visualized in this paper to help the reader understand our results.

4.1 Congestion effects at signalised junctions

In [Figure 4](#) we visualize the vehicle density in all segments at three specific times as color-coded overlays on the aerial images of the intersection at those times. As can be seen, the first segment of each entry leg has the highest vehicle density. The figure also illustrates the variation in vehicle density within the first segment of the four intersection entry legs over a 250-second period from Sequence 4, captured on a Thursday early afternoon. Purple dashed lines on the plots mark the three specific time points for which the density overlays are visualized. The visualized overlays show that there are particularly long queues on the WE leg, suggesting a higher traffic flow that the signal timing may not be able to handle effectively. This is supported by the increasing density in the WE direction over a longer period of time, exceeding 200 seconds.

For our analysis, we identified the accumulation and decay phases of vehicle density within each segment by observing the density patterns, with the accumulation and decay periods highlighted in grey and yellow, respectively. The accumulation period is defined as the interval during which the density continuously increases without decreasing, while the decay period includes the remaining time. During the decay period, when vehicles are allowed to cross the intersection, slight increases in density may occur due to vehicle flow, causing minor fluctuations in the plots. In addition, due to variations in driver behaviour and traffic conditions, start and stop times may differ, resulting in slight variations in the duration of the phases in different cycles. It is noticeable that the accumulation and decay times for the EW, NS and SN legs follow similar patterns over the cycles. This is in contrast to WE, where the derived lengths of the phases and frequency of cycles appear to vary greatly and are sometimes very short. A more in-depth investigation is required here; therefore, [Section 4.2](#) investigates the individual phases for each direction of driving.

4.2 Signal timing plans and traffic anomaly detection

Each leg of the intersection is equipped with a pre-sorting area, which allows vehicles to enter the correct lane dependent on their direction of travel. For WE, the number of lanes per direction is displayed in [Figure 5](#). Through visual inspection of the aerial imagery, we determined the start and end times of the red phase per traffic light for each sorting area. During the remaining time, vehicles are allowed to proceed through the intersection, which we indicated with a green color. Many traffic signals also provide drivers with cues, the use of amber lights, flashing lights, countdowns, etc. to indicate an approaching red phase, but this information is not visible in aerial imagery and is therefore not included in our analysis. We define the start of the red phase as the moment when the first vehicle stops at the stop line, although the actual transition may occur slightly earlier. Similarly, the start of the green phase is marked by the first vehicle starting to move, although driver reaction times may introduce a slight delay. While these transitions may differ

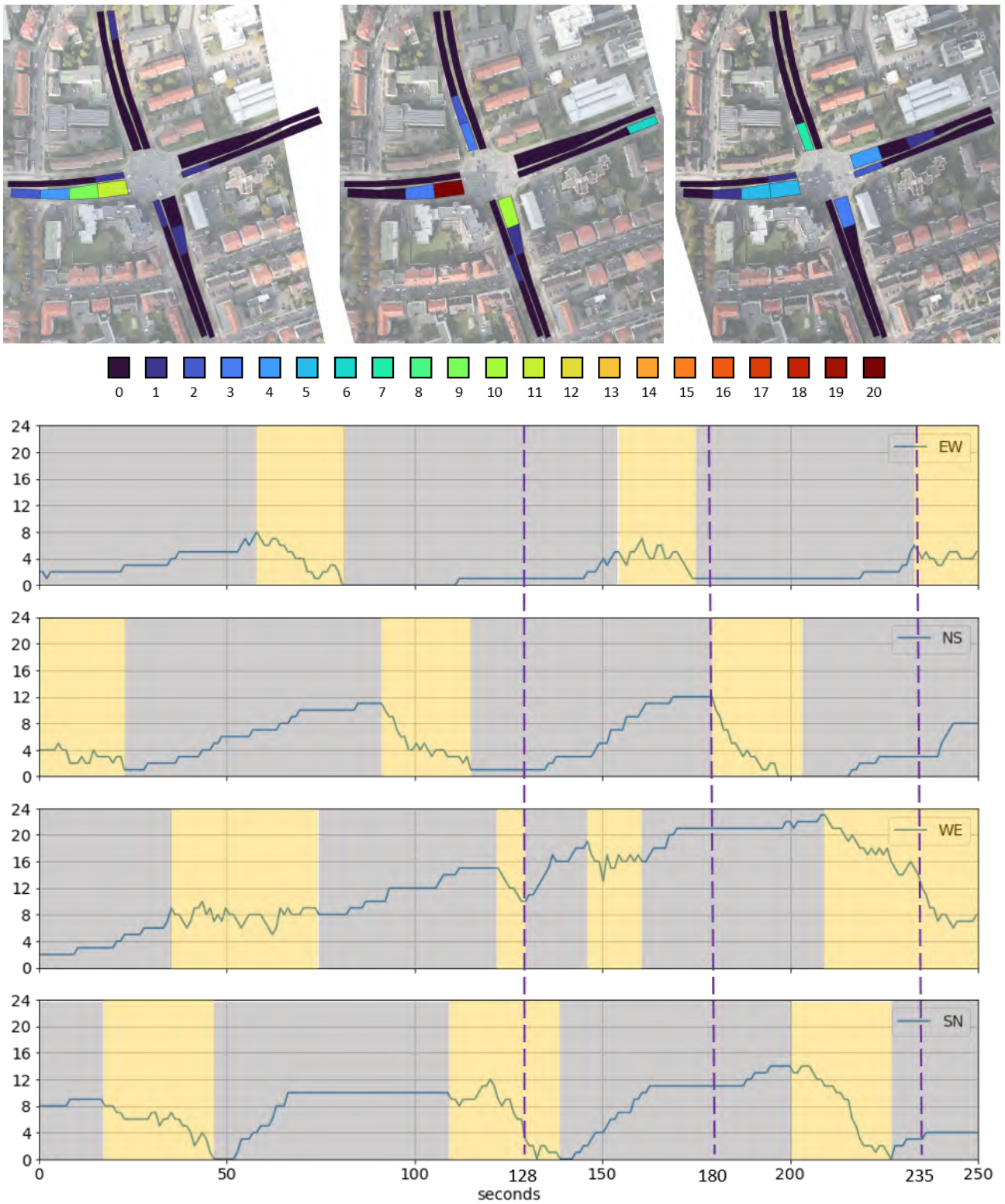


Figure 4. Vehicle densities in the incoming and outgoing legs of the junction at three points in time (marked by purple dashed lines). The total number of vehicles in the first segment of each incoming leg is shown below. The accumulation and decay times are indicated in grey and yellow, respectively.

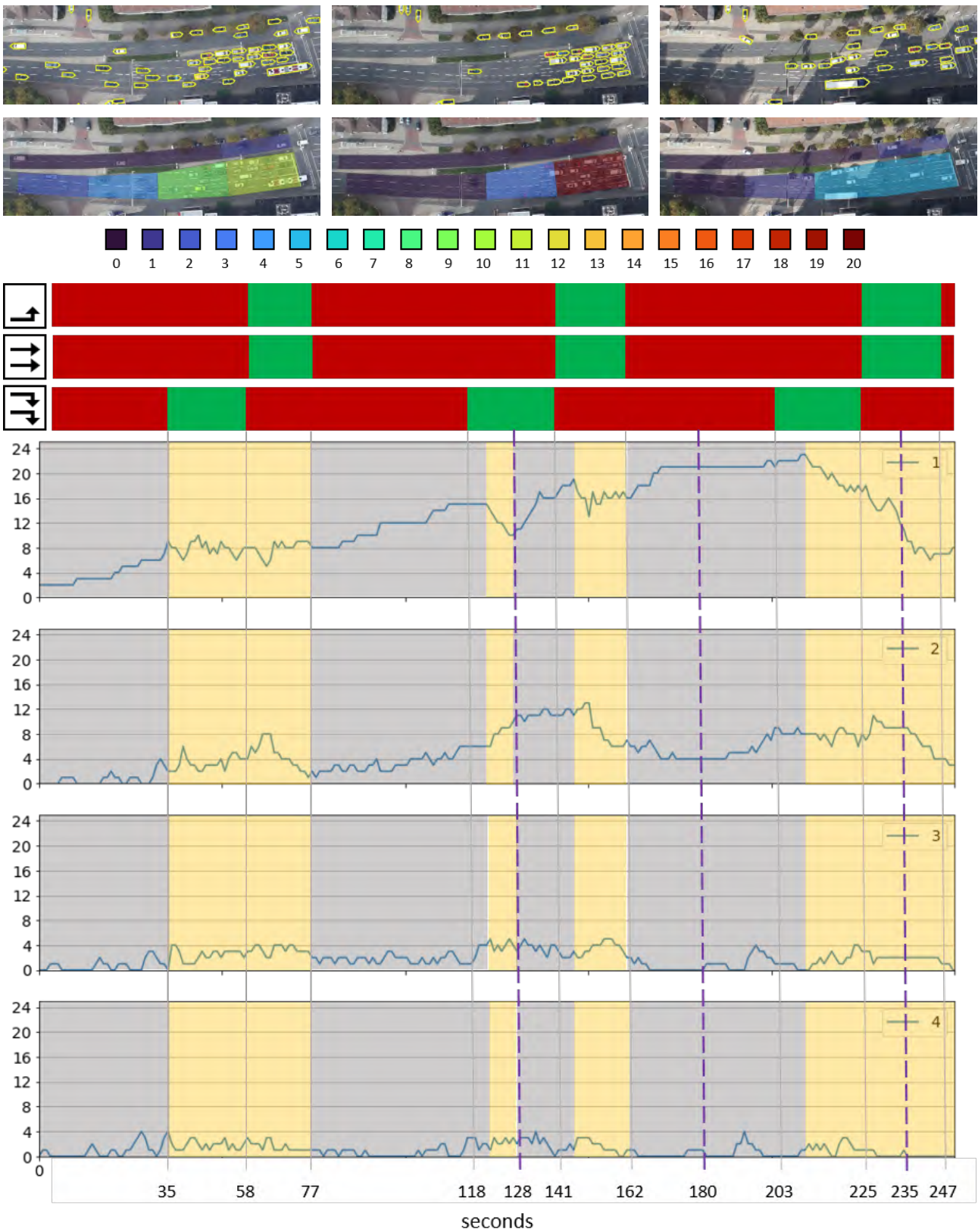


Figure 5. Vehicle detections and derived densities in the segments of the WE leg. The signal timing plan per direction and the number of vehicles in the individual segments over time is displayed below. The vehicles approach the junction by travelling through segments 4, 3, 2 and 1.

slightly from the exact times, any minor discrepancies do not affect the overall accuracy of our analysis.

A comparison of the traffic light cycles with the previously determined accumulation and decay times shows a general agreement. Since not all directions are switched to green at the same time, the density of vehicles in the first segment increases despite the green phase for right-turning vehicles in the second 128 of the observation. This phenomenon can be attributed to the fact that vehicles from segments 2, 3, and 4 enter segment 1 with a delay. This delay can be seen in both the density plots for these segments and the visualization of detected vehicles in the upper left of the figure. Since the flow of vehicles through WE is controlled by another upstream traffic signal, there is potential for adjustment to increase vehicle throughput and prevent congestion. The significant increase in vehicles entering the first segment between seconds 128 and 141 suggests that a greater proportion of these vehicles are traveling straight or making left turns.

When looking at the vehicle densities per segment, it can also be seen that the inflow via segments 3 and 4 is similarly high over time, although there is a congestion effect. The increased vehicle density is initially only visible in segment 1, later also in segment 2, resulting in fewer vehicles being able to pass through the junction than the number of arriving vehicles. A closer analysis of the aerial images reveals that this is mainly due to left-turning traffic (see aerial images at the center and right side). This traffic first has to wait for oncoming traffic before it can continue its journey. As a result, around three vehicles can enter the junction area before the corresponding traffic light switches to red. Due to the oncoming traffic in the EW direction, the waiting vehicles can only leave the junction area afterwards. If the intention of traffic engineers is to prevent the build-up, it could be examined whether the green phase for these vehicles can be reduced, given the traffic load from the oncoming direction. It could also be checked whether the left-turn green phase can be extended. Another option could be to introduce a second left-turn lane; however, this would imply infrastructural adjustments to the outgoing northern arm. In order to be able to refer to a larger data basis for decision makers, Section 4.3 investigates the potential of long-term airborne observations.

4.3 Recognisability of changes in signal timing plans by long-term observation

As shown in Table 1, Sequence 4 consists of multiple consecutive recordings, allowing for a long-term analysis of traffic patterns. Figure 6 displays the vehicle densities over the course of almost 66 min, whereby the recordings are paused by the helicopter regularly turning off, reducing the load on the engines and then approaching the intersection again. The general characteristics as described in Section 4.1 can also be observed here: for legs EW, NS and SN, the vehicle density is reduced to small number of vehicles remaining in the green phases, while for leg WE the demand exceeds the capacity for most of the observed cycles.

Since the time interval between the local maxima of the vehicle densities (which corresponds to the cycle length) and the duration of accumulation and decay remain steady, it can be concluded that – at least during the time covered by Sequence 4 – no significant changes in the signal timing plan were taken, e.g. by traffic-actuated traffic lights. Section 4.4 investigates whether such patterns are observable during the course of multiple days.

4.4 Robustness in the event of increased traffic flows over the course of several days

In Section 4.2 we concluded that the left-turn appears to be the bottleneck given the traffic flows per direction of travel. For leg SN, the left-turn design capacity is higher as two lanes allowing for this movement (see Figure 3a). In addition, the signal timing is designed so that vehicles can turn left directly without waiting for oncoming traffic (see Figure 3b).

Figure 7 exhibits the traffic flow for the SN leg over the course of several days on different days of the week. The vehicle densities between the different observations differ in some cases by a factor of more than 2, although the green phase appears to be sufficiently long at all observation times, so that only few vehicles remain in segment 1. It is noticeable that the green phase is always of the same length, but the speed at which vehicles clear the segment varies. Based on the overlaid change rates in Figure 7, the vehicle speed appears to be lower for shorter queues, while vehicles drive through the junction area faster when there are more vehicles.

5. Summary and Outlook

This paper demonstrates the potential of aerial image sequences for analyzing vehicle flows at urban intersections, providing valuable insights for adaptive traffic signal timing and improving urban mobility. The results show that significant traffic characteristics can be extracted even from short observation periods. By combining multiple image sequences taken at different times of the day and week, a comprehensive understanding of traffic patterns emerges.

However, there are limitations to this approach. While aerial imagery can be scaled effectively to cover larger areas than traditional camera systems, the high effort required to analyze just a few intersections suggests that this method is best suited for initial assessments and periodic monitoring. More in-depth analysis could then be conducted using stationary sensor technology to collect more precise data.

Looking forward, future work could extend this approach to multiple intersections and explore the interactions between traffic light cycles. The use of semantic segmentation methods can help to scale the analysis to large areas with multiple intersections. In addition, lane-wise segmentation allows a more detailed analysis of the traffic situation. Integration of this approach with other data sources such as traffic sensors or GPS data could be explored to demonstrate the complementary effect of this approach. In addition, the inclusion of other modes of transport could lead to the development of comprehensive traffic volume maps for intersections. The integration of object tracking techniques could enable automated movement counts that differentiate by intersection leg and direction. Aerial image sequences taken over extended periods of time can provide a wealth of information; the application of Multi-Object Tracking (MOT) algorithms could facilitate advanced traffic engineering applications, including corridor progression analysis, capacity allocation, and safety assessments related to potential conflict points. Such developments will enhance the capacity for data-driven decision making in urban traffic management, ultimately contributing to more efficient and safer urban environments.

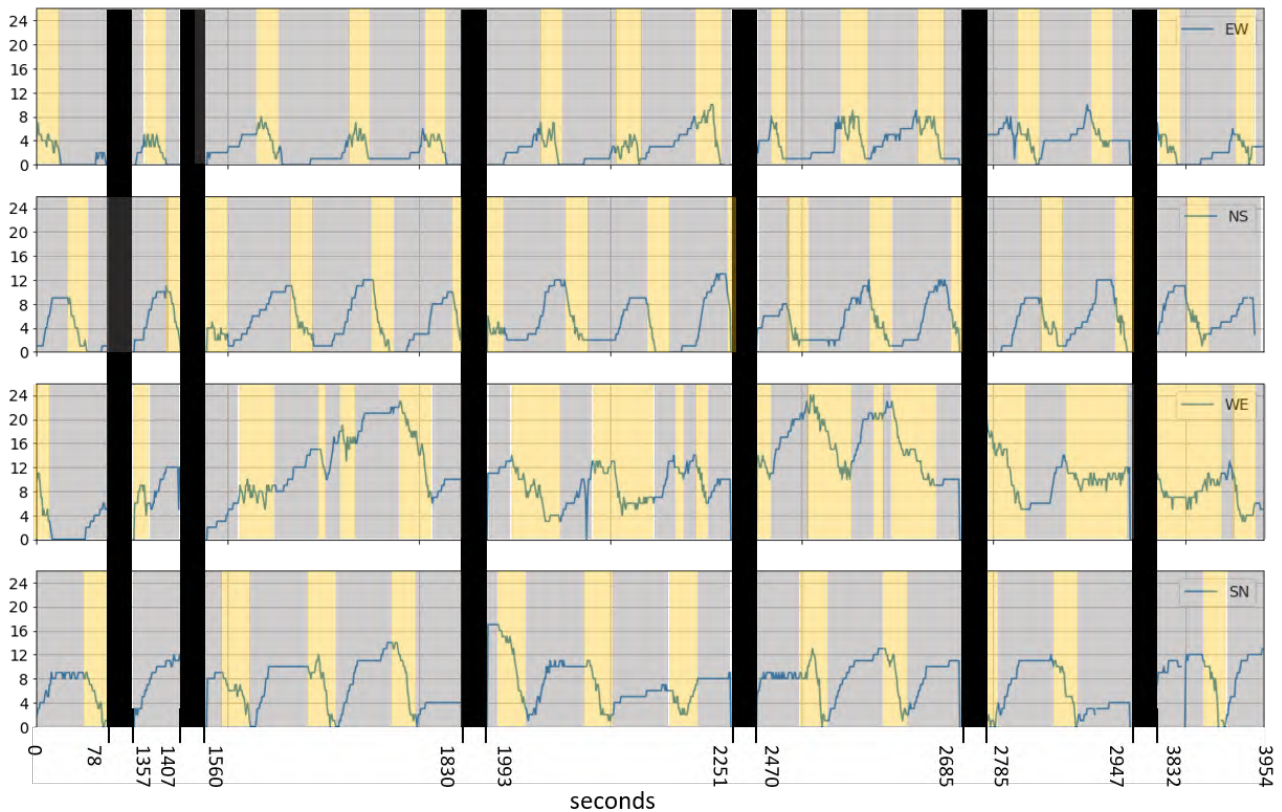


Figure 6. Vehicle densities in the first segment of the incoming legs throughout multiple subsequent recordings, resulting in a total time span of almost 66 minutes.

References

- Azimi, S. M., Bahmanyar, R., Henry, C., Kurz, F., 2021. EAGLE: Large-Scale Vehicle Detection Dataset in Real-World Scenarios using Aerial Imagery. *Proceedings of the International Conference on Pattern Recognition (ICPR)*, IEEE.
- Bouguettaya, A., Zarzour, H., Kechida, A., Taberkit, A. M., 2022. Vehicle Detection From UAV Imagery With Deep Learning: A Review. *IEEE Transactions on Neural Networks and Learning Systems*, 33(11), 6047-6067.
- Hellekes, J., Mühlhaus, M., Bahmanyar, R., Azimi, S., Kurz, F., 2024. VETRA: A Dataset for Vehicle Tracking in Aerial Imagery - New Challenges for Multi-Object Tracking. *Proceedings of the European Conference on Computer Vision (ECCV)*, Springer.
- Kurz, F., Rosenbaum, D., Meynberg, O., Mattyus, G., Reinartz, P., 2014. Performance of a real-time sensor and processing system on a helicopter. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-1, 189-193.
- Kurz, F., Türmer, S., Meynberg, O., Rosenbaum, D., Runge, H., Reinartz, P., Leitloff, J., 2012. Low-cost optical Camera Systems for real-time Mapping Applications. *Photogrammetrie - Fernerkundung - Geoinformation*, 2012(2), 159-176.
- Merkle, N. M., Gstaiger, V., Schröter, E., d'Angelo, P., Azimi, S., Kippnich, U., Barthel, C., Kurz, F., 2020. Real-time aerial imagery for crisis management: Lessons learned from an european civil protection exercise. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII, Copernicus Publications, 1243-1249.
- Mühlhaus, M., Kurz, F., Guridi Tartas, A. R., Bahmanyar, R., Azimi, S. M., Hellekes, J., 2023. Vehicle Classification in Urban Regions of the Global South from Aerial Imagery. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*.
- National Academies of Sciences, Engineering, and Medicine, 2015. *Signal Timing Manual - Second Edition*.
- Tzika-Kostopoulou, D., Nathanail, E., Kokkinos, K., 2024. Big data in transportation: a systematic literature analysis and topic classification. *Knowledge and Information Systems*, 1-26.
- Wieland, M., Merkle, N., Schneibel, A., Henry, C., Lechner, K., Yuan, X., Azimi, S. M., Gstaiger, V., Martinis, S., 2023. Ad-hoc situational awareness during floods using remote sensing data and machine learning methods. *IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium*, 1166-1169.
- Zhang, H., Li, F., Liu, S., Zhang, L., Su, H., Zhu, J., Ni, L. M., Shum, H.-Y., 2022. DINO: DETR with Improved DeNoising Anchor Boxes for End-to-End Object Detection.

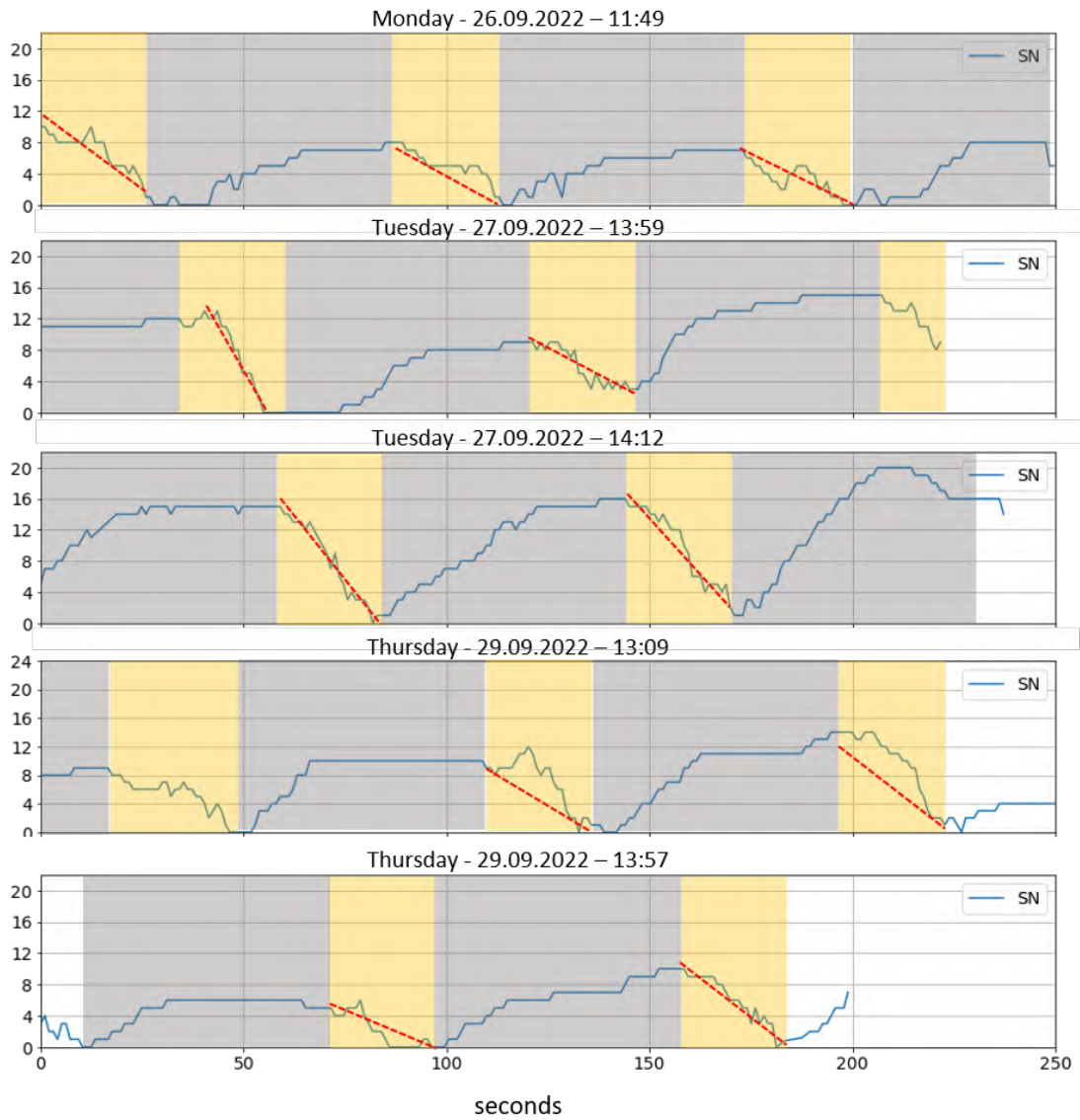


Figure 7. Vehicle densities in the first segment of the SN leg in the course of multiple days. The density change rate is indicated by red dotted lines.