

# Real-time Lane-wise Traffic Monitoring in Optimal ROIs

Mei Qiu, Wei Lin, Lauren Ann Christopher, Stanley Chien\*, Yaobin Chen, Shu Hu  
Purdue University Indianapolis, IN, USA

**Abstract**—In the US, thousands of Pan, Tilt, and Zoom (PTZ) traffic cameras monitor highway conditions. There is a great interest in using these highway cameras to gather valuable road traffic data to support traffic analysis and decision-making for highway safety and efficient traffic management. However, there are too many cameras for a few human traffic operators to effectively monitor, so a fully automated solution is desired. This paper introduces a novel system that learns the locations of highway lanes and traffic directions from these camera feeds automatically. It collects real-time, lane-specific traffic data continuously, even adjusting for changes in camera angle or zoom. This facilitates efficient traffic analysis, decision-making, and improved highway safety.

## I. INTRODUCTION

Numerous Pan, Tilt, and Zoom (PTZ) traffic cameras are installed along highways in the USA, allowing operators to monitor traffic conditions. However, human operators oversee hundreds of cameras, making it impossible to watch them all simultaneously. Thus, there is interest in using Artificial Intelligence (AI) to analyze footage in real time, providing valuable traffic data and alerting operators to issues.

Traffic monitoring systems come in various types, broadly classified into two groups: traditional systems and intelligent management systems [1], [2]. AI and big data analytics have revolutionized traffic monitoring and management. Intelligent Transportation Systems (ITS), equipped with advanced sensors, radars, and license plate recognition cameras play a vital role in detecting and deterring traffic rule violations [3], [4]. These systems also aid in alleviating traffic congestion by managing traffic scenarios through real-time data visualization [5], [6].

However, these existing traffic monitoring systems need to annotate images from a large video surveillance data. Most of them can monitor two directions of traffic status but are not lane-specific. There is some work with lane-wise counting, but others rely on road lane marking detection, which fails in various lighting, weather, and ground conditions [7]. Vehicle motion trajectories can be used to learn lanes [8] but the performance depends on vehicle detection and long tracking accuracy. The innovation in this work is to perform detection and tracking only in optimal regions of the road. Our new result also solves the problem of existing traffic flow estimation that requires fixed zoom and viewing angles [7], [9], [10].

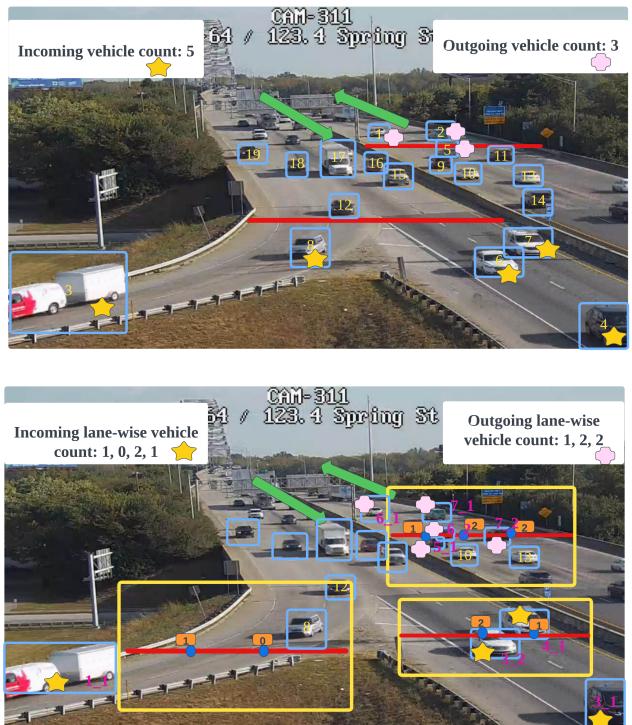


Fig. 1: (Top) **Traditional System:** Vehicles are detected, tracked, and counted across the entire frame, with manually defined counting lines, leading to suboptimal performance. (Bottom) **Our system:** Vehicle detection is performed across the entire frame. Vehicle tracking is concentrated in several automatically learned Regions of Interest (ROIs), optimizing detection and tracking results. Counting is lane-specific, with each vehicle assigned a unique LaneID, moving away from bi-directional counting. **Best view in color.**

This paper introduces a Real-Time Automatic Lane Learning and Lane-Specific Traffic Status Monitoring System, building upon our prior research [11], [12]. In our previous work, we developed a method for automatically identifying highway locations, lane boundaries, and traffic flow directions from video footage captured by cameras. Leveraging this acquired knowledge, the current study focuses on the real-time collection and reporting of traffic data for each lane, including metrics such as vehicle counts (illustrated in Fig. 1), flow rates, and congestion estimates derived from vehicle count statistics. Furthermore, this system is equipped to autonomously detect changes in the camera's angle or zoom level, subsequently reinitiating the road and lane learning process to maintain

\*Corresponding Author

continuous and accurate traffic status monitoring.

The contributions of our work can be summarized as follows:

- Building on our earlier contributions in LCD [11] and MRLL [12], this paper introduces an advanced, adaptive system capable of **lane-wise** vehicle counting, flow rate calculation, and traffic status detection that operates continuously, around the clock. This system is designed to accommodate cameras of varying resolutions and frame rates and is robust against a wide range of weather and traffic conditions, utilizing real-time video streams as its primary input source.
- We have developed a novel, standalone module named “Camera View Checking” that operates continuously to monitor for any changes in the camera’s angle or view. It accomplishes this by comparing the current camera view with an initial, well-defined reference view. Should any deviation in the camera angle or view be detected, this module automatically triggers the system to re-initiate the learning process for road and lane parameters.
- We devised a novel approach, termed “Video Rate-Computer Speed Synchronization” to dynamically adjust the input frame rate in accordance with the system’s processing capabilities, thereby ensuring the maintenance of real-time performance.
- We have developed a region-based tracking system leveraging an enhanced DeepSort architecture, specifically designed for real-time tracking in the most effective Regions of Interest (ROIs) to ensure optimal vehicle detection. Key improvements include the integration of our “Video Rate-Computer Speed Synchronization” method, which dynamically adjusts input frame rates based on processing speed to sustain real-time operations. We have also transitioned from using the traditional Intersection Over Union (IOU) metric to the more advanced Complete IOU (CIOU) distance, significantly improving the precision of detection and tracking data association. Furthermore, the introduction of an adaptive matching threshold has been instrumental in optimizing the system’s tracking accuracy.
- We have thoroughly evaluated the effectiveness of our proposed lane-wise vehicle counting system by comparing it with manual counting in 9 videos.

## II. RELATED WORKS

### A. Object Detection

Traditional object detection methods like background subtraction [14] and image difference [15] have been replaced by deep learning, due to superior detection accuracy and speed. Deep learning methods fall into two categories: two-stage detectors (e.g., Faster R-CNN, Mask R-CNN [16], [17]) and one-stage detectors (e.g., YOLO, YOLOv4 [18], [19]). Two-stage detectors offer superior accuracy, while one-stage detectors provide faster inference without pre-generated region proposals. These deep learning-based object detection algorithms are extensively employed in traffic surveillance videos, delivering satisfactory performance [20], [21].

### B. Multiple Object Tracking

Most object-tracking algorithms fall into two categories: detection-based tracking (tracking-by-detection) and detection-free tracking. In recent years, tracking-by-detection methods have been predominant. These methods involve performing object detection on each frame to obtain detection results, which are then associated with adjacent frames to form trajectories. Popular data association methods include the Hungarian algorithm and the Kuhn-Munkres algorithm [22]. Post-processing techniques like soft non-maximum suppression (NMS) [23] are often applied to smooth and refine the trajectories. The Intersection Over Union (IOU) [24] metric is commonly used to associate detection results based on bounding box overlaps between frames. SORT [25] and Deep SORT [26] are two widely used tracking methods in traffic surveillance videos [27], [28]. SORT combines Kalman Filter with the Hungarian algorithm and relies on bounding box size and position for motion estimation and data association. Deep SORT utilizes a CNN to extract appearance information, enhancing the association metric with motion information. This allows Deep SORT to track objects more effectively during occlusions, reducing ID switches but also increasing computational costs.

## III. METHOD

This section summarizes our system, described in [11], [12], that analyzes highway surveillance camera footage in real-time to detect changes in camera angles or zoom levels. It then learns the number of lanes and their locations, providing lane-wise vehicle counting, flow rate, and congestion estimation in optimal ROIs with best vehicle detection. The system, depicted in **Fig. 2**, runs on a local computer with a single GPU and comprises two main components: **Environment Learning** for adapting to unpredictable camera settings and **Road Condition and Traffic Status Detection** for analyzing lane-based traffic information.

### A. Environment Learning

The goal of this component includes finding the locations of the lanes and optimal locations to detect and count vehicles on the lanes. Instead of relying on non-robust lane-markings in highways, we adopted a method that uses the vehicle motion information on the video to find the lane locations because most vehicles on the highway are driven within their lanes [11], [12].

**Camera View Checking** The program monitors for changes in camera angle or zoom level which would invalidate learned lane information. The efficient ImageHash algorithm [29] is used to assess significant and rapid background changes in the top fifth of the image frame (the sky and far side of the road), to signal potential camera adjustments. If the Hash Distance value exceeds a certain threshold, it is considered a camera angle or zoom level change, prompting a relearning process. This check occurs every 50 frames.

The algorithm is proven to be very robust in various camera motions and lighting conditions. The example testing cases are shown in Fig. 3, such as when the cloud background changed

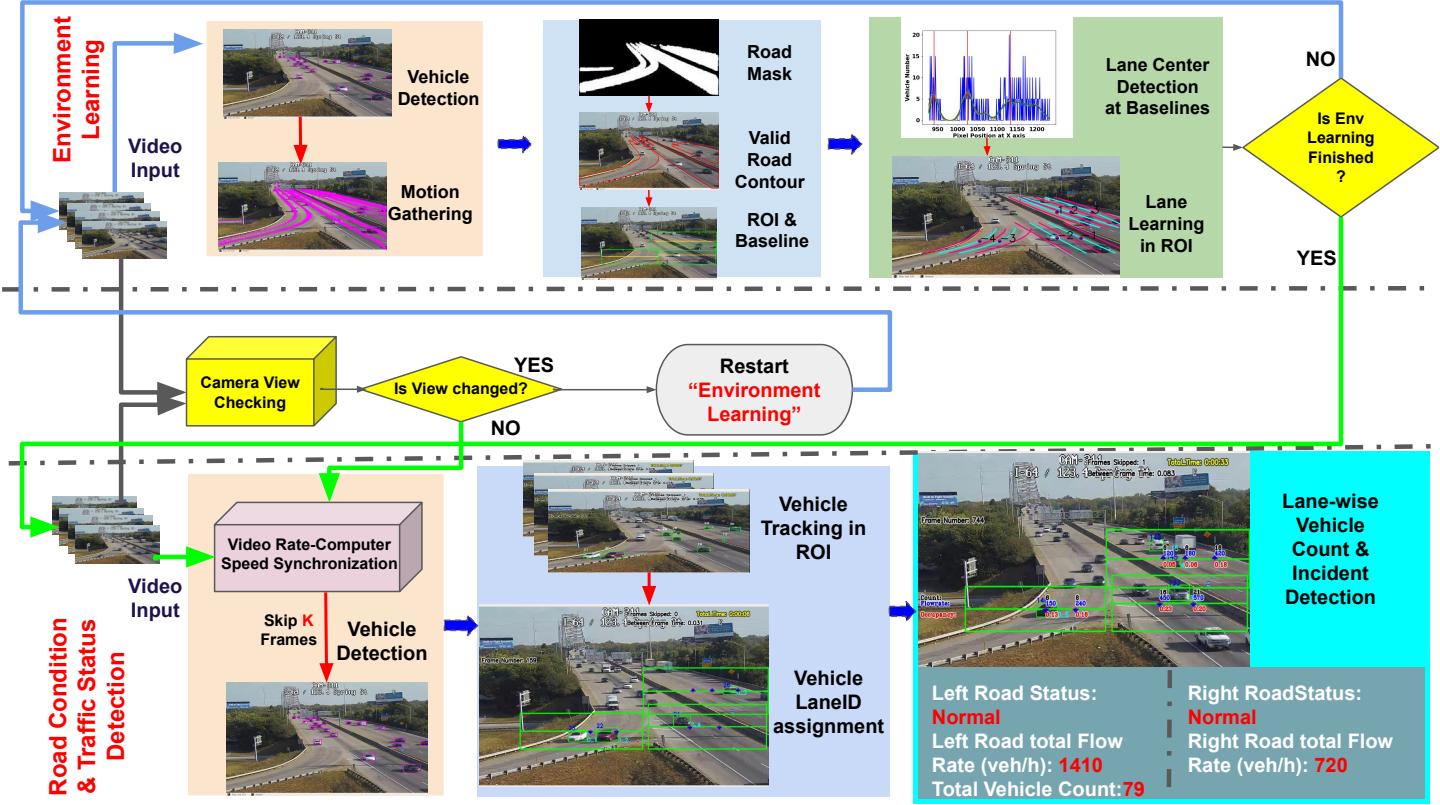


Fig. 2: The system we designed is running on every local computer with a single GPU. This system gets real-time stream input from a single camera. Since cameras are not fixed on highways, the **Environment Learning** module will learn the road and lane information based on vehicle motion trajectories (the details can be found in our previous work: [11]–[13]). When these parameters about road and lanes are learned for a particular view (can be defined by customer), the system will go to the next stage, to do **Road Condition and Traffic Status Detection** via each lane. An independent module (i.e., **Camera View Checking**) keeps running all the time to check camera angle changes or not by comparing with the first, well-defined view. Once the system detects the camera angle/view is changed, the **Road Condition and Traffic Status Detection** module stops working immediately and the system goes to the **Environment Learning** to learn a new set of road and lane information parameters. More details are explained in Section III.

significantly, when the two input videos were taken over a long time apart, or when the camera zoomed out or zoom in, under all of these tough PTZ camera situations, our algorithm gives the correct detection with the default Hash threshold. This algorithm detects the large camera angle shifts and also detects those less apparent camera parameter changes. There are some cases that our algorithm cannot handle well, as shown in Fig. 4. In summary, our camera viewing checking algorithm works well in the majority of our ITS environments. The system monitors the camera's viewing angle. If it detects changes in angle or zoom level, it halts and restarts the environment learning stage to relearn the road and lane parameters from the video.

#### B. Road Condition and Traffic Status Detection

Learning the ROI and lane boundaries makes lane-based traffic status detection possible. The issues faced and solutions in lane-based traffic status detection are described in the following subsections.

1) **Video Rate-Computer Speed Synchronization:** The goal of this system development is that it can be used for all highway surveillance cameras. Since the frame rate (frame per second:  $fps$ ) and resolution of these cameras can be quite different, the processing time for image processing can vary significantly. We reduce the  $fps$  processed based on the computer computation speed (assume the time cost of each frame is  $\delta t$  with the unit of second) by adaptively skipping  $K$  frames for every frame processed during the process following Equation (1):

$$K = \begin{cases} \frac{1}{\delta t} - \frac{1}{fps} & \text{if } \frac{1}{\delta t} - \frac{1}{fps} < 0 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

We designed a Video Rate-Computer Speed Synchronization method that observes the computer processing time for each input frame and the incoming frames per second. Then we use this information to determine how many frames should be skipped for each frame processed so there will be no backlog frames. Since the processing time for each frame

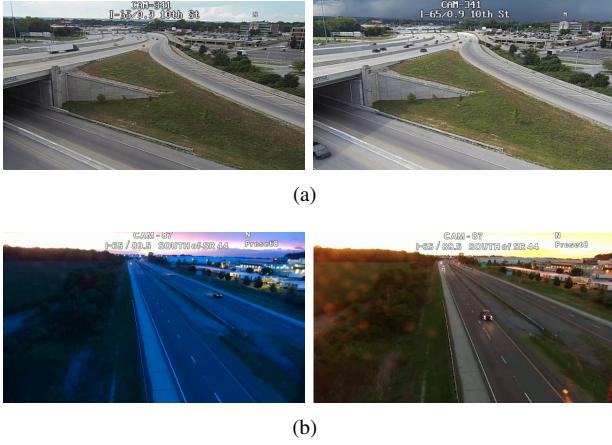


Fig. 3: *Correct Camera View Change Detection Cases.* (a) The cloud background of the right frame changed significantly from the left frame but did not cause false detection. The system did not report the camera view change. This is **correct detection**. (b) The left frame is taken at dawn and the right frame in the afternoon. The system did not report the camera view changes. This is **correct detection**.



Fig. 4: *Wrong Camera View Change Detection Cases.* (a) The left frame is taken in the morning and the right frame is taken at night. The algorithm gives false detection because of the flare in the second frame is background. The Hash value distance returned by our algorithm is larger than the threshold. This is **wrong detection**. (b) The algorithm predicts camera angle is changed in the right frame from the left frame while it is not. But it is very tough for humans as well to predict in this kind of scenario. This is **wrong detection**.

is affected by the vehicles captured and processed in the frame, the number of frames skipped fluctuates over time.

2) *Vehicle Detection:* The pre-trained YOLOv4 is used in this work and the input to vehicle detection is no longer continuous but with  $K$  frames skipped.

3) *Online Multi-target Tracking with Adaptive Kalman Filter:* Before tracking, we filter out invalid detected vehicles,

removing those with bounding box sizes more than twice the median truck size and those outside any ROI or roads. NMS is applied to eliminate redundant bounding box overlaps. DeepSort [26] is then employed for online multiple vehicle tracking on the remaining valid detection in each frame. However, the classic tracking pipeline has limitations in our application. Firstly, real-time performance on a single GPU is essential, but feature extraction from a deep learning model at each frame is time-consuming. Secondly, our proposed skipping frame strategy renders the IOU distance and fixed threshold in DeepSort inadequate, as detection and tracklets may not overlap. Therefore, we modify DeepSort to meet our application's requirements.

#### We modified Deepsort: By removing object appearance feature matching and using CIOU distance instead of IOU distance in Cascade Matching step in DeepSort

By importing the frame-skipping strategy, the whole system achieves real-time processing status without much tracking accuracy loss. However, in the default, Deepsort, the IOU distance used in the last step for matching detected vehicles and tracks will not meet the requirement if the new detected vehicles do not overlap. So the distance between the bounding box's centers and the consistency of the aspect ratio has the potential to solve this problem. Based on the work CIOU loss [30], we replace the IOU distance with CIOU distance. The original IOU distance is defined as Equation (2) and IOU is the intersection-over-union of bounding boxes in detections and tracks. The modified CIOU distance is defined as shown in Equation (3).

$$d_{IOU} = 1 - IOU \quad (2)$$

$$d_{CIOU} = 1 - IOU + D + \alpha * V \quad (3)$$

In CIOU, the normalized central point distance  $D$  is designed to measure the distance of two boxes as calculated in Equation (4),

$$D = \frac{\rho^2(p^d, p^t)}{c^2} \quad (4)$$

where  $p^d = [x_d, y_d]^T$  and  $p^t = [x_t, y_t]^T$  are the central points of boxes in detections and tracks,  $c$  is the diagonal length of each box in the track, and  $\rho$  is specified as the Euclidean distance function.  $V$  is defined as the consistency of the aspect ratio and calculated as:  $\frac{4}{pt^2}(\arctan \frac{w^d}{h^d} - \arctan \frac{w^t}{h^t})$ . The trade-off parameter  $\alpha$  is defined the same as with the default CIOU. The CIOU distance mitigates the missing match when detections are not overlapping with predicted tracks and are more robust to the scale of bounding boxes.

#### The matching distance threshold was adjusted with the skipped frame number $K$ in Cascade Matching step in DeepSort

Based on the above analysis, a fixed IOU distance threshold does not work well because of different frames are skipped in each iteration. So we adjust the CIOU distance threshold to  $prethre * (K + 1)$ , where  $K$  donates the skipped frame number and  $prethre$  is the default fixed matching threshold in Deepsort, if no frame is skipped,  $K$  equals 0.



Fig. 5: 9 ITS traffic scenes recorded in sunny, rainy, snowy, nighttime, and congestion traffic conditions.

After tracking, each vehicle passing through the ROI is assigned a unique  $Fid$ .

4) *Lane-wise Vehicle Counting:* The result of vehicle counting with time information can be proportionally converted to flow rate and valuable traffic status information. To realize lane-wise vehicle counting, we associate every tracked vehicle  $i$  in the ROI with  $Fid_i$  to a unique lane ID  $Lid_i$  using the lane boundary information obtained in the environment learning stage. When assigning the lane ids to each vehicle in the current frame, only the previous frame is compared. This strategy will not only reduce the time cost and also reduce the id switch when we assign the lane ids if the id switch happens in FIDs in a longer trace. If the center of a tracked vehicle is within the boundary of a lane and within the average car length at the baseline, the vehicle is counted for that lane. A tracked vehicle is only counted one time.

5) *Traffic Status Detection:* The system can generate traffic status and incident reports for the users in three steps: (i) determine the flowrate of each lane based on the vehicle counts over a specific time interval periodically (ii) estimate the percentage of pixels occupied by all vehicles on each lane within the ROI as occupancy rate as shown in the **Equation (5)**, and (iii) traffic status checking based on the combination of the flow rate and occupancy.

$$Occp_l = \frac{\sum(h_l)}{H} \quad (5)$$

, here  $l$  is the lane ID, and  $h_l$  is the sum of all vehicles' bounding box height in lane  $l$ . For every  $T$  minute, the instantaneous flow rate  $Fr_l$ (vehicle per hour) of lane  $l$ , is calculated based on the vehicle counting  $C_l$  of lane  $l$  during  $T$ :  $Fr_l = \frac{C_l * 60}{T}$ . The time interval for the flow rate calculation cannot be too long since it will not show the short-term traffic status changes. On the other hand, the time interval for the flow rate calculation cannot be too short since it will fluctuate too much to understand its meaning.

#### IV. EXPERIMENTS

##### A. Experimental Settings

**Datasets.** For vehicle count testing purpose, we created another 9 videos and each video lasts about 2 minutes. These

testing videos have unique camera views and various weather, traffic density and visibility as shown in Fig. 5. We assume each video has finished the ROIs and lane learning including: lane center locations, lane directions, lane boundaries in all the ROIs.

**Evaluation Metrics.** To estimate our system's performance in vehicle counting in each lane, we manually counted 9 videos from four scenes including sunny, rainy, night, and congested traffic, and each video lasts 2 minutes. We counted vehicles in each lane separately when every vehicle passes the baseline. The flowrate of ground truth is also estimated with the counting ground truth. The total counting accuracy of one video is defined as the corrected counting percentage:  $\frac{\text{total system count}}{\text{total ground truth count}} * \%$ . We consider the smaller count as the numerator and the higher the value means the higher the counting accuracy. We design two other matrices to estimate the road level flow rate accuracy compared with the flow rate generated from the ground truth counting: **MEA**(mean estimation accuracy):  $\frac{\sum(Fr_l)}{\sum(FrGT_l)}$  and **RMSE**(root mean square error):  $\sqrt{\frac{\sum(Fr_l - FrGT_l)^2}{m}}$ . We define the  $FrGT_l$  as the ground truth flow rate at lane  $l$  and  $m$  as the total number of lanes. The MAE has a similar definition of counting accuracy, we regard the smaller value as the numerator and the higher value means the estimation is closer to the ground truth. The RMSE estimates the whole error of estimation, and the lower value means better counting and flow rate estimation. To check the traffic status, we defined an estimate rule based on the flow rate and occupancy rate of each lane as shown in Equation (6). Based on the rules we defined, our system gets 100% accurate traffic status (normal, slow, and jam) reports of all the lanes of total 9 videos.

$$\text{status}_l = \begin{cases} \text{Jam} & \text{if } Fr_l < 600 \text{ and } Occp_l > 0.6 \\ \text{Slow} & \text{if } 600 < Fr_l < 900 \text{ and } 0.4 < Occp_l < 0.6 \\ \text{Normal} & \text{otherwise} \end{cases} \quad (6)$$

**Manual Count.** We hire 2 master students to count the actual vehicles crossing the counting line manually. For each video, they count the vehicles in each video on a regular interval which is 30 seconds and totally count 4 intervals. And we also record flow rate value, and occupancy in each lane during the same interval calculated automatically by our designed system. Finally, we analyze how these two indicators change over time and do the comparison.

**Implementation Details.** We utilized YOLOv4 [19], a popular deep learning object detector, to detect vehicles in camera frame streams. During the reference phase, we set detection confidence score thresholds to 0.25 and IOU threshold to 0.45. For tracking, we employed Deep Sort framework, excluding the CNN feature extractor. Cosine distance metric facilitated track association in each frame, with an initial IOU distance threshold of 0.35. Parameters like  $max\_age$  (30) and  $n\_init$  (3) were set to control track deletion and initialization phases, respectively. Coupled with our frame-skipping strategy, the

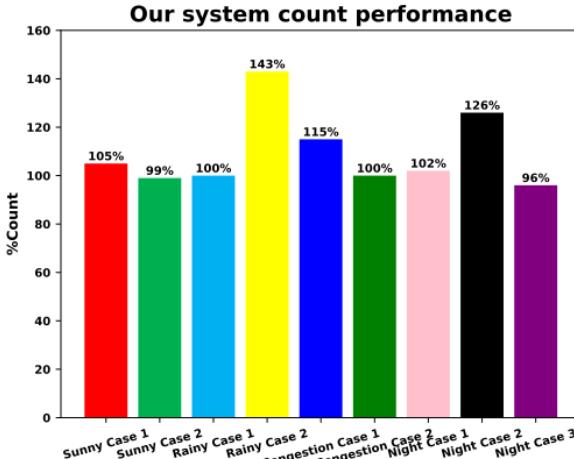


Fig. 6: The overall count percentage for all vehicles from 9 cases by our counting system. Lower than 100% means miss count, larger than 100% means over count, and 100% means count by our system is totally the same as the ground truth.

entire system ran in real-time on a single NVIDIA Quadro RTX 5000 GPU.

## B. Results

**Lane-wise Real-time Vehicle Counting.** The lane detection in this system performs with acceptable accuracy when not in totally dark conditions. Experimental results on our own collected highway data demonstrates the effectiveness of our proposed framework in lane-wise vehicle counting by comparing with manual ground truth. The system count of 7 of 9 videos is close to the ground truth, and lowest counting error is 2% as shown in Fig.6. However, the performance decreases in low lighting conditions, poor viewing angles, or with occlusions.

**Lane-wise Flow Rate Estimation.** When estimating our system’s lane-wise flow rate estimation accuracy, we only consider the cases in which all the lanes are correctly learned. The flow rate estimation accuracy and error results of the 8 videos are shown in TABLE I.

These results prove our system achieves good flow rate estimation accuracy and the highest MAE reaches **0.95** and it reaches 0.86 even in night cases. The average RMSE is lowest in the sunny and daytime cases which means the lighting is still the major factor affecting the detection and counting.

**Road Condition and Traffic Status Detection.** Based on the rules we defined, our system gets 100% accurate traffic status (normal, slow, and jam) reports of all the lanes of all 9 videos.

## V. CONCLUSION

Highway surveillance cameras face challenges due to unpredictable changes in viewing direction and zoom level. To address this, a lane-based automatic traffic monitoring system with a lane learning component has been developed and proved its success using real-time Indiana Highway data. The paper outlines the system’s architecture and highlights

TABLE I: Accuracy of lane-wise flow rate estimation in various scenarios.

Video Name	MEA ↑	RMSE ↓
Sunny 1	<b>0.95</b>	76.49
Sunny 2	0.93	54.38
Rainy 1	0.93	47.43
Rainy 2	0.77	228.47
Night 1	0.86	118.59
Night 2	0.93	<b>36.74</b>
Congestion 1	0.86	354.96
Congestion 2	0.9	129.03
Average	0.89	130.63

the importance of vehicle motion for lane detection, vehicle counting and traffic status estimation.

**Limitations.** The performance of our system depends critically on the accuracy of its various sub-modules, which include lane learning, vehicle detection, tracking, LaneID assignment, and counting. Specifically, lane learning, which is primarily informed by video input and vehicle motion, tends to be less effective in situations such as traffic jams or when vehicles remain stationary. The task of vehicle detection faces significant challenges in the Intelligent Transportation System (ITS) environment, where adverse weather conditions and low-light conditions at night can significantly impair detection capabilities.

In Multi-Object Tracking (MOT), one of the major hurdles is dealing with occlusions, particularly in dense traffic scenarios where bounding boxes of different vehicles overlap, making it difficult to track individual vehicles accurately. Moreover, the use of heavy-weight computational architectures for tracking hampers the ability to perform real-time tracking, necessitating the development of lighter-weight alternatives that do not sacrifice accuracy.

To accurately assess the overall performance of our system, it is essential to conduct tests across a broader range of scenarios to validate the system’s robustness. This entails incorporating more diverse test cases that can effectively simulate the wide variety of real-world conditions under which the system must operate.

**Future Work.** To improve the overall accuracy of the system, it is crucial to make incremental improvements to each individual module. Additionally, employing a more diverse collection of test videos will be instrumental in accurately estimating the system’s performance.

**Acknowledgment.** This work was supported by the Joint Transportation Research Program (JTRP), administered by the Indiana Department of Transportation and Purdue University, Grant SPR-4436. The authors would like to thank all INDOT Study Advisory Committee members, including Jim Sturdevant, Ed Cox, Tim Wells, for their guidance and advice throughout the project. The authors would also thank Zhengming Ding, Stephan Gerve, Upadhyay, Aniket Pankaj, Prathamesh Somkant Panat, Anup Atul Mulay, Baudouin Ramsey, Kunal Mandil, Arjun Narukkanchira Anilkumar, and Kavya Prasad for their contributions in various part of this project.

## REFERENCES

- [1] Tamer Nadeem et al. Trafficview: A scalable traffic monitoring system. In *IEEE International Conference on Mobile Data Management, 2004. Proceedings.* 2004, pages 13–26. IEEE, 2004.
- [2] Neeraj Kumar Jain et al. A review on traffic monitoring system techniques. *Soft computing: Theories and applications: Proceedings of SoCTA 2017*, pages 569–577, 2019.
- [3] Mahmoud Abbasi et al. Deep learning for network traffic monitoring and analysis (ntma): A survey. *Computer Communications*, 170:19–41, 2021.
- [4] Igor Bisio et al. A systematic review of drone based road traffic monitoring system. *IEEE Access*, 2022.
- [5] Pallavi A Mandhare et al. Intelligent road traffic control system for traffic congestion: a perspective. *International Journal of Computer Sciences and Engineering*, 6(07):2018, 2018.
- [6] Vishal Mandal et al. Artificial intelligence-enabled traffic monitoring system. *Sustainability*, 12(21):9177, 2020.
- [7] Xingchen Zhang et al. Monocular visual traffic surveillance: a review. *IEEE TITS*, 2022.
- [8] Adriel Isaiah Valeroso Amoguis et al. Road lane segmentation using vehicle trajectory tracking and lane demarcation lines. In *Proceedings of the 2023 6th International Conference on Machine Vision and Applications*, pages 64–71, 2023.
- [9] Aleksandr Fedorov et al. Traffic flow estimation with data from a video surveillance camera. *Journal of Big Data*, 6:1–15, 2019.
- [10] Li Li et al. Trajectory data-based traffic flow studies: A revisit. *Transportation Research Part C: Emerging Technologies*, 114:225–240, 2020.
- [11] Mei Qiu et al. Intelligent highway lane center identification from surveillance camera video. In *ITSC*, pages 2506–2511. IEEE, 2021.
- [12] Mei Qiu et al. Intelligent highway adaptive lane learning system in multiple rois of surveillance camera video. *IEEE TITS*, 2024.
- [13] Mei Qiu et al. Attention mechanism improves yolov5x for detecting vehicles on surveillance videos. In *2022 IEEE AIPR*, pages 1–8. IEEE, 2022.
- [14] Quin Cai et al. Tracking human motion in an indoor environment. In *ICIP*, volume 1, pages 215–218. IEEE, 1995.
- [15] Daniel J Dailey et al. An algorithm to estimate mean traffic speed using uncalibrated cameras. *TITS*, 1(2):98–107, 2000.
- [16] Shaoqing Ren et al. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015.
- [17] Kaiming He et al. Mask r-cnn. In *ICCV*, pages 2961–2969, 2017.
- [18] Joseph Redmon et al. You only look once: Unified, real-time object detection. In *CVPR*, pages 779–788, 2016.
- [19] Alexey Bochkovskiy et al. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*, 2020.
- [20] Hang Shi et al. Anomalous driving detection for traffic surveillance video analysis. In *ICIST*, pages 1–6. IEEE, 2021.
- [21] Soumi Mitra et al. Towards an optimised vehicle detection algorithm for multi-object tracking in traffic surveillance. In *AICS*, pages 200–211, 2021.
- [22] Harold W Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955.
- [23] Navaneeth Bodla et al. Soft-nms-improving object detection with one line of code. In *ICCV*, pages 5561–5569, 2017.
- [24] Erik Bochinski, Volker Eiselein, and Thomas Sikora. High-speed tracking-by-detection without using image information. In *2017 14th IEEE international conference on advanced video and signal based surveillance (AVSS)*, pages 1–6. IEEE, 2017.
- [25] Alex Bewley et al. Simple online and realtime tracking. In *2016 IEEE ICIP*, pages 3464–3468. IEEE, 2016.
- [26] Nicolai Wojke et al. Simple online and realtime tracking with a deep association metric. In *2017 IEEE ICIP*, pages 3645–3649. IEEE, 2017.
- [27] Chong Liu et al. City-scale multi-camera vehicle tracking guided by crossroad zones. In *CVPR*, pages 4129–4137, 2021.
- [28] Lijun Yu et al. Traffic danger recognition with surveillance cameras without training data. In *AVSS*, pages 1–6. IEEE, 2018.
- [29] David Marr and Ellen Hildreth. Theory of edge detection. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 207(1167):187–217, 1980.
- [30] Zhaohui Zheng et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation. *IEEE Trans. Cybern.*, 52(8):8574–8586, 2021.