TrafficNight: An Aerial Multimodal Benchmark For Nighttime Vehicle Surveillance

Guoxing Zhang[1], Yiming Liu[2], Xiaoyu Yang[1], Hailong Huang[1], and Chao Huang[1]

[1] The Hong Kong Polytechnic University, Hong Kong, China
[2] Southeast University, Nanjing, China

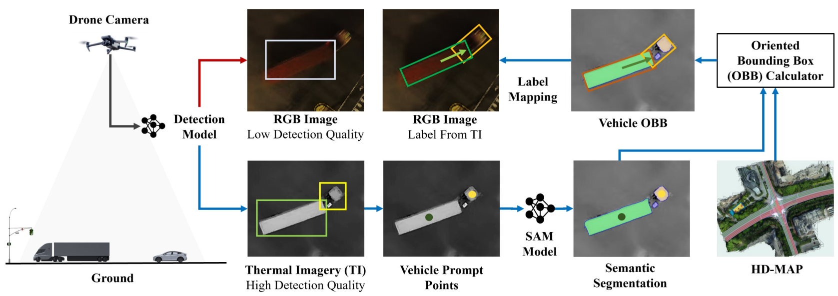


Fig. 1: Annotation Process Overview. Data is collected using UAVs with thermal and sRGB sensors. The process starts with detecting Horizontal Bounding Boxes(HBB) in thermal images, followed by the Segment Anything Model(SAM) outlining vehicles accurately. Oriented Bounding Boxes(OBB) are then created using the OBB Calculator and high-definition maps(HD-MAP) for precise spatial context. The TrafficNight dataset process results in a comprehensive collection of synchronized sRGB and thermal images, OBB annotations, HD-map data, and raw video files.

Abstract. In autonomous simulation and surveillance, realistic scenarios are crucial for advancing object detection algorithms. Existing aerial datasets suffer from sample class imbalance, especially in larger vehicles like trucks, and unrealistic lighting conditions. This hampers progress in driving behavior analysis and imitation. To address these limitations, we introduce a novel multimodal vehicle surveillance dataset, integrating aerial thermal infrared and sRGB imagery. It contributes: (1) A novel thermal infrared vehicle detection benchmark, ensuring robust object detection in nighttime lighting conditions. (2) Thermal infrared surveillance videos paired with corresponding HD-MAPs for improved multi-vehicle tracking. (3) Specialized annotations for semi-trailers, precisely documenting their movement trajectories and physical coordinates. TrafficNight significantly advances understanding of larger vehicles in traffic dynamics, serving as a benchmark for enhancing Autopilot systems and traffic surveillance in challenging environments. [3]

Keywords: Nighttime · Thermal infrared · Traffic flow

[3] See TrafficNight project webpage for the code and more.

1 Introduction

Recent advancements in autonomous vehicle (AV) and Simulation systems owe significantly to the evolution of training datasets,particularly aerial observations. These offer unique insights into traffic dynamics, vehicle behaviors, and spatial relationships, often missed in ground-level data [4, 8, 10, 23]. The aerial perspective is key for algorithmic accuracy in applications ranging from traffic management to autonomous navigation.

However, existing aerial surveillance datasets [16,30,31] face two main limitations: a bias towards ideal lighting conditions and a skewed representation of vehicle types. The former can compromise the safety and efficiency of autonomous systems through detection errors. Despite efforts in data augmentation and algorithmic adjustments, the nuances of natural lighting are often inadequately replicated. The underrepresentation of larger vehicles, affects the understanding of traffic flow. Remedies like oversampling and synthetic sample generation [5,20] have been used but often introduce new biases or inadequately capture complex behaviors. Our response to these challenges is a comprehensive aerial multimodal dataset, integrating vertical RGB and thermal infrared imaging. It is designed to address diverse lighting conditions and ensure balanced vehicle representation, aiming to provide a realistic foundation for advancing object detection research. To address limitations in existing aerial surveillance datasets, we present TrafficNight—a comprehensive aerial multimodal dataset that integrates vertical RGB and thermal infrared imaging. In Table 1, we compare TrafficNight with other widely used datasets, including B3D, Citysim, INO, and LLVIP. Our dataset excels in providing a balanced representation across diverse lighting scenarios and vehicle types, which is crucial for training robust computer vision models under real-world conditions.

Table 1: Compare with existing data content. TrafficNight excels in multimodal representation. Simultaneously, it demonstrates strengths in semi-trailers, nighttime, and 3D tools, crucial for detailed object detection in diverse real-world conditions

	Data	Semi-trailer	Nighttime	3D Mapping
B3D [3]	sRGB	✗	✗	✗
Citysim	sRGB	✗	✗	✗
INO	sRGB,cIF	✗	✗	✗
LLVIP [12]	sRGB,cIF	✗	✗	✗
Our	sRGB,cIF,IF	✓	✓	✓

We addressed key challenges in annotation, overcoming conventional methods' limitations in scalability and precision. Our innovations include: (1) An Automated Annotation Framework, integrating a target detection algorithm with Meta's SAM model for efficient vehicle segmentation in infrared images, as in Fig.1. (2) Thermal infrared surveillance videos with corresponding HD-MAPs

(a) Vehicle Type Interference **(b)** U-turn at intersection **(c)** Multi-Veh Curve Trailing

Fig. 2: Nighttime typical scenarios. TrafficNight enriched with semi-trailers sample. Encompasses a large number of scenarios with mutual interference and vehicle following behaviors on curved road segments.

enhance multi-vehicle tracking. (3) A specialized model for accurately labeling semi-trailers, identifying tractor and carriage sections separately. (4) A algorithm for converting 2D image coordinates into 3D spatial coordinates, crucial for calibrating Car Following Models (CFM) parameters.

2 TrafficNight

2.1 Dataset Construction And Structure

Data collection, focused in Shenzhen, China, utilized the DJI Mavic 3T drone equipped with RTK technology for accurate geospatial coordinates and real-time acquisition. Each scene is uniquely identified by a SceneID, featuring synchronized sRGB and thermal image pairs. Details of video durations and image counts per scene are listed in Table. 2. The dataset preserves original thermal image data, allowing developers to select their preferred color mapping techniques. Areas with high truck and semi-trailer traffic, particularly in low-light conditions, were prioritized for data collection. These locations are urban areas representative of typical regions where large vehicles, such as trailers, are frequently found. This approach allowed us to capture diverse urban traffic scenarios, including those involving numerous semi-trailers.

The dataset includes critical scenarios for traffic management and autonomous driving research. For instance, a multi-vehicle interference scenario, shown in Fig.2a, demonstrates interactions among various vehicles, aiding traffic flow model development. A scenario in Fig.2b, capturing a semi-trailer executing a U-turn, offers insights for behavior prediction algorithms in intersection situations. Additionally, Fig. 2c depicts semi-trailers navigating turns, crucial for trajectory prediction in dynamic environments. These scenarios are instrumental in advancing trajectory and behavior prediction research, improving traffic flow models, and enhancing the safety and efficiency of autonomous driving systems.

Raw Videos: The raw video data in this dataset comprises synchronized sRGB and thermal images, captured using the DJI M3T drone's integrated sensors, as depicted in Fig. 3. The sensor configuration consists of a Complementary

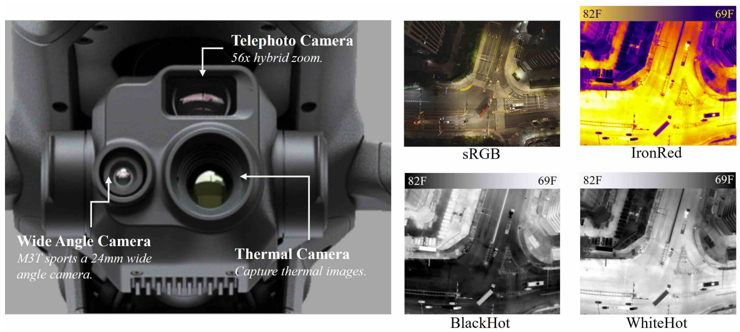


Fig. 3: Sensor intergretion (left) and **Color palette of raw video** (right). Visualization results after using different color palettes on the same thermal infrared data.

Table 2: All collection result

SenceID	Video's ID	Duration	ImagePair
TN01	TN10262209	18min	237
TN02	TN10272135,TN10272206	32min	405
TN03	TN10290421,TN10290446	30min	388
TN04	TN10281958,TN10282031	36min	237
TN05	TN10292044,TN10292118	36min	134
TN06	TN10312124,TN10312150	35min	179
TN07	TN11012013,TN11012037	28min	400

Metal Oxide Semiconductor (CMOS) for sRGB data and an uncooled Vanadium Oxide (VOx) sensor for thermal imaging. sRGB videos are recorded in low-light mode at 3840x2160 resolution and 30 frames per second (fps), while the thermal videos, captured in low gain mode (0-500°C) to enhance thermal detail, have a resolution of 1280x1024 at 30fps.

R-JPEG Images: This format preserves the original temperature data of objects, which can inspire other researchers to develop higher-performance visual algorithms based on this raw temperature data. It is significant to mention that there are no standard post-processing tools to modify the color grading of these thermal videos,as existing tools are primarily designed for image data. However, our development kit includes the ColorPalette.py script, which addresses this gap. Using ColorPalette.py script to represent temperatures in thermal imaging data is shown in Fig. 3.

HDMAP: In developing our High-Definition Mapping (HDMAP) annotations, we utilized high-accuracy point cloud data, acquired through multi-view reconstruction techniques using the Mavic M3T drone equipped with Real-Time Kinematic (RTK) systems. RTK technology enables the capture of highly precise point cloud data, crucial for detailed mapping. The point cloud's color visualization, shown in Fig.4, facilitated the identification of lane markings for

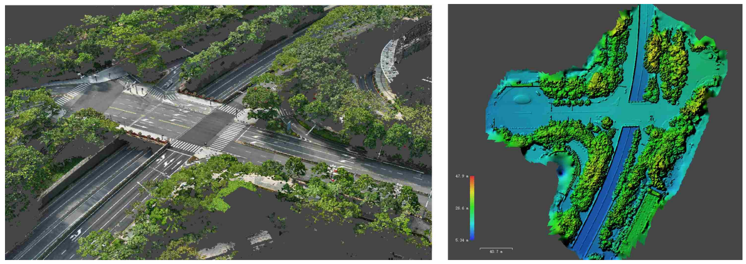


Fig. 4: The color visualization of the point cloud and the Digital Surface Model (DSM)

HDMAP. Additionally, we generated a digital surface model (DSM) from the point cloud to represent terrain height variations, as illustrated in Fig.4. Employing this high-fidelity point cloud data, we used the RoadRunner tool for map creation. Renowned for its precision, RoadRunner enabled us to produce detailed HDMAPs that accurately represent the physical layout of scenes and include essential topological information like lane positions and relationships, vital for research in autonomous driving and traffic flow analysis.

File Structure: The dataset comprises seven traffic scenes, encompassing intersections and road segments. It contains a comprehensive array of original videos, totaling 150GB from both thermal infrared and sRGB cameras. Additionally, it includes 4GB of RGB and thermal infrared image data with corresponding annotations. The Oriented Bounding Box (OBB) annotations are carefully aligned with the image files, adhering to consistent naming conventions for easy access and reference. Moreover, the dataset is enriched with 3D High-Definition Map (HDMAP) data, offering detailed insights into lane positioning and topology within each scene. Also included is a suite of development toolkits, facilitating the use of this dataset in training computer vision algorithms. The dataset's structure is detailed in Fig. 5.

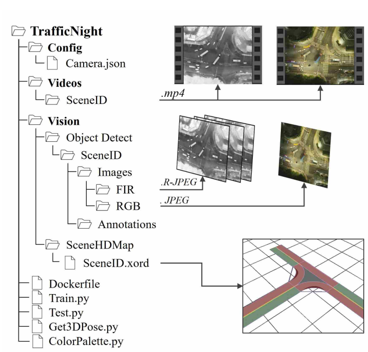
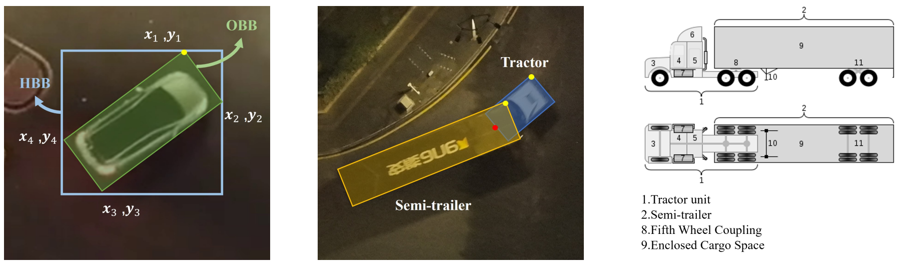


Fig. 5: Dataset structure overview

2.2 Dataset Annotation

Category Definition: The dataset encompasses annotations for various vehicle types such as cars, buses, trucks, tractors, semi-trailers, and empty semi-trailers. A key feature is the distinct labeling of detachable towing semi-trailer objects and tractors. This separation is vital for analyzing and predicting the movement trajectories of each component of these vehicles.



(a) Obb label example **(b)** Semi-trailer split results **(c)** Each part of the semi-trailer

Fig. 7: Annotation geometry definition. All objects in the dataset, from individual vehicles to semi-trailer parts, with semi-trailers distinctly segmented into tractors and trailers.

Different vehicle types often exhibit unique driving styles, which is critical in traffic flow and driving behavior studies. To address this, our dataset establishes a tree-like hierarchy of vehicle categories, as illustrated in the Fig. 6. Special attention is given to differentiating between Tractor Units, Cargo space semi-trailers, and Empty semi-trailers. This distinction is essential for accurate visual perception and trajectory prediction, enabling more precise modeling of various vehicle dynamics and behaviors.

Geometry Definition: An Oriented Bounding Box (OBB) is a rectangular bounding box that can be rotated, unlike Horizontal Bounding Boxes (HBB) which align parallel to base axes. OBBs' rotational capability allows for a more accurate representation of objects in various orientations and poses.

Within our dataset, bounding boxes are defined by the coordinates of their four corners, as shown in Fig. 7a. Each bounding box is represented by the coordinates of its top-left (x_1, y_1), top-right (x_2, y_2), bottom-right (x_3, y_3), and bottom-left (x_4, y_4) corners in the image plane. These coordinates precisely outline the object's boundaries within the image. For semi-trailer vehicles, annotations are based on the definitions in Fig. 7c, as demonstrated in Fig. 7b.

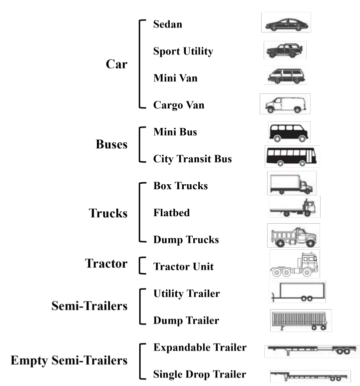


Fig. 6: Category definition in TrafficNight

Employing OBBs for object geometry definition significantly enhances annotation accuracy and application effectiveness. This precise object representation, regardless of orientation, improves the dataset's utility for complex computer vision tasks, offering more accurate and descriptive annotations than traditional HBBs, especially for irregularly shaped or variously angled objects.

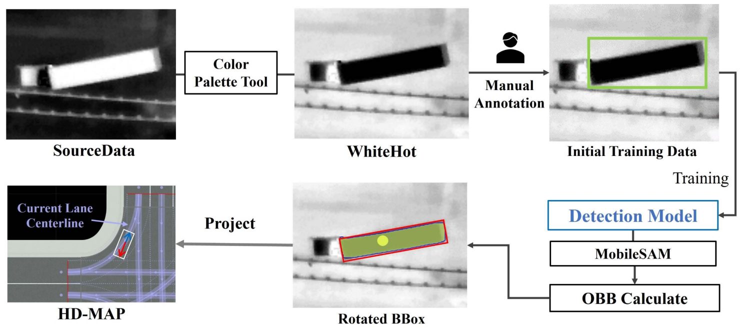


Fig. 8: Automated labeling pipeline

Oriented Bounding Box Generation. To reduce the costs associated with manual data labeling, we introduced an automated labeling pipeline, depicted in Fig. 8. This pipeline includes four main components: a simplified color palette tool from our development kit, the YoloV8 Horizontal Bounding Box (HBB) detector, the MobileSAM [15, 28] model for vehicle mask generation, and an Oriented Bounding Box (OBB) calculator.

HDMAP. In this dataset, high-definition (HD) maps for each scene were precisely developed using RoadRunner and aligned with the WGS84 (EPSG 4326) reference system, ensuring accuracy. Stored in xord format and created from spatial point cloud data, the maps provide detailed three-dimensional representations. The inclusion of these HD maps is essential, as they improve the understanding of context in complex traffic environments. This aligns with sophisticated predictive models like VectorNet [9, 11, 19], enhancing accuracy in multi-agent system behavior analysis.

2.3 Dataset Statistics

The dataset encompasses 2200 pairs of annotated thermal and sRGB image data, with a relatively even distribution across different vehicle categories, including semi-trailers and trucks. The distribution of labeled categories, spatial locations, and sizes on the image is depicted in Fig. 9. Thermal images are captured in floating-point format, stored in R-JPEG format (16 bits) at 1280x1024 resolution. In contrast, sRGB images are recorded at 4000x3000 resolution in standard JPEG format (8 bits). Annotations are formatted similarly to those in DOTA-v2 [7].

The sRGB and thermal images are captured independently, not extracted from raw video, allowing researchers to apply preferred color palettes using tools like the DJI Thermal Analysis Tool. For sRGB data, camera settings automatically adjust for optimal image quality. The thermal sensor operates in low-gain mode, adjusted for ambient temperature conditions during capture.

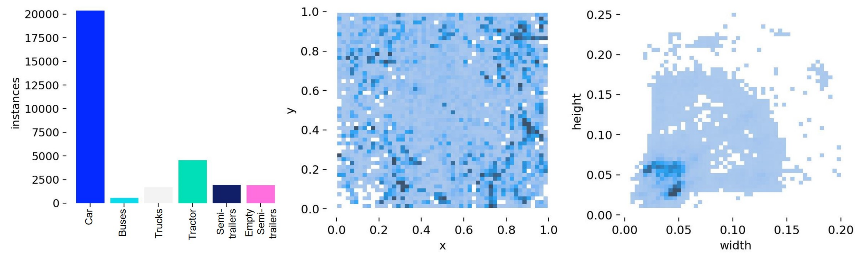


Fig. 9: Distribution of categories, spaces and size

3 Toolkit

3.1 Pre-training Model

In autonomous vehicle technology, thermal infrared imagery offers significant advantages over traditional RGB imaging, especially in low-visibility situations. For our pre-training model, we selected thermal infrared images, post-processed with a color palette to highlight thermal signatures.

The accompanying toolkit for this dataset includes pre-trained models and specialized tools for coordinate transformation. It features the YoloV8-obb object detection network, tailored for thermal infrared vehicle images. These pre-trained models are designed to enhance vehicle detection and tracking in thermal videos, and to improve night-time vision algorithms with RGB data.

Furthermore, the toolkit contains the script, enabling the conversion of pixel coordinates from images to physical space latitude and longitude coordinates. This feature is critical for applications requiring accurate physical dimensions, such as calibrating driving models and advancing trajectory prediction research.

Fig. 10: Tracking vehicles in thermal infrared videos.

We compiled a dataset of 2,200 thermal images, each precisely annotated with oriented bounding boxes (OBB) for accurate object localization. This dataset underpins the training of our YoloV8-obb model, an advanced object detection framework. The model has been meticulously fine-tuned to recognize and categorize various objects in thermal imagery, focusing on living beings and vehicles, typically difficult to detect in poor visibility. The pre-trained YoloV8-obb model enables efficient object detection from unique thermal signatures, crucial for high-accuracy applications in challenging environments. Its output can be integrated with sophisticated tracking algorithms like BoT-SORT [1] or ByteTrack [29]. These trackers, used post-detection, aid in robust trajectory tracking, augmenting the model's applicability in dynamic scenarios prevalent in autonomous vehicle navigation and surveillance.

3.2 3DMapping Toolkit

Contrasting with aeronautical datasets like NGSIM [21] and Citysim [30], which predominantly rely on affine transformations for vehicle coordinate conversion, our script employs the Perspective-n-Point (PNP) algorithm [22], focusing on minimizing reprojection error for more precise physical coordinate determination.

While affine transformation offers simplicity, it fails to account for parallax errors in images, leading to inaccuracies in vehicle positioning. In comparison, our PNP-based approach more accurately reflects real-world scenarios, providing a significant improvement in positional accuracy.

The script in our toolkit is crucial for converting vehicle coordinates from image frames to geographic coordinates. It employs the Perspective-n-Point (PNP) algorithm [22] to determine the camera's external parameters: rotation matrix R and translation vector T, as defined in Eq.1. These parameters are calculated using the UAV camera's internal parameters K and the world coordinates C_{gn} of a reference point on the Digital Surface Model (DSM), along with its corresponding image projection coordinates P_{gn}.

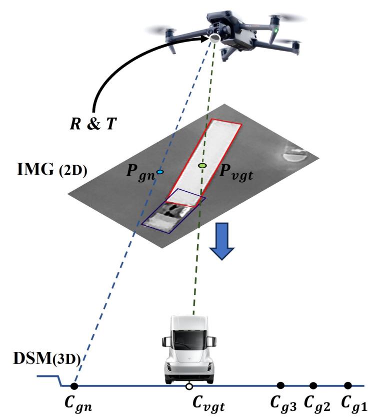


Fig. 11: Use RayTrack method to mapping 2d to 3d.

$$[R, T] = f_{pnp}(K, C_{\mathrm{gn}}, P_{\mathrm{gn}}) \tag{1}$$

After determining the external parameters, we can convert the camera coordinates P_{vgt} of any vehicle on the image to C_{ve} the coordinates on the DSM, by tracking the light path form vehicle.

$$P_{\mathrm{vgt}} = K \cdot (R \cdot C_{\mathrm{vgt}} + T) \tag{2}$$

4 Application

4.1 Perception Algorithm Enhancement

Vehicle detection in visual algorithms faces challenges from low-light conditions, dynamic shadows, and headlight glare, which impair image quality and algorithm performance.

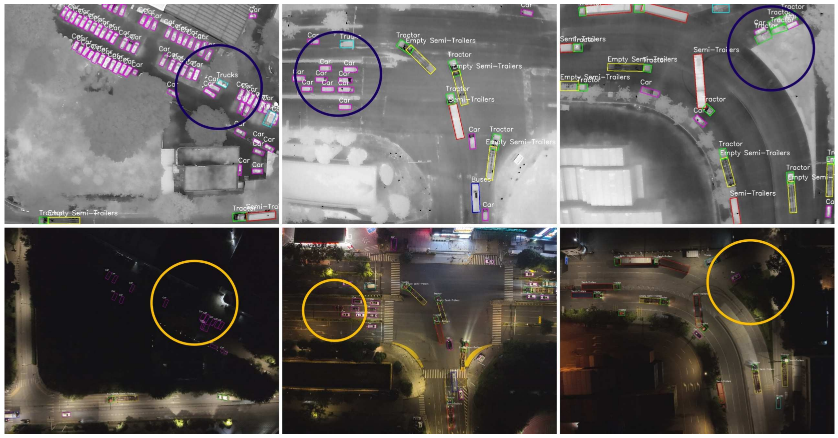


Fig. 12: YoloV8 inference results

Our dataset addresses these challenges by focusing on nighttime scenarios and incorporating thermal imaging data. Thermal imaging, unaffected by low-light conditions, reliably detects vehicles using heat signatures. The dataset also ensures a balanced representation of larger vehicles, enriching training inputs.

We conducted rotating object detection experiments using sRGB and cIF (Colorlize Infrared) data as inputs, testing SOTA methods (LSKNet [18], RTMDet-R [2], and YoloV8-OBB [24]) on our dataset (75.0% training, 25.0% testing). The training parameters were optimized using grid search on two 3090Ti GPUs. The results, shown in Tab. 3, indicate that all networks perform better with cIF than sRGB, enabling detection of some vehicles in darkness. Notably, as (b) in Fig. 12 illustrated, networks trained on our data can detect vehicles in extremely low-light conditions even with sRGB input. Additionally, our dataset's vehicle annotations, based on Colorlize Infrared Images, prove more effective than human annotations in low-light conditions, making it a robust nighttime benchmark.

Table 3: Experiment results of SOTA Method (OBB) on TrafficNight. AP50 meansthe AP at loU threshold of 0.5, AP75 means the AP at loU threshold of 0.75.

	LSKNet-S* [18]		RTMDet-R-l [2]		Yolov8-obb [24]	
	AP50	AP75	AP50	AP75	AP50	AP75
sRGB	79.45	53.44	79.33	52.25	77.83	51.71
cIF	94.46	86.35	93.10	85.69	91.10	85.32

Another key strength is the precision of our annotations. In difficult conditions, accurate annotations are essential for developing robust algorithms. Our dataset, using both RGB and thermal data for labeling, guarantees annotations that are not only precise but also applicable to real-world situations. This meticulous labeling, along with the dataset's distinctive features, makes it a crucial resource for advancing nighttime vehicle detection [6, 17, 26] and tracking algorithms [25, 27].

4.2 Calibration of CFM and Semi-Trailer Parameters

Calibrating the parameters of Car Following Models (CFM) is crucial for enhancing the accuracy and reliability of traffic flow predictions. Proper calibration aligns the model's behavior with real-world observations, leading to precise traffic simulations and forecasts, essential for traffic management, infrastructure planning, and autonomous driving systems development. However, existing research data often lack nighttime and low-visibility environmental conditions. Our dataset addresses this gap with an emphasis on nighttime scenarios. Using our toolkit, developers can extract vehicle trajectories from nighttime videos, enabling the calculation of traffic flow parameters through various methods [13,14]. This nighttime data ensures comprehensive CFM parameter calibration, improving the model's predictive accuracy in diverse lighting conditions. The Fig.13 depicts vehicle speed distribution, derived from tracking results in an intersection scene within the dataset . By using the tracking results, we can calculate the speed distribution of different vehicle models, and use the average speed of each vehicle model as a parameter to set up CFM for traffic flow simulation.

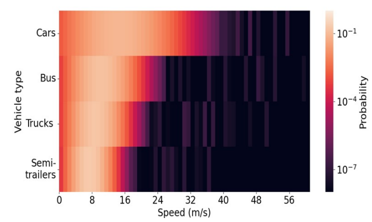


Fig. 13: Use tarcking result to calibrating of CFM. Vehicle speed distribution in an intersection scence.

5 Conclusion

This dataset marks a significant advancement in Autopilot and surveillance systems, overcoming the constraints of existing datasets. It accurately represents traffic scenarios, with a focus on semi-trailers and low-light conditions. The combination of thermal infrared imaging and oriented bounding boxes greatly improves object detection in varied lighting situations. Our automated annotation framework, incorporating sophisticated algorithms and Meta's SAM model, optimizes the labeling process. The use of the PNP algorithm for 3D coordinate transformation and specialized annotations for articulated trucks further enhances its application in spatial analysis and trajectory prediction. However, we acknowledge that thermal imaging sensors might introduce unknown noise under extreme weather conditions. Additionally, our current data is collected from tropical regions.

In future work, we plan to supplement it with data from colder regions and expand the dataset to include unstructured traffic scenarios and high-speed environments. This expansion aims to offer a more comprehensive resource for researchers, increasing the dataset's applicability in developing detection techniques for complex traffic conditions. It will also address the growing demand for diverse data in autonomous vehicle technology, maintaining the dataset's critical role in advancing the field.

Acknowledgements

This work was supported in part by the Research Institute for Artificial Intelligence of Things (RIAIoT) through the project P0050293, and by the Smart Traffic Fund (project ID: PSRI/53/2210/PR).

References

1. Aharon, N., Orfaig, R., Bobrovsky, B.Z.: Bot-sort: Robust associations multi-pedestrian tracking. arXiv preprint arXiv:2206.14651 (2022)
2. et al., C.L.: Rtmdet: An empirical study of designing real-time object detectors (2022)
3. et al., F.W.: Decentralized vehicle coordination: The berkeley deepdrive drone dataset (2022)
4. Cordts, M., Omran, M., Ramos, S., Scharwächter, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B.: The cityscapes dataset. In: CVPR. vol. 2 (2015)
5. Dablain, D., Krawczyk, B., Chawla, N.V.: Deepsmote: Fusing deep learning and smote for imbalanced data. IEEE Transactions on Neural Networks and Learning Systems (2022)
6. Deng, X., Wang, P., Lian, X., Newsam, S.: Nightlab: A dual-level architecture with hardness detection for segmentation at night. In: CVPR. pp. 16938–16948 (2022)
7. Ding, J., Xue, N., Xia, G.S., Bai, X., Yang, W., Yang, M., Belongie, S., Luo, J., Datcu, M., Pelillo, M., Zhang, L.: Object detection in aerial images: A large-scale benchmark and challenges. IEEE Transactions on Pattern Analysis and Machine Intelligence pp. 1–1 (2021). https://doi.org/10.1109/TPAMI.2021.3117983
8. Ettinger, S., Cheng, S., Caine, B., Liu, C., Zhao, H., Pradhan, S., Chai, Y., Sapp, B., Qi, C.R., Zhou, Y., et al.: Large scale interactive motion forecasting for autonomous driving: The waymo open motion dataset. In: ICCV. pp. 9710–9719 (2021)
9. Gao, J., Sun, C., Zhao, H., Shen, Y., Anguelov, D., Li, C., Schmid, C.: Vectornet: Encoding hd maps and agent dynamics from vectorized representation. In: CVPR. pp. 11525–11533 (2020)
10. Geiger, A., Lenz, P., Stiller, C., Urtasun, R.: Vision meets robotics: The kitti dataset. The International Journal of Robotics Research **32**(11), 1231–1237 (2013)
11. Gu, J., Sun, C., Zhao, H.: Densetnt: End-to-end trajectory prediction from dense goal sets. In: ICCV. pp. 15303–15312 (2021)
12. Jia, X.e.a.: Llvip: A visible-infrared paired dataset for low-light vision. In: CVPR. pp. 3496–3504 (2021)

13. Ke, R., Li, Z., Kim, S., Ash, J., Cui, Z., Wang, Y.: Real-time bidirectional traffic flow parameter estimation from aerial videos. IEEE Transactions on Intelligent Transportation Systems **18**(4), 890–901 (2016)
14. Ke, R., Li, Z., Tang, J., Pan, Z., Wang, Y.: Real-time traffic flow parameter estimation from uav video based on ensemble classifier and optical flow. IEEE Transactions on Intelligent Transportation Systems **20**(1), 54–64 (2018)
15. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. ICCV (2023)
16. Krajewski, R., Bock, J., Kloeker, L., Eckstein, L.: The highd dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC). pp. 2118–2125. IEEE (2018)
17. Li, G., Zhang, S., Yang, J.: Nighttime pedestrian detection based on feature attention and transformation. In: 2020 25th International Conference on Pattern Recognition (ICPR). pp. 9180–9187. IEEE (2021)
18. Li, Y.e.a.: Large selective kernel network for remote sensing object detection. In: ICCV (October 2023)
19. Liang, M., Yang, B., Hu, R., Chen, Y., Liao, R., Feng, S., Urtasun, R.: Learning lane graph representations for motion forecasting. In: ECCV. pp. 541–556. Springer (2020)
20. Mullick, S.S., Datta, S., Das, S.: Generative adversarial minority oversampling. In: ICCV. pp. 1695–1704 (2019)
21. Punzo, V., Borzacchiello, M.T., Ciuffo, B.: On the assessment of vehicle trajectory data accuracy and application to the next generation simulation (ngsim) program data. Transportation Research Part C: Emerging Technologies **19**(6), 1243–1262 (2011)
22. Schweighofer, G., Pinz, A.: Globally optimal o (n) solution to the pnp problem for general camera models. In: BMVC. pp. 1–10 (2008)
23. Sun, P., Kretzschmar, H., Dotiwalla, X., Chouard, A., Patnaik, V., Tsui, P., Guo, J., Zhou, Y., Chai, Y., Caine, B., et al.: Scalability in perception for autonomous driving: Waymo open dataset. In: CVPR. pp. 2446–2454 (2020)
24. Team, U.: Yolov8-obb project
25. Tian, W., Chen, L., Zou, K., Lauer, M.: Vehicle tracking at nighttime by kernelized experts with channel-wise and temporal reliability estimation. IEEE Transactions on Intelligent Transportation Systems **19**(10), 3159–3169 (2017)
26. Xu, Q., Ma, Y., Wu, J., Long, C., Huang, X.: Cdada: A curriculum domain adaptation for nighttime semantic segmentation. In: ICCV. pp. 2962–2971 (2021)
27. Ye, J., Fu, C., Zheng, G., Paudel, D.P., Chen, G.: Unsupervised domain adaptation for nighttime aerial tracking. In: CVPR. pp. 8896–8905 (2022)
28. Zhang, C., Han, D., Qiao, Y., Kim, J.U., Bae, S.H., Lee, S., Hong, C.S.: Faster segment anything: Towards lightweight sam for mobile applications. arXiv preprint arXiv:2306.14289 (2023)
29. Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., Wang, X.: Bytetrack: Multi-object tracking by associating every detection box. In: ECCV. pp. 1–21. Springer (2022)
30. Zheng, O., Abdel-Aty, M., Yue, L., Abdelraouf, A., Wang, Z., Mahmoud, N.: Citysim: A drone-based vehicle trajectory dataset for safety-oriented research and digital twins. Transportation Research Record (2023). https://doi.org/10.1177/03611981231185768

31. Zhu, P., Wen, L., Du, D., Bian, X., Fan, H., Hu, Q., Ling, H.: Detection and tracking meet drones challenge. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(11), 7380–7399 (2021)