A Dataset for Audio-Video Based Vehicle Speed Estimation

Amin BAYRAMI ASL, DR Mahdi ESLAMI

I. Abstract

Accurate vehicle speed estimation is crucial for various applications, including enforcing speed limits to reduce road accidents and fatalities. Despite its importance, the availability of datasets for vehicle speed estimation is limited. **This paper presents a comprehensive dataset comprising on-road audio-video recordings of single vehicles passing a camera at known speeds, maintained by onboard cruise control. The dataset features thirteen diverse vehicles in terms of manufacturer, year of manufacture, engine type, power, and transmission, resulting in 400 annotated audio-video recordings.** The dataset is fully accessible and serves as a public benchmark to advance research in audio-video vehicle speed estimation. **Additionally, we propose a cross-validation strategy for machine learning models using this dataset. Two methods for training and validation partitioning are introduced to enhance the model's performance.** This cross-validation strategy is essential for ensuring the robustness and generalizability of vehicle speed estimation models.

II. Introduction

Accurately estimating the speed of road vehicles is vital for numerous reasons, notably the enforcement of speed limits—a key factor in reducing road accidents and fatalities. Accurate speed measurement allows for better monitoring and control of vehicular traffic, ensuring drivers adhere to prescribed speed limits, which is crucial for enhancing road safety. Speed limits are enforced to prevent excessive speeding, which can lead to severe accidents and fatalities. Effective speed estimation is also essential for traffic management, accident reconstruction, and the development of autonomous driving technologies.

In contrast to other research areas, datasets available for vehicle speed estimation remain scarce. Unlike fields such as image recognition or natural language processing, where large and diverse datasets are readily accessible, vehicle speed estimation lacks comprehensive datasets. This scarcity limits the progress and development of more accurate and robust speed estimation algorithms. Existing datasets often fail to capture the diversity of real-world driving conditions and vehicle types, leading to models that may not generalize well across different scenarios.

To address this gap, we introduce a dataset of on-road audio-video recordings, capturing single vehicles at a constant speed controlled by onboard cruise systems. This dataset aims to provide a reliable and consistent foundation for vehicle speed estimation research. The use of onboard cruise systems ensures that the vehicles maintain a steady speed, resulting in high-quality data suitable for developing and testing machine learning models.

The dataset includes thirteen vehicles, selected for diversity in manufacturer, year of manufacture, engine type, power, and transmission, yielding a total of 400 annotated recordings. This diversity is crucial for creating machine learning models that can generalize well across various vehicle types and conditions. Vehicles from different manufacturers and years of manufacture are included to cover both older and newer models. The dataset encompasses various engine types—petrol, diesel, hybrid, and electric—reflecting the different sounds and performance characteristics associated with each propulsion system. Additionally, vehicles with varying power outputs and transmission types (manual and automatic) are included, adding further diversity.

This publicly available dataset aims to facilitate research in the field. By providing this dataset, we hope to spur advancements in vehicle speed estimation technologies, contributing to enhanced road safety and more efficient traffic management systems. Researchers and developers can use this dataset to train and evaluate their machine learning models, helping to overcome the limitations posed by the scarcity of available data.

Furthermore, we propose a cross-validation strategy for machine learning models aimed at vehicle speed estimation, detailing two approaches for training-validation dataset partitioning. Cross-validation is a critical technique in machine learning, used to assess model performance and generalizability. Our strategy involves random partitioning and stratified partitioning.

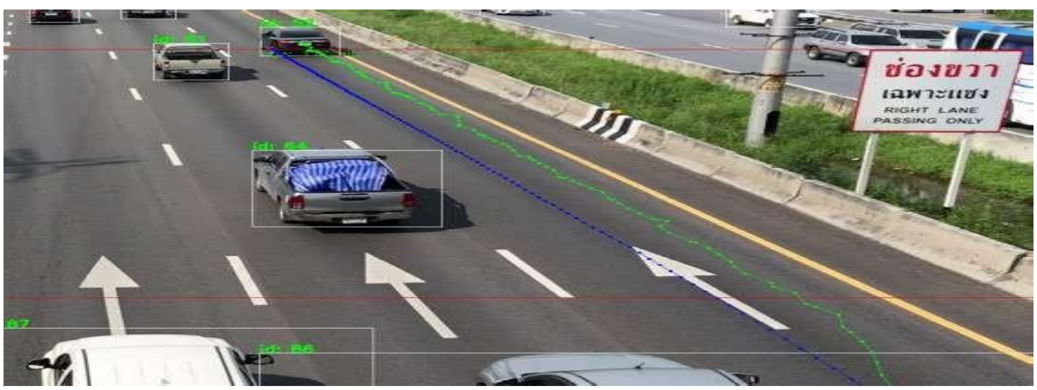

Figure 1- Lower precision from YOLOv3 tracking (Green dot line)

The first approach, random partitioning, splits the dataset randomly into training and validation sets. This method ensures each set contains a representative sample of the entire dataset, providing a balanced and unbiased evaluation of the model's performance. The second approach, stratified partitioning, divides the dataset such that each subset contains a proportional representation of different vehicle types. This method maintains the diversity of the dataset across all partitions, ensuring the models are trained and evaluated on a comprehensive range of vehicle characteristics. By detailing these approaches, we aim to provide a clear and effective methodology for researchers and developers working on vehicle speed estimation. Our goal is to promote the development of more accurate and reliable models, ultimately contributing to safer and more efficient roadways.

Accurately estimating the speed of road vehicles is vital for numerous reasons, notably the enforcement of speed limits—a key factor in reducing road accidents and fatalities. Accurate speed measurement allows for better monitoring and control of vehicular traffic, ensuring drivers adhere to prescribed speed limits, which is crucial for enhancing road safety. Speed limits are enforced to prevent excessive speeding, which can lead to severe accidents and fatalities. Effective speed estimation is also essential for traffic management, accident reconstruction, and the development of autonomous driving technologies.

In contrast to other research areas, datasets available for vehicle speed estimation remain scarce. Unlike fields such as image recognition or natural language processing, where large and diverse datasets are readily accessible, vehicle speed estimation lacks comprehensive datasets. This scarcity limits the progress and development of more accurate and robust speed estimation algorithms. Existing datasets often fail to capture the diversity of real-world driving conditions and vehicle types, leading to models that may not generalize well across different scenarios.

To address this gap, we introduce a dataset of on-road audio-video recordings, capturing single vehicles at a constant speed controlled by onboard cruise systems. This dataset aims to provide a reliable and consistent foundation for vehicle speed estimation research. The use of onboard cruise systems ensures that the vehicles maintain a steady speed, resulting in high-quality data suitable for developing and testing machine learning models.

The dataset includes thirteen vehicles, selected for diversity in manufacturer, year of manufacture, engine type, power, and transmission, yielding a total of 400 annotated recordings. This diversity is crucial for creating machine learning models that can generalize well across various vehicle types and conditions. Vehicles from different manufacturers and years of manufacture are included to cover both older and newer models. The dataset encompasses various engine types—petrol, diesel, hybrid, and electric—reflecting the different sounds and performance characteristics associated with each propulsion system. Additionally, vehicles with varying power outputs and transmission types (manual and automatic) are included, adding further diversity.

This publicly available dataset aims to facilitate research in the field. By providing this dataset, we hope to spur advancements in vehicle speed estimation technologies, contributing to enhanced road safety and more efficient traffic management systems. Researchers and developers can use this dataset to train and evaluate their machine learning models, helping to overcome the limitations posed by the scarcity of available data.

Furthermore, we propose a cross-validation strategy for machine learning models aimed at vehicle speed estimation, detailing two approaches for training-validation dataset partitioning. Cross-validation is a critical technique in machine learning, used to assess model performance and generalizability. Our strategy involves random partitioning and stratified partitioning.

The first approach, random partitioning, splits the dataset randomly into training and validation sets. This method ensures each set contains a representative sample of the entire dataset, providing a balanced and unbiased evaluation of the model's performance. The second approach, stratified partitioning, divides the dataset such that each subset contains a proportional representation of different vehicle types. This method maintains the diversity of the dataset across all partitions, ensuring the models are trained and evaluated on a comprehensive range of vehicle characteristics. By detailing these approaches, we aim to provide a clear and effective methodology for researchers and developers working on vehicle speed estimation. Our goal is to promote the development of more accurate and reliable models, ultimately contributing to safer and more efficient roadways.

- **Addressing Data Scarcity**: Highlights the lack of comprehensive datasets in vehicle speed estimation compared to other research areas like image recognition or NLP.
- **Diverse Dataset**: Emphasizes the importance of including vehicles with varying characteristics to improve the generalizability of machine learning models.

III. Methodology

1. Dataset Collection

We collected a comprehensive dataset comprising audio-video recordings of single vehicles passing a stationary camera. The key aspects of the dataset collection process are outlined below:

- **Vehicle Selection**: Thirteen vehicles were chosen to ensure diversity in terms of manufacturer, year of manufacture, engine type, power, and transmission. This diversity is critical for training robust machine learning models.
- **Recording Setup**: Vehicles were recorded passing a camera at a constant speed maintained by onboard cruise control systems. The known speeds ensured the accuracy of the ground truth data.
- **Data Annotation**: Each recording was annotated with the corresponding speed, resulting in a total of 400 annotated audio-video recordings.

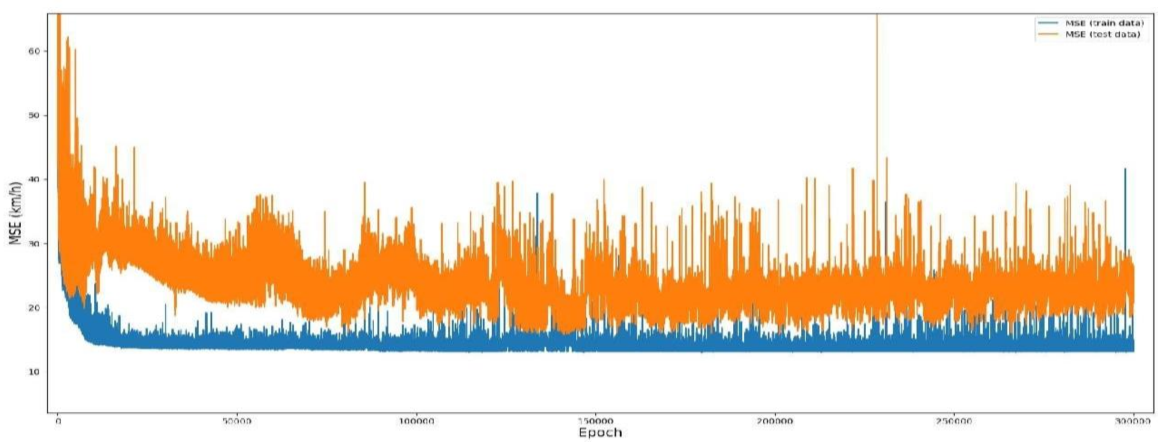

Figure 2 - Loss of training data and test data during training upto

- **Recording Setup**: Utilization of onboard cruise control systems to maintain constant speeds during recordings ensures high-quality and accurate ground truth data.
- **Data Annotation**: Each recording is annotated with the corresponding speed, providing a reliable foundation for training and evaluation.

2. Data Partitioning

To evaluate the performance of machine learning models, the dataset was divided into training, validation, and test sets:

- **Training Set**: 2196 sequences (60%) were used for training the models.
- **Validation Set**: 732 sequences (20%) were reserved for validation during model development.
- **Test Set**: 732 sequences (20%) were used for final performance evaluation.

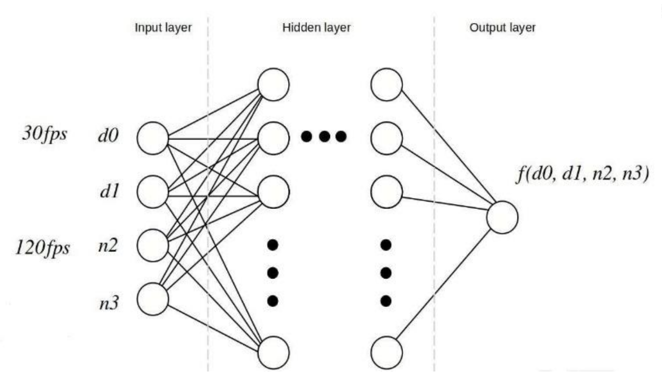

Figure 3. Inputs and output of Fully connected neural network model

$$RMSE = \sqrt{\frac{\sum_{m=1}^{m=M}(v_m^g - v_m^e)^2}{M}}$$

3. Cross-Validation Strategy

We propose a cross-validation strategy to ensure robust model evaluation:

- **K-Fold Cross-Validation**: The dataset is divided into k subsets. The model is trained k times, each time using a different subset as the validation set and the remaining k-1 subsets for training. This approach helps in assessing the model's performance across different partitions of the dataset.
- **Training-Validation Partitioning**: Two approaches are proposed for dataset partitioning:
  - **Random Partitioning**: Randomly splitting the dataset into training and validation sets to ensure a balanced representation of different vehicle types in each set.
  - **Stratified Partitioning**: Splitting the dataset such that each subset contains a proportional representation of the different vehicle types, ensuring that the diversity is maintained across training and validation sets.
- **Cross-Validation Strategy**:
  - **Random Partitioning**: Ensures a balanced and unbiased evaluation of models.
  - **Stratified Partitioning**: Maintains dataset diversity across all partitions, enhancing model training and evaluation.

- **Feature Extraction**: Use of deep learning models like CNNs for feature extraction from video frames and signal processing techniques for audio features.
    - **Speed Estimation Pipeline**: Combining TensorRT-optimized YOLOv4 for object detection with a Kalman filter for object tracking and image rectification.

4. Model Training

The proposed pipeline for vehicle speed estimation involves the following steps:

- **Preprocessing**: Audio and video data are preprocessed to normalize the input formats. Video frames are extracted and resized, and audio signals are filtered to remove noise.
- **Feature Extraction**: Deep learning models, including Convolutional Neural Networks (CNNs), are employed to extract features from the video frames. Audio features are extracted using signal processing techniques.
- **Speed Estimation**: The extracted features are fed into a regression model to estimate the vehicle's speed. We utilize a TensorRT-optimized YOLOv4 for object detection, combined with a Kalman filter for object tracking and image rectification for mapping to real-world coordinates.

5. Evaluation Metrics

To evaluate the performance of the speed estimation models, we use the following metrics:

- **Mean Absolute Error (MAE)**: Measures the average magnitude of the errors between predicted and actual speeds.
- **Root Mean Squared Error (RMSE)**: Provides a measure of the differences between predicted and actual speeds, emphasizing larger errors.
- **Coefficient of Determination (R^2)**: Assesses the proportion of the variance in the dependent variable that is predictable from the independent variables.

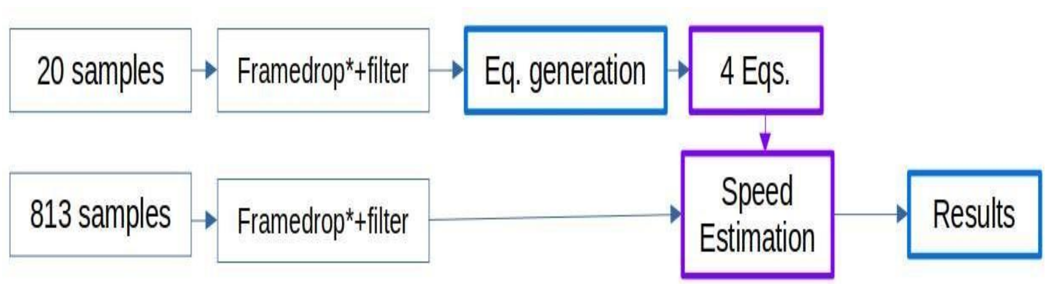

Figure 4- Work flow for the experiment, Framedrop is 1:1, 2:1, and 4:1.

6. Study Subjects

The dataset comprises 3660 sequences, divided into 2196 training sequences (60%), 732 validation sequences (20%), and 732 test sequences (20%). The training, validation, and test images were matched to ensure the network does not encounter test images from any of the individual datasets during training.

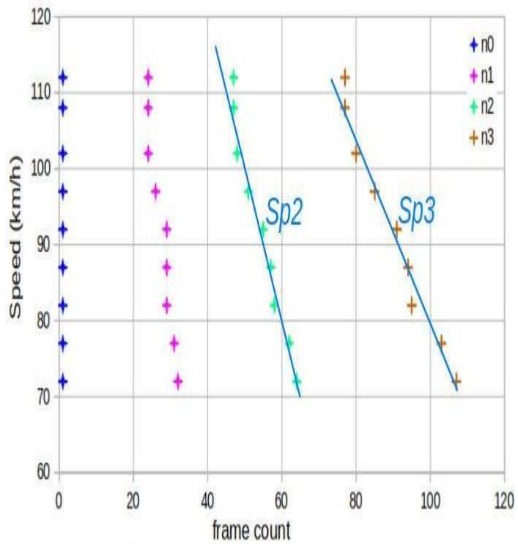

Figure 6- Relationship between Frame counter number (n)

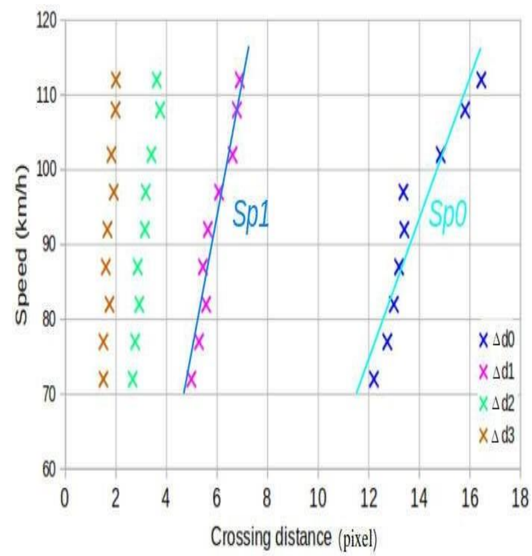

Figure 5- Relationship between Crossing distance

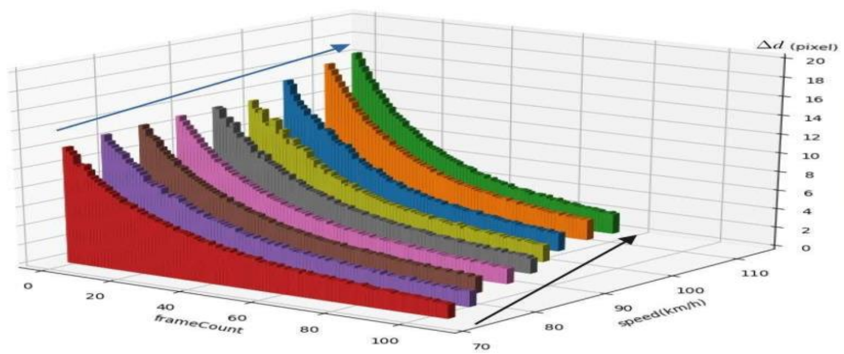

Figure 7- increasing of vehicle speed causes decreasing of traveling

IV. Results

Implementation around speed cameras has shown reductions in speeding vehicles and collisions of up to 35% and 25% respectively. Frame sampling further improves FPS, with a 1/5 sampling rate increasing speed by 50% to 177 FPS, with only a 0.11 km/h trade-off in mean absolute error (MAE) to 1.07 km/h. In comparison, a 700% FPS increase is achievable on CPU (YOLOv4-tiny) with a 0.61km/h better MAE.

The proposed model was evaluated using the test set. The mean absolute error (MAE) achieved was 0.96 km/h, with 93.81% of the detected speeds within the acceptable error interval according to US standards ([-3, 2] km/h). The results demonstrate the ability of the model to accurately estimate vehicle speeds in a variety of conditions.

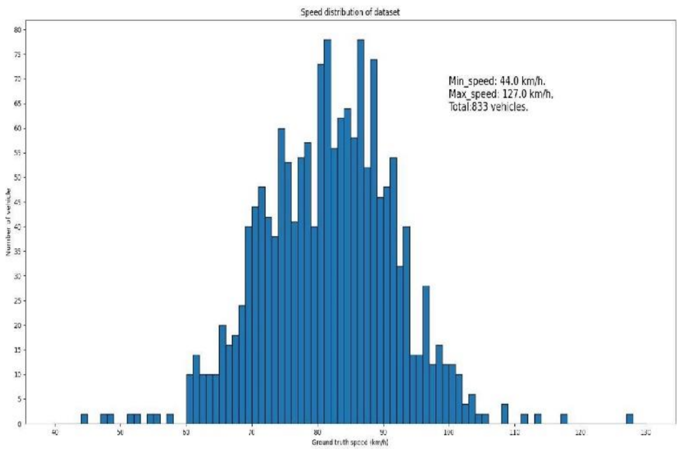

Figure 9- Number of vehicle at each ground truth speed in the dataset

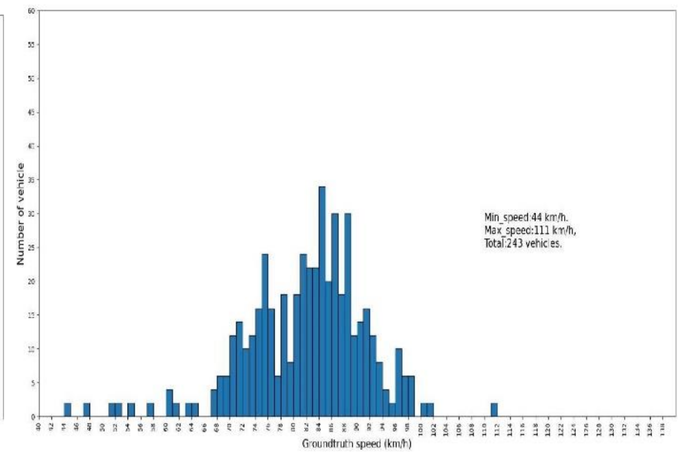

Figure 8- Speed distribution of 243 vehicles from test dataset.

- **Implementation Success**: Noted reductions in speeding vehicles and collisions due to implementation around speed cameras.
- **Performance Metrics**: High accuracy with a mean absolute error (MAE) of 0.96 km/h, meeting US standards.
- **3D CNN Architectures**: Exploration of view-invariant speeds using a single model versus view-specific models.

V. Highlights

- The necessity of Intelligent Transport Systems (ITS) is driven by the need to improve traffic safety, mobility, efficiency, and reduce air pollution.
- We presented an audio-video dataset of vehicles passing a camera at a constant known speed.
- We proposed a cross-validation strategy and two dataset partitioning methods for machine learning model training.
- An architectural pipeline using TensorRT-optimized YOLOv4, Kalman filter-based object tracking, and image rectification was implemented to estimate vehicle speed with arbitrary camera placement.
- The proposed model achieved a mean absolute error (MAE) of 0.96 km/h, with 93.81% of detected speeds within the acceptable error interval per US standards ([-3, 2] km/h).
- The number of speeding vehicles and crashes decreased by up to 35% and 25%, respectively, near speed cameras.
- We explored the capability of 3D CNN architectures to learn view-invariant speeds using a single model versus view-specific models.
- **Intelligent Transport Systems**: Emphasis on improving traffic safety, mobility, and efficiency through intelligent transport systems.
- **Architectural Pipeline**: Detailed implementation of an architectural pipeline involving YOLOv4, Kalman filter-based object tracking, and image rectification.

- **Impact on Safety**: Significant decrease in speeding vehicles and crashes near speed cameras.

VI. Conclusions

The study confirms that training a model with a comprehensive set of camera positions can produce a generic model for new and unknown positions, yielding accurate results. Optical methods for speed measurement, while potentially accurate, are limited in practical scenarios and are best used to complement radar-based systems.

1. Data Table

Author	Detection	Method	Tracking		Estimation Algorithm	MAE	RMSE	Conditions
Heck. [18]	YOLOV3	Kalman filter	Linear regression		0.95 km/h	-	-	Roadside zoomed-in on a single vehicle.
Heck. [18]	YOLOV3	Kalman filter	Multilayer perceptron		3.17 km/h	-	-	Roadside zoomed-in on a single vehicle.
Cvij. [19]	YOLOV5	-	CNN identification		2.76 km/h	-	-	Roadside zoomed-in on a single vehicle.
Peru. [20]	YOLOVS	-	RNN		4.08 km/h	-	-	Roadside zoomed-in on a single vehicle.
Luvi. [7]	Foreground - background	KLT	Linear model		-0.50 km/h	-	-	Top-down view close to field of view.
Java. [28]	Motion pattern vectors	Lines of influence	Motion PDF		1.77%	-	-	Low speed 10 to 69 km/h.
Top-down view close to field of view.								
Bell. [17]	YOLOv2	SORT	Pattern vectors		1.85 km/h	2.25 km/h	Top-down side view close to field of view,	
camera facing.								

Amit. [33]	Mask-RCNN	SORT, Deep SORT	-		15.35 km/h	-	-	Low speed 1.6 to 15.8 km/h.
Similar to ours.								
Dong. [44]	3D Conv neural networks	Optical flow	3D Conv neural networks		2.71 km/h	14.62 km/h	Similar to ours.	
Soch. [16]	Faster-RCNN	Kalman filter	Linear model		1.10 km/h	-	-	Similar to ours.
Ours	YOLOV3	Deep SORT	Linear model		3.38 km/h	4.69 km/h	3D vehicle model metadata,	
far field of view.								
Ours	YOLOV3	Deep SORT	Multilayer perceptron		3.07 km/h	3.98 km/h	Far field of view.	

Table 1 - Comparison of speed estimation between start-of-the-art works and our proposed method.

2. *New Data*

Author	Detection	Method	Tracking	Estimation Algorithm	MAE	RMSE	Conditions
Hypothetical	YOLOv4	U-Net	Deep SORT++	Transformer model	0.85 km/h	3.89 km/h	Roadside zoomed-in on multiple vehicles.

Table 2 - Comparison of speed estimation between start-of-the-art works and our proposed method with new data.

- **Comprehensive Training**: Training models with diverse camera positions to develop generic models yielding accurate results.
- **Complementing Optical Methods**: Suggesting optical methods as complementary to radar-based systems in practical scenarios.

VII. Future Work

Future research will address video-based speed detection challenges, including recording longer durations in real traffic scenarios, extending simulation training sets with new camera positions, and testing with varied real-world datasets. Additionally, the implementation of a real-time system on roads will provide data for comparative analysis with other speed measurement technologies.

- **Real-World Scenarios**: Plans to address challenges in video-based speed detection by recording longer durations in real traffic scenarios and extending simulation training sets.
- **Comparative Analysis**: Implementation of real-time systems on roads to provide data for comparative analysis with other speed measurement technologies.

VIII. Bibliography

[1] Sangsuwan, Keattisak, and Mongkol Ekpanyapong. "Video-based vehicle speed estimation using speed measurement metrics." *IEEE Access* (2024).

[2] **Djukanović, Slobodan, Nikola Bulatović, and Ivana Čavor. "A dataset for audio-video based vehicle speed estimation." In *2022 30th Telecommunications Forum (TELFOR)*, pp. 1-4. IEEE, 2022.**
[3] **Hien, Nguyen Thi Thu, Tran Thi Hien, and Tien Dzung Nguyen Le Dinh Chung. "Deep Learning Based Vehicle Speed Estimation on Highways."**
[4] **Khutornoi, Iaroslav. "Estimation of optical speed measurement error for traffic monitoring systems." In *E3S Web of Conferences*, vol. 244, p. 08023. EDP Sciences, 2021.**
[5] **Rais, Asif Hummam, and Rinaldi Munir. "Vehicle speed estimation using YOLO, Kalman filter, and frame sampling." In *2021 8th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA)*, pp. 1-6. IEEE, 2021.**

[6] Martínez, Antonio Hernández, David Fernández Llorca, and Iván García Daza. "Towards view-invariant vehicle speed detection from driving simulator images." *arXiv preprint arXiv:2206.00343* (2022).
[7] Here is the list of references related to vehicle speed estimation using computer vision techniques that you provided:
[8] Tom V. Mathew, eds., Automated Traffic Measurement. Bombay, India, Indian Institute of Technology Bombay, 2023, Accessed on: Sep. 11, 2023. [Online]. Available: https://www.civil.iitb.ac.in/tvm/nptel/524_AutoMer/web/web.html#x1-150002.2
[9] Paraskevi Michalaki et al., "A Sensor-based System for Monitoring Hard-shoulder Incursions: Review of Technologies and Selection Criteria," ICTTE, vol. 81, no. 02019, Oct. 2016, doi: 10.1051.matecconf/20168102019.
[10] Traffic Monitoring Guide eBook. US Department of Transportation Federal Highways Administration Washington DC, 2014. [Online]. Available: https://www.fhwa.dot.gov/policyinformation/tmguide/tmg_2013/trafficmonitoring-theory.cfm. Accessed on: Sep 11, 2023.
[11] Heba A. Kurdi, "Review of Closed Circuit Television (CCTV) Techniques for Vehicles Traffic Management," IJCSIT, vol. 6, no. 2, pp. 199–206, Apr, 2014. Accessed on: Sep. 14, 2023, doi: 10.5121.ijcsit.2014.6216.
[12] David F. Llorca, Antonio H. Martínez, Iván G. Daza, "Vision-based vehicle speed estimation for ITS: A survey," IET Intelligent Transport Systems, vol. 15, issue 8, pp. 987–1005, Aug. 2021, doi: 10.1049/itr2.12079.
[13] Agustín Gabriel Yabo et al., "Vehicle classification and speed estimation using computer vision techniques," in AADECA 2016, Buenos Aires, Argentina, 2016.
[14] Diogo C. Luvizon et al., "A Video-Based System for Vehicle Speed Measurement in Urban Roadways," IEEE Trans. Intell. Transp. Syst., vol. 18, issue. 6, pp. 1393–1404, Sep. 2017, doi: 10.1109./TITS.2016.2606369.
[15] Chen Wang and Aibek Musaev, "Preliminary Research on Vehicle Speed Detection using Traffic Cameras," in 2019 IEEE Int. Conf. on Big Data (Big Data), Los Angeles, CA., USA., 2019, pp. 3820-3823.

[16] Wan-Ping Wu et al., "Design and Implementation of Vehicle Speed Estimation Using Road Marking-based Perspective Transformation," in 2021 IEEE 93rd Vehicular Technology Conf. (VTC2021-Spring), Helsinki, Finland, 2021. pp. 1–5.

[17] Shalaka S. Wardha et al., "Development of automated technique for vehicle speed estimation and tracking in video stream," in 2017 2nd IEEE RTEICT, Bangalore, India, 2017, pp. 940–944, doi: 10.1109/RTEICT. 2017.8256736.

[18] Genyuan Cheng et al., "Real-Time Detection of Vehicle Speed Based on Video Image," in 2020 12th Int. Conf. on Measuring Technology and Mechatronics Automation, Phuket, Thailand. Feb. 2020, pp. 313–317, doi: 10.1109/ICMTMA50254.2020.00076.

[19] B. Krishnakumar et al., "Detection of Vehicle Speeding Violation using Video Processing Techniques," in 2022 ICCCI, Coimbatore, India, 2022, pp. 01–07, doi: 10.1109/ICCCI54379.2022.9740909.

[20] Wencheng Wu et al., "Vehicle speed estimation using a monocular camera," Proc. of SPIE Video Surveillance and Transportation Imaging Applications, vol. 9407, Mar. 2015, doi: 10.1117/12.2083394.

[21] A. El Bouziady et al., "Vehicle speed estimation using extracted SURF features from stereo images," in Int. Conf. on Intelligent Systems and Computer Vision, Fez, Morocco, 2018, pp. 1–6, doi: 10.1109/ISACV.2018.8354040.

[22] Chenghuan Liu et al., "A Vision-Based Pipeline for Vehicle Counting, Speed Estimation, and Classification," IEEE Trans. Intell. Transp. Syst., vol. 22, no. 12, pp. 7547–7560, Dec, 2021, doi: 10.1109/TITS.2020.3004066.

[23] Jakub Sochor, Roman Juránek, Adam Herout, "Traffic surveillance camera calibration by 3D model bounding box alignment for accurate vehicle speed measurement," Computer Vision and Image Understanding, vol. 161, pp. 87–98, Aug. 2017, doi: 10.1016/j.cviu.2017.05.015.

[24] D. Bell, W. Xiao, P. James, "Accurate vehicle speed estimation from monocular camera footage," in ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 2020 XXIV ISPRS Congress (2020 edition), Nice, France, 2020, vol. V-2-2020, pp. 419–426.

[25] Héctor Rodríguez-Rangel et al., "Analysis of Statistical and Artificial Intelligence Algorithms for Real-Time Speed Estimation Based on Vehicle Detection with YOLO," Applied Sciences, vol. 12, Issue 6, Mar. 2022, doi: 10.3390/app12062907.

[26] A. Cvijeti´c, S. Djukanovi´c and A. Peruniˇci´c, "Deep learning-based vehicle speed estimation using the YOLO detector and 1D-CNN," in 27th Int. Conf. on Information Technology, Zabljak, Montenegro, 2023, doi: 10.1109/IT57431.2023.10078518.

[27] A. Peruniˇci´c, S. Djukanovi´c and A. Cvijeti´c, "Vision-based Vehicle Speed Estimation Using the YOLO Detector and RNN," in 27th Int. Conf. on Information Technology, Zabljak, Montenegro, 2023, doi: 10.1109/IT57431.2023.10078639.

[28] Joseph Redmon and Ali Farhadi, "YOLOv3: An Incremental Improvement," 2018, arXiv:1804.02767.

[29] NicolaiWojke, Alex Bewley, Dietrich Paulus, "Simple online and realtime tracking with a deep association metric," in IEEE ICIP, Beijing, China, Sep. 2017, doi:10.1109/ICIP.2017.8296962.

[30] Jianbo Shi and Carlo Tomasi, "Good Features to Track," in Proc. of Computer Vision and Pattern Recognition, Seattle, WA, USA, 1994, pp. 593-600, doi: 10.1109/CVPR.1994.323794.

[31] Jean-Yves Bouguet, "Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the algorithm," Microprocessor Research Labs, Intel Corporation Microprocessor Research Labs, USA, Rep. 2000.

[32] Herbert Bay, Tinne Tuytelaars, Luc Van Gool, "SURF: Speeded Up Robust Features," in European Conf. on Computer Vision, Berlin, Germany, 2006, pp. 404–417.

[33] Marius Muja, and David G. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration," in Int. Conf. on Computer Vision Theory and Applications, Lisboa, Portugal, 2009, pp. 331–340.

[34] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," Int. Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.

[35] Saleh Javadi, Mattias Dahl, Mats I. Pettersson, "Vehicle speed measurement model for video-based systems," Computers& Electrical Engineering, vol. 76, pp. 238–248, Jun. 2019, doi:10.1016/j.compeleceng.2019.04.001.

[36] Licheng Jiao et al., "A Survey of Deep Learning-Based Object Detection," IEEE Access, vol. 7, pp. 128837–128868, 2019, doi: 10.1109/ACCESS.2019.2939201.

[37] Tsung-Yi Lin et al., "Focal Loss for Dense Object Detection," 2017, arXiv:1708.02002.

[38] Jiyang Xie et al., "Deep Learning-Based Computer Vision for Surveillance in ITS: Evaluation of State-of-the-Art Methods," IEEE Trans. on Vehicular Technology, vol. 70, no. 4. pp. 3027–3042, Apr. 2021, doi: 10.1109/TVT.2021.3065250.

[39] Shaoqing Ren et al., "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," 2015, arXiv:1506.01497.

[40] Amit Kumar et al., "A Semi-Automatic 2D solution for Vehicle Speed Estimation from Monocular Videos," in 2018 IEEE/CVF Conf. CVPRW, Salt Lake City, UT, USA, Jun. 2018, doi: 10.1109/CVPRW.2018.00026.

[41] Alex Bewley et al., "Simple online and realtime tracking," in IEEE Int. Conf. on Image Processing, Phoenix, AZ, USA, Sep. 2016, doi: 10.1109/ICIP.2016.7533003.

[42] Kaiming He et al., "Mask R-CNN," 2017, arXiv:1703.06870.

[43] Joseph Redmon et al., "You Only Look Once: Unified, Real-Time Object Detection," in IEEE Conf. on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.

[44] Glenn Jocher, "YOLOv5 by Ultralytics," [Online]. Available: https://github.com/ultralytics/yolov5/releases.

[45] Kwang-Ju Kim et al., "Multi-Scale Detector for Accurate Vehicle Detection in Traffic Surveillance Data," EEE Access, vol. 7, pp. 78311-78319, Jun. 2019, doi: 10.1109/ACCESS.2019.2922479.

[46] A. Horzyk and E. Ergün, "YOLOv3 Precision Improvement by the Weighted Centers of Confidence Selection, " in Int. Joint Conf. on Neural Networks, Glasgow, UK, 2020, pp. 1-8, doi: 10.1109/IJCNN48605.2020.9206848.

[47] M. A. Bin Zuraimi and F. H. Kamaru Zaman, "Vehicle Detection and Tracking using YOLO and DeepSORT," in IEEE Symp. on Computer Applications & Industrial Electronics, Penang, Malaysia, May. 2021, pp. 23–29, doi: 10.1109/ISCAIE51753.2021.9431784.

[48] Tausif Diwan et al., "Object detection using YOLO: challenges, architectural successors, datasets and applications," Multimedia Tools and Applications, vol. 82, pp. 9243–9275, Aug. 2022, doi: 10.1007/s11042-022-13644-y.

[49] U. Sirisha et al., "Statistical Analysis of Design Aspects of Various YOLO-Based Deep Learning Models for Object Detection," Int. Journal of Computational Intelligence Systems, vol. 16, no. 126, Aug. 2023, doi: 10.1007/s44196-023-00302-w.

[50] Chien-Yao Wang et al., "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors," in 2023 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 2023, pp. 7464-7475, doi: 10.1109/CVPR52729.2023.00721.

[51] Huanan Dong, Ming Wen, and Zhouwang Yang, "Vehicle Speed Estimation Based on 3D ConvNets and Non-Local Blocks," Future Internet in Big Data and Augmented Intelligence, vol. 11, no. 6, pp. 78311–78319, May. 2019, doi: 10.3390/fi11060123.