ARTIFICAL INTELLIGENCE, REAL GAINS. USING MACHINE LEARNING TO MONITOR OVER SIZE/OVER MASS HEAVY VEHICLE TRAFFIC ON BRIDGES



M. BUCKLEY
Professional Data Scientist at the National Transport Research Organisation. Obtained his Bachelor of Information Technology (Data Science) from Macquarie University

A. SINHA
Senior Professional Researcher at National Transport Research Organisation. Obtained her Ph.D from UNSW, Sydney.

K. LEWIS
Senior Professional at National Transport Research Organisation. Obtained his Masters (Civil Engineering) from Swinburne University of Technology.

M. BERENI
Chief Executive Officer at Traffic Management Association of Australia. Is a Fellow of the Institute of Engineers, with Masters of Science (Civil) and Masters of Science (Transport).

T. WANG
Professional Engineer at National Transport Research Organisation in Brisbane, Australia. Graduated in Civil and Environmental Engineering from the University of Queensland.

S. LYNCH
Manager of Structures Strategic Asset Management at the Department of Transport and Main Roads.

M. HALDANE
Director of Structures Management at the Department of Transport and Main Roads. Obtained her Masters of Engineering from Queensland University of Technology.

Abstract

Increasing demand for access by larger and heavier vehicles on aging road networks has heightened concerns about structures. To minimise the impact of Over Size Over Mass (OSOM) heavy vehicles (HVs), specific travel conditions are imposed based on mass and dimensions, including reduced speeds and specific lane positioning over bridges. The paper is based on the research outcomes as part of the National Asset Centre of Excellence (NACOE) Project S60: *Smart Network Monitoring.* The research project developed three machine learning models for the automated collection of data at a test site in Central Queensland, Australia. The models include a HV detection and tracking model, a speed estimation model, and a lane positioning model. An off-the-shelf text detection model was also utilised to detect oversize signs on HVs. Results show that the HV detection model achieved an 80% accuracy rate in detecting and tracking HVs across the bridge. The speed detection model performed impressively, with an error of 0.74 km/h on OSOM vehicles. Additionally, the lane positioning model and text detection model correctly identified the lane positioning of each vehicle. This study demonstrates the feasibility of using artificial intelligence in the transportation sector to monitor heavy vehicles on road assets, with the findings suggesting that AI technology could be a valuable resource in protecting such assets.

Keywords: Machine Learning, Artificial Intelligence, Heavy vehicle monitoring

1. Introduction

The impact of heavy vehicles (HVs) on bridges and culverts has been a major consideration for transportation engineers for many years, with the Queensland Department of Transport and Main Roads managing more than 3,000 bridges alone. Class 1 Over Size Over Mass (OSOM) heavy vehicles, due to their weight and size, have a significant impact on the structural integrity of bridges and culverts. The National Heavy Vehicle Regulator (NHVR) (n.d.) identify that the common Class 1 vehicles include special purpose vehicles, agricultural vehicles and OSOM vehicles. Thus, a need for an efficient way to monitor and track these vehicles across road assets is needed to better understand the interactions of these vehicles on these structures. This information may then be used to ensure it aligns with asset management strategy and investment plans.

One solution for the monitoring of heavy vehicles on bridges at a network level is through the implementation of a smart network monitoring solution that utilises video analytics based on Artificial Intelligence. This approach has been proposed in several studies, including one by Liu et al. (2020). They developed a video-based system that uses deep learning algorithms to detect and track heavy vehicles on bridges. The system achieved high accuracy in detecting and tracking heavy vehicles, demonstrating its potential for monitoring heavy vehicles on bridges and culverts. Another study by Dong et al. (2020) proposed a similar system that uses a convolutional neural network (CNN) to detect and track vehicles on bridges. The system also achieved high accuracy in detecting and tracking heavy vehicles. In addition to monitoring heavy vehicles, video analytics can also provide insights into the behaviour of these vehicles on bridges and culverts. For example, Liu et al. (2020) used their system to analyse the speed and lane occupancy of heavy vehicles on a bridge, providing valuable information for bridge maintenance and design.

Through the National Asset Centre of Excellence, a collaborative research agreement between the Queensland Department of Transport and Main Roads, and the Australian Road Research Board, the potential use of video analytics and other technologies to provide insight into OSOM HVs, including detection, tracking, and identifying their mass, dimensions, speed, and lateral positioning were explored. Lateral position in this paper is in reference to the vehicle's lane position. This information will enable road managers to understand better the interactions of these heavy vehicles with road assets. The proposed system will enable road managers to gain a better understanding of how OSOM HVs interact with road assets, which can inform decision-making regarding the asset management of the network. By providing insights into the mass, dimensions, speed, and lateral positioning of these HVs, the system can facilitate informed management of the network and minimise unexpected impacts on structures. The project examines the ability of video analytics techniques using 2D (monocular) camera footage, identifies the information that can be obtained and discusses the method’s limitations. Recommendations on enhancing video analytics capabilities through integration with other technologies are also provided in Section 6.

The paper showcases the development of the proof of concept and assesses its feasibility using current technology. Section 2 of the paper provides a description of the data used in the study. Section 3 outlines the methodology employed to develop the proof of concept (PoC). Section 4 provides a detailed discussion of the models developed during the study. Section 5 presents a comprehensive analysis of the results obtained from the study. The final section of the paper draws conclusions, identifies the limitations of the study, and presents opportunities for future research.

2. Data

The camera data used for the research project was collected by TMR in the Isaac Region of Central Queensland. The cameras were positioned at a bridge on the Peak Downs Highway, with Camera 1 facing north and Camera 2 facing south. The cameras were triggered when strains exceeded a threshold corresponding (approximately) to a 10 t heavy vehicle, with the resolution of the strain gauges being 0.1 microstrain. Footage from each triggered event is presented in a TMR and EngAnalysis dashboard in mp4 format. The frame rate of the captured videos was reduced from 25 FPS to 10 FPS and resolution reduced from 1920x1080 to 854x480 due to limited LTE bandwidth at the site. The image quality available for edge computing processes is higher. The aim was to use a pre-trained model, so there was no need for any additional data cleaning or annotation. This method allowed a fast test of the object detection models and an evaluation of their performance on the collected data. The large number of videos captured each day provided a rich dataset to work with, in the order of 15,000 frames of footage from heavy vehicle events each day, ensuring that the results were representative and accurate.

The data used in this study has several limitations. The data was only collected from a single test site, and raises concerns about the generalisability of the results, as one single test site may not represent the diversity of environments and scenarios in which heavy vehicles traverse bridges. The camera footage was recorded with a FPS of 10, which is considered low, and this could have resulted in inaccuracies in the object detection or other model results. A higher frame rate and improve resolution may enhance the accuracy of the object detection and other model results. These limitations should be considered when interpreting the results of this study, and future studies should aim to include more test sites in diverse locations, and cameras with a higher FPS to further validate the findings of this research. Figure 2.1 shows a screenshot of a sample video captured at 7:37 AM on the 15th of January 2022 from the north and south directions respectively.

Figure 2.1 –Video Output of Camera

3. Methodology

The proof of concept (PoC) had a scope that included implementing several functionalities such as HV detection, speed detection, lateral positioning estimation, vehicle mass estimation, vehicle dimension estimation, and license plate recognition. The researchers were successful in developing a system that could perform HV detection, speed detection and lateral positioning estimation. However, they were unable to develop vehicle mass estimation and vehicle dimension estimation functionality using vision-based methods due to the need for more detailed information about the vehicle, which cannot be reliably inferred from camera footage alone. Lastly, license plate recognition is found to be potentially achievable using the current methodology, but difficult to achieve due to the limitations of the current video quality.

Figure 3.1 shows the framework and overarching concept of PoC with the subtasks involved. The video analytics model takes monocular video footage as input and estimates features in the frame. The tasks in the green boxes (Detect HV Speed, Detect lateral positioning of HV) were achieved using monocular camera footage. The task in the yellow-coloured boxes (Collect license plate data, Estimate HV dimensions) can be potentially achieved but are difficult to achieve with the current video quality and methodology. The task in the red coloured box (Estimate HV Mass) is possible by a combination of sensors but not possible to achieve using only video data. Furthermore, two independent models (not integrated into the end-to-end model) were also developed and tested. The two models are a License Plate Detection Model and a Signage Detection model.

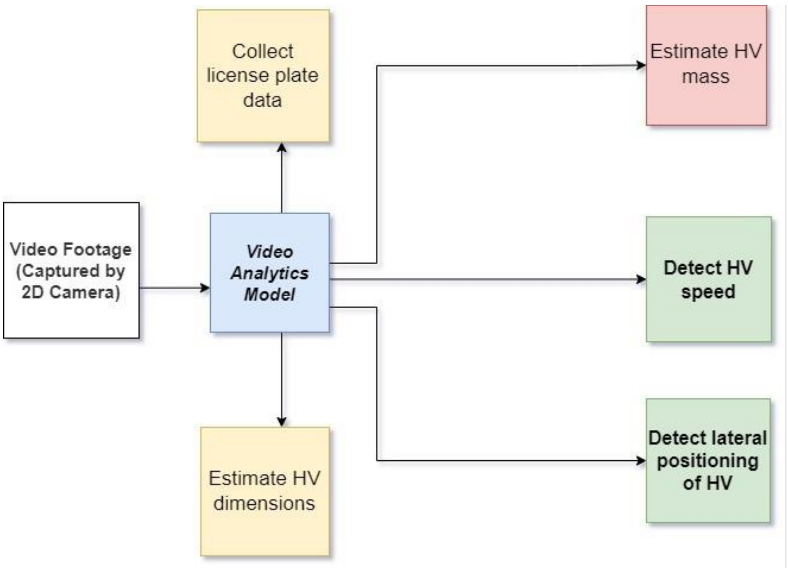

Figure 3.1 – Framework of System

Figure 3.2 depicts the end-to-end integrated model (methodology and models) developed. There were three sub-models developed for this end-to-end model: HV Detection Model (Deep

Learning Model), Speed Detection/Estimation Model, and Lateral Positioning Estimating Model. All three models use input data from Camera 1 only. In addition to the end-to-end model, the investigation also included two additional independent models which are License Plate Detection and Signage Detection model. The models are discussed in detail in Section 4.

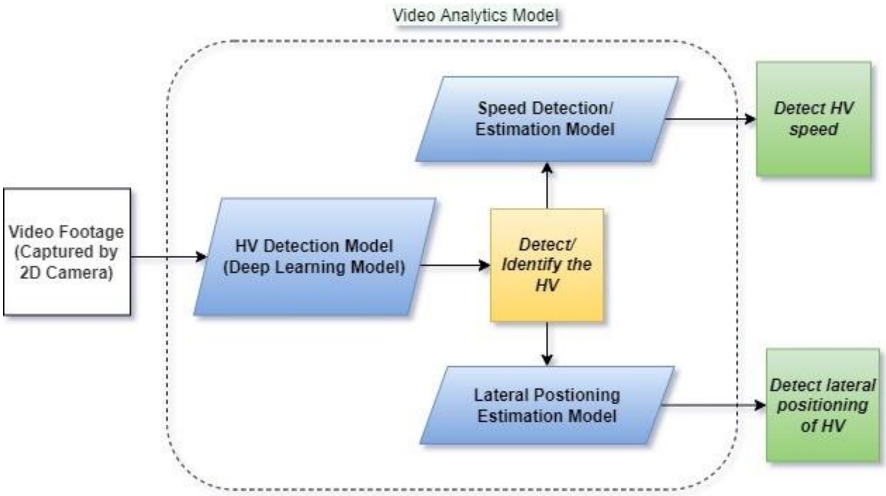

Figure 3.2 – Methodology Flowchart of System

4. Models

4.1. HV Detection and Tracking Model

The HV detection system uses a YOLOv4 deep learning model (Bochkovskiy, Wang, and Liao 2020) to perform object detection. The YOLOv4 model does not require any specialised devices and can be run on conventional graphics processors, making it accessible for use. The model was pretrained using the COCO dataset (Lin et al. 2015), which provides a large collection of labelled images of everyday objects, making it ideal for training object detection models. The implementation of the YOLOv4 model enables real-time object detection with visual output by leveraging its optimised architecture, which allows for fast and efficient processing of video input, meaning it can be an effective tool for detecting objects in real-time. A robust detection and tracking system is evidenced by the ability to detect vehicles from a considerable distance, ensure accurate vehicle detection and classification, and sustain an unchanging tracking ID as long as the vehicle remains within the frame.

4.2. Speed Estimation Model

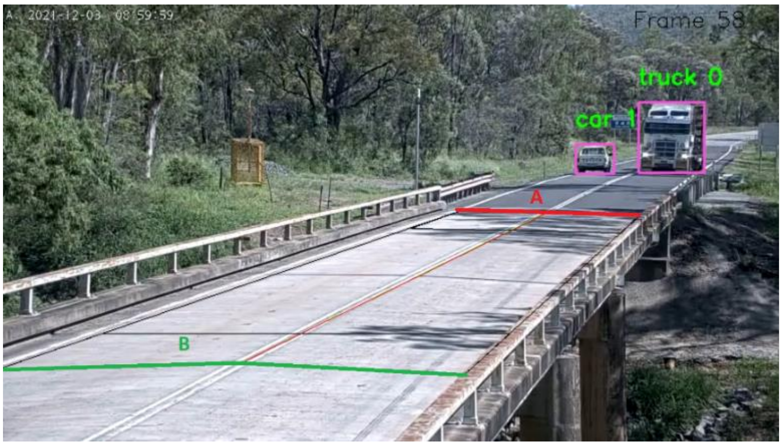

Figure 4.1 - Output of System showing Start A and End B for Speed Estimation

The method of calculating a vehicle's speed was to identify the time taken for the vehicle to cross the known length of the bridge. Figure 4.1 is an example frame extracted from the data, where the distance between ***line A*** and ***line B*** is used to estimate the speed. The distance between ***line A*** and ***line B*** was measured on site for the calibration of this model. This translates that this speed detection model is required to be calibrated for every site. The Speed Estimation Model is a computer vision geometric model and doesn't require a training image dataset. However, as previously mentioned, the model must be calibrated with a known distance in the frame. By counting the number of frames it took for a vehicle to cross the known distance, it was possible to compare that to the frames per second (FPS) of the input video, which for this research was 10 FPS, to get the time taken for the vehicle to cross the bridge. Using this known distance and the time taken for a vehicle to cross it, a vehicle's speed could be identified. Figure 4.1 shows a sample output of the system. For the test bridge site, ***line A*** indicates the first line for the distance calculation while ***line B*** indicates the second line for the distance calculation. The actual (physical) distances between these lines are 54.66 metres. Therefore, if it takes 24 frames for a vehicle to cross this distance, at 10 frames a second, it is possible to estimate that the vehicle was travelling at approximately 22.7 m/s or 82 km/h.

4.3. Lateral Positioning Model

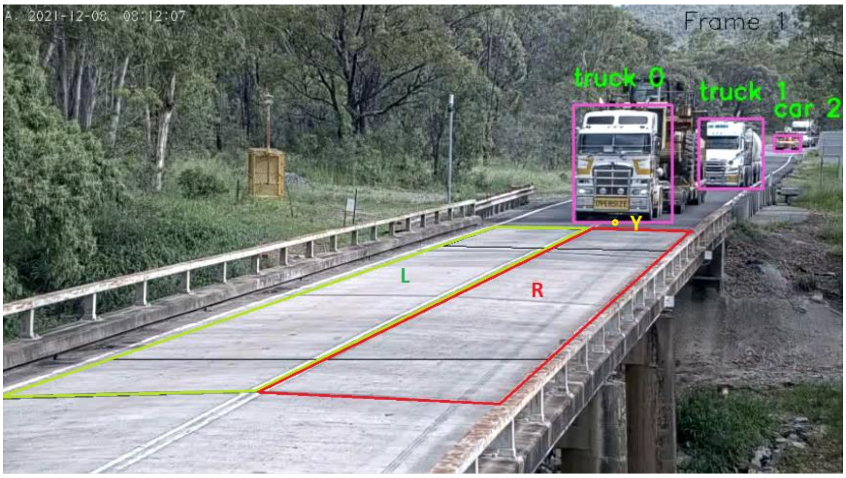

Figure 4.2 - Output of System showing Box L and Box R for Lane Identification and Point Y

Lateral position in this paper is in reference to the vehicle's lane position. The Lateral Positioning Estimation Model uses computer vision and geometric techniques to estimate a vehicle's lateral position on a bridge without the need for field calibration. The model draws two bounding boxes over each lane as shown in Figure 4.2 and calculates the position of multiple points on a vehicle (Point Y) relative to the boxes (Box L and Box R). If a point falls within a specific box, the vehicle is determined to be in that lane. This approach indicates which lane the truck is in without the need for fieldwork.

4.4. Oversize Sign Recognition Model

The detection of oversize (OS) signs on a HV is crucial in validating if they are within regulation dimensions, as any HV could be OSOM due to their dimension or weight. Therefore, OS signs are an important identifier in inferring if a vehicle is an OSOM HV without extracting dimension or weight information. It is important to note that an OS sign only indicates that a vehicle is likely to be a Class 1 or Class 3 OSOM vehicle, and any positively identified vehicle may or may not be subject to bridge travel restrictions. To achieve this, a pre-trained Optical Character Recognition (OCR) model was integrated into the system, to detect the keywords "over" and "size." For the OCR process, the EasyOCR (Jaided AI, 2022) library was used, which made it simple to run the OCR on the bounding boxes of detected vehicles and extract the text, which was then searched for the targeted keywords. The use of OS signs as a means of validation helps to ensure that the detected vehicles are indeed Class 1 or Class 3 OSOM HVs.

5. Results

This section discussed the results and the performance of the models developed.

5.1. HV Detection Model

The HV Detection model correctly detected and tracked HVs 72% of the time in low-speed bins (<15 km/h) and 93.7% of the time in high-speed bins (>70 km/h) from a validation dataset with a sample size of 46. The model's performance in detecting and tracking HVs is indicated by the accuracy of the results, however, a notable difference between slower and faster speeds was observed. Further inspection of the test videos was done to investigate the difference in model performance for high-speed and low-speed bins. It was noticed that a higher portion of the low-speed bin vehicles are not classic-looking tautliner HVs or flatbed HVs which the model is likely to be trained upon, but instead, many were low loaders carrying oversized mining equipment. Figure 5.1 shows a frame where the model misses a detection of an OSOM HV carrying mining equipment. The lack of identification can be traced to the high number of Class 1 OSOM HVs[1] which can provide difficulties in detecting OSOM HVs, hence the inaccuracies in detecting HVs in the low-speed bin.

This means that the model requires training with a more diverse or specific dataset. This kind of behaviour from models may be due to the training data (COCO dataset) being diverse and not prepared for a specific task, in addition to fewer training images of OSOM vehicles, as the indivisible loads they transport are not consistent dimensions. To further improve performance, the YOLO model can be trained with a custom dataset. The current accuracy (72%) is good at detecting and tracking vehicles over bridges. Since the time of the study, more advanced architecture, such as YOLOv8 (Ultralytics, 2023), has been developed, which could improve the model's accuracy further.

Figure 5.1 – Output of System Showing a Missed Detection of an OSOM HV

5.2. Speed Estimation Model

We were able to estimate the speed of vehicles on the bridge with positive results. At low speeds, an error of 0.7 km/h was calculated, while at higher speeds, an error of 2.8 km/h was calculated. This translates to the model exhibiting higher accuracy at lower speeds. With the goal of comparing the speeds of Class 1 HVs to the speed limit on road assets, which is 10km/h on the test bridge for certain platform configurations, the lower variance at low speeds is more

[1] Many of the low-speed events were associated with Class 1 HVs carrying mining equipment and which are variable in appearance due to the different payloads. The poorer detection performance of the pre-trained model for lower speed events likely results from limitations of the pre-trained model variable based on common vehicle class distribution rather than Class 1 HV's carrying mining equipment.

important to our program, showing that we have a higher precision in detecting the speeds of these vehicles. The estimated speed's error of 0.7 km/h indicates that our model performance is impressive, with 68% of the results in the low-speed bin being within 0.7 km/h from the speeds that were detected by bridge sensors, and 95% being within 1.5 km/h away from the sensor speeds.

5.3. Lateral Positioning Model

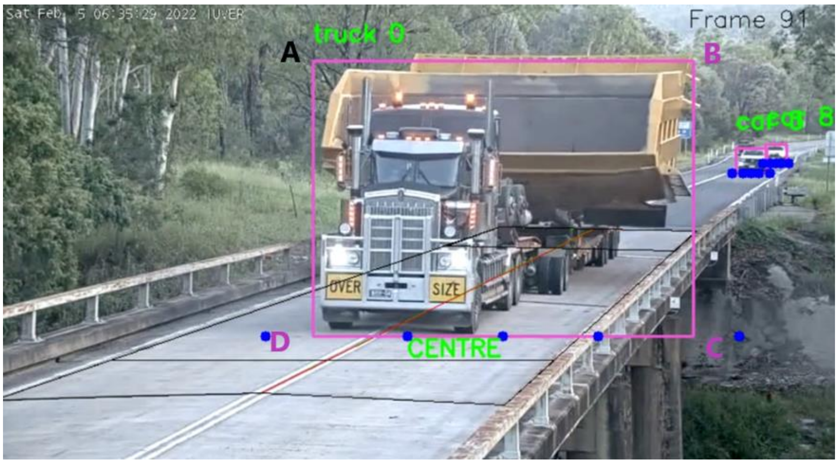

Figure 5.2 – Output of System Showing Bounding Box ABCD

The research has successfully designed a system that could identify a vehicle's lateral position on the bridge using only video data recorded from a monocular camera. The system can very accurately and confidently identify a vehicle's lateral position on the bridge with little error. In both our test datasets (low and high-speed), the vehicle's lateral position was estimated correctly 100% of the time based on manual review, identifying itself as an excellent model in estimating the lateral positioning of vehicles. However, the robustness of the model and our methodology must be further tested on new data and at new sites with differing camera angles.

The system predicts the lateral positioning of a detected vehicle for each frame in the video. After an entire video is analysed, the postprocessing system outputs the most common (mode) driveline for each vehicle. Figure 5.2 is a screenshot of the video output of the program. It detects the vehicle and recognises it as a truck. It then uses the blue points shown on the bottom of the bounding box ABCD (edge CD) to dictate the vehicle's estimated driveline by identifying the positioning of the points in relation to the drawn polygons on the frame. Additionally, a tracking id is shown above the bounding box, allowing the object to be tracked throughout frames.

5.4. Oversize Sign Recognition Model

We tested our Signage Detection Model on both our low and high-speed test datasets, where we collected sample data of 18 low-speed vehicles, and 16 high-speed vehicles. 100% of the low-speed vehicles (18) contained OS signs and 31% of the high-speed vehicles (5) contained OS signs, totalling 23 sample OS signs out of a total sample size of 34. In the 23 videos containing OS signs, our system outputted a detection 100% of the time. In the 11 videos which didn't contain OS signs, our system didn't output any false positives. This results in both a precision and recall of 100%, which shows itself to be an excellent model for detecting OS signs on vehicles. This value may indicate that our test dataset is too small to have a high degree of confidence in this result, meaning that while the results are extremely promising, more validation should be undertaken.

5.5. Final Output and performance for End-to-End Model

Each model was combined into a single end-to-end program which allowed for each model to be run concurrently, outputting the results in near real-time. Table 1 shows the sample CSV output of the program, reporting each detected vehicle's class, driveline, speed, and OSOM signage.

Table 1 – CSV Output of System on Test Video

Id	Class	Driveline	Speed	OSOM
0	truck	CENTRE	75.7	TRUE
5	car	RIGHT	81.9	FALSE
8	car	RIGHT	85.6	FALSE

Table 2 – Features and their Performance Metrics

Model	Performance
Heavy Vehicle Detection	Recall Low Speed: 0.7 High Speed: 0.9
Speed Detection (km/h)	Error Low Speed: 0.7 High Speed: 2.8

Table 2 summarises the two models and their performance. The Heavy Vehicle detection model showed a recall of 0.7 for low speeds and 0.9 for high speeds. Additionally, the speed detection model showed a error of 0.7 km/h for low speeds and 2.8 km/h for high speeds. Recall is a metric used to evaluate the proportion of true positives that a model correctly detects out of all positive instances. For instance, a recall of 1 means that a model has detected all positive instances without any false negatives. The lateral positioning and OSOM signage detection models had a recall of 1, indicating that more testing should be conducted before we can be confident in that result, as the likelihood of a 100% accurate model is unlikely.

6. Closing Remarks

The developed system provides valuable insights into the feasibility of using video analytics to detect and monitor Class 1 OSOM HVs on road assets. The findings indicate that speed estimation, OSOM vehicle detection, and lateral positioning estimation are achievable with current monocular object detection techniques at a high degree of accuracy. However, it was also identified that vehicle mass estimation and vehicle dimension estimation weren't feasible with the current methodologies, although could be possible with the integration of monocular video and other sensors.

The vehicle mass estimation and vehicle dimension estimation functionality cannot be developed using vision-based methods because they require more detailed information about the vehicle than can be captured using cameras alone. In particular, accurate mass estimation requires knowledge of the vehicle's weight distribution, which cannot be reliably inferred from visual data. Similarly, accurate dimension estimation requires precise measurements of features such as tyre size, which may not be visible in camera footage or may be subject to perspective distortion. Therefore, alternative methods such as weighing scales or manual measurements are typically used to obtain accurate vehicle mass and dimension information. A combination of such systems can provide more accurate and robust information than a single monocular camera. A study by Chen et al. (2019) proposed a method for detecting overweight vehicles on bridges using video analytics. The system uses a machine learning algorithm to classify vehicles based on their weight and determine whether they exceed the legal weight limit. The system achieved a high accuracy in detecting overweight vehicles and can be used to improve monitoring capabilities of heavy vehicles on bridges. Similarly, Li et al. (2019) proposed a system for identifying overloaded vehicles using video analytics. The system uses a convolutional neural network (CNN) to detect overloaded vehicles based on their axle weight distribution. The system achieved a high accuracy in detecting overloaded vehicles and can be used to improve monitoring capabilities of heavy vehicles on bridges and culverts. Recently, Zhang et al. (2021) proposed a system that combines video analytics with structural health monitoring (SHM) for real-time monitoring of bridges. The system uses machine learning algorithms to analyse video data and detect abnormal vehicle behaviour, while also using SHM data to monitor the structural health of the bridge. The system can provide real-time alerts to transportation authorities when abnormal vehicle behaviour is detected, enabling them to collect valuable data necessary to prevent damage to the bridge.

The developed system and models demonstrate a high accuracy level and performance, but it does have some limitations. As the object detection model is based on video footage, the system relies on adequate contrast to detect features. To work at night, either low-light cameras or adequate lighting must be installed. Night-vision cameras could be used to monitor OSOM vehicles in low-light conditions, as demonstrated in a study by Song et al. (2018) that used infrared cameras to detect and track heavy vehicles on a bridge. Additionally, the object detection model has poor performance when detecting OSOM HVs when they have oversize or irregularly shaped loads, such as oversize mining equipment. This can lead to missed detections which means that certain OSOM HVs may not be tracked. Furthermore, the data used in the study is limited, as it was collected from only one test site and may be improved with a higher frame-rate. This could potentially impact the generalizability of the findings and result in inaccuracies in object detection or other model results. Therefore, it is important to consider these limitations when interpreting the results. Future research should aim to include data from more test sites in diverse locations and higher FPS to further validate the findings.

As the use of video analytics continues to evolve and advance, it is important to consider the potential future implications and applications of using this technology to detect and monitor OSOM vehicles on road assets. Future work should be undertaken to use more modern object detection and tracking models, particularly in researching different tracking techniques, as minor tracking errors can lead to missed detections in our models. Future work should be undertaken in developing a custom dataset on OSOM vehicles to be able to custom train an AI model on OSOM vehicles and their non-standard loads. Additionally, research should be completed on using different camera types, such as depth or night-vision cameras to improve the performance of OSOM vehicle monitoring. For example, depth cameras have shown potential in accurately estimating the size and shape of objects in traffic scenes (Li et al., 2021). By developing further research in this field, cheaper and smarter ways to monitor our freight networks and obtain assurance data can be developed that lead to better asset management decisions.

7. References

Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

Chen, H., Wu, J., Du, Y., Huang, C., & Wang, J. (2019). Detection of overweight vehicles on bridges using deep learning-based computer vision. Journal of Bridge Engineering, 24(6), 04019032.

Dong, J., Ma, Y., Wang, X., & Zhu, L. (2020). A vehicle detection and tracking system for bridge inspection using convolutional neural network. IEEE Access, 8, 122116-122126.

Jaided AI. (2022). EasyOCR. GitHub. Retrieved from https://github.com/JaidedAI/EasyOCR

Li, J., Wang, K., Yao, Y., Yang, K., & Zeng, L. (2021). Depth Video-based Detection and Tracking of Vehicles in Traffic Scenes. IEEE Transactions on Intelligent Transportation Systems.

Li, X., Wang, H., Zhu, S., Liu, S., & Chen, X. (2019). Overloaded vehicle identification on bridges based on deep convolutional neural networks. IEEE Transactions on Intelligent Transportation Systems, 20(5), 1755-1764.

Lin, T. Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., ... & Dollar, P. (2015). Microsoft COCO: Common Objects in Context. arXiv preprint arXiv:1405.0312.

Liu, S., Wang, H., Li, H., Li, X., & Chen, X. (2020). Video-based vehicle detection and tracking system for bridge safety monitoring. Journal of Computing in Civil Engineering, 34(1), 04019059.

National Heavy Vehicle Regulator (n.d.), Class 1, NHVR, accessed 21 April 2023, <https://www.nhvr.gov.au/road-access/mass-dimension-and-loading/classes-of-heavy-vehicles/class-1>.

Song, Q., Jin, W., Liu, H., & Zhang, S. (2018). Infrared thermal imaging-based vehicle detection and tracking for bridge safety monitoring. Sensors, 18(7), 2096.

Ultralytics. (2023). YOLOv8. GitHub. Retrieved from https://github.com/ultralytics/ultralytics

Zhang, S., Peng, W., Liu, Y., & Qiu, X. (2021). Video-assisted vehicle behavior monitoring and structural health monitoring for real-time bridge safety assessment. Automation in Construction, 122, 103506.