DEEP LEARNING-BASED ROAD TRAFFIC DENSITY ANALYSIS AND MONITORING USING SEMANTIC SEGMENTATION

[1*]ADITHYA KUSUMA WHARDANA, [2]PARMA HADI RENTELINGGI

[1]Department of Informatics Engineering, Tanri Abeng University

Jl. Swadarma Raya No.58, Ulujami, Kec. Pesanggrahan, Kota Jakarta Selatan

[2]Graduate School Science and Technology, Keio University, Japan

2-15-45 Mita, Minato City, Tokyo 108-0073, Japan

e-mail: [1]adithya@tau.ac.id, [2]parma@ohtsuki.ics.keio.ac.id

[*]Corresponding author

ABSTRACT

Due to factors such as a growing population, more people using private vehicles, and outdated transportation infrastructure, Jakarta, the capital city of Indonesia, suffers from chronic traffic congestion. The environment, citizens' safety, productivity, and quality of life are all negatively impacted by these interruptions. In response to these difficulties, this study proposes a novel method for traffic monitoring. By combining YOLOv5, optical flow, and recurrent neural networks (RNN) with image processing and artificial neural networks, a unified traffic monitoring system can be achieved. We went with YOLOv5 because of how well it identifies various automobiles. The number of vehicles is counted between video frames using Optical Flow, and then the traffic density is classified using RNN. With an accuracy of 87% following testing, RNN was clearly a winner when it came to vehicle density classification. The goals of this research are to lessen the societal and environmental toll of traffic congestion, increase our knowledge of and ability to control Jakarta's traffic, and lay the groundwork for the creation of more advanced traffic monitoring systems. The growing traffic issues in the nation's capital are anticipated to be alleviated with this strategy.

Keywords: *Deep Learning, Optical flow, Recurrent Neural Network, Semantic Segmentation*

1. INTRODUCTION

Jakarta, Indonesia's central city, has significant difficulties in controlling its extreme traffic congestion. Jakarta, one of the most populated cities in the world, has serious traffic issues due to rising private automobile usage, greater population mobility, and inadequate transportation infrastructure. In addition to decreasing productivity, traffic congestion endangers people's safety, quality of life, and environment. In this situation, traffic monitoring requires different strategies [1].

The advancement of digital technology and image processing presents novel traffic control and monitoring prospects. Because Recurrent Neural Networks (RNN) can process sequential data, they can be a helpful solution. RNN can leverage optical flow technology to gather digital image data related to traffic. By integrating RNN into traffic monitoring systems, we can expect a better understanding of traffic movements and more efficient congestion handling. RNN in integration with Optical Flow [2]–[4] and YOLOv5 (You Only Look Once) technology [5]–[7].

YOLOv5 remains a superior choice for detecting objects and identifying different types of vehicles [5]. The main objective of this research is to explore how Optical Flow, RNN, and YOLOv5 can be integrated into an integrated traffic monitoring system. Before the model is run, the video taken as test data will first go through a segmentation process using the semantic method so that objects are more accessible to recognize. Through the combination of the strengths of these technologies, this research can provide deeper insights into traffic monitoring. On traffic management in Jakarta. The results of this research are expected to serve as a foundation for the development of more sophisticated and responsive traffic monitoring systems in the future, with a better understanding of traffic movements to reduce their negative impacts on society and the environment.

2. RESEARCH METHODOLOGY

2.1 Flowchart Research

Semantic segmentation and recursive neural networks are employed in this research, which is innovative for its application. Innovative application in the study carried out as presented in Figure 1; it shows the progression of techniques used in the study. shows the progression of the research's methodologies.

The study's proposed methodology is depicted in Figure 1, which begins with the collection of data in the form of videos taken directly on the Daan Mogot and Jakarta roads. Each video is then labeled, and the data is processed for semantic segmentation. Following semantic segmentation, each motorcycle or vehicle that wishes to be processed using the Yolo model is given a boundary box. Finally, a recurrent neural network (RNN) is used to train the Yolo process. The RNN model is used to test the data following the model training procedure.

2.2 Semantic Segmentation

In the realm of digital image processing, one segmentation technique is semantic segmentation. Vehicles in particular can be recognized by semantic systems, and each one is then given a color label [8]. Semantic will explain how each pixel in an image is grouped. Every frame of the movie used as test data will be processed [9]–[12]. The semantic segmentation method is illustrated in figure 2 and involves processing the image to make it easier for the created model to recognize and process.

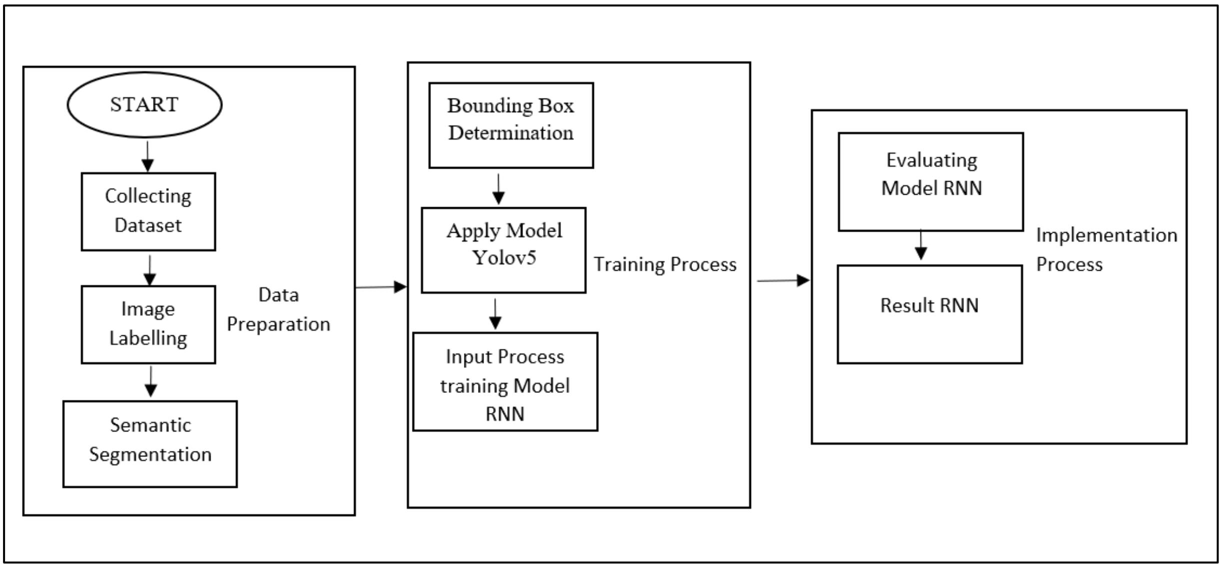

Figure 1. Flowchart Proposed Method

Figure 2. Semantic Segmentation

 P-ISSN: 2528-0260 E-ISSN: 2579-5392
p.1-8

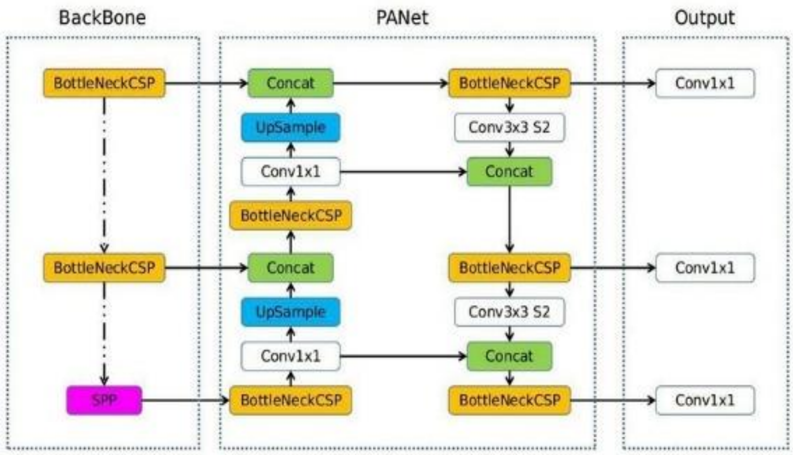

Figure 3. YOLOv5 model

Semantic segmentation is a computer vision problem that involves segmenting an image into many regions or segments based on the semantic significance of individual pixels. To put it another way, it entails recognizing each pixel in an image and categorizing it into predefined groups or categories, such "person," "road," "car," "tree," etc. Compared to basic object identification, which merely identifies objects in an image and draws bounding boxes around them, semantic segmentation assigns a name to each pixel in an image, providing a more thorough understanding of the scene.

2.3 YOLO (You Only Look Once)

The year 2021 saw the creation of the YOLOv5 algorithm. When it came to finding the front cab of big trucks, the results demonstrated that the taught algorithm was very accurate. Kasper-Eulaers was able to use this method to find winter rest spots for huge vehicles [13], [14]. Here are four variants of the YOLOv5 code that are available to the public: YOLOv5x, YOLOv5l, YOLOv5m, and YOLOv5s [5]–[7]. In terms of feature map breadth and depth, YOLOv5s is the most compact. Ultralytics LLC suggested YOLOv5, which stands for "You Only Look Once," in May 2020. Lighter in design, it possesses real-time image recognition capabilities at a video frame rate of 140 fps. Just over 27 MB in size, or one-ninth of YOLOv4's total, is YOLOv5. The variations of YOLOv5 are YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5 [15], [16], with each variant corresponding to a different feature map width and network depth.

Among popular image data analysis tools, Yolov5 stands out for its high degree of accuracy and fast detection models. The yolov5 model makes research easier by automatically merging several input photos into a single image when the suggested algorithm identifies things that can be detected with a detection time of just 2 milliseconds per picture.

One effective and widely used real-time object identification model is Yolov5. Developed by Glenn Jocher and his colleagues at Ultralytics, it is the latest version of the YOLO (You Only Look Once) series of models. Some important things to note about Yolov5 are: Instant Object Recognition: Applications requiring quick response times, such vehicle recognition on the road, can benefit from Yolov5's real-time object detection capabilities. In order to teach Yolov5 to detect certain things, you can use a variety of datasets. In a number of tests, Yolov5's ability to detect things was found to be quite good.

2.4 Optical Flow

One way to monitor the progression of items in a video or picture sequence from one frame to the next is by using optical flow [2]–[4]. The goal of optical flow is to represent the interframe variation in pixel intensity as a translation in two-dimensional space. When these two ideas are integrated, a system can use YOLO for object detection and optical flow for object movement estimation in the same scene. Possible uses include intrusion detection, traffic monitoring, and video monitoring for motion [17]–[19]. Robotics, motion estimation, computer vision, and video analysis are among fields where optical flow is used. This is how it functions: Pixel-based motion estimation. Optical flow methods analyze the variations in pixel positions between frames by studying the intensity patterns in consecutive frames. The approach determines the direction and magnitude of motion for each pixel by comparing the brightness of pixels between frames. Optical flow methods often rely on two assumptions: the spatial coherence assumption, which states that neighboring pixels move in a similar way, and the brightness constancy assumption, which states that pixel intensities remain constant between consecutive frames. These assumptions allow the software to accurately deduce motion vectors. Mathematical Representation: Optical flow is commonly visualized as a 2D vector field, where each vector corresponds to the velocity of a pixel. The goal of optical flow is to reduce the disparity between predicted and observed pixel intensities. This can be formulated as an optimization problem or a partial differential equation (PDE).

2.5 Recurrent Neural Network

Recurrent Neural Networks (RNNs) are most appropriate for jobs that involve sequences, such as speech and natural language processing. These networks have the ability to identify sequential relationships since they operate on the principle of retaining previous inputs [20]–[22]. Recurrent Neural Networks (RNNs) are effective at tasks that involve input sequences of varying lengths. RNNs are effective for analyzing time series data due to the significance of the temporal sequence of the data points. RNNs have versatile applications in several domains, particularly in time-dependent tasks, as they possess the ability to model the interdependence among sequentially arranged data points. RNNs are classified as deep learning models because they perform computations on incoming data across numerous levels of processing.RNNs have a higher parameter count compared to traditional machine learning models, which may lead to training difficulties. In order to effectively utilize Recurrent Neural Networks (RNNs), it is necessary to possess proficient training [23]–[25]. Equation (1) is the formula for a fundamental Recurrent Neural Network (RNN).

$$ht = \tanh(Wih \cdot xt + Whh \cdot ht - 1) \quad (1)$$

ht the concealed state at the specified time step ***t.*** tanh The hyperbolic tangent function is a mathematical function that scales its input to a specific range of -1 to 1, represents the input-to-hidden-state connection weight matrix. Whh refers to the weight matrix that connects hidden states. *xt* represents the data input at the current time step *t. ht-1* what the concealed condition is at time step t-1.

3. RESULTS AND DISCUSSIONS

3.1 Segmentation

The segmentation method employed is a semantic approach that assigns color labels to items in every frame. The segmentation procedure is performed independently using the model specifically designed to get more efficient outcomes. This process is influenced by the device's image processing capabilities. The training and labeling datasets are obtained from the COCO and Pascal VOC databases [26]–[28]. The dataset labeling assigns the number label 7 to vehicles and 14 to motorcycles. The footage was captured at JPO Jelambar, specifically on the Daan Mogot highway. Each video has a duration of approximately 5 seconds, consisting of approximately 150 to 170 frames.

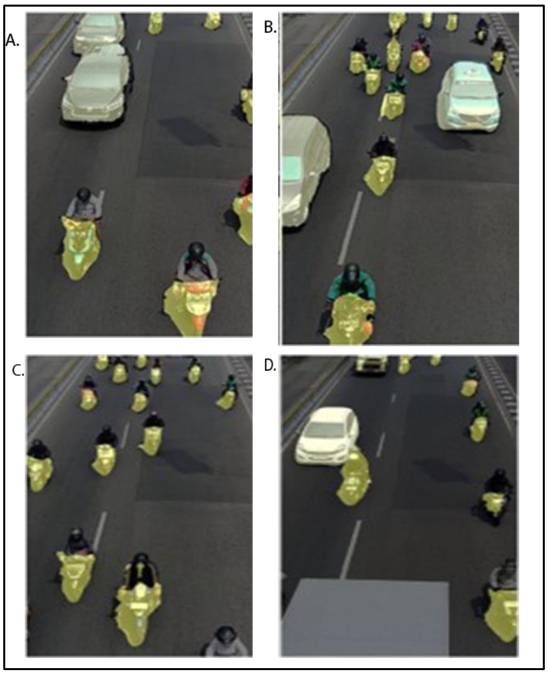

Figure 4. Result Semantic Segmentation

Figure 5. Object Detection and Labeling Result YOLOv5

Figure 4 displays the outcome of semantic segmentation, which involves segmenting multiple video datasets that have been captured. Figure 4 (a) shows the semantic segmentation of the first video, which has a moderate level of density. Figure 4(b) and Figure 4(c) also depict the semantic segmentation of datasets with a moderate level of density, with Figure 4(d) specifically focusing on motorbikes crossing the Daan Mogot area. Figure 4(d), on the other hand, represents the semantic segmentation of data with a very high level of density, featuring numerous motorbikes and cars crossing the video capture area.

3.2 Object Detection and Labeling

The YOLO algorithm is utilized for the purpose of object recognition and labeling. The YOLOv5 version of the YOLO model has been imported into the research using the PyTorch library. OpenCV is used to analyze traffic footage by examining each frame individually. Subsequently, each frame is provided as input to the YOLOv5 model for the purpose of object detection, specifically targeting cars within the frame. The detection results consist of a bounding box that specifies the geographic position of the object, a confidence value that indicates the level of certainty that the object exists, and the class ID of the identified object. YOLOv5 exhibits much superior performance compared to its predecessors. Earlier iterations of YOLO required more time to analyze videos of same length [29].

YOLOv5 demonstrates good performance in both in object detection and labeling. However, there are difficulties in object detection with the label 'motorcycle' because it is covered by the color of the previous segmentation result. Although detection and labeling of the 'motorcycle' object cannot be done, it can be overcome or handled by the detection of the motorcyclist so that the number of vehicles detected in each frame will still match those in the testing video. The results can be seen in Figure 5.

The Optical Flow technique is employed to estimate traffic density by quantifying the number of cars in motion inside the video. Optical Flow is utilized on grayscale frames to compute the flow of pixel intensity between consecutive frames. Optical Flow is a technique used to monitor the displacement of objects between consecutive frames. The utilization of Optical Flow is crucial in vehicle counting, as it incorporates both prior and current frames to detect alterations in vehicle position. This enables the recording of cars that enter or exit the monitoring area.

Traffic density predictions are computed via a Recurrent Neural Network (RNN) model. The Recurrent Neural Network (RNN) was trained utilizing traffic density data that was gathered using Optical Flow technology. Optical Flow is utilized to produce predictions in each iteration through a forward pass. The Mean Squared Error (MSE) function is used to calculate the loss by quantifying the difference between the projected output and the actual goal. Subsequently, a reverse pass is performed to calculate the gradient. Stochastic Gradient Descent (SGD) is used as the optimizer to modify the model parameters [30], [31].

Once the training process is completed, the Recurrent Neural Network (RNN) model is employed to forecast forthcoming traffic density by leveraging the information from the provided sample data points. Subsequently, these forecasts are incorporated into the data points for the subsequent iteration. The final outcome is a mean loss value that provides an indication of the model's ability to accurately forecast future traffic density. This ability is referred to as the model's accuracy, which is measured at 87%.

4. CONCLUSION

By incorporating segmentation, object detection, and traffic density prediction algorithms into current methodologies, a more comprehensive comprehension of traffic dynamics can be achieved. This device can serve as a highly efficient instrument for monitoring traffic in diverse settings, offering exceptional speed and precision.

REFERENCES

[1] K. Khazukov *et al.*, (2020), "Real-time monitoring of traffic parameters," *Journal of Big Data*, vol. 7, no. 1, pp. 1–20, doi:10.1186/S40537-020-00358-X/FIGURES/21.

[2] Y. Jiao, G. Shi, and T. D. Tran, (2021), "Optical Flow Estimation Via Motion Feature Recovery," in *Proceedings - International Conference on Image Processing, ICIP*, pp. 2558–2562.

[3] S. Ando and T. Kindo, (2022), "Direct Imaging of Stabilized Optical Flow and Possible Anomalies from Moving Vehicle," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 12, pp. 24044–24056, doi:10.1109/TITS.2022.3199203.

[4] J. Yuan, T. Jiang, X. He, S. Wu, J. Liu, and D. Guo, (2023), "Dynamic obstacle detection method based on U–V disparity and residual optical flow for autonomous driving," *Scientific Reports*, vol. 13, no. 1, pp. 1–10, doi:10.1038/s41598-023-34777-6.

[5] Y. Fan *et al.*, (2022), "Application of Improved YOLOv5 in Aerial Photographing Infrared Vehicle Detection," *Electronics*, vol. 11, no. 15, pp. 1–20, doi:10.3390/ELECTRONICS11152344.

[6] M. Kasper-Eulaers, N. Hahn, P. E. Kummervold, S. Berger, T. Sebulonsen, and Ø. Myrland, (2021), "Short Communication: Detecting Heavy Goods Vehicles in Rest Areas in Winter Conditions Using YOLOv5," *Algorithms*, vol. 14, no. 4, pp. 1–11, doi:10.3390/A14040114.

[7] L. Shao, H. Wu, C. Li, and J. Li, (2023), "A Vehicle Recognition Model Based on Improved YOLOv5," *Electronics*, vol. 12, no. 6, pp. 1–14, doi:10.3390/ELECTRONICS12061323.

[8] Y. Cai, L. Dai, H. Wang, L. Chen, and Y. Li, (2022), "DLnet With Training Task Conversion Stream for Precise Semantic Segmentation in Actual Traffic Scene," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 11, pp. 6443–6457, doi:10.1109/TNNLS.2021.3080261.

[9] S. Lu, Z. Luo, F. Gao, M. Liu, K. Chang, and C. Piao, (2021), "A Fast and Robust Lane Detection Method Based on Semantic Segmentation and Optical Flow Estimation," *Sensors*, vol. 21, no. 2, p. 400, doi:10.3390/S21020400.

[10] S. Liu, J. Cheng, L. Liang, H. Bai, and W. Dang, (2021), "Light-Weight Semantic Segmentation Network for UAV Remote Sensing Images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 8287–8296, doi:10.1109/JSTARS.2021.3104382.

[11] H. Ghandorh, W. Boulila, S. Masood, A. Koubaa, F. Ahmed, and J. Ahmad, (2022), "Semantic Segmentation and Edge Detection—Approach to Road Detection in Very High Resolution Satellite Images," *Remote Sensing*, vol. 14, no. 3, p. 613, doi:10.3390/RS14030613.

[12] M. S, N. P, and M. S. S, (2021), "Instance Segmentation for Autonomous Vehicle," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 9, pp. 565–570.

[13] T. H. Wu, T. W. Wang, and Y. Q. Liu, (2021), "Real-Time Vehicle and Distance Detection Based on Improved Yolo v5 Network," *2021 3rd World Symposium on Artificial Intelligence, WSAI 2021*, pp. 24–28, doi:10.1109/WSAI51899.2021.9486316.

[14] Z. Chen, L. Cao, and Q. Wang, (2022), "YOLOv5-Based Vehicle Detection Method for High-Resolution UAV Images," *Mobile Information Systems*, vol. 2022, doi:10.1155/2022/1828848.

[15] A. Li, S. Sun, Z. Zhang, M. Feng, C. Wu, and W. Li, (2023), "A Multi-Scale Traffic Object Detection Algorithm for Road Scenes Based on Improved YOLOv5," *Electronics*, vol. 12, no. 4, pp. 1–16, doi:10.3390/ELECTRONICS12040878.

[16] N. Al-Qubaydhi *et al.*, (2022), "Detection of Unauthorized Unmanned Aerial Vehicles Using YOLOv5 and Transfer Learning," *Electronics*, vol. 11, no. 17, pp. 1–18, doi:10.3390/ELECTRONICS11172669.

[17] K. Blachut and T. Kryjak, (2022), "Real-Time Efficient FPGA Implementation of the Multi-Scale Lucas-Kanade and Horn-Schunck Optical Flow Algorithms for a 4K Video Stream," *Sensors*, vol. 22, no. 13, doi:10.3390/S22135017.

[18] M. Yao, J. Wang, J. Peng, M. Chi, and C. Liu, (Oct. 2023), "FOLT: Fast Multiple Object Tracking from UAV-captured Videos Based on Optical Flow," *MM 2023 - Proceedings of the 31st ACM International Conference on Multimedia*, pp. 3375–3383, doi:10.1145/3581783.3611868.

[19] S. Shen, L. Kerofsky, and S. Yogamani, (2023), "Optical Flow for Autonomous Driving: Applications, Challenges and Improvements," *ArXiv*, doi:10.48550/ARXIV.2301.04422.

[20] J. M. Ackerson, R. Dave, and J. Seliya, (2021), "Applications of Recurrent Neural Network for Biometric Authentication & Anomaly Detection," *Information*, vol. 12, no. 7, doi:10.3390/INFO12070272.

[21] U. Mittal and P. Chawla, (Jan. 2023), "Acoustic Based Emergency Vehicle Detection Using Ensemble of deep Learning Models," *Procedia Computer Science*, vol. 218, pp. 227–234, doi:10.1016/J.PROCS.2023.01.005.

[22] I. Ullah and Q. H. Mahmoud, (2022), "Design and Development of RNN Anomaly Detection Model for IoT Networks," *IEEE Access*, vol. 10, pp. 62722–62750, doi:10.1109/ACCESS.2022.3176317.

[23] K. Griesbach, M. Beggiato, and K. H. Hoffmann, (2022), "Lane Change Prediction With an Echo State Network and Recurrent Neural Network in the Urban Area," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 6473–6479, doi:10.1109/TITS.2021.3058035.

[24] G. Kocher and G. Kumar, (2021), "Machine learning and deep learning methods for intrusion detection systems: recent developments and challenges," *Soft Computing 2021 25:15*, vol. 25, no. 15, pp. 9731–9763, doi:10.1007/S00500-021-05893-0.

[25] Y. Jeong, S. Kim, and K. Yi, (2020), "Surround vehicle motion prediction using lstm-rnn for motion planning of autonomous vehicles at multi-lane turn intersections," *IEEE Open Journal of Intelligent Transportation Systems*, vol. 1, no. 1, pp. 2–14, doi:10.1109/OJITS.2020.2965969.

[26] G. Cheng and J. Y. Zheng, (2020), "Semantic segmentation for pedestrian detection from motion in temporal domain," *Proceedings - International Conference on Pattern Recognition*, pp. 6897–6903, doi:10.1109/ICPR48806.2021.9411958.

[27] M. Imad, O. Doukhi, and D. J. Lee, (2021), "Transfer Learning Based Semantic Segmentation for 3D Object Detection from Point Cloud," *Sensors*, vol. 21, no. 12, doi:10.3390/S21123964.

[28] M. Colley, B. Eder, J. O. Rixen, and E. Rukzio, (2021), "Efects of semantic segmentation visualization on trust, situation awareness, and cognitive load in highly automated vehicles," *Conference on Human Factors in Computing Systems - Proceedings*, doi:10.1145/3411764.3445351/SUPPL_FILE/3411764.3445351_VIDEOFIGURE.MP4.

[29] Z. Deng, C. Yao, and Q. Yin, (2023), "Safety Helmet Wearing Detection Based on Jetson Nano and Improved YOLOv5," *Advances in Civil Engineering*, vol. 2023, no. 1, doi:10.1155/2023/1959962.

[30] M. Liu, B. Liao, C. Wang, Y. Wang, and Y. Wang, (Nov. 2020), "Real-Time Vehicle Taillight Recognition Based on Siamese Recurrent Neural Network," *Journal of Physics: Conference Series*, vol. 1673, no. 1, doi:10.1088/1742-6596/1673/1/012056.

[31] Y. Luo, Y. Xiao, L. Cheng, G. Peng, and D. D. Yao, (2021), "Deep Learning-based Anomaly Detection in Cyber-physical Systems," *ACM Computing Surveys (CSUR)*, vol. 54, no. 5, doi:10.1145/3453155.