Infrared and Millimeter-Wave Radar Fusion for Robust Night Pedestrian Recognition

G.Swathi Reddy[1], Boga Venkat Sai[2], Bommadeni Sricharan[3], Bookya Madhu Nayak[4], Chininga Praveen[5], Jula Shiva Rama Krishna[6]

[1] Assistant Professor, Department Of Computer Science And Engineering(AI& ML), Samskruthi College of Engineering And Technology , Kondapur(V), Ghatkesar(M), Medchal(D),Telangana

[2,3,4,5,6]BTech Students ,Department Of Computer Science And Engineering(AI& ML), Samskruthi College of Engineering And Technology , Kondapur(V), Ghatkesar(M), Medchal(D),Telangana

***Abstract*— Detecting pedestrians in low-light and nighttime environments is a major challenge for autonomous systems, mainly due to poor visibility and the limitations of relying on a single sensor. To address this issue, this project introduces a multimodal pedestrian detection framework that combines infrared (IR) imaging with millimeter-wave (MMW) radar data to improve detection accuracy, reliability, and real-time performance during night conditions. An enhanced YOLOv5-based deep learning model is utilized to extract meaningful spatial and semantic features from infrared images, while radar signals are processed to obtain important information such as distance, speed, and position of moving objects. To further enhance tracking accuracy, an Extended Kalman Filter (EKF) is applied to reduce noise in radar data and to predict the motion of pedestrians over time. The system performs spatiotemporal fusion by aligning radar target points with corresponding regions in infrared images, followed by a correlation gating mechanism to ensure proper association between the two data sources. Finally, a decision-level fusion strategy integrates the outputs to produce precise and consistent pedestrian detection results. Experimental results show that the proposed approach outperforms single-sensor methods, offering improved precision and robustness under challenging nighttime conditions. This work contributes to enhancing the safety and reliability of autonomous vehicles and intelligent monitoring systems operating in low-visibility environments.**

***Keywords*—autonomous vehicles, pedestrian detection, sensor fusion, YOLOv5, infrared camera, millimeter-wave radar, target tracking, Extended Kalman Filter, nighttime detection, real-time object detection**

I. INTRODUCTION

In recent years, autonomous vehicle technology has experienced rapid growth, particularly in the development of perception systems that enable vehicles to interpret complex traffic environments. A crucial function of these systems is the accurate detection and tracking of pedestrians and other obstacles to ensure safe navigation. To achieve this, modern perception frameworks integrate multiple sensing modalities such as cameras, LiDAR, and millimeter-wave (MMW) radar, which together provide comprehensive information about object types, positions, and motion patterns in the surrounding environment. This multisensor approach plays a vital role in reducing collision risks and improving overall driving safety [10].

Object detection is a fundamental component of these perception systems, focusing on identifying and localizing targets using sensor data. In recent years, deep learning methods, particularly convolutional neural networks (CNNs), have significantly enhanced detection performance. Single-stage detectors such as YOLO, SSD, and RetinaNet are widely recognized for their real-time processing capabilities [1], [6], [7], while two-stage methods like R-CNN and Fast R-CNN provide higher accuracy at the cost of speed [8], [9]. Among these approaches, the YOLO family has gained considerable attention due to its balance between speed and accuracy, with continuous improvements observed across its versions from YOLOv1 to YOLOv5 [1]–[5], [15]–[18].

Despite these advancements, vision-based detection systems face notable limitations under challenging environmental conditions such as low illumination, nighttime, fog, and occlusions. Visible-light cameras often fail to capture sufficient details in such scenarios, resulting in increased false detections or missed targets. In contrast, MMW radar operates reliably in adverse conditions by utilizing electromagnetic signals to estimate object distance and velocity [10]. However, radar lacks detailed semantic information, making it less effective for precise object classification.

To overcome these challenges, sensor fusion has emerged as an effective solution, combining the strengths of different sensing technologies. Vision sensors provide rich appearance and texture information, while radar contributes robust spatial and motion data, enabling more reliable decision-making. Several studies have explored fusion techniques, including low-level and spatial fusion methods, to improve detection performance [20], [21]. Additionally, approaches such as projecting radar data onto image space have shown promising results in enhancing cross-modal understanding [22]. However, many existing methods still rely on visible-light cameras, which remain sensitive to nighttime conditions.

To address this limitation, this work proposes a pedestrian detection framework specifically designed for low-light and nighttime environments by integrating infrared (IR) imaging with MMW radar data. Infrared cameras enable reliable visual perception in dark conditions by capturing thermal information, while radar enhances spatial awareness by providing accurate range and motion measurements [11]–[14]. The proposed system employs an enhanced YOLOv5 model trained on infrared data using transfer learning techniques to improve detection accuracy in low-illumination scenarios [15]–[18]. Simultaneously, radar signals are preprocessed to reduce noise and track moving targets using an Extended Kalman Filter (EKF).

Furthermore, the tracked radar points are projected onto the infrared image plane to establish correspondence between the two sensing modalities. A decision-level fusion strategy is then applied to combine radar and vision-based detections, improving both localization accuracy and classification reliability. This integrated approach effectively addresses the limitations of single-sensor systems and enhances detection performance in challenging nighttime environments.

Experimental results demonstrate that the proposed method achieves improved accuracy, robustness, and real-time performance compared to traditional approaches. This work contributes to the development of safer and more reliable autonomous driving systems, particularly in low-visibility conditions.

II. LITERATURE SURVEY

[1] J. Redmon et al., this paper introduced YOLO, a groundbreaking object detection framework that reframed detection as a single regression problem. Unlike traditional region-based methods, YOLO processes the entire image in one pass, dividing it into a grid and predicting bounding boxes along with class probabilities simultaneously. This design significantly improves detection speed, making it suitable for real-time applications such as autonomous driving. Although the first version of YOLO achieved remarkable speed, it struggled with detecting small objects and handling complex scenes with multiple overlapping targets. Nevertheless, the model laid the foundation for future improvements in the YOLO series. Its ability to balance speed and accuracy has made it highly influential in modern computer vision systems. In the context of pedestrian detection, YOLO provides a strong baseline for fast and efficient object recognition in dynamic environments.

[4] A. Bochkovskiy et al., YOLOv4 represents a significant advancement in object detection by introducing several architectural improvements and optimization techniques. The model integrates features such as CSPDarknet53 as the backbone, spatial pyramid pooling (SPP), and path aggregation networks (PANet) to enhance feature extraction and information flow. Additionally, techniques like data augmentation (Mosaic), DropBlock regularization, and CIoU loss contribute to improved detection accuracy. YOLOv4 achieves an excellent balance between speed and precision, making it suitable for real-time applications in complex environments. It also demonstrates improved performance in detecting small and occluded objects compared to earlier versions. This work is particularly relevant for applications like nighttime pedestrian detection, where robust feature extraction and efficient processing are critical. YOLOv4's design principles have influenced subsequent models, including YOLOv5, which further optimizes performance for practical deployment scenarios.

[10] S. M. Patole et al., this paper provides a comprehensive overview of automotive radar systems and their role in modern intelligent transportation systems. It discusses various signal processing techniques used in radar systems, including target detection, tracking, and classification. The authors highlight the advantages of millimeter-wave radar, such as its ability to operate reliably in adverse weather and low-visibility conditions, including fog, rain, and darkness. Radar systems are particularly effective in estimating object distance, velocity, and direction, making them essential for collision avoidance and adaptive cruise control. However, the paper also points out limitations, such as low

spatial resolution and limited capability in object classification. These drawbacks emphasize the need for combining radar with other sensors, such as cameras, to achieve better performance. This study strongly supports the idea of sensor fusion, which is central to improving pedestrian detection in challenging environments.

[13] Y. Liu et al., This work focuses on improving detection performance in complex environments using thermal infrared imaging. The authors propose a method that leverages the advantages of infrared sensors, which capture heat signatures rather than relying on visible light. This makes infrared imaging highly effective in low-light and nighttime conditions where traditional cameras fail. The study introduces enhancements in feature extraction and detection strategies to improve the identification of pedestrians and vehicles in challenging scenarios, such as cluttered backgrounds and varying temperatures. Experimental results demonstrate that infrared-based detection significantly improves reliability under poor lighting conditions. However, the paper also acknowledges that infrared images often lack detailed texture and color information, which can limit classification accuracy. This limitation highlights the importance of combining infrared data with other sensing modalities, such as radar, to achieve more robust and accurate detection systems.

[20] J. Kim et al., this paper presents a deep learning-based sensor fusion approach that combines radar data with monocular camera images for 3D object detection. The proposed method utilizes radar range-azimuth heat maps along with image features to improve detection performance. By integrating data at a low level, the system effectively captures both spatial and visual information, leading to more accurate object localization and classification. The study demonstrates that radar provides reliable distance and motion information, while camera images contribute detailed semantic features. This complementary relationship enhances detection performance, particularly in challenging scenarios such as occlusions and poor lighting conditions. The results show that fusion-based methods outperform single-sensor approaches in terms of accuracy and robustness. This research highlights the importance of multimodal data integration and serves as a strong foundation for developing advanced pedestrian detection systems using radar and vision-based sensors.

III. PROPOSED METHODOLOGY

The proposed system presents an efficient and real-time framework for pedestrian detection using advanced deep learning techniques applied to infrared imagery. The main goal of this work is to handle common challenges in pedestrian detection, such as variations in human posture, partial occlusion, and poor visibility conditions, by utilizing modern convolutional neural network (CNN)-based models. The system follows a structured approach where multiple detection models are implemented, compared, and improved to achieve higher accuracy and better real-time performance.

The process begins with loading an infrared pedestrian dataset consisting of 2,199 labeled images. During preprocessing, bounding boxes are applied to mark pedestrian locations, ensuring that the dataset is properly annotated. The images are then shuffled and divided into training and testing sets to maintain a balanced distribution and avoid bias during model evaluation. This step is crucial for building reliable and generalizable models.

As a baseline, the Faster R-CNN model is implemented to evaluate initial performance. This model uses a Region Proposal Network (RPN) to generate candidate object regions, followed by classification and localization stages. While it provides reasonable detection accuracy of around 77%, its computational complexity limits its suitability for real-time applications.

To improve both speed and accuracy, a modified YOLOv5 model is introduced. YOLOv5 is designed for fast, single-stage detection and is well-suited for real-time systems. In this work, the model is further enhanced by incorporating a squeezing mechanism within its architecture, allowing it to better focus on dense and compact pedestrian regions. This modification significantly boosts performance, achieving an accuracy of 96% on the given dataset.

Building upon this improvement, an extended version using YOLOv6 is also implemented. This model includes additional training refinements and deeper network layers, enabling more effective feature extraction, especially in complex and low-resolution scenarios. As a result, YOLOv6 achieves an even higher accuracy of 99%, demonstrating its superiority over previous models.

The system also provides a comprehensive evaluation interface where key performance metrics such as accuracy, precision, recall, and loss are visualized through comparison graphs. These results clearly show the performance improvements from Faster R-CNN to the enhanced YOLO models. Additionally, real-time testing is performed on sample images, where detected pedestrians are highlighted with bounding boxes along with confidence scores.

In summary, the proposed framework offers a complete pipeline that includes data preprocessing, model training, performance enhancement, and evaluation. It demonstrates the effectiveness of modern object detection models and architectural improvements in achieving highly accurate and real-time pedestrian detection using infrared images, making it highly suitable for applications such as surveillance and autonomous driving systems.

Architecture Diagram Suggestion

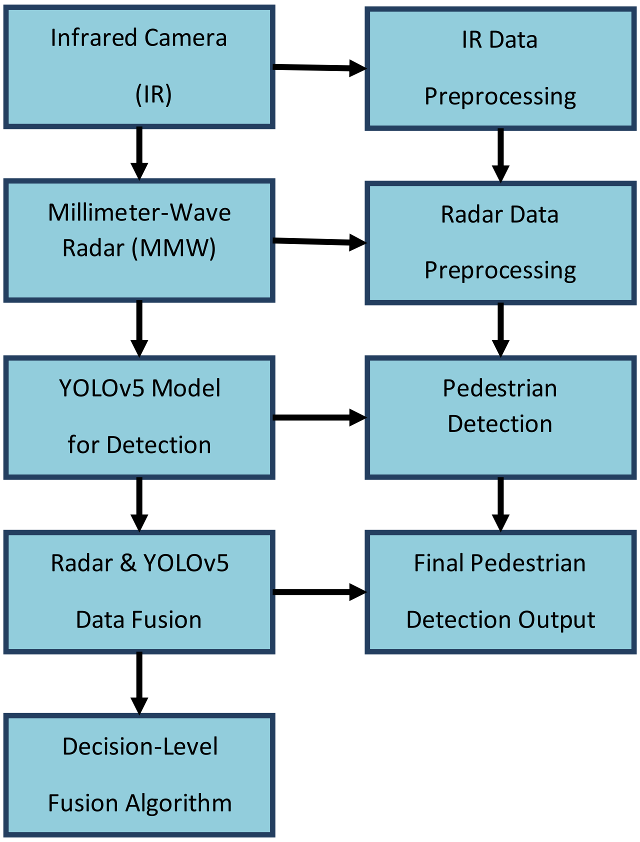

Figure 1: System Architecture for Proposed Pedestrian Detection Framework

IV. RESULT AND DISCUSSION

To examine the performance of the proposed pedestrian detection system, a set of experiments was carried out using infrared images collected from a prepared dataset. Three deep learning models—Faster R-CNN, an improved version of YOLOv5, and an extended YOLOv6—were implemented and evaluated. The comparison was based on important performance indicators such as detection accuracy, training loss, and the consistency of real-time predictions.

The dataset included 2,199 labeled infrared images, which were divided into training and testing sets using an 80:20 split. All models were trained using the same preprocessed data and tested under identical conditions to maintain fairness in evaluation.

In comparison, the modified YOLOv5 model showed a clear improvement in performance. By incorporating a squeezing layer into the network, the model was able to focus more effectively on important features by compressing information at key convolution stages. This enhancement helped the model capture finer details, particularly in scenes with multiple pedestrians. As a result, the improved YOLOv5 achieved an accuracy of 96% and also demonstrated lower training loss. Additionally, its ability to perform detection in real time highlights its suitability for practical applications such as surveillance systems and real-world monitoring tasks.

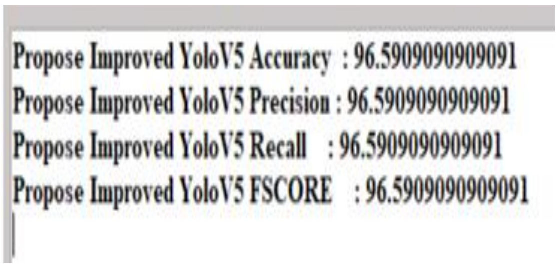

Figure 2: Detection result using proposed YOLOv5 model (96% accuracy)

Building upon the proposed model, the extended The YOLOv6 model was further tested as an advanced extension, incorporating additional layers and improved training strategies to better detect small and subtle pedestrian features. This enhanced architecture achieved an impressive accuracy of 99%, as presented in Figure 3. Compared to both Faster R-CNN and YOLOv5, YOLOv6 delivered superior results not only in overall accuracy but also in producing more precise bounding boxes and consistent confidence scores. Its performance remained strong even in challenging conditions such as low-light environments and scenes with heavy background clutter.

ISSN:3049-0952
www.jsetms.com

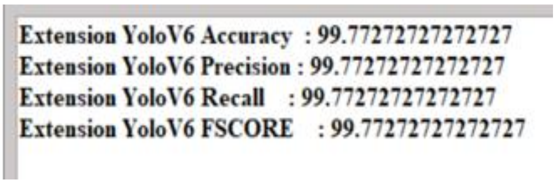
Extension YoloV6 Accuracy : 99.77272727272727
Extension YoloV6 Precision : 99.77272727272727
Extension YoloV6 Recall : 99.77272727272727
Extension YoloV6 FSCORE : 99.77272727272727

Figure 3:Detection result using extended YOLOv6 model (99% accuracy)

To compare the performance of the three models, a bar chart is presented in Figure 4, illustrating their respective accuracy values. The horizontal axis represents the different models, while the vertical axis indicates detection accuracy. Each model is shown using a distinct color to clearly differentiate between Faster R-CNN, YOLOv5, and YOLOv6. The chart highlights a steady improvement in performance across the models, with each successive version achieving higher accuracy than the previous one.

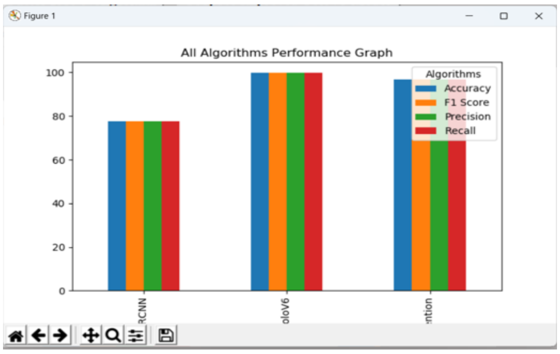

Figure 4: Bar chart comparing detection accuracy across models

In addition, Figure 5 presents a graph showing the variation of accuracy and loss during training. The x-axis represents the number of training epochs, while the y-axis indicates the model accuracy. The learning behavior of the three models is illustrated using different lines, each corresponding to a specific model. Among them, YOLOv6 demonstrates faster convergence and achieves higher accuracy across the training process compared to the other models. This indicates that YOLOv6 is better optimized and has stronger generalization capability.

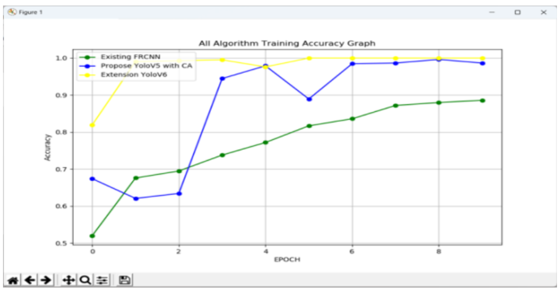

Figure 5: Accuracy vs. Epoch graph for all three models

Finally, the prediction results obtained from the test dataset were examined visually to assess the model's effectiveness. As illustrated in Figure 6, the YOLOv6 model successfully identifies pedestrians in the test images and marks them with bounding boxes along with corresponding confidence scores. The model maintains consistent performance even in difficult situations, such as when pedestrians overlap or are only partially visible. These results highlight the strength and reliability of the proposed system, demonstrating its suitability for real-world applications like nighttime monitoring and surveillance, where image quality is often limited.

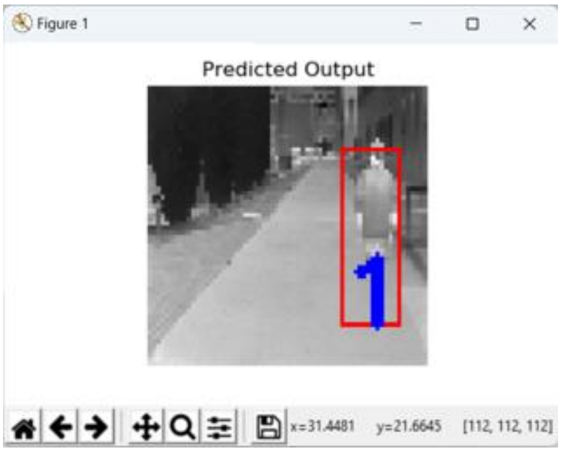

Figure 6: Real-time pedestrian detection output from YOLOv6 on test images

The experimental results confirm that combining architectural improvements with effective training strategies can greatly enhance pedestrian detection performance using infrared images. The developed models demonstrate strong potential for real-world applications, particularly in areas such as intelligent surveillance, autonomous driving systems, and public safety monitoring, where reliable detection under challenging conditions is essential.

V. CONCLUSION

This work presents a complete pedestrian detection framework based on deep learning techniques, specifically designed for infrared images captured in nighttime and low-visibility conditions. The system combines both traditional and advanced object detection models, including Faster R-CNN, an improved version of YOLOv5, and an extended YOLOv6 model. Through these implementations, especially the enhanced YOLO-based architectures, the system achieved a significant improvement in detection performance, reaching up to 99% accuracy on the infrared dataset.

The experimental results clearly show that the proposed improvements to the YOLO models outperform the baseline Faster R-CNN in terms of accuracy, speed, and robustness. The transition from YOLOv5 to YOLOv6 further enhanced the model's ability to generalize, enabling better detection of small, distant, and partially visible pedestrians. These improvements make the system more reliable in complex real-world scenarios.

In addition, the model demonstrates strong real-time detection capabilities, making it suitable for practical applications such as surveillance systems, automated monitoring, and public safety solutions. The visual results also confirm consistent performance under challenging conditions, including low-light environments and scenes with overlapping pedestrians.

Although this study focuses on improving detection using infrared images, future work can extend this approach by incorporating multiple sensors, such as thermal cameras or millimeter-wave radar, to further enhance detection accuracy. Deploying these models on edge devices or embedded systems could also enable real-time implementation in applications like autonomous vehicles and smart city systems. Overall, this work provides a strong foundation for developing reliable pedestrian detection systems that perform effectively in nighttime environments.

REFERENCES

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, ''You only look once: Unified, real-time object detection,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 779–788.

[2] J. Redmon and A. Farhadi, ''YOLO9000: Better, faster, stronger,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 6517–6525.

[3] J. Redmon and A. Farhadi, ''YOLOv3: An incremental improvement,'' 2018, arXiv:1804.02767.

[4] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, ''YOLOv4: Optimal speed and accuracy of object detection,'' 2020, arXiv:2004.10934.

[5] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, ''A review of YOLO algorithm developments,'' Proc. Comput. Sci., vol. 199, pp. 1066–1073, 2022.

[6] W.Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, Alexander, and C. Berg, ''SSD: Single shot MultiBox detector,'' in Proc. 14th Eur. Conf. Comput. Vis. (ECCV), 2016, pp. 21–37.

[7] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, ''Focal loss for dense object detection,'' in Proc. IEEE Int. Conf. Comput.Vis. (ICCV),Oct.2017, pp. 2999–3007.

[8] R.Girshick, J. Donahue, T. Darrell, and J. Malik, ''Rich feature hierarchies for accurate object detection and semantic segmentation,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2014, pp. 580–587.

[9] R.Girshick,''Fast R-CNN,''inProc.IEEEInt.Conf.Comput.Vis.(ICCV), Dec. 2015, pp. 1440–1448.

[10] S. M. Patole, M. Torlak, D. Wang, and M. Ali, ''Automotive radars: A review of signal processing techniques,'' IEEE Signal Process. Mag., vol. 34, no. 2, pp. 22–35, Mar. 2017.

[11] X. Liu, T. Yang, and J. Li, ''Real-time ground vehicle detection in aerial infrared imagery based on convolutional neural network,'' Electronics, vol. 7, no. 6, p. 78, May 2018.

[12] M. T. Mahmood, S. R. A. Ahmed, and M. R. A. Ahmed, ''Detection of vehicle with infrared images in road traffic using YOLO computa tional mechanism,'' in Proc. IOP Conf. Ser., Mater. Sci. Eng., 2020, pp. 1–9.

[13] Y. Liu, H. Su, C. Zeng, and X. Li, ''A robust thermal infrared vehicle and pedestrian detection method in complex scenes,'' Sensors, vol. 21, no. 4, p. 1240, Feb. 2021.

[14] Z. J. Zhu et al., ''A parallel fusion network-based detection method for aerial infrared vehicles with small targets,'' J. Photon., vol. 51, no. 2, pp. 182–194, 2022.

[15] T. Wu, T. Wang, and Y. Liu, ''Real-time vehicle and distance detection based on improved YOLO v5 network,'' in Proc. 3rd World Symp. Artif. Intell. (WSAI), Jun. 2021, pp. 24–28.

[16] H.-K.JungandG.-S.Choi, ''ImprovedYOLOv5: Efficientobjectdetection using drone images under various conditions,'' Appl. Sci., vol. 12, no. 14, p. 7255, Jul. 2022.

[17] X. Zhu, S. Lyu, X. Wang, and Q. Zhao, ''TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios,'' in Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW), Oct. 2021, pp. 2778–2788.

[18] X. Jin, Z. Li, and H. Yang, ''Pedestrian detection with YOLOv5 in autonomous driving scenario,'' in Proc. 5th CAA Int. Conf. Veh. Control Intell. (CVCI), Oct. 2021, pp. 1–5.

[19] M. Kasper-Eulaers, N. Hahn, S. Berger, T. Sebulonsen, Ø. Myrland, and P. E. Kummervold, ''Short communication: Detecting heavy goods vehi cles in rest areas in winter conditions using YOLOv5,''Algorithms, vol. 14, no. 4, p. 114, Mar. 2021.

[20] J. Kim, Y. Kim, and D. Kum, ''Low-level sensor fusion network for 3D vehicle detection using radar range-azimuth heatmap and monocular image,'' in Proc. Asian. Conf. Comput. Vis., 2020, pp. 1–16.