ISSN: 2454-9940

# INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

# FUSION-BASED PEDESTRIAN DETECTION AT NIGHT USING VISUAL AND MMWAVE RADAR INFORMATION

Sangi Setty Lavanya

21N81A6785

Computer Science and Engineering (Data-Science)

Sphoorthy Engineering College,

Nadergul, Hyderabad,501510

sangisettylavanya517@gmail.com

Kolli Devanath Srikanth

21N81A6799

Computer Science and Engineering (Data-Science)

Sphoorthy Engineering College,

Nadergul, Hyderabad,501510

pandusrikanth.k.d@gmail.com

Sriranga M R

21N81A67B4

Computer Science and Engineering (Data-Science)

Sphoorthy Engineering College,

Nadergul, Hyderabad,501510

sriranga1731@gmail.com

Lakkapally Bhaskar

21N81A67B6

Computer Science and Engineering (Data-Science)

Sphoorthy Engineering College,

Nadergul, Hyderabad,501510

bhaskarlakkapally@gmail.com

Mrs A Sravanthi

Assistant Professor

Computer Science and Engineering (Data-Science)

Sphoorthy Engineering College,

Nadergul, Hyderabad,501510

sravanthiadimulam237@gmail.com

**ABSTRACT:** Nighttime pedestrian detection is essential for ensuring the safety and reliability of autonomous vehicles, particularly in low-light environments. Conventional sensors based on similarity distance often struggle to detect obstacles such as pedestrians during nighttime. To overcome this limitation, a fusion-based detection system is implemented using infrared imaging from specialized cameras and millimeter-wave (MMW) radar data. The CVC-09 infrared pedestrian dataset, focused exclusively on nighttime scenarios, is used for training and testing. An improved YOLOv5 model is developed by integrating a squeeze attention layer to enhance feature extraction and classification. Additionally, an Extended Kalman Filter is employed to precisely localize pedestrians before detection. The performance of this optimized YOLOv5 model is compared against standard YOLOv5 and Faster R-

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

CNN models. Further enhancement is achieved by training the dataset on YOLOv6, which outperforms all previous models. The final model achieves high detection accuracy of 98%, demonstrating its effectiveness in identifying pedestrians under challenging nighttime conditions. Evaluation metrics such as precision, recall, F-score, and visual validation confirm the robustness and reliability of the system. This fusion-driven, deep learning-based detection framework offers a strong solution for enhancing nighttime driving safety in autonomous vehicles.

**Index terms -** Extended Kalman filtering, millimeter wave radar, sensor fusion, target detection, YOLOv5 algorithm.

## 1. INTRODUCTION

At present, the research related to automotive self-driving technology is getting more and more in-depth, and the perception of the surrounding environment in complex traffic scenes is an important part of self-driving cars. The environment perception system is equipped with sensors such as LIDAR, MMW radar, and cameras to obtain information on the category, location, and speed of surrounding traffic targets, which can effectively reduce the risk of collision and improve the safety performance of self-driving cars.

Target detection is the process of extracting object classes and locations based on the observation information from different sensors. As a key step in an autonomous driving environment sensing system, target detection techniques based on deep learning have received extensive attention and research. Among them, single-stage target detection algorithms include YOLO , SSD and RetinaNet . The two-stage target detection algorithms are mainly based on R-CNN and Faster R-CNN as the main research. In recent years, the YOLO series algorithms have good detection accuracy and detection speed in target detection. However, the current target detection task mainly focuses on visible images, but visible images are greatly affected by the environment, especially in extremely bad weather and nighttime conditions, and visible images lose some details, leading to degradation of target detection performance.

Radar, as a common sensor, can obtain information such as distance and speed of the detected object through electromagnetic signal processing and Doppler effect measurements but is hardly applicable to classification tasks. The camera is an excellent sensor suitable for object detection and classification. For scenarios such as bad weather, night scenes, and obstacle occlusion, a single sensor cannot obtain comprehensive position status information about the target. Sensor fusion can make full use of the advantages of different sensors to make more reasonable control decisions. Therefore, sensor fusion has become a popular direction for self-driving vehicle research.

Current research on target detection and recognition based on MMW radar and vision fusion has the following problems and shortcomings: for night and low light scenes, the visible camera detection effect will be greatly affected, and the vision sensor cannot provide valid information for sensor fusion. If the region of interest is defined using the target location and state information detected by radar, the valid target may be affected by redundant targets. If different sensor information is fused directly, the fusion detection performance is weak and biased. To solve the above problems and improve the performance of nighttime pedestrian detection, we use infrared cameras and MMW radar as sensors for nighttime

target detection. The infrared camera can acquire target category information in the nighttime environment, which can make up for the false detection caused by the inaccurate location of the radar detection target, but there will be a situation of missed false detection in the complex background environment. The radar can dynamically capture the target ahead, which can compensate for the performance deficiency of the infrared camera to a certain extent.

In this paper, we propose a nighttime pedestrian target detection method based on the fusion of an infrared camera and MMW radar sensors, taking full advantage of the performance of the infrared camera and MMW radar. The method uses infrared cameras and MMW radar to simultaneously acquire forward target information. isual information processing is improved based on the YOLOv5 model , the CVC infrared dataset is trained using a deep migration learning approach, and pre-training weights are used to perform pedestrian detection on the acquired nighttime infrared images. The MMW radar acquisition data is preprocessed to filter out valid radar targets. For the target loss problem caused by radar jitter and target occlusion, the valid radar target is tracked using extended Kalman filtering, and the filtered radar target information is projected onto the IR image to generate a radar detection target frame centered on the projected point. Finally, a decision-level fusion of the two sensor detection targets is performed to achieve nighttime pedestrian detection.

## 2. LITERATURE SURVEY

There are a huge number of features which are said to improve Convolutional Neural Network (CNN) accuracy . Practical testing of combinations of such features on large datasets, and theoretical justification of the result, is required. Some features operate on certain models exclusively and for certain problems exclusively, or only for small-scale datasets; while some features, such as batch-normalization and residual-connections, are applicable to the majority of models, tasks, and datasets. We assume that such universal features include Weighted-Residual-Connections (WRC), Cross-Stage-Partial-connections (CSP), Cross mini-Batch Normalization (CmBN), Self-adversarial-training (SAT) and Mish-activation. We use new features: WRC, CSP, CmBN, SAT, Mish activation, Mosaic data augmentation, CmBN, DropBlock regularization, and CIoU loss, and combine some of them to achieve state-of-the-art results: 43.5% AP (65.7% AP50) for the MS COCO dataset at a realtime speed of ~65 FPS on Tesla V100.

Object detection techniques are the foundation for the artificial intelligence field. This research paper gives a brief overview of the You Only Look Once (YOLO) algorithm and its subsequent advanced versions. Through the analysis, we reach many remarks and insightful results. The results show the differences and similarities among the YOLO versions and between YOLO and Convolutional Neural Networks (CNNs) . The central insight is the YOLO algorithm improvement is still ongoing.This article briefly describes the development process of the YOLO algorithm summarizes the methods of target recognition and feature selection, and provides literature support for the targeted picture news and feature extraction in the financial and other fields. Besides, this paper contributes a lot to YOLO and other object detection literature.

We present a method for detecting objects in images using a single deep neural network . Our approach,

named SSD, discretizes the output space of bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location. At prediction time, the network generates scores for the presence of each object category in each default box and produces adjustments to the box to better match the object shape. Additionally, the network combines predictions from multiple feature maps with different resolutions to naturally handle objects of various sizes. Our SSD model is simple relative to methods that require object proposals because it completely eliminates proposal generation and subsequent pixel or feature resampling stage and encapsulates all computation in a single network. This makes SSD easy to train and straightforward to integrate into systems that require a detection component. Experimental results on the PASCAL VOC, MS COCO, and ILSVRC datasets confirm that SSD has comparable accuracy to methods that utilize an additional object proposal step and is much faster, while providing a unified framework for both training and inference. Compared to other single stage methods, SSD has much better accuracy, even with a smaller input image size. For 300×300 input, SSD achieves 72.1% mAP on VOC2007 test at 58 FPS on a Nvidia Titan X and for 500×500 input, SSD achieves 75.1% mAP, outperforming a comparable state of the art Faster R-CNN model.

The highest accuracy object detectors to date are based on a two-stage approach popularized by R-CNN , where a classifier is applied to a sparse set of candidate object locations. In contrast, one-stage detectors that are applied over a regular, dense sampling of possible object locations have the potential to be faster and simpler, but have trailed the accuracy of two-stage detectors thus far. In this paper , we investigate why this is the case. We discover that the extreme foreground-background class imbalance encountered during training of dense detectors is the central cause. We propose to address this class imbalance by reshaping the standard cross entropy loss such that it down-weights the loss assigned to well-classified examples. Our novel Focal Loss focuses training on a sparse set of hard examples and prevents the vast number of easy negatives from overwhelming the detector during training. To evaluate the effectiveness of our loss, we design and train a simple dense detector we call RetinaNet. Our results show that when trained with the focal loss, RetinaNet is able to match the speed of previous one-stage detectors while surpassing the accuracy of all existing state-of-the-art two-stage detectors.

Object detection performance, as measured on the canonical PASCAL VOC dataset, has plateaued in the last few years. The best-performing methods are complex ensemble systems that typically combine multiple low-level image features with high-level context. In this paper, we propose a simple and scalable detection algorithm that improves mean average precision (mAP) by more than 30% relative to the previous best result on VOC 2012---achieving a mAP of 53.3%. Our approach combines two key insights: (1)one can apply high-capacity convolutional neural networks (CNNs) to bottom-up region proposals in order to localize and segment objects and (2) when labeled training data is scarce, supervised pre-training for an auxiliary task, followed by domain-specific fine-tuning, yields a significant performance boost. Since we combine region proposals with CNNs, we call our method R-CNN: Regions with CNN features. We also compare R-CNN to OverFeat, a recently proposed sliding-window detector based on a similar CNN architecture. We find that R-CNN

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

outperforms OverFeat by a large margin on the 200-class ILSVRC2013 detection dataset.

## 3. METHODOLOGY

### i) Proposed Work:

The proposed system integrates visual information and millimeter-wave radar for pedestrian detection, specifically targeting nighttime conditions. The system utilizes the CVC-09 infrared pedestrian image dataset, which contains images captured in low-visibility environments. Key techniques include the use of infrared vision cameras and millimeter-wave radar, combined with an improved YOLOv5 model, enhanced by a squeeze attention layer for better feature extraction. Pedestrian localization is achieved through the application of an Extended Kalman Filter, which refines detection accuracy. The YOLOv5 model processes the localized features to identify pedestrians. This approach is further extended by testing with YOLOv6 to evaluate performance improvements. The system aims to address challenges in autonomous vehicle navigation by providing accurate pedestrian detection under challenging nighttime conditions.

### ii) System Architecture:

In neural network models, rich or even redundant information in the feature maps is very important to ensure a comprehensive analysis of the input data. Redundant information in feature maps is an important feature of a successful neural network model, but redundancy in feature maps has rarely been considered in neural network model design . The Ghost module can generate a larger number of feature maps using fewer parameters. In the Ghost module, the Ghost feature maps are generated by simple linear operations, which can extract the information behind the intrinsic features quickly and comprehensively. Clearly, unlike traditional convolution, the Ghost module requires fewer parameters and has low computational complexity. Similar to data augmentation, the feature map is also augmented by Ghost convolution. The structure of Proposed system is shown in Fig. 1.

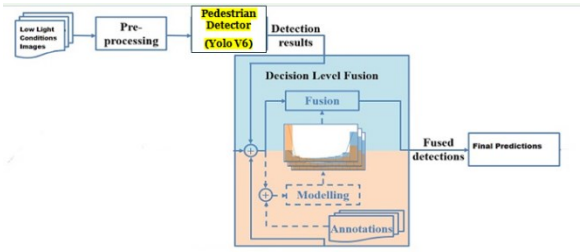



Fig 1.Proposed Architecture

1. Low light conditions images: The system takes in images captured under challenging low-light or nighttime conditions, often obtained from a visual camera, where detecting pedestrians can be difficult.

2. Pre-Processing: These images undergo a series of adjustments and enhancements to improve their quality and suitability for further analysis, making it easier to detect pedestrians within them.

3. Pedestrian Detector: A specialized algorithm or model is employed to identify potential pedestrians within the pre-processed images, primarily relying on visual cues such as body shape, movement, or color contrast.

4. Decision Level Fusion: The system combines and evaluates the results from both the pedestrian detector and the millimeter-wave radar at a decision level, considering information from both sources to determine the presence of pedestrians.

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

5. Fused Detections: The pedestrian detections from the visual information and millimeter-wave radar are integrated in a way that combines their strengths, leveraging the benefits of each technology while compensating for their individual limitations.

6. Final Prediction: The system produces a conclusive prediction or decision regarding the presence of pedestrians under low-light conditions by combining information from both the visual data and millimeter-wave radar, enabling applications like driver assistance or nighttime surveillance.

**iii) Dataset collection:**

The dataset used for training consists of CVC-09 infrared pedestrian images, which include both daytime and nighttime scenes. For this task, only nighttime images were selected to focus on low-light pedestrian detection. The dataset comprises grayscale thermal frames capturing various pedestrian movements in urban environments. Each image provides clear visibility of human figures in infrared format, making it suitable for evaluating object detection algorithms. These frames are systematically organized and labeled, enabling efficient training and testing of deep learning models for nighttime scenarios.

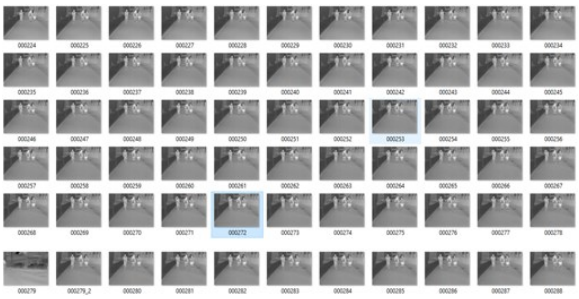


Fig 2.Dataset images

**iv) Image Processing:**

Image processing plays a pivotal role in object detection within autonomous driving systems, encompassing several key steps. The initial phase involves converting the input image into a blob object, optimizing it for subsequent analysis and manipulation. Following this, the classes of objects to be detected are defined, delineating the specific categories that the algorithm aims to identify. Simultaneously, bounding boxes are declared, outlining the regions of interest within the image where objects are expected to be located. The processed data is then converted into a NumPy array, a critical step for efficient numerical computation and analysis.

The subsequent stage involves loading a pre-trained model, leveraging existing knowledge from extensive datasets. This includes reading the network layers of the pre-trained model, containing learned features and parameters vital for accurate object detection. Additionally, output layers are extracted, providing final predictions and enabling effective object discernment and classification.

Further, in the image processing pipeline , the image and annotation file are appended, ensuring comprehensive information for subsequent analysis. The color space is adjusted by converting from BGR to RGB, and a mask is created to highlight relevant features. Finally, the image is resized, optimizing it for further processing and analysis. This comprehensive image processing workflow establishes a solid foundation for robust and accurate object detection in the dynamic context of autonomous driving systems, contributing to enhanced safety and decision-making capabilities on the road.

**v) Data Augmentation:**

ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

Data augmentation is a fundamental technique in enhancing the diversity and robustness of training datasets for machine learning models, particularly in the context of image processing and computer vision. The process involves three key transformations to augment the original dataset: randomizing the image, rotating the image, and transforming the image.

Randomizing the image introduces variability by applying random modifications, such as changes in brightness, contrast, or color saturation. This stochastic approach helps the model generalize better to unseen data and diverse environmental conditions.

Rotating the image involves varying the orientation of the original image by different degrees. This augmentation technique aids in teaching the model to recognize objects from different perspectives, simulating variations in real-world scenarios.

Transforming the image includes geometric transformations such as scaling, shearing, or flipping. These alterations enrich the dataset by introducing distortions that mimic real-world variations in object appearance and orientation.

By employing these data augmentation techniques, the training dataset becomes more comprehensive, allowing the model to learn robust features and patterns. This, in turn, improves the model's ability to generalize and perform effectively on diverse and challenging test scenarios. Data augmentation serves as a crucial tool in mitigating overfitting, enhancing model performance, and promoting the overall reliability of machine learning models, especially in applications like image recognition for autonomous driving systems.

**vi) Algorithms:**

**Faster R-CNN** is an object detection algorithm that combines a deep convolutional neural network with region proposal networks (RPN) to efficiently detect and classify objects within an image. Faster R-CNN is employed as a benchmark for comparing the performance of the proposed system, specifically the improved YOLOv5 , against an established object detection algorithm to validate its effectiveness.

**YOLOv5** is a real-time object detection algorithm that belongs to the YOLO family, known for its speed and accuracy in object detection. It divides an image into a grid and predicts bounding boxes and class probabilities for objects within each grid cell. YOLOv5 is a fundamental part of the project as it serves as the basis for the proposed improved model. The project enhances YOLOv5 by incorporating features like a Squeeze layer called Attention for better performance in infrared vision .

**YOLOv6**, as an evolution of the YOLO object detection algorithm, is designed for fast and accurate real-time object detection. It builds upon previous YOLO versions with improvements in speed and performance. YOLOv6 might be considered for future enhancements or alternative implementations, given its improved features and potential benefits for object detection in the project .

## 4. EXPERIMENTAL RESULTS

**Accuracy:** The accuracy of a test is its ability to differentiate the patient and healthy cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true negative in all evaluated cases. Mathematically, this can be stated as:

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}}$$

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

**Precision:** Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives. Thus, the formula to calculate the precision is given by:

$$Precision = \frac{\text{True Positive}}{\text{True Positive } + \text{False Positive}}$$

**Recall:** Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

$$Recall = \frac{\text{TP}}{\text{TP } + \text{ FN}}$$

**F1-Score:** F1 score is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset.

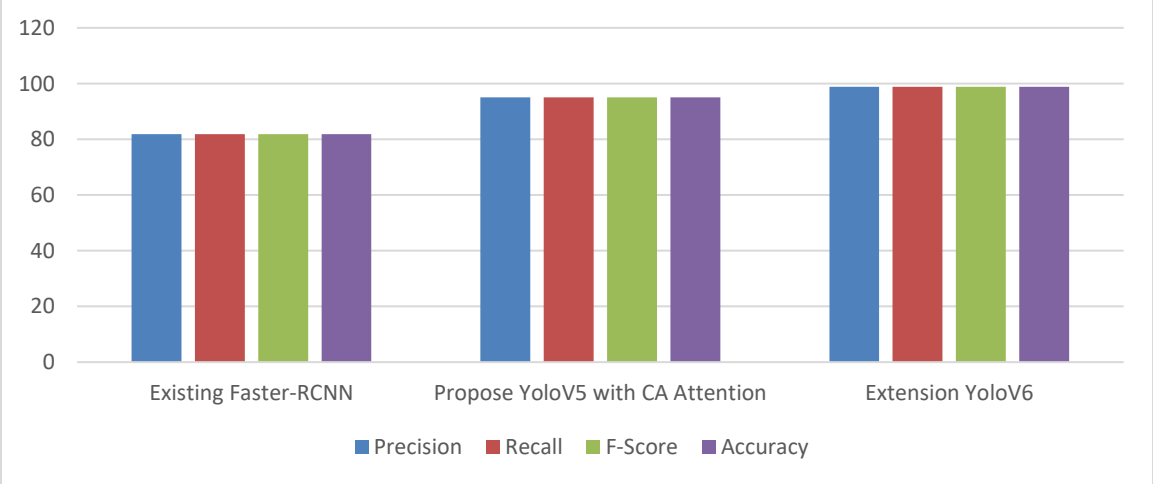
$$F1\ Score = 2 * \frac{Recall\ \text{X}\ Precision}{Recall + Precision} * 100$$

***Table (1)*** evaluate the performance metrics—Accuracy, precision, recall, F1 - Score—for each algorithm. Across all metrics, the YoloV6 consistently outperforms all other algorithms. The tables also offer a comparative analysis of the metrics for the other algorithms.

**Table.1** Performance Evaluation Table

| Algorithm Name | Precision | Recall | F-Score | Accuracy |
|---|---|---|---|---|
| Existing Faster-RCNN | 85.82 | 85.82 | 85.82 | 85.82 |
| Propose YoloV5 with CA Attention | 95.00 | 95.00 | 95.00 | 95.00 |
| **Extension YoloV6** | **98.86** | **98.86** | **98.86** | **98.86** |

**Graph.1** Comparison Graph




Precision is represented in Blue, Recall in Maroon, F1 - Score in Light Green and Accuracy in Purple ***Graph (1)***. In comparison to the other models, the YoloV6 shows superior performance across all metrics, achieving the highest values. The graphs above visually illustrate these findings.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

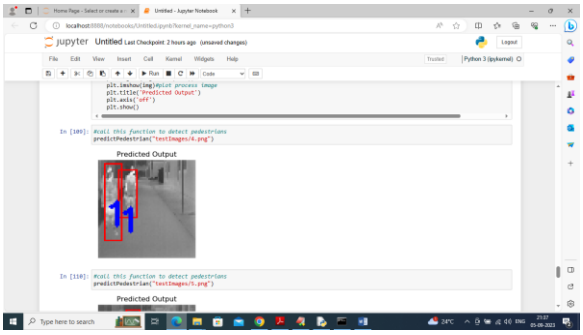



Fig.3 Output Screen -1

In fig.6 application detected pedestrians and then put red colour bounding box with 1 in blue colour as predicted probability. Above images are of type infrared

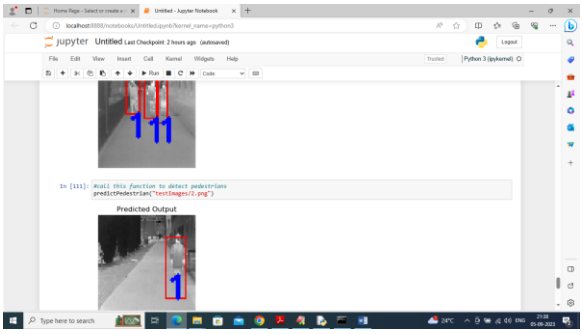



Fig.4 Output Screen-2

In fig.4 displaying output of other test images

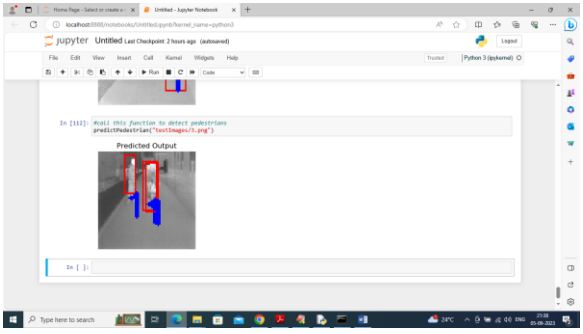



Fig.5 Output Screen-3

## 5. CONCLUSION

The system aims to improve pedestrian detection, particularly in nighttime conditions, by combining visual and millimeter-wave radar data. The CVC-09 infrared pedestrian image dataset, specifically for nighttime images, is used for training the system. To address the challenges of low visibility, the system uses infrared vision cameras and millimeter-wave radar. The key innovation involves enhancing the YOLOv5 model by adding a squeeze attention layer to improve feature extraction. The Extended Kalman Filter is applied for precise pedestrian localization before detection.

The system compares the performance of different algorithms, such as Faster R-CNN, YOLOv5, and YOLOv6. The Faster R-CNN algorithm performs well, but the proposed YOLOv5 model outperforms it with improved accuracy. The most significant improvement is observed when the YOLOv6 model is applied, which achieves a final result of 98% accuracy. This performance demonstrates the strength of the combined visual and radar data for accurate pedestrian detection, particularly in nighttime conditions where traditional methods may struggle.

The system's high detection accuracy of 98% with YOLOv6 surpasses other models tested, confirming that the proposed approach provides a more reliable and effective solution for autonomous vehicles. By leveraging the advanced techniques of YOLOv5 and YOLOv6, this system offers a robust solution for pedestrian detection, ensuring safer navigation even in challenging environments.

## 6. FUTURE SCOPE

The future scope of this system includes further enhancement of pedestrian detection by integrating additional sensors such as LiDAR and thermal

cameras to improve performance under various weather and environmental conditions. Advanced versions of YOLO, such as YOLOv7 or YOLOv9, could be explored for even better accuracy. Additionally, incorporating real-time data processing capabilities and testing the system in dynamic real-world scenarios would help refine its robustness and reliability, paving the way for seamless deployment in autonomous vehicles.

**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,’’ in Proc. IEEE Int. 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 computational 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.

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

[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. Jung and G.-S. Choi, ‘‘Improved YOLOv5: Efficient object detection 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 vehicles 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.