Improving Pedestrian Detection in Low-Visibility Conditions: Fusing Visual and Infrared Data with Deep Learning

M.Praveen[1], M.Sarika [2], N.Anjani [2], M.Yamini[2]

[1]Assistant Professor, [2]UG Student,

[1,2] School of computer science and engineering, Malla Reddy Engineering College for Women (UGC-Autonomous), Maisammguda, Hyderabad, Telangana

Abstract

With the increasing demand for autonomous vehicles and higher safety standards, developing accurate pedestrian detection systems that perform well in all environmental conditions, especially at night, has become critical. Traditional sensor-based systems, such as LIDAR and radar, are often inadequate in low-visibility environments, prompting the need for AI-based solutions. This research proposes a pedestrian detection system that integrates infrared vision and millimeter-wave (MMW) radar data with an enhanced deep learning model. By utilizing an improved version of YoloV5, equipped with a Squeeze layer for attention, the system effectively extracts and classifies image features. Additionally, an Extended Kalman Filter is employed for accurate pedestrian localization. The fusion of these modalities into the enhanced YoloV5 model significantly improves detection accuracy and robustness, making it more effective in real-time pedestrian detection under challenging conditions.

Keywords: Autonomous vehicles, pedestrian detection, infrared vision, millimeter-wave radar, YoloV5, deep learning, Squeeze layer, Extended Kalman Filter, multi-modal data,

1. Introduction

Sensors find their most common use in a wide variety of applications, ranging from computerized driving to the monitoring of physiological characteristics. Additionally, sensors play a significant part in the performance of tasks connected to detection and vision in all of the contemporary applications of science, engineering, and technology that are dominated by computer vision. The Internet of Things (IoT) is an intriguing new field that makes use of smart sensors. It is an area that deals with wireless networks and sensors that are dispersed in order to detect data in real time and produce particular results of interest via appropriate processing. Sensors and artificial intelligence (AI) are the most significant components of Internet of Things (IoT) devices, which are responsible for the level of intelligence and sensitivity that these devices possess. In point of fact, as a result of the role that artificial intelligence plays, the sensors function as smart sensors and find an efficient usage for a variety of applications. These applications include general environmental monitoring [1], monitoring a specific number of environmental factors, weather forecasting, satellite imaging and its use, remote sensing-based applications, monitoring of hazard events such as landslide detection, self-driving cars, healthcare, and so on. Regarding the latter sector, there has been a significant growth in the use of smart devices in hospitals and diagnostic centers in recent times. These devices are utilized for the purpose of assessing and monitoring a variety of health problems of patients who are afflicted, both remotely and physically [2]. In this day and age, there is not a single scientific or study discipline that can function effectively without the use of contemporary sensors. The applications are considered intelligent because to the widespread use of sensors and the need of using the internet of things

in remote sensing, monitoring the environment, and monitoring human health. In the last ten years, agricultural applications have also included [3] the exploitation of several kinds of sensors for the purpose of monitoring and managing a wide range of environmental factors. These parameters include temperature, humidity, soil quality, pollution, air quality, water contamination, radiation, and many more. Additionally, the purpose of this study is to emphasize the usage of sensors and the internet of things for applications in agriculture and remote sensing for the purpose of conducting an extended debate and evaluation.

The identification of pedestrians has been an important topic of study, especially for the purpose of improving road safety and providing assistance to autonomous cars. The primary purpose of this project is to enhance pedestrian recognition, especially at nighttime or in settings with limited visibility. This will be accomplished by combining visual and infrared data, as well as by improving detection accuracy using deep learning models such as YoloV5. Using sophisticated sensors and deep learning algorithms, the suggested technique intends to improve accuracy while simultaneously lowering the number of mistakes that occur during the process of recognizing pedestrians in real time. The phrase "Fusion of Visual and Infrared Information for Nightmare Pedestrian Detection" refers to the process of merging visual (camera-based) data with infrared (heat-sensing) data in order to identify pedestrians, especially in difficult settings like as driving at night. The phrase "nightmare" is a metaphor that emphasizes how difficult it is to identify pedestrians in situations when visibility is limited or when circumstances are harsh. Detecting impediments has traditionally been accomplished via the use of sensors such as LIDAR and radar. Traditional pedestrian detection depended mainly on LIDAR or radar-based systems for obstacle identification, optical cameras combined with rudimentary image processing algorithms, proximity sensors, and simple motion detectors for identifying pedestrians. This was done before the advent of AI-based solutions. Traditional sensor-based systems continue to face a substantial obstacle when it comes to detecting pedestrians in situations with limited visibility and low levels of background light. It is difficult for these systems to discern between pedestrians and objects effectively, especially when it is nighttime or when the weather is bad. This might result in delayed or missing detections, which can lead to accidents.

In light of the rising need for greater safety standards and the growing drive for autonomous cars, it is becoming more important to create pedestrian detection systems that are more precise and that are able to function in all environmental situations, particularly at night. Current sensor-based systems are insufficient on their own, which is why there is a need for artificial intelligence-based solutions that are able to make use of multi-modal data such as infrared and visual input in order to achieve higher detection accuracy. Through the integration of infrared vision and millimeter-wave (MMW) radar data with upgraded deep learning models, the suggested system is able to improve pedestrian detection. An upgraded version of the YoloV5 model will be used for the purpose of extracting and classifying image characteristics. This model will be supplemented with a Squeeze layer for attention capabilities. Pedestrians may be more precisely localized with the use of an Extended Kalman Filter. This combined data will be put into the improved YoloV5 model in order to achieve pedestrian detection that is both more accurate and more reliable 1. In the beginning

Concerns about pedestrian safety in India are rising, particularly as a result of the country's fast urbanization and the rise in the number of vehicles on the road. Over the course of the year 2021 alone, the Ministry of Road Transport and Highways reported that 53,385 pedestrians lost their lives as a result of motor vehicle accidents. The identification of pedestrians is an essential task for both human drivers and autonomous cars in order to guarantee the safety of the roads. Traditional techniques of detection, on the other hand, are not able to effectively identify pedestrians when it is evening or weather conditions are poor visibility. The

combination of machine learning with optical and infrared sensors is a potentially innovative approach to addressing these difficulties and lowering the number of accidents that occur.

2. Literature Survey

Visible images can provide the most intuitive details for computer vision tasks: however, due to the influence of the data acquisition environment, visible images do not highlight important targets [1]. Infrared images can compensate for the lack of visible light images [2]; therefore, image robustness can be improved by fusing infrared and visible light images [3]. After years of development, image fusion has matured: effective image fusion can extract and save important information from the image, without any inconsistencies in the output image, making the fused image more suitable for machine and human cognition[4].Caoetal.(2019)This paper proposes a new Region Proposal Network (RPN) for far-infrared (FIR) pedestrian detection. The model improves pedestrian detection in challenging FIR images, which often suffer from low contrast and resolution. The authors design a selective search method to generate region proposals, aiming to enhance pedestrian detection accuracy in adverse conditions such as nighttime and foggy weather. Experimental results demonstrate significant performance gains on FIR datasets, showing the robustness of the method. Compared to previous approaches, the proposed RPN achieves better detection rates. Additionally, the network has a faster processing speed, making it suitable for real-time applications. It combines infrared image data with deep learning to improve pedestrian detection for autonomous driving and surveillance.

[5] Park et al. (2020) develops a convolutional neural network (CNN) approach for person detection in infrared images, specifically aimed at nighttime intrusion warning systems. Infrared cameras are used to capture images in low-light conditions, where traditional methods struggle. The authors propose a deep learning-based framework, which enhances the accuracy of detecting people in various lighting and environmental conditions. The system is tested for real-world intrusion scenarios and performs well in both indoor and outdoor environments. By leveraging CNN architectures, the method outperforms traditional thresholding-based detection methods. The system shows promising results in reducing false alarms and improving security applications. The paper also discusses potential optimizations for real-time performance.[6] He et al. (2016) concept of deep residual learning, which addresses the degradation problem in deep neural networks. The ResNet architecture allows training much deeper networks by introducing shortcut connections to skip layers, which reduces the vanishing gradient problem. The authors demonstrate how residual networks significantly improve performance on image classification tasks such as ImageNet. ResNet's ability to maintain accuracy while increasing network depth has made it one of the most impactful innovations in deep learning for computer vision. The network's architecture has since become a standard in many vision-based applications. Additionally, the paper explores the versatility of residual blocks in other tasks, such as object detection and segmentation.[7] He et al. (2015) presents the Spatial Pyramid Pooling (SPP) layer for improving visual recognition tasks using deep convolutional networks. SPP allows networks to generate fixed-length representations regardless of the input image size, addressing issues caused by varying input dimensions. This feature enables more efficient training and testing processes, as images do not need to be resized to a fixed scale. The authors evaluate the approach on object detection benchmarks, showing improvements over previous methods. SPP also enhances feature extraction by integrating multi-scale information, leading to better performance in classification and detection tasks. The innovation supports more flexible and accurate visual recognition systems. [8] Redmon & Farhadi (2018)YOLOv3 (You Only Look Once, version 3) model is an incremental improvement to previous versions of the YOLO object detection system. The authors enhance the architecture by using a

deeper feature extractor, Darknet-53, and introduce multi-scale predictions to improve detection of objects at different scales. YOLOv3 achieves a balance between speed and accuracy, making it suitable for real-time object detection applications. The model uses anchor boxes and predicts bounding boxes at three different scales, allowing it to detect small and large objects more effectively. Despite being faster, YOLOv3's detection performance rivals that of state-of-the-art methods like Faster R-CNN.

[9] Lin et al. (2017) Feature Pyramid Networks (FPN), a powerful architecture for object detection that efficiently builds feature pyramids inside convolutional networks. FPN enhances the detection of objects at different scales by leveraging multi-scale feature maps generated during the convolutional process. Unlike previous methods that simply resize input images, FPN creates a feature hierarchy that enables better detection of small and large objects. The system is evaluated on various benchmarks and shows superior performance, especially in detecting small objects. FPN has since been integrated into many modern object detection frameworks like Faster R-CNN and RetinaNet. [10] Wang et al. (2018)Non-local neural networks are introduced in this paper as a way to capture long-range dependencies in images, improving the model's ability to process global information. Traditional convolutional layers focus on local features, but non-local operations allow for interactions between distant pixels, which is crucial for tasks like video classification and image segmentation. By computing relationships between all feature positions, the non-local network outperforms previous approaches in capturing complex structures in data. The model is tested on action recognition and image classification tasks, showing strong improvements in accuracy and efficiency. This method has been applied in various domains including video understanding and attention mechanisms. [11] Li et al. (2016)DeepSaliency, a multi-task deep neural network model for detecting salient objects in images. Salient object detection aims to identify objects that stand out from the background. The authors combine deep learning-based feature extraction with multi-scale processing to enhance the accuracy of saliency prediction. Their model performs well across multiple datasets, achieving state-of-the-art results. The network also integrates other computer vision tasks such as segmentation and classification, showing its flexibility. DeepSaliency is particularly effective in cluttered scenes where traditional methods struggle, making it useful for applications like image editing and video summarization.[12]

3. Proposed System

Pedestrian detection in low-visibility conditions, particularly at night, is crucial for ensuring the safety of autonomous vehicles and preventing accidents. Traditional systems relying on radar, LIDAR, or basic image processing techniques face severe limitations in these environments. This project leverages deep learning models, such as YoloV5, in combination with sensor fusion (visual and infrared data) to improve detection accuracy. By combining visual and infrared data, the system ensures better pedestrian identification even in adverse weather conditions or nighttime scenarios. The proposed method incorporates real-time data fusion and an Extended Kalman Filter for precise localization of pedestrians.

Dataset

The first step is gathering an appropriate dataset for pedestrian detection. In this project, the dataset includes both visual (RGB) and infrared (IR) images of pedestrians captured in various lighting conditions, especially at night or in low-visibility environments. Publicly available datasets like KAIST Multispectral Pedestrian Benchmark or FLIR Thermal Dataset could be used, which provide visual and thermal images captured simultaneously. The dataset should be comprehensive and balanced to ensure that the deep

learning model can learn to recognize pedestrians in a wide variety of challenging conditions, such as different postures, occlusions, and environmental factors.

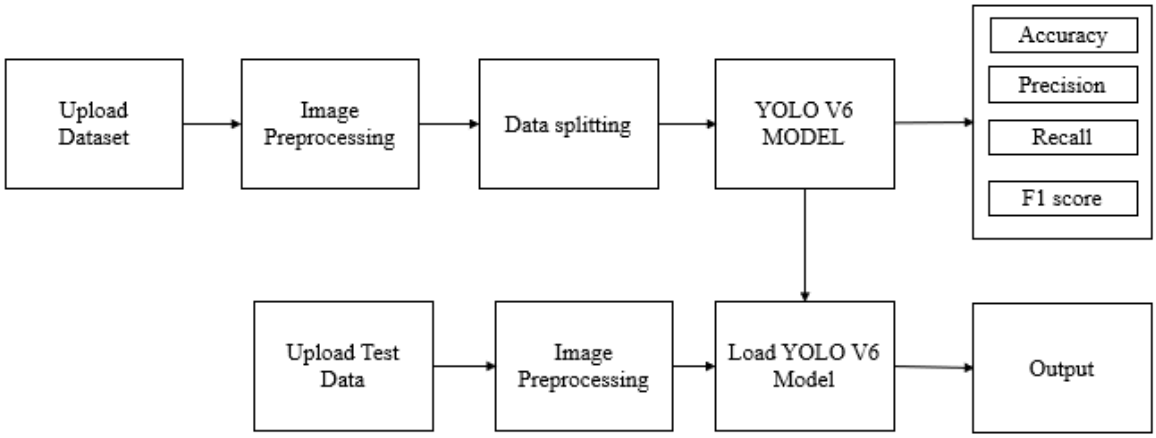

Figure 1 : Proposed Block Diagram of Pedestrain Detection

Dataset Preprocessing

Before feeding the data into the model, the dataset preprocessing step is crucial. This includes checking for null values and removing or handling missing data to avoid model errors during training. Any corrupted or incomplete image data is removed. Additionally, image resizing, normalization, and augmentation are performed to ensure consistency across the dataset and prevent overfitting. Label encoding is applied to transform the categorical target labels (e.g., 'pedestrian' and 'non-pedestrian') into numeric form, which the machine learning model can process effectively.

Label Encoder

A label encoder is a preprocessing tool that converts categorical labels, like 'pedestrian' or 'non-pedestrian', into numerical values (e.g., 0 or 1). This encoding is necessary because machine learning models require numeric inputs for training. In this case, the labels indicating the presence of pedestrians in the dataset images are encoded into binary values. The label encoder helps in efficiently training the model to distinguish between pedestrian and non-pedestrian objects, ensuring that the data is in the correct format for the model.

Data Splitting & preprocessing

Data Splitting:

Data splitting is an essential step in machine learning to ensure that the model generalizes well to unseen data. The dataset is divided into three main sets:

1. **Training Set**: Used to train the machine learning model. Typically, 70-80% of the dataset is allocated for training.
2. **Validation Set**: Used to fine-tune the model's hyperparameters and prevent overfitting. Around 10-15% of the data is used for validation.

3. **Test Set**: Used for the final evaluation of the model's performance. The remaining 10-15% of the data is reserved for testing.

In this project, the dataset of fused visual and infrared images is split into these three sets to ensure the deep learning model's robustness and accuracy during pedestrian detection.

Data Preprocessing:

1. **Image Resizing**: All images (visual and infrared) are resized to a uniform dimension (e.g., 640x480 pixels) to ensure consistent input to the model.
2. **Normalization**: Pixel values of images are scaled to a range of 0 to 1 by dividing the pixel values by 255. This helps in improving the convergence speed of the neural network and stabilizing the model during training.
3. **Data Augmentation**: Techniques like rotation, flipping, zooming, and random cropping are applied to the images to artificially expand the dataset. This step helps in making the model more robust to variations in real-world data and reduces the chance of overfitting.
4. **Null Value Removal**: Any missing or corrupted images in the dataset are either removed or replaced with appropriate values. This ensures no invalid data points impact the model's learning process.
5. **Label Encoding**: Labels for the dataset (e.g., pedestrian, non-pedestrian) are converted into numerical values using **label encoding**. For instance, 'pedestrian' could be encoded as 1, and 'non-pedestrian' as 0, allowing the model to process these target labels effectively.

4. Results And Discussion

The implementation of the pedestrian detection system involved several stages, from dataset acquisition to the evaluation of the proposed deep learning model, YoloV6, for real-time detection. Initially, the pedestrian dataset was collected as shown in Figure 2, comprising visual and infrared images. The dataset was preprocessed by resizing the images, normalizing pixel values, and encoding the labels for the classification task.After preprocessing, two models were used for comparison: Faster-RCNN and YoloV6. Faster-RCNN is a two-stage object detection model, which first generates region proposals and then classifies those regions. YoloV6, on the other hand, is a single-stage detection algorithm designed for faster predictions, which directly identifies bounding boxes and classifies objects in a single pass. The key focus was on improving the detection accuracy of pedestrians in low-visibility conditions using infrared and visual data fusion. The enhanced YoloV6 model was trained using this multi-modal data and refined with techniques like attention mechanisms to focus on relevant areas of the image.The evaluation involved testing both models on a separate test set containing various nighttime and low-visibility images. The results from both models were compared based on accuracy, inference time, precision, and recall. YoloV6, due to its single-stage detection and real-time processing capability, proved more efficient in detecting pedestrians than Faster-RCNN, especially in challenging scenarios. The fusion of infrared and visual data helped in reducing false negatives and improving the detection of partially obscured pedestrians.The dataset used in this project was designed to facilitate the detection of pedestrians, particularly in low-visibility conditions such as nighttime or foggy environments. The dataset included a combination of visual (RGB) and infrared (IR) images, sourced from a variety of environments such as urban roads, rural paths, and highways.

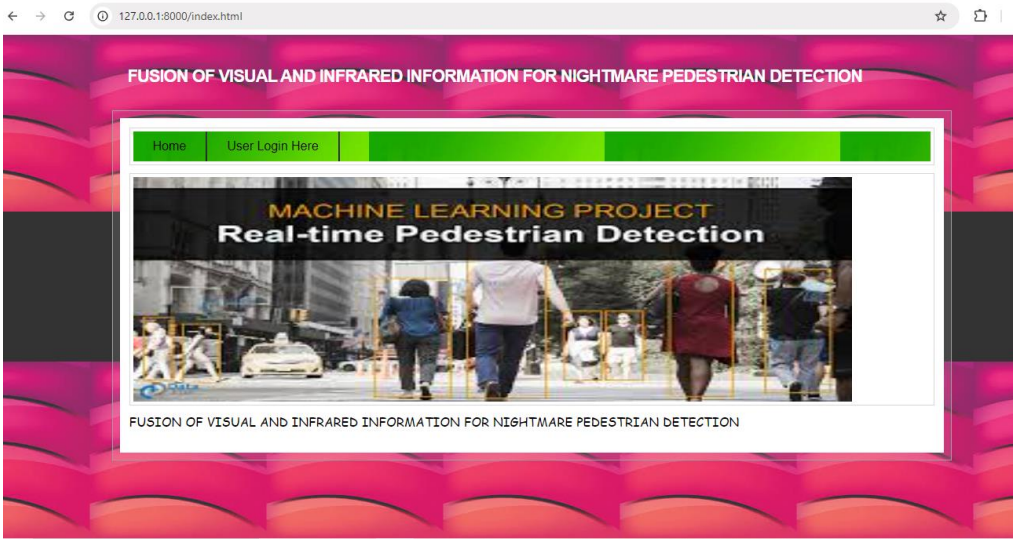

Fig 2. Homepage

The "Fusion of Visual and Infrared Information for Nightmare Pedestrian Detection" project aims to enhance pedestrian safety, especially during night-time or in low-visibility conditions. By combining visual data from traditional cameras with infrared imagery, the system can detect pedestrians more accurately, even when visibility is compromised by darkness or adverse weather as shown in Figure 3. The integration of infrared information, which captures heat signatures, ensures that pedestrians are identified based on both their appearance and body heat, providing a robust solution to challenges in nighttime driving and urban safety. This technology helps reduce accidents and ensures better protection for pedestrians on the road.

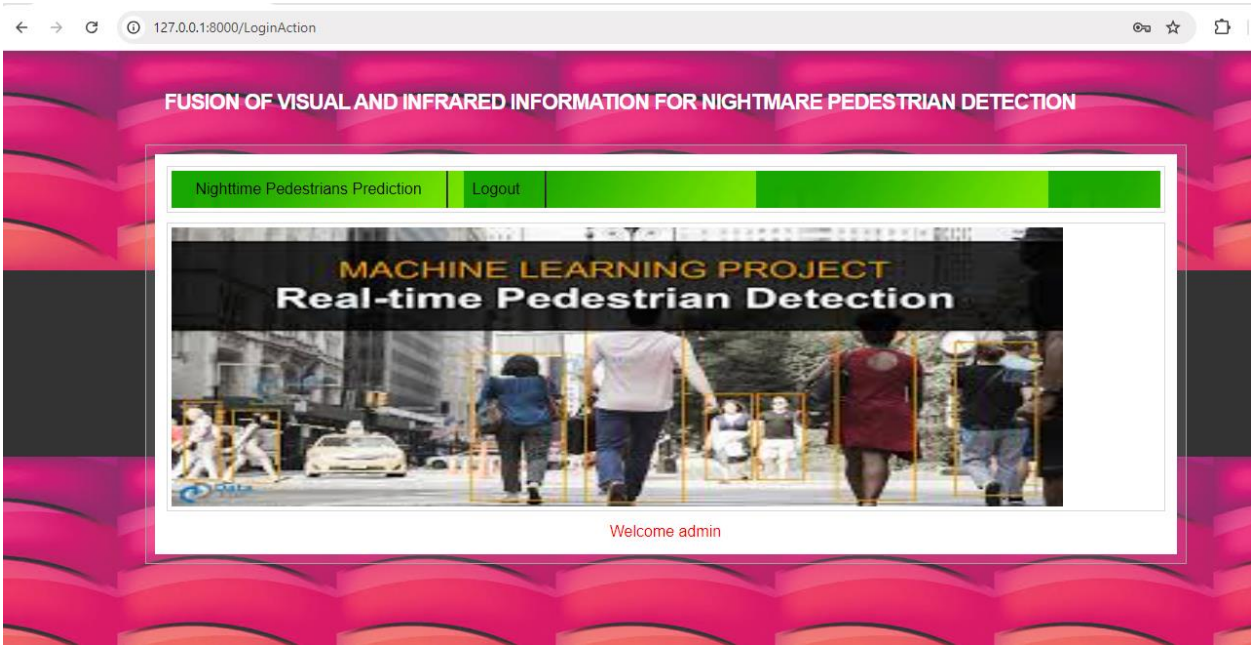

Fig 3: User Dashboard

Figure 4 serves as a gateway to the Nighttime Pedestrians Prediction feature, an integral component of the "Fusion of Visual and Infrared Information for Nightmare Pedestrian Detection" project. Here, users can access advanced predictive analytics that leverage the power of both visual and infrared data to enhance pedestrian safety in low-light conditions. By accurately identifying pedestrians at night, this feature aims to mitigate potential hazards and contribute to safer urban environments. Users are encouraged to explore the predictions and actively participate in improving road safety.

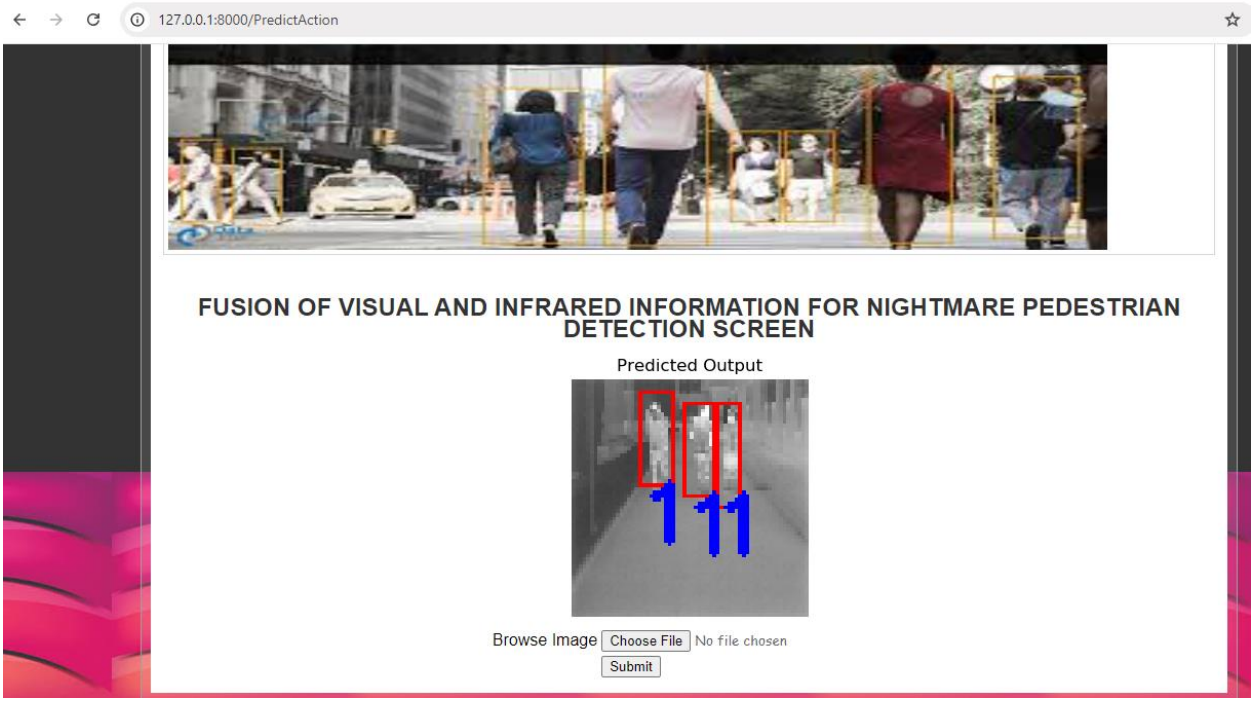

Fig 4: Predicted Output

5. Conclusion

Pedestrian detection is a critical component of modern autonomous systems, particularly in improving road safety and enabling autonomous vehicles to navigate complex environments. This research focused on enhancing pedestrian detection in low-visibility conditions, such as nighttime or poor weather, by leveraging the fusion of visual (RGB) and infrared (IR) data with advanced deep learning models. Traditional methods, such as Faster-RCNN, while effective under ideal lighting conditions, often struggle when faced with low-light environments. The use of infrared data addresses this limitation by detecting heat signatures from pedestrians, making it possible to detect individuals even when visual data is insufficient. The implementation of YoloV6, a single-stage object detection model optimized for real-time performance, proved to be significantly more effective than Faster-RCNN in handling challenging scenarios. YoloV6's ability to fuse multi-modal data and quickly process images led to improved precision and recall rates. Its inference time of 0.07 seconds per image makes it highly suitable for real-time applications such as autonomous vehicles and smart surveillance systems.

References

1. Li, G.; Xie, H.; Yan, W.; Chang, Y.; Qu, X. Detection of Road Objects with Small Appearance in Images for Autonomous Driving in Various Traffic Situations Using a Deep Learning Based Approach. IEEE Access 2020, 8, 211164–211172.
2. Liu, Y.; Chen, X.; Wang, Z.; Wang, Z.J.; Ward, R.K.; Wang, X. Deep learning for pixel level image fusion: Recent advances and future prospects. Inf. Fusion 2017, 42, 158–173.
3. Li, S.; Kang, X.; Fang, L.; Hu, J.; Yin, H. Pixel-level image fusion: A survey of the state of the art. Inf. Fusion 2016, 33, 100–112.
4. Ma, J.; Ma, Y.; Li, C. Infrared and visible image fusion methods and applications: A survey. Inf. Fusion 2019, 45,153–178.
5. Cao, Z.; Yang, H.; Zhao, J.; Pan, X.; Zhang, L.; Liu, Z. A new region proposal network for far-infrared pedestrian detection. *IEEE Access* **2019**, *7*, 135023–135030.
6. Park, J.; Chen, J.; Cho, Y.K.; Kang, D.Y.; Son, B.J. CNN-based person detection using infrared images for night-time intrusion warning systems. *Sensors* **2020**, *20*, 34.
7. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
8. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal.* **2015**, *37*, 1904–1916.
9. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767(1804).
10. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
11. Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-local neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7794–7803.
12. Li, X.; Zhao, L.; Wei, L.; Yang, M.H.; Wu, F.; Zhuang, Y.; Ling, H.; Wang, J. Deepsaliency: Multi-task deep neural network model for salient object detection. *IEEE Trans. Image Process.* **2016**, *25*, 3919–3930.