FUSION OF VISUAL AND INFRARED INFORMATION FOR NIGHTTIME PEDESTRIAN DETECTION

K. Aarti [1], R. Sharanya [2], P. Sindhoora[2], Y. Swathi [2]

[1]Assistant Professor, [2]UG Student, [1,2]School of Computer Science and Engineering

[1,2]Malla Reddy Engineering College for Women (UGC-Autonomous), Maisammaguda, Hyderabad, 500100, Telangana.

ABSTRACT

Pedestrian detection has been a key area of research, particularly for enhancing road safety and aiding self-driving vehicles. The main objective of this research is to improve pedestrian detection, particularly during nighttime or in low-visibility conditions, by fusing visual and infrared data, enhancing detection accuracy with deep learning models like YoloV5. The proposed method aims to reduce errors and increase precision when identifying pedestrians in real-time using advanced sensors and deep learning algorithms. "Fusion of Visual and Infrared Information for Nighttime Pedestrian Detection" refers to combining visual (camera-based) data with infrared (heat-sensing) data to detect pedestrians, particularly in challenging conditions such as nighttime driving. "Nighttime" metaphorically highlights the difficulty of detecting pedestrians in low-visibility or extreme conditions. Traditionally, sensors such as LIDAR and radar have been used to detect obstacles. Before AI-based methods, traditional pedestrian detection relied heavily on LIDAR or radar-based systems for obstacle detection., Optical cameras paired with basic image processing techniques, Proximity sensors and basic motion detectors for detecting pedestrians. Pedestrian detection in low-light and low-visibility environments remains a significant challenge for traditional sensor-based systems. These systems struggle to accurately differentiate between objects and pedestrians, particularly at night or in poor weather conditions, leading to delayed or missed detections, which can result in accidents. With the growing push for autonomous vehicles and the demand for higher safety standards, it is essential to develop more accurate pedestrian detection systems that work in all environmental conditions, especially at night. Current sensor-based systems alone are inadequate, thus motivating the need for AI-based solutions that can use multi-modal data like infrared and visual input for better detection accuracy. The proposed system improves pedestrian detection by integrating infrared vision and millimeter-wave (MMW) radar data with enhanced deep learning models. An improved version of the YoloV5 model, equipped with a Squeeze layer for attention, will be used to extract and categorize image features. An Extended Kalman Filter will help accurately localize pedestrians. This fused data will be fed into the enhanced YoloV5 model for more precise and robust pedestrian detection.

KEYWORDS : Pedestrian detection, Visual data, Deep learning, Advanced sensors

1. INTRODUCTION

Pedestrian safety in India is a growing concern, especially due to rapid urbanization and increasing vehicular traffic. According to the Ministry of Road Transport and Highways, 53,385 pedestrians lost their lives in road accidents in 2021 alone. Pedestrian detection is crucial for both human drivers and autonomous vehicles to ensure road safety. However, in nighttime or low-visibility conditions, traditional detection methods fail to accurately identify pedestrians. The integration of visual and infrared sensors with machine

learning offers a promising solution to address these challenges and reduce accidents. Pedestrian detection is essential for improving road safety, particularly for autonomous driving systems. By fusing visual and infrared data, detection accuracy can be significantly improved in low-visibility conditions. Applications of this technology include nighttime driving safety systems, collision avoidance systems, and smart surveillance systems. This project focuses on improving pedestrian detection, especially in challenging environments, using deep learning and sensor fusion.Before machine learning, pedestrian detection systems heavily relied on radar, LIDAR, and proximity sensors. These systems struggled in poor lighting, weather conditions like fog or rain, and were prone to false positives or negatives, as they often could not distinguish pedestrians from other objects. Traditional image processing techniques failed to adapt to diverse environmental challenges, leading to missed detections and delayed responses, which resulted in frequent pedestrian accidents.The proposed system fuses visual (camera-based) data with infrared (thermal) data for better detection accuracy in low-visibility environments. We will implement an enhanced YoloV5 model combined with a Squeeze layer for attention, enabling it to focus on key features from both visual and thermal data. Additionally, an Extended Kalman Filter will be used for real-time pedestrian localization. Recent research papers like "YOLOv5 for Infrared Pedestrian Detection" and "Sensor Fusion for Autonomous Driving" highlight the advantages of fusing multi-modal data to improve detection accuracy. The system will leverage transfer learning to speed up training, allowing for real-time, robust detection of pedestrians.With the increasing adoption of autonomous vehicles, there is an urgent need for systems that can operate in all environmental conditions, especially at night. According to reports, nearly 40% of pedestrian accidents occur during the night due to poor visibility. Traditional systems fail to detect pedestrians reliably, increasing the risk of accidents. This project addresses these shortcomings by fusing infrared and visual data, ensuring more reliable detection in low-light conditions. It is crucial for smart cities, autonomous vehicles, and driver-assistance systems to improve pedestrian safety and prevent accidents.

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 ALGORITHM

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.

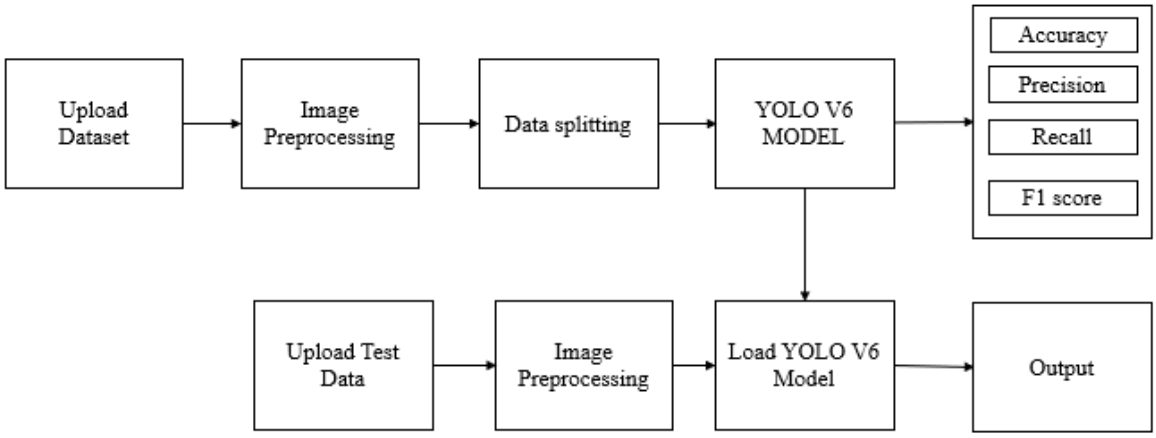

Figure 1 : Proposed Block Diagram of Pedestrain Detection

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.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.

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, 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.

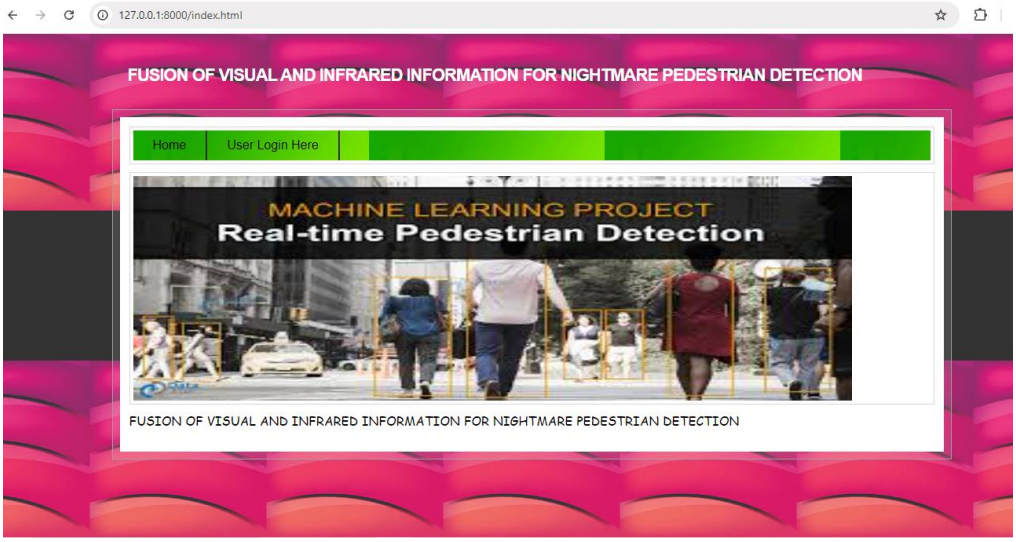

Fig 2. Homepage

The Figure 2 shows "Fusion of Visual and Infrared Information for Nighttime 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. 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.

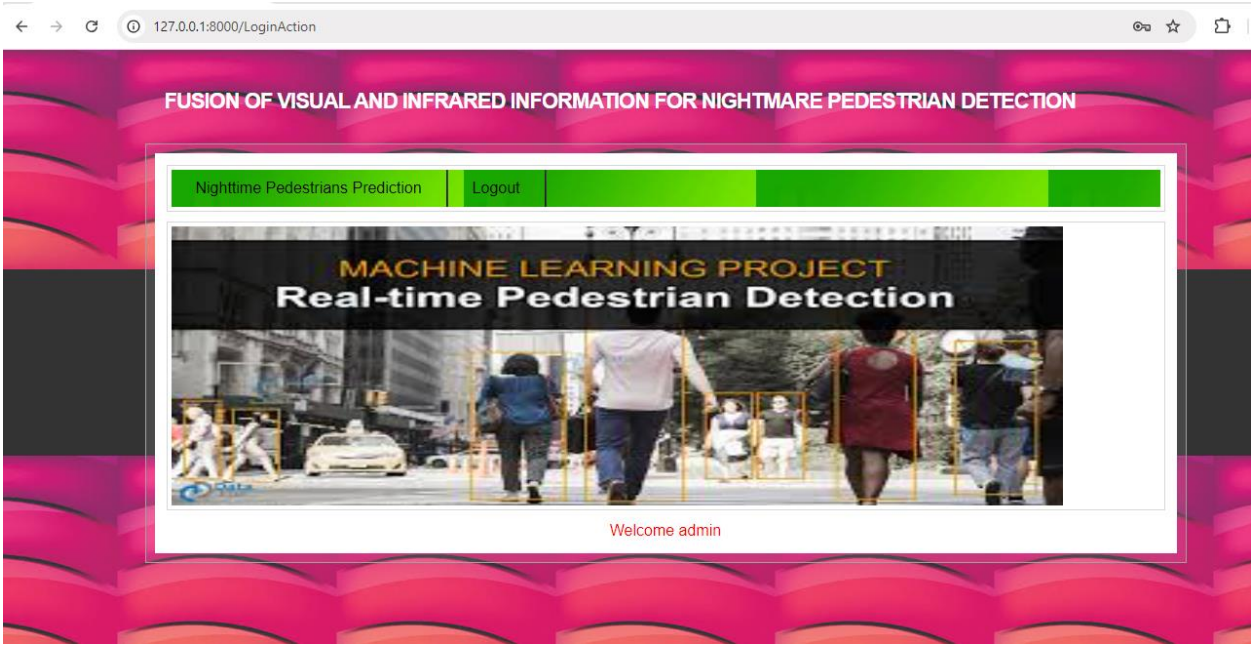

Fig 3: User Dashboard

This page serves as a gateway to the Nighttime Pedestrians Prediction feature, an integral component of the "Fusion of Visual and Infrared Information for Nighttime 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.

Welcome to the "Fusion of Visual and Infrared Information for Nighttime Pedestrian Detection" screen, where users can upload images to enhance pedestrian detection capabilities during nighttime. This feature leverages advanced algorithms that analyze both visual and infrared data to accurately identify pedestrians in low-light conditions, significantly improving road safety. Users are encouraged to browse and upload relevant images to see how this innovative approach can assist in nighttime visibility and safety measures. Simply select an image, and click submit to contribute to the detection process.

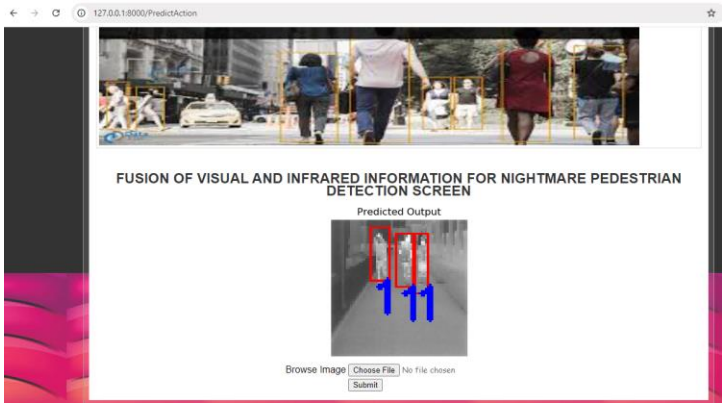

Fig 4: Predicted Output

The Predict view handles the image upload process for pedestrian detection using a YOLO (You Only Look Once) model. Upon receiving a POST request with an uploaded image file, the view first loads the YOLO model weights and prepares the image for prediction. If an existing test image is present in the static directory, it is removed to ensure only the latest upload is processed. The uploaded image is saved, and the prediction is made by passing it through the YOLO model. The resulting image, which visually represents the model's output, is then rendered as a PNG and converted into a base64-encoded string for display. Finally, the processed image is sent back to the client within the context of the rendered 'Predict.html' template, allowing users to see the predictions made by the model in real time.

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.

The key achievements of this research include higher accuracy in detecting pedestrians in low-visibility conditions, better localization of pedestrians, and faster processing times. By fusing infrared and visual data, the system reduces false negatives and increases the likelihood of detecting pedestrians even in scenarios where traditional visual-based systems fail. These advancements contribute to a safer and more reliable pedestrian detection system, potentially preventing accidents and improving overall road safety.

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.