

# Harmonizing Habitat: P-Yolov5 Enhanced Computer Vision For Mitigating Human-Wildlife Conflicts In Rural Areas

Jayasheelan Palanisamy<sup>1\*</sup>, Dr. S. Devaraju<sup>2</sup>

<sup>1\*</sup>Research Scholar, Dept of Computer Science, Sri Krishna Arts and Science College, Coimbatore, Tamil Nadu, India, sheelan.jsp@gmail.com

<sup>2</sup>Senior Assistant Professor, School of Computing Science and Engineering (SCSE), VIT Bhopal University, Bhopal Madhya Pradesh, devamcet@gmail.com

**Citation:** Jayasheelan Palanisamy, et al (2024), Harmonizing Habitat: P-Yolov5 Enhanced Computer Vision For Mitigating Human-Wildlife Conflicts In Rural Areas, *Educational Administration: Theory and Practice*, 30(6), 2263-2272,

Doi: 10.53555/kuey.v30i6.4767

## ARTICLE INFO ABSTRACT

In rural areas, the incursion of forest animals onto roads and into villages poses a substantial safety risk to residents. To address this issue, our research introduces an innovative variant of the You Only Look Once (YOLO) model, designated as P-YOLOv5. This model utilizes a MobileNetV3 backbone and a Feature Pyramid Network (PANet) neck to enhance real-time object detection and image recognition. P-YOLOv5 excels in balancing speed and accuracy, making it particularly suitable for applications like autonomous driving, surveillance, and robotics. The MobileNetV3 backbone provides an efficient framework for feature extraction, while the PANet neck enhances the model's capacity to capture contextual and spatial information across various scales. Experimental results showcase exceptional object detection performance, revealing high precision and recall rates on a meticulously pre-processed dataset. The Precision-Recall Curve further emphasizes the model's ability to strike a balance between accuracy and false positive rates, highlighting its practical applicability. Noteworthy is P-YOLOv5's achievement of a remarkable 0.93 mAP@0.5 for all classes, underscoring its robust capabilities for real-world object detection tasks in situations where forest animals pose a threat to human safety.

**Keywords:** *Animal Monitoring, YOLOV5, Object Detection, MobilenetV3, Modified YOLOV5 and Animal Surveillance*

## I. Introduction

The escalating incidents of wildlife encroachment into rural areas, with animals entering villages and traversing roads, pose substantial safety challenges for residents [1]. In response to this prevalent issue, our research introduces an innovative computer vision approach [2]. We propose a variant of the You Only Look Once (YOLO) model, denoted as P-YOLOv5, which leverages a MobileNetV3 backbone and a Feature Pyramid Network (PANet) neck to enhance real-time object detection and image recognition [3]. Chosen for its optimal balance between speed and accuracy, this technology finds application in critical domains such as autonomous driving, surveillance, and robotics.

The basis of our approach lies in meticulously curated datasets featuring instances of tigers and elephants infiltrating villages, sourced from diverse open platforms on Kaggle. These animals were selected due to their frequent interactions with human settlements, underlining the importance of understanding and mitigating human-wildlife conflicts[4]. Illustrated in Figure 1, the dataset captures pivotal scenarios through CCTV images, offering valuable insights into the complex dynamics between these animals and human environments.

To achieve our goals, we delve into the YOLO model and its evolution with P-YOLOv5, utilizing the efficiency of MobileNetV3 as its backbone and incorporating the innovative PANetneck[5]. This creates a lightweight yet powerful framework for processing input data. Subsequent sections elaborate on the components of the YOLO algorithm, the MobileNetV3 backbone, and the Feature Pyramid Network, emphasizing their collective role in achieving accurate and swift object detection.

Our experimental results underscore the effectiveness of P-YOLOv5 in object detection, showcasing high precision and recall rates[6]. The Precision-Recall Curve further emphasizes the model's ability to balance accuracy and false positive rates, reinforcing its practical applicability. Noteworthy is P-YOLOv5's achievement of an impressive 0.93 mAP@0.5 for all classes, highlighting its robust capabilities in real-world scenarios where wildlife poses a threat to human safety[7]. This research aims to contribute to the advancement of computer vision solutions addressing critical challenges in human-wildlife interactions, particularly in rural settings.

## II. Related Works

Deep learning, an effective machine learning approach that involves multiple layers of data representations, has proven to be highly successful in a variety of domains, with particular achievements in image classification, segmentation, and object detection. Recent advancements in deep learning have yielded promising results, especially in the challenging task of fine-grained image classification, which involves distinguishing subtle differences between subcategories. In this paper, we present a comprehensive examination of various deep architectures and models, focusing on their distinctive attributes [8]. We begin by explaining the functionality of Convolutional Neural Network (CNN) architectures and their components, followed by an in-depth exploration of a range of CNN models, spanning from classical LeNet to AlexNet, ZFNet, GoogleNet, VGGNet, ResNet, ResNeXt, SENet, DenseNet, Xception, and PNAS/ENAS. The primary emphasis is on the application of these deep learning architectures in three key areas: (i) wildlife detection, (ii) small arms detection, and (iii) human detection. For each model, we provide a comprehensive review summary encompassing system details, databases used, applications, and reported accuracy, offering valuable insights for future research in these application domains.

Amid escalating concerns regarding global threats from terrorism and illegal migration, the imperative to leverage cutting-edge technology to bolster security measures and safeguard both people and property has grown significantly [9]. Thermal cameras have assumed a pivotal role in advanced video surveillance systems due to their ability to operate effectively in challenging conditions, including low-light settings and adverse weather, where conventional RGB cameras may prove inadequate. This study delves into the automated detection of individuals in thermal imagery by adapting convolutional neural network models originally tailored for RGB image detection. The research evaluates the performance of leading object detectors, such as Faster R-CNN, SSD, Cascade R-CNN, and YOLOv3, post-retraining on a dataset of thermal images extracted from videos that simulate illicit movements along borders and within secured areas, encompassing diverse environmental conditions and movement patterns. Notably, YOLOv3 emerged as a standout choice, offering a compelling blend of speed and competitive performance. The study explores different training dataset configurations to determine the minimum number of thermal images required for achieving robust detection results, achieving impressive accuracy across various test scenarios with a relatively modest training dataset. The trained model is also subjected to evaluation using established thermal imaging datasets. Furthermore, the study addresses the recognition of both humans and animals in thermal imagery, a vital consideration for scenarios involving covert activities and border security. Additionally, the researchers introduce their original thermal dataset, comprising surveillance videos captured under various environmental conditions.

Camera traps are commonly used in wildlife surveys and biodiversity monitoring, often generating a substantial number of images or videos. To automate the identification of wildlife in these images and expedite analysis, deep learning techniques have been suggested. However, there is a lack of research validating and comparing the suitability of different object detection models in real field monitoring scenarios. In this study, we established a wildlife image dataset for Northeast Tiger and Leopard National Park (NTLNP dataset). We then assessed the recognition performance of three prevalent object detection architectures, comparing training models on day and night data separately and combined [10]. The results demonstrated satisfactory performance when training on both day and night data, with an average of 0.98 mAP for animal image detection and 88% accuracy for animal video classification. Notably, the one-stage YOLOv5m model achieved the highest recognition accuracy. The application of AI technology allows ecologists to efficiently extract valuable information from large image datasets, saving considerable time.

Animal-Vehicle Collision (AVC) poses a significant challenge on both urban and rural roads and highways, mainly due to the fast movement of vehicles and animals, cluttered environments, image noise, and occluded animals. While deep learning has been applied to animal-related issues, it often requires large training datasets, leading to complex models. In this study, we introduce an animal detection system to address AVC. Our system combines sparse representation and deep features optimized with FixResNeXt. We utilize a feature-efficient learning algorithm called Sparse Network of Winnows (SNoW) to represent deep features extracted from different parts of the animals sparsely [11]. Experimental results demonstrate the system's robustness to variations in viewpoint, partial occlusion, and illumination. On benchmark datasets, our system achieves an impressive average accuracy of 98.5%.

The ocean serves as a vital ecosystem, and aquatic creatures play a crucial role in the biological world, particularly in aquaculture. Precisely and intelligently recognizing and detecting aquatic animals is a pressing challenge in underwater biological research [12]. The widespread use of artificial intelligence (AI), particularly deep learning (DL), presents both opportunities and challenges for efficiently exploring aquatic animals. While DL has been extensively applied in the visual recognition and detection of land animals, its application in the underwater domain is still in its infancy due to the complexities of the underwater environment and data acquisition difficulties. This article provides an overview of the current state of DL application for aquatic animals, discussing potential challenges and future directions. Key advancements in DL algorithms for visual recognition and detection of aquatic animals are presented, encompassing datasets, algorithms, and performance. The article summarizes DL applications in aquatic animal research, including image and video detection, species classification, biomass estimation, behaviour analysis, and food safety. Moreover, challenges in object recognition and detection for aquatic animals are classified, followed by a discussion of future research directions and conclusions. Understanding the strides made in DL for recognizing and detecting aquatic animals is instrumental in expanding its use in marine biological exploration.

Animal-vehicle collisions represent a common hazard on highways, particularly during nocturnal driving. This increased risk is attributed not only to reduced visibility at night but also to the unpredictable behaviour of animals in close proximity to vehicles [13]. While thermal imaging has been studied extensively as a means to mitigate night time visibility issues, limited attention has been given to forecasting animal actions based on their specific postures in the presence of moving vehicles. This paper introduces an innovative system that combines a two-dimensional convolutional neural network (2D-CNN) with thermal images to evaluate the danger posed by animals in various postures to passing automobiles during night time hours. The proposed system was subjected to testing using thermal images depicting real-life scenarios of animals in specific postures near roadways, achieving an accurate classification of these postures in 82% of cases. Overall, this system provides a strong foundation for the development of automotive tools designed to reduce animal-vehicle collisions during night time driving.

The following are the novelties of this article are:

1. Optimizing the performance of object identification in an video frames.
2. Modifying the backbone characteristics enhances YOLO's ability to extract deeper features.
3. By modifying the backbone, YOLO's computation becomes progressively lighter, enhancing its Overall performance.

### III. Data Pre-Processing and Computer Vision Models

#### Data Collection

The animal dataset was carefully curated from multiple open-source platforms on Kaggle, with a specific emphasis on documenting instances of tigers and elephants invading villages [14]. This dataset comprises two primary categories: Tigers and Elephants. These classes were selected due to their frequent appearances in Indian villages, making them critical subjects for understanding and addressing human-wildlife conflicts. The images within the dataset offer valuable insights into the dynamic relationship between these magnificent animals and human settlements, contributing to ongoing efforts to manage and mitigate such interactions.



Tiger Image	Elephant Image
	
A) Tiger in CCTV in Village	B) Elephants on CCTV

Fig 1 Dataset Image from the CCTV Fields

### Data Pre-processing

In the process of data pre-processing, we standardized the image sizes by resizing them to a consistent 640x640 resolution. Subsequently, we applied a series of augmentation techniques to diversify the dataset. These techniques included shearing, 90-degree rotation, scaling with variances of up to 20%, zooming images by up to 23%, and converting them to grayscale. After the augmentation procedures, Table 1 represent the partitioned the dataset into three segments, allocating 70% for training, 20% for validation, and 10% for testing purposes. This meticulous approach significantly contributed to improving the model's accuracy in our analysis.

**Table 1 Dataset Splitting for Proposed Methodology**

Class Names	Training	Validation	Testing
Elephant	721	183	81
Tiger	684	206	91

### Object Localization

Object localization stands as a pivotal component within the realm of computer vision, focusing on the precise identification and positioning of objects within images or video frames. Unlike mere object recognition, object localization extends its scope to not only ascertain the presence of objects but also to pinpoint their specific locations, often accomplished through the establishment of bounding boxes or pixel-level segmentation masks [15]. This discipline finds application across a spectrum of fields, including object tracking, augmented reality, and autonomous navigation, furnishing machines with the ability to comprehend and engage with their surroundings with a heightened spatial awareness. The advent of deep learning models like Faster R-CNN and Mask R-CNN has significantly elevated the accuracy and efficiency of object localization, rendering it an indispensable tool in real-world scenarios.

### Object Detection

Object detection stands as a pivotal computer vision task, aimed at the identification and localization of objects within images or video frames. It surpasses mere image classification, as it not only recognizes the presence of objects but also pinpoints their exact positions and outlines. This technology finds extensive utility across diverse domains, including autonomous driving, surveillance, robotics, and retail [16]. Object detection techniques frequently leverage deep learning methodologies, particularly convolutional neural networks (CNNs), to process visual data, creating bounding boxes around detected objects, often enriched with class labels to specify the objects in question. The advent of real-time and highly accurate object detection has brought about transformative advancements in numerous industries, offering enhanced safety, efficiency, and automation in a wide array of applications.

## IV. PROPOSED METHODOLOGY

### YOU ONLY LOOK ONCES

The YOU ONLY LOOK ONCE (YOLO) model stands out as a revolutionary approach to real-time object detection in the realm of computer vision [17]. YOLO takes a unique approach by processing images as a unified entity, as opposed to the traditional two-step method of region proposals and subsequent classification. This design enables YOLO to swiftly and precisely detect objects in a single pass, distinguishing it with exceptional speed and accuracy. Its performance is characterized by its remarkable efficiency, making it particularly well-suited for applications demanding real-time object detection, such as autonomous vehicles and surveillance systems. YOLO maintains high standards of object recognition and localization, ensuring its reliability and utility. Its capability to handle multiple object classes simultaneously, even in complex scenarios, further underscores its significance in the field of computer vision, cementing its status as a preferred choice for applications where speed and precision are paramount.

### PROPOSED YOU ONLY LOOK ONCES V5 (P-YOLOv5)

The YOLO (You Only Look Once) version 5 model represents a significant step forward in the realm of real-time object detection and image recognition. YOLO V5 has gained recognition for its remarkable speed and accuracy, as it has optimized its architecture and incorporated efficient techniques to streamline the process of object detection[18]. This model excels in detecting objects with precision and swiftness, making it a valuable choice for real-time applications like autonomous driving, surveillance, and robotics. YOLO V5 delivers top-tier performance by achieving high accuracy in object detection while preserving the agility needed for swift decision-making and response across various industries. Its continuous development and improvements have established it as a prominent solution for object detection tasks.

The YOLO (You Only Look Once) algorithm consists of three key components: the backbone, neck, and head. The backbone serves as the initial feature extractor, analyzing the input image through a series of



convolutional and pooling layers to create feature maps that encode important image information [19]. These feature maps are then handed over to the neck, which acts as an intermediary component. The neck enhances the feature maps by fusing them and pooling them spatially, ensuring that the model can effectively detect objects of varying sizes and accurately handle objects distributed across the image. The final component, the head, is responsible for generating object detection predictions. It includes prediction layers that analyze the enhanced feature maps to produce bounding box coordinates, objectness scores, and class probabilities. The head's non-maximum suppression (NMS) step filters redundant predictions, resulting in a list of non overlapping bounding boxes representing detected objects.

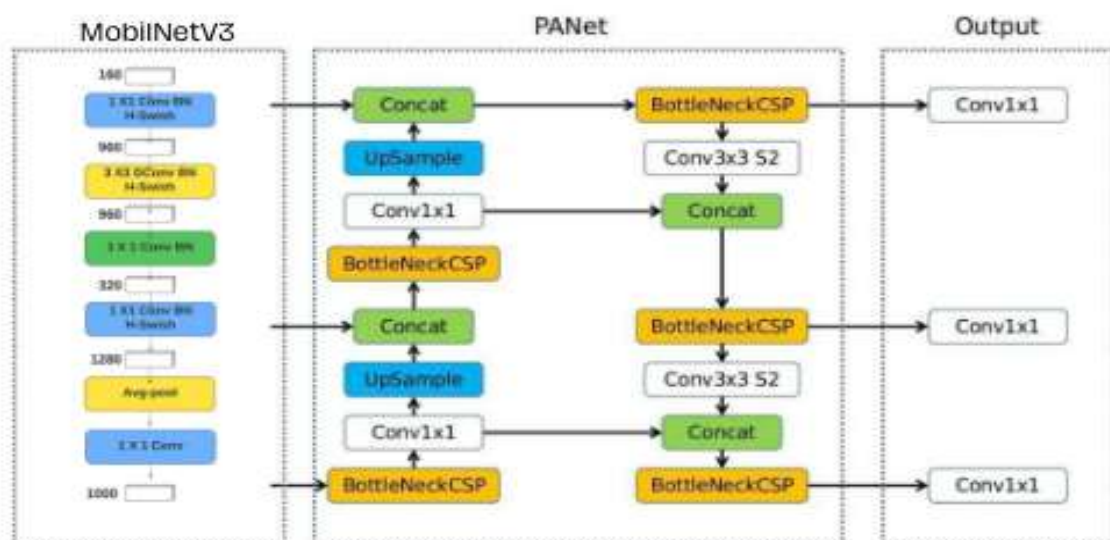


Fig 2 Proposed YOU ONLY LOOK ONCES V5

Figure 2 offers a detailed illustration of the novel methodology introduced in YOLOv5 M, which is a variant of YOLO version 5 featuring a MobileNetV3 backbone and a PANet neck. This altered architecture represents a significant advancement in object detection algorithms. The MobileNetV3 backbone, a pivotal component, has been adapted to serve as the foundational element for feature extraction, providing a lightweight yet efficient framework for processing input data. Additionally, the seamless integration of the PANet neck into the system augments the network's capacity to capture contextual and spatial information across various scales.

#### BackBone Network

The YOLOv5 architecture benefits from MobileNetV3 as its underlying backbone, a notably lightweight and efficient neural network design. MobileNetV3 is engineered to perform feature extraction with a strong emphasis on optimizing computational resources. Its distinctive characteristics include the utilization of inverted residual blocks with linear bottlenecks, facilitating the efficient capture of features across various scales. MobileNetV3 is organized into multiple stages, each featuring channel reduction, and it incorporates the Squeeze-and-Excitation (SE) module to enhance feature representation by recalibrating channel-wise responses[20].

#### Feature Pyramid Network (FPN) – Neck Network

Incorporated into YOLOv5 is a Feature Pyramid Network, labelled as PANet, aimed at improving the model's object detection capabilities across various scales[21]. This pyramid architecture effectively captures both fine-grained details and broader contextual information, enabling a more comprehensive understanding of the scene.

#### Detection Head

YOLOv5 employs a detection head that consists of multiple prediction layers, each responsible for detecting objects at a specific scale. These prediction layers predict bounding boxes, objectness scores, and class probabilities for multiple anchor boxes per grid cell.

#### Anchor Boxes

YOLOv5 uses anchor boxes to predict the size and position of objects within an image. These anchor boxes are pre-defined shapes that the model adjusts during training to better match the actual objects' sizes and proportions.

$$(x, y, w, h) = (\sigma(t_x) + c_x, \sigma(t_y) + c_y, p_w * e^{(t_w)}, p_h * e^{(t_h)})$$

**Non-Maximum Suppression (NMS)**

After object detection, YOLOv5 applies non-maximum suppression to remove duplicate or highly overlapping bounding boxes, retaining only the most confident predictions. This step helps reduce redundancy in the output. YOLOv5's post-processing phase refines the detected bounding boxes and class predictions, accounting for image resolutions and anchor box dimensions. This ensures that the final predictions align with the original image's coordinates.

**BackBone – Mobilenetv3**

MobileNetV3, specifically the Small variant, stands out as a noteworthy architecture in the realm of deep learning, particularly for applications where efficiency and lightweight design are paramount. Its architectural components are finely tuned to achieve an optimal balance between model size and performance [22]. At the core of MobileNetV3 Small's architecture lies the innovative concept of depth wise separable convolutions. These convolutions break down the operation into two distinct steps: depth wise convolution, which extracts spatial features, and pointwise convolution, which projects these features into a higher-dimensional space. This approach drastically reduces the number of parameters and computational costs compared to traditional convolutions while retaining the model's capacity to capture essential features within the input data.

In addition to depth wise separable convolutions, MobileNetV3 Small incorporates inverted residual blocks, which introduce non-linearity and richer feature representations [23]. These blocks are comprised of a lightweight bottleneck layer, reducing the number of channels, followed by depth wise separable convolution, and finally, an expansion layer that restores the channel dimensions. This structural design optimizes the trade-off between model size and performance. Furthermore, the architecture integrates the Squeeze-and-Excitation (SE) module, which adaptively recalibrates channel-wise responses by evaluating the importance of each feature map. This recalibration significantly enhances the model's discriminative power and its ability to focus on the most relevant features.

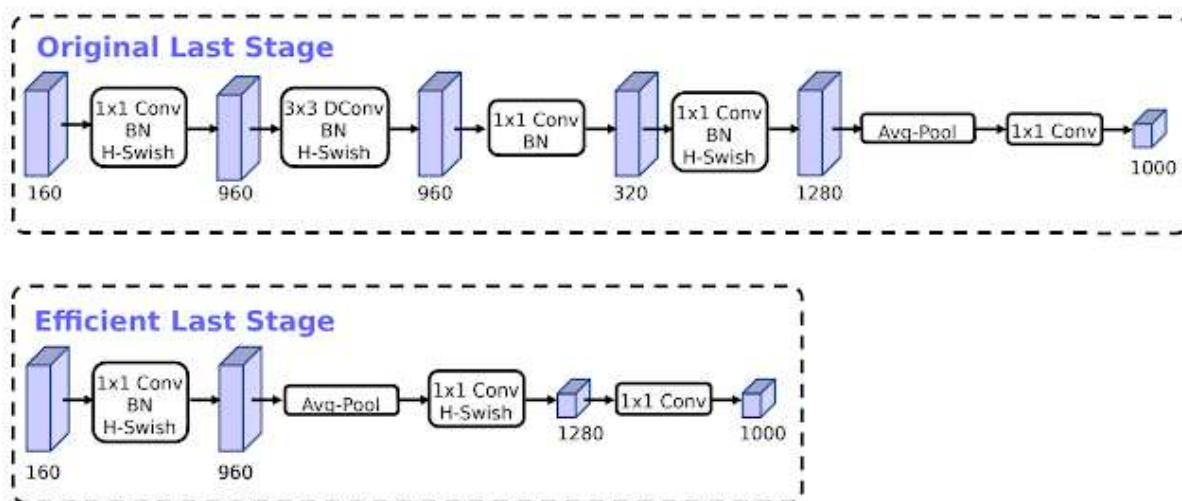


Fig 3 MobileNetV3 Architecture

In Fig 2 display MobileNetV3 Small typically consists of multiple stages, each containing varying numbers of inverted residual blocks. These stages are thoughtfully designed to capture features at different scales and complexities. The hierarchical approach empowers the model to comprehend input data comprehensively and adapt to a wide range of computer vision tasks. MobileNetV3 Small's efficient design and emphasis on speed make it particularly well-suited for real-time applications and edge devices where computational resources are limited. Moreover, its adaptability and versatility allow it to excel in various computer vision domains, including image classification, object detection, and semantic segmentation [24]. This architectural prowess positions MobileNetV3 Small as a valuable asset in the toolkit of deep learning practitioners.

The YOLOv5 architecture with the MobileNetV3 backbone yields a final output that consists of a list of bounding boxes, each accompanied by associated class labels and confidence scores. These bounding boxes accurately delineate the detected objects within the input image. YOLOv5, with its MobileNetV3 backbone, is renowned for its exceptional balance between speed and accuracy. When evaluated in terms of mean average precision (mAP) at an intersection over union (IOU) of 0.5, YOLOv5 holds its ground alongside Focal Loss while delivering processing speeds approximately four times faster. An important advantage is that the

model's size can be fine-tuned to tailor the trade-off between speed and accuracy without requiring retraining, rendering it a versatile and adaptable choice for a wide array of applications.

## V. Performance Metrics with Experimental Result

### Performance Metrics

#### F1 Confidence Curve

The F1 Confidence Curve depicts the relationship between F1 score and confidence thresholds in classification. It helps assess the precision-recall trade-off at various confidence levels.

$$F1ConfidenceCurve = 2 * (precision \times recall) / (precision + recall)$$

#### Precision Confidence Curve

The Precision Confidence Curve shows how well a model performs at different confidence levels in classifying things. It helps us see how the precision, or accuracy, changes as we adjust confidence levels.

$$Precision = True\ Positives / (True\ Positives + False\ Positives)$$

#### Recall Confidence Curve

The Recall Confidence Curve shows how the recall rate, which measures the ability to capture relevant instances, varies at different confidence thresholds in classification tasks. It visually depicts the changes in recall across various confidence levels, offering valuable insights into how well the model performs.

$$Recall = True\ Positives / (True\ Positives + False\ Negatives)$$

#### Precision – Recall Curve

The Precision-Recall Curve shows how precision and recall are related at different decision thresholds in classification. It visually demonstrates how precision and recall values evolve with changing confidence levels.

### Experimental Result

The experimental results for P-YOLOv5 demonstrate its strong performance in animal surveillance tracking. The proposed YOLOv5 achieved an impressive Mean Average Precision (mAP) of 0.93 at an intersection over union (IoU) threshold of 0.5 for object tracking. Figure 4 illustrates the training and validation loss, along with performance metrics, of the proposed model across different epochs. This includes the class loss, which helps determine the type of target object, and other relevant losses. The figure highlights that the backbone MobileNetV3 effectively reached a favourable local minimum during both the training and feature extraction phases.

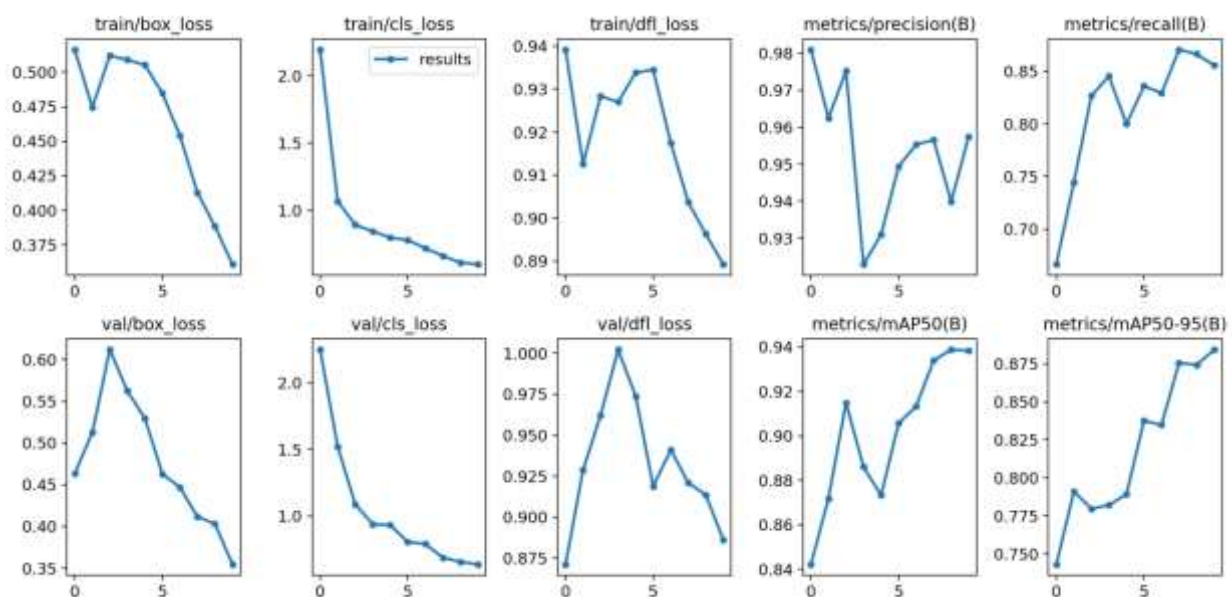


Fig 4 Proposed YOLOv5 Result for Training and Validation with losses

In Fig 5 represent the F1 confidence curve Recall confidence curve for the P-YOLOv5. In the wildlife animal class are Elephant and Tiger. In the F1 Confidence curve denote the F1 score for the all class image are at 0.90 at 0.604 is deonte that the hermoic mean of the precision and recall in the all class and same as Recall Confidence curve obtain the all class value for the P-YOLOv5 model is 0.98 at 0.000 it clearly declare the various threshold changes in the precision and recall.

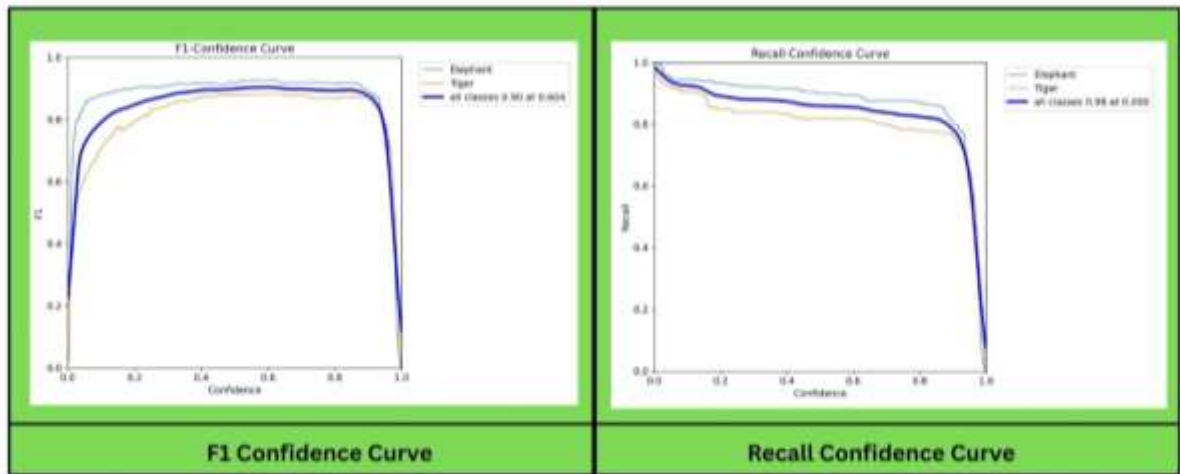


Fig 5 F1 and Recall Confidence Curve for the P-YOLOv5 Survaling and Tracking the wildlife Monitoring.

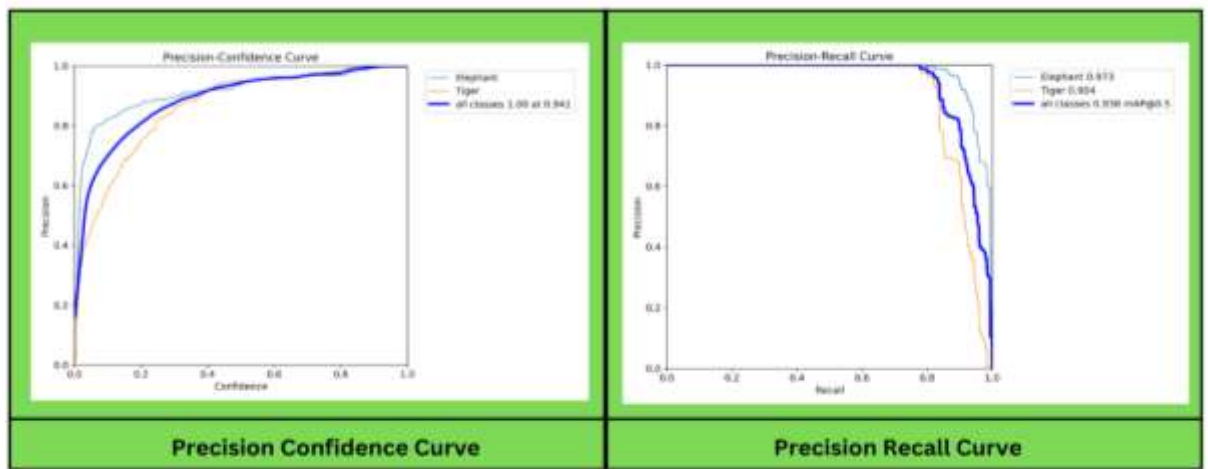


Figure 6 – Precision Confidence Curve and Precision-Recall Curve for P-YOLOv5 in Animal Monitoring

Figure 6 displays both the Precision Confidence Curve and Precision Recall Curve for the proposed surveillance and tracking system. In the Precision Confidence Curve, the accuracy for all classes reaches 1.00 at a confidence threshold of 0.941, highlighting the precision accuracy of the proposed model. The Precision Recall Curve illustrates the accuracy of precision and recall thresholds through the Mean Average Precision (mAP). Notably, Figure 6 distinctly presents the Mean Average Precision for two specific classes, and the overall mAP for all classes is reported as 0.93 at an intersection over union (IoU) threshold of 0.5. This clear representation indicates that the proposed model achieves high accuracy in surveillance and tracking tasks.

Table 2 Performance comparison of the various YOLO algorithms

Model	Precision Recall Value	Recall confidence value	Accuracy
YOLOv3-s	89.1@mAP	0.91	89
YOLOv5-s	91.7@mAP	0.95	92
Proposed YOLOv5	93.8@mAP	0.98	93

Table 2 compares YOLOv3, YOLOv5, and a modified version of YOLOv5 for wild animal detection. The analysis clearly favours the modified YOLOv5, showing it outperforms the other models in accuracy and efficiency. These results highlight the potential of the modified YOLOv5 for wildlife monitoring and conservation.

### VI. Conclusion

Our investigation delves into the urgent issue of wildlife intrusion into rural areas, posing a direct hazard to the safety of local residents. The adoption of the P-YOLOv5 model, integrating a MobileNetV3 backbone and a Feature Pyramid Network (PANet) neck, emerges as a robust solution for real-time object detection and



image recognition, particularly in scenarios involving tigers and elephants. The meticulously assembled dataset, documenting instances of human-wildlife conflicts, yields valuable insights into these interactions, deepening our comprehension of the challenges posed by such encounters. Our experiments affirm the exceptional performance of P-YOLOv5 in object detection, with high precision and recall rates. The model's adept management of accuracy and reduction of false positives, demonstrated by the Precision-Recall Curve, solidify its practical relevance across diverse settings, including autonomous driving, surveillance, and robotics. Notably, achieving an impressive 0.93 mAP@0.5 for all classes underscores the model's robust capabilities in real-world situations where wildlife poses a threat to human safety.

Looking forward, the knowledge derived from this research holds the promise of enhancing safety measures in rural areas grappling with human-wildlife conflicts. The proposed computer vision methodology, particularly leveraging P-YOLOv5, introduces avenues for improved monitoring and early intervention systems. This research not only contributes to the advancement of computer vision solutions but also underscores the critical need to address and mitigate challenges arising from the convergence of human and wildlife habitats. The findings highlight the importance of fostering harmonious coexistence between humans and the environment, marking a meaningful stride toward a safer and more sustainable future.

### Reference

1. Chandrakar, R., Raja, R., & Miri, R. (2021). Animal detection based on deep convolutional Neural networks with genetic segmentation. *Multimedia Tools and Applications*, 1-14.
2. Wu, W., Liu, H., Li, L., Long, Y., Wang, X., Wang, Z., ... & Chang, Y. (2021). Application of local fully Convolutional Neural Network combined with YOLO v5 algorithm in small target detection of remote sensing image. *PloS one*, 16(10), e0259283.
3. Yar, H., Khan, Z. A., Ullah, F. U. M., Ullah, W., & Baik, S. W. (2023). A modified YOLOv5 architecture for efficient fire detection in smart cities. *Expert Systems with Applications*, 231, 120465.
4. Robinson, N. B., Krieger, K., Khan, F. M., Huffman, W., Chang, M., Naik, A., & Gaudino, M. (2019). The current state of animal models in research: A review. *International Journal of Surgery*, 72, 9-13.
5. Kim, J. H., Kim, N., Park, Y. W., & Won, C. S. (2022). Object detection and classification based on YOLO-V5 with improved maritime dataset. *Journal of Marine Science and Engineering*, 10(3), 377.
6. Gu, Y., Wang, S., Yan, Y., Tang, S., & Zhao, S. (2022). Identification and analysis of emergency behavior of cage-reared laying ducks based on YoloV5. *Agriculture*, 12(4), 485.
7. Madhumathi, C. S., Naveen, V., Akshay, N., Kumar, M. S., & Aslam, M. M. (2023, April). Advanced Wild Animal Detection and Alert System using YOLO V5 Model. In *2023 7th International Conference on Trends in Electronics and Informatics (ICOEI)* (pp. 365-371). IEEE.
8. Dhillon, A., & Verma, G. K. (2020). Convolutional neural network: a review of models, methodologies and applications to object detection. *Progress in Artificial Intelligence*, 9(2), 85- 112.
9. M. Krišto, M. Ivasic-Kos and M. Pobar, "Thermal Object Detection in Difficult Weather Conditions Using YOLO," in IEEE Access, vol. 8, pp. 125459-125476, 2020, doi: 10.1109/ACCESS.2020.3007481.
10. Tan, M., Chao, W., Cheng, J. K., Zhou, M., Ma, Y., Jiang, X., ... & Feng, L. (2022). Animal detection and classification from camera trap images using different mainstream object detection architectures. *Animals*, 12(15), 1976.
11. Meena, S. D., & Loganathan, A. (2020). Intelligent animal detection system using sparse multi discriminative-neural network (SMD-NN) to mitigate animal-vehicle collision. *Environmental Science and Pollution Research*, 27(31), 39619-39634.
12. Li, J., Xu, W., Deng, L., Xiao, Y., Han, Z., & Zheng, H. (2023). Deep learning for visual recognition and detection of aquatic animals: A review. *Reviews in Aquaculture*, 15(2), 409- 433.
13. Mowen, D., Munian, Y., & Alamaniotis, M. (2022). Improving road safety during nocturnal hours by characterizing animal poses utilizing CNN-based analysis of thermal images. *Sustainability*, 14(19), 12133.
14. Nad, C., Roy, R., & Roy, T. B. (2022). Human elephant conflict in changing land-use land- cover scenario in and adjoining region of Buxa tiger reserve, India. *Environmental Challenges*, 7, 100384.
15. Zhang, D., Han, J., Cheng, G., & Yang, M. H. (2021). Weakly supervised object localization and detection: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(9), 5866-5885.
16. Wu, X., Sahoo, D., & Hoi, S. C. (2020). Recent advances in deep learning for object detection. *Neurocomputing*, 396, 39-64.
17. Cao, C. Y., Zheng, J. C., Huang, Y. Q., Liu, J., & Yang, C. F. (2019). Investigation of a promoted you only look once algorithm and its application in traffic flow monitoring. *Applied Sciences*, 9(17), 3619.
18. P. T. R. Thangaraj, P. P. U. R. M and B. Vadivelu, "Real-Time Handgun Detection in Surveillance Videos based on Deep Learning Approach," *2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, Salem, India, 2022, pp. 689-693, doi: 10.1109/ICAAIC53929.2022.9793288.

19. Sun, X. M., Zhang, Y. J., Wang, H., & Du, Y. X. (2022, February). Research on ship detection of optical remote sensing image based on Yolo V5. In *Journal of Physics: Conference Series* (Vol. 2215, No. 1, p. 012027). IOP Publishing.
20. Zhao, L., & Wang, L. (2022). A new lightweight network based on MobileNetV3. *KSII Transactions on Internet & Information Systems*, 16(1).
21. Chen, S., Zhao, J., Zhou, Y., Wang, H., Yao, R., Zhang, L., & Xue, Y. (2023). Info-FPN: An Informative Feature Pyramid Network for object detection in remote sensing images. *Expert Systems with Applications*, 214, 119132.
22. Li, G., Fan, H., Jiang, G., Jiang, D., Liu, Y., Tao, B., & Yun, J. (2023). RGBD-SLAM Based on Object Detection With Two-Stream YOLOv4-MobileNetv3 in Autonomous Driving. *IEEE Transactions on Intelligent Transportation Systems*.
23. Jia, L., Wang, T., Chen, Y., Zang, Y., Li, X., Shi, H., & Gao, L. (2023). MobileNet-CA- YOLO: An Improved YOLOv7 Based on the MobileNetV3 and Attention Mechanism for Rice Pests and Diseases Detection. *Agriculture*, 13(7), 1285.
24. Deng, T., & Wu, Y. (2022). Simultaneous vehicle and lane detection via MobileNetV3 in car following scene. *Plos one*, 17(3), e0264551.