

Research Paper

NOVEL ANIMAL DETECTION SYSTEM USING IMPROVED YOLO

VANGALA SWATHI¹, MD. TUSHAARA², M. AJAY KUMAR³, M. SWAROOP⁴,
MD. FAYAZ⁵

¹Assistant Professor, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam,
Telangana, India

^{2, 3, 4, 5} B.Tech, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam,
Telangana, India

ABSTRACT— Wild animal intrusion in residential areas and sudden road crossings pose significant risks to both humans and wildlife. Leveraging advanced technologies, this research proposes a reliable wildlife detection system based on a cascaded YOLOv11 approach. The system plays a crucial role in monitoring biodiversity, understanding animal behaviour, and supporting global conservation efforts. To enhance detection accuracy, the input dataset is pre-processed using adaptive histogram equalization for contrast enhancement. Following this, segmentation is performed using a super-pixel-based Fast Fuzzy C-Means (FCM) clustering technique to identify regions of interest. Features are subsequently extracted and fed into the

cascaded YOLOv11 framework for accurate detection and classification of wild animals. The proposed approach ensures robust and efficient wildlife monitoring, offering a valuable solution to mitigate human-wildlife conflicts while contributing to environmental sustainability and wildlife conservation.

Index Terms – Animal Detection, Improved YOLO, Computer Vision, Deep Learning, Real-Time Object Detection, Wildlife Monitoring, Image Processing, Smart Surveillance, Artificial Intelligence, Machine Learning.

I. INTRODUCTION

Wild animal intrusion into residential areas and sudden road crossings present significant

threats to human safety, wildlife preservation, and ecosystem stability. These challenges necessitate the development of advanced wildlife detection systems that can accurately identify and monitor animal movement in real time. This research introduces a cascaded YOLOv11-based approach for wildlife animal detection, which provides a robust and efficient solution for mitigating human-wildlife conflicts, monitoring biodiversity, and supporting global conservation initiatives.

The proposed system is designed to achieve high accuracy in wildlife detection by integrating multiple advanced preprocessing, segmentation, and detection techniques. The input dataset initially undergoes Adaptive Histogram Equalization (AHE) for contrast enhancement, which improves the visibility of features in low-light or high-contrast environments commonly encountered in wildlife imagery. AHE dynamically adjusts image contrast, ensuring that critical details within the data, such as animal contours and distinguishing features, are preserved and highlighted for subsequent analysis.

Following contrast enhancement, the enhanced images are subjected to super-pixel-based Fast Fuzzy C-Means (FCM) clustering for segmentation. This method

effectively partitions the input images into homogeneous regions (super-pixels) while maintaining a balance between segmentation accuracy and computational efficiency. The FCM-based segmentation groups image pixels with similar characteristics into clusters, isolating regions of interest (e.g., wild animals) from the background. This step significantly reduces noise and enhances the precision of feature extraction, which is critical for accurate detection.

II. RELATED WORK

A. Visual Instruction Tuning

This paper introduces a novel approach called Visual Instruction Tuning (VIT), which aims to enhance the capabilities of visual models in understanding and executing complex visual instructions. Traditional visual models are primarily trained on large-scale datasets with minimal task-specific guidance, which can lead to suboptimal performance in specialized tasks. Visual instruction tuning addresses this limitation by incorporating task-specific visual cues during the model's training process. This allows the model to better interpret nuanced visual cues and instructions, improving its accuracy in tasks such as object detection, image captioning, and semantic segmentation. The authors also explore the potential

applications of VIT in multi-modal AI systems where understanding and generating tasks from both visual and textual data is crucial. Experimental results show significant performance improvements when using VIT on several benchmark datasets, demonstrating the efficacy of visual instruction as a means to bridge the gap between vision and language-based tasks. This research opens new avenues for developing more flexible and task-aware visual models.

B. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation

The paper introduces the concept of rich feature hierarchies for the dual tasks of object detection and semantic segmentation, significantly advancing the state-of-the-art in both areas. Traditional object detection approaches struggled with low detection accuracy and the inability to handle complex background clutter. The authors propose a solution through the integration of multiple levels of convolutional feature hierarchies. By leveraging deep convolutional networks (CNNs), they enhance the ability of the model to recognize objects at various scales, which is crucial for real-world applications where objects may appear in different sizes or orientations. The paper introduces R-

CNN, a model that extracts region proposals from an image and classifies each proposal using CNNs, yielding better detection accuracy. Additionally, the authors combine CNN-based features with robust object proposal algorithms like selective search to significantly improve the detection process. The method's success demonstrated the power of deep learning in the field of object detection and laid the foundation for further developments such as Fast R-CNN and Faster R-CNN, which optimize and accelerate the detection process.

C. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks

Faster R-CNN represents a groundbreaking advancement in real-time object detection by introducing Region Proposal Networks (RPNs), which allow for joint training of the region proposal generation and object detection processes. Unlike previous models that used external algorithms like selective search to generate region proposals, Faster R-CNN generates proposals directly from the feature map of the convolutional layers. This unified approach eliminates the need for costly external region proposal algorithms, thus reducing computational overhead and enabling faster processing times. RPNs use a small sliding window to propose regions of

interest, which are then classified and refined by the subsequent layers of the Faster R-CNN pipeline. The paper shows that this method not only speeds up object detection but also maintains or even improves accuracy compared to traditional methods. The authors validate their model on several challenging datasets, demonstrating that Faster R-CNN achieves state-of-the-art performance in real-time object detection. This paper marks a significant shift towards integrated, end-to-end learning systems in computer vision, influencing future research in object detection architectures.

III. PROPOSED METHODOLOGY

The proposed **YOLOv11** model aims to address the limitations of previous versions, such as YOLOv7 and YOLOv8, by incorporating advanced techniques to improve detection accuracy, robustness, and computational efficiency. YOLOv11 introduces an enhanced backbone network with **improved feature extraction capabilities** using hybrid **transformer-based attention mechanisms** and efficient convolutional modules. These enhancements allow the model to better capture fine-grained details, making it highly effective in detecting **tiny and occluded objects** across complex backgrounds. To optimize inference speed, the model incorporates **lightweight**

modules and quantization techniques, making it suitable for real-time applications on resource-constrained devices. Moreover, YOLOv11 integrates a **dynamic anchor-free detection mechanism** with adaptive grids, reducing computational overhead while maintaining high localization accuracy. This proposed approach enhances detection precision, improves generalizability across diverse datasets, and provides a scalable solution for real-world applications, including **wildlife monitoring, surveillance systems**, and other critical detection tasks.

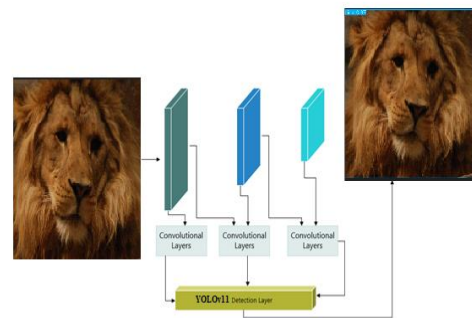


Fig. 1: System Overview

3.1 Input Data Module

- Handles image and video datasets collected from wildlife monitoring sources such as CCTV cameras, drones, or other surveillance systems.

3.2 Preprocessing Module

- Implements Adaptive Histogram Equalization for contrast enhancement.
- Standardizes image size and format for consistency.

- Displays results with bounding boxes and detection confidence scores.
- Provides visual outputs in the form of marked images, videos, or real-time feeds.

3.3 Segmentation Module

- Utilizes Super-Pixel-Based Fast Fuzzy C-Means (FCM) clustering to segment animal regions from the background efficiently.

3.4 Feature Extraction Module

- Incorporates hybrid transformer-based attention mechanisms and convolutional layers to extract significant features like edges, textures, and shapes.

3.5 YOLOv11 Detection Module

- Deploys the cascaded YOLOv11 model for multi-stage detection.
- Leverages multi-scale feature fusion (FPN and PAN) for improved detection across small, medium, and large objects.
- Uses anchor-free detection for reducing computational overhead and enhancing accuracy.

Output and Visualization Module



Fig. 2: Home Page

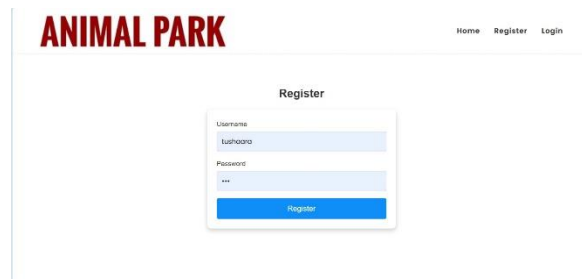


Fig. 3: Registration

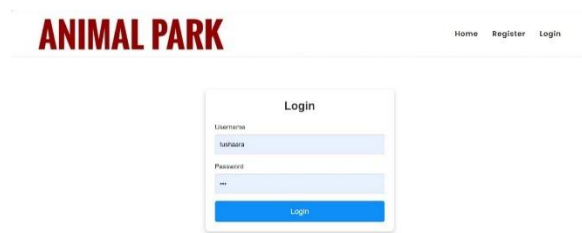


Fig. 4: Login

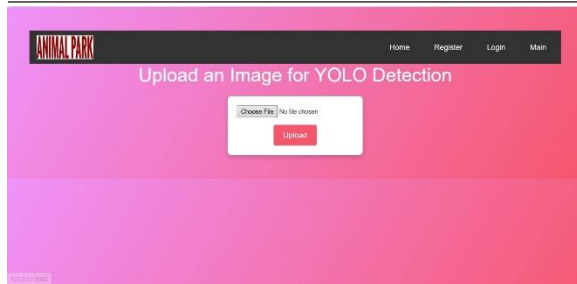


Fig. 5: Upload Image

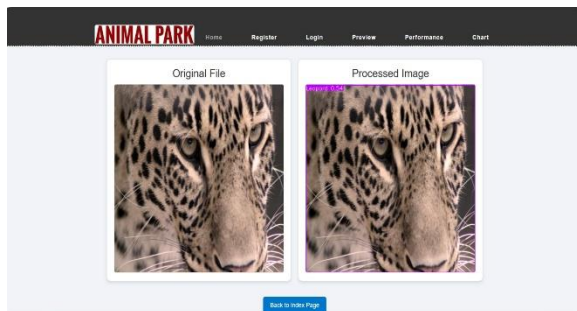


Fig. 6: Detect Animal

V. CONCLUSION

In conclusion, the YOLOv11-based wildlife detection system offers a highly efficient and accurate solution for monitoring wildlife and preventing human-animal conflicts in real-time. By integrating advanced techniques like transformer-based attention mechanisms, multi-scale feature fusion, and anchor-free detection, the system overcomes many of the limitations found in previous YOLO versions, ensuring better handling of occlusions, small objects, and complex environments. The proposed system's real-time performance and scalability make it a valuable tool for wildlife conservation, biodiversity monitoring, and reducing

accidents caused by animal intrusions. While the current system provides significant improvements, future enhancements, such as multi-modal data integration, deep learning-based tracking, and user-friendly interfaces, will further elevate its impact. Ultimately, this system serves as a vital step toward leveraging technology for the protection of wildlife and the preservation of natural ecosystems.

REFERENCES

- [1] WHO road traffic deaths http://www.who.int/gho/road_safety/mortality/en/
- [2] J.L. DSouza, Road Safety: A Critical Issue Requiring Government Attention. International Journal of Education and Science Research review, Volume-2, Issue-2, April, 2015.
- [3] J. Pang, I. Singh, Accelerometer Based Real-Time Remote Detection and Monitoring of Hand Motion. Proceedings of the World Congress on Engineering and Computer Science 2011 Vol II WCECS 2011, October 19-21, 2011, San Francisco, USA.



[4] Author: Scott J. Weiner, Feasibility of a 802.11 VANET Based Car Accident Alert System.

[5] halil, U., Javid, T., & Nasir, A. (2017).

Automatic road accident detection techniques: A brief survey. 2017 International Symposium on Wireless Systems and Networks (ISWSN).

[6] Faiz, A. B., Imteaj, A., & Chowdhury, M. (2015). Smart vehicle accident detection and alarming system using a smartphone. 2015 International Conference on Computer and Information Engineering (ICCIE).

AUTHORS Profile



Ms. VANGALA SWATHI has received her B.Tech from JNTUH, Hyderabad in 2015, M.Tech degree in Computer science

from JNTUH, Hyderabad in 2017. She is dedicated to teaching field from the last 7 years. Her research area Deep Learning. At present she is working as Assistant Professor in CSE Department at Guru Nanak

Institutions Technical Campus ,Ibrahimpatnam, Telangana, India.

Ms. MD. TUSHAA-RA pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. M. AJAY KUMAR pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions

Technical Campus affiliated to JNTUH in 2022-2026.



M. SWAROOP pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus

affiliated to JNTUH in 2022-2026.



Mr. MD. FAYAZ pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions

Technical Campus affiliated to JNTUH in 2022-2026.