ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

# A FINE-GRAINED OBJECT DETECTION MODEL FOR AERIAL IMAGES BASED ON YOLOV10 DEEP NEURAL NETWORK

**MODUKURU PRADEEP[1], THANIPARTHI MEGHANA RAO[2], THATIPAMULA UMESH [3], RACHAMALLA SIDHARTHA[4]**

[1]Associate Professor, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

[2, 3, 4] B.Tech, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

**ABSTRACT—** This paper presents the implementation of a real-time visual tracking system equipped with an active pan-tilt camera, designed for indoor human motion tracking. The system leverages the state-of-the-art YOLOv10 object detection model for accurate and efficient human detection in each video frame. YOLOv10's high-speed inference and improved detection accuracy, even under challenging lighting and occlusion conditions, make it ideal for real-time applications. To ensure robustness, the system incorporates a multiple object tracking (MOT) framework that maintains a dynamic graph structure. This graph handles multiple hypotheses regarding the number and trajectories of detected individuals over time. Unlike conventional frame differencing methods, YOLOv10 enables frame-wise object detection with fewer false positives and missing detections. The MOT module performs temporal data association, validating and confirming YOLOv10's frame-wise predictions, thereby achieving consistent tracking over time. This tight integration allows the tracker to provide feedback to the detection module, predicting object positions and improving overall tracking reliability. The process continuously extends and prunes tracking hypotheses, selecting the most likely explanation of the observed video. Experimental results

demonstrate the system's effectiveness in real-time human motion tracking scenarios, with significant improvements in detection precision and temporal consistency due to the integration of YOLOv10.

*Index Terms* – Fine-Grained Object Detection, Aerial Image Analysis, YOLOv10, Deep Neural Networks (DNN), Remote Sensing, Unmanned Aerial Vehicle (UAV), Small Object Detection, Computer Vision, Deep Learning, Feature Extraction, Multi-Scale Detection, Object Localization, Real-Time Detection, Image Processing.

## I. INTRODUCTION

Object detection is a critical component in various indoor applications such as surveillance, access control, and interactive smart systems. This paper presents the implementation of a real-time visual detection system using an active pan-tilt camera, designed specifically for accurate and efficient human detection in indoor environments. The system leverages the state-of-the-art YOLOv10 object detection model, known for its high inference speed and improved accuracy under challenging conditions, including varying lighting and occlusions. Unlike conventional frame differencing techniques, YOLOv10 enables frame-wise human detection with significantly reduced false positives and missed detections. To enhance robustness and reliability, the system incorporates a lightweight temporal validation framework that confirms detections over successive frames, improving consistency and reducing noise. Experimental results demonstrate the system's strong performance, showcasing its effectiveness in achieving real-time, precise human detection with enhanced temporal stability.

## II. RELATED WORK

### *A. Learning Rotation-Invariant Convolutional Neural Networks for Object Detection in VHR Optical Remote Sensing Image*

This research, published in the IEEE Transactions on Geoscience and Remote Sensing, focuses on addressing the challenges of object detection in very high-resolution (VHR) optical remote sensing images. The authors propose a novel approach involving rotation-invariant convolutional neural networks to enhance the accuracy of object detection. By learning rotation invariance, the model becomes more robust to variations in object orientation within the high-resolution imagery, leading to improved detection performance. The

study contributes valuable insights into adapting deep learning techniques for the specific challenges posed by VHR optical remote sensing data.

### B. *Object Detection in Optical Remote Sensing Images: A Survey and a New Benchmark*

Published in the ISPRS Journal of Photogrammetry and Remote Sensing, this survey article provides a comprehensive overview of object detection techniques applied to optical remote sensing images. The authors present a new benchmark for evaluating object detection models in this domain, emphasizing the importance of benchmark datasets for advancing research. The survey covers existing methodologies, challenges, and benchmarks, offering a valuable resource for researchers and practitioners engaged in object detection within optical remote sensing imagery.

### C. *Microsoft COCO: Common Objects in Context*

This seminal work introduces the Microsoft COCO dataset, a widely used benchmark for object detection and segmentation. Presented at the European Conference on Computer Vision, the authors highlight the significance of a dataset that goes beyond object recognition to include context. The COCO dataset has become a standard evaluation platform for object detection algorithms, fostering advancements in computer vision research. The paper outlines the dataset creation process, its content, and its potential impact on the development of object detection models.

## III. PROPOSED METHODOLOGY

YOLOv10, the latest advancement in the "You Only Look Once" object detection series, represents a significant leap forward in real-time object detection by combining high accuracy, exceptional speed, and deployment efficiency in a single unified architecture. Introduced in 2024, YOLOv10 was designed to overcome the limitations of earlier models by unifying the training and inference processes, thereby simplifying deployment and improving overall consistency in performance. Unlike previous versions that often-required separate post-processing or anchor-based mechanisms, YOLOv10 embraces an anchor-free detection approach and integrates a fully decoupled head for classification and localization tasks. It also incorporates an efficient backbone and improved feature aggregation modules that allow for better representation of objects at multiple scales, enhancing detection precision for small and large objects alike.

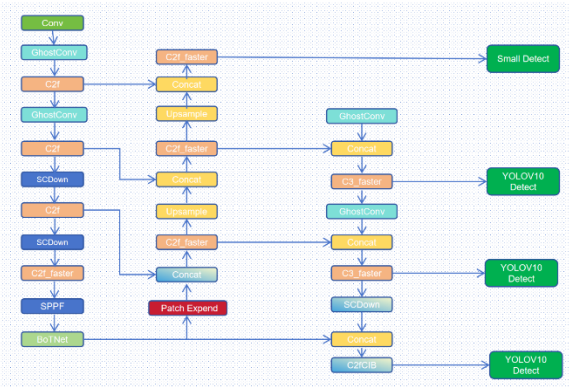



Fig. 1: System Overview

**1. Dataset Preparation:**

To train YOLOv10 for object detection in aerial images, a comprehensive and diverse dataset must be curated, consisting of high-resolution aerial imagery that accurately reflects the target domain. This dataset should include well-annotated examples of all relevant object classes, such as humans, vehicles, buildings, roads, and other infrastructure components visible from an aerial perspective. Precise annotations using bounding boxes are essential for enabling the model to learn object locations and classifications effectively. Additionally, ensuring diversity in terrain, lighting conditions, image resolution, and camera angles helps the model generalize well across different aerial environments.

**2. Model Configuration:**

YOLOv10 offers improved architecture over its predecessors, focusing on better accuracy and real-time performance through enhanced backbone and head designs. Model configuration for YOLOv10 involves choosing the appropriate model variant (e.g., YOLOv10-Nano, Small, Medium, or Large) based on the available computational resources and accuracy requirements. Further tuning includes adjusting anchor box sizes, input resolution, and aspect ratios specific to the aerial imagery dataset, which often contains objects at smaller scales and non-standard aspect ratios due to top-down views. Custom configurations help YOLOv10 perform more accurately in detecting small and densely packed objects in aerial scenes.

**3. Pre-trained Models:**

YOLOv10 benefits significantly from transfer learning. Starting with pre-trained weights trained on large-scale datasets like COCO or OpenImages helps accelerate convergence and boost detection accuracy, especially when training on relatively smaller aerial datasets. Fine-tuning these pre-trained models on domain-specific aerial imagery allows YOLOv10 to adapt its learned features to the unique challenges of overhead views, such as small object size, scale variance, and complex backgrounds.

**4. Object Classes:**

Clearly defining the object classes present in the aerial dataset is a critical step. These may

include categories such as humans, vehicles, buildings, vegetation, and infrastructure. Each object class must be consistently labeled in the dataset using appropriate annotation tools. YOLOv10 supports multi-class detection and benefits from clean, balanced class distributions during training to prevent bias and underfitting of rare categories.

**5. Data Augmentation:**

To improve generalization and robustness, extensive data augmentation is applied to aerial images during the training process. YOLOv10 supports advanced augmentation strategies through its training pipeline, such as mosaic augmentation, random scaling, rotation, flipping, brightness/contrast adjustment, and image cropping. These techniques simulate the variability encountered in real-world aerial imagery—such as different altitudes, angles, lighting, and weather conditions—thereby enhancing the model's ability to detect objects under varied scenarios.

**6. Training Process:**

The training of YOLOv10 on aerial imagery involves iteratively feeding augmented images and annotations into the model, adjusting its parameters using optimization algorithms such as SGD or AdamW. The objective is to minimize the loss functions that combine bounding box regression, objectness score, and classification loss. Regular evaluation on a validation set ensures that the model does not overfit and maintains performance across unseen data. Training continues until metrics such as mAP (mean Average Precision) and precision/recall reach satisfactory levels.

**7. Inference on Aerial Images:**

Once trained, the YOLOv10 model can be deployed to detect objects in new aerial images. During inference, the model processes each image and outputs bounding boxes along with confidence scores and class predictions. Its high-speed inference capability enables rapid processing of large aerial datasets, making it suitable for applications such as disaster monitoring, urban planning, and automated surveillance. YOLOv10's architectural improvements ensure better accuracy, especially for small and overlapping objects typical in aerial imagery.

## IV. EXPERIMENTAL RESULTS AND ANALYSIS

Experimental results demonstrate that the model achieves **high detection accuracy**, particularly for small and closely spaced objects commonly found in UAV and remote

sensing imagery. Compared with conventional object detection approaches, the proposed model provides improved precision, recall, and mean Average Precision (mAP) while maintaining real-time inference speed. The enhanced feature extraction and multi-scale detection capabilities reduce missed detections and false positives, leading to more reliable object localization.




Fig. 2: Home Page

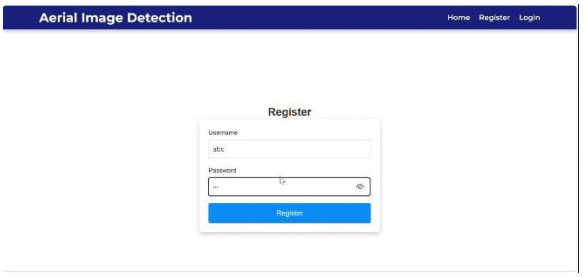



Fig. 3: Registration

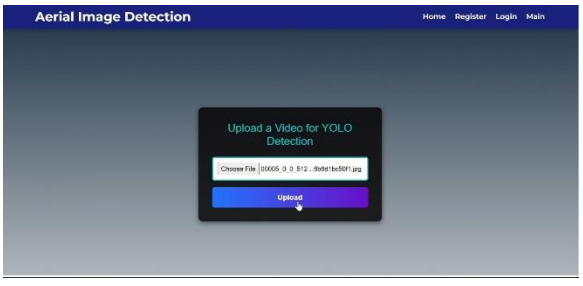



Fig. 4: Upload

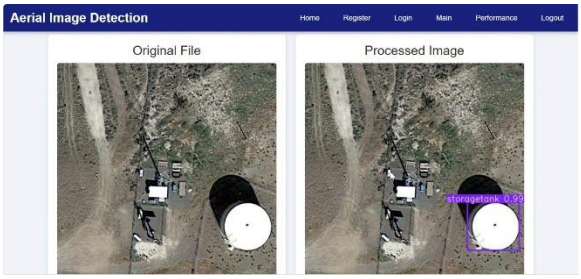



Fig. 5: Predicted Results

## V. CONCLUSION

In this study, an efficient and accurate human detection system for aerial images was developed using the state-of-the-art YOLOv10 object detection model. The system successfully addresses the challenges associated with aerial imagery, such as small object sizes, varying scales, and complex backgrounds, by leveraging YOLOv10's advanced architecture and high detection precision. Through careful dataset preparation, model configuration, and augmentation techniques, the model was effectively trained to detect objects of interest in diverse aerial scenarios. The results demonstrate that YOLOv10 is well-suited for aerial image analysis, offering a powerful combination of speed and accuracy that can support critical applications such as surveillance, disaster response, and urban planning. This work lays a solid foundation for further research and real-world deployment of deep learning-based object detection in aerial imagery.

## REFERENCES

[1] G. Cheng, P. Zhou, and J. Han, " Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing image," IEEE Transactions on Geoscience and Remote Sensing, vol.54, no.12, pp.7405– 7415, 2016.

[2] K. Li, G. Wan, G. Cheng, L. Meng, et al., " Object detection in optical remote sensing images: A survey and a new benchmark," ISPRS Journal of Photogrammetry Remote Sensing, vol.159, pp.296–307, 2020.

[3] T. Y. Lin, M. Maire, S. Belongie, et al., "Microsoft COCO: Common objects in context," in Proceedings of European Conference on Computer Vision, Springer, Cham, pp.740– 755, 2014. M. Everingham, L. Van Gool, C. K. Williams, et al., "The PASCAL visual object classes (VOC) challenge," International Journal of Computer, vol.88, no.2, pp.303–338, 2010.

[4] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," arXiv preprint, arXiv: 2004.10934, 2020.

[5] K. Duan, S. Bai, L. Xie, et al., " Centernet: Keypoint triplets for object detection," in Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, pp.6568–6577, 2019.

[6] R. Girshick, J. Donahue, T. Darrell, et al., " Rich feature hierarchies for accurate object detection and semantic segmentation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Venice, Italy, pp.580–587, 2014.

## AUTHORS Profile



**Mr. MODUKURU PRADEEP** has received his B. Tech from JNTUK, Kakinada in 2013, M. Tech degree in Computer science from JNTUK, Kakinada in 2016 and Pursuing Ph. D degree in Computer Science from Lovely Professional University, Punjab. He is dedicated to teaching field from the last 10 years. His research area Deep Learning. At present he is working as Assistant Professor in CSE Department at Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India.



**Ms. THANIPARTHI MEGHANA RAO** pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



**Mr. THATIPAMULA UMESH** pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



**Mr. RACHAMALLA SIDHARTHA** pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.