Deep Learning based enhanced aerial object detection

Avinash Chouhan, Dibyajyoti Chutia, Biswarup Deb and S. P. Aggarwal
North Eastern Space Applications Centre, Umiam Meghalaya, India
Email: avinash.chouhan@nesac.gov.in

(Received on 12 November 2024; in Final Form on 28January 2025)

***DOI**: https://doi.org/10.58825/jog.2025.19.1.197*

Abstract: In congested urban environments, accurate detection and counting of humans and vehicles provide valuable insights for optimizing traffic flow, identifying congestion hotspots, and designing efficient transportation systems. By leveraging computer vision algorithms, such as deep learning based object detection models, real-time monitoring of pedestrian and vehicular traffic can be achieved with high accuracy and granularity. The ability to precisely quantify pedestrian and vehicle movements enables urban planners and policymakers to make data-driven decisions regarding infrastructure development, road maintenance, and public transit planning. In this work, we enhanced the existing deep learning based network architecture for object detection using UAV images. The proposed work enhances the super-resolution pipeline, which contributes to the overall performance improvement of the object detection model. The enhanced network architecture can detect and give a count of the number of objects for any particular area in the image.

Keywords: Object detection, UAV, Super resolution

.

1 Introduction

Detecting objects in images captured by Unmanned Aerial Vehicles (UAVs) using deep learning methods has become a crucial research focus, with wide-ranging applications such as monitoring, emergency response, and infrastructure analysis. Convolutional neural networks (CNNs), a subset of deep learning models, have proven highly effective in recognizing and locating objects in UAV imagery by leveraging patterns learned from large, annotated datasets. These models excel in distinguishing objects within complex environments, despite challenges like varying lighting conditions. A significant hurdle in this area is the limited availability of labeled datasets for training. UAVs, also known as drones, are remotely controlled or can operate autonomously via pre-programmed flight paths. Equipped with advanced sensors like high-resolution cameras and LiDAR, they capture detailed aerial imagery, which provides valuable spatial data for object detection. The rapid development of UAV technologies has led to increasing interest in object detection from aerial perspectives, which this paper explores in depth. In this work we enhanced the existing object detection method to effectively detect small objects using aerial inputs. We also curated a drone image dataset for object detection task which can be shared for future research. The improved object detection model can be used to identify aerial objects with smaller sizes in occluded environments precisely. It is crucial for applications like autonomous drones and disaster management, where it effectively handles challenges such as scale variability and cluttered backgrounds.

2 Literature Review

In (Zhang et al. 2023), the author introduces SuperYOLO, an efficient and precise object detection method that integrates multimodal data. It performs detection across multiple objects using assisted Super Resolution (SR) learning, optimizing both object detection accuracy and computational efficiency. In (Papageorgiou et al. 1998), the author introduces a novel framework for object detection in complex scenes, this method uses a comprehensive set of basic functions along with techniques from statistical learning. In (Pathak et al. 2018), the author examines the use of CNN-based deep learning methods for object detection, providing an overview of available frameworks, benchmark datasets, and key applications. Additionally, a comparison of state-of-the-art deep learning approaches is presented. Surveys deep learning-based techniques for object detection and tracking in UAV-acquired data, while (Zhou et al. 2022) introduces FasterX, a lightweight detector for UAV imagery. The PixSF head, proposed within this framework, includes a position encoder-decoder to enhance small object detection by incorporating positional embedding in the head layer. In (Mittal et al. 2020), a review of Deep Learning (DL) based object detection algorithms focuses on UAV datasets at low-level altitudes, which are less studied than standard datasets. The study compares two-stage detectors like faster region convolutional neural networks and Cascade region convolutional neural networks with one-stage models such as YOLO and Single Shot MultiBox Detector (SSD) and also explores advanced detectors like CornerNet. The analysis highlights research gaps, particularly in improving detection accuracy and performance for UAV applications. In (Zhang et al. 2020), the author introduces a multi-task symmetric network, which integrates a model using dilated convolutions to enhance detection performance, especially in dense scenes. Tested on VisDrone and UAVDT datasets, GDF-Net can be applied to other object detection models to improve recall and precision. In (Liu et al. 2020), the author proposes a Multi-branch Parallel Feature Pyramid Network (MPFPN) aimed at improving small object detection in UAV images. MPFPN recovers features lost in deeper layers and applies a supervised spatial attention module (SSAM) to reduce background interference. A downward architecture is utilized in the faster region convolutional neural network stage to improve localization accuracy, and experiments on the VisDrone-DET dataset demonstrate competitive performance against state-of-the-art models. In (Albaba et al. 2021), the author introduces SyNet, an ensemble network that combines both multi-stage and single-stage detection approaches to minimize false negatives and enhance detection quality. SyNet integrates CenterNet and Cascade R-CNN, leveraging pretrained feature extractors and an ensemble strategy to improve object detection performance by capitalizing on the advantages of both detection methods. In (Bejiga et al. 2016), the author presents a technique to support search and rescue (SAR) operations using UAVs equipped with smart cameras. The UAV captures images of debris-avalanche, which are processed by a pretrained CNN for feature extraction, followed by a linear SVM classifier to distinguish objects of interest. This method increases the efficiency of SAR missions. In (Xiao et al. 2020), a method is proposed for detecting arbitrary-oriented and densely packed objects in UAV images. Evaluated on DOTA, VEDAI, and VisDrone datasets, the method enhances multiscale feature extraction and employs a rotating region network to capture distinctive features effectively. In (Li et al. 2023), the author introduces an improved version of YOLOv5, called YOLOv5-VTO, designed to enhance the detection of obscured and small vehicles in aerial imagery. This model adds a detection branch for small targets, replaces PAN with BiFPN for better multiscale feature fusion, and incorporates Soft-NMS to improve detection under occlusion. Experiments demonstrate that YOLOv5-VTO surpasses the original YOLOv5s. In (Liu et al. 2020), the author presents an automated deep learning (DL) method utilizing Faster-RCNN to detect and count oil palm trees from UAV drone imagery. By cropping and sampling images into sub-images, the model was trained to accurately locate individual trees, achieving high detection accuracy. In (Hosseinpoor et al. 2020), the study introduces a modified UNet encoder-decoder model for building extraction from high-resolution remote sensing data. By integrating ResNet-101 into the encoder and applying embedded feature fusion in the decoder, the model enhances segmentation accuracy, as demonstrated on a publicly available urban scene dataset. In (Aswini et al. 2021), a deep learning approach for frontal object detection using monocular drone camera images is proposed, comparing three pre-trained neural networks based on speed and accuracy. In (Meng et al. 2020), an approach consisting of UAV remote-sensing with deep learning for the detection of ground excavators is introduced, resulting in the UAV-ED system, which includes modules for UAV control, detection, geolocation, and real-time information transfer. In (Zhang et al. 2020), a system combining ultramicro oblique and infrared thermal imaging is proposed for detecting structural damage and cracks in buildings post-disaster. By constructing 3D remote sensing images and analyzing 2D plane images, the system extracts structural information while infrared thermal images help detect temperature anomalies, enabling rapid post-disaster damage assessment. In (Zhang et al. 2020), the Global Density Fused Convolutional Network (GDF-Net) is proposed for object detection in UAV images, optimized for high-density object scenarios. The model integrates a Global Density Model (GDM) using dilated convolutions, improving recall and precision, especially in congested scenes, and has been tested on VisDrone and UAVDT datasets. In (Maharjan et al. 2022), an automatic method for mapping plastic waste in rivers using UAVs and deep learning models is proposed. YOLOv5s, tested at two sites in Laos and Thailand, performed best, achieving a high mean Average Precision (mAP) for plastic detection. In (Micheal et al. 2022), a deep learning-based framework for accurate multi-object tracking in UAV videos is introduced. The system uses a tiny object detector for detection and an LSTM-based network tracker for tracking, reducing identity switching and improving performance over existing methods. In (Bazi et al. 2018), a novel Convolutional Support Vector Machine (CSVM) network is introduced, combining convolutional layers with SVMs for feature extraction and classification. Using forward supervised learning instead of backpropagation, the method shows promising results for vehicle and solar panel detection in UAV imagery. In (Wu et al. 2021), a CNN-based modeling method for single object detection on UAVs is proposed, without using transfer learning. An inducing neural network is included to accelerate training convergence, enabling the model to select and distinguish specific objects within the same category. In (Sun et al. 2022), an algorithm based on YOLOv3 is proposed. It improves the accuracy of detection by leveraging shallow feature maps for location prediction, fusing local and global features in a Feature Pyramid Network (FPN), adaptively weighting FPN outputs, and refining the Squeeze-and-Excitation attention mechanism for precise feature response adjustments.In (Zeng et al. 2023), the author introduces YOLOv7-UAV, a real-time small object detection algorithm optimized for UAV aerial imagery. Key improvements include the removal of the second downsampling layer, the introduction of module for enhanced multi-scale feature extraction, optimized anchor allocation through binary K-means clustering, and the use of weighted Gaussian Wasserstein distance for sample assignment. These modifications boost detection speed, increase mean Average Precision (mAP) compared to YOLOv7, and reduce model complexity. In (Saetchnikov et al. 2021), the author presents a comparative study of deep neural networks for detecting objects with similar patterns, using limited pre-trained datasets. In (Wu et al. 2022), the author introduces the Cross Channel Reconstruction (CCR) module for remote sensing image classification. CCR enables the features from different modalities extracted by the CNN to be fused more compactly. This approach often results in improved classification performance compared to traditional fusion methods based on concatenation or alignment. In (Wu et al. 2023), UIU-Net was proposed for infrared small object detection, combining a small U-Net within a larger backbone for multi-scale feature learning, with RM-DS and IC-A modules enhancing contrast. In (Li et al. 2023), LRR-Net introduced a deep learning-based hyperspectral anomaly detector, integrating trainable low-rank representation to improve scalability and reduce manual tuning.

3 Dataset

Creating a high-quality dataset is a critical step in the object detection pipeline, as the model's performance and accuracy largely depend on the dataset. This phase is essential for evaluating and improving various algorithms. The advent of large-scale datasets, featuring millions of images, has been instrumental in achieving significant advancements in object detection capabilities.

NESAC dataset: We utilized a drone to capture UAV images at a resolution of 4000 x 3000 pixels. These images were primarily taken from urban areas in the Meghalaya and Assam regions of North-East India. Although we aimed to collect a larger dataset, we limited the scope to the images we could acquire. To optimize the training process, we resized the images to a resolution of 400 x 300 pixels, which helped reduce training time. We also manually filtered out images that lacked any identifiable objects, leaving us with a final dataset of 10,000 image patches. The dataset was preprocessed and structured using Python to align with standard model architectures.
We also used the VEDAI (Razakarivony et al. 2016) and VASAI (Wang et al. 2022) benchmark dataset for fair comparison.

4 Method Used

4.1 Enhanced Super-YOLO

In figure 1, the block diagram of our modified Enhanced Super-YOLO is shown. It consists of a Backbone for feature information and extraction, a Detector to calculate bounding box and regression loss, and a Super Resolution pipeline to recreate the original input. We enhanced the existing Super-YOLO (Zhang et al. 2023) with improvement in the super-resolution pipeline which used EDSR network (Lim et al. 2017). Densely Residual Laplacian Network (DRLN) (Anwar et al. 2022) based network is used to improve the performance of super-resolution pipeline which further improves the feature representation of the backbone. Path Aggregation Network (PANet) for fast image segmentation and Feature Pyramid Networks (FPN) for generic feature extraction are utilized

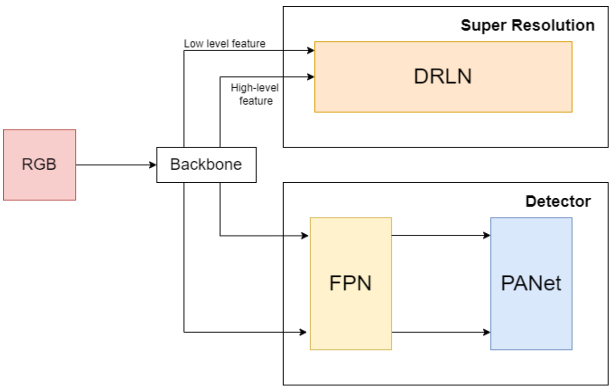


Figure 1. Enhanced Super-YOLO framework

4.2 Loss Function

In object detection, the loss function combines localization loss (e.g., Smooth L1, GIoU, CIoU) to assess bounding box accuracy and classification loss (e.g., Cross-Entropy, Focal Loss) to evaluate object classification, with Focal Loss addressing the class imbalance. By optimizing these combined loss functions, the model improves its ability to detect objects more accurately, both in terms of their positions and classifications. The overall loss of our network comprises of two components: the detection loss L_o and super resolution reconstruction loss L_s represented as:

$$L_{total} = c_1 L_o + c_2 L_s$$

In this framework, c_1 and c_2 serve as coefficients that balance the two training tasks. For calculating the super resolution reconstruction loss L_s is defined as the difference between input image X and the super-resolved output S, we utilize L1 loss instead of L2 loss. This is expressed mathematically as:

$$L_s = \|S - X\|_1$$

Detection loss consists of three key components: the loss associated with object presence L_{obj}, the loss related to object localization L_{loc}, and the loss for object classification L_{cls}. The overall detection loss L_o can be represented as follows:

$$L_o = \lambda_{loc}\sum_{l=0}^{2} + a_l L_{loc} + \lambda_{obj}\sum_{l=0}^{2} + a_l L_{loc} + \lambda_{cls}\sum_{l=0}^{2} + c_1 L_{cls}$$

In this equation, *l* denotes the output layer in the network's head, while a_l, b_l, and c_l represent the weights assigned to the different layers for each of the three loss functions. The weights λ_{loc}, λ_{obj}, and λ_{cls} help regulate the emphasis placed on errors related to bounding box coordinates, dimensions, object presence, absence, and classification accuracy.
In figure 2, observations at lower altitudes, the images exhibited a more detailed and intimate view of the landscape. Fine textures and small-scale features were prominently visible, providing a closer connection to the subject matter.

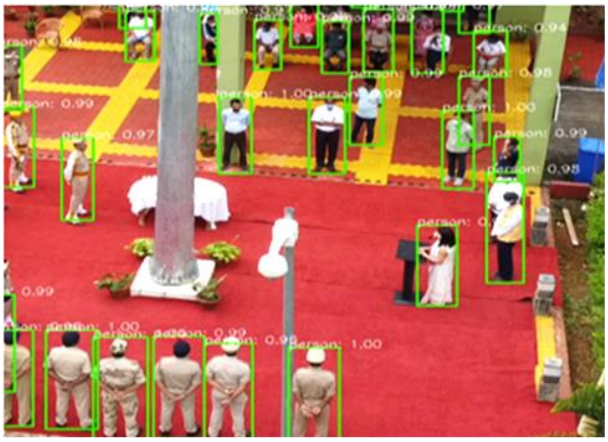


Figure 2. Low Height Fly

In figure 3, observations at medium altitude flights struck a balance between detail and overall landscape composition. The images captured from this height showcased a broader perspective while retaining sufficient detail, making it an ideal choice for a variety of applications.

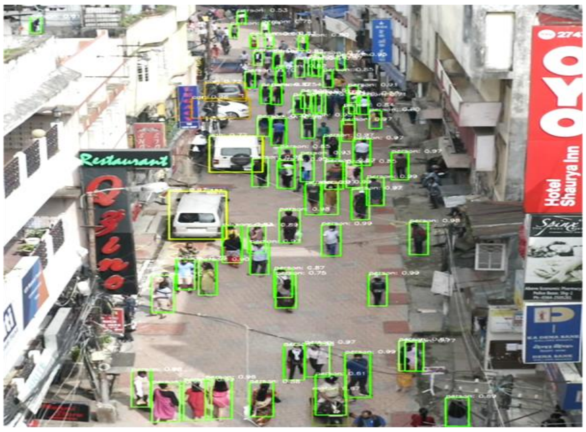


Figure 3. Medium Height Fly

In figure 4, observations at high-altitude flights revealed a sweeping, panoramic view of the landscape. While details at the ground level were less pronounced, the images captured from this height provided a unique and comprehensive overview of the area, highlighting large-scale patterns and topographical feature

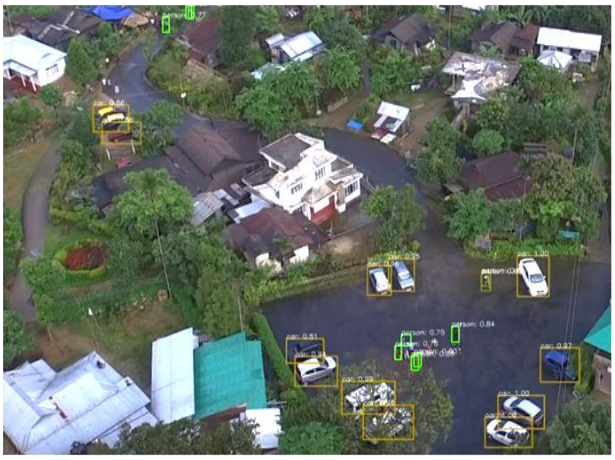
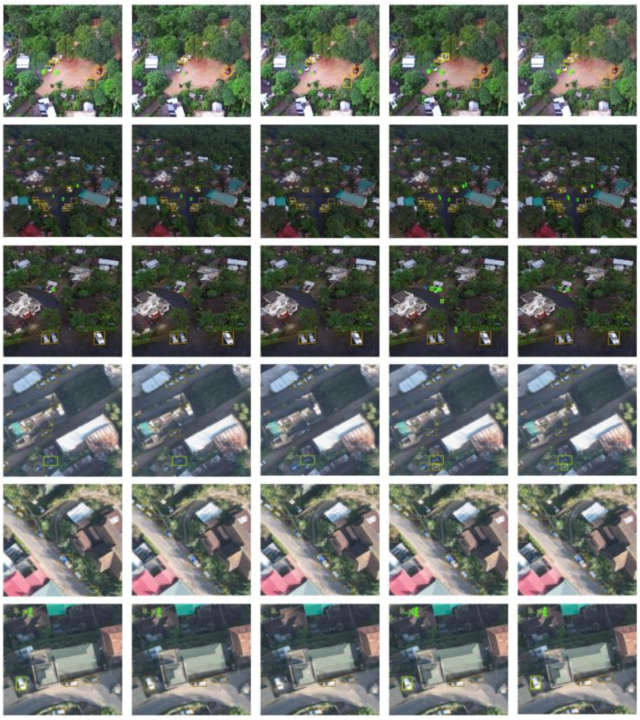


Figure 4: High Height Fly

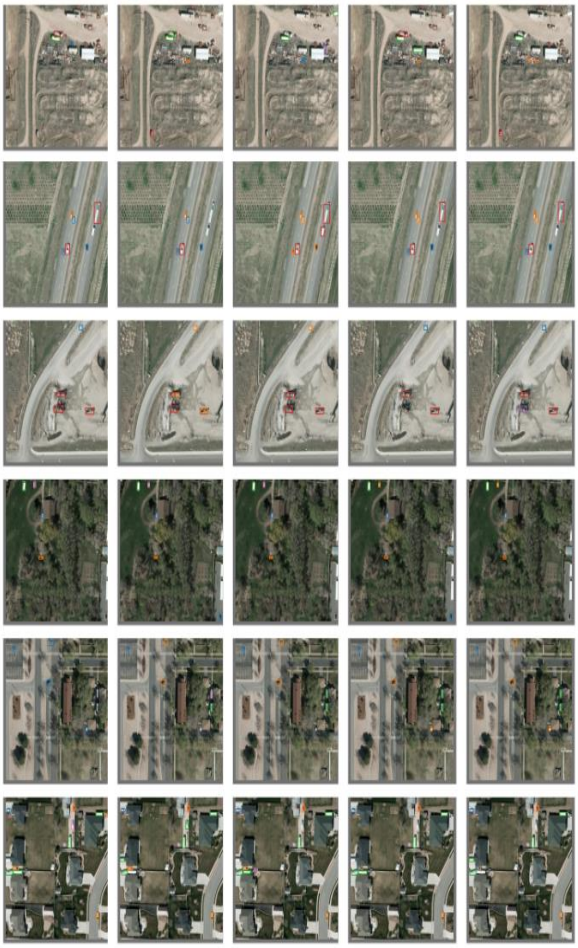
5 Results

In figure 5, From left to right, each column represents: [a] ground truth and predicted results [b] SuperYOLO, [c] ShuffleNet, [d] YOLOrs, and [e] Our model. Vehicles are accurately detected in the second and fourth input rows, while missed detections occur in the first input [c], third input [c], fifth input [c], and the last inputs [b] & [c]. Person detection is largely missed across most inputs, but overall, vehicles and persons are accurately detected in [e]. In figure 6, From left to right, each column represents: [a] ground truth, predicted results [b] SuperYOLO, [c] ShuffleNet, [d] YOLOrs, and [e] Our model. Vehicles, such as cars and pickups, are accurately detected in the third and sixth input rows, with missed detections occurring in the first input [c] and [d], as well as the fifth input [a]. Overall, vehicles like cars and pickups, along with other classes, are accurately detected in [e].



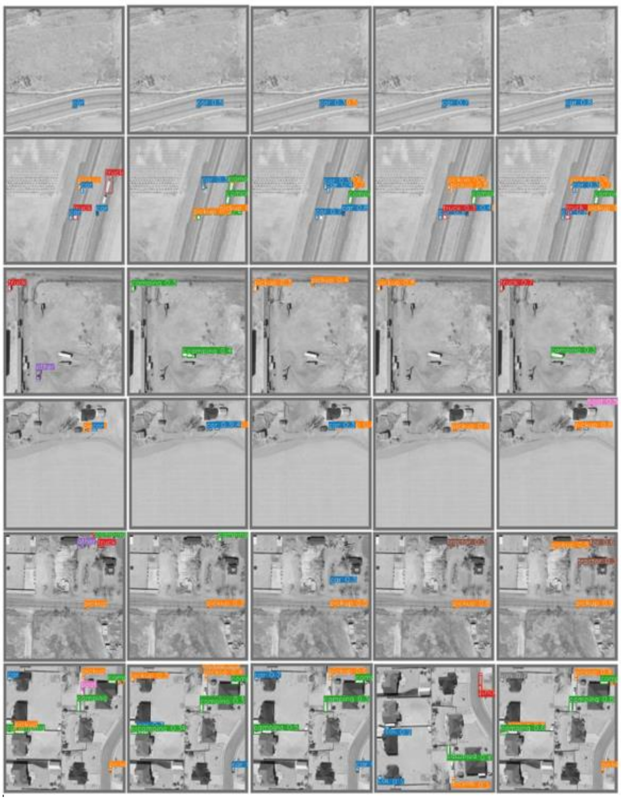
(a) (b) (c) (d) (e)

Figure 5. Predicted results of the proposed model on UAV dataset. Detections labeled with green rectangles are persons and those with orange rectangles are vehicle.



In figure. 7, From left to right, each column represents: [a] ground truth, predicted results, [b] SuperYOLO, [c] ShuffleNet, [d] YOLOrs, and [e] Our model. In fifth row vehicles, such as cars and truck, are accurately detected in [d][e] and missed on [b][c]. In forth row, truck is detected only on [e] and missed on all other. Overall, vehicles like cars and pickups, along with other classes, are accurately detected in [e]. figure. 8, presents the visual outcomes of our model applied to various scenes and respective ground truth from the VSAI (Wang et al. 2022) dataset. It can able to detect small and large vehicles captured using drones precisely.

(a) Ground truth



(a) (b) (c) (d) (e)

Figure 7. Comparison results of various models on VEDAI IR (Razakarivony et al. 2016) dataset.

(b) Our model results

Figure 8. Visualization result produced by Our method on VSAI (Wang et al. 2022) dataset

6 Accuracy & Analysis

The accuracy assessment evaluates the alignment between detection results and the reference mask, using precision, recall, and mean Average Precision (mAP) as key metrics, where precision measures the proportion of true positive detections among all positive detections.

Precision and Recall are calculated as,

$$\text{Precision} = \frac{TP}{TP + FP} \ \& \ \text{Recall} = \frac{TP}{TP + FN}$$

True Positives (TP) refer to objects that are correctly identified, whereas False Positives (FP) denote incorrectly identified objects. False Negatives (FN) indicate objects that were not detected. Precision and recall together offer insights into detection accuracy, contributing to the mean Average Precision (mAP) for overall model performance across classes and thresholds.

$$\text{mean Average Precision} = \frac{AP}{N} = \frac{\int_0^1 \blacksquare p(r)dr}{N}$$

Here, P is precision, r is recall, and N the number of categories.

In table 1, we observe the performance of our model across various classes on the VEDAI RGB dataset (Razakarivony et al. 2016). The overall results indicate strong performance in detecting common vehicle types like cars and pickups, while less common categories such as boats and vans show comparatively lower metrics

In table 2, we observe a comparison of our model's performance with other models on the VEDAI RGB dataset (Razakarivony et al. 2016). Our model here demonstrates a balanced performance across all metrics, surpassing SuperYOLO (Zhang et al. 2023) and ShuffleNet (Zhang et al. 2018) in mAP@50 and achieving higher overall precision compared to YOLOrs (Sharma et al. 2021) and ShuffleNet (Zhang et al. 2018).

In table 3, we observe a comparison of our model's performance with other models on the VEDAI IR dataset (Razakarivony et al. 2016). The car class has the highest precision and mAP, indicating strong performance. However, the van class shows significantly low recall, suggesting poor detection consistency for that category.

In table 4, we observe a comparison of our model's performance with other models on the NESAC dataset. Our model here outperforms SuperYOLO (Zhang et al. 2023), ShuffleNet (Zhang et al. 2018), and YOLOrs (Sharma et al. 2021) in all metrics, demonstrating higher precision and recall, as well as better mAP, indicating more accurate detection and classification on the NESAC dataset.

Table 1. Classwise Precision (P), Recall (R), and Mean Average Precision (mAP) for Our model on the VEDAI (Razakarivony et al. 2016) RGB dataset

Class	Images	Labels	P	R	mAP@50	mAP@5:95
all	121	364	72.37	67.33	74.84	45.11
car	121	134	86.18	82.09	89.56	58.23
pickup	121	95	79.74	78.73	84.37	55.52
camping	121	39	63.61	76.22	65.87	45.55
truck	121	30	83.0	70.0	83.52	50.85
other	121	20	61.33	50.00	65.47	29.80
tractor	121	19	86.31	78.95	87.48	44.15
boat	121	17	68.94	52.94	60.15	35.58
van	121	10	49.82	49.68	62.26	41.20

Table 2. Comparison of SuperYOLO, ShuffleNet, and YOLOrs with Our model on the VEDAI RGB dataset

Model	P	R	mAP@50	mAP@5:95
Our	72.37	67.33	74.84	45.11
SuperYOLO	79.21	57.19	69.00	40.51
ShuffleNet	68.29	58.92	64.4	37.58
YOLOrs	55.56	58.25	57.51	33.67

Table 3. Classwise Precision (P), Recall (R), and Mean Average Precision (mAP) for Our model on the VEDAI IR dataset

Class	Images	Labels	P	R	mAP@50	mAP@5:95
all	121	364	63.67	51.21	56.83	34.22
car	121	134	91.92	67.94	83.01	53.45
pickup	121	95	69.43	68.42	75.31	46.94
camping	121	39	63.36	58.97	64.65	41.31
truck	121	30	56.65	56.67	51.08	30.09
other	121	20	57.92	25	31.56	12.69
tractor	121	19	54.35	47.37	43.87	20.55
boat	121	17	58.7	35.29	43.15	27.07
van	121	10	57.04	5	61.97	41.65

Table 4. Comparison of SuperYOLO, ShuffleNet, and YOLOrs with Our model on the NESAC dataset

Model	P	R	mAP@50
Our	61.06	56.11	61.06
SuperYOLO	59.92	55.22	56.31
ShuffleNet	56.21	51.24	53.56
YOLOrs	54.43	50.15	51.87

7 Conclusion

Real-time applications have witnessed a significant transformation due to deep learning-based object detection. In sectors like autonomous vehicles, surveillance, and medical imaging, these models play a crucial role in enabling real-time decision-making. The ability to process information quickly and accurately is vital for ensuring the safety and efficiency of systems relying on object detection. We showcased the application of an enhanced object detection algorithm for aerial object detection. As research continues, addressing challenges and refining methodologies, the future holds promise for even more sophisticated, interpretable, and efficient deep learning models in object detection.

References

Aswini N. and U. S V (2021) "Drone Object Detection Using Deep Learning Algorithms," 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, pp. 1187-1192, doi: 10.1109/ICIRCA51532.2021.9544983.

Albaba B. M. and S. Ozer (2021) "SyNet: An Ensemble Network for Object Detection in UAV Images," 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, pp. 10227-10234, doi: 10.1109/ICPR48806.2021.9412847.

Anwar S. and N. Barnes (2022) "Densely Residual Laplacian Super-Resolution," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 3, pp. 1192-1204, doi: 10.1109/TPAMI.2020.3021088.

Bazi Y. and F. Melgani (2018), "Convolutional SVM Networks for Object Detection in UAV Imagery," in IEEE Transactions on Geoscience and Remote Sensing, vol. 56, no. 6, pp. 3107-3118, doi: 10.1109/TGRS.2018.2790926

Bejiga M. B. A. Zeggada and F. Melgani (2016) "Convolutional neural networks for near real-time object detection from UAV imagery in avalanche search and rescue operations," IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, pp. 693-696, doi: 10.1109/IGARSS.2016.7729174.

Hosseinpoor H. and F. Samadzadegan (2020) "Convolutional Neural Network for Building Extraction from High-Resolution Remote Sensing Images," International Conference on Machine Vision and Image Processing (MVIP), Iran, pp. 1-5, doi: 10.1109/MVIP49855.2020.9187483.

Li C., B. Zhang, D. Hong, J. Yao and J. Chanussot, (2023) "LRR-Net: An Interpretable Deep Unfolding Network for Hyperspectral Anomaly Detection," in IEEE Transactions on Geoscience and Remote Sensing, vol. 61, pp. 1-12, Art no. 5513412, doi: 10.1109/TGRS.2023.3279834

Lim B., S. Son, H. Kim, S. Nah, and K. Mu Lee (2017) “Enhanced deep residual networks for single image super-resolution,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), pp. 136–144.

Liu Y., F. Yang and P. Hu (2020) "Small-Object Detection in UAV-Captured Images via Multi-Branch Parallel Feature Pyramid Networks," in IEEE Access, vol. 8, pp. 145740-145750, doi: 10.1109/ACCESS.2020.3014910.

Maharjan N., H. Miyazaki, B. M. Pati, M. N. Dailey, S. Shrestha and T. Nakamura (2022) "Detection of River Plastic Using UAV Sensor Data and Deep Learning" Remote Sensing 14, no. 13: 3049. https://doi.org/10.3390/rs14133049

Meng L., Z. Peng, J. Zhou, J. Zhang, Z. Lu, A. Baumann and Y. Du (2020) "Real-Time Detection of Ground Objects Based on Unmanned Aerial Vehicle Remote Sensing with Deep Learning: Application in Excavator Detection for Pipeline Safety" Remote Sensing 12, no. 1: 182. https://doi.org/10.3390/rs12010182

Micheal A. A., K. Vani (2022) Deep Learning-Based Multi-class Multiple Object Tracking in UAV Video. J Indian Soc Remote Sens 50, 2543–2552. https://doi.org/10.1007/s12524-022-01615-7

Mittal P., R. Singh and A. Sharma (2020) "Deep learning-based object detection in low-altitude UAV datasets: A survey." Image and Vision computing 104, 104046.

Papageorgiou C. P., M. Oren and T. Poggio (1998) "A general framework for object detection," Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271), Bombay, Indiam, pp. 555-562, doi: 10.1109/ICCV.1998.710772.

Pathak A. R., M. Pandey and S. Rautaray (2018) “Application of Deep Learning for Object Detection”, Procedia Computer Science, Volume 132, Pages 1706-1717, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2018.05.144

Razakarivony S. and F. Jurie (2016) “Vehicle detection in aerial imagery: A small target detection benchmark,” J. Vis. Commun. Image Represent., vol. 34, pp. 187–203, 2016

Saetchnikov I. V., E. A. Tcherniavskaia and V. V. Skakun (2021) "Object Detection for Unmanned Aerial Vehicle Camera via Convolutional Neural Networks," in IEEE Journal on Miniaturization for Air and Space Systems, vol. 2, no. 2, pp. 98-103, doi: 10.1109/JMASS.2020.3040976

Sharma M. et al., (2021) “YOLOrs: Object detection in multimodal remote sensing imagery,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 14, pp. 1497–1508

Sun W., L. Dai, X. Zhang et al. (2022) RSOD: Real-time small object detection algorithm in UAV-based traffic monitoring. Appl Intell 52, 8448–8463. https://doi.org/10.1007/s10489-021-02893-3

Wang J., X. Teng, Z. Li, Q. Yu, Y. Bian and J. Wei (2022) "VSAI: A Multi-View Dataset for Vehicle Detection in Complex Scenarios Using Aerial Images" Drones 6, no. 7: 161. https://doi.org/10.3390/drones6070161

Wu X., D. Hong and J. Chanussot (2022) "Convolutional Neural Networks for Multimodal Remote Sensing Data Classification," in IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-10, Art no. 5517010, doi: 10.1109/TGRS.2021.3124913

Wu X., D. Hong and J. Chanussot (2023) "UIU-Net: U-Net in U-Net for Infrared Small Object Detection," in IEEE Transactions on Image Processing, vol. 32, pp. 364-376, doi: 10.1109/TIP.2022.3228497

Wu X., W. Li, D. Hong, R. Tao and Q. Du (2021). Deep Learning for UAV-based Object Detection and Tracking:

A Survey. ArXiv. https://doi.org/10.1109/MGRS.2021.3115137

Xiao J., S. Zhang, Y. Dai, Z. Jiang, B. Yi and C. Xu (2020) "Multiclass Object Detection in UAV Images Based on Rotation Region Network," in IEEE Journal on Miniaturization for Air and Space Systems, vol. 1, no. 3, pp. 188-196, doi: 10.1109/JMASS.2020.3025970.

Zeng Y., T. Zhang, W. He and Z. Zhang (2023) "YOLOv7-UAV: An Unmanned Aerial Vehicle Image Object Detection Algorithm Based on Improved YOLOv7" Electronics 12, no. 14: 3141. https://doi.org/10.3390/electronics12143141

Zhang J., J. Lei, W. Xie, Z. Fang, Y. Li and Q. Du (2023) "SuperYOLO: Super Resolution Assisted Object Detection in Multimodal Remote Sensing Imagery," in IEEE Transactions on Geoscience and Remote Sensing, vol. 61, pp. 1-15, Art no. 5605415, doi: 10.1109/TGRS.2023.3258666.

Zhang X., X. Zhou, M. Lin and J. Sun (2018) “ShuffleNet: An extremely efficient convolutional neural network for mobile devices,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit, pp. 6848–6856.

Zhang R., H. Li, K. Duan, S. You, K. Liu, F. Wang and Y. Hu. (2020). "Automatic Detection of Earthquake-Damaged Buildings by Integrating UAV Oblique Photography and Infrared Thermal Imaging" Remote Sensing 12, no. 16: 2621. https://doi.org/10.3390/rs12162621

Zhou W., X. Min, R. Hu, Y. Long and H. Luo (2022) FasterX: Real-Time Object Detection Based on Edge GPUs for UAV Applications. ArXiv. /abs/2209.03157