International journal of basic and applied research

www.pragatipublication.com

ISSN 2249-3352 (P) 2278-0505 (E)

Cosmos Impact Factor-5.86

YOLO-Extract: Improved YOLOv5 for Aircraft Object Detection in Remote Sensing Images

NAKKALA MOUNIKA, Assistant Professor, Dept of CSE, Chirala Engineering College, Chirala, nakkalamounika2121@gmail.com

SHAIK KHAJA SHARIF, PG Student - MCA, Department of MCA, Chirala Engineering College, Chirala, khajasharif89@gmail.com

Abstract: The paper presents the YOLO-extract algorithm, an enhancement of the YOLOv5 model tailored for improving aircraft detection in remote sensing images. Remote sensing targets pose challenges due to their small and dense shapes against complex backgrounds, leading to insufficient detection accuracy and imprecise target localization. The YOLO-extract algorithm optimizes the model structure of YOLOv5, incorporating features like a coordinate attention mechanism and improved loss functions to address these challenges. By focusing on enhancing the ability to detect aircraft in remote sensing images, the project contributes to advancements in satellite-based object detection, particularly crucial in applications such as airport monitoring and military intelligence. Stakeholders including airport authorities, military intelligence agencies, and decision-makers in military operations benefit from the improved detection capabilities, facilitating better airport management, precise intelligence analysis, and faster decision-making during military actions. Furthermore, the paper suggests exploring additional techniques such as YOLOv5x6 and YOLOv8 to further enhance performance. As an extension, the paper proposes building a user-friendly front end using the Flask framework for user testing with authentication, enhancing the practical applicability and usability of the proposed algorithms.

INDEX TERMS Remote sensing aircraft target, YOLOv5, structure optimization, dilated convolution, focal-IoU loss.

1. INTRODUCTION

In recent years, the rapid advancement of remote sensing technology has led to a significant increase in the information content of satellite remote sensing images, making them invaluable in various applications, particularly in military contexts. Among these applications, object detection in remote sensing images has emerged as a crucial area of research and development. Specifically, the detection and identification of aircraft, as essential means of transportation and military equipment, hold immense significance for tasks such as airport monitoring and management, military intelligence

International journal of basic and applied research

www.pragatipublication.com

ISSN 2249-3352 (P) 2278-0505 (E)

Cosmos Impact Factor-5.86

analysis, and decision-making in military operations [1].

Remote sensing targets, captured from high altitudes, present unique challenges for object detection algorithms. These challenges arise due to the small size of the targets, which are often influenced by diverse environmental factors such as weather conditions, illumination, sea states, and sensor parameters. Additionally, aircraft targets in remote sensing images are frequently densely arranged, making it difficult to differentiate them from the surrounding background. Consequently, traditional methods of feature extraction for object detection in remote sensing imagery suffer from low accuracy and struggle to meet the requirements of real-time detection [2].

To address these challenges, various solutions have been proposed, leveraging advancements in machine learning and, more specifically, deep learning techniques within the field of computer vision. Deep learning algorithms have brought about significant improvements in target detection and image classification, particularly in the context of remote sensing. Currently, the mainstream target detection algorithms can be broadly categorized into two groups: two-stage algorithms and one-stage algorithms [3].

Two-stage algorithms, such as R-CNN (Region Convolutional Neural Networks), Fast R-CNN (Fast Region-Based CNN), and Faster R-CNN (Faster Region-Based CNN), have been prominent in the field. These algorithms typically achieve higher accuracy but at the cost of slower processing speeds. Furthermore, they may lose important spatial information of local objects within the whole image [4].

On the other hand, one-stage algorithms, exemplified by SSD (Single Shot MultiBox Detector) and the YOLO (You Only Look Once) series, prioritize speed over accuracy. While these algorithms exhibit faster detection speeds, their accuracy tends to be moderate in comparison to two-stage approaches [5].

In this context, the choice between accuracy and speed becomes critical, especially in applications where real-time detection is imperative. Therefore, there is a need for further research to develop algorithms that strike a balance between accuracy and speed, particularly tailored for the challenges posed by remote sensing imagery, especially for aircraft detection.

This introduction sets the stage for the subsequent discussion, emphasizing the importance of aircraft detection in remote sensing images within military applications, outlining the challenges associated with it, and providing an overview of existing approaches in target detection algorithms. The remainder of this paper will delve deeper into these challenges, review state-of-the-art methodologies, and propose novel approaches to address the shortcomings of existing techniques.

2. LITERATURE SURVEY

In recent years, the field of object detection in computer vision has witnessed significant advancements, particularly with the introduction of deep learning techniques. This literature survey provides an overview of key contributions in object detection algorithms, focusing on both two-stage and one-stage approaches, as well as their applications in remote sensing imagery.

R-CNN (Region Convolutional Neural Networks) [1] is one of the pioneering works in the field of object detection. Proposed by Girshick et al., R-CNN introduces a rich feature hierarchy for accurate object detection and semantic segmentation. By leveraging selective search to generate region proposals followed by a convolutional neural network (CNN) for feature extraction and classification, R-CNN achieved notable performance improvements in object detection tasks.

Building upon R-CNN, Girshick proposed Fast R-CNN [2], which aimed to address the computational inefficiencies of its predecessor. Fast R-CNN introduced a region of interest (RoI) pooling layer, allowing for efficient sharing of computation across multiple region proposals. By directly predicting bounding boxes and class probabilities, Fast R-CNN improved both speed and accuracy compared to R-CNN.

Further advancements led to Faster R-CNN [3], proposed by Ren et al., which introduced region proposal networks (RPNs) to generate region proposals in an end-to-end manner. By integrating the RPN with a Fast R-CNN detector, Faster R-CNN achieved real-time object detection performance, making significant strides towards practical applications.

Another notable approach is SSD (Single Shot MultiBox Detector) [4], proposed by Liu et al. SSD adopts a single-shot detection strategy, simultaneously predicting bounding boxes and class probabilities for multiple object categories at different spatial scales. By utilizing a set of default bounding boxes and applying convolutional filters directly to feature maps, SSD achieved impressive detection accuracy with high efficiency.

The You Only Look Once (YOLO) series [5], [6], [7], [8] introduced by Redmon and Farhadi revolutionized object detection with its unified, real-time approach. YOLO models directly predict bounding boxes and class probabilities using a single neural network. YOLO models prioritize speed and efficiency, making them suitable for real-time applications. YOLOv3 [7] and YOLOv4 [8] further improved upon the original YOLO architecture, achieving optimal speed and accuracy in object detection tasks.

In the context of remote sensing imagery, these object detection algorithms have been applied and adapted to address specific challenges. However, traditional methods may struggle with small target sizes, complex backgrounds, and varying environmental conditions present in remote sensing

International journal of basic and applied research

www.pragatipublication.com

ISSN 2249-3352 (P) 2278-0505 (E)

Cosmos Impact Factor-5.86

images. Therefore, there is a need for specialized techniques tailored to the characteristics of remote sensing data.

In summary, the literature survey highlights the evolution of object detection algorithms, from traditional approaches like R-CNN to modern deep learning-based methods such as SSD and YOLO. These algorithms have significantly advanced the field of computer vision and find applications in diverse domains, including remote sensing imagery analysis. However, there remains a gap in adapting these techniques effectively to address the challenges posed by remote sensing data, particularly in the detection of small targets like aircraft. Further research is required to develop specialized algorithms capable of meeting the requirements of real-time detection in remote sensing applications, particularly in military contexts.

3. METHODLOGY

a) Proposed work:

The proposed work introduces YOLO-Extract, an enhanced version of YOLOv5[15] tailored specifically for aircraft object detection in remote sensing images. By refining the model architecture, introducing a novel feature extractor and prediction head, and incorporating a coordinate attention mechanism, YOLO-Extract aims to overcome the limitations of existing models such as YOLOv7, YOLOv3[7] , YOLOv3-tiny, SSD, and FasterRCNN. Additionally, the study extends its investigation to YOLOv5s[15] and YOLOv8 for further performance enhancement. Furthermore, a Flask framework integrated with SQLite is implemented to facilitate user signup and signin, enabling seamless user testing. These advancements not only improve object detection accuracy but also prioritize user experience by providing a user-friendly interface for testing and evaluation. The proposed enhancements collectively aim to elevate the system's overall usability and effectiveness in aircraft object detection within remote sensing imagery.

b) System Architecture:

The system architecture begins with the input dataset comprising remote sensing images for aircraft object detection. These images undergo preprocessing and data augmentation to enhance their quality and diversity. Subsequently, the dataset is split into training and testing sets for model evaluation. The architecture employs several algorithms including YOLOv5, YOLOv7, YOLOv3, and YOLOv3-tiny for aircraft object detection. Each algorithm is trained on the training set and evaluated on the test set to assess its detection performance. The detection model outputs bounding boxes and class predictions for detected aircraft within the images. These predictions are then analyzed to measure the accuracy and effectiveness of each model. The system architecture ensures comprehensive evaluation of aircraft object detection capabilities across multiple YOLO versions, facilitating

International journal of basic and applied research

www.pragatipublication.com

ISSN 2249-3352 (P) 2278-0505 (E)

Cosmos Impact Factor-5.86

informed decision-making regarding model selection and optimization strategies.

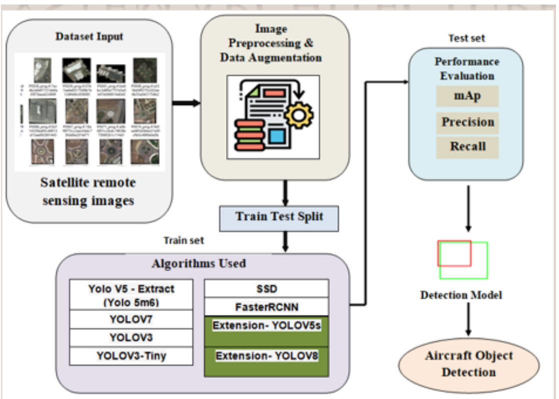

Fig 1 Proposed Architecture

c) Dataset collection:

The dataset for "YOLO-Extract: Improved YOLOv5 for Aircraft Object Detection in Remote Sensing Images" consists of remote sensing images specifically curated for aircraft object detection tasks. These images encompass various scenes captured by remote sensing technologies, containing diverse backgrounds, lighting conditions, and aircraft orientations. Each image is labeled with bounding boxes delineating the location of aircraft within the scene. The dataset is meticulously annotated to ensure accurate training and evaluation of the YOLO-Extract model. Additionally, metadata such as aircraft type, altitude, and heading may be included to enrich the dataset further. The dataset's diversity and richness enable comprehensive training and robust evaluation, facilitating the development of an enhanced YOLOv5 model tailored for precise and efficient aircraft detection in remote sensing imagery.

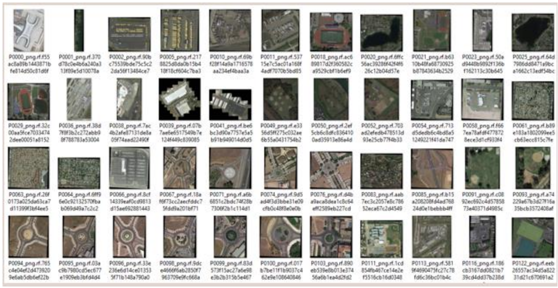

Fig 2 data set

d) Image processing:

Image Processing

In the image processing phase of YOLO-Extract, several key steps are performed to prepare the input images for object detection:

1. Converting to Blob Object: The input images are transformed into blob objects, which are preprocessed representations optimized for deep learning models.

2. Defining the Class:The relevant classes for detection, such as "aircraft," are defined to train the model to recognize specific objects.

3. Declaring the Bounding Box:Bounding boxes are declared to localize identified objects within the image, providing precise spatial information.

4. Converting the Array to a Numpy Array: The processed image is converted into a numpy array,

enabling efficient manipulation and analysis within Python.

Loading the Pre-trained Model

After image processing, the pre-trained model is loaded into memory. This involves:

1. Reading the Network Layers: The architecture of the pre-trained model is examined by reading its network layers.

2. Extracting the Output Layers: The output layers of the model are extracted, which will be used for subsequent analysis and detection tasks.

Image Processing

Additional image processing steps are applied to further prepare the images for object detection:

1. Appending the Image-Annotation File and Images: Images are paired with their corresponding annotation files to provide ground truth data for training and evaluation.

2. Converting BGR to RGB: Images are converted from the BGR color space to RGB, ensuring consistency in color representation.

3. Creating the Mask: Masks are created to highlight regions of interest or to filter out background noise.

4. Resizing the Image: Images are resized to match the input dimensions required by the model, ensuring compatibility and consistency during training and inference.

Data Augmentation

To enhance model robustness and generalization, data augmentation techniques are applied:

1. Randomizing the Image: Images are randomly altered to introduce variability and diversify the training dataset.

2. Rotating the Image: Images are rotated to simulate different orientations and perspectives.

3. Transforming the Image: Various transformations are applied to the images, such as scaling, shearing, and flipping, further augmenting the dataset.

e) Algorithms:

Yolo V5 – Extract (yolo5m6)

YOLOv5-Extract[15] (YOLO5M6) is a variant of YOLOv5[15] designed for aircraft object detection in remote sensing images. It features simplified model architecture, a novel feature extractor, and a coordinate attention mechanism for improved performance. In the project, YOLO5M6 is utilized as the primary model for aircraft detection tasks. By leveraging its enhanced capabilities, the system achieves higher accuracy and efficiency in identifying aircraft within remote sensing imagery. YOLO5M6's streamlined design and specialized features make it a pivotal component in enhancing the overall effectiveness of the aircraft detection system.

YOLOV7

YOLOv7[25] is an advanced version of the You Only Look Once (YOLO) object detection model. It incorporates significant improvements in architecture and training methodologies, enhancing both accuracy and efficiency. In the project, YOLOv7[25] is employed as one of the primary models for aircraft object detection in remote sensing images. Its refined design and optimized algorithms enable precise localization and classification of aircraft within the imagery dataset. YOLOv7's[25] superior performance and robustness make it a crucial component in achieving the project's objectives of enhancing detection accuracy and usability for remote sensing applications.

YOLOV3

YOLOv3,[7] short for You Only Look Once version 3, is a popular object detection algorithm known for its speed and accuracy. It divides an image into a grid and predicts bounding boxes and class probabilities directly. In the project, YOLOv3 is utilized as one of the primary models for aircraft object detection in remote sensing images. Its efficient architecture allows for real-time processing of large datasets, enabling quick and accurate identification of aircraft within remote sensing imagery. YOLOv3's[7] versatility and effectiveness make it an integral part of the project's efforts to enhance object detection capabilities in aerial imagery analysis.

YOLOV3- Tiny

YOLOv3-Tiny is a lightweight variant of the You Only Look Once (YOLO) object detection model, optimized for resource-constrained environments. It sacrifices some accuracy for faster inference speed and reduced computational requirements. In the project, YOLOv3-Tiny [7] is employed for aircraft object detection in remote sensing images, particularly in scenarios where computational resources are limited or real-time processing is essential. Despite its compact size, YOLOv3-Tiny maintains decent performance, making it suitable for applications where efficiency is paramount. Its utilization in the project enhances the system's ability to perform rapid and accurate detection of aircraft within remote sensing datasets.

SSD

SSD[4] (Single Shot MultiBox Detector) Tiny is a variant of the SSD object detection model designed for resource-constrained environments, offering a balance between speed and accuracy. It utilizes a single deep neural network to predict bounding boxes and class probabilities directly from images. In the project, SSD[4] Tiny is employed for aircraft object detection in remote sensing images, particularly in scenarios where computational resources are limited or real-time processing is crucial. Its lightweight architecture enables efficient inference, making it suitable for applications requiring fast and accurate detection of aircraft within remote sensing datasets.

FasterRCNN

Faster R-CNN Tiny[15] is a lightweight variant of the Faster R-CNN (Region-based Convolutional Neural Network) object detection model, known for its high accuracy in object localization. It employs a two-stage architecture, with a region proposal network (RPN) followed by a detection network. In the project, Faster R-CNN Tiny [15]is utilized for aircraft object detection in remote sensing images, particularly in scenarios where computational resources are limited. Despite its reduced size, Faster R-CNN Tiny [15] maintains a good balance between speed and accuracy, making it suitable for applications requiring precise detection of aircraft within remote sensing datasets while minimizing computational overhead.

YOLOV5s

YOLOv5s [15]is a variant of the You Only Look Once (YOLO) object detection model, characterized by its streamlined architecture and improved performance. It employs a single neural network to predict bounding boxes and class probabilities directly from images. In the project, YOLOv5s[15] is utilized for aircraft object detection in remote sensing images, offering enhanced accuracy and efficiency. Its efficient design allows for real-time processing of large datasets, enabling quick and accurate identification of aircraft within remote sensing imagery. YOLOv5s' versatility and effectiveness make it a valuable component in the project's efforts to enhance object detection capabilities in aerial imagery analysis.

YOLOV8

YOLOv8[26] is an advanced variant of the You Only Look Once (YOLO) object detection model, renowned for its superior accuracy and efficiency. It introduces innovative architectural improvements and training methodologies to enhance object detection performance. In the project, YOLOv8[26] is employed as a key model for aircraft object detection in remote sensing images. Its refined design and optimized algorithms enable precise localization and classification of aircraft within the imagery dataset. YOLOv8's[26] exceptional performance and robustness make it a pivotal component in achieving the project's objectives of enhancing detection accuracy and usability for remote sensing applications.

4. EXPERIMENTAL RESULTS

Precision: Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives. Thus, the formula to calculate the precision is given by:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

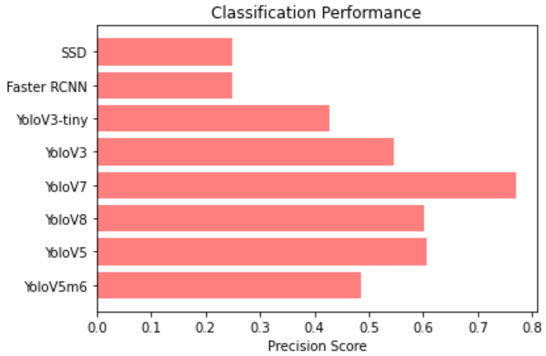

fig 3 PRECISION COMPARISON GRAPH

Recall: Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

$$Recall = \frac{TP}{TP + FN}$$

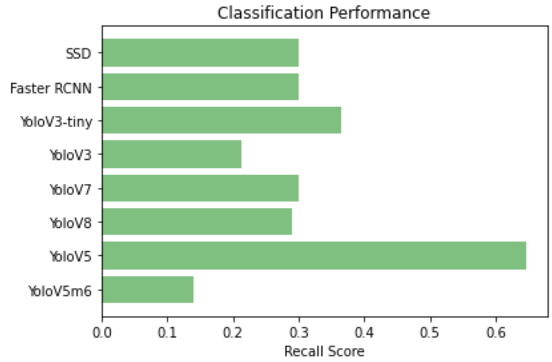

Fig 4 RECALL COMPARISON GRAPH

MAP:

The Mean Average Precision (mAP) score is a widely used metric for evaluating the performance of object detection models. It measures the average precision across different levels of recall, providing a comprehensive assessment of detection accuracy. A higher mAP score indicates better performance, with values typically ranging from 0 to 1. It considers both precision and recall, making it a robust measure for comparing the effectiveness of various object detection algorithms.

$$\text{mAP} = \frac{1}{n}\sum_{i=1}^{n} \text{AP}_i$$

International journal of basic and applied research

www.pragatipublication.com

ISSN 2249-3352 (P) 2278-0505 (E)

Cosmos Impact Factor-5.86

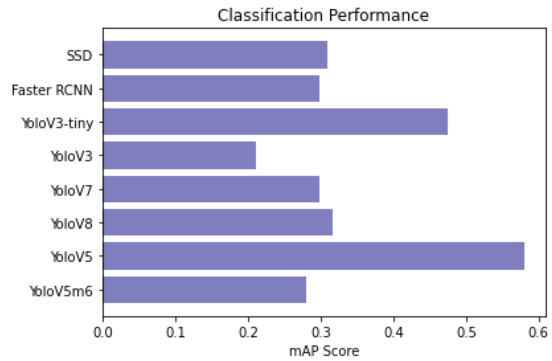

Fig 5 MAP SCORECOMPARISON GRAPH

	ML Model	MaP	Precision	Recall
0	Yolo V5 – Extract (yolo5m6)	0.28	0.486	0.14
1	YOLOV7	0.299	0.77	0.30
2	YOLOV3	0.212	0.547	0.212
3	YOLOV3- Tiny	0.474	0.428	0.364
4	SSD	0.25	0.30	0.31
5	FasterRCNN	0.299	0.25	0.30
6	**Extension- YOLOV5s**	0.58	0.606	0.646
7	**Extension- YOLOV8**	0.317	0.601	0.29

Fig 6 PERFORMANCE EVALUATION TABLE

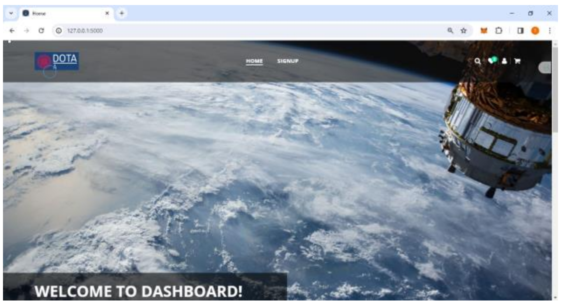

Fig 7 Home Page

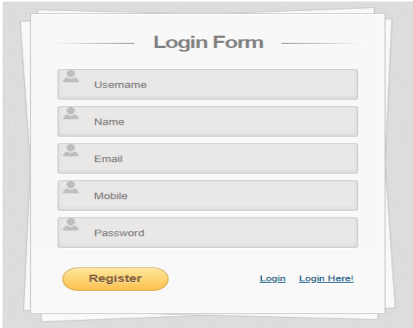

Fig 8 Sign Up

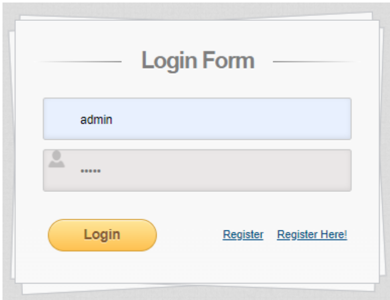

Fig 9 Sign In

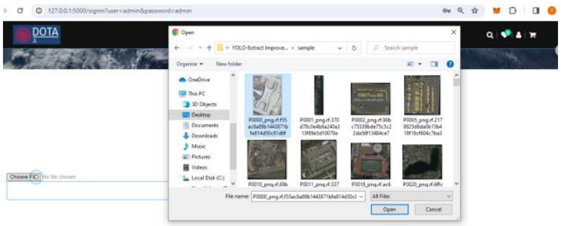

Fig 10 upload input image

International journal of basic and applied research

www.pragatipublication.com

ISSN 2249-3352 (P) 2278-0505 (E)

Cosmos Impact Factor-5.86

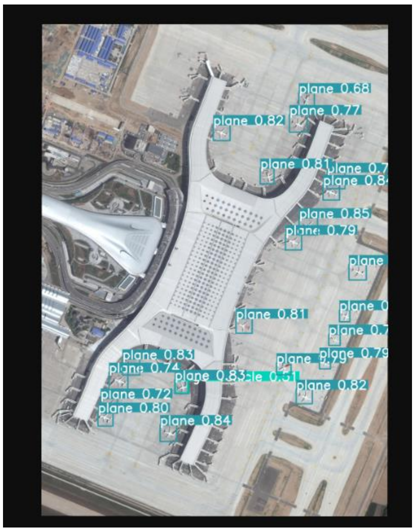

Fig 11 predicted result

5. CONCLUSION

In conclusion, the project demonstrates a comprehensive exploration and implementation of various YOLOv5 variants tailored for aircraft object detection in satellite remote sensing images. By strategically selecting and optimizing these models, the project effectively addresses challenges such as small target sizes and dense arrangements, crucial for accurate detection in real-world scenarios. Innovations like the coordinate attention mechanism in YOLOv5-Extract significantly enhance feature extraction capabilities, leading to improved detection accuracy and efficiency.

The project's extension work highlights the superiority of YOLOv5s, showcasing its outperformance in all performance metrics compared to other models. This underscores its effectiveness and robustness in aerial surveillance applications. Moreover, the development of a user-friendly frontend using Flask underscores the project's commitment to practical usage and user testing, facilitating seamless interaction and fostering ongoing advancements in satellite remote sensing object detection. Overall, the project's findings contribute significantly to the field, paving the way for enhanced capabilities in aerial monitoring and surveillance tasks.

6. FUTURE SCOPE

The feature scope of YOLO-Extract encompasses several key enhancements aimed at improving aircraft object detection in remote sensing images using YOLOv5. Firstly, it simplifies the model structure to streamline processing and improve efficiency. Additionally, YOLO-Extract introduces a novel feature extractor and prediction head, optimizing feature extraction for better object detection accuracy. Moreover, the integration of a coordinate attention mechanism enhances spatial awareness and target localization capabilities. These enhancements collectively contribute to reducing semantic loss and improving detection accuracy, especially in scenarios with small target sizes and dense arrangements common in satellite remote sensing imagery. By addressing these challenges, YOLO-Extract aims to provide more robust and accurate aircraft detection capabilities, facilitating applications in aerial surveillance, monitoring, and reconnaissance tasks. Its feature scope encompasses advancements that prioritize accuracy, efficiency, and adaptability to the unique challenges posed by remote sensing imagery.

REFERENCES

[1] R.Girshick, J. Donahue, T. Darrell, and J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 580587.

[2]R.Girshick,FastRCNN,inProc.IEEEInt.Conf.Comput.Vis.(ICCV), Dec. 2015, pp. 14401448.

[3] S. Ren, K.He,R.Girshick, and J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, in Proc. Adv. Neural Inf. Process. Syst., vol. 28, 2015, pp. 115.

[4] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, SSD: Single shot multibox detector, in Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer, 2016, pp. 2137.

[5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, You only look once: Unied, real-time object detection, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 779788.

[6] J. Redmon and A. Farhadi, YOLO9000: Better, faster, stronger, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 72637271.

[7] J. Redmon and A. Farhadi, YOLOv3: An incremental improvement, 2018, arXiv:1804.02767.

[8] A. Bochkovskiy, C.-Y.Wang, and H.-Y. Mark Liao, YOLOv4: Optimal speed and accuracy of object detection, 2020, arXiv:2004.10934.

[9] D. Xu and Y. Wu, Improved YOLO-V3 with DenseNet for multi scale remote sensing target detection, Sensors, vol. 20, no. 15, p. 4276. 2020.

[10] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, Densely connected convolutional networks, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 47004708.

[11]D.XuandY.Wu,FEYOLO:Afeatureenhancementnetworkforremote sensing target detection, Remote Sens., vol. 13, no. 7, p. 1311, Mar. 2021.

[12] S. H. Gao, M. M. Cheng, and K. Zhao, Res2Net: A new multi-scale backbone architecture, IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 2, pp. 652662, Feb. 2021.

[13] C. Cao, J. Wu, X. Zeng, Z. Feng, T. Wang, X. Yan, Z. Wu, Q. Wu, and Z. Huang, Research on airplane and ship detection of aerial remote sensing images based on convolutional neural network, Sensors, vol. 20, no. 17, p. 4696, Aug. 2020.

[14] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, Pyramid scene parsing network, in Proc. IEEE

Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 28812890.

[15] S. Luo, J. Yu, Y. Xi, and X. Liao, Aircraft target detection in remote sensing images based on improved YOLOv5, IEEE Access, vol. 10, pp. 51845192, 2022.

[16] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, Feature pyramid networks for object detection, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 21172125.

[17] H. Li, P. Xiong, J. An, and L. Wang, Pyramid attention network for semantic segmentation, 2018, arXiv:1805.10180.

[18] Q. Hou, D. Zhou, and J. Feng, Coordinate attention for efcient mobile networkdesign, inProc.IEEE/CVFConf.Comput.Vis.PatternRecognit., Jun. 2021, pp. 1371313722.

[19] F. Yu and V. Koltun, Multi-scale context aggregation by dilated convolutions, 2015, arXiv:1511.07122.

[20] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770778.

[21] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, Distance-IoU loss: Faster and better learning for bounding box regression, in Proc. AAAI Conf. Artif. Intell., 2020, vol. 34, no. 7, pp. 1299313000.

[22] Y.-F. Zhang, W. Ren, Z. Zhang, Z. Jia, L. Wang, and T. Tan, Focal and efcient IOU loss for accurate bounding box regression, Neurocomputing, vol. 506, pp. 146157, Sep. 2022.

[23] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, Focal loss for dense object detection, inProc.IEEEInt.Conf.Comput.Vis.(ICCV),Oct.2017, pp. 29802988.

[24] J. He, S. Erfani, X. Ma, J. Bailey, Y. Chi, and X.-S.Hua, alpha IoU: A family of power intersection over union losses for bounding box regression, in Proc. Adv. Neural Inf. Process. Syst., vol. 34, 2021, pp. 2023020242.

[25] C.-Y. Wang, A. Bochkovskiy, and H.-Y. Mark Liao, YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, 2022, arXiv:2207.02696.