IJITCE

ISSN 2347- 3657

International Journal of
Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 12, Issue 3, 2024

A FINE-GRAINED OBJECT DETECTION MODEL FOR AERIAL IMAGES BASED ON YOLOV5 DEEP NEURAL NETWORK

1 Nadia Anjum, Assistant Professor, Department of CME, Stanley College of Engineering & Technology for Women, Telangana, India, nadiaanjum@stanley.edu.in

2 Khansa Nazeer, Department of CME, Stanley College of Engineering & Technology for Women, Telangana, India, khansanazeer03@gmail.com

3 Hamda Nadeem, Department of CME, Stanley College of Engineering & Technology for Women, Telangana, India, hamda.nadeem.2002@gmail.com

4 Syeda Saniya Razvi, Department of CME, Stanley College of Engineering & Technology for Women, Telangana, India, viqarrazvi14@gmail.com

Abstract: This research aims to solve the common task of highly precise object detection in remote sensing images, which existing approaches that were developed for natural scenes are not suitable for. In this process we are using Circular Smooth Label (CSL), which does angle regression into a class-type, reducing losses due to angle periodicity. YOLOv5 will be taken as the original model for which we add the CSL and an attention mechanism module to strive for higher detection accuracy for small objects with arbitrary orientations. On the FAIR1M dataset, our YOLOv5-CSL resulting in an average 0.72 mAP. Besides that, examining the types such as YOLOv5x6 yields a measuring points, with the mAP increasing by 0.80% or above. This note allows to assert that a remote sensing object detection enhancement certainly belongs to a kind of the research topics in the future.

***Index terms** - Fine-grain object detection, High-resolution aerial images, Oriented object detection, YOLOv5*.

1. INTRODUCTION

In recent years, the field of computer vision has witnessed significant advancements in recognizing multiple classes of fine-grained objects in high-resolution remote sensing images. Object recognition, being one of the core tasks in computer vision, serves as the cornerstone for various cutting-edge applications including automatic driving, robot vision, and video surveillance [1]. However, unlike natural scenes where object orientation is typically horizontal, objects in remote sensing images exhibit arbitrary orientations. This inherent characteristic poses several challenges to remote sensing image object detection, such as dense alignment, complex background, and the prevalence of small targets [2].

Rotation detectors have emerged as indispensable tools in both civil and military domains, encompassing applications such as military defense, intelligent transportation, geological disaster monitoring, maritime surveillance, and urban planning [1], [2]. These detectors provide precise orientation and scale information for fine-grained objects in high-resolution remote sensing images, thus enabling accurate object recognition and classification.

In recent years, convolutional neural networks (CNNs) have revolutionized deep learning techniques, empowering them with robust feature extraction and representation capabilities. This has led to a significant improvement in the accuracy of object detection [3]. While classical object detectors leveraging deep learning methods have demonstrated promising results on natural scene images such as the COCO dataset and the VOC benchmark test [4]–[10], applying these generic detection algorithms directly to aerial remote sensing images encounters various difficulties and challenges.

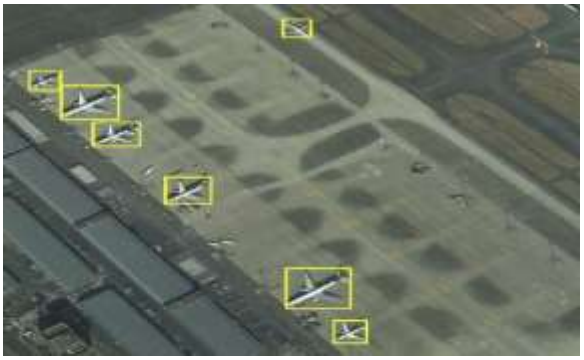

Fig 1 Object Detection Aerial Image

The fundamental differences between remote sensing images and natural images present hurdles for generic detection algorithms. Consequently, researchers have devised rotating object detectors to address these challenges. Although satisfactory results have been achieved on some publicly available large datasets [11]–[13], several fundamental issues persist.

To accurately locate and classify rotating targets in remote sensing images, existing studies typically rely on angular representations of rotating bounding boxes based on five parameters: rotation angle, center position, width, and height. However, many of these methods suffer from boundary discontinuity, which arises from periodic angle and regression inconsistencies. This instability in the training process can lead to inaccurate predictions of orientation, thereby compromising detection accuracy.

The accurate prediction of the angle is crucial for the detection of rotating objects. Therefore, addressing the challenges associated with angle prediction is imperative for improving the performance of rotating object detectors in remote sensing images.

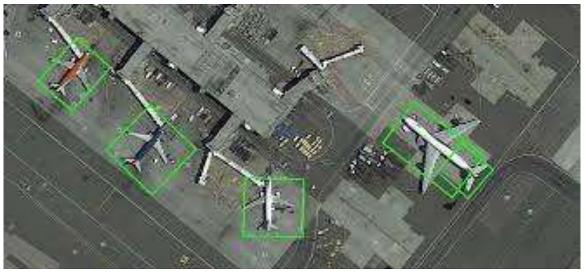

Fig 2 Aerial Image

In this paper, we delve into the intricacies of fine-grained object detection in high-resolution remote sensing images. We investigate the limitations of existing detection algorithms when applied to remote sensing images and explore the significance of rotation detectors in overcoming these challenges. Additionally, we examine the shortcomings of current rotating object detection methods, particularly concerning angle prediction accuracy. Through our research, we aim to contribute to the advancement of object detection in remote sensing images, ultimately facilitating the development of more robust and accurate detection systems for various real-world applications.

II. LITERATURE SURVEY

Object detection in very high resolution optical remote sensing images is a fundamental problem faced for remote sensing image analysis. Due to the advances of powerful feature representations, machine-learning-based object detection is receiving increasing attention [1]. Although numerous feature representations exist, most of them are handcrafted or shallow-learning-based features. As the object detection task becomes more challenging, their description capability becomes limited or even impoverished. More recently, deep learning algorithms, especially convolutional neural networks (CNNs), have shown their much stronger feature representation power in computer vision. Despite the progress made in nature scene images, it is problematic to directly use the CNN feature for object detection in optical remote sensing images because it is difficult to effectively deal with the problem of object rotation variations. To address this problem, this paper proposes a novel and effective approach to learn a rotation-invariant CNN (RICNN) model for advancing the performance of object detection[21], which is achieved by introducing and learning a new rotation-invariant layer on the basis of the existing CNN architectures[30,31,32]. However, different from the training of traditional CNN models that only optimizes the multinomial logistic regression objective, our RICNN model is trained by optimizing a new objective function via imposing a regularization constraint, which explicitly enforces the feature representations of the training samples before and after rotating to be mapped close to each other, hence achieving rotation invariance. To facilitate training, we first train the rotation-invariant layer and then domain-specifically fine-tune the whole RICNN network to further boost the performance. Comprehensive evaluations on a publicly available ten-class object detection data set demonstrate the effectiveness of the proposed method.

A new dataset with the goal of advancing the state-of-the-art in object recognition by placing the question of object recognition in the context of the broader question of scene understanding. This is achieved by gathering images of complex everyday scenes containing common objects in their natural context. [7,13,15] Objects are labeled using per-instance segmentations to aid in precise object localization. Our dataset contains photos of 91 objects types that would be easily recognizable by a 4 year old. With a total of 2.5 million labeled instances in 328k images, the creation of our dataset drew upon extensive crowd worker involvement via novel user interfaces for category detection, instance spotting and instance segmentation. We present a detailed statistical analysis of the dataset in comparison to PASCAL, ImageNet, and SUN. Finally, we provide baseline performance analysis for bounding box and segmentation detection results using a Deformable Parts Model.

There are a huge number of features which are said to improve Convolutional Neural Network (CNN) accuracy. Practical testing of combinations of such features on large datasets, and theoretical justification of the result, is required [32,33]. Some features operate on certain models exclusively and for certain problems exclusively, or only for small-scale datasets; while some features, such as batch-normalization and residual-connections, are applicable to the majority of models, tasks, and datasets [5]. We assume that such universal features include Weighted-Residual-Connections (WRC), Cross-Stage-Partial-connections (CSP), Cross mini-Batch Normalization (CmBN), Self-adversarial-training (SAT) and Mish-activation. We use new features: WRC, CSP, CmBN, SAT, Mish activation, Mosaic data augmentation, CmBN, DropBlock regularization, and CIoU loss, and combine some of them to achieve state-of-the-art results: 43.5% AP (65.7% AP50) for the MS COCO dataset at a realtime speed of ~65 FPS on Tesla V100.

In object detection, keypoint-based approaches often suffer a large number of incorrect object bounding boxes, arguably due to the lack of an additional look into the cropped regions. This paper presents an efficient solution which explores the visual patterns within each cropped region with minimal costs. We build our framework upon a representative one-stage keypoint-based detector named CornerNet[6]. Our approach, named CenterNet, detects each object as a triplet, rather than a pair, of keypoints, which improves both precision and recall. Accordingly, we design two customized modules named cascade corner pooling and center pooling, which play the roles of enriching information collected by both top-left and bottom-right corners and providing more recognizable information at the central regions, respectively. On the MS-COCO dataset, CenterNet achieves an AP of 47.0%, which outperforms all existing one-stage detectors by at least 4.9%. Meanwhile, with a faster inference speed, CenterNet demonstrates quite comparable performance to the top-ranked two-stage detectors[6].

Object detection performance, as measured on the canonical PASCAL VOC dataset, has plateaued in the last few years. The best-performing methods are complex ensemble systems that typically combine multiple low-level image features with high-level context. In this paper, we propose a simple and scalable detection algorithm that improves mean average precision (mAP) by more than 30% relative to the previous best result on VOC 2012---achieving a mAP of 53.3%. Our approach combines two key insights: (1) one can apply high-capacity convolutional neural networks (CNNs) to bottom-up region proposals in order to localize and segment objects and (2) when labeled training data is scarce, supervised pre-training for an auxiliary task, followed by domain-specific fine-tuning, yields a significant performance boost. Since we combine region proposals with CNNs, we call our method R-CNN: Regions with CNN features. [32,33] We also compare R-CNN to OverFeat, a recently proposed sliding-window detector based on a similar CNN architecture. We find that R-CNN outperforms OverFeat by a large margin on the 200-class ILSVRC2013 detection dataset.

ANALYSIS OF TECHNIQUES DISCUSSED IN LITERATURE SURVEY

S.NO	YEAR	AUTHOR	TITLE	DATASET	METHODOLOGY	RESULT
1	2022	Xian Sun, Peijin Wang, Zhiyuan Yan, et.al.,	FAIR1M: A benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery	FAIR1M dataset	The methodology involves collecting remote sensing images with resolutions of 0.3m to 0.8m from various platforms, annotating objects with oriented bounding boxes for 5 categories and 37 sub-categories, and proposing novel evaluation methods.	The FAIR1M dataset presents a more challenging environment for fine-grained object detection compared to DOTA, as evidenced by a significant difference in mAP scores.
2	2022	Wentong Li; Yijie Chen; Kaixuan Hu; Jianke Zhu	Oriented RepPoints for Aerial Object Detection	DOTA, HRSC2016, UCAS-AOD and DIOR-R	The proposed approach for aerial object detection utilizes adaptive points learning to handle non-axis aligned targets. Three oriented conversion functions aid classification and localization. Quality assessment and sample assignment scheme select representative samples, with a spatial constraint to penalize	Experimental results on DOTA, HRSC2016, UCAS-AOD, and DIOR-R datasets validate the efficacy of the proposed adaptive points learning approach for aerial object

					outliers for robust learning.	detection.
3	2021	Jiaming Han, Jian Ding, et.al.,	ReDet: A Rotation-equivariant Detector for Aerial Object Detection	DOTA-v1.0, DOTA-v1.5 and HRSC2016	The proposed Rotation-equivariant Detector (ReDet) incorporates rotation-equivariant networks to extract rotation-equivariant features, enabling accurate orientation prediction and reducing model size. Rotation-invariant RoI Align (RiRoI Align) adaptively extracts rotation-invariant features from equivariant features based on RoI orientation.	ReDet achieves state-of-the-art performance in aerial object detection while significantly reducing model parameters.
4	2021	Jiaming Han, Jian Ding, Jie Li, Gui-Song Xia	Align Deep Features for Oriented Object Detection	DOTA and HRSC2016	The proposed Single-shot Alignment Network (S2A-Net) comprises a Feature Alignment Module (FAM) and an Oriented Detection Module (ODM). FAM refines anchors with an Anchor Refinement Network and aligns convolutional features using Alignment Convolution.	S2A-Net achieves state-of-the-art performance on DOTA and HRSC2016 datasets while maintaining high efficiency.

					ODM utilizes active rotating filters for orientation encoding and produces orientation-sensitive features to improve classification and localization accuracy. Additionally, we explore object detection in large-size images for improved speed and accuracy.	
5	2021	Xue Yang, Junchi Yan, Ziming Feng, et.al.,	R 3Det: Refined Single-Stage Detector with Feature Refinement for Rotating Object	DOTA, HRSC2016, UCAS-AOD, ICDAR2015	Our method proposes an end-to-end refined single-stage rotation detector for fast and accurate object detection, utilizing progressive regression for coarse to fine granularity. It incorporates a feature refinement module to address feature misalignment, enhancing detection performance by re-encoding position information through pixel-wise feature interpolation. Additionally, an approximate SkewIoU loss is introduced for	Experiments on DOTA, HRSC2016, UCAS-AOD, and ICDAR2015 datasets demonstrate the effectiveness of our rotation detection approach.

					more accurate rotation estimation.	
6	2021	Jingru Yi, Pengxiang Wu, Bo Liu, et.al.,	Oriented Object Detection in Aerial Images with Box Boundary-Aware Vectors	DOTA, HRSC2016	We extend a horizontal keypoint-based detector for oriented object detection in aerial images. Center keypoints are first detected, followed by regression of box boundary-aware vectors (BBAVectors) to capture oriented bounding boxes, distributed in four quadrants. To handle imbalance, boxes are classified as horizontal or rotational.	The proposed method, utilizing box boundary-aware vectors, outperforms baseline approaches in oriented object detection on aerial images, demonstrating competitive performance against state-of-the-art methods in experiments.
7	2021	Xue Yang, Xiaojiang Yang, et.al.,	Learning High-Precision Bounding Box for Rotated Object Detection via Kullback-Leibler Divergence	DOTA, UCAS-AOD, HRSC2016, ICDAR2015, MLT and MSRA-TD500	We propose a novel rotation regression loss design for rotated object detection, based on the deduction methodology from the relation between rotation and horizontal detection. This loss utilizes Kullback-Leibler Divergence (KLD) between 2-D Gaussian	Experimental results demonstrate the superiority of the proposed Kullback-Leibler Divergence (KLD) rotation regression

					distributions, dynamically adjusting parameter gradients for adaptive optimization.	loss in high-precision detection on seven datasets, with publicly available code.
8	2021	Xue Yang, Junchi Yan, Qi Ming, et.al.,	Rethinking Rotated Object Detection with Gaussian Wasserstein Distance Loss	DOTA, UCAS-AOD, HRSC2016, ICDAR2015, ICDAR 2017 MLT	We propose a novel regression loss based on Gaussian Wasserstein distance (GWD) for rotating object detection. This loss converts rotated bounding boxes into 2-D Gaussian distributions, enabling efficient learning via gradient back-propagation. GWD addresses boundary discontinuity and square-like problems, enhancing detection accuracy across various datasets.	Experimental results across five datasets using various detectors demonstrate the effectiveness of our Gaussian Wasserstein distance-based regression loss for rotating object detection. Codes are publicly available.
9	2021	Aravind Srinivas; Tsung-Yi Lin; Niki	Bottleneck Transformers for Visual Recognition	COCO	BoTNet replaces spatial convolutions with global self-attention in the final three	BoTNet achieves 44.4% Mask AP and

		Parmar; Jonathon Shlens; Pieter Abbeel; Ashish Vaswani			bottleneck blocks of a ResNet, enhancing performance in image classification, object detection, and instance segmentation tasks. This design aligns ResNet bottleneck blocks with Transformer blocks, achieving state-of-the-art results across tasks.	49.7% Box AP on COCO Instance Segmentation, surpassing previous best results. It achieves 84.7% top-1 accuracy on ImageNet, outperforming EfficientNet in speed.
10	2020	Ke Li, Gang Wan, Gong Cheng, et.al.,	Object Detection in Optical Remote Sensing Images: A Survey and A New Benchmark	DIOR dataset	The methodology involves conducting a comprehensive review of deep learning-based object detection advancements, identifying shortcomings in existing datasets, and proposing a large-scale benchmark dataset called DIOR, comprising 23463 images and 192472 instances across 20 object classes, to address these limitations.	The DIOR dataset, with 23463 images and 192472 instances across 20 classes, offers significant scale, diversity, and complexity, facilitating robust evaluation of object detection methods.

III. GENERAL METHODOLOGY

The proposed system, YOLOv5_csl, is devised to enhance fine-grained object detection in aerial and remote sensing images by integrating the YOLOv5 algorithm as a baseline and introducing the Circular Smooth Label (CSL) method [18]. The system incorporates an angle classification module (CSL) and an attention mechanism module to leverage global context effectively. Additionally, advanced techniques such as YOLOv5x6, YOLOv6, and YOLOv7 are explored, with YOLOv5x6 achieving a notable mean Average Precision (mAP) of 69.4% [19]. To facilitate user engagement and testing, a frontend is developed using the Flask framework, providing a user-friendly interface for fine-grained object detection model evaluation based on the YOLOv5 deep neural network. Authentication integration ensures secure access, offering a comprehensive solution for both performance improvement and real-world user testing scenarios.

The system architecture, as depicted in Figure 3, utilizes the YOLOv5 algorithm as the baseline for capturing the angles of arbitrarily oriented targets in aerial images through the circular smooth label method. Integration of a self-attention mechanism expands the perceptual field of the model, enabling better fitting of targets in the dataset and enhancing detection accuracy, particularly for subtle targets. Extensive experiments conducted on widely available public evaluation datasets such as DOTA and FAIR1M demonstrate the effectiveness of the proposed YOLOv5_csl method for aerial object detection [19].

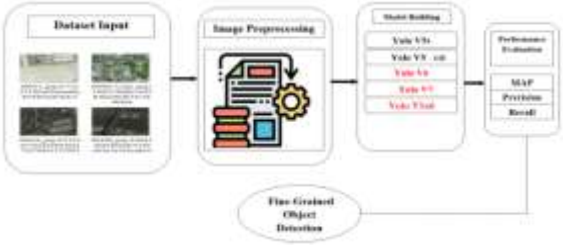

Fig 3 Proposed Architecture

Image processing constitutes a crucial component of the system, involving various preprocessing steps to prepare the aerial images for object detection. This includes converting images into blob objects, resizing, normalization, and defining object classes and bounding boxes. The processed data is then utilized to train a neural network model, typically based on YOLO or similar frameworks capable of object localization within images. Performance evaluation metrics such as mAP, Precision, and Recall are employed to assess the trained model's effectiveness.

Data augmentation techniques play a vital role in enhancing the diversity and robustness of the training dataset. Randomizing, rotating, and transforming images introduce variability and simulate real-world scenarios, thereby improving the model's ability to generalize and perform effectively on diverse test data [1].

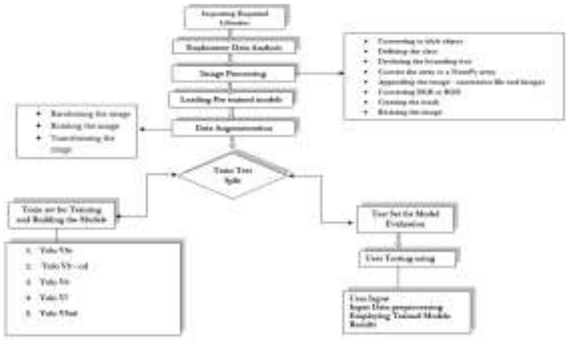

Fig 4 Flow for General Methodology

IV. DATASET

The object detection process commences with the acquisition of a dataset containing aerial images, which serve as the primary input for subsequent stages. Upon obtaining the dataset, the initial step involves reading the images and conducting an exploratory analysis. This exploration entails plotting the images to glean insights into the dataset's content and structure. By visualizing the images, researchers gain a preliminary understanding of the objects present, their distribution, and any inherent patterns or variations. This initial exploration is crucial for informing subsequent preprocessing steps and model development. Additionally, it aids in identifying any potential challenges or anomalies within the dataset, allowing for appropriate adjustments and optimizations to be made. Overall, this preliminary analysis sets the foundation for the effective implementation of the object detection system and ensures the subsequent stages proceed smoothly and effectively.

Fig 4 Dataset Images

DISCUSSION & CHALLENGES:

In the discussion of fine-grained object detection in aerial and remote sensing images, several key points emerge regarding the proposed methodology, its implications, and associated challenges.

Firstly, the utilization of the YOLOv5 algorithm as a baseline, augmented with the Circular Smooth Label (CSL) method, showcases a promising approach for addressing the challenges posed by arbitrary object orientations in remote sensing images. This methodological fusion not only enhances the detection accuracy but also improves the model's ability to handle small and subtle targets.

However, despite the promising results achieved with YOLOv5_csl, several challenges remain. One notable challenge is the complexity of remote sensing images, characterized by dense alignment, complex backgrounds, and the predominance of small targets. These factors contribute to the difficulty in accurately detecting and classifying fine-grained objects, necessitating further refinement and optimization of the proposed methodology.

Moreover, the integration of advanced techniques such as YOLOv5x6 and YOLOv6 introduces additional complexities and considerations. While these techniques have shown improved performance, their implementation may require significant computational resources and expertise, posing practical challenges for deployment in real-world scenarios.

Additionally, the development of a user-friendly interface and the integration of authentication mechanisms present challenges in terms of usability, security, and scalability. Balancing these aspects while ensuring seamless interaction and controlled access to the system is essential for its practical viability and adoption.

Overall, while the proposed methodology shows promise for fine-grained object detection in remote sensing images, addressing these challenges will be critical for realizing its full potential and applicability in real-world contexts.

V. CONCLUSION

The project successfully addressed limitations in existing object detection algorithms [1,2,6,7,8], extending their applicability to fine-grained objects in remote sensing scenarios. The introduction of an attention mechanism module in YOLOv5 enhanced the model's ability to perform fine-grained object detection. This improvement resulted in better recognition accuracy, especially for small objects, without sacrificing computational efficiency. The extended YOLOv5x6 algorithm excelled in fine-grained object detection, achieving an impressive 69.45% mAP. With advanced architectures, diverse datasets, real-time optimization, and continuous adaptation, YOLOv5x6 stands as a robust solution, promising significant advancements in remote sensing applications. The integration of the Flask framework, coupled with SQLite for user authentication, established a user-friendly interface. This allowed users to input images, and the system seamlessly processed and displayed the final outcomes, showcasing the practical application of the developed models. Researchers [13,15,16,18], industry professionals, and government agencies gain from enhanced object detection in remote sensing for improved analysis and decision-making. The technology's advancements also benefit technology developers, end users in remote sensing applications, and the general public through increased accuracy and safety measures..

FUTURE SCOPE

Explore the integration of state-of-the-art neural network architectures and attention mechanisms to further enhance the model's capability for fine-grained object detection in remote sensing images. Extend the project's scope by incorporating more diverse and extensive datasets for training, allowing the model to generalize better across various real-world scenarios and improve its performance on a broader range of remote sensing applications. Focus on optimizing the algorithm for real-time processing and deployment on edge devices, enabling its use in applications such as autonomous systems, surveillance, and disaster response with minimal.

REFERENCES

[1] 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.

[2] 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.

[3] 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.

[7] T.Y. Lin, P. Goyal, R. Girshick, et al., “Focal loss for dense object detection,” in Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, pp.2980– 2988, 2017.

[8] W. Liu, D. Anguelov, D. Erhan, et al., “ SSD: Single shot multibox detector,” in Proceedings of the European Conference on Computer Vision, Springer, Cham, pp.21–37, 2016.

[9] J. Redmon, S. Divvala, R. Girshick, et al., “You only look once: Unified, real-time object detection,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, pp.779–788, 2016.

[10] S. M. Azimi, E. Vig, R. Bahmanyar, et al., “Towards multiclass object detection in unconstrained remote sensing imagery,” in Proceedings of Asian Conference on Computer

[11] Vision, Springer, Cham, pp.150–165, 2019. G. Zhang, S. Lu, and W. Zhang, “ CAD-Net: A contextaware detection network for objects in remote sensing imagery,” IEEE Transactions on Geoscience Remote Sensing, vol.57, no.12, pp.10015–10024, 2019.

[12] J. Han, J. Ding, N. Xue, et al., “ReDet: A rotation-equivariant detector for aerial object detection,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, pp.2768–2795, 2021.

[13] X. Yang, J. Yang, J. Yan, et al. “ SCRDet: Towards more robust detection for small, cluttered and rotated objects,” in Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, pp.8231–8240, 2019.

[14] J. Ding, N. Xue, Y. Long, et al., “Learning roi transformer for oriented object detection in aerial images,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, pp.2844–2853, 2019.

[15] J. Ding, N. Xue, Y. Long et al., "Learning roi transformer for oriented object detection in aerial images", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2844-2853, 2019.

[16] J. Han, J. Ding, J. Li, et al., “Align deep features for oriented object detection,” IEEE Transactions on Geoscience and Remote Sensing, vol.60, pp.1–11, 2021.

[17] X. Yang, J. Yan, Z. Feng, et al., “ R3Det: Refined singlestage detector with feature refinement for rotating object,” in Proceedings of the 35th AAAI

Conference on Artificial Intelligence, Virtual Event, pp.3163–3171, 2021.

[18] X. Yang and J. Yan. " Arbitrary-oriented object detection with circular smooth label," in Proceedings of European Conference on Computer Vision 2020, LNCS, vol.12353, Springer, Cham, pp.677–694, 2020.

[19] G. S. Xia, X. Bai, J. Ding, et al., " DOTA: A large-scale dataset for object detection in aerial images," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, pp.3974–3983, 2018.

[20] X. Sun, P. Wang, Z. Yan, et al., "FAIR1M: A benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery," ISPRS Journal of Photogrammetry Remote Sensing, vol.184, pp.116–130, 2022.

[21] X. Yang, H. Sun, K. Fu, et al., "Automatic ship detection in remote sensing images from google earth of complex scenes based on multiscale rotation dense feature pyramid networks," Remote Sensing , vol.10, no.1, article no.132, 2018.

[22] K. Fu, Z. Chang, Y. Zhang, et al., " Rotation-aware and multi-scale convolutional neural network for object detection in remote sensing images," ISPRS Journal of Photogrammetry Remote Sensing, vol.161, pp.294–308, 2020.

[23] Z. Liu, H. Wang, L. Weng, et al., "Ship rotated bounding box space for ship extraction from high-resolution optical satellite images with complex backgrounds," IEEE Geoscience Remote Sensing Letters, vol.13, no.8, pp.1074– 1078, 2016.

[24] S. Ren, K. He, R. Girshick, et al., " Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.39, no.6, pp.1137–1149, 2017.

[25] L. Zhou, H. Wei, H. Li, et al., " Objects detection for remote sensing images based on polar coordinates," arXiv preprint, arXiv: 2001.02988, 2020.

[26] J. Yi, P. Wu, B. Liu, et al., "Oriented object detection in 62 Chinese Journal of Electronics 2023 aerial images with box boundary-aware vectors," in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, pp.2149– 2158, 2021.

[27] W. Li, Y. Chen, K. Hu, et al., "Oriented reppoints for aerial object detection," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, Louisiana, pp.1829–1838, 2022.

[28] X. Yang, X. Yang, J. Yang, et al., "Learning high-precision bounding box for rotated object detection via kullback-leibler divergence," Advances in Neural Information Processing Systems, vol.34, pp.18381–18394, 2021.

[29] X. Yang, J. Yan, Q. Ming, et al., " Rethinking rotated object detection with Gaussian Wasserstein distance loss," in Proceedings of the International Conference on Machine Learning, Vienna, Austria, pp.11830–11841, 2021.

[30] J. Hu, L. Shen, and G. Sun, " Squeeze-and-excitation networks," in Proceedings of the IEEE

Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, pp.7132–7141, 2018.

[31] S. Woo, J. Park, J. Y. Lee, et al., " CBAM: Convolutional block attention module," in Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, pp.3–19, 2018.

[32] H. Zhao, J. Jia, and V. Koltun, "Exploring self-attention for image recognition," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, pp.10073–10082, 2020.

[33] A. Srinivas, T. Y. Lin, N. Parmar, et al., "Bottleneck transformers for visual recognition," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, pp.16514–16524, 2021.

[34] A. F. Agarap, " Deep learning using rectified linear units (ReLU)," arXiv preprint, arXiv: 1803.08375, 2018.

[35] B. Xu, N. Wang, T. Chen, et al., "Empirical evaluation of rectified activations in convolutional network," arXiv preprint, arXiv: 1505.00853, 2015.

[36] D. Misra, " Mish: A self regularized non-monotonic activation function," arXiv preprint, arXiv: 1908.08681, 2019.

[37] J. Deng, W. Dong, R. Socher, et al., "A large-scale hierarchical image database," in Proceedings of IEEE Computer Vision and Pattern Recognition, Miami, FL, USA, pp.248–255, 2009.

[38] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in Proceedings of the International Conference on Machine Learning, Long Beach, California, USA, pp.6105–6114, 2019.

[39] N. Ma, X. Zhang, M. Liu, et al., "Activate or not: Learning customized activation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, pp.8028–8038, 2021.