UtilitasMathematica

ISSN 0315-3681 Volume 122 (2), 2025

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

[1] Kanaparthy Neeraja, [2] Vinit Gunjan,

Associate Professor

[1, 2] Department of Computer Science and Engineering, CMR Institute of Technology

Abstract: Research seeks to triumph over the reduction of state-of-the-art algorithms of item reputation, which can be on the whole optimized for the herbal surroundings, via introducing strategies specially evolved to discover pleasant -grained gadgets in far flung taking pictures photographs. Current strategies of recognizing faraway survey items regularly come upon issues with horizontal detection due to headaches in attitude regression, resulting in vast injury to the getting to know of the model. The cyclic nature of angles prevents regression strategies and prevents accurate identification in far flung sensing programs. The proposed approach represents a circular smooth label (CSL) method to transform an perspective regression into a classification layout. This new method offers with the problem of angular regression and offers a extra efficient way to recognize far flung survey gadgets. The progressed version takes advantage of the benefits of Yolov5 as a basis integrates many modules and strategies to increase the accuracy of detection, mainly in small items. The use of CSL increases the ability of the model to understand the angles of desires with any orientation. The improved model achieves efficiency and simplicity, minimizes hardware requirements and denotes a potential trajectory for future progress in fine -grained object recognition in long -distance survey images. In addition, the studies have used advanced approaches such as Yolov5x6, Yolov6 and Yolov7, with Yolov5x6 achieved significant mean average precision (mAP) 69.4%. The front end was built using a flask frame to strengthen the user's interaction and offering a user -friendly interface to test the fine -grained model recognition of objects in aerial photographs with a deep neural network Yolov5. Incorporation of verification ensures safe and regulated access to the system and provides a thorough solution to increase performance and users' evaluation in practical situations.

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

I. INTRODUCTION

This study addresses the current challenge in the computer vision of identifying many categories of fine -grained objects inside the high -resolution remote exploration. The identification of the object, one of the 3 fundamental obligations of pc imaginative and prescient, serves as a technological foundation for superior applications such as automatic control, robotic vision and video supervision. Remote sensing pics often display a random orientation of gadgets, in contrast to herbal settings, which can be usually horizontal. Recognition of gadgets in far off survey pixel is limited by dense alignment, complicated backdrops and small desires. Rotation detectors offer accurate orientation and scale for complicated objects in high -decision far off -decision pictures, essential in army defense, smart shipping, geological disaster tracking, maritime supervision, urban planning and other applications [1], [2].

Recently convolutional neural networks (CNN) rejuvenated deep getting to know by way of facilitating sturdy extraction and illustration of factors and therefore elevated the accuracy of the identity of the object. However, most conventional item detectors using deep gaining knowledge of methodologies have shown efficiency in the photos of natural scenes, as evidenced through the coconut data record and VOC test. Given the good sized differences among lengthy -distance survey pictures and herbal snap shots, traditional popularity algorithms maintain to stumble upon problems while they are without delay carried out to the photos of the lengthy -distance survey antenna.

As a end result, scientists have created numerous rotating item detectors; However, no matter achieving fine outcomes on publicly available large statistics files [11] - [13], primary demanding situations persist. Numerous research used the angular representations of spinning border bins defined by five parameters - the angle of rotation, the vicinity of the middle, the width and height - to efficaciously become aware of and classify revolving items in remote sensing pics. Many of those methodologies

display discontinuity or loss that may be attributed to a periodic perspective and regression irregularities, resulting in instability at some stage in the schooling procedure and probably threatening the prediction of orientation. The accuracy of the prediction of the attitude is essential to become aware of the spinning gadgets.

II. RELATED WORK

An important challenge in the processing of a remote exploration image is to recognize objects in high resolution optical images. The identification of an object based on machine learning is becoming increasingly popular due to robust representations [1]. Despite several features, most are either produced or based on shallow learning. As the object detection becomes more demanding, the capacity of the description decreases. Deep learning methodologies, especially convolutional neural networks (CNN), have recently demonstrated excellent representation of features in computer vision. Regardless of the fact that it is progress in the nature scene, the use of CNN functions to recognize objects in optical remote shooting of photos is a challenge due to changes in rotation of objects. This research represents an innovative and effective approach to the development a rotationinvariant CNN (RICNN) for the rotation of the invariant to identify objects by inclusion and training of a new layer based on established CNN architectures. Unlike the CNN conventional models, which only optimize the multinomial logistical regression goal, our RICNN model increases the new objective function by integrating the regulatory restrictions, which explicitly aligns representation of the elements of training samples before and after rotation. First we train a rotary invariant layer and then fine - tune the RICNN network for certain domains to increase performance. The proposed strategy will prove to be successful through comprehensive evaluation of a publicly available data set of identifying ten classes.

A new data file that contextualizes identification of items inside the scene interpretation to support the discipline. This is achieved by photographing complex daily situations with common things in their authentic environment. [7, 13, 15]
Segmentation on an instance facilitates the location of items. Our data file includes photos of 91 things that four years can identify. Our data file includes 2.5 million annotated instances of 328,000 photos generated by several user interfaces for category identification, instance detection, and DAVE instance segmentation. We perform statistical analysis of the data set and compare it to Pascal, Imagenet and the Sun. We end with the performance of the baseline of the model of deformable parts for detecting the border of boxing and segmentation.

Numerous factors increase CNN accuracy. The combination of these features must be empirically evaluated on extensive data sets and theoretically documented [22,23]. Although batch normalization and residual connections are appropriate for most fashions, responsibilities and information sets, a few factors are limited to certain models, demanding situations or restrained facts sets [5]. It is thought that SAT, MISH activation, Weighted-ResidualConnections (WRC), Cross-Stage-PartialConnections (CSP), and Cross mini-Batch Normalization (CmBN) have conventional attributes. We integrate modern capabilities including WRC, CSP, CMBN, SAT, MISH
ACTIVATION, MOSAIC DATA AUGMENTATION, Dropblock Regularization and Loss of Ciou to get tremendous results. It reached 43.5% of the common accuracy (65.7% AP50) to the Tesla V100 at approximately 65 frames according to 2nd the use of the MS Coco statistics record.

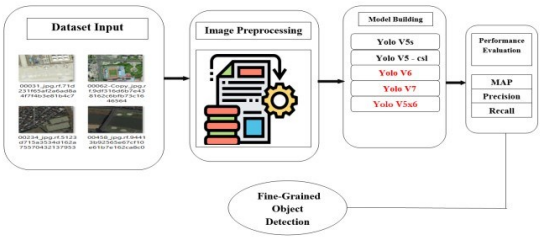
The identification of object -based objects often generates several unrective bounding boxes bounding objects due to insufficient control of cropped areas. This work represents a cost effective method for analyzing visual patterns inside the cropped zones. Our solution is based on Corner net, a one -step keyboard detector [6]. Centernet identifies each item as a triplet of key points instead of a pair, increasing accuracy and induction. As a result, we are developing two specialized modules, Cascade Corner association and center association to improve the left and lower corner data and improve the recognition of basic areas. Centernet exceeds all one-stage detectors by at least 4.9% on the MS-Coco data file, which is AP 47.0%. Centernet has a comparable power with the management of two stage detectors, facilitated by its accelerated inference speed.

In latest years, the overall performance of the object identification on the traditional Pascal VOC information set has stagnated. Comprehensive record systems that integrate low -degree visual data with a high stage context provide top-quality

overall performance. This take a look at introduces an instantaneous and scalable detection technique that increases mAP by way of greater than 30% as compared to the preceding great bring about VOC 2012 and reaches 53.3%. Our methodology integrates simple findings: (1) High -potential CNN can successfully find and phase items from bottom to top and (2) oversees preschool tasks for auxiliary tasks associated with a pleasant given domain. We check with our method as R-CNN: regions with CNN traits due to integration of CNN place design. [22, 23] R-CNN is likewise as compared with overwhelming, currently proposed sliding windows based totally on CNN. R-CNN drastically exceeds immoderate advent at the Ilshvrc2013 detection facts set in the study room.

III. MATERIALS AND METHODS

The proposed approach, designated by Yolov5_cSL, seeks to improve the recognition of a fine -grained object in aerial and long -distance shooting images. It uses the Yolov5 algorithm as the basis and represents a circular smooth label methodology (CSL) [18]. The system includes an angle classification module (CSL) and the attention mechanism module to increase the use of global context. In addition, sophisticated approaches such as Yolov5x6, Yolov6 and Yolov7 have been integrated, with a significant mAP 69.4%. The front end was built using a flask frame to increase user interaction and offer a user -friendly interface to test the fine -grained model recognition of objects on aerial photographs based on the deep neural network Yolov5. Incorporating the verification ensures a safe and regulated access to the system and provides holistic solutions to increase the performance and evaluation of the user in practical contexts.



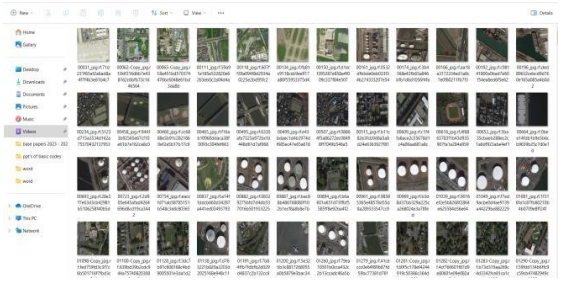
"Fig.1 Proposed Architecture"

This method identifies detailed objects in aerial images with Yolov5. The procedure begins with a

collection of aerial photographs. The images are then pre -processed to improve quality. Several Yolov5 models such as Yolov5, Yolov5-Csi, Yolov6, Yolov7 and Yolov5x6 were built and trained. The effectiveness of these models is evaluated by measures such as map, accuracy and download. This technology seeks to provide accurate and comprehensive recognition of objects in aerial images for purposes such as supervision and urban planning.

A) Dataset Collection:

The method begins with obtaining a data set consisting of aerial images. These photos act as input data for the following phases of the object detection system. Photos of data file are analyzed and preliminary examination includes mapping of these images to understand the content and structure of the data file.



"Fig.2 Dataset Collection" **B)**

Image Processing:

Image processing is essential for recognizing objects in autonomous driving systems, including many critical phases. The first step means transforming the input image to the Blob object, so they optimize it for further analysis and editing. Subsequently, categories of items to be recognized are determined, which specifies the exact classifications that intend to recognize the algorithm. At the equal time, the boundary bins are decided to define regions of hobby inside the picture, wherein the gadgets are predicted to be located. The analyzed facts are then converted right into a Numpy field, a main step for a fast numerical calculation and evaluation.

Another section is loading a pre -educated version the use of statistics derived from complicated information sets. This way exploring the network layers of a pre -planed version that includes discovered characteristics and parameters essential for the right identification of the object. In addition, output layers are developed, providing convincing predictions and facilitating powerful identity and categorization of gadgets.

In addition, in the photo processing piping, pix and annotations are combined and guarantee whole data for destiny evaluation. The color space is modified by way of transmission from BGR to RGB and a mask is generated to emphasize the elements. The photograph is sooner or later modified to boom its suitability for further processing and evaluation. This precise workflow of photo processing offers a robust basis for reliable and accurate popularity of gadgets in the growing surroundings of independent riding structures, improving street safety and choice - making.

C) Data Augmentation: Increasing facts is a key technique for improving variety and resistance of facts documents education for gadget getting to know models, mainly in photo processing and laptop imaginative and prescient. The technique includes three simple modifications for growing the original data set: picture randomization, picture rotation, and photograph transformation.

Randomization of the photograph increases the sort of random modifications, consisting of brightness, contrast or saturation. This stochastic method will increase the capability of the model to generalize to unknown statistics and diverse instances of the environment.

Turn the photograph way converting the orientation of the original picture at one-of-a-kind levels. This approach of augmentation helps in training the version to discover gadgets from special components and imitating variability inside the real world.

Image transformation consists of geometric changes such as scaling, reduce and rolling. These changes increase the statistics document through consisting of distortion that reflect the actual world deviations in the advent and orientation of the item.

The use of those records approaches for statistics augmentation increases the complexity of the facts set of training and lets in the model to obtain strong capabilities and patterns. This will increase the potential of the version to generalize and characteristic freely throughout special and traumatic check situations. Increasing information is an vital method to reduce excessive impact, improve the performance of the model and increase the overall reliability of device mastering fashions, specially in packages which includes photograph popularity for self reliant using structures. **D) Algorithms:**

Yolov5s, or "you look only one version 5 small" is a real -time object identification system recognized for its efficiency and speed. This particular edition of Yolov5 is optimized for lower computing requirements while maintaining real - time speed. Yolov5, which is selected for its efficiency, is particularly suitable for environments with limited computational means, which is optimal for use in applications such as remote sensing, where hardware restrictions may be present. "$LossYOLOv5s$" = "$Lbbox + Lobj + Lcls$"(1)

Where:

“Lbbox: Bounding box regression loss. Lobj: Objectness loss (confidence score). Lcls: Classification loss for detected objects”.

Yolov5-CSL is a modified variation of Jolov5, which integrates the technique of a circular smooth label (CSL) [18]. This modification increases the recognition of objects in high - resistant photographs with high resolution improvement accuracy of calculating the orientation of spinning objects.

Yolov5-CSL is specially designed for fine-grained object recognition, effectively solving problems associated with angle regression, and therefore facilitates detection of objects with variable orientation in a specified environment. "$LYOLOv5 - csl$" = " − i" = "1∑$Nyilog(y\^i)$"(2)

Where:
“yi: Smoothed ground truth label for angle classification. y^i: Predicted probability for each angle class. N: Total number of discrete angle classes”.

Yolov6 is a sophisticated framework for identifying objects adapted to industrial applications and achieves balance between speed and accuracy. The project includes specialized features: Bi-directional Concatenation (BIC) promotes the integration of functions, Anchor-aided Training (AAT) optimizes predictions, and improved spine achieves the latest power. Yolov6 offers many model versions to meet different calculation requirements, which provides an effective and accurate solution for identifying industrial objects.

"$FYOLOv6$" = "$Concat(Flow, Up(Fhigh))$"(3)

Where:

"Flow: Low-level feature map. Fhigh: High-level feature map. Up(·): Upsampling operation. Concat(·): Feature fusion by concatenation".

Yolov7 is the latest iteration in the Yolo series, which is characterized by exceptional speed and accuracy in identifying the object in real time. Yolov7, which has been selected for the project due to its unrivaled performance, excels in real - time processing, achieves excellent results on benchmarks such as Coco, and efficiently managing devices limited to resources. Its emphasis on strengthening small objects increases its adaptability, making it an impressive tool for applications such as video supervision and car without a driver.

Yolov5x6 is a variation of Yolov5, a one -stage model of object detection, which uses a CSPDarknet to extraction of elements, a panel for fusion functions and a head of Yolov5. Advantages include real -time recognition, increased accuracy and lightweight design, which is optimal for fine grained objects in aerial images. Optimized for mobile devices shows excellent efficiency, accuracy and speed.

IV. RESULTS AND DISCUSSION

Precision: Precision quantifies the percentage of efficiently identified positive cases or samples. Precision is decided by using the components:

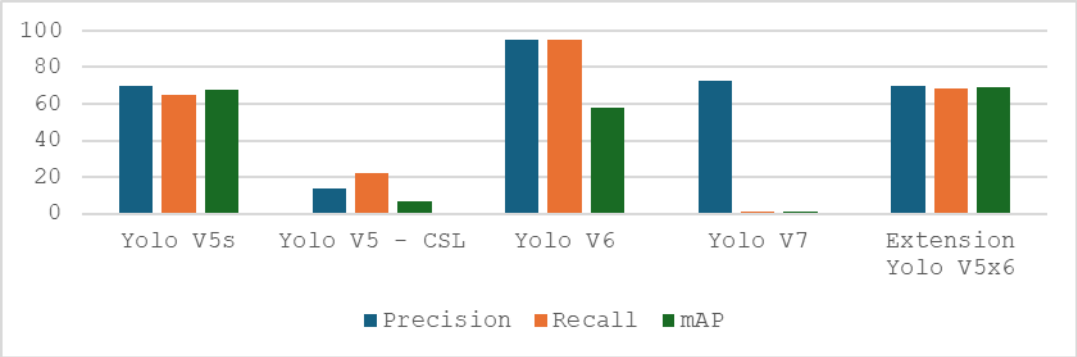
$$"Precision" = \frac{"True\ Positive"}{"True\ Positive + False\ Positive"} \quad (4)$$

Recall: ML recall assesses a model's potential to choose out all relevant times of a class. It demonstrates a version's efficacy in encapsulating times of a class by using comparing nicely anticipated high satisfactory observations to the general variety of positives.

Table.1 Performance Evaluation Table

ML Model	Precision	Recall	mAP
Yolo V5s	70.0	64.9	67.4
Yolo V5 - CSL	14.1	22.0	6.57
Yolo V6	95.0	95.0	57.8
Yolo V7	72.3	0.97	0.8
Extension Yolo V5x6	69.7	68.5	69.4

Graph.1 Comparison Graphs – Classification



$$"Recall" = \frac{"TP"}{"TP + FN"} \quad (5)$$

mAP: Mean Average Precision (MAP) is a way to compare how good rankings are. It looks at how many good suggestions appear and where they are in a list. MAP at K is the average of the AP at K for all users or searches.

$$"mAP""AP_k"(6) \quad = "\frac{1}{n}" \sum_{"k=1"}^{"k=n"}$$

Table 1 presents the "performance metrics—Precision, Recall, and mAP"—assessed for each method. The Yolo V5x6 attains the greatest scores. Metrics from other methods are also provided for comparison.

In Graph 1, precision is shown in blue, recall in orange, and mAP in green. Compared to the other models, the Extension Yolo V5x6 demonstrates greater performance, attaining the highest values across all metrics. The graphs above graphically represent these results.

V. CONCLUSION

The study effectively overcomes the hassle of modern item reputation techniques and extends its software to first-class -grained objects in far flung sensing contexts. The integration of the eye of the attention mechanism into Yolov5 advanced the capacity of the model for best -grained item popularity. This improvement led to increased identification accuracy, especially in small items, while maintaining the calculation efficiency. The improved Yolov5x6 algorithm has shown excellent performance in fine -grained object recognition with a remarkable 69.45%mAP. Yolov5x6, characterized by sophisticated architectures, various data sets, real -time optimization and continued adaptation, is an impressive solution that offers significant progress in long -distance exploration applications. Using a flask frame together with SQLite to authenticate users has created an intuitive user interface. This allowed users to record photographs and analyzed the system efficiently and presented the final results, showing the practical use of established models. Scientists, experts in the field and government bodies benefit from recognizing the extended object during remote survey, facilitate excellent analysis and decision -making. Development in technology helps developers, end users in remote survey applications and the general public by increasing accuracy and security measures.

Explore the inclusion of advanced neural network designs and attention processes to improve the knowledge of the model in fine -grained items inside the remote survey photos. Expand the scope of the project by integrating more diverse and comprehensive data sets for training, allowing the model more efficiently generalizing in many situations in the real world and increasing its performance in a wider range of long -distance sensing applications. Concentrate on strengthening the algorithm for processing and implementation on marginal devices in real time and facilitate its use in domains such as autonomous systems, supervision and response to low latency disasters.

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 rotationequivariant 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 largescale 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 highresolution 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.