ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-4485-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICASSP48485.2024.10447000

PSEKD: PHASE-SHIFT ENCODED KNOWLEDGE DISTILLATION FOR ORIENTED OBJECT DETECTION IN REMOTE SENSING IMAGES

Chao Wang[1], *Yubiao Yue*[2], *Bingchun Luo*[3], *Yujie Chen*[1], *Jun Xue*[1,*]

[1]School of Computer Science and Technology, Anhui University, Hefei 230601, China
[2] School of Biomedical Engineering, Guangzhou Medical University, Guangzhou, 511436, China
[3] School of Computer Science and Technology, Harbin Institute of Technology, Weihai,264200, China.

ABSTRACT

With the vigorous development of computer vision, oriented object detection has gradually been featured. However, angle boundary discontinuity and its knowledge distillation have been the bottleneck for rotating detection distillation design. In this paper, a novel knowledge distillation method named Phase-shift encoded knowledge distillation (PseKD) is proposed to improve the accuracy of predicting object orientation. Specifically, we design a phase-shift encoded module (PSEM) to solve the problem of angle periodicity due to rotational symmetry in oriented object detection by mapping angles to phases of different frequencies. Secondly, an angle knowledge distillation strategy (AKDS) is designed to guide lightweight models to learn angle knowledge from high-performance models. Experiments on public datasets demonstrate that the proposed method can effectively solve the various periodic fuzzy problems caused by rotational symmetry in remote sensing oriented object detection.

***Index Terms*—** Remote Sensing, Knowledge Distillation, Object Detection

1. INTRODUCTION

With the richness of remote sensing resources and the rapid development of deep learning methods, many methods have been proposed, such as Oriented RCNN [1] and S^2ANet [2]. In order to improve feature extraction and characterization capabilities, these methods are designed to be more complex, which leads to higher demands on computing resources. Especially, rotated object detection is a relatively emerging but challenging area, due to the difficulties of locating arbitrary-oriented objects and separating them effectively from the background. The contradiction between complex angle configurations and limited computing power restricts the application of rotated object detection in remote sensing.

In recent years, several lightweight methods[3] have attracted a lot of attention in the field of deep learning by simplifying convolutional neural network (CNN) structures to reduce model parameters. For example, MobileNet [4] uses lightweight convolution operations to improve the computational speed of the model. However, these rough method of model compression greatly damages the performance of the model. To solve this problem, Han et al. [3] proposed a separable convolution method, analyzed and summarized the efficiency problem in lightweight model design in detail, and provided a solution with both performance and efficiency. Still, there is a significant performance gap between these lightweight models compared to larger models.

Knowledge distillation (KD) [5] achieves more efficient feature refinement by guiding lightweight models to simulate feature extraction of high-performance complex models. Therefore, in addition to compressing the model structure, it is also important to refine the feature representation function of the model. In the object detection task, Li et al.[6] designed supervision of high-level features sampled from the region proposals of the teacher model. Zheng et al. [7] proposed a localization distillation that refined the flexible localization information for the lightweight model. Wang et al. [8] use category information in optical images to guide student networks to fully exploit localization knowledge in teacher networks. These methods provide useful ideas and practical approaches for the refinement and optimization of the model.

However, the above Knowledge Distillation methods designed to focus on common object detection tasks do not adequately consider the unique characteristics presented by objects in remotely sensed images. Firstly, objects in remote sensing images are presented in a top-down perspective, so the multi-directional nature of targets must be fully considered in the detection process. Secondly, targets involved in remote sensing images are usually depicted in the form of oriented bounding boxes, so the bounding box design needs to be reconceptualized to achieve effective knowledge distillation for specific angle localization. Therefore, we proposed a novel knowledge distillation method named phase shifting encoder knowledge distillation (PseKD) to improve the accuracy of the student model in predicting the target direction. The contributions of this work are summarized as follows:

- Our proposed PseKD can reduce the performance gap between lightweight and complex models without any additional inference cost.

* Corresponding authors.

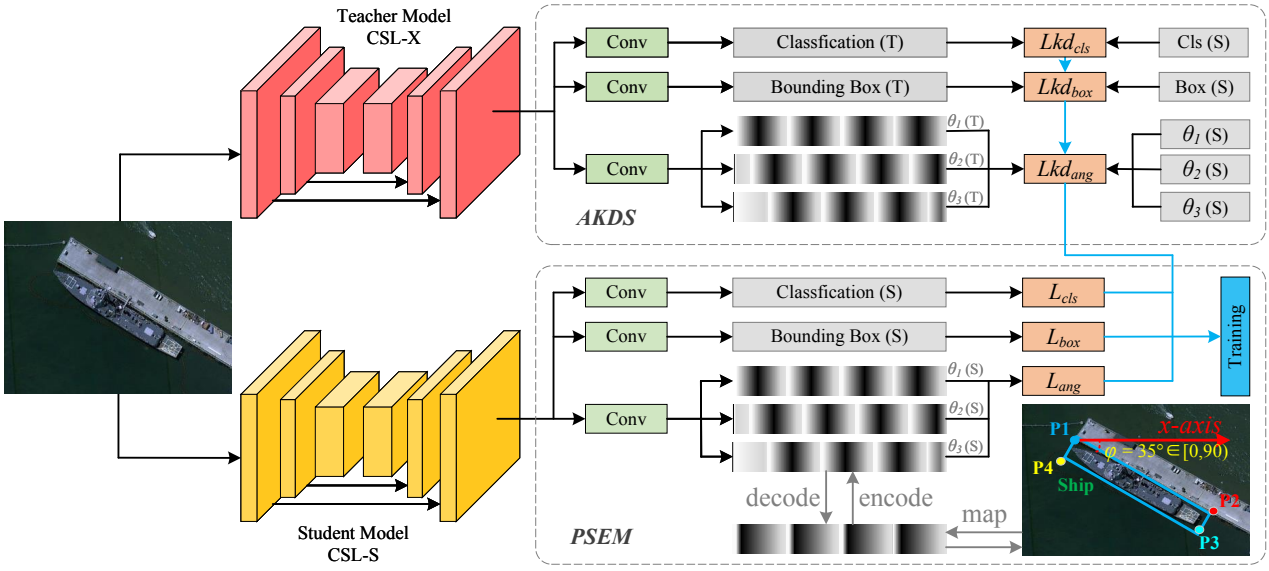

Fig. 1. The framework of the proposed PseKD, where PSEM donates phase-shift encoded module and AKDS donates angle knowledge distillation strategy. CSL-X and CSL-S indicate that CSL[9] uses different backbone networks(CSPDarkNet-X, CSPDarkNet-S).

- In the proposed PseKD, a phase-shift encoded module (PSEM) is developed to extract the arbitrarily oriented angles of remote sensing objects.
- We design an angle knowledge distillation strategy (AKDS) for lightweight models to learn angle knowledge from high-performance models.

2. PROPOSED METHOD

The proposed PseKD consists of two parts: the PSEM, which enhances the direction angle extraction of remote sensing objects, and the AKDS strategy, which distills remote sensing objects' angles from the high-performance model to the lightweight model.

2.1. Phase-shift Encoded Module

In the overhead-view remote sensing images, objects have arbitrary directions. Many studies, such as CSL [9], KFIoU [10], etc., have shown that this directional feature plays an important role in the remote sensing object detection task. However, these methods consider the problems from different perspectives and each has its pros and cons. For example, CSL is simple and stable, but not able to solve the square-like problem, and its performance could be greatly affected by hyper-parameters; KFIoU solve both problems elegantly, but their prediction is relatively inaccurate. Inspired by PSC [11], we propose an idea to utilize phase-shift coding for angle prediction in oriented object detection. The method has the potential to solve both boundary discontinuity and square-like problems.

The overall flowchart of the phase-shift encoded module (PSEM) is shown in Fig. 1. Taking the "long edge 90" angle definition as an example for illustration, symbols can be defined as follows:

- θ: Orientation angle, in range [-π/2,π/2)
- φ: Principal phase, in range [-π,π)
- X: Encoded data, $X = \{x_n | n = 1, 2, \ldots, N_{step}\}$

Map: The cycle of sin or cos is 2 π, whereas a rectangle box is identical to itself when rotated by π, thus a mapping is required to match them, as follows:

$$\varphi = 2\theta \tag{1}$$

Encode: The formula of encoding φ into X can be described as:

$$x_n = \cos\left(\varphi + \frac{2n\pi}{N_{step}}\right) \tag{2}$$

where $n = 1, 2, \ldots, N_{step}$.

To simplify the subsequent description, Eq. 2 is also denoted as $X = f_{enc}(\varphi)$.

Decode: The formula of decoding φ from X can be described as:

$$\varphi = -\arctan \frac{\sum_{n=1}^{N_{step}} x_n \sin\left(\frac{2n\pi}{N_{step}}\right)}{\sum_{n=1}^{N_{step}} x_n \cos\left(\frac{2n\pi}{N_{step}}\right)} \tag{3}$$

The arctan in the formula should be implemented by the arctan2 function so that its output is in the range (-π,π]. Equ. 3 is also denoted as $\varphi = f_{dec}(X)$.

2.2. Angle Knowledge Distillation Strategy

The oriented bounding box is represented by five parameters (x, y, w, h, θ), denoting the box's center coordinates, width, height, and angle, respectively. (x, y, w, h) employs the Dirac delta distribution solely for ground-truth locations, which cannot account for the ambiguity in bounding boxes. We represent the bounding box with a probability distribution, which appropriately captures the uncertainty in its location. Assume $y \in B$ is an edge of a bounding box. The model may provide us with the expected value $\widehat{y}$:

$$\widehat{y} = \int_{y_0}^{y_n} x \times p(x)dx \tag{4}$$

where x is the regression coordinate and $p(x)$ is the corresponding probability. Given the range of the label y with minimum y_0 and maximum $y_n (y_0 \leqslant \widehat{y} \leqslant y_n, n \in N)$, By quantizing the continuous regression range $[y_0, y_n]$ into the uniform discretized variable $[y_0, y_1, \cdots, y_n]$ with n subintervals. Given the discrete distribution property $\sum_{i=0}^{n} p(y_i) = 1$, the estimated regression value $\widehat{y}$ can be represented as:

$$\widehat{y} = \sum_{i=0}^{n} y_i \times p(y_i) \tag{5}$$

Thus, The bounding box with no angle distillation loss is as follows:

$$Lkd_{box} = KLD\left(p^T \| p^S\right) \tag{6}$$

where KLD is KL-Divergence. Define the classification output of the teacher model and student model as c^T and c^S. Then, the classification distillation loss is

$$Lkd_{cls} = KLD\left(c^T \| c^S\right) \tag{7}$$

Afterward, the loss of the angle distillation can be calculated with KLD loss:

$$Lkd_{ang} = \sum_{i=1}^{n} KLD\left({x_i}^T \| {x_i}^S\right) \tag{8}$$

2.3. Loss Function

Oriented object detection contains two sub-tasks, i.e., regression and classification. Deep learning methods deploy two subnetworks to deal with two sub-tasks respectively. Therefore, we use multi-task loss to balance classification, localization, and distillation tasks. Finally, the loss function of PseKD is defined as

$$\begin{aligned} L_{total} =& \lambda_1 * L_{cls} + \lambda_2 * L_{box} \\ &+ \lambda_3 * (Lkd_{cls} + Lkd_{box} + Lkd_{ang}) \end{aligned} \tag{9}$$

where L_{cls}, L_{box} indicate classification and regression loss. For consistency, we set $\lambda_1 = \lambda_2 = 0.4$, $\lambda_3 = 0.2$.

3. EXPERIMENTS AND DISCUSSIONS

3.1. Experimental Conditions

Training setup and Evaluation Metrics. In order to evaluate the performance of the proposed PseKD comprehensively, an experimental platform is used, including a computer with an NVIDIA GeForce RTX 3090 GPU (24GB). We use a stochastic gradient descent (SGD) optimizer with an initial learning rate of 0.01, a momentum of 0.937, and a weight decay of $5{\times}10^{-4}$. Then the entire network is trained in an end-to-end manner. In this paper, all models are evaluated with two object detection metrics: AP_{50} and mean Average Precision (mAP).

Datasets. The widely-used public datasets HRSC2016[12], OGSOD[8] and SRSSD[13] are used. HRSC2016 is a ship detection dataset with a total of 1061 images from 300 $\times$ 300 pixels to 1500 $\times$ 900 pixels, in which the training set, validation set, and testing set include 436, 181, and 444 images, respectively. OGSOD dataset contains 18,331 images with the size of 256 $\times$ 256 pixels, of which more than 48,000 objects. SRSSD dataset contains 666 SAR images with the size of 1024 $\times$ 1024 pixels, in which more than 2,884 objects falling into 6 categories are annotated.

Baseline models. In this work, the high-performance model CSL[9](CSPDarkNet-X) is used as the teacher model, and the lightweight model CSL[9](CSPDarkNet-S) is used as the student model.

Table 1. Ablation experiments on the HRSC2016 dataset

Modules	Teacher CSL-X[9]	Student CSL-S[9]	PSEM	AKDS	mAP
Selectecd Moudule(s)	✓				<u>88.2</u>
		✓			83.9
		✓	✓		85.1
		✓		✓	85.9
		✓	✓	✓	**90.8**

[1] **Best** and <u>*Second best*</u>, Higher mAP is better.

3.2. Ablation Experiments and Discussions

The ablation results of each component of the proposed PseKD on the HRSC2016 dataset are listed in Table 1. The proposed PSEM improves the mean Average Precision (mAP) of CSL-S by 1.20%. The distillation strategy of AKDS and PSEM + AKDS increases the mAP of CSL-S by 2.64% and 4.16%, respectively. The introduced PseKD narrows down the performance disparity between the teacher model and the lightweight student model from 4.3% to -2.6%, thereby facilitating a remarkable 6.9% mAP augmentation for CSL-S without incurring any speed-related trade-offs. The conducted ablation experiments substantiate the efficacy of individual components within the proposed PseKD framework from a quantitative standpoint.

3.3. Comparison with the State-of-the-Arts

Table 2, Table 3, and Table 4 lists the comparative experimental results of the proposed PseKD and high-performance models on the HRSC2016[12] dataset, OGSOD[8] dataset, and SRSSD[13] dataset, respectively.

As shown in Table 2, Roi Trans[14] and R3Det[15] have high accuracy (89.50% mAP, 89.23% mAP), but their inference speed (7.83 FPS, 12.00 FPS) is slow due to their complex structure. KLD[16] and CSL[9] have fast inference speeds (72.34 FPS, 73.53 FPS), but particularly low accuracy (86.40% mAP, 83.90% mAP). As shown in Table 3, our method PseKD achieves the best results (88.3% AP_{50}) in the total 3 categories.

Table 2. Comparative results on the HRSC2016 dataset

Methods	Backbone	mAP	Speed (fps)
Roi Trans[14]	ResNet50	89.50	7.83
Gliding Vertex[17]	ResNet101	87.33	10.52
R3Det[15]	ResNet101	89.23	12.00
KLD[16]	CSPDarkNet-S	86.40	72.34
GGHL[18]	DarkNet53	87.30	42.30
CSL[9] (Baseline)	CSPDarkNet-S	83.90	**73.53**
PSEC (Ours)	CSPDarkNet-S	**90.80**	**73.53**

[1] **Best** and *Second best*, Higher mAP is better.

To verify our proposed PseKD not only in optical remote sensing images but also in SAR images, We set up experiments on the SRSSD[13] dataset as shown in Table 4. The categories of objects in SRSSD include Ore Oil (OO), Fishing (FH), LawEnforce (LE), Dredger (DD), Cell Container (CC), and Container (CT). As the experiments show, the S^2ANet[2] and Gliding Vertex[17] can barely reach 51.0% and 50.2% AP_{50}. However, CSL[9] only achieve 44.1% AP_{50}. This demonstrates SRSSD[13] is challenging for recently oriented detectors and has a huge room for growth. Compared with other methods, PseKD achieves dominant accuracy of 56.3% AP_{50}, which verifies the robustness and generalization of our methods.

Our method PseKD not only achieves good performance on the evaluation metrics mAP and AP_{50} but also gains significant improvements in perceptual quality. As shown in Fig. 2, our PseKD can obtain more guidance information from different knowledge in the high-performance model, and reduce false negative and false positive. Especially, our method

Table 3. Comparative results on the OGSOD(RGB) dataset

Methods	Backbone	Oil Tank	Bridge	Harbor	AP_{50}
Oriented RCNN[1]	ResNet-50	62.0	88.7	90.8	80.5
S^2ANet[2]	ResNet-50	60.9	89.4	90.8	80.4
R3Det[15]	ResNet-50	70.5	88.8	90.4	83.3
Gliding Vertex[17]	ResNet-50	61.6	78.9	89.2	76.5
KFIoU[10]	ResNet-50	57.7	89.2	90.8	79.2
CSL [9](Baseline)	CSPDarkNet-S	74.1	87.4	93.7	85.1
PseKD (Ours)	CSPDarkNet-S	**78.2**	**90.9**	**95.9**	**88.3**

[1] **Best** and *Second best*, Higher AP_{50} is better.

Table 4. Comparative results on the SRSSD dataset

Methods	OO	FH	LE	DD	CC	CT	AP_{50}
S^2ANet[2]	54.9	27.9	27.3	71.5	79.7	44.7	51.0
Gliding Vertex[17]	45.1	**30.4**	27.3	71.3	**80.0**	47.2	50.2
Oriented RCNN[1]	57.8	21.8	27.3	78.4	78.1	**55.9**	53.2
KFIoU[10]	55.7	27.6	3.4	76.6	73.5	51.2	48.0
Oriented Reppoints[19]	**58.3**	18.4	20.0	77.9	72.5	42.6	48.2
CSL[9] (Baseline)	20.3	13.6	60.8	75.1	62.3	32.5	44.1
PseKD (Ours)	41.1	25.3	**78.6**	**82.1**	67.1	43.4	**56.3**

[1] **Best** and *Second best*, Higher AP_{50} is better.

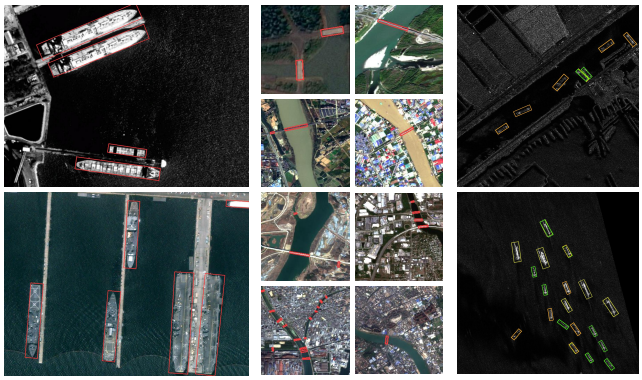

Fig. 2. The visual experimental results of the proposed PseKD on (a) the HRSC2016 dataset, (b) the OGSOD dataset, and (c) the SRSSD dataset..

achieves a huge improvement in SAR images. In addition, the only hyper-parameter N_{step} is an integer greater than or equal to 3, so we evaluate several N_{step} values most commonly used in the phase-shift technique, and the results, which are obtained on PseKD, are shown in Table 5. Overall, increasing the value of N_{step} does not yield substantial advantages.

Table 5. Performance under different N_{step} values

Metrics	N_{step}			
	step = 3	step = 4	step = 5	step = 6
HRSC2016 (mAP)	90.8	90.6	90.5	89.9

4. CONCLUSIONS

In this paper, a novel knowledge distillation method PseKD has been proposed to improve the performance of remote sensing object detection models. In the proposed PseKD, the developed PSEM enhances the arbitrary-oriented angle extraction of the lightweight model, and the designed AKDS guides the lightweight model to learn the angle knowledge from the high-performance model. The experiments on public datasets have demonstrated that each component proposed in PseKD is valid, and the detection performance of the lightweight model can be greatly improved without any inference costs by using the proposed PseKD.

5. REFERENCES

[1] Xingxing Xie, Gong Cheng, Jiabao Wang, Xiwen Yao, and Junwei Han, "Oriented r-cnn for object detection," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 3520–3529.

[2] Jiaming Han, Jian Ding, Jie Li, and Gui-Song Xia, "Align deep features for oriented object detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–11, 2021.

[3] Zhanchao Huang, Wei Li, Xiang-Gen Xia, Hao Wang, Feiran Jie, and Ran Tao, "Lo-det: Lightweight oriented object detection in remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2022.

[4] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.

[5] Ji-Hoon Bae, Doyeob Yeo, Junho Yim, Nae-Soo Kim, Cheol-Sig Pyo, and Junmo Kim, "Densely distilled flow-based knowledge transfer in teacher-student framework for image classification," *IEEE Transactions on Image Processing*, vol. 29, pp. 5698–5710, 2020.

[6] Quanquan Li, Shengying Jin, and Junjie Yan, "Mimicking very efficient network for object detection," in *Proceedings of the ieee conference on computer vision and pattern recognition*, 2017, pp. 6356–6364.

[7] Zhaohui Zheng, Rongguang Ye, Qibin Hou, Dongwei Ren, Ping Wang, Wangmeng Zuo, and Ming-Ming Cheng, "Localization distillation for object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[8] Chao Wang, Rui Ruan, Zhicheng Zhao, Chenglong Li, and Jin Tang, "Category-oriented localization distillation for sar object detection and a unified benchmark," *IEEE Transactions on Geoscience and Remote Sensing*, 2023.

[9] Xue Yang and Junchi Yan, "Arbitrary-oriented object detection with circular smooth label," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VIII 16*. Springer, 2020, pp. 677–694.

[10] Xue Yang, Yue Zhou, Gefan Zhang, Jitui Yang, Wentao Wang, Junchi Yan, Xiaopeng Zhang, and Qi Tian, "The kfiou loss for rotated object detection," *arXiv preprint arXiv:2201.12558*, 2022.

[11] Yi Yu and Feipeng Da, "Phase-shifting coder: Predicting accurate orientation in oriented object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 13354–13363.

[12] Zikun Liu, Liu Yuan, Lubin Weng, and Yiping Yang, "A high resolution optical satellite image dataset for ship recognition and some new baselines," in *International conference on pattern recognition applications and methods*. SciTePress, 2017, vol. 2, pp. 324–331.

[13] Songlin Lei, Dongdong Lu, Xiaolan Qiu, and Chibiao Ding, "Srsdd-v1. 0: A high-resolution sar rotation ship detection dataset," *Remote Sensing*, vol. 13, no. 24, pp. 5104, 2021.

[14] Jian Ding, Nan Xue, Yang Long, Gui-Song Xia, and Qikai Lu, "Learning roi transformer for oriented object detection in aerial images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2849–2858.

[15] Xue Yang, Junchi Yan, Ziming Feng, and Tao He, "R3det: Refined single-stage detector with feature refinement for rotating object," in *Proceedings of the AAAI conference on artificial intelligence*, 2021, vol. 35, pp. 3163–3171.

[16] Xue Yang, Xiaojiang Yang, Jirui Yang, Qi Ming, Wentao Wang, Qi Tian, and Junchi Yan, "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.

[17] Yongchao Xu, Mingtao Fu, Qimeng Wang, Yukang Wang, Kai Chen, Gui-Song Xia, and Xiang Bai, "Gliding vertex on the horizontal bounding box for multi-oriented object detection," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 4, pp. 1452–1459, 2020.

[18] Zhanchao Huang, Wei Li, Xiang-Gen Xia, and Ran Tao, "A general gaussian heatmap labeling for arbitrary-oriented object detection," *arXiv e-prints*, pp. arXiv–2109, 2021.

[19] Wentong Li, Yijie Chen, Kaixuan Hu, and Jianke Zhu, "Oriented reppoints for aerial object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 1829–1838.