Learning to Reduce Information Bottleneck for Object Detection in Aerial Images

Yuchen Shen, Dong Zhang, Zhihao Song, Xuesong Jiang, and Qiaolin Ye, Member, IEEE

Abstract—Object detection in aerial images is a critical and essential task in the fields of geoscience and remote sensing. Despite the popularity of computer vision methods in detecting objects, these methods have been faced with significant limitations of aerial images such as appearance occlusion and variable object sizes. In this work, we explore the limitations of conventional neck networks in object detection by analyzing information bottlenecks. We propose an enhanced neck network to address the information deficiency issue in current neck networks. Our proposed neck network, which serves as a bridge between the backbone network and the head network, comprises a global semantic network (GSNet) and a feature fusion refinement module (FRM). The GSNet is designed to perceive contextual surroundings and propagate discriminative knowledge through a bidirectional global pattern. The FRM is developed to exploit different levels of features to capture comprehensive location information. We validate the efficacy and efficiency of our approach through experiments conducted on two challenging datasets, DOTA and HRSC2016. Our method outperforms existing approaches in terms of accuracy and complexity, demonstrating the superiority of our proposed method.

Index Terms—Aerial image processing, information bottleneck, object detection, remote sensing scene recognition.

I. INTRODUCTION

The task of object detection in aerial images is a significant research challenge in the fields of computer vision and remote sensing [1]. This task involves classifying objects and assigning a unique bounding box to each object in the given image, which is critical for numerous practical applications, such as environmental monitoring and urban planning [2]. However, the existing computer vision methods for object detection in aerial images suffer from an unsatisfactory performance, largely due to the specific characteristics of aerial images, such as occlusion, variable size, and highly complex backgrounds [3], [4].

In this letter, as illustrated in Fig. 1, we aim to address the information bottleneck problem in the aerial image detection task, which is caused by the insufficient feature representation in conventional neck networks. Specifically, we propose an enhanced neck network [5] based on the feature pyramid network [6], comprising a global semantic network (GSNet) and a feature fusion refinement module (FRM). GSNet uses large-size kernels to enlarge the receptive field via a global bilateral scanning operation, encouraging nondiscriminative features to gain discriminative capability through sensing adjacent regions with a high response [7]. FRM leverages the obtained multiscale localization maps to extract features of the target object accurately, improving the quality of detection predictions. To demonstrate the effectiveness and efficiency of our approach, we conduct experiments on two datasets, namely DOTA [8] and HRSC2016 [9], using Fast R-CNN [10] and RetinaNet [11]. Our proposed GSNet with FRM achieves state-of-the-art performance in both datasets, with an mAP of 79.37% and 74.49% for DOTA and an mAP of 90.50% and 90.47% for HRSC2016. Our contributions include the identification of the limitations of conventional neck networks, the proposal of GSNet and FRM to reduce the information bottleneck and improve the feature representation ability of the enhanced FPN-based neck network, and the achievement of competitive performance on two challenging datasets.
represented as a Markov chain, where the input variable $X$ is transformed into a series of latent representations $Y_m$ by $M$ layers of the network, and ultimately to the output variable $Z$. The mutual information between $X$ and $Y_m$ in the chain satisfies the data processing inequality. The information bottleneck approach aims to learn an optimal representation $Y^*$ that maximizes the mutual information between $Y$ and $Z$, while minimizing the mutual information between $Y$ and $X$ [12]. This is achieved by solving an optimization problem with a tradeoff parameter $\beta$. The information bottleneck approach is useful for the compression of input images and preservation of task-relevant information in training detection models. However, the localization map generated by the detector tends to highlight only small discriminative parts of the target object, such as the dog’s head and paws, leading to low-quality detection predictions due to the omission of nondiscriminative information that is essential for localization [13], [14]. Several prior works have employed information bottlenecks to improve the performance of detection models.

B. Reducing Information Bottleneck

The insufficiency of the effective receptive field of conventional neck networks, particularly small kernels, causes an information bottleneck, which fails to capture contextual dependencies and prevents the transmission of discriminative information to adjacent regions with low response [15]. To address this issue, we propose an enhanced FPN-based neck approach that aims to learn an optimal representation $Y^*$ to address this issue, we propose an enhanced FPN-based neck approach that aims to learn an optimal representation $Y^*$ to highlight only small discriminative parts of the target object, such as the dog’s head and paws, leading to low-quality detection predictions due to the omission of nondiscriminative information that is essential for localization [13], [14]. Several prior works have employed information bottlenecks to improve the performance of detection models.

II. METHODOLOGY

A. Revisiting Information Bottleneck

The study presented in this letter revisits the concept of information bottleneck, which was previously introduced to explore the information flow between layers in a convolutional neural network (CNN) model [7]. The CNN model can be represented as a Markov chain, where the input variable $X$ is transformed into a series of latent representations $Y_m$ by $M$ layers of the network, and ultimately to the output variable $Z$. The mutual information between $X$ and $Y_m$ in the chain satisfies the data processing inequality. The information bottleneck approach aims to learn an optimal representation $Y^*$ that maximizes the mutual information between $Y$ and $Z$, while minimizing the mutual information between $Y$ and $X$ [12]. This is achieved by solving an optimization problem with a tradeoff parameter $\beta$. The information bottleneck approach is useful for the compression of input images and preservation of task-relevant information in training detection models. However, the localization map generated by the detector tends to highlight only small discriminative parts of the target object, such as the dog’s head and paws, leading to low-quality detection predictions due to the omission of nondiscriminative information that is essential for localization [13], [14]. Several prior works have employed information bottlenecks to improve the performance of detection models.

B. Reducing Information Bottleneck

The insufficiency of the effective receptive field of conventional neck networks, particularly small kernels, causes an information bottleneck, which fails to capture contextual dependencies and prevents the transmission of discriminative information to adjacent regions with low response [15]. To address this issue, we propose an enhanced FPN-based neck network, depicted in Fig. 2. Typically, the backbone of the network, following [16], is the ImageNet [17] pretrained ResNet [18]. Our proposed GSNet (in Section II-C) employs large-size kernels through a bilateral scanning operation to significantly expand the receptive fields [19], enabling the acquisition of rich contextual features. We also introduce FRM (in Section II-C) to better utilize the high-quality object localization maps obtained at different scales. The final prediction results are generated using head networks such as Faster R-CNN [10] and RetinaNet [11].

C. Global Semantic Network

We propose GSNet, which is depicted in Fig. 2(b), to reduce feature loss and improve the model’s ability to highlight nondiscriminative regions [19]. GSNet consists of two steps.

Step-1: To expand discriminative object parts, we use larger kernels to increase the effective receptive field. We decompose the large-size 2-D kernel into two parallel branches, each of which is composed of 1-D convolutions along two different directions (lengthwise and widthwise) to reduce complexity. The complementary features provided by each branch are merged using elementwise addition to generate a refined feature map. Besides, we remove pooling layers and fully connected layers to retain as much location information in the feature maps as possible. The symmetric, linear 1-D filters, which are transformed from large 2-D kernels, overcome the limitations of additional computational overhead brought by large-kernel CNNs and enable the network to perceive high-response context, transferring discriminative knowledge to adjacent object regions with low response. The first step can be formulated as

$$F(X) = \text{Conv}(\text{Conv}(X)^T) + \text{Conv}(\text{Conv}(X))^T$$

where $X \in \mathbb{R}^{H \times W \times C}$ is the input feature map extracted from the backbone feature pyramid, Conv represents the 1-D convolution, and $+$ is the elementwise addition.

Step-2: Small objects in aerial images are difficult to detect. Kernels of small size focus on catching details and structural features that are necessary for the detection of small objects. We employ two successive standard kernels (e.g., $3 \times 3$ convolutions) to capture fine object boundaries and improve learned feature representations. To boost performance, we form the residual structure in the network by reusing the output from the backbone feature pyramid $X$ and from the first step $F(X) \in \mathbb{R}^{H \times W \times 16}$. Formally,

$$F^G(X) = F(X) + F^R(X) + X$$

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
\[ F^R(X) = \text{Conv} \ 2D(\delta(\text{Conv} \ 2D(F(X)))) \]  
(3)

where \( F^R(\cdot) \) is the residual branch, and \( \delta \) is the ReLU function \([20]\). In summary, GSNet enables the network to access features of the target object more precisely by mining both global and local spatial information, producing dense localization maps of different scales.

D. Fusion Refinement Module

As shown in Fig. 2(c), FRM is achieved by integrating the outputs of GSNet. FRM first concatenates the localization maps with different scales to learn channel-wise dependencies, which could prevent corrupting features that are beneficial for prediction. Specifically, a feature map \( F^G_{i+1}(X) \) from the \((i+1)\)th level of GSNet is resized via bilinear interpolation with a factor of 2 to obtain a resized feature map \( F^G_i(X) \) for \( i = 2, 3, 4 \). The feature map \( X_i \) from the corresponding layer of the backbone is then concatenated with \( F^G_i(X) \) and \( F^R_i(X) \) along the channel dimension, resulting in the concatenated feature map \( F^{Cat}_i \) given by

\[ F^{Cat}_i = \begin{bmatrix} F^G_{i+1}(X); F^G_i(X); X_i \end{bmatrix} \ i = 2, 3, 4 \]  
(4)

where \([\cdot; \cdot; \cdot] \) denotes the concatenation operation. To reduce the number of model parameters, two \( 1 \times 1 \) convolution layers are adopted at intervals to reduce the channel dimension sequentially to 256 and 64. Subsequently, a series of convolutions are implemented to remove the aliasing effect resulting from the interpolation. This can be expressed mathematically as

\[ F^F_i = f^{1 \times 1}(f^{3 \times 3}(f^{1 \times 1}(f^{3 \times 3}(f^{1 \times 1}(F^{Cat}_i)))))) \]  
(5)

where \( f^{1 \times 1} \) and \( f^{3 \times 3} \) denote \( 1 \times 1 \) and \( 3 \times 3 \) convolutions, respectively. The merged prediction maps are fed into a \( 3 \times 3 \) convolution layer to construct the enhanced feature pyramid \( F^F_1, F^F_2, F^F_3, F^F_4 \). FRM refines the initial maps generated by the backbone, which only cover sparse parts of the object, into accurate and reliable object localization maps that can be beneficial for dense prediction tasks in practice.

III. EXPERIMENTS

A. Datasets and Evaluation Metric

DOTA \([8]\) contains 2806 aerial images of varying sizes and 15 categories, while the HRSC2016 \([9]\) has 1061 high-resolution images. We randomly choose 1/2 for training, 1/6 for validation, and 1/3 for testing, after cropping them into \(1024 \times 1024 \) patches. We evaluate the model using mAP(%) and three metrics: giga floating-point operations per second (GFLOPs), Params(M), and frames per second (FPS).

B. Settings

1) Baselines: We adopt two well-known object detection models, Faster R-CNN \([10]\) and RetinaNet \([11]\), as our baselines, with ResNet101 and ResNet50 used as backbone networks. RoI-Transformer \([16]\) and RotatedRetinaNet \([11]\) are used to develop the rotated head. All settings are based on official codes for a fair comparison.

2) Training Details: Standard SGD with an initial learning rate of 0.005 and 0.0025 is used for the two baselines, respectively. We set the weight decay to 0.0001 and the momentum to 0.9. The models are trained on RTX 3060 with a batch size of 2 for DOTA and HRSC2016 datasets.

![Fig. 3. Class activation map comparisons with the baseline model. (a) Input image. (b) Baseline. (c) Ours.](image)

![Fig. 4. Visualization comparison of oriented detection results between the baseline (blue boxes) and “Faster R-CNN + Ours” (red boxes) on DOTA \([8]\).](image)

| ABLATION STUDY RESULTS ON DOTA \([8]\) AND HRSC2016 \([9]\). R-101/-50 Denotes ResNet-101/-50, Respectively. “Fast” and “Reti” Are Faster R-CNN \([10]\) and RetinaNet \([11]\), Respectively. |
|-----------------|-------------|-------------|-------------|-------------|-------------|
| Methods         | Backbone    | GSNet | FRM | DOTA      | HRSC       | Params | GFLOPs   |
|-----------------|-------------|-------|-----|-----------|------------|--------|----------|
| Fast [10]       | R-101       | x     | x   | 73.09     | 88.36      | 74.12  | 289.19   |
| Fast [10]       | R-101       | √     |     | 74.49 [0.44, 70.88] | 74.61      | 293.98 |
| Fast [10]       | R-101       | √     |     | 75.50 [0.24, 78.93] | 76.45      | 325.66 |
| Fast [10]       | R-101       | √     |     | 76.23 [0.14, 89.50] | 77.91      | 338.14 |
| Reti [11]       | R-50        | x     | x   | 66.79     | 86.90      | 36.42  | 215.92   |
| Reti [11]       | R-50        | √     |     | 70.78 [1.99, 87.43] | 79.13      | 221.54 |
| Reti [11]       | R-50        | √     |     | 71.10 [2.31, 87.79] | 80.40 [1.25] | 38.34  | 226.03   |
| Reti [11]       | R-50        | √     |     | 71.61 [2.82, 88.40] | 81.50      | 39.69  | 231.66   |

C. Ablation Study

1) Effectiveness of the Proposed Modules: The effectiveness of our proposed components is validated on a two-stage detector Faster R-CNN \([10]\). For results on DOTA \([8]\), as shown in Table I, it is observed that the bounding box mAP is improved by 1.40%, 2.41%, and 3.14% after adding GSNet, FRM, and “GSNet + FRM” separately. Similarly, for HRSC \([9]\), it can be also seen that the mAP is improved by 0.88%, 0.93%, and 1.14% when GSNet, FRM, and “GSNet + FRM” are used, respectively.

2) Effectiveness on Different Baselines: To further validate the versatility of the proposed units, we equipped them with another one-stage detector RetinaNet \([11]\). From the sixth and seventh rows of Table I, we observe that the same situation occurs in RetinaNet. For DOTA \([8]\), after adding GSNet, FRM, and “GSNet + FRM,” mAP is significantly improved by 1.99%, 2.31%, and 2.82%, respectively. For HRSC \([9]\), the improvements are 0.51%, 0.89%, and 1.50% mAP. From a general analysis of these improvements, it can be demonstrated that our proposed modules play different but complementary roles. Concretely, the GSNet module perceives the contextual surroundings via a global bilateral scanning operation, which allows the model to highlight nondiscriminative regions, such that features of the target object can be captured more precisely. Furthermore, the FRM module makes better use of the obtained features at different levels, encouraging each layer to preserve the comprehensive location information. As a result,
these two proposed modules complement each other and can jointly produce reliable detection results.

3) Complexity Analysis: The evaluation of model efficiency requires the use of metrics such as Params and GFLOPs. A decrease in the number of Params and GFLOPs results in higher efficiency and faster inference speed. As presented in Table I, the GSNet module increases the Params and GFLOPs slightly in comparison to the baseline model. This is because GSNet transforms the large-size 2-D kernel into a pair of 1-D ones, which effectively controls the increase in model parameters and computational costs. The use of GSNet and FRM increases the computational complexity, with an average increment of 3.53 M Params and 32 GFLOPs on these two detectors. Nonetheless, we believe the performance improvement justifies the cost. Our model outperforms TransConvNet [21] and rotated varied-size attention (RVSA) [22] with a significantly higher performance of 79.37% mAP, fewer Params of 77.91 M, and lower computational costs (338.14 GFLOPs). Specifically, TransConvNet [21] achieves 78.41% mAP, 90 M Params, and 752 GFLOPs, whereas RVSA [22] achieves 78.75% mAP, 114 M Params, and 413.29 GFLOPs.

4) Visualizations: The comparison of attention maps is depicted in Fig. 3. It is observed that the attention map in Fig. 3(b) highlights only the most discriminative regions, such as the airplane’s head, whereas the attention map in Fig. 3(c) obtained with the aid of GSNet and FRM focuses on more object-related regions, providing a full view of the airplane. This result signifies the efficacy of our proposed approach in expanding the network’s attention to nondiscriminative regions, which aids in dense prediction in object detection. The visual comparisons presented in Fig. 4 demonstrate the effectiveness of our improved model over the Faster R-CNN [10] in addressing issues such as incorrect detection of bounding boxes on a harbor and a tennis court, and detecting occluded small vehicles. To sum up, our approach enhances the localization ability of the model by encouraging each layer to preserve more features and thereby reducing the information bottleneck.

D. Comparisons With State-of-the-Arts

According to the results presented in Table II, our proposed “GSNet + FRM” model achieves the highest performance among some of the state-of-the-art methods, with 79.37% mAP for two-stage oriented object detection and 74.49% mAP for one-stage oriented object detection. These results are indicative of our model’s superior feature representation capabilities. With an input image size of 1024 × 1024, our model attains a frame rate of 14.0 and 20.0 FPS on two RTX 3060 GPUs, respectively. The oriented object detection results on the DOTA test set are visualized in Fig. 5. As can be seen from Table III, our results on the HRSC data set also outperform the state-of-the-art methods, achieving the highest performance with comparable inference speed (i.e., 90.50% mAP in 14.0 FPS and 90.47% mAP in 25.1 FPS).
IV. CONCLUSION

We introduce GSNet and FRM to enhance the neck network for object detection by addressing the information bottleneck issue, achieving superior performance on challenging datasets. However, the proposed network still faces difficulty detecting tiny targets. Future work will focus on addressing this limitation and applying our methods to other computer vision tasks like semantic segmentation and video object detection. In addition, “GSNet + FRM” can be flexibly integrated as a block into any CNNs with the information bottleneck, including the backbone and head network, which makes it more friendly to various applications.

REFERENCES

[1] Z. Wei et al., “Learning calibrated-guidance for object detection in aerial images,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 15, pp. 2721–2733, 2022.
[2] D. Zhang, N. Li, and Q. Ye, “Positional context aggregation network for remote sensing scene classification,” IEEE Geosci. Remote Sens. Lett., vol. 17, no. 6, pp. 943–947, Jun. 2020.
[3] Q. Zhang, S. Lu, and W. Zhang, “CAD-Net: A context-aware detection network for objects in remote sensing imagery,” IEEE Trans. Geosci. Remote Sens., vol. 57, no. 12, pp. 10015–10024, Dec. 2019.
[4] L. Fu, D. Zhang, and Q. Ye, “Recurrent thrifty attention network for remote sensing scene recognition,” IEEE Trans. Geosci. Remote Sens., vol. 59, no. 10, pp. 8257–8268, Oct. 2021.
[5] D. Zhang, H. Zhang, J. Tang, M. Wang, X. Hua, and Q. Sun, “Feature pyramid transformer,” in Proc. ECCV, 2020, pp. 323–339.
[6] Y. Zhao, R. Han, and Y. Yao, “A new feature pyramid network for object detection,” in Proc. Int. Conf. Virtual Reality Intell. Syst. (ICVRIS), Sep. 2019, pp. 428–431.
[7] A. M. Saxe et al., “On the information bottleneck theory of deep learning,” J. Stat. Mech., Theory Exp., vol. 2019, no. 12, Dec. 2019, Art. no. 124020.
[8] G.-S. Xia et al., “DOTA: A large-scale dataset for object detection in aerial images,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 3974–3983.
[9] Z. Liu, L. Yuan, L. Weng, and Y. Yang, “A high resolution optical satellite image dataset for ship recognition and some new baselines,” in Proc. 6th Int. Conf. Pattern Recognit. Appl. Methods, 2017, pp. 324–331.
[10] 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., 2015, pp. 1–9.
[11] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 2999–3007.
[12] N. Tishby and N. Zaslavsky, “Deep learning and the information bottleneck principle,” in Proc. IEEE Inf. Theory Workshop (ITW), May 2015, pp. 1–5.
[13] D. Zhang, H. Zhang, J. Tang, X.-S. Hua, and Q. Sun, “Causal intervention for weakly-supervised semantic segmentation,” in Proc. NIPS, 2020, pp. 655–666.
[14] G. Zhao, J. Wang, and Z. Zhang, “Random shifting for CNN: A solution to reduce information loss in down-sampling layers,” in Proc. 26th Int. Joint Conf. Artif. Intell., Aug. 2017, pp. 3476–3482.
[15] W. Luo, Y. Li, R. Urtasun, and R. Zemel, “Understanding the effective receptive field in deep convolutional neural networks,” in Proc. NIPS, 2016, pp. 1–9.
[16] J. Ding, N. Xue, Y. Long, G.-S. Xia, and Q. Lu, “Learning RoI transformer for oriented object detection in aerial images,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 2844–2853.
[17] O. Russakovsky et al., “ImageNet large scale visual recognition challenge,” Int. J. Comput. Vis., vol. 115, no. 3, pp. 211–252, Dec. 2015.
[18] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770–778.
[19] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun, “Large kernel matters—Improve semantic segmentation by global convolutional network,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 1743–1751.
[20] X. Glorot, A. Bordes, and Y. Bengio, “Deep sparse rectifier neural networks,” in Proc. 14th Int. Conf. Artif. Intell. Statist., 2011, pp. 315–323.
[21] X. Liu, S. Ma, L. He, C. Wang, and Z. Chen, “Hybrid network model: TransConvNet for oriented object detection in remote sensing images,” Remote Sens., vol. 14, no. 9, p. 2090, Apr. 2022.
[22] D. Wang et al., “Advancing plain vision transformer towards remote sensing foundation model,” IEEE Trans. Geosci. Remote Sens., early access, Nov. 21, 2022, doi: 10.1109/TGRS.2022.322818.
[23] R. Zhang, X. Zhang, Y. Zheng, D. Wang, and L. Hua, “MScNet: A multilevel stacked context network for oriented object detection in optical remote sensing images,” Remote Sens., vol. 14, no. 20, p. 5066, Oct. 2022.
[24] J. Zhou, K. Feng, W. Li, J. Han, and F. Pan, “TSNet: Two-stage sample selective strategy for rotating object detection,” Neurocomputing, vol. 501, pp. 753–764, Aug. 2022.
[25] X. Xie, G. Cheng, J. Wang, X. Yao, and J. Han, “Oriented R-CNN for object detection,” in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2021, pp. 3500–3509.
[26] W. Qian, X. Yang, S. Peng, X. Zhang, and J. Yan, “RSDet++: Point-based modulated loss for more accurate rotated object detection,” IEEE Trans. Circuits Syst. Video Technol., vol. 32, no. 11, pp. 7869–7879, Nov. 2022.
[27] X. Yang, J. Yan, W. Liao, X. Yang, J. Tang, and T. He, “SCRDet++: Detecting small, cluttered and rotated objects via instance-level feature denoising and rotation loss smoothing,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 2, pp. 2384–2399, Feb. 2023.
[28] J. Yi, P. Wu, B. Liu, Q. Huang, H. Qu, and D. Metaxas, “Oriented object detection in aerial images with box boundary-aware vectors,” in Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV), Jan. 2021, pp. 2149–2158.
[29] T. W. Zhang et al., “Performance releaser with smart anchor learning for arbitrary-oriented object detection,” CAAI Trans. Intell. Technol., Oct. 2022.
[30] X. Pan et al., “Dynamic refinement network for oriented and densely packed object detection,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2020, pp. 11204–11213.
[31] Q. Ming, L. Miao, Z. Zhou, and Y. Dong, “CFC-Net: A critical feature capturing network for arbitrary-oriented object detection in remote-sensing images,” IEEE Trans. Geosci. Remote Sens., vol. 60, 2022, Art. no. 560814.
[32] L. Hou, K. Lu, and J. Xue, “Refined one-stage oriented object detection method for remote sensing images,” IEEE Trans. Image Process., vol. 31, pp. 1545–1558, 2022.
[33] X. Yang, Q. Liu, J. Yan, A. Li, Z. Zhang, and G. Yu, “R3Det: Refined single-stage detector with feature refinement for rotating object,” in Proc. AAAI, 2021, pp. 3163–3171.
[34] Q. Ming, L. Miao, Z. Zhou, X. Yang, and Y. Dong, “Optimization for arbitrary-oriented object detection via representation invariance loss,” IEEE Geosci. Remote Sens. Lett., vol. 19, pp. 1–5, 2022.
[35] Q. Ming, Z. Zhou, L. Miao, H. Zhang, and L. Li, “Dynamic anchor learning for arbitrary-oriented object detection,” in Proc. AAAI, 2021, pp. 2355–2363.
[36] J. Zhang, Huang, Y. Li, and B. Pan, “Oriented ship detection based on intersecting circle and deformable RoI in remote sensing images,” Remote Sens., vol. 14, no. 19, p. 4749, Sep. 2022.
[37] J. Han, J. Ding, J. Li, and G.-S. Xia, “Align deep features for oriented object detection,” IEEE Trans. Geosci. Remote Sens., vol. 60, 2022, Art. no. 5602511.
[38] L. Hou, K. Lu, J. Xue, and Y. Li, “Shape-adaptive selection and measurement for oriented object detection,” in Proc. AAAI, 2022, pp. 923–932.
[39] W. Li, Y. Chen, K. Hu, and J. Zhu, “Oriented RepPoints for aerial object detection,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2022, pp. 1819–1828.
[40] J. Han, J. Ding, N. Xue, and G.-S. Xia, “ReDet: A rotation-equivariant detector for aerial object detection,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2021, pp. 2785–2794.