

基于自适应空间融合的遥感图像目标检测器

余高枫¹, 叶佩华²

¹温州大学计算机与人工智能学院, 浙江 温州

²湖州师范学院信息技术中心, 浙江 湖州

收稿日期: 2023年3月18日; 录用日期: 2023年4月17日; 发布日期: 2023年4月24日

摘要

随着遥感技术的蓬勃发展, 日益庞大的遥感数据使得传统的图像处理手段已经不足以满足需求, 因此, 将深度学习目标检测技术应用于遥感领域成为了最优解决方案。然而, 由于遥感图像小目标排列密集、尺度变化剧烈, 使得直接将适用于自然场景的目标检测算法迁移到遥感领域的效果不佳。本文选择YOLOX为基础网络进行改进, 在特征提取网络后加入自适应空间特征融合结构, 将深层特征信息与浅层特征信息融合, 提升小目标的识别率。此外, 本研究对样本标签匹配策略进行了优化, 以解决高长宽比类目标物体角度偏移敏感的问题, 并通过修改损失函数降低正负样本不平衡问题带来的影响。我们在遥感图像数据集DOTA上进行训练和测试, 实验结果表明, 改进的YOLOX算法检测效果更好, *mAP*达到了79.07%, 比YOLOX提高了2.75%。另外, 在HRSC2016数据集上也进行了实验, 实验证明了模型具有优秀的鲁棒性。

关键词

遥感图像, 目标检测, 自适应空间融合结构, 样本标签匹配策略

A Remote Sensing Image Object Detector Based on Adaptive Spatial Fusion

Gaofeng Yu¹, Peihua Ye²

¹College of Computer Science and Artificial Intelligence, Wenzhou University, Wenzhou Zhejiang

²Information Technology Center, Huzhou University, Huzhou Zhejiang

Received: Mar. 18th, 2023; accepted: Apr. 17th, 2023; published: Apr. 24th, 2023

Abstract

With the rapid development of remote sensing technology, the increasingly large remote sensing data has rendered traditional image processing techniques inadequate to meet the demands. As a result,

applying deep learning object detection techniques to the remote sensing field has become the optimal solution. However, the dense arrangement of small objects and drastic scale variations in remote sensing images make the direct migration of object detection algorithms applicable to natural scenes to the field of remote sensing ineffective. In this paper, we improve the YOLOX algorithm by incorporating an adaptive spatial feature fusion structure after the feature extraction network, which fuses deep and shallow features to enhance the recognition rate of small objects. In addition, this study optimized the sample label assignment strategy to address the issue of angle offset sensitivity for narrow class target objects with high aspect ratios. Furthermore, the impact of the imbalance between positive and negative samples was mitigated by modifying the loss function. We conduct experiments on the DOTA remote sensing image dataset for training and testing. The results demonstrate that the improved YOLOX algorithm achieves better detection performance, with a *mAP* of 79.07%, which is 2.75% higher than that of YOLOX. Furthermore, we conduct experiments on the HRSC2016 dataset, which confirms the model's excellent robustness.

Keywords

Remote Sensing Image, Object Detection, Adaptive Spatial Fusion Structure, Sample Label Assignment Strategy

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

遥感技术是指从高空获取有关地球及其环境的信息,通常是通过飞行器或者卫星来收集[1]。近年来,随着高性能视觉传感器和航天航空技术的研究和发展,遥感技术也有了很大的进步,例如,遥感卫星的分辨率和数据获取速度得到了显著提高。由于遥感数据具有全局性、实时性和大规模等优势,因此在军事防御、灾害监测、城市规划和环境保护等各个领域都有应用。

遥感图像目标检测是遥感技术中的一个重要研究领域,主要通过自动化的方式,对遥感图像中的目标进行快速准确的识别和定位。传统的遥感图像目标检测方法由于需要人工进行大量的参数调整,难以进一步进行应用,因此基于深度学习的遥感图像目标检测方法逐渐成为主流,众多学者在这方面做出了杰出的研究成果。Zhu 等人通过设计一个新型预测头来处理目标的尺度差异,提高遥感图像中小目标的检测精度[2]。Fu 等人提出了一种特征融合架构,通过生成多尺度特征融合来处理多尺度物体,不同层次的特征组合有助于检测不同尺度的物体[3]。Hou 等人设计了一种用于任意方向物体检测的形状自适应检测器,根据不同物体形状对 *IoU* 阈值的要求不同,设计了一种针对遥感图像的样本匹配策略[4]。胡俊颐等人在 YOLOv4 的基础上,加入多模态的想法,构建了新型小目标检测模型,提高了遥感图像小目标的检测精度[5]。

综上所述,虽然目标检测技术发展迅速,将自然场景下的检测模型迁移应用于遥感图像检测也取得了一定成果。然而,由于遥感图像的特殊性仍然存在许多挑战,尤其是目标尺度变化剧烈、小目标检测等问题。因此,本文提出了一种基于自适应空间融合结构的网络模型,本文工作如下:

- 1) 针对遥感图像小目标排列密集、数量众多的难点,提出自适应空间融合结构,使用模型学习的权重来融合深层特征信息和浅层特征信息,提高模型对小目标的检测精度。
- 2) 针对单阶段检测器由于缺乏候选区域选择阶段而产生的正负样本不平衡问题,修改了置信度损失

函数。

3) 针对遥感图像中存在着大量高长宽比的目标物体, 优化了样本匹配策略。

2. 改进的 YOLOX 算法

YOLOX [6]是一种高效的检测算法, 它是 YOLO 系列中的最新成员, 它结合了目标检测近年来的一些重要研究热点, 比如无锚框检测器、先进的标签分配策略和端到端检测器等, 从而达到了较好的检测性能。本文在网络结构和样本匹配策略方面进行改进, 用于提高遥感图像目标检测性能。本文算法框架如图 1。

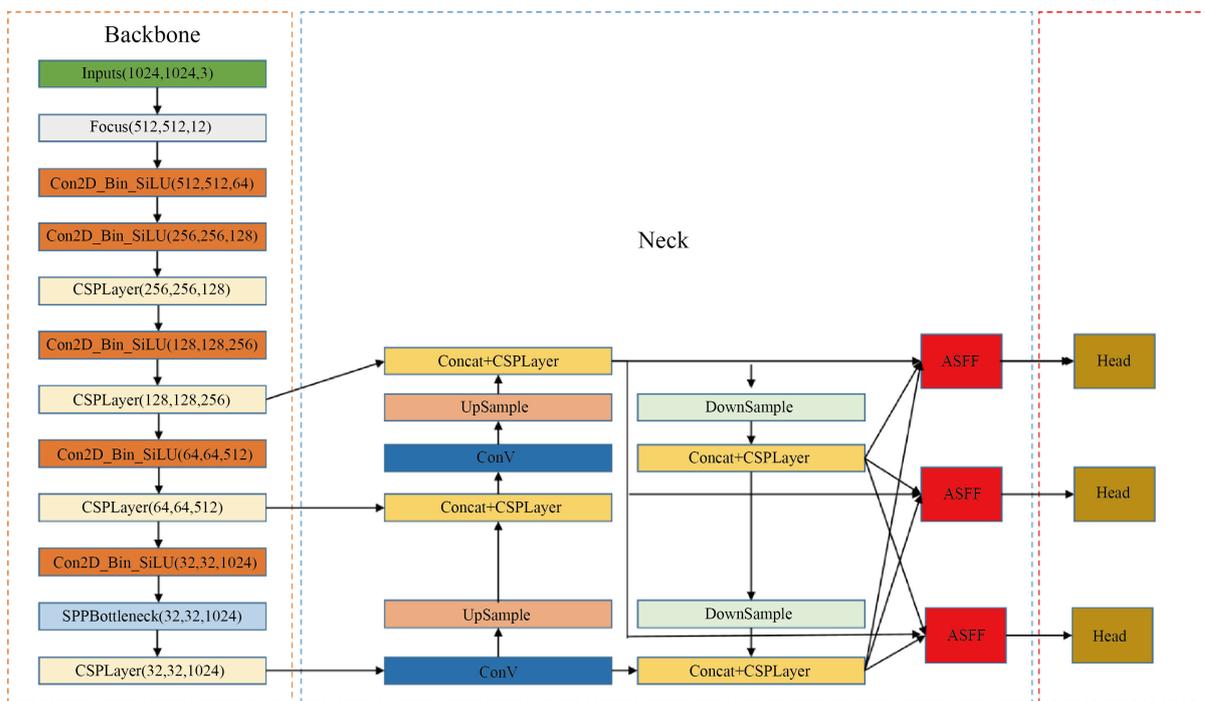


Figure 1. Improved network structure diagram of YOLOX algorithm

图 1. 改进后的 YOLOX 算法网络结构图

2.1. 自适应空间融合结构

金字塔特征表示法(FPN) [7]是目标检测模型在面对尺度变化较大时常用的解决方法。对于融入 FPN 的单阶段检测器来说, 不同特征尺度之间的不一致是其主要限制, 若是直接融合各层次特征就显得并不合理, 所以本文提出了自适应空间特征融合结构, 它通过在空间上过滤冲突信息来抑制梯度反转的时候不一致的情况, 改善了特征的比例不变性。模块结构如图 2 所示。

以 Fusion-3 为例, 将来自 level 1、level 2、level 3 的特征 χ^1 、 χ^2 、 χ^3 与来自不同层的权重参数 α_3 、 β_3 和 γ_3 相乘并最后相加, 就能得到自适应融合特征 Fusion-3。公式如下:

$$y_{ij}^l = \alpha_{ij}^l \cdot x_{ij}^{1 \rightarrow l} + \beta_{ij}^l \cdot x_{ij}^{2 \rightarrow l} + \gamma_{ij}^l \cdot x_{ij}^{3 \rightarrow l} \quad (1)$$

2.2. 平衡损失函数设计

在整个训练过程中, 损失函数式由 $loss_{cls}$ 、 $loss_{obj}$ 、 $loss_{reg}$ 三部分组成, $loss_{cls}$ 是指在取出特征点的类别预测结果后, 根据真实框的类别与特征点的类别预测结果做出的交叉熵损失, $loss_{obj}$ 是根据正负样本

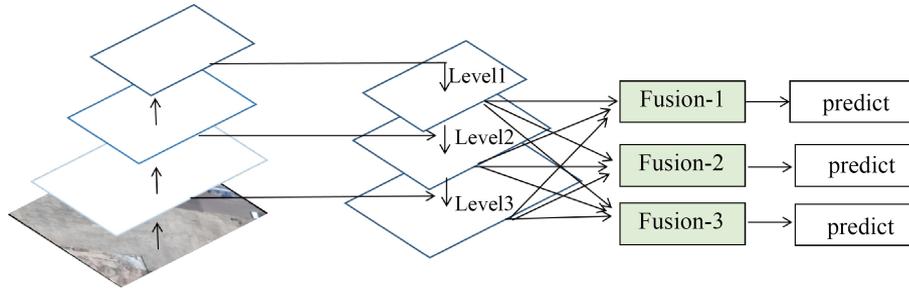


Figure 2. Adaptive spatial fusion structure diagram
图 2. 自适应空间融合结构图

和特征点是否包含目标物体的预测结果做出的交叉熵损失, $loss_{reg}$ 是预测框和真实框的 IoU 损失。具体如公式 2, 其中 λ 人为可以设置调整的权重参数。

$$loss = \lambda loss_{reg} + loss_{cls} + loss_{obj} \quad (2)$$

由于我们的模型是一个单阶段的检测器, 没有提前生成候选区域的过程, 它将对从网络中提取的每个特征进行预测, 并对每个位置的类别进行区分。这种过于密集的预测方法会产生过多负样本, 导致产生正负样本不平衡问题, 因此我们使用 Focalloss [8] 来代替原先的 BCE loss, 公式如下:

$$p_t = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{otherwise} \end{cases}, \quad FL(p_t) = -\alpha_t (1 - p_t)^\gamma \log(p_t) \quad (3)$$

上式中, y 的取值为 1 和 -1, 分别代表前景和背景, p 的取值范围为 0~1, 是模型预测属于前景的概率。 α_t 和 γ 都是权重参数, 通过 α_t 可以抑制正负样本的数量失衡, 通过 γ 可以控制难易区分样本数量失衡。

2.3. 样本标签匹配策略

目标检测算法往往都需要解决预测出来的框过多, 如何去除低质量框的问题, 也称为正负样本匹配策略问题。常见的策略大多是使用预测框和真实框直接算 IoU , 根据 IoU 阈值去划分正负样本, 大于 IoU 阈值的为正样本, 低于 IoU 阈值的则视为负样本或者忽略样本。YOLOX 的标签匹配策略为 SimOTA 技术, 它将匹配问题看成一个最优传输的问题, 即如何将预测框与真实框进行匹配使得代价最低。具体流程如图 3 所示。

在上述的过程, 预测框与真实框之间形成的 IoU 起着决定性的作用, 它决定了这个样本是否能够成为正样本。然而, 对于遥感图像这种目标排列密集, 物体尺度变化剧烈的对象来说, 直接使用 IoU 来作为评判标准并不严谨。在面对长宽比较大的狭长类物体如轮船、港口等时, 直接比较 IoU 来判断正负样本时, 就容易漏掉一些潜在高质量正样本。如图 4 右图相较于左图仅仅是旋转角度 θ 有一点变化, 整体的 IoU 就急剧下降, 但是从图 4 中可以看出, 右图其实是一个潜力框, 并不应该被直接否定。因此, 本文使用一种形状自适应阈值来代替传统的 IoU 阈值。

相比传统的静态 IoU 阈值需要人为进行不断调整从而达到最优效果, 我们的方法可以随着目标物体自适应调整, 无疑更具先进性。具体如公式 4 所示:

$$T_i^{IoU} = f(\gamma_i) * (\mu + \sigma) \quad (4)$$

其中, T_i^{IoU} 表示第 i 个真实标准框它所对应的阈值, $f(\gamma_i)$ 是计算目标物体长宽比的函数, μ 和 σ 则分别是平均值和标准差, 其公式如下。

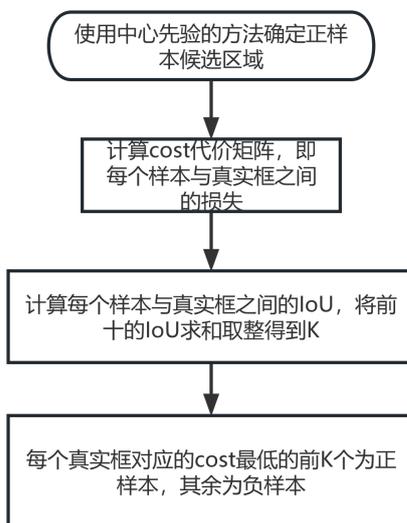


Figure 3. SimOTA sample assignment strategy
图 3. SimOTA 样本匹配策略



Figure 4. Effect of angular offset on objects with high aspect ratio
图 4. 角度偏移对高长宽比物体的影响

$$\mu = \frac{1}{j} \sum_{j=1}^j IoU_{ij} \quad (5)$$

$$\sigma = \sqrt{\frac{1}{j} \sum_{j=1}^j (IoU_{ij} - \mu)^2} \quad (6)$$

其中, IoU_{ij} 指第 j 个预测框与第 i 个真实框的 IoU , γ_i 的具体计算方式为长边与短边之比, 根据分析, 整体 IoU 阈值的大小应该随着长宽比的增大而减小, 因此 $f(\gamma_i)$ 被我们设计成一个递减的函数, 其公式如下:

$$f(\gamma_i) = e^{-\frac{\gamma_i}{\omega}} \quad (7)$$

其中, ω 是一个人为设置的权重参数, 根据实验可知, 当数据集中狭长类目标物体占比更大时, 更大的 ω 往往会取得更好的性能。

3. 实验结果与分析

3.1. 实验设置

实验采用深度学习框架 Pytorch 1.11.0, NVIDIA GeForce GTX 1080Ti 显卡, E5-2620 v4 CPU, 64 GB 内存, 64 位 Window 10 操作系统。实验输入图像大小设置为 640*640, Batch 大小设置为 4, epoch 设置为 100, 学习率设置为 0.0001, 使用余弦退火策略调整学习率。

3.2. 数据集

我们在武汉大学发布的 DOTA 数据集上进行了训练和预测, DOTA 数据集的内容丰富, 包含多种检测类别, 且各种目标之间的尺度存在差异、目标排列紧密且方向任意, 是一个具有挑战性的大规模遥感图像数据集。目标类别分别是: 飞机(PL)、棒球场(BD)、桥梁(BR)、田径场(GTF)、大型车辆(LV)、小型车辆(SV)、船舶(SH)、网球场(TC)、篮球场(BC)、储存罐(ST)、足球场(SBF)、环岛(RA)、港口(HA)、游泳池(SP)、直升机(HC)。

此外, 我们还在 HRSC2016 数据集上进行实验来验证本文方法的鲁棒性。

3.3. 评价指标

本实验采用 mAP 作为模型性能的评价指标, 它能够全面地评估目标检测模型的性能, 包括定位准确度和分类准确度, 公式如(8)~(11):

$$P = \frac{TP}{TP + FP} \quad (8)$$

$$R = \frac{TP}{TP + FN} \quad (9)$$

$$AP = \int_0^1 P(R) dR \quad (10)$$

$$mAP = \frac{\sum_{k=1}^n AP^k}{n} \quad (11)$$

其中, TP 为真正例即算法预测正确的正样本个数, FP 为算法预测错误但是判断为正样本的个数, FN 为算法预测错误且被认为是负样本的个数, P 为准确率, R 为召回率, AP 是平均准确度。 mAP 是所有类别 AP 的平均值。

3.4. 消融实验

为了验证本文改进网络的有效性, 在相同实验条件下, 对各个方法模块进行实验, 实验结果如表 1, 可以看出本文提出的各个改进模块都对整体模型的性能提升起到了不错的效果。

Table 1. Ablation experiments on the DOTA dataset

表 1. 在 DOTA 数据集上的消融实验

Fusion	Improved SimOTA	Focal loss	mAP (%)
×	×	×	76.32
×	×	√	77.38
×	√	×	77.89
×	√	√	78.46
√	√	√	79.07

3.5. 对比实验

上述实验验证了本文方法各个模块的有效性, 接下来在 DOTA 数据集和 HRSC2016 数据集上分别进行实验, 并与主流方法 RetinaNet、SCR Det [9]、F3Net [10]和原始 YOLOX 算法进行对比。实验结果如表 2 所示, 我们的算法相较于原始 YOLOX 不仅小目标的 AP 明显提升, 而且对于高长宽比的目标物体如船舶等的识别效果更好, 算法的 mAP 相较于原始算法提升了 2.75%, 达到了 79.07% 的优秀水平。在 DOTA 数据集上的可视化结果如图 5 所示, 可以看出相较于原始算法, 本文算法的漏检率更低, 识别更精确。

Table 2. Comparison with dominant methods on the DOTA dataset

表 2. 在 DOTA 数据集上与主流方法的对比

Category	RetinaNet	SCR Det	F3Net	原 YOLOX	本文修改后的网络
			AP (%)		
PL	88.67	89.98	88.89	90.30	90.44
BD	77.62	80.65	78.48	74.69	71.23
BR	41.81	52.09	54.62	58.10	66.91
GTF	58.17	68.36	74.43	67.06	69.04
SV	74.58	68.36	72.80	82.54	83.31
LV	71.64	60.32	77.52	87.30	87.28
SH	79.11	72.41	87.54	88.28	88.71
TC	90.29	90.85	90.87	90.29	90.59
BC	82.18	87.94	87.64	67.73	70.65
ST	74.32	86.86	85.63	86.59	88.27
SBF	54.75	65.02	63.80	60.12	69.54
RA	60.60	66.68	64.53	67.38	74.35
HA	62.57	66.25	78.06	58.50	67.41
SP	69.67	66.25	72.36	85.22	86.94
HC	60.64	68.24	63.19	80.74	81.40
mAP	69.77	72.68	76.02	76.32	79.07

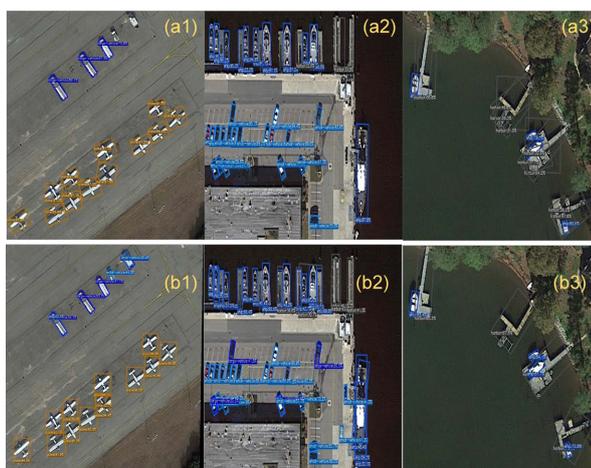


Figure 5. Visualization results on the DOTA dataset (a is the original YOLOX visualization, b is the visualization results of the algorithm in this paper)

图 5. 在 DOTA 数据集上的可视化结果(a 为原始 YOLOX 可视化, b 为本文算法可视化结果)

为了证明本文方法的鲁棒性, 在 HRSC2016 数据集也进行了实验, 并分别与 RoI Transformer [11]、Gliding Vertex [12]、R3Det [13]、SLA [14]、DAL [15]等优秀方法做了对比, 如图 6 所示, 证明本文算法的性能最优, 在该数据集上 mAP 达到了 90.14%。

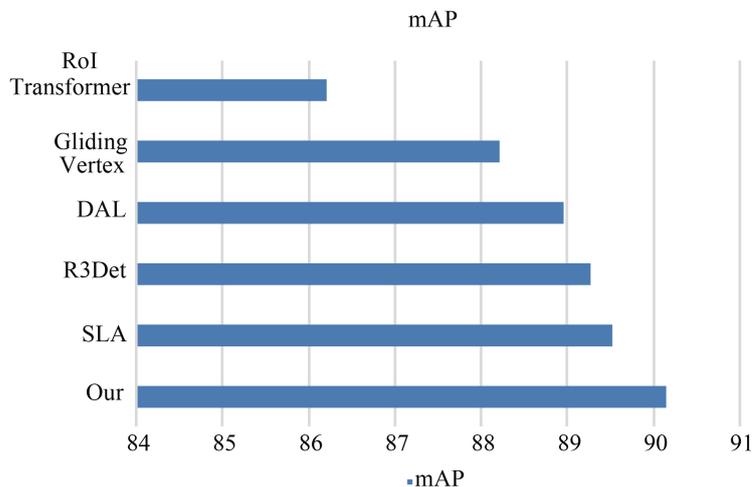


Figure 6. Comparative experiments on the HRSC2016 dataset
图 6. 在 HRSC2016 数据集上的对比实验

4. 结论

针对遥感图像中小目标排列密集、高长宽比目标物体普遍存在、正负样本失衡的问题, 本文在 YOLOX 算法的基础上, 引入了自适应特征融合结构, 该结构在特征金字塔的基础上, 通过学习不同特征图之间的潜在关系, 将不同级别的特征自适应地融合在一起, 有效提高了小目标的检测精度; 使用 Focal loss 替换掉原先的 BCE loss, 通过减小易分类样本的权重来缓解正负样本不平衡问题; 并提出形状自适应 IoU 阈值来优化样本匹配策略, 该阈值通过对目标的形状进行分析, 结合目标的长宽比和尺度信息, 自适应调整 IoU 阈值解决高长宽比目标对于角度偏移敏感的问题。实验结果表明, 改进后的 YOLOX 算法在遥感图像目标检测方面拥有不错的性能, 同时具有一定的鲁棒性。但是, 本文的改进还存在一些不足之处, 在数据量较小的情况下, 模型的泛化能力有待进一步提升, 未来可以考虑引入更加先进的网络模型来提升遥感图像目标检测器的性能。

参考文献

- [1] Luo, S., Yu, J., Xi, Y., et al. (2022) Aircraft Target Detection in Remote Sensing Images Based on Improved YOLOv5. *IEEE Access*, **10**, 5184-5192. <https://doi.org/10.1109/ACCESS.2022.3140876>
- [2] Zhu, X., Lyu, S., Wang, X., et al. (2021) TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-Captured Scenarios. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Montreal, 11-17 October 2021, 2778-2788. <https://doi.org/10.1109/ICCVW54120.2021.00312>
- [3] Fu, K., Chang, Z., Zhang, Y., et al. (2020) Rotation Aware and Multi-Scale Convolutional Neural Network for Object Detection in Remote Sensing Images. *ISPRS Journal of Photogrammetry and Remote Sensing*, **161**, 294-308. <https://doi.org/10.1016/j.isprsjprs.2020.01.025>
- [4] Hou, L., Lu, K., Xue, J., et al. (2022) Shape-Adaptive Selection and Measurement for Oriented Object Detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, **36**, 923-932. <https://doi.org/10.1609/aaai.v36i1.19975>
- [5] 胡俊, 王秋红. 基于遥感图像的多模态小目标检测[J]. 图学学报, 2022, 43(2): 197-204.
- [6] Ge, Z., Liu, S., Wang, F., et al. (2021) Yolox: Exceeding Yolo Series in 2021. ArXiv: 2107.08430.

-
- [7] Lin, T.Y., Dollár, P., Girshick, R., *et al.* (2017) Feature Pyramid Networks for Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 936-944. <https://doi.org/10.1109/CVPR.2017.106>
- [8] Lin, T.Y., Goyal, P., Girshick, R., *et al.* (2017) Focal Loss for Dense Object Detection. *Proceedings of the IEEE International Conference on Computer Vision*, Venice, 22-29 October 2017, 2980-2988. <https://doi.org/10.1109/ICCV.2017.324>
- [9] Yang, X., Yang, J., Yan, J., *et al.* (2019) SCRDet: Towards More Robust Detection for Small, Cluttered and Rotated Objects. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Seoul, 27 October-2 November 2019, 8231-8240. <https://doi.org/10.1109/ICCV.2019.00832>
- [10] Wei, J., Wang, S. and Huang, Q. (2020) F³Net: Fusion, Feedback and Focus for Salient Object Detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, **34**, 12321-12328. <https://doi.org/10.1609/aaai.v34i07.6916>
- [11] Ding, J., Xue, N., Long, Y., *et al.* (2019) Learning Roi Transformer for Oriented Object Detection in Aerial Images. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 2844-2853. <https://doi.org/10.1109/CVPR.2019.00296>
- [12] Xu, Y., Fu, M., Wang, Q., *et al.* (2020) Gliding Vertex on the Horizontal Bounding Box for Multi-Oriented Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **43**, 1452-1459. <https://doi.org/10.1109/TPAMI.2020.2974745>
- [13] Yang, X., Yan, J., Feng, Z., *et al.* (2021) R³Det: Refined Single-Stage Detector with Feature Refinement for Rotating Object. *Proceedings of the AAAI Conference on Artificial Intelligence*, **35**, 3163-3171. <https://doi.org/10.1609/aaai.v35i4.16426>
- [14] Ming, Q., Miao, L., Zhou, Z., *et al.* (2021) Sparse Label Assignment for Oriented Object Detection in Aerial Images. *Remote Sensing*, **13**, Article 2664. <https://doi.org/10.3390/rs13142664>
- [15] Ming, Q., Zhou, Z., Miao, L., *et al.* (2021) Dynamic Anchor Learning for Arbitrary Oriented Object Detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, **35**, 2355-2363. <https://doi.org/10.1609/aaai.v35i3.16336>