

DOI:10.16136/j.joel.2022.06.0724

基于改进 YOLOv4 的遥感图像目标检测

叶玉伟, 任彦*, 高晓文, 王佳鑫

(内蒙古科技大学 信息工程学院, 内蒙古 包头 014010)

摘要:针对遥感图像在复杂背景下因特征提取和表达能力不足而存在漏检和检测效果不佳的问题,提出一种优化特征提取网络的 YOLOv4 算法模型。该改进模型引入了一种新的 Dense-PANet 结构以获取更高的分辨率特征,并通过在特征提取网络中嵌入注意力机制以适应遥感图像因视野范围大而导致复杂背景下小目标漏检和检测效果不佳的问题。为了证明本文所提方法的有效性,针对 DIOR 遥感数据源进行了对比实验,结果表明,本文算法平均准确率(mean average precision, mAP)为 86.55%,相比原算法提高了 2.52%,较 YOLOv3、RetinaNet 提高了 6.58%、14.09%,验证了所改进算法的有效性。

关键词:目标检测; 遥感图像; YOLOv4; 注意力机制

中图分类号:TP751.1 **文献标识码:**A **文章编号:**1005-0086(2022)06-0607-07

Remote sensing image target detection based on improved YOLOv4

YE Yuwei, REN Yan*, GAO Xiaowen, WANG Jiaxin

(School of Information Engineering, Inner Mongolia University of Science and Technology, Baotou, Inner 014010, China)

Abstract: Aiming at the problem of missed detection and poor detection effect of remote sensing images due to insufficient feature extraction and expression capabilities in complex backgrounds, a YOLOv4 algorithm model that optimizes feature extraction network is proposed. The improved model introduces a new Dense-PANet structure to obtain higher resolution features, and embeds the attention mechanism in the feature extraction network to adapt to remote sensing images due to the large field of view, which leads to the missed detection of small targets in complex backgrounds and the problem of poor detection results. In order to prove the effectiveness of the method proposed in this paper, a comparative experiment was conducted on DIOR remote sensing data sources. The results show that the average accuracy (mean average precision, mAP) of the algorithm in this paper is 86.55%, which is an increase of 2.52% compared to the original algorithm. YOLOv3 and RetinaNet increased by 6.58% and 14.09%, which verifying the effectiveness of the improved algorithm.

Key words: target detection; remote sensing image; YOLOv4; attention mechanism

1 引言

目标检测是计算机视觉和图像处理领域研究的热点问题,已经成为社会安全、农业农村和军事国防等各个领域的基础性课题之一^[1]。在远距离拍摄的航空航天遥感影像或视频监控中,与自然场景下的图像相比,遥感图像的背景复杂度更高、

目标尺寸更小且外观较为模糊。在复杂背景中检测小目标是目标检测领域的一个重要研究方向,小目标在原图中所占像素数较少,携带的信息有限,缺乏纹理、形状、颜色等外观信息。基于深度学习的目标检测算法对小目标检测精度较低且对复杂背景下目标适应能力弱,容易发生漏检和误检现象,严重影响检测结果,因此小目标检测在计

* E-mail: ren0831@imust.edu.cn

收稿日期: 2021-10-22 修订日期: 2021-11-19

基金项目: 国家自然科学基金(62063027)、内蒙古自治区科技计划项目(2020GG0048)、内蒙古自然科学基金(2019MS06002)和内蒙古自治区高等学校青年科技人才发展计划项目(NJYT22057)资助项目

计算机视觉任务中一直是一个亟待解决的难点问题^[2,3]。AUDEBERT等^[4]在遥感图像中将语义分割和目标检测进行结合,利用深度全卷积网络对遥感图像中的车辆进行了精确分割,将背景和待检测目标区别开,可以有效地提高检测性能^[5]。文献[6]在YOLOv3网络的深层特征中增加检测尺度,以获得较小的接受场,提高网络对小目标的灵敏度;CUI等^[7]提出了一种上下文感知块网络(context-aware block network, CAB Net),CAB利用金字塔扩张卷积来合并多级上下文信息,通过构建高分辨率和强语义特征图来改善小物体检测。在文献[8]中,提出了一种内外网检测器,其中设计了双向特征融合模块(bidirectional feature fusion module, Bi-FFM)、上下文推理模块(contextual reasoning module, CRM)和上下文特征增强模块(context feature enhancement module, CFEM)3个模块,从特征提取、建议定位、分类等方面对小目标检测进行了改进。

使用YOLOv4目标检测算法对遥感数据集进行检测,对于复杂背景情况下小目标的检测效果不尽理想^[9,10],因小目标具有分辨率低、特征提取困难,语义信息不足等特点,导致最终的漏检率很高。为了降低复杂背景下小目标的漏检情况,本文在YOLOv4原有框架基础上,引入了密集连接来获取更多的特征;为提高复杂背景下的检测精度,在特征提取网络中嵌入注意力机制模块,以实现更好的区别目标和背景;为了进一步提高网络的检测性能,优化了颈部特征提取网络中的激活函数。

2 YOLOv4 目标检测算法

YOLOv4算法在之前算法的基础上增加了许多优化方法,包括骨干网络、激活函数、损失函数、网络训练等方面的优化。YOLOv4的网络结构主要有3个部分:骨干网络(CSPDarkNet53)、颈部(Neck)、头部(Head)^[11]。

YOLOv4的骨干网络是在Darknet53的基础上结合CSPnet^[12]形成的一种结构。CSP模块将基础层的特征映射划分为两部分,然后通过跨阶段层次结构将它们合并起来,如图1所示,这样不仅可以保

证模型的准确率,而且可以降低计算量。并在批量归一化(batch normalization, BN)层后使用Mish^[13]激活函数,增强深层信息的传播。算法的颈部是在YOLOv3中特征金字塔网络FPN^[14]的基础上增加了一个自底向上的路径聚合网络(path aggregation network, PAN^[15]),从主干网络的不同层聚合不同的特征,增强网络的特征提取能力。同时增加空间金字塔池化(spatial pyramid pooling, SPP^[16])模块增大网络的感受野,同时减少过拟合现象。

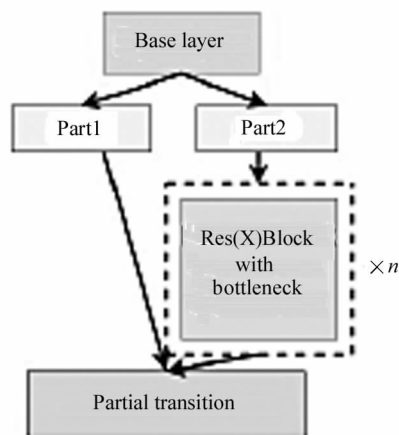


图1 CSPResNet(X)t

Fig. 1 CSPResNet(X)t

在头部结构中,YOLOv4继承了YOLOv3的多尺度预测,提高了对不同尺度目标的检测性能。使用CIOU(complete intersection over union)误差替换均方误差作为损失函数:

$$L_{CIOU} = 1 - IOU_{(a,b)} + \frac{\rho^2(a_{dr}, b_{dr})}{d^2} + \alpha v, \quad (1)$$

$$\alpha = \frac{v}{(1 - IOU_{(a,b)}) + v}, \quad (2)$$

$$v = \frac{4}{\pi^2} (\arctan \frac{w_{gt}}{h_{gt}} - \arctan \frac{w}{h})^2, \quad (3)$$

式中, $IOU_{(a,b)}$ 是真实框预测框的交并比, $\rho^2(a_{dr}, b_{dr})$ 是真实框预测框中心点的欧氏距离, d 是包含真实框预测框的最小框的对角线距离, w_{gt} 、 w 、 h_{gt} 、 h 分别是真实框和预测框的宽、高。

3 对YOLOv4算法的改进

本文主要在以下3方面对现有YOLOv4算法进行改进:1)引入密集连接实现特征重用,增强小目标的定位能力;2)嵌入注意力机制,解决复杂背景下目标检测的问题;3)改进激活函数,进行微调。改进后的网络结构如图2所示。

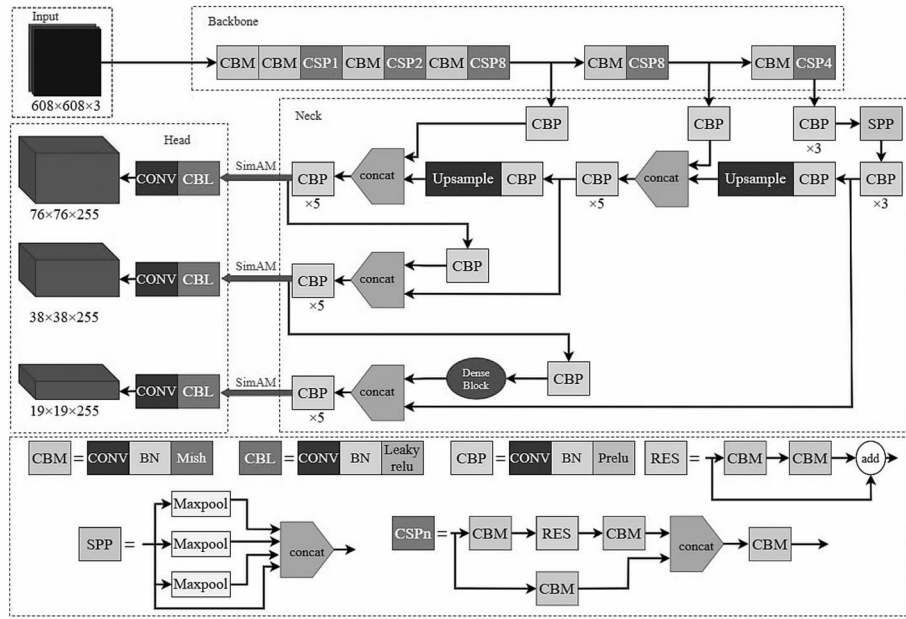


图 2 改进后的 YOLOv4 算法
Fig. 2 Improved YOLOv4 algorithm

3.1 引入密集连接

在特征提取网络中,低层特征语义信息相对稀缺,但对象定位准确。相反,高层特征语义信息丰富,但对象定位比较粗糙。受到 DenseNet^[17] 网络的启发,将密集连接结构引入 PANet 中,这种密集连接

的模式使网络的当前层接收其之前的所有层特征作为输入,因此,后面的特征层会获取前面所有层的特征图,实现特征重用,使得高层特征图在保证丰富的语义信息的同时增强对小目标的定位能力。密集连接结构展开图如图 3 所示。

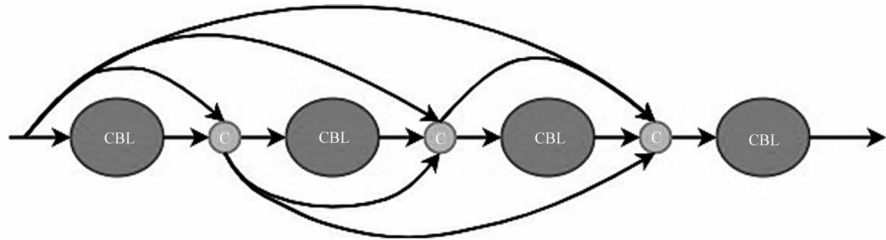


图 3 密集连接结构
Fig. 3 Densely connected structure

3.2 嵌入 SimAM 注意力机制模块

SimAM^[18] 注意力机制(3D)通过关注通道维度和空间维度信息,同时对权值进行统一处理,使模型可以自动学习到不同通道和空间特征的重要程度,从而进一步增强神经网络的特征提取能力。相对于常用的通道注意力^[19](1D)、空间注意力^[20](2D)来说更具有灵活性。如图 4 所示。为了更好地实现注意力,需要对每个神经元的重要性进行评估,定义了如式(4)所示的能量函数:

$$e_t(w_t, b_t, y, x_i) = (y_t - \hat{t})^2 + \frac{1}{M-1} \sum_{i=1}^{M-1} (y_0 - \hat{x}_i)^2, \quad (4)$$

式中, $\hat{t} = w_t t + b_t$; $\hat{x}_i = w_t x_i + b_t$, t 和 x_i 分别指输入特

征 X 的目标神经元和其他神经元, i 指空间维度上的索引, w_t 和 b_t 分别指某个神经元变换时的“权重”和“偏差”,采用二值标签,并添加正则项对式(4)进行最小化处理,最终的能量函数定义如式(5)所示:

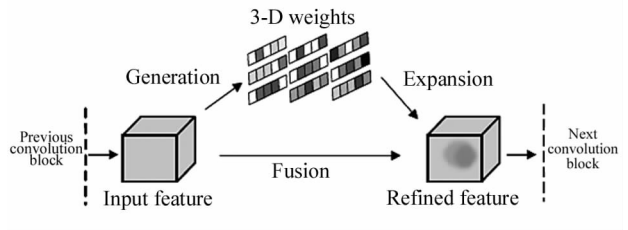


图 4 3D 注意力机制
Fig. 4 3D attention mechanism

$$e_t(w_t, b_t, y, x_i) = \frac{1}{M-1} \sum_{i=1}^{M-1} (-1 - (w_t x_i + b_t))^2 + (1 - (w_t t + b_t))^2 + \lambda w_t^2, \quad (5)$$

上式具有如下解析解:

$$w_t = -\frac{2(t - u_t)}{(t - u_t)^2 + 2\sigma_t^2 + 2\lambda}, \quad (6)$$

$$b_t = -\frac{1}{2}(t + u_t)w_t, \quad (7)$$

式中, $u_t = \frac{1}{M-1} \sum_{i=1}^{M-1} x_i$, $\sigma_t^2 = \frac{1}{M-1} \sum_{i=1}^{M-1} (x_i - u_t)^2$, 因此, 最小能量如式(8):

$$e_t^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{u})^2 + 2\hat{\sigma}^2 + 2\lambda}, \quad (8)$$

由式(8)可知: 能量越低, 神经元 t 与周围神经元的区别越大, 重要性越高。因此, 神经元的重要性可以通过 $1/e_t^*$ 得到。式(4)一式(8)推导出了能量函数并挖掘了神经元的重要性, 按照注意力机制的定义, 需要对重要的特征进行增强处理:

$$\tilde{X} = \text{sigmoid}(\frac{1}{E}) \odot X. \quad (9)$$

由于 SimAM 是一个不会引入额外参数的模块, 并且大部分操作均基于所定义的能量函数选择, 避

免了过多的结构调整, 通过实验证明该模块的加入可以更好地区分目标和背景, 提高了网络的检测性能。

3.3 改进激活函数

将颈部结构中使用的激活函数更换为 PRelu, LeakyRelu 和 Prelu 都是在 Relu 的基础上改进而来。Relu 是将所有的负值都设为零, 相反, LeakyRelu 是给所有负值赋予一个固定的非零斜率, 而在 Prelu 中, 负值部分的斜率是根据数据来定的, 而非预先定义的, 这样就会根据数据在动态过程中学习到一个最优参数, 对网络的最终结果进行优化。

4 实验与分析

4.1 实验数据

为评估本文所改进 YOLOv4 算法的检测性能, 实验采用 DIOR 遥感图像数据集, 涵盖 20 个对象类, 幅面大小约为 $800 \text{ pixel} \times 800 \text{ pixel}$ 的光学遥感影像共计 23 463 张, 共含有 19 247 2 个实例, 训练集、验证集、测试集按 7 : 1 : 2 进行划分, 数据集部分样本示例如图 5 所示。

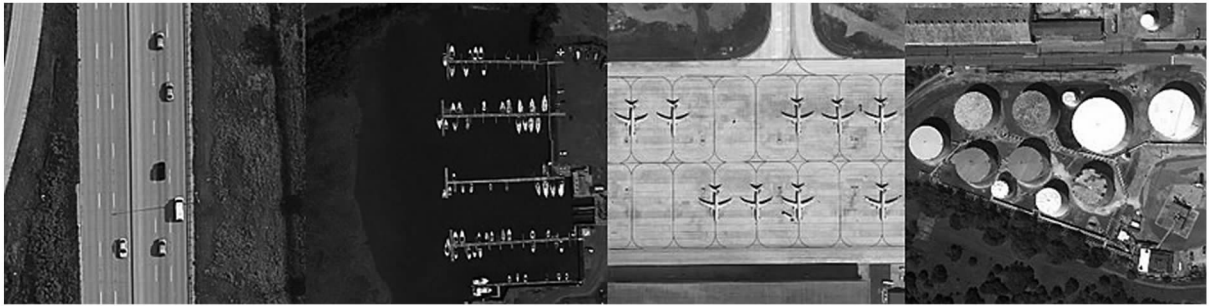


图 5 数据集部分样本示例

Fig. 5 Examples of partial samples of the data set

4.2 评估标准及训练方法

本文采用已被广泛应用于目标检测评估的平均精度均值(mean average precision, mAP)作为衡量模型的标准, mAP 是指多个类别精度均值(average precision, AP)的平均值, 每个类别都可以根据准确率(precision)和召回率(recall)绘制一条 PR 曲线, 其在 0—1 区间范围内绘制的曲线与坐标轴所围成的面积即为 AP , 具体可表示为 $AP = \int_0^1 p(r) dr$, 准确率和召回率可表示为:

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (10)$$

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (11)$$

式中参数意义如表 1 所示。

表 1 样本分类的混淆矩阵

Tab. 1 Confusion matrix for sample classification

TrueSituation	Forecast	
	Positive example	Negative example
Positive example	True positive(TP)	False negative(FP)
Negative example	False positive(FP)	True negative(FN)

本文所改进 YOLOv4 算法在 TUF RTX 3 090 显卡上采用 Pytorch 开源深度学习框架进行训练和测试, 利用在 ImageNet 数据集上预训练的模型进行

迁移学习,初始学习率设置为 0.001,采用余弦退火衰减法对学习率进行优化,经过 120 次训练最终损失图线如图 6 所示,并且采用冻结训练对主干部分的训练权重进行冻结,使得时间和资源利用都能得到很大改善。

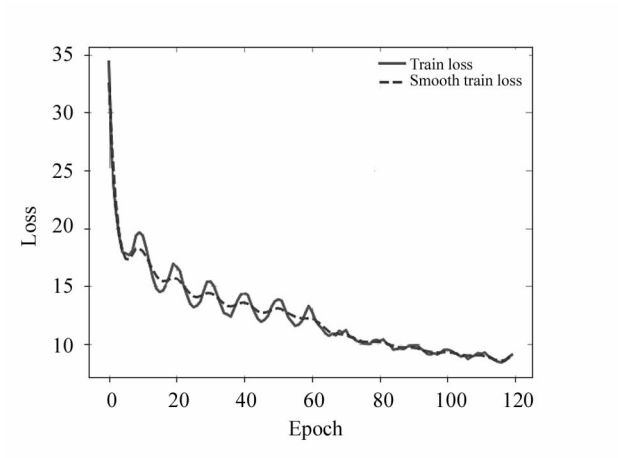


图 6 训练损失图

Fig. 6 Training loss graph

4.3 实验结果与分析

表 2 是基于 DIOR 数据集的不同目标检测算法

表 2 各算法在遥感图像物体的检测精度

Tab. 2 The detection accuracy of various algorithms in remote sensing image objects

Method	Backbone	<i>mAP</i>
RetinaNet	ResNet50	72.46%
YOLOv3	DarkNet53	79.97%
YOLOv4	CSPDarkNet53	84.02%
Ours	CSPDarkNet53	86.55%

实验结果。可以看出,在大小为 608×608 分辨率的输入图像下,本文目标检测模型的 *mAP* 值达到了 86.55%,在一阶段目标检测算法中,该模型的 *mAP* 值比 YOLOv3^[21] 高出 6.58%;比 YOLOv4 原算法高出 2.53%,比二阶段检测算法 RetinaNet^[22] 提高了 14.09%。可以看出本文所改进方法在检测精度上有一定的提升。

为了更好地理解本文改进算法中各个模块对检测效果的影响,表 3 是基于 DIOR 数据集以 YOLOv4 为基准模型逐步增加注意力机制、密集连接结构和激活函数进行的消融实验。由表 3 可以看到,结合密集连接结构可以使网络对目标的定位和分类更加准确,改进后模型的 *mAP* 值比 YOLOv4 提升了 1.59%,结合注意力机制后,可以更好地区分目标和背景,使 *mAP* 值比 YOLOv4 提升了 2.52%,改变激活函数对结果进行微调,使得最终准确率达到 86.55%。

表 3 模块消融实验

Tab. 3 Module ablation experiment

YOLOv4	Simam	Dense	Prelu	<i>mAP</i>
✓	—	—	—	84.02%
✓	✓	—	—	85.66%
✓	—	✓	—	85.61%
✓	✓	✓	—	86.50%
✓	✓	✓	✓	86.55%

图 7 是几种常用检测算法与本文改进算法在复杂场景下小目标的检测效果对比图,由于廊桥的形状与飞机的形状相似,停机坪的颜色与飞机的颜色

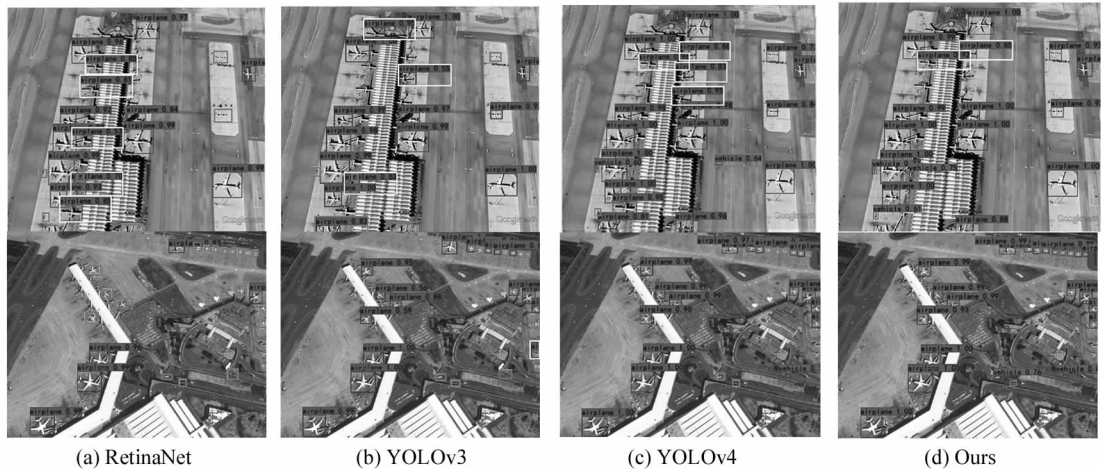


图 7 4 种算法在复杂背景下检测小目标物体的差异

Fig. 7 The difference between the four algorithms to detect small target objects in a complex background

相近,所以很容易造成误检(白色手动标注框)或漏检(绿色手动标注框)情况的发生。图8是针对4种算法漏检率的对比,可以更加直观地看出改进算法

的效果。实验结果表明,本文改进的YOLOv4算法能够有效地提高复杂背景下遥感图像的目标检测精度,尤其是对小目标物体的检测更具优势。

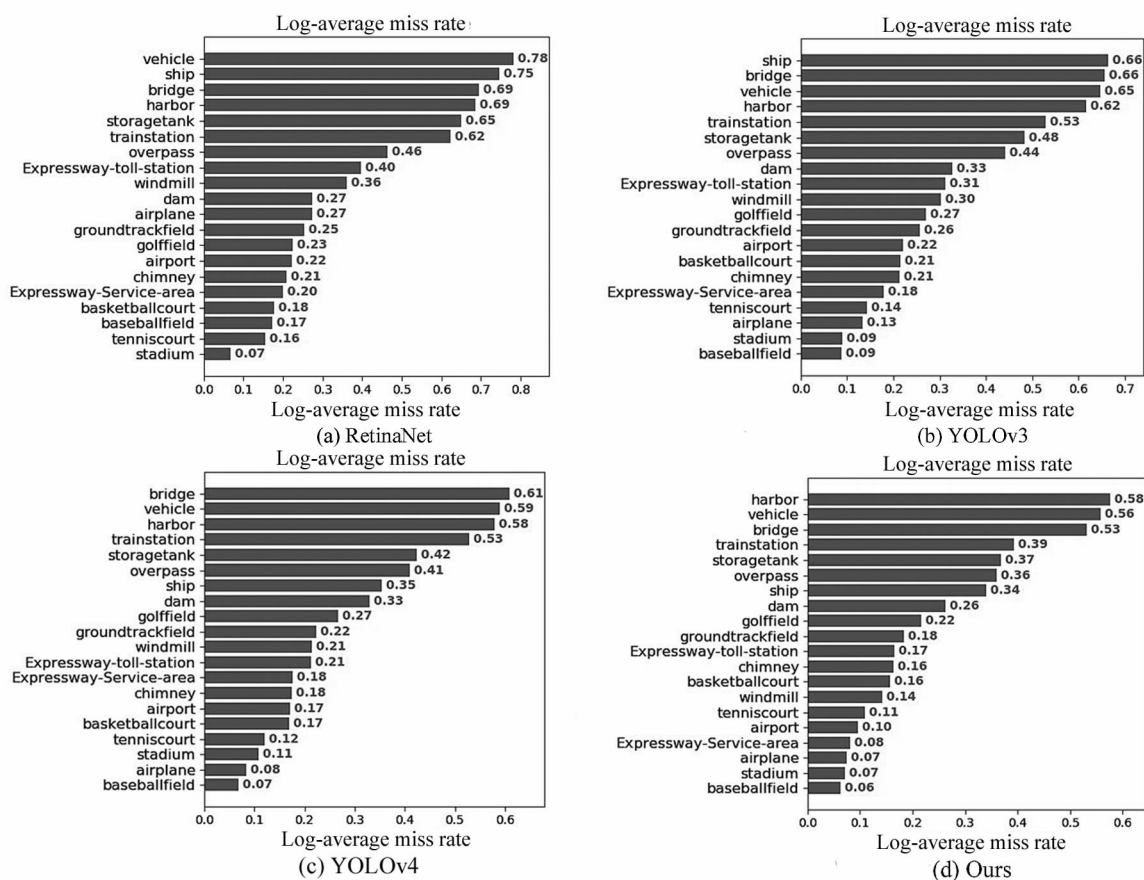


图8 4种算法误检率的对比

Fig. 8 Comparison of missed detection rates of four algorithms

5 结论

本文针对常用目标检测算法在复杂背景情况下检测精度差,尤其是对于小目标漏检严重的问题提出了改进的YOLOv4目标检测算法。该算法在颈部特征提取网络部分引入了密集连接,将低层丰富的语义信息直接传递到高层特征图,以增强对小目标的定位能力;嵌入Simam注意力机制,极大程度地区分目标和背景;同时将网络中的激活函数替换为Prelu函数,可以根据数据集对参数进行优化。在训练时使用了余弦退火学习率、标签平滑以及Drop-Block正则化方法,有效加强了对小目标的检测。在DIOR数据集上,该算法在满足实时检测的前提下,平均精度指标 mAP 比YOLOv4提升了2.52%,但是该算法仍然存在一些漏检情况,这依然是需要继续深化研究的问题。

参考文献:

- [1] FAN L L, ZHAO H W, ZHAO H Y, et al. Survey of target detection based on deep convolution neural networks[J]. Optical and Precision Engineering, 2020, 28(5): 1152-1164.
范丽丽, 赵宏伟, 赵浩宇, 等. 基于深度卷积神经网络的目标检测研究综述[J]. 光学精密工程, 2020, 28(5): 1152-1164.
- [2] LI H G, YU R N, DING W R. Research development of small object tracking based on deep learning[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(7): 107-125.
李红光, 于若男, 丁文锐. 基于深度学习的小目标检测研究进展[J]. 航空学报, 2021, 42(7): 107-125.
- [3] LIU Y, LIU H Y, FAN J L, et al. A survey of research and application of small object detection based on deep

- learning[J]. Acta Electronica Sinica, 2020, 48(3): 590-601.
- 刘颖,刘红燕,范九伦,等. 基于深度学习的小目标检测研究与应用综述[J]. 电子学报, 2020, 48(3): 590-601.
- [4] AUDEBERT N, LE SAUX B, LEFÈVRE S. Segment-before-detect: vehicle detection and classification through semantic segmentation of aerial images[J]. Remote Sensing, 2017, 9(4): 368.
- [5] JAING B, QU R K, Li Y D, et al. Survey of object detection in unmanned aerial vehicle imagery based on deep learning[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(4): 137-151.
- 江波,屈若崑,李彦冬,等. 基于深度学习的无人机航拍目标检测研究综述[J]. 航空学报, 2021, 42(4): 137-151.
- [6] CAO C, WU J, ZENG X, et al. Research on airplane and ship detection of aerial remote sensing images based on convolutional neural network[J]. Sensors, 2020, 20(17): 4696.
- [7] CUI L S, LV P, JIANG X H, et al. Context-aware block net for small object detection[J]. IEEE Transactions on Cybernetics, 2020, 52(4): 2300-2312.
- [8] LENG J X, REN Y H, JIANG W, et al. Realize your surroundings: Exploiting context information for small object detection[J]. Neurocomputing, 2021, 433: 287-299.
- [9] LIU J H. Multi-target detection method based on YOLOv4 convolutional neural network[J]. Journal of Physics: Conference Series, 2021, 1883(1): 012075.
- [10] SHI R P, JIANG D N, FANG Q. Aircraft target detection in remote sensing image based on YOLOv4[J]. Bulletin of Surveying and Mapping, 2021(S1): 134-138.
- 史瑞鹏,蒋丹妮,方青. 基于 YOLOv4 的遥感影像飞机目标检测[J]. 测绘通报, 2021(S1): 134-138.
- [11] BOCHKOVSKIY A, WANG C Y, LIAO H. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23) [2021-10-22]. <http://arxiv.org/abs/2004.10934v1>.
- [12] WANG C Y, LIAO H, WU Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 13-19, 2020, Seattle, WA, USA. New York: IEEE, 2020: 390-391.
- [13] MISRA D. Mish: a self regularized non-monotonic neural activation function [EB/OL]. (2019-08-23) [2021-10-22]. <http://arxiv.org/abs/1908.08681>.
- [14] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 2117-2125.
- [15] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 8759-8768.
- [16] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37(9): 1904-1916.
- [17] HUANG G, LIU Z, VAN DER MAATEN L, et al. DenseNet: connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 2261-2269.
- [18] YANG L, ZHANG R, et al. SimAM: a simple, parameter-free attention module for convolutional neural networks [C]//Proceedings of the 38th International Conference on Machine Learning, July 18-24, 2021, On Line. Virtual Event: IMLS, 2021, 139: 11863-11874.
- [19] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2017: 7132-7141.
- [20] ZHU X, CHENG D, ZHANG Z, et al. An empirical study of spatial attention mechanisms in deep networks [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea. New York: IEEE, 2019: 6687-6696.
- [21] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08) [2021-10-22]. <https://arxiv.org/abs/1804.02767>.
- [22] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2018, 42(2): 318-327.

作者简介:

任彦 (1977—), 女, 教授, 硕士生导师, 主要研究方向为遥感图像处理。