

基于改进 YOLOv3 的自然场景下鸟类检测算法

宋子盈^{1,2*}, 杨奎河², 张宇²

¹北京交通大学计算机与信息技术学院, 北京 100044;

²河北科技大学信息科学与工程学院, 河北 石家庄 050018

摘要 针对自然场景下鸟类检测任务中存在模型参数量大、计算量高和正负样本严重不平衡的问题, 提出了 YOLOBIRDS 算法。一方面, 对特征提取网络模型进行修改, 将标准卷积神经网络结构修改为深度可分离残差模型。另一方面, 对损失函数进行修改, 将目标框大小与位置损失函数由均方误差修改为广义交并比(CIoU), 在置信度损失函数中增加正负样本控制参数。实验结果表明, 在衡水湖鸟类数据集中, YOLOBIRDS 算法的平均精度均值(mAP)达 87.12%, 比原算法高 2.71 个百分点; 参数个数达 12425917, 比原算法低 79.88%; 速度达 32.67 frame/s, 比原算法高 19.98%。采用该算法训练得到的新模型对鸟类检测的精度更高, 检测速度更快, 对鸟类检测的整体识别率大幅度提高, 平衡了正负样本的损失权重。

关键词 目标检测; 深度可分离残差模型; 广义交并比损失函数; YOLOv3 算法

中图分类号 TP391

文献标志码 A

DOI: 10.3788/LOP202259.1815013

Bird Detection Algorithm in Natural Scenes Based on Improved YOLOv3

Song Ziyi^{1,2*}, Yang Kuihe², Zhang Yu²

¹School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China;

²School of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang 050018, Hebei, China

Abstract This study proposes the YOLOBIRDS algorithm to solve the challenges of several model parameters, high amount of calculation, and a considerable imbalance of positive and negative samples in bird detection tasks in natural scenes. The feature extraction network model was modified, and the standard convolution neural network structure was modified to the depthwise separable residual model. Additionally, the loss function was modified, and the object box size and position loss function were modified from mean square error to generalized intersection over union (CIoU). The confidence loss function includes the positive and negative sample control parameters. The experimental results show that in the Hengshui Lake bird dataset, the mean average precision (mAP) of the YOLOBIRDS algorithm reaches 87.12%, which is 2.71 percentage points higher than that of the original algorithm. Moreover, number of parameters reaches 12425917, which is 79.88% lower than that of the original algorithm. Finally, the speed reaches 32.67 frame/s, which is 19.98% higher than that of the original algorithm. The new model trained by the proposed algorithm has higher accuracy and faster detection speed, which greatly improves the overall recognition rate of bird detection and balances the loss weight of positive and negative samples.

Key words object detection; depthwise separable residual model; generalized intersection over union loss function; YOLOv3 algorithm

1 引言

河北衡水湖国家级自然保护区是众多鸟类越冬

迁徙的必经之地。此前公园内的鸟类检测, 仅仅依靠人工, 存在工作量大、人力成本高和数据不准确的问题。

收稿日期: 2021-07-23; 修回日期: 2021-08-07; 录用日期: 2021-08-23

基金项目: 中国留学基金委地方合作项目(201808130283)、中国教育部人工智能协同育人项目(201801003011)、河北科技大学
校立课题(82/1182108)

通信作者: *songziying1997@gmail.com

随着基于深度学习的目标检测算法^[1-3]的迅速发展,无人监守的鸟类检测成为可能。自然场景下进行鸟类检测任务时存在很多困难。一方面,背景特征和鸟类目标特征很相似,鸟类目标和背景往往难以分清楚,并且鸟类目标往往只占到整个图像很小的一部分,小目标鸟类难以检测,存在鸟类目标和背景类不平衡的问题^[4]。另一方面,目前的目标检测算法为了更好地拟合鸟类目标,增加卷积层,导致参数量和计算量很大,检测速度不能达到鸟类目标检测任务的实时要求^[5]。

针对鸟类的目标检测,国内外研究人员提出了多种解决方法。2010年,尹永宜等^[6]采用帧间差分法对鄱阳湖的鸟类进行检测。通过估计背景运动的方式去除背景的噪声干扰,在背景简单和场景单一的条件下提高了整体的检测精度但并没有很好地解决鸟类目标和背景类不平衡的问题。2012年,Shakeri等^[7]采用高斯混合模型进行背景去除。该模型满足实时性的要求,但是由于去除背景的缘故,非常多不易检测的鸟类目标也一同被去除。2014年,陈灿等^[8]使用背景剪除法 and 连续适应性均值移动算法对青海湖鸟类监控视频进行了斑头雁目标的检测和跟踪,所采用的算法具有较低的计算复杂度和较高的鲁棒性,但是仅仅对图像中的大目标鸟类进行了检测,无法解决复杂环境导致正负样本不平衡的问题。2017年,翁雨辰等^[9]提出了一种基于深度区域网络的卷积神经网络架构,即VGG16+RCNN,该架构解决了细粒度鸟类检测的问题,整体检测精度非常高,但是由于采用了标准卷积的主干网络结构,计算量和参数量非常高,并不能满足鸟类的实时性检测要求。2018年,李新叶等^[10]提出了一种基于Faster RCNN联合语义提取和检测的分类方法。通过在卷积层引入自上而下的候选区域建议方法,该方法在鸟类细粒度检测上取得了非常好的检测效果,但是采用候选区域建议,导致计算量非常大并且检测速度并没有达到鸟类目标检测的实时性要求。2020年,陈咏秋等^[11]提出了基于YOLOv3算法的输电线路鸟类检测模型。该模型使用残差网络作为主干网络,采用多尺度检测策略提高对小目标鸟类的检测,满足了目标检测实时性的要求,但是并没有解决复杂环境下正负样本不平衡的问题。

针对目前目标检测算法进行鸟类检测任务时所存在的问题,本文提出了一种针对自然场景下鸟类的多目标检测算法(YOLOBIRDS)。一方面,对特征提取网络模型进行修改。针对主干网络存在参数量和计算量大的问题,对深度可分离卷积进行残差连接,对深层网络特征和浅层网络特征进行结合,既降低了参数量和计算量,也大幅度提升了检测精度。另一方面,对损失函数进行修改。目标框大小与位置损失函数由均方误差修改为广义交并比(CIoU),置信度损失函数增加正负样本控制参数,解决均方误差易受到目标尺度影

响和正负样本不平衡的问题。

2 YOLOBIRDS 算法

2.1 YOLOBIRDS 网络结构

采用轻量级网络模型^[12-14]提取图像特征,即深度可分离卷积神经网络,如图1所示。与传统卷积运算相比,深度可分离卷积参数量少并且运算成本低^[15]。传统卷积算法中,输出特征大小为 $H \times W \times C$,参数量为 $N \times k \times k \times C$,计算量为 $H \times W \times k \times k \times N \times C$ ^[16]。

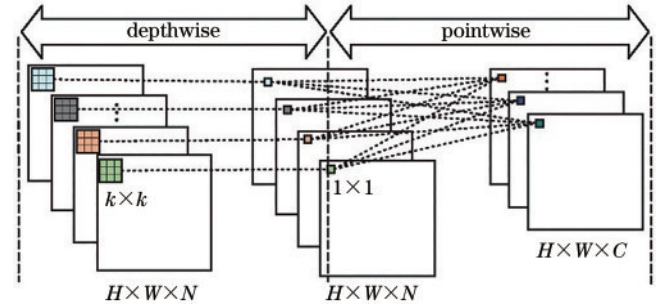


图1 深度可分离卷积过程

Fig. 1 Depthwise separable convolution process

深度可分离卷积参数量为 $N \times 1 \times 1 \times C$,计算量为 $H \times W \times 1 \times 1 \times N \times C$,与标准卷积相比,参数量和计算量同时下降 $k \times k$ 。深度可分离卷积由两部分组成:

- 1) 深度卷积,同一个信道中输入 $k \times k$ 的卷积,信道是独立的,信道之间不会进行信息的交互和计算,参数量为 $N \times k \times k$,计算量为 $H \times W \times k \times k \times N$;
- 2) 逐点卷积,采用 $1 \times 1 \times N$ 的 C 卷积融合信道间的信息,信道不是独立的,可以进行信道之间的通信和计算^[17]。

深度可分离残差模型在深度可分离卷积基础上引入残差连接,引入残差连接的目的是将深层特征和浅层特征结合,解决深层卷积发散的问题并学习浅层特征的具体物理特征^[18]。

YOLOBIRDS的整个网络结构如图2所示。模型的输入形状是(608, 608, 3),分别表示宽度、高度和通道数。选择608作为输入图像的宽度和高度的原因是YOLOBIRDS有19个网格,608=32×19。通过对传统目标检测算法的经验分析,32像素是作为一个栅格长宽大小的最佳选择^[19]。对输入图像进行深度可分离残差模型处理,卷积核为(3, 3, 32),3表示卷积核的宽度和高度,32表示卷积核的个数。在深度可分离残差模型运算后,输出形状分别为(608, 608, 32)、(304, 304, 64)、(152, 152, 128)、(76, 76, 256)、(38, 38, 512)、(19, 19, 1024)。最后,需要融合多尺度特征的最后3层,即(76, 76, 256)、(38, 38, 512)、(19, 19, 1024),分别代表小尺度、中尺度、大尺度目标检测。

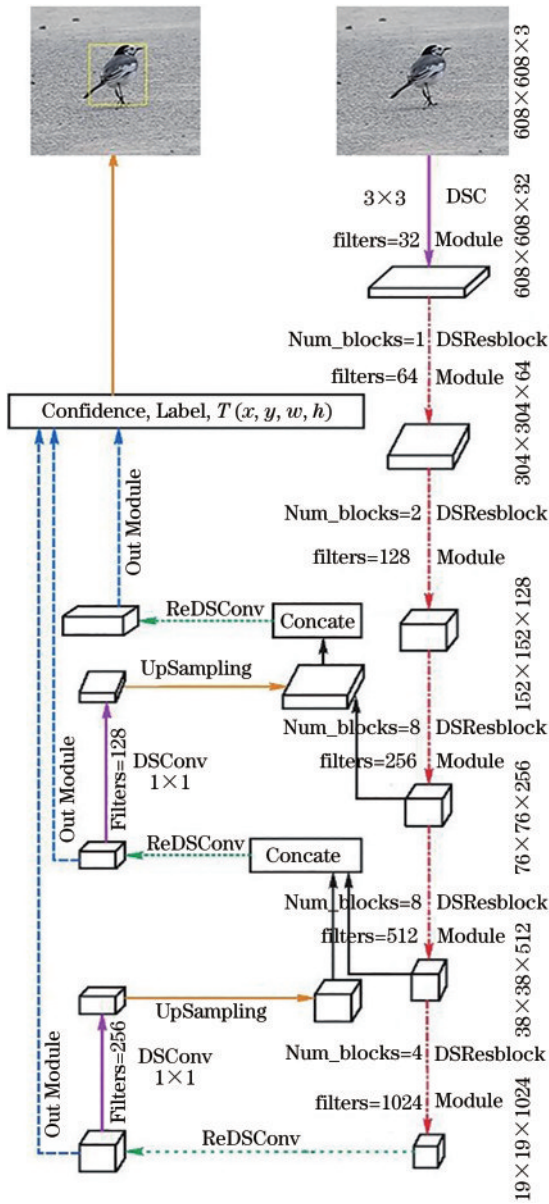


图2 YOLOBIRDS的整个网络结构
Fig. 2 Schematic of YOLOBIRDS structure

2.2 YOLOBIRDS 损失函数

YOLOv3算法的损失函数由3部分组成。

1) 目标框大小与位置损失函数,采用均方误差损失函数进行回归。此函数无法反映预测目标框和真实目标框的检测结果,对目标框的尺度非常敏感,易受到大尺度目标框的影响,导致反向传播回去的权值更偏向大目标的物体,不利于对小目标物体的检测。

2) 类别损失函数,采用交叉熵损失函数进行目标分类。

3) 置信度损失函数,采用交叉熵损失函数确定目标框的置信度,置信度损失函数由两部分组成,分别是正样本(目标类)损失函数和负样本(背景类)损失函数^[20]。此函数无法控制正负样本权重,会带来正负样本不平衡,由于背景类(负样本)权重太大,部分背景类检测为目标类(正样本)。

针对均方误差不足,本文采用CIoU^[21]作为目标框大小与位置损失函数,将重叠面积、中心点距离和长宽比加以考虑。CIoU损失函数表达式为

$$R_{IoU} = \frac{I}{U}, \quad (1)$$

$$R_{DioU} = \frac{r_E^2(\mathbf{b}, \mathbf{b}^{gt})}{c^2}, \quad (2)$$

$$R_{CIoU} = R_{DioU} + a\vartheta, \quad (3)$$

$$\mathcal{L}_{CIoU} = 1 - R_{CIoU}, \quad (4)$$

式中: U 为两目标框面积的并集; I 为两目标框面积的交集; r_E 代表欧氏距离; \mathbf{b} 代表预测值的目标框的中心点; \mathbf{b}^{gt} 代表真实值的目标框的中心点; c 是目标框真实值和预测值之间最小交集框的对角线长度; $\vartheta = \frac{4}{\pi^2} \left(\arctan \frac{\omega^{gt}}{h^{gt}} - \arctan \frac{\omega}{h} \right)$ 度量长宽比的一致性, ($\omega, h, \omega^{gt}, h^{gt}$)分别代表预测目标框的宽度、预测目标框的高度、真实目标框的宽度、真实目标框的长度;

$a = \frac{\vartheta}{(1 - R_{IoU}) + \vartheta}$ 是一个正的折中参数。

在目标检测任务中,样本分为正负样本。其中正样本表示真实目标框以内的图像区域,即前景类;负样本表示真实目标框的图像区域,即背景类^[22]。对于目标检测算法,主要需要关注的是对应真实物体的正样本,在训练时会根据其损失值来调整网络参数。相比之下,负样本对应着图像的背景,如果有大量的负样本参与训练,会使得损失函数更受到负样本特征的影响,从而降低网络收敛效率与检测精度^[23]。对背景类置信度进行学习的目的是增强模型的泛化能力,学习到非目标的特征,忽略背景中像目标而不是目标的特征。

在自然场景的鸟类检测数据中,绝大多数的目标检测样本是小目标而且图像背景复杂,影响对鸟类目标的检测^[24]。小目标意味着负样本的信息明显多于正样本,从而使得目标检测算法学习到了更多背景的信息。自然场景下正负样本不平衡的影响如图3所示,算法想正确地检测出左上侧白鹳目标,却错误地将右下侧的目标认为是白鹳目标。为了不检测背景中非目标的目标框,引用Focal loss,以此控制正负样本



图3 自然场景下正负样本不平衡的影响
Fig. 3 Influence of positive and negative sample imbalance in natural scene

的权重,减小大量背景类的权重,更侧重于前景类的检测。

一阶段检测器通过卷积神经网络滑动窗口进行卷积运算,这带来了一个严重的缺点,即学习到很多背景类的特征,受背景类的影响,由于图像中背景类面积占比过大,出现严重的正负样本不均衡^[25]。采用 Focal loss 作为置信度的损失函数,通过超参数 a_1 和 r 来控制正负样本权重^[26]。本质上来说,交叉熵损失函数是 Focal loss 的特殊形式,即 $a_1=1, r=1$ ^[27]。本文采用 $a_1=0.25, r=2$, 目前实验测试最好的值^[28]。Focal loss 使模型在训练时能够避免被一阶段算法存在的大量负样本转移注意力^[29]。Focal loss 函数为

$$L(p_i) = -a_1(1-p_i)^r \log(p_i), \quad (5)$$

式中: a_1 是一个控制正负样本权重的超参数; r 是一个控制样本分类难度的超参数,即调制系数; p_i 是预测样本的概率。

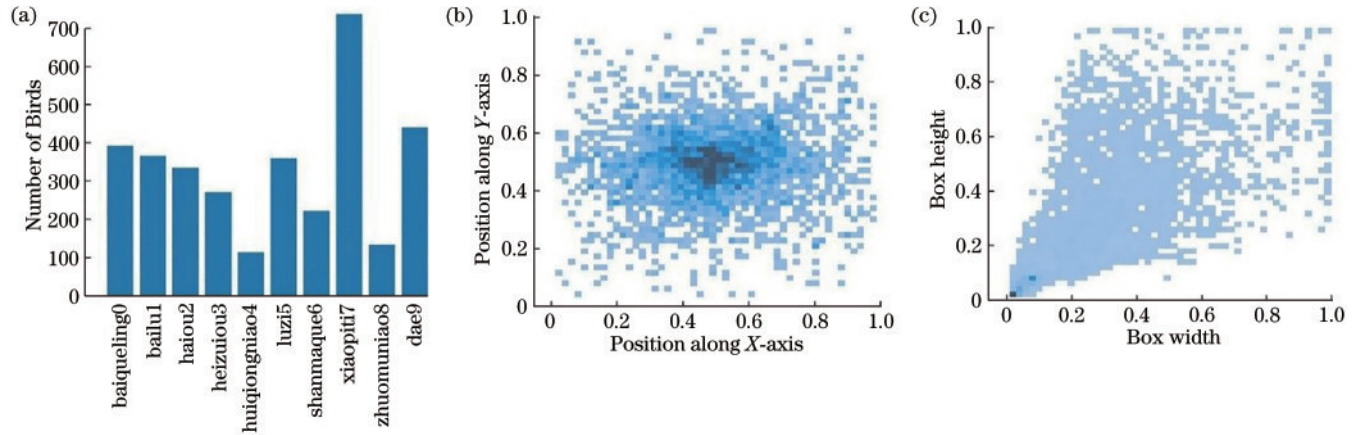


图4 鸟类目标统计结果。(a)鸟类数量分布图;(b)鸟类位置分布图;(c)鸟类目标框大小

Fig. 4 Statistical results of bird objects. (a) Statistic of number of birds; (b) bird location; (c) bird box size

3.2 实验结果与分析

实验在 Windows10 下进行。深度学习框架是 TensorFlow。使用的图形卡配置为 2 Nvidia GeForce RTX 2080 Ti。显卡为 22 GB。

在 $R_{iou}=0.5$ 时,利用 10 种鸟类对 YOLOBIRDS 算法与 YOLOv3、SSD300、Faster RCNN 算法进行了比较,对各种鸟类的平均精确度(AP)如表 1 所示。YOLOBIRDS 在 10 种鸟类中对白鹤鸽、黑嘴鸥、小鸊鷉和灰琼鸟这 4 种鸟类的 AP 较好,并且在黑嘴鸥上比 YOLOv3 算法高 8.53 个百分点。说明 YOLOBIRDS 算法对单个鸟类的 AP 表现非常好。

同时在 $R_{iou}=0.5$ 时,4 种算法的检测精度、检测速度和模型尺寸的比较结果如表 2 所示。YOLOBIRDS 算法的平均精确度的均值(mAP)可以达 87.12%,与 YOLOv3、SSD300、Faster RCNN 相比,分别提高了 2.71 个百分点、15.89 个百分点、0.76 个百分点。在检测速度方面, YOLOBIRDS 算法远远优于 Faster

3 实验数据集与结果分析

3.1 实验数据集

为了评价 YOLOBIRDS 算法在真实自然环境中的检测性能,本文的鸟类检测数据集来自高清摄像头拍摄的衡水湖图像,采用 10 种鸟类进行检测。对数据集的鸟类目标进行统计,中国衡水湖鸟类数据集鸟类目标统计结果如图 4 所示。其中,图 4(a)是鸟类目标的数量统计图,小鸊鷉(xiaopiti7)的数量是最多的,啄木鸟(zhuomuniao8)和灰琼鸟(huiqiongniao4)是最少的,图 4(a)鸟类目标可以真实反映中国衡水湖鸟类的分布和摄像头拍摄的情况,因此没有对鸟类目标的数量进行平均处理。图 4(b)是鸟类目标的位置分布图,鸟类目标位置以中间位置为主。图 4(c)是鸟类目标框所占图像的面积,反映了小目标占比最多。采用 Cutmix 数据增强策略^[30],增加了输入图像的可变性,丰富了图像的特征信息。最后,每个类别的数据量超过 3000 个。

表 1 不同的算法对鸟类的 AP

Table 1 AP of different algorithms for different birds unit:%

Bird class	SSD300	YOLOv3	Faster RCNN	YOLOBIRDS
baiqueling0	71.85	87.78	88.54	91.15
bailu1	73.78	87.65	90.12	88.57
haiou2	87.82	88.97	86.32	82.12
heizuiou3	69.49	84.32	83.54	92.85
huiqiongniao4	90.56	91.12	92.39	91.32
luzi5	63.89	87.75	89.43	86.48
shanmaque6	61.43	78.57	82.45	79.58
xiaopiti7	79.33	85.67	88.23	91.82
zhuomuniao8	60.64	75.34	83.44	87.67
dae9	53.51	76.93	79.14	79.64

RCNN 算法。与 YOLOv3 算法相比, YOLOBIRDS 算法的模型大小只有原 YOLOv3 算法的 1/5 左右。浮点运算数(FLOPs)衡量目标检测算法的时间复杂度,

表 2 不同算法的整体性能对比

Table 2 Overall model performance comparison of different algorithms

Method	mAP / %	Speed / (frame · s ⁻¹)	Access inventory / MB	Number of parameters	FLOPs / 10 ⁹
SSD300	71.23	45.64	107.51		
YOLOv3	84.41	27.23	254.03	61760674	60.68
Faster RCNN	86.36	12.52	588.59		
YOLOBIRDS	87.12	32.67	132.72	12425917	24.22

YOLOBIRDS 算法的 FLOPs 仅为 YOLOv3 的 39.91%。对于模型的空间复杂度,即访存量(access inventory),YOLOBIRDS 的访存量仅为 YOLOv3 的 52.25%。通过改进主干网络的网络结构,YOLOBIRDS 算法的复杂度充分降低,加快了模型的推理速度。

YOLOBIRDS 和 YOLOv3 系列的对比如表 3 所示。改变 YOLOv3 的主干网络,与不同的特征提取网络进行性能对比,YOLOBIRDS 的 mAP 最好,但运行速度仅次于 VGG19。

表 3 YOLOBIRDS 与 YOLOv3 系列对比
Table 3 Comparison between YOLOBIRDS and YOLOv3 series

Method	mAP / %	Speed / (frame · s ⁻¹)
YOLOv3 (origin, darknet53)	84.41	27.23
YOLOv3 (VGG19)	78.69	38.15
YOLOv3 (Resnet50)	79.58	24.59
YOLOv3 (Inceptionv4)	82.89	26.25
YOLOv3 (DenseNet)	84.51	27.97
YOLOv3 (SENet)	82.54	30.18
YOLOv3 (DualPathNet)	86.46	31.65
YOLOBIRDS	87.12	32.67

当 IoU 分别为 0.5, 0.6, 0.7 时,比较了 YOLOBIRDS 算法与 SSD300、YOLOv3、Faster RCNN 算法的检测精度,比较结果如表 4 所示。从表 4 可以看出:随着 IoU 阈值的提高,YOLOBIRDS 算法在检测精度上比

表 4 不同 IoU 下不同算法的 mAP 对比

Table 4 Comparison of mAP of different algorithms under different IoU thresholds unit: %

Method	R _{IoU} =0.5	R _{IoU} =0.6	R _{IoU} =0.7
SSD300	71.23	62.39	55.72
YOLOv3	84.41	72.33	67.32
Faster RCNN	86.36	77.38	69.33
YOLOBIRDS	87.12	76.27	70.34

其他三种算法有越来越明显的优势,在 IoU 阈值较高的情况下,仍然可以保持较高的 mAP 值;另外 4 种算法的 mAP 随着 IoU 阈值的增加而急剧下降。这表明 YOLOBIRDS 算法得到的预测框与真实目标框有较高的吻合度,物框定位更准确。与 SSD300、YOLOv3 和 Faster RCNN 相比,YOLOBIRDS 在不同 IoU 下表现更好,更稳定。

当大于置信度阈值时,模型认为检测中存在目标,在检测框中,首先用置信度阈值进行过滤。当检测框的置信度大于该阈值时,认为检测框中有目标;否则,认为没有目标。当置信度阈值为 0.25、0.45、0.65 时,不同模型的指标参数如表 5 所示。从表 5 可以看出,YOLOBIRDS 的 F1_Score 略好于 Faster RCNN 等算法。

从测试集中选择了具有代表性的小目标鸟类图片展开对比分析,如图 5 所示。图 5(a)对照组 1 是受光照影响严重的图像,图 5(b)对照组 2 是受大雾影响的图像,图 5(c)对照组 3 是正常拍摄的图像。

表 5 不同置信度阈值下不同算法的指标参数

Table 5 Index parameters of different algorithms under different confidence thresholds

Method	Precision / %			Recall / %			F1_Score / %		
	0.25	0.45	0.65	0.25	0.45	0.65	0.25	0.45	0.65
SSD300	69.86	72.76	81.83	65.09	73.28	83.29	67.39	73.02	82.55
YOLOv3	62.47	74.87	82.28	69.12	76.27	84.91	65.63	75.56	83.57
Faster RCNN	63.39	78.63	83.09	66.93	78.92	86.53	65.11	78.77	84.78
YOLOBIRDS	66.34	79.37	85.13	69.42	78.20	85.91	67.85	78.78	85.52

对比图 5(a)中的实验结果。YOLOBIRDS 的正确检测结果最多,其次是 Faster RCNN 和 YOLOv3, SSD300 的正确检测结果最少。受光照的影响,SSD300 有 1 个错误的检测,将整张图片错误地检测为白鹭,仅有 1 个正确的检测;YOLOv3 和 Faster RCNN 分别有 2 个和 3 个正确的检测;YOLOBIRDS 正确检

测的数量最多,没有错误检测的目标,未检测的鸟类目标是 13 个。结果表明,YOLOBIRDS 算法在具有代表性的小目标检测图像上有明显的提升效果。对比图 5(b)中的实验结果。YOLOBIRDS 的正确检测结果最多,其次是 Faster RCNN 和 YOLOv3, SSD300 的正确检测结果最少。与 YOLOv3 相比,YOLOBIRDS

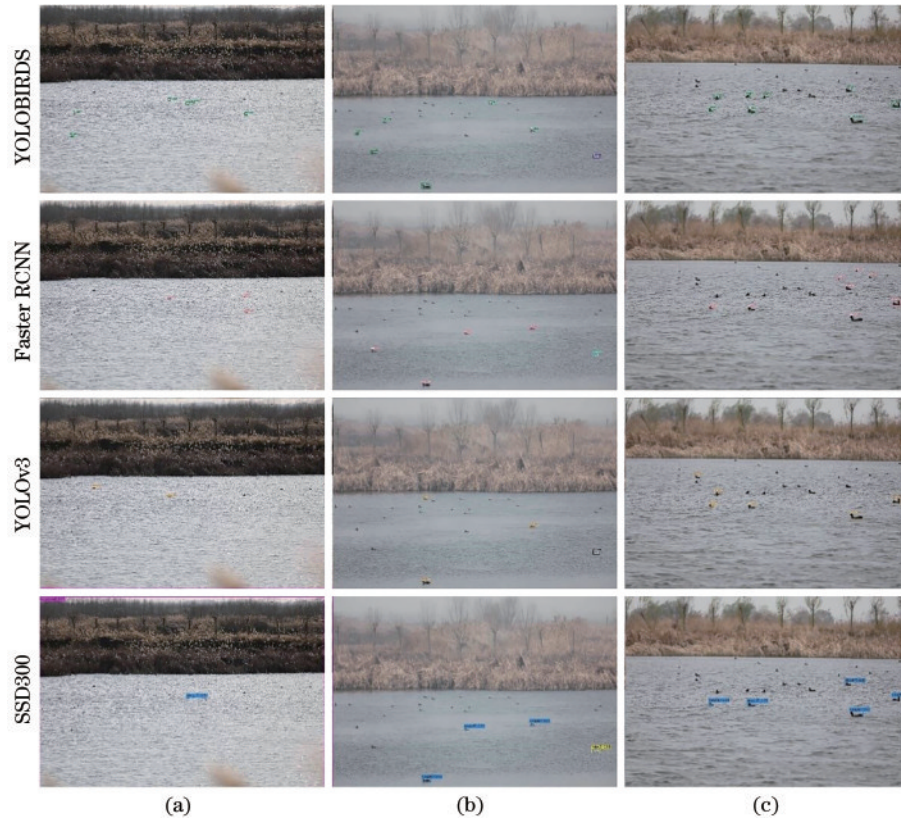


图 5 不同算法的实验结果对比。(a)受光照影响严重的图像;(b)受大雾影响的图像;(c)正常拍摄的图像

Fig. 5 Comparison of experimental results of different algorithms. (a) Images seriously affected by light; (b) images affected by fog; (c) images taken normally

在正确检测上多 3 个目标,未检测的少 3 个,都没有出现错误检测的目标。结果表明,YOLOBIRDS 算法在具有代表性的小目标检测图像上有明显的提升效果。对比图 5(c)中的实验结果。YOLOBIRDS 和 Faster RCNN 的正确检测结果最多,其次是 YOLOv3, SSD300 的正确检测结果最少。相比 YOLOv3, Faster RCNN 算法在正确检测上多 3 个目标。然而, YOLOBIRDS 算法与 Faster RCNN 算法相比并没有提升。结果表明, YOLOBIRDS 算法在具有代表性的小目标检测图像上有明显的提升效果。

4 结 论

针对自然场景下鸟类的多目标检测算法出现的问题,提出 YOLOBIRDS 算法。该算法解决了鸟类检测任务中模型参数量大、计算量高和正负样本严重不平衡的问题。一方面,引入轻量级主干网络,减少计算量和参数量,增加检测速度。另一方面,对损失函数进行改进,目标框大小与位置损失函数由均方误差修改为 CIoU,置信度损失函数增加正负样本控制参数。实验数据表明,采用 YOLOBIRDS 算法训练得到的新模型降低了模型参数量和计算量,平衡了正负样本权重。但当图像中遮挡部分是较大的目标时,新模型不能进行精确的检测,后续工作是将通过对抗生成网络生成遮挡较大鸟类的特征信息,将其特征信息作为数据集

的补充加入到鸟类特征训练中,以提高对遮挡较大鸟类的检测效果。

参 考 文 献

- [1] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [2] 段仲静, 李少波, 胡建军, 等. 深度学习目标检测方法及其主流框架综述[J]. 激光与光电子学进展, 2020, 57(12): 120005.
Duan Z J, Li S B, Hu J J, et al. Review of deep learning based object detection methods and their mainstream frameworks[J]. Laser & Optoelectronics Progress, 2020, 57(12): 120005.
- [3] 韩永赛, 马时平, 李帅, 等. 基于优化区域卷积神经网络的机场区域检测[J]. 激光与光电子学进展, 2020, 57(10): 101021.
Han Y S, Ma S P, Li S, et al. Airport area detection based on optimized regional convolutional neural network [J]. Laser & Optoelectronics Progress, 2020, 57(10): 101021.
- [4] 刘万军, 王凤, 曲海成. 融合多尺度特征的目标检测模型[J]. 激光与光电子学进展, 2019, 56(23): 231007.
Liu W J, Wang F, Qu H C. Object detection model based on multi-scale feature integration[J]. Laser & Optoelectronics Progress, 2019, 56(23): 231007.
- [5] Redmon J, Divvala S, Girshick R, et al. You only look

- once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.
- [6] 尹永宜, 李艳. 基于综合匹配的鸟类视频中运动物体的检测[J]. 遥感信息, 2010, 25(5): 35-39.
Yin Y Y, Li Y. The object detection in birds video based on synthesized matching method[J]. Remote Sensing Information, 2010, 25(5): 35-39.
- [7] Shakeri M, Zhang H. Real-time bird detection based on background subtraction[C]//Proceedings of the 10th World Congress on Intelligent Control and Automation, July 6-8, 2012, Beijing, China. New York: IEEE Press, 2012: 4507-4510.
- [8] 陈灿, 熊妍, 阎保平. 野外科研视频中鸟类目标的检测和跟踪技术研究[J]. 科研信息化技术与应用, 2014, 5(1): 53-58.
Chen C, Xiong Y, Yan B P. Wild birds target detection and tracking technology in video system[J]. E-Science Technology & Application, 2014, 5(1): 53-58.
- [9] 翁雨辰, 田野, 路敦民, 等. 深度区域网络方法的细粒度图像分类[J]. 中国图象图形学报, 2017, 22(11): 1521-1531.
Weng Y C, Tian Y, Lu D M, et al. Fine-grained bird classification based on deep region networks[J]. Journal of Image and Graphics, 2017, 22(11): 1521-1531.
- [10] 李新叶, 王光陆. 基于卷积神经网络语义检测的细粒度鸟类识别[J]. 科学技术与工程, 2018, 18(10): 240-244.
Li X Y, Wang G B. Fine-grained bird recognition based on convolution neural network semantic detection[J]. Science Technology and Engineering, 2018, 18(10): 240-244.
- [11] 陈咏秋, 孙凌卿, 张永泽, 等. 基于 YOLO v3 的输电线路鸟类检测技术研究[J]. 计算机工程, 2020, 46(4): 294-300.
Chen Y Q, Sun L Q, Zhang Y Z, et al. Research on bird detection technology for electric transmission line based on YOLO v3[J]. Computer Engineering, 2020, 46(4): 294-300.
- [12] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 21-37.
- [13] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(6): 1137-1149.
- [14] Azimi S M, Vig E, Bahmanyar R, et al. Towards multi-class object detection in unconstrained remote sensing imagery[M]//Jawahar C V, Li H D, Mori G, et al. Computer vision-ACCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11363: 150-165.
- [15] Yang X, Yang J R, Yan J C, et al. SCRDet: towards more robust detection for small, cluttered and rotated objects[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 8231-8240.
- [16] Duan K W, Bai S, Xie L X, et al. CenterNet: keypoint triplets for object detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 6568-6577.
- [17] Mao Q C, Sun H M, Liu Y B, et al. Mini-YOLOv3: real-time object detector for embedded applications[J]. IEEE Access, 2019, 7: 133529-133538.
- [18] 李竺强, 朱瑞飞, 马经宇, 等. 联合连续学习的残差网络遥感影像机场目标检测方法[J]. 光学学报, 2020, 40(16): 1628005.
Li Z Q, Zhu R F, Ma J Y, et al. Airport detection method combined with continuous learning of residual-based network on remote sensing image[J]. Acta Optica Sinica, 2020, 40(16): 1628005.
- [19] Wu Z F, Shen C H, van den Hengel A. Wider or deeper: revisiting the ResNet model for visual recognition[J]. Pattern Recognition, 2019, 90: 119-133.
- [20] Howard A G, Zhu M L, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17)[2021-02-03]. <https://arxiv.org/abs/1704.04861>.
- [21] Rezatofghi H, Tsoi N, Gwak J Y, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 658-666.
- [22] 江金洪, 鲍胜利, 史文旭, 等. 基于 YOLO v3 算法改进的交通标志识别算法[J]. 计算机应用, 2020, 40(8): 2472-2478.
Jiang J H, Bao S L, Shi W X, et al. Improved traffic sign recognition algorithm based on YOLO v3 algorithm [J]. Journal of Computer Applications, 2020, 40(8): 2472-2478.
- [23] 王殿伟, 方浩宇, 刘颖, 等. 一种基于改进 SiameseRPN 的全景视频目标跟踪算法[J]. 激光与光电子学进展, 2020, 57(24): 241008.
Wang D W, Fang H Y, Liu Y, et al. Algorithm for panoramic video tracking based on improved SiameseRPN [J]. Laser & Optoelectronics Progress, 2020, 57(24): 241008.
- [24] Alom M Z, Yakopcic C, Nasrin M S, et al. Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network [J]. Journal of Digital Imaging, 2019, 32(4): 605-617.
- [25] 睢丙东, 张湃, 王晓君. 一种改进 YOLOv3 的手势识别算法[J]. 河北科技大学学报, 2021, 42(1): 22-29.
Sui B D, Zhang P, Wang X J. A gesture recognition algorithm based on improved YOLOv3[J]. Journal of Hebei University of Science and Technology, 2021, 42(1): 22-29.
- [26] 陈科圻, 朱志亮, 邓小明, 等. 多尺度目标检测的深度学习研究综述[J]. 软件学报, 2021, 32(4): 1201-1227.
Chen K Q, Zhu Z L, Deng X M, et al. Deep learning for multi-scale object detection: a survey[J]. Journal of Software, 2021, 32(4): 1201-1227.

- [27] Woo S, Park J, Lee JY, et al. CBAM: Convolutional Block Attention Module[C]//Proceedings of the Proceedings of the European Conference on Computer Vision (ECCV), September 8-14, 2018, Munich, Germany. New York: IEEE Press, 2018: 3-19.
- [28] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2999-3007.
- [29] Tao R S, Wei Y L, Jiang X J, et al. Towards real-world X-ray security inspection: a high-quality benchmark and lateral inhibition module for prohibited items detection [EB/OL]. [2021-08-23]. <https://arxiv.org/abs/2108.09917>.
- [30] Yang X, Sun H, Fu K, et al. Automatic ship detection in remote sensing images from google earth of complex scenes based on multiscale rotation dense feature pyramid networks[J]. Remote Sensing, 2018, 10(1): 132.