

针对遥感影像的 MSA-YOLO 储油罐目标检测*

李 想^{1,2}, 特日根^{1,2}, 赵宇恒^{1,2}, 陈文韬^{1,2}, 徐国成³

(1. 长光卫星技术股份有限公司, 吉林 长春 130000;

2. 吉林省卫星遥感应用技术重点实验室, 吉林 长春 130000;

3. 吉林大学 材料科学与工程学院, 吉林 长春 130000)

摘 要: 原油作为一种重要的战略物资, 在我国经济和军事等多个领域均起到重要作用。提出一种算法 MSA-YOLO (MultiScale Adaptive YOLO), 该算法在 YOLOv4 算法的基础上进行优化, 并基于以吉林一号光学遥感卫星影像为主的遥感图像数据集进行实验, 对特定监控区域内的储油罐进行识别与分类。算法优化内容包括: 为简化储油罐监测模型同时保证模型的效率, 对 YOLOv4 的网络结构中的多尺度识别模块进行修剪; 使用 k-means++ 聚类算法进行初始锚框的选取, 使模型加速收敛; 使用基于 CIoU-NMS 的优化, 进一步提升推理速度和准确度。实验结果表明, 与 YOLOv4 相比, MSA-YOLO 模型参数数量减少 25.84%; 模型尺寸减少 62.13%; 在 Tesla V100 的 GPU 环境下, 模型的训练速度提升 6 s/epoch, 推理速度提升 15.76 F/s; 平均精度为 95.65%。与此同时, MSA-YOLO 算法在多种通用目标识别算法进行的对比实验中均体现出了更高效的特点。MSA-YOLO 算法对储油罐进行准确且实时的识别具有通用可行性, 可为遥感数据在能源期货领域提供技术参考。

关键词: 计算机视觉; 目标检测; 深度学习; YOLO; 储油罐检测

中图分类号: TP75

文献标识码: A

DOI: 10.16157/j.issn.0258-7998.223191

中文引用格式: 李想, 特日根, 赵宇恒, 等. 针对遥感影像的 MSA-YOLO 储油罐目标检测[J]. 电子技术应用, 2022, 48(11): 24-32, 40.

英文引用格式: Li Xiang, Te Rigen, Zhao Yuheng, et al. MSA-YOLO oil storage tank target detection for remote sensing images[J]. Application of Electronic Technique, 2022, 48(11): 24-32, 40.

MSA-YOLO oil storage tank target detection for remote sensing images

Li Xiang^{1,2}, Te Rigen^{1,2}, Zhao Yuheng^{1,2}, Chen Wentao^{1,2}, Xu Guocheng³

(1. Chang Guang Satellite Technology Co., Ltd., Changchun 130000, China;

2. Main Laboratory of Satellite Remote Sensing Technology of Jilin Province, Changchun 130000, China;

3. School of Materials Science and Engineering, Jilin University, Changchun 130000, China)

Abstract: Crude oil, as an important strategic material, plays an important role in many fields such as my country's economy and military. This paper proposes an algorithm MSA-YOLO (MultiScale Adaptive YOLO), which is optimized on the basis of the YOLOv4 algorithm, and is experimented based on the remote sensing image dataset mainly based on Jilin-1 optical remote sensing satellite images, to make identification and classification of oil storage tanks. The algorithm optimization contents include: in order to simplify the oil storage tank monitoring model and ensure the efficiency of the model, prune the multi-scale identification module in the network structure of YOLOv4; use the k-means++ clustering algorithm to select the initial anchor frame to accelerate the convergence of the model; use CIoU-NMS-based optimization to further improve inference speed and accuracy. The experimental results show that compared with YOLOv4, the number of parameters of MSA-YOLO model is reduced by 25.84%; the model size is reduced by 62.13%; in the GPU environment of Tesla V100, the training speed of the model is increased by 6 s/epoch, and the inference speed is increased by 15.76 F/s; the average accuracy is 95.65%. At the same time, the MSA-YOLO algorithm shows more efficient characteristics in the comparative experiments of various general target recognition algorithms. The MSA-YOLO algorithm has universal feasibility for accurate and real-time identification of oil storage tanks, and can provide technical reference for remote sensing data in the field of energy futures.

Key words: computer vision; target recognition; deep learning; YOLO; storage tank detection

* 基金项目: 国家重点研发计划(2019YFE0127000, SQ2020YFA070264); 吉林省科技发展计划项目(BZYYBDKZ2020010101, BZYYBDKZ2020010102); 海南省重大科技计划项目(ZDKJ2019007); 吉林省重点研发项目(20200401094GX)

0 引言

近年来,随着高分辨率光学卫星遥感影像处理技术的快速发展,基于遥感影像的目标识别取得了大量成果。其中,对地表自然形成或人造物体进行识别一直是从业人员的关注重点之一。储油罐是在石油、天然气等石化行业中使用的设备,用于储存在环境温度下为液态的原油或者其他化工产品度下为液态的原油或者其他化工产品。按照储油罐的不同用途,分为固定顶型和外浮顶型。利用遥感影像的太阳高度角和内外阴影参数,可以对外浮顶储油罐的满油率进行估算,通过满油率数据在能源期货价格的预测模型中进行回归分析,不但可以为能源期货交易机构提供参考,还能对我国原油的采购及存储等起到指导作用。而在上述工作中,首要任务是在高分辨率遥感影像中实现固定顶和外浮顶储油罐的高效识别与分类。

对于储油罐检测算法,国内外研究人员在传统图像处理领域已经取得很多成果。早在 2013 年, Kushwaha 等提出了一种基于形态学方法和图像拆分/合并的分割技术,对明亮的固定顶圆形储油罐进行检测^[1]。在 2018 年, Jing 等提出了一种基于形状导向的显著性的储油罐检测框架,提取轮廓形状提示作为目标信息来指导选择性显著值的计算,消除高对比度背景的干扰,准确地定位到储油罐^[2]。随着卷积神经网络(Convolutional Neural Network, CNN)和机器学习技术的迅速发展^[3],许多基于卷积神经网络的模型在储油罐检测中表现出较好的结果。2015 年 Zhang 等人提出一种将卷积神经网络与方向梯度直方图特征相结合的方法,使用方向梯度直方图提取局部区域的形状信息,并使用卷积神经网络提取储油罐的周围特征^[4]。Girshick 等提出基于区域卷积神经网络^[5](Region-based Convolutional Neural Networks, R-CNN)和升级版本 Fast R-CNN^[6],取得了较高的检测精度。Ren 和 Girshick 等人又提出了 Faster R-CNN 算法^[7],进一步提高了目标检测效果,但仍存在检测速度缓慢的弊端。针对此类问题 Joseph Redmon 等在 2015 年提出了 YOLO 算法^[8],该算法将对象检测框架划为空间分隔边界框的回归问题,仅需要一步操作就可以完成目标检测任务。在之后的 2016 年和 2018 年又推出 YOLO9000^[9]和 YOLOv3^[10]两次升级版本,在检测的速度和准确率上都取得较好效果^[11-13]。在此基础上 Bochkovski 等人于 2020 年发布的 YOLOv4 对 YOLOv3 提出了五种改进方式和二十余个改进技巧,在平均精度和速度上也远超 YOLOv3 版本^[14],但由于高分辨率的遥感影像尺寸过大,同时 YOLOv4 模型复杂,推理时间仍比较慢,因此在储油罐检测工程化应用中存在适用性不足的问题。

针对传统的储油罐检测算法效率低下且达不到工程化标准,以及深度学习算法在使用场景上有欠缺的问题,本文提出了一种储油罐检测算法 MSA-YOLO,该算

法基于 YOLOv4 算法进行改进:对 YOLOv4 算法的网络层进行修剪,降低网络复杂度,在保证识别准确率的情况下提升推理速度;同时通过自适应锚框和优化 NMS 的方式,进一步提升模型收敛速度和识别准确率。

1 算法模型

1.1 YOLOv4 算法

Bochkovski 等人在 2020 年 4 月发布的 YOLOv4 算法,其 Backbone 使用了 CSPDarknet53 结构,该结构取消了算力消耗较高的计算瓶颈结构,同时减少了参数使之更容易训练。Neck 部分主要采用了空间金字塔池化^[15]和路径聚合网络^[16]的方式,可以更好地提取融合特征。Head 部分将网络的输出数据与真实标注数据对接,进行损失值的计算,同时重新组织数据格式,对特征图进行解码,并对原始网格单元中的 (x, y, w, h) 值分别做相应的激活等。

YOLOv4 中的多尺度检测策略通过上采样生成三个尺度的融合特征图,分别用来识别大、中、小三类物体,如表 1 所示,输入图像通过下采样生成 13×13 的大尺度特征图、 26×26 的中尺度特征图和 52×52 的小尺度特征图。以 COCO 数据集^[17]为例,大尺度特征图对应的锚框为 $(142, 110)$ 、 $(192, 243)$ 和 $(459, 401)$,中尺度特征图对应的锚框为 $(36, 75)$ 、 $(76, 55)$ 和 $(72, 146)$,小尺度特征图对应的锚框为 $(12, 16)$ 、 $(19, 36)$ 和 $(40, 28)$ 。

表 1 多尺度预测

特征图	感受野	锚框
13×13	大	$(142, 110)$, $(192, 243)$, $(459, 401)$
26×26	中	$(36, 75)$, $(76, 55)$, $(72, 146)$
52×52	小	$(12, 16)$, $(19, 36)$, $(40, 28)$

如图 1 所示,9 个默认锚框分别以蓝色边框(大)、绿色边框(中)和红色边框(小)进行区分,蓝色边框和红色



图 1 YOLOv4 默认锚框

边框的尺寸与真实储油罐尺寸差距较大,绿色边框的尺寸与真实储油罐尺寸较为一致。因此对于尺寸较为固定的储油罐检测中,默认锚框与目标的尺寸相差较大,不利于模型的加速收敛。与此同时,随着锚框数量的增加,预测滤波器中的卷积滤波器数量线性增加,这将导致网络规模变大,并增加训练时间。

在目标检测算法中,IoU(Intersection over Union)和NMS(Non-Maximum Suppression)具有重要的意义。IoU代表的是“预测边框”和“真实边框”的交集和并集的比值^[18]。在YOLOv4优化损失函数的过程中使用边界框回归的方式,IoU损失公式如式(1)所示:

$$\mathcal{L}_{IoU} = 1 - \frac{|BIB^{gt}|}{|BUB^{gt}|} \quad (1)$$

NMS用于确保出现在多个候选框中的物体仅被计算一次,滤出其他可预测同一对象的边界框,保留最高的置信度^[19]。NMS算法的步骤如下:

- (1)将所有边界框进行置信度大小排序;
- (2)保留置信度最高的边界框,移除与该边界框IoU大于阈值的边界框;
- (3)对于剩下的所有边界框执行步骤(2),直到无法继续移除边界框为止,算法结束。

在本文所述的高分辨率大尺寸遥感影像检测中,有两个步骤均使用到了NMS:

- (1)在预测过程中,使用NMS过滤检测框;
- (2)在将检测框融合到大尺寸原始图片上时,使用NMS过滤相邻小尺寸图片重叠部分的检测框。

基于IoU的NMS在过滤检测框的过程中存在着明显的缺点,即当两个检测框不相交时, \mathcal{L}_{IoU} 的值始终为1,无法给出优化的方向,回归速度慢。

1.2 遥感影像的目标识别

现阶段,在高分辨率大尺寸遥感影像中检测较小物体仍是一个重大挑战,以吉林一号光学遥感卫星拍摄的2021年6月的美国库欣区遥感影像为例,其尺寸为14 166像素×10 742像素,分辨率为0.75 m。该影像覆盖范围超过100平方公里,包含1.5亿个像素,而其中的待检测储油罐对象尺寸仅为几十像素。YOLOv4网络可识别的图像输入像素大小多为416×416、608×608等,与遥感影像的尺寸差距过大,无法直接进行高分辨率大尺寸遥感影像的识别。

2 算法优化

2.1 MSA-YOLO网络结构

在YOLOv4的检测过程中,对于尺寸不同的目标采取多尺度检测的方式。对于标准大小的储油罐,直径从6.1 m到91.5 m不等,直径超过61 m的储油罐被称为中大型储油罐^[20]。经调查研究,监控区域的储油罐大小直径从20 m到70 m不等,以中大型储油罐为主。在尺寸为416×416、分辨率为0.75 m的图片上,储油罐罐顶尺

寸为26.67像素至93.3像素,通过计算真实标注框数据,带有阴影的储油罐尺寸所在区间为54像素至133像素,该区域基本符合中尺度特征图的预测区间。由于储油罐的尺寸相对固定,因此为了降低模型的复杂度,减少计算量,加快推理速度,以应对大尺寸遥感影像的快速检测,需对YOLOv4的网络层执行剪枝操作^[21]。

为了适合储油罐尺寸特性,可将YOLOv4头部多尺度检测中用来识别大尺度和小尺度的分支进行剪枝,仅保留识别中尺度的分支,并将路径聚合网络的上采样模块与高层语义信息进行结合,建立储油罐的特征检测层。如图2所示,在MAS-YOLO算法的网络结构图中,训练数据经过CSPDarknet53骨干网络,提取目标特征;通过基于剪枝的YOLOv4头部模块,在中尺度特征图上训练出储油罐识别模型。

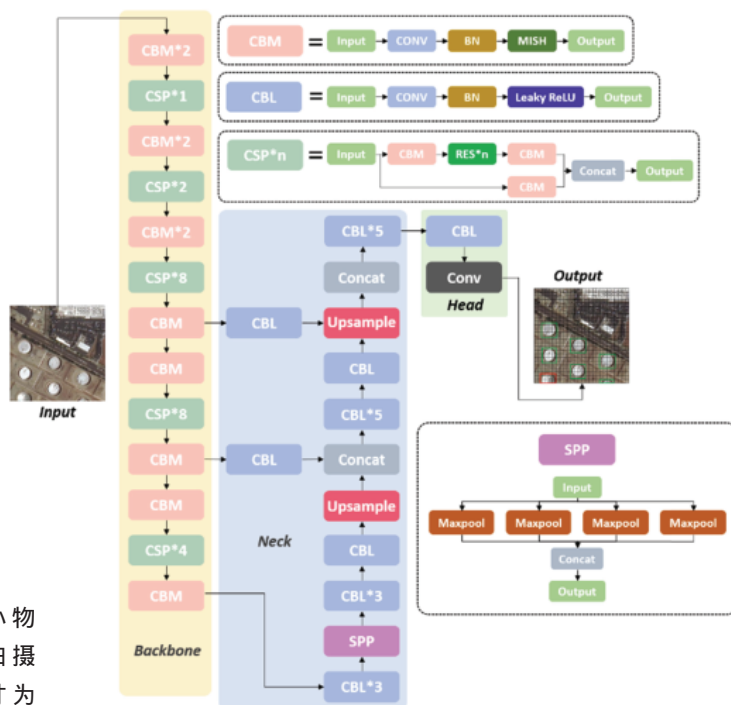


图2 MSA-YOLO算法网络结构图

2.2 自适应锚框

为了降低模型的复杂度,同时适应储油罐的尺寸特性,MAS-YOLO算法仅需保留中尺度特征图对应的3个锚框。在校正锚框尺寸的过程中,使用k-means++聚类算法^[22]。未使用k-means算法^[23],主要是因为两者对于聚类中心初始化的方式不同。其中k-means算法中随机初始化k个聚类中心,而k-means++算法生成k个相互距离更远的聚类中心点,最大程度地避免了聚类中心初始化敏感度的问题。聚类中心数量(下文用k表示)与IoU和IoU变化率的对应关系如图3所示。

- (1)IoU与k值呈正相关,k值越大,IoU越大。
- (2)当k<3时,IoU值较小且变化幅度较大;当k≥3

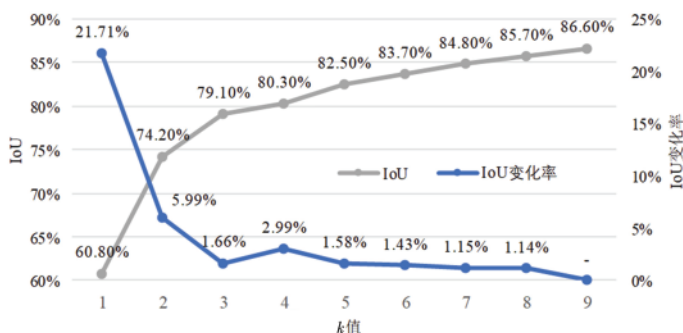


图3 不同聚类中心数量的平均IoU与平均IoU变化率

时,平均IoU大且变化幅度较小。

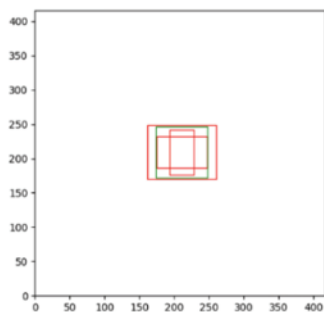
(3)当 $k < 3$ 时,IoU变化率大且变化幅度较大;当 $k \geq 3$ 时,平均IoU变化率小且变化幅度区域稳定。

由于随着 k 的增加,锚框尺寸趋于接近,因此取 $k=3$ 时,既可以减少锚框带来的误差,也可以加快训练的收敛速度。如表2所示,在初始锚框选择的应用中,当 $k=3$ 时, k -means算法的IoU为78.72%, k -means++算法的IoU为79.13%; k 为1至9时, k -means算法的平均IoU为78.74%, k -means++算法的平均IoU为79.74%。综上所述, k -means++平均IoU更大,精确度更高。

表2 聚类算法比较

聚类算法	$k=3$	\bar{k}
k -means	78.72	78.74
k -means++	79.13	79.74

因此,对数据集进行聚类中心为3的聚类分析,得到3个自适应锚框,尺寸分别为(60,32)、(42,67)和(72,91),在图4(a)中分别用红色边框表示;同时对数据集内全部对象尺寸进行统计,得到储油罐对象的平均尺寸为(68.2,68.3),在图4(a)中用绿色边框表示。如图4(a)所示,3个尺寸的自适应锚框均有宽(或高)度与平均尺寸保持一致,同时还保留自身的另一个高(或宽)度的独立性,以适应各种尺寸的储油罐对象。如图4(b)所示,三个自适应锚框基本可以分别覆盖大、中、小三种



(a) 自适应锚框尺寸与对象平均尺寸



(b) 自适应锚框与储油罐对应情况

图4 自适应锚框示意图

尺寸的储油罐,锚框的形状非常接近储油罐的真实形状,较默认锚框相比更为准确,以便MSA-YOLO网络更容易地学习对锚框的小幅调整,创建更准确的边界框。

2.3 CIOU-NMS

2.3.1 CIOU-NMS算法

针对两个检测框不相交时,传统IoU回归速度慢的缺点,GIoU^[24](Generalized Intersection over Union)对此进行了优化,如式(2)所示:

$$\mathcal{L}_{\text{GIOU}} = 1 - \text{IoU} + \frac{|C - \text{BUB}^{\#}|}{|C|} \quad (2)$$

其中 C 表示包含两个框的最小矩形,可以对两框不相交的情况进行优化。而当两个候选框完全相交时,GIoU损失则退化到IoU损失,收敛速度同样会减慢。DIOU^[25](Distance Intersection over Union)在GIOU的基础上,对于不同距离、方向、面积和比例的预选框都能做到较好的回归,如式(3)所示:

$$\mathcal{L}_{\text{DIOU}} = 1 - \text{IoU} + \frac{\rho^2(b, b^{\#})}{c^2} \quad (3)$$

其中 b 和 $b^{\#}$ 分别代表预选框和目标框的中心点, $\rho^2(b, b^{\#})$ 代表两个中心点的欧式距离, c 代表能同时覆盖两框的最小矩形的对角线距离。DIOU可以在预选框与目标框不重叠时,为预选框提供移动方向,也可以最小化两个目标框的距离。但当两个框中心点重合时, $\rho^2(b, b^{\#})$ 和 c 值都不变,DIOU此时需引入框的宽高比,即CIOU(Complete Intersection over Union),宽高比所占的比例与IoU呈正相关,最终,CIOU如式(4)所示:

$$\mathcal{L}_{\text{CIOU}} = 1 - \text{IoU} + \frac{\rho^2(b, b^{\#})}{c^2} + \alpha \nu \quad (4)$$

其中 α 是权重函数, ν 用来度量宽高比的一致性,如式(5)和式(6)。

$$\alpha = \frac{\nu}{(1 - \text{IoU}) + \nu} \quad (5)$$

$$\nu = \frac{4}{\pi} \left(\arctan \frac{\omega^{\#}}{h^{\#}} - \arctan \frac{\omega}{h} \right)^2 \quad (6)$$

图5为同一储油罐对象的真实边框与多种情况的预测边框示意图,其中 G 为真实边框,其余为预测边框,对每一种情况分别计算不同的IoU损失如表3所示。

结论如下:

- (1)对于图5(a),真实边框与预测边框有相交部分, $\mathcal{L}_{\text{IoU}}(A, G) \neq 1$,IoU可提供优化方向;
- (2)对于图5(b),真实边框与预测边框不相交, $\mathcal{L}_{\text{IoU}}(B, G) = 1$, $\mathcal{L}_{\text{GIOU}}(B, G) \neq 1$,IoU无法提供优化方向;
- (3)对于图5(c)、5(d)和5(e),真实边框与预测边框完全相交, $\mathcal{L}_{\text{GIOU}}(C, G) = \mathcal{L}_{\text{GIOU}}(D, G) = \mathcal{L}_{\text{GIOU}}(E, G)$, $\mathcal{L}_{\text{DIOU}}(C, G) \neq \mathcal{L}_{\text{DIOU}}(D, G) \neq \mathcal{L}_{\text{DIOU}}(E, G)$,GIOU无法提供优化方向;
- (4)对于图5(e)和5(f),两框中心点重合但预测边框宽高比不同, $\mathcal{L}_{\text{CIOU}}(F, G) \neq \mathcal{L}_{\text{DIOU}}(F, G) \neq \mathcal{L}_{\text{GIOU}}(F, G) \neq \mathcal{L}_{\text{IoU}}(F, G)$,

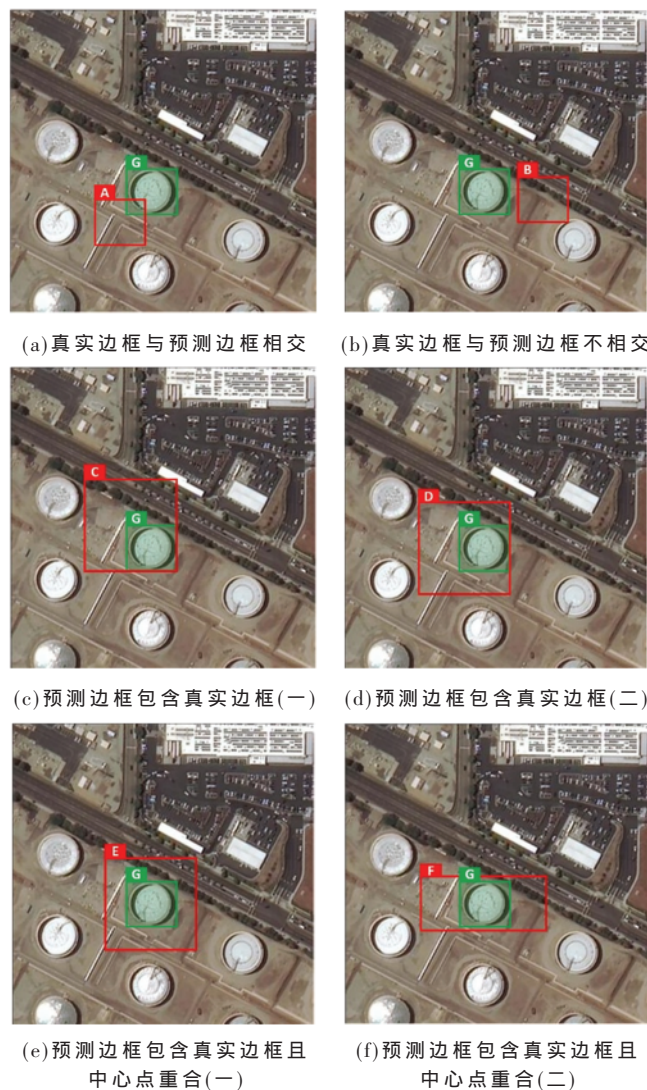


图5 同一储油罐对象的真实边框与多种预测边框

DIoU 无法提供优化方向,此时仅 CIoU 可以提供优化方向。与其他几种方法相比,CIoU 更符合目标框回归的机制,将目标与锚框之间的距离、重叠度和尺度都涵盖在内,使得目标框回归更加稳定,最大程度地避免了训练

表 3 不同的 IoU 方式与结果				
IoU 方式	\mathcal{L}_{IoU}	\mathcal{L}_{CIoU}	\mathcal{L}_{DIoU}	\mathcal{L}_{CIoU}
(A, G)	0.96	0.71	0.87	0.87
(B, G)	1.00	0.74	0.76	0.76
(C, G)	0.77	0.77	0.82	0.82
(D, G)	0.77	0.77	0.79	0.79
(E, G)	0.77	0.77	0.77	0.77
(F, G)	0.71	0.71	0.71	0.72

过程中的发散问题。

2.3.2 CIoU-NMS 在预测过程中的使用

对于尺寸为 416×416 的输入图像,经过目标识别算法网络,最终生成 26×26×2=1 352 个边界框。通过 NMS,从多个预测的边界框中保留最佳边界框。如图 6 所示,以测试图片中的两个储油罐为例,使用 CIoU-NMS,得到置信度最高的检测结果。

2.3.3 CIoU-NMS 在融合过程中的使用

在遥感影像识别过程中,需将全部 416×416 尺寸的小图检测到的边界框融合到原始图像上,如图 7 所示,

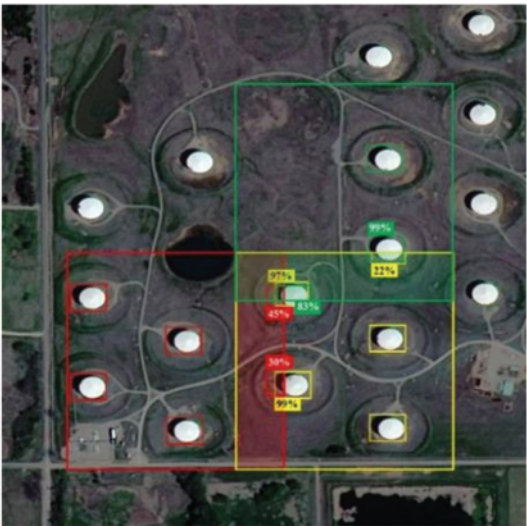


图7 CIoU-NMS 在融合过程中的应用

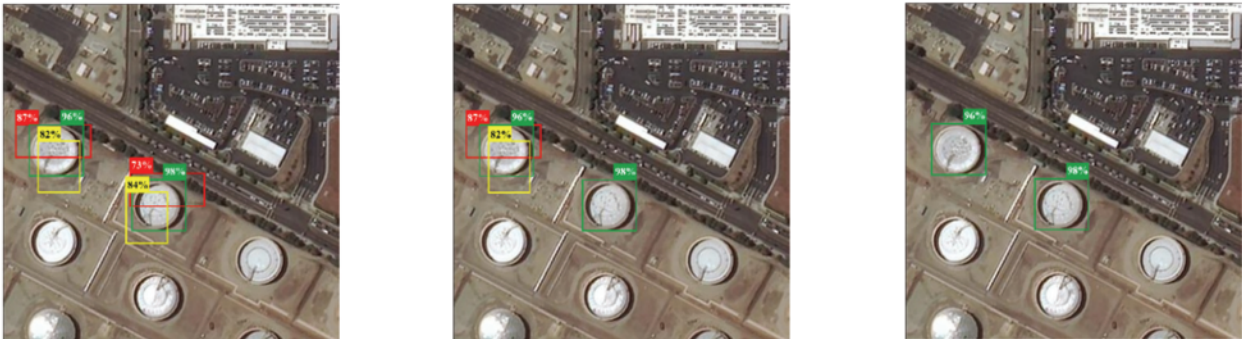


图6 CIoU-NMS 在预测过程中的应用

尺寸较大的矩形框代表三张相邻的 416×416 尺寸的检测图像,尺寸较小的多个矩形框分别代表每张检测图像上的储油罐检测结果。在 20% 尺寸的重叠区域中,三个储油罐分别被三张检测图像检测了 7 次,生成 7 个预测框,此时需要使用 CIoU-NMS,保留置信度最高的三个边界框,得到最终的识别结果。

2.4 高分辨率大尺寸遥感影像的处理

针对储油罐检测的过程中无法直接把高分辨率大尺寸遥感影像输入网络进行识别的问题,本文基于 SIM-RDWN^[26]项目编写识别程序,对遥感影像进行处理,步骤如下:

(1) 程序读取原始尺寸的遥感影像。

(2) 程序使用 scikit-image^[27]对原始影像通过滑动窗口的方式进行切割。如图 8 所示,为了不遗漏待检测对象,相邻图片带有 20% 尺寸的重叠部分,最终将原始图像生成多幅尺寸为 416×416 的图像。

(3) 将全部图像输入 MSA-YOLO 网络进行目标识别,输出每一幅图像对应的储油罐识别结果文本文件。

(4) 将全部识别结果文本文件作为输入,执行坐标转换函数,输出为每个储油罐在原始图像大小的真实坐标值 (x, y, w, h) 。

(5) 将全部储油罐的坐标值使用 OpenCV-Python 的 cv2.imwrite() 方法绘制到原始图像上,识别程序结束。

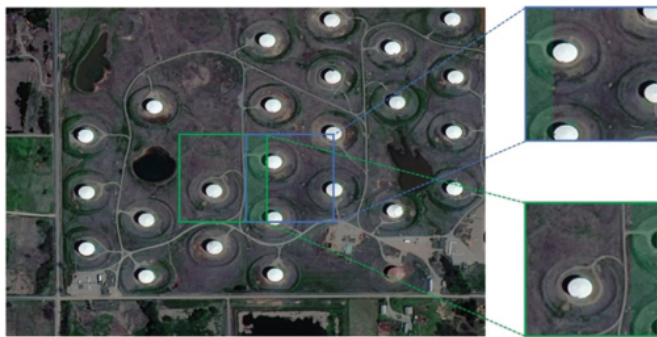


图 8 通过滑动窗口进行图像切割

通过该程序的处理,解决了深度学习算法对高分辨率大尺寸遥感影像的识别问题,具有现实意义。

3 数据集

3.1 MSA-YOLO 数据集

在 MSA-YOLO 数据集的制作中,主要使用了吉林一号光学遥感影像数据,同时也包括了开源遥感数据集,如 NWPU VHR-10 数据集^[28-30]、DOTA 数据集^[31-32]、RSOD 数据集^[33-34]、AID 数据集^[35]等。数据集中共包含 2 025 幅原始遥感图像,通过顺时针旋转 60° 、 90° 、 180° 和 270° 扩充至 10 125 幅,其中 20% (2 025 幅) 作为测试集,20% (2 025 幅) 作为验证集,其余 60% (6 075 幅) 作为训练集,共包含 60 760 个固定顶储油罐和 137 320 个外浮顶储

油罐,共计 198 080 个储油罐。

3.2 高分辨率大尺寸遥感影像数据集

高分辨率大尺寸遥感影像数据集包含从 2020 年 8 月至 2021 年 6 月吉林一号卫星星座系列拍摄的美国库欣区 156 幅遥感影像。单幅影像尺寸均在 $10\,000 \times 10\,000$ 像素以上。该数据集均作为测试集,用以验证 MSA-YOLO 模型在高分辨率大尺寸遥感影像上的识别效果。

4 实验配置与结果分析

4.1 实验配置与评价标准

本研究实验配置的 CPU 为 Intel Xeon, GPU 为 Tesla V100 SXM2, 内存为 26 GB, 操作系统为 Ubuntu 18.04。优化算法训练参数设置如表 4 所示。

表 4 MSA-YOLO 算法训练参数设置

参数名称	参数值
Batch size	16
Learning rate	2.61×10^{-3}
Epoches	300
Momentum	0.949
Input size	416×416
Classes	2

为了验证模型的性能,本文采用精确率、召回率、mAP@.5 和 F1-score 四个技术指标和模型大小、推理速度和训练速度三个性能指标进行评价,如式(7)~式(10)所示。

$$\text{Precision} = \frac{TP}{TP + FP} \quad (7)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (8)$$

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N \text{AP}_i \quad (9)$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

其中, TP、FP 和 FN 分别代表真阳性(True Positive, TP)、假阳性(False Positive, FP)和假阴性(False Negative, FN); N 代表检测类别的数量,由于检测类别分别为固定顶储油罐(类别序号为 0, 标签为“Fixed_Roof_Tanks”)和外浮顶储油罐(类别序号为 1, 标签为“Floating_Roof_Tanks”),因此在本文中 $N=2$ 。Model-size 即为模型生成的最优权重文件,主要存放训练过程中需要学习的权重和系数,以及包括训练次数、训练超参数等在内的其他参数信息,一般来说,在同样的网络结构下,Model-size 与模型复杂度成正比,Model-size 越大,模型越复杂。

4.2 实验结果与分析

4.2.1 MSA-YOLO

为了验证 MSA-YOLO 算法的有效性,对 MSA-YOLO 数据集中 2 025 幅测试图像中的储油罐进行了测试,结果如表 5 所示。MSA-YOLO 算法的 Precision 和 Recall 分

表 5 MSA-YOLO 与 YOLOv4 算法检测结果

算法	Precision/%	Recall/%	mAP@.5/%	F1-score/%	Model-size/MB	Detection-speed/(F/s)	Training-speed/(s/epoch)
YOLOv4	90.35	93.40	95.75	91.85	404	404.41	53.10
MSA-YOLO	89.25	95.20	95.65	92.13	153	420.17	47.00

别为 89.25%和 95.20% ,mAP@.5 为 95.65% ,F1-score 为 92.13%,Model-size 为 153 MB,Detection-speed 为 420.17 F/s, Training-speed 为 47 s/epoch。与 YOLOv4 相比 ,mAP@.5 低 0.1% ,F1-score 高 0.28% ,指标基本持平 ;而 Model-size 减少了 62.13% ,Detection-speed 提升 15.76 F/s, Training-speed 减少了 6.1 s/epoch。实验结果表明 ,与 YOLOv4 算法相比 ,经过剪枝操作的 MSA-YOLO 算法在保证了准确率的前提下 ,模型的推理速度和训练速度更快 ,模型更小。

为了验证 MSA-YOLO 方法的有效性 ,本研究比较了四种目标检测算法 ,包括 YOLOv3、YOLOv4、YOLOv5 和 EfficientDet-D0^[36]。在相同环境和数据集的情况下 ,五种算法的测试结果如表 6 所示。

测试结果表明 ,这五种检测算法的 mAP@.5 分别为 94.10%、95.75%、94.25%、81.50%和 95.65% ;Model-size 分别为 117 MB、404 MB、170 MB、15 MB和 153 MB;Detection-speed 分别为 434.78 F/s、404.41 F/s、294.12 F/s、268.72 F/s和 420.17 F/s;Training-speed 分别为58.9 s/epoch、53.1 s/epoch、68.9 s/epoch、184.8 s/epoch和 47.0 s/epoch。

从测试结果分析中可以看出 ,MSA-YOLO 算法的检测精度较高 ,mAP@.5 仅较 YOLOv4 低 0.10% ,比其他三种算法分别高 1.55%、1.45%和 14.15% ;F1-score 仅低于 YOLOv5 算法 0.77% ,比其他三种算法分别高 1.56%、0.28%和 10.68%。在 Detection-speed 方面 ,MSA-YOLO 算法仅慢于 YOLOv3 算法 14.61 F/s ,比其他三种算法分别快 15.76 F/s、126.05 F/s 和 151.45 F/s ;在 Training-speed

方面 ,MSA-YOLO 算法比其他四种算法分别快 11.90 s/epoch、6.10 s/epoch、21.90 s/epoch 和 137.80 s/epoch ;在 Model-size 方面 ,MSA-YOLO 算法的模型比 YOLOv3 算法的模型大 36 MB ,比 EfficientDet-D0 算法的模型大 138 MB ,小于其他两种算法模型251 MB 和 17 MB。但 EfficientDet-D0 算法比 MSA-YOLO 算法 mAP@.5 低 14.15% ,Detection-speed 慢 151.45 F/s ;YOLOv3 虽然 Detection-speed 仍比 MSA-YOLO 算法快 13.61 F/s ,但 mAP@.5 比 MSA-YOLO 算法低 1.45% ,从成本效益的综合角度来看 ,MSA-YOLO 算法仍更为高效。

结合实验数据 ,可得到结论 :

(1)MSA-YOLO 算法在与其它四种算法的比较中 ,mAP@.5 与 YOLOv4 和 YOLOv5 算法基本持平 ,而在 Model-size、Detection-speed 和 Training-speed 上均占优 ;

(2)YOLOv3 与 MSA-YOLO 算法相比 ,虽然 Model-size 和 Detection-speed 均占优 ,但 mAP@.5 和 Training-speed 差距较大 ;

(3)MSA-YOLO 算法各项指标均优于 EfficientDet-D0。

综上所述 ,从成本效益的综合角度来看 ,MSA-YOLO 算法更为高效。本文提出了一种基于 YOLOv4 算法的其它算法进行剪枝算法的原因是其原本的 mAP@.5 低于 MSA-YOLO 算法 ,而剪枝算法通常不能提高模型的准确性^[37]。五种算法的检测结果显示如图 9 所示。

本文提出的 MSA-YOLO 算法在保证了检测精度的前提下 ,改善了模型参数过多、体积过大的问题 ,同时大幅缩短了模型训练时间 ,提升了检测速度 ,具有一定的

表 6 不同的目标检测算法检测结果

算法	Precision/%	Recall/%	mAP@.5/%	F1-score/%	Model-size/MB	Detection-speed/(F/s)	Training-speed/(s/epoch)
YOLOv3	87.65	93.70	94.10	90.57	117	434.78	58.90
YOLOv4	90.35	93.40	95.75	91.85	404	404.41	53.10
YOLOv5	92.80	93.00	94.25	92.90	170	294.12	68.90
EfficientDet-D0	81.50	81.40	81.50	81.45	15	268.72	184.80
MSA-YOLO	89.25	95.20	95.65	92.13	153	420.17	47.00



图 9 5 种算法的检测结果

现实意义。

4.2.2 自适应锚框检测结果

如图 10 所示,以训练过程中的前 200 批次为例,使用默认锚框与使用自适应锚框的两次模型训练的平均损失均从初始的 1 902 开始减少,在 100 批次附近基本达到一致。在 1 至 100 批次的训练过程中,使用自适应锚框的模型训练,其平均损失收敛速度更快,更有助于

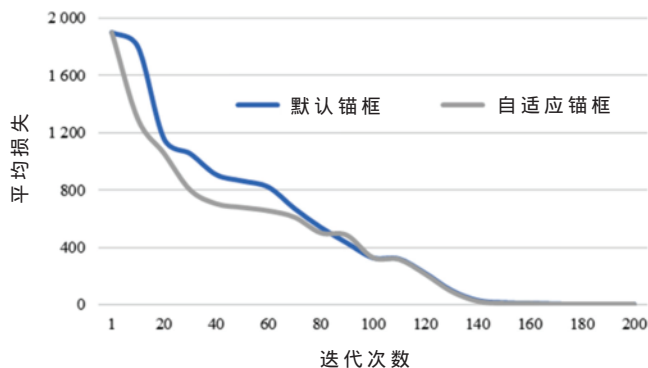


图 10 默认锚框与自适应锚框的目标检测结果

模型的加速收敛。

4.2.3 使用 CIoU-NMS 对 MSA-YOLO 进行预测框优化结果

在实验中,通过设置 nms_kind 参数,将 NMS 方式从 GreedyNMS 调整为 CIoU-NMS,将两种配置分别进行训练,实验结果如表 7 所示,使用 CIoU-NMS 的 Precision 和 Recall 分别为 91.35%和 94.40%,比使用 IoU 方式的 Greedy-NMS 算法均提高 1%,同时 mAP@.5 提升 0.42%,F1-score 提升 1%。实验结果表明,与基于 IoU 的 GreedyNMS 相比,基于 CIoU 的 NMS 具有更高的精度,有利于生成效果更好

表 7 使用 IoU 与使用 CIoU 的 NMS 算法检测结果

(%)				
算法	Precision	Recall	mAP@.5	F1-score
GreedyNMS	90.35	93.40	95.75	91.85
CIoU-NMS	91.35	94.40	96.17	92.85

的模型。

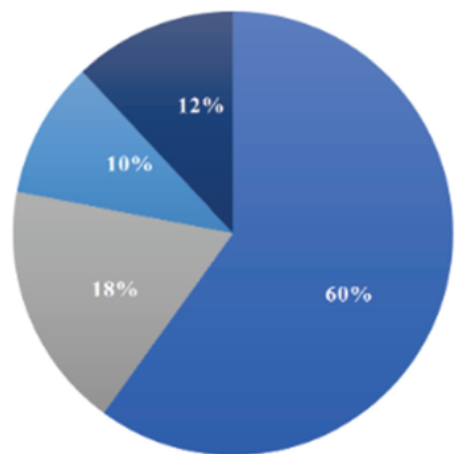
4.2.4 高分辨率大尺寸遥感影像的目标检测

本文对高分辨率大尺寸遥感影像数据集进行检测与统计,生成检测结果图片,如图 11 所示。图 11 为吉林一号高分 03 号卫星拍摄的具有 0.75 m 分辨率的美国库欣区遥感影像(局部),在该图中共有储油罐 349 个,用蓝色边框表示;其中 TP 为 344 个,用绿色边框表示;FP 为 4 个,用红色边框表示;FN 为 1 个,用黄色边框表示。该图 Precision 为 98.85%,Recall 为 99.71%。

对于全部 156 幅测试影像的总体识别结果如图 12 所示,其中满足 Precision $\geq 95\%$ 的图片占 60.31%;满足 $95\% > \text{Precision} \geq 90\%$ 的图片占 17.56%; 满足 $90\% > \text{Pre-}$

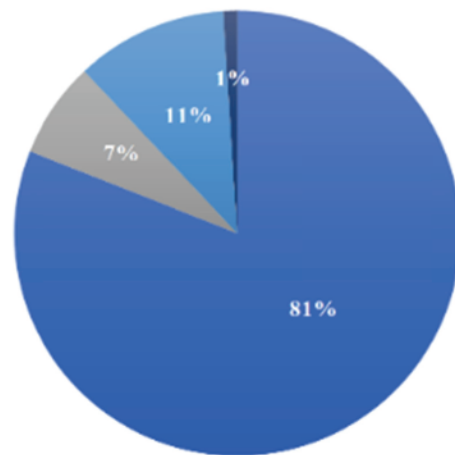


图 11 高分辨率大尺寸遥感影像的目标检测结果



■ 95% < Precision < 100% ■ 90% < Precision < 95%
■ 80% < Precision < 90% ■ 0% < Precision < 80%

(a) Precision 检测结果



■ 95% < Recall < 100% ■ 90% < Recall < 95%
■ 80% < Recall < 90% ■ 0% < Recall < 80%

(b) Recall 检测结果

图 12 高分辨率大尺寸遥感影像的目标检测结果统计

cision $\geq 80\%$ 的图片占 9.92%；满足 Precision $\leq 80\%$ 的图片占 12.21%。满足 Recall $\geq 95\%$ 的图片占 81.34%；满足 $95\% > \text{Recall} \geq 90\%$ 的图片占 7.46%；满足 $90\% > \text{Recall} \geq 80\%$ 的图片占 10.45%；满足 Recall $\leq 80\%$ 的图片占 0.75%。识别每幅影像的平均耗时为 59.05 s，对每幅影像的预处理及后处理平均耗时为 17.37 s，每幅影像平均总耗时为 76.42 s。结合识别准确率与识别时间，该算法具有可行性。

5 结论

为了实现对美国库欣区遥感影像中储油罐的准确、实时地智能检测，本研究提出了基于优化 YOLOv4 算法的储油罐检测方法 MSA-YOLO 算法，并对其进行实验。主要结论如下：

(1) MSA-YOLO 算法通过对监控区域的环境调研，在保证模型检测精度的前提下，将目标识别模型的多尺度检测 Head 层进行修剪，仅保留符合储油罐真实尺寸的输出尺度，同时使用 k-means++ 算法对锚框进行优化，生成自适应锚框；同时使用 CIoU 进行 NMS 优化。经实验表明，MSA-YOLO 算法的平均精度为 95.65%，检测速度可达 420.17 F/s。

(2) 在对比实验中，将 MSA-YOLO 算法与当前主流目标识别算法进行对比，结果表明，MSA-YOLO 算法具有相对最高的准确率，最快的训练速度和检测速度，以及较小的模型尺寸，在能准确检测储油罐的前提下，可以满足实时性的要求。

(3) 在对高分辨率大尺寸遥感影像的检测中，将遥感影像经过预处理后进入 MSA-YOLO 识别网络，得到识别结果后，经过后处理生成检测结果，此过程准确率高，识别时间短，可为基于高分辨率大尺寸的遥感卫星影像中储油罐检测提供技术参考。

在快速准确地识别美国库欣区遥感影像中储油罐的前提下，进而通过区域内原油储量的计算和回归分析，可以挖掘出遥感数据在能源期货领域的巨大的应用价值，更早做出合理投资决策，为交易者和策略制定者提供信息优势。

参考文献

- [1] KUSHWAHA N K, CHAUDHURI D, SINGH M P. Automatic bright circular type oil tank detection using remote sensing images[J]. Defence Science Journal, 2013, 63(3): 298-304.
- [2] JING M, ZHAO D, ZHOU M, et al. Unsupervised oil tank detection by shape-guide saliency model[J]. IEEE Geoscience and Remote Sensing Letters, 2018, 16(3): 477-481.
- [3] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Advances in Neural Information Processing Systems, 2012, 25: 1097-1105.
- [4] ZHANG L, SHI Z, WU J. A hierarchical oil tank detector with deep surrounding features for high-resolution optical satellite imagery[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2017, 8(10): 1-15.
- [5] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [6] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [7] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]//Advances in Neural Information Processing Systems, 2015: 91-99.
- [8] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.
- [9] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271.
- [10] JOSEPH R, FARHADI A. YOLOv3: an incremental improvement[J]. arXiv preprint arXiv: 1804.02767, 2018.
- [11] 王均成, 贺超, 赵志源, 等. 基于 YOLOv3-tiny 的视频监控目标检测算法[J]. 电子技术应用, 2022, 48(7): 30-33, 39.
- [12] 王林, 黄三丽. YOLOv3 网络在车标检测中的应用[J]. 电子技术应用, 2020, 46(9): 34-37, 42.
- [13] 张富凯, 杨峰, 李策. 基于改进 YOLOv3 的快速车辆检测方法[J]. 计算机工程与应用, 2019, 55(2): 12-20.
- [14] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[J]. arXiv preprint arXiv: 2004.10934, 2020.
- [15] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [16] Liu Shu, Qi Lu, Qin Haifang, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.
- [17] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context[C]//European Conference on Computer Vision. Springer, Cham, 2014: 740-755.
- [18] Zhou Dingfu, Fang Jin, Song Xibin, et al. IoU loss for 2d/3d object detection[C]//2019 International Conference on 3D Vision(3DV), IEEE, 2019.
- [19] Liu Songtao, Di Huang, Wang Yunhong. Adaptive NMS: refining pedestrian detection in a crowd[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1050-1059.

(下转第 40 页)

- [16] 王宗巍,杨玉超,蔡一茂,等.面向神经形态计算的智能芯片与器件技术[J].中国科学基金,2019(6):656-662.
- [17] 鄢贵海.DPU:以数据为中心的专用处理器[J].中国计算机学会通讯,2021,17(10).
- [18] 辛思达.面向粗粒度动态可重构处理器的通用领域算法实现与优化[D].长沙:国防科学技术大学,2015.
- [19] 魏少军,李兆石,朱建峰,等.可重构计算:软件可定义的计算引擎[J].中国科学(信息科学),2020,50(9):1407-1426.

- [20] 李锟,曹荣荣,孙毅,等.基于忆阻器的感存算一体技术研究进展[J].微纳电子与智能制造,2019(4):87-102.

(收稿日期:2022-03-10)

作者简介:

朱晶(1984-),女,硕士,高级经济师,主要研究方向:集成电路、新一代信息技术、技术经济学。



扫码下载电子文档

(上接第32页)

Recognition, 2019.

- [20] STANDARD A P I. Welded tanks for oil storage[S]. 2013.
- [21] He Yihui, Zhang Xiangyu, Sun Jian. Channel pruning for accelerating very deep neural networks[C]//Proceedings of the IEEE International Conference on Computer Vision, 2017.
- [22] ARTHUR D, VASSILVITSKII S. K-means++: the advantages of careful seeding[C]//SODA'07, 2006.
- [23] KRISHNA K, MURTY M N. Genetic K-means algorithm[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 1999, 29(3): 433-439.
- [24] REZATOFIGHI H, TSOI N, GWAK J, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019.
- [25] Zheng Zhaohui, Wang Ping, Liu Wei, et al. Distance-IoU loss: faster and better learning for bounding box regression[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2020.
- [26] VAN ET TEN A. Satellite imagery multiscale rapid detection with windowed networks[C]//2019 IEEE Winter Conference on Applications of Computer Vision(WACV), IEEE, 2019.
- [27] GARY B, KAEHLER A. Learning OpenCV: computer vision with the OpenCV library[M]. O'Reilly Media, 2008.
- [28] CHENG G, HAN J, ZHOU P, et al. Multi-class geospatial object detection and geographic image classification based on collection of part detectors[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2014, 98(dec.): 119-132.
- [29] CHENG G, HAN J. A survey on object detection in optical remote sensing images[J]. ISPRS Journal of Photogrammetry & Remote Sensing, 2016, 117: 11-28.
- [30] CHENG G, ZHOU P, HAN J. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(12): 7405-7415.
- [31] Xia Guisong, Bai Xiang, Ding Jian, et al. DOTA: a large-

scale dataset for object detection in aerial images[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.

- [32] Xia Guisong, Bai Xiang, Ding Jian, et al. DOTA: a large-scale dataset for object detection in aerial images[C]//The IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2018.
- [33] LONG Y, GONG Y, XIAO Z, et al. Accurate object localization in remote sensing images based on convolutional neural networks[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017(5): 1-13.
- [34] XIAO Z, LIU Q, TANG G, et al. Elliptic fourier transformation-based histograms of oriented gradients for rotation-invariant object detection in remote-sensing images[J]. International Journal of Remote Sensing, 2015, 36(2): 618-644.
- [35] Xia Guisong, Hu Jingwen, Hu Fan, et al. AID: a benchmark data set for performance evaluation of aerial scene classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(7): 3965-3981.
- [36] Tan Mingxing, Pang Ruoming, LE Q V. Efficientdet: scalable and efficient object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020.
- [37] He Yihui, Zhang Xiangyu, Sun Jian. Channel pruning for accelerating very deep neural networks[C]//Proceedings of the IEEE International Conference on Computer Vision, 2017.

(收稿日期:2022-07-15)

作者简介:

李想(1992-),男,硕士研究生,工程师,主要研究方向:深度学习、图像处理。

特日根(1987-),男,博士研究生,副研究员,主要研究方向:深度学习、大数据分析。

赵宇恒(1992-),男,硕士研究生,实习研究员,主要研究方向:深度学习、图像处理。



扫码下载电子文档

版权声明

经作者授权，本论文版权和信息网络传播权归属于《电子技术应用》杂志，凡未经本刊书面同意任何机构、组织和个人不得擅自复印、汇编、翻译和进行信息网络传播。未经本刊书面同意，禁止一切互联网论文资源平台非法上传、收录本论文。

截至目前，本论文已经授权被中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普网）、DOAJ、美国《乌利希期刊指南》、JST 日本科技技术振兴机构数据库等数据库全文收录。

对于违反上述禁止行为并违法使用本论文的机构、组织和个人，本刊将采取一切必要法律行动来维护正当权益。

特此声明！

《电子技术应用》编辑部

中国电子信息产业集团有限公司第六研究所