

doi: 10.6046/zrzyyg.2023139

引用格式: 于森,荆虹波,王翔,等. 基于深度特征的多方向目标检测研究[J]. 自然资源遥感, 2024, 36(3): 267-271. (Yu M, Jing H B, Wang X, et al. Multi-directional target detection based on depth features[J]. Remote Sensing for Natural Resources, 2024, 36(3): 267-271.)

基于深度特征的多方向目标检测研究

于森¹, 荆虹波¹, 王翔², 李兴久¹

(1. 北京城建勘测设计研究院有限责任公司, 北京 100101; 2. 新兴华安智慧科技有限公司, 北京 100160)

摘要: 近年来目标检测成为计算机视觉技术的重要分支,广泛应用于医学、军事、城轨等领域,随着卫星和遥感技术的进步,其获取的图像蕴含着丰富的信息,因此对这些图像中目标自动检测和理解变得至关重要。但是遥感影像中目标方向随意、密集等,传统目标检测方法容易导致漏检错检,针对此问题,该文提出多卷积核特征组合自适应区域生成网络(multi-convolution kernel feature combination adaptive region proposal network, MFCARPN)算法进行多方向检测,该算法引入多个不同卷积核提取特征,可以根据目标的差异性自适应地学习每个卷积核特征的权重参数,得到和目标更加匹配的特征图,同时通过结合目标原始特征使分类回归模型参数可以依据目标之间的差异性动态变化,提高区域生成网络(region proposal network, RPN)自适应能力。实验表明其在DOTA标准数据集的平均精度均值(mean average precision, mAP)达到75.52%,相较于GV算法提高0.5个百分点,由此证明了该算法的有效性。

关键词: 遥感影像; 自适应; MFCARPN; 多方向检测

中图法分类号: TP 391.4 **文献标志码:** A **文章编号:** 2097-034X(2024)03-0267-05

0 引言

近年来计算机视觉发展迅速,其主要是为了识别和理解图像和视频中包含的人类需要的有效信息。目标检测是结合定位与分类技术,旨在发现图像中一切感兴趣的目标,确定它们的位置、大小、方向以及类别的一项计算机视觉基本任务。

目标检测在轨道交通、城市规划、病灶检测等领域发挥着非常重要的作用,例如地铁沿线环境复杂,违章施工等各类风险源快速准确识别,防止事故发生,保障地铁安全运营。卫星和遥感技术^[1]的进步使人们每天都能获取大量高清晰度图像,这些无约束场景中的图像蕴含着丰富的信息,但是远远超过了人工操作和处理的范畴。为了使这些图像在实际安全管理(例如施工监测^[2]、病害管理^[3]等)中发挥价值,对其中风险源的自动检测和理解变得至关重要。

随着深度神经网络^[4]的发展,目标检测已经取得了较明显的进步,根据适应的场景不同分为水平目标检测^[5-6]和多方向目标检测^[7-8]。相较于自然场景,遥感影像这类无约束场景中风险源方向随意、

尺度变化大^[9]、背景复杂、宽高比差异明显、密集等因素的存在使其成为具有挑战性的研究课题之一。水平目标检测使用标准矩形进行边框表示,对密集多方向目标进行检测会导致目标之间的区域发生重叠,不利于后续的非极大值抑制,造成漏检,降低了目标的召回率。同时水平边框不能够准确地提供该目标方位和尺度信息,不能够满足实际需求,为了解决无约束场景目标检测中存在的各种挑战,多方向目标检测算法逐渐成为研究热点。Jiang等^[7]提出了R²CNN多方向目标检测框架,该算法用2个坐标点以及高度来表示倾斜边框,但是边框形状被限制为矩形,难以表示不规则形状目标;Xia等^[8]为了推进地球观测和遥感中的目标检测研究而引入了DOTA1.0数据集以解决遥感稀缺性的问题,同时提出了FR-O算法,该算法最终对边框的4个角点坐标进行预测,能够表示不规则的目标边框,但是坐标顺序容易发生混淆,导致检测精度下降;Bai提出了Gliding Vertex^[10]算法,其利用水平边框作为约束,在预测水平边框的同时预测倾斜边框在水平边框各条边上的偏移,加强倾斜边框表示的准确性,该算法提出了一种新颖的定向边框表示方式,避免了角度预测和角点预测的弊端,但是其未针对无约束场景

收稿日期: 2023-05-17; 修订日期: 2023-10-13

第一作者: 于森(1984-),女,博士,高级工程师,主要围绕城市轨道交通行业精细化管理和运营开展信息化数字化产品研发工作。Email: yumiao4503210@126.com。

多方向目标的特点对特征提取网络做出适应性的改变;Pan 等^[11]提出了 Dynamic Refinement Network 多方向目标检测框架,该算法在特征提取以及边框预测方面提出巨大的改进,提高了网络检测能力,但是边框表示仍然采用角度方式,同样受矩形边框限制。

综上所述,相比于水平目标检测,利用多方向目标检测可以表示不规则的边框,进而更加准确地获取目标方向以及尺度等信息。但是在多方向目标检测领域,特征提取以及边框表示等都是决定检测精度的重要因素。目前存在的多方向目标检测算法有待进一步改进,检测精度仍然有很大的提升空间。基于此,本文通过对现有算法进行改进,提出多卷积核特征组合自适应区域生成网络(multi-convolution kernel feature combination adaptive region proposal network, MFCARPN)算法,该算法设计多卷积特征组合以及池化自适应网络(pooled adaptive network, PAN),致力于获取包含精确目标信息的深度特征图以及候选区域,提高区域生成网络(region proposal network, RPN)网络自适应能力,进而提高遥感影像中的多方向目标检测精度。

1 研究方法

1.1 多方向目标检测技术路线

基于遥感影像的多方向目标检测研究路线如图 1 所示,包括模型训练、AI 识别、综合管理业务应用等模块。

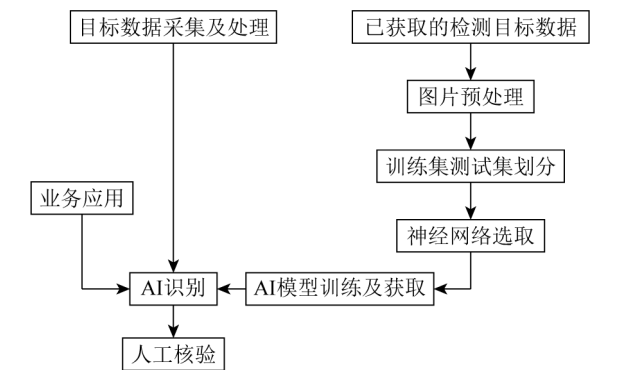


图 1 风险源识别研究路线

Fig. 1 Research route of risk source identification

具体步骤为:

- 1)通过收集遥感影像等方式获取图片并对其进行预处理,减少噪声等因素的影响。
- 2)按照一定的规则对数据集进行划分,形成训练集和测试集。
- 3)对比各类卷积神经网络的特点,分析风险源识别的特征,选择适合的神经网络进行特征提取,训练得到风险源识别模型。

4)利用训练好的模型对新采集的遥感影像进行识别,得到用户所需要的信息。

5)分析图片信息,将其与实际业务进行结合,进一步满足业务需求。

1.2 算法改进

本文详细分析遥感影像中目标的特点,重点关注该场景下目标角度随意、尺度变化大、背景复杂、纵横比差异明显等问题,研究了多种基于深度卷积神经网络的多方向目标检测算法,归纳总结了各类算法在网络结构、特征提取以及边框预测等方面的差异性、局限性、可行性。为了进一步解决遥感影像目标检测中存在的各类挑战因素,实现更加精确快速的检测效果,本文在已有的深度神经网络算法基础上针对特征提取、卷积方式、区域生成等方面提出改进,致力于获取包含精确目标信息的深度特征图以及候选区域。

为了能够使卷积核感受野以及检测模型参数随着目标的变化而变化,本文基于 Faster-RCNN^[12] 卷积神经网络进行改进,设计了 MFCARPN 算法,其网络结构如图 2 所示。

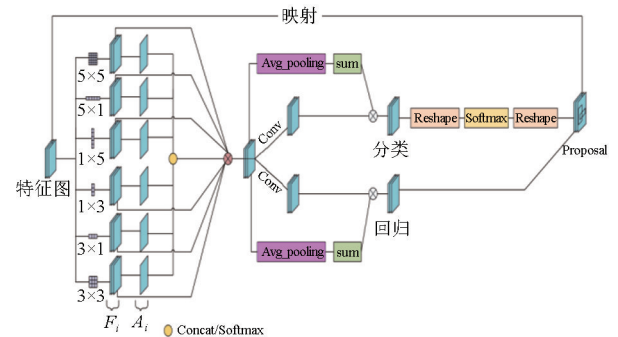


图 2 MFCARPN 网络结构

Fig. 2 MFCARPN network structure

1.2.1 多卷积核特征组合

在通用的卷积神经网络中常用于提取深度特征的卷积核大小为 3×3 ,固定大小的卷积核感受野固定,不能随着目标动态变化。为了能够减轻目标与神经感受野失配的问题,受到文献[6]的启发,引入了多卷积核特征组合模块,该模块可以根据目标的差异性自适应地学习每个卷积核特征的权重参数,聚集使用不同大小、形状卷积核提取的有效信息,得到和目标更加匹配的特征图从而进行后续预测。具体为,根据对轨道交通沿线遥感影像中目标的尺度差异性分析,将模块分别设置 6 个不同尺寸的卷积核,分别为 $3 \times 3, 1 \times 3, 3 \times 1, 1 \times 5, 5 \times 1, 5 \times 5$,每个卷积分支负责不同的感受野。首先将骨干网络特征图($W \times H \times C$)输入该模块,通过 6 个不同卷积分支得到包含不同信息的 6 个特征图 $F_i(i = 1, 2, \dots, 6)$,

大小仍然为 $W \times H \times C$ 。接下来使用 1×1 的卷积核对 F_i 操作,目的是为了对 F_i 进行降维,得到通道数为 1 的特征图 A_i 。最后在通道方向上使用 Softmax 函数连接 A_i ,获得各个分支的归一化权重参数,将 F_i 与其对应的权重参数相乘得到占比不同的各个分支特征,之后将分支特征进行融合,输出多卷积核组合特征图。为了保持网络输出相似性,该特征图没有进行通道扩展,和骨干网络特征图大小保持一致。多卷积核特征组合可以根据目标的差异性训练权重参数,选择适合目标形状的卷积核组合方式提取特征,例如,当目标的宽高比接近 $1:5$,则 1×5 卷积核分支权重占比大,这样就可以提取出与目标更加匹配的特征。同时卷积核的设置可以随着数据集的差异性而变化,当数据集中存在大纵横比目标时,可以增加更大比例卷积核分支,网络容易扩展。

1.2.2 PAN 网络

PAN 网络结构如图 3 所示。给定一个输入特征图,大小为 $W \times H \times C$,首先通过 Avg_pooling 得到全局特征值,其输出大小 $1 \times 1 \times C$,然后通过 sum 操作将每个通道特征叠加,得到包含有整张图片信息的特征值 F 。其次,输入特征图分别通过分类分支和边框回归分支,使用 1×1 的卷积算子处理得到 F_{cls} 和 F_{reg} ,大小分别为 $W \times H \times 18$ 和 $W \times H \times 36$ 。最后 F 分别和 F_{cls} 与 F_{reg} 执行 mul 操作,得到分类预测和边框回归预测结果。RPN 网络直接使用 F_{cls} 和 F_{reg} 作为输出,在测试过程中 1×1 卷积参数是提前训练好的,数值固定不变。本自适应预测结合了 F 值, F

值会随着输入图像的变化而变化,在该模块中充当动态参数,预测结果会随着该参数改变,从而适应目标的特殊性。同时池化操作相较于卷积操作不会引入额外的参数,计算量小,不会增加网络复杂度。

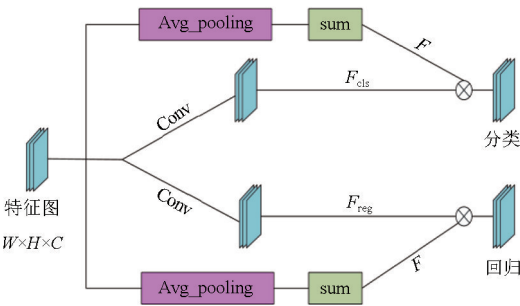


图 3 池化自适应网络
Fig.3 Pooled adaptive network

2 实验结果分析

为了证明 MFCARPN 方法的有效性,本文选用 DOTA 作为标准数据集,其收集了来自多种传感器和平台的不同分辨率航空影像多方向物体检测的数据,总计 2 806 张图像,其中 1 411 张为训练集,458 张为评估集,937 张为测试集。每张图片分辨率在 800×800 到 $4\,000 \times 4\,000$ 范围内,包含 15 类常见目标(飞机、舰船、储罐、棒球场、网球场、篮球场、田径场、海港、桥梁、大型车辆、小型车辆、直升机、环岛、足球场、游泳池),188 282 个不同尺度、方向和形状的实例,目标影像如表 1 所示。

表 1 DOTA 数据集目标影像示例
Tab.1 Target image of DOTA dataset

类别	大型车辆	游泳池	直升飞机	桥梁	飞机	田径场	小型车	海港
示例								
类别	棒球场	网球场	篮球场	足球场	储罐	舰船	环岛	
示例								

在保证实验条件相同的情况下,将 MFCARPN 和其他经典的多方向目标检测算法进行比较,其定量实验结果如表 2 所示。从表中数据对比可知,MFCARPN 网络在飞机、大型车辆、篮球场等类别的检测精度均高于其他算法,在 DOTA 数据集上平均精度均值 (mean average precision, mAP) 达到 75.52%,相较于 GV 算法 mAP 提高 0.5 个百分点。

该网络在 RPN 中引入多个不同卷积核提取特征,根据目标之间的差异性选择与其匹配的感受野,进而得到更加准确的深度特征,同时结合原始特征进行自适应分类预测,使得分类回归模型的参数可以结合目标原始特征,随着图片的变化而变化,能够自适应每个样本的独特性,提高检测网络的泛化能力。

表 2 定量实验结果

Tab. 2 Results of quantitative experiments (%)

类别	FR-O	R ² CNN	R-trans	GV	MFCARPN
飞机	79.42	80.94	88.64	89.64	90.57
棒球场	44.13	65.67	78.52	85.00	82.95
桥梁	17.70	35.34	43.44	52.26	52.20
田径场	64.05	67.44	75.92	77.34	68.04
小型车辆	35.30	59.92	68.81	73.01	61.91
大型车辆	38.02	50.91	73.60	73.14	78.98
游泳池	37.16	55.81	83.59	86.82	80.07
网球场	89.41	90.67	90.74	90.74	90.84
篮球场	69.64	66.92	77.27	79.02	88.70
储罐	59.28	72.39	81.46	86.81	87.91
足球场	50.30	55.06	58.39	59.55	59.86
环岛	52.91	52.23	53.54	70.91	71.83
海港	47.89	55.14	62.83	72.94	74.07
舰船	47.40	53.35	58.93	70.86	66.87
直升机	46.30	48.22	47.67	57.32	77.97
mAP	54.13	60.67	69.56	75.02	75.52

本文所提出的目标检测算法不仅适用于普遍物体的分类与定位,也具有良好普适性,可以根据实际需求制作数据集,同时进行参数调整,检测用户所关注的目标信息。例如在轨道交通领域,运营过程中风险源类型众多,研究人员可以通过需求调研确认风险源类型,然后采集遥感影像,按照要求进行风险源标注,同时调整算法中类别、图片分辨率等参数,最后进行算法训练,得到和识别目标匹配的模型,进而运用到实际研究领域。

3 总结与展望

本文提出了 MFCARPN 网络进行多方向目标检测,主要结论如下:

- 1)该网络通过设置多个不同尺度的卷积核进行特征提取,根据物体的形状调整感受野,选择与目标匹配的卷积分支,使得网络可以适用于不同大小的目标。
- 2)提出了池化自适应预测网络,可以结合目标的特征得到动态参数从而使得模型参数随着输入图片的变化而变化,结合目标的唯一性得到更加有针对性的分类、回归结果。
- 3)实验表明 MFCARPN 可以有效提高多方向目标检测精度,能够获得更加精确的目标位置、形状以及方向等信息,使得目标检测能够得到更加广泛地应用于轨道交通沿线风险源检测、病灶检测等领域。
- 实际需求中用户关注的目标类别多,遥感影像分辨率不同,导致其检测精度会随之发生变化,未来将针对不同识别对象所采用的影像精度和算法匹配进一步的量化研究。同时该算法网络层次加深导致

检测速度降低,之后需要着眼于速度这一性能,设计出既精确又快速的检测算法。

参考文献 (References):

[1] Tang T,Zhou S,Deng Z,et al. Arbitrary-oriented vehicle detection in aerial imagery with single convolutional neural networks[J]. Remote Sensing,2017,9(11):1170.

[2] 郭俊,宁志勇,王亚涛,等. 基于视频大数据的施工安全监测研究及应用[J]. 科技与创新,2023(6):153-159.

Guo J,Ning Z Y,Wang Y T,et al. Research and application of construction safety monitoring based on video big data[J]. Science and Technology & Innovation,2023(6):153-159.

[3] 江思阳. 基于计算机视觉的城市轨道交通弓网磨耗病害检测技术研究[D]. 北京:北京交通大学,2020.

Jiang S Y. Research on abrasion defect detection technology of urban rail transit pantograph and catenary based on computer vision [D]. Beijing:Beijing Jiaotong University,2020.

[4] Krizhevsky A,Sutskever I,Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communication of the ACM,2017,60(6):84-90.

[5] Redmon J,Farhadi A. YOLOv3: An incremental improvement [J/OL]. arXiv,2018(2018-04-08). <https://arxiv.org/abs/1804.02767>.

[6] Tian Z,Shen C,Chen H,et al. FCOS: Fully convolutional one-stage object detection[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE,2020:9627-9636.

[7] Jiang Y,Zhu X,Wang X,et al. R²CNN:Rotational region CNN for orientation robust scene text detection[J/OL]. arXiv,2017(2017-06-29). <https://arxiv.org/abs/1706.09579>.

[8] Xia G S,Bai X,Ding J,et al. DOTA: A large-scale dataset for object detection in aerial images[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE,2018:3974-3983.

[9] 王振华,谭智联,李静,等. Re-YOLOX:利用 Resizer 改进的 YOLOX 近岸海域监测目标识别模型[J]. 自然资源遥感,2023,35(3):10-16. doi:10.6046/zrzyyg.2022425.

Wang Z H,Tan Z L,Li J,et al. Re-YOLOX: A YOLOX model for identifying nearshore monitoring targets improved based on the Resizer model[J]. Remote Sensing for Natural Resources,2023,35(3):10-16. doi:10.6046/zrzyyg.2022425.

[10] Xu Y,Fu M,Wang Q,et al. Gliding vertex on the horizontal bounding box for multi-oriented object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2021,43(4):1452-1459.

[11] Pan X,Ren Y,Sheng K,et al. Dynamic refinement network for oriented and densely packed object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE,2020:11204-11213.

[12] 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,2017,39(6):1137-1149.

Multi-directional target detection based on depth features

YU Miao¹, JING Hongbo¹, WANG Xiang², LI Xingjiu¹

(1. *Beijing Urban Construction Survey, Design and Research Institute Co., Ltd., Beijing 100101, China;*

2. *Emerging Hua'an Smart Technology Co., Ltd., Beijing 100160, China*)

Abstract: In recent years, target detection, as an important branch of computer vision technology, has been widely applied in fields such as medicine, military affairs, and urban rail transit. As satellite and remote sensing technologies advance, images obtained using these technologies contain abundant information. This makes it crucial to conduct automatic target detection and understanding of these images. However, due to the random directions and dense distribution of targets in remote sensing images, conventional methods are prone to lead to missing or incorrect detection. In response, this study proposes a multi-convolution kernel feature combination-based adaptive region proposal network (MFCARPN) algorithm for multi-directional detection. This algorithm introduces multiple convolution kernel features for target extraction. The weight parameters of these convolution kernel features can be determined through adaptive learning according to the differences between the targets, yielding the characteristic patterns that match better with targets. Meanwhile, in combination with the original features of the targets, the parameters of the classification and regression model vary dynamically according to the difference between targets. Thus, the RPN's adaptive ability can be improved. The experimental results indicate that the mAP of the standard dataset DOTA reached up to 75.52%, which is 0.5 percentages higher than that of the baseline algorithm GV. Therefore, the MFCARPN algorithm proposed in this study proves effective.

Keywords: remote sensing image; adaptive ability; MFCARPN; multi-directional detection

(责任编辑: 陈 庆)