

计算机仿真

Computer Simulation

ISSN 1006-9348, CN 11-3724/TP

《计算机仿真》网络首发论文

题目：基于改进 YOLO V5s 模型的遥感图像目标检测及应用
作者：张晨光，滕桂法，丁文卿
网络首发日期：2023-04-19
引用格式：张晨光，滕桂法，丁文卿. 基于改进 YOLO V5s 模型的遥感图像目标检测及应用[J/OL]. 计算机仿真.
<https://kns.cnki.net/kcms/detail/11.3724.tp.20230418.1822.018.html>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

基于改进 YOLO V5s 模型的遥感图像目标检测及应用

张晨光^{1,2}, 滕桂法¹, 丁文卿³

- (1. 河北农业大学信息科学与技术学院, 河北 保定 071000;
2. 沧州交通学院计算机与信息技术学院, 河北 沧州 061100;
3. 河北农业大学渤海校区, 河北 沧州 061100)

摘要: 利用改进YOLO V5s模型实现遥感图像目标检测并用于地域贫困评估。针对现有模型提出了三点改进: 加强PAN结构、基于bounding box的RIOU_Loss回归损失函数、协同注意力机制。同时将遥感图像目标作为表征, 计算连续时间节点内的贫困率变化。实验结果表明, 改进模型的P、R、mAP@0.5、mAP@0.5:0.95 值存在不同程度的提升, 而Loss值有所下降。因此, 与原模型相比, 改进模型具备更精准的目标检测能力。同时, 与传统的统计数据方法相比, 改进模型为地域贫困评估提供了一种等效的无数据评估思路。

关键词: 遥感; 目标检测; YOLO V5s

中图分类号: TP391.9

文献标识码: B

Application of Improved YOLO V5s Model for Regional Poverty Assessment using Remote Sensing Image Target Detection

Zhang Chenguang^{1,2}, Teng Guifa¹, Ding Wenqing³

- (1. College of Information Science and Technology, Agricultural University of Hebei, Baoding Hebei 071000, China;
2. School of Computer and Information Technology, Cangzhou Jiaotong College, Cangzhou Hebei 061100, China;
3. Bohai College, Agricultural University of Hebei, Cangzhou Hebei 061100, China)

ABSTRACT: This study aims at applying the improved YOLO V5s model for the assessment of regional poverty using remote sensing image target detection. For this purpose, three improvements were made to the model. So, a new enhanced PAN structure was proposed. Accordingly, a new RIOU_Loss regression loss function of bounding box was proposed. Furthermore, a new collaborative attention mechanism was put forward. In addition, while objects in the remote sensing images were used as the representations of poverty status, the changes in the images were considered to evaluate the regional poverty rate in a continuous time interval. The results showed that the values of P, R, mAP@0.5 and mAP@0.5:0.95 of the model were improved, while the Loss value was decreased. Hence, compared with the original model, the application of the improved model led to the more accurate detection of the targets. In other words, being cheap, efficient, accurate and objective, as well as using no data. The proposed model can be considered as a promising approach in the assessment of regional poverty as it also provides comparable results with the traditional statistical methods.

KEYWORDS: Remote sensing; Target detection; YOLO V5s

1 引言

目标检测是高光谱遥感图像的重要应用, 属于高维数据的研究范畴^[1,2]。贫困数据的获取成本高, 耗时长, 极易受到人为因素的影响。将遥感图像目标作为贫困因素的表征进行检测, 能够以无数据方式客观地反映出地域经济的发展情况。

传统的检测算法普遍基于 Neyman-Pearson

规则, 在背景数据中将稀疏的目标区域进行二值分类^[3]。然而, 地表的采样距离通常大于目标的实际尺寸, 导致研究方法受限在亚像素级别, 给检测过程造成困难^[4,5]。同样, 诸如背景干扰、拍摄角度、噪声等也会成为不利因素^[6]。

特征提取是目标检测的关键环节^[7], 可以降低数据维度^[8], 并将数据关系映射到低维空间, 关键是实现从输入数据到特征向量之间的

基金项目: 沧州市科技局基于大数据推荐算法及遥感的市域经济发展研究规划 (213102003); 河北省省属高等学校基本科研业务费研究项目 (KY2021052); 河北省人力资源和社会保障厅面向就业市场的新工科建设中应用型大数据人才培养研究 (JRSHZ-2022-02037)

最佳转换^[9]。传统算法包括支持向量机、人工神经网络、小波变换及主成分分析等^[10-13]，然而，由于受到算法性能、计算能力及数据维度等因素的限制，部分技术已不能适应目前的实际需求^[14]。

近年来，深度学习在遥感图像目标检测领域发挥了重要作用，并拓展了相关知识^[15]。该技术基于卷积神经网络，其设计方式和性能都在不断完善^[16]。基于深度学习的目标检测包括基于候选区域和基于回归两种方式^[17]。前者又被称为“两阶段”方法，在提取候选区域的基础上进行目标检测，包括 R-CNN、Fast R-CNN、Faster R-CNN 及 Mask R-CNN^[18-21]。后者又被称为“一阶段”方法，去除了候选区域的提取过程，直接利用回归得到目标检测结果，包括 YOLO 及 SSD^[22,23]。

YOLO V5 是一阶段方法的最新版本，具备实时检测、准确度高及计算量小等优势，包括 S、M、L 和 X 四种结构，其中，S 结构具备最基本的功能，常被用作研究及改进对象。

目前，YOLO V5 模型已被公认为最优的检测工具之一，并被广泛应用于遥感图像目标检测过程。文献[24]针对合成孔径雷达图像中存在噪声的问题，提出了一种改进 N-YOLO 模型。首先利用分类组件 NLC 对噪声进行推导和分类，然后利用目标提取模块 STPAE 获取目标的完整区域，最后将两部分结果融合。通过与 CNN 算法对比证明了该模型性能更优，最终实现了对船舶目标的识别。文献[25]针对遥感图像中飞行器识别速度及精度不足的问题，提出了一种改进 YOLOV5-Aircraft 模型。首先在批处理过程中添加定心及缩放操作，增加了模型对特征的提取能力。然后将交叉熵损失函数替换为平滑 K-L 算法。最后以 CSandGlass 模块替代了残差模块，并去除了低分辨率特征层。通过与原模型对比证明了改进模型具备更优的速度、精度及收敛性。文献[26]提出了一种基于改进 YOLO V5s 模型的多尺度船舶检测方法。首先结合注意力机制模块 CAM 将改进残差模块 Res2Net 引入新建模块 CSPMRes2，使其在深度和尺度维度上同时具备特征提取能力。然后新建 FC-FPN 模块，为特征的最佳融合方案提供系数。最终实现了对船舶对象的多尺度检测。

与现有研究成果相比，本文的主要特点及贡献体现在以下几点：第一，数据来源广泛。以往的地域贫困评估普遍基于统计学方法或仅利用大数据技术，受人为因素的影响较大。本文利用遥感图像目标检测结果进行辅助验证，使判定过程更加客观。第二，模型改进全面。以往的改进方法普遍基于组件，形式相对单一，对目标检测能力的提升程度有限。本文的改进方式兼顾了模型结构、损失函数及组件设计，通过增加 PAN 过程的深度提升了模型的特征定位能力；将 CIOU 损失函数中 GT 框与 P 框的位置关系替换为二者的角度关系，提升了检测性能；引入注意力机制并适当改进，同时结合多通道卷积技术，增强了模型针对不同类型特征的提取能力。第三，检测结果详实。衡量 YOLO V5 模型性能的指标主要包括：P、R、mAP@0.5、mAP@0.5:0.95、Loss。目前，绝大多数的文章只公布了部分参数的检测结果，不能为读者提供全面的参考，而本文检测了全部参数，结果更详实。

本文利用改进 YOLO V5s 模型实现遥感图像目标检测并用于地域贫困评估。对原模型进行了三点改进。第一，提出了一种加强 PAN 结构，在保持 FPN 功能的同时增加了 PAN 过程的深度。第二，提出了一种 RIOU_Loss 回归损失函数，用于 bounding box 的检测判定。第三，提出了一种协同注意力机制，分别连接 Backbone 与 Neck，以及 Neck 与 Prediction。与原模型相比，改进模型可以更精准地检测目标。实验数据表明，P、R、mAP@0.5 及 mAP@0.5:0.95 的值在 COCO128 数据集上分别从 0.6852 提升至 0.7346、从 0.9729 提升至 0.9802、从 0.9664 提升至 0.9763、从 0.797 提升至 0.8541。而在自定义遥感图像数据集上，分别从 0.9217 提升至 0.9507、从 0.9753 提升至 0.9958、从 0.9795 提升至 0.9931、从 0.8433 提升至 0.8909。两个数据集上的 Loss 值分别从 0.0661 下降至 0.0645、从 0.0515 下降至 0.0396。与传统的统计数据方法相比，遥感图像目标检测可以作为一种无数据检测方法被用于地域贫困评估过程。

2 材料和方法

2.1 数据集准备

通过 BIGEMAP 软件获取遥感图像 842,304

张，分辨率为0.5米。选定区域并调整时间节点，可实现所需图像的下载，如图1所示。



图 1 遥感图像选择

遥感图像为JPG格式的谷歌瓦片，属于高光谱数据，相关属性如表1所示。

表 1 遥感图像的下载格式

等级	比例尺	分辨率	格式
18	1:1757	0.5m	*.jpg

LabelImg¹工具使用Python及Qt开发,用于遥感图像的标注，如图2所示。

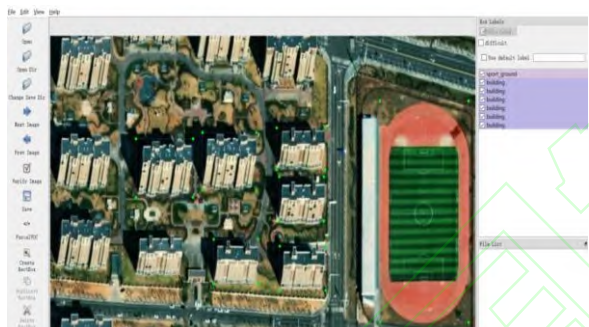


图 2 遥感图像标注

2.2 贫困因素及表征关系

将遥感图像中的目标作为地域经济发展因素的表征，通过对其进行分析来验证发展状态。贫困因素及表征关系如表2所示。

表 2 贫困因素及表征关系

贫困因素	表征
食品	大棚+耕种面积
住房	平房+楼房
交通	公路
教育	操场

2.3 YOLO V5s 环境配置

使用Anaconda3²软件作为代码的运行环境。在<https://pytorch.org>下载并配置Pytorch，安装指令为 `pip3 install Torch ==1.10.1+ CU113`

¹ Downloaded from <https://github.com/tzutalin/labelImg>
² Downloaded from <https://mirrors.tuna.tsinghua.edu.cn/anaconda/archive/>

TorchVision == 0.11.2+ CU113 TorchAudio ==0.10.1+ CU113 - F
https://download.pytorch.org/whl/cu113/torch_stable.html，官方代码下载地址为 <https://github.com/ultralytics/yolov5>。

YOLO V5s模型的结构如图3所示，包括输入端、Backbone、Neck及Prediction。与之前的版本相比，新增了自适应锚框计算、自适应图像缩放、改进Neck结构及SiLU激活函数等内容。

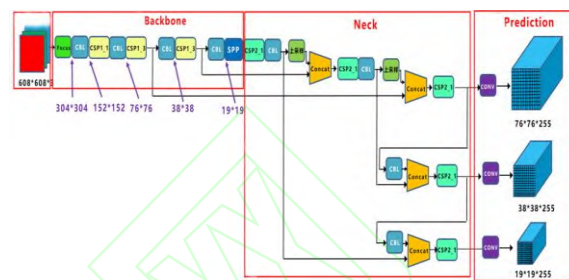


图 3 YOLO V5s 模型结构

3 YOLO V5s 模型改进

工程运行所需的硬件配置及主要参数如表3所示。

表 3 硬件配置及主要运行参数

指标项	值
GPU	NVIDIA GeForce GTX 1060
conf 阈值	0.5
iou 阈值	0.5

3.1 加强 PAN

原模型的PAN+FPN+PAN结构如表4所示。FPN为上采样过程，可以将特征提取结果逐层放大并用于语义识别。而PAN为下采样过程，可以使定位信息逐层精准。

表 4 YOLO V5s 模型结构

PAN	FPN	PAN
0:input:[32,128,128]		
1:CBS:[64,64,64]		
2:C3:[64,64,64]	17:C3:[128,32,32]	18:CBS:[128,16,16]
3:CBS:[128,32,32]	16:Concat	19:Concat
	4:[256,32,32]	14:[256,16,16]

续表 4 YOLO V5s 模型结构

PAN	FPN	PAN
4:C3:[128,32,32]	15:Up:[128,32,32]	20:C3:[256,16,16]
5:CBS:[256,16,16]	14:CBS:[128,16,16]	21:CBS:[256,8,8]
		22:Concat
6:C3:[256,16,16]	13:C3:[256,16,16]	10:[512,8,8]
7:CBS:[512,8,8]	12:Concat	23:C3:[512,8,8]

	6:[512,16,16]	
8:SPP:[512,8,8]	11:Up:[256,16,16]	24:Detect:(17,20,23)
9:C3:[512,8,8]	10:CBS:[256,8,8]	

和原模型相比，改进模型存在以下几点变化，如表5所示。

表 5 改进模型的加强 PAN 结构

PAN	FPN	PAN
0:input:[32,128,128]		
1:CBS:[64,64,64]		
2:C3:[64,64,64]		
3:CBS:[128,32,32]		
4:C3:[128,32,32]	15:Up:[128,128,128]	16:CBS:[64,64,64]
		17:Concat
5:CBS:[256,16,16]	14:CBS:[128,16,16]	2:[128,64,64]
6:C3:[256,16,16]	13:C3:[256,16,16]	18:CBS:[128,32,32]
	12:Concat	19:Concat
7:CBS:[512,8,8]	6:[512,16,16]	4:[256,32,32]
8:SPP:[512,8,8]	11:Up:[256,16,16]	20:C3:[128,32,32]
9:C3:[512,8,8]	10:CBS:[256,8,8]	21:CBS:[128,16,16]
		22:Concat
		14:[256,16,16]
		23:C3:[256,16,16]
		24:CBS:[256,8,8]
		25:Concat
		10:[512,8,8]
		26:C3:[512,8,8]
		27:Detect:(20,23,26)

第一，yolov5s.yaml文件的层数由24扩展到27。第二，缩短了FPN过程。同时，第15层中的图像尺寸由32提升至128，该层既是FPN过程的结束，也是第二个PAN过程的开始，从而增加了PAN过程的深度。第三，新增第16层至18层，并利用Concat操作在通道维度上将特征提取的中间结果与第2层结合。两个PAN过程都实现了从128到8的下采样。在保持FPN功能的基础上，模型的整体深度也因PAN过程的加深而得到扩展，不仅增强了特征提取能力，也提升了特征定位的准确性。

3.2 RIOUS_Loss 回归损失函数

目前常用的CIOU_Loss回归损失函数可由式（1）（2）（3）表示。

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w^p}{h^p} \right)^2 \quad (1)$$

$$\alpha = \frac{v}{1 - iou + v} \quad (2)$$

$$CIOU = 1 - \left(iou - \frac{D_{-2}^2}{D_{-C}^2} - v \cdot \alpha \right) \quad (3)$$

现有函数考虑到ground truth (GT)框与prediction (P)框的重叠面积、中心点距离、宽高比等因素，但忽略了二者的位置差异。以图4为例， P_1 和 P_2 内嵌于GT，三者具有相同的宽高比。

因此， $v = 0$ 在式（1）中恒成立，同时 $v \cdot \alpha = 0$

在式（3）中成立。当

$$iou_1 - \frac{D_{-2_1}^2}{D_{-C}^2} = iou_2 - \frac{D_{-2_2}^2}{D_{-C}^2} \quad (4)$$

成立时， P_1 和 P_2 的CIOU值相等且无法区分，然而由 P_1 和 P_2 所围成的面积并不相等。由于 P_1 距离GT中心点较远而 P_2 距离较近，因此 D_{-2_1} 大于 D_{-2_2} 。如果 iou_1 同时大于 iou_2 ，则存在一定概率使式（4）成立。

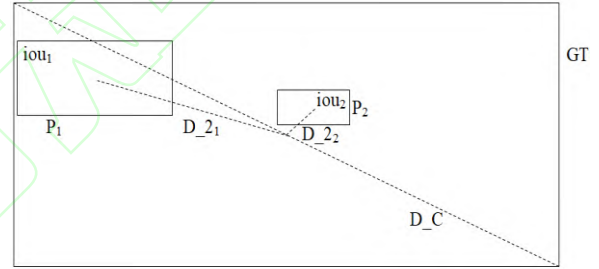


图 4 CIOU 损失函数失效的情况

如图5所示，O表示GT框的中心点，A表示GT框的左上点，B表示P框的中心点。以O为圆心，以 D_{-2} 为半径可以形成圆R，即为B的运动轨迹。也就是说，IOU、 D_{-C} 、 D_{-2} 的值都是固定的，但B出现的位置是随机的，且B与R重合。在 $\triangle OAB$ 中，OA与OB的长度是固定的，但AB的长度会随着B的移动而变化，同时各内角随之变化。根据正弦定理，当

$$BO / \sin \angle BAO = AO / \sin \angle ABO \quad (5)$$

成立，有

$$\sin \angle BAO / \sin \angle ABO = BO / AO \quad (6)$$

成立，则 GT 和 P 之间的位置关系可转换为角度关系。 $\angle BAO$ 反映了 B 在 R 上的移动情况，即 GT 与 P 的相对位置关系。B 可在 0 到 360

度之间变化，涵盖了每一种可能。然而存在一种特殊情况，当B沿着R移动到E（即B穿过R并延伸到E）， $\angle BAO$ 的值没有发生变化，此时 $\angle BOA$ 可被用于辅助性证明。参考式（5）和（6），当

$$AB / \sin \angle AOB = AO / \sin \angle ABO \quad (7)$$

成立，有

$$\sin \angle AOB / \sin \angle ABO = AB / AO \quad (8)$$

成立。BC和BD是B到GT的垂直距离，则公式

$$BC = |x_B - x_A| \quad (9)$$

$$BD = |y_B - y_A| \quad (10)$$

$$AB = \sqrt{BC^2 + BD^2} \quad (11)$$

成立。综上，RIOU_Loss回归损失函数可表示为

$$\begin{aligned} iou &= (BO / AO) \times (AB / AO) \\ &= iou - (4 \times D_2 \times AB / D_C^2) \end{aligned} \quad (12)$$

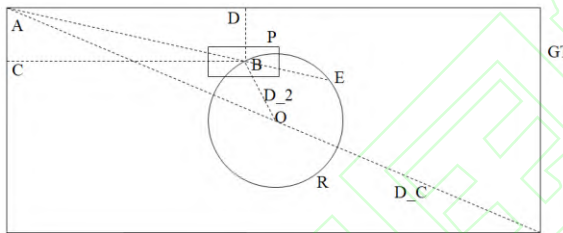


图5 RIOU_Loss回归损失函数示意图

3.3 协同注意力机制

如果将YOLO V5s模型与注意力机制进行简单结合，并不能有效提升检测结果的性能。表6列举了四种常用注意力机制^[27-30]与模型结合后的检测结果，与Step 2相比，各项指标（P, R, mAP@0.5 和mAP@0.5:0.95）的性能不仅没有增加，反而降低了。

表6 YOLO V5s 模型结合常用注意力机制

	Model Layers	Yaml Layers	Time/ Image(ms)	Weight
Step 2 (Baseline)	310	27	91.927	15.3M 7518909

续表6 YOLO V5s 模型结合常用注意力机制

	Model Layers	Yaml Layers	Time/ Image(ms)	Weight
+SE	316	28	105.525	15.4M 7551677

+CBAM	321	28	105.173	15.4M 7551775
+ECA	314	28	106.383	15.3M 7518912
+CA	320	28	105.891	15.4M 7544557

P	R	mAP @0.5	mAP @.5:.95	GFLOPS (B)
0.7143	0.98	0.9746	0.8425	23.9
0.6544	0.9714	0.9649	0.817	24
0.6691	0.9715	0.9655	0.8041	24
0.6672	0.9687	0.9631	0.8114	23.9
0.6611	0.9758	0.9671	0.8003	24

可能的原因有两个。第一，传统的注意力机制一般需要在通道维度或宽高维度上执行池化，导致处理结果的整体维度与原模型不符，造成数据丢失。第二，在数据增强过程中，Sigmoid函数一般被用于系数生成，该函数实现了从线性分布到非线性分布的转换。如图6所示，在 $[a, b]$ 区间内，转换效果优于最初的线性函数 $y = x + n$ ，然而在区间 $[-\infty, a]$ 及 $[b, +\infty]$ 内，转换结果无限接近于极值，不能很好地反映分布情况。实验数据显示，在训练过程的初期，检测结果会出现较大的波动。而且在后期，即使增加训练次数也无法使模型的检测性能得到持续提升，因为特征提取能力有限并已达到瓶颈。

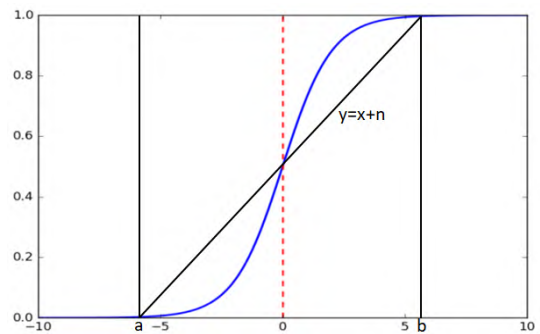


图6 Sigmoid函数与线性函数对比

新的协同注意力机制CAM由两部分组成。第一，提出了一种反向填充注意力机制组件RFAM。与传统的结合方式不同，该组件并不是直接插入到已有模型中，而是替代了第10层。也就是说，在Backbone与Neck的连接处，以RFAM组件替代了CBS组件。第二，为了提取不同的特征，在Step 2的基础上，多通道卷积组件MCC被分别添加到20、23、26层之后，用于连接Neck

与Prediction。同时，为了缓解下采样过程中的数据丢失，22及26层中的CBS组件被替换为膨胀卷积组件DC，如表7所示。

表 7 改进模型的协同注意力机制

PAN	FPN	PAN
0:input:[32,128,128]		
1:CBS:[64,64,64]		
2:C3:[64,64,64]		
3:CBS:[128,32,32]		
4:C3:[128,32,32]	15:Up:[128,128,128]	16:CBS:[64,64,64]
		17:Concat
5:CBS:[256,16,16]	14:CBS:[128,16,16]	2:[128,64,64]
6:C3:[256,16,16]	13:C3:[256,16,16]	18:CBS:[128,32,32]
	12:Concat	19:Concat
7:CBS:[512,8,8]	6:[512,16,16]	4:[256,32,32]
8:SPP:[512,8,8]	11:Up:[256,16,16]	20:C3:[128,32,32]
9:C3:[512,8,8]	10:RFAM:[256,8,8]	21:MCC:[128,32,32]
		22:DC:[128,16,16]
		23:Concat
		14:[256,16,16]
		24:C3:[256,16,16]
		25:MCC:[256,16,16]
		26:DC:[256,8,8]
		27:Concat
		10:[512,8,8]
		28:C3:[512,8,8]
		29:MCC:[512,8,8]
		30:Detect:(21,25,29)

RFAM组件位于Backbone与Neck的连接处，结构如图7所示。其中，AP表示均值池化（AdaptiveAvgPool2d）操作，Re表示复制（Repeat）操作，M表示通道维度上的均值（Mean）操作，Sub表示减法（Subtraction）操作。有3条序列被同时执行。第一条，将输入数据完全保留而不进行任何操作，类似于残差网络中的输入数据。第二条，顺次执行通道计算及像素计算。在通道计算过程中，首先利用AP操作将输入数据的尺寸由 $h \times w$ 转换为 1×1 ，然后利用Re操作将尺寸逆向填充并恢复为 $h \times w$ 。也就是说，在任意通道上，每个像素点的值都是相同的。最后，将该值作为衡量标准，与输入数据进行Sub计算，即当前结果与输入数据在每个像素点上的差异，该操作可以在特征提取过程中增加各目标之间的差异。为了进一步提取特征，CBS₁和CBS₂

模块被执行，同时实现了通道数量的压缩及解压缩。而在像素计算过程中，M操作的执行使通道数量由c转换为1，再利用Re操作逆向填充为c，并利用Sub操作计算与输入数据之间的差值。同样，为了进一步提取特征，卷积核为7的CBS₃模块被执行。第三条与第二条相似，只是交换了通道计算与像素计算的顺序。最后，将三条序列的结果合并，执行CBS模块。

RFAM组件从不同角度对输入数据进行通道计算及像素计算，与传统的注意力机制不同，用于系数生成的Sigmoid函数被替换为逆向填充操作，在缓解数据丢失的同时，也为特征提取提供了必要条件。

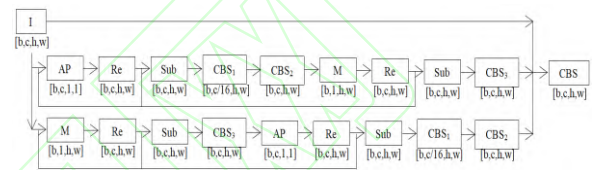


图7 RFAM组件结构

然而，目前的模型性能并没有达到最优。与表6中的检测结果相比，结合了RFAM组件的模型性能虽然优于其它几种注意力机制，但仍然低于Baseline的水平，如表8所示。

表 8 YOLO V5s 模型结合 RFAM 组件

	Model	Yaml	Time/	Weight
	Layers	Layers	Image(ms)	
+RFAM	329	27	155.264	41.1M
P	R	mAP @0.5	mAP @.5:.95	GFLOPS (B)
0.6894	0.9758	0.9692	0.8317	44.6

可能的原因是，将YOLO V5s模型与RFAM组件结合，虽然可以在训练过程的初期缩短收敛时间并提升检测性能，但模型的特征提取能力并没有得到整体提升。表9展示了RFAM组件与Step 2在不同时间节点上的性能对比。将检测结果按照每50个一组进行划分，共10组。由数据可知，RFAM在0到150次中保持了较快的增长趋势且各项性能都优于Step 2，但之后被逐步反超。

表 9 RFAM 与 Step 2 性能对比

	epochs	P	R	mAP @0.5	mAP @.5:.95
Step 2	50	0.0582	0.0324	0.0501	0.0165
	100	0.1928	0.2642	0.2665	0.1061

	150	0.2908	0.5241	0.4361	0.1618
	200	0.3331	0.7149	0.6565	0.3255
	250	0.3998	0.8072	0.7324	0.3916
	300	0.4362	0.854	0.8081	0.4338
	350	0.5026	0.8458	0.8307	0.4734
	400	0.4874	0.8957	0.8592	0.5358
	450	0.5018	0.9027	0.8799	0.562
	500	0.5155	0.9343	0.8892	0.5665
	50	0.1239	0.0839	0.1066	0.0343
	100	0.371	0.3392	0.3565	0.1582
	150	0.3859	0.543	0.5372	0.2721
	200	0.3629	0.6275	0.5978	0.3259
+RFAM	250	0.3442	0.6779	0.6769	0.3862
	300	0.424	0.732	0.74	0.4452
	350	0.4442	0.761	0.752	0.4638
	400	0.4871	0.8114	0.7881	0.486
	450	0.4973	0.8274	0.8222	0.5525
	500	0.5023	0.8495	0.8591	0.5654

RFAM组件是Backbone与Neck之间的过渡，在保持卷积操作的基础上，新增了注意力机制，从而改变了最初的特征提取效果。实验数据显示，利用不同的卷积核对特征进行拆分，可以更有效地提升模型的整体性能，而多通道卷积组件MCC能够满足要求，如图8所示。

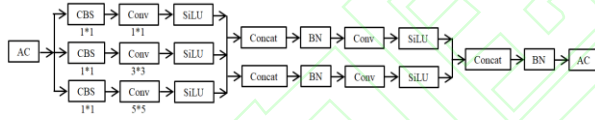


图8 MCC组件结构

该组件位于Neck与Prediction的连接处。首先利用非对称卷积组件ACNet将输入数据分为3条序列，执行 1×1 的CBS模块操作。随后分别执行 1×1 、 3×3 、 5×5 的Conv卷积及SiLU激活操作，得到不同的特征提取结果。利用Concat操作合并相邻的两条序列，并继续执行BatchNorm2d+Conv+SiLU操作，进一步提取合并结果的特征，此时的序列由3条减少为2条。最终合并序列并执行BatchNorm2d+ACNet操作。在此过程中，ACNet可以增强提取特征的鲁棒性。

为了进一步缓解下采样过程中的数据损失，将CBS模块替换为 $k=3, s=2, p=2, d=2$ 的膨胀卷积，式（13）（14）展示了卷积过程。

$$f = k + (k - 1)(d - 1) \quad (13)$$

$$out = (in - f + 2p + 1) / s \quad (14)$$

4 实验结果及论述

针对YOLO V5s模型进行了3点改进，检测结果及性能提升如下。COCO128数据集被用于检测阶段，自定义遥感图像数据集被用于验证阶段。

4.1 检测阶段

4.1.1 加强 PAN 结构

加强 PAN 结构与原模型的性能对比如表10所示。

表 10 加强 PAN 结构与原模型性能对比

	Model	Yaml	Time/	Weight
	Layers	Layers	Image(ms)	
Baseline	283	24	59.063	14.1M
				7276605
+EPAN	310	27	83.602	15.3M
				7518909
P	R	mAP	mAP	GFLOPS
		@0.5	@.5:.95	(B)
0.6852	0.9729	0.9664	0.797	17.1
0.7039	0.979	0.9716	0.8348	23.9

检测结果如图9所示，其中，Box表示预测框损失，Objectness表示目标检测损失，Classification表示分类损失，三者共同构成了模型损失值Loss。

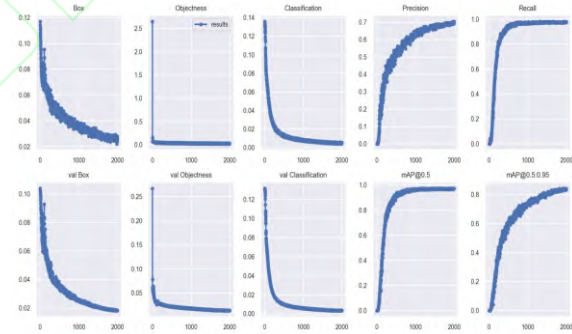


图9 加强 PAN 检测结果

准确率可表示为式（15），即正确分类为正的样本占全部正样本检测结果的比例。

$$precision = TP / (TP + FP) \quad (15)$$

召回率可表示为式（16），即正确分类为正的样本占全部正样本实际结果的比例。

$$recall = TP / (TP + FN) \quad (16)$$

mAP@0.5 和 mAP@0.5:0.95 表示均值准确率，由 Prediction 和 Recall 共同获得。表 11 展示了改进模型与原模型的损失值对比。

表 11 加强 PAN 与原模型的损失值对比

	Box	Objectness	Classification	Loss
Baseline	0.02941	0.03177	0.00495	0.06613
+EPAN	0.02651	0.02738	0.005304	0.05919

4.1.2 RIOU_Loss 回归损失函数

RIOU_Loss 损失函数与原模型的性能对比如表 12 所示。

表 12 RIOU_Loss 与原模型性能对比

	Model	Yaml	Time/	Weight
	Layers	Layers	Image(ms)	
Baseline	283	24	59.063	14.1M
				7276605
+EPAN	310	27	83.602	15.3M
				7518909
+RIOU_Loss	310	27	91.927	15.3M
				7518909
P	R	mAP	mAP	GFLOPS
		@0.5	@.5:.95	(B)
0.6852	0.9729	0.9664	0.797	17.1
0.7039	0.979	0.9716	0.8348	23.9
0.7143	0.98	0.9746	0.8425	23.9

检测结果如图 10 所示。

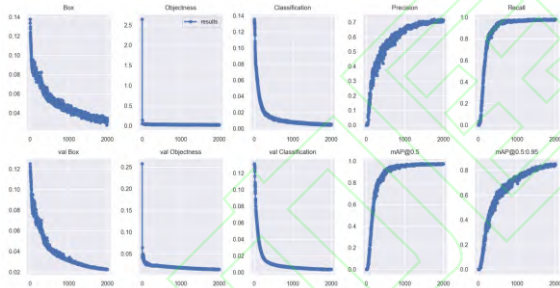


图 10 RIOU_Loss 检测结果
损失值如表 13 所示。

表 13 RIOU_Loss 与原模型的损失值对比

	Box	Objectness	Classification	Loss
Baseline	0.02941	0.03177	0.00495	0.06613
+EPAN	0.02651	0.02738	0.005304	0.05919
+RIOU_Loss	0.03305	0.02782	0.005284	0.06616

4.1.3 协同注意力机制

协同注意力机制与原模型的性能对比如表 14 所示。

表 14 协同注意力机制与原模型性能对比

	Model	Yaml	Time/	Weight
	Layers	Layers	Image(ms)	
Baseline	283	24	59.063	14.1M

				7276605
+EPAN	310	27	83.602	15.3M
				7518909
+RIOU_Loss	310	27	91.927	15.3M
				7518909
+CAM	461	30	218.756	76.7M
				38139645

P	R	mAP	mAP	GFLOPS
		@0.5	@.5:.95	(B)
0.6852	0.9729	0.9664	0.797	17.1
0.7039	0.979	0.9716	0.8348	23.9
0.7143	0.98	0.9746	0.8425	23.9
0.7346	0.9802	0.9763	0.8541	77.1

检测结果如图 11 所示。

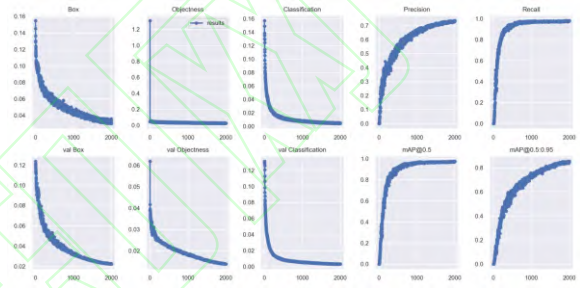


图 11 协同注意力机制检测结果
损失值如表 15 所示。

表 15 协同注意力与原模型的损失值对比

	Box	Objectness	Classification	Loss
Baseline	0.02941	0.03177	0.00495	0.06613
+EPAN	0.02651	0.02738	0.005304	0.05919
+RIOU_Loss	0.03305	0.02782	0.005284	0.06616
+CAM	0.03334	0.02618	0.004939	0.06445

4.2 验证阶段

改进模型与原模型在自定义遥感图像数据集上的性能对比如表 16 所示。

表 16 改进模型与原模型性能对比

	Model	Yaml	Time/	Weight
	Layers	Layers	Image(ms)	
Baseline	283	24	67.612	14.4M
				7085118
improved	461	30	209.845	76.3M
				37948158

续表 16 改进模型与原模型性能对比

P	R	mAP	mAP	GFLOPS
		@0.5	@.5:.95	(B)
0.9217	0.9753	0.9795	0.8433	16.5
0.9677	0.9958	0.9927	0.951	76.5

验证结果如图 12 所示。

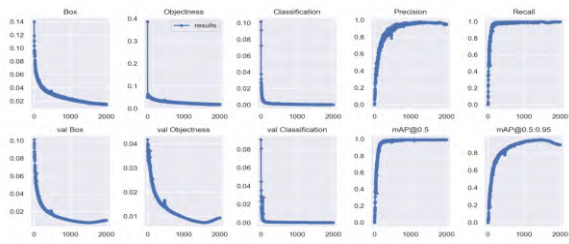


图 12 改进模型验证结果

损失值如表 17 所示。

表 17 改进模型与原模型的损失值对比

	Box	Objectness	Classification	Loss
Baseline	0.02008	0.03071	0.000738	0.05153
improved	0.01835	0.0209	0.000381	0.03963

4.3 检测数据

2014 年到 2019 年的表征目标检测结果如表 18 所示。

表 18 2014-2019 年表征目标检测结果

年份	高速	平房 1	平房 2	楼房
2014	4717	11569	3964	353
2015	8171	17948	4656	457
2016	9717	16463	5521	743
2017	15074	18241	5156	1122
2018	16812	19462	5087	1192
2019	16881	25890	4657	1988

操场	大棚	梯田 1	耕地	梯田 2
12	1148	1773	290	220
21	1377	2616	327	311
22	1659	3823	478	519
25	2565	4587	529	673
25	2827	4614	555	731
30	3080	5291	550	855

检测结果 95% 的回归分析如表 19 所示。

表 19(a) 回归分析结果

Covariance Items	Values
Multiple R	0.999676
R Square	0.999352
Adjusted R Square	0.998272
Standard Error	368.8552

表 19(b) 方差分析结果

Variance Items	df	SS
Regression Analysis	5	6.29E+08
Residual	3	408162.5
Total	8	6.3E+08
MS	F	Significance F

1.26E+08 925.0926 5.6E-05
136054.2

数据显示回归效果显著。表征目标 2014 年到 2019 年的变化情况如图 13 所示。

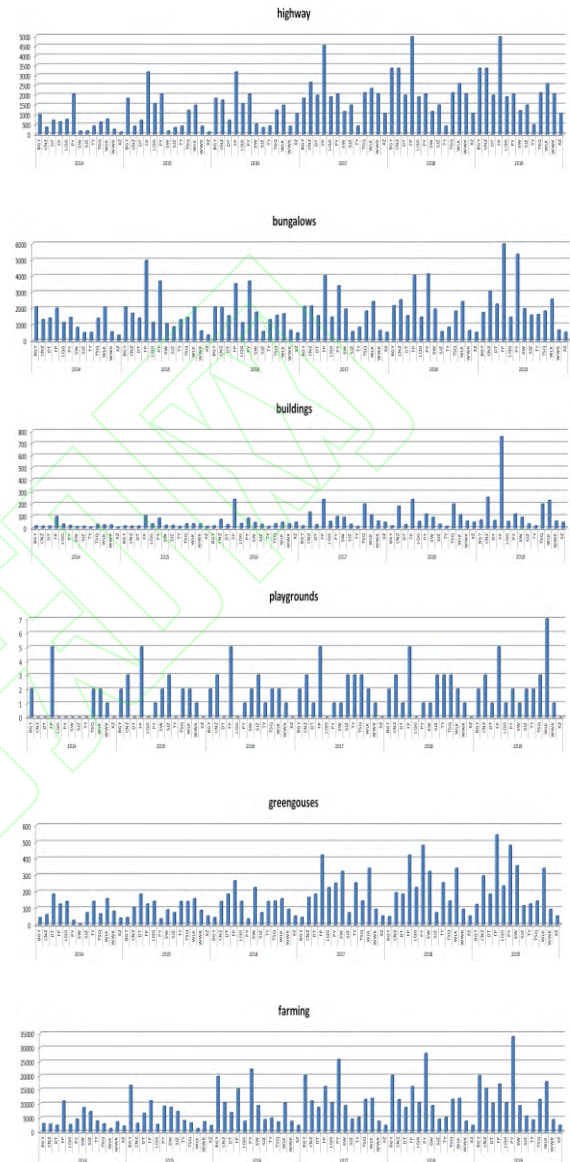


图 13 表征目标历年变化情况。依次为：高速、平房、楼房、操场、大棚、耕地

通过贫困率表示地域经济发展情况。如表 20 所示，结合现有数据，遥感图像目标检测结果可用于预测 2020 年的贫困率。由结果可知贫困率小于 0，则该地区已脱贫。

表 20 表征目标及贫困率计算结果

年份	高速	平房 1	平房 2	楼房	操场
2014	4717	11569	3964	353	12
2015	8171	17948	4656	457	21

2016	9717	16463	5521	743	22
2017	15074	18241	5156	1122	25
2018	16812	19462	5087	1192	25
2019	16881	25890	4657	1988	30
2020	21105	26055	5279	2052	32
大棚	梯田 1	耕地	梯田 2	贫困率	
1148	1773	290	220	0.5437	
1377	2616	327	311	0.3004	
1659	3823	478	519	0.186	
2565	4587	529	673	0.1378	
2827	4614	555	731	0.0693	
3080	5291	550	855	0.0045	
3601	6219	550	879	-0.1367	

以传统统计数据作为辅助验证，如表 21 所示。各项贫困因素在 2020 年的预测值都小于 0，验证了表征目标检测结果的正确性。

表 21 传统统计数据分析结果

年份	食品	教育	住房	交通
2014	2.7953	2.2208	2.8920	2.1978
2015	1.5799	1.2442	1.5229	1.2422
2016	1.1272	0.9437	1.2436	1.0093
2017	0.8692	0.7763	0.8194	0.8608
2018	0.4920	0.4587	0.4796	0.4720
2019	0.0346	0.0327	0.0321	0.0317
2020	-1.1821	-0.2223	-0.5615	-0.2696

4.4 目标检测效果

将改进模型（右）与原模型（左）的检测效果对比，如图 14 所示。可知改进模型对目标的识别能力更强，且标注范围更精准。

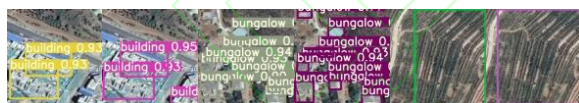


图 14 原模型与改进模型检测效果对比

图中每 2 张图片一组，分别为楼房、平房、梯田

4.5 同类型实验结果对比

为了验证本文改进模型的性能，将实验结果与同类型文章对比。文献[31]利用夜间遥感图像中的渔船作为自定义数据集，提出了一种 TASFF-YOLOv5 算法，添加了小目标识别层和自适应空间特征融合网络。文献[32]利用合成孔径雷达图像作为自定义数据集，将注意力机制引入 YOLO V5 模型，提出了一种轻量级端到端目标检测框架。文献[33]以航空飞行器遥

感图像为自定义数据集，通过改进 YOLO V5 模型的信道注意模块，提出了 ECAP-YOLO 模型，利用新增检测层替代了大目标检测模块，并利用转置卷积替代了上采样。对比结果如表 22 所示，通过数据可知，本文的改进模型不仅可以得到全面的评价指标，且性能更优。

表 22 同类型实验结果对比

	P	R	mAP @0.5	mAP @.5:.95	Loss
[31]	0.952	0.931	0.949	0.509	—
[32]	0.956	0.956	0.978	0.658	—
[33]	0.729	0.493	0.476	0.176	—
本文	0.9677	0.9958	0.9927	0.951	0.0396

5 总结

利用改进 YOLO V5s 模型实现遥感图像目标检测，并用于地域贫困评估。传统算法普遍基于统计数据，获取成本高、耗时久、易受人为因素影响。相反，遥感图像具备廉价、高效、精准、客观等优势。因此，遥感图像目标检测作为一种无数据检测方法，可以成为研究地域经济发展的新途径。模型的改进涉及结构及算法等多个方面。检测阶段利用 COCO128 数据集，P 值从 0.6852 增至 0.7346，提升 7.21%；R 值从 0.9729 增至 0.9802，提升 0.75%；mAP@0.5 值从 0.9664 增至 0.9763，提升 1.024%；mAP@0.5:0.95 值从 0.797 增至 0.8541，提升 7.164%。而损失值从 0.0661 降至 0.0645，下降 2.481%。验证阶段利用自定义遥感图像数据集，P 值从 0.9217 增至 0.9507，提升 3.146%；R 值从 0.9753 增至 0.9958，提升 2.102%；mAP@0.5 值从 0.9795 增至 0.9931，提升 1.388%；mAP@0.5:0.95 值从 0.8433 增至 0.8909，提升 5.644%。而损失值从 0.0515 降至 0.0322，下降 37.435%。因此，与原模型相比，改进模型具备更精准的目标检测能力。另外，利用贫困率分析地域经济发展情况，遥感图像目标检测方法和传统的统计数据方法获得了相同的结论，从而证明了改进模型的有效性。

参考文献:

- [1] Bitar A W, Cheong L F, Ovarlez J P. Sparse and low-rank matrix decomposition for automatic target detection in hyperspectral imagery[J]. IEEE Transactions on Geoscience

- and Remote Sensing, 2019,57(8):5239-5251.
- [2] Peng M, Zhang L, Sun X, et al. A Fast Three-Dimensional Convolutional Neural Network-Based Spatiotemporal Fusion Method (STF3DCNN) Using a Spatial-Temporal-Spectral Dataset[J]. Remote Sensing, 2020,12(23):3888.
 - [3] Nasrabadi N M. Hyperspectral target detection: An overview of current and future challenges[J]. IEEE Signal Processing Magazine, 2013,31(1):34-44.
 - [4] Chiang S S, Chang C I, Ginsberg I W. Unsupervised target detection in hyperspectral images using projection pursuit[J]. IEEE Transactions on Geoscience and Remote Sensing, 2001, 39(7): 1380-1391.
 - [5] Reed I S, Gagliardi R M, Stotts L B. Optical moving target detection with 3-D matched filtering[J]. IEEE Transactions on Aerospace and Electronic Systems, 1988,24(4):327-336.
 - [6] Li J, Dai Y, Li C, et al. Visual detail augmented mapping for small aerial target detection[J]. Remote Sensing, 2018,11(1):14.
 - [7] Na G S, Chang H. Unsupervised Subspace Extraction via Deep Kernelized Clustering[J]. ACM Transactions on Knowledge Discovery from Data (TKDD), 2021,16(1):1-15.
 - [8] Kumar G, Bhatia P K. A detailed review of feature extraction in image processing systems[C]. 2014 Fourth international conference on advanced computing & communication technologies. IEEE, 2014: 5-12.
 - [9] Storcheus D, Rostamizadeh A, Kumar S. A survey of modern questions and challenges in feature extraction[C]. Feature Extraction: Modern Questions and Challenges. PMLR, 2015:1-18.
 - [10] Razaque A, Ben Haj Frej M, Almi'ani M, et al. Improved support vector machine enabled radial basis function and linear variants for remote sensing image classification[J]. Sensors, 2021,21(13):4431.
 - [11] Li S, Xu Y, Zhu M, et al. Remote sensing airport detection based on end-to-end deep transferable convolutional neural networks[J]. IEEE Geoscience and Remote Sensing Letters, 2019,16(10): 140-1644.
 - [12] Li D, Che X, Luo W, et al. Digital watermarking scheme for colour remote sensing image based on quaternion wavelet transform and tensor decomposition[J]. Mathematical Methods in the Applied Sciences, 2019,42(14):4664-4678.
 - [13] Azam F, Rizvi A, Khan W Z, et al. Aircraft Classification Based on PCA and Feature Fusion Techniques in Convolutional Neural Network[J]. IEEE Access, 2021,9:161683-161694.
 - [14] Jianchang R, Zabalza J, Marshall S, et al. Effective feature extraction and data reduction in remote sensing using hyperspectral imaging[J]. IEEE Signal Processing Magazine, 2014,3(4):149-154.
 - [15] Li Y, Zhang H, Xue X, et al. Deep learning for remote sensing image classification: A survey[J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2018,8(6):e1264.
 - [16] Yuan Q, Shen H, Li T, et al. Deep learning in environmental remote sensing: Achievements and challenges[J]. Remote Sensing of Environment, 2020,241:111716.
 - [17] Zheng Y, Dong-Ye C, Wang P, et al. Improved YOLOv5 Network Method for Remote Sensing Image Based Ground Objects Recognition[J]. 2022. DOI: 10.21203/rs.3.rs-1224458/v1.
 - [18] Yan J, Wang H, Yan M, et al. IoU-adaptive deformable R-CNN: Make full use of IoU for multi-class object detection in remote sensing imagery[J]. Remote Sensing, 2019,11(3):286.
 - [19] Girshick R. Fast r-cnn[C]. Proceedings of the IEEE international conference on computer vision. 2015:1440-1448.
 - [20] Ji H, Gao Z, Mei T, et al. Improved faster R-CNN with multiscale feature fusion and homography augmentation for vehicle

- detection in remote sensing images[J]. IEEE Geoscience and Remote Sensing Letters, 2019,16(11):1761-1765.
- [21] Zhang L, Wu J, Fan Y, et al. An efficient building extraction method from high spatial resolution remote sensing images based on improved mask R-CNN[J]. Sensors, 2020,20(5):1465.
- [22] Xu D, Wu Y. Improved YOLO-V3 with DenseNet for multi-scale remote sensing target detection[J]. Sensors, 2020,20(15): 4276.
- [23] Qu J, Su C, Zhang Z, et al. Dilated convolution and feature fusion SSD network for small object detection in remote sensing images[J]. IEEE Access, 2020,8: 82832-82843.
- [24] Tang G, Zhuge Y, Claramunt C, et al. N-Yolo: A SAR ship detection using noise-classifying and complete-target extraction[J]. Remote Sensing, 2021,13(5):871.
- [25] Luo S, Yu J, Xi Y, et al. Aircraft Target Detection in Remote Sensing Images Based on Improved YOLOv5[J]. IEEE Access, 2022.
- [26] Zhou K, Zhang M, Wang H, et al. Ship Detection in SAR Images Based on Multi-Scale Feature Extraction and Adaptive Feature Fusion[J]. Remote Sensing, 2022,14(3): 55.
- [27] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018:7132-7141.
- [28] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]. Proceedings of the European conference on computer vision (ECCV). 2018:3-19.
- [29] Qilong Wang, Banggu Wu, Pengfei Zhu, et al. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
- [30] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 13713-13722.
- [31] Shao J, Yang Q, Luo C, et al. Vessel detection from nighttime remote sensing imagery based on deep learning[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021,14:12536-12544.
- [32] Xie F, Lin B, Liu Y. Research on the Coordinate Attention Mechanism Fuse in a YOLOv5 Deep Learning Detector for the SAR Ship Detection Task[J]. Sensors, 2022,22(9):3370.
- [33] Kim M, Jeong J, Kim S. ECAP-YOLO: Efficient Channel Attention Pyramid YOLO for Small Object Detection in Aerial Image[J]. Remote Sensing, 2021,13(23): 4851.

[作者简介]

张晨光（1984-），男（汉族），河北省沧州市人，博士研究生，主要研究领域为大数据及人工智能。

滕桂法（1963-），男（汉族），河北省保定市人，教授，博士研究生导师，主要研究领域为农业信息化技术。

丁文卿（1989-），女（汉族），河北省沧州市人，硕士，主要研究领域为大数据处理及分析。