基于深度学习的遥感图像小目标检测技术研究

余 江,杨晓青*

南昌职业大学校长办公室, 江西 南昌

收稿日期: 2025年6月25日; 录用日期: 2025年7月22日; 发布日期: 2025年7月30日

摘要

本文针对遥感图像小目标检测中存在的特征提取困难、背景干扰严重以及检测精度和速度难以平衡等问题,深入探讨了多种创新方法。通过对相关算法如YOLOv11系列的改进,引入新型模块与机制,显著提升了小目标检测性能。在多个公开数据集及自建数据集上的实验表明,改进后的算法在小目标检测的准确率、召回率等指标上有显著提升,为遥感图像小目标检测的实际应用提供了有力的技术支持与参考。

关键词

深度学习,遥感图像小目标,测试技术

Research on Small Object Detection Technology in Remote Sensing Images Based on Deep Learning

Jiang Yu, Xiaoqing Yang*

The President Office of Nanchang Vocational University, Nanchang Jiangxi

Received: Jun. 25th, 2025; accepted: Jul. 22nd, 2025; published: Jul. 30th, 2025

Abstract

This article explores various innovative methods to address the difficulties in feature extraction, severe background interference, and difficulty in balancing detection accuracy and speed in small object detection in remote sensing images. By improving related algorithms such as the YOLOv11 series and introducing new modules and mechanisms, the performance of small object detection has been significantly improved. Experiments on multiple public and self-built datasets have shown that the improved algorithm significantly improves the accuracy and recall of small object detection, providing strong

*通讯作者。

technical support and reference for the practical application of small object detection in remote sensing images.

Keywords

Deep Learning, Small Targets in Remote Sensing Images, Testing Technology

Copyright © 2025 by author(s) and Hans Publishers Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). http://creativecommons.org/licenses/by/4.0/

1. 引言

随着科技的迅猛发展,遥感技术作为现代科技的重要组成部分,正逐渐改变着我们的生活。遥感技术通过传感器收集地球表面的信息,包括图像、光谱等数据,为农业、环保、城市规划等领域提供了便捷、高效的监测手段。在农业领域,遥感技术可以实时监测作物生长环境,为农业灾害预警和农田管理提供有力支持。在环保领域,遥感技术可以捕捉海洋、陆地、大气等自然环境的实时变化,为环境保护和污染治理提供重要依据。在城市规划领域,遥感技术可以为城市规划提供全面、准确的地理信息和环境监测数据,助力城市可持续发展。

遥感图像的目标检测[1]是一个备受关注的研究方向。国内外学者已经取得了一些研究成果。 Lindeberg [2]提出了一种尺度不变特征变换(SIFT)算法,该算法对图像域内的平移、旋转和缩放变换具有 不变性,对适度透视变换和光照变化具有鲁棒性。Tomasi [3]提出了一种方向直方图(Histograms of Oriented Gradient, HOG)算法,采用中心差分法对已划分的大小固定图像进行梯度近似计算,并在每个单元格内累 积梯度方向的直方图,再归一化处理以降低对比度的变化。Cheng等人[4]提出了一种学习旋转不变 HOG 特征的方法,通过改进一种新的目标函数来学习旋转不变模型,使旋转前后的训练样本具有相似特征, 从而实现旋转不变性。Cai 等人[5]提出了一种面向对象的遥感图像变化检测方法,通过融合 Gabor 纹理 与马尔可夫随机场纹理等多种特征,并利用加权模糊 c 均值算法和 Relief 算法确定的特征权重,提高变 化检测的准确性,降低单一权重特征带来的不确定性。Peng 等人[6]提出了一种差异增强密集注意力卷积 神经网络(DDCNN),旨在对双时相光学遥感图像中的变化进行有效识别。Cao 等人[7]提出了一种基于 Swin-Transformer 和 YOLOv5 的融合模型,通过引入完整的交并比和改进 K-means 聚类算法,优化锚框 生成,增强全局上下文信息提取,并采用坐标注意力机制提升特征检测精度。Chen 等人[8]提出了一种全 局上下文空间注意力网络(GCSANet),通过整合全局上下文空间注意力机制与密集连接的卷积网络,有效 地提取了遥感场景中的多尺度全局特征。Ye 等人[9]提出了一种基于自适应注意力融合机制的卷积网络 (AAFM),该网络模型基于 EfficientDet [10]的骨干网络,旨在自动融合最优特征信息,并克服对多尺度对 象的敏感性,提高卷积神经网络在遥感目标检测中的性能。

2. 遥感图像小目标检测算法

YOLOv11 模型结构及工作原理

1、YOLOv11 模型结构

YOLOv11 [11]是 Ultralytics 公司于 2024 年官方发布的最新一款基于图像的深度学习模型。YOLOv11 摒弃传统的卷积神经网络(CNN)骨干,转而采用 Transformer 架构。Transformer 通过自注意力机制,能够有效捕捉长距离依赖关系,克服了 CNN 感受野的限制,尤其在检测小目标时表现出色。比如在复杂场景

的遥感图像中,微小的建筑物或车辆等小目标,以往的 CNN 骨干网络可能难以精准识别,而 YOLOv11 的 Transformer 骨干网络可通过全局上下文信息,更准确地定位和分类这些小目标。其性能指标相较于前代产品有质的提高,如图1所示。



Figure 1. Performance comparison diagram between YOLOv11 and other YOLO models 图 1. YOLOv11 与其他 YOLO 模型性能对比图

相较于传统的 CNN 骨干网络,如在早期 YOLO 系列中广泛应用的 Darknet 系列, Transformer 骨干 网络具有多方面的优势。CNN 的卷积操作虽然在局部特征提取方面表现出色,但随着网络层数的加深, 梯度消失和梯度爆炸问题逐渐凸显,限制了模型对全局信息的获取能力。此外, Transformer 在参数共享 方式上也与 CNN 不同,它通过多头注意力机制,能够同时关注图像的不同方面,学习到更丰富、更具代 表性的特征,进一步提升了模型的泛化能力和检测性能。YOLOv11 的网络结构如下图 2 所示。



Figure 2. Network architecture diagram of YOLOv11 图 2. YOLOv11 的网络结构图

YOLOv11 作为流行的目标检测算法家族中较新的一员, Jocher 和他的团队[12]进行了多项改进和优

化。主要有以下三个创新点:

1) 提出 C3K2 机制: YOLOv11 引入了 C3K2 机制,其中 C3K2 模块是由 C2F 模块演变而来的。当 参数 C3k 设置为 False 时,C3K2 模块相当于 C2F 模块,如图 3 所示。

#测试参数为 False 的情况 setup_false = "from your_module import C3K2" stmt_false = "C3K2(arg1, False)" time_false = timeit.timeit(stmt_false, setup_false, number=1000) print (f"参数 True 时的执行时间: {time_true: .6f} 秒") print (f"参数 False 时的执行时间: {time_false:6f} 秒") print (f"性能差异: {time_false / time_true - 1) * 100:2f96")

Figure 3. The second parameter of C3K2 is set to False diagram 图 3. C3K2 第二个参数被设置为 False 图



当 C3k 设置为 True 时, Bottleneck 模块被替换为 C3k 模块, C3k 机制的网络结构图如图 4 所示。

2) 创造 C2PSA 机制: C2PSA 是对 C2f 模块的扩展,结合了 PSA (Position-Sensitive Attention: 位置 信息的注意力机制)模块,旨在通过多头注意力机制和前馈神经网络来增强特征提取能力。它可以选择性 地添加残差结构(shortcut)以优化梯度传播和网络训练效果。同时,使用 FFN 可以将输入特征映射到更高 维的空间,捕获输入特征的复杂非线性关系,允许模型学习更丰富的特征表示,如图 5 所示。

3) 引用深度可分离卷积(DWConv): YOLOv11 在其分类检测模块集成了两层深度卷积(Depthwise Convolution),该设计显著削减了参数数量与计算负荷。DWConv 作为一种高效卷积运算,凭借对各通道的独立处理降低了运算量,并与1×1的点态卷积组合,形成深度可分离的卷积(具体见图 6)。

在分类检测头中,YOLOv11 插入了两个 DWConv,这样的做法大幅度减少参数量和计算量,原有的 两个普通的 Conv 卷积和由 3 变为了 1。

2、基于 ACMix 的自注意力机制的遥感图像小目标检测算法

ACMix 乃是一种融合型架构,它巧妙地结合了自注意力单元与卷积运算各自的长处。其核心理念在于,传统卷积处理及自注意力模块中的多数运算均可借由1×1卷积高效完成。ACMix 在初始运作时,使用1×1卷积来投射输入的特征图谱,进而生成一系列中间状态的特征表达。之后,它会依据自注意力或卷积这类相异的计算范式,对这些中间特征分别进行复用与整合。如此设计使得 ACMix 既能利用自注意力的全局感知优势,又能通过卷积捕捉局部细微特征,从而在控制运算成本的同时,提升模型整体性能。整体算法结构如图7所示。

Figure 4. C3k mechanism diagram 图 4. C3k 机制图



Figure 5. C2PSA mechanism diagram 图 5. C2PSA 机制图



Figure 6. Comparison diagram of decoupling heads between YOLOv8 and YOLOv11 图 6. YOLOv8 与 YOLOv11 解耦头对比图



Figure 7. Structure schematic diagram of YOLO-ACMix algorithm 图 7. YOLO_ACMix 算法结构示意图

ACMix 模型关键的改良机制可归纳为两大方面。其一为自注意力与卷积的协同整合。此法将自注意 力技术及卷积技术加以融合,旨在实现两者优点的结合。其二则是运算的解析及重塑。此概念是指将传 统的卷积运算和自注意力运算予以拆分,随后重新构建为效率更高的形式。

1) 自注意力与卷积的协同整合

自注意力与卷积的协同整合具体途径包括:

(a) 特征的拆解: 自注意力机制中的查询(query)、键(key)、值(value)以及卷积操作均通过 11 卷积进 行特征拆解。

(b) 运算资源共享:卷积路径和自注意力路径共用相同的1×1卷积运算,借此减少了重复计算。

(c) 特征的融汇: 在 ACMix 模型内部,由卷积与自注意力产生的特征通过求和操作进行合并,以此 增强模型的特征提取效能。

(d) 模块化的设计理念: ACMix 因其模块特性,能够灵活植入到多样化的网络体系之中,并有助于 优化网络的特征展现能力。



Figure 8. Convolution, self-attention, and ACMix construction and computational overhead diagram 图 8. 卷积、自注意力及 ACMix 构造与计算开销图

图 8 阐明了 ACMix 的核心理念,它对卷积、自注意力机制以及 ACMix 自身的结构和计算开销进行 了对比分析。图中内容细分如下:

(a) 卷积模块:展示了其标准运作流程,其中包含一个 K 平方(K²)规模的 1 × 1 卷积环节,这代表了 卷积核的尺寸范畴及卷积运算的聚合效应。

(b) 自注意力机制: 阐释了此种自注意力构造,其关键在于三次使用1×1卷积,这种设计体现了多 头架构下各个头部的线性变换过程,并且整合了自注意力机制的聚合步骤。

(c) ACMix:介绍了 ACMix (即本研究提出的方法),它融合了卷积与自注意力的聚合功能,关键之处 在于 1×1 卷积在这两种路径间实现了共享,目的在于降低运算耗费并集成轻量化的聚合操作。

从整体上看,ACMix 致力于通过共享计算组件(特别是1×1卷积)并结合两种相异的聚合方式,来优化特征通道层面的计算复杂度。

2) 运算的解析及重塑

在 ACMix 架构体系内,所谓运算的解析及重塑,其核心要义在于对传统的卷积操作及自注意力计算 过程进行拆解,并将其转变成一种更具运行效率的实现形式。此方法主要通过下列步骤得以达成:

(a) 对卷积与自注意力进行分解处理:此举将标准的卷积核心拆分为数个1×1卷积单元,各单元用

以处理不同的特征区域;此外,自注意力组件中用以产生查询(query)、键(key)与数值(value)的过程,也被转换为基于1×1卷积的运算。

(b) 重构为集成的混合模块:接着,这些经过分解的运算被重新构建为一个一体化的混合式模块,该 模块既保有卷积获取空间细节的专长,也融合了自注意力执行全局信息聚合的独特优势。

(c) 提升整体运算效能:此种分解及重塑的方案,由于有效地减少了重复性的计算步骤,因此显著增强了运算的综合效率,并同步减轻了模型构建的内在复杂性。



Figure 9. Construction diagram of ACMix mixing module 图 9. ACMix 混合模块构造图

图 9 清晰地展示了 ACMix 所提出的混合模块的构造。该图具体阐释如下:

(a) 卷积模块:关于卷积操作,其中尺寸为3×3的卷积是借由1×1卷积方法被拆解的,此部分揭示了特征映射的变换流程。

(b) 自注意力机制:说明了输入信号首先被变换为查询(query)、键(key)以及数值(value),此过程是运用1×1卷积来达成的,且经由相似性的匹配来运算注意力分配的权重。

(c) ACMix: 阐述了 ACMix 本身, 它融合了前述(a)与(b)两部分的特性; 其初始步骤运用三次独立的 1×1 卷积来对输入特征映射执行投射操作, 接下来的阶段则将两条不同路径所获得的特征予以加和, 以 此形成最终的输出结果。

图 8 揭示了 ACMix 单元的运作流程,着重突出了两种核心机制的有机结合,并且给出了各个运算单 元所对应的计算开销。

3. 数据集构建与处理

3.1. 数据集的获取

在机器学习任务中,特别是在目标侦测方面,合适的数据集选择至关重要。数据集不仅直接影响着 模型的表现与精确度,还会对实际应用效果产生长远影响。为完成目标识别任务,本实验中使用的是 DOTA 数据集[13],是由 Google earth 等卫星拍摄,数据集由 12,013 张图像组成,包含 188,282 个实例, 共有 15 种分类。如图 10 所示,该数据集有 15 种不同的目标分类,分别为飞机、棒球场、桥梁、操场跑 道、小型车辆、大型车辆、轮船、网球场、篮球场、储油罐、足球场、环路、港口、游泳池和直升机。图 片的分辨率范围从 400 × 400 至 4000 × 4000,涵盖了各种方向和形状的目标。其中,训练集、验证集和 测试集之间是严格独立的。



Figure 10. Distribution diagram of the number of categories in the dataset 图 10. 数据集中各类别数量分布图

3.2. 数据集的预处理与标注

1) 数据集的预处理

数据集预处理是保证模型训练效果的关键步骤,特别是在处理图像数据时,统一图片尺寸有助于提升模型的稳定性和性能。为此,本研究利用 Roboflow 平台对所有图片进行了尺寸规范化处理成 416×416,确保输入数据的一致性,从而优化训练过程和模型表现。如图 11 所示。

@ #		C Versions					+ Cre	ate New Version
•	Sciuridae i	Versions		model to use feature	s like Label Assist and Mod our auto-scaling API and e	el Evaluation and deployn dge device support.	nent options like	
ela	DATA	2025-04-11 8:18am			∳ Custom	Train		
*	Upload Data Upload Data	Stretch to			Uses Robofiow Credits	View usage 2		
8	S Annotate				How to Upload C	ustom Weights		
	Dataset 987							
	D Versions Train							
	@ Analytics		987 Total Im	ages			Vie	w All Images →
	IE Classes & Tags		17	Martin Comments				
	MODELS							
	-b¢ Models							
	LS, Visualize		Dataset Split	Concerned and the				
	DEPLOY			TRAIN SET	VALID SET		TEST SET	
ø				987 Images	0 Images		o imagés	
4			Preprocessing	Auto-Orient: Applied Resize: Stretch to 416x416				
Z			Augmentations	No augmentations were appli	ed.			
				and a second sec				

Figure 11. Preprocessed data diagram of Roboflow platform 图 11. Roboflow 平台预处理数据图

2) 数据集标注

本文标注采取人工标注。标注工具是 LabelImg。其操作界面简洁,通过手动绘制矩形框,精确框选 遥感图像中的小目标,并赋予相应的类别标签。在标注过程中,需严格遵循统一的标注规范,例如对于 车辆目标,明确规定无论车辆处于静止还是行驶状态,只要能清晰识别轮廓,均需标注。人工标注虽然 耗时耗力,但能够保证标注的高质量。标注界面如图 12 所示。



Figure 12. LabelImg labeling diagram 图 12. LabelImg 标注图

3.3. 数据集的划分与分析

本文通过随机划分的方式,将资料集划分为训练、验证及测试三组,各组比例定为7:2:1。数据分配 详情如下:训练集部分涵盖了8408幅图片资料,验证部分持有2402幅,而测试部分则包括1203幅。采 用此种划分方案,其目的在于确保各个数据组别均能充分地代表整体样本并且维持分布上的均衡状态。

标签	Car	Track	Plane	Ship	Tennis
训练集	4918	700	700	1400	690
验证集	1405	200	200	400	197
测试集	703	100	100	200	100
总计	7026	1000	1000	2000	987

Table	1. Statistics	table of the	e number	of various	labels in	the	dataset
表1.	数据集各类	标签数量组	充计表				

在表1中,各类别的样本数量表现不均衡,尤其是 car 和 ship 类样本比较多,而 track、plane 和 tennis 类样本比较少,符合小目标种类多样性特征,适合运用于多任务学习场景。验证集所含样本的规模,约 占训练样本总量的百分之二十五。整体图像资料数量相当可观,特别是在 car 和 ship 这两个类别上,为 模型学习过程奠定了坚实的数据基础。此数据集适于构建能够辨识不同类别的模型,进而有助于提升其

泛化性能。验证与测试部分的样本规模合理,这将有助于精确衡量所建模型的实际效能。配合适宜的预 处理技术及训练策略,可有效利用此图像资料在模型构建与成效评估环节的价值。

4. 实验分析

1) 算法性能分析比较

为了进一步验证改进算法在遥感图像目标检测任务中的性能,本实验中将 YOLOv5 [14]、YOLOv11 和本文算法作为对比实验,为确保比较的公正性,各项实验中的 YOLO 模型均使用了相同的数据资料来 源及一致的划分标准,并且训练阶段的超参数设定也维持不变。实验结果如下图 13 所示。

各模型的实验评估趋势在图 11 中得以呈现。图中清晰列示了关键性能指标,例如召回率、mAP 和 精度。其中 YOLOv11 的 mAP 达到最高,但其召回率和精度并不是最高。从图中可以看出从 YOLOv5 至 YOLOv11_ACMix 模型的精确度不断提高,YOLOv11_ACMix 的精度最高。



Figure 13. Comparison diagram of YOLOv5, YOLOv11 and YOLOv11_ACMix models 图 13. YOLOv5、YOLOv11 和 YOLOv11_ACMix 模型对比图

模型	指标	综合	Car	Track	Plane	Ship	Tennis
VOL Out	mAP	0.923	0.993	0.812	0.902	0.995	0.913
101003	Recall	0.894	0.981	0.765	0.862	0.988	0.875
	mAP	0.925	0.991	0.831	0.891	0.995	0.915
TOLOVIT	Recall	0.892	0.975	0.779	0.857	0.998	0.854
VOLOU11 ACMin	mAP	0.924	0.993	0.827	0.890	0.995	0.917
TOLOWIT_ACMIX	Recall	0.899	0.981	0.787	0.847	0.995	0.885

 Table 2. Detailed experimental results table of animal identification

 表 2. 动物识别详细实验结果表

表 2 中列出了所有实验详细数据信息,从综合指标上看,YOLOv11 的 mAP 最高,达到 92.5%,而 YOLOv11_ACMix 的召回率和精确度最高,分别为 89.9%和 96.3%。YOLOv11_ACMix 模型表现效果最 好。具体从召回率(recall)和平均精确度均值(mAP)两个方面进行分析。在 mAP 指标上,YOLOv11 表现最 佳,总体 mAP 达到 92.5%。此外,YOLOv11_ACMix 总体 mAP 数值仅次于 YOLOv11,达到了 92.4%。 而 YOLOv5 的整体 mAP 也达到了 92.3%。在召回率方面,YOLOv11_ACMix 的综合召回率排名最高, 达到 89.9%。

根据表 3 的数据,对以上三种模型的浮点数和参数量展开对比分析。YOLOv5 在 FLOPs 上高于 YOLOv11 和 YOLOv11_ACMix,但 Parameters 低于 YOLOv11 和 YOLOv11_ACMix,而 YOLOv11 与 YOLOv11_ACMix 的 FLOPs 相等,但 YOLOv11 的 Parameters 低于 YOLOv11_ACMix 的 Parameters。较低的参数量和较低的 FLOPs 意味着模型更加轻量化,计算效率更高,运行速度更快。综合分析表 2 与表 3 的数据可见,YOLOv11_ACMix 在识别准确性与运算速度两方面均有出色建树,因此更适合应用于小目标识别的应用场景之中。

Table 3. Comparison table of floating-point numbers and parameter quantities for YOLOv5, YOLOv11, and YOLOv11_ACMix models

表 3. YOLOv5	、YOLOv11 和	YOLOV11 ACM	ix 模型浮点致和参致重对比表
		_	

选取模型	FLOPs (G)	Parameters
YOLOv5	7.1	2,503,319
YOLOv11	6.3	2,583,127
YOLOv11_ACMix	6.3	2,593,359

2) 可视化结果分析

图 14 展示了 YOLOv11_ACMix 模型在训练阶段的收敛表现曲线。观察此图可知,该模型的收敛进程 展现出平稳且迅速的特性,约在 100 个训练周期之后趋于稳定。此种现象表明,YOLOv11_ACMix 模型在 应对动物识别类任务时,具备了良好的运行稳定性与卓越的数据拟合效果。另外,模型能够迅速收敛也反 映出其在学习阶段可以高效捕捉数据集内的核心特征,从而减小了发生过拟合的可能性。这不仅增强了模 型在实际应用中的可靠度,亦为后续的改进与功能拓展奠定了基础。借由对数据资料集的优化或对训练参 数的精细调整,有望进一步提升其检测的精确度及泛化适应性,从而更好地契合实际应用场景的需要。



Figure 14. YOLOv11_ACMix model training convergence diagram 图 14. YOLOv11_ACMix 模型训练收敛图

图 15 展示了应用本文提出的算法得到的一系列可视化结果,这些结果证明了算法在处理多样化背景 下的车辆检测任务上的卓越性能。无论是穿梭于城市道路的轿车、停泊在停车场的轿车、隐藏在住宅区 间的汽车,还是停靠在路边的各类车辆,算法均能够以高度的准确性进行识别和定位。



Figure 15. Visualization diagram of test results 图 15. 检测结果可视化图

5. 结束语

本文深入探讨了遥感图像小目标检测问题,并基于YOLOv11网络开发了一种先进的目标检测网络: YOLOv11_ACMix,提出了一种基于ACMix的自注意力机制的遥感图像小目标检测算法,这种算法在处 理具有复杂背景的水平和有向标注数据集时,展现出了优异的检测性能。

基金项目

2022年江西省教育厅科学技术研究项目(GJJ2204508)。

参考文献

- Zou, Z., Chen, K., Shi, Z., Guo, Y. and Ye, J. (2023) Object Detection in 20 Years: A Survey. *Proceedings of the IEEE*, 111, 257-276. <u>https://doi.org/10.1109/jproc.2023.3238524</u>
- [2] Lindeberg, T. (2012) Scale Invariant Feature Transform. *Scholarpedia*, 7, Article 10491. <u>https://doi.org/10.4249/scholarpedia.10491</u>
- [3] Tomasi, C. (2022) Histograms of Oriented Gradients. Computer Vision Sampler, 1-6.
- [4] Cheng, G., Zhou, P., Yao, X., Yao, C., Zhang, Y. and Han, J. (2016) Object Detection in VHR Optical Remote Sensing Images via Learning Rotation-Invariant HOG Feature. 2016 4th International Workshop on Earth Observation and Remote Sensing Applications (EORSA), Guangzhou, 4-6 July 2016, 433-436. <u>https://doi.org/10.1109/eorsa.2016.7552845</u>
- [5] Cai, L., Shi, W., Hao, M., Zhang, H. and Gao, L. (2018) A Multi-Feature Fusion-Based Change Detection Method for Remote Sensing Images. *Journal of the Indian Society of Remote Sensing*, 46, 2015-2022. <u>https://doi.org/10.1007/s12524-018-0864-1</u>

- [6] Peng, X., Zhong, R., Li, Z. and Li, Q. (2021) Optical Remote Sensing Image Change Detection Based on Attention Mechanism and Image Difference. *IEEE Transactions on Geoscience and Remote Sensing*, 59, 7296-7307. <u>https://doi.org/10.1109/tgrs.2020.3033009</u>
- [7] Cao, X., Zhang, Y., Lang, S. and Gong, Y. (2023) Swin-Transformer-Based YOLOv5 for Small-Object Detection in Remote Sensing Images. *Sensors*, 23, Article 3634. <u>https://doi.org/10.3390/s23073634</u>
- [8] Chen, W., Ouyang, S., Tong, W., Li, X., Zheng, X. and Wang, L. (2022) GCSANet: A Global Context Spatial Attention Deep Learning Network for Remote Sensing Scene Classification. *IEEE Journal of Selected Topics in Applied Earth Observations* and Remote Sensing, 15, 1150-1162. <u>https://doi.org/10.1109/jstars.2022.3141826</u>
- [9] Ye, Y., Ren, X., Zhu, B., Tang, T., Tan, X., Gui, Y., et al. (2022) An Adaptive Attention Fusion Mechanism Convolutional Network for Object Detection in Remote Sensing Images. *Remote Sensing*, 14, Article 516. https://doi.org/10.3390/rs14030516
- [10] Tan, M., Pang, R. and Le, Q.V. (2020) EfficientDet: Scalable and Efficient Object Detection. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, 13-19 June 2020, 10778-10787. https://doi.org/10.1109/cvpr42600.2020.01079
- [11] <u>https://github.com/ultralytics/ultralytics</u>
- [12] Jegham, N., Chan, Y.K. and Marwan, A. (2024) Evaluating the Evolution of YOLO (You Only Look Once) Models: A Comprehensive Benchmark Study of YOLO11 and Its Predecessors. arXiv: 2411.00201v1.
- [13] Xia, G., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., et al. (2018) DOTA: A Large-Scale Dataset for Object Detection in Aerial Images. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, 18-23 June 2018, 3974-3983. <u>https://doi.org/10.1109/cvpr.2018.00418</u>
- [14] Liu, Y., Shao, Z. and Hoffmann, N. (2021) Global Attention Mechanism: Retain Information to Enhance Channel-Spatial Interactions. arXiv: 2112.05561.