



(12) 发明专利申请

(10) 申请公布号 CN 121837788 A

(43) 申请公布日 2026. 04. 10

(21) 申请号 202610087851.7

(22) 申请日 2026.01.22

(71) 申请人 北京科技大学

地址 100083 北京市海淀区学院路30号

(72) 发明人 李诗媛 詹文豪 王传浩 赵文哲

(74) 专利代理机构 北京金智普华知识产权代理有限公司 11401

专利代理师 岳野

(51) Int. Cl.

G06V 10/764 (2022.01)

G06V 10/25 (2022.01)

G06V 10/22 (2022.01)

G06V 10/26 (2022.01)

G06V 10/52 (2022.01)

G06V 20/13 (2022.01)

G06V 10/80 (2022.01)

G06V 10/82 (2022.01)

G06N 3/045 (2023.01)

G06N 3/0464 (2023.01)

G06N 3/08 (2023.01)

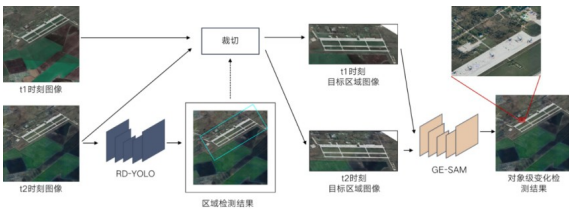
权利要求书2页 说明书10页 附图1页

(54) 发明名称

一种基于视觉大模型的目标级变化检测方法、装置、设备及介质

(57) 摘要

本发明涉及一种基于视觉大模型的目标级变化检测方法、装置、设备及介质,方法包括:针对大范围遥感影像,采用RD-YOLO旋转目标检测网络进行精确提取得到包括大尺度目标场景的双时相感兴趣区域ROI;对双时相感兴趣区域ROI采用OD-YOLO目标检测模型进行定位检测,生成包括候选目标边界框的目标区域图像;将候选目标边界框采用视觉大模型SAM的GE-SAM模型进行分割,生成分割掩模;对分割掩模进行匹配和交并比计算,通过预设的阈值来判断,最终输出对小尺寸目标的变化检测结果。本方法无需大量特定场景标注数据即可实现目标级变化检测,跨场景适应性强,且检测精度与效率有所提升,应用范围更广。



1. 一种基于视觉大模型的目标级变化检测方法,其特征在于,所述方法包括:
 - S1. 针对大范围遥感影像,采用RD-YOLO旋转目标检测网络进行精确提取得到包括大尺度目标场景的双时相感兴趣区域ROI;
 - S2. 对双时相感兴趣区域ROI采用OD-YOLO目标检测模型进行定位检测,生成包括候选目标边界框的目标区域图像;
 - S3. 将候选目标边界框采用视觉大模型SAM的GE-SAM模型进行分割,生成分割掩模;
 - S4. 对分割掩模进行匹配和交并比计算,通过预设的阈值来判断,最终输出对小尺寸目标的变化检测结果。
2. 根据权利要求1所述的方法,其特征在于,步骤S1具体包括:
 - S11. 采用骨干网络对所述遥感影像进行分层特征提取,得到不同尺度特征图;
 - S12. 采用骨干网络中的特征融合模块对尺度特征图中的低层级特征图和高层级特征图进行深度融合,生成融合特征图;
 - S13. 将融合后的特征图进行旋转目标检测与坐标映射,得到包括大尺度目标场景的双时相感兴趣区域ROI。
3. 根据权利要求1所述的方法,其特征在于,步骤S2包括:
 - S21. 对双时相感兴趣区域ROI的不同场景类别动态设置滑动窗口的尺寸;
 - S22. 采用滑动窗口在双时相感兴趣区域ROI上按设定步长进行滑动,生成两个局部切片图像集;
 - S23. 将两个局部切片图像集中的每一个局部切片图像分别输入到两个OD-YOLO网络中进行处理,得到两个包括候选目标边界框的目标区域图像。
4. 根据权利要求1所述的方法,其特征在于,所述S3包括:
 - S31. 对目标区域图像的候选目标边界框提出其几何中心点;
 - S32. 对 t_1 或 t_2 时刻的目标区域图像经滑动窗口处理后得到的图像切片采用图像编码器进行处理,并提取高维图像特征嵌入;
 - S33. 将几何中心点映射为点提示嵌入向量;
 - S34. 采用交叉注意力机制融合高维图像特征嵌入与点提示嵌入向量,根据点提示的位置信息,从高维图像特征中分割出与该点提示语义相一致的连通区域;
 - S35. 将得到的所有连通区域作为分割掩模。
5. 根据权利要求1所述的方法,其特征在于,所述S4包括:
 - S41. 对于分割掩模的前一时相中的每一个掩模,遍历后一时相中的所有掩模,并计算前一时相中的每一个掩模与后一时相中的所有掩模之间的交并比;
 - S42. 根据交并比与预先设置的阈值的关系来判断目标的变化。
6. 根据权利要求5所述的方法,其特征在于,所述S42具体包括:对于前一时相的某个掩模,如果在后一时相中找不到任何一个掩模使得所计算的交并比大于等于预先设置的阈值,则判定该目标消失;或者,对于后一时相的某个掩模,如果在前一时相中找不到任何一个掩模使得计算得到的交并比大于等于预先设置的阈值,则判定该目标为新出现。
7. 根据权利要求5所述的方法,其特征在于,所述交并比由两个掩模区域相交部分的面积绝对值除以该两个掩模区域合并部分的总面积绝对值得到。
8. 一种基于视觉大模型的目标级变化检测装置,其特征在于,所述装置用于实现权利

要求1-7任一项所述的方法,包括:

提取模块,用于针对大范围遥感影像,采用RD-YOLO旋转目标检测网络进行精确提取得到包括大尺度目标场景的双时相感兴趣区域ROI;

定位检测模块,用于对双时相感兴趣区域ROI采用OD-YOLO目标检测模型进行定位检测,生成包括候选目标边界框的目标区域图像;

分割模块,用于将候选目标边界框采用视觉大模型SAM的GE-SAM模型进行分割,生成分割掩模;

计算模块,用于对分割掩模进行匹配和交并比计算,通过预设的阈值来判断,最终输出对小尺寸目标的变化检测结果。

9.一种计算机存储介质,其特征在于,所述介质上存储有计算机程序,所述计算机程序被处理器执行实现权利要求1-7任一项所述的方法。

10.一种电子设备,其特征在于,所述电子设备包括:

存储器,存储有可执行指令;

处理器,所述处理器运行所述存储器中的所述可执行指令,以实现权利要求1-7中任一项所述的方法。

一种基于视觉大模型的目标级变化检测方法、装置、设备及介质

技术领域

[0001] 本发明属于遥感图像处理与计算机视觉技术领域,具体涉及一种基于视觉大模型的目标级变化检测方法、装置、设备及介质。

背景技术

[0002] 随着遥感成像、视频监控等数据获取手段的快速发展,多时相图像数据规模不断增加,目标变化检测技术在能源环保、国土资源监测、城市规划、交通管理以及军事侦察等领域具有广泛的应用价值。现有的变化检测方法大多基于像素级或区域级特征,通过差值、阈值、深度学习等手段来识别图像间的变化。目前,现有的变化检测技术主要包括基于像素差异、特征提取与分类以及深度学习的方法。

[0003] 在基于像素差异的方法中,已有研究提出通过图像差分、比值运算或变化向量分析等手段,对不同时相图像的像素值进行差异计算,并结合阈值分割得到变化区域。该类方法实现简单,计算效率较高,但容易受到光照变化、成像噪声和季节性差异的影响,误检率较高,难以区分真实目标的变化与背景干扰。

[0004] 在基于特征提取与分类的方法中,有研究利用主成分分析、纹理特征提取以及光谱特征融合等方式,将多时相图像的差异转化为特征向量,并借助支持向量机或随机森林分类器对其进行分类,以实现变化区域识别。这类方法在一定程度上提高了检测精度,但依赖手工设计特征,泛化能力有限,在复杂场景下鲁棒性不足。

[0005] 近年来,深度学习方法逐渐成为变化检测的主流。部分研究提出基于卷积神经网络的端到端模型,例如利用孪生网络提取两幅时相图像的特征并计算差异,或基于U-Net结构实现像素级变化分割;也有研究尝试采用Transformer架构,通过全局注意力机制提升跨区域特征建模能力。这些方法在公开遥感数据集(如LEVIR-CD、WHU-CD等)上取得了较好的效果,但普遍依赖大规模标注数据,训练成本高,且在不同传感器、不同地物场景下迁移性较差。

[0006] 此外,还有部分研究尝试开展多源遥感数据的变化检测,例如融合光学图像与SAR图像进行联合分析,通过特征对齐与融合提升复杂环境下的适应性。然而,该类方法往往需要额外的多模态配准与融合策略,计算复杂度较高,实用性有限。因此,现有的变化检测技术主要存在以下缺点:

对目标级语义信息利用不足:传统方法多停留在像素差异或低层特征对比上,难以准确区分真实目标变化与噪声干扰。

[0007] 泛化能力有限:基于特定场景或小规模数据集训练的模型,在面对复杂多样的目标类型及多源成像条件时,检测精度显著下降。

[0008] 训练成本高:深度学习方法通常需要大规模标注数据和长时间训练,难以快速迁移到新领域或新任务。

[0009] 鉴于上述不足,现有技术存在的技术问题是如何在无需大量特定场景训练的情况

下,只要充分利用视觉大模型的跨任务泛化能力,即可实现面向多时相、多场景、多目标类型的目标级变化检测,在保证检测精度的同时,提高方法的鲁棒性与适用性,从而满足复杂遥感监测和智能感知应用的需求。

发明内容

[0010] 为了克服现有技术存在的上述问题,本发明提供一种基于视觉大模型的目标级变化检测方法、装置、设备及介质,用于解决现有技术中存在的上述问题。

[0011] 一种基于视觉大模型的目标级变化检测方法,所述方法包括:

S1. 针对大范围遥感影像,采用RD-YOLO旋转目标检测网络进行精确提取得到包括大尺度目标场景的双时相感兴趣区域ROI;

S2. 对双时相感兴趣区域ROI采用OD-YOLO目标检测模型进行定位检测,生成包括候选目标边界框的目标区域图像;

S3. 将候选目标边界框采用视觉大模型SAM的GE-SAM模型进行分割,生成分割掩模;

S4. 对分割掩模进行匹配和交并比计算,通过预设的阈值来判断,最终输出对小尺寸目标的变化检测结果。

[0012] 如上所述的方面和任一可能的实现方式,进一步提供一种实现方式,步骤S1具体包括:

S11. 采用骨干网络对所述遥感影像进行分层特征提取,得到不同尺度特征图;

S12. 采用骨干网络中的特征融合模块对尺度特征图中的低层级特征图和高层级特征图进行深度融合,生成融合特征图;

S13. 将融合后的特征图进行旋转目标检测与坐标映射,得到包括大尺度目标场景的双时相感兴趣区域ROI。

[0013] 如上所述的方面和任一可能的实现方式,进一步提供一种实现方式,步骤S2包括:

S21. 对双时相感兴趣区域ROI的不同场景类别动态设置滑动窗口的尺寸;

S22. 采用滑动窗口在双时相感兴趣区域ROI上按设定步长进行滑动,生成两个局部切片图像集;

S23. 将两个局部切片图像集中的每一个局部切片图像分别输入到两个OD-YOLO网络中进行处理,得到两个包括候选目标边界框的目标区域图像。

[0014] 如上所述的方面和任一可能的实现方式,进一步提供一种实现方式,所述S3包括:

S31. 对目标区域图像的候选目标边界框提出其几何中心点;

S32. 对 t_1 或 t_2 时刻的目标区域图像经滑动窗口处理后得到的图像切片采用图像编码器进行处理,并提取高维图像特征嵌入;

S33. 将几何中心点映射为点提示嵌入向量;

S34. 采用交叉注意力机制融合高维图像特征嵌入与点提示嵌入向量,根据点提示的位置信息,从高维图像特征中分割出与该点提示语义相一致的连通区域;

S35. 将得到的所有连通区域作为分割掩模。

[0015] 如上所述的方面和任一可能的实现方式,进一步提供一种实现方式,所述S4包括:

S41. 对于分割掩模的前一时相中的每一个掩模,遍历后一时相中的所有掩模,并

计算前一时相中的每一个掩模与后一时相中的所有掩模之间的交并比；

S42.根据交并比与预先设置的阈值的关系来判断目标的变化。

[0016] 如上所述的方面和任一可能的实现方式,进一步提供一种实现方式,所述S42具体包括:对于前一时相的某个掩模,如果在后一时相中找不到任何一个掩模使得所计算的交并比大于等于预先设置的阈值,则判定该目标消失;或者,对于后一时相的某个掩模,如果在前一时相中找不到任何一个掩模使得计算得到的交并比大于等于预先设置的阈值,则判定该目标为新出现。

[0017] 如上所述的方面和任一可能的实现方式,进一步提供一种实现方式,所述交并比由两个掩模区域相交部分的面积绝对值除以该两个掩模区域合并部分的总面积绝对值得到。

[0018] 本发明还提供了一种基于视觉大模型的目标级变化检测装置,所述装置用于实现所述的方法,包括:

提取模块,用于针对大范围遥感影像,采用RD-YOLO旋转目标检测网络进行精确提取得到包括大尺度目标场景的双时相感兴趣区域ROI;

定位检测模块,用于对双时相感兴趣区域ROI采用OD-YOLO目标检测模型进行定位检测,生成包括候选目标边界框的目标区域图像;

分割模块,用于将候选目标边界框采用视觉大模型SAM的GE-SAM模型进行分割,生成分割掩模;

计算模块,用于对分割掩模进行匹配和交并比计算,通过预设的阈值来判断,最终输出对小尺寸目标的变化检测结果。

[0019] 本发明提供了一种计算机存储介质,所述介质上存储有计算机程序,所述计算机程序被处理器执行实现所述的方法。

[0020] 本发明提供了一种电子设备,所述电子设备包括:

存储器,存储有可执行指令;

处理器,所述处理器运行所述存储器中的所述可执行指令,以实现所述的方法。

[0021] 本发明的有益效果

本发明的基于视觉大模型的目标级变化检测方法,所述方法包括:针对大范围遥感影像,采用RD-YOLO旋转目标检测网络进行精确提取得到包括大尺度目标场景的双时相感兴趣区域ROI;对双时相感兴趣区域ROI采用OD-YOLO目标检测模型进行定位检测,生成包括候选目标边界框的目标区域图像;将候选目标边界框采用视觉大模型SAM的GE-SAM模型进行分割,生成分割掩模;对分割掩模进行匹配和交并比计算,通过预设的阈值来判断,最终输出对小尺寸目标的变化检测结果。本方法具有如下优点:

(1) 无需大量特定场景标注数据:有深度学习方法通常依赖大规模人工标注数据集进行训练,迁移到新场景时需要重新训练或微调。本发明利用预训练视觉大模型的跨任务泛化能力,在缺乏标注样本的条件下仍能保持较好的检测性能,显著降低了数据依赖。

[0022] (2) 实现目标级变化检测:统深度学习方法大多停留在像素级变化检测,本发明通过在视觉大模型特征空间中提取目标实例特征,并进行多时相目标对齐与差异建模,实现目标级变化判别,更符合实际应用需求。

[0023] (3) 跨场景适应性强:现有方法在不同传感器、分辨率或地物类型下泛化能力不

足。本发明利用视觉大模型的全局上下文建模能力,有效提升了跨场景、跨数据源的鲁棒性。

[0024] (4)检测精度与效率提升:过大模型语义特征提取与目标级差异分析,本发明能够减少噪声干扰和虚警,同时避免复杂的多源融合与重新训练过程,在保证精度的同时提高检测效率。

[0025] (5)应用范围更广:本发明方法不仅适用于高分辨率遥感影像,还可扩展至视频监控、多源遥感(光学+SAR)等场景,具有更强的通用性和推广价值。

附图说明

[0026] 图1是本发明的方法流程示意图;
图2是本发明的RD-YOLO网络结构示意图;
图3是本发明的GE-SAM网络结构示意图。

具体实施方式

[0027] 为了更好的理解本发明的技术方案,本发明内容包括但不限于下文中的具体实施方式,相似的技术和方法都应该视为本发明保护的范畴之内。为使本发明要解决的技术问题、技术方案和优点更加清楚,下面将结合附图及具体实施例进行详细描述。

[0028] 应当明确,本发明所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其它实施例,都属于本发明保护的范围。

[0029] 在本发明实施例中使用的术语是仅仅出于描述特定实施例的目的,而非旨在限制本发明。在本发明实施例和所附权利要求书中所使用的单数形式的“一种”、“所述”和“该”也旨在包括多数形式,除非上下文清楚地表示其他含义。

[0030] 本发明公开了一种基于视觉大模型的目标级变化检测方法,所述方法包括:

S1.针对大范围遥感影像,采用RD-YOLO旋转目标检测网络进行精确提取得到包括大尺度目标场景的双时相感兴趣区域ROI;

S2.对双时相感兴趣区域ROI采用OD-YOLO目标检测模型进行定位检测,生成包括候选目标边界框的目标区域图像;

S3.将候选目标边界框采用视觉大模型SAM的GE-SAM模型进行分割,生成分割掩模;

S4.对分割掩模进行匹配和交并比计算,通过预设的阈值来判断,最终输出对小尺寸目标的变化检测结果。

[0031] 具体来说,如图1所示,本发明的基于视觉大模型的目标级变化检测方法,旨在解决传统变化检测方法中存在的像素级“伪变化”干扰、目标漏检误检以及对旋转不规则目标区域适应性差等问题。本发明通过引入一种多阶段、由粗到精的检测策略,该策略有机结合了为不同尺度目标专门优化的深度学习模型与视觉基础模型,实现了对遥感影像中关键目标的高精度、高鲁棒性的变化检测。

[0032] 本发明的整体流程包括以下步骤:

第一步.面向特定场景的感兴趣区域(Region of Interest, ROI)精确提取:通过

光学卫星或航空遥感平台获取的高分辨率光学影像,影像内容通常覆盖数十平方公里,包含复杂的背景信息以及本发明所关注的大尺度目标场景,如机场跑道区域、大型港口码头区域。针对大尺度目标场景,本发明对相邻两个 t_1, t_2 时刻的图像中的 t_2 后一时刻的图像,即 t_2 时刻图像采用用于大尺度目标检测优化的、名为RD-YOLO的旋转目标检测网络,通过融合该图像中的高层语义特征与低层细节特征,精确提取出如机场、港口等大尺度、方向性强的感兴趣区域ROI。大范围影像中背景像素占据绝大部分比例,若直接全图处理,计算量极其巨大且容易引入大量背景噪声。本发明采用RD-YOLO网络预先提取包含关键设施的ROI,可以过滤掉无关背景区域,显著降低后续计算量;将后续处理聚焦于特定语义场景,减少虚警率,提高检测的针对性和效率。对 t_2 时刻图像进行上述处理后得到的感兴趣区域ROI及其坐标,随后将该坐标映射至 t_1 时刻图像进行对应区域裁切,得到关于 t_1 时刻的感兴趣区域ROI。 t_1 时刻的感兴趣区域ROI和 t_2 时刻的感兴趣区域ROI合在一起称为双时相ROI,即对同一地理区域生成的两个不同时刻感兴趣区域。两个时刻 t_1, t_2 之间可间隔不同的时间段,如一分钟,一年或数年均可。

[0033] 第二步.基于滑动窗口与目标检测器的候选目标初步筛查:对双时相ROI,采用一个为小目标检测优化的、名为OD-YOLO的目标检测模型进行裁切,通过增强高分辨率特征图的利用,对双时相ROI中的飞机、舰船等小型目标进行快速、精确的定位,生成具有候选目标的边界框区域的 t_1 时刻目标图像和 t_2 时刻目标区域图像。小尺寸目标是指在双时相的ROI的图像中像素占比极小,例如相对原图比例低于1%的目标。对小目标进行定位的原因及效果:此类目标在经过多层卷积下采样后,其特征信息极易丢失,采用传统模型易漏检。本发明采用裁切,直接将小目标进行专门进行定位是为了在变化检测前先确立“关注对象”,避免直接比对像素带来的误差,显著提升对细微变化的捕捉能力。

[0034] 第三步.基于视觉大模型的实例分割与精确掩模生成:将OD-YOLO检测到的 t_1 时刻目标区域图像和 t_2 时刻目标区域图像中的候选目标边界框的几何中心点作为精确提示,均输入到一个名为GE-SAM的分割模块中。该模块利用视觉大模型SAM强大的分割能力,为每个候选目标边界框生成高精度的两个实例分割掩模。

[0035] 第四步.基于掩模交并比IoU的目标级变化判定:对两个实例分割掩模,进行匹配和交并比计算,通过预设的IoU阈值来判定小型目标是否发生“变化”(如出现、消失、位移等),前面步骤一至三是对两个不同时刻单张图进行目标检测,第四步是对两张图上检测出的目标进行变化检测(判断是否变化),最终输出目标级的变化检测结果。

[0036] 下面来具体描述前述步骤,第一步,面向特定场景的感兴趣区域ROI精确提取:

在处理高分辨率、大范围遥感影像时,直接进行全图变化检测会消耗巨大的计算资源,且背景区域的复杂变化(如光照、季节变化)会严重干扰目标检测。本步骤旨在从原始大范围遥感影像中,高效且精确地检测出如机场跑道、港口区域等大尺度、具有旋转特性的感兴趣区域。目的是检测飞机舰船等小尺寸目标的变化,这里的大尺度是相对飞机、舰船可能出现的场景,先检测出这些小尺寸目标场景的位置,这类目标兼具丰富的细节纹理(低层特征)和明确的整体结构(高层特征)。为兼顾这两点,本发明提出一种名为RD-YOLO (Region Detection YOLO)的网络,来生成双时相ROI。

[0037] RD-YOLO网络包括骨干网络BACKBONE,其中设置有Transformer特征融合模块,通过其中的Transformer特征融合模块对骨干网络BACKBONE提取的低层级和高层级特征图进

行深度融合,生成一个既包含丰富细节又具备全局视野的融合特征图,再基于此进行旋转框检测。其结构如图2所示,本步骤的具体实现包括:

对t2时刻(后时相)图像,以其影像为基准进行ROI提取。首先在t2时刻影像上检测并锁定感兴趣区域及其坐标作为区域检测结果,将检测结果中的感兴趣区域作为t2时刻感兴趣区域,随后将检测结果中的该坐标映射至t1时刻(前时相)影像进行对应区域裁剪,得到t1时刻感兴趣区域,该策略旨在以最新时刻的目标分布为导向,通过对比前后不同时刻但是同一空间位置的历史状态,实现对目标动态变化的精准观测。

[0038] 为达到上述目的,对t2时刻图像进行粗粒度缩放,将t2时刻图像作为原始图像,其长宽均缩小至原来的1/100。这一步旨在极大压缩空间分辨率,使网络能够在极小的计算代价下进行机场、港口等区域提取,之后在下采样后的图像周围进行填充操作,将画布尺寸扩大为原来的1.2倍。这种1.2倍的扩充是为了旋转目标检测 (Oriented Object Detection) 做准备。当后续对图像中的目标(如倾斜的跑道)进行旋转或裁剪时,边缘填充可以防止目标超出边界而导致特征丢失,确保留出足够的旋转余量。

[0039] 分层特征提取:将填充后的图像采用骨干网络进行处理,首先通过其一个标准的CNN骨干网络(如CSPDarknet),提取出特征图C1,之后以该特征图C1为基础,依次在前一特征图上提取出不同尺度的特征图C2-C5,其中,C1、C2特征图为浅层特征,分辨率较高,主要包含目标的边缘、纹理、颜色等几何细节信息;C3、C4特征图为中层特征,包含目标的局部部件信息;C5特征图为深层特征,分辨率为原图的1/32,但感受野最大,富含丰富的高级语义信息,能反映目标的整体类别属性。

[0040] Transformer特征融合模块:此模块是RD-YOLO网络的关键部分。它通过自注意力Self-Attention机制捕捉每个特征尺度内的全局依赖,再通过交叉注意力Cross-Attention机制实现跨尺度信息的交互与对齐,具体包括如下步骤:

浅层特征编码:网络前端通过卷积层Conv提取得到高分辨率的浅层特征图,记为C2',其尺寸为输入尺寸的1/4,即 $\frac{h}{4} \times \frac{w}{4} \times C_2$,该特征图经过调整维度Reshape操作后输入Transformer编码器,通过自注意力机制对局部纹理和边缘细节进行编码,生成富含细节信息的特征序列Key和Value,简称K,V。

[0041] 深层特征解码网络深层继续通过C3K2模块和卷积层下采样,得到低分辨率的深层特征图,记为C5',其尺寸为输入尺寸的1/32,即 $\frac{h}{32} \times \frac{w}{32} \times C_5$ 。该特征图经过调整维度Reshape操作后输入Transformer 解码器作为查询项Query,简称Q。

[0042] 上述Q,K,V的自注意力机制的计算表示为:

$$Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

经过上述线性变换得到三个向量:查询Q、键K、值V。通过计算Q和K的点积来衡量序列中任意两个位置的相似度,除以 $\sqrt{d_k}$ 进行缩放以稳定梯度, d_k 为键向量的维度。然后通过softmax函数将相似度得分归一化为权重,最后将权重应用于值V向量,得到加权后的特征表示。这使得网络能够聚焦于对当前任务最重要的特征区域。

[0043] 特征融合:在TransformerDecoder解码器中,深层语义特征Query与编码后的浅层细节特征Key/Value进行交叉注意力计算Cross-Attention,这一过程使得网络能够利用深层的语义信息去“索引”浅层的精确位置细节,融合后的特征经过YOLOv11原生的C3K2模块进一步精炼,输出最终的融合特征图。

[0044] 旋转目标检测与坐标映射:将融合特征图送入旋转框检测头Head。该检测头采用YOLOv11 OBB Head,分别预测目标的类别分数和旋转边界框参数。

[0045] 损失函数优化:网络的训练目标是最小化复合损失函数 L ,其公式如下:

$$L = \lambda_{cls} L_{cls} + \lambda_{reg} L_{reg}$$

其中, L_{cls} 为分类损失(如Varifocal Loss),用于判断特征点处是否存在感兴趣的大尺度目标(如机场); L_{reg} 为旋转框回归损失(如Rotated CIoU Loss),用于惩罚预测框参数与真实标注框(Ground Truth)之间的差异。预测框参数包括 (x, y, w, h, θ) , x, y 分别代表预测框中心点坐标、 w, h 表示预测框的宽度和高度, θ 表示预测框中心点的旋转角度。

[0046] 结果上采样与输出:如图2所示,检测头Head输出的检测结果是基于缩小后的特征图尺度。为了得到区域目标检测结果,即最终的ROI区域,需进行坐标上采样Upsampling,将预测出的旋转框坐标按比例映射回预处理后的图像尺度 $1.2H \times 1.2W$,最终裁剪出包含大尺度目标的区域检测结果,作为 t_2 时刻感兴趣区域ROI。

[0047] 通过此步骤,可以从 t_1 时刻和 t_2 时刻的原始影像中裁剪出对齐的、包含最少背景噪声的双时相ROI图像,即 t_1 时刻感兴趣区域ROI和 t_2 时刻感兴趣区域ROI。这种基于旋转框的精确裁剪,不仅大大减少了后续处理的数据量,也为后续的目标检测提供了规范化的输入,显著提升了检测的准确性。

[0048] 第二步,基于滑动窗口与目标检测器的候选目标初步筛查:

在第一步提取出的双时相ROI图像内部,检测出其中的高效检测飞机、舰船、车辆等小尺寸目标。此处所述的小尺寸目标,是指在整幅ROI图像中面积占比低于1%的离散目标。检测这些目标的原因及效果:由于ROI图像(如整个机场)依然具有较大的分辨率,若直接将ROI缩放至网络输入尺寸,则上述小目标将退化为噪点或完全消失。因此,必须采用高分辨率切片的方式进行检测,以保留目标的形状和纹理特征,确保后续变化检测的召回率。

[0049] 为此,在本步骤本发明提出一种名为OD-YOLO(Object Detection YOLO)的检测器。OD-YOLO在网络架构上与前述步骤一中的RD-YOLO保持一致(即同样采用图2所示的Backbone+Head结构),其核心区别在于输入数据的构造方式采用了滑动窗口策略。本步骤的具体实现包括:

1. 基于场景自适应的滑动窗口切片生成:针对高分辨率的ROI图像,本发明采用一种场景自适应的切片策略,即根据步骤一中识别出的ROI中的场景类别动态设定滑动窗口的尺寸,包括:

针对机场场景:考虑到飞机目标排列相对紧凑且尺寸中等,设置滑动窗口尺寸为 1024×1024 像素,以平衡检测视野与小目标细节。

[0050] 针对港口场景:考虑到舰船及集装箱堆场尺度较大且分布跨度广,设置滑动窗口

尺寸为 2048×2048 像素,以确保完整覆盖大尺度舰船目标并保留足够的上下文环境。

[0051] 重叠与生成:无论采用上述哪一种场景所设置的尺寸,为了避免位于窗口边缘的目标被截断导致漏检,均设置相邻窗口之间存在20%的重叠率。通过在两个ROI图像上按设定步长滑动,分别生成一系列覆盖整个ROI且尺度适宜的两个高分辨率局部切片图像集。

[0052] 2.基于复用架构的特征提取与检测:将两个局部切片图像集中的每一个局部切片图像分别输入到两个OD-YOLO网络中。

[0053] 特征提取:图像经过与步骤1相同的CNN骨干网络提取浅层特征C2”和深层特征C5”。

[0054] Transformer特征交互:利用图2所示的Transformer模块,以切片图像内的深层语义特征为Query,检索浅层细节特征(Key/Value),得到融合后的特征,增强对小目标微弱特征的捕捉能力。

[0055] 旋转框预测:融合后的特征送入旋转检测头,输出该切片图像内所有候选目标的类别概率和旋转边界框坐标。

[0056] 损失函数:训练阶段同样采用由分类损失 L_{cls} 和回归损失 L_{reg} 构成的复合损失函数 L ,但训练数据更换为包含小目标的两个局部切片图像集。

[0057] 3.坐标还原与冗余去除NMS:由于使用了重叠的滑动窗口,同一个目标可能在多个相邻切片图像的检测结果中出现。因此,需要进行后处理:

全局坐标映射:首先根据滑动窗口在ROI图像中的偏移量,将切片内的局部坐标转换为ROI图像坐标系下的全局坐标;

非极大值抑制NMS:对所有切片产生的全局候选框进行汇总,应用非极大值抑制算法,当两个候选框的交并比IoU高于设定阈值且类别相同时,保留置信度最高的候选框,剔除冗余框。

[0058] 采用此步骤,利用滑动窗口策略结合高精度的Transformer检测架构,能够高效地在感兴趣区域内生成高召回率的候选目标集、得到包括候选目标边界框的 t_1 时刻目标区域图像和 t_2 时刻目标区域图像,为下一阶段的精细化分割提供了可靠的输入,有效避免了因目标尺寸过小而导致的漏检问题。

[0059] 第三步,基于视觉大模型的实例分割与精确掩模生成:

获得候选目标的边界框后,为实现精确的变化判定,需要获取目标的像素级轮廓。本步骤引入一个GE-SAM的模块,其利用视觉大模型SAM完成此任务。如图3所示,该模块的核心策略是不对SAM模型本身进行微调,而是利用前一步骤OD-YOLO生成的边界框几何信息作为强先验提示,激发SAM的零样本分割能力。

[0060] 本步骤对 t_1 时刻和 t_2 时刻的目标区域图像并行执行相同的处理流程,具体实现如下:

1.提示构建(Prompt Engineering):对于步骤二中OD-YOLO输出的每一个候选目标边界框 $B_{box} = (x, y, w, h)$,提取其几何中心点坐标 $P_{center} = (x + w/2, y + h/2)$ 。该中心点将被用作SAM的单点提示Single-Point Prompt,指示目标在图像中的主体位置。

[0061] 2.视觉大模型推理:将包含小尺寸目标的图像切片输入至标准的SAM模型中,如图3中间部分的框图所示,SAM的处理流程包括:

图像编码:对 t_1 时刻的目标区域图像经滑动窗口处理后得到的图像切片输入至图像编码器Image Encoder,通过ViT架构提取高维图像特征嵌入;

提示编码:将前述提取的几何中心点坐标 P_{center} 输入至点提示编码器Points Encoder,被映射为点提示嵌入向量;

掩模解码:掩模解码器Mask Decoder利用交叉注意力机制,融合高维图像特征嵌入与点提示嵌入向量,根据点提示的位置信息,从高维图像特征中分割出与该点提示语义一致的连通区域。

[0062] 3.掩模生成:将所有的图像切片进行上述操作,得到的所有连通区域作为最终关于目标的输出高精度二值化实例分割掩模Instance Segmentation Mask。对 t_1 时刻和 t_2 时刻的所有图像切片重复上述过程,最终分别得到两个时刻的精确目标掩模集合 $\{M_{t1}\}, \{M_{t2}\}$ 。

[0063] 通过此步骤,目标的表示形式从粗糙的矩形框升级为精细的像素级轮廓,有效剔除矩形框内非目标背景像素的干扰,为后续的高精度比对奠定基础。

[0064] 第四步,基于掩模交并比IoU的目标级变化判定:

本步骤对前后时相的精确目标掩模进行比较,以判定目标变化。此步骤的核心是采用交并比(Intersection over Union, IoU)作为衡量目标相似性的度量标准,本步骤的具体实现如下:

交并比计算:输入前后两个时相的精确目标掩模集合 $\{M_{t1,i}\}_{i=1}^{N_1}$ 和 $\{M_{t2,j}\}_{j=1}^{N_2}$, N_1 , N_2 分别为集中的掩模个数,对于 t_1 时相中的每一个掩模 $M_{t1,i}$,遍历 t_2 时相中的所有掩模 $M_{t2,j}$,计算它们的交并比为:

$$IoU(M_A, M_B) = \frac{|M_A \cap M_B|}{|M_A \cup M_B|}$$

$|M_A \cap M_B|$ 代表两个掩模区域相交(重叠)部分的面积(像素数量), $|M_A \cup M_B|$ 代表两个掩模区域合并(并集)部分的总面积。IoU的值域为 $[0, 1]$,值越大表示两个区域的重合度越高。

[0065] 变化判定:设定一个IoU阈值 τ 。对于 t_1 时相的某个目标掩模 $M_{t1,i}$,如果在 t_2 时相中找不到任何一个掩模 $M_{t2,j}$ 使得它们的IoU大于等于阈值 τ ,则判定该目标消失,表达式如下:

$\forall j \in [1, N_2], IoU(M_{t1,i}, M_{t2,j}) < \tau \Rightarrow \text{Disappeared}$,其中, j 为变量,取值为 $1, 2, \dots, N_2$ 。

[0066] 反之,对于 t_2 时刻的某个目标掩模 $M_{t2,j}$,如果在 t_1 时相中找不到任何一个掩模 $M_{t1,i}$ 使得它们的IoU大于等于阈值 τ ,则判定该目标为新出现。

[0067] $\forall i \in [1, N_1], IoU(M_{t1,i}, M_{t2,j}) < \tau \Rightarrow \text{Appeared}$,其中, i 为变量,取值为 $1, 2, \dots, N_1$ 。

[0068] 最终的变化检测结果,以小尺寸目标为单位进行标注,清晰地指出哪些小尺寸目标是新增的、哪些些小尺寸目标是消失的。利用对精确掩模的量化比对,能够客观、准确地判定每个些小尺寸目标的变化状态,最终生成清晰、可靠的些小尺寸目标级变化图谱,有效避免了传统像素级变化检测方法所产生的“伪变化”问题。

[0069] 综上所述,本发明通过RD-YOLO、OD-YOLO、GE-SAM和IoU判定四个精心设计的串联步骤,构成了一个完整、高效且精确的目标级变化检测技术方案,在理论和实践上均具有显著的创新性和实用价值。

[0070] 作为本发明公开的实施例,本发明还提供了一种基于视觉大模型的目标级变化检测装置,所述装置用于实现所述的方法,包括:

提取模块,用于针对大范围遥感影像,采用RD-YOLO旋转目标检测网络进行精确提取得到包括大尺度目标场景的双时相感兴趣区域ROI;

定位检测模块,用于对双时相感兴趣区域ROI采用OD-YOLO目标检测模型进行定位检测,生成包括候选目标边界框的目标区域图像;

分割模块,用于将候选目标边界框采用视觉大模型SAM的GE-SAM模型进行分割,生成分割掩模;

计算模块,用于对分割掩模进行匹配和交并比计算,通过预设的阈值来判断,最终输出对小尺寸目标的变化检测结果。

[0071] 作为本发明公开的实施例,本发明还提供了一种计算机存储介质,所述介质上存储有计算机程序,所述计算机程序被处理器执行实现所述的方法。

[0072] 作为本发明公开的实施例,本发明还提供了一种电子设备,所述电子设备包括:

存储器,存储有可执行指令;

处理器,所述处理器运行所述存储器中的所述可执行指令,以实现所述的方法。

[0073] 上述说明示出并描述了本发明的若干优选实施例,但如前所述,应当理解本发明并非局限于本文所披露的形式,不应看作是对其他实施例的排除,而可用于各种其他组合、修改和环境,并能够在本文所述申请构想范围内,通过上述教导或相关领域的技术或知识进行改动。而本领域人员所进行的改动和变化不脱离本发明的精神和范围,则都应在本发明所附权利要求书的保护范围内。

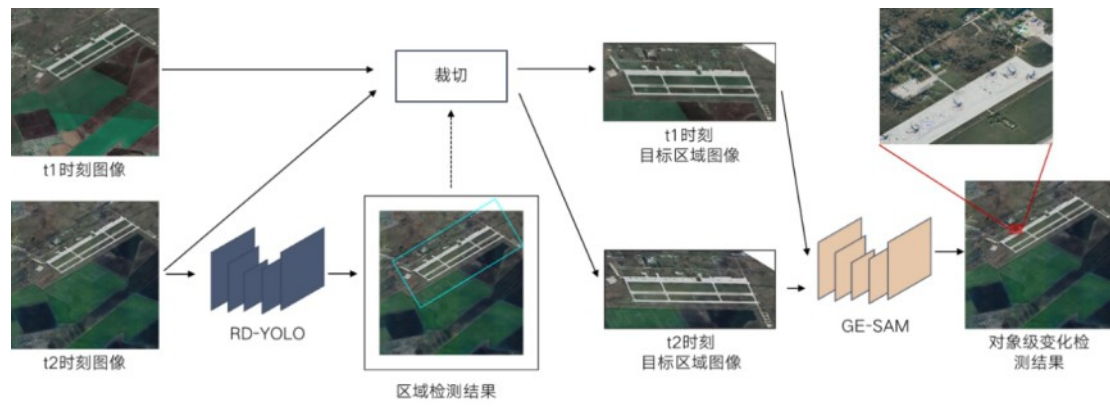


图1

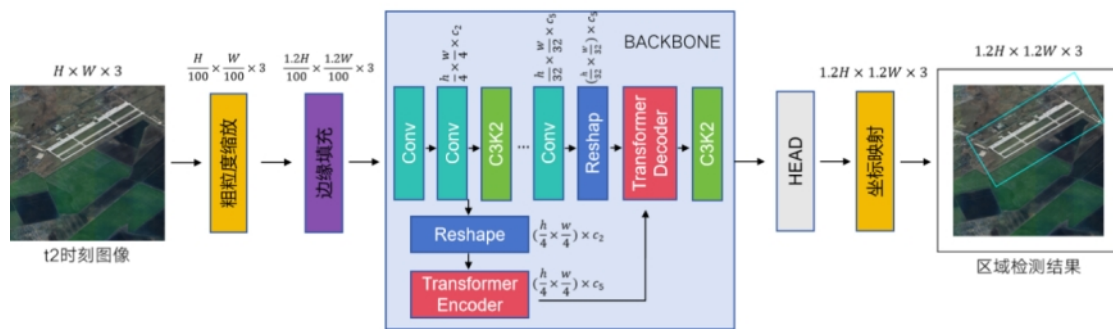


图2

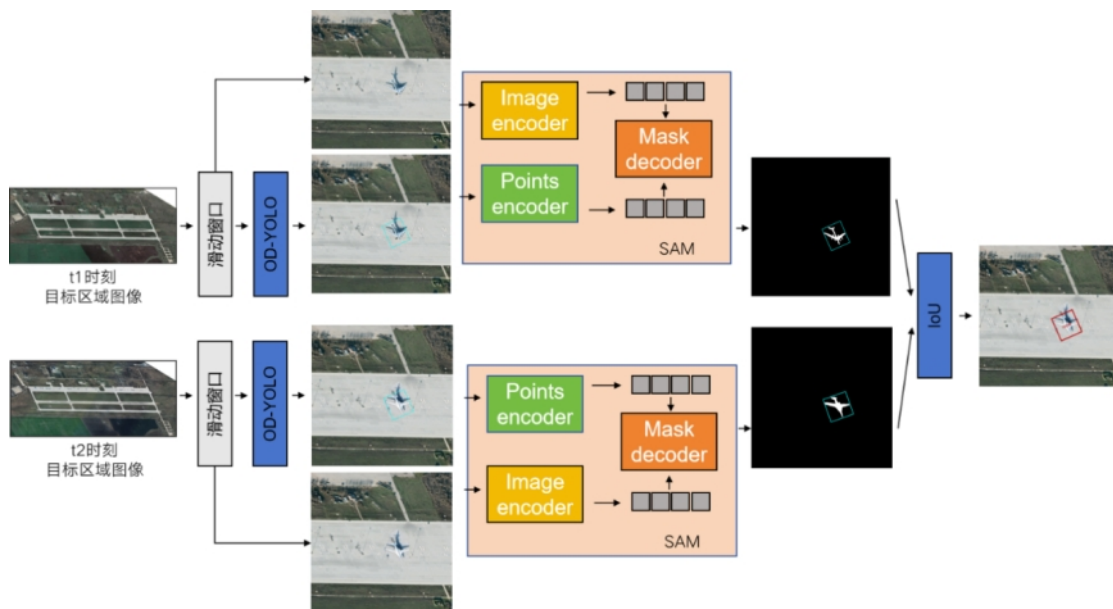


图3