



(12) 发明专利申请

(10) 申请公布号 CN 122067120 A

(43) 申请公布日 2026. 05. 19

(21) 申请号 202610087847.0

G06N 3/045 (2023.01)

(22) 申请日 2026.01.22

G06N 3/0464 (2023.01)

(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 20/13 (2022.01)

G06V 10/764 (2022.01)

G06V 10/52 (2022.01)

G06V 10/80 (2022.01)

G06V 10/82 (2022.01)

G06T 3/4038 (2024.01)

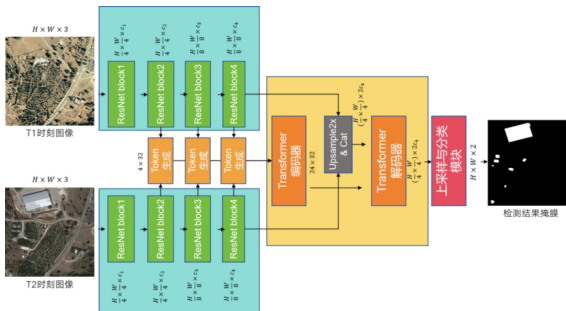
权利要求书2页 说明书9页 附图2页

(54) 发明名称

一种融合多尺度特征的Transformer地物变化检测方法、装置、设备及介质

(57) 摘要

本发明涉及一种融合多尺度特征的Transformer地物变化检测方法、装置、设备及介质,包括:对两个时相的遥感影像进行提取,得到不同层级的多尺度特征信息;将多尺度特征信息压缩并符号化为一系列语义令牌;将所有语义令牌进行编码得到编码后的令牌信息;将编码后的令牌信息进行解码拼接,得到高信息含量的差异特征图;对差异特征图进行处理,得到最终的变化检测结果。通过引入基于注意力的多层语义令牌生成机制,来解决多尺度特征向序列信息高效转化的问题。本发明通过构建统一的Transformer编码器,来解决贯穿多尺度特征的全局时空上下文建模问题;并通过Transformer解码器结构,来解决差异特征生成机制简单的问题,从而全面提升变化检测的准确性、鲁棒性和精细化水平。



1.一种融合多尺度特征的Transformer地物变化检测方法,其特征在于,所述方法包括:

- S1.对两个时相的遥感影像进行提取,得到不同层级的多尺度特征信息;
- S2.将所述多尺度特征信息压缩并符号化为一系列语义令牌;
- S3.将所有语义令牌进行编码得到编码后的令牌信息;
- S4.将编码后的令牌信息进行解码拼接,得到高信息含量的差异特征图;
- S5.对差异特征图进行处理,得到最终的变化检测结果。

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:

对经过预处理的同一地理区域的第一双时相遥感影像和第二双时相遥感影像分别采用卷积神经网络ResNet进行提取,从卷积神经网络ResNet的不同网络深度中提取得到关于第一时相的遥感影像或第二时相的遥感影像的浅层特征图、中层特征图和深层特征图。

3.根据权利要求2所述的方法,其特征在于,步骤S2包括:

S21.将第一双时相遥感影像或第二双时相遥感影像的浅层特征图和中层特征图的空间分辨率上采样,使得空间分辨率与深层特征图相同,得到尺度统一的三层特征图;

S22.对三层特征图采用第一卷积层将其通道维度投影到一个预设的令牌维度,得到投影后的特征;

S23.采用第二卷积层为三层特征图中的每一层级的特征图生成N个注意力图,得到注意力权重图,其中正整数N为每层生成的令牌数量;

S24.将注意力权重图进行归一化处理,得到归一化后的注意力图;

S25.将归一化后的注意力图作为空间权重,对投影后的特征进行加权求和,得到长度为N的第一语义令牌或第二语义令牌;

S26.对第一语义令牌或第二语义令牌按序列维度进行拼接,得到第一令牌序列或第二令牌序列。

4.根据权利要求3所述的方法,其特征在于,所述S3包括:

S31.将第一令牌序列和第二令牌序列进行拼接,得到令牌长序列;

S32.将所述令牌长序列采用多头自注意力机制进行线性变换,得到查询、键和值三个矩阵;

S33.对所述三个矩阵进行计算注意得分并加权求和,得到编码令牌。

5.根据权利要求4所述的方法,其特征在于,所述S4包括:

S41.将第一时相的深层特征图和第二时相的深层特征图进行拼接,得到两时相的高级语义信息特征图;

S42.将所述编码令牌作为编码器记忆;

S43.将深层特征图作为查询矩阵、将编码令牌作为键矩阵和值矩阵输入解码器进行解码进行处理,得到差异特征图。

6.根据权利要求5所述的方法,其特征在于,所述S5具体包括:S51.对差异特征图进行融合与降维处理,得到整合后的差异特征图;

S52.对整合后的差异特征图的分辨率放大,得到全分辨率的差异特征图;

S53.将全分辨率的差异特征图进行分类预测,得到最终的变化检测图。

7.根据权利要求6所述的方法,其特征在于,所述分类检测具体包括:将全分辨率的差

异特征图进行分类,得到不同类别的特征图;

对不同类别的特征图中的每个像素判断属于各个类别的概率,根据所述概率值来得到最终的变化检测图。

8.一种融合多尺度特征的Transformer地物变化检测装置,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,用于对两个时相的遥感影像进行提取,得到不同层级的多尺度特征信息;

压缩模块,用于将所述多尺度特征信息压缩并符号化为一系列语义令牌;

编码模块,用于将所有语义令牌进行编码得到编码后的令牌信息;

拼接模块,用于将编码后的令牌信息进行解码拼接,得到高信息含量的差异特征图;

处理模块,用于对差异特征图进行处理,得到最终的变化检测结果。

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。

一种融合多尺度特征的Transformer地物变化检测方法、装置、设备及介质

技术领域

[0001] 本发明属于计算机视觉与人工智能技术领域,具体涉及一种融合多尺度特征的Transformer地物变化检测方法、装置、设备及介质。

背景技术

[0002] 变化检测(Change Detection, CD)是遥感影像智能解译领域的一项基本且关键的任务。其目的是通过分析同一地理区域在不同时间点获取的多时相遥感影像,识别并定位地表发生变化的区域。该技术在国土资源规划、城市扩张监测、自然灾害评估、环境保护、农业估产等众多领域具有极其重要的应用价值。目前,针对遥感影像的地物变化检测技术,根据其核心原理与发展阶段,主要可归纳为以下几类技术方案:

首先,是基于像素代数运算的传统方法。该类方法通过对两个时相的影像进行直接的像素级操作,如图像差分法、图像比值法或变化向量分析法,来获取像素值的差异信息。随后,通过设定一个或多个阈值来分割图像,从而区分变化区域与未变化区域。此类方案的优点在于原理简单、计算速度快,但其检测结果极易受到影像间辐射差异、传感器噪声、太阳高度角变化及季节性因素的干扰,导致变化检测结果中存在大量的伪变化信息,鲁棒性较差。

[0003] 其次,是基于传统特征提取与机器学习分类的方法。为克服像素级运算的局限性,有研究提出先从影像中提取人工设计的特征,例如基于主成分分析PCA的变换特征、基于灰度共生矩阵GLCM的纹理特征,或是基于尺度不变特征变换SIFT的关键点特征。然后,将这些特征输入到传统的机器学习分类器中,如支持向量机SVM、决策树或随机森林,通过训练分类器来识别变化地物。这类方法因利用了更丰富的图像信息,在一定程度上提升了抗干扰能力,但其性能高度依赖于人工设计特征的有效性,特征的泛化能力有限,且在面对地物类型复杂、变化形式多样的场景时,其检测精度和自动化水平均难以满足实际应用需求。

[0004] 近年来,基于深度学习的方法已成为变化检测领域的主流技术路线。其中,以卷积神经网络CNN为基础的模型占据主导地位。例如,有研究采用孪生网络Siamese Network结构,通过两个权值共享的CNN分支分别提取双时相影像的深度特征,再对特征进行融合比对以识别变化;亦有研究借鉴U-Net等编码器-解码器架构,通过跳跃连接融合多尺度特征,以实现像素级的精细分割。近期,部分研究开始引入Transformer架构,利用其自注意力机制来捕获特征间的长距离依赖关系,以期提升对全局和复杂场景变化的理解能力。尽管这些深度学习方法显著提升了变化检测的性能,但它们通常存在以下不足:一是多数模型在特征融合时未能充分挖掘不同尺度特征间的深层关联,易忽略对检测结果至关重要的细节信息;二是卷积网络的固有局部性限制了其对大范围、结构化变化的建模能力,而现有Transformer的应用方式也未能与多尺度特征进行有效协同。

[0005] 当前的地物变化检测技术仍存在以下亟待解决的技术问题:

1. 多尺度特征信息向全局上下文模型转化的机制缺失:现有技术虽然能够通过

CNN提取多尺度特征,但在将其用于后续的全局建模(如Transformer)时,通常采用简单的展平或分块策略。这种方式未能有效保留和提炼每个特征层级的核心语义信息,导致在序列化过程中丢失了大量有价值的特征表达。因此,现有技术缺少一种能将不同尺度的二维特征图高效、无损地转化为具有代表性的一维序列信息的机制,这限制了全局模型捕捉精细变化的能力。

[0006] 2.全局时空上下文建模未能贯穿多尺度特征:基于CNN的模型受限于卷积核的局部感受野,难以有效捕获图像中远距离像素之间的依赖关系。虽然部分研究引入了Transformer来增强全局建模能力,但其通常仅在网络的单一层级(多为深层)应用,未能构建一个能够同时融合浅层空间细节、中层结构信息和深层语义信息的统一时空上下文模型。现有技术无法对来源于不同时相、不同特征层级的全部信息进行一体化的全局关联性分析。

[0007] 3.差异特征生成机制简单,对复杂变化关系表达不足:多数现有方法依赖于对双时相特征进行简单的代数运算(如相减、拼接)来生成差异特征。这种方式是一种浅层的特征比对,难以捕捉地物之间复杂的非线性变化关系,例如“建筑A被绿地B取代”这类涉及语义类别转换的变化。现有技术缺少一种能够根据全局时空上下文来动态生成差异特征的智能机制,导致其对复杂变化的表达能力不足。

发明内容

[0008] 为了克服现有技术存在的上述问题,本发明提供一种融合多尺度特征的Transformer地物变化检测方法、装置、设备及介质,用于解决现有技术中存在的上述问题。

[0009] 一种融合多尺度特征的Transformer地物变化检测方法,所述方法包括:

- S1.对两个时相的遥感影像进行提取,得到不同层级的多尺度特征信息;
- S2.将所述多尺度特征信息压缩并符号化为一系列语义令牌;
- S3.将所有语义令牌进行编码得到编码后的令牌信息;
- S4.将编码后的令牌信息进行解码拼接,得到高信息含量的差异特征图;
- S5.对差异特征图进行处理,得到最终的变化检测结果。

[0010] 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:

对经过预处理的同一地理区域的第一双时相遥感影像和第二双时相遥感影像分别采用卷积神经网络ResNet进行提取,从卷积神经网络ResNet的不同网络深度中提取得到关于第一时相的遥感影像或第二时相的遥感影像的浅层特征图、中层特征图和深层特征图。

[0011] 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.将第一双时相遥感影像或第二双时相遥感影像的浅层特征图和中层特征图的空间分辨率上采样,使得空间分辨率与深层特征图相同,得到尺度统一的三层特征图;

S22.对三层特征图采用第一卷积层将其通道维度投影到一个预设的令牌维度,得到投影后的特征;

S23.采用第二卷积层为三层特征图中的每一层级的特征图生成N个注意力图,得到注意力权重图,其中正整数N为每层生成的令牌数量;

S24.将注意力权重图进行归一化处理,得到归一化后的注意力图;

S25.将归一化后的注意力图作为空间权重,对投影后的特征进行加权求和,得到长度为N的第一语义令牌或第二语义令牌;

S26.对第一语义令牌或第二语义令牌按序列维度进行拼接,得到第一令牌序列或第二令牌序列。

[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,所述S3包括:

S31.将第一令牌序列和第二令牌序列进行拼接,得到令牌长序列;

S32.将所述令牌长序列采用多头自注意力机制进行线性变换,得到查询、键和值三个矩阵;

S33.对所述三个矩阵进行计算注意得分并加权求和,得到编码令牌。

[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,所述S4包括:

S41.将第一时相的深层特征图和第二时相的深层特征图进行拼接,得到两时相的高级语义信息特征图;

S42.将所述编码令牌作为编码器记忆;

S43.将深层特征图作为查询矩阵、将编码令牌作为键矩阵和值矩阵输入解码器进行解码进行处理,得到差异特征图。

[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S5具体包括:S51.对差异特征图进行融合与降维处理,得到整合后的差异特征图;

S52.对整合后的差异特征图的分辨率放大,得到全分辨率的差异特征图;

S53.将全分辨率的差异特征图进行分类预测,得到最终的变化检测图。

[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,所述分类检测具体包括:将全分辨率的差异特征图进行分类,得到不同类别的特征图;

对不同类别的特征图中的每个像素判断属于各个类别的概率,根据所述概率值来得到最终的变化检测图。

[0016] 本发明提供了一种融合多尺度特征的Transformer地物变化检测装置,所述装置用于实现所述的方法,包括:

提取模块,用于对两个时相的遥感影像进行提取,得到不同层级的多尺度特征信息;

压缩模块,用于将所述多尺度特征信息压缩并符号化为一系列语义令牌;

编码模块,用于将所有语义令牌进行编码得到编码后的令牌信息;

拼接模块,用于将编码后的令牌信息进行解码拼接,得到高信息含量的差异特征图;

处理模块,用于对差异特征图进行处理,得到最终的变化检测结果。

[0017] 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。

[0018] 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。

[0019] 本发明的有益效果

本发明的融合多尺度特征的Transformer地物变化检测方法,包括:对两个时相的遥感影像进行提取,得到不同层级的多尺度特征信息;将多尺度特征信息压缩并符号化为一系列语义令牌;将所有语义令牌进行编码得到编码后的令牌信息;将编码后的令牌信息进行解码拼接,得到高信息含量的差异特征图;对差异特征图进行处理,得到最终的变化检测结果。通过引入基于注意力的多层语义令牌生成机制,来解决多尺度特征向序列信息高效转化的问题。本发明通过构建统一的Transformer编码器,来解决贯穿多尺度特征的全局时空上下文建模问题;并通过创新的Transformer解码器结构,来解决差异特征生成机制简单的问题,从而全面提升变化检测的准确性、鲁棒性和精细化水平。本方法具有如下优点:

(1) 在特征转化方式上,本发明的技术方案具有更强的语义表示能力。所述现有技术采用简单的展平或分块方式将单一尺度的特征图转化为序列,此过程可能破坏特征的原始空间邻接关系且未能有效提炼关键信息。本发明提案采用的基于注意力加权的多尺度特征令牌化方法,通过可学习的注意力权重对特征进行加权池化,生成能够概括该尺度核心信息的语义令牌。此方法为一种有监督的信息提炼过程,相较于现有技术的无差别序列化,能够为后续的Transformer模型提供信息密度更高、代表性更强的输入。

[0020] (2) 在上下文建模范围上,本发明的技术方案更能结合不同层级信息。所述现有技术仅在单一深层特征尺度上应用Transformer,因而无法捕获浅层细节特征中的长距离依赖,也无法直接建模不同尺度特征间的交互关系。本发明提案构建的贯穿多尺度与双时相的统一上下文编码架构,将所有时相、所有尺度的信息令牌纳入统一的全局建模框架。此架构使得模型能够学习并整合跨越不同时间点和不同特征层级的关联信息,从而形成对场景变化更为完整的理解。

[0021] (3) 在差异生成机制上,本发明的技术方案采用了一种新的技术路径。所述现有技术在经过Transformer处理后,仍需依赖后续的卷积解码器对特征进行解码,以间接推断变化区域。本发明提案采用的基于Transformer解码器的动态差异特征生成方法,利用交叉注意力机制,将拼接后的双时相特征作为查询,去检索并融合由编码令牌提供的全局变化上下文信息。此机制使得差异特征的生成过程是基于全局信息引导的,能够直接生成语义信息更为丰富的差异表达,相较于现有技术,更能有效处理复杂的语义变化场景。

附图说明

[0022] 图1是本发明的方法流程示意图;

图2是本发明的多层语义令牌生成结构示意图;

图3是本发明的Transformer编码器与解码器模块的内部结构示意图。

具体实施方式

[0023] 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。

[0024] 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。

[0025] 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。

[0026] 本发明公开了一种融合多尺度特征的Transformer地物变化检测方法,所述方法包括:

- S1.对两个时相的遥感影像进行提取,得到不同层级的多尺度特征信息;
- S2.将所述多尺度特征信息压缩并符号化为一系列语义令牌;
- S3.将所有语义令牌进行编码得到编码后的令牌信息;
- S4.将编码后的令牌信息进行解码拼接,得到高信息含量的差异特征图;
- S5.对差异特征图进行处理,得到最终的变化检测结果。

[0027] 具体来说,本发明提出一种融合多尺度特征的Transformer地物变化检测方法。图1中展示了数据的维度变化流程:输入的双时相影像维度为 $H \times W \times 3$; 经过ResNet特征提取后,分别获得不同尺度的特征图(如Block2输出 $\frac{H}{4} \times \frac{W}{4} \times c_2$ 等); 随后通过令牌生成模块(Token Generate)将每层特征压缩为语义令牌,最终生成总计 24×32 的令牌序列送入Transformer编码器; 解码部分将深层特征上采样并拼接后得到的特征序列 ($\frac{H}{4} \times \frac{W}{4} \times 2c_4$), 结合编码器输出进行解码,最终经上采样模块恢复至 $H \times W \times 2$ 的变化检测结果掩膜。本发明首先通过一个权值共享的双分支卷积神经网络从两个时相的遥感影像中提取不同层级的深度特征; 其次,通过一个多层语义令牌生成模块将多尺度的特征信息压缩并符号化为一系列语义令牌; 再次,利用Transformer编码器对两个时相的所有语义令牌进行联合建模,深度挖掘特征间的时空关联性; 然后,将编码后的令牌信息输入至Transformer解码器,对拼接后的深层特征图进行信息增强,从而直接生成高信息含量的差异特征图; 最后,通过上采样与分类模块得到像素级的最终变化检测结果。

[0028] 下面,将详细阐述本发明所包括的具体步骤:

第一步:双时相影像的多尺度特征提取

本步骤旨在从输入的不同时相遥感影像中,提取出能够表征地物在不同抽象层次上的特征。该步骤由图1所示的双分支特征提取网络完成,具体实现包括:

数据输入:输入为两张已经过预处理(例如,大气校正、几何配准、数据归一化等)的同一地理区域的第一时相遥感影像和第二时相遥感影像,记为第一和第二影像 I_{T1} 和 I_{T2} 。同一地理区域是指两幅影像覆盖的地表范围完全一致,且已完成像素级的配准,这是进行变化检测的前提。双时相影像不仅在获取时刻上存在前后差异(如T1为变化前,T2为变化后),其内容可能因季节变化、光照条件、传感器差异等因素呈现辐射差异,而本发明的目标是剔除这些无关干扰,仅检测出感兴趣的地物变化(如建筑物的增加、拆除或道路的修建等)。其数据格式为张量,维度为 $[B, C, H, W]$, 其中B为批处理大小,C为影像通道数(例如RGB影像为3),H和W分别为影像的高度和宽度。

[0029] 特征提取主干网络:本发明采用一个具有Siamese结构的卷积神经网络ResNet作为特征提取的骨干网络。该网络包含两个权值共享的分支,分别独立处理第一和第二影像 I_{T1} 和 I_{T2} 。权值共享的机制确保了对两张影像的特征提取遵循统一的度量标准,有利于后续进行准确的差异比较。令特征提取网络表示为函数 Φ_{CNN} 。

[0030] 多尺度特征获取:在所述主干网络的前向传播过程中,从其不同的网络深度(阶段或层)中提取三个不同尺度的特征图集:

浅层特征 F_{T1}^1, F_{T2}^1 :从ResNet的layer2(Block2)输出,其维度为 $[B, C_2, H/4, W/4]$ 。该层特征分辨率较高,保留了丰富的空间纹理、颜色和边缘等细节信息。

[0031] 中层特征 F_{T1}^2, F_{T2}^2 :从ResNet的layer3(Block3)输出,其维度为 $[B, C_3, H/8, W/8]$ 。该层特征具有中等的语义抽象级别,能够表征地物的部件或小范围组合。

[0032] 深层特征 F_{T1}^3, F_{T2}^3 :ResNet的layer4(Block4)输出,其维度为 $[B, C_4, H/8, W/8]$ 的特征图,该层特征具有高度抽象的全局语义信息。该过程表示为:

$$\{F_{Tx}^1, F_{Tx}^2, F_{Tx}^3\} = \Phi_{CNN}(I_{Tx}), \quad Tx \in \{T1, T2\}, x \text{ 取1或2},$$

通过此步骤,对于第一影像 I_{T1} 和第二影像 I_{T2} 中的每张输入影像,均可获得一套包含浅、中、深三个层次的特征图集。

[0033] 第二步:基于注意力加权的多层语义令牌Token生成

本步骤旨在将不同尺度的二维空间特征图高效地转化为一系列低维度的语义令牌序列。如图2所示,该步骤由多层语义令牌生成模块完成,具体实现包括:

特征图尺寸统一:为方便后续处理,将中层特征图 F^2 和深层特征图 F^3 的空间分辨率上采样至与浅层特征图 F^1 一致,即 $\frac{H}{4} \times \frac{W}{4}$,得到尺度统一后的特征图集。统一尺度的作用在于消除不同层级特征在空间维度上的差异,确保后续生成的空间注意力权重图具有相同的分辨率,从而能够精确对应到图像的同一空间位置进行加权聚合。

[0034] 特征投影与注意力图生成:对尺寸统一后的三层特征图 $\{F^1, F^2, F^3\}$,并行地进行两种操作:

特征投影:使用独立的 1×1 卷积层 $Conv_P^l$ (即第一卷积层)将其通道维度投影到一个预设的令牌维度 d_{token} , d_{token} 为正整数,本实施例中取值为32,得到用于生成令牌内容的特征 $P^l \in \mathbb{R}^{B \times d_{token} \times h \times w}$,其中符号 $\mathbb{R}$ 表示实数集,指代该张量的数据类型为实数张量。

[0035] 注意力权重图生成:使用另一组独立的 1×1 卷积层 $Conv_A^l$ (即第二卷积层),该卷积层与特征投影层的参数独立,虽卷积核大小相同,但输出通道数为 N ,为每一层级的特征图生成 N 个注意力图,得到注意力权重图 $A^l \in \mathbb{R}^{B \times N \times h \times w}$ 。其中 N 为每层生成的令牌数量,取正整数,本发明中取 $N = 4$ 。

[0036] 注意力归一化:对生成的注意力权重图 A^l 在空间维度上应用Softmax函数进行归一化,使得每个令牌在空间上的注意力权重之和为1。对于第 k 个令牌的注意力图 $\hat{A}_{b,k,i,j}^l$,其归一化过程如下:

$$\hat{A}_{b,k,i,j}^l = \frac{\exp(A_{b,k,i,j}^l)}{\sum_{i'=1}^h \sum_{j'=1}^w \exp(A_{b,k,i',j'}^l)}, \text{ 其中, } b \text{ 表示批次索引, } k \text{ 表示令牌索引, } 1 < k < N,$$

(i, j) 和 (i', j') 均表示特征图的空间坐标索引, $1 < i < H, 1 < j < W, i', j'$ 与 i, j 的取值范围相同, l 表示特征层级索引。 $A_{b,k,i,j}^l$ 表示在 l 特征层级下 b 批次的第 k 个令牌索引在 (i, j) 点位的取值。

[0037] 语义令牌生成:将归一化后的注意力图 $\hat{A}^l$ 作为空间权重,对令牌内容的特征 P^l 进行加权求和,从而为每一层特征图生成 N 个长度为 d_{token} 的语义令牌。第 k 个令牌 $T_k^l \in \mathbb{R}^{B \times d_{token}}$ 的计算过程如下:

$$T_k^l = \sum_{i=1}^h \sum_{j=1}^w \hat{A}_{:,k,i,j}^l \cdot P_{:,i,j}^l, \text{ 其中 } P_{:,i,j}^l \text{ 表示在空间位置 } (i, j) \text{ 处的特征向量 (包含所有批次和通道), } \hat{A}_{:,k,i,j}^l \text{ 是对应的标量权重。公式中的点乘与求和操作对应于爱因斯坦求和约定Einsum,即在空间维度 } (H, W) \text{ 上进行加权聚合。此过程等效于一种注意力池化,将空间特征信息依据学到的重要性分布汇聚成紧凑的向量表示。}$$

有批次和通道), $\hat{A}_{:,k,i,j}^l$ 是对应的标量权重。公式中的点乘与求和操作对应于爱因斯坦求和约定Einsum,即在空间维度 (H, W) 上进行加权聚合。此过程等效于一种注意力池化,将空间特征信息依据学到的重要性分布汇聚成紧凑的向量表示。

[0038] 令牌序列构建:将从影像 I_{T1} 的三层特征中生成的所有令牌 $\{T^1, T^2, T^3\}$,本发明中 T^1, T^2, T^3 取值大小均为 4×32 ,沿序列维度进行拼接,形成一个包含 $3 \times N$ 个令牌的完整序列第一令牌 T_{T1} ,同理,对影像 I_{T2} 进行完全相同的操作,得到第二令牌序列 T_{T2} 。

[0039] 第三步:基于Transformer编码器的时空上下文信息建模

本步骤利用Transformer强大的序列建模能力,学习两个时相、多个尺度特征令牌之间的复杂依赖关系。如图3左半部分所示,该步骤由Transformer编码器完成,具体实现包括:

序列拼接与位置编码:将第一影像 I_{T1} 和第二影像 I_{T2} 的第一令牌序列 T_{T1} 和第一令牌序列 T_{T2} 进行拼接,形成一个包含所有时相、所有尺度信息的长序列 $T_{cat} \in \mathbb{R}^{B \times (6N) \times d_{token}}$,其大小为 24×32 ,为该序列添加可学习的位置编码序列 E_{pos} ,以使模型能够区分不同令牌的来源(时相及特征层级),则有:

$T_{in} = T_{cat} + E_{pos}$,其中 T_{in} 表示添加位置编码后输入到Transformer编码器的最终令牌序列。

[0040] 编码器处理:将 T_{in} 输入到一个标准的多层Transformer编码器中。其核心为多头自注意力(Multi-head Self-Attention, MHSA)机制。对于输入序列,首先通过线性变换得到查询Query(Q)、键Key(K)、值Value(V)三个矩阵:

$Q = T_{in}W_Q, K = T_{in}W_K, V = T_{in}W_V$,其中 W_Q, W_K, W_V 分别表示用于生成查询、键和值的可学习线性投影权重矩阵。

[0041] 然后计算注意力得分并加权求和:利用点积计算查询矩阵 Q 与键矩阵 K 之间的相似度,经缩放和Softmax归一化后得到注意力权重,再对值矩阵 V 进行加权求和。计算公式如下:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \text{ 其中, } d_k \text{ 表示键向量的维度, } \sqrt{d_k} \text{ 作为分母}$$

用于缩放点积结果,以防止数值过大导致梯度消失;softmax()函数计算出的结果即为注意力权重(Attention Score),该权重作用于值矩阵 V 上完成加权求和,从而捕获序列内部的全局依赖关系。

[0042] 将上述自注意力机制的输出通过层归一化(Layer Norm)和多层感知机(MLP)模块进行非线性变换和特征提炼,最终得到编码器的输出。编码器的输出是与输入序列等长的一组经过深度信息交互和提炼后的编码令牌 $M \in \mathbb{R}^{B \times (6N) \times d_{token}}$,大小为 24×32 。

[0043] 第四步:基于Transformer解码器的差异特征图生成

本步骤使用Transformer解码器直接生成差异特征。如图3右半部分所示,该步骤由Transformer解码器完成,具体实现包括:

解码器输入准备:

查询(Query):将第一和第二两个时相的深层特征图 F_{T1}^3 和 F_{T2}^3 在通道维度上进行拼接,形成一个融合了两时相高级语义信息的特征图 $F_{cat} \in \mathbb{R}^{B \times 2C_4 \times h/4 \times w/4}$ 。该特征图将作为解码器的查询输入。

[0044] 键(Key)与值(Value):将第三步中编码器输出的编码令牌序列 M 作为解码器的记忆(Memory)。在此过程中,利用图3所示的线性投影层Linear将编码令牌 M 投影到解码器查询Q的高维特征空间(本实施例中为 24×64),以增强特征的表达能力并匹配解码器的内部维度,投影后的令牌序列分别作为交叉注意力机制中的“键”和“值”。

[0045] 交叉注意力解码:将 F_{cat} (作为Query)和 M (作为Key和Value)输入Transformer解码器。在解码器的多头交叉注意力(Multi-head Cross-Attention)层中,Query来自 F_{cat} 的空间位置,而Key和Value来自编码令牌 M 。这使得解码器能够利用 F_{cat} 中的每个像素位置去“查询”全局时空变化上下文信息(编码在令牌 M 中),并将这些信息动态地、有选择地注入回二维的特征图中,此时产生与输入 F_{cat} 维度相同的初步差异特征图。

[0046] 输出差异特征:将上述交叉注意力机制的输出,即初步差异特征图通过层归一化(Layer Norm)和多层感知机MLP模块进行非线性变换和特征提炼,最终得到解码器的输出。解码器的输出是一个与输入查询特征图尺寸相同,但通道信息已经被令牌中的变化信息深度调制的特征图 F_{diff} 。该特征图直接表征了地物的差异信息,其语义层次比简单的特征相减更为丰富和可靠。

[0047] 第五步:特征融合、上采样与变化分类

本步骤旨在将解码器生成的差异特征图恢复至原始分辨率,并进行最终的像素级分类。该步骤由图1中的上采样与分类模块完成。

[0048] 特征融合:使用一个卷积模块对解码器输出的差异特征图 F_{diff} 进行通道融合与降维,进一步整合信息,得到通道数更少、特征表达更紧凑的整合差异特征图。

[0049] 上采样恢复分辨率:通过一系列的上采样模块(如转置卷积或插值+卷积),将整合差异特征图的分辨率逐步放大,直至恢复到与输入影像相同的原始尺寸 $[B, C_{out}, H, W]$,得到全分辨率的差异特征图。进行此操作的原因是变化检测是像素级的预测任务,必须将特征图恢复到与原图一致的大小,才能为原始影像中的每一个像素点提供分类预测。

[0050] 分类预测:将全分辨率的差异特征图输入到一个最终的分器中(为一个 1×1 卷积层),该分器将特征图的通道数调整为类别 N_{class} (对于二元变化检测,通常为2,即“变化”与“未变化”)。该操作生成的结果是未经归一化的分类得分图Logits,其中每个像素位置包含两个数值,分别代表该点属于“未变化”和“变化”的原始置信度得分。

[0051] 输出结果:对分类器的输出在通道维度上应用Softmax函数,将得分转换为概率。Softmax函数的计算表达式如下:

$$P(c, i, j) = \frac{e^{x_{c,i,j}}}{\sum_{k=1}^{N_{class}} e^{x_{k,i,j}}}$$

其中, $x_{c,i,j}$ 表示在空间位置 (i, j) 处第 c 类的原始得分, N_{class} 为总类别数。通过该计算,得到双时相影像在同一地理坐标位置上的每个像素点属于各个类别的概率图

$P_{out} \in \mathbb{R}^{B \times N_{class} \times H \times W}$ 。需要说明的是,概率图 P_{out} 并非最终的变化检测图。为了得到最终的二值化结果,再对概率图在通道维度上进行Argmax(求最大值索引)操作,从而生成最终的二值化变化检测掩膜图(Change Map),其中像素值为1表示变化,0表示未变化。

[0052] 综上所述,本发明通过多尺度特征提取、注意力加权令牌化、Transformer编解码器结构,构建了一个端到端的深度学习模型,能够有效融合不同时相遥感影像的多层次特征,精确地检测出地物变化区域,具有较高的鲁棒性和准确性。

[0053] 作为本发明公开的实施例,本发明还提供了一种融合多尺度特征的Transformer地物变化检测装置,所述装置用于实现所述的方法,包括:

提取模块,用于对两个时相的遥感影像进行提取,得到不同层级的多尺度特征信息;

压缩模块,用于将所述多尺度特征信息压缩并符号化为一系列语义令牌;

编码模块,用于将所有语义令牌进行编码得到编码后的令牌信息;

拼接模块,用于将编码后的令牌信息进行解码拼接,得到高信息含量的差异特征图;

处理模块,用于对差异特征图进行处理,得到最终的变化检测结果。

[0054] 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。

[0055] 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。

[0056] 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。

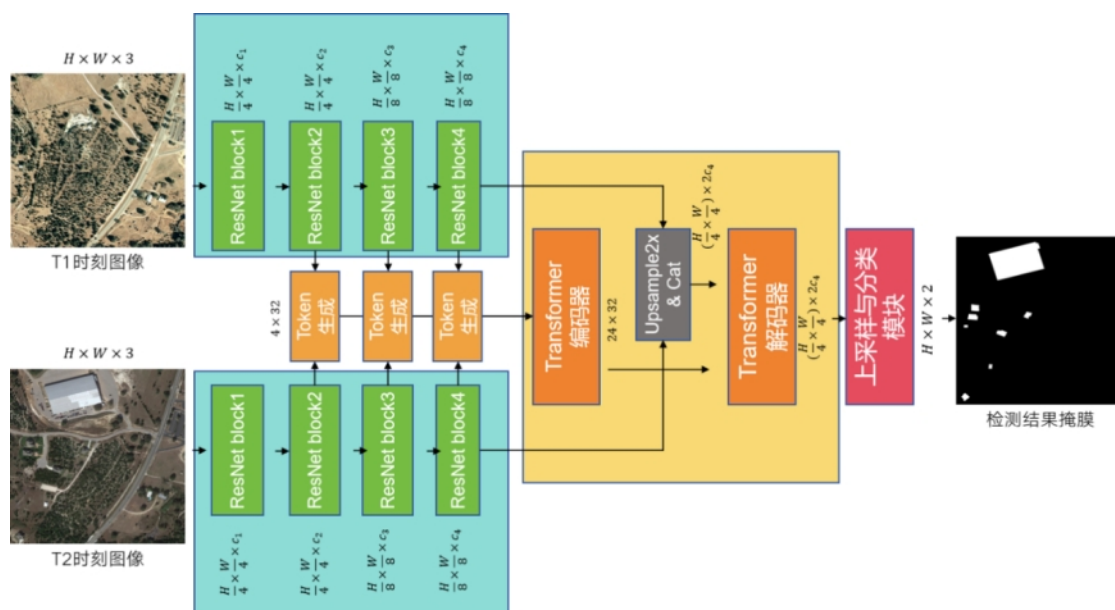


图1

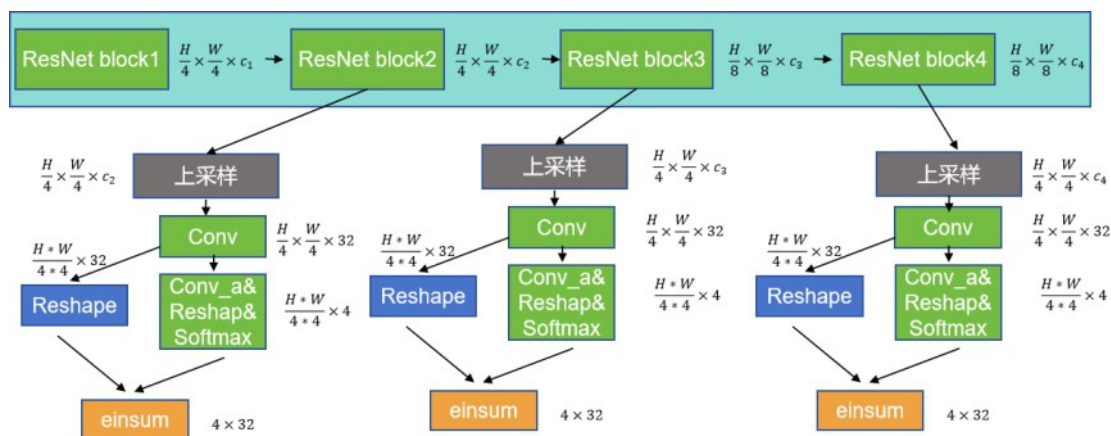


图2

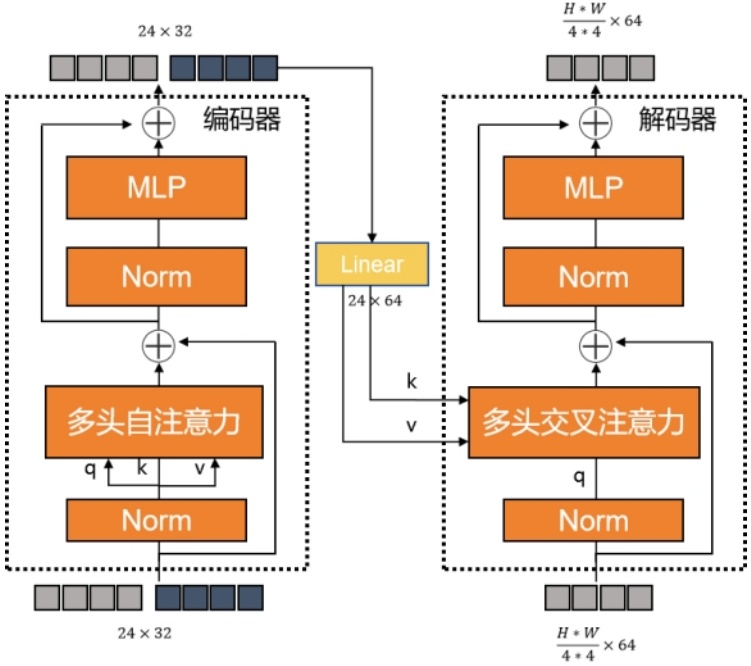


图3