



引文格式: 厉芳婷, 张过, 周方栋, 等. 基于双流并行全向扫描Mamba的遥感影像建筑物变化检测[J]. 武汉大学学报(信息科学版), 2024, 49(9):1574-1585. DOI:10.13203/j.whugis20240270

Citation: LI Fangting, ZHANG Guo, ZHOU Fangdong, et al. Building Change Detection Based on Dual-Stream Parallel Omnidirectional Scan Mamba Network[J]. Geomatics and Information Science of Wuhan University, 2024, 49(9): 1574-1585. DOI: 10.13203/j.whugis20240270

基于双流并行全向扫描Mamba的遥感影像 建筑物变化检测

厉芳婷^{1,2} 张过¹ 周方栋³ 王上³ 梁思² 李志炜⁴

1 武汉大学测绘遥感信息工程国家重点实验室, 湖北 武汉, 430079

2 湖北省航测遥感院, 湖北 武汉, 430074

3 华中师范大学人工智能教育学部, 湖北 武汉, 430079

4 91001部队, 北京, 100841

摘要: 采用深度学习技术对海量遥感影像进行建筑物变化检测与解译分析, 可以为自然资源要素管理和国土资源节约集约化提供关键科学依据和数据支持, 对自然资源动态监测和空间治理现代化具有重要意义。状态空间模型Mamba结合了全局感受野与动态加权, 在时间序列预测任务中能够更好地捕捉趋势和周期性信息, 然而目前Mamba在遥感影像密集预测的研究尚不充分, 对高分辨率遥感影像变化检测存在计算复杂度高、检测效率低等问题。对影响Mamba算法参数量的关键因素进行了深入分析, 提出了双流并行全向扫描Mamba网络, 构建通道内的数据并行化处理方法, 在通道总数不变的情况下, 有效减少单个Mamba块中数据的通道数, 实现网络总体参数量的下降和高效的特征提取。在LEVIR-CD和WHU-CD两个数据集上进行了对比实验, 结果表明, 所提网络各项指标均优于非Mamba架构的网络, 与基准网络对比训练效率分别提升了19.45%和8.26%, 且在网络参数量与计算复杂度方面显著降低, 分别降低了35.8%和22.4%。

关键词: 高分辨率遥感影像; 变化检测; 深度学习; 状态空间模型; 双流并行

中图分类号: P237

文献标识码: A

收稿日期: 2024-03-12

DOI: 10.13203/j.whugis20240270

文章编号: 1671-8860(2024)09-1574-12

Building Change Detection Based on Dual-Stream Parallel Omnidirectional Scan Mamba Network

LI Fangting^{1,2} ZHANG Guo¹ ZHOU Fangdong³ WANG Shang³ LIANG Si² LI Zhiwei⁴

1 State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

2 Hubei Institute of Photogrammetry and Remote Sensing, Wuhan 430074, China

3 Faculty of Artificial Intelligence in Education, Central China Normal University, Wuhan 430079, China

4 91001 Troops, Beijing 100841, China

Abstract: Objectives: The state space model Mamba, which combines global sensing field and dynamic weighting, can better capture the trend and periodic information in the time series prediction task, and has become an important direction in the current deep learning technologies for remote sensing image analysis and interpretation. However, the current research of Mamba in the dense prediction of remote sensing images is insufficient, and there are problems such as high computational complexity and low detection efficiency for the change detection of high-resolution remote sensing images. **Methods:** We conduct an in-depth analysis of the key factors affecting the number of parameters of the Mamba layer, and propose a dual-stream parallel

基金项目: 国家自然科学基金(62177017)。

第一作者: 厉芳婷, 博士, 高级工程师, 主要研究方向为遥感影像数据处理技术和应用。lifangting@whu.edu.cn

通讯作者: 张过, 博士, 教授。guozhang@whu.edu.cn

周方栋, 博士生。fdzhou@mails.cnu.edu.cn

omnidirectional scan Mamba (DSPOSM) model architecture to construct an in-channel data parallelization processing method. The proposed DSPOSM model can effectively reduce the number of channels of the data in single Mamba block while the total number of channels remains unchanged, in order to achieve highly efficient feature extraction based on the decrease of the overall number of parameters of the model. **Results:** Comparative experimental results on the LEVIR-CD and WHU-CD datasets show that the proposed method outperforms the non-Mamba architecture method in all metrics and significantly reduces the number of parameters and computational complexity of the network by 35.8% and 22.4%, and improves the training efficiency by 19.45% and 8.26%, compared with the benchmark method, respectively. **Conclusions:** The proposed DSPOSM method can significantly reduce the number of parameters and computational complexity of Mamba-based networks and improve training efficiency.

Key words: high-resolution remote sensing image; change detection; deep learning; state space model; dual-stream parallel

城市化和经济社会的快速发展使得土地利用的问题日益突出^[1],自然资源部早在 2014 年出台规定实行严格的耕地保护和节约集约用地制度,提升土地资源对经济社会发展的承载能力。对地观测技术的快速发展极大地提升了遥感影像数据的获取质量与数量,如何更有效、准确地利用海量的高分辨率遥感数据进行变化检测与解译分析,已成为遥感技术应用的核心问题,而遥感影像中建筑物变化信息的提取能为城市规划、土地覆盖变化以及国土资源管理等重大工程和应用提供科学决策依据^[2-3]。

目前的遥感影像建筑物自动提取方法主要分为传统方法和基于深度学习方法。传统的建筑物变化检测方法主要包括基于像素、特征、目标和时序变化的检测方法^[4-6]。传统方法需要大量的人工投入,且对不同条件的遥感影像泛化性较弱,随着数据量的增加和对检测结果准确度要求的不断提高,传统方法难以满足海量高分辨率遥感影像对复杂场景中建筑物变化的提取需求^[7]。深度学习的出现极大地提高了变化检测的效率和准确性。其中,卷积神经网络(convolutional neural network, CNN)通过多层卷积和池化操作,自动学习图像的特征表示,并且具有较强的非线性建模能力,能够更好地适应遥感影像中的复杂地面环境,被广泛应用于高分辨率遥感影像的建筑物变化检测^[8-10]。但是 CNN 存在模型结构上固有的问题,有限的感受野使得模型无法很好地获取像素之间的长期相关性,在处理不同时空分辨率中复杂的多时相场景时仍存在不足^[11]。有学者将 Transformer^[12]引入变化检测任务,有效解决了上述问题。Transformer 利用自注意力机制来建模遥感影像中所有像素之间空间关系,能够捕捉到建筑物变化的细节和上下文信

息,获取多时相特征之间的时空关系^[13]。然而,自注意力机制需要对所有像素点进行计算,导致计算复杂度较高,特别是在处理大规模高分辨率遥感影像时,需要较高的计算资源和时间。同时,基于 Transformer 的模型在预处理阶段会将原始图像划分为小块,这导致其仅包含图像的一部分,提供的上下文信息有限。

状态空间模型(state space model, SSM)是将状态空间变化引入深度学习的新架构。状态空间可看作对地图的简化描述,通过当前所处位置及其他描述状态的变量表示为当前状态,SSM 根据输入预测下一步可能的状态,即可能的位置。结构化状态空间序列(structured state space sequence, S4)模型^[14]是线性时不变(linear time-invariant, LTI)的 SSM,通过数据离散化兼顾 CNN 的并行训练和循环神经网络(recurrent neural network, RNN)的快速推理,同时引入隐藏态来处理长程依赖关系。因此,S4 模型能够较好地处理连续长序列数据,如文本、音频和传感器数据。Mamba^[15]将时变参数集成到 S4 模型中,允许模型过滤信息,同时使用硬件感知算法加快训练和推理过程,凭借强大的上下文建模能力和线性复杂度,有望在语言建模领域替代 Transformer 的主流地位。与文本和音频这类序列数据不同,图像是一种非定向的数据,文献[16]通过交叉扫描机制在二维图像上进行选择性扫描,提升有效感受野,将图像任务转变为长序列任务,在相同参数量的情况下对图像分类、目标检测和语义分割等视觉任务取得更佳优异的性能。此外,文献[17]表明 Mamba 更适用于目标检测、语义分割等具有长序列和自回归特征的任务。遥感数据特别是高分辨率光学影像中存在丰富的地表语义信息,这些语义信息在时间和空间上都表现出长

距离依赖特性, Mamba架构在这类大规模的密集预测和时序预测任务上具有非常大的潜力。文献[18-19]将Mamba架构应用于遥感影像密集检测任务, 在语义分割和变化检测中展现了优异的性能。RS-Mamba^[18]针对遥感影像中地表覆盖在空间上的不确定性, 优化交叉扫描机制, 对高分辨率遥感图像在多个方向上进行全局上下文建模, 并深入研究遥感图像尺寸对密集预测任务性能的具体影响; ChangeMamba^[19]针对不同变化检测任务的需求, 提出3种基于Mamba架构的变化检测网络, 并结合Mamba架构提出时空关系建模机制实现多时态特征的时空交互, 从而获得精确的变化预测结果。然而, 这些研究在网络中使用复杂的机制提高网络的准确度, 显著增加了网络的参数量和计算复杂度。在训练效率和计算复杂度方面, Mamba架构并未完全展现出预期的显著优势。

在此基础上, 本文为进一步研究Mamba架构在遥感影像变化检测任务中的潜力, 提出了用于高分辨率遥感建筑物变化检测的双流并行全向扫描Mamba网络(dual-stream parallel omnidirectional scan Mamba network, DSPOSM); 通过深度分析Mamba架构中影响参数数量的因素, 提出用于提取深层特征的双流并行全向扫描(dual-stream parallel omnidirectional scan, DSPOS)块, 在保持数据通道总数不变的情况下, 实现参数量的大幅下降和高效的变化特征提取。

1 研究方法

1.1 SSM基本原理

近年来, SSM由于其在序列长度上线性扩展的特性在深度学习领域受到广泛关注。传统的SSM可以看作是线性时不变系统, 其通过隐状态 $h(t) \in \mathbf{R}^{d \times L}$, 将输入序列 $x(t) \in \mathbf{R}$ 映射到输出响应 $y(t) \in \mathbf{R}$, 其中 d 表示状态维度的大小, L 为序列长度。该过程可以用线性常微分方程表示:

$$\begin{cases} h'(t) = Ah(t) + Bx(t) \\ y(t) = Ch(t) + Dx(t) \end{cases} \quad (1)$$

式中, t 为当前的输入时间; $h'(t)$ 表示当前输入 $x(t)$ 的隐藏态; $h(t)$ 表示上一时间的隐藏态; $A \in \mathbf{R}^{d \times d}$ 为演化参数; $B \in \mathbf{R}^{d \times L}$ 和 $C \in \mathbf{R}^{L \times d}$ 为投影参数; $D \in \mathbf{R}^{L \times L}$ 为用于残差操作的动态参数。但在推理过程中, A 、 B 、 C 固定为训练时的参数, 不随输入不同而变化, 无法针对输入做针对性推理。

深度学习中使用SSM必须首先将数据离散化, 其通过将连续常微分方程离散化来实现。文献[14]提出的S4模型是连续SSM的离散化应用, 通过步长参数 Δ 将连续参数 A 和 B 转换为离散参数 \bar{A} 和 \bar{B} 。SSM实现离散化最常用的方法是零阶保持法(zero-order hold, ZOH), 其计算式为:

$$\begin{cases} \bar{A} = e^{(\Delta A)} \\ \bar{B} = (\Delta A)^{-1} (e^{(\Delta A)} - I) \Delta B \end{cases} \quad (2)$$

式中, I 表示单位矩阵。由此得到离散化后SSM方程为:

$$\begin{cases} h_p = \bar{A}h_{p-1} + \bar{B}x_p \\ y_p = Ch_p + Dx_p \end{cases} \quad (3)$$

式中, h_p 是时间步长为 p 时的隐藏态; x_p 、 y_p 分别是时间步长为 p 时的输入和输出序列。

文献[15]提出的Mamba架构在全局建模能力上与Transformer近似, 同时还能保持线性增长率的计算开销。Mamba块的结构如图1(a)所示, 其采用SiLU作为非线性激活函数。S4模型由于其LTI的特性, 在推理过程中 A 、 B 、 C 为不随输入动态改变的固定参数。Mamba在S4模型的基础上提出了选择性机制, 该机制的核心思想在于将离散参数 \bar{A} 、 \bar{B} 和 \bar{C} 变为输入依赖的参数, 允许在推理过程中基于输入序列和历史信息对参数 B 和 D 进行动态调整, 因此Mamba中SSM能够对输入信息有选择性处理, 从而更有效地对复杂的时间信息进行建模。Mamba通过输入序列 x 建模 B 、 C 和 Δ 的函数, 计算式为:

$$\begin{cases} B = S_B(x) \\ C = S_C(x) \\ \Delta = \tau_\Delta(S_\Delta(x) + b) \end{cases} \quad (4)$$

式中, x 为输入序列; $S_B(x)$ 、 $S_C(x)$ 均表示对 x 进行投影维度为 d 的线性投影操作; $S_\Delta(x)$ 表示对 x 进行投影维度为1的线性投影操作; τ_Δ 表示Softplus激活函数; b 为偏置项。在此基础上, Mamba使用简化后的ZOH法离散化参数, 计算式为:

$$\begin{cases} \bar{A} = e^{(\Delta A)} \\ \bar{B} = \Delta B \end{cases} \quad (5)$$

由式(5)可知, \bar{A} 仍使用ZOH法, \bar{B} 使用ZOH法的一阶近似形式减少计算量。Mamba中SSM的整体流程如图1(b)所示, Mamba通过步长参数 Δ 决定是否关注当前输入, 即允许SSM根据输入的内容选择性地记住或忽略信息。此外, Mamba采用并行扫描算法替代S4模型中的使用卷积进行模型的循环计算算法, 从而优化GPU内存布局。这些改进使得Mamba在保持SSM线性扩展

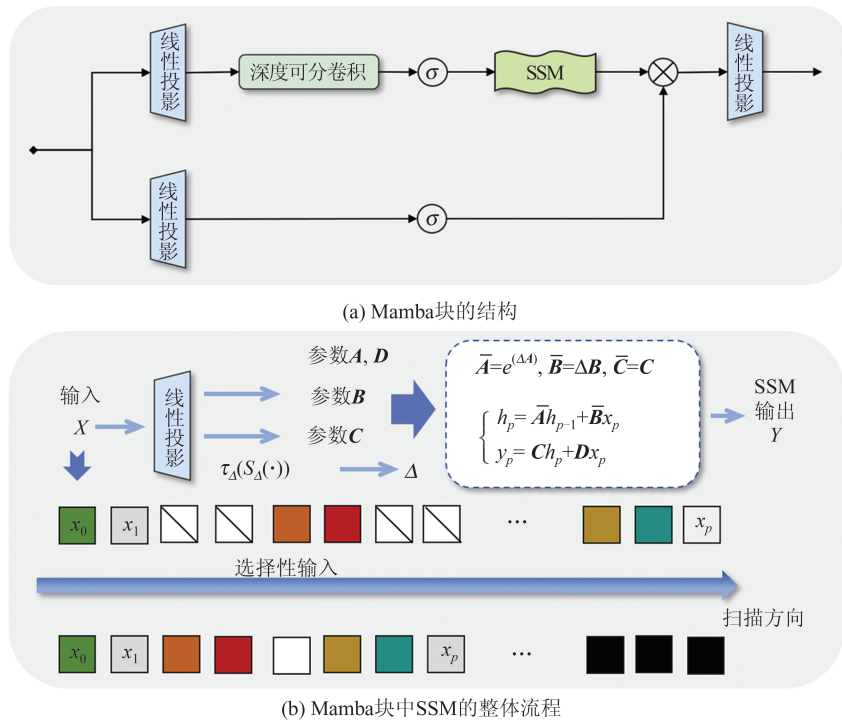


图 1 Mamba 块和 SSM 的图示

Fig. 1 Illustration of Mamba Block and SSM

特性的同时,能够在具有长序列特征的任务上取得优异的效果。

然而上述模型仅能处理单向序列数据,无法处理没有特定方向的图像数据。文献[16]提出了通用二维视觉模型视觉 Mamba (vision Mamba, VMamba),通过引入交叉扫描模块在水平和垂直方向扫描输入图像数据,使得模型能够在二维图像空间中进行一维选择性扫描,进而增强模型的全局有效感受野。此外,VMamba 在设计架构时做了进一步的改进,提高了模型的性能及推理速度,在将计算复杂度降到线性复杂度的同时,实现超过视觉 Transformer (vision Transformer, ViT)^[20]的性能。

1.2 DSPOS 网络

本文提出的 DSPOS 网络结构如图 2 所示,其采用 Siamese 编码器-解码器网络结构,主要包括特征提取、解码和特征融合 3 个部分。其中,输入图像尺寸为 $H \times W \times 3$, H 、 W 分别表示图像的高、宽,3 为图像的通道数;特征提取层采用双流设计的 DSPOS 块,提高高分辨率遥感影像中建筑物特征的提取效率;解码阶段中 Conv+BN+ReLU 表示依次经过 1×1 卷积、批标准化和 ReLU 激活函数的运算。输入的双时相高分辨率遥感影像首先被转换为双时相图像块序列,编码阶段中输入序列送入到 4 组共享权重的特征提取层中提取特征。

为了避免引入过多的计算量,特征提取层中 DSPOS 块的层数 N_i 设置为 $\{2, 2, 9, 2\}$ ($i=1, 2, 3, 4$),特征图的通道数 C_i 设置为 $\{96, 192, 384, 768\}$;在特征融合模块中,特征提取层提取的双时相特征图通过通道维度卷积融合,获取两个时相中丰富的语义特征;融合后的特征图通过两次卷积平滑边缘输入到解码器中;解码阶段与编码阶段对应,在解码层中,上层解码特征图经过上采样操作后,与对应编码层融合特征图进行级联,融合高层语义信息和低层次细节信息,其中,第一个解码层为最后一级特征提取层的融合特征图与上一特征提取层融合特征图在解码层级联;特征图在解码器中逐步上采样,得到最后的变化结果后输出图像。

1.3 Mamba 参数量分析

Mamba 对输入的数据在通道维度上并行处理,其参数量大小受输入数据的通道数、SSM 状态维度以及卷积核等多方面的影响,其中 SSM 状态维度和深度可分卷积中卷积核的大小在实际中因网络具体任务设定为固定参数。为此,本节针对输入数据的通道数对 Mamba 中参数量的影响进行分析。

首先,Mamba 在架构中使用深度可分卷积代替传统卷积层,从而减少参数量和计算量。对于通道数为 c_m 的图片,其尺寸为 $H \times W$,分别对其进行卷积核为 3×3 的常规卷积和深度可分卷积,输

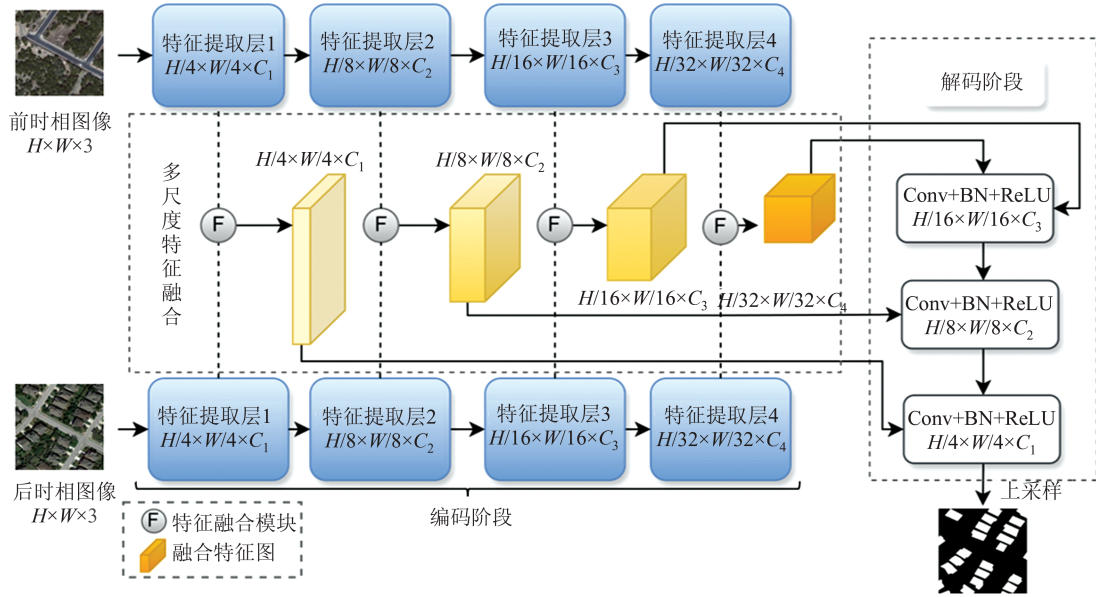


图2 DSPOS网络结构

Fig. 2 Structure of Dual-Stream Parallel Omnidirectional Scan Mamba Network

出通道数为 c_{out} 。常规卷积的参数量为 $c_{in} \times c_{out} \times 3 \times 3$, 计算量为 $c_{in} \times c_{out} \times 3 \times 3 \times H \times W$; 深度可分卷积的参数量为 $c_{in} \times 3 \times 3 + c_{out} \times c_{in} \times 1 \times 1$, 计算量为 $c_{in} \times 3 \times 3 \times H \times W + c_{out} \times 3 \times H \times W$, 卷积操作的参数量和计算量从常规卷积的关于通道数的平方数量级减少到关于通道数的线性复杂度。因此, 深度可分卷积的参数量只与输入数据通道总数相关。

其次, Mamba 中存在 2 个输入投影层和 1 个输出投影层, 输入投影层的参数量为 $c_{in} \times c_{inner} \times 2$, 输出投影层的参数量为 $c_{inner} \times c_{out}$ 。内部投影通道数 c_{inner} 由投影扩展倍数 r 和输入通道数 c_{in} 的乘积确定, 本文将 r 设置为固定参数 2。投影层的参数量计算式为:

$$\begin{cases} P_{in_proj} = c_{in} \times c_{inner} \times 2 \\ P_{out_proj} = c_{inner} \times c_{out} \times 2 \end{cases} \quad (6)$$

式中, P_{in_proj} 以及 P_{out_proj} 分别表示输入投影层与输出投影层的参数量。Mamba 块的输入和输出通道数相同, 即 $c_{in} = c_{out}$, 且内部投影通道数 c_{inner} 为输入通道数 c_{in} 的固定倍数, 因此输入投影层和输出投影层的参数量只与输入通道数 c_{in} 相关, 且投影层的参数量随着通道数的增加成二次方比例增加。此外, SSM 中的线性投影层参数也与输入数据通道数 c_{in} 相关, 其参数量计算式为:

$$P_{SSM_proj} = c_{inner} (k + d_{state} \times 2) \quad (7)$$

式中, P_{SSM_proj} 表示 SSM 内部的参数量; k 为输入依赖补偿的秩, 大小为 c_{in} 的 1/16; d_{state} 为状态维的大小, 本文固定为 16。

综上, 由于 Mamba 内部存在较多的投影层, 在保持其他参数不变的情况下, 减少单个 Mamba 输入数据的通道数对 Mamba 的参数量有显著影响。在此基础上, 本文提出了使用双流并行全向扫描块作为特征提取层的 DSPOS 网络, 在保持优异性能的前提下有效降低网络的参数量。

1.4 双流并行全向扫描块设计

基于 §1.3 中输入数据通道数对 Mamba 参数影响的分析, 本文设计了一个全新的用于遥感影像建筑物特征提取的 DSPOS 块。其在保持输入数据的通道总数不变的情况下, 能够在取得优异性能的同时大幅减少参数量。DSPOS 块使用与在视觉状态空间 (visual state space, VSS) 块^[16]的基础上进行扫描方向改进的全向状态空间 (omnidirectional state space, OSS) 块^[19]相同的结构, 在 OSS 块中引入双流并行方法。VSS 块和 OSS 块的选择性扫描机制如图 3 所示, OSS 块中全向扫描模块通过增加扫描方向捕获更加复杂的空间关系, 提升全局建模能力。

DSPOS 块结构如图 4 所示, 其中定向扫描 (oriented scanning Mamba, OSM) 模块的扫描策略与 OSS 块相同。在 OSM 模块中, Mamba 在通道维度上并行提取特征。在多流并行操作中, 单个 OSM 模块输入数据通道数减少, 但数据的总通道数保持不变, Mamba 的计算效率并未受到影响。OSM 模块输入数据通道数减少, 线性投影层对输入特征数据进行线性映射时仅包含当前输入数据的信息, 可能会影响网络学习复杂的跨通道特征表示。在 DSPOS 块中, 数据经由两个独

立的 OSM 模块,可能会学习到不同的特征表示,这些特征在通道维度上拼接互补从而保持网络较好的性能。本文选择使用双流并行操作构建 DSPOS 块,在减少参数数量和复杂度的同时最大限度减少并行操作对性能的影响。双流并行的具体操作如下:在 DSPOS 块中,通道数为 C 的影像特征 X 在经过层归一化操作后被分为通道数为 $C/2$ 的影像特征 $X_1^{C/2}$ 、 $X_2^{C/2}$,然后将每个影像特征送入 OSM 模块内进行计算;得到的深层影像特征通过残差连接优化远程空间信息获取能力,最后经过通道维度连接得到输出 X_{out} 。

OSM 模块的方法流程如下:首先,数据通过线性投影调整维度进行 3×3 深度可分卷积,在每个输入通道上并行操作,提取空间特征,保证在不牺牲网络性能的前提下显著减少参数量、提升计算效率;其次,经过卷积和 SiLU^[21] 激活函数之后的数据在全向扫描 Mamba 模块中分别沿水平、垂直、对角线以及反对角线的前向与反向共 8 个方向进行定向扫描,确保特征的全面捕获,扫描得到的特征序列按并行进行 Mamba 计算,然后将处理后的序列按通道维度进行相加操作。最后,全向扫描模块的输出与经过线性投影和

SiLU 激活函数处理的原始输入进行特征融合,通过输出线性投影调整维度得到 OSM 模块的最终输出。

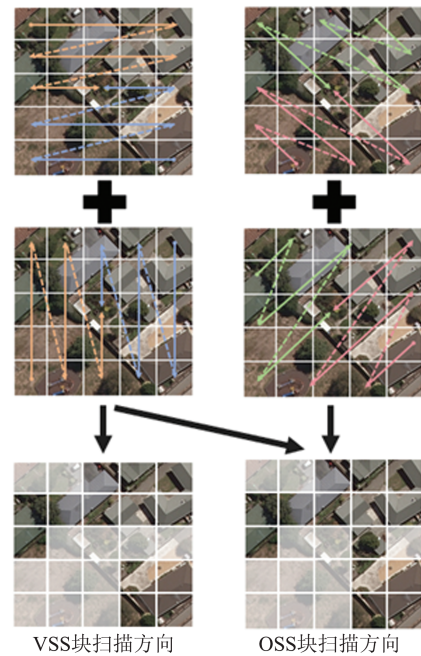


图 3 VSS 块和 OSS 块的选择性扫描机制

Fig. 3 Illustration of the Selective-Scan Mechanism for VSS Block and OSS Block

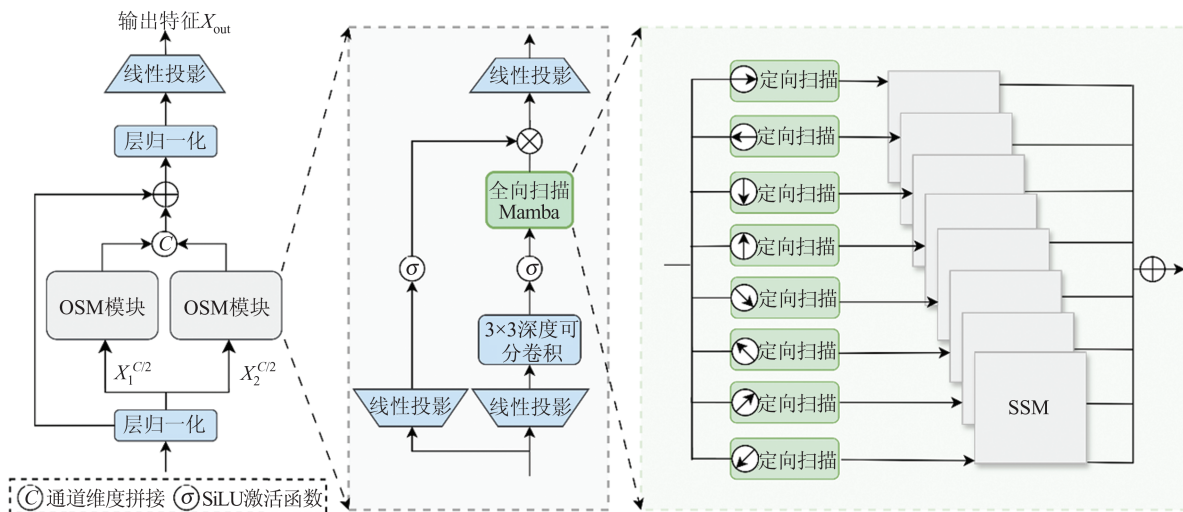


图 4 DSPOS 块的结构

Fig. 4 Architecture of DSPOS Block

1.5 特征融合模块

输入影像在编码阶段进行特征提取的过程中,由于每一级特征提取层都有下采样操作,提取的特征图所蕴含的细节信息在编码阶段会存在一定的损失。为此,本文将各级特征提取层得到的双时相特征图通过特征融合模块进行特征融合,融合后的特征图通过跳跃连接与解码阶段的特征相结合,以提高变化检测结果中的细节特征精度。图 5 展示了特征融合模块的整

体流程,其中 H_i 、 W_i 和 C_i 分别表示四级特征图的高、宽和通道数 ($i=1, 2, 3, 4$), 双时相特征图通过 1×1 卷积进行特征融合获取丰富的细节信息和语义信息。对于分辨率较高的浅层特征,简单的卷积操作可以获得足够的细节信息,同时减少语义信息的损失。最后,特征图像经过两次 3×3 卷积平滑特征边缘,得到输出的融合特征图,并将其送入相应的编码层进行特征级联。

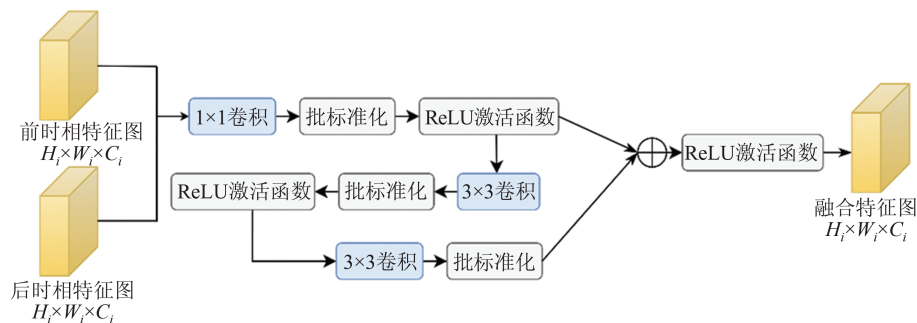


图5 特征融合模块

Fig. 5 Feature Fusion Module

2 实验与分析

2.1 数据集

本文实验使用的 LEVIR-CD 数据集^[22]为 2020 年公开的建筑物变化检测数据集,取景于美国得克萨斯州,包含 637 对空间分辨率为 0.5 m、尺寸为 $1\,024 \times 1\,024$ 像素的遥感影像;WHU-CD 数据集^[23]为 2018 年公开的建筑物变化检测数据

集,取景于新西兰 Christchurch,包含一组空间分辨率为 0.3 m、尺寸为 $15\,354 \times 32\,507$ 像素的航空图像大图。本文将两个数据集的原始图像统一裁剪为 256×256 像素的影像,并且按照 7:1:2 的比例划分为训练集、验证集和测试集。图 6(a)与图 6(b)分别展示了 LEVIR-CD 和 WHU-CD 数据集裁剪后的前、后时相影像以及对应的建筑物变化标签数据样例。



(a) LEVIR-CD数据集

(b) WHU-CD数据集

图6 变化检测数据集部分样例

Fig. 6 Examples of Change Detection Datasets

2.2 实验环境与参数设置

本文实验均在 64 位 Ubuntu 18.04 环境下进行,CPU 为 Intel Xeon Silver 4210R @2.0 GHz,显卡为 NVIDIA RTX A6000,所有实验均在 Python 3.9+PyTorch 1.11.0+CUDA 11.3 的深度学习环境下实现。实验训练过程选择 Adam^[24]作为优化器,初始学习率设为 0.000 1,衰减率 β_1 设置为 0.9, β_2 设置为 0.999,批量大小设置为 16,训练轮次设置为 200。本文网络采用二元交叉熵损失函数。

2.3 评价指标

本文使用精确度 P 、召回率 R 、F1 分数和交并比 (intersection over union, IoU) 作为评价变化检测网络性能的指标。各指标的值越大,表明算法的变化检测结果精度越高。上述指标的计算式分别为:

$$P = \frac{TP}{TP + FP} \times 100\% \quad (8)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (9)$$

$$F1 = \frac{2RP}{R + P} \times 100\% \quad (10)$$

$$IoU = \frac{TP}{TP + FN + FP} \times 100\% \quad (11)$$

式中, TP(true positive)表示正确检测影像中变化的像素数; TN(true negative)表示正确检测影像中未变化的像素数; FN(false negative)表示影像中变化、预测结果为未变化的像素数; FP(false positive)表示影像中未变化、预测结果为变化的像素数。

2.4 实验结果与分析

为了验证 DSPOSM 在高分辨率遥感影像变化检测任务的有效性, 本文选择了 8 种变化检测网络进行对比分析, 分别为 FC-EF^[25]、FC-Siam-Diff^[25]、FC-Siam-Conc^[25]、DTCDSCN^[26]、BIT^[12]、HSAA-CD^[27]、ChangeMamba^[19] 及 RSM-CD^[18]。其中 FC-EF、FC-Siam-Diff、FC-Siam-Conc 和 DTCDSCN 为经典的 CNN 变化检测网络; BIT 为经典的 Transformer 变化检测网络, HSAA-CD 通过多尺度特征融合和残差操作优化特征边缘; ChangeMamba 和 RSM-CD 为最新提出的 Mamba 架构网络。基于 Mamba 架构的网络均未使用经过预训练的权重, 本文将 RSM-CD 作为本文的基准网络。

2.4.1 LEVIR-CD 数据集实验结果与分析

LEVIR-CD 数据集定量评价结果见表 1。由表 1 可以看出, 相较于 8 种对比方法, 本文提出的 DSPOSM 在 R 、 $F1$ 分数和 IoU 3 个指标上均取得了最佳结果, 比基准网络 RSM-CD 分别高出了 1.97%、0.52% 和 0.15%, P 仅次于 RSM-CD, 且基于 Mamba 的 RSM-CD 和 DSPOSM 均取得了较其他非 Mamba 网络更优异的表现。

表 1 LEVIR-CD 数据集上各网络变化检测的定量结果/%

Tab. 1 Quantitative Evaluation of Different Networks on LEVIR-CD Dataset/%

网络	P	R	$F1$	IoU
FC-EF ^[25]	86.91	80.17	83.40	71.53
FC-Siam-Diff ^[25]	88.46	83.31	86.31	75.91
FC-Siam-Conc ^[25]	87.99	76.77	83.69	71.96
DTCDSCN ^[26]	88.16	86.50	87.32	77.35
BIT ^[12]	88.67	88.66	87.63	79.51
HSAA-CD ^[27]	89.14	88.83	88.56	79.84
ChangeMamba ^[19]	90.87	88.45	89.80	81.49
RSM-CD ^[18]	93.79	88.71	90.64	83.26
DSPOSM	91.64	90.68	91.16	83.41

图 7 展示了不同对比网络在 LEVIR-CD 数据集中 6 组影像的变化检测结果。从图 7 可以看出, DSPOSM 预测的变化建筑物区域的边界更为完整、光滑。在影像 4 中存在极小的建筑物变化区域, 对比法的预测结果大多出现漏检或检测不准确的情况, DTCDSCN 未能检测到该变化区域, 其他方法也未能较为完整检测到变化, 仅有本文方法能够较准确地对变化区域进行检测。在影像 5 中, 当建筑物屋顶与地面光谱特征相似时, 仅 RSM-CD 出现了漏检情况, 且 ChangeMamba 和 DSPOSM 预测结果边界较为模糊。说明基于 Mamba 的变化检测网络在对光谱特征相似的区域较易混淆, 造成建筑物变化区域的漏检与误检。

2.4.2 WHU-CD 数据集实验结果与分析

WHU-CD 数据集定量结果见表 2。由表 2 可以看出, DSPOSM 的 IoU 最佳, R 次佳, 在 R 和 IoU 2 个指标上比基准网络 RSM-CD 分别高出了 0.11% 和 0.61%, P 和 $F1$ 分数仅次于 RSM-CD。基于 Mamba 的 ChangeMamba、RSM-CD 和 DSPOSM 在总体定量结果上显著高于非 Mamba 的网络, 此结果与表 1 中定量评价结果一致。由表 1 和表 2 可以看出, 在两个不同的数据集上, 基于 Mamba 的网络在特征提取方面不仅准确性高, 而且具有较强的鲁棒性, 能够在处理复杂地物类别时保持高效的检测性能。

在 WHU-CD 数据集中, 由于地物类别的复杂性增加, 导致变化检测的难度显著提升。为了进一步验证本文所提出网络的效果, 图 8 展示了更多的复杂类型建筑物区域中不同网络的变化检测结果对比。由图 8 可以看出, 总体上 DSPOSM 比大部分网络的变化检测结果更好。在影像 3 中, 区域存在大量密集分布的建筑物, 造成对比方法预测的变化区域大多存在漏检情况。ChangeMamba 和 RSM-CD 能够正确预测变化建筑物区域, 但建筑边缘的贴合情况较差, DSPOSM 预测的变化建筑物区域边缘与真值差异最小; 在影像 6 中, 前、后时相影像中道路和建筑的光谱特征相似, 在该较易混淆区域中, 基于 Mamba 的方法能准确定位区域并且正确检测区域内建筑物的变化, 识别出的变化区域更加完整连续。

2.4.3 消融实验

为进一步验证本文所提方法的有效性, 本文对提出的并行方法进行消融实验和分析。选择 RSM-CD 作为基准网络, 以下称为 Baseline,

Baseline中Mamba输入数据通道数为 C ;实验1在Baseline的基础上使用输入数据通道数变为 $C/2$ 的双流并行Mamba;实验2在Baseline的基础上使用输入数据通道数变为 $C/4$ 的多流并行

Mamba;实验3在Baseline的基础上使用输入数据通道数变为 $C/8$ 的多流并行Mamba。消融实验选择在WHU-CD数据集上进行,定量结果见表3。

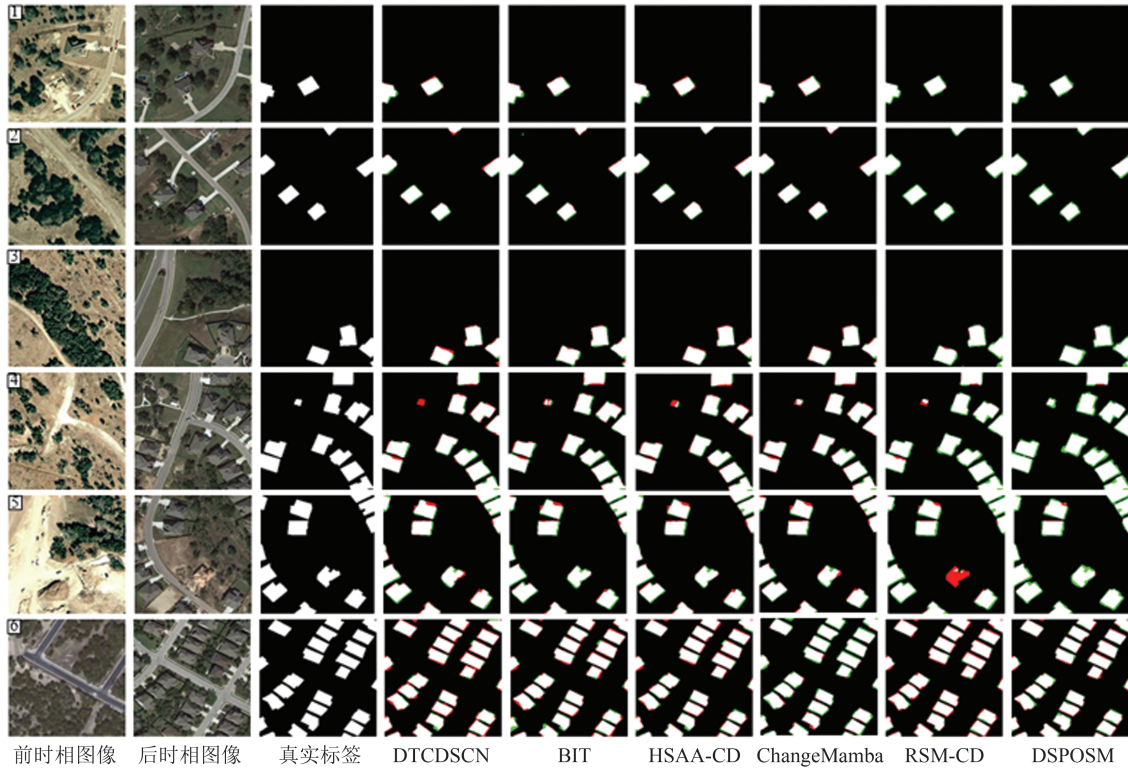


图7 LEVIR-CD数据集上各网络变化检测结果对比

Fig. 7 Experimental Results of Different Networks on LEVIR-CD Dataset

表2 WHU-CD数据集上各网络变化检测的定量结果/%

Tab. 2 Quantitative Evaluation of Different Networks on WHU-CD Dataset/%

网络	P	R	$F1$	IoU
FC-EF ^[25]	78.50	78.33	78.41	64.53
FC-Siam-Diff ^[25]	81.73	80.43	80.56	71.91
FC-Siam-Conc ^[25]	78.99	78.32	78.65	70.96
DTCDCSCN ^[26]	80.74	81.20	80.97	70.77
BIT ^[12]	84.98	82.64	83.77	72.90
HSAA-CD ^[27]	85.55	83.54	84.29	73.41
ChangeMamba ^[19]	93.07	89.70	91.35	84.08
RSM-CD ^[18]	93.58	87.59	90.49	86.88
DSPOSIM	93.40	87.70	90.46	87.49

由表3可以看出,实验2和实验3的参数量和计算量更低,但性能开始降低,其中实验3对比Baseline,参数量和计算量分别降低了48.99%、35.63%,但 $F1$ 分数和IoU分别下降了5.49%、4.45%;实验1对比Baseline,在参数量和计算量均降低的同时,仍保持了较好的性能,参数量和计算量分别下降了约35.8%和21.4%, $F1$ 分数仅

降低了0.03%,IoU指标提升了0.61%。因此,本文采用实验1的设置作为DSPOSIM块的结构。

2.4.4 网络复杂度及效率对比分析

采用网络参数量、计算量以及在不同数据集上训练200轮的时间作为评价指标,对本文网络与其他遥感影像变化检测网络进行定量评价,定量对比结果见表4。基于卷积的方法DTCDCSCN网络的参数量和计算复杂度接近于DSPOSIM,但该方法的检测性能远低于基于Mamba的方法;基于Transformer的网络BIT和HSAA-CD仅采用ResNet18作为编码器,整体参数量较小,但由于其注意力机制需要对所有像素进行运算,因此计算量在对比方法里最大,在数据集上训练时间花销最高;基于Mamba的网络RSM-CD和DSPOSIM训练时间显著降低,并且与RSM-CD相比,DSPOSIM网络的参数量下降了约35.8%,计算量下降了约21.4%,在WHU-CD和LEVIR-CD数据集上,训练时间分别下降了约19.45%和8.25%,表明本文方法能够显著降低基于Mamba的网络参数量与计算复杂度,提高训练网络的效率。

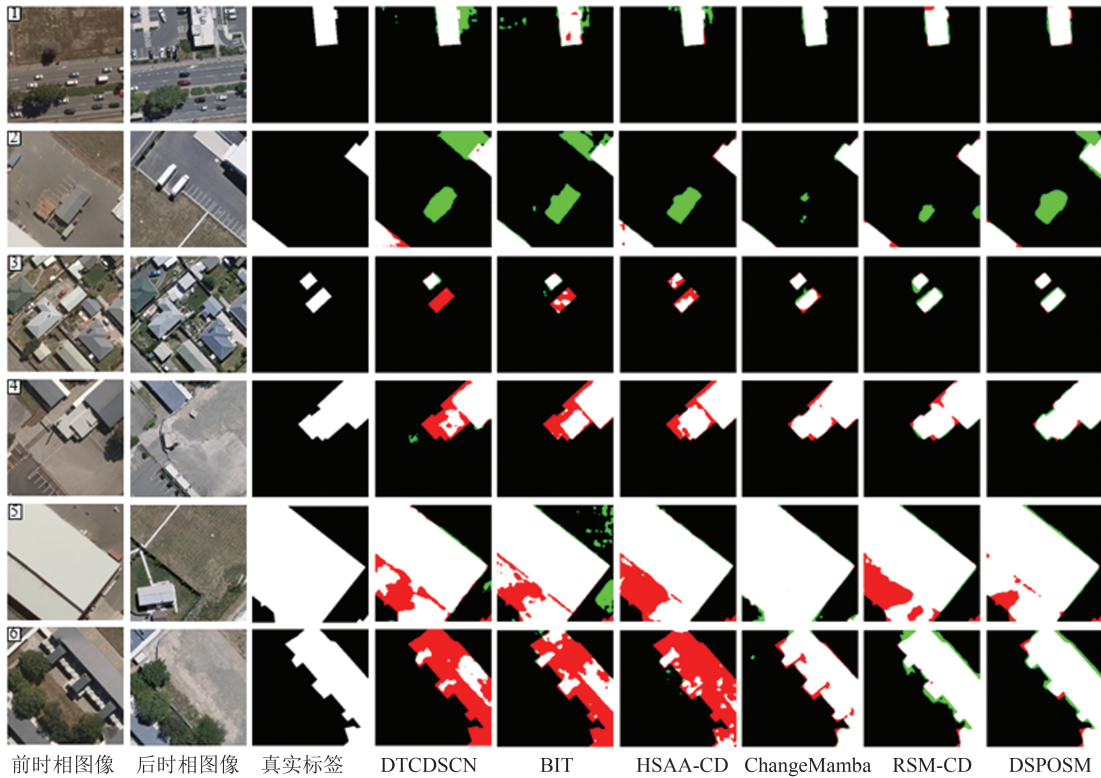


图 8 WHU-CD数据集上各网络变化检测结果对比

Fig. 8 Experimental Results of Different Networks on WHU-CD Dataset

表 3 不同并行方法在 WHU-CD 数据集上的消融实验

Tab. 3 Ablation Experiments of Different Parallel Methods on WHU-CD Dataset

实验	参数量/M	FLOPs/G	F1/%	IoU/%
Baseline	49.97	35.64	90.49	86.88
1	32.08	28.03	90.46	87.49
2	27.04	24.64	89.68	86.31
3	25.49	22.94	85.00	82.43

表 4 不同变化检测网络的网络复杂度和算法效率对比

Tab. 4 Comparison of Network Complexity and Algorithmic Efficiency of Different Change Detection Networks

网络	参数量/M	FLOPs/G	训练时间/h	
			WHU-CD	LEVIR-CD
FC-EF ^[25]	1.35	7.15	5.22	6.63
FC-Siam-Diff ^[25]	1.35	9.45	5.40	7.15
FC-Siam-Conc ^[25]	1.54	10.66	5.56	6.41
DTCDCSCN ^[26]	41.07	26.44	9.24	15.37
BIT ^[12]	11.50	53.07	11.92	17.25
HSAA-CD ^[27]	21.61	75.41	11.84	16.55
ChangeMamba ^[19]	53.96	57.41	11.67	15.86
RSM-CD ^[18]	49.97	35.64	8.33	11.52
DSPOSM	32.08	28.03	6.71	10.57

3 结 语

针对当前高分辨率遥感影像变化检测方法中存在的计算复杂度高、检测效率低下的问题, 本文采用 Mamba 框架构建了变化检测网络, 并对影响 Mamba 算法参数数量的关键因素进行了深入分析。基于此, 本文提出了双流并行全向扫描 Mamba 网络。本文方法在 LEVIR-CD 和 WHU-CD 两个遥感影像数据集上与其他变化检测网络进行对比验证, 实验结果表明本文方法展现出优异的性能, 在多个评价指标上取得了最佳结果, 证明了该方法在减少网络参数的同时, 仍能保持良好的检测性能。同时, 基于 Mamba 架构的方法 ChangeMamba、RSM-CD 和 DSPOSM 在复杂地物类别上的检测准确性高、鲁棒性强, 显示出 Mamba 架构在遥感影像变化检测领域的巨大潜力。通过消融实验进一步验证了本文方法的有效性, 并在网络复杂度及效率对比分析中, 展示了 DSPOSM 在降低参数量、计算复杂度和提高训练效率方面的优势。但本文方法会小幅影响模型在较为复杂类型的场景下的检测性能, 在未来工作中, 考虑进一步研究减少本文方法在复杂场景下对性能的影响, 以应对包括非农领域在内的更多复杂场景下的遥感影像变化检测需求。

参 考 文 献

- [1] Liu Yaolin, Tong Zhaomin, Liu Sui, et al. Progress and Prospects of Research on Optimal Land-Use Allocation Modeling[J]. *Geomatics and Information Science of Wuhan University*, 2022, 47(10): 1598-1614. (刘耀林, 全照民, 刘岁, 等. 土地利用优化配置建模研究进展与展望[J]. 武汉大学学报(信息科学版), 2022, 47(10): 1598-1614.)
- [2] Sui Haigang, Feng Wenqing, Li Wenzhuo, et al. Review of Change Detection Methods for Multi-temporal Remote Sensing Imagery[J]. *Geomatics and Information Science of Wuhan University*, 2018, 43(12): 1885-1898. (眭海刚, 冯文卿, 李文卓, 等. 多时相遥感影像变化检测方法综述[J]. 武汉大学学报(信息科学版), 2018, 43(12): 1885-1898.)
- [3] Guo H N, Shi Q, Marinoni A, et al. Deep Building Footprint Update Network: A Semi-supervised Method for Updating Existing Building Footprint from Bi-temporal Remote Sensing Images[J]. *Remote Sensing of Environment*, 2021, 264: 112589.
- [4] Wu C, Du B, Zhang L P. Slow Feature Analysis for Change Detection in Multispectral Imagery[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2014, 52(5): 2858-2874.
- [5] Nielsen A A. The Regularized Iteratively Reweighted MAD Method for Change Detection in Multi- and Hyper- Spectral Data[J]. *IEEE Transactions on Image Processing*, 2007, 16(2): 463-478.
- [6] Shen Shaohong, Wan Youchuan, Gong Hao, et al. An Adaptive Threshold Segmentation Method Based on Spatial Statistic Theory to High-Resolution Remote Sensing Change Detection [J]. *Geomatics and Information Science of Wuhan University*, 2009, 34(8): 902-905. (申邵洪, 万幼川, 龚浩, 等. 遥感影像变化检测自适应阈值分割的Kriging方法[J]. 武汉大学学报(信息科学版), 2009, 34(8): 902-905.)
- [7] Yuan Zhou, Guo Haitao, Lu Jun, et al. High-Resolution Remote Sensing Image Change Detection Technology Based on UNet++ and Attention Mechanism [J]. *Journal of Geomatics Science and Technology*, 2021, 38(2): 155-159. (袁洲, 郭海涛, 卢俊, 等. 融合UNet++网络和注意力机制的高分辨率遥感影像变化检测算法[J]. 测绘科学技术学报, 2021, 38(2): 155-159.)
- [8] Wang Y F, Yu X D, Wen X Y, et al. Learning a 3D-CNN and Convolution Transformers for Hyperspectral Image Classification [J]. *IEEE Geoscience and Remote Sensing Letters*, 2024, 21: 5504505.
- [9] Zhang Zhuoer, Pan Jun, Shu Qidi. Building Extraction Based on Dual-Stream Detail-Concerned Network[J]. *Geomatics and Information Science of Wuhan University*, 2024, 49(3): 376-388. (张卓尔, 潘俊, 舒奇迪. 基于双路细节关注网络的遥感影像建筑物提取[J]. 武汉大学学报(信息科学版), 2024, 49(3): 376-388.)
- [10] Gao Xianjun, Ran Shuhao, Zhang Guangbin, et al. Building Extraction Based on Multi-feature Fusion and Object-Boundary Joint Constraint Network[J]. *Geomatics and Information Science of Wuhan University*, 2024, 49(3): 355-365. (高贤君, 冉树浩, 张广斌, 等. 基于多特征融合与对象边界联合约束网络的建筑物提取[J]. 武汉大学学报(信息科学版), 2024, 49(3): 355-365.)
- [11] Wang Xin, Li Yingying, Zhang Xiangliang. Ghost-Guided UNet++ for High-Resolution Remote Sensing Image Change Detection [J]. *Journal of Image and Graphics*, 2024, 29(05): 1460-1478. (王鑫, 李莹莹, 张香梁. Ghost引导UNet++的高分遥感影像变化检测[J]. 中国图象图形学报, 2024, 29(05): 1460-1478.)
- [12] Chen H, Qi Z P, Shi Z W. Remote Sensing Image Change Detection with Transformers [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 5607514.
- [13] Li Tao, Wang Hairui. Remote Sensing Image Change Detection of Construction Land Based on Siamese Attention Network[J]. *Computer Science*, 2023, 50(6A): 220500040-5. (李滔, 王海瑞. 基于孪生注意力网络的建设用地遥感影像变化检测[J]. 计算机科学, 2023, 50(6A): 220500040-5.)
- [14] Gu A, Goel K, Ré C. Efficiently Modeling Long Sequences with Structured State Spaces [EB/OL]. [2021-01-21] <https://arxiv.org/abs/2111.00396v3>.
- [15] Gu A, Dao T. Mamba: Linear-Time Sequence Modeling with Selective State Spaces [EB/OL]. [2024-06-08] <https://arxiv.org/abs/2312.00752>.
- [16] Liu Y, Tian Y, Zhao Y, et al. Vmamba: Visual State Space Model [EB/OL]. [2024-06-08] <https://arxiv.org/abs/2401.10166>.
- [17] Yu W H, Wang X C. MambaOut: Do We Really Need Mamba for Vision? [EB/OL]. [2024-06-08] <https://arxiv.org/abs/2405.07992>.
- [18] Zhao S J, Chen H, Zhang X L, et al. RS-Mamba for Large Remote Sensing Image Dense Prediction [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 5633314.
- [19] Chen H, Song J, Han C X, et al. ChangeMamba: Remote Sensing Change Detection with Spatiotem-

- poral State Space Model[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 4409720.
- [20] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An Image Is Worth 16×16 Words: Transformers for Image Recognition at Scale [J]. ArXiv e-Prints, 2020: arXiv: 2010.11929.
- [21] Elfving S, Uchibe E, Doya K. Sigmoid-Weighted Linear Units for Neural Network Function Approximation in Reinforcement Learning [J]. *Neural Networks*, 2018, 107: 3-11.
- [22] Chen H, Shi Z W. A Spatial-Temporal Attention-Based Method and a New Dataset for Remote Sensing Image Change Detection [J]. *Remote Sensing*, 2020, 12(10): 1662.
- [23] Ji S P, Wei S Q, Lu M. Fully Convolutional Networks for Multisource Building Extraction from an Open Aerial and Satellite Imagery Data Set [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(1): 574-586.
- [24] Kingma D P, Ba J. Adam: A Method for Stochastic Optimization[EB/OL]. [2024-06-08] <https://arxiv.org/abs/1412.6980v9>
- [25] Caye Daudt R, Le Saux B, Boulch A. Fully Convolutional Siamese Networks for Change Detection [C]//The 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 2018.
- [26] Liu Y, Pang C, Zhan Z Q, et al. Building Change Detection for Remote Sensing Images Using a Dual-Task Constrained Deep Siamese Convolutional Network Model [J]. *IEEE Geoscience and Remote Sensing Letters*, 2021, 18(5): 811-815.
- [27] Li F T, Zhou F D, Zhang G, et al. HSAA-CD: A Hierarchical Semantic Aggregation Mechanism and Attention Module for Non-agricultural Change Detection in Cultivated Land [J]. *Remote Sensing*, 2024, 16(8): 1372.