arXiv:2603.01161v1 [cs.CV] 1 Mar 2026

GRAD-Former：用于变化检测的门控鲁棒注意力差分 Transformer

Durgesh Ameta*、Ujjwal Mishra*、Praful Hambarde 和 Amit Shukla，IEEE 会员

摘要—遥感变化检测（CD）旨在识别不同时期拍摄的卫星图像之间的语义差异。尽管深度学习极大地推动了这一领域的发展，但现有基于卷积神经网络（CNN）、Transformer 和选择性状态空间模型（SSM）的方法，仍难以精确描绘变化区域。尤其是，传统的基于 Transformer 的方法应用于超高分辨率（VHR）卫星图像时，会面临二次计算复杂度；而且在训练数据有限时往往表现不佳，导致 VHR 图像中丰富的空间信息未得到充分利用。我们提出了 GRAD-Former，这是一种新型框架，能够在缩减模型大小、保持高效的同时增强上下文理解能力。该框架由一个新型编码器、自适应特征相关性与细化（AFRAR）模块、融合模块和解码器模块组成。AFRAR 通过两个提出的组件实现全局与局部上下文感知：选择性嵌入增强（SEA）模块和全局-局部特征细化（GLFR）模块。SEA 和 GLFR 分别利用门控机制和差分注意力，生成多个 softmax 堆，以捕捉重要特征，同时尽量减少对无关特征的捕捉。在三个具有挑战性的 CD 数据集（LEVIR-CD、CDD、DSIFN-CD）上进行的多项实验表明，与现有方法相比，GRAD-Former 的性能更优。值得注意的是，GRAD-Former 使用更少参数的同时，在所有数据集的全部指标上均优于当前最先进的模型。我们的框架为遥感变化检测性能树立了新的基准。代码将发布于：https://github.com/Ujjwal238/GRAD-Former

索引词——变化检测（CD）、高分辨率遥感图像、差分注意力、Transformer。

I. 引言

THE 地球表面在自然过程和人类活动的共同作用下持续变化，因此需要一种可靠、系统的监测方法来评估这些变化 [1]。变化检测（CD）是利用遥感多时相数据监测某一地理区域，识别并描述感兴趣现象或对象的变化特征，并对差异进行定量分析。变化检测是城市土地管理 [2], [3], 、可持续资源



图1. 示例图像展示了变化检测任务中的不同挑战及其真实变化掩膜。尤其难以忽略无关变化（以红色框标出），例如（a）阴影和光照差异、（b）季节变化，以及（c）行驶中的汽车和屋顶变化，同时准确检测（d）大尺度与小尺度的有意义的变化（以绿色框标出）。

利用 [4], 和应急响应决策 [5],等领域的重要工具，可为传统的人工现场调查提供更高效的替代方案。

卫星技术的进步提升了遥感数据的质量和多样性。然而，现有模型和技术的性能提升并未跟上这些进步。随着分辨率提高，噪声和背景信息变得更加明显 [6]。此外，如图1(a)和图1(b)所示，双时相数据中的图像会受到照明干扰和季节变化的影响。此外，可移动物体（如车辆）的位置变化、屋顶颜色的改变，以及变化区域大小的显著差异也很常见；多个相邻变化区域会导致特征高度不一致，如图1(c)和图1(d)所示，这使得区分相关特征与无关特征极具挑战性。

因此，为了准确识别变化区域，研究人员提出了许多具有代表性的传统变化检测算法，例如基于判别投影的方法 [7], 慢特征分析 [8], 变化向量分析 [9] 以及基于变换的技术 [10]。然而，这些经典方法未能捕捉多样化的特征，也忽略了空间上下文 [11]。这促使基于深度学习的

*这些作者对本研究做出了同等贡献。Durgesh Ameta就职于印度知识体系与心理健康应用（IKSMHA）中心，IIT Mandi，曼迪，喜马偕尔邦 175005，印度（durgesgameta@gmail.com）。Ujjwal Mishra就职于计算学院，IIIT Una，喜马偕尔邦177209，印度（ujjwalmishra238@gmail.com）。Praful Hambarde与Amit Shukla就职于人工智能与机器人中心（CAIR），IIT Mandi，曼迪，喜马偕尔邦 175005，印度（praful@iitmandi.ac.in，amitshukla@iitmandi.ac.in）。

模型不断发展，这些模型采用卷积自编码器（CAEs）[12], 基于像素的深度神经网络（DNNs） [13], 循环神经网络（RNNs） [14], 卷积神经网络（CNNs） [15] ，并展现出对数据噪声的鲁棒性，且能够高效获取时空上下文信息。

随着深度学习方法仍占据主导地位，孪生架构继续被广泛用于解决变化检测问题[16]。通常，这些方法会将多尺度特征提取、空洞卷积和注意力机制等架构要素纳入其框架。基于CNN的架构 [17], [18], [19]长期以来一直是变化检测任务的标准方法。然而，它们最近正被使用注意力机制 [23]的基于变换器的架构 [20], [21], [22]所取代，以解决卷积神经网络在捕获全局上下文表示方面的困难。尽管简单的基于变换器的架构相较于卷积神经网络具有优势，但也存在显著局限。它们难以捕获局部上下文信息，而这对于沿复杂且不规则边界准确分割细微变化至关重要。此外，其计算复杂度呈二次增长，导致参数数量庞大且内存需求很高。这使得它们不太适合实际的遥感变化检测应用。因此，一种最优方法应致力于消除不必要的特征，同时捕获重要的局部和全局上下文信息，以便以最少的参数检测图像对之间的显著变化和结构变化。

为应对上述挑战，我们提出了 GRAD-Former，这是一种高效的变化检测框架。我们设计的关键创新在于自适应特征相关性与精炼（AFRAR）模块，它能有效滤除甚高分辨率（VHR）卫星图像中的噪声和无关背景信息，使模型能够专注于关键的局部细节和全局上下文细节。我们的方法充分利用高分辨率光学影像，以高精度捕获细粒度局部特征和广泛的全局模式。

为此，AFRAR模块配备了选择性嵌入增强模块（SEA），该模块采用门控技术进行稳健的特征提取，确保仅保留相关信息。此外，我们的全局-局部特征精炼（GLFR）模块利用差分注意力 [24] 提取全局上下文信息，同时最大限度地减少噪声并降低计算开销。我们在三个不同的变化检测基准数据集上对 GRAD-Former 进行了严格评估，结果表明其优于最先进（SOTA）方法。公开数据集上的出色定量和定性结果证明，我们的模型达到了同类最佳性能。

本研究的主要贡献总结如下：

- 我们提出了 GRAD-Former，这是一种稳健的基于孪生网络的变化检测框架，能够有效抑制噪声和无关背景信息，从而准确检测双时相卫星图像之间的语义差异。
- 我们在 AFRAR模块中的 GLFR模块和 SEA模块中引入了结合门控机制的差分注意力概念，用于变化检测；此外，还在差分融合（DA）模块中采用多尺度的基于差异的融合，将差异特征与编码特征相结合，以增强对变化区域的关注。
- 在三个公开变化检测数据集上的大量实验验证了我们方法的有效性，展示了其鲁棒性与效率，并取得了最先进的性能。

本文其余部分组织如下。第二节概述相关工作。第三节详细介绍为变化检测提出的GRAD-Former架构。模型的实验分析在第四节给出。最后，第五节对本文进行总结。

II. 相关工作

在本节中，我们回顾变化检测任务的多种方法，包括基于CNN、基于 Transformer以及基于状态空间模型（SSM）的方法，并讨论它们各自的优势与局限。

A. 基于CNN的变化检测方法

在各种深度学习架构中，卷积神经网络（CNN）凭借其能够有效提取图像中的空间特征和层次特征，已成为一种强大的工具。卷积神经网络通过卷积滤波器擅长捕捉局部信息，因此特别适用于语义分割和遥感变化检测（CD）等像素级任务。这一能力使其在计算机视觉的众多领域得到广泛应用，包括医学图像分割 [25],目标检测 [26], 以及环境监测[27]。

早期方法采用了全卷积网络（FCN）和 U-Net 架构 [28], , 通过使用编码器-解码器结构提取多尺度空间特征，为后续研究奠定了坚实基础。然而，U-Net 由于采用线性压缩而遭遇特征丢失。为解决这一问题，研究者采用了诸如 UNet++ [29]中的跳跃连接等改进方法，以优化特征融合并保留特征。SNUnet [30]。基于 U-Net 框架的 BiDateNet [31], 通过整合多尺度特征表示，提升了变化检测性能。

基于CNN的变化检测模型的发展见证了孪生网络的出现。孪生网络采用双流架构，能够有效处理双时相遥感图像。Daudt 等人 [32] 提出了三种基于孪生网络的模型：FC-EF、FC-Siam-Conc 和 FC-Siam- Di，每种模型均采用不同的融合策略，以有效比较时间变化。此外，Khan 等人 [33]提出了一种基于 CNN 的弱监督方法，无需密集标注数据即可直接定位变化检测区域。

尽管具有诸多优势，基于CNN的方法仍存在显著局限。它们难以捕捉长程

依赖关系 [34], , 因此难以有效建模时间和空间关系。此外，它们对照明变化、云层覆盖和环境条件的敏感性，也给现实世界中的变化检测任务带来了挑战。

B. 基于 Transformer 的变化检测方法

Transformer 推动了计算机视觉（CV）的显著发展，在图像分类、分割和目标检测等多种任务中表现出色。其强大的全局上下文关系建模能力，使其尤其适用于遥感（RS）变化检测（CD）任务 [35]。与主要捕获局部空间结构的卷积神经网络不同，Transformer 通过自注意力机制聚合局部和全局特征，从而提升双时相图像间的特征表征能力 [36]。

早期基于 Transformer 的变化检测模型由类 UNet 架构发展而来，并利用层次化特征提取。例如，SwinSUNet [35] 将层次化 Swin Transformer 结构集成到孪生 UNet 框架中，以增强局部和全局特征融合。但这种方法使用了大量参数。为提升多尺度特征利用率，TFI-GR [37] 和 A2Net[38] 等方法采用轻量级的基于 Transformer 的架构，并通过监督注意力机制突出变化区域。与此同时，双时相注意力 Transformer [39] 利用交叉注意力机制，优化双时相特征之间的交互，从而更准确地定位变化。

其他方法结合了卷积神经网络与 Transformer 的优势。Hybrid Transformer [40] 和 MFiNet [41], 等混合网络将卷积神经网络的局部特征提取与 Transformer 的全局上下文建模相结合，从而取得了更优的变化检测性能。Noman 等人在 ELGC-Net [42]中通过在单个模块中集成卷积神经网络和 PT 注意力来实现这一点，两者各处理一半的通道。ChangeFormer[21] 增强了对长期依赖关系的建模，确保在多个时间尺度上进行稳健的特征提取。此外，Trans-MAD [43] 和 WSMsFNet[44] 等方法集成了基于 Transformer 的空间与通道注意力机制，以改进对变化区域的感知。Wu 等人近期的工作 CGMNet [45] 探索了用于高光谱图像变化检测任务的基于门控机制的注意力网络。近期的变化检测研究还进一步探索了即插即用框架的开发。具体而言，AdvCP [46] 提出了一种对抗式类别提示方案，旨在弱监督学习场景下，利用基于 SAM（Segment Anything Model）的基线模型，揭示图像级分类模型所学习到的共现像素级表征。

尽管 Transformer 能够捕捉长程依赖关系，但由于 VHR 双时相图像中存在过多噪声和高度冗余的信息，它们难以从中找到相关特征。

C. 基于状态空间模型的变化检测方法

状态空间模型（SSM）由控制系统理论发展而来，最初用于对线性时不变系统中的输入响应关系进行建模。它们被引入计算机视觉和遥感源于高效处理长程序列数据、同时缓解传统模型计算复杂度的需求。与因自注意力而呈现二次复杂度的Transformer不同，诸如Mamba之类的状态空间模型具有线性时间复杂度，因而在大规模视觉任务中具有高度可扩展性 [47]。

Mamba 及其变体已成功应用于多种任务，包括图像分类、目标检测和医学图像分割。Vision Mamba（VMamba）通过引入四向扫描机制（左上、右下、右上、左下），扩展了 Mamba 的能力，使其能够在保持计算效率的同时，高效遍历空间特征 [48]。类似地，RS-Mamba 引入多方向扫描方法，以增强遥感图像中的特征提取 [49]。

在变化检测（CD）领域，CDMamba [50] 和 LCCDMamba [51] 等基于 Mamba 的方法已被开发出来，以更好地捕捉双时相图像之间的时空差异。CDMamba 采用残差结构和局部-全局特征提取来有效检测变化，而 LCCDMamba 进一步融合多尺度融合技术，以保留高维语义。ChangeMamba 基于 VMamba 的扫描机制，增强时空特征关系，在密集预测任务中取得了显著进展 [52]。近期的进展包括 SpectralMamba [53] 和 S2Mamba [54], , 它们将光谱-空间融合技术用于高光谱变化检测；此外，CM-UNet [55] 和 RS3Mamba [56], 则增强了不同时间帧之间的特征对齐，从而提高精度。此外，LocalMamba [57] 等新兴研究致力于动态优化不同层的扫描方案，以解决效率与细粒度特征提取之间的权衡。Qu等人在其近期工作 MambaFedCD [58] 中，探索了采用状态空间模型骨干网络进行变化检测任务的联邦主动学习策略。

尽管取得了这些进展，SSM 仍存在明显的局限性。虽然它们擅长捕获长程依赖，但在精细的局部特征提取方面往往表现不佳，而这对于准确的变化检测至关重要。现有的大多数方法侧重于改进扫描技术或利用预训练骨干网络，而非解决局部表征固有的挑战。此外，基于 SSM 的模型在需要精确边界 delineation 的任务中可能存在敏感性问题，从而限制了其在高分辨率遥感中的适用性。

III. 方法

我们的基础框架是一个基于 Transformer 的孪生网络。它由三个主要部分组成：编码器、融合模块和解码器 [36]。孪生编码器处理变化前后图像，并在四个阶段提取多尺度特征图。每个阶段提取的特征图分别记为 $\hat{\mathrm{F}}_{pre}^{i}$ 和 $\hat{\mathrm{F}}_{post}^{i}$，其中 $i \in [1,2,3,4]$，随后传入融合模块，以实现双时相特征交互。所有阶段融合后的特征均输入

解码器，并通过多个卷积层和转置卷积层进行组合与细化，生成高分辨率特征图。随后，预测层对这些上采样后的特征图进行处理，生成最终变化图。

此外，本节描述了网络的总体架构，并详细解释了AFRAR模块。

A. 总体架构

所提出的变化检测（CD）框架（称为GRAD-Former）的架构如图2(a)所示。与基础框架类似，GRAD-Former由编码器、融合模块和解码器组成。编码器以变化前和变化后卫星图像作为输入，并生成多尺度特征图$\hat{\mathrm{F}}_{pre}^{i}$和$\hat{\mathrm{F}}_{post}^{i}$，其中$i \in [1,2,3,4]$。在每个阶段，特征图首先通过图块嵌入层进行下采样，然后经过一系列编码器块进行处理。编码器块的结构如图2(c)所示，它包含所提出的新型自适应特征相关性与细化（AFRAR）模块以及卷积MLP层；在将输入传递给AFRAR模块和卷积MLP层之前，会在通道上执行实例归一化。图2(d)详细描绘了所提出的AFRAR模块，其中包含用于相关特征提取的SEA和GLFR模块，并将它们各自的输出进行拼接以得到$\bar{\mathrm{F}}_{norm}^{i}$。

在每个阶段，与基线框架一致，从编码器块中提取的特征对通过所提出的新型差分融合（DA）模块进行融合。如图3所示，DA模块通过沿通道维度拼接$\hat{\mathrm{F}}_{pre}^{i}$、$\hat{\mathrm{F}}_{post}^{i}$及二者之差，对语义与差分特征进行编码。随后，DA模块使用卷积层减少通道数，再经过激活层，得到$\hat{\mathrm{F}}_{fuse}^{i}$，其中$i \in [1,2,3,4]$作为输出。可概括为：

$$\hat{\mathcal{F}}_{\mathrm{fuse}}^{i} = \sigma\left(\mathrm{Conv}_{1\times1}\left(\mathrm{Conc}\left(\hat{\mathcal{F}}_{\mathrm{pre}}^{i},\ \hat{\mathcal{F}}_{\mathrm{post}}^{i},\ [\hat{\mathcal{F}}_{\mathrm{post}}^{i} - \hat{\mathcal{F}}_{\mathrm{pre}}^{i}]\right)\right)\right) \quad (1)$$

其中，$Conc$ ($Concatination$)沿通道维度进行，$\sigma(\cdot)$表示GELU激活函数。

The随后，将F_{fuse}^{i}输入解码器，以预测二值变化图（M）。

解码器架构：如图2(b)所示，GRAD-Former 采用由多个卷积层和转置卷积层组成的解码器。来自四个阶段的多尺度融合特征（$\hat{\mathrm{F}}_{fuse}^{i}$where $i \in [1,2,3,4]$）沿通道维度拼接后，输入一个 1×1卷积层。随后，应用转置卷积逐步提高特征图的空间分辨率。为进一步细化这些特征，加入一个包含两个 3×3 卷积层的残差块。转置卷积与残差块的组合应用两次，以恢复空间维度，使其与输入大小相匹配。最后，通过一个卷积层生成预测分数，其包含两个输出通道：一个表示未变化类别，另一个表示变化类别。沿通道维度执行 argmax 操作，得到最终的二值变化图。

接下来，我们将详细介绍自适应特征相关性与细化（AFRAR）模块。

B. 自适应特征相关性与细化（AFRAR）模块

AFRAR模块旨在高效捕获相关的全局和局部上下文信息，同时忽略噪声及无关的背景信息，并将计算复杂度降至低于传统自注意力的水平。受 Inception 模块 [59], 的启发（该模块采用不同卷积核尺寸的并行卷积操作），我们的方法将输入通道划分为不同的组，并对每个分组分别应用不同的操作（见图2(d)）。具体而言，输入特征按通道划分，并通过两个独立的上下文聚合分支进行处理，以提取重要的全局和局部上下文表征。AFRAR模块的核心包含两个新颖模块：选择性嵌入增强（SEA）（图4(b)）和全局-局部特征细化（GLFR）（图4(a)），它们分别专注于筛选和捕获关键信息以及全局-局部上下文。

设$\mathrm{F}^{i} \in \mathrm{R}^{H^i \times W^i \times C^i}$为编码器块的输入特征。经过归一化后，它变为$\mathrm{F}_{norm}^{i}$，并作为 i^{th} 阶段 AFRAR 的输入，其中（H^i，W^i，C^i）分别表示输入图像的高度、宽度和通道数。

首先，我们对归一化后的输入特征$\mathcal{F}^{i}$范数按通道拆分，得到$\mathcal{F}^{i}$GLFR和$\mathcal{F}^{i}$SEA，其中$\mathcal{F}^{i}$GLFR、$\mathcal{F}^{i}$SEA$\in \mathcal{R}^{H^i \times W^i \times \frac{C^i}{2}}$。它们分别作为 SEA 模块和 GLFR 模块的输入。也就是说，

$$\mathcal{F}_{\mathrm{GLFR}}^{i}, \mathcal{F}_{\mathrm{SEA}}^{i} = \mathrm{Split}(\mathcal{F}_{norm}^{i}). \quad (2)$$

接下来，我们将详细介绍 AFRAR 模块中的 SEA 模块和 GLFR 模块。

1）选择性嵌入增强（SEA）模块：受 [60], [61], [62], [63] 启发，这些方法在解码器模块中采用门控机制，以提升从输入映射到输出的质量，但它们未能利用输入本身的关键特征。为了增强通道特征的表达能力，同时尽量减少模型参数的数量，我们在编码器块中引入 SEA 模块，以最大限度地提升所捕获信息的质量。输入特征$\mathrm{F}_{\mathrm{SEA}}^{i}$ 首先经过 $L2$ 归一化，然后与一个可学习参数 α 相乘。该参数用于学习和评估各通道的重要性。在 L2 范数处理过程中，引入常数 ϵ（$0 \le \epsilon \le 10^{-5}$）以防止数值不稳定性。该嵌入（$E$）表示为：

$$E = \alpha \cdot \sqrt{\sum_{H,W} {\mathcal{F}_{\mathrm{SEA}}^{i}}^{2} + \epsilon} \quad (3)$$

计算嵌入 E 后，我们求得 RMS 的值 E，并引入另一个可学习参数 γ，用于学习

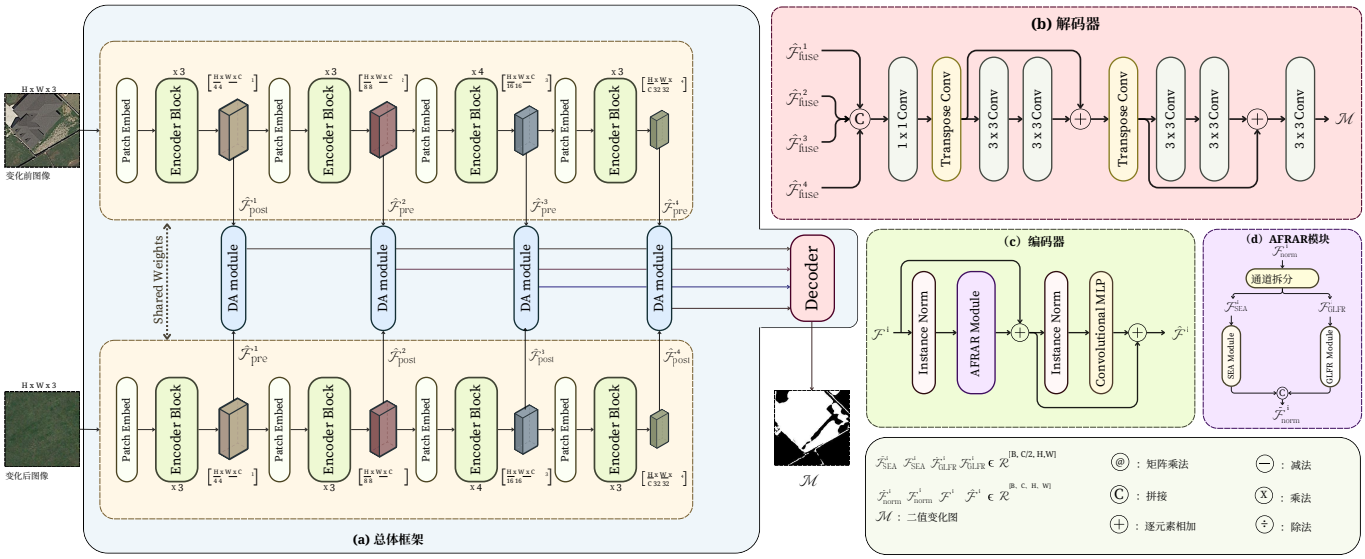


图2. 所提出的 GRAD-Former 的整体设计如下所示：(a) 概述了完整的网络结构，展示了通过共享编码器处理变化前和变化后的输入图像。该编码器分四个阶段提取特征，随后使用 DA模块对这些特征进行融合。在每个阶段，融合后的特征图都会在解码器中进一步合并；解码器采用多个卷积层和转置卷积层生成上采样特征。这些上采样特征图最终在预测层中用于生成最终的变化图。(b) 解码器的详细架构：它通过沿通道维度拼接，从四个阶段提取融合后的特征图 $\hat{F}^{i}_{fuse}$（$i = 1, 2, 3, 4$）。随后，它应用一个 1×1 卷积对特征进行投影。接着，通过两次上采样将空间维度恢复至与模型输入相匹配，并使用一个级联转置卷积后接一个由两个卷积层组成的残差块。最后，采用一个卷积层生成预测分数。(c) 展示了编码器块结构，其中包含新颖的自适应特征相关性与细化（AFRAR）模块以及卷积MLP层。在输入送入AFRAR层和MLP层之前，先沿通道进行实例归一化。(d) 展示了AFRAR模块的详细架构：首先沿通道拆分输入F^{i}_{norm}，然后结合SEA和GLFR模块提取相关特征，并将其输出拼接形成 $\bar{F}^{i}_{norm}$。

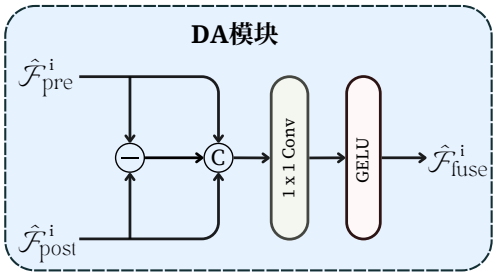


图3. 所提出的差异融合（DA）模块的详细示意图。该模块通过沿通道维度拼接$\hat{F}^{i}_{pre}$和$\hat{F}^{i}_{post}$及其差值，捕获语义与差分特征。随后，在DA模块中通过卷积层减少通道数，并接入激活层，得到$\hat{F}^{i}_{fuse}$，其中 $i \in [1, 2, 3, 4]$。))。

通道之间的竞争关系。我们将此 γ 除以计算得到的均方根，以获得归一化因子 N。

$$N = \frac{\gamma}{\sqrt{\mathbb{E}_c[E^2] + \epsilon}} \tag{4}$$

其中 $\mathbb{E}_c[\cdot]$ 表示通道维度上的均值。引入ϵ 以确保数值稳定性。随后引入另一个可学习参数 β ，以增强重要通道的贡献并降低无关通道的影响。在门控函数中使用该 β ，如下所示：

$$G = 1 + \tanh(E \cdot N + \beta) \tag{5}$$

其中， G 是门控函数的输出。选择 $1 + tanh(x)$ 作为门控函数，可以对通道权重进行非线性调整，从而避免梯度爆炸。最后，将 G 与输入 F^{i} 逐元素相乘，得到 SEA $\bar{\mathrm{F}}^{i}_{SEA}$ 的输出。

$$\bar{\mathcal{F}}^{i}_{SEA} = \mathcal{F}^{i}_{SEA} \cdot G \tag{6}$$

该 G 矩阵包含所有重要且相关的特征，因此这些特征在输出中会得到增强，从而提高模型的鲁棒性。由于甚高分辨率图像可能包含大量噪声，而相关信息可能非常稀疏，因此有必要对这些特征进行编码。

*2)*全局-局部特征细化（GLFR）模块：另一半的 $\mathcal{F}^{i}$范数由该模块处理，得到 $\mathcal{F}^{i}{}_{\mathrm{GLFR}} \in \mathcal{R}^{H^i \times W^i \times \frac{C^i}{2}}$。现有基于Transformer的变化检测方法的一个关键问题是，它们往往会过度分散注意力，经常聚焦于无关信息，从而影响性能，导致处理速度变慢、输出无关信息。随着卫星图像质量提高，由于额外背景信息和噪声的加入，这一问题会愈发严重。为解决这一问题，我们提出了GLFR模块。受[24], 启发，我们采用多头差分注意力来有效捕获全局相关信息。为此，我们首先通过以下方式得到查询（Q）、键（K）和值（V）矩阵：

$$Q = W_Q \mathcal{F}^{i}_{\mathrm{GLFR}}, \quad K = W_K \mathcal{F}^{i}_{\mathrm{GLFR}}, \quad V = W_V \mathcal{F}^{i}_{\mathrm{GLFR}} \tag{7}$$

其中 Q, $K \in \mathrm{R}^{B \times C/2 \times H \times W}$和 $V \in \mathrm{R}^{B \times C/4 \times H \times W}$。在传统Transformer中，注意力机制通过单个Softmax函数将关注分配到序列中的所有词元。这往往会使重要内容被无关上下文淹没，尤其是在较长或复杂的输入中。

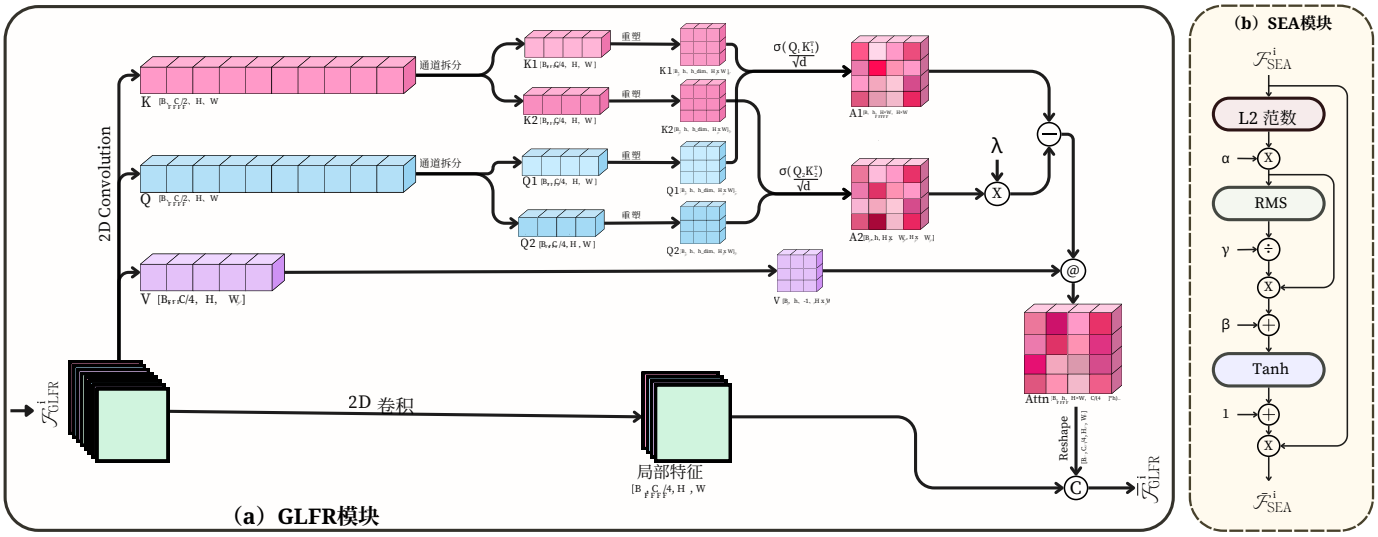


图4。（a）所提出的全局-局部特征细化（GLFR）模块概览：该注意力机制通过实现差分多头注意力，旨在克服Transformer中注意力焦点分散的问题。该模块根据输入$\mathrm{F}^{i}_{\mathrm{GLFR}}$生成 Q、 K、 V矩阵，并将 Q、 K拆分成成对矩阵，以计算双重Softmax注意力图（A_1、 A_2）。通过计算差值 $A = A_1 - \lambda \cdot A_2$并结合可学习缩放因子 λ，该模块生成稀疏注意力模式，仅关注相关特征，同时滤除噪声。所得注意力输出与局部特征相结合，在无需承担Transformer架构通常所需计算开销的情况下，实现全局上下文与局部细节之间的高效平衡。（b）展示了所提出的选择性嵌入增强（SEA）模块，该模块通过门控机制选择性地放大重要特征。输入特征$\mathrm{F}^{i}_{\mathrm{SEA}}$经过 $L2$归一化，并与可学习参数 α相乘。利用另一个可学习参数 γ计算归一化因子。随后，门控函数（$G(x) = 1 +$ 双曲正切$(x + \beta)$）根据每个通道的重要性对其进行自适应加权，其中 β是另一个可学习标量。最终输出$\bar{\mathrm{F}}^{i}_{\mathrm{SEA}} = \mathrm{F}^{i}_{\mathrm{SEA}} \cdot G$增强相关特征，同时抑制噪声，使模型能够抵抗高分辨率卫星影像中信息稀疏性的影响。

为应对这一问题，我们计算两个Softmax映射：一个用于聚焦相关词元，另一个用于表示输入中的噪声或干扰；二者之差将产生稀疏注意力模式，使注意力分数仅集中在相关词元上，从而使模型更加高效和准确。具体做法是沿通道将 Q和 K矩阵分别拆分为 Q_1， Q_2和 K_1， K_2，然后按如下方式计算相应的注意力分数：

$$Q_1, Q_2 = \mathrm{split}(Q), \quad K_1, K_2 = \mathrm{split}(K) \tag{8}$$

这里 $Q_1, Q_2, K_1, K2 \in \mathrm{R}^{B\times C/4\times H\times W}$随后重塑为适用于注意力的 $\mathrm{R}^{B\times h\times h_{dim}\times (H\times W)}$，其中 h表示注意力头的数量，在我们的模型中设为 4；h_{dim}表示每个注意力头的维度（通道数），并根据输入调整如下：

$$h_{dim} = \left\lfloor \frac{input\ dimension}{4 \times head} \right\rfloor \tag{9}$$

值（V）矩阵也会重塑为 $V \in \mathrm{R}^{B\times h\times -1\times (H\times W)}$。我们分别计算 Q_1的 Softmax堆 A_1 和 A_2，以及 K_1 的 Softmax堆 $Q_2 K2$：

$$A_1 = \sigma\left(\frac{Q_1^T K_1}{\sqrt{h_{dim}}}\right), \quad A_2 = \sigma\left(\frac{Q_2^T K_2}{\sqrt{h_{dim}}}\right) \tag{10}$$

其中 $\sigma(x)$ 是 Softmax函数。现在，得到的 Softmax堆之间的差值应包含相关信息，同时抑制重叠噪声。这种相减方式类似于降噪耳机通过抵消共同的声波来消除环境噪声。最终的 Softmax堆（A）计算如下：

$$\mathrm{A} = (A_1 - \lambda \cdot (A_2)) \tag{11}$$

其中 $A \in \mathrm{R}^{B\times h\times (H\times W)\times (H\times W)}$。 λ 是一个可学习标量，用于控制第二个 softmax 映射（表示噪声）的影响。其计算方式如下：

$$\lambda = \lambda_1 - \lambda_2 + \lambda_{\mathrm{init}} \tag{12}$$

其中，

$$\lambda_1 = e^{(\lambda_{q_1} \cdot \lambda_{k_1})} \tag{13}$$

$$\lambda_2 = e^{(\lambda_{q_2} \cdot \lambda_{k_2})} \tag{14}$$

其中， $\lambda_{\mathrm{init}} = 0.8$ 和λ_{q_1}， λ_{k_1}， λ_{q_2}， $\lambda_{k_2} \in \mathrm{R}^{h_{dim}}$是可学习向量。最后，我们求得 $Attn$ 矩阵：

$$Attn = \mathrm{A}@V^T \tag{15}$$

其中 @ 表示矩阵乘法运算。随后将 $Attn$ 重塑为 $\mathrm{R}^{B\times C/4\times H\times W}$，并与通过以下方式获得的局部特征（$l_{feat}$）拼接

$$l_{feat} = reshape(\mathcal{F}^{i}_{\mathrm{GLFR}}) \tag{16}$$

$l_{feat} \in \mathrm{R}^{B\times C/4\times H\times W}$。我们通过仅在大幅缩减的通道尺寸上计算注意力（见图4(a)），并有效地执行该计算，克服了Transformer的计算开销问题，从而获得更稳健且无误差的结果。该模块的输出 $\bar{\mathrm{F}}^{i}$GLFR由沿通道拼接 l_{feat}和 $Attn$ 得到。

$$\bar{\mathcal{F}}^{i}_{GLFR} = \mathrm{concat}(l_{\mathrm{feat}}, Attn) \tag{17}$$

表 I 不同方法在CDD、DSIFN-CD、和LEVIR-CD 数据集上的性能比较。最佳结果以绿色标示，次佳结果以-标示为红色。

Type	方法	发表年份	CDD			DSIFN-CD			LEVIR-CD		
			F_1	IoU	总体精度（%）	F_1	IoU	总体精度（%）	F_1	IoU	总体精度（%）
基于CNN	FC-EF [19]	ICIP-2018	65.38	48.57	93.07	64.38	47.47	87.55	81.16	68.29	98.12
	FC-Sima-diff [19]	ICIP-2018	70.37	54.29	94.31	62.03	44.96	88.17	83.77	72.08	98.44
	FC-Sima-conc [19]	ICIP-2018	72.65	57.05	94.47	65.42	48.61	87.60	86.26	75.85	98.60
	DTCDSCN [64]	LGRS-2021	94.14	88.93	98.63	87.21	76.58	95.59	89.75	81.41	98.98
	SNUNet [30]	LGRS-2022	95.34	91.11	98.90	79.64	66.17	93.48	88.59	79.51	98.85
	ISNet [65]	TGRS-2022	95.63	91.84	98.35	78.42	67.01	88.69	90.32	82.35	99.04
	SARAS-Net [66]	AAAI-2023	93.19	87.33	98.26	67.58	51.04	89.01	88.37	79.16	98.84
基于 Mamba	ChangeMamba [52]	TGRS-2024	94.91	90.32	98.81	90.21	82.17	94.39	90.18	82.07	99.01
	RS-Mamba [49]	LGRS-2024	93.59	87.96	98.49	86.67	76.48	95.56	89.56	81.06	98.96
	CDMamba [50]	arXiv-2024	94.84	90.20	98.83	85.81	75.13	95.17	90.41	82.25	99.02
	MF-VMamba [67]	TGRS-2024	95.69	91.58	98.97	88.05	78.73	96.01	90.64	82.89	99.07
基于 Transformer	BIT [36]	TGRS-2022	94.37	89.34	98.68	71.49	55.64	90.74	89.23	80.55	98.94
	ChangeFormer [68]	国际地球科学与遥感研讨会-2022	94.21	89.06	98.63	87.34	77.53	95.76	90.46	82.58	99.04
	ScratchFormer [69]	TGRS-2024	96.12	92.54	99.09	67.62	45.13	89.74	88.94	80.08	98.90
	ELGC-Net [42]	TGRS-2024	92.32	85.74	98.22	73.76	58.42	92.68	89.11	80.36	98.88
	Meta-SGNet [70]	TGRS-2024	-	-	-	-	-	-	89.20	80.50	98.93
	RAHFF-Net [71]	JSTARS-2025	94.46	87.96	98.64	-	-	-	89.41	79.72	98.92
	AMDANet [72]	TGRS-2025	-	-	-	-	-	-	90.32	82.34	99.04
	CICD [73]	JSTARS-2025	-	-	-	-	-	-	91.23	83.87	-
	AdvCP [46]	TIP-2026	-	-	-	58.90	43.10	85.75	70.06	50.45	98.95
	GRAD-Former	-	97.57	95.26	99.43	93.14	87.16	97.65	91.52	84.36	99.14

IV. 实验分析

本节全面概述了数据集、评估指标和实验设置。本节还分析了GRAD-Former在多个数据集上的性能，通过定性和定量评估将其与以往最先进模型进行比较，并提出消融研究以评估模型中各组件的重要性。

A. 数据集

本研究选取了三个样本分布不均的公开数据集，即 LEVIR-CD 数据集 [20], 、DSIFN-CD 数据集 [18], 和 CDD 数据集[74]，以全面验证 GRAD-Former 的有效性。

1）LEVIR-CD *[20]*：该数据集是用于建筑物变化检测的大规模基准数据集，包含 637 对高分辨率图像，空间分辨率为每像素 0.5 米，图像尺寸为 1024 × 1024 像素。这些图像来源于谷歌地球，记录了跨度为 5 至 14 年的变化。该数据集专门用于检测建筑物的建设与拆除。在实验中，我们使用该数据集的裁剪版本，提取尺寸为 256 × 256 像素且互不重叠的图像块。按照 [68], 中所述的标准协议，我们采用默认的数据划分，其中训练集包含 7120 对图像，验证集包含 1024 对，测试集包含 2048 对。

2) DSIFN-CD *[75]*：该二元变化检测数据集源自谷歌地球，包含来自中国六座主要城市——西安、重庆、北京、成都、武汉和深圳——的高分辨率双时相影像。数据集共包含 3988 对遥感图像对，涵盖水体、耕地、建筑物和道路等多种土地覆盖变化。数据集提供裁剪后的子图像对，尺寸为256 × 256 像素，并划分为训练集、验证集和测试集，分别包含 14,400、1,360 和 192 对图像。

3) CDD *[74]*：CDD数据集是从谷歌地球获取的公开数据集，由11个原始图像对组成，分辨率范围为0.03至1米/像素。其中，七对描绘季节性变化（4,725× 2,700），四对捕捉与灾害相关的变化（1,900 × 1,000），覆盖多样的地理区域和景观。每一对都包含云层覆盖、阴影、光照差异和旋转等复杂变化。该数据集主要关注自然要素的季节性变化，从单株树木到广阔的森林区域，较小部分则突出从车辆到大型建筑物的结构变化。它还以裁剪 256 × 256 格式提供，分别包含10,000、3,000和3,000个图像对，用于训练、验证和测试。

B. 评估指标

我们使用交并比（IoU）、 F_1 分数和总体准确率（总体精度（%））在所有数据集上评估变化检测结果，其中交并比和 F_1 分数针对变化类别计算。

C. 实验设置

我们使用PyTorch[1] 深度学习框架实现了所提出的 GRAD-Former，并由一台

[1]https://pytorch.org

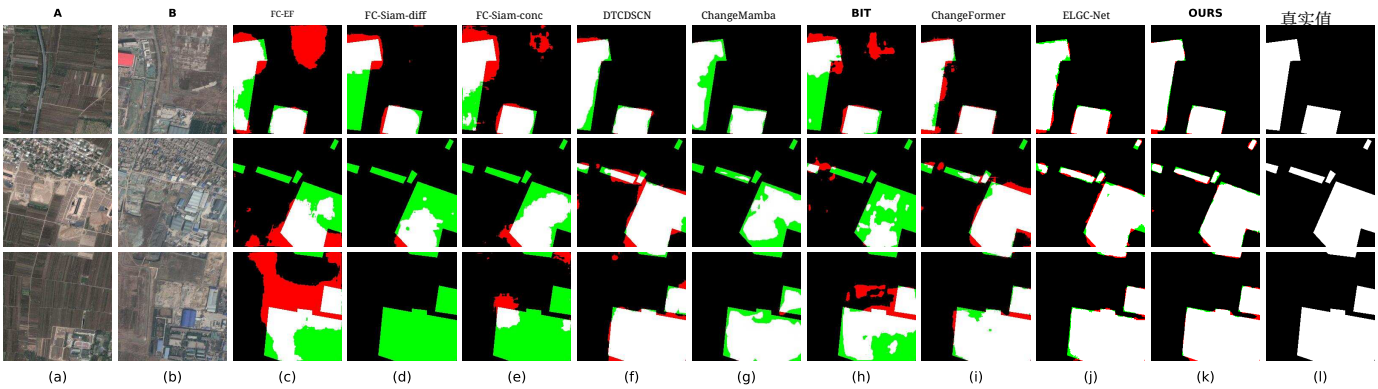


图5. DSIFN-CD 数据集上的定性比较。我们展示了与文献中八种最佳现有变化检测方法的比较，这些方法的代码库均已公开。(a) 变化前图像 (A)，(b) 变化后图像 (B)，(c) FC-EF [19], (d) FC-Siam-diff [19], (e) FC-Siam-conc [19], (f) DTCDSCN [64], (g) ChangeMamba [52], (h) BIT [36], (i) ChangeFormer [36], (j) ELGC-Net [42], (k) GRAD-Former（我们的方法）和 (l) 真实值。配色方案：白色表示真正例（即“发生变化”），黑色表示真负例（即“未发生变化”），红色表示假正例，绿色表示假负例。

配备16 GB显存的NVIDIA RTX A4000 GPU在训练和测试阶段提供算力。模型优化采用AdamW优化器，初始学习率设为0.0001。学习率遵循线性衰减策略，并在第100和200个epoch按计划衰减。训练过程共进行400个epoch，以确保模型得到充分学习。批量大小配置为2，以适应训练约束。对于模型评估，每个epoch均执行验证，并根据验证性能选择最佳模型，以确保最优泛化能力。随后使用专用评估器进入测试阶段，在测试集上评估模型性能。

GRAD-Former 编码器由四个阶段组成，分别包含 3、3、4 和 3 个编码器块，生成通道维度分别为 64、96、128 和 256 的特征图。编码器处理大小为 $256 \times 256 \times 3$ 的输入图像对。另一方面，解码器生成保持原始空间分辨率的二值变化图。为增强模型泛化能力，训练期间采用了多种数据增强技术，包括随机翻转、缩放、裁剪、颜色抖动和高斯模糊。推理期间，模型权重保持冻结，输入图像对被送入网络。模型生成包含两个通道的概率变化图，并沿通道维度执行 argmax 操作，以此得到二值变化图。

1) 损失函数：在训练阶段，我们采用逐像素交叉熵损失来评估网络的性能。交叉熵损失函数的数学定义如下：

$$\text{Loss} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)\right] \qquad (18)$$

其中 y 表示实际类别标签，$\hat{y}$ 表示预测概率，N 表示像素总数。

D. 与最先进方法的定量比较

我们对所提出的方法与采用开放代码库的最先进（SOTA）方法进行了定量比较，涵盖 LEVIR-CD、DSIFN-CD 和 CDD-CD 数据集上的关键指标（F_1、总体精度（%）和交并比），比较时间跨度为过去 8 年，结果见表 I。然而，由于我们开展研究时，最新模型（RAHFF-Net [71], AMDANet [72], Meta-SGNet [70], CICD [73],DgFA [76]）的官方代码库尚不可用，因此无法在所有数据集和指标上进行全面评估。为确保比较公平且具有代表性，我们改为报告其各自论文中公布的结果，并仅比较与本研究使用的数据集重合的部分。为突出检测性能更优的方法，各项评估指标中的最佳值和次佳值分别以绿色和红色标示。表 II 对比了我们的模型与其他 SOTA 模型在 DSIFN-CD 数据集上的参数量、GFLOPS 和 F_1 值。

GRAD-Former在每个数据集的所有评估指标上均优于所有其他模型（表I）。在CDD数据集上，我们的模型取得了97.57%的 F_1 分数、95.26%的交并比和99.43%的总体精度，超越了以往所有方法。与最佳的基于Transformer的模型ScratchFormer相比，我们的方法将 F_1 分数提高了1.45%，将交并比提高了2.72%，表明其具有更优越的特征提取能力。与最佳的基于CNN的模型SNUNet相比，我们的方法将交并比提高了2.23%，总体精度提升了0.53%，凸显了其更有效地捕捉重要空间和上下文关系的能力，如表中所示。

在 DSIFN-CD 数据集上，我们的方法依然保持优势，取得了最高的 F_1 score（93.14%）、87.16%的交并比（IoU）以及97.65%的总体精度（OA）。与该数据集上表现第二好的模型 Change-Mamba 相比，我们的模型在 F_1 score 上提升了2.93%，交并比（IoU）提高了约5.0%。此外，与领先的基于 Transformer 的模型 ChangeFormer 相比，我们的模型总体精度（OA）高出1.89%。这些结果表明，在卷积神经网络（CNN）和基于状态空间模型（SSM）的架构表现不佳的复杂变化及较低空间分辨率场景下，我们的模型能够进行有效处理。

在 LEVIR-CD 数据集上，我们的方法继续展现出卓越的性能，F_1 得分为 91.52%，交并比为 84.36%，总体精度为 99.14%。与表现最佳的基于 Transformer 的模型 CICD 相比，我们的方法在

交并比上高出 0.49%，总体精度上高出 0.10%。与基于 Mamba 的模型相比，我们的方法在交并比上比 CDMamba 高出 1.27%， F_1 得分高出 1.11%，这表明我们的改进进一步增强了长程依赖建模和相关特征聚合能力。

值得注意的是，GRAD-Former 未使用任何预训练骨干特征提取器，却仍以更少的可训练参数超越了最先进方法，展现出在不同数据集上优异的泛化能力。这些结果凸显了我们的方法在各类甚高分辨率数据集中捕获相关空间和上下文关系的稳健性。

E. 与最先进方法的定性比较

图5、图6和图7分别展示了多种 SOTA 方法在 DSIFN-CD、LEVIR-CD 和 CDD数据集上的对比可视化结果，如 [67] 所示。结果表明，基于 U-Net 的方法更容易出现假负例和假正例。不过，融入注意力机制后，这些问题显著减少。尽管这些方法通常能够识别所有变化区域，但在检测小面积样本以及准确识别变化方面仍面临困难。这一问题可能源于季节差异和光照条件带来的无关变化。此外，由于包含车辆等目标的图像样本相对较少，检测出的变化区域往往不完整且不准确。

GRAD-Former解决了伪变化问题，同时为检测到的对象提供了更清晰的边界，并提升了对象间语义一致性。三个数据集上的结果表明，与其他模型相比，GRAD-Former取得了更优的视觉效果。所提出的方法通过利用门控机制识别重要特征并消除噪声，同时借助注意力进行全局上下文建模，有效减少了假负例。观察图5可以发现，我们的模型生成的二值变化图几乎与真实值无法区分，并且能够检测清晰的边缘，区分小范围和大范围变化区域。在图6中可以观察到，我们的模型将假正例降至最低，并检测出场景3图6中变化区域之间的细微细节。如图7所示，与其他SOTA模型相比，我们的模型能够忽略所有季节性变化，生成高度准确的变化图（场景1图7）；我们的模型也是唯一能够准确识别左侧河流（场景2图7）的模型，并且能够检测出许多较小的变化区域，这体现了模型的鲁棒性，以及其聚焦于相关变化区域并忽略不重要特征和噪声的能力。

F. 消融研究

1) AFRAR模块中特征提取模块与融合模块的消融实验：我们在 LEVIR-CD 数据集上评估了所提出方法的影响，包括 SEA、GLFR 和 DA模块。为此，我们系统地采用这些技术的不同组合，并报告相应的参数、交并比、 F_1 分数和总体精度（%）

表 III 选择性嵌入增强（SEA）模块、全局-局部特征细化（GLFR）模块和差异融合（DA）模块的组合及其性能指标的消融研究。

模块			参数（M）	**IoU**	F_1	总体精度（%）
SEA	**GLFR**	**DA**				
✓	×	×	9.99	79.69	87.93	96.98
×	×	✓	10.72	80.30	88.9	97.09
✓	×	✓	10.71	81.22	88.99	97.80
×	✓	×	10.19	82.92	90.63	97.60
×	✓	✓	10.20	83.10	90.79	99.08
✓	✓	×	10.24	83.92	90.94	99.04
✓	✓	✓	**10.90**	**84.36**	**91.52**	**99.14**

如表 III 所示。我们最终的网络整合了所有上下文聚合方法，取得了最高的交并比分数 84.36%，且参数效率更高，如表 III 最后一行所示。这表明，我们的模块能够有效消除甚高分辨率数据集中的噪声，并行整合多种上下文聚合策略。

表 II 不同方法在模型大小、计算需求和性能指标方面的比较。我们提出的方法以适中的参数量取得了最高的 F_1 得分（93.14%）。

方法	参数(M)	GFLOPs	F_1
FC-EF [19]	1.35	14.31	64.38
FC-EF-conc [19]	1.54	21.32	62.03
FC-EF-diff [19]	1.35	18.91	65.42
DTCDSCN [64]	31.25	52.89	87.21
SNUNet [30]	12.03	21.33	79.64
ISNet [65]	14.36	23.58	78.42
SARAS-Net [66]	56.89	139.90	67.58
ChangeMamba [52]	85.53	179.32	90.21
RS-Mamba [49]	52.73	95.74	86.67
CDMamba [50]	12.71	151.23	85.81
MF-Vmamba [67]	57.84	94.06	88.05
BIT [36]	3.49	42.53	71.49
ChangeFormer [68]	41.03	211.15	87.34
ScratchFormer [69]	36.92	196.59	67.62
ELGC-Net [42]	10.57	123.59	73.76
GRAD-Former	**10.90**	**129.50**	**93.14**

2) GLFR模块中不同注意力机制的消融：我们的研究证实了差分注意力机制的有效性（如表IV所示）。在对比测试中，当我们用标准自注意力替换差分注意力时，发现差分注意力不仅所需参数更少，而且提高了评估分数。基于差分注意力的模块在评估矩阵和效率方面甚至优于声称可线性扩展的基于PT注意力的模块。

3）GRAD-Former 训练过程中不同损失函数的比较：我们还进行了消融研究，以考察不同损失函数对所提出模型训练的影响。表V中的结果清楚地表明，

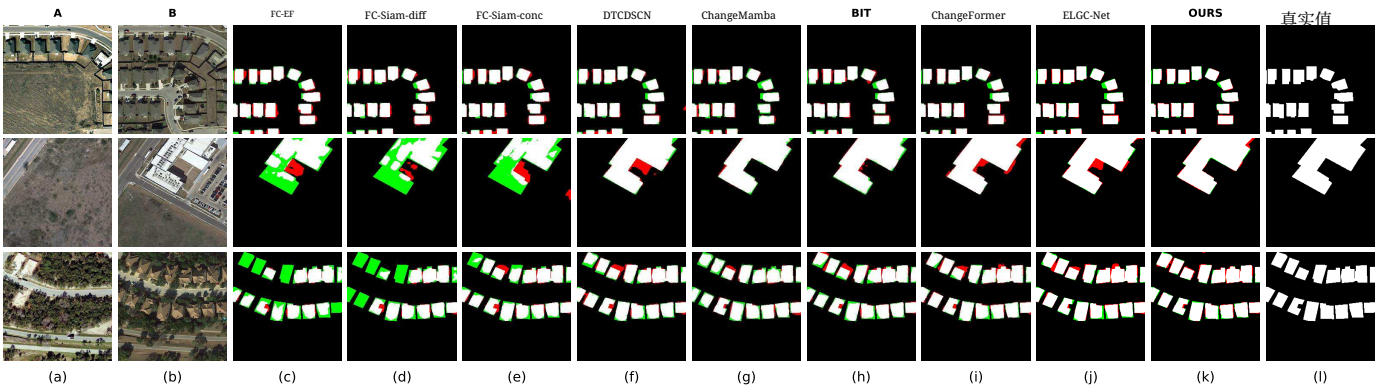


图6. LEVIR-CD 数据集上的定性比较。我们展示了与文献中现有的八种最佳变化检测方法的比较，这些方法的代码库均已公开。(a) 变化前图像（A），(b) 变化后图像（B），(c) FC-EF [19], (d) FC-Siam-diff [19], (e) FC-Siam-conc [19], (f) DTCDSCN [64], (g) ChangeMamba [52], (h) BIT [36], (i) ChangeFormer [36], (j) ELGC-Net [42], (k) GRAD-Former（我们的方法）和 (l) 真实值。颜色方案：白色表示真正例（即“发生变化”），黑色对应真负例（即“未发生变化”），红色表示假正例，绿色表示假负例。

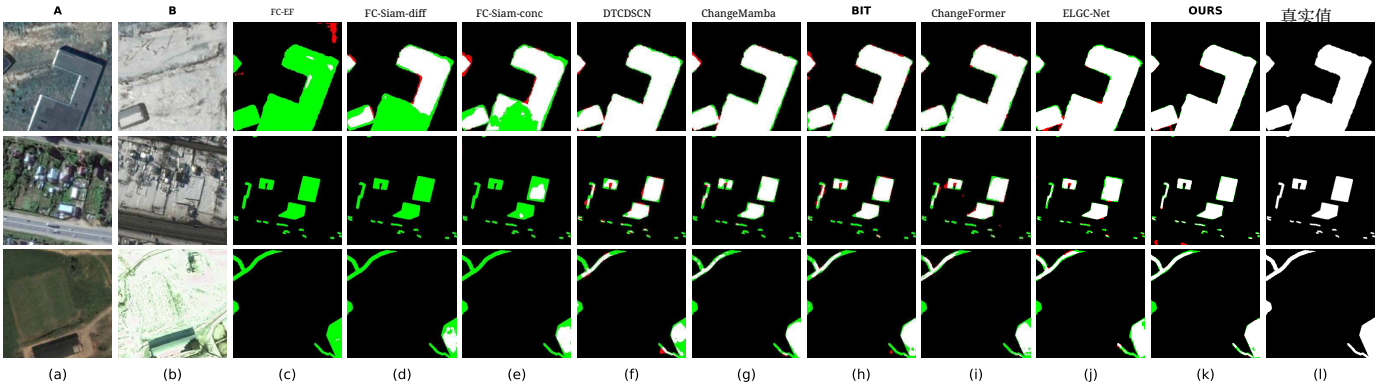


图7. CDD数据集上的定性比较。我们展示了与文献中八种最佳现有变化检测方法的比较，这些方法的代码库均已公开。(a) 变化前图像 (A)，(b) 变化后图像 (B)，(c) FC-EF [19], (d) FC-Siam-diff [19], (e) FC-Siam-conc [19], (f) DTCDSCN [64], (g) ChangeMamba [52], (h) BIT [36], (i) ChangeFormer [36], (j) ELGC-Net [42], (k) GRAD-Former（我们的方法）和 (l) 真实值。配色方案：白色表示真正例（即“发生变化”），黑色表示真负例（即“未发生变化”），红色表示假正例，绿色表示假负例。

表IV GLFR模块中不同注意力机制在LEVIR-CD 数据集上的比较。

注意力类型	参数（M）	F_1	IoU	总体精度（%）
简单注意力	10.98	84.32	72.89	98.18
池化转置注意力	10.83	91.08	83.62	99.1
差分注意力	**10.90**	**91.52**	**84.36**	**99.14**

表V 在 LEVIR-CD 数据集上进行模型训练时采用不同损失函数的比较分析，最优性能结果以粗体标出。

损失函数	F_1	IoU	总体精度（%）
焦点损失	84.32	72.89	98.18
平均交并比损失	48.93	47.90	95.80
交叉熵损失	**91.52**	**84.36**	**99.14**

与替代损失函数相比，交叉熵（CE）损失能为我们的模型训练带来更优异的性能。正如 Noman 等人也指出的那样， [42], 交叉熵稳定的梯度动态和逐像素优化与孪生架构的空间特征提取有效契合。相反，焦点损失会过度强调困难负样本，导致训练不稳定性；而当变化区域稀疏时，平均交并比损失会出现梯度不稳定性，导致变化检测任务的收敛效果欠佳。

V. 结论

本文提出了 GRAD-Former，这是一种专为遥感变化检测设计的模型，旨在解决.....不足。

通过其新颖的AFRAR模块，利用超高分辨率（VHR）卫星图像中的相关信息。GRAD-Former高效整合全局-局部上下文特征聚合器和门控模块，以增强语义信息，同时将可训练参数的数量保持在最低水平。值得注意的是，该模型无需预训练骨干特征提取器即可运行。大量实验验证了我们方法的有效性：GRAD-Former在三个具有挑战性的数据集上取得了最先进的性能，并创下新的基准。未来的工作将致力于进一步提升效率，并使模型适用于边缘设备上的实时变化检测。

参考文献

[1] C. Fichera、G. Modica和M. Pollino，“利用多时相遥感影像和景观指数进行土地覆盖分类与变化检测分析”，《欧洲遥感杂志》，第45卷，第1–18页，2012年4月。[2] X. Huang、Y. Cao和J. Li，“一种利用时间序列多视角高分辨率光学卫星影像监测新建建筑区的自动变化检测方法”，《环境遥感》，第244卷，第111802页，2020年7月。[3] R. Liu、M. Kuffer和C. Persello，“采用基于CNN的变化检测方法研究贫民窟的时间动态”，《遥感》，第11卷，第2844页，2019年11月。[4] F. Deng、W. Luo、Y. Ni、X. Wang、Y. Wang和G. Zhang，“Umit-net：一种用于利用遥感图像精确提取道路的U型混合Transformer网络”，《IEEE地球科学与遥感汇刊》，第61卷，第1–1页，2023年1月。[5] X. Wang、X. Fan、Q. Xu和P. Du，“通过扩展形态学剖面和集成策略进行基于变化检测的同震滑坡制图”，《ISPRS摄影测量与遥感杂志》，第187卷，第225–239页，2022年5月。[6] A. Ulrichsen、P. Murray、S. Marshall、M. Gabbouj、S. Kiranyaz、M. Yamaç和N. Aburaed，“用于参数高效高光谱单图像超分辨率的运算神经网络”，《IEEE应用地球观测与遥感专题期刊》，第PP卷，第1–17页，2023年1月。[7] N. Gupta、S. Ari和N. Panigrahi，“利用无监督学习和基于RBF的聚类进行Landsat影像变化检测”，《IEEE计算智能新兴主题汇刊》，第5卷，第2期，第284–297页，2021年。[8] C. Wu、B. Du和L. Zhang，“用于多光谱影像变化检测的慢特征分析”，《IEEE地球科学与遥感汇刊》，第52卷，2014年4月。[9] Y. Li、X. Li、J. Song、Z. Wang、Y. He和S. Yang，“在后验概率空间中利用变化向量分析进行基于遥感的变化检测：一种上下文敏感贝叶斯网络方法”，《IEEE应用地球观测与遥感专题期刊》，第PP卷，第1–21页，2023年1月。[10] S. Seydi和R. Shahhoseini，“用于全极化遥感图像变化检测的变换算法”，载于《ISPRS——国际摄影测量、遥感与空间信息科学档案》，第XLII-4/W18卷，2019年10月，第963–967页。[11] M. Hussain、D. Chen、A. Cheng、H. Wei和D. Stanley，“遥感图像变化检测：从基于像素的方法到基于对象的方法”，《国际摄影测量与遥感杂志》，第80卷，第91–106页，2013年6月。[12] L. Bergamasco、S. Saha、F. Bovolo和L. Bruzzone，“利用卷积自编码器多分辨率特征进行无监督变化检测”，《IEEE地球科学与遥感汇刊》，第PP卷，第1–1页，2022年1月。[13] B. Hou、Y. Wang和Q. Liu，“基于深度特征和低秩的变化检测”，《IEEE地球科学与遥感快报》，第PP卷，第1–5页，2017年11月。[14] L. Mou和X. X. Zhu，“用于多光谱图像土地覆盖变化检测的循环卷积神经网络”，载于《IGARSS 2018——2018年IEEE国际地球科学与遥感研讨会》，2018年，第4363–4366页。[15] J. Vinholi、B. Gregory Palm、D. Silva、R. Machado和M. Pettersson，“利用波长分辨率合成孔径雷达图像堆栈，通过卷积神经网络进行变化检测”，《IEEE地球科学与遥感汇刊》，第PP卷，第1 1页，2022年1月—。[16] T. Bai、L. Wang、D. Yin、K. Sun、Y. Chen、W. Li和D. L. and，“遥感变化检测的深度学习：综述”，《地球空间信息科学》，第26卷，第3期，第262–288页，2023年。[17] Z. Gong、R. Jing和H. Guan，“基于多尺度SLIC-CNN和SCAE特征的VHR卫星影像土地覆盖变化检测”，《IEEE Access》，第8卷，第1–1页，2020年12月。[18] C. Zhang、P. Yue、D. Tapete、L. Jiang、B. Shangguan、L. Huang和G. Liu，“用于高分辨率双时相遥感图像变化检测的深度监督图像融合网络”，《ISPRS摄影测量与遥感杂志》，第166卷，第183–200页，2020年8月。[19] R. Daudt、B. Saux和A. Boulch，“用于变化检测的全卷积孪生网络”，载于《IEEE国际图像处理会议》，2018年10月，第4063–4067页。

[20] H. He 和 Z. Shao，“利用 Transformer 和 CNN 进行遥感图像变化检测的时空注意力方法及新数据集”，《Remote Sensing》，第12卷，第1662页，2020年5月。[21] P. Yuan、Z. Qingzhan、X. Zhao、X. Wang、X. Long 和 Y. Zheng，“一种用于遥感图像语义变化检测的基于 Transformer 的孪生网络及开放光学数据集”，《国际数字地球杂志》，第15卷，第1506–1525页，2022年9月。[22] X. Song、Z. Hua 和 J. Li，“基于双特征混合注意力的遥感图像变化检测 Transformer 网络”，《IEEE地球科学与遥感汇刊》，第PP卷，第1–1页，2022年1月–。[23] A. Vaswani、N. Shazeer、N. Parmar、J. Uszkoreit、L. Jones、A. N. Gomez、L. u. Kaiser 和 I. Polosukhin，“Attention Is All You Need”，载于《神经信息处理系统进展》，第30卷。Curran Associates, Inc.，2017年。[24] T. Ye、L. Dong、Y. Xia、Y. Sun、Y. Zhu、G. Huang 和 F. Wei，“差分 Transformer”，2024年。[25] D. Müller 和 F. Kramer，“Miscnn：一种结合卷积神经网络与深度学习的医学图像分割框架”，《BMC 医学影像》，第21卷，2021年1月。[26] A. Tripathi、T. Kumar、T. Dhansetty 和 J. Kumar，“使用卷积神经网络进行实时目标检测”，《International Journal of Engineering and Technology(UAE)》，第7卷，第33–36页，2018年4月。[27] X. Wu、D. Hong 和 J. Chanussot，“用于多模态遥感数据分类的卷积神经网络”，《IEEE地球科学与遥感汇刊》，第PP卷，第1–1页，2021年11月。[28] O. Ronneberger、P. Fischer 和 T. Brox，“U-Net：用于生物医学图像分割的卷积网络”，载于《医学图像计算与计算机辅助干预国际会议》，第9351卷，2015年10月，第234–241页。[29] Z. Zhou、M. M. Rahman Siddiquee、N. Tajbakhsh 和 J. Liang，“UNet++：用于医学图像分割的嵌套式 U-Net 架构”，载于《医学图像分析中的深度学习（DLMIA）研讨会》，2018年7月。[30] S. Fang、K. Li、J. Shao 和 Z. Li，“SNUNet-CD：用于甚高分辨率图像变化检测的密集连接孪生网络”，《IEEE地球科学与遥感快报》，第PP卷，第1–5页，2021年2月。[31] M. Papadomanolaki、S. Verma、M. Vakalopoulou、S. Gupta 和 K. Karantzalos，“利用循环神经网络检测多时相 Sentinel-2 数据中的城市变化”，载于《IEEE国际地球科学与遥感研讨会》，2019年7月，第214–217页。[32] R. Daudt、B. Saux 和 A. Boulch，“用于变化检测的全卷积孪生网络”，载于《IEEE国际图像处理会议（ICIP）》，2018年10月，第4063–4067页。[33] S. Khan、X. He、F. Porikli、M. Bennamoun、F. Sohel 和 R. Togneri，“面向弱监督变化检测的深度结构化网络学习”，载于《国际人工智能联合会议》，2017年8月，2008 2015 pp. –。[34] A. Dosovitskiy、L. Beyer、A. Kolesnikov、D. Weissenborn、X. Zhai、T. Unterthiner、M. Dehghani、M. Minderer、G. Heigold、S. Gelly、J. Uszkoreit 和 N. Houlsby，“一张图像胜过 16×16 个词：用于大规模图像识别的 Transformer”，载于《国际学习表征会议》，2021年。[35] F. Song、S. Zhang、T. Lei、Y. Song 和 Z. Peng，“MSTDSNet-CD：用于快速增长城市区域变化检测的多尺度 Swin Transformer 与深度监督网络”，《IEEE地球科学与遥感快报》，第1–1页，2022年4月。[36] H. Chen、Z. Shi 和 Z. Qi，“基于 Transformer 的遥感图像变化检测”，《IEEE地球科学与遥感汇刊》，第60卷，2021年7月。[37] Z. Li、C. Tang、L. Wang 和 A. Zomaya，“通过时序特征交互与引导式细化进行遥感变化检测”，《IEEE地球科学与遥感汇刊》，第PP卷，第1–1页，2022年1月–。[38] Z. Li、C. Tang、X. Liu、W. Zhang、J. Dou、L. Wang 和 A. Zomaya，“通过渐进式特征聚合与监督注意力实现轻量级遥感变化检测”，《IEEE地球科学与遥感汇刊》，第PP卷，第1–1页，2023年1月。[39] W. Lu、L. Wei 和 M. Nguyen，“用于建筑物变化检测与建筑物损毁评估的双时相注意力 Transformer”，《IEEE应用地球观测与遥感专题期刊》，第PP卷，第1–20页，2024年1月。[40] Q. Ke 和 P. Zhang，“Hybrid-TransCD：一种通过令牌聚合实现的混合 Transformer 遥感图像变化检测网络”，《ISPRS 国际地理信息杂志》，第11卷，第263页，2022年4月。

[41] W. Ren、Z. Wang、M. Xia 和 H. Lin，“MFiNet：用于高分辨率遥感图像变化检测的多尺度特征交互网络”，《Remote Sensing》，第16卷，第1269页，2024年4月。[42] M. Noman、M. Fiaz、H. Cholakkal、S. Khan 和 F. S. Khan，“ELGC-Net：用于遥感变化检测的高效局部-全局上下文聚合”，《IEEE地球科学与遥感汇刊》，2024年。[43] Y. Lin、S. Liu、Y. Zheng、X. Tong、H. Xie、H. Zhu、K. Du、H. Zhao 和 J. Zhang，“一种用于甚高分辨率遥感图像变化检测的无监督基于 Transformer 的多变量变化检测方法”，《IEEE应用地球观测与遥感专题期刊》，第PP卷，第1–10页，2024年1月。[44] B. Wang、X. Jiang、P. Qin 和 J. Zeng，“WSMsFNet：结合全监督与多尺度融合网络进行遥感图像变化检测”，《IEEE应用地球观测与遥感专题期刊》，第PP卷，第1–14页，2024年1月。[45] L. Wu、J. Peng、B. Yang、W. Sun 和 M. Huang，“CGMNet：一种用于高光谱变化检测的中心像素与基于门控机制的注意力网络”，《IEEE图像处理汇刊》，第35卷，第1951–1965页，2026年。[46] Z. Zhao、C. Wu、D. Wang、H. Chen、C. Chen、Z. Zheng、B. Du 和 L. Zhang，“通过对抗式类别提示推进卫星图像中的弱监督变化检测”，《IEEE图像处理汇刊》，第34卷，第7065–7078页，2025年。[47] A. Gu 和 T. Dao，“Mamba：采用选择性状态空间的线性时间序列建模”，arXiv 预印本 arXiv:2312.00752，2023年。[48] Y. Liu、Y. Tian、Y. Zhao、H. Yu、L. Xie、Y. Wang、Q. Ye 和 Y. Liu，“Vmamba：视觉状态空间模型”，arXiv 预印本 arXiv:2401 10166，2024年。。[49] K. Chen、B. Chen、C. Liu、W. Li、Z. Zou 和 Z. Shi，“Rsmamba：采用状态空间模型的遥感图像分类”，《IEEE地球科学与遥感快报》，第21卷，第1–5页，2024年。[50] H. Zhang、K. Chen、C. Liu、H. Chen、Z. Zou 和 Z. Shi，“CDMamba：将局部线索融入 Mamba 以实现遥感图像二元变化检测 ”，《 IEEE 地球科学与遥感汇刊》，第PP卷，第1–1页，2025年1月。[51] J. Huang、X. Yuan、C.-T. Lam、Y. Wang 和 M. Xia，“LCCDMamba：用于甚高分辨率遥感图像土地覆盖变化检测的视觉状态空间模型”，《IEEE应用地球观测与遥感专题期刊》，第18卷，第5765–5781页，2025年。[52] H. Chen、J. Song、C. Han、J. Xia 和 N. Yokoya，“ChangeMamba：采用时空状态空间模型的遥感变化检测”，《IEEE地球科学与遥感汇刊》，第62卷，第1–20页，2024年。[53] J. Yao、D. Hong、C. Li 和 J. Chanussot，“SpectralMamba：用于高光谱图像分类的高效 Mamba”，2024年。[54] G. Wang、X. Zhang、Z. Peng、T. Zhang 和 L. Jiao，“S 2 Mamba：用于高光谱图像分类的空谱状态空间模型”，《IEEE地球科学与遥感汇刊》，第PP卷，第1–1页，2025年1月–。[55] M. Liu、J. Dan、Z. Lu、Y. Yu、Y. Li 和 X. Li，“CM-UNet：用于遥感图像语义分割的混合卷积神经网络-Mamba U-Net”，2024年。[56] X. Ma、X. Zhang 和 M.-O. Pun，“RS3Mamba：用于遥感图像语义分割的视觉状态空间模型”，《IEEE地球科学与遥感快报》，第21卷，第1–5页，2024年。[57] T. Huang、X. Pei、S. You、F. Wang、C. Qian 和 C. Xu，“LocalMamba：采用窗口化选择性扫描的视觉状态空间模型”，arXiv 预印本 arXiv:2403.09338，2024年。[58] J. Qu、J. Zhao、W. Dong、L. Zhang 和 Y. Li，“MambaFedCD：基于 Mamba 的空-谱-时协同主动联邦高光谱变化检测”，《IEEE图像处理汇刊》，第35卷，第1478–1492页，2026年。[59] C. Szegedy、W. Liu、Y. Jia、P. Sermanet、S. Reed、D. Anguelov、D. Erhan、V. Vanhoucke 和 A. Rabinovich，“深入卷积网络”，载于《2015年IEEE计算机视觉与模式识别会议（CVPR）》，2015年，第1–9页。[60] S. Zhou、G. Yuan、Z. Hua 和 J. Li，“Dgfeg：用于遥感变化检测的动态门控融合与边缘图感知网络”，《IEEE应用地球观测与遥感专题期刊》，第18卷，第3581–3598页，2025年。[61] L.-L. Li、Z.-H. You、S.-B. Chen、L.-L. Huang、J. Tang 和 B. Luo，“Gateformer：用于遥感图像变化检测的 Transformer 门控注意力 U-Net”，《 IEEE 应用地球观测与遥感专题期刊》，第17卷，第871–883页，2024年。[62] W. Liu、J. Li、H. Wang、R. Tan、Y. Fu 和 Q. Tian，“Lcd-net：一种结合特征融合与门控机制的轻量级遥感变化检测网络”，《IEEE应用地球观测与遥感专题期刊》，第PP卷，第1–应用地球观测与遥感，第 PP 卷，第 1–13 页，2025年1月。[63] Deepanshi、R. Barkur、D. Suresh、S. Lal、C. S. Reddy 和 P. G. Diwakar，“Rscdnet：一种用于从双时相高分辨率遥感图像中进行变化检测的鲁棒深度学习架构”，IEEE计算智能新兴主题汇刊，第 7 卷，第 2 期，第 537–551 页，2023年。[64] Y. Liu、C. Pang、Z. Zhan、X. Zhang 和 X. Yang，“使用双任务约束深度孪生卷积网络模型进行遥感图像建筑物变化检测”，IEEE地球科学与遥感快报，第 18 卷，第 5 期，第 811–815 页，2021年。[65] G. Cheng、G. Wang 和 J. Han，“ISNet：提升遥感图像变化检测可分性的研究”，IEEE地球科学与遥感汇刊，第 60 卷，第 1–11 页，2022年。[66] C.-P. Chen、J.-W. Hsieh、P.-Y. Chen、Y.-K. Hsieh 和 B.-S. Wang，“SARAS-Net：用于变化检测的尺度与关系感知孪生网络”，AAAI人工智能会议论文集，第 37 卷，第 12 期，第 14 187–14 195 页，2023年6月。[67] Z. Zhang、X. Fan、X. Wang、Y. Qin 和 J. Xia，“一种基于多层状态空间模型的新型遥感图像变化检测方法”，IEEE地球科学与遥感汇刊，第 62 卷，第 1–14 页，2024年。[68] W. G. C. Bandara 和 V. M. Patel，“一种用于变化检测的基于 Transformer 的孪生网络”，载于 IGARSS 2022——2022年IEEE国际地球科学与遥感研讨会，2022年，第 207–210 页。[69] M. Noman、M. Fiaz、H. Cholakkal、S. Narayan、R. Muhammad Anwer、S. Khan 和 F. Shahbaz Khan，“使用从头训练的变换器进行遥感变化检测”，IEEE地球科学与遥感汇刊，第 62 卷，第 1–14 页，2024年。[70] J. Deng、G. Gong、G. Zhou、M. Yan 和 L. Li，“一种用于遥感图像建筑物变化检测的门控特征融合网络，配备元编码器和自校准交叉模块”，IEEE地球科学与遥感汇刊，第 62 卷，第 1–16 页，2024年。[71] B. Wang、K. Zhao、T. Xiao、P. Qin 和 J. Zeng，“RAHFF-Net：用于遥感图像变化检测的召回可调层次特征融合网络”，IEEE应用地球观测与遥感专题期刊，第 18 卷，第 176–190 页，2025年。[72] Y. Su、P. Ma、W. Wang、S. Wang、Y. Wu、Y. Li 和 P. Jing，“AMDANet：用于图像变化检测的增强型多尺度差异聚合网络”，IEEE地球科学与遥感汇刊，第 1–1 页，2025 年。[73] W. Zheng 、 J. Yang 、 J. Chen 、 J. He 、 P. Li、D. Sun、C. Chen 和 X. Meng，“通过颜色分布归一化进行跨时相知识注入，实现遥感变化检测”，IEEE应用地球观测与遥感专题期刊，第 18 卷，第 6249–6265 页，2025年。[74] M. A. Lebedev、Y. V. Vizilter、O. V. Vygolov、V. A. Knyaz 和 A. Y. Rubis，“使用条件对抗网络进行遥感图像变化检测”，《国际摄影测量、遥感与空间信息科学档案》，第 XLII-2 卷，第 565–571 页，2018年。[75] J. Wu、C. Xie、Z. Zhang 和 Y. Zhu，“一种用于遥感图像变化检测的深度监督注意力高分辨率网络”，《遥感》，第 15 卷，第 1 期，2023年。[76] F. Zhou、X. Zhang、H. Shuai、R. Hang、S. Zhu 和 T. Geng，“用于遥感图像变化检测的双粒度特征对齐”，IEEE应用地球观测与遥感专题期刊，第 18 卷，第 4487–4497 页，2025年。

GRAD-Former：用于变化检测的门控鲁棒注意力差分变换器

补充材料

1 引言

本文档提供了有关我们提出的 GRAD-Former 变化检测（CD）模型的补充结果与分析。GRAD-Former 框架通过缩减模型规模，在保持效率的同时增强上下文理解能力，其架构由一个包含自适应特征相关性与细化（AFRAR）模块的新型编码器、融合模块和解码器模块组成。本补充材料全面地从定量和定性两方面评估了该模型与最先进方法相比的性能，并特别关注 WHU-CD 数据集。

2 数据集描述

WHU-CD 数据集 [1] 是一个专注于建筑物变化检测的开源数据集，由武汉大学的研究人员利用来自新西兰基督城的航空影像构建。该数据集包含 2012 年至 2016 年间拍摄的变化前后图像，空间分辨率为 0.3 米/像素。数据集覆盖 20.5 平方千米的区域，记录了城市改造情况：2012 年有 12,796 栋建筑物，2016 年有 16,077 栋，反映了地震后的变化。实验中，图像被裁剪为 256×256 像素，并分别划分 5,947 对、744 对和 743 对图像用于训练、验证和测试。

3 定量性能分析

表 S 1中，我们对 GRAD-Former 与现有最先进的遥感变化检测方法在 WHU-CD 数据集上的表现进行了全面比较评估。实验结果表明，在所有评估指标上，GRAD-Former 均持续优于其他模型（F_1：91.29%，交并比：83.98%，总体精度：99.31%）。我们的模型仅包含 10.90M 个参数，与 BIT（57.84M）、RS-Mamba（52.73M）和 ChangeMamba（85.53M）等其他高性能架构相比，展现出显著的计算效率优势。

与表现排名第二的 RS-Mamba 相比，我们的方法使 F_1 分数提高了 0.26%，交并比提高了 0.55%，表明其特征提取能力更强。与交并比排名第二的 BIT 相比，GRAD-Former 的交并比提高了 0.54%，凸显了其更有效地捕获重要空间和上下文关系的能力。这些结果证明，所提出的自适应特征相关性与细化（AFRAR）模块能够在保持计算效率的同时，增强模型的判别能力。

4 计算效率—性能权衡分析 **s**

图 S 1 通过帕累托效率图直观展示了各对比变化检测模型之间的效率—性能权衡。GRAD-Former 位于帕累托前沿上，表明其在架构复杂度与性能之间实现了最佳平衡。尽管具备出色的判别能力，我们的模型仍将计算需求控制在可接受范围内（129.5 GFLOPs），从而便于在资源受限环境中进行实际部署。

分析可视化结果表明，许多竞争性方法要么需要显著增加参数量才能达到相当的性能，要么在模型复杂度相当甚至更高的情况下仍表现出较差的指标。ChangeMamba [7] 所需参数约为 GRAD-Former 的 8 倍，但取得的F_1 分数更低，而 BIT [9] 则需要约 5.5 倍的参数量才能达到相当但仍逊色的性能。这种显著的参数效率在实际遥感场景中尤为重要，因为

*这些作者对本项工作作出了同等贡献。

Type	方法	参数（M）	发表时间	WHU-CD		
				F_1	IoU	总体精度（%）
基于CNN	FC-EF [2]	1.35	ICIP-18	80.64	67.56	98.12
	FC-Sima-diff [2]	1.54	ICIP-18	67.64	51.1	95.82
	FC-Sima-conc [2]	1.35	ICIP-18	64.66	47.78	95.32
	DTCDSCN [3]	31.25	LGRS-21	87.79	78.23	98.89
	SNUNet [4]	12.03	LGRS-22	90.29	82.3	99.05
	ISNet [5]	14.36	TGRS-22	83.24	72.47	98.38
	SARAS-Net [6]	56.89	LGRS-23	87.09	77.14	98.89
基于 Mamba	ChangeMamba [7]	85.53	TGRS-24	84.11	78.05	98.86
	RS-Mamba [8]	52.73	LGRS-24	91.03	83.43	99.22
基于 Transformer	BIT [9]	57.84	TGRS-22	90.97	83.44	99.11
	ChangeFormer [10]	41.03	IGARSS-22	85.31	79.02	99.00
	ScratchFormer [11]	36.92	TGRS-24	87.78	78.22	98.93
	ELGC-Net [12]	10.57	TGRS-24	87.54	77.85	98.84
	MSFNet [13]	N/A	JSTARS-24	90.87	82.8	99.07
	RaHFF-Net [14]	16.48	JSTARS-25	90.43	81.72	99.23
	AMDANet [15]	25.83	TGRS-25	89.3	80.68	99.03
	AdvCP [16]	N/A	TIP-26	77.25	62.39	98.51
	GRAD-Former	10.90	-	91.29	83.98	99.31

表 S 1：WHU 数据集上不同遥感变化检测方法的对比 F_1分数、交并比和总体准确率（OA（%）），其中 F_1 交并比是针对变化类别计算的。 最高和次高值分别以高亮显示在各指标列中。

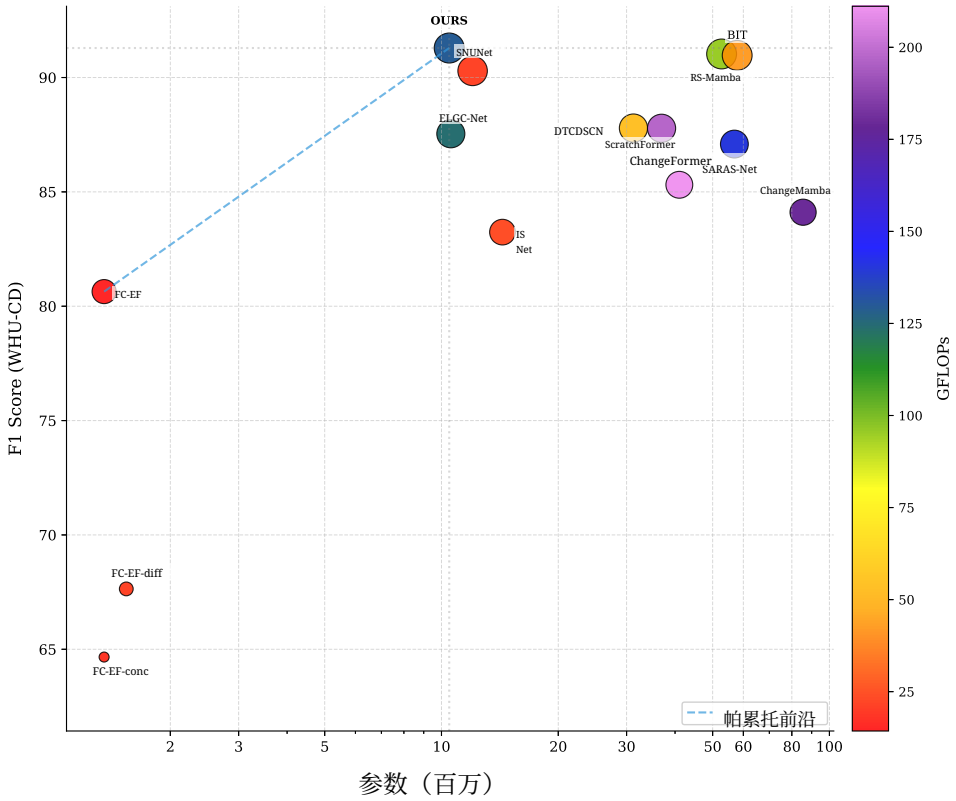


图 S 1：根据表 1 对变化检测模型的性能进行比较。该图展示了模型规模（百万级参数）与性能（F_1 WHU-CD 数据集上的得分）之间的权衡。气泡颜色表示计算复杂度（GFLOPs）。蓝色虚线表示最优模型的帕累托前沿。我们的模型位于帕累托前沿上，实现了最先进的 F_1 得分（91.29%），同时保持了合理的计算效率（129.5 GFLOPs）。

计算资源可能有限。计算复杂度与性能之间的关系进一步验证了我们的方法，证明 GRAD-Former 是一种计算效率高且性能出色的架构，适用于遥感变化检测应用。

5 定性结果比较

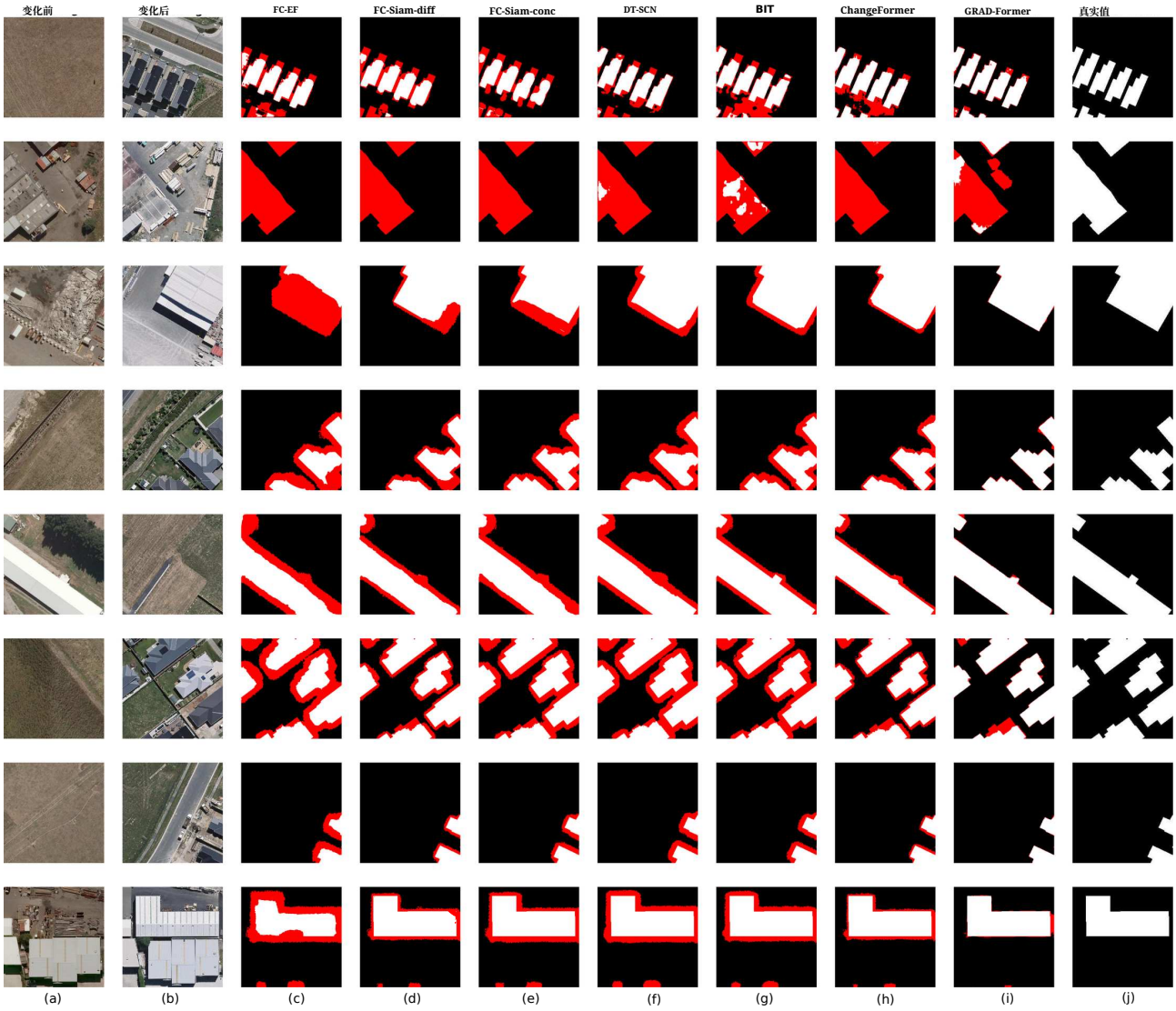


图 S 2：在 WHU-CD 数据集上应用多种最先进的变化检测方法进行的可视化比较：(a) 变化前图像，(b) 变化后图像，(c) FC-EF [2], **(d) FC-Siam-diff [2], (e) FC-Siam-conc [2], (f)DT-SCN [3], (g) BIT [9], (h) ChangeFormer [10],** (i) GRAD-Former，以及 (j) 真实值。结果中，白色像素表示真正例（变化区域），黑色像素表示真负例（未变化区域），红色像素则表示假正例和假负例。

图 S 2 展示了使用 WHU-CD 数据集中的示例图像，对 GRAD-Former 与多种最先进的变化检测方法进行的定性比较。与先前的方法（FC-EF [2],FC-Siam-diff [2], FC-Siam-conc [2]）相比，GRAD-Former 在精确勾勒变化区域的几何边界方面表现出更强的能力；这些先前方法存在明显的边界定位误差。与替代方法相比，我们提出的架构显著减少了假正例和假负例（以红色像素表示），表明其判别可靠性有所提高。

GRAD-Former能够有效捕捉其他方法常常忽略的复杂结构变化，尤其是在语义变化细微的区域。定性分析表明，该模型在

空间分布各异的区域中均能保持稳定的检测性能。相比之下，其他基于 Transformer 的方法会因语义变化的复杂程度不同而表现不一。较早的基于 CNN 的方法难以捕捉细粒度变化，而较新的基于Transformer的模型虽有所改进，但仍往往会产生边界不精确的问题。GRAD-Former生成的结果与真实值高度吻合，建筑物边缘清晰，且误检极少。这凸显了该模型通过所提出的AFRAR模块有效融合局部和全局上下文信息的能力。

6 结论

本文呈现的补充分析为 GRAD-Former 在遥感变化检测任务中的卓越性能提供了全面的实证证据。通过严谨的定量比较评估、计算效率分析和定性评估，我们证明，所提出的架构框架为该领域树立了新的基准。新颖的自适应特征相关性与细化（AFRAR）模块由选择性嵌入增强（SEA）和全局-局部特征细化（GLFR）组件构成，其集成使 GRAD-Former 在保持参数效率的同时，实现了最先进的性能。

这些实证发现验证了我们通过整合全局与局部特征来增强上下文理解的方法，并表明 GRAD-Former 能够有效解决在超高分辨率卫星影像中精确勾勒变化区域的方法学挑战。

参考文献

[1] S. Ji, S. Wei, and M. Lu, “用于从开放航拍与卫星影像数据集中提取多源建筑物的全卷积网络,” IEEE地球科学与遥感汇刊, 第57卷, 第1期, 第574–586页, 2019年。[2] R. Daudt, B. Saux, and A. Boulch, “用于变化检测的全卷积孪生网络,” 载于 IEEE国际图像处理会议, 2018年10月, 第4063–4067页。[3] Y. Liu, C. Pang, Z. Zhan, X. Zhang, and X. Yang, “基于双任务约束深度孪生卷积网络模型的遥感图像建筑物变化检测,” IEEE地球科学与遥感快报, 第18卷, 第5期, 第811–815页, 2021年。[4] S. Fang, K. Li, J. Shao, and Z. Li, “SNUNet-CD：用于甚高分辨率图像变化检测的密集连接孪生网络,” IEEE地球科学与遥感快报, 第PP卷, 第1–5页, 2021年2月。[5] G. Cheng, G. Wang, and J. Han, “ISNet：提升遥感图像变化检测的可分性,” IEEE地球科学与遥感汇刊, 第60卷, 第1–11页, 2022年。[6] C.-P. Chen, J.-W. Hsieh, P.-Y. Chen, Y.-K. Hsieh, and B.-S. Wang, “SARAS-Net：用于变化检测的尺度与关系感知孪生网络,” AAAI人工智能会议论文集, 第37卷, 第12期, 第14 187–14 195页, 2023年6月。 [在线]。获取地址：https://ojs.aaai.org/index.php/AAAI/article/view/26660[7] H. Chen, J. Song, C. Han, J. Xia, and N. Yokoya, “ChangeMamba：基于时空状态空间模型的遥感变化检测,” IEEE地球科学与遥感汇刊, 第62卷, 第1–20页, 2024年。[8] K. Chen, B. Chen, C. Liu, W. Li, Z. Zou, and Z. Shi, “Rsmamba：基于状态空间模型的遥感图像分类,” IEEE地球科学与遥感快报, 第21卷, 第1–5页, 2024年。[9] H. Chen, Z. Shi, and Z. Qi, “基于 Transformer 的遥感图像变化检测,” IEEE地球科学与遥感汇刊, 第60卷, 2021年7月。[10] W. G. C. Bandara and V. M. Patel, “用于变化检测的基于 Transformer 的孪生网络,” 载于 IGARSS 2022 - 2022年IEEE国际地球科学与遥感研讨会, 2022年, 第207–210页。[11] M. Noman, M. Fiaz, H. Cholakkal, S. Narayan, R. Muhammad Anwer, S. Khan, and F. Shahbaz Khan, “使用从头训练的变换器进行遥感变化检测,” IEEE地球科学与遥感汇刊, 第62卷, 第1–14页, 2024年。[12] M. Noman, M. Fiaz, H. Cholakkal, S. Khan, and F. S. Khan, “ELGC-Net：用于遥感变化检测的高效局部-全局上下文聚合,” IEEE地球科学与遥感汇刊, 2024年。[13] Z. Guo, H. Chen, and F. He, “MSFNet：用于遥感变化检测的多尺度空频特征融合网络,” IEEE应用地球观测与遥感专题期刊, 第18卷, 第1912–1925页, 2025年。[14] B. Wang, K. Zhao, T. Xiao, P. Qin, and J. Zeng, “RAHFF-Net：用于遥感图像变化检测的召回率可调层次特征融合网络,” IEEE应用地球观测与遥感专题期刊, 第18卷, 第176–190页, 2025年。[15] Y. Su, P. Ma, W. Wang, S. Wang, Y. Wu, Y. Li, and P. Jing, “AMDANet：用于图像变化检测的增强型多尺度差异聚合网络,” IEEE地球科学与遥感汇刊, 第1–1页, 2025年。[16] Z. Zhao, C. Wu, D. Wang, H. Chen, C. Chen, Z. Zheng, B. Du, and L. Zhang, “通过对抗式类别提示推进卫星图像的弱监督变化检测,” IEEE图像处理汇刊, 第34卷, 第7065–7078页, 2025年。