

基于时空感知与级联对齐的遥感影像变化检测框架

杨景玉^{1,2}, 张凯伟¹, 党建武^{1,2,3}, 王锋⁴, 火久元¹

1. 兰州交通大学 电子与信息工程学院, 兰州 730070;

2. 甘肃省人工智能与图形图像处理工程研究中心, 兰州 730070;

3. 兰州交通大学 轨道交通信息与控制国家级虚拟仿真实验教学中心, 兰州 730070;

4. 甘肃路桥飞宇交通设施有限公司, 兰州 730070

摘要: 遥感影像变化检测 CD (Change Detection) 是对地球表面进行动态监测的关键技术。然而, 因光照、季节变化等因素导致的时间不一致性与影像间几何失配导致的空间不一致性所形成的时空复合干扰, 严重混淆真实地物变化与伪变化, 降低检测精度。针对时空复合干扰问题, 本文提出了一种采用级联对齐策略的时空感知对齐与多尺度特征流 SAMFF (Spatio-temporal Aware Multi-scale Feature Flow) 的变化检测网络。首先, 在特征提取阶段, 设计跨时相注意力对齐模块 CTAAM (Cross-Temporal Attentional Alignment Module), 该模块以交互与重构的方式对齐双时相特征的通道语义与局部上下文表征, 增强对真实地物变化的感知能力, 有效抑制由光照、季节变化等非地物因素引起的光谱伪变化。其次, 针对几何失配问题, 设计了特征增强的流对齐差分模块 FADM (Flow-Aligned Difference Module), 该模块通过增强后的特征在不同尺度上对齐双时相特征, 缓解配准误差等带来的空间干扰。然后, 利用跨尺度注意力特征增强模块 CSAFE (Cross-Scale Attention Feature Enhancement) 对已对齐的时空特征进行多尺度特征增强, 以突出真实变化信息。最后, 利用置信度流引导的特征融合模块 CFGFM (Confidence Flow Guided Feature Fusion Module) 自适应地融合多层级对齐特征, 抑制由时间表现差异与几何失配复合引起的残余伪变化响应。在 SYSU-CD、WHU-CD 与 GZ-CD 3 个公开数据集上的实验表明, 本方法 F1 分数和交并比 IoU (Intersection over Union) 分别达到 83.83% 和 72.16%、93.44% 和 87.69% 和 88.40% 和 79.23%, 优于当前主流方法。

关键词: 变化检测, 遥感影像, 时空感知, 特征对齐, 多尺度融合, 深度学习

中图分类号: TP751/P2

引用格式: 杨景玉, 张凯伟, 党建武, 王锋, 火久元. 2026. 基于时空感知与级联对齐的遥感影像变化检测框架. 遥感学报, 30(7): 2227-2248

Yang J Y, Zhang K W, Dang J W, Wang F and Huo J Y. 2026. A spatiotemporal-aware and cascaded alignment framework for remote sensing image change detection. National Remote Sensing Bulletin, 30(7): 2227-2248 [DOI: 10.11834/jrs.20265478]

1 引言

遥感变化检测 RSCD (Remote Sensing Change Detection) 旨在通过分析同一地理区域不同时期获取的遥感影像, 识别地表目标的动态变化信息, 从而为理解地球表面的时空演变过程提供关键数据支撑。作为遥感领域的核心研究方向, RSCD 技术已广泛应用于城市扩张监测、土地资源管理、灾害评估、生态环境保护及土地利用分析等领域 (程堃等, 2025)。

然而, 随着遥感数据的激增与地表覆盖的复杂化, 现有方法在精确建模复杂时空变化模式方面仍面临根本性挑战 (付琨等, 2024)。具体而言, 成像条件差异 (如光照、太阳高度角和季节变化) 引入的时间不一致性, 与多视角成像几何配准误差导致的空间不一致性, 在实际数据中常同时出现, 共同构成时空复合干扰。这两类干扰源的物理成因独立, 但在观测与计算链路中叠加存在, 导致误差在特征差分、对齐与融合等环节中传播放大, 严重制约检测性能。此类干扰不仅引

收稿日期: 2025-11-17; 预印本: 2009-05-27

基金项目: 国家自然科学基金(编号:62367005); 甘肃省自然科学基金(编号:24JRRA253); 中央引导地方科技发展资金项目(编号:24ZYQA051)

第一作者简介: 杨景玉, 研究方向为遥感图像处理及遥感大数据应用。E-mail: yangjy@mail.lzjtu.cn

入伪变化噪声与对齐偏差,还会遮蔽真实变化信号(图1)。具体而言:(1)成像时间差异导致的环境变化,使同一地物呈现表观异质性(如图1(a)中屋顶因光照或季节变化产生的光谱差异),易被误判为真实变化。(2)传感器视角差异引起建筑物立面位移与轮廓畸变(图1(b)),若未充分几何校正,则会生成几何伪变化。在算法流程中,这

两类独立的物理误差会相互交织并产生级联影响:一方面,时间表观差异降低特征相似性,增大匹配不确定性,从而放大初始配准误差;另一方面,空间几何失配导致异质像元被错误关联,进而传播并放大表观差异在差分域中的伪变化。这种在算法层面相互加剧的混合干扰(图1(c)),最终显著降低了检测结果的可靠性。

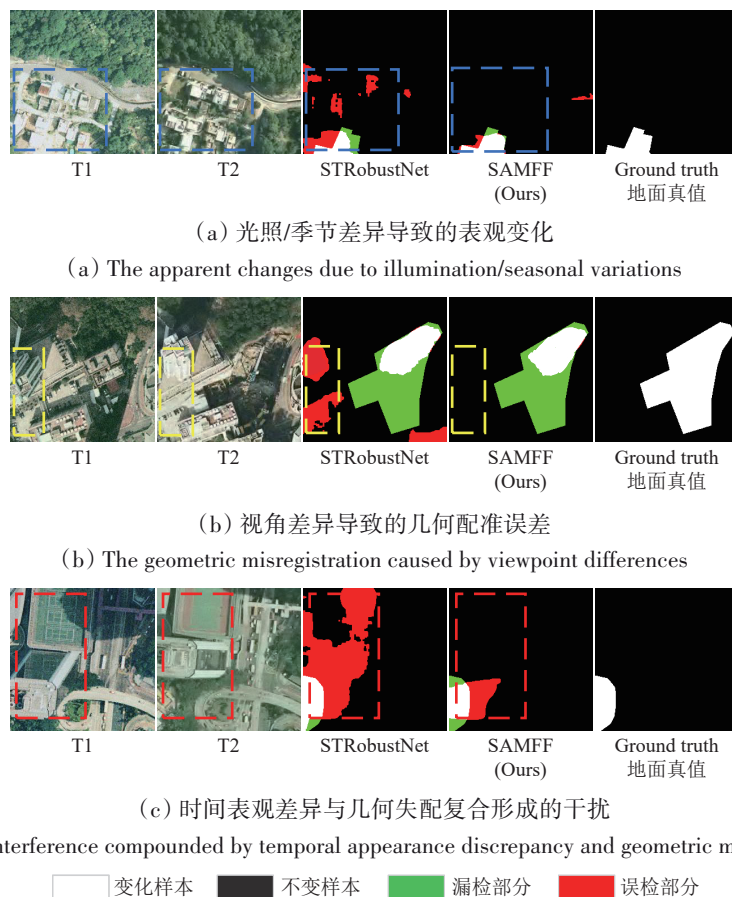


图1 遥感变化检测中导致误检的无关变化示例(虚线框代表需要重点关注的区域)

Fig.1 Exemplars of irrelevant changes leading to false detections in remote sensing change detection

(The dotted box represents the area that requires special attention)

传统变化检测方法,包括基于代数运算、变换(如主成分分析PCA(Principal Component Analysis)(Celik, 2009)、缨帽变换KT(Kauth-thomas Transformation)(Han等, 2007)、变化向量分析CVA(Change Vector Analysis)(Saha等, 2019))及分类(如支持向量机SVM(Support Vector Machine)(Negri等, 2021)、K近邻算法KNN(K-nearest neighbors)(Sun等, 2021))等技术,其核心在于依赖人工设计特征与经验性决策规则。然而,面对上述复合干扰,这些方法因缺乏对深层语义与几何关系的建模能力,难以实现有效抑制。

近年来,深度学习推动该领域进入端到端表征学习阶段。以卷积神经网络CNN(Convolutional Neural Network)(Ye等, 2023)和Transformer(Chen等, 2022b)为代表的网络,提升了对变化特征的判别能力。早期深度学习主要采用早期融合EF(Early Fusion)(Peng等, 2019)与后期融合LF(Late Fusion)(Zhao等, 2023)两类范式。EF在输入端进行通道拼接,易使光谱伪变化与几何配准误差在特征提取初期即深度交织,干扰真实变化辨识;LF虽先独立编码,但在特征提取阶段缺乏必要的交互机制,难以实现时序特征的有效对

齐与重构。由此可见，二者均未能在特征层面实现充分、及时的双时相交交互，因此在处理时空复合干扰时存在明显局限。

为应对上述挑战，现有研究主要沿着3条技术主线演进：面向时间差异的时相一致性建模、面向几何失配的空间对齐校正，以及尝试对二者进行联合建模与流程化约束。

在时相一致性建模方面，研究始于共享权重的孪生网络结构（如FC-Siam系列）（Dauidt等，2018），旨在通过参数共享提升特征一致性。随后，注意力机制与Transformer被广泛引入，以增强全局上下文与长程时序依赖的建模能力。例如双时相影像Transformer BIT（Bitemporal Image Transformer）（Chen等，2022b）利用Transformer解码器与语义令牌，对高层次语义的上下文交互进行编码。近期，特征交互时机向编码阶段前移：Changer网络通过元交互机制促进特征交换与重构（Fang等，2023）；IDA-SiamNet在编码器间引入交互与动态融合模块，以提升建筑物变化检测的鲁棒性（Li等，2024）。但上述交互模块本质上仍属于后融合范式（Jiang等，2023），难以在特征提取早期有效隔离由光照、季节等因素引起的表观伪变化，导致残留噪声随网络深度累积。更关键的是，这类方法主要聚焦于光谱差异鲁棒性，缺乏对像素级几何位移的显式建模与校正能力，因此在建筑物边缘等几何敏感区域，常因未校正的配准误差而产生误检。

在空间对齐校正方面，早期工作依赖互信息（Pluim等，2003）、尺度不变特征变换SIFT（Scale Invariant Feature Transform）（Lowe，2004）等传统特征匹配方法，但其在辐射差异大、纹理稀疏场景下稳定性有限。深度学习的引入推动了对齐技术向端到端范式演进。其中，参数回归网络（Rocco等，2018；Li等，2022）可预测全局变换，但对局部形变建模不足；而基于光流的方法（Dosovitskiy等，2015；Sun等，2018）能估计密集位移场，更适用于像素级非线性校正。在变化检测中，此类方法已形成多种技术路径：分辨率与对齐感知变化检测网络RACDNet（Resolution and Alignment-aware Change Detection Network）（Tian等，2022）利用可变形卷积适应几何形变；Li等（2024）利用光流场实现特征重采样；双时相特征对齐网络BiFA（Bitemporal Feature Alignment）（Zhang等，

2024）提出差分流场对齐；多注意力路径神经网络MAPNet（Multi Attending Path neural Network）（Zhao等，2024）利用光流引导注意力实现多尺度渐进融合；Liu等（2025）将光流概念拓展至多尺度特征空间以解决语义错位问题；同时，一些工作将对齐与差分表征联合设计以增强变化表征，如双时相对齐与差分特征融合的变化检测ChangeAD（Change Detection via Bi-Temporal Alignment and Differential Feature Integration）（Meng等，2024）。然而，这些方法存在一个共性局限：其对齐操作通常直接作用于受时间不一致性污染的特征。在未有效抑制表观伪变化的情况下，光流或形变估计难以区分真实几何位移与表观差异，从而导致对齐场偏差，甚至可能引入新的空间伪变化。

部分研究尝试对时空干扰进行联合建模。例如在编码器各层同步执行时域与空间对齐，即“边提取、边对齐”策略（Zhang等，2024）。但此策略引入了误差交叉传播的隐患：浅层未充分抑制的时域噪声（如残留光照伪影）会误导后续空间对齐过程；而早期遗留的配准误差也会干扰深层时域特征的语义对齐。此类误差在编码过程中沿层级不断累积与相互放大，最终严重制约网络对真实地物变化的判别能力。

此外，跨时相语义一致性与全局一致性学习成为抑制复杂噪声的关键路径。二者存在内在关联：语义一致性旨在保持无变化区域的表征稳定，缓解光照等表观差异引发的语义漂移；全局一致性则通过广域上下文提供全局约束，在多尺度融合中抑制局部噪声传播。相关研究已从局部交互转向全局约束：Zhu等（2022）在孪生网络中引入全局学习机制以增强语义一致性；Chen等（2022a）提出语义感知的密集表示学习，提升深层特征对齐能力；Cheng等（2024）通过内容净化过滤光照等非语义差异；Holail等（2025）借助视觉Transformer表明，全局一致性可为局部精细结构提供可靠语义先验。综上，全局一致性是实现跨时相语义对齐并抑制伪变化扩散的重要基础。然而，高分辨率影像中时间表观差异与几何失配常叠加形成时空复合干扰，现有方法多在混合特征空间中隐式追求对齐，缺乏对二者级联传播机制的显式建模，尤其在局部强几何畸变区域，易产生对齐模糊与伪变化残留，从而限制检测结果的稳健性。

针对上述因时间表观差异与几何失配在算法

流程中叠加并引发误差传播所形成的时空复合干扰问题, 本文提出逐级对齐的变化检测网络时空感知对齐与多尺度特征流 SAMFF (Spatio-temporal Aware Multi-scale Feature Flow)。其核心思想是将复合干扰分解为时间差异抑制与几何误差校正两个子任务, 构建级联处理架构, 实现逐级干扰抑制与判别性特征增强, 并在多尺度融合中施加几何与语义双一致性约束, 以抑制残留噪声传播。具体而言, 网络首先通过跨时相注意力对齐模块 CTAAM (Cross-Temporal Attentional Alignment Module) 和流对齐差分模块 FADM (Flow-Aligned Difference Module), 分别从通道语义交互与像素级几何配准两个维度完成初步伪变化抑制与跨时相特征对齐; 进而通过跨尺度注意力特征增强模块 CSAFE (Cross-Scale Attention Feature Enhancement) 聚合多尺度上下文信息, 建立跨尺度全局一致性, 提升差分表征的判别性与鲁棒性; 最后通过置信度流引导的特征融合模块 CFGFFM (Confidence Flow Guided Feature Fusion Module), 以双向流场与置信度调制完成多层次对齐特征在几何与语义上的双重一致性融合, 抑制残留噪声与语义错位, 输出高精度变化检测结果。本文通过协同设计的级联对齐框架, 系统实现了对时空复合干扰的逐级抑制与变化特征的鲁棒提取。

2 本文方法

2.1 整体框架

给定双时相遥感影像对 I_1 与 I_2 , 变化检测的核心目标是生成反映地表真实变化 Δ 的二值变化图 M 。然而, 实际高分辨率影像常受复杂的时空复合干扰, 其物理过程可抽象建模为 $I_2 = G(R(I_1 + \Delta)) + \epsilon$ 。其中, R 表示由季节、光照等因素引起的时间维辐射畸变, G 表示由视角、地形等因素导致的空间几何形变, ϵ 为噪声; R 与 G 在物理成因上相互独立, 但在成像结果中复合叠加, 导致直接学习映射 $F(I_1, I_2) \rightarrow M$ 时, 极易因干扰的非线性与复杂性而产生歧义或陷入局部最优。

特征对齐是实现高精度检测的关键环节: 只有在辐射与几何特征均得到充分对齐的特征空间中, 差异计算才能有效捕捉真实的地物语义变化; 同时, 系统化的对齐操作能够抑制误差在特征提取、差异计算与特征融合过程中的累积与传递,

从而为提升网络的判别能力与跨场景泛化性能提供重要基础。

为此, 本文提出级联对齐优化范式, 将上述复合干扰的抑制分解为两个序贯子任务, 并构建 SAMFF 网络予以实现 (图 2)。该网络首先进行时域特征对齐, 旨在学习特征映射 ΦT , 使双时相特征在辐射风格上趋于一致, 该过程不直接求解复杂的辐射逆变换, 而是通过特征间的深层交互, 隐式地学习一种对辐射畸变 R 具有不变性的特征表示, 从而在特征空间最小化非变化区域的特征差异 $\|\Phi T(I_1) - \Phi T(I_2)\|$ 。此任务由 CTAAM 实现, 其通过跨时相信号提纯输出蕴含真实变化的特征流 $\{F_{I_1}^i\}_{i=1}^4$ 和 $\{F_{I_2}^i\}_{i=1}^4$, 从而实现时域层面的特征对齐。继而进行空间几何对齐, 由 FADM 基于光流估计显式校正几何形变 G , 生成几何一致的多尺度差分特征流 $\{D^i\}_{i=1}^4$ 。之后, CSAFE 通过多尺度上下文聚合与注意力机制, 强化对齐特征中的判别性时空差异模式, 最终输出增强的多尺度特征流 $\{D_{eng}^i\}_{i=1}^4$ 。最后, $\{D_{eng}^i\}_{i=1}^4$ 被送入 CFGFFM 解码器, 该解码器通过其流引导的置信度融合机制, 自适应地对深、浅层特征进行空间对齐与加权融合, 有效整合多尺度上的时空变化信息, 并抑制由复合干扰残留导致的低置信度区域噪声, 最终解码为具有空间一致性与细节完整性的像素级变化检测图 M 。

2.2 跨时相注意力对齐模块 (CTAAM)

在双时相遥感影像变化检测中, 成像时间、视角、光照及季节差异会导致显著的光谱不一致性, 从而引入大量伪变化。尤其在多视角条件下, 时间不一致性与几何配准误差相互交织, 形成难以抑制的时空复合干扰, 严重影响网络对真实变化的感知。为抑制此类干扰, 本文提出 CTAAM, 其核心思想是在特征提取的每一层级, 通过通道级特征交互与融合, 挖掘并整合双时相特征间的时空依赖关系, 提升网络对真实变化的判别能力。该模块由双向注意力聚合与融合特征增强两大机制构成: 前者借助跨时相注意力 CTA (Cross-Temporal Attentional) 实现非局部空间关联, 在时域对齐的同时完成错位特征的隐式匹配, 缓解早期空间校正缺失带来的对齐偏差; 后者通过双分支增强结构优化特征质量, 其中通道语义增强分支 CSEB (Channel Semantic Enhancement Branch) 显式抑制

对时域干扰敏感的通道响应, 局部上下文感知分支 LCB (Local Context-aware Branch) 则借助空间感知能力强化细节保持与噪声鲁棒性。该设计缓解了时间不一致性导致的光谱干扰, 为后续空间

对齐模块提供了去噪后的特征输入, 有效抑制了时域误差在层级间的累积传播, 从而改善了网络在复杂场景下的变化感知鲁棒性, 其整体结构如图 3 所示。

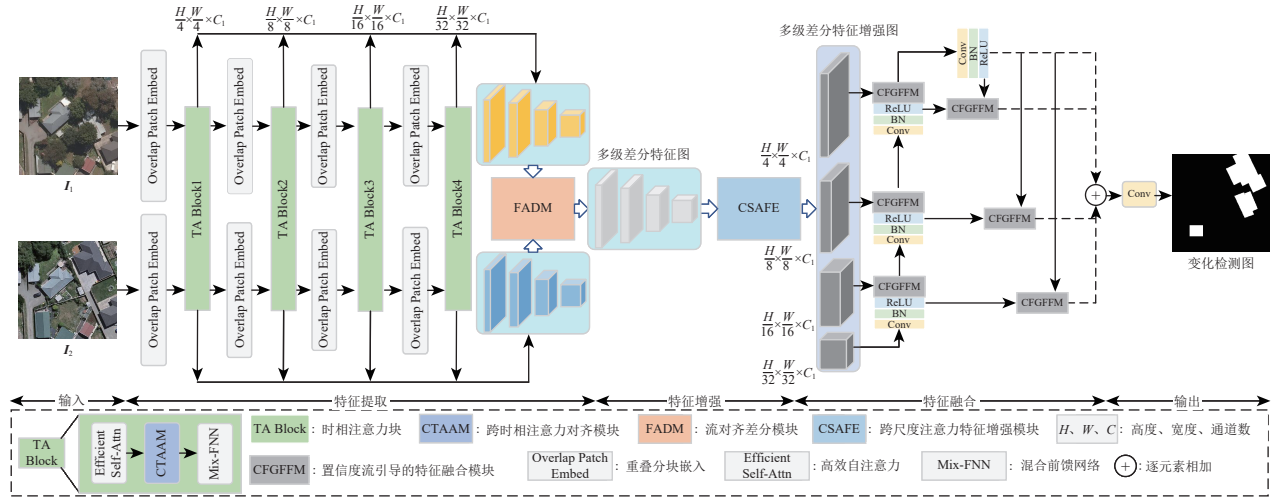


图2 SAMFF网络结构

Fig.2 Architecture of the SAMFF network

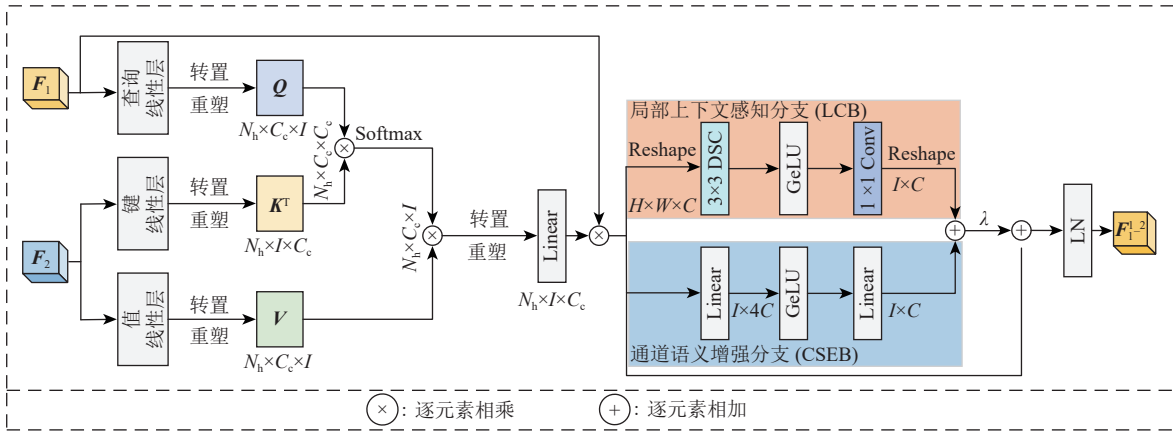


图3 跨时相注意力对齐模块

Fig.3 Cross-temporal attentional alignment module

以 F_2 引导 F_1 为例 (省略阶段标记), 给定同层级双时相特征 $F_1 \in \mathbb{R}^{H \times W \times C}$ 与 $F_2 \in \mathbb{R}^{H \times W \times C}$ 。将 F_1 经线性投影、重塑与转置生成查询向量 $Q \in \mathbb{R}^{N_h \times C_c \times I}$, F_2 并行经两独立线性投影层生成键向量 $K \in \mathbb{R}^{N_h \times C_c \times I}$ 和值向量 $V \in \mathbb{R}^{N_h \times C_c \times I}$ 。其中, $I = H \times W$, 为序列长度; N_h 为头数; C_c 为单头通道维度。计算 Q 与 K^T 的点积, 经缩放归一化后应用 Softmax 得到注意力权重图, 该权重图与 V 通过矩阵乘法进行聚合。最后, 多头输出拼接后经线性投影生成聚合特征 F_a , 并通过残差连接与原始输入特征 F_1 相加得到融合特征 F_{add} 。上述过程如下:

$$Q = F_1 W^q \quad (1)$$

$$K = F_2 W^k \quad (2)$$

$$V = F_2 W^v \quad (3)$$

$$\text{head}_j = \sigma \left(\frac{QK^T}{\sqrt{C_c}} \right) V \quad (4)$$

$$F_a = \text{Concat}(\text{head}_1, \dots, \text{head}_j) W^0 \quad (5)$$

$$F_{add} = F_1 + F_a \quad (6)$$

式中, W^q 、 W^k 和 $W^v \in \mathbb{R}^{C \times C}$, 分别为查询、键和值对应的可学习线性投影矩阵; $\text{Concat}(\cdot)$ 表示拼接操作; C_c 是单头通道维度; head_j 表示注意力头, j 为

注意力头索引, 表示第 j 个注意力头; $\mathbf{W}^0 \in \mathbf{R}^{N_h \times C_c \times C_c}$, 为输出线性投影矩阵; $\sigma(\cdot)$ 表示 Softmax 操作。

完成初步的时空信息聚合后, 为进一步增强特征判别性, 增强网络对真实变化的感知能力, 本文引入了一种并行的双分支增强机制, 旨在对融合特征 $\mathbf{F}_{\text{add}}$ 进行深度提纯, 突出变化敏感特征并抑制噪声, 从而同时捕捉全局长程依赖与局部细节特征。

针对遥感影像中由光照、视角变化等因素引发的局部伪变化问题, LCB 分支通过 3×3 深度可分离卷积显式建模局部邻域空间依赖, 经 GeLU 激活后, 通过 1×1 逐点卷积融合空间拓扑与通道信息, 生成联合表征 $\mathbf{F}_{\text{local}} \in \mathbf{R}^{I \times C}$, 该过程可有效区分真实地物结构性变化与表观伪变化, 提升网络对局部细节的辨识鲁棒性与感知精度。上述过程如下:

$$\mathbf{F}_{\text{local}} = \text{Conv}_{1 \times 1} \left(\text{GeLU} \left(\text{DSC}_{3 \times 3}(\mathbf{F}_{\text{add}}) \right) \right) \quad (7)$$

式中, $\text{Conv}_{1 \times 1}(\cdot)$ 表示 1×1 卷积; $\text{GeLU}(\cdot)$ 表示激活函数处理; $\text{DSC}_{3 \times 3}(\cdot)$ 表示 3×3 深度可分离卷积。

为抑制季节更替与成像条件差异引发的光谱伪变化, CSEB 分支首先通过降维投影压缩特征至低维空间, 抑制冗余并提取全局语义先验; 继而利用 GeLU 激活函数进行非线性变换以解耦复杂模式; 最后经升维操作恢复通道维度, 输出语义增强特征 $\mathbf{F}_{\text{semantic}} \in \mathbf{R}^{I \times C}$ 。此过程自适应地抑制对时域变化敏感的通道响应, 增强特征的特征鲁棒性。上述过程如下:

$$\mathbf{F}_{\text{semantic}} = \text{Linear}_{\text{up}} \left(\text{GeLU} \left(\text{Linear}_{\text{down}}(\mathbf{F}_{\text{add}}) \right) \right) \quad (8)$$

式中, $\text{Linear}_{\text{down}}(\cdot)$ 表示降维操作; $\text{GeLU}(\cdot)$ 表示激活函数处理; $\text{Linear}_{\text{up}}(\cdot)$ 表示升维操作。

最终, 两分支输出经可学习标量权重 λ 与 $\mathbf{F}_{\text{add}}$ 进行自适应融合, 最终以残差连接的方式与注意力阶段的输出相结合, 得到融合特征 $\mathbf{F}_{\text{fused}} \in \mathbf{R}^{I \times C}$, 上述过程如下:

$$\mathbf{F}_{\text{fused}} = \lambda(\mathbf{F}_{\text{local}} + \mathbf{F}_{\text{semantic}}) + \quad (9)$$

式中, $\lambda(\cdot)$ 表示权重融合操作。

最终在通道维度上实现了时域特征对齐, 并输出多尺度特征流, 从而在特征提取阶段抑制时间不一致性干扰, 为后续空间层面的特征对齐提供鲁棒的特征基础。

2.3 流对齐差分模块(FADM)

在双时相遥感影像变化检测中, 多视角成像

固有的几何配准误差常导致像素级空间位移, 引入伪变化噪声, 干扰真实变化区域识别。为缓解几何失配问题并抑制误差在层级间的传递, 本文提出 FADM。该模块以 CTAAM 输出的时域对齐特征作为输入, 首先通过 MSCA 预增强机制, 对特征进行几何结构强化, 抑制时域对齐过程中可能残留的噪声, 从而为后续光流估计提供判别性特征。然后, 将双时相特征的空间偏移建模为密集时空差分流场 (spatio-temporal differential flow field)。该流场表征了从一时相特征到另一时相特征的像素级位移映射, 并通过基于流场的形变变换实现精确的空间对齐。上述分阶段的特征增强与对齐策略, 提高了几何位移的建模精度, 缓解了空间未对齐误差向解码阶段的传递, 增强了网络对真实变化区域的空间感知一致性, 缓解了配准误差被误判为语义变化的问题, 最终输出多尺度差分特征流。

网络具体结构如图4所示。为简化表述, 本文以第一阶段为例进行阐述, 输入的双时相特征分别表示为 $\mathbf{F}_1^{1,2}$ 与 $\mathbf{F}_1^{2,1}$ 。首先, 经 MSCA 模块增强特征表示, 以确保后续流场估计的准确性与鲁棒性。对于输入特征 $\mathbf{F}_{\text{in}} \in \mathbf{R}^{I \times C}$, 通过 1×1 卷积升维, 得到升维特征 $\mathbf{F}_{\text{m}} \in \mathbf{R}^{I \times C}$ 。随后, 并行采用两个不同扩张率 ($r = 1, 2$) 的 3×3 深度可分离卷积捕获多尺度上下文并融合输出, 得到融合特征 $\mathbf{F}'_{\text{fuse}} \in \mathbf{R}^{I \times C}$ 。为增强对关键通道的感知, 引入通道注意力机制对特征进行调制, 得到特征 $\mathbf{F}_{\text{ch}} \in \mathbf{R}^{I \times C}$ 。最后, 通过 1×1 卷积恢复通道维度并与输入残差连接, 输出增强特征 $\mathbf{F}_{\text{out}}^e \in \mathbf{R}^{I \times C}$ 。上述过程可以表示如下:

$$\mathbf{F}_{\text{m}} = \text{GeLU} \left(\text{BN} \left(\text{Conv}_{1 \times 1}(\mathbf{F}_{\text{in}}) \right) \right) \quad (10)$$

$$\mathbf{F}'_{\text{fuse}} = \text{GeLU} \left(\text{BN} \left(\text{DSC}_{3 \times 3}^{r=1}(\mathbf{F}_{\text{m}}) + \text{DSC}_{3 \times 3}^{r=2}(\mathbf{F}_{\text{m}}) \right) \right) \quad (11)$$

$$\mathbf{F}_{\text{ch}} = \sigma \left(\text{Conv}_{1 \times 1} \left(\text{ReLU} \left(\text{Conv}_{1 \times 1} \left(\text{AAP}(\mathbf{F}'_{\text{fuse}}) \right) \right) \right) \right) \quad (12)$$

$$\mathbf{F}_{\text{out}}^e = \text{BN} \left(\text{Conv}_{1 \times 1}(\mathbf{F}'_{\text{fuse}} \times \mathbf{F}_{\text{ch}}) \right) + \mathbf{F}_{\text{in}} \quad (13)$$

式中, $\text{BN}(\cdot)$ 为批归一化; $\text{GeLU}(\cdot)$ 表示激活函数处理; $\text{Conv}_{1 \times 1}(\cdot)$ 表示 1×1 卷积; $\text{DSC}_{3 \times 3}^{r=1}(\cdot)$ 表示 $r = 1$ 的 3×3 深度可分离卷积; $\text{DSC}_{3 \times 3}^{r=2}(\cdot)$ 表示 $r = 2$ 的 3×3 深度可分离卷积; $\sigma(\cdot)$ 表示 Sigmoid 操作; $\text{ReLU}(\cdot)$ 表示修正线性单元; $\text{AAP}(\cdot)$ 表示自适应平均池化操作。

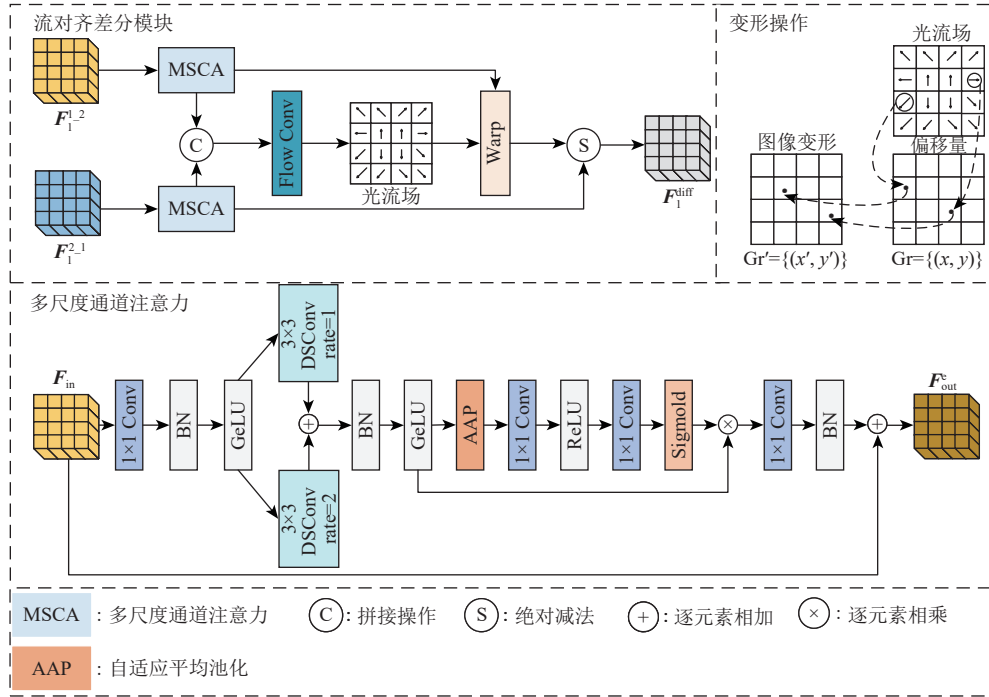


图4 流对齐差分模块

Fig.4 Flow-aligned difference module

经过 MSCA 模块处理后，原始的双时相特征 $F_1^{1,2}$ 与 $F_1^{2,1}$ 被增强为语义信息更鲁棒、感知能力更强的特征 F_{out1}^e 与 F_{out2}^e 。随后，将增强后的特征拼接在一起，并通过 3×3 卷积生成两通道的时空差分场，分别代表特征图的水平位移和垂直位移：

$$F_{out1}^e = \text{MSCA}(F_1^{1,2}) \quad (14)$$

$$F_{out2}^e = \text{MSCA}(F_1^{2,1}) \quad (15)$$

$$\mathbf{F}\mathbf{l} = \text{Conv}_{\text{flow}}(\text{Concat}(F_{out1}^e, F_{out2}^e)) \quad (16)$$

式中， $\text{MSCA}(\cdot)$ 表示 MSCA 模块处理； $\mathbf{F}\mathbf{l} \in \mathbf{R}^{2 \times H \times W}$ 表示时空差分场； $\text{Conv}_{\text{flow}}(\cdot)$ 表示 3×3 卷积； $\text{Concat}(\cdot)$ 表示拼接操作。

在光流形变阶段，基于时空差分场提供的位移信息构建标准空间网格 $\mathbf{G}\mathbf{r}=\{(x, y)|x \in [1, h], y \in [1, w]\}$ ，其高度 h 与宽度 w 沿垂直轴与水平轴在归一化区间 $[-1, 1]$ 内均匀离散化。通过笛卡尔积生成网格坐标后，对网格点 (x, y) 施加光流场形变变换：

$$(x, y) \rightarrow (x', y') = (x, y) + (u(x, y), v(x, y)) \quad (17)$$

式中， $u(x, y)$ 与 $v(x, y)$ 分别表示时空差分场的水平和垂直位移分量，变形后网格 $\mathbf{G}\mathbf{r}'=\{(x', y')\}$ 定义了特征重映射坐标。具体地，位移场 $\mathbf{F}\mathbf{l}$ 编码了特征 F_{out1}^e 相对于标准网格 $\mathbf{G}\mathbf{r}$ 的偏移量。因此，通过将网格点 $p_1 \in \mathbf{G}\mathbf{r}$ 的初始位置与 $\mathbf{F}\mathbf{l}$ 在该点的偏

移量叠加得到扭曲坐标 $\hat{p}$ ：

$$\hat{p} = p_1 + \mathbf{F}\mathbf{l} \quad (18)$$

为了消除坐标偏移导致的采样空缺，采用可微分双线性插值机制计算扭曲特征 F_{wrap} ，最终通过对齐特征 F_{wrap} 与 F_{out2}^e 时相特征执行逐元素绝对减法运算，生成像素级对齐的双时相差分特征 F_1^{diff} ：

$$F_{wrap}(p_{wrap}) = \sum_{p \in N(\hat{p})} w_p F_{out1}^e \quad (19)$$

$$F_1^{diff} = S(F_{wrap}, F_{out2}^e) \quad (20)$$

式中， p_{wrap} 表示 F_{wrap} 中点的位置； $N(\hat{p})$ 表示 $\hat{p}$ 的四邻域采样点（左上、左下、右上、右下）； w_p 为扭曲空间网格上的双线性核权重； $S(\cdot, \cdot)$ 表示绝对减法运算。

FADM 模块通过显式建模与补偿几何位移，有效缓解了空间失配干扰，为后续特征增强与变化识别提供了判别性更强、几何鲁棒的差分特征流。

2.4 跨尺度注意力特征增强模块(CSAFE)

在通过 CTAAM 模块与 FADM 模块分别完成时域与空间特征对齐后，所得的多尺度差分特征仍面临两个关键挑战：其一，各尺度特征语义判别力不足，难以区分真实变化与残余干扰；其二，深层特征语义丰富但空间细节模糊，浅层特征定位精确但语义表征薄弱，且缺乏跨尺度交互机制。传

统差分方法因特征模糊与上下文信息缺失，难以建模复杂变化模式。为此，本文提出 CSAFE，如图 5 所示。该模块一方面通过多尺度空洞卷积金字塔增强上下文感知，另一方面引入自顶向下的跨

尺度注意力融合机制，实现深层语义向浅层特征的定向传递，从而提升变化区域的语义一致性、结构完整性与跨尺度关联表达能力，同时抑制残余噪声。

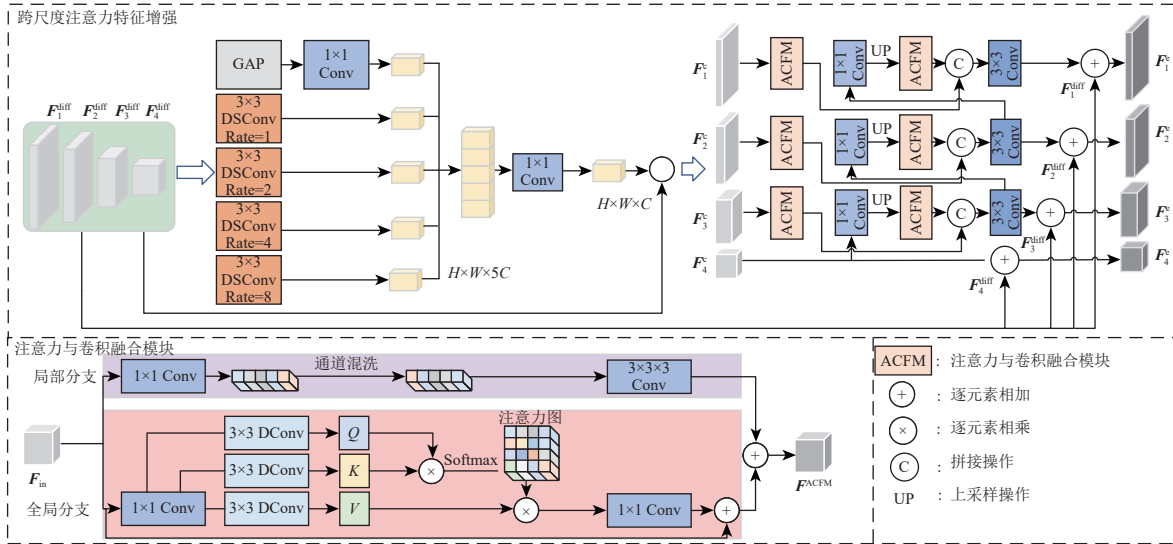


图 5 跨尺度注意力特征增强模块

Fig.5 Cross-scale attention feature enhancement module

对于输入的多尺度差分特征流 $\{F_i^{diff}\}_{i=1}^4$ ，CSAFE 模块首先采用多尺度空洞卷积金字塔增强各层级特征的上下文感知能力。以第 i 层的差分特征 F_i^{diff} 为例，通过并行多分支结构提取多尺度上下文信息：

$$F_i^e = PE(F_i^{diff}) + F_i^{diff} \quad (21)$$

式中， $PE(\cdot)$ 表示五分支并行的多尺度特征提取操作。后四条路径采用不同空洞率 ($r = 1, 2, 4, 8$) 的深度可分离卷积捕获多感受野特征；第一条路径则通过全局平均池化建模全局语义信息。随后，所有分支的输出沿通道拼接后经 1×1 卷积下采样，再与原始差分特征 F_i^{diff} 逐元素相加得到增强后的特征 F_i^e 。该过程有效融合多尺度空间上下文，增强对不同尺度变化目标的感知能力。相关公式如下：

$$PE(F_i^{diff}) = \text{Conv}_{1 \times 1} \left(\text{Concat} \left(F_{i,r=1}^{DSC}, F_{i,r=2}^{DSC}, F_{i,r=4}^{DSC}, F_{i,r=8}^{DSC}, F_i^{GAP} \right) \right) \quad (22)$$

式中， $F_{i,r}^{DSC}$ 表示空洞率为 ($r = 1, 2, 4, 8$) 的深度可分离卷积输出特征； F_i^{GAP} 表示全局平均池化分支特征； $\text{Conv}_{1 \times 1}(\cdot)$ 表示 1×1 卷积； $\text{Concat}(\cdot)$ 表示拼接操作。

为促进深层高级语义特征与浅层低级语义特

征间的有效信息传递，CSAFE 模块采用自顶向下的跨尺度注意力融合策略，从最高层 $i = 4$ 开始，逐层向下传递并融合增强特征。对于第 3、2、1 层，其最终增强特征 F_i^e 的生成过程如下：

$$F_i^e = \text{Fusion}(F_i^e, F_{i+1}^e) + F_i^{diff} \quad (23)$$

融合函数 $\text{Fusion}(\cdot)$ 基于 ACFM (Chen 和 Lu, 2025) 实现高效的特征交互。具体而言，高层特征 F_{i+1}^e 经 1×1 卷积与上采样对齐至当前尺度，随后与当前层特征 F_i^e 分别输入 ACFM 模块增强后拼接，再经 3×3 卷积实现跨尺度融合。具体过程如下：

$$F_{i+1}^{up} = \text{UP}(\text{Conv}_{1 \times 1}(F_{i+1}^e)) \quad (24)$$

$$F_{i,e}^{ACFM} = \text{ACFM}(F_i^e) \quad (25)$$

$$F_{i+1,up}^{ACFM} = \text{ACFM}(F_{i+1}^{up}) \quad (26)$$

$$\text{Fusion}(F_i^e, F_{i+1}^e) = \text{Conv}_{3 \times 3} \left(\text{Concat}(F_{i,e}^{ACFM}, F_{i+1,up}^{ACFM}) \right) \quad (27)$$

式中， $\text{UP}(\cdot)$ 表示上采样； $\text{ACFM}(\cdot)$ 表示 ACFM 模块； $\text{Conv}_{1 \times 1}(\cdot)$ 表示 1×1 卷积； $\text{Conv}_{3 \times 3}(\cdot)$ 表示 3×3 卷积； $\text{Concat}(\cdot)$ 表示拼接操作。

ACFM 模块采用并行的局部卷积与全局注意力双分支架构实现高效特征交互。全局分支采用注意力机制强化长程信息交互：输入特征 F_{in} 首先通

过 1×1 卷积与 3×3 深度可分离卷积生成查询 Q 、键 K 和值 V 张量，其维度均为 $\hat{H} \times \hat{W} \times \hat{C}$ ($\hat{H}$ 和 $\hat{W}$ 为空间尺寸， $\hat{C}$ 为通道数)。将 Q 重塑为 $\hat{Q} \in \mathbb{R}^{\hat{H} \times \hat{W} \times \hat{C}}$ ， K 重塑为 $\hat{K} \in \mathbb{R}^{\hat{C} \times \hat{H} \times \hat{W}}$ 。为避免传统注意力图 $\mathbb{R}^{\hat{H} \times \hat{W} \times \hat{H} \times \hat{W}}$ 的高计算开销，对 $\hat{Q}$ 和 $\hat{K}$ 进行 Softmax 归一化，生成轻量化注意力图 $A \in \mathbb{R}^{\hat{C} \times \hat{C}}$ 。该分支计算过程表述为

$$f_{\text{att}} = \text{Conv}_{1 \times 1} \left(\text{Attention}(\hat{Q}, \hat{K}, \hat{V}) \right) + F_{\text{in}} \quad (28)$$

$$\text{Attention}(\hat{Q}, \hat{K}, \hat{V}) = \hat{V} \text{Softmax} \left(\frac{\hat{Q} \hat{K}}{\alpha} \right) \quad (29)$$

式中， f_{att} 为输出特征； $\text{Conv}_{1 \times 1}(\cdot)$ 表示 1×1 卷积； F_{in} 表示输入特征； α 调节点积幅值。

局部分支首先通过 1×1 卷积调整通道维度，随后沿通道轴划分特征组并进行通道混洗，在各组内应用深度可分离卷积以促进通道交互。组间输出沿通道轴拼接后，经 $3 \times 3 \times 3$ 卷积完成特征提取：

$$f_{\text{conv}} = \text{Conv}_{3 \times 3 \times 3} \left(\text{CS} \left(\text{Conv}_{1 \times 1} (F_{\text{in}}) \right) \right) \quad (30)$$

式中， f_{conv} 为输出特征； $\text{Conv}_{3 \times 3 \times 3}(\cdot)$ 为 $3 \times 3 \times 3$ 三维卷积；CS 表示通道混洗操作。

这种双分支设计使模块能够同时关注变化的边

界细节和整体结构，其输出是两个分支结果的加和：

$$F^{\text{ACFM}} = f_{\text{att}} + f_{\text{conv}} \quad (31)$$

最终，融合后的特征与原始差分特征进行残差连接，形成增强后的多层级差分特征表示 F_i^e ，有效减缓了原始变化信息的丢失并促进了梯度的稳定传播。CSAFE 模块通过上述多尺度空洞卷积金字塔与自顶向下的跨尺度注意力融合机制，将经过时空对齐后的多尺度差分特征转化为语义连贯、判别性强的增强特征流，为后续解码阶段提供了高质量的特征输入。

2.5 置信度流引导的特征融合模块 (CFGFFM)

传统双线性插值融合方法基于均匀邻域假设，难以建模经前序模块优化后多尺度特征间的复杂跨尺度语义依赖，易引发特征回传错位。此外，前序对齐模块（如 FADM）在弱纹理或遮挡区域的对齐结果存在不确定性，若将此类误差直接引入融合阶段，将导致变化信息的累积损失，并降低最终结果可靠性。

针对上述问题，本文设计 CFGFFM，其结构如图 6 所示。该模块包含 3 个阶段：空间注意力引导的特征预处理、置信度感知的特征光流变形以及门控注意力融合。

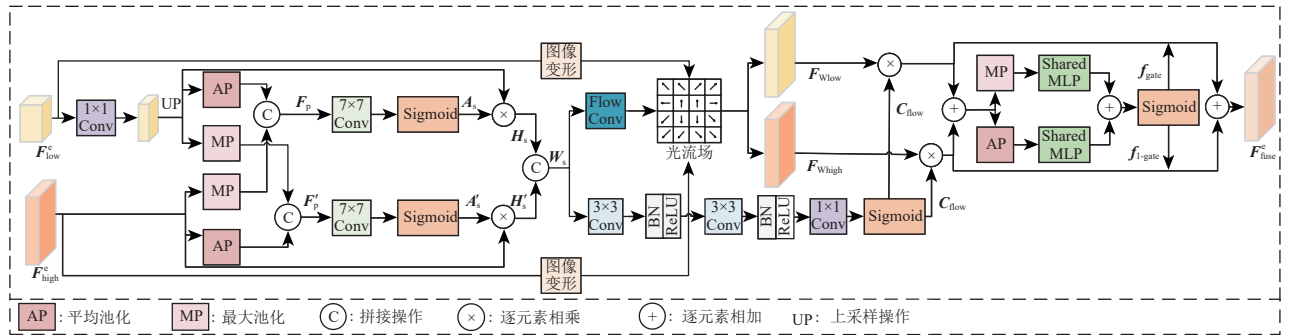


图6 置信度流引导的特征融合模块

Fig.6 Confidence flow guided feature fusion module

在空间注意力引导的特征预处理阶段，为增强对关键变化区域的感知能力并抑制无效信息干扰，首先通过 1×1 卷积将低分辨率特征 F_{low}^e 的通道数调整至高分辨率特征 F_{high}^e 维度，并经双线性插值初步上采样得到 $F_{\text{low}}^{\text{up}}$ 。随后对 F_{high}^e 与 $F_{\text{low}}^{\text{up}}$ 分别执行通道维度的最大池化 MP (Max Pooling) 与平均池化 AP (Average Pooling)，通过交叉融合策略构建跨尺度上下文特征 F_p 与 F'_p ，再经 7×7 卷积与 Sigmoid 激活生成空间特征权重 A_s 与 A'_s ，以实现

不同空间位置的差异化加权。将权重 A_s 与特征图 $F_{\text{low}}^{\text{up}}$ 逐元素相乘，得到加权特征 H_s ，同理可得 H'_s 。最终，将 H_s 与 H'_s 沿通道维拼接，形成用于光流估计与置信度估计的特征图 W_s 。该过程的表述为

$$F_{\text{low}}^{\text{up}} = \text{UP} \left(\text{Conv}_{1 \times 1} (F_{\text{low}}^e) \right) \quad (32)$$

$$F_p = \text{Concat} \left(\text{MP} (F_{\text{high}}^e), \text{AP} (F_{\text{low}}^{\text{up}}) \right) \quad (33)$$

$$A_s = \sigma \left(\text{Conv}_{7 \times 7} (F_p) \right) \quad (34)$$

$$H_s = A_s \times F_{\text{low}}^{\text{up}} \quad (35)$$

$$\mathbf{W}_s = \text{Concat}(\mathbf{H}_s, \mathbf{H}'_s) \quad (36)$$

式中, $\text{UP}(\cdot)$ 表示上采样; $\text{Conv}_{1 \times 1}(\cdot)$ 表示 1×1 卷积; $\text{Concat}(\cdot)$ 表示拼接操作; $\sigma(\cdot)$ 表示 Sigmoid 函数; $\text{Conv}_{7 \times 7}(\cdot)$ 表示 7×7 卷积; $\text{AP}(\cdot)$ 表示平均池化; $\text{MP}(\cdot)$ 表示最大池化。

在置信度感知的特征光流变形阶段, 为应对弱纹理、遮挡等场景下光流估计可靠性下降的问题, 引入了置信度感知机制。该机制旨在量化几何对齐过程的可信度, 并抑制由低质量对齐区域引入的干扰。具体而言, 特征图 $\mathbf{W}_s$ 通过 3×3 卷积生成

$$\mathbf{C}_{\text{flow}} = \sigma \left(\text{Conv}_{1 \times 1} \left(\text{ReLU} \left(\text{BN} \left(\text{Conv}_{3 \times 3} \left(\text{ReLU} \left(\text{BN} \left(\text{Conv}_{3 \times 3}(\mathbf{W}_s) \right) \right) \right) \right) \right) \right) \right) \quad (38)$$

$$\mathbf{F}_{\text{wlow}}^c = \text{Wrap}(\mathbf{F}_{\text{low}}^e, \mathbf{F}_{\text{low}}^c) \times \mathbf{C}_{\text{flow}} \quad (39)$$

$$\mathbf{F}_{\text{whigh}}^c = \text{Wrap}(\mathbf{F}_{\text{high}}^e, \mathbf{F}_{\text{high}}^c) \times \mathbf{C}_{\text{flow}} \quad (40)$$

式中, $\text{Conv}_{\text{flow}}(\cdot)$ 与 $\text{Conv}_{3 \times 3}(\cdot)$ 表示 3×3 卷积; $\mathbf{C}_{\text{flow}}$ 表示置信度图; $\sigma(\cdot)$ 表示 Sigmoid 函数; $\text{BN}(\cdot)$ 表示批归一化; $\text{ReLU}(\cdot)$ 表示修正线性单元; $\text{Conv}_{1 \times 1}(\cdot)$ 表示 1×1 卷积; $\text{Wrap}(\cdot)$ 表示变形操作; $\mathbf{F}_{\text{wlow}}^c$ 与 $\mathbf{F}_{\text{whigh}}^c$ 分别表示经置信度加权后的低分辨率特征图与高分辨率特征图。

在门控注意力融合阶段, 为实现对齐后多尺度特征的自适应融合, 引入门控机制。该机制首先将置信度加权后的特征相加, 形成联合特征 $\mathbf{F}_{\text{wsum}}$ 。随后, $\mathbf{F}_{\text{wsum}}$ 被输入通道注意力模块。该模块并行执行全局 AP 与全局 MP, 通过共享的多层感知机提取通道注意力权重, 并经 Sigmoid 函数激活生成门控系数 f_{gate} 。在特征融合过程中, f_{gate} 与 $1 - f_{\text{gate}}$ 别用于调节低分辨率语义特征和高分辨率细节特征的响应强度, 使网络能够自适应平衡全局语义信息与局部细节信息, 从而获得更具判别性的融合特征。最终, 通过加权求和得到融合特征 $\mathbf{F}_{\text{fuse}}^e$, 该特征能同时捕捉全局语义与局部细节。该过程表述如下:

$$\mathbf{F}_{\text{wsum}} = \mathbf{F}_{\text{wlow}}^c + \mathbf{F}_{\text{whigh}}^c \quad (41)$$

$$f_{\text{gate}} = \sigma \left(\text{MLP} \left(\text{MP}(\mathbf{F}_{\text{wsum}}) \right) \right) + \text{MLP} \left(\text{AP}(\mathbf{F}_{\text{wsum}}) \right) \quad (42)$$

$$\mathbf{F}_{\text{fuse}}^e = \mathbf{F}_{\text{wlow}}^c \times f_{\text{gate}} + (1 - f_{\text{gate}}) \times \mathbf{F}_{\text{whigh}}^c \quad (43)$$

式中, $\sigma(\cdot)$ 表示 Sigmoid 函数; $\text{AP}(\cdot)$ 表示平均池化; $\text{MP}(\cdot)$ 表示最大池化; $\text{MLP}(\cdot)$ 表示多层感知机。

如图 2 所示, CFGFFM 模块同时用于相邻层级与跨层级特征融合。在相邻层级融合中, 特征在相邻层级间逐步传递与精化。为进一步提升特征

四通道光流场 $\mathbf{F}\mathbf{I}$, 其中前两通道与后两通道分别编码高、低分辨率特征图的水平和垂直位移。同时, $\mathbf{W}_s$ 通过一个光流置信度网络处理, 生成值域在 $[0, 1]$ 之间的单通道置信度图 $\mathbf{C}_{\text{flow}}$, 以表征每个像素位置光流估计的可靠性。在变形阶段, 采用网格采样技术, 根据光流场中的位移对原始特征 $\mathbf{F}_{\text{wlow}}$ 与 $\mathbf{F}_{\text{whigh}}$ 进行重采样 (Wrap)。变形后的特征与置信度图 $\mathbf{C}_{\text{flow}}$ 逐元素相乘, 从而抑制低置信度区域的特征响应, 增强对齐鲁棒性。该过程表述如下:

$$\mathbf{F}\mathbf{I} = \text{Conv}_{\text{flow}}(\mathbf{W}_s) \quad (37)$$

表达能力并稳定数值分布, 每个 CFGFFM 模块的输出均连接一个标准卷积块 (结构为: 卷积—批量归一化—激活函数)。跨层级融合则将不同层级的优化特征与经由相邻层级传播获得的最高分辨率特征进行对齐与整合, 从而实现多尺度信息的高效融合。最终, 4 个层级经对齐与融合后得到的细粒度特征被用于生成判别性更强的最终变化图。

3 实验

3.1 数据集

实验采用 3 个公开可获取的 RSCD 数据集进行训练、验证与测试, 分别为: SYSU-CD 数据集 (Shi 等, 2022)、WHU-CD 数据集 (Ji 等, 2019) 和 GZ-CD 数据集 (Yan 等, 2023)。

(1) SYSU-CD 数据集由中山大学发布, 包含 20000 对 0.5 m 分辨率的航空影像 (256 像素 $\times$ 256 像素), 配有二进制变化标注图。该数据集影像拍摄于 2007 年—2014 年的香港地区, 涵盖建筑、港口、植被、道路等多种变化类型, 复杂性高。数据集分为训练集、验证集和测试集, 分别包含 12000 对训练样本、4000 对验证样本和 4000 对测试样本, 是变化检测领域权威基准数据集。

(2) WHU-CD 数据集由武汉大学发布, 包含 2012 年与 2016 年新西兰地区的两期航空影像, 分辨率为 0.2 m。该数据集原始图像尺寸为 32507 像素 $\times$ 15354 像素, 裁剪为 256 像素 $\times$ 256 像素块后, 按 8 : 1 : 1 的比例划分为 5948 对训练样本、744 对验证样本、745 对测试样本。数据集标注了不同时期建筑物数量的变化, 为建筑物变化检测提供了高质量、高分辨率的基准数据。

(3) GZ-CD数据集是高分辨率卫星影像变化检测的基准数据集之一，涵盖2006年—2019年广州郊区的0.55 m分辨率影像，包含19对存在季节差异的影像对。原始图像尺寸介于1006像素×1168像素至4936像素×5224像素。所有影像均裁剪为256像素×256像素的非重叠图像块，并按8:1:1的比例划分训练集2505对、验证集314对、测试集314对。

3.2 实验设置

本文所提出的网络基于PyTorch框架实现，并在单块NVIDIA GeForce RTX 3090显卡进行训练与测试。采用Adam优化器，其参数设置为：初始学习率为 1×10^{-4} ， β_1 和 β_2 分别为0.9和0.999。训练过程中，批处理大小设为16，总训练轮数为200。损失函数为交叉熵CE (Cross-Entropy) 损失与Dice损失之和 (Cardoso等, 2017)，公式如下：

$$LE = \lambda_1 L_{CE} + \lambda_2 L_{Dice} \quad (44)$$

$$L_{CE} = -\frac{1}{N} \sum_{i=1}^N y_i \ln(\hat{y}_i) \quad (45)$$

$$L_{Dice} = 1 - \frac{2 \sum_{i=1}^N y_i \hat{y}_i}{\sum_{i=1}^N y_i + \sum_{i=1}^N \hat{y}_i} \quad (46)$$

式中，LE为本文网络所用的混合损失策略； L_{CE} 为交叉熵损失； L_{Dice} 为Dice损失； λ_1 和 λ_2 为损失函数的权重系数； y_i 表示第*i*个像素的真实标签； $\hat{y}_i$ 表示网络预测的第*i*个像素为变化类的概率； N 为图像总像素数。

为定量评估网络性能，采用了5项指标：总体精度OA (Overall Accuracy)、交并比IoU (Intersection over Union)、精确率Pre (Precision)、召回率Rec (Recall)和F1分数(F1)。这5项指标的定义如下：

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \quad (47)$$

$$IoU = \frac{TP}{TP + FN + FP} \quad (48)$$

$$Pre = \frac{TP}{TP + FP} \quad (49)$$

$$Rec = \frac{TP}{TP + FN} \quad (50)$$

$$F1 = \frac{2 \times Pre \times Rec}{Pre + Rec} \quad (51)$$

式中，TP (True Positive) 为真阳性；表示真实变化区域被正确检测为变化的像素数量；TN (True Negative) 为真阴性，表示真实无变化区域被正确

检测为无变化的像素数量；FP (False Positive) 为假阳性，表示真实无变化区域被误检测为变化的像素数量；FN (False Negative) 为假阴性，表示真实变化区域被漏检（误检测为无变化）的像素数量。上述指标的取值范围均为0—1，数值越高表明方法性能越好。

3.3 对比方法

为全面评估所提出方法的有效性，本文将其与11种基于深度学习的代表性变化检测方法进行比较。这些方法包括：全卷积早期融合网络FC-EF (Fully Convolutional Early Fusion) (Daudt等, 2018)、密集连接孪生变化检测网络SNUNet (densely connected siamese network for change detection of vhr images) (Fang等, 2022)、基于transformer的孪生变化检测网络ChangeFormer (a transformer-based siamese network for change detection) (Bandara等, 2022)、双时相图像transformer网络BIT (Bitemporal Image Transformer) (Chen等, 2022b)、基于轴向交叉注意力的双分支融合网络ACABFNet (bibranch fusion network based on axial cross attention) (Song等, 2023)、双分支多层级跨期变化检测网络DMINet (Dual-branch Multilevel Intertemporal Network) (Wu等, 2023)、时空增强与层间融合网络SEIFNet (Spatiotemporal Enhancement and Interlevel Fusion Network) (Huang等, 2024)、混合变化编码器ChangeBind (a hybrid change encoder for remote sensing change detection) (Noman等, 2024)、时空特征融合与引导聚合网络STFF-GA (Spatio-Temporal Feature Fusion and Guide Aggregation network) (Wei等, 2024a)、基于时空交互注意力的鲁棒变化检测网络CDNeXt (robust change detection for remote sensing images based on temporospatial interactive attention module) (Wei等, 2024b)、基于时空鲁棒表示的高效变化检测网络STRobustNet (efficient change detection via spatial-temporal robust representations) (Zhang, 2025)。

(1) 表1展示了各方法在SYSU-CD数据集上的定量结果。本文所提出的SAMFF在除Pre外的评估指标上均取得最优性能：相较于次优方法，其OA提升0.49%，Rec提升1.49%，F1提升1.38%，IoU提升1.99%。结果表明，SAMFF能更准确地识别变化区域。

表1 SYSU-CD数据集不同方法的定量对比结果

Table 1 Quantitative comparison of different methods on the SYSU-CD Dataset

方法	年份	OA/%	F1/%	IoU/%	Pre/%	Rec/%
FC-EF	2018	89.21	76.79	62.33	77.96	75.66
SNUNet	2022	88.50	76.00	61.29	74.82	77.22
ChangeFormer	2022	90.91	79.92	66.56	83.44	76.69
BIT	2022	89.38	77.49	63.25	77.49	77.49
ACABFNet	2023	91.19	80.82	67.81	83.09	78.67
DMINet	2023	90.41	79.02	65.32	81.62	76.58
SEIFNet	2024	<u>91.57</u>	81.24	68.41	85.46	77.42
ChangeBind	2024	91.41	<u>81.39</u>	<u>68.63</u>	83.25	79.62
STFF-GA	2024	91.42	80.22	66.97	87.88	73.78
CDNeXt	2024	92.13	82.45	70.17	86.89	78.45
STRobustNet	2025	90.70	80.11	66.82	80.80	<u>79.43</u>
本文方法(SAMFF)	—	92.62	83.83	72.16	<u>86.75</u>	81.11

注：最高值用黑体显示，第二高值用斜体显示，第三高值用下划线显示。

如图7所示，在视觉对比实验中，本文所提方法在所有对比方法中表现出最优性能。在S1与S2

样本中，时间不一致性导致屋顶颜色变化，形成显著的伪变化干扰，多数对比方法均出现假阳性检测结果，而SAMFF对颜色与光照变化展现出更强的鲁棒性。在S3与S4样本中，因观测角度差异引起的建筑物立面位移与轮廓畸变未能在配准过程中完全校正，导致几何伪变化产生，所有对比方法均出现明显漏检与误检现象，而SAMFF仍保持最高检测准确率。在S5与S6样本中，时间不一致性与几何失配形成复合干扰，增加了变化检测难度，导致多数方法出现误检与漏检，而SAMFF仍能有效区分真实变化区域与背景干扰。

(2) 表2展示了WHU-CD数据集上的实验结果。SAMFF在除Rec外的评估指标上均取得最优性能：相比次优方法，OA提升0.02%，F1提升0.19%，IoU提升0.33%，Pre提升0.28%。尽管部分指标提升幅度较小，但SAMFF在F1、IoU及Pre等关键指标上均优于现有方法，体现出其整体表现更稳定。

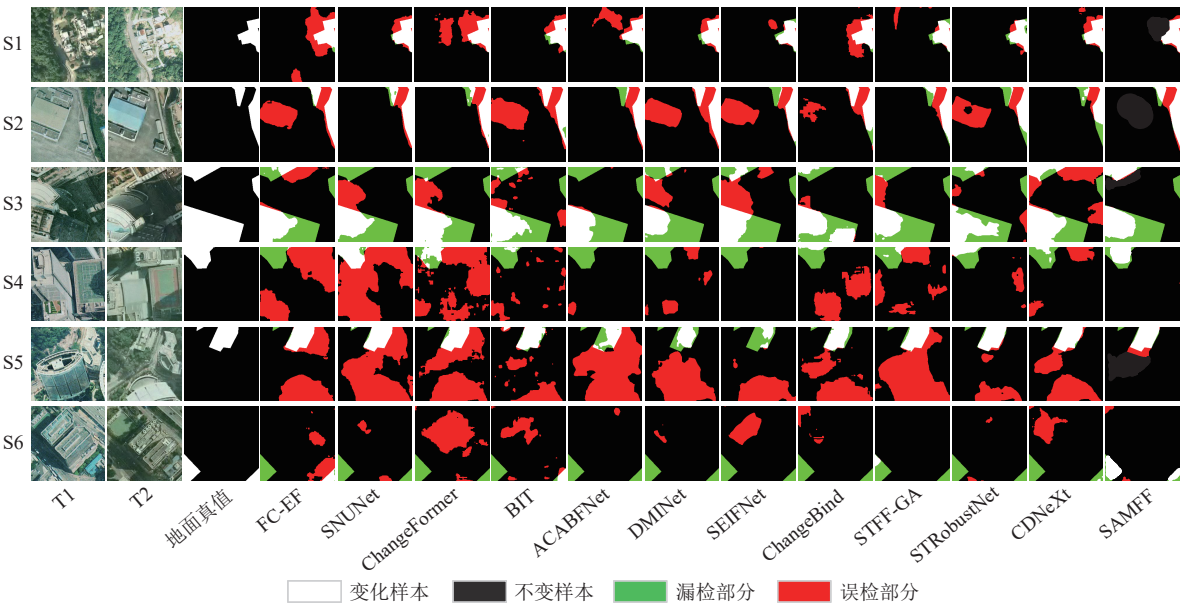


图7 各对比算法在SYSU-CD数据集上的可视化结果

Fig.7 Visualization results of comparative algorithms on the SYSU-CD dataset

如图8所示，在WHU-CD数据集上的定性对比结果表明，本文方法在不同挑战性场景下均表现出优越性能。在建筑密集的S1样本中，多数方法虽能检测整体变化区域，但边界误检明显，而SAMFF在边缘细节保持方面误检较少且边界更贴合标注。对于视角差异导致建筑偏移的S2样本，SAMFF仍能清晰识别真实变化区域，有效缓解几何错位干扰。在阴影与伪变化共存的S3样本中，

所有对比方法均出现不同程度漏检，SAMFF保持最高检测准确率。S4样本中，光照引起的建筑屋顶颜色差异形成强烈伪变化干扰，多数对比方法产生假阳性响应，而SAMFF对此类光谱变化展现出更强鲁棒性。在阴影与遮挡共存的S5样本中，SAMFF检测结果优于所有对比方法，进一步验证其对复杂干扰的抑制能力。受时间不一致性与几何失配复合干扰的S6样本，变化检测难度显著增加，

SAMFF对真实变化与背景干扰的分离更清晰。

(3) 表3展示了GZ-CD数据集上的性能对比，SAMFF在所有评估指标上均取得最优结果：相比次优方法，OA提高0.09%，F1提高0.47%，IoU提高0.78%，Pre提高0.06%，Rec提高0.61%。尽管部分指标的绝对值提升有限，但SAMFF在F1、IoU等关键指标上的一致性优势，仍表明了其有效性。

在该数据集上，本文选取了6组具有代表性的样本进行可视化分析，结果如图9所示。相较于其他对比方法，SAMFF在各样本组中均展现出更优的感知效果。在S1、S3和S5样本中，存在大范围地表植被变化，现有方法普遍出现误检，而本文方法能够更完整地识别真实变化区域。在S2与S4样本中，复杂背景环境增加了检测难度，导致边缘判别模糊，所有对比方法均存在不同程度的漏检与误检现象。可视化结果表明，尽管SAMFF在极少数区域也存在轻微误识别，但其检测结果与真实标注的整体相似度最高。在S6样本中，建筑

物屋顶与道路颜色高度相似，易引发伪变化干扰，本文方法仍能够较好地识别真实变化区域。

表2 WHU-CD数据集不同方法的定量对比结果

Table 2 Quantitative comparison of different methods on the WHU-CD dataset

方法	年份	OA/%	F1/%	IoU/%	Pre/%	Rec/%
FC-EF	2018	99.10	88.64	79.53	89.51	87.71
SNUNet	2022	99.28	90.52	82.68	93.32	87.89
ChangeFormer	2022	99.14	87.18	77.27	92.70	82.28
BIT	2022	99.24	90.40	82.48	91.38	89.44
ACABFNet	2023	99.28	90.57	82.76	94.45	86.99
DMINet	2023	99.19	91.49	84.31	92.65	90.35
SEIFNet	2024	99.04	89.99	81.80	90.11	89.88
ChangeBind	2024	<u>99.42</u>	92.48	86.01	<u>94.76</u>	90.30
STFF-GA	2024	99.41	<u>92.66</u>	<u>86.32</u>	92.03	93.30
CDNeXt	2024	99.47	93.25	87.36	95.20	<u>91.38</u>
STRobustNet	2025	99.39	92.23	85.57	93.56	90.94
本文方法(SAMFF)	—	99.49	93.44	87.69	95.48	<u>91.48</u>

注：最高值用黑体显示，第二高值用斜体显示，第三高值用下划线显示。

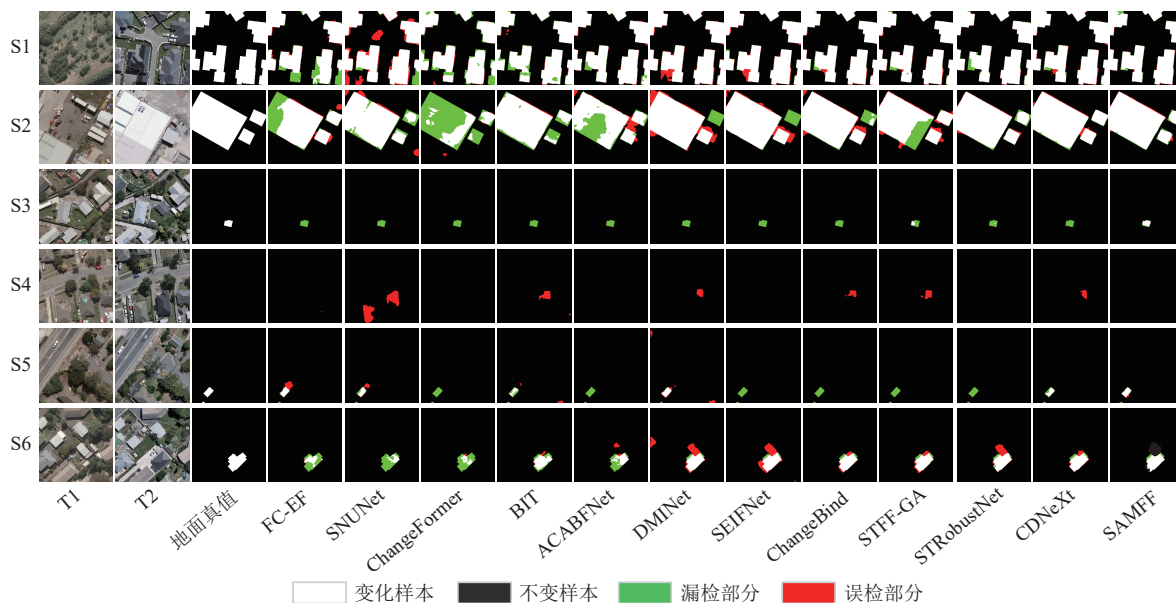


图8 各对比算法在WHU-CD数据集上的可视化结果

Fig.8 Visualization results of comparative algorithms on the WHU-CD dataset

3.4 模型计算效率

模型计算效率是评估其实际应用价值的关键。本文以256像素×256像素图像为输入，通过参数量(Params)与浮点运算次数(FLOPs)两项指标进行综合评价。如图10所示，早期网络(如FC-EF、SNUNet)虽复杂度低，但检测性能有限；后续引入Transformer等模块的网络(如ChangeFormer、

ACABFNet)性能有所提升，却伴随参数与计算量的大幅增长。部分近期网络(如DMINet、STRobustNet)则在效率与性能之间取得了较好的平衡。

相比之下，本文所提网络在计算效率上优势显著：参数量仅为14.28 M，低于ChangeFormer(41.02 M)与STFF-GA(22.37 M)；浮点运算次数低至3.65 G，为所有对比方法中最低。与此同时，

网络在SYSU-CD、WHU-CD和GZ-CD数据集上的F1分数均位列第一，分别达到83.83%、93.44%和88.40%。该结果充分表明，本文网络在维持优异检测精度的同时，实现了更高的计算效率，具备更好的实际部署潜力。

3.5 消融实验

为进一步验证各模块对变化检测性能的影响，本文开展了消融实验。

(1) 为验证本文所提出的SAMFF网络中关键模块的有效性，本文基于SegFormer-B0架构构建基准网络（Baseline），并设计了共计16组消融实验组合，涵盖各模块的单独及联合测试。定量结果（表4）表明，所有引入模块的实验设置均优于基准网络。在SYSU-CD、WHU-CD和GZ-CD 3个数据集上，完整网络的F1分数相较基准网络分别提升3.06%、2.20%和1.39%，IoU分别提升4.42%、3.81%和2.22%。上述结果一致表明，各模块均能

有效提升变化检测性能，验证了本文所提模块设计的有效性。

表3 GZ-CD数据集不同方法的定量对比结果
Table 3 Quantitative comparison of different methods on the GZ-CD dataset

方法	年份	OA/%	F1/%	IoU/%	Pre/%	Rec/%
FC-EF	2018	94.45	65.67	48.89	73.82	59.14
SNUNet	2022	97.07	84.25	72.79	86.82	81.82
ChangeFormer	2022	96.70	80.30	67.09	86.51	74.93
BIT	2022	96.81	81.87	69.31	83.46	80.35
ACABFNet	2023	96.72	83.19	71.21	87.58	79.21
DMINet	2023	96.93	84.25	72.79	88.86	80.10
SEIFNet	2024	96.88	84.33	72.90	86.91	81.89
ChangeBind	2024	97.58	87.93	78.45	89.76	86.17
STFF-GA	2024	97.50	87.53	77.82	89.44	85.69
CDNeXt	2024	97.46	87.03	77.04	90.01	83.24
STRobustNet	2025	97.02	82.88	70.76	84.63	81.20
本文方法(SAMFF)	—	97.67	88.40	79.23	90.07	86.78

注：最高值用黑色显示，第二高值用斜体显示，第三高值用下划线显示。

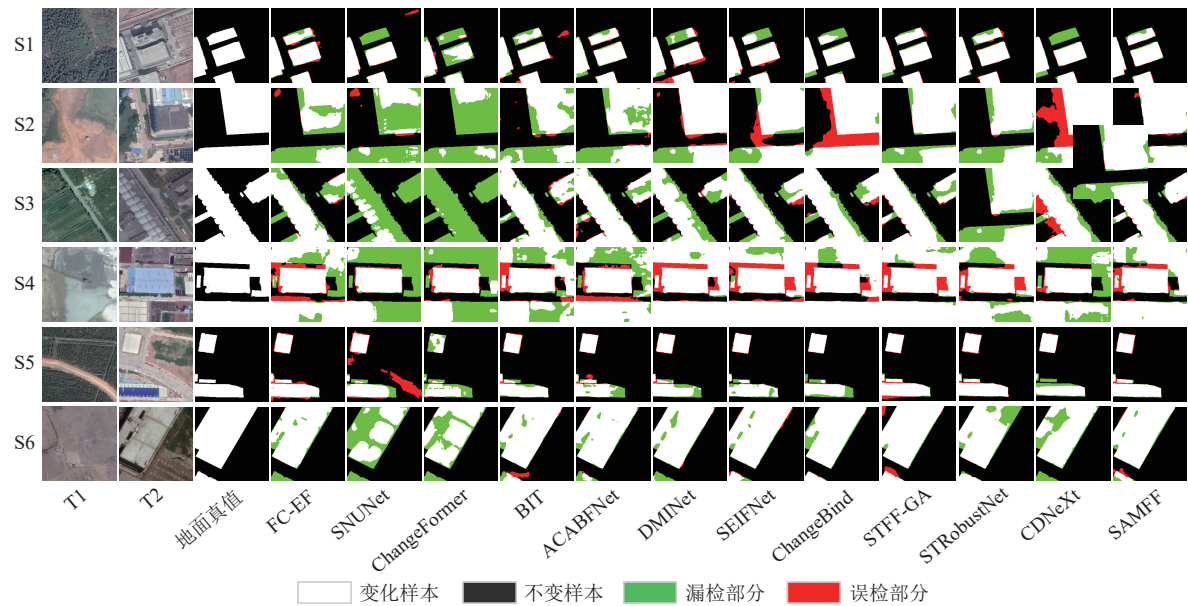


图9 各对比算法在GZ-CD数据集上的可视化结果
Fig.9 Visualization results of comparative algorithms on the GZ-CD dataset

(2) CTAAM模块的消融。为验证CTAAM模块的作用，本文开展了系统消融实验（表5）。实验中，“w/o CTAAM-1”代表不包含CTAAM的基准网络（即SAMFF-11），“S1—S4”对应4个特征提取阶段。结果表明，随着CTAAM从浅层（S1）到深层（S4）逐步引入，网络的F1分数与IoU指标持续提升，证明该模块能通过多尺度通道交互有效抑制光照、季节等因素引起的时序不一致性干扰，增强对真实变化的感知能力。

为进一步分析CTAAM的作用机制，本文基于WHU-CD数据集可视化了各阶段的差分特征图（图11）。在浅层（S1）引入CTAAM能提升空间细节的完整性，使建筑边缘更清晰、背景噪声更少；在深层（S3、S4）引入则强化了网络对语义级时序变化的响应能力，使真实变化区域的激活更为准确。上述结果表明，多阶段嵌入CTAAM可实现层次化的时序干扰抑制与特征对齐，从而提升检测的鲁棒性。

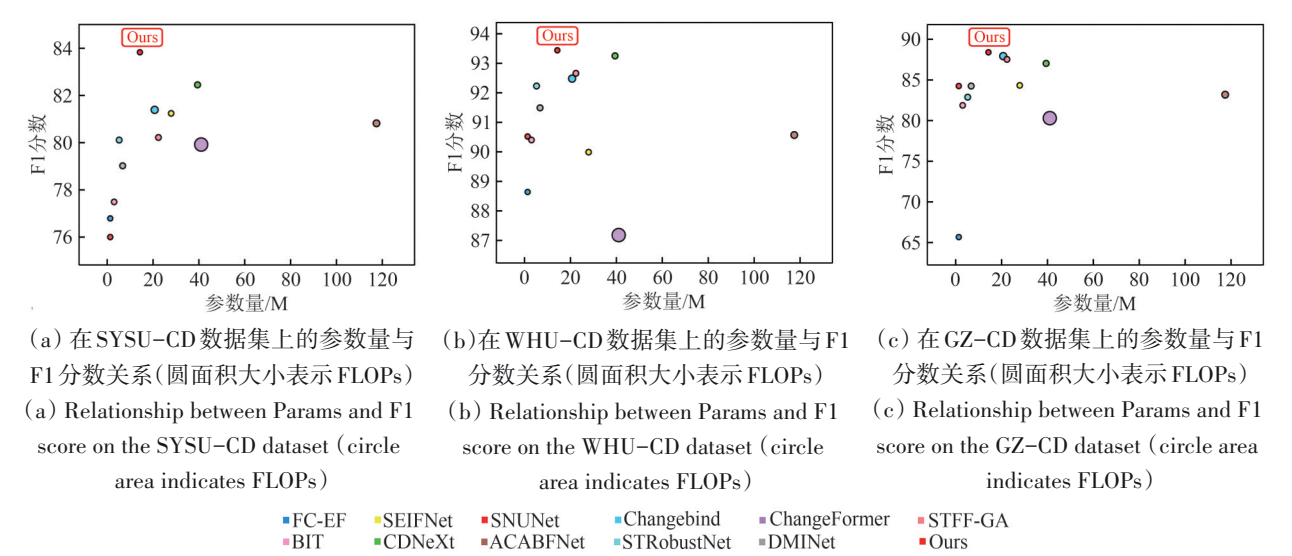


图 10 不同方法在 SYSU-CD、WHU-CD 和 GZ-CD 数据集上的性能、参数量与计算复杂度对比

Fig.10 Comparison of method performance, Params, and FLOPs among different methods on SYSU-CD, WHU-CD, and GZ-CD datasets

表 4 3 个数据集上的定量消融实验结果

方法	CTAAM	FADM	CSAFE	CFGFFM	SYSU-CD		WHU-CD		GZ-CD	
					F1/%	IoU/%	F1/%	IoU/%	F1/%	IoU/%
基准网络	×	×	×	×	80.77	67.74	91.24	83.88	87.01	77.01
SAMFF-1	√	×	×	×	81.51	68.79	91.93	85.07	87.48	77.75
SAMFF-2	×	√	×	×	81.63	68.96	91.85	84.93	87.39	77.60
SAMFF-3	×	×	√	×	81.39	68.62	91.71	84.69	87.18	77.27
SAMFF-4	×	×	×	√	81.42	68.66	91.75	84.76	87.13	77.20
SAMFF-5	√	√	×	×	82.42	70.10	92.43	85.93	87.79	78.24
SAMFF-6	√	×	√	×	81.75	69.13	91.99	85.17	87.57	77.89
SAMFF-7	√	×	×	√	81.79	69.19	91.95	85.11	87.47	77.73
SAMFF-8	×	√	√	×	81.46	68.72	91.92	85.05	87.63	77.98
SAMFF-9	×	√	×	√	81.64	68.98	91.89	85.01	87.51	77.79
SAMFF-10	×	×	√	√	81.58	68.89	91.83	84.89	87.42	77.65
SAMFF-11	×	√	√	√	82.91	70.81	92.32	85.74	87.95	78.49
SAMFF-12	√	×	√	√	83.24	71.29	92.46	85.98	88.19	78.87
SAMFF-13	√	√	×	√	83.56	71.76	92.81	86.58	88.27	79.01
SAMFF-14	√	√	√	×	83.69	71.95	92.93	86.79	88.31	79.07
SAMFF	√	√	√	√	83.83	72.16	93.44	87.69	88.40	79.23

注：“√”代表引入该模块，“×”代表未引入该模块，加粗数字代表最优值。

表 5 CTAAM 模块不同阶段的消融实验结果

方法	S1	S2	S3	S4	SYSU-CD		WHU-CD		GZ-CD	
					F1/%	IoU/%	F1/%	IoU/%	F1/%	IoU/%
w/o CTAAM-1	×	×	×	×	82.91	70.81	92.32	85.74	87.95	78.49
w/o CTAAM-2	√	×	×	×	83.21	71.25	92.84	86.64	88.07	78.68
w/o CTAAM-3	√	√	×	×	83.54	71.73	93.02	86.95	88.19	78.87
w/o CTAAM-4	√	√	√	×	83.69	71.95	93.21	87.28	88.24	78.95
SAMFF	√	√	√	√	83.83	72.16	93.44	87.69	88.40	79.23

注：“√”代表该阶段引入 CTAAM，“×”代表该阶段未引入 CTAAM，加粗数字代表最优值。

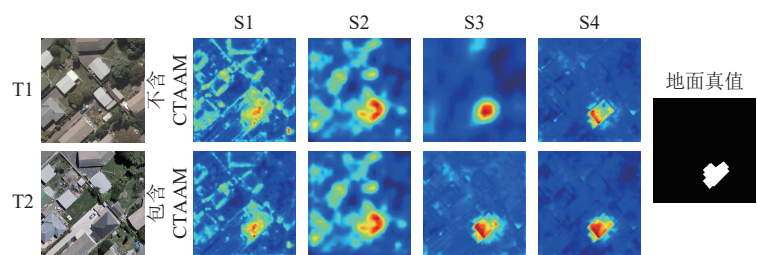


图 11 WHU-CD数据集上差分特征图的可视化结果(红色表示较高的注意力值,蓝色表示较低的注意力值)
Fig.11 Visualization results of differential feature maps on the WHU-CD Dataset (Red color indicates higher attention values , while blue color indicates lower attention values)

此外，通过模块内部消融实验（表6）验证了CTAAM中两个核心组件的有效性。其中，“w/o Model A”为未引入CTA、LCB、CSEB的基线模型。仅使用CTA（w/o Model B）即显著提升性能，表明时域交互的基础作用。在此基础上，分别加入LCB（w/o Model C）或CSEB（w/o Model D）均能

进一步提升：LCB增强边缘细节捕捉能力，CSEB通过抑制光谱差异提升语义鲁棒性。完整的CTAAM模块（SAMFF）在所有测试中均达最优性能，证明注意力对齐与特征增强机制具有协同互补效应，可共同提升网络在复杂场景下的感知精度。

表 6 CTAAM模块内部机制的消融实验结果
Table 6 Ablation study results for internal mechanisms of the CTAAM module

方法	CTA	LCB	CSEB	SYSU-CD		WHU-CD		GZ-CD	
				F1/%	IoU/%	F1/%	IoU/%	F1/%	IoU/%
w/o Model A	×	×	×	82.91	70.81	92.32	85.74	87.95	78.49
w/o Model B	√	×	×	83.11	71.12	92.71	86.41	88.15	78.81
w/o Model C	√	√	×	83.32	71.41	92.96	86.85	88.27	78.99
w/o Model D	√	×	√	83.44	71.57	92.82	86.60	88.33	79.11
SAMFF	√	√	√	83.83	72.16	93.44	87.69	88.40	79.23

注：“√”代表引入该组件，“×”代表未引入该组件，加粗数字代表最优值。

（3）FADM模块的消融。为验证FADM模块在不同特征层级中的作用，本文基于SAMFF网络设置了分阶段嵌入该模块的消融实验，定量结果如表7所示。实验中，“w/o FADM-1”表示未引入FADM的基准网络，“S1—S4”代表4个层级特征提取阶段。实验结果表明，随着FADM模块从浅

层至深层的逐级引入，网络在所有数据集上的F1与IoU指标均呈现稳定提升，最终完整网络（全部4阶段引入FADM）在所有数据集上均达到最优性能。该现象说明，FADM模块能够在不同特征层级中有效建模像素级空间偏移，从而提升双时相特征的对齐质量与判别能力。

表 7 FADM模块不同阶段的消融实验结果
Table 7 Ablation study results for different stages of the FADM module

方法	S1	S2	S3	S4	SYSU-CD		WHU-CD		GZ-CD	
					F1/%	IoU/%	F1/%	IoU/%	F1/%	IoU/%
w/o FADM-1	×	×	×	×	83.24	71.29	92.46	85.98	88.19	78.87
w/o FADM-2	√	×	×	×	83.37	71.48	92.79	86.55	88.24	78.95
w/o FADM-3	√	√	×	×	83.63	71.87	92.97	86.86	88.29	79.04
w/o FADM-4	√	√	√	×	83.71	71.98	93.19	87.25	88.34	79.12
SAMFF	√	√	√	√	83.83	72.16	93.44	87.69	88.40	79.23

注：“√”代表该阶段引入FADM，“×”代表该阶段未引入FADM，加粗数字代表最优值。

为系统验证 FADM 模块对几何配准误差的抑制能力及其在级联框架中的关键作用，本文在严格控制变量的条件下（仅替换对齐模块，骨干网络、训练策略及 SAMFF 其余模块均保持一致），将引入 FADM 的完整网络（SAMFF）及未引入 FADM 的网络（SAMFF-12）与其他 3 种代表性流场对齐模块进行同框架对比：双时相对齐 BTA（Bi-Temporal Alignment）（Meng 等，2024）在特征金字塔集成光流对齐以捕获多尺度运动、流与注意力引导的双时相对齐模块 FA-BAFM（Flow-and Attention-guided Bitemporal Alignment Module）（Zhao 等，2024）引入注意力引导机制优化光流估计，动态可变形双对齐融合 D³AF（Dynamic-Deformable Dual-alignment Fusion）（Li 等，2024）采用显式流场估计与可微分重采样补偿视角差异。定量结果（表 8）显示，SAMFF 在 3 个数据集上均取得最佳性能。尤其在几何敏感的 WHU-CD 数据集上，其 IoU（87.69%）较次优的 D³AF 提升 0.49%；在 SYSU-CD 数据集上，*F1* 分数也领先次优的 D³AF 0.24%。性能优势源于 FADM 独特的“先增强、后对齐”设计：通过 MSCA 模块预先增强特征的几何结构，

抑制时相伪变化，从而生成更精准的差分流场。可视化结果（图 12）进一步印证：在 SYSU-CD 数据集视角倾斜场景中，未引入 FADM 的网络（SAMFF-12）存在边缘重影误检；对比模块虽改善配准，但弱纹理区域细节仍有丢失；而集成 FADM 的 SAMFF 网络通过特征增强与精准光流对齐的级联操作，有效校正了几何误差，检测结果与真实标注一致性最高，验证了 FADM 在提升特征判别力与抑制空间干扰方面的有效性。

表 8 替换不同对齐模块后的网络性能对比

Table 8 Network performance comparison after replacing different alignment modules

方法	SYSU-CD		WHU-CD		GZ-CD	
	<i>F1</i> /%	IoU/%	<i>F1</i> /%	IoU/%	<i>F1</i> /%	IoU/%
SAMFF-12	83.24	71.29	92.46	85.98	88.19	78.87
BTA	83.30	71.38	92.89	86.72	88.21	78.91
FA-BAFM	83.48	71.64	93.07	87.04	88.26	78.99
D ³ AF	83.59	71.81	93.16	87.20	88.32	79.08
SAMFF	83.83	72.16	93.44	87.69	88.40	79.23

注：加粗数字代表最优值；SAMFF-12 为未引入 FADM 的基准网络；BTA、FA-BAFM、D³AF 为替换为对应对齐模块的变体；SAMFF 为引入 FADM 的完整网络。

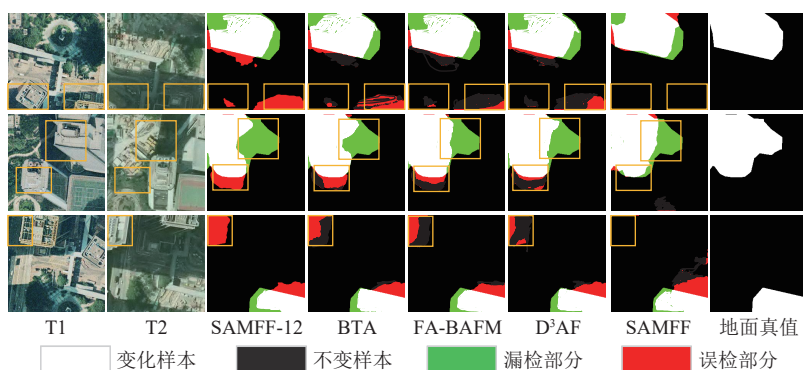


图 12 FADM 有效性的可视化对比

Fig.12 Visual comparison of FADM effectiveness

（4）CSAFE 模块消融。该模块旨在对经前序模块时空对齐后的差分特征进行深度语义增强。通过多尺度上下文提取与跨尺度注意力融合机制，该模块能够整合变化区域的语义与结构信息，从而增强网络对复杂变化模式的辨识能力，并保持空间连续性。消融实验（表 4）表明，引入 CSAFE 模块后，网络在 SYSU-CD、WHU-CD 与 GZ-CD 数据集上性能均显著提升：相较于未引入该模块的 SAMFF-13，完整 SAMFF 的 *F1* 分别提升 0.27%、0.63% 和 0.13%，IoU 分别提升 0.40%、1.11% 和

0.22%。

为验证其有效性，本文在各数据集上进行了定性比较（图 13）。集成 CSAFE 的网络（SAMFF）在变化区域识别精度上明显提升，尤其在边缘细节完整性方面表现突出。以 WHU-CD 为例，未集成 CSAFE 的网络（SAMFF-13）在建筑物边缘漏检明显，表明基础差分特征对细长目标结构表达不足；SAMFF 网络借助 CSAFE 的多尺度上下文感知与注意力机制，能够更完整地捕捉全局结构特征，从而提升线性地物变化识别的完整度。

(5) CFGFFM 模块消融。CFGFFM 作为解码器，负责将增强后的多尺度特征对齐融合为最终变化图。该模块通过流引导的置信度融合机制，自适应地对齐并加权融合深层和浅层特征，有效整合多尺度时空信息，最终解码出具有空间一致性与细节完整性的像素级变化图。如表4中SAMFF-14与SAMFF的对比所示，采用CFGFFM替换基准解码器后，网络各项指标均提升，验证了其有效性。

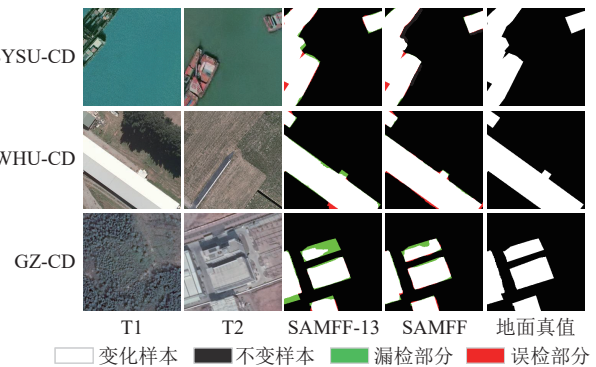


图 13 CSAFE 有效性的可视化对比
Fig.13 Visual comparison of CSAFE effectiveness

在特征融合中，CFGFFM 充分利用不同分辨率特征的互补性：高分辨率特征擅长捕捉小目标及局部细微变化，低分辨率特征利于识别大范围环境变化。通过跨层级对齐融合，该模块能够综合这些互补信息，构建更具判别性的变化特征表达。可视化结果（图 14）进一步表明，CFGFFM 在跨层级融合中增强了对不同规模变化区域的感知与重建能力，提升了网络对复杂场景的适应性与检测精度。

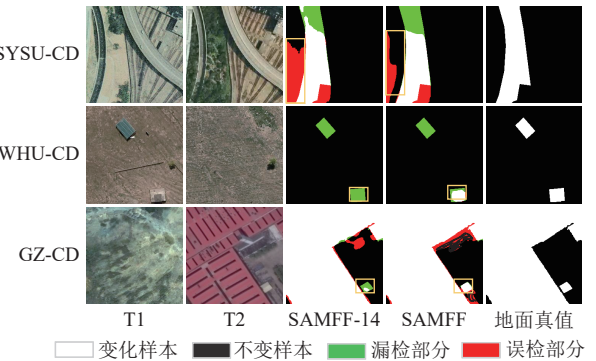


图 14 CFGFFM 有效性的可视化对比
Fig.14 Visual comparison of CFGFFM effectiveness

3.6 模型泛化能力分析

鉴于真实遥感应应用场景的复杂性与不可预测性，为评估网络对未知数据分布的适应能力及其实际部署潜力，本文设计了跨数据集零样本推理实验。该实验模拟了网络在训练数据与测试数据存在显著差异时的场景：网络仅在SYSU-CD数据集上训练，随后不经任何微调，直接在WHU-CD数据集上进行预测。这一设定引入了多维度差异，包括影像分辨率（0.5 m与0.2 m）、辐射特征（不同传感器与时相采集条件）以及变化目标类别（多类别综合变化与单一建筑变化）。

如表9所示，所有对比方法在该设定下性能均出现下降，但本文提出的SAMFF网络展现出较强的跨场景泛化能力，其F1分数保持在72.81%（表9），优于其他方法。可视化结果（图 15）进一步表明，SAMFF在WHU-CD数据集上能更完整地检出建筑变化，并有效抑制由分辨率与辐射差异引起的伪变化。这得益于其级联对齐架构的针对性设计：该架构通过分阶段处理，先抑制由传感器与时相差引起的光谱差异，再校正因分辨率与视角变化导致的几何偏移，从而使网络能够学习到更具适应性的变化特征，在数据分布差异下仍能保持稳定性能。

表 9 各对比算法在 WHU-CD 数据集上的
跨数据集推理性能对比

Table 9 Comparison of cross-dataset inference performance of comparative algorithms on WHU-CD Dataset						
方法	年份	OA/%	F1/%	IoU/%	Pre/%	Rec/%
SEIFNet	2024	95.41	69.51	53.26	74.86	64.93
ChangeBind	2024	96.39	70.36	54.27	76.40	65.19
STFF-GA	2024	96.19	70.72	54.70	75.91	66.20
CDNeXt	2024	96.69	71.36	55.48	76.35	66.97
STRobustNet	2025	95.91	70.14	54.01	75.32	65.64
本文方法(SAMFF)	—	97.96	72.81	57.23	77.55	68.59

注：加粗为本文方法。

3.7 损失函数系数分析

为验证不同损失函数系数对网络性能的影响，本文在SYSU-CD、WHU-CD和GZ-CD数据集上开展了相应实验，实验结果如表10所示。对于SYSU-CD和GZ-CD数据集，当交叉熵损失系数（ λ_1 ）与

Dice 损失系数 (λ_2) 相对平衡时, 网络取得最优性能; 对于 WHU-CD 数据集, 由于变化区域边界较规整, 当 λ_1 与 λ_2 取相对平衡时, 网络在 F1 与 IoU 上同

时达到最优, 而不均衡权重配置会导致性能下降。因此, 将系数 λ_1 和 λ_2 均设为 0.5 的混合损失函数, 在 3 个数据集上均展现出较好的稳定性与适用性。

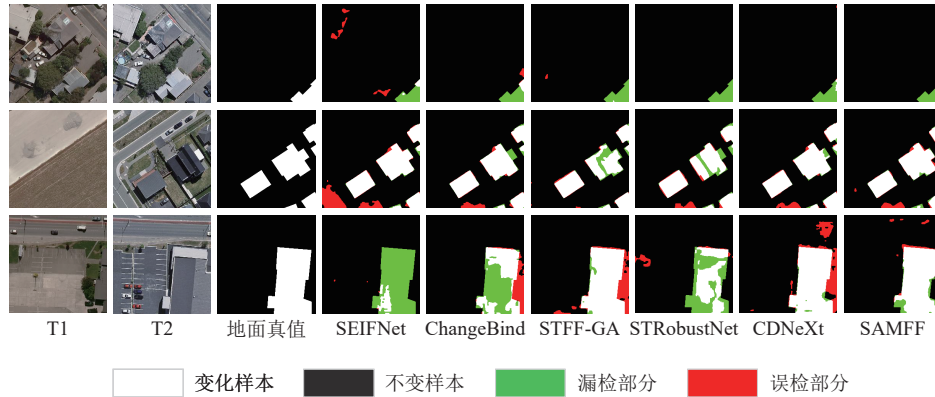


图 15 各对比算法在 WHU-CD 数据集上的跨数据集推理可视化结果

Fig.15 Cross-dataset inference visualization results of compared methods on the WHU-CD Dataset

表 10 不同损失函数系数设置下的网络性能评估结果

Table 10 Performance evaluation of the proposed network under different loss function coefficients

λ_1	λ_2	SYSU-CD		WHU-CD		GZ-CD	
		F1/%	IoU/%	F1/%	IoU/%	F1/%	IoU/%
1	0	82.10	69.64	92.15	85.45	87.23	77.36
0	1	81.52	68.80	91.85	84.92	86.95	76.92
0.5	0.5	83.83	72.16	93.44	87.69	88.40	79.23
0.5	1	83.25	71.32	93.08	87.05	88.02	78.61
1	0.5	83.18	71.21	93.12	87.13	87.95	78.49

注: 加粗数字代表最优值。

4 结 论

本文针对高分辨率遥感影像变化检测中由时间不一致性与几何配准误差形成的复合干扰问题, 提出 SAMFF 方法。该方法通过 CTAAM 在多层级特征上进行时域对齐, 抑制时序干扰; 再借助 FADM 实现高精度空间配准, 补偿像素级位移。该设计有效缓解了传统方法中误差的交叉传播与累积。在此基础上, 通过 CSAFE 强化差分特征, 并利用 CFGFFM 模块自适应融合多级特征, 最终输出空间一致且细节完整的变化检测图。在 SYSU-CD、WHU-CD 和 GZ-CD 数据集上, SAMFF 在 F1 与 IoU 指标上均达到先进水平, 且在参数量与计算复杂度上具有显著优势, 实现了检测精度与效率的平衡。进一步分析表明, 网络在强视差遮挡区域及异物同谱等复杂场景下存在少量误检, 同时受限于深层特征下采样与光流平滑约束, 对极微小目

标的检测存在漏检风险。本文提出的级联对齐策略主要面向时间表观差异与几何失配并存, 且易在算法链路中产生误差叠加的典型双时相遥感场景。针对辐射差异较显著的情形, 光流估计所依赖的跨时相特征相似性假设易受破坏, 此时优先提升跨时相一致性可降低表观扰动对后续位移估计的干扰, 从而增强几何校正的稳定性。对于辐射差异较弱但视角几何畸变更突出的情形 (如短时相内多视角成像), 则可将空间对齐前置作为补充策略, 以优先建立像素对应关系。

后续研究将从以下 3 个方面展开: (1) 聚焦于数据驱动的自适应对齐, 通过量化评估输入影像辐射差异与几何失配的相对主导性, 动态优化时域与空域校正的级联顺序, 以增强网络在复杂成像条件下的适用性; (2) 引入遮挡感知与时空一致性约束, 融合语义先验与细粒度保持机制, 以提升网络对复杂场景的判别能力与几何鲁棒性; (3) 探索半监督及无监督学习范式, 以提高网络在开放场景中的泛化能力与实际应用价值。

参考文献 (References)

- Bandara W G C and Patel V M. 2022. A transformer-based Siamese network for change detection//IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium. Kuala Lumpur, Malaysia: IEEE: 207-210 [DOI: 10.1109/IGARSS46834.2022.9883686]
- Cardoso M J, Arbel T, Carneiro G, Syeda-Mahmood T, Tavares J M R

- S, Moradi M, Bradley A, Greenspan H, Papa J P, Madabhushi A, Nascimento J C, Cardoso J S, Belagiannis V and Lu Z. 2017. Deep learning in medical image analysis and multimodal learning for clinical decision support//Third International Workshop, DL-MIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017. CityQuébec, QC, Canada: Springer International Publishing [DOI: 10.1007/978-3-319-67558-9]
- Celik T. 2009. Unsupervised change detection in satellite images using principal component analysis and k-Means clustering. *IEEE Geoscience and Remote Sensing Letters*, 6(4): 772-776 [DOI: 10.1109/LGRS.2009.2025059]
- Chen H, Li W Y, Chen S and Shi Z W. 2022a. Semantic-aware dense representation learning for remote sensing image change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 5630018 [DOI: 10.1109/TGRS.2022.3203769]
- Chen H, Qi Z P and Shi Z W. 2022b. Remote sensing image change detection with Transformers. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 5607514 [DOI: 10.1109/TGRS.2021.3095166]
- Chen Z L and Lu S N. 2025. CAF-YOLO: a robust framework for multi-scale lesion detection in biomedical imagery//ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Hyderabad, India: IEEE: 1-5 [DOI: 10.1109/ICASSP49660.2025.10888358]
- Cheng G, Wang G X and Han J W. 2025. Deep learning for change detection in remote sensing: a review and new outlooks. *National Remote Sensing Bulletin*, 29(6): 1587-1597 (程焜, 王光兴, 韩军伟. 2025. 深度学习遥感变化检测综述: 典型算法及发展趋势. *遥感学报*, 29(6): 1587-1597) [DOI: 10.11834/jrs.20254441]
- Cheng M F, He W, Li Z H, Yang G Y and Zhang H Y. 2024. Harmony in diversity: content cleansing change detection framework for very-high-resolution remote-sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 218: 1-19 [DOI: 10.1016/j.isprsjprs.2024.09.002]
- Daudt R C, Le Saux B and Boulch A. 2018. Fully convolutional Siamese networks for change detection//25th IEEE International Conference on Image Processing. Athens: IEEE: 4063-4067 [DOI: 10.1109/ICIP.2018.8451652]
- Dosovitskiy A, Fischer P, Ilg E, Häusser P, Hazirbas C, Golkov V, van der Smagt P, Cremers D and Brox T. 2015. FlowNet: learning optical flow with convolutional networks//2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE: 2758-2766 [DOI: 10.1109/ICCV.2015.316]
- Fang S, Li K Y and Li Z. 2023. Changer: feature interaction is what you need for change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 5610111 [DOI: 10.1109/TGRS.2023.5610111].
- Fang S, Li K Y, Shao J Y and Li Z. 2022. SNUNet-CD: a densely connected Siamese network for change detection of VHR images. *IEEE Geoscience and Remote Sensing Letters*, 19: 8007805 [DOI: 10.1109/LGRS.2021.3056416]
- Fu K, Lu W X, Liu X Y, Deng C B, Yu H F and Sun X. 2024. A comprehensive survey and assumption of remote sensing foundation modal. *National Remote Sensing Bulletin*, 28(7): 1667-1680 (付琨, 卢宛萱, 刘小煜, 邓楚博, 于泓峰, 孙显. 2024. 遥感基础模型发展综述与未来设想. *遥感学报*, 28(7): 1667-1680) [DOI: 10.11834/jrs.20233313]
- Han T, Wulder M A, White J C, Coops N C, Alvarez M F and Butson C. 2007. An efficient protocol to process Landsat images for change detection with tasselled cap transformation. *IEEE Geoscience and Remote Sensing Letters*, 4(1): 147-151 [DOI: 10.1109/LGRS.2006.887066]
- Holail S, Saleh T, Xiao X W, Zahran M, Xia G S and Li D R. 2025. Edge-CVT: edge-informed CNN and vision transformer for building change detection in satellite imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 227: 48-68 [DOI: 10.1016/j.isprsjprs.2025.05.021]
- Huang Y Y, Li X H, Du Z S and Shen H F. 2024. Spatiotemporal enhancement and interlevel fusion network for remote sensing images change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 5609414 [DOI: 10.1109/TGRS.2024.3360516]
- Ji S P, Wei S Q and Lu M. 2019. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. *IEEE Transactions on Geoscience and Remote Sensing*, 57(1): 574-586 [DOI: 10.1109/TGRS.2018.2858817]
- Jiang B, Wang Z T, Wang X X, Zhang Z Y, Chen L, Wang X and Luo B. 2023. VcT: visual change transformer for remote sensing image change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 2005214 [DOI: 10.1109/TGRS.2023.3327139]
- Li L Z, Han L, Ding M T, Liu Z H and Cao H Y. 2022. Remote sensing image registration based on deep learning regression model. *IEEE Geoscience and Remote Sensing Letters*, 19: 8002905 [DOI: 10.1109/LGRS.2020.3032439]
- Li Y C, Lei S, Liu N Q, Li H C and Du Q. 2024. IDA-SiamNet: interactive- and dynamic-aware Siamese network for building change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 5628213 [DOI: 10.1109/TGRS.2024.3410977]
- Liu Y T, Li S H, He Z and Liu K T. 2025. Edge and flow guided iterative CNN for remote sensing image change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 63: 4406422 [DOI: 10.1109/TGRS.2025.3550973]
- Lowe D G. 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2): 91-110 [DOI: 10.1023/B:VISI.0000029664.99615.94]
- Meng JT, Xu X Y, Li P Y and Zhang Z. 2024. ChangeAD: enhanced remote sensing change detection via Bi-temporal alignment and differential feature integration//2024 IEEE International Geoscience and Remote Sensing Symposium. Qingdao, China: IEEE: 1-4 [DOI: 10.1109/CAC63892.2024.10865446]
- Negri R G, Frery A C, Casaca W, Azevedo S, Dias M A, Silva E A and Alcântara E H. 2021. Spectral-spatial-aware unsupervised change detection with stochastic distances and support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 59(4):

- 2863-2876 [DOI: 10.1109/TGRS.2020.3009483]
- Noman M, Fiaz M and Cholakkal H. 2024. ChangeBind: a hybrid change encoder for remote sensing change detection//IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium. Athens, Greece: IEEE: 8417-8422 [DOI: 10.1109/IGARSS53475.2024.10640559]
- Peng D F, Zhang Y J and Guan H Y. 2019. End-to-end change detection for high resolution satellite images using improved UNet++. *Remote Sensing*, 11(11): 1382 [DOI: 10.3390/rs11111382]
- Pluim J P W, Maintz J B A and Viergever M A. 2003. Mutual-information-based registration of medical images: a survey. *IEEE Transactions on Medical Imaging*, 22(8): 986-1004 [DOI: 10.1109/TMI.2003.815867]
- Rocco I, Arandjelovic R and Sivic J. 2018. End-to-end weakly-supervised semantic alignment//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, UT, USA: IEEE: 6917-6925 [DOI: 10.1109/CVPR.2018.00723]
- Saha S, Bovolo F and Bruzzone L. 2019. Unsupervised deep change vector analysis for multiple-change detection in VHR images. *IEEE Transactions on Geoscience and Remote Sensing*, 57(6): 3677-3693 [DOI: 10.1109/TGRS.2018.2886643]
- Shi Q, Liu M X, Li S C, Liu X P, Wang F and Zhang L P. 2022. A deeply supervised attention metric-based network and an open aerial image dataset for remote sensing change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 5604816 [DOI: 10.1109/TGRS.2021.3085870]
- Song L, Xia M, Weng L G, Lin H F, Qian M and Chen B Y. 2023. Axial cross attention meets CNN: bibranch fusion network for change detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16: 21-32 [DOI: 10.1109/JSTARS.2022.3224081]
- Sun D Q, Yang X D, Liu M Y and Kautz J. 2018. PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, UT, USA: IEEE: 8934-8943 [DOI: 10.1109/CVPR.2018.00931]
- Sun Y L, Lei L, Guan D D and Kuang G Y. 2021. Iterative robust graph for unsupervised change detection of heterogeneous remote sensing images. *IEEE Transactions on Image Processing*, 30: 6277-6291 [DOI: 10.1109/TIP.2021.3093766]
- Tian J, Peng D F, Guan H Y and Ding H Y. 2022. RACDNet: resolution- and alignment-aware change detection network for optical remote sensing imagery. *Remote Sensing*, 14(18): 4527 [DOI: 10.3390/rs14184527]
- Wei H G, Wang N, Liu Y, Ma P G, Pang D D, Sui X B and Chen Q. 2024a. Spatio-temporal feature fusion and guide aggregation network for remote sensing change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 5642216 [DOI: 10.1109/TGRS.2024.3470314]
- Wei J J, Sun K M, Li W Z, Li W B, Gao S, Miao S X, Zhou Q H and Liu J Y. 2024b. Robust change detection for remote sensing images based on temporospatial interactive attention module. *International Journal of Applied Earth Observation and Geoinformation*, 128: 103767 [DOI: 10.1016/j.jag.2024.103767]
- Wu X, Chen Y, Yan Q Y, Zhao Y H, Zhao J L, Zhang X Q, Higashita R and Liu J. 2023. DMINet: a lightweight dual-mixed channel-independent network for cataract recognition//2023 International Joint Conference on Neural Networks (IJCNN). Gold Coast, Australia: IEEE: 1-8 [DOI: 10.1109/IJCNN54540.2023.10191292]
- Yan T Y, Wan Z F, Zhang P P, Cheng G and Lu H C. 2023. TransY-Net: learning fully transformer networks for change detection of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 4410012 [DOI: 10.1109/TGRS.2023.3327253]
- Ye Y X, Wang M M, Zhou L, Lei G Y, Fan J W and Qin Y. 2023. Adjacent-level feature cross-fusion with 3-D CNN for remote sensing image change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 5618214 [DOI: 10.1109/TGRS.2023.3305499]
- Zhang H, Teng Y H, Li H J and Wang Z H. 2025. STRobustNet: efficient change detection via spatial-temporal robust representations in remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 63: 5612215 [DOI: 10.1109/TGRS.2025.3540794]
- Zhang H T, Chen H, Zhou C Y, Chen K Y, Liu C Y, Zou Z X and Shi Z W. 2024. BiFA: remote sensing image change detection with bi-temporal feature alignment. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 5614317 [DOI: 10.1109/TGRS.2024.3376673]
- Zhao F, Zhu Z L, Liu H Q and Zhang J J. 2024. Multiscale alignment and progressive feature fusion network for high resolution remote sensing images change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 6700113 [DOI: 10.1109/TGRS.2024.3433373]
- Zhao S J, Zhang X L, Xiao P F and He G J. 2023. Exchanging dual-encoder-decoder: a new strategy for change detection with semantic guidance and spatial localization. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 4508016 [DOI: 10.1109/TGRS.2023.3327780]
- Zhu Q Q, Guo X, Deng W H, Guan Q F, Zhong Y F, Zhang L P and Li D R. 2022. Land-use/land-cover change detection based on a Siamese global learning framework for high spatial resolution remote sensing imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 184: 63-78 [DOI: 10.1016/j.isprsjprs.2021.12.005]

A spatiotemporal-aware and cascaded alignment framework for remote sensing image change detection

YANG Jingyu^{1,2}, ZHANG Kaiwei¹, DANG Jianwu^{1,2,3}, WANG Feng⁴, HUO Jiuyuan¹

1. School of Electronics and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China;

2. Gansu Provincial Engineering Research Center for Artificial Intelligence and Graphics & Image Processing, Lanzhou 730070, China;

3. National Virtual Simulation Experimental Teaching Center for Rail Transit Information and Control, Lanzhou Jiaotong University, Lanzhou 730070, China;

4. Gansu Luqiao Feiyu Transportation Facilities Co., Ltd, Lanzhou 730070, China

Abstract: High-resolution remote sensing image Change Detection (CD) identifies land-cover changes from bi-temporal images of the same geographic area and is essential for urban monitoring, disaster assessment, ecological surveillance, and land resource management. However, practical CD remains challenging because real bi-temporal images are often affected simultaneously by temporal appearance discrepancies and spatial geometric misalignment. Illumination changes, seasonal variations, shadows, and radiometric inconsistencies may produce spectral pseudo-changes, while viewpoint differences and imperfect registration may introduce building displacement, boundary distortion, and local spatial offsets. Although these two types of interference are physically different, they become coupled during feature extraction, differencing, alignment, and fusion, causing error propagation and making it difficult to separate true semantic changes from irrelevant variations. To address this problem, this study proposes a Spatiotemporal-Aware Multi-scale Feature Flow (SAMFF) network for robust remote sensing image CD under compound spatiotemporal interference.

The proposed SAMFF follows a cascaded alignment strategy that decomposes compound interference suppression into four stages: temporally consistent feature representation, spatial geometric correction, multi-scale enhancement, and confidence-guided fusion. First, a cross-temporal attentional alignment module (CTAAM) is embedded in the feature extraction stage to enable deep interaction between bi-temporal features. By combining cross-temporal attention with local context-aware and channel semantic enhancement branches, CTAAM aligns channel semantic and local contextual representations, thereby reducing pseudo-change responses caused by illumination and seasonal variations. Second, a flow-aligned difference module (FADM) is introduced to handle geometric mismatch. It first enhances temporally aligned features through multi-scale channel attention and then estimates a dense differential flow field to warp and align bi-temporal features at multiple scales. This design improves pixel-level correspondence and suppresses false responses caused by registration errors. Third, a Cross-Scale Attention Feature Enhancement (CSAFE) module aggregates multi-scale contextual information through a dilated convolution pyramid and top-down attention-based feature interaction, strengthening the discriminative representation of real change regions. Finally, a Confidence Flow Guided Feature Fusion Module (CFGFFM) acts as the decoder. It estimates confidence-aware flow fields and adaptively fuses high-level semantic features with low-level spatial details, reducing residual noise in low-confidence regions and producing a spatially consistent change map.

Experiments are conducted on three public remote sensing CD datasets, namely, SYSU-CD, WHU-CD, and GZ-CD. SAMFF achieves $F1/IoU$ of 83.83%/72.16%, 93.44%/87.69%, and 88.40%/79.23%, respectively, on the three datasets, outperforming representative CNN-, Transformer-, and alignment-based comparison methods in overall CD accuracy. Qualitative results further show that SAMFF can better suppress pseudo-changes caused by roof color differences, shadows, vegetation appearance variations, building parallax, and local misregistration. Ablation studies verify the effectiveness of CTAAM, FADM, CSAFE, and CFGFFM, demonstrating that the proposed temporal-to-spatial cascaded alignment design consistently improves performance. In addition, SAMFF achieves a good balance between accuracy and efficiency, with 14.28 M parameters and 3.65 G FLOPs for 256×256 image inputs.

SAMFF provides an effective cascaded framework for high-resolution remote sensing image CD by explicitly addressing the coupled influence of temporal appearance discrepancy and spatial geometric misalignment. Instead of directly learning changes from mixed noisy features, the proposed method progressively suppresses temporal pseudo-changes, corrects geometric offsets, enhances multi-scale change representations, and fuses features with confidence guidance. Experimental results confirm its robustness, accuracy, and deployment potential in complex bi-temporal remote sensing scenarios.

Key words: change detection, remote sensing, spatiotemporal awareness, feature alignment, multiscale fusion, deep learning

Supported by National Natural Science Foundation of China (No. 62367005); Natural Science Foundation of Gansu Province, China (No. 24JRRA253); Central Government Guides Local Science and Technology Development Foundation (No. 24ZYQA051)