

MS-SwinCE: 融合多尺度 Swin Transformer 与对比特征增强的遥感图像变化检测^①



朱永进, 李 熹, 李永胜, 苏湘粤

(广西民族大学 人工智能学院, 南宁 530006)

通信作者: 李 熹, E-mail: 40708970@qq.com

摘 要: 遥感图像变化检测在城市扩张与灾害监测等领域中具有重要意义, 然而, 现有方法在特征提取能力、抗伪变化干扰以及多尺度融合方面仍存在不足之处. 本文提出一种融合多尺度 Swin Transformer、对比特征增强模块 (contrastive feature enhancement module, CFEM) 与双向特征金字塔网络 (bi-directional feature pyramid network, BiFPN) 的变化检测模型 MS-SwinCE. 该模型利用局部窗口和移位机制增强长程依赖建模能力, CFEM 精准提取变化差异并抑制噪声, BiFPN 实现多尺度语义信息的高效融合. 实验结果表明, MS-SwinCE 在 LEVIR-CD 数据集上, 相较于 ChangeFormer, *IoU* 提升了 1.18%, *F1* 分数提升了 0.70%, *Precision* 提升了 0.32%, *Recall* 提升了 1.06%; 在 WHU-CD 数据集上, 相较于 BIT, *IoU* 提升了 1.84%, *F1* 分数提升了 1.06%, *Precision* 提升了 0.47%, *Recall* 提升了 1.66%. 此外, 在保持较高精度的同时, 模型参数量为 31.66M, 明显低于精度相近的 ChangeFormer (41.03M), 在精度与效率间实现了良好权衡. 消融实验进一步验证了各模块的有效性与协同增益.

关键词: 遥感图像变化检测; Swin Transformer; 对比特征增强模块; 多尺度特征融合; 双向特征金字塔网络

引用格式: 朱永进, 李熹, 李永胜, 苏湘粤. MS-SwinCE: 融合多尺度 Swin Transformer 与对比特征增强的遥感图像变化检测. 计算机系统应用, 2025, 34(12):67-74. <http://www.c-s-a.org.cn/1003-3254/10022.html>

MS-SwinCE: Remote Sensing Image Change Detection Integrating Multi-scale Swin Transformer and Contrastive Feature Enhancement

ZHU Yong-Jin, LI Xi, LI Yong-Sheng, SU Xiang-Yue

(School of Artificial Intelligence, Guangxi Minzu University, Nanning 530006, China)

Abstract: Remote sensing image change detection plays a vital role in urban expansion and disaster monitoring. However, existing methods still exhibit limitations in feature extraction capability, resistance to pseudo-change interference, and multi-scale feature fusion. To this end, this study proposes MS-SwinCE, a change detection model that integrates multi-scale Swin Transformer, a contrastive feature enhancement module (CFEM), and a bi-directional feature pyramid network (BiFPN). The model enhances long-range dependency modeling via the local window and shifted mechanism, with CFEM accurately capturing change differences and suppressing noise, and BiFPN efficiently fusing multi-scale semantic information. Experimental results demonstrate that MS-SwinCE outperforms ChangeFormer on the LEVIR-CD dataset, with an improvement of 1.18% in *IoU*, 0.70% in *F1* score, 0.32% in *Precision*, and 1.06% in *Recall*. On the WHU-CD dataset, MS-SwinCE achieves an increase of 1.84% in *IoU*, 1.06% in *F1* score, 0.47% in *Precision*, and 1.66% in *Recall* compared to BIT. Additionally, while maintaining high accuracy, MS-SwinCE has a parameter count of 31.66M, notably lower than ChangeFormer (41.03M) with similar accuracy. Thus, an effective balance between accuracy and efficiency is achieved. Ablation studies further confirm the effectiveness and synergistic gain of each module.

^① 收稿时间: 2025-05-13; 修改时间: 2025-06-12; 采用时间: 2025-06-30; csa 在线出版时间: 2025-10-21

CNKI 网络首发时间: 2025-10-22

Key words: remote sensing image change detection; Swin Transformer; contrastive feature enhancement module (CFEM); multi-scale feature fusion; bi-directional feature pyramid network (BiFPN)

图像变化检测 (change detection, CD) 通过对比不同同时相遥感图像以识别地表变化, 广泛应用于灾害监测、城市扩张等领域^[1,2]。尽管深度学习方法, 特别是卷积神经网络 (convolutional neural network, CNN), 在该任务中取得显著进展, 但其局限于局部感受野, 难以捕捉长程依赖和全局语义, 易受伪变化和噪声干扰^[3,4]。

1 图像变化检测

Li 等^[1]通过多尺度卷积结构增强对细节变化的感知, Yin 等^[2]进一步引入空间和通道注意力机制动态加权特征, 提升检测精度, 但其对复杂背景仍较敏感。

Transformer 架构以其全局建模能力为变化检测提供了新路径。ChangeFormer^[5]基于层次化 Transformer 编码器与轻量级多层感知机 (multi-layer perceptron, MLP) 解码器的 Siamese 架构, 有效提取了多尺度变化特征。Noman 等^[6]提出从零训练的 Transformer 模型, 打破对预训练权重的依赖, 在遥感变化检测中验证了自监督初始化的有效性。Swin Transformer^[7]通过局部窗口自注意力与移位机制, 在保持计算效率的同时增强了上下文建模能力。Zhang 等^[8]提出多阶段 Swin Transformer 检测网络, 实现细节与全局信息融合, 但当前模型在特征融合与噪声抑制方面仍有不足。Lin 等^[9]提出双向 FPN 结构以融合局部与全局特征, 改善语义表达能力。此外, 自监督学习与对比学习策略逐渐应用于缓解标注不足问题。Chen 等^[4]提出对比损失以约束前后时相特征差异, 有效提升模型对变化区域的判别能力^[4,10]。

本文提出一种基于多尺度 Swin Transformer 与对比特征增强模块 (CFEM) 的变化检测模型 MS-SwinCE, 结合局部窗口自注意力机制与多尺度特征金字塔网络 (BiFPN), 实现高效特征融合。特别的, CFEM 模块通过计算跨时相特征的余弦相似度并结合双通道注意力机制生成动态权重, 从而精准增强变化区域的语义表达并抑制噪声, 显著提高了对细节变化的敏感度和准确性。MS-SwinCE 利用多尺度 Swin Transformer 增强了长程依赖建模能力, Swin Transformer 通过局部窗口自注意力机制与移位操作, 增强了全局上下文建模能力, 能够有效捕捉遥感图像中的远距离变化信息。此外, 在

特征提取过程中, squeeze-and-excitation block (SE-Block) 模块通过全局平均池化动态调整通道权重, 进一步增强了关键特征的表达能力。SEBlock 的引入有助于优化不同通道之间的依赖关系, 使得模型能够更加精确地捕捉变化信息, 提升模型对细节变化的表现。并通过 BiFPN 高效融合多尺度特征, BiFPN 通过自上而下与自下而上的双向路径传递信息, 结合可学习加权融合机制, 动态调整不同尺度特征的权重, 进一步提升变化检测的空间预测一致性, 尤其在细节和小目标检测方面表现突出。

2 方法

2.1 模型概述

图 1 展示了模型整体架构, 输入为变化前后的图像。模型包括基于 Swin Transformer 的编码器、对比特征增强模块 (CFEM)、双向特征金字塔网络 (BiFPN) 和轻量化解码器。编码器分为 4 个阶段, 第 1 阶段通过 patch embedding 进行下采样, 再输入 Swin Transformer block。后续阶段通过 patch merging 进行下采样后进入 Swin Transformer block。Swin Transformer block 利用局部窗口自注意力机制提取多尺度特征, 并结合 SEBlock 增强通道间的依赖关系, 提升特征表达能力。

CFEM 模块通过跨层特征差异计算, 增强图像变化的语义信息并抑制噪声, 同时校准不同尺度的特征, 提高变化检测的敏感度和细节感知能力。

BiFPN 模块通过双向信息流动高效融合多尺度特征, 实现低层次细节与高层次语义特征的互补, 提升变化检测的准确性与空间预测一致性。

2.2 Swin Transformer 特征提取器

图 2 展示了 Swin Transformer block 的核心结构。其由多个关键模块组成, 旨在通过分层设计高效建模全局和局部信息。W-MSA 表示窗口注意力机制 (window attention mechanism), SW-MSA 表示移动窗口注意力机制 (shifted window attention mechanism)。局部窗口自注意力机制在每个窗口内计算特征的相关性, 降低了计算复杂度, 同时保留了窗口内的上下文信息。为了实现跨窗口的信息交互, Swin Transformer 通过移位操作

在每层调整窗口位置, 确保相邻窗口的特征能互相联系, 从而捕捉长程依赖关系^[7].

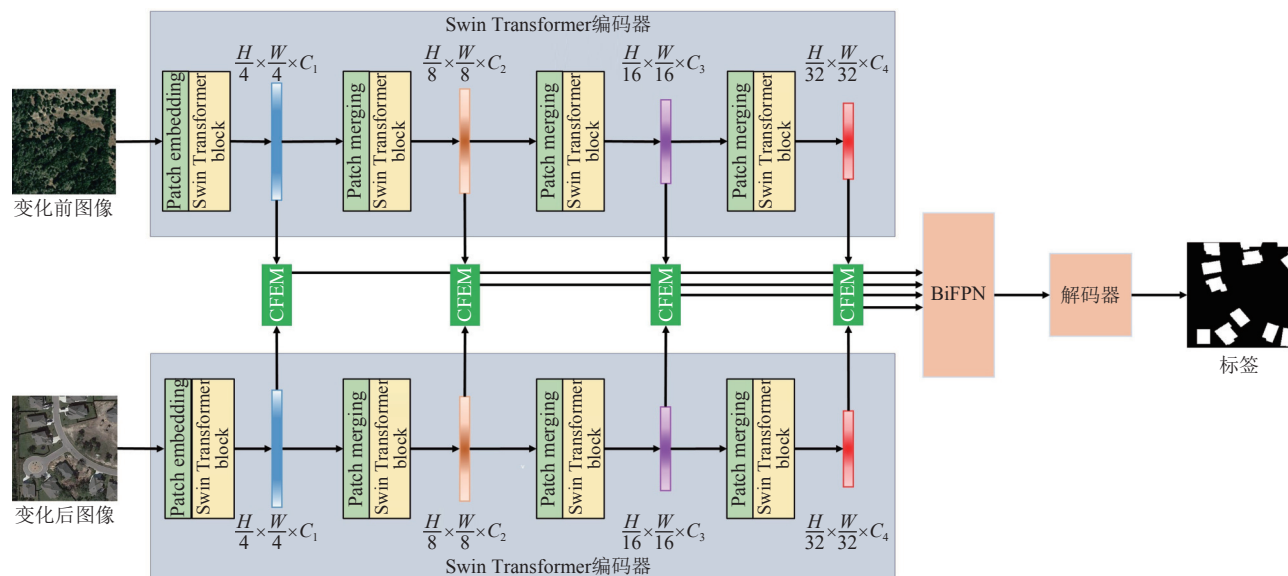


图1 MS-SwinCE 结构图

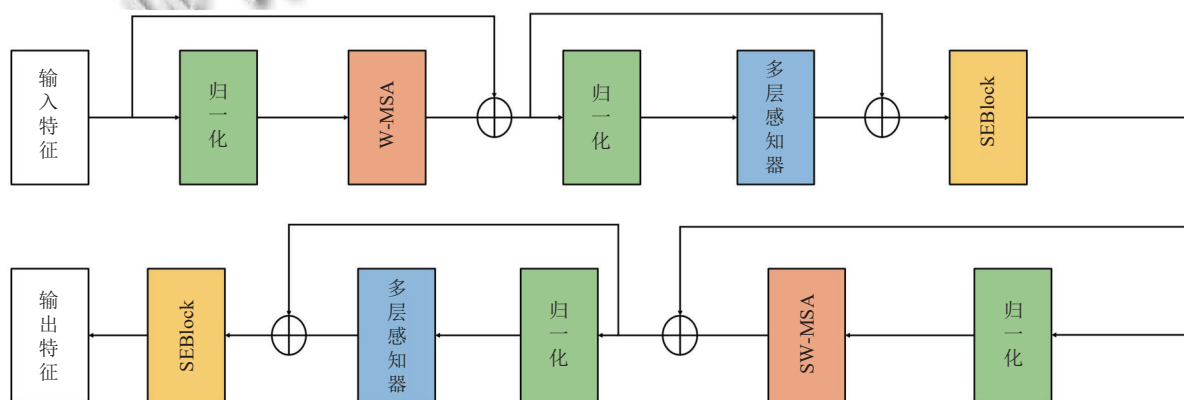


图2 Swin Transformer block 结构

为了进一步增强通道特征表达能力, Swin Transformer block 引入 SEBlock 模块, 通过全局平均池化 (global average pooling, GAP) 提取每个通道的全局信息, 并动态调整通道权重^[11]. 公式如下:

$$F_{\text{out}} = F_{\text{in}} \cdot \text{Sigmoid}(W_2(\text{ReLU}(W_1(\text{GAP}(F_{\text{in}})))))) \quad (1)$$

2.3 对比特征增强模块 (CFEM)

在遥感图像变化检测中, 高分辨率遥感数据的多光谱波段特性以及多时相获取机制常引入复杂的背景干扰, 如建筑阴影位移和植被光谱季节性变异. 此外, 真实变化区域通常表现为局部细微的光谱特征突变, 这些变化极易受到云层遮挡、地物光谱噪声等时相差异因素的影响, 导致其被淹没或无法有效检测. 为了解决这一问题, 本文设计了对比特征增强模块 (CFEM).

该模块针对遥感数据中跨时相光谱相关性衰减和伪变化区域光谱相似性干扰的挑战, 通过余弦相似度计算变化前后特征的通道级差异, 精确捕捉真实变化区域的光谱特征突变点. 同时, CFEM 模块结合注意力机制, 动态抑制背景噪声通道, 实现对遥感图像中细微变化区域的定向增强与噪声过滤. 图3展示了 CFEM 模块通过特征相似性计算与双通道注意力加权实现变化区域强化的具体架构.

本研究提出的 CFEM 以变化前后的特征作为输入, 首先通过余弦相似度 (cosine similarity, CosSim) 计算通道维度的特征相关性. 余弦相似度公式为:

$$\text{CosSim}(F_{\text{pre}}^i, F_{\text{post}}^i) = \frac{F_{\text{pre}}^i \cdot F_{\text{post}}^i}{\|F_{\text{pre}}^i\| \|F_{\text{post}}^i\| + \varepsilon} \quad (2)$$

其中, ϵ 为数值稳定性的平滑参数, 用于防止除零错误^[12,13]. 通过此操作, 可以精准捕捉变化区域的通道级特征差异, 减少背景区域的影响. 接下来, 将前后特征图通过

逐元素加法操作进行融合, 得到初始差异特征图, 综合两组特征的局部相似性, 为后续全局特征的提取奠定了基础^[6].

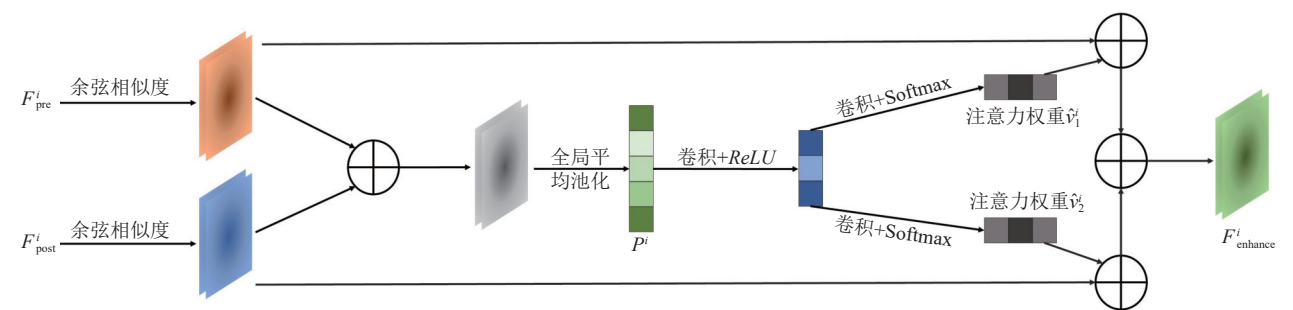


图3 CFEM 模块架构

初始差异特征图通过全局平均池化 (GAP) 操作, 将空间维度压缩为通道级特征向量, 为通道注意力权重的生成提供依据^[6]. 之后, 通过卷积操作和 ReLU 激活函数对特征向量进行非线性变换, 生成高维表示向量, 并通过两个独立卷积层生成两组注意力权重, 最后通过 Softmax 函数归一化权重, 确保通道之间的权重分布平衡.

为了充分利用注意力权重对变化区域的敏感性, 模块对变化前后特征分别进行加权, 具体来说, 将注意力权重与变化前特征和变化后特征逐元素相乘, 得到增强后的前后特征^[6]. 该增强特征不仅保留了变化区域的显著信息, 还有效抑制了背景噪声的干扰, 优于传统的简单特征差分方法. CFEM 模块通过动态调整特征的权重, 能够更加准确地捕捉细节变化, 同时提升对复杂场景的适应能力.

2.4 双向特征金字塔模块 (BiFPN)

在遥感图像变化检测中, 高分辨率影像常包含尺

度差异显著的变化目标 (如城市扩张中的大型建筑群与农田边缘的细微调整), 而传统 FPN 的单向信息流难以同时兼顾高分辨率细节与低分辨率语义. BiFPN 是对传统 FPN 的改进结构, 具备高效的多尺度特征融合能力, 广泛应用于目标检测任务. 其核心设计在于引入自上而下与自下而上的双向信息流, 并结合自适应加权机制, 有效增强不同分辨率特征间的信息交互与语义表达^[9,14]. 对于遥感数据中的跨时相多尺度特征 (不同季节植被覆盖变化的尺度差异), BiFPN 通过双向路径实现互补: 自上而下路径传递高分辨率细节以捕捉细微变化; 自下而上路径聚合低分辨率语义以理解大型地物变化逻辑. 结合动态权重分配机制, 可针对遥感场景中目标尺度非均匀分布特性 (数据集中建筑变化尺寸跨度达数十至数百像素) 强化关键尺度特征, 提升多尺度变化目标的感知能力与定位精度. 图 4 展示了 BiFPN 的结构示意, 其中输入特征图经双向融合路径与权重优化后生成增强特征.

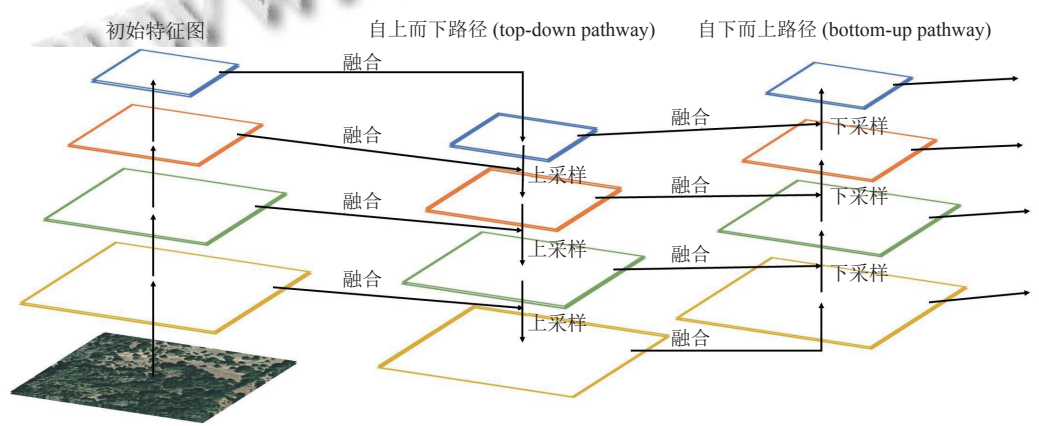


图4 BiFPN 模块架构

具体而言, BiFPN 通过自上而下 (top-down) 和自下而上 (bottom-up) 双向路径实现高效信息流动: 高分辨率特征向下传递细节信息, 低分辨率特征向上传递语义信息, 从而增强多尺度目标的感知能力^[9,14]. 在每一融合节点, 输入特征图先经 1×1 卷积统一通道数, 再通过深度可分离卷积进行压缩处理, 提升融合效率与匹配性^[9].

同时, BiFPN 引入自适应加权机制 (adaptive weighting mechanism), 为不同来源的特征图分配可学习权重, 以强化关键尺度信息的贡献. 权重经训练动态调整^[9,14]. 其融合公式如下:

$$F_{\text{fused},i} = \omega_1 \cdot P_i + \omega_2 \cdot P_{i+1} \quad (3)$$

其中, ω_1 和 ω_2 为每一层特征图的权重, P_i 和 P_{i+1} 分别是低分辨率和高分辨率的特征图. 该机制能够突出关键尺度信息, 有效提升特征融合质量与检测精度^[9,14].

2.5 训练优化策略

为提升模型的泛化能力并抑制过拟合, 本文在训练阶段引入了多种数据增强方法, 包括随机水平翻转、垂直翻转、 90° 旋转与归一化操作. 上述增强操作在保持图像语义一致性的基础上, 有效扩大了训练样本空间, 提高了模型对姿态与视角变化的适应能力.

模型训练采用 AdamW (Adam with weight decay) 优化器, 初始学习率设为 1×10^{-4} , 权重衰减因子为 0.01. 为实现更平滑的收敛过程, 引入多项式衰减 (polynomial decay) 学习率调度策略, 其形式为:

$$\alpha_t = \alpha_0 \cdot \left(1 - \frac{t}{T}\right)^{0.9} \quad (4)$$

其中, α_0 为初始学习率, t 表示当前训练轮次, T 为最大训练轮数.

在损失函数设计方面, 为兼顾样本类别不平衡与样本难度差异, 本文将加权交叉熵损失 (weighted cross-entropy loss) 与焦点损失 (focal loss) 进行加权融合. 交叉熵损失函数广泛用于分类任务, 可通过度量预测概率分布与真实标签之间的差异, 有效驱动模型学习正确的判别边界^[15], 其定义为:

$$L_{\text{CE}}(p, y) = -(y \cdot \log(p) + (1 - y) \cdot \log(1 - p)) \quad (5)$$

其中, p 为模型预测的正类概率, y 为真实标签.

Focal loss 引入焦点因子 γ 以降低易分类样本对损失的主导作用, 从而聚焦于难分类样本^[16]. 其公式为:

$$L_{\text{Focal}}(p_t) = -\alpha_t \cdot (1 - p_t)^\gamma \cdot \log(p_t) \quad (6)$$

其中, α_t 是类别权重, 用于平衡正负类的影响. γ 是焦点因子 (本文取 $\gamma = 2$), p_t 为当前预测概率.

综合考虑两者的优势, 本文最终损失函数定义为:

$$L = \lambda \cdot L_{\text{CE}} + (1 - \lambda) \cdot L_{\text{Focal}} \quad (7)$$

其中, $\lambda \in [0, 1]$ 为损失加权系数, 控制两种损失函数的贡献比例. 该组合策略旨在提升模型的整体性能^[4].

3 实验分析

3.1 数据集描述与评价指标

本研究使用两个典型的遥感图像变化检测数据集: LEVIR-CD 和 WHU-CD. LEVIR-CD 包含 637 对 1024×1024 的高分辨率时序影像, 涵盖建筑增减与城市扩展等典型场景. 为适应模型输入并提升样本量, 原始图像被裁剪为 256×256 的图像块, 并按 7:1:2 的比例划分为训练集、验证集和测试集. WHU-CD 则由大量已标注的 256×256 图像块构成, 具有丰富的地物类型和高质量的变化标注, 广泛用于模型泛化能力评估, 同样采用 7:1:2 的比例进行数据划分.

为全面评估模型在变化检测任务中的性能, 本文采用 5 项常用评价指标: 总体精度 (overall accuracy, OA)、准确率 (Precision)、召回率 (Recall)、F1 分数 (F1) 和交并比 (intersection over union, IoU). 基于 4 类分类结果: 真正例 (true positive, TP) 为模型正确识别的正类样本数; 真负例 (true negative, TN) 为模型正确识别的负类样本数; 假正例 (false positive, FP) 为模型误判为正类的负类样本数; 假负例 (false negative, FN) 为模型误判为负类的正类样本数.

总体精度 (OA) 是所有预测结果中正确分类的样本所占的比例, 用于衡量模型在所有类别上的总体分类效果. 其计算公式为:

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

准确率 (Precision) 衡量的是模型预测为正类的样本中, 实际为正类的比例, 反映了模型在预测正类时的准确性. 其计算公式为:

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

召回率 (Recall) 表示的是模型能够正确识别的所有正类样本的比例, 反映了模型对正类样本的捕获能力. 其计算公式为:

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

$F1$ 分数 ($F1$) 是准确率和召回率的调和平均值, 是对两者综合考虑的评价指标. 其计算公式为:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (11)$$

交并比 (IoU) 是用于衡量预测区域与真实变化区域重叠度的指标, 通常用于评估变化检测的精度. 其计算公式为:

$$IoU = \frac{TP}{TP + FP + FN} \quad (12)$$

3.2 结果与分析

表 1 和表 2 展示了不同方法在 LEVIR 和 WHU-CD 数据集上的性能对比, 实验结果表明, 本文方法在所有指标上均表现最优, 显著优于次优方法的性能. 在 LEVIR 数据集上, 本文方法相比 ChangeFormer 在 OA 上提升了 0.06%, $Precision$ 提升 0.32%, $F1$ 提升 0.70%, $Recall$ 提升 1.06%, IoU 提升 1.18%, 并且比 BIT 方法在 IoU 上提升了 2.98%. 在 WHU-CD 数据集上, 相比 BIT 方法, OA 提升 0.10%, $Precision$ 提升 0.47%, $F1$ 提升 1.06%, $Recall$ 提升 1.66%, IoU 提升 1.84%. 综合来看, 本文方法在 LEVIR 和 WHU-CD 数据集上均优于现有方法, 特别是在 IoU 和 $F1$ 等关键指标上取得了显著提升, 验证了其在多尺度变化检测任务中的准确性与适用性. 此外, 尽管保持较高精度, 本文方法的参数量为 31.66M, 显著低于 ChangeFormer (41.03M), 在保持优越性能的同时展现了效率与性能之间的良好平衡.

表 1 LEVIR 数据集对比结果

Methods	OA (%)	$Precision$ (%)	$F1$ (%)	$Recall$ (%)	IoU (%)	Parameters (M)
FC-EF ^[17]	98.39	86.91	83.40	80.17	71.53	1.35
FC-Siam-Di ^[17]	98.67	89.53	86.31	83.31	75.92	1.35
FC-Siam-Conc ^[17]	98.69	91.99	83.69	76.77	71.96	1.55
DTCDSCN ^[18]	98.77	88.53	87.67	86.83	78.05	41.04
BIT ^[4]	98.92	89.24	89.31	89.37	80.68	12.4
ChangeFormer ^[5]	99.04	92.05	90.40	88.80	82.48	41.03
Ours	99.10	92.37	91.10	89.86	83.66	31.66

注: 加粗结果表示最优

表 2 WHU-CD 数据集对比结果

Methods	OA (%)	$Precision$ (%)	$F1$ (%)	$Recall$ (%)	IoU (%)	Parameters (M)
FC-EF ^[17]	98.05	84.40	78.31	73.04	64.36	1.35
FC-Siam-Di ^[17]	98.32	89.62	80.97	73.84	68.03	1.35
FC-Siam-Conc ^[17]	98.50	86.90	83.95	81.19	72.34	1.55
DTCDSCN ^[18]	99.03	93.58	89.58	85.91	81.13	41.04
BIT ^[4]	99.10	92.46	90.53	88.62	82.65	12.4
ChangeFormer ^[5]	99.00	92.79	90.39	88.12	79.49	41.03
Ours	99.20	92.93	91.59	90.28	84.49	31.66

注: 加粗结果表示最优

图 5 展示了不同方法的变化检测可视化结果. 对于 LEVIR 数据集, 本文方法能更准确地检测新增建筑物并减少虚假检测; 而对于 WHU-CD 数据集, 能够更精确地保持目标轮廓, 减少误检和漏检, 相较于 BIT 和 ChangeFormer 表现更优. 总体而言, 本文方法在两个数据集上的表现突出, 有效减少了误检、漏检, 并提高了目标区域完整性, 验证了其优越性.

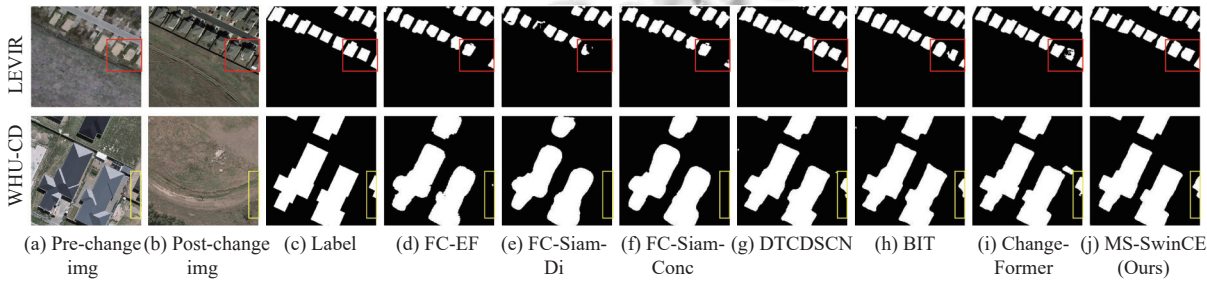


图 5 不同方法在 LEVIR 和 WHU 数据集上的变化检测结果对比

3.3 消融实验

表 3 和图 6 展示了各模块对模型性能和变化检测效果的影响. 完整模型 (Exp-1, Baseline) 在各项指标上表现最优, OA 为 99.10%, $F1$ 为 91.10%, IoU 为 83.66%, 且能准确识别变化区域. 移除 SEBlock 模块 (Exp-2) 后, $Precision$ 和 IoU 分别下降 0.67% 和 0.77%, 说明

SEBlock 能有效增强特征通道表征, 提升精度和细节分割效果. 去除 CEFM 模块 (Exp-3) 时, $Recall$ 和 $F1$ 下降 5.02% 和 3.04%, IoU 降至 78.66%, 证明 CEFM 在多尺度特征融合和上下文信息提取中的关键作用, 提升了召回率和分割完整性. 移除 BiFPN 模块 (Exp-4) 后, $Precision$ 、 $Recall$ 和 IoU 有所下降, 特别是 IoU 减少

2.29%, 验证了 BiFPN 在融合跨尺度特征和提升边缘细节检测中的重要性. 仅保留 Backbone (Exp-5) 时, 性能大幅下降, IoU 降至 78.81%, 出现明显误检和漏检, 显示各辅助模块对性能提升至关重要.

综上, SEBlock、CEFM 和 BiFPN 在精确度、召回率和细节识别能力方面共同提升了模型的性能, 三者协同作用确保了模型在变化检测任务中的优越性.

表 3 消融实验结果 (%)

实验编号	OA	$Precision$	$F1$	$Recall$	IoU
Exp-1 (Baseline)	99.10	92.37	91.10	89.86	83.66
Exp-2 (-SEBlock)	99.05	91.70	90.64	89.61	82.89
Exp-3 (-CEFM)	98.82	91.53	88.06	84.84	78.66
Exp-4 (-BiFPN)	98.97	91.18	89.73	88.32	81.37
Exp-5 (Backbone)	98.83	91.16	88.15	85.32	78.81

注: 加粗表示最优结果

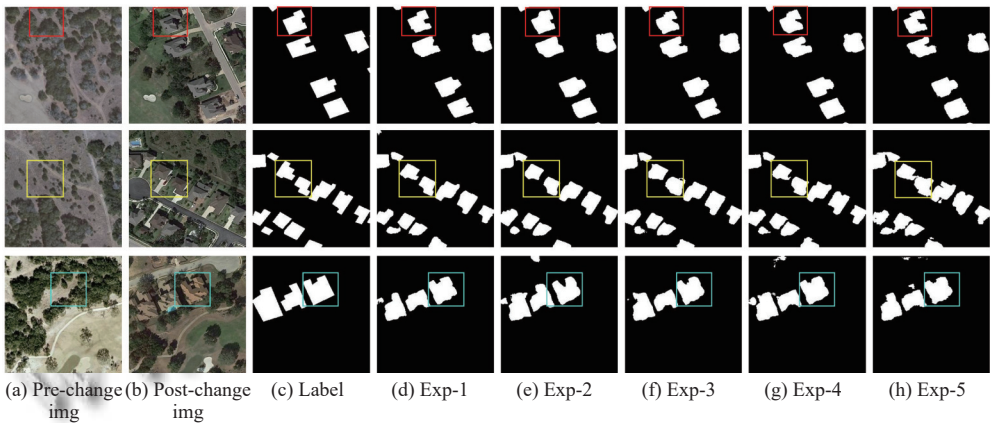


图 6 LEVIR 数据集上模型消融实验结果对比

4 结论与展望

本文所提出的 MS-SwinCE 模型通过引入多尺度 Swin Transformer、对比特征增强模块 (CFEM) 和双向特征金字塔网络 (BiFPN), 有效提升了遥感图像变化检测的精度与空间一致性. 模型结合局部窗口与全局依赖的特征提取能力、特征差异增强与多尺度特征融合, 显著提高了变化区域的检测能力. 适用于城市扩张监测、自然灾害前后的地物变化识别、生态保护区植被覆盖动态分析等场景. 实验结果表明, MS-SwinCE 在 LEVIR-CD 和 WHU-CD 数据集上均优于现有模型, 在 IoU 、 $F1$ 、 $Precision$ 和 $Recall$ 等关键指标上取得全面提升. 消融实验验证了各模块设计对性能提升的贡献. 未来将探索模型在大规模数据集和不同遥感场景下的泛化能力, 扩展其实际应用.

参考文献

1 Li WM, Xue LH, Wang XQ, *et al.* ConvTransNet: A CNN-Transformer network for change detection with multiscale global-local representations. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 5610315. [doi: 10.1109/TGRS.2023.3272694]

2 Yin MM, Chen ZB, Zhang CJ. A CNN-Transformer network

combining CBAM for change detection in high-resolution remote sensing images. *Remote Sensing*, 2023, 15(9): 2406. [doi: 10.3390/rs15092406]

3 Song XY, Hua Z, Li JJ. PSTNet: Progressive sampling Transformer network for remote sensing image change detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2022, 15: 8442–8455. [doi: 10.1109/JSTARS.2022.3204191]

4 Chen H, Qi ZP, Shi ZW. Remote sensing image change detection with Transformers. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 5607514. [doi: 10.1109/TGRS.2021.3095166]

5 Bandara WGC, Patel VM. A Transformer-based Siamese network for change detection. *Proceedings of the 2022 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2022)*. Kuala Lumpur: IEEE, 2022. 207–210. [doi: 10.1109/IGARSS46834.2022.9883686]

6 Noman M, Fiaz M, Cholakkal H, *et al.* Remote sensing change detection with Transformers trained from scratch. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 4704214. [doi: 10.1109/TGRS.2024.3383800]

7 Liu Z, Lin YT, Cao Y, *et al.* Swin Transformer: Hierarchical vision Transformer using shifted windows. *Proceedings of the 2021 IEEE/CVF International Conference on Computer*

- Vision. Montreal: IEEE, 2021. 9992–10002. [doi: [10.1109/ICCV48922.2021.00986](https://doi.org/10.1109/ICCV48922.2021.00986)]
- 8 Zhang XF, Cheng SL, Wang LJ, *et al.* Asymmetric cross-attention hierarchical network based on CNN and Transformer for bitemporal remote sensing images change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 2000415. [doi: [10.1109/TGRS.2023.3245674](https://doi.org/10.1109/TGRS.2023.3245674)]
- 9 Lin TY, Dollár P, Girshick R, *et al.* Feature pyramid networks for object detection. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017. 936–944. [doi: [10.1109/CVPR.2017.106](https://doi.org/10.1109/CVPR.2017.106)]
- 10 Li QY, Zhong RF, Du X, *et al.* TransUNetCD: A hybrid Transformer network for change detection in optical remote-sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 5622519. [doi: [10.1109/TGRS.2022.3169479](https://doi.org/10.1109/TGRS.2022.3169479)]
- 11 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 7132–7141. [doi: [10.1109/CVPR.2018.00745](https://doi.org/10.1109/CVPR.2018.00745)]
- 12 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas: IEEE, 2016. 770–778.
- 13 Wang XL, Girshick R, Gupta A, *et al.* Non-local neural networks. *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 7794–7803. [doi: [10.1109/CVPR.2018.00813](https://doi.org/10.1109/CVPR.2018.00813)]
- 14 Tan MX, Pang RM, Le QV. EfficientDet: Scalable and efficient object detection. *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle: IEEE, 2020. 10778–10787. [doi: [10.1109/CVPR42600.2020.01079](https://doi.org/10.1109/CVPR42600.2020.01079)]
- 15 Li XL, Zhang XT, Huang W, *et al.* Truncation cross entropy loss for remote sensing image captioning. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 59(6): 5246–5257. [doi: [10.1109/TGRS.2020.3010106](https://doi.org/10.1109/TGRS.2020.3010106)]
- 16 Lin TY, Goyal P, Girshick R, *et al.* Focal loss for dense object detection. *Proceedings of the 2017 IEEE International Conference on Computer Vision*. Venice: IEEE, 2017. 2999–3007. [doi: [10.1109/ICCV.2017.324](https://doi.org/10.1109/ICCV.2017.324)]
- 17 Caye Daudt R, Le Saux B, Boulch A. Fully convolutional siamese networks for change detection. *Proceedings of the 25th IEEE International Conference on Image Processing (ICIP)*. Athens: IEEE, 2018. 4063–4067. [doi: [10.1109/ICIP.2018.8451652](https://doi.org/10.1109/ICIP.2018.8451652)]
- 18 Liu Y, Pang C, Zhan ZQ, *et al.* Building change detection for remote sensing images using a dual-task constrained deep siamese convolutional network model. *IEEE Geoscience and Remote Sensing Letters*, 2021, 18(5): 811–815. [doi: [10.1109/LGRS.2020.2988032](https://doi.org/10.1109/LGRS.2020.2988032)]

(校对责编: 王欣欣)