

# 用于遥感影像变化检测的频时注意力网络

于春燕, IEEE 高级会员, 李浩波, 胡亚斌, 张强, IEEE 会员, 宋美平, IEEE 会员, 王玉磊, IEEE 会员

**摘要** — 遥感影像中的变化检测 (CD) 被认为是对地观测领域的关键任务, 而它通常面临数据复杂和微小变化的困境。为了应对上述挑战, 这封信提出了一种创新的 CD 频时注意力网络 (FTAN), 它包含两个高级模块, 包括多维卷积频率注意力模块 (MCFA) 和交互式注意力模块 (IAM)。具体来说, MCFA 模块对于通过合并多尺度空间和频域特征来提高 CD 的灵敏度至关重要。作为 MCFA 的补充, IAM 聚合了与类别相关的令牌, 并处理来自不同时间段的交叉关注信息。MCFA 和 IAM 的无缝集成使 FTAN 网络具有更强大的功能, 可以准确检测次要区域和边缘。在 LEVIR-CD 和 DSIFN-CD 等数据集上的实验表明, 在 F1 分数和 IoU 指标方面优于现有模型, 从而表现出卓越的性能。我们的代码和预训练模型将于 <https://github.com/chirsycy/FTAN> 日发布。

**索引术语** — 对抗性训练、域适应、高光谱图像 (HSI) 分类、迁移学习。

## I. I

遥感影像 (RSI) 的 CHANGE 检测 (CD) [1] 是指自动检测同一场景的多时相影像的差异或变化, 对土地覆盖发展、城市数据采集和环境监测具有重要意义。一般来说, RSI 的特点是非线性特征, 包括空间和光谱可变性。传统方法在处理上述复杂性时遇到了挑战。近年来, 深度学习 (DL) 方法 [2]、[3]、[4]、[5]、[6]、[7] 由于在遥感应用领域具有自动特征提取和卓越的性能, 取得了长足的进步。如今, CD 的基于注意力的模型 [8] 促进了对特定变化对象的识别。其中, 自我注意机制, 例如

基于 Transformer 的网络 [8] 捕获长距离依赖关系, 提供了一种更有效的空间信息编码方法。通常, 许多模型采用时空注意力机制来改进和改进 CD 实现的功能。TinyCD [9] 结合了来自全局时间和局部空间信息的 CD 的低级特征, 用于时空特征融合。ChangeFormer [10] 直接从输入图像中提取 CD 相关信息, 并进行上下文建模, 以获得高效、准确的检测结果。最近, 一些 CD 模型专注于解决边缘检测挑战。特别是, EGDE-Net [11] 提出了用于边界信息细化的边缘感知模块。Changer [12] 通过双时间特征之间的交互构建了一个新的 CD 模式, 这对细节和边缘是有益的。最近, 在计算机视觉领域, 频域学习已成为增加通道注意力的一种流行方式, 它可以捕获全局模式和长期依赖性。尽管现有的基于边缘的 CD 模型产生了令人印象深刻的结果, 但频域信息的缺失导致边界检测和微小的变化不完整或不准确。涉及频域的 RS-CD 方法有效地过滤了噪声并增强了细微变化的重要特征。

在本文中, 我们提出了一种新的 CD 频时注意力网络 (FTAN), 其中多维卷积频率注意力模块 (MCFA) 巧妙地将多尺度空间特征与频域特征相结合。值得注意的是, 作为 MCFA 核心的基于频域掩码的自适应卷积 (FDM-AC) 擅长降低噪声和突出关键信号, 这对于提高 CD 的灵敏度至关重要。此外, 交互式注意力模块 (IAM) 通过自适应注意力机制合并交互式空间数据, 丰富局部和全局特征表示, 有效实现区域连续性并促进检测连通性。本研究的主要贡献总结如下:

- 1) 与以前的模型不同, 我们提出了一种新的频率-时空-时间注意力网络来增强 RSI CD 的边缘表示。据我们所知, 这是在 CD 框架中将时空提取的频率信息组合在一起的第一次尝试。值得注意的是, 核心 MCFA 机制将多尺度注意力与频域信息相结合, 以加强时空特征, 有利于次要目标检测。
- 2) 与以前的方法不同, 所提出的 IAM 捕获交互式远程依赖关系。作为模型的关键部分, IAM 聚合了与类别相关的

接收日期 2024 年 6 月 26 日;修订于 2024 年 9 月 9 日;2024 年 9 月 27 日接受。发布日期: 2024 年 10 月 10 日;当前版本的日期 2024 年 10 月 24 日。这项工作部分得到了 Grant 62471079 和 Grant 62401095 下的中国国家自然科学基金的支持, 部分得到了 Grant 3132017124 下的中央大学基本研究基金的支持。(通讯作者: 张强)

Chunyan Yu、Haobo Li、Qiang Zhang、Meiping Song 和 Yulei Wang 就职于中国大 116026 连海事大学信息与技术学院高光谱成像中心 (CHIRS) (电子邮件: yucy@dlmu.edu.cn、lihaobo1998@gmail.com、qzhang95@dlmu.edu.cn、smping@163.com、wangyulei@dlmu.edu.cn)。Yabin 胡 就职于中国青岛 266061 自然资源部第一海洋研究所和海洋遥测技术创新中心 (电子邮件: huyabin@fio.org.cn)。

数字对象标识符 10.1109/LGRS.2024.3477991

1558-0571 © 2024 IEEE。允许个人使用, 但重新发布/重新分发需要 IEEE 许可。  
有关更多信息, 请参阅 <https://www.ieee.org/publications/rights/index.html>。

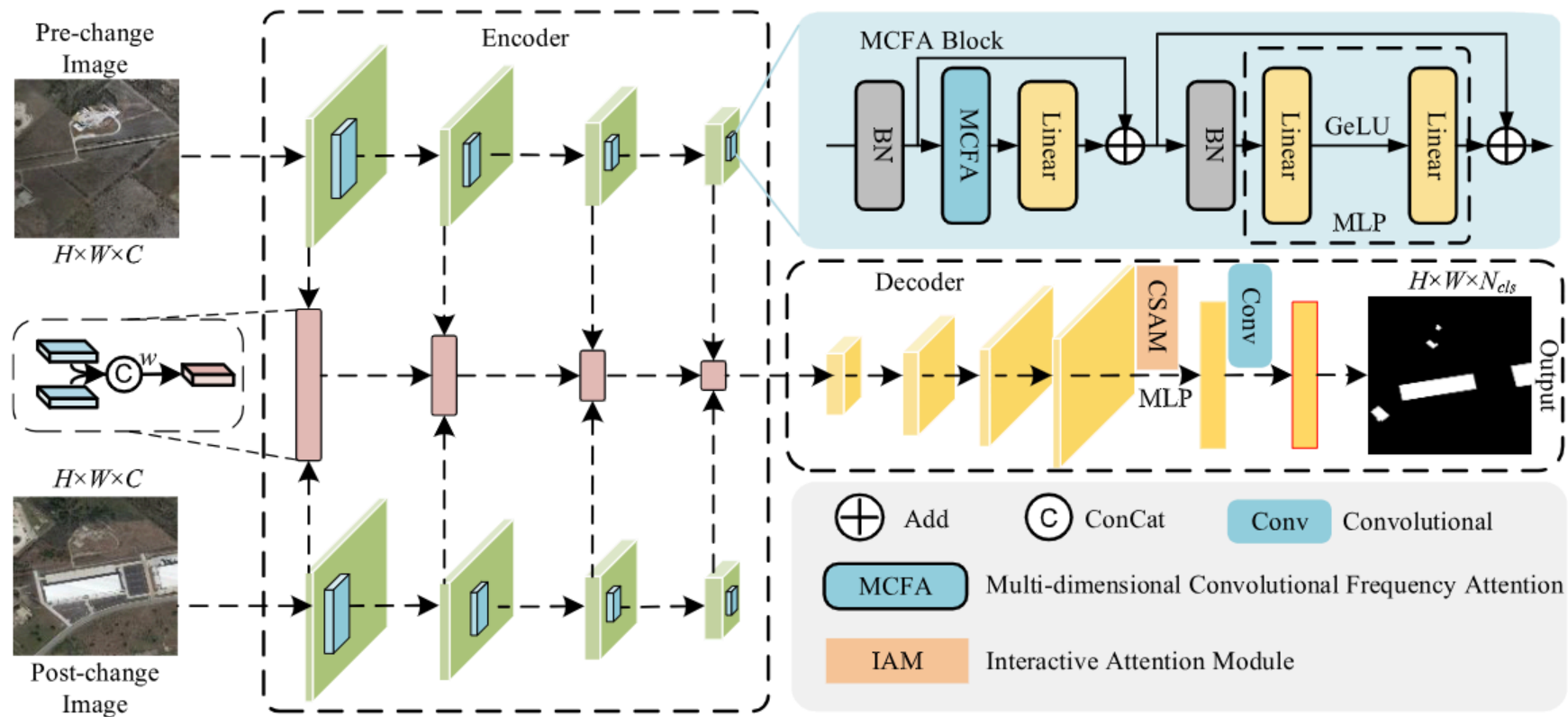


图 1. 建议的 FTAN 架构图示。

令牌，并适应时空背景。

II. PA 拟议的 FTAN 的整体架构如图 1 所示。正如所观察到的，编码器利用新颖的频率感知卷积注意力来提取显著性，带有 IAM 的解码器块负责在部分区域提炼和生成连续检测图。更多详细信息在第 II-A 节和第 II-B 节中概述。

### A. 多维卷积频率注意力

1) 编码器设计：我们采用卷积显著性提取代替注意力提取。总的来说，我们方法的编码器包含四个阶段进行精细的显著性提取，分辨率降低为  $H/4 \times W/4$ 、 $H/8 \times W/8$ 、 $H/16 \times W/16$ 、 $H/32 \times W/32$ ，其中  $H$  和  $W$  分别表示输入图像的高度和宽度。该模式保持了微小变化的显著特征，并带有卷积显著性提供的自我关注。

2) MCFA：如图 2 所示，MCFA 的主要元件包括 FDM-AC、部分卷积运算 (PConv) 和多分支深度条带卷积块。具体来说，FDM-AC 中内置了四个并行模块。在每个模块中，空间域图像  $X$  首先通过快速傅里叶变换 (FFT) 转换为频域表示，如后续方程所示

$$X = F(X) \text{ 的 } F. \quad (1)$$

然后，创建由预定义参数缩放的频率掩码以匹配  $X$  的大小，掩码定义如下：

$$\text{面具} = \begin{cases} 1, & u - \frac{M}{2} < \text{scale} \text{ 和 } v - \frac{N}{2} < \text{scale} \\ 0, & \text{否则} \end{cases} \quad (2)$$

其中  $M$  和  $N$  表示行和列， $u$  和  $v$  表示频率的水平和垂直坐标

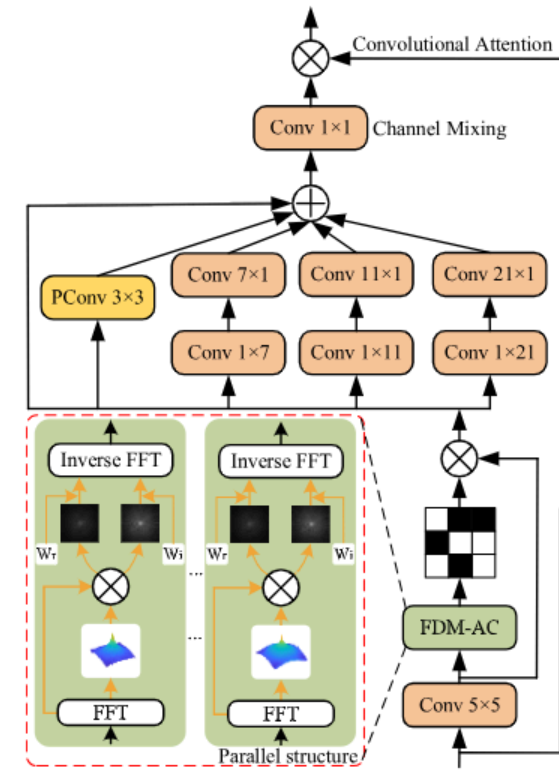


图 2. MCFA 插图。首先，FDM-AC 生成频域混淆特征，显示在并行结构中，PConv 保留来自剩余通道的信息，多分支深度条带卷积块使用  $1 \times n$ 、 $n \times 1$  个内核进行空间细节捕获。

域。使用掩码，我们获得了 RSI 的过滤频域图像，用以下公式表示为  $X_{\text{masked}}$ ：

$$X_{\text{masked}} = X \cdot \text{面具}. \quad (3)$$

根据 FFT 的固有性质， $X$  由实  $X$  和虚构的  $X_{\text{parts}}$  分解。为了进一步挖掘 RSI 的频域信息，两个不同权重的卷积运算分别表示为魔杖器来处理实部和虚部。值得注意的是， $w$ ， $w \in \mathbb{C}$ ，其中  $\mathbb{C}$  表示复数集， $c$  是通道数。

在融合运算中，输出嵌入  $X_{\text{of}}$  FDM-AC 由 (4) 产生，其中  $*$  表示卷积运算， $F$  表示逆 FFT

$$X = F \left( \frac{1}{4} \sum_{i=1}^4 X * w + i \right) \quad (4)$$



在 MCFA 中, 采用 Pconv 来保留显著性信息, 以改进边界和细节预测, 因为卷积核仅在输入张量通道的子集上。Pconv 的实现公式如下:

$$W \otimes X = \text{连接} (W(X), X) \quad (5)$$

其中  $X$  表示  $X \otimes X$ ,  $\otimes$  是矩阵乘法运算符,  $W$  表示  $3 \times 3$  卷积运算。 $X$  表示  $X$  的通道的前面部分, 而  $X$  是指  $X$  的其余通道。

此外, RSI 中的深度条带卷积利用  $1 \times n$ 、 $n \times 1$  个内核来捕获更改对象中的空间细节。具体来说, 从 MCFA 中提取的显著性信息使用以下公式获得:

$$X = W \otimes X \quad (6)$$

$$\tilde{X} = X \oplus W(X) \oplus \text{宽}(W(X)) \quad (7)$$

其中  $W$  表示核大小为  $5 \times 5$  的卷积运算,  $X$  表示  $W$  applied 到  $X$  的输出,  $\otimes X$  是具有多个分支的串联映射。 $\oplus$  是串联作,  $W$  represents Pconv。此外,  $W$  and  $W$  分别表示  $1 \times n$  和  $n \times 1$ ,  $n \in \{7, 11, 21\}$  的卷积运算。

最后,  $X$  的优化特征是通过由  $1 \times 1$  卷积层引导的注意力机制实现的

$$H = W \otimes \tilde{X} \otimes X \quad (8)$$

其中  $H$  表示 MCFA 的输出,  $\otimes$  是矩阵乘法运算符,  $W$  表示  $1 \times 1$  运算。

### B. 互动注意力模块

如图 3 所示, IAM 模块位于 FTAN 解码器的末端。在结构上, 解码器采用四个上采样阶段的级联来逐步提高空间分辨率, 并结合残差块来优化特征。具体来说, Token 序列由 Inter Tokens 和 Cls Tokens 组成, 其中 Inter Token 是通过特征图的线性变换得到的, 而 Cls Tokens 是在初始化过程中随机初始化的类别相关 Token。最终, 解码器通过  $3 \times 3$  卷积层处理细化的特征图。值得注意的是, Cls 令牌被用作交互式桥梁, 并有效地集成和比较来自不同时间图像的数据。此外, 它仅作为自注意力计算中的键和值, 使模型能够直接访问和处理时空数据, 便于处理复杂的时空信息。自我情境注意力的计算方式如下:

$$H = \text{Softmax}_T \left( \frac{Q(s) \times K(s)}{C} \cdot V(s) \right) \quad (9)$$

其中  $s$  来自标记间的输入,  $t$  表示像素位置,  $q$  是从  $s$  到  $t$  的线性映射,  $k$  表示从  $s$  到  $t$  的键的线性映射,  $v$  表示从  $s$  到  $t$  的值的线性映射,  $T$  是像素总数,  $C$  是输入特征图  $s$  的通道数,  $H$  表示自我注意机制的中间特征。

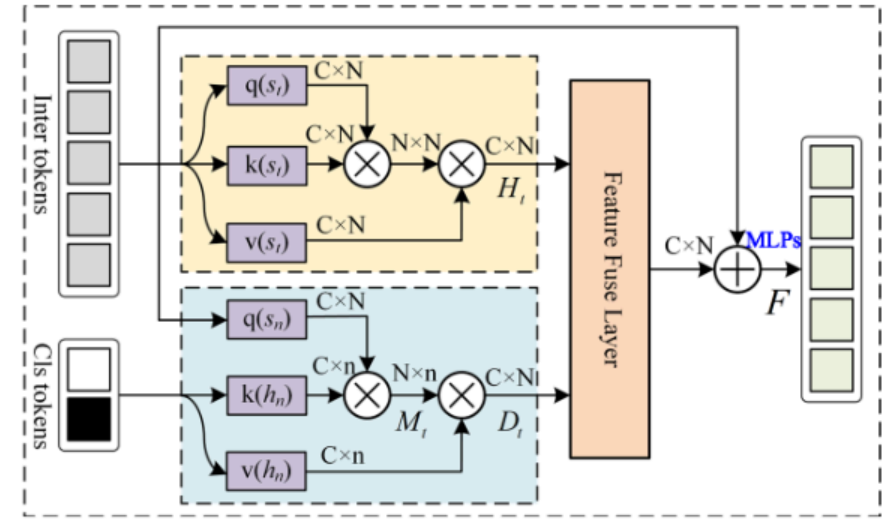


图 3. IAM 图示。首先, 获取 Inter tokens 和 CLS token。接下来, 自我语境和交互式时空信息增强由 (9) – (12) 实现。随后, 特征融合层负责特征聚合并获取 map 以进行后续变化预测。

具体来说, 跨时空信息增强的计算计算如下:

$$M = \text{Softmax}_N \left( \frac{Q(s) \times K(H)}{C} \right) \quad (10)$$

$$Z = \sum_{n=1}^N M \quad (11)$$

$$D = \frac{1}{Z} \sum_{n=1}^N X \otimes v(h) \quad (12)$$

其中  $n$  表示特征像素位置。  $h$  表示 CLS 标记的输入特征,  $k$  表示从  $h$  到键的线性映射,  $v$  表示从  $h$  到值的线性映射。  $M$  表示与位置  $t$  处的像素相对应的相似度。  $Z$  表示使用像素  $t$  计算的  $n$  对相似性的聚合,  $\odot$  是矩阵的 Hadamard 积。  $D$  表示交叉注意力机制的中间特征。经过特征融合层处理后,  $D$  and  $H$  的输出加上输入 Inter 标记, 构成最终的 IAM 特征图, 表示为  $F$ 。

随后,  $F$  最初通过 MLP 层进行处理, 并未采样为  $H \times W$  的大小。最后, 将得到的特征图通过另一个 MLP 层进行处理, 以预测分辨率为  $H \times W \times N$  的变化掩码。

通常, 在 CD 任务中将  $N$  设置为 2。

## III. 时代

### A. 数据说明

在本节中, 我们使用两个流行的 RSI 数据集评估了拟议的 FTAN。

- 1) LEVIR-CD: 该数据集由 637 对组成, 涵盖不同的建筑类型, 分辨率为  $1024 \times 1024$ 。我们通过裁剪原始图像获取了大小为  $256 \times 256$  的样本进行模型训练。在以下实验中, 训练集、验证集和测试集分别包含 7120、1024 和 2048 个样本。
- 2) DSIFN-CD: 该数据集涉及中国城市地区收集的六类卫星 RSI 对, 原始分辨率为  $512 \times 512$ 。同样, 我们将图像裁剪成分辨率为  $256 \times 256$  的样本, 训练集、验证集和测试集分别包括 14 400、1360 和 192 对。

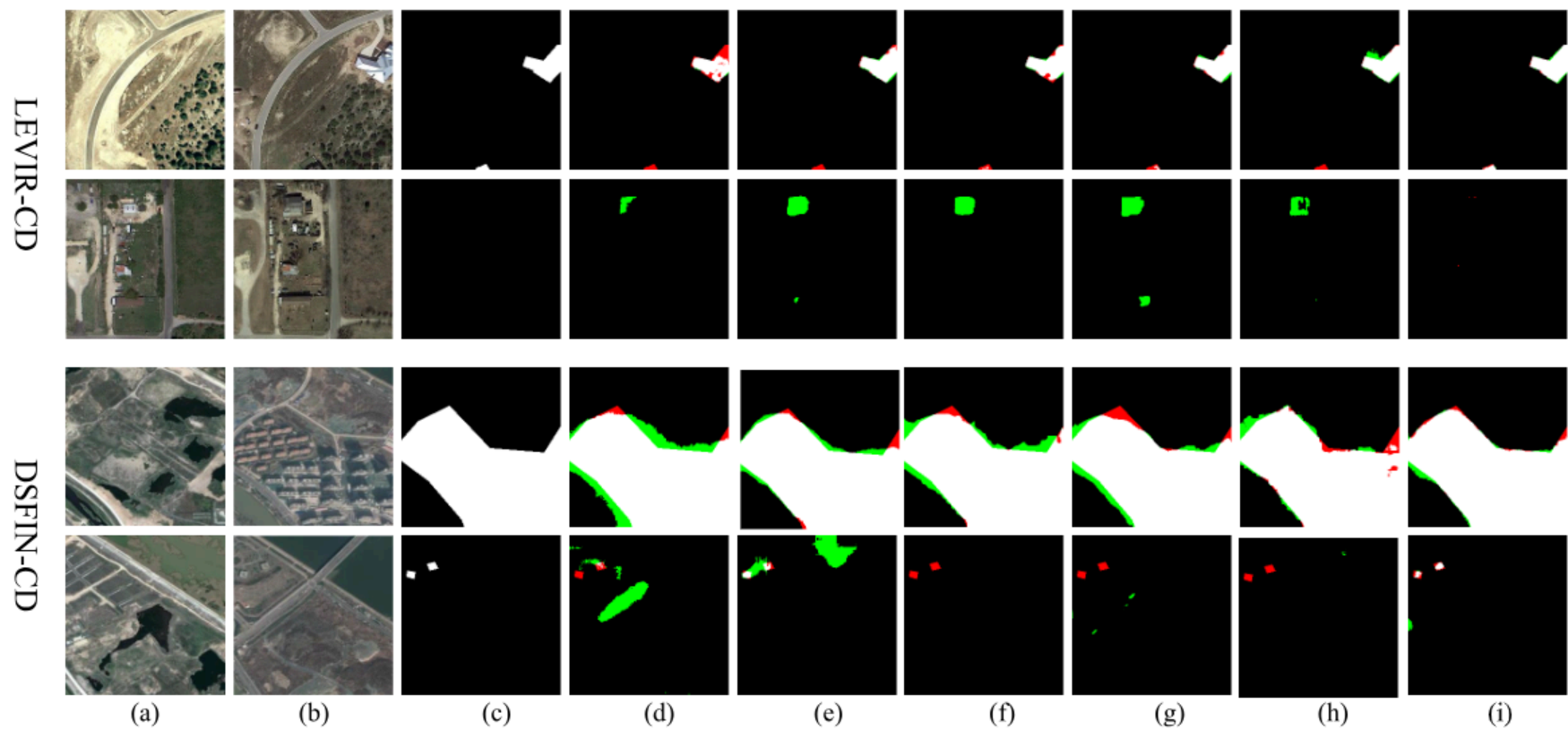


图 4.LEVIR-CD (前两行) 和 DSIFN-CD (后两行) 上不同 CD 方法的比较结果。(a) 前 IMG。(b) IMG 后。(c) 标签。(d) FC-EF。(e) DTCDSCN。(f) 比特。(g) 小 CD。(h) ChangeFormer。(i) FTAN。(绿色表示误检, 红色表示漏检)。

B. 实验设置

所有实验均在 PyTorch 平台的计算机上进行, 并使用 NVIDIA Quadro RTX 8000 GPU 进行训练。在训练阶段, 我们通过随机翻转进行数据增强, 重新缩放作通过在 0.8 到 1.2 之间随机选择的因子调整图像大小, 裁剪、高斯模糊和随机颜色抖动, 以增加训练样本的数量。此外, 在实验执行中采用了交叉熵损失和 AdamW 优化器。此外, 学习率最初设置为 0.0001, 并且批量大小固定为 24。特别是, 所有 CD 方法都没有预训练模型, 以便在以下实验中进行公平比较。

C. 结果和分析

首先, 我们验证性能以证明 FTAN 的优越性。首先, 我们利用了一系列 SOTA 方法与我们提出的方法进行比较。通常, TinyCD 采用孪生 U-Net 架构, 该架构采用低级功能来实现高效的 CD。ChangeFormer 模型构建了一个分层的 Transformer 编码器和 MLP 解码器来生成检测图。表 I 和图 4 报告了 LEVIR-CD 和 DSIFN-CD 上所有 CD 方法的实验结果。可以观察到, FTAN 生成最高的 F1 分数和 IoU。具体来说, LEVIR-CD 数据集上的 F1 分数和 IoU 分别为 90.51% 和 82.78%。而对于 DSIFN-CD 数据集, 值分别为 89.56% 和 81.10%。如图 4 所示, FTAN 的绿色区域和红色区域都是最少的, 这表明错误和漏检率最低。与当前最先进的 ChangeFormer 模型相比, 所提出的新 CD 范式有利于对不同尺度和斑驳区域的变化目标的敏感性。所有结果和分析都证明了我们模型的有效性和稳定性。

接下来, 进行消融研究以验证 FTAN 中两个模块的贡献。单独

表 I 和表 II 显示了消融结果。如表 I 所示, 关于 LEVIR-CD 数据集, 所提出的模型同时采用 MCFA 和 IAM 产生了最佳性能, 例如精度、召回率、F1 和 IoU。使用 IAM 的实现比没有 IAM 的模型更具竞争力, 准确率分别为 91.54% 和 92.12%。尽管使用 MCFA 的实施可以提高检测能力, 但该方法会导致检测能力的提高, 并且同时配备 MCFA 和 IAM 的方法在四个标准上产生最佳性能。具体来说, 精度、召回率、F1 和 IoU 值分别达到 92.41%、88.82%、90.51% 和 82.78%。对于 DSIFN-CD 数据集, 消融研究揭示了类似的趋势, 其中 MCFA 和 IAM 模块的融合显着增强了性能指标。如表 II 所示, 双模块配置实现了 Recall 和 IoU 的提升, 分别达到 89.56% 和 81.10%。结果有力地验证了 MCFA 和 IAM 的结合不仅增强了对数据集中细微差异的检测能力, 而且还显着降低了误报率。简而言之, 正如所观察到的, 对于 LEVIRCD 数据集, MCFA 模块在精度和 IoU 方面提供了更大的优势。对于这两个数据集, 所有评估指标都通过包含 IAM 模块而得到增强, 如表 II 和表 III 的第一行和第三行所示。此外, 我们进一步进行了一项研究, 以评估 FDM-AC 在 MCFA 成分中的疗效。图 5 演示了通过 GradCAM 算法生成的激活图 [16]。可以观察到, 图 5 (d) 中的视觉图显示了 FDM-AC 促进的增强检测精度, 它将激活图与预测结果保持一致。与没有 FDM-AC 的模型相比, 使用 FDM-AC 的方法产生了更有趣的 CD 面积, 并加强了目标和噪声之间的差异。如图所示, 通过有效地放大特征表示, FDM-AC 使 MCFA 能够识别涉及多个信息的细微区别, 从而提高 RSI CD 的准确性。



表 I

AQRWDCD MTDS								
Model	LEVIR- CD				DSIFN-CD			
	Precision	Recall	F1	IoU	Precision	Recall	F1	IoU
FC-EF[13]	89.31	84.81	87.01	77.00	70.16	68.45	69.30	53.02
FC-Siam-Diff[13]	87.89	75.91	81.46	68.72	72.26	49.56	58.80	41.64
FC-Siam-Conc[13]	85.93	80.02	82.87	70.75	55.18	72.95	62.83	45.80
DTCDCN [14]	90.35	85.82	88.03	78.62	80.37	83.88	82.09	69.62
BIT[15]	90.61	88.25	89.41	80.86	81.83	75.88	78.74	64.94
ChangeFormer[10]	91.54	87.04	89.23	80.56	89.40	85.93	87.63	77.98
TinyCD[9]	90.74	<b>89.03</b>	89.88	81.62	77.48	77.60	77.54	63.32
FTAN (ours)	<b>92.41</b>	88.82	<b>90.51</b>	<b>82.78</b>	<b>90.54</b>	<b>88.61</b>	<b>89.56</b>	<b>81.10</b>

表 II

阿斯利韦-CD					
MCFA	IAM	Precision	Recall	F1	IoU
X	X	91.54	87.04	89.23	80.56
✓	X	92.12	87.70	89.85	81.58
X	✓	91.88	88.51	90.17	82.09
✓	✓	<b>92.41</b>	<b>88.82</b>	<b>90.51</b>	<b>82.78</b>

表 III

ASDSIFN-CD					
MCFA	IAM	Precision	Recall	F1	IoU
X	X	89.40	85.93	87.63	77.98
✓	X	88.79	<b>89.01</b>	88.99	80.17
X	✓	89.60	87.92	88.75	79.78
✓	✓	<b>90.54</b>	88.61	<b>89.56</b>	<b>81.10</b>

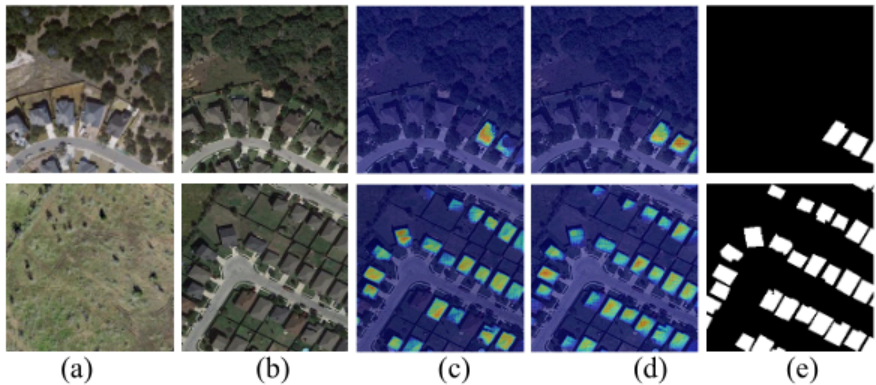


图 5.MCFA 功效的 Grad-CAM 可视化。(a) 前 IMG。(b) IMG 后。(c) 不使用 FDM-AC 的 Grad-CAM 激活 (d) 使用 FDM-AC 的 Grad-CAM 激活。(e) 标签。

IV. C 这封信提出了一个 CD 模型，它利用了空间特征、频率信息和通道间显著性的协作。所提出的 FTAN 网络结合了 MCFA 模型和 IAM 块，有效地融合了多尺度空间特征与频域特征，以提高 CD 精度。值得注意的是，MCFA 专注于通过多尺度卷积注意力提取特征。此外，将 IAM 集成到 FTAN 的解码器中，通过有效地聚合与类别相关的时间数据来改善特征区分。广泛的实验和分析表明，与现有的高级模型相比，性能更胜一筹。未来，我们计划结合多源 RSI 以充分利用互补信息并进一步提高 CD 准确性。

R

[1] M. Gong, J. Zhao, J. Liu, Q. Miao, and L. Jiao, “基于深度神经网络的合成孔径雷达图像变化检测”, IEEE Trans. Neural Netw.学习. Syst., 第 27 卷, 第 1 期, 第 125-138 页, 2016 年 1 月。[2] L. Zhang, M. Lan, J. Zhang, and D. Tao, “用于遥感图像道路分割的阶段性无监督域适应与对抗性自我训练”, IEEE Trans. Geosci.遥感, 第 60 卷, 第 1-13 页, 2021 年。[3] C. Yu, Y. Zhu, M. Song, Y. Wang, and Q. Zhang, “看不见的特征提取: 使用光谱压缩网络进行空间映射扩展以进行高光谱图像分类”, IEEE Trans. Geosci.遥感, 第 62 卷, 2024 年, Art.不 5521915。[4] L. Zhang 和 L. Zhang, “用于遥感数据分析的人工智能: 挑战和机遇回顾”, IEEE Geosci. Remote Sens. Mag., 第 10 卷, 第 2 期, 第 270-294 页, 2022 年 6 月。

[5] C. Yu, M. Xu, Q. Zhang, and X. Lu, “用于高光谱图像分类无监督域自适应的双重干预约束掩码-对抗框架”, IEEE Geosci.Remote Sens. Lett., 第 21 卷, 第 1-5 页, 2024 年。[6] H. Yu, H. Yang, L. Gao, J. 胡, A. Plaza, and B. Zhang, “基于具有光谱相似性过滤的门控光谱-空间-时间注意力网络的高光谱图像变化检测”, IEEE Trans. Geosci.遥感, 第 62 卷, 第 1-15 页, 2024 年。

[7] Y. Wang, X. Chen, E. Zhao, C. Zhao, M. Song, and C. Yu, “用于高光谱目标检测的基于无监督动量对比学习的变压器网络”, IEEE J. Sel. Topics Appl. Earth Observ.遥感, 第 17 卷, 第 9053-9068 页, 2024 年。[8] A. Vaswani 等人, “注意力就是你所需要的”, Proc. Adv. Neural Inf. 过程. 系统, 第 30 卷, 2017 年。

[9] A. Codegoni, G. Lombardi 和 A. Ferrari, “TINYCD: 用于变化检测的(不是如此)深度学习模型”, 神经计算. Appl., 第 35 卷, 第 11 期, 第 8471-8486 页, 2023 年 4 月。[10] W. G. C. Bandara 和 V. M. Patel, “用于变化检测的基于变压器的孪生网络”, IEEE 国际地球科学论文集. 遥感. Symp. (IGARSS), 2022 年 7 月, 第 207-210 页。[11] Z. Chen 等人, “EGDE-Net: 一种建筑变化检测方法用于基于边缘制导和差分增强的高分辨率遥感图像”, ISPRS J. Photogramm.遥感, 第 191 卷, 第 203-222 页, 2022 年 9 月。

[12] S. Fang, K. Li 和 Z. Li, “Changer: 功能交互就是您所需要的用于变化检测”, IEEE Trans. Geosci.遥感, 第 61 卷, 2023 年, Art.不 5610111。[13] R. C. Daudt, B. Le Saux, and A. Boulch, “全卷积孪生网络进行变化检测”, 第 25 届 IEEE 国际会议论文集图像过程. (ICIP), 2018 年 10 月, 第 4063-4067 页。[14] Y. Liu, C. Pang, Z. Zhan, X. Zhang, and X. Yang, “Building change (构建变革) 使用双任务约束的深度孪生卷积网络模型检测遥感图像”, IEEE Geosci. 遥感. Lett., 第 18 卷, 第 5 期, 第 811-815 页, 2021 年 5 月。[15] H. Chen, Z. Qi, and Z. Shi, “遥感图像变化检测”带变压器”, IEEE Trans. Geosci.遥感, 第 60 卷, 第 1-14 页, 2021 年。

[16] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, 和 D. Batra, “Grad-CAM: 通过基于梯度的定位从深度网络进行视觉解释”, IEEE Int. Conf. Comput. 论文集. 可见. (ICCV), 2017 年 10 月, 第 618-626 页。