

面向细粒度的多尺度上下文聚合 CNN-Transformer 网络

农田变化检测

刘梦曦, IEEE 学生委员, 柴卓群, 邓浩军, 刘荣

非农化事件严重威胁当地农业生态系统和全球粮食安全。遥感变化检测(Remote sensing change detection, CD)可以为及时发现和预防此类事件提供有效途径。然而,现有的 CD 方法难以处理高分辨率图像中农田变化的类内差异较大的问题。此外,传统的基于 CNN 的模型受到远程上上下文信息丢失和 deconvlayer 带来的高计算复杂度的困扰。因此,在本文中,我们提出了一种具有多尺度上下文聚合的 CNN-变压器网络(MSCANet),它结合了 CNN 和变压器的优点,以实现高效的农田 CD。在 MSCANet 中,首先利用基于 CNN 的特征提取器捕获分层特征,然后设计基于变压器的 msctet 对上下文信息进行编码和聚合。最后,利用一个包含三个 CNN 分类器的多分支预测头来获取变化图,以增强对深层的监督。此外,针对缺乏细粒度感兴趣农田变化的 CD 数据集,我们还提供了一个新的农田变化检测数据集,该数据集包含 600 对 512×512 双时间图像,空间分辨率为 $0.5-2m$ 。MSCANet 在高分辨率语义 CD 数据集上的 F1 值最高,为 64.67%,在 CLCD 数据集上的 F1 值最高,为 71.29%。

索引词变化检测(CD), 耕地, 深度学习(DL), 遥感, transformer。

我的介绍。

农业生产是世界粮食安全的保障[1]。然而，受到近期快速的

人口的增长和急剧的气候变化，耕地，作为农业活动的基本单位，遭受了许多不利的变化，包括植树造林，挖湖，扩大储备，非法建筑。这些非农业化事件不仅扰乱了当地的农业

稿件收到于 2022 年 3 月 26 日;修订 2022 年 4 月 18 日;审定 2022 年 5 月 18 日。发布日期 2022 年 5 月 23 日;当前版本发布日期 2022 年 6 月 3 日。国家自然科学基金项目 (61976234) 和中山大学中央高校基本科研业务费专项 (22qntd2001) 资助。(通讯作者:刘荣)

作者与中山大学地理与规划学院广东省城市化与地理模拟重点实验室,
广 东 广 州 510275 (e-mail:
liumx23@mail2.sysu.edu.cn; chaizhq@mail2.sysu.edu.cn; denghj5@mail2.sysu.e
du.cn; liurong25@mail. sysu.edu.cn).

文章中的代码和数据集可从以下链接下载
<https://github.com/liumency/CropLand-CD>。

数字对象标识符 10.1109/JSTARS.2022.3177235

这项工作是在知识共享归属 4.0 许可下授权的。更多信息，请参见 <https://creativecommons.org/licenses/by/4.0/>

生态系统,还威胁全球粮食供应[3]。因此,为了获得及时的农田信息,保证农田生产和粮食安全[4],对农田的快速动态变化检测(CD)是极其重要的[5]。

由于农田分布广泛,通过人工田间调查获取农田动态费时费力。随着卫星影像的广泛应用,遥感技术在地形分类[7]、建筑物足迹提取[8]、土地覆盖 CD[9]等方面都成为了一种有效而现实的方法。传统的 CD 方法主要基于多光谱图像,提取丰富的光谱、纹理和结构特征,以实现快速的像元或物体变化结果。如变化向量分析(CVA)[10]、主成分分析[10]、多变量变化检测[12]、[13]因其特征表达简洁、变化提取快速等优点,在 CD 研究中得到了广泛应用。

然而, 由于简单的特征难以满足多样化和高精度的变化提取需求, 因此将基于手工特征工程的机器学习方法应用于 CD 任务[14], [15]。例如, Vries 等人[16]将基于随机森林的后分类和传统的 CVA 方法结合到年度耕地变化制图的更新中。然而, 这些基于 ml 的方法需要事先的专业知识来手动构建和选择特征, 在不同的区域和数据集[17]上泛化性能较低。此外, 由于多光谱图像[18]的空间分辨率有限, 难以获得细粒度的动态结果。

由于人工智能技术和遥感平台的快速发展, CD 研究的重点已经转向深度学习(DL)模型和高分辨率图像(HRIs)[19], [20]。这些深度学习模型基于卷积神经网络(convolutional neural networks, cnn)结构, 通过重构经典的 UNet[22]、[23]、DeepLab[24]、[25]、ResNet[26]、[27]等经典模型, 能够自动学习人力资源 *ris*[21]中的多层次变化信息。虽然 CD 结果容易受到图像之间的季节、辐射和大气扰动的影响, 但最近的 CD 网络中引入了许多新技术, 以更好地感知来自双时间图像的变化, 包括多尺度特征融合[28]、[29]、注意力机制[30]、[31]、循环神经网络[32]–[34]等, 这些技术已被证明可以有效地增强

模型的特征提取能力。然而，这些传统的基于 cnn 的方法仍然面临两个瓶颈:一是特征编码和解码过程中的信息丢失问题，二是随着层数和数据量的增加，其计算量呈指数级增长的问题。

最近，最初为自然语言处理任务[35]而设计的 transformer 在计算机视觉领域也受到了广泛关注，如图像分类[36]、分割[37]、物体识别[38]和图像描述[39]等。与 CNN 相比，transformer 已经显示出强大的建模全局依赖关系的能力，以减轻远程信息的损失[40]。在这些工作的启发下，Chen 等人[41]将 transformer 引入到 CD 任务中，并实现了一个双时态图像 transformer (BIT)，它在差分基 CD 框架中将输入图像编码为上下文丰富的语义标记。

尽管现有方法在遥感 CD 方面取得了巨大的成就，但实现细粒度农田 CD 仍然存在挑战。深度学习模型的性能在很大程度上依赖于训练数据集，许多之前的工作已经为 CD 提供了良好的标注数据集，如用于语义 CD 的高分辨率语义变化检测数据集(HRSCD)[42]和 SECOND[43]，用于二进制 CD 的 SYSU-CD[44]和 SVCD[45]，以及用于构建 CD 的 BCDD[46]和 levird -CD[31]等。到目前为止，还没有专门针对农田变化的数据集，这极大地限制了农田 CD 模型的发展和应用。因此，如何高效有效地对双时态图像之间的多尺度信息进行建模是快速农田 CD 任务中的迫切需求。

为了解决上述问题，本文提出了一个多尺度上下文聚合网络(MSCANet)和一个高分辨率耕地变化检测数据集(CLCD)。MSCANet 首先采用 CNN 骨干网从双时图像中捕获多尺度特征;然后利用多尺度上下文聚合器(MSCA)通过 transformer 架构对丰富的上下文信息进行建模和聚合;最后，应用多分支预测头(multibranchpredictionhead, MBPH)获取变化映射，进一步增强对隐藏层的特征提取和学习。MSCANet 基于 CNN-transformer 结构构建，充分结合了 CNN 和 transformer 的优点，满足了快速、准确的农田 CD 的迫切需求。CLCD 由 600 对标注了各种农田变化的双时相图像组成，可为基于 dl 的模型进行农田 CD 任务提供基准。本文的贡献可以归纳为三点。

- 1)提出了一种基于 CNN-transformer 混合架构的农田 cd MSCANet，其中 hamscet 用于多尺度上下文信息编码，MBPH 用于深度特征学习。
- 2)提供了一个高分辨率的 CLCD，包含 600 对空间分辨率为 0.5-2 m 的 512 ×512 图像。
- 3)与 6 个最先进的(SOTA)进行对比实验

CD 模型在 HRSCD、[42]和 CLCD 上的实验结果说明了这一点

所提出的 MSCANet 可以获得最高的 F1 分数，分别为 64.67%和 71.29%。

文章其余部分组织如下。第二节揭示了方法论的细节结构，而第三节给出了实验设置。实验结果将在第四节中进行演示和分析，消融研究和模型效率将在第五节中讨论。最后，第六节对本文进行总结。

2 方法

如图 1 所示，MSCANet 包含三个部分:CNN 特征提取器、MSCA和 MBPH。各部分的详细信息将在下文介绍。

A.特征提取器

MSCANet 采用 CNN 骨干网作为特征提取器，该特征提取器是在 ResNet-18[47]的基础上改进的，去掉了初始的全连接层。因此，特征提取器包含一个 7 ×7 的卷积层，以及四个残差块(ResBlocks)。第一个卷积层的步幅为 2，用于提取半大小的浅层特征。然后进一步使用 stride 2 的 3×3 max-pool 层来捕获原始图像四分之一大小的特征，目的是过滤重要特征并减少参数数量。

每个 ResBlock 包含两个 3 ×3 卷积层，一个批处理归一化层[48]和一个整流线性单元(ReLU)函数[49]。在输入到 ReLU 层之前，通过元素加法将特征与原始输入特征融合。由于 ResBlock-1 和 ResBlock-4 中的第一个卷积层采用的步长为 1，因此特征输入前后的大小保持不变，而 ResBlock-2 和 ResBlock-3 中的第一个卷积层采用的步长为 2，因此输出的特征大小减半。最后，ResBlock-4 的输出特征大小为原始输入图像的 1/16。每个 ResBlock 的输出通道分别为 64、128、256、512。

为了获得多尺度农田信息，ResBlock-1、2、4 的多尺度输出将被转发到后续模块中。在此之前，将对选定的特征应用一个 3 ×3和一个 1 ×1的卷积层，将其通道大小统一为 32。

B.多尺度上下文聚合器

为了进一步对特征提取器中的多尺度信息进行建模和融合，在 MSCANet 中设计了一个 MSCA。MSCA 使用基于变压器架构的三个令牌编码器和三个令牌解码器来捕获和聚合来自三个不同大小特征的多尺度上下文信息。

1) Token 编码器:Token 编码器旨在通过空间注意力模块和 transformer 模块编码特征的全局上下文信息。考虑到计算和存储上的限制，首先采用空间注意力模块将输入特征转换为目标大小的三维 token 嵌入，供后续的 transformer 模块使用。根据图 2(a)，给定一个输入特征，简称为

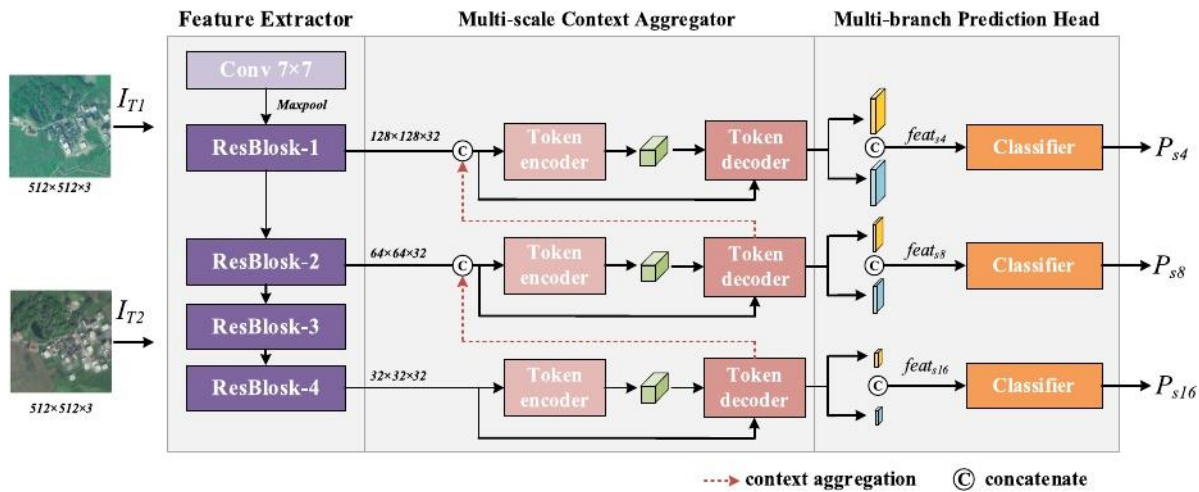


图 1 所示。拟议的 MSCANet 概述。

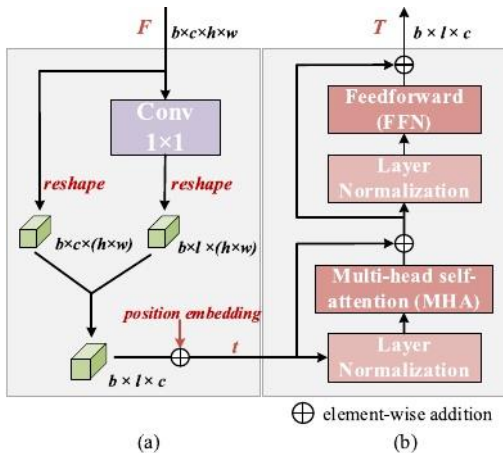


图 2 所示。token 编码器的架构。(a)空间注意力模块;(b) Transformer 编码器。

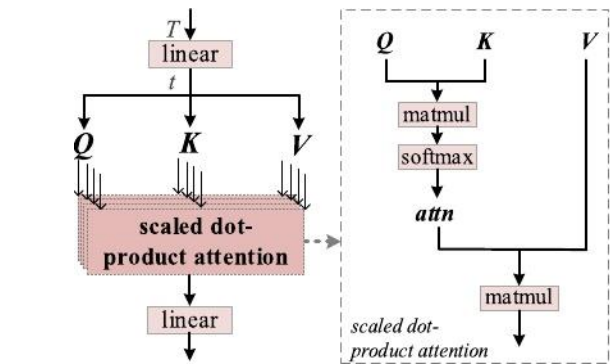


图 3 所示。MHA 块的体系结构。

$F \in \mathbb{R}^{b \times c \times h \times w}$, 空间注意力模块采用 1×1 卷积层获得中间特征, 简称 F' 。那么, F 和 F' 会被重塑成

3D tokens, referred as $f \in \mathbb{R}^{b \times c \times (h \times w)}$ and $f' \in \mathbb{R}^{b \times l \times (h \times w)}$,

分别。最后, f 和 f' 会通过 einsum 操作, 可表示为

$$\begin{matrix} TBL \\ \Rightarrow fbl(hw) \end{matrix} \quad \begin{matrix} 流化 \\ 块块块 \end{matrix} \quad (1)$$

其中 b 、 c 、 h 、 w 分别表示输入特征 F 的批量大小、通道数、高度、宽度; L 是 token 长度, 在模型中设置为 4。

此后, transformer 编码器被用来对 token 中的上下文信息进行建模, 其中一组可训练参数第一个元素被明智地添加到 token t 中以进行位置嵌入 (PE)。变器编码器具有标准 transformer 结构[40], 其中包含一个 MHA 块和一个前馈 (FFN) 块, 在每个块之前应用层归一化 (LN) 层。

MHA 块的架构如图 3 所示。MHA 首先将 t 扩展成一个新的嵌入 t' 通过线性层,

可以表示为

$$t' = tW^I, t' \in \mathbb{R}^{b \times l \times (n \times d \times 3)} \quad (2)$$

其中, w_i 为线性层的权值, n 为 MHA 的头数, d 为后续张量的维数。 N 和 d 分别设为 8 和 64。

那么, 嵌入 t' 将转发给 MHA 的不同负责人。各头之间不共享参数。每个头包含两个步骤: 线性变换和尺度点积注意 (SDPA)。应用了三层线性变换来映射 t' 转化为 query ($Q \in \mathbb{R}^{b \times n \times l \times d}$)、key ($K \in \mathbb{R}^{b \times n \times l \times d}$)、value ($V \in \mathbb{R}^{b \times n \times l \times d}$), 可以表示为

$$Q, K, V = t'W^Q, t'W^K, t'W^V \quad (3)$$

其中 W_Q , W_K , 和 W_V 分别表示映射 Q , K 和 V 的线性层的权重。

然后, 在 SDPA 中, 通过点积运算和 Softmax 激活计算 Q 和 K 之间的相关性, 生成注意图, 作为权重

SDPA 中的过程可以表示为

$$\text{SDPA}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V. \tag{4}$$

在施加线性层之前，将每个头部的输出连接在一起，然后我们将得到 MHA 的最终输出，可以用以下公式表示：

$$\text{head}_i = \text{SDPA}\left(t'W_i^Q, t'W_i^K, t'W_i^V\right), i \in (0, n] \tag{5}$$

$$\text{MHA}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_n)W^O \tag{6}$$

其中， w_{iQ} 、 w_{iK} 、 w_{iV} 分别表示第 i 个头部的映射 Q 、 K 、 V 的线性层的权重， w_o 是 MHA 中最后一个线性层的权值。

在 FFN 中，使用两个线性层和一个高斯误差线性单元激活[50]来进一步变换 MHA 的学习令牌。

2)令牌解码器:令牌解码器接收两个输入，分别是来自特征提取器的卷积特征 F 和来自令牌编码器的令牌嵌入，记为 T 。在令牌解码器中，首先为 PE 在 F 中添加一个可训练参数，然后将 F 和 T 输入到变压器解码器中，目的是将令牌嵌入重新投影到像素空间，增强 F 中的上下文信息。

在变压器解码器中，在使用 MHA 模块之前，对 F 和 T 应用权重共享 LN 层。MHA 与令牌编码器中的 MHA 类似。区别主要在于 Query 是从 F 映射的，而 Key 和 Value 是从 t 映射的。注意，这里没有应用掩码机制。

在不同尺度的 token 编解码过程中，引入了上下文聚合连接(见图 1 虚线)，将较高层次的 transformer 特征聚合为较低层次的卷积特征。

C.多分支预测头

为了更好地利用已有的多尺度特征，MBPH 采用三种架构相同的基于 cnn 的分类器生成变化结果，监督深层特征学习，帮助提取更多有用的 CD 特征。

通过特征提取器和 MHCA 获得输入图像的多尺度特征后，相同尺度的特征将通过级联的方式融合在一起并插值到原始图像大小。然后应用分类器从多尺度特征中获得三个变化图。每个分类器包含两个 3×3 卷积层。

MBPH 输出多级预测图，称为 P_{s4} 、 P_{s8} 和 P_{s16} ，用于对 MSCANet 进行深入监督，这为模型在多级捕获更有效的特征以进行后续预测提供辅助帮助。在训练过程中，MSCANet 将通过三个变化图之间的交叉熵损失和地面真值 Y 进行优化。交叉熵损失的公式化可以表示为

$$L_{CE}(P, Y) = -[Y\log P + 1 - Y\log(1 - P)]. \tag{7}$$

因此，MSCANet 的总损耗可以表示为

$$L = L_{CE}(P_{s4}, Y) + L_{CE}(P_{s8}, Y) + L_{CE}(P_{s16}, Y). \tag{8}$$

从目标函数可以看出，对 MSCANet 中的隐藏层进行了深度监督，生成了更多的可区分特征。而在测试过程中，只使用 P_{s4} 来获得最终的更改结果。

3 实验设置

答:数据集

1)高分辨率语义变化检测数据集:HRSCD[42]是一个语义 CD 数据集，包含 291 对大小为 10000×10000 的 0.5 m RGB 航空图像，每张图像都有相应的土地覆盖信息，包括 5 种类型的人工地表、农业区、森林、湿地和水域。所有图像均采集自法国雷恩和卡昂的城市和农村地区。为了获得细粒度的农田兴趣变化，我们重新对双时间图像的原始标签进行分类，将“农业区”类别划分为 1，其余类别划分为 0。然后，通过比较重新分类的双时态标签，可以获得农田的变化标注。

为方便模型训练，我们以不重叠的行为对原始图像进行裁剪，得到农田 CD 的 4398 对 512×512 样本，这些样本以 6:2:2 的比例进行分离训练、验证和测试。in HRSCD 的示例如图 4 所示。

2)耕地变化检测:CLCD 数据集由 600 对耕地变化样本图像组成，其中 320 对用于训练，120 对用于验证，120 对用于测试。CLCD 中的双时间图像分别由中国广东省高分 2 号于 2017 年和 2019 年采集，空间分辨率范围为 0.5 ~ 2 m。每组样本由两幅 512×512 和一幅 a 的图像组成

对应的耕地变化二值标签。如图 5 所示，CLCD 中标注的主要变化类型包括建筑物、道路、湖泊和裸地等。

B.比较方法

在我们的实验中，双颞 CD 采用 SOTA 方法进行比较。

- 1) FC-EF[23]是一种基于 unet 的 CD 方法，它接收双时相图像的拼接作为输入，将它们作为单独的通道。
- 2) FC-Siam-conc[23]是 FC-EF 的一种变体，采用共享权值的暹罗结构获取多级特征，并将其连接起来合并变化信息。
- 3) DTCDSCN[51]是一种基于暹罗 fcn 的方法，具有注意机制，同时考虑空间和通道的变化信息，以提取更多的上下文特征。
- 4)多方向融合路径网络(MFPNet)[29]是一种多方向特征融合方法，它利用了具有多方向信息的 multiscalefusionnetwork

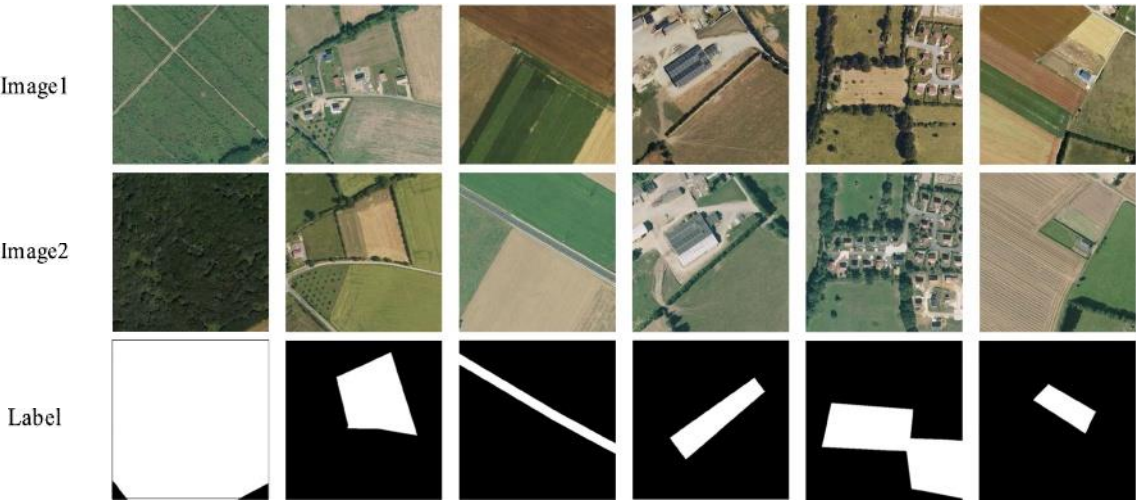


图 4 HRSCD 数据集中大小为 512 ×512 的示例。



图 5 CLCD 数据集中大小为 512 ×512 的示例。CLCD 中标注的主要变化类型包括建筑物、道路、湖泊和裸地。

- flowformingdatapropagation 更容易，同时突出重要的特征。
- 5)深度监督图像融合网络(DSIFN)[28]使用差异判别网络进行 CD，通过注意力机制将多级特征与图像差异图融合。
- 6) BiT[41]是一种基于变压器的特征融合方法，它将 Siamese 标记器和变压器编码器-解码器结构集成到公共 CD 网络中，从而能够在全局特征空间中捕获更有意义和有效的上下文概念。

C.参数和度量标准

所提出的模型和所有涉及的实验在 PyTorch 中实现。使用 Adam 优化器对所有模型进行训练，⁴批大小为 8，学习率为 1。训练过程持续 100 epoch，而数据增强策略则随机应用于训练集以避免

过度拟合，包括垂直和水平翻转，以及随机旋转。

选择四种常见的指标，准确率(Pre)、召回率(Rec)、F1-score 和交并比(IoU)进行准确率评估。它们可以被定义为:

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

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

$$F1 = \frac{2Pre \times Rec}{Pre + Rec}$$
 (11)

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

其中 TP、TN、FP 和 FN 分别表示真阳性、真阴性、假阳性和假阴性。

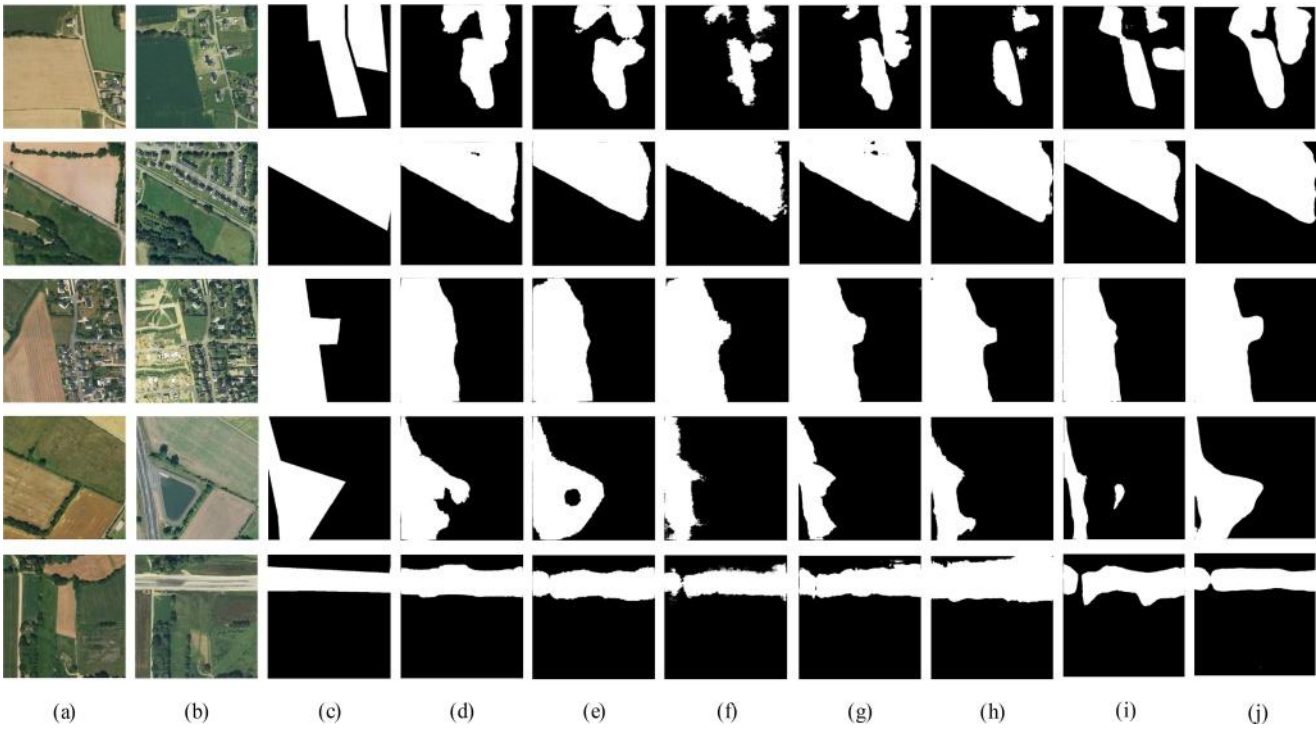


图 6 所示。HRSCD 数据集实验结果可视化。(一)Image1。(b) Image2。(c)标签。(d) FC-EF。FC-Siam-conc (e)。DTCDSCN (f)。(g) BiT。MFPNet (h)。(我)DSIFN。MSCANet (j)。

表我

HRSCD 的实验结果

Method	Pre(%)	Rec(%)	F1(%)	IoU(%)
FC-EF	72.75	50.30	59.48	42.33
FC-Siam-conc	72.23	47.53	57.34	40.19
DTCDSCN	75.79	48.83	59.39	42.24
BiT	71.30	52.23	60.30	43.16
MFPNet	76.42	54.98	63.95	47.01
DSIFN	77.00	54.27	63.66	46.70
MSCANet	70.17	59.97	64.67	47.79

表 2CLCD 的实验结果

Method	Pre(%)	Rec(%)	F1(%)	IoU(%)
FC-EF	71.70	47.60	57.22	40.07
FC-Siam-conc	73.27	52.91	61.45	44.35
DTCDSCN	54.49	66.23	59.79	42.64
BiT	61.42	62.75	62.08	45.01
MFPNet	83.20	60.74	70.22	54.11
DSIFN	79.07	63.79	70.61	54.58
MSCANet	75.36	67.64	71.29	55.39

四、结果与分析

A. HRSCD 实验

HRSCD 的 Pre、Rec、F1 和 IoU 结果见表 1。FC-Siam-conc 的 F1 和 IoU 是所有方法中最低的，分别为 57.34%和 40.19%，而完全卷积-早期融合(FC-EF)的 F1 和 IoU 略高，分别为 59.48%和 42.33%。双任务约束深度孪生卷积网络(DTCDSCN)的性能介于 fc - siam -conc 和 fc - ef 之间，f1of59.48%，其次是 BiT，获得了 60.30%的 F1。总的来说，MFPNet 和 DSIFN 的性能明显优于上述模型，F1 分别为 63.95%和 63.66%。MSCANet 的最优召回率、F1 和 IoU 值分别为 59.97%、64.67%和 47.79%，分别比排名第二的 MFPNet 高 4.99%、0.72%和 0.78%。

图 6 为不同方法在 HRSCD 数据集上不同场景下的实验结果。在农田变化为人工地表、裸地和道路方面，它们是

外观差异明显，大多数模型都能取得相对较好的识别效果。从图 6 的第 3 行可以看出，我们提出的模型在很多方法失败的情况下，可以很好地提取农田到草地的变化。此外，对于挖湖的变化(见图 6 第 4 行)，由于相关样本数量相对较少，大多数方法的检测结果相当有限。在这种情况下，MSCANet 仍然可以完全识别这些变化。

B. CLCD 上的实验

所有方法对 CLCD 的定量结果见表 II。与 HRSCD 的结果不同，带有 Siamese 编码器和特征拼接的 FC-Siam-conc 比 FC-EF 和 DTCDSCN 效果更好，F1 为 61.45%。下图为 BiT，显示了变压器结构比传统 UNet 模型的先进性。MFPNet 和 DSIFN 在 CLCD 上的性能仍然很亮眼，将 CLCD 上的 F1 提升到了 70%。这可以归因于 MFPNet 和 DSIFN 中使用的多尺度特征融合策略，而类内尺度

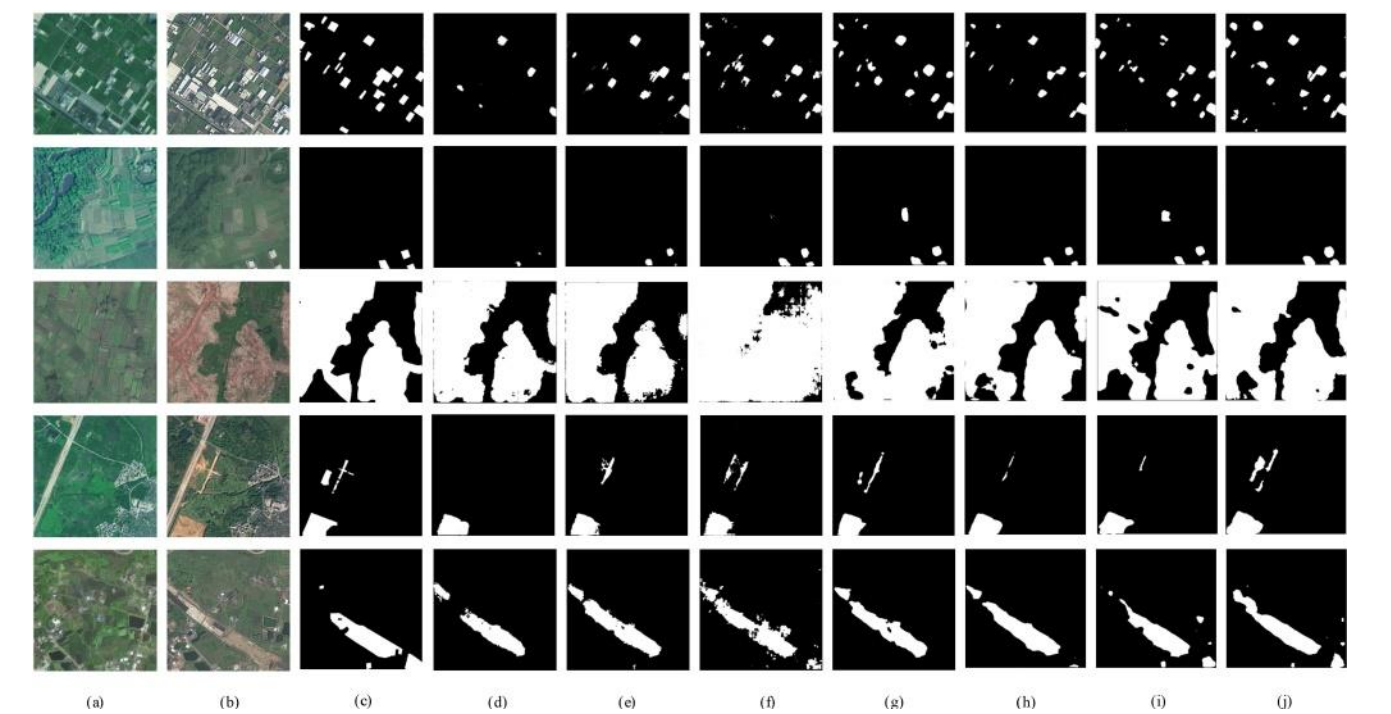


图 7 所示。在 CLCD 数据集上的实验结果可视化。(一)Image1。(b)图像二。(c)标签。(d) FC-EF。(e) FC-Siam-conc。(f) DTCDSCN。(g) BiT。(h) MFPNet。(i) DSIFN。(j) MSCANet。

CLCD 的差异远大于 HRSCD。其中，MSCANet 在 Rec、F1 和 IOU 中效果最好，分别达到 67.64%、71.29%和 55.39%，分别比 DSIFN 高 1.41%、0.68%和 0.81%。

在 CLCD 上的可视化对比如图 7 所示。与 FC-EF 相比，FC-Siam-conc 可以更准确地识别变化区域。DTCDSCN 的变化结果比较碎片化，并且由于光照和物候的差异存在严重的误分类，这与 DTCDSCN 的高查全率和低查全率相呼应，如表 2 所示。由于采用多尺度特征融合策略，MFPNet 和 DSIFN 可以很好地处理各种尺度的农田 CD。尽管如此，在 Transformer 结构编码语义上下文信息的帮助下，BiT 在复杂场景(如图 7 第 1 行)的农田 CD 上表现更好。总体而言，我们的 MSCANet 优于所有比较方法，不仅在大尺度变化的边缘保存能力更强，而且在小尺度变化(如田间道路和建筑物)的检测更完整，这与它的最高召回率一致，如表 2 所示。

诉讨论

A. 消融研究

在本节中，我们对 CLCD 进行消融研究，进一步验证 MSCA 和 MBPH 在 MSCANet 中整合的意义。“基础”模型是用于比较的基本模型，没有任何花招。“+MSCA”表示带 MSCA 的“基础”模型，“MBPH”表示带 MBPH 的“基础”模型。消融研究结果见表三。与 F1 值为 68.71%的“基础”模型相比，F1 值为

表三 CLCD消融研究

Method	Pre(%)	Rec(%)	F1(%)	IoU(%)
Base	70.34	67.15	68.71	52.15
+ MSCA	66.28	72.71	69.35	53.08
+ MBPH	73.67	68.42	70.95	54.98
MSCANet	75.36	67.64	71.29	55.39

“+MSCA”模型和“+MBPH”模型分别提高了 0.64%和 2.24%，初步证明了 MSCA 和 MBPH 模型的有效性。“+MSCA”的召回率最高，为 72.71%，也就是说，MSCA 的加入有利于减少 CD 中的遗漏，这对农田 CD 任务非常重要。MSCANet 在烧蚀实验中取得了最好的效果，充分表明了 MSCA 与 MBPH 相结合的可行性。

图 8 提供了消融结果的可视化对比。从示例结果可以看出，“+MSCA”的 CD 结果的边缘更接近原始标签，尽管有一些伪变化。这表明 MSCA 模块可以有效地对特征之间的多尺度上下文信息进行编码和聚合，从而提高结果的语义表示。与“base”模型的结果相比，“+MBPH”模型的 CD 结果虚警率更低。这表明，MBPH 模块通过监督深层隐藏层的学习，有助于提取更多的判别特征，减少伪变化。毫无疑问，MSCANet 结合了 MSCA 和 MBPH 的优点，在边界提取和减少误报方面都具有优势。

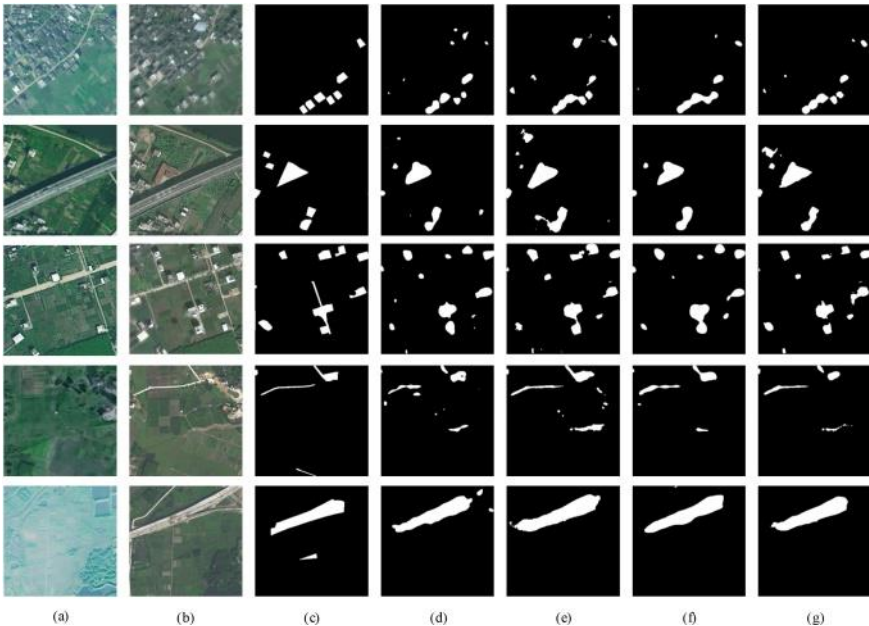


图 8 对 CLCD数据集消融研究的可视化。(一)Image1。(b)图像二。(c)标签。(d)基地。(e) +MSCA。(f) +MBPH。(g) MSCANet。

表 4

不同方法的模型效率

Method	FLOPs(G)	Params(M)
FC-EF	14.29	1.35
FC-Siam-conc	21.29	1.55
DTCDSCN	52.83	31.26
BiT	42.37	3.49
MFPNet	514.93	85.97
DSIFN	329.06	50.443
MSCANet	59.08	16.42

B. 模型效率

为了深入了解实际应用中的不同 CD 模型，我们采用两个指标，浮点运算(FLOPs)和参数数(参数)，进一步比较所有比较方法的模型效率。FLOPs 通过计算乘法和加法运算的次数来衡量模型的计算复杂度，单位为 109(G)。参数为

模型训练过程中需要学习的参数数量，对应于模型的空间复杂度，单位为 106(M)。

给定两个双时输入大小为 $1 \times 3 \times 512 \times 512$ ，所有方法的 FLOPs 和 Params 如表 4 所示。结合前面的分析可以看出，在所有模型中，FC-EF、FC-Siam-conc 的 FLOPs 和 Params 最低。然而，由于使用了复杂的多尺度特征融合策略，MFPNet 和 DSIFN 在 HRSCD 和 CLCD 数据集上都具有优异的性能，具有最高的 FLOPs 和 Params。凭借 cnn -变压器混合架构的优势，MSCANet 可以在相对较低的 FLOPs 和 Params 下实现最先进的 CD 性能，反映了其在快速 CD 应用中的可行性和潜力。

六. 结论

本文提出了一种用于农田 CD 的 MSCANet 和一种新的高分辨率数据集(CLCD)。MSCANet 采用 cnn -变压器结构，其中采用预训练的 ResNet-18 提取分层特征。然后，设计了基于变压器的 MSCA 模块，对多尺度特征中的上下文信息进行编码和解码，并利用上下文聚合连接实现不同层次的特征融合和聚合。最后，使用 MBPH 来帮助增强特征学习并捕获更多有用的特征。

HRSCD 和 CLCD 的实验证明了所提出的 MSCANet 和 CLCD 在农田 CD 上的可行性。CLCD 的消融研究进一步验证了 MSCA 和 MBPH 集成的有效性。更具体地说，MSCA 有助于获得变化对象在边缘和形态方面的语义属性，而 MBPH 可以减少结果中的伪变化。通过对 offlops 和 Params 的比较，进一步证明了 MSCANet 在空间和计算复杂度方面的优势。这些结果充分证明了 MSCANet 在高效农田 CD 中的能力。

参考文献

[1]张志强, 《中国粮食安全问题研究进展》, 《科学》, 2010 年第 4 期, 第 1 - 4 页。
[2]张晓明, 张晓明, “气候变化对全球粮食安全的影响”, 《粮食安全》, 第 10 卷, 第 7 期。1、2021 年, 艺术。不。e261。
[3]李志强, “基于 gis 的全球耕地数据分析”, 《农业科学》, 2011。粮食安全。阿格利司。经济政策。环境。, vol. 4, pp. 37-45, 2015。
[4]陈志强, “农业可持续集约化:前提与政策”, 《科学》, 2013 年第 3 期, 第 33-34 页。

[5]张晓明, 张晓明, 张晓明, “基于 gis 的土地利用遥感影像研究”, *遥感环境科学*., vol. 156, pp. 490-499, 2015.

[6]史文, 张明, 张明, 陈顺生, 詹志强, “基于人工智能的变化检测技术研究进展与挑战”, *《遥感》*, 第 12 卷, 第 1 期。2020 年 5 月 10 日, 《艺术》。不。1688.

[7]张建军, 张建军, 张建军, “基于地形特征的机载激光雷达地形分类方法研究”, *遥感学报. 遥感技术*, 第 77 卷, 第 21-30 页, 2013.

[8]刘鹏等, “基于空间残差初始卷积神经网络的建筑足迹提取”, *《遥感》*, vol. 11, no. 11. 7, 第 830 页, 2019 年 4 月。

[9]胡明, 吴春春, 张明, 杜斌, “基于自编码器的高光谱异常变化检测”, *电子学报*, 2011,31(4):444 - 444。主题:。地球 Observ. 遥感, 第 14 卷, 第 3750-3762 页, 2021。

[10]李建军, 王小明, 王小明, “基于多光谱分析的土地覆被监测方法研究”, *Int. 《遥感》*, 第 19 卷, 第 411-426 页, 1998。

[11]张晓明, 张晓明, 张晓明, “基于遥感数据的土地覆被变化分析”, *《遥感技术》*., vol. 10, pp. 175-184, 1980。

[12]王小明, 王小明, 王小明, 等。基于多光谱图像的多尺度变化检测方法研究[j]. *遥感技术*., 第 64 卷, 第 1-19 页, 1998 年。

[13]张晓明, 张晓明, 张晓明, “基于尺度空间滤波的城市变化监测”, *Int. J.阿普尔.地球 Observ. 《地理信息》*, vol. 15, pp. 38-48, 2012。

[14]张晓明, 张晓明, 张晓明, 等。基于 SVM 的遥感突变检测方法[j]. *地理信息学报*, 2014, (4):448 - 448。遥感, 左., 第六卷, 第。3, 第 606-610 页, 2009 年 7 月。

[15]王晓东, “基于随机森林分类器的土地覆盖地图快速更新”, *《遥感技术》*, vol.8, no. 15. 2016 年 10 月 11 日, Art.:不。888.

[16] B. De Vries, M. Dec uperer, J. Verbesselt, A. Zeileis, M. Herold, S. Joseph, “基于 landsat 时间序列的热带森林扰动-再生动态监测”, *遥感环境*., vol. 169, pp. 320-334, 2015。

[17] B. De Vries, M. Dec uperer, J. Verbesselt, A. Zeileis, M. Herold, S. Joseph, “基于 Landsat 时间序列的热带森林扰动-再生动态监测”, *遥感环境*., vol. 169, pp. 320-334, 2015。

[18]高峰等, “基于 Landsat 和 MODIS 影像融合的农田尺度作物进展制图”, *《遥感环境》*., vol. 188, pp. 9-25, 2017。

[19]李志勇, 刘新, J. A. Benediktsson, N. Falco, “基于高分辨率光学图像的土地覆盖变化检测技术研究进展”, *电子工程学报格奥西. 遥感杂志*, 第 10 卷, 第 10 期。1, 第 44-63 页, 2022 年 3 月。

[20]谭凯, 肖军, 王晓霞, 王新, 梁晓霞, 杜鹏, “基于多分类器和光谱空间特征的高分辨率遥感图像变化自动检测”, *电子工程学报*, 。主题:。地球 Observ. 遥感技术, 第 9 卷, 第 9 期。8, pp. 3439-3451, 2016 年 8 月。

[21]刘志强, 刘志强, “深度学习”, *《自然》*, 第 521 卷, 第 436-444 页, 2015。

[22]彭大东, 张磊, 关辉, “基于改进 UNet++ 的高分辨率卫星图像端到端变化检测”, *遥感*, vol. 11, 2019。

[23]张晓明, 李晓明, 李晓明, “基于卷积神经网络的城市变化特征分析”, *地球科学进展, 国际版. 格奥西. 遥感*., 2018, pp. 2115-2118。

[24]李志刚, “基于深度扩展学习网络的遥感图像语义分割方法研究”, *中文信息学报*, 2016。列托人., vol. 51, pp. 2355-2377, 2020。

[25]王新等, “基于端到端图像分割的高分辨率遥感图像变化检测”, *国际版. J.阿普尔.地球 Observ. Geoinf.*, 第 104 卷, 2021 年, 艺术。不。102582.

[26]柯强, 张磊, “基于层次分割注意力的跨连体变更检测网络”, *电子工程学报*, 2011,44(4):448 - 448。主题:。地球 Observ. 遥感技术, vol. 14, pp. 9987-10002, 2021。

[27] Venugopal, “基于样本选择的扩展网络学习的遥感图像变化检测”, *遥感学报*., vol. 20, pp. 1-22, 2019。

[28]张志刚, “一种基于深度监督的遥感图像融合方法”, *遥感学报*, 2014, (4):448 - 448。遥感, vol. 166, pp. 183 - 200,2020。

[29]徐建军, 罗翀, 陈辉, 魏生, 罗翀, “基于多方向自适应特征融合和感知相似性的遥感变化检测”, *遥感*, vol. 13, no. 29. 2021 年 8 月 15 日, 艺术。不。3053.

[30]刘新, 施强, 何德明, 刘新, 张磊, “基于多分辨率的图像变化检测方法”, *电子工程学报*, 2014。格奥西. 遥感, vol. 60, 2022, Art. 不。4403718.

[31]陈辉, 施强, “基于时空关注的遥感图像变化检测方法”, *遥感*, vol. 12, no. 31. 2020 年 5 月 10 日, 《艺术》。不。1662.

[32]王小明, 王小明, 王小明, “基于多尺度神经网络的城市变化监测方法研究”, *《IEEE》, 进程. Geosci. 遥感*., 2019, pp. 214-217。

[33]朱晓霞, 朱晓霞, 牟磊, “基于卷积神经网络的多光谱图像变化检测方法”, *《IEEE Trans》. Geosci. 《遥感技术》*, 第 57 卷, 第 57 期。2, 第 924-935 页, 2019 年 2 月。

[34]陈达, 吴春春, 杜斌, 张翀, 王磊, “基于深度连体卷积多层递归神经网络的多源 VHR 图像变化检测”, *IEEE Trans. Geosci. 《遥感技术》*, 第 58 卷, 第 5 期。4, 第 2848-2864 页, 2020 年 4 月。

[35]王志强, “注意力是你所需要的”, *《进程》, 《神经网络》. 系统. , 为. Conf.神经网络. Process. 系统.*, I Guyon 等人, Eds. 美国加利福尼亚州长滩:Curran Associates, 2017, 页 5998-6008。

[36]刘涛, 张翀, 白斌, 张翀, “基于交叉部分学习的细粒度图像分类”, *IEEE Trans. 图像的过程*., vol. 31, pp. 748-758, 2021。

[37]王小明, 王小明, “一种基于网络的各向异性三维医学图像分割方法”, *中国医学工程学报(自然科学版)*., 2021, pp. 8857-8861。

[38]陈达, 谢宏华, 刘涛, “基于自适应图像变换的单次目标检测”, *IEEE/CVF Conf, 进程. Vis.模式识别*., 2021, 第 12247-12256 页。

[39]张晓明, 张晓明, 张晓明, “基于网格存储的图像处理方法”, *电子工程学报/CVF 会议, 计算, 程序. Vis.模式识别*., 2020, pp. 10578-10587。

[40]张晓明, 张晓明, 张晓明, “基于视觉变换的遥感图像分类”, *《遥感技术》*, vol. 13, 2013。

[41]陈辉, 祁志强, 施强, “基于变压器的遥感图像变化检测”, *电子工程学报. 格奥西. 《遥感》*, 第 60 卷, 第 1-14 页, 2022 年, Art. 不。5607514.

[42]张晓明, 张晓明, 张晓明, “基于多任务学习的语义变化检测”, *计算. Vis. Image Understanding*, vol. 187, 2019, Art. 不。102783.

[43]杨锴, “航空图像语义变化检测的非对称连体网络”, *《电子工程学报》. 格奥西. 《遥感》*, 第 60 卷, 第 1-18 页; 2022 年, 艺术。不。5609818.

[44]施强, 刘勇, 李生, 刘勇, 王峰, 张, “基于深度监督关注度量的遥感图像变化检测”, *电子工程学报. 格奥西. 《遥感》*, 第 60 卷, 第 1-16 页, 2022 年。不。5604816.

[45]张晓明, 张晓明, “航空图像变化检测中的约束光流”, *《电子工程学报国际版》, 程序. 格奥西. 遥感*., 2011, 第 4176-4179 页。

[46]吕明, 魏世民, 季生, “基于卷积神经网络的多源建筑图像提取方法”, *《电子工程学报》. 格奥西. 《遥感技术》*, 第 57 卷, 第 57 期。1, 第 574-586 页, 2018 年 1 月。

[47]何凯, 张晓, 任淑娟, 孙军, “基于深度残差学习的图像识别”, *IEEE 会议, 计算, 程序. Vis.模式识别*., 2016, pp. 770-778。

[48]王小明, 王小明, “一种基于深度神经网络的内部协变量移位算法”, *《程序》, 国际版. 会议. 学习*., 2015, 第 448-456 页。

[49]张晓明, 张晓明, “深度稀疏整流神经网络”, *程序, 计算机科学与技术*, 2011, pp. 387 - 387。

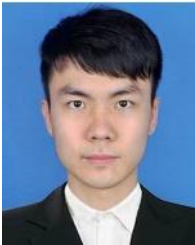
[50]王小明, 王小明, 王小明, 等。高斯误差线性单位 (Gaussian error linear units, GELUs) [j]. *计算机工程学报*, 2016,35(4):555 - 557。

[51]刘勇, 詹志强, 张晓明, 张, 杨锴, “基于深度连体卷积网络的遥感图像变化检测”, *电子工程学报, 格奥西. 遥感, 左*., 卷十八, 第。5, 第 811-815 页, 2021 年 5 月。



刘荣 (IEEE 学生会会员)，2019 年毕业于中国中山大学广州分校，获地理信息科学学士学位。现就读于地理与规划学院，攻读地图学与地理信息系统博士学位。

她的研究兴趣包括遥感图像的智能理解、变化检测和领域自适应。



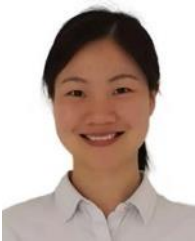
邓浩军，2020 年毕业于中国广州中山大学地理信息科学专业，获理学士学位。目前在地理与规划学院攻读地图学与地理信息系统专业硕士学位。

他的研究方向是物候学和城市化中的机器学习和深度学习。



柴卓群目前在中山大学广州地理与规划学院攻读理学士学位。

他的研究方向是基于深度学习的城市空间分析和图像处理。



刘荣 2013 年毕业于中国武汉大学测绘工程专业理学士，2018 年毕业于武汉大学测绘与遥感信息工程国家重点实验室，获摄影测量与遥感专业博士学位。

她现为中国中山大学地理与规划学院副教授。遥感技术博士后或高级研究员

2018 年至 2021 年，在德国航空航天中心，以及慕尼黑工业大学地球观测信号处理研究所工作。她的研究兴趣包括遥感图像处理和进化计算。