Dsfer-Net：一种利用现代Hopfield网络进行双时相变化检测的深度监督与特征检索网络

Shizhen Chang，*IEEE* 会员，Michael Kopp，*IEEE* 会员，Pedram Ghamisi，*IEEE* 高级会员Bo Du，*IEEE* 高级会员

arXiv:2304.01101v2 [cs.CV] 4 Jun 2024

摘要——**变化检测是高分辨率遥感图像的一项重要应用，旨在监测和分析地表随时间发生的变化。随着高分辨率遥感数据量迅速增长，且纹理特征日益复杂，研究者提出了多种定量的基于深度学习的方法。这些方法通过提取深层特征并结合时空信息，优于传统的变化检测方法。然而，对于深层特征如何提升检测性能，目前仍缺乏合理的解释。在我们的研究中，我们发现现代Hopfield网络层能够显著增强语义理解。本文提出一种用于双时相变化检测的深度监督与特征检索网络（Dsfer-Net）。具体而言，通过全卷积孪生网络联合提取双时相图像中具有高度代表性的深层特征。基于双时相图像的序列地理信息，我们设计了一个特征检索模块，以提取差异特征，并以深度监督的方式利用判别信息。此外，我们观察到，深度监督特征检索模块为所提出网络在深层中的语义理解提供了可解释的证据。最后，我们的端到端网络通过聚合不同层的检索特征和特征对，构建了一种新颖的框架。在三个公开数据集（LEVIR-CD、WHU-CD和CDD）上开展的实验，证实了所提出的Dsfer-Net优于其他最先进的方法。与表现最佳的DSAMNet相比，Dsfer-Net取得了显著提升，F1 分数分别提高了4.7%、5.9%和2.3%。此外，与我们此前提出的FrNet相比，Dsfer-Net同样取得了显著提升，在三个数据集上的F1 分数分别提高了2.0%、1.4%和4.5%。代码将在网上发布（https://github.com/ShizhenChang/Dsfer-Net）。**

索引词——**变化检测、高分辨率光学遥感、现代 Hopfield 网络、孪生卷积网-络。**

I. 引言

S. Chang 和 M. Kopp 任职于奥地利维也纳 Landstraßer Hauptstraße 5, 1030 的人工智能高级研究所（电子邮件：szchang@ieee.org；michael.kopp@iarai.ac.at）。

P. Ghamisi 任职于奥地利维也纳 Landstraßer Hauptstraße 5, 1030 的人工智能高级研究所，同时也任职于德国弗赖贝格 Chemnitzer Str. 40, D-09599 的亥姆霍兹德累斯顿-罗森多夫研究中心（HZDR）弗赖贝格亥姆霍兹资源技术研究所（HIF）机器学习组（电子邮件：pghamisi@gmail.com）。

B. Du 任职于中国武汉市武汉大学计算机科学学院人工智能研究院多媒体软件国家工程研究中心及湖北省多媒体与网络通信工程重点实验室，邮编 430079（电子邮箱：dubo@whu.edu.cn）。

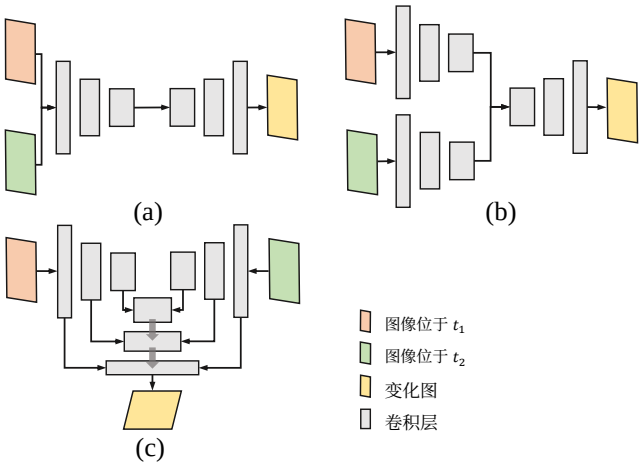


图 1. 基于深度学习的变化检测方法的基本架构。(a) 单分支架构。(b) 和 (c) 双分支架构。

召回率 EMOTE

基于遥感（RS）影像的变化检测（CD） [1], [2] 旨在识别和分析地理位置准确匹配的多时相影像中的土地覆盖变化。这项重要的图像解译任务广泛应用于地球表面时空演变分析，例如城市规划 [3], 灾害评估 [4],视觉监控 [5],以及自然资源管理 [6],等领域。

过去几十年来，双时相变化检测受到了广泛关注，并涌现出大量方法，包括传统方法 [7] 和基于深度学习的方法 [8]。在深度学习得到广泛应用之前，变化检测方法主要侧重于通过统计分析 [9], [10],、信号处理 [11], [12], 或机器学习 [13] 方法提取可解释的人工设计特征。然而，在实际应用中很难确定哪种方法最合适，尤其是面对规模相对较大的数据集时。此外，随着高分辨率光学传感器（如 WorldView-3、GeoEye- 1、QuickBird 和高分二号）的快速发展 [14],，可获取的高分辨率遥感图像日益增多，由此推动了众多基于深度学习的变化检测方法的发展，使其在复杂数据分析中具有更高的精度和更可靠的地理空间理解能力。

作为一种强大的深度学习架构，卷积神经网络（CNNs）的应用日益广泛，

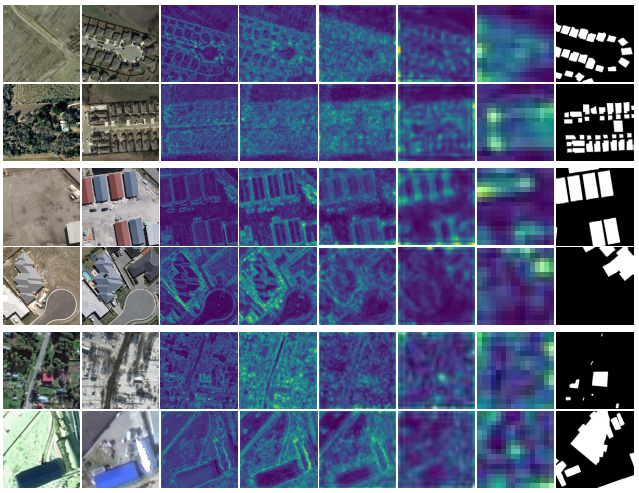


图2. 以 VGG-16 为骨干网络的孪生变化检测网络中，从浅层到深层的差异特征示例。(a) 来自 t_1 的图像。(b) 来自 t_2 的图像。(c) 第一阶段的特征。(d) 第二阶段的特征。(e) 第三阶段的特征。(f) 第四阶段的特征。(g) 第五阶段的特征。(h) 真实标签。

这是因为它们能够自动提取层次化多级特征，并学习携带变化区域判别信息的高度抽象语义上下文。与单图像处理任务不同，设计合适的图像对融合架构可以提高特征提取和变化诊断的鲁棒性 [15]。如图 1 所示，特征融合主要分为两类：单分支架构[16] 和双分支架构 [17]。单分支架构首先通过图像拼接或计算图像差分来融合两个输入，再将其作为一个整体输入网络 [18]。该架构采用早期融合（EF）策略，随后将融合后的图像转换为高层特征。然而，这种方法只能融合部分信息，无法充分考虑个体信息 [19]。双分支架构采用后期融合策略，首先学习单时相特征，然后将学习到的特征合并，作为解码器的输入。从本质上说，采用共享权重的孪生神经网络参数更少，收敛更快。尽管双分支架构优于单分支的图像级融合方法，但它不可避免地受到梯度消失现象的限制，导致较浅层无法有效学习有用特征。因此，合理结合深层与较浅层的特征，更有助于有效识别原始图像中导致变化的特征。

为解决上述问题并准确提取特征，研究人员设计了更深、更复杂的基于卷积神经网络的网络。众所周知，浅层特征主要涵盖图像的空间结构，包括纹理、边界和颜色；而深层特征则包含更抽象的内容和具有代表性的语义概念 [20]。因此，多尺度融合策略（图 1(c) 直观展示了这一策略）已被用于整合不同层的特征。研究人员提出了多种增大感受野的增强型网络，这些网络通常还会结合生成对抗网络（GAN） [21]，、长短期记忆网络（LSTM） [22]，以及空间与通道注意力机制 [23]等架构组件。然而，简单地融合由孪生架构的极深层学习到的特征，并不一定能提升变化检测方法的性能，因为这往往会导致参数量和计算内存需求激增。图2展示了采用 VGG-16 作为骨干网络的孪生网络的性能。值得注意的是，中间层对变化区域相关特征的保留效果最佳。虽然浅层保留了丰富的纹理细节，却难以滤除来自未变化区域的信息。相反，由于反复池化，高层特征往往变得过于抽象，从而妨碍网络准确识别感兴趣的变化。此外，卷积神经网络（CNN）的架构限制意味着，高层特征本质上与对应分支的浅层紧密关联，因此限制了对图像对之间时序信息的探索。

为应对这些挑战，我们提出了一种用于双时相变化检测的深度监督与特征检索网络，简称 Dsfer-Net。为提升双时相变化检测中深层特征提取的有效性，从而提高识别不同时刻空间与语义变化的性能和准确率，我们的框架引入了一个受深度学习中现代Hopfield网络（MHN）启发的深度监督特征检索（DSFR）模块。该模块利用双Hopfield层聚合时空特征，从而能够通过两条流对图像对中的变化进行序贯理解与跟踪。具体而言，DSFR模块中融入了深度监督，以确保对高层特征中的语义概念和空间结构进行全面恢复与检索。检索操作旨在提升双时相变化检测中深层特征提取的有效性，从而提高识别不同时刻空间与语义变化的性能和准确率。随后，解码器对拼接后的特征对进行解码，并采用多尺度融合策略将其与检索特征连接，以实现精确定位。最后，为优化训练，我们采用混合损失函数，以尽量减少网络传播和特征检索中的误差。

本文的主要贡献总结如下：

1) 我们提出了一种新颖的变化检测网络，将深度监督与特征检索理念相结合，用于双时相遥感图像。在此基础上，我们进一步尝试将 Hopfield 层引入变化检测，并阐释深层特征检索与聚合的有效性。所提出的网络通过孪生架构采用带权重策略的多尺度融合，从浅层到深层全面捕获感兴趣的变化。

2) 为了帮助高层特征保留变化信

息，同时泛化语义内容，DSFR模块专为深层设计。它通过序列式理解来解析全局上下文，并能在特征高度抽象时检索差异。我们加入了一种混合损失，以平衡整个网络和Hopfield模块的误差。

3）大量实验表明，所提出的 Dsfer-Net 在保持变化目标形状和边界方面表现出优越性能。在不同数据集的比较中，所提出的模型明显优于其他最先进的模型。

本文其余部分的结构安排如下：第二节回顾相关工作。第三节介绍所提出的 Dsfer-Net 方法。第四节介绍实验并讨论研究结果。最后，第五节总结全文。

II. 相关工作

A. 基于深度学习的变化检测方法

根据变化检测目的的不同分析，基于深度学习的变化检测方法可分为三类：基于像素的方法、基于对象的方法和基于深层特征的方法 [24]。基于像素的方法通常使用深度神经网络将图像映射到深层特征空间，然后逐像素比较特征，以获得所需的变化检测结果。例如，Zhang等人 [25] 使用深度置信网络（DBN）提取用于判别的关键信息。随后，将双时相特征映射到二维极坐标域，以表征变化信息，并通过采用监督聚类算法获得变化图。Du等人 [11] 设计了一个深度慢特征分析模块，先将输入图像映射到深层特征空间，再使用慢特征分析度量计算特征差异，从而获得变化图。Chen等人[26] 使用预训练的 VGG-16 架构和注意力机制提取特征；最终通过计算加权双边界对比损失得到变化图。

基于对象的方法的主要思想是对输入图像进行分割以提取特征，然后识别分割后目标的状态变化 [19]。Ji等人 [27]提出了一种混合型基于CNN的变化检测框架，将建筑物视为目标，并通过 Mask R-CNN 提取建筑物。Zhang等人 [28] 提出了一种端到端的超像素增强变化检测网络，该网络采用超像素采样网络来学习特定任务的超像素，并使用基于UNet的孪生网络来利用多层次变化线索。Liu等人 [29] 提出了一种基于对象的变化检测框架，该框架利用CNN提取图像块的特征，并结合不同的融合策略来区分变化。我们认为，对于基于像素和基于对象的变化检测方法，DBN、LSTM、注意力机制和卷积神经网络等深度学习模块都被用作工具，将原始图像转换为高层特征，同时保留与所关注的变化相关的最具判别性的信息并抑制未变化区域。

由于在特征提取方面表现出色，基于卷积神经网络的深度学习网络已广泛用于高分辨率遥感变化检测任务。为了从现有标注数据中提取更可靠的信息，研究人员开发了端到端深度学习架构——全卷积网络（FCN），用于基于深度特征的变化检测。在 [30],Daudt等人提出了三种基于 UNet 模型的 FCN：FC-EF、FC-Siam-conc 和 FC-Siam-diff。FC-EF 方法采用早期融合策略，将图像对拼接后作为网络的一个输入；FC-Siam-conc 和 FC-Siam-diff 则是 FC-EF 的孪生扩展。与传统孪生网络一样，FC-Siam-conc 和 FC-Siam-diff 的编码层分为两个结构相同、权重共享的分支。这两种模型的区别在于，FC-Siam-conc 在解码器中拼接两条跳跃连接，而 FC-Siam-diff 则计算特征差异的绝对值，并在解码阶段将其与跳跃连接拼接。 [31]中也同时提出了类似架构。在后续研究中， [32],Daudt等人修改了 FC-EF 架构，加入带有跳跃连接的残差块，以提高结果的空间精度。Peng等人 [33]提出了一种改进的 UNet++网络，该网络首先将配准图像对拼接作为输入，然后采用多侧输出融合策略，结合来自不同语义层级的变化图。

尽管这些方法展现出良好的变化检测能力，但未能充分利用空间上下文信息以及从浅层到深层的内部关系，因此容易受到噪声、角度、阴影和上下文等因素引发的伪变化影响。为此，研究人员提出了增强算法来弥补这些不足。此外，这些方法还利用输入数据之间的时间关系，以生成更加稳健的输出。Sun 等人 [22] 将 LSTM 集成到 UNet 架构中，并采用卷积长短期记忆网络作为核心卷积结构。该方法结合了卷积与循环结构，并证明循环神经网络能够有效捕捉图像间的时间关系。Tang 等人 [34] 提出了一种用于目标细粒度变化检测（OFCD）的频率解耦交互（FDINet）方法，并引入小波交互模块（WIM）来建模时空相关性。Wang 等人 [35] 采用空谱交叉融合模块，通过执行各种卷积操作来增强语义特征。Lei 等人 [36] 提出了一种高效的超轻量空谱特征协作网络（USSFC-Net），该网络采用空谱特征协作策略，可直接学习三维空谱注意力权重，而无需引入额外参数。

近年来，注意力机制在图像处理领域展现出巨大潜力：它们能够根据特征自身像素之间的相关性自动增强特征。Peng等人 [18] 提出了一种差异增强密集注意力卷积神经网络（DDCNN），该网络由多个上采样注意力单元组成

，用于建模高层特征与低层特征之间的内部相关性。Patil等人 [37] 开发了一种优化架构，该架构采用基于孪生网络的预训练编码器和基于注意力的UNet解码器，以重建细粒度特征图。Lv等人 [38] 将空间-光谱注意力机制与多尺度膨胀卷积模块相结合，以更积极地捕获变化。Zhao等人 [39] 提出了一种基于特征交互和多任务学习的变化检测方法（FMCD），该方法采用混合注意力块（MAB）来提高模型对变化的敏感性。与用作特征提取器的传统卷积神经网络相比，近期变化检测算法的发展引入了基于Transformer的方法。例如， [40], 提出的GeoFormer采用非局部孪生编码器，将几何卷积与Transformer相结合。这种创新组合增强了模型为远程上下文特征提供局部几何表示信息的能力。

注意力机制的结构存在局限：特征只能在特定尺度（通道、空间或位置）上得到增强，却忽略了两条流之间的序列关系。因此，受 MHN 启发，我们将深度监督特征检索（DSFR）模块引入变化检测框架。

B. 图像检索与特征聚合

基于内容的图像检索旨在通过分析图像内容的颜色、纹理或形状，从大规模数据库中查找与给定查询图像相关的图像 [41]。视觉数据的指数级增长推动了基于内容的图像检索迅速发展；该方法通过分析颜色、纹理或形状，从大规模数据库中检索相关图像 [41]。基于内容的图像检索涵盖实例和语义任务 [42], ，需要在大规模图像集合中准确、高效地检索具有判别性的特征表示。

过去二十年来，随着深度学习在特征提取中的应用，图像检索取得了惊人的进展。在深度学习取得突破之前，人工设计特征（如视觉词袋（BoW）[43] 和尺度不变特征变换（SIFT） [44]）被广泛使用。此后，能够直接从数据中学习多层级特征表示的深度学习，在这一领域展现出无可争议的优越性能。众所周知，深度学习模型中的卷积层能够捕获与图像局部区域相关的结构细节，其范围由特定感受野决定。为增强从深度卷积神经网络中提取的激活值的判别能力，并帮助检索特定实例，图像检索中引入了特征嵌入和聚合策略。在 [45],, Noh 等人提出了一种基于卷积神经网络、用于大规模地标检索的注意力局部特征描述子，可据此选择关键点特征。Chaudhuri 等人提出了一种孪生图卷积网络，利用对比损失最小化最具判别性的嵌入空间，以进行遥感图像检索 [46]。其他聚合指标也与现成的深度学习模型相结合，用于图像检索例如局部聚合描述子向量（VLAD） [47]，选择性匹配核（SMK）[48], 深度哈希嵌入 [49],等。

与前面提到的图像检索系统不同，我们提出了一种DSFR模块，能够直接在深度卷积层中执行检索过程。需要注意的是，该模块利用现代 Hopfield 网络，在监督下检索深度差异特征，并对检索到的特征对进行聚合。随后，这些特征对将被传递用于双时相遥感变化检测。

C. 现代*Hopfield*网络

Hopfield网络于20世纪80年代提出 [50], [51], ，是一类基于能量的二值神经网络，旨在作为联想记忆，专门用于存储和检索二值数据。在 [52]–[57] 提出在能量函数中使用多项式项的研究基础上， [58]证明，使用指数项可以显著提高这些经典二值 Hopfield 网络的存储容量。与这些二值联想记忆网络不同，我们采用了 [59] 提出的现代 Hopfield 网络，该网络处理连续输入，也就是说，可以存储和检索实值数据。重要的是，与经典网络不同，这些现代 Hopfield 网络（MHN）具有可微性，因此可以通过梯度下降进行训练。它们还保留了二值网络的一项关键能力：存储数量呈指数级增长的样本，并且只需一次更新即可检索。MHN的成功应用包括解决大规模多实例学习任务 [60],、预测少样本和零样本化学反应模板 [61] ，以及开展层次化原型学习并改进用于遥感图像生成任务的文本-图像嵌入表示 [62]。

现代Hopfield网络擅长联想记忆任务，因此在关联环境特征随时间发生的变化方面也极具价值。MHN的学习机制使深度学习网络能够根据观测数据的反馈，持续提升检测和分类变化的能力。为了保留深层特征的表征与检索能力，并更好地理解和利用遥感数据中的真实变化，本文在先前研究 [63] 的基础上，提出了一种深度监督特征检索模块。此外，为了实现MHN所学习的可解释深层特征，我们进一步探索了原型Hopfield层在语义分割中的适当应用。

III. 方法

A. 概述

所提出的 Dsfer-Net 总体架构如图 3 所示。将配对图像输入孪生卷积神经网络，以提取多层深度特征。编码器基于带有预训练参数的 VGG-16 构建，并移除了其最后的全连接层和最后一个池化层。为解码提取的多尺度特征图，我们提出了理解融合（CF）策略，并设计了深度监督特征检索（DSFR）模块。为捕获深层中的语义和序列信息，

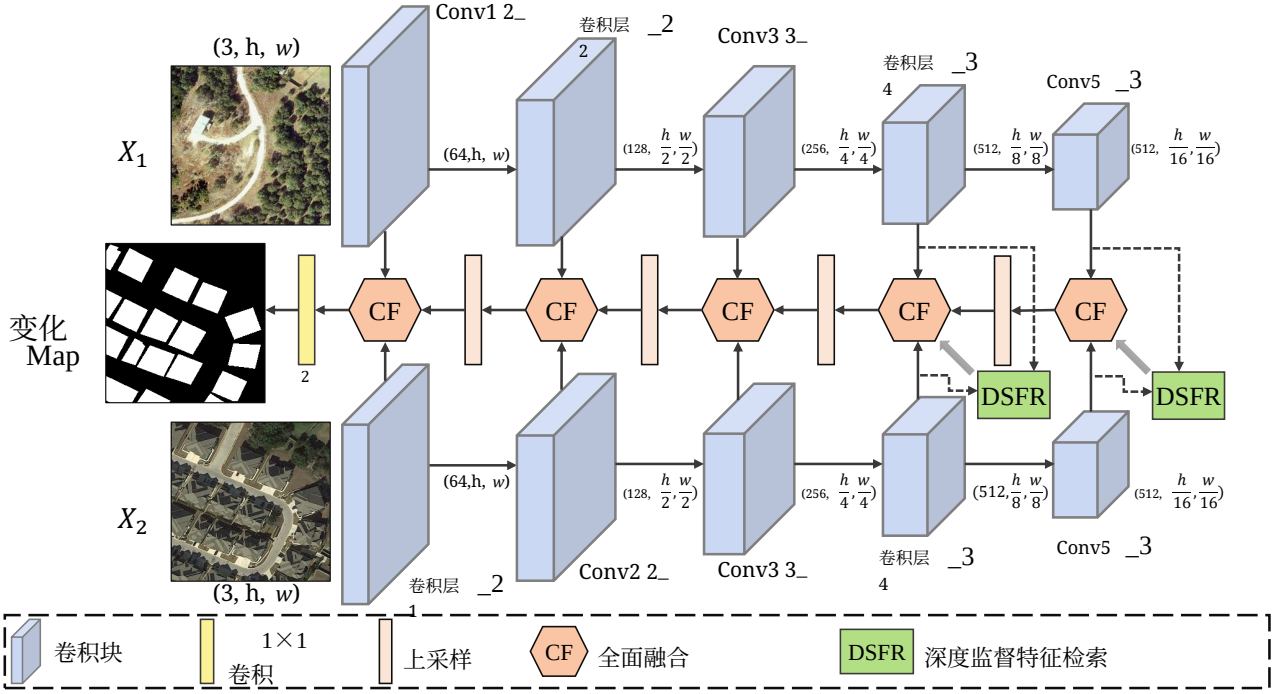


图3. Dsfer-Net 的整体架构。特征提取器采用共享权重的孪生架构。使用 pool5 之前的 VGG-16 作为骨干网络。DSFR模块用于深层的特征聚合与检索。CF 块全面融合多尺度特征。

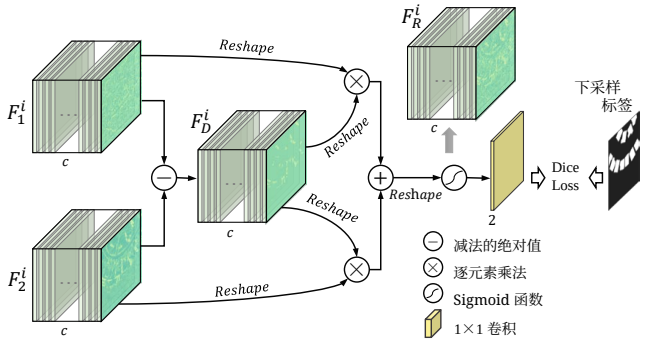


图4. 深度监督特征检索（DSFR）模块。(F_1^i, F_2^i) 表示特征提取器提取的第 i 对特征图。对于所提出的网络，DSFR模块应用于第四和第五对特征，其中 $i = 4$, 5。

采用两个 DSFR 模块来聚合并检索隐藏层中的特征差异，并通过所提出的 CF 策略，将来自卷积层（卷积层1 2、Conv2 2、Conv3 3、Conv4 3、Conv5 3）的多尺度特征与检索特征进行层次化融合。

设 $X_1 \in \mathbb{R}^{3\times h\times w}$ 和 $X_2 \in \mathbb{R}^{3\times h\times w}$ 分别表示在不同时间（t_1 和 t_2）采集的配对图像，Y 表示标注变化图，其中 h 和 w 分别表示图像的高度和宽度。这项二值变化检测任务的主要步骤可简要描述如下：

1）首先，将图像对输入具有共享权重的特征提取器，通过该提取器可获得五对多尺度原始特征图 $\{(F_1^i,\ F_2^i)\}_{i=1}^5$。
2）对于更高层次的特征，即 $i = 4$, 5，应用 DSFR模块以获得能够保持感兴趣变化的中间变化图。基于变化图和下采样标签计算 Dice 损失。

3) 同时，两个分支的特征图通过 CF 策略拼接并解码，DSFR 模块的双Hopfield层检索得到的变化图 F_R^i 则作为拼接特征的权重掩码传递。随后，将多尺度特征融合并解码至相同维度，并根据融合特征图和真实标签计算交叉熵损失。

4) 最后，通过反向传播，使用交叉熵损失与 Dice 损失的组合来优化整个网络。

B. 深度监督特征检索

对于提取的深层特征图，我们认为特征对之间的变化也应得到保留并且可追溯。然而，正如图2所示，高层特征通常过于抽象且语义化，难以保留与关注变化相关的重要纹理信息。

为利用双时相图像特征中的语义鸿沟并有效聚合序列信息，我们引入了深度监督特征检索（DSFR）模块。如图4所示，DSFR模块主要由双Hopfield层构成，其设计旨在利用并整合输入图像中的序列信息。为了准确捕捉更深层网络中特征之间的变化信息，DSFR模块采用

监督策略，并最终检索出与变化图相关的图。该架构配置使我们的方法能够增强深层特征的检索与合成，从而更全面地理解图像数据中的时序变化和语义细微差异。给定一对深层特征（F_1^i，F_2^i），DSFR模块计算差异特征的绝对值，对特征进行重塑，并通过Hopfield 型层检索差异特征，该层将两个相应的深层特征视为存储模式。随后，将检索特征融合并归一化，得到深层变化图。

Hopfield 层的原型最初于 [59], 中提出，能够在具有多种结构和可变应用领域的深度神经网络中学习或检索状态模式。为使 MHN 更好地适用于所提出的二值变化检测任务，我们将差异特征定义为状态模式，以帮助激活与原始特征中存储的变化信息高度相关的记忆节点。在传入 Hopfield 层之前，先将给定特征重塑为二维矩阵，其中 $F_D^i \in \mathbb{R}^{(h'w')\times c}$ 表示二维绝对差异特征，原始特征 $F_j^i \in \mathbb{R}^{(h'w')\times c}, j = 1, 2$ 表示存储模式，其中 c、h' 和 w' 分别表示特征的深度、高度和权重。对于每个 Hopfield 层，检索得到的差异特征的输出表示为：

$$F_{R_j}^i = \text{softmax}(\beta F_D^i W_D W_S^\top F_j^{i\top}) F_j^i W_S, \tag{1}$$

其中，softmax(·)表示按行计算的 softmax 函数，β 为缩放参数。 W_D和 W_S 分别是可学习参数，用于将状态模式和存储模式投影到统一空间。为进一步提升双 Hopfield层的效果，将检索特征合并并归一化为一张变化图：

$$F_R^i = \sigma(F_{R_1}^i + F_{R_2}^i), \tag{2}$$

其中，σ 表示 Sigmoid 函数。

DSFR模块的作用有两个方面：1）检索得到的深层变化图直接传递至 CF 块，作为权重掩码，用于突出从较浅层提取的特征中的变化区域；2）检索得到的深层变化图输入 1×1 卷积层，以生成中间变化图 $M_i \in \mathbb{R}^{2\times h'\times w'}$，并根据该中间变化图和下采样标签计算 Dice 损失。

C. 全面融合

融合特征图最直接的方法
该策略将两个流的特征与来自更深层的特征直接拼接为单个特征图。然而，引入 DSFR模块后，必须考虑检索到的、反映变化权重的特征。直接组合异质特征会增加训练难度。为应对这一挑战，更有效地将从 DSFR模块检索到的深层变化图融入原始特征图，同时高效融合多尺度特征，我们设计了一种称为 CF 块的全面融合策略，并在解码器中实现。

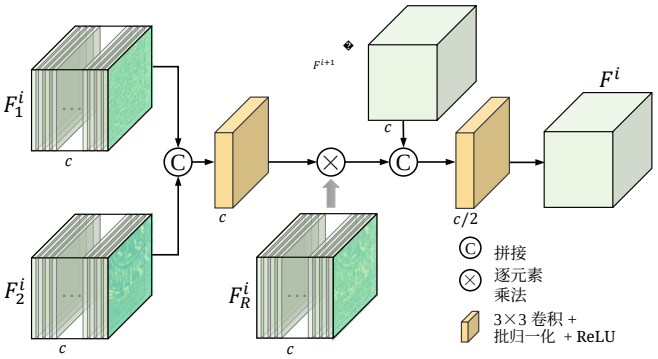


图 5. 全面融合（CF）策略。(F_1^i, F_2^i) 表示由特征提取器提取的第 i 对特征图。如果 $i = 4, 5$，则将拼接后的特征与检索到的特征 F_R^i 相乘。$\tilde{F}^{i+1}$ 表示从更深层的 CF 块中获得的上采样特征。

有关 CF策略的详细示意，请参阅图 5。给定一对原始特征（F_1^i，F_2^i），CF 块首先将这两个特征拼接起来，并将其输入一个 3×3卷积层。对于最深的层，即 $i = 4$ 和 5，检索到的变化图 F_R^i 将作为拼接特征的掩码输入。相乘后的特征会突出变化信息并抑制未变化区域。随后，将相乘后的特征与来自更深层的上采样 CF 特征 $\tilde{F}^{i+1}$ 拼接，并通过另一个 3×3卷积层对拼接后的特征进行解码，作为整个网络中第 i个 CF 块的输出图。形式上，第 i阶段的 CF 特征图可表示为

$$F^i = f_2^{(3\times3)}(\textbf{Concat}(f_1^{(3\times3)}(\mathcal{F}^i) \odot F_R^i, \tilde{F}^{i+1})), \tag{3}$$

其中，每个 $f^{(3\times3)}(\cdot)$ 表示一个卷积层，其卷积核大小为 3×3，并具有与图 5 所示的 i 相连的适当数量的通道，后接批归一化层（BN）和 ReLU 激活层；**拼接**$(\cdot,\cdot)$ 表示两个特征的拼接，$\mathcal{F}^i$ =**拼接**(F_1^i, F_2^i) 表示 F_1^i 和 F_2^i 拼接后的特征，而 $\odot$ 表示逐元素乘法。

CF 块按顺序运行，将检索到的变化图与原始特征图融合以增强变化信息，随后再与来自更深层的特征融合。与直接拼接相比，这种全面融合策略能够有效利用反映变化的信息，同时融合来自较浅层和较深层的多个特征图。

D. 网络训练

已有研究证明，将softmax层视为逐像素分类问题时，二元交叉熵损失函数对二分类问题具有良好的效果 [64]。然而，对于典型的变化检测数据集，变化像素与未变化像素的类别分布通常严重不平衡。为缓解这一类别不平衡问题，我们采用加权二元交叉熵损失函数来训练模型。该函数可表示如下

$$\mathcal{L}_{wbce} = -\frac{1}{hw}\sum_{l=1}^{hw}(w_1 Y_l \log \hat{Y}_l + w_0(1-Y_l)\log(1-\hat{Y}_l)), \quad (4)$$

其中， w_0 和 w_1 分别表示未变化类别和变化类别的权重， Y_l 和 Y_l 分别表示标注变化图和预测变化图中的第 l个像素。

在 DSFR模块中，还会对更深层的卷积特征计算 Dice 损失。令 i 表示骨干网络的第 i个基本阶段；因此，下采样后的标注变化图和中间预测变化图分别可表示为 Y^i 和 $\hat{Y}^i$，Dice 损失可写为

$$\mathcal{L}^i_{dice} = 1 - 2 \cdot \frac{\sum_{m=1}^{h'w'} Y^i_m \hat{Y}^i_m}{\sum_{m=1}^{h'w'}(Y^i_m + \hat{Y}^i_m)}, \quad (5)$$

其中， Y^i_m 和 $\hat{Y}^i_m$ 分别表示下采样标签和中间预测的第 m个像素。在本研究中，原始标注变化图与中间图之间的高度和宽度关系可推导为

$$\begin{aligned} h' &= 2^{1-i} \cdot h, \\ w' &= 2^{1-i} \cdot w. \end{aligned} \quad (6)$$

最后，用于训练所提出的网络的目标函数可表示为:

$$\mathcal{L}_{total} = \mathcal{L}_{wbce} + \lambda \cdot (\mathcal{L}^4_{dice} + \mathcal{L}^5_{dice})/2, \quad (7)$$

其中， λ 为用于调整各损失项的加权因子。

IV. 实验与讨论

本研究在三个广泛使用的变化检测数据集上对所提出的 Dsfer-Net 进行实验，并与八种先进算法进行比较。下文将详细介绍实现方法和评估指标。最后，我们将给出实验结果并进行详细讨论。

A. 数据集描述与评估指标

我们在三个变化检测数据集上进行了实验:

1) **LEVIR-CD** 是一个大规模建筑物变化检测数据集，涵盖美国德克萨斯州的 20 个不同区域。该数据集包含 637 对超高分辨率（VHR）Google Earth（GE）图像块，尺寸为 1024×1024 像素，空间分辨率为 0.5 米/像素。按照 [65],，我们将图像对划分为训练集、验证集和测试集，比例约为 7:1:2。原始图像被裁剪为 16 个尺寸为 256×256 的小图块。因此，训练集共包含 7,120 对图像，验证集包含 1,024 对，测试集包含 2,048 对。

2) **WHU-CD** 是 WHU 建筑物数据集的一个子集 [66], 其中包含航空与卫星影像。建筑物变化子集覆盖新西兰基督城 20.5 km^2 的区域。2011年2月，该地区发生了 6.3 级地震，并在随后几年重建。一组成对图像的尺寸为 32,507×15,354 像素，空间分辨率为 0.075 米/像素，分别采集于 2012年 和 2016年。为开展实验，我们将原始图像裁剪为 8189 个不重叠图块，每个图块含有 256×256 像素，并按 6:1:3 的比例将裁剪后的图块划分为训练集、验证集和测试集。因此，训练集共包含 4,572 对图像，验证集包含 762 对，测试集包含 2,286 对。

3) **CDD** [67] 是一个季节变化遥感数据集，采集自 Google Earth（DigitalGlobe），包含七对人工创建的季节变化图像及其真实标注，每对图像的像素数为 4725×2700 ；此外还包含四对变化较小、分辨率为 1900×1000 像素的季节变化图像。图像的空间分辨率为 0.03∼1 米/像素。将原始图像裁剪为大小为 256×256 的图块，并随机旋转图块（0∼2π）后，共获得 16,000 对图像，其中 10,000 对用于训练，3,000 对用于验证，其余 3,000 对用于测试。

为定量评估所提出方法的性能，选取精确率（P）、召回率（R）、F1 分数、总体精度（OA）和交并比（IoU）作为评估指标，定义如下:

$$\begin{aligned} \mathrm{P} &= \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \\ \mathrm{R} &= \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \\ \mathrm{F1} &= \frac{2\mathrm{PR}}{\mathrm{P}+\mathrm{R}} \\ \mathrm{OA} &= \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{FP}+\mathrm{TN}+\mathrm{FN}} \\ \mathrm{IoU} &= \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}+\mathrm{FN}} \end{aligned}$$

其中，TP、TN、FP 和 FN 分别表示真正例、真负例、假正例和假负例的数量。

B. 对比方法与实现设置

为验证我们模型的性能，我们将提出的 Dsfer-Net 与八种最先进的方法进行了比较。以下简要介绍这些对比方法。

- FC-Siam-conc [30] 是一种基于 U-Net 的架构，采用孪生编码流分别从两幅图像中提取深层特征；随后将提取的特征与两条跳跃连接进行拼接，并传递至解码流。

- FC-EF-Res [32] 是一种早期融合架构，首先将双时相图像拼接在一起，再输入网络。与传统卷积层不同，该架构加入了残差块，以提升网络性能并促进训练。

- STANet [65]设计了一种自注意力机制，用于建模双时相图像的时空关系并捕获不同尺度的依赖关系。随后计算特征之间的注意力权重，以生成更具判别性的特征。

- BiT [68] 是一种基于 Transformer 的模块，利用视觉词来表示高层次的时空变化

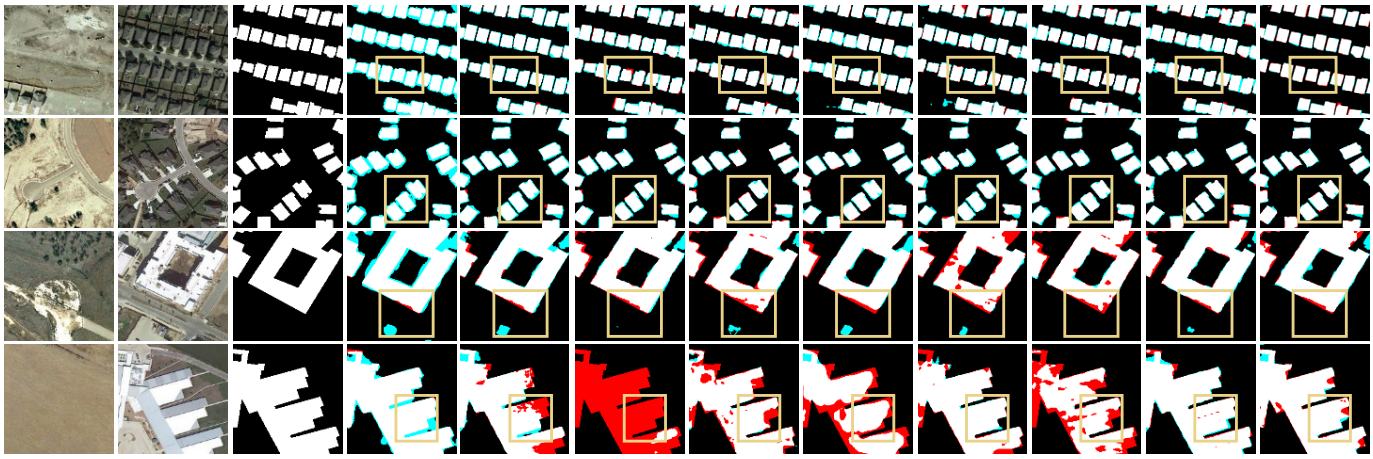


图6. 不同方法在 LEVIR-CD 数据集上的可视化示例。(a) 来自 t_1 的图像。(b) 来自 t_2 的图像。(c) 真实标注。(d) FC-Siam-conc。(e) FC-EF-Res。(f) STANet。(g)BiT。(h) DSAMNet。(i) ChangeFormer。(j)VcT。(k) FrNet。(l)Dsfer-Net。真正例（TP）、假正例（FP）、真负例（TN）和假负例（FN）分别以白色、青色、黑色和红色表示。

概念。双时相图像 Transformer（BiT）被集成到一种基于深层特征差分的变化检测网络中。

- DSAMNet [69] 采用两个深度监督层来辅助隐藏层的学习，并设计了一个度量模块，用于堆叠不同尺度的缩放后特征。该方法采用卷积块注意力模块（CBAM），使特征更具判别性。
- ChangeFormer [70] 是一种基于 Transformer 的网络，在孪生网络架构中将分层 Transformer 编码器与多层感知机（MLP）解码器相结合。这种设计能够高效捕获多尺度长程细节。
- VcT [71] 采用视觉变化 Transformer 作为其骨干网络，通过捕获双时相图像中可靠标记之间的依赖关系，有效利用图像内和图像间线索。
- FrNet [63] 是一项利用 MHN 融合深层特征的初步研究。Hopfield 池化层与最深的卷积层结合使用，并选用 VGG-16 作为特征提取器。

所有实验均使用 PyTorch 实现，并在 NVIDIA A100 GPU 上训练。为确保对比方法的效率比较公平，我们同时考虑原始文献中指定的参数设置和实际性能，并报告最佳性能。对于所提出的 Dsfer-Net，我们将 Hopfield 层的投影维度设为 512。我们采用 Adam 优化器，在 LEVIR-CD 和 WHU-CD 数据集上的初始学习率为 $1e-5$，在 CDD 数据集上的初始学习率为 $1e-4$。学习率将在 30,000 次迭代后线性衰减至 0。我们将权重衰减设为 $1e-5$，批量大小设为 32。每 500 次迭代进行一次验证，并使用验证集上表现最佳的模型在测试集上进行评估。三个数据集均采用相同的训练、验证和测试设置。

C. 基准比较与分析

本节将所提出的模型与三种基准数据集上的八种最先进的方法进行比较。可视化结果见图 6 至图 8。在 LEVIR-CD 和 WHU-CD 数据集中，建筑物是待评估的变化目标。可以观察到，大多数方法都能在 LEVIR-CD 和 WHU-CD 数据集中正确检测出小型建筑物。然而，与其他方法相比，FC-Siam-conc 和 FC-EF-Res 未能区分建筑物周围未发生变化的土地覆盖类型，导致虚警率较高。这表明，直接拼接深层特征或特征差异，无法有效地为网络训练提供判别信息。对于较大型建筑物，如方框所示，STANet、BiT、DSAMNet 和 VcT 无法正确检测这些目标，尤其是形状较为复杂的目标。结果表明，尽管广受认可的注意力机制能够重新整合并强调深层特征中的重要信息，但其应用也暴露出自检索机制的局限性，可能无法突出时相图像之间的真实差异。与 FrNet 相比，所提出的模型能够更好地控制虚警，并更清晰地勾勒出变化目标的边界。在所有方法中，我们的方法在检测变化目标方面表现最佳，并能返回最准确的建筑物轮廓。我们得出结论，所设计的 DSFR模块有助于从时间序列数据中提取真实变化，并尽可能保留接近真实情况的细节。

CDD 数据集记录了地面在不同时期的季节性变化和目标变化。所提出的方法优于其他对比方法。如图 8 所示，选择建筑物、道路和汽车作为可视化的变化目标，以反映受时间、极端天气和人类活动影响的变化。可视化结果表明，我们的模型能够从 CDD 数据集中提取小尺寸和大尺寸变化，并且与其他方法相比，假正例（FP）和假负例（FN）最少

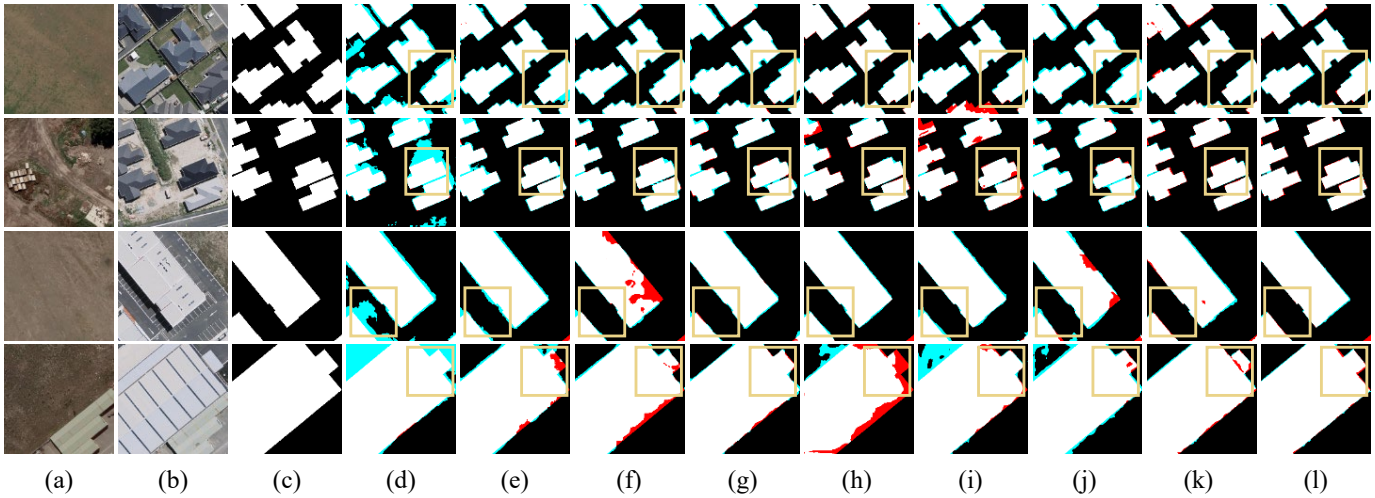


图 7. 不同方法在 WHU-CD 数据集上的可视化示例。(a) 来自 t_1 的图像。(b) 来自 t_2 的图像。(c) 真实标注。(d) FC-Siam-conc。(e) FC-EF-Res。(f) STANet。(g) BiT。(h) DSAMNet。(i) ChangeFormer。(j) VcT。(k) FrNet。(l) Dsfer-Net。真正例、假正例、真负例和假负例分别用白色、青色、黑色和红色表示。

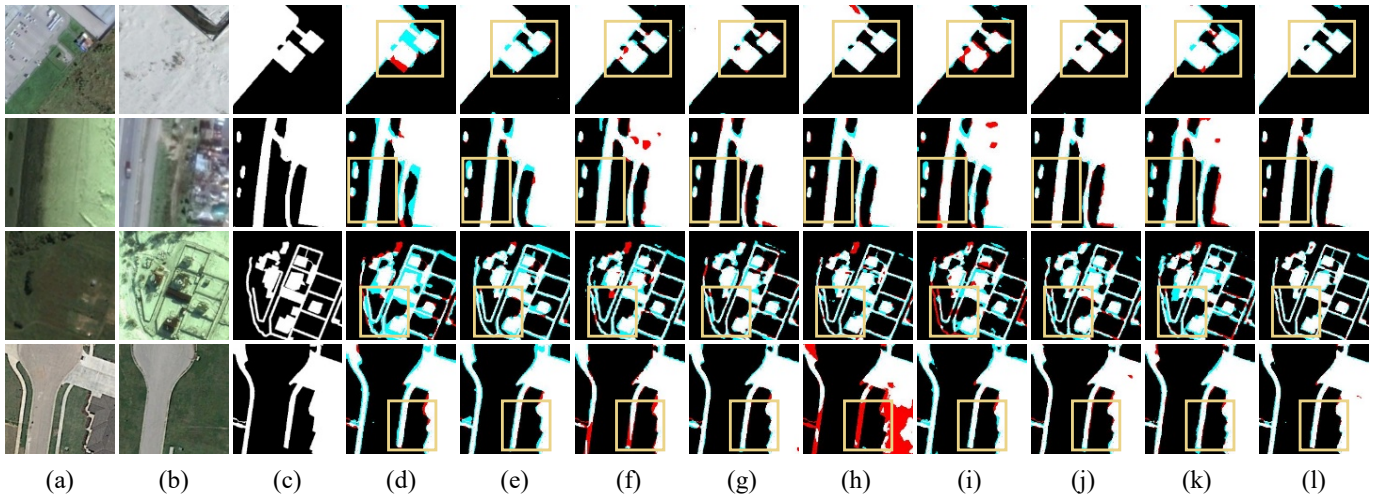


图 8. 不同方法在 CDD 数据集上的可视化示例。(a) 来自 t_1 的图像。(b) 来自 t_2 的图像。(c) 真实标注。(d) FC-Siam-conc。(e) FC-EF-Res。(f) STANet。(g) BiT。(h) DSAMNet。(i) ChangeFormer。(j) VcT。(k) FrNet。(l) Dsfer-Net。真正例、假正例、真负例和假负例分别用白色、青色、黑色和红色表示。

。即使在相对复杂的语义场景中，所提出的模型也能准确检测真实变化。相比之下，FC-Siam-conc 和 FC-EF-Res 在检测变化时无法避免虚警。此外，它们在保留狭窄道路和小目标的边界方面表现不佳。尽管 STANet、BiT、ChangeFormer 和 FrNet 能够提取面积相对较大的变化目标，但在检测细粒度变化方面仍表现不佳。作为一种深度监督和基于注意力的方法，DSAMNet 在识别狭窄且细小的变化方面表现良好，但会漏检大尺寸目标，且无法保持目标的内部紧致性。在所有可比较的方法中，作为一种基于 Transformer 的方法，VcT 在识别狭窄目标和面积相对较大的变化区域时，检测性能较为均衡。

所有方法的定量评估结果如表 I 所示。鉴于我们提出的方法针对二元变化在变化检测中，我们仅报告变化类别的准确性。在这三个数据集上，我们提出的模型在精确率、F1分数、总体精度和交并比方面均表现出色。在 LEVIR-CD 数据集上，Dsfer-Net 的 F1分数和交并比分别比表现最佳的对比方法高 1.9 和 3.2 个百分点。在 WHU-CD 数据集上，我们的模型在 F1分数和交并比分别比 FrNet 高 1.4 和 2.4 个百分点。在 CDD 数据集上，所提出的方法的 F1分数和交并比分别比 DSAMNet 高 2.2 和 3.9 个百分点。此外，我们观察到这些方法中的精确率与召回率呈负相关，这可以直观地理解为虚警和漏检通常不会同时发生。因此，变化检测方法需要在检测更多正例与控制虚警之间取得平衡。在不同数据集上的定量比较中，我们提出的方法表现更优，虚警更少，漏检数量也较为合理

表 I 三个数据集上的定量比较结果。

方法	LEVIR-CD					WHU-CD					CDD				
	P	R	F1	OA	IoU	P	R	F1	OA	IoU	P	R	F1	OA	IoU
FC-Siam-conc	68.29	**96.93**	80.12	97.55	66.84	64.95	**93.57**	76.68	97.90	62.18	77.31	88.94	82.72	95.61	70.53
FC-EF-Res	79.71	93.62	86.11	98.46	75.61	87.92	86.49	87.20	99.06	77.31	85.66	94.85	90.02	97.52	81.86
STANet	86.55	87.99	87.27	98.69	77.41	90.55	83.75	87.02	99.08	77.02	88.92	90.83	89.86	97.58	81.59
BiT	86.30	90.17	88.19	98.77	78.87	89.57	85.72	87.60	99.10	77.94	92.49	91.22	91.85	98.09	84.92
DSAMNet	90.00	82.49	86.08	98.52	75.56	83.49	90.18	86.71	99.05	76.53	93.27	92.54	92.91	98.32	86.76
ChangeFormer	81.36	87.77	84.45	98.35	73.08	85.01	79.61	82.22	98.73	69.80	85.12	90.07	87.53	96.97	77.82
VcT	85.67	88.14	86.89	98.64	76.81	84.04	88.84	86.37	98.96	76.02	93.19	89.38	91.25	97.98	83.90
FrNet	86.80	90.91	88.81	98.83	79.87	93.76	88.72	91.17	99.37	83.77	89.83	91.52	90.67	97.78	82.92
Dsfer-Net	**89.42**	92.15	**90.76**	**99.04**	**83.09**	**94.17**	91.04	**92.58**	**99.46**	**86.18**	**93.70**	**96.69**	**95.17**	**98.84**	**90.79**

*注：所有分数均以百分比 (%) 表示，最佳结果以**粗体**标出。

表 II LEVIR-CD 数据集上的参数量和推理时间（秒）。

方法	参数量(M)	推理时间(S)
FC-Siam-conc	1.55M	16.60
FC-EF-Res	1.10M	18.68
STANet	16.93M	40.33
BiT	11.94M	39.60
DSAMNet	16.95M	27.91
ChangeFormer	43.99M	107.27
VcT	12.41M	82.23
FrNet	15.00M	15.22
Dsfer-Net	34.08M	21.27

表 III DSFER-NET 中不同策略的性能贡献。

	CF	DSFR	LEVIR-CD	WHU-CD	CDD
基线+拼接			89.97	91.87	93.91
基线+CF	✓		90.05	91.88	94.70
基线+拼接+DSFR		✓	90.28	92.34	94.33
基线+CF+DSFR	✓	✓	**90.76**	**92.58**	**95.17**

* 注：所有分数均以百分比（%）报告，最佳 F1 分数以**粗体**突出显示。

误报和合理的漏检。

为更好地了解模型的效率，我们比较了所提出模型与其他对比方法在 LEVIR-CD 数据集上的参数量（Params）和推理时间。所有实验均在配备 NVIDIA A100 GPU 的服务器上进行。如表 II 所示，我们提出的 Dsfer-Net 模型参数量为 34.08M。这主要是因为模型包含 Hopfield 层，其需要更多计算内存。尽管参数量较高，我们提出的方法仍展现出有竞争力的推理时间，表现与基于注意力的方法相当。

D. 消融研究

为评估我们设计的CF策略和DSFR模块的有效性，我们在三个数据集上进行了消融研究：Base+拼接、Base+CF、Base+拼接+DSFR，

表IV LEVIR-CD 数据集上的参数量和推理时间（秒）。

骨干网络	F1	OA	参数量（M）	推理时间（秒）
ResNet-18	88.24	98.75	22.90M	18.12
Transformer	86.98	98.64	59.24M	57.08
VGG-16	90.76	99.04	34.08M	21.27

以及Base+CF+DSFR。顾名思义，Base+拼接是一个以VGG-16为骨干网络、将拼接后的特征作为解码器输入的孪生网络。Base+CF采用CF策略融合多尺度特征。在Base+拼接+DSFR中，DSFR模块从更深层提取变化信息，并通过拼接融合各卷积块的特征。Base+CF+DSFR对应于所提出的Dsfer-Net。不同配置的F1分数如表 III所示。可以观察到，与经典的特征拼接过程相比，CF策略在三个数据集上均带来了小幅提升。此外，与单独使用特征拼接相比，将CF策略与DSFR模块结合可进一步提升变化检测性能。

为进一步评估我们方法的有效性，我们使用 Dsfer-Net 模型，在 LEVIR-CD 数据集上采用不同的骨干网络架构进行了基准实验。表IV展示了采用三种经典骨干网络 ResNet-18、Transformer 和 VGG-16 提取双时相深层特征所得的结果。为保持一致性，我们沿用了此前使用 VGG-16 架构开展实验时的参数设置。

我们的研究结果表明，采用 VGG-16 架构时，Dsfer-Net 可取得最佳性能；与 ResNet-18 和基于 Transformer 的编码器相比，其性能更优，同时在参数量和推理时间之间保持了良好的平衡。值得注意的是，在将我们的方法与表 I 中列出的其他方法进行比较时，尤其是与采用 ResNet-18 作为骨干网络的方法（如 STANet、BiT 和 DSAMNet）比较时，Dsfer-Net 在 LEVIR-CD 数据集上展现出更优的性能。此外，与采用基于 Transformer的编码器的 ChangeFormer 和 VcT 相比，我们配备基于Transformer的编码器的模型同样表现出更好的性能。这些结果凸显了我们的方法在不同

表V 加权因子对损失函数的敏感性分析。

λ	LEVIR-CD		WHU-CD		CDD	
	F1	OA	F1	OA	F1	OA
0.001	90.69	99.04	92.58	99.46	95.01	98.80
0.01	90.73	99.04	**92.58**	**99.46**	94.83	98.75
0.1	**90.76**	**99.04**	92.56	99.46	94.99	98.79
1	90.66	99.03	92.34	99.45	**95.17**	**98.84**
10	90.12	98.96	92.04	99.43	94.75	98.75
100	89.47	98.88	92.75	99.48	94.49	98.68

*注：所有分数均以百分比（%）表示，最佳结果以**粗体**突出显示。

骨干网络架构下的有效性，并彰显了其在变化检测任务中的竞争优势。

E. 敏感性分析

如第三节所述，所提出的网络的损失函数由交叉熵损失和 Dice 损失构成，并通过加权因子进行平衡。为探究不同 λ 取值对所提出网络训练的影响，我们在三个数据集上设置不同的 λ 取值进行了对比实验。由于采用了DSFR模块来提升特征提取能力并保留深层中的纹理信息，因此 λ的最佳取值代表了训练过程中检索特征最合适的占比。请注意，当 λ 设为 0 时，该网络等同于上一小节中的第二个基线模型“Base+CF”。如表V所示， λ 取值相对较小时更有利于提升性能。众所周知，经过重复下采样后，变化目标的形状会略有变化，细节中的纹理信息也会有所丢失。为深度监督模块赋予适当的权重有助于捕捉跨场景变化目标的位置，但权重过高可能会在细粒度地物覆盖上造成虚警，导致性能下降。具体而言，对于 LEVIR-CD、WHU-CD 和 CDD 数据集，当 λ 的取值分别为 0.1、0.01 和 1 时，性能最佳。

*F. Hopfield*特征可视化

为进一步评估DSFR模块的有效性，我们展示了三个数据集中的检索特征可视化结果。DSFR模块主要由双Hopfield层构成，并应用于第四和第五卷积阶段。因此，我们将聚合Hopfield特征与尺寸相同的下采样图像对和标签进行比较。

如图9至图11所示，在这三个数据集中，图像对的下采样结果均存在明显差异
数据集（包括颜色、光照的变化，甚至一些未变化的目标）。然而，深度监督特征检索模块通过强调中间变化图与下采样标签之间的一致性，增强了网络在更深层捕捉纹理结构的能力，从而提高了网络对真实变化的鲁棒性。在 LEVIR-CD 和 CDD 数据集中，DSFR模块

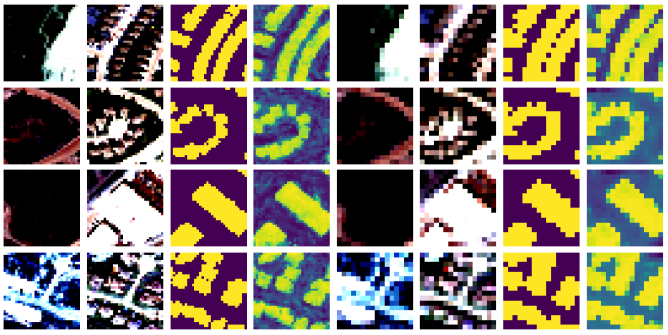


图9. LEVIR-CD 数据集中DSFR模块检索到的深层特征可视化示例。(a) t_1 中的图像，下采样8倍。(b) t_2 中的图像，下采样8倍。(c) 下采样8倍的真实标签。(d) 第四阶段检索到的特征。(e) t_1 中的图像，下采样16倍。(f) t_2 中的图像，下采样16倍。(g) 下采样16倍的真实标签。(h) 第五阶段检索到的特征。

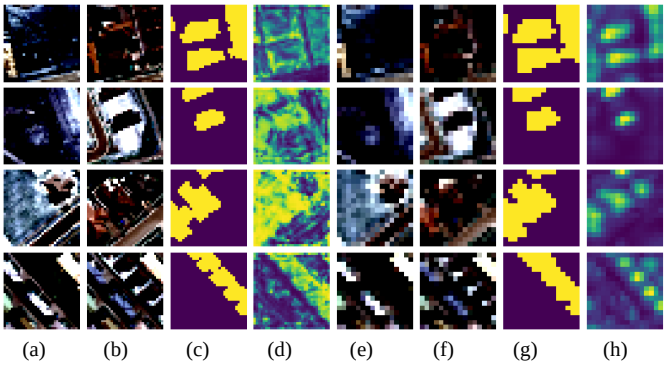
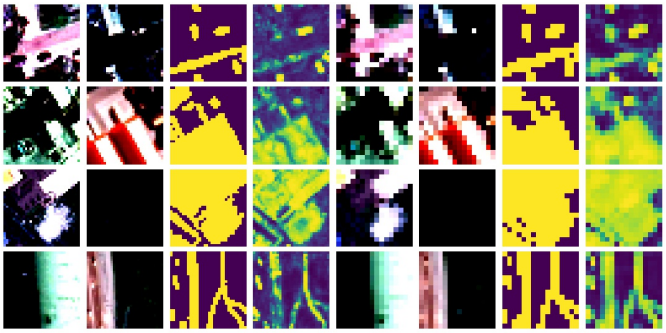


图 10. WHU-CD 数据集中 DSFR模块检索到的深层特征可视化示例。(a) t_1 中的图像下采样 8 倍。(b) t_2 中的图像下采样 8 倍。(c) 真实标签下采样 8 倍。(d) 第四阶段的检索特征。(e) t_1 中的图像下采样 16 倍。(f) t_2 中的图像下采样 16 倍。(g) 真实标签下采样 16 倍。(h) 第五阶段的检索特征。

能在更深层有效突出关注的变化，且中间变化图与下采样真实标注保持一致。在 CDD 数据集中，双Hopfield层还能有效检索狭窄道路和汽车等小目标。然而，在 WHU-CD 数据集中，第四阶段检索到的差异特征准确性较低。这种差异主要是因为 WHU 数据集采集了地震前后的目标区域影像，而灾后快速重建和城市扩张导致新建筑物、道路、广场和植被发生了显著变化。因此，检索到的差异特征与建筑物变化标签存在显著差异。相较之下，可以观察到，随着网络深入，DSFR模块在第五阶段能够更准确地定位变化信息。

V. 结论

本文提出了一种用于遥感图像变化检测的深度监督与特征检索网络。受现代 Hopfield 网络启发，我们检索差异特征并聚合序列信息



(a) (b) (c) (d) (e) (f) (g) (h)

图 11. CDD 数据集中 DSFR模块检索的深层特征可视化示例。(a) 来自 t_1 的图像，下采样 8 倍。(b) 来自 t_2 的图像，下采样 8 倍。(c) 真实标签，下采样 8 倍。(d) 第四阶段检索的特征。(e) 来自 t_1 的图像，下采样 16 倍。(f) 来自 t_2 的图像，下采样 16 倍。(g) 真实标签，下采样 16 倍。(h) 第五阶段检索的特征。

，以深度监督的方式进行学习。为合理拼接多层级特征，我们在解码器中采用了全面融合策略。在三个广泛使用的变化检测数据集上进行了大量实验，以验证所提出方法的有效性。在两个建筑物变化检测数据集中，所提出的模型能够更好地保留建筑物轮廓：对于密集建筑物，可避免虚警；对于不规则且相互独立的建筑物，则能控制漏检。对于季节变化数据集，所提出的模型能够在伪季节性变化中准确捕捉真实变化，并检测小型和狭窄目标。通过消融研究和敏感性分析，我们验证了 DSFR 模块对骨干网络的提升作用，并确定了合理的参数设置。最后，可视化示例展示了 Hopfield 层在差异特征检索与聚合方面的性能。通过实验我们发现，当场景中出现许多非预期变化时，特征检索模块在较深层中可能无法准确定位感兴趣的变化。因此，我们将继续设计合适的结构，以克服伪变化带来的副作用。此外，我们计划在未来的工作中分析不同训练比例对检测精度的影响。这项分析或将进一步揭示我们方法的可扩展性与实用性。

致谢

作者谨向分享代码的所有对比方法的作者、LEVIR-CD、WHU-CD 和 CDD 数据集的贡献者、提出宝贵意见并参与讨论的徐永浩博士，以及提供支持的人工智能高级研究所（IARAI）表示感谢。

参考文献

[1] H. Jiang、M. Peng、Y. Zhong、H. Xie、Z. Hao、J. Lin、X. Ma 和 X. Hu，“基于深度学习的高分辨率遥感图像变化检测综述”，*Remote Sens.*，第 14 卷，第 7 期，文章编号 1552，2022 年。

[2] Z. Lv、H. Huang、X. Li、M. Zhao、J. A. Benediktsson、W. Sun 和 N. Falco，“利用异构遥感图像进行土地覆盖变化检测：综述、进展与展望”，Proc. IEEE，2022年。[3] J. Prendes、M. Chabert、F. Pascal、A. Giros 和 J.-Y. Tourneret，“一种用于检测由同质和异构传感器采集图像变化的新型多变量统计模型”，IEEE Trans. Image Process.，第24卷，第3期，第799–812页，2014年。[4] Z. Zheng、Y. Zhong、J. Wang、A. Ma 和 L. Zhang，“基于深度面向对象语义变化检测框架的建筑物损毁快速评估：从自然灾害到人为灾害”，Remote Sens. Environ.，第265卷，第112636页，2021年。[5] M. J. Carlotto，“遥感影像变化的检测与分析及其在广域监视中的应用”，IEEE Trans. Image Process.，第6卷，第1期，第189–202页，1997年。[6] B. Descl´ee、P. Bogaert 和 P. Defourny，“基于统计型面向对象方法的森林变化检测”，Remote Sens. Environ.，第102卷，第1–2期，第1–11页，2006年。[7] R. Touati、M. Mignotte 和 M. Dahmane，“利用无监督的基于像素对的马尔可夫随机场模型进行遥感图像多模态变化检测”，IEEE Trans. Image Process.，第29卷，第757–767页，2019年。[8] L. Ru、B. Du 和 C. Wu，“基于相关性融合的多时相场景分类与场景变化检测”，IEEE Trans. Image Process.，第30卷，第1382–1394页，2020年。[9] E. F. Lambin 和 A. H. Strahlers，“多时相空间中的变化向量分析：利用高时间分辨率卫星数据检测并分类土地覆盖变化过程的工具”，Remote Sens. Environ.，第48卷，第2期，第231–244页，1994年。[10] Z. Lv、M. Zhang、W. Sun、J. A. Benediktsson、T. Lei 和 N. Falco，“利用高光谱遥感图像进行土地覆盖变化检测的空间上下文信息利用框架”，IEEE Trans. Geosci. Remote Sens.，第61卷，第1–11页，2023年。[11] B. Du、L. Ru、C. Wu 和 L. Zhang，“用于多时相遥感图像变化检测的无监督深度慢特征分析”，IEEE Trans. Geosci. Remote Sens.，第57卷，第12期，第9976–9992页，2019年。[12] G. Liu、Y. Yuan、Y. Zhang、Y. Dong 和 X. Li，“基于风格变换的无监督变化检测空间-光谱特征学习”，IEEE Trans. Geosci. Remote Sens.，第60卷，第1–15页，2020年。[13] Z. Liu、G. Li、G. Mercier、Y. He 和 Q. Pan，“通过同质像素变换检测异构遥感图像中的变化”，IEEE Trans. Image Process.，第27卷，第4期，第1822–1834页，2017年。[14] C. Zhang、P. Yue、D. Tapete、L. Jiang、B. Shangguan、L. Huang 和 G. Liu，“用于高分辨率双时相遥感图像变化检测的深度监督图像融合网络”，ISPRS J. Photogramm. Remote Sens.，第166卷，第183–200页，2020年。[15] D. Wang、J. Zhang、B. Du、G.-S. Xia 和 D. Tao，“遥感预训练的实证研究”，IEEE Trans. Geosci. Remote Sens.，2022年。[16] M. Lin、G. Yang 和 H. Zhang，“变化是一个过程：面向超高分辨率遥感图像的成对至视频变化检测网络”，IEEE Trans. Image Process，第32卷，第57–71页，2022年。[17] Q. Li、R. Zhong、X. Du 和 Y. Du，“TransUNetCD：一种用于光学遥感图像变化检测的混合Transformer网络”，IEEE Trans. Geosci. Remote Sens.，第60卷，第1–19页，2022年。[18] X. Peng、R. Zhong、Z. Li 和 Q. Li，“基于注意力机制和图像差分的光学遥感图像变化检测”，IEEE Trans. Geosci. Remote Sens.，第59卷，第9期，第7296–7307页，2020年。[19] Y. Lei、X. Liu、J. Shi、C. Lei 和 J. Wang，“利用深层特征进行多尺度超像素分割以实现变化检测”，IEEE Access，第7卷，第36600–36616页，2019年。[20] J. Yosinski、J. Clune、A. Nguyen、T. Fuchs 和 H. Lipson，“通过深度可视化理解神经网络”，arXiv预印本 arXiv:1506.06579，2015年。[21] B. Hou、Q. Liu、H. Wang 和 Y. Wang，“从 W-Net 到 CDGAN：基于深度学习技术的双时相变化检测”，IEEE Trans. Geosci. Remote Sens.，第58卷，第3期，第1790–1802页，2019年。[22] S. Sun、L. Mu、L. Wang 和 P. Liu，“L-UNet：一种用于遥感图像变化检测的LSTM网络”，IEEE Geosci. Remote. Sens. Lett.，2020年。[23] Y. Zhou、F. Wang、J. Zhao、R. Yao、S. Chen 和 H. Ma，“基于时空信息的遥感图像变化检测多头自注意力机制”，IEEE Trans. Circ. Syst. Video Tech.，第32卷，第10期，第6615–6626页，2022年。[24] Y. Sun、L. Lei、D. Guan 和 G. Kuang，“用于异构遥感图像无监督变化检测的迭代鲁棒图”，IEEE Trans. Image Process.，第30卷，第6277–6291页，2021年。

[25] H. Zhang、M. Gong、P. Zhang、L. Su 和 J. Shi，“利用深度表示与特征变化分析进行多光谱影像特征级变化检测”，IEEE Geosci. Remote. Sens. Lett.，第13卷，第11期，第1666–1670页，2016年。[26] J. Chen、Z. Yuan、J. Peng、L. Chen、H. Huang、J. Zhu、Y. Liu 和 H. Li，“DASNet：用于高分辨率卫星图像变化检测的双重注意力全卷积孪生网络”，IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.，第14卷，第1194–1206页，2020年。[27] S. Ji、Y. Shen、M. Lu 和 Y. Zhang，“利用卷积神经网络和模拟样本从大规模航空影像中检测建筑实例变化”，Remote Sens.，第11卷，第11期，第1343页，2019年。[28] H. Zhang、M. Lin、G. Yang 和 L. Zhang，“Escnet：一种用于超高分辨率遥感图像的端到端超像素增强变化检测网络”，IEEE Trans. Neural Netw. Learn. Syst.，2021年。[29] T. Liu、L. Yang 和 D. Lunga，“利用面向对象的图像分析进行深度学习变化检测”，Remote Sens. Environ.，第256卷，第112308页，2021年。[30] R. C. Daudt、B. Le Saux 和 A. Boulch，“用于变化检测的全卷积孪生网络”，载于 Proc. - Int. Conf. Image Process. (ICIP)，2018年，第4063–4067页。[31] Y. Chen、X. Ouyang 和 G. Agam，“Mfcnet：用于图像变化检测的端到端方法”，载于 Proc. - Int. Conf. Image Process. (ICIP)，2018年，第4008–4012页。[32] R. C. Daudt、B. Le Saux、A. Boulch 和 Y. Gousseau，“用于大规模语义变化检测的多任务学习”，Comput. Vis. Image Underst.，第187卷，第102783页，2019年。[33] D. Peng、Y. Zhang 和 H. Guan，“利用改进的 UNet 进行高分辨率卫星图像端到端变化检测++”，Remote Sens.，第11卷，第11期，第1382页，2019年。[34] Y. Tang、S. Feng、C. Zhao、Y. Fan、Q. Shi、W. Li 和 R. Tao，“一种基于频率解耦交互的高分辨率遥感图像目标细粒度变化检测方法”，IEEE Trans. Geosci. Remote Sens.，第62卷，第1–13页，2023年。[35] J. Wang、F. Liu、L. Jiao、H. Wang、H. Yang、X. Liu、L. Li 和 P. Chen，“SSCFNet：一种用于遥感变化检测的空间-光谱交叉融合网络”，IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.，2023年。[36]T. Lei、X. Geng、H. Ning、Z. Lv、M. Gong、Y. Jin 和 A. K. Nandi，“一种用于遥感图像变化检测的超轻量空间-光谱特征协同网络”，IEEE Trans. Geosci. Remote Sens.，第61卷，第1–14页，2023年。[37] P. S. Patil、R. S. Holambe 和 L. M. Waghmare，“EffCDNet：利用深度注意力网络迁移学习，实现高空间分辨率卫星图像变化检测”，Digit. Signal Process.，第118卷，第103250页，2021年。[38] Z. Lv、F. Wang、G. Cui、J. A. Benediktsson、T. Lei 和 W. Sun，“利用变化幅度图像引导空间-光谱注意力网络，通过遥感图像进行土地覆盖变化检测”，IEEE Trans. Geosci. Remote Sens.，第60卷，第1–12页，2022年。[39] C. Zhao、Y. Tang、S. Feng、Y. Fan、W. Li、R. Tao 和 L. Zhang，“基于特征交互与多任务学习的高分辨率遥感双时相图像变化检测”，IEEE Trans. Geosci. Remote Sens.，2023年。[40] J. Zhao、L. Jiao、C. Wang、X. Liu、F. Liu、L. Li 和 S. Yang，“Geo-Former：一种用于变化检测的几何表示变换器”，IEEE Trans. Geosci. Remote Sens.，第61卷，第1–17页，2023年。[41] W. Chen、Y. Liu、W. Wang、E. Bakker、T. Georgiou、P. Fieguth、L. Liu 和 M. S. Lew，“实例检索的深度学习：综述”，arXiv预印本 arXiv:2101.11282，2021年。[42] B. Barz 和 J. Denzler，“深度学习时代的基于内容的图像检索与语义鸿沟”，载于 Proc. - Int. Conf. Pattern Recognit.，2021年，第245–260页。[43] J. Sivic 和 A. Zisserman，“Video Google：一种用于视频中目标匹配的文本检索方法”，载于 Proc. IEEE国际计算机视觉会议，第3卷，2003年，第1470页。[44] D. G. Lowe，“来自尺度不变关键点的显著图像特征”，Int. J. Comput. Vis.，第60卷，第2期，第91–110页，2004年。[45] H. Noh、A. Araujo、J. Sim、T. Weyand 和 B. Han，“利用注意力深度局部特征进行大规模图像检索”，载于 Proc. IEEE国际计算机视觉会议，2017年，第3456–3465页。[46] U. Chaudhuri、B. Banerjee 和 A. Bhattacharya，“用于基于内容的遥感图像检索的孪生图卷积网络”，Comput. Vis. Image Underst.，第184卷，第22–30页，2019年。[47] Y. Gong、L. Wang、R. Guo 和 S. Lazebnik，“深度卷积激活特征的多尺度无序池化”，载于 Eur. Conf. Comput. Vis.

[48] J. Yang、J. Liang、H. Shen、K. Wang、P. L. Rosin 和 M.-H. Yang，“用于图像检索的深度卷积特征动态匹配核”，IEEE Trans. Image Process.，第 27 卷，5302，2018 11，5288 no. pp. – 。[49] T.-T. Do、D.-K. Le Tan、T. T. Pham 和 N.-M. Cheung，“用于大规模图像搜索的特征同步聚合与哈希”，载于 IEEE国际计算机视觉会议，2017 年，第 6618–6627 页。[50] J. J. Hopfield，“具有涌现集体计算能力的神经网络与物理系统”，美国国家科学院院刊，第 79 卷，第 8 期，第 2554–2558 页，1982 年。[51] ——，“具有分级响应的神经元具有与双态神经元类似的集体计算特性。”美国国家科学院院刊，第 81 卷，第 10 期，第 3088–3092 页，1984 年。[52] H. H. Chen、Y. C. Lee、G. Z. Sun、H. Y. Lee、T. Maxwell 和 C. L. Giles，“联想记忆的高阶相关模型”，AIP会议论文集，第 151 卷，第 1 期，第 86–99 页，1986 年。[53] D. Psaltis 和 H. P. Cheol，“非线性判别函数与联想记忆”，AIP会议论文集，第 151 卷，375，1986 1，370no. pp. – 。[54] P. Baldi 和 S. S. Venkatesh，“自旋玻璃与高阶神经网络的稳定点数量”，Phys. Rev. Lett.，第 58 卷，第 913–916 页，1987 年。[55] E. Gardner，“多连接神经网络模型”，J. Phys. A，第 20 卷，第 11 期，第 3453–3464 页，1987 年。[56] L. F. Abbott 和 Y. Arian，“广义网络的存储容量”，Phys. Rev. A，第 36 卷，第 5091–5094 页，1987 年。[57] D. Krotov 和 J. J. Hopfield，“用于模式识别的稠密联想记忆”，神经信息处理系统进展，第 29 卷，2016年。[58] M. Demircigil、J. Heusel、M. L¨owe、S. Upgang 和 F. Vermet，“关于一种具有巨大存储容量的联想记忆模型”，J. Stat. Phys.，第 168 卷，第 2 期，第 288–299 页，2017 年。[59] H. Ramsauer、B. Sch¨afl、J. Lehner、P. Seidl、M. Widrich、L. Gruber、M. Holzleitner、M. Pavlovi´c、G. K. Sandve、V. Greiff、D. Kreil、M. Kopp、G. Klambauer、J. Brandstetter 和 S. Hochreiter，“Hopfield网络就是你所需要的一切”，载于国际学习表征会议，2021 年。[60] M. Widrich、B. Sch¨afl、M. Pavlovi´c、H. Ramsauer、L. Gruber、M. Hol- zleitner、J. Brandstetter、G. K. Sandve、V. Greiff、S. Hochreiter 等，“用于免疫受体库分类的现代Hopfield网络与注意力”，神经信息处理系统进展，第 33 卷，18 845，2020 18 832pp. – 。[61] P. Seidl、P. Renz、N. Dyubankova、P. Neves、J. Verhoeven、M. Segler、J. K. Wegner、S. Hochreiter 和 G. Klambauer，“用于少样本与零样本反应模板预测的现代Hopfield网络”，arXiv 预印本 arXiv:2104.03279，2021 年。[62] Y. Xu、W. Yu、P. Ghamisi、M. Kopp 和 S. Hochreiter，“Txt2Img-MHN：利用现代Hopfield网络从文本生成遥感图像”，IEEE图像处理汇刊，第 32 卷，第 5737–5750 页，2023 年。[63] S. Chang、M. Kopp 和 P. Ghamisi，“用于双时相遥感图像变化检测的深度特征检索网络”，2022 年。[64] Y. Lei、D. Peng、P. Zhang、Q. Ke 和 H. Li，“用于街景变化检测的分层配对通道融合网络”，IEEE图像处理汇刊，第 30 卷，第 55–67 页，2020 年。[65] H. Chen 和 Z. Shi，“一种用于遥感图像变化检测的时空注意力方法及新数据集”，Remote Sens.，第 12 卷，第 10 期，文章编号 1662，2020 年。[66] S. Ji、S. Wei 和 M. Lu，“利用开放航空与卫星影像数据集进行多源建筑物提取的全卷积网络”，IEEE Trans. Geosci. Remote Sens.，第 57 卷，第 1 期，第 574–586 页，2018 年。[67] M. Lebedev、Y. V. Vizilter、O. Vygolov、V. Knyaz 和 A. Y. Rubis，“使用条件生成对抗网络进行遥感图像变化检测。”Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.，第 42 卷，第 2 期，2018 年。[68] H. Chen、Z. Qi 和 Z. Shi，“利用变换器进行遥感图像变化检测”，IEEE Trans. Geosci. Remote Sens.，第 60 卷，14，2021 1pp. – 。[69] Q. Shi、M. Liu、S. Li、X. Liu、F. Wang 和 L. Zhang，“用于遥感变化检测的深度监督注意力度量网络及开放航空影像数据集”，IEEE Trans. Geosci. Remote Sens.，第 60 卷，第 1–16 页，2021 年。[70] W. G. C. Bandara 和 V. M. Patel，“用于变化检测的基于变换器的孪生网络”，载于国际地球科学与遥感研讨会 (IGARSS)，2022 年，第 207–210 页。[71] B. Jiang、Z. Wang、X. Wang、Z. Zhang、L. Chen、X. Wang 和 B. Luo，“Vct：用于遥感图像变化检测的视觉变化 Transformer”，IEEE Trans. Geosci. Remote Sens.，2023 年。