

Transformer与CNN融合的双分支网络在遥感图像阴影检测中的应用

王一帆, 黄鲜, 王建林, 周彤, 周文俊, 彭博

西南石油大学 计算机与软件学院, 成都 610500

摘要: 遥感图像中的阴影对图像解读和地物要素提取至关重要, 但它们也可能成为干扰因素。当前, 尽管深度学习在遥感图像阴影检测中取得了一定进展, 但仍面临漏检、误检以及网络模型复杂等挑战。为此, 本研究提出了一种融合Transformer与卷积神经网络(CNN)优势的双分支阴影检测网络。该方法采用双分支结构, 从全局和局部两个层面深入挖掘信息, 实现特征的有效整合, 以更精确地捕捉阴影特征。此外, 通过引入阴影预测模块和联合损失函数, 进一步提升了阴影检测的准确性。在航空影像阴影数据集AISD(Aerial Imagery Shadow Dataset)上的实验结果表明, 该方法显著提高了阴影检测的准确率, 达到97.112%, 并成功降低了误检率(BER降低了0.389), 充分验证了其有效性。总之, 本研究提出的双分支遥感图像阴影检测网络为减少小面积阴影的漏检与误检提供了新的有效手段, 对遥感图像解译和地物要素提取具有重要意义。

关键词: 遥感图像, 阴影检测, 语义分割, 双分支网络, 特征融合, Transformer, CNN, ResNet

中图分类号: TP751

引用格式: 王一帆, 黄鲜, 王建林, 周彤, 周文俊, 彭博. XXXX. Transformer与CNN融合的双分支网络在遥感图像阴影检测中的应用. 遥感学报, XX(XX): 1-11

WANG Yifan, HUANG Xian, WANG Jianlin, ZHOU Tong, ZHOU Wenjun, PENG Bo. XXXX. Application of Dual-Branch Network Integrated with Transformer and CNN in Remote Sensing Image Shadow Detection. National Remote Sensing Bulletin, DOI: 10.11834/jrs.20243358]

1 引言

随着科技的不断进步, 高分辨率卫星光学遥感技术已成为获取地球表面信息的重要手段。自20世纪60年代E.L.Pruitt(Walker2006)提出卫星遥感技术的概念以来, 这项技术已广泛应用于卫星和航空领域。遥感技术的发展不仅提升了我们对地球的观测能力, 还带来了高级图像处理技术的迫切需求。在遥感图像处理中, 一个关键挑战是阴影的存在。高分辨率遥感图像中普遍存在的阴影会显著减少图像的信息量, 从而影响图像的可视化、解译以及地物要素的定量提取。根据形成原理, 阴影可分为投射阴影和自身阴影(Li等, 2013; Guan等, 2022)。投射阴影是由物体阻挡光线形成的, 它会导致图像光谱变化减少,

进而使得相关性失效。相比之下, 自身阴影是由物体表面与光源方向的不一致导致的(Zali等, 2022), 通常受到周围物体反射光的影响而显得更亮。在遥感图像中, 自身阴影的特征通常是不均匀的, 这增加了检测其形状的难度。因此, 为了准确解析遥感图像并提取地物要素, 必须有效处理这些阴影。如何准确检测和处理阴影已成为遥感图像处理领域的重要研究方向。未来的研究需要不断探索新的方法和技术, 以应对阴影带来的挑战, 进一步提升遥感图像的解析精度和应用价值。

近年来, 深度学习技术, 特别是卷积神经网络(Convolutional Neural Network, CNN), 在计算机视觉任务中取得了显著的进展。这类方法通过端到端的学习方式, 能够自动提取图像信息的特

收稿日期: XXXX-XX-XX; 预印本: XXXX-XX-XX

基金项目: 四川省科技计划(编号:2023NSFSC1393, 2023NSFSC0504), 西南石油大学“启航计划”项目(编号:2021QH001)

第一作者简介: 王一帆, 研究方向图像处理。E-mail: yifan.wang@swpu.edu.cn

通信作者简介: 王一帆, 研究方向图像处理。E-mail: yifan.wang@swpu.edu.cn

征, 展现出强大的特征提取能力。相较于传统方法, 深度学习避免了大量先验知识和繁琐步骤的需求, 因而在图像处理领域得到广泛应用。在高分辨率遥感图像处理中, 深度学习同样展现出其潜力。特别是在阴影检测任务中, 一些基于深度学习的方法已被成功引入。此外, 深度学习在遥感领域的其他应用也取得了诸多成果, 包括高光光谱图像分类 (Li 等, 2022)、语义分割 (Xu 等, 2021)、目标检测 (Liu 等, 2022) 和图像重建 (Shao 等, 2019) 等。这些进展表明深度学习在遥感图像处理中的广泛应用前景。

阴影检测任务在本质上可以视作一种图像分割任务。众多成功的语义分割方法, 如 FCN (Long 等, 2015)、U-Net (Ronneberger 等, 2015)、SegNet (Badrinarayanan 等, 2017)、以及 PSPNet (Zhao 等, 2017) 和 DeepLabv3+ (Firdaus-Nawi 等, 2011) 都已被应用于阴影检测, 并在多个方面展现出优于传统方法的效果。然而, 在遥感图像领域, 阴影的多样性、复杂场景以及标记数据的缺乏都构成了独特的挑战。为了应对这些挑战, 近年来研究者提出了一些专门的深度学习模型。例如, Luo 等人 (Luo 等, 2020) 引入了 DSDDNet 模型, 该模型采用编码器-解码器架构并结合有监督的深度残差融合模块。值得一提的是, 他们还构建了首个公开的高分辨率航空影像阴影数据集 AISD (Aerial Imagery Shadow Dataset)。Jin 等人 (Jin 等, 2020) 则通过整合空间注意模块和新型损失函数进一步提升了阴影检测的准确性。同样值得关注的是 Zhang 等人 (Zhang 等, 2020) 提出的循环阴影注意力网络 RSAM (Recurrent Shadow Attention Model), 该模型利用注意力机制逐步优化检测结果, 并配套构建了 SSAD (Shadow Semantic Annotation Database) 数据集以供进一步研究使用。

随着技术的持续进步, 罗等人 (Luo 等, 2021) 进一步提出了边缘感知空间金字塔融合网络 (ESPFNet)。该网络创新性地结合了平行空间金字塔结构 (Parallel Spatial Pyramid, PSP)、边缘检测模块 (Edge Detection Module, EDM) 和边缘感知多分支集成 (Edge-aware Multi-branch Integration, EMI)。这些组件的协同作用显著提升了阴影检测的效果。同时, Liu 等人 (Liu, Zhang, 等, 2021) 基于多尺度空间注意机制, 提

出了 MSASDNet 模型。该模型专门针对航空遥感影像的阴影检测进行设计, 并成功解决了在使用航空图像数据集 AISD 时显著阴影不完全提取的问题。这些新方法的涌现不仅为遥感图像阴影检测领域注入了新的活力, 还为后续研究提供了宝贵的思路和技术支持。

最近, Shuktepe 等人 (Ufuktepe 等, 2021) 提出了一种颇具创新性的阴影掩码估计方法。该方法巧妙地结合了三维点云技术和已知的太阳角度, 以生成用于阴影检测的训练数据——阴影掩模。这种自监督的阴影掩模生成策略为基于深度学习的阴影检测架构带来了新的发展机遇。与此同时, Zhu 及其团队 (Zhu 等, 2022) 在最近的研究中提出了一种名为上下文细节感知网络 (CDANet) 的方法。该方法在阴影提取任务中表现出色, 取得了令人满意的效果。这些研究成果不仅充分展示了深度学习在遥感图像阴影检测领域的广泛应用和持续发展势头, 也为该领域的后续研究提供了宝贵的启示和技术支持。

在遥感图像阴影检测领域, 尽管已涌现出众多方法, 但仍面临一系列挑战。首要挑战在于遥感图像场景的高复杂度, 这使得准确区分阴影像素变得极为棘手。由于遥感图像的高分辨率特性, 阴影呈现出多样化的亮度和尺寸, 这无疑增加了检测算法的设计难度。为提升分割准确性而采用的复杂模型设计往往伴随着计算开销的显著增加。此外, 传统阴影检测方法在应对遥感图像时展现出局限性。遥感图像与普通图像在特征上的差异导致许多在普通图像任务中表现优异的网络模型在遥感图像上表现不佳。特别是, 遥感图像中的阴影常呈现不均匀分布和碎片化特点, 使得小面积阴影的检测变得尤为困难, 漏检现象频发。再者, 当阴影与图像中的其他特征 (如更亮或更暗的区域) 重叠时, 它在水体、植被以及深色物体等地物间的区分变得更为复杂。这种光谱上的混淆不仅容易导致非阴影区域被误判为阴影, 还可能使小面积阴影区域被完全忽略。因此, 如何有效区分阴影与其他地物, 并准确检测各种尺寸的阴影, 仍是该领域亟待解决的关键问题。

本文针对遥感图像阴影检测任务中的现有挑战, 提出了一种创新的双分支网络结构, 结合了 Transformer 与 CNN, 以实现更准确有效的遥感图像阴影检测。本方法通过 Transformer 网络, 有效

提取了遥感图像的全局信息,并利用自注意力机制捕获了图像中的长程依赖关系;而CNN分支则捕捉图像的局部特征信息和阴影细节纹理。为实现两个分支的优势互补,我们设计了一个全局-局部特征融合模块,确保模型能够同时关注阴影的精细结构和整体上下文。这种融合策略有望显著提升阴影检测的精确度。在训练过程中,我们设计了一个联合损失函数,该函数包括一个主损失函数和两个辅助损失函数,均以交叉熵损失为基础。这种设计不仅有助于加快模型的收敛速度,还能更好地指导模型的训练过程,从而实现更为理想的分割效果。通过优化这个联合损失函数,模型能够全面关注各个分支的学习情况,确保能够充分捕捉遥感图像中的阴影信息。

2 研究方法或原理

为了克服遥感图像阴影检测方法的现有不足,本文创新地提出了一种基于CNN卷积神经网络与Transformer的双分支遥感图像阴影检测网络。该网络设计的核心理念在于全面而有效地利用遥感图像的局部和全局特征信息。为实现这一目标,网络结合了CNN在局部特征提取方面的优势和Transformer在全局信息捕捉方面的能力。为了验证所构建网络的性能,本文在AISD遥感图像阴影检测数据集上进行了广泛的实验。实验的主要目的是评估网络的准确性和精确率,具体将通过一系列评估指标来衡量,如准确率(Precision)和召回率(Recall)等。此外,我们还将展示阴影检测的可视化效果,并对实验结果进行深入分析,以全面评估所提出网络在遥感图像阴影检测任务中的性能。

2.1 双分支遥感图像阴影检测网络

本文提出的双分支遥感图像阴影检测网络结构如图1所示,主要由五个关键部分构成:Transformer分支、CNN分支、特征融合模块、阴影预测模块以及损失函数。在这个网络中,本方法采用ResNet作为CNN分支的骨干网络进行特征提取,生成了四个不同分辨率的特征图 C_1 、 C_2 、 C_3 和 C_4 ,其分辨率依次为原始图像的1/4、1/8、1/16、1/32,且通道数依次成倍增加。这些特征图提供了丰富的图像纹理和局部信息。同时,Transformer分支则采用Swin Transformer作为特征

提取网络,同样生成了四个特征图 T_1 、 T_2 、 T_3 和 T_4 ,其分辨率与CNN分支相同,但通道数增加了一倍,从而包含了更多的图像全局信息。通过这样的设计,CNN分支能够提供详细的局部特征,而Transformer分支则能够捕捉全局上下文,共同为阴影检测任务提供全面而有效的特征表示。

在获取了Transformer和CNN的特征后,本方法通过全局-局部特征融合模块(Global-Local Fusion Model, GLFM)对 C_1 和 T_1 、 C_2 和 T_2 、 C_3 和 T_3 、 C_4 和 T_4 的特征进行了融合。为了实现特征的逐层融合与精炼,我们将上一层的GLFM输出经过一次转置卷积操作,使其通道数减半,同时宽高扩大到原来的两倍,再与下一层的GLFM输出进行特征相加。这种策略确保了CNN分支的局部信息和Transformer分支的全局信息得到有效保留与互补。具体来说,我们将Transformer的全局特征融入CNN分支,以增强其全局感知能力;同样地,也将CNN的局部特征融入Transformer分支,以提升其局部感知能力。融合后得到的 CT_1 、 CT_2 、 CT_3 、 CT_4 特征图兼具了两者的优势,从而更全面地描绘了遥感图像的特性。随后,这些融合后的特征信息被送入阴影预测模块中。该模块主要由 1×1 、 3×3 、 1×1 的卷积层与一个残差块构成,并引入了Sigmoid非线性激活函数对特征图进行逐像素的非线性变换。最终,通过像素二分类化处理得到精准的阴影检测图。这一流程不仅保留了全局与局部的关键信息,还通过特征交互与精炼提升了阴影检测的准确性。

最后,为了确保CNN和Transformer两个分支都能学习到有效的特征表示,本文分别引入了两个辅助损失函数来分别监督它们的学习过程(使用 $Loss_{aux_c}$ 来监督CNN分支的学习,以及 $Loss_{aux_t}$ 来监督Transformer分支的学习)。这样的设计能够确保两个分支都能够提取到具有良好表达能力的特征,从而使融合后的特征更加全面和强大,进一步提升网络的性能。这种设计思路不仅能够提升模型的学习效果,还能够显著增强模型对遥感图像阴影检测的准确性。

2.2 Swin Transformer

Swin Transformer (Liu, Lin, 等, 2021)是一种新型的层级式Transformer架构,成功地将在Transformer的全局信息建模能力与移动窗口技术

相结合。通过移动窗口设计，Swin Transformer 能够实现跨窗口的连接，使模型能够捕捉到相邻窗口间的关联信息，并执行有效的跨窗口特征交互。这一特性显著增大了模型的感受野，提高了特征信息的提取效率。

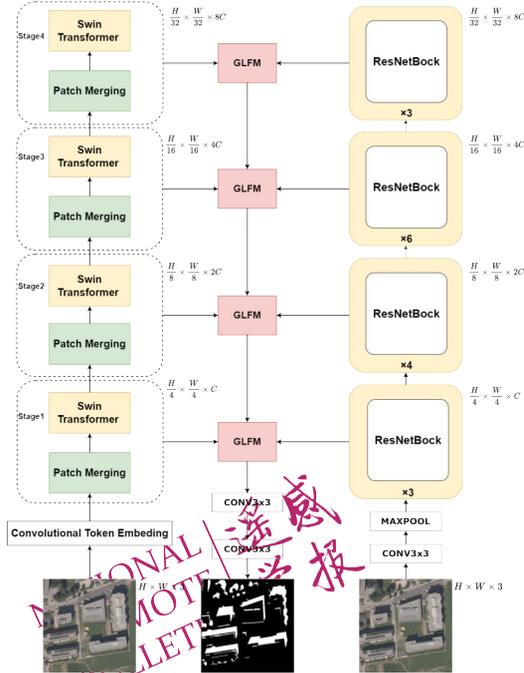


图1 网络结构图

Fig.1 The network structure diagram

传统的 Vision Transformer 方法通常将图像划分为较大的 Patch 以实现序列化，但这样做容易导致像素级信息的丢失。若对图像中的每个像素进行序列建模，将带来巨大的计算负担。为解决这一问题，Swin Transformer 采用了层级式的设计架构，避免了上述弱点。与 Vision Transformer 的 Patch 模式不同，Swin Transformer 逐层进行下采样，模拟了 CNN 中的滑动窗口、层次结构和池化等核心思想，同时保留了 Transformer 的全局建模能力。因此，在特征提取和信息建模方面，Swin Transformer 展现出了高效性和准确性，为遥感图像阴影检测等任务提供了强大的支持。其独特的架构和特性使得模型能够更好地捕捉遥感图像中的复杂阴影模式，从而实现更精确的阴影检测。

2.3 ResNet

为了解决深度 CNN 神经网络中的退化问题，Kaiming He 等人 (He 等, 2016) 提出了残差网络 ResNet。该网络通过引入残差连接，有效地将不

同卷积层的特征信息从浅层传播到深层。这种设计不仅减轻了梯度消失或爆炸的问题，还使得训练更深的网络模型成为可能。

如图 1 所示，ResNet 的结构从输入 $H \times W \times 3$ (3 表示 RGB 通道) 的图像开始，首先经过一个包含卷积层、批量归一化 (Batch Normalization, BN)、整流线性单元 (ReLU) 和 maxpool 的初始处理模块。随后，特征图依次经过四个阶段，生成 C_1 、 C_2 、 C_3 和 C_4 四个不同层次的特征。在每个阶段，特征图的分辨率通过降采样操作减半，而通道数则相应地增加一倍，以捕获更丰富的特征信息。

ResNet 的核心在于其残差块的设计。在每个阶段内，多个残差块串联起来，每个残差块都将输入特征与主分支输出的特征进行相加。这种连接方式有助于保留浅层特征信息，并在深层网络中实现有效的特征传播。此外，为了增强网络的非线性表达能力，ReLU 激活函数被广泛应用于残差块中。特别是在每个阶段的第一个残差块中，降采样卷积和带有 BN 的降采样残差连接被精心设计，以确保在降低特征图尺寸的同时能够提取到更多的有用特征。

瓶颈结构和残差连接的巧妙设计使得 ResNet 模型更易于训练，并能够成功地提升神经网络模型的深度和准确率。以 ResNet-50 为例 (如图 1 右侧所示)，它在每个阶段分别配置了 3、4、6 和 3 个残差块，实现了性能与计算复杂性的良好平衡。因此，在本文中，我们选择 ResNet-50 作为 CNN 分支的基础架构，以在处理复杂任务时保证模型性能的同时合理利用计算资源。

2.4 全局-局部特征融合模块

全局上下文信息由于能够捕获图像中的整体结构和语义关系，因此在复杂遥感图像的语义分割任务中发挥着至关重要的作用。然而，局部上下文信息同样至关重要，因为它能够保存丰富的空间细节并为准确的阴影检测提供关键的线索。为了充分利用这两种上下文信息，本文提出了一种高效的全局-局部特征融合模块 GLFM，旨在有效地融合全局和局部特征，以实现更精确的语义分割和阴影检测。

如图 2 所示，双分支遥感图像阴影检测网络构建了两个并行分支，分别专注于提取图像的全局和局部特征。这两个分支分别是 Swin Transformer

网络和 CNN 分支网络。为了实现全局和局部信息的有效交互,本文引入了 GLFM。GLFM 接收来自 Swin Transformer 和 CNN 分支的特征作为输入,通过其内部的卷积层、多头注意力模块、归一化层和激活函数等结构进行处理,实现了特征的高效融合。在上采样过程中,通过 GLFM 的融合引导, Swin Transformer 网络能够更精准地捕捉图像的局部细节特征信息,同时 CNN 分支网络也能够从 Swin Transformer 网络学习到丰富的全局特征信息。这样的设计不仅促进了两个分支之间的互补学习,还确保了全局和局部信息的全面利用。

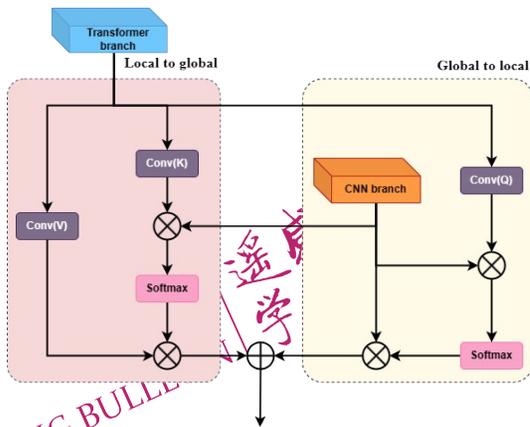


图2 全局-局部特征融合模块
Fig.2 Global-local fusion module

具体而言, GLFM 模块包含多头注意力模块以及全局和局部交互模块。它接收来自 CNN 分支(携带局部信息)和 Transformer 分支(携带全局信息)的特征作为输入。首先,这些特征经过 1×1 卷积和 Sigmoid 函数处理,将像素值归一化到 0 到 1 的范围内来消除特征值之间的极端差异。同时,通过额外的 1×1 卷积来对齐通道维度,确保特征的一致性。在全局到局部(Global to Local)的过程中,Transformer 分支的特征经过 Conv(Q) 卷积作为查询(Query),而 CNN 分支的特征则分别作为键(Key)和值(Value)。通过计算查询和键的点积,得到注意力图,该图表示了 Transformer 分支的每个特征对 CNN 分支的每个特征的重要性。这个注意力图随后与 CNN 分支的值进行乘法运算,得到加权后的特征,实现了全局信息对局部特征的加权。相对应地,在局部到全局(Local to Global)的过程中, CNN 分支的特征被用作查询,而 Transformer 分支的特征则经过 Conv(K) 和

Conv(V) 卷积操作后分别作为键和值。类似地,通过计算查询和键的点积来生成注意力图。这两个过程(Global to Local 和 Local to Global),同时进行,并进行加权融合。最后,应用 1×1 卷积得到输出特征。这种设计在 CNN 和 Transformer 之间建立了有效的连接,促进了局部和全局信息的交互融合,从而有助于优化各自分支的性能,并最终提升了阴影检测的准确性。

2.5 阴影预测模块

阴影预测模块是本文提出的关键组件之一,其详细结构如图 3 所示。该模块主要由一系列 1×1 、 3×3 、和 1×1 的卷积层构成,并引入了一个简单的残差模块以增强特征表示。在经过双分支网络(包含全局和局部特征)处理后,我们得到了由多个 GLFM 输出特征相加、通道数为 128 的特征图,这些特征图蕴含了丰富的全局和局部上下文信息。阴影预测模块的核心任务是把这些特征图恢复到原始的分辨率,并生成一个 256×256 大小的阴影预测图。为实现这一目标,本文采用了特定的解码器上采样策略。在解码阶段,为了更好地保留细节特征信息,利用 1×1 卷积运算来调整通道数以实现特征的跨通道交互和信息整合。此外,本文引入 Sigmoid 非线性激活函数对特征图进行逐像素的非线性变换,以增强网络的表达能力。通常,我们将 Sigmoid 输出概率值大于 0.5 的像素判定为阴影区域,而输出概率值小于或等于 0.5 的像素则视为非阴影区域。

最后,通过像素二分类处理,将连续型的概率图转换为离散的阴影检测效果图。这一步骤可视为对网络输出的后处理操作,旨在使结果更加直观且易于理解。通过这样的处理流程,阴影预测模块能够有效地利用全局和局部特征进行阴影检测,并输出高质量的预测结果。

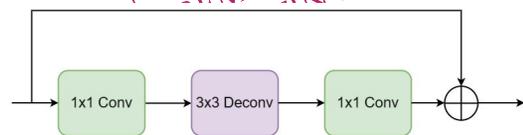


图3 阴影预测模块
Fig.3 Shadow prediction module

2.6 联合损失函数

损失函数在深度学习模型的优化过程中起着

至关重要的作用，对于阴影检测任务的最终分割效果具有决定性的影响。为了更有效地训练和优化模型，本文精心设计了一个联合损失函数。该函数由一个主损失函数和两个辅助损失函数组成。主损失函数主要负责监督整合后的网络模型的整体训练和优化过程，确保模型能够准确地学习阴影特征。而辅助损失函数则分别用于监督各自分支的优化过程，帮助模型更好地捕捉全局和局部上下文信息。具体而言，每个分支的损失函数均采用交叉熵损失函数。交叉熵损失函数在分类任务中广泛应用，能够有效地度量模型预测概率分布与真实概率分布之间的差异。通过最小化交叉熵损失，我们可以加速模型的收敛速度并提高阴影检测的准确性。公式(1)展示了交叉熵损失函数的具体形式：

$$Loss_{CT} = -\frac{1}{N} \sum_{i=1}^N y_i \cdot \log(p(y_i)) + (1 - y_i) \log(1 - p(y_i)) \quad (1)$$

其中， N 代表图像中像素的总数量； y_i 代表遥感图像中第 i 个像素的标签值，正类（阴影像素）为1，负类（非阴影像素）为0， $y_i \in \{0, 1\}$ 代表遥感图像中的第 i 个阴影像素， $p(y_i)$ 代表第 i 个像素预测为阴影像素的概率。

此外，我们选择焦点损失作为辅助损失函数，如公式(2)所示：

$$Loss_{aux_c} = Loss_{aux_r} = -y_i(1 - p_i)^\gamma \cdot \log p_i - (1 - y_i)p_i^\gamma \log(1 - p_i) \quad (2)$$

其中， γ 是聚焦参数，用于调整权重变化较大的分类样本，在本文的实验中 γ 设定为2。CNN分支的辅助损失函数（ $Loss_{aux_c}$ ）和Transformer分支的辅助损失函数（ $Loss_{aux_r}$ ）作为局部损失函数，共同监督各自分支的优化过程。

最终，整个模型的损失函数（ $Loss_{All}$ ）如公式(3)所示，它是主损失函数和辅助损失函数的加权和：

$$Loss_{All} = Loss_{CT} + Loss_{aux_c} + Loss_{aux_r} \quad (3)$$

这一联合损失函数的设计使得我们可以在网络的不同分支上施加不同的监督作用，有助于加速训练过程并提高模型的鲁棒性和泛化能力。通过这种策略，我们期望在保证模型精度的同时，也使其更具实用性和适应性。

3 实验过程及结果分析

3.1 数据集

本文的网络模型在专门为遥感图像阴影检测任务设计的AISD数据集上进行了训练和测试。AISD数据集以其高分辨率和多样的地表覆盖类别而著称，涵盖了奥斯汀、芝加哥、蒂罗尔、维也纳和因斯布鲁克这五个世界上不同地区的城市结构特征和照明条件。这种多样性对提升网络训练的鲁棒性至关重要。

为了有效地划分训练和验证数据，AISD数据集按照8:1:1的比例被划分为训练集、验证集和测试集。具体而言，训练集包含412张图像及其对应的阴影掩码，用于模型的训练；验证集和测试集则各包含51张图像及其对应的阴影掩码，分别用于模型的验证和性能评估。

认识到大规模数据集在训练高性能深度学习模型中的关键作用，本文对AISD数据集进行了广泛的数据增强处理。这包括图像的翻转、缩放、平移、旋转、裁剪以及多种滤波处理。经过这些增强操作后，训练集图像数量增至11836张，验证集图像数量增至1100张。这种数据增强策略有助于提升模型的鲁棒性和泛化能力，使网络模型在更广泛的场景中实现更有效的阴影检测。

在实验过程中，所有输入图像均被统一调整为 $3 \times 256 \times 256$ 的大小。实验基于PyTorch框架在Ubuntu平台上进行，并利用NVIDIA GeForce RTX 3090 24G显卡加速计算。为了提高训练效率和速度，实验中设置了batch size为8。这些实验细节为复现本文结果提供了重要参考。

3.2 评价指标

为了全面评估本文提出的遥感图像阴影检测方法的性能，我们采用了多个评价指标进行衡量。具体如下：

(1) TP (True Positive)、TN (True Negative)、FP (False Positive)、FN (False Negative)

这些基础指标用于描述阴影和非阴影部分的检测情况。其中，TP表示正确检测到的阴影像素数量；TN表示正确检测到的非阴影像素数量；FP表示错误检测到的阴影像素数量；FN表示错误检测到的非阴影像素数量。这些指标为后续的评价

提供了基础数据。

(2) 准确度 (Accuracy)

准确度反映了模型在识别阴影像素时的整体性能,即正确检测的阴影像素与总像素的比例。

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

(3) 精确度 (Precision)

精确度衡量了模型预测为阴影像素的像素中,实际确实为阴影的比例。高精度意味着模型较少地将非阴影像素误判为阴影。

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

(4) 召回率 (Recall)

召回率又称为查全率,表示模型正确检测到的阴影像素与所有真实阴影像素的比例。高召回率意味着模型能够检测到更多的真实阴影像素。

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

(5) F1分数 (F1-Score)

F1分数是精确度和召回率的调和平均值,提供了一个综合性的性能指标。它同时考虑了精度和召回率,评价了模型的整体性能。

$$F1\ score = \frac{2 * Precision * Recall}{Precision + Recall} \times 100\% \quad (7)$$

(6) 平衡错误率 (Balance Error Rate, BER)

平衡错误率是阴影检测领域的公认评价标准,BER的值越低,说明模型的阴影检测结果越好,即模型的错误率越低。BER提供了一个标准化的方式,以比较不同方法在同一数据集上的性能。

$$BER = \left(1 - 0.5 \times \left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP} \right) \right) \times 100\% \quad (8)$$

通过这些评价指标,我们能够全面、深入地分析本文提出的遥感图像阴影检测方法的性能,并与其他方法进行对比。

3.3 实验结果对比分析

为了全面评估本文提出的基于Transformer和CNN双分支遥感图像阴影检测网络的有效性,本文进行了一系列对比实验。参与对比的方法包括了一些具有代表性的深度神经网络,如FCN、SegNet、U-Net这三种经典的语义分割模型。选择这些模型进行对比是因为阴影检测在本质上可以被视为一种语义分割问题。此外,本文还对比了一种流行的深度学习阴影检测方法,即FSDNet

(Hu等,2021),以及专门针对航空图像阴影检测的三种先进网络方法:CAD、MSASDNet(Liu等,2021)和DSSDNet。所有对比实验均在AISD数据集上进行,并采用了相同的评价指标以确保公平性。实验结果如图4所示,我们的方法在小阴影区域的检测精度上明显优于其他对比方法。具体来说,我们的方法能够有效区分光源下的黑暗物体和真实阴影下的光亮物体,展现出较高的精确度和召回率。此外,在处理弱光照区域与阴影真实区域时,我们的方法也能够正确辨别,进一步证明了其卓越的性能。这一系列对比实验的结果充分证明了本文所提出方法在阴影检测任务上的显著优势。

为了使对比结果更具说服力,本文进一步进行了定量的比较分析。具体数据如表1所示,其中包括了多个关键评价指标的对比结果。通过分析这些数据可以看出,本文提出的基于Transformer与CNN融合的双分支遥感图像阴影检测方法在多个评价指标上都显著优于大多数现有阴影检测方法。这一结果证明了本文所提出方法的有效性和优越性。

具体而言,在与其他阴影检测方法的比较中,三种经典的语义分割网络(FCN、SegNet、U-Net)的评价指标明显较低。它们的F1-score分别仅为0.826、0.832和0.852,而本文提出的方法的F1-score达到了0.931,提升了3.7%至5.9%。在BER方面,这些方法的数值分别为8.238、9.312和10.095,而本文的方法的BER值仅为4.301,降低了3.937至5.794,显示出在减少错误率方面的显著优势。

与近两年流行的遥感图像阴影检测方法(CAD、DSSDNet、MSASDNet)相比,本文提出的方法同样表现出色。具体而言,在F1-score指标上,我们的方法提高了0.7%至1.3%,在Accuracy指标上提高了0.116%至1.545%,而在BER指标上降低了0.363至1.520。这些结果进一步证明了本文提出的方法在处理复杂和多样化阴影问题方面的优越性。

值得注意的是,本文的方法在精度与召回率的差值上仅为0.851%,而其他方法的这一差值范围为1.352%至8.883%。这一结果充分显示了我们的方法在平衡阴影检测的精度和召回率方面的卓越性能。这也从侧面证明了我们的双分支网络设

计以及损失函数设计的有效性。

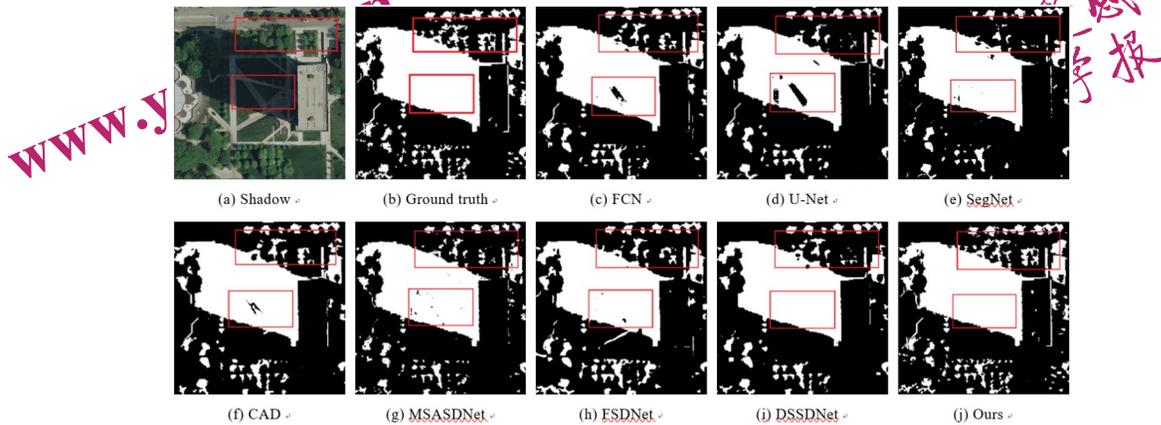


图4 不同网络模型在 AISD 数据集上阴影检测定性比较

Fig.4 Qualitative comparison of shadow detection on AISD dataset of different network models

表 1 不同网络模型在 AISD 数据集上阴影检测定量比较

Table 1 Quantitative comparison of shadow detection on AISD dataset of different network models

方法	F1	Accuracy	Precision	Recall	BER
FCN	0.826	92.349	88.416	80.487	9.238
SegNet	0.832	93.186	88.254	82.891	9.312
U-Net	0.852	93.778	90.971	84.478	10.095
CAD	0.924	96.236	93.037	91.239	5.046
MSASDNet	0.922	96.996	93.455	91.103	4.664
FSDNet	0.881	95.567	92.012	89.553	6.242
DSSDNet	0.918	95.732	93.122	90.584	5.821
Ours	0.931	97.112	92.972	93.121	4.275

综上所述，通过对这些评价指标的详细比较，我们充分证明了本文提出的方法在遥感图像阴影检测任务中具有出色的性能。

3.4 消融实验结果分析

为了更深入地探讨 Transformer 与 CNN 融合的双分支在遥感图像阴影检测中的贡献，我们进行了细致的消融研究。通过逐步叠加网络模型的各个模块，我们详细观察了每个模块对最终阴影检测结果的影响。消融实验的具体结果如表 2 所示。

表 2 AISD 数据集上的消融实验

Table 2 Ablation experiments on the AISD dataset

方法	F1	Accuracy	BER
SwinTransformer	0.912	96.405	5.212
ResNet	0.879	94.168	8.312
SwinTransformer+ResNet	0.923	96.993	5.019
SwinTransformer+ResNet+GLFM	0.931	97.112	4.275

实验数据表明，单独依赖 Transformer 网络或 CNN 网络时，其准确率明显低于两者融合后的模型。在融合后，F1 值和准确率都有显著的提升，这充分证明了双分支融合的有效性。此外，当在融合模型中引入 GLFM 后，准确率进一步提升了 0.707%，而 BER 值降低了 0.778，这进一步体现了在网络中引入 GLFM 的优越性。

从表中的各项评价指标来看，本文提出的方法在 F1 值和准确率上都展现出了优越的性能，较融合前有显著的提高；同时，BER 值的降低也体现了模型在阴影检测任务上的高效性。

此外，消融实验效果如图 5 所示，通过图中的绿色方形框可以直观看出，经过完整的双分支融合网络处理后，模型对小阴影的检测能力得到了显著增强。与未引入 GLFM 的模型相比，误检率与漏检率也有明显的降低。

总结来说，本次消融研究通过实验数据和图像结果两个方面全面验证了本文设计的各个网络模块的有效性，并揭示了它们在提升阴影检测性能方面的重要性。

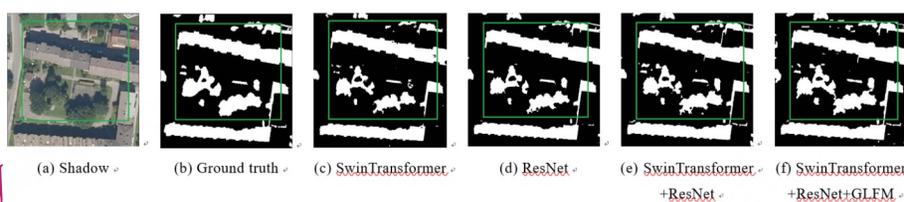


图5 在遥感图像阴影检测 AISD 数据集上的消融性实验效果图

Fig.5 Rendering of ablation experiments on an AISD dataset for shadow detection in remote sensing images

4 结论

本文致力于提高遥感图像阴影检测的实际效果,降低漏检率和误检率,并解决网络复杂性等挑战。针对卷积神经网络在提取图像信息时的局限性,我们创新地提出了一种结合 Transformer 与 CNN 的双分支遥感图像阴影检测方法。该方法使得 CNN 网络能够从 Transformer 网络获取全局信息特征,而 Transformer 网络则能从 CNN 网络获取局部信息特征。此外,我们还引入了一个图像全局-局部特征融合模块,以促进遥感图像的局部信息和全局信息之间的交互学习,进而提升阴影检测的准确性。在 AISD 遥感图像数据集上的实验结果显示,我们的方法显著降低了遥感图像阴影检测的错误率和漏检率,提高了阴影检测的准确性。具体来说,网络的准确率最高可达 97.112%, 错误率最低为 4.275% (详见实验部分), 这些指标均优于当前其他现有模型。此外,我们还通过消融实验验证了双分支遥感图像阴影检测网络的有效性 (详见消融研究部分)。然而,我们也意识到本文所提方法仍有提升空间。当前遥感图像阴影检测任务中仍存在网络参数量大、计算复杂度高等问题。未来,我们将集中精力解决上述问题,以期进一步提升遥感图像阴影检测的性能和效率。总体而言,该算法在遥感图像阴影检测任务中展现出了优秀的性能表现和实际应用价值。

参考文献 (References)

Badrinarayanan V, Kendall A, Cipolla R. 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12): 2481-2495 [DOI: 10.1109/TPAMI.2016.2644615]

Firdaus-Nawi M, Noraini O, Sabri M Y, Siti-Zahrah A, Zamri-Saad M and Latifah H. 2011. DeepLabv3+ _encoder-decoder with Atrous separable convolution for semantic image segmentation. *Pertanika Journal of Tropical Agricultural Science*, 34(1): 137-143.

Guan Y, Chen X A, Tian J and Tang Y. 2022. Shadow Detection for Remote Sensing Images//Proceedings of the 2022 6th International Conference on Compute and Data Analysis (ICCA). Shanghai, ACM: 68-74. [DOI: 10.1145/3523089.3523101]

He K, Zhang X, Ren S and Sun J. 2016. Deep residual learning for image recognition//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, IEEE: 770-778. [DOI: 10.1109/CVPR.2016.90]

Hu X, Wang T, Fu C W, Jiang Y, Wang Q and Heng P A. 2021. Revisiting shadow detection: A new benchmark dataset for complex world. *IEEE Transactions on Image Processing*, 30: 1925-1934. [DOI: 10.1109/TIP.2021.3049331]

Jin Y, Xu W, Hu Z, Jia H, Luo X and Shao D. 2020. GSCA-UNet: Towards automatic shadow detection in urban aerial imagery with global-spatial-context attention module. *Remote Sensing*, 12(17): 2864. [DOI: 10.3390/rs12172864]

Li H, Zhang L, Shen H. 2013. An adaptive nonlocal regularized shadow removal method for aerial remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 52(1): 106-120. [DOI: 10.1109/TGRS.2012.2236562]

Li R, Zheng S, Duan C, Wang L and Zhang C. 2022. Land cover classification from remote sensing images based on multi-scale fully convolutional network. *Geo-spatial Information Science*, 25(2): 278-294. [DOI: 10.1080/10095020.2021.2017237]

Liu D, Zhang J, Liu K and Zhang Y. 2022. Aerial remote sensing image cascaded road detection network based on edge sensing module and attention module. *IEEE Geoscience and Remote Sensing Letters* 19: 1-5. [DOI: 10.1109/LGRS.2022.3190495]

Liu D, Zhang J, Wu Y and Zhang Y. 2021. A shadow detection algorithm based on multiscale spatial attention mechanism for aerial remote sensing images. *IEEE Geoscience and Remote Sensing Letters* 19: 1-5. [DOI: 10.1109/LGRS.2021.3100294]

Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z and Guo B. 2021. Swin transformer: Hierarchical vision transformer using shifted windows//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, IEEE: 10012-10022. [DOI: 10.1109/ICCV48922.2021.00986]

Long J, Shelhamer E, Darrell T. 2015. Fully convolutional networks for semantic segmentation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, IEEE: 3431-3440. [DOI: 10.1109/CVPR.2015.7298965]

Luo S, Li H, Shen H. 2020. Deeply supervised convolutional neural network for shadow detection based on a novel aerial shadow im-

- agery dataset. ISPRS Journal of Photogrammetry and Remote Sensing, 167: 443-457. [DOI: 10.1016/j.isprsjprs.2020.07.016]
- Luo S, Li H, Zhu R, Gong Y and Shen H. 2021. ESPFNet: An edge-aware spatial pyramid fusion network for salient shadow detection in aerial remote sensing images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14: 4633-4646. [DOI: 10.1109/JSTARS.2021.3066791]
- Ronneberger O, Fischer P, Brox T. 2015. U-net: Convolutional networks for biomedical image segmentation//Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015): 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, 234-241. [DOI: 10.1007/978-3-319-24574-4_28]
- Shao Z, Wang L, Wang Z and Deng J. 2019. Remote sensing image super-resolution using sparse representation and coupled sparse auto-encoder. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12(8): 2663-2674. [DOI: 10.1109/JSTARS.2019.2925456]
- Ufuktepe D K, Collins J, Ufuktepe E, Fraser J, Krock T and Palaniappan K. 2021. Learning-based shadow detection in aerial imagery using automatic training supervision from 3D point clouds//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, IEEE: 3926-3935. [DOI: 10.1109/ICCV54120.2021.00439]
- Walker H J. 2006. Evelyn Lord Pruitt, 1918 - 2000. Annals of the Association of American Geographers, 96(2): 432-439. [DOI: 10.1111/j.1467-8306.2006.00487.x]
- Xu Z, Zhang W, Zhang T, Yang Z and Li J. 2021. Efficient transformer for remote sensing image segmentation. Remote Sensing, 13(18): 3585. [DOI: 10.3390/rs13183585]
- Zali S A, Mat-Desa S, Che-Embi Z and Mohd-Isa W N. 2022. Post-Processing for Shadow Detection in Drone-Acquired Images Using UNET. Future Internet, 14(8): 231. [DOI: 10.3390/fi14080231]
- Zhang Y, Chen G, Vukomanovic J, Singh K K, Liu Y, Holden S and Meentemeyer R K. 2020. Recurrent Shadow Attention Model (RSAM) for shadow removal in high-resolution urban land-cover mapping. Remote Sensing of Environment, 247: 111945. [DOI: 10.1016/j.rse.2020.111945]
- Zhao H, Shi J, Qi X, Wang X and Jia J. 2017. Pyramid scene parsing network//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, IEEE: 2881-2890. [DOI: 10.1109/CVPR.2017.660]
- Zhu Q, Yang Y, Sun X and Guo M. 2022. CDANet: Contextual detail-aware network for high-spatial-resolution remote-sensing imagery shadow detection. IEEE Transactions on Geoscience and Remote Sensing, 60: 1-15. [DOI: 10.1109/TGRS.2022.3143886]

Application of Dual-Branch Network Integrated with Transformer and CNN in Remote Sensing Image Shadow Detection

WANG Yifan, HUANG Xian, WANG Jianlin, ZHOU Tong, ZHOU Wenjun, PENG Bo

School of Computer Science and Software Engineering, Southwest Petroleum University, Chengdu 610500, P.R. China

Abstract: (Objective) Shadows in remote sensing imagery play a crucial role in image interpretation and feature extraction but are also known to introduce significant challenges in image analysis. Traditional methods often struggle with complex shadow scenarios, leading to missed or false detections. This paper introduces a novel approach that enhances shadow detection accuracy and reliability in high-resolution remote sensing images. (Method) The proposed dual-branch network synergistically combines the strengths of Transformer and Convolutional Neural Networks (CNNs) to tackle the challenges of shadow detection. The network leverages a Transformer branch to capture global contextual relationships and a CNN branch to emphasize local textural details. This architecture is designed to exploit the complementary nature of global and local information, providing a comprehensive feature representation. This method also introduces a shadow prediction module that integrates these features for effective shadow segmentation. A joint loss function, comprising a primary loss and auxiliary losses, is utilized to refine learning and accelerate convergence, thereby enhancing the detection accuracy. (Result) The proposed method was rigorously tested on the Aerial Imagery Shadow Dataset (AISD), demonstrating substantial improvements in shadow detection metrics. It achieved a shadow detection accuracy of 97.112% and significantly reduced the false detection rate, with a balance error rate (BER) decrease of 0.389. These results not only validate the effectiveness of the dual-branch architecture but also showcase the advantages of integrating global and local features through our innovative network design. (Conclusion) The dual-branch network provides a robust solution to the perennial challenges of shadow detection in remote sensing imagery. By effectively minimizing missed and false detections, the network holds significant promise for enhancing the interpretability and utility of high-resolution satellite images in various applications, such as urban planning and environmental monitoring. The future work will focus on optimizing the network architecture and exploring its applicability to other complex imaging conditions.

Key words: remote sensing image, shadow detection, semantic segmentation, dual-branch network, feature integration, Transformer, CNN,

ResNet

Supported by Supported by Sichuan Science and Technology Program (No.2023NSFSC1393, No.2023NSFSC0504), Scientific Research Starting Project of SWPU (No.2021QHZ001)

www.ygxb.ac.cn

NATIONAL
REMOTE
SENSING BULLETIN | 遥感学报

NATIONAL
REMOTE
SENSING BULLETIN | 遥感学报

www.ygxb.ac.cn

www.ygxb.ac.cn

NATIONAL
REMOTE
SENSING BULLETIN | 遥感学报