引文格式：王南,钱雨卉,魏维轩. 城市公园植被高精度多光谱遥感识别模型构建[J]. 地理空间信息, 2024, 22(4): 71-75. doi:10. 3969/j.issn. 1672-4623.2024.04.016

城市公园植被高精度多光谱遥感识别模型构建

王　南[1]，钱雨卉[1]，魏维轩[2*]
（1. 南京工业大学建筑学院，江苏 南京 211800；2. 上海济致建筑规划设计有限公司，上海 200040）

摘　要：利用无人机与多光谱相机的组合航测获取可见光与多光谱影像，标注多光谱GNDVI图像和三通道可见光图像，采用图像融合网络、双重注意力网络和特征解码网络实现双重注意力卷积神经网络模型运算，并以上海泡泡公园做实证研究。基于早期融合（early fusion）和特征融合（feature fusion）的双重注意力卷积神经网络模型的整体植被识别率可达91.8%。先进设备的组合和双重注意卷积神经网络学习可实现高精度识别的模型构建，提升城市公园植被高精度识别的能力，可为后续的公园碳汇测算与种植评估提供更为精准的科学技术支撑。

关键词：城市公园植被；多光谱遥感；高精度识别模型；双重注意力卷积神经网络；碳汇
中图分类号：P237　　**文献标志码：**B　　**文章编号：**1672-4623（2024）04-0071-05

High-precision Multi-spectral Remote Sensing Detection Modeling of Urban Park Vegetation

WANG Nan[1], QIAN Yuhui[1], WEI Weixuan[2]
(1. School of Architecture, Nanjing Tech University, Nanjing 211800, China;
2. Jizhi Architecture Planning and Design Co., Ltd., Shanghai 200040, China)

Abstract: With the assistance of unarmed aerial vehicle and multi-spectral camera, we labelled GNDVI image and three-channel image. And then, we used image fusion network, dual attention network and feature decoding network to implement dual attention convolutional neural network (DANet) modeling. Finally, taking Shanghai Bubble Park for example, we conducted empirical research. The results show that DANet model based on early fusion and feature fusion can identify 91.8% of total vegetation. The combination of advanced equipment and DANet can construct high-precision detection model of urban park vegetation, and provide more accurate scientific and technological support for the follow-up park carbon sink estimation and planting evaluation.

Key words: urban park vegetation, multi-spectral remote sensing, high-precision detection model, DANet, carbon sink

1　评估的重要性

“增加生态碳汇”的自然调节是实现双碳目标的主要途径，但在碳汇测算中往往以林木生长量为主[1]，忽视了作为城市重要绿色基础设施的公园绿地中植被的碳汇功能[2]。与较为成熟的林木碳汇测算相比，城市公园绿地植被的碳汇测算与评估理论、方法、模型、实践相对匮乏，尤其表现在公园植被识别方面，但先进设备的组合使用与智能算法的深度优化或可为公园植被的识别与后续的碳汇测算、种植设计等提供有力的数理依据，快速响应双碳政策，发挥城市公园绿地生态效能。

2　公园植被高精度识别的研究进展

2.1　公园植被识别的主要方法

公园植被的识别与传统的林地识别方法有较大差异。林地植被往往是单一种或以成熟乔木林为主，成组团化，边界清晰、要素明确、植被识别相对简单[3]，而公园绿地中的植被常常大小不一，种植种类多、分布复杂、界面多样，故需借助更为准确的模型以实现高精度识别。

现有识别方法中，现场测量法精度较高但工作量大，设计图纸测量法工作量小，但因植物实际的生长范围变化和勘察技术限制导致误差较大，均不能较好

收稿日期：2022-11-13。
项目来源：国家自然科学基金青年科学基金资助项目（72104106）。
第一作者简介：王南（1984—），博士，硕士生导师，研究方向为风景园林规划设计，E-mail：sindywang1215@qq.com。
通信作者：魏维轩（1989—），博士研究生，高级工程师，研究方向为风景园林规划设计与机器学习，E-mail：430146@qq.com。

地评估各类植物的生长状态，降低了精确度，进而影响了后续碳汇测算的可靠性[4]。

2.2 多光谱遥感在植被识别中的应用

自1970年代起，随着多光谱遥感技术和无人机与高精度相机等设备的出现，通过研究SPOT和TM的多光谱图像各波段的相关性及最佳波段组合[5]，多光谱图像逐渐应用于地质灾害管理、矿产资源、城市建设等领域，并向精细农业[6]、环境评价、数字城市等新兴行业应用转移。近年来，无人机多光谱遥感影像开始应用于农业检测[7]和植被生态识别[8]，结合植被覆盖指数对植被威胁区域进行预警，融合框架、算法以生成具有高空间质量的多光谱图像[9]，但考虑到公园植被的复杂性与动态性，在公园植被识别中的应用较少。

2.3 多光谱遥感植被图像的机器学习

多光谱遥感获得的图像需经数据格式转换和针对性分析后才可为植被识别所用，主要包括基于人工设计的人工描述子分类算法和基于深度学习的图像分类算法，后者更能够排除差异、适应干扰。自2006年深度置信网络提出以来[10]，卷积神经网络的图像识别功能已可逐渐实现多光谱遥感的影像分割[11]。

但在这些场景识别中常采用的多路卷积神经网络，仅从近红外图像和三通道彩色图像多通道输入识别[12]，对城市公园植被识别的场景适用性不强。因此，基于双重注意力机制，针对不同通道所代表特征的重要性程度和不同局部区域的显著性程度处理遥感图像[13]，或可成为高精度识别城市公园植被的新路径。

3 模型构建

3.1 适用多光谱识别的植被分析指数

绿化植被相对于公园内其他非植物要素具备较好的多光谱识别优势，通常采用NDVI指数进行分析，但NDVI对土壤背景的变化较为敏感，对公园非植物识别存在客观偏差，故采用更为稳定的GNDVI指数对多波段数据进行整合，构建用于辅助三通道可见光图像识别的GNDVI图像。

$$GNDVI=(NIR-GREEN)/(NIR+GREEN) \quad (1)$$

式中，NIR为840 nm近红外波段；GREEN为555 nm绿色波段。

3.2 双重注意力卷积神经网络建模

卷积神经网络的注意力机制能够让网络内神经元专注于指定特征属性，有指向性地进行网络参数优化，而双重注意力的卷积神经网络结构具备了自适应地集成局部特征和全局依赖的能力，更适应三通道彩色图像中多光谱和广域空间的两方位信息。

在通道注意力和空间注意力2个模块中：通道注意力模块强调图层之间信息关联性，对于特征的每一个通道都使用专门的检测器进行分析，之后通过池化或通道压缩方案综合所有通道的信息，既实现了通道信息的独立化提取，又保证通道关联性不丢失。空间注意力模块在强调了卷积核局部内空间关系的基础卷积神经网络基础上，提出了范围更广的空间感知方案，在每一层图像空间范围内进行自卷积和点乘，计算空间内每个像素与其他像素的相关性，让神经网络认识到重点特征位置，提高卷积层内关注点的表现力，抑制不必要的特征。

模型将2个模块并行组合，在对基础特征网络提取的特征以向量模式重构后，注意力机制为不同向量和空间的像素赋予不同的注意力权重，利用神经网络的自我学习能力分析信息关联性。对多光谱以及三通道可见光图像的识别任务，双重注意力神经网络既强调了多光谱图像的光谱信息，又降低了广域信息分析的难度，提高了特征差异性（图1）。

4 双重注意力网络高精度神经网络模型构建

4.1 神经网络模型构建

网络结构分为三部分（图2）：图像融合网络、双重注意力网络以及特征解码网络。整体结构按编解码结构设计，自顶向下编码器块对输入特征图谱进行提取，而自底向上的解码器将输入特征优化，生成符合训练要求的分割解译结果。考虑到图像的多尺度特性，本网络利用特征提取基础网络本身带有的层次性

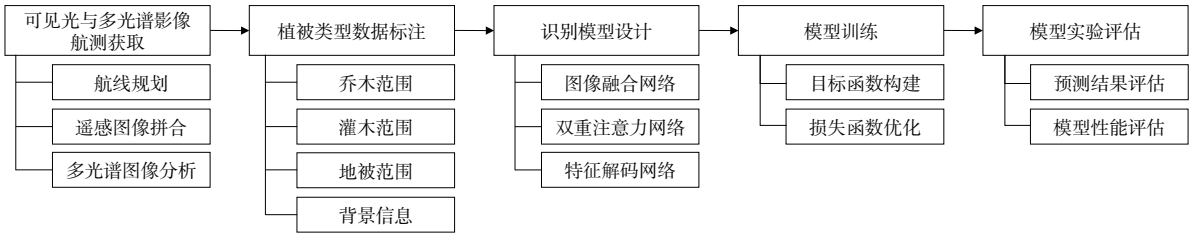


图1 双重注意力网络高精度识别城市公园植被的技术路径

语义特征，借助特征金字塔网（feature pyramid network）分离并优化基础特征，构建特征金字塔。通过侧向连接方式融合编、解码器的特征，利用顶层抽象语义特征和底层的高分辨率细节特征，实现语义分割结果精细化。

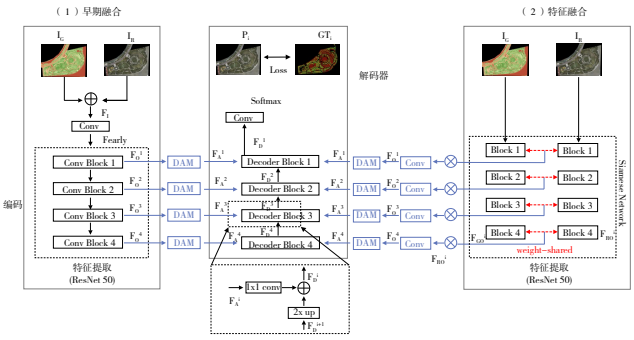


图2　神经网络整体结构图

4.1.1　图像融合网络

图像融合网络的作用是对多光谱图像生成的GNVDI图像，以及可见光图像基础特征的提取与融合。早期融合（early fusion）数据模块是在2个图像基于通道进行融合后，将特征输入基础特征网络；特征融合（feature fusion）数据模块是在基础网络后，对提取的基础特征进行融合。

4.1.2　双重注意力网络

由基础特征网络ResNet块提取的输出 F_O^i，无论是来自早期融合模块或特征融合模块，都被输入到一个双重注意模块（dual attention module）内。多个注意力模块形成单侧链接链路，搭建多尺度特征金字塔，实现编码器特征与解码器特征的跳层连接（skip connection）。每个双重注意模块由空间注意力模块和一个通道注意力模块组成（图3）。

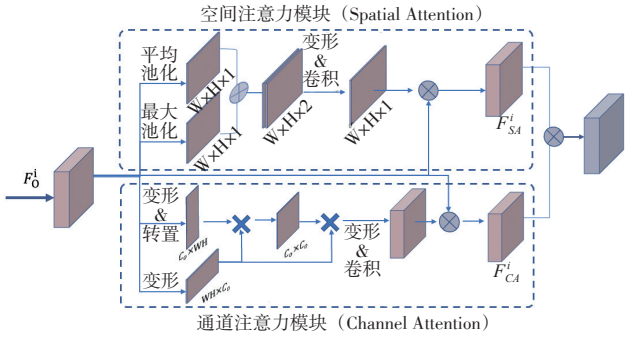


图3　双重注意力网络结构图

4.1.3　特征解码网络

位置注意力模块将所有像素位置纳入分析过程，通过特征的加权和来选择性地聚合每个位置的特征。无论距离如何，类似的特征都将彼此相关。通道注意力模块通过整合所有通道映射之间的相关特征来选择性地强调存在相互依赖关系的通道映射。最后，2个注意力模块的输出被相加后经过数个3×3的卷积神经层处理，生成该层级的解码器对应的输入特征 F_A^i：

$$F_A^i = H\left(F_{SA}^i \oplus F_{CA}^i\right) \tag{2}$$

不同层级的注意力特征 F_A^i 首先经过1×1的卷积改变特征图的通道数，之后通过上采样改变特征图空间大小，与底层特征相加融合并卷积整理。此过程重复操作，实现自下向上的特征融合，如此操作可以逐步改进特征表示，优化分割结果：

$$F_D^i = \begin{cases} H\left(H\left(F_A^i\right) \oplus F_D^{i+1}\right), i \neq \text{bottom} \\ F_A^i, i = \text{bottom} \end{cases} \tag{3}$$

之后获得每一层级的输出 F_D^i，经过一个 $1 \times 1 \times C$ 的卷积层和softmax激活函数层，获得每一类目标在该像素出现的概率图 P^i，该概率图与输入的原始标注大小相同，其中 C 为需要区分的目标类别数，i 为特征层级。

4.2　模型训练与损失函数

基于监督学习方案进行模型的训练为实现多尺度损失获取，每一级融合后的解码特征图 F_D^i 与目标标注图 GT^i 进行对比，构建多尺度损失函数，其中 GT^i 为经过下采样处理的标注产品。目标函数 $Q\left(P^i, GT^i\right)$ 构造为如下所示：

$$Q\left(P^i, GT^i\right) = -\frac{1}{N \times L}\sum_i^L \sum_n^N p^{i,n} \log\left(GT^{i,n}\right) + \left(1 - p^{i,n}\right)\log\left(1 - GT^{i,n}\right) \tag{4}$$

式中，$p^{i,n}$ 为第 i 层特征中，第 n 种目标的概率；$GT^{i,n}$ 为进行下采样后该位置标注是否已标注为第 n 类，$GT^{i,n} = 1$；L 为基础网络包含的卷积块总数；N 为总类别数。

在训练过程中，使用Adam函数进行目标函数优化，算法学习率采用余弦退火（cosine annealing）进行更新，初始学习率为 $1e^{-5}$，权值下降设定为 $5e^{-4}$，共迭代训练60 000次，批量为4。

考虑到样本不平衡问题，使用加权交叉熵思想优化损失函数，即对于不同目标使用不同的损失权重。对于小数目样本如乔木，提高其权重，加强它对损失函数的贡献；对于大数目的样本，比如环境背景（包括周围环境和裁剪中的黑色像素）减少其对损失函数的贡献。优化损失函数如下，主要在正样本上进行加权处理：

$$w^n(GT)=\frac{GT^n}{\sum_n^N GT^n} \tag{5}$$

$$Q\left(P^i,GT^i\right)=-\frac{1}{N\times L}\sum_i^L\sum_n^N w^n\left[p^{i,n}\log\left(GT^{i,n}\right)\right]+$$
$$\left(1-p^{i,n}\right)\log\left(1-GT^{i,n}\right) \tag{6}$$

式中，w 为自适应权重参数，根据每张训练图中对应类别的像素比例计算而得；w^n 中 n 为不同的类别。

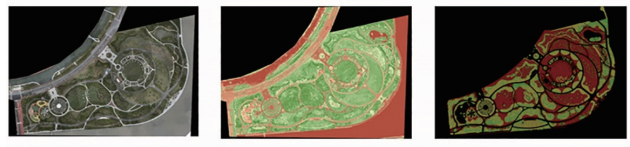
5 多光谱遥感高精度识别模型实证

5.1 案例选择与数据获取

上海泡泡公园位于上海市奉贤区，占地约20 hm^2，是植被覆盖丰富的城市公共绿地。针对19.25 hm^2的核心区域，于2021年5、6、7月开展了无人机可见光与多光谱航测，筛选出最优天气下的成像数据。采集范围为450、555、660、720、750、840 nm 6个波段，巡航高度为100 m，航测后图像数据拼合精度为2.54 cm。本研究航测选取精度要求：①遥测区域无云层阴影影响且光照条件充足；②乔木树冠边界清晰，常绿与落叶乔木区分明显；③灌木与地被空间分界明显且可区分类型，如木本、草本、成片灌木与独立灌木等。

5.2 数据标注过程

实验图像包括多光谱GNDVI图像和三通道可见光图像2种，尺寸均为10 486×7 328 px，重点关注乔木（tree）、灌木（bush）、地被（grass）3类要素。使用标准化标注软件Lableme对于原始图像进行标注，未标注目标被统一归为背景要素（background），如图4所示。



a三通道可见光原图　b多光谱GNDVI图像（假彩色图）　c标注图像（红色为地被，黄色为灌木，绿色为乔木

图4 采集图像

评估与验证过程在数据集的构建方面与传统实验不同。首先，从原图中随机切割了2个互不重叠的1 500×1 500 px图像切片作为测试集的一部分；之后，在排除以上图片切片的基础上，进行有重叠的512×512 px的图像切割。考虑到最终识别准确性评估必须模拟真实图像，重叠部分的切片融合过程一并纳入训练样本。切割后图像切片按比例进行划分，保证全部训练集、验证集、测试集的比例为8∶1∶1。为提高训练数据多样性，使用缩放、旋转、颜色偏移、高斯噪声加噪等方式对生成的切片以及对应标注图像进行数据增广。

5.3 模型实验结果

使用Precision（P），Recall（R）和F1-score综合评定模型性能，Precison代表模型能否准确区分类别的能力，Recall代表模型能否找到关注目标的能力：

$$P=\frac{TP}{TP+FP} \tag{7}$$

$$R=\frac{TP}{TP+FN} \tag{8}$$

$$F1=\frac{2(P\times R)}{(P+R)} \tag{9}$$

式中，TP为预测到目标且预测类别正确的像素数量；FN为预测到是背景但实际是目标的像素数量；FP为预测到是目标但实际是背景的像素数量，计算后得到性能评估表（表1）。

表1 模型性能评估表/%

	单入网络						双入网络					
	P	R	$F1$	P	R	$F1$	P	R	$F1$	P	R	$F1$
区分背景	0.974	0.986	0.979	0.973	0.960	0.967	0.973	0.983	0.979	0.982	0.979	0.982
乔木	0.845	0.761	0.839	0.733	0.698	0.782	0.888	0.758	0.846	0.891	0.819	0.886
灌木	0.804	0.920	0.890	0.787	0.862	0.873	0.820	0.925	0.899	0.857	0.931	0.920
地被	0.924	0.841	0.900	0.821	0.859	0.888	0.913	0.844	0.907	0.914	0.872	0.926
平均值	0.876	0.851	0.887	0.797	0.834	0.866	0.882	0.854	0.895	0.894	0.881	0.918

相对于仅使用可见光影像，由于缺失材质等信息，P、R、$F1$在仅使用GNDVI时表现均不佳，但在区分背景和目标上有较好表现，而双入网络（double in network）性能明显优于单入网路（single in network）。如乔木，早期融合相对仅可见光影像输入的方案精度提高了5.09%，相对仅多光谱GDVI融合提高了21.15%；而

特征融合更是在此基础上更加提升，相对单可见光获得了5.44%的提升。可视化和数字化结果均印证了2种图像结合输入网络的效果更佳，而在特征层面融合效果较图像层面直接融合效果更好，其应用达到了91.8%的公园植被高精度识别。

6 讨　论

在实验中，将融合与仅使用可见光图像在相同的编解码网络中进行消融实验比对，实验结果表明，单纯使用可见光图像尽管可以实现多类别植被目标的分离工作，但是其效果相对于融合后结果存在明显差异。尤其是乔木类的结果，其 *R* 以11.83%的差异远低于平均水平，这就导致了大量的乔木样本缺失。对于灌木，其 *P* 则以8.95%低于平均水平，该类样本易被错误分类为其他类别。

以上数据可证明仅使用可见光图像存在检测的局限性。相对来说，在融合方案中（以feature fusion为例）乔木 *R* 的平均偏差比例改善为7.57%，灌木 *P* 的平均偏差改善为4.32%，在单一指标上有了明显提升。在综合评定指标 *F*1 上，乔木、灌木、地被3类别分别实现了5.60%、3.37%、2.89%的数据提升，整体上提升了3.49%。重点关注到，乔木检测在信息融合后有明显提升，可以合理推测近红外光谱和绿色光

谱互信息的加入改善了网络识别植被的能力并提高了网络对类别间差异信息的响应，迫使网络更好地挖掘了红蓝光谱的信息。

为了验证GNDVI信息本身就提供了较可见光图像更多的信息量，也进行了类似以上的实验。实验结果是否定的，从总体试验结果看仅GNDVI图像作为输入的结果反而比仅可见光图像作为输入的结果差2.37%，对于仅GNDVI图像结果均差于可见光图像的原因不难分析，主要是GNDVI图像仅包含绿色和近红外两部分的光谱信息，其初始信息量远低于包含三光谱信息的可见光图像，尤其对于树木，绿色光谱的相似性使得树木难以区分于周围植被。但是，对于灌木和地被类的内部性能提升，推测GNDVI信息可以将网络关注力置于大面积覆盖的灌木和地被区分上。

7 结　语

在双碳政策的引导下，城市公园绿地的各类植被的碳汇研究愈发重要，其评估的关键点在于公园植被的高精度识别，以提升碳汇能力、促进节能减排、改善人居环境。通过DJI M300RTK航测无人机与MS600Pro机载多光谱相机的设备组合可实现可见光与多光谱影像的航测获取，结合双重注意卷积神经网络学习可实现高精度识别的模型构建，优化了传统遥感航片在像素模糊、边界不清、识别效率低下的问题。在上海泡泡公园的模型训练与实验评估表明，通过早期融合和特征融合的组合模型方案，乔木、灌木、草本识别度分别可达88.6%、92.0%和92.6%，综合识别度可达91.8%，为后续碳汇测算、植被评估提供了有力的科学技术支撑。

参考文献

[1] 刘滨谊,王南. 应对气候变化的中国西部干旱地区新型人居环境建设研究[J]. 中国园林,2010,26(8):8-12

[2] 王南,魏维轩,刘滨谊. 公园游步道设计提升健步功效的研究:以南京鱼嘴湿地公园游步道设计为例[J]. 中国园林,2021,37(2):48-53

[3] 牟怀义. 基于DEM的森林植被空间分析与变化检测[J]. 地理空间信息,2018,16(7):96-100

[4] 俞天宇,倪文俭,刘见礼,等. 大兴安岭林区无人机可见光影像散发枯立木识别算法[J]. 遥感学报,2021,25(3):725-736

[5] 杜政,方耀. 结合随机森林的高分一号分类最优组合研究[J]. 地理空间信息,2017,15(2):15-18

[6] 宋勇,陈兵,王琼,等. 无人机遥感监测作物病虫害研究进展[J]. 棉花学报,2021,33(3):291-306

[7] 周恩生,魏维轩,王南. 城市公园碳汇计量中RTK多光谱遥感技术的应用研究[J]. 现代园艺,2022,45(9):47-49

[8] 刘凤莲,曹永兴,高润明,等. 基于高分辨率卫星影像的输电走廊植被生长预警[J]. 电力科学与技术学报,2021,36(3):188-194

[9] 朱金山,宋珍珍,赵露露. 结合底质分类与SVR算法的多光谱影像测深[J]. 地理空间信息,2019,17(11):44-46

[10] Hinton G E, Osindero S, Teh Y W. A Fast Learning Algorithm for Deep Belief Nets[J]. Neural Computation, 2006, 18(7):1 527-1 554

[11] 魏麟. 基于高分辨率遥感影像的震灾建筑物损毁检测[J]. 地理空间信息,2022,20(3):68-71

[12] 江泽涛,秦嘉奇,胡硕. 基于多路卷积神经网络的多光谱场景识别方法[J]. 计算机科学,2019,46(9):265-270

[13] 徐从安,吕亚飞,张筱晗,等. 基于双重注意力机制的遥感图像场景分类特征表示方法[J]. 电子与信息学报,2021,43(3):683-691