

# Hyperspectral Image Instance Segmentation Using Spectral–Spatial Feature Pyramid Network

## 摘要

高光谱图像具有更丰富的光谱信息。普通网络模型在特征提取过程中通常更关注图像的空间信息，因此对于 HSI 来说，对其光谱维度的关注缺失将导致光谱特征缺失。在图像的特征提取中，空间信息和光谱信息可以被视为互补特征。考虑到以上因素，我们需要设计一个既能够关注到空间信息，又能够捕捉光谱信息的尺度金字塔模块。为了增强光谱特征的提取并捕捉细微的差异，在原网络的基础上，我们引入了一种在三个尺度上操作的尺度金字塔特征提取模块：像素、超像素和子图。这设计旨在全面提取图像的多尺度特征，以更好地适应树木在不同尺度上的表现。本文创新内容是和叶松鑫师兄一起工作的，为了验证我们的方法，我们从深圳大学沧海校区（SZU-CH）收集了高光谱数据，并对该数据集进行了详细的训练和验证。

**关键词：**特征融合；高光谱图像（HSI）实例分割，特征金字塔网络

## 1 引言

随着遥感技术的发展，HSI 的质量和分辨率都有了显著的提高，许多重要的领域，如民用地图的校正、城市的三维重建等，都需要更加精细的遥感影像分类的支持 [5]。实例分割不仅是对每个像素进行分类，还可以将同一类别中的所有个体区分开来，可以更加精确地利用 HSI 信息。实现端到端的 HSI 实例分割可以消除传统分类方法的冗余操作，为 HSI 处理提供新的契机。

语义分割可以为图像中的每个像素分配一个对应的类别，但它并不区分同一类别中的对象 [4]。作为计算机视觉中的一个重要方向，实例分割结合了目标检测和语义分割两个任务，能够根据特定类别区分不同的实例，并在实例级目标定位的结果上实现不同类别的像素级分割 [6]。随着深度卷积神经网络的应用，许多实例分割框架被提出，如 Instance FCN [7]、Mask R- CNN [8]、MS R- CNN [12]、Point Rend [1] 等。

论文中引入了尺度金字塔模块，分别在像素、超像素和子图三个尺度上操作。这种超像素分割策略不仅对分割和特征提取任务有帮助，而且能够减少潜在的计算量。处理每个块所需的计算量更为有效，因为每个块代表一个感知单元，相较于逐像素处理，有助于提高分割方法的效率，特别是在大规模高光谱图像的处理中。子图尺度特征提取旨在获取相邻超像素之间的信息。通过对所有超像素图进行特征提取，得到整个图像的节点特征矩阵，有效地利

用每个节点的邻域信息，增强对节点环境及其周围上下文的理解。这对于节点分类和边缘预测任务尤为重要。为了缓解特征分布在迭代中的扭曲，引入了批量归一化（BN）来归一化输出，使其更接近于零均值和单位方差的分布。

## 2 相关工作

### 2.1 实例分割

实例分割在计算机视觉中扮演着关键角色，结合了对对象检测和语义分割，能够在像素级别上对图像进行更精细的分割。相较于语义分割 [4]，实例分割能够为每个像素分配相应的类别并区分同一类别中的不同对象 [6]。根据涉及的阶段数量，实例分割方法可分为两阶段 [8] 和单阶段两大类。双阶段方法首先进行对象检测 [2]，然后在检测到的对象上进行进一步的实例分割，而单阶段方法直接关联对象检测和实例分割，避免了繁琐的后处理步骤，提高了处理速度。

两阶段实例分割方法的代表是 mask R-CNN [8]，通过扩展边界盒回归分支实现更精细的 mask 预测。尽管这些方法在任务上取得了显著成果，但存在后处理繁琐、计算成本高、速度慢的问题 [3]。相对而言，单阶段方法在速度和精度方面实现了更好的平衡，通过直接关联对象检测和实例分割 [11]，提高了实时性和效率。Faster R-CNN 等单阶段方法借鉴了先进的对象检测算法中的候选区域思想，为实例分割任务提供了更灵活和高效的解决方案。

未来的研究可能集中在进一步提高单阶段方法的准确性，以适应更广泛的应用场景。单阶段实例分割模型简单而灵活，极大地影响了实例分割领域的研究。其主要过程包括直接输出类别和掩码，不包含目标检测头，从而在简化流程的同时提高了实时分割的速度。单阶段实例分割的主要过程如下所示：

$$Z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (1)$$

### 2.2 超像素分割

在计算机视觉中，图像分割是将图像划分为一系列超像素的过程，其中超像素是由相似特征的相邻像素组成的小区域。这种分割过程有助于降低图像处理的复杂度，并且超像素通常保留了对进一步图像分析有价值的信息。在高光谱图像处理中，超像素分割被视为关键步骤，它将图像划分为感知意义明显的块 [10]，提供了更有效的数据表示。

本文采用了简单线性迭代聚类（SLIC）[9] 算法进行超像素分割预处理。SLIC 利用 k-means 算法生成超像素，通过在与超像素大小成正比的区域内限制搜索空间，有效减少了距离计算的次数。此外，SLIC 采用了加权距离度量，结合颜色和空间邻近性，对超像素的大小和紧凑性进行更好的控制，具有高计算效率和理想的分割结果。

整个高光谱图像处理流程中，超像素分割首先对图像进行了划分，产生了感知意义明显的超像素块。这有助于捕获局部结构信息，并为后续处理提供了更有效的数据表示。超像素分割的结果经过预处理构图，为构建图结构提供了基础。随后，构图后的超像素输入尺度金字塔模块，以提取多尺度的图结构特征，从而更全面地捕捉图像的信息。

### 3 本文方法

1) 本论文中使用了 CBAM 模型, CBAM 模块 (包括通道注意力模块和空间注意力模块两部分, 用于进行频谱和空间特征的注意力融合), 旨在增强网络对图像中有信息的区域的感知能力, 如图1所示。

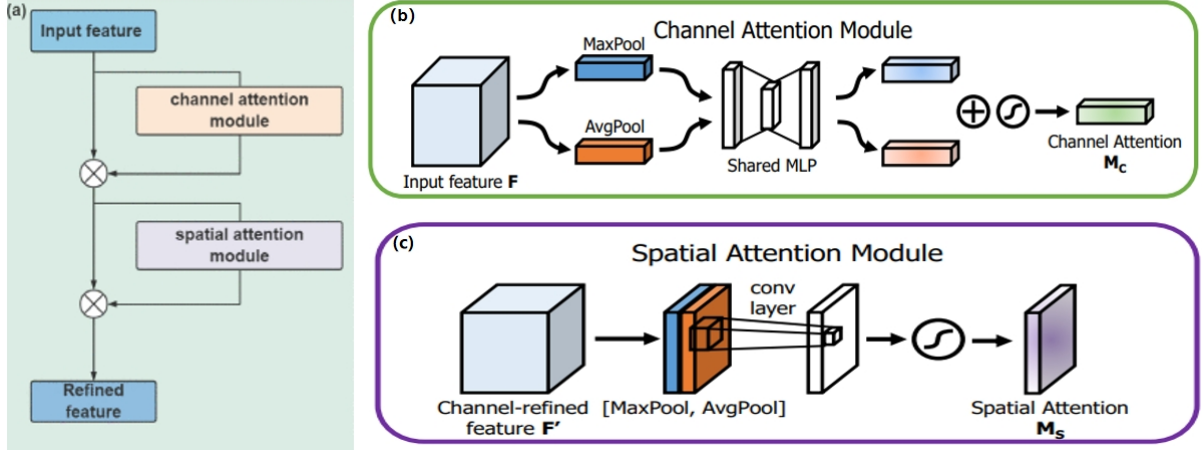


图 1. CBAM 模块总体结构: ( a ) CBAM 模块整体流程, ( b ) 通道注意力模块概述, ( c ) 空间注意力模块概述。

通道注意力机制可以表示为如下形式:

$$M_c(F) = \sigma(MLP(AvgPool(F))) + MLP(MaxPool(F)) \quad (2)$$

式中:  $\sigma$  表示 sigmoid 函数。在通道注意力模块的计算过程中, 我们同时使用平均池化操作和最大池化操作来聚合空间信息。共享网络由权重共享多层感知器 ( Multilayer Perception, MLP ) 组成, 共享 MLP 的核心部分用  $1 \times 1$  卷积完成信息提取。同样地, 在空间注意力模块的计算过程中, 沿着通道轴应用平均和最大池化操作来获得两个特征图, 然后基于通道进行级联。然后进行  $7 \times 7$  的卷积操作, 将通道数减少为 1, 最后使用 sigmoid 函数生成空间注意力特征图。上述过程可表示为如下形式:

$$M_s(F') = \sigma(f^{(7 \times 7)}(AvgPool(F'); MaxPool(F')))) \quad (3)$$

2) PAN 模块 (包含自上而下路径 (Top-Down Pathway) 和自下而上路径 (Bottom-Up Pathway), 利用多尺度金字塔结构来提取不同层次的语义和空间信息), 能够有效地融合不同层次的特征信息, 从而提升网络对实例分割任务的感知能力和表达能力, 如图 2 所示:

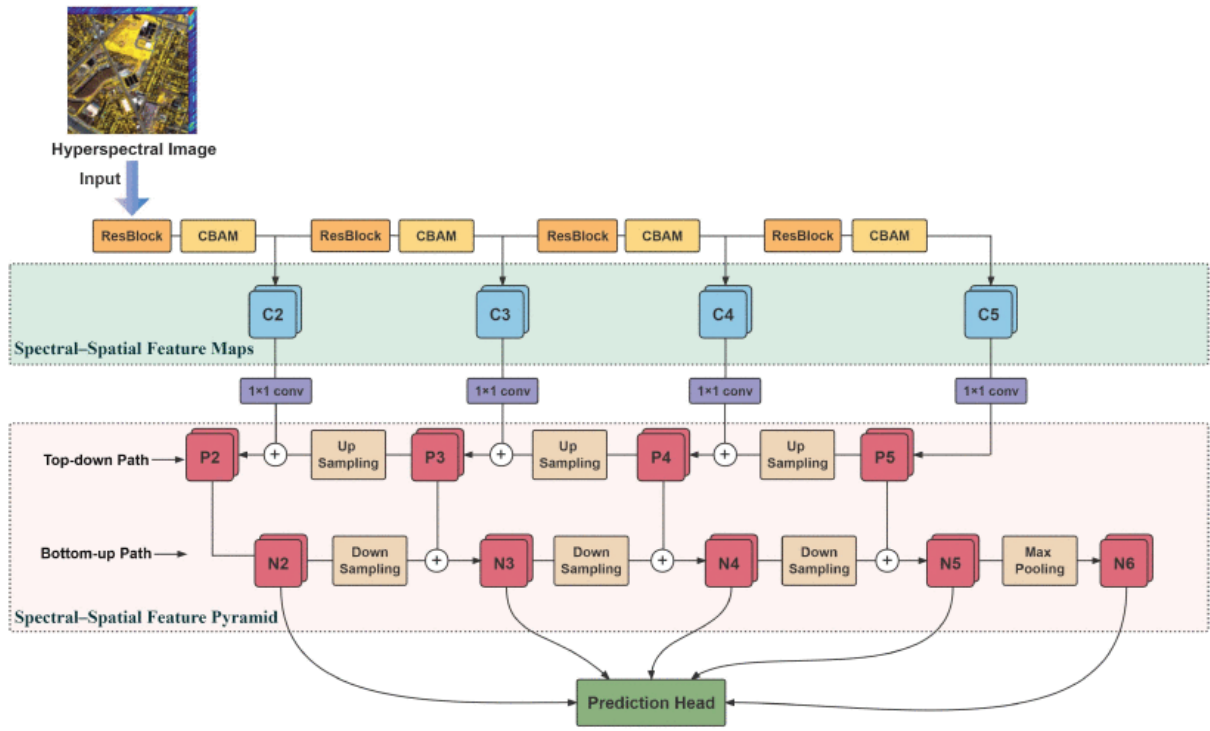


图 2. PAN 模块详见 Spectral-Spatial FPN 主要结构

3) 实例分割（两阶段网络模型，单阶段网络模型），模块结构如图3所示。

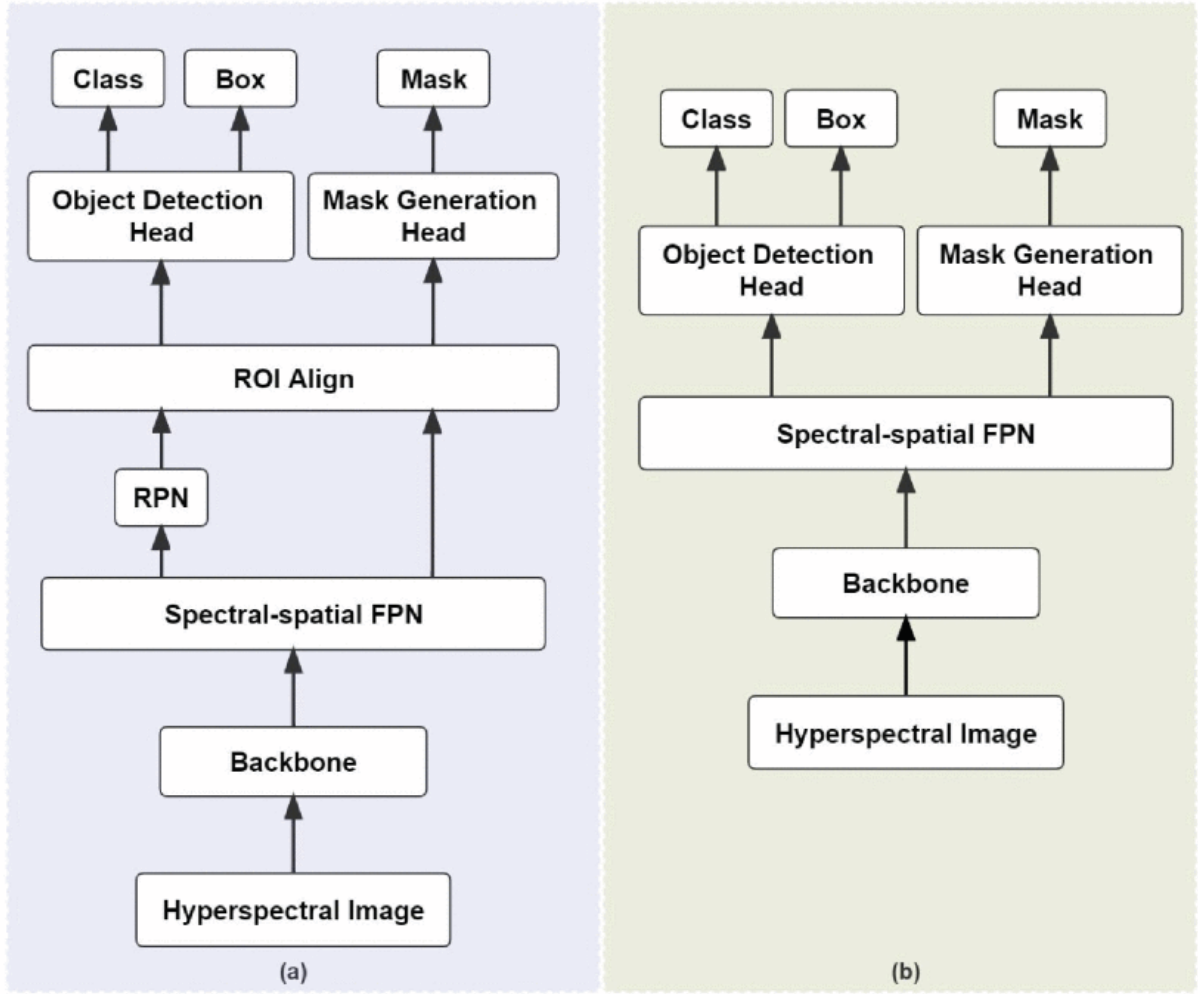


图 3. 实例分割 (a) 两阶段网络模型和 (b) 单阶段网络模型

### 3.1 本文方法概述

此部分对本文将要复现的工作进行概述，我们通过引入三个尺度操作的尺度金字塔特征提取模块，包括像素、超像素和子图，我们旨在增强光谱特征的提取，可以更全面地捕捉图像的上下文信息，同时确保光谱信息得到足够的关注。我们从深圳大学沧海校区（SZU-CH）收集了高光谱数据，并对该数据集进行了详尽的训练和验证，以验证我们提出的尺度金字塔网络的有效性。

### 3.2 尺度金字塔特征提取模块

本文的主要贡献在于引入了图网络的尺度金字塔模块。本文构建的图结构包含了不同层次的信息。为了更好地挖掘每一层的潜在信息，我们提出了一个用于图网络的尺度金字塔模块。如图4所示，尺度金字塔模块从像素、超像素和子图三个尺度层次提取特征。该研究的主要创新点在于引入了尺度金字塔模块，为图结构构建提供了多层次的信息。图网络的设计包含了不同层次的数据，为了充分挖掘每一层的潜在信息，研究者提出了专门为图网络设计的尺度金字塔模块。该模块从像素、超像素和子图三个尺度层次提取特征，以更全面、多维度地捕捉图像信息。



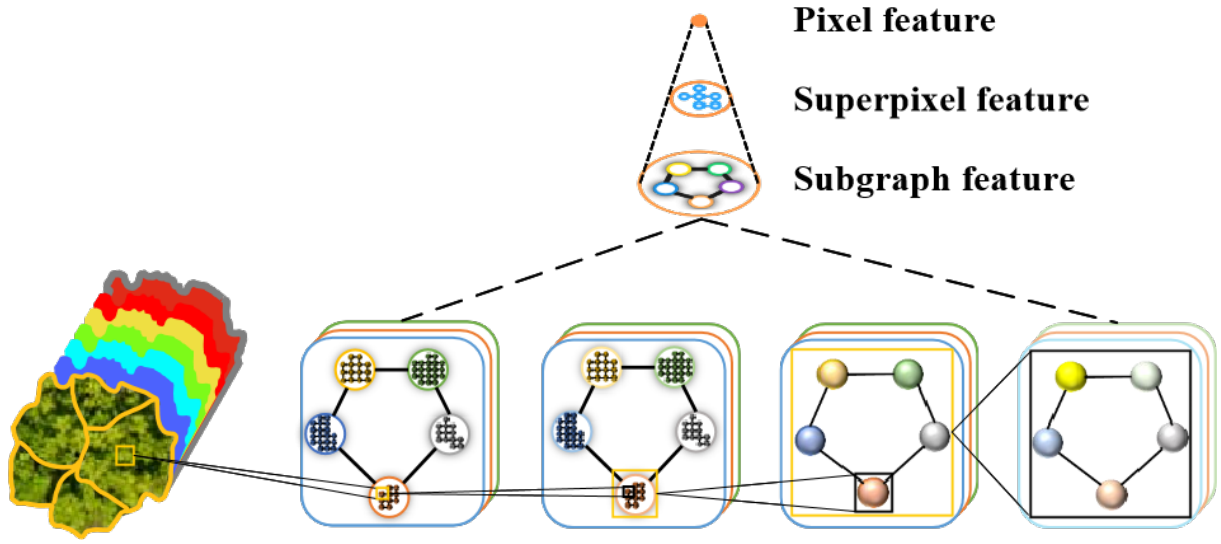


图 4. 尺度金字塔模块示意图

### 1) 像素尺度特征提取

像素尺度特征提取模块是用于图卷积的高效通道注意力模块。随着光谱分辨率的提高, 光谱信息存在一定的冗余。据介绍, 存在于各种树种之间的光谱信息变化显著。此外, 虽然属于同一类别的两棵树可能表现出可比较的光谱信息, 但在特定的波段中可以观察到差异。因此, 设计通道注意力来捕获不同的波段是至关重要的。

$$\omega_s = \sigma(C1D_k(g(A_s))) \quad (4)$$

其中  $\sigma$  是通道全局平均池 (GAP),  $g$  是 Sigmoid 函数,  $C1D_k$  是核大小为  $k$  的 1D 卷积), 并一个设计用于将  $s$  扩展到与  $a_s$  相同大小的函数, 并且是逐元素乘积。

### 2) 超像素尺度特征提取

超像素尺度特征提取旨在学习超像素图的结构表示。我们在超像素尺度特征提取模块中使用线性图卷积, 从每个超像素图中提取一般信息, 并缓解由无监督聚类方法引起的超像素分割偏差的影响。当在超像素尺度上提取特征时, 我们使用全局平均池化作为非线性操作的替代, 以聚合来自超像素内所有节点的信息。与基于像素的方法相比, 基于超像素的个体树木分割方法更为合适。通过进行超像素分割, 图像被划分为块状区域, 每个块都包含单个感知意义的信息。这种分块的方式使得分割后的特征更能够代表树冠层面的光谱独特性, 从而更有效地捕捉到树木的形态和光谱特征。

### 3) 子图尺度特征提取

子图尺度特征提取旨在获取相邻超像素之间的信息。对所有超像素图进行超像素特征提取, 得到代表整幅图像的节点特征矩阵  $X_g = [h_1, \dots, h_S]^T$ 。在这种全局背景下, 有效地利用每个节点的邻域信息, 可以增强对节点所处环境及其周围上下文的理解。在提取的特征中包含邻域信息在节点分类和边缘预测任务中起着至关重要的作用。传播过程中线性和非线性操作的迭代应用可能会扭曲特征分布, 从而为网络收敛带来挑战。为了缓解这一挑战, 我们引入批量归一化 (BN) 对输出进行归一化处理, 使其更接近均值和单位方差为零的分布。

## 4 复现细节

### 4.1 与已有开源代码对比

该论文已有开源代码，复现过程如下：本文设计了尺度金字塔模块，专注于提取像素、超像素和子图三个尺度上的图结构特征，利用多尺度特征提取方法有效地捕获高光谱图像的光谱属性。

### 4.2 实验环境搭建

在本节中，我们将介绍我们收集的 HSI 数据集 SZUCH 以及实验环境。1) 数据集。我们的 HSI 数据集是在中国深圳的深圳大学沧海校区收集的，我们对研究区域的树种进行了实地调查。根据调查结果，对单株木及其对应的树种进行了识别。如图5所示，共标注了 12 个类别，其他类别包括非树木对象。

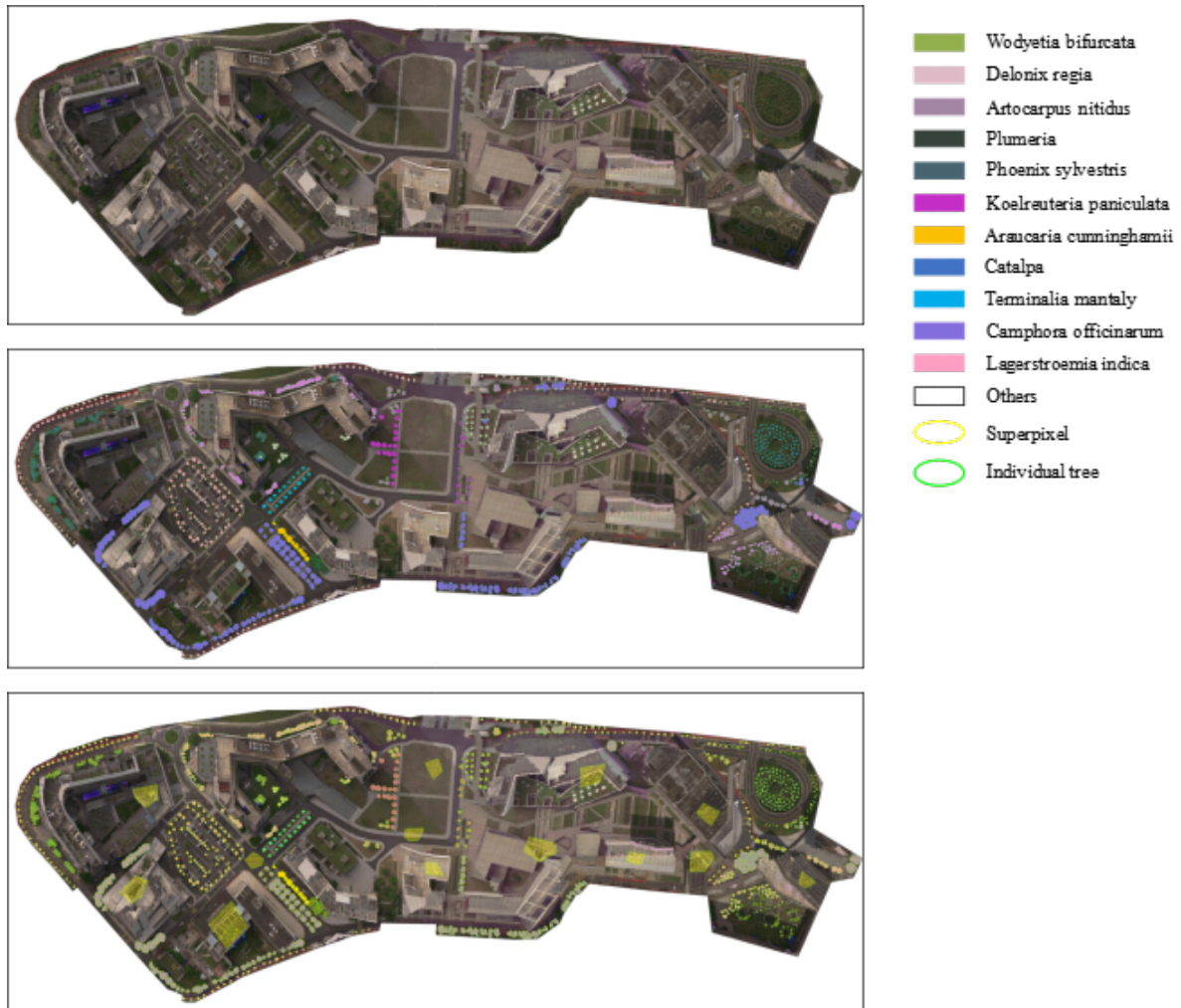


图 5. (a) 描绘了研究区域的高光谱图像的伪彩色表示，其中 (b) 表示手动注释的标签。(b) 中的绿色圆圈表示单独的树木，每种颜色代表图例中所示的特定类别。(c) 显示为超级像素生成的标签，每个黄色圆圈代表一个不同的超级像素块。

2) 实验环境。在 linux 操作系统上，使用 anaconda 作为环境管理器，并安装好 pytorch 相关环境，并安装 numpy 和 matplotlib。然后，使用与 RGB 相对应的三个波段进行超像素

分割，使用该数据集来评估该创新。我们使用现有的检测和分割方法比较数据集某些区域中 RGB 图像的预测。此外，总体准确度 (OA)、平均准确度 (AA) 和 kappa 被用作度量模型的节点分类性能的指标。关于该模型中的参数，为了节省资源消耗，将 7 中的 K 设置为 3，并且将所有层中的隐藏单元设置为 128。此外，考虑到单个树像素块的大小，SLIC 中的 k 的值被设置为 100。在培训过程中，使用 Adam 优化器，1000 次迭代的学习率为 1000。

### 4.3 创新点

该部分是和叶松鑫师兄一起工作下完成的。通过引入基于超像素构造的尺度金字塔图网络，本文旨在利用图结构中的节点分类来实现对高光谱数据中单株树木的精确分割。为了充分捕捉图像特征，我们设计了尺度金字塔模块，专注于提取像素、超像素和子图三个尺度上的图结构特征。为了验证提出模型的性能，我们在采集的一个基于无人机的高光谱数据集，即在 SZU-CH 数据集上对创新点进行验证，用于全面评估模型在实际场景中的表现。实验结果清晰地表明，提出的模型在单独的树分割任务中表现出色。通过对树木的准确分割，我们证实了该模型在高光谱图像分析中的有效性。

## 5 实验结果分析

本部分对实验所得部分结果进行分析，详细对实验内容进行说明，实验结果进行描述并分析。

首先通过该模型完成了对树种的分类任务以及单株分割任务，我们可以在下面给出表1，即节点分类精确度，kappa，边缘预测精确度以及 AUC 的结果展示。

本实验实现了该模型对单树识别的分割，图2展示了实验结果的部分可视化，可以看到，该模型的分割结果对比 groundtruth 十分相近。并且在多物体以及物体边缘处都取得比较好的分割结果。在我们的比较实验中，我们选择了 DeepForest 和 Detectree2 两种模型进行对比，以及考虑到分段任意模型 (SAM) 的最新工作，也将其纳入比较范围。由于缺乏注释的 RGB 图像数据，我们并未对比较模型进行调整，而是仅对来自本地收集区域的 RGB 图像进行了预测。



表 1. 表 1 SPGN 在 SZU 数据集上的实验结果展示

Class	Abbrev.	SPGN-base	SPGN
C0	OT	95.17	99.09
C1	WB	45.93	80.83
C2	DR	72.11	87.00
C3	AN	78.47	92.92
C4	PL	82.27	87.75
C5	PS	85.35	94.11
C6	KP	84.34	93.81
C7	AC	90.12	97.47
C8	CA	84.66	97.67
C9	TM	92.92	96.45
C10	CO	84.16	88.54
C11	LI	82.83	94.97
OA(%)		80.42	82.84
AA(%)		82.37	87.66
kappa		0.84	0.94
Link 0	different	78.55	92.83
Link 1	same	80.52	82.49
AUC		0.85	0.92

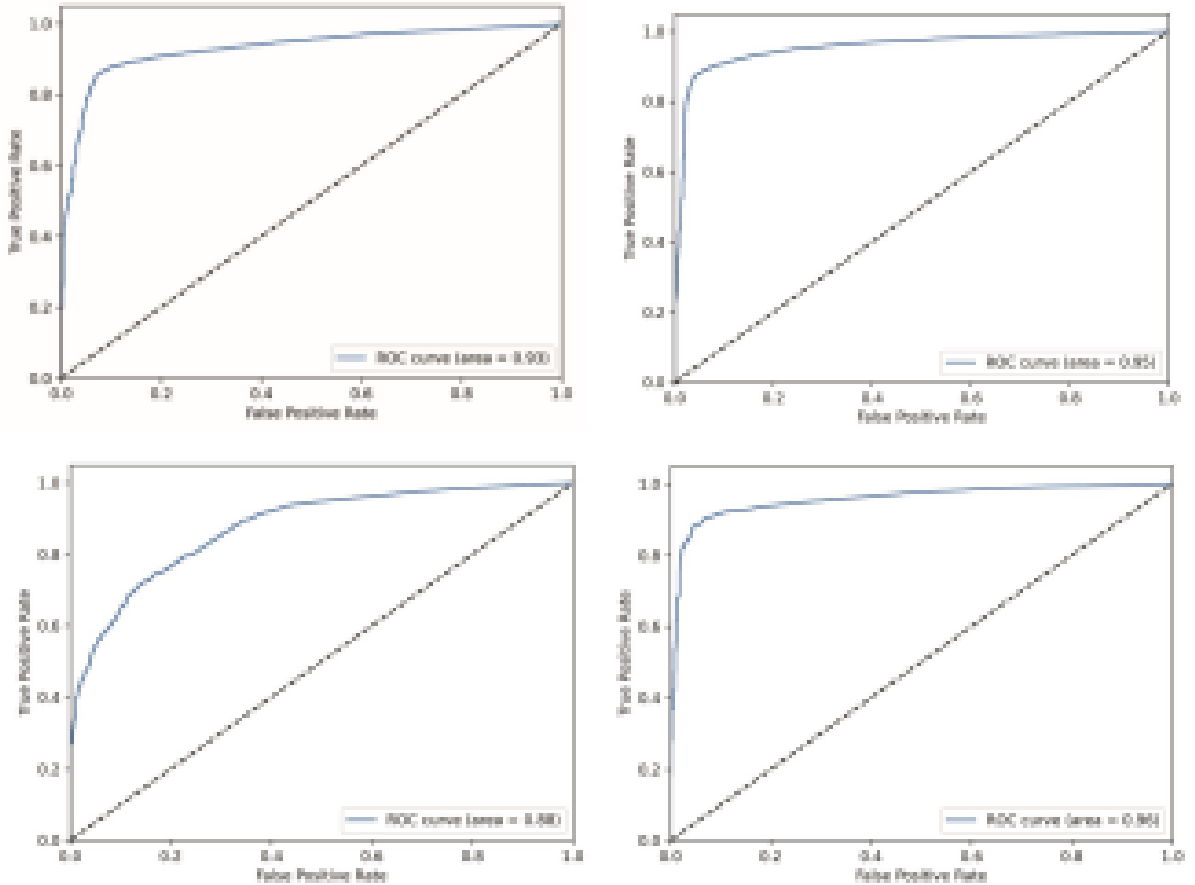


图 6. 真阳性和假阳性的 ROC (Receiver Operating Characteristic) 曲线

上图是真阳性和假阳性的 ROC 曲线，如图6。在 ROC 曲线中，横轴表示假阳性率 (False Positive Rate, FPR)，纵轴表示真阳性率 (True Positive Rate, TPR)。TPR 也称为灵敏度或召回率，计算方式是真正例的数量除以真正例的总数量。FPR 是假正例的数量除以真负例的总数量。通过观察 ROC 曲线，我们可以得到以下信息：该曲线都是靠近左上角，这说明了模型性能越好。左上角表示高的真阳性率和低的假阳性率，即模型在保持高召回率的同时，尽量减少错误的正例预测。同时 ROC 曲线下方的面积 (Area Under the Curve, AUC) 也是一个重要的性能指标。AUC 越接近 1，表示模型性能越好。

## 6 总结与展望

在 SZU-CH 数据集上对创新点进行验证是本研究的一项关键工作。另外通过一系列消融实验，我们深入证实了在从不同尺度提取高光谱图像特征方面的有效性，同时也验证了子图位置编码在增强边缘预测任务中的空间信息方面的重要性。这次实验的主要任务是对 FPN 网络框架进行复现，并有针对性地引入了创新的多尺度金字塔模块，以进一步优化网络性能。然而，与论文中的结果相比，我们在基础框架的复现上尚存在一定的差距。值得注意的是，尽管基础框架的复现存在一些差距，但在创新的多尺度金字塔模块的引入下，实验结果仍然验证了创新点的可行性和有效性。这反映了该模块对从不同尺度提取高光谱图像特征的良好性能，并且子图位置编码在增强边缘预测任务中的重要作用也得到了进一步的验证。

然而，文中并未详细探讨实现过程中的不足之处。未来的研究方向可能包括进一步改进模型性能、优化计算效率，以及在更广泛的应用场景中验证模型的通用性。这对于全面理解该模型在实际应用中的潜力至关重要。通过深入分析模型在复现过程中存在的差距和不足之处，我们可以为未来的研究提供有力的方向。在进一步优化模型性能方面，我们可以考虑引入更先进的网络结构或调整现有结构的参数，以实现更好的性能表现。此外，优化计算效率也是一个关键的研究方向，可以通过模型压缩、轻量级网络设计等方式来减少计算负担，提高模型的实际应用价值。

通过这次实验，我逐渐体验到了科研的曲折与充实。一开始，面对环境搭建的一系列失败，我深感科研的不易。在解决问题的过程中，我学到了更多实践技能，也锻炼了不畏困难的勇气。原文复现的过程中，我深刻领悟到了理论与实际的差距。有时候，看似简单的任务也可能隐藏着许多未知的难题。然而，正是这些困扰让我更加专注地去深挖问题本质，不断调整实验方案，寻找最优解。科研之路充满了挑战，但也蕴藏着无限的乐趣，科研不仅仅是为了解决某个问题，更是一次次的自我挑战和突破的过程。正因如此，每一次的成功都变得格外珍贵。在这个过程中，我不仅学到了专业知识，更锻炼了坚韧不拔、勇攀高峰的品质。通过这次实验，我不仅收获了研究的成果，更收获了成长的喜悦。这段科研之旅或许有挫折，但正是这些挫折让我变得更加坚韧，更加成熟。未来的科研之路充满了未知，我将继续秉持着初心，迎接更多的挑战，追求更高的学术境界。在叶松鑫师兄的一同努力下，成功地完成了这项工作，这门课程真的大大提升了我的思考能力，编程能力以及心理素质，也希望自己继续加油。

## 参考文献

- [1] K. He A. Kirillov, Y. Wu and R. Girshick. Pointrend: Image segmentation as rendering. In *Proc. IEEE/CVF Conf. on Computer Vision & Pattern Recognition*, page 250–269, 2017.

- [2] R. Girshick B. Hariharan, P. Arbeláez and J. Malik. Simultaneous detection and segmentation. In *Proc. Euro. Conf. on Computer Vision*, page 297–312, 2014.
- [3] F. Xiao D. Bolya, C. Zhou and Y. J. Lee. Yolact++ better realtime instance segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(2):1108–1121, Feb. 2022.
- [4] J. Hu et al. Istr: End-to-end instance segmentation with transformers. 2021.
- [5] J. Yue et al. Optical remote sensing image understanding with weak supervision: Concepts, methods, and perspectives. *IEEE Geosci. Remote Sens. Mag.*, 10(2):250–269, Jun. 2022.
- [6] A. M. Hafiz and G. M. Bhat. A survey on instance segmentation: State of the art. *Int. J. Multimedia Inf. Retr.*, 9(3):171–189, Jul. 2020.
- [7] Y. Li S. Ren J. Dai, K. He and J. Sun. Instance-sensitive fully convolutional networks. In *Proc. Euro. Conf. on Computer Vision*, page 534–549, 2016.
- [8] P. Dollár K. He, G. Gkioxari and R. Girshick. Mask r-cnn. In *Proc. Int. Conf. on Computer Vision*, page 250–269, 2017.
- [9] K. Smith A. Lucchi P. Fua R. Achanta, A. Shaji and S. S` usstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274–2282, 2012.
- [10] Ren and Malik. Learning a classification model for segmentation. page 10–17, 2003.
- [11] J. Xing L. Huang C. Huang S. Wang, Y. Gong and W. Hu. Rdsnet: A new deep architecture for reciprocal object detection and instance segmentation. in *Proc. AAAI Conf. Artif. Intell. (AAAI)*, 34(7):12208–12215, 2022.
- [12] Y. Gong C. Huang Z. Huang, L. Huang and X. Wang. Mask scoring r-cnn. In *Proc. IEEE/CVF Conf. on Computer Vision & Pattern Recognition*, page 6409–6418, 2019.