



工学硕士学位论文

**基于特征解耦和多模态数据融合的滨海湿地
高光谱遥感精细化制图研究**

作 者 温家祥

导 师 崔宾阁 教授

山东科技大学

二〇二五年五月

中图分类号 TP751

学校代码 10424

UDC 004

密 级 公开

山东科技大学

工学硕士学位论文

基于特征解耦和多模态数据融合的滨海湿地 高光谱遥感精细化制图研究

作 者 温家祥 入学时间 2022 年 9 月

导 师 崔宾阁 职 称 教 授

申请学位 工学硕士 所在学院 计算科学与工程学院

学科（类别） 计算机应用技术 方向（领域） 人工智能

答辩日期 2025 年 5 月 28 日 提交日期 2025 年 6 月 6 日

学位论文使用授权声明

本人完全了解山东科技大学有关保留、使用学位论文的规定，同意本人所撰写的学位论文的使用授权按照学校的管理规定处理。

作为申请学位的条件之一，学校有权保留学位论文并向国家有关部门或其指定机构送交论文的电子版和纸质版；有权将学位论文的全部或部分内容编入有关数据库发表，并可以以电子、网络及其他数字媒体形式公开出版；允许学校档案馆和图书馆保留学位论文的纸质版和电子版，可以使用影印、缩印或扫描等复制手段保存和汇编学位论文；为教学和科研目的，学校档案馆和图书馆可以将公开的学位论文作为资料在档案馆、图书馆等场所或在校园网上供校内师生阅读、浏览。

（保密的学位论文在解密后适用本授权）

作者签名：

日 期： 年 月 日

导师签名：

日 期： 年 月 日

学位论文原创性声明

本人呈交给山东科技大学的学位论文，除所列参考文献和世所公认的文献外，全部是本人攻读学位期间在导师指导下的研究成果。除文中已经标明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

若有不实之处，本人愿意承担相关法律责任。

学位论文作者签名：

年 月 日

学位论文审查认定书

研究生 在规定的学习年限内，按照培养方案及个人培养计划，完成了课程学习，成绩合格，修满规定学分；在我的指导下完成本学位论文，论文中的观点、数据、表述和结构为我所认同，论文撰写格式符合学校的相关规定，同意将本论文作为申请学位论文。

导师签名：

日期：

摘要

滨海湿地作为海陆交界的生态过渡带，在抵御海岸侵蚀、维系生物多样性和调节全球碳循环中发挥着不可替代的作用，精准解析滨海湿地植被群落的空间分布是评估湿地健康状况和制定生态修复策略的关键。随着遥感技术的突破，高光谱卫星图像光谱分辨率高、波段众多，能够获取地物几乎连续的光谱曲线，为滨海湿地光谱易混淆物种的精细识别提供了可能；激光雷达则通过主动探测技术精确量化植被高程、冠层密度等地形形态特征，弥补了光学影像在垂直维度信息缺失的短板。然而滨海湿地植被精细制图仍面临如下挑战：

（1）滨海湿地植被类内影像特征变化较大导致传统遥感分类方法精度受限；（2）湿地植被斑块破碎化，单一数据源难以实现形态-光谱协同解析。主要研究内容如下：

（1）滨海湿地植被类内影像特征变化较大导致传统分类模型易受成像条件影响，制约了滨海湿地高光谱图像精细制图的精度。针对这一问题，本文提出一种基于特征解耦与原型学习的双分支深度学习网络。首先，构建空谱特征分支与域特征分支的双通道架构：空谱特征分支通过原型归纳模块将高维特征映射至类别原型空间，显式建模植被类别的可区分性表征；域特征分支采用变分自编码器重构输入图像，通过隐变量约束提取与分类无关的干扰特征。其次，设计互信息最小化准则实现双分支特征解耦，强制空谱特征与域特征在信息空间正交。进一步，在原型学习过程中，通过最大化空谱特征与原型向量的互信息，同时约束其与域特征的统计独立性，从而增强滨海湿地地物分类特征的可解释性与鲁棒性。

（2）滨海湿地受到潮沟带和互花米草等入侵物种的影响，植被破碎化，单一遥感数据源难以全面表征滨海湿地植被的形态与光谱特性。本文提出一种高光谱-激光雷达数据自适应融合网络，实现几何结构与光谱特征的多层次协同。首先，设计可变形卷积多尺度编码器，通过动态调整卷积核偏移量，自适应捕捉湿地植被的斑块状分布与不规则边界特征。其次，提出几何引导光谱注意力模块，利用激光雷达高程信息生成空间权重图，动态校准高光谱特征通道的响应强度，增强冠层高度差异显著区域的光谱分辨力。进一步，创新渐进式融合机制，通过门控权重网络实现从局部细节到全局语义的特征层次化交互，避免早期融合中的信息淹没问题。最后，采用动态路由分类头，根据输入特征复杂度自动选择最优的专家网络组合进行决策。

关键词：滨海湿地；高光谱遥感；特征解耦；多模态融合；激光雷达

Abstract

Test v0.9

目 录

摘 要	6
图清单	I
表清单	II
变量注释表	III
1 绪 论	1
1.1 研究背景与意义 (Research background and significance)	1
1.2 国内外研究现状 (Research status at home and abroad)	2
1.2.1 基于像素的分类方法	2
1.2.2 面向对象的分类方法	2
1.2.3 基于深度学习的分类方法	3
1.3 问题的提出 (Statement of problem)	4
1.4 主要研究内容 (Main research content)	4
1.5 组织结构及章节安排 (Structure and chapter arrangement)	5
2 相关理论与技术	7
2.1 深度学习基础	7
2.1.1 卷积神经网络	7
2.1.2 注意力机制	9
2.2 表征学习	10
2.2.1 解耦表征学习 (Disentangled Representation Learning)	10
2.2.2 原型学习 (Semantic segmentation methods)	11
2.3 图卷积 (Graph Convolutional Network)	11
2.4 混合专家 (Mixture of Experts)	12
2.4 本章小结 (Summary)	13
3 基于特征解耦与原型学习的滨海湿地精细制图方法	14
3.1 引言 (Foreword)	14
3.2 MID-ProNet: 互信息解耦原型对齐网络 (MID-ProNet: Mutual Information Disentangled Prototype Alignment Network)	14
3.3.1 MID-ProNet 结构	14
3.2.2 双分支特征解耦器	16
3.2.2 动态原型对齐机制	20
3.3 实验与分析 (Experiments and analysis)	21
3.3.1 实验环境	22
3.3.2 实验数据集	22
3.3.3 评价指标	25
3.3.4 实验结果	26

3.3.5 消融实验.....	30
3.3.6 模型复杂性评估	30
3.4 本章小结 (Summary)	31
4. 基于多模态的滨海湿地植被精细制图方法.....	32
4.1 引言.....	32
4.2 主要流程	32
4.2.1 研究区域.....	32
4.2.2 实验数据.....	33
4.2.3 数据预处理	33
4.2.4 对比方法.....	34
4.3 几何-光谱联合动态形变网络 (Geometric-Spectral Joint Dynamic Deformation Network, GSJDD-Net)	34
4.3.1 可变形多模态编码器 (Deformable Multi-scale Alignment Module)	35
4.3.2 几何引导注意力模块 (Geometry-guided Spectral Attention)	35
4.3.3 自适应渐进融合模块 (Adaptive Progressive Fusion)	36
4.3.4 动态路由分类模块 (Dynamic Routing Classification)	37
4.4 实验结果与分析	38
4.4.1 实验结果.....	38
4.4.2 消融实验.....	39
4.4.3 模型复杂性评估	40
4.5 本章小结 (Summary)	40
5 总结与展望.....	41
5.1 总结 (Summary)	41
5.2 展望 (Prospect)	41
参考文献.....	43

参考文献

作者简介

致谢

学位论文集

Contents

摘 要	6
图清单	I
表清单	II
变量注释表	III
1 绪 论	1
1.1 研究背景与意义 (Research background and significance)	1
1.2 国内外研究现状 (Research status at home and abroad)	2
1.2.1 基于像素的分类方法	2
1.2.2 面向对象的分类方法	2
1.2.3 基于深度学习的分类方法	3
1.3 问题的提出 (Statement of problem)	4
1.4 主要研究内容 (Main research content)	4
1.5 组织结构及章节安排 (Structure and chapter arrangement)	5
2 相关理论与技术	7
2.1 深度学习基础	7
2.1.1 卷积神经网络	7
2.1.2 注意力机制	9
2.2 表征学习	10
2.2.1 解耦表征学习 (Disentangled Representation Learning)	10
2.2.2 原型学习 (Semantic segmentation methods)	11
2.3 图卷积 (Graph Convolutional Network)	11
2.4 混合专家 (Mixture of Experts)	12
2.4 本章小结 (Summary)	13
3 基于特征解耦与原型学习的滨海湿地精细制图方法	14
3.1 引言 (Foreword)	14
3.2 MID-ProNet: 互信息解耦原型对齐网络 (MID-ProNet: Mutual Information Disentangled Prototype Alignment Network)	14
3.3.1 MID-ProNet 结构	14
3.2.2 双分支特征解耦器	16
3.2.2 动态原型对齐机制	20
3.3 实验与分析 (Experiments and analysis)	21
3.3.1 实验环境	22
3.3.2 实验数据集	22
3.3.3 评价指标	25
3.3.4 实验结果	26

3.3.5 消融实验.....	30
3.3.6 模型复杂性评估	30
3.4 本章小结 (Summary)	31
4. 基于多模态的滨海湿地植被精细制图方法.....	32
4.1 引言.....	32
4.2 主要流程	32
4.2.1 研究区域.....	32
4.2.2 实验数据.....	33
4.2.3 数据预处理	33
4.2.4 对比方法.....	34
4.3 几何-光谱联合动态形变网络 (Geometric-Spectral Joint Dynamic Deformation Network, GSJDD-Net)	34
4.3.1 可变形多模态编码器 (Deformable Multi-scale Alignment Module)	35
4.3.2 几何引导注意力模块 (Geometry-guided Spectral Attention)	35
4.3.3 自适应渐进融合模块 (Adaptive Progressive Fusion)	36
4.3.4 动态路由分类模块 (Dynamic Routing Classification)	37
4.4 实验结果与分析	38
4.4.1 实验结果.....	38
4.4.2 消融实验.....	39
4.4.3 模型复杂性评估	40
4.5 本章小结 (Summary)	40
5 总结与展望.....	41
5.1 总结 (Summary)	41
5.2 展望 (Prospect)	41
参考文献.....	43

References

Author's Resume

Acknowledgements

Thesis Data Collection

图清单

[illegible]

变量注释表

变量	注释	初现页
----	----	-----

1 绪 论

1 Introduction

1.1 研究背景与意义（Research background and significance）

滨海湿地是陆地与海洋生态系统间的过渡地带，在调节气候、维持生物多样性及固碳等方面发挥着不可替代的作用[1]。根据联合国千年生态系统[2]评估，湿地是全球价值最高的生态系统，凭其独特的涵养水源、调蓄洪水和固碳释氧特性被誉为“地球之肾”。其植被与沉积物通过减缓海浪冲击，能够有效保护海岸线免受风暴侵蚀，同时复杂的土壤水盐梯度形成了独特的植被群落，这些群落对水土保持、盐碱地改良及生态景观维持至关重要[2][3]。然而，在过去很长的一段时间内，全球经济发展对湿地生态资源的保护缺乏足够重视，导致大量湿地被开发为城市或农业用地，湿地生态功能逐渐退化，全球范围内的湿地退化问题日益严峻。全球约 50%的湿地已经消失[5,6]。我国是世界上湿地资源最丰富的国家之一，湿地种类齐全且分布广泛。但是，我国正面临严重的湿地流失与功能退化问题。近 30 年来，我国湿地面积已减少约 33%，对生态系统稳定性和生物多样性构成重大挑战，滨海湿地已成为全球受威胁最严重的生态系统之一[7][8]。实现滨海湿地的精细制图，可以进一步分析滨海湿地的发展规律、退化原因以及所面临的问题，有助于专家提出解决滨海湿地保护问题的方案，为滨海湿地的生态修复提供科学支撑,对于维护我国海洋和陆地生态系统的平衡具有重要意义。

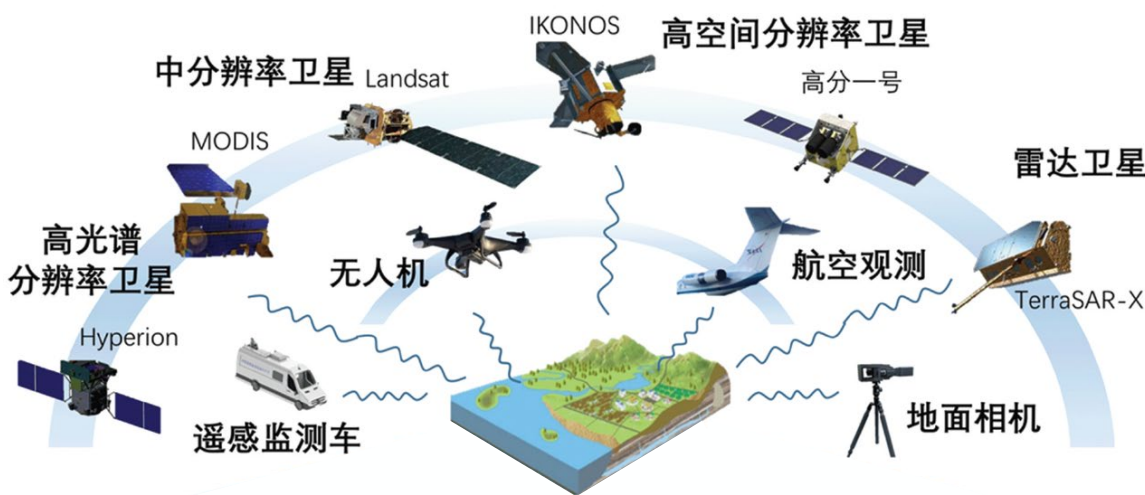


图 1.1 多层级遥感观测平台与传感器分类示意图[9]

传统的湿地野外调查方法在数据获取过程中面临诸多挑战，不仅需要投入大量人力和经济资源，还可能对湿地生态系统造成一定程度的破坏。此外，对于地形复杂、植被茂密或人类难以进入的区域，该方法难以及时、有效地获取和更新数据，导致采

集的信息具有局限性，无法全面反映湿地整体特征[9]。同时，现场数据的处理依赖于经验丰富的专家进行目视解译和手动标注，这一过程耗时且难以满足湿地实时监测的需求。如图 1.1 所示，随着遥感观测设备的不断发展以及遥感数据类型的日益丰富，多源异构遥感数据获取体系已逐步形成完整的观测能力，涵盖全色影像、多光谱影像、高光谱影像以及激光雷达数据等。此外随着深度学习技术在遥感领域的应用和不断发展，滨海湿地植被精细制图也有了新的扩展。相比传统踏勘方法，通过遥感技术和深度学习技术可以快速获取大范围、长时序的湿地信息，实现湿地边界划定、类型识别、植被分布、土壤水分含量等关键参数提取，在降低人力与经济成本，提高湿地制图与动态监测的时效性与精度方面有着独到的优势[10-14]。因此，结合深度学习技术开展滨海湿地植被精细制图有着重要的意义

1.2 国内外研究现状（Research status at home and abroad）

滨海湿地植被具有类型多样、生境复杂、分布广阔、生态脆弱等特点，不同植被类型之间在湿度和盐度条件驱动作用下常交杂混生，且极易发生变化，导致植被破碎化且边界模糊。此外，湿地植被具有较大的类内差异和类间相似性，进一步增加了精准区分不同植被类型的难度。实现滨海湿地植被的精细制图仍然是一项具有挑战性的任务。在目前的滨海湿地植被精细制图方法，除了传统的目视解译外，大致可以分为三种：

1.2.1 基于像素的分类方法

基于像素的分类方法通过提取像素光谱特征或人工设计特征组合[28, 38, 47-48]，结合传统机器学习算法实现湿地地物识别。典型研究如 Mahdianpari 等[53]集成 Sentinel-1/2 多源遥感数据，采用支持向量机（Support Vector Machine, SVM）算法构建纽芬兰首套 10 米分辨率湿地分类体系；Cao 等[54]融合无人机高光谱数据与数字表面模型，基于随机森林（Random Forest, RF）分类器实现红树林物种级精细分类。谢静等[47]利用多时相 HJ-1 影像光谱特征构建面向对象分类模型，开展三江平原湿地类型识别；Wang 等[48]通过 Google Earth Engine 平台整合时间序列 Landsat 影像，采用支持向量机算法完成中国滨海湿地动态制图。基于像素的分类方法该方法虽在湿地分类中应用广泛，但其特征表达主要依赖独立像素的光谱响应特性[29, 55]，缺乏对地物空间纹理特征及邻域上下文关系的建模能力[56]，制约了复杂湿地景观的分类精度。

1.2.2 面向对象的分类方法

面向对象的分类方法以同质性像元组成的“对象”为基本处理单元，通过融合对象的光谱响应特征、几何形态、纹理结构及空间尺度特征[28, 38, 47]实现地物识别。

该方法虽与基于像素的分类器架构相似，但通过对象化特征表达能获得更规整的分类边界与更优的语义一致性，因此在湿地遥感领域得到广泛应用。典型应用如 Mao 等[28]基于 Landsat 8 OLI 影像，采用面向对象分层分类方法构建中国首套国家级湿地专题产品；谢静等[47]结合多时相 HJ-1 影像光谱与纹理特征，建立面向对象的湿地分类模型并应用于三江平原湿地监测。然而，尽管该方法使用的分类器与面向像素的分类方法相近，但是面向对象的分类方法可以提取到更完整的第五特征，所以被广泛的应用到湿地遥感分类中。但是面向对象的分类方法高度依赖图像分割质量与特征工程的有效性[28, 57]，Fu 等[57]通过对比 GF-1 高分辨率影像与 SAR 数据的分类实验表明，面向对象方法在提升湿地植被制图精度的同时，显著增加了特征计算与模型调优的复杂度。

1.2.3 基于深度学习的分类方法

基于深度学习的滨海湿地分类方法通过构建多层神经网络架构，实现了对多源遥感影像的深层特征自适应提取与高精度分类。相较于传统机器学习方法，深度学习模型能够通过端到端训练机制融合光谱、纹理及时空上下文信息，显著提升了复杂湿地地物的识别能力。典型技术路径包括卷积神经网络（CNN）与全卷积网络（FCN）的应用，例如 Hu 等[34]结合 CHRIS 影像的光谱与纹理特征，构建深度卷积网络（DCNN）对黄河口湿地高光谱影像进行分类，有效解决了植被群落的光谱混淆问题；朱玉玲等[35]则针对互花米草识别难题，提出融合植被指数与浅层特征的 V-DCNN 模型，其分类精度较传统 SVM 方法提升显著。为进一步保留空间细节，Cui 等[36]基于改进 U-Net 架构设计 WetlandNet 模型，通过引入反卷积模块增强特征分辨率，在黄河口滨海湿地典型地物分割中实现了边界更规整的分类结果。

随着滨海湿地遥感数据源的多元化发展，单一传感器或时相数据已难以满足复杂生境分类需求。为突破这一瓶颈，研究者将深度学习方法与多源数据融合技术，通过构建自适应网络架构充分挖掘光谱-空间-时序特征的协同表达潜力。

在多源数据融合与模型创新方面，研究者提出多种适应性方案。Feng 等[37]集成多时相 Sentinel-1/2 数据，构建多分支 CNN 模型提取时序光谱与雷达特征，为黄河三角洲湿地动态监测提供新方法；Dang 等[38]将 ResNet 嵌入 U-Net 框架，联合 Sentinel-2 影像与地形数据（ALOS-DEM、NOAA-DEM），在越南滨海湿地分类中实现了跨区域模型迁移。Liu 等[30]通过联合高光谱与多光谱数据的判别性回归变换，设计多流 DCNN 模型，其提取的紧凑特征在滨海湿地分类中展现出更强的鲁棒性。尽管深度学习方法在特征表达能力上具有优势，其应用仍面临显著挑战：模型训练需依赖大量标注样本，而湿地复杂的地物边界导致标注成本高昂，像素级标注需求制约了其大范围应用[33]。

1.3 问题的提出 (Statement of problem)

滨海湿地作为海陆交界的重要生态过渡带，在抵御海岸侵蚀、维系生物多样性以及调节全球碳循环方面发挥着不可替代的作用。精准获取滨海湿地植被的空间分布信息对于评估湿地生态健康、制定科学的生态修复策略以及推动区域可持续发展具有重要意义。然而，由于滨海湿地环境的复杂性及其植被类型的多样性，利用遥感影像进行高精度分类仍面临诸多挑战，主要体现在以下两个方面：

(1) 滨海湿地植被存在同物异谱和同谱异物现象。由于数据采集设备的设置以及光照、环境温度、空气湿度和大气折射率等自然因素的影响，相同类别地物的光谱信息呈现多样性，不同类别地物的光谱信息表现出局部相似性，增加了高光谱遥感图像分类的复杂性。滨海湿地植被类内影像特征变化较大加剧了类别混淆问题，降低了传统分类方法在滨海湿地植被精细制图领域的适用性。

(2) 滨海湿地植被呈现破碎化和斑块状分布，其边界形态受潮流、沉积物输运及复杂环境因素的影响，存在类别重叠问题，边界往往较为模糊，给遥感影像分类带来了额外挑战。这种边界的不规则性和模糊性不仅增加了区域分割的难度，还可能导致提取结果出现边缘粗糙和形状失真等问题。传统像素级分类方法难以有效建模滨海湿地植被的复杂空间结构，制约了高精度分类的实现。

1.4 主要研究内容 (Main research content)

滨海湿地植被分类面临光谱特征高度相似、植被斑块破碎化等挑战，传统方法在分类精度和鲁棒性方面存在一定局限性。为此，本文分别基于单一高光谱数据和高光谱-激光雷达融合数据，提出两种新颖的深度学习新方法，以提升滨海湿地植被分类的精度和鲁棒性。主要研究内容如下：

(1) 基于特征解耦与原型学习的滨海湿地植被高光谱图像分类方法

滨海湿地不同植被群落的光谱特征高度相似，使得传统分类模型容易受到背景干扰，从而限制高光谱影像的分类精度。针对这一问题，本文提出一种融合特征解耦与原型学习的双分支深度网络，以增强高光谱植被分类的可分性和鲁棒性。首先，设计空谱特征分支与域特征分支的双通道架构。其中，空谱特征分支通过原型归纳模块，将高光谱特征映射到类别原型空间，显式建模植被类别的可分性表征；域特征分支采用变分自编码器 (VAE) 重构输入影像，提取与分类无关的干扰特征。其次，引入互信息最小化准则，实现空谱特征与域特征的解耦，从而降低背景干扰，提高分类的稳健性。此外，在原型学习过程中，采用动量更新机制优化类别原型，并约束其与域特征的统计独立性，以增强分类特征的可解释性和泛化能力。

(2) 基于多模态特征融合的滨海湿地植被协同解析方法

滨海湿地植被受潮沟带、入侵物种（如互花米草）等影响，景观呈现高度破碎化

和不规则边界，单一高光谱数据难以全面表征其形态和光谱特征。针对这一问题，本文引入激光雷达（LiDAR）数据，并提出一种基于多尺度动态对齐与几何引导注意力（MSDAM-GGSA）的融合框架。该方法充分利用 LiDAR 的空间结构信息与高光谱数据的光谱特征，实现精准的模态互补与信息对齐。首先，本文设计了多尺度动态对齐模块（MSDAM），通过可变形卷积（Deformable Convolution）在不同尺度上对高光谱数据进行变换，同时利用 LiDAR 数据进行特征对齐。MSDAM 采用不同膨胀率（dilation）的卷积提取局部与全局特征，并结合自适应池化（Adaptive Pooling）进行尺度归一化，以减少模态差异带来的不匹配问题。其次，本文提出几何引导光谱注意力（Geometry-Guided Spectral Attention, GGSA），借助 LiDAR 提取的几何信息（如边缘与梯度），指导高光谱数据的特征学习。该模块通过计算高光谱特征的查询（Query）与 LiDAR 辅助信息的键（Key），自适应调整光谱特征的注意力分布，从而增强空间-光谱一致性。最后，本文构建了自适应渐进式融合（Adaptive Progressive Fusion, APF）机制，以逐步融合不同尺度的特征。APF 采用多级门控（Gating Mechanism）调节高光谱和 LiDAR 之间的信息流动，并结合动态路由（Dynamic Routing）策略，根据不同区域特征的重要性选择最优分类专家（Expert）。

1.5 组织结构及章节安排（Structure and chapter arrangement）

本文的章节安排如下所述：

第一章：绪论。 阐述滨海湿地生态保护与遥感制图的研究背景与意义，系统梳理国内外滨海湿地植被分类方法的研究现状，包括基于像素、面向对象及深度学习的分类技术，总结当前高光谱遥感制图中存在的“同物异谱”“植被破碎化”等核心挑战，并提出本文的研究思路与创新点。最后概述全文的章节结构与逻辑框架。。

第二章：相关技术。 介绍支撑本文研究的理论基础与技术方法，包括深度学习基础（卷积神经网络、注意力机制）、表征学习（解耦表征学习、原型学习）、图卷积网络及混合专家系统，重点分析特征解耦与多模态融合的技术原理，为后续章节的方法设计提供理论支撑。

第三章：基于特征解耦与原型学习的滨海湿地精细制图方法。 针对滨海湿地高光谱影像中植被光谱混淆问题，提出互信息解耦原型对齐网络（MID-ProNet）。首先设计双分支架构分离地物固有特征与干扰特征，通过互信息最小化实现特征解耦；其次构建动态原型学习模块优化类别表征的鲁棒性；最后在黄河口滨海湿地数据集上开展实验，验证模型在复杂场景下的分类性能及泛化能力。

第四章：基于多模态数据融合的滨海湿地植被精细制图方法。 针对植被斑块破碎化与边界模糊问题，提出几何-光谱联合动态形变网络（GSJDD-Net）。结合高光谱与激光雷达数据，设计可变形多模态编码器、几何引导注意力模块及自适应渐进融合策略，实现光谱-几何特征的多层次协同；通过动态路由分类模块提升复杂区域的

决策精度。实验结果表明，该方法有效增强了边界敏感区域的分类一致性。。

第五章：总结与展望。 总结本文在滨海湿地高光谱遥感制图中的研究成果，分析现有方法的局限性，包括复杂共生植被的光谱混淆问题及跨模态数据对齐精度不足等。未来计划结合时序遥感数据与多层级样本库，探索动态场景下的湿地植被监测，进一步提升模型的解释性与生态应用价值。

2 相关理论与技术

2 Related theories and techniques

本章主要介绍本文研究方法所依赖的有关理论基础和相关技术，这些内容涵盖了卷积神经网络、特征解耦、原型学习、图卷积、边缘检测及注意力机制。

2.1 深度学习基础

深度学习是一种机器学习的分支，通过多层神经网络来模拟和学习复杂的数据表示。在高光谱图像处理中，深度学习通过构建深层神经网络，利用多层次的抽象特征来实现对图像内容的理解和表征。这些神经网络通过训练过程自动学习图像中的关键特征，使其能够高效地执行任务如分类、分割和检测。对于高光谱图像，深度学习可以学习光谱信息的复杂关系，从而提高对不同地物和材料的准确分类能力。深度学习在图像处理领域取得了显著的成功，为解决复杂的图像分析问题提供了强大的工具和方法。本节将介绍本文使用的深度学习相关知识。

2.1.1 卷积神经网络

卷积神经网络（Convolutional Neural Networks, CNN）最早于 1998 年由 LeCun 等人[12]提出，但是起初并没有得到重视。随着深度学习数据集的发展和硬件设备算力的提升，特别是 AlexNet[13]在 2012 年 ImageNet 竞赛中取得突破性成果后，CNN 逐渐成为计算机视觉领域的核心算法框架。卷积神经网络通过层次化的结构设计实现了对图像特征的自动提取与抽象。传统的 CNN 通常由输入层、卷积层、池化层、全连接层和输出层组成，各组件通过级联方式形成端到端的特征学习框架，如图 2.1 所示。

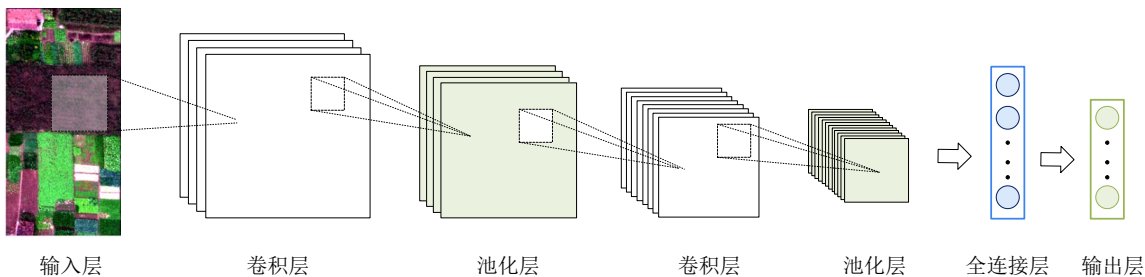


图 2.1 卷积神经网络示意图

Fig. 2.1 Schematic diagram of convolutional neural network

卷积层（Convolutional Layer）是 CNN 的核心组成部分，其主要作用是利用局部感受野（Receptive Field）提取特征，逐层构建从低级到高级的特征表示。卷积操作的本质是通过滑动窗口在输入图像上执行加权求和，并叠加偏置项，从而获得特征映射（Feature Map）。其计算公式如下：

$$\mathbf{F}^{(l)} = f\left(\sum_{k=1}^{C_{in}} \mathbf{W}^{(l,k)} * \mathbf{X}^{(k)} + b^{(l)}\right)$$

其中 $\mathbf{X}^{(k)}$ 表示第 k 个输入特征图， $\mathbf{W}^{(l,k)}$ 为第 l 个滤波器在

第 k 通道的权重矩阵， $*$ 表示二维卷积运算， $b^{(l)}$ 为偏置项， $f(\cdot)$ 为非线性激活函数（如 ReLU）。卷积操作通过权值共享机制减少模型参数，提升计算效率，同时保留局部空间结构信息。如图 2.2 所示，卷积核在输入特征图上滑动，并生成新的特征表示。如图 2.2 所示，卷积核在输入特征图上滑动，并生成新的特征表示。

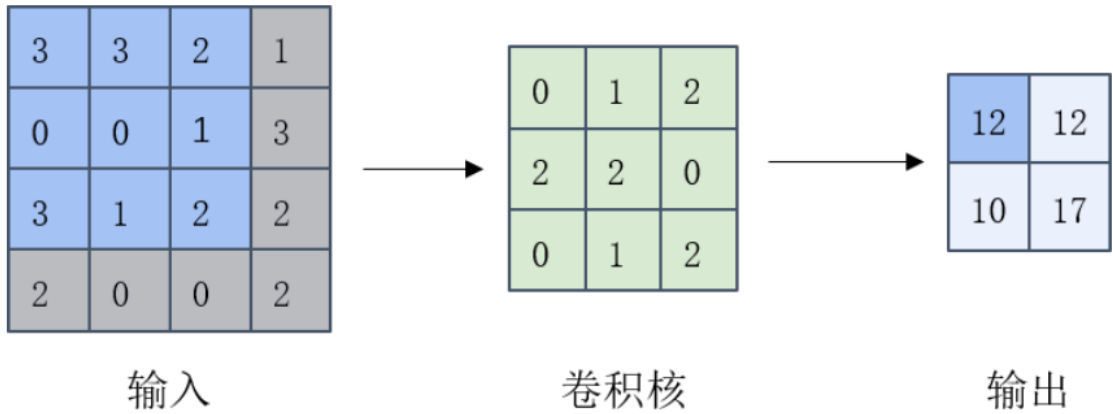


图 2.2 卷积示意图

Fig. 2.2 Schematic diagram of convolutional

池化层（Pooling Layer）是一种降维操作，其目的是减少特征图的空间维度，同时保留关键信息，提高模型的计算效率并增强模型对微小变换的鲁棒性。常见的池化方式包括最大池化（Max Pooling）和平均池化（Average Pooling），操作过程如图 2.3 所示

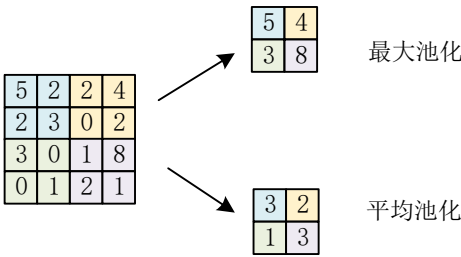


图 2.3 池化示意图

Fig. 2.3 Schematic diagram of Pooling

虽然池化操作降低了计算复杂度并增强了模型的平移不变性，但由于其不可逆性，可能导致部分位置信息丢失。因此，在高精度任务中，研究人员引入 自适应池化 和 注意力机制 以提升信息保留能力。

经过卷积和池化等线性变换后，输出层可视为输入层特征的线性组合。然而，单纯的线性映射无法有效逼近复杂的数据分布，因此 CNN 在输出层之前通常引入非线性激活函数，常见的激活函数如图 2.4 所示

全连接层（Fully Connected Layer, FC）通常位于图像分类网络的最后几层，其作用是将最终卷积层输出的特征图展平为一维向量，并与下一层的所有神经元建立全连接，以整合整个网络提取到的深层特征。此前，卷积层、池化层以及非线性激活函

数已对原始数据进行分布式特征学习，并将其映射至隐层特征空间，而全连接层进一步对这些特征进行变换，使其对应于样本的类别标签，从而完成最终的分类任务。

在 CNN 结构设计中，全连接层的维度通常是固定的，因此网络只能接收特定尺寸的输入图像。由于全连接层包含大量参数，不仅会显著增加计算和存储开销，还可能导致模型训练速度下降，并提高过拟合的风险。

2.1.2 注意力机制

注意力机制源于对人类视觉认知的研究。人类在观察场景时，会自发地聚焦于关键区域而忽略次要信息，受此启发，研究者将注意力机制引入人工智能领域，赋予模型动态分配计算资源的能力，使其能够自主识别输入数据的关键区域或特征[1]。在自然语言处理领域，注意力机制首次在机器翻译任务中展现出革命性价值。2014 年，Bahdanau 等人[50]首次提出基于注意力的序列到序列模型，通过动态对齐源语言与目标语言词汇的关联权重，显著提升了长距离依赖建模能力。Vaswani 等人[51]进一步提出基于自注意力的 Transformer 架构，彻底取代传统的循环神经网络（RNN），成为自然语言处理的主流框架。Transformer 通过并行化全局注意力计算，实现了对长序列依赖关系的高效建模，为大规模预训练模型（如 BERT[52]、GPT[53]）的发展奠定了基础。

随着注意力机制在自然语言处理中的成功，研究者开始探索其在计算机视觉领域的应用潜力。相较于传统卷积操作的局部感知特性，注意力机制能够建立像素或特征通道间的全局关联，从而更高效地捕获复杂视觉模式。在视觉系统中，注意机制可以看作是一个动态的选择过程，通过根据输入的重要性自适应地加权特征来实现。

在计算机视觉领域主要衍生出三种典型实现形式：通道注意力、空间注意力和混合注意力。Hu 等人[54]提出的 SE（Squeeze-and-Excitation）模块是通道注意力的代表，通过全局平均池化获取通道维度统计量，利用全连接层建立通道间非线性交互，自适应校准特征通道的重要性权重，结构如图 2.12 所示。。空间注意力则聚焦于特征图的空间维度相关性，典型如空间变换网络（STN）[55]，通过生成二维空间权重图来增强关键区域的特征响应。Woo 等人[56]提出的 CBAM（Convolutional Block Attention Module）创新性地将通道与空间注意力进行级联融合，通过顺序执行通道维度压缩和空间维度聚焦，实现了对特征图的双重校准。

自注意力（Self-Attention）机制作为 Transformer 架构的核心组件，在视觉任务中展现出超越传统卷积的全局建模能力。其核心思想是通过计算特征图内部所有位置对的相似度，建立全局上下文依赖关系，结构如图 2.15 所示。给定输入特征矩阵 $X \in \mathbb{R}^{n \times d}$ ，自注意力机制通过线性变换生成查询、键和值矩阵：

$$Q = XW_Q, K = XW_K, V = XW_V$$

其中 $W_Q, W_K, W_V \in \mathbb{R}^{d \times d_k}$ 为可学习参数矩阵。注意力权重通过缩放点积计算，并作

用于值矩阵实现特征重构:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

相较于传统卷积的局部感受野限制,自注意力机制通过全局关联建模,能够捕获长距离像素间依赖关系。这一机制使得模型能够有效地识别图像中不同区域之间的关联,尤其对于长距离像素间的依赖关系,它表现出了优越的建模能力。相比于传统卷积神经网络仅依赖局部感受野,自注意力机制消除了距离限制,使得模型能够关注到更远的像素信息,从而在处理复杂的图像任务时,提升了特征提取的精度和泛化能力。此外,自注意力机制的计算方式使得模型能够在不同的尺度和语境下自适应地调整注意力权重,进一步增强了对多样化输入的适应能力和表达能力。

2.2 表征学习

表征学习旨在通过自动化的特征提取过程,将原始数据映射到低维潜在空间,使得学习到的表示能够有效捕获数据的内在结构与语义信息。其核心目标是实现特征的可分离性 (Separability) 与鲁棒性 (Robustness), 从而提升下游任务的性能。深度学习的兴起推动了表征学习的突破性进展, 尤其是通过层级非线性变换 (如卷积操作、注意力机制) 实现高阶语义特征的逐层抽象。本节将重点讨论表征学习的两个重要分支: 解耦表征学习和原型学习, 这些方法在提升模型的可解释性和性能方面具有重要作用。

2.2.1 解耦表征学习 (Disentangled Representation Learning)

解耦表征学习 (Disentangled Representation Learning) 要求将数据中的潜在生成因子 (如物体形状、纹理、光照条件) 分离至相互独立的子空间中。可表述为最小化隐变量间的互信息这种分解特性使得模型能够显式地控制数据的生成过程。

解耦表征学习旨在通过特征解耦技术揭示数据底层语义结构, 将复杂数据分解为具有独立物理含义的隐变量, 从而提高模型的可解释性与泛化能力[1][6]。近近年来, 研究者们广泛关注在潜在空间中对特征进行解耦, 以提取跨领域共享的模型[2]。一种常见的实现方法是通过对抗学习进行特征解耦[7][11][15][28], 但对抗学习的优化过程较为复杂, 难以稳定训练[12]。因此, 另一种有效的方法是利用能够显式衡量特征间共享信息的度量准则, 如互信息 (Mutual Information, MI), 其用于衡量两个随机变量之间的共享信息量[10]。

互信息已经被广泛应用于提升分类网络的泛化能力。在视觉识别[3][13][16]和超声图像分类[14]等任务中, 互信息的正则化策略能够有效提升模型的鲁棒性。在滨海湿地植被精细分类任务中, 由于不同数据模态 (如光谱信息与结构信息) 的统计特性

不同, 直接进行模态融合可能引入冗余或无关信息。因此, 通过计算互信息来约束特征解耦过程, 可以增强网络对不同模态特征的适应能力, 提升分类性能。

此外, 在医学图像分割任务中, 许多研究已经验证了解耦表征学习对于跨领域泛化能力的提升[4][15][28]。通过提取解耦的域特征(图像风格)和解剖特征(形状)来提高超声图像分割的泛化能力, 并减少因采集设备或参数变化带来的性能下降[28]。类似地, 在滨海湿地植被分类任务中, 针对不同遥感数据源的特性, 通过解耦表征学习来分离环境背景信息和植被类别特征, 可有效减少数据分布偏差带来的性能下降, 提高分类模型的适应性。

综上所述, 解耦表征学习为多模态遥感数据的融合与分类提供了一种有效的方法, 能够提升模型的泛化能力, 减少数据分布差异的影响。结合互信息优化策略, 能够进一步增强特征解耦的效果, 使模型在不同数据模态之间保持稳定性和鲁棒性, 从而提升滨海湿地植被精细分类的准确性。

2.2.2 原型学习 (Semantic segmentation methods)

原型学习是一种基于认知心理学的机器学习方法。认知心理学研究表明, 人们在学习解决问题时, 会以过去的案例作为模型[1,88,127]。在众多机器学习算法中, 从基于经典统计的方法到支持向量机, 再到多层感知机等, 原型学习因其示例驱动的特性和直观的解释性而受到关注[9,32,34,98]。在原型学习中, 原型通常是从训练数据中选择的一组代表性样本, 每个原型代表一个类别或簇。它通过将观察结果与代表性示例直接比较来进行分类, 使得分类过程更加直观和易于理解。

原型学习的基础是最近邻法则, 这是最早的原型学习方法之一[24]。基于这一基础, 许多著名的非参数分类器被提出, 如学习向量量化(LVQ)[62]、广义LVQ[96]以及邻域成分分析[38,94]。

近年来, 将深度学习与原型学习相结合的研究逐渐兴起, 并展现出良好的潜力。这种结合不仅在少样本[100]、零样本[55]和无监督学习[118,122]中表现突出, 还在监督分类[39,84,117,125]和可解释网络[66]等领域取得了显著成果。例如, 在少样本语义分割任务中, 许多模型可以被视为基于原型的网络[30,108,111], 这为两个分割领域的更紧密合作提供了可能性。

通过以往的经验, 原型学习能够为新问题提供直观的解决方案, 这在复杂问题的解决中具有重要的意义。此外, 将深度学习与原型学习相结合, 不仅提高了模型的性能, 还为模型的解释性提供了新的思路, 这在当前机器学习领域中是一个重要的研究方向。

2.3 图卷积 (Graph Convolutional Network)

图卷积网络(Graph Convolutional Network, GCN)通过将卷积操作从欧几里得空

间拓展至非结构化图数据，实现了对复杂拓扑关系的高效建模[42]。其核心创新源于谱图理论与深度学习的融合，提出了一种基于空间域邻接节点特征聚合的简化框架。相较于传统方法依赖人工设计图核函数或复杂频域变换，GCN 采用局部谱近似理论，将图卷积定义为节点及其邻域的加权特征传播（如公式 2.11 所示）。该框架通过端到端学习机制自适应捕获节点属性与拓扑结构的联合表征，为图结构数据的半监督学习奠定了理论基础。

传统卷积神经网络（CNN）依赖规则的局部感受野，而图卷积的核心挑战在于如何定义非规则图结构上的卷积操作。GCN 通过邻接矩阵 A 和节点特征矩阵 X 的联合建模，提出特征传播规则。

$$H^{(l+1)} = \sigma \left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)} \right)$$

其中 $\tilde{A} = A + I$ 为添加自连接的邻接矩阵， \tilde{D} 为其度矩阵， $W^{(l)}$ 为可训练参数。

图 2.4 展示了图卷积神经网络（GCN）的示意图。首先，将整张待处理的图数据输入网络。在第一层图卷积（Graph Convolution）中，每个节点从其邻居节点聚合信息，即执行一次卷积操作，并用卷积结果更新自身特征。随后，特征经过非线性激活函数 ReLU，再依次通过另一层图卷积和激活函数。该过程不断重复，直到网络达到预设的深度。最终，网络输出更新后的节点特征。

2.4 混合专家（Mixture of Experts）

混合专家系统（Mixture of Experts, MoE）是一种在机器学习领域中广泛应用的模型架构，它通过结合多个专家网络的输出来提高模型的性能和灵活性，能够有效的提高计算效率的同时降低资源消耗。混合专家的基本思想是利用门控网络来动态地选择和组合不同专家网络的输出，从而针对不同的输入数据选择最合适的专家进行处理。

MoE 由多个专家网络和一个或多个门控网络组成。专家网络通常为前馈神经网络，负责从输入数据中提取和学习不同的特征和模式。门控网络则根据输入数据的特性，为每个专家分配相应的权重，决定各专家在最终输出中的贡献比例。图 2.7 展示了 MoE 结构的一种实现方式。通过这种动态的专家选择和权重分配机制，MoE 能够灵活地适应不同任务和数据分布，从而提升模型的泛化能力和整体性能。

混合专家系统通过动态选择最合适的专家来处理不同的数据样本，显著提高了模型的表现力和灵活性。MoE 的核心优势在于其“稀疏激活”机制，只激活部分专家网络，而不是同时调用所有专家。这种机制不仅有效减少了计算资源消耗，还提高了训练和推理的效率。此外，MoE 通过门控网络自动分配权重，确保每个专家能够针对不同的输入数据进行优化，进一步提升了模型在处理复杂任务时的适应性和泛化能力。得益于其独特的结构和机制，MoE 在处理大规模数据集和多样化任务时表现出色，成为提升模型效率和准确性的有效工具。

2.4 本章小结 (Summary)

本章系统地介绍了本文研究工作所依赖的理论基础和相关技术。首先深入介绍了深度学习基础，涵盖卷积神经网络的架构细节，包括卷积层、池化层、激活函数及全连接层的作用与计算方式，以及注意力机制在视觉领域的三种典型实现形式和自注意力机制的原理与优势。接着探讨了解耦表征学习和原型学习这两种表征学习的重要分支，阐述了解耦表征学习通过特征解耦提升模型泛化能力的原理，以及原型学习基于过去案例进行分类的直观特性。随后介绍了图卷积网络，阐述了其通过邻接矩阵和节点特征矩阵的联合建模实现图结构数据高效建模的方式。最后阐述了混合专家系统的架构与核心优势，即通过门控网络动态选择专家网络以提升模型性能和灵活性，同时降低资源消耗。这些理论与技术共同为后续章节的研究工作奠定了坚实基础，提供了多维度的思路与方法支撑。

3 基于特征解耦与原型学习的滨海湿地精细制图方法

3 Attention-enhanced fine mapping method for coastal wetland

3.1 引言 (Foreword)

近年来,受自然和人为因素影响,滨海湿地本土植物群落持续退化,严重威胁到迁徙鸟类的食源和栖息地。利用卫星遥感数据绘制不同土地覆被类型的地图,对滨海湿地自然资源的恢复与保护和生态服务具有重要意义。高光谱遥感技术为滨海湿地植被分类提供了独特的光谱信息优势,但在实际应用中仍面临显著挑战。由于土壤盐度和湿度变化范围较大,碱蓬、芦苇等物种分化为不同的生态型,在遥感影像中光谱呈现高变异性传统分类方法难以准确捕捉植被类别间的细微光谱差异,从而导致分类精度下降。因此,如何解决复杂背景干扰导致的滨海湿地地物同物异谱现象下提取具有判别性的固有光谱特征,以提升滨海湿地精细制图的分类准确性和鲁棒性,成为当前滨海湿地精细制图研究的重要科学问题。

针对上述问题,本章节提出了一种用于滨海湿地精细制图的互信息解耦原型对齐网络 (Mutual Information Disentangled Prototype Alignment Network, MID-ProNet),旨在从特征解耦与原型学习的角度优化滨海湿地高光谱地物分类性能。MID-ProNet 由类别固有特征学习分支和类别无关特征重建分支两个分支组成。其中,类别固有特征学习分支专注于提取不受光照条件、观测角度等外部因素影响的滨海湿地地物固有特征。类别无关特征重建分支则着重提取与光照条件、传感器特性强相关的图像领域表征。通过互信息约束实现两类特征的有效解耦,并结合原型学习策略建立稳健的类别知识表示,从而显著提升滨海湿地复杂环境下的分类精度和泛化能力。为验证方法有效性,本研究选取黄河口滨海湿地的高光谱影像进行实验分析,结果表明,相较于主流分类模型,该方法在复杂场景下的特征鉴别能力与分类一致性均有显著提升,尤其在光谱渐变区域与生境异质区域展现出更优的适应性。

3.2 MID-ProNet: 互信息解耦原型对齐网络 (Mutual Information Disentangled Prototype Alignment Network , MID-ProNet)

3.2.1 MID-ProNet 结构

本章提出了一种面向滨海湿地植被分类的互信息解耦原型对齐网络 (Mutual Information Disentangled Prototype Alignment Network, MID-ProNet),其整体架构如图 3.1 所示。MID-ProNet 由地物固有特征提取器 (Intrinsic Feature Extractor, IFE)、类

别无关特征提取器（Category-Independent Feature Extractor, CIFE）、噪声重建模块（Noise Reconstruction Module, NRM）和动态原型学习模块（Dynamic Prototype Learning Module, DPLM）四部分组成。MID-ProNet 通过互信息计算实现特征解耦，并结合动态原型学习，优化类别判别能力，从而提升滨海湿地地物分类的准确性。

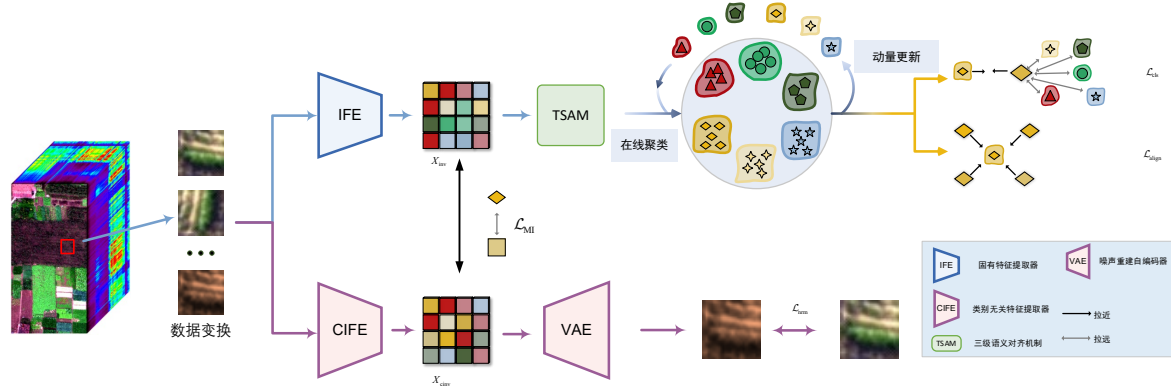


图 3.1 MID-ProNet 网络结构图

Fig. 3.1 Data preprocessing process

具体而言，对于原始高光谱遥感影像 $X \in \mathbb{R}^{H \times W \times B}$ 可以形式化一个高光谱遥感影像数据集 $\mathcal{D} = \{(X_i, Y_i)\}_{i=1}^N$ ，其中 $X \in \mathbb{R}^{H \times W \times B}$ 代表高光谱图像， H 和 W 分别表示图像的高度和宽度， B 是光谱波段数， $N = H \times W$ 为总像素数。对于每个像素，其对应的类别标签 Y 被表示为 one-hot 向量 $Y_i \in \mathbb{R}^C$ ，其中 C 为类别数。数据集被划分为训练集 $\mathcal{D}_{\text{train}}$ 和测试集 $\mathcal{D}_{\text{test}}$ 。为了提高模型对复杂环境的适应能力，训练集 $\mathcal{D}_{\text{train}}$ 先进行数据增强，包括随机旋转、添加高斯噪声等，以模拟光照条件变化与传感器噪声，从而构造增强后的训练集 $\mathcal{D}'_{\text{train}}$ 。数据增强操作可表示为变换函数 $T(\cdot)$ ：

$$\mathcal{D}'_{\text{train}} = \{(T(X_i), Y_i) | (X_i, Y_i) \in \mathcal{D}_{\text{train}}\}$$

增强后的高光谱影像被输入 MID-ProNet 进行特征提取和分类。MID-ProNet 采用 H 型双分支结构，分别为滨海湿地地物原型学习分支和类别无关表征学习分支。其中，滨海湿地地物原型学习分支由地物固有特征提取器（IFE）和动态原型学习模块（DPLM）组成，负责提取地物的固有表征，并通过原型学习生成类别原型，以确保分类的稳定性和可辨识性。类别无关表征学习分支则类别无关特征提取器（CIFE）和噪声重建模块（NRM）组成，负责提取类别无关的背景特征，并利用 NRM 进行噪声重建，以增强模型对复杂环境的适应能力。MID-ProNet 通过计算 IFE 提取的地物固有特征与 CIFE 提取的类别无关表征之间的互信息，实现特征解耦，确保地物类别信息不受环境因素干扰。此外，在训练过程中，MID-ProNet 通过 DPLM 动态更新类别原型，使其逐步逼近真实类别的固有表征，从而提升模型的泛化能力。如图 3.2 所示，MID-ProNet 包括训练和推理两个阶段。在训练过程中，MID-ProNet 计算 IFE 提取的地物固有特征与 CIFE 提取的类别无关表征之间的互信息，从而实现特

征解耦，确保地物的类别信息不受环境因素干扰，并且动态更新类别原型。

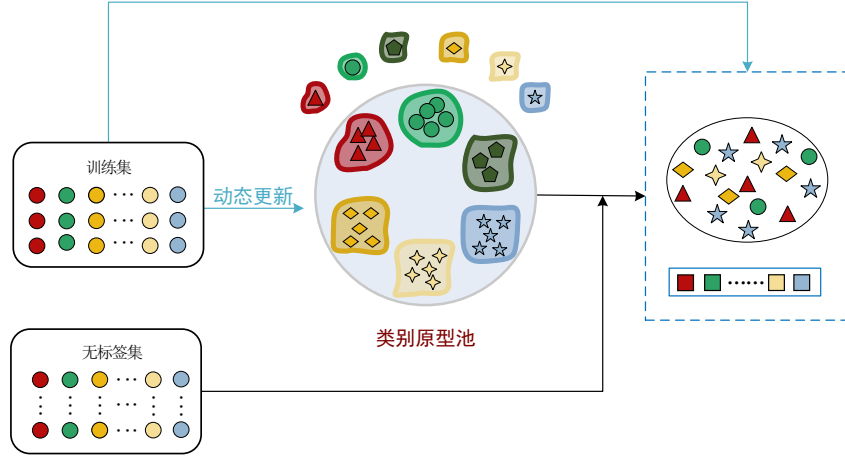


图 3.2 MID-ProNet 网络训练和测试流程图

Fig. 3.2 Data preprocessing process

在推理阶段，待分类的高光谱影像 $\mathcal{D}_{\text{test}}$ 直接送入 MID-ProNet 的滨海湿地地物原型学习分支，利用地物固有特征提取器 (IFE) 提取测试样本的固有特征表示 X_{IFE} 。基于训练阶段学习到的类别原型，采用多级语义对齐机制生成当前样本的初始原型。通过类别原型库的类别原型引导测试样本的初始原型，使其更接近真实类别的表征，并通过动态调整完成最终的语义对齐。最终，对齐后的样本原型被送入分类器进行类别预测，获得最终的分类结果 \hat{Y} 。推理过程中，MID-ProNet 不再更新类别原型，而是依靠训练阶段学习到的类别先验信息进行分类，从而确保分类的稳定性和泛化能力。

3.2.2 数据预处理

在训练数据构建过程中，为了增强模型对环境变化的鲁棒性，本文对训练集 $\mathcal{D}_{\text{train}}$ 进行数据增强操作，构造增强后的数据集 $\mathcal{D}'_{\text{train}}$ 。数据增强策略包括随机旋转 ($\pm 90^\circ$)、翻转、随机高斯噪声 $\epsilon \sim \mathcal{N}(0, \sigma^2)$ 、光谱通道扰动等，以模拟实际遥感观测条件下的环境变化，提高模型在不同光照条件、成像角度和噪声水平下的泛化能力。数据增强过程可形式化表达为变换函数 $T(\cdot)$ ：

$$\mathcal{D}'_{\text{train}} = \{ \{ T(X_i), Y_i \} | (X_i, Y_i) \in \mathcal{D}_{\text{train}} \}$$

其中， $T(\cdot)$ 为数据增强操作集合，通过不同方式对原始样本进行变换。经过预处理后的高光谱影像数据被送入 MID-ProNet 进行特征提取和分类，确保模型能够有效适应滨海湿地复杂环境，提高分类稳定性和鲁棒性。

3.2.3 滨海湿地地物原型学习分支 (Prototype Learning Branch, PLB)

滨海湿地地物原型学习分支 (Prototype Learning Branch, PLB) 旨在提取地物

的固有光谱特征，并通过动态原型学习建立稳健的类别表征，以提升分类的鲁棒性和泛化能力。由于滨海湿地的光谱特征受土壤水分、盐度以及物种生态型分化等因素影响较大，传统分类方法难以准确捕捉类别间的细微光谱差异。因此，本研究通过构建地物固有特征提取器（Intrinsic Feature Extractor, IFE）和动态原型学习模块（Dynamic Prototype Learning Module, DPLM），从光谱特征解耦与类别表征优化的角度提升滨海湿地植被分类的准确性。

3.2.3.1 地物固有特征提取器（IFE）

地物固有特征提取器（Intrinsic Feature Extractor, IFE）被设计用来提取不受外界环境影响的固有地物特征，确保分类结果的稳定性和可辨识性。在滨海湿地高光谱遥感影像中，不同类别的地物具有各自独特的光谱反射特性，但由于外部因素（如光照强度、湿度、土壤成分等）的影响，相同类别的地物在不同条件下可能表现出较大的光谱变化，形成“同物异谱”现象。IFE 通过深度特征提取网络，学习地物类别的核心光谱特征，去除与环境因素相关的干扰信息。

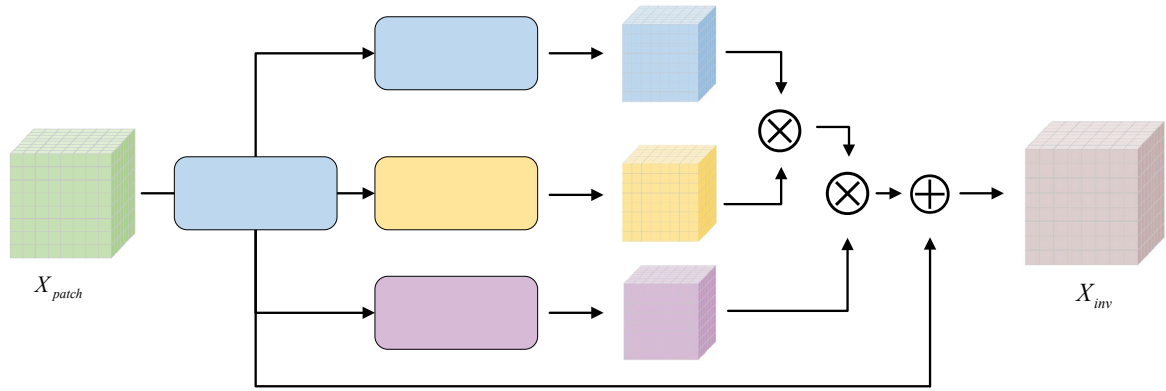


图 3.3 数据预处理流程

Fig. 3.3 Data preprocessing process

IFE 的结构如图 3.3 所示，IFE 采用多个层级的卷积神经网络进行特征提取。具体而言，从经过数据增强后的训练集 D'_{train} 中获得以标记像素为中心的高光谱图像块

$X_{\text{patch}} \in \mathbb{R}^{s \times s \times B}$ ，并将其输入到 IFE，其中 $s \times s$ 是空间大小， B 为光谱波段数量。首先，通过 1×1 卷积压缩光谱维度至 $B/4$ ，减少冗余信息的同时增强跨波段特征交互。

同时，为增强模型对滨海湿地复杂光谱特征的表达能力，网络引入多尺度深度可分离卷积结构，通过 3×3 、 5×5 、 7×7 三种卷积核并行提取局部细节特征、中程关联特征及全局上下文特征。深度可分离卷积结构如图 3.4 所示，该结构通过将标准卷积分解为深度卷积和逐点卷积两个步骤，有效降低了计算复杂度和参数量。并参考自

注意力机制进行不同尺度特征之间的融合，进一步捕捉跨波段的长期依赖关系，增强模型对细微光谱差异的敏感性，以获取稳定的类别特征，这个过程可以形式化表示为：

$$F_{dw}^k = \mathcal{W}_{dw}^k \odot F_c, F_{pw}^k = \mathcal{W}_{pw}^k * F_{dw}^k$$

其中 $k \in \{3, 5, 7\}$ 为卷积核尺寸， \odot 表示深度卷积操作。多尺度特征通过哈达玛乘积（Hadamard Product）进行融合，强化跨波段长程依赖关系的建模能力：

$$F_{multi} = F_{dw}^3 \circ F_{dw}^5 \circ F_{dw}^7$$

为避免深层网络梯度退化，设计残差连接结构将多尺度特征 F_{multi} 与压缩特征 F_c 融合。融合后的特征经 1×1 卷积调整通道维度后，通过逐元素相加与归一化操作生成最终输出：

$$X_{inv} = \text{GELU}\left(\text{BN}\left(\mathcal{W}_f * F_{multi} + F_c\right)\right)$$

该架构通过光谱压缩抑制环境噪声，利用多尺度卷积捕捉空间-光谱联合特征，并结合残差机制保留浅层判别信息，最终实现地物本质特征的稳定表达。

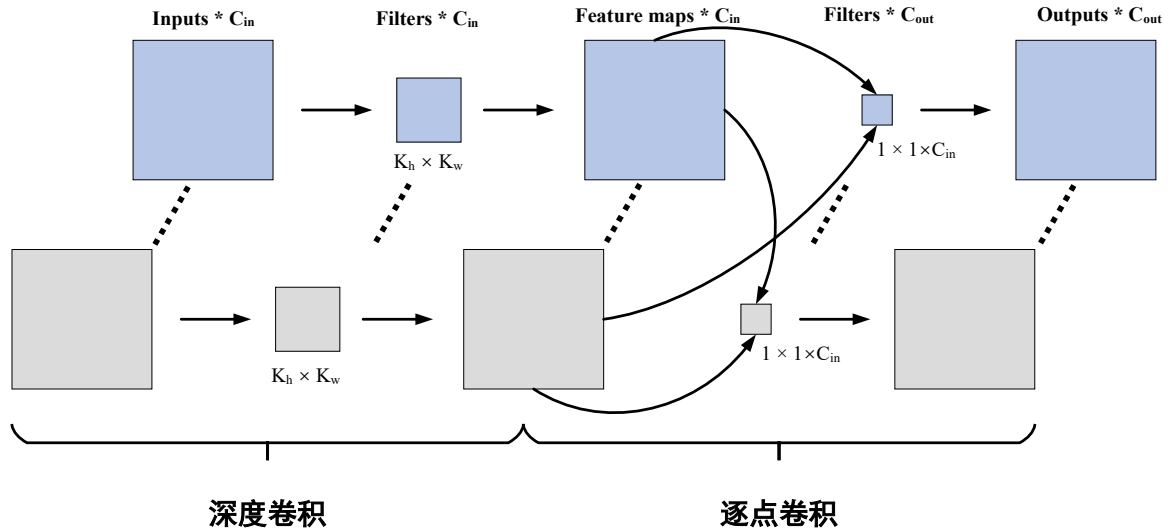


图 3.4 数据预处理流程

Fig. 3.3 Data preprocessing process

3.2.3.2 动态原型学习模块（DPLM）

动态原型学习模块（Dynamic Prototype Learning Module, DPLM）用于建立稳健

的类别表征，并在训练过程中动态更新类别原型，以优化模型的分类性能。传统的监督学习方法通常依赖于固定的类别中心进行分类，但在高光谱遥感影像分类任务中，由于“同物异谱”和“异物同谱”问题的存在，固定类别中心难以充分表征类别特征。因此，DPLM 通过在线更新类别原型，使其逐步逼近真实类别的固有表征，从而提升分类的鲁棒性。

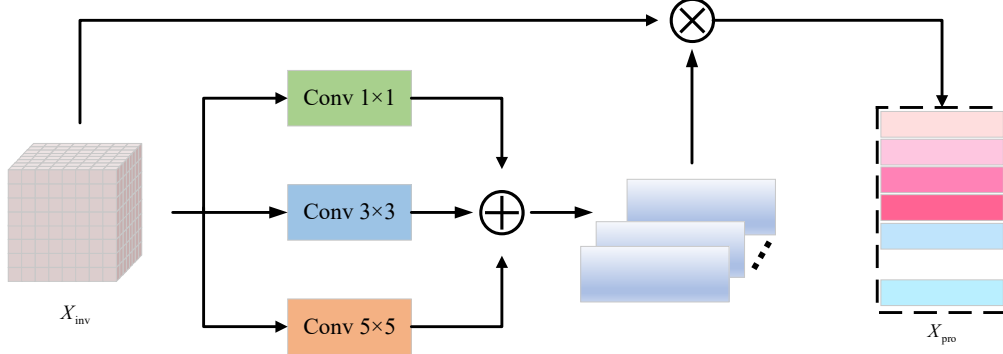


图 3.4 多级语义编码器

Fig. 3.3 Data preprocessing process

在 DPLM 中，引入了多级语义编码器，构建了一种层次化动态原型表征框架，以实现从像素特征到类别原型的层次化映射。DPLM 通过多级语义编码器提取多尺度上下文特征，采用 1×1 、 3×3 和 5×5 三种不同感受野的卷积核并行捕获输入特征的局部语义信息，并通过特征拼接与形状变换操作将其投影至潜在语义空间，生成初级类别语义 X_{pro} 。该语义表征通过在线聚类机制动态维护原型池，其更新过程采用动量平滑策略，可以表示为：

$$p_t = \alpha p_{t-1} + (1 - \alpha) \frac{1}{N} \sum_{i=1}^N f_i$$

其中 p_t 表示当前时刻原型， α 为动量系数， f_i 为当前批次样本特征。该策略通过历史语义与当前特征的渐进融合，抑制单批次数据噪声对语义表征的影响，增强原型池的时序稳定性。训练阶段中，DPLM 基于固有特征提取器（IFE）输出的地物本质特征，构建初级类别语义与原型池的交互机制。通过设计原型引导模块，从原型池中检索历史语义 p^{mem} ，并与当前初级类别语义 p^{cur} 进行跨批次语义融合：

$$p^{enh} = \lambda \cdot p^{cur} + (1 - \lambda) \cdot \text{softmax}(p^{cur} p^{mem\top}) p^{mem}$$

其中 λ 为自适应权重参数，通过注意力机制筛选历史语义中与当前特征相关的全局信息，弥补初级语义的局部偏差。同时，为了约束语义表征 X_{pro}^c 与真实类别中心 μ_c 的空间一致性，优化特征分布的可分性，本章节设计了一个原型对齐损失

$$\mathcal{L}_p = \sum_{c=1}^C \|\mu_c - X_{pro}^c\|_2^2$$

其中， C 表示类别， c 表示具体的类别。推理阶段则固定原型池，通过多级语义编码生成测试样本的初级语义，与存储的原型进行特征空间匹配后送入分类器，确保模型在复杂光谱场景下的分类鲁棒性。

3.2.4 类别无关表征学习分支

类别无关表征学习分支旨在学习领域特征，并通过与类别固有特征进行互信息计算，实现特征解耦。该分支由类别无关特征提取器（Category-Independent Feature Extractor, CIFE）和噪声重建模块（Noise Reconstruction Module, NRM）组成。通过提取不依赖类别信息的领域特征，并利用互信息最小化约束与类别固有特征的相关性，该分支能够有效增强模型的特征分离能力，从而提升高光谱遥感分类的精度。

类别无关特征提取器（CIFE）主要用来从经过数据增强的高光谱图像中学习类别无关的光照等领域特征，从而促进滨海湿地地物原型学习分支学习到更精确的地物固有特征。CIFE 利用浅层卷积神经网络提取局部特征，并结合 Transformer 结构捕捉长距离相关性。CIFE 通过计算 IFE 之间的互信息，约束其输出与类别固有特征提取器提取特征之间的互信息最小化，以确保该分支提取的仅为类别无关信息。

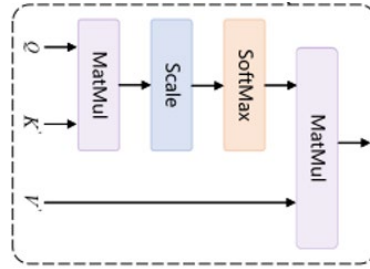


图 3.4 自注意力机制编码器

Fig. 3.3 Data preprocessing process

在具体实现过程中，从经过数据增强后的训练集 D'_{train} 中获得以标记像素为中心的高光谱图像块 $X_{\text{patch}} \in \mathbb{R}^{s \times s \times B}$ ，首先针对 X_{patch} 中心像元 $X_{\text{centra}} \in \mathbb{R}^{1 \times 1 \times B}$ 提取局部特征，并通过扩展操作，提高维度至 $X'_{\text{centra}} \in \mathbb{R}^{s \times s \times B}$ ，之后 X'_{centra} 作为自注意力机制中的 Q， X_{patch} 作为 K 和 V 输入到 Transformer 结构中，计算中心像元作为自注意力机制与输入图像块之间的长程依赖，并将特征映射到潜在空间得到 X_{cinv} ，去除类别信息的影响。最后，通过互信息最小化约束类别无关特征与类别固有特征的相关性，从而实现特征解耦。该模块的设计使得高光谱影像在复杂背景下依然能够保持稳定的类别判别能力。

噪声重建模块（NRM）基于自编码器架构组成，输入为 CIFE 提取的类别无关特

征 $X_{\text{cinv}} \in \mathbb{R}^{H \times W \times C}$ 。为了提高模型对于领域特征的提取能力，重建添加的环境噪声，解码器通过反卷积层逐步重建原始特征，其目标函数定义为：

$$\mathcal{L}_{\text{nrm}} = \| \text{Decoder}(F_{\text{noisy}}) - F_{\text{domain}} \|_2 + \lambda \cdot \text{KL}(q(z | F_{\text{noisy}}) \| p(z))$$

式中首项约束重建误差，次项通过 KL 散度约束隐变量 z 的分布与标准正态先验的匹配性， λ 为平衡因子。

3.2.5 互信息计算与特征解耦

为实现地物类别特征与混合环境特征的有效解耦，本文通过互信息（Mutual Information, MI）约束类别固有特征提取器（IFE）和类别无关特征提取器（CIFE）的优化过程。互信息定义为 IFE 提取的类别特征 X_{inv} 与 CIFE 提取的混合特征 X_{cinv} 的联合分布与边缘分布乘积之间的 Kullback-Leibler (KL) 散度。通过最小化互信息，可抑制两类特征之间的冗余相关性，确保 IFE 专注于地物固有光谱表征，而 CIFE 仅捕获环境干扰因素。具体而言，基于 Donsker-Varadhan 表示，互信息的下界可通过神经网络 $T_{\theta_{\text{MINE}}}$ 进行估计，其优化目标为最大化联合分布样本与边缘分布样本的期望差异。

利用蒙特卡洛方法，从训练批次中分别采样联合分布对 $(X_{\text{inv}}, X_{\text{cinv}})$ 和边缘分布对 $(X'_{\text{inv}}, X'_{\text{cinv}})$ ，并通过梯度上升更新 $T_{\theta_{\text{MINE}}}$ 的参数以逼近互信息的真实值。最终，互信息损失定义为负的下界估计值，反向传播至 IFE 和 CIFE 网络以驱动特征解耦。

在特征解耦过程中，IFE 通过多尺度深度可分离卷积提取地物固有光谱特征，而 CIFE 结合浅层卷积与 Transformer 结构建模环境相关的长程依赖特征。两类特征的互信息损失约束其联合分布趋近于边缘分布的独立性假设，从而消除环境噪声对类别判别特征的干扰。此外，动态原型学习模块（DPLM）通过原型对齐损失进一步优化类别表征的类内紧致性与类间可分性，而噪声重建模块（NRM）则通过自编码器重构环境噪声以增强 CIFE 的鲁棒性。

MID-ProNet 的总损失函数由分类损失、互信息解耦损失、原型对齐损失及噪声重建损失四部分构成，具体形式化表示为：

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{cls}} + \alpha \mathcal{L}_{\text{MI}} + \beta \mathcal{L}_{\text{align}} + \gamma \mathcal{L}_{\text{nrm}}$$

其中， \mathcal{L}_{cls} 为交叉熵分类损失， \mathcal{L}_{MI} 为互信息解耦损失， $\mathcal{L}_{\text{align}}$ 为原型对齐损失， \mathcal{L}_{nrm} 为噪声重建损失， α 、 β 、 γ 为各损失项的平衡系数。该多任务损失框架通过联合优化特征解耦、原型对齐及噪声抑制，显著提升了滨海湿地复杂场景下的地物分类精度与鲁棒性。

3.3 实验与分析（Experiments and analysis）

3.3.1 实验环境

MID-ProNet 网络模型是基于 PyTorch 深度学习框架构建的。所有的实验工作均在 Ubuntu 18.04 操作系统环境下进行,本文利用 GeForce RTX 2080Ti GPU 以及 Intel(R) Xeon(R) Silver 4214R CPU 来完成网络模型的训练与测试。关于实验环境的详细信息,如表 3.1 所示。

表 3.1 实验环境详细信息

Table 3.5 The experimental environment details

实验环境	参数
系统	Ubuntu 18.04.5 LTS
CPU	Intel(R) Xeon(R) Silver 4214R CPU @ 2.40GHz
GPU	GeForce RTX 2080Ti
RAM	128GB

3.3.2 实验数据集

为了评估所提方法的有效性,采用了两个公开可用的黄河口滨海湿地高光谱图像数据集: Coast 数据集、NC16 数据集和公开数据集 Indian Pines (IP), 分别随机抽取 10%、2.16%和 10%样本作为训练集。下面详细介绍每个数据集。

Coast 数据集源自 2018 年 8 月 2 日 Zhuhai-1 卫星对山东省东营市垦利区黄河口滨海湿地的观测成果。该卫星搭载的成像传感器在 400-1000 nm 波长范围内进行光谱成像,光谱分辨率为 2.5 nm, 空间分辨率为 10 米。图像尺寸为 1024×1024 像素, 涵盖了湿地生态系统中多种地物类型, 如滩涂、水体、植被等。整个数据集包含 1024 个光谱波段, 经过预处理后, 实际用于分析的波段数量为 1024 个。图 3.2 展示了该高光谱数据集的假彩色合成图及其对应的地面真值参考图。

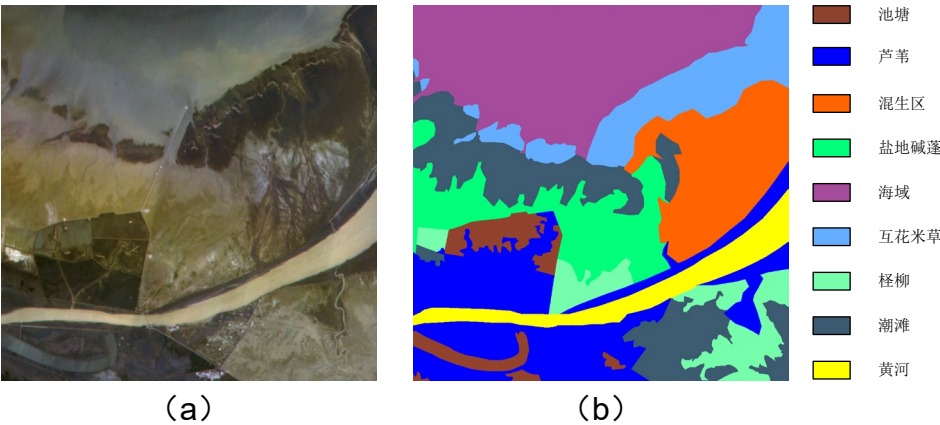


图 2.2 Coast 数据集 (a)假彩色图像 (b)地面真值图
Fig. 2.12 Coast dataset (a) False color composite (b) Ground truth

表 3.2 Coast 数据集上每个类别的训练和测试集的数量

Table 3.2 Number of training and test samples used in Coast datasets

类别编号	地面覆盖类型	样本数量
1	池塘	96395
2	芦苇	182815
3	混生区	63759
4	盐地碱蓬	101347
5	海域	117997
6	互花米草	145829
7	怪柳	57664
8	潮滩	40158
9	黄河	242612
Total		1048576

NC12 数据集源自于 2020 年 9 月 23 日对某地区的高光谱观测。该数据集的获取时间范围为 12:27 - 13:10，天气晴朗无云，为数据采集提供了良好的大气条件。无人机飞行高度为 300 m，使得 NC12 数据集的空间分辨率达到了约 0.182 m。数据集中涵盖了 12 种类型的土地覆盖，为地物分类和环境研究提供了丰富的样本。其光谱成像范围广泛，能够捕捉到不同地物在多个波段下的特征，为后续的分析 and 研究提供了坚实的数据基础。

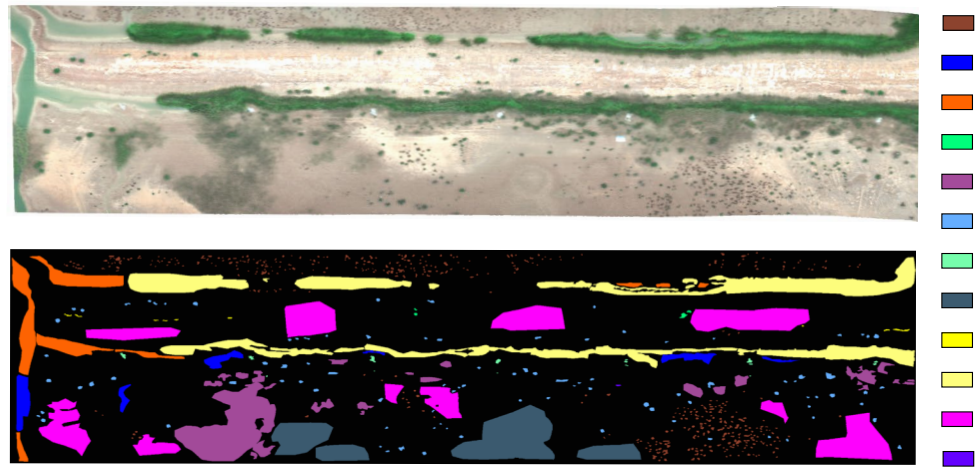


图 2.2 YRD-NC12 数据集 (a)假彩色图像 (b)地面真值图

Fig. 2.12 YRD-NC12 dataset (a) False color composite (b) Ground truth

表 3.2 YRD-NC12 数据集上每个类别的训练和测试集的数量

Table 3.2 Number of training and test samples used in YRD-NC12 datasets

类别编号	地面覆盖类型	样本数量
1	Suaeda glauca	14117
2	Reed	15603
3	Water	36291
4	White cloth	428

5	Withered reed	61284
6	Tamarix	9257
7	Stone	1167
8	Moist soil	68497
9	Mudflat	634
10	Spartina alterniflora	124320
11	Dry soil	120073
12	Standard reflectance cloth	75
Total		451746

Indian Pines (IP) 数据集由机载可见/红外成像光谱仪于 1992 年获取的，它主要对美国印第安纳州西北部的印度松树进行了成像。该数据集的空间尺寸为 145×145 ，涵盖了 220 个光谱反射波段，波长范围为 $0.4\text{-}2.5\mu\text{m}$ ，空间分辨率达到 20 米/像素 (*mpp*)。然而，由于噪声和吸水性，去除了 20 个光谱波段，其余 200 个波段作为研究的对象。图 2.11 展示了 IP 数据集的假彩色图及其对应的地面真值参考图。

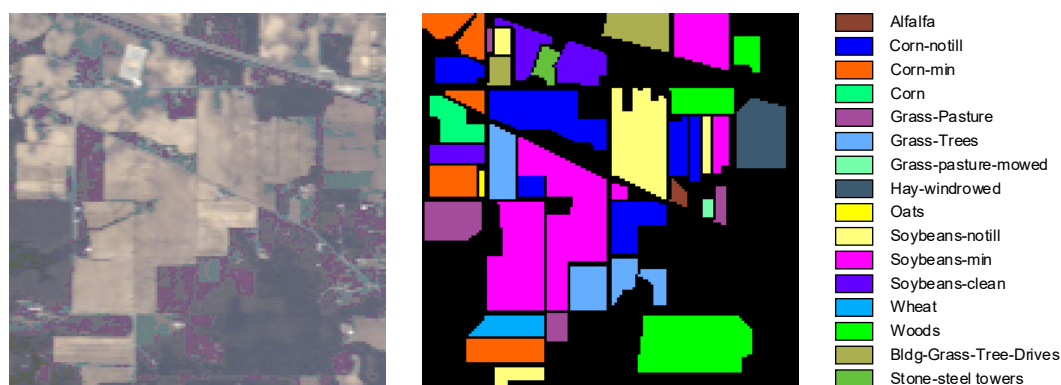


图 2.11 Indian Pines 数据集 (a) 假彩色图像 (b)地面真值图

Fig. 2.11 Indian Pines dataset (a) False color composite (b) Ground truth

图 2.2 Indian Pines 数据集 (a)假彩色图像 (b)地面真值图

Fig. 2.12 Indian Pines dataset (a) False color composite (b) Ground truth

表 3.2 Indian Pines 数据集上每个类别的训练和测试集的数量

Table 3.2 Number of training and test samples used in Indian Pines datasets

类别编号	地面覆盖类型	样本数量
1	Alfalfa	46
2	Corn-notill	1428
3	Corn-min	830
4	Corn	237
5	Grass/Pasture	483
6	Grass/Trees	730
7	Grass/pasture-mowed	28

8	Hay-windrowed	478
9	Oats	20
10	Soybeans-notill	972
11	Soybeans-min	2455
12	Soybeans-clean	593
13	Wheat	205
14	Woods	1265
15	Bldg-Grass-Tree-Drives	386
16	Stone-steel towers	93
Total		10249

3.3.3 评价指标

混淆矩阵是评估分类模型性能的重要工具。它清晰地呈现了模型预测结果与实际类别之间的关系，并为计算多个关键评估指标提供了基础。如表 3.1 所示，混淆矩阵的每一行对应一个真实类别，每一列对应一个预测类别。对角线上的元素表示正确分类的样本数量，而非对角线上的元素则表示错误分类的样本数量。。

表 3.1 混淆矩阵

Table3.1 The confusion matrix

预测类别 真实类别	类别 1	类别 2	...	类别 n	总计
类别 1	x_{11}	x_{12}	...	x_{1n}	x_{n+}
类别 2	x_{21}	x_{22}	...	x_{2n}	x_{2+}
...
类别 n	x_{n1}	x_{n2}	...	x_{nn}	x_{n+}
总计	x_{+1}	x_{+2}	...	x_{+n}	N

总体准确率（Overall Accuracy, OA）是衡量分类模型性能的一个常用指标，它反映了模型正确预测的样本数在总样本数中的占比。OA 值越高，通常意味着模型的预测效果越好。具体计算公式如下：

$$OA = \frac{\sum_{i=1}^n \text{正确预测的样本数}}{\text{总样本数}}$$

平均准确率（Average Accuracy, AA）则是另一种评估分类模型性能的指标，它计算的是所有类别准确率的平均值。与 OA 相比，AA 能够更全面地反映模型在各个类别上的表现差异。其计算公式如下：

$$AA = \frac{1}{C} \sum_{i=1}^C \frac{\text{正确预测的第}i\text{类样本数}}{\text{第}i\text{类总样本数}}$$

其中， C 表示类别总数。

Kappa 系数（Kappa Coefficient）是一种统计指标，用于评估分类模型的一致性和准确性，特别是在处理分类任务时。它不仅考虑了模型预测与实际观察之间的一致性，还对随机预测进行了修正。Kappa 值通常在 0 到 1 之间，值越大表示模型的预测结果与实际结果的一致性越高。Kappa 系数的计算公式如下：

$$\kappa = \frac{OA - P_e}{1 - P_e}$$

其中， P_e 是随机一致性的期望概率，计算公式为：

$$P_e = \sum_{i=1}^C \left(\frac{\text{第}i\text{类实际样本数}}{\text{总样本数}} \times \frac{\text{第}i\text{类预测样本数}}{\text{总样本数}} \right)$$

3.3.4 实验结果

为了验证本章提出的 MID-ProNet 在滨海湿地高光谱图像上的有效性，在实验中采用典型的机器学习支持向量机（SVM）和深度学习方法与 3DCNN、HyperSN、morphFormer、SSFTT、HSIFormer 与 MID-ProNet 进行比较。不同方法的预测结果如图 3.7 所示。

表 3.-VIII 中显示了 3 个数据集上拟定方法和所有比较方法的单类准确度、OA、AA 和 κ ，其中最佳结果以粗体突出显示。

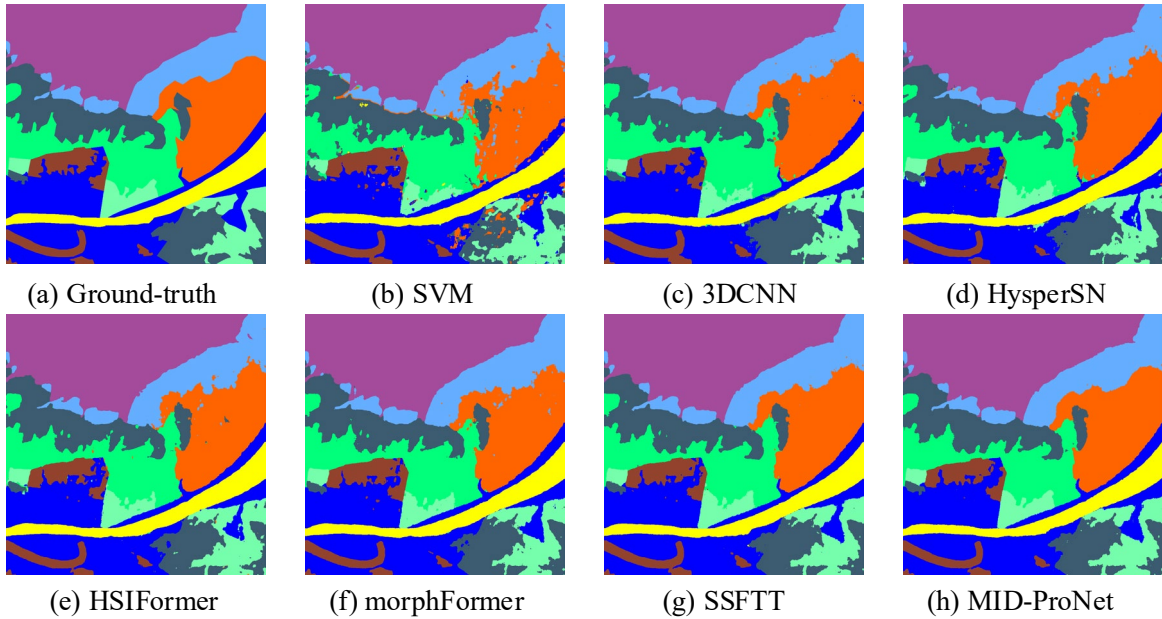


图 3.7 Coast 数据集上的分类结果图

Fig. 3.7 Classification maps for the Coast dataset

第一个实验是在 Coast 数据集上开展的，从图 3.4 中可以看出，SVM 的分类结果呈现出破碎化，含有很多的噪声，这是因为 SVM 只使用了原始图像的光谱信息，忽略了地物之间空间结构信息。为了引入空间信息，部分研究者使用 3D 卷积处理空间和光谱信息，如 3DCNN、HybirdSN 和 DDCD。但是从图像中我们可以看到，尽管分类结果要优于 SVM，但是这些基于 CNN 的分类结果仍然呈现出景观破碎化的现象。这是因为基于 CNN 的方法受限于感受野的大小，无法处理地物之间的长程依赖。MorphFormer 和 SSFTT 的分类图更加平滑和完整，但是仍然在潮滩、混生区和盐地碱蓬的交界地带效果并不理想，存在边界模糊的问题。这是因为高光谱图像中存在的邻近像元效应导致的边界模糊问题。相对而言，本章提出的方法得到了最好的结果。

表 3.6 Coast 分类结果

Table 3.6 Classification result of the Coast dataset							
类别	SVM	3DCNN	DDCD	HybirdSN	MorphFormer	SSFTT	SSCANet
1	84.14	93.12	98.08	88.58	99.16	94.77	97.10
2	92.00	94.93	94.80	97.98	94.38	97.82	98.71
3	84.83	98.14	95.71	94.36	98.28	97.46	98.88
4	88.42	96.29	96.18	94.52	94.63	97.61	98.48
5	98.63	99.70	99.49	99.34	99.51	99.78	99.75
6	85.69	88.26	94.98	97.29	96.81	98.75	98.83
7	76.98	92.40	84.61	88.04	94.99	96.93	96.57
8	95.37	92.48	95.91	94.06	97.25	95.80	97.85
9	96.32	97.79	95.77	96.12	98.73	98.88	98.05
OA(%)	90.38	94.45	95.01	95.80	97.06	97.89	98.61
AA(%)	90.19	95.39	94.62	95.53	95.76	97.36	98.39
$\kappa \times 100$	89.38	94.63	95.02	95.10	96.57	97.54	98.37

在表 3.2 中，可以观察到，尽管基于 CNN 的方法已经实现了良好的分类准确度，但是由于缺乏建模全局上下文信息导致在大部分的单类别精度上要弱于基于 Transformer 的方法。在大部分的类别上，我们的方法取得了最高的数值，特别是对于较难分辨的互花米草和怪柳这两个类别，表明我们的方法面对邻近像元效应导致的边界模糊问题上取得了最好的结果。此外，我们提出的方法在获得了最高数量的最佳单类精度的基础上，还在三个其他的指标 AO、AA 和 kappa 系数上都获得了最高的精度，进一步验证了我们提出的方法的优越性。然而，我们提出的方法也有一些局限性。

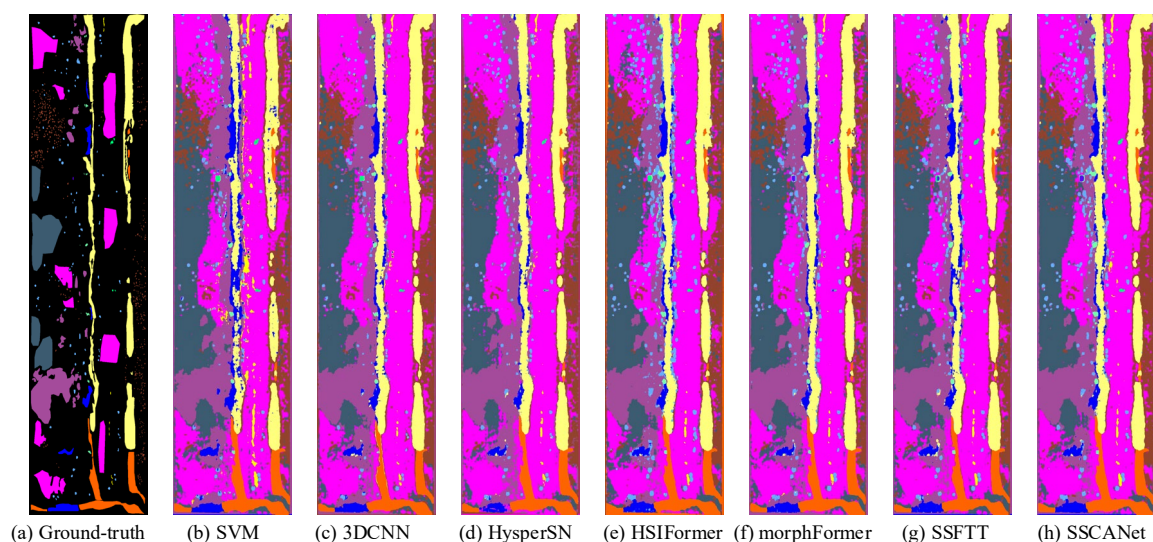


图 3.7 YRD-NC12 数据集上的分类结果图

Fig. 3.7 Classification maps for the YRD-NC12 dataset

第二个数据集是在 NC12 数据集上开展的，不同于 Coast 数据集，NC12 数据集是一个无人机航空遥感影像，空间分辨率要明显优于 Coast 数据集。从分类结果上，我们能够明显看到大部分的方法在 OA 和 kappa 系数上取得了较好的结果，但是 AA 普遍较差。这是因为 NC12 数据集中存在小样本类别，特别是标准反射布。现有的方法无法处理滨海湿地复杂的地物情况下兼顾小样本类别，但是从表中，我们可以看到，我们的结果取得了最优的乘积，在每个小样本类别和其他地物类别上均取得了较优的结果。

表 3.6 YRD-NC12 分类结果

Table 3.6 Classification result of the YRD-NC12 dataset

类别	SVM	3DCNN	HybirdSN	HSIFormer	MorphFormer	SSFTT	MID-ProNet
1	98.22	99.33	99.84	99.71	99.92	99.95	99.96
2	61.20	94.65	98.30	98.56	99.55	97.68	98.90
3	99.93	100.00	99.97	99.94	99.99	99.91	99.91
4	84.22	83.77	100.00	84.10	100.00	92.52	100.00
5	97.81	98.60	98.27	99.84	99.70	99.73	99.89
6	96.42	100.00	99.25	92.05	99.42	99.80	99.46
7	95.31	99.80	100.00	99.73	96.21	100.00	99.82
8	98.49	99.57	99.86	98.01	99.87	99.80	99.85
9	50.70	99.62	100.00	100.00	99.65	99.67	99.84
10	99.98	99.83	99.70	99.88	99.56	99.78	99.90
11	99.83	99.10	98.17	99.86	99.64	99.91	99.88
12	0.00	0.00	64.86	0.00	26.87	100.00	100.00
OA(%)	96.96	99.22	99.08	99.34	99.64	99.75	99.84
AA(%)	87.50	86.75	87.44	89.68	92.66	95.28	99.12
$\kappa \times 100$	96.23	99.03	98.85	99.18	99.55	99.69	99.80

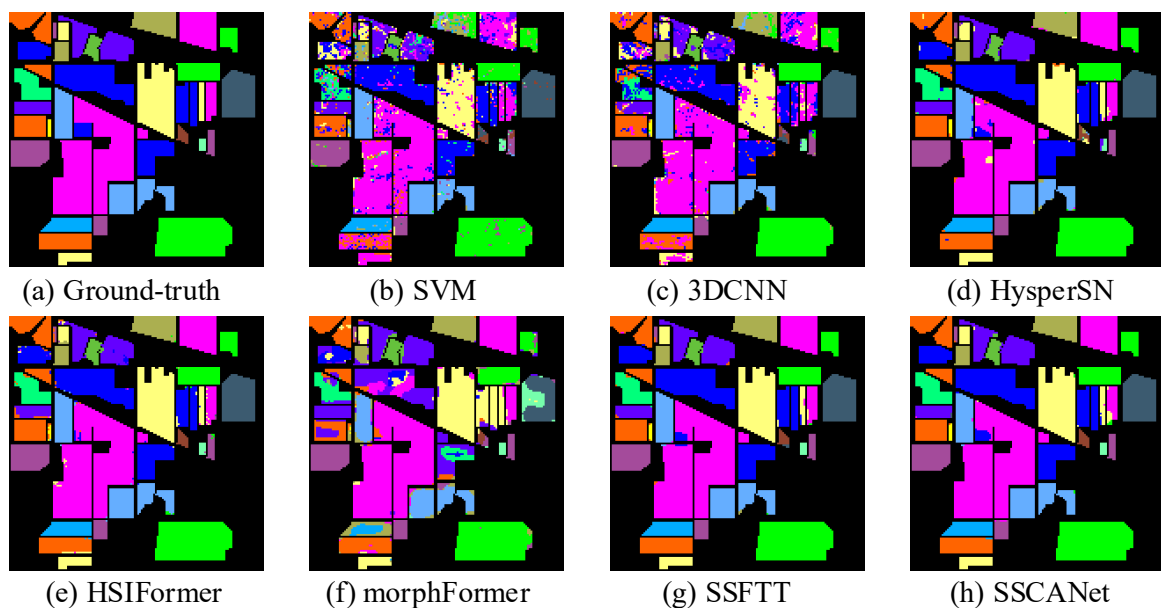


图 3.7 数据集上的分类结果图

Fig. 3.7 Classification maps for the Indian Pines dataset

为了验证网络的泛化能力，第三个对比试验选择在经典的高光谱地物分类数据集 Indian Pines 数据集上进行。通过与其他对比方法比较分析，本文得到的分类图最接近真值图。传统的机器学习方法 SVM 的效果较差，图像中存在大量的噪声。在 CNN 作为骨干的网络中，3D-CNN 存在一定程度的噪声，并且对边缘类别的分类效果较差，HyperSN 通过结合 2D 卷积和 3D 卷积的空间特征，比 3D-CNN 产生了更好的视觉效果。相比之下，融合了 CNN 和 Transformer 的分类方法得益于提取的丰富空谱特征，分类效果显著优于基于 CNN 方法，并且表现出更优的分类性能。SSFTT 方法基于 Transformer 提取全局空谱信息优于只关注全局光谱信息的 HSIFormer。值得注意的是，每张图像中蓝色标记的“Corn-notill”区域与橙色的“Corn-mintill”区域，两者在视觉上极为相似且难以区分，其他的对比方法都或多或少错误地将其分类。本文的方法在更大程度上识别了蓝色区域与橙色区域，这也证明了本文方法的优异性能

表 3.6 Indian Pines 分类结果

Table 3.6 Classification result of the Indian Pines dataset

类别	SVM	3DCNN	HybirdS N	HSIForme r	MorphFormer	SSFTT	MID- ProNet
1	68.03	100.00	88.73	90.00	100.00	100.00	100.00
2	85.94	76.13	88.89	98.59	94.60	99.40	99.41
3	82.21	81.83	88.08	93.21	72.77	98.55	99.74
4	75.36	88.61	81.91	97.18	79.67	99.29	99.08
5	94.60	95.01	92.16	99.42	92.94	99.21	99.89
6	98.10	96.86	94.92	99.39	100.00	99.62	99.77
7	78.85	97.92	100.00	94.65	54.41	98.08	100.00

8	98.96	95.66	99.05	99.31	97.99	98.32	99.66
9	80.56	100.00	81.82	83.34	75.00	100.00	94.74
10	81.33	74.17	80.84	95.98	73.96	98.41	97.80
11	91.09	81.56	92.56	96.80	74.14	97.69	98.25
12	84.90	82.06	78.58	96.63	63.66	98.49	99.43
13	98.15	100.00	85.88	97.50	96.47	99.73	98.94
14	95.52	95.09	91.56	99.61	98.52	99.44	99.18
15	76.07	91.30	92.90	93.90	76.39	98.43	99.01
16	89.39	97.50	95.24	95.37	87.17	93.73	91.10
OA(%)	80.40	85.22	89.65	97.24	77.99	98.63	98.97
AA(%)	75.49	79.62	79.15	94.78	74.02	97.14	98.54
$\kappa \times 100$	77.54	83.06	88.17	96.86	74.65	98.45	98.83

本章提出的 MID-ProNet 方法通过对光谱信息进行重用，并自适应地增强不同地物的敏感通道，强调光谱相似的地物之间的细微差异，使得整体预测结果较好，改善了混淆分类现象。

3.3.5 消融实验

表 3.5 MID-ProNet 消融实验的精度和效率

Table 3.4 Ablation evaluation accuracies and efficiencies of MID-ProNet

	IFE	MI	DPLM	OA	AA	$\kappa \times 100$
1	×	√	√	98.12	96.45	97.86
2	√	×	√	97.85	97.42	97.55
3	√	√	×	95.90	91.41	95.32
4	√	√	√	98.89	98.64	98.74

3.3.6 模型复杂性评估

如表 3.6 所示，使用 FLOPs 衡量模型的复杂度，FLOPs 值越小表示模型复杂度越低，相应的计算量和算力要求就越小。MID-ProNe 的 FLOPs 值与对比方法相比处于中等水平，且每个 epoch 的训练时间略有增加。但是，如图 3.10 所示，MID-ProNe 的收敛速度比其他模型快得多且取得的分类精度最高，这可以弥补每个 epoch 运行时间更长的缺点。

为了验证 MID-ProNe 网络模型中的冗余性，对网络模型进行剪枝操作。由于 MID-ProNe 中使用 SE Block 与稠密连接结合，导致模型的复杂度较大，本节对其进行剪枝操作，对 MID-ProNe 模型进行剪枝的程度及模型的精度、FLOPs 变化如图 3.10 所示。

在 MID-ProNe 模型训练中加入 L1 正则化,使得权重稀疏化,按特征输出的绝对值大小来评估其重要性,然后将权重较小的部分裁剪掉^[57]。随着模型剪枝程度的增加,模型的整体精度和 FLOPs 相应下降。剪枝程度为 20%时,模型的精度下降了 2.63%,FLOPs 下降了 7.1M,此时模型的 FLOPs 接近对比实验中 FLOPs 最优的 WetlandNet,精度高出其 0.48%,这说明在裁剪了部分冗余的连接后,MID-ProNet 在保持较低模型复杂度的情况下仍能取得最高的精度,证明了模型具有较好的特征学习能力和鲁棒性。这也表明 MID-ProNe 模型存在少量冗余节点,大部分连接参与主要的计算。

3.4 本章小结 (Summary)

本章节详细说明了提出的基于深度语义分割的滨海湿地高光谱遥感影像精细制图方法,该方法主要包括遥感数据预处理、模型训练和模型预测三个步骤。本章节提出注意力增强的语义分割网络 MID-ProNe,综合提取低层细节特征与高层抽象特征,自适应增强地物的显著特征表示,缓解了高光谱影像在语义分割方法中提取精度不高的问题。MID-ProNe 可以封装为全自动的端到端处理程序,在黄河口滨海湿地的高光谱遥感影像上的实验结果表明,相比于机器学习方法及其他语义分割方法,该方法适合于地物类型丰富且分布复杂的滨海湿地,提高了滨海湿地精细制图的准确性。

4. 基于多模态数据融合的滨海湿地植被精细制图方法

4.1 引言

滨海湿地在调节气候、维持生物多样性等方面具有重要生态功能。然而，由于潮汐作用、植被斑块破碎化及入侵物种的影响，湿地植被的光谱和几何特征表现出高度复杂性，使得传统高光谱分类方法在此类场景中精度受限。高光谱遥感能够提供连续的光谱信息，有助于区分湿地植被的细微光谱差异，而激光雷达则提供精确的几何结构信息，弥补了光谱数据在空间结构上的不足。因此，如何充分利用多源遥感数据，实现对滨海湿地植被的精准解析，仍是亟待解决的难题。

针对这一问题，本章提出了一种几何-光谱联合动态形变网络 (Geometric-Spectral Joint Dynamic Deformation Network, GSJDD-Net)，旨在通过深度特征融合提升滨海湿地植被分类的精度与鲁棒性。GSJDD-Net 由四个核心模块组成，分别针对湿地环境的复杂性和植被形态的多样性进行优化。首先，可变形多尺度对齐模块通过引入可变形卷积，自适应调整卷积感受野，使特征提取能够适应局部几何形态，同时利用多尺度并行卷积增强对不同大小植被斑块的表征能力，实现跨尺度特征对齐。其次，几何引导光谱注意力模块结合高程数据的梯度信息构建几何先验，以缓解光谱特征在边界模糊区域的信息损失问题。该模块通过多层感知机 (MLP) 提取全局光谱描述子，并引入几何注意力机制，提升模型对复杂地形的适应性。第三，自适应渐进融合模块采用三级融合策略，先通过直接拼接光谱与激光雷达特征获得初步融合表示，再利用可学习门控机制动态调整光谱与几何信息的融合权重，最终引入基于图卷积的拓扑结构保持机制，增强破碎植被斑块间的空间一致性。最后，动态路由分类模块利用异构专家网络，根据输入特征的复杂度自适应选择最优专家进行分类决策，并通过空间感知机制计算路由权重，提高分类精度。实验基于黄河口滨海湿地的高光谱和激光雷达数据进行验证，结果表明，相较于传统分类方法，GSJDD-Net 在复杂湿地环境下展现出更优的特征分辨能力，特别是在光谱渐变区域和边界模糊区域表现出更高的分类一致性和适应性，为滨海湿地植被的精准分类提供了有效方案。。

4.2 主要流程

4.2.1 研究区域

本章研究区域位于中国山东省东营市的黄河口滨海湿地 ($37^{\circ} 45' \sim 37^{\circ} 51' \text{N}$, $119^{\circ} 06' \sim 119^{\circ} 14' \text{E}$)，北临渤海，东接莱州湾，地理位置优越 (如图 X 所示)。该湿地总面积约 110 km^2 ，是世界少有的河口湿地生态系统之一，在维持生态平衡、生物多样性保护及湿地资源管理方面具有重要价值。黄河口滨海湿地是中国暖温带地区面积

最大、保存最完整的滨海湿地之一，具有典型的河口湿地生态特征。湿地内植被类型丰富，主要包括互花米草、柽柳、碱蓬和芦苇等，植被覆盖度较高。受地势低洼和潮汐影响，部分区域土壤盐碱化程度较高，为耐盐植物提供了适宜的生长环境。然而，近年来，受黄河来水量减少、人类活动加剧以及全球气候变化等因素影响，该地区正面临湿地面积缩减、景观破碎化、生物多样性下降等生态挑战，对湿地监测和保护提出了更高要求。

4.2.2 实验数据

本文所用数据为 2022 年 9 月 18 日采集于山东省东营市垦利区黄河口滨海湿地上空的 chengliu 数据集，涵盖高光谱图像与激光雷达图像。高光谱图像由 DJI M600 无人机搭载德国 CUBERT s185 高光谱成像传感器获取，包含 450 - 950nm 的 125 个波段，空间分辨率为 0.5m。激光雷达图像由飞马 D LiDAR 2000 采集，具有 905nm 波段、最大 220 points/m² 的点云密度及三回波类型，空间分辨率约为 0.5m。无人机飞行高度为 300m。该数据集为研究区域的生态环境、土地利用类型等提供了丰富的信息。

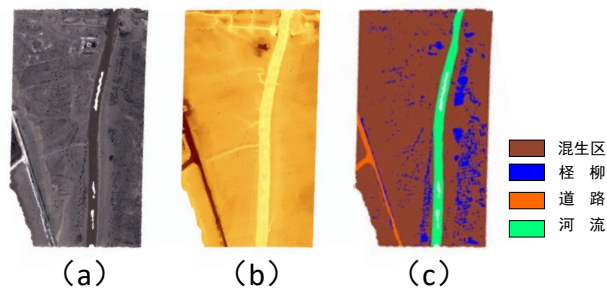
同时为了验证方法的泛化性，

Trento 数据集取自意大利南部城市 Trento 的农村地区。Trento 数据集的空间分辨率为 1 m,场景包含六种不同的土地覆盖:苹果树、建筑、地面、木材、葡萄园和道路。

LiDAR DSM 数据由 Optech ALTM 3100EA 传感器采集，HSI 数据由 AISA Eagle 传感器采集，两个传感器的空间分辨率均为 1m，空间尺寸均为 600×166 像素。HSI 具有 402.89 ~ 989.09 nm 范围内的 63 个光谱通道，其中光谱分辨率为 9.2 nm。

4.2.3 数据预处理

对于黄河口柽柳多模态遥感数据，首先进行图像预处理、随后执行图像标注，最后完成数据集的构建。图像预处理主要利用 ENVI 软件对两个模态数据进行一系列操作，包括图像校正和空间对齐等步骤。接着，使用 ArcGIS 软件对黄河口滨海湿地遥感影像进行矢量化标注，并将这些矢量图转换成栅格图以生成真值图。数据集制作指从遥感图像中选取柽柳等滨海湿地植被密集分布的片段，用作训练数据和测试数据。



4.2.4 对比方法

为了评估所提出模型的性能，本文将 GSJDD-Net 与其他经典的多模态融合方法 TB-CNN、HRWN、EndNet、MDL-RS、MFT、GLT-Net 进行了比较。

4.3 几何-光谱联合动态形变网络（Geometric-Spectral Joint Dynamic Deformation Network, GSJDD-Net）

GSJDD-Net 旨在充分挖掘光谱信息与地形几何特征的互补性，以提升滨海湿地分类的精度与鲁棒性。该网络由可变形多模态编码器（Deformable Multi-modal Encoder）、几何引导注意力模块（Geometry-guided Spectral Attention）、自适应渐进融合模块（Adaptive Progressive Fusion）和动态路由分类模块（Dynamic Routing Classification）四部分组成。首先，GSJDD-Net 采用可变形卷积与多尺度几何增强单元，对高光谱与 LiDAR 数据进行特异性编码，以捕捉复杂地物的边界信息和局部形态变化。随后，通过几何引导注意力机制对齐跨模态特征，减少传感器成像差异带来的空间错位。进一步地，引入自适应渐进融合策略，结合门控权重网络与图卷积，实现不同模态信息的层次化融合，增强特征表达能力。最终，借助动态专家路由机制，自适应调整分类决策，提高复杂场景下的泛化能力。GSJDD-Net 通过多模态协同建模，确保光谱与几何特征的互补性，能够有效提升滨海湿地分类的准确性和稳健性。

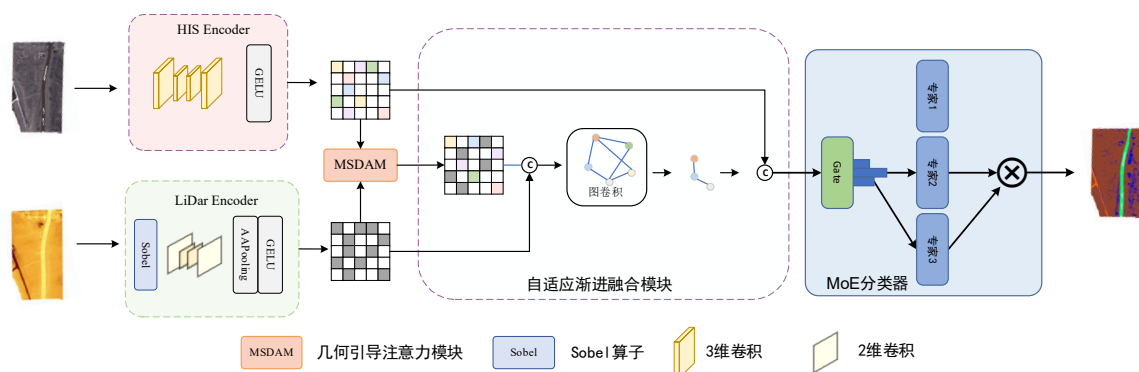


图 4.2

4.3.1 可变形多模态编码器 (Deformable Multi-scale Alignment Module)

光谱信息和地形几何特征在滨海湿地分类任务中具有互补性,但由于光谱特征易受边界模糊的影响,直接使用可能导致特征融合过程中的信息损失。为解决该问题,本模块采用双路径交互机制,结合几何信息引导光谱特征的增强。

本章设计可变形多模态编码器 (Deformable Multi-modal Encoder), 其结构如图 4.5 所示。该模块采用双分支架构分别处理高光谱与激光雷达数据。

在复杂的滨海湿地环境中,植被和地形的形态变化较为剧烈,传统的固定感受野卷积无法充分捕捉这些区域的局部几何特征。在高光谱分支引入了可变形卷积 (Deformable Convolution), 通过学习可变形卷积核的空间偏移量, 使特征提取适应局部地物的形态变化。同时, 考虑到滨海湿地破碎地块的尺度差异, 设计不同膨胀率的并行支路, 以增强模型对不同尺度特征的捕获能力。高光谱分支引入三维可变形卷积核 $\mathcal{K}_d \in \mathbb{R}^{8 \times 3 \times 3}$, 通过动态偏移量预测网络生成空间采样位置:

$$\Delta p_n = \text{Conv}_{3 \times 3}(\text{ReLU}(\text{Conv}_{3 \times 3}(x^{(l)})))$$

其中偏移量预测器由两层 3×3 卷积构成, $x^{(l)}$ 为第 l 层输入特征。该设计使卷积核能自适应调整空间感受野, 有效捕捉破碎化地物的不规则边界特征。

LiDAR 图像拥有丰富的空间信息, 能够为滨海湿地植被分类提供有效的边界信息。在 Lidar 分支由几何增强单元组成, 采用改进的各向异性 Sobel 算子提取高程梯度特征:

$$G = \sqrt{(\mathcal{K}_x \otimes I)^2 + (\mathcal{K}_y \otimes I)^2}$$

式中 \mathcal{K}_x , \mathcal{K}_y 为方向敏感卷积核, 增强线性结构与边缘特征的响应强度。

本模块通过可变形卷积的动态调节能力与几何增强操作, 实现多模态特征的特异性编码, 为后续融合提供具有强表征力的基元特征。

4.3.2 几何引导注意力模块 (Geometry-guided Spectral Attention)

异源数据受到不同传感器成像条件的影响, 尽管可能是同一日期的数据, 但是仍然存在领域差异导致的特征错位问题, 本章提出几何引导多尺度对齐模块 (Geometry-guided Multi-scale Alignment), 其工作流程如图 4.6 所示。构建多尺度特征金字塔:

$$\mathcal{F}_k = \text{DeformConv}(\mathcal{F}_{k-1}; d_k = 2^{k-1})$$

其中 d_k 为第 k 级的扩张率, 形成覆盖局部细节到全局语义的多级特征表达。通过几何相似性度量实现跨模态特征对齐:

$$\mathcal{S}(i, j) = \frac{\phi(\mathcal{F}_i^{HSI}) \cdot \phi(\mathcal{F}_j^{LiDAR})}{\|\phi(\mathcal{F}_i^{HSI})\|_2 \|\phi(\mathcal{F}_j^{LiDAR})\|_2}$$

式中 $\phi(\cdot)$ 为通道压缩函数。设置相似度阈值 $\tau = 0.7$ ，当 $\mathcal{S}(i, j) > \tau$ 时建立跨模态特征关联，生成几何约束下的对齐特征图。

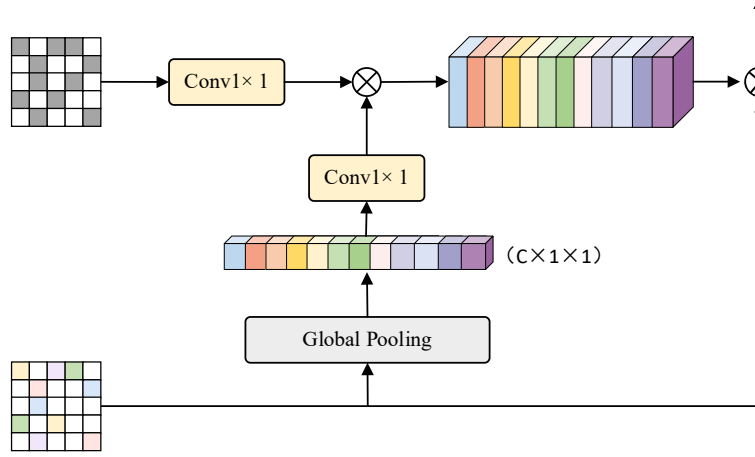


图 4.4 几何引导注意力模块

该模块通过多级可变形卷积与几何相似性约束，有效解决因传感器成像差异导致的空间错位问题，为后续融合提供几何一致的特征表达。

4.3.3 自适应渐进融合模块 (Adaptive Progressive Fusion)

为充分融合高光谱与 LiDAR 数据的互补信息，本研究设计了一种层次化的自适应融合机制 (Hierarchical Adaptive Fusion)，其整体结构如图 4.7 所示。该机制采用递进式融合策略，分为初级融合和高级融合两个阶段，以确保多模态特征的高效交互与信息保留。

在初级阶段，通过通道拼接 (Channel Concatenation) 整合高光谱与 LiDAR 特征，并采用 1×1 卷积进行特征压缩，以减少维度冗余并增强特征表达能力。其计算方式如：

$$\mathcal{F}_{fuse}^{(1)} = \mathcal{W}_c \left(\left[\mathcal{F}^{HSI} \oplus \mathcal{F}^{LiDAR} \right] \right)$$

其中 \mathcal{W}_c 为 1×1 卷积实现的通道压缩层， \oplus 表示通道拼接操作，将 64 维双模态特征压缩至 32 维。

在高级阶段，引入门控权重网络 (Gated Weight Network) 和图卷积 (Graph Convolutional Network, GCN) 进行特征交互，以提升多模态特征的融合适应性。首先，门控权重网络根据输入特征自适应生成融合比例：

$$g = \sigma \left(\mathcal{W}_g * \left[\mathcal{F}^{HSI}, \mathcal{F}^{LiDAR} \right] \right)$$

式中 $\sigma(\cdot)$ 为 sigmoid 激活函数， \mathcal{W}_g 为权重映射矩阵，门控机制能够动态调整各模态

特征的贡献比例，避免信息冗余或模态间不均衡问题。为了建模局部到全局的特征关联，利用基于坐标距离构建的邻接矩阵 $\mathcal{A} \in \mathbb{R}^{N \times N}$ 进行图卷积传播：

$$\mathcal{F}_{gc} = \text{GCN}(\mathcal{F}^{HSI}, \mathcal{A})$$

融合表征可以表示为：

$$\mathcal{F}_{fuse}^{(2)} = g \odot \mathcal{F}^{HSI} + (1 - g) \odot \mathcal{F}^{LiDAR} + \mathcal{F}_{gc}$$

其中 \odot 表示逐像素乘法，该公式确保融合特征在局部几何细节与全局光谱信息之间达到动态平衡，提高模型在复杂滨海湿地场景中的适应能力。实现了从局部到全局的信息整合，相较于传统拼接融合方式，能够有效缓解信息淹没（Information Overwhelming）问题，提高多模态特征的表征能力。

4.3.4 动态路由分类模块（Dynamic Routing Classification）

为提高复杂场景下的分类鲁棒性，本章节基于混合专家系统（MoE）设计了一种动态专家决策机制，其整体架构如图 4.8 所示。该系统采用专家网络（Expert Networks）与动态路由（Dynamic Routing）机制相结合的策略，以实现自适应的特征分类。

该机制包含三组具有不同特征学习能力的专家网络，每个专家网络 $E_k(\cdot)$ 通过深度可分离卷积（Depthwise Separable Convolution, DSConv）提取特定模式的特征表示，其计算形式如下：

$$f_{expert}(x) = DWConv(x)$$

其中， W_k 为第 k 个专家网络的参数矩阵， \otimes 表示卷积运算。深度可分离卷积能够降低计算复杂度，同时提升对局部特征的捕捉能力，使各专家网络专注于不同类别或特征模式。

为确保模型能够根据输入数据特性选择最合适的专家网络，引入路由网络（Routing Network）来计算每个专家的选择概率。首先，对输入特征 F 进行全局平均池化（Global Average Pooling, GAP），提取全局信息：

$$p_k = \frac{\exp(W_r^k \otimes \text{GAP}(F))}{\sum_{i=1}^3 \exp(W_r^i \otimes \text{GAP}(F))}$$

其中， W_r^k 为路由网络的权重， p_k 代表第 k 个专家网络被选择的概率。Softmax 归一化确保概率总和为 1，使得决策过程具有可解释性和自适应性。

最终，系统依据各专家网络的权重分配，计算分类决策输出：

$$P = \sum_{k=1}^3 p_k \cdot E_k(F)$$

通过动态路由机制，该系统能够根据输入数据的特性自适应调整分类策略，从而在不同地物模式下选择最优的专家组合。这种方式既保证了模型的计算效率，又增强了对复杂地物类别的判别能力，相较于固定架构的分类网络，在不增加显著计算开销的前提下，提高了分类的精准性与泛化能力。

4.4 实验结果与分析

4.4.1 实验结果

为了验证本章提出的 MID-ProNet 在滨海湿地高光谱图像上的有效性，在实验中采用典型的机器学习方法随机森林（RF）、支持向量机（SVM）和语义分割方法（PSPNet^[38]、SegNet^[40]、U-Net^[38]、MultiResUNet^[41]、WetlandNet^[21]）与 AD-UNet 进行比较。不同方法的预测结果如图 3.7 所示。

第三个实验是在 Indian Pines 在表 V 中，可以观察到，尽管基于 CNN 的方法已经实现了良好的分类准确度，但是由于缺乏建模全局上下文信息导致在大部分的单类别精度上要弱于基于 Transformer 的方法。我们的方法在面对不同生态型的同种地物取得了最好的结果，在免耕玉米、少耕玉米和玉米的当类精度都取得 99% 以上的准确率，在少耕大豆和大豆上都取得了最优的分类结果。此外，我们提出的方法在获得了最高数量的最佳单类精度的基础上，还在三个其他的指标 AO、AA 和 kappa 系数上都获得了最高的精度，进一步验证了我们提出的方法的优越性。然而，我们提出的方法也有一些局限性。例如，对于免耕大豆的精度略低于 SSFTT。这可能是因为我们方法在进行特征解耦时的尺度是固定的，当处理边缘类别时更多的关注到了主体类别，导致了误分现象。Salinas 数据集的分类难度相对较小，从表 VI 中可以明显看出，对于即使是传统的机器学习方法也可以取得一个不错的精度。即使是 Salinas 中的几个小样本类，例如莴苣 romaine 6 wk 和莴苣 romaine 7 wk，2DCNN 方法都表现出了良好的性能。Salinas 数据集含有的较少的噪声，因此我们的方法没有足够的优势。

如表 VII 所示，相比于在前两个数据集的分类表现，morphFormer 的性能获得了明显提升。这可能在于 Whu-Hi-Honghu 数据集提供了更多的可用于训练的数据的同时，数据集复杂度明显提升。morphFormer 倾向于利用更多的数据量来激活 Transformer 层，从而建模训练样本的长距离相关性，影响了其在样本量较小时的分类性能。但是尽管有所提升，morphFormer 方法在处理小样本类别或者混合像元比较严重的类别时，分类效果仍然不太理想。我们的方法相对于 morphFormer 更多的关注对于混合像元的处理，对于噪声比较多像元能够更好的提取出主体相关的类别，同时引入记忆机制减轻 Transformer 层对于数据量要求的依赖，对于较难分类的 11 类互花米草和 12 类苔藓，本文所提出的方法在区分这些土地覆盖类型方面表现出色。

在表 VIII 中,对于2类水泥路 Cement road 和8类桤柳 Tamarix,相较于基于 Transformer 的最佳方法分别提高了 1.19%和 2.27%。总的来说,该方法在所有数据集上的分类性能均优于其他方法。以 OA 为例,在 PU、Salinas、WHU-Hi-Honghu 和 NC16 数据集上, MiFormer 的准确率分别比最佳方法高出 0.34%、0.93%、1.6%和 0.09%。

本章提出的 AD-UNet 方法通过对光谱信息进行重用,并自适应地增强不同地物的敏感通道,强调光谱相似的地物之间的细微差异,使得整体预测结果较好,改善了混淆分类现象。

表 3.4 不同方法的分类精度和效率

Table 3.3 The classification accuracies and efficiencies for different methods							
F1 (%) 方法 类别	SVM	PSPNet	SegNet	U-Net	Multi- ResUNet	WetlandNet	AD-UNet
桤柳	54.30	66.67	62.55	63.87	70.66	93.98	94.55
混生区	60.75	61.97	66.28	69.07	64.29	98.97	98.58
黄河	91.21	94.61	87.22	93.56	94.57	94.08	94.95
道路	70.13	85.61	87.09	85.99	84.05	97.28	98.12
OA (%)	75.21	79.39	82.17	82.63	83.42	83.54	86.97
Kappa	0.714	0.759	0.792	0.797	0.801	0.808	0.849
AA	-	29.59	29.23	24.33	26.29	17.81	25.85

不同方法的分类精度和时间效率如表 3.4 所示,其中加粗数值为所有方法中的最优值。由于滨海湿地地物种类较多,且 PSPNet 没有充分利用地物的光谱信息,因此精度相比于机器学习方法提升不大,总体精度比 RF、SVM 分别高出 4.41%和 4.18%,在盐地碱蓬上的精度仅为 20.77%。与 PSPNet 相比, SegNet 更注重空间局部信息,总体精度提高了 2.78%,在盐地碱蓬、混生区等部分地物上的分类结果也有所改善。MultiResUNet 与 U-Net 的分类结果相比,精度有所提升,但 FLOPs 值也相应提高,表明模型的计算量随之增大。WetlandNet 在互花米草和混生区上取得了所有方法中的最优分类精度, FLOPs 值最小。本章提出的注意力增强方法 AD-UNet 的 FLOPs 值比 WetlandNet、U-Net 高,比 PSPNet、SegNet 和 MultiResUNet 相比较低,但是 AD-UNet 方法在大部分地物上取得了最高的分类精度和 Kappa 系数,适用于精度要求较高的滨海湿地精细分类任务。

4. 4. 2 消融实验

表 3.5 AD-UNet 消融实验的精度和效率

Table 3.4 Ablation evaluation accuracies and efficiencies of AD-UNet						
	Conv	MI	Tokenizer	OA	AA	$\kappa \times 100$
1	×	√	√	98.12	96.45	97.86

2	√	×	√	97.85	97.42	97.55
3	√	√	PE	95.90	91.41	95.32
4	√	√	√	98.89	98.64	98.74

4.4.3 模型复杂性评估

如表 3.6 所示，使用 FLOPs 衡量模型的复杂度，FLOPs 值越小表示模型复杂度越低，相应的计算量和算力要求就越小。AD-UNet 的 FLOPs 值与对比方法相比处于中等水平，且每个 epoch 的训练时间略有增加。但是，如图 3.10 所示，AD-UNet 的收敛速度比其他模型快得多且取得的分类精度最高，这可以弥补每个 epoch 运行时间更长的缺点。

为了验证 AD-UNet 网络模型中的冗余性，对网络模型进行剪枝操作。由于 AD-UNet 中使用 SE Block 与稠密连接结合，导致模型的复杂度较大，本节对其进行剪枝操作，对 AD-UNet 模型进行剪枝的程度及模型的精度、FLOPs 变化如图 3.10 所示。

在 AD-UNet 模型训练中加入 L1 正则化，使得权重稀疏化，按特征输出的绝对值大小来评估其重要性，然后将权重较小的部分裁剪掉^[57]。随着模型剪枝程度的增加，模型的整体精度和 FLOPs 相应下降。剪枝程度为 20% 时，模型的精度下降了 2.63%，FLOPs 下降了 7.1M，此时模型的 FLOPs 接近对比实验中 FLOPs 最优的 WetlandNet，精度高出其 0.48%，这说明在裁剪了部分冗余的连接后，AD-UNet 在保持较低模型复杂度的情况下仍能取得最高的精度，证明了模型具有较好的特征学习能力和鲁棒性。这也表明 AD-UNet 模型存在少量冗余节点，大部分连接参与主要的计算。

4.5 本章小结 (Summary)

本章节详细说明了提出的基于深度语义分割的滨海湿地高光谱遥感影像精细制图方法，该方法主要包括遥感数据预处理、模型训练和模型预测三个步骤。本章节提出注意力增强的语义分割网络 AD-UNet，综合提取低层细节特征与高层抽象特征，自适应增强地物的显著特征表示，缓解了高光谱影像在语义分割方法中提取精度不高的问题。AD-UNet 可以封装为全自动的端到端处理程序，在黄河口滨海湿地的高光谱遥感影像上的实验结果表明，相比于机器学习方法及其他语义分割方法，该方法适合于地物类型丰富且分布复杂的滨海湿地，提高了滨海湿地精细制图的准确性。

5 总结与展望

5 Summary and prospect

5.1 总结 (Summary)

本文为充分利用高光谱遥感图像的特征，分别从语义分割和分类方法对滨海湿地精细制图进行研究。主要分析了滨海湿地遥感影像解译在国内外的研究现状，并详细介绍了研究内容所涉及的理论基础和相关技术。在训练样本不足的情况下，深度学习技术需要充分挖掘高光谱遥感数据中的光谱和空间特征，提取出高层特征指导滨海湿地精细制图任务，从而达到精确的地物分类效果。本文提出了两种滨海湿地精细制图方法用于滨海湿地智能解译，并且都可以封装成端到端的程序，自动化处理遥感图像，输出预测结果。本文的研究内容主要包括：

(1) 注意力增强的滨海湿地精细制图方法

该方法从语义分割网络中常用的编码器-解码器出发，构建了适用于高光谱图像的滨海湿地精细制图网络 AD-UNet。为了充分利用高光谱图像中丰富的光谱特征，引入密集连接的思想来进行特征重用，鼓励特征传播；同时加入注意力机制，自适应的对通道维的特征图进行加权，保留地物的显著光谱特征。在编码器-解码器之间使用跳跃连接进行特征互补，实现对高光谱图像的语义分割任务。通过在黄河口滨海湿地的 Zhuhai-1 高光谱图像上进行实验，证明了提出方法的有效性。

(2) 微场景嵌入的滨海湿地精细制图方法

该方法基于场景信息，提出微场景嵌入的滨海湿地精细制图方法 TSE-Net。该方法将地物目标周围的几十个像素描述为一个微场景，基于地物的微场景表示对通道进行加权，提高不同地物像元的光谱分辨力；选取 DenseNet 作为基线网络，在其中加入注意力机制来自适应提取显著且鲁棒的特征。通过学习地物之间的光谱依赖性，可以有效区分不同种类的地物，实现滨海湿地的精细分类。实验表明了基于微场景嵌入的方法在小样本条件下仍取得精确的分类结果，证明了模块之间的互补性。

5.2 展望 (Prospect)

滨海湿地的精细分类仍然是一项具有挑战性的工作。尽管本文提出的自适应特征增强的语义分割和分类方法都有较好的表现，但是其应用仍然存在一定的局限性。对于滨海湿地植被的精细分类，本文存在的不足和需进一步研究之处包含：

(1) 针对整体分类效果，由于滨海湿地地理环境复杂，部分区域不同植被共生，从根本上造成植被光谱混淆现象。下一步考虑结合多源多时相的遥感影像数据，充分提取滨海湿地植被的固有特征，将其融合到深度学习模型中，辅助滨海湿地地物分类，提升深度学习技术的可解释能力，完成长时间序列的滨海湿地动态监测。

(2) 针对单种植被的识别效果, 芦苇、碱蓬等植被在土壤盐度含量不同的区域具有不同的分布密度和生态学特征, 如碱蓬在含盐量较低的区域分布密集、植株高大、多呈绿色, 在含盐量较高的区域分布稀疏、植株矮小、多呈红色。受影像分辨率的限制, 长势稀疏的植被信号弱, 难以被识别, 在自动分类和解译时容易遗漏。下一步考虑构建基于现场观测和无人机遥感的植被多层级样本库, 归纳、提取植被的多种生态学特征, 对滨海湿地展开更有针对性的研究。

参考文献

- [1] E. Adam, O. Mutanga, D. Rugege. Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: A review[J]. *Wetlands Ecol Manage*, 2010, 18(3): 281–296.
- [2] 解雪峰, 项琦, 吴涛, 蒋国俊, 孙晓敏, 朱明, 濮励杰. 滨海湿地生态系统土壤微生物及其影响因素研究综述[J]. *生态学报*, 2021, 41(1): 1-12.
- [3] 廖丽蓉, 武明月, 戴子熠, 梁嘉慧, 左平. 基于文献计量的滨海湿地碳收支研究进展分析[J]. *海洋环境科学*, 2022, 41(1): 32-39+58.
- [4] 张华兵, 刘红玉, 李玉凤, 安静, 薛星宇, 侯明行. 自然条件下盐城海滨湿地土壤水分/盐度空间分异及其与植被关系研究[J]. *环境科学*, 2013, 34(2): 540-546.
- [5] 韩广轩, 王光美, 毕晓丽, 王传远等. 黄河三角洲滨海湿地演变机制与生态修复[M]. 北京: 科学出版社, 2018.
- [6] Yan C, Qing-Wen Q I, Gui-Shan Y. Main characteristics, progress and prospect of international wetland science research[J]. *Progress in Geography*, 2011, 21(2): 111-120.
- [7] 徐启薇. 建设项目用海面积控制指标体系研究[D]. 浙江大学, 2019.
- [8] 王建步, 张杰, 马毅, 任广波. 基于GF-1 WFV影像的黄河口湿地植被盖度估测及分析[J]. *海洋学报*, 2018, 40(6): 40-50.
- [9] 张永辉. 基于语义分割的滨海湿地遥感图像信息提取[D]. 山东科技大学, 2020.
- [10] 张兵. 高光谱图像处理与信息提取前沿[J]. *遥感学报*, 2016, 20(5): 1062-1090.
- [11] 李先怡, 范海生, 潘申林, 蒋晓华, 吴佳奇. Zhuhai-1高光谱卫星数据及应用概况[J]. *卫星应用*, 2019(8): 12-18.
- [12] 张顺, 龚怡宏, 王进军. 深度卷积神经网络的发展及其在计算机视觉领域的应用[J]. *计算机学报*, 2019, 42(3): 453-482.
- [13] 刘环宇. 基于图像深度学习的地面目标识别应用系统研制[D]. 哈尔滨工业大学, 2017.
- [14] 刘润红, 梁士楚, 赵红艳, 漆光超, 李丽香, 姜勇, 牛振国. 中国滨海湿地遥感研究进展[J]. *遥感技术与应用*, 2017, 32(6): 998-1011.
- [15] 李哲, 宫兆宁, 刘先林, 关晖, 王颖. 基于面向对象多端元混解模型的植被覆盖度反演及其时空分布研究[J]. *遥感技术与应用*, 2018, 33(6): 1149-1158.
- [16] Wang X, Gao X, Zhang Y, Fei X, Zhao H. Land-cover Classification of Coastal Wetlands using the RF Algorithm for Worldview-2 and Landsat 8 Images[J]. *Remote Sensing*, 2019, 11(16): 1927-1947.
- [17] Yang J, Ren G, Ma Y, Fan Y. Coastal Wetland Classification Based on High Resolution SAR and Optical Image Fusion[C]. 2016 IEEE International Geoscience and Remote Sensing

Symposium (IGARSS), 2016:886-889.

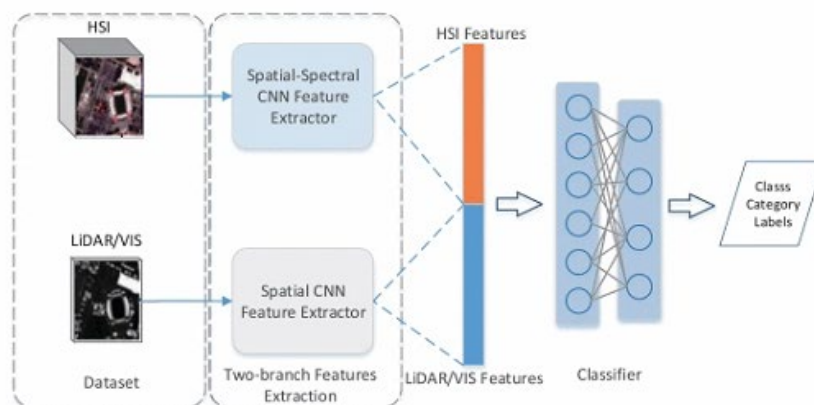
- [18] 张磊, 宫兆宁, 王启为, 金点点, 汪星. Sentinel-2影像多特征优选的黄河三角洲湿地信息提取[J]. 遥感学报, 2019, 23(2): 313-326.
- [19] Hu Y, Zhang J, Ma Y, Li X, Sun Q, An J. Deep Learning Classification of Coastal Wetland Hyperspectral Image Combined Spectra and Texture Features: A case study of Huanghe (Yellow) River Estuary wetland[J]. *Acta Oceanologica Sinica*, 2019, 38(5): 142–150.
- [20] 朱玉玲, 王建步, 王安东, 王锦锦, 赵晓龙, 任广波, 胡亚斌, 陈晓英, 马毅. 融合浅层特征的深度卷积神经网络互花米草遥感监测方法[J]. 海洋科学, 2019, 43(7): 12-22.
- [21] Cui B, Zhang Y, Li X. WetlandNet: Semantic Segmentation for Remote Sensing Images of Coastal Wetlands via Improved U-Net with Deconvolution[C]. *International Conference on Genetic and Evolutionary Computing*. Springer, Singapore, 2019: 281-292.
- [22] Feng Q, Yang J, Zhu D, Liu J, Guo H. Integrating Multitemporal Sentinel-1/2 Data for Coastal Land Cover Classification Using a Multibranch Convolutional Neural Network: A Case of the Yellow River Delta[J]. *Remote Sensing*, 2019, 11(5): 1006-1028.
- [23] Dang K, Nguyen M, Nguyen D, Phan T, Giang T, Pham H, Nguyen T, Tran T, Bui D. Coastal Wetland Classification with Deep U-Net Convolutional Networks and Sentinel-2 Imagery: A Case Study at the Tien Yen Estuary of Vietnam[J]. *Remote Sensing*, 2020, 12(19): 3270-3296.
- [24] Liu C, Tao R, Li W, et al. Joint Classification of Hyperspectral and Multispectral Images for Mapping Coastal Wetlands[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2020, 14: 982-996.
- [25] Gao Y, Gao F, Dong J, Du Q, Li H. Synthetic Aperture Radar Image Change Detection via Siamese Adaptive Fusion Network[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021, 14: 10748-10760.
- [26] 凌晨, 张鑫彤, 马雷. 基于Mask R-CNN算法的遥感图像处理技术及其应用[J]. *计算机科学*, 2020, 47(10): 151-160.
- [27] Liu H, Peng C, Yu C, Wang J, Liu X, Yu G, Jiang W. An End-to-End Network for Panoptic Segmentation[J]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, 6172-6181.
- [28] Durand, Thibaut, Mehra, Nazanin, Mori, Greg. Learning a Deep ConvNet for Multi-label Classification with Partial Labels[J]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, 647-657.
- [29] Belgiu M, Stein A. Spatiotemporal Image Fusion in Remote Sensing[J]. *Remote Sensing*, 2019, 11(7): 818-838.

- [30] Li P, Chen X, Shen S. Stereo R-CNN based 3D Object Detection for Autonomous Driving[J]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, 7644-7652.
- [31] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [32] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[J]. Proceedings of the 32nd International Conference on Machine Learning, 2015: 448-456.
- [33] 康旭琴, 吴偶, 王磊,等. 面向疾病相关关联抽取的深度语义特征研究[J]. 计算机工程与应用, 2018, 54(8): 260-265.
- [34] Roy S K, Krishna G, Dubey S R, et al. HybridSN: Exploring 3D-2D CNN Feature Hierarchy for Hyperspectral Image Classification[J]. IEEE Geoscience and Remote Sensing Letters, 2020, 17(2):277-281.
- [35] Criminisi A, Shotton J. Classification Forests for Semantic Segmentation of Brain Lesions in Multi-channel MRI[J]. Springer London, 2013, 10.1007/978-1-4471-4929-3(Chapter 17):245-260.
- [36] Arbeláez P, Hariharan B, Gu C, et al. Semantic segmentation using regions and parts[C]. 2012 IEEE conference on computer vision and pattern recognition, 2012: 3378-3385.
- [37] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2015: 3431-3440.
- [38] Zhao H, Shi J, Qi X, et al. Pyramid Scene Parsing Network[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2017: 6230-6239.
- [39] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[C]. International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015: 234-241.
- [40] Badrinarayanan V, Kendall A, Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(12): 2481-2495.
- [41] Ibtehaz N, Rahman M S. MultiResUNet: Rethinking the U-Net Architecture for Multimodal Biomedical Image Segmentation[J]. Neural Networks, 2020: 74-87.
- [42] 许凤晖, 慕晓冬, 赵鹏, 等. 利用多尺度特征与深度网络对遥感影像进行场景分类[J]. 测绘学报, 2016, 45(7): 834-840.
- [43] Zhang X, Zou J, He K, et al. Accelerating Very Deep Convolutional Networks for

- Classification and Detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016, 38(10): 1943-1955.
- [44] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[J]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2016, 770-778.
- [45] Huang G, Liu Z, Laurens V, et al. Densely Connected Convolutional Networks[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017: 2261-2269.
- [46] Cui B, Zhang H, Jing W, Liu H, Cui J. SRSe-Net: Super-Resolution-Based Semantic Segmentation Network for Green Tide Extraction[J]. Remote Sensing, 2022, 14(3): 710-730.
- [47] Jie H, Li S, Gang S, et al. Squeeze-and-Excitation Networks[J]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [48] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]. Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.
- [49] 杨伟, 陈沈良. 黄河三角洲河口区滨海湿地面积动态变化与影响因素分析[J]. 海洋科学, 2011, 35(7): 61-66.
- [50] 杨远陶, 刘瑞, 曹礼刚, 杨梅, 陈景珏. 基于随机森林算法的Zhuhai-1高光谱影像土地利用信息提取[J]. 物探化探计算技术, 2021, 43(06): 818-824.
- [51] 王建步, 张杰, 马毅, 任广波. 基于GF-1 WFV的黄河口湿地植被碳储量估算研究[J]. 海洋科学进展, 2019, 37(1): 75-83.
- [52] 田晟, 张裕天, 张剑锋, 许凯. 基于深度神经网络的交通出行方式选择模型[J]. 北华大学学报(自然科学版), 2019, 20(1): 109-113.
- [53] 张宇倩, 李国辉, 雷军, 何嘉宇. FF-CAM: 基于通道注意机制前后端融合的人群计数[J]. 计算机学报, 2021, 44(2): 304-317.
- [54] Ketkar N. Introduction to keras[M]. Deep learning with Python. Apress, Berkeley, California, 2017: 97-111.
- [55] Abadi M, Barham P, Chen J, et al. Tensorflow: A system for large-scale machine learning[C]. 12th Symposium on Operating Systems Design and Implementation, 2016: 265-283.
- [56] Hu Y, Zhang J, Ma Y, et al. Hyperspectral Coastal Wetland Classification Based on a Multiobject Convolutional Neural Network Model and Decision Fusion[J]. IEEE Geoscience and Remote Sensing Letters, 2019: 1-5.
- [57] Zhuang L, Li J, Shen Z, et al. Learning Efficient Convolutional Networks through Network Slimming[C]. ICCV, 2017: 292-308.
- [58] 张继贤, 顾海燕, 杨懿, 等. 高分辨率遥感影像智能解译研究进展与趋势[J]. 遥感学报, 2021, 25(11): 2198-2210.

- [59] Zheng Z, Zhong Y, Wang J, et al. Foreground-Aware Relation Network for Geospatial Object Segmentation in High Spatial Resolution Remote Sensing Imagery[J]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 4096-4105.
- [60] Zhao J, Hu L, Dong Y, et al. A combination method of stacked autoencoder and 3D deep residual network for hyperspectral image classification[J]. International Journal of Applied Earth Observation and Geoinformation, 2021, 102(4):1-17.
- [61] Erturk A, Gullu M K, Erturk S. Hyperspectral image classification using empirical mode decomposition with spectral gradient enhancement[J]. IEEE Transactions on Geoscience & Remote Sensing, 2013, 51(5):2787-2798.

-
- [62] X. Xu, W. Li, Q. Ran, Q. Du, L. Gao and B. Zhang, "Multisource Remote Sensing Data Classification Based on Convolutional Neural Network," in IEEE Transactions on Geoscience and Remote Sensing, vol. 56, no. 2, pp. 937-949, Feb. 2018, doi: 10.1109/TGRS.2017.2756851.



- [63] Xin Wu, Danfeng Hong, Jocelyn Chanussot. Convolutional Neural Networks for Multimodal Remote Sensing Data Classification, IEEE Transactions on Geoscience and Remote Sensing, 2022, vol. 60, pp. 1-10, DOI: 10.1109/TGRS.2021.3124913.

