



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
遥感精细化制图研究

Fine-Scale Remote Sensing Mapping of Coastal Wetlands via
Feature Decoupling and Multimodal Data Fusion

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

Fine-Scale Remote Sensing Mapping of Coastal Wetlands via Feature
Decoupling and Multimodal Data Fusion

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，滨海湿地植被精细制图对于评估湿地健康状态和制定生态修复策略至关重要。然而，传统遥感分类方法在应对滨海湿地植被精细制图时面临诸多挑战：一方面，滨海湿地植被类内光谱特征变化较大，导致分类精度受限；另一方面，湿地植被斑块破碎化严重，单一数据源难以实现形态与光谱信息的有效协同解析。随着遥感技术的不断进步，高光谱卫星影像和激光雷达数据为解决这些问题提供了新的机遇。高光谱影像能够获取地物几乎连续的光谱曲线，有助于识别光谱易混淆物种；激光雷达则在垂直维度上弥补了光学影像的不足，能够精确量化植被的高程和冠层密度等形态特征。融合这两种数据，在滨海湿地植被精细制图中将展现出更大的潜力。针对上述问题，本研究的主要内容如下：

(1) 本文针对滨海湿地植被类内光谱特征变化较大导致的光谱混淆问题，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。该网络通过构建空谱特征分支与域特征分支的双通道架构，利用原型归纳模块和变分自编码器分别提取植被类别的可区分性表征和与分类无关的干扰特征。同时，设计互信息最小化准则实现双分支特征解耦，并在原型学习过程中增强滨海湿地地物分类特征的可解释性与鲁棒性，从而提高分类精度。

(2) 针对滨海湿地植被破碎化导致单一遥感数据源难以全面表征其形态与光谱特性的问题，本文提出一种高光谱-激光雷达数据自适应融合网络。该网络通过可变形卷积多尺度编码器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。同时，利用几何引导光谱注意力模块，结合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。此外，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，进一步提升分类精度。

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

Coastal wetlands, as ecological transition zones at the land-sea interface, play an irreplaceable role in resisting coastal erosion, maintaining biodiversity, and regulating the global carbon cycle. Fine-scale vegetation mapping of coastal wetlands is crucial for assessing wetland health and formulating ecological restoration strategies. However, traditional remote sensing classification methods face numerous challenges in fine-scale vegetation mapping of coastal wetlands. On one hand, the large spectral variability within vegetation classes limits classification accuracy. On the other hand, the fragmentation of wetland vegetation patches makes it difficult for a single data source to effectively integrate morphological and spectral information. With the continuous advancement of remote sensing technology, hyperspectral satellite imagery and LiDAR data offer new opportunities to address these issues. Hyperspectral imagery provides nearly continuous spectral curves of objects, helping to distinguish spectrally confusable species, while LiDAR compensates for the shortcomings of optical imagery in the vertical dimension, accurately quantifying vegetation elevation, canopy density, and other morphological features. The fusion of these two data types shows great potential for fine-scale vegetation mapping of coastal wetlands. This study primarily focuses on the following:

(1) To address the issue of spectral confusion caused by large spectral variability within coastal wetland vegetation classes, this paper proposes a dual-branch deep learning network based on feature disentanglement and prototype learning. The network constructs a dual-channel architecture with hyperspectral feature and domain feature branches, using a prototype inductive module and a variational autoencoder to extract discriminative representations for vegetation categories and interference features unrelated to classification. At the same time, mutual information minimization criteria are designed to achieve feature disentanglement between the two branches, enhancing the interpretability and robustness of coastal wetland vegetation classification features during prototype learning, thereby improving classification accuracy.

(2) To address the issue of vegetation fragmentation in coastal wetlands, which makes it difficult for a single remote sensing data source to comprehensively represent both morphological and spectral characteristics, this paper proposes a hyperspectral-LiDAR data adaptive fusion network. The network dynamically adjusts the convolution kernel offset using a deformable convolution multi-scale encoder to adaptively capture the patchy distribution and irregular boundary features of wetland vegetation. Simultaneously, a geometry-guided spectral attention module is used, combining LiDAR elevation information to generate spatial weight maps that dynamically calibrate the response intensity of hyperspectral feature channels, enhancing spectral resolution in areas with significant canopy height differences. Additionally, an innovative

progressive fusion mechanism is introduced, utilizing a gated weight network to achieve hierarchical feature interaction from local details to global semantics, avoiding the information dilution problem in early fusion. Finally, a dynamic routing classification head is employed, which automatically selects the optimal expert network combination for decision-making based on the complexity of the input features, further improving classification accuracy.

Keywords: Coastal Wetlands; Hyperspectral Remote Sensing; Feature Disentanglement; Multi-Modal Fusion; LiDAR

目 录

图清单	1
表清单	3
变量注释表	4
1 绪 论	1
1.1 研究背景与意义	1
1.2 国内外研究现状	2
1.3 问题的提出	4
1.4 主要研究内容	4
1.5 组织结构及章节安排	5
2 相关理论与技术	6
2.1 深度学习基础	6
2.2 表征学习	9
2.3 图卷积神经网络	10
2.4 混合专家	11
2.4 本章小结	12
3 基于特征解耦与原型学习的滨海湿地精细制图方法	13
3.1 引言	13
3.2 互信息解耦原型对齐网络	13
3.3 实验准备	20
3.4 实验与分析	26
3.5 本章小结	32
4. 基于多模态数据融合的滨海湿地植被精细制图方法	33
4.1 引言（Foreword）	33
4.2 滨海湿地多模态数据植被精细制图主要流程	33
4.3 几何-空谱联合动态形变网络	35
4.4 实验及分析结果	39
4.5 本章小结	45
5 总结与展望	46
5.1 总结	46
5.2 展望	46
参考文献	

Content

List of Figures	1
List of Tables	3
List of Variables	4
1 Introduction	1
1.1 Research background and significance	1
1.2 Research status at home and abroad	2
1.3 Statement of problem	4
1.4 Main research content	4
1.5 Structure and chapter arrangement	5
2 Related theories and techniques	6
2.1 Fundamentals of Deep Learning	6
2.2 Representation Learning	9
2.3 Graph Convolutional Network	10
2.4 Mixture of Experts	11
2.4 Summary	12
3 Feature Decoupling and Prototype Learning-Based Fine-Grained Mapping Method for Coastal Wetlands	13
3.1 Foreword	13
3.2 Mutual Information Decoupling Prototype Alignment Network	13
3.3 Experimental Preparation	20
3.4 Experiments and analysis	26
3.5 Summary	32
4 Multimodal Data Fusion-Based Fine-Grained Vegetation Mapping Method for Coastal Wetlands	33
4.1 Foreword	33
4.2 Main Workflow of Multimodal Data-Based Fine-Grained Vegetation Mapping in Coastal Wetlands	33
4.3 Geometric-Spectral Joint Dynamic Deformation Network	35
4.4 Experimental Results and Analysis	39
4.5 Summary	45
5 Summary and prospects	46
5.1 Summary	46
5.2 Prospect	46
References	

图清单

图序号	图名称	页码
图 1.1	多层次遥感观测平台与传感器分类示意图	1
Fig.1.1	Multi-tier remote sensing platforms and sensor taxonomy	1
图 2.1	卷积神经网络示意图	6
Fig.2.1	Schematic diagram of a convolutional neural network	6
图 2.2	卷积操作示意图	7
Fig.2.2	Schematic diagram of convolution operation	7
图 2.3	池化示意图	7
Fig.2.3	Schematic diagram of pooling	7
图 2.4	常见激活函数示意图	8
Fig.2.4	Common Activation Functions	8
图 2.5	图神经网络计算过程	11
Fig.2.5	Computation Process of Graph Neural Networks	11
图 2.6	混合专家示意图	11
Fig.2.6	Mixture of Experts (MoE) Schematic	11
图 3.1	MID-ProNet 网络结构图	14
Fig.3.1	Architecture of MID-ProNet	14
图 3.2	MID-ProNet 网络训练推理流程图	15
Fig.3.2	Training and Inference Flowchart of MID-ProNet Network	15
图 3.3	地物固有特征提取器模块结构示意图	16
Fig.3.3	Architecture of the Interactive Feature Extraction Module	16
图 3.4	深度可分离卷积结构示意图	17
Fig.3.4	Architecture of Depthwise Separable Convolution	17
图 3.5	多级语义编码器结构	18
Fig.3.5	Architecture of Multi-Level Semantic Encoder	18
图 3.6	自注意力机制编码器结构示意图	19
Fig.3.6	Architecture of Self-Attention Encoder	19
图 3.7	Coast 数据集 (a)假彩色图像 (b)地面真值图	21
Fig.3.7	Coast dataset (a) False color composite (b) Ground truth	21
图 3.8	YRD-NC12 数据集 (a)假彩色图像(b)地面真值图	22
Fig.3.8	YRD-NC12 dataset (a) False color composite (b) Ground truth	22
图 3.9	Indian Pines 数据集 (a)假彩色图像 (b)地面真值图	23
Fig.3.9	Indian Pines dataset (a) False color composite (b) Ground truth	23
图 3.10	Coast 数据集上的分类可视化图	26
Fig.3.10	Classification visualization map on the Coast dataset	26
图 3.11	YRD-NC12 数据集上的分类可视化图	27
Fig.3.11	Classification visualization map on the YRD-NC12 dataset	27
图 3.12	Indian Pines 数据集上的分类可视化图	28
Fig.3.12	Classification maps for the Indian Pines dataset	28
图 3.13	混淆矩阵可视化结果 (a)未进行特征解耦(b)进行特征解耦	30

Fig.3.13	Visualization results of the confusion matrix(a) Without feature decoupling(b) With feature decoupling	30
图 3.14	不同空间输入大小的对分类性能的影响	31
Fig.3.14	Impact of different spatial input sizes on classification performance	31
图 4.1	滨海湿地多模态数据植被精细制图主要流程图	34
Fig.4.1	Main flowchart of fine vegetation mapping using multimodal data in coastal wetlands	34
图 4.2	GSJDD-Net 网络结构示意图	36
Fig.4.2	Schematic diagram of GSJDD-Net network structure	36
图 4.3	几何引导注意力模块结构示意图	37
Fig.4.3	Structure diagram of the geometry-guided attention module	37
图 4.4	Chengliu 数据集 (a) 高光谱假彩色图像 (b) Lidar 图像 (c)地面真值图	39
Fig.4.4	Chengliu dataset (a) Hyperspectral false-color image (b) LiDAR image (c) Ground truth map	39
图 4.5	Trento 数据集 (a) 高光谱假彩色图像 (b) Lidar 图像 (c)地面真值图	40
Fig.4.5	Trento dataset (a) Hyperspectral false-color image (b) LiDAR image (c) Ground truth	40
图 4.6	Chengliu 数据集上的分类可视化图	42
Fig.4.6	Classification visualization map on the Chengliu dataset	42
图 4.7	Trento 数据集上的分类可视化图	44
Fig.4.7	Classification visualization map on the Trento dataset	44
图 4.8	混淆矩阵可视化结果 (a)未使用几何引导注意力(b)使用几何引导注意力	45
Fig.4.8	Confusion matrix visualization results (a) Without geometry-guided attention (b) With geometry-guided attention	45

表清单

表序号	表名称	页码
表 3.1	实验环境详细信息	21
Table 3.1	The experimental environment details	21
表 3.2	Coast 数据集上每个类别样本的数量	22
Table 3.2	Number of samples per category in the Coast dataset	22
表 3.3	YRD-NC12 数据集上每个类别的样本数量	22
Table 3.3	Number of samples per category in YRD-NC12 datasets	22
表 3.4	Indian Pines 数据集上每个类别的样本数量	23
Table 3.4	Number of samples per category in the Indian Pines dataset	23
表 3.5	混淆矩阵	24
Table 3.5	The confusion matrix	24
表 3.6	Coast 数据集上的分类结果	26
Table 3.6	Classification results on the Coast dataset	26
表 3.7	YRD-NC12 数据集上的分类结果	28
Table 3.7	Classification result on the YRD-NC12 dataset	28
表 3.8	Indian Pines 数据集上的分类结果	29
Table 3.8	Classification result on the Indian Pines dataset	29
表 3.9	MID-ProNet 消融实验的精度	30
Table 3.9	Ablation evaluation accuracies of MID-ProNet	30
表 4.1	Chengliu 数据集上每个类别的数量	40
Table 4.1	Number of samples per category in the Chengliu dataset	40
表 4.2	Trento 数据集上每个类别的数量	40
Table 4.2	Number of samples per category in the Trento dataset	40
表 4.3	Chengliu 数据集上的分类结果	42
Table 4.3	Classification results on the Chengliu dataset	42
表 4.4	Trento 数据集上的分类结果	43
Table 4.4	Classification results on the Trento dataset	43
表 4.5	GSJDD-Net 消融实验的精度	44
Table 4.5	Ablation evaluation accuracies of GSJDD-Net	11

变量注释表

变量	注释	初现页
Q	查询向量	8
K	键向量	8
V	值向量	8
w_Q, w_K, w_V	可学习的参数矩阵	8
$H^{(l)}$	1 层的节点特征矩阵	10
\tilde{A}	添加自连接的邻接矩阵	10
\tilde{D}	\tilde{A} 度矩阵	10
$W^{(l)}$	GCN 中可训练参数	10
\mathcal{D}	高光谱遥感影像数据集	13
X	原始高光谱遥感影像	13
Y	类别标签	13
$\mathcal{D}_{\text{train}}$	训练集	13
$\mathcal{D}_{\text{test}}$	测试集	13
$\mathcal{D}'_{\text{train}}$	经过图像变换后的训练集	13
ϵ	高斯噪声	14
T	变换函数	14
X_{patch}	高光谱图像块	15
F_{multi}	多尺度	16
F_c	压缩特征	16
p_t	t 时刻原型	17
f_i	当前批次样本特征	17
α	动量系数	17
p^{mem}	历史语义信息	17
p^{cur}	初级类别语义	17
X_{pro}^c	类别 c 的语义表征	17
μ_c	类别 c 的真实类别中心	17
λ	噪声重建模块平衡因子	18
$(X_{\text{inv}}, X_{\text{cinv}})$	训练批次中的采样联合分布对	19
$(X'_{\text{inv}}, X'_{\text{cinv}})$	训练批次中的边缘分布对	19
X_{gpatch}	激光雷达图像块	34
F_s^0	经过 DME 得到空谱特征	34
F_g^0	经过 DME 得到集合空间特征	34
F_s^1	经过 GSA 得到的空谱特征	36
$\mathcal{F}_{\text{fuse}}^1$	APF 模块得到的初级融合特征	37
$\mathcal{F}_{\text{fuse}}^2$	APF 模块得到的高级融合特征	37
f_{expert}	专家模块输出	38
P	模型输出预测结果	38

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^{[3][4]}。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。

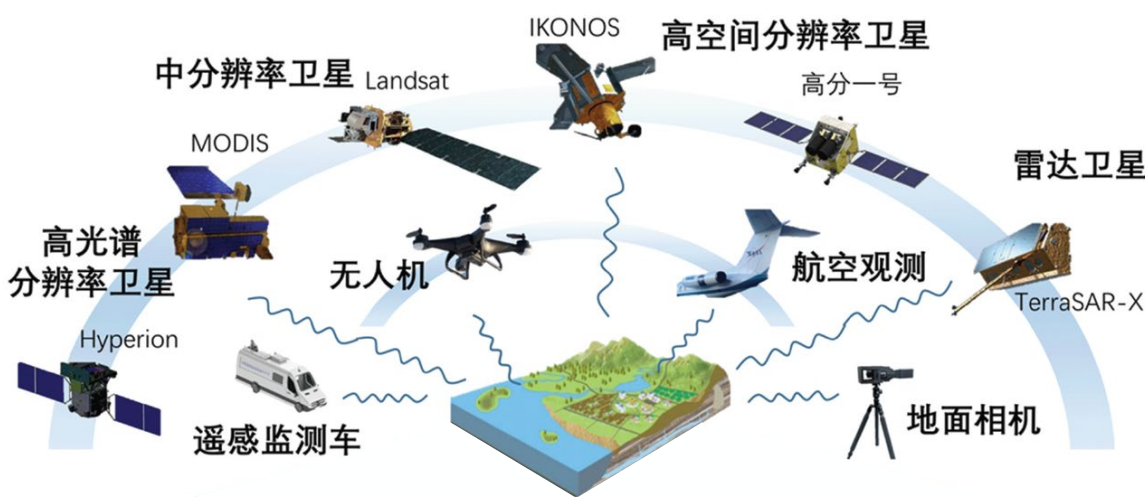


图 1.1 多层次遥感观测平台与传感器分类示意图^[9]
Fig.1.1 Multi-tier remote sensing platforms and sensor taxonomy

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^[10-16]。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,大致可以分为基于像素的滨海湿地制图方法、面向对象的滨海湿地制图方法和基于深度学习的滨海湿地制图方法:

1.2.1 基于像素的滨海湿地制图方法 (Pixel-based coastal wetland mapping method)

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^{[17][20]},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^[21]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^[22]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^[19]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^[20]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^{[23][24]},缺乏对地物空间纹理特征及邻域上下文关系的建模能力^[25],制约了复杂湿地景观的分类精度。

1.2.2 面向对象的滨海湿地制图方法 (Object-oriented coastal wetland mapping method)

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^{[17][19]}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^[17]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^[19]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^[17],Fu 等^[26]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 基于深度学习的滨海湿地制图方法 (Deep learning-based coastal wetland mapping method)

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^[27]采用模糊隶属度规则的多目标卷积神经网络决策融合策略较为有效的解决了植被类型混生的问题;朱玉玲等^[28]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^[29]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^[30]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^[31]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^[32]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^[9]。

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)

滨海湿地作为海陆交互的复杂生态系统, 其植被分类对生态保护与修复至关重要。然而, 由于光谱特征相似性高、植被斑块破碎化等挑战, 传统分类方法难以准确获取植被信息。本文分别基于单一高光谱数据和高光谱-激光雷达融合数据角度, 提出两种深度学习方法, 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

针对不同植被群落光谱特征相似的问题, 本文设计融合特征解耦与原型学习的双分支深度网络 (MID-ProNet)。通过空谱特征分支的原型归纳模块, 显式建模植被类别可分性表征; 域特征分支采用变分自编码器重构输入影像, 提取干扰特征。引入互信息最小化准则实现特征解耦, 并通过动量更新机制优化类别原型, 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。针对单一高光谱数据难以全面表征植被形态和光谱特征的局限，本文引入激光雷达数据，提出几何-光谱联合动态形变网络（GSJDD-Net）。设计多尺度动态对齐模块，利用可变形卷积和自适应池化减少模态差异；构建几何引导光谱注意力模块，借助 LiDAR 几何信息优化特征学习；并通过自适应渐进式融合机制和多级门控融合机制，实现多尺度特征的动态集成，最终提升滨海湿地植被分类的精度。

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源遥感数据，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（Fundamentals of Deep Learning）

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^[34]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^[35]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。

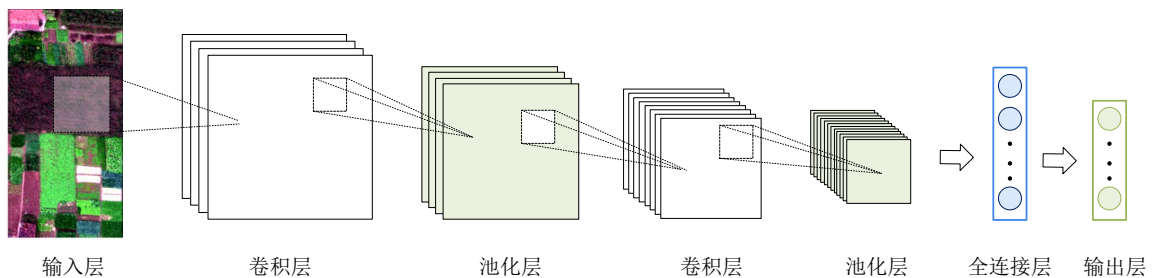


图 2.1 卷积神经网络示意图

Fig. 2.1 Schematic diagram of a 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。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。

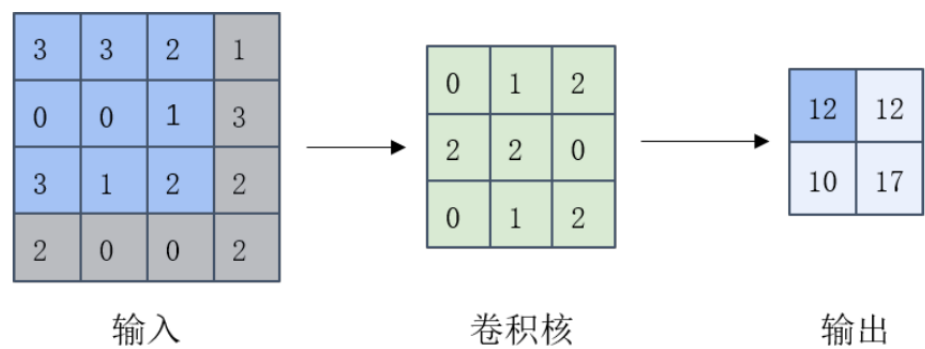


图 2.2 卷积示意图

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。

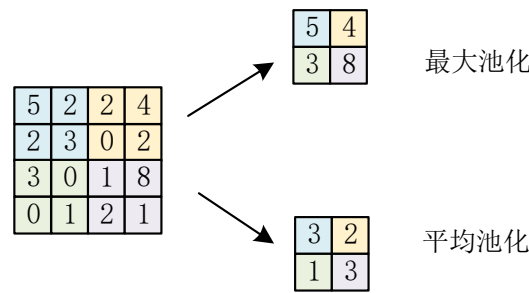


图 2.3 池化示意图

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。

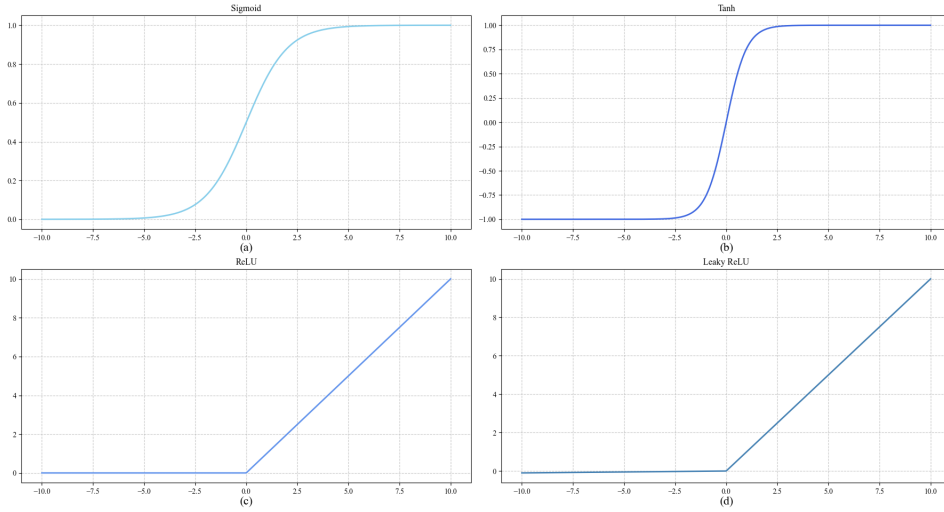


图 2.4 常见的激活函数

Fig. 2.4 Common Activation Functions

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^[36]。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。**Bahdanau** 等人^[37]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^[38]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，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**^[39]、**GPT**^[40]）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^[41]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。空间注意力则聚焦于特征图的空间维度相关性，**wang** 等人^[42]提出了。**Woo** 等人^[43]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^[44]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。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、键和价值矩阵：

$$Q = XW_Q, K = XW_K, V = XW_V \quad (2.1)$$

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：

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2.2)$$

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（Representation Learning）

表征学习通过自动化的特征提取将原始数据映射到低维空间，捕获数据内在结构与语义信息，其核心目标是提升特征可分离性与鲁棒性以增强下游任务性能。深度学习的非线性层级变换推动了该领域进展，实现高阶语义特征的逐层抽象。本节聚焦解耦表征学习与原型学习，二者对模型可解释性及性能提升具有关键作用。

2.2.1 解耦表征学习

解耦表征学习通过分离数据潜在生成因子（如形状、纹理）至独立子空间，构建具有物理可解释性的隐变量表征，其核心目标在于揭示数据底层语义结构以提升模型泛化能力^[45]。当前研究主要采用对抗学习和互信息度量两类方法：前者通过特征解耦实现隐变量分离^{[46][49]}，但存在训练稳定性缺陷^[50]；后者利用互信息（MI）显式量化特征关联性^[51]，有效规避对抗训练的优化难题。

在视觉识别^{[52][54]}和超声图像分类^[55]任务中，MI 正则化通过约束特征独立性抑制冗余信息干扰。在医学图像分割领域中，通过分离图像风格与形状特征降低设备差异影响^[56]显著的提升了分类结果。

基于互信息的解耦表征方法为多模态遥感数据融合提供了新范式。通过约束特征子空间独立性,既可增强模型对数据异质性的适应能力,又能建立可追溯的决策依据,对滨海湿地植被精细分类具有显著工程价值。

2.2.2 原型学习

原型学习源于认知心理学的知识表征理论,其核心思想是通过典型样本构建类别表征^{[57][58]}。与支持向量机、多层感知机等传统机器学习方法相比,该方法凭借其实例驱动的学习机制和显式可解释性特征,在模式识别领域展现出独特优势^{[59][61]}。其基本范式是通过度量待测样本与预设原型的相似度实现分类,其中原型通常定义为训练集中具有类别代表性的特征向量。

该方法的理论基础可追溯至最近邻分类法则^[62],后续发展出学习向量量化 LVQ^[63]、广义 LVQ^[64] 和邻域成分分析^[65]等经典算法。当前研究热点聚焦于深度学习与原型学习的融合范式,在少样本学习^[66]、零样本识别^[67]和无监督聚类^[68]等场景中取得显著成效,同时增强了监督分类^[69]和可解释网络^[70]的模型透明度。

2.3 图卷积神经网络 (Graph Convolutional Network)

图卷积网络 (Graph Convolutional Network, GCN) 是一种将卷积操作从欧几里得空间扩展到非结构化图数据的深度学习模型。由于传统方法难以有效建模图数据的复杂拓扑关系,GCN 通过结合图信号处理与谱图理论,实现了对节点及其邻域信息的高效整合,从而提升了对图数据的表征能力^[71]。其核心思想是利用图的邻接关系进行特征传播,使得每个节点能够聚合其邻域的信息,逐层提取高阶特征。

GCN 通过定义基于谱图理论的图卷积操作来适应图结构数据。其基本更新规则可以表示如下:

$$H^{(l+1)} = \sigma \left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)} \right) \quad (2.3)$$

其中, $H^{(l)}$ 表示 l 层的节点特征矩阵, $\tilde{A} = A + I$ 为添加自连接的邻接矩阵, \tilde{D} 为 \tilde{A} 度矩阵, $W^{(l)}$ 为可训练参数。通过归一化的邻接矩阵对节点特征进行加权平均,使得每个节点的特征不仅包含自身的信息,还融合了其邻域节点的信息,从而实现特征的平滑化和拓扑结构的联合学习。

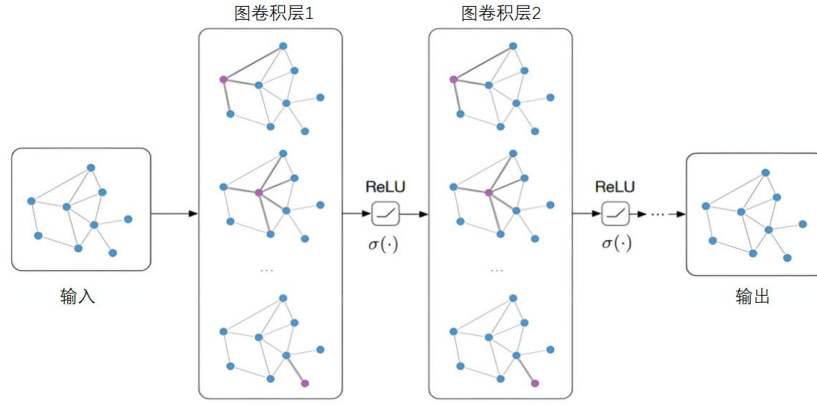


图 2.5 图神经网络计算过程

Fig. 2.5 Computation Process of Graph Neural Networks

图 2.5 展示了 GCN 的计算流程。将整个图的邻接矩阵及节点特征矩阵输入网络。在第一层图卷积中，每个节点利用邻接矩阵进行特征传播，并结合自身特征进行更新。更新后的特征通过 ReLU 激活函数，引入非线性能力，以增强模型的表达能力。随后，该过程在更深层的 GCN 中继续迭代，逐步提取高阶特征，增强对全局拓扑结构的感知能力。

2.4 混合专家（Mixture of Experts）

混合专家模型（Mixture of Experts, MoE）是一种集成学习方法^[72]，通过组合多个专门化的子模型（即“专家”）来提升整体模型的性能。每个专家模型专注于处理特定类型的数据或子任务，门控网络根据输入数据的特征动态地分配权重，决定哪些专家参与计算及其贡献程度。这种动态路由机制使 MoE 能够有效地处理复杂且多样化的任务。

MoE 的核心优势在于其“稀疏激活”机制^[73]，即在每次推理过程中，仅激活部分专家网络，而非同时调用所有专家。这种选择性激活不仅降低了计算资源的消耗，还提高了训练和推理的效率。此外，门控网络的动态分配确保了每个专家针对不同的输入数据进行优化，从而增强了模型的适应性和泛化能力。

图 2.6 展示了 MoE 结构的一种实现方式。通过动态的专家选择和权重分配机制，MoE 能够灵活适应不同任务和数据分布，提升模型的泛化能力和整体性能。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，已广泛应用于自然语言处理、计算机视觉等领域。

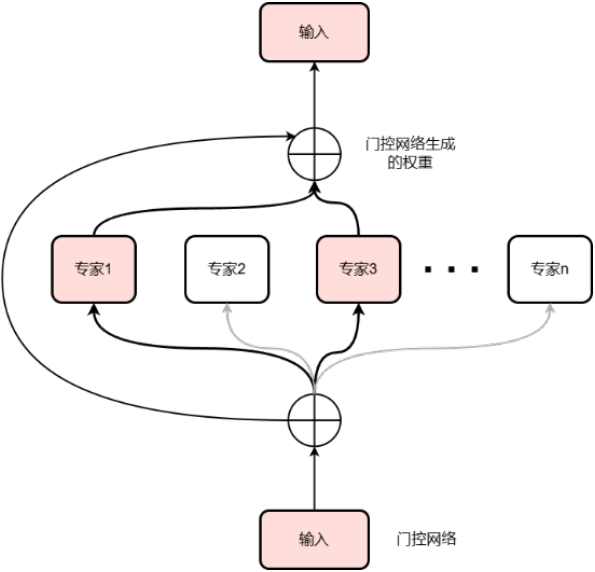


图 2.6 混合专家示意图

Fig. 2.6 Mixture of Experts (MoE) Schematic

2.4 本章小结（Summary）

本章对研究工作涉及的核心理论与技术框架进行了概括性阐述。首先概述了卷积神经网络的核心组件及其工作机制，并分析了自注意力机制在视觉任务中的独特优势。其次梳理了表征学习领域的两大方向，阐释了解耦表征对模型泛化能力的提升作用以及原型学习的认知可解释特性。在此基础上，说明了图卷积网络处理非欧式空间数据的建模方法，以及混合专家系统通过动态路由机制实现计算资源优化的创新思路。这些理论框架和技术路线共同构成了后续创新研究的基石。

3 基于特征解耦与原型学习的滨海湿地精细制图方法

3 Feature Decoupling and Prototype Learning-Based Fine-Grained Mapping Method for Coastal Wetlands

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 互信息解耦原型对齐网络 (Mutual Information Decoupling 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。

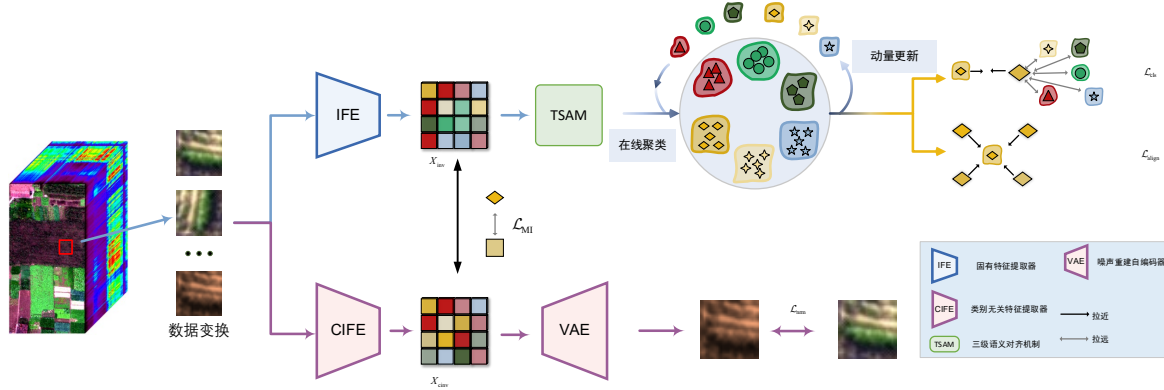


图 3.1 MID-ProNet 网络结构图

Fig. 3.1 Architecture of MID-ProNet

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 $\mathcal{D}'_{\text{train}}$:

$$\mathcal{D}'_{\text{train}} = \{(T(X_i), Y_i) | (X_i, Y_i) \in \mathcal{D}_{\text{train}}\} \quad (3.1)$$

数据增强操作可表示为变换函数 $T(\cdot)$, 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度。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。

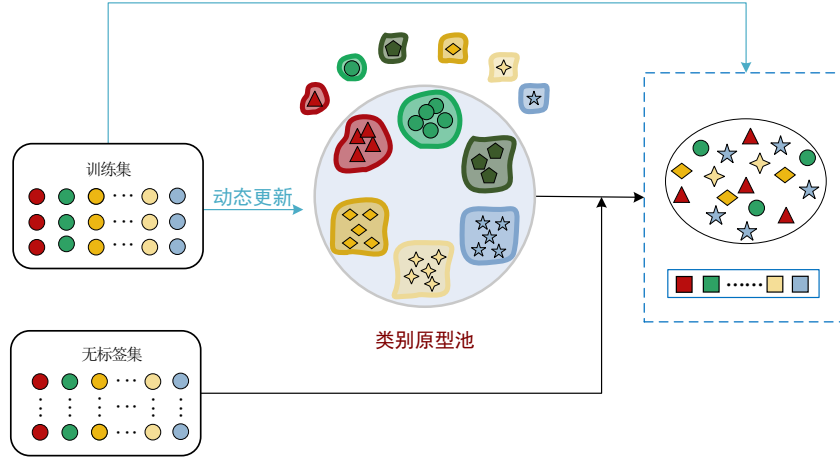


图 3.2 MID-ProNet 网络训练和测试流程图

Fig. 3.2 Training and Testing Pipeline of MID-ProNet

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 $T(\cdot)$ ：

$$\mathcal{D}'_{\text{train}} = \{(T(X_i), Y_i) | (X_i, Y_i) \in \mathcal{D}_{\text{train}}\} \quad (3.2)$$

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。

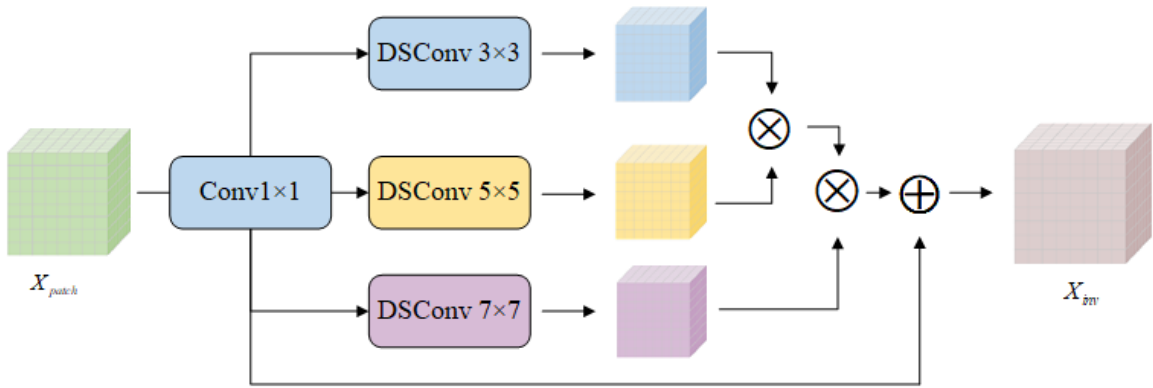


图 3.3 地物固有特征提取器模块结构示意图

Fig. 3.3 Architecture of the Interactive Feature Extraction Module

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 $X_{patch} \in \mathbb{R}^{s \times s \times B}$, 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。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。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。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：

$$F_{multi} = F_{dw}^3 \odot F_{dw}^5 \odot F_{dw}^7 \quad (3.3)$$

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：

$$X_{inv} = \text{GELU}\left(\text{BN}\left(\mathcal{W}_f \odot F_{multi} + F_c\right)\right) \quad (3.4)$$

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。

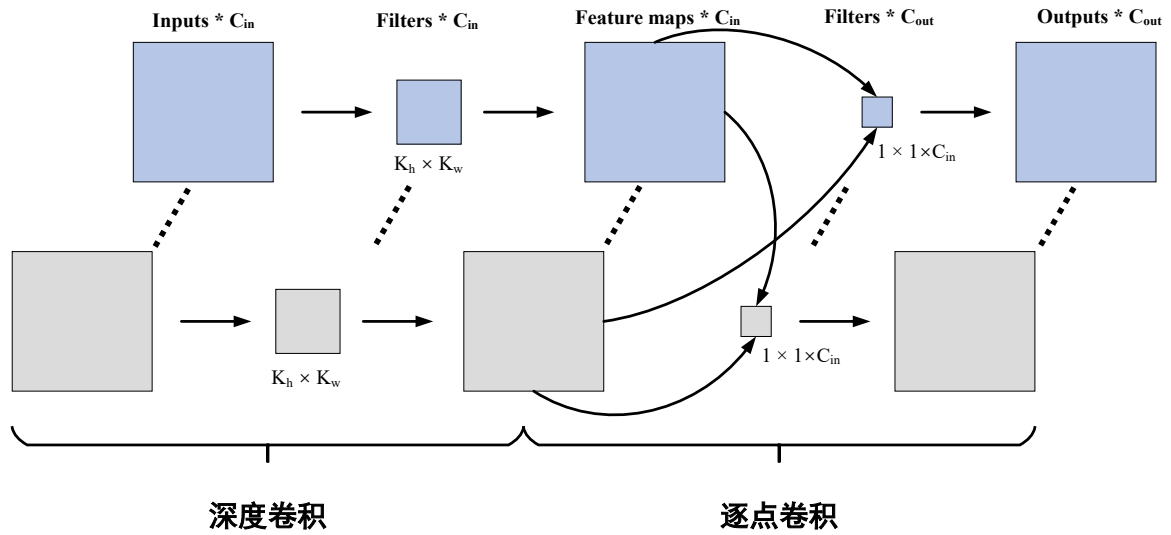


图 3.4 深度可分离卷积结构示意图

Fig. 3.4 Architecture of Depthwise Separable Convolution

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。

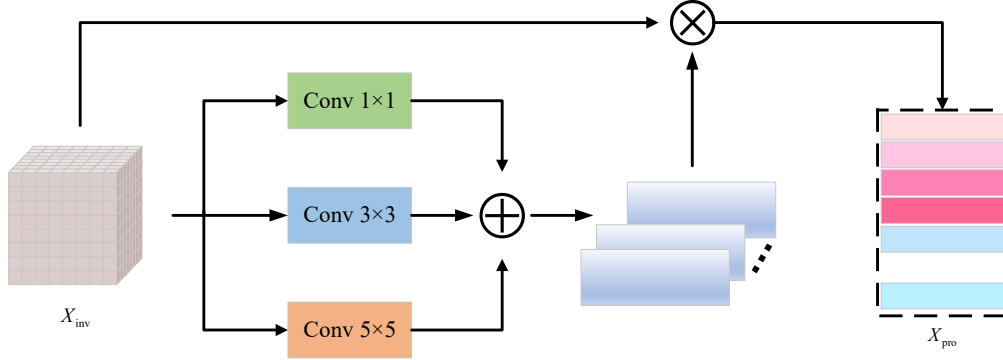


图 3.5 多级语义编码器

Fig. 3.5 Architecture of Multi-Level Semantic Encoder

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：

$$p_t = \alpha p_{t-1} + (1 - \alpha) \frac{1}{N} \sum_{i=1}^N f_i \quad (3.5)$$

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。通过设计的原型引导模块，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：

$$p^{enh} = \lambda \cdot p^{cur} + (1 - \lambda) \cdot \text{softmax}(p^{cur} p^{mem\top}) p^{mem} \quad (3.6)$$

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

$$\mathcal{L}_p = \sum_{c=1}^C \|\mu_c - X_{pro}^c\|_2^2 \quad (3.7)$$

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。

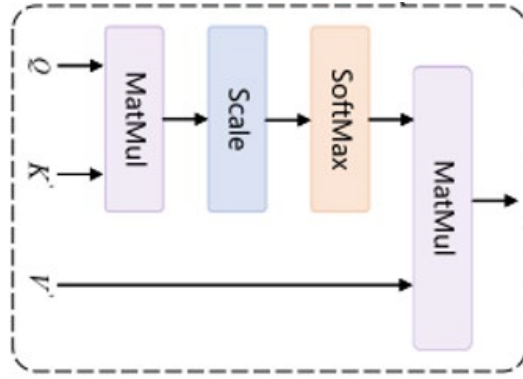


图 3.6 自注意力机制编码器结构示意图

Fig. 3.6 Architecture of Self-Attention Encoder

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：

$$\mathcal{L}_{\text{nrm}} = \| \text{Decoder}(F_{\text{noisy}}) - F_{\text{domain}} \|_2 + \lambda \cdot \text{KL}(q(z | F_{\text{noisy}}) \| p(z)) \quad (3.8)$$

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一致性与类间可分性, 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:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{cls}} + \alpha \mathcal{L}_{\text{MI}} + \beta \mathcal{L}_{\text{align}} + \gamma \mathcal{L}_{\text{nrm}} \quad (3.9)$$

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 实验准备 (Experimental Preparation)

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 3.1 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。

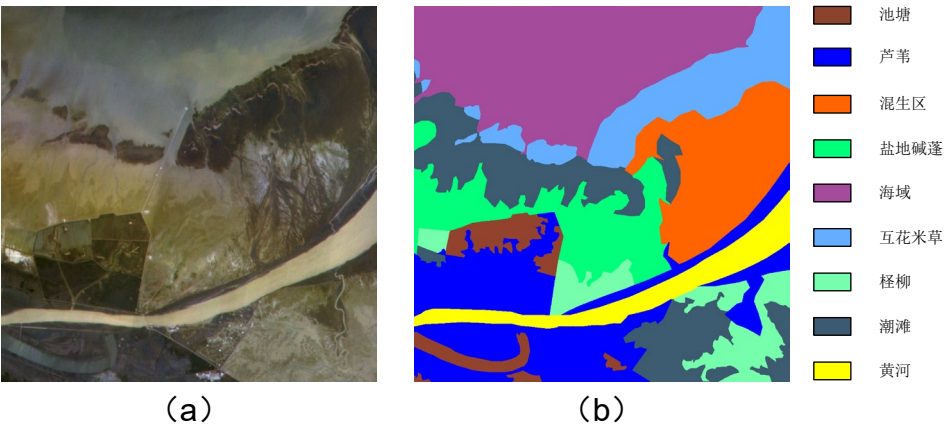


图 3.7 Coast 数据集 (a)假彩色图像 (b)地面真值图
Fig. 3.7 Coast dataset (a) False color composite (b) Ground truth

表 3.2 Coast 数据集上每个类别样本的数量

Table 3.2 Number of samples per category in the Coast dataset		
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

表 3.2 Coast 数据集上每个类别样本的数量（续）

Table 3.2 Number of samples per category in the Coast dataset (Continued)

类别编号	地面覆盖类型	样本数量
4	盐地碱蓬	101347
5	海域	117997
6	互花米草	145829
7	柽柳	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，为后续的分析研究提供了坚实的数据基础。

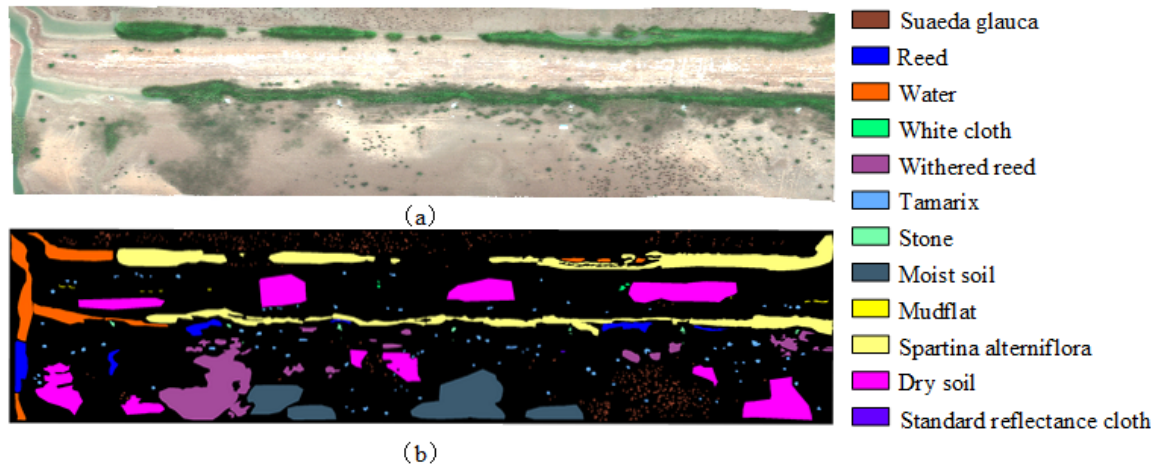


图 3.8 YRD-NC12 数据集 (a)假彩色图像 (b)地面真值图

Fig. 3.8 YRD-NC12 dataset (a) False color composite (b) Ground truth

表 3.3 YRD-NC12 数据集上每个类别的样本数量

Table 3.3 Number of samples per category in the 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

表 3.3 YRD-NC12 数据集上每个类别的样本数量（续）

Table 3.3 Number of samples per category in the YRD-NC12 datasets (Continued)

类别编号	地面覆盖类型	样本数量
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 0.4-2.5um，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。

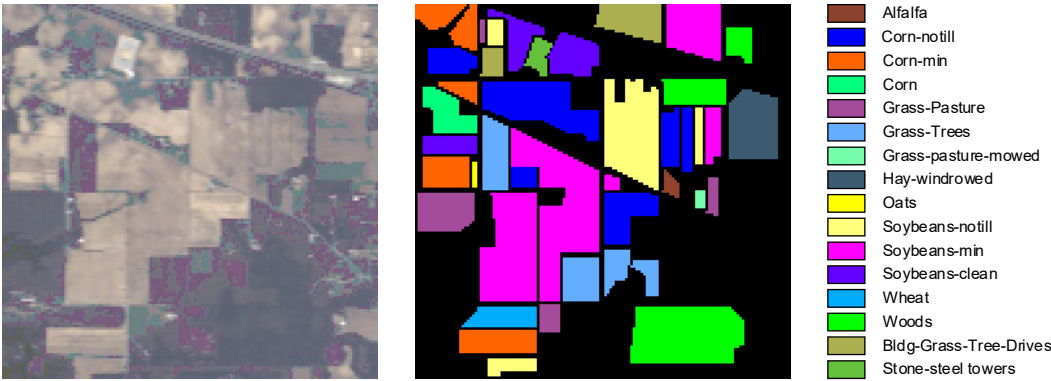


图 3.9 Indian Pines 数据集 (a) 假彩色图像 (b)地面真值图

Fig. 3.9 Indian Pines dataset (a) False color composite (b) Ground truth

表 3.4 Indian Pines 数据集上每个类别的样本数量

Table 3.4 Number of samples per category in the Indian Pines dataset

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

表 3.4 Indian Pines 数据集上每个类别的样本数量（续）
Table 3.4 Number of samples per category in the Indian Pines dataset (Continued)

类别编号	地面覆盖类型	样本数量
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 3.5 混淆矩阵

Table3.5 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：

$$OA = \frac{\sum_{i=1}^n \text{正确预测的样本数}}{\text{总样本数}} \tag{3.10}$$

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：

$$AA = \frac{1}{C} \sum_{i=1}^C \frac{\text{正确预测的第}i\text{类样本数}}{\text{第}i\text{类总样本数}} \quad (3.11)$$

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:

$$\kappa = \frac{OA - P_e}{1 - P_e} \quad (3.12)$$

其中, P_e 是随机一致性的期望概率, 计算公式为:

$$P_e = \sum_{i=1}^C \left(\frac{\text{第}i\text{类实际样本数}}{\text{总样本数}} \times \frac{\text{第}i\text{类预测样本数}}{\text{总样本数}} \right) \quad (3.13)$$

3.3.4 实验对比方法

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^[74]和深度学习方法与 3DCNN^[75]、HyperSN^[76]、DBDA^[77]、SSFTT^[78]、HSIFormer^[79]与 MID-ProNet 进行比较。

(1) SVM: SVM 的核函数能够将数据映射到更高维的空间, 从而解决非线性分类问题。

(2) 3DCNN: 通过将 HSI 切割成多个三维立方体, 构建三维 CNN 模型来提取图像中的空间光谱信息, 相较于仅使用单一光谱域的 1D-CNN 方法相比, 通常能够获得更好的结果。

(3) HyperSN: 3D-CNN 和 2D-CNN 的混合模型。该模型分别使用 3D-CNN 和 2D-CNN 提取光谱空间特征。

(4) DBDA: 具有位置注意和通道注意, 能够获得任意两个通道和任意两个像素之间的相关性, 以增强特征依赖性, 实现了特征提取和分类的端到端训练。

(5) SSFTT: 基于 CNN 和 Transformer 混合结构, 同时引入了高斯加权特征标记器进行特征变换, 来捕获频谱空间特征和高级语义特征。

(6) HSIFormer: 通过结合 3D 和 2D 卷积神经网络 (CNN) 作为特征提取器, 并引入局部窗口注意力 (Local Window Attention, LWA) 机制, 有效地处理了高光谱数据的分类问题

3.3.5 实验参数设置

在训练阶段，优化器采用 Adam 优化器，同时设定批大小为 64，学习率及权重衰减分别为 $1e-4$ 和 $1e-6$ 。为了自适应地调整参数学习率并加速网络权重的更新，初始学习率定为 0.003。训练迭代次数设置为 100 次，将空间大小设置为 13×13 ，并使用交叉熵损失函数进行反向传播。为了保证结果的准确性，各方法的分类结果都被执行了 10 次，最终的分类结果是 10 次实验的平均值。

3.4 实验与分析 (Experiments and analysis)

3.4.1 实验结果

表 3.5-3.7 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。

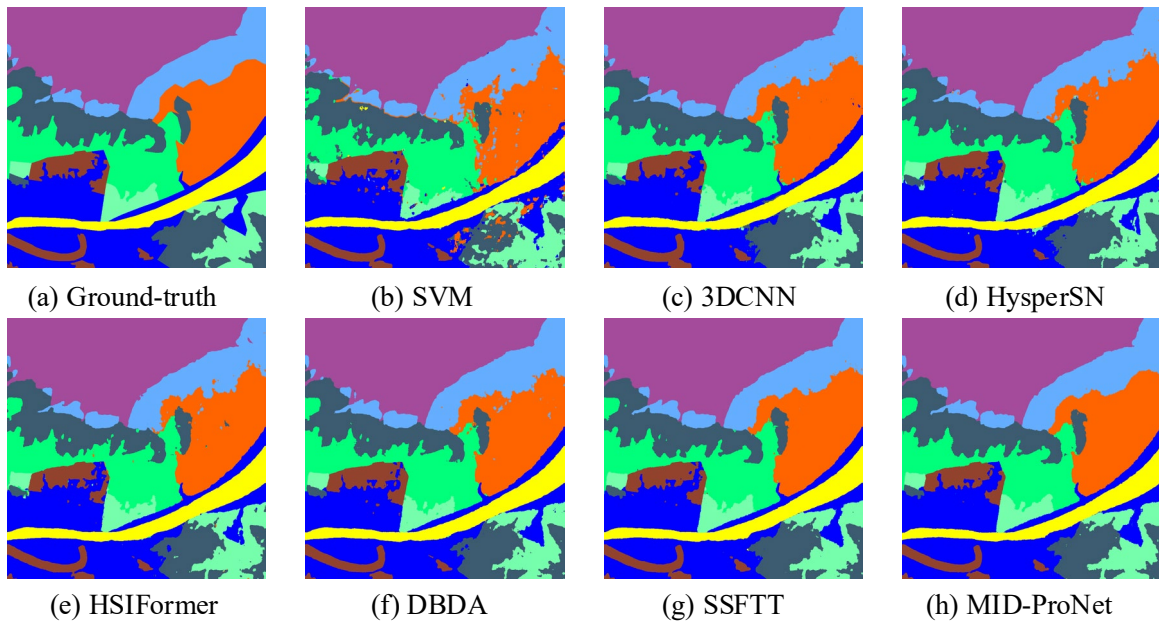


图 3.10 Coast 数据集上的分类可视化图

Fig. 3.10 Classification visualization map on 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。但是从图像中可以看到，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。HSIFormer 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 数据集上的分类结果

Table 3.6 Classification results on the Coast dataset

类别	SVM	3DCNN	HybirdSN	HSIFormer	DBDA	SSFTT	MID-ProNet
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，本章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，表明本章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，本章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，进一步验证了本章提出的方法的优越性。

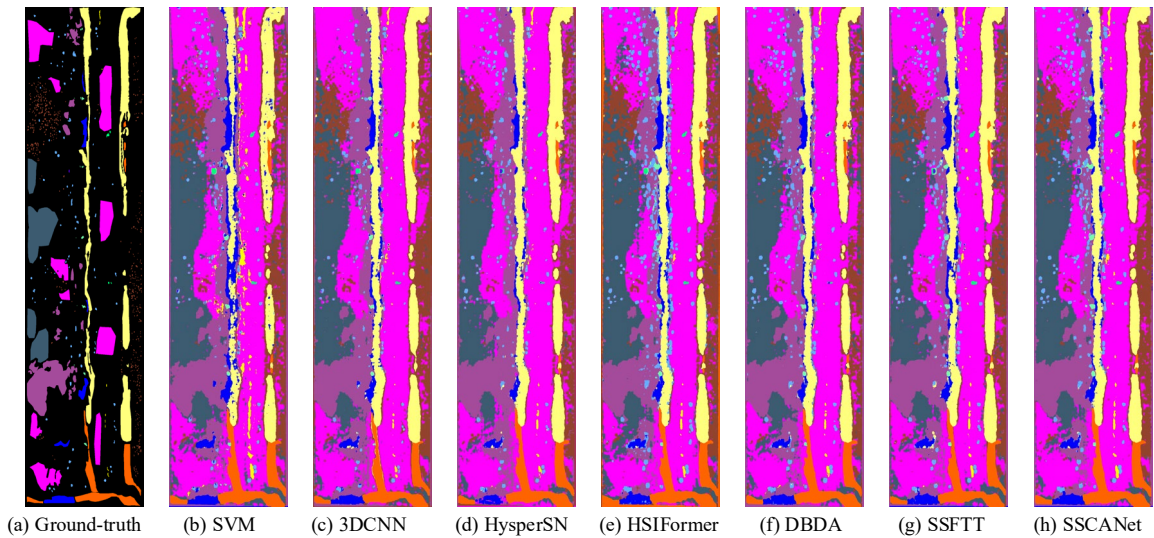


图 3.11 YRD-NC12 数据集上的分类可视化图

Fig. 3.11 Classification visualization map on 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，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，但是从表中可以看到，本章方法取得了最优的分类结果，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 3.7 YRD-NC12 数据集上的分类结果

Table 3.7 Classification result on the YRD-NC12 dataset

类别	SVM	3DCNN	HybirdSN	HSIFormer	DBDA	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	99.82	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

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。

本章提出的 MID-ProNet 方法通过对地物光谱进行特征解耦，分离出地物的本质表征，使得整体预测结果较好，改善了混淆分类现象。但是本章的方法仍存在部分局限性，例如，Indian Pines 数据集上“Soybeans-notill”的精度略低于 SSFTT。这可能是因为本章的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。

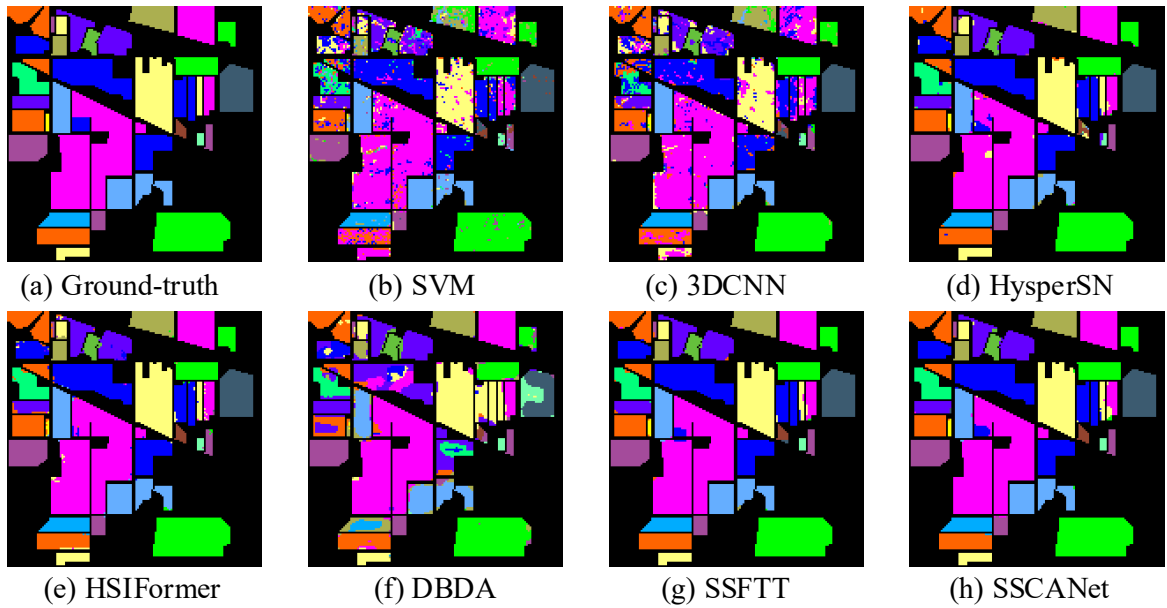


图 3.12 Indian Pines 数据集上的分类可视化图

Fig. 3.12 Classification maps for the Indian Pines dataset

表 3.8 Indian Pines 分类结果

Table 3.8 Classification visualization map on the Indian Pines dataset

类别	SVM	3DCNN	HybirdSN	HSIFormer	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

3.4.2 消融实验

为了验证所提出的 IFE 模块、DPLM 模块和基于互信息特征解耦模块 MI 的有效性，在 MID-ProNet 上通过移除特定组件进行消融实验以进行比较。表 3.11 显示了在 IP 数据集上的消融实验，当 MID-ProNet 移除基于互信息特征解耦 MI，AA 为 96.45%，性能下降了 2.19%。这一结果表明，MI 模块在提升分类性能方面发挥了重要作用。其主要原因在于 MI 模块利用互信息进行特征解耦，从而提取地物的本质特征，提高模型对类别间混淆的辨别能力。MID-ProNet 移除 DPLM 时，OA、AA 和 Kappa 分别为 95.90%、91.41%和 95.32%。主要原因是 DPLM 通过原型学习和多级语义编码器学习地物的类别原型，并且利用已有的类别原型来补充当前类别的信息，使提取的特征可以及时进行修正。通过移除 IFE，替换为 2 维卷积，能够明显看到现有 OA、AA 和 Kappa 分别降低了 1.09%，1.22%和 1.19%。当使用所有的模块时，MID-ProNet 方法 OA、AA 和 Kappa 分别为 98.89%、98.64%和 98.74%，性能达到了最优。

表 3.9 MID-ProNet 消融实验的精度

Table 3.9 Ablation evaluation accuracies of MID-ProNet

	MI	IFE	DPLM	OA	AA	$\kappa \times 100$
1	×	√	√	97.12	96.45	96.86
2	√	×	√	97.85	97.42	97.55
3	√	√	×	95.90	91.41	95.32
4	√	√	√	<u>98.89</u>	<u>98.64</u>	<u>98.74</u>

3.4.3 特征解耦有效性分析

为了进一步验证互信息特征解耦的有效性，本文在 Coast 数据集上对比了进行特征解耦和未进行特征解耦两种情况下混淆矩阵的可视化结果，分别如图 4.10 (a)和 4.10 (b)所示。在这两张图中，纵坐标表示真实的类别标签，横坐标表示预测的类别标签，对角线表示预测的类别标签与真实的类别标签相同的概率。通过观察和对比，相较于图 4.10 (a)中未进行特征解耦融合后的准确分类效果，图 4.10 (b)中进行特征解耦的分类结果存在更少的错误分类情况。这一显著差异直观地展现了特征解耦在改善分类精度方面的积极作用。

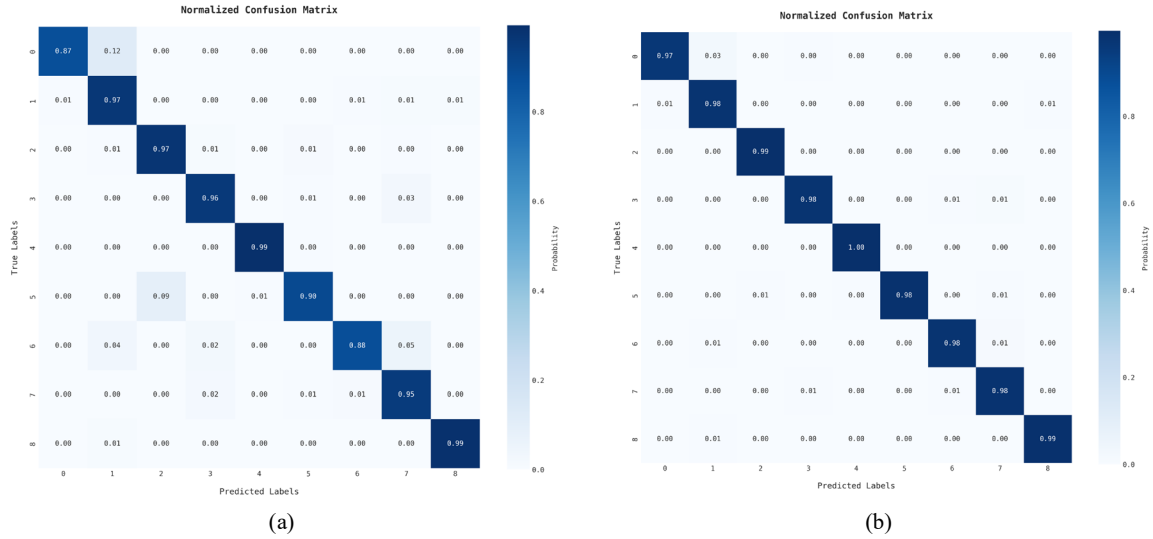


图 3.13 混淆矩阵可视化结果 (a)未进行特征解耦(b)进行特征解耦

Fig.3.13. Visualization results of the confusion matrix (a) With spatial-spectral feature fusion (b) Without spatial-spectral feature fusion

3.4.4 超参数分析

空间输入大小是影响高光谱图像分类性能的关键超参数，直接决定了像素分类过程中可利用的邻域空间信息量。为系统评估该参数对 MID-ProNet 模型性能的影响，本研究在 **Indian Pines** 数据集上设计了控制变量实验，选取 6 种不同尺度的图像块尺寸（ 9×9 、 11×11 、 13×13 、 15×15 、 17×17 和 19×19 ）进行测试，并在图 3.12 中展示了不同输入尺寸下模型性能的变化趋势。实验结果表明，随着空间输入大小的增加，整体精度（OA）、平均精度（AA）及 Kappa 系数均显著提升，表明适当扩大空间输入有助于增强分类性能。然而，当输入块尺寸超过 13×13 时，分类性能在上述三个指标上均出现下降趋势。因此，本文最终将图像块大小设定为 13×13 ，以充分利用空间信息，同时避免因输入规模过大导致的性能退化。该现象的主要原因在于，当输入块较小时，模型所能提取的空间信息有限，难以充分捕捉像素间的关联性与上下文信息，从而影响分类效果；而当输入块过大时，尽管可获取更多空间信息，但同时增加了计算复杂度，并可能导致模型过拟合，进而削弱其泛化能力。

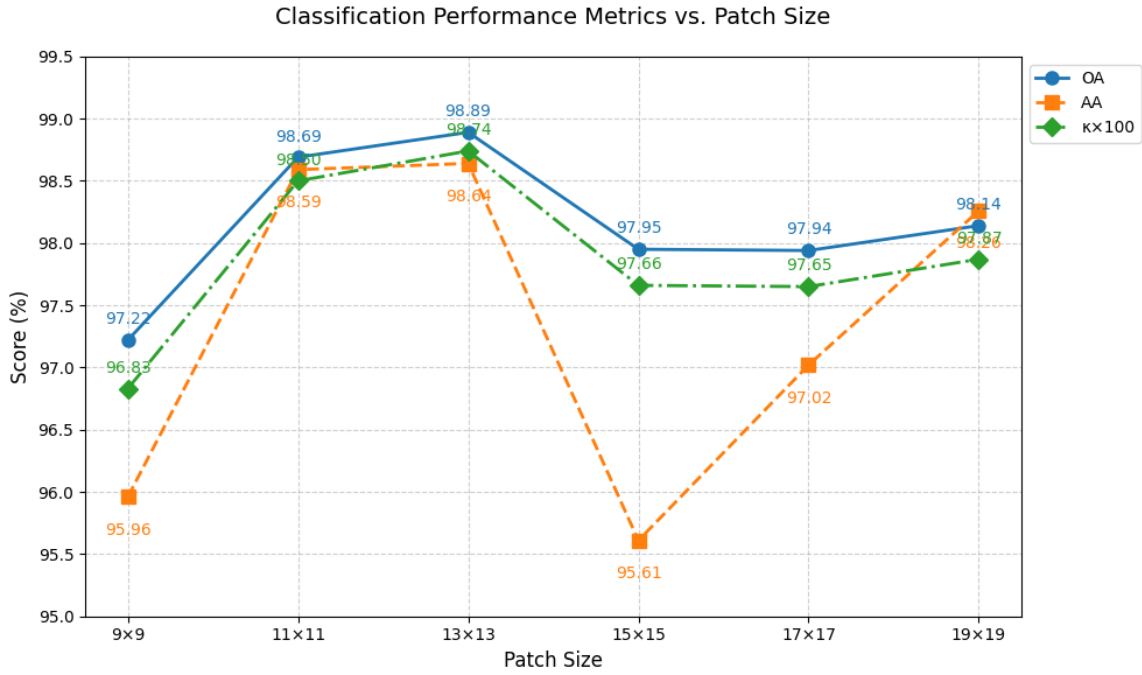


图 3.14 不同空间输入大小的对分类性能的影响

Fig.3.14 Impact of different spatial input sizes on classification performance

3.5 本章小结 (Summary)

本章节提出了一种基于特征解耦与原型学习的滨海湿地植被精细分类方法 (MID-ProNet)。该方法通过互信息约束的双分支网络结构,实现了地物固有特征与类别无关特征的有效解耦,解决了复杂环境下光谱特征混淆的关键问题;同时引入动态原型学习机制,通过多级语义编码和在线原型更新策略,增强了模型对滨海湿地植被光谱变异性的适应能力。实验结果表明,该方法在多个湿地数据集和通用数据集上均展现出优越的分类性能,特别是互信息解耦模块的引入显著提升了模型的判别能力,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 Multimodal Data Fusion-Based Fine-Grained Vegetation Mapping Method for Coastal Wetlands

4.1 引言 (Foreword)

滨海湿地在调节气候、维持生物多样性和缓解海岸侵蚀方面发挥着重要的生态功能。然而,由于潮汐作用、植被斑块破碎化以及入侵物种扩散等因素的影响,滨海湿地植被在光谱和几何特征上呈现出高度复杂性,导致传统的高光谱分类方法在此类场景下的适用性受限。高光谱图像虽能捕捉连续光谱特征以识别细微差异,却受限于空间分辨率不足,难以精确刻画植被边界,激光雷达可构建厘米级三维地形模型,但缺乏有效的光谱鉴别能力。两种数据在光谱响应与空间结构表征维度上存在显著异构性,致使传统单模态方法难以协同利用多维信息。因此,如何有效融合多模态遥感数据,提升滨海湿地植被的精细分类能力,仍是当前亟待解决的关键问题。

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)。受多模态特征渐进融合机制启发,可变形多模态编码器通过可变形卷积的动态感受野调整,实现植被几何形态自适应感知,配合多尺度并行卷积完成跨模态特征对齐。在此基础上,设计几何引导注意力模块,融合 Sobel 算子与高程梯度构建双通道几何先验,通过 MLP 生成光谱描述子建立空间-光谱耦合矩阵,有效缓解地形起伏导致的光谱混淆问题。自适应渐进融合模块,该模块采用门控加权策略动态调节高光谱与 LiDAR 特征贡献度,通过图卷积网络构建植被斑块拓扑关联,显著提升破碎植被区域的空间一致性表达。最终,动态路由分类模块引入专家选择机制,依据特征复杂度自适应配分类器参数,结合空间感知路由权重计算,实现复杂场景下的精准决策。

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 滨海湿地多模态数据植被精细制图主要流程 (Main Workflow of Multimodal Data-Based Fine-Grained Vegetation Mapping in Coastal Wetlands)

本章基于多模态数据融合的方法对滨海湿地遥感影像进行精细分类,具体流程如图 4.1 所示,包括三个基本阶段:(1)遥感影像数据预处理,对获取的多源遥感

影像进行预处理，包括数据对齐、辐射校正、大气校正等，以确保数据的空间一致性和光谱精度。随后，利用 ArcGIS 软件对影像进行几何校正，裁剪出地物类别丰富、特征明显且适合分类的区域，然后基于 LabelMe 软件制作模型训练需要的真值解译图；（2）网络模型训练，通过不断学习训练数据的特征分布，更新模型参数，优化训练目标；（3）网络模型测试，使用迭代优化收敛后的网络模型对测试集的遥感数据进行预测。

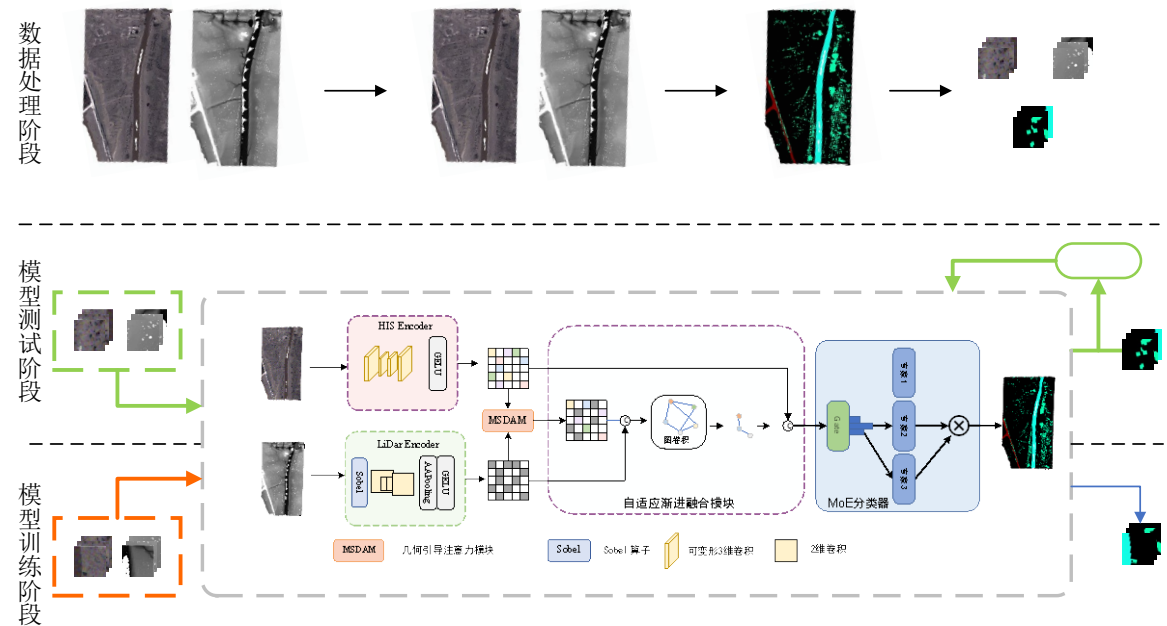


图 4.1 滨海湿地多模态数据植被精细制图主要流程图

Fig. 4.1 Main flowchart of fine vegetation mapping using multimodal data in coastal wetlands

本章构建了一种用于滨海湿地植被精细制图的多模态数据融合网络 GSJDD-Net，通过提取滨海湿地植被几何信息引导高光谱图像信息，有效的解决了滨海湿地精细分类。与其他方法相比，该方法能够充分利用激光雷达图像中的空间几何信息和高光谱图像中的空谱特征，提高了滨海湿地植被精细制图的准确率。

4.2.1 研究区域及数据预处理

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°45'~37°51'N，119°06'~119°14'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 km²，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。

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。

4.2.2 训练及处理流程

在 GSJDD-Net 网络模型训练阶段,将训练样本输入到 GSJDD-Net 中,最终输出层的激活函数选择了 Softmax 函数,以生成预测概率图。利用预测结果与对应的标签图像计算损失,使用误差反向传播对网络参数进行更新。滨海湿地精细分类属于多分类任务,因此使用交叉熵作为模型损失函数。此外,模型训练过程中利用自适应矩估计 Adam 优化算法对网络参数进行迭代更新。

GSJDD-Net 网络模型的测试阶段则是用网络中已经训练好的权重参数去预测新数据的一个过程。由于语义分割网络是一种端到端的网络模型,将用于测试的卫星遥感图像输入训练好的 GSJDD-Net 网络模型,可以直接获得滨海湿地精细分类结果。

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

针对滨海湿地植被分类中光谱特征与几何特征融合困难的问题,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)。网络结构如图 4.1 所示,由可变形多模态编码器 (Deformable Multimodal Encoder, DME)、几何引导注意力模块 (Geometric-Guided Attention, GSA)、自适应渐进融合模块 (Adaptive Progressive Fusion Module, APFM) 和动态路由分类模块 (Dynamic Routing Classification Module, DRCM) 四部分组成。

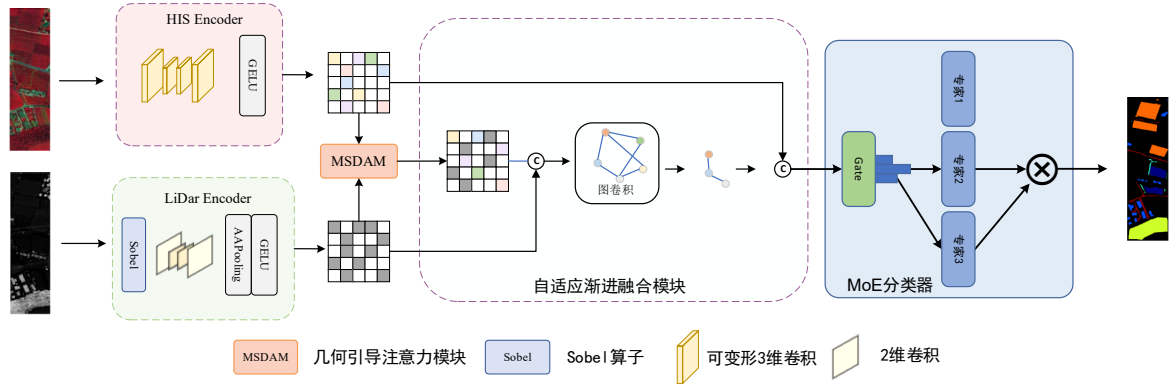


图 4.2 GSJDD-Net 网络结构示意图

Fig. 4.2 Schematic diagram of GSJDD-Net network structure

GSJDD-Net 采用 Patch 图像块处理机制，输入高光谱数据 $X_{patch} \in \mathbb{R}^{s \times s \times B}$ 和激光雷达数据 $X_{gpatch} \in \mathbb{R}^{s \times s \times 1}$ ，其中 s 为局部图像块尺寸， B 为高光谱波段数。通过滑动窗口从整幅图像中提取局部块，每个样本包含中心像素及其邻域的空间信息。首先，可变形多模态编码器（DME）结合可变形卷积网络（DCN）与多尺度几何增强单元（MGEU），分别从高光谱和激光雷达局部块中提取空谱特征 F_s^0 和几何特征 F_g^0 。可变形卷积能够自适应调整采样位置，以更精准地建模不规则植被边界。同时，多尺度几何增强单元（MGEU）结合不同尺度的并行卷积，进一步丰富地形特征的表达。随后，几何引导注意力模块（GSA）从激光雷达数据中提取梯度信息，构建几何先验，并生成注意力权重 W_m ，用于修正空谱特征 F_s^0 以抑制地形起伏导致的边缘模糊现象。接下来，自适应渐进融合模块（APFM）采用两级融合机制，实现不同模态数据之间的融合。此外，该模块利用图卷积网络（GCN）建模像素间的空间拓扑关系，以增强破碎植被斑块的局部一致性。最后，动态路由分类模块（DRCM）依据局部块特征复杂度计算专家分类器权重 W_r^k ，并通过加权融合多个分类器输出，实现自适应分类决策，有效提升分类精度。

4.3.1 可变形多模态编码器

滨海湿地植被受潮汐交替影响，边界通常呈现模糊过渡与不规则形态，传统固定感受野的卷积操作难以精确建模此类复杂形态。针对此问题，本研究设计了可变形多模态编码器（Deformable Multi-modal Encoder, DME），通过动态调整感受野分布，增强对不同形态植被斑块的自适应表征能力。

针对输入的高光谱数据 $X_{patch} \in \mathbb{R}^{s \times s \times B}$ 和激光雷达数据 $X_{gpatch} \in \mathbb{R}^{s \times s \times 1}$ ，DME 首先进行特征映射：

$$F_s^0 = \phi_s(X_{patch}), F_g^0 = \phi_g(X_{gpatch}) \quad (4.1)$$

其中， ϕ_s 和 ϕ_g 为初始特征提取映射函数，将原始数据转换为初始特征

$F_s^0 \in \mathbb{R}^{s \times s \times C_s}$ 和 $F_g^0 \in \mathbb{R}^{s \times s \times C_g}$ 。

随后，DME 采用可变形卷积（DCN）学习位置偏移量 Δp ，实现感受野的动态调整，以提高对复杂植被边界的表征能力。多尺度几何增强单元（MGEU）进一步采用不同尺度的并行卷积路径（ 3×3 、 5×5 、 7×7 ），增强对不同粒度地形特征的感知能力。

4.3.2 几何引导注意力模块

滨海湿地地形起伏变化显著，导致相同植被在不同地形条件下的光谱特征存在较大差异，从而增加了分类难度。为缓解地形变化导致的光谱混淆问题，本研究提出几何引导注意力模块（GSA），利用几何信息引导光谱特征提取过程。

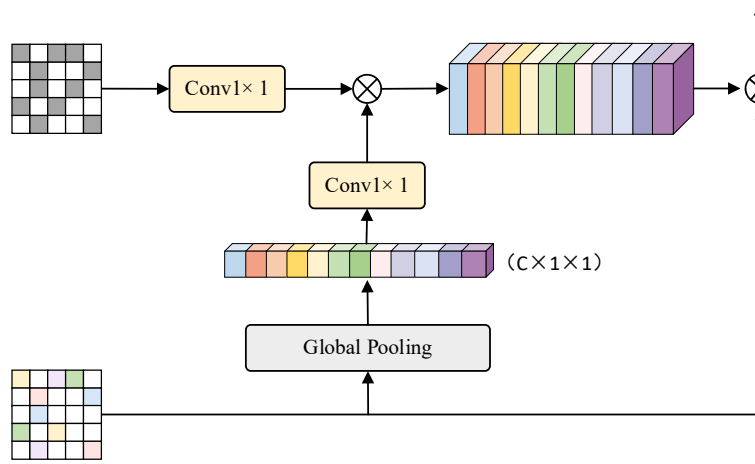


图 4.3 几何引导注意力模块结构示意图

Fig. 4.3 Structure diagram of the geometry-guided attention module

网络结构如图 4.3 所示，几何引导注意力模块（GSA）利用地形信息优化光谱特征提取，以降低滨海湿地复杂地形条件下的光谱混淆。几何引导注意力模块通过从多模态编码器处理后的激光雷达数据 F_g^0 中提取的地形梯度信息，引导同样经过编码的高光谱影像（HSI）空谱特征 F_s^0 以获取地物空间边缘信息。 F_g^0 经过 1×1 卷积处理，获取梯度特征，并通过 ELU 激活函数增强表达能力。 F_s^0 经过全局池化，并通过多层感知机计算空谱描述子，以提取更具判别力的空谱特征 F_s^1 。空谱描述子经维度扩展，与地形信息对齐，形成跨模态信息交互。

在注意力计算过程中， F_s^0 经过 1×1 卷积生成查询（Query）张量，LiDAR 先验信息与光谱描述子拼接后作为键（Key）张量。查询张量与键张量在通道维度拼接，并通过 1×1 卷积计算能量分布，经 Sigmoid 激活函数生成最终的注意力权重 W_m 。最终，增强后的空谱特征通过可学习的缩放因子进行加权求和得到几何引导后的空谱特征 F_s^1 ，实现几何信息对空谱特征的显式引导，从而提高分类任务在复杂地形区域的稳健性。

4.3.3 自适应渐进融合模块

自适应渐进融合模块整体结构如图 4.1 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，通过通道拼接整合 GSA 得到的高光谱空谱特征 F_s^1 与 DME 提取的 LiDAR 特征 F_g^0 ，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：

$$\mathcal{F}_{fuse}^1 = \mathcal{W}_c \left(\left[F_s^1 \oplus F_g^0 \right] \right) \quad (4.2)$$

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，将通道维度保持与拼接前一致。

在高级阶段，引入图卷积（Graph Convolutional Network, GCN）进行特征交互，以提升多模态特征的融合适应性和几何保持能力。利用基于坐标距离构建的邻接矩阵 $\mathcal{A} \in \mathbb{R}^{N \times N}$ 进行图卷积传播得到 \mathcal{F}_{gc} ，建模局部到全局的特征关联：

$$\mathcal{F}_{gc} = \text{GCN} \left(\mathcal{F}_{fuse}^1, \mathcal{A} \right) \quad (4.3)$$

为了 APF 引入可学习的参数动态的调整融合比例，融合表征可以表示为：

$$\mathcal{F}_{fuse}^2 = \left[F_s^0 \oplus \mathcal{F}_{gc} \right] \quad (4.4)$$

其中 \oplus 表示通道拼接操作，APF 模块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，相较于传统单次拼接融合方式，能够有效缓解信息淹没问题，提高多模态特征的代表能力。

4.3.4 动态路由分类模块

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 4.1 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 $E_k(\cdot)$ 通过不同尺度深度可分离卷积（Depthwise Separable Convolution, DSConv）提取特定模式的特征表示，其计算形式如下：

$$f_{expert} \left(\mathcal{F}_{fuse}^2 \right) = \text{DWConv} \left(\mathcal{F}_{fuse}^2 \right) \quad (4.5)$$

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，引入路由网络来计算每个专家的选择概率。首先，对输入特征 \mathcal{F}_{fuse}^2 进行

全局平均池化，提取全局信息：

$$p_k = \frac{\exp(W_r^k \otimes \text{GAP}(\mathcal{F}_{fuse}^2))}{\sum_{i=1}^3 \exp(W_r^i \otimes \text{GAP}(\mathcal{F}_{fuse}^2))} \quad (4.6)$$

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：

$$P = \sum_{k=1}^3 p_k \cdot E_k(\mathcal{F}_{fuse}^2) \quad (4.7)$$

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 实验及分析结果 (Experimental Results and Analysis)

4.4.1 实验数据

本章所用数据为 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空空的 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。

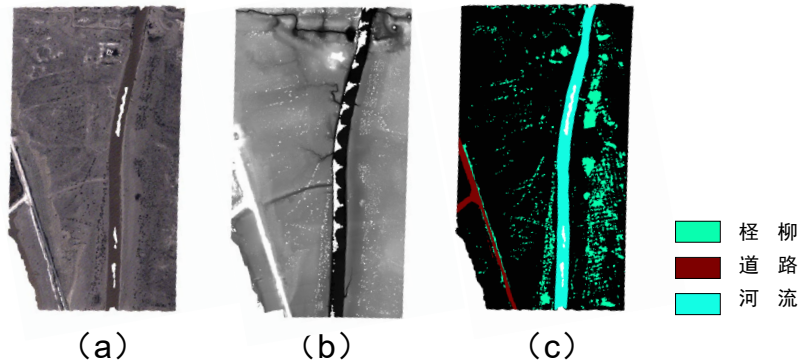


图 4.4 Chengliu 数据集 (a) 高光谱假彩色图像 (b) Lidar 图像 (c) 地面真值图

Fig. 4.4 Chengliu dataset (a) Hyperspectral false-color image (b) LiDAR image (c) Ground truth map

表 4.1 Chengliu 数据集上每个类别的数量

Table 4.1 Number of samples per category in the Chengliu dataset

类别编号	地面覆盖类型	样本数量
1	桉柳	55811
2	道路	12252
3	河流	44979
Total		113042

此外，为了验证方法的泛化性，本章节在 His 和 Lidar 多模态 Trento 数据集同样开展了对比实验。Trento 数据集采自意大利南部城市 Trento 的农村地区，空间分辨率为 1 m，涵盖六种不同的土地覆盖类型，包括苹果园、建筑物、裸地、木材、葡萄园和道路。LiDAR DSM 数据由 Optech ALTM 3100EA 传感器获取，而高光谱数据由 AISA Eagle 传感器采集，二者的空间分辨率均为 1 m，影像尺寸为 600 × 166 像素。Trento 数据集的高光谱影像覆盖 402.89 ~ 989.09 nm 波段，共包含 63 个光谱通道，光谱分辨率为 9.2 nm。

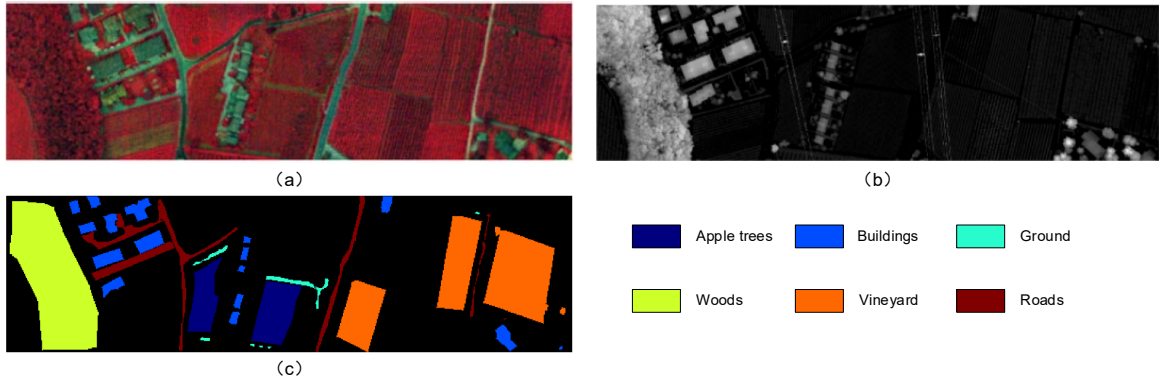


图 4.5 Trento 数据集 (a) 高光谱假彩色图像 (b) Lidar 图像 (c)地面真值图

Fig. 4.5 Trento dataset (a) Hyperspectral false-color image (b) LiDAR image (c) Ground truth map

表 4.2 Trento 数据集上每个类别的数量

Table 4.2 Number of samples per category in the Trento dataset

类别编号	地面覆盖类型	样本数量
1	Buildings	4034
2	woods	2903
3	Roads	479
4	Apples	9123
5	Ground	10501
6	Vineyard	3174
Total		30214

本文使用每个数据集的不同比例的训练样本来验证本文提出的模型的有效性。在 chengliu 数据集和 Trento 两个数据集中，随机抽取 2%和 2%的数据作为训练样本。表 3.1-3.4 详细说明了所有数据集的样本划分。

4.4.2 实验对比方法

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 TBCNN^[80]、CCRNet^[81]、Glt-Net^[82]、EndNet^[83]、CALC^[84]、MAHiDFNet^[85]进行了比较。

TBCNN: 提出基于卷积神经网络的多源遥感数据分类框架, 通过设计并行 CNN 分支分别提取不同模态数据的空间-光谱特征, 并采用全连接层实现多特征决策级融合。

CCRNet: 构建面向多模态遥感数据的统一 CNN 架构, 利用共享权重卷积层提取跨模态共性特征, 同时保留模态特异性特征, 通过自适应特征选择机制提升异源数据兼容性。

Glt-Net: 提出全局-局部 Transformer 网络, 使用多头注意力机制捕获跨模态全局上下文信息, 结合局部卷积模块提取细粒度空间特征, 实现多源特征协同建模。

EndNet: 设计深度编码器-解码器网络结构, 编码器采用多尺度卷积提取多层次特征, 解码器通过跳跃连接实现特征重构, 最终融合光谱高度信息与 LiDAR 高程特征。

CALC: 采用耦合对抗学习框架实现高光谱与 LiDAR 数据融合分类, 通过构建双流特征提取器和对抗判别器, 促使网络学习模态不变性特征表示。

MAHiDFNet: 构建多尺度空间-光谱跨模态注意力网络, 通过并行多分支结构提取不同尺度的空谱特征, 设计跨模态注意力模块动态校准高光谱与 LiDAR 特征的空间对应关系。

4.4.3 实验参数设置

本章所使用的实验环境与评价指标与第三章相同。在参数设置上, 本章在训练阶段, 优化器采用 Adam 优化器, 同时设定批大小为 64, 学习率及权重衰减分别为 $1e-4$ 和 $1e-6$ 。为了自适应地调整参数学习率并加速网络权重的更新, 初始学习率定为 0.003。训练迭代次数设置为 100 次, 将空间大小设置为 11×11 , 并使用交叉熵损失函数进行反向传播。为了保证结果的准确性, 各方法的分类结果都被执行了 10 次, 最终的分类结果是 10 次实验的平均值。

4.4.4 实验结果

第一个实验是在 chengliu 数据集上开展的。表 4.3 表明了 6 种对比方法和 GSJDD-Net 在自建的 chengliu 数据集上的实验结果。从表 4.3 中可以明显看出, GSJDD-Net 在 OA、AA 和 Kappa 方面优于其他方法。Chengliu 数据集不同土地覆盖类别的训练

样本存在不平衡。其中怪柳的数量约占 49.37%，将近一半，而道路的总样本数量只有 10.84%，这在一定程度上对道路的分类造成影响，所有的方法在道路的分类结果上表现普遍都不是特别好，最高的准确率是本章的方法实现的，准确率达到了 95.88%。这表明即使在数据样本不均衡的情况下，该方法仍能取得出色的分类性能。

表 4.3 Chengliu 数据集上的分类结果

Table 4.3 Classification results on the Chengliu dataset							
类别	TBCNN	CCRNet	Glt-Net	EndNet	CALC	MAHiDFNet	GSJDD-Net
1	98.74	98.72	98.76	98.96	98.62	98.54	99.10
2	94.10	94.39	93.25	94.38	93.92	94.49	95.88
3	99.68	99.30	99.61	99.73	99.46	99.68	99.93
OA (%)	98.61	98.48	98.50	98.77	98.45	98.55	99.03
AA (%)	97.51	97.47	97.21	97.69	97.34	97.57	98.26
$\kappa \times 100$	97.63	97.41	97.44	97.90	97.35	97.53	98.35

图 4.6 展示了在 Chengliu 数据集上的实验结果。从分类结果图可以明显看出，滨海湿地植被分类存在破碎化的分布特征，这给分类任务带来了一定挑战。尤其是在道路与怪柳相邻区域，常出现误分类现象。从图中可以观察到，Glt-Net 和 EndNet 在怪柳与道路交界处的分类结果存在较多误分。造成这一现象的主要原因有两点：一方面，这两个类别的数据存在样本不均衡的问题；另一方面，怪柳呈现破碎化分布，导致与道路类别之间产生邻近像元效应。这些现象进一步验证了本章所提出方法在应对复杂场景方面的优势。

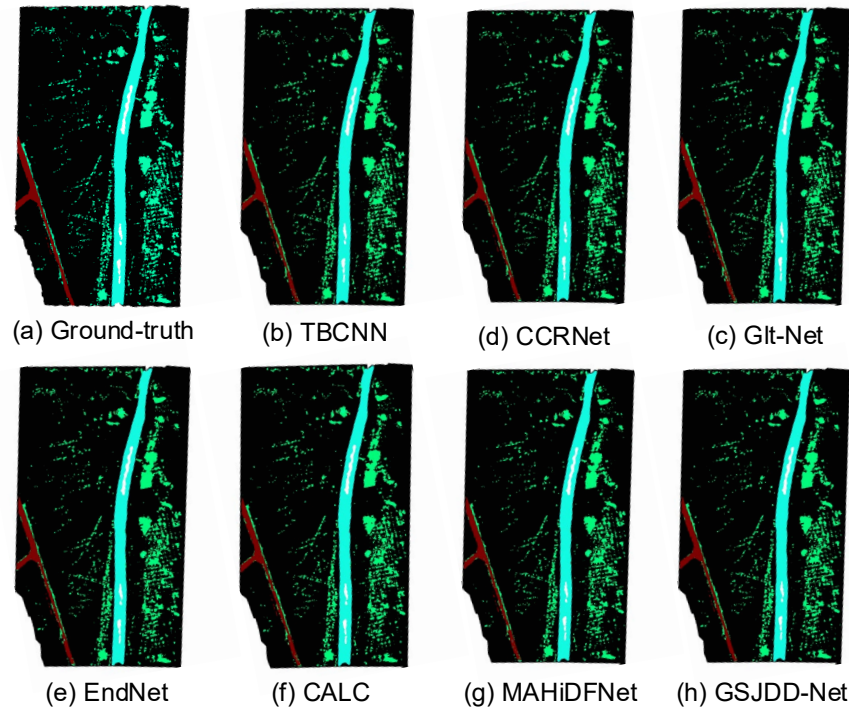


图 4.6 Chengliu 数据集上的分类可视化图

Fig. 4.6 Classification visualization map on the Chengliu dataset

第二个实验是在 Trento 数据集中开展的，传统的多模态方法更多的是按照数据集官方提供的训练样本数量来进行的，本章节没有遵循传统的训练样本数量，因为这在一定程度避免了样本不均衡对于模型的影响。表 4.4 表明了 6 种对比方法和 GSJDD-Net 在 Trento 数据集上的实验结果。从表 4.4 中可以明显看出，GSJDD-Net 在 OA、AA 和 Kappa 方面优于其他方法。特别是 Trento 数据集中不同土地覆盖类别的训练样本存在不平衡。本章的方法在大部分的类别上都取得了最优的训练结果，特别是在“roads”类别，本章的方法要明显的领先其他的对比方法至少 2.86%，证明了本章提出方法的优越性。从表中可以看到 Glt-Net 取得的结果并不是很理想，特别是在“roads”和“Vineyard”这两个类别上，明显低于其他的模型，这可能是因为 Glt-Net 是基于全局-局部 Transformer 网络，现有训练样本的参数量并不能完全的激活 Transformer 网络。相对于 Transformer 和架构的方法，CNN 架构的方法 CCRNeth 和 TB-CNN 要明显优于 Glt-Net，更进一步验证了猜想。

表 4.4 Trento 数据集上的分类结果

Table 4.4 Classification results on the Trento dataset

类别	TBCNN	CCRNeth	Glt-Net	EndNet	CALC	MAHiDFNet	GSJDD-Net
1	98.65	98.03	98.44	96.93	97.34	99.86	99.97
2	97.34	98.30	98.26	97.62	98.54	99.57	99.81
3	78.92	78.28	66.24	88.60	77.85	91.98	94.84
4	100.00	100.00	99.99	100.00	100.00	100.00	100.00
5	99.03	99.00	99.01	98.70	98.99	100.00	99.99
6	98.70	96.23	96.01	97.08	97.50	99.31	99.39
OA (%)	98.75	98.49	98.32	98.42	98.54	99.74	99.83
AA (%)	95.44	94.97	92.99	96.49	95.04	98.45	99.00
Kappa	98.34	97.98	97.76	97.90	98.05	99.65	99.77

图 4.7 显示了在 Trento 数据集上所有方法的可视化分类结果。可以看出 GSJDD-Net 的分类图上虽然少数类别有噪声，但是绝大多数类别的分类结果比较好，且优于其他方法。虽然 MAHiDFNet 方法的分类图与本文的相似但是能够明显得分看得到在“bulidings”这个类别上，本章的方法在边界保持上要明显优于 MAHiDFNet。CALC 相对于 MAHiDFNet 边缘保持问题有所缓解，主要是由于 CALC 通过耦合对抗学习框架来实现高光谱和图像融合分类，学习不变特征保持，但是仍然存在边缘错误分类问题。GSJDD-Net 通过空间几何引导空谱特征几何细化，进一步使用图卷积保持集合空间位置信息，能够有效补充高光谱图像空间分辨率导致的邻近像元效应，在一定程度上缓解边缘分类容易出错的问题。

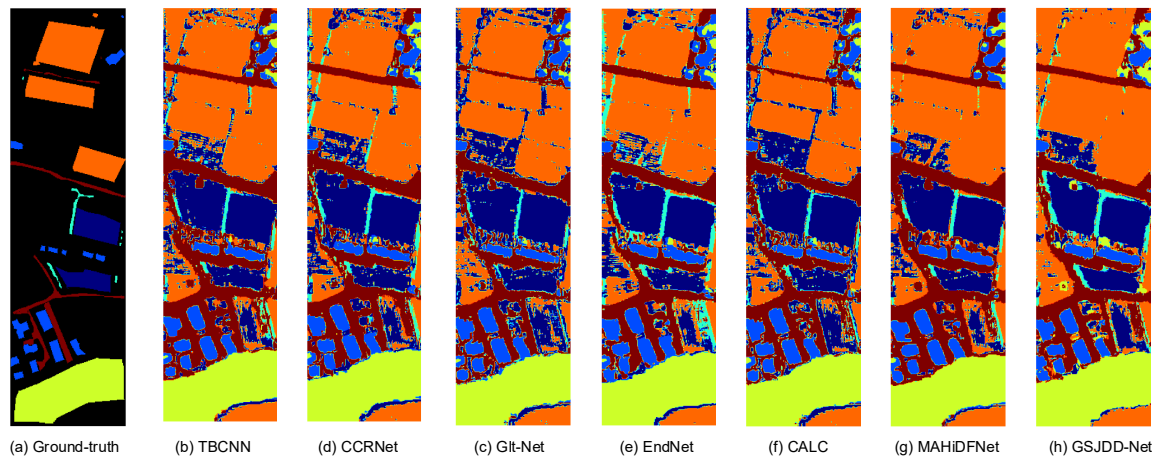


图 4.7 Trento 数据集上的分类可视化图

Fig. 4.7 Classification visualization map on the Trento dataset

4.4.5 消融实验

为了验证所提出的 Sobel 算子、APF 模块和 DRC 模块的有效性，在 GSJDD-Net 上通过移除特定组件进行消融实验以进行比较。表 4.5 显示了在 Trento 数据集上的消融实验，当 GSJDD-Net 移除 Sobel 算子时 OA、AA 和 Kappa 性能分别下降 1.51%、1.97%和 1.62%，表明其利用边缘检测算子建模植被斑块的几何轮廓，通过多尺度梯度特征增强植被边界感知的能力；移除 APF 模块时 OA、AA 和 Kappa 分别降低至 96.33%、95.15%和 96.42%，这可能是 APF 模块采用图卷积网络构建植被斑块拓扑关联，通过两级渐进融合策略实现高光谱与 LiDAR 特征的高效交互。缺乏该模块会导致多模态特征融合过程中几何-光谱信息失配，尤其在破碎化植被区域产生特征混淆。当采用普通卷积替代 DRC 模块时，OA、AA 和 Kappa 下降 3.31%、3.59%和 3.45%，表明动态路由机制通过自适应分类器参数显著提升了复杂地形下的分类鲁棒性。当完整采用所有模块时，GSJDD-Net 在 OA、AA 和 Kappa 指标上分别达到 99.83%、99.00%和 99.77%的最优结果，充分验证了各模块在滨海湿地植被分类任务中的必要性和互补性。

表 4.5 GSJDD-Net 消融实验的精度
Table 4.5 Ablation evaluation accuracies of GSJDD-Net

	Sobel	APF	DRC	OA	AA	$\kappa \times 100$
1	×	√	√	98.32	97.03	98.15
2	√	×	√	96.33	95.15	96.42
3	√	√	×	96.52	95.41	96.32
4	√	√	√	<u>99.83</u>	<u>99.00</u>	<u>99.77</u>

4.4.6 几何引导注意力有效性评估评估

为了进一步验证几何引导注意力的有效性，本文在 Trento 数据集上对比了采用与未采用几何引导注意力两种情况下的混淆矩阵可视化结果，分别如图 4.10(a) 和图 4.10(b) 所示。在混淆矩阵中，纵轴表示真实类别标签，横轴表示预测类别标签，对角线元素反映了预测类别与真实类别匹配的概率。对比分析表明，相较于图 4.10(a) 中未使用几何引导注意力的分类结果，图 4.10(b) 采用几何引导注意力后，错误分类数量显著减少，分类精度得到提升。这一结果直观地验证了几何引导注意力在优化分类性能方面的有效性，并进一步证明了其在几何-空谱联合动态形变框架中的关键作用。

(b)

图 4.7 混淆矩阵可视化结果 (a)未使用几何引导注意力(b)使用几何引导注意力

Fig. 4.7 Confusion matrix visualization results (a) Without geometry-guided attention (b) With geometry-guided attention

4.5 本章小结 (Summary)

本章提出几何-光谱联合动态形变网络 (GSJDD-Net)，以应对滨海湿地植被精细分类中高光谱与 LiDAR 数据融合的挑战。同时，基于黄河口滨海湿地典型植被斑块与地形梯度样本，构建了高光谱-LiDAR 多模态怪柳数据集，GSJDD-Net 通过可变形多模态编码、几何引导注意、渐进融合策略和动态路由分类模块构建多维协同机制。可变形编码器自适应捕捉植被边界，并结合多尺度几何增强单元优化地形特征。几何引导注意模块利用 LiDAR 地形梯度优化光谱特征，抑制地形起伏导致的光谱混淆。渐进融合模块采用门控加权策略平衡多模态贡献，并结合图卷积网络 (GCN) 强化植被斑块的拓扑一致性。动态路由分类模块基于专家选择机制匹配分类器参数与特征复杂度，提高分类精准度。实验表明，GSJDD-Net 在黄河口滨海湿地多模态数据集和通用多模态数据及上显著提升分类精度，尤其在光谱渐变区和边界模糊区表现优越，为湿地植被制图和多源遥感数据融合提供了新思路。

5 总结与展望

5 Summary and prospect

5.1 总结 (Summary)

滨海湿地植被的精细制图对于生态保护与资源管理具有重要意义。滨海湿地的生态环境复杂,植被类型丰富,然而,受潮汐、水文、气候等因素的影响,滨海湿地植被精细制图任务在该场景下存在显著的光谱变异性与类别混淆问题,主要表现为:

(1) 同物异谱与同谱异物现象加剧了类内光谱变异与类间光谱相似的问题,导致传统分类方法难以有效建模;(2) 滨海湿地植被呈现破碎化和斑块状分布,边界模糊,进一步增加了类别判别的难度。此外,由于滨海湿地遥感数据的标注成本较高,有限的人工标注样本与高维光谱特征之间的不平衡,使得现有深度学习方法泛化能力受限,难以在复杂的湿地环境中实现高精度分类。针对滨海湿地植被精细制图的上述问题,本文的主要贡献总结如下:

(1) 特征解耦与原型学习的滨海湿地精细制图方法

该方法提出互信息解耦原型对齐网络(MID-ProNet),通过空谱特征分支与域特征分支的双通道架构,实现植被类别特征的显式建模与干扰特征的提取。采用互信息最小化准则实现特征解耦,并通过动量更新机制优化类别原型,增强分类特征的可解释性与泛化能力。通过在黄河口滨海湿地高光谱图像和通用高光谱图像上进行实验,证明了提出方法的有效性。

(2) 多模态数据融合的滨海湿地精细制图方法

该方法提出几何-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、几何引导光谱注意力模块及自适应渐进融合机制,实现光谱-几何特征的多层次协同。通过动态路由分类模块可以有效区分不同种类的地物,实现滨海湿地的精细分类。实验表明了基于多模态数据融合的方法在复杂数据上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)

滨海湿地的精细制图仍然是一项具有挑战性的工作。尽管本文提出的基于特征解耦与多模态数据融合的方法在提升分类精度方面取得了显著成效,但在实际应用中仍存在一些局限性,需要进一步研究和改进。针对当前方法存在的不足,未来的研究方向可以从以下几个方面展开:

(1) 滨海湿地的生态环境复杂多变,不同季节、不同年份的光谱特征存在一定差异,这使得模型在跨时空数据上的泛化能力仍然有限。未来研究可考虑引入多源多时相遥感数据,通过时序分析和自监督学习方法,增强模型对长期动态变化的适应

性。同时，可结合物候学特征，构建基于多时相数据的动态原型库，以提高模型对滨海湿地植被长期演变的理解能力。

（2）针对高精度植被分类，滨海湿地的植被类型复杂，受潮汐、水动力及人为干扰等因素影响较大，传统遥感影像的时间和空间分辨率难以满足精细分类需求。未来可以结合无人机-卫星协同观测策略，利用无人机影像的高空间分辨率优势，弥补卫星遥感影像的局部细节不足，并构建多尺度植被分类体系。此外，探索边缘计算与实时监测技术，在无人机平台上集成轻量级深度学习模型，实现滨海湿地植被的实时分类与动态监测，提高数据处理效率，为湿地保护与管理提供更精准的技术支持。

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] Assessment ME. *Ecosystems and human well-being: wetlands and water*. Washington, DC: World resources institute, 2005.
- [3] 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.
- [4] 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): 32-3958.
- [5] Nicholls RJ. Coastal flooding and wetland loss in the 21st century: changes under the SRES climate and socio-economic scenarios. *Global Environmental Change*, 2004, 14(1): 69—86.
- [6] Davidson NC. How much wetland has the world lost? Long-term and recent trends in global wetland area. *Marine and Freshwater Research*, 2014, 65(10): 934—941
- [7] 牛振国, 张海英, 王显威, 等. 1978—2008年中国湿地类型变化. *科学通报*, 2012, 57(16): 1400—1411.
- [8] 韩月, 柯樱海*, 王展鹏, 梁德印, 周德民.(2023). 资源一号 02D 卫星高光谱数据黄河三角洲湿地景观分类, *遥感学报* 27: 1387–1399.
- [9] 钟燕飞;吴浩;刘寅贺;湿地遥感制图研究现状与展望[J];中国科学基金;2022年03期
- [10] 李建平, 张柏, 张冷, 等. 湿地遥感监测研究现状与展望. *地理科学进展*, 2007, 26(1): 33—43.
- [11] T. Liu and A. E. Amr, Multi-view object-based classification of wetland land covers using unmanned aircraft system images [J], *Remote Sensing of Environment*, 2018, 216(15):122-138.
- [12] H. Xu, M. E. Hodgson, S. E. Piovan et al., The potential of using LiDAR and color-infrared aerial imagery for palustrine wetland typology and change [J], *Gience Remote Sensing*, 2018, 55(4): 477-501.
- [13] 张磊,宫兆宁,王启为等. 2019.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*,23(2):313-326
- [14] 丛毓, 邹元春, 吕宪国, 等. 湿地资源调查与湿地监测的比较研究. *湿地科学*, 2021, 19(3): 277—28
- [15] 张兵.2016.高光谱图像处理与信息提取前沿.遥感学报,20(5):1062-1090
- [16] Lefsky M A , Cohen W B , Parker G G ,et al.Lidar Remote Sensing for Ecosystem Studies

- Lidar, an emerging remote sensing technology that directly measures the three-dimensional distribution of plant canopies, can accurately estimate vegetation structural attributes and should be of particular interest[J]. *Psychological Reports*, 2002, 46(1):927-930.DOI:10.2466/pr0.1980.46.3.927.
- [17] MAO D H, WANG Z M, DU B J, et al. National wetland mapping in China: a new product resulting from object-based and hierarchical classification of Landsat 8 OLI images[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, 164: 11-25.
- [18] HU T G, LIU J H, ZHENG G, et al. Quantitative assessment of urban wetland dynamics using high spatial resolution satellite imagery between 2000 and 2013[J]. *Scientific Reports*, 2018, 8: 7409.
- [19] 谢静, 王宗明, 毛德华, 等. 基于面向对象方法和多时相HJ-1影像的湿地遥感分类——以完达山以北三江平原为例[J]. *湿地科学*, 2012, 10(4): 429-438.
- [20] WANG X X, XIAO X M, ZOU Z H, et al. Mapping coastal wetlands of China using time series Landsat images in 2018 and Google Earth Engine[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, 163: 312-326.
- [21] MAHDIANPARI M, SALEHI B, MOHAMMADIMANESH F, et al. The first wetland inventory map of Newfoundland at a spatial resolution of 10 m using sentinel-1 and sentinel-2 data on the google earth engine cloud computing platform[J]. *Remote Sensing*, 2018, 11(1): 43.
- [22] CAO J J, LENG W C, LIU K, et al. Object-based mangrove species classification using unmanned aerial vehicle hyperspectral images and digital surface models[J]. *Remote Sensing*, 2018, 10(1): 89.
- [23] GONG P, NIU Z G, CHENG X, et al. China's wetland change (1990-2000) determined by remote sensing[J]. *Science China Earth Sciences*, 2010, 53(7): 1036-1042.
- [24] GONG P, LIU H, ZHANG M, et al. Stable classification with limited sample: transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017[J]. *Science Bulletin*, 2019, 64(6): 370-373.
- [25] BUNTING P, ROSENQVIST A, LUCAS R, et al. The global mangrove watch—A new 2010 global baseline of mangrove extent[J]. *Remote Sensing*, 2018, 10(10): 1669.
- [26] Fu J, et al. Dual attention network for scene segmentation[C]. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019.
- [27] HU Y B, ZHANG J, MA Y, et al. Hyperspectral coastal wetland classification based on a multi-object convolutional neural network model and decision fusion[J]. *IEEE Geoscience and Remote Sensing Letters*, 2019, 16(7): 1110-1114.

- [28] 朱玉玲, 王建步, 王安东, 等. 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.
- [29] 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.
- [30] 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.
- [31] 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.
- [32] 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.
- [33] 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.
- [34] LeCun Y, et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998.
- [35] Krizhevsky A, et al. ImageNet classification with deep convolutional neural networks. NeurIPS 2012.
- [36] Posner M I. Orienting of attention[J]. Quarterly journal of experimental psychology, 1980, 32(1): 3-25.
- [37] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint arXiv:1409.0473, 2014.
- [38] Vaswani A, et al. Attention is all you need[C]. Advances in neural information processing systems, 2017.
- [39] Devlin J, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.
- [40] Radford A, et al. Language models are unsupervised multitask learners[J]. OpenAI blog, 2019, 1(8): 9.
- [41] Hu J, et al. Squeeze-and-excitation networks[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018.
- [42] Wang X, et al. Non-local neural networks[C]. Proceedings of the IEEE conference on computer

- vision and pattern recognition, 2018.
- [43] Woo S, et al. Cbam: Convolutional block attention module[C]. Proceedings of the European conference on computer vision, 2018.
- [44] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc., Red Hook, NY, USA, 6000–6010.
- [45] Bengio, Y., Courville, A., Vincent, P.: Representation learning: A review and new perspectives. IEEE transactions on pattern analysis and machine intelligence 35(8), 1798–1828
- [46] Huang, X., Liu, M.Y., Belongie, S., Kautz, J.: Multimodal unsupervised image-to-image translation. In: Proceedings of the European conference on computer vision (ECCV). pp. 172–189
- [47] Lee, H.Y., Tseng, H.Y., Huang, J.B., Singh, M., Yang, M.H.: Diverse image-to-image translation via disentangled representations. In: Proceedings of the European conference on computer vision (ECCV). pp. 35–51 (2018)
- [48] Ning, M., Bian, C., Wei, D., Yu, S., Yuan, C., Wang, Y., Guo, Y., Ma, K., Zheng, Y.: A new bidirectional unsupervised domain adaptation segmentation framework. In: IPMI. pp. 492–503. Springer (2021)
- [49] Zhao, Q., Lyu, S., Bai, W., Cai, L., Liu, B., Wu, M., Sang, X., Yang, M., Chen, L.: A multi-modality ovarian tumor ultrasound image dataset for unsupervised cross-domain semantic segmentation. arXiv preprint arXiv:2207.06799 (2022)
- [50] Lezama, J.: Overcoming the disentanglement vs reconstruction trade-off via jacobian supervision. In: International Conference on Learning Representations (2018)
- [51] Kraskov, A., Stögbauer, H., Grassberger, P.: Estimating mutual information. Physical review E 69(6), 066138 (2004)
- [52] Cha, J., Lee, K., Park, S., Chun, S.: Domain generalization by mutual-information regularization with pre-trained models. In: European Conference on Computer Vision. pp. 440–457. Springer (2022)
- [53] Liu, X., Yang, C., You, J., Kuo, C.C.J., Kumar, B.V.: Mutual information regularized feature-level frankenstein for discriminative recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 44(9), 5243–5260 (2021)
- [54] Peng, X., Huang, Z., Sun, X., Saenko, K.: Domain agnostic learning with disentangled representations. In: International Conference on Machine Learning. pp. 5102–5112. PMLR (2019)

- [55] Meng, Q., Matthew, J., Zimmer, V.A., Gomez, A., Lloyd, D.F., Rueckert, D., Kainz, B.: Mutual information-based disentangled neural networks for classifying unseen categories in different domains: Application to fetal ultrasound imaging. *IEEE Trans. Med. Imag.* 40(2), 722–734 (2020)
- [56] Chen, C., Dou, Q., Chen, H., Qin, J., Heng, P.A.: Unsupervised bidirectional crossmodality adaptation via deeply synergistic image and feature alignment for medical image segmentation. *IEEE Trans. Med. Imag.* 39(7), 2494–2505 (2020)
- [57] AGNAR AAMODT, ENRIC PLAZA. Case-based reasoning: Foundational issues, methodological variations, and system approaches[J]. *AI Communications*, 1994, 7(1): 39-59.
- [58] NEWELL A, SIMON H A. Human problem solving[M]. Prentice Hall, 1972.
- [59] BISHOP C M. Pattern recognition[M]//Machine learning. 2006: 128(9).
- [60] DUDA R O, HART P E. Pattern classification and scene analysis[M]. New York: Wiley, 1973.
- [61] FRIEDMAN J H. The elements of statistical learning: Data mining, inference, and prediction[M]. Springer, 2009.
- [62] COVER T, HART P. Nearest neighbor pattern classification[J]. *IEEE Transactions on Information Theory*, 1967, 13(1): 21-27.
- [63] KOHONEN T. The self-organizing map[J]. *Neurocomputing*, 1998, 21(1-3): 1-6.
- [64] SATO A, YAMADA K. Generalized learning vector quantization[C]//NeurIPS. 1995.
- [65] GOLDBERGER J, HINTON G E, ROWEIS S, et al. Neighbourhood components analysis[C]//NeurIPS. 2004.
- [66] SNELL J, SWERSKY K, ZEMEL R S. Prototypical networks for few-shot learning[C]//NeurIPS. 2017.
- [67] JETLEY S, ROMERA-PAREDES B, JAYASUMANA S, et al. Prototypical priors: From improving classification to zero-shot learning[EB/OL]. arXiv:1512.01192, 2015.
- [68] WU Z, XIONG Y, YU S X, et al. Unsupervised feature learning via non-parametric instance discrimination[C]//CVPR. 2018.
- [69] METTES P, VAN DER POL E, SNOEK C. Hyperspherical prototype networks[C]//NeurIPS. 2019.
- [70] LI O, LIU H, CHEN C, et al. Deep learning for case-based reasoning through prototypes: A neural network that explains its predictions[C]//AAAI. 2018.
- [71] Kipf T N, Welling M. Semi-Supervised Classification with Graph Convolutional NetworksJ/OL. *International Conference on Learning Representations (ICLR)*, 2017. (2016-09-09)2024-12-28. <https://arxiv.org/abs/1609.02907>
- [72] Jacobs R A, Jordan M I, Nowlan S J, et al. Adaptive mixtures of local expertsJ. *Neural*

- computation, 1991, 3(1): 79-87.
- [73] SHAZEER N, MIRHOSEINI A, MAZIARZ K, 等. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer[C]// Proceedings of the International Conference on Learning Representations (ICLR). 2017.
- [74] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 8, pp. 1778–1790, 2004.
- [75] Z. Zhong, J. Li, Z. Luo, and M. Chapman, "Spectral–spatial residual network for hyperspectral image classification: A 3-d deep learning framework," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 2, pp. 847–858, 2018.
- [76] S. K. Roy, G. Krishna, S. R. Dubey, and B. B. Chaudhuri, "Hybridsn: Exploring 3-d–2-d cnn feature hierarchy for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 2, pp. 277–281, 2020.
- [77] Li, Rui, Shunyi Zheng, Chenxi Duan, Yang Yang, and Xiqi Wang. Classification of Hyperspectral Image Based on Double-Branch Dual-Attention Mechanism Network[J]. *Remote Sensing*, 2020, 12(3): 582. DOI: 10.3390/rs12030582.
- [78] Sun, L., Zhao, G., Zheng, Y., & Wu, Z. Spectral–Spatial Feature Tokenization Transformer for Hyperspectral Image Classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 1-14. DOI: 10.1109/TGRS.2022.3144158
- [79] Alkhatib, M. Q., & Jamali, A. HSIFormer: An Efficient Vision Transformer Framework for Enhanced Hyperspectral Image Classification Using Local Window Attention[C]//2024 14th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS). Helsinki, Finland, 2024: 1-5. DOI: 10.1109/WHISPERS65427.2024.10876533.
- [80] Xu, X., Li, W., Ran, Q., Du, Q., Gao, L., & Zhang, B. Multisource Remote Sensing Data Classification Based on Convolutional Neural Network[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(2): 937-949
- [81] Wu, X., Hong, D., & Chanussot, J. Convolutional Neural Networks for Multimodal Remote Sensing Data Classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 1–10.
- [82] Hong, D., Gao, L., Hang, R., Zhang, B., & Chanussot, J. Deep Encoder–Decoder Networks for Classification of Hyperspectral and LiDAR Data[J]. *IEEE Geosci. Remote Sens. Lett.*, 2022, 19: 1–5.
- [83] Ding, K., Lu, T., Fu, W., Li, S., & Ma, F. Global–Local Transformer Network for HSI and LiDAR Data Joint Classification[J]. *IEEE Trans. Geosci. Remote Sens.*, 2022, 60: 1–13.

-
- [84] Lu, T., Ding, K., Fu, W., Li, S., & Guo, A. Coupled Adversarial Learning for Fusion Classification of Hyperspectral and LiDAR Data[J]. Inf. Fusion, 2023, 93: 118–131.
- [85] Wang, X., Feng, Y., Song, R., Mu, Z., & Song, C. Multi-attentive Hierarchical Dense Fusion Net for Fusion Classification of Hyperspectral and LiDAR Data[J]. Inf. Fusion, 2022, 82: 1–18.