

华北水利水电大学

North China University of Water Resources and Electric Power

毕业论文

题目：基于跨分辨率对比聚类视觉特征学习的 SAR 地物分类

学 院	信息工程学院
专 业	人工智能
姓 名	高树林
学 号	202018526
指导老师	杨阳蕊、王畅
完成时间	2024 年 6 月

独立完成与诚信声明

本人郑重声明：所提交的毕业设计（论文）是本人在指导教师的指导下，独立工作所取得的成果并撰写完成的，郑重确认没有剽窃、抄袭等违反学术道德、学术规范的侵权行为。文中除已经标注引用的内容外，不包含其他人或集体已经发表或撰写过的研究成果。对本文的研究做出重要贡献的个人和集体，均已在文中作了明确的说明并表示了谢意。本人完全意识到本声明的法律后果由本人承担。

毕业设计（论文）作者签名：

指导导师签名：

签字日期：

签字日期：

毕业设计（论文）版权使用授权书

本人完全了解华北水利水电大学有关保管、使用毕业设计(论文)的规定。特授权华北水利水电大学可以将毕业设计(论文)的全部或部分内容公开和编入有关数据库提供检索，并采用影印、缩印或扫描等复制手段复制、保存、汇编以供查阅和借阅。同意学校向国家有关部门或机构递交毕业设计(论文)原件或复印件和电子文档(涉密的成果在解密后应遵守此规定)。

毕业设计(论文)作者签名:

导师签名:

签字日期:

签字日期:

目录

摘要	I
ABSTRACT	III
第1章 绪论	1
1.1 研究背景	1
1.2 国内外研究现状	1
1.2.1 SAR 地物分类研究现状	1
1.2.2 对比学习研究现状	4
1.3 本文主要工作与结构安排	4
第2章 相关理论概述	7
2.1 SAR 图像介绍	7
2.1.1 SAR 图像成像原理	7
2.1.2 SAR 图像特征	8
2.2 机器学习介绍	9
2.2.1 传统学习与深度学习	9
2.2.2 监督学习和无监督学习	10
2.3 块分类介绍	12
2.4 块分类任务评价指标	13
2.5 分割任务评价指标	14
2.6 本章小结	16
第3章 跨分辨率对比学习地物分类的方法	19
3.1 SAR 数据处理	19
3.1.1 裁剪块的确定	19
3.1.2 数据裁剪	21
3.1.3 SAR 数据集分割	23
3.1.4 SAR 数据增强	23
3.1.5 跨分辨率 SAR 图像对	24
3.2 模型设计	25
3.3.1 整体流程图	25
3.3.2 代理任务	25
3.3.3 下游任务	27
3.3 分割任务	30
3.3.1 块分类粒度	30
3.3.2 条件随机场	31
3.4 训练设置	32
3.5 本章小结	33
第4章 实验结果与分析	35
4.1 分类任务结果和分析	35
4.1.1 裁剪块尺寸	35
4.1.2 下游任务网络层数	38
4.2 分割任务结果和分析	38
4.2.1 块分类粒度	38
4.2.2 全连接条件随机场	40

4.3 消融实验	41
4.4 和有监督的模型的比较	42
4.5 预训练分割任务结果分析	46
4.6 本章小结	47
第 5 章 总结与展望	49
5.1 本文工作总结	49
5.2 工作展望	50
参考文献	51
致 谢	57

基于跨分辨率对比聚类视觉特征学习的 SAR 地物分类

摘要

合成孔径雷达（SAR）是一种能够生成高分辨率图像的雷达，其成像过程不受天气等自然条件的干扰。SAR 图像地物分类是 SAR 图像解译过程中的关键环节，尽管近年来基于深度学习的技术在这一领域取得了显著进展，SAR 图像地物分类仍面临诸多挑战，包括高分辨率导致的大数据量、地物类型复杂多样、以及高昂的标注成本。针对这些挑战，本文提出了一种两阶段的基于跨分辨率对比学习的自监督方法 SWAV-ResSegNet，该方法包括代理任务和下游任务两个阶段。

代理任务通过在线聚类和交换预测机制进行特征提取过程。它结合不同分辨率下同一区域内的图像并使用数据增强创建图像的不同视图，然后提取特征表示，并将其映射到一组可训练的原型向量上形成聚类编码。代理任务的核心在于利用这些聚类编码预测其他视图的编码，而非直接比较特征。代理任务鼓励网络学习区分不同图像的特征表示，同时确保同一图像的不同视图在特征空间中彼此接近，而不同图像的视图则彼此远离，进而够捕捉图像本质特征的表示，而无需依赖于显式的标签信息。下游任务将代理任务学到的特征应用于 SAR 图像地物分类。利用迁移学习，这些特征被用于设计深度学习分类器并进行微调，以提升模型性能。最终利用少量标签就能达到接近于有监督学习的水平。

本文通过多次实验确定图像裁剪块的最佳尺寸和分割块的最佳粒度，同时引入全连接条件随机场来更好地对图像进行后处理并用实验数据证明了该方法的可行性。此外本文还简要介绍了下游任务为分割任务的网络 SWRU-Net，并且在极少量数据下达到和有监督预训练模型 Unet 相接近的 Miou 和 PA 效果。本文还介绍了基于 SWAV-ResSegNet、SWRU-Net、Unet、Deeplabv3 四种模型开发的实时 SAR 图像解译系统，能够自动化解译分割结果和生成检测报告。

本文的工作不仅在理论上证明了自监督学习在 SAR 图像处理中的有效性，而且在实际应用中也提供了一个基于 SAR 地物分类算法的实时分割系统。

关键词：SAR 图像， 地物分类， 自监督学习， 对比学习， 跨分辨率， 聚类

SAR Land Cover Classification Based on Cross-Resolution Contrastive Clustering Visual Feature Learning

ABSTRACT

Synthetic aperture radar (SAR) is a kind of radar that is capable of generating high-resolution images, and its imaging process is not interfered by natural conditions such as weather, etc. The feature classification of SAR images is a key link in the process of SAR image interpretation, and despite the significant progress made by deep-learning-based techniques in this field in recent years, the interpretation of SAR images still faces many challenges, including large data volume due to high resolution, complexity and variety of feature types, and high labeling cost. data volume due to high resolution, complexity and diversity of feature types, and high labeling cost. To address these challenges, this paper proposes a two-stage self-supervised learning method SWAV-ResSegNet based on cross-resolution comparative learning, which consists of two stages: the proxy task and the downstream task.

The proxy task simplifies the feature extraction process through online clustering and exchange prediction mechanisms. It combines images within the same region at different resolutions and uses data augmentation to create different views of the image, then extracts feature representations and maps them onto a set of trainable prototype vectors to form a clustering code. The core of the agent task is to use these clustering codes to predict the codes of other views rather than comparing features directly, encouraging the network to learn to differentiate between the feature representations of different images while ensuring that different views of the same image are close to each other in feature space and views of different images are far away from each other, which in turn is enough to capture a representation of the essential features of the image without relying on explicit labeling information. The downstream task applies the features learned by the agent task for SAR image classification. Using migration learning, these features are used to design deep learning classifiers and fine-tune them

to improve model performance. Eventually a level close to supervised learning can be achieved using a small number of labels.

In this paper, we determine the optimal size of the image cropping block and the optimal granularity of the segmentation block through several experiments, and introduce a fully connected conditional random field to better post-process the image and demonstrate the feasibility of the method with experimental data. In addition, this paper also briefly introduces SWRU-Net, a network based on the segmentation task as a downstream task, and achieves Miou and PA results close to the supervised pre-training model Unet with a very small amount of data. This paper also presents a real-time SAR image decoding system developed based on four models, SWAV-ResSegNet, SWRU-Net, Unet, and Deeplabv3, which is capable of automated decoding of segmentation results and generating detection reports.

The work in this paper not only proves the effectiveness of self-supervised learning in SAR image processing theoretically, but also provides a real-time segmentation system based on SAR feature classification algorithm in practical applications.

Key Words: SAR images, Terrain Classification , Self-Supervised Learning , Contrastive Learning, Cross-Resolution, Clustering

第1章 绪论

1.1 研究背景

合成孔径雷达^[1]（Synthetic Aperture Radar, SAR）是一种利用合成孔径原理实现高分辨率微波成像的主动遥感技术，它具有全天时、全天候的观测能力，不受光照和天气条件的限制，能够提供地表的高分辨率图像，是获取地物信息的一种重要手段^[2-6]。SAR系统通过在飞行平台上发射电磁波并接收从地面目标反射回来的信号来工作。由于雷达自带照射源，它可以在黑夜中出色地工作，而使用的电磁波波段几乎可以无失真地穿透水汽云层。这种特性使得SAR在环境监测^[7-9]、资源调查^[10-13]、军事侦察等^[14-16]多个领域具有广泛的应用价值。

为了更充分利用SAR数据的潜力，需要依赖先进的图像处理技术，因此SAR图像解译^[17]是当前研究的热点，其中SAR图像解译中地物分类是一门值得深入研究的课题。随着与深度学习技术^[18-23]的融合，SAR图像地物分类的自动化特征提取、泛化能力和准确率得到了显著提升，进一步推动了SAR图像解译技术的发展和应用。

然而，由于SAR图像的场景通常非常广阔，且现代SAR系统的成像分辨率日益提高，众多问题^[24]使得获取大量标注样本变得困难和昂贵：高分辨率的SAR图像意味着每个图像都包含大量的像素点，因此即使是相对较小的场景，其图像数据量也可能非常庞大。此外，需要专业知识来准确地标注SAR图像中的不同地物类型，这通常由有丰富行业经验的专家完成，这不仅耗时而且成本高昂。再者，SAR图像中的场景可能包含多种复杂的地物类型和结构，这些可能难以通过自动化工具精确标注。因此如何利用少量标签或者不用标签就能够学习到图像的有效特征至关重要。对比学习（Contrastive Learning）是一种自监督学习方法，它通过比较数据的不同视图来学习数据的表示而无需显式标签。相对于传统的有监督学习方法，利用对比学习对SAR图像进行地物分类具有实际意义。

1.2 国内外研究现状

1.2.1 SAR 地物分类研究现状

当前，SAR地物分类研究正处于快速发展阶段，其研究方法主要分为两种，

传统方法和深度学习方法。

按照不同的处理方式，传统SAR图像地物分类方法可以分为以下几类：

(1) 边缘信息提取、多尺度分割与规则集构建：汪等^[25]通过改进的Canny算子提取影像边缘信息，结合多尺度分割与规则集构建，实现对五类地物的高效分类。该方法需要较高的专业知识和经验，视觉一致性有待提升。

(2) 基于上下文信息增强：符等^[26]提出方向性上下文协方差矩阵(DCCM)方法，增强单通道SAR图像的分类。尽管在特定数据集上取得显著提升，但对SAR图像特定属性依赖性强，泛化能力有待验证。李^[30]引入区域类别模糊概率(RCFP)，提高SAR图像分类精度，但主要依赖图像强度特征，对纹理、颜色和极化特征等信息缺乏考虑。

(3) 极化特征与纹理特征分析：王等^[27]利用极化SAR图像的极化特征与纹理特征，通过LightGBM和SLIC算法进行初级和二级分类，有效抑制相干斑噪声。这种方法在处理复杂特征时可能遇到挑战，并对超像素生成算法和分类模型依赖较强。赵等^[29]通过互易处理和滤波技术降低斑点噪声，利用极化分解方法提取特征，并使用支持向量机进行湿地分类。该方法计算成本较高。

(4) 基于机器学习的分类：魏等^[28]利用随机森林分类器，通过二元特征捕获极化SAR图像中的地物信息，准确率和运行效率高，但对算法参数选择敏感，随机性可能影响结果一致性。

(5) 多特征融合方法：王^[31]利用极化目标分解和数学形态学提取特征，通过核函数和Mercer定理进行特征融合，提高分类准确性，但在处理大规模数据集时计算量较大。

在深度学习方法上，赵等^[32]通过多尺度融合技术增强了图像的特征信息量，并降低了噪声干扰，最后将处理后的数据输入到DeepLabV3网络结构中进行训练和分类。但是该方法使用了特定的极化分解方法和小波变换技术，该方法可能在处理其他类型的SAR图像或不同极化特征时需要调整参数或方法。谢等^[33]提出了一种融合深度可分离卷积和多尺度残差结构的UNet网络——MSR-Unet，在小样本数据集上表现出更高的分类准确率和鲁棒性。然而，这种方法在多样化环境下的泛化能力不能得到保证，此外该方法强烈依赖大量标注数据。刘等^[34]通过调整卷积核大小、减少全连接层参数、加入池化层有针对性地改进AlexNet网络结

构，相较原始AlexNet得到一定提升，但这种提升是具有一定限度的，且易受数据影响。楚等^[35]提出了一种结合人工特征工程与深度神经网络的遥感影像地物分类方法。通过提取纹理、结构等多尺度特征图并结合原始图像，构建高维输入提升分类准确度。改进DeepLab v3 网络结构实现端到端分类。该方法能有效提升分类精度和抗干扰性，但是该方法涉及到人工提取特征，需要一定的专业知识。刘等^[36]提出了一种结合随机森林特征选择和超像素分割的极化SAR图像分类方法。通过随机森林降低特征维度，超像素分割校正边界像素错误，解决了“维度灾难”和噪声干扰问题，提高了极化SAR图像分类的效率和准确性。但该方法仅使用了有限的极化特征参数。孙等^[37]提出了一种基于深度卷积网络的Segnet模型，通过Gamma MAP滤波、地理定标和数据增强减少图像噪声、校正地理定位并增加数据量，但是该方法对噪声敏感，导致结果出现较多噪声。张等^[38]利用深度迁移学习进行遥感影像多标签分类的方法，主要是通过训练不同的ResNet模型并选择最佳模型来识别地表的不同地物。

总体来看，传统方法主要包括基于纹理、统计特征和分类器的手工设计特征提取和分类方法。传统方法往往基于多年研究积累的专业知识，具有较强的理论基础，这些方法在特定场景中往往表现出色、简单且鲁棒，处理高维特征空间能力强。但是需要较高的专业知识和经验。而深度学习技术，特别是卷积神经网络（CNN）在SAR地物分类中取得了显著进展，能够自动学习特征表示并实现更准确的分类结果，减少对专业知识的依赖。然而，这种依赖仅限于在特征选取时不需要专业知识和经验，面对有监督的深度学习算法时，仍然需要大量标注数据进行训练，导致数据获取成本高。此外，在多样化环境下，某些深度学习方法的泛化能力不能得到保证。

为了克服传统方法中手动特征提取的复杂性和深度学习方法对大量标注数据的依赖性，对比学习方法作为一种能够自动学习数据有效表示的自监督学习框架逐渐受到关注。对比学习不需要大量标注数据，而是通过在无监督或自监督环境中学习数据的内部结构，能有效提升模型的特征表示能力，通过对比学习，模型能够学习到更加鲁棒和泛化的特征表示，有助于提高精度和稳定性。特别是对于SAR这种场景复杂性、类别多样性、成像参数多样性的数据，对比学习具有得天独厚的优势，是未来SAR地物分类研究的一个重要方向。

1.2.2 对比学习研究现状

对比学习最早出现可追溯到 1949 年的 Donald Hebb^[40]提出了 Hebb 学习规则，这是对比学习的一个早期形式，强调了相邻神经元在学习过程中的关联性。2015 年，Florian 等^[41]在 FaceNet 中使用三元组损失来学习人脸的深度特征表示，它通过确保来自同一身份的两个图像（正样本）在特征空间中的距离小于来自不同身份的图像（负样本）之间的距离来训练网络，这是对比学习在计算机视觉领域的一个里程碑。其训练过程中使用了一种特殊的损失函数，即三元组损失。之后 Ting Chen 等^[42]人提出了 SimCLR 框架，它通过使用数据增强和对比损失来学习无监督的特征表示，这为未来的无监督和半监督学习研究提供了新的方向，极大地推动了对比学习在深度学习领域的应用。2020 年，He 等^[43]人提出了 MoCo (Momentum Contrast)，通过对比学习来构建一个动态的字典用于在训练过程中进行编码器的查找任务。MoCo 利用队列和移动平均编码器 (momentum encoder) 来构建这个字典，从而实现在无监督学习中的有效对比。Jean-Bastien 等^[44]提出 BYOL，其包括在线网络和目标网络。在线网络负责预测目标网络表示的图像视图。目标网络的参数是在线网络参数的指数移动平均，这为学习过程提供了稳定性。通过这种方式，BYOL 能够避免使用负样本对，从而简化了训练过程，并减少了计算和内存需求。Mathilde 等^[45]提出了一种新的无监督学习方法 SwAV，用于学习视觉特征。它通过对比学习框架来学习数据的有效表示，而无需依赖于有标签的监督信息。此外文章还提到了一种数据增强策略，即多裁剪。它从同一图像中提取多个裁剪，这些裁剪作为独立的视图输入到模型中。这种方法可以帮助模型学习到更加鲁棒和区分性的特征表示，因为它迫使模型从不同的视角理解和表示同一物体或场景。

1.3 本文主要工作与结构安排

本文旨在探索合成孔径雷达图像地物分类的新方法，特别是针对 SAR 图像的高分辨率、多视角特性以及标注数据稀缺的问题。为了实现这一目标，本文提出了一种基于跨分辨率对比学习的自监督学习方法，以优化特征提取过程并提升分类性能。本文的主要工作和结构安排如下：

第一章详细介绍了 SAR 图像地物分类的重要性和应用背景。SAR 图像由于其

全天候、全天时的成像能力，广泛应用于环境监测、农业调查、城市规划等领域。然而，SAR图像数据的高分辨率和多视角特性带来了巨大的数据处理量，增加了计算复杂度和存储需求。此外，地物类型的复杂性，如农田、水体、城市建筑等，增加了分类的难度。同时，获取高质量的标注数据成本高昂且时间耗费巨大，进一步限制了有监督学习方法的应用。因此，现有的SAR图像分类方法在处理这些挑战时表现出一定的局限性。

第二章提供了本文的理论基础，涵盖SAR图像的成像原理和特性，以及机器学习和深度学习在图像处理中的应用，还介绍了分类和分割任务的评价指标以及块分类方法，为后续的研究方法和实验设计提供了理论基础。

第三章主要介绍了本文的实验方法，包括数据的处理、模型的设计、实验参数的设置。在特征学习方面，采用对比聚类方法，通过构建正负样本对来学习有区分度的特征表示，增强同类地物特征的一致性和不同类地物特征的区分性。这种方法通过对比学习和聚类技术的结合，能够更好地捕捉SAR图像中的重要特征。基于此，该章构建了一种新的深度学习分类模型，并设计有效的训练策略和损失函数，以优化模型性能。

第四章主要展示了实验结果，并分析不同粒度、不同裁剪尺寸、不同分辨率、是否使用后处理方法的结果和相应指标，并且和利用当前主流深度学习网络得到的分割结果做了比较，此外还做了消融实验以探究传统的数据增强对SAR的训练是否有正向作用，并将本文提出的两种模型和有监督图像分割模型进行了对比。

第五章是针对本文的总结和未来工作的展望。该部分总结了本文方法论部分包括数据增强对SAR图像的影响、以及自监督未来在SAR图像的地物分类研究的潜在价值和重要性。同时对本文实验得到的一些结论进行深化和推广，指出通过上述研究方向的探索，可以进一步提升SAR图像解译技术的自动化水平，拓宽遥感技术的应用范围。本文的结构图如图 1-1 所示。

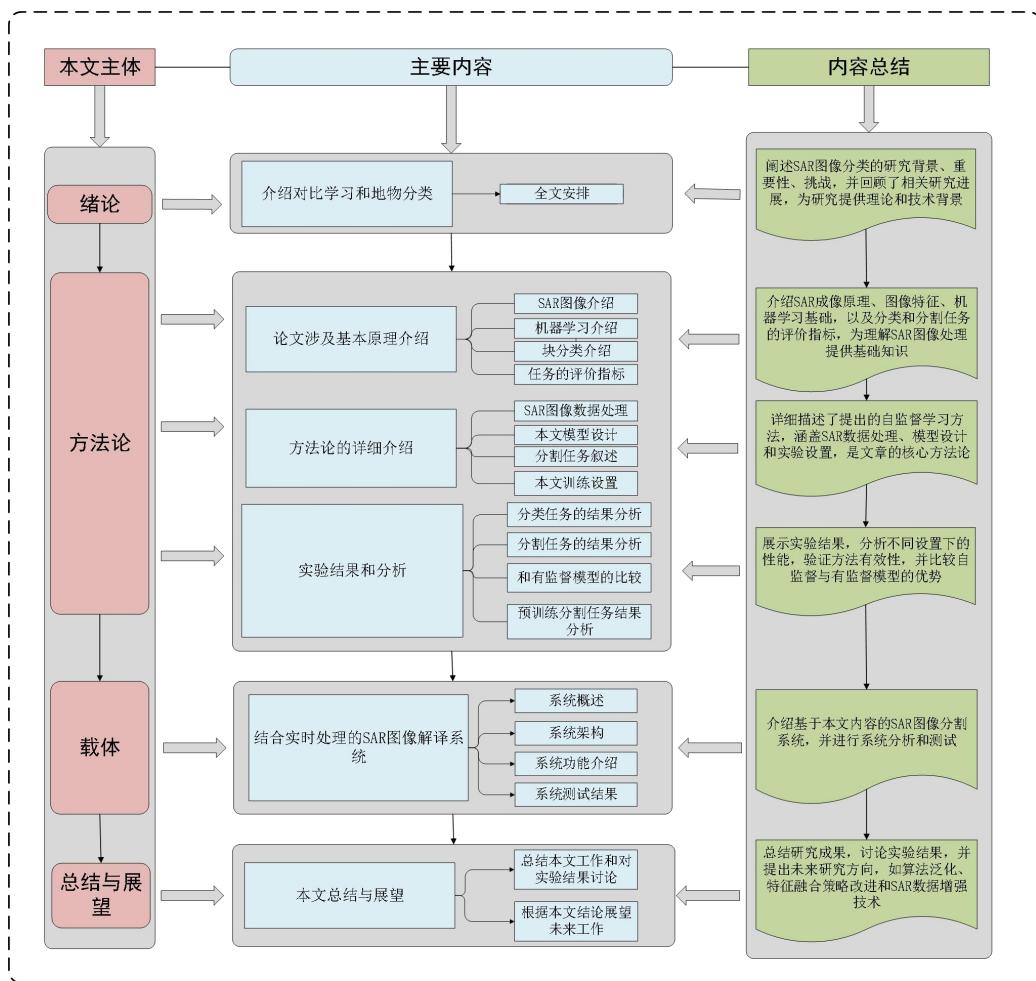


图 1-1 本文结构图

第 2 章 相关理论概述

2.1 SAR 图像介绍

2.1.1 SAR 图像成像原理

SAR系统是一种具备主动探测能力的雷达系统，其工作原理是将雷达搭载于飞行平台如飞机或卫星上，利用合成天线技术和信号的后处理，模拟出一个巨大的天线并侧向发射连续的垂直于航迹方向的电磁波脉冲，这些脉冲在遇到地面对目标时会产生散射并形成后向散射信号，随后雷达天线捕捉这些回波信号的多普勒相位，并存储起来，利用合成孔径技术在飞行过程中积累数据，形成高分辨率的带状影像^[46]。SAR图像的分辨率由距离向分辨率和方位向分辨率共同决定。其中，距离向分辨率取决于雷达系统发射脉冲的持续时间，脉冲越短则分辨率越高；而方位向分辨率则通过合成孔径技术实现，即随着平台的移动，SAR通过在飞行路径上不同位置发射和接收信号，合成一个较大的虚拟天线，从而提高成像分辨率。SAR成像中，天线的相位中心在不同时间位于不同的位置，形成一个合成孔径。合成孔径的长度 L_s 可以表示为：

$$L_s = \frac{v \cdot T}{2} \quad (2-1)$$

其中 v 是平台的速度， T 是合成孔径的时间长度。距离分辨率与发射信号的带宽和雷达天线的长度有关，公式为：

$$\Delta R = \frac{c}{2 \cdot B} \quad (2-2)$$

其中 c 是光速， B 是信号的带宽。方位分辨率与合成孔径的长度有关，公式为：

$$\Delta \theta = \frac{\lambda}{L_s} \quad (2-3)$$

其成像原理如图 2-1 所示。

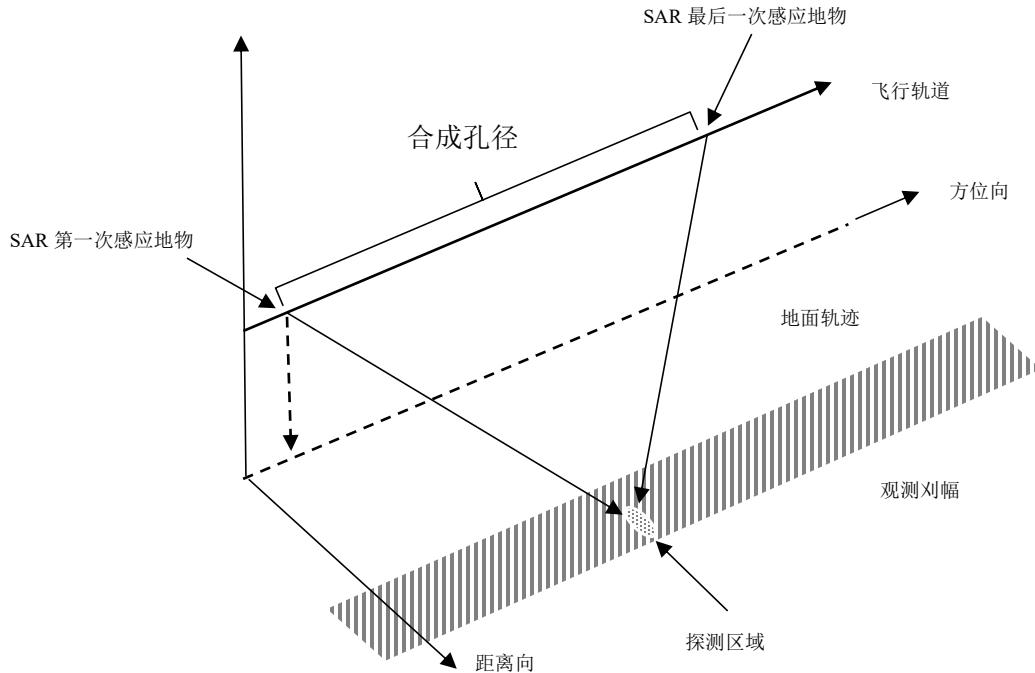


图 2-1 SAR 成像原理图

2.1.2 SAR 图像特征

合成孔径雷达成像和光学遥感图像是两种不同的遥感技术，它们在成像原理、应用领域以及图像特点上存在显著的区别。SAR 利用微波信号进行成像，不依赖于光照条件，能够穿透云层和某些植被，提供高分辨率的灰度图像，而光学遥感依赖于可见光或近红外光的反射，需要光照条件，能够提供彩色图像，反映地表的真实颜色。部分数据集上的 SAR 图像和光学遥感图像如图 2-2~3 所示。

图 2-2 是 Li 等^[61]采集的湖北地区的 SAR 图像和光学遥感图像，明显看出光学遥感图像极易受到天气影响导致清晰度不高，这种现象在图 2-3 在陕西汉中地区采集的光学遥感图像也有所体现，这是由于在多云多雾的天气下，信号传输的过程中发生了瑞利散射，当信号抵达传感器时，信号强度大幅度降低，导致成像模糊不清。反观 SAR 图像则并不会受这种因素影响，由于瑞利散射的严重程度与波长的四次方成反比，较长波长的 SAR 信号受天气影响较晴天相差无几，因此无论在何种条件下都能够生成高分辨率的图像，但是由于成像原理的复杂，SAR 图像只能呈现出灰度图像，正因如此，SAR 图像解译是现代图像解译的热点和难点。

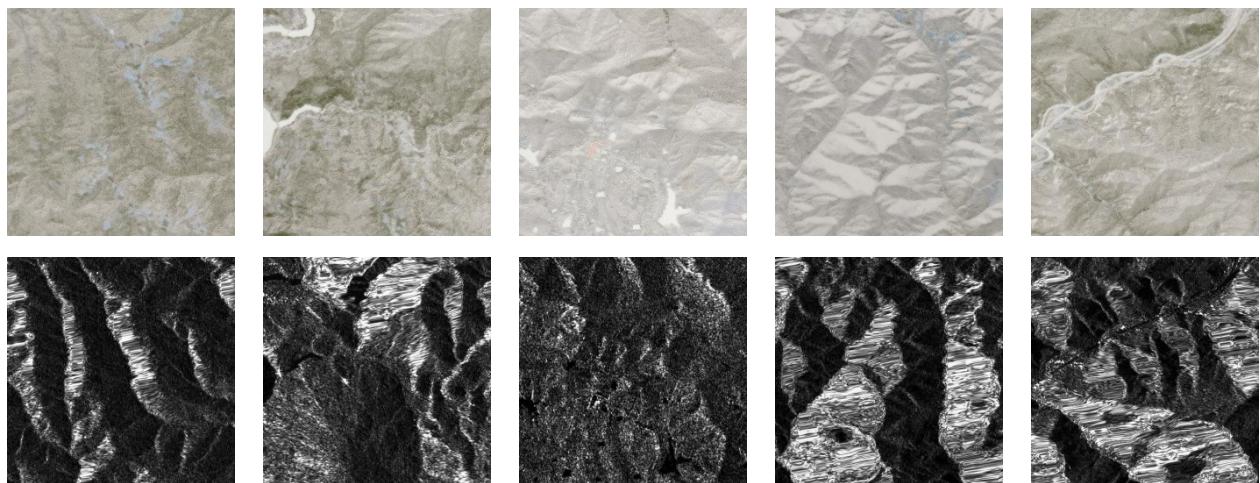


图 2-2 whu-opt-sar 的光学遥感图像（上）和 SAR 图像（下）

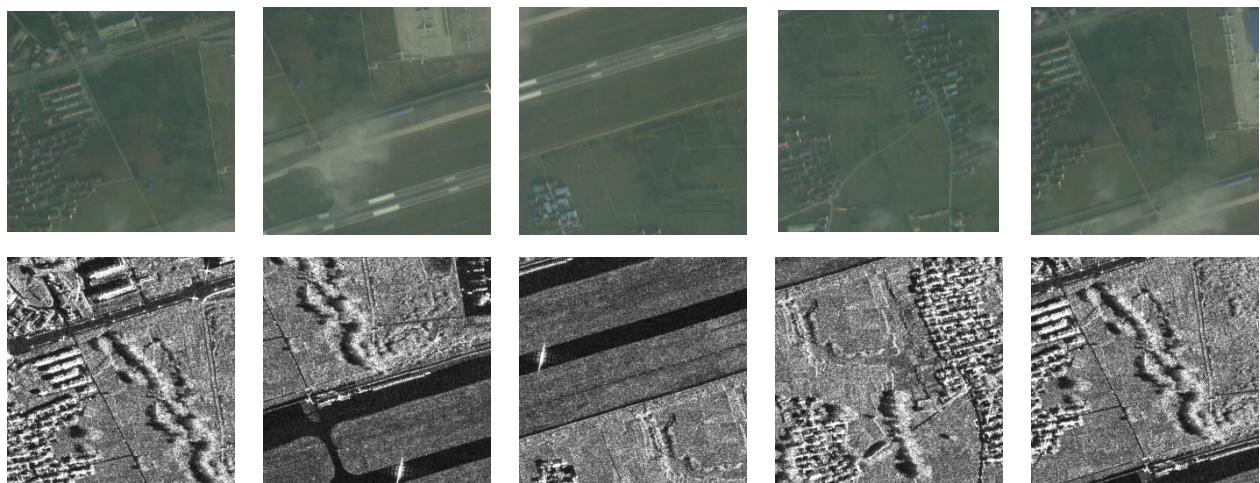


图 2-3 汉中地区 1m 分辨率的光学遥感图像（上）和 SAR 图像（下）

2.2 机器学习介绍

2.2.1 传统学习与深度学习

机器学习是人工智能的一个重要分支，其目标是通过从数据中学习模式和规律来使计算机系统具有智能化能力，而不需要明确地进行编程。机器学习可以分为传统机器学习和深度学习两大类。

传统机器学习是人工智能领域中较早发展起来的一系列算法和技术，在深度学习被广泛应用之前，就已经被广泛地应用于各种数据挖掘和预测任务。传统机器学习侧重于使用相对简单的数学模型来识别数据中的模式和关系。这些模型通常包括参数，这些参数从数据中学习得到，以便进行预测或分类。在深度学习之

前，一般使用决策树、支持向量机（SVM）等算法分析医学图像和临床数据，进行疾病预测和诊断；利用随机森林（RF）、梯度提升机（GBM）等集成学习算法评估贷款审批的业务风险，检测欺诈行为；通过朴素贝叶斯等算法构建聊天机器人，做文本分类、垃圾邮件过滤、情感分析等应用；利用隐马尔可夫模型（HMM）和高斯混合模型（GMM）进行语音识别等。

然而，传统机器学习需要手工设计特征，这意味着在应用传统机器学习算法之前，需要人工选择和提取最相关的特征。这种过程需要领域专家的知识和经验，并且可能耗费大量时间和精力。而且，手工设计的特征可能无法充分地捕捉数据中的复杂模式和关系，导致模型性能受限。

在传统机器学习方法百花齐放之际 Yang 等^[49] 提出了具有卷积神经网络（Convolutional Neural Network, CNN）结构的 Le-Net，能自动提取特征而不需手工设计或者降维提取特征，这是普遍意义上最早的卷积神经网络，是深度学习领域中的重要里程碑之一。2006 年，Hinton 等^[38] 正式提出了深度学习的概念，并提出逐层初始化方法解决了深度神经网络的梯度消失问题，这一突破性进展极大促进了神经网络和深度学习研究的复兴。

AlexNet^[47]是一个具有划时代意义的深度卷积神经网络，它提出了 ReLU 激活函数、局部响应归一化、Dropout 技术和多种数据增强技术，打响了神经网络往深度发展的第一枪。VGG^[62]提出了更小的卷积核，使网络深度进一步增加。GoogLeNet^[63]引入了 Inception 模块，通过并行应用不同尺寸的卷积核，捕捉多尺度特征并且减少计算量，推动了多尺度特征融合技术的发展。He 等^[64]人提出的 ResNet 引入了残差学习解决了深度网络中的梯度消失问题，将网络深度提高空前的深度。ResNet 的成功证明了残差学习的有效性，并且对后续的网络设计产生了深远的影响。时至今日，即便在此文撰写的过程中，仍有不少网络是以 ResNet 为基础模型改进的，有不少工作是受到 ResNet 启发开展的。

2. 2. 2 监督学习和无监督学习

深度学习的成功依赖于海量数据的支持，其中根据数据的标记与否，可以分为监督学习和无监督学习两种学习范式。

当前，监督学习在深度学习中占据核心地位，尤其是在语音识别、自然语言处理、计算机视觉等方向上已经有了非常成熟的应用。然而监督学习需要对海量

的数据进行标记，这需要标记者拥有较为丰富的领域知识，因此获取大量准确标记的数据在实际中往往是昂贵且不可行的。类比于卷积神经网络能够自动提取特征的思想，如果机器拥有通过自己学习到如专家用以区分不同类别特征的能力，此时不需要耗费巨大的资源做标记，这将为许多标记样本稀少，标记难度大的任务带来福音。因此学者们提出了无监督学习。

无监督的思想是在未标记的数据上挖掘出同类数据的相同特性，进而学习数据内在的结构和模式。根据是否生成伪标签，无监督又分为传统无监督学习和自监督学习。在传统无监督学习中，算法纯粹依赖于数据本身的结构，而不生成任何伪标签。这些方法试图发现数据的内在模式、分布或结构特征，常见的算法包括聚类算法如 Kmeans、层次聚类、DBSCAN、线性判别分析（LDA），密度估计算法如高斯混合模型（GMM）。与传统无监督学习不同，自监督学习通过从数据中生成伪标签来训练模型。根据学习方式的不同，自监督学习分为生成式学习和对比式学习。

生成式学习以自编码器，以数据生成数据，使其在整体或者高级语义上与训练数据相近。这类学习的代表如 GAN^[65]，和 MAE^[66]。GAN 通过生成器和判别器相互竞争学习数据分布，类似于造假者和检验者之间的较量。生成器努力制造以假乱真的数据，而判别器则努力揭露这些伪造。MAE 则利用掩码技术学习图像的细节特征。对比式学习注重同类实例的共同特征和不同类实例的差异，通过最小化类内距离、最大化类间距离来优化模型。相较于生成式学习，对比式学习更注重抽象语义级别的特征，使模型更简单且具有更强的泛化能力。

对比式学习的目标是学习一个编码器，此编码器对同类数据进行相似的编码，并使不同类的数据的编码结果尽可能的不同。在大量研究中，学者大都利用对比学习设计上游的代理任务（Pretext Task）来获取能够提取高度抽象的高维特征的编码器，之后冻结该编码器，利用少量有标签的数据做微调。在 Moco^[42]、SimCLR^[41] 模型中，代理任务是让模型学会区分同一样本的不同视图，对于不同样本的视图要表现出差异性，以此通过学习到相似和不相似的图像表示来提取特征，然而这种特征级上的比较计算量巨大，资源开销包括大量数据点的成对特征比较、和存储内存库（memory bank）和动量网络（momentum network）的维护。在 SWAV^[45]模型中，将原代理任务由匹配问题转为预测问题，预测借助聚类

实现。正常情况下，通过代理任务得到的编码器具有极强的特征抽取能力，可以运用在多种下游任务上，包括但不限于目标检测、图像分类和图像分割。对比学习的一般学习范式如图 2-4 所示。

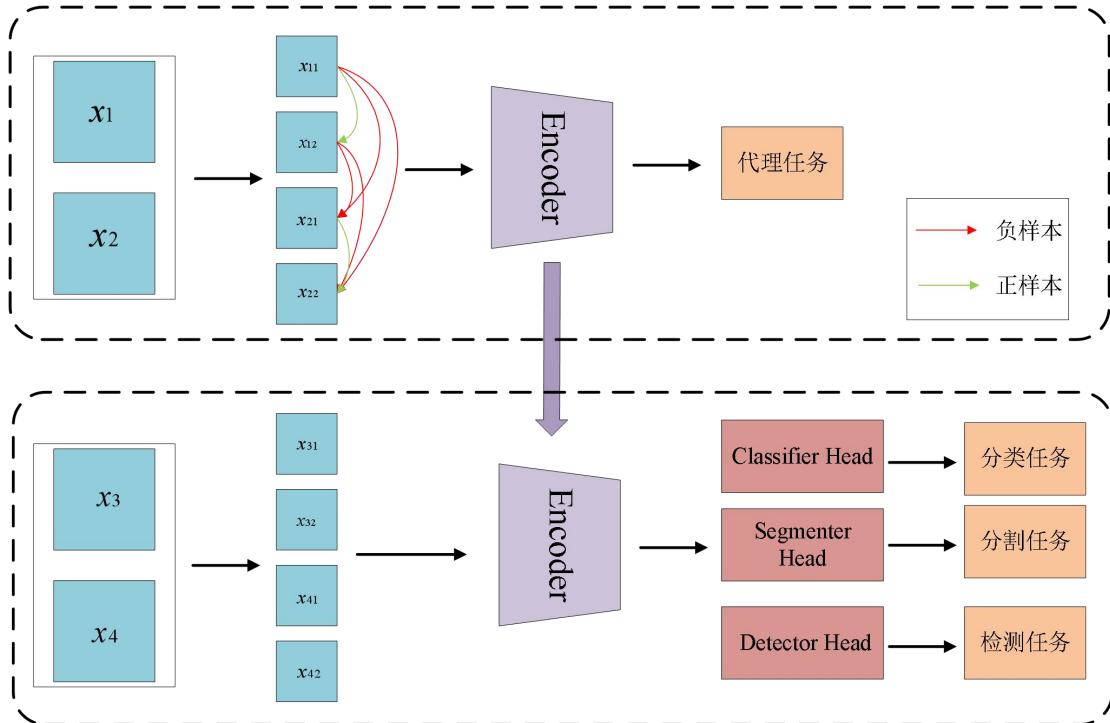


图 2-4 对比学习的一般范式

2.3 块分类介绍

当前主流的 SAR 图像分割算法绝大多数是基于大场景下的分割算法，主流的有 Unet^[72-75]、Deeplabv^[67-68] 系列算法、Transformer^[69-71] 算法，在精确度上取得了不错的效果，然而这些算法对于大场景下的局部小场景检测的效果不佳^[76-77]，这是因为其结构的局限性导致的在下采样式导致某些重要特征丢失。其次，这些算法是逐像素的分类方法，通常需要较多的内存和较高计算能力。块分类则很好的解决了这一问题。

块分类方法是一种高效的图像分割技术，其核心思想是利用图像中局部区域的同质性将图像划分为较小的独立块来显著降低计算复杂度，适用于大规模图像数据的处理。该方法不仅保留了图像如纹理和边缘信息的局部特征，而且通过设计的块尺寸选择，能够匹配图像中物体的相对大小，从而提高识别的准确度。此外，块分类策略允许并行化处理，进一步加速了分割过程。尽管主要关注局部信息，该方法亦能通过先进的图分割算法或条件随机场等技术，整合块间的上下文

信息，优化全局分割结果。

2.4 块分类任务评价指标

在块分类任务中，通常的评价指标包括准确率(Accuracy)、精准率(Precision)、召回率(Recall)、F1 分数(F1score)度量。这些指标可以通过块分类任务混淆矩阵(Confusion Matrix)的值相互运算得到。分类任务混淆矩阵的定义如表 2-1 所示。

表 2-1 分类任务混淆矩阵

混淆 矩阵		预测结果 (单位: 样本)	
真实值	正例	正例 (TP)	反例 (FN)
	反例	假正例 (FP)	真反例 (TN)

准确率(Accuracy)用于衡量分类正确的样本数占总样本数的比例，用于评估模型在所有样本中分类正确的程度，计算公式为：

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2-4)$$

精准率(Precision)用于衡量在所有被分类为正例的样本中，确实是正例的比例，用于衡量模型在预测为正例的样本中的准确程度，计算公式为：

$$Precision = \frac{TP}{TP + FP} \quad (2-5)$$

召回率(Recall)：在所有真正例中，被正确分类为正例的比例，用于衡量模型找出所有正例的能力，计算公式为：

$$Recall = \frac{TP}{TP + FN} \quad (2-6)$$

F1 度量(F1score)：精准率和召回率的调和平均数，F1 度量综合考虑了精准率和召回率，适用于不平衡类别的情况，计算公式为：

$$F1score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (2-7)$$

上指标通常适用于二分类任务的评估，但对于本次的 SAR 地物分类任务，涉及到 8 种地物类型。在这种情况下，True Positive 及类似概念不适用，因此无法直接计算 Precision、Recall 和 F1 指标。通常的做法是采用 Macro-F1、Micro-F1 或 Weighted-F1，将任务视为若干二分类任务，以便分别计算每个类别的 F1 指标。

Macro-F1 是 F1 分数的一个变体，它计算每个类别的 F1 分数，然后取这些分数的平均值，不考虑每个类别的样本数量。这种方法对所有类别都是平等的，无论它们的样本数量如何。计算公式如下：

$$\text{Macro-F1} = \frac{1}{c} \sum_{i=1}^c F1_i \quad (2-8)$$

其中， c 是类别的数量， $F1_i$ 是类别 i 的 F1 分数。*Macro-F1* 提供了一种不受类别样本数量影响的评估方法。它确保了模型在所有类别上的平均性能都得到了考虑，而不是仅仅关注数量较多的类别。

Weight-F1 是另一种 F1 分数的变体，它考虑了类别不平衡的情况。在多分类问题中，如果某些类别的样本数量远多于其他类别，直接计算 F1 分数可能会导致模型对这些大类别的偏好。*Weight-F1* 通过给每个类别的 F1 分数赋予权重来解决这个问题，这些权重与每个类别的样本数量成反比。计算公式如下：

$$\text{Weight-F1} = \frac{\sum_{i=1}^c w_i \times F1_i}{\sum_{i=1}^c w_i} \quad (2-9)$$

其中， w_i 是类别 i 的权重，通常为 $F1_i = \frac{N}{n_i}$ ， N 是所有类别的样本总数， n_i 是类别 i 的样本数量。*Weight-F1* 提供了一种在类别不平衡情况下评估模型性能的方法。它通过给稀有类别更高的权重，使得模型不仅要在大类别上表现良好，还要在小类别上表现良好，从而更全面地评估模型性能。

2.5 分割任务评价指标

在分割任务评价指标中，通常的评价指标包括像素准确率 (PA)、类别平均像素准确率 (MPA)、交并比 (IoU) 平均交并比 (MIoU)、频权交并比 (FWIoU)、以及 F1 分数 (*F1score*)。这些指标可以通过分割任务混淆矩阵 (Confusion Matrix) 的值相互运算得到。分割任务混淆矩阵的定义如表 2-2 所示：

表 2-2 分割任务混淆矩阵

混淆		预测结果 (单位: 像素)	
矩阵		正例	反例
真实值	正例	真正例 (TP)	假反例 (FN)
	反例	假正例 (FP)	真反例 (TN)

PA 的含义是预测类别正确的像素数占总像素数的比例。它对应分类任务的准确率 (*Accuracy*)，计算公式如下：

$$PA = \frac{\sum_{i=0}^k p_{ii}}{\sum_{i=0}^k \sum_{j=0}^k p_{ij}} \quad (2-10)$$

或

$$PA = \frac{TP + TN}{TP + TN + FP + FN} \quad (2-11)$$

上式中， k 表示任务的总类别数， i 表示真实地物标签， j 表示真实预测地物类别。 p_{ij} 表示标签为 i 被预测为 j 的像素数量。它衡量了整个图像中正确分类的像素比例，但是对于样本不均衡的情况，背景与标记样本之间的比例往往严重失衡，用这种方法进行度量并不适合。

MPA 是 PA 的一种简单提升，计算每个类内被正确分类像素数的比例，之后求所有类的平均，其计算公式为：

$$MPA = \frac{1}{k} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij}} \quad (2-12)$$

IoU 衡量的是预测的分割区域与真实分割区域之间的重叠部分占两者并集的比例，它能够衡量模型预测的准确性，其计算公式为：

$$IOU = \frac{TP}{TP + FP + FN} \quad (2-13)$$

$MIOU$ 是计算所有类别的 IoU 的平均值，是一个综合的评估指标，可以比较不同模型或算法在多个类别上的整体性能，计算公式为：

$$MIOU = \frac{1}{k} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}} \quad (2-14)$$

$FWIoU$ 是 $MIOU$ 的一种提升，它考虑了不同类别之间的样本分布差异，根据

每个类出现的频率为其设置权重，适合于类别不均衡的数据集上，其计算公式为：

$$FWIoU = \frac{1}{\sum_{i=0}^k \sum_{j=0}^k p_{ij}} \sum_{i=0}^k \frac{\sum_{j=0}^k p_{ij} p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}} \quad (2-15)$$

*CPA*是真正例占全部预测为正例的比例，计算公式为：

$$CPA = \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + p_{ii}} \quad (2-16)$$

*Recall*是正确预测为正的像素数量占全部正样本的比例，计算公式为：

$$Recall = \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ji} + p_{ii}} \quad (2-17)$$

*F1score*是*CPA*和*Recall*的调和平均值，计算公式为：

$$F1score = \frac{2 \times CPA \times Recall}{CPA + Recall} \quad (2-18)$$

此外，遥感图像的分类解译结果往往需要一定的精度评价指标进行精度验证，只有进行精度验证，才能知道分类解译的结果是否准确可靠。*Kappa*系数就经常被用于影像分类的空间一致性检验，是一种衡量分类精度的指标，其计算公式为：

$$\kappa = \frac{p_o - p_e}{1 - p_e} \quad (2-19)$$

其中， p_o 在数值上等于 *PA*，且

$$p_e = \frac{(TP + FP)(TP + FN) + (TN + FP)(TN + FN)}{(TP + TN + FP + FN)^2} \quad (2-20)$$

2.6 本章小结

本章介绍了 SAR 图像地物分类的理论基础，全面审视了 SAR 图像的成像原理、特性以及在多个关键领域的应用潜力，深入分析了 SAR 图像的高分辨率特性以及由此带来的数据处理挑战。随后本章对机器学习在图像解译中的应用进行了系统性的阐述，特别是深度学习技术如何通过自动提取特征来推动图像处理技术的进步。本章详细比较了传统机器学习方法与深度学习方法的差异，突出了深度学习在处理大规模、高维度数据集时的显著优势，也点明深度学习需要大量数据标注的不足。基于此，在监督学习与无监督学习的讨论中，本章特别强调了无

监督学习在处理未标记数据时的重要性，这对于 SAR 图像地物分类尤其关键，因为获取大量精确标注的 SAR 图像数据既耗时又昂贵。在无监督学习中，本章详细介绍了自监督学习框架及其两种范式，对比式学习和生成式学习，并介绍了如何在缺乏显式标签信息的情况下，利用多视角样本的聚类和类簇区分来学习地物特征。此外，本章引入了块分类方法，这为图像分割任务带来了新的视角。块分类方法通过将图像分成小块进行独立处理，有效降低了计算复杂度，同时保留了图像的局部特征。这种方法特别适合于 SAR 图像的分割，因为它能够适应图像中地物的局部变化，提高分割的精度和效率，第 3 章据此展开。最后本章还对分类和分割任务的评价指标进行了全面的介绍，包括准确率、精准率、召回率、F1 分数等，这些指标对于量化评估模型性能至关重要，是第 4 章的理论支撑。

第3章 跨分辨率对比学习地物分类的方法

3.1 SAR 数据处理

本文数据来自西安电子科技大学智能感知教育部重点实验室，选取汉中地区分辨率为1m和3m的两组数据。该图像是由GF3卫星采集，极化方式为HH，C波段，图像尺寸分别是10011*7710和3337*2570。

3.1.1 裁剪块的确定

地物分类任务的实质是对图像的分割，属于语义分割范畴。块分类方法能够从亚像素级别上提取地物特征，通常情况下，在能获得相同语义的前提下，图像块的尺寸决定了模型学到地物特征的强弱：若图像块的尺寸过小，模型无法学习到有效的语义信息，而更偏重于像素信息，该效果等效于用像素值来作为评判依据，这并非理想效果；若图像块的尺寸过大，在同一图像块中会有多种地物，而块分类方法已经假定每一个图像块中只含有一种地物，因此选取一个合适的块尺寸尤为重要。本文调研了目前常用的用于深度学习图像分类任务的数据集^[46-56]，并将其数据量及尺寸总结如表3-1所示：

表3-1 常见数据集统计

数据集名称	数据量	数据大小	图像尺寸	发布时间	概述
STL10 ^[46]	113,000	2.64GB	96x96x3	2011年	分类&检测
CIFAR10 ^[48]	60,000	163MB	32x32x3	2009年	分类
CIFAR100 ^[48]	60,000	160MB	32x32x3	2009年	分类
MNIST ^[49]	70,000	11MB	28x28x1	1998年	分类
ImageNet ^[50]	数百万	数GB	224*224	2009年	分类
COCO ^[51]	328,000	数百GB	可变	2014年	分类&检测&分割
CelebA ^[52]	202,599	1.33GB	178x218x3	2015年	识别
FashionMNIST ^[53]	70,000	29MB	28x28x1	2017年	分类&识别
SVHN ^[54]	600,000	1.6GB	32x32x3	2011年	识别
Omniglot ^[55]	1623	11MB	105x105x1	2005年	识别
Food101 ^[56]	101,000	5GB	512x512x3	2014年	分类

综合上述内容，本文选取了32*32、64*64、128*128、256*256、512*512（下

记裁剪方式 I~V) 作为备选裁剪的图像块尺寸，并分析按照这样的尺寸裁剪后得到的数据集的纯度：即每一个图像块都有一种地物超过 50% 的数量占总数据集的百分比，将这个值定义为 α ， α 数值的计算公式如下：

$$\alpha = \frac{\max\left(\sum_{i=0,j=0}^{i=w,j=h} P_{0ij}\right)}{\sum\left(\sum_{i=0,j=0}^{i=w,j=h} P_{0ij}, \sum_{i=0,j=0}^{i=w,j=h} P_{1ij}, \sum_{i=0,j=0}^{i=w,j=h} P_{2ij}, \dots, \sum_{i=0,j=0}^{i=w,j=h} P_{nij}\right)} \quad (3-1)$$

其中， n 表示整个数据集上类别总数， w 表示图像块的长度， h 表示图像块的宽度，用 P_{nij} 判断在当前这张图像块像素位置为 (i, j) 属于标签为 n 的地物，若判断为真，将该值赋 1，否则赋 0。

$$P_{nij} = \begin{cases} 1 & (i, j) \text{ 点的像素是属于 } n \\ 0 & (i, j) \text{ 点的像素不属于 } n \end{cases} \quad (3-2)$$

按照上述 5 种裁剪方法，汉中地区分辨率为 1m 的图像按照无重叠裁剪，统计得到如图 1 所示的结果。最优结果是当图像裁剪成 32*32 大小，值为 0.79。说明在整个数据集中有约 79% 的图像块可以选取图像中超过 50% 的类别当作他的标签使用。

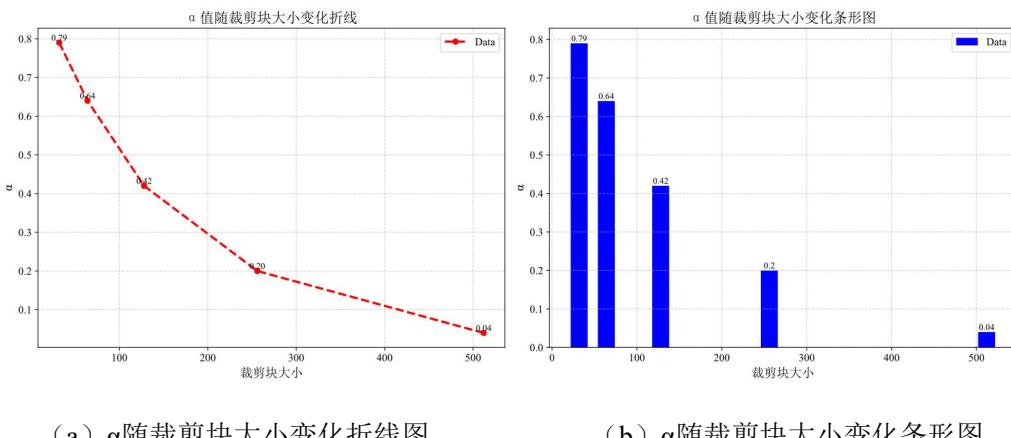


图 3-1 裁剪方式 I~V 得到的 α 值

考虑到，图像裁剪越小，可能会导致模型在训练的过程中过分关注像素信息，更少关注或者不关注语义信息，因此还需要进一步探究尺寸为多大的图像才是训练模型的最佳数据集，这一点将在第 4 章进行分析。

3.1.2 数据裁剪

本文采用汉中地区两类不同分辨率的SAR影像的尺寸分别为10011*7710和3337*2570，分辨率为1m和3m。在3.1.1中已经介绍了要按照裁剪方式I~V的策略裁剪数据。裁剪策略分为有重叠裁剪和无重叠裁剪。无重叠裁剪即每裁剪一次，滑动窗口每次往右或者往下移动图像块大小的像素值。若按照无重叠裁剪，得到的1m和3m的图像数量如表3-2所示：

表3-2 按照无重叠裁剪对1m、3m数据裁剪

裁剪方式	方式I	方式II	方式III	方式IV	方式V
1m	74880	18720	4680	1170	285
3m	8320	2880	520	130	30

如果按照无重叠裁剪，根据表3-2结果，随着图像块尺寸的增大，数据集的规模逐渐减小，特别是当3m分辨率的图像经过方式V的无重叠裁剪只得到了30张数据，难以作为对比学习的训练数据。因此需要按照有重叠的方法裁剪。如果采用有重叠裁剪，重叠裁剪的步长是一个重要的因素。步长能直接影响获取的数据集的规模。而且不同大小的图像块使用的步长也应该随着图像块的尺寸相应的变化，这样避免了较大的图像块使用过小的步长会导致需要裁剪的块数增多，从而增加了计算资源的消耗。

本文提出了一种方法来研究如何确定步长。本文提出了3种方法分别是平移块长度的1/2、1/3、1/4，分别几位步长I、步长II、步长III，得到如下表3-3~表3-8所示的数据规模、处理时间、以及存储大小（计算步长时采用向下取整方法）。

表3-3 按照步长I对1m数据裁剪

数据	数据规模（张）	处理时间（s）	存储大小（MB）
1m-32*32-1/2步长	299520	188987.35s	305 MB
1m-64*64-1/2步长	74329	49752.80s	281.56 MB
1m-128*128-1/2步长	18445	12799.98s	271.86 MB
1m-256*256-1/2步长	4543	3192.52s	265.03 MB
1m-512*512-1/2步长	1102	758.32s	256.43 MB

表 3-4 按照步长II对 1m数据裁剪

数据	数据规模 (张)	处理时间 (s)	存储大小 (MB)
1m-32*32-1/3 步长	766464	516827.19s	792.36 MB
1m-64*64-1/3 步长	173010	129696.12s	655.36 MB
1m-128*128-1/3 步长	42716	30136.41s	629.59 MB
1m-256*256-1/3 步长	10120	7339.09s	590.38 MB
1m-512*512-1/3 步长	2408	1861.75s	560.33 MB

表 3-5 按照步长 III 对 1m 数据裁剪

数据	数据规模 (张)	处理时间 (s)	存储大小 (MB)
1m-32*32-1/4 步长	1198080	856045.52s	1238.55 MB
1m-64*64-1/4 步长	297316	252324.80s	1126.23 MB
1m-128*128-1/4 步长	73233	51888.32s	1079.38 MB
1m-256*256-1/4 步长	17901	13423.09s	1044.31 MB
1m-512*512-1/4 步长	17901	13423.09s	1044.31 MB

表 3-6 按照方步长 I 对 3m 数据裁剪

数据	数据规模 (张)	处理时间 (s)	存储大小 (MB)
3m-32*32-1/2 步长	32913	3654.02s	33.18 MB
3m-64*64-1/2 步长	8137	714.15s	29.19 MB
3m-128*128-1/2 步长	1989	176.11s	27.80 MB
3m-256*256-1/2 步长	475	52.00s	26.20 MB
3m-512*512-1/2 步长	108	12.38s	23.44 MB

表 3-7 按照方步长 II 对 3m 数据裁剪

数据	数据规模 (张)	处理时间 (s)	存储大小 (MB)
3m-32*32-1/3 步长	84074	7283.24s	84.75 MB
3m-64*64-1/3 步长	18720	1518.81s	67.16 MB
3m-128*128-1/3 步长	4543	400.94s	63.50 MB
3m-256*256-1/3 步长	1036	99.36s	57.15 MB
3m-512*512-1/3 步长	221	21.64s	47.97 MB

表 3-8 按照方步长 III 对 3m 数据裁剪

数据	数据规模 (张)	处理时间 (s)	存储大小 (MB)
3m-32*32-1/4 步长	131652	10756.45s	132.71 MB
3m-64*64-1/4 步长	32185	3645.53s	115.47 MB
3m-128*128-1/4 步长	7777	656.17s	108.70 MB
3m-256*256-1/4 步长	1813	153.23s	100.01 MB
3m-512*512-1/4 步长	391	47.26s	84.88 MB

通过对分辨率为 1m、3m 的图像按照裁剪块大小为 32*32、64*64、128*128、256*256、512*512 设置步长为裁剪块的 1/2、1/3 和 1/4 一共 30 组实验数据的规模分析、存储分析、裁剪时间分析，得到结果：使用过小的步长会增加计算资源消耗，同时增加处理时间和内存需求，并导致裁剪块之间重叠过多，增加了冗余信息和计算量，这在大尺寸图像表现更加明显；使用过大的步长会导致图像数量不足。因此在本文中，折中选择按照 1/3 步长来进行有重叠裁剪。此外，考虑到 512*512 的图像块在裁剪的结果中的数据量不足的问题可能会导致设计的模型出现过拟合的情况，因此 512*512 大小的图像块也不再考虑，同时，考虑到步长为 1/3 裁剪块大小计算带来的取整问题，以及将 1m 和 3m 分辨率下的相对裁剪问题，本文重新设计了裁剪块尺寸以更好的契合任务。重新裁剪块选择的依据是：尽可能和原裁剪块大小相接近；在 3m 分辨率的图像上做相对裁剪的时候尽可能不会造成整除问题。综合考虑，本文选取了 27*27、63*63、126*126、252*252 的大小作为新的裁剪块大小，将其命名为裁剪方式 VI~XI。

3.1.3 SAR 数据集分割

为了保证每一类数据比例平衡，本文选择在每一个类别上进行数据集的划分。将训练集和测试集划分为 8:2 的比例。具体操作为：给每一类中所有图像随机编写序号，以图像编号对 5 求余，若余数为 0 到 4 随机初始化的某个值，则以该编码为名称的图像块将被划分到测试集，该轮中其他图像块被划分到训练集；一轮结束后，再重新初始化数值。

3.1.4 SAR 数据增强

根据当前研究对比学习的研究^[43-45]，本文设置了高斯变换、颜色变换、图

像翻转、多裁剪等数据增强的方法，这些方法在传统的自然图像上能够发挥重要作用，使模型对输入数据的变化更加鲁棒。然而由于SAR图像的特殊性，这样的变换不一定适用于SAR图像的数据增强。因此本文在第4章做了详细的消融实验。

3.1.5 跨分辨率 SAR 图像对

要训练一个优质的模型，采用数据增强往往会有比较好的效果，而相对于传统的通过加噪、翻转、裁剪等数据增强的方法，采用多视角的方法也开始被引入到深度学习中。Tian 等^[58]人通过对比不同视角的特征表示来学习共享信息的特征，黄等^[59]人提出了多视角融合的方法，通过结合局部和全局特征来提高分类性能，苏等^[60]人探讨了多视角多标签学习问题，并构建了偏最小二乘回归模型来探索训练实例和测试实例之间的相关性。

基于以上工作的可行性以及对比学习的特殊性，本文采用两种方法生成多视角样本。首先，采用图像随机裁剪技术，通过在一定范围内对图像进行裁剪，获取原图像中目标的局部信息，这种方法是从单幅图像自身出发的多视角样本（如图 a 所示）。此外，还可以从不同成像参数出发的多视角样本构建，如本文设计了将 1m 和 3m 分辨率的图像的相对位置作为一组对比对，来获取更加精确的语义信息（如图 b 所示）。这种方法有助于后续基于聚类的判别式自监督代理任务，从地物目标的局部与整体关系中学习。

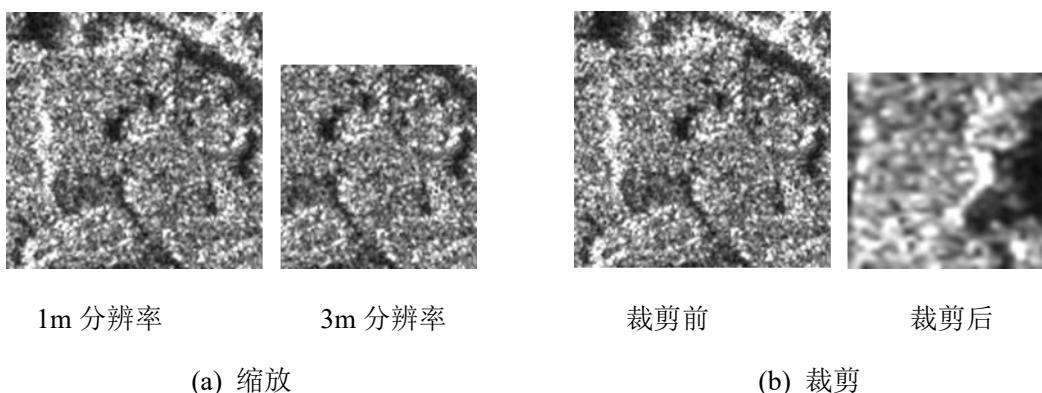


图 3-2 图像的多视角样本构建

本文的数据整体处理方式为：首先对一个 1m 分辨率下的 SAR 图像块样本进行复制，生成 8 个副本。然后，对每个副本进行尺寸调整，将图像放大至 256×256 的尺寸。接着，随机选取两个副本并分别随机裁剪 224×224 的子区域，用以提取整体的 SAR 语义信息，同时将剩余 6 个副本裁剪为 96×96 的小区域，

以获取局部的 SAR 细节信息。此外，针对每个副本添加高斯噪声，以模拟实际环境中的信号干扰和噪声影响。同时，为了增加数据的多样性和鲁棒性，添加颜色变化，调整图像的亮度、对比度或其他颜色属性。此外，随机翻转图像以增加数据的随机性和泛化能力。同时，对应的 3m 分辨率下的图像也做如此操作，之后将两类数据堆叠在一起，如此，对于一个 SAR 图像样本及其副本，一共生成了 16 个增强样本，这些样本都用于描述同一块地物。

3.2 模型设计

3.3.1 整体流程图

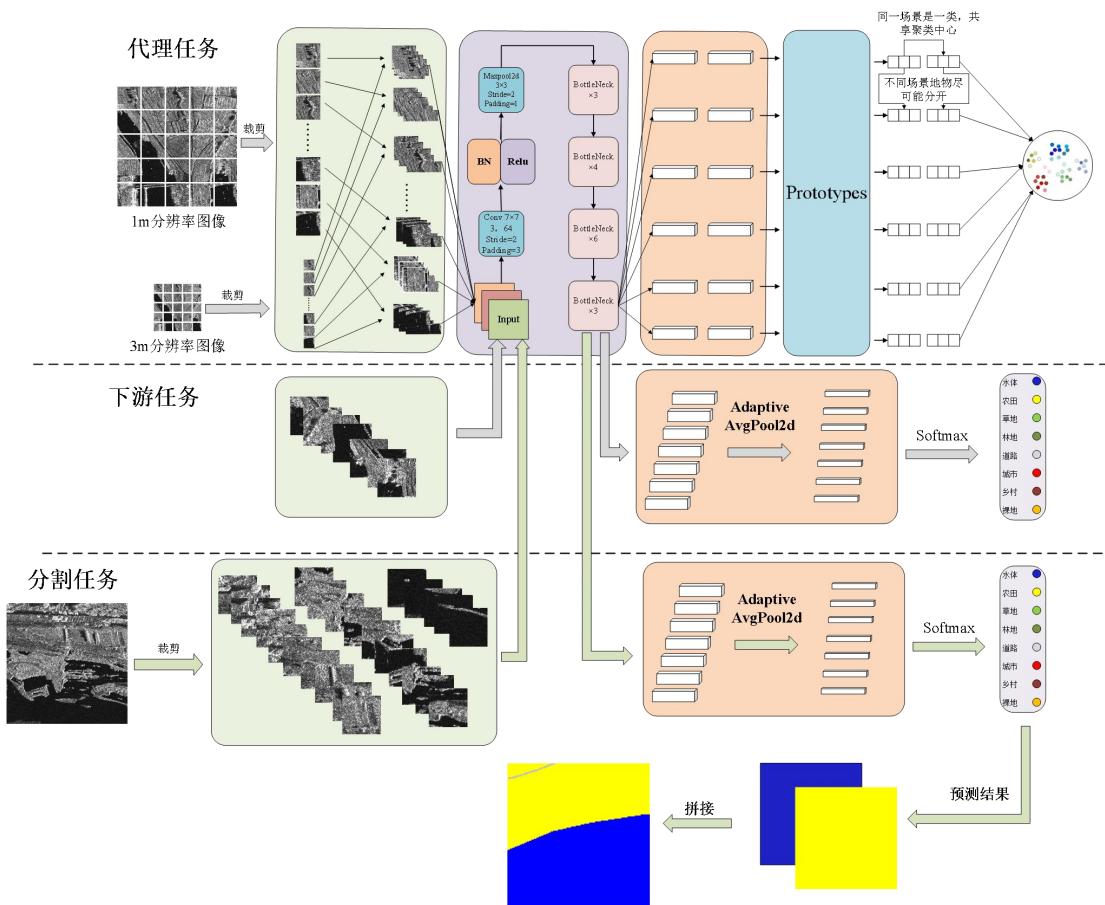


图 3-3 整体流程图

3.2.2 代理任务

在对比学习的上下文中，特征提取过程通常涉及将输入数据通过一系列编码器层映射到一个特征空间。在这个空间中，相似的输入应该更接近，而不相似的

输入应该更远离。基于此理念，本文提出了基于自监督学习的对比聚类框架 SWAV-ResSegNet，旨在探索不同成像条件下 SAR 图像地物的高层语义。该方法的核心步骤包括以下几点：首先，对同类地物的多视角样本进行聚类，以形成类簇；然后，通过区分各类的类簇，学习各类地物特征的本质表示。这一框架有助于在缺乏显式标签的情况下，从 SAR 图像中学习到有用的地物语义信息。

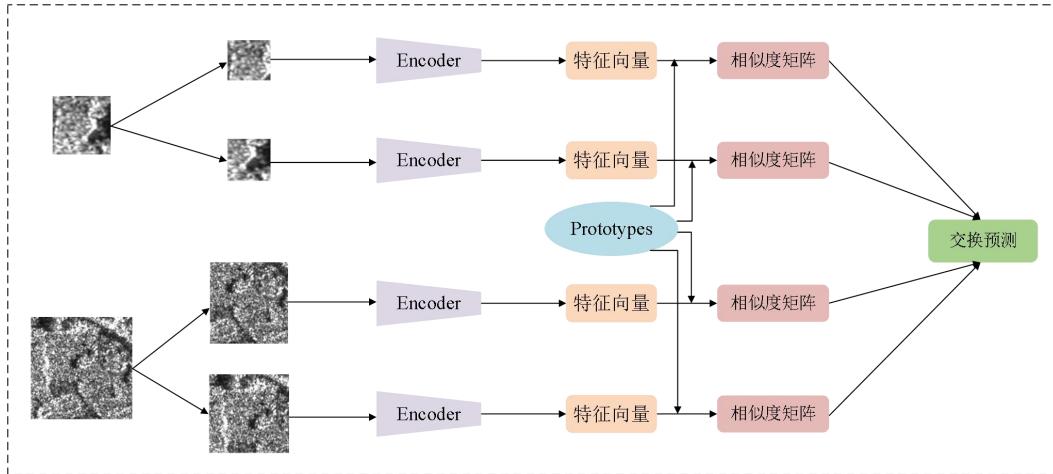


图 3-4 基于对比聚类的判别式代理任务框架

代理任务框架运行逻辑是：图像经过数据增强之后得到若干张增强样本 $\{x_1, x_2, x_3, \dots, x_n\}$ ，这些样本经过同一个特征提取器编码器 $encoder$ 后，得到若干个特征 $\{z_1, z_2, z_3, \dots, z_n\}$ ，通过将得到的特征与设定的 K 个原型集 $\{c_1, c_2, c_3, \dots, c\}$ 匹配得到特征对应的编码 $\{q_1, q_2, q_3, \dots, q_n\}$ ，继而设置一个具有如下损失函数的交换预测问题，为了更好理解，取 $n = 2$ 的情况举例：

$$L(z_t, z_s) = \ell(z_t, q_s) + \ell(z_s, q_t) \quad (3-3)$$

上式包含从特征 z_t 预测编码 q_s 及从特征 z_s 预测编码 q_t 的两项，每项表示编码和概率的交叉熵损失，其中概率由特征 z_i 与原型集中所有原型的点积得到。

其中 $\ell(z, q)$ 用于度量特征 z 与编码 q 间的距离，具体形式如下：

$$\ell(z_t, q_s) = - \sum_k q_s^{(k)} \log p_t^{(k)} \quad (3-4)$$

其中，

$$p_t^{(k)} = \frac{\exp\left(\frac{1}{\tau} z_t^T c_k\right)}{\sum_k \exp\left(\frac{1}{\tau} z_t^T c_k\right)} \quad (3-5)$$

τ 表示温度参数。

上述表示训练样例集中一对样例的情况。令 τ 表示多视角样例变换集合, x_{nt} 表示图像样本 x_n 的第 t 个视角样例, 对应得到的图像特征向量及编码分别是 z_{nt} 和 q_{nt} 。将(3)式应用于 N 幅图像及成对的数据扩充, 可得如下交换预测问题的整体损失函数:

$$-\frac{1}{N} \sum_{n=1}^N \sum_{s,t \sim T} \left[\frac{1}{\tau} z_{nt}^T c_{q_{ns}} + \frac{1}{\tau} z_{ns}^T c_{q_{nt}} - \log \sum_{k=1}^K \exp\left(\frac{z_{nt}^T c_k}{\tau}\right) - \log \sum_{k=1}^K \exp\left(\frac{z_{ns}^T c_k}{\tau}\right) \right] \quad (3-6)$$

通过对原型集及图像编码器参数的联合最小化得到解。和其他两两特征对比方法不同的是, 该方法并不把用编码作为学习目标, 而是通过同类地物不同的视角的编码需要保持一致来进行学习, 因此本方法可以理解为是一张图片的多个不同视角需要对应相同的聚类中心而不是直接用它们的特征来做。

3. 2. 3 下游任务

在深入研究了对比学习框架中的代理任务, 并了解了如何通过这些任务学习到有用的特征表示之后, 现将重点转向学习到的特征在 SAR 图像处理中的应用。代理任务的学习结果如何在实际下游任务中发挥作用是一个关键问题。在对比学习的背景下, 下游任务通常是指最终的目标任务, 如图像分类、目标检测等, 这些任务是对比学习模型训练完成后所要解决的实际问题。在本文中, 总体任务是对 SAR 图像的分割, 下游任务可以基于图像分类任务和图像分割任务来进行。

1. 基于图像块分类的下游任务

在面对 SAR 图像分类任务时, 可以采用传统机器学习算法和深度学习方法两种不同的途径。传统机器学习算法如支持向量机 (SVM) 和朴素贝叶斯 (Naive Bayes) 通常基于手工设计的特征和浅层模型, 这需要大量领域知识和经验。此外, 这些方法可能需要根据不同的数据集进行参数调整, 增加了调参的复杂性。另外, SVM 在处理大规模数据时可能会受到内存和计算资源的限制, 因为它需

要在整个数据集上计算核矩阵，导致计算复杂度较高，并且对噪声和异常值敏感，模型的鲁棒性较差。

相比之下，深度学习方法依赖于神经网络，具有自动学习特征表示和处理大规模数据的能力。神经网络能够自动学习数据中的特征表示，无需手动设计特征，从而适应不同数据的复杂特征关系。在处理大规模数据和复杂模式识别任务上，神经网络表现出色，例如语音识别、图像处理等。此外，神经网络能够构建非线性模型，并且在深度神经网络中可以学习到更加复杂的非线性关系，从而提高分类任务的性能。

因此，本小节选择利用神经网络来构建基于跨分辨率对比学习 SAR 图像的下游分类器。这种方法能够充分利用对比学习框架中学到的特征表示，从而提高 SAR 图像分类的准确性和效率。构建的下游任务模型框架如图 3-5 所示：

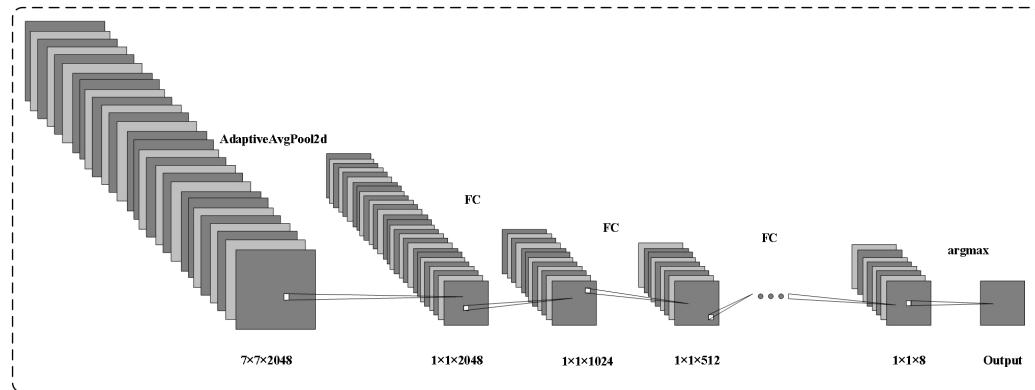


图 3-5 下游任务的网络结构

下游任务框架运行逻辑是：特征提取网络将每个 SAR 图像样本映射到一个高维特征空间中的点后要将每个做进一步处理。首先，输入图像经过自适应平均池化层，该层将输入特征图的尺寸调整到 $H \times W$ ，以便于后续全连接层的处理，其数学表达式为：

$$z_{pool} = P_{av}(F_{in}) = \frac{1}{H \times W} \sum_{h=1}^H \sum_{w=1}^W F_{in}(h, w) \quad (3-7)$$

其中， $P_{av}(F_{in})$ 是池化后的特征向量。池化后的特征向量随后进入全连接层。每个全连接层 f_c 可以表示为一个线性变换，后跟一个非线性激活函数。第 l 层的全连接操作可以表示为：

$$Z^l = W^l X^l + b^l \quad (3-8)$$

$$X^l = Z^{l-1} \quad (3-9)$$

其中, W^l 是第 l 层的权重矩阵, b^l 是偏置向量, X^l 是第 $l-1$ 层的输出 (对于第一层, X^l 就是池化后的 $P_{av}(F_{in})$), Z^l 是第 l 层的线性输出。

利用 ReLU 激活函数将全连接层的线性输出转换为非线性输出, 定义为:

$$A^l = \text{ReLU}(Z^l) = \max(0, Z^l) \quad (3-10)$$

这表示如果 Z^l 中的元素小于 0, 那么在 A^l 中这个元素将被置为 0。

最后一个全连接层 F_c 的输出 Z^k 被用作图像样本 X^i 的特征表示:

$$Z_k = W^z Z^{k-1} + b^z \quad (3-11)$$

其中 k 表示设计了 k 个全连接层。 Z_k 代表含义为映射到每个类别的概率, 对 Z_k 进行求解得到该 SAR 图像块经过神经网络后被预测的类别序号 $pred$, 计算公式为:

$$pred = \text{argmax}(Z_k) \quad (3-12)$$

2. 基于图像分割的下游任务

当前的图像分割模型普遍依赖于先进的特征提取器来实现对图像内容的精确识别和区分。从这个角度出发, 可以将主流的图像分割网络看作“预训练+微调”的学习范式。这种范式首先通过在大规模数据集上进行预训练, 学习到丰富的特征表示, 然后在特定任务上进行微调, 以适应特定的分割需求。要训练一个优质的特征提取器, 往往需要海量的数据, 而 SAR 地物分类提供的数据往往不能提供满足此需求。

本文提出的代理任务正好可以解决这一问题, 利用代理任务中已经训练好的特征提取器, 将其应用于下游任务的特征提取阶段。这种方法的优势在于, 预训练的特征提取器已经具备了对 SAR 图像特征的高度抽取能力, 能够捕捉到复杂的图像特征。当这些特征被用于下游任务时, 可以显著提升分割的准确性和效率。由于本文主要以基于图像块分类为下游任务进行图像分割, 因此这一部分不再详细展开, 本文只在第 4 章第 5 小节记录了对比实验及结果。

3.3 分割任务

3.3.1 块分类粒度

本节采用一种分块分类策略，其在地物类别识别任务中被广泛应用。该策略通过将输入图像分割成多个预测块，并针对每个块进行独立的有重叠的预测。随后，通过整合各个预测块的信息，以确定整体图像中特定地物类别的位置和分布。这种方法的优势在于能够有效地处理大规模图像数据，并提高对复杂地物类别的识别准确性。根据上述内容，将块分类粒度定义为：将图像块均等分割 N 个小图像块，通过这 N 个小图像块来预测其中 1 个小图像块的类别，将该窗口所在的的地物范围的一部分标记为该类别，之后按照步长滑动，直到滑到图像两侧，具体算法如下：

算法流程

Input: 图像块 block、滑动窗口 window 大小、步长 step

Output: 分割图 result

1. 根据设定的滑动窗口和步长创建一个填充后的表示结果的图像块 block'， $block'.size > block.size$ ；
 2. for 纵向移动 do
 3. for 横向移动 do
 4. 对滑动窗口 window 进行预测
 5. 将 window 所在区域的相对位置标记在 result 指定区域标记为识别的类
 6. 按步长迭代直到滑动窗口到达 block' 横向边缘
 7. end for
 8. 按步长迭代直到滑动窗口到达 block' 纵向边缘
 9. end for
-

为了探究块分类粒度对分割结果的影响，本节设计了 5 组实验，分别记作粒度 I、粒度 II、粒度 III、粒度 IV、粒度 V。其中粒度 I 表示无重叠预测，即步长等于滑动窗口长宽大小；粒度 II 表示用 4 个图像块来预测一个图像块每次把 1 号位右下角、2 号位左下角、3 号位左上角、4 号位右上角各 1/4 大小的区域标记为类别，粒度 III 表示用 9 个图像块来预测一个图像块，每次把中心位置（5 号

位) 标记为类别中心; 粒度 IV 和 V 的做法和粒度 II 和 III 类似, 分别把 6 号位右下角、7 号位左下角、10 号位左上角、11 号位右上角各 1/4 大小的区域和 13 号区域标记为类别, 粒度 I~V 相对位置关系如图 3-6 所示:

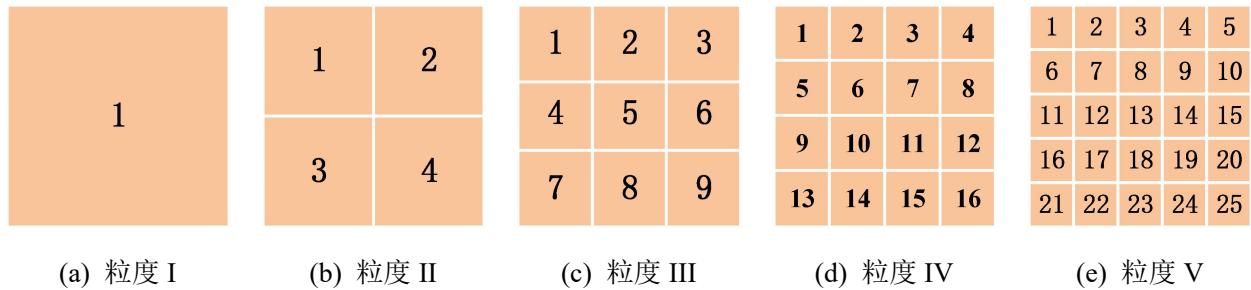


图 3-6 五种粒度的图像块编号示意图

3.3.2 条件随机场

采用基于块的预测方法进行地物要素的自动提取时, 其结果的准确性和连贯性受到多种因素的影响。尽管通过减小图像块的尺寸以提高预测粒度或者利用更细的粒度预测, 得到的分割结果在视觉上仍存在不连续性, 特别是在河流等线性地物的提取上, 这种情况表现得更为明显。这是因为使用这种补丁级的地物要素提取方法难以避免会产生一些误差并且在不同区域边缘处会产生锯齿状结构, 因此需要采用后处理的方法提高精度获得高质量的地物要素提取结果图。本章引入了全连接条件随机场 (FC-CRF) 作为优化手段, 以增强分割的精度。FC-CRF 作为一种先进的条件随机场模型, 其核心优势在于能够捕捉并建模图像中任意像素对之间的关系, 从而利用全局的图像上下文进行像素级的分类决策。这种全局性的考量可以显著提升分割任务的精确度, 特别是在复杂的场景和边界细节的处理上表现突出。其具体过程如下:

假设目标 SAR 图像和预测图像有 n 个像素, 对于每个像素构建一个随机场 $X = \{X_1, X_2, \dots, X_n\}$, 其中 X_k 代表分配给第 k 个像素的标签, 这同样适用于像素的 RGB 颜色向量。同时为像素的 RGB 向量创建随机场 $I = \{I_1, I_2, \dots, I_n\}$, 其中 I_k 表示第 k 个像素的 RGB 向量。此外假设 FC-CRF 模型遵循吉布斯分布:

$$P(X|I) = \frac{1}{Z(I)} e^{-E(X|I)} \quad (3-13)$$

其中 x 是观测值, $E(x|I)$ 是能量函数, 由一元势函数和二元势函数组成, 可以写

为：

$$E(x|I) = \sum_i \psi_u(x_i) + \sum_{i,j} \psi_p(x_i, x_j) \quad (3-14)$$

一元势函数由深度学习模型的输出特征图经过 *softmax* 函数转换得到的概率分布图表示。该概率分布图反映了每个像素点属于各个类别的概率。而二元势函数融合了原始图像的空间位置信息和颜色信息，用于描述像素点之间的关系，促使相邻像素采取相似的类别标签以增强图像的一致性。二元势函数可以写为：

$$\psi_p(x_i, x_j) = u(x_i, x_j) \sum w^m K_G^m(f_i, f_j) \quad (3-15)$$

其中， $\psi_p(x_i, x_j)$ 是标签兼容项，它约束了像素间传导的条件，只有在相同标签条件下，能量才可以相互传导。

3.4 训练设置

在代理任务模型设置上，选取前两个裁剪块用于计算样本分配；设置温度参数为 0.1，正则化参数设定为 0.05，设置迭代次数为 3，来实现每一步迭代的精细调控。此外，设置特征维度为 128，原型的数量设定为 3000。上述参数均与 Caron^[45]等人的工作保持一致；在代理任务训练设置中，设置初始学习率为 0.2，动量为 0.9，权重衰减为 1e-6。采用 SGD 优化器并将其包装在 LARC 中，LARC 根据每层梯度情况动态调整学习率，信任系数为 0.001，且禁用梯度裁剪。采用余弦退火学习率调度策略，通过余弦函数的形状动态调整学习率。在下游任务中，设置 epoch 为 100，学习率设置为 0.05，使用 SGD 优化器，权重衰减为 1e-6，采用余弦退火学习率调度策略，通过余弦函数的形状动态调整学习率。本文训练环境配置表见表 3-9。

表 3-9 实验环境配置

环境	版本配置
深度学习框架	Pytorch
操作系统	Linux
语言及其版本	Python 3.7
CPU	鲲鹏 920 aarch64 架构
NPU	HUAWEI Ascend 910PremiumA

3.5 本章小结

本章详细介绍了基于跨分辨率对比学习的自监督学习地物分类的方法论。包括数据处理，模型设计，实验设置。在数据处理任务中，介绍了 SAR 图像的预处理，包括块分类、数据裁剪、数据增强以及跨分辨率图像对的构建。提出了一个包含代理任务和下游任务的自监督学习框架。在代理任务中，采用了对比学习框架，通过构建一个对比聚类框架来探索 SAR 图像在不同成像条件下的高层语义。该框架利用多视角样本的聚类和类簇的区分，学习地物特征的本质表示。在下游任务中，本章对比了传统机器学习算法和深度学习方法在 SAR 图像分类上的优劣，并最终选择了基于深度学习的分类器，以充分利用学习到的复杂特征表示。最终设计的下游任务模型框架包括特征提取、自适应平均池化、全连接层和分类输出等步骤，其中 ReLU 激活函数引入了非线性，增强了模型的表达能力。此外，本章还探讨了如何将分类任务的方法应用于图像分割任务，并设计了不同块分类粒度对分割结果的影响实验。此外，引入了条件随机场（CRF）作为后处理步骤，以增强分割结果的连贯性和准确性。

第 4 章 实验结果与分析

4.1 分类任务结果和分析

本节旨在研究如何使模型性能达到最佳水平，重点探讨了裁剪块尺寸和下游任务网络层数对实验结果的影响，并对网络提取的特征进行了定性评估。

4.1.1 裁剪块尺寸

在第 3 章中确定了方法 VI、VII、VIII、IX 四种裁剪方法用于制作数据集，取这四个数据集的所有数据进行代理任务，并且在四个数据集的训练集中分别随机取出 10% 的数据作为各自下游任务的训练数据，用各自的测试集作为对训练的模型性能的测试。表 4-1 记录了在不同裁剪块尺寸下下游任务各训练 100 轮后代理任务的准确率。

表 4-1 不同图像块尺寸下代理任务和下游任务准确率

裁剪 方式	下游任务指标(%)		
	Accuracy	Micro-F1score	Weight-F1score
方式 VI	68.2	22.0	61.2
方式 VII	72.1	34.4	67.2
方式 VIII	84.2	77.6	84.2
方式 IX	83.8	59.7	83.4

上表中，方式 VI 是 27*27 裁剪方式，方式 VII 是 63*63 裁剪方式，方式 VIII 是 126*126 裁剪方式，方式 IX 是 252*252 裁剪方式。实验结果表明，裁剪块尺寸对模型性能有显著影响。随着裁剪块尺寸的增大，模型的准确率先升后降，最佳性能出现在 126*126 尺寸。这可能是因为较小的裁剪块使模型过于关注像素特征，而较大的裁剪块则可能忽略细节。因此，选择适当的裁剪块尺寸是平衡特征学习的关键。通过可视化分析，我们验证了这一假设。

图 4-1 显示，在裁剪方式 VIII 下，模型关注的地方确为目标地物。这表明裁剪方式 VIII 在提取目标地物特征方面具有优势，而其他裁剪方式在学习过程中可能出现了一定程度的偏差导致模型在对目标地物的关注上不够准确或全面。故裁剪方式 VIII 为最佳。图 4-2 是利用裁剪方法 VI~IX 获得的数据集的测试结果。

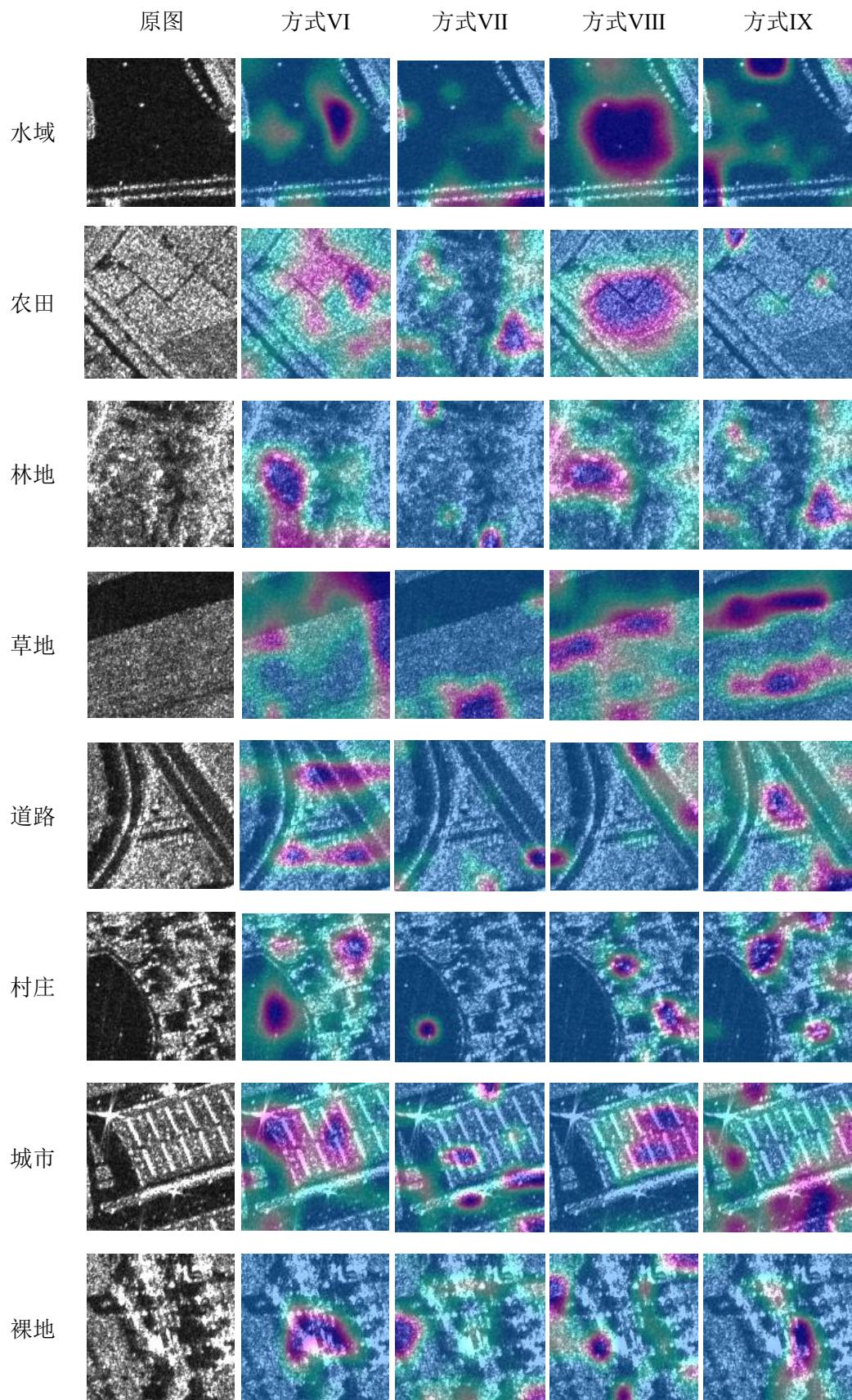


图 4-1 方式VI~IX代理任务的 encoder 提取特征可视化

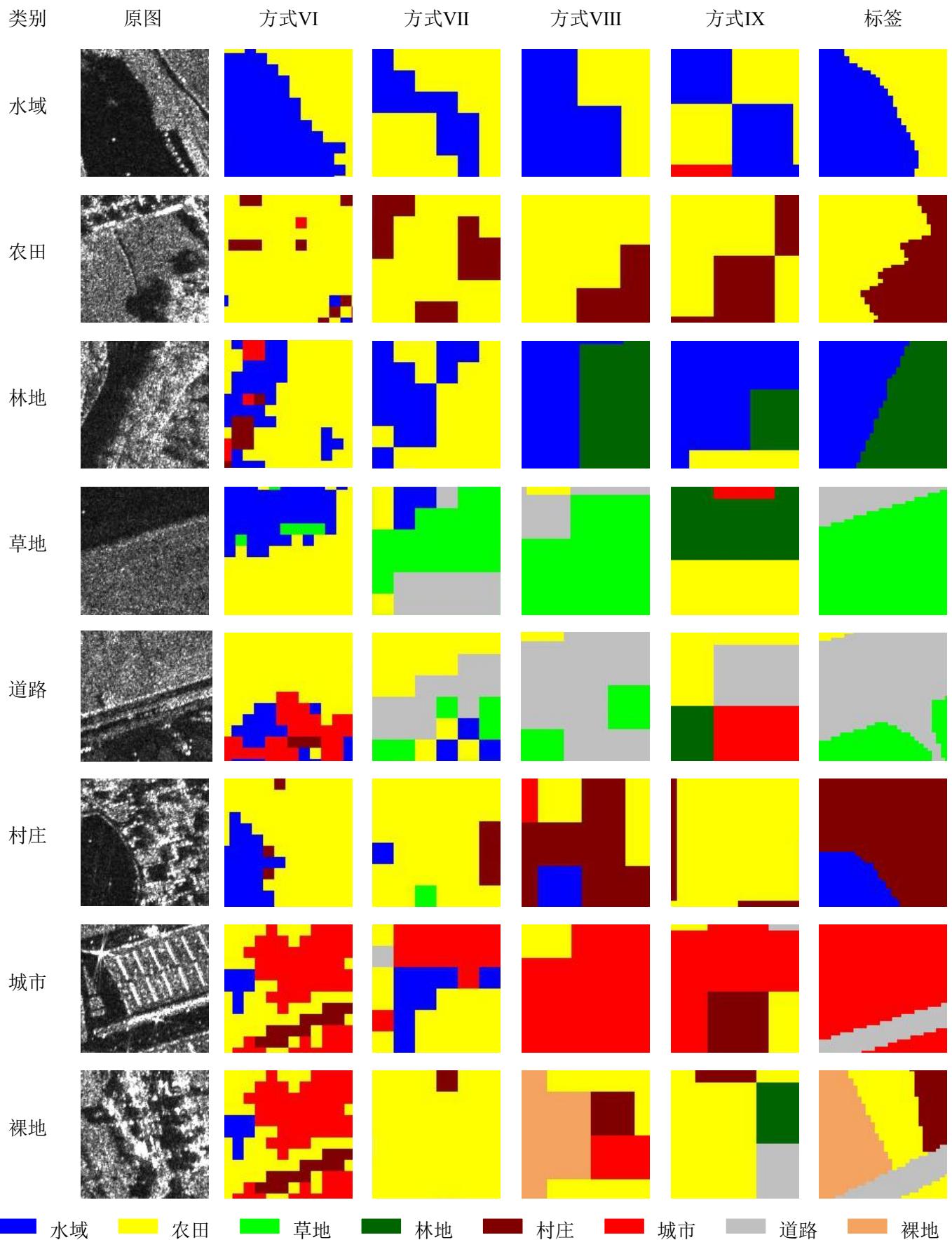


图 4-2 在汉中 1m 数据上的测试结果图

4.1.2 下游任务网络层数

本文在第3章提出了利用全连接网络作为下游任务的分类器，在训练过程中，全连接网络的层数能直接影响下游任务的精确度。因此本节提出了8种不同的网络结构，从1层全连接层到8层全连接层。数据集采用裁剪方式VIII得到的数据集，模型采用方式III得到的数据集训练的模型，在下游任务上训练100轮，分别记录其在第50和第100轮的精确度得到如表4-2所示的结果：

表4-2 不同全连接层数下的准确度表

全连接层数	维度变化	50轮精确度	100轮精确度
1	2048 - 8	83.0	84.3
2	2048 - 512 - 8	84.4	85.4
3	2048 - 512 - 64 - 8	85.5	87.4
4	2048 - 1024 - 256 - 64 - 8	86.2	87.8
5	2048 - 1024 - 512 - 256 - 64 - 8	86.7	88.1
6	2048 - 1024 - 512 - 256 - 128 - 64 - 8	86.8	88.4
7	2048 - 1024 - 512 - 256 - 128 - 64 - 32 - 8	86.2	88.2
8	2048 - 1024 - 512 - 256 - 128 - 64 - 32 - 16 - 8	86.0	87.9

从上表结果可以看出，当下游任务使用6层全连接层时，在50轮和100轮的训练结果，其分类的准确度在同等训练设置不同网络层数的任务中均达到大值。这可能是因为在6层全连接层的情况下，模型能够更好地平衡复杂度和泛化能力。增加全连接层的深度会增加模型的复杂度，同时也会增加过拟合的风险，这种现象在SAR图像这种数据量少，且噪声较多的数据上更为明显。6层全连接层可能正好能够捕获到适当层次的抽象特征，既能够保留输入数据的原始特征信息，又能够学习到足够复杂的表示来区分不同类别。而增加层次可能会使模型过度依赖高阶特征，导致模型在面对新样本时泛化能力下降。

4.2 分割任务结果和分析

本节旨在研究如何使模型在分割任务上达到最佳水平，重点探讨了预测块颗粒度和利用条件随机场（CRF）对实验结果的影响。

4.2.1 块分类粒度

在第3章中，本文提出了5种不同的粒度，本节采用方法VIII裁剪数据集，选择6层全连接层，以探究块分类粒度对语义分割的影响。表4-3记录了对应的各粒度下的相关指标数据。

表4-3 分割任务在1m的整体数据上各项指标

粒度	PA	MPA	MIOU	FWIOU	Kappa
粒度 I	0.791	0.616	0.498	0.663	0.627
粒度 II	0.800	0.639	0.509	0.674	0.642
粒度 III	0.806	0.683	0.556	0.628	0.653
粒度 IV	0.805	0.669	0.540	0.681	0.651
粒度 V	0.704	0.328	0.238	0.567	0.499

上表中，开始时随着粒度的细化，块分类模型在分割任务上的各项指标逐渐提升，且PA、MPA、MIOU和Kappa系数在粒度III时达到最高值，这表明，适度增加图像的粒度可以提供更多的上下文信息，帮助模型更准确地预测中心物体的类别。然而，随着细化的深入，这种增加趋势不会持续，尤其是在粒度V下，所有指标显著降低，这是由于过多的上下文信息引入了更多的噪音和干扰，反而降低了模型的预测精度。因此，本实验最佳的粒度是粒度III。

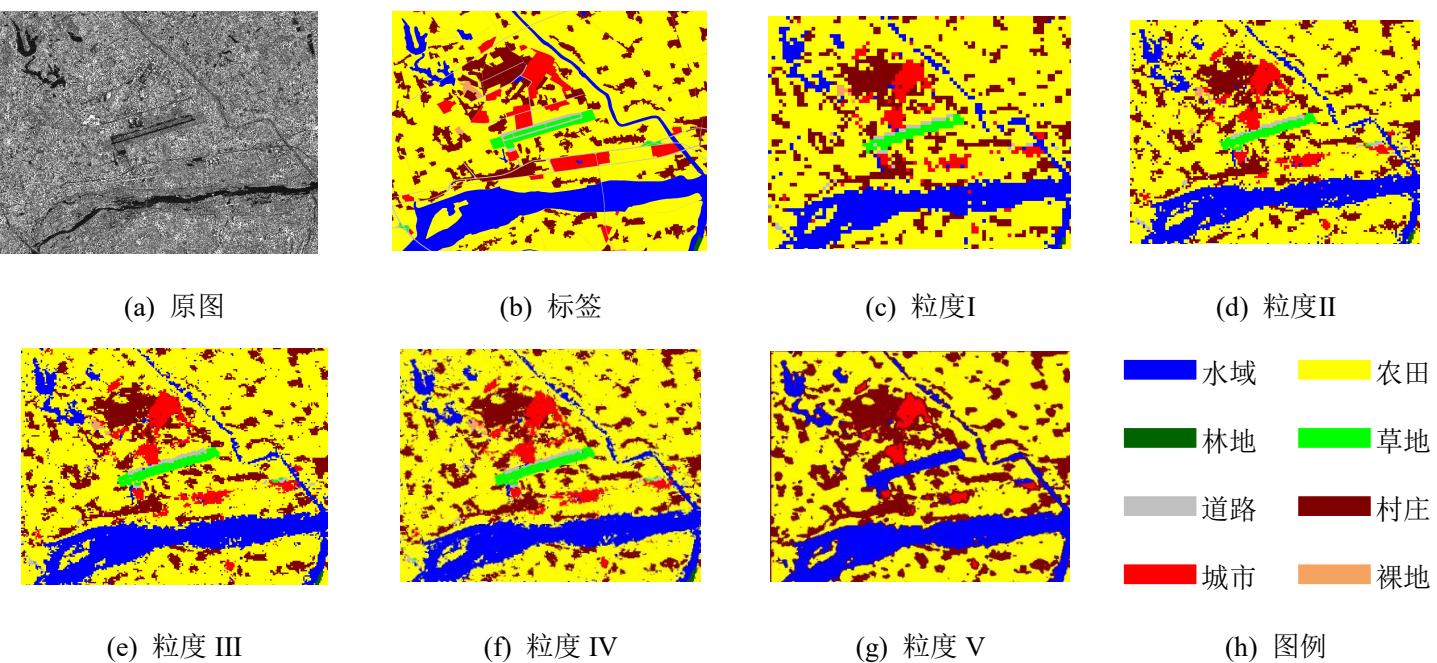


图4-3 在汉中1m数据上的整体测试结果图

4.2.2 全连接条件随机场

在第3章中，本文提出了利用全连接条件随机场作为初步分割结果的后处理方法以消除结果的锯齿效果。本节设置一元势能的目标概率为0.5，将整个预测的分割图裁剪为若干个512*512大小的分割图，在每个条件随机场中迭代5次，并记录各种粒度下利用全连接条件随机场处理结果的各种指标，得到如表4-4所示的结果。

表4-4 各粒度利用CRF算法在1m的整体数据上各项指标

粒度	PA	MPA	MIOU	FWIOU	Kappa
粒度 I	0.795	0.674	0.540	0.670	0.631
粒度 II	0.817	0.643	0.517	0.683	0.655
粒度 III	0.837	0.694	0.567	0.646	0.654
粒度 IV	0.823	0.661	0.543	0.635	0.673
粒度 V	0.714	0.307	0.212	0.543	0.509

上表中，粒度I到粒度IV的实验结果，使用CRF后的MIOU、FWIOU和Kappa值普遍高于未使用CRF的结果，特别是在粒度III下，使用CRF后，PA、MPA、MIOU、FWIOU和Kappa值均达到了最高值，这表明CRF在这种粒度下对降噪和增强图像分割边界清晰度的显著效果。从结果图（见图4.4）的直观角度看，粒度III经过CRF处理后的结果更加光滑。此外，使用CRF后，粒度III的PA从0.806提高到0.837，MPA从0.683提高到0.694，MIOU从0.556提高到0.567，Kappa系数从0.653提高到0.654。这些指标的提升表明CRF在保持全局一致性的同时，提高了分割的精度。

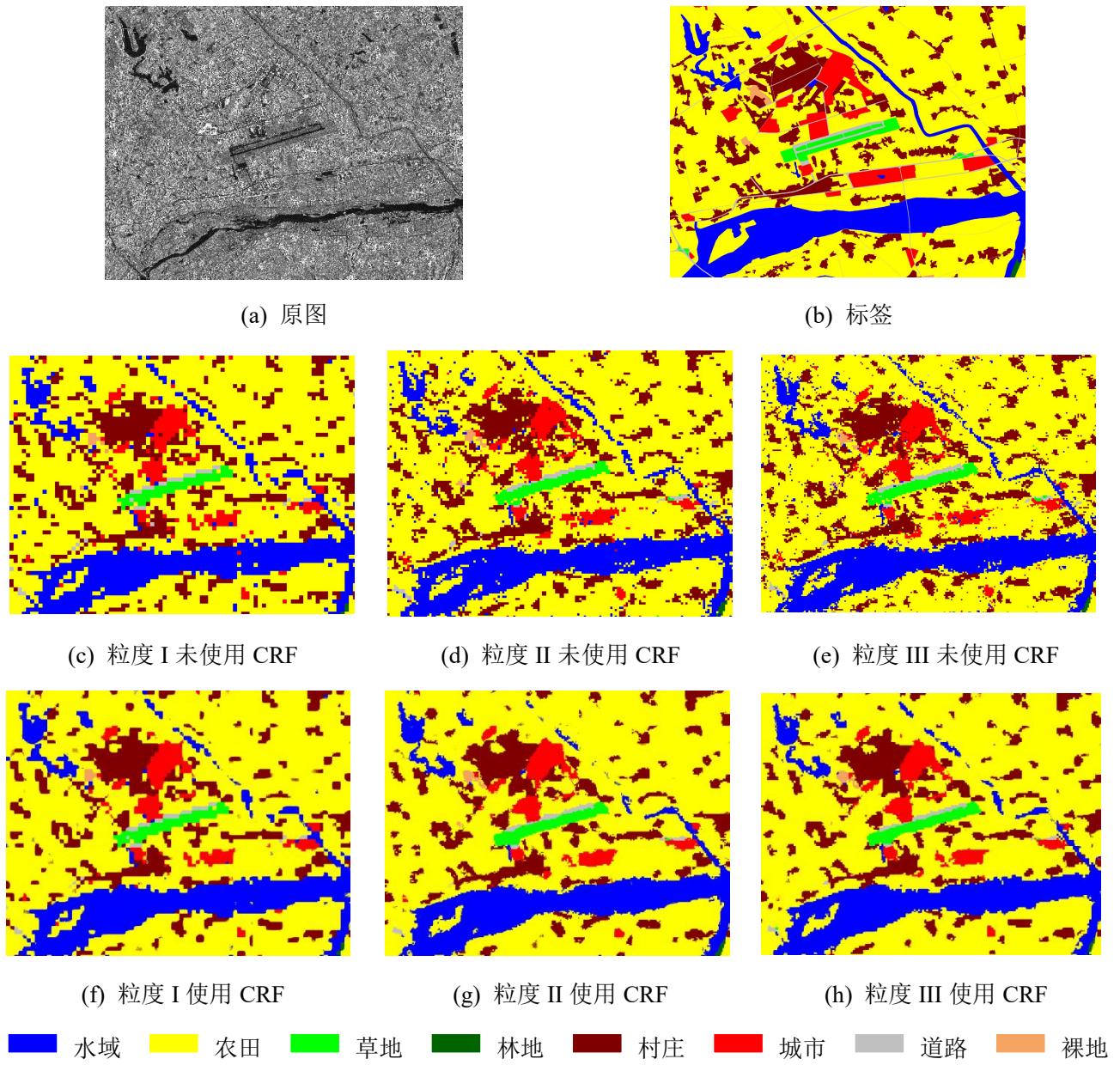


图 4-4 在汉中 1m 数据上的整体测试结果图

4.3 消融实验

在第 2 章中设置了几种不同的数据增强方法，但是由于 SAR 图像的特殊性，这几种增强方法不一定都对模型的训练有正向作用，因此本文探究了在不同数据增强下，模型的性能，并将下游任务和分割任务的指标记录在表 4-5 中。

表 4-5 消融实验表

数据增强 方式	下游任务指标		分割任务指标				
	ACC	Weight-F1score	PA	MPA	MIOU	FWIOU	Kappa
无颜色变化	0.850	0.843	0.818	0.482	0.405	0.692	0.650
无随即反转	0.843	0.835	0.808	0.562	0.383	0.675	0.623
无高斯噪声	0.847	0.840	0.821	0.482	0.406	0.695	0.653
无颜色失真	0.857	0.851	0.807	0.458	0.380	0.675	0.623
无对比对	0.756	0.712	0.693	0.259	0.199	0.510	0.313
均使用	0.842	0.776	0.806	0.683	0.556	0.628	0.653
仅使用对比对	0.901	0.874	0.860	0.714	0.620	0.755	0.735

上述结果是在汉中地区 1m 图像按照裁剪方式VIII裁剪得到的数据集按照粒度 III 预测的结果。采用预训练+微调的方法，即代理任务采用无标记的数据，下游任务采用已划分好的有标记的 10%训练集的数据。消融实验结果表明，使用颜色变化、随机翻转、添加高斯噪声、进行颜色失真等来对数据做增强并不能很好地提高模型在下游任务上的能力，甚至在一些指标上有了严重的下滑现象，因此传统的大众化数据增强方法并不适合 SAR 数据。在实际应用中，未使用数据增强的方法反而表现更优。这一结果表明，在针对 SAR 数据进行处理时，应谨慎选择和设计数据增强方法，而非盲目采用传统的增强手段。

4. 4 和有监督的模型的比较

为了探究本文模型和基于有监督的深度学习模型在预测相同地物块时的效率，本节选取了目前有监督图像分割领域主流的两类模型，Unet 和 Deeplabv3，在本节中，二者以 Resnet 作为特征提取器，将原尺寸为 10011*7710 的图像无重叠地裁剪为若干 128*128 大小地图像块，共计获得 5460 个地物块，分别统计 SWAV-ResSegNet、CRF+SWAV-ResSegNet、Resnet+Unet、Resnet+Deeplabv3 四种模型在其上地各项分割指标，并统计推理时间计算推理效率，得到表 4-6 的结果。

表 4-6 本文模型和有监督模型的比较

模型	PA	MPA	MIOU	FWIOU	Kappa	推理效率 张/s
Resnet50+Deeplabv3	0.559	0.299	0.188	0.416	0.303	15.68
Restnet50+Unet	0.593	0.248	0.162	0.443	0.283	23.74
SWAV-ResSegNet	0.791	0.616	0.498	0.663	0.627	28.00
CRF+SWAV-ResSegNet	0.795	0.674	0.540	0.670	0.631	24.53

据表 4-6 显示，SWAV-ResSegNet 模型在地物块预测任务中表现出色。SWAV-ResSegNet 和 CRF+SWAV-ResSegNet 在所有分割性能指标上显著优于 Resnet50+Deeplabv3 和 Resnet50+Unet，表明 SWAV-ResSegNet 模型在自监督学习任务中的有效性，以及结合 CRF 后能进一步提升分割性能。同时 SWAV-ResSegNet 模型的推理效率最高，达到 28 张/秒，显著优于 Resnet50+Unet 和 Resnet50+Deeplabv3。尽管 CRF+SWAV-ResSegNet 的分割性能略优于 SWAV-ResSegNet，但推理效率稍低。在需要高效率推理的应用场景中，SWAV-ResSegNet 模型提供了最佳的性能和效率平衡。因此，SWAV-ResSegNet 模型在地物块预测任务中表现最优，推荐优先使用，而在对分割精度要求更高的场景下，可以考虑使用 CRF+SWAV-ResSegNet。利用本文方法得到的分割结果在各项指标上都超过了有监督学习，并且在推理效率上也超过了有监督学习。

然而，由于 SWAV-ResSegNet 是以块分类为基础的，因此当分类块很大时，同一个地块包含的地物远远超过一种，此时不宜用该地区内的最大概率的地物类别代表整个地块的类别，因此 SWAV-ResSegNet 方法具有一定的局限性，适用于小场景下的细粒度图像分割。但 SWAV-ResSegNet 仍然可以采用更细的粒度对图像进行近似分割。而 Deeplabv3 和 Unet 这种有监督的图像分割网络以标签为约束，不受地块尺寸影响，在更大的地块上可以展现出更高的精确度。然而正是由于这种模型对于大地块关注多而导致在某些大场景内的小地物容易被忽略。本文在现有模型和数据的基础上，不断调整地块的裁剪尺寸，通过多次迭代获得 Unet 和 Deeplabv3 模型在汉中 1m 图像上最适合的裁剪尺寸（见表 4-7 和表 4-8）。在各个模型最适合的裁剪尺寸上进行进一步的各类地物指标的定量分析，将结果记录在表 4-9 中。

表 4-7 Unet+Resnet50 在不同尺寸图像块下的指标

裁剪尺寸	PA	MPA	MIOU	FWIOU	Kappa
100*100	0.499	0.202	0.118	0.360	0.173
200*200	0.786	0.401	0.325	0.652	0.591
300*300	0.879	0.533	0.475	0.786	0.773
400*400	0.900	0.578	0.522	0.821	0.815
500*500	0.903	0.591	0.532	0.826	0.822
600*600	0.903	0.597	0.538	0.827	0.823
700*700	0.900	0.594	0.532	0.823	0.819
800*800	0.895	0.587	0.519	0.814	0.809
900*900	0.885	0.569	0.501	0.797	0.788
1000*1000	0.877	0.559	0.488	0.785	0.774

表 4-8 Deeplabv3 在不同尺寸图像块下的指标

裁剪尺寸	PA	MPA	MIOU	FWIOU	Kappa
100*100	0.312	0.191	0.083	0.192	0.102
200*200	0.841	0.542	0.443	0.731	0.710
300*300	0.942	0.667	0.608	0.892	0.895
400*400	0.946	0.672	0.615	0.900	0.903
500*500	0.939	0.664	0.602	0.888	0.890
600*600	0.933	0.651	0.590	0.878	0.880
700*700	0.921	0.636	0.573	0.857	0.858
800*800	0.910	0.619	0.551	0.838	0.837
900*900	0.894	0.599	0.523	0.813	0.808
1000*1000	0.880	0.571	0.493	0.790	0.781

根据上述结果，利用 Unet+Resnet50 网络在划分地块大小为 600*600 的时候效果最佳，在同组实验中，各项指标均达到最大值，因此 Unet+Resnet50 组的最佳尺寸块大小为 600*600。同理，Deeplabv3+Resnet50 组的最佳尺寸为 400*400。基于这两个尺寸，本节设计了在最佳裁剪尺寸组内的各类别的指标评估。评估指

标选取了各个类别的 IOU 值、CPA 值、Recall 值和 F1-Score 分数，实验结果见表 4-9。

表 4-9 Unet+Resnet50、Deeplabv3+Resnet50、SWAV-ResSegNet 方法下各类别指标

模型	类别	IOU	CPA	Recall	F1Score
Deeplabv3+ Resnet	水域	0.841	0.906	0.921	0.914
	农田	0.854	0.941	0.902	0.921
	林地	0	0	0	0
	草地	0.644	0.900	0.694	0.784
	道路	0.254	0.298	0.636	0.406
	村庄	0.644	0.752	0.817	0.783
	城市	0.698	0.823	0.820	0.822
	裸地	0	0	0	0
Unet+Resnet50	水域	0.873	0.939	0.936	0.932
	农田	0.879	0.948	0.024	0.936
	林地	0	0	0	0
	草地	0.763	0.842	0.890	0.866
	道路	0.309	0.376	0.635	0.473
	村庄	0.687	0.812	0.817	0.815
	城市	0.793	0.863	0.908	0.885
	裸地	0	0	0	0
SWAV-ResSegNet+ Resnet50	水域	0.711	0.888	0.782	0.831
	农田	0.759	0.857	0.863	0.863
	林地	0.643	0.743	0.783	0.783
	草地	0.628	0.819	0.772	0.772
	道路	0.132	0.150	0.233	0.233
	村庄	0.476	0.675	0.645	0.645
	城市	0.519	0.654	0.683	0.683
	裸地	0.581	0.676	0.735	0.735

上表中，在 SWAV-ResSegNet 模型中，裸地和林地类别有非零的 IOU 值，

而该指标在 Deeplabv3+Resnet50 和 Unet+Resnet50 在裸地上的表现则为 0, 表明 SWAV-ResSegNet 在捕捉地物变化和小地物方面的能力较强。上述结果表明, 自监督模型在识别小地物时表现出了一定的优势。

总体来看, 自监督模型具有较强的小地物识别能力和出色的泛化能力, 能够在不同地物类型上提供稳定的性能。虽然在大地块上的精度略低于有监督模型, 但胜在少量标记, 且在复杂场景中的表现更加均衡, 尤其在细节捕捉和适应不同图像分布方面具有优势。

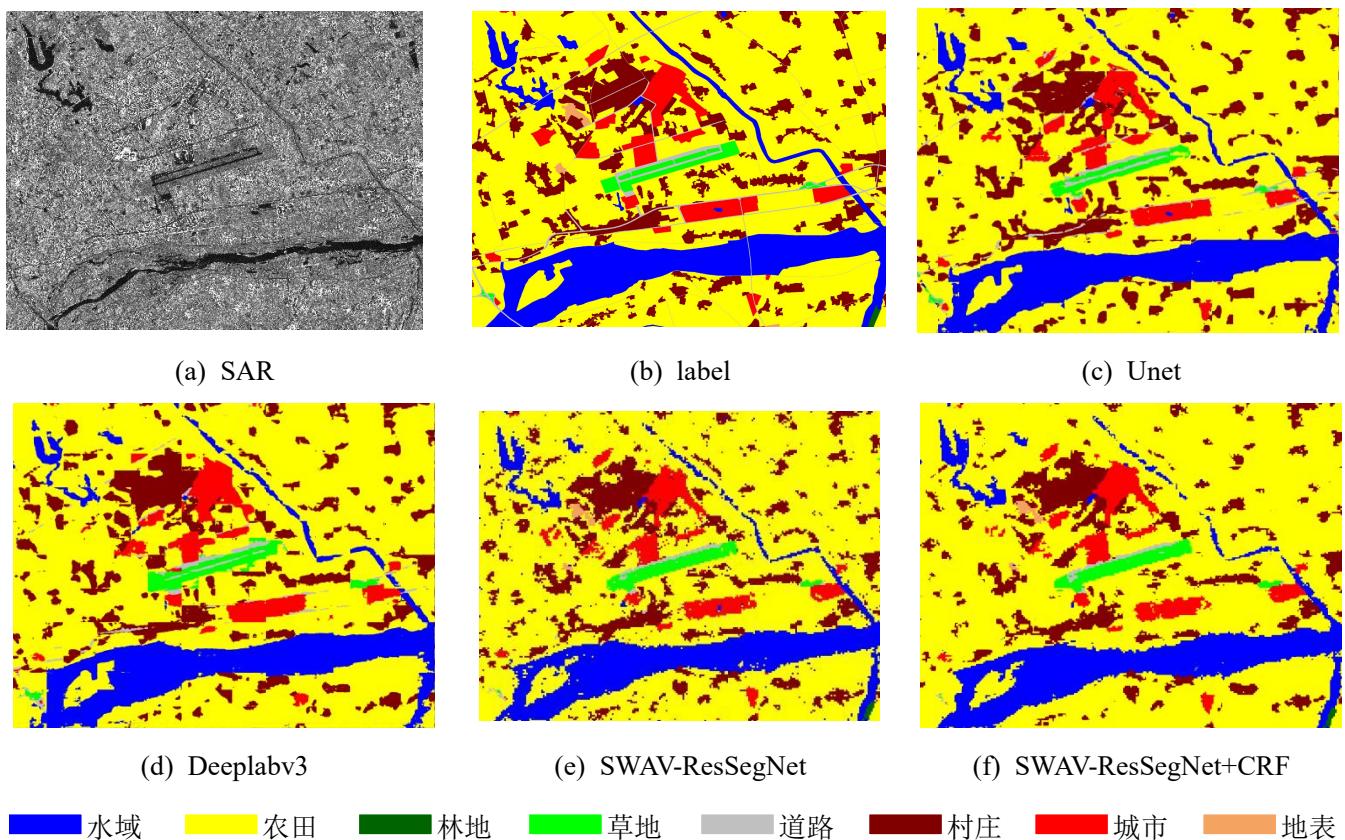


图 4-5 有监督和无监督的测试结果图

4.5 预训练分割任务结果分析

受 Unet+Resnet50 和 Deeplabv3+Resnet50 启发, 本节提 SWRU-Ne 模型。该模型基于代理任务的特征提取器, 将下游任务游分类任务改为利用 Unet 网络进行分割, 将上游的 backbone 作为分割任务的预训练模型进行分割任务, 分别使用训练集的 0.1、0.2、03、0.4、0.5、0.6、07、0.8 作为下游任务的训练数据, 记

作 0.1train~0.8train 测试集不变。数据裁剪大小为 512*512 大小，其他处理方法与 4.4 节中有监督网络数据处理方法保持一致，训练 100 轮后得到各模型在测试集上的 PA 和 Miou。并记录在表 4-10 中。

表 4-10 SWRU-Net 和 Unet+Resnet50 的 PA 和 Miou 比较

指标	0.1train	0.2train	0.3train	0.4train	0.5train	0.6train	0.7train	0.8train	Unet+Resnet50
PA	85.32%	94.3%	95.31%	96.87%	97.06%	97.45%	97.68%	97.94%	96.05%
Miou	66.45%	70.9%	73.01%	76.40%	77.62%	78.54%	78.87%	79.11%	75.45%

通过对比 SWRU-Net 模型与 U-Net+ResNet50 模型的性能数据不难得出，SWRU-Net 只需要 40% 训练集的数据量就能够达到较有监督模型 Unet+Resnet50 更好的效果。即便在较少训练数据的情况下（0.1train~0.2train）能够达到 Miou 指标下 66% 的效果。随着训练数据量的增多，SWRU-Net 能达到 Unet+ResNet50 相近甚至更优的性能。这表明自监督学习能够有效地利用未标注的数据，提高模型在图像分割任务中的性能。

4.6 本章小结

本章通过一系列实验，全面评估了本文提出的基于跨分辨率对比学习的自监督框架 SWAV-ResSegNet 在 SAR 图像地物分类任务中的性能。首先，本章研究了裁剪块尺寸对模型性能的影响。结果表明，适中的裁剪块尺寸能够平衡模型对像素级细节和整体语义信息的捕捉能力，从而在下游任务中实现最佳的分类准确率。同时，本章分析了不同网络层数对分类任务的影响。研究发现，六层全连接网络结构在保持模型复杂度和泛化能力之间取得了最佳平衡，实现了最高的分类准确度。其次，本章探究了不同块分类粒度下的图像分割情况，在本实验中，当块分类粒度为粒度 III 设置的时候效果达到最佳，过大或过小的粒度都会降低分割结果的指标。同时，本章验证了条件随机场（CRF）作为后处理步骤在分割任务中的有效性。CRF 通过考虑像素间的全局关系，显著提高了分割结果的连贯性和准确性。实验结果表明，该框架能够有效地从无标签的 SAR 图像中学习到有用的特征表示，并在分类和分割任务中均取得了出色的表现。然后，本章进行了数据增强方法的有效性消融实验。结果显示，传统的数据增强技术（如颜色变

换、高斯噪声等)并不适用于SAR图像,而未增强的数据在模型训练中表现更佳。此外,本章还比较了自监督模型与当前主流的有监督模型(如Unet和Deeplabv3)的性能。实验结果表明,尽管自监督模型在大地块的分类精度上略逊一筹,但在小场景下的细粒度分割任务中,自监督模型具有更强的识别能力和更高的推理效率。最后,本章根据“预训练+微调”的学习范式将代理任务作为Unet网络的特征提取器,提出了SWRU-net网络并与Unet+Resnet50进行了对比实验。实验结果表明,自监督模型通过自监督学习有效利用未标注数据,在图像分割任务中即使在较少训练数据的情况下也能达到与有监督相近甚至更优的性能。

第 5 章 总结与展望

5.1 本文工作总结

SAR 图像由于其全天时、全天候的观测能力，在环境监测、资源调查、军事侦察等多个领域具有重要的应用价值。然而，SAR 图像的解译面临诸多挑战，包括图像的高分辨率导致数据量大、地物类型复杂多样、标注成本高昂等。这些挑战限制了 SAR 图像在实际应用中的潜力。

为了克服上述挑战，本文提出了一个自监督学习框架，该框架包括代理任务和下游任务两个阶段。在代理任务中，通过对比学习框架，利用多视角样本的聚类和类簇的区分来学习地物特征。这一阶段不依赖于显式的标签信息，能够从无标记的 SAR 图像中挖掘有用的监督信息。在代理任务中，本文设计了一个基于对比聚类的判别式代理任务框架。该框架通过对同类地物的多视角样本进行聚类，形成类簇，并通过区分这些类簇来学习地物特征的本质表示。这一过程涉及到图像增强、特征提取、原型匹配和损失函数设计等关键步骤。在下游任务中，本文设计了一个基于深度学习的分类器，用于对 SAR 图像中的图像块进行分类。该分类器利用了在代理任务中学到的特征表示，通过自适应平均池化层和全连接层来实现分类。本文工作总结如下：

(1) 探究了自监督在 SAR 图像上的有效性。本文通过对比学习策略，成功地从大量未标记的 SAR 图像中提取出有助于地物分类的特征，不仅证明了自监督学习在 SAR 图像处理中的有效性，而且为处理大规模未标记和少标记数据集提供了一种可行的解决方案。

(2) 传统数据增强对 SAR 的作用。通过消融实验结果发现传统数据增强技术(如颜色变换、高斯噪声等)在 SAR 图像上的效果并不如预期。因此在 SAR 图像处理中应更加谨慎地选择和设计数据增强策略。

(3) 实现了一个集成了多种算法的图像分割解译系统。通过该系统可以得到上传图像中各地物的分布和比例，可以提高 SAR 图像解译的自动化水平，减少人工干预，提高解译效率和准确性。

5.2 工作展望

本文在合成孔径雷达图像分类领域取得了显著成果，但仍存在一些局限性和未来改进的空间。以下是对未来研究方向的详细展望：

(1) 算法泛化能力的提升。在本文中，虽然提出的自监督学习框架在特定数据集上表现出色，但算法在不同成像条件下的泛化能力尚未得到充分验证。未来的研究可以探索算法在多样化 SAR 图像数据上的表现，包括不同传感器、不同成像模式、不同地理区域的数据。这需要构建更加全面和具有代表性的 SAR 图像数据集，以及设计更加鲁棒的特征学习策略。

(2) 多尺度特征融合策略的改进。在本文中，虽然采用了多尺度特征融合策略，但未来的研究可以进一步改进这一策略，包括设计更加有效的特征融合方法，以及探索不同尺度特征的重要性。细粒度分类的探索。SAR 图像中的地物类型可能非常复杂，未来的研究可以探索细粒度的分类方法，以区分更加详细的地物类别。这需要设计更加复杂的网络结构和学习策略，以捕获更加精细的地物特征。

(3) 自监督和无监督在 SAR 地物分类上的应用。本文的研究表明，自监督学习是 SAR 图像地物分类的一个有前景的研究方向。未来的研究可以在此基础上进一步探索，如利用半监督学习框架结合少量标记数据进行训练，或者开发无监督学习方法来完全摆脱对标记数据的依赖。。除了自监督学习，未来的研究还可以探索半监督学习和无监督学习在 SAR 图像分类中的应用。这包括利用少量标记数据进行训练的半监督方法，以及完全无需标记数据的无监督方法。

(4) SAR 的数据增强。在本文消融实验中发现传统数据增强的效果并不能带来较大的提升，反而增加了计算量。鉴于传统数据增强方法在 SAR 图像上的效果有限，未来的工作需要探索和开发更适合 SAR 图像特性的数据增强技术，可能包括模仿不同成像条件下的 SAR 图像效果，或者模拟不同季节和时间的成像差异。此外 SAR 图像的获取过程受到物理因素的影响，如地形、大气条件等。未来的数据增强可以结合物理模型，生成更加真实的 SAR 图像变体，以提高模型对于实际应用中多变条件的适应性。

(5) 开发的图像分割解译系统还具备一定的可扩展性，可以根据需要集成更多的算法和功能，以适应不同的应用场景和需求。

参考文献

- [1] 焦李成,侯彪,唐旭,等.人工智能,类脑计算与图像解译前沿[M].陕西: 西安电子科技大学出版社,2020.
- [2] 邓云凯, 赵凤军, 王宇. 星载 SAR 技术的发展趋势及应用浅析[J]. 雷达学报, 2012, 1(1): 1–10.
- [3] MOREIRA A, PRATS-IRAO LA P, YOUNIS M, etc. A tutorial on synthetic aperture radar[J]. IEEE Geoscience and Remote Sensing Magazine, 2013, 1(1): 6–43.
- [4] 魏钟铨. 合成孔径雷达卫星[M]. 北京: 科学出版社, 2001.
- [5] REIGBER A, SCHEIBER R, JAGER M, etc. Very-high resolution airborne synthetic aperture radar imaging: Signal processing and applications[J]. Proceedings of the IEEE, 2013, 101(3): 759–783
- [6] 范剑超, 王德毅, 赵建华, 等. 高分三号 SAR 影像在国家海域使用动态监测中的应用[J]. 雷达学报, 2017, 6(5): 456–472.
- [7] 邓云凯,张衡,范怀涛,等.面向综合环境监测的星载 SAR 技术发展[J].中国空间科学技术,2023,43(02):32-46.
- [8] 黄登冕,张聪,姚晓军,等.矿山环境遥感监测研究进展[J].遥感技术与应用,2022, 37(05):1043-1055.
- [9] 王樱洁,王宇,禹卫东,等.See-Earth: 高频时序多维地球环境监测 SAR 星座[J].雷达学报,2021,10(06):842-864.
- [10] 刘羽.合成孔径雷达在林业调查中的应用概述[J].内蒙古林业调查设计, 2019, 42(3):3.
- [11] 张少伟.基于多源数据的内蒙古大兴安岭林区森林资源变化监测研究[D].北京: 中国林业科学研究院,2019
- [12] 张康宇.基于 C 波段 SAR 的海面风场反演方法与近海风能资源评估[D].浙江: 浙江大学,2019.
- [13] 贾慧,杨柳,郑景飚.无人机遥感技术在森林资源调查中的应用研究进展[J].浙江林业科技,2018,38(04):89-97.

- [14] 魏英杰.发展海洋监视技术铸就大国海上天眼——《星载 SAR 与 AIS 综合的海洋目标信息处理技术》书评[J].雷达学报,2017,6(03):324-325.
- [15] 程玉鑫,袁凌峰.机载 SAR 发展现状[J].电子测试,2016(08):19+6
- [16] 陈艺天,李斌兵.合成孔径雷达的主要发展方向和军事应用[J].科技信息,2011(33):73-74.
- [17] 于飞,隋正伟,邱凤婷,等.SAR 图像智能解译样本数据集构建进展综述[J].网络安全与数据治理,2023,42(S1):97-105.
- [18] 叶子琦,肖夏阳,李刚田,等.面向小样本的 SAR 图像飞机目标分类方法[J].遥感信息,2023,38(06):60-67.
- [19] 吴游宇,刘德强,余飞,等.融合 SAR 分支深度学习的隧道病害识别与成像[J].公路,2023,68(12):320-328.
- [20] 张睿,王梓祺,李阳,等.任务感知的多尺度小样本 SAR 图像分类方法[J].计算机科学,2024: 1 - 10..
- [21] 黄岩,任沂斌.基于双极化 SAR 图像的 U-Net 海冰多分类模型[J].海洋与湖沼,2023,54(06):1551-1563.
- [22] 周一鸣,滕旭阳.基于多尺度混合注意力机制的 SAR 图像房屋分割方法[J].电
脑知识与技术,2023,19(17):23-26.
- [23] 黄岩.基于深度学习和 SAR 遥感影像的北极波弗特海海冰分类研究[D].北京:
中国科学院大学(中国科学院海洋研究所),2023.
- [24] 吕小玲,仇晓兰,俞文明,等.基于无监督域适应的仿真辅助 SAR 目标分类方法
及模型可解释性分析[J].雷达学报,2022,11(01):168-182.
- [25] 汪荃,陈军军,董翔.基于分割耦合规则集的城乡遥感地物分类[J].中南农业科
技,2024,45(03):79-82+89.
- [26] 符婷,陈思伟.SAR 图像方向性上下文协方差矩阵构建方法及地物分类应用
[J].遥感学报,2024,28(03):730-746.
- [27] 王懿泽,孙吉利,闫成杰,等.基于超像素与 LightGBM 的极化 SAR 图像地物分
类[J].中国科学院大学学报,2023,40(05):658-669.
- [28] 魏鹏超,方向忠.基于随机森林的极化 SAR 图像地物分类研究[J].信息技术,2023,
47(02):81-85.

- [29] 赵泉华,冯林达,李玉.基于最优极化特征组合的 SAR 影像湿地分类[J].地球信息科学学报,2021,23(04):723-736.1
- [30] 李渊.复杂地形 SAR 图像分类方法研究[D].河南: 河南大学,2021.
- [31] 王贤圆.极化合成孔径雷达图像特征表示与目标分类方法研究[D].四川: 电子科技大学,2021.
- [32] 赵文慧,俞宣,杨中傲,等.利用多尺度特征的极化 SAR 深度学习分类[J].现代雷达,2023,45(04):48-54.
- [33] 谢雯,王若男,羊鑫,等.融合深度可分离卷积的多尺度残差 UNet 在 PolSAR 地物分类中的研究[J].电子与信息学报,2023,45(08):2975-2985.
- [34] 刘鹏,谢春华,安文韬,等.改进 AlexNet 的高分三号全极化海岛地物分类[J].遥感信息,2021,36(05):142-147.
- [35] 楚博策,高峰,帅通,等.基于特征图集合的遥感影像深度学习地物分类研究[J].无线工程,2022,52(04):630-637.
- [36] 刘浩然,刘秀清,王春乐.基于随机森林和超像素的极化 SAR 图像分类[J].国外电子测量技术,2021,40(09):29-35.
- [37] 孙盛,董晓冬,水文泽,等.基于深度卷积网络的 SAR 图像地物分类研究[J].计算机仿真,2021,38(07):456-460.
- [38] 张博涵,徐晓敏.基于深度迁移学习的多标签遥感影像地物分类[J].北京信息科技大学学报(自然科学版),2023,38(06):95-100.
- [39] Hinton G E , Salakhutdinov R R .Reducing the Dimensionality of Data with Neural Networks[J].Science, 313(5786):504-507.
- [40] Attneave, Fred, M. B. and Donald Olding Hebb. “The Organization of Behavior: A Neuropsychological Theory.” (1949).
- [41] Schroff F , Kalenichenko D , Philbin J .FaceNet: A Unified Embedding for Face Recognition and Clustering[J].Proceedings of the IEEE conference on computer vision and pattern recognition,2015,10(1109):815-823
- [42] Chen T , Kornblith S , Norouzi M ,etc..A Simple Framework for Contrastive Learning of Visual Representations[J]. PMLR, 2020,1597-1607

- [43] He K , Fan H , Wu Y ,etc..Momentum Contrast for Unsupervised Visual Representation Learning[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).IEEE,2020,9726-9735
- [44] Grill J B , Strub F ,Altché, Florent,etc..Bootstrap Your Own Latent: A New Approach to Self-Supervised Learning[J]Arxiv,2020,2006(07733).
- [45] Caron M , Misra I , Mairal J ,etc.Unsupervised Learning of Visual Features by Contrasting Cluster Assignments[J]Arxiv,2020,2006(09882)
- [46] Coates, A., Ng, A. Y., Lee, H. . An analysis of single-layer networks in unsupervised feature learning[J] In Proceedings of the fourteenth international conference on artificial intelligence and statistics,2011,215-223.
- [47] A. Krizhevsky, I. Sutskever and G. E. Hinton.ImageNet classification with deep convolutional neural networks[J],Communications of the ACM,2017,60(6):84-90.
- [48] Krizhevsky, A., & Hinton, G. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009,1(4): 7.
- [49] LeCun, Y., Cortes, C., & Burges, C. (1998). The MNIST database of handwritten digits. <http://yann.lecun.com/exdb/mnist/>.
- [50] Deng J , Dong W , Socher R ,etc..ImageNet: A large-scale hierarchical image database[J].Proc of IEEE Computer Vision & Pattern Recognition, 2009:248-255.
- [51] Lin T Y , Maire M , Belongie S ,etc.Microsoft COCO: Common Objects in Context[J].Springer International Publishing, 2014.740-755.
- [52] Liu Z , Luo P , Wang X ,etc..Deep Learning Face Attributes in the Wild [J].IEEE international conference on computer vision 2016.3730-3738.
- [53] Xiao H , Rasul K , Vollgraf R .Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms[J]ArXiv,2017,1708(07747).
- [54] Netzer, Y., Wang, etc. Y. (2011). Reading digits in natural images with unsupervised feature learning. [J]NIPS workshop on deep learning and unsupervised feature learning,2(2011): 1-9.

- [55] Lake, B. M., Salakhutdinov, etc. Human-level concept learning through probabilistic program induction.[J]Science,2015,350(6266):1332-1338.
- [56] Bossard, L., Guillaumin, etc.. Food-101-mining discriminative components with random forests. In European conference on computer vision,2014,446-461.
- [57] Z. Wu, Y. Xiong, S. X. Yu etc. Unsupervised Feature Learning via Non-parametric Instance Discrimination.2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 3733-3742,
- [58] Tian Y , Krishnan D , Isola P .Contrastive Multiview Coding[J]ECCV,2019.
- [59] 黄伟锋,张甜,常东良,等.基于多视角融合的细粒度图像分类方法[J].信号处理,2020,36(09):1607-1614.
- [60] 苏可政,肖燕珊,刘波.基于实例的多视角多标签学习算法[J].计算机科学与应用, 2022, 12(4):12.
- [61] Li X , Zhang G , Cui H ,etc.MCANet: A joint semantic segmentation framework of optical and SAR images for land use classification[J].Int. J. Appl. Earth Obs. Geoinformation, 2022, 106:102638.
- [62] Simonyan K , Zisserman A .Very Deep Convolutional Networks for Large-Scale Image Recognition[J].Computer Science, 2014.1409:1556.
- [63] Szegedy C , Liu W , Jia Y ,etc. Going Deeper with Convolutions[J].IEEE Computer Society, 2014.1-9
- [64] He K , Zhang X , Ren S ,etc.Deep Residual Learning for Image Recognition[J].IEEE, 2016,770-778
- [65] Goodfellow I , Pouget-Abadie J , Mirza M ,etc. Generative Adversarial Networks[J].MIT Press, 2014,(2):2672 - 2680
- [66] He K , Chen X , Xie S ,et al.Masked Autoencoders Are Scalable Vision Learners[J]CVPR,2021,15979-15988

- [67] 陆妍如,毛辉辉,贺琰,等.基于 Deeplab v3+的高分辨率遥感影像地物分类研究[J].地理空间信息,2022,20(06):1-6.
- [68] 王云艳,罗冷坤,王重阳.Deeplab 网络的极化合成孔径雷达图像分类[J].测绘科学,2020,45(06):110-117.
- [69] 刘琦,杨小华,陈柏行,等.基于 Transformer 的成都市地物分类技术[J].地理空间信息,2023,21(12):51-53.
- [70] 李薇,樊彦国,周培希.改进 Transformer 的高光谱图像地物分类方法——以黄河三角洲为例[J/OL].自然资源遥感:1-9[2024-05-12].
- [71] 高峰,孟德森,解正源,等.基于 Transformer 和动态 3D 卷积的多源遥感图像分类[J].北京航空航天大学学报,2024,50(02):606-614.
- [72] 申传庆,王凯,王文杰.基于 ResNet-UNet 的地表覆盖自动分类技术研究[J].地理空间信息,2023,21(06):21-23+27.
- [73] He X , Zhou Y , Zhao J ,etc.Swin Transformer Embedding UNet for Remote Sensing Image Semantic Segmentation[J].IEEE Transactions on Geoscience and Remote Sensing, 2022, 60:1-15.
- [74] Ye H , Liu S , Jin K ,etc.CT-UNet: An Improved Neural Network Based on U-Net for Building Segmentation in Remote Sensing Images[J].ICPR, 2021,166-172.
- [75] Yan, C.; Fan, X.; Fan, J.; etc. Improved U-Net Remote Sensing Classification Algorithm Based on Multi-Feature Fusion Perception. Remote Sens. 2022, 14, 1118.
- [76] Wang Z , Zou N , Shen D ,etc.Non-Local U-Nets for Biomedical Image Segmentation[J].Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(4):6315-6322.
- [77] 马文琪,石颉,吴宏杰.深度卷积神经网络语义分割综述[J/OL].微电子学与计算机,2023,(09):55-64.

致 谢

在论文付梓之际，我万分感谢所有老师、同学、朋友、家人们在我本科期间给予的教导、关心、支持和帮助。感谢华水，感谢信息工程学院，感谢 2020185 班。如果可以，我希望未来还有机会回到华水，继续深耕，发散光热。特别感谢杨阳蕊老师一直以来无微不至的关怀、照顾，在四年的亦师亦友的生活中我得到了许多成长。还要感谢西安电子科技大学人工智能学院的任仲乐老师提供给我的课题。没有你们，我无法完成这篇论文，感谢您们！

附:致谢名单(按姓氏笔画排序，排名不分先后)

马 斌	王 畅	王晓东	王海周	王朝霞
王聚源	王蕊珠	孔祥瑞	冯 岭	朱亚萍
朱 凯	朱政霖	任中云	任仲乐（西电）	刘 宁
刘丽华	闫 修	闫淑卿	关利南	许 梅
孙全红	孙志鹏	孙明明	孙树勋	李帅星
李雨洋	李 果	李 洋	李 勇	李素萍
李浩南	李海瑞	李 雷	李 鹏	杨阳蕊（导师）
杨雨倩	杨学颖	吴慧欣	余桂霞	宋晓燕
张月莹	张庆锋	张红艳	张良有	张 昭
张 速	张瑞霞	张 纶	陈思思	欧阳宁康
欧 瑞	周 丹	周美慧	赵 磊	郝梓同
娄亚鑫	袁进霞	贾小标	贾 敏	高勋刚
高淑芳	郭彦溥	黄士杰	菅朋朋	曹海松
康 凯	韩光辉	韩红玲	谢昊洋	雷冬雪
蔡 昊	裴嘉丰	翟丽萍	薛素静	魏志楠

