

摘要

耕地与经济和民生息息相关，是不可或缺的重要资源。然而，在我国城市化和工业化的快速发展中，耕地资源正逐步被占用，威胁到“耕地红线”政策的目标。因此，快速且准确的对耕地信息进行提取对保证农业可持续发展和国家粮食安全有着十分重要的意义。

遥感影像能够提供地表物体的丰富信息，遥感影像分类是耕地信息提取的重要手段。随着遥感影像质量的不断提升，高分辨率遥感影像带来了丰富的地物信息，但对其高精度、高效率的分类提取也是一个极大的挑战。深度学习中的语义分割算法凭借着较高的分类精度，以及较强的自动学习能力，逐渐成为处理高分辨率遥感影像的重要手段之一。本文将深度学习语义分割模型应用于高分辨率遥感影像耕地信息提取，探索适用于耕地提取的高精度、高效率方法。

经典语义分割模型通常具有大量的训练参数，进行影像分割时存在分割不准确和效率低的问题，且大多数耕地提取的方法只适用于某些特定的数据集或者特定的研究区，模型的泛化能力较差。针对这些问题，本文的主要研究工作和结论如下：

(1) 对 DeepLabv3+ 模型改进，得到 MST-DeepLabv3+ 模型。使用轻量级网络 MobileNetV2 替代 DeepLabv3+ 模型的特征提取网络 Xception，以减少模型参数量，提高训练速度；对比不同的注意力机制模块，选取通道注意力机制 SENet 加入模型，弥补轻量级网络带来的精度损失，提高遥感影像语义分割的精度；引入迁移学习，将 ImageNet 数据集上训练的特征提取网络用做预训练模型，增强模型获取特征的能力，提升网络分割精度。在 ISPRS Vaihingen 数据集与 GID 数据集上进行多模型对比实验的结果显示，MST-DeepLabv3+ 模型具有优秀的分割性能，能够有效解决经典模型分割结果中存在的不同程度误分、漏分及过度分割问题，改善地物边界及轮廓分割不准确的现象，为后续耕地信息提取提供帮助。此外，消融实验也说明了改进方法对模型的影响。

(2) 耕地信息提取研究。以周口市太康县作为研究区，使用高分一号遥感卫星的 PMS 影像制作太康耕地数据集。为进一步探索适用于耕地提取的方法，在 MST-DeepLabv3+ 模型的基础上，提出改变迁移模型的 MST-DeepLabv3+(GID) 与改变影像波段组合的 MST-DeepLabv3+(NirRG) 两种方法。将这两种方法与 MST-DeepLabv3+、

PSPNet、UNet 和 DeepLabv3+模型应用于太康耕地数据集。实验结果表明，本文提出的 MST-DeepLabv3+(GID)、MST-DeepLabv3+(NirRG)与 MST-DeepLabv3+三种方法的各项评价指标值均高于 PSPNet、UNet 和 DeepLabv3+模型，可以很好的完成耕地区域完整分割，具有平滑分割边界，消除粘连现象的能力，能够高效且高精度的实现耕地信息提取。其中 MST-DeepLabv3+模型精度最高，整体效果更好。

实验证实了本文提出的三种模型均能有效的提取耕地信息，能够为今后的高分辨率遥感影像耕地信息提取提供参考。其中 MST-DeepLabv3+模型不仅在耕地提取任务中具有最佳的分割性能，在其他数据集上也能得到较好的分割结果，泛化能力强。

关键词: 高分辨率遥感影像，语义分割，轻量级网络，注意力机制，迁移学习

目 录

摘 要.....	I
ABSTRACT.....	III
1 绪论.....	1
1.1 研究背景与意义.....	1
1.2 国内外研究现状.....	2
1.2.1 基于机器学习的遥感影像分类方法.....	2
1.2.2 基于深度学习的遥感影像分类方法.....	3
1.2.3 耕地提取研究现状.....	5
1.3 研究内容与研究方法.....	7
1.4 本文结构安排.....	8
2 遥感影像语义分割模型.....	10
2.1 经典模型.....	10
2.1.1 FCN.....	10
2.1.2 SegNet.....	11
2.1.3 UNet.....	11
2.1.4 PSPNet.....	12
2.1.5 DeepLabv3+.....	13
2.2 评价指标.....	14
2.3 本章小结.....	15
3 研究区与实验数据.....	16
3.1 研究区概况.....	16
3.1.1 研究区地理位置与自然特征.....	16
3.1.2 研究区耕地情况.....	16
3.2 实验数据.....	17
3.2.1 ISPRS 数据集.....	17
3.2.2 GID 数据集.....	18
3.2.3 高分一号遥感影像.....	19
3.3 本章小结.....	20
4 MST-DeepLabv3+：一种改进的遥感影像语义分割方法.....	21

4.1 改进机制	21
4.1.1 主干网络	22
4.1.2 注意力机制	23
4.1.3 迁移学习	24
4.2 实验配置	24
4.3 分类结果对比分析	25
4.3.1 数据预处理	25
4.3.2 ISPRS Vaihingen 数据集分类结果与分析	25
4.3.3 GID 数据集分类结果与分析	28
4.4 消融实验及模型参数对比	31
4.5 本章小结	31
5 基于 MST-DeepLabv3+的周口市太康县耕地信息提取研究	33
5.1 太康耕地数据集建立	33
5.2 耕地提取方法与结果分析	35
5.2.1 耕地提取方法	35
5.2.2 耕地提取整体精度多模型对比分析	35
5.2.3 耕地提取各类别精度多模型对比分析	36
5.2.4 耕地提取可视化结果对比	37
5.3 本章小结	39
6 总结与展望	41
6.1 本文总结	41
6.2 创新点	42
6.3 研究展望	42
参考文献	44
致 谢	50
攻读学位期间发表的学术论文目录	51

1 绪论

1.1 研究背景与意义

民以食为天，耕地是粮食生产的重要载体，也是影响经济和农业可持续发展的主要因素。然而我国城市化和工业化的发展，快速消耗着土地资源，使得我国耕地资源总量正逐步减少。为保证国家粮食安全，维护社会稳定，中国于 2013 年提出坚守 18 亿亩“耕地红线”政策。2009 年底第二次全国土地调查数据显示我国耕地总量达 203077 万亩^[1]，到 2019 年底，第三次全国土地调查显示耕地为 191792.79 万亩^[2]，十年间减少了 11284.21 万亩，耕地总量的变化趋势正逐步威胁到“耕地红线”政策的目标。因此能否准确及时地获取耕地面积对农业发展评估与国家宏观调控决策具有重要意义。

传统的土地利用管理与耕地面积统计方法基于实地调查信息采集的方式，数据更新滞后且采集过程极大地消耗人力、物力。1972 年，美国发射了第一颗遥感卫星，也就是如今更名后的陆地卫星 Landsat，提供了应用遥感技术对地观测的新方式。

遥感技术具有监测范围广、数据采集快、获取信息量大的优点，能够更好地帮助人们探索 and 了解周围的世界。遥感监测满足全面并精准获取地面各类信息的条件，逐渐取代了传统的人工区域调查方法，广泛应用于土壤研究^[3]、地质工程^[4]、国土资源^[5]等相关领域。遥感技术是进行耕地面积提取的主要手段。二十世纪八十年代初，我国开始在土地资源管理工作中应用遥感技术^[6]。

在遥感技术早期应用阶段，研究人员利用卫星和航拍图像研究地球表面，人工判读是分析这些图像的主要方法。随着对地遥感观测平台的发展，遥感影像数据量大幅增加，影像分辨率得到有效提高，依旧依靠人工目视解译，不仅耗费大量人力财力，效率低下，而且识别精度过分依赖解译者的主观经验，因此开发利用计算机自动且有效的影像分析方法变得越来越重要。

机器学习的发展给遥感图像分类带来了重大变化，能有效解决“同物异谱”和“异物同谱”现象^[7]。然而，传统机器学习算法仅使用像素级信息作为对遥感图像进行分类的基础，忽略了空间和上下文信息^[8]，不能详细的反映数据深层细节。而且当处理大量层次复杂、范围较大的遥感图像时，容易出现拟合现象，分类精度较差。随着遥感卫星发展，从高分辨率遥感影像中提取耕地信息正日益成为现代农业数字化、信息化的重要

途径。传统机器学习算法对包含复杂地物信息的高分辨率遥感影像信息提取的效果不好，应用需求和遥感数据的增加对耕地语义分割提出了更高的要求。

深度学习在计算机视觉应用中的成功，使得以卷积神经网络（Convolutional Neural Network, CNN）^[9]为基础的方法逐渐应用于遥感影像解译领域，为遥感影像耕地提取识别带来了巨大的突破。CNN 在遥感领域的应用使得影像分类准确度大幅提升。面对场景复杂的遥感影像，进行耕地信息提取时，能够自适应地提取浅层和深层特征，无需人工参与。经过特征融合，能够高效挖掘丰富的语义信息，使分割任务的准确性得到显著提高。但经典的 CNN 模型在批处理过程中存在冗余计算，会产生较高的内存消耗，造成较低的分割效率^[10]。目前已存在的 CNN 模型变体虽然取得了卓越的成果，但对高分辨率遥感影像耕地信息提取时，仍存在类别粘连、区域不完整，分割效率低等问题。

本文在高分辨率遥感影像的基础上，探究不同深度学习模型对耕地信息提取的能力，并对已有的深度学习模型进行改进以提高耕地提取的精度和效率。

1.2 国内外研究现状

遥感影像地物分类根据多种特征将影像中的每个像素划分为不同类型，是获取丰富的土地利用信息的主要手段。遥感影像耕地信息提取的研究等同于遥感影像的分类识别研究，耕地提取能力也就是模型对遥感影像耕地类别的分割能力。本节先对遥感影像分类方法的发展进行总结，然后对使用遥感技术进行耕地提取的研究进行阐述。

1.2.1 基于机器学习的遥感影像分类方法

机器学习以概率统计为理论根据，为遥感图像的分类方法提供了一些可行性思路，主要有非监督分类与监督分类^[11]。

非监督分类又称聚类分析，主要根据像元之间的特征相似度进行归类合并，不需要先验知识，自动化程度较高。主要有 K 均值（K-means, KM）算法^[12]以及迭代自组织数据分析算法(Iterative Self-organizing Data Analysis Techniques Algorithm, ISODATA)等^[13]。钟燕飞等人^[14]探究了适合遥感影像 K 均值聚类的最优初始化方法。Kirsi Karila 等人^[15]采用 ISODATA 算法区分不同的耕地农作物、水产养殖和果树类别，获得较高的总体准确度。李鹏等人^[16]使用不同方法对高分二号影像中的农业信息进行提取，根据对比结果发现非监督分类的优势在于原理简单，但存在信息提取精度偏低的劣势。

监督分类是在已知样本区类别信息的基础上，有效地判别非样本区。由于利用了先验知识，相较于非监督分类往往具有更高的分类精度^[16]。监督分类方法包括支持向量机

(Support Vector Machine, SVM)^[17]、条件随机场 (Conditional Random Field, CRF)^[18]、决策树 (Decision Tree, DT)^[19]、随机森林 (Random Forest, RF) 等^[20]。支持向量机是结合核函数和优化理论的线性分类器,通过有限的样本信息寻找最优折衷点,从而解决复杂数据的分类问题^[21]。为了检测遥感图像中农业及其他多类特征, Li 等人^[22]提出将颜色特征与支持向量机分类器相结合的方法。Volpi 等人^[23]使用结构化支持向量机对城市场景进行分类。张浩^[24]以高分二号遥感图像为基础,研究了 SVM 在建筑物识别中的应用。决策树的分类方式基于归纳推理,以分层分类思想作为指导原则,对较为繁杂的多分类问题进行简化^[11]。郭力娜等人^[25]建立决策树分类规则,提高中分影像土地利用分类效率。Xu 等人^[26]利用决策树生成的先验知识为郑州市土地利用研究提供支持。随机森林由许多决策树集成而来,通过随机抽取部分样本进行决策树建模,组合多棵决策树的预测结果,投票得出最终预测结果^[27]。条件随机场是判别式模型算法,具有较强的上下文信息建模能力^[28]。宁淼福^[29]结合随机森林、集成学习和全连接条件随机场在高分辨率影像上提取招弧角。Sun 等人^[30]使用随机森林集成学习技术对遥感影像进行分类,然后使用改进的条件随机场增强分类结果,与非基于 CNN 的方法相比,在 ISPRS 数据集上实现了最高的准确率,但与基于 CNN 的方法相比,仍有明显差距。

监督分类与非监督分类方法可以根据具体分类任务进行选择,二者的结合使用,也可能达到更好的效果。刘咏梅等人^[31]针对黄土丘陵沟壑地区的地形情况,采用监督分类与非监督分类结合的集成方法,较好地提取了研究区域的坡耕地信息。

大多数基于传统机器学习的遥感图像解译算法都采用特征提取、特征分析,对于特定的场景和数据集,解译效果良好^[32]。然而,传统机器学习算法提取特征的能力有限^[8],无法准确捕捉输入数据的细节。当待处理遥感图像的背景层次复杂,且目标尺度变化较大时,模型分割精度易受到影响,会出现欠拟合或过拟合现象。

1.2.2 基于深度学习的遥感影像分类方法

高分辨率遥感影像信息复杂多样、特征丰富、尺度大,可以提供丰富的特征信息,精细地呈现地物的空间结构和纹理特征^[33]。然而高分辨率遥感影像带来更多数据与信息的同时,也给遥感影像的解译带来了诸多挑战,如解译成本高,信息提取耗时费力,不能满足当前社会经济发展对资源信息快速提取与更新的迫切需求。

语义分割是像素级的分类任务,图像中的各种类别由不同像素组合而成,语义分割能够对不同像素赋予相应的类别标签^[34]。相比于传统的机器学习方法,语义分割可以通

通过对复杂图像深层特征的自动学习，达到提高分割精度的目的。

基于深度学习的遥感影像语义分割主要利用卷积神经网络（CNN）及其变体模型自动学习图像特征并进行信息分类。CNN 由多层组成，主要包括卷积层、池化层和全连接层，能够在不同层次上学习和提取图像特征。CNN 算法在遥感影像分类中取得了优异的结果。朱祺夫等人^[35]使用 GoogLeNet 模型提取城市建筑用地的影像信息，并与传统的支持向量机作对照，验证了深度学习方法具备对遥感影像地物提取的优势。Jadhav 等人^[36]采用 ResNet101 网络对高分辨率遥感影像的土地覆盖和作物类型进行语义分割，实现了主要作物的准确分类。Kussul 等人^[37]证明了具有卷积神经网络集合的架构相较于传统的全连接多层感知器(MLP)的架构，能够更好地区分遥感影像作物类型。

与传统的机器学习技术相比，CNN 显著提高了遥感图像的分割精度，但经典的 CNN 模型在进行卷积和池化时会丢失部分图像信息，不断积累的冗余计算会造成分割效率较低^[10]。Long 等人^[38]用卷积层代替了 CNN 的全连接层，并采用反卷积操作进行上采样，提出了全卷积网络（Fully Convolutional Networks, FCN）。Fu 等人^[39]使用空洞卷积优化 FCN 模型，并采用条件随机场对分割数据进行后处理，对高分二号影像处理时，分割精度得到显著提高。为了提高算法精度并减少噪声的影响，Chen 等人^[40]以 shortcut-block 结构为基础，使用重叠 SNFCN 和 SDFCN 语义分割框架的方法，显著提高了城市地区遥感影像的分割精度。

然而，在上采样时 FCN 没有考虑像素之间的关系，分割结果仍然粗糙。为了提高分割精度，大量以 FCN 为改进基础的模型相继问世。Ronneberger 等人^[41]提出具有 U 型结构的编码器—解码器网络 UNet，使用跳跃连接融合深层抽象特征和浅层细节特征，能够组合比 FCN 更多的低级特征，从而实现了更加精确的像素边界定位，显著提升了分割精度。Singh 等人^[42]将 UNet 网络用于印度某地的 Sentinel-2 影像中，绘制不同农业土地利用类型。Badrinarayanan 等人^[43]提出的 SegNet 网络同样是编码器—解码器结构。不同于 FCN 网络直接复制特征图，其解码器的上采样过程采用池化索引的方式，训练参数较少，在存储和计算效率方面都极有优势。Weng 等人^[44]将可分离残差模块应用于 SegNet，进行遥感图像水体分割，精度与 FCN 相比有显著提升。Zhao 等人^[45]提出了具有金字塔池化结构的 PSPNet 网络，将特征图分成不同层级和不同子区域，通过聚合不同区域的上下文信息，完成多层次的语义特征融合，充分挖掘全局信息。

DeepLab 系列网络^[46-49]是由 Google 研究团队开源的一系列深度学习网络。DeepLab

系列网络先后引入空洞卷积^[50]、条件随机场^[18]和空洞空间金字塔池化（Atrous Spatial Pyramid Pooling, ASPP）模块^[51]，充分利用特征图的多尺度信息，增强模型捕获精细细节的能力，使得语义分割网络的性能不断上升到新的水平。DeepLabv3+模型以DeepLabv3为基础，通过添加解码器模块而来。解码器能够增强模型处理边界信息的能力，较好保留分割目标的细节信息。

目前，在遥感影像语义分割实验中，除了使用对单一模型进行改进的方法之外，也有许多研究是基于多种算法结合的方式：结合传统方法和深度学习算法、结合多种深度学习算法等，这些方法能够不同程度地提高分类精度。何红术^[52]在改进的UNet网络中引入条件随机场，用于细化水体分割结果。么嘉棋^[53]结合SegNet和条件随机场，较好保留了目标的边缘轮廓。

国内外现阶段趋向于使用基于卷积神经网络的变体模型对高分辨率遥感影像进行提取和分类，其中DeepLabv3+是目前最优秀的通用分割网络之一，分割边缘光滑，而且分割精度在多个公开数据集上取得了领先优势^[48]。但DeepLabv3+模型在高分辨率遥感影像的语义分割方面仍存在一些不足，如因模型参数量大而造成分割效率低，分割过程中存在相似物体错分、细小目标漏分等现象，从而导致分割精度不高。

本文将以DeepLabv3+模型为基础的改进模型引入到高分辨率遥感影像耕地提取中，对提高耕地提取精度和效率的研究进一步探索。

1.2.3 耕地提取研究现状

随着遥感技术的进步，遥感影像的数据来源逐渐从低分辨率影像向高分辨率影像过渡，遥感影像耕地提取的方式也随着机器学习算法和深度学习网络的发展不断发生改变。下面从耕地提取的影像数据来源和不同阶段耕地信息的提取方式两个方面进行研究总结。

（1）影像数据来源

遥感影像数据源是地物信息提取的基础。不同的遥感卫星影像具有不同的适用条件与应用领域。遥感卫星的覆盖周期、空间分辨率会对耕地信息提取的结果产生一定影响。目前对遥感影像的分辨率层级没有具体划分，本文参照邓书斌在《ENVI遥感图像处理方法》^[54]中提到的分类方式，结合遥感影像在耕地提取的实际应用，对目前存在的遥感影像来源进行总结。

用于耕地提取中低分辨率遥感影像有 Landsat^[55, 56]、MODIS^[57, 58]、Sentinel^[59]等, 高分辨率遥感影像有 GF-1^[60-62]、GF-2^[63]、QuickBird^[64]、WorldView^[55, 65]、SPOT^[55, 66]等。此外, 许多研究结合了不同空间分辨率的遥感卫星影像, 以多源遥感影像作为耕地信息提取的基础^[67-70]。多源遥感影像可以充分利用各种遥感影像的优点, 能够更深层地对耕地信息进行分类识别。

(2) 耕地信息提取方式的发展

遥感技术应用之前, 研究人员对耕地信息的获取主要依靠实地勘察的方式, 周期长且更新慢。遥感技术的普遍应用, 使得利用遥感卫星影像获取耕地信息成为可能。初期, 人们对遥感影像的地物识别依赖于目视解译。目视解译能够通过影像中耕地呈现的纹理、形态以及地物之间的关联, 实现对耕地信息的准确提取^[71], 但处理大尺度范围的耕地影像时, 提取过程会耗费大量时间和人力成本。迫于农业信息统计对及时精确获取耕地信息的需要, 影像的自动识别分类方法极具探索价值。

机器学习方法为遥感影像耕地提取带来了新的思路。李昌俊等人^[72]使用支持向量机模型实现了对耕地类别较为准确快速的识别。Lewis-Gonzales^[65]使用监督分类方法在非洲国家的 Landsat 和 WorldView-2 影像上进行耕地提取和监测。李奕等人^[64]对比了监督分类与 ISODATA 算法在高分辨率 QuickBird 数据中提取耕地信息的表现, 证实监督分类效果更理想。

深度学习方法能够自主学习并不断挖掘图像的深层特征, 相比传统机器学习方法, 性能更为优秀。卷积神经网络是深度学习的基础模型, 后续许多应用于耕地信息提取的模型都基于此网络进行改进。陈玲玲等人^[73]汲取轻量级网络与 UNet 模型的优点, 有效改善了传统模型提取耕地结果较为零碎的情况。程锐^[74]通过集成 FCN、SegNet、PSPNet 和 UNet 网络提出 EDL 模型, 能有效解决耕地地块提取过程中出现的过拟合问题。Xu 等人^[75]将深度可分卷积应用于 UNet 网络, 在提取高分辨率遥感影像的耕地信息时表现突出。魏青迪等人^[76]以 DeepLab 网络为基础, 提出 ECLDeepLab 模型, 有效提高了高分二号影像上耕地与其他地物的提取精度。

综上所述, 基于深度学习对高分辨率遥感影像处理是目前较为流行的方法, 在遥感影像耕地信息提取上也取得了一些成果, 但仍存在一些问题, 如:

(1) 语义分割模型的参数量大小影响着训练效率。现有的模型改进大多数以提高分割精度为主，忽略了模型参数的调整，当处理场景复杂的高分辨率遥感影像时，训练时间长、效率较低，难以满足及时获取耕地信息的需求。

(2) 现有的深度学习模型在进行耕地提取时，存在轮廓提取不准确、小斑块漏分、误分和类别边界粘连的现象，分割精度仍有待进一步提高。

(3) 目前大部分耕地提取的方法大多只针对某一个地区或某一种数据，很少考虑影像不同波段组合对分割结果的影响。此外，迁移学习中预训练模型的不同也会在一定程度上影响分割效果。

因此，探索更适用于耕地提取的高精度、高效率、泛化能力强的语义分割模型，是本文的主要研究方向。

1.3 研究内容与研究方法

为了解决目前耕地语义分割存在信息提取不准确和分割效率低，且分割模型泛化性较差的问题，本文以 DeepLabv3+ 模型为基础，提出一种改进语义分割模型 MST-DeepLabv3+ 用于高分辨率遥感影像耕地信息的分类识别；并提出使用不同波段组合与不同迁移模型的方法，进一步探索能够实现耕地信息更高精度、更高效率的提取方法。本文技术路线图见图 1-1，主要研究内容与方法如下：

(1) DeepLabv3+ 模型的特征提取网络较为复杂、参数量大。MST-DeepLabv3+ 模型使用计算量小的轻量级网络作为主干网络，能够大幅减少模型参数，提高模型训练速度，同时可以避免由于参数较多而造成的过拟合风险。

(2) 轻量级网络的添加可以有效提高模型分割效率，但也会造成一部分的精度损失。通道注意力机制能够对特征通道权重重新分配，使网络专注于具有较高重要性的特征通道，进而实现遥感影像语义分割精度的提升。迁移学习通过加载预训练模型，能够增强模型获取特征的能力，有效提高模型分割性能。MST-DeepLabv3+ 模型引入注意力机制与迁移学习，以获得较高的分割精度。并通过多模型在 ISPRS Vaihingen 与 GID 两个遥感影像公开数据集上的实验结果对比，验证 MST-DeepLabv3+ 模型的有效性。

(3) 以高分一号遥感卫星影像为数据源，制作太康耕地数据集。新的耕地提取模型参考 GID 与 ISPRS Vaihingen 数据集各自的特征，将含有耕地类别的 GID 数据集的训练模型作为迁移模型，构建 MST-DeepLabv3+(GID) 模型；对包含近红外、红色和绿色波段的太康耕地数据集处理，建立 MST-DeepLabv3+(NirRG) 模型。与经典模型对比，分析

MST-DeepLabv3+(GID)、MST-DeepLabv3+(NirRG)与 MST-DeepLabv3+模型对耕地信息的提取效果。

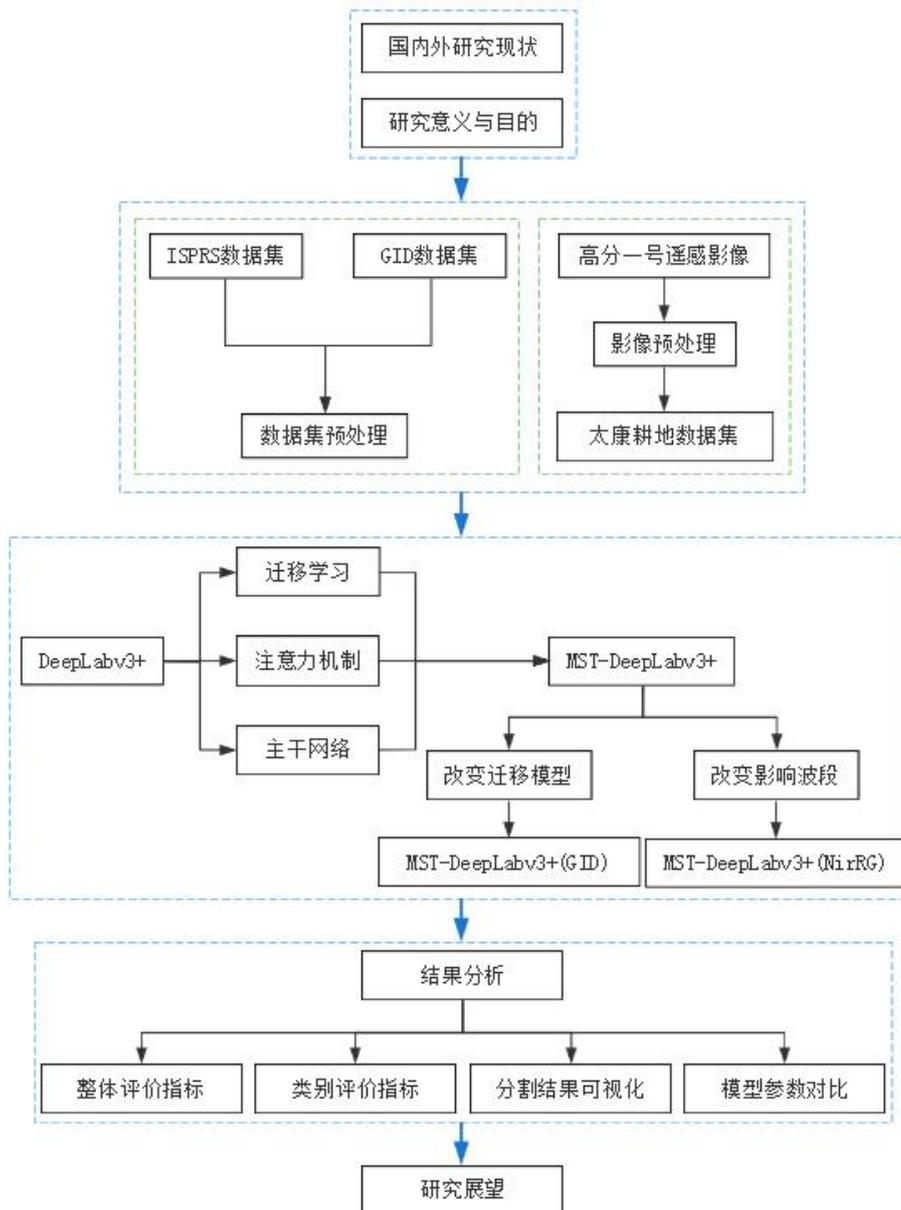


图 1-1 技术路线图

1.4 本文结构安排

本文主要分为六个章节，各章节的主要内容安排如下：

第一章主要介绍了遥感影像耕地提取的背景和意义，概括了国内外遥感影像提取的发展与现状，强调了深度学习对遥感影像提取的重要意义。结合耕地提取研究目前存在的不足，对本文研究内容和研究思路进行概括。

第二章详细介绍了目前较为常见的部分经典语义分割模型的基本结构和原理，包括

FCN、SegNet、UNet、PSPNet 与 DeepLabv3+模型。并对本文实验所用语义分割评价指标进行概括。

第三章依次对本文研究区地理位置、耕地信息与实验数据来源进行介绍，说明研究区的选取原因。

第四章首先对 MST-DeepLabv3+模型优化方法的相关理论和基本原理进行具体讲解，包括轻量级网络、注意力机制和迁移学习。描述实验环境配置之后对不同模型在 ISPRS Vaihingen 和 GID 两个公共数据集的分类结果进行定量和定性两种方式的对比分析，结合消融实验与模型参数的对比，从多方面验证 MST-DeepLabv3+模型的性能。

第五章包括太康耕地数据集的建立和耕地提取实验结果的对比分析。在第四章的模型基础上，增加 MST-DeepLabv3+(GID)和 MST-DeepLabv3+(NirRG)两种模型，探索更优秀耕地信息提取方法。

第六章对本文研究内容进行整体概括，总结本文创新点并阐述对下一步研究的展望。

2 遥感影像语义分割模型

在遥感影像领域，语义分割模型仍有较大的优化空间。本章对部分经典语义分割模型 FCN、SegNet、PSPNet、UNet 与 DeepLabv3+ 进行汇总，从模型结构和原理入手，进行整体分析，总结各模型的特点，作为本文下一步实验的参考。

2.1 经典模型

2.1.1 FCN

FCN 网络基于 CNN 网络，由 Long 等人 2015 年提出^[38]。CNN 网络主要由卷积层、池化层和全连接层构成。图 2-1 为 CNN 网络架构图，其核心是卷积层，池化层在卷积层之后，也称为下采样层，通过对特征图降维来降低模型复杂度。网络末端是全连接层，整合卷积层与池化层提取到的特征信息^[77]。CNN 网络处理细节丰富的高分辨率影像时，会出现丢失信息严重的情况，无法得到精准的分类结果。

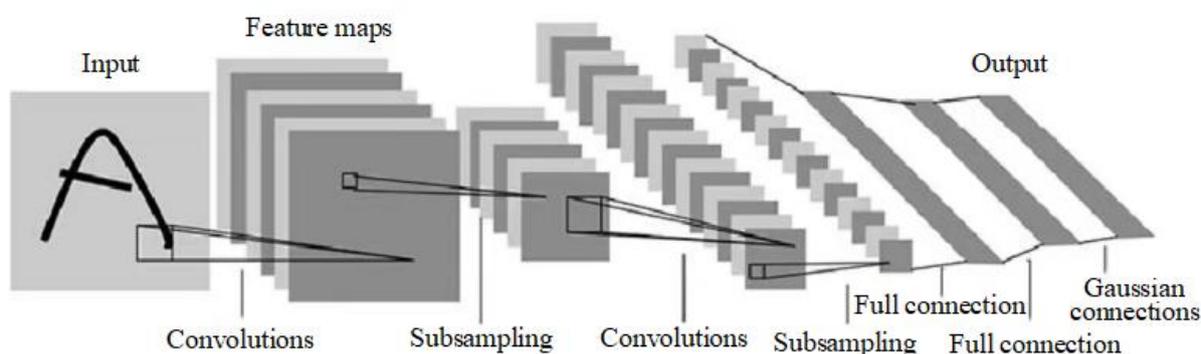
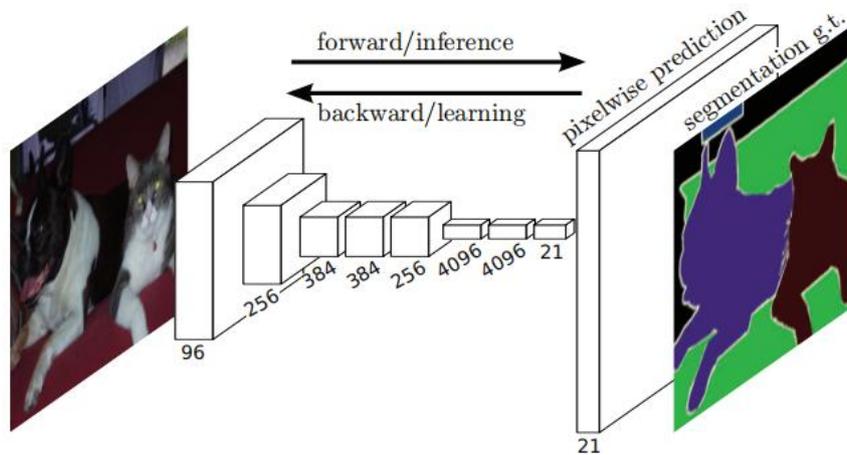


图 2-1 CNN 网络架构^[9]

FCN 网络将 CNN 网络中的全连接层替换为卷积层，使用全卷积架构可以处理任意大小的输入图像，并进行端到端、像素到像素的训练。引入可以融合多尺度特征的跳跃结构，结合深层信息和浅层信息，有效地提高了信息的利用率。其中浅层信息为位置信息，深层信息为语义信息，能够获知每个像素属于什么类别。FCN 的网络架构见图 2-2。

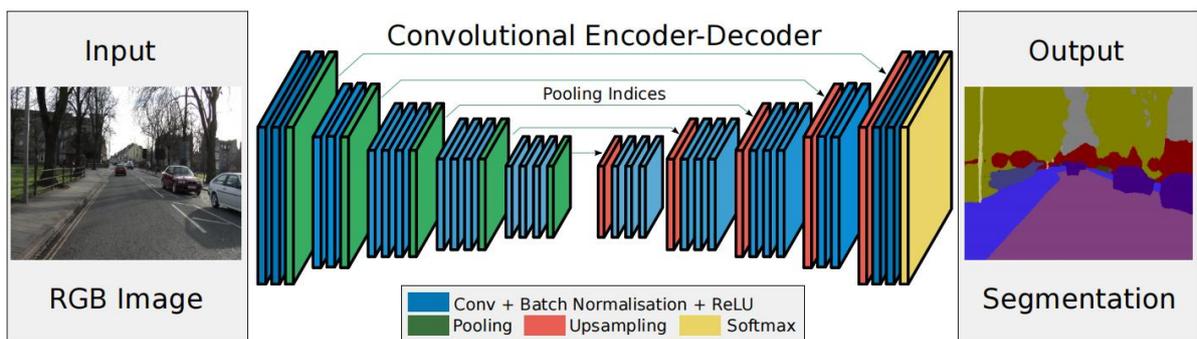
作为第一个具有编码器—解码器结构的语义分割网络，FCN 网络通过池化层降低特征空间的维数，减小图像特征大小，同时避免网络出现过拟合^[78]。然后，通过反卷积操作，将尺寸缩小的特征图恢复为原始图像大小，得到每个像素所属的类别。然而，FCN 的上采样过程无法充分利用空间上下文信息，在分割中会出现边界不完整，结果粗糙的问题。

图 2-2 FCN 网络架构^[38]

2.1.2 SegNet

Badrinarayanan 等人 2017 年提出的 SegNet 模型由编码器、解码器和像素级分类器组成，最初用于道路场景和室内场景的语义分割^[43]。SegNet 的网络架构如图 2-3 所示。

SegNet 模型构建的整体思路与 FCN 类似，但又针对 FCN 网络的缺陷做了一些改进。首先，SegNet 网络将编码器部分的全连接层直接删除，比 FCN 网络更小且更容易训练。其次，不同于 FCN 网络的池化层将大量内存用于保留整个特征图的方法，SegNet 网络通过保留池化索引的方式，以最少的内存空间和计算时间来最大限度地保留边界信息。最后，在解码器部分，与 FCN 反卷积上采样相比，SegNet 采用反池化的方式进行上采样，通过调用最大池化操作的索引，恢复图像纹理信息和位置信息。

图 2-3 SegNet 网络架构^[43]

2.1.3 UNet

UNet 网络结构形似“U”，是一种编码器——解码器架构，于 2015 年由 Ronneberger 等人提出^[41]。最初 UNet 网络是为数据量较少的生物学图像分割任务设计的，后来被广泛应用于计算机视觉任务。UNet 的网络结构如图 2-4 所示。

(1) 编码器部分(左侧)遵循典型的卷积网络架构, 与 VGG^[79]相似, 为卷积和最大池化的堆叠, 每个下采样模块均由两个卷积核大小为 3×3 的卷积层, ReLU 函数与一个步长为 2, 滤波器为 2×2 的最大池化层的组成, 进行特征提取和特征压缩。池化操作使得特征图像大小减小, 通道数翻倍。

(2) 解码器部分(右侧)对特征图进行四次上采样操作以恢复图像信息。每个上采样模块由两个卷积核大小为 3×3 的卷积层和一个大小为 2×2 反卷积组成。反卷积操作能够增加特征图像大小, 减少通道数。在解码器最后一层, 利用一个 1×1 卷积调整最终特征层的通道数为所需分类类别数。

(3) UNet 通过拼接编码器和解码器的相应层, 将浅层的位置信息与深层的语义信息相融合, 恢复图像细节, 使模型能够捕获局部和全局信息。

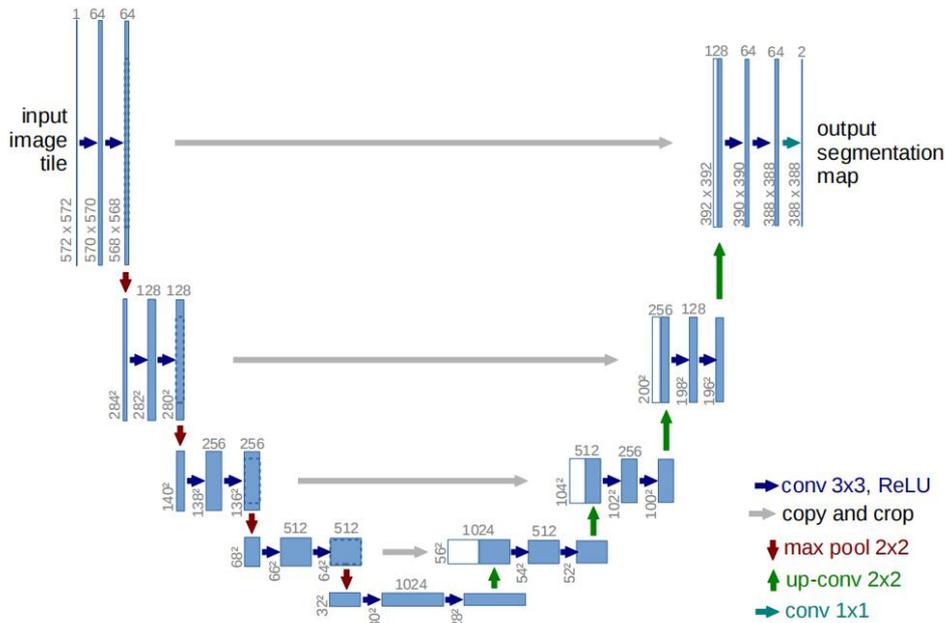


图 2-4 UNet 模型结构图^[41]

UNet 网络相比于其他分割模型, 只需要较少带有注释的图像就可以实现较高的分类准确率^[40], 在多种类型的遥感影像和遥感数据集上取得了卓越的分割效果^[80-82]。

2.1.4 PSPNet

PSPNet 网络由 Zhao 等人在 2017 年提出, 目的为优化 FCN 网络在语义分割任务中存在的对各类别之间关联性缺少关注, 对细小类别识别能力不足等问题, 主要包括主干特征提取网络和金字塔池化模块(Pyramid Pooling Module, PPM)^[45]。网络结构如图 2-5 所示。

PSPNet 以 ResNet 作为主干特征提取网络, 通过卷积操作进行初步特征提取得到特

征图，并使用辅助分支优化学习过程。金字塔池化模块赋予了 PSPNet 捕获不同尺度局部特征的能力，用于加强特征提取。将输入的特征层划分为 6×6 , 3×3 , 2×2 , 1×1 的子区域，通过分级和平均池化来提取不同尺度和分辨率的特征，使用 1×1 的卷积将通道数降为 $1/4$ ，之后通过上采样操作恢复特征图的空间分辨率^[83]。将 4 种经过池化提取的特征图与卷积操作得到的初步特征图通过维度拼接的方式进行整合，有助于 PSPNet 捕获多尺度的上下文信息，从而提高获取全局信息的能力。

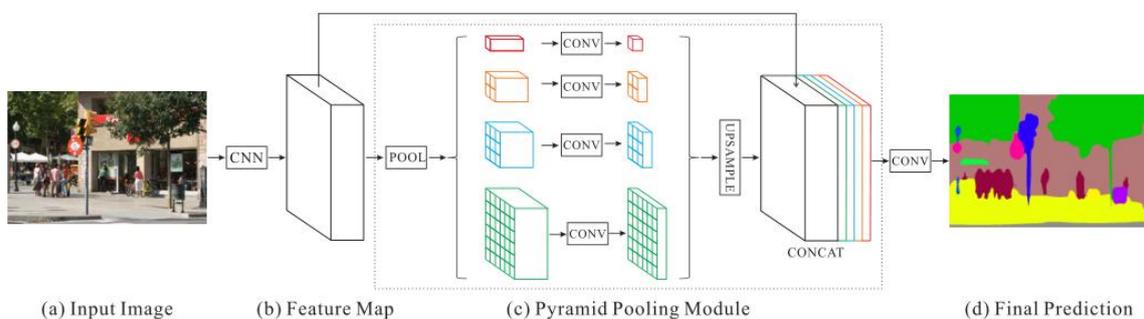


图 2-5 PSPNet 模型结构^[45]

2.1.5 DeepLabv3+

DeepLabv3+网络由 Chen 等人在 2018 年推出，采用编码器——解码器的结构^[48]。DeepLabv3+模型将 DeepLabv3 整体作为网络的编码器，使用一系列卷积层和池化层对输入图像进行下采样并捕获不同尺度的特征，对目标边缘信息的提取效果进行优化。采用解码器融合来自主干网络的低级特征和来自 ASPP 模块的高级特征，进行特征信息恢复，提升分割效果，较好地保留目标的边缘细节信息。DeepLabv3+模型以 Xception 做为主干网络，并在 ASPP 和解码器模块中应用深度可分离卷积，从而产生分割效果更好的编码器——解码器网络。DeepLabv3+在许多数据集上展示了优秀的性能^[84-86]，能够捕获多尺度特征并保留细节信息，非常适合分析高分辨率遥感影像。DeepLabv3+模型结构如图 2-6 所示。

(1) Encoder 部分：包含主干特征提取网络 (DCNN) 与空洞空间金字塔池化模块。特征提取网络里使用串行空洞卷积，处理后的结果分别传入 Decoder 与 ASPP 模块进行特征提取。ASPP 模块使用不同空洞率的空洞卷积来捕捉多尺度特征信息，提高网络感受野。深层特征经过 ASPP 模块提取后进行堆叠，经过 1×1 卷积调整通道数，进入 Decoder 部分。

(2) Decoder 部分：浅层特征通过 1×1 卷积降维后，融合由 Encoder 部分输出并通

过双线性插值方法进行 4 倍上采样操作的深层特征，使用 3×3 卷积进行特征提取。最后通过 4 倍上采样操作，获得与输入图像尺寸相同的预测结果。

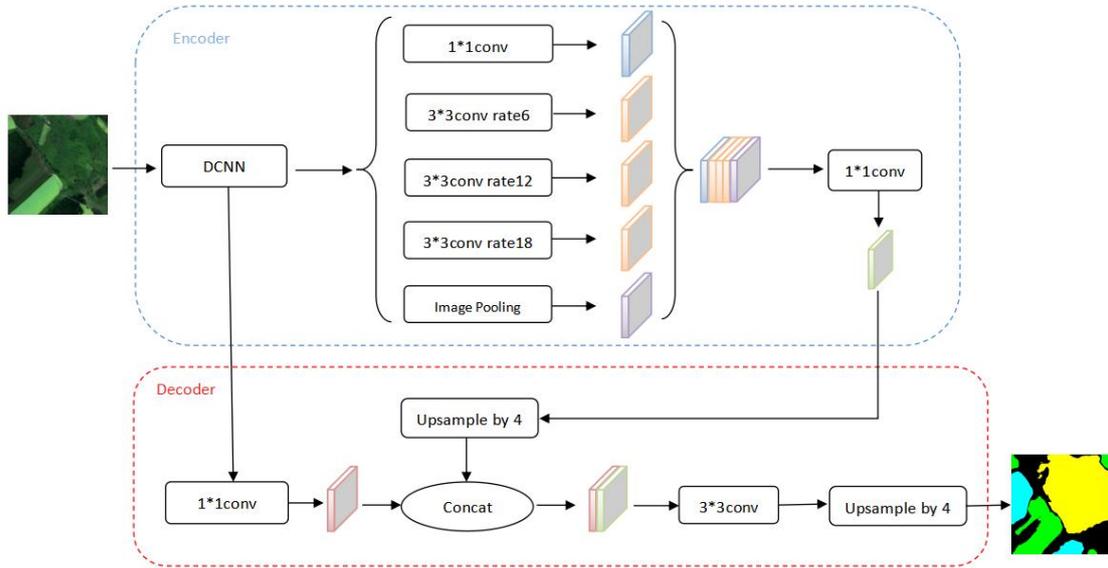


图 2-6 DeepLabv3+模型结构

2.2 评价指标

为了有效评估模型的分割性能，探讨不同模型在遥感影像语义分割中的优缺点，本文使用平均交并比(Mean Intersection over Union, MIoU)、总体精度 (Overall Accuracy, OA)、Precision、Recall 与 F1 分数 (F1-Score) 作为评价指标，综合评价模型分割性能。

MIoU 即平均交并比，是目前语义分割领域使用频率最高的标准度量指标^[87]。首先根据各类别的真实值和预测值两个集合，分别计算交集与并集的比值，也就是交并比。之后对全部类的交并比求平均值，即可得到 MIoU。见式 (2-1)：

$$MIoU = \frac{1}{K+1} \sum_{i=0}^k \frac{TP}{FN + FP + TP} \quad (2-1)$$

OA 是指正确分类的像素数与总像素数的比值，可以表示模型总体精度。见式(2-2)：

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

Precision 为精确率，又称查准率，是指预测为正例的像素中,预测正确的像素占比，见式 (2-3)：

$$Precision = \frac{TP}{FP + TP} \quad (2-3)$$

Recall 指召回率，又称查全率，是模型正确识别为正例的像素占总的正例像素的比

值。计算见式 (2-4)。

$$Recall = \frac{TP}{TP + FN} \quad (2-4)$$

F1-Score 是一个更加全面的综合性模型精度评价指标, 为 Precision 和 Recall 的调和平均值。能够解决当各类别像素数量偏差较大时, OA 指标不能对具体分类结果进行准确评价的问题, 见式 (2-5) :

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

上述公式中 $k+1$ 表示数据类别数, 包含背景类; TP (True Positive) 表示真正例, 即模型预测为正例, 实际是正例; FP (False Positive) 表示假正例, 即模型预测为正例, 实际是反例; FN (False Negative) 表示假反例, 即模型预测为反例, 实际是正例; TN (True Negative) 表示真反例, 即模型预测为反例, 实际是反例。

2.3 本章小结

本章先详细介绍了几种目前应用广泛且分割效果较好的语义分割模型的结构及基本原理, 包括 FCN、SegNet、UNet、PSPNet 与 DeepLabv3+模型, 为后续对比实验中的模型选择提供依据。第二节对本文选用的语义分割评价指标简要说明, 包括 MIoU、OA、Precision、Recall 与 F1-Score。通过多个指标, 能够更加全面的评估实验结果。

3 研究区与实验数据

3.1 研究区概况

3.1.1 研究区地理位置与自然特征

周口市太康县位于河南省东部、周口市北部，地跨东经 $1140^{\circ} 32'$ ~ $1150^{\circ} 08'$ 、北纬 $330^{\circ} 54'$ ~ $340^{\circ} 17'$ ，海拔 $45.5\sim 59.5\text{m}$ ，呈现西北高、东南低的地势特征，高差约为 12.9m 。太康县东与柘城县、鹿邑县接壤，南与淮阳区、西华县相连，西靠扶沟县，北临通许县、杞县、睢县，县域东西长约为 55.25km ，南北宽约为 44.25km ，下辖 15 个镇、8 个乡，总面积约为 1759km^2 ^[88, 89]。太康县属于暖温带季风气候，全县平均气温约为 15.1°C ，四季分明，降雨量随季节变化，全县平均降雨量约为 729.9mm ^[90]。太康县境内地势平坦，土地肥力良好，适宜农业生产，是传统农业大县，农作物主要包括小麦、大豆、红薯等 60 余种，拥有“中国粮食生产先进县”、“中国优质棉基地县”等称号^[89]。图 3-1 为太康县位置图。

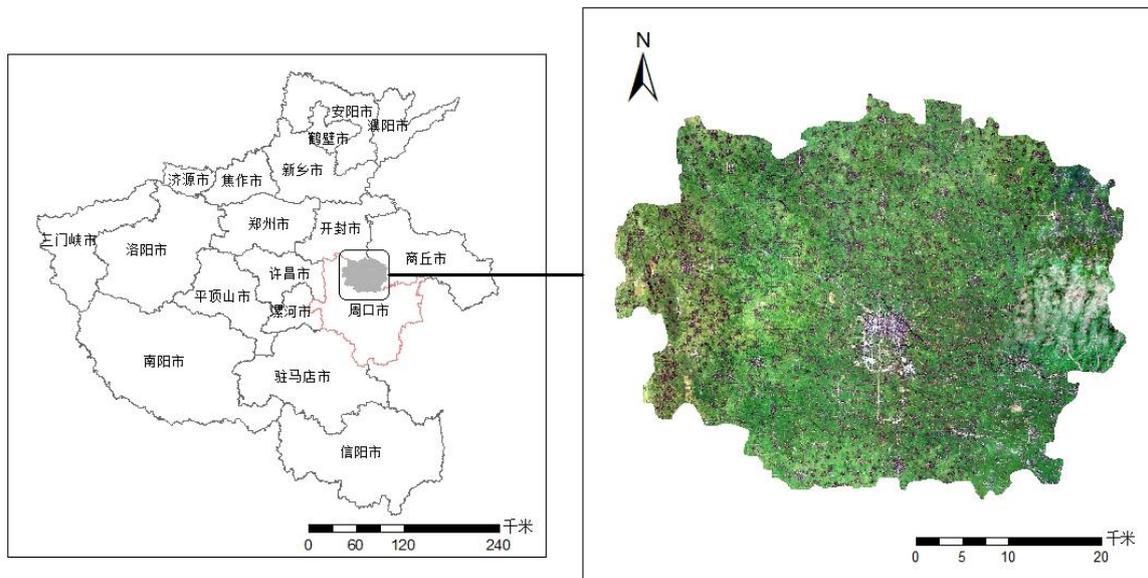


图 3-1 太康县位置图

3.1.2 研究区耕地情况

据太康县第三次全国国土调查主要数据公报数据显示，以 2019 年 12 月 31 日为标准时点，太康县耕地约为 190 万亩^[91]。太康县耕地面积约占县域总面积的 72%，约占周口市总耕地面积的 16%，是重要的粮食生产地区。

太康县的土壤类型以河流沉积物形成的潮土为主，一般呈中性至弱碱性，表现为地势平坦、土层深厚^[92]。太康县粮食耕作结构较为单一，全部为水浇地，且空间分布较为集中，利于耕作^[93]。根据 2022 年周口统计年鉴的统计信息显示，太康县粮食作物总播种面积为 201.31 千公顷，夏收粮食占 111.69 千公顷，秋收作物占 89.62 千公顷。谷物种植面积共 188.17 千公顷，其中小麦种植面积约占 111.69 千公顷，约为谷物种植总面积的 59%。玉米种植面积为 76.47 千公顷，仅次于小麦。小麦与玉米是太康县种植的主要粮食作物。据报道，2023 年小麦种植面积达到了 172 万亩。

本文选取太康县作为研究区，对平原地区耕地信息的提取研究具有较好的代表性。

3.2 实验数据

3.2.1 ISPRS 数据集

ISPRS 数据集由国际摄影测量与遥感学会创建^[94]，包含 Vaihingen 和 Postdam 两个子数据集，这两个子数据集区域都涵盖了大部分城市场景。Vaihingen 是一个较小的村庄，有许多独立的建筑，包含 33 幅不同大小的遥感图像，每幅图像都是从一个更大的顶层正射影像图片中提取而来，空间分辨率为 9cm。遥感图像由近红外、红色和绿色 3 个波段组成，示例图见图 3-2。Postdam 是一个典型的历史城市，有着大的建筑群、狭窄的街道和密集的聚落结构，共 38 张尺寸为 6000 像素×6000 像素无人机影像，分辨率为 5cm，示例图见图 3-3。这两个数据集均包含 6 个常见的土地覆盖类别：不透水面，建筑物，低矮植被，树木，汽车，背景。本文选择 Vaihingen 数据集进行后续实验。

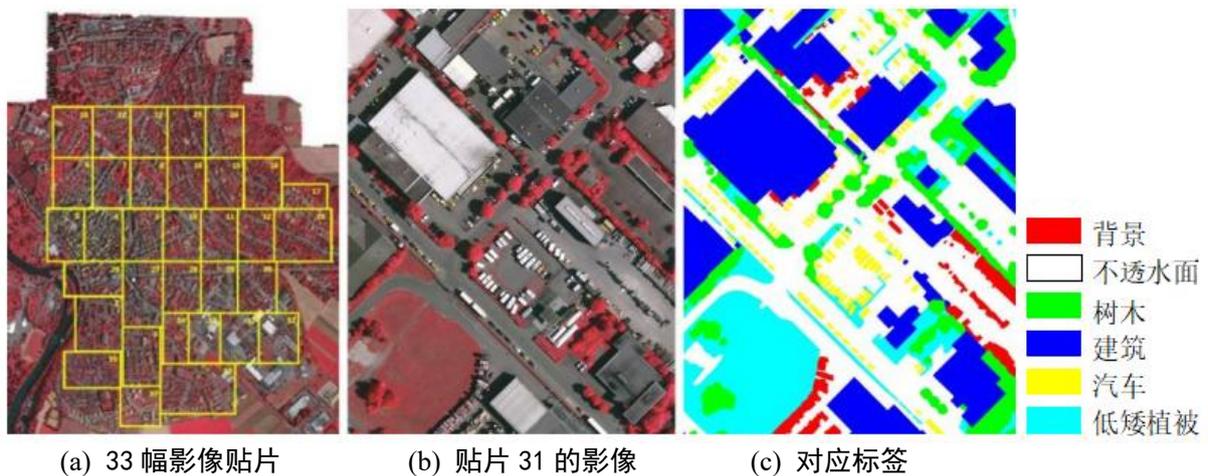


图 3-2 Vaihingen 数据集示例图

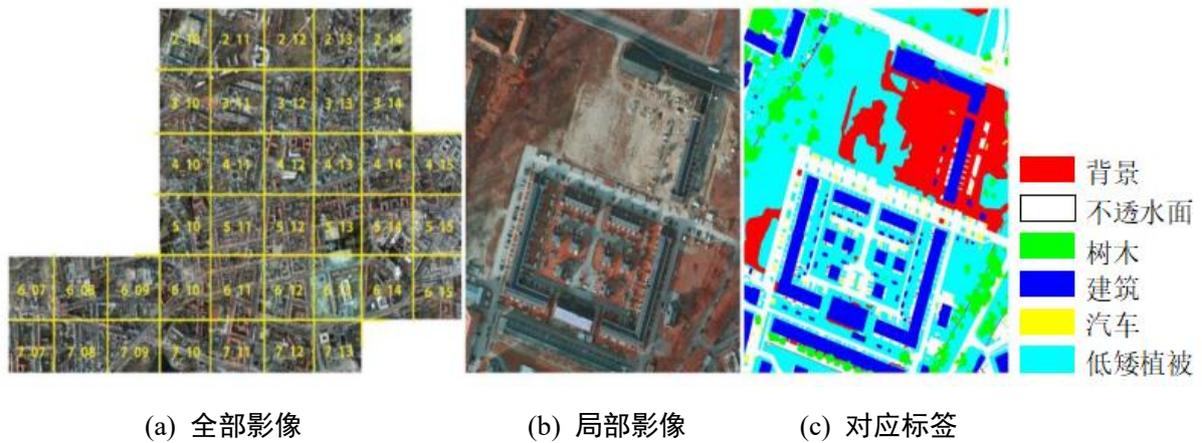


图 3-3 Postdam 数据集示例图

3.2.2 GID 数据集

GID 数据集是根据中国高分二号卫星数据构建的大规模高分辨率遥感图像土地覆盖数据集^[95]，其中高分二号卫星包含 1m 分辨率的全色图像和 4m 分辨率的多光谱图像。GID 数据集包括大规模分类集（图 3-4 为示例图）和精细土地覆盖集（图 3-5 为示例图）两个部分，图像清晰无云，十分贴近真实的地物分布特性。大规模分类集包含来自 60 多个不同城市的 150 张尺寸为 6800 像素×7200 像素的高品质遥感影像，整体图像的地理区域范围超过 50000 平方公里。共有水体、建筑、耕地、草地和森林 5 个土地覆盖类别。精细土地覆盖集由高分二号卫星影像切割处理得到，包含稻田、灌溉地、旱地、园地、乔木林、灌木林、自然草地、人工草地、工业用地、城市住宅、农村住宅、交通用地、河流、湖泊、池塘 15 个类别，共计有 30000 个图像块。尺寸类别有 56 像素×56 像素,112 像素×112 像素,224 像素×224 像素。

GID 数据集的大规模分类集包含耕地类别，选用该数据集进行实验能够为之后的高分辨率遥感影像耕地提取研究奠定基础。

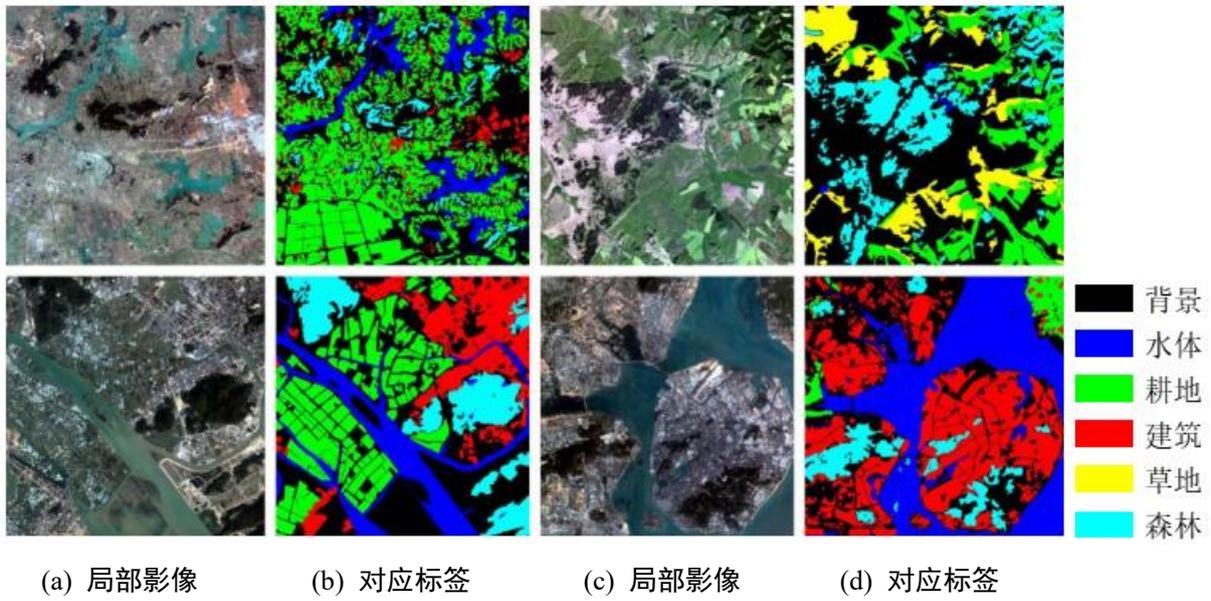


图 3-4 大规模分类集示例图

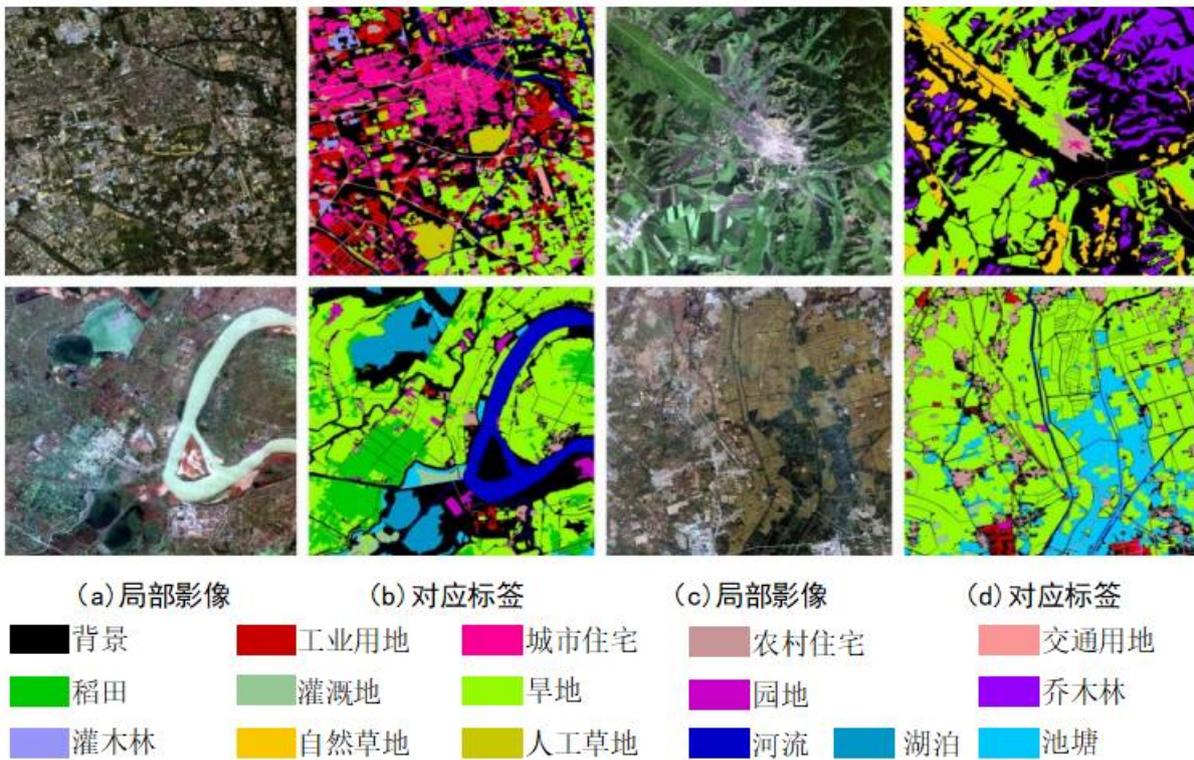


图 3-5 精细土地覆盖集示例图

3.2.3 高分一号遥感影像

本文研究区的高分一号遥感卫星影像来源于高分湖北中心^[96]。高分一号卫星是中国自主研发的地球观测遥感卫星，于 2013 年 4 月在中国酒泉卫星发射中心成功发射，是中国高分辨率对地观测系统重大专项天基系统中的第一颗卫星，为土地监测^[97]、灾害评估^[98, 99]、生态安全^[100]等提供服务。高分一号遥感卫星影像具有高空间分辨率、高稳定度

姿态、高时间分辨率、寿命大于五年，且 5 年末可靠度不低于 0.6 等优点，提高了我国高分辨率影像数据的自给率，为利用国产卫星数据进行高分辨率遥感影像的土地利用分类研究提供了极大的帮助^[101]。

高分一号卫星平均轨道高度为 644.5km，寿命为 5~8 年，配备有两台 2m 分辨率全色、8m 分辨率多光谱的 PMS 相机和四台 16m 分辨率 WFV 相机，覆盖周期分别为 41 天和 4 天。PMS 相机地面幅宽大于 60km，有五个波段，分别为全色（波长 0.45-0.90 μm ）、蓝（Band1, 0.45-0.52 μm ）、绿色（Band2, 0.52-0.59 μm ）、红色（Band3, 0.63-0.69 μm ）和近红外（Band4, 0.77-0.89 μm ），WFV 相机地面幅宽大于 800km，有四个波段，分别是近红外、红、绿和蓝波段^[102]。

本文选用分辨率较高的 PMS 影像进行实验，按照耕地纹理清晰可见，且云层覆盖较少的标准筛选影像。根据研究区耕地农作物的生长及成熟周期，最终选取 2017 年 2 月份与 5 月份的两幅影像作为源数据，场景 ID 分别为 3350241 与 3661472。全色与多光谱影像的传感器分别为 PAN1 和 MSS1，统一投影坐标系为 WGS_1984_UTM_Zone_50N。

3.3 本章小结

本章介绍了本文所使用的研究数据，共有三种数据，一是波段信息为近红外、红色和绿色的 ISPRS Vaihingen 数据集；二是具有耕地类别的 GID 数据集；三是高分一号遥感影像。此外，还介绍了所选研究区周口市太康县的基本地理信息、气候条件概况和主要农作物的种植状况。

4 MST-DeepLabv3+: 一种改进的遥感影像语义分割方法

目前在语义分割领域，DeepLabv3+网络是最优秀的通用分割网络之一。但由于特征提取网络结构复杂，会造成 DeepLabv3+模型训练时内存占用较大，模型拟合速度慢，分割效率低的问题；其次，DeepLabv3+模型在高分辨率的语义分割方面较难准确捕捉地物细节，导致分割精度不高。本章对 DeepLabv3+模型进行改进，提出 MST-DeepLabv3+模型，以达到提高分割精度，降低参数数量的目的。其中，MST 为 MobileNetV2、SENet 与 Transfer learning 的首字母组合得到。在 ISPRS Vaihingen 数据集和 GID 数据集上测试其性能，并与几种经典语义分割算法的精度结果对比，验证改进模型的有效性。

4.1 改进机制

与 DeepLabv3+模型相比，本文提出的 MST-DeepLabv3+模型进行了三点改进：（1）DeepLabv3+模型的编码器部分所用的特征提取模块 Xception 网络^[103]结构复杂，参数量及内存占用较大，会造成模型拟合速度慢，分割效率低的问题。因此将其更换为轻量级网络 MobileNetV2^[104]，以减少语义分割模型中的参数量，提高模型训练效率；（2）引入通道注意力机制 SENet^[105]，通过对特征通道权重的重新分配，改善目标被漏分与被误判的情况，从而提高分割精度；（3）运用迁移学习^[106]，将从 ImageNet 数据集^[107]上获得的模型用作预训练模型，以优化网络训练速度，提高模型捕获特征的能力，在降低模型复杂度的同时提高分割性能。图 4-1 为 MST-DeepLabv3+模型结构。

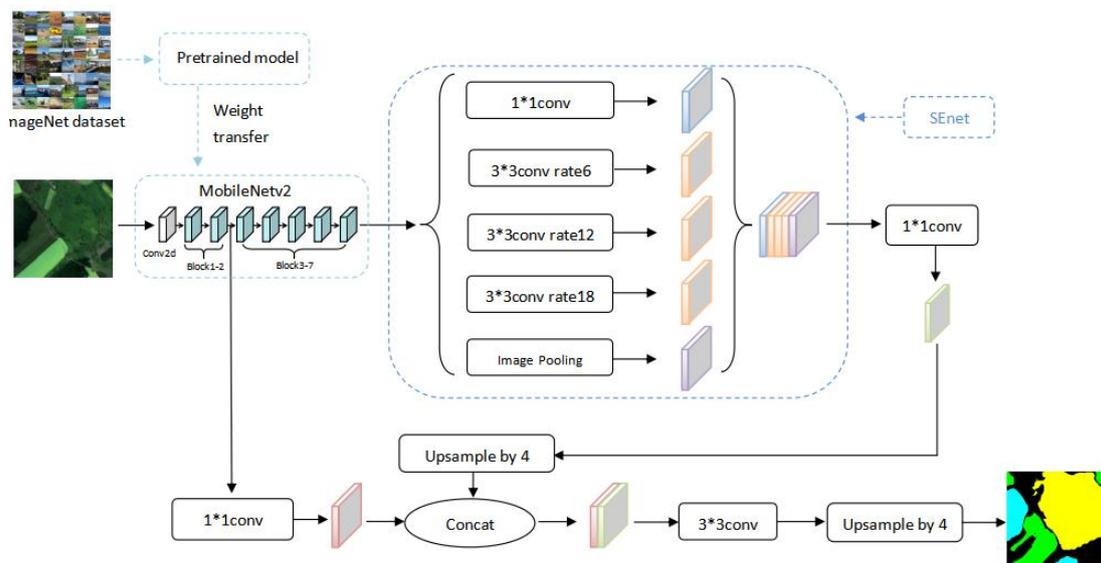


图 4-1 MST-DeepLabv3+模型结构

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/526104134232011002>