



自然资源标准项目建议书

中文名称	卫星遥感影像解译样本生产技术规程				
英文名称	Code of practice for sample production of satellite remote sensing image interpretation				
标准层级	<input type="checkbox"/> 国家标准 <input checked="" type="checkbox"/> 行业标准	标准性质		<input checked="" type="checkbox"/> 推荐性标准 <input type="checkbox"/> 强制性标准	
标准类型	<input type="checkbox"/> 环保 <input checked="" type="checkbox"/> 方法 <input type="checkbox"/> 产品 <input type="checkbox"/> 安全 <input type="checkbox"/> 管理 <input type="checkbox"/> 其他 <input type="checkbox"/> 基础 <input type="checkbox"/> 卫生				
制定/修订	<input checked="" type="checkbox"/> 制定 <input type="checkbox"/> 修订		被修订标准号		
采用国际标准	<input checked="" type="checkbox"/> 无 <input type="checkbox"/> ISO 确认的国际标准 <input type="checkbox"/> ISO <input type="checkbox"/> ISO/IEC <input type="checkbox"/> IEC <input type="checkbox"/> ITU		采用程度		<input type="checkbox"/> 修改 <input type="checkbox"/> 等同 <input type="checkbox"/> 非等效
采标号			采标名称		
ICS	07.040		CCS	A77	
技术归口单位	全国地理信息标准化技术委员会 卫星应用分技术委员会(TC230SC3)				
牵头起草单位	自然资源部国土卫星遥感应用中心				
参加起草单位	武汉大学、中国科学院空天信息创新研究院、湖南省第二测绘院、重庆市地理信息和遥感应用中心、广东省国土资源技术中心、辽宁省自然资源事务服务中心、内蒙古自治区测绘地理信息中心、湖北省航测遥感院				
联系人	姓名	王光辉	职务/职称	副处长	手机 18610588358
	电话	010-68412617	邮箱	wanggh@lasac.cn	
	所在单位	自然资源部国土卫星遥感应用中心			
项目周期	24个月	是否采用快速程序	<input type="checkbox"/> 是 <input checked="" type="checkbox"/> 否	快速程序代码	<input type="checkbox"/> B2 <input type="checkbox"/> C3 <input type="checkbox"/> B3 <input type="checkbox"/> B4 <input type="checkbox"/> B1
目的、意义	<p>以深度学习为代表的人工智能技术因其海量数据的学习和应用能力，已成为卫星遥感影像解译发展的新方向。类型丰富、种类齐全的遥感影像解译样本数据可有效提升深度学习算法模型的准确性与自适应性。但现有样本数据类别少、且都是基于特定目标设定，无法有效支撑各类信息的快速、准确提取；样本尺寸固定，无法满足不同深度学习算法模型对样本尺寸的要求；样本数据量小，无法支撑深度学习算法模型训练、优化及迭代，制约着人工智能技术在遥感影像解译领域的广泛应用。因此，急需开展卫星遥感影像解译样本采集技术规范研究，约束和规范解译样本采集方式、规格、策略等内容，形成一体化的作业标准，指导解译样本数据的采集与制作，支撑遥感影像智能算法模型的构建、迭代与优化。此项工作的开展具有重要的现实意义及广阔的应用前景。</p>				
范围和主要技术内容	<p>范围：本文件规定了卫星遥感影像解译样本数据生产的总体要求、样本数据采集、样本集生产等内容。适用于卫星遥感影像解译样本生产工作。</p> <p>主要技术内容： 1) 总体要求：包括空间参考、时间参考、精度、完整性等；</p>				

	<p>2) 样本数据本采集: 包括: 数据源准备、样本要素采集、样本元数据录入及样本质量检查;</p> <p>3) 样本集生产: 包括样本集数据生成、样本集元数据录入及样本集质量检查。</p>		
国内外情况简要说明	<p>随着深度学习等人工智能技术方法在遥感解译领域的快速发展, 大规模、高质量、多形态的遥感影像解译样本数据已成为卫星遥感影像解译的主要瓶颈之一。武汉大学、中科院、DigitalGlobe 等众多国内外高校、科研院所、企事业单位相继建立各类样本数据集, 但国内外尚未形成统一的解译样本采集技术规范或指导性文件。自然资源部直属单位及相关省市单位, 围绕业务需求, 已逐步开展了相关基于人工智能的遥感信息提取技术研究, 初步探索开展了解译样本采集等相关工作。综合相比, 本次计划编制的《卫星遥感影像解译样本生产技术规程》聚焦自然资源应用领域, 结合自然资源监测监管、国土空间规划、自然资源督察执法等业务需求, 充分发挥光学、SAR、高光谱等多源遥感影像数据优势, 形成一体化的解译样本数据采集与制作作业标准, 支撑遥感解译技术发展, 服务自然资源调查、监测、评价工作。</p>		
有关法律法规和强制性标准的关系	<p>本技术规范遵守现行法律、法规要求, 无冲突内容。本技术规范与有关的国家标准保持一致。</p>		
经费预算说明	<p>经费采用自筹方式解决。预算共 6 万元, 其中专家咨询费 4 万元, 劳务费 1 万元, 印刷费 1 万元。</p>		
时间进度安排	<p>2023.7-2024.4 完成标准征求意见稿和标准编制说明。</p> <p>2024.4-2025.1 完成标准送审稿、编制说明和征求意见汇总处理表。</p> <p>2025.1-2025.5 完成标准报批稿、编制说明、意见汇总处理表等报批资料。</p>		
标准涉及的产品清单	<p>无</p>		
是否有国家级科研项目支撑	<input type="checkbox"/> 是 <input checked="" type="checkbox"/> 否	科研项目编号及名称	
是否涉及专利	<input type="checkbox"/> 是 <input checked="" type="checkbox"/> 否	专利号及名称	
是否由行标/地标/团标转化	<input type="checkbox"/> 是 <input checked="" type="checkbox"/> 否	转化标准号及名称	
是否属于军民通用的标准项目	<input type="checkbox"/> 是 <input checked="" type="checkbox"/> 否	协调情况	
牵头起草单位意见(加盖公章)		推荐单位意见(加盖公章)	
备注			

CH

中华人民共和国测绘行业标准

XX/T XXXXX—XXXX
代替 XX/T

卫星遥感影像解译样本生产技术规范

Code of practice for sample production of satellite remote sensing image
interpretation

(点击此处添加与国际标准一致性程度的标识)

(草案)

XXXX - XX - XX 发布

XXXX - XX - XX 实施

中华人民共和国自然资源部 发布

目 次

前 言	I
引 言	II
1 范围	3
2 规范性引用文件	3
3 术语和定义	3
4 总体要求	3
5 样本采集数据源	4
5.1 遥感影像数据	4
5.1.1 光学影像	4
5.1.2 高光谱影像	4
5.1.3 SAR 影像	4
5.2 已有成果数据	5
5.3 辅助数据	5
6 样本生产工作流程	5
7 样本数据采集	6
7.1 样本采集要求	6
7.1.1 拓扑要求	6
7.1.2 要素边界要求	6
7.1.3 地物连通性要求	6
7.2 样本采集	6
7.2.1 准备数据源	6
7.2.2 样本要素采集	7
7.2.3 样本元数据录入	7
7.3 样本质量检查	7
7.3.1 质量检查内容	7
7.3.2 质量检查流程	8
7.4 样本数据组织	8
8 样本集生产	9
8.1 样本集制作要求	9
8.2 样本集制作	9
8.2.1 样本集数据生成	9
8.2.2 元数据录入	10
8.2.3 样本集质量检查	10
8.3 样本集成果组织	10
附 录 A (资料性附录) 卫星遥感影像解译样本元数据填写规范	12
附 录 B (资料性附录) 卫星遥感影像解译样本集小片元数据内容及规范	14
参 考 文 献	15

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由中华人民共和国自然资源部提出。

本文件由全国地理信息标准化技术委员会卫星应用分技术委员会（SAC/TC230/SC3）归口。

本文件起草单位：自然资源部国土卫星遥感应用中心、武汉大学、中国科学院空天信息创新研究院、湖南省第二测绘院、重庆市地理信息和遥感应用中心、广东省国土资源技术中心、辽宁省自然资源事务服务中心、内蒙古自治区测绘地理信息中心、湖北省航测遥感院。

本文件主要起草人：

引 言

以深度学习为代表的人工智能技术因其海量数据的学习和应用能力,已成为卫星遥感影像解译发展的新方向。类型丰富、种类齐全的遥感影像解译样本数据可有效提升深度学习算法模型构建的准确性与自适应性,本文件对卫星遥感影像解译样本生产过程中的数据源内容、样本采集流程、样本集生产等内容进行全面细化和完善,满足解译样本生产各项工作中的流程化、标准化、规范化的工作需求。

卫星遥感影像解译样本生产技术规程

1 范围

本文件规定了卫星遥感影像解译样本生产的总体要求、工作流程、样本数据本采、样本集生产等内容。

本规范适用于卫星遥感影像解译样本生产工作。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T41149-2021 基础地理信息数据质量要求与评定
GB/T39608-2020 基础地理信息数字成果元数据
GB/T37151-2018 基于地形图标准分幅的遥感影像产品规范
GB/T35643-2017 光学遥感测绘卫星影像产品元数据
GB/T39613-2020 地理国情监测成果质量检查与验收
GB/T18316-2008 数字测绘成果质量检查与验收
GB/T13923-2006 基础地理信息要素分类与代码
GB/T13989-2012 国家基本比例尺地形图分幅和编号
CH/T9012-2011 基础地理信息数字成果 数据组织及文件命名规则

3 术语和定义

下列术语和定义适用于本文件。

元数据 Metadata

关于数据的数据。即数据的标识、覆盖范围、质量、空间和时间模式、空间参照系等信息。

样本集 Image Sample Set

指深度学习所用的一组特定样本类别和大小的样本小片数据的集合。

样本小片 Tile Sample

用于深度学习任务中按规范生产的特定样本类别和大小的遥感影像样本，样本小片包括影像小片和标签小片两部分。

影像小片 Image Tile

影像小片指样本小片中的样本范围内的影像数据。

标签小片 Label Tile

标签小片指样本小片中标记地物类别信息的数据，数据格式可以为矢量、栅格或文本等数据格式。

有效范围 Virtual Region

指样本采集中实际采集样本要素数据的空间范围。

4 总体要求

(1) 空间参考系：解译样本数据的坐标系统应采用 2000 国家大地坐标系，高程基准采用 1985 国家高程基准，采用的标准分幅数据的分幅方式和编号应符合 GB/T 13989 的规定。

(2) 时间参考：解译样本数据集中的日期应采用公历纪元，时间应采用北京时间。日期和时间属性值的格式应正确。各类数据的时间度量应准确、可靠，符合数据现势性要求。

(3) 逻辑一致性：解译样本数据的存储组织形式、格式和命名应符合规定要求，数据文件应不为空、

无缺少、无多余、无损坏、可读取。

(4) 位置准确度：解译样本采集采用的各项影像数据应符合规定的平面精度要求。采集的样本要素应与影像套和，偏移值不应超过规定的限差。

(5) 属性精度：样本采集成果数据的属性项的定义和属性值填写应符合规定要求，接边要素的属性值应一致。

(6) 完整性：样本采集矢量数据中要素的选取应符合样本类别要求,不应遗漏或多余。

5 样本采集数据源

5.1 遥感影像数据

5.1.1 光学影像

(1) 数据内容

光学影像数据指经过正射校正处理的单景或镶嵌裁切的卫星影像、航空影像实体数据及光学影像服务，影像数据需要包含影像信息的元数据文件。

(2) 数据要求

1) 影像质量：用于采集各类地物的影像，整体云量 $\leq 2\%$ ，影像中非常年性积雪等覆盖面积一般小于影像面积的 10%，且覆盖区域不能影响重要地物的判读，无影响影像信息判读和造成几何精度损失的纹理不清、噪声、模糊、扭曲、错位、漏洞等影像缺损现象。

用于采集各类影像质量问题样本的影像可按需采用各类问题影像，对一些原始数据质量问题的影像不做正射处理要求。

2) 空间参考：采用章节 4 规定的平面坐标系。

3) 格式要求：影像实体数据采用.tif或.img等常用影像格式存储，影像服务将服务访问地址存储在.txt文件中并以服务名称命名，元数据文件采用.xml格式存储。

5.1.2 高光谱影像

(1) 数据内容

高光谱影像数据指经过正射校正处理的单景或镶嵌裁切的高光谱卫星影像及机载影像实体数据，影像数据需要包含影像信息的元数据文件。

(2) 数据要求

1) 影像质量：影像清晰，整体云量 $\leq 2\%$ ，无影响影像信息判读和造成几何精度损失的纹理不清、噪声、模糊、扭曲、错位、漏洞等影像缺损现象。

2) 空间参考：采用章节 4 规定的平面坐标系。

3) 格式要求：影像实体数据采用.tif格式存储，元数据文件采用.xml格式存储。

5.1.3 SAR 影像

(1) 数据内容

SAR 影像数据指经过多视处理、斑点滤波、正射校正处理的单景或镶嵌裁切的 SAR 卫星影像和机载影像实体数据，影像数据需要包含影像信息的元数据文件。

(2) 数据要求

1) 影像质量：影像整体清晰，无影响影像信息判读和造成几何精度损失的纹理不清、噪声、模糊、扭曲、错位、漏洞等影像缺损现象。

2) 空间参考：采用章节 4 规定的平面坐标系。

3) 格式要求：影像实体数据采用.tif格式存储，元数据文件采用.xml格式存储。

5.2 已有成果数据

(1) 数据内容

已有成果数据指第三次土地调查、地理国情监测、土地变更调查等历史调查成果，包括矢量成果数据和影像成果数据和对应的元数据文件。

(2) 数据要求

1) 矢量成果内容：明确已有成果数据的数据内容，包括成果分类体系、成果类别定义、成果采集时间。

2) 影像成果内容：如果成果中包括对应的影像数据，明确影像数据的影像元数据信息，包括分辨率、影像类型、波段数、波段组合、卫星、传感器等。

3) 空间参考：矢量成果数据和影像数据应范围一致，空间坐标系调整至章节 4 规定采用的平面坐标系。

4) 格式要求：矢量文件采用.shp 文件格式，影像文件采用.tif 或.img 文件格式。

5.3 辅助数据

(1) 数据内容

辅助数据指样本采集过程中用于获取相应参考信息或属性信息的数据，包括 DEM、DSM 等数据。

(2) 数据要求

1) 空间参考：采用章节 4 规定的平面坐标系。

2) 高程基准：采用章节 4 规定的高程基准。

3) 格式要求：辅助影像数据采用.tif 或.img 格式存储。

6 样本生产工作流程

卫星遥感影像解译样本生产整体工作流程见图 6-1。

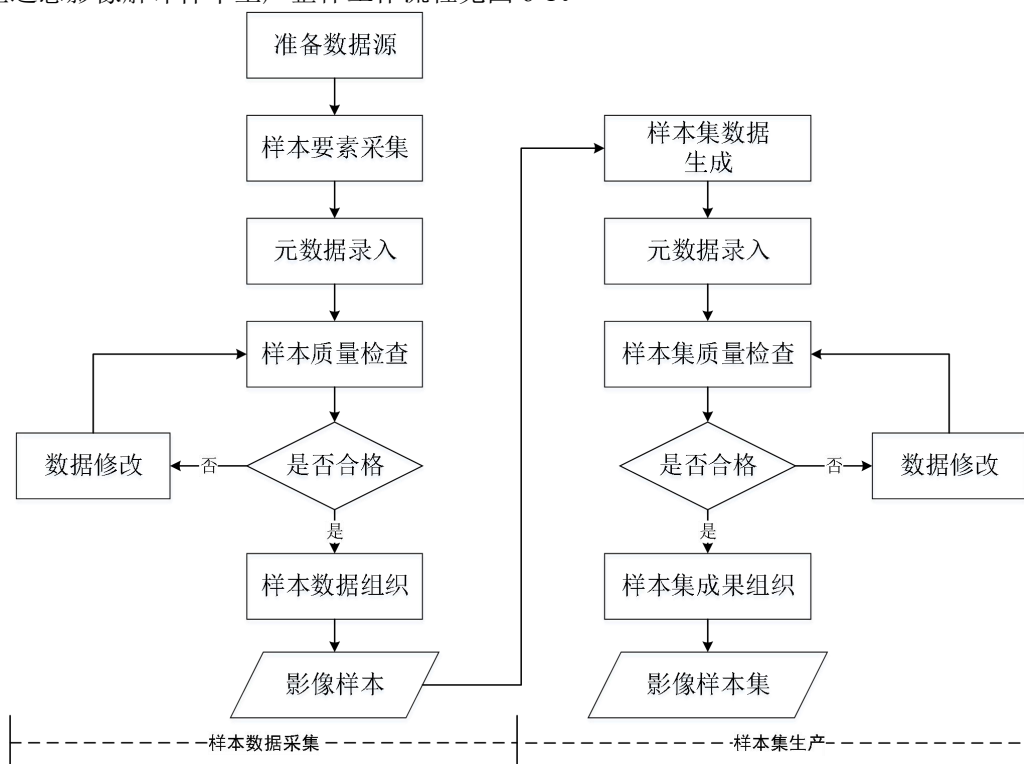


图 6-1 样本生产工作流程

7 样本数据采集

7.1 样本采集要求

7.1.1 拓扑要求

拓扑要求是对样本矢量要素采集的通用要求，主要包括以下三点：

(1) 要素拓扑关系正确，无不合理悬挂点、伪节点，不同要素间共点、共线、共面时公共部分应严格重合，相邻面间重叠和裂隙应小于规定容差值，要素不应存在不合理极小面、折刺、自相交等几何图形。

(2) 要素类中的多边形内部不重叠，多边形间可以共用顶点和边，或者完全不相连。

(3) 同一类别样本中的线要素不能重叠，线状要素可以相交或者交叉，但不能共用某一段。

7.1.2 要素边界要求

样本采集要素边界要求有以下几点：

(1) 边界与要素边缘基本吻合，边界尽量避免锯齿状和明显扭曲，特殊情况除外。

(2) 对于一些人工建设目标或较小的地物目标，边界精度在 2 个像素以内。

(3) 对于固定结构地物（如光伏、风车、飞机等），采取核心点和边界相结合的方式采集。

(4) 对于多独立要素组合成的场景要素，要以该场景的最大范围作为外边界。

(5) 对于大范围自然地表覆盖要素，受制于不同地物要素间过渡区域的界定困难，边界与要素边缘吻合度控制在 5 个像素以内。

(6) 不同季节、不同时间的自然地表覆盖要素可能存在重要的边缘、纹理、色调差异，要素边界要与影像数据对应，不允许采用不同时相影像数据采集同一套样本矢量要素。

(7) 对于地表变化要素，受制于前后影像上地物要素间过渡区域的影像分辨率，边界精度控制在 3 个像素以内。

(8) 地表变化要素的变化范围、变化程度与影像时相密切相关，要素边界要与前后期参考影像一一对应，不同时相间边界范围参考意义不大。

7.1.3 地物连通性要求

(1) 交通、水利等线状地物采集时需保持地物的连通性。

(2) 交通、水利等线状地物被权属界线分割的，按照不同图斑对待。

7.2 样本采集

7.2.1 准备数据源

基于待采集样本类别，选取划定样本采集区。样本采集区选取完成后，生成样本采集区范围矢量文件，数据格式采用.shp 数据格式。

根据确定的样本采集区，按照章节 5 规定的技术要求，准备相应区域样本采集所需的数据源。数据源选取要求如下：

(1) 如果已有成果数据只包括矢量数据，则需要按待采集样本类别或时相要求匹配相应的影像实体数据或影像服务，选取的影像数据分辨率应与该成果采集时的影像源分辨率一致，两期成果匹配的影像数据影像类型应一致。

(2) 各类数据源影像实体数据和已有成果数据需裁剪至样本采集区的各最小采集单元范围大小，影像服务数据不做裁剪要求。

7.2.2 样本要素采集

7.2.2.1 新数据采集

新数据采集即基于获取的影像数据或影像服务，采集规定的单类或多类样本矢量。

样本采集时，需要按照“应采尽采”原则，采集样本区内所有该类别样本要素，确保无遗漏；样本图斑需严格遵守各类型样本地物边界精度要求无明显拓扑错误；半自动采集得到的样本矢量，需要对未满足样本采集精度要求的图斑通过人工编辑方式进行数据修改至满足精度要求。

7.2.2.2 已有成果采集

已有成果采集即基于已有的矢量数据成果和对应影像数据，采集生成样本数据。已有成果采集的不同类型有：

(1) 已有成果数据包含影像成果或匹配了与成果采集时间一致的影像数据，则按照自动采集方式提取待采集样本类别图斑，提取的成果图斑精度不符合待采集样本精度要求时，需人工编辑修改样本图斑，保证样本数据精度。

(2) 已有成果数据不包括影像成果但匹配了与成果采集时间不一致的影像数据，则先按照自动采集方式提取样本图斑，基于影像数据人工采集图斑变化部分，生成与影像时间一致的样本图斑。

(3) 已有成果数据包含影像成果同时匹配了新时相影像数据，则先按照自动采集方式选取符合类别的图斑信息，基于新时相影像数据人工采集图斑变化部分生成变化类样本。

(4) 两期同类型已有成果数据包含影像成果或匹配了两期相应的影像数据，按照自动采集方式选取符合类别样本图斑，基于两期矢量成果自动生成变化类样本，提取的成果图斑精度不符合待采集样本精度要求时，需人工编辑修改样本图斑，保证样本数据精度。

7.2.3 样本元数据录入

采集的样本数据需要记录相应元数据，收集整理各项元数据信息，并录入的样本元数据文件中，样本元数据包括样本信息、影像信息、区域信息、制作信息等，具体内容如下：

- (1) 样本信息包括样本名称、样本类别名称、类别编码、样本时间、空间参考等。
- (2) 影像信息包括影像类型、分辨率、波段数、像素类型、拍摄平台、影像卫星、影像传感器等。
- (3) 区域信息包括样本省、市、县、样本区地形、气候等。
- (4) 制作信息包括制作单位、采集人、采集时间等。

7.3 样本质量检查

7.3.1 质量检查内容

依据 GB/T 17941，质量检查的主要内容如表 7-1 所示。

表 7-1 样本质检内容

质量元素	质量子元素	检查项	检查内容	适用成果
空间参考系	大地基准	平面坐标系	检查平面坐标系是否符合要求	所有成果
位置精度	平面精度	边界精度	图斑边界与影像套和精度	样本矢量成果
拓扑一致性	拓扑关系	重叠	要素图斑间重叠情况	样本矢量成果
		相接	图斑相接错误，如错误悬挂点	样本矢量成果

质量元素	质量子元素	检查项	检查内容	适用成果
		缝隙	检查图斑相接部分缝隙情况	样本矢量成果
完整性	完整性	要素多余	要素多余的情况	样本矢量成果
		要素遗漏	要素遗漏的情况	样本矢量成果
属性精度	属性正确性	属性项	属性项定义是否符合要求（名称、类型、长度等）	样本矢量成果
		属性值	属性值填写错误情况	样本矢量成果
元数据	元数据	项错误	元数据项错误或遗漏	元数据文件
		内容错误	元数据内容错误或遗漏	元数据文件
逻辑一致性	格式一致性	数据组织	数据组织方式是否符合要求	所有成果
		数据格式	数据格式是否符合要求	所有成果
		数据文件	数据文件是否缺失、损坏、	所有成果
		文件命名	数据文件名称是否符合要求	所有成果
影像质量	影像特性	色彩特征	检查影像色调不正常、明显失真	样本影像
		信息缺失	检查影像纹理不清、噪声、模糊、扭曲等情况	样本影像

7.3.2 质量检查流程

样本采集完成后，需要对样本数据进行质量检查，对于样本采集成果质检存在问题的数据部分，需记录错误类型，标记位置，并返回到样本采集作业端进行修改，直到满足各项质量检查要求，完成样本质检工作。

7.4 样本数据组织

采集完成的样本数据，按照采集任务和样本区层级组织文件目录，整理存储：

- (1) 首先按照采集任务样本文件夹，文件夹以采集任务名称命名，文件夹下存放各样本区文件夹。
 - (2) 各样本区文件夹以样本区命名，文件夹下存放该样本采集区数据文件。
 - (3) 样本采集数据包括样本矢量数据、样本影像数据、样本元数据文件和样本采集有效范围文件，文件以样本名称加后缀区分，各文件组织要求如下：
 - 1) 样本矢量文件采用.shp 格式。
 - 2) 样本影像文件采用.tif 或.img 通用格式，如果数据源为影像服务，则将影像服务地址记录到文件中，并保存为.txt 文档。
 - 3) 样本元数据采用.xml 格式。
 - 4) 样本采集有效范围文件存储该样本区范围矢量文件，采用.shp 格式。
- 样本采集成果文件组织如图 7-1。

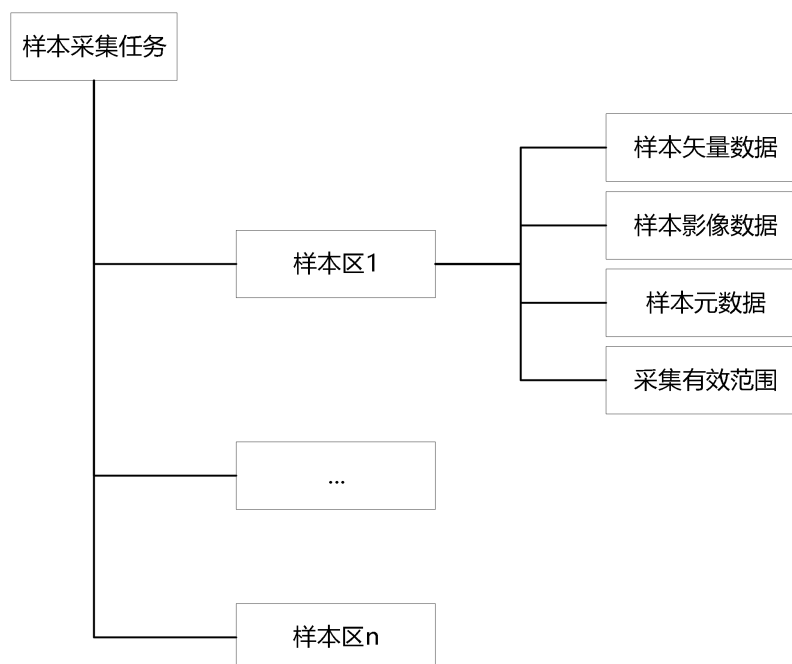


图 7-1 样本采集成果组织目录

8 样本集生产

样本集生产即将采集完成的样本成果数据裁切为规则大小的一组样本集。

8.1 样本集制作要求

样本裁切需要设置的各项条件如下：

(1) 样本尺寸：样本尺寸制定需考量不同样本类别地物特征及深度学习网络结构，设定为边长不小于 512 像素的正方形规则大小。

(2) 样本重叠度：样本重叠度代表单张样本小片间的重叠程度，设定为样本尺寸的一半。

(3) 样本有效占比：样本有效占比为标签小片中规定样本类别区域占整张样本小片大小的比例，制作时筛选有效占比高于 5% 的样本小片。

(4) 标签类别：标签小片的类别标注依据标准 onehot 编码，如果样本是单类地物样本，标签图 0 为背景，1 为目标地类，如果样本是多类地物，标签从 0 开始编号，其中 0 为背景，1 为第 1 类别的标注，以此类推。

(5) 样本小片名称：样本小片名称通常由样本集名称+样本小片编号组成，样本小片编号从 1 开始，位数不够时前面补 0。

(6) 样本数据格式：影像小片格式为.tif 或.jpg，.jpg 格式需要附带同名.aux.xml 文件记录小片空间范围信息，保证数据空间定位；标签小片格式为.png，需要附带同名.aux.xml 文件记录标签小片空间范围信息，保证空间定位。

8.2 样本集制作

8.2.1 样本集数据生成

样本集制作基于采集得到的影像样本数据，按照样本集制作的各项条件进行样本集小片裁切生成，具体裁切过程为：

(1) 空间范围裁切：根据遥感影像行列数，按照样本大小和样本步长，可以计算得到每张样本范

围的行列号并转换为空间范围，行列号转换为空间坐标具体计算公式如下：

$$X_{geo} = Geo[0] + X_{pixel} * Geo[1] + Y_{pixel} * Geo[2]$$

$$Y_{geo} = Geo[3] + X_{pixel} * Geo[4] + Y_{pixel} * Geo[5]$$

其中， X_{geo} 、 Y_{geo} 代表地理空间坐标， X_{pixel} 、 Y_{pixel} 代表影像行列号， $Geo[0]$ 和 $Geo[3]$ 代表图像左上角坐标（地理坐标）， $Geo[1]$ 和 $Geo[5]$ 代表影像分辨率（东西方向为正，南北方向为负）， $Geo[2]$ 和 $Geo[4]$ 一般取 0。

（2）样本小片生成：根据单张样本的空间范围，裁剪对应范围的影像数据和标注矢量数据，基于裁切得到的小片范围的影像数据和标注矢量数据，根据小片样本集制作格式要求，导出生成规定格式和命名规范的影像小片和标签小片。

（3）样本小片元数据生成：基于样本小片元数据的内容，计算该张小片元数据信息并记录小片元数据文件。

（4）循环制作过程直到所有样本裁剪完成，获得该遥感影像初始样本数据集。

（5）按照样本有效占比条件筛选样本集数据，将不符合条件的样本小片数据剔除。

8.2.2 元数据录入

样本小片裁切生成后，需要整理记录样本集元数据，并录入到样本集元数据文件中，元数据.xml 文件中的各项元数据项标记采用元数据项英文名称。元数据内容基于样本采集制作的影像样本元数据内容，并增加各类小片路径、样本尺寸等样本信息。

8.2.3 样本集质量检查

样本集制作完成后，需要对样本集数据进行质量检查，如果样本集数据质检存在问题，则修改或重新生成样本集数据，直到满足各项质量检查要求，完成样本质检工作。样本集质检项主要包括空间参考系、元数据、逻辑一致性内容，具体质检项内容参照章节 7.3.1。

8.3 样本集成果组织

样本集由 n 组样本小片和样本集元数据组成。

（1）样本小片：一组样本小片包括影像小片、标签小片和小片元数据。

1) 影像小片：裁切后的规则大小的影像数据，地表变化类影像小片包括前后时期两份影像小片。

2) 标签小片：与影像小片对应的栅格化的样本类别标签数据。

3) 小片元数据：记录单组小片数据的名称、样本有效占比等相关样本小片属性信息。

（2）样本集元数据：记录整组样本集的对应关系及相应的影像信息和分类信息的样本集元数据信息，以便数据建库管理，样本集元数据文件为.xml 格式。

样本集以文件夹形式组织存储，影像小片和标签小片和小片元数据以同名形式存储在不同文件夹中，地表变化类影像小片在影像小片文件夹中分前期后期存储在不同文件夹中，样本集以采集任务+样本区命名，样本数据组织形式如图 8-1。

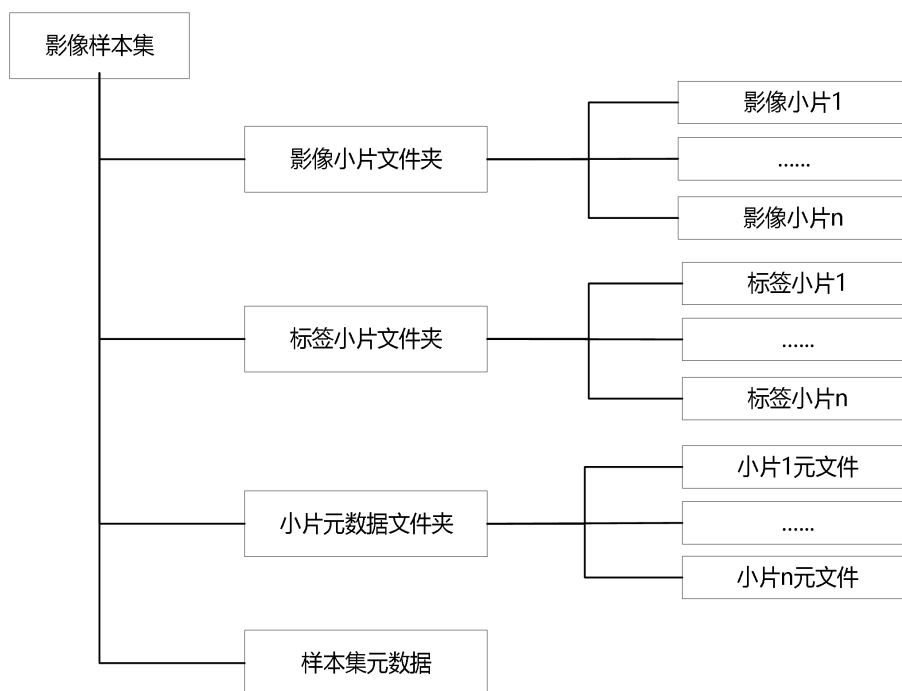


图 8-1 样本集组织目录

附录 A
(资料性附录)

卫星遥感影像解译样本元数据填写规范

样本集具体元数据内容及填写规则见表A.1。

表 A.1 影像样本小片集元数据

组	项名	英文名称	定义	约束/条件	数据类型	域
样本信息	名称	SetName	样本集名称	M	字符串	自由文本
	影像小片	ImagePath	影像小片存储路径	M	字符串	自由文本
	影像小片旧	ImagePathOld	地表变化类样本旧时相影像小片路径	C/地表变化类旧时相	字符串	自由文本
	标签小片	LabelPath	标签小片存储路径	M	字符串	自由文本
	样本小片元数据	MetaPath	样本小片元数据文件存储路径	M	字符串	自由文本
	样本尺寸	SampleSize	样本尺寸大小	M	整型	>0
	分类编码	ClassCode	样本所属分类编码	M	字符串	依据样本分类体系
	分类名称	ClassName	样本所属分类名称	M	字符串	依据样本分类体系
	样本时间	SampleDate	影像数据获取时间(地表变化类样本为新时相时间)	M	日期型	CCYY-MM-DD (GB/T 7408-2005)
	样本时间旧	SampleDateOld	地表变化类样本为旧时相时间	C/地表变化类旧时相	日期型	CCYY-MM-DD (GB/T 7408-2005)
空间参考	SpatialRef	影像采用的空间参照系	M	字符串	'EPSG:' + 坐标系代号, 例如CGCS2000为'EPSG:4490'	
影像信息	影像类型	ImageType	遥感影像数据类型	M	字符串	光学、SAR、高光谱
	影像分辨率	Resolution	影像地面分辨率	M	实型	>0.0, 单位为米
	影像分辨率旧	Resolution	影像地面分辨率	M	实型	>0.0, 单位为米
	波段数	BandCount	影像包含的波段数量	M	整型	>0
	波段组合	BandStru	影像波段的组成信息	M	字符串	波段颜色大写表示, 如RGB, RGBNir
	像素类型	PixelType	影像像素存储的像素类型	M	字符串	整数像素位深加类型小写, 例如

组	项名	英文名称	定义	约束/条件	数据类型	域
						8usinged
	拍摄平台	Platform	影像拍摄载荷平台	M	字符串	星载、机载
	影像卫星	Satlite	影像获取的卫星平台	O	字符串	通用缩写大写，通用全称首字母大写，例如 GF1、WorldView-4
	影像卫星旧	SatliteOld	影像获取的卫星平台	C/地表变化类旧时相卫星影像	字符串	通用缩写大写，通用全称首字母大写，例如 GF1、WorldView-4
	影像传感器	Sensor	影像获取的传感器类型	O	字符串	大写英文
	影像传感器旧	SensorOld	影像获取的传感器类型	C/地表变化类旧时相卫星影像	字符串	大写英文
区域信息	省	Province	样本数据所在省名称	M	字符串	标准全称
	市	City	样本数据所在市名称	M	字符串	标准全称
	县	County	样本数据所在县名称	O	字符串	标准全称
	地形	Landform	样本区地形特征	M	字符串	依据样本分区信息填写
	气候	Climate	样本区气候特征	M	字符串	依据样本分区信息填写
制作信息	制作单位	ProductionUnit	样本制作单位	M	字符串	自由文本
	采集人	Producer	样本制作人员	M	字符串	自由文本
	采集时间	CollectionTime	样本采集时间	M	字符串	CCYY-MM-DD(GB/T 7408-2005)

附 录 B
(资料性附录)
卫星遥感影像解译样本集小片元数据内容及规范

遥感影像解译样本集具体小片元数据内容及填写规则见表B.1。

表 B.1 样本小片元数据内容

项名	英文名称	定义	约束/条件	数据类型	域
样本小片编号	TileID	样本小片序号	M	字符串	自由文本
样本大小	TileSize	样本小片尺寸大小	M	整型	>0
有效占比	RegionRatio	样本类别有效区域比例	M	实型	>0.0

参 考 文 献
