

样本库设计方案

**苏州中科天启遥感科技有限公司**

**2019 年 3 月 25 日**

**文档历史**

|  |  |  |  |
| --- | --- | --- | --- |
| 文档版本 | 发布时间 | 修改内容 | 修改人员 |
| 0.1 | 2019.03.25 | 创建 | 陶明明 |

# 1、样本库设计原则

一个样本由一个样本影像、1个/多个样本矢量标签、1个/多个模板标签图像构成。为了合理有效对样本数据进行管理，建立样本数据之间的对应关联关系，便于对样本进行入库、查询、查看、导出、转换等管理。

# 2、样本库的组织结构

根据地物要素类别，样本库以树状结构的形式进行描述，根据实际需要可以对样本库的组织结构分支节点进行扩充。样本库的存储结构示例如图 1所示：



图 1样本库的组织结构图

样本库的存储结构通过地物类别描述字典的形式进行描述，并以xml的形式进行管理，其描述示例如下：

<feature\_type>

<object\_category>

<subcategory >.....</subcategory>

</object\_category>

</feature\_type>

在实际应用中，样本库类别字典通过地物类别字典编码的形式进行检索和访问。地物类别字典编码通过6位的十进制数字来表示，从高到低，第一位表示地物要素类别，第二和第三位表示地物类别，后三位表示具体的地物类型，各字段的值为该地物在地物类别描述字典的顺序值（以1作为起始顺序），如图 2所示，耕地的地物类别标签字典编码为203001。



图 2 地物类别标签字典编码定义示例

# 3、样本库的目录定义

根据样本库的组织结构，以及样本标签类型，将样本存储在分支节点所在的目录，该目录为父节点所在目录的子目录。样本库的目录共有四级，分别如下：

第一级：根目录，data

第二级：地物要素类型，point/line/polygon

第三级：地物类别，以point为例，用地物类别名称的拼音作为目录名称，如yancong/kuangjing/tianwentai/…

第四级：样本数据。

# 4、样本信息描述

样本数据由样本影像、样本矢量、标签图像以及样本信息描述文件组成。其中，样本信息描述文件用于对样本基本信息的描述，以xml格式保存在地物类别目录中，其元素字段包括以下基本信息：

1. 样本影像描述信息（image\_info）：包括字段信息如下
   1. 影像相对路径（image\_path）；
   2. 影像时相（time\_phase）;
   3. 卫星平台信息（satellite）；
   4. 空间分辨率（spatial\_resolution）;
   5. 影像分辨率（image\_ resolution）;
   6. 影像高度（image\_hight）;
   7. 影像宽度（image\_width）;
   8. 影像波段数（band\_count）;
   9. 影像波段组合（band\_combination）。
2. 标签图像描述信息（label\_info），包括字段信息如下：
   1. 标签图像相对路径（image\_path）；
   2. 图像高度（image\_hight）;
   3. 图像宽度（image\_width）。
3. 标签图像类别描述信息（class\_info），包括字段信息如下：
   1. 类别标签字典编码（class\_code）；
   2. 类别灰度值（class\_value）。

样本信息描述文件示例如下：

<sample\_info>

<image\_info>

<image\_path >…</image\_path>

<time\_phase >…</time\_phase>

<satellite >…</satellite>

<spatial\_resolution>…</spatial\_resolution>

<image\_resolution>…</image\_resolution>

<image\_height>…</image\_width>

<image\_width>…</image\_width>

<band\_count>…</band\_count>

<band\_combination>…</band\_combination>

</image\_info>

<label\_info>

<image\_path >…</image\_path>

<image\_height>…</image\_width>

<image\_width>…</image\_width>

</label\_info>

<class\_info>

<class\_code>…</class\_code>

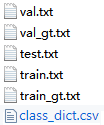
<class\_value>…</class\_value>

</class\_info>

</sample\_info>

# 5、训练样本导出文件

根据特定的任务及样本描述约束条件，从样本库中自动检索需要的样本影像与标签图像，并生成样本列表，包括训练集的样本列表及对应的ground truth列表、验证集的样本列表及对应的ground truth列表、测试集样本列表、目标类别属性信息，包括以下文件：



# 6、样本库管理配套工具组成及接口关系

样本库管理配套工具及接口关系如图 3所示。



图 3样本库管理配套工具及接口关系图