

CN-MSLU-DEMO-1K Dataset Description

Based on [CN-MSLU-100K](#), we have created the example dataset CN-MSLU-DEMO-1K for you to better understand the characteristics and applicability of the dataset. In this description file, we provide basic information about the dataset, as well as sample code for exploring the data.

Overview

The CN-MSLU-100K dataset consists of over 100,000 irregular remote-sensing land parcel images. Combining the "[Classification and Planning Standards for Urban Land Use](#)" (GB 50137-2011) and Alibaba's "[AMAP POI](#)", we have categorized the main features depicted in the remote sensing images into 5 major classes as "Residential Districts", "Commercial Zones", "Industrial Land", "Public Services", "Agriculture and Nature". Each major category is subdivided into secondary categories, totaling 22 sub-categories.

In addition, during the labelling process, we also obtained a smaller number of "Transportation Facilities", and large amount of "Unknow Landuse" categories which are difficult to judge due to insufficient information on land parcels, and included them in the dataset. The final dataset contains 7 categories and 28 sub-categories, please refer to the table in the appendix for the description and count number of first and second level categories.

In the CN-MSLU-100K dataset, we extracted 200 images from each of the five main categories to produce the CN-MSLU-DEMO-1K dataset for a better understanding of the characteristics and applicability of the dataset.

Dataset stats

File directory structure

The file structure is shown in Figure 1 as follows:

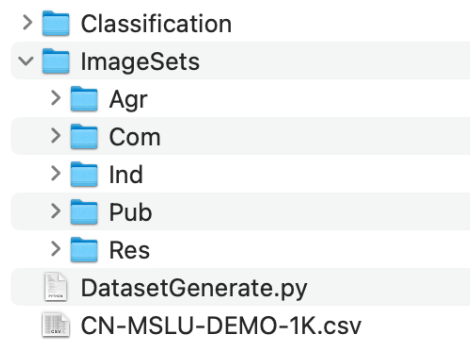


Figure 1 Dataset File Structure

The description for each folder and file is shown in Table 1

Table 1 The description for each folder and file

Folder or file name		Format	Description
Classification		Folder	All data description files in xml format, containing information such as category, path, image size, etc.
ImageSets		Folder	Remote-sensing land parcel images images of datasets saved by category
	Agr	Folder	Agriculture and Nature
	Com	Folder	Commercial Zones
	Ind	Folder	Industrial Land
	Pub	Folder	Public Services
	Res	Folder	Residential Districts
DatasetGenerate.py		Python Scrpit	Code for organizing all XML files into CSV format dataset
CN-MSLU-DEMO-1K.csv		csv	A csv table generated using DatasetGenerate.py to store dataset information. Contains all data categories, file names, storage paths, image widths, image heights, geographic information, first level class names, second level class names

Demo code

The code `DatasetGenerate.py` reads the xml file in a multi-threaded way, extracts the basic information such as file name, storage path, land use category, etc. of each piece of data described in it, and saves it as a csv file. When making the dataset, the data can be quickly manipulated by reading `CN-MSLU-DEMO-1K.csv` directly.

When using the code, the number of threads can be adjusted to run faster depending on the configuration of the computer being used.

Dataset presentation

The data in the **Classification** folder is saved as a structured XML format file as shown in Figure 2. The XML file contains basic information such as the image path of the parcel, the size of the parcel image, the geographic information of the parcel, as shown in Figure 3.

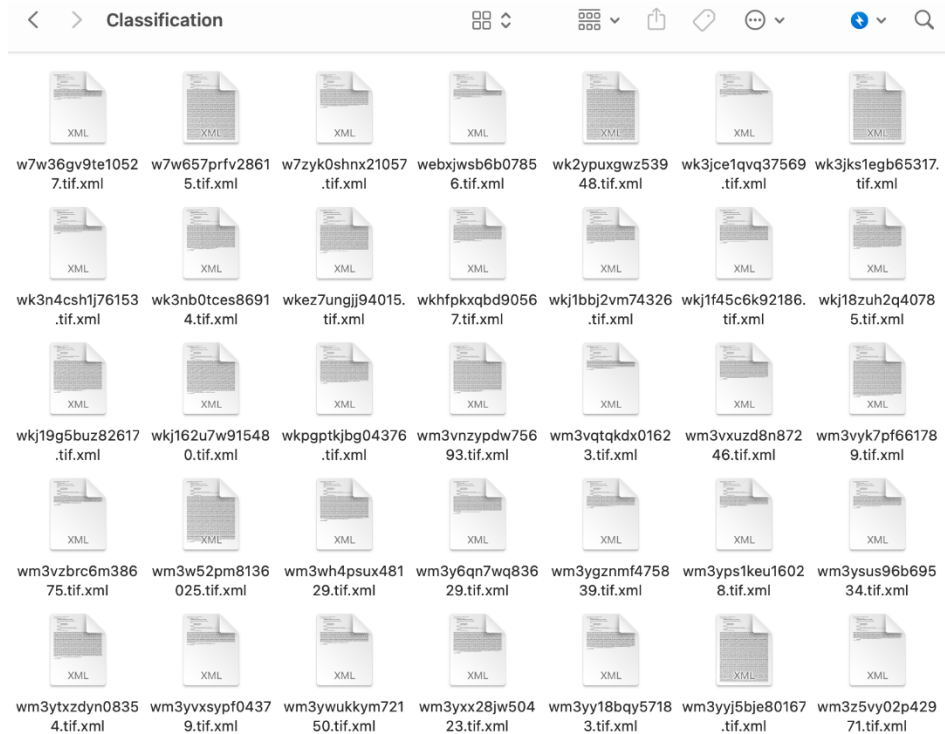


Figure 2 Overview of xml files in Classification folder

```
1 <?xml version="1.0" encoding="utf-8"?>
2 <classification>
3   <folder>Agr</folder>
4   <filename>w7w36gv9te10527.tif</filename>
5   <source>
6     <database>CN-MSLU-DEMO-1K</database>
7   </source>
8   <size>
9     <width>121</width>
10    <height>123</height>
11    <depth>3</depth>
12  </size>
13  <category>
14    <firstlevel>Agriculture and Nature</firstlevel>
15    <secondlevel>Forestland and Grassland</secondlevel>
16  </category>
17  <geoinfo>
18    <coord>GCJ02</coord>
19    <polygon>110.1664151,19.9286467;110.1664217,19.9286553;110.1664273,19.9286673;110.1664342,19.9287142;
110.1664382,19.9287241;110.1664466,19.9287361;110.1664484,19.9287416;110.1664629,19.9289011;110.1664
65,19.928912;110.1664722,19.9289269;110.1665626,19.9290575;110.1665693,19.9290654;110.1665774,19.929
0719;110.1667532,19.9291876;110.1668248,19.929238;110.1668425,19.9292462;110.1668778,19.9292496;110.
1689608,19.9293838;110.1689765,19.9293825;110.1689912,19.9293767;110.1690036,19.9293669;110.1690101,
19.9293585;110.1690164,19.9293441;110.1690182,19.9293337;110.169015,19.9291995;110.168999,19.9287737;
110.169031,19.927771;110.1690649,19.9271387;110.1690633,19.9271225;110.169057,19.9271076;110.169050
4,19.9270989;110.1690423,19.9270918;110.1690328,19.9270864;110.1690171,19.9270821;110.1665748,19.926
8008;110.1665643,19.926801;110.1665491,19.926805;110.1664252,19.9268579;110.1664129,19.9268671;110.1
664036,19.9268793;110.1663995,19.9268887;110.1663969,19.9269038;110.1663761,19.9276046;110.1664012,1
9.9286119;110.1664026,19.9286227;110.1664061,19.9286331;110.1664151,19.9286467</polygon>
20    <center>110.16771178050843,19.92810001608732</center>
21  </geoinfo>
22 </classification>
```

Figure 3 Overview of xml format file record data content

The data in the `ImageSets` folder is shown in Figure 4. Remote sensing image parcels are stored in different folders according to their categories, and the name of each folder is an abbreviation of the corresponding category.



Figure 4 Overview of Remote sensing image parcels in `ImageSets` folder

Related projects

The following link will go into more detail about the dataset we are using and the projects we are working on based on that dataset. It also includes the email addresses of the people responsible, so if you have any questions, please feel free to contact us!

[CN-MSLU-100K: Land Use Classification Dataset at Block Scale for Multi-source Spatio-temporal Data - 城市之光 - City of Light \(urbancomp.net\)](https://urbancomp.net/)

Appendice: The description and count number of first and second level categories of CN-MSLU-100K Dataset

First Level Category	Second Level Category	The amount of data
Residential Districts (Res) 40682	Rural Homestead	1549
	Rural Architecture and Farmland	14148
	High-rise Residential Buildings	20884
	Villas and High-end Residences	1864
	Urban Villages	2237
Commercial Zones (Com) 6684	Business Tower	978
	Commercial Entertainment	588
	Office Campus	2708
	Commercial Market	1125
	Shopping Center and Commercial Street	1125
Industrial Land (Ind) 24498	Hotel	160
	Industrial Park and Factory	21593
Public Services (Pub) 6286	Construction Site	2904
	Party and Government Institutions	719
	Non-profit Public Institutions (Museum; Stadium; Hospital)	917
	Educational and Research Institutions	2580
Agriculture and Nature (Agr) 21411	Parks and Squares	2070
	Mountain	2484
	Forestland and Grassland	6916
	Water	2260
	Farmland	7293
Transportation Facilities (Tra) 799	Wasteland	2458
	Transport facilities (Car Park; Gas Station; Service Station)	290
	Transportation hub (Subway; Bus or Train Station; Airport)	366
Unknow Landuse (Unk) 25069	Highway & Track	143
	Lack of Information	5753
	Invalid Land Parcel (Small-sized & Narrow)	2776
	Mixed Landuse	16540