Towards a crowdsourcing platform for labeling remote sensing images online

Computer Network Information Center
Chinese Academy of Sciences
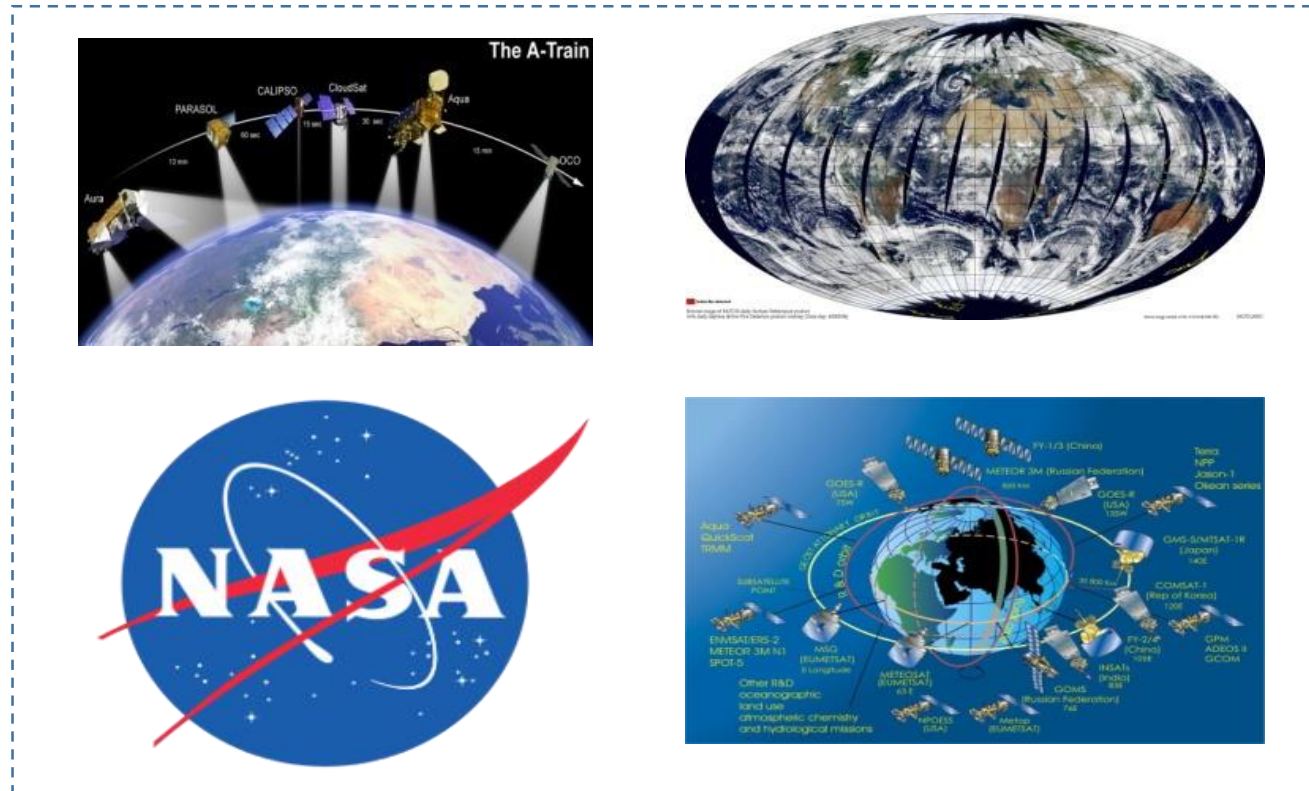
Jianghua Zhao(zjh@cnic.cn)

March 21, 2018

Remote Sensing is never a strange concept.
More than 1000 operational satellites are currently in orbit around Earth, generating large amounts of remote sensing data.

In fact, nowadays, there are thousands of man-made satellites orbiting the earth.

From 2008, the United States Geological Survey(USGS) began opening all geospatial datasets to the public for free.

1 Background

- Remote sensing data are associated with geographic locations, so they provide particularly critical information.
- They are applied widely and are even transforming diverse fields.
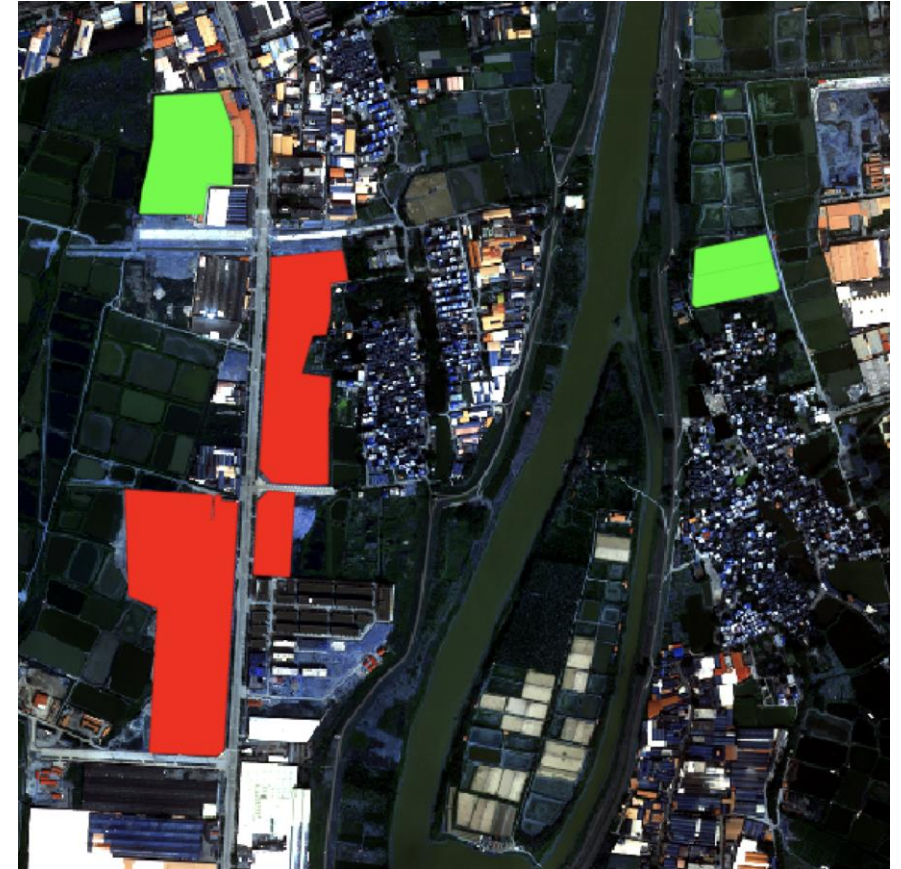

Global change detection

Monitoring ecosystem

Disaster Prevention and Mitigation

City planning

- The RS images need to be processed to extract land cover information. In computer vision, it is called semantic segmentation.

Satellite images and land approval records are stacked together. The red/green polygons are different land development projects approved by the government in 2015.

1 Background

Semantic segmentation research

input image   filter bank   texton map (colors ⇔ texton indices)

clustering and assignment
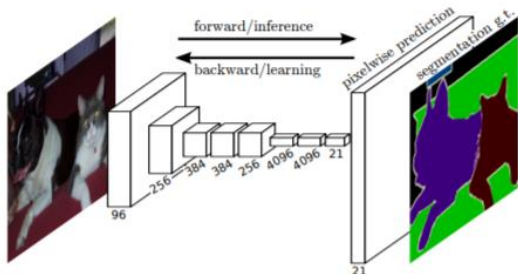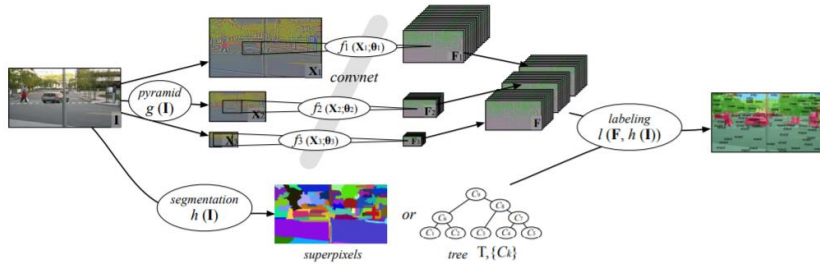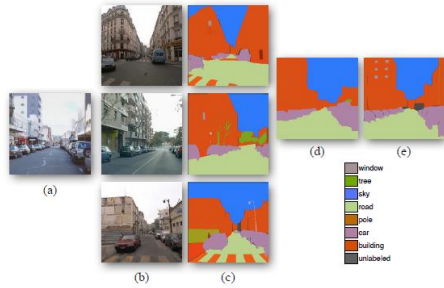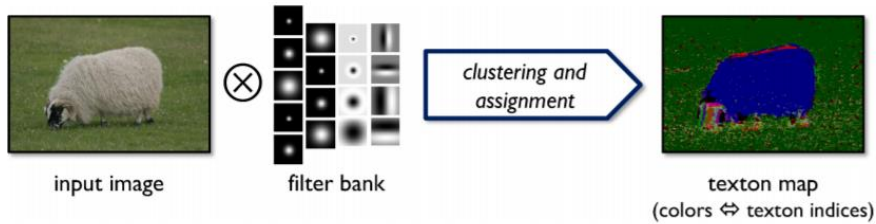
TextonBoost--parametric method representative

Shotton, Jamie, et al. "Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context." IJCV, 2009.

Label Transfer—non-parametric method representative

Liu C, Yuen J, Torralba A. Nonparametric scene parsing via label transfer[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(12): 2368-2382.

Learning Hierarchical Features for Scene Labeling-- the first algorithm to apply deep learning in semantic segmentation
Farabet, C., Couprie, C., Najman, L., & LeCun, Y. (2013). Learning hierarchical features for scene labeling. *IEEE transactions on pattern analysis and machine intelligence(*TPAMI*), 35(8), 1915-1929.*

Fully Convolutional Networks for Semantic Segementation -- the best paper candidates for CVPR
Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." *Proceedings of the IEEE conference on computer vision and pattern recognition.* **2015**.

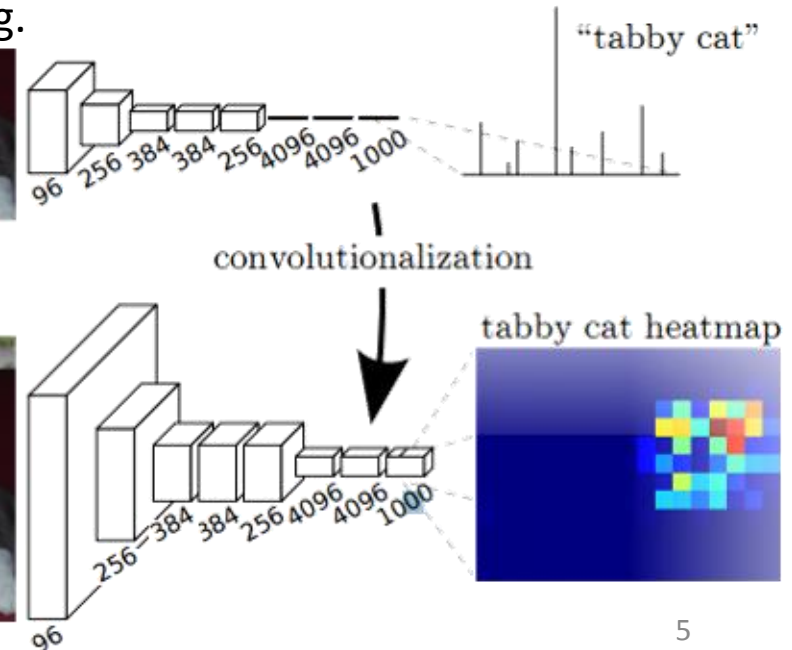segNet, DeconvNet, U-net, PSPNet, dCNN, etc.

4

1 Background

Deep convolutional neural networks are suitable for image interpretation task.

A deep convolutional neural network (dCNN) stacks many convolutional layers. Each layer acts as a feature-detector.

- Early layers will recognize simple, locally-confined features like lines, edges and patches of colour.
- Each successive layer will recognize features that are visually more and more complex.
- The output layer produces the classification of the whole image.
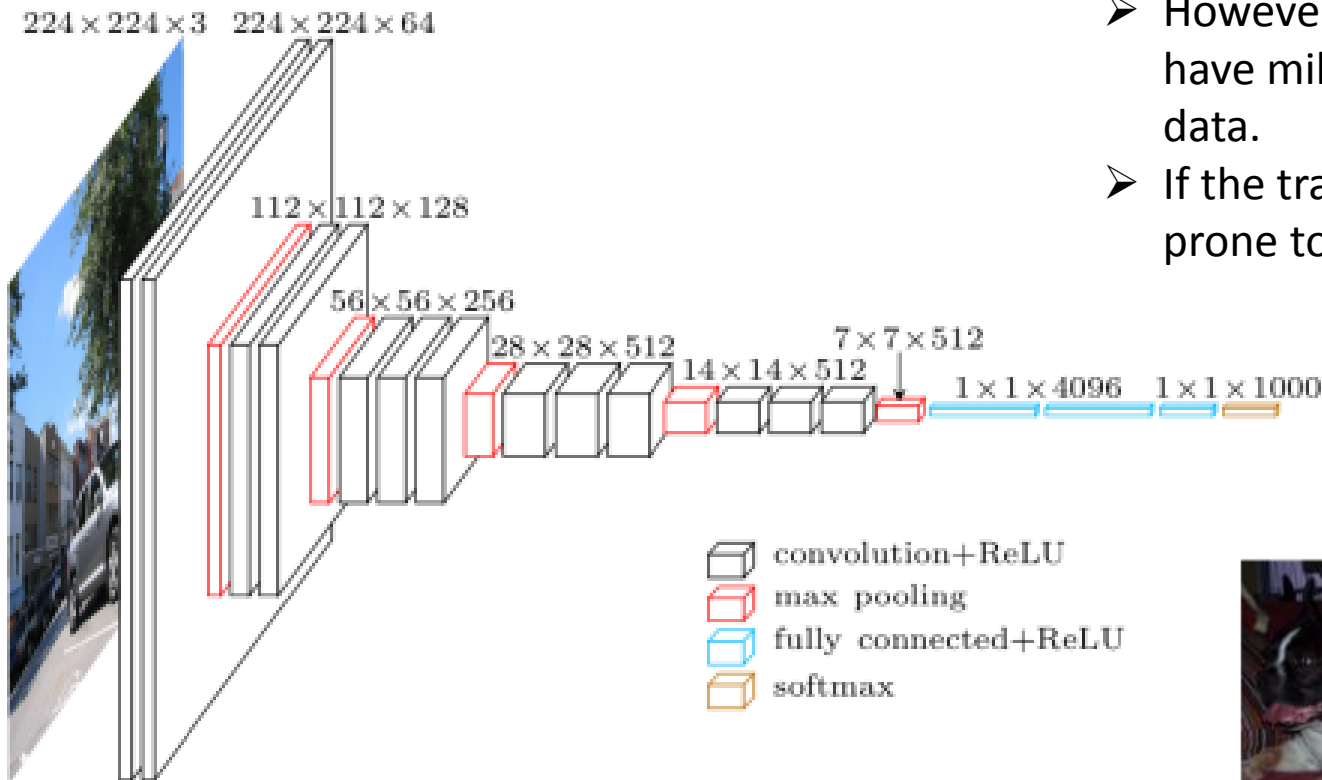
➢ However, very deep networks with many hidden layers have millions of parameters. So it requires a lot of training data.

➢ If the training set is not large enough, the networks are prone to be over-fitting.

Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//
IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2015:3431-3440.

1 Background

Crowdsourcing is the act of outsourcing work to a large group of unknown people (Howe, 2006).

As each human brain has great image-processing abilities, crowdsourcing seems to be a promising solution for generating labeled training dataset.

- The ImageNet dataset have been widely used in computer vision.
- Over 14 million URLs of images have been hand-annotated by ImageNet to indicate what objects are pictured.
- In at least one million of the images, bounding boxes are also provided.
- All these are completed on Amazon Mechanical Turk, a crowdsourcing platform

#Y	#N	Conf Cat	Conf BCat
0	1	0.07	0.23
1	0	0.85	0.69
1	1	0.46	0.49
2	0	0.97	0.83
0	2	0.02	0.12
3	0	0.99	0.90
2	1	0.85	0.68

User 1	Y	Y	Y
User 2	N	Y	Y
User 3	N	Y	Y
User 4	Y	N	Y
User 5	Y	Y	Y
User 6	N	N	Y

Figure 7: Left: Is there a Burmese cat in the images? Six randomly sampled users have different answers. **Right:** The confidence score table for "Cat" and "Burmese cat". More votes are needed to reach the same degree of confidence for "Burmese cat" images.

Amazon Mechanical Turk

Deng, Jia, et al. "Imagenet: A lar... uter Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009.

➢ LabelMe is an online annotation tool that allows sharing and labeling of images for computer vision research.

➢ The tool has been online since August 2005 and has accumulated over 400 000 annotated objects.

➢ The online tool provides functionalities for drawing polygons to outline the spatial extent of object in images, querying for object annotations, and browsing the database

LabelMe is Crowdsourcing-based labeling platform.

Bryan C. Russell，LabelMe: A Database and Web-Based Tool for Image Annotation，2007
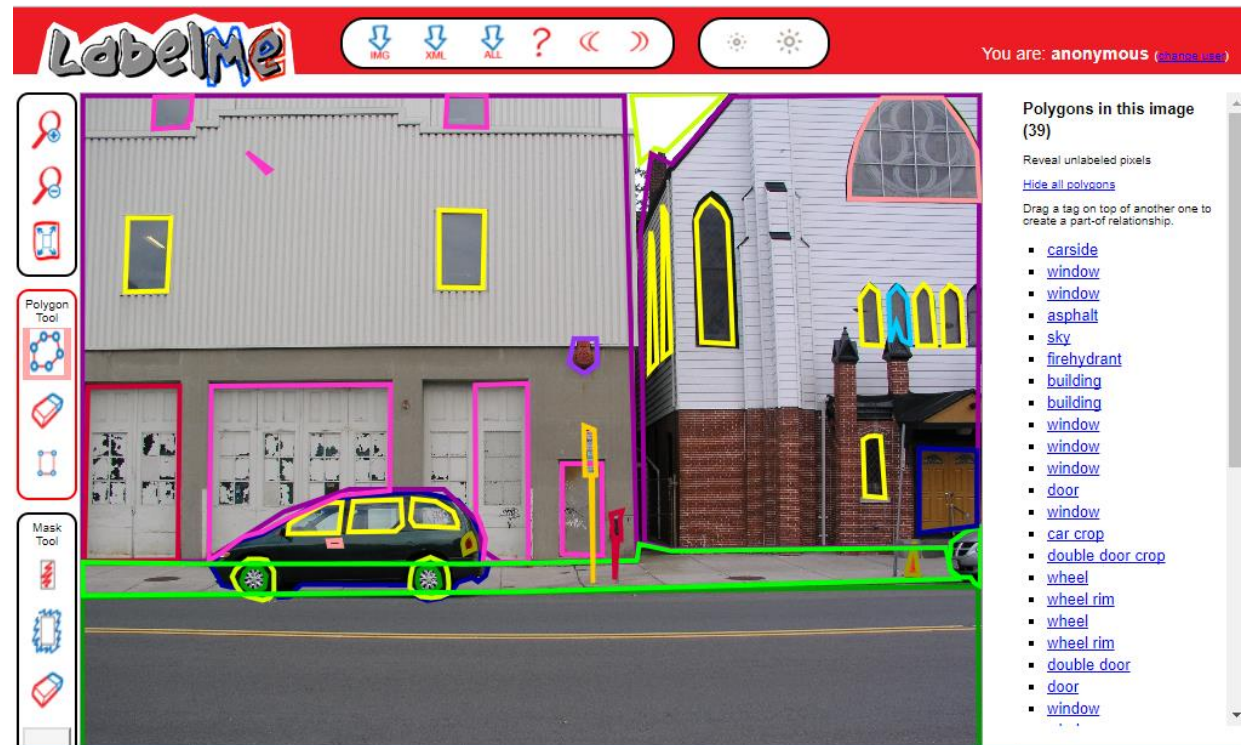
1 Background

DIYlandcover is an open-sourced platform to extract landcover types based on crowdsourcing.


http://mappingafrica.princeton.edu/
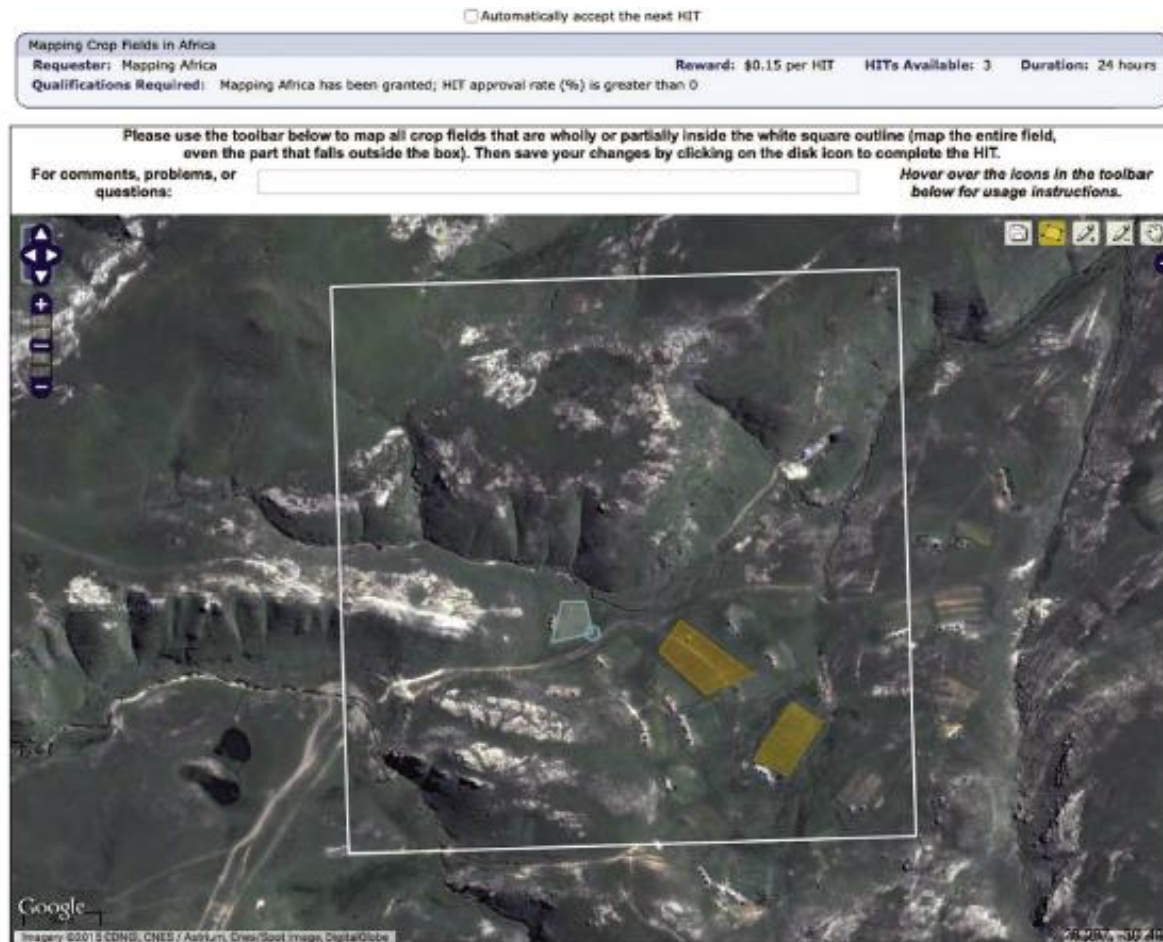
➤ It serves representative samples of high resolution imagery to an online job market Mechanical Turk (MT).
➤ Workers delineate individual landcover features.
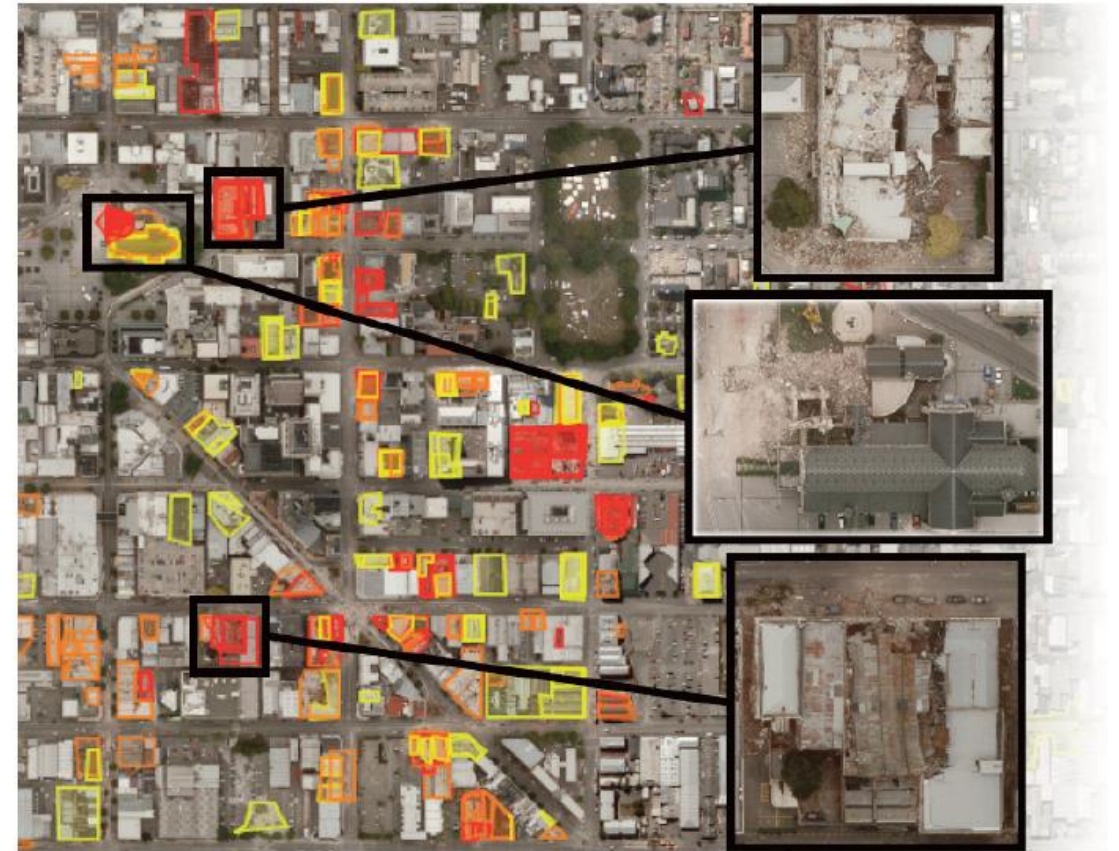➤ Their mapping skills are frequently assessed.

L.D. Estes et al. / Environmental Modelling & Software 80 (2016) 41–53
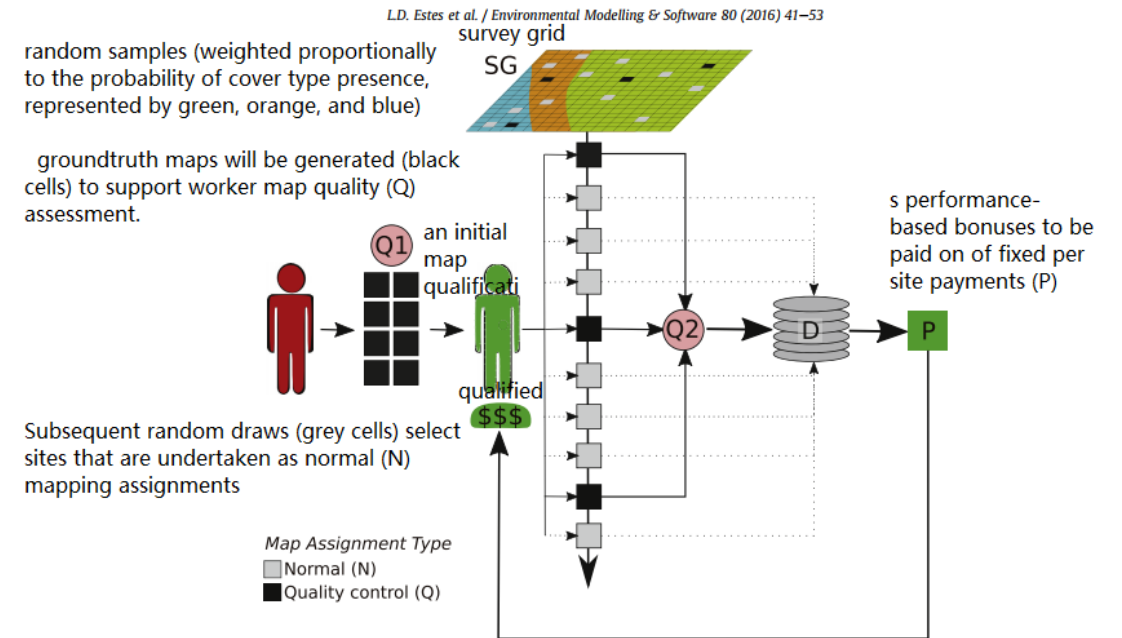
The authors also obtained South African cropland map through DIYlandcover 91% accuracy, which exceeds the accuracy of current generation global landcover products.

1 Background

Geo-Wiki is a web-based geospatial portal with open access to Google Earth.
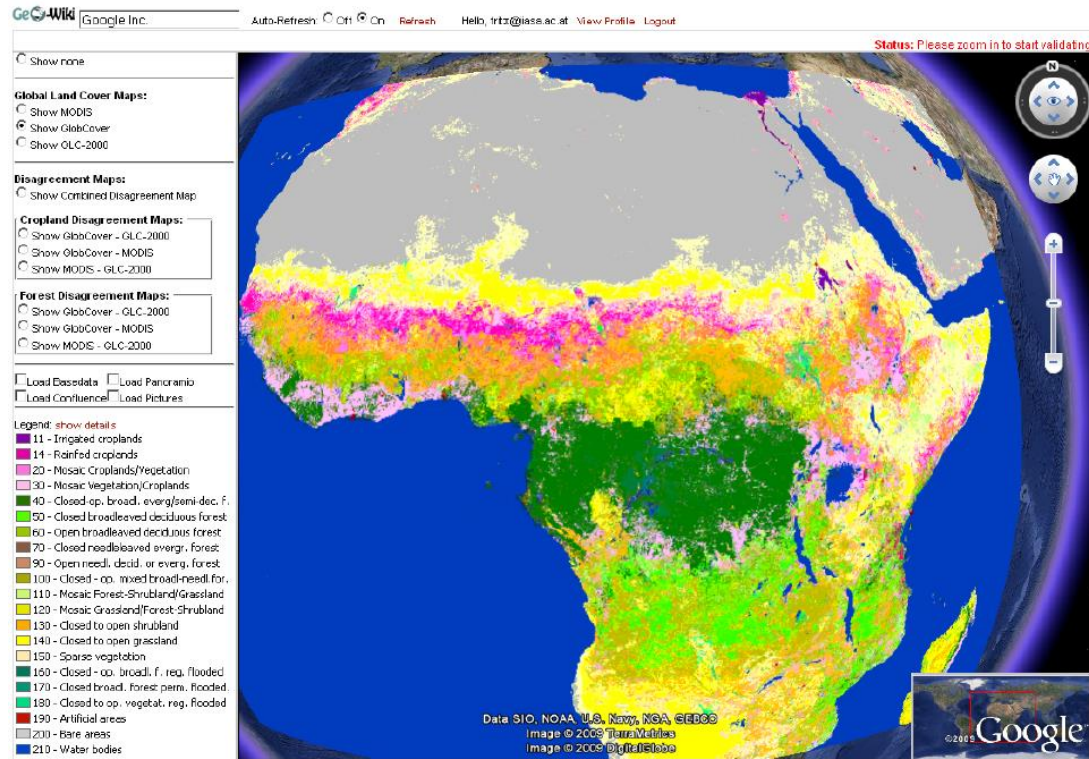
Virtual Disaster Viewer is a social networking tool using crowdsourced analysis of remote sensing imagery for earthquake impact and damage assessment.

Experts or the public can use the high resolution satellite imagery from Google Earth or Bing maps to validate land cover products such as GLC-2000, MODIS, and GlobCover on the web. Thus it is able to create a land cover product with higher accuracy.

Sample of crowdsourced damage map of Christchurch, New Zealand. Each polygon was drawn by a crowd contributor and marked with a damage level (red: complete destruction, orange: very heavy damage, yellow: substantial damage).

9

1 Background

Training datasets for image semantic segmentation

Categories: 23
images: 591
Contributor: Microsoft(J. Shotton et al.)

Categories: 183
images: 30369
Contributor: B. Russell et al.

Categories: 32
images: 701
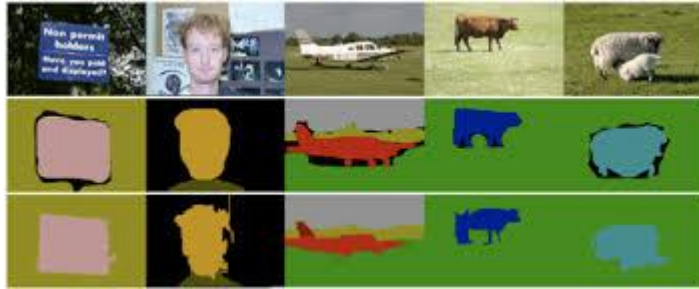Contributor: G. Brostow et al.

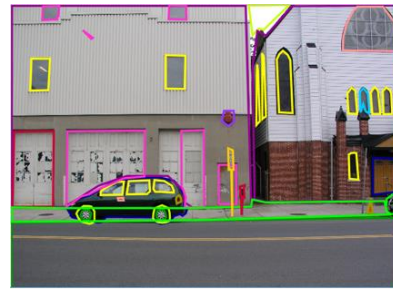# Categories: 33
images: 2688
Contributor: C. Liu et al.

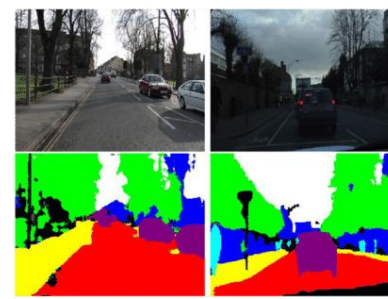# Categories: 21
images: 2913
Contributor: M. Everingham e

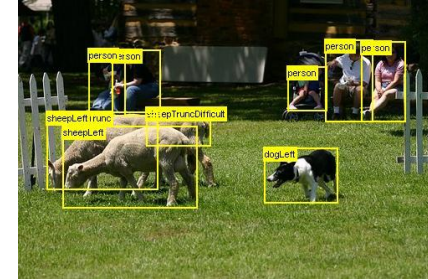| MSRC	LabelMe	CamVid	SIFTFLOW	PASCAL VOC
2005	2008	2009	2011	2012

2014		2015	2017
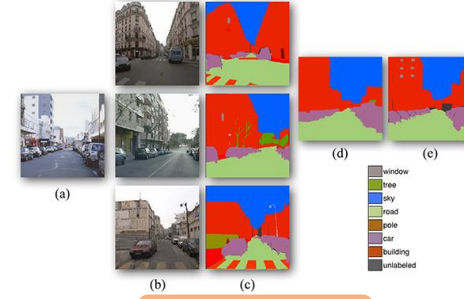
PASCAL PART	Microsoft COCO	PASCAL CONTEXT	Cityscape	ADE20K

Categories: 20
images: 10103
Contributor: R. Mottaghi et al.

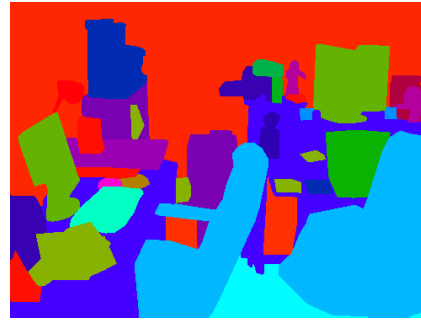# Categories: 91
images: 123387
Contributor: TY. Lin et al.

Categories: 540
images: 10103
Contributor: J. Xiao et al.

Categories: 540
images: 10103
Contributor: J. Xiao et al.

Categories: 2603
images: 22210
Contributor: B. Zhou et al.

Training datasets for RS image semantic segmentation

University of California

#images: 2,100;

#categories: 21(100 images per class)

tile size: 256*256 pixels

Image source: USGS National Map (0.3m)

UC Merced Land-Use dataset

Aerial Image Dataset (AID)

 # images: 210,000 ;

 # categories:30(200-420 images per class)

Multiple spatial resolution: 0.5-8m

Tile size: 600*600 pixels

Image source: GOGLE EARTH (0.2m)

Northwestern Polytechnical University

#images: 31,500 ;

#categories: 45(700 images per class)

Expert classification:(30m-0.2m)

NWPU-RESISC45 dataset

#images: 2,800;

#categories: 7 (400 images per class)

Tile size: 400*400 pixels

Image source: GOGLE EARTH (0.2m)

RSSCN7 dataset

11

1

Data volume are not large enough
- compared to the millions of parameters in a convolutional neural network, the number of images in those datasets are far not enough..
- It is hard to find download link.

2

`Inconsistent time`
- Remote sensing images have various spatial, temporal and spectral resolution, so training dataset should be obtained at the same spatial and temporal region.
- Single data source: Google Earth

So, large volume of high-quality RS training dataset are still needed.

Geospatial Data Cloud (GSCloud)

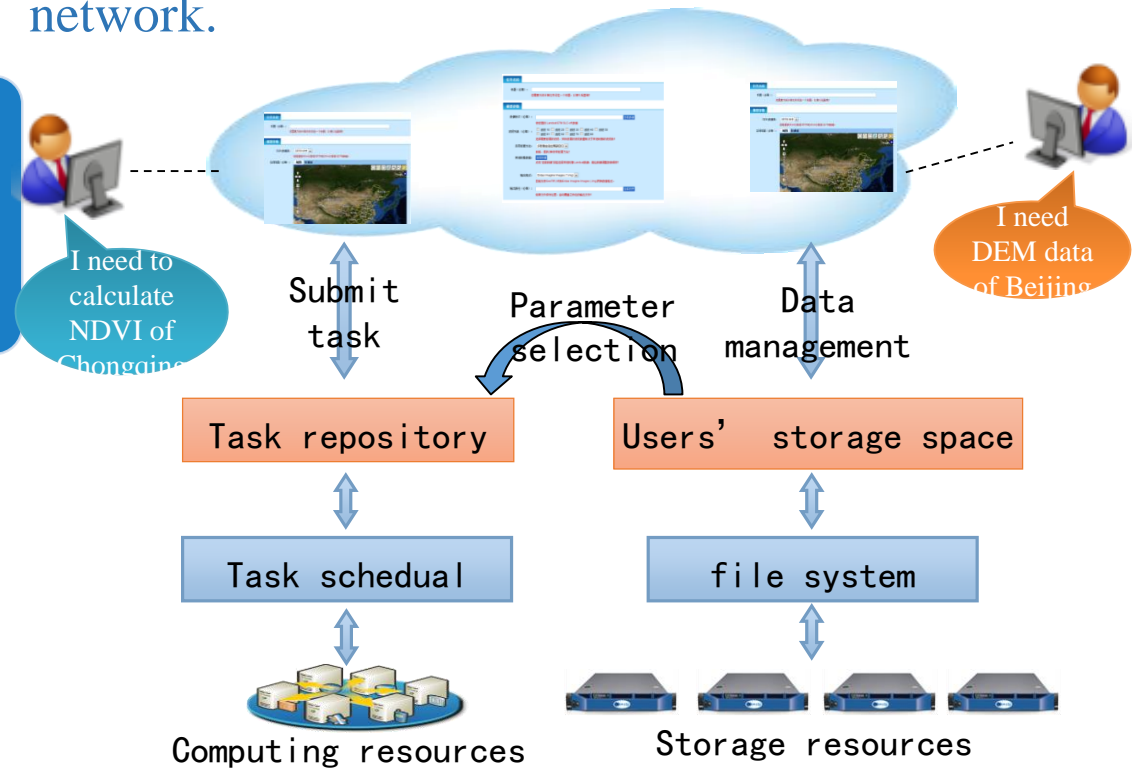**—A platform for massive remote sensing data aggregation, retrieving, downloading and processing.**

➢ Large data resources:

Diverse and huge data sources are available in GSCloud. In fact, there are over 400TB data entities. Datasets including Landsat, MODIS, DEM and so on are all provided for free.
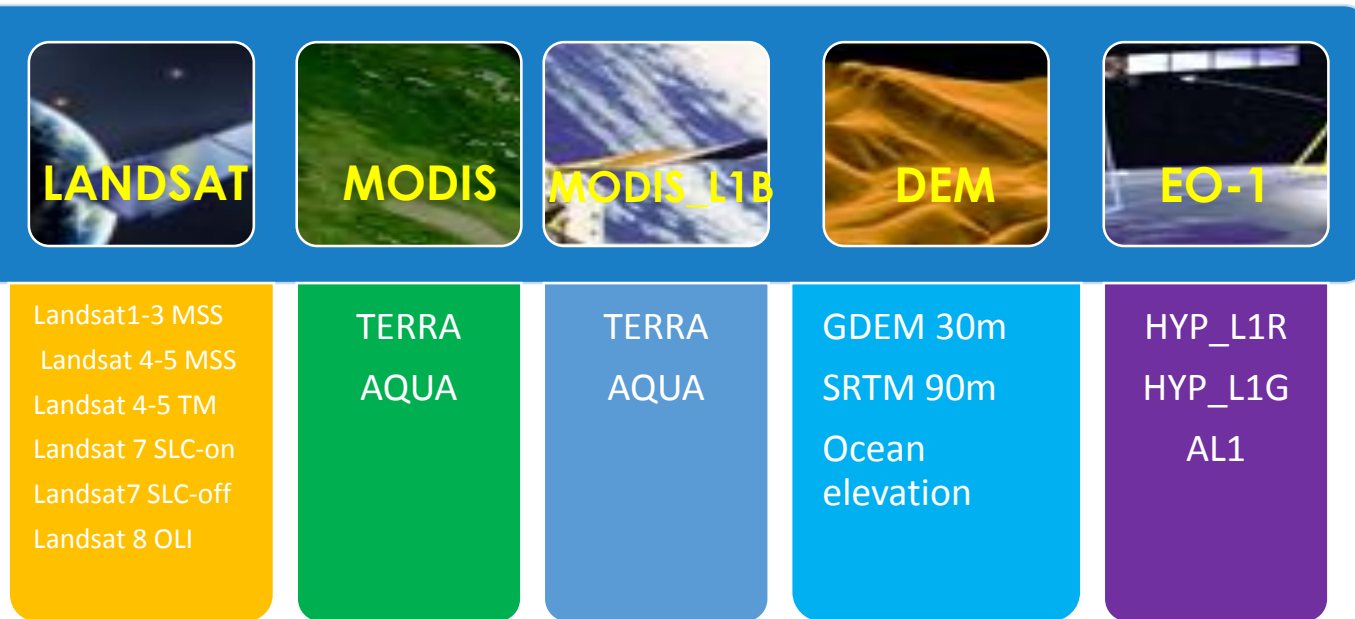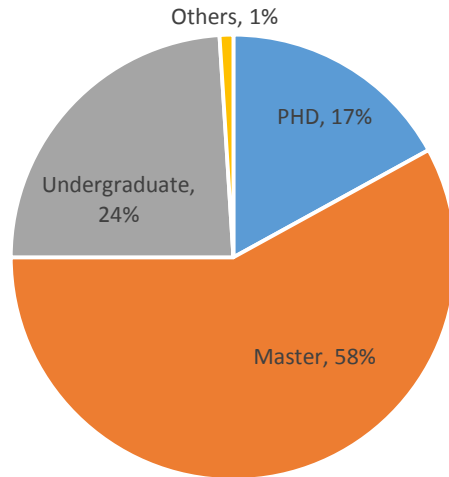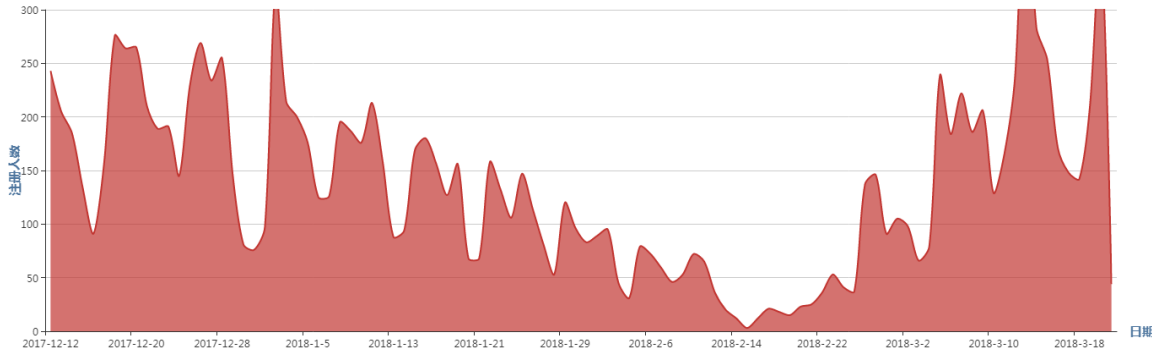
➢ Data processing Services:

To satisfy users' online processing requests, GSCloud provides a simple, fast and interactive online model computing service based on the network.

LANDSAT	MODIS	MODIS_L1B	DEM	EO-1
Landsat1-3 MSS Landsat 4-5 MSS Landsat 4-5 TM Landsat 7 SLC-on Landsat7 SLC-off Landsat 8 OLI	TERRA AQUA	TERRA AQUA	GDEM 30m SRTM 90m Ocean elevation	HYP_L1R HYP_L1G AL1

14

A lot of RS image interpretation requests

Up to now: 203,303 registered users

Users' education distribution

Task id	tasks	Publish time
T2015050001	基于Landsat 5/Landsat7的2000年青藏高原西北部地区湖泊提取	2015年5月13日
T2015050002	基于Landsat 5/Landsat7的2000年青藏高原西南部地区湖泊提取	2015年5月13日
T2015050003	基于Landsat 5/Landsat7的2000年青藏高原东部地区湖泊提取	2015年5月13日
T2015050004	基于Landsat 5的2010年青藏高原西北部地区湖泊提取	2015年5月27日
T2015050005	基于Landsat 5的2010年青藏高原西南部地区湖泊提取	2015年5月27日
T2015050006	基于Landsat 5的2010年青藏高原东部地区湖泊提取	2015年5月27日
T2015050007	基于Landsat 5的2005年青藏高原区域湖泊提取	2015年5月27日
T2015060001	2013年石家庄市土地利用分类	2015年6月30日
T2015070001	基于Landsat的1995年青藏高原西北部地区湖泊提取	2015年7月16日
T2015070002	基于Landsat的1995年青藏高原西南部地区湖泊提取	2015年7月16日
T2015070003	基于Landsat的1995年青藏高原东部地区湖泊提取	2015年7月16日
T2015080001	基于高分辨率影像的河南黄河流域湿地类型提取	2015年8月14日
T2015080002	Landsat影像去云处理	2015年8月20日
T2015100001	基于Landsat8影像的2015年武汉市土地利用类型解译	2015年10月22日
T2015100002	基于Landsat影像的2009年及2015年青海省玛多县湿地和植被覆盖度解译	2015年10月28日
T2015120001	基于Landsat8的2015年青藏高原西北部地区湖泊提取	2015年12月2日
T2015120002	基于Landsat8的2015年青藏高原西南部地区湖泊提取	2015年12月2日
T2015120003	基于Landsat8的2015年青藏高原东部地区湖泊提取	2015年12月2日

15

As crowdsourcing has been proved to be an effective way to tackle simple tasks, plus we have so many qualified users, our platform is designed to be a crowdsourcing platform.
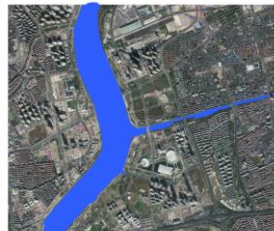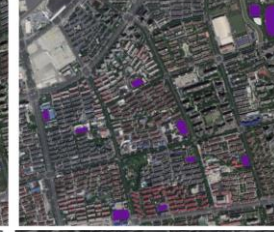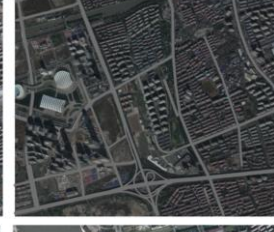
Remote sensing images are displayed on the web, and users can directly label the images by just drawing polygons and add tags.
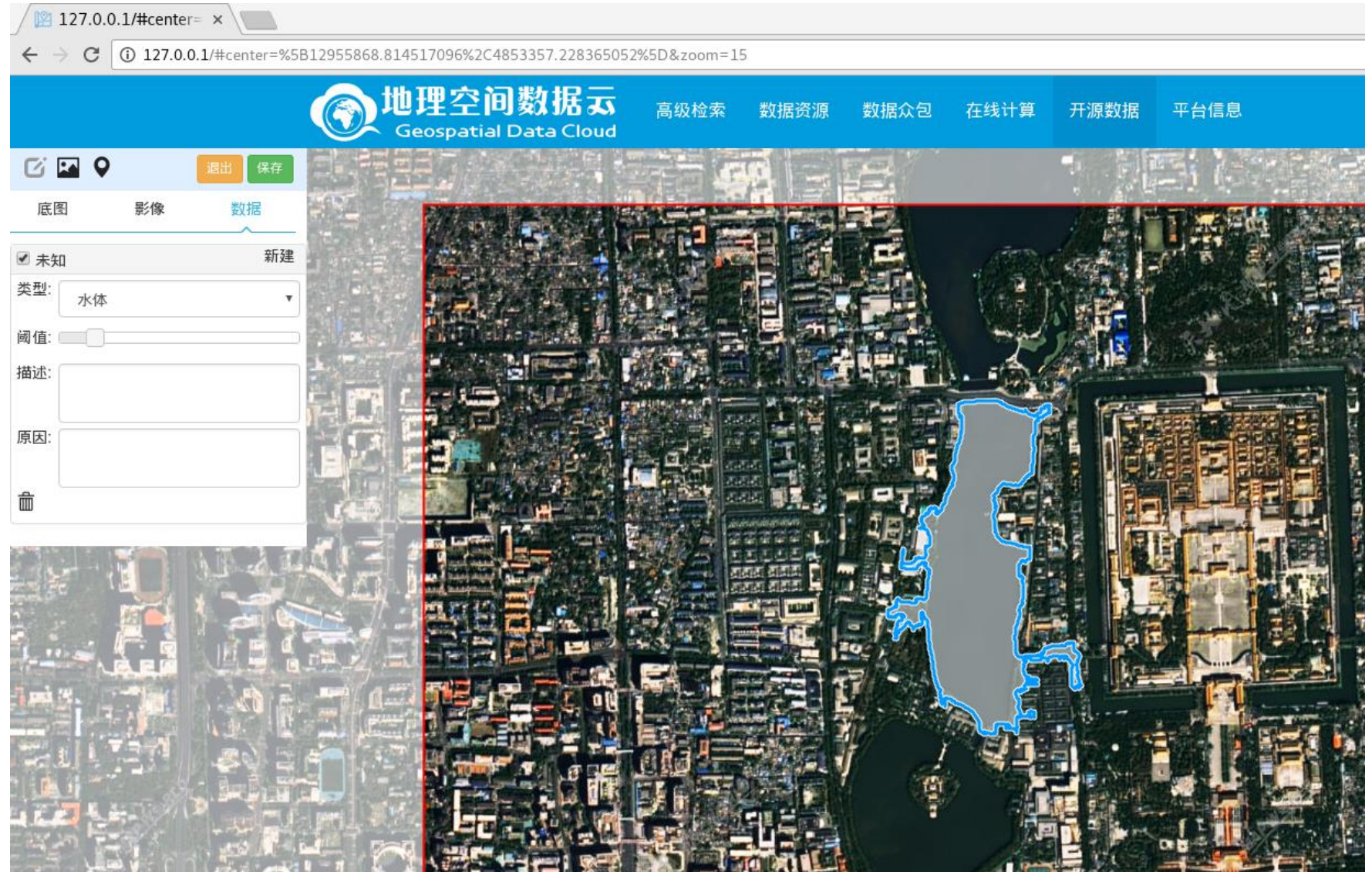
3 Ongoing study--RSLabel

There are a variety of classification standards:

- --Common Six categories: Forest, Grass, Farmland, Bare land, Wet land, water, residential

- --LUCC classification system of Chinese Academy of Sciences

- --classification system adopted by some specific research project

- --National Standards: GB_T21010-2007

- --《土地利用现状分类体系》of 1984

- --Wetland Resource Survey Technical Regulations

- -- special classification requirements

Water	rivers, water body in parks
Vegetation	trees in park, residential region, or along the street
Sports ground	playground of elementary and junior high school, stadium runway
Roads	Urban Roads, Overpasses, Elevated Roads Excluding: Shadows on roads, street trees, building shelter
Construction site	areas around the building, parking lot non-green floor in the landscape
Bare land	large area of open space, non-green vacant land, farmland

➤ Task-targeted

- Tasks are released by Administrator;
- Interpretation standards are given;
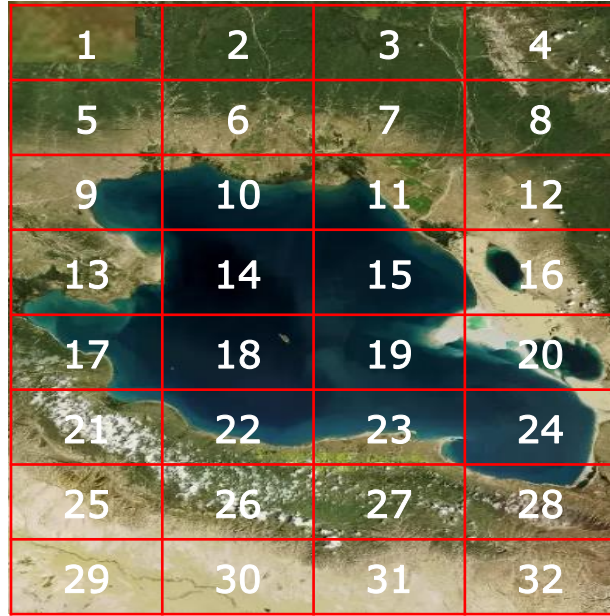
➤ Tile as a unit of editing

- Administrator can control the editable region
- Grid lock：Solve the conflict caused by multiple users editing the same area at the same time
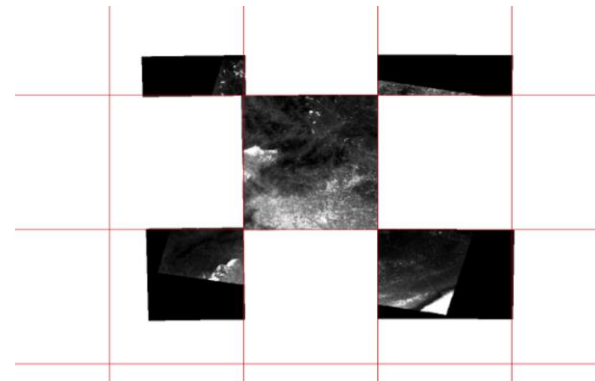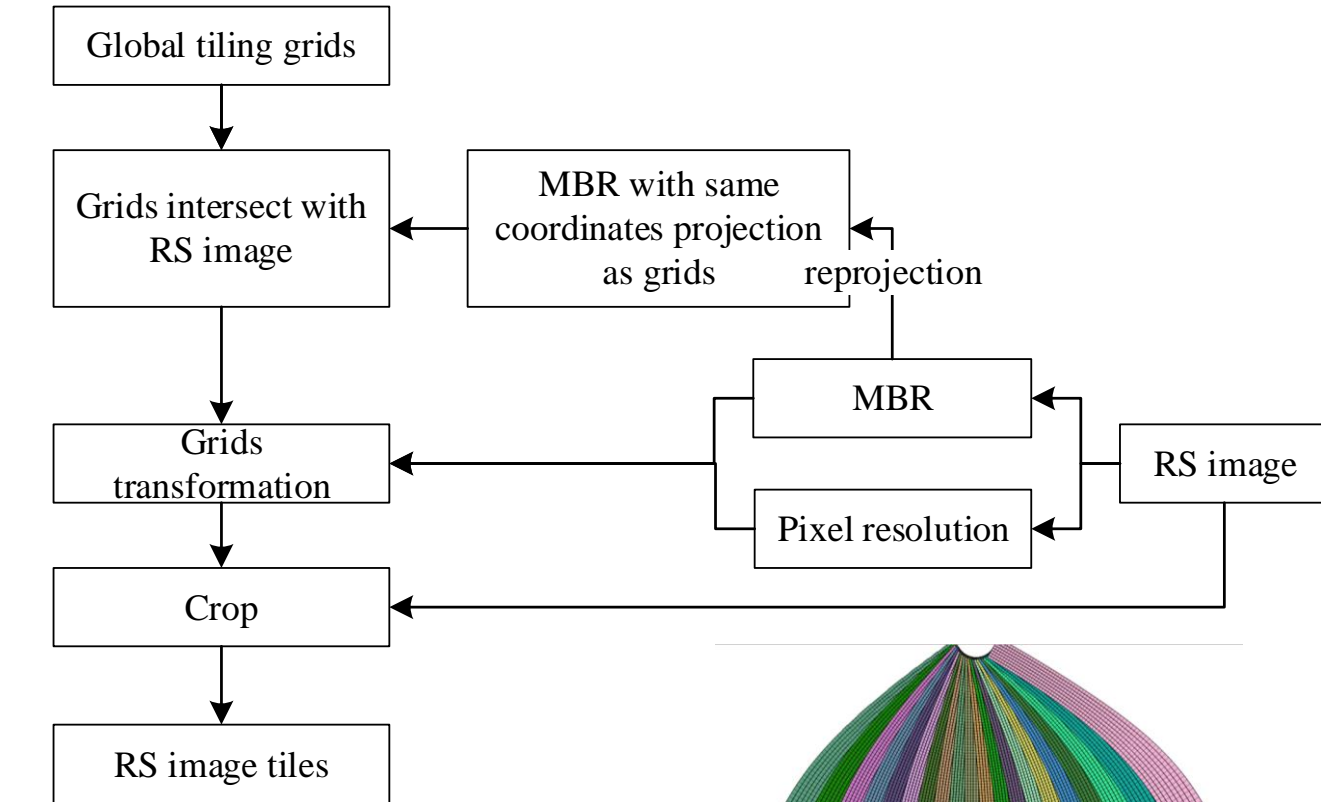
UTM grids from 60E to 180E

labeling process

Select tile

Delineate AOI

Input tag

Real-time rendering in browser

Is the modification completed?

No

Yes

Submit

The server performs the modification

data Storage

Database

Client

Server

19

3 Ongoing study--RSLabel

➢ OpenStreetMap is launched by Steve Coast in 2004, which intend to generate the global vector data through crowdsourcing.

➢ The OpenStreetMap (OSM) has a rich set of vector data provided by volunteers at a global scale.

OSM data problem:
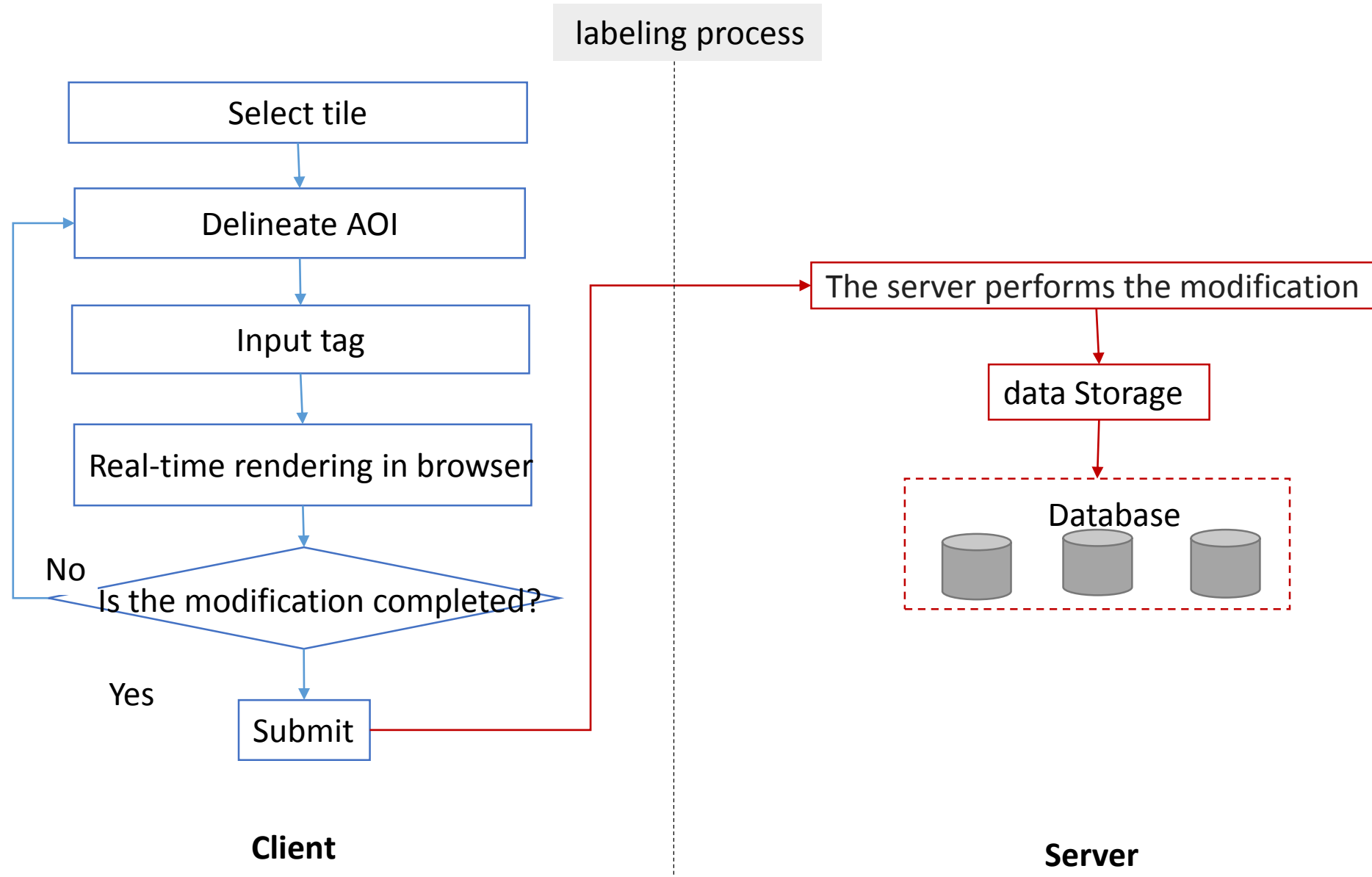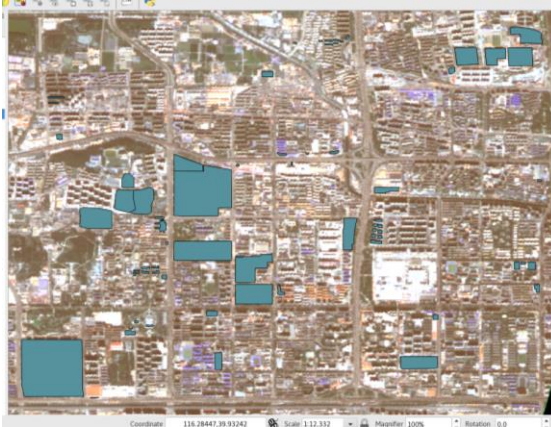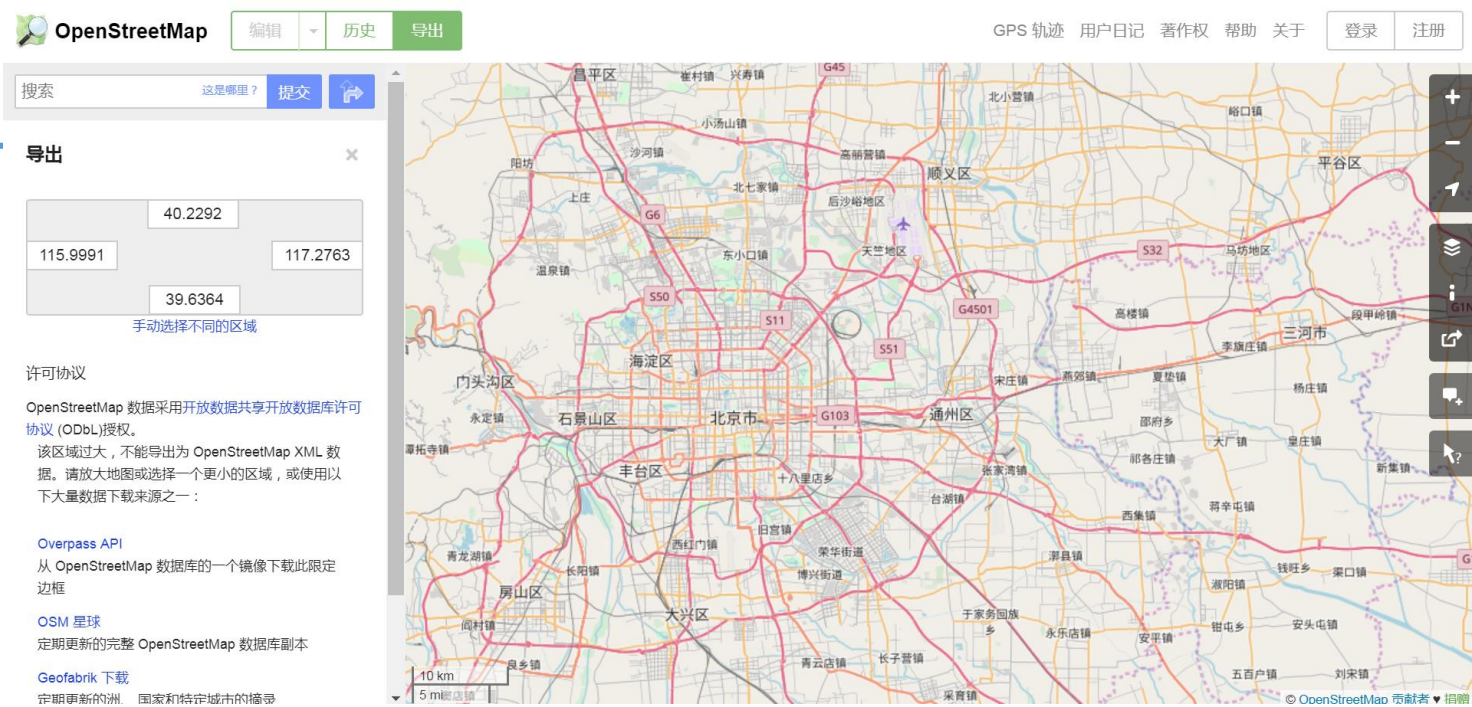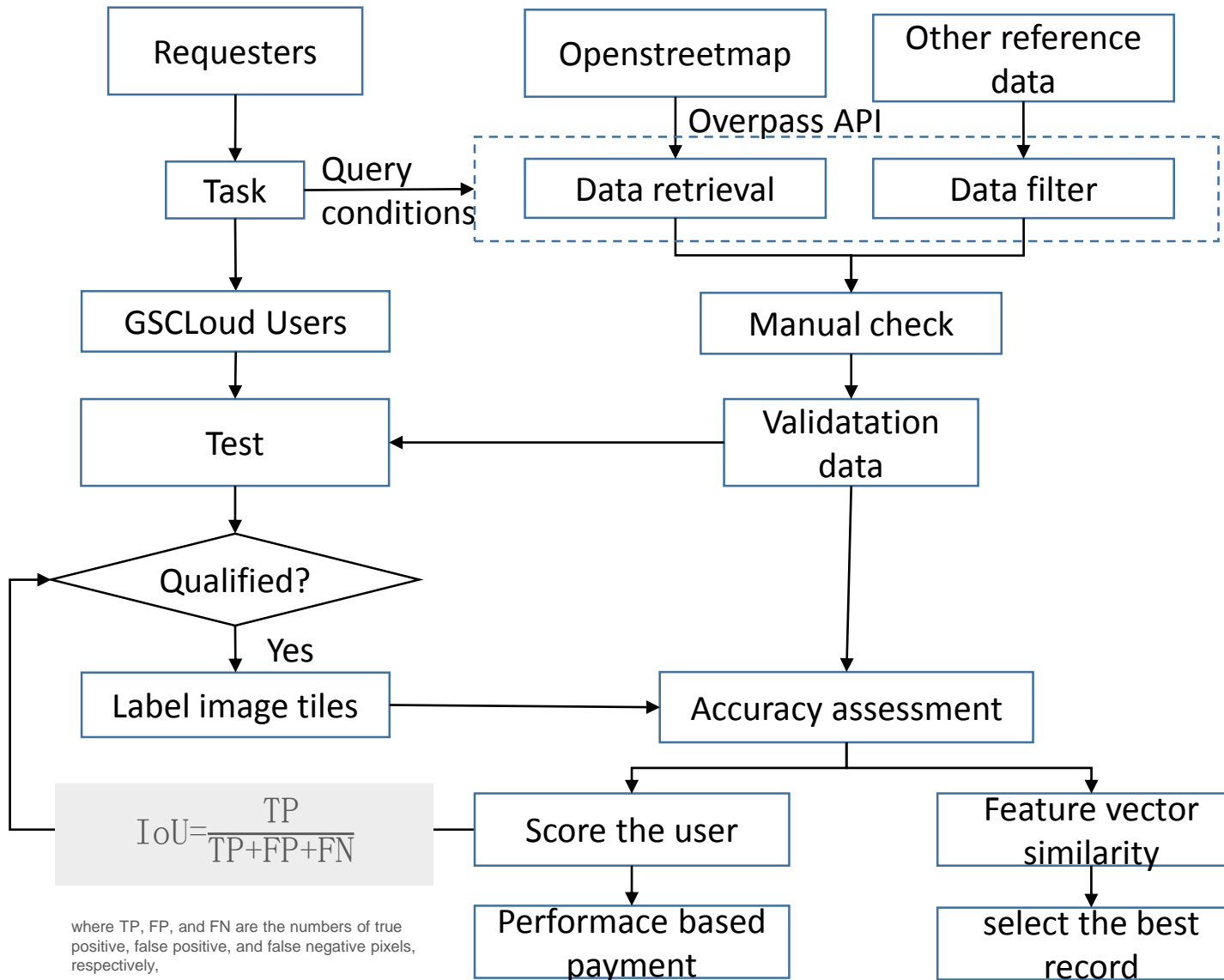1. Inconsistent time between OSM and RS image;
2. Uneven distribution;
3. Label standards are too complex；

➢ Convert the OSM data into LULC map, manual check is still necessary.

1.1 Aerialway
1.2 Aeroway
1.3 Amenity
   1.3.1 Sustenance
   1.3.2 Education
   1.3.3 Transportation
   1.3.4 Financial
   1.3.5 Healthcare
   1.3.6 Entertainment, Arts & Culture
   1.3.7 Others
1.4 Barrier
   1.4.1 Linear barriers
   1.4.2 Access control on highways
1.5 Boundary
   1.5.1 Attributes
1.6 Building
   1.6.1 Accommodation
   1.6.2 Commercial
   1.6.3 Civic/Amenity
   1.6.4 Other Buildings
   1.6.5 Additional Attributes
1.7 Craft
1.8 Emergency
   1.8.1 Medical Rescue
   1.8.2 Firefighters

1.8.3 Lifeguards
1.8.4 Others
1.9 Geological
1.10 Highway
   1.10.1 Roads
   1.10.2 Link roads
   1.10.3 Special road types
   1.10.4 Paths
   1.10.5 Lifecycle
   1.10.6 Attributes
   1.10.7 Other highway features
1.11 Historic
1.12 Landuse
1.13 Leisure
1.14 Man_made
1.15 Military
1.16 Natural
   1.16.1 Vegetation or surface related
   1.16.2 Water related
   1.16.3 Landform related
1.17 Office
1.18 Place
   1.18.1 Administratively declared places
   1.18.2 Populated settlements, urban
   1.18.3 Populated settlements, urban and rural

1.18.4 Other places
1.19 Power
1.20 Public Transport
1.21 Railway
   1.21.1 Tracks
   1.21.2 Additional features
   1.21.3 Stations and Stops
   1.21.4 Other railways
1.22 Route
1.23 Shop
   1.23.1 Food, beverages
   1.23.2 General store, department store, mall
   1.23.3 Clothing, shoes, accessories
   1.23.4 Discount store, charity
   1.23.5 Health and beauty
   1.23.6 Do-it-yourself, household, building materials, gardening
   1.23.7 Furniture and interior
   1.23.8 Electronics
   1.23.9 Outdoors and sport, vehicles
   1.23.10 Art, music, hobbies
   1.23.11 Stationery, gifts, books, newspapers
   1.23.12 Others
1.24 Sport
1.25 Tourism
1.26 Waterway

20

3 Ongoing study--RSLabel

Requesters

Task — Query conditions →

Openstreetmap — Overpass API →

Other reference data

Data retrieval

Data filter

GSCLoud Users

Manual check

Test ← Validatation data

Qualified? — Yes

Label image tiles → Accuracy assessment

$$IoU=\frac{TP}{TP+FP+FN}$$

where TP, FP, and FN are the numbers of true positive, false positive, and false negative pixels, respectively,

Score the user

Performace based payment

Feature vector similarity

select the best record

Six categories/classes have been defined:

1. Impervious surfaces (RGB: 255, 255, 255)
2. Building (RGB: 0, 0, 255)
3. Low vegetation (RGB: 0, 255, 255)
4. Tree (RGB: 0, 255, 0)
5. Car (RGB: 255, 255, 0)
6. Clutter/background (RGB: 255, 0, 0)

21

Thank You!

Any proposals about GSCloud are welcome. Please contact

Email: data@cnic.cn Phone: 010-58812534

Any questions about this report. Please contact zjh@cnic.cn

our website: http://www.gscloud.cn