CAPTAIN
COMPUTATIONAL AND PHOTOGRAMMETRIC VISION

# RS IMAGE INTERPRETATION FROM DATA PERSPECTIVE

# Diversity, Richness, Scalability (DiRS) :

## *On Benchmarking Remote Sensing Image Interpretation*


**Gui-Song Xia**

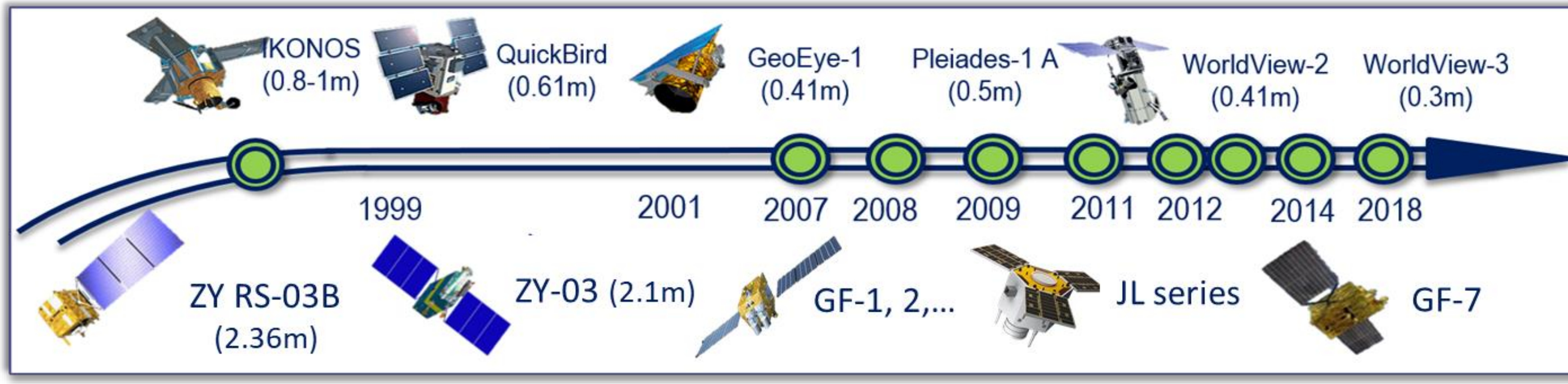
guisong.xia@whu.edu.cn

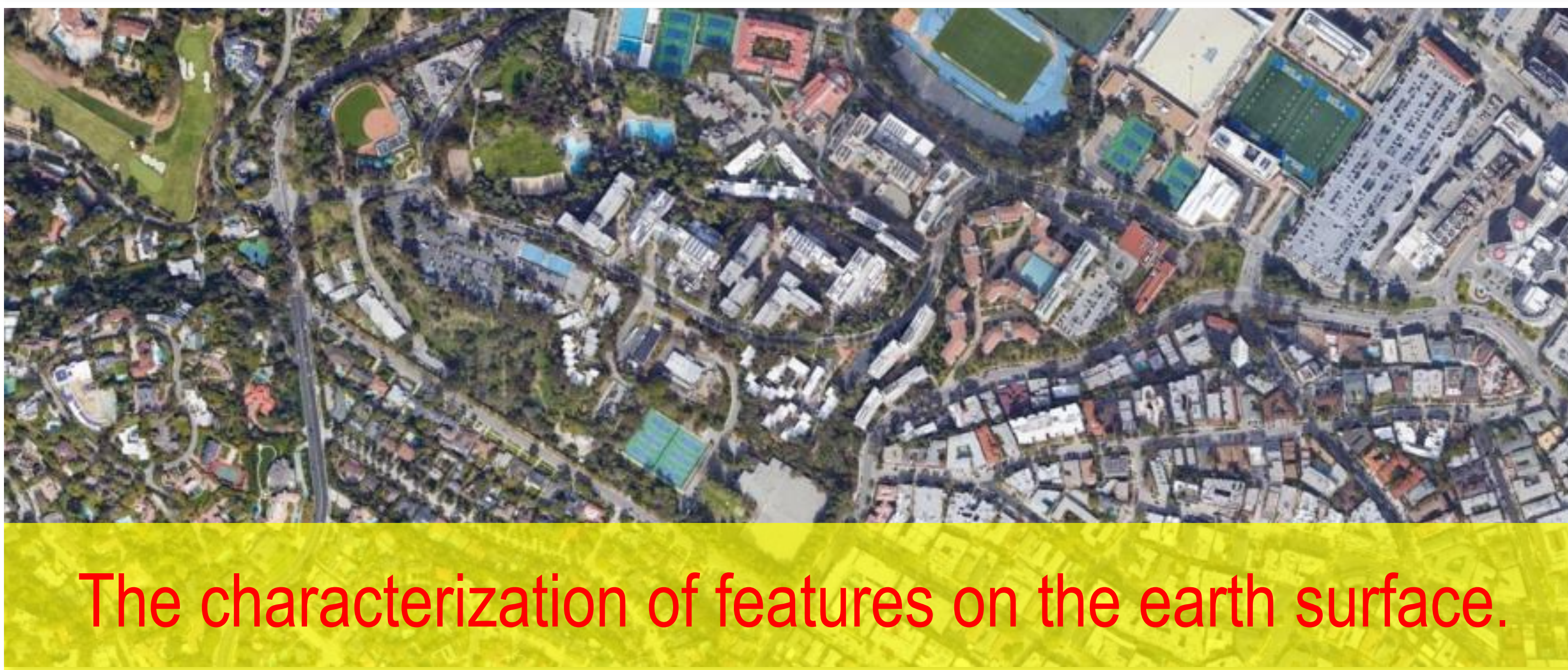
**School of Computer Science, Wuhan University**
**Institute of Artificial Intelligence, Wuhan University**
**State Key Lab. LIESMARS, Wuhan University**


**Jun. 1, 2020**

# Advanced RS Technology

RS technology has significantly improved the earth observation ability.



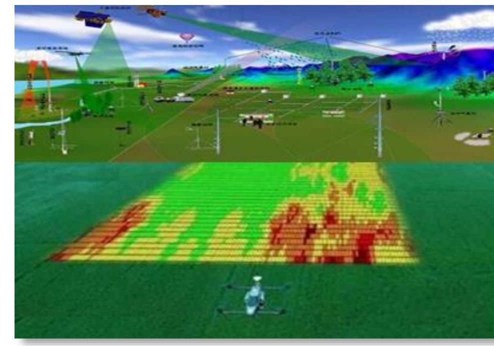
The characterization of features on the earth surface.

# Applications of RS Images

**Interpretation of RS images plays important roles in many real-world applications.**

National security

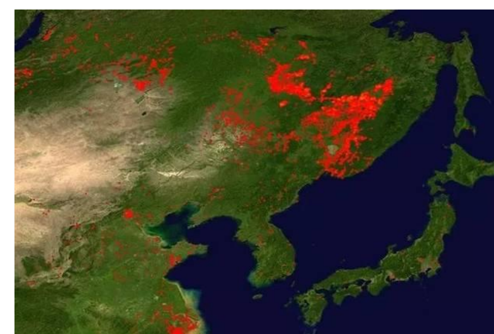
High definition map



Precision agriculture

Smart city

Disaster assessment



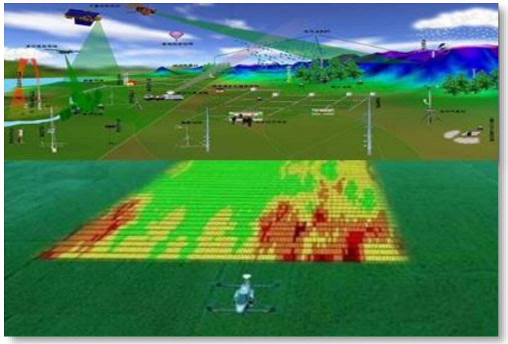
Environ. monitoring

# Applications of RS Images

**Interpretation of RS images plays important roles in many real-world applications.**

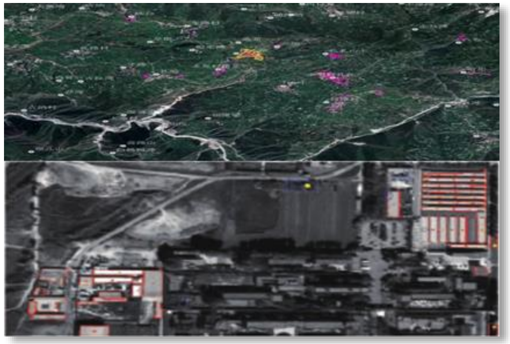
National security

High definition map

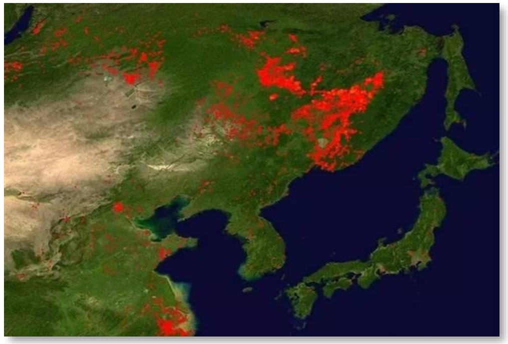


Precision agriculture

Smart city



Disaster assessment



Environ. monitoring

# Interpretation of RS Images

**Current situation:** **Increasing demands for automatic interpretation**

Satellites on-orbit

- **Variation:** difference in spectral, spatial, and temporal properties

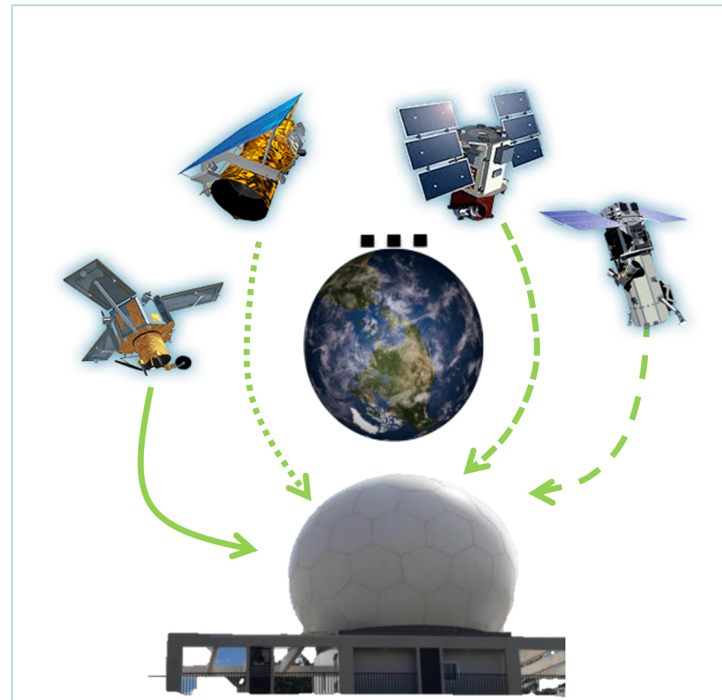
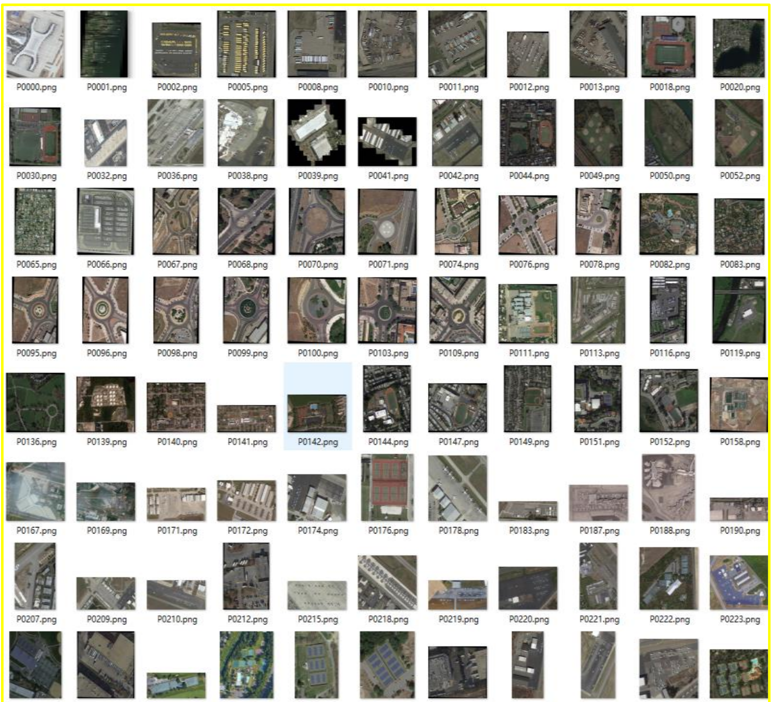


Image acquisition

- **Inconsistency:** multi-modal, multi-source RS images

# Interpretation of RS Images

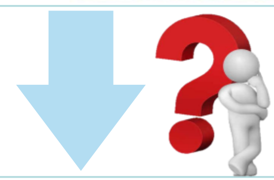
## **Current situation:** Increasing demands for automatic interpretation



Large volume of images

- **Challenge:** geometrical shapes, textural attribute, structural characteristic …

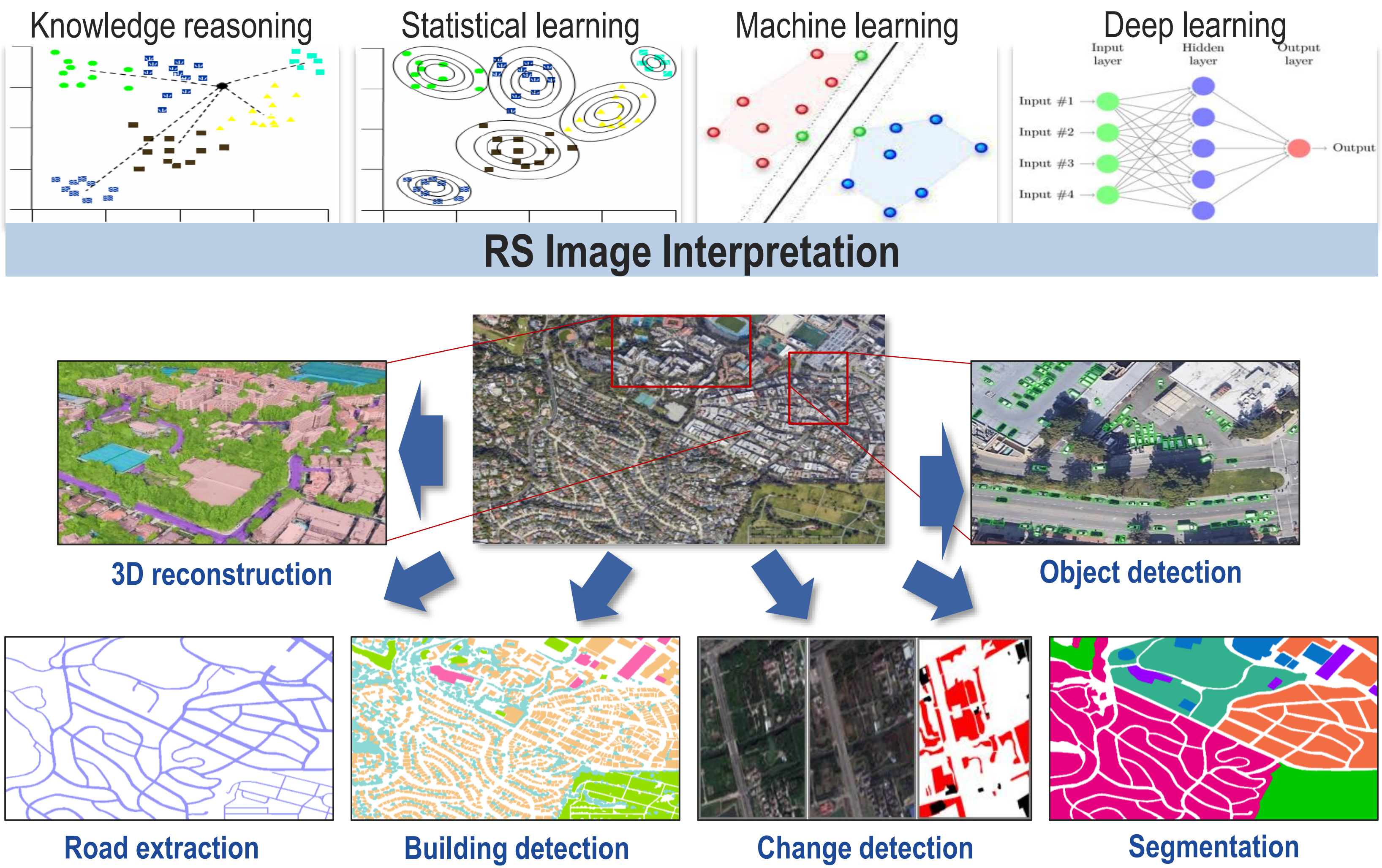
Manual interpretation

*Automatic interpretation*

# Blooming Data-driven methods

Content interpretation: data-driven methods for RS image interpretation.



# Motivation

- **Huge-volume RS images** ***v.s. limited data with labels***
- **Increasing number of datasets with** ***different purposes and standards***

  - The **ever-growing volume of RS images** is acquired while very **few of them are annotated** with valuable information.
  - The **generalization ability of algorithms** for interpreting RS images is of great urgency to be enhanced.
  - The **Representative and large-scale RS image datasets** with accurate annotations is demanded to narrow *the gap between algorithm development and real applications*.
  - There is a **lack of public platforms** for systematic evaluation and fair comparison among different algorithms.

# Outline

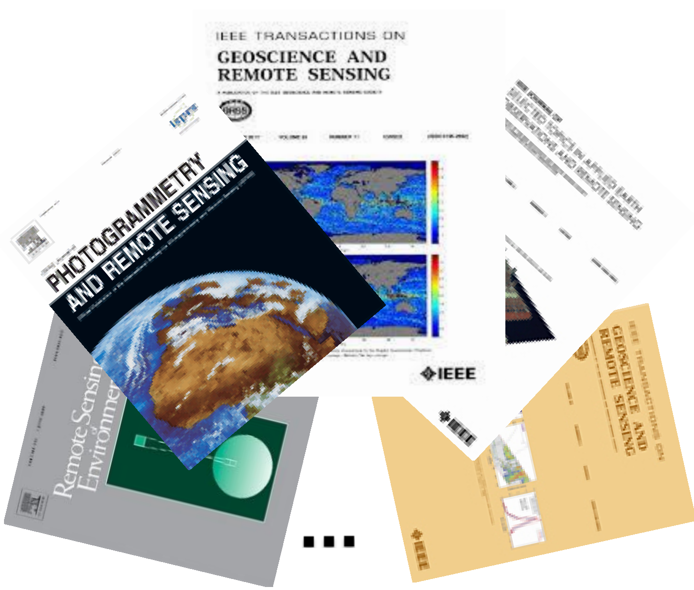
- Background
- **Research Focus in the Past Decade**
- Principles to Benchmark RS Image Interpretation
- An Example： Million-AID
- Challenges and Perspectives
- Conclusions

# Focus in the Past Decade

CAPTAIN
COMPUTATIONAL AND PHOTOGRAMMETRIC VISION

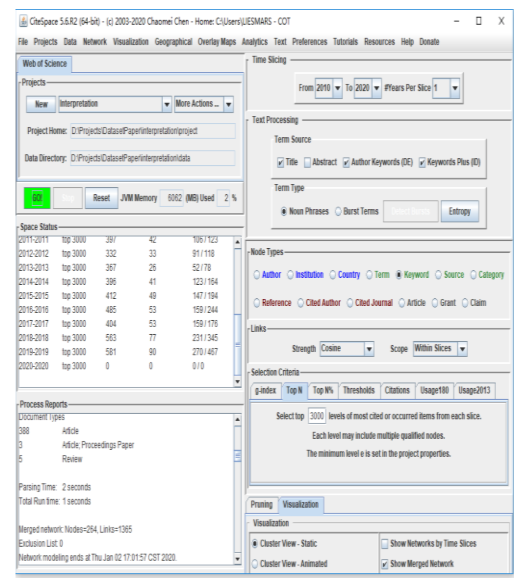
## ■ A systematic investigation to the literature

- **Journals with good reputation:** *ISPRS J. P&RS, RSE, TGRS …*
- **Meta-data for analysis:** 11, 337 retrieved articles over the past decade
- **Bibliometric analysis:** title/topic/keywords … concerning image interpretation




Selected journals

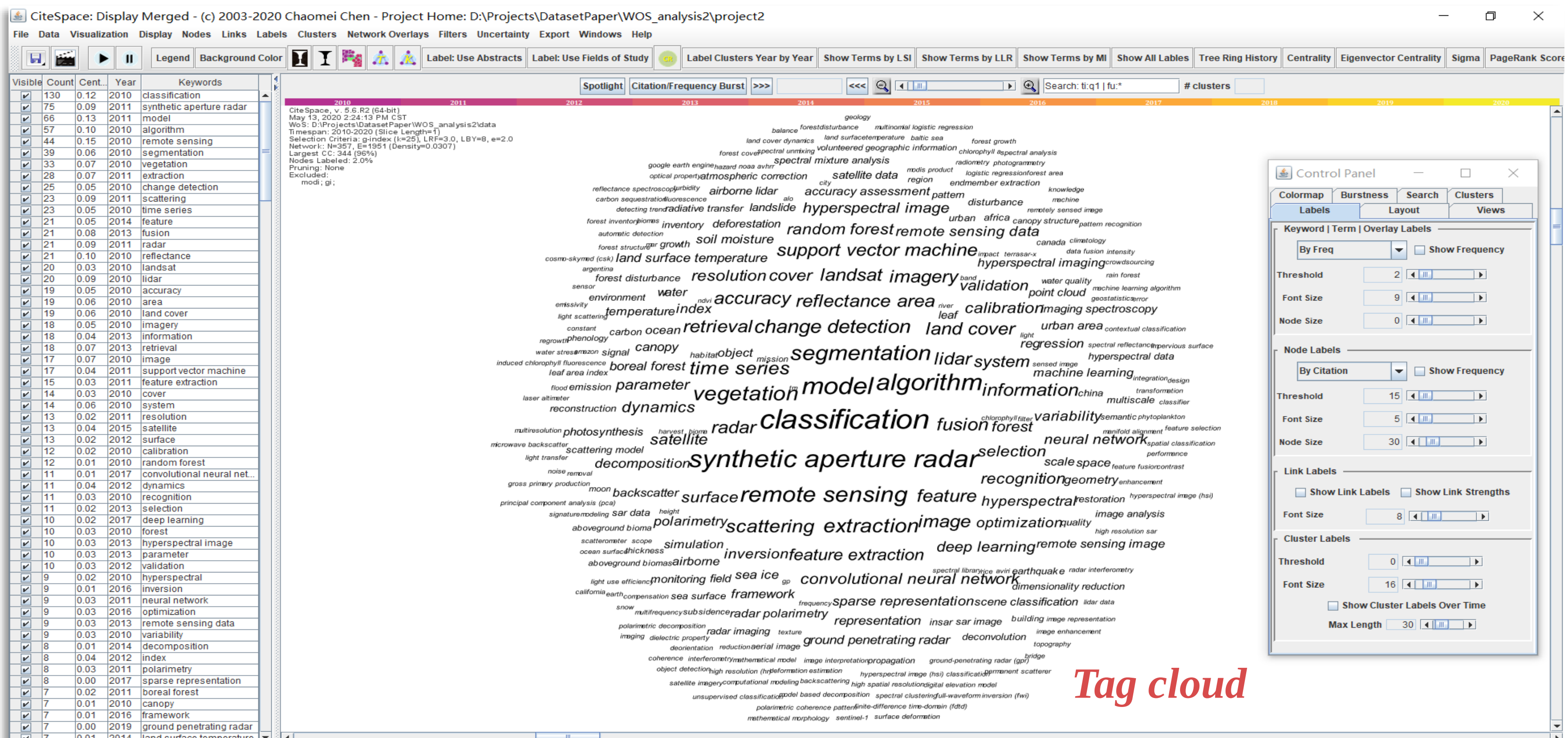
Meta-data



Bibliometric analysis

# Frequency Terms

- Interpretation mainly focus on ***classification*** tasks (*scene*, *object*, *land cover* …)
- ***Segmentation*** and ***change detection*** occupy prominent positions
- ***Algorithm*** and ***model*** play significant roles in RS image interpretation



# Citation Bursts

- ***Deep learning/CNN*** climb to the top for RS image interpretation since 2017
- ***Data Set*** confirms its key role in data-driven interpretation schemes since 2019

***Top 5 Keywords with the Strongest Citation Bursts***

| Year | Keywords | Strength | Duration | Year | Keywords | Strength | Duration | Year | Keywords | Strength | Duration |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2011 | atmospheric correction | 2.27 | | 2014 | surface | 2.08 | | 2017 | convolutional neural network | 3.19 | |
| | endmember extraction | 1.86 | | | scattering model | 1.95 | | | deep learning | 2.90 | |
| | segmentation | 1.63 | | | landslide | 1.62 | | | sparse representation | 2.31 | |
| | object detection | 1.24 | | | extraction | 1.59 | | | field | 2.27 | |
| | reflectance spectroscopy | 1.24 | | | reflectance | 1.23 | | | radiative transfer | 1.81 | |
| 2012 | hyperspectral data | 2.24 | | 2015 | time series | 2.17 | | 2018 | regression | 2.70 | |
| | signal | 1.87 | | | hyperspectral imaging | 2.10 | | | random forest | 2.45 | |
| | backsactter | 1.41 | | | remote sensing | 2.00 | | | dimentionality reduction | 2.20 | |
| | model | 1.37 | | | subsidence | 1.74 | | | reconstruction | 2.20 | |
| | optical property | 1.28 | | | water quality | 1.74 | | | machine learning | 1.91 | |
| 2013 | object | 2.45 | | 2016 | inversion | 2.96 | | 2019 | ground penetrating radar | 3.72 | |
| | urban area | 2.45 | | | simulation | 2.84 | | | water | 3.19 | |
| | monitoring | 2.04 | | | information | 2.44 | | | emission | 2.12 | |
| | hyperspectral image | 1.85 | | | point cloud | 2.37 | | | temperature | 1.79 | |
| | algorithm | 1.68 | | | restoration | 2.28 | | | propagation | 1.59 | |

# Available Datasets for Interpretation

CAPTAIN
COMPUTATIONAL AND PHOTOGRAMMETRIC VISION

- RS *scene classification* datasets

| Dataset | #Cat. | #Images per cat. | #Images | Resolution (m) | Image size | Year |
|---|---|---|---|---|---|---|
| UC-Merced | 21 | 100 | 2,100 | 0.3 | 256 × 256 | 2010 |
| WHU-RS19 | 19 | 50 to 61 | 1,013 | up to 0.5 | 600 × 600 | 2012 |
| RSSCN7 | 7 | 400 | 2,800 | – | 400 × 400 | 2015 |
| SAT-4 | 4 | 89,963 to 178,034 | 500,000 | 1 to 6 | 28 × 28 | 2015 |
| SAT-6 | 6 | 10,262 to 150,400 | 405,000 | 1 to 6 | 28 × 28 | 2015 |
| BCS | 2 | 1,438 | 2,876 | – | 600 × 600 | 2015 |
| RSC11 | 11 | ∼100 | 1,232 | ∼0.2 | 512 × 512 | 2016 |
| SIRI-WHU | 12 | 200 | 2,400 | 2 | 200 × 200 | 2016 |
| NWPU-RESISC45 | 45 | 700 | 31,500 | 0.2 to 30 | 256 × 256 | 2016 |
| AID | 30 | 220 to 420 | 10,000 | 0.5 to 8 | 600 × 600 | 2017 |
| RSI-CB256 | 35 | 198 to 1,331 | 24,000 | 0.3 to 3 | 256 × 256 | 2017 |
| RSI-CB128 | 45 | 173 to 1,550 | 36,000 | 0.3 to 3 | 128 × 128 | 2017 |
| RSD46-WHU | 46 | 500 to 3,000 | 117,000 | 0.5 to 2 | 256 × 256 | 2017 |
| EuroSAT | 10 | 2,000 to 3,000 | 27,000 | 10 | 64 × 64 | 2018 |
| PatternNet | 38 | 800 | 30,400 | 0.06 to 4.7 | 256 × 256 | 2018 |

- RS *object detection* datasets

| Datasets | Annot. | #Cat. | #Instances | #Images | Image width | Year |
|---|---|---|---|---|---|---|
| TAS | HBB | 1 | 1,319 | 30 | 792 | 2008 |
| SZTAKI-INRIA | OBB | 1 | 665 | 9 | ∼800 | 2012 |
| NWPU-VHR10 | HBB | 10 | 3,651 | 800 | ∼1,000 | 2014 |
| DLR 3k | OBB | 2 | 14,235 | 20 | 5,616 | 2015 |
| UCAS-AOD | OBB | 2 | 14,596 | 1,510 | ∼1,000 | 2015 |
| VEDAI | OBB | 9 | 3,640 | 1,210 | 512/1,024 | 2016 |
| COWC | CP | 1 | 32,716 | 53 | 2,000–19,000 | 2016 |
| HRSC2016 | OBB | 26 | 2,976 | 1,061 | ∼1,100 | 2016 |
| RSOD | HBB | 4 | 6,950 | 976 | ∼1,000 | 2017 |
| CARPPK | HBB | 1 | 89,777 | 1,448 | 1280 | 2017 |
| LEVIR | HBB | 3 | 11,028 | 22,000 | 800 | 2018 |
| VisDrone | HBB | 10 | 54,200 | 10,209 | 2,000 | 2018 |
| xView | HBB | 60 | 1,000,000 | 1,413 | ∼3,000 | 2018 |
| DOTA-v1.0 | OBB | 15 | 188,282 | 2,806 | 800–4,000 | 2018 |
| HRRSD | HBB | 13 | 55,740 | 21,761 | 152–10,569 | 2019 |
| DIOR | HBB | 20 | 192,472 | 23,463 | 800 | 2019 |
| DOTA-v1.5 | OBB | 16 | 402,089 | 2,806 | 800–13,000 | 2019 |
| DOTA-v2.0 | OBB | 18 | 1,488,666 | 11,067 | 800–20,000 | 2020 |

# Available Datasets for Interpretation

CAPTAIN
COMPUTATIONAL AND PHOTOGRAMMETRIC VISION

- RS *semantic segmentation* datasets

| Datasets | #Cat. | #Images | Resolution (m) | #Bands | Image size | Year |
|---|---|---|---|---|---|---|
| Kennedy Space Center | 13 | 1 | 18 | 224 bands | 512×614 | 2005 |
| Botswana | 14 | 1 | 30 | 242 bands | 1476×256 | 2005 |
| Salinas | 16 | 1 | 3.7 | 224 bands | 512×217 | – |
| University of Pavia | 9 | 1 | 1.3 | 115 bands | 610×340 | – |
| Pavia Centre | 9 | 1 | 1.3 | 115 bands | 1096×492 | – |
| ISPRS Vaihingen | 6 | 33 | 0.09 | IR,R,G,DSM,nDSM | ∼2,500×2,500 | 2012 |
| ISPRS Potsdam | 6 | 38 | 0.05 | IR,RGB,DSM,nDSM | 6,000×6,000 | 2012 |
| Massachusetts Buildings | 2 | 151 | 1 | RGB | 1,500×1,500 | 2013 |
| Massachusetts Roads | 2 | 1,171 | 1 | RGB | 1,500×1,500 | 2013 |
| Indian Pines | 16 | 1 | 20 | 224 bands | 145×145 | 2015 |
| Zurich Summer | 8 | 20 | 0.62 | NIR, RGB | 1,000×1,150 | 2015 |
| Inria Dataset | 2 | 360 | 0.3 | RGB | 1,500×1,500 | 2017 |
| EVlab-SS | 10 | 60 | 0.1 to 2 | RGB | 4,500×4,500 | 2017 |
| RIT-18 | 18 | 3 | 0.047 | 6 bands | 9,000×6,000 | 2017 |
| WHU Building-Aerial Imagery | 2 | 8,189 | 0.3 | RGB | 512×512 | 2019 |
| WHU Building-Satellite Imagery I | 2 | 204 | 0.3 to 2.5 | RGB | 512×512 | 2019 |
| WHU Building-Satellite Imagery II | 2 | 17,388 | 2.7 | RGB | 512×512 | 2019 |
| So2Sat LCZ42 | 17 | 400,673 | 10 | 10 bands | 32×32 | 2019 |
| SEN12MS | 33 | 180,662 triplets | 10 to 50 | up to 13 bands | 256×256 | 2019 |
| UAVid | 8 | 420 | – | RGB | ∼4,000×2,160 | 2020 |
| GID | 15 | 150 | 0.8 to 10 | 4 bands | 6,800×7,200 | 2020 |

- RS *change detection* datasets

| Datasets | #Cat. | #Image pairs | Resolution (m) | #Bands | Image size | Year |
|---|---|---|---|---|---|---|
| SZTAKI AirChange | 2 | 13 | 1.5 | RGB | 952×640 | 2009 |
| AICD | 2 | 1000 | 0.5 | 115 bands | 800×600 | 2011 |
| Taizhou Data | 4 | 1 | 30 | 6 bands | 400×400 | 2014 |
| Kunshan Data | 3 | 1 | 30 | 6 bands | 800×800 | 2014 |
| Yancheng | 4 | 2 | 30 | 242 bands | 400×145 | 2018 |
| Urban-rural boundary of Wuhan | 20 | 1 | 4/30 | 4/9 bands | 960×960 | 2018 |
| Hermiston City area, Oregon | 5 | 1 | 30 | 242 bands | 390×200 | 2018 |
| OSCD | 2 | 24 | 10 | 13 bands | 600×600 | 2018 |
| Quasi-urban areas | 3 | 1 | 0.5 | 8 bands | 1,200×1,200 | 2018 |
| WHU Building-Change Detection | 2 | 1 | 0.2 | RGB | 32,207×15,354 | 2018 |
| Season-varing Dataset | 2 | 16,000 | 0.03 to 0.1 | RGB | 256×256 | 2018 |
| ABCD | 2 | 4,253 | 0.4 | RGB | 160×160 | 2018 |
| HRSCD | 6 | 291 | 0.5 | RGB | 10,000×10,000 | 2019 |
| MtS-WH | 9 | 1 | 1 | NIR, RGB | 7,200×6,000 | 2019 |
| LEVIR-CD | 2 | 637 | 0.5 | RGB | 1,024×1,024 | 2020 |
| SCDN | 30 | 4,214 | 0.5 to 3 | RGB | 512×512 | 2020 |

# Some Critical Reviews

## Categories involved in interpretation

- ***Small number*** of categories, content interpretation for ***certain objects***
- Categories with ***equal relationship***, chaotic management for semantic information
- Complex semantic categories and relationships in real applications, e.g., LULC

## Dataset annotation

- Nearly all ***manually annotated*** by experts, extensive ***labor remains to relieve***
- ***Visualization for large scale, high spectral RS images*** annotation is demanded
- ***Lack of interchange*** with application departments for efficient data annotation


UC Merced, 21 classes, 2100 images


NPWU VHR, 800 images, manually annotated

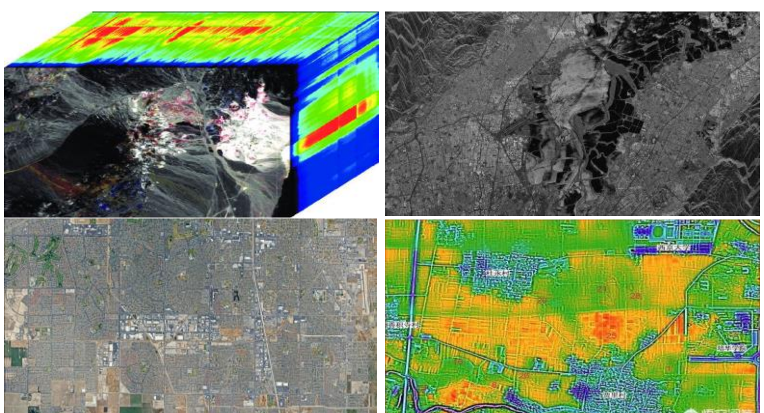
# Some Critical Reviews

## Image source

- Optical images (Google Earth) as data standard since spatial pattern, visual texture, structural information are more concerned (e.g., for scene/object recognition)
- High spectral, SAR images for abnormal object detection by the physical property

## Dataset scale

- ***Limited number***, ***size of chipped images***, ***performance saturation*** of algorithms
- ***Lack of image variation***, ***sample diversity***, and ***content representation***, causing ***weak generalization ability*** of interpretation algorithms



Multi-modal image source

Simple scenes and complex reality

# Outline

- Background
- Research Focus in the Past Decade
- **Principles to Benchmark RS Image Interpretation**
- An Example： Million-AID
- Challenges and Perspectives
- Conclusions

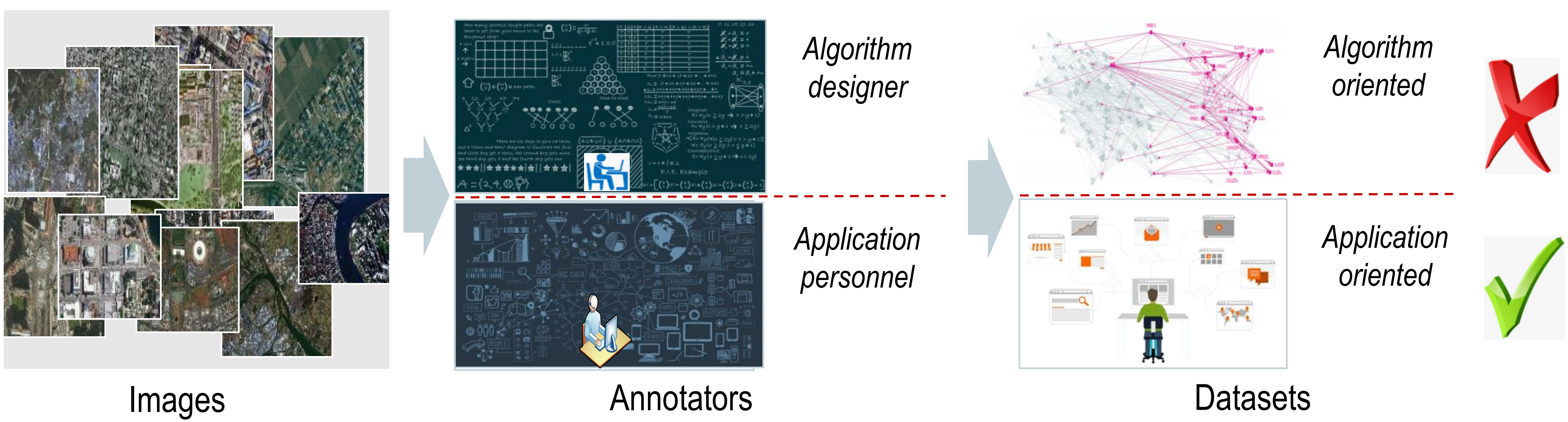
# Principles to Benchmark RS II

## Toward real-world scenarios ***rather than specific algorithms***

- Model training, testing, and screening for practical applications
- Rich samples of variation in background, scale, imaging conditions, …

## Annotation by application sides ***rather than algorithm developers***

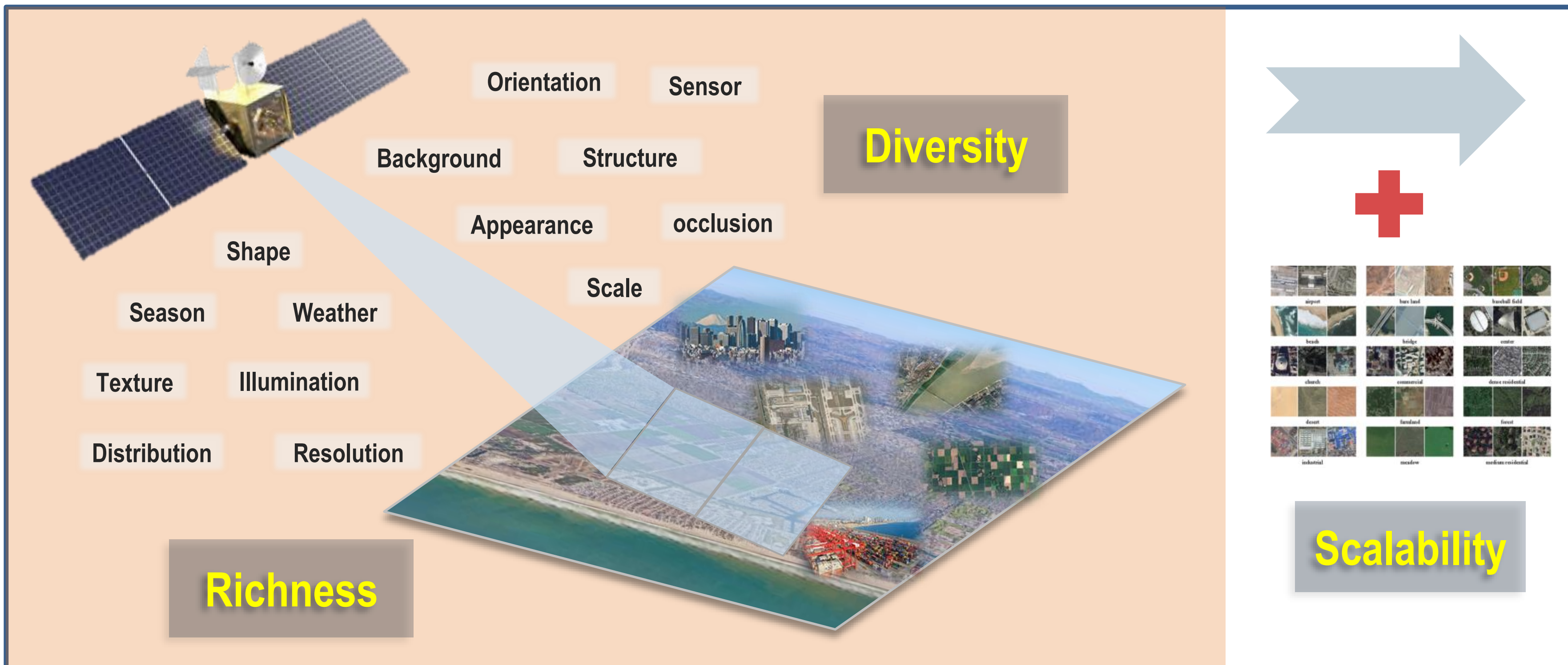
- Label images and samples considering practical challenges in application



# DiRS: Principles to Benchmark RS II

## DiRS for dataset construction

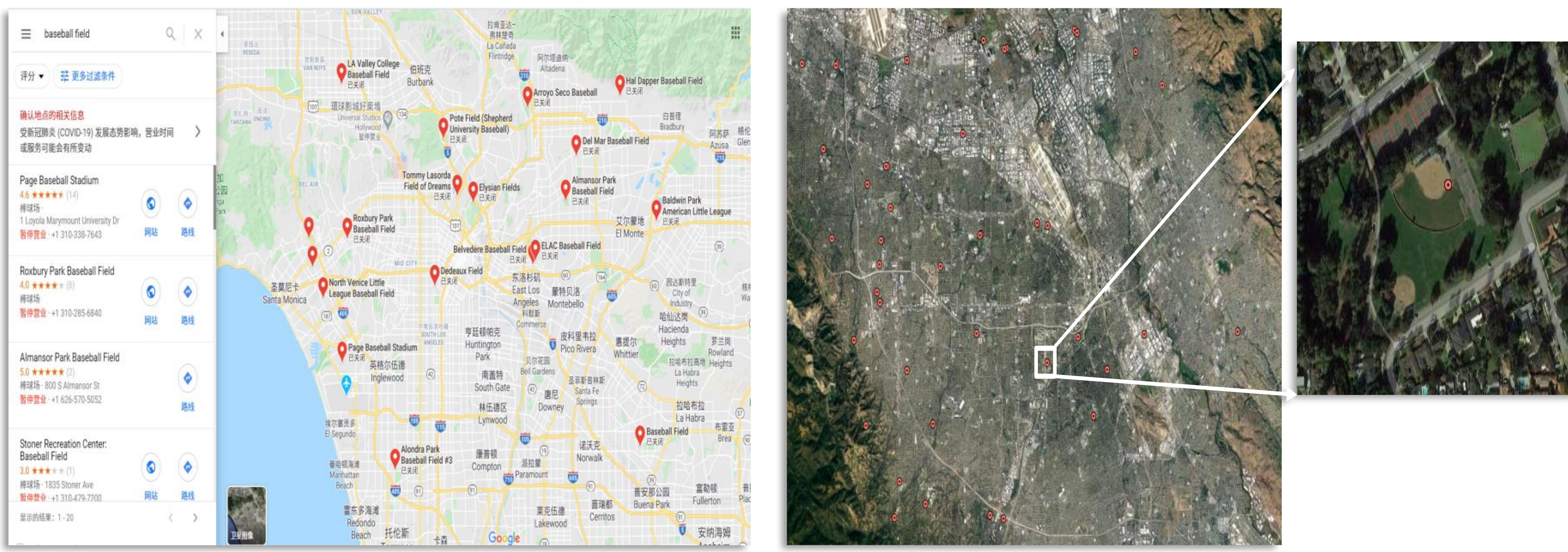
- ***Diversity***: between-/within-class diversity, complementarity of features
- ***Richness***: large-scale images, sufficient samples, diverse characteristics …
- ***Scalability***: sufficient space for new data involvement, sustainable availability



# Geographic Information Integration

CAPTAIN
COMPUTATIONAL AND PHOTOGRAMMETRIC VISION

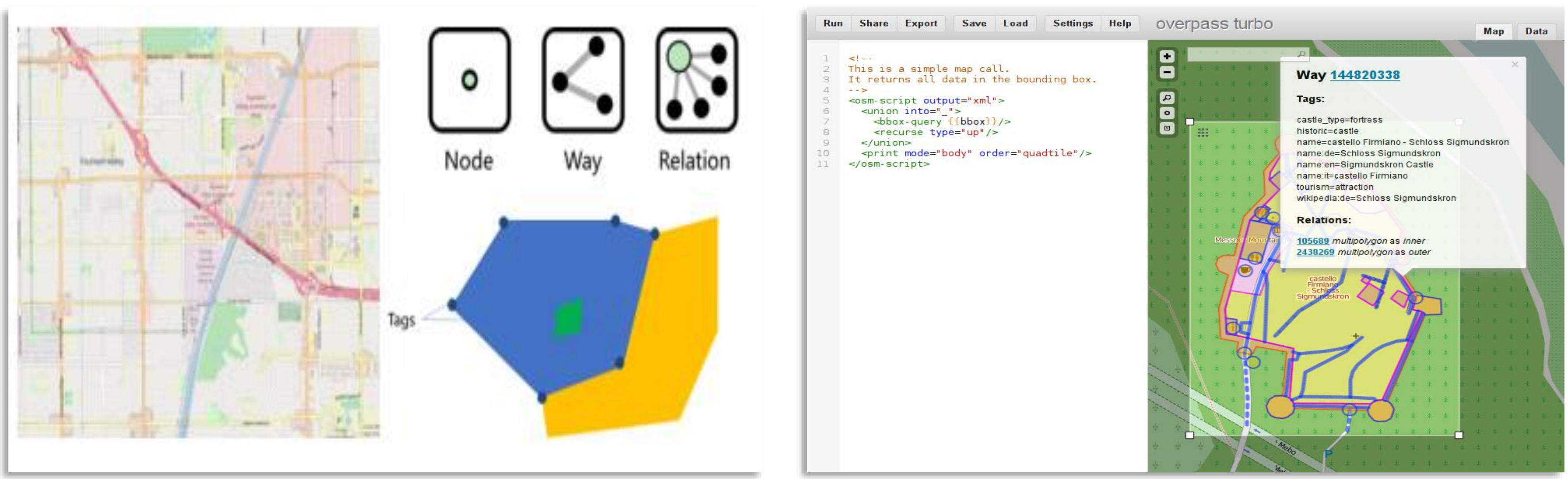
- Coordinates Collection for RS Image Acquisition
  - **Geographic information utilization**
    - Rich positional data with millions of point, line and region objects
    - Inherent semantic tags for interested image extraction with Map API




Searched baseball fields using Google Map API

# Geographic Information Integration

- Coordinates Collection for RS Image Acquisition
  - **Open source data**
    - Geographic data with rich semantic information that is timely updated, low cost and with large amount, e.g., OSM,WikiMapia …
    - Excellent interface for data customization, elements aligned with different maps




Elements of interest extracted from OSM

# Geographic Information Integration

- Coordinates Collection for RS Image Acquisition
  - **Geodatabase integration**
    - Public geodatabases released by the state institutions and communities
    - Domain-specific geodatabase that is publicly available


National Bridge Inventory

Transit Stations

Interstate Highways

Water System

Protestant Churches

ESRI Open Data Hub

……

Public geodatabases available for image coordinates collection

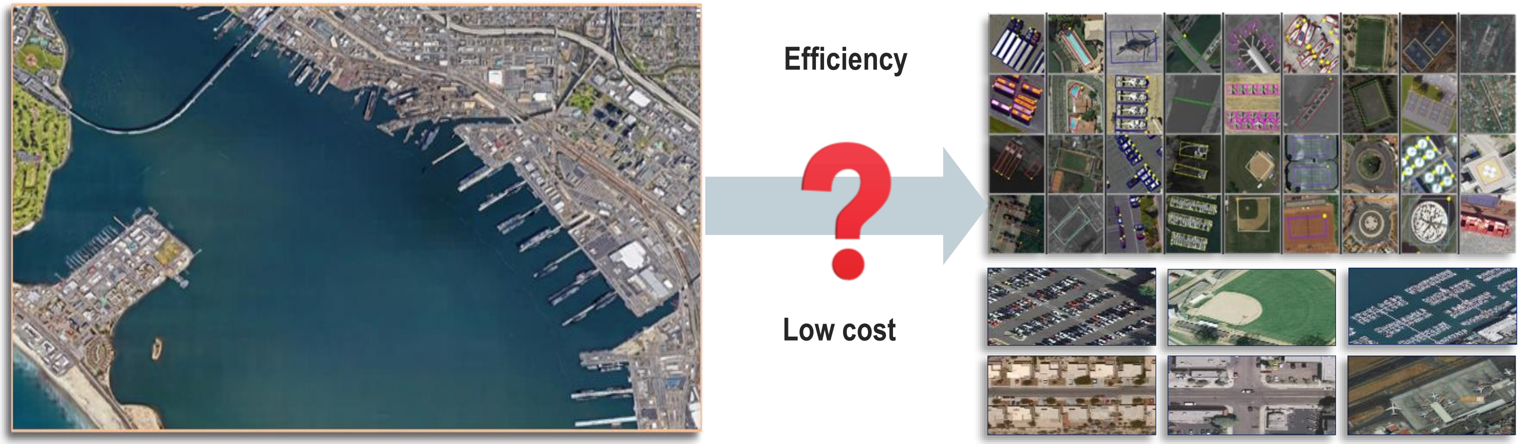
# Annotation methodology

## Manual Annotation

- Accuracy guarantee, but labor-intensive and time-consuming
- Hard to meet the scale requirements particularly for data-driven methods

## Automatic Annotation

- Reduce the cost of annotation by leveraging learning models
- Bias problem deriving from the initialization data and model capability



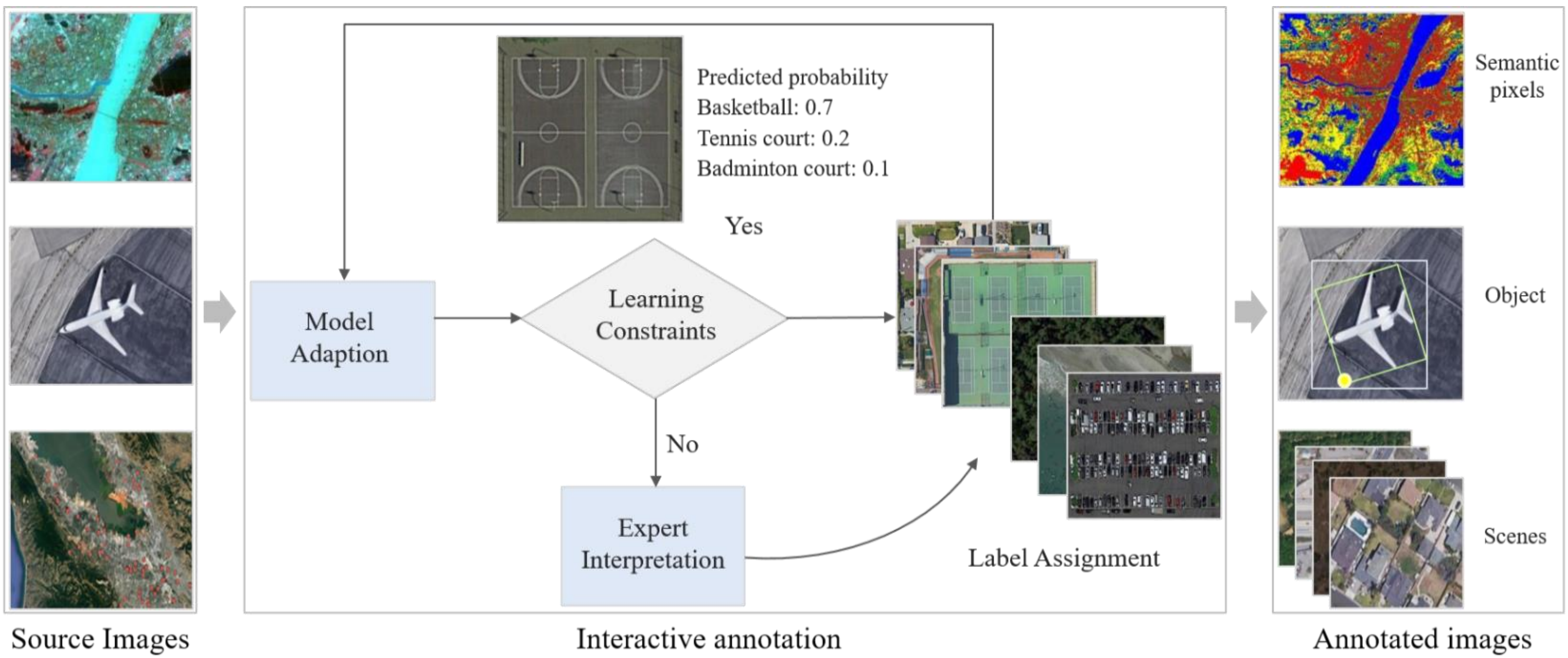

Source image

Annotated samples

# Annotation methodology

## Interactive Annotation

- Annotation with human-computer interaction, semi-automatic annotation
- Guarantee for quality and efficiency, toward large-scale dataset construction




General workflow of Semi-automatic annotation in RS images

# Quality Assessment

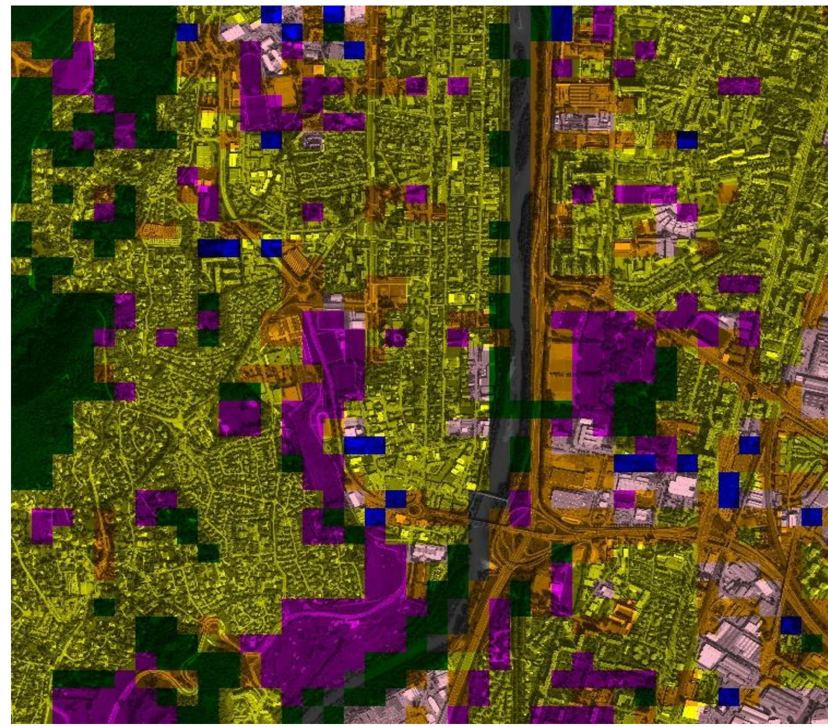
- ***Rules and Samples:*** annotation without ambiguity, specific samples for instructions
- ***Training of Annotators:*** well-qualified annotators for dataset quality guarantee
- ***Multi-stage Pipeline:*** annotation task decomposition
- ***Grading and Reward:*** incentive mechanism for incompetent/competent annotators
- ***Multiple Annotations:*** merge multiple accurate annotations
- ***Annotation Review:*** expert/peer review and quality rating
- ***Spot Check and Assessment:*** gold data for annotation quality assurance

# Outline

- Background
- Research Focus in the Past Decade
- Principles to Benchmark RS Image Interpretation
- **An Example： Million-AID**
- Challenges and Perspectives
- Conclusions

# Million-AID

- Scene Classification
  - **High-level knowledge expression to RS image contents**
  - **Semantic information recognition to local areas of RS images**



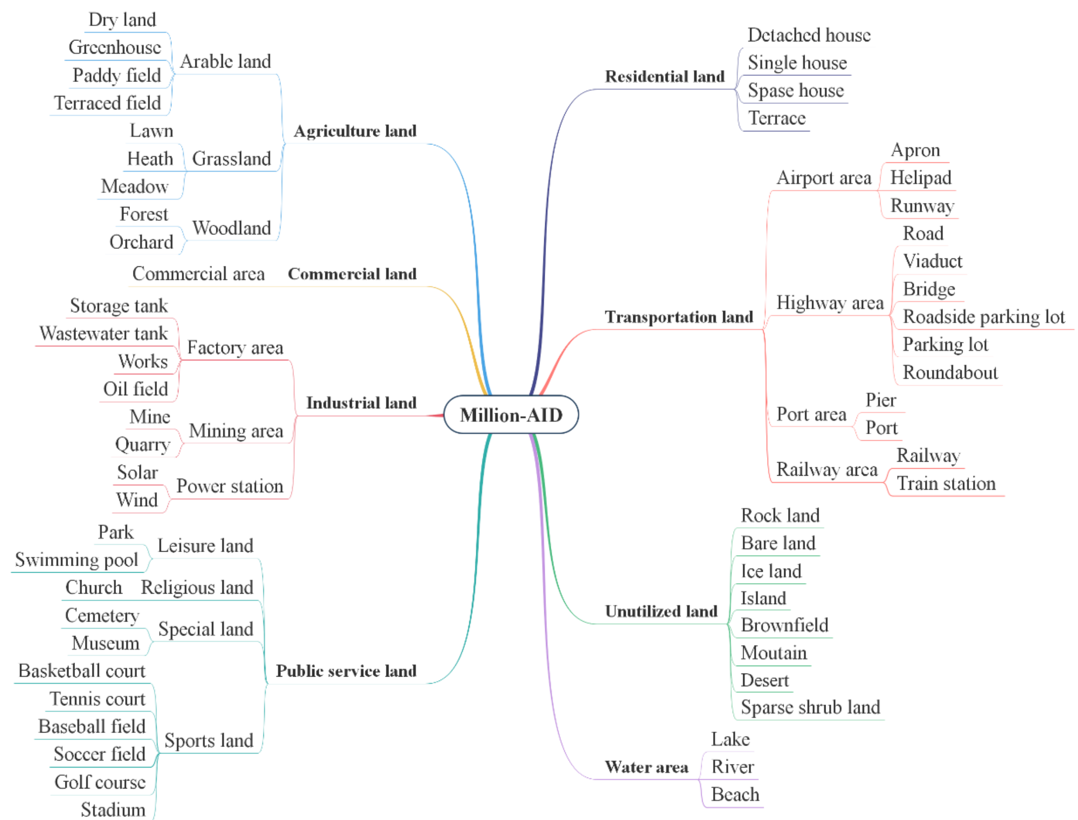
| Forest | Grass land | Parking lot | Resid. area | Indus. area | Water | Other |
|---|---|---|---|---|---|---|

# Million-AID

## Category Organization

***Chinese Land Use Classification Criteria***


***8 major categories with 57 sub-categories***

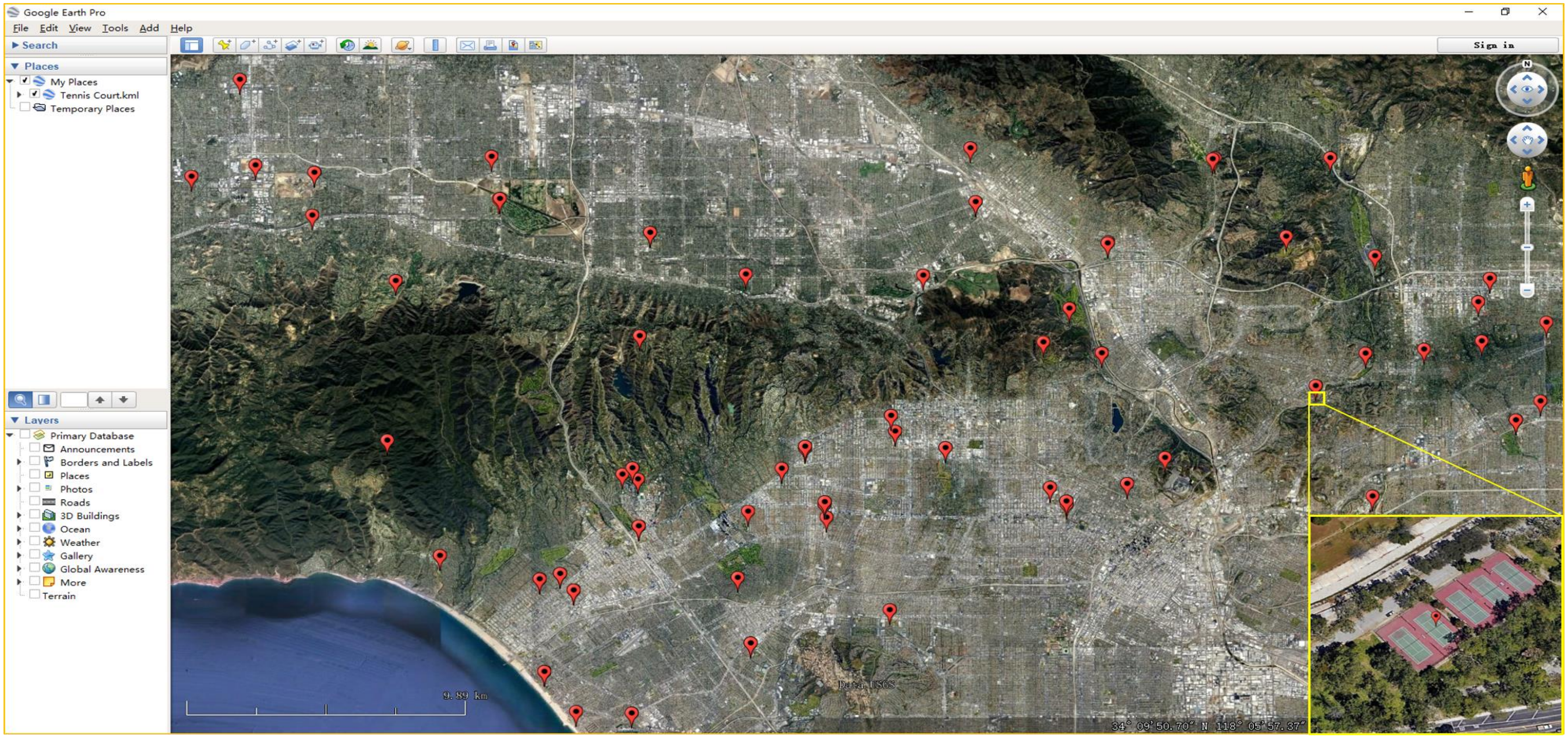



The hierarchical scene category network of Million-AID

# Million-AID

CAPTAIN
COMPUTATIONAL AND PHOTOGRAMMETRIC VISION

- Semantic Coordinates Collection
  - **Point coordinates using Google Map API**

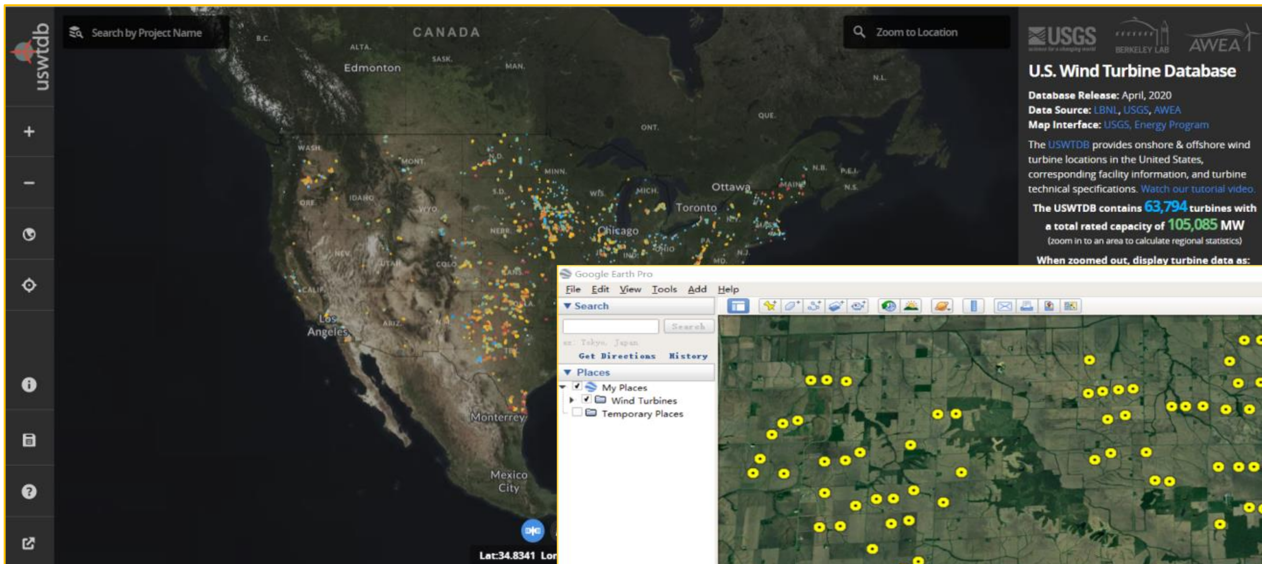



The points of searched tennis courts shown in Google Earth. We consider the tennis courts as point ground features. In this case, we use Google Map API for coordinates collection.

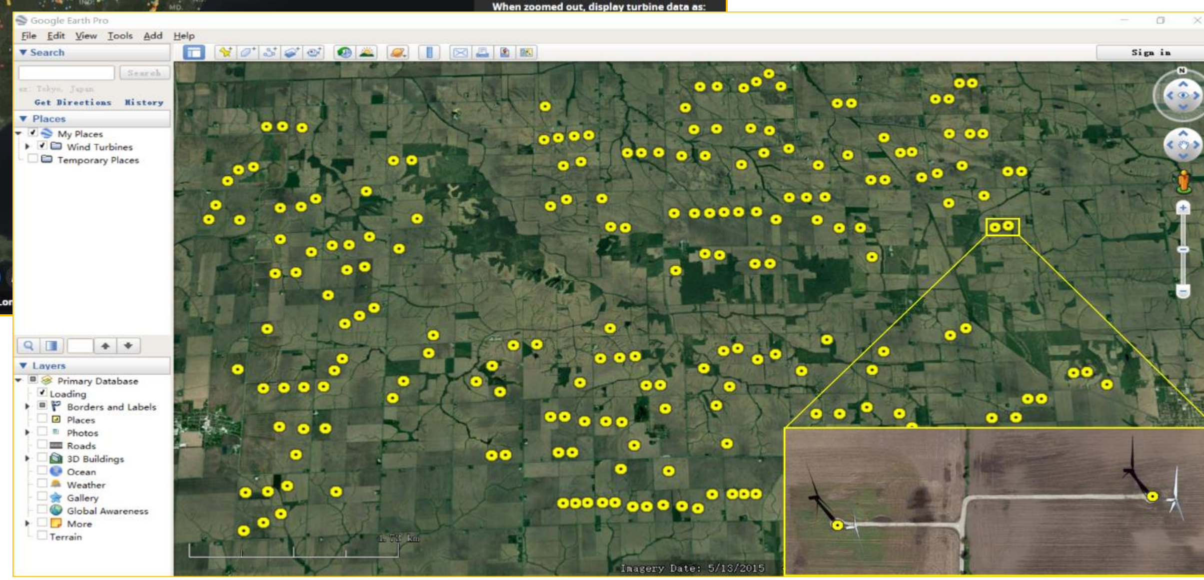
# Million-AID

## Semantic Coordinates Collection

### Point coordinates integrated from Geodatabase




U.S. Wind Turbine Database

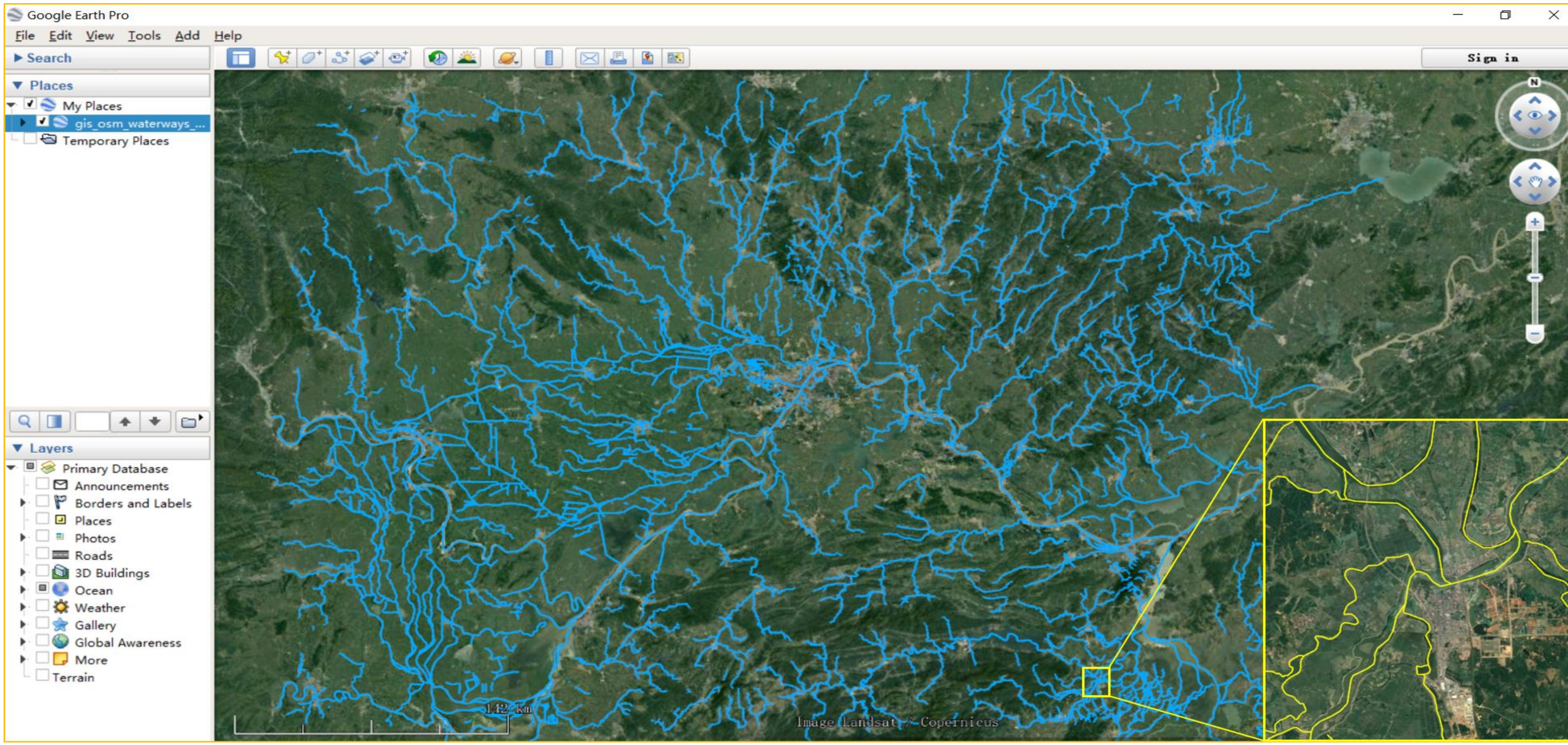



*Integrated W.T. data*

The points of wind turbines extracted from USWTDB and integrated in Google Earth. Generally, over 60, 000 objects of wind turbines can be collected by the database.

# Million-AID

- Semantic Coordinates Collection
  - **Line features extracted from OSM**




The river lines within a local area of China collected from OSM and displayed in Google Earth

# Million-AID

CAPTAIN
COMPUTATIONAL AND PHOTOGRAMMETRIC VISION

## ■ Semantic Coordinates Collection

### ■ Plane features customized on OSM


(a)

| | |
|---|---|
| aeroway | aerodrome |
| ele | 37 |
| iata | VRC |
| icao | RPUV |
| length | 1886 |
| name | Virac Airport |
| operator | Civil Aviation Authority of the Philippines |
| source | Wikipedia |
| wikidata | |
| wikipedia | |

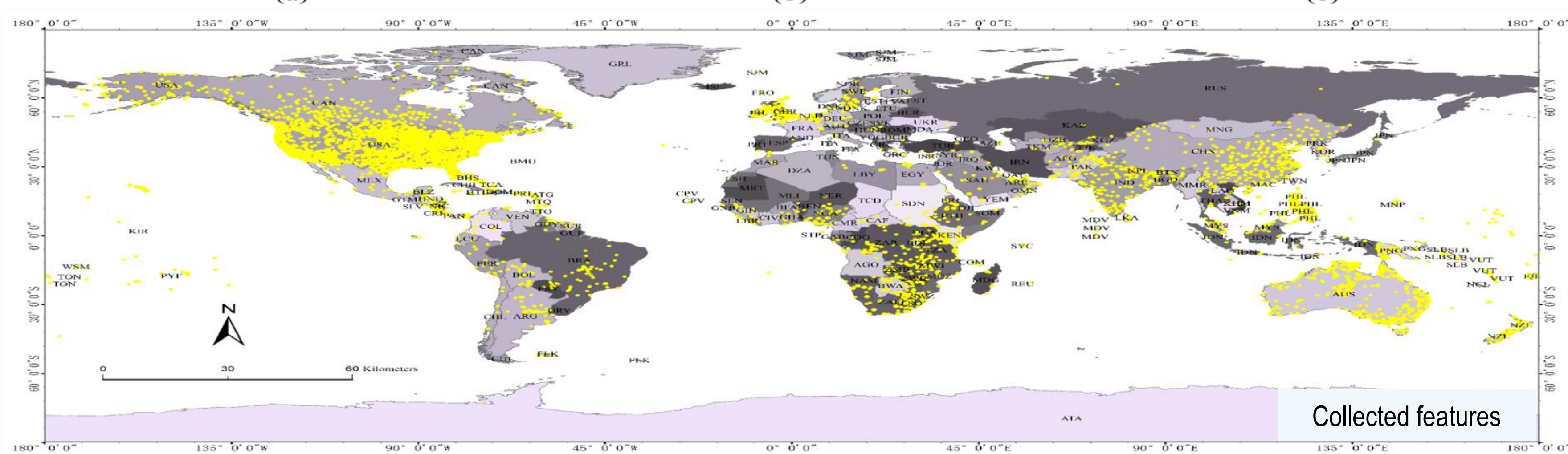
Semantic tags

(b)

```
<osm-script>
  <union>
    <query type="way">
      <has-kv k="aeroway" regv="aerodrome"/>
      <has-kv k="name" regv="[a|A]irport$|机场$"/>
    </query>
    <query type="relation">
      <has-kv k="aeroway" regv="aerodrome"/>
      <has-kv k="name" regv="[a|A]irport$|机场"/>
    </query>
  <recurse type="relation-way"/>
  </union>
  <union>
    <item/>
    <recurse type="down"/>
  </union>
  <print/>
</osm-script>
```

Data customization

(c)




The illustration of searching scenes of airports around the world. An airport in OSM contains a large amount of tags, which can be employed to search airports with specific semantic key-value labels.

# Million-AID

## Scene Image Acquisition

### Image blocks by the line, point, and plane data

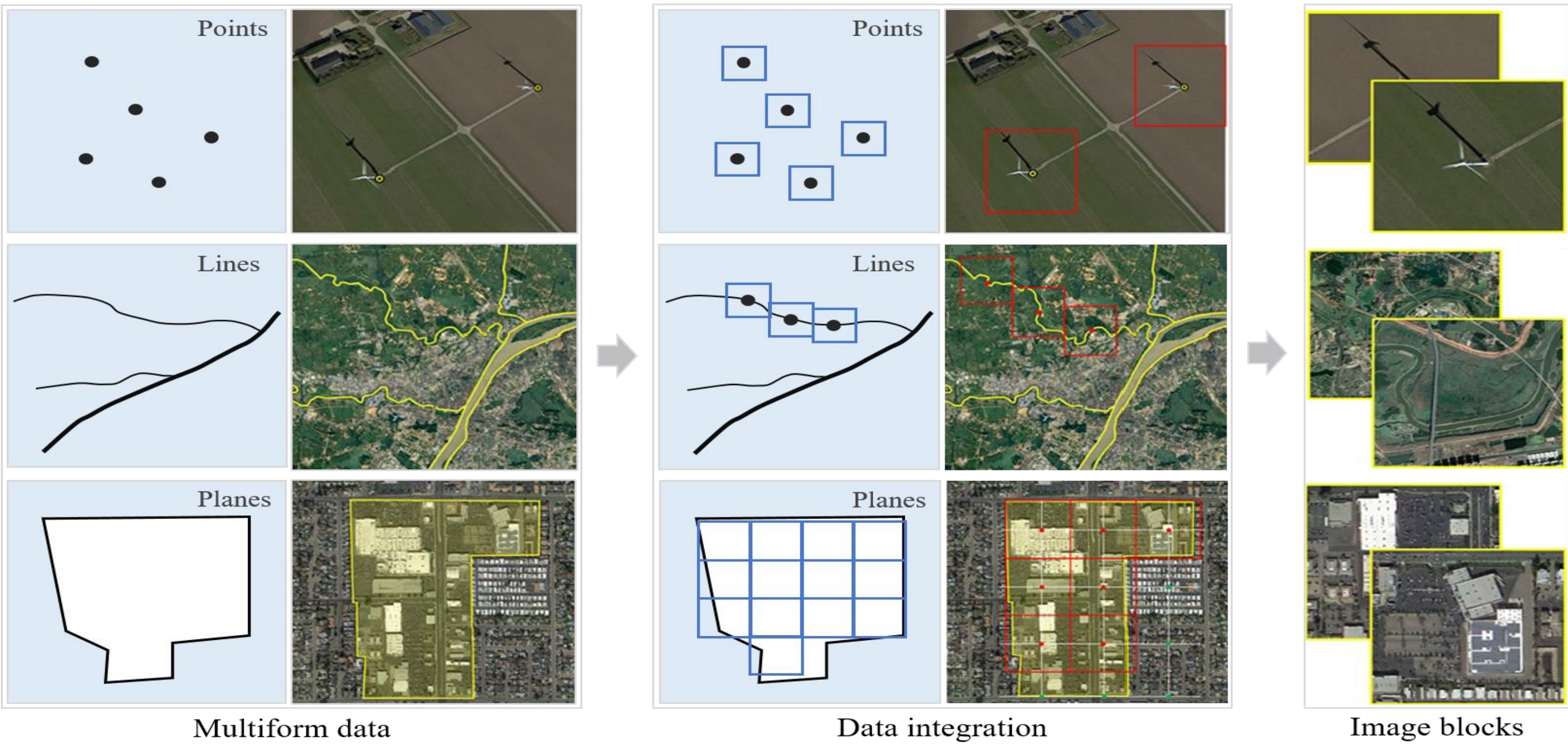


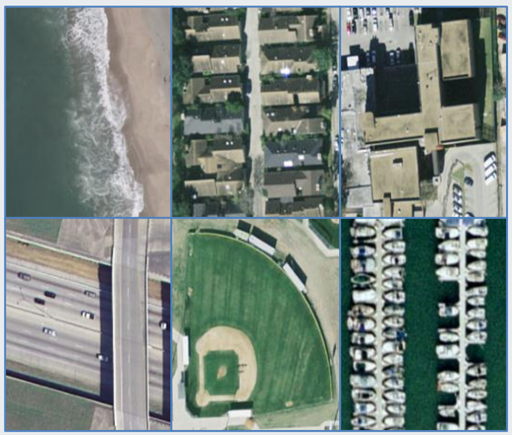

Illustration of the acquisition of RS scene images based on the collected geographic point, line and area data. ***Points***: centers of scene blocks. ***Line***: sampled by intervals. ***Plane***: sampled by mesh grids.

# Million-AID

## A Glimpse of Comparison

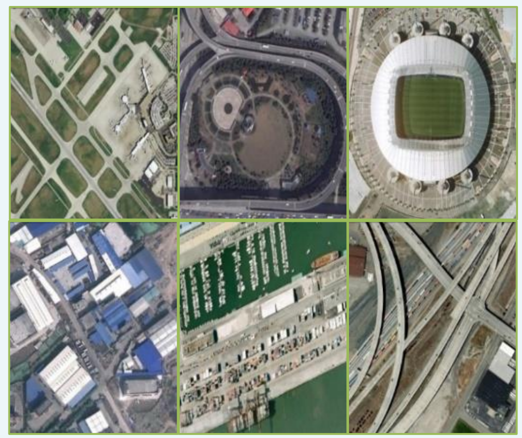
### Million-AID: DiRS, better approximate real applications

**UC-Merced**



- Categories: 21
- Image size: 256x256
- Resolution: ~ 0.3m
- Number of images: 2100

**WHU-RS19**



- Categories: 19
- Image size: 600x600
- Resolution: 0.2 ~ 10m
- Number of images: 950

**Million-AID**

- **Categories: 57**
- **Image size: 114~1024**
- **Resolution: 0.2 ~ 10m**
- **Number of images: 100M**

# Outline

- Background
- Research Focus in the Past Decade
- Principles to Benchmark RS Image Interpretation
- An Example： Million-AID
- **Challenges and Perspectives**
- Conclusions

# Challenges and Perspectives

- How to speed up the annotation process?

- **Visualization technology for RS image Annotation**
  - ***High-spectral images:*** several selected bands, band transformation …
  - ***Large-scale images*****:** efficient display technology helps to catch the content essence
  - ***SAR images:*** signal expression via physical means

- **Annotation Efficiency and Quality Improvement**
  - ***Cooperation with application departments:*** convert production data to algorithms
  - ***Annotation tools:*** open-sourced and professional tools for RS image annotation
  - ***Noisy annotations:*** noise cleansing, performance impact, tolerant algorithms

# Challenges and Perspectives

- Speed up the annotation process

Annotation tools for image dataset construction

| No. | Name | Year | Description |
|---|---|---|---|
| 1 | LabelMe | 2008 | An online image annotation tool that supports various annotation primitives, including polygon, rectangle, circle, line and point. |
| 2 | Video Annotation Tool from Irvine, California (VATIC) | 2012 | An online tool that efficiently scaling up video annotation with crowdsourced marketplaces (*e.g.*, AMT). |
| 3 | LabelImg | 2015 | A popular graphical image annotation application that labels objects in images with bounding boxes. |
| 4 | Visual Object Tagging Tool (VOTT) | 2017 | An open source annotation and labeling tool for image and video assets, extensible for importing/exporting data to local or cloud storage providers, including Azure Blob Storage and Bing Image Search. |
| 5 | Computer Vision Annotation Tool (CVAT) | 2018 | A universal data annotation approach for both individuals and teams, supporting large-scale semantic annotation for scene classification, object detection and image segmentation. |
| 6 | Image Tagger | 2018 | An open source online platform to create and manage image data and diverse labels (*e.g.*, bounding box, polygon, line and point), with friendly support for collaborative image labeling. |
| 7 | Polygon RNN++ | 2018 | A deep learning-based annotation strategy, producing polygonal annotation of objects segmentation interactively using humans-in-the-loop. |
| 8 | Makesence.AI | 2019 | An open source and online image annotation platform, using different artificial model to give recommendations as well as automate repetitive and tedious labeling activities. |
| 9 | VGG Image Annotator (VIA) | 2019 | A simple and standalone manual annotation software for image and video, providing rich labels like point, line, polygon as well as circle and ellipse without project management. |

# Outline

- Background
- Research Focus in the Past Decade
- Principles to Benchmark RS Image Interpretation
- An Example： Million-AID
- Challenges and Perspectives
- **Conclusions**

# Conclusions

- **A review of annotated datasets for RS image interpretation**
  - Covering literature published over the past decade
  - A systematic review of the existing RS image datasets concerning the current mainstream of RS image interpretation tasks
- **Principles to build RS image benchmarks**
  - DiRS: on creating benchmark datasets for RS image interpretation
  - A picture of coordinates collection, methodology for RS image dataset construction
- **An example for construction : Million-AID**
  - A large-scale benchmark dataset for RS image scene classification

# Dataset References

## Scene Classification

- **UC-Merced:** Y. Yang and S. Newsam. Bag-of-visual-words and spatial extensions for land-use classification. in *Proc. 18th SIGSPATIAL Int. Conf. Adv. GIS.* ACM, 2010, pp. 270–279.
- **WHU-RS19:** G. Sheng, W. Yang, T. Xu, et. al. High-resolution satellite scene classification using a sparse coding based multiple feature combination. *Int. J. Remote Sens.*, vol. 33, no. 8, pp. 2395–2412, 2012.
- **RSSCN7:** Q. Zou, L. Ni, T. Zhang, et. al. Deep learning based feature selection for remote sensing scene classification. *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 11, pp. 2321–2325, 2015.
- **SAT-4/6:** S. Basu, S. Ganguly, S. Mukhopadhyay, et. al. Deepsat: A learning framework for satellite imagery. in *Proc. 23th SIGSPATIAL Int. Conf. Adv. GIS*, 2015, pp. 1–10.
- **BSC:** O. A. B. Penatti, K. Nogueira, and J. A. dos Santos. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? in *Proc. IEEE Conf. CVPRW*, 2015, pp. 44–51.
- **RSC11:** L. Zhao, P. Tang, and L. Huo. Feature significance-based multibag-of-visual-words model for remote sensing image scene classification. *J. App. Remote Sens.*, vol. 10, no. 3, p. 035004, 2016.
- **SIRI-WHU:** Q. Zhu, Y. Zhong, B. Zhao, et. al. Bag-of-visualwords scene classifier with local and global features for high spatial resolution remote sensing imagery. *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 6, pp. 747–751, 2016.
- **NWPU-RESISC45:** G. Cheng, J. Han, and X. Lu. Remote sensing image scene classification: Benchmark and state of the art. *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, 2017.
- **RSD46-WHU:** Z. Xiao, Y. Long, D. Li, et. al. High-resolution remote sensing image retrieval based on cnns from a dimensional perspective. *Remote Sens.*, vol. 9, no. 7, p. 725, 2017.

# Dataset References

## Scene Classification

- **AID:** G.-S. Xia, J. Hu, F. Hu, et. al. AID: A benchmark data set for performance evaluation of aerial scene classification. *IEEE Trans. Geosci. Remote Sens*., vol. 55, no. 7, pp. 3965–3981, 2017.
- **AID++:** P. Jin, G.-S. Xia, F. Hu, et. al. AID++: An updated version of aid on scene classification. In *Proc. IEEE IGARSS*, 2018, pp. 4721–4724.
- **EuroSAT:** P. Helber, B. Bischke, A. Dengel, et. al. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens*.,vol. 12, no. 7, pp. 2217–2226, 2019.
- **PatternNet:** W. Zhou, S. Newsam, C. Li, et. al. Patternnet: A benchmark dataset for performance evaluation of remote sensing image retrieval. *ISPRS J. Photogrammetry Remote Sens*., vol. 145, pp. 197–209, 2018.
- **RSI-CB:** H. Li, X. Dou, C. Tao, et. al. RSI-CB: A large-scale remote sensing image classification benchmark using crowdsourced data. *Sensors*, vol. 20, no. 6, pp. 1594, 2020.

# Dataset References

## Object Detection


- **TAS:** G. Heitz and D. Koller. Learning spatial context: Using stuff to find things. in *Proc. ECCV*, 2008, pp. 30–43.
- **SZTAKI-INRIA:** C. Benedek, X. Descombes, and J. Zerubia. Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics. *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 33–50, 2012.
- **NWPU-VHR10:** G. Cheng, J. Han, P. Zhou, et al. Multi-class geospatial object detection and geographic image classification based on collection of part detectors. *ISPRS J. Photogrammetry Remote Sens.*, vol. 98, pp. 119–132, 2014.
- **DLR 3k:** K. Liu and G. Mattyus. Fast multiclass vehicle detection on aerial images. *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 9, pp. 1938–1942, 2015.
- **UCAS-AOD:** H. Zhu, X. Chen, W. Dai, et al. Orientation robust object detection in aerial images using deep convolutional neural network. in *Proc. IEEE ICIP*, 2015, pp. 3735–3739.
- **VEDAI:** S. Razakarivony and F. Jurie. Vehicle detection in aerial imagery: A small target detection benchmark. *J Vis. Commun. Image R.*, vol. 34, pp. 187–203, 2016.
- **COWC:** T. N. Mundhenk, G. Konjevod, W. A. Sakla, et al. A large contextual dataset for classification, detection and counting of cars with deep learning. in *Proc. ECCV*, 2016, pp. 785–800.
- **HRSC2016:** Z. Liu, H. Wang, L. Weng, et al. Ship rotated bounding box space for ship extraction from high-resolution optical satellite images with complex backgrounds. *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 8, pp. 1074–1078, 2016.

# Dataset References

## Object Detection


- **CARPPK:** M.-R. Hsieh, Y.-L. Lin, and W. H. Hsu. Drone-based object counting by spatially regularized regional proposal networks. in *Proc. IEEE ICCV*, 2017, pp. 4145–4153.
- **RSOD:** Y. Long, Y. Gong, Z. Xiao, et al. Accurate object localization in remote sensing images based on convolutional neural networks. *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2486–2498, 2017.
- **LEVIR:** Z. Zou and Z. Shi. Random access memories: A new paradigm for target detection in high resolution aerial remote sensing images. *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1100–1111, 2018.
- **VisDrone:** P. Zhu, L. Wen, X. Bian, et al. Vision meets drones: A challenge. *arXiv preprint* arXiv:1804.07437, 2018.
- **xView:** D. Lam, R. Kuzma, K. McGee, et al. Xview: Objects in context in overhead imagery. *arXiv preprint* arXiv:1802.07856, 2018.
- **HRRSD:** Y. Zhang, Y. Yuan, Y. Feng, et al. Hierarchical and robust convolutional neural network for very high-resolution remote sensing object detection. *IEEE Trans. Geosci. Remote Sens*., vol. 57, no. 8, pp. 5535–5548, 2019.
- **DIOR:** K. Li, G. Wan, G. Cheng, et al. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS J. Photogrammetry Remote Sens*., vol. 159, pp. 296 – 307, 2020.
- **DOTA:** G.-S. Xia, X. Bai, J. Ding, et. al. DOTA: A large-scale dataset for object detection in aerial images. in *Proc. IEEE Conf. CVPR*., 2018, pp. 3974–3983.

# Dataset References

## Semantic Segmentation


- **KSC/Botswana:** J. Ham, Y. Chen, M. M. Crawford, et. al. Investigation of the random forest framework for classification of hyperspectral data. *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 492–501, 2005.
- **Salinas/University of Pavia/Pavia Centre:** Pavia dataset. [Online]. Available: *http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes*
- **ISPRS Vaihingen/Potsdam:** ISPRS 2D semantic labeling contest. [Online]. Available: *http://www2.isprs.org/commissions/comm3/wg4/semantic-labeling.html*
- **Massachusetts Buildings/Roads:** V. Mnih. Machine learning for aerial image labeling. *Ph.D. dissertation*, University of Toronto, 2013.
- **Indian Pines:** M. F. Baumgardner, L. L. Biehl, and D. A. Landgrebe. 220 band aviris hyperspectral image data set: June 12, 1992 indian pine test site 3. Sep 2015. [Online]. Available: *https://purr.purdue.edu/publications/1947/1*
- **Zurich Summer:** M. Volpi and V. Ferrari. Semantic segmentation of urban scenes by learning local class interactions. In *Proc. IEEE Conf. CVPRW*, 2015, pp. 1–9.
- **Inria Dataset:** E. Maggiori, Y. Tarabalka, G. Charpiat, et. al. Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark. in *Proc. IEEE Int. Geosci. Remote Sens. Symposium*, 2017, pp. 3226–3229.
- **EVlab-SS:** M. Zhang, X. Hu, L. Zhao, et. al. Learning dual multi-scale manifold ranking for semantic segmentation of high-resolution images. *Remote Sens.*, vol. 9, no. 5, p. 500, 2017.

# Dataset References

## Semantic Segmentation


- **RIT-18:** R. Kemker, C. Salvaggio, and C. Kanan. Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. *ISPRS J. Photogrammetry Remote Sens.*, vol. 145, pp. 60–77, 2018.
- **WHU Building:** S. Ji, S. Wei, and M. Lu. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 574–586, 2018.
- **SAR-building:** W. Yang and G. Xia. Sar-building: Building detection dataset in sar images. 2019. [Online]. Available: http://captain.whu.edu.cn/WUDA-RSImg/sarb.html
- **SEN12MS:** M. Schmitt, L. H. Hughes, C. Qiu, et al. Sen12ms–a curated dataset of georeferenced multi-spectral sentinel-1/2 imagery for deep learning and data fusion. *arXiv preprint* arXiv:1906.07789, 2019.
- **So2Sat LCZ42:** X. X. Zhu, J. Hu, C. Qiu, et al. So2sat lcz42: A benchmark dataset for global local climate zones classification. *IEEE Geosci. Remote Sens. Mag.*, 2020.
- **GID:** X.-Y. Tong, G.-S. Xia, Q. Lu, et. al. Land-cover classification with high-resolution remote sensing images using transferable deep models. *Remote Sens. Environ.,* vol. 237, p. 111322, 2020.
- **UAVid:** Y. Lyu, G. Vosselman, G.-S. Xia, et al. Uavid: A semantic segmentation dataset for uav imagery. *ISPRS J. Photogrammetry Remote Sens.*, 2020.

# Dataset References

## Change Detection

- **SZTAKI AirChange:** C. Benedek and T. Sziranyi. Change detection in optical aerial images by a multilayer conditional mixed markov model. *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 10, pp. 3416–3430, 2009.
- **AICD:** N. Bourdis, D. Marraud, and H. Sahbi. Constrained optical flow for aerial image change detection. in *Proc. IEEE IGARSS*, 2011, pp. 4176–4179.
- **Taizhou/Kunshan Data:** W. Zhang, X. Lu, and X. Li. A coarse-to-fine semi-supervised change detection for multispectral images. *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3587–3599, 2018.
- **Yancheng:** A. Song, J. Choi, Y. Han, et al. Change detection in hyperspectral images using recurrent 3d fully convolutional networks. *Remote Sens.*, vol. 10, no. 11, p. 1827, 2018.
- **Urban-rural boundary of Wuhan:** D. He, Y. Zhong, and L. Zhang. Land cover change detection based on spatial-temporal sub-pixel evolution mapping: A case study for urban expansion. in *Proc. IEEE IGARSS*, 2018, pp. 1970–1973.
- **Hermiston City area, Oregon:** J. Lopez-Fandino, A. S. Garea, D. B. Heras, et al. Stacked autoencoders for multiclass change detection in hyperspectral images. in *Proc. IEEE IGARSS*, 2018, pp. 1906–1909.
- **OSCD:** R. C. Daudt, B. Le Saux, A. Boulch, et al. Urban change detection for multispectral earth observation using convolutional neural networks. in *Proc. IEEE IGARSS*, 2018, pp. 2115–2118.
- **Quasi-urban areas:** S. Saha, F. Bovolo, and L. Brurzone. Unsupervised multiplechange detection in vhr optical images using deep features. in *Proc. IEEE GARSS*, 2018, pp. 1902–1905.

# Dataset References

## Change Detection

- **WHU-Building:** S. Ji, S. Wei, and M. Lu. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 574–586, 2018.
- **Season-varing:** M. Lebedev, Y. V. Vizilter, O. Vygolov, V. Knyaz, and A. Y. Rubis. Change detection in remote sensing images using conditional adversarial networks. *Int. arch. photogramm. remote sens. spat. inf. sci.*, vol. 42, no. 2, 2018.
- **ABCD:** A. Fujita, K. Sakurada, T. Imaizumi, et al. Damage detection from aerial images via convolutional neural networks. In *Proc. IAPR Int. Conf. Machine Vision Applications*, 2017, pp. 5–8.
- **MtS-WH:** C. Wu, L. Zhang, and B. Du. Kernel slow feature analysis for scene change detection. *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 4, pp. 2367–2384, 2017.
- **LEVIR-CD:** H. Chen and Z. Shi. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. *Remote Sens.*, vol. 12, no. 10, p. 1662, 2020.
- **CSDN:** K. Yang, Z. Liu, G.-S. Xia, and L. Zhang. CSDN: A cross spatial difference network for semantic change detection in remote sensing images. In *Proc. IEEE IGARSS*, 2020.

CAPTAIN

COMPUTATIONAL AND PHOTOGRAMMETRIC VISION

# THANKS


*School of Computer Science, Wuhan University*

*Institute of Artificial Intelligence, Wuhan University*

*State Key Lab. LIESMARS, Wuhan University*

Gui-Song Xia (guisong.xia@whu.edu.cn)