Highly Scalable Machine Learning Methods on Sunway TaihuLight

Haohuan Fu

haohuan@tsinghua.edu.cn

High Performance Geo-Computing (HPGC) Group

http://www.thuhpgc.org

清華大學
Tsinghua University
地球系统科学研究中心
清华大学全球变化研究院
国家超级计算无锡中心
National Supercomputing Center in Wuxi

Earth Science and Supercomputers

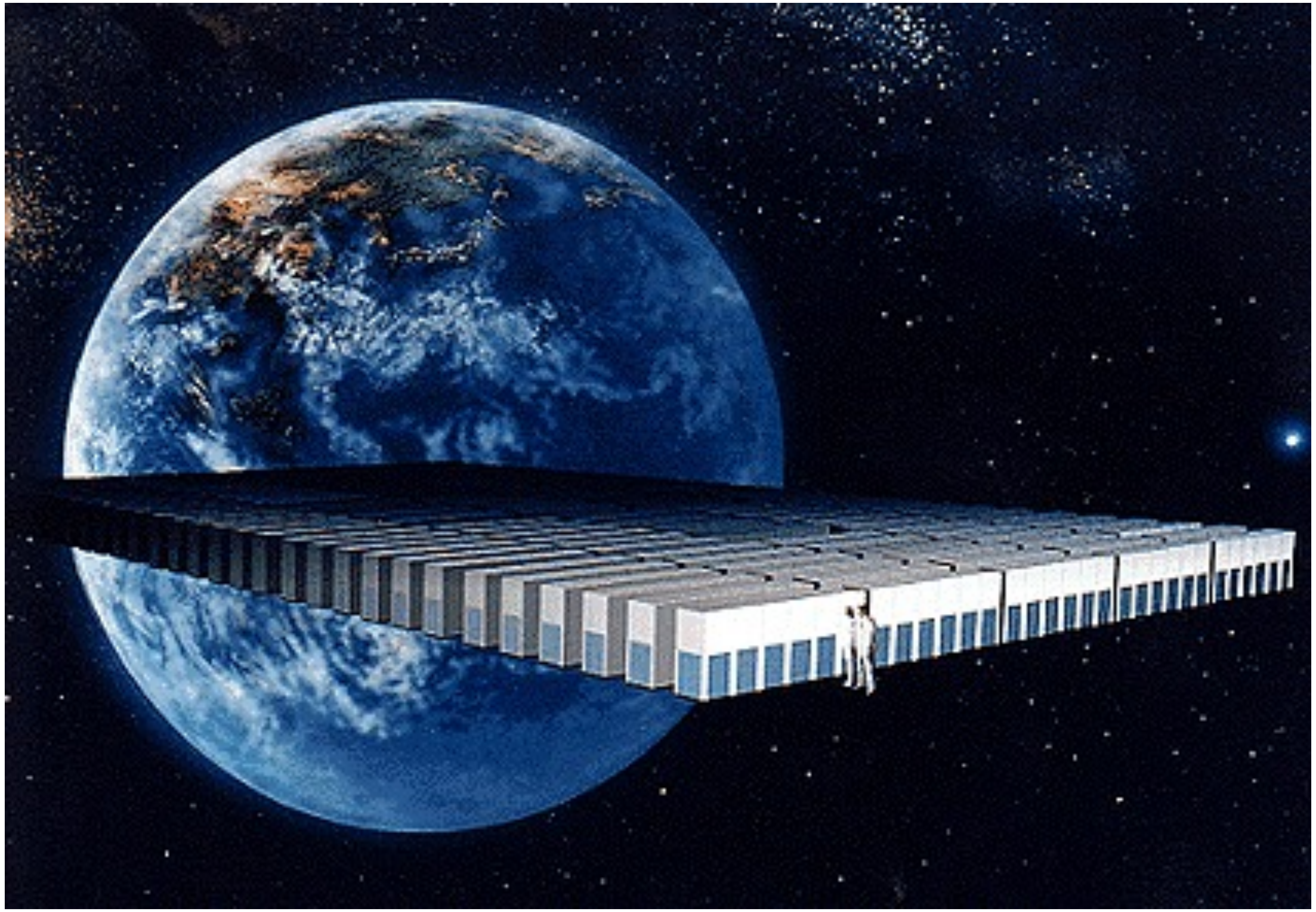


figure credit: Earth Simulator, JPN

- Create a digital earth, so as to:
  - simulate
  - analyze
  - understand
  - predict and mitigate

Two Major Functions

Simulation

Data Analysis

Two Major Functions

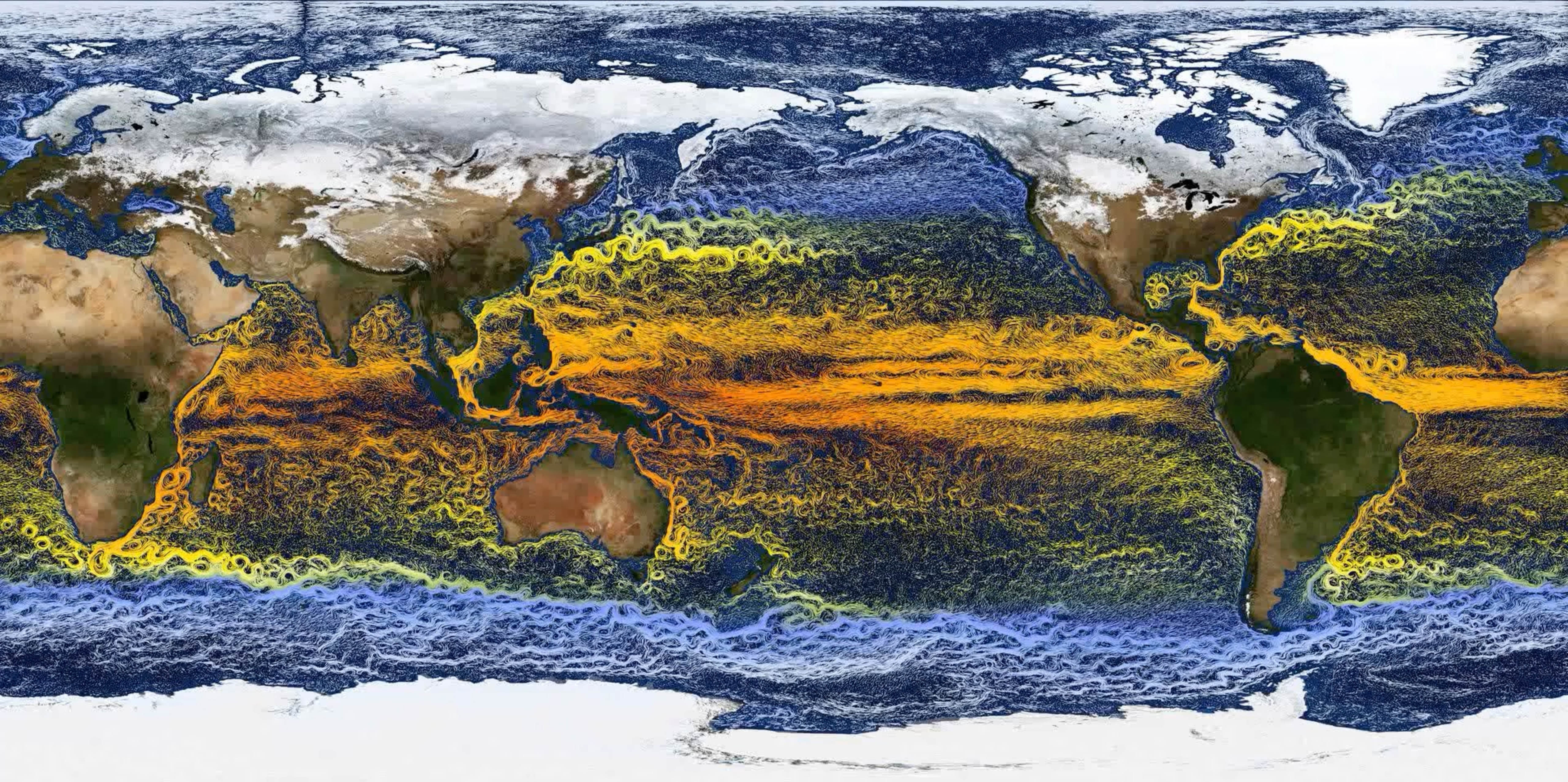
To design highly efficient and highly scalable simulation applications

To develop intelligent data mining methods for the analysis of BIG scientific DATA

Two Major Functions

To design highly efficient and highly scalable simulation applications

To develop intelligent data mining methods for the analysis of BIG scientific DATA

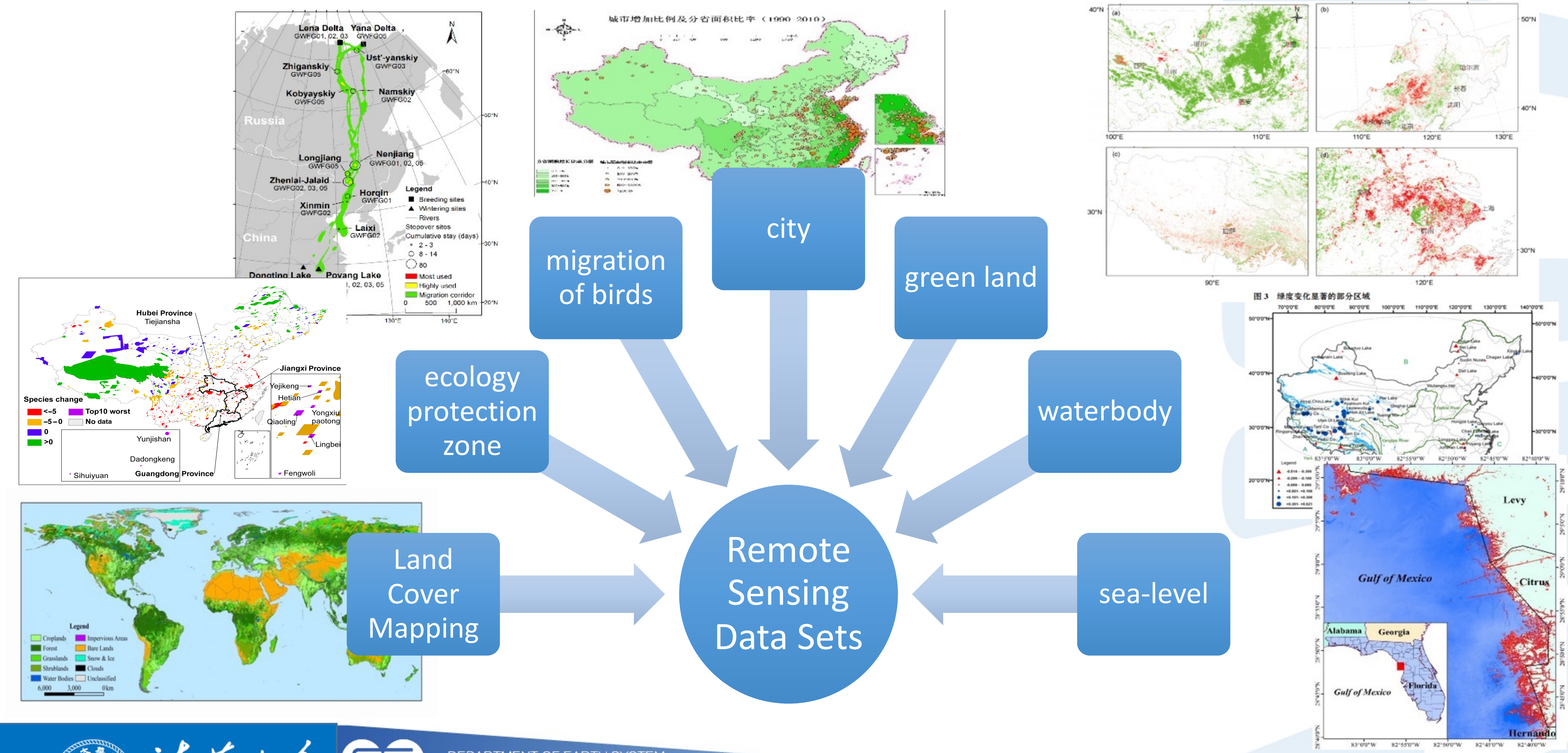


Two Major Functions

To design highly efficient and highly scalable simulation applications

To develop intelligent data mining methods for the analysis of BIG scientific DATA

Look Ahead: Data-Driven Modeling and Prediction



Tsinghua University
DEPARTMENT OF EARTH SYSTEM
清华大学地球系统科学系

Potential of data: meter-level resolution, study of specific birds or trees, a huge help for models

Efforts on Sunway TaihuLight

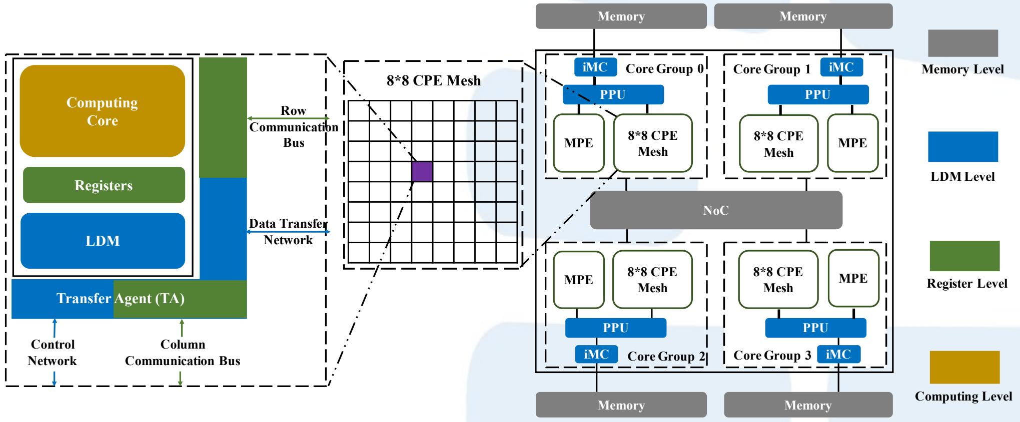
Application

AI-Software

Hardware

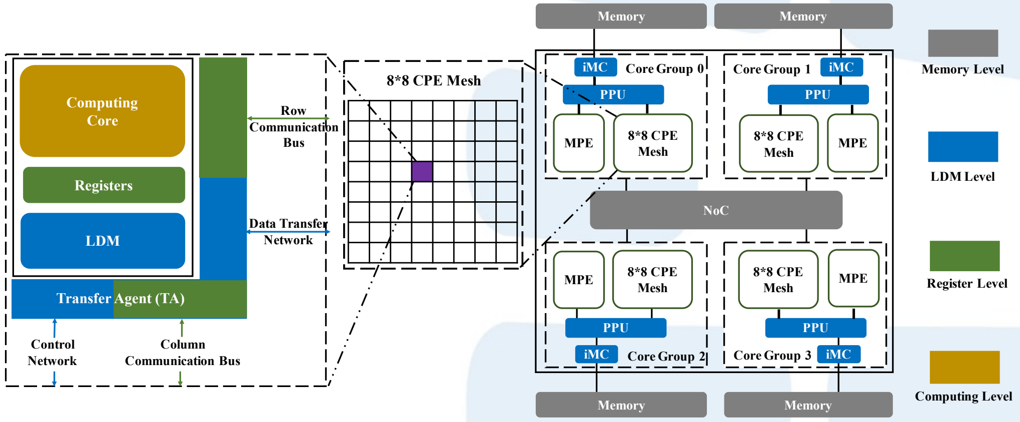
Sunway TaihuLight

- Heterogeneity within the chip
- Top 1 in Top500 (2016-2017)
- 125 Pflops
- Over 10 million cores



Sunway TaihuLight

- Heterogeneity within the chip
- Top 1 in Top500 (2016-2017)
- 125 Pflops
- Over 10 million cores



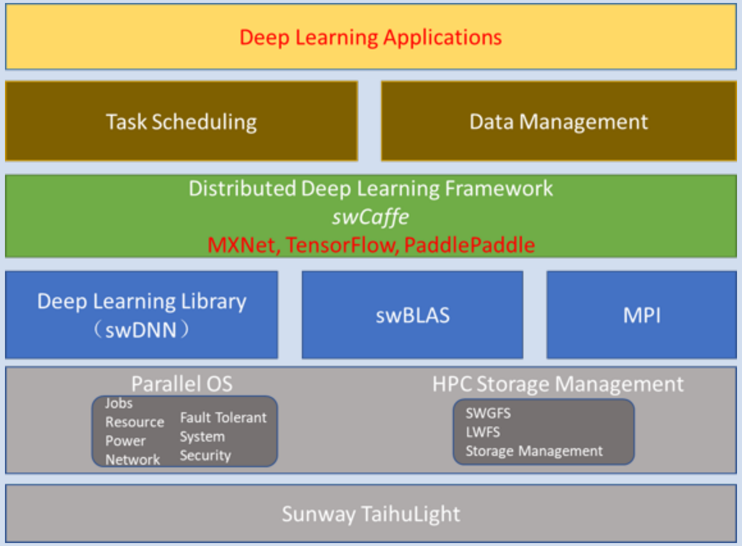
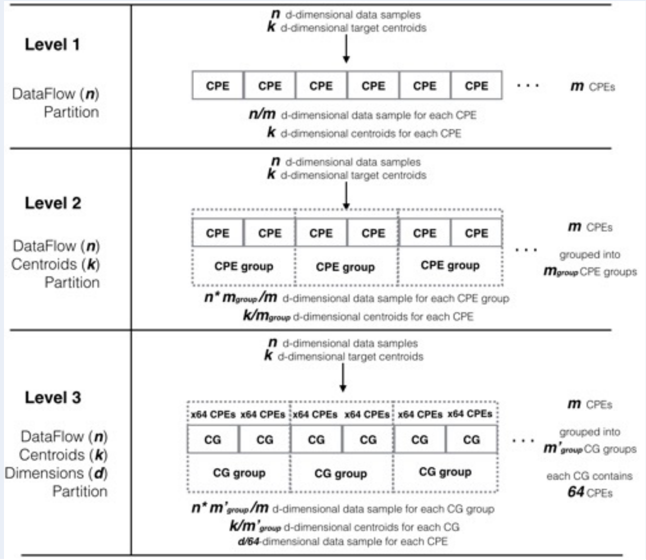
Deep Learning performance is decided as an important metric to benchmark coming Exa-Scale systems.

Efforts on Sunway TaihuLight

Application

AI-Software

Hardware

Domain	Software	Scale	
Machine Learning	swDNN + swCaffe	256 to 1,024 nodes	
	k-means	Up to 10,240 nodes (2 million cores)	

Sunway DNN Software Stack

SWCAFFE swCaffe
Auto swDNN swAutoDNN
SWDNN swDNN
SWGEMM swGEMM

swDNN v2.0

MLP	LSTM	Explicit Conv / Im2col/ col2im	implicit Conv	Winograd Conv	Pooling	Tensor Transfor-mation	Batch-Norm

54% ➡**123%**

Conv efficiency

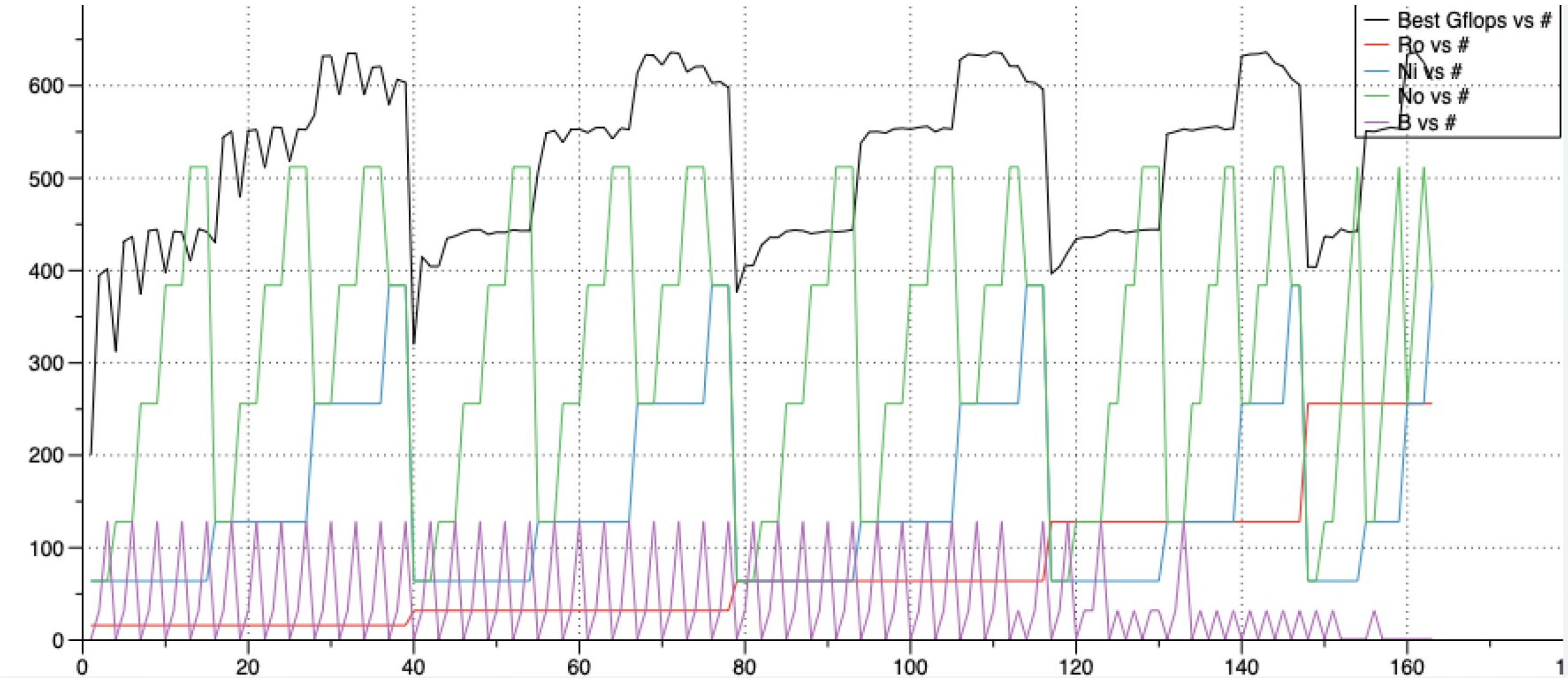
1.6 Tflops ➡**3.64** Tflops

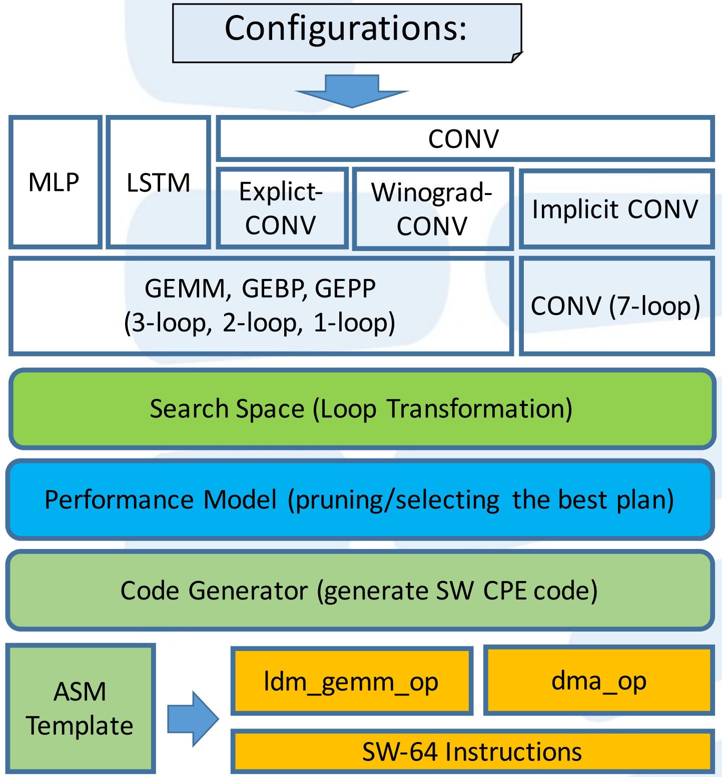
Conv performance

80%

Memory bandwidth utilization

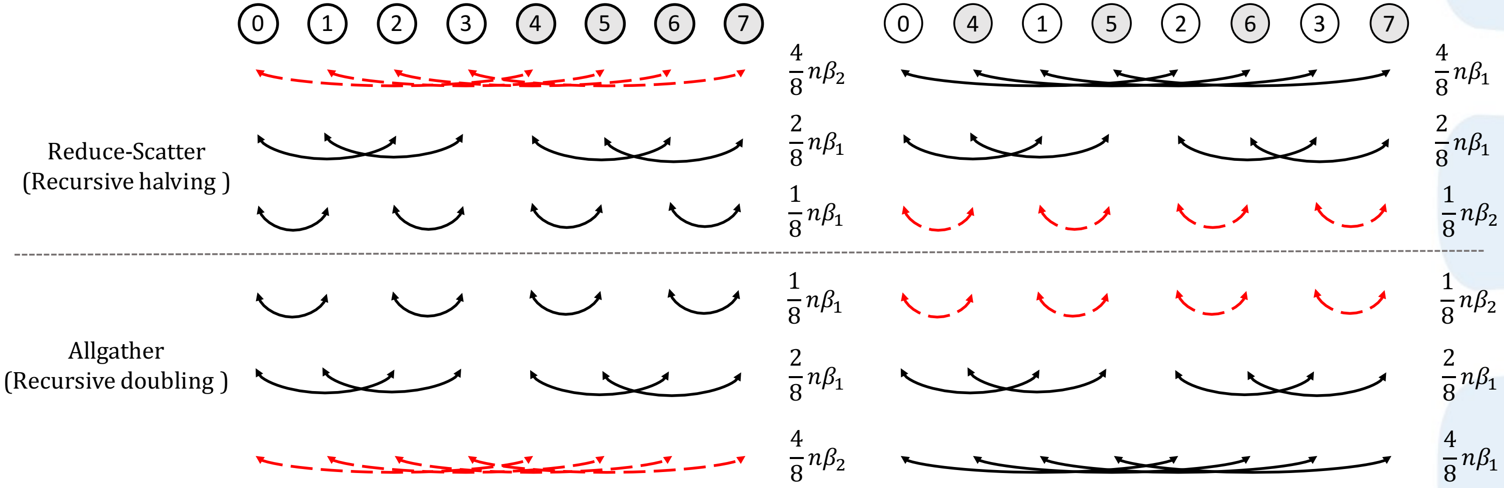
AutoDNN: auto tuning for DNN





Topology-aware Allreduce

- Rabenifner Algi.+ Reorder the logical number according to topology position



Original All-Reduce Cost : $6\alpha + \frac{7}{8}\gamma + \frac{3}{4}n\beta_1 + n\beta_2$ Improved All-Reduce Cost : $6\alpha + \frac{7}{8}\gamma + \frac{3}{2}n\beta_1 + \frac{1}{4}n\beta_2$

$$t_{allreduce} = t_{reduce-scatter} + t_{allgather} \quad (2)$$

$$t_{reduce-scatter} = \log p\alpha + (q-1)\beta_1\frac{n}{p} + (p-q)\beta_2\frac{n}{p} + \frac{p-1}{p}n\gamma \quad (3)$$

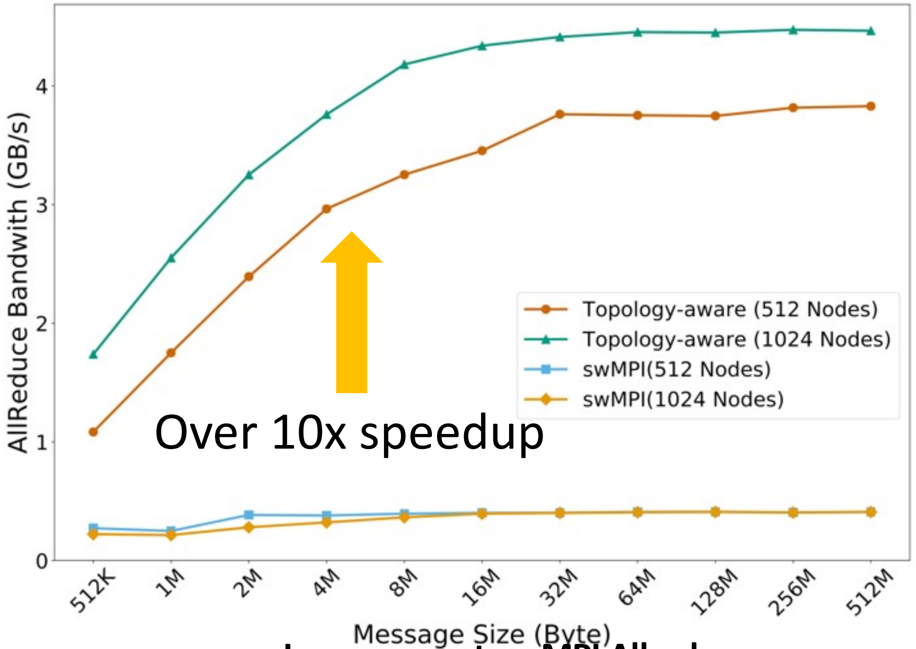
$$t_{allgather} = \log p\alpha + (q-1)\beta_1\frac{n}{p} + (p-q)\beta_2\frac{n}{p} \quad (4)$$

$$t_{new-reduce-scatter} = \log p\alpha + (p-\frac{p}{q})\beta_1\frac{n}{p} + (\frac{p}{q}-1)\beta_2\frac{n}{p} + \frac{p-1}{p}n\gamma \quad (5)$$

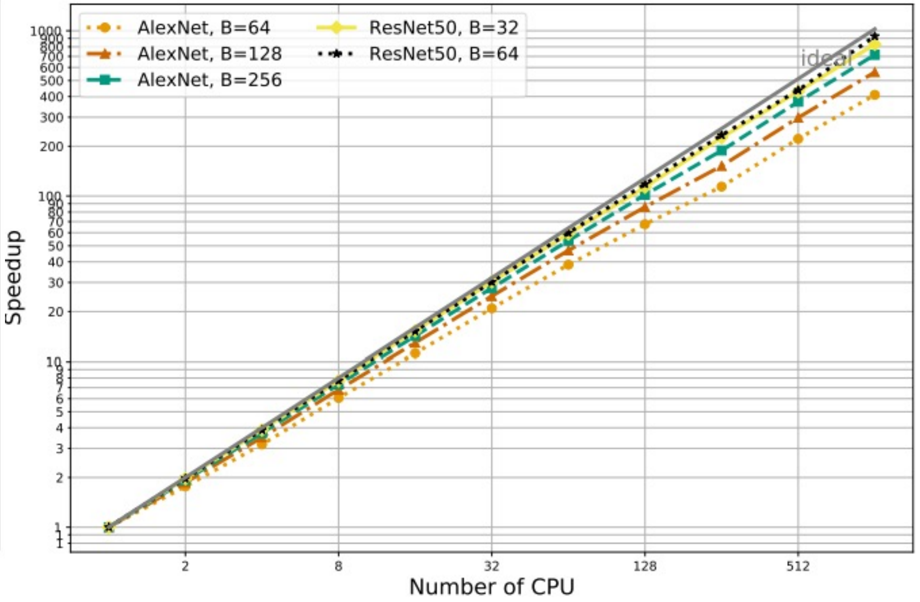
$$t_{new-allgather} = \log p\alpha + (p-\frac{p}{q})\beta_1\frac{n}{p} + (\frac{p}{q}-1)\beta_2\frac{n}{p} \quad (6)$$

Scalability with 1,024 Nodes

	Batch	Speedup		Batch	Speedup
AlexNet	128	561.58	ResNet50	32	828.32
	64	409.50		64	928.15

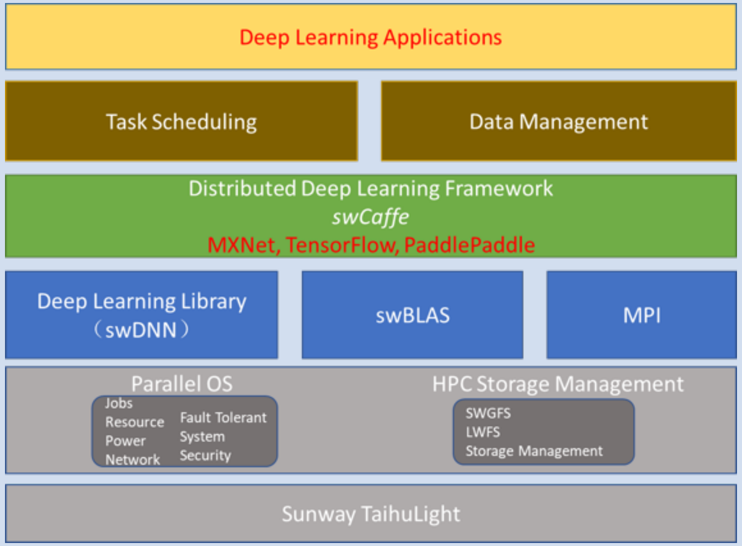
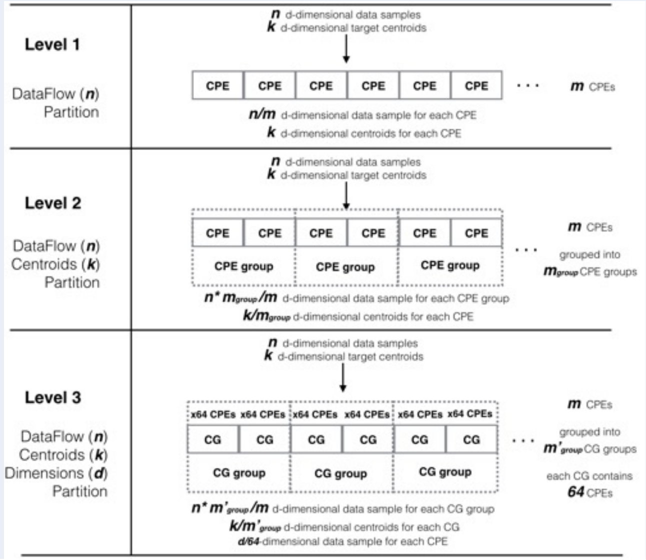


Improvement on MPI Allreduce

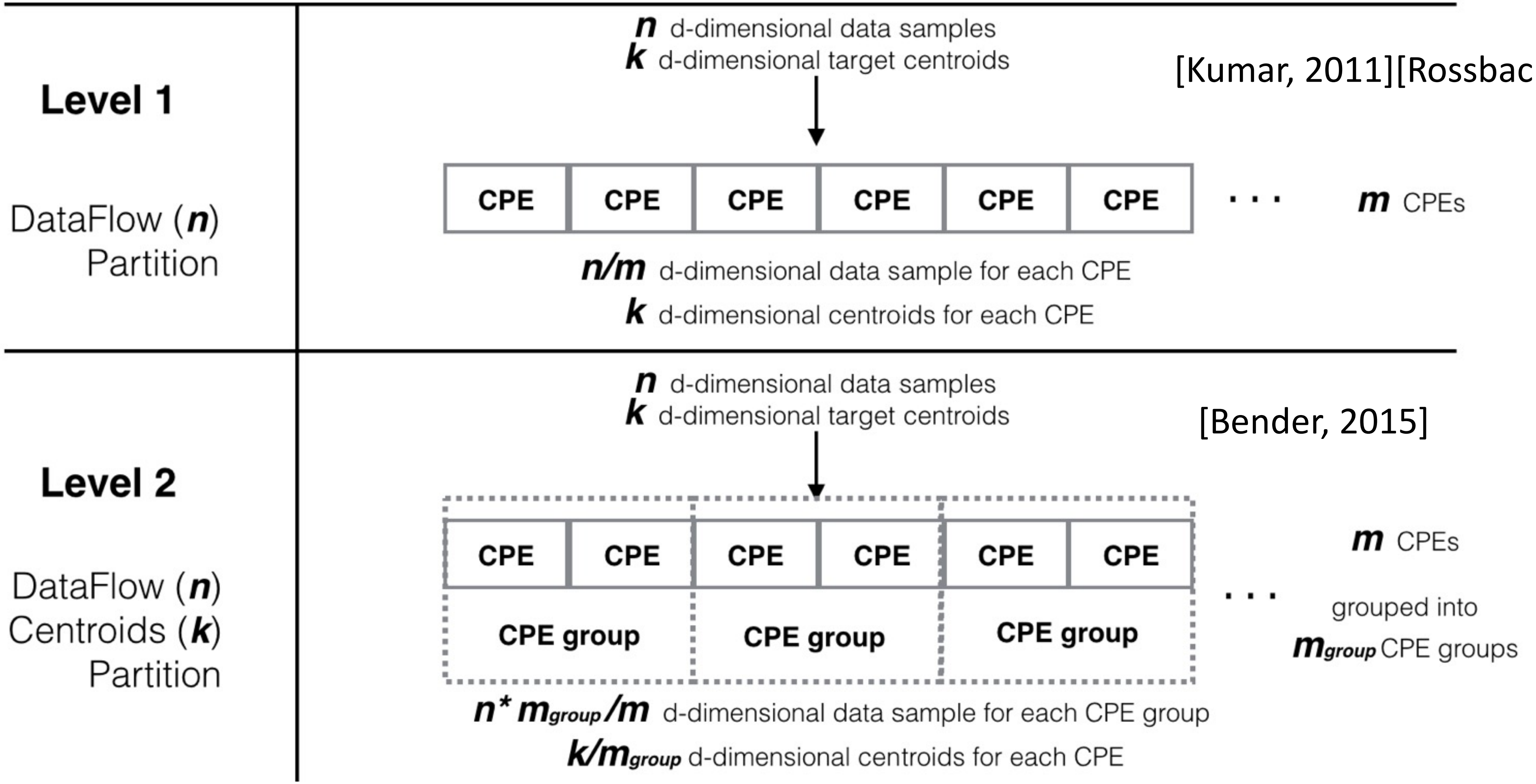


Scalability

清華大學
Tsinghua University
DEPARTMENT OF EARTH SYSTEM SCIENCE,TSINGHUA UNIVERSITY
清华大学地球系统科学系

<table>
<tr><th>Domain</th><th>Software</th><th>Scale</th><th></th></tr>
<tr><td>Machine Learning</td><td>swDNN + swCaffe</td><td>256 to 1,024 nodes</td><td>
</td></tr>
<tr><td></td><td>k-means</td><td>Up to 10,240 nodes (2 million cores)</td><td>
</td></tr>
</table>

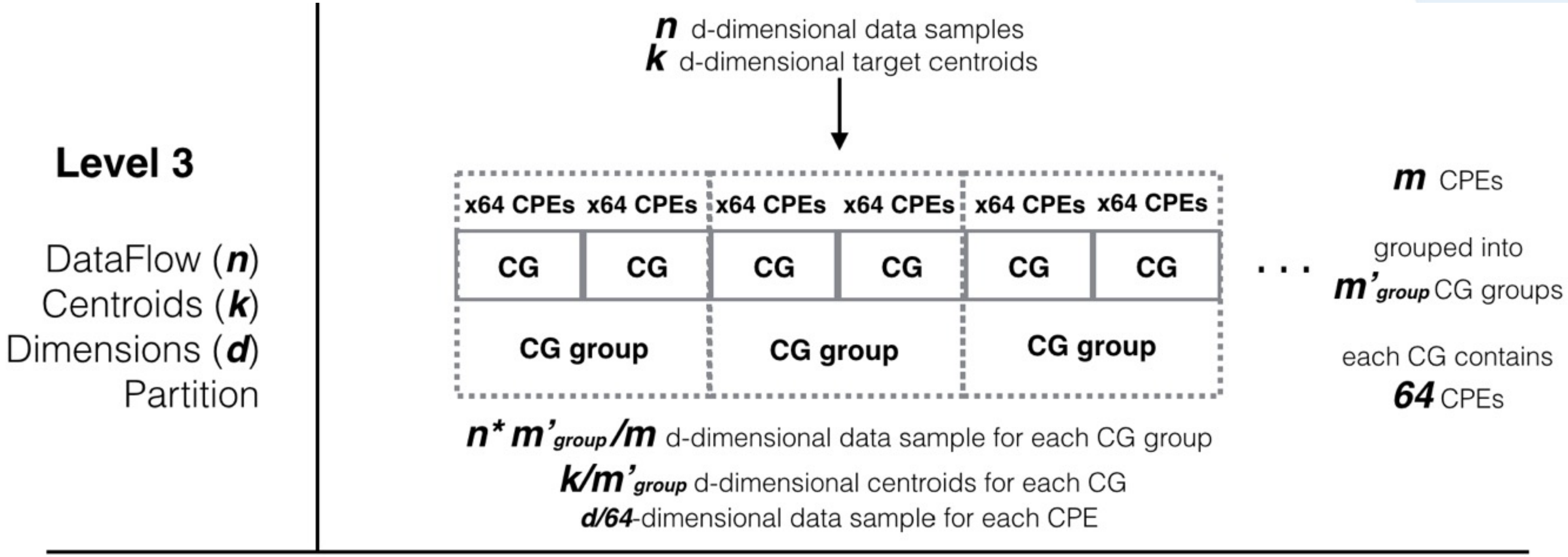
Existing Parallel k-means Designs



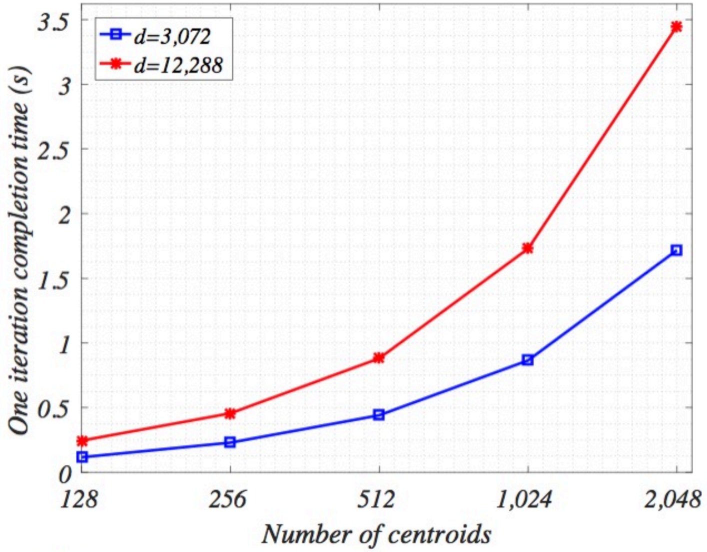
[Kumar, 2011][Rossbach, 2013][Cai, 2015]

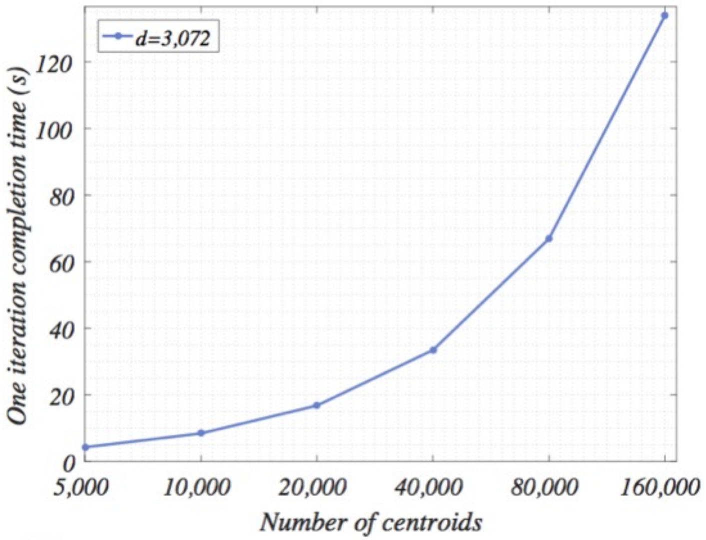
[Bender, 2015]

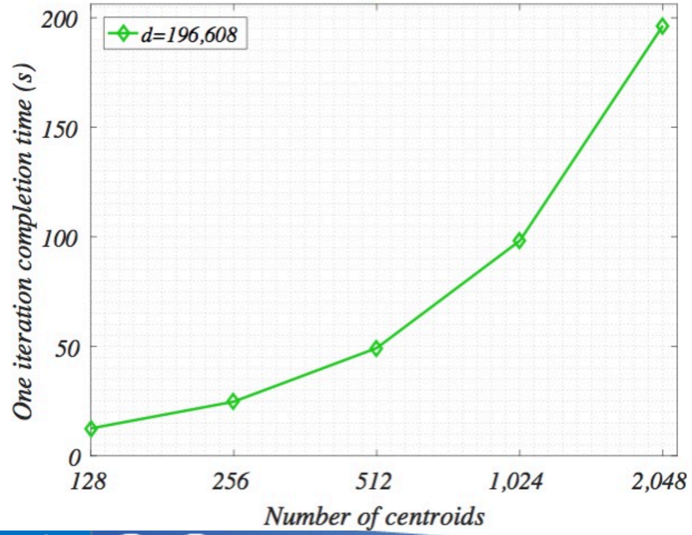
Hierarchical Data Partition for k-means

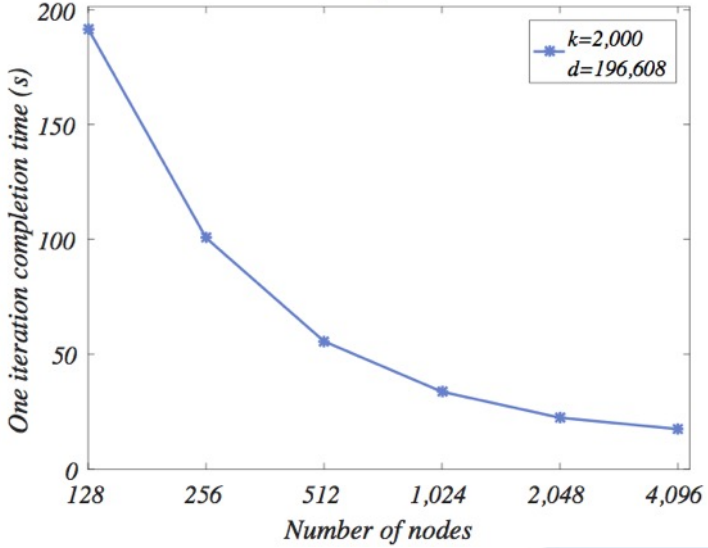


Experimental Results

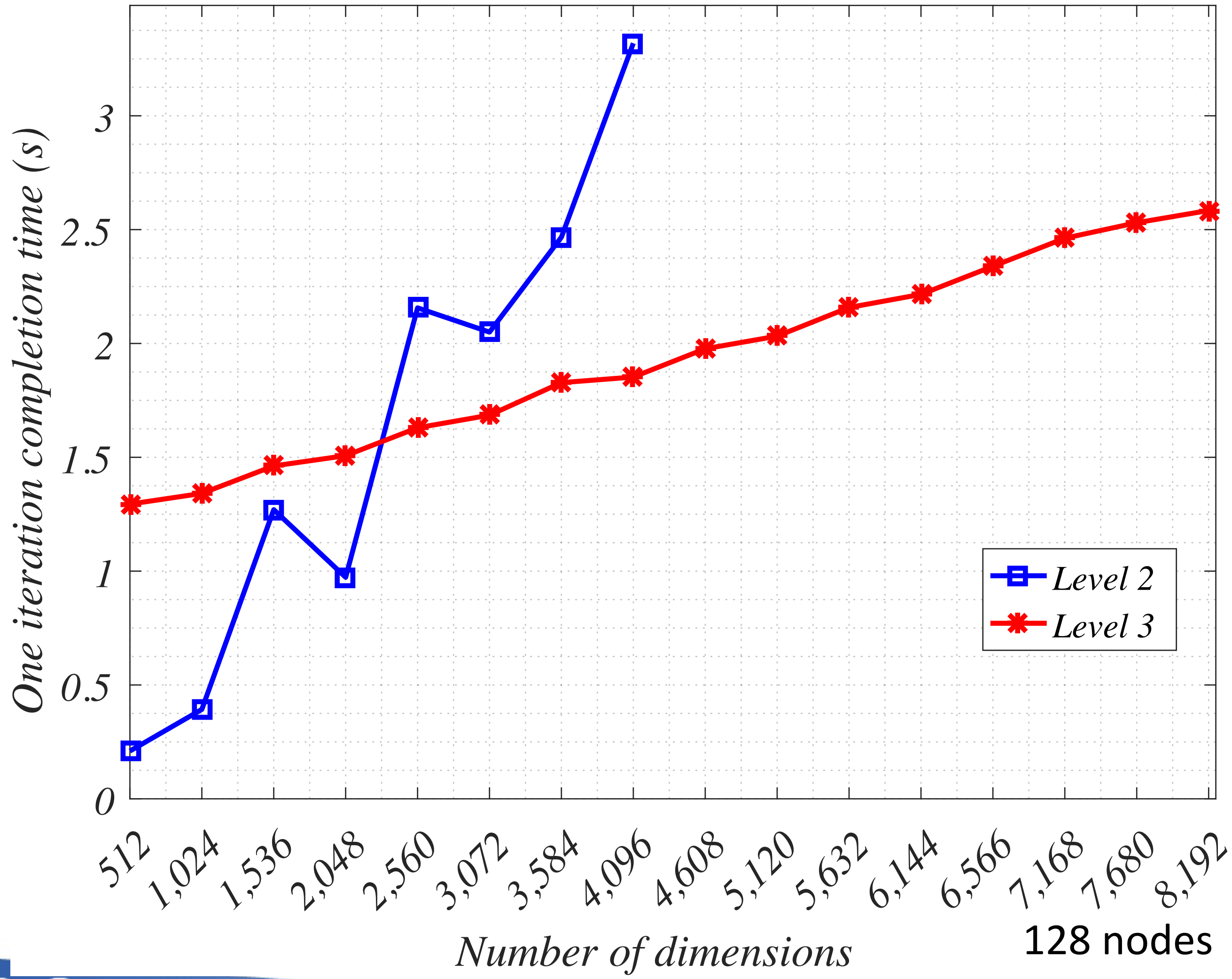








Experimental Results



128 nodes

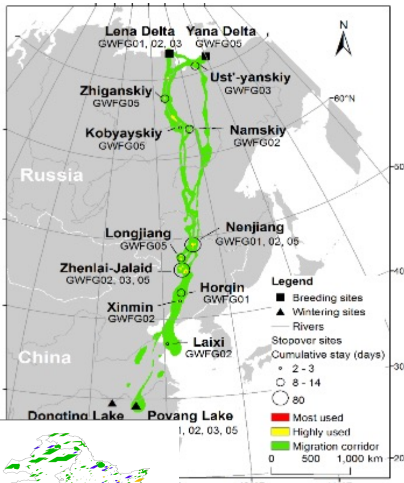
Efforts on Sunway TaihuLight

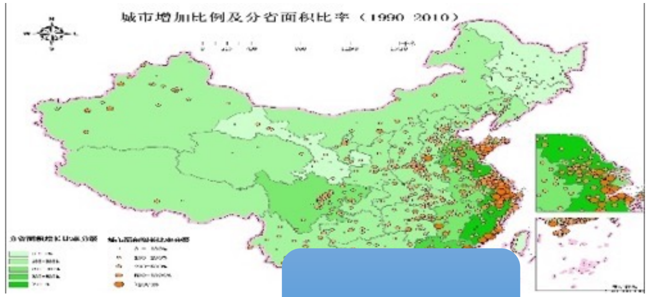
Application

AI-Software

Hardware

Look Ahead: Data-Driven Modeling and Prediction





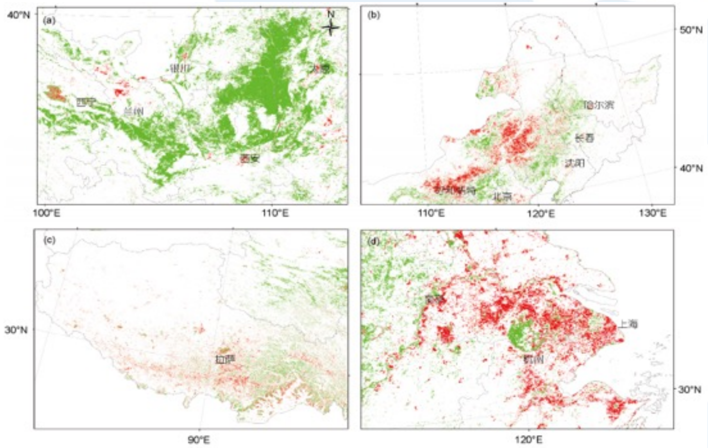
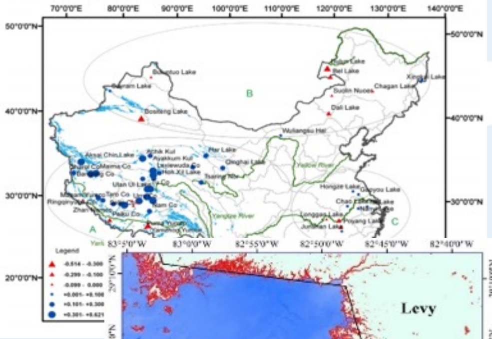
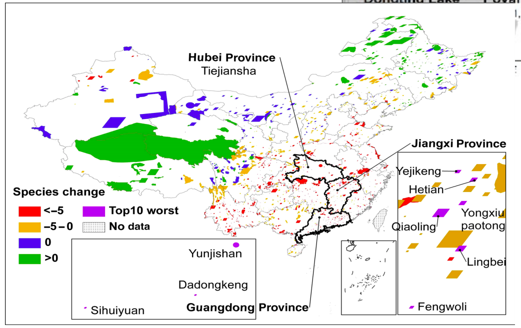
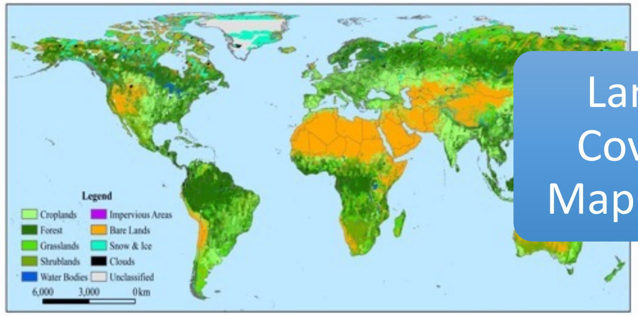


图 3 绿度变化显著的部分区域







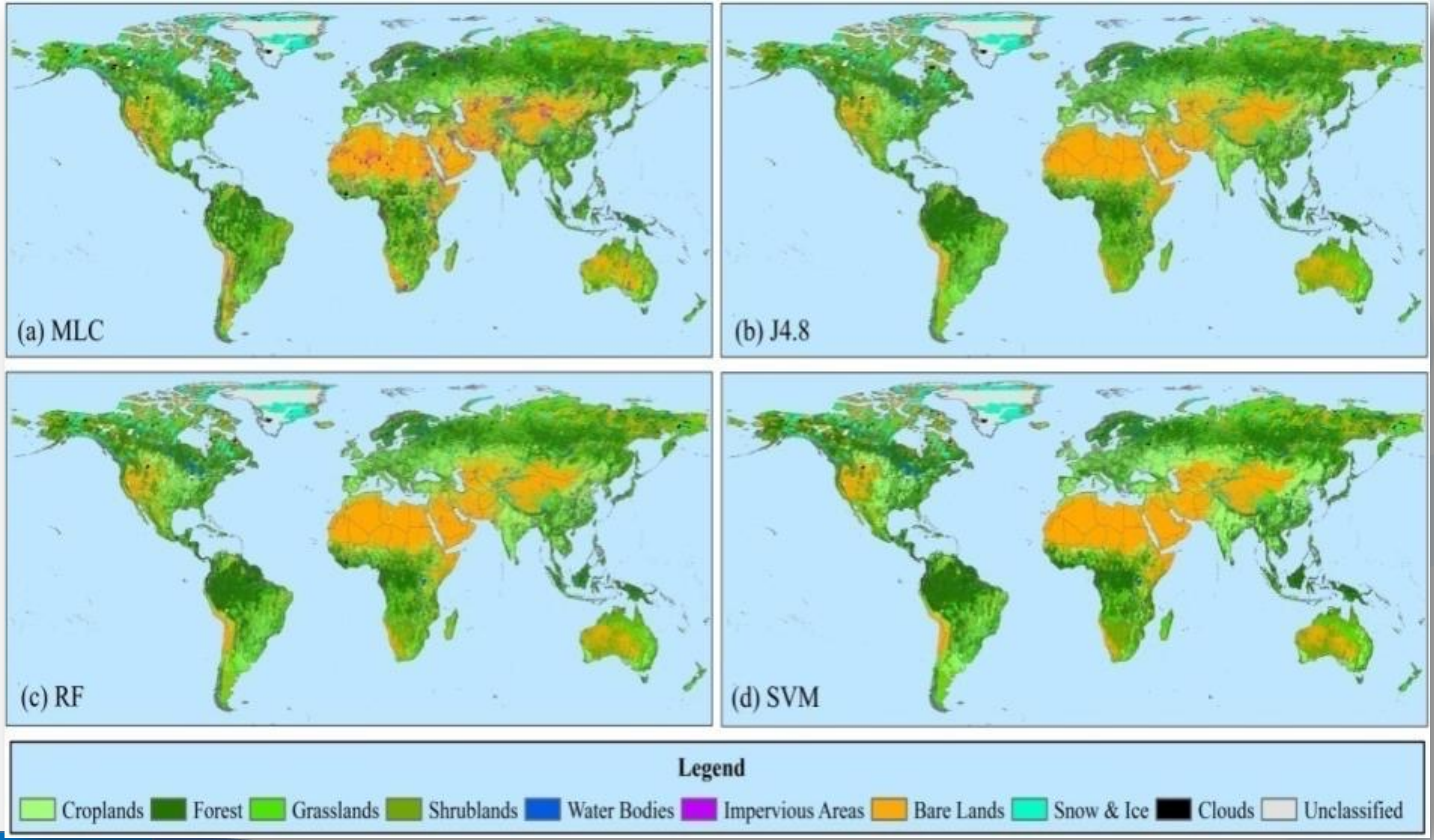
- migration of birds
- city
- green land
- ecology protection zone
- waterbody
- Land Cover Mapping
- sea-level

→ Remote Sensing Data Sets

Potential of data: meter-level resolution, study of specific birds or trees, a huge help for models

Example 1: Global Land Cover Mapping

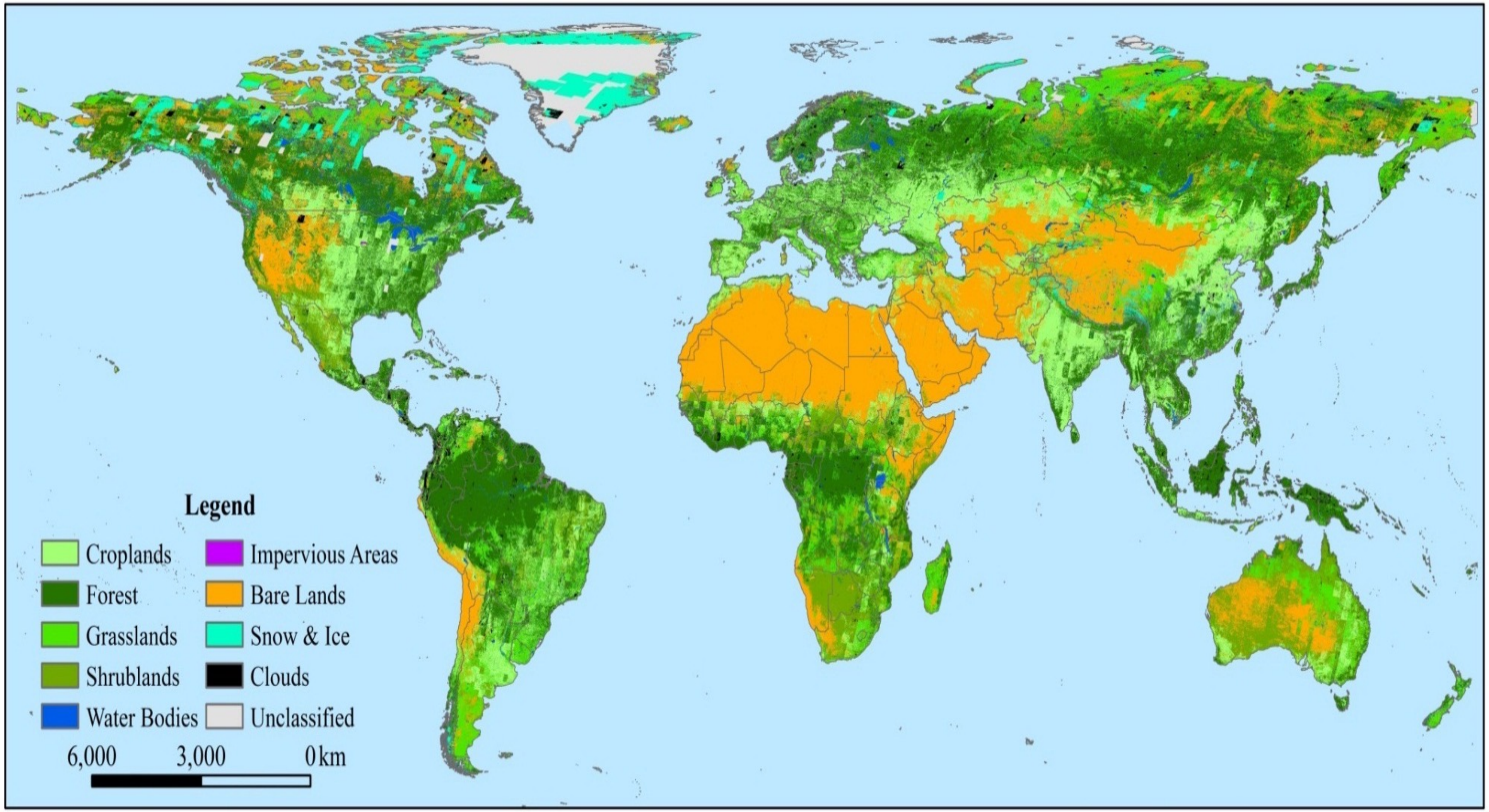
First 30 m resolution global land cover maps



Running SVM on 8900 Landsat scenes on a supercomputer, achieved result in one day.

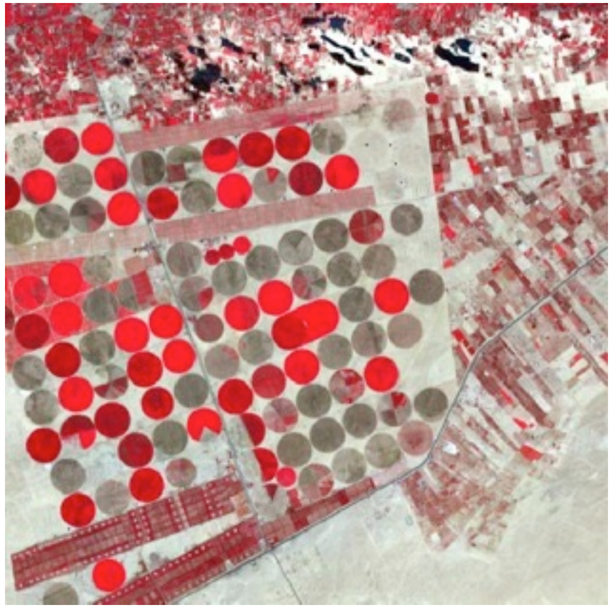
Gong et al., 2013, IJRS

FROM-GLC (Accuracy: 63.72%)

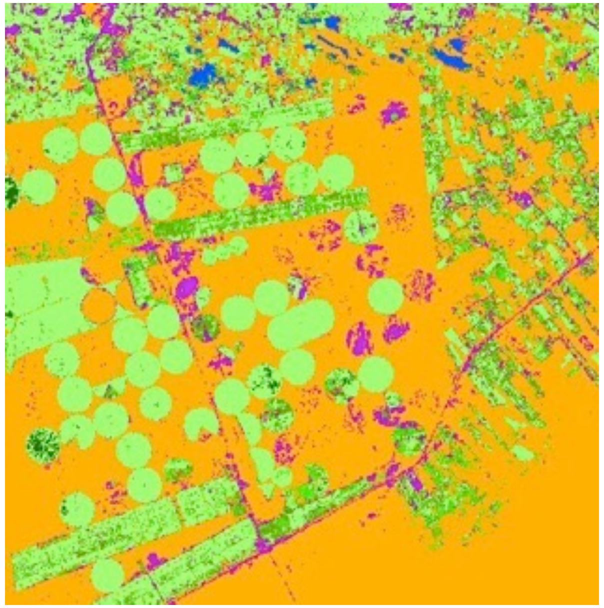


Gong et al., 2013. IJRS

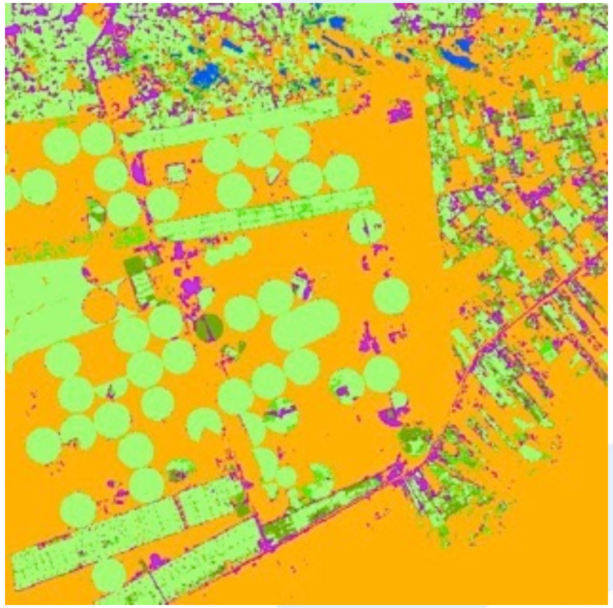
A Starting Point: Direct Application of Stacked AutoEncoder



Landsat Image



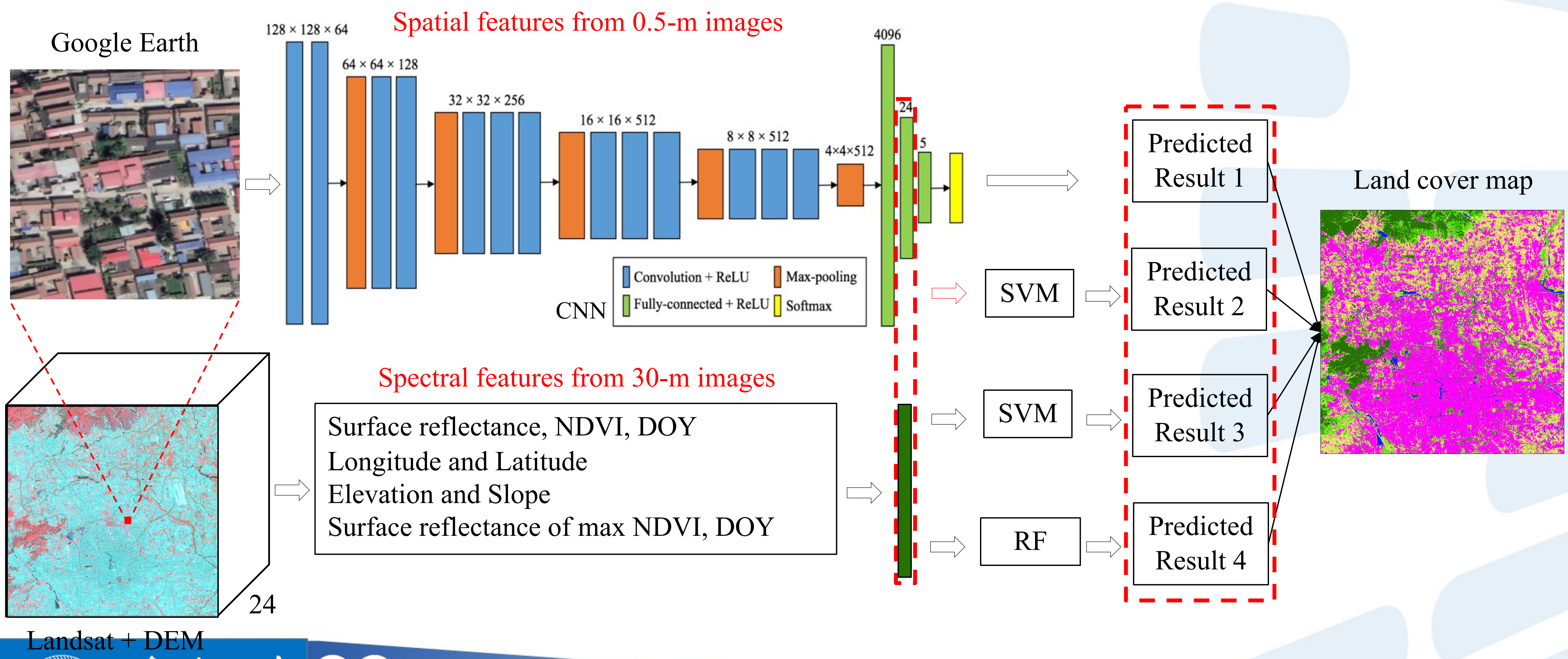
RF Image



SAE Image

	RF	SVM	ANN	SAE
Overall Accuracy	76.03%	77.74%	77.86%	78.99%
Mapping Time	33.605 ± 0.183	16344.188	4.014 ± 0.003	13.250 ± 0.042

Integrating Google Earth Image



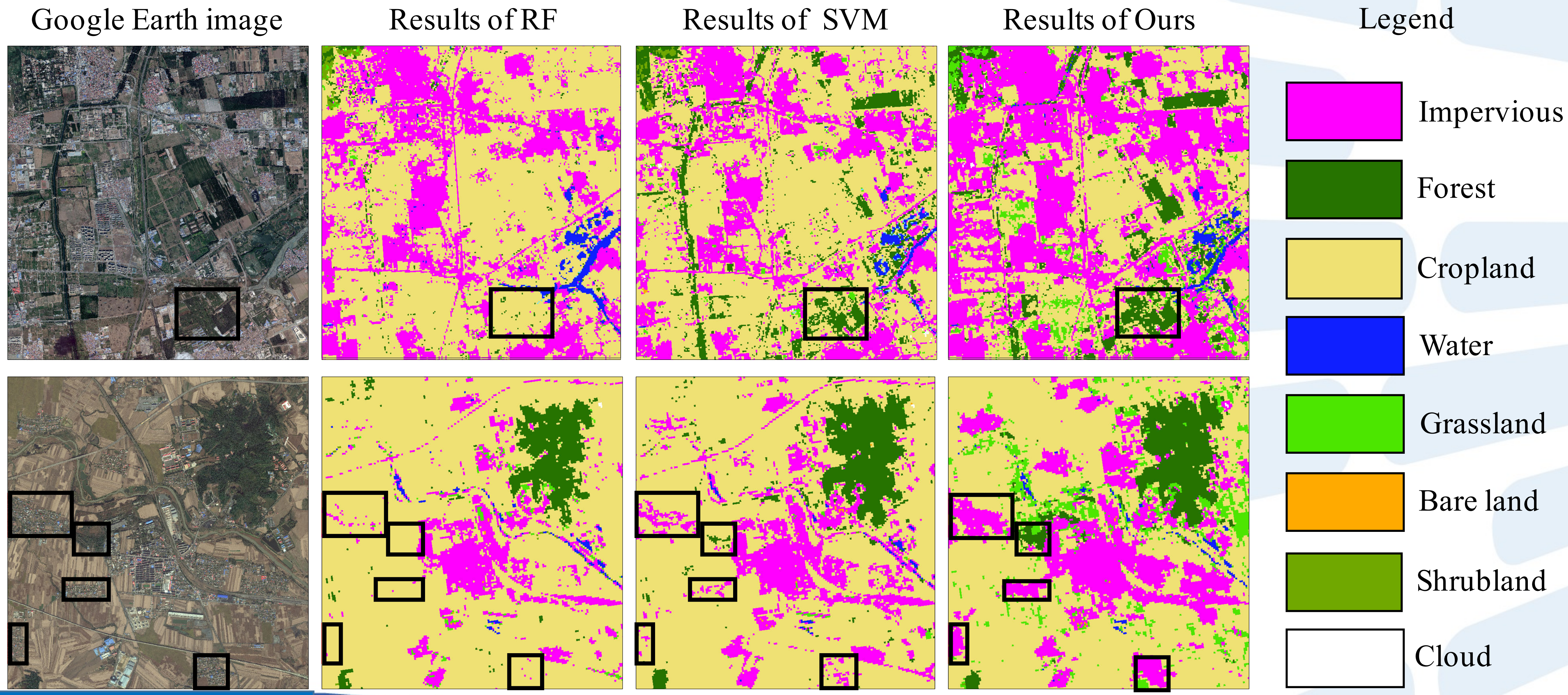
Integrating Google Earth Image

Method	RF			SVM			CNN-30M			CNN-Multi-resolution		
Land cover type	UA	PA	AA	UA	PA	AA	UA	PA	AA	UA	PA	AA
Cropland	75.66	69.73	72.69	78.91	70.52	74.71	79.11	72.31	75.71	80.08	79.97	80.03
Forest	86.69	78.38	82.53	87.13	77.85	82.49	86.94	78.78	82.86	88.14	80.83	84.48
Grassland	64.13	70.48	67.30	61.26	74.08	67.67	63.35	74.11	68.73	77.80	75.62	76.71
Shrubland	5.14	29.51	17.33	3.43	24.00	13.71	3.71	26.00	14.86	11.71	30.60	21.16
Wetland	3.77	33.33	18.55	11.32	42.86	27.09	9.43	71.43	40.43	7.55	80.00	43.77
Water	94.34	73.53	83.93	92.45	73.13	82.79	97.17	74.10	85.64	93.40	76.74	85.07
Tundra	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Impervious	52.58	53.33	52.96	62.44	62.15	62.30	62.91	68.02	65.47	84.51	67.42	75.96
Bare land	95.35	89.28	92.32	94.86	88.12	91.49	96.65	88.32	92.48	96.11	96.11	96.11
Snow/Ice	88.77	93.47	91.12	90.55	92.04	91.29	92.17	92.58	92.38	91.88	91.88	91.88
Cloud	92.09	90.16	91.12	90.95	89.49	90.22	92.71	91.68	92.20	92.21	91.52	91.87
OA (%)		79.90			80.20			81.31			84.40	

Slight increase using
30-meter resolution images

Great increase using
Multi-resolution images

Integrating Google Earth Image



Fewer confusions among various land cover types in the results of our proposed method.

30m to 10m

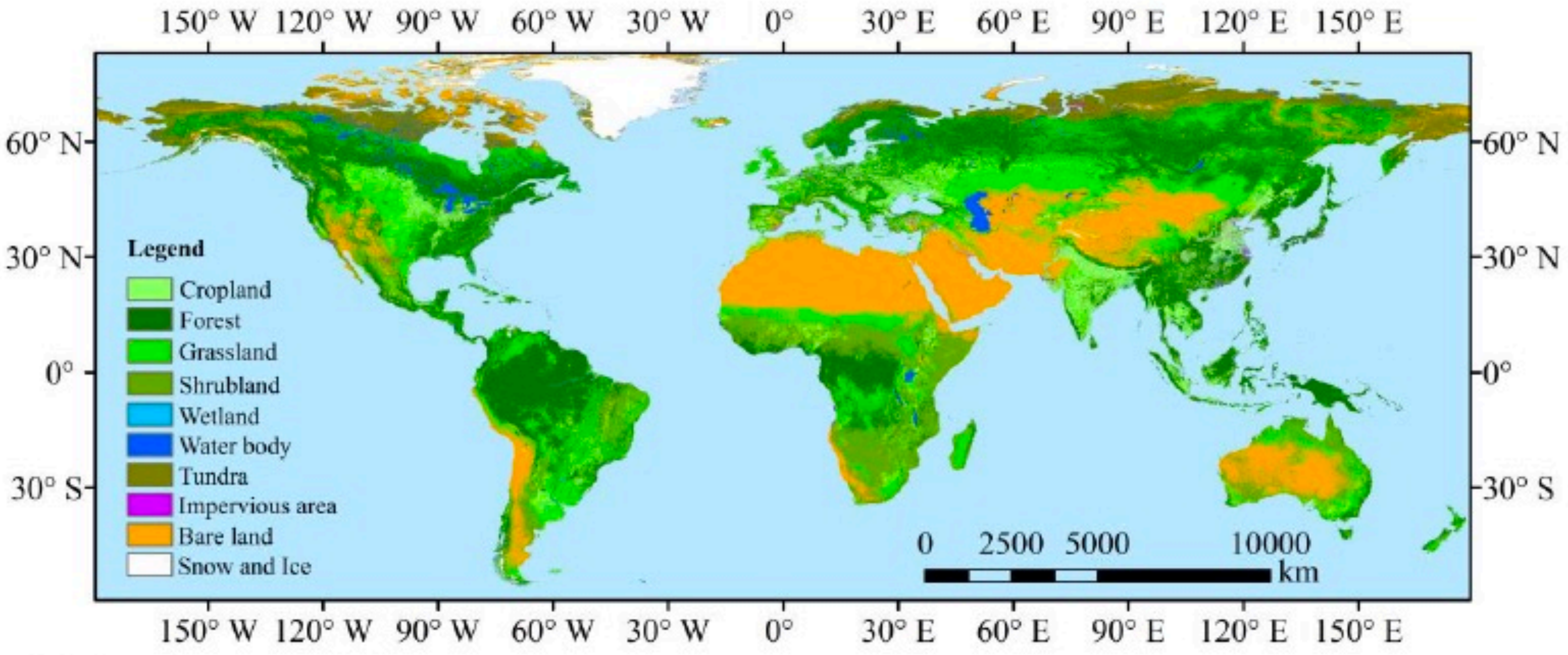
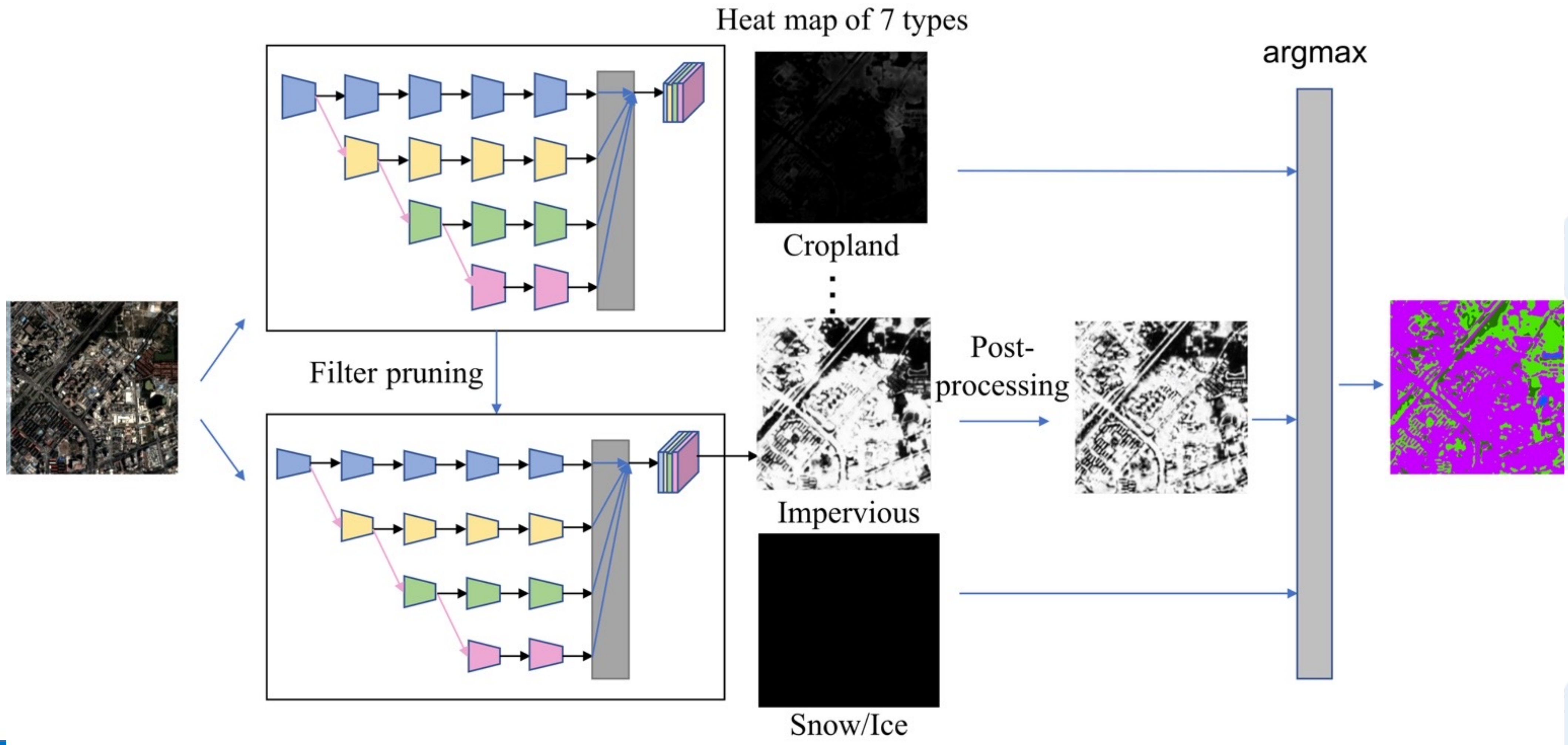


Fig. 2. Global land cover map, FROM-GLC10, based on 10 m resolution Sentinel-2 data acquired in 2017

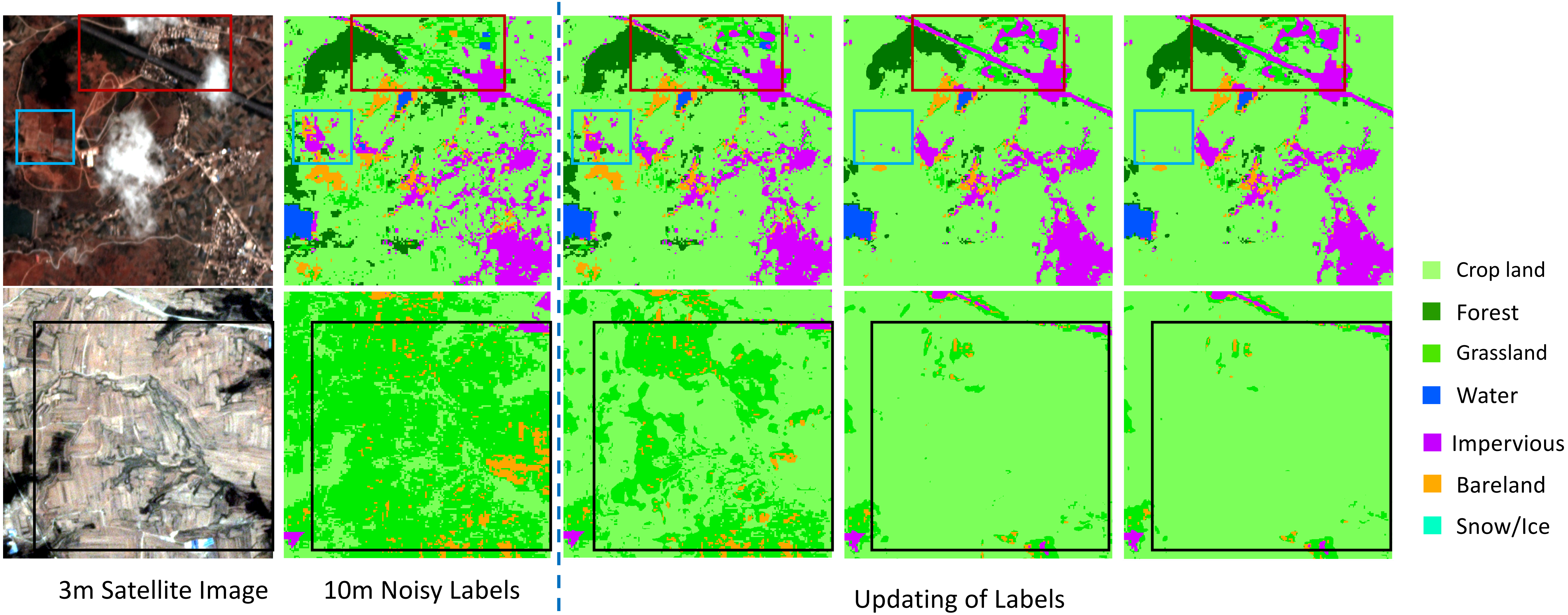
Gong P., et al., 2019. Stable classification with limited sample: transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017,Science Bulletin.

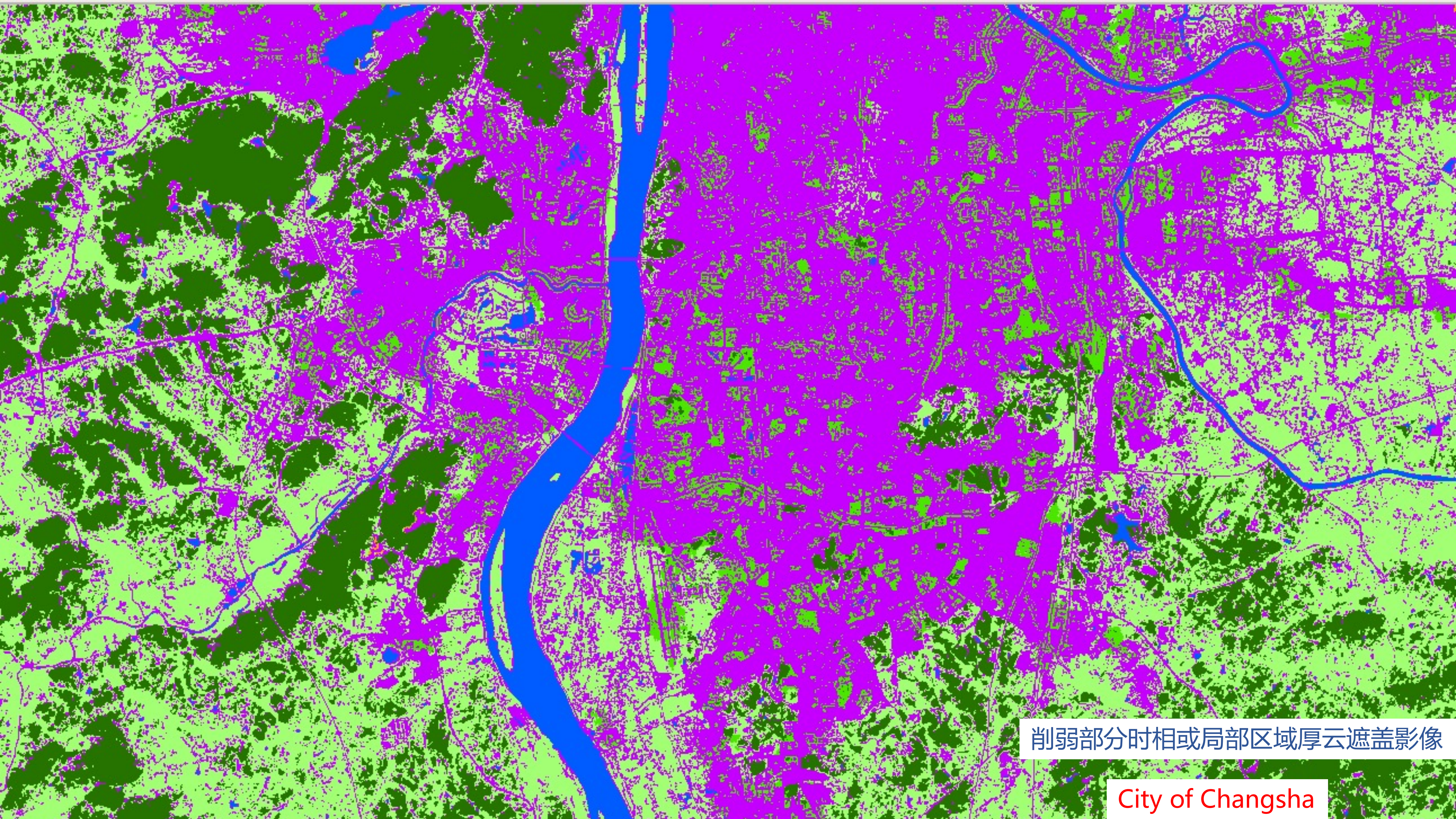
10m to 3m



Heat map of 7 types
argmax
Cropland
Filter pruning
Post-
processing
Impervious
Snow/Ice

Update of Labels

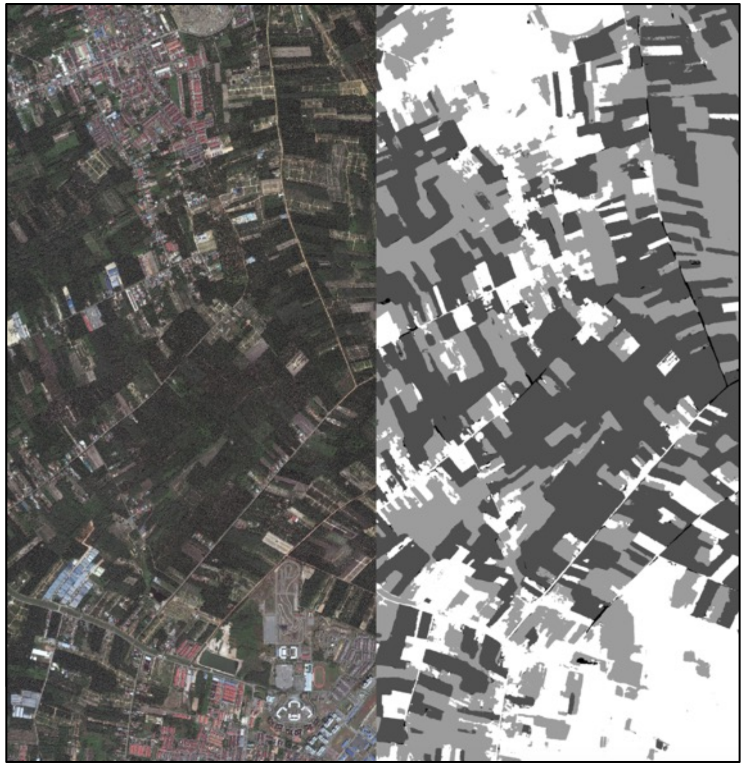



削弱部分时相或局部区域厚云遮盖影像
City of Changsha

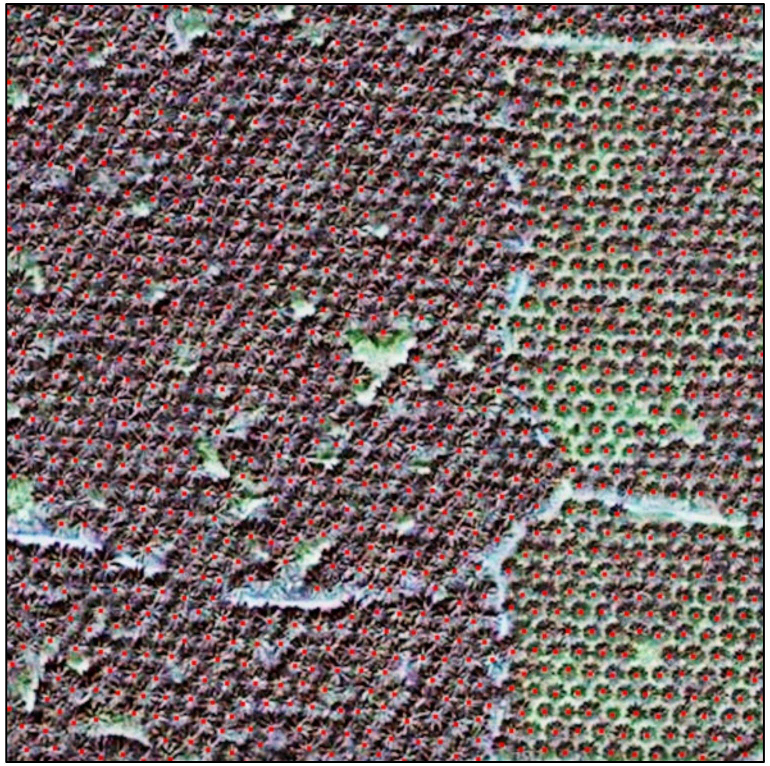
Example 2: Detection of Oil Palm Trees

Case 2: Intelligent monitoring of oil palm trees

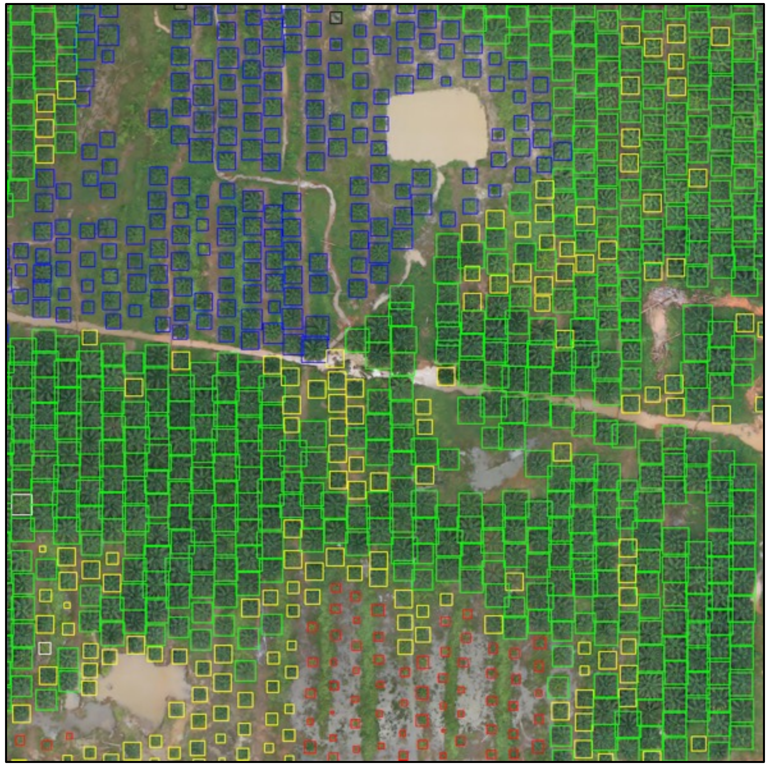
Mapping of oil palm trees using high-resolution satellite images



Detection of oil palm trees using high-resolution satellite images



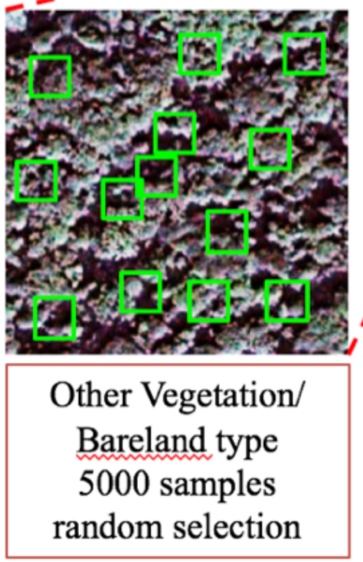
Detection and classification of oil palm trees using UAV images

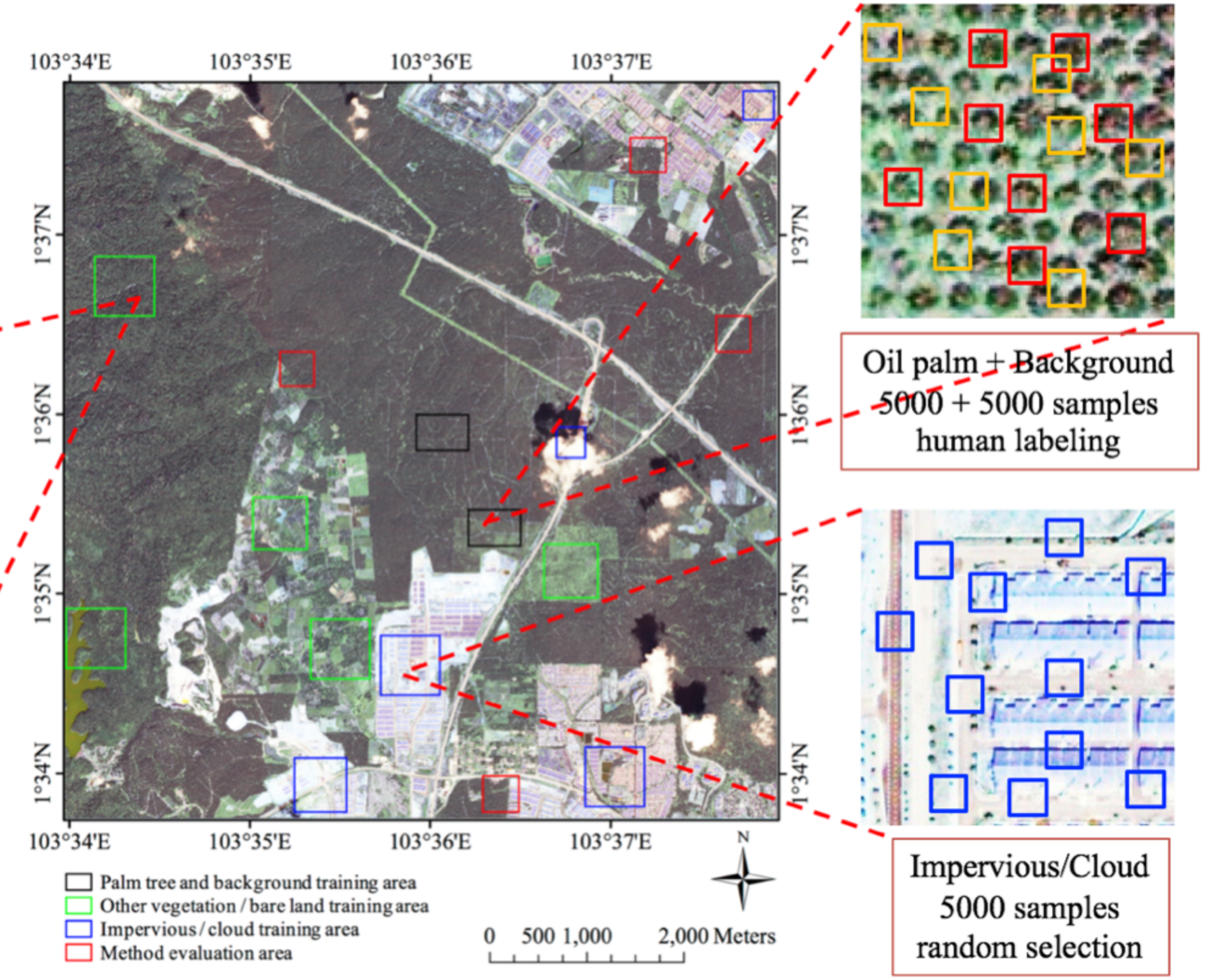


CNN based large-scale oil palm tree detection

Demand for oil palm

- The oil palm is the most rapidly expanding crops in tropical countries
- The palm oil is the most consuming vegetable oil in the world (35%)

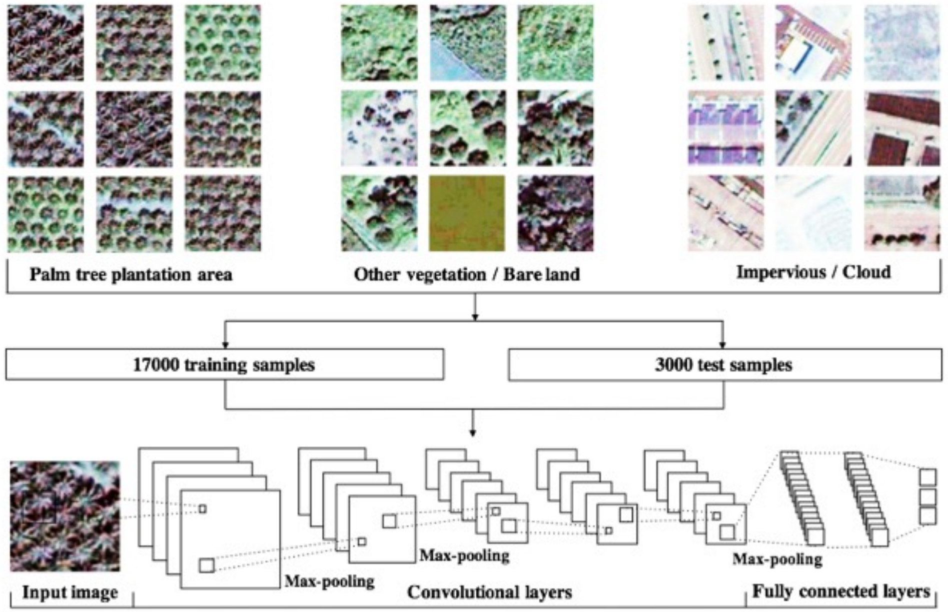




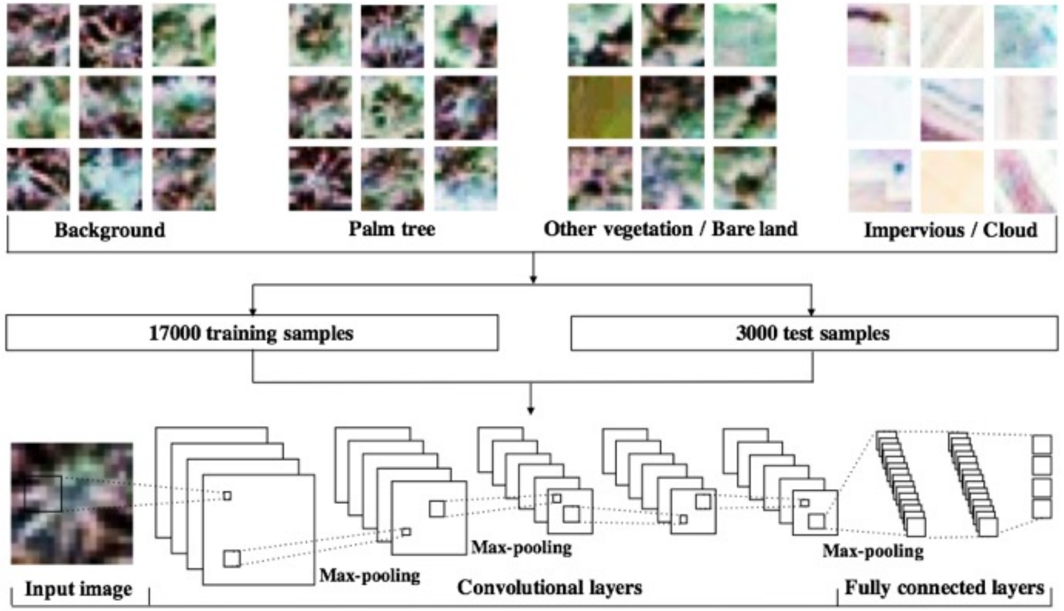
CNN based large-scale oil palm tree detection

Multi-level CNN training and optimization

- The first CNN is used for land cover classification to locate the oil palm plantation area, including three types of samples (oil palm plantation area, other vegetation / bare land, and impervious/cloud).
- The second CNN is used for object classification to identify the oil palms, including for types of samples (oil palm, background, other vegetation / bare land, and impervious/cloud).
- The two CNNs are trained and optimized independently based on 17,000 training samples and 3000 validation samples.



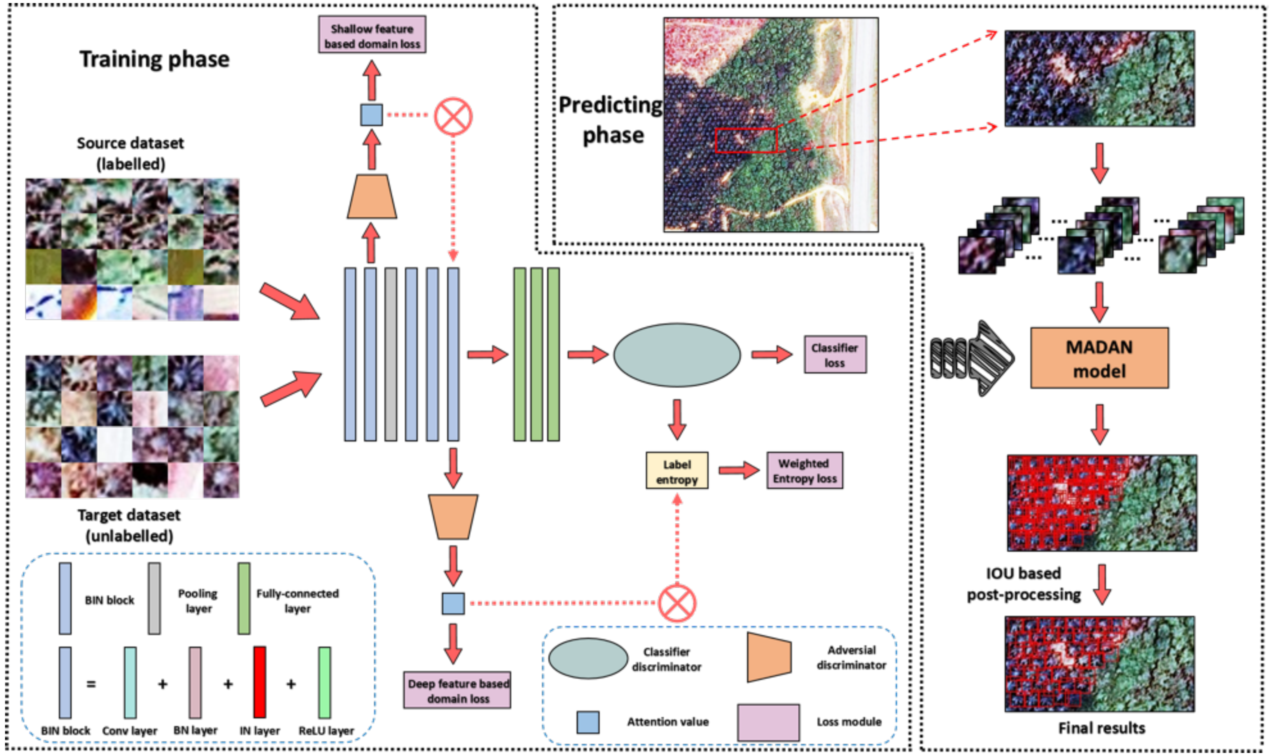
CNN-1: Land cover classification



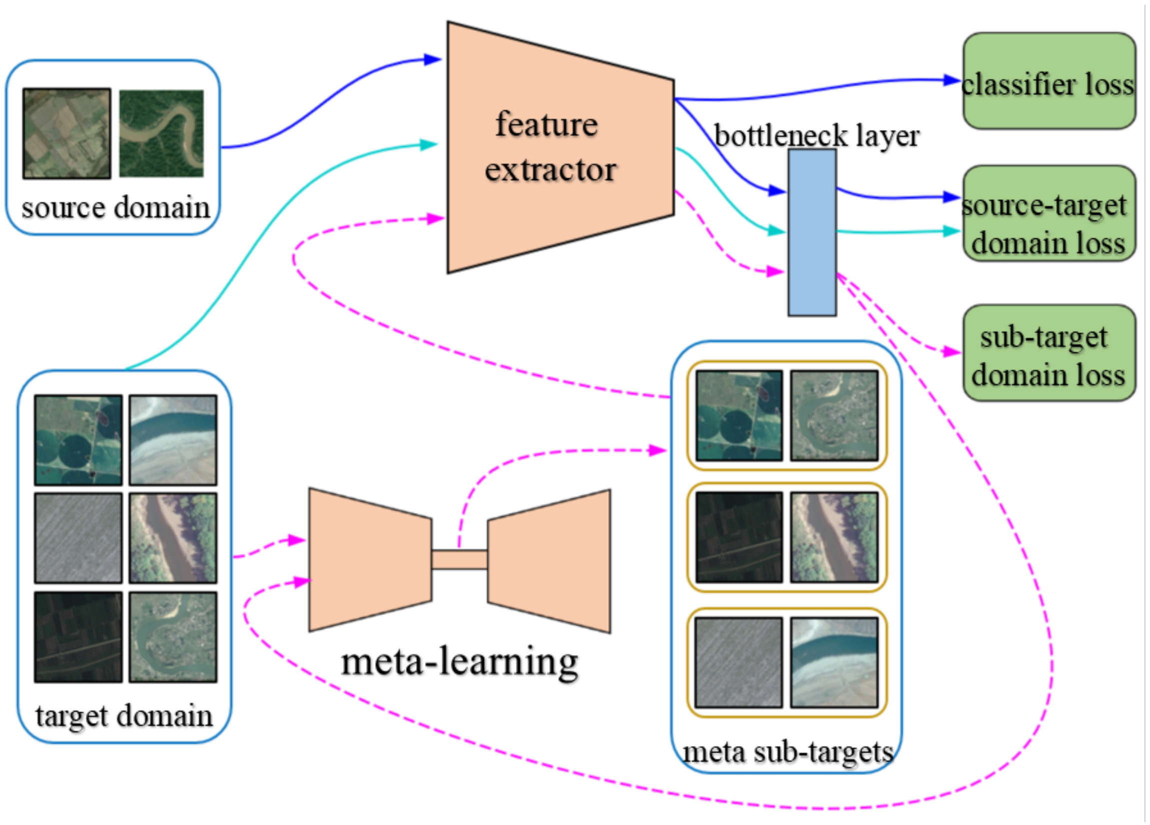
CNN-2: Object classification

Transfer Learning

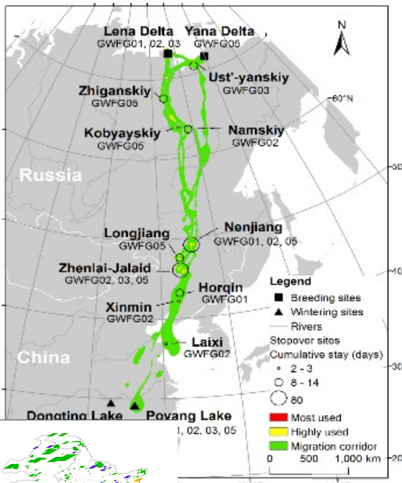
One Source Domain to One Target Domain

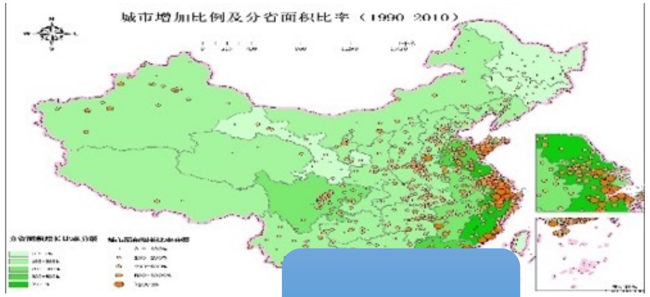


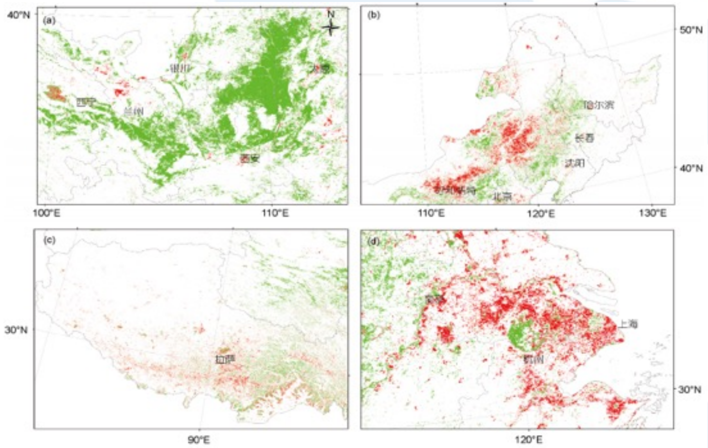
One Source Domain to Multi Target Domain

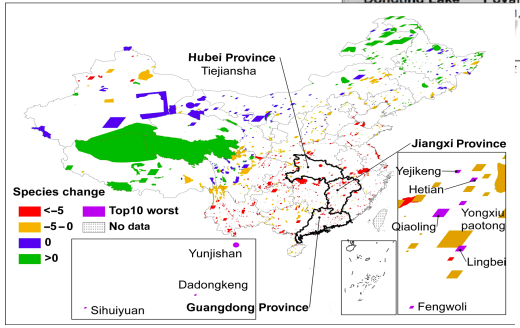


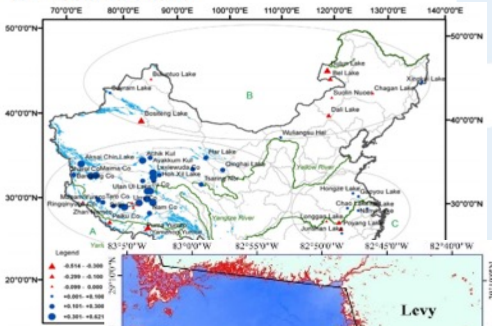
Look Ahead: Data-Driven Modeling and Prediction

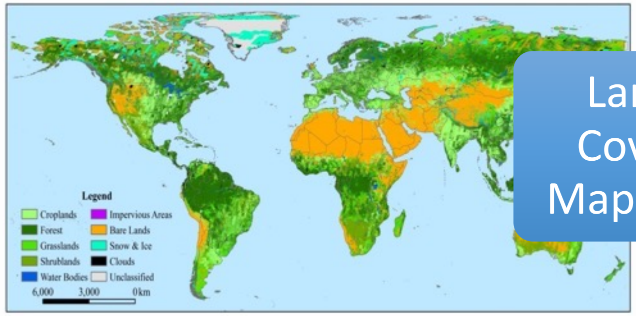












- migration of birds
- city
- green land
- ecology protection zone
- waterbody
- Land Cover Mapping
- sea-level

→ **Remote Sensing Data Sets**

Potential of data: meter-level resolution, study of specific birds or trees, a huge help for models