Supplementary materials

1. Introduction of BDF-Net

The model structure, as shown in Figure S1. Irregular parcels with varying shapes and sizes are transformed into multiple fixed-sized parcels using Poisson disk sampling. The deep semantic information of the fixed-sized parcels is then extracted by a remote sensing feature module, which is based on the Swin-Transformer (Liu et al. 2021). Word2Vec (Mikolov et al. 2013) for POI embedding (Yao et al. 2017) is employed to extract semantic information from the POI (Point of Interest) data. Finally, the model combines the POI and remote sensing imagery feature vectors through weighted fusion and feeds them into fully connected layers and a SoftMax layer for classification, resulting in the final output.

The experiment was conducted on an A4000 GPU using the PyTorch framework and Python 3.10. The hyperparameters used were: learning rate of 0.001, learning rate decay coefficient of 0.01, 300 epochs, and a batch size of 8. The Adam optimizer and the Cross Entropy loss function were employed as the optimizer and loss function, respectively.

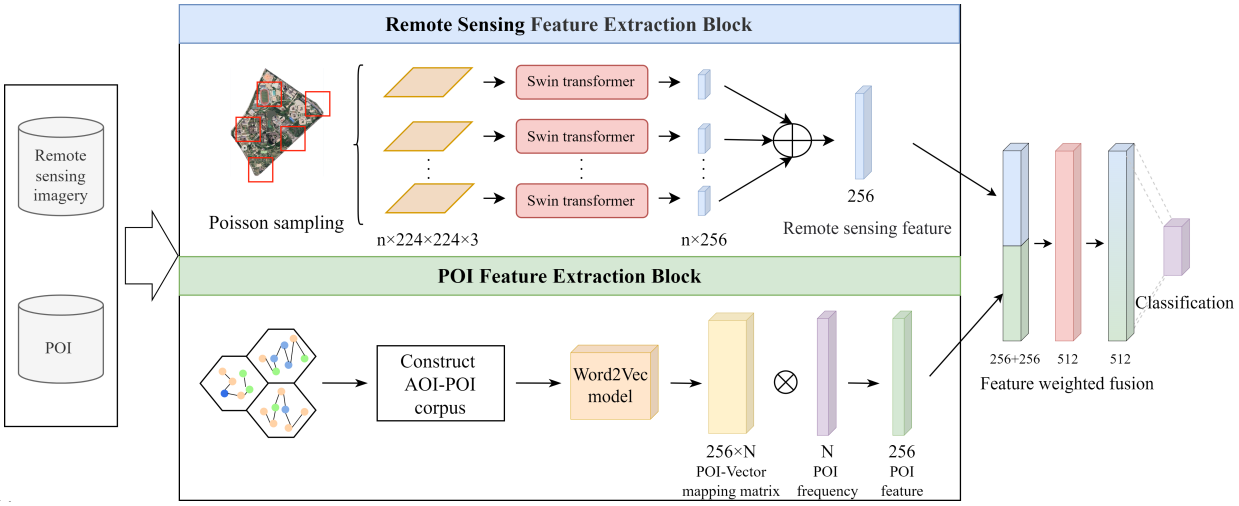

Figure S1. Structure diagram for BDF-Net.

2. Efficiency of BDF-Net

Table S1. Results of ablation analysis experiments for BDF-Net.

Model name	Sampling methods for remote sensing images			Data used for feature extraction		Models used for remote sensing feature extraction		Evaluation indices	
	Poisson sampling	random sampling	non-sampling	RSI	POI	Swin-T	ResNet	OA	Kappa
BDF-Net	√			√	√	√		0.881	0.878
R-BDF-Net		√		√	√	√		0.858	0.857
N-BDF-Net			√	√	√	√		0.632	0.619
Rs-Swin-Net	√			√		√		0.778	0.716
POI2Vec-Net	√				√			0.501	0.432
BDF-ResNet	√			√	√		√	0.787	0.734

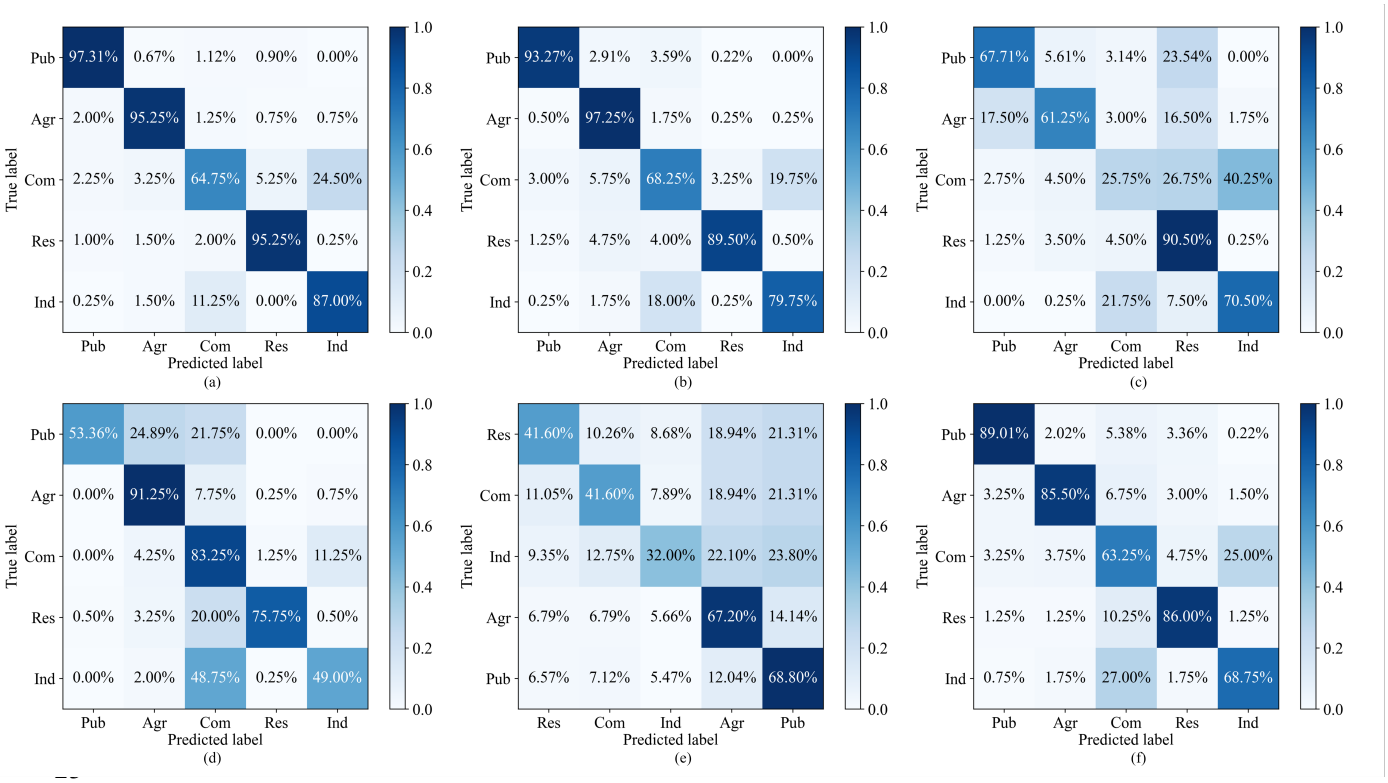

Figure S2. Confusion matrix plots for models in ablation analysis experiments. (a) BDF-Net, (b) R-BDF-Net, (c) N-BDF-Net, (c) Rs-Swin-Net, (e) POI2Vec-Net, (f) BDF-ResNet.

3. Supplementary data for the study area

Table S2. 81 representative cities in the study area and their administrative levels

Number	Province or autonomous region	City	Administrative level
1	-	Beijing	municipality
2	-	Shanghai	municipality
3	-	Tianjin	municipality
4	-	Chongqing	municipality
5	Anhui	Anqing	prefecture-level region
6	Anhui	Chizhou	prefecture-level
7	Anhui	Chuzhou	prefecture-level
8	Anhui	Hefei	provincial capital/prefecture-level region
9	Anhui	Maanshan	prefecture-level region
10	Anhui	Tongling	prefecture-level region
11	Anhui	Wuhu	prefecture-level region
12	Anhui	Xuancheng	prefecture-level region
13	Fujian	Fuzhou	provincial capital/prefecture-level region
14	Fujian	Quanzhou	prefecture-level region
15	Fujian	Xiamen	municipality with independent planning status
16	Gansu	Lanzhou	prefecture-level region
17	Guangdong	Dongguan	prefecture-level region
18	Guangdong	Foshan	prefecture-level region
19	Guangdong	Guangzhou	provincial capital/prefecture-level region
20	Guangdong	Huizhou	prefecture-level region
21	Guangdong	Jiangmen	prefecture-level region
22	Guangdong	Shenzhen	municipality with independent planning status
23	Guangdong	Zhaoqing	prefecture-level region
24	Guangdong	Zhongshan	prefecture-level region
25	Guangdong	Zhuhai	prefecture-level region
26	Guangxi Zhuang Autonomous Region	Nanning	provincial capital/prefecture-level region
27	Guizhou	Guiyang	provincial capital/prefecture-level region

Number	**Province or autonomous region**	**City**	Administrative level
28	Hainan	Haikou	provincial capital/prefecture-level region
29	Hebei	Baiding	prefecture-level region
30	Hebei	Hengshui	prefecture-level region
31	Hebei	Langfang	prefecture-level region
32	Hebei	Qinhuangdap	prefecture-level region
33	Hebei	Shijiazhuang	provincial capital/prefecture-level region
34	Hebei	Tangshan	prefecture-level region
35	Hebei	Zhangjiakou	prefecture-level region
36	Henan	Luoyang	prefecture-level region
37	Henan	Zhengzhou	provincial capital/prefecture-level region
38	Heilongjiang	Haerbin	provincial capital/prefecture-level region
39	Hubei	Wuhan	provincial capital/prefecture-level region
40	Hunan	Xiangtan	prefecture-level region
41	Hunan	Changsha	provincial capital/prefecture-level region
42	Hunan	Zhuzhou	prefecture-level region
43	Jilin	Changchun	provincial capital/prefecture-level region
44	Jiangsu	Changzhou	prefecture-level region
45	Jiangsu	Nanjing	provincial capital/prefecture-level region
46	Jiangsu	Nantong	prefecture-level region
47	Jiangsu	Suzhou	prefecture-level region
48	Jiangsu	Taizhou	prefecture-level region
49	Jiangsu	Wuxi	prefecture-level region
50	Jiangsu	Xuzhou	prefecture-level region
51	Jiangsu	Yancheng	prefecture-level region
52	Jiangsu	Yangzhou	prefecture-level region
53	Jiangsu	Zhenjiang	prefecture-level region
54	Jiangxi	Nanchang	provincial capital/prefecture-level region
55	Liaoning	Dalian	municipality with independent planning status
56	Liaoning	Shenyang	provincial capital/prefecture-level region
57	Inner Mongolia Autonomous Region	Hohhot	provincial capital/prefecture-level region

Number	Province or autonomous region	City	Administrative level
58	Ningxia Hui Autonomous Region	Yinchuan	provincial capital/prefecture-level region
59	Qinghai	Xining	provincial capital/prefecture-level region
60	Shandong	Dongying	prefecture-level region
61	Shandong	Jinan	provincial capital/prefecture-level region
62	Shandong	Jining	prefecture-level region
63	Shandong	Linyi	prefecture-level region
64	Shandong	Qingdao	municipality with independent planning status
65	Shandong	Weifang	prefecture-level region
66	Shandong	Yantai	prefecture-level region
67	Shandong	Zibo	prefecture-level region
68	Shanxi	Taiyuan	prefecture-level region
69	Shaanxi	Xian	provincial capital/prefecture-level region
70	Sichuan	Chengdu	provincial capital/prefecture-level region
71	Tibet Autonomous Region	Lhasa	provincial capital/prefecture-level region
72	Xinjiang Uyghur Autonomous Region	Urumqi	provincial capital/prefecture-level region
73	Yunnan	Kunming	provincial capital/prefecture-level region
74	Zhejiang	Hangzhou	provincial capital/prefecture-level region
75	Zhejiang	Huzhou	prefecture-level region
76	Zhejiang	Jiaxing	prefecture-level region
77	Zhejiang	Jinhua	prefecture-level region
78	Zhejiang	Ningbo	municipality with independent planning status
79	Zhejiang	Shaoxing	prefecture-level region
80	Zhejiang	Taizhou	prefecture-level region
81	Zhejiang	Zhoushan	prefecture-level region

4. Data for filtering samples based on size and location of parcel

Table S3. Statistical data table for parcels grouped by area.

Group number	Area range (m^2)		Percentage of quantities in each land use category					Results of evaluation indices			
	Minimum area	Maximum area	Agr	Com	Ind	Res	Pub	*Gini*	P_{nm}	P_{iu}	S_{avg}
1	598.34	38931.01	55.32%	10.20%	5.33%	23.91%	5.25%	0.621	0.954	0.835	0.722
2	38931.31	51408.89	25.29%	15.45%	13.63%	36.86%	8.77%	0.750	0.883	0.949	0.764
3	51410.46	61480.50	17.27%	14.59%	17.48%	40.45%	10.22%	0.744	0.859	0.945	0.691
4	61481.59	71594.68	15.24%	14.23%	20.61%	40.54%	9.39%	0.741	0.852	0.968	0.677
5	71595.33	82294.70	14.97%	12.22%	22.80%	40.26%	9.76%	0.739	0.857	0.961	0.683
6	82295.31	93767.39	15.42%	12.13%	23.54%	40.02%	8.89%	0.738	0.851	0.949	0.661
7	93767.96	106489.90	16.26%	10.24%	25.96%	38.90%	8.64%	0.737	0.852	0.942	0.661
8	106489.92	121832.36	13.92%	10.24%	29.13%	37.41%	9.30%	0.737	0.851	0.923	0.650
9	121832.36	138862.36	17.27%	9.05%	29.85%	36.34%	7.48%	0.735	0.846	0.925	0.637
10	138863.61	160206.25	17.15%	8.26%	33.18%	34.25%	7.15%	0.731	0.853	0.910	0.644
11	160209.72	187310.21	16.48%	7.56%	34.71%	34.68%	6.57%	0.722	0.848	0.880	0.605
12	187325.79	224526.00	17.01%	7.02%	36.30%	32.88%	6.79%	0.722	0.844	0.839	0.579
13	224526.00	272085.10	18.57%	6.38%	38.15%	30.07%	6.83%	0.721	0.848	0.775	0.566
14	272091.88	340493.10	20.79%	5.30%	37.44%	30.02%	6.45%	0.720	0.854	0.742	0.569
15	340512.33	467345.18	22.84%	4.71%	35.99%	29.53%	6.93%	0.724	0.851	0.573	0.508
16	467347.38	676818.47	23.10%	3.63%	30.71%	35.08%	7.48%	0.722	0.851	0.378	0.436
17	676818.47	1112381.54	31.72%	3.60%	22.57%	34.81%	7.31%	0.721	0.876	0.303	0.472
18	1112419.94	1934031.04	30.83%	2.78%	12.85%	47.63%	5.90%	0.657	0.898	0.217	0.409
19	1934166.00	4333233.59	34.28%	1.88%	5.20%	55.73%	2.91%	0.568	0.929	0.088	0.313
20	4334050.63	64690413.16	35.63%	2.25%	1.66%	59.20%	1.27%	0.522	0.970	0.031	0.333

Table S4. Variation of *Gini*, P_{nm}, P_{iu} and mean score $S_{avg}$with dispersion coefficient *d*.

Dispersion coefficient *d*	Evaluation indices			Mean score S_{avg}
	Gini	P_{nm}	P_{iu}	
0	0.000	0.682	0.998	0.335
0.1	0.096	0.681	0.996	0.348
0.2	0.742	0.689	0.996	0.410
0.3	0.299	0.690	0.994	0.480
0.4	0.438	0.696	0.995	0.528
0.5	0.579	0.695	0.989	0.587
0.6	0.860	0.704	0.985	0.713

Dispersion coefficient d	Evaluation indices			Mean score S_{avg}
	Gini	P_{nm}	P_{iu}	
0.7	0.793	0.701	0.965	0.644
0.8	1.000	0.716	0.925	0.752
0.9	0.840	0.717	0.869	0.610
1	0.935	0.727	0.821	0.641

5. CN-MSLU-DEMO Dataset Description

We created the CN-MSLU-DEMO dataset by extracting typical samples from each of the five major categories in CN-MSLU-100K. This allows interested researchers to gain a better understanding of the dataset's characteristics and applicability. In this description, we provide basic information about the dataset, as well as sample code for exploring the data.

5.1. Introduction of CN-MSLU-100K

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 Group's "AMAP POI" classification system, we categorized the land use types of the parcels into five major categories as "Residential Districts (Res)", "Commercial Zones (Com)", "Industrial Land (Ind)", "Public Services (Pub)", "Agriculture and Nature (Agr)". Each major category is subdivided into a total of 22 secondary categories.

In addition, during the labelling process, we also obtained a smaller number of "Transportation Facilities (Tra)", and large amount of "Unknow Landuse (Unk)" categories which are difficult to judge due to insufficient information on land parcels.

These data are also included in the dataset. The final dataset contains 7 categories and 28 sub-categories. Refer to Table S5 for the land use classification system of the parcels in this study and the number of samples in each category.

Table S5. The land use classification system of the parcels in this study and the number of samples in each category.

Category type	First Level Category	Second Level Category	Amount
Major categories	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
		Hotel	160
	Industrial Land (Ind) 24498	Industrial Park and Factory	21593
		Construction Site	2904
	Public Services (Pub) 6286	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
		Wasteland	2458
Additional categories	Transportation Facilities (Tra) 799	Transport facilities (Car Park; Gas Station; Service Station)	290
		Transportation hub (Subway; Bus or Train Station; Airport)	366

Category type	First Level Category	Second Level Category	Amount
	Unknow Landuse (Unk) 25069	Highway & Track	143
		Lack of Information	5753
		Invalid Land Parcel (Small-sized & Narrow)	2776
		Mixed Landuse	16540

5.2. Dataset stats

5.2.1. File directory structure

The description for each folder and file is shown in Table S6.

Table S6. 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 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 organize the information of all the data from the XML files into a CSV table.
CN-MSLU-DEMO.csv		CSV	A CSV table generated using DatasetGenerate.py to store dataset information. Contains categories, file names, storage paths, image widths, image heights, geographic information, first level class names, second level class names for all data.

5.2.2. Demo code

A python script file “DatasetGenerate.py” is provided to read the XML file in a multi-threaded manner. It can extract the descriptive information in the XML files of all remote sensing images, including file name, storage path, land use category, etc.,

and then store this information as a CSV file “CN-MSLU-DEMO.csv”. When using the remote sensing image dataset, you can read the CSV file to perform quick operations on the data.

References

Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 10012-10022.

Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space[J]. arXiv preprint arXiv:1301.3781, 2013.

Yao Y, Li X, Liu X, et al. Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model[J]. International Journal of Geographical Information Science, 2017, 31(4): 825-848.