High-Resolution Land-Cover Mapping Based on a Cross-Resolution Deep Learning Framework and Available Low-Resolution Labels

Xiaoman Qi, Junhuan Peng, Yuebin Wang, *Member, IEEE*, Xiaotong Qi, and Yun Peng

***Abstract*—High-resolution land cover mapping (LCM) is pivotal in numerous disciplines but still challenging to be acquired because traditional supervised methods require a substantial number of high-resolution labels that is labouring and expensive. To this issue, abundantly available low-resolution land-cover maps are regarded as alternative label sources, but the mismatch of spatial resolution and the misidentification of different land-cover categories introduced an amount of noisy labeled samples. This study introduces a novel cross-resolution deep learning framework, termed CRNN, to generate high-resolution LCM by leveraging low-resolution mapping products. First, a high-resolution backbone is proposed to safeguard the preservation of output resolution while simultaneously retaining the deep feature extraction capability of the network. Furthermore, an attention module is incorporated into the CRNN framework to alleviate the adverse impact of imbalanced samples. More importantly, to address the label noise issue, a weakly supervised loss based on feature similarity is proposed and calculated for obtaining dependable supervision information from low-resolution LCM products. The qualitative and quantitative results demonstrate that the CRNN framework surpasses several state-of-the-art methods. Moreover, based on the CRNN, the 10-m resolution land-cover maps of Beijing and Shanghai for 2020 are produced using 30-m resolution LCM products as reference data. As a further application, CRNN provides a viable and promising approach for reusing existing data products, contributing to some extent toward achieving sustainability goals.**

***Index Terms*—Attention module, deep learning (DL), land cover mapping (LCM), noisy label, weakly supervised learning.**

I. Introduction

HIGH-RESOLUTION land cover mapping (LCM) provides necessary information for detailed national land source surveys, spatial planning, and many sustainability-related applications [1], [2], [3]. The maturation and development of satellite and airborne remote sensing (RS) platforms have increased the availability of imagery with high spatial resolution. This abundance of data allows for the precise and periodic production of LCM [1]. Hence, exploring suitable RS data and mapping algorithms becomes particularly significant for efficiently updating and refining large-scale LCM with high spatial resolution.

Most existing LCM products, e.g., GLC_FCS30 [4], FROM_GLC10 [5], and European Space Agency (ESA) global LCM product at 10-m resolution for 2020 (called ESA10) [6], were produced by supervised classification strategies, such as support vector machine (SVM) [7], decision tree (DT) [8], and random forest (RF) [9]. With the advancement of deep learning (DL) techniques, convolutional neural networks (CNNs) have brought about a transformative impact on the realm of automatic learning, particularly concerning the acquisition of hierarchical features for pixelwise classification tasks, thus, enabling achieving accurate and efficient LCM products [1], [10], e.g., Esri 10-m LCM product during 2018–2021 (called Esri10) [11]. These supervised classification methods depend highly on numerous reference data with high accuracy [12]. In general, augmenting the number and diversity of training samples tends to yield more accurate and robust classification results [13]. Nonetheless, annotating numerous training samples is widely acknowledged as inefficient and expensive [12], [14].

As an alternative approach, several studies have explored the utilization of available low-resolution LCM products to achieve refined and accurate LCM results [1], [2], [15], [16], [17]. Although these products contain a considerable amount of noisy labels due to the mismatch of spatial resolution or misidentification of large-scale land types, they also encompass a wealth of existing knowledge and information, which benefits the model training for identifying the land cover types [15]. In previous studies, Kaiser et al. [18] demonstrated that large-scale products could serve as a substitute for manually annotated high-quality training data. And it was proved that the network could still achieve reasonable performance based on these imperfect large-scale data. Notably, the endeavor of generating high-resolution and accurate LCM through low-resolution labels holds the potential to enhance the utilization of Earth observation data in achieving sustainable development goals [16]. Nevertheless,

Manuscript received 12 September 2023; revised 27 November 2023; accepted 11 December 2023. Date of publication 14 December 2023; date of current version 28 December 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 42074004, in part by Shanxi Transportation Holdings Group Company Limited under Grant 19-JKKJ-3, and in part by the Natural Science Foundation of the Higher Education Institutions of Jiangsu Province, China under Grant 21KJB420002. *(Corresponding authors: Junhuan Peng.)*

Xiaoman Qi, Junhuan Peng, and Yuebin Wang are with the School of Land Science and Technology, China University of Geosciences, Beijing 100083, China (e-mail: qxm@email.cugb.edu.cn; pengjunhuan@163.com; wangyuebin@cugb.edu.cn).

Xiaotong Qi is with the School of Marine Technology and Geomatics, Jiangsu Ocean University, Lian Yungang 222005, China (e-mail: qixiaotong08@163.com).

Yun Peng is with the POWERCHINA Zhongnan Engineering Corporation Limited, Changsha 410014, China (e-mail: zsertyf@126.com).

Digital Object Identifier 10.1109/JSTARS.2023.3342994

along with these advantages, a significant challenge arises in the form of potential classification errors present in public LCM products and discrepancies in spatial resolution between these products and the high-resolution RS data to be classified [4]. Moreover, most of these solutions for high-resolution LCM based on low-resolution available products have not adequately considered the impact of sample imbalance on the training of models and also suffer from computational burden when the complex frameworks are utilized.

To overcome the above deficiencies, we proposed a cross-resolution CNN framework for high-resolution LCM, termed CRNN. First, multisource high-resolution satellite imageries are fused by a data fusion module, based on which the more discriminative information is extracted and utilized for accurate LCM. Second, a feature extraction module is utilized to extract feature information, thus, preventing the resolution reduction during the deep feature extraction. Followingly, an attention module is introduced in this network to alleviate the impact of imbalanced training samples. Finally, a weakly supervised loss is designed based on feature similarity to alleviate the adverse effects caused by noisy labeled samples contained in low-resolution LCM products. The CRNN framework aims to provide an alternative approach for automatically updating the large-scale high-resolution LCM without relying on high-resolution labels or ancillary information.

The main contributions of this study are as follows.

1) A novel CRNN for high-resolution LCM is designed, introducing an attention module to alleviate the impact of imbalanced samples.
2) Considering the feature similarity, a weakly supervised loss is utilized in this framework, thus, alleviating the adverse effects caused by noisy labeled samples present in low-resolution LCM products.
3) A series of extensive experiments are conducted to verify the tremendous potential of the proposed framework, and the high-resolution (10 m) LCM products of Beijing and Shanghai cities are obtained based on the high-resolution satellite RS data and available low-resolution (30 m) LCM products.

The rest of this article is organized as follows. Section II reviews the related works. Section III introduces the study area and data. Section IV represents motivation and the proposed CRNN framework. Section V gives the qualitative and quantitative results of the proposed framework and compared methods, and ablation analysis is also given in this section. Finally, Section VI concludes this article.

II. Related Work

In this section, we introduce some related works for both existing LCM products and cross-resolution LCM methods. As for existing LCM products, we review some outstanding products at both global and national scales. Moreover, we introduce different cross-resolution LCM methods that tend to solve the problem of lacking high-resolution accurate reference data.

A. Development of Public LCM Products

Due to the constraints of the sensors, the imagery accessible during the initial phases was predominantly acquired at low to moderate spatial resolutions [e.g., moderate resolution imaging spectroradiometer (MODIS)]. Over the past few decades, a multitude of approaches have been employed to leverage the abundant spectral information present in imagery for robust pixel classification. These approaches have widespread adoption in producing large-scale coarse (to 30 m) LCM products [19]. Examples of those products include 500-m MODIS land cover type product (MODIS LCM) [20], 300-m ESA climate change initiative land cover product [21], and 100-m Copernicus global land cover [22].

Subsequently, to describe the land surface more finely, several global land-cover products with 30-m spatial resolution were produced using Landsat data and numerous training samples. For instance, Gong et al. [14] produced the first 30-m resolution global LCM (FROM_GLC30) based on four supervised classifiers, including the maximum likelihood classifier (MLC), DT, RF, and SVM. Meanwhile, Chen et al. [23] developed a pixel- and object-based method with knowledge (POK-based approach) to automatically produce a global LCM at 30-m resolution, called GlobeLand30. Recently, Zhang et al. [4] introduced GLC_FCS30, providing an exemplary classification system with high accuracy. These datasets have been widely used in numerous tasks, e.g., environmental change studies, sustainable development, etc.

The accessibility of high-resolution RS data, coupled with advancements in computing and storage capacity, has facilitated the emergence of global LCM products with significantly improved resolution, reaching up to 10 m. For instance, Gong et al. [3] produced the first 10-m resolution global LCM product in 2017, named FROM_GLC10, based on Sentinel-2 images. Moreover, the ESA provided a global LCM product at 10-m resolution for 2020 (ESA10) based on Sentinel-1 and Sentinel-2 data [6]. Meanwhile, Karra et al. [11] developed a DL segmentation model utilizing Sentinel-2 data to generate a global LCM at a spatial resolution of 10 m, called Esri10.

As for the national scale, numerous outstanding high-resolution LCM products are continuously produced [24], [25], [26], [27], [28], [29]. For instance, the United States Geological Survey published the 30-m resolution National Land Cover Database (NLCD) encompassing the entirety of the United States [27], [28]. Using Sentinel data, Marston et al. [29] produced the annual U.K. LCM at 10-m resolution based on the RF classifier. For China, some studies have released high-quality national-scale LCM. For example, using Landsat images on the Google Earth engine (GEE) platform, Yang and Huang [24] produced the 30-m resolution LCM of China (CLCD) from 1985 to 2019 based on the RF classifier. Moreover, Gong et al. [26] released a novel urban land use map for the entire China (EULUC-China) based on 10-m resolution Sentinel-2 A/B images and other auxiliary data. More importantly, Li et al. [25] developed the first 1-m resolution LCM of China, SinoLC-1, using a DL-based method and high-resolution Google Earth

TABLE I
DETAILS OF DIFFERENT GLOBAL- AND NATIONAL-SCALE LCM PRODUCTS WITH A SPATIAL RESOLUTION OF 30 m OR 10 m

Products	Scale	Spatial resolution (m)	Main data sources	Method	Land cover types	years	References
FROM_GLC30	Global	30	Landsat	MLC, DT, SVM, RF	Level 1: 10; Level 2: 28	-	[14]
GlobeLand30	Global	30	Landsat	POK-based approach	10	2000, 2010, 2020	[23]
GLC_FCS30	Global	30	Landsat-8	RF	30	2015	[4]
FROM_GLC10	Global	10	Sentinel-2	RF	10	2017	[3]
ESA10	Global	10	Sentinel-1 and Sentinel-2	RF	11	2020	[6]
Esri10	Global	10	Sentinel-2	DL	10	2017–2021	[11]
NLCD	National	30	Landsat	DT	Level 1: 8; Level 2: 16	2001–2016	[28]
UK LCM	National	10	Sentinel-2	DL	10	2017–2021	[29]
CLCD	National	30	Landsat	RF	9	1990–2019	[24]
EULUC-China	National	10	Sentinel-2	RF	Level 1: 5; Level 2: 12	2018	[26]
SinoLC-1	National	1	GE imagery	DL	11	2021	[25]

(GE) imagery. The details about these high-resolution global and national scale LCM products are represented in Table I.

B. Development of Cross-Resolution LCM Methods

As mentioned above, numerous LCM products, e.g., GlobeLand30 [23], GLC_FCS30 [4], ESA10 [6], etc., have been produced by supervised classification strategies, such as SVM [7], DT [8], and RF [9]. However, these supervised classification methods depend highly on numerous reference data with high accuracy [12]. In general, annotating numerous high-quality training samples is widely acknowledged as laborious and expensive [12], [14]. To this issue, several studies have explored the utilization of available low-resolution LCM products to achieve refined and accurate LCM results [1], [2], [15], [16], [17]. Nevertheless, a significant challenge arises in the form of potential classification errors present in public LCM products and discrepancies in spatial resolution between these products and the high-resolution RS data to be classified [4].

To overcome the above deficiencies, Zhang and Roy [30] derived a continent-wide LCM at a spatial resolution of 30 m by leveraging the 500-m MODIS LCM product. Lee et al. [31] utilized an enhanced version of the Bayesian updating of land cover method to fuse unsupervised Landsat classifications to GlobCover2009, effectively enhancing the spatial resolution of classification from 300 to 30 m. Moreover, the 10-m spatial resolution FROM_GLC10 was obtained by a classifier trained directly on the 30-m resolution freely accessible data [3]. Meanwhile, Schmitt et al. [32] provided a large-scale dataset fusing the high-resolution (10 m) Sentinel-2 image data and low-resolution (250 m–1 km) MODIS land cover products. The relative experiment of CNN-based semantic segmentation revealed that the resolution of the MODIS-derived LCM was effectively enhanced, enabling the retrieval of more detailed information. However, it should be noted that the studies mentioned above directly train their models using imperfect low-resolution data without explicitly addressing the negative impact caused by the presence of significant noise. In general, classification accuracy depends not only on the classifier but also on the quality of the reference dataset [33].

Therefore, some studies utilized a small number of high-quality samples to refine the classifier trained on imperfect samples [34], [35], [36]. For instance, Hermosilla et al. [34] produced annual LCM from the Landsat time series by training the model on existing LCM products and refining it using lidar data. Moreover, Maggiori et al. [35] addressed the issue of imperfect training data by initializing the CNNs using many potentially imperfect reference data and then refining the CNNs using a smaller set of accurately annotated data. Besides, Robinson et al. [36] integrated a 30-m resolution publicly available LCM product with 1-m resolution reliable labels to enhance the generalization ability of the model and produced the first 1-m resolution LCM product of the contiguous US. In general, annotating high-resolution label data is often challenging and resource-intensive, significantly limiting the practical implementation of the abovementioned approaches.

Therefore, some studies attempted to utilize some noise-robust methods for mitigating the impact of noise and enhancing the accuracy of the LCM process [1], [25], [37], [38], [39]. For instance, Schmitt et al. [37] explored the utilization of weakly supervised learning strategies, by which valuable information was extracted for accurate high-resolution LCM. Besides, Li et al. [38] introduced four different high-resolution LCM methods using only low-resolution and noisy LCM products as training labels, all awarded as winners of the DFC2021 Track MSD. Remarkably, these methods involved different strategies, including specialized architectures, multimodel fusion, semisupervised learning approaches, etc. Subsequently, Li et al. [1] introduced a low-to-high network (L2HNet) designed to generate a high-resolution LCM based on high-resolution satellite data and low-resolution LCM products across the entire state of Maryland in the United States. L2HNet innovatively incorporated a DL backbone to extract high-resolution features from satellite imagery. In addition, it devised a low-to-high (L2H) loss function to acquire dependable supervised information from low-resolution labels. Based on this framework, Li et al. [25] established the first 1-m resolution national-scale LCM of China (SinoLC-1) based on GE imagery and other open-access data. Moreover, Huang et al. [39] adopted a simplified L2HNet (called SL2H in this study) to produce the 1-m resolution LCM of Wuhan city in China using high-resolution GE images and low-resolution LCM products.

On the other hand, the solution that corrects noisy labels at the training phase becomes a novel attempt for cross-resolution LCM. For instance, Dong et al. [15] proposed a noise correction method to rectify noisy samples based on the

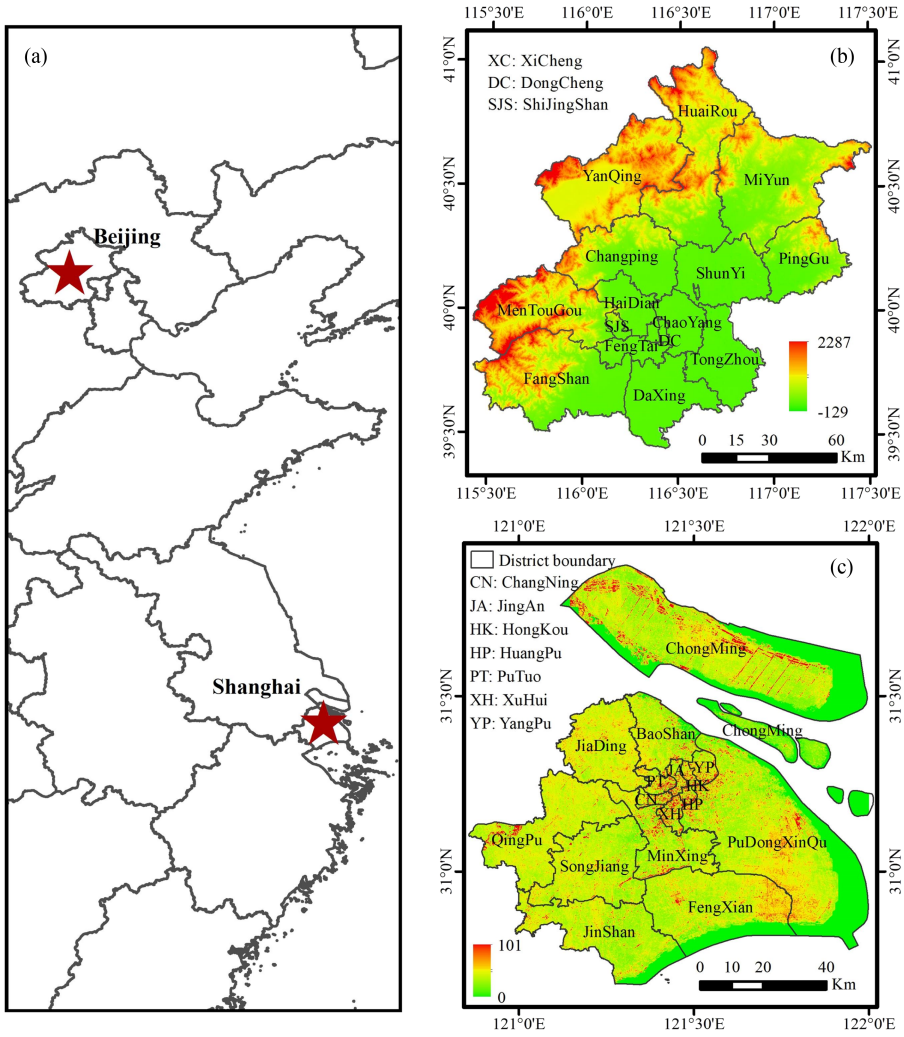


Fig. 1. Illustration of the study areas. (a) Location of two study areas. (b) Digital elevation model (DEM) maps of Beijing. (c) DEM maps of Shanghai.

predicted distribution of the CNN backbone. This proposed method successfully generated a high-resolution (3 m) LCM product using the high-resolution satellite data and a low-resolution (10 m) LCM product in China. In addition, Liu et al. [2] introduced a cross-resolution LCM framework to complete accurate LCM of China with the 10-m spatial resolution by taking the 30-m resolution historical product as the basis. They introduced a conditional random field model to refine the low-resolution reference labels.

III. Study Area and Data

In this section, a comprehensive description is provided for the experimental materials, including the study area as well as the dataset.

A. Study Area

In this study, two study areas, i.e., Beijing and Shanghai, were selected to perform the experiments, as shown in Fig. 1. Beijing (39°54′N, 116°23′E), located in the northern region of the North China Plain, serves as the capital city of China. It encompasses a total area of 16 807.8 km^2, with approximately 62% mountainous terrain and 38% plains. It comprises 16 administrative country-level subdivisions visually depicted in Fig. 1(b). As of the year 2020, the population of the Beijing metropolitan region was over 20 million. Moreover, Beijing is characterized by a temperate continental climate strongly affected by the wet monsoon. There are distinct seasonal variations, with shorter spring and autumn seasons and relatively more prolonged winter and summer seasons. The mean annual temperature hovers around 10–12 °C in Beijing. The winter season is generally dry and cold, owing to the penetration of the cold Siberian air mass southward over the Mongolian Plateau. At the same time, the summer is marked by hot and humid conditions caused by warm and moist monsoon winds originating from the southeast. The annual precipitation in Beijing is around 640 mm, with the majority (60%–70%) occurring in July and August.

Shanghai (31°12′N, 121°30′E), located on the eastern coast of China, occupies a prominent position in national and global contexts. Shanghai boasts a vast urban expanse covering approximately 6340.5 km^2. As one of the most populated cities globally, the population of the Shanghai metropolitan region

in 2020 was over 27 million, making it a bustling hub of cultural diversity and economic activity. The municipal boundary consists of 16 administrative country-level subdivisions visually depicted in Fig. 1(c). Shanghai experiences a humid subtropical climate, which is characterized by distinct seasons. Summers are hot and humid, whereas winters are generally temperate to cold and damp. The average annual temperature is 17.6 °C. Precipitation is moderate throughout the year, with the heaviest rainfall in summer. Shanghai's terrain primarily comprises flat coastal plains, and the few hills of the city lie to the southwest. Shanghai boasts a wealth of rivers, canals, streams, and lakes, making it renowned for its abundant water resources as an integral part of the Lake Tai drainage basin.

TABLE II
DESCRIPTION OF BANDS FOR THE SENTINEL-2 DATA USED IN THIS STUDY

Bands	Description	Central Wavelength (μm)	Spatial Resolution (m)
Band 1	Blue	0.490	10
Band 2	Green	0.560	10
Band 3	Red	0.665	10
Band 4	NIR	0.842	10
Band 5	SWIR	1.610	20
Band 6	SWIR	2.190	20
Band 7	NDVI	-	10
Band 8	EVI	-	10
Band 9	NDWI	-	10

B. Data and Preprocessing

1) Sentinel-1 Data: Synthetic aperture radar (SAR) imageries offer considerable advantages, especially in adverse meteorological conditions where acquiring optical data becomes challenging. Unlike optical satellites, SAR satellites have the capability to provide cloud-free images, making them a valuable alternative in such situations. Numerous studies have demonstrated that combining SAR data with optical data can improve land cover classification results [40], [41], [42].

In this study, the Sentinel-1 data in interferometric wide-swath mode was utilized. Specifically, the Sentinel-1 ground range detected (GRD) data were collected from the GEE platform. Each scene obtained from the GEE platform has undergone preprocessing with the Sentinel-1 toolbox. This preprocessing includes thermal noise removal, radiometric calibration, and terrain correction to ensure the accuracy and suitability of the data for further analysis. It should be noted that the Sentinel-1 GRD data on the GEE platform has been resampled to 10-m spatial resolution.

Sentinel-1 data is available in two polarization channels: vertical–vertical (VV) and vertical–horizontal (VH). Each polarization channel provides valuable and specific information about different land cover types [43]. Therefore, combining the VV and VH polarization channels is expected to produce superior LCM results compared with using either channel independently. In this study, the backscatter amplitude for both VH and VV polarizations served as the R and G channels for the Sentinel-1 satellite data. As the B channel, a VH-to-VV amplitude ratio (cross-pol ratio) was calculated [44]. This dataset is commonly referred to as the RGB SAR dataset. The normalized channel values for RGB SAR were scaled to (0, 255) to ensure compatibility with the expected RGB pixel values [45].

2) Sentinel-2 Data: Sentinel-2 provides multispectral RS imageries with high resolution for land monitoring, covering various aspects such as vegetation, soil, water cover, coastal areas, and emergency rescue services [46]. In this study, six bands with a resolution of either 10 m or 20 m and three most regularly used spectral indicators, i.e., the normalized difference vegetation index (NDVI) [47], enhanced vegetation index (EVI) [48], and normalized difference water index (NDWI) [49], were utilized as training data. The NDVI, EVI, and NDWI have been frequently utilized in the task of LCM [50], [51], [52].

In this study, Sentinel-2 satellite data were collected from the GEE platform. Given the GEE platform lacks abundant surface reflectance (SR) data, the study resorted to using top-of-the-atmosphere (TOA) reflectance data as a substitute. The Sentinel-2 TOA data on the GEE platform was meticulously corrected radiometrically and geometrically [53]. Although TOA reflectance is not as persuasive as SR data, earlier studies have revealed that it is a suitable substitute for assessing land cover and conducting further analysis without SR data [54].

The preprocessing of Sentinel-2 satellite data encompassed several vital steps: clouds and shadow masks, indexes calculation, spatial resampling, and temporal aggregation. First, cloudy observations were evaluated using the QA60 band, and then corresponding cloudy and shaded pixels were removed. Then, the three spectral indicators were calculated and integrated into the Sentinel-2 data as additional bands. Furthermore, all the bands were resampled to 10-m spatial resolution using bicubic interpolation. This resampling step ensured consistency and compatibility among the different input variables during further analysis. Following these, cloud-free imagery for the study area was created by calculating the median value of one-year satellite data, thus, effectively addressing the spatial heterogeneity in the observed data. Finally, single-composed satellite data with nine potential features were acquired by preprocessing, as shown in Table II.

3) Reference Data: The low-resolution (30 m) LCM product for 2020 was obtained from the publication of Zhang et al. [4], which was termed GLC_FCS30. This dataset was considered a groundbreaking achievement as it marked the first global LCM dataset with an exemplary classification system (total of 30 classes) comprising 16 comprehensive global land-cover types and 14 additional detailed and regional land-cover types. As the description in [4], the exemplary classification system of GLC_FCS30 can be described with the GlobeLand30 [23] Level 0 classification system (nine land-cover types). For GLC_FCS30 data, the preprocessing mainly comprised resampling, reprojection, and reclassifying. To ensure consistency with satellite data, the 30-m LCM production was resampled to a 10-m spatial resolution and, then, reprojected to the World Geodetic System 1984. Referring to the similar study of Li et al. [1], we merged the nine level-0 classes of the low-resolution LCM product into four classes for conducting the quantitative assessment to demonstrate the effectiveness of the CRNN framework, which was

TABLE III
DEFINITIONS AND CODES OF LAND COVER TYPES USED IN THIS STUDY

Classes of GLC_FCS30	Classes of ESA	Classes of SinoLC-1	Target classes	code
Cropland	Cropland	Cropland	Low vegetation	0
GrassLand	GrassLand	Grassland		
Wetlands	Herbaceous wetland	WetLand		
Bare areas	Bare/sparse vegetation	Barren and sparse vegetation		
Shrubland	Shrubland	Shrubland		
	Moss and lichen	Moss and lichen		
Forest	Tree cover	Tree cover	Tree canopy	1
	Mangroves			
Water body	Permanent water bodies	Water	Water	2
Impervious surfaces	Built-up	Building	Impervious	3
		Traffic route		

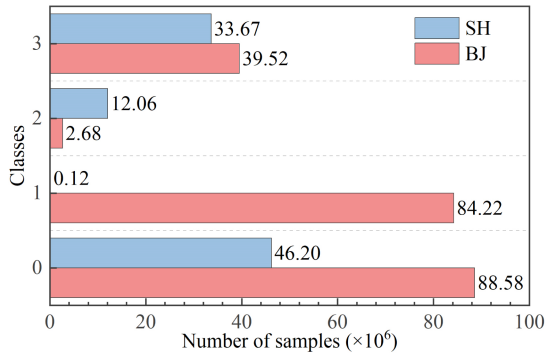


Fig. 2. Illustration of the class imbalance of GLC_FCS30 data for Beijing (BJ) and Shanghai (SH).

shown in Table III. It should be noted that the class "permanent ice and snow" is absent within the study area and will not be considered in the subsequent discussion.

To illustrate the class imbalance of GLC_FCS30, the number of pixels in each class is shown in Fig. 2. It is worth noting that class imbalance is inherent in the reference data of both study areas. As for Beijing city, the low vegetation and tree canopy are majority classes, but the water class is the minority class with few training samples. However, due to the urbanization of Shanghai, a few samples belong to the tree canopy class in the reference data. Faced with this situation, the problem of class imbalance should been considered during the design of the LCM method for improving the accuracy of LCM products.

4) Validation Data: For evaluating the performance of different methods, 3662 ground survey samples for Beijing and 1360 samples for Shanghai were obtained by visual interpretation of high-resolution satellite imagery, as shown in Fig. 3. Meanwhile, the numbers of annotated samples for different classes are also listed. All the validation samples have been released at GitHub: https://github.com/cugbrs/ValidationData_Beijing-Shanghai.

Moreover, we attempted to produce LCM products consistent with the public high-resolution LCM products based on available low-resolution LCM products. Therefore, the public high-resolution LCM products were also selected as validation data to assess and compare the performance of various LCM methods. Zanaga et al. [6] developed and validated the first global LCM products, termed ESA10, for the years 2020 and 2021 at a high resolution of 10 m. Benefiting the public of SinoLC-1 data [25], the satisfactory spatial resolution (1 m) of LCM is available for evaluating the performance of high-resolution LCM. Therefore, the ESA10 and the SinoLC-1 datasets can also be regarded as the reference data in this study. As shown in Table III, the 11 classes of the ESA10 and the SinoLC-1 were reclassified to four classes consistent with the low-resolution training label data. It should be noted that the class "snow and ice" is absent within the study area and will not be considered in the subsequent discussion.

As publicly available products, ESA10 and SinoLC-1 data contain partial classification errors. According to the validation results of [6] and [25], it is known that the ESA10 and the SinoLC-1 achieved an overall accuracy (OA) of 75% and 73.61%, respectively. Therefore, we compare the consistency of the classification results with these data only as an auxiliary result to demonstrate the validity of the methodology proposed in this study. Also discuss whether the method can generate LCM results that are comparable with publicly available product quality using only noisy labels.

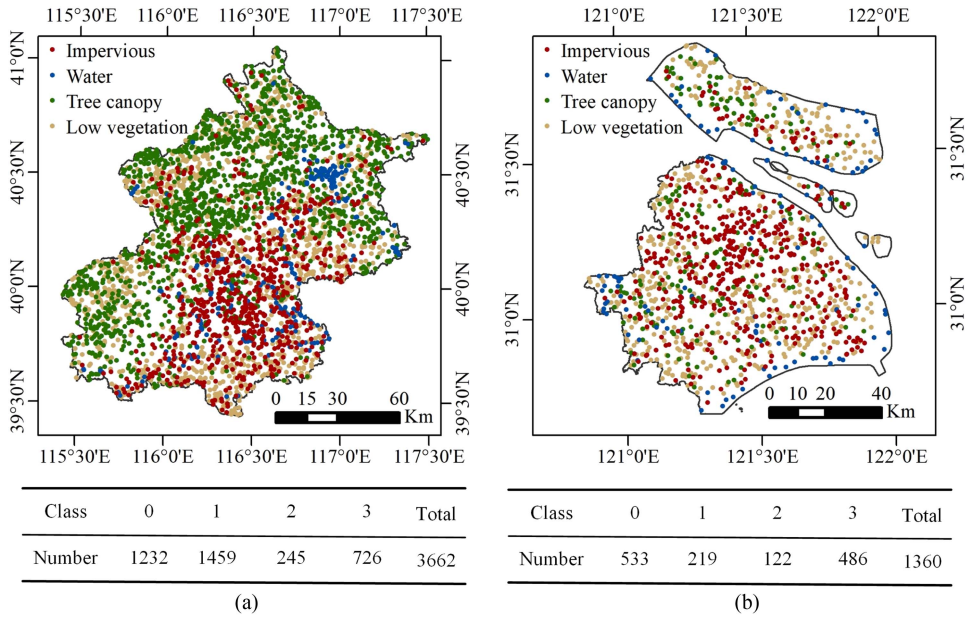
IV. METHODOLOGY

A. Motivation

Assuming that each RS image $X = \{x_1, x_2, \ldots, x_n\}$ has a label map $Y = \{y_1, y_2, \ldots, y_n\}$, where x_i is the ith pixel in image X, and n represents the total number of pixels in a single image, $y_i \in [C]$ represents the corresponding label of the pixel x_i, and C denotes the number of land cover types. Given a training dataset $D = \{(X^k, Y^k)\}_{k=1}^{N}$, where N represents the total number of training samples in the dataset. This study attempts to develop a model that can predict the accurate label $\hat{Y} = \{\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n\}$ of one image, and $\hat{Y} = f(X, \theta)$, where f refers to the function of generating prediction from input data.

B. Proposed Method

In the common segmentation task, U-Net, constructed with a deep encoder–decoder structure, is widely used and has achieved well-done segmentation performance [55], [56]. Inspired by the



Class	0	1	2	3	Total
Number	1232	1459	245	726	3662

(a)

Class	0	1	2	3	Total
Number	533	219	122	486	1360

(b)

Fig. 3. Distribution and numbers of ground survey samples. (a) Beijing. (b) Shanghai.

U-Net structure, a new simple semantic segmentation model for land use mapping (called CRNN) was proposed in this study, as shown in Fig. 4, which aims to preserve the deep feature extraction capability while simultaneously retaining a high output spatial resolution. This section provides a sequential introduction to the CRNN framework, its key components, and the loss function used for training the model. The proposed CRNN framework is a trainable DL network for producing high-resolution LCM results based on high-resolution satellite imageries and available low-resolution labels.

1) Backbone Network: As shown in Fig. 4, the proposed network can be partitioned into three distinct components: data fusion, feature extraction, and attention module.

The data fusion module can extract abundant features from multiple data and enhance the semantic information for classification. In this module, the features of Sentinel-1 and Sentinel-2 data were extracted by 3×3 convolutional layer and a batch normalization layer, respectively. Then, they were fused by concatenation in depth dimension, as shown in Fig. 4(a).

The feature extraction module contains high-resolution and deep feature extraction, as shown in Fig. 4(b). In the part of high-resolution features extraction, the resolution of the feature map remains constant to ensure consistency with the input high-resolution image. This component involves extracting spatial and spectral information through a series of 3×3 convolutional layers, followed by batch normalization, and activation layers. In the deep feature extraction part, the feature maps were scaled down in resolution to extract deeper semantic information with a larger receptive field and then scaled up in resolution to keep the resolution consistent with the initial image, similar to the encoder–decoder structure of the U-Net model. In this part, a 2×2 max pooling layer and a 2×2 average pooling layer were utilized for downsampling simultaneously. Once the two pooling operations were applied, we combined them through concatenation to produce a deep fusion feature, which has been acknowledged as an effective method for feature concentration [1]. Within the part, a concatenation operation was performed with the corresponding feature map from the same depth. This step ensures the preservation of shallow features and facilitates the application of residual learning techniques [57]. After these, the feature maps generated by the high-resolution feature extraction part and deep feature extraction part were concatenated. Then, the concatenated feature was fused by applying a sequence of 3×3 convolutional layers, followed by the batch normalization layer and activation layer. This process enables the integration of different features for enhanced representation.

Following, an attention module was utilized to make the network pay more attention to uncertain pixels, as shown in Fig. 4(c). Class imbalance causes the problem that the model tends to overfit the easily classified class, making it challenging to learn the discriminative information from the hard classified class [58]. Therefore, land cover types with fewer training samples are often challenging to learn adequately, resulting in lower prediction probabilities. To overcome this problem, information entropy was introduced in the attention module. Information entropy is an index to measure the level of information clutter. The specific and unitary information tends to have small entropy, and the uncertain and chaotic information tends to have large entropy. Therefore, information entropy can effectively guide the selection of learned features before classification. That is, the pixels with large entropy need to incorporate more semantic features to improve the accuracy of classification, which can mitigate the negative effect of imbalanced samples.

This attention module first used a 1×1 convolutional layer (conv) to map the fused feature $F = \{f_1, f_2, \ldots, f_n\}$ to different categories, then normalized the results of the different categories

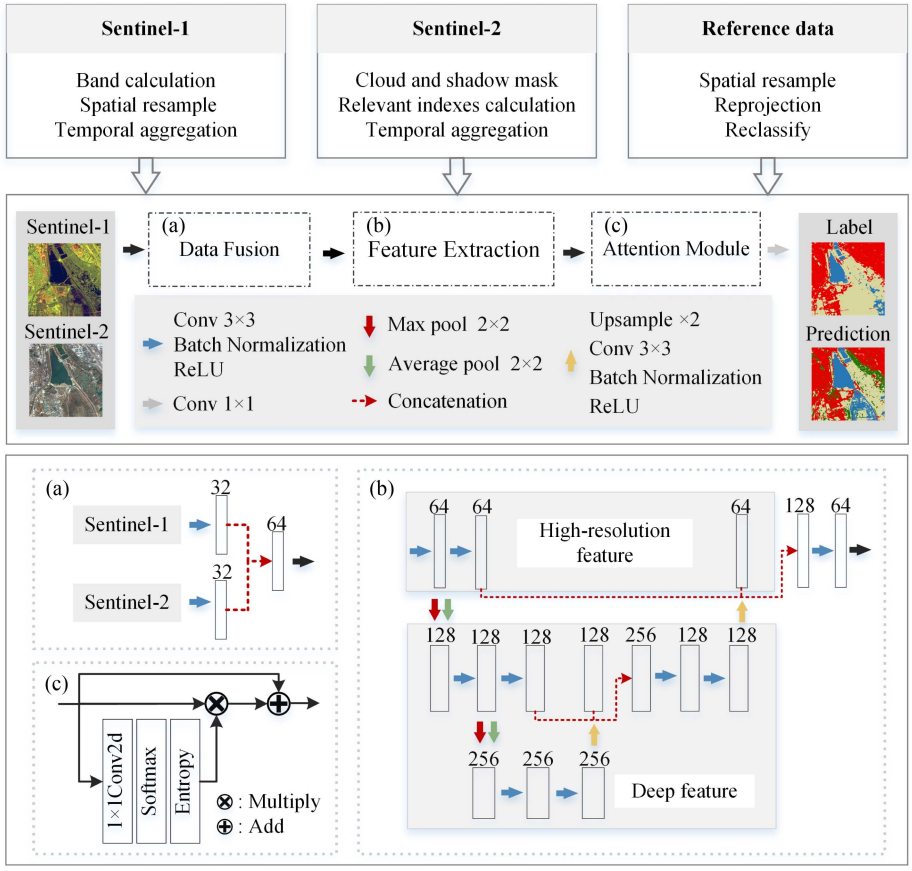


Fig. 4. Structure of the proposed backbone network. (a) Data fusion. (b) Feature extraction. (c) Attention module.

to [0, 1] using the softmax function, i.e.

$$P = \text{Softmax}(\text{conv}(F)) \tag{1}$$

$P = \{p_1, p_2, \ldots, p_n\}$, and $p_i \in [0, 1]$ represent the probability of different class for ith pixel. Subsequently, the information entropy of ith pixel e_i was calculated based on (2), and the result was used as the spatial attention weight

$$e_i = -\sum_{j=1}^{C} p_i^j \log p_i^j. \tag{2}$$

Finally, the feature was multiplied by the spatial attention weight in a pointwise manner to generate the weighted feature, and the fusion feature and weighted feature were fused by a pointwise summation, as shown in the following:

$$O_i = f_i + f_i \times e_i. \tag{3}$$

At the end of the CRNN, the classification probability was obtained through a 1×1 convolutional layer.

2) Loss Function: Due to the challenge of obtaining accurate supervision information from low-resolution labels for high-resolution mapping results, the optimization criterion in this task differs significantly from the conventional LCM task. In particular, the coarse labels contain a significant number of noisy samples due to the mismatch between high-resolution RS data and low-resolution labels, along with the misidentification of various land cover types for low-resolution LCM products. Hence, the low-resolution LCM products cannot be considered a reliable supervised source, as they may hinder the effective training process of the network. Nevertheless, despite their limitations, the coarse labels still exhibit a certain degree of reliability in matching corresponding parts of high-resolution imagery [3].

To this issue, a weakly supervised loss was employed to identify and retain the reliable portions of the coarse labels. This loss effectively alleviated the influence of noisy labels during the model training process, enhancing the robustness and accuracy of the network's performance. First, the confident area was calculated according to the input features since pixels of the same class often have more similar feature information [1], [17]. Formally, the input features for pixel (i, j) can be represented as I_{ij}, where $i \in [0, H]$ and $j \in [0, W]$, H and W represent the height and width of each image, respectively.

Furthermore, due to the majority of the annotations in the coarse labels being correct, we calculated the average vector of input features $\hat{I}^{(l)}$ for each special class l in a minibatch, which was regarded as the standard feature representation

$$\hat{I}^{(l)} = \frac{1}{N^{(l)}} \sum_{i=1}^{H} \sum_{j=1}^{W} \hat{I}_{i,j}^{(l)} \tag{4}$$

where $N^{(l)}$ is the number of pixels for each special class l. Then, the distance $S_{i,j}^{(l)}$ between the standard vector $\hat{I}^{(l)}$ and the vectors for each pixel $I_{i,j}^{(l)}$ was calculated

$$S_{i,j}^{(l)} = \left\| I_{i,j}^{(l)} - \hat{I}^{(l)} \right\|_2. \tag{5}$$

Finally, a threshold $t^{(l)}$, which was calculated by averaging the distance values for specific class l, was utilized to obtain a selective mask for the confident area (mask_{CA}). It is crucial to emphasize that the threshold used in this process is class-conditional, thereby preventing difficult and rare classes from being incorrectly identified as label noise. The abovementioned processes can be represented as

$$t^{(l)} = \frac{1}{N^{(l)}} \sum_{i=1}^{H} \sum_{j=1}^{W} S_{i,j}^{(l)} \tag{6}$$

$$\text{mask}_{\text{CA}} = \left\{ (i,j) | S_{i,j}^{(l)} < t^{(l)} \right\}. \tag{7}$$

The reliability of the pixels included in the confident area is sufficiently high, and the training loss could be calculated as

$$L_W = -\sum_{i=1}^{H} \sum_{j=1}^{W} \left[y_{(i,j)\in \text{mask}_{\text{CA}}} \log \hat{y}_{(i,j)\in \text{mask}_{\text{CA}}} + \left(1 - y_{(i,j)\in mask_{\text{CA}}}\right) \log \left(1 - \hat{y}_{(i,j)\in \text{mask}_{\text{CA}}}\right) \right] \tag{8}$$

Meanwhile, due to the coarse labels still containing partial reliable semantic information, the original noisy label Y was also incorporated for training purposes, and L_N was formulated as

$$L_N = -\sum_{i=1}^{H} \sum_{j=1}^{W} [y_{(i,j)} \log \hat{y}_{(i,j)} + \left(1 - y_{(i,j)}\right) \log \left(1 - \hat{y}_{(i,j)}\right)]. \tag{9}$$

Therefore, a joint loss function L_{joint} was applied in this study, and it could be expressed as

$$L_{\text{joint}} = L_W + \alpha L_N \tag{10}$$

where $\alpha = 0.3$ represents a hyperparameter that balances the two loss terms during the training phase.

Previous research has demonstrated that DL models tend to learn from simple and clean labels before adapting to label noise [2]. Therefore, the traditional cross-entropy loss with all coarse labels, i.e., L_N, was used to warm up the model, after which the model was trained using the joint loss L_{joint}.

C. Experimental Settings

In this study, the high-resolution satellite imageries, the resampled low-resolution reference data, and the ESA10 product were all cropped into patches of 128×128 pixels with 20% overlap. Moreover, we fused the satellite data and the resampled low-resolution reference data to build the training datasets. The high-quality ground truth samples and high-resolution LCM products were utilized to validate the performance of different models. The CRNN was first trained for 20 epochs at a fixed learning rate of 0.01 with L_N, after which the network was trained for 80 epochs at a fixed learning rate of 0.001 with L_{joint}. The PyTorch framework was employed to train all the networks, utilizing the AdamW optimizer. The algorithms were executed on an NVIDIA GeForce RTX 2080Ti GPU for efficient processing and computation.

D. Compared Methods

In this study, seven outstanding classification methods in the task of LCM were chosen as comparison methods, i.e., SVM [7], RF [9], U-Net [57], SegNet [60], NCLCM [15], L2HNet [1], and SL2H [39]. SVM and RF, which are traditional machine learning methods, are commonly used baseline models for RS due to their capability to handle high-dimensional input variables [1], [17].

Ronneberger et al. [57] introduced a U-Net model with an encoder–decoder architecture designed for biomedical image segmentation tasks. The U-Net model enhances the reusability of features in image segmentation tasks by utilizing the concatenation of multilevel feature maps with matching dimensions. Numerous studies have proven that the U-Net model is efficient in the task of LCM [10], [17], [59].

The SegNet model was introduced for semantic pixelwise segmentation, whose distinctive feature lies in its approach to upsampling the lower resolution input feature map(s) within the decoder. In particular, the decoder leverages the pooling indices obtained during the maxpooling step in the corresponding encoder for nonlinear upsampling, eliminating the need to explicitly learn the upsampling process [60]. Moreover, SegNet has been widely utilized in RS image classification tasks [31], [61].

To obtain the high-resolution LCM based on low-resolution and low-accuracy products, Dong et al. [15] proposed a noise correction method, i.e., NCLCM, to rectify noisy samples based on the predicted distribution of CNN backbone. This proposed method successfully generated a high-resolution (3 m) LCM product using the high-resolution satellite data and a low-resolution (10 m) LCM product in China.

Moreover, to overcome the barrier of unmatching resolution between satellite images and label data, Li et al. [1], [25] introduced a novel network architecture called L2HNet to obtain high-resolution LCM results automatically and produced a national high-resolution LCM product. L2HNet accomplished this without relying on accurate high-resolution annotated labels during the large-scale land-cover map-updating process. Besides, Huang et al. [39] adopted the SL2H method to produce the 1-m resolution LCM covering Wuhan city of China based on high-resolution satellite data.

E. Evaluated Metrics

To evaluate the LCM performance of various methods, OA, mean producer accuracy (mPA), mean user accuracy (mUA), mF_1, and the mean intersection over union (mIoU), all of which are commonly used in relative tasks, were selected to evaluate the classification results derived by different methods in this study. Assume that TP represents the true positive sample, TN is the true negative sample, FP is the false positive sample, and FN is the false negative sample. The formulas of the metrics

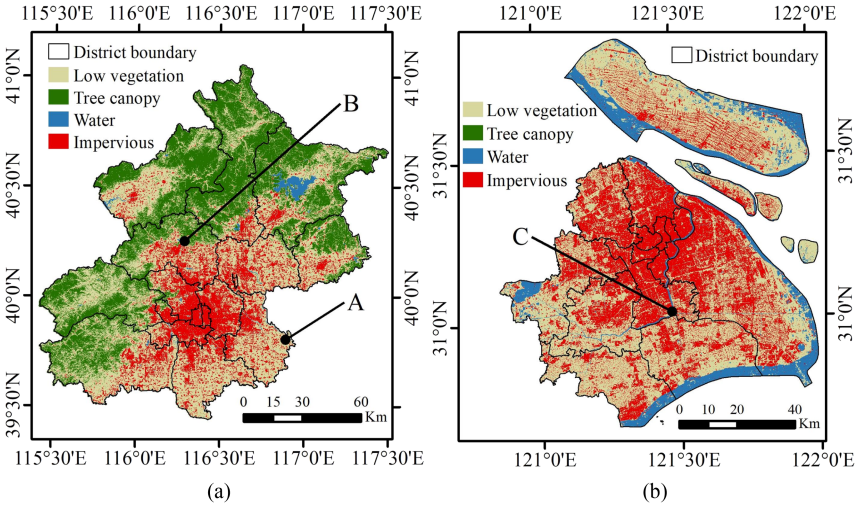


Fig. 5. LCM of proposed CRNN methods in (a) Beijing and (b) Shanghai for the year 2020. A, B, and C are three regions of interest (ROI-A, ROI-B, and ROI-C) for showing details in the following discussion.

mentioned above are shown as follows:

$$\text{OA} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{11}$$

$$\text{PA} = \frac{\text{TP}}{\text{TP+FN}} \tag{12}$$

$$\text{UA} = \frac{\text{TP}}{\text{TP+FP}} \tag{13}$$

$$F_1 = \frac{2 \times \text{PA} \times \text{UA}}{\text{PA} + \text{UA}} \tag{14}$$

$$\text{IoU} = \frac{\text{TP}}{\text{TP+FP+FN}} \tag{15}$$

PA, UA, F_1, and IoU were computed independently for each class and then averaged to obtain the final value, i.e., mPA, mUA, mF_1, and mIoU.

V. Results and Discussion

In this section, qualitative and quantitative analyses were utilized to evaluate the performance of the proposed CRNN framework. Moreover, a series of ablation experiments were performed to prove the efficiency of the modules and settings of the CRNN framework.

A. Qualitative Results

For the qualitative comparison, the LCM results of the CRNN method in Beijing and Shanghai for 2020 are described in Fig. 5. Based on Figs. 1(b) and 5(a), it can be observed that impervious surfaces in Beijing are predominantly distributed in the relatively flat central-southern and southeastern regions, whereas the northern and western high-altitude areas are primarily covered by vegetation. Moreover, impervious surfaces, mainly found in urban residential areas, are often accompanied by a significant amount of low-growing vegetation. As shown in Figs. 1(c) and 5(b), Shanghai has a flat terrain, with the primary land cover types being impervious surfaces and low vegetation. The central urban area is characterized by dense buildings, whereas small and medium-sized towns are scattered in the suburbs. Agricultural land, bare land, and other low vegetation are mainly distributed on the outskirts of the city. Overall, the above findings are consistent with other studies [62], [63]. For a detailed visual comparison of the results obtained by different methods, we selected three regions of interest (ROI-A, ROI-B, and ROI-C) to showcase the mapping results, as depicted in Figs. 6–8, respectively.

As depicted in Figs. 6 and 7, ROI-A and ROI-B encompass all land cover classes considered in this study, enabling a comprehensive assessment of the classification performance of different methods. Through the comparison of Figs. 6(f) and (g) and 7(f) and (g), the classification results of two traditional machine learning methods (SVM and RF) exhibit good capability in predicting roads; however, they encounter difficulties in accurately predicting the distribution of water bodies and small-scale impervious surfaces. Figs. 6(h) and (i), and 7(h) and (i) illustrate the classification results of two classic DL methods, i.e., U-Net and SegNet. The classification results of U-Net and SegNet are rough and close to the low-resolution labels. The results indicate that due to the influence of the deep downsampling structure, these two methods struggle to accurately depict finer road distributions or the course of rivers. Meanwhile, in the subplot of Figs. 6(j) and (k), and 7(j) and (k), the results of two cross-resolution LCM methods (NCLCM and L2HNet) show the difficulty in distinguishing the classes with limited samples, e.g., water. This result can be attributed to the confident area selection of both methods based on whole class samples rather than class-conditional, which leads to the model neglecting the classes with limited samples. As shown in Figs. 6(l) and 7(l), SL2H as a simple version of L2HNet without the L2H loss shows the difficulty in recognizing fine features, such as road.

As depicted in Figs. 6(m) and 7(m), the proposed CRNN framework demonstrates its superiority in qualitative comparison. First, leveraging the high-resolution semantic

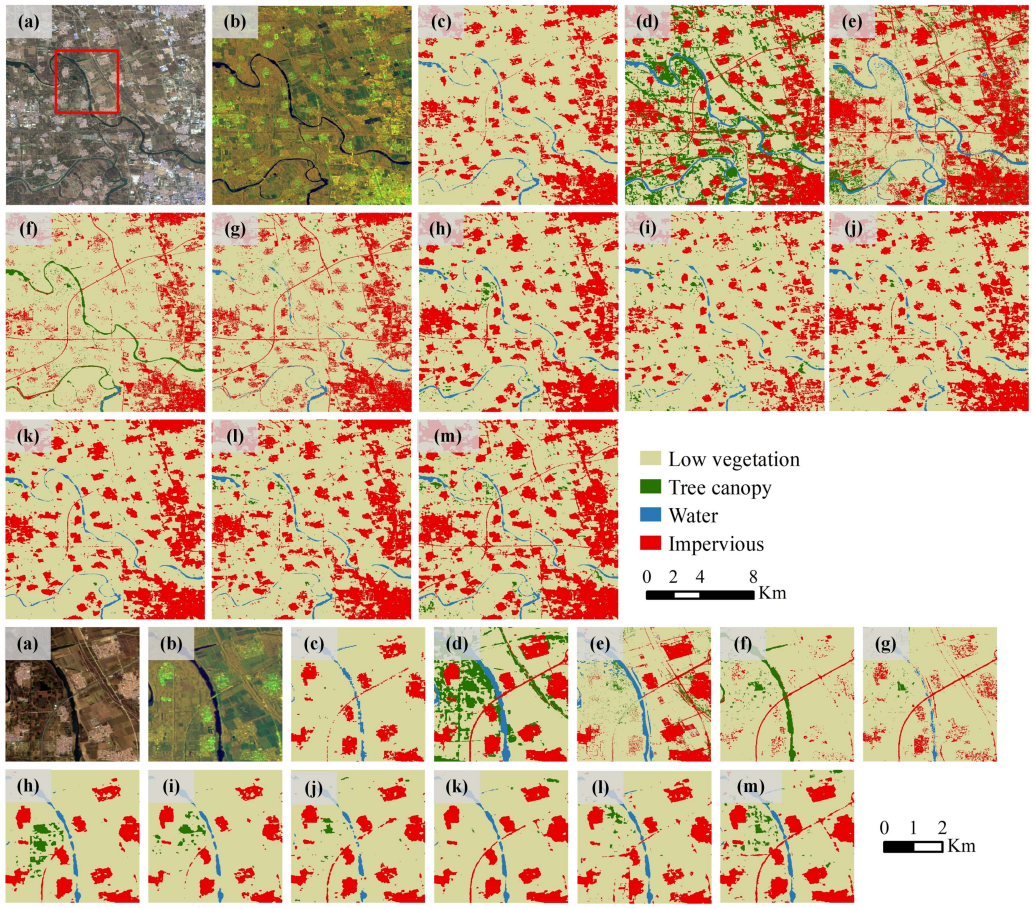


Fig. 6. Comparison of the classification results of different methods for ROI-A. (a) Sentinel-2 RGB imagery (R: Band 4; G: Band 3; B: Band 2). (b) Sentinel-1 false RGB imagery (R: VH; G: VV; B: VH/VV). (c) GLC_FCS30_2020. (d) ESA10_2020. (e) SinoLC-1. (f) SVM. (g) RF. (h) U-Net. (i) SegNet. (j) NCLCM. (k) L2HNet. (l) SL2H. (m) CRNN.

backbone network, the mapping results maintain a high consistency with the input images, thereby preserving more detailed information about the land cover. Second, benefitting from the attention module, the proposed method can somewhat alleviate the issue of poor model classification performance caused by imbalanced training samples. Thus, the water category could be described better by the CRNN than most of the other methods. Finally, weakly supervised learning can mitigate the impact of noise in low-resolution labeled data, thereby obtaining classification products that are more consistent with the input imagery data. However, there is a substantial difference in the distribution of tree class between the predicted map of CRNN and the ESA product, while the distribution of CRNN maps is generally consistent with SinoLC-1 maps [Figs. 6(e) and 7(e)]. This disparity may be due to the lack of training samples for the tree class in the training labels, i.e., the GLC_FCS30 dataset. In addition, it is also possible that the ESA product contains some classification errors. Nevertheless, based on low-resolution products, we have generated high-resolution products that are generally consistent with ESA10, proving the efficiency of the CRNN method and providing reliable insights for practical production needs.

Fig. 8 displays the qualitative comparison of different methods for a bridge spanning a river. As can be seen from Fig. 8(f) and (g), although SVM and RF can calculate labels for each pixel more detailedly, their classification accuracy was not satisfactory due to the influence of noisy labels. For example, SVM misclassified water as low vegetation, whereas RF misclassified the bridge as low vegetation. As shown in Fig. 8(h)–(l), the five DL methods struggled to distinguish smaller land objects accurately because of the limited capability for high-resolution feature extraction or learning from noisy labels. Notably, CRNN outperformed two traditional machine learning methods and gave comparable results with five DL methods.

In summary, the two traditional machine learning methods appear to have accurate predictions for fine details such as roads. However, they exhibit weak discriminative power for confusable land cover types, such as water bodies. On the other hand, the two classic DL segmentation models are susceptible to network structure and noisy labels, often leading to a loss of detailed information in the prediction results. Although the NCLCM and L2HNet methods can partially mitigate the impact of noisy labels, their ability to handle imbalanced samples is limited due to the lack of consideration for small-sample categories. SL2H makes obtaining accurate LCM based on low-resolution training labels challenging due to the lack of noise-robust loss function. The proposed CRNN method demonstrates a balanced performance compared with other methods, producing mapping results that showcase both accurate land cover classification and

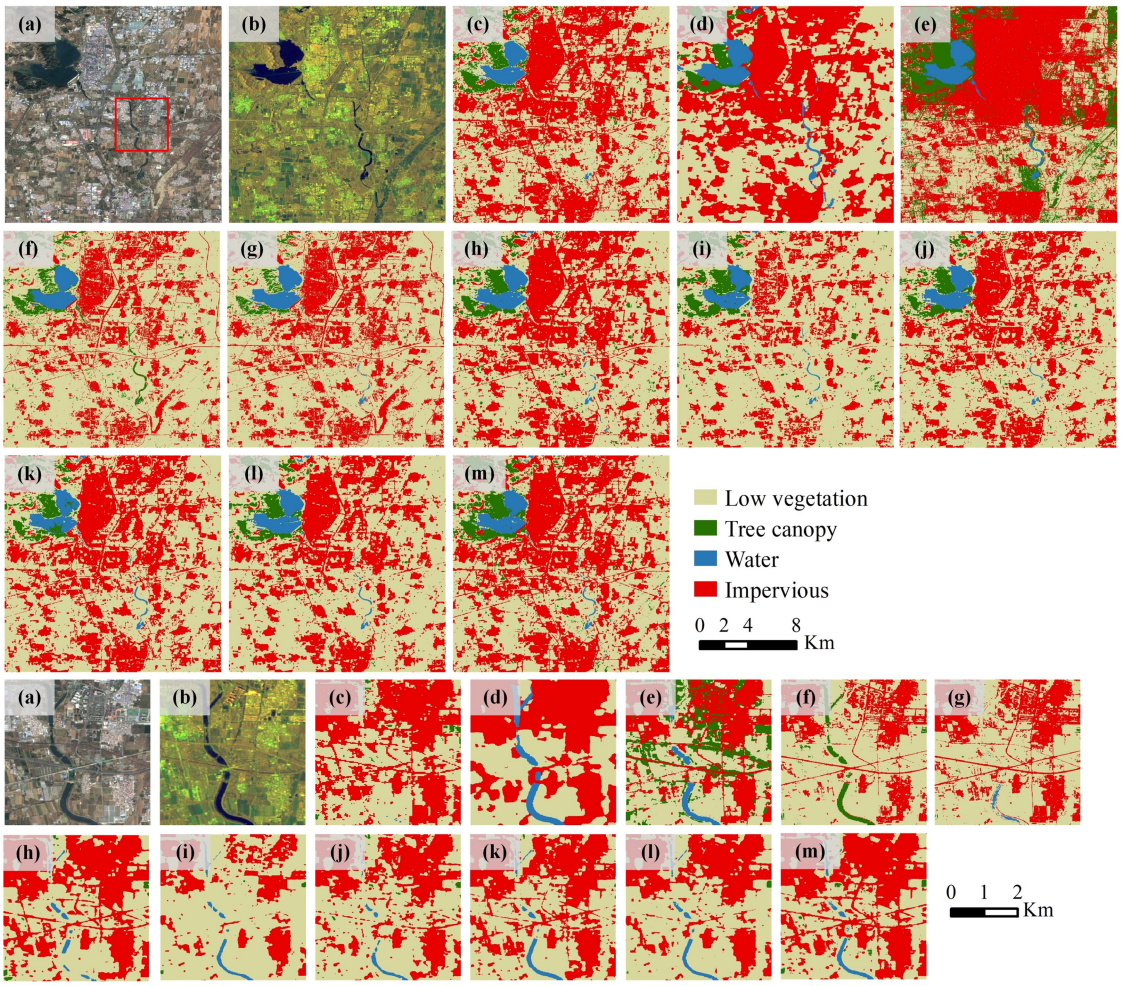


Fig. 7. Comparison of the classification results of different methods for ROI-B. (a) Sentinel-2 RGB imagery (R: Band 4; G: Band 3; B: Band 2). (b) Sentinel-1 false RGB imagery (R: VH; G: VV; B: VH/VV). (c) GLC_FCS30_2020. (d) ESA10_2020. (e) SinoLC-1. (f) SVM. (g) RF. (h) U-Net. (i) SegNet. (j) NCLCM. (k) L2HNet. (l) SL2H. (m) CRNN.

precise detail information, which may stem from the advantage of the attention module and the weakly supervised loss function.

B. Quantitative Results

During the quantitative comparison, all metrics were calculated based on the ground survey samples and the ESA10 datasets, respectively. The pixelwise OA, mPA, mUA, mF_1, and mIoU of different methods are represented in Tables IV and V. Due to the SinoLC-1 data was developed by the L2HNet method, SinoLC-1 was only used for qualitative comparisons.

As shown in Table IV, two traditional machine learning methods (SVM and RF) obtained unsatisfactory performance compared with most DL methods. Specifically, they obtained about 76.09% and 67.79% for the average values of OA for Beijing and Shanghai, respectively, indicating that these two methods are unsuitable for the cross-resolution LCM task. For two classical DL methods (U-Net and SegNet), the improvement of OA was significant, with an increase of about 3.54% in OA for Beijing and about 3.76% for Shanghai, which may be attributed to the powerful feature learning capability of DL models. However, due to the impact of noisy labels and imbalanced samples, these two simple segmentation models struggled to accurately classify small-sample land cover types, leading to lower mIoU value (about 69.74% and 49.43% for the average values of mIoU for Beijing and Shanghai, respectively). Commonly, deep encoder–decoder semantic networks tend to excessively down-sample the feature maps, thereby, promoting consistency with low-resolution LCM labels. As a result, these semantic methods may not be well-suited for accurately mapping intricate land-cover details when confronted with a substantial number of incorrectly labeled samples. As for NCLCM and L2HNet, benefiting the series L2H classification module, its prediction obtained satisfactory OA results, averaging 81.09% for Beijing and 73.02% for Shanghai. However, due to NCLCM and L2HNet using the same threshold for all land cover types when calculating confident areas, both methods were susceptible to the influence of imbalanced samples, resulting in poorer mUA values (about an average of 85.33% for Beijing and 58.97% for Shanghai). Compared with L2HNet, SL2H yielded poorer classification performance (81.19% OA for Beijing and 73.29%

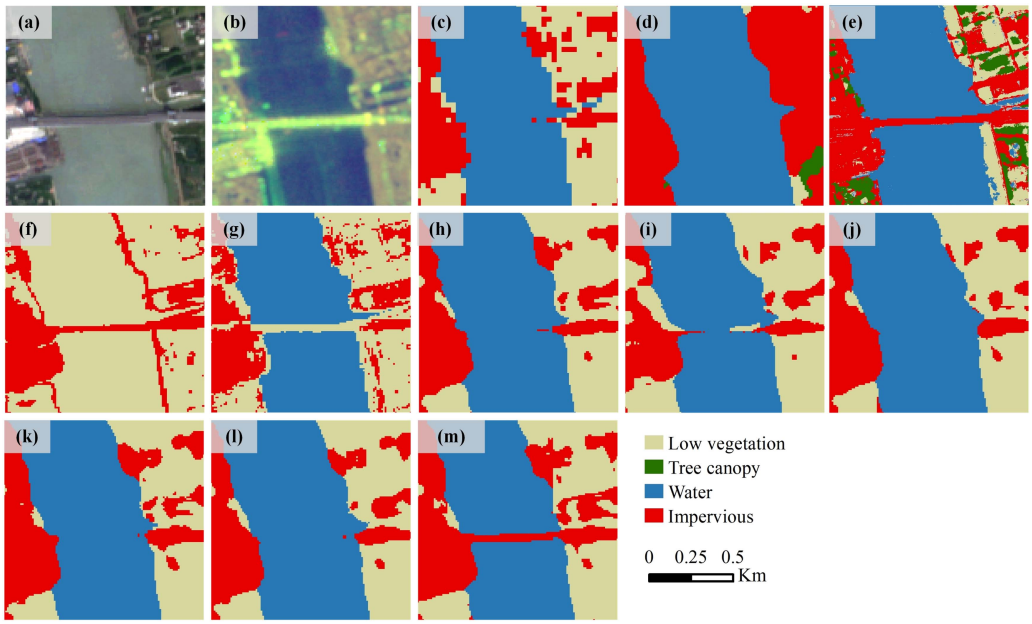


Fig. 8. Comparison of the classification results of different methods for ROI-C. (a) Sentinel-2 RGB imagery (R: Band 4; G: Band 3; B: Band 2). (b) Sentinel-1 false RGB imagery (R: VH; G: VV; B: VH/VV). (c) GLC_FCS30_2020. (d) ESA10_2020. (e) SinoLC-1. (f) SVM. (g) RF. (h) U-Net. (i) SegNet. (j) NCLCM. (k) L2HNet. (l) SL2H. (m) CRNN.

TABLE IV
QUANTITATIVE RESULTS OF THE COMPARISON METHODS AND THE PROPOSED FRAMEWORK BASED ON THE GROUND SURVEY SAMPLES

Study Area	Methods	OA(%)	mUA(%)	mPA(%)	mF_1(%)	mIoU(%)
BJ	SVM	73.91	78.27	68.98	72.13	56.99
	RF	78.26	**87.36**	77.72	80.96	68.90
	U-Net	81.22	85.37	83.70	84.12	73.09
	SegNet	78.03	85.95	76.00	79.55	66.38
	NCLCM	80.90	85.79	80.94	82.88	71.04
	L2HNet	81.28	84.86	82.61	83.38	71.98
	SL2H	81.19	84.66	82.30	83.25	71.60
	CRNN	**82.72**	86.40	**84.39**	**85.15**	**74.55**
SH	SVM	70.19	59.72	64.61	61.68	52.45
	RF	65.38	57.55	62.86	59.43	51.30
	U-Net	71.47	56.58	64.46	59.53	49.47
	SegNet	71.62	**60.89**	63.59	59.87	49.38
	NCLCM	72.51	59.65	63.49	61.51	51.89
	L2HNet	73.53	58.29	65.93	60.82	51.29
	SL2H	73.29	59.10	65.49	60.13	53.57
	CRNN	**74.26**	60.24	**66.72**	**61.99**	**54.35**

BJ:Beijing; SH:Shanghai; The best results are shown in bold type.

OA for Shanghai) because of the lack of noise-robust L2H loss. The proposed CRNN method generally demonstrated a well-balanced performance compared with the other methods and achieved the highest OA, mF_1 scores, and mIoU. As displayed in Table IV, the CRNN method achieved 82.72% OA and 74.55% mIoU for Beijing city and 74.26% OA and 54.35% mIoU for Shanghai City, representing an improvement of around 0.73%–8.88% in OA and 1.46%–17.56% in mIoU compared with baseline methods.

Moreover, it can be seen from Table V that CRNN also obtained superior performance compared with baseline methods when the ESA10 was regarded as validation data. In particular, CRNN achieved the best OA compared with other methods, with an increase of an average of 1.73% and 2.37% for Beijing and Shanghai, respectively. Therefore, it proved that the classification results of CRNN are consistent basically with the existing high-resolution LCM product, indicating the proposed framework is able to produce high-quality LCM product based on available low-resolution LCM products rather than numerous field survey samples.

To further discuss the LCM performance of various methods for different categories, taking Beijing as an example, we

TABLE V
QUANTITATIVE RESULTS OF THE COMPARISON METHODS AND THE PROPOSED FRAMEWORK BASED ON ESA10

Study Area	Methods	OA(%)	mUA(%)	mPA(%)	mF_1(%)	mIoU(%)
BJ	SVM	76.70	77.06	76.70	76.71	63.26
	RF	76.50	77.18	76.50	76.38	63.88
	U-Net	79.47	80.02	73.53	76.64	61.83
	SegNet	79.49	**80.78**	72.13	76.21	60.49
	NCLCM	79.95	80.04	73.11	76.42	61.59
	L2HNet	78.74	79.46	70.92	74.95	59.38
	SL2H	78.68	78.02	68.71	73.50	58.59
	CRNN	**80.20**	80.26	**76.95**	**78.57**	**64.67**
SH	SVM	63.56	64.27	**72.96**	68.12	46.58
	RF	65.26	70.91	65.26	60.52	48.03
	U-Net	70.69	81.07	59.12	68.38	47.22
	SegNet	71.17	81.17	59.97	68.98	47.27
	NCLCM	71.16	82.12	59.51	69.01	47.72
	L2HNet	71.14	**82.92**	58.71	68.74	47.03
	SL2H	70.88	81.56	53.08	64.30	43.86
	CRNN	**71.20**	81.56	65.29	**72.42**	**48.11**

BJ:Beijing; SH:Shanghai; The best results are shown in bold type.

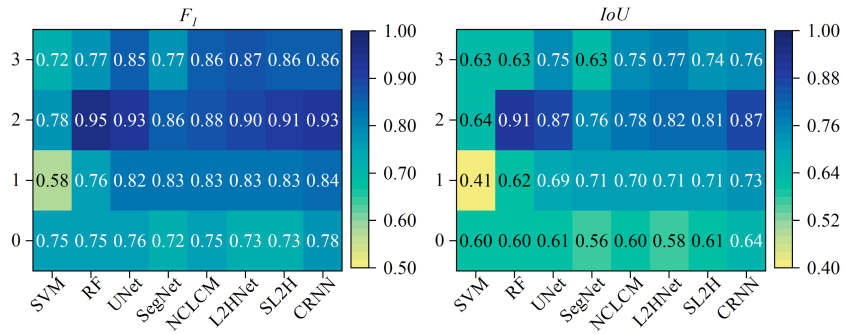


Fig. 9. Heatmaps of F_1 and IoU for Beijing. Horizontal axis represents different methods, and the vertical axis represents different land cover types.

exhibited the F_1 heatmap and IoU heatmap of all the methods for the four land cover types in Fig. 9. It can be observed that the SVM and RF showed poorer and more unbalanced classification performance compared with four DL-based methods, with lower F_1 and IoU for most of the classes. Remarkably, the improvement of six DL-based methods in F_1 and IoU for tree canopy and water classes was particularly noticeable, with an increase of about 8%–26% in F_1 and 12%–32% in IoU, respectively. Moreover, the CRNN method exhibited a well-balanced performance compared with other methods. It achieved robust and satisfactory results for each class, with the highest F_1 and IoU scores on low vegetation and tree canopy. This result demonstrates the superiority of the CRNN method, which benefits from the attention module. This module pays more attention to the uncertain pixels and improves the classification accuracy of the CRNN method. Furthermore, it can be noticed that low vegetation is the most challenging class to be recognized by the models. This result may be due to the fact that low vegetation consists of more different land cover types and introduces more complex information, which can confuse the models.

In summary, the comparison methods were unable to achieve the desired accuracy for LCM due to the challenges posed by imbalanced and noisy labels. SVM and RF, two traditional machine learning methods, obtained the low OA suffering from the imperfect training label. U-Net and SegNet, two deep semantic segmentation networks, tended to be consistent with the distribution of the original label and were relatively unsuitable for the accuracy mapping based on noisy training samples. Furthermore, NCLCM and L2HNet acquired poorer classification results, possibly due to the impact of imbalanced samples. Besides, it is difficult for the SL2H method to alleviate the impact of noisy samples in low-resolution LCM products due to the lack of noise-robust loss function. The proposed CRNN method achieved the best OA, mF_1 scores, and mIoU. This may be because the model considers high-resolution detail representation and the impact of imbalanced samples and mitigates the influence of noisy labels through a loss function.

C. Ablation Experiments

To verify whether the proposed CRNN framework and corresponding modules are effective for the LCM process, several ablation experiments were performed for both study areas. We demonstrated the quantitative and qualitative results of the

TABLE VI
QUANTITATIVE RESULTS OF ABLATION EXPERIMENTS

Study Area	Methods	OA(%)	mUA(%)	mPA(%)	mF_1(%)	mIoU(%)
BJ	CRNN^0	76.34	84.61	77.58	79.50	66.38
	CRNN^{AT}	78.71	85.47	74.10	77.86	64.11
	CRNN^{WL}	78.41	84.65	77.98	79.85	66.66
	CRNN^{S2}	79.69	83.05	82.39	82.32	70.56
	CRNN	**82.72**	**86.40**	**84.39**	**85.15**	**74.55**
SH	CRNN^0	71.34	52.23	59.12	55.46	47.31
	CRNN^{AT}	73.65	55.89	62.42	58.97	53.56
	CRNN^{WL}	73.65	56.01	62.61	59.13	51.78
	CRNN^{S2}	71.30	51.63	63.70	57.03	50.66
	CRNN	**74.26**	**60.24**	**66.72**	**61.99**	**54.35**

BJ:Beijing; SH:Shanghai; The best results are shown in bold type.

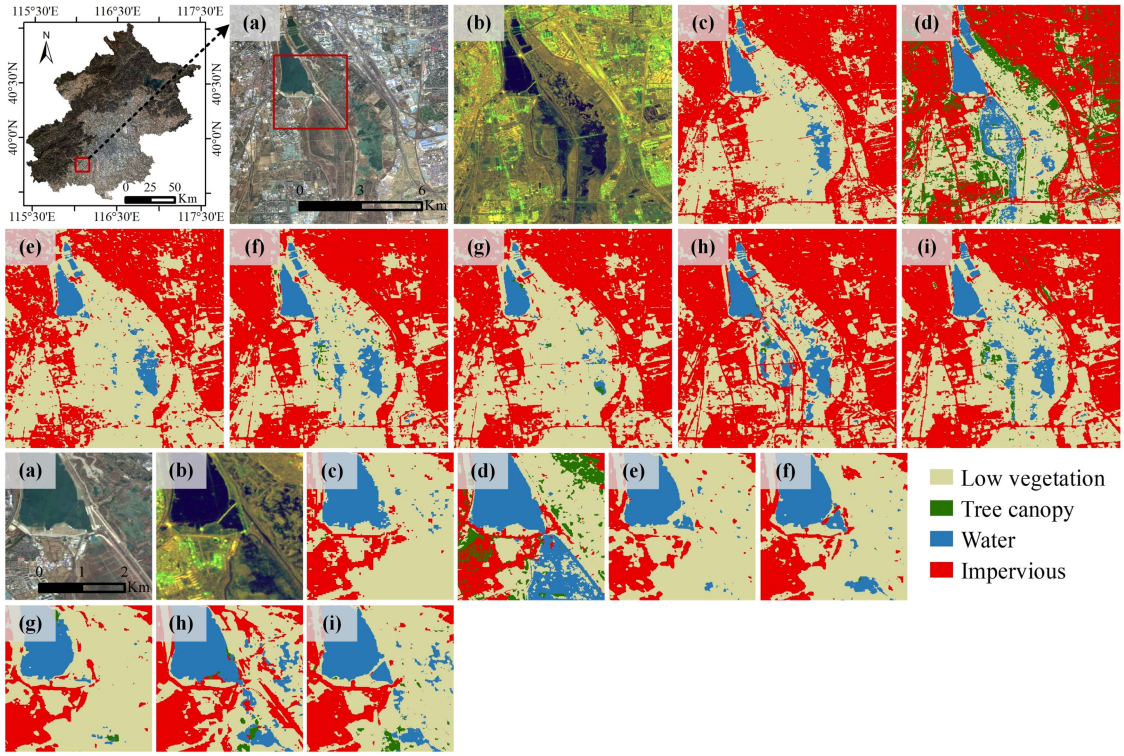


Fig. 10. Comparison of the classification results of ablation experiments for Beijing. (a) Sentinel-2 RGB imagery, (b) Sentinel-1 false RGB imagery. (c) GLC_FCS30_2020. (d) ESA10_2020. (e) CRNN^0. (f) CRNN^{AT}. (g) CRNN^{WL}. (h) CRNN^{S2}. (i) CRNN.

CRNN framework trained with various settings, respectively. The definition of different models is as follows:

CRNN^0: Backbone model trained with CrossEntropy loss, i.e., CRNN without attention module, and the weakly supervised loss function.

CRNN^{AT}: Backbone model with attention module and CrossEntropy loss, i.e., CRNN without the weakly supervised loss.

CRNN^{WL}: Backbone model with the weakly supervised loss, i.e., CRNN without attention module.

CRNN^{S2}: Our proposed model with single Sentinel-2 data, i.e., CRNN without data fusion.

CRNN: Our proposed model.

The quantitative results of the ablation experiments are shown in Table VI, and the qualitative results are shown in Figs. 10 and 11.

From the results shown in Table VI, CRNN^0 obtained acceptable quantitative results with 76.34% and 71.34% OA for Beijing and Shanghai, respectively, which indicated that the proposed backbone model could extract enough discriminating information and acquire accurate mapping results based on coarse label. Furthermore, the CRNN^{AT} obtained more incredible classification results than CRNN^0 with an average of 2.34% increase in OA for both study areas because the attention module can pay more attention to unreliable pixels and, then, improve the whole classification accuracy. From the quantitative

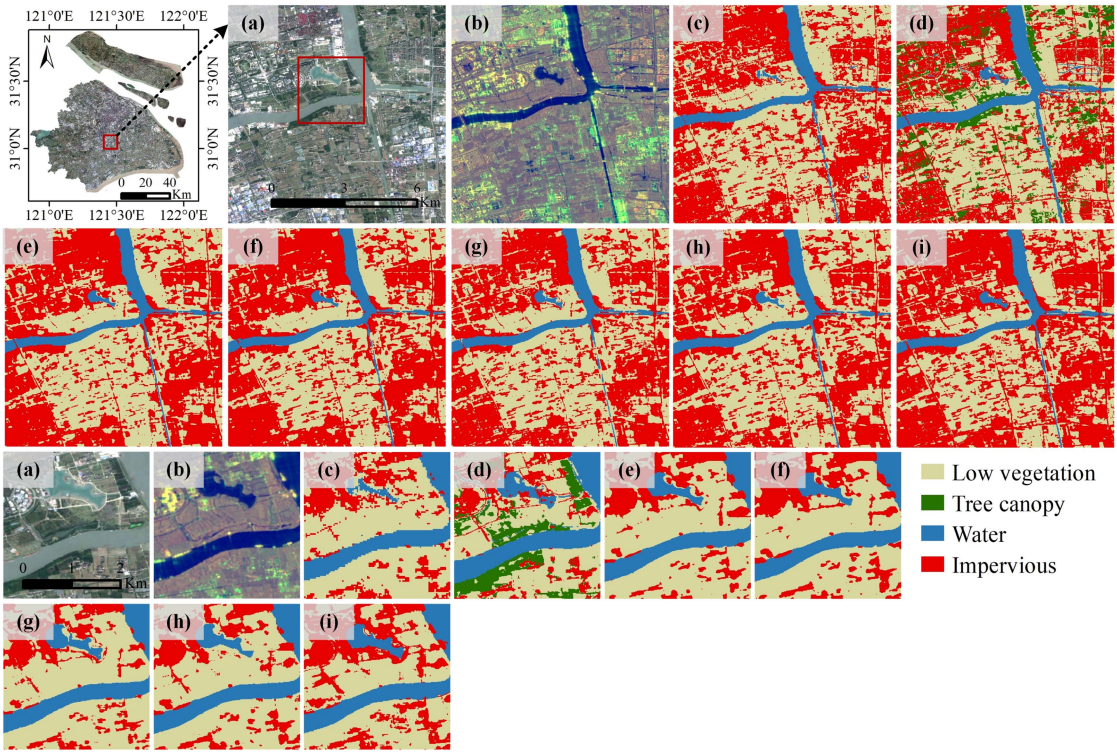


Fig. 11. Comparison of the classification results of ablation experiments for Shanghai. (a) Sentinel-2 RGB imagery, (b) Sentinel-1 false RGB imagery. (c) GLC_FCS30_2020. (d) ESA10_2020. (e) CRNN^0. (f) CRNN^{AT}. (g) CRNN^{WL}. (h) CRNN^{S2}. (i) CRNN.

results, when the backbone model was trained with a weakly supervised loss, CRNN^{WL} was more difficulty misguided by the coarse labels and acquired better classification results compared with the results of CRNN^0, with an increase of an average of 2.19% in OA. Moreover, CRNN showed the best overall results, benefiting from applying both the attention module and weakly supervised loss function. Without a data fusion module, that is, only Sentinle-2 data was utilized to train the CRNN, the CRNN^{S2} produced lower quantitative results than CRNN, with a decrease of an average of 3.00% in OA and 3.84% in mIoU, proving that multiple satellite data fusion benefits the feature extraction and land-cover classification.

As for the qualitative results, when the attention module was trained in the CRNN framework, the result of CRNN^{AT} shown in Fig. 10(f) more easily identified the class with fewer training samples, such as the water class, compared with the result of CRNN^0 shown in Fig. 10(e). This also proved that the attention module is efficient for improving the classification performance of the model trained on imbalanced samples. Meanwhile, as can be seen from Fig. 10(e) and (g), it is evident that CRNN^{WL} could discern more successive roads than CRNN^0, which benefits from the weakly supervised learning strategy. However, the absence of the attention module in CRNN^{WL} caused difficulty in identifying the water class for the CRNN^{WL} model, further showing the superiority of the attention module. Besides, as shown in Fig. 10(h), the mapping result of CRNN^{S2} heavily misclassified the low vegetation as impervious because of the lack of Sentinel-1 features, which contain more discerning feature information about the impervious class. Similarly, as shown in Fig. 11, the CRNN method acquired satisfactory results in accurately portraying the water and impervious classes.

In general, the results of ablation experiments indicate that the CRNN framework is effective in high-resolution LCM tasks, and the various settings employed in the CRNN model are reasonably designed. This model and its settings effectively utilize reliable information from low-resolution labels, producing accurate and stable high-resolution land cover maps.

VI. Conclusion

With the increasing availability of high-resolution satellite data, the demand for high-resolution LCM has become imperative and urgent. However, generating accurate and current high-resolution land cover maps remains challenging, primarily due to the substantial requirement of high-resolution labels in traditional supervised methods. To overcome this challenge, this study presents a novel CNN-based framework, termed CRNN, specifically designed to produce high-resolution land cover maps using only low-resolution LCM products as training labels. The qualitative and quantitative results showcase the superiority of the CRNN method over several state-of-the-art approaches in classification performance. Moreover, the ablation study reveals that each module of the CRNN framework can effectively exploit the coarse labels to supervise the training process more reasonably. In addition, leveraging the CRNN framework, a 10-m resolution land cover map for Beijing and Shanghai is generated by utilizing 30-m resolution LCM products as training labels. Furthermore, as a compelling application, CRNN demonstrates

its potential in reusing existing data products, contributing to the advancement of sustainable development goals in the field of LCM.

ACKNOWLEDGMENT

The authors would like to thank the ESA for providing the Sentinel-1 and Sentinel-2 data. The authors would also like to thank the European Union for providing the ESA10 data and the team led by Prof. Liangyun Liu at the University of Chinese Academy of Sciences for providing GLC_FCS30 products.

REFERENCES

[1] Z. Li, H. Zhang, F. Lu, R. Xue, G. Yang, and L. Zhang, "Breaking the resolution barrier: A low-to-high network for large-scale high-resolution land-cover mapping using low-resolution labels," *ISPRS J. Photogrammetry Remote Sens.*, vol. 192, pp. 244–267, 2022.

[2] Y. Liu, Y. Zhong, A. Ma, J. Zhao, and L. Zhang, "Cross-resolution national-scale land-cover mapping based on noisy label learning: A case study of China," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 118, 2023, Art. no. 103265.

[3] P. Gong et al., "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," *Sci. Bull.*, vol. 64, pp. 370–373, 2019.

[4] X. Zhang, L. Liu, X. Chen, Y. Gao, S. Xie, and J. Mi, "GLC_FCS30: Global land-cover product with fine classification system at 30m using time-series Landsat imagery," *Earth System Sci. Data*, vol. 13, no. 6, pp. 2753–2776, 2021.

[5] B. Chen et al., "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," *Sci. Bull.*, vol. 64, pp. 370–373, 2019.

[6] D. Zanaga et al., "ESA worldcover 10m 2020 v100," Zenodo, Oct. 2021, doi: 10.5281/zenodo.5571936.

[7] M. Pal and P. M. Mather, "Support vector machines for classification in remote sensing," *Int. J. Remote Sens.*, vol. 26, no. 5, pp. 1007–1011, 2005.

[8] M. A. Friedl and C. E. Brodley, "Decision tree classification of land cover from remotely sensed data," *Remote Sens. Environ.*, vol. 61, no. 3, pp. 399–409, 1997.

[9] M. Belgiu and L. Drăguţ, "Random forest in remote sensing: A review of applications and future directions," *ISPRS J. Photogrammetry Remote Sens.*, vol. 114, pp. 24–31, 2016.

[10] Z. Han et al., "Comparing fully deep convolutional neural networks for land cover classification with high-spatial-resolution Gaofen-2 images," *ISPRS Int. J. Geo- Inf.*, vol. 9, no. 8, 2020, Art. no. 478.

[11] K. Karra, C. Kontgis, Z. S.-Weil, J. C. Mazzariello, M. Mathis, and S. P. Brumby, "Global land use/land cover with Sentinel 2 and deep learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 4704–4707.

[12] L. Yu et al., "FROM-GLC plus: Toward near real-time and multi-resolution land cover mapping," *GIScience Remote Sens.*, vol. 59, no. 1, pp. 1026–1047, 2022.

[13] M. F.-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321–331, 2018.

[14] P. Gong et al., "Finer resolution observation and monitoring of global land cover: First mapping results with landsat TM and ETM+ data," *Int. J. Remote Sens.*, vol. 34, no. 7, pp. 2607–2654, 2013.

[15] R. Dong, W. Fang, H. Fu, L. Gan, J. Wang, and P. Gong, "High-resolution land cover mapping through learning with noise correction," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022.

[16] C. Robinson et al., "Global land-cover mapping with weak supervision: Outcome of the 2020 IEEE GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3185–3199, 2021.

[17] X. Qi, Y. Wang, J. Peng, L. Zhang, W. Yuan, and X. Qi, "The 10-meter winter wheat mapping in Shandong province using Sentinel-2 data and coarse resolution maps," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 9760–9774, 2022.

[18] P. Kaiser, J. D. Wegner, A. Lucchi, M. Jaggi, T. Hofmann, and K. Schindler, "Learning aerial image segmentation from online maps," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6054–6068, Nov. 2017.

[19] M. C. Hansen and T. R. Loveland, "A review of large area monitoring of land cover change using Landsat data," *Remote Sens. Environ.*, vol. 122, pp. 66–74, 2012.

[20] S. P. Abercrombie and M. A. Friedl, "Improving the consistency of multitemporal land cover maps using a hidden Markov model," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 2, pp. 703–713, Feb. 2016.

[21] A. Mousivand and J. J. Arsanjani, "Insights on the historical and emerging global land cover changes: The case of ESA-CCI-LC datasets," *Appl. Geography*, vol. 106, pp. 82–92, 2019.

[22] M. Buchhorn, M. Lesiv, N.-E. Tsendbazar, M. Herold, L. Bertels, and B. Smets, "Copernicus global land cover layers—collection 2," *Remote Sens.*, vol. 12, no. 6, 2020, Art. no. 1044.

[23] J. Chen et al., "Global land cover mapping at 30m resolution: A pok-based operational approach," *ISPRS J. Photogrammetry Remote Sens.*, vol. 103, pp. 7–27, 2015.

[24] J. Yang and X. Huang, "The 30m annual land cover dataset and its dynamics in China from 1990 to 2019," *Earth System Sci. Data*, vol. 13, no. 8, pp. 3907–3925, 2021.

[25] Z. Li, W. He, M. Cheng, J. Hu, G. Yang, and H. Zhang, "SinoLC-1: The first 1-meter resolution national-scale land-cover map of China created with the deep learning framework and open-access data," *Earth System Sci. Data Discuss.*, vol. 2023, pp. 1–38, 2023.

[26] P. Gong et al., "Mapping essential urban land use categories in China (EULUC-China): Preliminary results for 2018," *Sci. Bull.*, vol. 65, no. 3, pp. 182–187, 2020.

[27] J. Wickham, S. V. Stehman, D. G. Sorenson, L. Gass, and J. A. Dewitz, "Thematic accuracy assessment of the NLCD 2016 land cover for the conterminous United States," *Remote Sens. Environ.*, vol. 257, 2021, Art. no. 112357.

[28] L. Yang et al., "A new generation of the United States national land cover database: Requirements, research priorities, design, and implementation strategies," *ISPRS J. Photogrammetry Remote Sens.*, vol. 146, pp. 108–123, 2018.

[29] C. G. Marston, A. W. O'Neil, R. D. Morton, C. M. Wood, and C. S. Rowland, "LCM2021–the UK land cover map 2021," *Earth System Sci. Data Discuss.*, vol. 15, pp. 4631–4649, 2023.

[30] H. K. Zhang and D. P. Roy, "Using the 500m MODIS land cover product to derive a consistent continental scale 30 m Landsat land cover classification," *Remote Sens. Environ.*, vol. 197, pp. 15–34, 2017.

[31] J. Lee, J. A. Cardille, and M. T. Coe, "BULC-U: Sharpening resolution and improving accuracy of land-use/land-cover classifications in Google Earth engine," *Remote Sens.*, vol. 10, no. 9, 2018, Art. no. 1455.

[32] M. Schmitt, L. Hughes, C. Qiu, and X. Zhu, "SEN12MS—a curated dataset of georeferenced multi-spectral sentinel-1/2 imagery for deep learning and data fusion," *ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 42, pp. 153–160, 2019.

[33] M. C.-Loor, M. Hadjikakou, and B. A. Bryan, "High-resolution wall-to-wall land-cover mapping and land change assessment for Australia from 1985 to 2015," *Remote Sens. Environ.*, vol. 252, 2021, Art. no. 112148.

[34] T. Hermosilla, M. A. Wulder, J. C. White, and N. C. Coops, "Land cover classification in an era of big and open data: Optimizing localized implementation and training data selection to improve mapping outcomes," *Remote Sens. Environ.*, vol. 268, 2022, Art. no. 112780.

[35] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Convolutional neural networks for large-scale remote-sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 645–657, Feb. 2017.

[36] C. Robinson et al., "Large scale high-resolution land cover mapping with multi-resolution data," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 12 726–12 735.

[37] M. Schmitt, J. Prexl, P. Ebel, L. Liebel, and X. Zhu, "Weakly supervised semantic segmentation of satellite images for land cover mapping-challenges and opportunities," *ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 5, pp. 795–802, 2020.

[38] Z. Li et al., "The outcome of the 2021 IEEE GRSS data fusion contest—track MSD: Multitemporal semantic change detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1643–1655, 2022.

[39] Y. Huang, Y. Wang, Z. Li, Z. Li, and G. Yang, "Simultaneous update of high-resolution land-cover mapping attempt: Wuhan and the surrounding satellite cities cartography using L2HNet," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 2492–2503, 2023.

[40] B. Hu et al., "Improving urban land cover classification with combined use of Sentinel-2 and Sentinel-1 imagery," *ISPRS Int. J. Geo- Inf.*, vol. 10, no. 8, 2021, Art. no. 533.

[41] Q. Weng, "Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends," *Remote Sens. Environ.*, vol. 117, pp. 34–49, 2012.
[42] Y. Zhang et al., "Mapping annual forest cover by fusing PALSAR/PALSAR-2 and MODIS NDVI during 2007–2016," *Remote Sens. Environ.*, vol. 224, pp. 74–91, 2019.
[43] N. Clerici, C. A. V. Calderón, and J. M. Posada, "Fusion of Sentinel-1A and Sentinel-2A data for land cover mapping: A case study in the lower Magdalena region, Colombia," *J. Maps*, vol. 13, no. 2, pp. 718–726, 2017.
[44] A. Mercier et al., "Evaluation of Sentinel-1 and 2 time series for land cover classification of forest–agriculture mosaics in temperate and tropical landscapes," *Remote Sens.*, vol. 11, no. 8, 2019, Art. no. 979. [Online]. Available: https://www.mdpi.com/2072-4292/11/8/979
[45] S. Šćepanović, O. Antropov, P. Laurila, Y. Rauste, V. Ignatenko, and J. Praks, "Wide-area land cover mapping with Sentinel-1 imagery using deep learning semantic segmentation models," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 10 357–10 374, 2021.
[46] M. Drusch et al., "Sentinel-2: ESA's optical high-resolution mission for GMES operational services," *Remote Sens. Environ.*, vol. 120, pp. 25–36, 2012.
[47] C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation," *Remote Sens. Environ.*, vol. 8, no. 2, pp. 127–150, 1979.
[48] A. Huete, H. Liu, K. Batchily, and W. V. Leeuwen, "A comparison of vegetation indices over a global set of TM images for EOS-MODIS," *Remote Sens. Environ.*, vol. 59, no. 3, pp. 440–451, 1997.
[49] B.-C. Gao, "NDWI—a normalized difference water index for remote sensing of vegetation liquid water from space," *Remote Sens. Environ.*, vol. 58, no. 3, pp. 257–266, 1996.
[50] R. Malinowski et al., "Automated production of a land cover/use map of Europe based on Sentinel-2 imagery," *Remote Sens.*, vol. 12, no. 21, 2020, Art. no. 3523.
[51] N. You et al., "The 10-m crop type maps in Northeast China during 2017–2019," *Sci. data*, vol. 8, no. 1, 2021, Art. no. 41.
[52] L. Zhong, L. Hu, and H. Zhou, "Deep learning based multi-temporal crop classification," *Remote Sens. Environ.*, vol. 221, pp. 430–443, 2019.
[53] N. You and J. Dong, "Examining earliest identifiable timing of crops using all available Sentinel 1/2 imagery and Google Earth engine," *ISPRS J. Photogrammetry Remote Sens.*, vol. 161, pp. 109–123, 2020.
[54] I. Emelyanova, O. Barron, and M. Alaibakhsh, "A comparative evaluation of arid inflow-dependent vegetation maps derived from Landsat top-of-atmosphere and surface reflectances," *Int. J. Remote Sens.*, vol. 39, no. 20, pp. 6607–6630, 2018.
[55] N. He, L. Fang, and A. Plaza, "Hybrid first and second order attention UNet for building segmentation in remote sensing images," *Sci. China Inf. Sci.*, vol. 63, pp. 1–12, 2020.
[56] J. McGlinchy, B. Johnson, B. Muller, M. Joseph, and J. Diaz, "Application of UNet fully convolutional neural network to impervious surface segmentation in urban environment from high resolution satellite imagery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 3915–3918.
[57] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.*, 2015, pp. 234–241.
[58] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020.
[59] M. Gomroki, M. Hasanlou, and P. Reinartz, "STCD-EffV2T Unet: Semi transfer learning EfficientNetV2 T-Unet network for urban/land cover change detection using Sentinel-2 satellite images," *Remote Sens.*, vol. 15, no. 5, 2023, Art. no. 1232.
[60] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
[61] D. G. Lee, Y. H. Shin, and D.-C. Lee, "Land cover classification using SegNet with slope, aspect, and multidirectional shaded relief images derived from digital surface model," *J. Sensors*, vol. 2020, pp. 1–21, 2020.
[62] H. Pan et al., "Updating of land cover maps and change analysis using GlobeLand30 product: A case study in Shanghai metropolitan area, China," *Remote Sens.*, vol. 12, no. 19, 2020, Art. no. 3147.
[63] S. Xie, L. Liu, X. Zhang, and J. Yang, "Mapping the annual dynamics of land cover in Beijing from 2001 to 2020 using Landsat dense time series stack," *ISPRS J. Photogrammetry Remote Sens.*, vol. 185, pp. 201–218, 2022.

Xiaoman Qi received the B.S. degree in engineering from China University of Geosciences, Beijing, China, in 2018, where she is currently working toward the Ph.D. degree in engineering with the School of Land Science and Technology.

Her research interests include deep learning and remote sensing image processing.

Junhuan Peng was born in Chongqing, China, in 1964. He received the B.S. degree in surveying and mapping and the M.S. degree in mine surveying from Central South University, Changsha, China, in 1985 and 1988, respectively, and the Ph.D. degree in geodesy and surveying engineering from Wuhan University, Wuhan, China, in 2003.

From 1988 to 2003, he was a Lecturer and then a Professor with the Guilin University of Technology, Guilin, China. From 2003 to 2005, he was a Postdoctoral Researcher in geodesy and surveying engineering with Shanghai Astronomical Observatory, China Academy of Sciences, Shanghai, China. From 2005 to 2007, he was a Professor with Chongqing University. Since 2007, he has been working as a Full Professor with the China University of Geosciences, Beijing, China. He has authored or coauthored more than 40 papers in Chinese and English in both international and national journals.

Dr. Peng was a Reviewer for several famous surveying and mapping and related academic journals.

Yuebin Wang (Member, IEEE) received the Ph.D. degree in engineering from the School of Geography, Beijing Normal University, Beijing, China, in 2016.

He was a Postdoctoral Researcher with the School of Mathematical Sciences, Beijing Normal University. He is currently an Associate Professor with the School of Land Science and Technology, China University of Geosciences, Beijing. His research interests include remote sensing imagery processing and 3-D urban modeling.

Xiaotong Qi received the Ph.D. degree in photogrammetry and remote sensing from the School of Geosciences and Surveying Engineering, China University of Mining and Technology, Beijing, China, in 2020.

She is currently a Lecturer with the School of Marine Technology and Geomatics, Jiangsu Ocean University, Lianyungang, China. Her research interests include hyperspectral remote-sensing image processing and application.

Yun Peng received the M.S. degree in engineering from China University of Geosciences, Beijing, China, in 2014.

He is currently a Senior Engineer with PowerChina Zhongnan Engineering Corporation Limited, Changsha, China. His research interests include photogrammetry and remote sensing, geographic information system, etc.