

Simultaneous Update of High-Resolution Land-Cover Mapping Attempt: Wuhan and the Surrounding Satellite Cities Cartography Using L2HNet

Yan Huang¹, Yuqing Wang¹, Zhanbo Li, Zhuohong Li¹, and Guangyi Yang¹

Abstract—Land-cover mapping is important for urban planning and management, and current land-cover mapping products are unable to meet the needs of cities due to frequent land surface changes. In this study, based on the low-to-high network (L2HNet) network, we generate a high-resolution land-cover mapping product for Wuhan and its surrounding areas. In this article, we adopt a simplified L2HNet by removing the confident area selection and the L2H loss module to shorten the cycle time of the entire mapping process. The mapping process used ESA LandCover (2021) as low-resolution labels and Google Maps as high-resolution remote sensing images. In the course of the experiment, we also calculate the four indicators mean intersection over union (MIOU), overall accuracy (OA), frequency weighted intersection over union (FWIoU), and Kappa, evaluate the accuracy of our product in predicting fine feature structure using a point-based test method, and compare it with six mainstream land-cover mapping products. The product achieves a 1m-resolution land-cover product in study areas while maintaining an over 75.21% MIOU. OA, FWIoU, and Kappa all maintain values above 85.00%, showing excellent prediction results. In quantitative analysis, compared to ESA LandCover(2021), the L2HNet product has a significant improvement in mapping accuracy for build-up and permanent water, including an exciting 21.08% improvement in permanent water accuracy and an amazing improvement in build-up. The comparison with mainstream products also shows the credibility and practicality of the product. The end result of this research fills a gap in Wuhan and its surrounding areas' 1m-resolution land-cover mapping product. While significantly improving the product's resolution, L2HNet makes time- and labor-saving periodic mapping a reality.

Index Terms—Image segmentation, land-cover mapping, low-to-high task, multiresolution.

I. INTRODUCTION

LAND-COVER refers to the physical coverage of the Earth's surface and is also regarded as one of the important observed variables. For instance, land-cover products

Manuscript received 11 November 2022; revised 16 December 2022; accepted 2 February 2023. Date of publication 8 February 2023; date of current version 13 March 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 42071322. (Corresponding author: Guangyi Yang.)

Yan Huang, Yuqing Wang, Zhanbo Li, and Guangyi Yang are with the Electronic Information School, Wuhan University, Wuhan 430072, China (e-mail: eis_hy@whu.edu.cn; wangsir@whu.edu.cn; zhanbo_li@whu.edu.cn; ygy@whu.edu.cn).

Zhuohong Li is with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430072, China (e-mail: ashelee@whu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3243281

could serve as the primary information sources for managing water resources [1] and the classification of forest species [2]. Land-cover mapping can directly provide the appearance characteristics of natural landscapes and an informative description of human development, which enjoys great significance in some applications, such as agricultural research, infrastructure planning, and resource development planning [3]. In many aspects, the land-cover is updated from the local to the global intermittently, while the demands for regularly updated high-precision land-cover information grows heavily [4]. For example, it is crucial to meet the urgent demand for the constantly updated high-resolution (HR) land-cover information because of climate change and natural disasters, which cause the land-cover maps to fall out of date rapidly. Therefore, efficient large-scale land-cover mapping has great significance for social development. In the early stage, the low-resolution (LR) land-cover products were generally, since the low/medium spatial resolution images were the only ones available. With the improvement of the spatial resolution of remote sensing images [5], [6], we are allowed to get much easier access to the high spatial resolution images, which makes it possible to constantly update the HR land-cover products and obviously increase the accuracy of land-cover classification.

Since the late 1980s, using multispectral images with LR for land-cover mapping has become a research focus area [4], and a variety of machine learning methods have been used to tackle the LR mapping tasks, including support vector machine (SVM) [7], decision tree (DT) [8], and random forest (RF) [9]. However, the spectral or texture information of the LR images was used as input in these pixel-based approaches, which does not consider enough neighborhood information and causes noise in the mapping process.

Because of the lack of neighborhood information, researchers found that these pixel-based methods are not suitable for HR mapping tasks. In order to reduce the generation of noise in HR land-cover mapping, the object-based image analysis (OBIA) method was proposed. In the OBIA method, an object's shape, size, spatial and spectral characteristics have been incorporated directly into the classification process, leading to greatly improved mapping accuracy. Though the OBIA method has been applied to some research, the low efficiency of the manually derived rule-based classification scheme hinders the application of the OBIA method to large-scale automatic land-cover mapping.

With the boom in deep learning techniques, land-cover mapping based on convolution neural networks (CNN) has been progressing at a dramatic pace [6], [10]. Using deep learning techniques, we can greatly reduce the economic and time consumption and obtain large-scale HR land-cover products with higher accuracy in the era of remote sensing Big Data [11]. In Wang et al.'s [12] research, OBIA is combined with CNN and a new object-scale adaptive convolutional neural network (OSA-CNN) for high-speed remote sensing image classification, which effectively improves the accuracy of image classification is proposed. In Huang et al.'s [13] study, a semitransfer deep convolutional neural network (STDCNN) approach is proposed to produce a land-cover mapping product with an overall accuracy (OA) of 91.25%, a Kappa coefficient of 0.903, and an OA of 80.00% and a Kappa of 0.78 for land-cover classification in Shenzhen, which has excellent mapping accuracy.

However, owing to supervised classification methods rely on training data collection, the mapping coverage of these methods which need manually annotated features is severely limited by the time-consuming and laborious manual annotation [14], [15]. In the experiments of Brown et al. [16], they proposed a new method to continuously generate HR land-cover maps at 10m-spatial resolution on a global scale by using deep learning and Sentinel-2 images. In order to train the fully convolutional neural network, a great effort was made in collecting reference data. About 4000 Sentinel-2 images were manually annotated by experienced image interpreters. All annotators were asked to annotate at least 70% of the images within a maximum time of 60 min per image. This problem is common in deep learning-based HR land-cover mapping. In Li et al.'s study [17], [18], change detection is a geospatial application for social good, and its development is then limited by the slow development of remote sensing image tagging technology and outdated classification labels. They proposed a change cross-detection approach for multitemporal semantic change detection with weak, noisy, and LR labels based on label improvements and multimodel fusion. The suggested approach won the first place prize in both phases of the 2021 Data Fusion Contest's MSD track. In Wang et al.'s [19] study, they proposed a new coarse-to-fine deep learning-based land-use change detection method, which also makes an attempt for fast and efficient land-cover mapping. In the study of Claudia et al. they proposed an operating system for fast, efficient, and unsupervised automatic generation of HR large-scale land-cover maps [20]. Therefore, the pursuit of fast and efficient land mapping methods has become a new research focus.

Nowadays, the available HR land-cover products covering the areas of China are still limited. Using the global sample set with a 30m-resolution around the 2015, Gong et al. created FROM-GLC30, which had reached an accuracy of 72.43% [21]. Lately, by employing a pixel-based classification system on Landsat imagery, Zhang et al. [22] made a global product with an accuracy of 82.5% called GLC_FCS30. Through the pixel- and OBIA-based classification method, GlobeLand30(2020) was made with an accuracy of 72.23% [23]. In the past three years, many institutions and scholars have proposed global land-cover products with higher resolution, which have the highest resolution were provided by ESRI_LandCover (2021) [24],

ESA_LandCover(2021) [25], and FROM_GLC10 [26]. In 2019, Gong et al. [21] employed the classifiers of RF to create a 10m-resolution global land-cover called FROM-GLC10 (2019), which reached an accuracy of 75.9%. These products have a wide range of application scenarios and high reference value, and have played an important role in many fields. Under the study of Mi et al. [27] GlobeLand30 has an important reference value for the updating of land-cover mapping used in change detection. In the study by Dong et al. [28], they demonstrated the contribution of FROM-GLC10 in understanding Chinese agricultural systems. However, the land-cover products with 30m- and 10m-resolutions still cannot meet the needs of local applications. For the city-scale, the applications of urban planning and surveys generally require land-cover products with higher resolution. The need for HR land-cover mapping is ubiquitous in the following application scenarios.

- 1) Urban land use mapping is an important but challenging task in the field of remote sensing. Despite the fact that a variety of classification techniques have been created to acquire land use information in urban areas, their accuracy and efficacy are insufficient to meet the needs of real-world applications like urban planning and land management [13]. Wu Shuosheng et al. [29] proposed a classification of detailed urban land use based on geometrical, textural, and contextual information of land parcels using a case study in Austin, Texas, where 50 parcel attributes were tested to classify nine urban land uses, overcoming the heterogeneity of urban land use categories in land use classification. However, because the GIS data necessary for land use classification are either not readily available or are too expensive to acquire, the classification may not be applicable to rapidly expanding metropolitan areas in developing countries. L2HNet enables rapid mapping and can provide the latest land-cover mapping products for developing countries.
- 2) Points of Interest (POI) data can link the informal world of everyday human discourse and the formal world of geographic information system (GIS). According to the survey, there are no reports on the use of POI data and satellite data together to produce detailed land use maps. To fill this gap, Hu Tengyun et al. [30] developed a protocol to map urban land use using medium-resolution satellite images and POI data to determine detailed land use categories in urban areas. The global pattern of urban land use distribution was well reflected. However, the accuracy of this method is relatively low for most secondary urban land use category applications. L2HNet can realize 1 m HR land-cover mapping with high accuracy, which can reflect the land use situation of urban areas relatively accurately, and the update cycle is short, which has more real-time value to promote urban development.
- 3) Land use/cover change is a key aspect of global environmental change and in a sense indicates the impact of human activities on the natural environment [31]. The University of Maryland Global Land Analysis and Discovery (GLAD) team developed and implemented an automated Landsat data processing system in 2020 that generates

globally consistent analysis-ready data (GLAD ARD) as input for land-cover and land use mapping and change analysis. Together, the GLAD ARD dataset and the ARD analysis and characterization tools provide an end-to-end solution for national and regional users for Landsat-based complimentary natural resource assessment and monitoring [32]. L2HNet enables rapid cycle mapping and real-time monitoring of land-cover changes to better reflect changes in the natural environment. L2HNet enables rapid cycle mapping and can monitor land cover changes in real time, thus reflecting changes in the natural environment in a more timely manner.

Wuhan is the capital of Hubei Province, the only subprovincial city and megacity in the six central provinces, an important industrial base, a science and education base, and a comprehensive transportation hub in China. This region is the location for the majority of Wuhan's universities as well as numerous high-tech businesses, well-known tourist attractions, a significant number of residential neighborhoods, and industrial zones, creating a complex urban environment that makes the land-cover product in this area more representative. In the process of Wuhan's development, urban planning, natural resource management, and epidemic prevention all need a comprehensive and detailed survey of land-cover. Therefore, it is significant to have a periodic HR land-cover product covering Wuhan and its surrounding cities.

As society's demand for HR land-cover mapping grows, the low-to-high task has advanced [17], [18]. In Dong et al.'s study, they proposed a solution that could produce 3m-resolution land-cover maps across the country without involving human effort, a major breakthrough in the low-to-high task [33]. In the context of the demand for higher resolution land-cover mapping, a low-to-high network for large-scale high-resolution (L2HNet) [34] is used in this article, which breaks the barrier of resolution between the HR remote sensing images and the LR training labels. Technically speaking, HR land-cover maps can be created just by using LR labeled data and HR remote images by using L2HNet. Because the mapping process doesn't need any HR labels or related auxiliary information, the method can accelerate the automatic updating pipeline of the large-scale HR land-cover maps, which greatly reduces the time consumption of manual annotating of data.

In this article, the Confident area selection (CAS) module and Low-to-high (L2H) loss of L2HNet method are removed while retaining the RP-backbone of L2HNet in order to improve the efficiency of land-cover mapping. Based on the simplified L2HNet method, we generated a 1m-resolution land-cover product covering Wuhan and its surrounding cities. In order to evaluate the performance of the product based on the condensed L2HNet, we constructed labeled maps of groundtruth (GT) and examined the classification quality of the product as well as attributes like accuracy and picture consistency. For more refined mapping results, a test point-based accuracy measurement method was employed to test more accurate mapping results. We contrasted the output with other widely used land-cover mapping to show its dependability. To the best of our knowledge, this is the first attempt to develop large-scale land-cover mapping since the L2HNet approach was proposed. Besides addressing the gap

in 1m-resolution land-cover mapping in China, we are able to save time and effort for large-scale land-cover mapping.

The rest of this article is organized as follows: Section II explains the working mechanism of L2HNet and the method of getting time-saving land-cover products with 1m-resolution. Section III, analyzes the performance of the obtained L2HNet land-cover mapping product and compare it with common land-cover mapping. Finally, Section IV concludes his article.

II. METHOD

The study focuses on four cities in Hubei, including one provincial capital and three large cities, which represent a variety of urban sizes, landscape characteristics, and urbanization intensities.

This section will outline the workflow of HR (1 m) land-cover mapping. The whole workflow is comprised of three stages, as illustrated in Fig. 1. The following subsections will provide an introduction to related work and Fig. 1's three phases.

A. Related Work

During the complete process of L2HNet mapping in Li's experiments, they set CAS module and L2H loss to utilize the unmatched LR labels more reasonably due to the mismatched resolution between their labels and the predicted product. The CAS module splits predictions into confident area and vague area according to the confidence probability maps for the subsequent calculation of the L2H loss. Finally, the L2H loss, including a modified weak-supervised cross entropy loss and a novel self-supervised dynamic vague area loss, is calculated based on the split areas and concentrated features. In the experiments, L2HNet has shown outstanding low-to-high mapping performance on sufficient experiments in the USA [34].

In our mapping process, we remove the CAS module and the L2H loss module in order to use a simplified L2HNet. We have improved the L2HNet method in our experiments, which are not exactly the same as Li's, and the specific optimization strategies and reasons for the improvements are as follows.

- 1) *Optimization Strategies*: In Li's experiment [34], in order to calculate the loss and further tweak the prediction outcomes, they separated the acquired prediction maps of the RP-backbone output into confidence areas and vague areas. In the simplified L2HNet method, the CAS module and the L2H loss module are removed, leaving only the RP-backbone. The output of the RP-backbone is then used as the actual prediction result.
- 2) *Reasons for the Improvements*: In Li's experiment [34], a 30m-resolution training label was used for the 1m-resolution land-cover mapping product. The resolution span between 30 and 10 m is large and there is a big difference in the products, therefore the prediction of the same part as the label will have a greater degree of confidence. The CAS module can extract the more confidence areas of the prediction results and pass the simulation results of the confidence areas into the L2H loss module for calculation, while the vague areas are not passed into the L2H loss module. After such an operation, it is possible to better

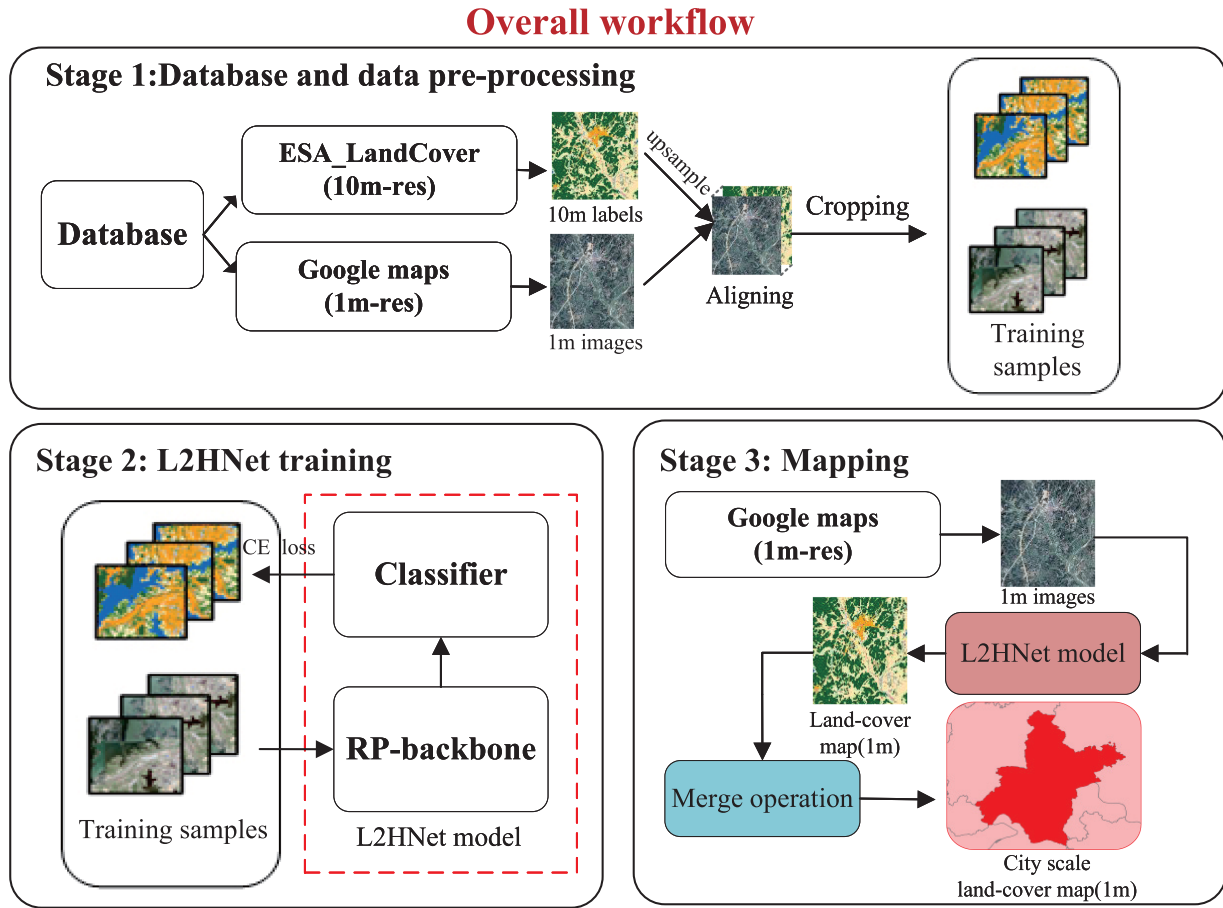


Fig. 1. Overall workflow of the high-resolution (1 m) land-cover mapping in Hubei.

retain the part with better prediction results and make a larger adjustment for the areas with less confidence. In this experiment, the training label is 10m-resolution, which has a good match with the prediction results of 1m-resolution. In the prediction results, most of them have a good degree of confidence, so even after dropping the CAS module and L2H loss module, we can still get good prediction results. After eliminating these two modules, the simplified L2HNet method is optimized in terms of operation time and is more practical.

In model training, the model is trained with a batch size of 16. The RP-backbone is made to reliably extract from the images hierarchical features with HR representations. The research establish numerous RP-blocks in the RP-backbone, as seen in Fig. 2. In each block, we set three convolution layers with sizes of 1×1 , 3×3 , and 5×5 to sample the input map. We refer to the input, middle, and fusion feature maps of the b th block as $I^{(b)}$, $M^{(b)}$, and $F^{(b)}$, respectively. As shown in Fig. 2, the kernel numbers of multiscale convolution layers are inversely proportional with their kernel sizes. The 1×1 kernels which have the largest kernel numbers encourage the block to extract feature details. Secondary, kernels with larger sizes (i.e., 3×3 and 5×5) provide necessary neighboring information. The convolution kernels in each layer are set to the proportion

in Fig. 2. Therefore, the parallel layers can scan them with appropriate receptive fields instead of the reduction of feature resolution caused by over down-sampling $I^{(b)}$. In particular, a 3×3 convolution input layer with four input channels (the bands R-G-B-NIR of images) and C_I output channels is used to create the input feature map of the first block.

In the input layer, RP-backbone collects information from Google Maps images to form a dense input feature map, and then a large number of RP-blocks continue to integrate features while maintaining the size, channel, and resolution of the map in the process.

B. Database and Data Preprocessing

To train our model in the initial stage of this research, LR labels and HR remote sensing images are required, but there is a lack of HR images that are suitable and available for China. HR satellite images from commercial US satellites like Worldview and Planet are available in large areas but expensive, while commercial Chinese satellites like Gaofen-2 and Jilin-1 are free in partial areas. However, their images have incomplete coverage. Google Maps with 1m-resolution, which can give vector maps comprising information on international urban regions, are chosen for HR remote sensing photos, and ESA_LandCover(2021)

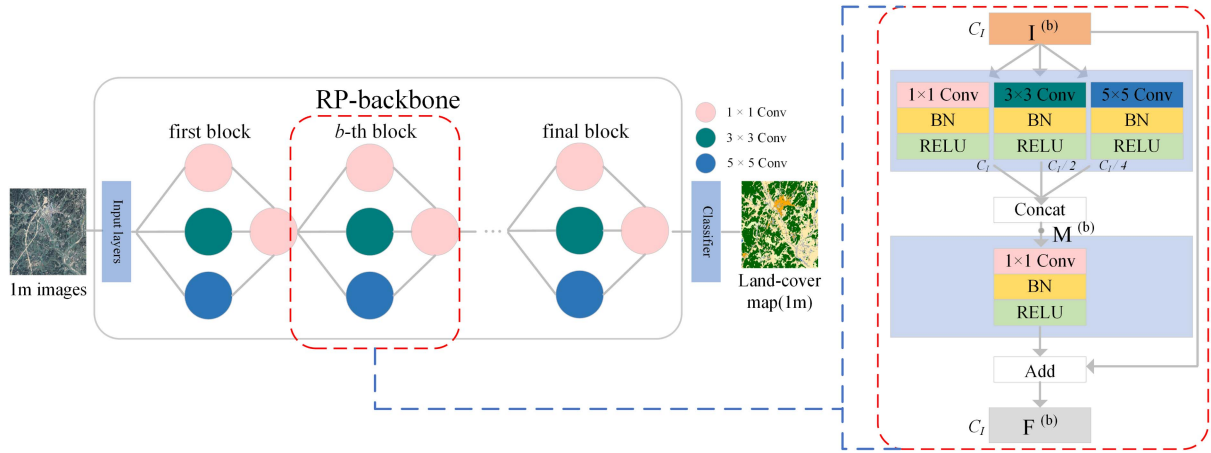


Fig. 2. Illustration of the RP-block. The input map $I^{(b)}$ is sampled by three convolution layers with sizes of 1×1 , 3×3 , and 5×5 .

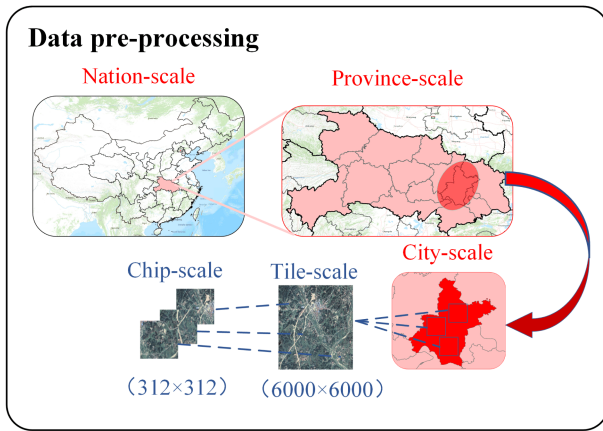


Fig. 3. Illustration of the data preprocessing.

are chosen as labels since it gives access to high-precision data for the most recent 10, 30, and 100 m products.

In the preprocessing module, since Wuhan and the surrounding cities are the focus of our investigation, we need to crop Google Maps to the province-scale first.

In the overall workflow, data preprocessing mainly includes two parts: 1) aligning and 2) cropping. The experiment first aligns Google Maps and ESA_LandCover(2021) for each city after upsampling 10m-labels for each city. As shown in Fig. 3, every city is cropped into several tiles with a size of 6000×6000 . Then, as training samples, every tile will be randomly cropped into 200 chips with a size of 312×312 . In order to avoid splicing gaps in the final map, our images' overlap is 156.

C. L2HNet Training

At this stage, each model is trained separately. In order to obtain a better fitting effect, the experiments train each model using the corresponding city, which takes the various development stages and geographic dispersion of each city into account. Therefore, training models with variable feature patterns can be eventually obtained by this training method.

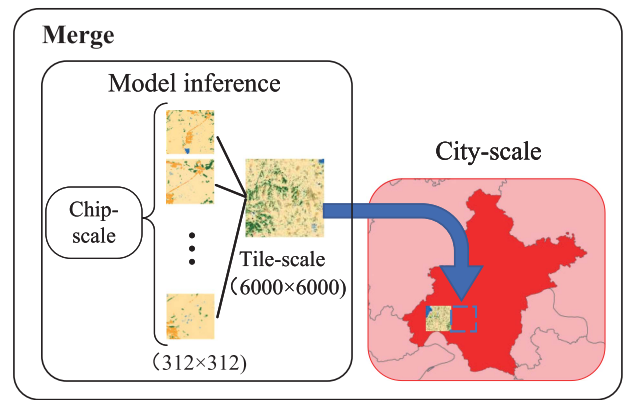


Fig. 4. Illustration of merging.

D. Mapping

After completing the model training, the experiment selects some 1m images from Google Maps as the input in the last stage. In model training, the experiment uses chips as input during model training. Therefore, to generate a city-scale image, these chips need to be merged to generate a city-scale image in the last stage. As Fig. 4 shows, during the inference process, the experiment merge chip-scale to tile-scale and finally arrange 6000×6000 tile-scale images in order to get city-scale images.

It will take two to three weeks to complete the mapping procedure at the province-scale. The only process that requires manual down/up loading is data download. The operations of preprocessing, training, and mapping are all highly automated which means a time-saving and labor-saving mapping process.

III. EXPERIMENT

The results for Wuhan and three additional representative cities are provided in Fig. 5. These products demonstrate that the characteristics of the land-cover can be well depicted by the L2HNet, leading to satisfactory classification results. Fig. 5 shows the L2HNet product of Wuhan and its surrounding cities. L2HNet product can be used to distinguish between smaller

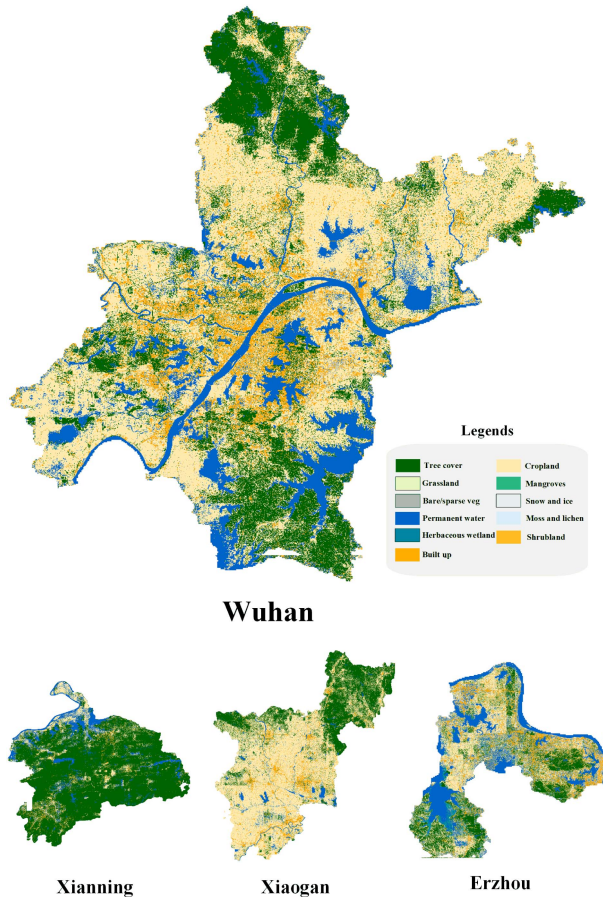


Fig. 5. L2HNet product of Wuhan and its surrounding cities.

or finer structures that are visible in contemporary urban environments. The product not only offers improved building recognition accuracy, but the higher resolution also enables the product to offer more details in land-cover products. For the growth of towns and cities as well as the usage of land, this is of paramount strategic importance.

A. Experiment Setting

Fig. 5 also displays the classification system applied in this investigation. Eleven land-cover classes are adopted, including tree cover, shrubland, grassland, cropland, built-up, bare/sparse veg, snow and ice, permanent water, herbaceous wetland, mangroves, moss and lichen. This legend is modified from the legend of ESA_Landcover (2021), which has the same land classes but different colors. In this study, the model is implemented on the GeForce RTX 3090 GPUs. The optimizer and the criterion are separately set as AdamW and cross-entropy loss. The weight for the loss function is determined based on the frequency of occurrence of each category. The initial learning rate is set at 0.01, and it will be reduced to 10% of the current learning rate when the loss stops decreasing after two epochs. During the training process, the model was trained for 80 epochs with a batch size of 16. Additionally, the default value of the random seed is set to 0 in order to ensure the training model is reproducible.

TABLE I
EXPERIMENT SETTING IN TRAINING

Experiment setting	
GPU	GeForce RTX 3090
Optimizer	AdamW
Criterion	Cross-Entropy
Initial learning rate	0.01
Learning rate decline adjustment	10% of the current learning rate every two undeclared epochs
Epochs	80
Batch size	16
Random seed	0

The experimental training process equipment and parameters are shown in Table I.

B. Experiment Results

1) *Qualitative Analysis*: In order to better demonstrate the effectiveness of the L2HNet-based land-cover mapping product for the Wuhan region, the following four metrics were computed for the product.

- Mean intersection over union (MIoU)*: MIoU is a standard metric for semantic segmentation and is an important evaluation metric in recently semantic segmentation problems. In the problem of semantic partitioning, intersection over union (IoU) is the ratio of the intersection between the true label and the predicted value of the class. And MIoU is the average of the intersection ratio of each class in the dataset.
- Overall accuracy (OA)*: Overall accuracy is the ratio between the number of correct predictions made by the model on all test sets and the overall number.
- Frequency weighted intersection over union (FWIoU)*: FWIoU is a slight enhancement on MIoU. Compared to MIoU, FWIoU sets weights for each class based on its frequency of occurrence, which is an important evaluation metric in practical semantic segmentation problems.
- Kappa coefficient*: Kappa Coefficient is a measure of classification accuracy that is used to measure the degree of agreement between two models when judging the same image, and the results range from 0 to 1. 0.0–0.20 very low agreement (slight), 0.21–0.40 fair agreement (fair), 0.41–0.60 moderate agreement (moderate), 0.61–0.80 high agreement (substantial), and 0.81–1 almost perfect agreement (almost perfect).

The MIoU and the OA indicate the accuracy in categories and pixels, respectively. The FWIoU combined the category occurrence frequencies to assess the accuracy of the prediction results. The Kappa Coefficient measures the extent to which two models agree when judging the same image. These four metrics have a wide range of applications in the quality analysis of semantic segmentation tasks [35], [36], [37]. During the evaluation we replaced one of the models with a GT result.

TABLE II
QUALITATIVE ANALYSIS OF 20 IMAGES

Images	MIou	OA	FWIoU	Kappa
Image1	70.78%	74.53%	87.16%	88.40%
Image2	74.81%	97.03%	79.75%	78.92%
Image3	78.56%	82.48%	92.01%	85.49%
Image4	79.84%	82.63%	91.76%	89.40%
Image5	70.68%	74.52%	88.67%	86.11%
Image6	73.87%	76.85%	92.45%	91.96%
Image7	76.76%	90.14%	93.44%	93.41%
Image8	71.73%	77.12%	84.64%	86.55%
Image9	70.92%	99.05%	92.40%	92.39%
Image10	76.49%	91.81%	86.81%	88.74%
Image11	76.19%	80.01%	87.79%	91.21%
Image12	74.14%	94.65%	84.91%	85.40%
Image13	74.20%	76.91%	90.98%	93.16%
Image14	74.61%	85.01%	95.84%	96.62%
Image15	77.01%	91.66%	58.75%	61.57%
Image16	76.95%	96.52%	84.52%	85.38%
Image17	74.87%	95.75%	80.77%	80.99%
Image18	76.99%	78.68%	95.97%	94.57%
Image19	79.90%	90.57%	88.72%	87.27%
Image20	75.08%	78.82%	88.44%	87.83%
AVERAGE	75.21%	85.73%	87.28%	87.26%
VAR	$7.6*10^{-4}$	$7.1*10^{-3}$	$6.5*10^{-3}$	$5.6*10^{-3}$

The research randomly selected 20 sets of images as test samples. These coefficients calculated from the test samples are provided in Table II. The accuracy throughout the 20 images is quite consistent, suggesting reasonably reliable mapping outcomes.

The average MIoU score across all images in these test samples is 75.21%, the average FWIoU and Kappa scores are always above 80%, with average values of 87.28% and 87.26%, respectively, and the average OA score is 85.23%. Individual images scored extremely well, demonstrating the excellent mapping results of the L2HNet product. Meanwhile, the overall variance of each indicator is at a low level, indicating that our mapping results have a stable play across land-cover classes. In order to visually represent the predicted results of the product, we have selected five out of test samples and set their close-up images, GT label in Fig. 6. In the five selected images, it is easy to visualize the prediction effect of the L2HNet product in towns as well as in the countryside. In densely built-up areas, L2HNet products offer a significant improvement in accuracy over LR land-cover mapping products. Since the accuracy for buildings is an important components of urban scenes, L2HNet product can be regarded as satisfactory. The average FWIoU (93.40%) and the Kappa scores (93.14%) in these five images are extremely good. This indicates that common land classes in urban scenes as well as in suburban scenarios can be predicted well in the L2HNet product. Although in the figure we still find some errors in the predictions for some water bodies, there is

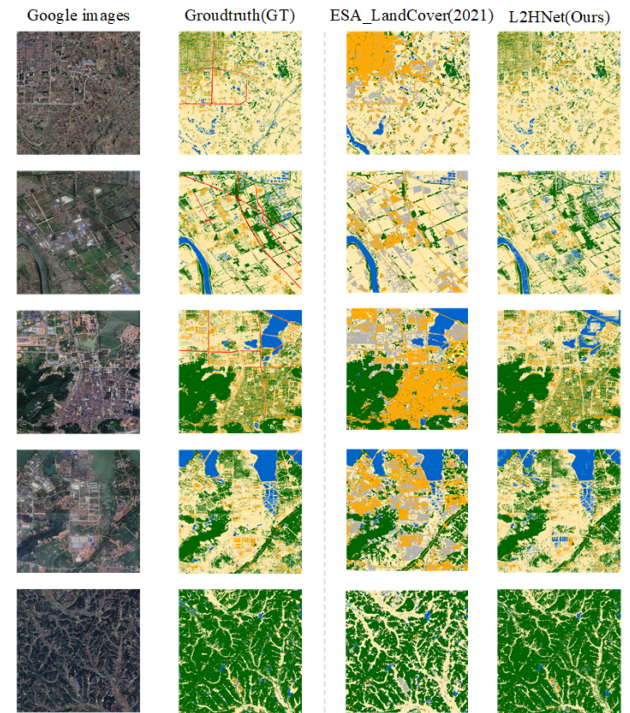


Fig. 6. Illustration of close-up images, GT label, and the corresponding L2HNet result for some sample areas in Wuhan.

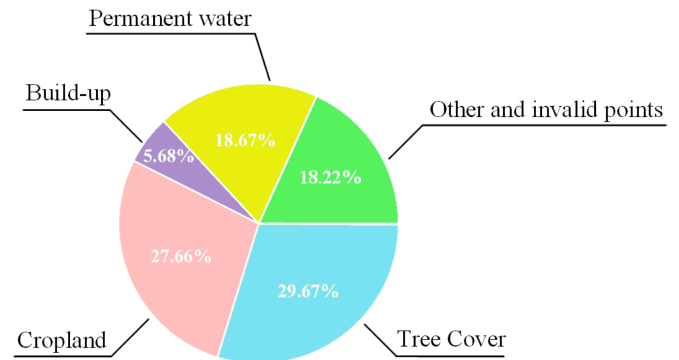


Fig. 7. Illustration of the proportion of each land-cover class in all test sites.

little variation in the accuracy among the 20 images, implying fairly robust mapping results. It is important to note that we discovered that the road will play a significant role in the 1m-resolution land-cover mapping when comparing GT labels with the anticipated images. As a result, we highlighted the obvious road in red on the GT labels, even though no such land class appeared in our prediction results.

2) *Quantitative Analysis*: In the four cities, a total of 900 test points were generated. Specifically, 300 test points were generated in Wuhan. The other three cities contain 200 test points each. To avoid large measurement errors due to the small sample size and to better verify the validity of the product, land-cover classes that account for a relatively large proportion of the experiment should be selected for testing. Fig. 7 shows the percentage of each land type among the 900 test sites. Based on the percentage, tree cover, build-up, cropland, and permanent

TABLE III
QUANTITATIVE ANALYSIS IN 4 CITIES

	ESA_LandCover(2021)				L2HNet(ours)			
	Tree cover	Cropland	Build-up	Permanent water	Tree cover	Cropland	Build-up	Permanent water
Wuhan	48.10%	95.00%	7.41%	58.67%	72.15%	80.00%	62.96%	70.67%
Xiaogan	88.57%	93.47%	70.00%	16.67%	88.57%	83.70%	80.00%	54.16%
Erzhou	67.00%	71.88%	50.00%	65.91%	79.07%	70.31%	100.00%	72.73%
Xianning	95.54%	76.47%	37.50%	48.00%	95.83%	48.48%	42.85%	76.00%
Accuracy	Wuhan	48.27%	Xiaogan	67.50%	Wuhan	67.43%	Xiaogan	65.00%
	Erzhou	61.50%	Xianning	75.50%	Erzhou	61.50%	Xianning	73.47%

The bolded data in the table indicate higher accuracy rates in L2HNet products than in ESA-LandCover (2021).

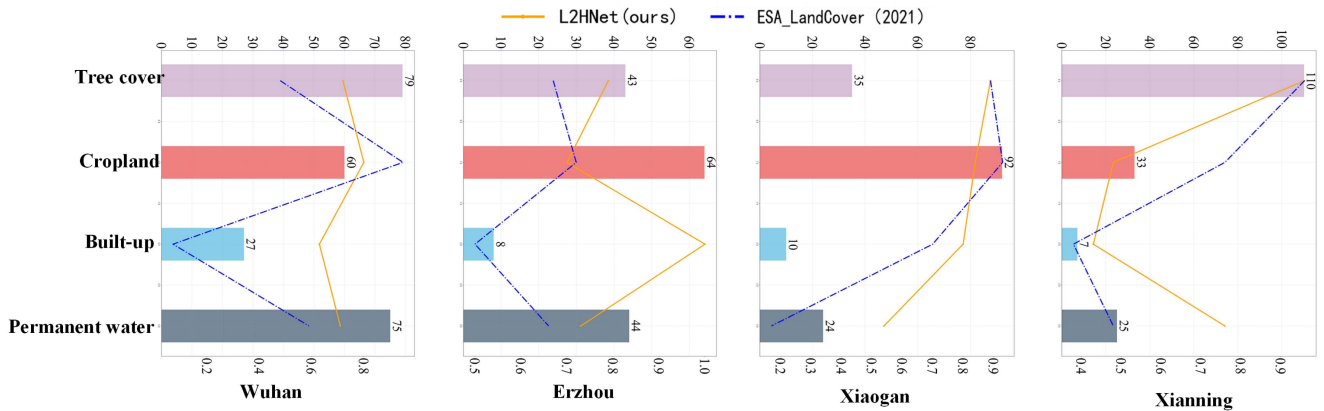


Fig. 8. Illustration of the statistical results of L2HNet and ESA_LandCover (2021) on four common land classes in Wuhan and its surrounding cities.

water can be obtained as the larger proportions of all land-cover classes. Since the point generation process in the experiment is random, a few test points will be generated at the edge of the land-cover mapping product, and these points are considered invalid points. The number of invalid points is small and has no effect on the experimental results.

Table III indicates the accuracy of these four land-cover classes in each city. The bolded part of the table indicates that the L2HNet product outperformed ESA_LandCover (2021) in this test result. The product exhibits a significant improvement in the accuracy of four land-cover classes that have a larger proportion of all classes in the data study in Wuhan region. Hubei’s provincial capital city, Wuhan, contains a wide variety of dense urban structures. Build-up’s features are frequently poorly detected in the LR product. In addition, when pursuing very fine point measurement processes, results frequently fail to meet expectations in the LR product. Build-up’s accuracy in the L2HNet offering is 63%, a significant improvement over ESA_LandCover(2021)’s accuracy of 7%. The accuracy rate for cultivated land has dropped, but it is still at a high level, and the result of 80% accuracy is still extremely outstanding.

The distribution of land-cover classes differs among cities, Fig. 8 shows the number of test points for each of the major land-cover classes in each city. Unlike Wuhan, which is the capital city of the province, the other three cities differed

significantly from Wuhan in terms of urban development and building density. In Wuhan, where building density is great, the L2HNet product is far ahead of ESA_LandCover (2021) in terms of OA. In places where building density is low, the L2HNet product nevertheless maintains a robust level of accuracy, with an accuracy difference of no more than 2.5% compared to ESA_LandCover (2021) in each city. L2HNet’s mapping outcomes in these cities continue to perform well in terms of building recognition, even hitting 100% accuracy in Ezhou. The data also demonstrate that the mapping outcomes for aquatic bodies have improved, with an average accuracy that is 21.08% greater than that of ESA_LandCover(2021). Comparing 16 indicators across four cities demonstrates that L2HNet has a strong capability for producing HR mapping findings. It is important to note that the L2HNet product delivers a 1m-resolution land-cover mapping product and can periodically map swiftly, producing a time-saving HR mapping result while preserving a more robust result.

3) Comparative Products: In Fig. 9, we illustrate five areas with a high number of fine features of Wuhan to demonstrate the mapping performance of the L2HNet product, such as dense areas of buildings, water bodies, forests, suburbs, etc., and compare them with the current more mainstream land-cover mapping products. In order to more intuitively represent the semantic segmentation effect of the products, we have made some adjustments to the legends of the mainstream products. For similar land

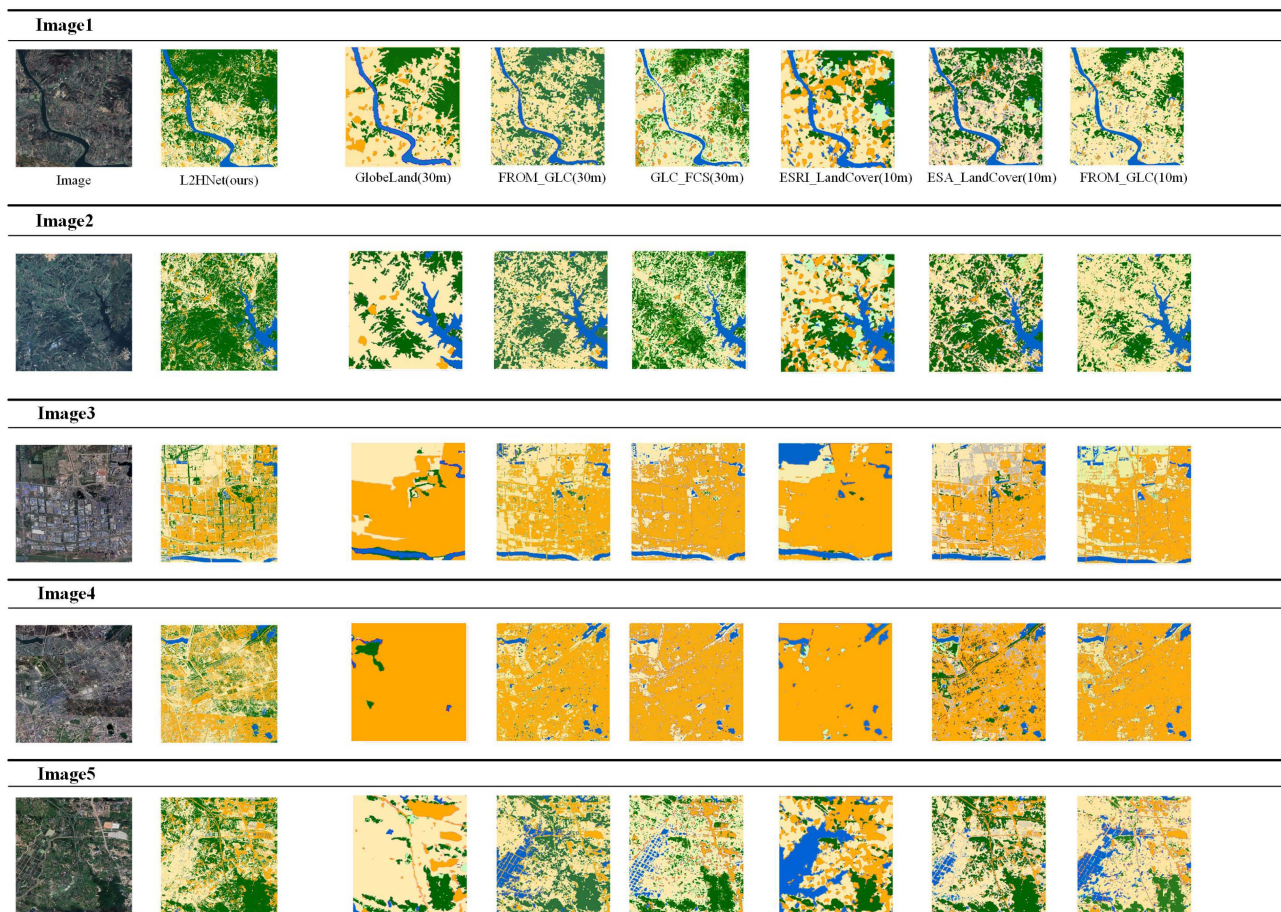


Fig. 9. L2HNet, Global_Land, FROM_GLC30 (2013), GLC_FCS30 (2021), ESRI_LandCover (2021), ESA_LandCover (2021), and FROM_GLC10 (2019) Insets for five representative sample areas in Wuhan and corresponding remote sensing satellite images.

classes, we use the colors corresponding to those in the L2HNet products to represent them. For land categories specific to other products, we kept the original legend colors. Fig. 10 shows the legend used for each product. By observing the results of Fig. 9, L2HNet product shows robust and accurate mapping capacity in different kinds of situations while the resolution is increased. The legend in Fig. 10 illustrates that the L2HNet product has a moderate number of classifications and provide more diverse land information based on higher resolution mapping results. The outline of build-ups can be clearly observed in the L2HNet product, and land-cover classes with larger areas, such as forests and cropland are represented with better effect. These classes are the main part of the land-cover classes in the city, which means that the L2HNet product has made great progress in the land-cover mapping of the city.

C. Discussion

Based on the Google Maps and ESA_LandCover (2021) dataset, L2HNet products of Wuhan and its surrounding are made along with a good discrimination effect in a relatively short time period. Compared to the land-cover products mentioned above, the L2HNet product maintains a high level of accuracy

while increasing resolution. The results suggest that the L2HNet product in Wuhan and its surrounding cities is capable of meeting the demand for the analysis of citispacial patterns.

To further evaluate the results of land-cover mapping in Wuhan area, Fig. 11 shows L2HNet product and ESA_LandCover (2021) mapping results in Wuhan area. The L2HNet product shows powerful and accurate mapping capabilities and retains a greater degree of land-cover detail in its predictions. In areas with more details, the L2HNet product not only has higher accuracy than other products but also has a strong fitting effect and a detailed display.

D. Limitations and Possible Improvements

Though images in Google Maps are free with a wide coverage, there still exists Mosaic data with inconsistent time in the database. As shown in Fig. 12, there are three different time points in one image tile. Inconsistent time points in the Google images would lead to incredible products.

Although Wuhan's land-cover mapping is somewhat representative of the distribution in most cities in China, there are differences in the development direction and management policies of individual cities, which can lead to differences

Legend of each product in Fig.9

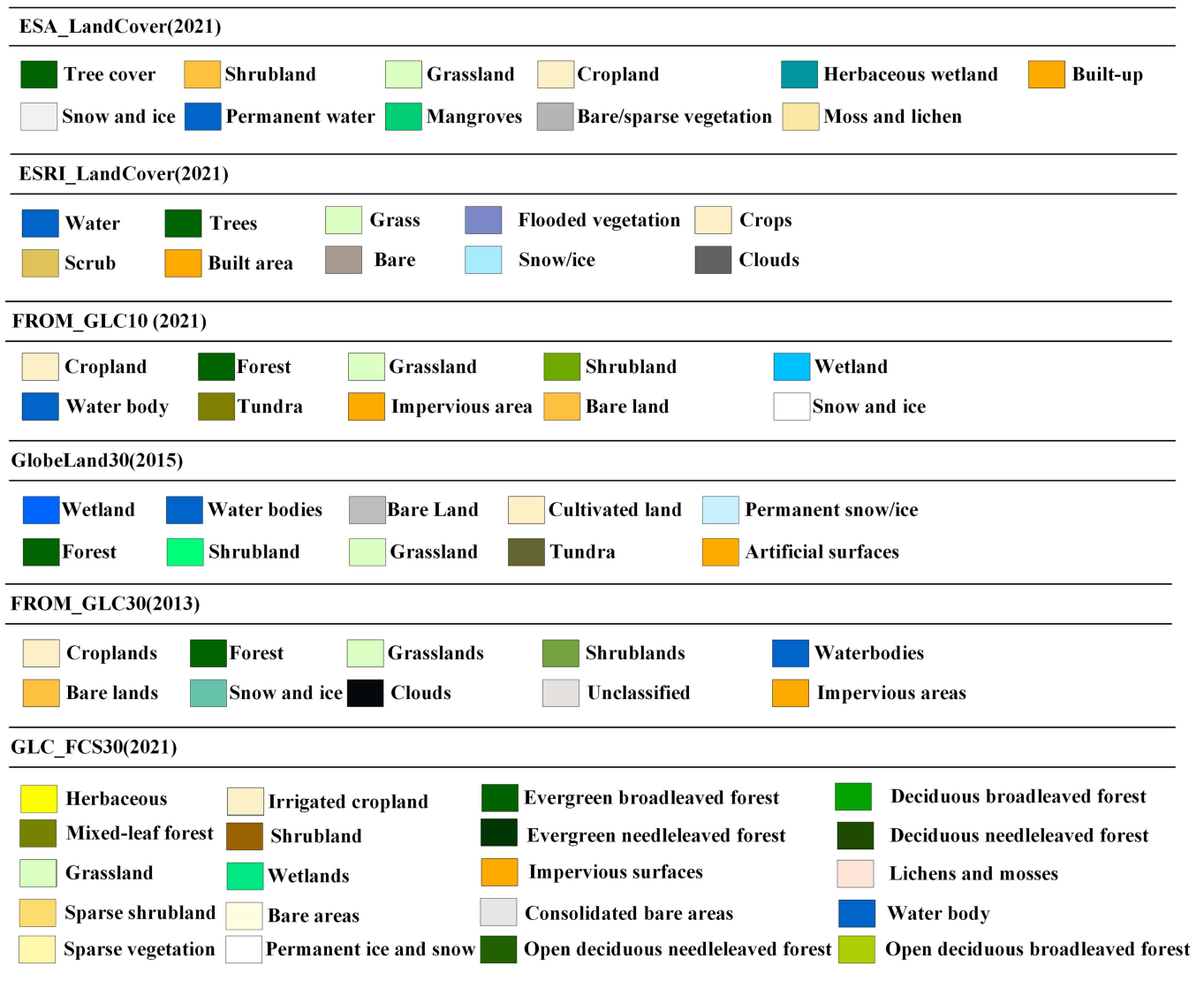


Fig. 10. Legend of each product in Fig. 9.

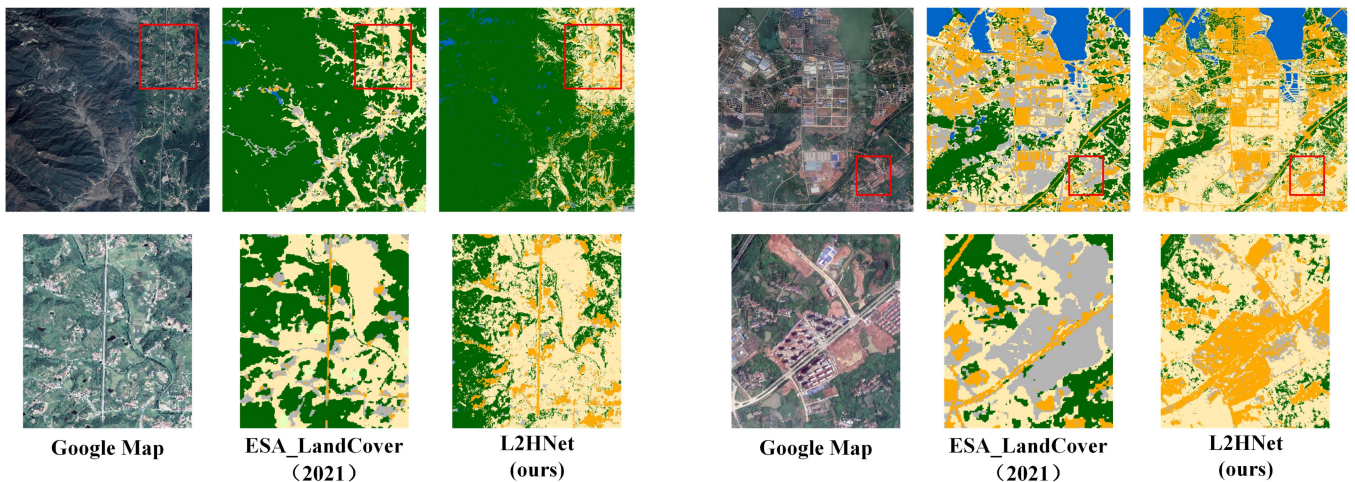


Fig. 11. Qualitative results for the Wuhan area, which shows the L2HNet product and ESA_LandCover (2021) mapping results.

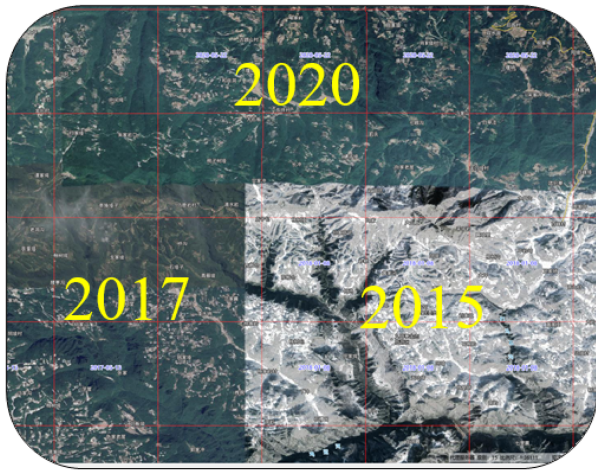


Fig. 12. Illustration of Google Maps dataset. There are three different time points in one image tile. Mosaic data with inconsistent time can result in inaccurate products from the mapping process.

in the distribution of land classes in each city's land-cover mapping product. For example, there are large differences in land distribution between highly modern cities and forest-dominated tourist cities. In the mapping process of this study, the weights of each land-cover classes are adjusted in the mapping process for different cities in order to achieve more accurate mapping results. In large-scale land-cover mapping, adjusting the weights for each city is still a time-consuming and laborious process.

The roads are commonly not considered in the product with 10 or 30 m resolution, but they are obvious in the 1m images. However, the L2HNet product does not achieve good classification results in the identification of road categories. Due to their similar spectral reflectance, the roads are misclassified as cropland or bare/sparse veg. Therefore, the open source street maps could be a valuable complement to the labels.

In the subsequent experiments, we hope to complete the mapping process of entire China after overcoming the shortcomings described above.

IV. CONCLUSION

Highly automated HR land-cover mapping is an important task as the demand for fine and extensive land-cover products grows dramatically. In this article, a land-cover mapping product is proposed based on a simplified L2HNet neural network and also describes the production process of the product. Compared to traditional networks that require properly labeled data sources, such as UNet, Deeplab v3+, and HRNet [38], mapping process based on L2HNet is more efficient without the huge demand for properly labeled data sources. During the mapping process, since the 10m land-cover maps still match well with the 1m images, the separate work of the RP-backbone ensures an efficient mapping process.

Comprehensive experiments on the Google Maps and ESA_LandCover (2021) datasets in the Wuhan area show the superiority of the proposed L2HNet compared to other mainstream land-cover mapping products. In L2HNet products, the

average MIOU reach 75.219%, while the OA reach 85.737%. Also 87.289% FWIoU and 87.267% kappa illustrate the accuracy and high consistency of land-cover mapping, respectively. In areas, such as dense buildings, L2HNet has finer semantic segmentation results, and the OA in quantitative analysis is improved by 20% compared to ESA_LandCover (2021).

Overall, based on the analysis and comparison in this article, the proposed L2HNet product is considered an effective HR land-cover mapping product. The mapping process in Wuhan and surrounding areas has proven the feasibility and high accuracy of the L2HNet network for land-cover mapping, filling the gap of 1m-resolution land-cover mapping in Wuhan. In the future, the prospect of work on the L2HNet network will be on national as well as global scale data. With high quality HR images and a wide selection of LR land-cover products, we will continue to produce HR land-cover mapping products for the China region.

REFERENCES

- [1] C. Toureiro, R. Serralheiro, S. Shahidian, and A. Sousa, "Irrigation management with remote sensing: Evaluating irrigation requirement for maize under Mediterranean climate condition," *Agricultural Water Manage.*, vol. 184, pp. 211–220, 2017.
- [2] Z. Xie, Y. Chen, D. Lu, G. Li, and E. Chen, "Classification of land cover, forest, and tree species classes with ZiYuan-3 multispectral and stereo data," *Remote Sens.*, vol. 11, no. 2, 2019, Art. no. 164.
- [3] T. R. Loveland et al., "Development of a global land cover characteristics database and IGBP discover from 1 km AVHRR data," *Int. J. Remote Sens.*, vol. 21, no. 6/7, pp. 1303–1330, 2000.
- [4] J. Cihlar, "Land cover mapping of large areas from satellites: Status and research priorities," *Int. J. Remote Sens.*, vol. 21, no. 6/7, pp. 1093–1114, 2000.
- [5] J. Li, Y. Ouyang, and T. Chen, "On the (M)-property of monotone measures and integrals on atoms," *Fuzzy Sets Syst.*, vol. 412, pp. 65–79, 2021.
- [6] V. Mazzia, A. Khaliq, and M. Chiaberge, "Improvement in land cover and crop classification based on temporal features learning from Sentinel-2 data using recurrent-convolutional neural network (R-CNN)," *Appl. Sci.*, vol. 10, no. 1, 2020, Art. no. 238.
- [7] S. Ullah et al., "Remote sensing-based quantification of the relationships between land use land cover changes and surface temperature over the lower Himalayan Region," *Sustainability*, vol. 11, no. 19, 2019, Art. no. 5492.
- [8] R. L. Lawrence and C. J. Moran, "The AmericaView classification methods accuracy comparison project: A rigorous approach for model selection," *Remote Sens. Environ.*, vol. 170, pp. 115–120, 2015.
- [9] H. K. Zhang and D. P. Roy, "Using the 500 m 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.
- [10] W. Bupphawat, T. Kasetkasem, I. Kumazawa, P. Rakwatin, and T. Chanwimaluang, "Super-resolution land cover mapping based on deep learning and level set method," in *Proc. 14th Int. Conf. Elect. Engi./Electron., Comput., Telecommun. Inf. Technol.*, 2017, pp. 557–560.
- [11] X. Yang, Z. Chen, B. Li, D. Peng, P. Chen, and B. Zhang, "A fast and precise method for large-scale land-use mapping based on deep learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 5913–5916.
- [12] J. Wang, Y. Zheng, M. Wang, Q. Shen, and J. Huang, "Object-scale adaptive convolutional neural networks for high-spatial resolution remote sensing image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 283–299, 2021.
- [13] B. Huang, B. Zhao, and Y. Song, "Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery," *Remote Sens. Environ.*, vol. 214, pp. 73–86, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0034425718302074>
- [14] J. R. Townshend et al., "Global characterization and monitoring of forest cover using Landsat data: Opportunities and challenges," *Int. J. Digit. Earth*, vol. 5, no. 5, pp. 373–397, 2012.

- [15] K. Nogueira, O. A. Penatti, and J. A. dos Santos, "Towards better exploiting convolutional neural networks for remote sensing scene classification," *Pattern Recognit.*, vol. 61, pp. 539–556, 2017.
- [16] C. F. Brown et al., "Dynamic world, near real-time global 10 m land use land cover mapping," *Sci. Data*, vol. 9, 2022, Art. no. 251.
- [17] Z. Li, F. Lu, H. Zhang, G. Yang, and L. Zhang, "Change cross-detection based on label improvements and multi-model fusion for multi-temporal remote sensing images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 2054–2057.
- [18] 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.
- [19] W. Mingchang, H. Zhang, W. Sun, S. Li, F. Wang, and G. Yang, "A coarse-to-fine deep learning based land use change detection method for high-resolution remote sensing images," *Remote Sens.*, vol. 12, 2020, Art. no. 1933.
- [20] C. Paris, L. Gasparella, and L. Bruzzone, "A scalable high-performance unsupervised system for producing large-scale hr land cover maps: The Italian country case study," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 9146–9159, 2022.
- [21] 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, no. 6, pp. 370–373, 2019.
- [22] X. Zhang, L. Liu, X. Chen, Y. Gao, and J. Mi, "GLC_FCS30: Global land-cover product with fine classification system at 30 m using time-series landsat imagery," *Earth Syst. Sci. Data*, vol. 13, pp. 2753–2776, 2020.
- [23] Q.-H. Zheng et al., "Accuracy comparison and driving factor analysis of LULC changes using multi-source time-series remote sensing data in a coastal area," *Ecological Inform.*, vol. 66, 2021, Art. no. 101457.
- [24] "ESRI_Landcover," [Online]. Available: <https://www.arcgis.com/apps/instant/media/index.html?appid=fc92d38533d440078f17678ebc20e8e2>
- [25] "ESA_Landcover," [Online]. Available: <https://viewer.esa-worldcover.org/worldcover/>
- [26] "FROM_GLC," [Online]. Available: <http://data.ess.tsinghua.edu.cn/fromglc2017v1.html>
- [27] J. Mi, L. Liu, X. Zhang, X. Chen, Y. Gao, and S. Xie, "Impact of geometric misregistration in GlobeLand30 on land-cover change analysis, a case study in China," *J. Appl. Remote Sens.*, vol. 16, no. 1, 2022, Art. no. 014516.
- [28] S. Dong, B. Gao, Y. Pan, R. Li, and Z. Chen, "Assessing the suitability of FROM-GLC10 data for understanding agricultural ecosystems in China: Beijing as a case study," *Remote Sens. Lett.*, vol. 11, no. 1, pp. 11–18, 2020.
- [29] S.-S. Wu, X. Qiu, E. L. Usery, and L. Wang, "Using geometrical, textural, and contextual information of land parcels for classification of detailed urban land use," *Ann. Assoc. Amer. Geographers*, vol. 99, no. 1, pp. 76–98, 2009.
- [30] T. Hu, J. Yang, X. Li, and P. Gong, "Mapping urban land use by using Landsat images and open social data," *Remote Sens.*, vol. 8, no. 2, 2016, Art. no. 151.
- [31] F. Mu, B. Zhu, and H. He, "Environmental dynamic monitoring of Olympic Green based on QuickBird and aerial images," *Proc. SPIE*, vol. 6200, 2006, Art. no. 620012.
- [32] P. Potapov et al., "Landsat analysis ready data for global land cover and land cover change mapping," *Remote Sens.*, vol. 12, no. 3, 2020, Art. no. 426.
- [33] R. Dong et al., "Improving 3-m resolution land cover mapping through efficient learning from an imperfect 10-m resolution map," *Remote Sens.*, vol. 12, 2020, Art. no. 1418.
- [34] 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.
- [35] Y. Liu, Q. Ren, J. Geng, M. Ding, and J. Li, "Efficient patch-wise semantic segmentation for large-scale remote sensing images," *Sensors*, vol. 18, no. 10, 2018, Art. no. 3232.
- [36] Z. Hao and X. Qi, "End-to-end concrete appearance analysis based on pixel-wise semantic segmentation and CIE Lab," *Cement Concrete Res.*, vol. 161, 2022, Art. no. 106926.
- [37] L. Cyr and K. Francis, "Measures of clinical agreement for nominal and categorical data: The kappa coefficient," *Comput. Biol. Med.*, vol. 22, no. 4, pp. 239–246, 1992.
- [38] G. Peng et al., "A new research paradigm for global land cover mapping," *Ann. GIS*, vol. 22, no. 2, pp. 87–102, 2016.



Yan Huang is currently working toward the bachelor's degree in electronic information engineering with Wuhan University, Wuhan, China.

His research interests include semantic segmentation, image processing, and deep learning.



Yuqing Wang is currently working toward the bachelor's degree in communication engineering with Wuhan University, Wuhan, China.

She joined the artificial intelligence special class, Wuhan University. Her research interests include deep learning and image processing.

Ms. Wang was the recipient of the first prize scholarship from the School of Electronic Information, Wuhan University.



Zhanbo Li is currently working toward the bachelor's degree in communication engineering with the College of Electronic Information Technology, Wuhan University, Wuhan, China.

His research interests include in-depth machine learning and land cover mapping.

Mr. Li was the recipient of the second prize scholarship from the School of Electronic Information, Wuhan University.



Zhuohong Li received the B.S. degree in communication engineering in 2020 from Wuhan University, Wuhan, China, where he is currently working toward the Ph.D. degree in photogrammetry and remote sensing with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing (LIESMARS).

His research interests include large-scale land cover mapping, semantic segmentation, and deep learning.

Mr. Li won first place in the 2021 IEEE GRSS Data Fusion Contest—Track MSD: Multitemporal Semantic Change Detection and second place in the 2022 IEEE GRSS Data Fusion Contest—Track SLM: Semi-supervised Land-cover Mapping.



Guangyi Yang received the Ph.D. degree in electronic engineering from Wuhan University, Wuhan, China, in 2018.

He is currently a Senior Engineer with the School of Electronic Information, Wuhan University. His research interests include machine vision and image processing.