



## Article

# Large-Scale Land Cover Mapping Framework Based on Prior Product Label Generation: A Case Study of Cambodia

Hongbo Zhu <sup>1,2</sup> , Tao Yu <sup>1</sup>, Xiaofei Mi <sup>1,\*</sup> , Jian Yang <sup>1</sup>, Chuanzhao Tian <sup>2</sup>, Peizhuo Liu <sup>1,3</sup>, Jian Yan <sup>1</sup>, Yuke Meng <sup>4</sup> , Zhenzhao Jiang <sup>4</sup> and Zhigao Ma <sup>4</sup>

<sup>1</sup> Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China; zhb999322@stumail.ncaie.edu.cn (H.Z.)

<sup>2</sup> School of Remote Sensing and Information Engineering, North China Institute of Aerospace Engineering, Langfang 065000, China

<sup>3</sup> School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing 100191, China

<sup>4</sup> School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo 454000, China

\* Correspondence: mixf@aircas.ac.cn

**Abstract:** Large-Scale land cover mapping (LLCM) based on deep learning models necessitates a substantial number of high-precision sample datasets. However, the limited availability of such datasets poses challenges in regularly updating land cover products. A commonly referenced method involves utilizing prior products (PPs) as labels to achieve up-to-date land cover mapping. Nonetheless, the accuracy of PPs at the regional level remains uncertain, and the Remote Sensing Image (RSI) corresponding to the product is not publicly accessible. Consequently, the sample dataset constructed through geographic location matching may lack precision. Errors in such datasets are not only due to inherent product discrepancies, and can also arise from temporal and scale disparities between the RSI and PPs. In order to solve the above problems, this paper proposes an LLCM framework for generating labels for use with PPs. The framework consists of three main parts. First, initial generation of labels, in which the collected PPs are integrated based on D-S evidence theory and initial labels are obtained using the generated trust map. Second, for dynamic label correction, a two-stage training method based on initial labels is adopted. The correction model is pretrained in the first stage, then the confidence probability (CP) correction module of the dynamic threshold value and NDVI correction module are introduced in the second stage. The initial labels are iteratively corrected while the model is trained using the joint correction loss, with the corrected labels obtained after training. Finally, the classification model is trained using the corrected labels. Using the proposed land cover mapping framework, this study used PPs to produce a 10 m spatial resolution land cover map of Cambodia in 2020. The overall accuracy of the land cover map was 91.68% and the Kappa value was 0.8808. Based on these results, the proposed mapping framework can effectively use PPs to update medium-resolution large-scale land cover datasets, and provides a powerful solution for label acquisition in LLCM projects.

**Keywords:** land cover; prior products; label generation; noisy label; confidence probability



**Citation:** Zhu, H.; Yu, T.; Mi, X.; Yang, J.; Tian, C.; Liu, P.; Yan, J.; Meng, Y.; Jiang, Z.; Ma, Z. Large-Scale Land Cover Mapping Framework Based on Prior Product Label Generation: A Case Study of Cambodia. *Remote Sens.* **2024**, *16*, 2443. <https://doi.org/10.3390/rs16132443>

Academic Editors: Mohamed Lamine Mekhalfi, Yakoub Bazi, Edoardo Pasolli and Mawloud Guermoui

Received: 27 May 2024

Revised: 23 June 2024

Accepted: 26 June 2024

Published: 3 July 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Regularly updated large-scale land cover mapping (LLCM) provides necessary information for land resource surveying, ecological environment assessment, urban spatial planning, crop growth monitoring, and other related applications. A large number of land cover classification prior products (PPs) have been made public to date. Large-scale low-resolution products based on MODIS images include MCD12Q1 [1] products of 500 m and MCD12C1 products of 0.05° for long time series, as well as Copernicus Global Land Services (CGLS) land cover products of 100 m for 2015–2019 [2,3]. In addition, 30 m land cover products based on Landsat series images include GlobeLand products for 2000, 2010, and 2020 [4] and GLC\_FCS30 Fine Dynamic products from 1985 to 2020 [5]. In recent

years, products with 10 m resolution based on Sentinel-2 images have been released, such as FROM\_GLC in 2017 [6], ESA WorldCover v100 in 2020 [7], ESA WorldCover v200 in 2021 [8], and ESRI LandCover [9], produced annually since 2017, as well as the near-real-time Dynamic World product [10]. The production of these products is mostly based on traditional random forest or object-oriented methods, which rely on artificially constructing features and require highly specialized knowledge, and as such cannot meet the efficiency and accuracy needs of LLCM [11].

In recent years, because deep learning method can automatically extract and learn features, it has been gradually applied to land cover mapping tasks in remote sensing images (RSI) [12,13], providing a new possibility for realizing large-scale and high-precision land cover mapping [14]. However, LLCM based on deep learning methods requires a large number of high-precision sample datasets for model training, and labeling in samples requires a high level of professional knowledge and rich interpretation experience of labeling personnel, which greatly increases the cost of labeling and sample collection. The limited availability of sample datasets poses challenges for regularly updating land cover products. For example, the newly released 2017–2023 10-m ESRI LandCover and Dynamic World v1 are based on training models and production of the National Geographic Society's Dynamic World Training dataset [15], which required a great deal of manpower and time.

In order to solve the problem of difficult label acquisition and repeated collection, studies have begun to use existing PPs as sample datasets required for label construction model training [16–18]. For example, using the 500-m MODIS land cover product to derive a consistent continental scale 30 m Landsat land cover classification [19]. The 2017 10-m FROM\_GLC product applied the 2015 all-season land cover mapping sample library [20] to Sentinel-2 images acquired in 2017, using a random forest classifier to generate a 10-m resolution global land cover map. The 2015 GLC\_FCS30 product was produced by taking training samples from the CCI\_LC [21] land cover product [22]. Although these studies have addressed the issue of label acquisition to an extent, the accuracy of PPs at the regional level is uncertain and remote sensing images (RSI) corresponding to products are not publicly available. In addition, there are differences in time and resolution between RSI and PPs for LLCM. As a result, datasets generated from existing public land cover products may contain a large number of labels with inaccurate noise. A dataset with noisy labels causes serious overfitting by the deep learning network, leading to reduced precision [23].

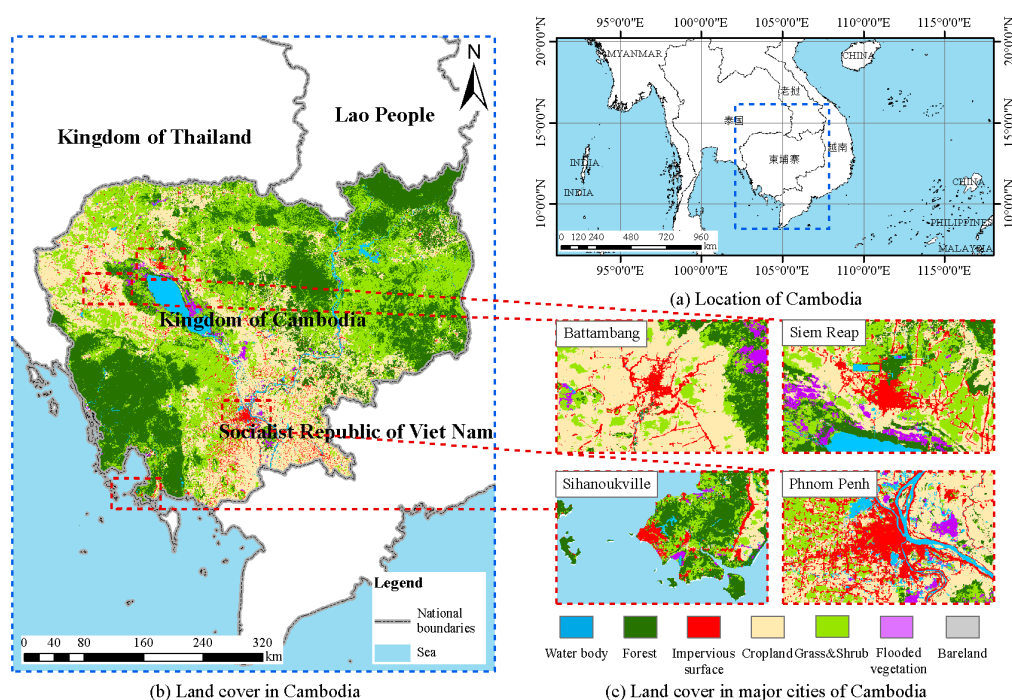
To solve those problems, an LLCM framework based on label generation for PPs is proposed in this paper, which solves the difficulties in obtaining LLCM labels and the problem of labels containing noise, allowing better use of PPs for generating labels and correcting the noise in labels to complete LLCM. In order to make use of multiple products to generate labels, D-S evidence theory is introduced. Based on the regional accuracy of PPs, the evidence of PPs is combined to generate a trusted label that integrates multiple products. In order to correct the noise in the labels, an online noise correction method is proposed which takes into account the confidence probability (CP) and spectral index of the model output to update the labels during the training process and then uses the united noise correction loss to train the model to recover the correct labels from the noise labels. Using the LLCM framework proposed in this paper, a 10-m land cover map of Cambodia for 2020 was produced.

## 2. Study Area and Materials

### 2.1. Study Area

The Kingdom of Cambodia, referred to as Cambodia, is located in the south of Indochina Peninsula in Asia. Its geographical position between latitude 10.5°N to 14.2°N and longitude 102.5°E to 107.5°E borders Laos and Thailand to the north, Vietnam to the east, and the Gulf of Thailand and Gulf of Siam to the south, as shown in Figure 1. The total area of Cambodia is about 181,035 km<sup>2</sup>, and its diverse terrain includes plains, mountains, plateaus, and coastal lowlands. Cambodia's climate is mainly tropical monsoon climate, with the year divided into two seasons: May to October for the rainy season, and November

to April for the dry season. It is warm and humid throughout the year with plenty of rainfall, which is conducive to the growth of various vegetation. Its ecological environment is complex and diverse; a variety of ecosystems meet here, from tropical rainforests to arid grasslands and from high mountains to coastal lowlands. Through in-depth research, we can not only better understand its ecological environment, but can provide scientific basis for resource management and environmental protection in Cambodia. However, due to the complex terrain of Cambodia, frequent clouds and rain, and serious interference from human behavior in some areas, the research encounters certain challenges.

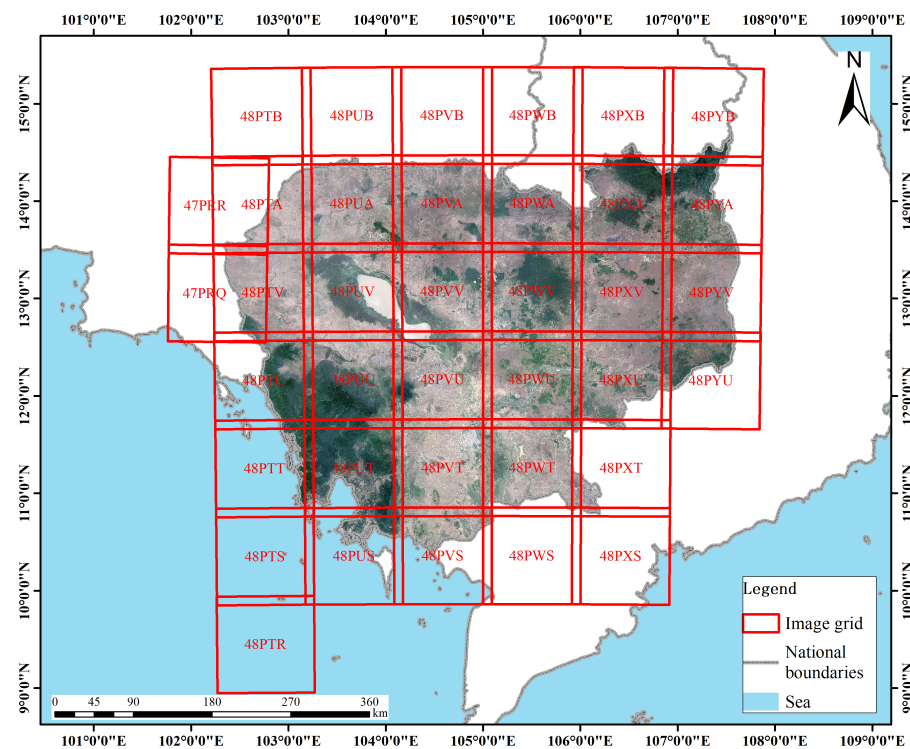


**Figure 1.** The geographical location of Cambodia, showing the overall and major urban land cover (ESRI LandCover): (a) location of Cambodia, (b) land cover in Cambodia, and (c) land cover in major cities of Cambodia.

## 2.2. Images and Preprocessing

In Cambodia, the weather makes it difficult to obtain images without cloud at the same time of year. Google Earth Engine (GEE) [24] is a cloud computing platform that processes satellite image data and other earth observation data. The platform provides global MODIS, Landsat, Sentinel, and other multi-source remote sensing data as well as terrain, climate, and other types of data. Its powerful cloud computing and storage capabilities greatly improve the efficiency of data processing, providing unprecedented opportunities for dynamic study of the Earth system. In this paper, Sentinel-2 L2A data are used. L2A data consist of the bottom of atmospheric reflectance after radiometric calibration and atmospheric correction. Sentinel-2 Cloud Probabilities (S2C), Cloud Displacement Index (CDI), and Directional Distance Transform (DDT) [25] for each cell in the image grid (Figure 2) was used to generate masks to reduce clouds and cloud shadows covering all available Sentinel-2 L2A-class images in the Cambodia region in 2020. The images were synthesized and mosaicked according to the spatial position. Finally, a total of 37 cloudless images of Cambodia in 2020 were obtained, covering about 181,000 km of land surface of Cambodia and some surrounding areas, including nine bands: B2, B3, B4, B5, B6, B7, B8, B11 and B12, all of which were sampled with the nearest neighbor sampling method to a resolution of 10 m. Based on the above process, image data can be better managed and processed to minimize the misclassification of ground objects caused by image quality, clouds, and cloud shadows. In order to facilitate model training, we first used the maximum

value of each band as the denominator to map the original data between 0 and 1, then used the mean value and standard deviation for normalization processing.



**Figure 2.** Sentinel-2 image grid of Cambodia.

### 2.3. PPs and LLCM Taxonomy

Among the existing PPs, most rely on accurately labeled training samples, and the labeling of these samples requires a large amount of labeling cost, which inevitably hinders the rapid updating of LLCM. By integrating multiple land cover products with a resolution of 10–30 m to generate training samples with relatively high accuracy and reliability on a global scale, the cost of obtaining a large number of training samples for LLCM can be greatly reduced while being more stable and reliable than a single product.

Therefore, in this paper we selected five global medium-resolution land cover products with similar primary LLCM taxonomy, three single-class products, and Open Street Map(OSM) data [26] (Open source data). Of the five land cover products, ESA WorldCover (European Space Agency, Paris, France), GLC\_FCS30 (Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China), and Globeland30 (National Geomatics Center of China, Beijing, China) are based on traditional machine learning models (random forests, multi-scale segmentation, etc.) which can provide finer class boundaries, while ESRI LandCover (Environmental Systems Research Institute, Inc., Redlands, CA, USA) and Dynamic World (Google Inc., Santa Clara, CA, USA) are products based on deep learning models that are more accurate in most regions. Three additional products, Global Impervious Surface GISD30 [27] (Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China), Global Flooded Vegetation GWL\_FCS30 [28] (Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China.) and Globe Cropland [29] (University of Maryland, College Park, MD, USA), were selected to improve accuracy. OSM is often used as supplementary data in land cover or land use mapping tasks. The data source, product year, spatial resolution and other information of the nine products are shown in Table 1.

**Table 1.** PPs information.

Reference Data	Image Data	Years	Resolution	Source
Dynamic World	Sentinel-2	2020	10 m	<a href="https://code.earthengine.google.com/">https://code.earthengine.google.com/</a> (accessed on 24 July 2022)
ESRI LandCover	Sentinel-2	2020	10 m	<a href="https://livingatlas.arcgis.com/landcover/">https://livingatlas.arcgis.com/landcover/</a> (accessed on 13 April 2022)
ESA WorldCover	Sentinel-1 Sentinel-2	2020	10 m	<a href="https://esa-worldcover.org/">https://esa-worldcover.org/</a> (accessed on 13 October 2021)
GLC_FCS30	Landsat	2020	30 m	<a href="https://zenodo.org/record/3986872">https://zenodo.org/record/3986872</a> (accessed on 13 October 2021)
Globeland30	Landsat HJ-1 GF-1	2020	30 m	<a href="http://www.globallandcover.com/">http://www.globallandcover.com/</a> (accessed on 21 November 2021)
GWL_FCS30	Sentinel-1 Landsat	2020	30 m	<a href="https://zenodo.org/record/6575731">https://zenodo.org/record/6575731</a> (accessed on 13 August 2021)
GISD30	Landsat	2020	30 m	<a href="https://zenodo.org/record/5220816">https://zenodo.org/record/5220816</a> (accessed on 13 August 2022)
Global cropland	Landsat	2019	30 m	<a href="https://glad.umd.edu/dataset/croplands">https://glad.umd.edu/dataset/croplands</a> (accessed on 13 August 2022)
Open Street Map	-	2020	-	<a href="https://master.apis.dev.openstreetmap.org/">https://master.apis.dev.openstreetmap.org/</a> (accessed on 13 September 2022)

As shown in Table 2, the LLM taxonomy used in this article references five prior products, Dynamic World, ESRI LandCover, ESA WorldCover, GLC\_FCS30, and Globeland30, which use Landsat and Sentinel images as primary data sources. As Cambodia is located in a tropical region, tundra, lichen, snow, and ice were removed from the LLM taxonomy. In addition, due to the relatively small and low accuracy of shrubland and grassland cover in GLC\_FCS30 and Globeland30 in Cambodia, shrubland and grassland were combined into the class of “Grass & Shrub”. Eventually, the LLM taxonomy was simplified into seven categories: water body, forest, impervious surface, cropland, Grass & Shrub, flooded vegetation, and bareland.

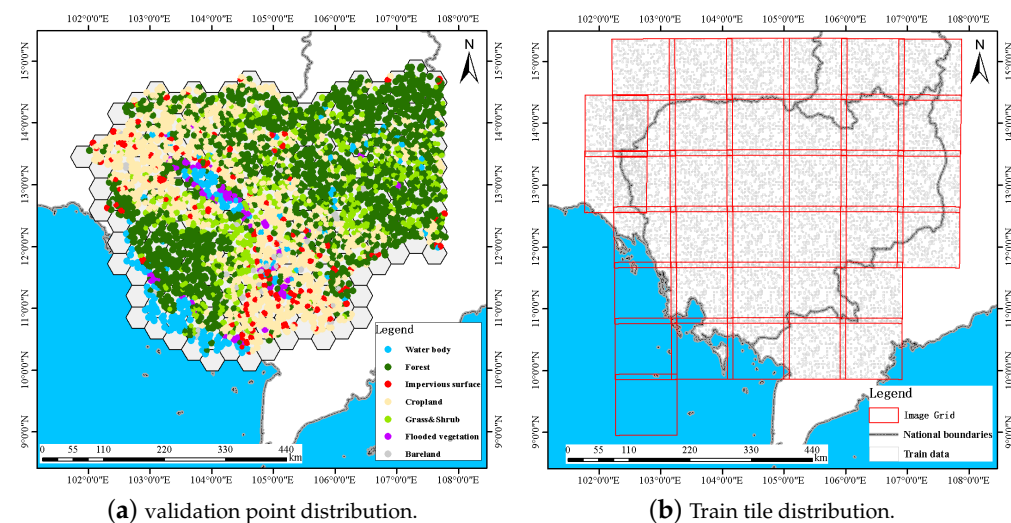
**Table 2.** Taxonomy of PPs and LLM.

LLCM	Dynamic World	ESRI LandCover	ESA WorldCover	GLC_FCS30	GlobeLand30
Water body	Water	Water	Permanent water bodies	Water body	Water bodies
Forest	Trees	Trees	Tree cover	Forest	Forest
Impervious surface	Built area	Built area	Built-up	Impervious surfaces	Artificial surfaces
Cropland	Crops	Crops	Cropland	Cropland	Cultivated Land
Grass & Shrub	Shrub & Scrub Grass	Rangeland	Shrubland Grassland	Shrubland Grassland	Shrubland Grassland
Flooded vegetation	Flooded vegetation	Flooded vegetation	Herbaceous Flooded vegetation Mangroves	Flooded vegetation	Wetland
Bareland	Bare ground	Bare ground	Bare/Sparse vegetation Moss and Lichen	Bare areas	Bareland

## 2.4. Validation and Training Dataset

To assess the accuracy of the land cover map for Cambodia, we annotated individual pixel points with mapping units of  $10 \times 10$  m ( $1 \times 1$  pixel). Cambodia was uniformly divided into hexagonal grids [30,31] with side lengths of  $0.2^\circ$ , and 20 verification points (corresponding to  $10 \times 10$  m pixels) were randomly selected in each hexagonal grid. In order to avoid repeated sampling in the same homogeneous material, the interval of each verification point was at least 2 km. All annotation was performed using Remote sensing data processing software, which provides vector editing tools to directly annotate Sentinel-2 images. It is easier to label categories such as water body, forest, cropland, impervious surface, and bareland at 10-m resolution in Sentinel-2 because these elements tend to occur in fairly uniform plots. The grassland, shrub, and flooded vegetation categories are more challenging to label, and are often confused with each other. Therefore, in addition to Sentinel-2 images, we obtained matched high-resolution satellite images through Google Maps and used ESRI LandCover as an aid for comprehensive judgment to label verification points. The above random sampling method can ensure that the collected verification points are evenly distributed in geographical space as much as possible and that all categories have a certain number of sample points. Finally, 3712 verification points were marked, as shown in Figure 3a.

The training dataset was based on a grid, and the generated initial labels and corresponding images were clipped according to a grid of size  $256 \times 256$  without overlapping. The same number of samples were randomly selected for each cell. For each grid, we randomly selected 20% of the slices as training data. The spatial distribution of the training dataset, which finally contained 13,869 data pairs, is shown in Figure 3b.



**Figure 3.** Distribution of dataset used for validation and training.

## 3. Methods

Figure 4 shows the methodological flow used to produce a land cover map in Cambodia. Based on publicly available land cover products and Sentinel-2 data, this paper completed a 10-m resolution land cover mapping of Cambodia for 2020. In the data processing part, fusion labels were first generated based on D-S evidence theory, then initial labels were obtained by selecting fusion labels using a synchronously generated trust graph. In the label correction part, the pretraining of the label correction model was carried out first, then the pretraining weight training model was loaded, and finally the classification graph and corresponding CP map were predicted during the training process. The designed CP label screening and NDVI label screening modules were used to screen and update the labels, and the joint loss function was calculated using the update label and the initial label, finally obtaining the corrected labels. In the training part of the classification model, the corrected labels and weighted loss function

were used to train the model. Finally, land cover classification and accuracy assessment were completed. The details are described in the following section.

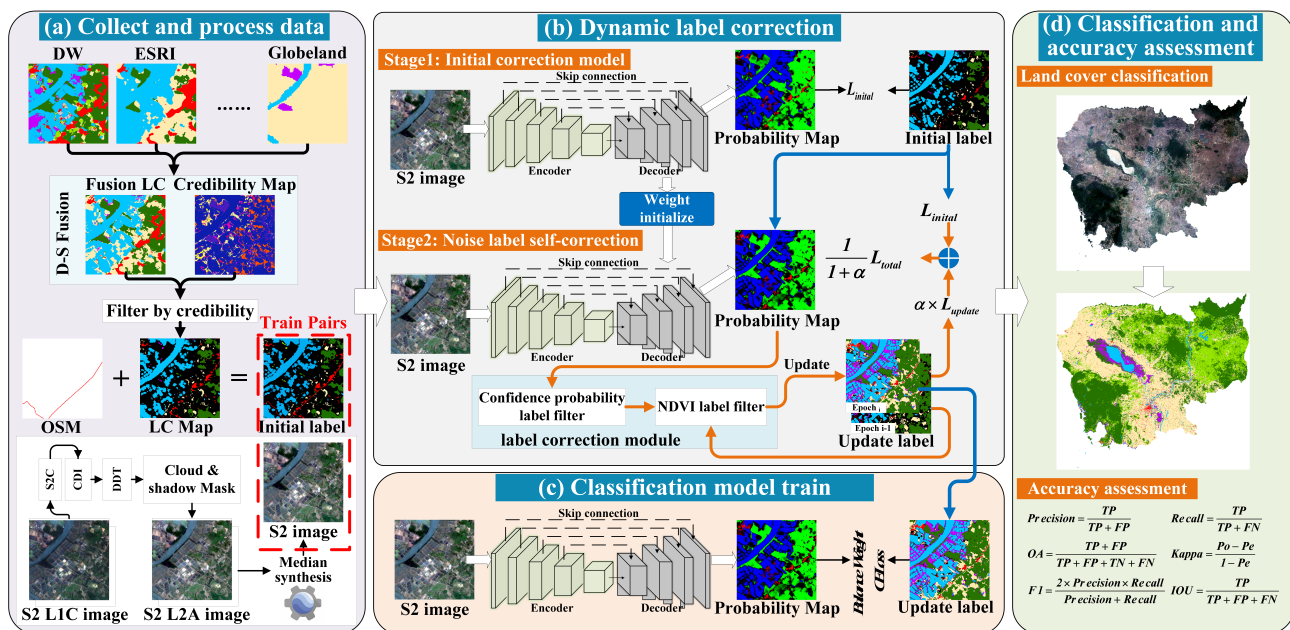


Figure 4. Mapping process of LLCM framework based on multi-source prior product label generation.

### 3.1. Label Generation Based on PPs

The sources of land cover data are diverse, and there are differences in accuracy, classification systems, and spatiotemporal scale. Moreover, there may be uncertain factors such as sensor error and classification algorithm error in the process of land cover data classification and acquisition. Dempster-Shafer (D-S) evidence theory can be used as a method of data fusion to effectively integrate data from these different sources in order to generate more reliable results. Therefore, this paper uses D-S evidence theory for fusion of PPs.

D-S evidence theory, a generalization of probability theory, can express random uncertainty as well as incomplete information and subjective uncertainty [32,33]. The principle of D-S evidence theory is to assume that  $D$  is the set of all possible values of variable  $X$  and that the elements in the set  $\Omega$  satisfy the mutually exclusive relationship. Then, the set  $\Omega$  is the identification framework of variable  $X$ . For the power set  $2^D$  of  $\Omega$ , it constitutes a proposition set. If the function  $m: 2^D \rightarrow [0, 1]$  satisfies  $m(\phi) = 0$  and  $\sum_{A \in D} m(A) = 1$ , then  $m$  is called the Basic Probability Assignment (BPA) and  $m(A)$  is the basic probability number of proposition  $A$ , representing the accuracy of the credibility assigned to  $A$ .

For each grid image, we used the collected validation data set to analyze Dynamic World Land Cover, ESRI LandCover, and ESA World Accuracy evaluation for each class of cover in the Globeland30, GLC\_FCS30, GISD30, GWL\_FCS30, and Global\_cropland products. The producer's accuracy and user's accuracy of eight products in each grid were obtained, and F1 scores calculated with both kinds of accuracy were assigned as the BPA:

$$m_i(T_j) = \frac{2 \times \text{Recall}_{ij} \times \text{Precision}_{ij}}{\text{Recall}_{ij} + \text{Precision}_{ij}} \quad (1)$$

In Equation (1),  $\text{Recall}_{ij}$  and  $\text{Precision}_{ij}$  are the producer's accuracy and user's accuracy of the  $i$ -th product for the target land cover type  $T_j$ , respectively,  $m_i(T_j)$  is the basic probability assignment of the  $i$ -th product for the target land cover classes  $T_j$  in the unit grid, and  $j$  indicates the eight land cover classes in this classification system, taking values from 1 to 8.

The essence of the evidence combination rule adopted by D-S evidence theory is the orthogonal sum of multiple pieces of evidence. Combined with the actual situation, Dynamic World Land Cover, ESRI LandCover, ESA WorldCover, Globeland30, GLC\_FCS30,

GISD30, GWL\_FCS30, and Global Cropland were fused to obtain the comprehensive probability  $m_i(T_j)$  of each class for each pixel:

$$k = \sum_{T_1 \cap T_2 \cap \dots \cap T_8 = D} m_1(T_j) m_2(T_j) \dots m_8(T_j) \quad (2)$$

$$m(T_j) = m_1(T_j) \oplus m_2(T_j) \dots \oplus m_8(T_j) = \frac{1}{(1-k)} \sum_{T_1 \cap T_2 \cap \dots \cap T_8 = T_j} m_1(T_j) m_2(T_j) \dots m_8(T_j) \quad (3)$$

where  $\oplus$  represents the orthogonal sum,  $m_1(T_j), m_2(T_j) \dots m_8(T_j)$  is the basic probabilistic assignment of the above products to the target land cover classes  $T_j$ , respectively, and  $k$  is the conflict coefficient.

In order to determine the final land cover classes of each pixel, it is necessary to judge the orthogonality of the evidence theory and the results obtained. In this paper, the maximum comprehensive probability is selected as the judgment criterion and the comprehensive probability  $m(T_j)$  of each pixel classes is compared. The class with the largest value is the final land cover class  $T$  of the pixel:

$$m(T_m) = \max_{j \in \{0,1,\dots,6,255\}} (m(T_j)), \quad (4)$$

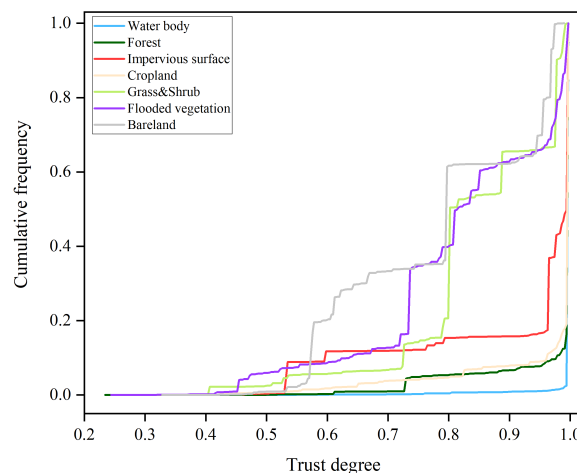
$$T = T_m. \quad (5)$$

In Equations (4) and (5),  $m(T_m)$  and  $T_m$  are the maximum comprehensive probability and the land cover type corresponding to the maximum value, respectively, and  $m(T_j)$  is the comprehensive probability of each classes. The initial Cambodia land cover labeling data were synthesized according to the decision principle. Finally the OSM data were superimposed on the D-S evidence theory fusion labeled data.

The trust degree of each pixel class can be calculated according to the trust function of D-S evidence theory. However, in the scenario of this paper, each class only has a subset of itself, meaning that the trust degrees of the classes obtained by the orthogonal sum are equal to their respective probability values:

$$Bel(T_m) = m(T_m) \quad (6)$$

In Equation (6), the value of  $Bel(T_m)$  ranges from 0 to 1; the greater the value of trust degree, the more reliable the fusion result. In order to select the regions with a high degree of trust and sufficient number of various types from the obtained labels as training labels, in this paper we divided the trust level into 255 levels and conducted cumulative histogram statistics. The lower limit value of the intermediate value of each trust level was taken as the threshold value and the fusion results were screened to obtain the initial labels, as shown in Figure 5.



**Figure 5.** The degree of trust.



### 3.2. Dynamic Label Correction

#### 3.2.1. Noise Label Correction Module

(1) CP Correction Module: Deep learning models are able to maintain efficient performance when confronted with a variety of different data inputs. They can accurately predict results even in the presence of noise, missing data, or other anomalies. Moreover, the probability of model prediction is shown to reflect the accuracy of model classification as confidence [34]. Therefore, in this paper, we use the classification results of the model output and the corresponding CP map [35] as the basis for correction labels and define the way in which the two thresholds are calculated.

We define the sample of high CP as  $P(y_i|x)$ . The closer  $P(y_i|x)$  is to 1, the higher the CP; specifically, given a sample  $X$ , there is a higher CP that  $x$  belongs to class  $y_i$  if  $P(y_i|x)$  is greater than the set threshold or closer to 1:

$$U_1 = P(\hat{y}_0|x). \quad (7)$$

In Equation (7),  $\hat{y}_0$  represents the class with the greatest possibility of  $x$ ; a larger  $U_1$  indicates that the class corresponding to the maximum CP of the pixel is more reliable.

Due to the fact that pixels in the model prediction results are easily classified into two categories with little difference in probability, in order to find pixels with high CP that are not easily confused in the model prediction results, the difference between the maximum and the second largest class probability of the model prediction is defined as the judgment threshold:

$$U_2 = P(\hat{y}_0|x) - P(\hat{y}_1|x). \quad (8)$$

In Equation (8),  $\hat{y}_0$  and  $\hat{y}_1$  represent the classes with the greatest possibility of  $x$  and second-greatest possibility of  $X$ , respectively; the difference between the two possibilities reflects the uncertainty of these two categories of the model. The larger the difference, the smaller the uncertainty, indicating that the two categories are not confused.

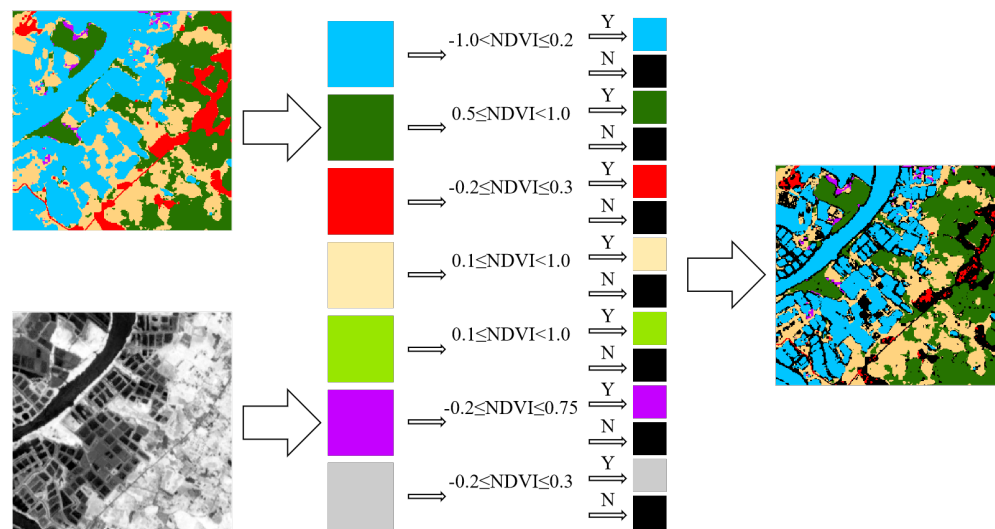
(2) Adaptive Threshold Control: This paper uses the CP  $U_1$  calculated by the maximum CP and the uncertainty  $U_2$  calculated by the maximum and second maximum probability to set a threshold to determine whether the pixels on the label should be updated. However, updating a tag using a fixed threshold to determine which pixels need to be updated results in different tags being updated to different degrees. Therefore, the threshold value of the batch image is obtained using the adaptive and phased method. The median values of  $U_1$  and  $U_2$  for each batch of images were calculated as their updated thresholds. In order to avoid restricting the threshold values of easily identifiable categories while relaxing them for more difficult categories, the thresholds were truncated with empirical thresholds. Finally, the two thresholds of each batch of images were expressed as follows.

$$\varphi_1 = \begin{cases} 0.9 & , \text{ if } \text{median}_{B \times 1 \times H \times W}(U_1) \geq 0.9 \\ \text{median}_{B \times 1 \times H \times W}(U_1) & , \text{ if } 0.9 > \text{median}_{B \times 1 \times H \times W}(U_1) > 0.5 \\ 0.5 & , \text{ if } \text{median}_{B \times 1 \times H \times W}(U_1) \leq 0.5 \end{cases} \quad (9)$$

$$\varphi_2 = \begin{cases} 0.5 & , \text{ if } \text{median}_{B \times 1 \times H \times W}(U_2) \geq 0.5 \\ \text{median}_{B \times 1 \times H \times W}(U_2) & , \text{ if } 0.5 > \text{median}_{B \times 1 \times H \times W}(U_2) > 0.2 \\ 0.2 & , \text{ if } \text{median}_{B \times 1 \times H \times W}(U_2) \leq 0.2 \end{cases} \quad (10)$$

Finally, if both  $\varphi_1$  and  $\varphi_2$  were satisfied, then the selected region was considered to be the high-confidence region of the label.

(3) NDVI Correction Module: The model typically predicts results with a high CP for the correct labels; however, the label is not completely correct, as there may be noisy regions in the label and the model will gradually fit the noisy label, resulting in a gradual increase in the CP of the label's noisy region. Therefore, we used the Normalized Difference Vegetation Index (NDVI) to screen those pixels in the label [36] that are incorrectly predicted. The process and threshold are shown in Figure 6.



**Figure 6.** Label filter by NDVI.

### 3.2.2. Label Correction Process

The process of noisy label correction aims to mitigate the potential presence of noise within the initial label, thereby diminishing the influence of these noisy labels on the classification model. In the early training of deep learning models, the network usually first learns those samples that are easy or correctly classified, which helps the model to establish good generalization ability. Later in the training, the model will gradually begin to fit to those samples containing noise or false labels [37]. Therefore, based on the above module, a two-stage dynamic label correction method is proposed in this paper, which dynamically corrects noise labels during deep learning model training rather than treating them as fixed labels. The label correction method consists of two stages: an initial model training stage and a noisy label self-correction stage. The two stages are described as follows:

**Stage 1: Initial correction model.** Although deep learning models have strong feature learning ability, it is easy for them to fit random noise, which greatly reduces the performance of the network. However, an interesting phenomenon is that deep learning models tend to learn correctly labeled samples early on, and start learning mislabeled samples only later [37]. Furthermore, when maintaining a high learning rate, deep learning models do not easily fit the wrong sample [38]. Therefore, in the initial label, a UNet [39] was trained utilizing the preliminary label and subsequently employed as the foundational classification model. UNet, a prevalent encoder–decoder architecture, progressively condenses feature maps to extract high-level semantic features. Concurrently, the decoder recuperates the spatial information of these feature maps, culminating in a prediction result of an equivalent size to the input. To augment the spatial details of the prediction outcome, the feature maps from the encoder are integrated into the decoder via skip connections.

**Stage 2: Noisy label self-correction.** We employed the initial network (Step 1) as the foundation for training. The parameters of the network and noisy labels were dynamically optimized throughout the training process. This iterative joint optimization can rectify mislabeled samples, decrease the dataset’s noise rate, and enhance model performance. The criterion for updating labels is based on the CP of the network prediction. Specifically, for each pixel, the existing label is adjusted to align with the model’s prediction result if the pixel’s predicted probability falls below a certain threshold; otherwise, it remains unaltered. In this study, the threshold value was adaptive and dynamically changed, and NDVI was used to further remove the incorrect regions in the label in order to avoid label update errors caused by network overfitting. In addition, most areas of the initial label are correct after confidence screening; if the loss is calculated based only on the updated label, the predicted results of the network may deviate completely from the initial label [23]. Therefore, in order to constrain the prediction results of the network, a joint loss function [40] is adopted in

Step 2 to simultaneously consider the loss of the initial label along with the update label and the model prediction probability:

$$L_{correct} = \frac{L_{initial}(P, Y) + \alpha \times L_{update}(P, \hat{Y})}{1 + \alpha}. \quad (11)$$

In Equation (11),  $Y$  represents the initial label,  $\hat{Y}$  represents the label corrected after the last training epoch,  $P$  is the probability of the network prediction,  $L_{initial}$  and  $L_{update}$  represent the initial loss function and the update loss function, respectively, both of which use the cross-entropy loss function, and is used to balance the two loss terms during training.  $\alpha$  is dynamically changed with training, as shown in Equation (12):

$$\alpha = \begin{cases} 0.5 & , \text{ if } current\_epoch + 1 > total\_epoch \\ \frac{current\_epoch + 1}{total\_epoch} \times 0.5 & , \text{ if } current\_epoch + 1 \leq total\_epoch \end{cases} \quad (12)$$

where  $total\_epoch$  is the total training round and  $current\_epoch$  is the current training round. By training the UNet with this loss function, the model can iteratively update and correct the labels.

### 3.3. Land Cover Mapping (LLCM)

#### 3.3.1. Classification Model Training

In the classification model training stage, the UNet model is retrained following the normal training process using the corrected labels. In particular, the ENet [41] class-based cumulative frequency method is used to calculate the weights and build a weighted cross-entropy loss function to balance the classes:

$$L_{classify} = -\frac{1}{N} \sum_{i=1}^N w \tilde{y}_i \ln(P(x_i)), \quad (13)$$

$$w_i = \frac{1}{\ln(c + p_i)}. \quad (14)$$

In Equation (13),  $\tilde{y}_i$  indicates whether the corrected label is in class  $i$  (if yes, it is 1; otherwise it is 0) and  $P(x_i)$  indicates the model's output probability. The weight  $w_i$  of class  $i$  is shown as Equation (14), where  $p_i$  represents the proportion of the number of pixels of class  $i$  in all pixel numbers. Here,  $c$  is set to 1.02, meaning that the class weight is limited to the interval [1, 50].

#### 3.3.2. Land Cover Mapping

In order to obtain a seamless land cover map of Cambodia, a seamless mapping and fusion strategy was used to process the RSI covering Cambodia in the process of trained network reasoning. Specifically, as shown in Figure 5, the process consisted of four steps. First, RSI tiles covering Cambodia were stitched together into the entire image. Second, in order to obtain a batch of data that could be processed by the model, the concatenated image was read into the memory in a sequence of  $256 \times 256$  patches with 64 overlapping pixels in the adjacent two patches. Patches were then passed batch-by-batch into a trained classification model to obtain land cover mapping results for the predicted batches. Although the input batches had 64 overlapping pixels, the overlapping regions had the same prediction results; thus, for the overlapping regions we used the prediction results of the later patch in the adjacent patch and seamlessly merged the prediction batches into the land cover map block. The hardware limitations of model prediction were reduced by reading data for specified positions and sizes in the image, and the patches were continuous between each other, reducing the impact of edge cracks between clipped prediction batches.

## 4. Results

### 4.1. Experimental Setup

The networks were all trained using AdamW while implementing the algorithms on NVIDIA 3090Ti GPUs with a total batch size of 16. In the first stage of dynamic correction, we took 0.01 as the initial learning rate, selected the cross-entropy loss function, and used the ReduceLROnPlateau strategy. When the minimum training loss of epochs did not decrease for ten consecutive years, the learning rate was scaled to 0.1 times the previous one and the training was stopped when the learning rate began to change for the first time. The weights before 10 epochs were selected as the weights for the second stage of the network initialization. In the second phase of dynamic correction, we initialized the UNet with the parameters pretrained in the first phase, used the constructed label dynamic correction method and our newly constructed loss function, fixed the learning rate at 0.01, and trained for 60 epochs.

When using the correction label to train the classification model, we used the ReduceLROnPlateau strategy. When the minimum training loss was not reduced for ten consecutive epochs, the learning rate was scaled to 0.1 times the previous value; the initial learning rate was 0.01 and 100 rounds of training were performed. In order to balance the classes, the weighted cross-entropy loss function mentioned above was used.

### 4.2. Mapping Results and Accuracy Assessment

In order to assess the efficacy of the proposed method on Land Cover tasks, we employed six widely recognized evaluation metrics. First, the user's accuracy(UA), also known as the precision, measures a model's ability to accurately classify an instance into a specific category. This is calculated by dividing the number of true positive instances (i.e., instances correctly classified as the target class) by the total number of instances predicted to belong to that class. The second metric is the producer's accuracy(PA), also referred to as the recall, which gauges a model's capacity to correctly identify a particular type of land cover. This is determined by dividing the number of true positive instances (i.e., instances correctly classified as the target class) by the total number of instances of that class in the ground truth. The third metric is the F1 score(F1), also known as the balanced score, which is defined as the harmonic mean of the precision and recall. The fourth metric is the intersection over union (IoU), which is commonly used to evaluate the performance of semantic segmentation tasks. The IoU is calculated by dividing the intersection area between the predicted segmentation and the ground truth by the union area between them. The fifth metric is the overall accuracy (OA), which is a frequently used as an evaluation index for classification models. It represents the samples correctly classified by the classifier in proportion to the total number of samples. Finally, Kappa is an indicator used to evaluate the performance of a classifier; it is typically used to measure the consistency between the classification result and the true value, and can also be employed to evaluate unbalanced samples.

The confusion matrix and classification result are shown in Table 3 and Figure 7 respectively. The confusion matrix shows that the accuracy was higher than 80% for all categories except Grass & Shrub and bareland, and was higher than 90% for water, forest, and impervious surfaces. The number of water samples is sufficient, and there is little noise or obvious features; thus, the accuracy is very high. The PA of impervious surfaces reached 99.38%, and the UA reached 96.41%, indicating that the model has strong ability to identify buildings. Due to the influence of different growth states of crops, the UA of cultivated land is relatively high and the PA is relatively low. As shown in Figure 7, the PA and UA of grass scrub species are relatively low due to the high degree of confusion between cultivated land and grass scrub species. Due to the large number of paddy fields and tidal flats in Cambodia, the PA of submerged vegetation is very low. Due to the confusion of bare soil with some impervious surfaces composed of gravel during the seeding process, and the confusion of bare soil with cultivated land, PA is lower in these cases. Overall, the OA reached 91.68% and the mF1 reached 0.8837, which is relatively high for the national scale land cover mapping with a resolution of 10 m.

Table 3. Confusion matrix of the LLM in Cambodia.

Mapped Class	Reference Class							Total	UA
	Water Body	Forest	Impervious Surface	Cropland	Grass & Shrub	Flooded Vegetation	Bareland		
Water body	191	0	0	0	0	1	0	192	0.9948
Forest	0	1626	0	10	14	5	0	1655	0.9825
Impervious surface	2	0	161	1	0	0	3	167	0.9641
Cropland	3	10	1	912	79	6	8	1019	0.8950
Grass & Shrub	1	6	0	134	407	0	0	548	0.7427
Flooded vegetation	7	1	0	6	9	82	0	105	0.7810
Bareland	0	0	0	2	0	0	24	26	0.9231
<b>Total</b>	<b>204</b>	<b>1643</b>	<b>162</b>	<b>1065</b>	<b>509</b>	<b>94</b>	<b>35</b>	<b>3712</b>	
<b>PA</b>	<b>0.9363</b>	<b>0.9897</b>	<b>0.9938</b>	<b>0.8563</b>	<b>0.7996</b>	<b>0.8723</b>	<b>0.6857</b>		

mF1 = 0.8837, mIOU = 0.8023, OA = 0.9168, Kappa = 0.8808

Note: mF1 = mean F1; mIOU = mean IOU.

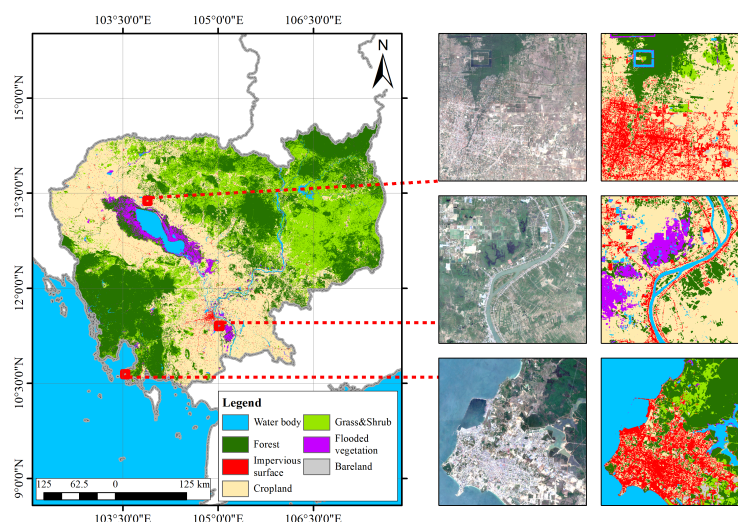


Figure 7. 10 m land cover map of Cambodia for 2020.

#### 4.3. Comparison with Existing PPs

Figure 8 shows the Sentinel-2 images and land cover classification maps for 2020 obtained with the Dynamic World(DW), ESRI LandCover(ESRI), ESRI LandCover(ESA), GLC\_FCS30(GLC), Globeland30(GLB) products. Compared with existing products, the cartographic process presented in this paper achieves better results in terms of vision and accuracy. Using the verification points collected in 2020 for quantitative comparison, the results in Table 4 show that the results for the method in this paper have the highest accuracy in most categories. Compared with ESRI LandCover, the F1-score for the hard-to-identify Grass & Shrub class was higher by 2.41%. Compared with ESA WorldCover, the F1-score for flooded vegetation increased by 18.19%, while for bare ground it increased by 9.94% compared to Dynamic World. The Dynamic World land cover map is based on the land cover situation of all images in the study area in 2020, and the land feature category with the most frequent occurrences is generated as the final result. Because the water body class experiences fewer changes in a year, the accuracy of the Dynamic World land cover map with the most frequent category as the final category is higher than that of the method presented in this paper. In general, the results obtained by this method have the highest accuracy, with an overall accuracy of 91.68%, which is 3.80% higher than that of ESRI LandCover, and a Kappa coefficient of 0.8808.

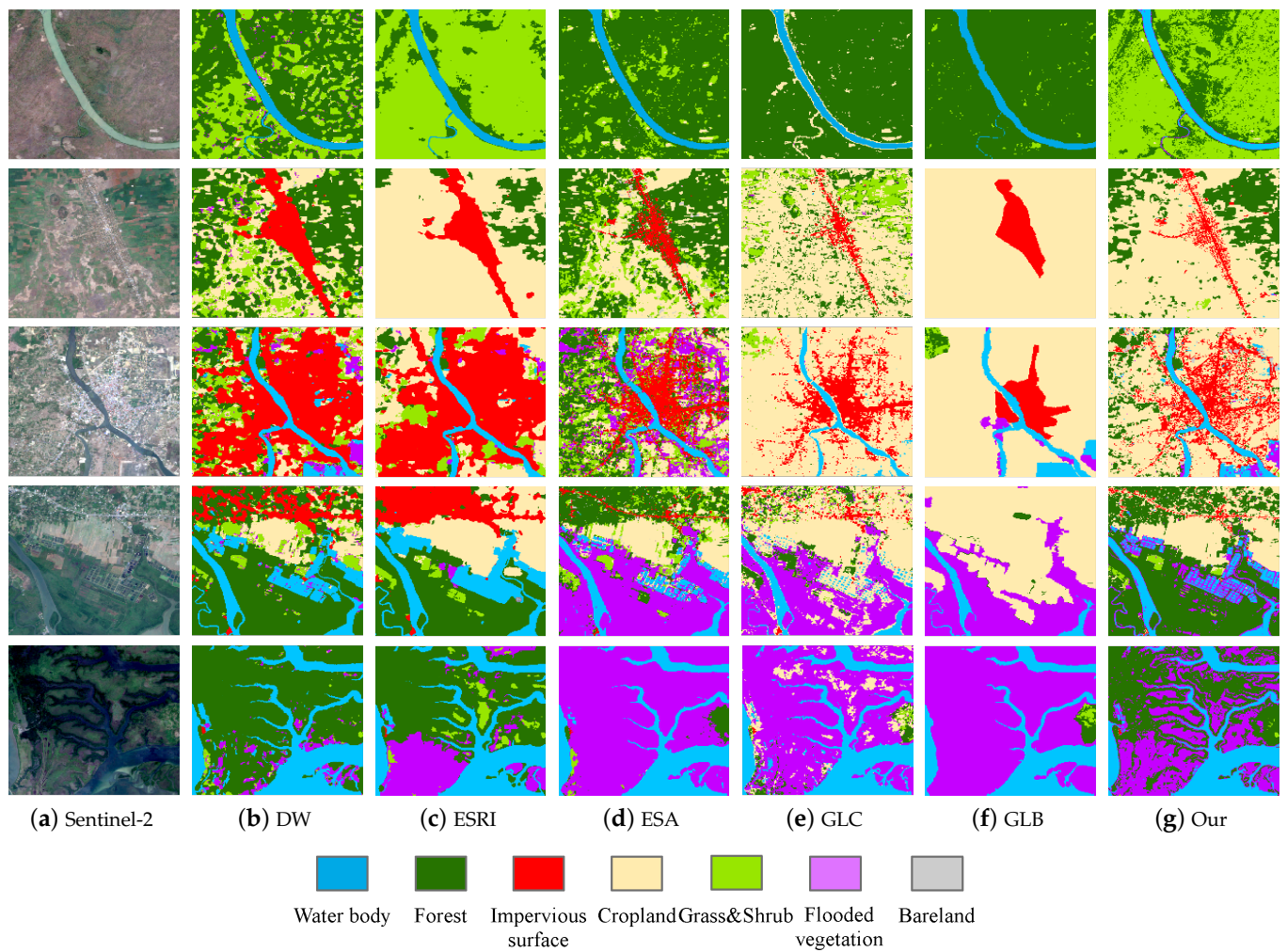


Figure 8. Comparison of the products obtained through our method and other PPs.

Table 4. Comparison with existing PPs.

Mapped Class	Metric	DW	ESRI	ESA	GLC	GLB	Our
Water body	F1	0.9703	0.9524	0.9072	0.9211	0.8238	0.9646
	IOU	0.9423	0.9091	0.8301	0.8538	0.7004	0.9317
Forest	F1	0.9637	0.9557	0.9550	0.7691	0.7556	0.9861
	IOU	0.9299	0.9151	0.9139	0.6249	0.6072	0.9725
Impervious surface	F1	0.9384	0.9388	0.8949	0.8737	0.5191	0.9787
	IOU	0.8840	0.8846	0.8098	0.7758	0.3506	0.9583
Cropland	F1	0.7432	0.8506	0.8101	0.7340	0.6547	0.8752
	IOU	0.5914	0.7400	0.6808	0.5798	0.4866	0.7782
Grass & Shrub	F1	0.6621	0.7460	0.5133	0.0267	0.1872	0.7701
	IOU	0.4949	0.5949	0.3452	0.0136	0.1033	0.6262
Flooded vegetation	F1	0.3978	0.5379	0.6422	0.1688	0.5323	0.8241
	IOU	0.2483	0.3679	0.4730	0.0922	0.3627	0.7009
Bareland	F1	0.6875	0.5882	0.4909	-	-	0.7869
	IOU	0.5238	0.4167	0.3253	-	-	0.6486
	mF1	0.7661	0.7956	0.7448	0.4991	0.4961	0.8837
	mIOU	0.6592	0.6900	0.6254	0.4200	0.3730	0.8023
	OA	0.8419	0.8788	0.8394	0.6781	0.6595	0.9168
	Kappa	0.7757	0.8292	0.7667	0.5283	0.4947	0.8808

## 5. Discussion

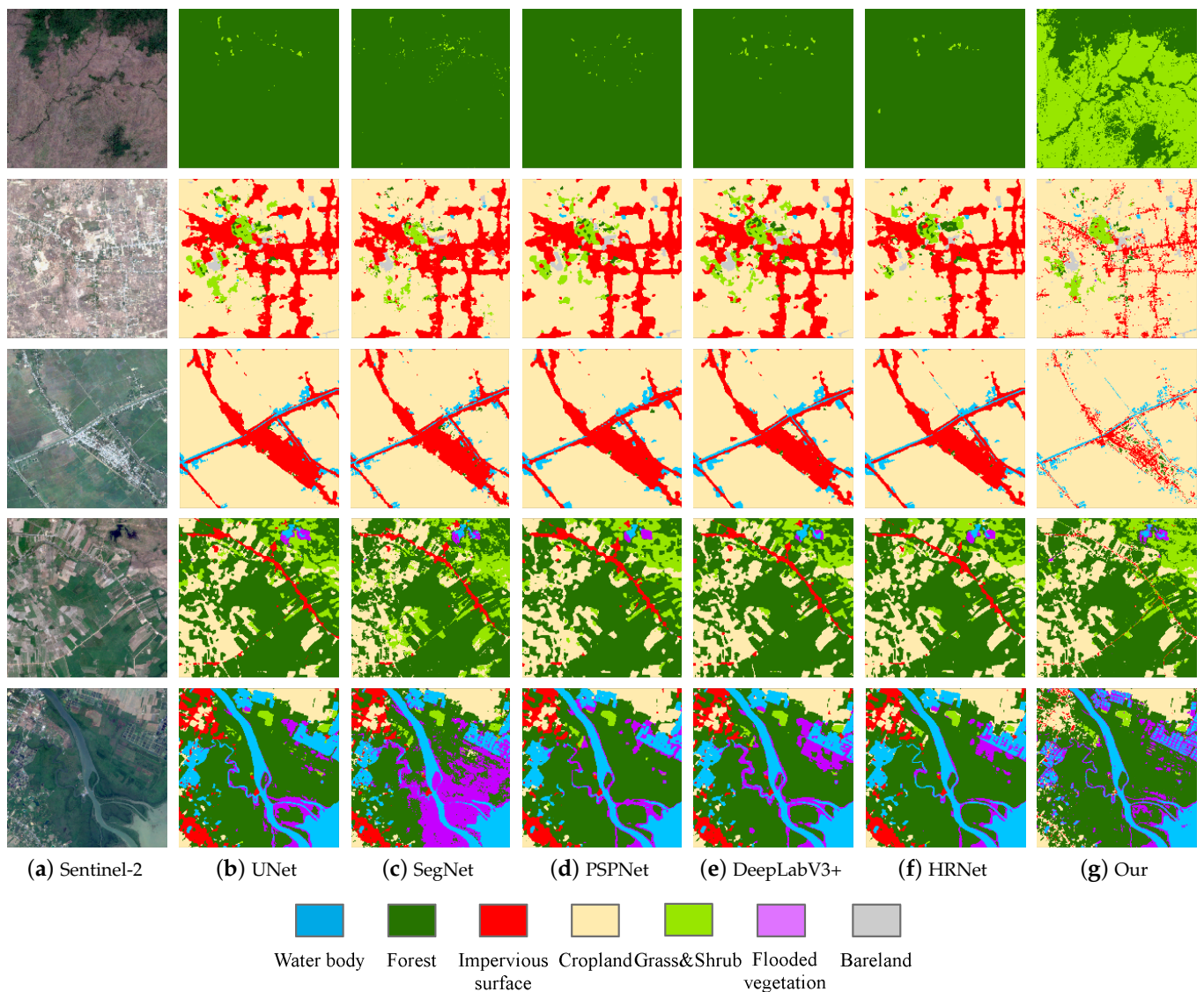
### 5.1. Classification Accuracy of Different Networks

In this section, the proposed mapping process is compared with the five most commonly used image classification methods: UNet, SegNet [42], PSPNet [43], DeepLabv3+ [44], and HRNet [45]. All of the training data, training parameters, loss functions, schedulers, optimizers, etc., were the same, and no pretraining parameters were loaded.

As shown in Table 5, the F1-scores and IoUs of all categories except cultivated land were the highest, and the overall accuracy was 91.68%. Compared with Table 4, it can be seen that the overall accuracy of the model trained with the initial labels is higher than that of the five compared land cover products, indicating the feasibility of using existing products as training labels. Figure 9 shows the classification results of the different models. The method proposed in this paper can distinguish forest land from the Grass & Shrub class well, while the other models misclassify Grass & Shrub into forest land. Compared with the results of other models, the method proposed in this paper can obtain finer results for the impervious surface class, and the surface boundary obtained by our method is clearer. There are a large number of paddy fields and aquaculture plots widely distributed across Cambodia. The aquatic vegetation in the interface area between these plots and the land is well extracted by our method, ensuring that these plots have obvious boundaries, while the results of other models are often wrongly divided into water bodies. In general, the noise is corrected and refined following the label generation and label correction processes described in this paper, significantly improving the classification accuracy of the model and fineness of the results. Compared with the UNet model trained without corrected labels, the overall accuracy of the model trained with the corrected labels is increased by 1.35%, and is improved by 3.8% compared to the highest overall accuracy of the public ESRI LandCover product.

**Table 5.** Comparison of models.

Mapped Class	Metric	UNet	SegNet	PSPNet	DeepLabv3+	HRNet	Our
Water	F1	0.9572	0.9521	0.9495	0.9471	0.9552	0.9646
	IOU	0.9179	0.9087	0.9038	0.8995	0.9143	0.9317
Forest	F1	0.9731	0.9607	0.9632	0.9710	0.9615	0.9861
	IOU	0.9476	0.9245	0.9289	0.9437	0.9258	0.9725
Impervious surface	F1	0.9501	0.9326	0.9280	0.9501	0.9529	0.9787
	IOU	0.9050	0.8736	0.8656	0.9050	0.9101	0.9583
Cropland	F1	0.8783	0.8689	0.8724	0.8768	0.8753	0.8752
	IOU	0.7830	0.7682	0.7737	0.7807	0.7783	0.7782
Grass & Shrub	F1	0.7073	0.6888	0.6869	0.7117	0.6644	0.7701
	IOU	0.5471	0.5253	0.5231	0.5525	0.4974	0.6262
Flooded vegetation	F1	0.7826	0.7175	0.7981	0.7822	0.7293	0.8241
	IOU	0.6429	0.5594	0.6640	0.6423	0.5739	0.7009
Bareland	F1	0.7500	0.7273	0.6538	0.6792	0.5106	0.7869
	IOU	0.6000	0.5714	0.4857	0.5143	0.3429	0.6486
	mF1	0.8569	0.8354	0.8360	0.8455	0.8070	0.8837
	mIOU	0.7634	0.7330	0.7350	0.7483	0.7061	0.8023
	OA	0.9033	0.8893	0.8936	0.9014	0.8920	0.9168
	Kappa	0.8603	0.8404	0.8460	0.8575	0.8425	0.8808



**Figure 9.** Comparison of land cover classification results for different models.

### 5.2. Evaluation of Each Part of the Framework

In ablation experiments, one or more components of the entire process are removed in order to understand how each part contributes to the overall process. Table 6 lists the accuracy of various combinations of the different steps in the mapping process. Based on the D-S evidence theory and integrating multiple products to generate labels, it includes three steps: “D-S Trust”, consisting of D-S evidence theory trust label screening; “Filter by NDVI”, consisting of NDVI label screening; and “Label Correction” based on CP.

All eight experiments were conducted on the basis of labels generated by the fusion of D-S evidence theory, and each row was trained on the accuracy of the land cover map produced after label processing. The first line indicates the precision of the land cover map obtained by the label without any processing. The second line is the result of the land cover map obtained by D-S evidence theory trust screening of the label, showing slightly improved precision compared with the first line. The third line is the accuracy of the land cover map after pixel-by-pixel screening by NDVI. Compared with the second line, the accuracy improvement is higher, indicating that the optimization effect of NDVI screening on labels is greater than that of trust screening. The results of the fourth line show that the accuracy with label noise correction is lower than no label correction when trust and NDVI screening are not used. In the fifth line, both confidence screening and NDVI screening



are carried out, and it can be seen that the accuracy is significantly improved compared with the first line. The sixth line uses trust screening followed by label noise correction, with a slight improvement in accuracy compared to the second line. The seventh line uses NDVI to filter the participating label noise correction; the accuracy is slightly lower than the results in the third line. The eighth line uses trust screening first, followed by NDVI screening for participating label noise correction, with higher accuracy than the previous lines. The above results show that while all three parts of the proposed cartographic process can improve the accuracy of classification, it is worth noting that the use of label noise correction must be used to filter the labels before the accuracy of the land cover map can be improved. For label noise correction using NDVI screening, it is necessary to first screen the labels for trust in order to achieve the best results. Finally, the effect of label noise correction with NDVI filtering is lower than that without NDVI filtering after trust screening.

**Table 6.** Influence of each combination of different steps on the accuracy of the cartographic process.

	D-S Trust	NDVI	Label Correction	mF1	mIOU	OA	Kappa
1				0.8569	0.7634	0.9033	0.8603
2	✓			0.8666	0.7766	0.9084	0.8686
3		✓		0.8656	0.7775	0.9133	0.8750
4			✓	0.8406	0.7471	0.9009	0.8554
5	✓	✓		0.8728	0.7841	0.9154	0.8778
6	✓		✓	0.8759	0.7921	0.9133	0.8747
7		✓	✓	0.8587	0.7676	0.9084	0.8669
8	✓	✓	✓	0.8837	0.8023	0.9168	0.8808

## 6. Conclusions

The difficulty of acquiring training data limits the updating of land cover products. In order to reduce the cost of acquiring labels, existing products can be used as labels; however, these labels contain noise. This paper proposes a land cover mapping framework based on multi-source prior product label generation. Existing land cover products are used to generate noise labels for medium-resolution remote sensing images. Through a three-stage model training process combining label correction with NDVI and confidence probability screening, a 10-m land cover map of Cambodia was completed based on existing products. The results show that the proposed method is effective and that the land cover map produced using the proposed mapping framework has higher precision and a better visual effect than existing land cover products with 10 m resolution. In general, the method presented in this paper does not require manually labeling samples, shortening the time needed to update land cover products and improving their accuracy. However, because the use of time information was not considered, the ability to identify flooded vegetation, grassland, shrubs, and other difficult categories remained limited, and the accuracy of the actual map was not reached. In future studies, we will further explore how to make better use of multi-modal and multi-temporal imagery, existing products, and publicly available statistics in order to achieve more accurate and faster updating of land cover maps.

**Author Contributions:** Conceptualization, H.Z., X.M. and P.L.; Data curation, H.Z., Y.M., Z.J. and Z.M.; Formal analysis, X.M.; Funding acquisition, T.Y.; Investigation, H.Z.; Methodology, H.Z. and J.Y. (Jian Yan); Project administration, X.M.; Resources, H.Z., X.M. and J.Y. (Jian Yang); Validation, H.Z. and X.M.; Visualization, H.Z.; Writing—original draft, H.Z.; Writing—review and editing, H.Z., X.M. and C.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Graduate innovation funding project of North China Institute of Aerospace Engineering (YKY-2022-60), the National Key R&D Program of China (2021YFE0117300), the Major Project of High Resolution Earth Observation System (30-Y60B01-9003-22/23), the Shandong Provincial Natural Science Foundation, China (Grant No.ZR2020QD012), and the Civil Aerospace Technology Pre-research Project of China's 14th Five-Year Plan.

**Data Availability Statement:** The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Friedl, M.A.; Sulla-Menashe, D.; Tan, B.; Schneider, A.; Ramankutty, N.; Sibley, A.; Huang, X. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sens. Environ.* **2010**, *114*, 168–182. [CrossRef]
- Buchhorn, M.; Smets, B.; Bertels, L.; De Roo, B.; Lesiv, M.; Tsendbazar, N.E.; Herold, M.; Fritz, S. Copernicus global land service: Land cover 100 m: Collection 3: Epoch 2019: Globe. *Zenodo* 2020, Version V3.0.1. Available online: <https://zenodo.org/records/3939050> (accessed on 1 January 2023).
- Buchhorn, M.; Lesiv, M.; Tsendbazar, N.E.; Herold, M.; Bertels, L.; Smets, B. Copernicus global land cover layers—Collection 2. *Remote Sens.* **2020**, *12*, 1044. [CrossRef]
- Chen, J.; Chen, J.; Liao, A.; Cao, X.; Chen, L.; Chen, X.; He, C.; Han, G.; Peng, S.; Lu, M.; et al. Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS J. Photogramm. Remote Sens.* **2015**, *103*, 7–27. [CrossRef]
- Zhang, X.; Liu, L.; Chen, X.; Gao, Y.; Xie, S.; Mi, J. GLC\_FCS30: Global land-cover product with fine classification system at 30 m using time-series Landsat imagery. *Earth Syst. Sci. Data* **2021**, *13*, 2753–2776. [CrossRef]
- Chen, B.; Xu, B.; Zhu, Z.; Yuan, C.; Suen, H.P.; Guo, J.; Xu, N.; Li, W.; Zhao, Y.; Yang, J.; 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.* **2019**, *64*, 3.
- Van De Kerchove, R.; Zanaga, D.; Keersmaecker, W.; Souverijns, N.; Wevers, J.; Brockmann, C.; Grosu, A.; Paccini, A.; Cartus, O.; Santoro, M.; et al. ESA WorldCover: Global land cover mapping at 10 m resolution for 2020 based on Sentinel-1 and 2 data. In Proceedings of the AGU Fall Meeting Abstracts, New Orleans, LA, USA, 13–17 December 2021; Volume 2021, pp. GC45I–0915.
- Zanaga, D.; Van De Kerchove, R.; Daems, D.; De Keersmaecker, W.; Brockmann, C.; Kirches, G.; Wevers, J.; Cartus, O.; Santoro, M.; Fritz, S.; et al. ESA WorldCover 10 m 2021 v200. Available online: <https://zenodo.org/records/7254221> (accessed on 8 August 2022)
- Karra, K.; Kontgis, C.; Statman-Weil, Z.; Mazzariello, J.C.; Mathis, M.; Brumby, S.P. Global land use/land cover with Sentinel 2 and deep learning. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, IEEE, Brussels, Belgium, 11–16 July 2021; pp. 4704–4707.
- Brown, C.F.; Brumby, S.P.; Guzder-Williams, B.; Birch, T.; Hyde, S.B.; Mazzariello, J.; Czerwinski, W.; Pasquarella, V.J.; Haertel, R.; Ilyushchenko, S.; et al. Dynamic World, Near real-time global 10 m land use land cover mapping. *Sci. Data* **2022**, *9*, 251. [CrossRef]
- Belgiu, M.; Drăguț, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]
- Maggiori, E.; Tarabalka, Y.; Charpiat, G.; Alliez, P. Convolutional neural networks for large-scale remote-sensing image classification. *IEEE Trans. Geosci. Remote Sens.* **2016**, *55*, 645–657. [CrossRef]
- Wambugu, N.; Chen, Y.; Xiao, Z.; Wei, M.; Bello, S.A.; Junior, J.M.; Li, J. A hybrid deep convolutional neural network for accurate land cover classification. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *103*, 102515. [CrossRef]
- Tong, X.Y.; Xia, G.S.; Lu, Q.; Shen, H.; Li, S.; You, S.; Zhang, L. Land-cover classification with high-resolution remote sensing images using transferable deep models. *Remote Sens. Environ.* **2020**, *237*, 111322. [CrossRef]
- Tait, A.M.; Brumby, S.P.; Hyde, S.B.; Mazzariello, J.; Corcoran, M. *Dynamic World Training Dataset for Global Land Use and Land Cover Categorization of Satellite Imagery*; PANGAEA: Wuhan, China, 2021. [CrossRef]
- Schmitt, M.; Hughes, L.H.; Qiu, C.; Zhu, X.X. SEN12MS—A Curated Dataset of Georeferenced Multi-Spectral Sentinel-1/2 Imagery for Deep Learning and Data Fusion. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *IV-2/W7*, 153–160. [CrossRef]
- Schmitt, M.; Prexl, J.; Ebel, P.; Liebel, L.; Zhu, X.X. Weakly supervised semantic segmentation of satellite images for land cover mapping—challenges and opportunities. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *V-3-2020*, 795–802. [CrossRef]
- Dong, R.; Li, C.; Fu, H.; Wang, J.; Li, W.; Yao, Y.; Gan, L.; Yu, L.; Gong, P. Improving 3-m resolution land cover mapping through efficient learning from an imperfect 10-m resolution map. *Remote Sens.* **2020**, *12*, 1418. [CrossRef]
- Zhang, H.K.; Roy, D.P. Using the 500 m MODIS land cover product to derive a consistent continental scale 30 m Landsat land cover classification. *Remote Sens. Environ.* **2017**, *197*, 15–34. [CrossRef]
- Li, C.; Gong, P.; Wang, J.; Zhu, Z.; Biging, G.S.; Yuan, C.; Hu, T.; Zhang, H.; Wang, Q.; Li, X.; et al. The first all-season sample set for mapping global land cover with Landsat-8 data. *Sci. Bull.* **2017**, *62*, 508–515. [CrossRef]
- Defourny, P.; Kirches, G.; Brockmann, C.; Boettcher, M.; Peters, M.; Bontemps, S.; Lamarche, C.; Schlerf, M.; Santoro, M. Land cover CCI. *Prod. User Guide Version* **2012**, *2*, 10–1016.
- Hua, T.; Zhao, W.; Liu, Y.; Wang, S.; Yang, S. Spatial consistency assessments for global land-cover datasets: A comparison among GLC2000, CCI LC, MCD12, GLOBCOVER and GLCNMO. *Remote Sens.* **2018**, *10*, 1846. [CrossRef]
- Yi, K.; Wu, J. Probabilistic end-to-end noise correction for learning with noisy labels. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 7017–7025.

24. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [[CrossRef](#)]
25. Frantz, D.; Haß, E.; Uhl, A.; Stoffels, J.; Hill, J. Improvement of the Fmask algorithm for Sentinel-2 images: Separating clouds from bright surfaces based on parallax effects. *Remote Sens. Environ.* **2018**, *215*, 471–481. [[CrossRef](#)]
26. Zhu, Q.; Lei, Y.; Sun, X.; Guan, Q.; Zhong, Y.; Zhang, L.; Li, D. Knowledge-guided land pattern depiction for urban land use mapping: A case study of Chinese cities. *Remote Sens. Environ.* **2022**, *272*, 112916. [[CrossRef](#)]
27. Zhang, X.; Liu, L.; Zhao, T.; Gao, Y.; Chen, X.; Mi, J. GISD30: Global 30 m impervious-surface dynamic dataset from 1985 to 2020 using time-series Landsat imagery on the Google Earth Engine platform. *Earth Syst. Sci. Data* **2022**, *14*, 1831–1856. [[CrossRef](#)]
28. Zhang, X.; Liu, L.; Zhao, T.; Chen, X.; Lin, S.; Wang, J.; Mi, J.; Liu, W. GWL\_FCS30: Global 30 m wetland map with fine classification system using multi-sourced and time-series remote sensing imagery in 2020. *Earth Syst. Sci. Data Discuss.* **2022**, *2022*, 1–31. [[CrossRef](#)]
29. Potapov, P.; Turubanova, S.; Hansen, M.C.; Tyukavina, A.; Zalles, V.; Khan, A.; Song, X.P.; Pickens, A.; Shen, Q.; Cortez, J. Global maps of cropland extent and change show accelerated cropland expansion in the twenty-first century. *Nat. Food* **2022**, *3*, 19–28. [[CrossRef](#)]
30. White, D.; Kimerling, J.A.; Overton, S.W. Cartographic and geometric components of a global sampling design for environmental monitoring. *Cartogr. Geogr. Inf. Syst.* **1992**, *19*, 5–22. [[CrossRef](#)]
31. Zhang, M.; Huang, H.; Li, Z.; Hackman, K.O.; Liu, C.; Andriamiarisoa, R.L.; Ny Aina Nomenjanahary Raherivelo, T.; Li, Y.; Gong, P. Automatic high-resolution land cover production in madagascar using sentinel-2 time series, tile-based image classification and google earth engine. *Remote Sens.* **2020**, *12*, 3663. [[CrossRef](#)]
32. Shafer, G. *A Mathematical Theory of Evidence*; Princeton University Press: Princeton, NJ, USA, 1976; Volume 42.
33. Dempster, A.P. Upper and lower probabilities induced by a multivalued mapping. In *Classic Works of the Dempster-Shafer Theory of Belief Functions*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 57–72.
34. Frénay, B.; Verleysen, M. Classification in the presence of label noise: A survey. *IEEE Trans. Neural Netw. Learn. Syst.* **2013**, *25*, 845–869. [[CrossRef](#)] [[PubMed](#)]
35. Lee, K.H.; He, X.; Zhang, L.; Yang, L. Cleannet: Transfer learning for scalable image classifier training with label noise. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 5447–5456.
36. Chen, Y.; Zhang, G.; Cui, H.; Li, X.; Hou, S.; Ma, J.; Li, Z.; Li, H.; Wang, H. A novel weakly supervised semantic segmentation framework to improve the resolution of land cover product. *ISPRS J. Photogramm. Remote Sens.* **2023**, *196*, 73–92. [[CrossRef](#)]
37. Arazo, E.; Ortego, D.; Albert, P.; O'Connor, N.; McGuinness, K. Unsupervised label noise modeling and loss correction. In Proceedings of the International Conference on Machine Learning, PMLR, Long Beach, CA, USA, 9–15 June 2019; pp. 312–321.
38. Tanaka, D.; Ikami, D.; Yamasaki, T.; Aizawa, K. Joint optimization framework for learning with noisy labels. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 5552–5560.
39. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015*; Proceedings, Part III 18; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
40. Wang, Y.; Ma, X.; Chen, Z.; Luo, Y.; Yi, J.; Bailey, J. Symmetric cross entropy for robust learning with noisy labels. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 322–330.
41. Paszke, A.; Chaurasia, A.; Kim, S.; Culurciello, E. Enet: A deep neural network architecture for real-time semantic segmentation. *arXiv* **2016**, arXiv:1606.02147.
42. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [[CrossRef](#)]
43. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the Computer Vision and Pattern Recognition, CVPR, Honolulu, HI, USA, 21–26 July 2017.
44. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
45. Sun, K.; Xiao, B.; Liu, D.; Wang, J. Deep high-resolution representation learning for human pose estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 5693–5703.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.