



## Article

# National-Standards- and Deep-Learning-Oriented Raster and Vector Benchmark Dataset (RVBD) for Land-Use/Land-Cover Mapping in the Yangtze River Basin

Pengfei Zhang <sup>1</sup>, Yijin Wu <sup>1,\*</sup>, Chang Li <sup>1,\*</sup>, Renhua Li <sup>2</sup>, He Yao <sup>2</sup>, Yong Zhang <sup>2</sup>, Genlin Zhang <sup>1</sup> and Dehua Li <sup>1</sup>

<sup>1</sup> Key Laboratory for Geographical Process Analysis & Simulation of Hubei Province, College of Urban and Environmental Science, Central China Normal University, Wuhan 430079, China; zhangpf@mails.ccn.edu.cn or pengf\_zhang@126.com (P.Z.); glz@mails.ccn.edu.cn (G.Z.); lidehua@mails.ccn.edu.cn (D.L.)

<sup>2</sup> Yangtze River Basin Monitoring Center Station for Soil and Water Conservation, Changjiang Water Resources Commission, Wuhan 430010, China; cwjczxz@163.com (R.L.); yaoxiaohe@163.com (H.Y.); yongzchn@163.com (Y.Z.)

\* Correspondence: wuyijin@ccnu.edu.cn (Y.W.); lcshaka@126.com or lichang@ccnu.edu.cn (C.L.)

**Abstract:** A high-quality remote sensing interpretation dataset has become crucial for driving an intelligent model, i.e., deep learning (DL), to produce land-use/land-cover (LULC) products. The existing remote sensing datasets face the following issues: the current studies (1) lack object-oriented fine-grained information; (2) they cannot meet national standards; (3) they lack field surveys for labeling samples; and (4) they cannot serve for geographic engineering application directly. To address these gaps, the national-standards- and DL-oriented raster and vector benchmark dataset (RVBD) is the first to be established to map LULC for conducting soil water erosion assessment (SWEA). RVBD has the following significant innovation and contributions: (1) it is the first second-level object- and DL-oriented dataset with raster and vector data for LULC mapping; (2) its classification system conforms to the national industry standards of the Ministry of Water Resources of the People's Republic of China; (3) it has high-quality LULC interpretation accuracy assisted by field surveys rather than indoor visual interpretation; and (4) it could be applied to serve for SWEA. Our dataset is constructed as follows: (1) spatio-temporal-spectrum information is utilized to perform automatic vectorization and label LULC attributes conforming to the national standards; and (2) several remarkable DL networks (DenseNet161, HorNet, EfficientNetB7, Vision Transformer, and Swin Transformer) are chosen as the baselines to train our dataset, and five evaluation metrics are chosen to perform quantitative evaluation. Experimental results verify the reliability and effectiveness of RVBD. Each chosen network achieves a minimum overall accuracy of 0.81 and a minimum Kappa of 0.80, and Vision Transformer achieves the best classification performance with overall accuracy of 0.87 and Kappa of 0.86. It indicates that RVBD is a significant benchmark, which could lay a foundation for intelligent interpretation of relevant geographic research about SWEA in the Yangtze River Basin and promote artificial intelligence technology to enrich geographical theories and methods.

**Keywords:** remote sensing dataset; deep learning; soil water erosion assessment; object-oriented image classification; land-use/land-cover mapping



**Citation:** Zhang, P.; Wu, Y.; Li, C.; Li, R.; Yao, H.; Zhang, Y.; Zhang, G.; Li, D. National-Standards- and Deep-Learning-Oriented Raster and Vector Benchmark Dataset (RVBD) for Land-Use/Land-Cover Mapping in the Yangtze River Basin. *Remote Sens.* **2023**, *15*, 3907. <https://doi.org/10.3390/rs15153907>

Academic Editors: Qian Du, Yanni Dong and Xiaochen Yang

Received: 8 June 2023

Revised: 3 August 2023

Accepted: 5 August 2023

Published: 7 August 2023



**Copyright:** © 2023 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

Soil water erosion has become a serious environmental hazard around the global world, which impacts climate change, agricultural production, and socio-economic-ecological sustainable development [1–3]. Land-use/land-cover (LULC) information reflects the interaction between human activities and natural ecosystems, which has been identified as a decisive factor to accelerate global land degradation and soil water erosion [4]. The People's Republic of China faces a critical challenge from soil water erosion [5,6]. Based on deep learning (DL) and remote sensing technology, high-accuracy LULC information could

be extracted to perform large-scale and extensive intelligent ground monitoring, which is beneficial to conducting soil water erosion assessment (SWEA) in a cost-efficient manner [7]. Compared with traditional models that rely on statistics or physical knowledge, DL networks are trained with massive samples that could automatically learn remote sensing parameter characteristics of ground objects [8–11]. This means that the performance of DL-based approaches strongly depends on the quality and quantity of the provided dataset [12]. Therefore, the construction of a high-quality remote sensing image interpretation dataset contributes to enhancing the generalization ability of DL networks and further improving the accuracy of LULC mapping.

At present, many research institutions and scholars are devoted to remote sensing dataset research for mapping LULC. There are two main categories of datasets: scene classification datasets and object detection datasets. Object detection datasets focus on recognizing ground objects with bounding boxes to predict the location and LULC categories [13], such as FAIR1M [12], TAS [14], ImageNet [15], PASCAL VOC [16], SZTAKI-INRIA [17], MSCOCO [18], UCAS-AOD [19], DLR 3K [20], NWPU VHR-10 [21], VEDAI [22], HRSC2016 [23], COWC [24], RSOD [25], LEVIR [26], ITCVD [27], DOTA [28], DIOR [14], and RSSOD [29]. Scene classification is for categorizing remote sensing images into a series of LULC categories with the image patches [29], such as UC Merced Land-Use [30], WHU-RS19 [31], RSSCN7 [32], Brazilian Coffee Scene [33], RSC11 [34], SIRI-WHU [35], NWPU-RESISC45 [29], RSD46-WHU [25], AID [36], AID++ [37], OPTIMAL-31 [38], PatternNet [39], OSAR [40], RSI-CB [41], DIOR [42], Eurosat [43], Bigearthnet [44], MLRSNet [45], BigEarthNet-MM [46], MRSB [47], AIFS-DATASET [48], MRSID [49], and LuoJiaSET [50]. The available datasets could offer great potential in mitigating the highly nonlinear and overparameterized restrictions of DL networks [51].

It is crucial and urgently necessary to construct thematic remote sensing datasets for conducting SWEA. There still exists an application gap between geographic research and specific geo-engineering applications. It is worth noting that different scholars or institutions have constructed diverse LULC classification systems to meet their respective research needs. However, this diversity of current datasets is not conducive to their universality across various geographical researches. To be specific, there is still a lack of thematic remote sensing datasets for conducting SWEA. Hence, it is crucial and urgently needed to construct datasets in accordance with a standard and authoritative LULC classification system. This could improve the application and universality of remote sensing datasets for SWEA research.

It is crucial and urgently necessary to construct refined fine-grained LULC datasets, which can significantly enhance the monitoring and assessing capacities of soil water erosion. LULC data are a significant factor for calculating the Chinese soil loss equation (CSLE) model [5,52]. It is also recognized as the authoritative model to quantitatively evaluate the magnitude and distribution of soil erosion in the People's Republic of China, and it adopts the LULC classification system conforming to the national industry standards of the Standards for Classification and Gradation of Soil Erosion (SL 190-2007) published by the Ministry of Water Resources of the People's Republic of China. It is generally regarded as the vegetation cover and biological practice factor, which reflects the impact of vegetation cover and biological practice on erosion rate under fallow conditions [47]. Therefore, improving the accuracy of LULC mapping directly affects the monitoring level of SWEA. The development of the fine-grained (i.e., second-level LULC) dataset is beneficial to implementing high-accuracy LULC mapping. It can further enhance the intra-class similarity and improve the inter-class variability, which promotes the development of the LULC classification level and, in turn, further improves the accuracy of SWEA.

It is crucial and urgently necessary to improve the quality of LULC labeling for the remote sensing dataset. The correction of sample labeling is very important in remote-sensing application research [53]. DL models rely on training using numerous labeled data to yield the high classification accuracy [54]. Due to the phenomenon of the same object with different spectrum and the foreign body with the same spectrum, different ground objects in

remote sensing images have similarities in respect of color, texture, size, shape, shadow, and distribution position. That generally results in errors of LULC labeling. To alleviate the issue, some researchers have delved into various aspects, as follows: on one hand, much research adopted quality control methods to reduce human errors [55–65]. For example, Qi et al. [45] rely on many technicians to train data labeling several times until reaching a predetermined and reliable confidence score. On the other hand, remote sensing interpretation keys have a great effect on image labeling [66–75]. Remote sensing interpretation keys are sampled through field survey, which can provide accurate interpretation reference and real LULC category information rather than relying on human visual interpretation. However, there are fewer studies that rely on field surveys to improve the quality of data labeling.

In conclusion, fine-grained remote sensing classification dataset research still faces the following scarcities:

- (1) Lacking object-oriented fine-grained datasets for DL-based LULC mapping. Current remote sensing scene classification datasets and object detection datasets primarily emphasize recognizing the LULC category and spatial position information. The fixed image patches generally contain heterogeneous objects, which means that the detailed geometrical information of ground objects, such as object boundaries, is still missing. That limits the progression of high-accuracy LULC mapping.
- (2) Lacking datasets conforming to national LULC classification standards. The classification systems of current remote sensing datasets are diverse and are formulated according to different research needs. This diversity in LULC classification systems hinders the broader applicability of these datasets in other related fields, such as agricultural production and socio-economic–ecologically sustainable development [1–3]. Thus, developing datasets based on universal and authoritative standards, such as national industry standards, is essential to enhance their universality and application value.
- (3) Lacking field surveys for LULC dataset labeling. Current remote sensing datasets generally depend on professional technicians to label the LULC category, which means that there is no process for field surveys to verify the correctness of the labeled LULC. The subjectivity of professional technicians and the complexity of remote sensing images contribute to the degradation of data labeling quality. Incorrect data labeling greatly influences the training of DL networks and reduces classification accuracy. Thus, improving the quality of image labeling will significantly enhance the quality of the dataset.
- (4) Lacking datasets meeting the engineering application requirement to conduct SWEA in the Yangtze River Basin. The application gap between current published datasets and SWEA application is still not addressed. It is worth noting that there is no thematic dataset available for conducting SWEA. In addition, some samples of representative LULC categories (e.g., sloping cropland) are not sampled in current datasets, which plays an important role in soil and water conservation.

To solve the above issues, a second-level standards-of-the-Ministry-of-Water-Resources-and-DL-oriented raster and vector benchmark dataset (RVBD) is the first to be established to perform LULC mapping for supporting SWEA in the Yangtze River Basin, which adheres to the notional standards published by the Ministry of Water Resources of the People's Republic of China and meets the requirements of DL. The main remarkable innovation and contributions are shown as follows:

- (1) A second-level object- and DL-oriented dataset with raster and vector data is first to be established for large-scale LULC mapping to the best of our knowledge. Different from the current datasets only containing remote sensing image patches, RVBD also includes vector data. In addition, image patches from open-source Google images are homogeneous objects with geometric boundary information, which can be directly applied for mapping LULC.
- (2) An LULC dataset conforming to the national industry standards is the first to be established to the best of our knowledge. The classification system of RVBD is

constructed following the water resources industry standard of the People's Republic of China, i.e., the Current Land Use Classification (GB/T 21010-2017). It is significant for improving the universality of RVBD and the application value.

- (3) A high-quality LULC labeling dataset with the assist of remote sensing interpretation keys is the first to be established to the best of our knowledge. Remote sensing interpretation keys are sampled through field surveys to facilitate the interpretation of LULC categories by indoor technicians. It is equally important that the correctness of sample labeling is verified through a field survey, which significantly improves the quality of sample labeling.
- (4) RVBD is the first to lay an intelligent foundation for high-accuracy LULC mapping to support SWEA to the best of our knowledge. It greatly improves the application value of RVBD. Particularly, geographical theories and methods are further enriched based on artificial intelligence (AI) technology.

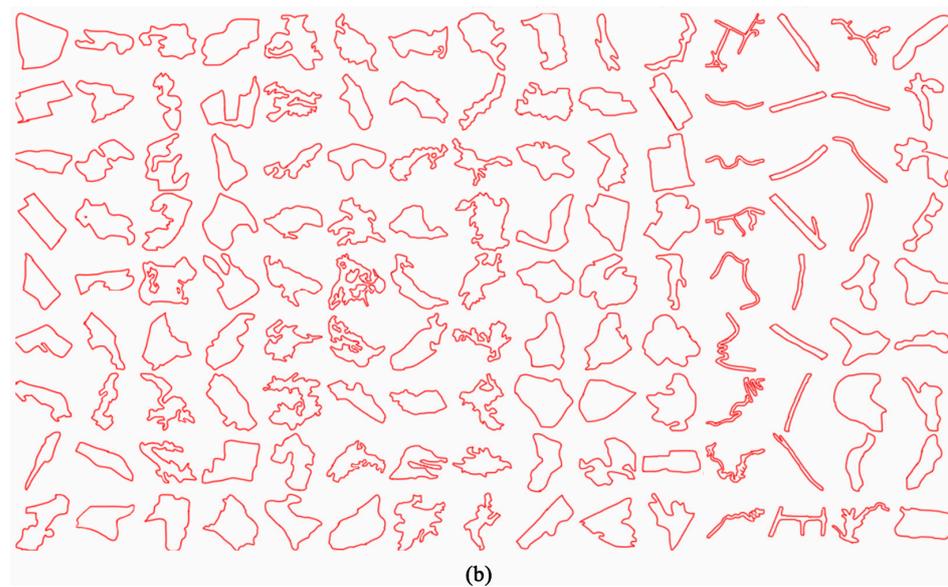
## 2. Raster and Vector Benchmark Dataset (RVBD)

### 2.1. Description of RVBD

RVBD is the first to be constructed to serve for SWEA in the Yangtze River Basin of the People's Republic of China. RVBD contains 23,300 pairs of labeled samples with corresponding remote sensing image patches and vector data. Vector data are utilized as masks to generate image patches by cropping remote sensing images with geometric shapes. The remote sensing images were downloaded from Google Earth with three spectral bands (i.e., red, green, and blue bands), which were obtained in the year 2020. The image resolution is 2 m. It is worth noting that the cloud cover area of the images does not exceed 10% of the overall regions. It includes 15 LULC categories: paddy land, dry land, sloping cropland, garden land, forest, shrub land, grassland, urban construction land, rural construction land, mining land, other construction land, rural road, other transportation land, water, and barren land. Some examples of each LULC category of RVBD are shown in Figure 1.



Figure 1. Cont.



**Figure 1.** Some examples of each LULC category in RVBD are shown: (a) represents the remote sensing image patches of samples and (b) represents the vector data of samples.

## 2.2. Classification System

In order to facilitate its application in remote sensing monitoring of soil and water loss in the Yangtze River Basin, as well as to provide high-accuracy LULC mapping information for relevant geographical research in the area, the classification system of RVBD could be developed in strict compliance with the industry standards set by the Ministry of Water Resources of China. The classification system of the dataset is implemented in accordance with the national standard (i.e., the Technical Specification of Soil and Water Conservation Monitoring by Remote Sensing (SL 592-2012)). It could be applicable for soil and water conservation monitoring at various scales within China, including national, watershed, and region scales. The standard reference, the Current Land Use Classification (GB/T 21010-2017), combined with the characteristics of the soil and water conservation industry, has formed a suitable LULC classification system for SWEA. This classification system could be applied to calculate soil and water loss and provide data support for ecological environment monitoring. Table 1 shows the classification system of each category in detail with some samples of remote sensing interpretation keys from field surveys.

**Table 1.** The LULC classification system of the RVBD is described in detail.

Level I	Level II	Remote Sensing Interpretation Key	
Name	Name	Remote Sensing Image	Photo
Cultivated land	Paddy land		
	Dry land		

**Table 1.** *Cont.*

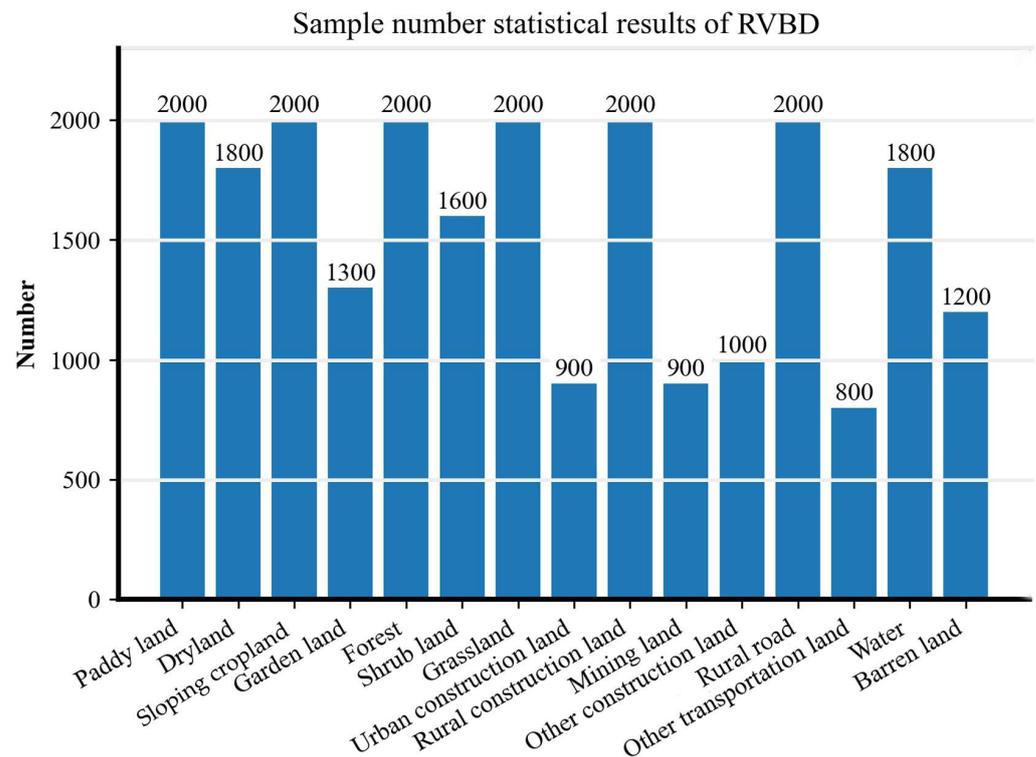
Level I Name	Level II Name	Remote Sensing Interpretation Key	
		Remote Sensing Image	Photo
Cultivated land	Sloping cropland		
			
Forest	Forest		
	Shrub land		
Grassland	Grassland		
		Construction land	Urban construction land
	Rural construction land		

Table 1. Cont.

Level I Name	Level II Name	Remote Sensing Interpretation Key	
		Remote Sensing Image	Photo
Construction land	Mining land		
	Other construction land		
Transportation land	Other transportation land		
	Rural road		
Water	Water		
Other land	Barren land		

### 2.3. Dataset Splits

The RVBD contains 23,300 pairs of samples with image patches and a vector dataset. The number of different LULC categories of RVBD varies greatly from 800 to 2000, and the sample number statistics for each LULC category is shown in Figure 2. The RVBD is divided into a training set with 13,980 pairs of objects, a validation set with 4660 pairs of objects, and a test set with 4660 pairs of objects, according to the ratio of 6:2:2, and remote sensing image patches are utilized for model training, model optimization, and model evaluation of DL baseline networks, respectively. The detailed splits are reported in Table 2.



**Figure 2.** The sample number of each LULC category in RVBD.

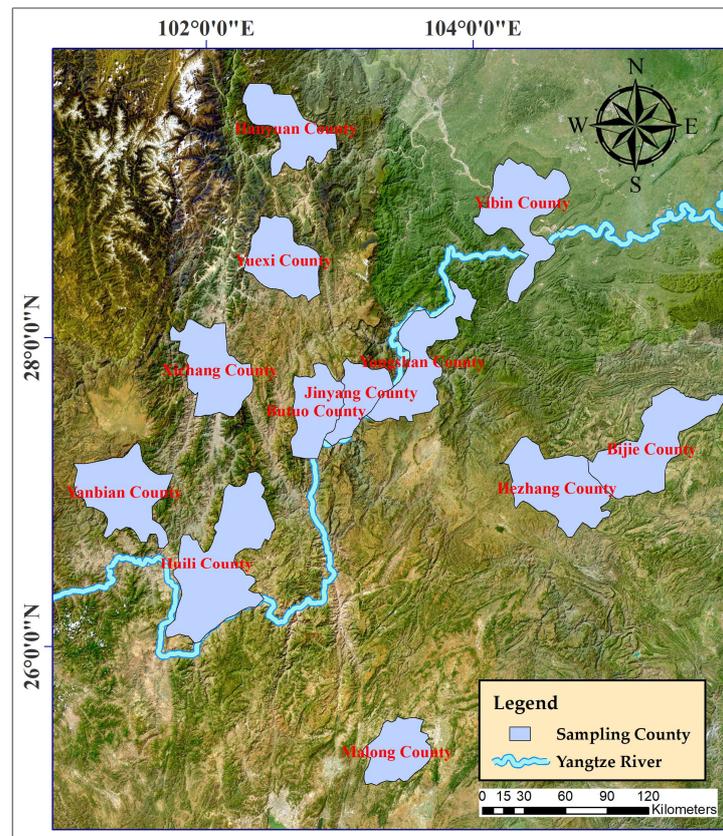
**Table 2.** The number of objects of RVBD for each category and dataset split.

LULC Category	Training Set	Validation Set	Test Set
Paddy land	1200	400	400
Dry land	1080	360	360
Sloping cropland	1200	400	400
Garden	780	260	260
Forest	1200	400	400
Shrub land	960	320	320
Grassland	1200	400	400
Urban construction land	540	180	180
Rural construction land	1200	400	400
Mining land	540	180	180
Other construction land	600	200	200
Other transportation land	1200	400	400
Rural road	480	160	160
Water	1080	360	360
Barren land	720	240	240
Total number	13,980	4660	4660

#### 2.4. Study Area

In this research, the study area is located in the middle and lower reaches of the Jinsha River Basin. It is one of the areas with serious soil water erosion in the Yangtze River Basin, and 13 counties are selected that run through Yunnan, Guizhou, and Sichuan provinces. Soil water erosion has become a critical factor hindering the development of the regional economy and society. The complex geological and geomorphic environment of the region, coupled with its various climatic conditions, have fostered a wide range of natural landscapes. In addition, the region is prosperous in both economy and culture, which is a key area of the Yangtze River Economic Belt. Thus, the complex interaction between natural and human activities has resulted in the formation of rich and diverse LULC patterns in this region, which is beneficial for selecting samples conforming to the

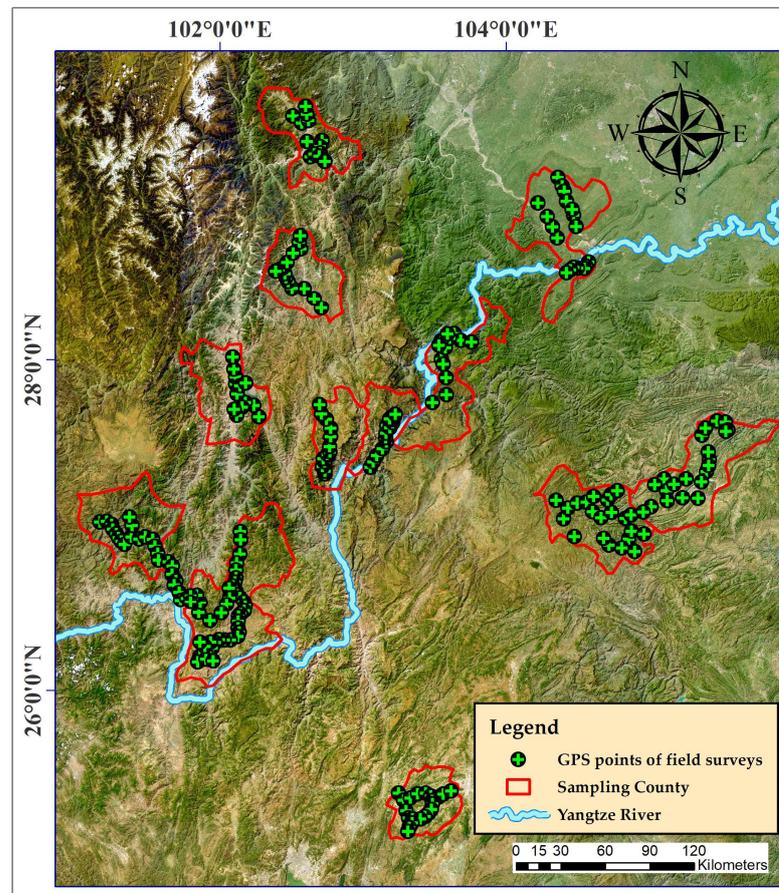
national industry standards and constructing a diverse and representative dataset. The study area is shown in Figure 3.



**Figure 3.** Study area.

### 2.5. Field Surveys

Field surveys are carried out to establish a representative, practical, and stable dataset in accordance with the industry standards, the Technical Specification of Soil and Water Conservation Monitoring by Remote Sensing (SL 592-2012) of the Ministry of Water Resources of the People’s Republic of China. It aims to (1) sample remote sensing interpretation keys for indoor professional visual interpretation and (2) verify the correctness of labeled samples. Field surveys are sampled as follows: (1) global positioning system (GPS) points are set up according to the principle of uniform distribution in space. It is worth noting that GPS points are selected in the areas with wide vision and covering a rich LULC category. (2) Field photos and remote sensing image samples are sampled or used to verify the LULC category information. Remote sensing interpretation keys are beneficial to verify the image characteristics (i.e., color, shadow, texture, size, shape, location, etc.) and could provide exact LULC category information to assist indoor technicians in improving the correctness of labeling LULC categories of images. Field verification is very meaningful for reducing the errors caused by the phenomenon of having the same objects with different spectra and foreign bodies with the same spectra. The GPS-sampled points of the field survey are shown in Figure 4.



**Figure 4.** Sampling points of field survey.

### 3. Methodology

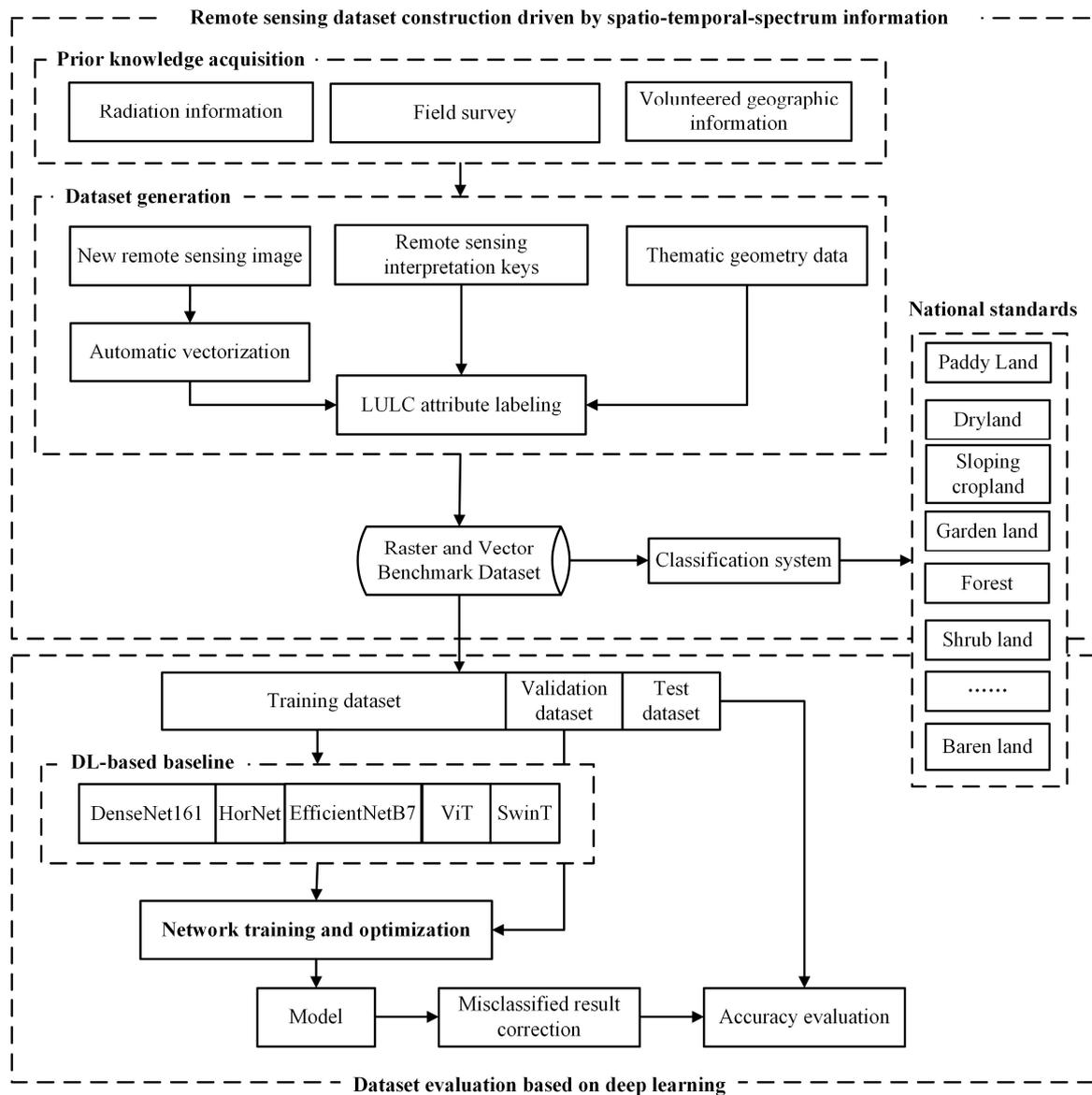
The overall workflow of the dataset construction method is shown in Figure 5, which mainly includes the following two contents:

(1) Remote sensing dataset construction driven by spatio-temporal spectrum information

Various pieces of geographic information driven by spatio-temporal spectrum big data are utilized to construct the RVBD. Radiation knowledge is provided from remote sensing images to implement large-scale monitoring for SWEA. Thematic geometry knowledge is provided from volunteered geographic information, which is regarded as the mask data to yield the high-accuracy geometry vector data, such as road, river, and construction data. Then, automatic vectorization is performed for unmasked areas based on the multi-resolution segmentation approach. Finally, LULC attribute information is labeled by professional technicians with the assist of expert knowledge from remote sensing interpretation keys.

(2) Dataset evaluation based on DL

Five outstanding DL networks obtained from two different architectures, i.e., convolutional neural network (CNN) and Transformer, are chosen to evaluate the effectiveness of the established RVBD. This dataset is divided into a training set, a validation set, and a test set for DL network training and parameter optimization. Then, the manual labeling errors and machine errors generated by the DL networks are correcting based on field surveys. Finally, the accuracy evaluation is performed to verify the effectiveness of the RVBD.



**Figure 5.** The overall workflow.

### 3.1. Remote Sensing Dataset Construction Driven by Spatio-Temporal Spectrum Information

#### (1) Prior knowledge acquisition from spatio-temporal spectral big data

Prior knowledge is conducive to alleviate the problem that the acquisition of large-scale labeled training data is laborious and expensive [76]. In the era of geographic big data, prior knowledge is obtained from spatio-temporal spectral big data to serve the construction of a remote sensing dataset, which includes various pieces of strictly calibrated geospatial information (e.g., radiation knowledge, thematic geometric knowledge, and expert knowledge). The spatio-temporal spectral big data provide the following data: remote sensing images are downloaded from open-source Google images and processed by strict geometric calibration and radiation calibration, high-accuracy volunteered geographic information from OpenStreetMap (OSM), and historical remote sensing interpretation keys sampled by field surveys. In addition, there are still much remote sensing data available that can be further explored, such as hyperspectral images. It is worth noting that they contain high-dimensional and multispectral information, which includes redundant data. This could increase the computational demands of DL networks and impede the convergence of DL training. Therefore, it is highly important to perform the dimensionality reduction methods, such as principal component analysis (PCA) [77].

## (2) Thematic geometry masking by volunteered geographic information

Volunteered geographic information (VGI) can provide high-accuracy geographic thematic data, which are widely recognized for the reliability, availability, and time efficiency of the data acquisition [78]. OpenStreetMap data are generally regarded as the most active and widely applied VGI data [79], which can be collected by both professional and amateur volunteers. Leveraging OSM data could save time and money on large-scale sample labeling. The accurate thematic vector data (i.e., road vector, river vector, and construction vector) could be yielded from OSM, and geographical registration is performed combining the aforementioned vector data with remote sensing images to minimize geometric errors. Then, remote sensing images can be masked with the above vector to obtain prior thematic vector regions.

## (3) Automatic vectorization based on multi-resolution segmentation

The multi-resolution segmentation approach is adopted to implement automatic vectorization for the remained unmasked regions [80]. It is a bottom-up region-merging approach to merge local homogeneity pixels to generate heterogeneous objects, and three crucial parameters are generally utilized to optimize the segmentation results, as follows: the scale parameter is used to optimize the segmented objects and solve the phenomenon of over-segmentation and under-segmentation; the shape parameter takes the relationship between spectral and spatial uniformity into account; the compactness parameter considers spatial heterogeneity from two geometric attributes of different objects such as the perimeter and their area, the perimeter, and bounding boxes.

## (4) Attribute labeling and dataset constructing

The LULC classification system is constructed for computing the model of CSLE, conforming to the national classification standard adopted by the Ministry of Water Resources of the People's Republic of China (i.e., GB/T 21010-2017). LULC samples are selected on the geometric vector data of ground objects combining the automatic vectorization result with OSM thematic vector data. Based on visual interpretation, the sample attribute is labeled by professional technicians with the assist of remote sensing interpretation keys from field surveys, which are available by field surveys and are beneficial to identify ground objects on remote sensing images. Quality control is implemented to check the correctness of the labeled LULC category based on the field survey, which is over 92%. Finally, the remote sensing images are cropped with the geometric vector data of selected samples to generate the established RVBD.

### 3.2. Dataset Evaluation Based on DL

#### 3.2.1. DL-Based Baseline

Current mainstream DL architectures are traditional CNN architectures and the more recent widely recognized Transformer architectures. The CNN and Transformer architectures have discrepant architecture designs and feature extraction capabilities. CNN architecture has a stronger local perceptual ability because of built-in inductive biases that apply local convolutional filters to enhance the spatial invariance [81]. Transformer architecture generally splits the input image into a sequence of patches to model sequence-to-sequence (long-range) relations for yielding the stronger global modeling ability. Several superior DL networks of CNN and Transformer architectures are selected to evaluate the established RVBD, such as HorNet [82], DenseNet161 [83], EfficientNetB7 [84], Vision Transformer (ViT) [85], and Swin Transformer (SwinT) [86], which are briefly introduced as follows.

EfficientNetB7 is one variant of EfficientNet and has been widely recognized for effectively evaluating the performance of remote sensing datasets [84], tackling the issue that traditional CNNs are scaled only in following individual means of increasing the numbers of layers, increasing numbers of channels, or adjusting the input image size. EfficientNet designs a new scaling-up CNN approach, which develops an innovative

compound coefficient to efficiently balance the scale relationships among depth, width, and resolution dimensions for yielding excellent whole-classification performance.

DenseNet161 is one variant of the Dense Convolutional Network and achieves a noteworthy improvement based on ResNet [87], which has the several following convincing benefits: alleviating the vanishing gradient issue, enhancing information flow, improving feature reuse, and compressing the number of parameters [83]. The dense block is the dominating feature extraction module in DenseNet, which regards any one of all the preceding layers as the input feature to directly access the subsequent layers. DenseNet enhances the reuse efficiency of the feature maps and yields good performance in remote sensing classification.

HorNet is inspired by the dot-product self-attention operation in Transformers and explicitly explores the spatial interaction relationship between local space and its circumjacent region [82]. Recursive gated convolution is designed to implement high-order and long-term spatial interactions with recursive gating convolution and large kernel convolutions. The network achieves efficient, extendable, and translation-equivariant performance for yielding remarkable power in image classification.

The design inspiration of ViT stems from the natural language processing (NLP) architecture. ViT is a pure and standard Transformer that splits an image into sequences of patches like sequences of word tokens in NLP, which leverages multi-head self-attention to capture the global dependency relationship for the patches [85]. In addition, position information remains based on the position-embedding module. It is a simple and scalable architecture that has shown superior performance in various visual tasks.

SwinT is different from pure visual Transformer architecture ViT, which introduces the additional visual inductive biases (i.e., locality, translation invariance, and hierarchy) with shifting windows to enhance the local modeling power [86]. Instead of using a fixed window to generate image patches, the shifted windows can compute self-attention crossing the boundaries of the previous windows for yielding stronger performance. It has lower latency and enables efficient processing of high-resolution images while maintaining a good balance between performance and computation.

### 3.2.2. Network Training Strategy

Transfer learning is an excellent training strategy and could achieve higher levels of generalization power with fewer training iterations, which could leverage pretrained weight parameters and can be applied to downstream tasks directly, such as remote sensing image classification [78]. For the aforementioned five DL networks for the established RVBD, the pretrained models stemming from ImageNet [15], which is a popular deep learning image classification dataset, are transferred to train with the RVBD. It is worth noting that network weights of all layers are fine tuned to yield better classification performance.

### 3.2.3. Misclassified Result Correction

Manual labeling can easily lead to misclassification. To address this problem, field surveys are carried out to correct the manual labeling errors and machine errors generated by DL networks.

- (1) Define unreliable classification results. Utilizing the softmax function, the DL-based classification probabilities are generated as output. If the top two highest classification probabilities are approximately equal (i.e., the difference is less than 0.1), they could be considered as unreliable classification results.
- (2) Correct the unreliable results by visual interpretation. The manual visual interpretation is employed to update the classification result with the assist of remote sensing interpretation keys.
- (3) Verify the results by field surveys. The aforementioned, easily misclassified objects are further verified by field surveys to correct the machine errors generated by DL-based classification and human errors generated by visual interpretation. Especially

in regions with terrain or potential hazards, we employ unmanned aerial vehicles (UAVs) to facilitate the manual validation [88].

### 3.2.4. Evaluation Metrics

To quantitatively evaluate the quality of the established RVBD, different evaluation metrics are adopted, such as overall accuracy (OA), Kappa coefficient (Kappa), precision, recall, and F<sub>1</sub> score. These metrics focus on the different facets of the classification capabilities of the selected DL network. In addition, confusion matrices of predicted results of all networks are also provided to represent more performance details, which is conducive to the analysis classification results for each LULC class.

The Kappa coefficient can be calculated by the following Equation (1):

$$\text{Kappa} = \frac{l_o - l_e}{1 - l_e}, l_o = \frac{m}{n}, l_e = \frac{\sum_{i=1}^c p_i q_i}{n^2} \quad (1)$$

where Kappa is the Kappa coefficient,  $m$  and  $n$  are the number of correctly classified samples and total samples, respectively, and  $p_i$  and  $q_i$  are the number of real samples and the predicted samples of the  $i$ th LULC category, respectively.

The OA can be calculated by the following Equation (2):

$$\text{OA} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

The precision can be calculated by the following Equation (3):

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

The recall can be calculated by the following Equation (4):

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

The F<sub>1</sub> score can be calculated by the following Equation (4):

$$F_1 = 2 \times \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

In Equations (2) and (3), true positive (TP) means that the true LULC category and predicted LULC category are both positive, true negative (TN) means that the true LULC category and the predicted LULC category are both negative, false positive (FP) means that the true LULC category is negative but the predicted LULC category is positive, and false negative (FN) means that the true LULC category is positive but the predicted LULC category is negative.

## 4. Experiments and Results

### 4.1. Experimental Settings

Five popular DL networks, i.e., HorNet [82], DenseNet161 [83], EfficientNetB7 [84], ViT [85], and SwinT [86], are chosen as the baseline networks to evaluate the classification performance for the constructed benchmark RVBD.

A fine-tuned training strategy is implemented to improve the generalization capability based on the pretrained weights of ImageNet [15]. All remote sensing image patches of samples are resized to  $256 \times 256$  pixels as the input for each network. Random horizontal and vertical flip operations are carried out for data augmentation. AdamW [89] is chosen as the optimizer with 100 epochs. The cosine annealing strategy is utilized as a learning scheduler, and the initial learning rate is set to 0.001. The parameter of the batch size of all networks is set to 32.

In addition, all aforementioned networks are implemented in the study on a work station equipped with, i.e., an Intel Core i7-8700 central processing unit (CPU) and four NVIDIA GeForce GTX 3090Ti Central Processing Unit graphics processing units (GPUs).

#### 4.2. Results and Analysis

In this section, five fine-tuned DL networks are performed as the baseline classification networks to evaluate the effectiveness of the constructed RVBD from the following two aspects: overall classification accuracy analysis and class-wise classification accuracy analysis.

##### (1) Overall classification accuracy analysis

The classification performance of the aforementioned five fine-tuned networks using RVBD is analyzed in detail from the overall classification view. As reported in Table 3, it is even more evident that all five networks consistently perform well, with OA and Kappa metric values both exceeding 0.80. In particular, the values with bold font indicate the highest value in the comparative networks. It is obvious that the ViT network achieves the best performance, and its OA and Kappa metric values are 0.87 and 0.86, respectively. In addition, the relatively worse classification network is HorNet, which also yields reliable accuracy with OA of 0.81 and Kappa of 0.80. The above-excellent OA and Kappa metric values of the five networks indicate that DL networks achieve significant classification ability and the constructed dataset is effective.

**Table 3.** The overall classification accuracy results for RVBD.

Network Name	Overall Accuracy	Kappa
DenseNet161 [83]	0.86	0.85
EfficientNetB7 [84]	0.84	0.83
HorNet [82]	0.81	0.80
SwinT [86]	0.83	0.82
ViT [85]	0.87	0.86

##### (2) Class-wise accuracy analysis

To represent the classification capacity of all the aforementioned networks for each LULC category of RVBD, class-wise accuracy analysis is carried out. The confusion matrixes of all networks are shown in Figure 6, which represent the classification details for each LULC category. We find that all networks have yielded outstanding classification performance for most LULC categories. Precision, recall, and  $F_1$ -score metrics are chosen to quantitatively evaluate the classification performance for each LULC category of RVBD, which are shown in Table 4, Table 5, and Table 6, respectively. It is worth noting that the highest classification values among the five aforementioned DL networks for each LULC category are annotated with bold font.

The results can be obviously observed as the following: (1) the mean values of all metrics (i.e., precision, recall, and  $F_1$  score) are basically higher than 0.80 for all chosen networks, which means that all networks represent the excellent and stable classification capacity for each LULC category; (2) the ViT network achieves the best performance on account of yielding the highest values of precision, recall, and  $F_1$ -score metrics among most LULC categories; (3) the HorNet network has the relatively worst classification capacity compared with other chosen networks; (4) the  $F_1$  score is the harmonic mean of the precision and recall metrics. From the view of the  $F_1$ -score metric, water, sloping cropland, and paddy land have the best classification effect for all chosen DL networks. Rural road, mining land, and shrub land are more easily misclassified.

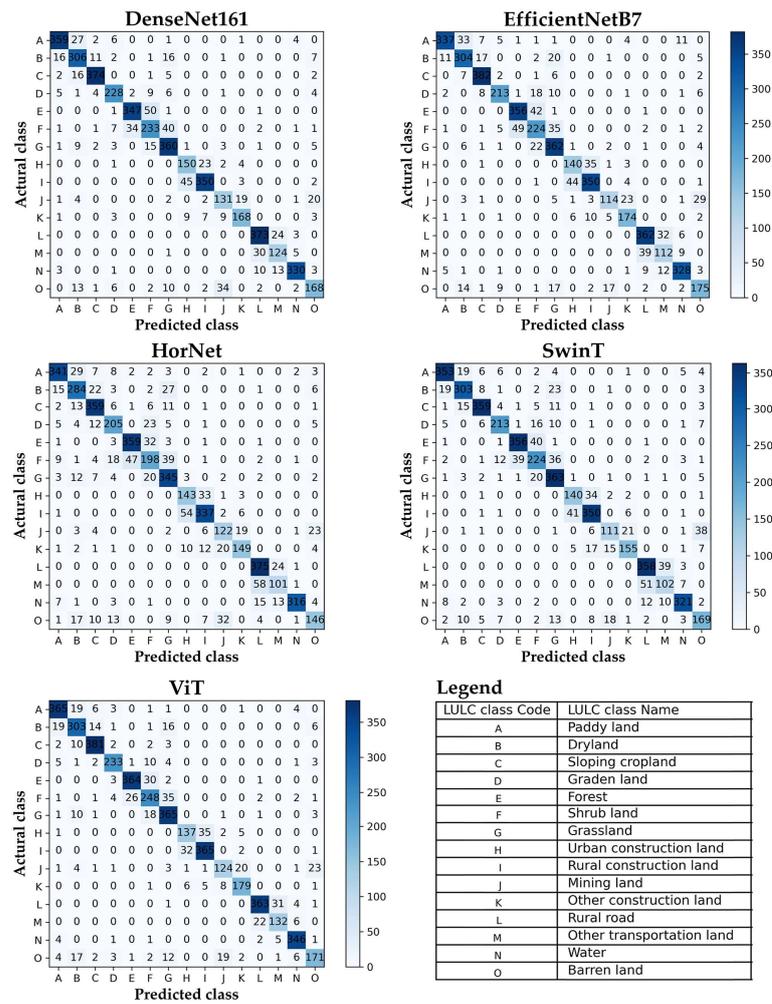


Figure 6. The confusion matrixes of classification results of all chosen DL networks are shown.

Table 4. The precision metric results of all networks for RVBD.

Network Name	DenseNet161 [83]	EfficientNetB7 [84]	HorNet [82]	SwinT [86]	ViT [85]
LULC Class					
Paddy land	0.92	0.94	0.88	0.90	0.91
Dry land	0.81	0.82	0.78	0.86	0.83
Sloping cropland	0.95	0.91	0.84	0.92	0.93
Garden	0.88	0.90	0.78	0.86	0.93
Forest	0.91	0.87	0.88	0.89	0.93
Shrub land	0.75	0.72	0.70	0.72	0.79
Grassland	0.81	0.79	0.78	0.79	0.83
Urban construction land	0.73	0.73	0.68	0.75	0.78
Rural construction land	0.91	0.87	0.84	0.84	0.90
Mining land	0.72	0.81	0.68	0.75	0.81
Other construction land	0.86	0.83	0.84	0.83	0.86
Other transportation land	0.89	0.87	0.82	0.84	0.93
Rural road	0.77	0.72	0.73	0.67	0.78
Water	0.95	0.91	0.98	0.93	0.94
Barren land	0.78	0.76	0.75	0.70	0.81
Mean Values	0.84	0.83	0.80	0.82	0.86

**Table 5.** The recall accuracy results of all networks for RVBD.

Network Name LULC Class	DenseNet161 [83]	EfficientNetB7 [84]	HorNet [82]	SwinT [86]	ViT [85]
Paddy land	0.90	0.84	0.85	0.88	0.91
Dry land	0.85	0.84	0.79	0.84	0.84
Sloping cropland	0.93	0.95	0.90	0.90	0.95
Garden	0.88	0.82	0.79	0.82	0.90
Forest	0.87	0.89	0.90	0.89	0.91
Shrub land	0.73	0.70	0.62	0.70	0.77
Grassland	0.90	0.90	0.86	0.91	0.91
Urban construction land	0.83	0.78	0.79	0.78	0.76
Rural construction land	0.87	0.87	0.84	0.87	0.91
Mining land	0.73	0.63	0.68	0.62	0.69
Other construction land	0.84	0.87	0.74	0.77	0.89
Other transportation land	0.93	0.90	0.94	0.89	0.91
Rural road	0.77	0.70	0.63	0.64	0.82
Water	0.92	0.91	0.88	0.89	0.96
Barren land	0.70	0.73	0.61	0.70	0.71
Mean Values	0.84	0.82	0.79	0.81	0.86

**Table 6.** The F<sub>1</sub> score accuracy results of all networks for RVBD.

Network Name LULC Class	DenseNet161 [83]	EfficientNetB7 [84]	HorNet [82]	SwinT [86]	ViT [85]
Paddy land	0.91	0.89	0.87	0.89	0.91
Dry land	0.83	0.83	0.78	0.85	0.84
Sloping cropland	0.94	0.93	0.87	0.91	0.94
Garden	0.88	0.86	0.78	0.84	0.91
Forest	0.89	0.88	0.89	0.89	0.92
Shrub land	0.74	0.71	0.66	0.71	0.78
Grassland	0.86	0.84	0.82	0.84	0.87
Urban construction land	0.78	0.75	0.73	0.76	0.77
Rural construction land	0.89	0.87	0.84	0.86	0.91
Mining land	0.73	0.71	0.68	0.68	0.74
Other construction land	0.85	0.85	0.79	0.80	0.88
Other transportation land	0.91	0.89	0.87	0.87	0.92
Rural road	0.77	0.71	0.68	0.65	0.80
Water	0.93	0.91	0.93	0.91	0.95
Barren land	0.74	0.75	0.67	0.70	0.76
Mean Values	0.84	0.83	0.79	0.81	0.86

#### 4.3. Discussion

##### (1) The effectiveness and superiority of RVBD

This paper performs comprehensive experiments based on several outstanding DL baselines to further demonstrate the effectiveness and superiority of the established RVBD. (1) Some reliable accuracy evaluation metrics are selected to evaluate the effectiveness of RVBD. Regardless of the following two aspects of analysis, overall accuracy or class-wise accuracy, the classification results also achieve stable and outstanding performance. (2) The RVBD represents an excellent classification ability with the assist of geometric information. Some LULC categories with representative geometric shapes generally obtain better classification accuracy. For example, water is banded and planar, construction land generally has regular shapes, and other rural transportation land is generally striped. The above-mentioned LULC categories all achieve better F<sub>1</sub> scores in all baselines. This indicates that the geometric information provided by vector data is beneficial to improving the classification accuracy. Meanwhile, our method has some substantive and reproducible practices for establishing relative datasets: (1) multiple-source vector data are beneficial for

providing references for manual visual interpretation and reducing a certain amount of the workload of sample labeling; (2) field surveys are conducive to sampling remote sensing interpretation keys as references for indoor manual visual interpretation. In addition, it is also conducive for verifying the classification results for correcting the machine errors generated by DL networks and human errors generated by visual interpretation.

## (2) The applicability of RVBD

The following advantages reflect the stronger applicability of the dataset: (1) the dataset is constructed based on the object-oriented approach. This means that it has complete geometric information, which makes it beneficial for mapping LULC [73,90,91]. (2) The classification system conforms to the national industry standard adopted by the Ministry of Water Resources of the People's Republic of China. This means that it has higher adaptability to be applied to other relative research. (3) The RVBD is a thematic dataset. This means that it can be applied to conduct SWEA in the Yangtze River Basin. (4) It is conducive to updating the ground-truth data comprehensively. The established dataset integrates raster and vector data, which makes it easy to update the classification results of ground-truth data based on the DL models to the vector data for achieving high-accuracy LULC mapping. (5) It is an excellent solution for a data application flexibility plan. On one hand, our dataset conforms to the national industry standards to construct a land-use/land-cover classification system, which makes it easily and directly applied in other research [92–95]; on the other hand, the transfer learning method [96] could be adopted to train effective DL models for achieving high-accuracy land-use/land-cover mapping in other geographical areas, only depending on a limited number of samples without huge cost.

## (3) The classification capacity of DL networks

DL networks are further verified to be effective in dealing with a remote sensing classification task. The aforementioned five DL networks are only selected to evaluate the established RVBD. However, the challenge still exists that samples of the established RVBD are easily misclassified, especially for similar LULC categories, because of the phenomenon of the same object with a different spectrum and a foreign body with the same spectrum. Hence, there is ample space for further optimizing DL networks to obtain better classification accuracy.

## 5. Conclusions

To the best of our knowledge, this paper is the first to construct an RVBD for conducting SWEA in the Yangtze River Basin based on the support of spatio-temporal spectral big data, which conforms the standards of the Ministry of Water Resources of the People's Republic of China and is based on DL. The RVBD includes 15 LULC categories and 23,300 pairs of object-based samples with corresponding image patches and vector data. It is sampled from Google images with 2 m resolution, which are spread across many countries of the Yangtze River Basin. It is worth noting that five DL networks are introduced to verify the effectiveness of the RVBD by evaluating the classification accuracy from two aspects: overall accuracy analysis and class-wise accuracy analysis. Experimental results verify the effectiveness of the RVBD. Every chosen network achieved remarkable performance, and the ViT network achieves the best classification performance with overall accuracy of 0.87 and Kappa of 0.86.

It is worth noting that the RVBD has broad and flexible applicability significance: (1) it could be utilized to provide high resolution LULC data, which contribute to the research in the Yangtze River basin, such as geographical conditions monitoring [95], prediction/simulation LULC change [94], ecosystem service [92], climate change [93], and so on; (2) it could be easily applied in other areas because of its inclusion of abundant and diverse LULC categories. On the one hand, the Yangtze River Basin basically covers all the terrain around the world. It has abundant topography with a multi-level and terraced distribution that spans from plateaus to plains; on the other hand, the Yangtze River Basin

exhibits diverse climate characteristics, which vary from plateau climate to subtropical monsoon climate to subtropical maritime climate. The above characteristics of natural conditions result in a rich variety of LULC categories except for deserts and glaciers, which allows our dataset to be easily transferred to any geographical area, such as the Yellow River Basin, and even to the world.

In addition to this, our research still contains some limitations: there are no LULC classes on the snow-covered plateau. Especially in the classification data verification of super-large national land, there still are some objective difficulties, as follows: (1) remote sensing images with limited spectral and spatial resolution pose difficulties for visual interpretation. The visual interpretation capacities of different technicians are different, which could inevitably result in misclassification results. (2) Sampling remote sensing interpretation keys by field surveys is difficult. Conducting field surveys is a challenging task for sampling remote sensing interpretation keys and verifying the accuracy of classification results because of the substantial human and financial resources required. In future, we will further improve this dataset in the following promising aspects: (1) the classification ability needs to be further explored and strengthened by incorporating or designing more advanced deep learning models; (2) the generalization ability should be further investigated and enhanced, especially under conditions of few-shot samples, to broaden its application potential. In particular, it lays an intelligent data foundation for SWEA in the Yangtze River Basin, which is beneficial for the promotion and development of the intelligent application of remote sensing.

**Author Contributions:** Conceptualization, P.Z. and C.L.; Methodology, P.Z. and C.L.; Software, G.Z. and D.L.; Formal analysis, P.Z.; Resources, Y.W., R.L., H.Y. and Y.Z.; Data curation, Y.W., H.Y. and Y.Z.; Writing—original draft, P.Z.; Writing—review & editing, P.Z. and C.L.; Visualization, G.Z. and D.L.; Supervision, Y.W., C.L. and R.L.; Project administration, Y.W.; Funding acquisition, C.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China (NSFC) (Grant No. 41771493 and 41101407) and the Fundamental Research Funds for the Central Universities (Grant No. CCNU22QN019).

**Data Availability Statement:** The national-standards- and deep-learning-oriented raster and vector benchmark dataset (RVBD) could be publicly available at <https://doi.org/10.5281/zenodo.8002595>.

**Acknowledgments:** The authors are grateful for the comments and contributions of the editors, anonymous reviewers, and members of the editorial team.

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

## References

- Jin, F.; Yang, W.; Fu, J.; Li, Z. Effects of vegetation and climate on the changes of soil erosion in the Loess Plateau of China. *Sci. Total Environ.* **2021**, *773*, 145514. [[CrossRef](#)] [[PubMed](#)]
- Vrieling, A. Satellite remote sensing for water erosion assessment: A review. *CATENA* **2006**, *65*, 2–18. [[CrossRef](#)]
- Lamane, H.; Moussadek, R.; Baghdad, B.; Mouhir, L.; Briak, H.; Laghlimi, M.; Zouahri, A. Soil water erosion assessment in Morocco through modeling and fingerprinting applications: A review. *Heliyon* **2022**, *8*, e10209. [[CrossRef](#)]
- Borrelli, P.; Robinson, D.A.; Fleischer, L.R.; Lugato, E.; Ballabio, C.; Alewell, C.; Meusburger, K.; Modugno, S.; Schütt, B.; Ferro, V. An assessment of the global impact of 21st century land use change on soil erosion. *Nat. Commun.* **2017**, *8*, 2013. [[CrossRef](#)]
- Liu, B.; Xie, Y.; Li, Z.; Liang, Y.; Zhang, W.; Fu, S.; Yin, S.; Wei, X.; Zhang, K.; Wang, Z. The assessment of soil loss by water erosion in China. *Int. Soil Water Conserv. Res.* **2020**, *8*, 430–439. [[CrossRef](#)]
- Wuepper, D.; Borrelli, P.; Finger, R. Countries and the global rate of soil erosion. *Nat. Sustain.* **2020**, *3*, 51–55. [[CrossRef](#)]
- Khatami, R.; Mountrakis, G.; Stehman, S.V. A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research. *Remote Sens. Environ.* **2016**, *177*, 89–100. [[CrossRef](#)]
- Long, C.; Li, X.; Jing, Y.; Shen, H. Bishift Networks for Thick Cloud Removal with Multitemporal Remote Sensing Images. *Int. J. Intell. Syst.* **2023**, *2023*, e9953198. [[CrossRef](#)]
- Dimitrovski, I.; Kitanovski, I.; Kocev, D.; Simidjievski, N. Current trends in deep learning for Earth Observation: An open-source benchmark arena for image classification. *ISPRS J. Photogramm. Remote Sens.* **2023**, *197*, 18–35. [[CrossRef](#)]

10. Shen, H.; Zhou, W.; Li, X. A Fast Globally Optimal Seamline Detection Method for High-Resolution Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 6003305. [[CrossRef](#)]
11. Tan, Z.; Gao, M.; Li, X.; Jiang, L. A Flexible Reference-Insensitive Spatiotemporal Fusion Model for Remote Sensing Images Using Conditional Generative Adversarial Network. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5601413. [[CrossRef](#)]
12. Sun, X.; Wang, P.; Yan, Z.; Xu, F.; Wang, R.; Diao, W.; Chen, J.; Li, J.; Feng, Y.; Xu, T.; et al. FAIR1M: A benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery. *ISPRS J. Photogramm. Remote Sens.* **2022**, *184*, 116–130. [[CrossRef](#)]
13. Yang, X.; Dong, M.; Wang, Z.; Gao, L.; Zhang, L.; Xue, J.-H. Data-augmented matched subspace detector for hyperspectral subpixel target detection. *Pattern Recognit.* **2020**, *106*, 107464. [[CrossRef](#)]
14. Li, K.; Wan, G.; Cheng, G.; Meng, L.; Han, J. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS J. Photogramm. Remote Sens.* **2020**, *159*, 296–307. [[CrossRef](#)]
15. Deng, J.; Dong, W.; Socher, R.; Li, L.; Li, K.; Li, F. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
16. Everingham, M.; Van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The pascal visual object classes (voc) challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [[CrossRef](#)]
17. Benedek, C.; Descombes, X.; Zerubia, J. Building Development Monitoring in Multitemporal Remotely Sensed Image Pairs with Stochastic Birth-Death Dynamics. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 33–50. [[CrossRef](#)]
18. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 740–755.
19. Zhu, H.; Chen, X.; Dai, W.; Fu, K.; Ye, Q.; Jiao, J. Orientation robust object detection in aerial images using deep convolutional neural network. In Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 27–30 September 2015; pp. 3735–3739.
20. Liu, K.; Mattyus, G. Fast multiclass vehicle detection on aerial images. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1938–1942.
21. Cheng, G.; Zhou, P.; Han, J. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 7405–7415. [[CrossRef](#)]
22. Razakarivony, S.; Jurie, F. Vehicle detection in aerial imagery: A small target detection benchmark. *J. Vis. Commun. Image Represent.* **2016**, *34*, 187–203. [[CrossRef](#)]
23. Liu, Z.; Wang, H.; Weng, L.; Yang, Y. Ship rotated bounding box space for ship extraction from high-resolution optical satellite images with complex backgrounds. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1074–1078. [[CrossRef](#)]
24. Mundhenk, T.N.; Konjevod, G.; Sakla, W.A.; Boakye, K. A large contextual dataset for classification, detection and counting of cars with deep learning. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 785–800.
25. Long, Y.; Gong, Y.; Xiao, Z.; Liu, Q. Accurate object localization in remote sensing images based on convolutional neural networks. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2486–2498. [[CrossRef](#)]
26. Zou, Z.; Shi, Z. Random access memories: A new paradigm for target detection in high resolution aerial remote sensing images. *IEEE Trans. Image Process.* **2017**, *27*, 1100–1111. [[CrossRef](#)]
27. Yang, M.Y.; Liao, W.; Li, X.; Rosenhahn, B. Deep learning for vehicle detection in aerial images. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 3079–3083.
28. Xia, G.; Bai, X.; Ding, J.; Zhu, Z.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; Zhang, L. DOTA: A large-scale dataset for object detection in aerial images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3974–3983.
29. Wang, Y.; Bashir, S.M.A.; Khan, M.; Ullah, Q.; Wang, R.; Song, Y.; Guo, Z.; Niu, Y. Remote sensing image super-resolution and object detection: Benchmark and state of the art. *Expert Syst. Appl.* **2022**, *197*, 116793. [[CrossRef](#)]
30. Yang, Y.; Newsam, S. Bag-of-visual-words and spatial extensions for land-use classification. In Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, San Jose, CA, USA, 2–5 November 2010; pp. 270–279.
31. Hu, J.; Jiang, T.; Tong, X.; Xia, G.-S.; Zhang, L. A benchmark for scene classification of high spatial resolution remote sensing imagery. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31 July 2015; pp. 5003–5006.
32. Zou, Q.; Ni, L.; Zhang, T.; Wang, Q. Deep learning based feature selection for remote sensing scene classification. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 2321–2325. [[CrossRef](#)]
33. Penatti, O.A.B.; Nogueira, K.; dos Santos, J.A. Do Deep Features Generalize From Everyday Objects to Remote Sensing and Aerial Scenes Domains. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Boston, MA, USA, 7–12 June 2015; pp. 44–51.
34. Zhao, L.; Tang, P.; Huo, L. Feature significance-based multibag-of-visual-words model for remote sensing image scene classification. *J. Appl. Remote Sens.* **2016**, *10*, 035004. [[CrossRef](#)]
35. Zhao, B.; Zhong, Y.; Xia, G.-S.; Zhang, L. Dirichlet-derived multiple topic scene classification model for high spatial resolution remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **2015**, *54*, 2108–2123. [[CrossRef](#)]

36. Xia, G.-S.; Hu, J.; Hu, F.; Shi, B.; Bai, X.; Zhong, Y.; Zhang, L.; Lu, X. AID: A benchmark data set for performance evaluation of aerial scene classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3965–3981. [[CrossRef](#)]
37. Jin, P.; Xia, G.-S.; Hu, F.; Lu, Q.; Zhang, L. AID++: An Updated Version of AID on Scene Classification. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 4721–4724.
38. Wang, Q.; Liu, S.; Chanussot, J.; Li, X. Scene classification with recurrent attention of VHR remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 1155–1167. [[CrossRef](#)]
39. Zhou, W.; Newsam, S.; Li, C.; Shao, Z. PatternNet: A benchmark dataset for performance evaluation of remote sensing image retrieval. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 197–209. [[CrossRef](#)]
40. Zhong, Y.; Su, Y.; Wu, S.; Zheng, Z.; Zhao, J.; Ma, A.; Zhu, Q.; Ye, R.; Li, X.; Pellikka, P. Open-source data-driven urban land-use mapping integrating point-line-polygon semantic objects: A case study of Chinese cities. *Remote Sens. Environ.* **2020**, *247*, 111838. [[CrossRef](#)]
41. Li, H.; Dou, X.; Tao, C.; Wu, Z.; Chen, J.; Peng, J.; Deng, M.; Zhao, L. RSI-CB: A Large-Scale Remote Sensing Image Classification Benchmark Using Crowdsourced Data. *Sensors* **2020**, *20*, 1594. [[CrossRef](#)]
42. Barrena-González, J.; Rodrigo-Comino, J.; Gyasi-Agyei, Y.; Pulido Fernández, M.; Cerdà, A. Applying the RUSLE and ISUM in the Tierra de Barros Vineyards (Extremadura, Spain) to Estimate Soil Mobilisation Rates. *Land* **2020**, *9*, 93. [[CrossRef](#)]
43. Helber, P.; Bischke, B.; Dengel, A.; Borth, D. EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 2217–2226. [[CrossRef](#)]
44. Sumbul, G.; Charfuelan, M.; Demir, B.; Markl, V. Bigearthnet: A large-scale benchmark archive for remote sensing image understanding. In Proceedings of the IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 5901–5904.
45. Qi, X.; Zhu, P.; Wang, Y.; Zhang, L.; Peng, J.; Wu, M.; Chen, J.; Zhao, X.; Zang, N.; Mathiopoulos, P.T. MLRSNet: A multi-label high spatial resolution remote sensing dataset for semantic scene understanding. *ISPRS J. Photogramm. Remote Sens.* **2020**, *169*, 337–350. [[CrossRef](#)]
46. Sumbul, G.; De Wall, A.; Kreuziger, T.; Marcelino, F.; Costa, H.; Benevides, P.; Caetano, M.; Demir, B.; Markl, V. BigEarthNet-MM: A Large-Scale, Multimodal, Multilabel Benchmark Archive for Remote Sensing Image Classification and Retrieval. *IEEE Geosci. Remote Sens. Mag.* **2021**, *9*, 174–180. [[CrossRef](#)]
47. Hong, D.; Hu, J.; Yao, J.; Chanussot, J.; Zhu, X.X. Multimodal remote sensing benchmark datasets for land cover classification with a shared and specific feature learning model. *ISPRS J. Photogramm. Remote Sens.* **2021**, *178*, 68–80. [[CrossRef](#)]
48. Li, L.; Yao, X.; Cheng, G.; Han, J. AIFS-DATASET for Few-Shot Aerial Image Scene Classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5618211. [[CrossRef](#)]
49. Li, Y.; Zhou, Y.; Zhang, Y.; Zhong, L.; Wang, J.; Chen, J. DKDFN: Domain Knowledge-Guided deep collaborative fusion network for multimodal unitemporal remote sensing land cover classification. *ISPRS J. Photogramm. Remote Sens.* **2022**, *186*, 170–189. [[CrossRef](#)]
50. Zhang, Z.; Zhang, M.; Gong, J.; Hu, X.; Xiong, H.; Zhou, H.; Cao, Z. LuoJiaAI: A cloud-based artificial intelligence platform for remote sensing image interpretation. *Geo-Spat. Inf. Sci.* **2023**, *1–24*. [[CrossRef](#)]
51. Papoutsis, I.; Bountos, N.I.; Zavras, A.; Michail, D.; Tryfonopoulos, C. Benchmarking and scaling of deep learning models for land cover image classification. *ISPRS J. Photogramm. Remote Sens.* **2023**, *195*, 250–268. [[CrossRef](#)]
52. Zhang, H.; Zhang, R.; Qi, F.; Liu, X.; Niu, Y.; Fan, Z.; Zhang, Q.; Li, J.; Yuan, L.; Song, Y. The CSLE model based soil erosion prediction: Comparisons of sampling density and extrapolation method at the county level. *CATENA* **2018**, *165*, 465–472. [[CrossRef](#)]
53. Dong, Y.; Liang, T.; Yang, C.; Luo, H.; Zhang, Y. Joint Distance Transfer Metric Learning for Remote-Sensing Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 6506205. [[CrossRef](#)]
54. Yang, C.; Dong, Y.; Du, B.; Zhang, L. Attention-Based Dynamic Alignment and Dynamic Distribution Adaptation for Remote Sensing Cross-Domain Scene Classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5634713. [[CrossRef](#)]
55. Hosseiny, B.; Rastiveis, H.; Homayouni, S. An Automated Framework for Plant Detection Based on Deep Simulated Learning from Drone Imagery. *Remote Sens.* **2020**, *12*, 3521. [[CrossRef](#)]
56. Northcutt, C.; Jiang, L.; Chuang, I. Confident Learning: Estimating Uncertainty in Dataset Labels. *J. Artif. Intell. Res.* **2021**, *70*, 1373–1411. [[CrossRef](#)]
57. Abdollahi, A.; Pradhan, B.; Shukla, N.; Chakraborty, S.; Alamri, A. Deep learning approaches applied to remote sensing datasets for road extraction: A state-of-the-art review. *Remote Sens.* **2020**, *12*, 1444. [[CrossRef](#)]
58. Lekki, J.; Deutsch, E.; Sayers, M.; Bosse, K.; Anderson, R.; Tokars, R.; Sawtell, R. Determining remote sensing spatial resolution requirements for the monitoring of harmful algal blooms in the Great Lakes. *J. Great Lakes Res.* **2019**, *45*, 434–443. [[CrossRef](#)]
59. Velmurugan, K.; Saravanasankar, S.; Venkumar, P.; Sudhakarapandian, R.; Bona, G.D. Hybrid fuzzy AHP-TOPSIS framework on human error factor analysis: Implications to developing optimal maintenance management system in the SMEs. *Sustain. Futur.* **2022**, *4*, 100087. [[CrossRef](#)]
60. Di Bona, G.; Falcone, D.; Forcina, A.; De Carlo, F.; Silvestri, L. Quality Checks Logit Human Reliability (LHR): A New Model to Evaluate Human Error Probability (HEP). *Math. Probl. Eng.* **2021**, *2021*, e6653811. [[CrossRef](#)]
61. Liang, X.; Liu, X.; Yao, L. Review—a survey of learning from noisy labels. *ECS Sens. Plus* **2022**, *1*, 021401. [[CrossRef](#)]

62. González-Rivero, M.; Beijbom, O.; Rodriguez-Ramirez, A.; Holtrop, T.; González-Marrero, Y.; Ganase, A.; Roelfsema, C.; Phinn, S.; Hoegh-Guldberg, O. Scaling up ecological measurements of coral reefs using semi-automated field image collection and analysis. *Remote Sens.* **2016**, *8*, 30. [[CrossRef](#)]
63. Chang, C.-M.; Lee, C.-H.; Igarashi, T. Spatial labeling: Leveraging spatial layout for improving label quality in non-expert image annotation. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan, 8–13 May 2021; pp. 1–12.
64. Bona, G.D.; Falcone, D.; Forcina, A.; Silvestri, L. Systematic human reliability analysis (SHRA): A new approach to evaluate human error probability (HEP) in a nuclear plant. *Int. J. Math. Eng. Manag. Sci.* **2021**, *6*, 345–362.
65. Gotovac, S.; Zelenika, D.; Marušić, Ž.; Božić-Štulić, D. Visual-based person detection for search-and-rescue with uas: Humans vs. machine learning algorithm. *Remote Sens.* **2020**, *12*, 3295. [[CrossRef](#)]
66. Gupta, E.; Das, S.; Rajani, M.B. Archaeological exploration in Srirangapatna and its environ through remote sensing analysis. *J. Indian Soc. Remote Sens.* **2017**, *45*, 1057–1063. [[CrossRef](#)]
67. Zhao, Y.; Feng, D.; Yu, L.; See, L.; Fritz, S.; Perger, C.; Gong, P. Assessing and improving the reliability of volunteered land cover reference data. *Remote Sens.* **2017**, *9*, 1034. [[CrossRef](#)]
68. Sajjad, H.; Kumar, P. Future challenges and perspective of remote sensing technology. In *Applications and Challenges of Geospatial Technology: Potential and Future Trends*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 275–277.
69. Fritz, S.; See, L.; Perger, C.; McCallum, I.; Schill, C.; Schepaschenko, D.; Duerauer, M.; Karner, M.; Dresel, C.; Laso-Bayas, J.-C.; et al. A global dataset of crowdsourced land cover and land use reference data. *Sci Data* **2017**, *4*, 170075. [[CrossRef](#)]
70. Wang, X. Information extraction of tourist geological resources based on 3D visualization remote sensing image. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *XLIII-3*, 1815–1820. [[CrossRef](#)]
71. Weiers, S.; Bock, M.; Wissen, M.; Rossner, G. Mapping and indicator approaches for the assessment of habitats at different scales using remote sensing and GIS methods. *Landsc. Urban Plan.* **2004**, *67*, 43–65. [[CrossRef](#)]
72. Duan, Y.; Li, X.; Zhang, L.; Chen, D.; Ji, H. Mapping national-scale aquaculture ponds based on the Google Earth Engine in the Chinese coastal zone. *Aquaculture* **2020**, *520*, 734666. [[CrossRef](#)]
73. Wu, Y.; Zhang, P.; Wu, J.; Li, C. Object-oriented and deep-learning-based high-resolution mapping from large remote sensing imagery. *Can. J. Remote Sens.* **2021**, *47*, 396–412. [[CrossRef](#)]
74. Zhao, Y.; Gong, P.; Yu, L.; Hu, L.; Li, X.; Li, C.; Zhang, H.; Zheng, Y.; Wang, J.; Zhao, Y.; et al. Towards a common validation sample set for global land-cover mapping. *Int. J. Remote Sens.* **2014**, *35*, 4795–4814. [[CrossRef](#)]
75. Behera, M.D.; Gupta, A.K.; Barik, S.K.; Das, P.; Panda, R.M. Use of satellite remote sensing as a monitoring tool for land and water resources development activities in an Indian tropical site. *Environ. Monit. Assess.* **2018**, *190*, 401. [[CrossRef](#)]
76. Zhang, M.; Zhao, X.; Li, W.; Zhang, Y.; Tao, R.; Du, Q. Cross-Scene Joint Classification of Multisource Data With Multilevel Domain Adaption Network. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, 1–13. [[CrossRef](#)]
77. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52. [[CrossRef](#)]
78. Su, Y.; Zhong, Y.; Zhu, Q.; Zhao, J. Urban scene understanding based on semantic and socioeconomic features: From high-resolution remote sensing imagery to multi-source geographic datasets. *ISPRS J. Photogramm. Remote Sens.* **2021**, *179*, 50–65. [[CrossRef](#)]
79. Johnson, B.A.; Iizuka, K. Integrating OpenStreetMap crowdsourced data and Landsat time-series imagery for rapid land use/land cover (LULC) mapping: Case study of the Laguna de Bay area of the Philippines. *Appl. Geogr.* **2016**, *67*, 140–149. [[CrossRef](#)]
80. Happ, P.N.; Ferreira, R.S.; Bentes, C.; Costa, G.; Feitosa, R.Q. Multiresolution segmentation: A parallel approach for high resolution image segmentation in multicore architectures. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2010**, *38*, C7.
81. Liu, Z.; Mao, H.; Wu, C.-Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A ConvNet for the 2020s. *arXiv* **2022**, arXiv:2201.03545. [[CrossRef](#)]
82. Rao, Y.; Zhao, W.; Tang, Y.; Zhou, J.; Lim, S.N.; Lu, J. Hornet: Efficient high-order spatial interactions with recursive gated convolutions. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 10353–10366.
83. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
84. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
85. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
86. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 10012–10022.
87. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
88. Li, C.; Yao, J.; Li, R.; Zhu, Y.; Yao, H.; Zhang, P.; Wei, D.; Zhao, S.; Li, Y.; Wu, Y. “3S” technologies and application for dynamic monitoring soil and water loss in the Yangtze river basin, China. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *43*, 1563–1567. [[CrossRef](#)]
89. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. *arXiv* **2017**, arXiv:1711.05101.

90. Martins, V.S.; Kaleita, A.L.; Gelder, B.K.; da Silveira, H.L.F.; Abe, C.A. Exploring multiscale object-based convolutional neural network (multi-OCNN) for remote sensing image classification at high spatial resolution. *ISPRS J. Photogramm. Remote Sens.* **2020**, *168*, 56–73. [[CrossRef](#)]
91. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. [[CrossRef](#)]
92. Li, C.; Fang, S.; Geng, X.; Yuan, Y.; Zheng, X.; Zhang, D.; Li, R.; Sun, W.; Wang, X. Coastal ecosystem service in response to past and future land use and land cover change dynamics in the Yangtze river estuary. *J. Clean. Prod.* **2023**, *385*, 135601. [[CrossRef](#)]
93. Xiao, R.; Cao, W.; Liu, Y.; Lu, B. The impacts of landscape patterns spatio-temporal changes on land surface temperature from a multi-scale perspective: A case study of the Yangtze River Delta. *Sci. Total Environ.* **2022**, *821*, 153381. [[CrossRef](#)]
94. Zhang, S.; Yang, P.; Xia, J.; Wang, W.; Cai, W.; Chen, N.; Hu, S.; Luo, X.; Li, J.; Zhan, C. Land use/land cover prediction and analysis of the middle reaches of the Yangtze River under different scenarios. *Sci. Total Environ.* **2022**, *833*, 155238. [[CrossRef](#)]
95. Zhang, J.; Li, W.; Zhai, L. Understanding geographical conditions monitoring: A perspective from China. *Int. J. Digit. Earth* **2015**, *8*, 38–57. [[CrossRef](#)]
96. Liu, Y.; Ding, L.; Chen, C.; Liu, Y. Similarity-Based Unsupervised Deep Transfer Learning for Remote Sensing Image Retrieval. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 7872–7889. [[CrossRef](#)]

**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.