



Article

Generating 10-Meter Resolution Land Use and Land Cover Products Using Historical Landsat Archive Based on Super Resolution Guided Semantic Segmentation Network

Dawei Wen ^{1,2}, Shihao Zhu ^{1,2}, Yuan Tian ³, Xuehua Guan ⁴ and Yang Lu ^{5,*}

¹ School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan 430205, China; daweiwen@wit.edu.cn (D.W.); 22207010107@stu.wit.edu.cn (S.Z.)

² Hubei Key Laboratory of Intelligent Robot, Wuhan Institute of Technology, Wuhan 430205, China

³ School of Systems Science, Beijing Normal University, Beijing 100875, China; 202031250017@mail.bnu.edu.cn

⁴ China Siwei Surveying and Mapping Technology Co., Ltd., Beijing 100044, China; guanxuehua@chinasiwei.com

⁵ School of Civil Engineering and Architecture, Wuhan Institute of Technology, Wuhan 430074, China

* Correspondence: 23101101@wit.edu.cn

Abstract: Generating high-resolution land cover maps using relatively lower-resolution remote sensing images is of great importance for subtle analysis. However, the domain gap between real lower-resolution and synthetic images has not been permanently resolved. Furthermore, super-resolution information is not fully exploited in semantic segmentation models. By solving the aforementioned issues, a deeply fused super resolution guided semantic segmentation network using 30 m Landsat images is proposed. A large-scale dataset comprising 10 m Sentinel-2, 30 m Landsat-8 images, and 10 m European Space Agency (ESA) Land Cover Product is introduced, facilitating model training and evaluation across diverse real-world scenarios. The proposed Deeply Fused Super Resolution Guided Semantic Segmentation Network (DFSRSSN) combines a Super Resolution Module (SRResNet) and a Semantic Segmentation Module (CRFFNet). SRResNet enhances spatial resolution, while CRFFNet leverages super-resolution information for finer-grained land cover classification. Experimental results demonstrate the superior performance of the proposed method in five different testing datasets, achieving 68.17–83.29% and 39.55–75.92% for overall accuracy and kappa, respectively. When compared to ResUnet with up-sampling block, increases of 2.16–34.27% and 8.32–43.97% were observed for overall accuracy and kappa, respectively. Moreover, we proposed a relative drop rate of accuracy metrics to evaluate the transferability. The model exhibits improved spatial transferability, demonstrating its effectiveness in generating accurate land cover maps for different cities. Multi-temporal analysis reveals the potential of the proposed method for studying land cover and land use changes over time. In addition, a comparison of the state-of-the-art full semantic segmentation models indicates that spatial details are fully exploited and presented in semantic segmentation results by the proposed method.

Keywords: deep learning; super resolution; land cover; Sentinel; Landsat; semantic segmentation; remote sensing



Citation: Wen, D.; Zhu, S.; Tian, Y.; Guan, X.; Lu, Y. Generating 10-Meter Resolution Land Use and Land Cover Products Using Historical Landsat Archive Based on Super Resolution Guided Semantic Segmentation Network. *Remote Sens.* **2024**, *16*, 2248. <https://doi.org/10.3390/rs16122248>

Academic Editors: Li Fang, Jian Yang and Yu Feng

Received: 14 May 2024

Revised: 15 June 2024

Accepted: 18 June 2024

Published: 20 June 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

Satellite-derived Land Use and Land Cover (LULC) products serve as vital data sources for an increasing number of applications, such as urban planning [1], ecological evaluation [2], climate change analysis [3], agricultural management [4], and policy making [5]. In light of a rapidly developing world, higher resolution and regularly updated LULC maps are required [6]. So far, annual LULC products have been produced even at a global scale, including the Moderate Resolution Imaging Spectroradiometer (MODIS) land cover-type product MCD12Q1 [7,8] (2001–2018, 500 m), European Space Agency Climate

Change Initiative (ESA-CCI) Land Cover [9,10] (1992–2020, 300 m), and Copernicus Global Land Service Dynamic Land Cover Map (CGLS-LC100) (2015–2019, 100 m). However, these global products have coarse spatial resolutions. With the free availability of high-resolution satellite data, e.g., Landsat (30 m in visible and near-infrared bands) and Sentinel (10 m), land cover maps with more spatial detail can be generated. Recently, higher-resolution satellite-derived global urban land cover maps are available. Gong et al. [11] generated and publicly released the first 30 m global land cover product in 2010, i.e., Finer Resolution Observation and Monitoring of Global Land Cover (FROM-GLC). More recently, 30 m FROM-GLC for the years 2010/2015/2017, 10 m FROM-GLC in 2017 [12], and 10 m ESA WorldCover Product for 2020 and 2021 [13] have been produced. In 2022, the first near real-time global product using Sentinel images, Dynamic World [14], with the spatial resolution of 10 m, was proposed. Although the spatial resolution and temporal frequency of these products are significantly improved, 10 m products only provide LULC data preceding 2015 since the 10 m Sentinel-2 satellite was launched on 27 June 2015. For longer time-series high-resolution satellites images (1984–current), the 30 m Landsat offers suitable alternatives but fails to provide fine-grained LULC information. Therefore, generating high-resolution LULC maps using relatively lower-resolution remote sensing images is of great importance.

In general, there are two kinds of approaches to generate high-resolution maps from lower-resolution images: stage-wise and end-to-end approaches (Table 1). Stage-wise approaches conduct semantic segmentation and super resolution (SR) independently [15]. For example, Cui et al. [16] proposed a novel stage-wise green tide extraction method for MODIS images. In this method, a super resolution model was trained with high spatial resolution Gaofen-1, and then semantic segmentation network for images with improved spatial resolution was introduced. Fu, et al. [17] proposed a method to classify marsh vegetation using multi-resolution multi-spectral and hyperspectral images, where independent super resolution and semantic segmentation techniques were employed and combined. Zhu et al. [18] proposed a CNN-based super-resolution method for multi-temporal classification. It was conducted in two stages: image super-resolution preprocessing and LULC classification. Huang, et al. [19] employed Enhanced Super-Resolution Generative Adversarial Networks for super-resolution image construction, subsequently integrating semantic segmentation models for the classification of tree species. Since restored images with better super resolution quality do not always guarantee better semantic segmentation results and super-resolution and semantic segmentation can promote each other [20], end-to-end approaches with interaction between these two tasks are introduced. Xu et al. [15] proposed an end-to-end network to extract high-resolution building maps from low-resolution, and high-resolution images are not needed in the training phase. In [21], the multi-task encoder-decoder network accepts low-resolution images as input and uses two different branches for super resolution and segmentation. Information interaction of these two subtasks is realized via shared decoder and overall weighted loss. Similar work in [22] used a multi-loss approach for training a multi-task network. In addition, a feature affinity loss was added to combine the learning of both branches. Salgueiro et al. [23] introduced a multi-task network designed to generate super-resolution images and semantic segmentation results utilizing freely available Sentinel-2 imagery. These studies jointly train super resolution and semantic segmentation using overall loss function, and information interaction between the two tasks relies on shared network structures. These kinds of networks attempt to solve the super resolution and semantic segmentation simultaneously, which can work well when high-resolution target images and low-resolution input images are just original images and their corresponding down-sampling results. The pivotal consideration is to attain an optimal balance in multi-task joint learning. However, in real-world situations, remote sensing images suffer from various image degradation factors such as blur and noise, due to imperfect illumination, atmospheric propagation, lens imaging, sensor quantification, temporal difference, etc. [24]. Therefore, super resolution for real multi-resolution satellite

images is more challenging. When coupled with the semantic segmentation task and treated equally, super resolution limits the final segmentation performance.

Table 1. A summary of approaches to generate high-resolution maps from lower-resolution images.

Category	Description	Characteristics	References
Stage-wise methods	Perform semantic segmentation and super resolution tasks independently.	Enhanced super-resolution quality in restored images does not consistently lead to improved semantic segmentation outcomes.	[16–19]
End-to-end methods	These two tasks are integrated within an end-to-end network.	Super-resolution and semantic segmentation can promote each other, but the balance of multi-task joint learning should be carefully handled.	[15,21–23]

Super resolution aided semantic segmentation is another research direction where semantic segmentation and super resolution are considered as main and auxiliary tasks, respectively. As stated in [25], main tasks are designed to produce final required output for an application, and auxiliary tasks serve for learning and supporting the main tasks. Following this research trajectory, in [26], the main task is land cover classification and the super resolution task aims to increase the resolution of remote sensing images. In [27], a target-guided feature super resolution network is proposed for vehicle detection, where features of small objects are enhanced for better detection results. However, a few existing methods have attempted to consider super resolution just as an auxiliary task by introducing a guidance module for main tasks under the constraint of super resolution. However, these methods still train the whole network in an end-to-end manner using a multi-task loss function. In addition, the employed datasets are just high-resolution remote sensing images and down-sampled low-resolution ones.

In summary, there are two main challenges in generating high-resolution LULC maps using relatively lower-resolution remote sensing images. Firstly, it is a predicament to guarantee the consistency of image degradation factors for real multi-resolution satellite images. Therefore, synthetic lower-resolution images generated by simple degradation models (e.g., bicubic down-sampling) can lead to domain gap, making models that do not generalize well to real-world data. At present, this domain gap between real lower-resolution and synthetic ones is very common and has not been solved. Second, higher super resolution accuracy does not guarantee better higher classification performance, especially for real multi-resolution satellite images. The general end-to-end training with multi-task loss function cannot deal with this problem. Therefore, how to treat the different tasks is very crucial. Based on these aforementioned problems, the deeply fused super resolution guided semantic segmentation network using real lower-resolution images is proposed. The main contributions of this article are reflected in the following aspects.

- (1) We introduce an open-source, large-scale, diverse dataset for super resolution and semantic segmentation tasks, situated within real-world scenarios. This dataset serves as a pivotal resource, directing the capturing of intricate land cover details from lower-resolution remote sensing images. In this dataset, the testing and training data are from different cities, which enables performance evaluation in terms of generalization capability and transferability.
- (2) We present a novel network architecture named the Deeply Fused Super Resolution Guided Semantic Segmentation Network (DFSRSSN), which combines super resolution and semantic segmentation tasks. By enhancing feature reuse of the super-resolution module, preserving detailed information, and fusing cross-resolution features, the finer-grained land cover information using lower-resolution remote sensing images was achieved. By integrating the super-resolution module into the semantic segmentation network, our model achieves generation of 10 m land cover maps using 30 m Landsat images, effectively enhancing spatial resolution. For training the entire

network, we leverage pre-trained super resolution parameters to initialize the weights of the semantic segmentation module. This strategy expedites model convergence and improves prediction accuracy.

An overview of the created dataset for the super resolution guided semantic segmentation method is provided in Section 2. The proposed method is presented in Section 3, followed by the experiments and analysis of the experimental results in Section 4. Finally, Section 5 concludes this article.

2. Dataset and Methodology

In this paper, we created a dataset for a super resolution guided semantic segmentation network. Based on this dataset, we propose a novel deeply fused super resolution guided semantic segmentation network, denoted as DFSRSSN, for the purpose of generating 10 m land cover product from Landsat imagery, as shown in Figure 1. The DFSRSSN comprises two fundamental components: (1) a super resolution module, i.e., Super Resolution Residual Network (SRResNet), and (2) a semantic segmentation module, i.e., Cross-Resolution Feature Fusion Network (CRFFNet). The performance of SR inherently affects the semantic segmentation module since SR features are fused in the semantic segmentation module. First, we train the super resolution module to achieve the maximum possible SR performance with the given architecture and data. Subsequently, pre-trained SR parameters, obtained through SRResNet, are leveraged as the initial weights during the training of the semantic segmentation module. This facilitates faster model convergence and higher prediction accuracy. This two-step training process enables the semantic segmentation module to learn to effectively leverage the predicted super resolution related information to enhance the classification of land cover types.

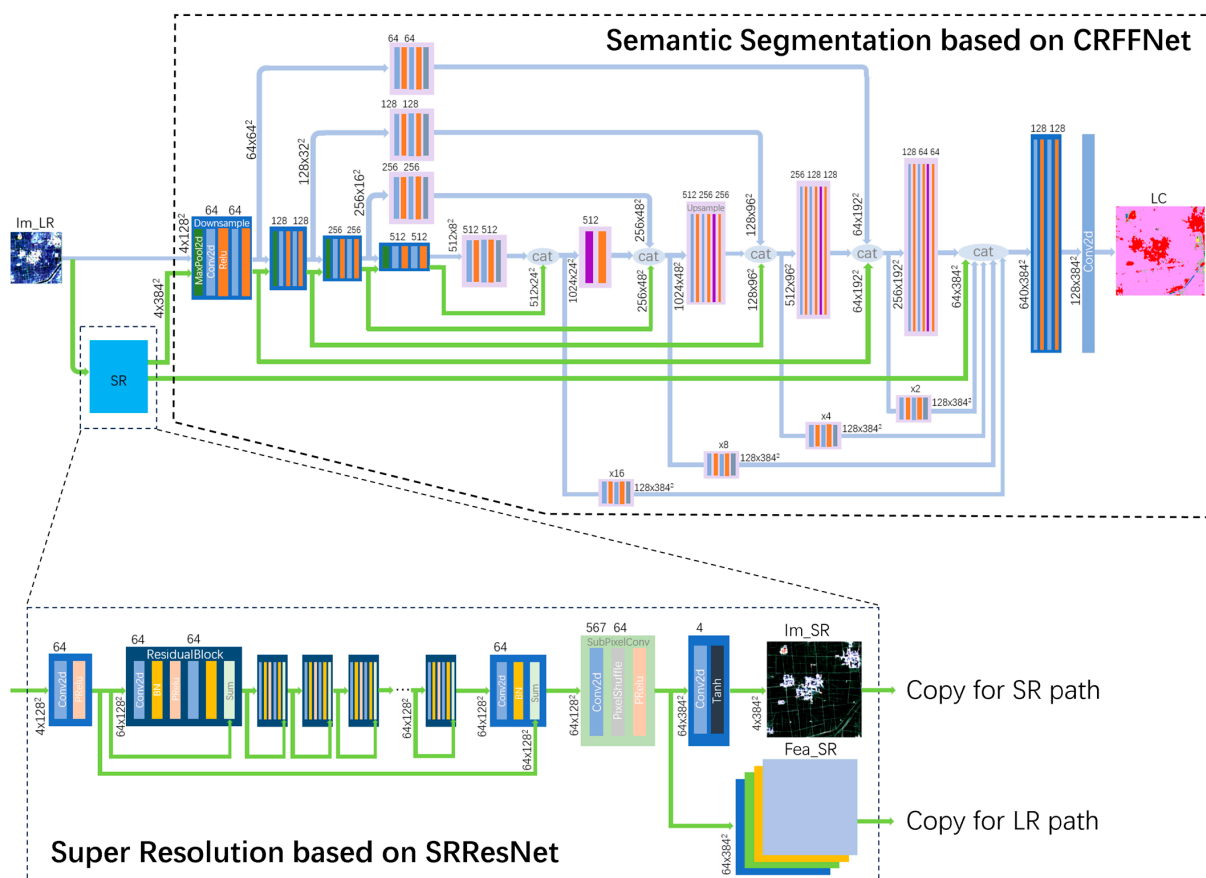


Figure 1. The network structure of the deeply fused super resolution guided semantic segmentation network (DFSRSSN): Super Resolution Residual Network (SRResNet) and Cross-Resolution Feature Fusion Network (CRFFNet).

2.1. Dataset

The proposed dataset contains a total of 10,375 image patches ($3.84 \text{ km} \times 3.84 \text{ km}$). This dataset includes 10 m Sentinel-2 images, 30 m Landsat-8 images, and 10 m European Space Agency (ESA) Land Cover Product. It should be noted that the ESA Land Cover products are mapped globally at a 10 m resolution for the years 2020 and 2021, utilizing Sentinel-1 and Sentinel-2 images. In addition, 8 categories of land cover (i.e., tree cover, shrubland, grassland, cropland, built-up, bare/sparse vegetation, permanent water bodies, and herbaceous wetland) are considered. In the experiments, Landsat-8 images with blue, green, red, and near-infrared bands (Band 2, Band 3, Band 4, and Band 5, respectively) were used as input, and Sentinel-2 images with the same four bands (B2, B3, B4, and B8, respectively) and 10 m ESA Land Cover Product of the year 2020 (ESA 2020) were used as output for super resolution and semantic segmentation tasks, respectively. The ESA 2020 achieved a global overall accuracy of 74.4%, as reported in [28].

For model training and validation, a total of 20 Chinese cities were selected (Table 2), including 2 first-tier cities, 1 second-tier city, 8 third-tier cities, 6 fourth-tier cities, and 3 fifth-tier cities. The 8707 image patches ($3.84 \text{ km} \times 3.84 \text{ km}$) of these Chinese cities were split into 80% for training and 20% for testing (dataset I). In addition, a total of 4 cities, i.e., Wuhan, China (II); Kassel, Germany (III); Aschaffenburg, Germany (IV); and Joliet, US (V), were selected as additional testing datasets. The percentages for different land cover types in ESA Land Cover Product for the five different testing sets are provided in Table 3. Dataset I has the same or similar data distribution as the training data since the data are from the same cities (i.e., 20 Chinese cities). The other four testing datasets (II–V) and training data are from different cities. Therefore, dataset I serves to evaluate model fitting performance on the training dataset, and datasets II–V can be used to properly evaluate the generalization capability and transferability of the model [29].

Table 2. Training and validation samples ($3.84 \text{ km} \times 3.84 \text{ km}$) for our study.

City Category	Cities	# of Samples
First-tier cities	Chengdu	784
	Changsha	170
Second-tier cities	Nanchang	529
Third-tier cities	Shangrao	324
	Jiujiang	414
	Zunyi	650
	Mianyang	405
	Xiangtan	567
	Zibo	275
	Cangzhou	289
	Baoding	425
Fourth-tier cities	Jingdezhen	169
	Yingtian	440
	Yibin	279
	Heze	196
	Binzhou	784
	Dezhou	588
Fifth-tier cities	Fengcheng	63
	Zhaotong	756
	Hengshui	600
Total		8707

Table 3. Percentages of the different land cover types (1–8) for datasets I–V.

Dataset	1	2	3	4	5	6	7	8
I	28.02	0.15	7.10	45.42	8.13	5.62	3.92	1.64
II	19.72	0.03	2.08	41.62	10.99	12.16	12.16	1.24
III	46.56	0.00	18.38	30.30	3.82	0.46	0.47	0.01
IV	54.04	0.00	18.77	22.37	3.92	0.46	0.43	0.01
V	14.41	0.00	8.72	69.81	2.81	1.35	2.71	0.19

Land cover types (%): 1 tree cover, 2 shrubland, 3 grassland, 4 cropland, 5 built-up, 6 bare/sparse vegetation, 7 permanent water bodies, 8 herbaceous wetland.

In dataset I, the land cover class distribution is extremely unbalanced, ranging from 0.15% for shrubland to 45.42% for cropland. Furthermore, the class distribution varied among the other four datasets, exhibiting diverse compositions across different land cover types. Specifically, the majority class type for datasets III and IV is tree cover (46.56% and 54.04%, respectively), and the majority class type for datasets II and V is cropland (41.62% and 69.81%, respectively). Shrubland and herbaceous wetland are extremely rare minority classes with extremely low percentages. Therefore, generating 10 m LULC products using 30 m Landsat images not only faces challenges related to spatial resolution gap and absence of true labels but also involves addressing the issue of an extremely imbalanced long-tailed distribution of land cover classes.

2.2. Super Resolution Module

The super resolution module, built upon the SRResNet architecture, is deployed with the primary objective of improving the spatial resolution of low-resolution images (denoted as Im_{LR}). The spatial resolution is improved by learning a mapping between low-resolution images (Im_{LR}) and high-resolution images (Im_{HR}) while preserving spectral information. The SRResNet, employed as the underlying framework, is characterized by a succession of residual blocks, local and global residual connections, and a pixel shuffle block [30]. The selection of SRResNet for this task is predicated on several rationale-driven considerations. First, the utilization of local and global residual learning can alleviate gradient vanishing problems in deep models, making it feasible to enhance the network performance in learning middle-frequency and high-frequency features and super-resolving images [31]. Secondly, the integration of the pixel shuffle block within SRResNet plays a pivotal role in suppressing the deleterious effects of redundant features on super resolution results. In comparison to the conventional transpose convolution methods, the pixel shuffle block proves to be superior in its ability to yield super resolution results [32]. It is noteworthy that the implementation of SRResNet in this study predominantly comprises 48 residual blocks and a shuffle block.

Peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) are two fundamental quantitative metrics for assessing the degree of similarity between super resolution images and their corresponding high-resolution counterparts. PSNR serves as a measure of image reconstruction quality with regard to noise. This is accomplished by comparing the grayscale values of corresponding pixels. On the other hand, SSIM delves into a more comprehensive evaluation by taking into account three key facets of image structure, namely brightness, contrast, and structure pattern [33].

To ensure the preservation of pixel grayscale values and fine spatial details of the image, both PSNR and SSIM are integrated into the loss function. It is imperative to note that larger PSNR and SSIM values correspond to a superior quality of super resolution images. Accordingly, the training objective of the SRResNet revolves around the minimization of the ensuing loss function, denoted as follows:

$$\text{Loss}(Im_SR, Im_HR) = -\text{PSNR}(Im_SR, Im_HR) - \text{SSIM}(Im_SR, Im_HR) \quad (1)$$

2.3. Semantic Segmentation Module

For the task of semantic segmentation, we have opted for the ResUnet architecture as our foundational framework. This architectural choice is particularly well-suited for facilitating precise pixel-to-pixel semantic segmentation. ResUnet achieves this by establishing connectivity between coarse multi-level features extracted from the initial convolutional layers within the encoder and the corresponding layers within the decoder through the utilization of skip connections, thus synergizing the strengths inherent in both the Unet and residual learning [34].

In this study, we have employed a ResUnet variant characterized by a nine-level architecture, indicating the depth of blocks in the network. This modified ResUnet architecture takes low-resolution images (Im_LR), super resolution images (Im_SR), and super resolution features (Fea_SR) as input, with the objective of producing a land cover (LC) map at a higher resolution. Consequently, adaptations to the classical ResUnet framework were imperative to address the challenge of cross-resolution feature fusion, resulting in the formulation of the CRFFNet.

As illustrated in Figure 1, the CRFFNet contains three primary components, namely, an encoder with multiple inputs, up-sampling blocks, and the multi-level feature fusion within the decoder. More specifically, the Siamese encoder is structured with four down-sampling blocks, which share weights and serve the purpose of feature extraction. The input to the encoder is an image pair, i.e., Im_LR and Im_SR . Subsequently, the down-sampled convolutional features are fused and fed into the Siamese decoder. However, it is essential to recognize that differences exist in the spatial dimensions of features derived from Im_LR and Im_SR at corresponding encoder levels. To address this incongruity, the up-sampling block (UP) is introduced for features derived from Im_LR , which is composed of two convolutions followed by bilinear interpolation up-sampling. In this manner, the cross-resolution feature fusion is described as follows:

$$F_{fuse}^i = \text{cat}(F_{SR}, \text{UP}(F_{LR}, sf), \text{de}(F_{fuse}^{i-1})) \quad (2)$$

where $F_{LR} \in \mathbb{R}^{c \times h \times w}$ and $F_{SR} \in \mathbb{R}^{c \times (sf \times h) \times (sf \times w)}$ denote the features derived from Im_LR and Im_SR , respectively, through the encoder. The variables c , h , and w represent the number of channels, height, and width of the features, while sf denotes the up-sampling factor. In this paper, sf is set as 3 since the spatial resolution of Im_LR is 3 times that in Im_SR . The term de represents deconvolution operations with a scale of 2 to up-sample the fused features at the previous level. In this way, the fused features at the i th level (F_{fuse}^i) of the decoder are concatenated (cat) by F_{SR} , UP features for F_{LR} with the scaling factor of sf , and deconvolution operations of features from the $(i - 1)$ th level (F_{fuse}^{i-1}).

Furthermore, to utilize the valuable features learned during the super resolution task and multi-scale features in the semantic segmentation task, the pixel shuffle outputs from the final residual blocks in the SRResNet and UP of different-level features in the decoder are concatenated with the output from the last up-sampling block in ResUnet. Ultimately, three convolution layers are employed to generate the final land cover map.

2.4. Implementation Details

Our model is implemented on the PyTorch framework, leveraging the computational power of a single NVIDIA RTX 4090 GPU. For optimization, we employed the Adam optimizer with a learning rate of 5×10^{-4} to minimize the loss function. The size of each input batch is 128×128 pixels with four spectral bands, and the batch size is set to 8. A total of 50 epochs is set for SRResNet and DFSRSSN to facilitate model convergence.

As for the evaluation of the proposed segmentation network, overall accuracy (OA) [35] and kappa coefficient (KC) [36] were used to evaluate the overall performance. We used user's accuracy (UA) [37], producer's accuracy (PA) [37], and F1 score [38] to evaluate the ability to classify different land cover types. These metrics are defined as follows:

$$OA = \frac{\sum_{i=1}^n x_{ii}}{N} \quad (3)$$

$$Kappa = \frac{N \sum_{i=1}^n x_{ii} - \sum_{i=1}^n (x_{i+} \times x_{+i})}{N^2 - \sum_{i=1}^n (x_{i+} \times x_{+i})} \quad (4)$$

$$UA = \frac{x_{ii}}{x_{i+}} \quad (5)$$

$$PA = \frac{x_{ii}}{x_{+i}} \quad (6)$$

$$F1 = 2 \frac{UA \times PA}{UA + PA} \quad (7)$$

where x_{ii} is the number of correctly classified pixels for the i th land cover type; x_{i+} and x_{+i} is the number of pixels for the i th land cover type in the classification result and reference data, respectively; n and N are the total number of land cover types and samples, respectively. OA is calculated as the ratio between the number of correctly classified pixels and the total number of pixels. KC is also used to measure the agreement between the classified result and the reference data but excluding chance agreement [35]. KC is more appropriate with imbalanced class distributions than OA. UA quantifies the proportion of i th land cover type in the classification result consistent with the reference data, while PA quantifies the probability of i th land cover type on the ground correctly classified by the classification result, thereby indicating commission errors and omission errors [39]. F1 score, the harmonic mean of UA and PA, is a tradeoff metric to quantify class-wise commission errors and omission errors [40].

3. Results

3.1. Overall Performance

The ResUnet with up-sampling block (ResUnet_UP) was compared to our proposed DFSRSSN in order to identify the role of super-resolution guided information to improve finer-resolution segmentation results. The ResUnet_UP only takes Im_{LR} (30 m) as input to the encoder, and each down-sampled convolutional feature is up-sampled by bilinear interpolation up-sampling to ensure the generation of 10 m land cover map. Figure 2 presents four representative examples of land cover mapping results using Landsat-8 images. Table 4 summarizes the OA and Kappa of the two methods. Compared to the ESA in Figure 2e, the proposed method can better identify different land cover types while retaining fine spatial details. For built-up areas mainly consisting of roads and buildings, with the aid of super-resolution information, more precise and accurate boundaries can be obtained. For tree cover, ResUnet_UP failed to capture accurate boundaries and contains more misclassification errors.

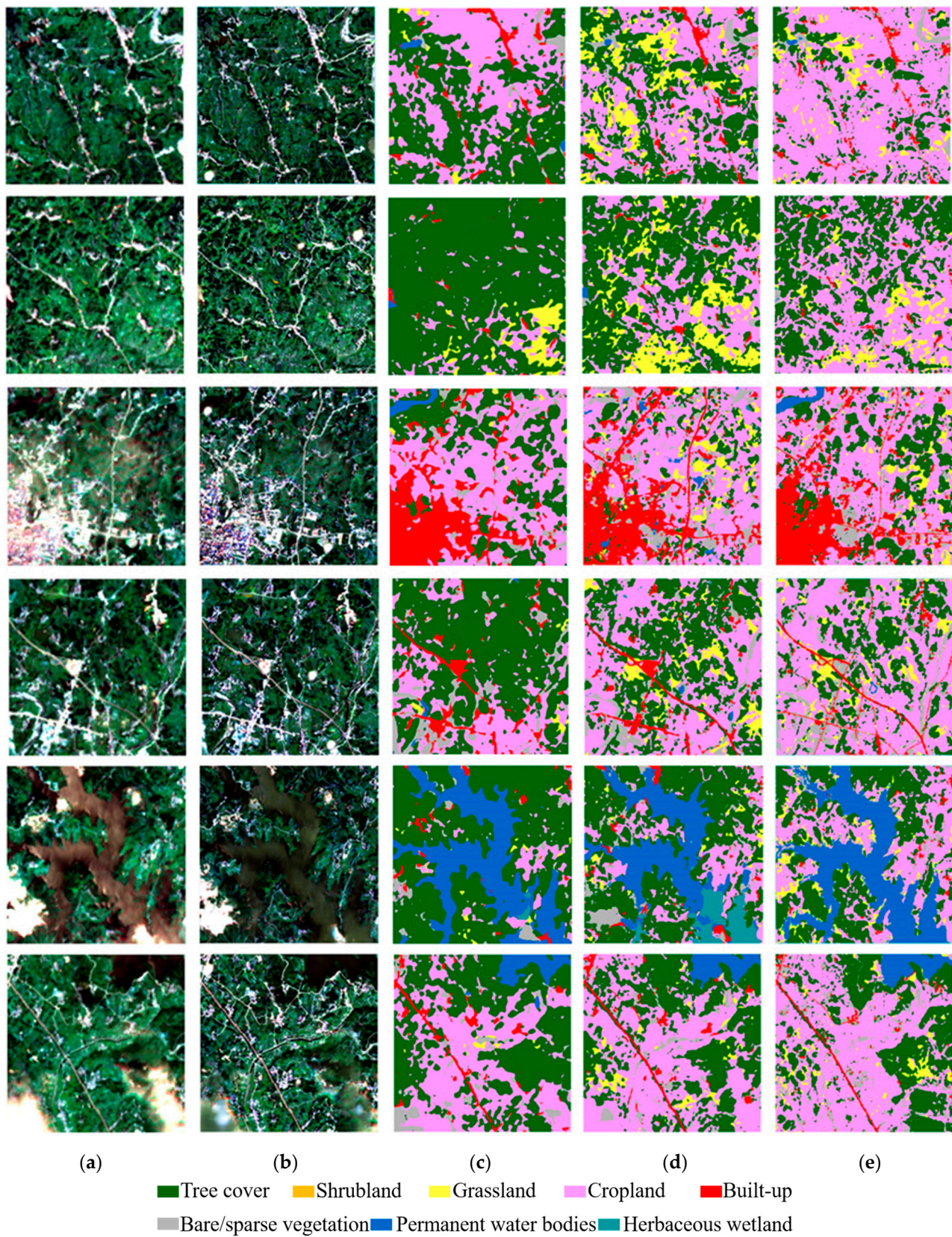


Figure 2. The 10 m land cover results using Landsat-8 images (a) Landsat-8 RGB images; (b) Super-resolution image; (c) ResUnet with up-sampling block (ResUnet_UP); (d) Deeply Fused Super Resolution Guided Semantic Segmentation Network (DFSRSN); and (e) ESA 2020.

From the quantitative accuracy analysis, DFSRSSN can improve *OA* from 77.49% to 83.29% and *Kappa* from 67.60% to 75.92% for dataset I. When the other four datasets were tested, DFSRSSN performed better in *OA* and *Kappa*. In the testing dataset II, there was a gain of 9.03% and 10.34% for *OA* and *Kappa*. For the three remaining testing datasets, significant increases in *Kappa* were observed, i.e., 43.97%, 29.44%, and 18.07%. Therefore, according to the comparison between ResUnet_UP and DFSRSSN in terms of *OA* and *Kappa*, the integration of super-resolution guided information in semantic segmentation networks is effective. However, a severe performance drop for testing datasets II–IV (*OA* ranging from 68.17% to 73.71%) can be observed compared to that of dataset I (*OA* of 83.29%), which is attributed to the fact that datasets II–IV have a quite different distribution from the training data.

Table 4. Overall accuracy (*OA*) (%) and *Kappa* coefficient (%) of ResUnet with up-sampling block (ResUnet_UP) and Deeply Fused Super Resolution Guided Semantic Segmentation Network (DFSRSSN).

Dataset	ResUnet_UP		DFSRSSN	
	<i>OA</i>	<i>Kappa</i>	<i>OA</i>	<i>Kappa</i>
I	77.49	67.60	83.29	75.92
II	59.14	45.74	68.17	56.08
III	39.44	16.05	73.71	60.02
IV	49.37	28.23	72.90	57.67
V	68.03	21.48	70.19	39.55

3.2. Results of Different Classes

The accuracy of different classes (e.g., *PA*, *UA*, and *F1* score values) is shown in Figure 3. The accuracy of the eight land cover types was completely different. In general, across different datasets, tree cover, cropland, built-up, and water had relatively higher *PA*, *UA*, and *F1* score values, while shrubland, grass land, bare/sparse vegetation, and herbaceous wetland need further improvement. The serious misclassification and omission of these four land cover classes is consistent with the conclusions reached by Ding, et al. [41] and Kang, et al. [42], who conducted the accuracy assessment of ESA 2020 in Southeast Asia and Northwestern China. The high label errors of ESA 2020 used for model training and testing can lead to an error accumulation especially for uneasily classified types. Therefore, it is a reasonable performance for our model with pseudo labels.

Compared to ResUnet_UP, the proposed method generally has higher *PA*, *UA*, and *F1* score values for most land cover types, which further demonstrates that the integration of super-resolution related information is effective in improving classification performance. As for the five different testing datasets, *PA*, *UA*, and *F1* score are the highest in dataset I. Datasets II–V with different cities in the training data have lower class-specific accuracy values. However, it should be noted that due to the diverse landscape pattern in different cities and different imaging conditions, spatial transferability of models trained on specific cities to a new city is very challenging [43]. To ensure a fair comparison of spatial transferability between ResUnet_UP and DFSRSSN, we used a relative change rate of corresponding *F1* score values between dataset I and the other datasets (II–V). In addition, the average of relative change rate for different datasets (II–V) was computed to evaluate the overall spatial transferability. Specifically, as for the more easily classified land cover types, i.e., tree cover, cropland, built-up, and water, *F1* scores are 85.54%, 88.27%, 79.66%, and 87.17%, respectively, in dataset I. An average relative drop rate of 6.98%, 19.55%, 19.29%, and 22.42% was observed for the four land cover types for DFSRSSN, while average relative drop rates for ResUnet_UP were 33.84%, 33.94%, 40.09%, and 25.67%. The significant smaller relative drop rate observed for DFSRSSN indicates better spatial transferability of our proposed method.

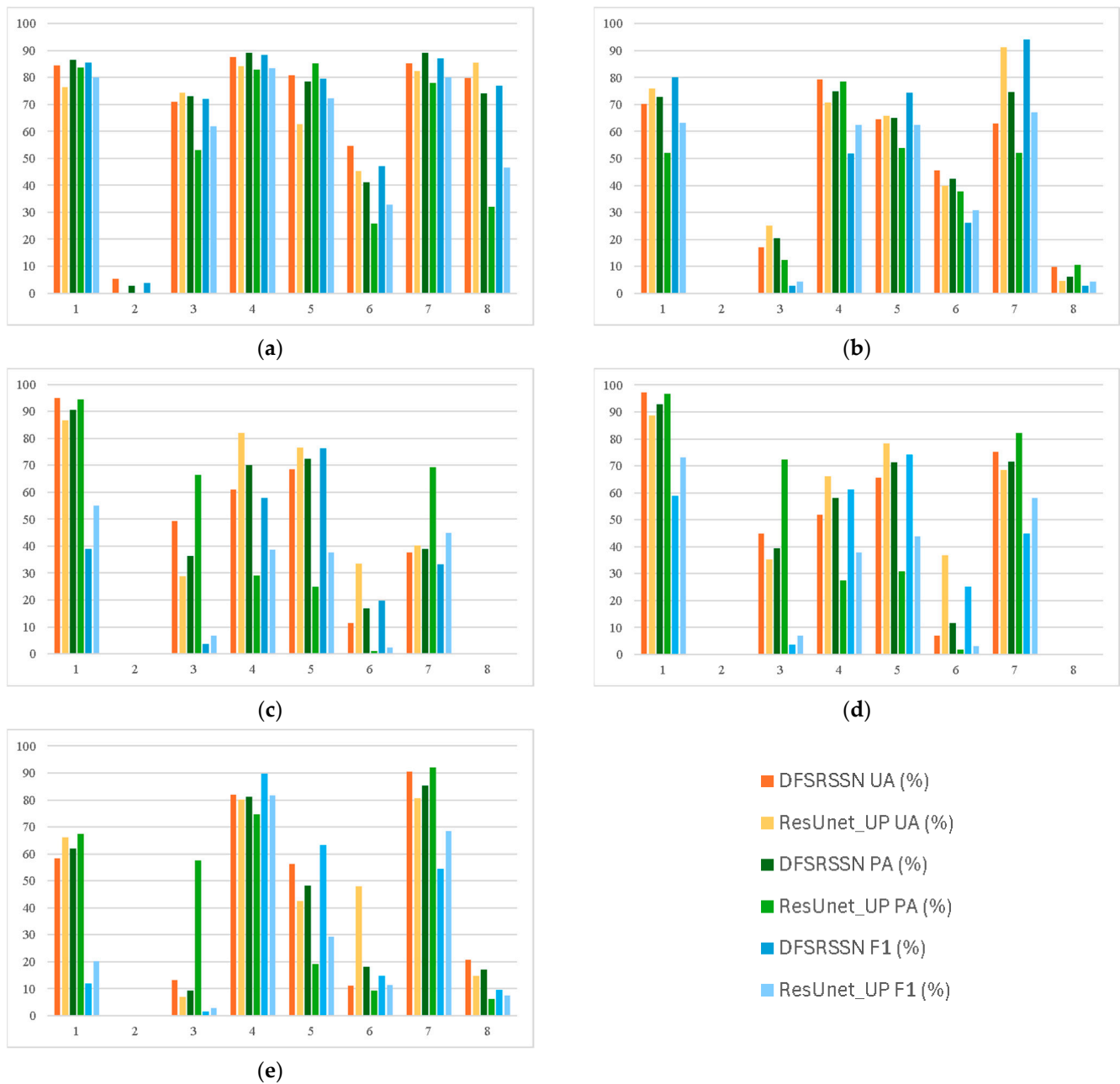


Figure 3. Comparison of user's accuracy (UA), producer's accuracy (PA), and F1 score for the different testing datasets: (a) Dataset I, (b) Dataset II, (c) Dataset III, (d) Dataset IV, and (e) Dataset V. Note: 1 tree cover, 2 shrubland, 3 grassland, 4 cropland, 5 built-up, 6 bare/sparse vegetation, 7 permanent water bodies, 8 herbaceous wetland.

4. Discussion

4.1. Multi-Temporal Results

The motivation of the proposed method was to generate 10 m land cover maps using 30 m Landsat images. This means that we can study 10 m level LULC change from 1984 (when 30 m Landsat was launched) until present day. The models were trained with images of 20 Chinese cities acquired in 2020, data in different cities and different years may significantly differ from the training data. In addition, for past scenarios preceding 2015, i.e., before the launch of 10 m Sentinel, it is of great significance to use the trained model to predict 10 m land cover maps. Therefore, multi-temporal results for Wuhan (not included in the 20 Chinese cities) in 2013 and 2020 were generated, as shown in Figure 4.

In addition, four zoomed regions of multi-temporal Landsat images, multi-temporal land covers, and ESA 2000 are selected to visually investigate the classification details. As shown in Figure 5, relatively obvious boundaries of houses and roads can be extracted by our proposed method. For some misclassification in ESA 2020, such as omission of some built-up area, our method can provide more refined result. When comparing multi-temporal results, small-scale changes such as buildings and road construction, can be captured even with possible errors. Based on the visual inspection, the multi-temporal land cover results extracted by the proposed model are reasonable for finer-grained change analysis.

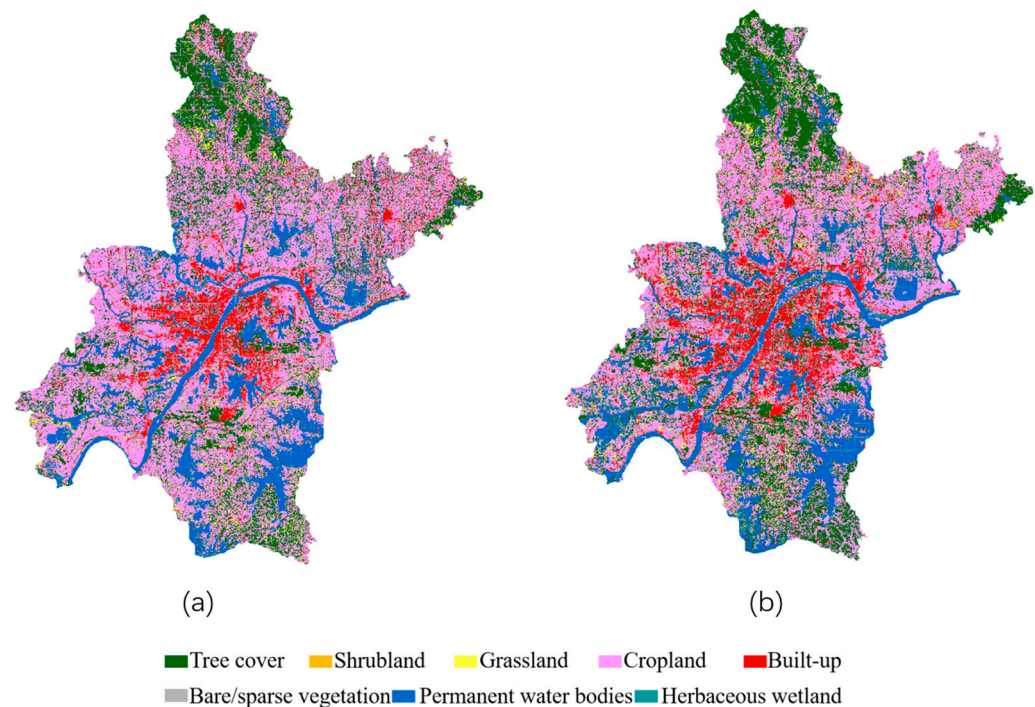


Figure 4. The land use and land cover map of Wuhan in (a) 2013 and (b) 2020.

Similar studies for multi-temporal analysis of Wuhan have been conducted extensively. For example, in [44], LULC change under urbanization in Wuhan, China from 2000 to 2019 based on continuous time series mapping using Landsat observation (30 m) was investigated. It was reported that built-up area increased at the expense of cropland reduction. The area of natural habitat, such as water, forest, and grassland, also declined. Shi et al. [45] focused on detecting the spatial and temporal change of urban lakes in Wuhan using a long time series of Landsat (30 m) and HJ-1A (30 m) remotely sensed data from 1987 to 2016 and revealed that most lakes in Wuhan had shrunk and changed to land. Xin et al. [46] monitored urban expansion using time series of Defense Meteorological Satellite Program's Operational Line-scan System (DMSP/OLS) data (1 km). However, finer-grained change analysis for Wuhan has seldom been conducted. Based on the multi-temporal results for Wuhan in 2013 and 2020, the LULC conversion matrix in the period 2013–2020 is presented in Table 5. It can be observed that 285.78 km² cropland was changed to built-up area due to urbanization. Under the press of urbanization, the areas of conversion from tree cover, grassland, bare/sparse vegetation, permanent water bodies, and herbaceous wetland to built-up area are 98.17 km², 12.98 km², 168.16 km², 18.88 km², and 0.98 km², respectively, which are much smaller than the conversion area from cropland. The major LULC conversion types are “from cropland to tree cover” and “from cropland to bare/sparse vegetation”, i.e., 732.90 km² and 425.93 km², which further cause a cropland reduction.

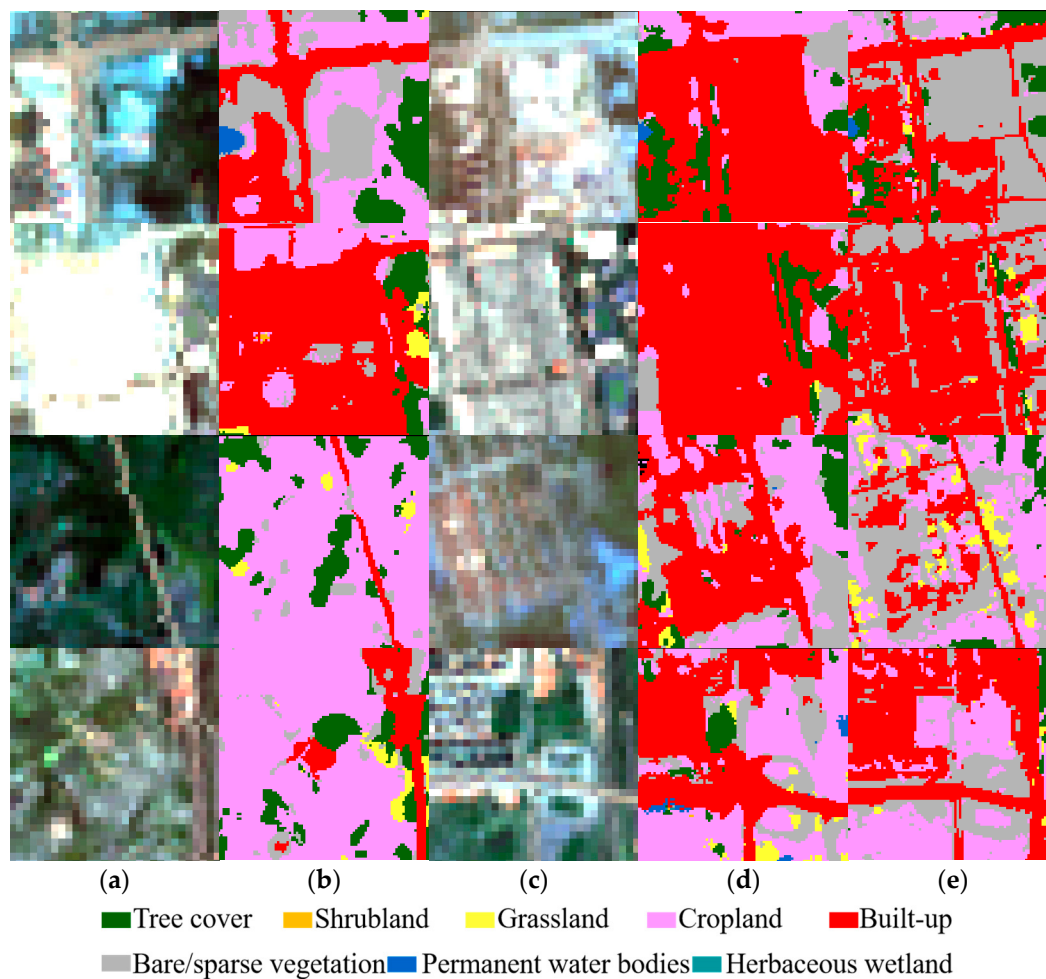


Figure 5. Representative scenes of 10 m multi-temporal results for Wuhan: (a) Landsat image in 2013; (b) Land cover map in 2013; (c) Landsat image in 2020; (d) Land cover map in 2020, and (e) ESA 2020.

Table 5. Land use and land cover conversion matrix of Wuhan in the period 2013–2020 (km²).

Area	1	2	3	4	5	6	7	8
1	872.53	2.04	59.93	512.69	98.17	141.88	59.66	11.13
2	2.20	0.06	0.34	1.59	0.32	0.38	0.72	0.11
3	89.83	0.46	30.00	97.19	12.98	24.70	21.35	3.77
4	732.90	3.19	152.20	1841.91	285.78	425.93	198.90	49.84
5	103.01	0.32	10.15	161.42	369.04	100.33	18.51	2.01
6	148.83	0.61	22.17	304.65	168.16	163.01	68.61	11.68
7	80.85	3.20	16.37	93.15	18.88	40.16	839.80	40.29
8	7.09	0.26	1.23	12.82	0.98	3.40	36.90	6.39

1: tree cover, 2: shrubland, 3: grassland, 4: cropland, 5: built-up, 6: bare/sparse vegetation, 7: permanent water bodies, 8: herbaceous wetland.

4.2. Comparison with Other Models

In order to further evaluate the performance of the proposed method, we conducted a comparative analysis with DFSRSSN which is optimized by multi-task loss (DFSRSSN_MTL) and three state-of-the-art semantic segmentation methods. DFSRSSN_MTL has the same network with DFSRRSN but was optimized in a different way. SR and semantic segmentation (SM) modules were trained simultaneously to minimize the total sum of SR and SM loss. The three well-known semantic segmentation methods involve Multiattention Network (MANet) [47], Pyramid Scene Parsing Network (PSPNet) [48], and DeepLabV3+ [49]. MANet utilizes self-attention modules to aggregate relevant contextual features hierarchi-

cally and a multi-scale strategy is designed to incorporate semantic information at different levels [47]. In PSPNet, the pyramid pooling module is designed to aggregate contextual information of different scales [50]. DeepLabV3+ employs dilated convolution to enlarge the receptive field of filters and introduces feature pyramid network to combine features at different spatial resolutions [51]. All these methods concentrate on extracting multi-scale information for better semantic segmentation performance. And they use bicubic Landsat images of 10 m as input. Overall performance in terms of *OA* and *Kappa* are in presented in Table 6. Bold values indicate the best accuracies, and the second-best ones are underlined. Our proposed model outperforms the three classics, fully segmentation methods that do not produce an SR image for dataset I. DFSRSSN_MTL has *OA* of 83.39% (the second best) and *Kappa* of 75.89% (the best). DFSRSSN achieved *OA* of 83.29% (the best) and *Kappa* of 75.92% (the best). While DFSRSSN_MTL is the best and DFSRSSN is the second best for dataset III (Kassel, Germany), PSPNet is the best for datasets II and IV. As for dataset V, MANet is the best, followed by DeepLabV3+ and DFSRSSN_MTL. Therefore, it is concluded that for overall performance our proposed method can achieve the best for data with similar distribution as training dataset and other more advanced semantic segmentation models have better transferability due to effective multi-scale feature extraction ability. The incorporation of our proposed architecture with these semantic segmentation models can be considered in further work.

Table 6. Overall accuracy (*OA*) (%) and *Kappa* (%) of the proposed and compared methods.

Dataset	DFSRSSN		DFSRSSN_MTL		MANet		PSPNet		DeepLabV3+	
	<i>OA</i>	<i>Kappa</i>	<i>OA</i>	<i>Kappa</i>	<i>OA</i>	<i>Kappa</i>	<i>OA</i>	<i>Kappa</i>	<i>OA</i>	<i>Kappa</i>
I	<u>83.29</u>	75.92	83.39	<u>75.89</u>	81.71	73.18	79.95	70.93	81.16	72.59
II	68.17	56.08	68.62	<u>56.03</u>	<u>68.87</u>	55.93	69.57	56.69	66.34	53.35
III	<u>73.71</u>	<u>60.02</u>	73.94	60.19	71.11	54.90	72.84	57.82	67.57	50.12
IV	<u>72.90</u>	<u>57.67</u>	<u>73.31</u>	<u>58.38</u>	72.57	55.51	76.04	60.86	66.27	45.73
V	70.19	39.55	72.77	39.84	80.65	52.18	70.56	41.41	<u>77.52</u>	<u>49.60</u>

Furthermore, to evaluate the performance of different classes, we present *F1* score values considering that *PA* and *UA* are often contradictory and *F1* score conveys the balance between *PA* and *UA*. *F1* score values of two major LULC types (tree cover and cropland), and built-up areas using different models are presented in Table 7. Compared with other LULC types, built-up areas have more abundant fine spatial details, i.e., scattered roads and isolated buildings, making it more imperative to use higher spatial resolution images to characterize its extent. Therefore, *F1* score values for built-up can reveal whether spatial details are fully exploited and presented in semantic segmentation results. DFSRSSN and DFSRSSN_MTL outperformed other models in tree cover classification for dataset I, III, and IV. As for cropland classification, the best is DFSRSSN_MTL in datasets I and III, PSPNet in datasets II and IV, and MANet in dataset V. Our proposed method generally has much higher *F1* score values for built-up area, for example 53.34% using DFSRSSN_MTL (the best) and 48.35% using DFSRSSN (the second best) vs. 41.75% using PSPNet (the third best).

Table 7. *F1* score values of the proposed and compared methods for tree cover, cropland, and built-up (%).

Dataset	DFSRSSN			DFSRSSN_MTL			MANet			PSPNet			DeepLabV3+		
	1	4	5	1	4	5	1	4	5	1	4	5	1	4	5
I	85.54	88.28	79.66	85.24	88.40	80.83	84.04	86.90	78.38	81.56	85.90	76.40	83.09	86.66	78.90
II	<u>72.94</u>	74.84	65.15	71.85	<u>75.87</u>	66.88	72.11	75.51	64.23	73.10	76.86	66.06	70.48	72.84	<u>66.18</u>
III	90.54	70.01	72.30	89.51	72.34	73.61	88.74	70.40	64.50	87.99	<u>72.28</u>	64.38	87.32	63.52	66.32
IV	92.84	58.15	<u>71.40</u>	<u>92.02</u>	64.72	72.61	91.86	<u>65.12</u>	64.37	90.35	70.33	65.87	89.81	43.85	66.18
V	61.98	81.05	<u>48.35</u>	55.93	83.43	53.34	70.98	89.29	38.12	59.03	81.29	41.75	<u>69.30</u>	<u>87.23</u>	37.81

1 tree cover, 4 cropland, 5 built-up.

5. Conclusions

In this study, we proposed a deeply fused super resolution guided semantic segmentation network for generating high-resolution land cover maps using relatively lower-resolution remote sensing images. We introduced a large-scale and diverse dataset consisting of 10 m Sentinel-2 images, 30 m Landsat-8 images, and 10 m European Space Agency (ESA) Land Cover Product, encompassing 8 land cover categories. Our method integrates a super resolution module (SRResNet) and a semantic segmentation module (CRFFNet) to effectively leverage super resolution guided information for finer-resolution segmentation results. Through extensive experiments and analyses, we demonstrated the efficacy of our proposed method. DFSRSSN outperformed the baseline ResUnet with pixel shuffle blocks in terms of overall accuracy and kappa coefficient, achieving significant improvements in land cover classification accuracy. Furthermore, our method exhibited better spatial transferability across different testing datasets, showcasing its robustness in handling diverse landscape patterns and imaging conditions. Moreover, the multi-temporal results demonstrated the spatiotemporal transferability of our proposed method, enabling the generation of 10 m land cover maps for past scenarios preceding the availability of 10 m Sentinel data. This capability opens up avenues for studying long-term LULC changes at a finer spatial resolution. In addition, the comparison of the state-of-the-art full semantic segmentation models indicate that spatial details are fully exploited and presented in semantic segmentation results using super-resolution guided information. In conclusion, our study contributes to the advancement of high-resolution land cover mapping using remote sensing data, offering a valuable tool for urban planning, ecological evaluation, climate change analysis, agricultural management, and policy making. The integration of super resolution guided information into semantic segmentation networks represents a promising approach for enhancing the spatial resolution and accuracy of land cover mapping, thereby facilitating informed decision-making and sustainable development initiatives. Nevertheless, it is imperative to acknowledge certain limitations of our methodology. Label errors in the ESA Land Cover Product may cause overfitting to mislabeled data. Therefore, pseudo label refinement can be considered in a further study.

Author Contributions: Conceptualization, D.W. and S.Z.; methodology, D.W., S.Z. and Y.L.; software, D.W. and S.Z.; validation, D.W., Y.L. and S.Z.; formal analysis, D.W. and S.Z.; investigation, D.W. and S.Z.; resources, S.Z.; data curation, S.Z.; writing—original draft preparation, D.W. and S.Z.; writing—review and editing, D.W., Y.L. and X.G.; visualization, S.Z.; supervision, D.W. and Y.L.; project administration, D.W., Y.T. and X.G.; funding acquisition, D.W. and S.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant 41901279, in part by the Science Foundation Research Project of Wuhan Institute of Technology of China under Grant K202239, and in part by the Graduate Innovative Fund of Wuhan Institute of Technology under Grant CX2023312.

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

Acknowledgments: The authors are grateful for the constructive comments from the anonymous reviewers and the editors.

Conflicts of Interest: Author Xuehua Guan was employed by the company China Siwei Surveying and Mapping Technology Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Mohamed, A.; Worku, H. Quantification of the land use/land cover dynamics and the degree of urban growth goodness for sustainable urban land use planning in Addis Ababa and the surrounding Oromia special zone. *J. Urban Manag.* **2019**, *8*, 145–158. [[CrossRef](#)]

2. Wen, D.; Ma, S.; Zhang, A.; Ke, X. Spatial Pattern Analysis of the Ecosystem Services in the Guangdong-Hong Kong-Macao Greater Bay Area Using Sentinel-1 and Sentinel-2 Imagery Based on Deep Learning Method. *Sustainability* **2021**, *13*, 7044. [\[CrossRef\]](#)
3. Naserikia, M.; Hart, M.A.; Nazarian, N.; Bechtel, B. Background climate modulates the impact of land cover on urban surface temperature. *Sci. Rep.* **2022**, *12*, 15433. [\[CrossRef\]](#)
4. Lu, J.; Fu, H.; Tang, X.; Liu, Z.; Huang, J.; Zou, W.; Chen, H.; Sun, Y.; Ning, X.; Li, J. GOA-optimized deep learning for soybean yield estimation using multi-source remote sensing data. *Sci. Rep.* **2024**, *14*, 7097. [\[CrossRef\]](#)
5. Chamling, M.; Bera, B. Spatio-temporal Patterns of Land Use/Land Cover Change in the Bhutan–Bengal Foothill Region Between 1987 and 2019: Study Towards Geospatial Applications and Policy Making. *Earth Syst. Environ.* **2020**, *4*, 117–130. [\[CrossRef\]](#)
6. Venter, Z.S.; Sydenham, M.A.K. Continental-Scale Land Cover Mapping at 10 m Resolution Over Europe (ELC10). *Remote Sens.* **2021**, *13*, 2301. [\[CrossRef\]](#)
7. Sulla-Menashe, D.; Gray, J.M.; Abercrombie, S.P.; Friedl, M.A. Hierarchical mapping of annual global land cover 2001 to present: The MODIS Collection 6 Land Cover product. *Remote Sens. Environ.* **2019**, *222*, 183–194. [\[CrossRef\]](#)
8. Friedl, M.; Sulla-Menashe, D. MCD12Q1 MODIS/Terra+ aqua land cover type yearly L3 global 500m SIN grid V006. *NASA EOSDIS Land Process. DAAC* **2019**, *10*, 200.
9. Defourny, P.; Kirches, G.; Brockmann, C.; Boettcher, M.; Peters, M.; Bontemps, S.; Lamarche, C.; Schlerf, M.; Santoro, M. Land cover CCI. In *Product User Guide Version*; ESA Climate Change Initiative: Paris, France, 2012; Volume 2, 10.1016.
10. Defourny, P.; Lamarche, C.; Marissiaux, Q.; Carsten, B.; Martin, B.; Grit, K. *Product User Guide Specification: ICDR Land Cover 2016–2020*; ECMWF: Reading, UK, 2021; 37p.
11. Gong, P.; Wang, J.; Yu, L.; Zhao, Y.; Zhao, Y.; Liang, L.; Niu, Z.; Huang, X.; Fu, H.; Liu, S.; et al. Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data. *Int. J. Remote Sens.* **2013**, *34*, 2607–2654. [\[CrossRef\]](#)
12. Chen, B.; Xu, B.; Zhu, Z.; Yuan, C.; Suen, H.P.; Guo, J.; Xu, N.; Li, W.; Zhao, Y.; Yang, J. 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*, 370–373.
13. Zanaga, D.; Van De Kerchove, R.; De Keersmaecker, W.; Souverijns, N.; Brockmann, C.; Quast, R.; Wevers, J.; Grosu, A.; Paccini, A.; Vergnaud, S.; et al. ESA WorldCover 10 m 2020 V100. *OpenAIRE* **2021**.
14. 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. [\[CrossRef\]](#)
15. Xu, P.; Tang, H.; Ge, J.; Feng, L. ESPC_NASUnet: An End-to-End Super-Resolution Semantic Segmentation Network for Mapping Buildings from Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 5421–5435. [\[CrossRef\]](#)
16. Cui, B.; Zhang, H.; Jing, W.; Liu, H.; Cui, J. SRSe-Net: Super-Resolution-Based Semantic Segmentation Network for Green Tide Extraction. *Remote Sens.* **2022**, *14*, 710. [\[CrossRef\]](#)
17. Fu, B.; Sun, X.; Li, Y.; Lao, Z.; Deng, T.; He, H.; Sun, W.; Zhou, G. Combination of super-resolution reconstruction and SGA-Net for marsh vegetation mapping using multi-resolution multispectral and hyperspectral images. *Int. J. Digit. Earth* **2023**, *16*, 2724–2761. [\[CrossRef\]](#)
18. Zhu, Y.; Geiß, C.; So, E. Image super-resolution with dense-sampling residual channel-spatial attention networks for multi-temporal remote sensing image classification. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *104*, 102543. [\[CrossRef\]](#)
19. Huang, Y.; Wen, X.; Gao, Y.; Zhang, Y.; Lin, G. Tree Species Classification in UAV Remote Sensing Images Based on Super-Resolution Reconstruction and Deep Learning. *Remote Sens.* **2023**, *15*, 2942. [\[CrossRef\]](#)
20. Liang, M.; Wang, X. A bidirectional semantic segmentation method for remote sensing image based on super-resolution and domain adaptation. *Int. J. Remote Sens.* **2023**, *44*, 666–689. [\[CrossRef\]](#)
21. Zhang, Q.; Yang, G.; Zhang, G. Collaborative Network for Super-Resolution and Semantic Segmentation of Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–12. [\[CrossRef\]](#)
22. Cai, Y.; Yang, Y.; Shang, Y.; Shen, Z.; Yin, J. DASRSNet: Multitask Domain Adaptation for Super-Resolution-Aided Semantic Segmentation of Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–18. [\[CrossRef\]](#)
23. Salgueiro, L.; Marcello, J.; Vilaplana, V. SEG-ESRGAN: A Multi-Task Network for Super-Resolution and Semantic Segmentation of Remote Sensing Images. *Remote Sens.* **2022**, *14*, 5862. [\[CrossRef\]](#)
24. Zhang, J.; Xu, T.; Li, J.; Jiang, S.; Zhang, Y. Single-Image Super Resolution of Remote Sensing Images with Real-World Degradation Modeling. *Remote Sens.* **2022**, *14*, 2895. [\[CrossRef\]](#)
25. Liebel, L.; Körner, M. Auxiliary tasks in multi-task learning. *arXiv* **2018**, arXiv:1805.06334.
26. Xie, J.; Fang, L.; Zhang, B.; Chanussot, J.; Li, S. Super Resolution Guided Deep Network for Land Cover Classification from Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–12. [\[CrossRef\]](#)
27. Li, J.; Zhang, Z.; Tian, Y.; Xu, Y.; Wen, Y.; Wang, S. Target-Guided Feature Super-Resolution for Vehicle Detection in Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [\[CrossRef\]](#)
28. Chaaban, F.; El Khattabi, J.; Darwishe, H. Accuracy Assessment of ESA WorldCover 2020 and ESRI 2020 Land Cover Maps for a Region in Syria. *J. Geovis. Spat. Anal.* **2022**, *6*, 31. [\[CrossRef\]](#)
29. Luo, M.; Ji, S.; Wei, S. A Diverse Large-Scale Building Dataset and a Novel Plug-and-Play Domain Generalization Method for Building Extraction. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 4122–4138. [\[CrossRef\]](#)

30. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 105–114.
31. Sdraka, M.; Papoutsis, I.; Psomas, B.; Vlachos, K.; Ioannidis, K.; Karantzas, K.; Gialampoukidis, I.; Vrochidis, S. Deep Learning for Downscaling Remote Sensing Images: Fusion and super-resolution. *IEEE Geosci. Remote Sens. Mag.* **2022**, *10*, 202–255. [\[CrossRef\]](#)
32. Razzak, M.T.; Mateo-García, G.; Lecuyer, G.; Gómez-Chova, L.; Gal, Y.; Kalaitzis, F. Multi-spectral multi-image super-resolution of Sentinel-2 with radiometric consistency losses and its effect on building delineation. *ISPRS J. Photogramm. Remote Sens.* **2023**, *195*, 1–13. [\[CrossRef\]](#)
33. Feng, C.-M.; Wang, K.; Lu, S.; Xu, Y.; Li, X. Brain MRI super-resolution using coupled-projection residual network. *Neurocomputing* **2021**, *456*, 190–199. [\[CrossRef\]](#)
34. Zhang, Z.; Liu, Q.; Wang, Y. Road Extraction by Deep Residual U-Net. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 749–753. [\[CrossRef\]](#)
35. Congalton, R.G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* **1991**, *37*, 35–46. [\[CrossRef\]](#)
36. Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas.* **1960**, *20*, 37–46. [\[CrossRef\]](#)
37. Liu, C.; Frazier, P.; Kumar, L. Comparative assessment of the measures of thematic classification accuracy. *Remote Sens. Environ.* **2007**, *107*, 606–616. [\[CrossRef\]](#)
38. Christen, P.; Hand, D.J.; Kirielle, N. A Review of the F-Measure: Its History, Properties, Criticism, and Alternatives. *ACM Comput. Surv.* **2023**, *56*, 73. [\[CrossRef\]](#)
39. Zhang, Z.; Xu, N.; Li, Y.; Li, Y. Sub-continental-scale mapping of tidal wetland composition for East Asia: A novel algorithm integrating satellite tide-level and phenological features. *Remote Sens. Environ.* **2022**, *269*, 112799. [\[CrossRef\]](#)
40. Bargiel, D. A new method for crop classification combining time series of radar images and crop phenology information. *Remote Sens. Environ.* **2017**, *198*, 369–383. [\[CrossRef\]](#)
41. Ding, Y.; Yang, X.; Wang, Z.; Fu, D.; Li, H.; Meng, D.; Zeng, X.; Zhang, J. A Field-Data-Aided Comparison of Three 10 m Land Cover Products in Southeast Asia. *Remote Sens.* **2022**, *14*, 5053. [\[CrossRef\]](#)
42. Kang, J.; Yang, X.; Wang, Z.; Cheng, H.; Wang, J.; Tang, H.; Li, Y.; Bian, Z.; Bai, Z. Comparison of Three Ten Meter Land Cover Products in a Drought Region: A Case Study in Northwestern China. *Land* **2022**, *11*, 427. [\[CrossRef\]](#)
43. Cao, Y.; Huang, X. A deep learning method for building height estimation using high-resolution multi-view imagery over urban areas: A case study of 42 Chinese cities. *Remote Sens. Environ.* **2021**, *264*, 112590. [\[CrossRef\]](#)
44. Zhai, H.; Lv, C.; Liu, W.; Yang, C.; Fan, D.; Wang, Z.; Guan, Q. Understanding Spatio-Temporal Patterns of Land Use/Land Cover Change under Urbanization in Wuhan, China, 2000–2019. *Remote Sens.* **2021**, *13*, 3331. [\[CrossRef\]](#)
45. Shi, L.; Ling, F.; Foody, G.M.; Chen, C.; Fang, S.; Li, X.; Zhang, Y.; Du, Y. Permanent disappearance and seasonal fluctuation of urban lake area in Wuhan, China monitored with long time series remotely sensed images from 1987 to 2016. *Int. J. Remote Sens.* **2019**, *40*, 8484–8505. [\[CrossRef\]](#)
46. Xin, X.; Liu, B.; Di, K.; Zhu, Z.; Zhao, Z.; Liu, J.; Yue, Z.; Zhang, G. Monitoring urban expansion using time series of night-time light data: A case study in Wuhan, China. *Int. J. Remote Sens.* **2017**, *38*, 6110–6128. [\[CrossRef\]](#)
47. Li, R.; Zheng, S.; Zhang, C.; Duan, C.; Su, J.; Wang, L.; Atkinson, P.M. Multiattention Network for Semantic Segmentation of Fine-Resolution Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–13. [\[CrossRef\]](#)
48. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6230–6239.
49. 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.
50. Yuan, W.; Wang, J.; Xu, W. Shift Pooling PSPNet: Rethinking PSPNet for Building Extraction in Remote Sensing Images from Entire Local Feature Pooling. *Remote Sens.* **2022**, *14*, 4889. [\[CrossRef\]](#)
51. Guo, S.; Yang, Q.; Xiang, S.; Wang, S.; Wang, X. Mask2Former with Improved Query for Semantic Segmentation in Remote-Sensing Images. *Mathematics* **2024**, *12*, 765. [\[CrossRef\]](#)

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.