remote sensing

MDPI

Article

Integration of Convolutional Neural Networks and Object-Based Post-Classification Refinement for Land Use and Land Cover Mapping with Optical and SAR Data

Shengjie Liu [1], Zhixin Qi [1,*], Xia Li [2] and Anthony Gar-On Yeh [3]

[1] Guangdong Provincial Key Laboratory of Urbanization and Geo-simulation, School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China; liushj23@mail2.sysu.edu.cn

[2] School of Geographic Sciences, Key Lab. of Geographic Information Science (Ministry of Education), East China Normal University, 500 Dongchuan Rd, Shanghai 200241, China; lixia@geo.ecnu.edu.cn or lixia@mail.sysu.edu.cn

[3] Department of Urban Planning and Design, The University of Hong Kong, Pokfulam Road, Hong Kong, China; hdxugoy@hkucc.hku.hk

* Correspondence: qizhixin@mail.sysu.edu.cn

Received: 15 February 2019; Accepted: 16 March 2019; Published: 22 March 2019

Abstract: Object-based image analysis (OBIA) has been widely used for land use and land cover (LULC) mapping using optical and synthetic aperture radar (SAR) images because it can utilize spatial information, reduce the effect of salt and pepper, and delineate LULC boundaries. With recent advances in machine learning, convolutional neural networks (CNNs) have become state-of-the-art algorithms. However, CNNs cannot be easily integrated with OBIA because the processing unit of CNNs is a rectangular image, whereas that of OBIA is an irregular image object. To obtain object-based thematic maps, this study developed a new method that integrates object-based post-classification refinement (OBPR) and CNNs for LULC mapping using Sentinel optical and SAR data. After producing the classification map by CNN, each image object was labeled with the most frequent land cover category of its pixels. The proposed method was tested on the optical-SAR Sentinel Guangzhou dataset with 10 m spatial resolution, the optical-SAR Zhuhai-Macau local climate zones (LCZ) dataset with 100 m spatial resolution, and a hyperspectral benchmark the University of Pavia with 1.3 m spatial resolution. It outperformed OBIA support vector machine (SVM) and random forest (RF). SVM and RF could benefit more from the combined use of optical and SAR data compared with CNN, whereas spatial information learned by CNN was very effective for classification. With the ability to extract spatial features and maintain object boundaries, the proposed method considerably improved the classification accuracy of urban ground targets. It achieved overall accuracy (OA) of 95.33% for the Sentinel Guangzhou dataset, OA of 77.64% for the Zhuhai-Macau LCZ dataset, and OA of 95.70% for the University of Pavia dataset with only 10 labeled samples per class.

Keywords: object-based post-classification refinement (OBPR); convolutional neural network (CNN); synthetic aperture radar (SAR); land use and land cover; object-based image analysis (OBIA)

1. Introduction

Land use and land cover (LULC) information is essential for forest monitoring, climate change studies, and environmental and urban management [1–4]. Remote sensing techniques are widely used for LULC investigation because of their capability to observe land surfaces routinely on a large scale. The most often used remotely sensed data are optical images, such as those from Landsat [5–7].

Synthetic aperture radar (SAR) images are also used for LULC classification because of their weather independence [8–12]. Unlike optical data, which contain spectral information, SAR data characterize the structural and dielectric properties of ground targets [13]. Combination of optical and SAR data results in a comprehensive observation of ground targets, and therefore, has a great potential to improve the accuracy of LULC classification [14].

The potential of the combination of optical and SAR data has been increasingly explored for LULC classification, especially after the Sentinel mission was initiated by the European Space Agency (ESA) for Earth observation, which provides free-of-charge optical and SAR data [15,16]. Thanks to the weather independence of radar remote sensing, Reiche et al. [17] improved forest mapping in a tropical region with heavy cloud coverage by fusion of optical and time series SAR imagery. Kussul et al. [18] applied the multi-layer perceptron (MLP) classifier for crop mapping in Ukraine and achieved accuracies of over 90% for major crop types using multitemporal optical and SAR data. Zhang et al. [19] reduced classification confusions among impervious surface, bare soil, shaded area, and water with fusion of optical and SAR images using RF classifier. Zhang and Xu [20] concluded that fusion of optical and SAR data for LULC mapping may be classifier-dependent. They found that SVM and RF had a better performance than maximum likelihood classifier and artificial neural network when using multisource data.

Object-based image analysis (OBIA), together with advanced machine learning algorithms, has been widely used for LULC classification, as it can delineate object boundaries and produce compact classification maps [21]. The spatial, textural, and contextual features extracted by OBIA have shown a great ability to boost the classification performance. For example, Wang et al. [22] produced a global map of build-up area with hierarchical object-based GLCM textures derived from Landsat images and showed a 2.8% improvement on OA compared with that from only spectral bands. Ruiz Hernandez and Shi [23] applied both GLCM texture metrics and spatial indices in a geographic OBIA framework with RF for urban land use mapping. Recently, Franklin and Ahmed [24] applied RF and object-based analysis for tree species classification on multispectral images captured by unmanned aerial vehicles (UAVs). Many studies have concluded that object-based spatial and textural features can significantly improve the classification [25–30].

Pixel-based spatial and textural features, rather than object-based textures, have been concluded very useful for LULC mapping as well. Huang et al. [31] found that pixel-based morphological profiles significantly outperformed object-based GLCM textures for forest mapping and species classification. Wang et al. [32] tested the Completed Local Binary Patterns (CLBP) textures originally designed for face recognition and found that the textures were suitable for classifying wetland vegetation using SVM. With recent advances in machine learning, deep learning models (e.g., CNNs) have achieved great success in computer vision and pattern recognition. Like OBIA, CNNs learn spatial, textural, and contextual information from images, which have been concluded very useful for LULC mapping [33–39]. Therefore, the integration of deep learning and OBIA is worth exploring.

Deep learning for remote sensing image classification can be classified into two categories [40]. One is the conventional LULC classification, in which we obtain a single satellite image and then randomly collect some labeled samples on it. The other is semantic segmentation, in which we collect a set of fully annotated images from the same sensor and then train a CNN to classify new images without any annotation. Semantic segmentation is based on a special kind of CNNs, the fully convolutional networks (FCNs) [41].

As FCNs do not require any annotation in the predicted image, they are extremely suitable for large-scale LULC classification [35,42–46]. Maggiori et al. [35] proposed a CNN with a fully convolutional architecture using a two step pre-training method to produce large-scale building maps. Kampffmeyer et al. [43] measured the uncertainty of FCNs and applied the median frequency balancing to adjust FCNs for imbalance classes and improved segmentation results of small objects. Yu et al. [45] proposed a FCN with the pyramid pooling module to capture features at multiple scales, and thus achieved accurate segmentation for multiple ground objects. As semantic segmentation often leads to blurry LULC boundaries, Marmanis et al. [46] designed a deep CNN that combined boundary detection and segmentation networks to delineate object boundaries.

FCNs have presented a great potential for large-scale LULC classification, but they highly depend on annotated images. For this reason, previous studies are often limited to some high-resolution benchmarks with only RGB channels, such as the ISPRS Vaihingen and Potsdam datasets. When the annotated maps are unavailable, for example, using Sentinel data for local climate zones (LCZs) classification [2,47,48], it is really difficult to use these semantic segmentation models. One exception is from the study of Liu et al. [49]. They successfully applied FCNs using training samples from a single image, where a training sample was the minimum bounding box of each object. However, pixels in the area outside the object and inside the bounding box must be manually labeled to boost the classification performance. Thus, patch-based CNNs are more suitable for LULC classification with single or few images than FCNs.

CNNs are originally designed for image recognition, and the input shall be a rectangular image [49–51]. Zhao et al. [52] applied a five-layer CNN to extract spatial features within an 18 × 18 window. Then, these features were combined with OBIA to produce the classification map based on the tan*h* classifier. Zhang et al. [50] carefully designed a novel object-based CNN to locate the convolutional center of an image object using the minimum bounding box. In their study, each image object was represented by a 128 × 128 image patch. However, objects delineated from satellite images vary widely in size. A large object results in a large minimum bounding box. A large-scale fixed representation may fail to capture small ground targets. For example, bridges on the water are very slender and a large part of the background is water, which might mislead a CNN to classify such image patch as water.

In this study, we present a novel yet simple method, namely object-based post-classification refinement (OBPR), to obtain object-based thematic maps produced by CNN using Sentinel multispectral and SAR data with very small input patches (e.g., 5 × 5). By using small input patches, small ground targets (e.g., high-rise buildings and roads) can be effectively captured. By post-classification processing, the classification maps are refined by object boundaries using majority voting. The proposed method was evaluated on two optical-SAR datasets and one hyperspectral dataset with diverse spatial resolutions. The three datasets are the Sentinel Guangzhou dataset with 10 m spatial resolution, the Zhuhai-Macau LCZ dataset with 100 m spatial resolution, and the University of Pavia dataset with 1.3 m spatial resolution. The remainder of this paper is organized as follows. Section 2 introduces the study area and the datasets. Section 3 explains the methodology, including details of CNNs and the proposed OBPR. Section 4 presents the results and discussion. Conclusions are drawn in Section 5.

2. Study Area and Data

2.1. The Optical-SAR Sentinel Guangzhou Dataset

The first dataset is the optical-SAR Sentinel Guangzhou dataset (available on Google Drive https://drive.google.com/open?id=1NoCHjqRmiYV1lijoHYFvWaqHxKCVrv8X). The study area is in the districts of Panyu and Haizhu in Guangzhou (Figure 1a), which is the center city of the Pearl River Delta in China. This study area features urban and country landscapes that include a variety of LULC categories. Therefore, it is an ideal site for testing the proposed method using optical and SAR data.

The optical Level-1C data were acquired on 1 November 2017 by Sentinel-2A (Figure 1b). The data consist of 13 spectral bands, including four bands with 10 m spatial resolution, six bands with 20 m spatial resolution, and three bands with 60 m spatial resolution. Those with 60 m spatial resolution were discarded in the study because they are not designed for land cover classification [16]. The Sentinel data were downloaded from the Open Access Hub of ESA (https://scihub.copernicus.eu/dhus/#/home). The detail spectral and spatial information of these spectral bands are shown in Table 1. As the input image of CNNs should have the same size, the 20 m resolution bands were resampled to 10 m resolution ones using the nearest neighbor interpolation method embedded in SNAP 5.0.

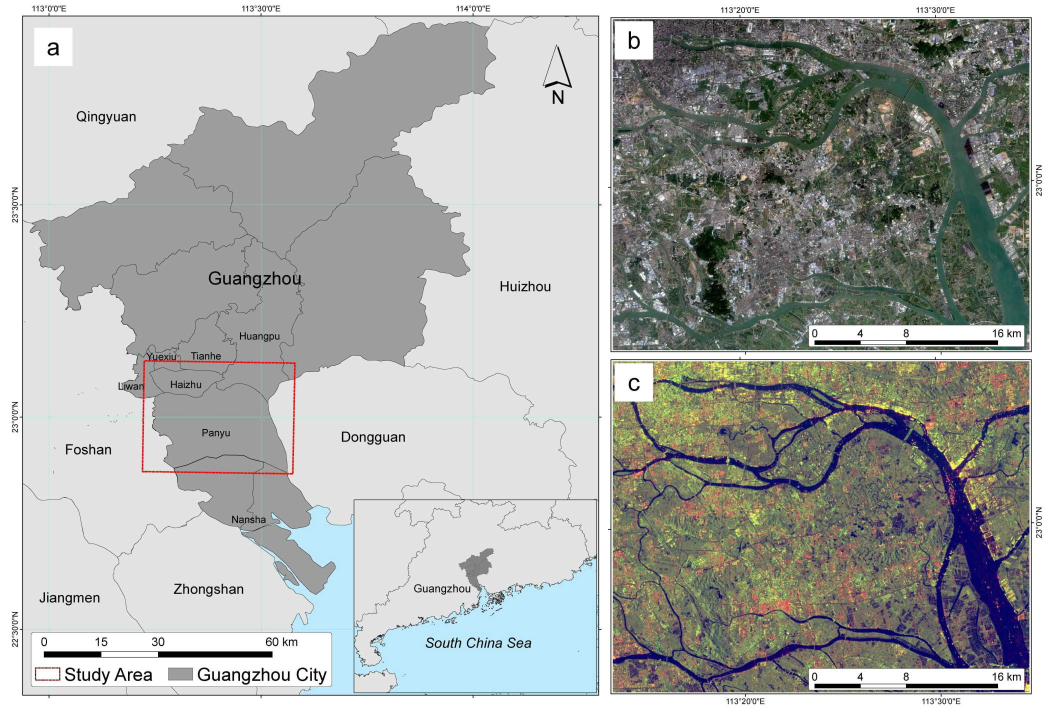


Figure 1. Study area and data. (**a**) Study area; (**b**) Sentinel-2A optical image (true color composition); and (**c**) Sentinel-1A SAR image (red: VV, green: VH, blue: VV/VH).

Table 1. Sentinel-1A SAR data and Sentinel-2A optical data used in this study.

Data Type	Image Channel	Wavelength	Spatial Resolution
Optical	Band 2	490 nm	10 m
	Band 3	560 nm	
	Band 4	665 nm	
	Band 8	842 nm	
	Band 5	705 nm	20 m
	Band 6	740 nm	
	Band 7	783 nm	
	Band 8a	865 nm	
	Band 11	1610 nm	
	Band 12	2190 nm	
SAR	VV	5.6 cm (C-Band)	20 m × 22 m
	VH	5.6 cm (C-Band)	

The C-Band SAR data were acquired by Sentinel-1A on 7 November 2017 (Figure 1c). The data are Level-1 interferometric wide-mode and Ground-Range-Detected High-resolution (GRDH) products in VV and VH polarizations. SNAP 5.0 was employed for SAR preprocessing. After radiometric calibration, the Lee sigma filter [53] with a 7 × 7 window and a 3 × 3 target window was implemented on the SAR data to suppress the speckle noise. The output data were geometrically corrected using Range-Doppler Terrain Correction embedded in SNAP 5.0 and then converted into decibel format (logarithmic scale) for classification. The optical and SAR data were clipped to a 3640 × 2890 size with a pixel size of 10 × 10 m.

The LULC types of the Sentinel Guangzhou dataset were categorized into 11 classes, namely, new town (NT), old town (OT), bare land (BL), port areas (P), green houses (GH), lawn (L), industry and commerce (IC), roads (R), villas (V), water (W), and trees (T). Old town included typically historic downtowns, urban villages, and villages in the suburbs. Industry and commerce were mostly large-area man-made buildings with high albedo, such as factories, conference centers, and high-speed railway stations. The samples were collected randomly through a visual interpretation of the high-resolution satellite images provided by Google Maps. The characteristics of these LULC classes in the high-resolution images are shown in Figure 2. To test the robustness of the proposed method, we constructed two subsets of training samples (50 and 10 object samples per class) randomly. The detailed numbers of the training and test samples are shown in Table 2.

Table 2. Numbers of training and test samples selected for each LULC class.

Class	Training Samples			Test Samples
	150 Objects	50 Objects	10 Objects	150 Objects
New town	4203	1512	288	4357
Old town	8869	2993	880	9394
Bare land	5312	1959	389	5250
Port area	8078	3102	582	6880
Green houses	12,321	3946	1162	11,098
Lawn	8184	2835	566	10,119
Industry and commerce	3362	1044	163	2911
Roads	8482	3265	436	8795
Villas	7513	2545	466	6189
Water	24,760	7844	2395	30,742
Trees	13,008	4753	1112	12,990

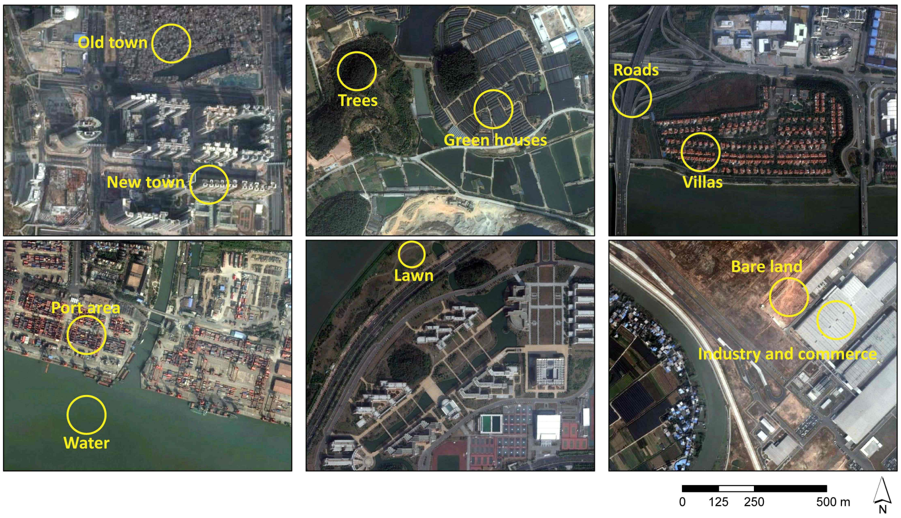


Figure 2. Typical LULC categories in the Guangzhou dataset.

2.2. *The Optical-SAR Zhuhai-Macau LCZ Dataset*

The second dataset is the optical-SAR Zhuhai-Macau LCZ dataset. The concept of LCZ is originally developed by Stewart and Oke [2] for urban heat island studies and now has attracted great interests in the remote sensing community, as it provides a standard classification system for urban land use mapping. For example, the 2017 IEEE GRSS Data Fusion Contest [48] was a task to perform classification of LCZs in nine cities worldwide with various urban environment. Under this context, LULC is categorized into 17 LCZs based on surface cover, structure, material, and human activity [2]. An ongoing project, the world urban database and access portal tools (WUDAPT) [54], is aimed to gather such climate relevant surface information using freely remotely sensed data (i.e., Landsat and Sentinel).

Based on the WUDAPT project, we collected a pair of Sentinel multispectral and SAR images to create the Zhuhai-Macau LCZ dataset. The multispectral imagery with zero cloud coverage (Figure 3) was collected on 21 March 2018, and the SAR imagery was collected on 19 March 2018. The study area fully covered the cities of Zhuhai and Macau and a small part of the neighboring cities (Zhongshan, Jiangmen, Shenzhen, and Hong Kong). After preprocessing of SAR image and registration, the images were resampled to 100 m spatial resolution using the nearest neighbor method. The study area with a true color composite is shown in Figure 3.

The reference data were collected on Google Earth and some of the reference data that were difficult to distinguish were checked in fields. The numbers of samples are shown in Table 3, and the samples captured on high spatial resolution satellite images are shown in Figure 4. The LCZ-7 class in this dataset was mainly green houses; the LCZ-C class is not taken into account as the study area is in the subtropics, resulting in a total of 16 classes. The dataset contains very limited labeled samples with a highly complex classification system, resulting a extremely difficult classification task. The size of the image is 1098 × 1098 with 12 channels.

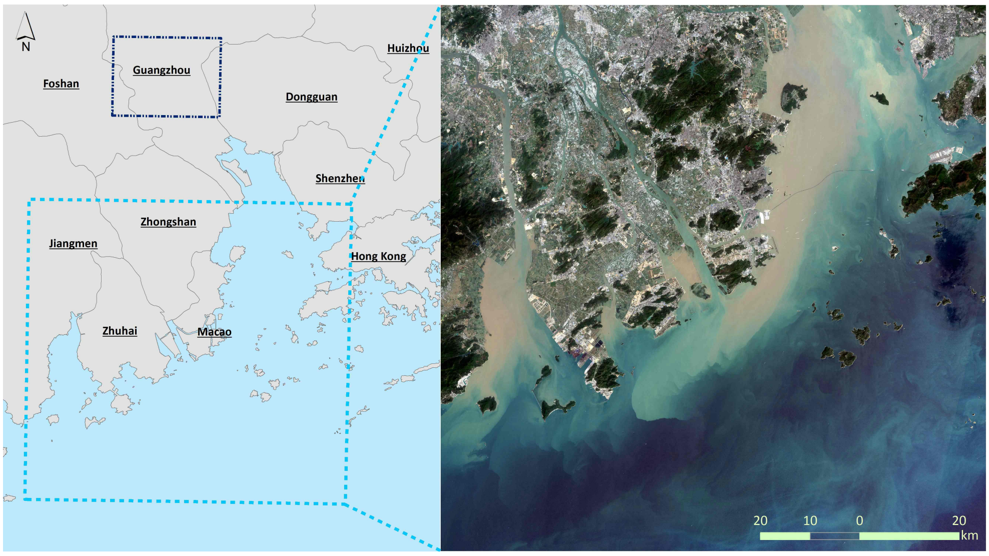


Figure 3. The Zhuhai-Macau LCZ dataset.

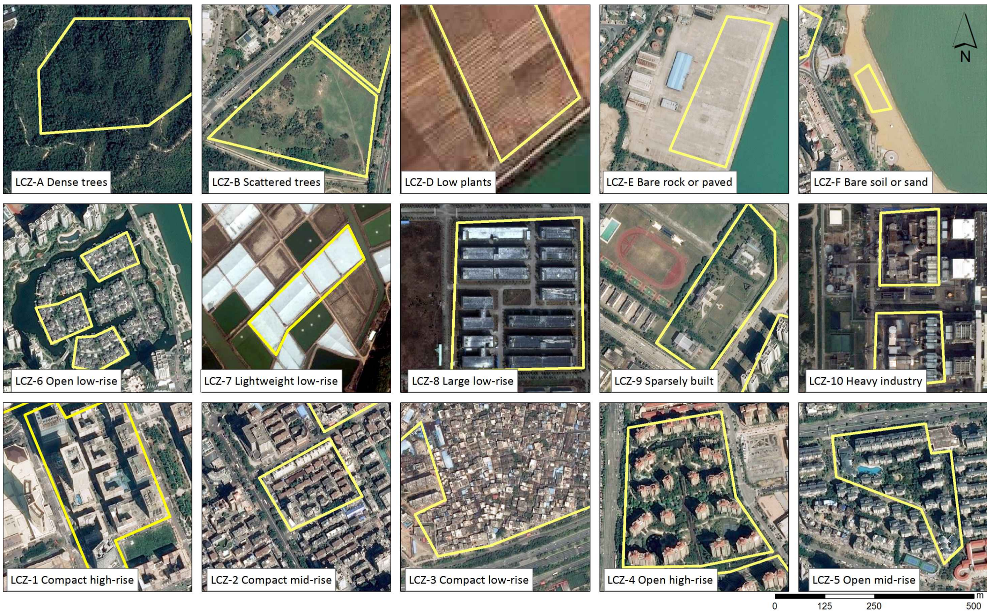


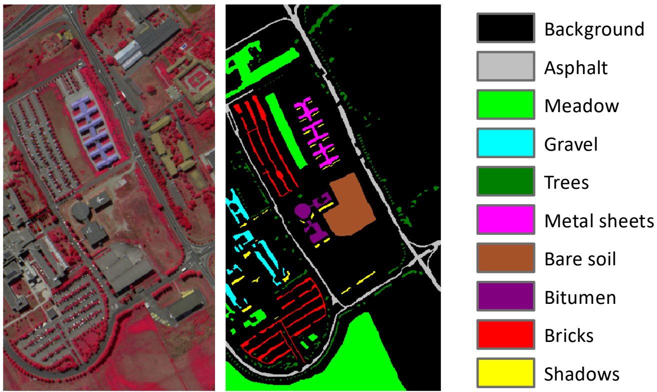
Figure 4. Local climate zones (except for water) in this study.

Table 3. Number of samples for the Zhuhai-Macau LCZ dataset and the University of Pavia dataset.

The Zhuhai-Macau LCZ Dataset					**The University of Pavia**	
Class		**Trainning Samples**	**Training Objects**	**Test Samples**	**Class**	**Samples**
LCZ-1	Compact high-rise	52	13	72	Asphalt	6631
LCZ-2	Compact mid-rise	26	10	36	Meadows	18,649
LCZ-3	Compact low-rise	168	118	251	Gravel	2099
LCZ-4	Open high-rise	141	81	98	Trees	3064
LCZ-5	Open mid-rise	50	31	32	Metal sheets	1345
LCZ-6	Open low-rise	73	41	55	Bare soil	5029
LCZ-7	Lightweight low-rise	127	46	157	Bitumen	1330
LCZ-8	Large low-rise	143	32	132	Bricks	3682
LCZ-9	Sparsely built	23	3	16	Shadows	947
LCZ-10	Heavy industry	19	1	40		
LCZ-A	Dense trees	88	43	76		
LCZ-B	Scattered trees	8	6	20		
LCZ-D	Low plants	36	4	103		
LCZ-E	Bare rock or paved	45	5	32		
LCZ-F	Bare soil or sand	59	39	23		
LCZ-G	Water	190	42	261		
Total		1248	515	1404		42,776

2.3. The University of Pavia Dataset

The University of Pavia (The data were downloaded from http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes), a popular benchmark hyperspectral dataset, is used to test the proposed method as well. This dataset was collected using the Reflective Optics System Imaging Spectrometer (ROSIS) sensor over the urban area of the University of Pavia, Italy, on 8 July 2002. The size of the data is 610 × 340 with a spatial resolution of 1.3 m and 103 spectral bands. There are nine classes in this dataset. The details of sample numbers are shown in Table 3 and the false color map with reference data is shown in Figure 5. In the experiment, we conducted the principal component analysis (PCA) and used the top three components for classification to remove redundant features and simulate common high spatial resolution imagery with only RGB channels.



(a) (b)

Figure 5. The University of Pavia dataset. (**a**) False color map; (**b**) Reference data.

3. Methods

3.1. Object-Based Classification Strategy

3.1.1. OBIA

For the Sentinel Guangzhou dataset, image objects were delineated from the optical and SAR images using the multiresolution segmentation algorithm embedded in eCognition [55]. Inspired by Qi et al. [9], who conducted image segmentation on the Pauli RGB composition image of polarimetric

SAR data, we performed image segmentation on four spectral bands, which provide the highest spatial resolution (10 m). We also slightly over-segmented the images to ensure the segmentation accuracy. The suitable parameters were determined through a heuristic process. A scale parameter of 30 was found suitable based on visual interpretation. The shape and compactness parameters were set as 0.10 and 0.80, respectively. The entire area was segmented into 329,725 image objects.

For OBIA, the mean values; standard deviation; and four categories of textural information, namely, gray-level co-occurrence matrix (GLCM) homogeneity, GLCM contrast, GLCM dissimilarity, and GLCM entropy, were extracted from each image channel. Eighteen indicators related to shape and extent, namely, border length, width, asymmetry, relative boarder to image border, elliptic fit, density, number of pixels, radius of smallest enclosing ellipse, rectangular fit, length, length/width, volume, radius of largest enclosed ellipse, shape index, compactness, roundness, area, and boarder index, were also extracted. These features were used in OBIA-RF.

For the Zhuhai-Macau LCZ dataset with 100 m spatial resolution, the segmentation algorithm was performed on all the channels. A scale of 20 was found suitable and the image was delineated into 71,583 image objects. For the University of Pavia dataset, the segmentation algorithm was performed on the top three principal components from PCA. The image was delineated into 5275 image objects under a scale of 8.

3.1.2. OBPR

To better clarify the proposed method, a systematic illustration of four approaches to obtain object-based thematic maps is presented in Figure 6. The conventional OBIA (Figure 6a) extracts spatial and textural features from image objects, and then uses one of the one-dimensional classifiers to classify image objects. However, we need to manually design these features to utilize the spatial information with OBIA.

The invention of CNNs simplifies the process of feature extraction, because CNNs can automatically learn spatial features for classification with back-propagation during training. To integrate deep learning with OBIA, we can treat a fixed-size image as the representation of an image object (Figure 6b) and then convert the problem of assigning irregular image objects with LULC types to a problem of classifying rectangular images to LULC types. Consequently, we can apply deep learning models from computer vision directly for LULC mapping. But the image needs to be large enough to cover the entire object, resulting inaccurate classification in small objects. The third solution is based on FCNs (Figure 6c), where an image is fed in an FCN and produce a segmentation map.

As shown in Figure 7, OBPR was performed after pixel-based classification. After classification, each pixel x has a predicted label c, where $c \in \{1, 2, ..., C\}$, and C is the number of LULC types. Based on the segmentation, each pixel with the predicted label (x, c) belongs to an image object R_k. Pixels in the same object should be classified as the same LULC type. Thus, for $R_k = \{(x_1, c_1), (x_2, c_2)..., (x_{n_k}, c_{n_k})\}$, where n_k is the number of pixels inside R_k, there should be only one corresponding label $\hat{c_k}$. Assuming $f(c_k)$ is the frequency of the predicted labels inside R_k, the assigned label $\hat{c_k}$ of the object is determined as:

$$\hat{c_k} = argmax(f(c_k)). \tag{1}$$

Figure 6d shows an example of the OBPR strategy. If more than one major label exist in an image object, a reasonable solution will be to assign the label with the highest occurrence in the entire image. As there are many pixels inside an object, we can randomly assign the smallest or largest integer as the object label without affecting the classification results. The analysis is presented in Section 4.7.2.

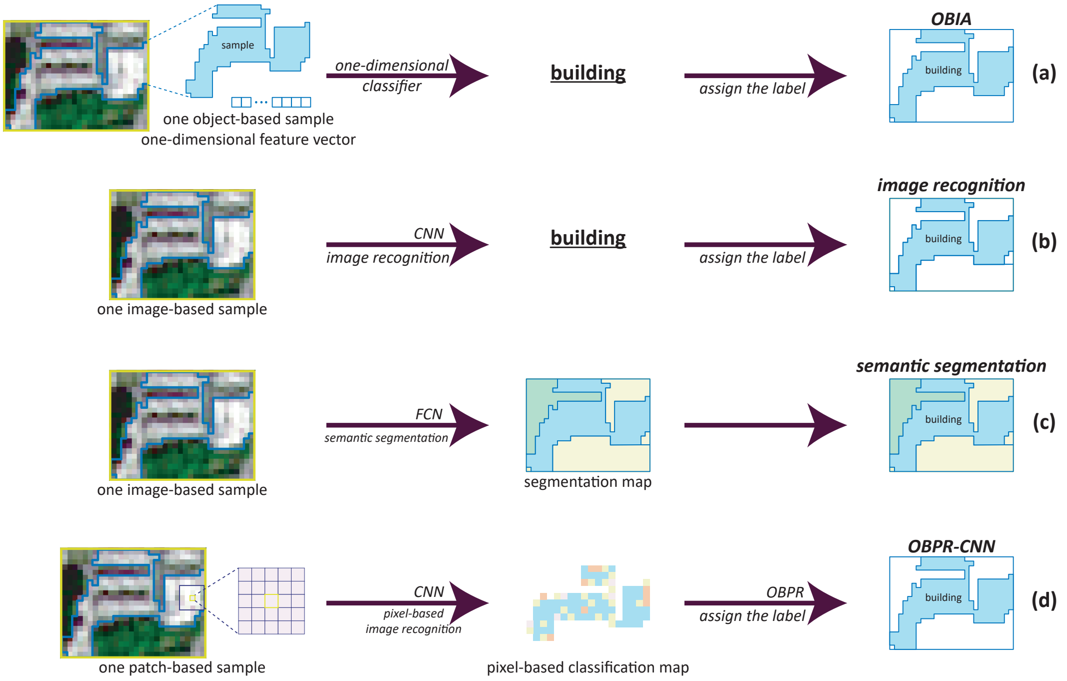


Figure 6. A systematic illustration of four different approaches to obtain object-based thematic maps.

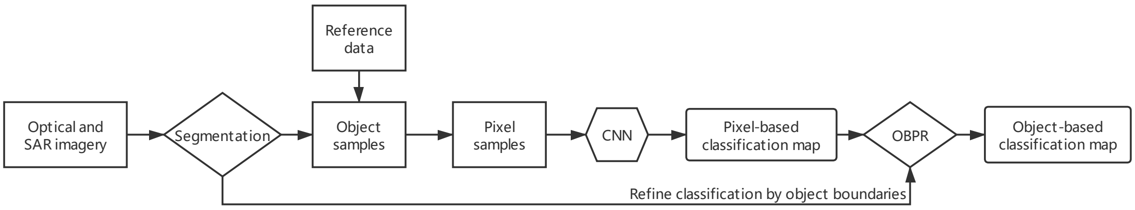

Reference data
Optical and SAR imagery
Segmentation
Object samples
Pixel samples
CNN
Pixel-based classification map
OBPR
Object-based classification map
Refine classification by object boundaries

Figure 7. Flowchart of the proposed OBPR-CNN. The image is first segmented into image objects. Based on reference data, we select object-based samples, and pixels inside image objects serve as training samples in the CNN. After obtaining the pixel-based classification map, object boundaries are applied to refine the classification result and obtain an object-based thematic map.

3.2. Machine Learning Algorithms

3.2.1. SVM

SVM is a competitive machine learning algorithm for its excellent generalization even with limited training samples. This is because SVM distinguishes training samples by finding the separate hyperplane related to the maximal margin and describes and specifies the hyperplane not by all the samples but only by the support vectors, which are the subset of samples. However, it may take considerable training time for huge datasets, especially when the most popular kernel Radial Basis Function is adopted. When applying pixel-based SVM, we randomly downsampled the number of samples to 400 per class to obtain the result within an acceptable time. The experiments with SVM were conducted on Python 3.6 using scikit-learn [56], which uses LibSVM [57] as its core algorithm. The parameters of C and γ were first coarse-grid-searched within $\{2^{-5}, 2^{-3}, ..., 2^{15}\}$ and $\{2^{-15}, 2^{-13}, ..., 2^{3}\}$. Then, the parameters were fine-grid-searched using the temporary best parameters $\hat{C}, \hat{\gamma}$ within $\hat{C}, \hat{\gamma} \times \{2^{-1.75}, 2^{-1.5}, ..., 2^{1.75}\}$. Fivefold cross-validation was performed to optimize the parameters.

3.2.2. RF

RF is one of the powerful ensemble learning algorithms. There are mainly three advantages of this algorithm. First, it can handle thousands of input features without feature selection. Second, it can estimate the importance of input features. Third, it is insensitive to noise and outliers. Given the aforementioned advantages, it has a great potential for LULC mapping with multi-source data. We implemented the experiments using scikit-learn [56] with Python 3.6. The number of trees was searched within $\{20, 40, ..., 200\}$ because previous studies showed that the optimal number of trees was within [20, 200] [58,59]. The number of max features was searched within $\{1, 2, ..., \sqrt{n} + 1\}$, where n is the number of input features.

3.2.3. CNN

Deep learning has achieved considerable successes in computer vision and natural language processing. CNN, as a successful deep learning architecture, has been applied to remote sensing image classification and achieved state-of-the-art results. A comparison of CNN and other machine learning algorithms is shown in Figure 8. CNN presents two advantages in remote sensing image classification [60]. First, the convolutional layers automatically learn useful textural and spatial features from the input patch-based samples. Second, the nonlinear layers, such as rectified linear unit and batch normalization, construct powerful functions for use in fusing and transforming the extracted features for the classification.

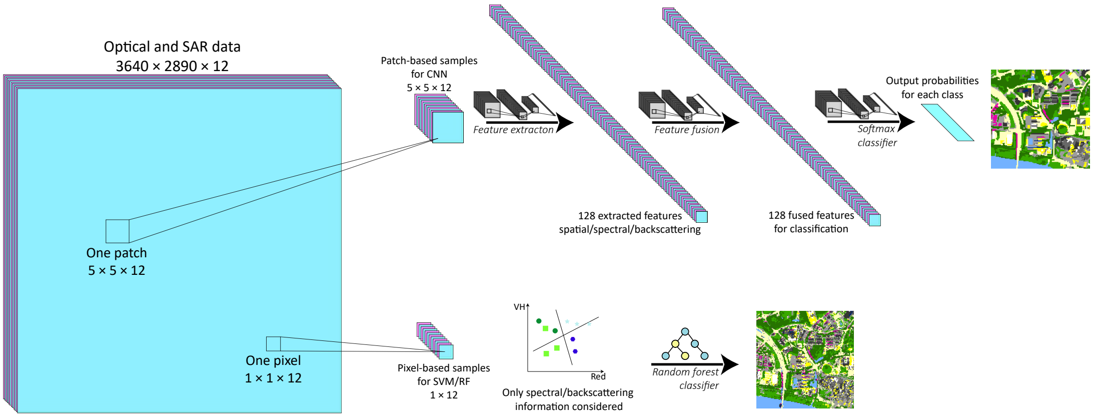


Figure 8. Classification systems.

In this study, CNN was implemented using the Keras [61] library with TensorFlow [62] back-end in Python 3.6. Wide contextual residual network (WCRN) modified from the contextual CNN by replacing AlexNet with ResNet was adopted because it can run on CPU and perform competitively [34,63]. Twenty percent of the training samples were separated as the validation set, and then the model that achieved the smallest loss in validation was used for classification. A total of 128 fused and transformed spatial-spectral-backscattering features were extracted. The output scores z of each class for one pixel were normalized as probabilities p by the softmax classifier:

$$p_j(z) = \frac{e^{z_j}}{\sum_k e^{z_k}} \tag{2}$$

where p_j is the probability of the pixel to belong to the j-th land cover category and k is the number of land cover classes.

4. Results and Discussion

The proposed method was evaluated on three datasets, two optical-SAR datasets with diverse spatial resolutions and one hyperspectral dataset. OBIA-SVM and OBIA-RF were selected as the benchmark methods. The experiments were conducted on a machine equipped with a 3.5 GHz Intel Xeon E3-1241 v3 CPU and 8G RAM.

4.1. Results on the Optical-SAR Sentinel Guangzhou Dataset

The classification results on the Sentinel Guangzhou dataset are shown in Table 4. The proposed method achieved the highest classification accuracy, with OA of 95.33% and κ of 0.94, considerably larger than those achieved by OBIA-SVM (OA of 90.22% and κ of 0.89) and OBIA-RF (OA of 88.20% and κ of 0.86). The classification accuracy (OA of 91.10% and κ of 0.90) obtained by the standard CNN was already larger than those by OBIA-SVM/RF. This result indicates that the spatial information extracted by CNN was helpful in LULC classification. Among the LULC classes, urban LULC categories, especially new town and roads, were better classified using the proposed method. New town is well planned mid-rise to high-rise buildings. Both new town and roads are very small in the image, surrounding by complicated urban structure. The spatial information for these LULC classes thus is very important for their accurate classification. Therefore, the ability of CNN to extract spatial features considerably helped the classification task.

Table 4. Classification accuracy (%) comparison among the proposed method and other competitors using 10 m and 20 m optical-SAR data. Experimental results with the same background color were produced by the same kind of classifier.

Classification Accuracy	SVM	OBIA-SVM	OBPR-SVM	RF	OBIA-RF	OBPR-RF	CNN	OBPR-CNN
New town	53.50	62.15	69.41	52.77	61.92	64.68	86.23	**93.71**
Old town	77.63	88.39	89.75	80.87	83.28	94.40	86.80	**94.72**
Bare land	90.29	92.19	92.69	90.04	91.20	91.68	94.15	**96.70**
Port area	77.38	86.57	82.46	77.67	83.97	80.80	88.92	**91.89**
Green houses	94.13	98.99	99.39	93.85	99.60	99.84	98.59	**100.00**
Lawn	81.22	84.88	85.56	81.55	82.13	85.56	80.96	**87.89**
Industry and commerce	89.04	93.10	91.93	90.86	89.63	92.17	93.61	**94.61**
Roads	55.52	67.98	61.99	53.38	58.76	54.78	76.12	**87.54**
Villas	66.76	81.40	79.95	71.89	78.87	84.07	84.12	**92.18**
Water	98.73	**100.00**	99.84	98.88	**100.00**	**100.00**	98.59	**100.00**
Trees	87.77	94.21	93.20	88.11	92.74	93.19	92.44	**95.25**
Overall accuracy (OA,%)	84.35	90.22	89.73	84.86	88.20	89.54	91.10	**95.33**
Kappa coefficient (κ)	0.82	0.89	0.88	0.82	0.86	0.88	0.90	**0.94**
Average accuracy (AA,%)	79.27	86.35	86.01	79.99	83.83	85.56	89.13	**94.05**

The proposed OBPR strategy remarkably improved the OA of CNN by 4.23%, indicating that the spatial constraint by object boundaries was very useful for LULC classification. When OBPR was combined with SVM/RF, the performance was as competitive as OBIA-SVM/RF (OA of 90.22% and OA of 88.20%), obtaining an OA of 89.73% and 89.64%, respectively. Most of the previous studies argued that the effectiveness of OBIA came from two aspects. One was that through OBIA we could obtain object-based classification maps. The other was that we could generate textural features from OBIA. Although the classification results in this study confirmed that OBIA-SVM/RF outperformed pixel-based SVM/RF, the superiority actually came from the spatial constraint that pixels inside one object should share the same label.

To evaluate the robustness of the proposed method, we constructed two subsets of training samples; the results are presented in Table 5. The classification results with the subsets of training samples were consistent with those using 150 object samples per class. The proposed method obtained OA of 93.76% with 50 object samples per class, 4.65% and 7.26% higher than those by OBIA-SVM and by OBIA-RF, respectively. When only 10 labeled objects per class available, the proposed method significantly outperformed other classification algorithms, achieving OA of 89.81%, 6.22% greater than that of OBIA-SVM and 7.38% greater than that of OBIA-RF. The margin between OBPR-CNN and OBIA-SVM/RF enlarged when the training samples became limited.

Previous studies demonstrated that sufficient samples (at least 50 samples per class) are need to construct the classification system for remote sensing image classification. Otherwise the performance of classifiers will be significantly degraded. As only 10 labeled objects per class were available, the classification was not satisfactory with OBIA. However, 10 objects contained at least 163 pixels in our study (Table 2). When the pixel samples were used, the number of pixel samples (163 per class) would be enough for classifiers to construct powerful classification systems. Therefore, we observed increases in OA of 3.03% and 3.62% from OBIA-SVM/RF to OBPR-SVM/RF.

Table 5. Overall accuracies (%) of the proposed method and the competitors. The best result of each dataset (row) is highlighted in bold, and the best result of each method (column) is underlined. NoS denotes number of training samples per class.

NoS	Optical	SAR	SVM	OBIA-SVM	OBPR-SVM	RF	OBIA-RF	OBPR-RF	MLP	OBPR-MLP	CNN	OBPR-CNN
150	10 m	*No*	77.86	84.59	86.70	76.96	82.33	87.71	75.58	81.53	88.50	**93.61**
	10 m	*Yes*	81.31	86.58	87.84	81.94	86.39	89.04	80.89	87.02	89.70	**94.57**
	10 m + 20 m	*No*	83.27	89.88	90.20	82.34	86.69	88.89	79.08	85.52	90.11	**94.43**
	10 m + 20 m	*Yes*	84.35	90.22	89.73	84.86	88.20	89.54	82.69	88.14	91.10	**95.33**
50	10 m	*No*	76.74	83.23	86.25	73.06	81.11	82.48	72.71	78.19	85.25	**91.10**
	10 m	*Yes*	78.40	84.96	86.74	79.84	85.50	86.55	79.23	85.58	87.57	**92.40**
	10 m + 20 m	*No*	79.68	88.94	88.23	79.77	85.74	86.36	73.68	80.26	87.20	**91.21**
	10 m + 20 m	*Yes*	82.65	89.11	89.40	83.18	86.50	88.80	79.46	85.01	88.91	**93.76**
10	10 m	*No*	71.19	77.48	77.85	69.22	71.40	76.89	68.98	72.87	84.12	**89.01**
	10 m	*Yes*	75.53	77.68	82.62	73.40	78.19	80.91	61.16	65.45	82.13	**86.46**
	10 m + 20 m	*No*	78.81	81.28	87.83	75.48	78.45	82.58	70.20	76.07	84.47	**89.70**
	10 m + 20 m	*Yes*	79.20	83.59	86.62	78.69	82.43	86.05	74.03	79.18	84.41	**89.81**

4.2. Results on the Zhuhai-Macau LCZ Dataset

The second experiment was conducted on the Zhuhai-Macau LCZ dataset, and the results are shown in Table 6. The proposed method outperformed other competitors in all cases. With full optical-SAR features, OBPR-CNN obtained OA of 77.64%, whereas the best non-CNN method OBPR-MLP only obtained OA of 70.94%, and the best OBIA method OBIA-RF achieved OA of 68.09%.

The best OA on this dataset was lower than 80%, indicating the complication of LCZ classification [47]. One of the crucial problems is that different LCZs might have the same material and result in the same spectral information in the satellite imagery. Thus, spatial information is essential to distinguish them. The comparison between OBPR-CNN and non-CNN method (77.64% versus 70.94%) indicated the advance of CNN, and the comparison between the proposed method and OBIA (77.64% versus 68.09%) illustrated the effectiveness of OBPR.

Table 6. Overall accuracies (%) of the Zhuhai-Macau LCZ dataset amongst the proposed method and the competitors. The best result of each dataset (row) is highlighted in bold, and the best result of each method (column) is underlined.

Optical	SAR	SVM	OBIA-SVM	OBPR-SVM	RF	OBIA-RF	OBPR-RF	MLP	OBPR-MLP	CNN	OBPR-CNN
10 m	*No*	58.40	61.61	63.32	63.75	57.05	67.09	60.40	62.68	70.30	**72.15**
10 m	*Yes*	64.32	68.02	67.17	65.03	59.90	67.88	62.89	66.45	67.88	**70.37**
10 m + 20 m	*No*	67.95	70.94	72.08	65.88	67.38	67.09	65.81	69.52	75.43	**76.92**
10 m + 20 m	*Yes*	68.09	66.74	70.80	67.38	68.09	68.02	66.17	70.94	75.21	**77.64**

4.3. Results on the University of Pavia Dataset

A popular hyperspectral dataset, the University of Pavia, was used to test the proposed method on high spatial resolution imagery. For this dataset, we randomly selected 5, 10, and 100 pixel samples per class, while the remaining samples served for validation. The image objects where these pixels lied in served as training samples in OBIA. All the pixels inside the training objects were used for training. In this manner, the OBPR strategy is in fact applying semi-supervised learning based on superpixels [64]. The OAs are presented in Table 7.

The proposed OBPR-CNN outperformed other methods among all sample sets. When training samples were sufficient (i.e., 100 per class), OBPR-CNN obtained OA of 96.32%, whereas the best non-CNN method OBPR-RF achieved OA of 94.90%. With the number of training samples decreasing, OBPR achieved OA of 95.70% using 10 sample per class and OA of 85.88% using five samples per class, whereas OBPR-RF obtained OAs of 78.82% and 67.28%, respectively. We can observe that OBPR-CNN was more superior when the samples were limited. This finding is contradictory to the common sense that deep learning models like CNNs need a large amount of training samples.

Table 7. Overall accuracies (%) of the University of Pavia dataset amongst the proposed method and the competitors. The best result of each sample set (row) is highlighted in bold. NoS denotes the number of training samples per class.

NoS	SVM	OBIA-SVM	OBPR-SVM	RF	OBIA-RF	OBPR-RF	MLP	OBPR-MLP	CNN	OBPR-CNN
100	72.26	85.66	78.57	90.28	89.92	94.90	79.80	86.27	93.32	**96.32**
10	63.04	67.75	69.13	73.26	75.66	78.82	68.40	76.18	90.04	**95.70**
5	62.39	58.51	67.23	63.96	61.91	67.28	69.30	75.55	83.26	**85.88**

Not only OBPR-CNN obtained higher OAs than conventional OBIA methods, but also non-CNN OBPR methods outperformed OBIA methods. Take RF for example as it is less sensitive to noisy features. OBPR-RF consistently obtained higher OAs compared with OBIA-RF, e.g., 94.90% versus 89.92% using 100 samples per class. Such results illustrate that we should rethink of the OBIA strategy, as we can obtain classification maps with higher OAs with OBPR. As mentioned before, the strategy of OBPR is indeed one kind of semi-supervised learning, which is based on superpixels and can increase the number of training samples. This explains why OBPR outperformed OBIA.

4.4. Contribution of SAR Data to LULC Classification

Results of two optical-SAR datasets indicated the effectiveness of SAR data when it comes to LULC classification. To analyze such effects, the difference of producer's accuracy (PA) and the user's accuracy (UA) between optical-only and optical-SAR data on the Sentinel Guangzhou dataset is presented in Figure 9. We found that the effects of SAR data depended on classifiers. When the RF classifier was used, the PAs and UAs of all the LULC classes were increased, especially for urban LULC types such as villas, roads, port areas, and new town. The improvement was less apparent when SAR data were combined with 10 m and 20 m optical data, but the differences of PAs and UAs of roads, port areas, and new town were still near 10%. Similar improvements were found in the classification results of SVM. The improvements made by CNN were not as significant as those using RF. The reason is that CNN can extract spatial information from the patch-based samples, whereas SVM and RF are pixel-based classifiers and lack the ability to make use of spatial features.

The contribution of SAR data to LULC classification was partly because of the side-looking imaging mode of radar remote sensing and the long wavelength of the C-band. The side-looking imaging mode and the C-band radar signals resulted in the low intensity of some ground targets, such as roads, which belong to the impervious surface and usually show high albedo in optical remote sensing images. In Figure 10, the SAR backscatter from roads was low and provided different physical information beyond optical remote sensing. The classification maps in Figure 10 show that with the addition of SAR data, the roads were identified accurately. Moreover, the classification maps produced by the combination of SAR and optical data presented minimal salt-and-pepper effects probably because the data from different sensors exhibited varied noises and the signals from radar remote sensing might have denoised the optical image.

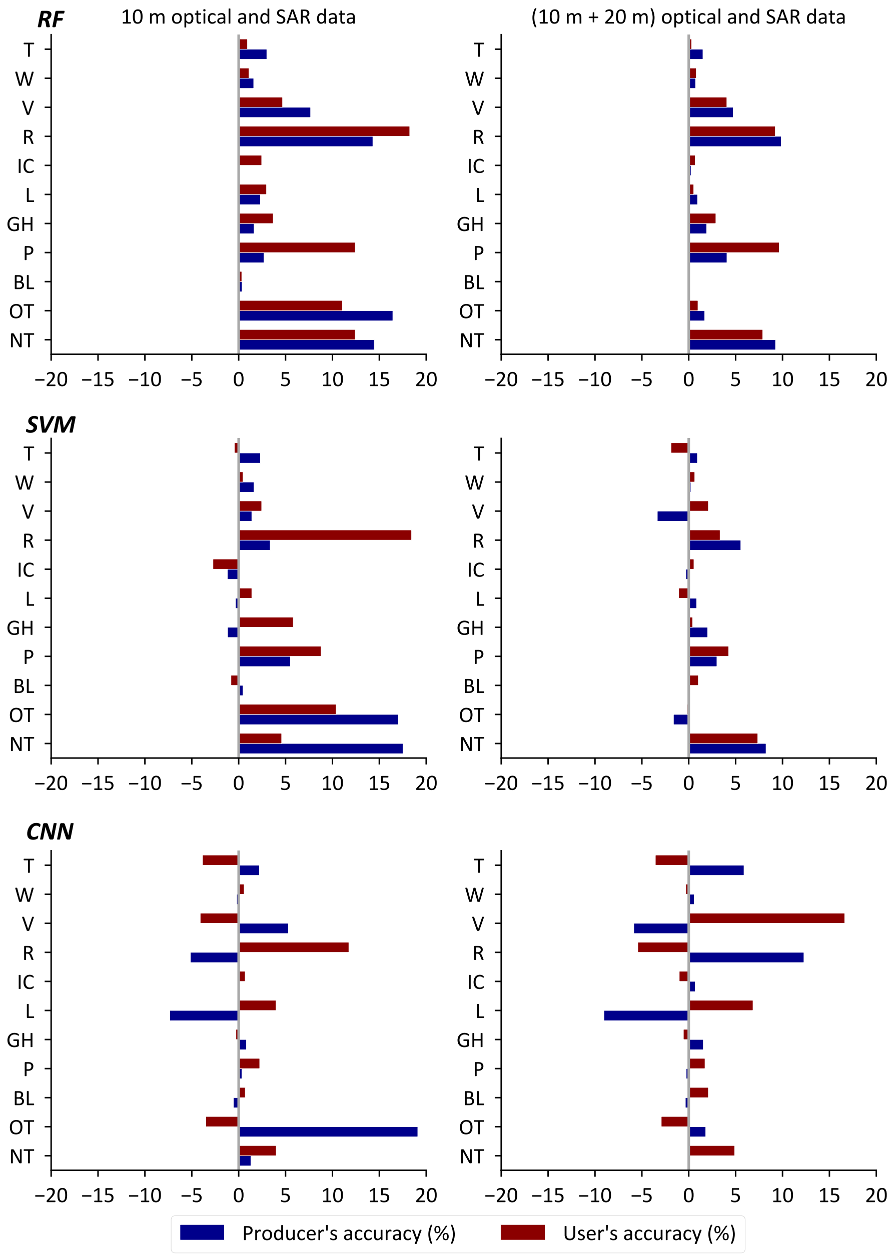


Figure 9. The contribution of SAR data for LULC classification on the Sentinel Guangzhou dataset. For RF and SVM, adding SAR data results in significant improvements on PAs and UAs of all the classes, showing the usefulness of SAR data. For CNN, the improvement is marginal, indicating that spatial information is more important for LULC classification than backscattering signals.

In Figure 11, the optical data suffered from shadow effects, which were severe in urban centers with city skylines. Most of the shadows were mistakenly classified as water, when the optical images were used alone, because the shadow effects were inevitable when a single data set from optical remote sensing was used. The radar backscatter from water was markedly lower than that from urban areas, due to the side-looking image mode of radar remote sensing and the complicated structure of the urban center. The differences in radar backscatter and textural features between water and urban areas could be extracted by CNN and resulted in accurate LULC classification, thus showing the advancement of the proposed OBPR-CNN.

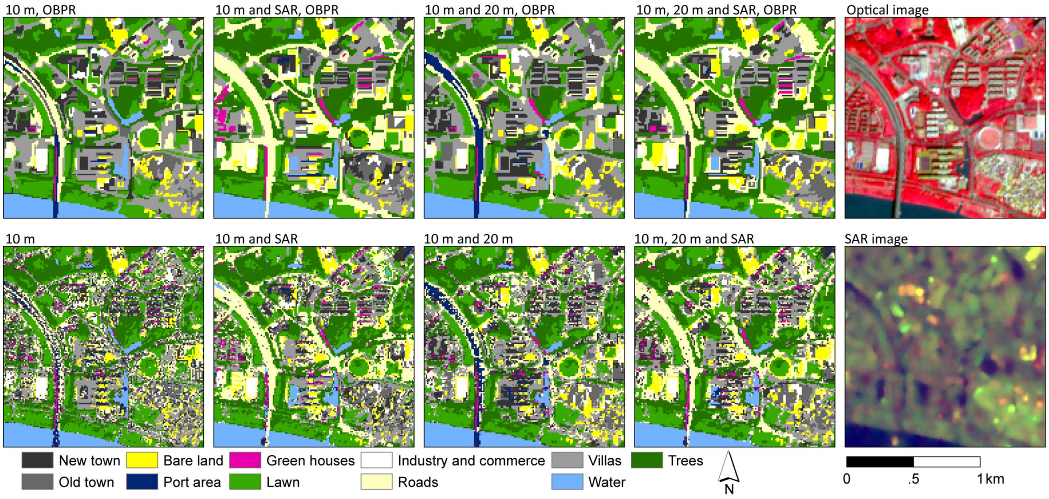


Figure 10. Classification maps obtained by RF. The addition of SAR data helped distinguish between roads and port area; the classification maps obtained with SAR data exhibited small salt-and-pepper effects; the classification maps processed with OBPR were compact.

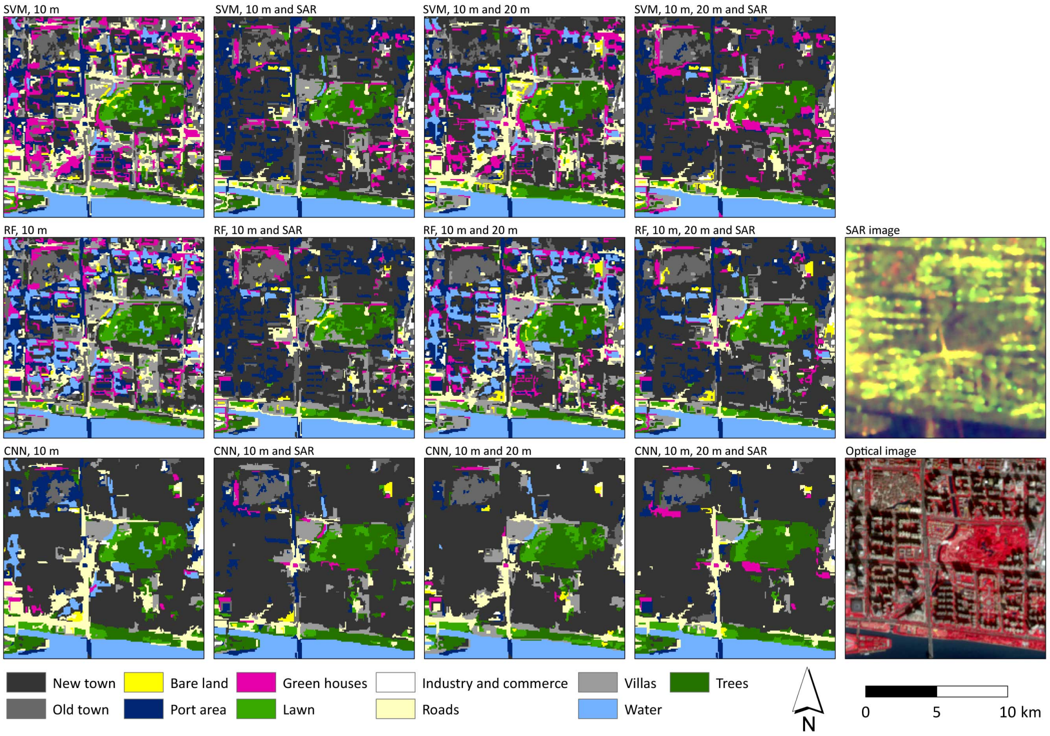


Figure 11. Classification maps obtained using OBPR-SVM, OBPR-RF, and OBPR-CNN. The addition of SAR data and the use of OBPR-CNN resulted in accurate LULC classification in urban areas with building shadows.

4.5. Feature Importance of the Sentinel Optical and SAR Data

The importance of features estimated by RF is presented in Figure 12 to illustrate the aforementioned conclusion that OBIA outperformed pixel-based algorithms mainly because it can obtain object-based thematic maps instead of utilizing textural features. The most important feature is the mean value of Band 12, which belongs to the middle infrared with a spatial resolution of 20 m. Other mean values of infrared spectral bands also played a significant role for LULC classification. This may be because the signals from infrared bands were less affected by atmosphere and were informative for LULC mapping. The mean values of 10 m spectral bands and the SAR backscatters

(VH and VV polarizations) were crucial for classification as well. However, the GLCM textures were not important (less than 2%) for classification as estimated by RF. The importance of features illustrated that GLCM textures were not as important as object boundaries for LULC classification using Sentinel optical and SAR data. Instead of using hand-crafted features, CNNs can learn spatial features automatically, which were optimized by back-propagation and had a better performance than hand-crafted GLCM textures.

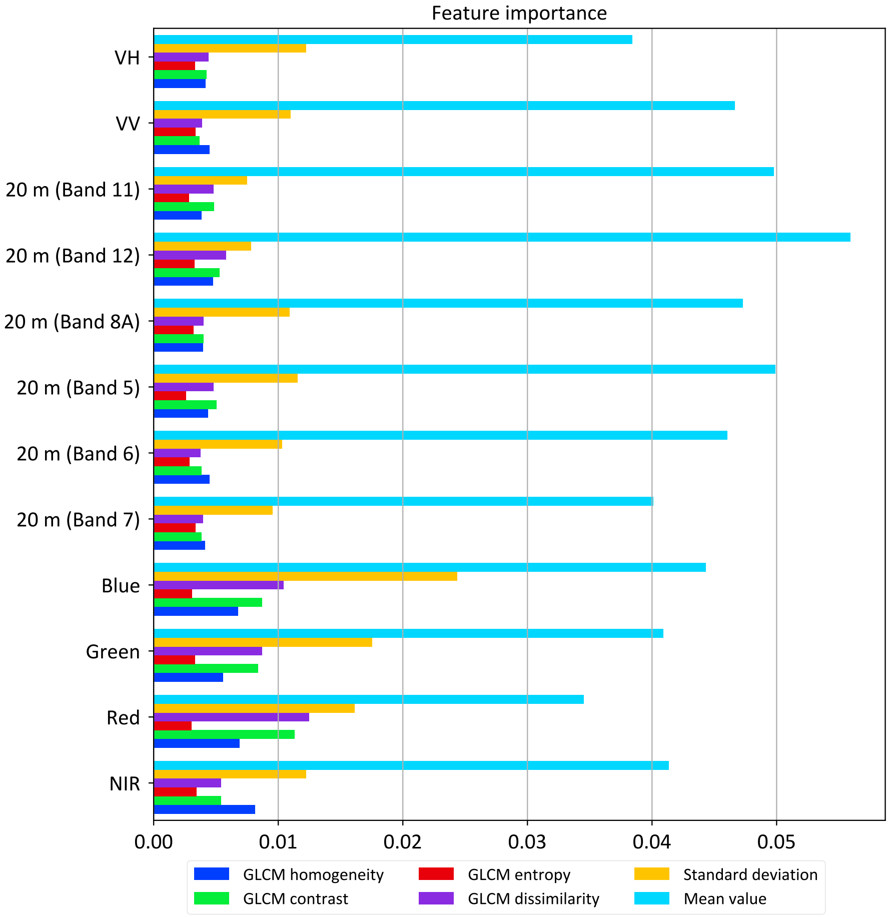


Figure 12. Feature importance estimated by RF classifier. The GLCM textural features have a very limited effect for LULC classification on the Sentinel Guangzhou dataset, whereas the mean value of each channel plays a significant role. The features generated by OBIA are not important enough and thus it might explain why OBIA is less competitive than OBPR-CNN.

4.6. CNN as Feature Extractor

The power of CNN lies in its capability to extract spatial features and fuses the spatial-spectral features into a high-dimensional feature space where the classifier can well distinguish the different classes. If the extracted features serve as input of SVM and RF, then the results of SVM and RF should be as competitive as those using CNN with the softmax classifier. As shown in Table 8, CNN-RF and CNN-SVM represent classification results by RF and SVM based on spatial-spectral features extracted by CNN. Notably, CNN indicates classification results based on the softmax classifier. The OAs using CNN as feature extractor for SVM, RF and softmax are competitive. Interestingly, the best OA of each dataset, including optical-only data and optical-SAR data, was obtained by RF. This may reflect the excellent generalization of RF that it can handle noisy and thousands of input features without feature selection.

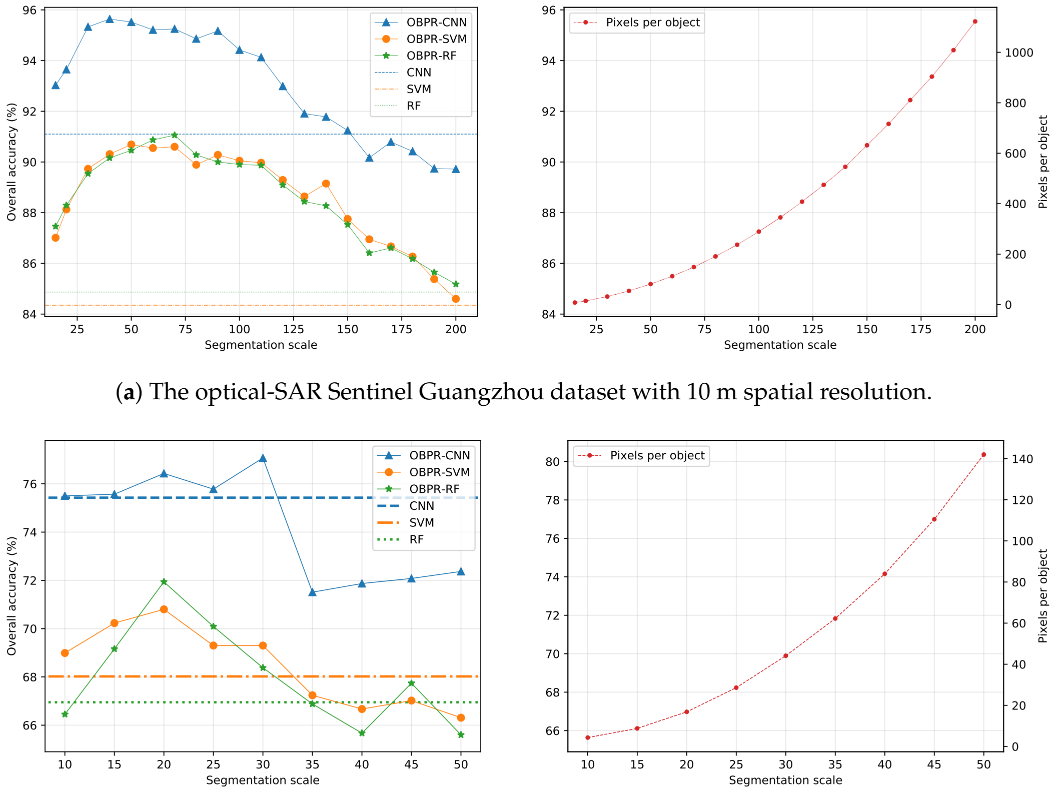
Table 8. Overall accuracy (%) of CNN (softmax), SVM and RF using CNN as feature extractor.

Optical		SAR	OBPR		
10 m	**20 m**	**10 m**	**CNN**	**CNN-SVM**	**CNN-RF**
√			93.61	94.65	**95.31**
√		√	**94.57**	94.44	**94.57**
√	√		94.43	94.7	**95.47**
√	√	√	95.33	95.47	**95.97**

4.7. Sensitivity Analysis

4.7.1. Sensitivity Analysis of the Segmentation Scale

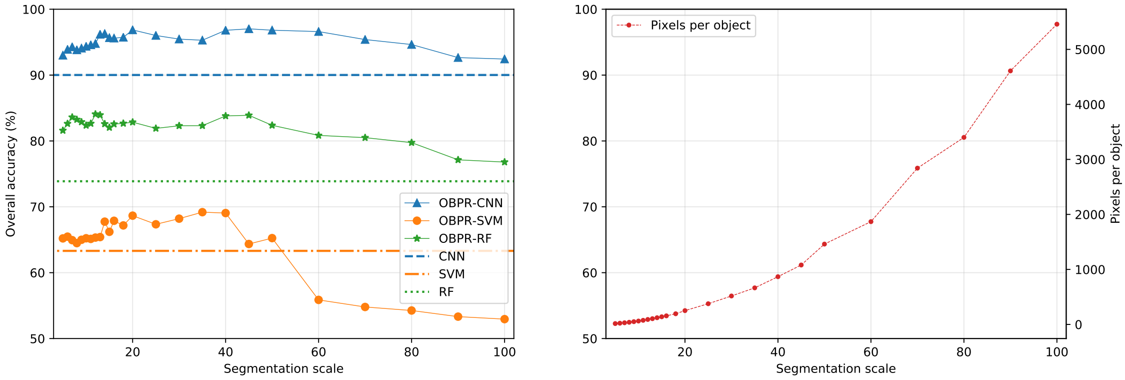
We have conducted experiments to analyze the sensitivity of the segmentation scale of OBPR. The results are presented in Figure 13. For the Sentinel Guangzhou dataset (Figure 13a), the performance of OBPR is effective. The optimal scale lies in the range of 40 to 80, whereas the average number of pixels per object varies approximately from 80 to 200 (8000–20,000 m^3). When the scale is very small and each object contains very little pixels, the improvement made by OBPR is limited yet observable. When the scale is very large (e.g., greater than 150), the performance of OBPR is degraded. Nevertheless, the improvement by OBPR is stable and effective as the scale of 30 to 120 is quite wide and safe. From the segmentation images (Figure 14). We can observe that a scale of 30 produces a very fragmented segmentation, whereas a scale of 120 leads to under-segmentation. With a heuristic process, one can easily find a proper segmentation scale in this range.



(**a**) The optical-SAR Sentinel Guangzhou dataset with 10 m spatial resolution.

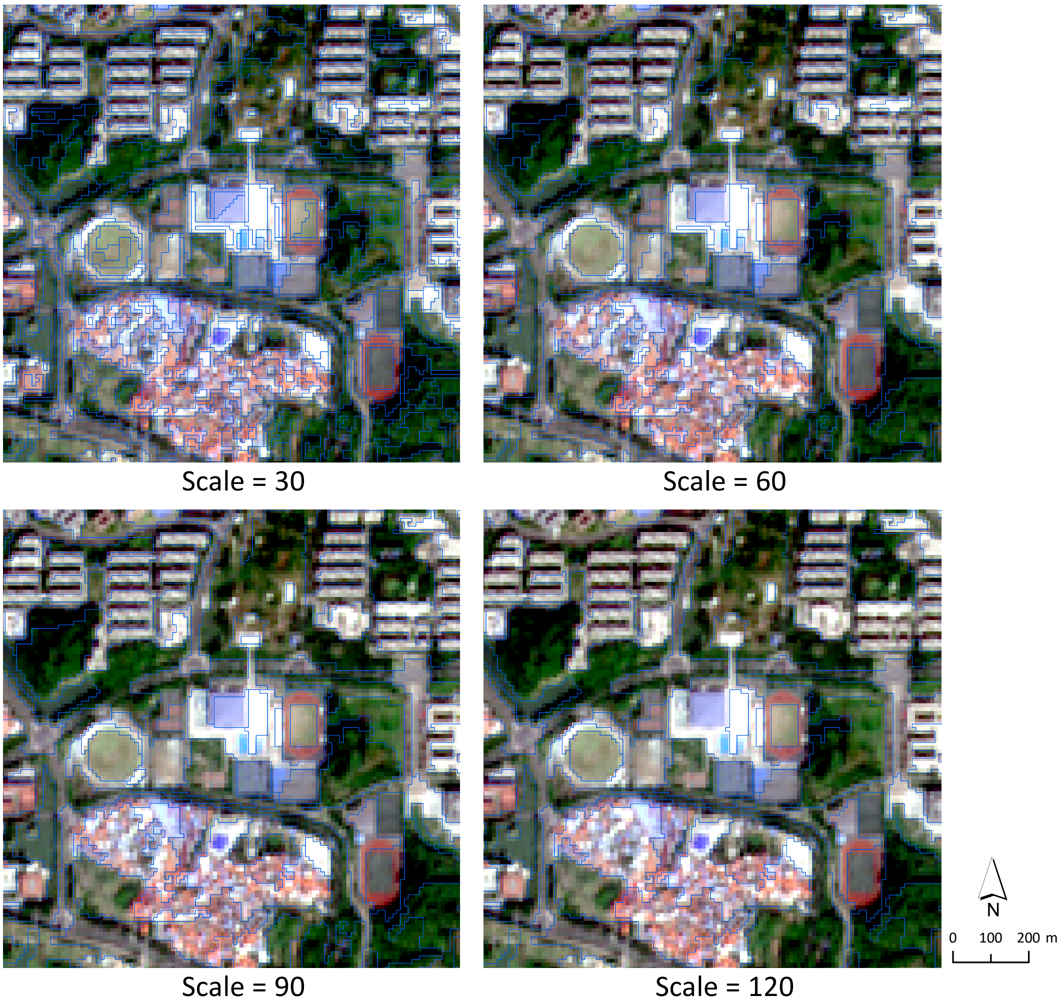
(**b**) The optical-SAR Zhuhai-Macau LCZ dataset with 100 m spatial resolution.

Figure 13. *Cont.*



(**c**) The University of Pavia dataset with 1.3 m spatial resolution.

Figure 13. Sensitivity analysis of the segmentation scale. (**Left**) OA as function of the segmentation scale. (**Right**) Average number of pixels per object as function of the segmentation scale.



Scale = 30 Scale = 60

Scale = 90 Scale = 120

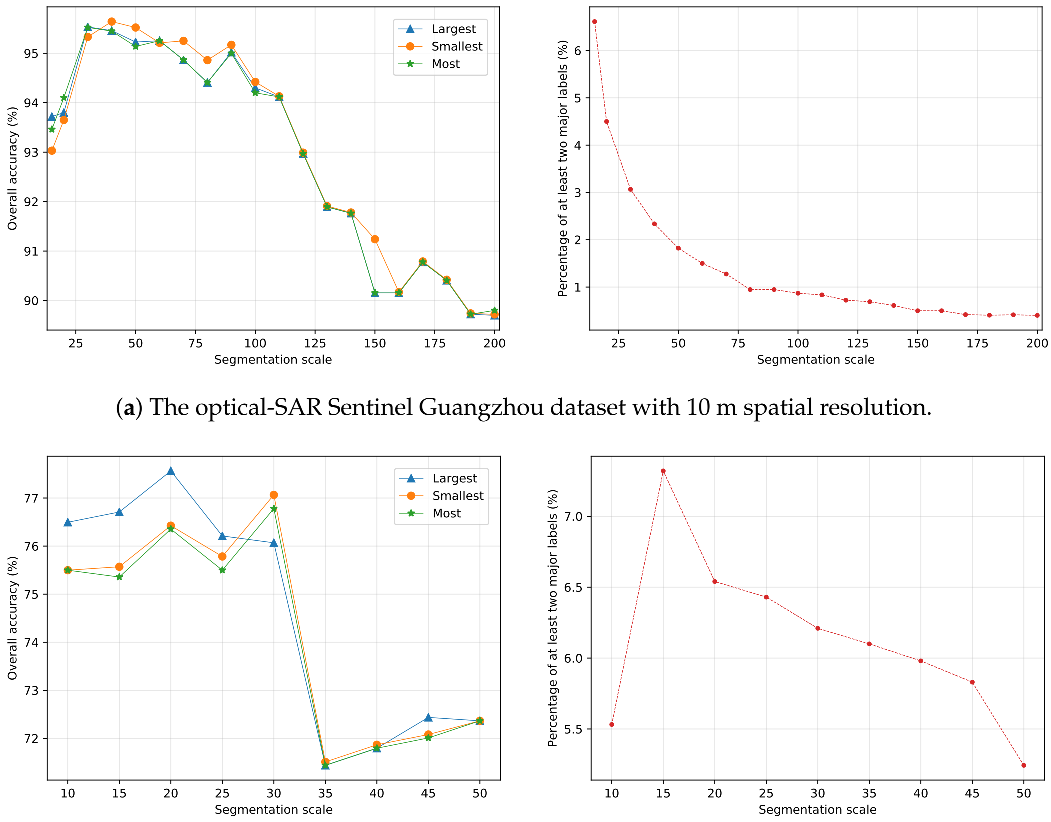
Figure 14. Segmentation maps with various segmentation scales. A scale of 30 leads to observable oversegmentaton, whereas a scale of 120 leads to undersegmentation. Still, the performance of OBPR is satisfactory in this range (Figure 13a).

For the Zhuhai-Macau LCZ dataset (Figure 13b), the improvement by OBPR is not as effective as that of the Sentinel Guangzhou dataset. The effective scale varies from 15 to 30 (0.1–0.4 km^3 per object). Since this dataset is with a very low spatial resolution (100 m), it might not be suitable for object-based classification.

The result of the University of Pavia dataset (Figure 13c) confirms that OBPR can obtain very satisfactory performance for high spatial resolution imagery. For classification maps by CNN and RF, OBPR can consistently improve the OA. For the classification map produced by SVM, OBPR degrades the result after the scale reaches to 60. A segmentation scale of 60 is extremely large, as each object contains almost 2000 pixels and the whole image is segmented to no more than 200 objects. An OA of 63.04% is quite low. When applying the majority voting strategy, incorrect classification could lead to a larger error. Nevertheless, with reasonable heuristic processing, it is easy to find a proper segmentation scale. The sensitivity analysis indicates that OBPR is less sensitive to the scale. OBPR can be very effective in a wide range of the segmentation scale with high to medium spatial resolution imagery.

4.7.2. The Choice of Three Majority Voting Strategies

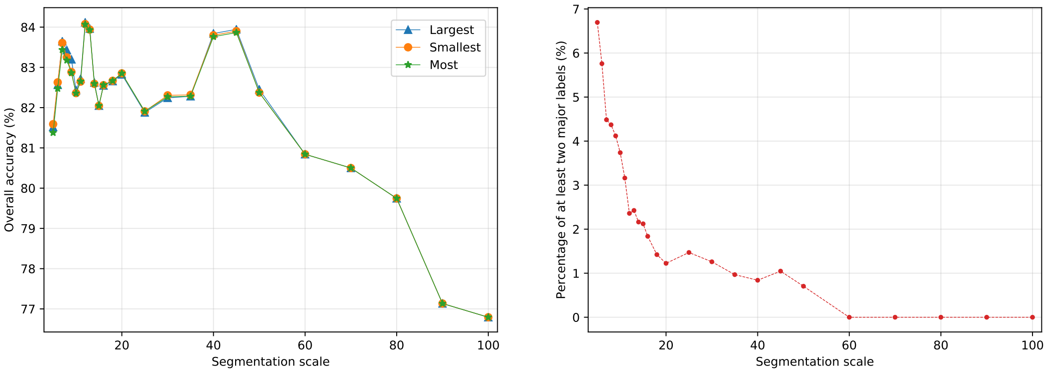
The results of three choices of majority voting strategies are presented in Figure 15. From the left side of Figure 15 we can observe that that the choice of the majority strategy has very limited effects (less than 0.5%) on OA. The result is expected because many pixels were present inside an object. Only a few objects encountered the situation in which at least two major labels were detected (right side of Figure 15). In addition, the randomness of the dominant LULC type can ease the problem.



(**a**) The optical-SAR Sentinel Guangzhou dataset with 10 m spatial resolution.

(**b**) The optical-SAR Zhuhai-Macau LCZ dataset with 100 m spatial resolution.

Figure 15. *Cont.*



(**c**) The University of Pavia dataset with 1.3 m spatial resolution.

Figure 15. (**Left**) OA as function of the segmentation scale with different majority voting strategies. Three choices of majority voting strategies. Largest: assign the largest integer as the object label. Smallest: assign the smallest integer as the object label. Most: assign the most frequent class in the candidates as the object label. (**Right**) percentage of at least two major labels as function of the segmentation scale.

4.7.3. The Effect of Patch Size

Classification maps of diverse patch sizes are shown in Figure 16. A large patch size (35× 35) results in inaccurate classification of small roads between mulberry fish ponds, whereas a small patch size (5× 5) better captures small objects in the image. In addition, using a small patch size is computationally efficient, providing users an opportunity to apply deep learning models on their personal laptops without expensive GPUs.

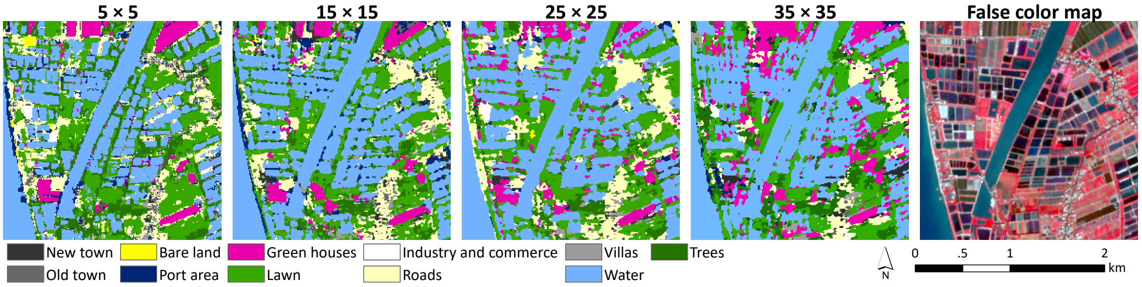


Figure 16. Classification maps produced by CNNs with different patch size. A large patch size leads to inaccurate classification of small roads between fish ponds.

4.7.4. Number of Trees in RF

Previous studies have a wide discussion on the optimal number of trees when using RF classifier [65]. We tested the RF classifier on the optical-SAR Sentinel Guangzhou dataset with the number of trees in the range of [20, 500] (Figure 17). The classification accuracy is insensitive to the number of trees as pointed out by Du et al. [59], especially after it is up to 60. In addition, OBPR significantly outperforms OBIA regardless of the number of trees.

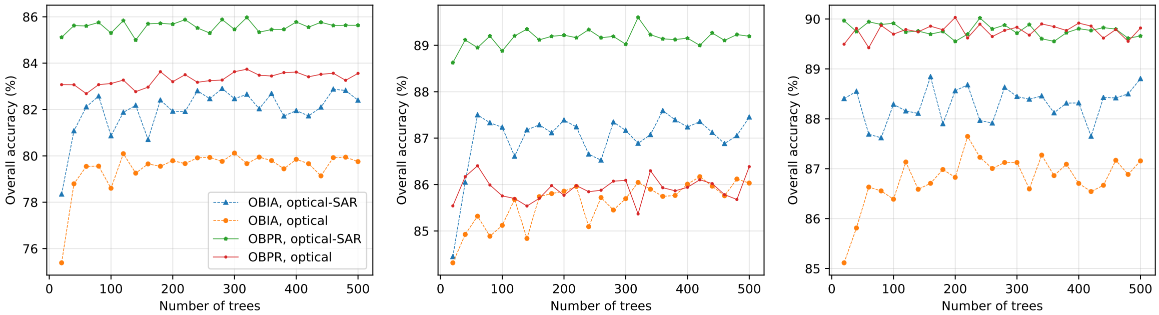


Figure 17. OA as function of number of trees in RF. We can observe that number of trees has very little effect on OA after it grows to 60. (**Left**) 10 samples per class. (**Middle**) 50 samples per class. (**Right**) 150 samples per class.

5. Conclusions

In this study, we developed a new method that equips CNNs with the ability to produce object-based thematic maps for LULC classification. Compared with other three methods, the proposed method OBPR-CNN can present promising results with limited labeled samples. Our method was tested on three datasets with diverse spatial resolutions and different classification systems. It obtained a remarkable result with OA of 95.33% and κ of 0.94 on the Sentinel Guangzhou dataset and a satisfactory result with OA of 77.64% with limited and imbalanced labeled samples on the Zhuhai-Macau LCZ dataset using Sentinel multispectral and SAR data. Our method also achieved a very competitive result (OA of 95.70%) on the popular hyperspectral dataset the University of Pavia with only 10 labeled samples per class. Such results outperformed traditional OBIA methods.

Through further studies, we found that object-based GLCM textures were less important for LULC mapping in this study. The performance of OBIA mainly lies in its capability to produce object-based classification maps rather than generating textural features. The hand-crafted GLCM textures were less superior than those learned by CNNs. Therefore, OBPR-CNN is better than OBIA to obtain object-based thematic maps. The combined use of optical and SAR data depended on classifiers. When CNNs were used, the addition of SAR data had limited improvement for LULC mapping, whereas the addition of SAR data played a significant role in distinguishing urban ground targets using one-dimensional classifier, i.e., SVM, RF and MLP. This study is the first to evaluate the performance of optical and SAR data using CNNs. From the results, we may conclude that in the era of deep learning, spatial information extracted by CNN is more crucial for LULC mapping than the combined use of optical and SAR data. Nevertheless, the addition of SAR data and the spatial information extracted by CNN helped distinguish urban LULC classes such as roads, new town, and port areas.

Future studies may explore high spatial resolution SAR imagery (e.g., TerraSAR-X) using the proposed method. The fusion of multimodal, multisource, and multitemporal data for complicated classification tasks such as LCZ classification is worth investigation as well.

Author Contributions: Conceptualization, S.L. and Z.Q.; Formal analysis, S.L. and Z.Q.; Funding acquisition, Z.Q., X.L. and A.G.-O.Y.; Investigation, S.L. and Z.Q.; Methodology, S.L.; Resources, X.L. and A.G.-O.Y.; Supervision, Z.Q.; Visualization, S.L.; Writing—original draft, S.L. and Z.Q.; Writing—review & editing, S.L., Z.Q., X.L. and A.G.-O.Y.

Funding: This study was supported by the National Natural Science Foundation of China (Grant No. 41601445), the Guangdong Natural Science Foundation (Grant No. 2016A030313230), the Young Scholar Research Fund of Sun Yat-sen University (Grant No. 16lgpy05), the Hong Kong Research Grants Council General Research Fund (HKU 17209214), the Chan To-Haan Endowed Professorship Fund and Distinguished Research Achievement Award of the University of Hong Kong, and the Science and Operational Applications Research - Education International (SOAR-EI) Initiative (Grant No. 5167).

Acknowledgments: The author would like to thank the editors and anonymous reviewers for their insightful comments and suggestions, which have significantly improved this paper.

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

Abbreviations

The following abbreviations are used in this manuscript:

SAR	synthetic aperture radar
OBIA	object-based image analysis
OBPR	object-based post-classification refinement
LULC	land use and land cover
FCN	fully convolutional network
CNN	convolutional neural network
SVM	support vector machine
RF	random forest
MLP	multi-layer perceptron
LCZ	local climate zone
OA	overall accuracy
PA	producer's accuracy
UA	user's accuracy
GLCM	gray-level co-occurrence matrix
VV	vertical transmit and vertical receive
VH	vertical transmit and horizontal receive
NoS	number of samples

References

1. Townshend, J.; Justice, C.; Li, W.; Gurney, C.; McManus, J. Global land cover classification by remote sensing: present capabilities and future possibilities. *Remote Sens. Environ.* **1991**, *35*, 243–255. [CrossRef]
2. Stewart, I.D.; Oke, T.R. Local climate zones for urban temperature studies. *Bull. Am. Meteorol. Soc.* **2012**, *93*, 1879–1900. [CrossRef]
3. Margono, B.A.; Potapov, P.V.; Turubanova, S.; Stolle, F.; Hansen, M.C. Primary forest cover loss in Indonesia over 2000–2012. *Nat. Clim. Chang.* **2014**, *4*, 730. [CrossRef]
4. Qi, Z.; Yeh, A.G.O.; Li, X.; Zhang, X. A three-component method for timely detection of land cover changes using polarimetric SAR images. *ISPRS J. Photogramm. Remote Sens.* **2015**, *107*, 3–21. [CrossRef]
5. Geneletti, D.; Gorte, B.G. A method for object-oriented land cover classification combining Landsat TM data and aerial photographs. *Int. J. Remote Sens.* **2003**, *24*, 1273–1286. [CrossRef]
6. Manandhar, R.; Odeh, I.O.A.; Ancev, T. Improving the Accuracy of Land Use and Land Cover Classification of Landsat Data Using Post-Classification Enhancement. *Remote Sens.* **2009**, *1*, 330–344. [CrossRef]
7. Khatami, R.; Mountrakis, G.; Stehman, S.V. Mapping per-pixel predicted accuracy of classified remote sensing images. *Remote Sens. Environ.* **2017**, *191*, 156–167. [CrossRef]
8. Li, X.; Yeh, A. Multitemporal SAR images for monitoring cultivation systems using case-based reasoning. *Remote Sens. Environ.* **2004**, *90*, 524–534. [CrossRef]
9. Qi, Z.; Yeh, A.G.O.; Li, X.; Lin, Z. A novel algorithm for land use and land cover classification using RADARSAT-2 polarimetric SAR data. *Remote Sens. Environ.* **2012**, *118*, 21–39. [CrossRef]
10. Qi, Z.; Yeh, A.G.O.; Li, X. A crop phenology knowledge-based approach for monthly monitoring of construction land expansion using polarimetric synthetic aperture radar imagery. *ISPRS J. Photogramm. Remote Sens.* **2017**, *133*, 1–17. [CrossRef]
11. Liu, W.; Yang, J.; Li, P.; Han, Y.; Zhao, J.; Shi, H. A Novel Object-Based Supervised Classification Method with Active Learning and Random Forest for PolSAR Imagery. *Remote Sens.* **2018**, *10*, 1092. [CrossRef]
12. Chen, W.; Gou, S.; Wang, X.; Li, X.; Jiao, L. Classification of PolSAR images using multilayer autoencoders and a self-paced learning approach. *Remote Sens.* **2018**, *10*, 110. [CrossRef]
13. Liesenberg, V.; de Souza Filho, C.R.; Gloaguen, R. Evaluating moisture and geometry effects on L-band SAR classification performance over a tropical rain forest environment. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 5357–5368. [CrossRef]

14. Van Beijma, S.; Comber, A.; Lamb, A. Random forest classification of salt marsh vegetation habitats using quad-polarimetric airborne SAR, elevation and optical RS data. *Remote Sens. Environ.* **2014**, *149*, 118–129. [CrossRef]
15. Torres, R.; Snoeij, P.; Geudtner, D.; Bibby, D.; Davidson, M.; Attema, E.; Potin, P.; Rommen, B.; Floury, N.; Brown, M.; et al. GMES Sentinel-1 mission. *Remote Sens. Environ.* **2012**, *120*, 9–24. [CrossRef]
16. Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; et al. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sens. Environ.* **2012**, *120*, 25–36. [CrossRef]
17. Reiche, J.; Verbesselt, J.; Hoekman, D.; Herold, M. Fusing Landsat and SAR time series to detect deforestation in the tropics. *Remote Sens. Environ.* **2015**, *156*, 276–293. [CrossRef]
18. Kussul, N.; Lemoine, G.; Gallego, F.J.; Skakun, S.V.; Lavreniuk, M.; Shelestov, A.Y. Parcel-Based Crop Classification in Ukraine Using Landsat-8 Data and Sentinel-1A Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 2500–2508. [CrossRef]
19. Zhang, Y.; Zhang, H.; Lin, H. Improving the impervious surface estimation with combined use of optical and SAR remote sensing images. *Remote Sens. Environ.* **2014**, *141*, 155–167. [CrossRef]
20. Zhang, H.; Xu, R. Exploring the optimal integration levels between SAR and optical data for better urban land cover mapping in the Pearl River Delta. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *64*, 87–95. [CrossRef]
21. Blaschke, T. Object based image analysis for remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 2–16. [CrossRef]
22. Wang, P.; Huang, C.; Tilton, J.C.; Tan, B.; de Colstoun, E.C.B. HOTEX: An approach for global mapping of human built-up and settlement extent. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017; pp. 1562–1565.
23. Ruiz Hernandez, I.E.; Shi, W. A Random Forests classification method for urban land-use mapping integrating spatial metrics and texture analysis. *Int. J. Remote Sens.* **2018**, *39*, 1175–1198. [CrossRef]
24. Franklin, S.E.; Ahmed, O.S. Deciduous tree species classification using object-based analysis and machine learning with unmanned aerial vehicle multispectral data. *Int. J. Remote Sens.* **2018**, *39*, 5236–5245. [CrossRef]
25. Cleve, C.; Kelly, M.; Kearns, F.R.; Moritz, M. Classification of the wildland–urban interface: A comparison of pixel-and object-based classifications using high-resolution aerial photography. *Comput. Environ. Urban Syst.* **2008**, *32*, 317–326. [CrossRef]
26. Laliberte, A.S.; Rango, A. Texture and scale in object-based analysis of subdecimeter resolution unmanned aerial vehicle (UAV) imagery. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 761–770. [CrossRef]
27. Kim, M.; Warner, T.A.; Madden, M.; Atkinson, D.S. Multi-scale GEOBIA with very high spatial resolution digital aerial imagery: scale, texture and image objects. *Int. J. Remote Sens.* **2011**, *32*, 2825–2850. [CrossRef]
28. Pu, R.; Landry, S.; Yu, Q. Object-based urban detailed land cover classification with high spatial resolution IKONOS imagery. *Int. J. Remote Sens.* **2011**, *32*, 3285–3308. [CrossRef]
29. Zhang, C.; Xie, Z. Combining object-based texture measures with a neural network for vegetation mapping in the Everglades from hyperspectral imagery. *Remote Sens. Environ.* **2012**, *124*, 310–320. [CrossRef]
30. Blaschke, T.; Feizizadeh, B.; Hölbling, D. Object-based image analysis and digital terrain analysis for locating landslides in the Urmia Lake Basin, Iran. *IEEE J. Sel. Top. App. Earth Obs. Remote Sens.* **2014**, *7*, 4806–4817. [CrossRef]
31. Huang, X.; Zhang, L.; Wang, L. Evaluation of morphological texture features for mangrove forest mapping and species discrimination using multispectral IKONOS imagery. *IEEE Geosci. Remote Sens. Lett.* **2009**, *6*, 393–397. [CrossRef]
32. Wang, M.; Fei, X.; Zhang, Y.; Chen, Z.; Wang, X.; Tsou, J.; Liu, D.; Lu, X. Assessing texture features to classify coastal wetland vegetation from high spatial resolution imagery using completed local binary patterns (CLBP). *Remote Sens.* **2018**, *10*, 778. [CrossRef]
33. Pesaresi, M.; Gerhardinger, A.; Kayitakire, F. A robust built-up area presence index by anisotropic rotation-invariant textural measure. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2008**, *1*, 180–192. [CrossRef]
34. Lee, H.; Kwon, H. Going deeper with contextual CNN for hyperspectral image classification. *IEEE Trans. Image Process.* **2017**, *26*, 4843–4855. [CrossRef]
35. Maggiori, E.; Tarabalka, Y.; Charpiat, G.; Alliez, P. Convolutional neural networks for large-scale remote-sensing image classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 645–657. [CrossRef]

36. Cheng, G.; Han, J.; Lu, X. Remote sensing image scene classification: benchmark and state of the art. *Proc. IEEE* **2017**, *105*, 1865–1883. [CrossRef]
37. Liu, T.; Abd-Elrahman, A. Deep convolutional neural network training enrichment using multi-view object-based analysis of Unmanned Aerial systems imagery for wetlands classification. *ISPRS J. Photogramm. Remote Sens.* **2018**, *139*, 154–170. [CrossRef]
38. Rezaee, M.; Mahdianpari, M.; Zhang, Y.; Salehi, B. Deep convolutional neural network for complex wetland classification using optical remote sensing imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 3030–3039. [CrossRef]
39. Mahdianpari, M.; Salehi, B.; Rezaee, M.; Mohammadimanesh, F.; Zhang, Y. Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery. *Remote Sens.* **2018**, *10*, 1119. [CrossRef]
40. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [CrossRef]
41. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
42. Marmanis, D.; Wegner, J.D.; Galliani, S.; Schindler, K.; Datcu, M.; Stilla, U. Semantic segmentation of aerial images with an ensemble of CNNs. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *3*, 473. [CrossRef]
43. Kampffmeyer, M.; Salberg, A.B.; Jenssen, R. Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1–9.
44. Alshehhi, R.; Marpu, P.R.; Woon, W.L.; Dalla Mura, M. Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks. *ISPRS J. Photogramm. Remote Sens.* **2017**, *130*, 139–149. [CrossRef]
45. Yu, B.; Yang, L.; Chen, F. Semantic segmentation for high spatial resolution remote sensing images based on convolution neural network and pyramid pooling module. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 3252–3261. [CrossRef]
46. Marmanis, D.; Schindler, K.; Wegner, J.D.; Galliani, S.; Datcu, M.; Stilla, U. Classification with an edge: Improving semantic image segmentation with boundary detection. *ISPRS J. Photogramm. Remote Sens.* **2018**, *135*, 158–172. [CrossRef]
47. Xu, Y.; Ren, C.; Cai, M.; Edward, N.Y.Y.; Wu, T. Classification of Local Climate Zones Using ASTER and Landsat Data for High-Density Cities. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 3397–3405. [CrossRef]
48. Yokoya, N.; Ghamisi, P.; Xia, J.; Sukhanov, S.; Heremans, R.; Tankoyeu, I.; Bechtel, B.; Le Saux, B.; Moser, G.; Tuia, D. Open Data for Global Multimodal Land Use Classification: Outcome of the 2017 IEEE GRSS Data Fusion Contest. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 1363–1377. [CrossRef]
49. Liu, T.; Abd-Elrahman, A.; Morton, J.; Wilhelm, V.L. Comparing fully convolutional networks, random forest, support vector machine, and patch-based deep convolutional neural networks for object-based wetland mapping using images from small unmanned aircraft system. *GISci. Remote Sens.* **2018**, *55*, 243–264. [CrossRef]
50. Zhang, C.; Sargent, I.; Pan, X.; Li, H.; Gardiner, A.; Hare, J.; Atkinson, P.M. An object-based convolutional neural network (OCNN) for urban land use classification. *Remote Sens. Environ.* **2018**, *216*, 57–70. [CrossRef]
51. Liu, T.; Abd-Elrahman, A. An Object-Based Image Analysis Method for Enhancing Classification of Land Covers Using Fully Convolutional Networks and Multi-View Images of Small Unmanned Aerial System. *Remote Sens.* **2018**, *10*, 457. [CrossRef]
52. Zhao, W.; Du, S.; Emery, W.J. Object-based convolutional neural network for high-resolution imagery classification. *IEEE J. Sel. Top. Appl.Earth Obs. Remote Sens.* **2017**, *10*, 3386–3396. [CrossRef]
53. Lee, J.S.; Wen, J.H.; Ainsworth, T.L.; Chen, K.S.; Chen, A.J. Improved sigma filter for speckle filtering of SAR imagery. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 202–213.

54. Bechtel, B.; Alexander, P.J.; Beck, C.; Böhner, J.; Brousse, O.; Ching, J.; Demuzere, M.; Fonte, C.; Gál, T.; Hidalgo, J.; et al. Generating WUDAPT Level 0 data–Current status of production and evaluation. *Urban Clim.* **2019**, *27*, 24–45. [CrossRef]
55. Benz, U.C.; Hofmann, P.; Willhauck, G.; Lingenfelder, I.; Heynen, M. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS J. Photogramm. Remote Sens.* **2004**, *58*, 239–258. [CrossRef]
56. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
57. Chang, C.C.; Lin, C.J. LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol. (TIST)* **2011**, *2*, 27. [CrossRef]
58. Li, C.; Wang, J.; Wang, L.; Hu, L.; Gong, P. Comparison of classification algorithms and training sample sizes in urban land classification with landsat thematic mapper imagery. *Remote Sens.* **2014**, *6*, 964–983. [CrossRef]
59. Du, P.; Samat, A.; Waske, B.; Liu, S.; Li, Z. Random forest and rotation forest for fully polarized SAR image classification using polarimetric and spatial features. *ISPRS J. Photogramm. Remote Sens.* **2015**, *105*, 38–53. [CrossRef]
60. Zhang, L.; Zhang, L.; Du, B. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geosci. Remote Sens. Mag.* **2016**, *4*, 22–40. [CrossRef]
61. Chollet, F. Keras. 2015. Available Online: https://github.com/fchollet/keras (accessed on 1 February 2019).
62. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. Tensorflow: A System for Large-Scale Machine Learning. In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), Savannah, GA, USA, 2–4 November 2016; pp. 265–283.
63. Liu, S.; Luo, H.; Tu, Y.; He, Z.; Li, J. Wide Contextual Residual Network with Active Learning for Remote Sensing Image Classification. In Proceedings of the 2018 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Valencia, Spain, 22–27 July 2018; pp. 7145–7148.
64. Wu, H.; Prasad, S. Semi-supervised deep learning using pseudo labels for hyperspectral image classification. *IEEE Trans. Image Process.* **2018**, *27*, 1259–1270. [CrossRef]
65. Belgiu, M.; Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]