Article

Deep Learning Semantic Segmentation for Land Use and Land Cover Types Using Landsat 8 Imagery

Wuttichai Boonpook [1,*], Yumin Tan [2], Attawut Nardkulpat [1,3], Kritanai Torsri [4], Peerapong Torteeka [5], Patcharin Kamsing [6], Utane Sawangwit [5], Jose Pena [7] and Montri Jainaen [8]

[1] Department of Geography, Faculty of Social Sciences, Srinakharinwirot University, Bangkok 10110, Thailand

[2] School of Transportation Science and Engineering, Beihang University, Beijing 100191, China

[3] Faculty of Geoinformatics, Burapha University, Chonburi 20131, Thailand

[4] Hydro-Informatics Institute, Ministry of Higher Education, Science, Research and Innovation, Bangkok 10900, Thailand

[5] National Astronomical Research Institute of Thailand, Chiang Mai 50180, Thailand

[6] Air-Space Control, Optimization and Management Laboratory, Department of Aeronautical Engineering, International Academy of Aviation Industry, King Mongkut's Institute of Technology, Ladkrabang, Bangkok 10520, Thailand

[7] Venezuela Space Agency (ABAE), Caracas 1010, Venezuela

[8] Faculty of Management Science, Kamphaeng Phet Rajabhat University, Kamphaeng Phet 62000, Thailand

[*] Correspondence: wuttichaib@g.swu.ac.th; Tel.: +66-892274740

Abstract: Using deep learning semantic segmentation for land use extraction is the most challenging problem in medium spatial resolution imagery. This is because of the deep convolution layer and multiple levels of deep steps of the baseline network, which can cause a degradation problem in small land use features. In this paper, a deep learning semantic segmentation algorithm which comprises an adjustment network architecture (LoopNet) and land use dataset is proposed for automatic land use classification using Landsat 8 imagery. The experimental results illustrate that deep learning semantic segmentation using the baseline network (SegNet, U-Net) outperforms pixel-based machine learning algorithms (MLE, SVM, RF) for land use classification. Furthermore, the LoopNet network, which comprises a convolutional loop and convolutional block, is superior to other baseline networks (SegNet, U-Net, PSPnet) and improvement networks (ResU-Net, DeeplabV3+, U-Net++), with 89.84% overall accuracy and good segmentation results. The evaluation of multispectral bands in the land use dataset demonstrates that Band 5 has good performance in terms of extraction accuracy, with 83.91% overall accuracy. Furthermore, the combination of different spectral bands (Band 1–Band 7) achieved the highest accuracy result (89.84%) compared to individual bands. These results indicate the effectiveness of LoopNet and multispectral bands for land use classification using Landsat 8 imagery.

Keywords: deep learning semantic segmentation; LoopNet; Landsat 8; land use dataset; land use extraction; multispectral bands; Thailand

1. Introduction

Land use and land cover (LULC) is a challenging problem for identification and monitoring due to the complexity of LULC, the many types of human activities, and seasonal changes. To overcome these problems, remote sensing technology and machine learning algorithms are used for generating LULC maps [1,2]. Furthermore, up-to-date and accurate LULC mapping is required for geographic information system (GIS) applications, urban planning, and the monitoring of land cover changes [3,4]. To generate LULC maps, ground surveys and remote sensing methods are traditionally applied to land use classification. Ground surveys can provide accurate and precise mapping; however, taking

measurements in the field is time-consuming, and requires considerable expense of resources and human labor. Remote sensing classification that applies pixel-based and object-based classifications can produce land use classification with high accuracy. However, this method always requires sample data for training and parameter adjustment.

Most remote sensing imagery used for land use classification is collected from passive sensors. Such sensors can produce multispectral or hyperspectral bands that can be used for indicating LULC types [5]. Furthermore, remote sensing imagery has different kinds of spatial resolution, from low to high. Low spatial resolution imagery (more than 100 m) is good for recording data in a wide area. It is suitable for pixel-based classification in order to classify LULC types as defined in the LULC Level 1 classification system for remote sensing data [6,7]. Medium spatial resolution imagery (10–100 m) is popularly used in pixel-based classification and object-based classification [8]. This resolution is useful for classifying LULC types as defined in the LULC Level 2 classification system [9]. However, it has limitations in classifying the spatially heterogenous or complex features of the LULC types. High spatial resolution imagery (30 cm–10 m) is capable of classifying and extracting complex LULC features that are unseen in low or medium resolution imagery [10]. This resolution has high potential for the classification of LULC types as defined in the LULC Level 3 classification system [11]. These kinds of image resolutions are mostly classified by image processing using machine learning algorithms.

Machine learning has played a major role in remote sensing classification for over a decade, and delivers good performance for LULC classification. Many machine learning algorithms, such as K-nearest neighbor (KNN), maximum likelihood estimator (MLE), support vector machine (SVM), random forest (RF), decision tree (DT), etc., can be applied for pixel-based and object-based classification [5,8]. To implement these algorithms for LULC classification, Ali Jamali, 2019 [12] evaluated and compared eight machine learning models, including multilayer perceptron, simple logistic, J48, Lazy IBK, random forest, decision tree, DTNB, and non-nested generalized exemplars, for LULC mapping using Landsat 8 OLI. The results show that machine learning can classify land use features accurately. Later, Mayani et al., 2021 [13] demonstrated machine learning techniques, including random forest classifier, minimum distance (Euclidean) classifier, gradient tree boost classifier, naïve Bayes classifier, and classification and regression tree classifier, for land use classification using remote sensing imagery. The results showed that random forest had the highest accuracy and was better than other methods. However, these algorithms require training samples and fine-tuning of parameters for supervised learning models.

Recently, deep learning algorithms have been of great practical interest for automatic classification of remote sensing imagery. They can manage big data problems and the complexity of land use features. These algorithms have been applied to image classification and object segmentation and have proven very successful for several remote sensing applications [14]. A comprehensive survey of remote sensing image classification based on CNNs for land use classification was performed by Song et al., 2019 [15]. The results showed better land use classification using deep learning algorithms due to the improvements on the CNN model they present in terms of how image classification tasks are implemented. Later, Alem et al., 2020, [16] described deep learning methods for the LULC classification based on remote sensing imagery. Their study reviewed deep learning applications and methods, including convolutional neural network (CNN), generative adversarial network (GAN), and recurrent neural network (RNN). Analyzing this research, they found that deep learning is sufficient, and better for remote sensing imagery for LULC classification. Furthermore, Vali et al., 2020 [17] reviewed deep learning for land use and land cover classification using a passive sensor with both multispectral and hyperspectral imagery. They showed that deep learning with remote sensing imagery involves an enormous number of technical challenges that come with the remote sensing applications. There is a growing interest in commercially viable applications, ground-truth scarcity, on-board data processing, the fast volume growth of data, and the

limitations in memory. Their review article confirmed that challenging deep learning models explores and improves the capability of the classification methods that can be applied to LULC applications. With the development of LULC classification, Kussul et al., 2017 [18] presented a multilevel deep learning architecture for image processing using Landsat 8 and Sentinel 1A imagery. The result of their research showed better performance of the 2D CNN in overall accuracy over the random forest, ensemble of multilayer perceptron, and 1D-CNN methods. Storie and Henry, 2018 [19] investigated deep learning neural networks (FCN-8 and VGG-16 network) to analyze LULC types using Landsat 5 and Landsat 7. The results showed high average accuracy and quickly produced land use mapping. Later, Akhassan et al., 2020 [20] evaluated a deep learning framework for LULC mapping using multispectral satellite imagery. They applied different models of adjustment FCN architecture, such as FCN-VGG16, FCN-ResNet-101, FCN-GoogLeNet, and two extensions (a context module and an adversarial network) to evaluate the performance of land use classification. The comparison showed that the FCN-ResNet-101 and two extensions had the highest performance compared to the other models. Gharbia et al., 2021 [21] proposed a deep convolutional neural network for land cover classification. This research evaluated AlexNet and the VGG-16 model to classify land cover classes such as water, urban, agriculture, road, and desert using the Landsat dataset. The result showed that the VGG-16 model was superior to the AlexNet model, with high accuracy and good classification results. Sun et al., 2020 [22] investigated a deep neural network (DNN) with pixel-wise classification to classify the crop types over remote sensing data such as barley, soybean, corn, spring wheat, dry beans, alfalfa, and sugar beets. The results illustrated that the deep learning technique can perform well on crop-type classification with high overall accuracy. To improve the performance of a deep learning algorithm to provide LULC classification, several researchers have proposed a combination method. Zhang et al., 2019 [23] proposed a joint deep learning model which combined multilayer perceptron (MLP) and CNN for land use classification over high spatial resolution imagery. The result shows that this combination method can achieve high overall accuracy and an accurate classification result. Later, Rousset et al., 2021 [24] demonstrated the XGBoost algorithm and deep learning technique for land cover detection and classification using SPOT6 satellite imagery. The result of this combination method showed improvement in land use classification. For a remote sensing dataset, Helber et al., 2019 [25] presented a novel dataset called the EuroSAT and a patch-based deep learning method for land use and land cover classification using Sentinel 2 imagery. The proposed method was confirmed as having the highest overall accuracy for land use classification. Later, Zhang et al., 2021 [26] evaluated the CNN cascade feature and McODM classifier to classify a land use map using the UC Merced land use dataset. This combination method showed an improvement in the learning ability for scene level classification with a highly accurate result.

Studies of LULC classification and deep learning semantic segmentation methods are a rapidly growing research trend. Kotaridis and Lazaridou, 2021 [27] performed a meta-analysis of advanced remote sensing image segmentation. They analyzed statistics and quantitative data in a segmentation algorithm, and determined that the deep learning model works well on high spatial resolution imagery due to the clear target features and patterns. However, the price for high spatial resolution imagery is high, and only covers a small area. Thus, medium spatial resolution imagery is a challenging type of data for LULC classification. This type of imagery provides a wide coverage area and is free of charge, which is suitable for land use monitoring. Poliyapram et al., 2019 [28] utilized a simplified variant of the U-Net DCNN structure for segmenting land use features such as water, ice, and land over medium resolution satellite images (Landsat 8). Their proposed structure outperformed U-Net and the DeepWaterMap model. Mutreha et al., 2020 [29] studied settlement classification with a segmentation method using Landsat 8 imagery. The deep learning model was able to detect and extract settlements with highly accurate results. Later, Du et al., 2021 [30] implemented a deep semantic segmentation model for paddy rice classification and mapping using multi-temporal data from Landsat imagery.

It was demonstrated that the segmentation model could segment the rice paddy over large areas accurately. Pereira et al., 2021 [31] presented a large-scale dataset for active fire detection from Landsat 8 and the U-Net network. The algorithms achieved better performance on active fire segmentation. Furthermore, Wang et al., 2022 [32] studied a deep learning semantic segmentation approach which used an ensemble network for land use extraction using Sentinel 2A and Gaofen PMS. The results showed that the U-Net++ architecture with the Timm-RegNetY-320 backbone achieved the highest performance compared with other backbones (VGG-16, EfficientNet-b7, ResNEt50, and Timm-ResNet101). Based on the literature, the improved capacity of LULC segmentation illustrates that deep learning semantic segmentation has the potential to detect and extract the variety and complexity of land use features accurately. However, land use features over medium spatial resolution imagery have medium level features, and the baseline network architectures that are used for the supervised learning model have deep and complex convolutional layers to learn on multidimensional data. Thus, the deep convolutional layers can cause a degradation problem, and multiple levels of deep steps can cause the loss of small object features.

Therefore, the aim of this research is to propose a deep learning semantic segmentation for LULC segmentation using Landsat 8 imagery. Our research innovation and contribution is to apply automatic land use segmentation using a deep learning algorithm over open-source data. The adjustment network architecture, which consists of a convolution loop, convolution block, and three levels of deep steps, is designed to learn and extract the variety of land use features from medium spatial resolution imagery. Then, it is proposed to use a land use dataset which consists of five main classes, including background, water, agriculture, urban, miscellaneous, and forest, to evaluate the performance of land use segmentation using Landsat 8 imagery. Moreover, the multispectral bands are used to evaluate its performance for detecting and extracting complex land use features.

2. Materials and Methods

This research investigates an automatic LULC segmentation using Landsat 8 imagery. The land use dataset is used for the supervised learning model. The multispectral bands are evaluated for their performance in identifying LULC types. Furthermore, an adjustment network is proposed to learn the complexity of land use patterns and features, and is compared to the state-of-the-art network architecture for semantic segmentation. The performance of the proposed network is calculated using accuracy assessment methods. This section presents the details of the land use dataset, a proposed network architecture, and the accuracy assessment algorithms.

2.1. Land Use Dataset

The study area is in Thailand, and has different spectral reflectance and texture features of land use types in each season. For example, the forest has different plants in abundance each season, urban growth is increasing year by year, water bodies expand in the rainy season, agriculture has planting activity that follows the crop calendar, etc. Thus, multiple perspective views of land use features are required for the supervised learning model. For land use types in Thailand, as reported by the land development department in 2021 [33], there are six main classes: urban and built-up land (UB); agricultural land (AC), which includes paddy field, field crop, perennial, orchard, horticulture, swidden cultivation, pasture and farmhouse, aquatic plants, aquaculture land, and integrated/diversified farms; forest land (FR), which includes disturbed forest and dense forest; water bodies (WT); miscellaneous land (MC), which includes other miscellaneous land, marsh, and swamp; and background (BG). The largest coverage of the LULC is 54% which is agricultural land; forest land is next, at 34%. The other types comprise 5% for built-up land, 4% for miscellaneous land, and 3% for water bodies.

This land use dataset presents the LULC types using Landsat 8 imagery. The data were collected from an operational land imager sensor installed on the Landsat 8 satellite.

There are 16 days in a repeating cycle for collecting data in the same area. The satellite collects the spectral reflectance with different spectrum wavelengths over eleven bands; seven bands are recorded from the visible, near infrared, and shortwave infrared portions of the spectrum, which have 30 m of spatial resolution along a 185 km swath. These include Band 1 (0.43–0.45 µm), Band 2 (0.450–0.51 µm), Band 3 (0.53–0.59 µm), Band 4 (0.64–0.67 µm), Band 5 (0.85–0.88 µm), Band 6 (1.57–1.65 µm), and Band 7 (2.11–2.29 µm). These spectral bands were selected for the land use dataset. Samples for the land use dataset were captured in Thailand from 2020 to 2022, and were from Path 126–131 and Row 46–56, as shown in Figure 1a. The total number of images used in this dataset was 50 scenes. Each image has a pixel size of 7591 × 7761 in width and height, as shown in Figure 1b.
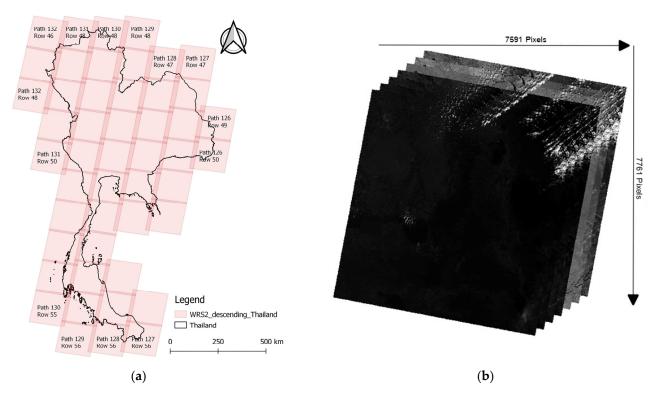

(**a**)   (**b**)

Figure 1. Landsat 8 WRS 2 Descending Path Row covering Thailand (**a**) and seven spectral bands of Landsat 8 (**b**).

Regarding the land use dataset preparation, the Landsat 8 imagery was clipped and split into images with 480 × 480 pixel size, which is the appropriate size for data samples to be fed into the network architecture. A sample set consisted of seven image bands (Band 1 to Band 7) and a labelled image. This research annotates the labeled images consisting of six semantic classes as background, forest, water, urban, miscellaneous, and agriculture. These classes relate to the LULC Level 1 from the USGS. The data sample set is shown in Figure 2. The total number of sample sets was 13,600 sets. These sample sets were divided into three groups named the training group, the validating group, and the testing group, as shown in Table 1.

Table 1. The number of samples for training, validating, and testing from the land use dataset.

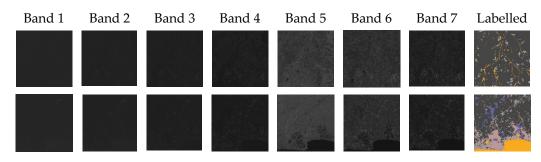| Dataset	Training	Validating	Testing
Land use dataset	8000	2000	3600

Figure 2. Sample sets of land use dataset showing the different spectral reflectance from Band 1 to Band 7 and labelled imagery.

2.2. Proposed Network Architecture

2.2.1. Convolution Block

This convolution block presents a learning feature algorithm that includes the convolutional layer, batch normalization, and activation function. The convolution layer learns the fundamental features with a set of filter banks, then batch normalization standardizes the feature map in order to stabilize the learning process and reduce the number of training epochs. The activation function determines the output value of the convolution and adapts the value to the specific length. This convolution block consists of two convolution sets. Each set includes a convolutional layer, a batch normalization layer, and an activation function. The convolution structure is shown in Figure 3.
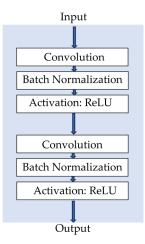

Figure 3. Convolution block.

2.2.2. Convolution Loop

The convolution loop proposes learning of features in different feature dimensions. This algorithm consists of a convolutional layer, a max-pooling layer, and an upsample layer. The structure is shown in Figure 4. In this process, the input is dumped to the convolutional layer. The output is sent to the max-pooling algorithm to reduce the feature size. The convolutional layers have increasing filter banks that are applied for learning ability. The upsampling algorithm is used for the upsampled feature map to achieve the target feature size. The concatenation function is combined between the first convolutional layer and the upsampling layer. The output is sent to the last convolutional layer in the convolution loop. The result of the last convolution layer is the output of the convolution loop.
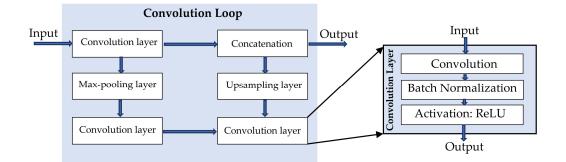
Figure 4. Convolution loop.

2.2.3. The Proposed Network Architecture

The proposed network architecture (LoopNet) aims to segment land use features over medium spatial resolution imagery. The network inherits the three-level modules to learn the multidimensional features and segment the complexity of land use types. The network architecture consists of four algorithms: the convolutional block, convolutional loop, max-pooling algorithm, and upsampling algorithm, as shown in Figure 5. The input data which has 480 × 480 pixels of feature sizes, feeds to the first convolutional layer and convolutes with a set of filter banks. The output moves to the convolution block in the first-level module. The convolution block learns and detects the object's features using the convolution algorithm. Later, the output of this block feeds to the convolution loop, which is designed to improve the learning ability in different feature dimensions using the max-pooling feature algorithm and optimizes the residual learning using the concatenation function. For the second-level module, the output of the first convolutional layer feeds to the max-pooling algorithm in order to compress the feature map to 240 × 240 pixels. This compression reduces the learning parameters and memory requirements. Convolution loops are applied to the learning features. This level consists of three convolutional loops, one concatenation function, and one upsampling algorithm. Due to the compression of the feature maps, the model can increase the number of filter banks to improve the learning ability of the object features. Later, in the third-level module, the model comprises two convolutional loops, a max-pooling algorithm, and an upsampling algorithm. The output of the max-pooling algorithm in the second-level module feeds to the first layer of this module. Then, the max-pooling algorithm is applied by resizing the object features to 120 × 120 pixels. The output feeds to the convolutional loop to learn the object features. The number of filter banks are increased in this convolution layer. The last layer of the third-level module is an upsampling algorithm that extends the feature map to double its size. Then, the feature map reverts to 240 × 240 pixels. The output of the third-level module is concatenated with the output of the convolutional loop of the second-level module. This method helps the model to optimize learning features when they are lost in the deep convolutional network. For the last layer in the second-level module, the feature map is applied to an upsampling algorithm in order to enlarge the feature map to 480 × 480 pixels. The output is concatenated with the output of the convolutional loop in the first level module. These feature maps pass through two convolution blocks to the last layer of this proposed network architecture. The softmax classifier calculates the class probabilities of the land use types over medium spatial resolution imageries.
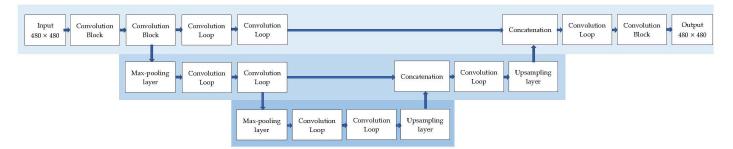
Figure 5. LoopNet network architecture.

2.3. Training

The supervised learning model for deep learning semantic segmentation was applied using the proposed network architecture and state-of-the-art network architectures. The weight training in the convolution layers was optimized by the Adam optimization function in the backpropagation algorithm. The best hyperparameters of each supervised learning model, the hyperparameters of SegNet, PSPnet, and DeeplabV3+, were set as 0.001 for learning rate, 0.999 for momentum, and 0.0001 for weight decay. For U-Net, ResU-Net, and U-Net++, the hyperparameters were set as 0.0001 for the learning rate, 0.99 for momentum, and 0.0001 for weight decay. The LoopNet set the hyperparameters as 0.0001 for learning rate, 0.999 for momentum, and 0.00001 for weight decay. Every network set the step size of learning at every 1000 iterations. The number of epochs was 100, and the batch size was defined as 10. Batch normalization and the dropout function were used in the model in order to prevent overfitting and underfitting. The early step techniques and L2 regularized logistic regression were used for the cost function. Weight balancing was applied to the loss function to ensure that all classes contributed equally to solving the unbalanced training data. The deep learning algorithm was implemented based on Tensorflow library version 2.4 with Python version 3.9 over Chalawan high-performance computing provided by the NARIT organization in Thailand.

2.4. Evaluation Metrics

Quantitative accuracy metrics such as overall accuracy, mean intersection over union, per class intersection over union, precision, and recall, were used to evaluate the performance of the supervised learning model. The accuracy assessment was conducted in three steps. The first step was to assess the learning process during the training model. The second step was to test the learning procedure during its iteration validating the sample sets. The third step evaluated the performance of the supervised learning model with a test sample set. The overall accuracy shows the performance of the supervised learning model. It calculates the sum of the true positive and true negative segmentations divided by the sum of the true positive, true negative, false positive, and false negative segmentations. The mean intersection over union (mIoU) presents the accuracy between the segmentation map and ground truth. It evaluates the ratio between the intersection of all correctly segmented pixels divided by the union of all correctly segmented pixels and all falsely segmented pixels. The per-class intersect over union computes the ratio of the target class between the intersection of correctly segmented pixels divided by the union of correctly segmented pixels and all falsely segmented pixels. Precision calculates the ratio of true positive value divided by the sum of the true positive values and false positive values. Recall measures the ratio of true positive values divided by the sum of true positive values and false negative values.

3. Experiments and Analysis

3.1. Experiment Designs

Following automatic land use segmentation by deep learning, this research creates a land use dataset from Landsat 8 imagery and implements the proposed network architecture (LoopNet) for a supervised learning model. In order to overcome the performance constraints of land use segmentation, the evaluation is set into three experiments. The first experiment assesses the performance of deep learning semantic segmentation and pixel-based classification by machine learning algorithms. The second experiment compares the performance of the proposed network with other state-of-the-art network architectures, including SegNet, U-Net, ResU-Net, U-Net++, PSPnet, and DeeplabV3+. Several performance metrics are used for these comparisons. The third experiment evaluates the performance of the land use dataset. Multiple spectral bands (Band 1 to Band 7) are evaluated for land use segmentation.

3.2. Experiment 1: Comparison of Pixel-Based Classification and Deep Learning Semantic Segmentation for Land Use Classification

This experiment compares the land use classification performance of pixel-based machine learning and semantic segmentation via deep learning. Table 2 lists the accuracy assessment and Figure 6 shows the extraction results. The results of the pixel-based machine learning algorithms show that the random forest algorithm achieves 74.69% overall accuracy and 26.14% mIoU. This performance is higher than the support vector machine algorithm and maximum likelihood estimation algorithm by about 1.83% and 6.57%, respectively, in terms of overall accuracy. In contrast, deep learning semantic segmentation proposes land use segmentation. Baseline networks such as SegNet and U-Net are applied in this evaluation. The results illustrate that the supervised learning model using SegNet architecture has 84.13% overall accuracy and 54.20% mIoU, which is higher accuracy than the U-Net architecture. As a result, the comparison between pixel-based machine learning and deep learning semantic segmentation demonstrates that the deep learning algorithm is superior to pixel-based classification. The overall accuracy, mIoU, precision, and recall present higher accuracy results than the pixel-based methods. In terms of per-class IoU accuracy, the highest performance for the deep learning algorithm is SegNet, with 66.55% in agriculture, 61.05% in forest, 30.27% in miscellaneous, 61.53% in urban, and 53.74% in water. These accuracy levels are higher than the random forest method, which has 73.65% in agriculture, 55.56% in forest, 18.47% in miscellaneous, 33.64% in urban, and 69.72% in water.

Table 2. The accuracy result (%) of pixel-based classification and deep learning semantic segmentation on the land use dataset. The highest values are highlighted in bold.
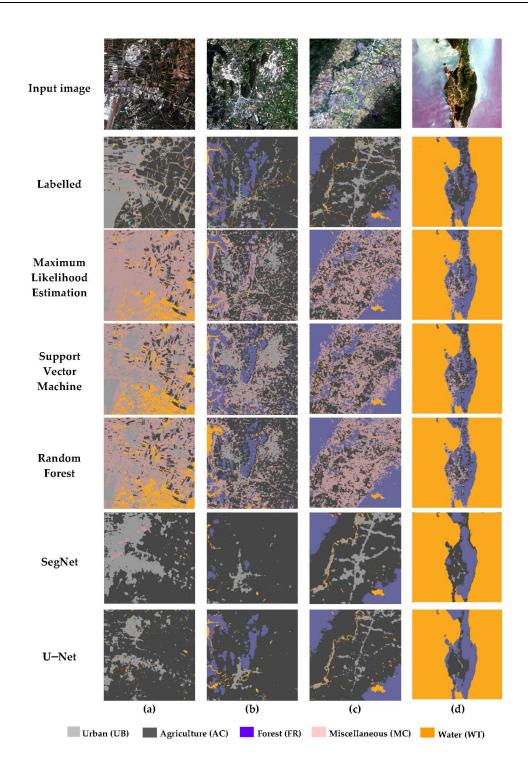
Method	Per Class IoU						Precision	Recall	mIoU	OA
	BG	AC	FR	MC	UB	WT				
Maximum likelihood estimation	71.89	52.12	41.13	15.45	30.23	67.94	52.23	53.37	21.23	68.12
Support Vector Machine	73.76	58.45	48.09	19.54	36.54	**70.98**	55.61	53.76	24.67	72.86
Random Forest	73.65	61.11	55.56	18.47	33.64	69.72	56.85	54.21	26.14	74.69
SegNet	87.39	66.55	**61.05**	30.27	**61.53**	53.74	**84.18**	**84.10**	54.20	**84.13**
U-Net	**88.33**	**67.57**	58.85	**30.75**	52.75	57.55	81.94	81.92	**55.09**	81.93

Figure 6. The classification results and segmentation results of land use features on the land use dataset. (**a**) urban and agriculture area; (**b**) urban, agriculture, and forest areas in countryside; (**c**) urban and agriculture areas in valley; (**d**) urban, agriculture and forest areas on island.

The comparison of classification images between pixel-based machine learning and deep learning semantic segmentation illustrates that the deep learning semantic segmentation has better performance than the pixel-based method. For the maximum likelihood estimation algorithm, the classification results present the lowest performance among pixel-based methods, showing confusion in classification between urban and miscellaneous classes as well as classification errors in the miscellaneous and water classes, as shown in Figure 6a,c (Maximum Likelihood Estimation). Furthermore, the support vector machine algorithm has fair performance in the pixel-based methods. There are classification

errors between the agriculture and water classes, as shown in Figure 6a (Support Vector machine). It lacks classification ability in the forest class, as shown in Figure 6b,c (Support Vector Machine). The random forest algorithm has better performance than the previous classification algorithms for land use classification, although there are classification errors in the agriculture area which misclassify the water class, as shown in Figure 6a (Random Forest). Furthermore, Figure 6c (Random Forest) illustrates a lack of classification in the agriculture and miscellaneous classes. Regarding the performance of pixel-based classification, it can be seen that there are a lot of errors in classification, such as pepper noise and classification errors. However, deep learning semantic segmentation, which is a feature learning algorithm, can detect and extract the land use types. Segmentation with baseline network architectures such as SegNet and U-Net presents fairly good results for land use segmentation, as shown in Figure 6a–d (SegNet, U-Net). The SegNet architecture is better at detecting and extracting the urban and agricultural land classes; however, it is lacking for the miscellaneous and forest classes, and omits a narrow river, as shown in Figure 6a–c (SegNet). The U-Net architecture performs better than SegNet in the forest and water classes, although it has errors in the urban and miscellaneous classes, as shown in Figure 6a–c (U-Net). The results in Figure 6 show that the deep learning semantic segmentation approach is superior to pixel-based classification for land use classification.

3.3. Experiment 2: Evaluation of the Proposed Network and State-of-the-Art Network Architecture for Land Use Segmentation

This experiment evaluates the performance of the proposed network architecture (LoopNet) with state-of-the-art network architectures for land use extraction using Landsat 8 imagery. Table 3 lists the accuracy assessment and Figure 7 shows the extraction results. The best performance for land use segmentation is LoopNet, which has 89.84% overall accuracy. The overall accuracy of the proposed network is higher than other network architectures by about 2.1% over U-Net++, 3.4% over ResU-Net, 3.62% over DeeplabV3+, 5.71% over SegNet, 5.72% over PSPnet, and 7.91% over U-Net. Regarding the precision and recall performance, LoopNet outperforms other networks with 87.88% precision and 87.81% recall, which is higher than U-Net++ by about 0.11% in precision and 0.09% in recall, DeeplabV3+ by about 0.73% in precision and 0.71% in recall, and PSPnet by about 1.63% in precision and 1.59% in recall. The lowest performance in precision and recall is SegNet, with 84.18% in precision and 84.10% in recall. For the mean intersection over union, which illustrates the overall effectiveness of feature shape extraction, the highest accuracy is found for LoopNet, with a 71.69% mIoU. The segmentation results for LoopNet are more accurate than U-Net++ by about 4.82%, DeeplabV3+ by about 5.52%, ResU-Net by about 5.94%, U-Net by about 6.6%, SegNet by about 7.49%, and PSPnet by about 9.76%. For the segmentation accuracy in each class, as listed in per class IoU, in the agriculture class, the highest IoU is LoopNet with 75.70%, which is more accurate than U-Net++ by about 1.95% and DeeplabV3+ by about 3.57%, while the lowest IoU is SegNet with 66.55%. For the forest class, the highest IoU is LoopNet with 73.41%, which is higher than U-Net++ by about 1.96%, DeeplabV3+ by about 3.63%, and ResU-Net by about 4.47%. The lowest performance in the forest class is U-Net, with 58.85% of IoU. For the miscellaneous class, the best performance is LoopNet, with 40.19%, followed by ResU-Net with 37.21% and U-Net++ with 35.58%. The lowest performance is U-Net, which has an IoU of 30.75%. Regarding the IoU for the urban class, the highest accuracy results are DeeplabV3+ with 46.60%, LoopNet with 43.12%, and U-Net++ with 42.07%. The lowest accuracy results are PSPnet with 28.30% and U-Net with 22.75%. For water segmentation, the highest segmentation performance is seen for LoopNet with 65.95%, which is higher than U-Net++ by about 4.14% and DeeplabV3+ by about 5.17%. The lowest performance on water segmentation is for SegNet, with 53.74%. Regarding the results of per-class IoU, it can be seen that the LoopNet achieves the best performance in land use segmentation, achieving the highest segmentation for the background, agriculture, forest, miscellaneous, and water land use classes.
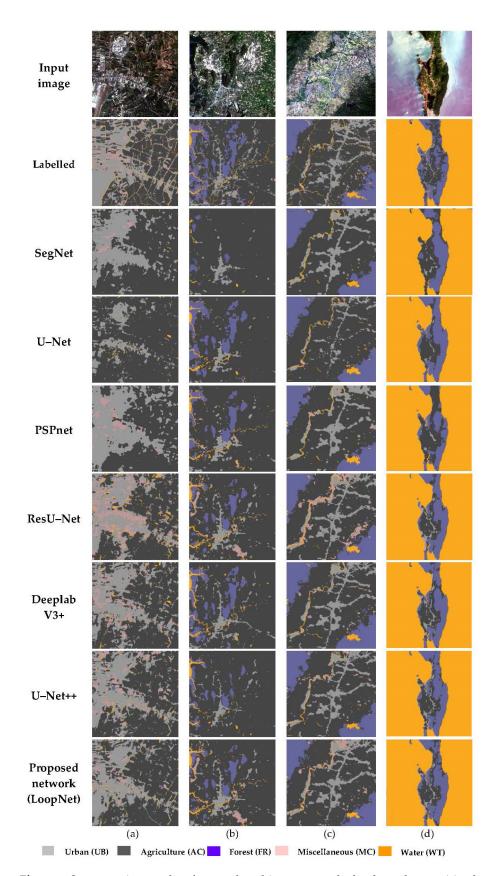
Figure 7. Segmentation results of network architectures on the land use dataset. (**a**) urban and agriculture area; (**b**) urban, agriculture, and forest areas in countryside; (**c**) urban and agriculture areas in valley; (**d**) urban, agriculture and forest areas on island.

Table 3. Accuracy results (%) of network architectures on the land use dataset.

Method	Per Class IoU						Precision	Recall	mIoU	OA
	BG	AC	FR	MC	UB	WT				
SegNet	87.39	66.55	61.05	31.27	31.53	53.74	84.18	84.10	64.20	84.13
U-Net	88.33	67.57	58.85	30.75	22.75	57.55	81.94	81.92	65.09	81.93
PSPnet	88.26	70.86	67.53	31.89	28.30	57.66	86.25	86.22	61.93	86.22
ResU-Net	88.68	70.10	68.94	37.21	39.65	58.85	85.50	85.40	65.75	85.44
DeeplabV3+	88.29	72.13	69.78	34.65	**46.60**	60.78	87.15	87.10	66.17	85.12
U-Net++	88.28	73.75	71.45	35.58	42.07	61.71	87.77	87.72	66.87	87.74
LoopNet	**90.22**	**75.70**	**73.41**	**40.19**	43.12	**65.95**	**87.88**	**87.81**	**71.69**	**89.84**

Regarding the segmentation results, deep learning illustrates that it can detect and extract the land use type accurately. For the baseline network, SegNet, it presents an accurate segmentation, though it lacks the ability to learn and extract small object features such as small urban areas and narrow rivers, as shown in Figure 7a–c (SegNet). The other baseline network, which is U-Net-based, shows better extraction for the forest class, as shown in Figure 7b (U-Net); however, it is less accurate than SegNet for on the urban land use class, as shown in Figure 7a,c (U-Net). The PSPnet network illustrates fair segmentation capability, though it omits the small object features and has segmentation errors for urban features, as shown in Figure 7a–c (PSPnet). For the improved networks, the ResU-Net architecture performs better segmentation for land use types. However, it performs less accurately for the miscellaneous class, as shown in Figure 7a,c (ResU-Net). The U-Net++ architecture has good segmentation capability, though it has extraction errors for the river and forest areas, as shown in Figure 7b,c (U-Net++). The performance of this network is similar to the DeeplabV3+ network. Although this network can detect and extract the land use types, there are errors in small object features in the agriculture and urban classes, as shown in Figure 7a,c (DeeplabV3+). The LoopNet architecture outperforms the others in land use segmentation, achieving better segmentation than the other network architectures. LoopNet can segment the urban, agriculture, and forest classes accurately, as shown in Figure 7a–d (LoopNet). However, its performance is lacking for the miscellaneous and water classes because of the complex features of miscellaneous land and small water bodies, including rivers and ponds.

3.4. Experiment 3: Evaluation of Spectral Bands for Land Use Segmentation

This experiment evaluates the effectiveness of the multispectral bands of Landsat 8 for land use segmentation. The performance accuracy results are shown in Table 4 and illustrate the land use segmentation results in Figure 8. These spectral bands show different spectral reflectance for the object features. The comparison shows that the best spectral band for land use segmentation is Band 5, which has 83.91% overall accuracy, 83.98% precision, and 83.86% recall. It has higher overall accuracy than the 0.51% of Band 4, 0.71% of Band 3, 1.32% of Band 6, 2.05% of Band 7, 6.59% of Band 1, and 9.13% of Band 2. Regarding the segmentation results in mIoU, the highest segmentation is for Band 6 with 54.96%, which is more accurate than Band 7 by about 0.18%, Band 2 by about 0.23%, Band 3 by about 2.51%, Band 4 by about 3.22%, Band 5 by about 3.29%, and Band 1 by about 4.73%. When looking at the per class IoU for the agriculture class, the best band is Band 4 which has 67.85%. This is higher than Band 7 by about 0.92%, Band 5 by about 1.01%, Band 3 by about 1.67%, Band 6 by about 2.51%, Band 1 by about 3.68%, and Band 2 by about 5.52%. For the forest class, the highest IoU is Band 6 with 66.04%. The lowest Band 2 with 48.65%.
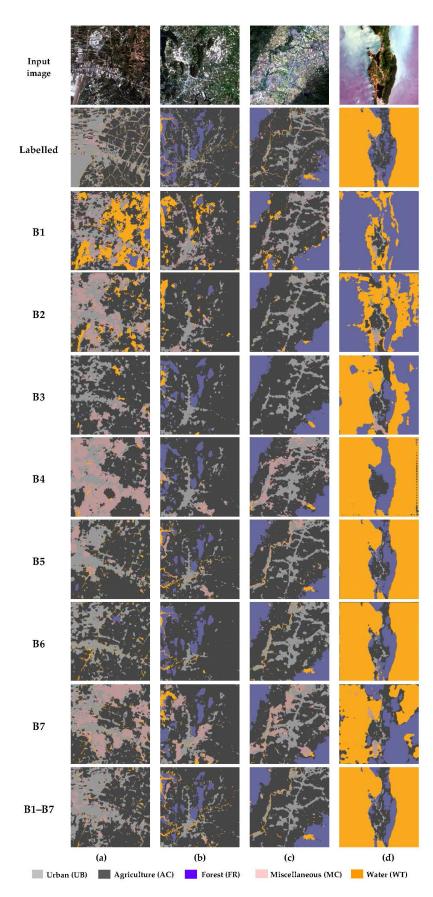
Figure 8. The segmentation results of spectral bands on the land use dataset. (**a**) urban and agriculture area; (**b**) urban, agriculture, and forest areas in countryside; (**c**) urban and agriculture areas in valley; (**d**) urban, agriculture and forest areas on island.

Table 4. The accuracy result (%) of spectral bands on the land use dataset.

Method	Per Class IoU						Precision	Recall	mIoU	OA
	BG	AC	FR	MC	UB	WT				
B1	**89.45**	64.17	49.76	28.02	48.79	38.57	77.89	76.36	50.23	77.32
B2	88.18	62.33	48.65	33.74	**62.91**	35.81	75.24	74.35	54.73	74.78
B3	87.43	66.18	61.26	31.08	54.56	46.16	83.96	**83.88**	52.45	83.12
B4	88.26	**67.85**	65.58	30.15	57.40	56.53	**84.88**	82.40	51.74	83.40
B5	88.24	66.84	63.54	32.33	54.96	**58.86**	83.98	83.86	51.67	**83.91**
B6	88.17	65.34	**66.04**	30.60	59.32	58.49	82.66	82.53	**54.96**	82.59
B7	88.23	66.93	62.77	**35.23**	62.41	50.89	81.93	81.80	54.78	81.86
B1-B7	**90.22**	**75.70**	**73.41**	**40.19**	**73.12**	**65.95**	**87.88**	**87.81**	**71.69**	**89.84**

The other results for IoU are Band 4 with 65.58%, Band 5 with 63.54%, Band 7 with 62.77%, Band 1 with 49.76%, and Band 2 with 48.65%. In the miscellaneous class, the band with the best performance is Band 7 with 35.23%, which is more accurate than Band 2 by about 1.49%, Band 5 by about 2.93%, Band 3 by about 4.15%, Band 6 by about 4.63%, Band 4 by about 5.08%, and Band 1 by about 7.21%. For the urban class, the best band is Band 2 with 62.91%, providing more accurate segmentation than Band 7 by about 0.5%, Band 6 by about 3.95%, Band 4 by about 7.95%, Band 5 by about 8.22%, Band 3 by about 8.35%, and Band 1 by about 14.12%. Lastly, for the water class the best band for water segmentation is Band 5 with 58.86%, followed by Band 6 with 58.49%, Band 4 with 56.53%, Band 7 with 50.89%, Band 3 with 46.16%, Band 1 with 38.57%, and Band 2 with 35.81%. In summary, the best bands with the highest performance in each class are Band 5 in the agriculture class, Band 6 in the forest class, Band 7 in the miscellaneous class, Band 2 in the urban class, and Band 5 in the water class. Due to the performance of each spectral band for detection and extraction of land use features, the combination of seven bands shows the highest performance, with 89% overall accuracy and 71.69% mIoU, and achieves the best accuracy in per-class IoU with 75.70% in the agriculture class, 73.41% in the forest class, 40.19% in the miscellaneous class, 73.12% in the urban class, and 65.95% in the water class.

Regarding the segmentation result, Band 1 shows segmentation errors in the water class, as shown in Figure 8a,d (B1). The water is segmented in the agriculture areas and the forest is extracted in the water areas. Regarding the visible bands, Band 2 has better segmentation results. The model can learn and extract the land use features well. Though it can reduce the segmented errors in the water class, narrow river and forest features are omitted in the segmentation results, as shown in Figure 8a–c (B2). Band 3 illustrates the segmented result accurately in the agriculture class. This band helps the model to differentiate between water and agricultural features. It provides better segmentation in water areas, as shown in Figure 8d (B3). However, urban features show more errors in this band, as shown in Figure 8a (B3). Band 4 shows accurate land use segmentation in the water class, although it has segmentation errors in the miscellaneous and urban features, as shown in Figure 8d (B4). Regarding the infrared bands, near-infrared in Band 5 provides accurate land use segmentation. The model can learn the land use features and segment the land use classes in the forest, agriculture, and water areas well, as shown in Figure 8a–d (B5). However, this band has limitations in detecting and extracting of miscellaneous features. For short-wave infrared, Band 6 has extraction errors for miscellaneous features. Although it was omitted in this learning model, this band is able to detect and segment narrow rivers, as shown in Figure 8a–c (B6). Lastly, Band 7 has limitations in terms of its ability to differentiate between the miscellaneous features and agriculture features. Its most common segmentation errors are for the miscellaneous, agricultural, water, and forest classes as shown in Figure 8a–d (B7). Furthermore, the multispectral bands show the differences in spectral reflectance and land use features, which has advantages for land use identification. To gain the advantages of these different spectral bands (Band 1–Band 7), Figure 8a–d (B1–B7) shows the better land use segmentation performance compared to

the supervised learning models trained with one band. The multi-band model can more accurately detect and extract the urban, forest, water, and agriculture classes. However, the miscellaneous class remains a limitation with regard to segmentation of complex features.

4. Discussion

Regarding the above-mentioned evaluations and comparisons, it was found that a deep learning semantic segmentation algorithm can be used to detect and extract land use features from medium spatial resolution imagery. The pixel-based machine learning algorithms (Maximum Likelihood Estimation, Support Vector Machine, Random Forest), which are statistical classification methods, consider spectral information. The per-pixel classification algorithms present fair classification results. These algorithms can classify land use features; however, they show classification errors when dealing with unclear and complex spectral features. Furthermore, the results are influenced strongly by salt-and-pepper effects. In contrast, deep learning algorithms can use the land use dataset to learn land use features and patterns from multiple perspectives. In evaluating deep learning performance, the supervised models which applied the baseline network architectures SegNet and U-Net have better results in land use segmentation using medium spatial resolution imagery. The supervised learning models are able to learn medium and complex land use features using the Landsat 8 imagery, then automatically detect and extract the land use classes accurately. The clear boundary of land use features shows good segmentation with a proper shape. However, small object features are omitted when fed through to the deep network. The unclear land use features and complexity of land cover types have an effect on extraction errors. The comparison results between pixel-based classification and semantic segmentation illustrate that deep learning semantic segmentation outperforms the pixel-based machine learning algorithms.

Due to segmentation errors in the deep steps of the network architecture when using medium spatial resolution imagery from Landsat 8, the multi-level features of the convolutional neural network may cause degradation problems, and medium resolution imagery appears with a medium level of object features. Thus, the convolution loop and convolution block in the LoopNet architecture are proposed for land use segmentation. The proposed network successfully segments the land use features using Landsat 8 imagery, showing better segmentation of land use shapes. The supervised learning model achieves 89.84% of overall accuracy and 71.69% of mIoU. Although there are extraction errors and misclassifications of land use classes, the proposed network produces better segmentation results on land use classes, with 75.70% for agriculture, 73.41% for forest, 40.19% for miscellaneous, 73.12% for urban, and 65.95% for water. The LoopNet network architecture uses a convolutional loop to its benefit in learning land use features in different feature dimensions. It solves degradation problems by feeding through the deep convolutional layers. This algorithm helps the model to segment land use features into a proper shape. The convolution block demonstrates its learning ability in the land use features. It supports the model in learning and differentiating the land use features into accurate classes. The three levels of deep steps in the convolutional neural network are designed to learn the land use features on a multidimensional level. This prevents degradation problems when the land use features are reduced to low level features in the bottom of the convolutional network. Regarding the results of land use segmentation shown in Figure 7a–d (Proposed Network), LoopNet is able to detect and extract the water, forest, and agriculture classes, though shows limitations in segmentation of the complex and unclear features of the miscellaneous, urban, and agriculture classes. Narrow river bodies are omitted in the segmentation results. Furthermore, based on the land use dataset, the performance of the proposed network architecture is evaluated and compared with baseline networks (SegNet, U-Net, PSPnet) and adjustment networks (ResU-Net, DeeplabV3+, U-Net++). Regarding the baseline networks, the supervised learning models based on the SegNet, U-Net, and PSPnet networks perform relatively well for land use segmentation,

although there are extraction errors in the water and miscellaneous land use classes. The deep stage convolutional neural network (SegNet) and pyramid scene parsing network (PSPnet) perform weakly on small and complex land use features. The U-shaped network (U-Net) lacks the ability to learn urban features. Regarding the adjustment networks, deep residual U-Net (ResU-Net) produces better segmentation results in the urban and water of land use classes, however, it over-segments the miscellaneous class. The improvement networks DeeplabV3 (DeeplabV3+) and nested U-Net (U-Net++) are able to overcome the extraction errors in the miscellaneous class. However, the performance of LoopNet exceeds that of the other network architectures in detecting and extracting the land use features accurately. It works well on medium spatial resolution imagery. The main classes, which show clear features, are segmented in accurate shapes. The limitation of this network is in its classification of the land use classes in the LULC classification system Levels 2 and 3. This is due to the low levels of differentiating features in the land use subclasses. The three levels of the deep steps are GPU-consuming, and limit the ability to learn multidimensional features.

Furthermore, based on the land use dataset, the comparison of spectral bands using the proposed network is effective at identifying land use features in different spectral wavelengths. Band 1, called the coastal/aerosol band, senses violets and deep blues. It has high contrast in shallow water and dust. The results show fair land use extraction, though it has limitations in terms of its ability to differentiate features between agriculture, forest, and deep-water areas. For visible bands (Band 2 (Blue), Band 3 (Green), and Band 4 (Red)), Band 2 measures the blue wavelength and enhances the water features. The extraction results provide better segmentation in the urban class. However, it has weaker performance in classifying similar texture features in the agriculture and forest classes, as well as in differentiating deep-water areas, which are the same errors as seen in Band 1. This is due to statistical learning between forest and water features. Band 3 presents the green wavelength, and emphasizes vegetation features for assessing plant vigor. This band improves learning ability for differentiating green features, and shows good segmentation in the agriculture and forest classes. Band 4 collects the red wavelength, and emphasizes vegetation slopes and plant chlorophyll status. It differentiates abundant plant features and dry areas. The results overcome deep water extraction issues, and provide more accurate segmentation results in the agriculture and forest classes. However, it has limitations in terms of its ability to detect and extract the complexity of miscellaneous features. For the infrared bands (Band 5 (near-infrared), Band 6 (SWIR), and Band 7 (SWIR)), Band 5 measures the near-infrared wavelength. This band contrasts the features between vegetation, water boundaries, and landforms. The results illustrate better segmentation of land use classes over other bands, especially for the water class. It is more accurate for the land use shape. However, it lacks narrow river extraction. Band 6 and Band 7 sense the shortwave infrared wavelength. They provide good presentation of wet and dry areas, plant drought stress, and strong contrasts between rocks and soils, and can delineate burnt areas. These bands show good land use extraction in forest and urban classes, though they are less accurate than Band 5 in differentiating water features, and there are more extraction errors in the miscellaneous and water classes. As a result, it can be seen that each spectral band presents different spectral reflectance for each land use feature. The segmentation performs land use extraction well. Each spectral band reflects an accurate segmentation result in each class. The performance of the supervised learning model, which applies all spectral bands, shows good segmentation for all the land use classes. The model takes advantage of different spectral reflectance to detect and extract the land use features accurately. However, identifying small and complex land use features is a limitation of deep learning semantic segmentation using medium spatial resolution imagery. To overcome these segmentation errors, the complexity of land use features in the land use dataset and the number of filter banks in the model can be increased.

According to these comparisons, deep learning semantic segmentation using the LoopNet architecture and land use dataset provides good the performance in learning

land use features from Landsat 8 imagery. Moreover, the supervised learning model is superior to other network architectures in detecting and extracting land use classes, as shown in Table 3 and Figure 7. The supervised learning model which applies all spectral bands raises the efficiency of accuracy assessment and segmentation results when using the land use dataset, as demonstrated in Table 4 and Figure 8.

5. Conclusions

This research proposed deep learning semantic segmentation to detect and extract land use features using medium spatial resolution imagery from the Landsat 8 satellite. The land use dataset, which has six main classes (background, agriculture, forest, miscellaneous, urban, and water) is intended for learning land use patterns and features. The LoopNet architecture, which comprises a convolutional loop and convolutional block, is implemented for the supervised learning model. Multiple spectral bands are evaluated for their performance to differentiate land use types. The experimental results show that the performance of deep learning semantic segmentation outperforms pixel-based machine learning. The proposed model can detect and extract land use features accurately. Furthermore, the proposed network architecture, LoopNet, demonstrates better performance than the baseline network and improvement network architectures, as confirmed by both the accuracy assessments and segmentation results. The LoopNet architecture is better at segmenting land use features. The supervised learning model can differentiate between urban and miscellaneous features, and can segment a narrow river accurately. Regarding the land use dataset, the performance of the multiple spectral bands illustrates that Band 5, which captures data in the near-infrared wavelength, shows better segmentation results over other bands. Moreover, the combination of multiple spectral bands (Band 1–Band 7) as a sample set achieves high performance, with 89.84% overall accuracy and 71.69% mIoU. For this reason, it is concluded that the deep learning algorithm applying the LoopNet architecture and the land use dataset can be used by the supervised learning model to detect and extract land use features using medium spatial resolution imagery. For future works, the number of data points with various land use features in the land use dataset can be increased, and time-series data could be applied to enhance the multiple perspective views of land use patterns.

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

References

1. Van Leeuwen, B.; Tobak, Z.; Kovács, F. Machine learning techniques for land use/land cover classification of medium resolution optical satellite imagery focusing on temporary inundated areas. *J. Environ. Geogr.* **2020**, *13*, 43–52.
2. Wang, J.; Bretz, M.; Dewan, M.A.A.; Delavar, M.A. Machine learning in modelling land-use and land cover-change (LULCC): Current status, challenges and prospects. *Sci. Total Environ.* **2022**, *822*, 153559.
3. Alademomi, A.S.; Okolie, C.J.; Daramola, O.E.; Akinnusi, S.A.; Adediran, E.; Olanrewaju, H.O.; Alabi, A.O.; Salami, T.J.; Odumosu, J. The interrelationship between LST, NDVI, NDBI, and land cover change in a section of Lagos metropolis, Nigeria. *Appl. Geomat.* **2022**, *14*, 299–314.
4. Fathizad, H.; Tazeh, M.; Kalantari, S.; Shojaei, S. The investigation of spatiotemporal variations of land surface temperature based on land use changes using NDVI in southwest of Iran. *J. Afr. Earth Sci.* **2017**, *134*, 249–256.
5. Macarringue, L.S.; Bolfe, E.; Pereira, P.R.M. Developments in land use and land cover classification techniques in remote sensing: A review. **2022**, *14*, 1–28.
6. Senf, C.; Hostert, P.; van der Linden, S. Using MODIS time series and random forests classification for mapping land use in South-East Asia. In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 July 2012; pp. 6733–6736.
7. Anderson, J.R. *A Land Use and Land Cover Classification System for Use with Remote Sensor Data*; US Government Printing Office: Washington, D.C., USA, 1976; Volume 964.
8. Qu, L.a.; Chen, Z.; Li, M.; Zhi, J.; Wang, H. Accuracy improvements to pixel-based and object-based lulc classification with auxiliary datasets from Google Earth engine. *Remote Sens.* **2021**, *13*, 453.
9. Amini, S.; Saber, M.; Rabiei-Dastjerdi, H.; Homayouni, S. Urban Land Use and Land Cover Change Analysis Using Random Forest Classification of Landsat Time Series. *Remote Sens.* **2022**, *14*, 2654.
10. Duro, D.C.; Franklin, S.E.; Dubé, M.G. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sens. Environ.* **2012**, *118*, 259–272.
11. Li, M.; Stein, A. Mapping land use from high resolution satellite images by exploiting the spatial arrangement of land cover objects. *Remote Sens.* **2020**, *12*, 4158.
12. Jamali, A. Evaluation and comparison of eight machine learning models in land use/land cover mapping using Landsat 8 OLI: A case study of the northern region of Iran. *SN Appl. Sci.* **2019**, *1*, 1448.
13. Mayani, M.B.; Itagi, R. Machine Learning Techniques in Land Cover Classification using Remote Sensing Data. In Proceedings of the 2021 International Conference on Intelligent Technologies (CONIT), Hubli, India, 25–27 June 2021; pp. 1–5.
14. 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.
15. Song, J.; Gao, S.; Zhu, Y.; Ma, C. A survey of remote sensing image classification based on CNNs. *Big Earth Data* **2019**, *3*, 232–254.
16. Alem, A.; Kumar, S. Deep learning methods for land cover and land use classification in remote sensing: A review. In Proceedings of the 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO), Noida, India, 4–5 June 2020; pp. 903–908.
17. Vali, A.; Comai, S.; Matteucci, M. Deep learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: A review. *Remote Sens.* **2020**, *12*, 2495.
18. Kussul, N.; Lavreniuk, M.; Skakun, S.; Shelestov, A. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 778–782.
19. Storie, C.D.; Henry, C.J. Deep learning neural networks for land use land cover mapping. In Proceedings of the IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 3445–3448.
20. Alhassan, V.; Henry, C.; Ramanna, S.; Storie, C. A deep learning framework for land-use/land-cover mapping and analysis using multispectral satellite imagery. *Neural Comput. Appl.* **2020**, *32*, 8529–8544.
21. Gharbia, R.; Khalifa, N.E.M.; Hassanien, A.E. Land cover classification using deep convolutional neural networks. In Proceedings of the International Conference on Intelligent Systems Design and Applications, Online, 12–15 December 2020; pp. 911–920.
22. Sun, Z.; Di, L.; Fang, H.; Burgess, A. Deep learning classification for crop types in north dakota. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 2200–2213.
23. Zhang, C.; Sargent, I.; Pan, X.; Li, H.; Gardiner, A.; Hare, J.; Atkinson, P.M. Joint Deep Learning for land cover and land use classification. *Remote Sens. Environ.* **2019**, *221*, 173–187.
24. Rousset, G.; Despinoy, M.; Schindler, K.; Mangeas, M. Assessment of deep learning techniques for land use land cover classification in southern new Caledonia. *Remote Sens.* **2021**, *13*, 2257.
25. Helber, P.; Bischke, B.; Dengel, A.; Borth, D. Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 2217–2226.

26. Zhang, Z.; Cui, X.; Zheng, Q.; Cao, J. Land use classification of remote sensing images based on convolution neural network. *Arab. J. Geosci.* **2021**, *14*, 1–6.

27. Kotaridis, I.; Lazaridou, M. Remote sensing image segmentation advances: A meta-analysis. *ISPRS J. Photogramm. Remote Sens.* **2021**, *173*, 309–322.

28. Poliyapram, V.; Imamoglu, N.; Nakamura, R. Deep learning model for water/ice/land classification using large-scale medium resolution satellite images. In Proceedings of the IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 3884–3887.

29. Mutreja, G.; Kumar, S.; Jha, D.; Singh, A.; Singh, R. Identifying Settlements Using SVM and U-Net. In Proceedings of the IGARSS 2020–2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; pp. 1217–1220.

30. Du, M.; Huang, J.; Chai, D.; Lin, T.; Wei, P. Classification and Mapping of Paddy Rice using Multi-temporal Landsat Data with a Deep Semantic Segmentation Model. In Proceedings of the 2021 9th International Conference on Agro-Geoinformatics (Agro-Geoinformatics), Shenzhen, China, 26–29 July 2021; pp. 1–6.

31. de Almeida Pereira, G.H.; Fusioka, A.M.; Nassu, B.T.; Minetto, R. Active fire detection in Landsat-8 imagery: A large-scale dataset and a deep-learning study. *ISPRS J. Photogramm. Remote Sens.* **2021**, *178*, 171–186.

32. Wang, L.; Wang, J.; Liu, Z.; Zhu, J.; Qin, F. Evaluation of a deep-learning model for multispectral remote sensing of land use and crop classification. *Crop J.* **2022**, *10*, 1435–1451.

33. department, L.D. Land Use 2562–2564. Available online: http://www1.ldd.go.th/web_OLP/result/landuse2562-2564.htm (accessed on 4 October 2022).