applied sciences

MDPI

Article

Integration of Object-Based Image Analysis and Convolutional Neural Network for the Classification of High-Resolution Satellite Image: A Comparative Assessment

Omer Saud Azeez [1], Helmi Z. M. Shafri [1,2,*], Aidi Hizami Alias [1] and Nuzul A. B. Haron [1]

[1] Department of Civil Engineering, Faculty of Engineering, University Putra Malaysia (UPM), Serdang 43400, Selangor, Malaysia
[2] Geospatial Information Science Research Centre (GISRC), Faculty of Engineering, Universiti Putra Malaysia (UPM), Serdang 43400, Selangor, Malaysia
* Correspondence: helmi@upm.edu.my or hzms2312@gmail.com

Abstract: During the past decade, deep learning-based classification methods (e.g., convolutional neural networks—CNN) have demonstrated great success in a variety of vision tasks, including satellite image classification. Deep learning methods, on the other hand, do not preserve the precise edges of the targets of interest and do not extract geometric features such as shape and area. Previous research has attempted to address such issues by combining deep learning with methods such as object-based image analysis (OBIA). Nonetheless, the question of how to integrate those methods into a single framework in such a way that the benefits of each method complement each other remains. To that end, this study compared four integration frameworks in terms of accuracy, namely OBIA artificial neural network (OBIA ANN), feature fusion, decision fusion, and patch filtering, according to the results. Patch filtering achieved 0.917 OA, whereas decision fusion and feature fusion achieved 0.862 OA and 0.860 OA, respectively. The integration of CNN and OBIA can improve classification accuracy; however, the integration framework plays a significant role in this. Future research should focus on optimizing the existing CNN and OBIA frameworks in terms of architecture, as well as investigate how CNN models should use OBIA outputs for feature extraction and classification of remotely sensed images.

Keywords: deep learning; convolutional neural networks; Object-Based Image Analysis; remote sensing; integration frameworks

check for updates

Citation: Azeez, O.S.; Shafri, H.Z.M.; Alias, A.H.; Haron, N.A. Integration of Object-Based Image Analysis and Convolutional Neural Network for the Classification of High-Resolution Satellite Image: A Comparative Assessment. *Appl. Sci.* **2022**, *12*, 10890. https://doi.org/10.3390/app122110890

Academic Editors: Tan-Hsu Tan, Mohammad Alkhaleefah and Yang-Lang Chang

Received: 14 September 2022
Accepted: 21 October 2022
Published: 27 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Image classification is one of the basic operations in remote sensing. In this operation, image pixels or image objects are clustered and labeled via automatic-learned features or hand-crafted samples. In general, several methods and techniques are employed for remote sensing image classification. In the last decade, deep learning techniques have attracted experts in the field of remote sensing due to their sophisticated architecture, including several learning steps and layers such as convolutional neural networks (CNNs). Deep learning methods, with their hierarchical architecture, learn deep and abstract image features, which are useful for the classification task. Pixel-based classification using ANN lacks efficient learning of spatial and contextual relationships among the image pixels and can lead to an extremely redundant computation. On the other hand, CNNs have outperformed most deep learning methods in terms of computer vision as well as pattern recognition. CNNs have revealed high efficiency for learning spatial, contextual, and textural information from remotely sensed images. However, the patch-wise CNN obtains artifacts on the boundary of the classified patches and usually results in blurred boundaries among ground surface objects [1,2] (Zhang et al. 2017a, 2017b). Consequently, introduces uncertainty during the image classification.

CNNs have excellent capabilities in extracting useful abstract features from remote sensing images; however, they are not directly applicable for performing classification at the object level. CNNs require image patches as input data to make inferences on a pixel's category label. This data structure is prone to errors when classifying the boundaries of ground objects [3] (Pan and Zhao 2019). Furthermore, the rectangular patch of CNNs is scarcely consistent with the non-homogeneous segment obtained from the OBIA process. Therefore, when applying CNN classifier on a segment at the boundary of a specified land cover, the CNN patch that includes the segment that is to be classified may also contain other segments belonging to a different type of land cover. In addition, these segments that belonged to other land cover categories may exhibit higher band values or notable patterns and texture. As a consequence, they will significantly influence the decision of CNNs. As a result of this inconsistency, classification errors will occur when we directly train the CNNs model on segments resulting from the OBIA method. This kind of error is called a jagged error, which will be seen between land cover borders as a shrink or over-expand forms in the resulting image. This issue has severely restricted the application of CNNs in OBIA classification. Generally, patch-based or pixel-level CNNs methods are not effectively designed respecting computational efficiency or precision to handle the complicated challenges of the classification of remotely sensed images [2] (Zhang 2018).

Unlike using deep learning on image pixels or patches, its use with OBIA is relatively more complex because of the data transformation requirements to specific data structures that deep learning models can be trained on. The literature contains four main frameworks that can be used to combine deep learning and OBIA, including (1) extracting deep and abstract features from OBIA attributes with deep learning, (2) OBIA-deep learning features fusion, (3) decision level fusion, and (4) heterogeneous segments filtering. While there are many research papers on the integration of CNNs and OBIA for remote sensing image classification, there is a significant lack of understanding of how these integration frameworks compare. Thus, this research aims to present a comparison of four common frameworks used to combine CNNs and OBIA for image classification. Our contribution to this research is summarized in some points;

- To the best of our knowledge, this is the first study that compares the common integration frameworks and provides an assessment of each framework by using a high-resolution satellite image dataset (worldview3) that has training and test areas, which could be a guideline for researchers in the field of the integration between deep learning and OBIA methods;
- A custom-made computational framework is developed in Python programming to combine the two main image processing frameworks i.e., (1) OBIA and (2) deep learning, which was necessary to avoid using multiple software that is considered complex and time-consuming.

The remainder of this paper presents related studies (Section 2), a methodology that includes research data, assessed integration frameworks, and the details of each method (Section 3), results and discussions on the main findings (Section 4), and conclusions and recommendations for future works (Section 5).

2. Previous Works

The idea behind integrating OBIA and deep learning for remote sensing image classification is to improve the accuracy and quality of the classification maps. Previous studies have indicated that such integration can improve classification results compared to any individual method [4] (Cui et al. 2018). Several integration frameworks have been proposed to combine OBIA with deep learning. This research groups those methods into four categories, including (1) training deep learning models on OBIA features, (2) OBIA-deep learning features fusion, (3) decision level fusion, and (4) heterogeneous patch filtering. Table 1 summarizes OBIA-CNN integration methods.

Table 1. The summary of OBIA-CNN integration methods.

OBIA-CNN Integration Method	References	Dataset	Pros	Cons
Training Deep Learning Models on OBIA Features	Jozdani et al. (2019) [5] Abdollahi et al. (2020) [6], Lam et al. (2020) [7]	High-resolution satellite image, aerial photo, unmanned aerial vehicle	It is considered a simple method that easily learns contextual relationships among the OBIA attributes.	This method lacks learning spatial characteristics of the image content (pixels and objects). As well as, it neglects the powerful capability of deep learning for extracting spatial and abstract features from the image data.
OBIA-Deep Learning Features Fusion	Zhao et al. (2017) [8], Majd et al. (2019) [9], Sutha et al. (2020) [10], Hong et al. (2020) [11], Tang et al. (2020) [12]	High-resolution satellite image, hyperspectral image	This method presents an advanced framework that utilized additional deep features that offer advantages over the first method.	This method requires heavy computations and hardware resources. On the other hand, it includes a duplication in some features extracted by OBIA and CNN, such as shape, texture, and color.
Decision Level Fusion	Guirado et al. (2021) [13] Lv et al. (2018) [14], Liu et al. (2019) [15], Abdi et al. (2018) [16], Robson et al. (2020) [17], Timilsina et al. (2020) [18], He et al. (2020) [19]	High-resolution satellite image, sentinel-2, LiDAR	It is a simple method in terms of construction and application.	This method does not fully utilize the OBIA method as no features are used for classification.
Heterogeneous Patch Filtering	Pan et al. (2019) [3], Fu et al. (2019) [20], Liu et al. (2018) [21], Liu et al. (2018) [22], Fu et al. (2019) [23], Ji et al. (2019) [24], Wang et al. (2020) [25]	High-resolution satellite image, unmanned aerial vehicle	It aims to improve the model's ability to classify the precise edge of ground objects correctly with some filtering methods that can be applied to image patches or image objects.	The challenge remains to best map image objects into image patches.

2.1. Training Deep Learning Models on OBIA Features

The first method uses any deep learning model such as CNN as a feature extractor to extract deep and abstract features from OBIA attributes. In other words, deep learning is applied to tabular data that contain information about the image segments and their related attributes. This method learns contextual relationships among the OBIA attributes. However, it lacks learning spatial characteristics of the image pixels and image objects. It also neglects the powerful capability of deep learning for extracting spatial and abstract features from the image data. Several studies have used this type of OBIA-CNN integration. Jozdani et al. (2019) [5] showed that such integrated models could outperform traditional machine learning methods for urban land cover classification in the United States. Integrated OBIA and CNN were also used by Abdollahi et al. (2020) [6] for road detection in orthophoto images. They used principal component analysis (PCA) to reduce the computation time of the model. In another study, Lam et al. (2020) [7] presented an integrated OBIA-CNN model for weed species identification and detection in a challenging grassland environment. They demonstrated the potential of such models for a semi- and full classification of weed species. The studies above indicate that such an integration technique could improve upon traditional machine learning or any of the OBIA or CNN

methods. However, deep learning models, especially CNN and its variants, have proven not suitable for tabular data because the spatial arrangement of the image objects is not considered in the modeling process.

2.2. OBIA-Deep Learning Features Fusion

Feature fusion is another approach to integrating OBIA and deep learning. This method combines attributes of OBIA and deep features of deep learning after extracting each feature set separately. The additional deep features utilized in this technique offer advantages over the first method. This approach is often implemented as a two-branch computational network, which contains a processing chain to perform segmentation and OBIA feature extraction and a network to learn deep and abstract features from the data. After combining the two feature sets, a classifier such as tree-based models or support vector machines (SVM) is used to obtain class labels for the image pixels [8,9] (Zhao 2017, Majd 2019). Sutha et al. (2020) [10] combined SVM and CNN to perform the classification of high-resolution remote sensing images, aiming to improve classification accuracy. Hong et al. (2020) [11] used the common multiscale segmentation algorithm to extract multiscale low-level image features and a CNN to obtain deep features from the low-level features at each scale, respectively. An approach to extract tree plantations from very high-resolution remote sensing images was proposed by Tang et al. (2020) [12]. They used an integrated OBIA-CNN framework to achieve that. They performed image segmentation to obtain OBIA features and a fine-tuned CNN to obtain deep features. To reduce the computation time of the model, they conducted feature selection based on the Gini index. The tea objects were then classified by a random forest (RF). The basic problem of this integration method is the heavy computations and the requirements of hardware resources [13] (Guirado et al. 2021). Other problems associated with this integration method include duplication in some features extracted by OBIA and CNN, such as shape, texture, and color.

2.3. Decision Level Fusion

The third technique depends on the decision of fusion or refining an initial classification map with some post-processing methods. In this integration, deep learning models such as CNNs are first applied to obtain a classification map of the area. Then, the image data is segmented using a segmentation algorithm. Finally, the classification map obtained by the CNN is refined by a majority filtering method or some other post-processing methods [14,15] (Lv 2018, Liu 2019,). Abdi et al. (2018) [16] proposed a method to refine a classification map produced by a CNN using image segmentation. They showed a significant improvement in classification accuracy over other traditional classifiers. Robson et al. (2020) [17] applied a combination of OBIA and CNN to identify rock glaciers in mountainous landscapes. Timilsina et al. (2020) [18] studied urban tree cover changes and their relationship with socioeconomic variables. They used data from satellite, Google Earth imageries, and light detection and ranging (LiDAR). In their approach, OBIA was used to refine and improve the tree heatmap obtained by a CNN. In addition, He et al. (2020) [19] incorporated the multiresolution segmentation into the classification layer of U-net and DenseNet architectures for land cover classification. They also used a voting method is applied to optimize the classification results. While studies have highlighted the significance of the decision level fusion techniques, this method does not fully utilize the OBIA method as no features are used for classification.

2.4. Heterogeneous Patch Filtering

In CNN, image patches often contain mixed land cover types, which affects the decision of the model as the output will reflect the more obvious land cover types. Studies have attempted to filter and refine the image patches or objects to achieve heterogeneous image patches or objects before classification tasks [3,20] (Pan 2019, Fu 2019). Liu et al. (2018) [21] presented a novel approach depending on multi-view unmanned aerial vehicle (UAV) orthophotos and using CNN-based adaptive patch windows and OBIA classification

to improve the classification of wetlands in the United States. Fu et al. (2018) [20] developed a model by integrating their models, namely, multiresolution segmentation, the center of gravity, and CNN. Their results improved the identification of irregular segmented objects from a very high-resolution remote sensing image, where their approach has successfully reduced the uncertainty associated with OBIA during classification in China. Liu et al. (2018) [22] compared CNN-OBIA and traditional machine learning models (i.e., SVM, ANN, and RF) in wetland classification and mapping in the United States. They found that CNN-OBIA achieved an accuracy higher than traditional models. Pan et al. (2019) [3] proposed an object-based heterogeneous filter integrated into a CNN to overcome the limitations of jagged errors at boundaries and the expansion/shrinkage of land cover areas originating from CNN-based models. Fu et al. (2019) [20] developed an approach based on the integration of CNN and OBIA with a majority overlapping region method to label the image segments. Ji et al. (2019) [23] showed that the integration of OBIA and CNN can improve image classification and change detection results compared to OBIA-based classification. Wang et al. (2020) [24] proposed adaptive patch sampling to map the object primitives into image patches along with the object primitive axes. The methods based on image patch filtering or image object filtering aim to improve the model's ability to classify the precise edge of ground objects correctly with some filtering methods that can be applied to image patches or image objects. While some studies have reported improvement in classification accuracy using this method, the challenge remains to best map image objects into image patches.

2.5. Research Gaps

Integrated deep learning and OBIA classification methods should preserve the capability of each of the individual methods. That is the powerful spatial abstract feature extraction capability of the deep learning models and the ability of OBIA methods to precise model edges of the ground objects. The methods discussed above have attempted to combine the strengths of deep learning and OBIA in single classification frameworks; however, they are still limited in taking full advantage of deep learning for feature extraction. In addition, there is no agreement on how deep learning and OBIA should be combined such that the complementary strengths of each of the individual methods are fully utilized. Future studies should therefore study the architectural design issues of integrating deep learning and OBIA for the classification of remote sensing images.

3. Data and Methodology

3.1. Training and Test Areas

The Worldview-3 (WV-3) satellite image used in this study were obtained over the Universiti Putra Malaysia (UPM) campus in Selangor, Malaysia (3°0′8.0181″ N, 101°43′1.2172″ E). The training and testing sites were chosen from the UPM site (Figure 1).

The WV-3 image was taken in November 2014 by the Digital Globe (Figure 2). The spatial resolution of the WV-3 image is 0.31 m and 1.24 m for the panchromatic band and multispectral bands, respectively. More specifically, the WV-3 image includes 8 bands as well as the panchromatic band with a radiometric resolution of 11 bits. These bands include (coastal, red, green, blue, yellow, red edge, near-infrared, and near-infrared). Digital Globe has more information on WV3's characteristics (2020). Two images have been extracted from the WV-3 image to implement the training and testing processes with a similar spatial resolution (0.31 m), where the image used for training covers 39.5 hectares, while the image used for testing covers 21 hectares.

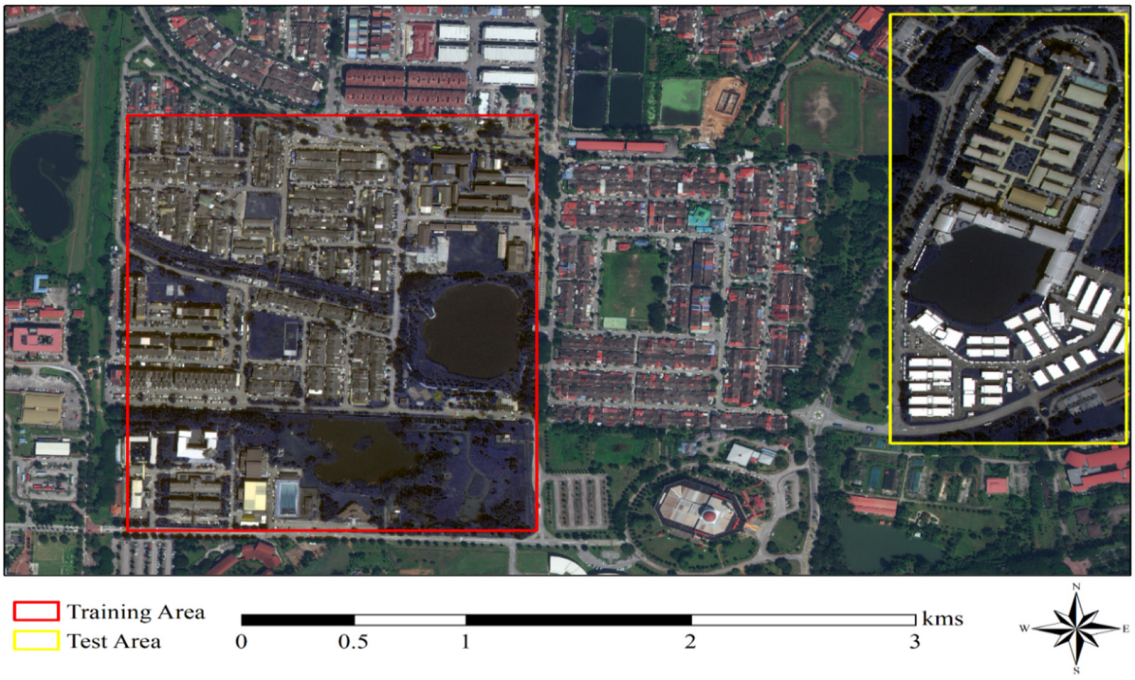


Figure 1. Location of the case study area and the training and test areas selected to develop and assess the proposed classification methods.

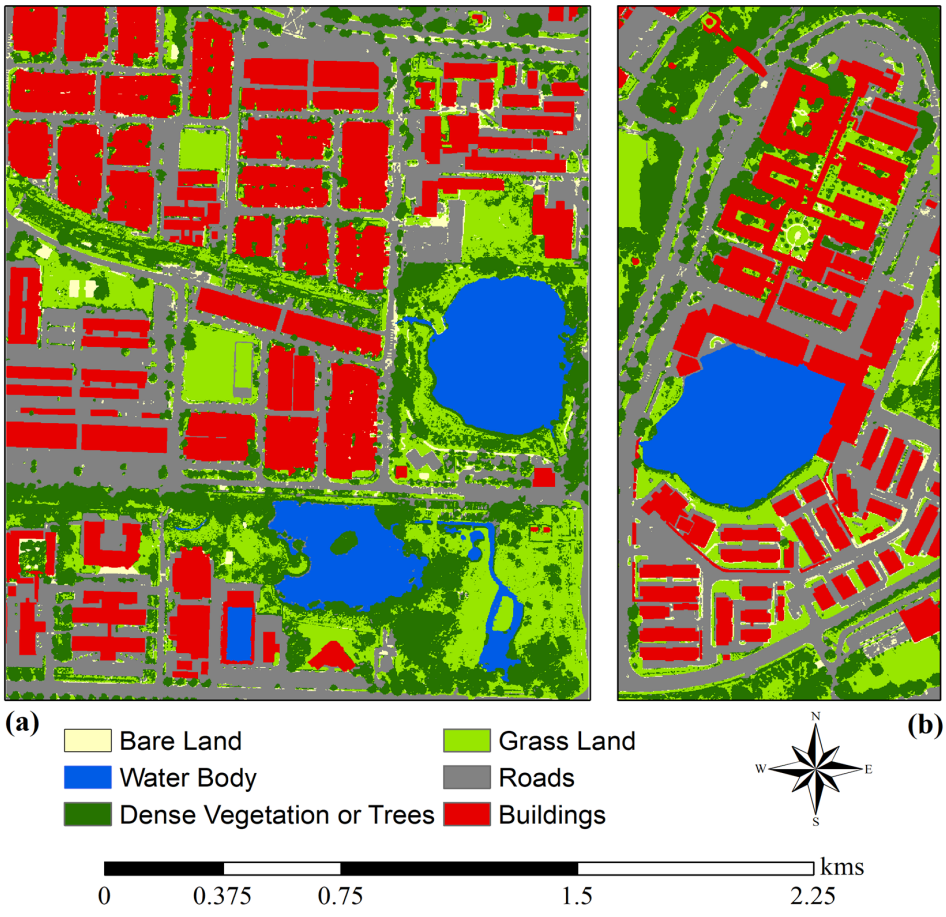


Figure 2. The ground truth data includes the main land cover types in the study area (Bare land,water body, dense vegetation or tress, grass land, roads, and buildings, (**a**) represents the training area and (**b**) represents the testing area.

The ground truth data were acquired as a land use and land cover (LULC) map in Geographic Information System (GIS) file format. The data were prepared by the Department of Survey and Mapping Malaysia (JUPEM) in 2015. The ground truth data of the training and test areas are presented in Figure 2. There are seven LULC types in the area, including grassland, road, urban/built-up, dense vegetation/trees, bare land, and water. The percentage of ground truth data used in training; grassland (16%), roads (27%), built-up area (26%), dense vegetation/trees (20%), water body (8%), and bare land (2%). On the other hand, the percentage of data used in testing grassland (14.5%), roads (31%), built-up area (24%), dense vegetation/trees (18.5%), water body (9%), and bare land (2.5%).

3.2. Research Methods

3.2.1. Object-Based Image Analysis (OBIA)

OBIA is a common classification approach in remote sensing that uses image objects through segmentation instead of image pixels for feature extraction. The classification is performed based on the features extracted for each image object with any statistical or machine learning classifiers such as ANN. OBIA has two main components, which are segmentation and classification. The segmentation divides the given image data into a set of image objects that have homogeneous features (including the spectrum, texture, and shape). For each image object, a set of spectral, spatial, and textural features can be extracted and used for the next stage of data processing, i.e., classification. The classification in OBIA is often performed with any statistical or machine learning methods, including SVM, ANN, and decision trees (DT).

- OBIA features

Several spectral, spatial, geometric, and textural features can be calculated for image objects and used for the classification of the image data. Spectral features such as minimum, maximum, mean, standard deviation, and range spectral information are the most common spectral features used in OBIA.On the other hand, spatial features including object area, object perimeter, elongation index, shape index, density, and rectangular fit are the common spatial/geometric features used in OBIA. In addition, several textural features such as contrast, dissimilarity, homogeneity, energy, correlation, and angular second moment are the most used textural features in OBIA studies.

This research uses the common features described above in the OBIA-related experiments. The feature extraction tool was implemented in Python based on libraries such as SciPy (https://www.scipy.org) (15 September 2022) and scikit-image (https://scikit-image.org) (15 September 2022).

- Segmentation Algorithm

The commonly used multiresolution segmentation algorithm (MRS) [25] (Chen et al. 2006) was utilized in this research to assess the classification methods that use OBIA for feature extraction. MRS generates image objects with greater geographical significance and strong adaptability [26] (Martha et al. 2011).

MRS acquires image objects by calculating the heterogeneity of the spectral and shape features and weighting the image layers. For image layer i, the heterogeneity is calculated as [27] (Chen et al. 2021):

$$f_i = w_i h_{shape} + (1 - w_i) h_{color} \quad (1)$$

where w_i is the user-defined weight of the shape heterogeneity of band i, $0 \leq w_i \leq 1$. The shape heterogeneity h_{shape} is calculated from the compactness h_{cp} and smoothness h_{sm} as:

$$h_{shape} = w_{cp} h_{cp} + (1 - w_{cp}) h_{sm} \quad (2)$$

where w_{cp} refers to the weight of the shape's compactness. The segmentation scale S, shape heterogeneity weight w_i, and compactness weight w_{cp} are the main parameters of the MRS algorithm.

3.2.2. Convolutional Neural Networks (CNN)

CNN is the most popular type of neural network that has achieved great success in computer vision tasks, including image classification and object detection [28–31] (Hongtao and Qinchuan 2016; Srinivas et al. 2016; Li et al. 2022, Han et al. 2022). It also achieved excellent results for remote sensing image classification [32–35] (Sharma et al. 2017; Shakya et al. 2021; Boulila et al. 2022, Chand et al. 2022). CNN was applied to solve remote sensing problems utilizing aerial photographs, multispectral images, and hyperspectral images. CNNs combine three basic architectural ideas, including local receptive fields, sharing weights, and subsampling, to achieve properties of the shift, scale, and distortion invariant, which are very important for image feature extraction. In general, local receptive fields provide each neuron within a specified connected convolutional layer with the only nearby region belonging to the previous layer, which assists the network in deriving the basic visual characteristics. Furthermore, the concept of sharing weights refers to the stability of the convolutional kernel's weight during the process of generating a feature map at a specified layer. Consequently, the number of trained parameters in CNNs will significantly decrease in comparison to the ANN method. Finally, the process of subsampling decreases the feature map's resolution that integrates with a convolution operator to obtain translation invariance.

The CNN architecture employed for the assessment of the integration of OBIA-CNN methods is shown in Figure 3, which is composed of 7 layers, including the input layer as well as the output layer. The image layer is used as input to the two-dimensional convolutional layers in order to learn the feature map. The extracted features are subjected to the next step, called the 2D max-pooling layer, which aims to repress unwanted information to optimize the computational efficiency of the framework. The extracted features from the learning process are then transformed into an individual feature vector based on the flattened layer. After that, a dense (fully connected) layer is used to learn the contextual information from the features and help the classification layer. The classification is performed with a softmax layer, and the output predictions are used to classify the given image.

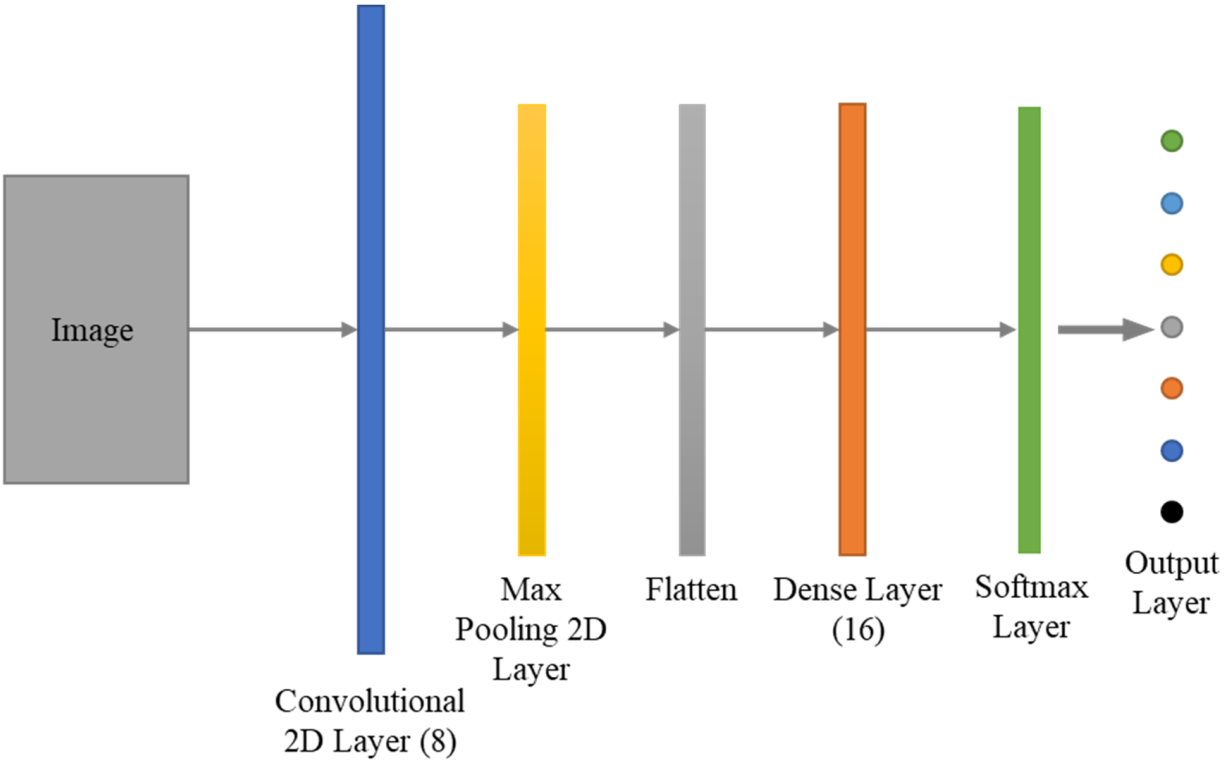


Figure 3. CNN architecture used in integrated OBIA-deep learning frameworks.

3.2.3. OBIA-CNN Integration Frameworks

The Theory of OBIA-CNN Integration

The main problem arises from the integration between OBIA-CNN due to the incompatibility between CNN input and OBIA segments, in general, the rectangular shape of CNN patch will never be compatible with the non-homogeneous shape of the segment (S) that resulted from OBIA method. For example, when performing OBIA classification based on patch-based CNN, the non-homogeneous segment is clipped based on a rectangular geometry of CNN patch, then CNN method is used to assign a specific category to this patch, then all pixels within the segment S will be assigned to the same category. Figure 4 shows the standard process of the integration between OBIA and CNN methods, two ideal situations may occur during the process of OBIA-CNN integration, the first situation arises when the segment (S) is larger than CNN patch, the patch is fully contained inside the segment (S), while the second situation arises when the CNN patch is larger than the segment (S) and the patch includes several segments represent the same land cover type, according to these situations, CNN patch is typically consistent with the segment (S), which can improve the accuracy of classification. The problem arises when the CNN patch covers multiple segments with different land cover categories. For example, a segment belongs to a specific type of land cover called (A); however, the CNN decision is mainly influenced by all of the pixels throughout the CNN patch, and the final output will represent the dominant land cover called (B) depending on several factors such as the dominant band value, larger areas, and structured textures. In this case, CNN classifier will make a wrong decision and obtain an incorrect classification for this segment [3]. Figure 4 shows the concept of the integration between OBIA and CNN.

OBIA-CNN Frameworks

Four OBIA-CNN integration frameworks have been identified from the literature review conducted for this research.

Table 2 summarizes the identified frameworks and their architectural concepts. The following subsections briefly describe each one of them:

- OBIA ANN

This method is the most basic integration of OBIA and deep learning. In this method, deep learning, i.e., ANN, is used to extract contextual and more abstract features based on OBIA features that are calculated with a typical OBIA procedure (image segmentation and spectral-spatial-textural features calculation) (Figure 4). OBIA features are organized in a tabular data structure, and the ANN is used to learn contextual information that may be present among the features. Finally, the classification is performed with a set of dense (fully connected) layers followed by a softmax layer. However, other classification models, such as SVM and DT, can be utilized to classify the contextual features and obtain the classification map of the given image. The current research uses a typical OBIA procedure, including MRS segmentation and spectral-spatial-textural features for the OBIA step, and an ANN as a classifier (Figure 5).

- Feature Fusion

Feature fusion for OBIA-CNN integration is another common framework used for remote sensing classification. In this method, OBIA is used to segment the given image and calculate several spectral, spatial, and textural features. Similarly, CNN is used to extract spatial abstract features from the data. The two obtained feature sets are then combined into a single feature vector. A classification layer is used to obtain the class labels using the feature vector that combines the OBIA and CNN features. See Figure 4 for the illustration of the typical OBIA-CNN feature fusion integration. This method may produce redundant features, such as shape and textural features, as they can be shared among the OBIA and CNN features. However, the problem can be tackled with network regularization or using dimensionality reduction techniques such as PCA.

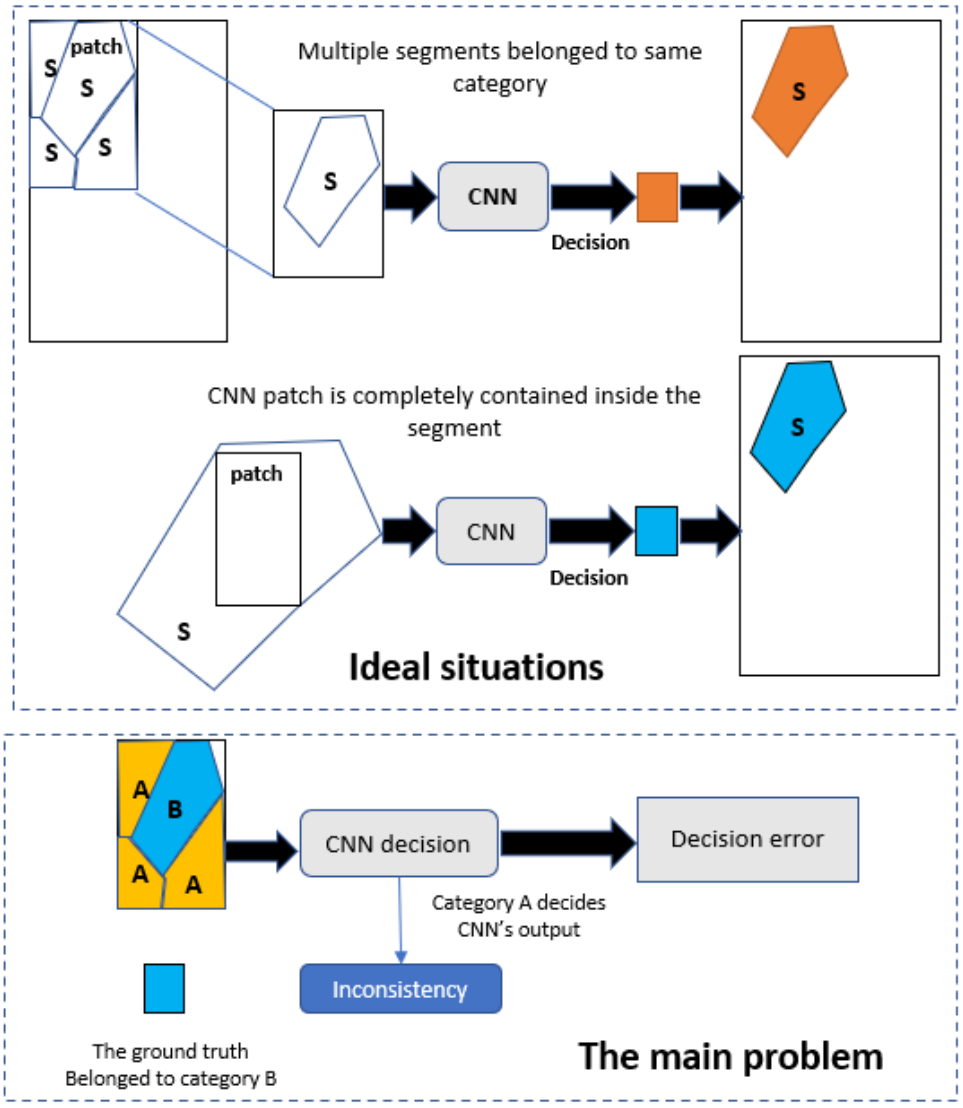


Figure 4. An example of the integration between OBIA and CNN for satellite image classification.

Table 2. Description of the integration frameworks assessed in this research.

Integration Framework	Base Models	Description
OBIA ANN	ANN	Applies ANN to OBIA features extracted at the image object level. This framework does not require training a CNN.
Feature Fusion	ANN + CNN	This framework has two computational branches. The first branch extracts OBIA features from the data. The second branch extracts deep features with a CNN from the data. Then, the obtained features are combined. Finally, ANN is applied to learn contextual features and perform classification.
Decision Fusion	CNN	This framework uses a CNN to obtain an initial classification map. Then, the image is segmented with a segmentation algorithm. Finally, the CNN map is refined based on the segmentation results with majority filtering.
Patch Filtering	CNN	Patch filtering uses image segmentation results to filter each image patch based on the dominant image object in the image patch. First, it produces image patches and image segmentation. Then, applies a variance filter to each image patch using the results of the segmentation. Finally, a CNN is applied to the obtained filtered image patches.

- Decision Fusion

This type of integration refines a classification map that is produced with a CNN or any other classification method based on the results of image segmentation (Figure 6). A majority filter is applied to make sure that each pixel in the image that belongs to the same image object is classified with the same label. This method is highly dependent on the accuracy of the CNN classification map and the results of the image segmentation. Other types of filtering may also be used, such as median and mean filters instead of the majority filtering.

- Patch Filtering

Patch filtering frameworks aim to filter image patches based on OBIA segmentation before using them in CNNs (Figure 4). For example, each image patch can be filtered based on the dominant image object that covers or contains the image patch with some filtering methods, such as variance filtering. The obtained filtered image patches are then used in CNNs as same as they are used with the traditional patch-based CNN methods.

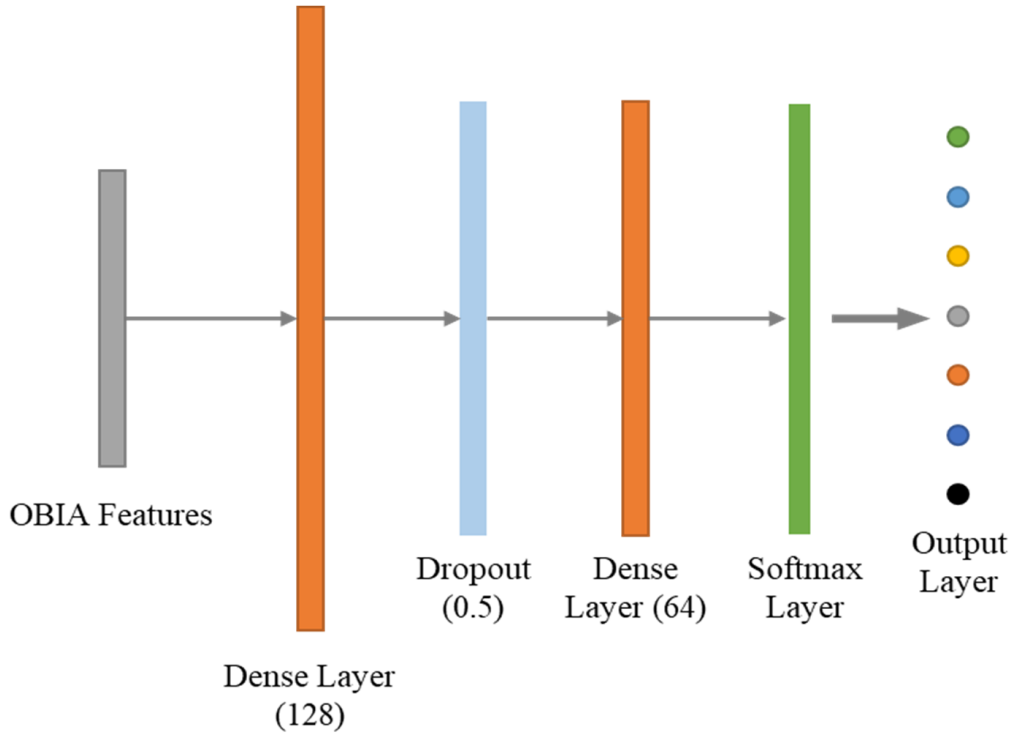


Figure 5. ANN architecture used in integrated OBIA-deep learning frameworks.

3.2.4. Training Parameters

Training deep learning models such as deep ANN and CNN require several hyperparameters that need to be configured carefully. In this research, the training parameters are set based on empirical experiments conducted on a subset taken from the whole dataset available to this research. The analyses showed some values to be suboptimal for the proposed classification models and the datasets used in the research. Table 3 summarizes the parameters used to train the ANN and CNN base models. For the ANN, the optimizer Nadam is Adam [36] (Kingma et al. 2014) with Nesterov momentum. The learning rate and learning rate decay were set as 0.001 and 0.001/100, respectively. The ANN models were trained for 500 epochs with early stopping criteria set to the patience of 15 epochs monitoring the validation loss. For the CNN, the Adam optimizer was found best with the same learning rate and learning rate decay as used for the ANN. The CNN models were also trained for 500 epochs with early stopping.

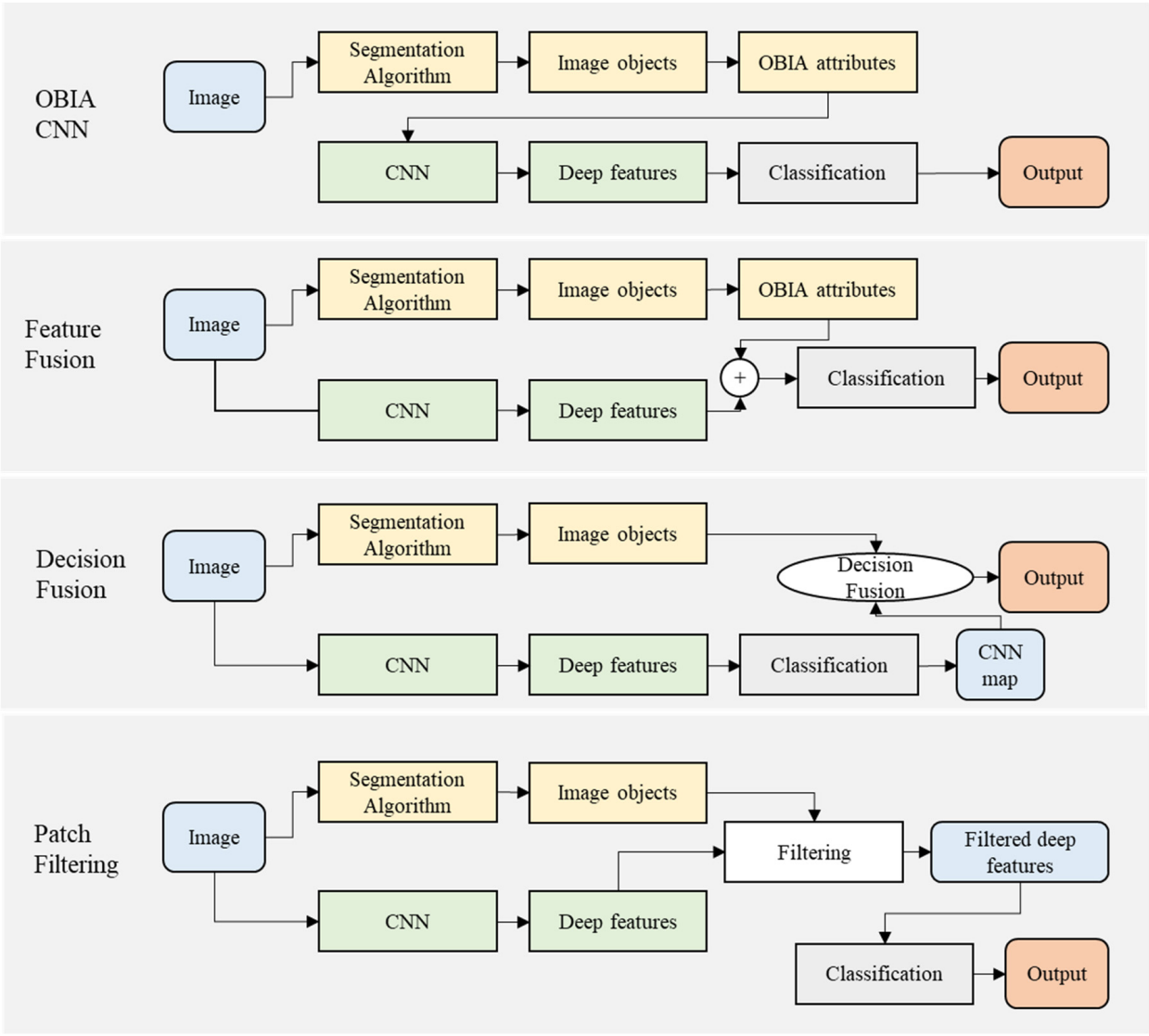


Figure 6. Typical architectures of the OBIA-CNN integration frameworks identified in this research.

Table 3. Parameters used to train the base models.

Model	Optimization	Learning Rate	Decay	Epochs
ANN	Nadam	0.001	0.001/100	500
CNN	Adam	0.001	0.001/100	500

In addition, Table 4 presents the hyperparameters of the ANN and CNN models used in this research. For the ANN, a dropout with a rate of 0.5 was used after the first dense layer to control the overfitting in the network. The hidden layer's activation was set as ReLU (rectified linear unit) as it is the common activation function used in many deep learning models for remote sensing applications. The loss function that was used to train the models with was categorical cross-entropy. The classification layer was a softmax layer. On the other hand, the CNN model had a few additional hyperparameters that needed to be configured to achieve the best classification results. The patch size of 5×5 was used as a sliding window size to extract image patches. The pool size in the max-pooling layers was set to 2×2. The size of the convolutional kernels was 3×3. The other configurations are the same as they were for the ANN models.

Table 4. Hyperparameters used in the base models.

Hyperparameter	Selected Value
Patch size	5×5
Pool size	2×2
Size of kernel filters	3×3
Dropout rate	0.5
Hidden layer activation	ReLU
Loss function	Categorical cross-entropy
Early stopping patience	15 epochs
Classification layer	Softmax

3.2.5. Accuracy Assessment Methods

The assessment of the classification accuracy was based on several accuracy metrics, including general assessment, overall accuracy (OA), Kappa index, F1-score and class-specific assessment, confusion matrix, and average class accuracy. For the training area, the accuracy is calculated on 20% of the samples that were not used during the training. For the test area, the models that were trained in the training area were used to generate the maps of the test area, and the assessment was performed using the ground truth data of the test area.

4. Results and Discussions

This section presents the results obtained from several experiments conducted in the current research to compare several integration frameworks for the classification of high-resolution aerial photographs. The accuracy of each classification model is presented and used to compare and assess the models. In addition, the classification maps are discussed regarding the visual quality and the distribution of the land covers in the study area.

4.1. Comparing Integration Frameworks

This research assessed four common OBIA-deep learning integration frameworks, namely, OBIA ANN, decision fusion, feature fusion, and patch filtering. In order to evaluate the feasibility of the integration, this research compared the integrated frameworks with pixel ANN and patch CNN models. This subsection presents the results obtained from the comparative experiments conducted in this research.

4.1.1. Comparing Integrated Models with Single Models (OBIA and CNN)

Two models were used to assess the integration feasibility in this research, including pixel ANN and patch CNN. Figure 1 shows how these models compare based on three accuracy matrices, i.e., OA, Kappa index, and F1-score, for the training and test areas. The results indicate that the patch CNN outperformed the pixel ANN on any accuracy metric and for both the training and test areas. For the training area, the patch CNN achieved 0.897 OA, 0.860 Kappa, and 0.900 F1-score, while the pixel ANN achieved 0.849 OA, 0.818 Kappa, and 0.850 F1-score. The accuracies of the patch CNN for the test area were 0.898 OA, 0.876 Kappa, and 0.890 F1-score. The pixel ANN for the test area achieved accuracies of 0.858 OA, 0.829 Kappa, and 0.860 F1-score.

In addition, Table 5 presents a detailed assessment of the four integration frameworks studied in this research. In general, the results indicate that patch filtering achieved the best classification results for both the training and test areas. Based on OA and Kappa index, the patch filtering achieved the best results for both the training (0.919 OA, 0.868 Kappa, and 0.920 F1-score) and the test areas (0.917 OA, 0.872 Kappa, and 0.910 F1-score). OBIA ANN performed slightly worse than patch CNN and pixel ANN. Using the test dataset, it obtained 0.842 OA, 0.784 Kappa, and 0.830 F1-score, whereas decision fusion and feature fusion achieved 0.862 OA and 0.860 OA, respectively.

Table 5. Accuracy assessment of the classification methods used in this research for the training and test areas.

Model	Training Area			Test Area		
	OA	Kappa	F1-score	OA	Kappa	F1-score
Pixel ANN	0.849	0.818	0.850	0.858	0.829	0.860
Patch CNN	0.897	0.860	0.900	0.898	0.876	0.890
OBIA ANN	0.838	0.763	0.840	0.842	0.784	0.830
Feature Fusion	0.869	0.807	0.870	0.860	0.788	0.850
Patch Filtering	0.919	0.868	0.920	0.917	0.872	0.910
Decision Fusion	0.864	0.793	0.850	0.862	0.793	0.850

4.1.2. Comparing the Classification Methods Using Per-Class Accuracies

Tables 6 and 7 present the per-class accuracies obtained for the assessed integration frameworks as well as the pixel ANN and the patch CNN.

Table 6. Per-class accuracies obtained for each classification method using the training area dataset.

Class Name	Average Accuracy					
	Pixel ANN	Patch CNN	OBIA ANN	Feature Fusion	Patch Filtering	Decision Fusion
Building	0.81	0.84	0.81	0.85	0.79	0.85
Road	0.75	0.82	0.72	0.78	0.76	0.75
Grass Land	0.90	0.93	0.75	0.65	0.88	0.70
Dense Vegetation or Trees	0.85	0.92	0.72	0.74	0.89	0.64
Water Body	0.94	0.93	0.88	0.94	0.89	0.93
Bare Land	0.86	0.69	0.70	0.88	0.93	0.89

Table 7. Per-class accuracies obtained for each classification method using the test area dataset.

Class Name	Average Accuracy					
	Pixel ANN	Patch CNN	OBIA ANN	Feature Fusion	Patch Filtering	Decision Fusion
Building	0.89	0.89	0.93	0.85	0.84	0.86
Road	0.80	0.88	0.79	0.74	0.84	0.79
Grass Land	0.78	0.88	0.55	0.71	0.82	0.60
Dense Vegetation or Trees	0.85	0.91	0.81	0.60	0.87	0.71
Water Body	0.98	0.97	0.90	0.93	0.97	0.98
Bare Land	0.84	0.68	0.72	0.86	0.89	0.85

For the training area, the grassland was best classified by the patch CNN (0.93), followed by the pixel ANN (0.9) and patch filtering (0.88). Roads were best identified by the patch CNN (0.82). Patch filtering achieved a road classification accuracy of 0.76, which is worse than the methods, i.e., patch CNN and feature fusion. Patch filtering also performed best for the bare land class with an accuracy of 0.93. For the building class, feature fusion and decision fusion (0.85) outperformed the other methods. Patch CNN achieved the best results for dense vegetation or trees (0.92). The water class was best classified by the pixel ANN and feature fusion (0.94).

For the test area, the grassland class was best classified by the patch CNN (0.88) followed by patch filtering (0.82). The OBIA ANN (0.55) and the decision fusion (0.60) achieved the worse results for the grassland class. OBIA ANN achieved the best results for the building class. In addition, bare land (0.89) was best classified by the patch filtering, and water (0.98) was best identified by the pixel ANN and decision fusion, respectively. Dense vegetation or trees were best classified with the patch CNN (0.91) and patch filtering (0.87).

Moreover, each land cover type can be described with certain spectral and spatial features. The integrated frameworks depend either on image segmentation, deep feature

extraction, or both. That is why some methods may perform well on certain land cover types compared to other types. For example, with the feature fusion method, no segmentation effect appears in the classification results. This method, therefore, may not be suitable for urban land cover with the complex geometry of buildings that exist in the area.

4.2. Results of Image Classification

Models such as OBIA ANN and patch CNN, as well as their integration by the four common frameworks, have been utilized to classify the image data of the training and test areas. Figures 7 and 8 show the classification maps obtained for the training and test areas using the assessed methods. The classification maps were produced according to six land cover classes, including grassland, road, building, dense vegetation or trees, and water. The training and test areas are mostly urban/built-up areas. The areas also consist of relatively complex road networks due to parking lots. Water bodies in the areas are mostly artificial small lakes. Both the training and test areas contain dense vegetation (or small trees) and grasslands.

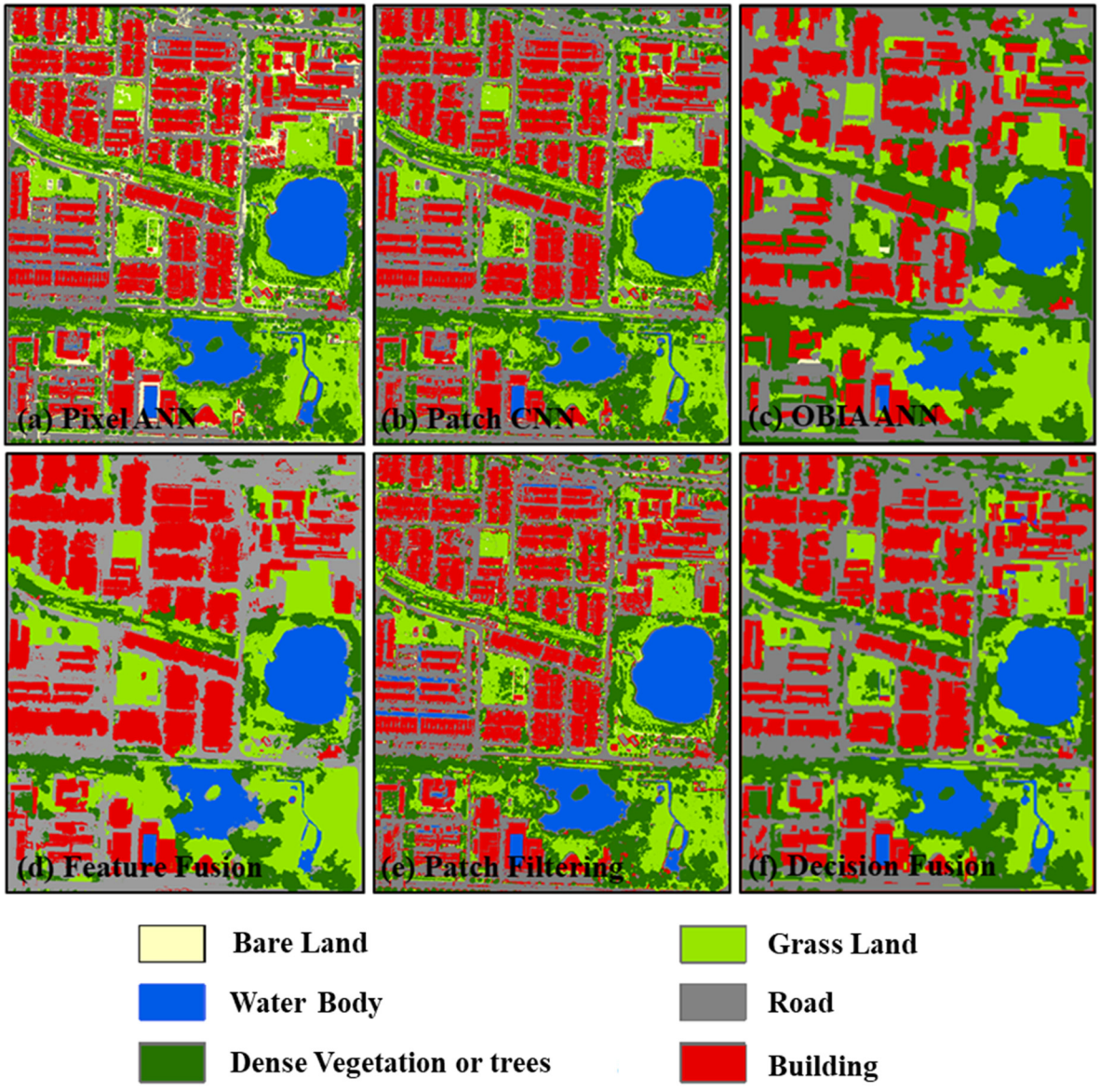


Figure 7. Classification maps obtained for the training area with the pixel ANN, patch CNN, and the four OBIA-CNN integration frameworks assessed in the current research.

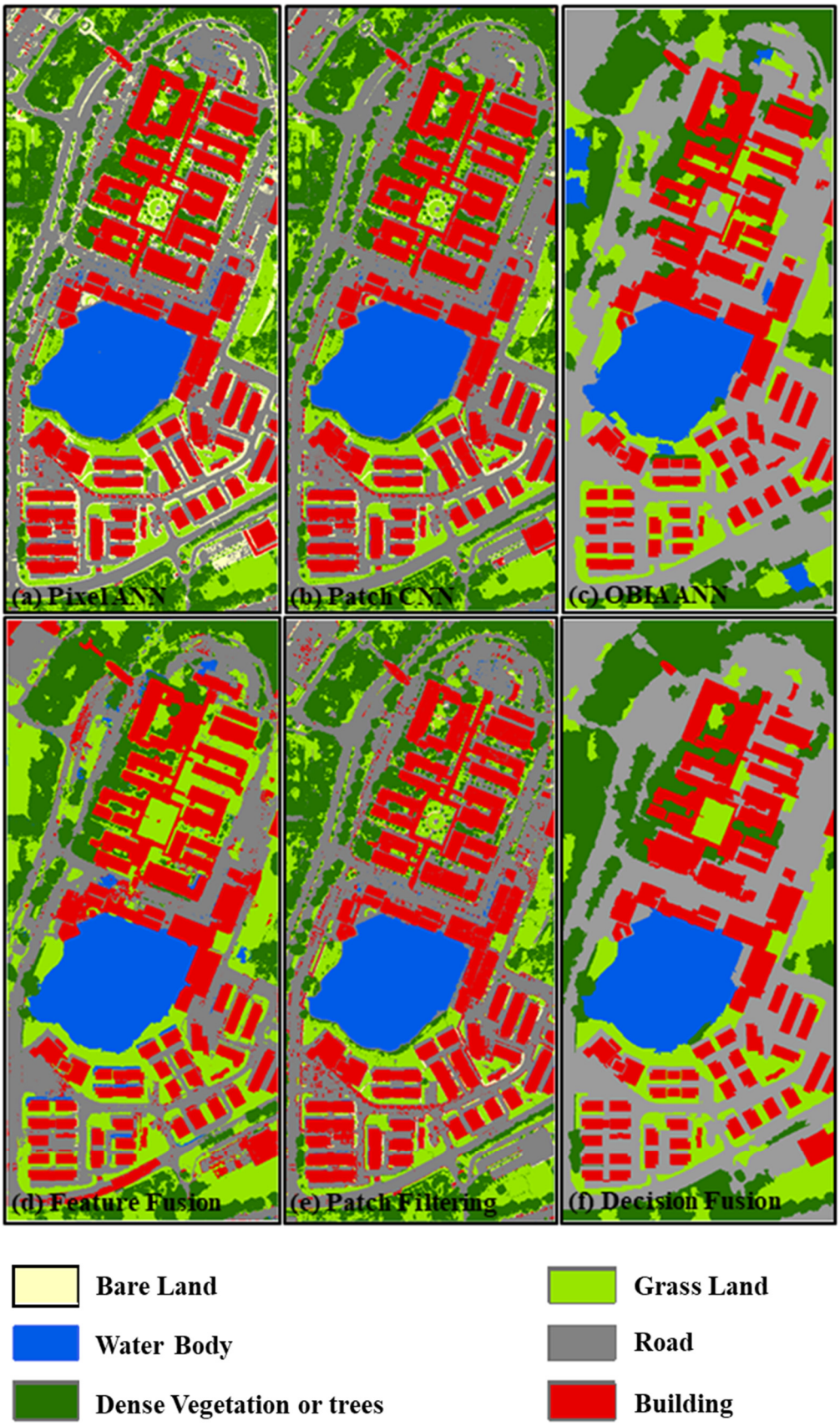


Figure 8. Classification maps obtained for the test area with the pixel ANN, patch CNN, and the four OBIA-CNN integration frameworks assessed in the current research.

Some classification maps present salt-and-pepper-like noise, while others contain less random noise, and the maps appear as vector maps. The maps of the pixel ANN contain random noise and misclassification between building and grassland and dense vegetation or tree classes. The maps of the patch CNN have less random noise and misclassifications. It also can be observed that the patch CNN improves the detection of the roads compared to the pixel ANN. The maps of the OBIA ANN contain very little random noise. However, they contain significant misclassification between building and road classes. It seems that the ANN could not learn useful contextual features from the OBIA features, or the latter is not enough to separate the building and road classes. The maps obtained by the patch filtering appear to have less random noise and misclassification compared to the patch CNN. The maps of the decision fusion are better than the OBIA ANN and with less misclassification between building and road classes.

4.3. Comparing with Recent Methods

In recent works on integrating OBIA and deep learning, contextual patches are commonly used. The methods, for example, center object-based CNN (Center OCNN) and random object-based CNN (Random OCNN), contextual patches using object center, or random points within an object, are used to extract patches for CNN feature extraction, respectively. The two methods are used as benchmark methods to compare with the standard integration frameworks used in this paper. The performance of these two methods is presented in Tables 8 and 9. The classification maps are presented in Figures 9 and 10 for the training and test areas, respectively. The OA for the center OCNN and random OCNN for the training area are 0.88 and 0.85, respectively. For the test area, the two methods achieved an OA of 0.9 and 0.89, respectively. The results indicate that the performance of these two recent methods was comparable to the standard integration methods and slightly worse than the best method, i.e., patch filtering.

Table 8. Overall accuracy and per-class accuracies of the proposed and benchmark classification methods based on samples from the training area.

Class	Center OCNN	Random OCNN
Buildings	0.91	0.8
Roads	0.82	0.78
Grass Land	0.9	0.89
Dense Vegetation/Trees	0.88	0.89
Water Body	0.93	0.93
Bare Land	0.93	0.93
OA	0.88	0.85
Kappa	0.86	0.82
F1-score	0.88	0.85

Table 9. Overall accuracy and per-class accuracies of the proposed and benchmark classification methods based on samples from the test area.

Class	Center OCNN	Random OCNN
Buildings	0.91	0.91
Roads	0.88	0.89
Grass Land	0.89	0.84
Dense Vegetation/Trees	0.91	0.88
Water Body	0.92	0.93
Bare Land	0.93	0.93
OA	0.9	0.89
Kappa	0.89	0.85
F1-score	0.9	0.88

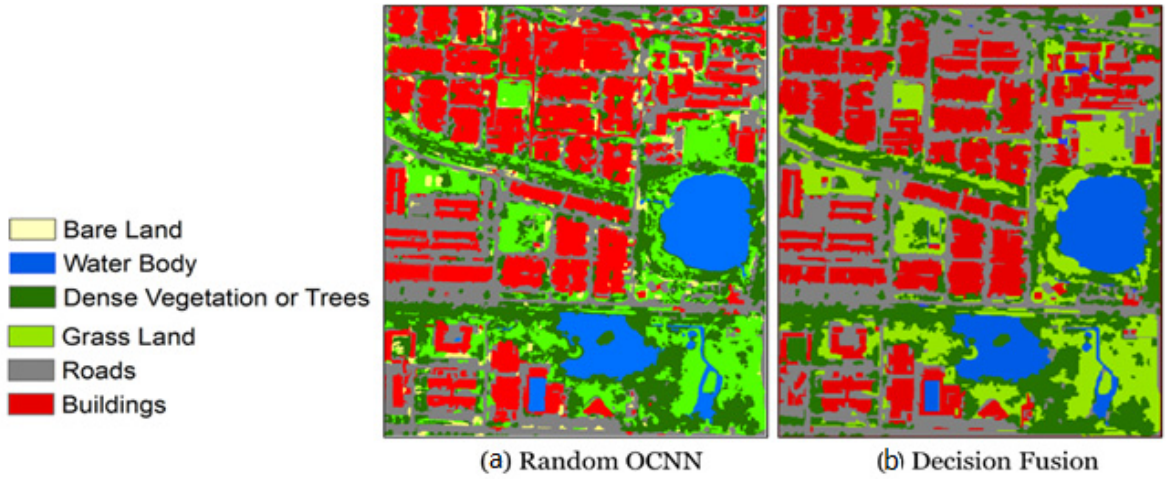


Figure 9. Classification maps obtained for the training area with the random OCNN and decision fusion.

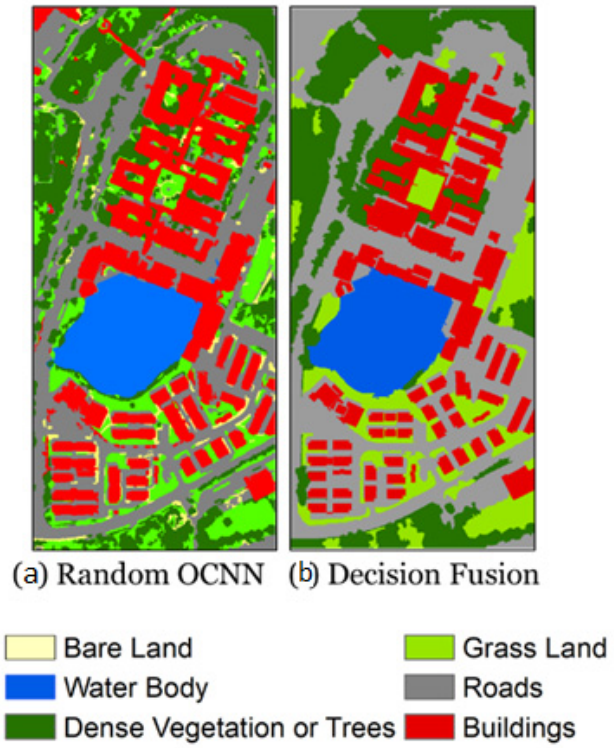


Figure 10. Classification maps obtained for the test area with the random OCNN and decision fusion.

4.4. Discussions

Deep learning methods such as CNNs are efficient feature extractors and have shown to be very successful for many computer vision applications, including image classification. In remote sensing, CNNs also achieved significant results compared to the traditional classification methods due to their ability to extract abstract features both from the spectral and the spatial domains. However, the challenges with deep learning, as presented in previous studies, are the problems of the artifacts in the class boundaries and the salt-and-pepper effect. Integration of OBIA and deep learning can solve the mentioned problems. However, the integration of OBIA into deep learning is not straightforward. This research assessed four common integration frameworks, namely, OBIA ANN, feature fusion, decision fusion, and patch filtering, for the classification of aerial photographs.

In deep learning models, pixels that belong to the same object are not enforced to be classified as the same semantic category. This reduces the artifacts in the class boundaries and the salt-and-pepper effect. However, deep learning has great capability in learning abstract features that are generalizable to new areas. It is; therefore, deep learning-based classification models often achieve high overall accuracy. On the other hand, OBIA, through segmentation, offers classification outputs that have few artifacts in the class boundaries and less salt-and-pepper effect. The combination of two methods can complement the strength of each method and result in classification models that have high accuracy and suitable quality.

The integration frameworks assessed in this research each attempt to combine the strength of each method differently. The OBIA ANN, for example, combines the strengths of OBIA, i.e., semantic segmentation and providing detailed spatial and textural features and the ability of ANN to learn contextual features such that the final models have an understanding of the relationship between the features. The main limitation of this approach is that ANN and similar models are not suitable for tabular data. The feature fusion method aims to extract features from each method and use the combined features for classification. This method can lead to redundant information as OBIA, and deep learning may learn the same spatial and textural features from the image data. In order to avoid the negative effects of redundant information in these models, dimensionality reduction techniques may be necessary. Alternatively, adding more layers with strong regularization after the concatenation layer combines the features from the OBIA and deep learning methods. Decision fusion uses only segmentation from OBIA and features from the deep learning method. Thus, there are no impacts of redundant information. However, some features of OBIA that may be useful for classification tasks are neglected. Decision fusion methods are easier to compute and optimize compared to the feature fusion method as there is no calculation of OBIA features. Finally, patch filtering methods are complex due to the requirement of matching between image objects and image patches. Moreover, there are several assumptions and abstractions in this method due to the matching process of the image objects and image patches. In addition, patch filtering methods not utilizing OBIA features as the decision fusion methods. This method uses local semantics based on image patches that may not take full advantage of the OBIA segmentation.

This research presented the advantages of combining OBIA and deep learning for very high-resolution satellite classification. Results showed that integrated frameworks, especially patch filtering, had a better result than the CNN-only methods and pixel-based ANN. Integrated models contributed to the reduction of artifacts in the class boundaries and the salt-and-pepper effect, which resulted in higher-quality classification maps.

5. Conclusions

This research assessed four OBIA-deep learning integration frameworks and compared them with a patch-based convolutional neural network (CNN) and a pixel-based artificial neural network (ANN) for the classification of high-resolution aerial photographs. The evaluated frameworks were OBIA ANN, OBIA-CNN feature fusion, decision fusion, and patch filtering. The best results were obtained by the patch filtering method for both the training and test areas.

Land cover mapping applications play a crucial role in many urban and environmental planning and management tasks. As those maps get more accurate, the plans made by the decision-makers for urban and environmental management can get better and become more efficient. The importance of integrating OBIA into deep learning for a high-resolution satellite image is highlighted in this research. However, there are still several challenges in this research area. For future works, the integration of OBIA-CNN should be more flexible by providing effective ways of combining image objects and image patches. Additional research should be carried out to improve the Patch filtering framework by finding better methods to merge image objects and image patches. Moreover, ensemble frameworks may also be applied. For example, OBIA features can be used in decision fusion methods by

refining with segmentation a classification map produced with an integrated model (e.g., feature fusion).

Author Contributions: H.Z.M.S. conceptualized, supervised, and obtained the grant for the research. O.S.A. and H.Z.M.S. collected and analyzed the data, performed the analyses and validation, wrote the manuscript, and contributed to the re-structuring and editing of the manuscript. H.Z.M.S., O.S.A., A.H.A. and N.A.B.H. professionally optimized the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by the Ministry of Higher Education Malaysia (MOHE) under the Fundamental Research Grant Scheme (FRGS) with project code: : FRGS/2/2014/TK02/UPM/02/2 (03-02-14-1529FR (Vote no:5524613)).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors acknowledge the resources and financial support provided by the Ministry of Higher Education Malaysia (MOHE). Universiti Putra Malaysia (UPM) is also acknowledged for the facilities provided, and the comments given by anonymous reviewers are highly appreciated.

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

References

1. Zhang, C.; Pan, X.; Li, H.; Gardiner, A.; Sargent, I.; Hare, J.; Atkinson, P.M. A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification. *ISPRS J. Photogramm. Remote Sens.* **2018**, *140*, 133–144. [CrossRef]
2. 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]
3. Pan, X.; Zhao, J.; Xu, J. An object-based and heterogeneous segment filter convolutional neural network for high-resolution remote sensing image classification. *Int. J. Remote Sens.* **2019**, *40*, 892–5916. [CrossRef]
4. Cui, W.; Zheng, Z.; Zhou, Q.; Huang, J.; Yuan, Y. Application of a parallel spectral–spatial convolution neural network in object-oriented remote sensing land use classification. *Remote Sens. Lett.* **2018**, *9*, 334–342. [CrossRef]
5. Jozdani, S.E.; Johnson, B.A.; Chen, D. Comparing deep neural networks, ensemble classifiers, and support vector machine algorithms for object-based urban land use/land cover classification. *Remote Sens.* **2019**, *11*, 1713. [CrossRef]
6. Abdollahi, A.; Pradhan, B.; Shukla, N. Road extraction from high-resolution orthophoto images using convolutional neural network. *J. Indian Soc. Remote Sens.* **2021**, *49*, 569–583. [CrossRef]
7. Lam, O.H.Y.; Dogotari, M.; Prüm, M.; Vithlani, H.N.; Roers, C.; Melville, B.; Zimmer, F.; Becker, R. An open source workflow for weed mapping in native grassland using unmanned aerial vehicle: Using Rumex obtusifolius as a case study. *Eur. J. Remote Sens.* **2021**, *54* (Suppl. 1), 71–88. [CrossRef]
8. 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]
9. Majd, R.D.; Momeni, M.; Moallem, P. Transferable object-based framework based on deep convolutional neural networks for building extraction. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 2627–2635. [CrossRef]
10. Sutha, J. Object based classification of high resolution remote sensing image using HRSVM-CNN classifier. *Eur. J. Remote Sens.* **2020**, *53* (Suppl. 1), 16–30.
11. Hong, L.; Zhang, M. Object-oriented multiscale deep features for hyperspectral image classification. *Int. J. Remote Sens.* **2020**, *41*, 5549–5572. [CrossRef]
12. Tang, Z.; Li, M.; Wang, X. Mapping tea plantations from VHR images using OBIA and convolutional neural networks. *Remote Sens.* **2020**, *12*, 2935. [CrossRef]
13. Guirado, E.; Blanco-Sacristán, J.; Rodríguez-Caballero, E.; Tabik, S.; Alcaraz-Segura, D.; Martínez-Valderrama, J.; Cabello, J. Mask R-CNN and OBIA fusion improves the segmentation of scattered vegetation in very high-resolution optical sensors. *Sensors* **2021**, *21*, 320. [CrossRef] [PubMed]
14. Lv, X.; Ming, D.; Lu, T.; Zhou, K.; Wang, M.; Bao, H. A new method for region-based majority voting CNNs for very high resolution image classification. *Remote Sens.* **2018**, *10*, 1946. [CrossRef]
15. Liu, S.; Qi, Z.; Li, X.; Yeh, A.G.O. Integration of convolutional neural networks and object-based post-classification refinement for land use and land cover mapping with optical and SAR data. *Remote Sens.* **2019**, *11*, 690. [CrossRef]
16. Abdi, G.; Samadzadegan, F.; Reinartz, P. Deep learning decision fusion for the classification of urban remote sensing data. *J. Appl. Remote Sens.* **2018**, *12*, 016038. [CrossRef]

17. Robson, B.A.; Bolch, T.; MacDonell, S.; Hölbling, D.; Rastner, P.; Schaffer, N. Automated detection of rock glaciers using deep learning and object-based image analysis. *Remote Sens. Environ.* **2020**, *250*, 112033. [CrossRef]
18. Timilsina, S.; Aryal, J.; Kirkpatrick, J.B. Mapping urban tree cover changes using object-based convolution neural network (OB-CNN). *Remote Sens.* **2020**, *12*, 3017. [CrossRef]
19. He, S.; Du, H.; Zhou, G.; Li, X.; Mao, F.; Zhu, D.E.; Xu, Y.; Zhang, M.; Huang, Z.; Liu, H.; et al. Intelligent mapping of urban forests from high-resolution remotely sensed imagery using object-based u-net-densenet-coupled network. *Remote Sens.* **2020**, *12*, 3928. [CrossRef]
20. Fu, Y.; Liu, K.; Shen, Z.; Deng, J.; Gan, M.; Liu, X.; Lu, D.; Wang, K. Mapping impervious surfaces in town–rural transition belts using China's GF-2 imagery and object-based deep CNNs. *Remote Sens.* **2019**, *11*, 280. [CrossRef]
21. 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]
22. 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]
23. Ji, S.; Shen, Y.; Lu, M.; Zhang, Y. Building instance change detection from large-scale aerial images using convolutional neural networks and simulated samples. *Remote Sens.* **2019**, *11*, 1343. [CrossRef]
24. Wang, J.; Zheng, Y.; Wang, M.; Shen, Q.; Huang, J. Object-scale adaptive convolutional neural networks for high-spatial resolution remote sensing image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *14*, 283–299. [CrossRef]
25. Chen, Y.; Feng, T.; Shi, P. Classification of remote sensing image based on object oriented and class rules. *Geomat. Inf. Sci. Wuhan Univ.* **2006**, *31*, 316–320.
26. Martha, T.R.; Kerle, N.; Van Westen, C.J.; Jetten, V.; Kumar, K.V. Segment optimization and data-driven thresholding for knowledge-based landslide detection by object-based image analysis. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 4928–4943. [CrossRef]
27. Chen, Y.; Chen, Q.; Jing, C. Multi-resolution segmentation parameters optimization and evaluation for VHR remote sensing image based on mean NSQI and discrepancy measure. *J. Spat. Sci.* **2021**, *66*, 253–278. [CrossRef]
28. Hongtao, L.; Qinchuan, Z. Applications of deep convolutional neural network in computer vision. *J. Data Acquis. Process.* **2016**, *31*, 1–17.
29. Srinivas, S.; Sarvadevabhatla, R.K.; Mopuri, K.R.; Prabhu, N.; Kruthiventi, S.S.; Babu, R.V. A taxonomy of deep convolutional neural nets for computer vision. *Front. Robot. AI* **2016**, *2*, 36. [CrossRef]
30. Li, Q.; Chen, Y.; Zeng, Y. Transformer with Transfer CNN for Remote-Sensing-Image Object Detection. *Remote Sen.* **2022**, *14*, 984. [CrossRef]
31. Han, Q.; Yin, Q.; Zheng, X.; Chen, Z. Remote sensing image building detection method based on Mask R-CNN. *Complex Intell. Syst.* **2022**, *8*, 1847–1855. [CrossRef]
32. Sharma, A.; Liu, X.; Yang, X.; Shi, D. A patch-based convolutional neural network for remote sensing image classification. *Neural Netw.* **2017**, *95*, 19–28. [CrossRef] [PubMed]
33. Shakya, A.; Biswas, M.; Pal, M. Parametric study of convolutional neural network based remote sensing image classification. *Int. J. Remote Sens.* **2021**, *42*, 2663–2685. [CrossRef]
34. Boulila, W.; Khlifi, M.K.; Ammar, A.; Koubaa, A.; Benjdira, B.; Farah, I.R. A Hybrid Privacy-Preserving Deep Learning Approach for Object Classification in Very High-Resolution Satellite Images. *Remote Sen.* **2022**, *14*, 4631. [CrossRef]
35. Chand, S. Semantic segmentation and detection of satellite objects using U-Net model of deep learning. *Multimed. Tools Appl.* **2022**, 1–20. [CrossRef]
36. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.