Journal of the Indian Society of Remote Sensing

Road Extraction from High-Resolution Orthophoto Images Using Convolutional Neural Network

--Manuscript Draft--

Manuscript Number:	ISRS-D-20-00656
Full Title:	Road Extraction from High-Resolution Orthophoto Images Using Convolutional Neural Network
Article Type:	Research Article
Corresponding Author:	Biswajeet Pradhan, Ph.D., University of Technology Sydney Faculty of Engineering and Information Technology Sydney, New South Wales AUSTRALIA
Order of Authors:	Abolfazl Abdollahi, PhD
	Biswajeet Pradhan, Ph.D.,
	Nagesh Shukla, PhD
Corresponding Author's Institution:	University of Technology Sydney Faculty of Engineering and Information Technology
Corresponding Author's Secondary Institution:	
First Author:	Abolfazl Abdollahi, PhD
First Author Secondary Information:	
Corresponding Author Secondary Information:	
Order of Authors Secondary Information:	
Funding Information:	University of Technology Sydney (Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS)) — Prof. Dr. Biswajeet Pradhan
Abstract:	Two of the major applications in geospatial information system (GIS) and remote sensing fields are object detection and man-made feature extraction (e.g., road sections) from high-resolution remote sensing imagery. Extracting roads from high-resolution remotely sensed imagery plays a crucial role in multiple applications, such as navigation, emergency tasks, land cover change detection, and updating GIS maps. This study presents a deep learning technique based on a convolutional neural network (CNN) to classify and extract roads from orthophoto images. We applied the model on five orthophoto images to specify the superiority of the method for road extraction. First, we used principal component analysis and object-based image analysis for pre-processing to not only obtain spectral information but also add spatial and textural information for enhancing the classification accuracy. Then, the obtained results from the previous step were used as input for the CNN model to classify the images into road and non-road parts and trivial opening and closing operation are applied to extract connected road components from the images and remove holes inside the road parts. For the accuracy assessment of the proposed method, we used measurement factors such as precision, recall, F1 score, overall accuracy and IOU. Achieved results showed that the average percentages of these factors were 91.09%, 95.32%, 93.15%, 94.44% and 87.21%. The results were also compared with those of other existing methods. The comparison ascertained the reliability and superior performance of the suggested model architecture for extracting road regions from orthophoto images.

Road Extraction from High-Resolution Orthophoto Images Using Convolutional Neural Network

Abolfazl Abdollahi[1], Biswajeet Pradhan[1,2,3]*, Nagesh Shukla[1]

[1]Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), University of Technology Sydney, NSW 2007, Australia; biswajeet.pradhan@uts.edu.au

[2]Center of Excellence for Climate Change Research, King Abdulaziz University, P. O. Box 80234, Jeddah 21589, Saudi Arabia

[3]Earth Observation Center, Institute of Climate Change, Universiti Kebangsaan Malaysia, 43600 UKM, Bangi, Selangor, Malaysia

* Correspondence: biswajeet24@gmail.com or Biswajeet.Pradhan@uts.edu.au

Funding

This research is supported by the Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS) in the University of Technology Sydney (UTS) under Grants 321740.2232335 and 321740.2232357.

Acknowledgments

The authors acknowledge and appreciate the provision of orthophoto images from airborne laser scanning data (LiDAR) by the Department of Planning.

Road Extraction from High-Resolution Orthophoto Images Using Convolutional Neural Network

Abstract

Two of the major applications in geospatial information system (GIS) and remote sensing fields are object detection and man-made feature extraction (e.g., road sections) from high-resolution remote sensing imagery. Extracting roads from high-resolution remotely sensed imagery plays a crucial role in multiple applications, such as navigation, emergency tasks, land cover change detection, and updating GIS maps. This study presents a deep learning technique based on a convolutional neural network (CNN) to classify and extract roads from orthophoto images. We applied the model on five orthophoto images to specify the superiority of the method for road extraction. First, we used principal component analysis and object-based image analysis for pre-processing to not only obtain spectral information but also add spatial and textural information for enhancing the classification accuracy. Then, the obtained results from the previous step were used as input for the CNN model to classify the images into road and non-road parts and trivial opening and closing operation are applied to extract connected road components from the images and remove holes inside the road parts. For the accuracy assessment of the proposed method, we used measurement factors such as precision, recall, F1 score, overall accuracy and IOU. Achieved results showed that the average percentages of these factors were 91.09%, 95.32%, 93.15%, 94.44% and 87.21%. The results were also compared with those of other existing methods. The comparison ascertained the reliability and superior performance of the suggested model architecture for extracting road regions from orthophoto images.

Keywords: CNN; deep learning; orthophoto images; OBIA; road extraction; remote sensing

1. Introduction

Space-borne, airborne, and drone-based sensors have obtained large amounts and different kinds of high-resolution images using recent advanced earth observation and remote sensing technologies; these images have been extensively utilized in several applications, such as urban planning (Abdullahi, Pradhan, & Jebur, 2015), disaster management (Youssef, Sefry, Pradhan, Alfadail, & Risk, 2016), and emergency tasks (Weng, 2012). Extracting road networks from remote sensing imagery plays a vital role in the improvement of transportation systems, such as traffic control, map updating, and automatic road navigation, for daily life and industrial applications (Z. Zhang, Liu, & Wang, 2018). Consequently, generating a novel technique to extract road regions from satellite images of high resolution and keeping road networks up to date is useful to geospatial information system (GIS) (W. Shi, Miao, & Debayle, 2014). High-resolution remote sensing imagery can produce massive scale data and has become the main data source to extract road regions and update geospatial database in real time (J. Zhang, et al., 2017). At present, road networks are changing more rapidly than ever. Acquiring precise road information from remote sensing data is also currently demanded (Abdullahi, et al., 2015). Therefore, extracting road features from satellite images of high-resolution has become a major research topic in the remote sensing field (Xia, Zhang, Liu, Luo, & Yang, 2018). Although road extraction from remote sensing imagery has received considerable attention in recent years, the task is still challenging because the sections and structure of roads are irregular and complex, respectively (Youssef, et al., 2016). Other features such as building roof, pedestrian areas, and car parking are similar in satellite images, thereby resulting in insufficient context of roads in these images. Meanwhile, vehicles on roads, shadows of trees, and buildings on roadsides can be identified from high-resolution satellite images (Bakhtiari, Abdollahi, & Rezaeian, 2017). Road class extraction from high-resolution remote sensing imagery is difficult because of the aforementioned issues.

Manual and traditional road extraction approaches from high-resolution remote sensing imagery are costly, time consuming, and full of errors because of human operators (J. Wang, Song, Chen, & Yang, 2015). Therefore, different road extraction approaches, such as supervised (Miao, Shi, Gamba, & Li, 2015) and unsupervised (Grinias, Panagiotakis, Tziritas, & sensing, 2016) techniques, have been suggested by researchers to extract road networks from high-resolution remote sensing images. These approaches use textural (Sghaier & Lepage, 2016), geometric, and photometric (He, Liao, Yang, Deng, & Liao, 2012) information to extract roads through classification (Cheng, Ding, Ku, & Sun, 2012).

2. Related Works

In this part, we discuss some unsupervised and supervised methods for road extraction from remote sensing images and then explain some early works related to deep learning methods to highlight the main contribution of deep learning approaches in extracting road sections. Unsupervised techniques use clustering algorithms to extract roads from remote sensing images. These methods are a form of pixel-based classification and computer automated (Xu, Xie, Feng, & Chen, 2018). Khesali, Zoej, Mokhtarzade, and Dehghani (2016) proposed a semi-automatic road extraction method by combining high-resolution IKONOS and TerraSAR-X images. They introduced two fusion approaches: knowledge-based fusion and neural network. First, they used various textural parameters and spectral features of optical images to implement neural networks on images and detect roads separately. Then, they applied the knowledge-based fusion method using thresholds of vegetation gray levels and narrow roads to extract road from every reference separately. Finally, the outputs were compared, and the benefits and drawbacks of each data source were examined. The experimental results demonstrated that the suggested approach can be implemented for road extraction. Unsalan and Sirmacek (2012) used a novel system based on probabilistic and graph theoretical approaches to extract road from high-resolution satellite images. They used a different type of images, namely, those from GeoEye, IKONOS, and QuickBird, to specify the weaknesses and strengths of the proposed method. The achieved outcomes proved that the suggested technique is effective and reliable in road extraction on such images. Supervised methods, such as support vector machine (SVM) (Abdollahi, Bakhtiari, & Nejad, 2018), random forest (Bedawi & Kamel, 2015), artificial neural network (Kirthika & Mookambiga, 2011), and deep learning, are more accurate than unsupervised methods. These approaches use labeled samples for training to extract features from remote sensing images (W. Wang, et al., 2016).

Abdollahi, et al. (2018) used a fusion method based on SVM and level set (LS) algorithms for road extraction from Google Earth images. First, SVM method was applied to classify the images, and then LS method was used to extract road sections from images. The empirical outcomes showed that the introduced technique can achieve excellent results in completeness and correctness values. However, the suggested approach misclassifies some objects that are similar to road class as false road sections.

Unsupervised methods rely on color features and are limited by color sensitivity (Panboonyuen, Vateekul, Jitkajornwanich, & Lawawirojwong, 2017). Therefore, if roads in remote sensing imagery have more than one color, then these segmentation algorithms will not attain excellent results and will not perform well in road extraction and classification. The current study focuses on the color sensitivity problem. In recent years, artificial intelligence algorithms have shown important developments in feature segmentation and extraction from remote sensing imagery and have encouraged researchers to identify road class from high-resolution remote sensing images because of the great efficiency of deep learning methods in various applications (Xu, Chen, Xie, & Wu, 2017).

One of the rapidly growing areas in machine learning is deep learning, which has become an optimistic tool for expediting image processing and object detection and has been strongly implemented to remote sensing images, especially in mapping of urban land cover with high accuracy results (Audebert, Le Saux, & Lefèvre, 2017). This section discusses previous works related to deep learning methods that have been applied on remote sensing images to extract road sections. J. Wang, et al. (2015) suggested a framework of neural dynamic model to extract road sections from VHR remote sensing imagery based on deep convolutional neural network (DCNN) and finite state machine (FSM). DCNN works as an important part to identify features from a complicated and dynamic atmosphere, whereas FSM changes the identified features to state for capturing their tracking habits. Their results indicated that the suggested approach outperforms the traditional approaches and is more accurate in extracting road section from images of high-resolution satellite data. Panboonyuen, et al. (2017) proposed an approach based on DCNN with landscape metrics and conditional random fields (CRF) for extracting road parts from high-resolution satellite images. They also applied a function of modern activation named exponential linear unit to modify the DCNN proficiency. They implemented the proposed approach on Thailand Earth Observation System satellite images and Massachusetts road aerial image datasets. Their suggested technique is accurate in road object segmentation on different kinds of remote sensing images in terms of recall, precision, and F1. The results attained for precision and F1 are 85% and 87% for aerial imagery and 75% and 64% for satellite imagery. Henry, Azimi, and Merkle (2018) proposed a deep fully convolutional neural networks to extract road from SAR images. They added spatial tolerance rules to the new networks to enhance their sensitivity towards thin objects. The experimental results show that their model can achieve good results and extract most of the road sections in their dataset.

Alshehhi and Marpu (2017) proposed a convolutional neural network (CNN) model to simultaneously extract roads and buildings from high-resolution remote sensing images. They used two challenging datasets (Massachusetts and Abu Dhabi) to illustrate the efficiency of the suggested network architecture. They integrated small features of roads and buildings of near areas with CNN to improve the performance of the model. They found that the introduced model has excellent performance in extracting road and building features from remote sensing imagery with high-resolution in urban regions.

In the current study, we developed a CNN model with regularization methods, such as dropout for road extraction from VHR orthophoto images. This research aims to use object-based image analysis (OBIA) and principal component analysis (PCA) and run the CNN model on several orthophoto images to explore the impact of CNN architecture on road extraction. In remote sensing, PCA methods have been implemented to improve classification, determine the trends in image data, and specify anomalies in outputs (Comber, Harris, & Tsutsumida, 2016). OBIA method can make a smart class decision based on class relationships utilizing image object, size, shape, and spectral, which can overcome color sensitivity and enhance the performance of the classifier on road extraction. Therefore, in this study, we introduced a technique based on CNN and spectral–spatial information to reduce the effect of color sensitivity and extract road from orthophoto images. In this paper, the segmented image using multiresolution segmentation algorithm was used as an input for CNN model for object-based image classification and then trivial opening and closing operation used to extract connected road components and fill holes in the road sections. Therefore, the main contribution of this study is to mix PCA and OBIA with CNN model to classify orthophoto images into road and non-road parts and then trivial opening is applied to make the binary image and extract road parts that has not done in the literature review. By applying this, the computation and training time were reduced, and the proposed model was trained by some samples (road and non-road segments) for only a few seconds, which is very less compared to aforementioned deep learning methods in this study while can achieve good results.

3. Methodology

In this section, the images, pre-processing, and the architecture of the suggested CNN model and training procedure are exhibited.

3.1. Data

In this research, data from orthophoto images from the Selangor State in Peninsular Malaysia were used (Figure 1). Selangor is one of the states in Malaysia and is situated in the western part of the country. The latitude and longitude of Selangor are 3.519863 and 101.538116, respectively. Orthophoto images were collected on November 2, 2015 with an airborne laser scanning of LiDAR system with an Optech Airborne Laser Terrain Mapper 3100 instrument with a flying height of 1510 m in a clear sky condition. These images were captured as RGB with a pixel resolution of 7 cm.

Fig. 1. Study area and the orthophoto images used in this study

3.2. Pre-processing

3.2.1. Geometric Correction

Pre-processing and removal of geometric distortion of remotely sensed data and specifying separate pixels in their properly denoted planimetric (x, y) map locations are necessary. Therefore, we can use geometrically corrected images to exploit polygon area, direction information, and precise distance (Aasen, Honkavaara, Lucieer, & Zarco-Tejada, 2018). For geometric correction, several ground control points were first collected from evidently recognizable points (e.g., solitary trees, corners, and road intersections) from the field. All the chosen points were well distributed throughout the images. Geometric calibration consists of three main steps: (1) recognition of transformation points in the image, (2) application of least square, and (3) accuracy evaluation process (Abdollahi, Pradhan, & Shukla, 2019). At this point, we applied the least square approach to determine the coefficient for the geometric rectification process. In addition, polynomial equations were used to specify the residuals and root mean square between the aligned X, Y coordinates and the reference X, Y coordinates.

3.2.2. Normalization

Data normalization is important because it enhances the progress of gradient descent optimization and activation functions. In this step, we used min–max normalization to normalize pixel values of the orthophoto images and avoid unusual gradient. Min–max normalization is also called feature scaling, where the range of numeric values of data is decreased between 0 and 1. This normalization can be computed using Equation (1).

$$z = \frac{x - min(x)}{max(x) - min(x)} \quad (1)$$

where z is the normalized data and min and max are the minimum and maximum values in x given its range.

3.3. Suggested method

This study aims to create an effective solution for extracting road sections from VHR orthophoto images using CNN model. The orthophoto image consists of m × n × d digital values, where m, n, and d are the image width, length, and depth. Common classification approaches use spectral information and a set of training examples to specify a label to each pixel on the image and classify the images (Sameen, Pradhan, & Aziz, 2018). In the present study, we not only used spectral information but also applied PCA and OBIA not only to achieve spectral information such as brightness, mean and standard deviation of each object but also obtain more information related to geometry and texture such as area, number of pixels, length, homogeneity, contrast, dissimilarity, entropy, correlation for improving the accuracy of classification. In the OBIA process, pixels are classified into objects on basis of either outside variable such as geological characteristics or spectral resemblance. Numerous variables may be assigned and categorized as spectral, shape, and neighborhood. Neighborhood variables include the mean variance of an object associated with dark ones; spectral variables include the standard deviation and mean value of a special spectral band; and shape variables include compactness, size, and perimeter. By combining several neighboring objects into one larger one, each object can be obtained (Blaschke, 2010). For the OBIA process, we applied multiresolution segmentation method to convert the images into superpixels and we tried set the scale, shape, and compactness parameters for the proposed segmentation method to 30, 0.2, and 0.5, respectively, to obtain high accuracy in the classification process. The proposed segmentation method is a region-based method, which reduces the non-homogeneous segments using spectral and shape characteristics. PCA method is a mathematical approach for dimension reduction of data (Ng, 2017). This method extracts the principal pattern on a linear system based on factoring matrix principle and maintains the main features of the image.

3.3.1. Architecture of Convolutional Neural Network

A typical CNN model includes alternatively piled convolution layers followed by dense (fully connected) ones. The convolution layers contain a series of layers such as convolution and pooling layers and dense and non-linear transformation functions. The convolution computes a dot product within the nearby region to produce each element of new images (feature maps), which is combined with a collection of weights (kernels) and the input feature maps. A non-linear function (e.g., Relu and tanh) and a pooling function are applied after this operator. Pooling function uses pre-defined functions (e.g., maximum and average) on a nearby area to fulfill down-sampling along the spatial dimensions of feature maps (Hu, Xia, Hu, & Zhang, 2015). A down-sampling method is doing sampling the image using principle local correlation of the image. This method can maintain effective information while reduce the amount of data processing and provides the features taken through convolution to have spatial invariability. For classifying and predicting the feature vector in the final output of network, a dense layer is applied. Using fully connected layer a set of number normalized to 0 and 1 is achieved, which the greater value of each sample belongs to a specific class. The fully connected layer links all neurons to every single neuron in its layer by taking them from the previous layers. Dropout regularization approach is performed to avoid overfitting in dense layers (Srivastava, Hinton, Krizhevsky, Krizhevsky, & Salakhutdinov, 2014). Dropout regularization method decreases the number of neurons of the network, which do not contribute anymore to the back-propagation and forward-pass (Nogueira, et al., 2016). To produce the probable output for every class, a sigmoid function, which is a logistic regression function, is used for binary classification in the last dense layer. The function maps any real value into another value between 0 and 1 (Krizhevsky, Sutskever, & Hinton, 2012).

W × H image patch with N-channels centered at x(i,j) and 2D filter-kernel $w_f \times h_f$ as input and output feature maps of $(W - w_f + 1) \times (H - h_f + 1)$ with K-channels are taken by a convolution layer. Each channel of the output image is named a filter site. A stride sf parameter is the distance, which is needed to slide down the convolution procedure in the input image. This stride parameter can affect the output of the convolution procedure (Hu, et al., 2015). The size of the output map from the convolution process is decreased to $((W - w_f)/ s_f + 1) \times ((H - h_f)/ s_f + 1)$ if $s_f > 1$. The convolution process is expressed using Equation (2).

$$x_k(ii, jj) = \sum_{n=1}^{N} \left\{ \sum_{p=0}^{W_f - 1} \sum_{q=0}^{h_f - 1} x_n(i \cdot s_f + p, j \cdot s_f + q) \cdot h_k(p, q) \right\} + b_k \tag{2}$$

where x_n (i,j) is the pixel value at (i,j) in the n-th channel of an input feature map, x_k(ii,jj) is the pixel value at (ii,jj) in the k-th filter site of the input map, h_k(p,q) is the weight value at (p,q) of the k-th filter, and bk is the bias parameter of the k-th filter that is shared among all locations (p,q).

A convolution process is accompanied by an activation function, which is a kind of transformation function. x_k(ii,jj) is used as input to the activation function of the neural network that is the output of convolution operation. b is a bias vector, and w is a weight vector. The activation function is defined using Equation (3).

$$Z(x_k(ii,jj)) = f\left(\sum_{k=1}^{k} x_k(ii,jj) \cdot w_k + b_k\right) \Leftrightarrow Z = f(X \cdot W + b) \tag{3}$$

For $f(\cdot)$, several alternative functions, such as rectified function, sigmoid, and tanh, can be used. Neurons work effectively with rectified function because this function induces sparsity in the hidden layers and avoid saturation during the learning process. Neurons also do not encounter gradient vanishing difficulty, which occurs when the gradient norm decreases after sequential updates in the back-propagation process (Zhou, Lapedriza, Xiao, Torralba, & Oliva, 2014). In this study, rectified linear unit (Relu) function is used for the first convolution and dense layers, which is defined using Equation (4).

$$A(x_k(ii,jj) = \max(0, Z(x_k(ii,jj)) \tag{4}$$

Pooling layers are utilized to apply down-sampling to reduce the number of parameters, amount of network computing and the size of image. In this study, we used max pooling to mix semantically comparable features into one (Maggiori, Tarabalka, Charpiat, & Alliez, 2017).

A classifier layer after a convolution and fully connected layer is used to predict class possibilities. In this study, we applied binary logistic regression algorithm to assign observations to a discrete set of classes. In order to solve the problem of binary classification in this study (road and non-road), we used sigmoid activation to map predicted values to probabilities Equation (5).

$$S(z) = \frac{1}{1+e^{-z}} \tag{5}$$

Where z is the input and S is the output between 0 and 1.

3.3.2. Model architecture

In this study, a simple CNN model was applied, which was created with two convolutional layers followed by two max-pooling operation, dropout, and two fully-connected layers (Figure 2). In this model, the convolutional kernel size was defined as 3×3 and 2×2 for pooling size in the max-pooling layer. To avoid overfitting, a dropout operator was implemented in the convolutional layer and the first fully connected layer with a drop probability of 0.25. Dropout is a technique, which neglect randomly chosen neurons during training. This means that their contribution to the activation of downstream neurons is temporally deleted on the forward pass and any weight updates are not used to the neurons on the backward pass. The entire process was performed on a CPU Core i7 2.11 GHz and memory RAM of 16 GB and GPU Nvidia Quadro P4000 with compute capability of 6.1 and memory of 8 GB under the framework of Keras with Tensorflow back-end. We used 2324 sample data, of which 70% were for training and validation (1626 samples), and 30% were for testing (698 samples). The number of nodes for the convolutional and max-pooling layers was set to 32, whereas that for the first dense layers was 256. Given that we had two classes (road and non-road), the number of nodes for the last dense layer was set to 2 depending on the number of class. A perfect optimization function is required to minimize the energy function and update the parameters of the model algorithm during training the model [44]. In this work, one of the most common optimizers (Adam) was used to minimize the losses and update the parameters, such as weights and biases. In this model, we used Adam optimizer for stochastic gradient descent to train deep learning models for reducing the loss function and for loss function a binary cross entropy function was used to quantifies the difference between two probability distributions. In this model, the adam configuration parameters such as learning rate, beta1 and beta2 defined as 0.001, 0.9 and 0.999 respectively. Table 1 shows a summary of the model layers.

3.3.3. Trivial Opening

Trivial operation was applied to extract connected road components on basis of some criteria. Assume that P(i) is the connected component, P is the image and T is main axis length, the trivial opening can be achieved using Equation 6 (Sujatha & Selvathi, 2015).

$$R_0 = \{P \mid \text{Long axis of minimum ellipse enclosing } P(i) \geq T\} \quad (6)$$

Where R_0 is the connected component. According to the condition T, trivial operation is utilized for suitable connected road components extraction. The whole region of connected road components is preserved if that component satisfied the condition T and it is removed if not satisfied the condition T. Since high-resolution remote sensing imagery were used in this paper, road sections emerged as long features and similar areas in these images, in which they can be simply filtered using trivial opening after classification by CNN. Also, closing morphological operation was applied to remove unwanted objects inside the extracted road parts and fill the holes.

Fig. 2. Proposed CNN model architecture

Table 1

Summary of the proposed CNN model layers

3.4. Performance Evaluation

Measurement factors, namely, precision, recall, F1 score, intersection over union (IOU) and overall accuracy (OA) (Equations (7)– (11)), are utilized to assess the efficiency of the introduced approach. Recall denotes the proportion of road pixels that are accurately classified among all actual road pixels. Precision describes the proportion of road pixels that are accurately classified among all anticipated pixels. F1 score is a composition of recall and precision. IOU calculates the number of pixels common between the prediction and target masks divided by the total number of pixels present across both masks. OA measures the precision of road and non-road pixels.

$$Precision = \frac{TP}{TP+FP} \tag{7}$$

$$Recall = \frac{TP}{TP+FN} \tag{8}$$

$$F1 = \frac{2\times Precision\times Recall}{Precision+Recall} \tag{9}$$

$$IOU = \frac{TP}{TP+FP+FN} \tag{10}$$

$$OA = \frac{Pos_r+Pos_n}{N} \tag{11}$$

where N is the number of pixels for test images and Pos_r and Pos_n are the positive number of road and non-road pixels at a pixel level.

4. Experimental results

Semantic segmentation, especially detection and extraction of objects (e.g., roads), from remote sensing images of high resolution play an essential role in several applications, such as traffic management, land cover analysis, urban planning, and emergency tasks. An increasing number of satellites are being launched as remote sensing technology develops. Therefore, accessing remote sensing imagery has become easier than before. This work also defined a model based on CNN and trivial opening methods to classify image into non-road and road areas and extract roads from orthophoto images. Arcmap 10.6, Ecognition Developer, and Python were used to perform the proposed model and calculate the performance accuracy for road extraction. Four images from various areas covered by vegetation and building were used to validate the performance of the suggested model in road extraction from orthophoto images in general (Figure 3). The figure has subfigures of three columns and five rows. The original, classified images and target road map (ground truth) are presented in the first, second and third columns, respectively. Figures 3(a), (c) and (j) show the image in an outer space of the city. The road section in the image was not surrounded by other features compared with the road section in Figures 3(e) and (g), which was completely covered by buildings, trees, and cars. As observed in the original images of Figure 3, the road and other features appeared with similar spectral characteristics, which introduced difficulty for the model in extracting road class accurately. Similar objects (noise) also appeared as road class in the extracted image. Therefore, we used PCA and OBIA to obtain additional information related to road objects, such as geometry, shape, and elongation, for modifying the performance of the suggested CNN model and extracting road class with high accuracy by eliminating non-road pixels and noises. After all the information was gathered, it was used as input for the proposed CNN model to identify road class from other objects in the orthophoto images. Figures 3(b), (d), (f), (h) and (k) show that the mixing CNN model and trivial opening could classify and extract the road section from VHR orthophoto images with high precision.

Fig. 3. Results of road extraction using the suggested CNN model: original orthophoto images (a, c, e , g, j), extracted road sections (b, d, f, h, k) and target road map (i, ii, iii, iv, v)

Figure 4 shows the performance accuracy of the proposed model with a dropout for 100 epochs on training and validation datasets. The model had learned effective characteristics to classify the images and extract road class depending on the increment in model accuracy and reduction in model loss over time. The accuracy of the model from one epoch to another fluctuated due to the use of dropout, which yielded a moderately different model at every epoch. The performance computation of the proposed CNN model was dependent on the hyperparameters such as image patch size and convolutional filters as well as dropout and other layers. The proposed CNN model with dropout took 174 second to be trained. Therefore, the performance calculation of the model is effective for the examined images while it will require more time for larger datasets.

Fig. 4. Model accuracy (a) and model loss (b) of the proposed CNN model

Road class extraction from high-resolution remotely sensed images can be considered binary classification. In this study, a confusion matrix was used to evaluate the performance of the proposed CNN model and assess the number of pixels belonging to road sections (positives) and other sections (negatives). Four important factors should be considered for calculating confusion matrix: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). TP indicates the number of precisely categorized road pixels, TN indicates the number of perfectly categorized non-road pixels, FP indicates the number of incorrectly categorized road pixels, and FN indicates the number of inaccurately categorized non-road pixels (Jan Dirk Wegner, Montoya-Zegarra, & Schindler, 2015). We calculated precision, recall, F1 score, IOU and OA based on confusion matrix parameters to calculate the efficiency of the introduced approach for road extraction. Table 2 exhibits the percentage of performance measures for calculating the accuracy of road extraction based on CNN model for each image separately.

Table 2

Precision, recall, f1 score, overall accuracy and IOU of the model for accuracy assessment

5. Discussion

As shown in Table 2, the percentage of precision, recall, F1 score, OA and IOU were 92.75%, 96.08%, 94.39%, 95.48%, and 89.37% respectively, for the image in Figure 3(b). The values were 93.64%, 96.19%, 94.90%, 95.91% and 90.29% for the image in Figure 3(d); 91.25%, 95.32%, 93.24%, 94.54% and 87.33% for the image in Figure 3(f); 90.48%, 95.10%, 92.73%, 94.10% and 86.44% for the image in Figure 3(h); and 87.35%, 93.91%, 90.51%, 92.21% and 82.66% for the image in Figure 3(k). As specified in Table 2, the proposed method obtained higher precision for road extraction from the image in Figure 3(d) than those in other figures for the entire assessment parameters. The same was observed for the image in Figure 3(b). Compared with the image in Figure 3(b), the proposed model could achieve higher precision and OA for the image in Figure 3(d). Therefore, the method could identify numerous pixels related to road section in this figure. In other words, the model recognized many pixels not belonging to the road region (FP) of the image in Figure 3(b). The accuracy of measurement parameters decreased slightly due to the presence of trees and car parking. The images in Figures 3(d) and (b) were captured in a non-complex area, where the road section was not completely covered by other objects, such as vegetation, cars, and buildings. Therefore, the suggested approach attained the highest accuracy for road extraction from the two images. By contrast, the image in Figure 3(f) was slightly surrounded by other features with a similar spectral reflectance to roads, especially car parking. As a result, these regions were identified as road sections and decreased the accuracy for road extraction of the image in this figure. The model failed to distinguish car parking sections from road sections in some parts. Therefore, in terms of performance measures, the extracted road of the image in Figure 3(f) had lower accuracy than that of the images in Figures 3(b) and (d). For the last image (Figure 3(k)), the precision of the introduced method for road extraction decreased dramatically compared with that for the images in other figures. For the entire accuracy assessment measures, the method achieved lower accuracy for extraction road class than that for the images in other figures. The image in Figure 3(k) was taken in a complex area, and the accuracy reduction in the figure was due to high similarities between road class and other features, such as shadow and car parking, in which the proposed method encountered difficulty in extracting road class and obtained less accuracy than for the images in other figures. In some parts, the road section has more similarity with car parking sections. Thus, the method produced plenty FPs in these sections. As a result, separating road regions from their environments was difficult because these sections had similar spectral reflectance to roads. Extraction of road parts was also difficult. Therefore, we applied OBIA to use spatial information for increasing the classification accuracy. However, using the OBIA, PCA, and CNN techniques concurrently determined that the suggested model had overall success for extracting road sections from orthophoto images. We performed PCA and OBIA to obtain additional information. The use of this information as input for the CNN model had resulted in an extremely precise road extraction. Figure 5 plots the precision, recall, f1 score, overall accuracy and IOU of the model for accuracy assessment of four different orthophoto images. All the images and performance measures are shown in x-axis and y-axis, respectively.

Fig. 5. Accuracy assessment factors of the suggested method for road extraction from four different orthophoto images

We compared the performance measure factors of this work with those of other works to show the advantages of the proposed approach in extracting road class from orthophoto images. We used five orthophoto images to run and show the effectiveness of the suggested method, whereas other works used several images. Therefore, we considered the average percentage of measurement factors for comparison. Jan D Wegner, Montoya-Zegarra, and Schindler (2013) developed a novel CRF based on PN-Potts model and represented a probabilistic presentation of network structure in the image for extracting road class from aerial orthoimages. They first segmented images into superpixels and then extracted a feature vector per superpixel. They fed the extracted feature vector to a random forest classifier to allocate a unary road probability for each superpixel. Next, promising candidate routes were generated, and superpixels were sampled randomly with high road possibility as seed nodes. Finally, superpixels of every candidate path created a higher-order category in CRF by connecting them with minimum cost paths. They performed their method on two various datasets, in which the average value of accuracy assessment measures was taken for comparison. Zhong, Li, Cui, and Jiang (2016) applied contemporary fully convolutional networks to extract road and building from high-spatial resolution images. They used Massachusetts dataset in training, validating, and testing the model and evaluated the accuracy of the proposed model using various parameters. They separated each image into nine uniformed 3×500×500-pixel image to make full use of the exiting pretrained models. The proposed model was directly fine-tuned on basis of FCN-16s-PASCAL model. They set the learning rate to 1×e-14 and the model was trained for 20,000 iterations. They found that the extraction accuracy rate of the suggested model improves remarkably by mixing the deep final-score layer with the shallow fine-grained pooling layer outcomes. They evaluated precision, recall, F1 score, and OA factors, which are taken for comparison with those of the proposed method in the present study. Wei, Wang, and Xu (2017) performed a road structure refined CNN (RSRCNN) method to extract road regions from aerial images. They designed deconvolutional and fusion layers in the architecture of RSRCNN to gain structured output of road extraction. For setting training, validation and test sets, they segmented every image into 16 nonoverlapping 375×375 images as an input to RSRCNN. Next, they applied RSRCNN in deep learning platform "Caffe" and for fine-tuning on their model, they used the pretrained of the 13 convolutional layers of VGG as the initial parameters. Then, a back propagation (BP) algorithm utilized for training RSRCNN model. For training RSRCNN, they used a new loss function based on geometric information of road structure, which is called road structure-based loss function. They evaluated the proposed method performance on basis of overall accuracy, F1, precision, and recall factors. The values are shown in Table 3 for comparison with those of our work.

Ardiyanto and Adji (2017) proposed a method for road part segmentation based on deep learning-based techniques, which is called deep residual coalesced convolutional network (RCC-Net). The RCC-net extracts relevant features by performing dimensionality reduction. For accuracy assessment, they applied precision, recall, F1, and OA measures. The average percentages of these factors are exhibited in Table 3 for comparison with those of the introduced work in this study. Table 3 shows the results achieved in this work compared with those in other works. In addition, we compared our model with modern deep learning methods such as generative adversarial networks (GANs) and fully convolutional networks (FCN) to show the validity and superior performance of the suggested model for road extraction from orthophoto images. Kestur, et al. (2018) proposed a model based on U-shaped FCN (UFCN) to extract road section from unmanned aerial vehicle (UAV) images. They evaluated precision, recall, f1 score and accuracy for classification performance, which are taken for comparison with those of the suggested approach in the current study. Q. Shi, Liu, and Li (2018) introduced a novel model of convolutional network based on end-to-end generative adversarial networks to extract road parts from remote sensing data with VHR images. They calculated measurement factors such as completeness (recall), correctness (precision) and quality percentage (equivalent to IOU) to evaluate the validity of proposed model for road extraction. Table 4 exhibits the average percentage of achieved results in the present work compared with new deep learning methods.

Table 3

Comparison of implementation factors of the suggested approach with other works (The best values are in bold)

Table 4

Comparison of implementation factors of the proposed method with other modern deep learning models (The best values are in bold)

Figure 6 plots the accuracy assessment factors for demonstrating the evident differences between the proposed model and other works. All the aforementioned works and corresponding values are shown in x-axis and y-axis, respectively. The percentages of precision, recall, and F1 for the first three works were considerably lesser than those of Xu (2018) and our work. Therefore, their methods were not as accurate as our work for road extraction. Given that Hu used encoder–decoder deep learning method for road extraction, it achieved high percentage in precision and f1 score factors with 93.4% and 93.7% respectively. However, for the remaining measurement factors (recall and OA), the proposed method obtained higher percentage than other methods. Therefore, the proposed approach was notably proficient for road extraction from orthophoto images.

Fig. 6. Accuracy measurement factors of the proposed method for road extraction compared with other works for road extraction

Furthermore, the accuracy assessment factors for showing the obvious differences between the suggested method in this study and other modern deep learning approached are plotted in Figure 7. Y-axis and x-axis show the corresponding values and mentioned works, respectively. As it observed, the percentages of precision and OA for the Kesture (2018) work are higher than those of Shi (2018) and our work. Therefore, their model was more accurate than others for road extraction. Given that Shi used GANs model for road extraction, it achieved less percentage in precision factor (88.31%), which means their model was not as accurate as the proposed method in this work (CNN) and Kesture (2018) work (FCN). The proposed method in this paper achieved higher percentage than other methods for the measurement factor (recall), which approved the suggested method was remarkably efficient for road extraction from orthophoto images.

Fig. 7. Accuracy assessment factors of the suggested model compared with other modern deep learning models for road extraction

6. Conclusion

In this study, our goal was to introduce a method to extract different types of roads from orthophoto images based on deep learning method (CNN). The proposed approach consists of the following steps: (1) geometric correction to assign a spatial coordinate system for the content of a map; (2) PCA for transforming a main correlated dataset into a considerably smaller collection of uncorrelated parameters, which describes nearly the entire information represented in the main dataset; and (3) OBIA to use spectral and spatial information, such as class relationship, size, and shape, for improving classification. All the obtained information was used as input for the combining CNN model and trivial opening for classification image into road and non-road sections and road extraction. We implemented the proposed model on five orthophoto images from different areas with dissimilar complexities to show the superiority of the model for road extraction. Performance measures such as precision, recall, F1, OA and IOU were calculated. The values achieved were 92.75%, 96.08%, 94.39%, 95.48%, and 89.37% for the image in Figure 3(b); 93.64%, 96.19%, 94.90%, 95.91% and 90.29% for the image in Figure 3(d); 91.25%, 95.32%, 93.24%, 94.54% and 87.33% for the image in Figure 3(f); and 90.48%, 95.10%, 92.73%, 94.10% and 86.44% for the image in Figure 3(h); and 87.35%, 93.91%, 90.51%, 92.21% and 82.66% for the image in Figure 3(k). These results confirmed the efficiency of the model for road extraction. In addition, the accuracy measurement factors of the suggested method were compared with those of other works. The plotted results verified the effectiveness of the method for road extraction from orthophoto images. The suggested model has some advantages, which make it suitable for road extraction from orthophoto images. The model can identify and extract not only junction and curved sections but also straight roads. The proposed method can also detect barriers, such as cars, tree shadows, and buildings. However, the results and images indicate that the model performance decreases when the complexity and the size of image increases. In other words, the method cannot extract road in complex areas where the road section is completely surrounded by other features. This inability is a disadvantage of the introduced model.

References

Aasen, H., Honkavaara, E., Lucieer, A., & Zarco-Tejada, P. (2018). Quantitative remote sensing at ultra-high resolution with uav spectroscopy: A review of sensor technology, measurement procedures, and data correction workflows. *Remote Sensing, 10*, 1091.

Abdollahi, A., Bakhtiari, H. R. R., & Nejad, M. P. (2018). Investigation of SVM and level set interactive methods for road extraction from google earth images. *Journal of the Indian Society of Remote Sensing, 46*, 423-430.

Abdollahi, A., Pradhan, B., & Shukla, N. (2019). Extraction of road features from UAV images using a novel level set segmentation approach. *International Journal of Urban Sciences*, 1-15.

Abdullahi, S., Pradhan, B., & Jebur, M. N. (2015). GIS-based sustainable city compactness assessment using integration of MCDM, Bayes theorem and RADAR technology. *J Geocarto International*, *30*, 365-387.

Alshehhi, R., & Marpu, P. R. (2017). Hierarchical graph-based segmentation for extracting road networks from high-resolution satellite images. *ISPRS Journal of Photogrammetry and Remote Sensing*, *126*, 245-260.

Ardiyanto, I., & Adji, T. B. (2017). Deep residual coalesced convolutional network for efficient semantic road segmentation. *IPSJ Transactions on Computer Vision Applications*, *9*, 6.

Audebert, N., Le Saux, B., & Lefèvre, S. (2017). Segment-before-detect: Vehicle detection and classification through semantic segmentation of aerial images. *Remote Sensing*, *9*, 368.

Bakhtiari, H. R. R., Abdollahi, A., & Rezaeian, H. (2017). Semi automatic road extraction from digital images. *The Egyptian Journal of Remote Sensing Space Science*, *20*, 117-123.

Bedawi, S. M., & Kamel, M. S. (2015). Road Detection in Urban Areas Using Random Forest Tree-Based Ensemble Classification. In *International Conference Image Analysis and Recognition, Niagara Falls, ON, Canada, 499-505. doi:*https://doi.org/10.1007/978-3-319-20801-5_55 (pp. 499-505).

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, *65*, 2-16.

Cheng, J., Ding, W., Ku, X., & Sun, J. (2012). Road Extraction from High-Resolution SAR Images via Automatic Local Detecting and Human-Guided Global Tracking. *International Journal of Antennas and Propagation*, *2012*, 1-10.

Comber, A. J., Harris, P., & Tsutsumida, N. (2016). Improving land cover classification using input variables derived from a geographically weighted principal components analysis. *ISPRS Journal of Photogrammetry and Remote Sensing*, *119*, 347-360.

Grinias, I., Panagiotakis, C., Tziritas, G., & sensing, r. (2016). MRF-based segmentation and unsupervised classification for building and road detection in peri-urban areas of high-resolution satellite images. *J ISPRS journal of photogrammetry*, *122*, 145-166.

He, C., Liao, Z.-x., Yang, F., Deng, X.-p., & Liao, M.-s. (2012). Road extraction from SAR imagery based on multiscale geometric analysis of detector responses. *J IEEE Journal of Selected Topics in Applied Earth Observations Remote Sensing*, *5*, 1373-1382.

Henry, C., Azimi, S. M., & Merkle, N. (2018). Road Segmentation in SAR Satellite Images With Deep Fully Convolutional Neural Networks. *IEEE Geoscience Remote Sensing Letters*, 1-5.

Hu, F., Xia, G.-S., Hu, J., & Zhang, L. (2015). Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sensing*, *7*, 14680-14707.

Kestur, R., Farooq, S., Abdal, R., Mehraj, E., Narasipura, O., & Mudigere, M. (2018). UFCN: A fully convolutional neural network for road extraction in RGB imagery acquired by remote sensing from an unmanned aerial vehicle. *Journal of Applied Remote Sensing*, *12*, 016020.

Khesali, E., Zoej, M. J. V., Mokhtarzade, M., & Dehghani, M. (2016). Semi automatic road extraction by fusion of high resolution optical and radar images. *Journal of the Indian Society of Remote Sensing*, *44*, 21-29.

Kirthika, A., & Mookambiga, A. (2011). Automated road network extraction using artificial neural network. In *International Conference on Recent Trends in Information Technology (ICRTIT), Chennai, Tamil Nadu, India, 1061-1065. doi:*https://doi.org/10.1109/ICRTIT.2011.5972323 (pp. 1061-1065).

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (Vol. 1, pp. 1097-1105).

Maggiori, E., Tarabalka, Y., Charpiat, G., & Alliez, P. (2017). Convolutional neural networks for large-scale remote-sensing image classification. *IEEE Transactions on Geoscience Remote Sensing*, *55*, 645-657.

Miao, Z., Shi, W., Gamba, P., & Li, Z. (2015). An object-based method for road network extraction in VHR satellite images. *J IEEE journal of selected topics in applied earth observations and remote sensing*, *8*, 4853-4862.

Ng, S. (2017). Principal component analysis to reduce dimension on digital image. *Procedia computer science*, *111*, 113-119.

Nogueira, K., Dos Santos, J. A., Fornazari, T., Silva, T. S. F., Morellato, L. P., & Torres, R. d. S. (2016). Towards vegetation species discrimination by using data-driven descriptors, 9th IAPR Workshop on Pattern Recogniton in Remote Sensing (PRRS), 1-6. https://doi.org/10.1109/PRRS.2016.7867024

Panboonyuen, T., Vateekul, P., Jitkajornwanich, K., & Lawawirojwong, S. (2017). An enhanced deep convolutional encoder-decoder network for road segmentation on aerial imagery. In *International Conference on Computing and Information Technology , Springer, 191-201. https://doi.org/10.1007/978-3-319-60663-7_18*.

Sameen, M. I., Pradhan, B., & Aziz, O. S. (2018). Classification of very high resolution aerial photos using spectral-spatial convolutional neural networks. *Journal of Sensors, 2018*, 1-12.

Sghaier, M. O., & Lepage, R. (2016). Road extraction from very high resolution remote sensing optical images based on texture analysis and beamlet transform. *J IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*, 1946-1958.

Shi, Q., Liu, X., & Li, X. (2018). Road detection from remote sensing images by generative adversarial networks. *IEEE access*, *6*, 25486-25494.

Shi, W., Miao, Z., & Debayle, J. (2014). An integrated method for urban main-road centerline extraction from optical remotely sensed imagery. *IEEE transactions on geoscience and remote sensing, 52*, 3359-3372.

Srivastava, N., Hinton, G., Krizhevsky, A., Krizhevsky, A., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, *15*, 1929-1958.

Sujatha, C., & Selvathi, D. (2015). Connected component-based technique for automatic extraction of road centerline in high resolution satellite images. *EURASIP Journal on Image Video Processing, 2015*, 8.

Unsalan, C., & Sirmacek, B. (2012). Road network detection using probabilistic and graph theoretical methods. *IEEE Transactions on Geoscience and Remote Sensing, 50*, 4441-4453.

Wang, J., Song, J., Chen, M., & Yang, Z. (2015). Road network extraction: A neural-dynamic framework based on deep learning and a finite state machine. *International Journal of Remote Sensing*, *36*, 3144-3169.

Wang, W., Yang, N., Zhang, Y., Wang, F., Cao, T., & Eklund, P. (2016). A review of road extraction from remote sensing images. *Journal of traffic transportation engineering*, *3*, 271-282.

Wegner, J. D., Montoya-Zegarra, J. A., & Schindler, K. (2013). A higher-order CRF model for road network extraction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA* (pp. 1698-1705).

Wegner, J. D., Montoya-Zegarra, J. A., & Schindler, K. (2015). Road networks as collections of minimum cost paths. *ISPRS Journal of Photogrammetry Remote Sensing, 108*, 128-137.

Wei, Y., Wang, Z., & Xu, M. (2017). Road structure refined CNN for road extraction in aerial image. *IEEE Geoscience Remote Sensing Letters*, *14*, 709-713.

Weng, Q. (2012). Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends. *Remote Sensing of Environment*, *117*, 34-49.

Xia, W., Zhang, Y.-Z., Liu, J., Luo, L., & Yang, K. (2018). Road Extraction from High Resolution Image with Deep Convolution Network—A Case Study of GF-2 Image. *the 2nd International Electronic Conference on Remote Sensing, china*, *2*, 325.

Xu, Y., Chen, Z., Xie, Z., & Wu, L. (2017). Quality assessment of building footprint data using a deep autoencoder network. *International Journal of Geographical Information Science, 31*, 1929-1951.

Xu, Y., Xie, Z., Feng, Y., & Chen, Z. (2018). Road Extraction from High-Resolution Remote Sensing Imagery Using Deep Learning. *Remote Sensing*, *10*, 1461.

Youssef, A. M., Sefry, S. A., Pradhan, B., Alfadail, E. A., & Risk. (2016). Analysis on causes of flash flood in Jeddah city (Kingdom of Saudi Arabia) of 2009 and 2011 using multi-sensor remote sensing data and GIS. *J Geomatics, Natural Hazards, 7*, 1018-1042.

Zhang, J., Chen, L., Wang, C., Zhuo, L., Tian, Q., & Liang, X. (2017). Road recognition from remote sensing imagery using incremental learning. *IEEE Transactions on Intelligent Transportation Systems*, *18*, 2993-3005.

Zhang, Z., Liu, Q., & Wang, Y. (2018). Road extraction by deep residual u-net. *J IEEE Geoscience Remote Sensing Letters*, *15*, 749-753.

Zhong, Z., Li, J., Cui, W., & Jiang, H. (2016). Fully convolutional networks for building and road extraction: Preliminary results. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 1591-1594). Beijing, China, 1591-1594. https://doi.org/10.1109/IGARSS.2016.7729406

Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., & Oliva, A. (2014). Learning deep features for scene recognition using places database. In *Proceedings of the 27th International Conference on Neural Information Processing Systems, Montreal, Canada, 1,487-495.*: MIT Press.

Figure

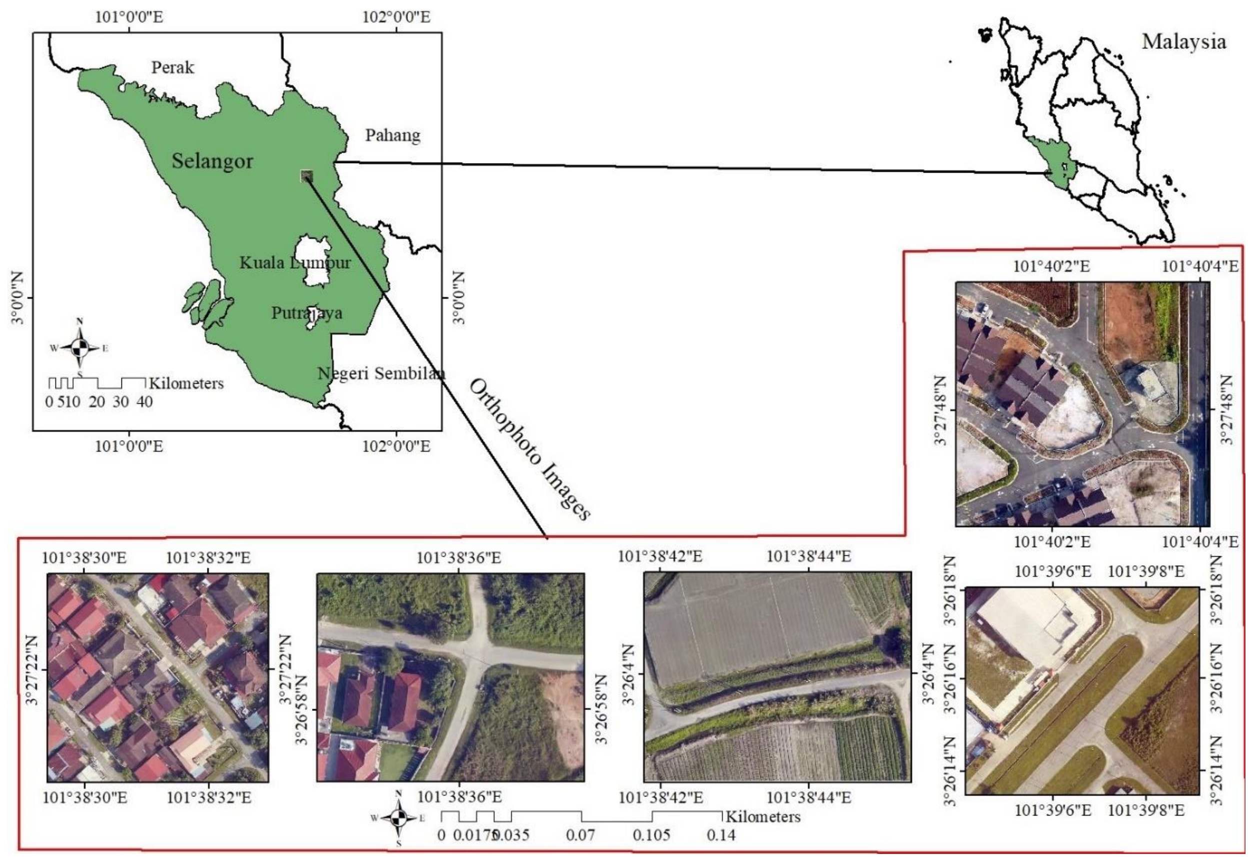

Fig. 1. Study area and the orthophoto images used in this study

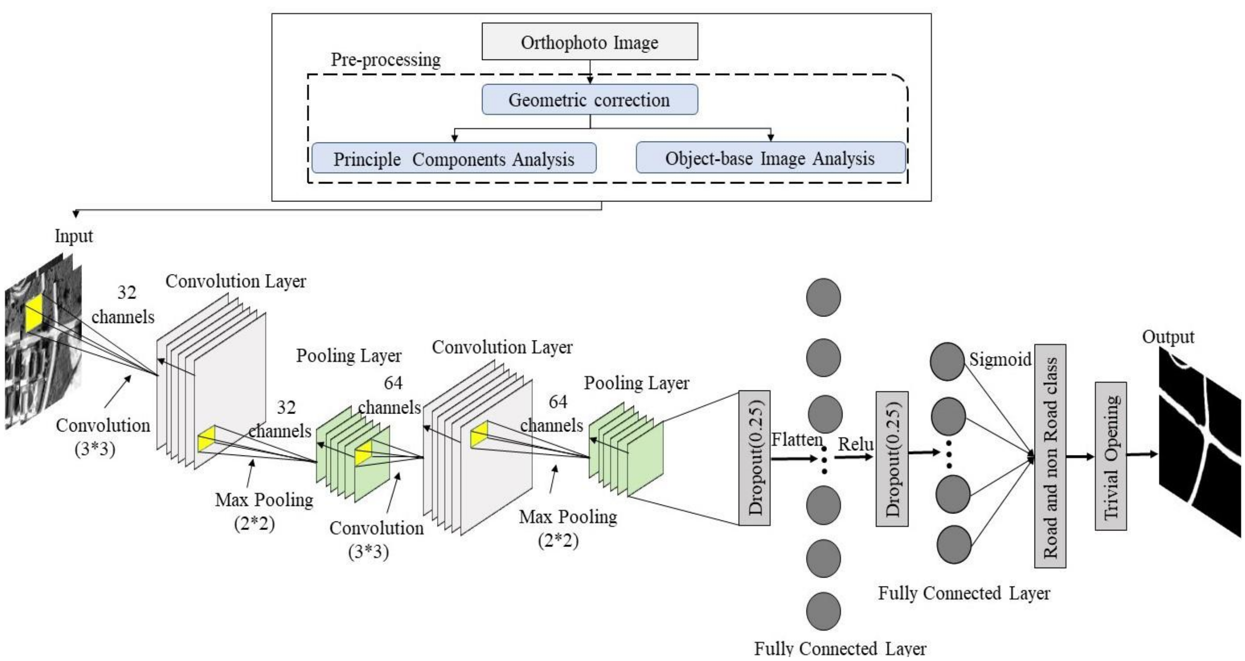

Fig. 2. Proposed CNN model architecture

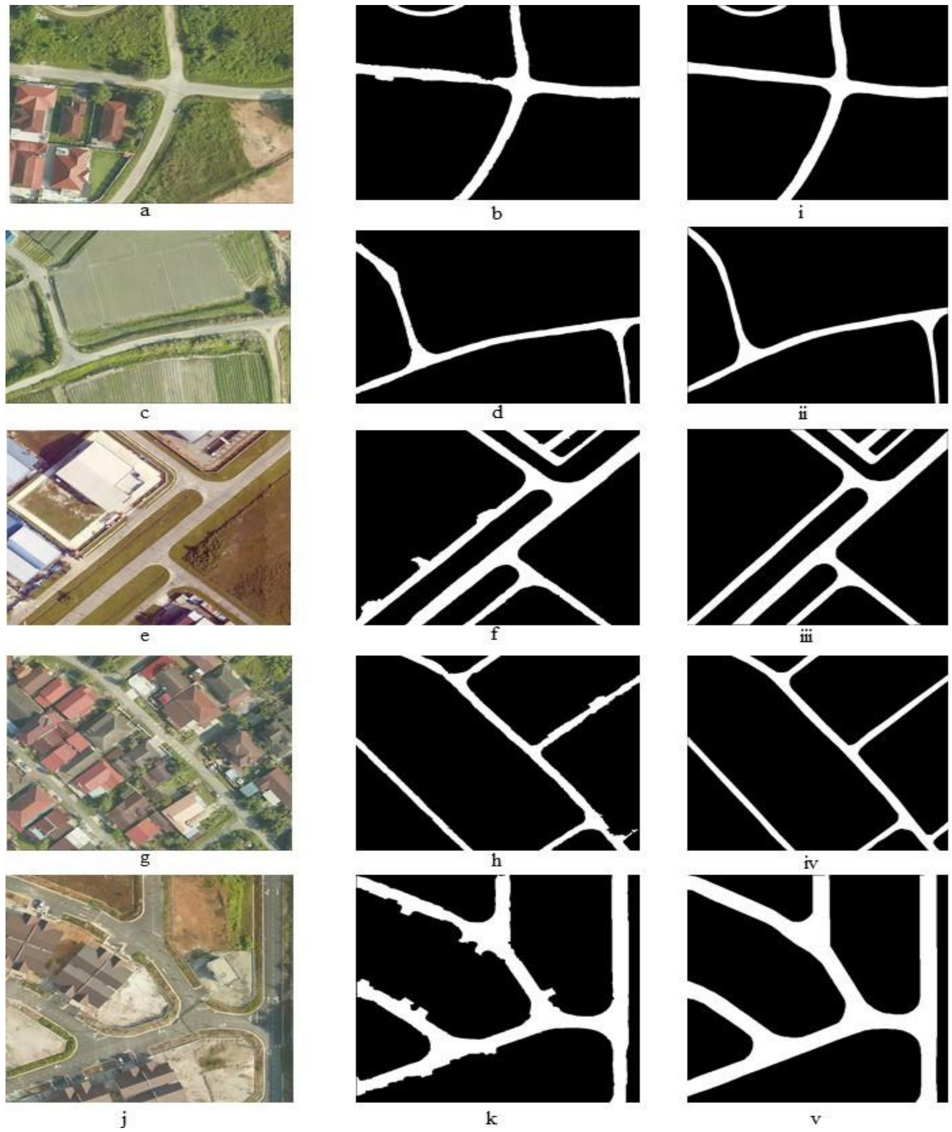

Fig. 3. Results of road extraction using the suggested CNN model: original orthophoto images (a, c, e , g, j), extracted road sections (b, d, f, h, k) and target road map (i, ii, iii, iv, v)

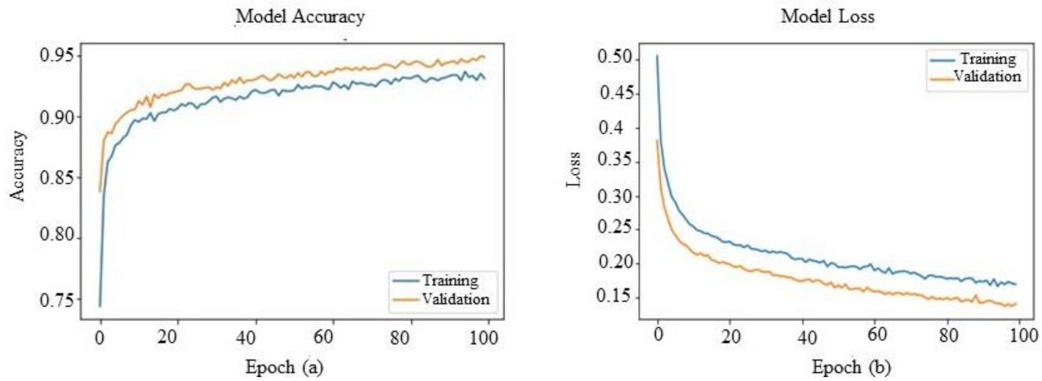

Fig. 4. Model accuracy (a) and model loss (b) of the proposed CNN model

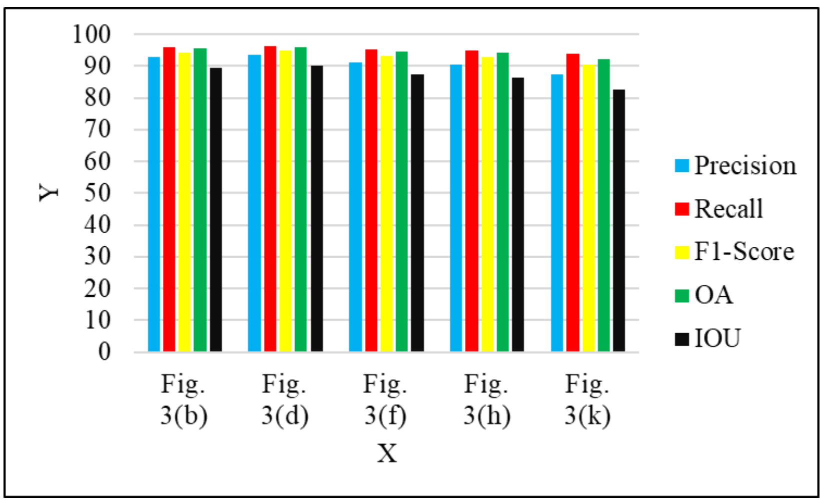

Fig. 5. Accuracy assessment factors of the suggested method for road extraction from four different orthophoto images

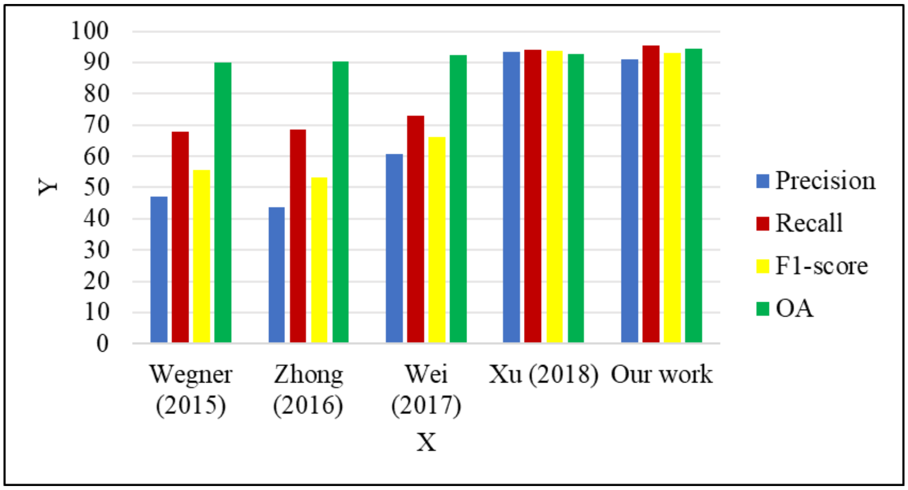

Fig. 6. Accuracy measurement factors of the proposed method for road extraction compared with other works for road extraction

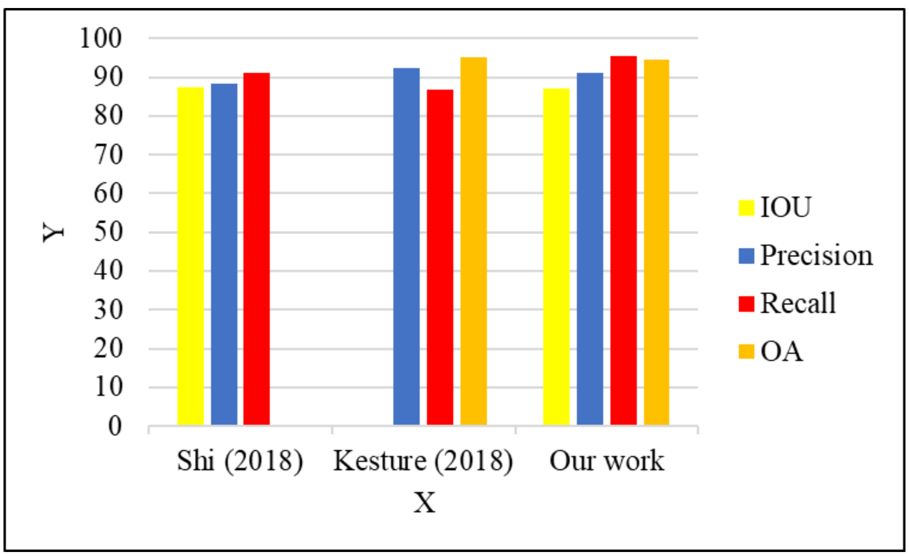

Fig. 7. Accuracy assessment factors of the suggested model compared with other modern deep learning models for road extraction

Table

Table 1. Summary of the proposed CNN model layers

Layer (type)	Output Shape	Parameter
Input	(None, 5, 5, 1)	0
Convolution2D	(None, 5, 5, 32)	320
Max-Pooling	(None, 2, 2, 32)	0
Convolution2D	(None, 2, 2, 64)	18496
Max-Pooling	(None, 1, 1, 64)	0
Dropout	(None, 1, 1, 64)	0
Flatten	(None, 64)	0
Dense	(None, 256)	16640
Dropout	(None, 256)	0
Dense (sigmoid)	(None, 2)	514

Table 2. Precision, recall, f1 score, overall accuracy and IOU of the model for accuracy assessment

Figures	Precision (%)	Recall (%)	F1 (%)	OA (%)	IOU (%)
Fig. 3(b)	92.75	96.08	94.39	95.48	89.37
Fig. 3(d)	93.64	96.19	94.90	95.91	90.29
Fig. 3(f)	91.25	95.32	93.24	94.54	87.33
Fig. 3(h)	90.48	95.10	92.73	94.10	86.44
Fig. 3(k)	87.35	93.91	90.51	92.21	82.66

Table 3. Comparison of implementation factors of the suggested approach with other works (The best values are in bold)

Methods	Precision (%)	Recall (%)	F1 score (%)	OA (%)
Wegner (2015)	47.1	67.9	55.6	89.9
Zhong (2016)	43.5	68.6	53.2	90.4
Wei (2017)	60.6	72.9	66.2	92.4
Xu (2018)	**93.4**	94	**93.7**	92.6
Our work	91.09	**95.32**	93.15	**94.44**

Table 4. Comparison of implementation factors of the proposed method with other modern deep learning models (The best values are in bold)

Methods	Precision (%)	Recall (%)	IOU (%)	OA (%)
Shi (2018)	88.31	91.01	**87.32**	-
Kestur (2018)	**92.5**	86.8	-	**95.2**
Our work	91.09	**95.32**	87.21	94.44

Click here to access/download
Supplementary Material
graphical abstract.jpg