ISSN: 2229-6956 (ONLINE)
DOI: 10.21917/ijsc.2023.0408

A DEEP LEARNING APPROACH FOR ROAD EXTRACTION FROM REMOTE SENSING IMAGERY

Md. Abdul Alim Sheikh[1], Tanmoy Maity[2] and Alok Kole[3]

[1]*Department of Electronics and Communication Engineering, Aliah University, India*
[2]*Department of Electrical Engineering, Indian Institute of Technology, Dhanbad, India*
[3]*Department of Electrical Engineering, RCC Institute of Information Technology, India*

Abstract

In recent years, Deep Learning (DL) is proving very successful set of tools for several image analysis, segmentation, and classification tasks. In this paper an automated Deep Learning Architecture (DLA) called the Deep Belief Neural Networks (DBN) stacked by Restricted Boltzmann Machines (RBMs), is designed, implemented, and experimentally evaluated for extracting semantic maps of roads in Remote Sensing (RS) images. Representative features are extracted by unsupervised pre-training of DBN and supervised fine-tuning phase. A Logistic Regression (LR) is added to the end of feature learning system to constitute a DBN-LR architecture. This LR classifier is employed to fine-tune the whole pre-trained network in a supervised way and classifies the patches from RS images. The features extracted from the image patches are fed to the architecture as input and it produces the class labels as a probability matrix as either a positive sample (road) or a negative sample (non-road). A math morphology algorithm is used to improve DBN performance during post processing. Experiments are conducted on a dataset of 970 RS scene images of urban and suburban areas to demonstrate the performance of the proposed network architecture. The proposed deep model resulted in an Overall Accuracy (OA) of 96.57% and F1-score of 0.9552. The results of the proposed architectures are compared with those of other network architectures. Experimental results demonstrate the effective performance of the proposed method for extracting roads from a complex scene.

Keywords:

Remote Sensing Imagery, Road Networks Extraction, Deep Learning, Deep Belief Network, Restricted Boltzmann Machine

1. INTRODUCTION

With the development of high-resolution and high-speed imaging sensors, RS imagery is becoming increasingly available nowadays [1]. Extraction of roads from RS images plays important roles for supporting several government activities and various Geographical Information System (GIS) applications such as map generation and update, urban planning, traffic management, automated vehicle navigation and guidance, emergency response, disaster management etc., [2]-[3]. In recent years, road extraction has also been extensively studied for terrain classification and ground vehicle navigation [4]. A stark increase in the amount of RS imagery available in recent years has made the relevant information extraction and interpretation of this high volume of data a challenging problem at scale. The traditional process of manually updating a road database is very tedious and time-consuming [2]. For this reason, automatic extraction of roads in high resolution remote sensing imagery has attracted a lot of attention in the photogrammetry and remote sensing community [1, 2].

Based on recent advances, DL is proving to be a very successful set of tools for several image analysis, segmentation, and classification tasks [5, 51]. In the wake of this success and increased availability of data and computational resources, the use of deep learning is finally taking off in remote sensing as well [6, 51]. DL methods have been used for road extraction tasks as well due to its superiority in modelling non-linear relationships among variables [52]. Recently, DBNs has gained lot of interest because of its efficient layer wise learning strategy. The DBN is constructed using a stack of probabilistic model called a RBM, which is used to extract layer of features at a time [7].

The main goal of this paper is to design, implement and experimentally evaluate an automated Deep Learning Architecture (DLA), the DBN stacked by RBMs, for extracting semantic maps of roads in RS images. First, the original images are pre-processed to cope with input images of varying quality, resolution, and channels. The input features of $M*N$ dimension are used as input in training the DBN model. A two-phase training has been done sequentially by: 1) pre-training of stacked RBM module in a layer-wise manner using unsupervised CD learning algorithm with 1 step of Gibbs sampling (CD-1) 2) fine-tuning, supervised learning with a classifier. The pre-trained RBMs are stacked up with a extra output layer LR classifier [9]. Finally, the whole DBN is fine-tuned in a supervised way with back-propagation algorithm. The specific structure of each layer is determined through experiments. All the pre-processed samples were treated as the input of the model, and the output of the trained model was the two-category classification maps which discriminate between "road" and "non-road" regions. The main advantage of training the model using DBN because of its unsupervised feature learning in pre-training which makes it prominent from other methods. Moreover, the proposed method is not more time consuming and outputs high accuracy. The contributions of the paper are summarized below.

- The major contribution of this work is the road extraction model based on the construction of deep network architecture DBN and the introduction of the RBM as the feature extractor.
- At the same time, this work designed the logistic unit in the expansive part of the model, as feature classifier.
- Compare different neural network structures specified in existing literature.
- Provide a feasibility study for application of deep learning method in the field of RS image analysis.

The paper is organized as follows. Section 1 covers the introduction. Section 2 briefly reviews the related works. Section 3 provides an overview of our approach with the DBN architecture stacked by RBM for roads extraction from RS imagery; experiments and discussions are reported in Section 4. Finally, conclusion and future direction is covered in Section 5.

2. RELATED WORK

Many different techniques have been proposed to deal with the road network extraction from RS imagery. Some comprehensive reviews on road objects extraction from RS imagery can be found in [10, 11]. Most of the road extraction techniques found in the literature can be clustered in two groups: classical methods and DL methods [3, 46]. DL uses much high-level and multi-scale information where classical methods use low-level features for road extraction [3, 6]. Road extraction from RS imagery is a classification problem including supervised and unsupervised classification methods [10]. Supervised classification methods include Artificial Neural Networks (ANN), and support vector machine (SVM), MRF, Maximum Likelihood (ML), and Decision Tree [11]. Unsupervised methods usually extract roads using clustering algorithms which do not need any training data. The most common unsupervised techniques used for road extraction from remote sensing images are k-means, spectral clustering, mean shift [12], line segment match method and graph theory [13, 14]. Some other classical methods include knowledge-based methods, mathematical Morphology and Active Contour model [10]. Road extraction on knowledge-based methods mainly depends on their own characteristics, such as spectrum and context features, and other knowledge related to their theory [15, 16]. Mathematical morphology has been used for road detection from RS images in [17]. Various active contour-based approaches have been proposed for road extraction from RS images and they are represented by snake and Level Set [2]. Some other classical methods for road extraction used methods such as perceptual grouping [49], and dynamic programming [54], Model based approach [18], Region Competition algorithm [19], Particle Filtering [20], Differential Geometry [21], object-based techniques [22], particle swarm optimization and fuzzy logic and fuzzy-object based ant colony optimization [23], texture analysis and beamlet transform [24], Tensor Voting and Geodesic method [25, 26] etc.

Recently, the uses of DL methods are taking off to extract roads from RS data [27, 28, 32, 51]. The Convolutional Neural Network (CNN) [29, 30] and the Fully Convolutional Neural (FCN) [31, 40] network are widely used for these purposes. Li et al. [32] used the CNN to identify whether a pixel belongs to a road, then the centerline of the road is extracted through some refinement process. Han et al. [55] proposed a method for Target Recognition Based on Deep Convolution Neural Network on Remote Sensing Image. D. Zhang et al. [56] used CNN for hyperspectral image classification using spatial and edge features. In another work, Saito et al. [30] used the CNN to extract roads and buildings simultaneously from RS images. Ruyi Liu et al. [33] proposed an automatic method to extract robust features based on CNN. Then Gabor filters and multiple directional non-maximum suppression are integrated to obtain accurate road network. Zhao et al. [34] developed an object-based deep learning technique for road extraction from RS images. The DBN model has been used to detect roads in airborne remote sensing images by Mnih and Hinton [35]. Sarhan et al. [36] proposed a framework called the CNN-Cellular Neural Network to extract roads from images using spectral and geometric characteristics of roads. LVQ ANN was used to extract roads from satellite images by Wijesingha [37]. Alshehhi, R. et al. [29] developed a single patch-based CNN architecture for roads and buildings extraction from RS data. In addition, CNN features and low-level features of roads and buildings are fused to improve the performance. In [27], authors proposed a neural-dynamic tracking framework to extract road networks based on deep Convolutional neural networks (DNN) and a finite state machine (FSM). Richer Convolutional Features (RCF) network are also presented in [11, 38].

Although the CNN has achieved certain results in road extraction, the local processing strategy still yields many errors in the extraction results. Compared with CNN, the features extracted from FCN network are high-level features that contain more abstract semantic information [6, 39]. Zhong et al. [40] proposed a technique using the FCN for road and buildings extraction from the RS data. Cheng et al. [41] proposed a cascaded end-to-end CNN (Cas-Net) for road network detection and centerline extraction. Fu et al. [31] developed a framework using the FCN and adopted Conditional Random Fields (CRFs) to refine output results. Panboonyuen et al. [6] proposed a method to extract high-level information of roads more accurately using Exponential Linear Unit (ELU), a deep convolutional encoder-decoder network (DCED), Landscape Metrics (LMs) and CRFs. Ramesh et al. [42] developed a U-shaped FCN (UFCN) to extract road by a stack of convolutions followed by deconvolutions with skip connections. Zhang et al. [28] proposed a method that combines the deep residual networks (ResNet) and U-Net for road area extraction. In [50], Bastani et al. developed a CNN- based iterative graph construction method called Roadtracer, which could directly identify road networks from aerial images.

Though many DL techniques have been developed for image segmentation and road extraction, they still suffer from one disadvantage. Existing methods have not fully exploited the spatial contexts of road images and may fail to extract in depth features. They are also harder to train, learning can be very slow with multiple hidden layers and overfitting can also be a serious problem for both generative models and discriminative models [7].

In this paper, we aim to tackle the problems by designing a DBN stacked by RBM which can be used to extract the in-depth features by efficient layer-by-layer learning strategy of the original image for road network extraction. Successful spatio-temporal mapping features can be extracted by using the proposed DBN architecture, to improve the accuracy of the classification. The proposed learning method for DBNs solves the inference problem and significantly decreases both the time taken for training and the overfitting.

3. PROPOSED ALGORITHM

A deep learning model is proposed to extract road network from remote sensing imagery. The data flow diagram of the proposed scheme is shown in Fig.1.

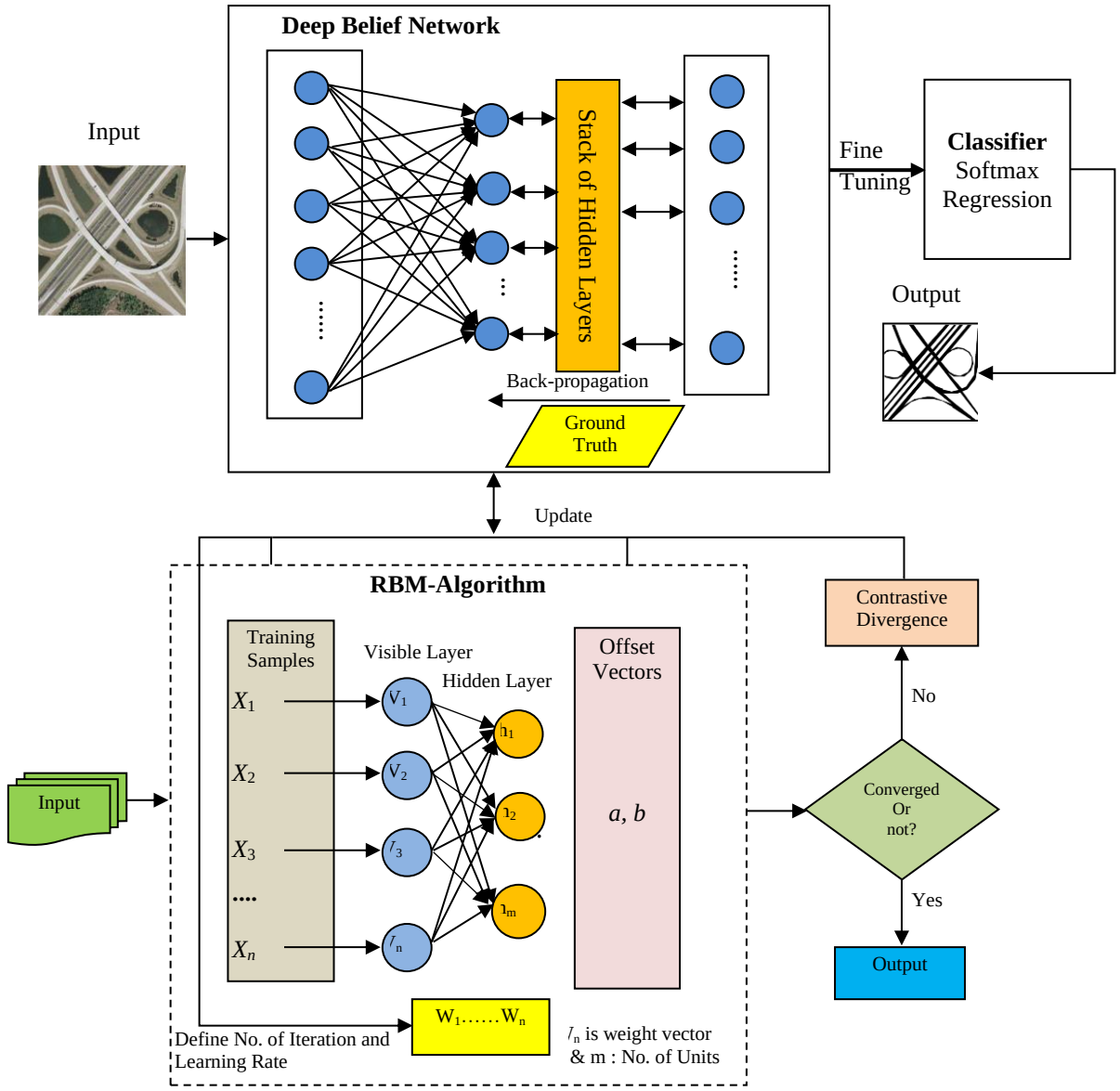


Fig.1. Proposed Scheme for Road Extraction

3.1 PRE-PROCESSING STEP

A pre-processing pipeline is adopted to cope with input images of varying quality, resolution, and channels to remove a large amount of noises and undesirable objects.

(a) (b) (c) (d)

Fig.2. (a) Input Image of suburban area (b) Result of bilateral Filtering (c) Input Image of Urban Area (d) Result of bilateral Filtering

The input images are filtered with the nonlinear bilateral filter [4] to minimize false alarms. Bilateral filter performs nonlinear smoothing on images by keeping the edge information. Fig.2(b) and 2(d) show the pre-processing result after bilateral filtering on the image shown in Fig.2(a) and 2(c).

3.2 DEEP BELIEF NEURAL NETWORK MODEL

DBN is composed by stacked of RBMs one by one each of which has its own visible layer and output layer [44]. The input image is applied as input to the DBN. The proposed DBN is used to extract the in-depth features by efficient layer-by-layer learning strategy of the original image for road network extraction. An overview of the network is shown in Fig.3.

3.2.1 Overview of Restricted Boltzmann Machine:

The topology of RBM is a two-layer stochastic graph which consists of one visible and one hidden layer [7] [8]. The input is fed to the visible layer represented by **v**, and the hidden layer represented by **h** is used to reconstruct the input as close to as possible**.** The visible layer is connected to the hidden layer, while there is no connection between the neurons within the same layer. An illustration of an individual RBM is given in Fig.3.

Given a training set $S=\{\mathbf{v}^1,\mathbf{v}^2,\mathbf{v}^3\ldots..,\boldsymbol{v}^{\boldsymbol{N}}\}$ that contains N samples, where $v^r=\left[v_1^r,v_2^r,v_3^r,...,v_n^r\right], r=1,2,...,N$ is r^{th} training sample. Then, RBM can be considered as an energy model [7]. Given a set of network states (**v, h**), the energy of certain joint configuration of the two layers is given by

$$E_\theta(\mathrm{v},\mathrm{h})=-\sum_{i=1}^{n}a_i v_i-\sum_{j=1}^{m}b_j h_j-\sum_{j=1}^{n}\sum_{j=1}^{m}v_j h_j w_{ij} \tag{1}$$

where θ denotes the parameters (i.e., **W,a,b**); $\boldsymbol{a}=[a_1,a_2,\ldots..a_n]$ is the bias of visible layer, $\boldsymbol{b}=[b_1,b_2,\ldots..b_m]$ is the hidden layer bias, and $W=\left[w_{ij}\right]$ denotes the weights between visible unit i and hidden unit j respectively.

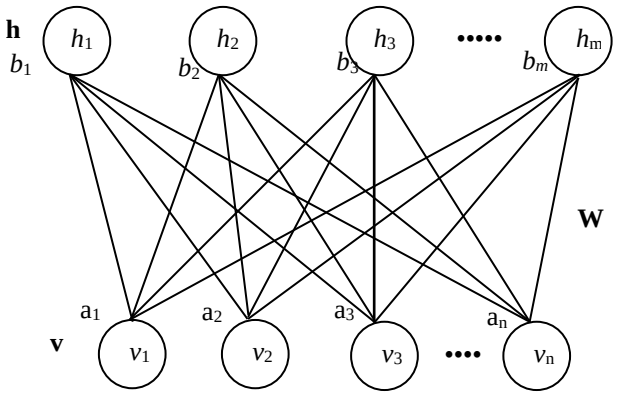


Fig.3. Structure of Restricted Boltzmann Machine (RBM)

Then, the joint probability distribution of a set of visible and hidden states can be obtained by:

$$p_\theta(v,h)=\frac{1}{Z_\theta}e^{-E_\theta(v,h)} \tag{2}$$

where Z_θ is the partition function which is given by:

$$Z_\theta=\sum_{v,h}e^{-E_\theta(v,h)}$$

Accordingly, the probability of the visible vector being assigned as **v** can be obtained by summing over all possible hidden vectors as follows:

$$p(v)=\frac{1}{Z_\theta}\sum_{h}e^{-E_\theta(v,h)} \tag{3}$$

Eq.(1)- Eq.(3) form a statistical mechanics model and can be used in the training of RBM.

RBM learning algorithms are based on gradient ascent on the log-likelihood. The log-likelihood of the model (3) for all training vector **v** is:

$$\begin{aligned}\ln L_{\theta,\mathrm{v}}&=\ln\prod_v p(\mathrm{v})=\ln\frac{1}{Z_\theta}\sum_h e^{-E_\theta(\mathrm{v},\mathrm{h})}\\&=\ln\sum_{\mathrm{h}}e^{-E_\theta(\mathrm{v},\mathrm{h})}-\ln\sum_{\mathrm{v},\mathrm{h}}e^{-E_\theta(\mathrm{v},\mathrm{h})}\end{aligned} \tag{4}$$

The goal is to find a set of parameters which can make $\ln\prod_v p(\mathrm{v})$ to its maximum. The parameters of RBM network are adjusted according to the principle of maximum likelihood. Maximizing the likelihood is the same as maximizing the log-likelihood function for all vectors **v** which is given by:

$$\ln L_{\theta,S}=\ln\prod_v p(\mathrm{v}^r)=\sum_{r=1}^{N}\ln p(\mathrm{v}^r) \tag{5}$$

The purpose of training RBM is to get the optimal value of parameter θ, that is:

$$\theta^*=\arg\max_\theta\left(\ln L_{\theta,S}\right) \tag{6}$$

where θ^* is the optimal value that makes the free energy of RBM system be minimum. Here the gradient decent method is used to find the maximum value of gradient $\ln L_{\theta,S}$ with respect to the parameter θ, we have

$$\frac{\partial\ln L_{\theta,S}}{\partial\theta}=\sum_{r=1}^{N}\frac{\partial\ln p(v)}{\partial\theta} \tag{7}$$

From Eq. (7), w.r.t. the parameters (**W, a, b**) we get

$$\frac{\partial\ln L_{\theta,S}}{\partial w_{i,j}}=\sum_{r=1}^{N}\left[p\left(h_j=1\middle|v^r\right)v_i^r-\sum_v p(v)\,p\left(h_j=1\middle|v\right)v_i\right] \tag{8}$$

$$\frac{\partial\ln L_{\theta,S}}{\partial a_i}=\sum_{r=1}^{N}\left[v_i^r-\sum_v p(v)v_i\right] \tag{9}$$

$$\frac{\partial\ln L_{\theta,v}}{\partial b_j}=\sum_{r=1}^{N}\left[p\left(h_j=1\middle|v^r\right)-\sum_v p(v)\,p\left(h_j=1\middle|v\right)\right] \tag{10}$$

where i = 1,2,3...n (No. of visible neuron), j =,2,3...m (No. of hidden neutrons)

Considering that there are no direct connections between the hidden units, the conditional probability of the binary state of unit h_j being set to 1 given visible vector **v** can be calculated as

$$p\left(h_j=1\middle|v\right)=sigmoid\left(b_j+\sum_{i=1}^{n}w_{i,j}v_i\right) \tag{11}$$

Also given a hidden vector **h**, the probability of the visible unit being 1 could be obtained as:

$$p\left(h_j=1\middle|h\right)=sigmoid\left(a_i+\sum_{i=1}^{m}w_{i,j}h_j\right) \tag{12}$$

Due to the high computational complexity of $\sum_v$, i.e., $O(2^n+2^m)$ updating the parameters based on these gradient formulas is not feasible. An efficient approximation method CD algorithm proposed by Hinton [7] has been adopted here. The gradients of the log-likelihood w. r. t parameters θ {W, a, b) for the training pattern **v** is then approximated by CD algorithm as follows:

$$\frac{\partial\ln L_{\theta,S}}{\partial w_{i,j}}\approx\sum_{r=1}^{N}\left[p\left(h_j=1\middle|v^{(r,0)}\right)v_i^{(r,0)}-p\left(h_j=1\middle|v^{(r,k)}\right)v_i^{(r,k)}\right] \tag{13}$$

$$\frac{\partial\ln L_{\theta,S}}{\partial a_i}\approx\sum_{r=1}^{N}\left[v_i^{(r,0)}-v_i^{(r,k)}\right] \tag{14}$$

$$\frac{\partial\ln L_{\theta,v}}{\partial b_j}=\sum_{r=1}^{N}\left[p\left(h_j=1\middle|v^{(r,0)}\right)-p\left(h_j=1\middle|v^{(r,k)}\right)\right] \tag{15}$$

where K is the number of sampling times in CD algorithm equal to 1 or 0 represents the starting point of sampling. Now, the parameters θ with (W, a, b) are updated using Eq.(13) to Eq.(15) as outlined in algorithm 1. An individual RBM can be trained efficiently using the learning rules in Eq.(13)-Eq.(15). The

procedures described above is the pre-training stage of the DBN and is performed unsupervised manner.

Algorithm 1: Parameter Updating Algorithm

Input: RBM; Training Set *S*
No. of Hidden Neurons *m*;
No. of Iteration: *t*;
Output: Gradient approximation Δw_{ij}, Δa_i, Δb_j for i=1,2,….n and j=1,2,…..m
Estimated parameters **(W, a, b)**

1. Initialization of Weight Matrix **W**, the visible layer bias **a** and the hidden layer bias **b** based on *S* and *n*
2. **do**
3. $v^{(0)} \leftarrow v$
4. **for** t=1 to k-1 **do**
5. **for** i = 1,2,3…n **do** sample $h_j^{(t)} \approx p\left(h_j \left| v^{(t)}\right.\right)$
   for j = 1,2,3……m **do** sample $v_i^{(t+1)} \approx p\left(v_i \left| h^{(t)}\right.\right)$
6. **for** i=1,2,3,...n; j=1,2,3,4.....m **do**
7. $\Delta w_{ij} \leftarrow \Delta w_{ij} + p\left(h_j = 1 \left| v^{(0,0)}\right.\right), v_i^{(0,0)} - p\left(h_j = 1 \left| v^{(0,k)}\right.\right), v_i^{(0,k)}$
8. $\Delta a_i \leftarrow \Delta a_i + v_i^{(0,0)} - v_i^{(0,k)}$
9. $\Delta b_j \leftarrow \Delta b_j + p\left(h_j = 1 \left| v^{(0,0)}\right.\right) - p\left(h_j = 1 \left| v^{(0,k)}\right.\right)$
10. **end**
11. **Return (**updated value**)**

Training of the DBN Model

DBN is one of the representative deep learning methods that composed by stacked of RBMs one by one each of which has its own visible layer and output layer [44]. As illustrated in Fig.4, the input is fed to the lower RBM. The hidden layer output of each layer is used as input to the visible layer of the subsequent RBM stage [8]. The joint probability distribution between the input data v and the layer hidden layer h^k in the visible layer is shown in Eq.(16).

$$p\left(v, h^1, h^2, h^3, ..., h^l\right) = \left(\sum_{k=0}^{l-2} p\left(h^{1k} \left| h^{k+1}\right.\right)\right) p\left(h^{l-1}, h^l\right) \quad (16)$$

where $p\left(h^{l-1}, h^l\right)$ is the joint probability distribution between the visible and hidden layers of the topmost RBM.

Since each layer of DBN is made as RBM, training each layer of DBN is the same as training a RBM. A two-phase training have been done sequentially by: 1) pre-training each unidirectional RBM module separately in an unsupervised manner using CD-1 [7]. 2) fine-tuning, supervised learning with a classifier. The pre-trained RBMs are then stacked up with a SR [9] classifier to form a four-layer deep neural network. The last hidden layer is the input to the SR classifier [9], i.e., the output layer. Finally, the whole DBN is fine-tuned in supervised manner with standard back propagation algorithm. The fine-tuning stage includes two phases: only the output layer is trained in the first phase; in the second phase, all layers are fine tuned. Fine-tuning is conducted in a supervised way with labelled data. At first, only the weights of the output layer were adjusted; and after few number of epochs, the weights of all layers were tuned sequentially. The obtained parameters can be directly applied to the incoming new data, which enables efficient object extraction.

Here the proposed DBN is built up with 3 RBMs to extract the features from the input images. After tuning, the number of the hidden layer nodes are selected to 50, 50 and 200, respectively. As the training samples are M×N dimension vectors, the number of nodes in the visible layer for the 1st RBM is M×N. The number of visible layer nodes for the 2nd and the 3rd RBMs is 50. The architecture of DBN is constructed by these three connected RBMs with a structure of 50-50-200-2. The two neuron in the output layer is used to represents the class of road and non-road structures from the input image. The number of pre-training epochs per layer is 100 with the learning rate of 0.01; the batch size is 128.

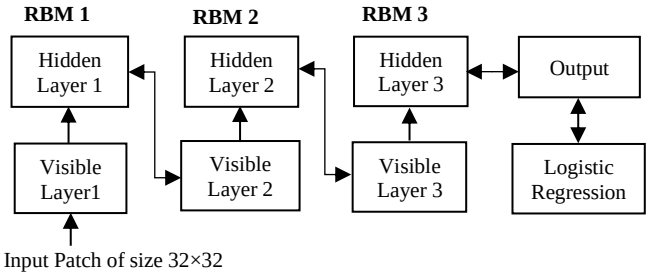


Fig.4. Training Process

Testing Scheme

Once the network is trained and network parameters are learned, image-patch data of a test data can be fed into the model to directly predict the road and non-road class. For the testing stage, the same pre-processing steps are applied on the testing data which were used in training. Images are created in the same fashion as in the training and these patches are then passed to the trained model. The patches from the RS images used in the training stage are not used in testing for the model. For the classification purpose LR is considered as a proficient way. As our problem is a binary classification problem, Sigmoid function is used in the output layer to show the probability map of road or non-road objects.

However, some pixels of the background are segmented as road because of high contrast among road, buildings. Moreover, some road pixels are missing from foreground. To remove such flaws (red circled) from the segmented images, we apply different morphological operations [28] as a post-processing step to obtain the final segmentation.

3.3 POST-PROCESSING

The objective of this refinement process is to eliminate the non-road regions or false segments which do not belong to roads. To eliminate these false segments, Connected Component Analysis (CCA) [45] is used at first to group pixels into different components based on pixel connectivity to extract the disjoint segments from the output of the previous stage. Components whose surface areas are smaller than a predefined threshold will be deleted. Region linking algorithm [45] is used to eliminate the discontinuities detected between road segments. Initially a dilation operation is performed on the CCA images to link the edge segments which are very close to each. A structural element of disk shape of radius 10 is used for the dilation operation

followed by morphological thinning operation. Morphological closing is then applied to remove small holes and noise from the road surface, while an opening operation is used to eliminate small pathways with a structuring element size that is smaller than the main roads width but larger than those of the pathways, resulting in the extracted road network as shown in Fig.5(c). The filtered image is shown in Fig.5(c), it can be clearly seen that all the misclassified objects unconnected to the main road network were removed.

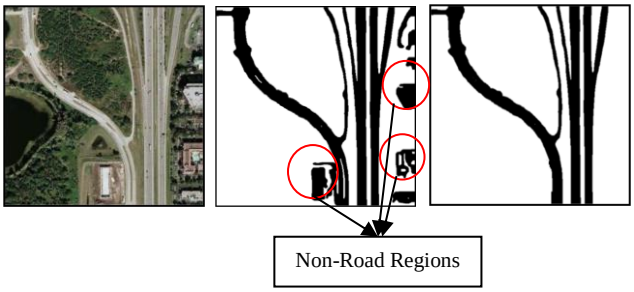


Fig.5. (a) Input Image (b) Classified Image (c) Final result after refinement process

4. EXPERIMENT AND RESULTS

The experiments and performance evaluation are carried out on RS images having different types of urban and suburban areas with various classes such as roads, buildings, trees, vegetation, and shadow. The proposed approach is implemented with MATLAB and tested on a Core i7 2.67GHz PC with 8GB RAM. The learning rate was set to 0.001 during the pre-training stage, and 0.5 was the momentum with 100 epochs for RBM training. Mini-batch size is set to 128 in the pre-training and fine-tuning stages where sigmoid was used as an activation function.

4.1 DATASET

To evaluate the proposed approach, the experiment is carried out on a dataset of 970 RS imagery of size 512× 512 each among which 500 images of urban and 470 images of suburban regions. For creating the data set, we use the benchmark dataset created by Cheng et al. [41] which consist 240 RS images collected from Google Earth. The current work also makes use of a part of database used in [4], images from Massachusetts Roads Dataset [35].

Table.1. Number of samples per class for the training, validation, and test set

Land Cover	Data set (970 image patches)		
	Training (60%)	Testing (20%)	Validation (20%)
Road	302	120	100
Non-Road	280	74	94

The dataset used were augmented by introducing random flips and rotation. In the present work, only 3-Band RGB image are utilized of all the datasets to extract the road networks. Fig.8 shows few samples from our dataset. For each image in the data set, a ground truth (road) map was also obtained using a human operator as shown in Fig.7. These labelled datasets are randomly divided into 60% for training data sets, 20% for validation data sets, and 20% for test data sets as shown in Table.1.

5. PERFORMANCE EVALUATION

The quantitative evaluation of the experimental results is achieved by comparing the automated (derived) results against a manually formed ground truth data. For assessing the performance of the proposed road extraction method, the following four evaluation metrics are used [1] [41]. The recall value indicates the percentage of the ground truth road pixels detected. The precision indicates the percentage of the correctly classified road pixels among all predicted pixels of the classifier. Finally, the quality value indicates the goodness of the result. The F1-Score indicates the harmonic average of Precision and Recall.

$$\text{Recall}=TP/(TP+FN) \quad (17)$$

$$\text{Precision}=TP/(TP+FP) \quad (18)$$

$$\text{Quality}=TP/(TP+FP+FN) \quad (19)$$

$$\text{F1-Score}=(2\times\text{Precision}\times\text{Recall})/(\text{Precision}+\text{Recall}) \quad (20)$$

where *TP* denotes the true positive; *FP* denotes the false positive, and *FN* denotes the false negative.

For evaluating the effectiveness of the proposed technique for road object extraction, the qualitative segmentation results are presented on the different dataset in Fig.6 and Fig.7. They contain all three bands (i.e., R, G, B) and rich information including roads, various buildings, vegetation etc. The images in first column are the original RS images of our database. Those in the third column indicate the automatic results of proposed framework. The final road extraction results can then be generated after post-processing, as presented in the images of the fourth column.

The results are assessed using the reference road extracted by human operator manually illustrated in second column. Road extraction results from the proposed technique on samples in the Massachusetts Road dataset are presented in Fig.6 for a qualitative assessment.

The four experimental results can highlight algorithm's efficiency for road detection. The Fig.7 shows a series of results generated in the process of extracting roads, including sample regions of each category of datasets, the reference road maps and the final road extraction results.

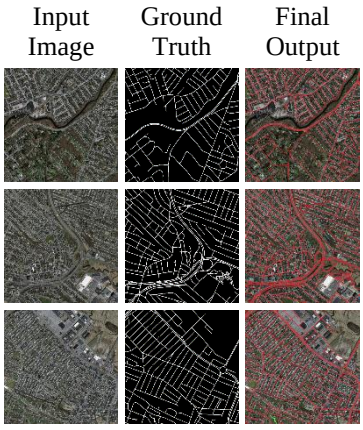


Fig.6. Road extraction Results from Massachusetts Road Dataset achieved by proposed method. The First and second column depict the input images and their corresponding ground truth

images from Massachusetts road dataset. The third column shows the building extraction results. The red, blue, and green colours represent TPs, FPs, and FNs, respectively

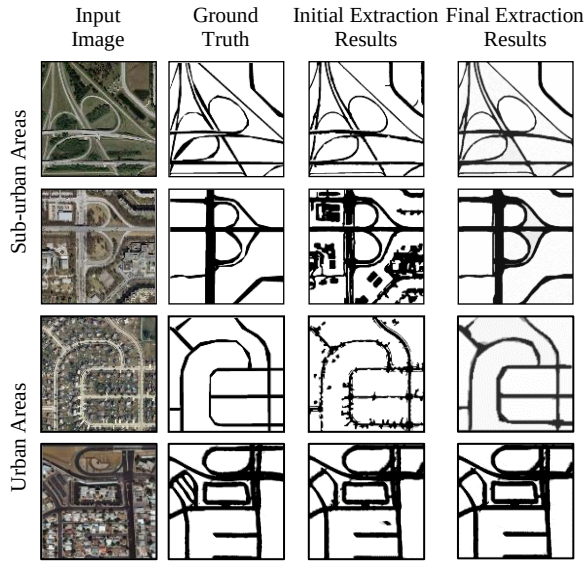
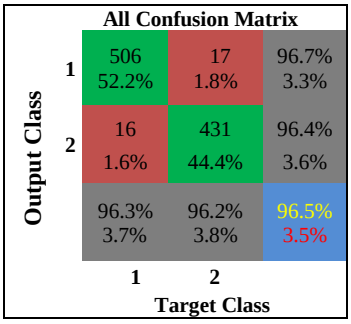
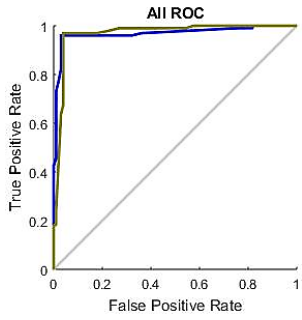


Fig.7. Road Extraction Results of Sub-urban and Urban areas from the data set achieved by the proposed method. First and second columns show the original images and corresponding ground truth. The third and fourth columns are results achieved before post-processing and after post-processing.

The performance of the proposed method is calculated by analysis of confusion matrix and the receiver operator characteristic curve (ROC). The confusion matrix and the ROCs of all data is shown in Fig.9.



(a) Confusion matrix



(b) ROC Plot

Fig.9. Confusion matrix and ROC plot of all data

Of total 970 samples, 937 samples were correctly classified, and 33 samples were misclassified by this network as shown in Fig.9(a). The proposed model resulted in an accuracy of 96.5%, recall of 96.31%, precision of 96.75%, quality 93.91%, F1-score 0.9552. ROC graph depicted in Fig.9(b) shows the plotting of true positive rate (recall) against false positive rate. ROC graph of this network for all data shows an excellent classification between the two categories as the curves lie between the diagonal and the upper-left corner.

The Table.2 list the overall Recall, Precision, Quality and F1-Score of the proposed method of training, testing, validation, and all data, respectively. As seen from the tables, the proposed system using deep learning technique incurs the acceptable level of performance with the mean values of no less than 96.31, 96.70, 93.65 and 95.52 for Recall, Precision, Quality and F1-Score, respectively.

To evaluate the quality of the network, cross-entropy (CE) is used. The Table.3 shows the overall Cross- Entropy (CE) and Percent Error values (E) for the training, validation, and the testing of the proposed deep model. Minimizing cross-entropy results in good classification. The error value indicates the fraction of samples which are misclassified.

Table.2. Performance analysis of proposed technique of Training, Testing Data, Validation, and all Data

DBN-LR deep learning				
	Recall	**Precision**	**Quality**	**F1-Score**
Training (%)	95.36	96.83	92.75	0.9524
Validation (%)	96.00	93.24	93.15	0.9530
Testing (%)	96.67	94.31	93.45	0.9577
All data (%)	96.31	96.75	93.91	0.9552

Table 3. Cross- Entropy (CE) and % Error (E) values

	Sample Size	**CE**	**%E**
Training	582	8.75586e-1	4.16666e-0
Validation	194	1.74173e-0	2.50000 e-0
Testing	194	1.81764e-0	2.50000 e-0

Validation performance based on the cross-entropy is shown in Fig.11. Training was stopped after iteration 69. The performance graph, presented in Fig.11, is showing how CE is minimized for good classification which proves the efficiency of the proposed model. Before epoch 69, the best validation performance of 0.178 was reached at epoch 63. The error histogram with 20 bins (bars) is also shown in Fig.12. It shows how the errors from the network on the testing instances are distributed.

The Table.4 presents the elapsed time of each section. All test images took 67.54 minutes which corresponds to an average processing time of approximately 21 seconds for each image. On the other hand, 80% - 85% of the total processing time is spent on during the training (sec.3.2.2) and classification process (sec. 3.2.3). A neglectable percentage of the total time is spent on the pre-processing step (Sec. 3.1) and post-processing (sec. 3.3). It is also noted that, the proposed approach is quite suitable for GPU programming.

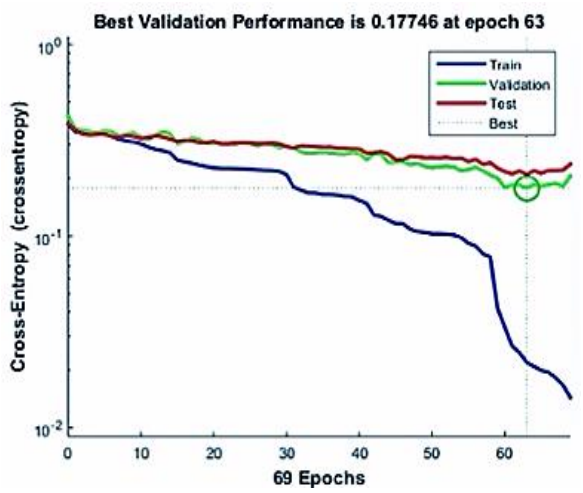


Fig.11. Best Validation Performance based on Cross-Entropy. Lower values are better. Zero means no error

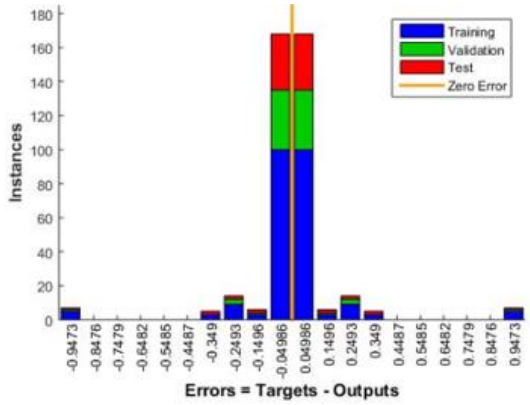


Fig.12. Error Histogram

Table 4. Elapsed time of each section of the proposed Road Extraction approach

Process (Section)	Elapsed Times min : sec (%)
Pre-processing stage (Sec. 3(1))	05:58 (08.48%)
Training (Sec. 3(2(2)))	31:35 (44.12%)
Classification (Sec.3(2(3)))	28:23 (40.30%)
Post-processing (Sec. 3(3))	05:04 (07.10 %)
Total	70:20 (100.00%)

To assess the relative significance of our approach, a comparison with 5 state-of the-art methods is undertaken as baseline which include ResUNet [3], Cascaded CNN [41], RoadCNN [50], FCN [53], RCNN-UNet [52] and Salient features and SVM [4]. Table 5 shows the comparative quantitative evaluation measured in terms of Recall, Precision, Quality and F1-score. In the table, the best performance is denoted in bold, and the second best is marked with underlines. We found that the proposed method can give the relatively high average performance of quality, recall rate, precision and F1-score better than the other four deep learning and one classical method. It should be noted that the results are the average performance of all images under test.

Table.5. Comparison and quantitative evaluation for complex urban areas shown in Fig.8 (2nd row). A higher value indicates a better performance

	Recall	Precision	Quality	F1-Score
Proposed	96.31	96.75	93.91	0.9552
ResUNet [3]	93.27	95.26	90..34	0.8834
Cascaded CNN [41]	93.78	92.47	88.89	0.9314
Road CNN [50]	95.77	94.47	89.05	0.9531
FCN [53]	76.2	85.31	86.56	0.8101
RCNN U-Net [52]	96.16	96.88	93.62	0.9523
Salient features [4]	85	84.67	74.32	0.8091

6. CONCLUSION

In this paper, a deep-learning-based feature-extraction and classification method has been proposed in a DBN architecture to extract road networks from RS images. The major contribution of this work is the road extraction model based on the construction of deep model DBN and the introduction of the RBM as the feature extractor. At the same time, this work designed the LR unit in the expansive part of the model, as feature classifier which improved the precision and F1 score. Experiments were carried out on a remote sensing imagery dataset containing 970 RS images of two categories (road/non-road) for evaluation. The roads were extracted successfully via the deep belief neural network proposed in this work, and the results showed the effectiveness and feasibility of the proposed framework in improving the performance of road extraction work in RS imagery. Quantitative comparisons were performed with some state-of-the-art such as the ResUNet [3], Cascaded CNN [41], RoadCNN [50], FCN [53], RCNN-UNet [52] and Salient features and SVM [4]. The proposed method in this work can obtain improvements in terms of the comprehensive evaluation metric, the F1 score, Recall, Precision and Quality over the methods. Experimental results validated the effectiveness of the proposed method.

Although our results are encouraging, the proposed method can be improved further by fusing deep features with structural ones in future studies. In the future work, we will use graphics processing unit to accelerate the feature learning process. It is observed that the road detection process produces a high degree of accuracy especially for the images of urban regions. In future, we will propose to extract deep features of road by DBN method to secure robustness of feature representation, low computational complexity, and the large variations in road structure. Moreover, the presence of buildings and other features like roads made the extraction process somewhat more difficult compared to the suburban case [57]. Road junction detection and modelling of shadows are issues to be addressed in future work. In addition, vectorization of the extracted road networks can also be a good extension of this work for GIS applications. Further road hypothesis verification and graph data structure are helpful to improve the results to form a complete road network representation.

REFERENCES

[1] Md. Abdul Alim Sheikh, Alok Kole and Tanmoy Maity, "A Multi-level Approach for Change Detection of Buildings using Satellite Imagery", *International Journal of Artificial Intelligence Tools*, Vol. 27, No. 8, pp. 1850031-1850045, 2018.

[2] Zhongbin Li, Wenzhong Shi, Qunming Wang and Zelang Miao, "Extracting Man-Made Objects from Remote Sensing Images via Fast Level Set Evolutions", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 53, No. 2, pp. 883-899, 2015.

[3] Z. Zhang, Q. Liu and Y.Wang, "Road Extraction by Deep Residual U-net", *IEEE Geoscience and Remote Sensing Letters*, Vol. 15, No. 5, pp. 749-753, 2018.

[4] Sukhendu Das, T. T. Mirnalinee, and Koshy Varghese, "Use of Salient Features for the Design of a Multistage Framework to Extract Roads from High-Resolution Multispectral Satellite Images", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 49, No.10, pp. 3906-3931, 2011.

[5] P.P. Singh and R.D. Garg., "A Two-Stage Framework for Road Extraction from High-Resolution Satellite Images by using Prominent Features of Impervious Surfaces", *International Journal of Remote Sensing*, Vol. 35, No. 24, pp. 8074-8107, 2014.

[6] T. Panboonyuen, K. Jitkajornwanich, S. Lawawirojwong, P. Srestasathiern and P. Vateekul, "Road Segmentation of Remotely-Sensed Images using Deep Convolutional Neural Networks with Landscape Metrics and Conditional Random Fields", *Remote Sensing*, Vol. 9, No. 7, pp 680-699, 2017.

[7] G.E. Hinton ,S. Osindero and Y.W. Teh, "A Fast Learning Algorithm for Deep Belief Nets", *Neural Computing*, Vol. 18, No. 7, pp. 1527-1554, 2006.

[8] G.E. Hinton and R.R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks", *Science*, Vol. 313, No. 5786, pp. 504-507, 2006.

[9] Na Lu, T. Li, X. Ren and H. Miao, "A Deep Learning Scheme for Motor Imagery Classification based on Restricted Boltzman Machines", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 25, No. 6, pp. 1-13, 2017.

[10] Weixing Wang, Nan Yang, Yi Zhang, Fengping Wang, Ting Cao, Patrik Eklund, "A review of road extraction from remote sensing images", *Journal of traffic and Transportation Engineering*, Vo. 3, No. 3, pp. 271-282, 2016.

[11] Zhaoli Hong, Dongping Ming, Keqi Zhou, Ya Guo, and Tingting Lu, "Road Extraction From a High Spatial Resolution Remote Sensing Image Based on Richer Convolutional Features", *IEEE Access*, Vol. 6, pp. 46988 – 47000, 2018. **DOI:** 10.1109/ACCESS.2018.2867210.

[12] Miao, Z.L., Wang, B., Shi, W., et al., "A semi-automatic method for road centerline extraction from VHR images", *IEEE Geoscience and Remote Sensing Letters*, Vol. 11, No. 11, pp. 1856-1860, 2014.

[13] Cem Unsalan and Beril Sirmacek, "Road Network Detection using Probabilistic and Graph Theoretical Methods", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 50, No.11, pp. 4441-4453, 2012.

[14] Rasha Alshehhi and Prashanth Reddy Marpu, "Hierarchical Graph-Based Segmentation for Extracting Road Networks from High-Resolution Satellite Images", *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 126, pp. 245-260, 2017.

[15] T. Peng, I. Prinet and J. Zerubia, "Incorporating Generic and Specific Prior Knowledge in a Multi-Scale Phase Field Model for Road Extraction from VHR Image", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 1, No. 2, pp.139-146, 2008.

[16] Z. Sun, H. Fang, M. Deng, A. Chen, P. Yue, and L. Di, "Regular Shape Similarity Index, A Novel Index for Accurate Extraction of Regular Objects from Remote Sensing Images", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 53, No. 7, pp. 3737-3748, 2015.

[17] D. Chaudhuri, N.K. Kushwaha and A. Samal, "Semi-Automated Road Detection from High Resolution Satellite Images by Directional Morphological Enhancement and Segmentation Techniques", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 5, No. 5, pp. 1538-1544, 2012.

[18] S. Leninisha and K. Vani, "Water Flow based Geometric Active Deformable Model for Road Network", *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 102, pp. 140-147, 2015.

[19] M. Amo, F. Martinez and M. Torre, "Road Extraction from Aerial Images using a Region Competition Algorithm", *IEEE Transactions on Image Processing*, Vol. 15, No. 5, pp. 1192-1201, 2006.

[20] S. Movaghati and A. Moghaddamjoo and A. Tavakoli, "Road Extraction from Satellite Images using Particle Filtering and Extended Kalman Filtering", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 48, No. 7, pp. 2807-2817, 2010.

[21] J. Liu, Q. Qin, J. Li and Y. Li, "Rural Road Extraction from High-Resolution Remote Sensing Images based on Geometric Feature Inference", *International Society for Photogrammetry and Remote Sensing*, Vol. 6, No. 10, pp. 314-337, 2017.

[22] Mehdi Maboudi, Jalal Amini, Michael Hahn and Mehdi Saati, "Object-Based Road Extraction from Satellite Images using Ant Colony Optimization", *International Journal of Remote Sensing*, Vol. 38, No. 1, pp. 179-198, 2017.

[23] Mehdi Maboudi, Jalal Amini, Shirin Malihi and Michael Hahn., "Integrating Fuzzy Object based Image Analysis and Ant Colony Optimization for Road Extraction from Remotely Sensed Images", *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 138, pp. 151-163, 2018.

[24] M.O. Sghaier and R. Lepage, "Road Extraction from very High Resolution Remote Sensing Optical Images based on Texture Analysis and Beamlet Transform", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 9, No. 5, pp. 1946-1958, 2016.

[25] Zelang M Iao, Bin Wang, Wenzhong Shi and Hao Wu, "A Method for Accurate Road Center line Extraction from a Classified Image", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 7, No. 12, pp. 4762-4771, 2014.

[26] C. Poullis, "Tensor-Cuts: A Simultaneous Multi-Type Feature Extractor and Classifier and its Application to Road Extraction from Satellite Images", *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 95, pp. 93-108, 2014.

[27] Jun Wang, Jingwei Song, Mingquan Chen & Zhi Yang, "Road Network Extraction, A Neural-Dynamic Framework based on Deep Learning and a Finite State Machine", *International Journal of Remote Sensing*, Vol. 36, No. 12, pp. 3144-3169, 2015.

[28] Zhengxin Zhang, Qingjie Liu and Yunhong Wang, "Road Extraction by Deep Residual U-Net", *IEEE Geoscience and Remote Sensing Letters*, Vol. 15, No. 5, pp. 749-753, 2018.

[29] R. Alshehhi and M.D. Mura, "Simultaneous Extraction of Roads and Buildings in Remote Sensing Imagery with Convolutional Neural Networks", *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 130 pp. 139-149, 2017.

[30] S. Saito, T. Yamashita and Y. Aoki, "Multiple Object Extraction from Aerial Imagery with Convolutional Neural Networks", *Electronic Imaging*, Vol. 60, No. 1, pp. 1-9, 2016.

[31] G. Fu, C. Liu, R. Zhou, T. Sun and Q. Zhang, "Classification for High Resolution Remote Sensing Imagery using a Fully Convolutional Network", *Remote Sensing*, Vol. 9, No. 5, pp. 498-519, 2017.

[32] A. Abdullah, B. Pradhan, N. Shukla, S. Chakraborty and A. Alarming, "Deep Learning Approaches to Remote Sensing Datasets for Road Extraction: A State-of-The - Art Review", *Remote Sensing*, Vol. 12, No. 2, pp. 1444-1458, 2020.

[33] Ruyi Liu, Qiguang Miao, Jianfeng Song, Yining Quan, Yunan Li, Pengfei Xu and Jing Dai, "Multiscale Road Center Lines Extraction from High-Resolution Aerial Imagery", *Neurocomputing*, Vol. 329, pp. 384-396, 2019.

[34] W. Zhao, S. Du and W.J. Emery, "Object-Based Convolutional Neural Network for High-Resolution Imagery Classification", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 10, No. 7, pp. 3386-3396, 2017.

[35] V. Mnih and G.E. Hinton, "Learning to Detect Roads in High-Resolution Aerial Images", *Proceedings of the European Conference on Computer Vision*, pp. 210-223, 2010.

[36] E. Sarhan, E. Khalifa and A.M. Nabil, "Road Extraction Framework by using Cellular Neural Network from Remote Sensing Images", *Proceedings of International Conference on Image Information and Processing*, pp. 1-5, 2011.

[37] J.S.J. Wijesingha, "Automatic Road Feature Extraction from High Resolution Satellite Images using LVQ Neural Networks", *Asian Journal of Geoinformatics*, Vol. 13, No. 1, pp. 30-36, 2013.

[38] Y. Liu, M.M. Cheng, X. Hu, K.Wang and X. Bai, "Richer Convolutional Features for Edge Detection", *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pp. 5872-5881, 2017.

[39] E. Maggiori and P. Alliez, "Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 55, pp. 645-657, 2017.

[40] Z. Zhong, J. Li, W. Cui and H. Jiang, "Fully Convolutional Networks for Building and Road Extraction: Preliminary Results", *Proceedings of International Conference on Geoscience Remote Sensing*, pp. 1591-1594, 2016.

[41] G. Cheng, Y. Wang, S. Xu, H. Wang, S. Xiang and C. Pan, "Automatic Road Detection and Centerline Extraction Via Cascaded End-to-End Convolutional Neural Network", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 55, No. 6, pp. 3322 -3337, 2017.

[42] R. Kestur and M. Mudigere, "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*, Vo. 12, pp. 1-14, 2018.

[43] T.T. Mirnalinee and Koshy Varghese, "An Integrated Multistage Framework for Automatic Road Extraction from High Resolution Satellite Imagery", *Journal of the Indian Society of Remote Sensing*, Vol. 39, No. 1, pp. 1-25, 2011.

[44] Xi Zhao Wang, T. Shang and R.Wang, "Noniterative Deep Learning: Incorporating Restricted Boltzmann Machine into Multilayer Random Weight Neural Networks", *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, Vol. 49, No. 7, pp. 1299-1308, 2019.

[45] R.C Gonzalez and R.E Woods, "*Digital Image Processing*", Prentice Hall, 2009.

[46] F. Hu, G.S. Xiua, J. Hu and L. Zhang, "Transferring Deep Convolutional Neural Networks for Scene Classification of High-Resolution Remote Sensing Imagery", *Remote Sensing*, Vol. 10, No. 11, pp. 14680-14707, 2015.

[47] M. Wang and J.C. Luo, "Extracting roads based on Gauss Markov Random Field Texture Model and Support Vector Machine from High-Resolution RS Image", *IEEE Transaction on Geoscience and Remote Sensing*, Vol. 9, No. 3, pp. 271-276, 2005.

[48] I. Laptev and C. Steger, "Automatic Extraction of Roads from Aerial Images based on Scale Space and Snakes", *Machine Vision and Applications*, Vol. 12, pp. 23-31, 2000.

[49] P. Gamba, F. Dell Acqua and G. Lisini., "Improving Urban Road Extraction in High-Resolution Images Exploiting Directional Filtering, Perceptual Grouping, and Simple Topological Concepts", *IEEE Geoscience and Remote Sensing Letters*, Vol. 3, No. 3, pp. 387-391, 2006.

[50] F. Bastani, S. He, S. Abbar, M. Alizadeh, H. Balakrishnan, S. Chawla, S. Madden and D. DeWitt, "Automatic Extraction of Road Networks from Aerial Images", *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pp. 1-13, 2018.

[51] L. Zhang, L. Zhang and B. Du, "Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art", *IEEE Geoscience and Remote Sensing Magazine*, Vol. 4, No. 2, pp. 22-40, 2016.

[52] Xiaofei Yang, Xutao Li, Yunming Ye and Raymond Y.K. Lau, "Road Detection and Centerline Extraction Via Deep Recurrent Convolutional Neural Network U-Net", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 57, No. 9, pp. 7209-7220, 2019.

[53] J. Long and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1337-1342, 2015.

[54] S. Valero, J. Benediktsson and Waske, B., "Advanced Directional Mathematical Morphology for the Detection of the Road Network in very High-Resolution Remote Sensing Images", *Pattern Recognition Letters*, Vol. 31, No. 10, pp. 1120-1127, 2010.

[55] Xiaofeng Han, Tao Jiang, Zifei Zhao and Zhongteng Lei, "Research on Remote Sensing Image Target Recognition Based on Deep Convolution Neural Network", *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 34, No. 5, pp. 1-14, 2020.

[56] Dexiang Zhang, Jingzhong Kang, Lina Xun and Yu Huang, "Hyperspectral Image Classification using Spatial and Edge Features Based on Deep Learning", *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 33, No. 09, pp. 1-13, 2019.

[57] Md. Abdul Alim Sheikh, Tanmoy Maity and Alok Kole, "Man-Made Object Extraction from Remote Sensing Images using Gabor Energy Features and Probabilistic Neural Networks", *ICTACT Journal on Image and Video Processing*, Vol. 13, No. 2, pp. 2849-2859, 2022.