Semantic Segmentation using K-means Clustering and Deep Learning in Satellite Image

Manami Barthakur
Department of Electronics and Communication Engineering
Gauhati University
Guwahati-781014, Assam, India
manamibarthakur@gmail.com

Kandarpa Kumar Sarma
Department of Electronics and Communication Engineering
Gauhati University
Guwahati-781014, Assam, India
kandarpaks@gmail.com

***Abstract*— In this paper, a deep learning based method, aided by certain clustering algorithm for use in semantic segmentation of satellite images in complex background is proposed. The work considers the formation and training of SegNet in which the output of K-means clustering algorithm is used as input and the label of the particular region of interest (ROI) in the image are used as target. The method does not require any feature extraction, region growing or splitting methods to configure and train the SegNet, which is a deep convolutional Encoder-Decoder architecture trained with (error) Back Propagation learning. The method is tested with different satellite images. The method is also compared with the results obtained when trained with SegNet without selecting the ROI using K-means algorithm and evaluated using metrics such as accuracy, mean IoU and weighted IoU. The normalized confusion matrix is also plotted. The experimental results show that the method is reliable and suitable for real world situations.**

Keywords— K-means clustering, semantic segmentation, deep learning, SegNet, Satellite Image.

I. INTRODUCTION

A versatile tool for exploring the Earth is remote sensing. To capture the spectral and spatial relations of objects and materials perceptible at a distance, different instruments or sensors are used. Satellite images, also known as remotely sensed images, are the data recorded by the sensors from a very small portion of the Earth's surface. The resolution of satellite images are increasing day by day, which in turn raises the level of detail and the heterogeneity of the scenes[2][3]. The geographic information systems which are used for classification of different regions of the earth's surface have used basic classification methods for years. Therefore, with these new high resolution images, these same basic methods cannot provide satisfactory results. Now a day, to classify different regions in satellite images, semantic segmentation is frequently used method. The goal of semantic image segmentation is to associate each pixel of an image with a class of what is being represented. It is essential for many image analysis tasks. The semantic segmentation differentiates between the objects of interest and their background or other objects. Semantic segmentation is also used in many applications such as Autonomous driving, Industrial inspection, Medical imaging analysis etc. [1]. There are several approaches such as semantic Texton Forest [4] and Random Forest based classifiers [5] for semantic segmentation. Most of these methods rely on the measured image characteristics and for this reason, they work well in certain cases and not in others. Moreover, the images are usually corrupted by several artifacts, such as image noise, missing or occluded parts, image intensity inhomogeneity or non-uniformity. Therefore, when dealing with complex image, some prior knowledge may be necessary to disambiguate the segmentation process. For that purpose learning and neuro-computing structures have been used extensively in the literature [1][3][4].

In present, Deep learning is very much popular since it is very useful for real-world applications due to the type of learning it performs. Many computer vision problems like semantic segmentation are performed using deep architectures, usually Convolutional Neural Networks (CNNs) [6] [7] [8], which are surpassing other approaches by a large margin in terms of accuracy and efficiency.

Several learning based methods for semantic segmentation have been developed until now. Regions with CNN [9] feature (RCNN) performs the semantic segmentation based on the object detection results. It first utilizes selective search [9] to extract a large quantity of object proposals and then computes CNN features for each of them. Long et al. [10] learned to combine coarse, high layer information with fine, low layer information. For bilinear up-sampling to pixel-dense outputs, the deconvolutional layers follow the multilayer outputs. In [11], a deep learning based vehicle detection method in satellite images is proposed.

In this paper, a deep learning based approach for semantic segmentation in satellite images is proposed. The work considers the formation and training a SegNet [1] in which the output of K-means clustering algorithm is used as input and the label of the particular region of interest (ROI) in the image are used as target. The method does not require any feature extraction, region growing or splitting methods to configure and train the SegNet, which is a deep convolutional Encoder-Decoder architecture trained with (error) Back Propagation learning. The experimental results show that the method is reliable and suitable for real world situations.

The paper is organized as follows. In Section II the review on SegNet and K-means clustering is presented. In Section III the proposed algorithm is discussed. Then in Section IV the

978-1-7281-0744-8/19/$31.00 ©2019 IEEE

experimental results are shown. At last, in Section V, the conclusion of the work is discussed.

II. A BRIEF OVERVIEW ON SEGNET AND K- MEANS CLUSTERING

A. SegNet

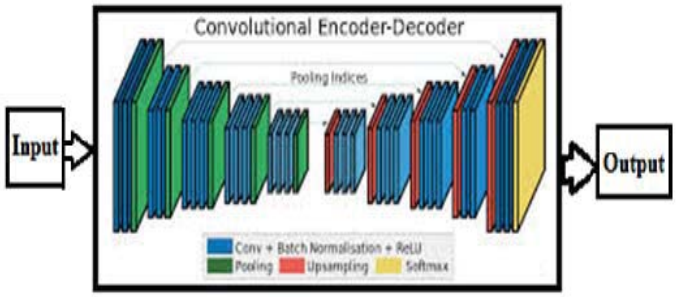

Fig. 1 Architecture of SegNet[1]

SegNet is a deep fully convolutional neural network architecture for semantic pixel-wise segmentation. It has an encoder network and a corresponding decoder network, followed by a final pixelwise classification layer as shown in the Fig. 1. The encoder network consists of 13 convolutional layers and each encoder layer has a corresponding decoder layer. Therefore, the decoder network also has 13 convolutional layers. To produce class probabilities for each pixel independently, the final decoder output is fed to a multi-class soft-max classifier. Each encoder performs convolution with a filter bank in the encoder network, to produce a set of feature maps. These are then batch normalized [12]. After that an element-wise rectified linear non-linearity (ReLU), (max (0, x)) is performed. Following that, max-pooling with a non-overlapping window of size 2 × 2 and stride 2 is performed. Then the resulting output is sub-sampled by a factor of 2. The max pooling indices are stored and used in the decoder network. The decoder network upsamples the input feature maps using the memorized max-pooling indices from the corresponding encoder feature map. This step produces sparse feature maps. In the decoder network, the feature maps are convolved with a trainable decoder filter bank to produce dense feature maps. Finally, the high dimensional feature representation of the output of the final decoder is fed to a trainable soft-max classifier. This soft-max classifies each pixel independently. The predicted segmentation corresponds to the class with maximum probability at each pixel [1].

B. K-Means Clustering

The aim of clustering analysis is to group data in such a way that similar objects are in one cluster and dissimilar objects are in different clusters. K-means clustering is an algorithm to classify or to group the objects K groups. K is a positive integer number. The grouping is done by minimizing the Euclidean distances between data and the corresponding cluster centroid. The steps of algorithm are as follows:

Step1: The value of K is initialized first. An initial set of K called centroids, i.e. in the data space, the virtual points are randomly created,

Step2: To the nearest centroid, every point in the data set is assigned.

Step3: Then the position of the centroid is updated by the means of the data points assigned to that cluster.

Until no centroid was shifted in one iteration, Steps 2 and 3 are performed. In practice, when the minimum shift is below a threshold, the algorithm is stopped. The k-means algorithm will then find the K groups of data that minimize the following function:

$$F = \sum_{i=1}^{K} \sum_{a_j \in S_i} (a_j - c_i)'(a_j - c_i) \quad (1)$$

where there are K clusters S_i, i = 1,2,…….K, and c_i is the centroid or mean point of all the points $a_j \in$ Si. [13][14].

.

III. THE PROPOSED METHOD

The block diagram for the proposed method is shown on Fig 2. The block diagram and the steps involved in the method are explained below.

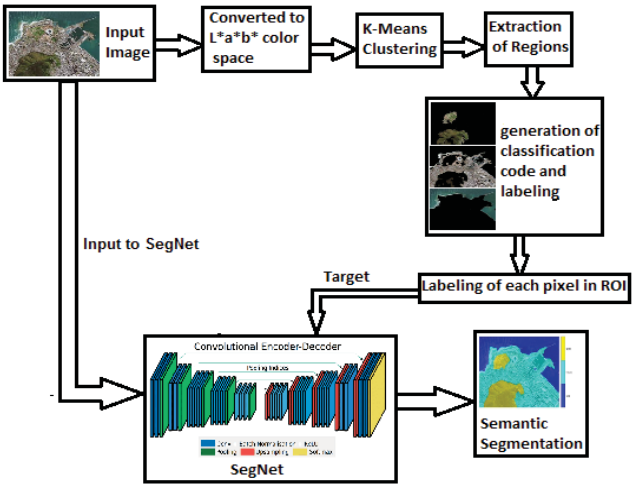

Fig. 2: Block diagram of the proposed method

Step 1: The input in this work is a Satellite image with complex background. The image is resized to 50% of its original image.

Step 2: The image is then converted to L*a*b* Color Space. The L*a*b* color space was derived from the CIE XYZ tristimulus values. The L*a*b* space consists of a luminosity layer 'L*', chromaticity-layer 'a*' which indicates where color falls along the red-green axis, and chromaticity-layer 'b*' indicates where the color falls along the blue-yellow axis. Since all of the color information is in the 'a*' and 'b*' layers, the 'a*' and 'b*' values of the pixels are taken for further processing.

Step 3: Then the k-means algorithm is applied to cluster different regions of the image. Here, the grouping is done by minimizing the Euclidean distances between data and the corresponding cluster center.

Step 4: The algorithm returns an index corresponding to a cluster in the image and with these index value every pixel in the image is labeled.

Step 5: Using the labeled pixels, the ROI of the image are separated**.**

Step 6: Each pixel of each ROI is labeled.

Step 7: Using the ROI as input and the label image of each ROI as target, the SegNet is trained with back-propagation algorithm. As the Batch training method accelerates the speed

of training and the rate of convergence of the MSE to the desired value, therefore this method is adopted here.

Step 8: The output of the SegNet will be a image where each pixel of the image is associated with a class of what is being represented.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

The method is tested using two images- "Satellite image 1" and "Satellite image 1". The images are shown in Fig. 3 and Fig. 7. The images are used to segment the "grass", "house" and "sea" region of the image. The regions such as "grass" region, "house" region and "sea" region obtained from k-means algorithm for the "Satellite image 1" are shown in Fig. 4(a), 4(b) and 4(c) respectively. Similarly, the regions such as "grass" region, "house" region and "sea region" obtained from k-means algorithm for the "Satellite image 2" are shown in Fig. 8(a), 8(b) and 8(c) respectively. The SegNet is trained with the ROI obtained from K-means algorithm and its label images. In the SegNet 13 number of convolution layer and deconvolution layer is used. The filter size is 3×3. The bias term in all convolutional layers are fixed to zero. The weights in the Convolution layers of the encoder and decoder subnetworks are initialized using the 'MSRA' weight initialization method. The output of the proposed method is shown in Fig. 5 and Fig. 9. The method is also compared with the results of SegNet when trained with the "Satellite image 1" and "Satellite image 2"and its labeled images. The experimental results are evaluate with the following metrics-

a. Accuracy:

It indicates the percentage of correctly identified pixels for each class. For each class, accuracy is the ratio of correctly classified pixels to the total number of pixels in that class, according to the ground truth. Therefore,

$$Accuracy = TP/(TP + FN) \quad (2)$$

Here, TP and FN are number of True positive and number of false negative respectively..

b. Intersection over union (IoU):

For each class, IoU is the ratio of correctly classified pixels to the total number of ground truth and predicted pixels in that class. Therefore,

$$IoU = TP/(TP + FP + FN) \quad (3)$$

Where, TP is the number of True positive and FN is the number of false negative and FP is false positive.

c. Weighted IoU:

It is the average IoU of each class, weighted by the number of pixels in that class.

Fig.3: Input for "Satellite image 1".

(a) Grass region (b) house region (c) Sea region

Fig.4: ROI obtained from k-means algorithm for "satellite image 1"

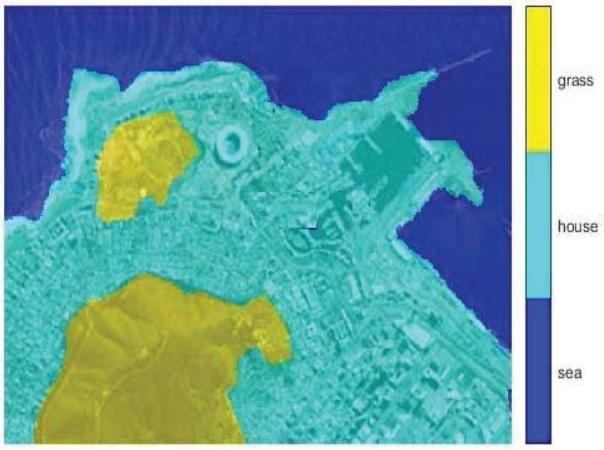

Fig. 5: Output of the proposed method for "Satellite image 1".

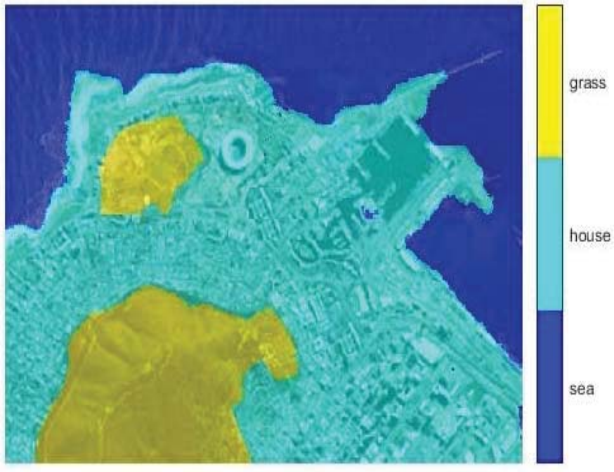

Fig.6. SegNet output for "Satellite image 1"

Fig.7: Input for “Satellite image 2”.

(a) Grass region (b) house region (c) Sea region

Fig.8: ROI obtained from k-means algorithm for “satellite image 2”

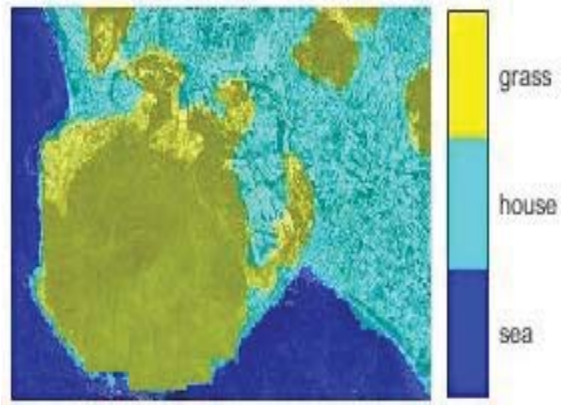

Fig.9: Output of the proposed method for “Satellite image 2”

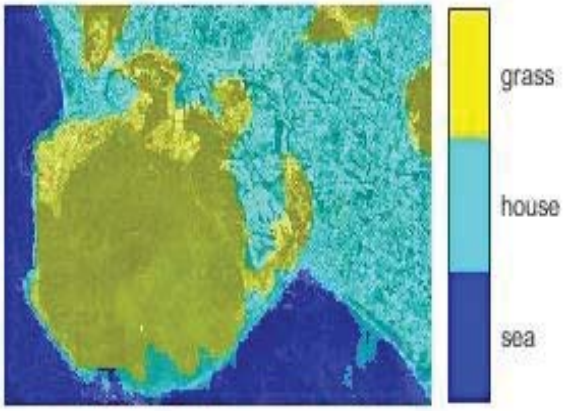

Fig.10. SegNet output for “Satellite image 2”.

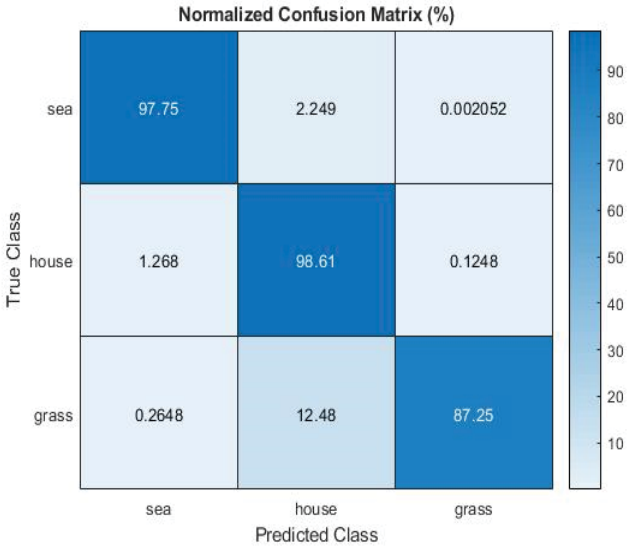

Fig.11. Confusion matrix of the proposed method for “satellite image 1”

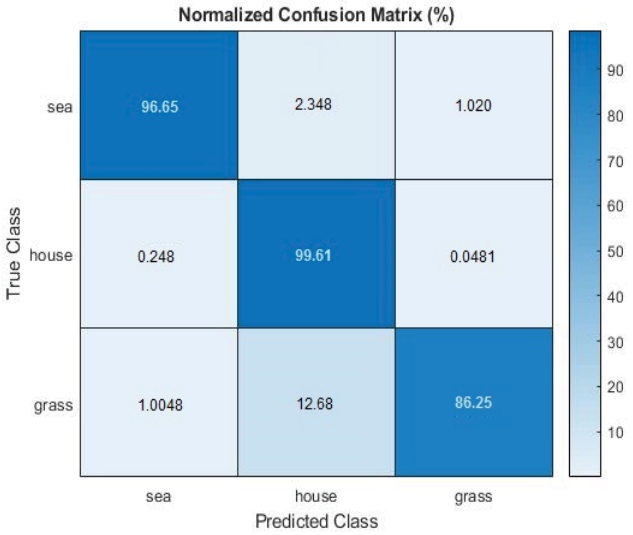

Fig.12. Confusion matrix for “satellite image 1” for SegNet method.

Table 1. Experimental results for “satellite image 1”

Images	Accuracy	Mean IoU	Weighted IoU
Proposed method	0.94356	0.50774	0.48297
SegNet	0.93265	0.50653	0.46843

Table 2. Experimental results for “satellite image 2”

Method	Accuracy	Mean IoU	Weighted IoU
Proposed method	0.92635	0.50646	0.47864
SegNet	0.92454	0.50321	0.46432

Table 1 shows the experimental results obtained for “Satellite image 1”. It was seen that the accuracy is 0.94356 for the proposed method and that for the SegNet is 0.93265. Again the mean IoU is 0.50774 for the proposed method and that for the SegNet is 0.50653. Finally the weighted IoU is

0.48297 for the proposed method and that for the SegNet is 0.46843.

Table 2 shows the experimental results obtained for "Satellite image 2". It was seen that the accuracy is 0.92635 for the proposed method and that for the SegNet is 0.92454. Again the mean IoU is 0.50646 for the proposed method and that for the SegNet is 0.50321. Finally the weighted IoU is 0.47864 for the proposed method and that for the SegNet is 0.46432. Thus for both the images the proposed method gives better result.

Fig. 11 shows the normalized confusion matrix for the proposed method when trained with "satellite image 1". It is seen that 97.75 % of the sea region, 98.61% of the house region and 87.25% of the grass region is predicted correctly. Thus the sea and the house region predicted properly than the grass region. In Fig. 12 the normalized confusion matrix for "satellite image 1" when trained with SegNet is shown. It is seen that 96.65% of the sea region, 99.61% of the house region and 86.25 % of the grass region is predicted correctly by the SegNet method. Thus when compared the two normalized confusion matrix, it is seen that the prediction result for sea and grass region in the proposed method is better than the SegNet method.

V. CONCLUSION

An deep learning based approach for semantic segmentation of setellite images in complex background is proposed in this paper. The work considers the formation and training an SegNet in which the output of K-means clustering algorithm is used as input and the label of the particular region of interest (ROI) in the image are used as target. The method does not require any feature extraction, region growing or splitting methods to configure and train the SegNet, which is a deep convolutional Encoder-Decoder architecture trained with (error) Back Propagation learning. The method is tested with different satellite images. The method is also compared with the results obtained when trained with SegNet method and evaluated using metrics such as accuracy, mean IoU and weighted IoU. The normalized confusion matrix is also plotted. From the experimental results it is seen that, though the time required for training is much higher but the method gives better result and it is reliable and suitable for real world situations.

ACKNOWLEDGEMENT

The authors are thankful to the support received under Visvesvaraya PhD scheme, MEITY, Govt. of India.

REFERENCES

[1] V. Badrinarayanan, A. Kendall, R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation", IEEE Transactions on Pattern Analysis and Machine Intelligence , vol-39, pp. 2481 - 2495, 2017.

[2] S.Praveena, S.P.Singh, I.V.Murali krishna, "An Approach for the Segmentation of Satellite Images using K-means, KFCM, Moving KFCM and Naive Bayes Classifier", in Proceedings of International Conference on Intelligent Computing, Sringer, pp. 21-26, Taiyuan, China, 2013

[3] X. Lin, X. Wang and W. Cui, "An Automatic Image Segmentation Algorithm Based on Spiking Neural Network Model", in Proceedings of

[4] International Conference on Intelligent Computing, Sringer, pp. 248-258, Taiyuan, China, 2014.

[5] J. Shotton, M. Johnson, R. Cipolla, "Semantic Texton Forests for Image Categorization and Segmentation", in Proceedings of IEEE conferece on computer vision and pattern recognition, pp. 1-6, Anchorage, AK, USA, 2008.

[6] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, A. Blake, "Real-Time Human Pose Recognition in Parts from Single Depth Images", in CPVR, pp. 1-8, Colorado Springs, CO, USA, 2016.

[7] A. Krizhevsky, I. Sutskever, G. Hinton, "Imagenet classification with deep convolutional neural networks", Advances in Neural Information Processing Systems, 2012.

[8] X. Chen, S. Xiang, C. Liu, C. Pan, "Vehicle Detection in Satellite Images by Parallel Deep Convolutional Neural Networks", Pattern Recognition (ACPR) 2013 2nd IAPR Asian Conference on, pp. 181-185, 2013.

[9] H. Zhu and J. Qi, "Using Genetic neural networks in image segmentation researching", in Proceedings of IEEE Intrnational Conference on Multimedia Technology (ICMT), pp. 1-6, Hangzhou, China, 2011.

[10] J.R.R Uijlings, V. D. Sande, T. Gevers, A.W. Smeulders, "Selective search for object recognition". In International Journal of Computer Vision Vol. 104 no.2 pp.154–171,2013.

[11] D. Eigen, R. Fergus, ''Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture''. In ICCV, 2015.

[12] Q. Jiang, L. Cao, M. Cheng, C. Wang, J. Li, "Deep neural networks-based vehicle detection in satellite images", In International Symposium on Bioelectronics and Bioinformatics (ISBB)pp. 1-6, Beijing, China, 2015.

[13] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," CoRR, vol. abs/1502.03167, 2015.

[14] I. Garg and B. Kaur, "Color based segmentation using K-mean clustering and watershed segmentation", Proceedings of in 3rd International Conference on Computing for Sustainable Global Development, New Delhi, India, 2016.

[15] M. Barthakur and K. K. Sarma "Complex Image Segmentation using K-means Clustering aided Neuro-Computing", 5th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, Delhi, 2018.