URBAN FEATURES EXTRACTION FROM HIGH RESOLUTION SATELLITE IMAGE USING DEEP LEARNING BASED SEMANTIC SEGMENTATION ALGORITHMS

Nikhil Kulkarni[1], Hina Pande [1], Poonam Tiwari [1], Shefali Agrawal [1]

nikhil.iirs@gmail.com, hina@iirs.gov.in, poonam@iirs.gov.in, shefali@iirs.gov.in

[1]Indian Institute of Remote Sensing (IIRS-ISRO), 4 Kalidas Road, Dehradun, India-248001

KEYWORDS: semantic segmentation, urban features, high resolution satellite image, deep learning.

ABSTRACT

Automated extraction of urban features from high resolution satellite imagery is a challenging task, especially extracting buildings and roads since such man-made features exhibit high intra-class variations and low inter-class variations. The conventional image processing techniques such as grey-level thresholding, surface-based segmentation, iterative pixel classification, edge detection, based on the fuzzy set, when used for high resolution satellite image segmentation, are not able to properly retain high-level features and details such as feature boundaries. The extensive range of applications for urban feature extraction includes automated map making, urban planning, and change detection for real-estate management, land use analysis, and disaster management. For this research, a deep learning-based semantic segmentation algorithm is discussed. In this study, several semantic segmentation models using convolutional neural network (CNN) are compared to achieve optimal accuracy for extracting urban features from high resolution worldview-2 satellite imagery.

Semantic segmentation of satellite and aerial images encounter difficulties due to factors such as relief displacement of high rise buildings, shadows of tall building and trees, etc. Accurate segmentation of buildings and roads into distinct classes from high resolution images becomes challenging because such man-made features have similar reflectance patterns over the visible range of the electromagnetic spectrum. To resolve these issues, an appropriate semantic segmentation model is to be chosen that will give optimal accuracy in extracting urban features (especially buildings and roads) from satellite imagery. Experimental results show that the developed model improves the accuracy of segmentation in comparison with conventional image processing techniques. Predicted results show more than 80% of overall accuracy is achieved using the proposed algorithm. The semantic segmentation model used in this research automatically extracts urban features especially buildings and roads with optimal accuracy.

1. INTRODUCTION

Semantic image segmentation is the task of clustering parts of the image together which belong to the same object class(Thoma, 2016). In this paper, a deep learning algorithm is used to extract the urban features from a remote sensing image. Deep Learning is a subset of Machine Learning, which is further a subset of Artificial Intelligence(Goodfellow, Bengio and Courville, 2016). For the last few years, one of the most difficult problems in computer vision has been image segmentation(Guo et al., 2018). Deep learning models have achieved remarkable results in computer vision(Krizhevsky et al., 2012) and speech recognition. Deep Learning uses brain like functioning using an artificial neural network called convolutional neural networks (CNN)**.**

A Convolutional Neural Network takes an input image, assign importance (learnable weights and biases) to features in the image. Such trained CNN can be used for various applications such as object detection, image segmentation, and feature extraction. In this paper, we have proposed a model based on existing algorithm which accurately extract urban features from the satellite imagery.

1.1 Motivation and problem statement:

Satellite image segmentation has its own challenges due to radiometric, geometric and atmospheric errors. Pixels representing urban features (especially buildings and roads) show high intra-class variations and low inter-class variation(Razeghi, 2015), hence difficult to segment into distinct classes with optimal accuracy. The segmentation of buildings and roads is been a major problem in the remote sensing domain. Buildings and roads extraction have significant importance in remote sensing applications. The extensive range of applications for building and road extraction includes automated map making, urban planning, and change detection for real-estate management, land use analysis, and disaster relief(Saito and Aoki, 2015). Conventional image segmentation techniques fail to classify all buildings and roads pixels into their respective classes. Use of deep learning image segmentation algorithms give an edge over conventional techniques in accuracy and processing larger dataset.

2. LITERATURE REVIEW

2.1 Concept of deep learning

AI imitates human intelligence to correlate the known values and predict unknown values. Machine learning is a subset of AI which uses logic, statistics, and probability for classification, clustering, regression of data and also for detection and synthesis of data. The major difference between machine learning and deep learning is the number of neuron layer(hidden layers) used in the algorithm.

Artificial neural networks (ANN) is inspired by the organization of cells in the animal visual cortex, where each neuron responds to stimuli in a restricted region of space, which is called the receptive field. Convolutional neural networks (CNN) which is an advanced version of ANN, utilize layers with convolving filters that are applied to local features (Yann LeCun et al.). The concept behind the convolutional neural network is image is filtered before training the deep neural network. A filter is a set of multipliers that perform convolution operation with the input image. The characteristics of the filter decide the nature of the outcome which represents features within the image used for specific applications. Figure 1 shows object identification technique.

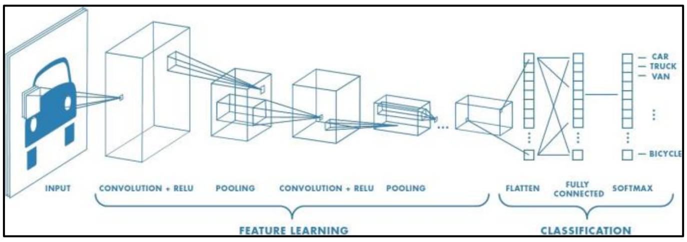


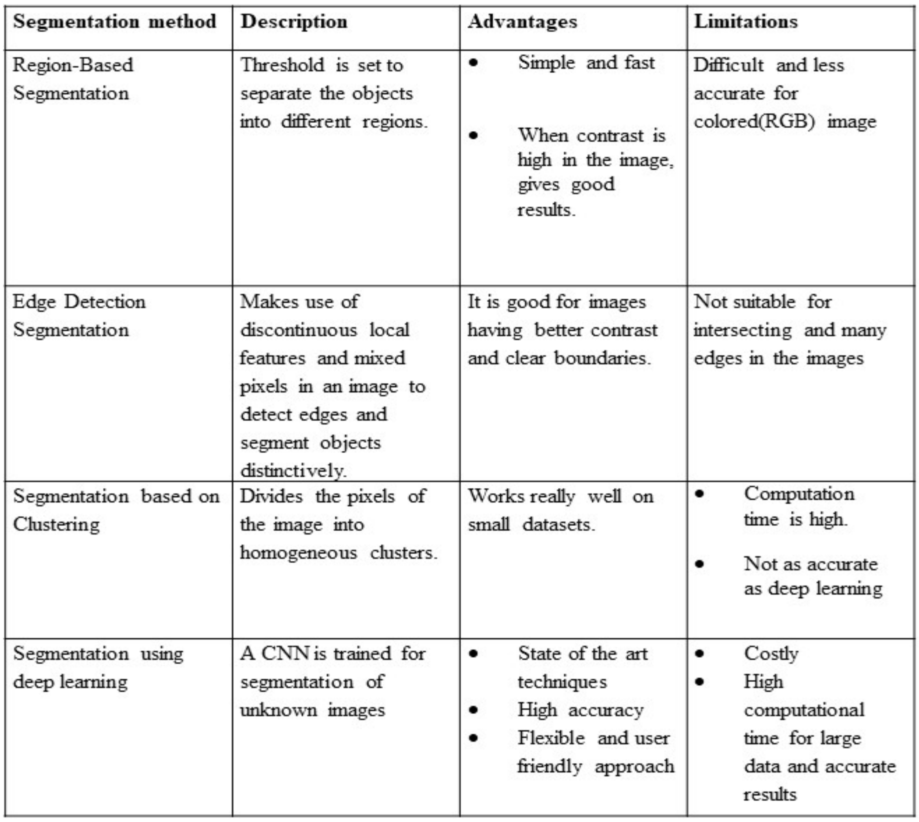
Figure 1. CNN layers for object identification

Convolution is a mathematical function and operated on a matrix with a smaller size matrix called kernel sliding over the input matrix. In a deep learning algorithm, the image is read as a matrix and convolution operations are performed. The size of the kernel is determined based on the expected outcome.

2.2 Segmentation

Segmentation is a process of partitioning the image into some non-intersecting regions such that each region is homogeneous and the union of no two adjacent regions is homogeneous (Pal and Pal, 1993). Image segmentation is different from image classification or object detection since it is not required to have prior knowledge of the visual concepts or objects (Guo et al., 2018). Many image segmentation techniques such as region growing or watershed that depend on iterative merging strategies (Brice and Fennema, 1970). There are many conventional techniques viz. region-based segmentation, edge detection segmentation, segmentation based on clustering, segmentation based on weakly supervised learning in CNN(Ryan, 1985), and state of the art deep learning techniques are used for image segmentation. Table 1 depicts the comparison of prominent and widely used segmentation methods. Deep learning algorithms for image segmentation are superior to conventional segmentation techniques. Therefore deep learning algorithms are more preferred in advanced image segmentation applications.

Table 1 Comparison of Segmentation methods



Segmentation method	Description	Advantages	Limitations
Region-Based Segmentation	Threshold is set to separate the objects into different regions.	• Simple and fast • When contrast is high in the image, gives good results.	Difficult and less accurate for colored(RGB) image
Edge Detection Segmentation	Makes use of discontinuous local features and mixed pixels in an image to detect edges and segment objects distinctively.	It is good for images having better contrast and clear boundaries.	Not suitable for intersecting and many edges in the images
Segmentation based on Clustering	Divides the pixels of the image into homogeneous clusters.	Works really well on small datasets.	• Computation time is high. • Not as accurate as deep learning
Segmentation using deep learning	A CNN is trained for segmentation of unknown images	• State of the art techniques • High accuracy • Flexible and user friendly approach	• Costly • High computational time for large data and accurate results

2.3 Semantic segmentation

Semantic image segmentation is the task of clustering parts of the image together which belong to the same object class(Thoma, 2016). Semantic segmentation includes two major tasks, image-level classification, and detection. Classification is grouping pixels showing similar characteristics in an identical category while detection refers to object localization and recognition(Liu, Deng and Yang, 2018). Semantic segmentation(Figure 2) is a pixel-level classification and discovers both semantics and location: global information resolves ‘what’, while local information resolves ‘where’ (Long, Shelhamer and Darrell, 2015).

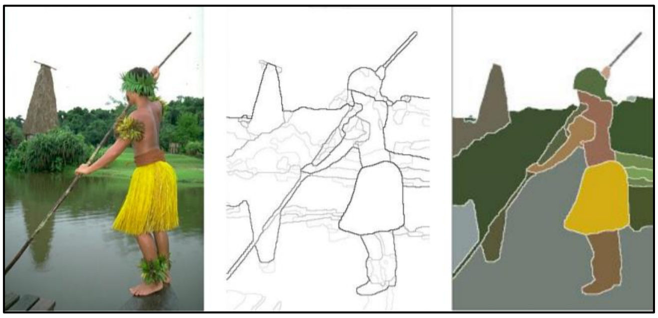


Figure 2.Semantic segmentation: Original Image, Localization of features and segmented outcome

2.4 Concept of U-Net

In U-Net(Ronneberger, Fischer and Brox, 2015) architecture, in the upsampling part, a large number of feature channels are constructed, as a result, the contextual information is propagated to higher resolution layers. Figure 3 shows the layers of U-Net.

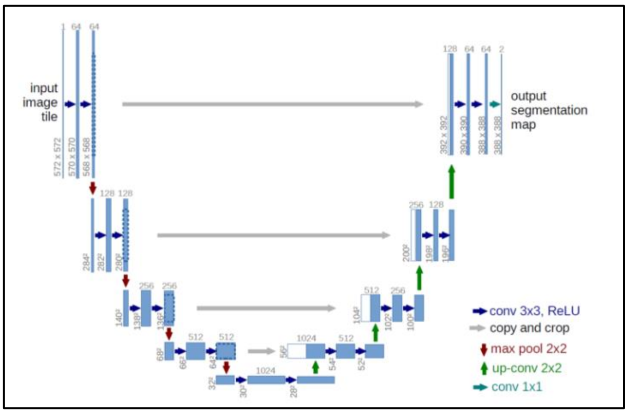


Figure 3.U-Net layers

This U-Net consists of three sections: The contracting(downsampling), the bottleneck, and the expanding section(upsampling).To maintain the symmetry, after each block, the number of feature maps used by the convolutional layer gets half. The higher resolution feature maps are concatenated from the downsampling path with the upsampled features to better learn the high-level characteristics.

3. STUDY AREA AND DATASET USED

A world view-2 satellite imagery of Gandhinagar sub-area, Gujrat, India is used for this project. The aim is to separate urban features from one another and segment them into distinct classes to extract from the satellite imagery therefore image from the urban area are chosen.

The satellite images show a part of Gandhinagar(Figure 4), a 1.51 sq.km area covered, which is a well-planned area with institutional setup, a residential area representing high inter variability in roads and buildings features. The multispectral satellite image, True color composite(TCC) of the Gandhinagar area (Figure 5) that is of 3 bands(R, G, B) with 1.85m spatial resolution.

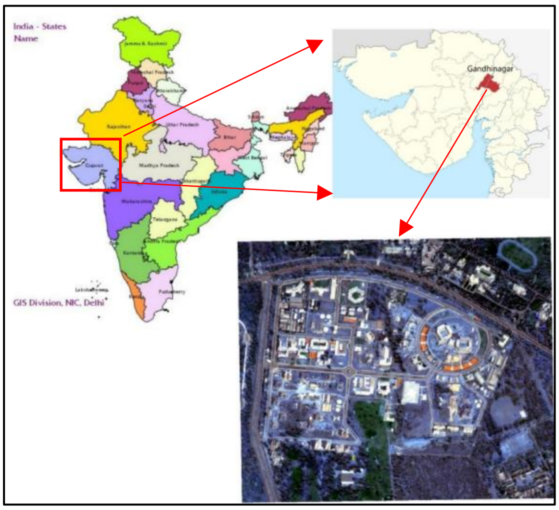


Figure 4.Gandhinagar area, Gujrat

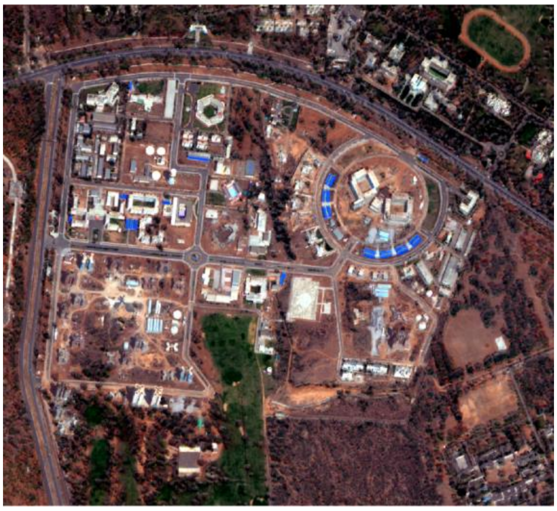


Figure 5. Worldview 2 image(TCC)

4. METHODOLOGY

To understand the methodology (Figure 6 shows a flowchart of the overall methodology), represented it in two parts, image pre-processing (represented by blue color) and deep learning(represented by green colour).

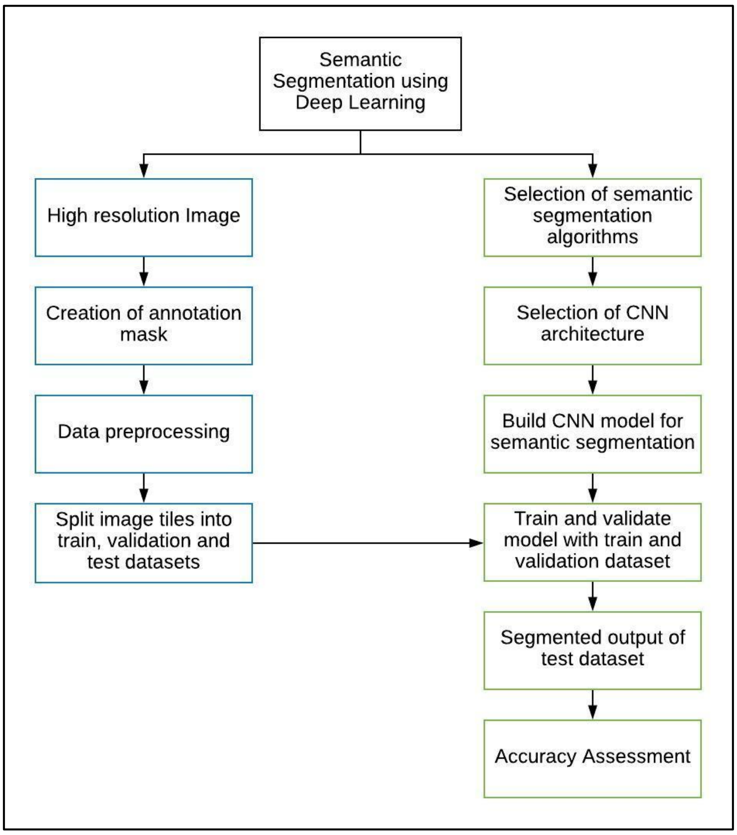

Semantic Segmentation using Deep Learning
High resolution Image
Creation of annotation mask
Data preprocessing
Split image tiles into train, validation and test datasets
Selection of semantic segmentation algorithms
Selection of CNN architecture
Build CNN model for semantic segmentation
Train and validate model with train and validation dataset
Segmented output of test dataset
Accuracy Assessment

Figure 6. Flowchart of Methodology

Data Preprocessing includes all the steps before feeding the dataset to the semantic segmentation model. The steps are as follows.

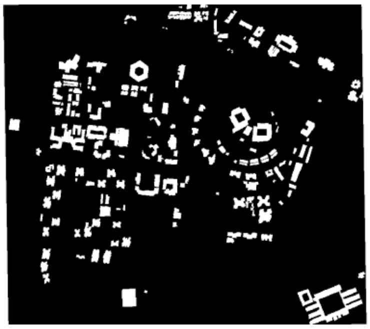
1. Generation of small image tiles from whole imagery(only for WV-2 dataset)
2. Nomenclature of image and corresponding annotation mask tiles
3. Image augmentation on image tiles
4. Associating image tiles to annotation mask tiles

4.1 The Architecture of the semantic segmentation model

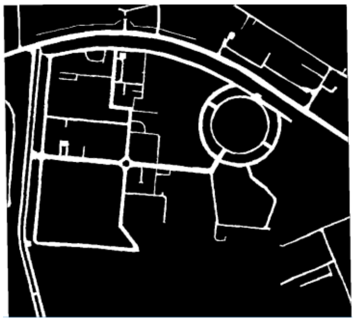
To extract the features from the satellite imagery, the Google Colab notebook is used to formulate the compatible semantic segmentation model using CNN. Fastai library on top of Pytorch library is used to compose segmentation model. There are many deep learning based semantic segmentation algorithms(Pal and Pal, 1993)(Guo *et al.*, 2018) available viz. SegNet, U-Net, DeepLab, PSPNet, etc. Contemplating the nature of the dataset, processing power of the hardware available, and the desired accuracy of the predicted results, the semantic segmentation algorithm for the project is finalized. U-Net gives good accuracy when the dataset is small(Ronneberger, Fischer and Brox, 2015) and requires less processing power. Backbone of ResNet-101 is used construct the model to improve the accuracy of the predicted results.

4.2 Classification schema

Classification schema defines the features that are to be segmented into a number of distinct classes. Here we classified into 5 distinct classes viz. Road, building, grass field, tree, void. Hence annotation mask of the original image comprises of these five classes(Figure 7). The individual annotation masks are created using vectorise-rasterize procedure using QGIS. These annotation masks are merged together to form a single annotation image representing all feature classes(Figure 8)



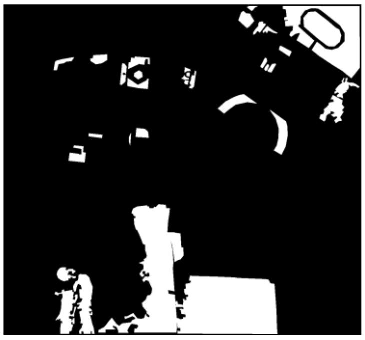
(a)Label for buildings



(b)Label for Roads

(c)Label for trees

(d)Label for void



(e)Label for grass fields

Figure 7. Annotation masks for 5 feature classes : (a)Label for buildings, (b)Label for Roads, (c)Label for trees, (d)Label for void, (e)Label for grass fields

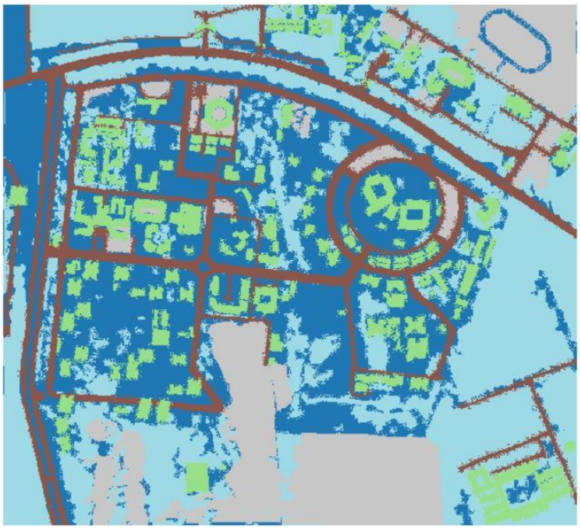


Figure 8. Annotation image

Classification schema of annotated image

🟫 Road 🟩 Building ⬜ Grass field 🟦 Void 🟦 Tree

4.3 Training of the semantic segmentation Model

The learning rate(L.R.) is a very important hyper-parameter and to find the ideal learning rate for the model lr_find(where lr is the learning rate variable) function is used that plots a graph between the loss and learning rate values. Figure 9 shows minimum loss is achieved when the learning rate is 0.001, therefore L.R. is taken as 0.001. Then the semantic segmentation model is trained and validated with the dataset (satellite image and its corresponding annotation image). 80% of the total data samples is used for training and validation while 20% are used for testing. The hidden layers of the model are frozen (model has imagenet weights) and only fully connected layers are allowed to change weights, this technique allowed to decrease the computational time for training the model.

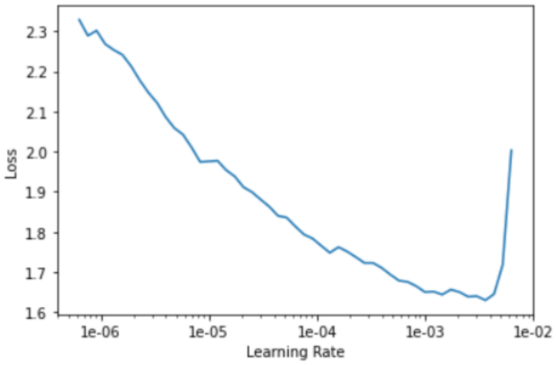


Figure 9. Learning rate vs Training loss curve

4.4 Accuracy assessment matrices

There are many accuracy assessment parameters used for calculating the accuracy of the model. Out of those matrices, pixel accuracy matrix and Intersection over union (IOU) or Jacobean index are used for accuracy assessment of the model.

- Pixel Accuracy:

It matches all the pixels in the reference image to the segmented image and calculates, how many pixels are represented right. So, if we have annotated one class in the reference image, the matrix will calculate how many pixels from that class and unclassed pixels are segmented correctly.

- Intersection over Union(IoU):

It is an object based accuracy matrix. The equation for IoU is the ratio of the area of overlap between reference and segmented image and area of union between the same (Figure 10).

The equation for IoU: IoU(A,B)=A∩B/A∪B

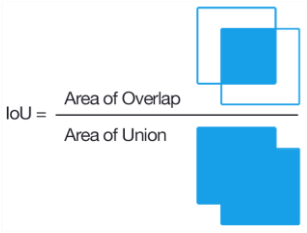


Figure 10 IoU calculation of two images

5. RESULTS AND DISCUSSION

5.1 Results

The trained semantic segmentation model is used to automatically extract the urban features from the test image samples. Table 2 represents the accuracy matrices values calculated with the predicted images for urban feature extraction by the model.

Table 2 Results of accuracy matrices

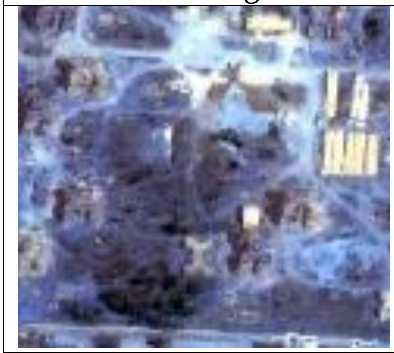
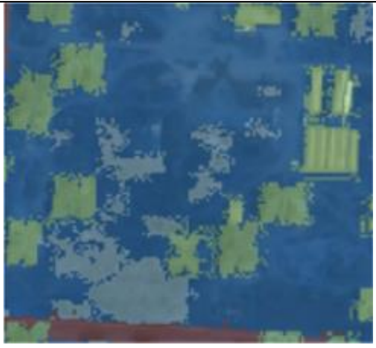
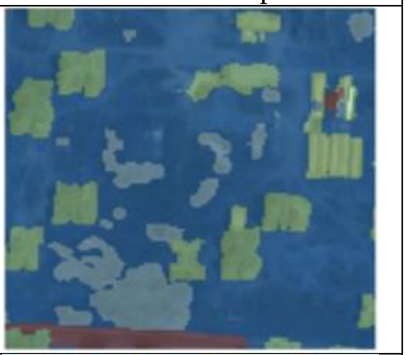
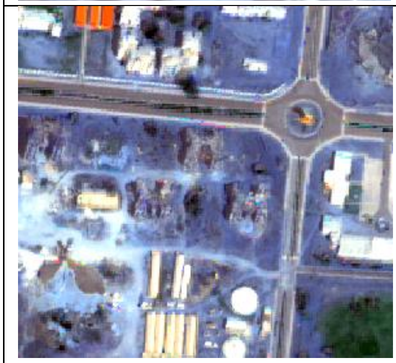
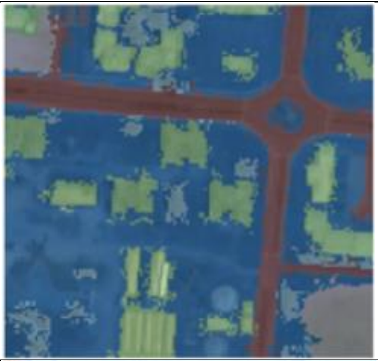
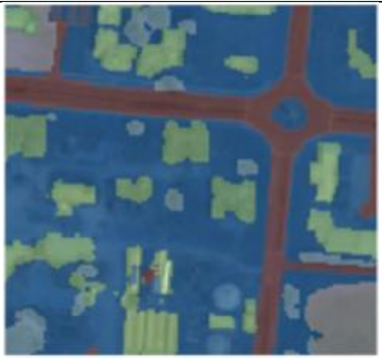
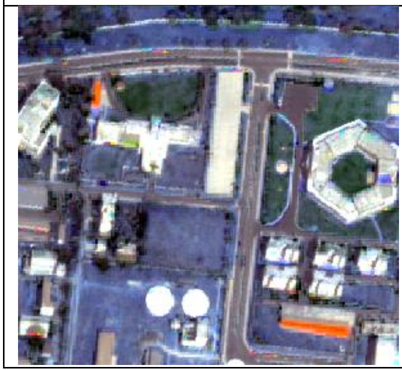
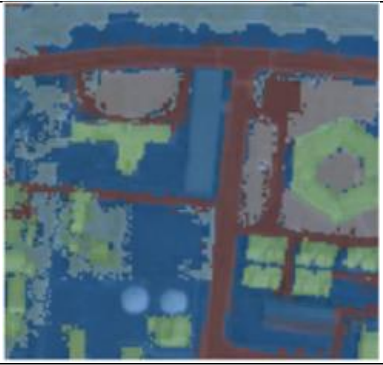
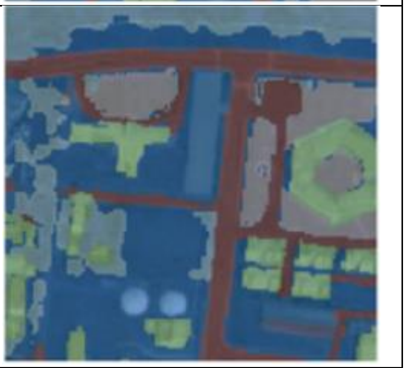
Model	No. of Epochs	Overall accuracy	IoU value
U-Net with ResNet 101	10	0.746	0.498
	500	0.890	0.800
	1000	0.880	0.827

Both overall accuracy and IoU are above 80%, which is considered as highly accurate segmentation results.

5.2 Analysis and discussion

This study formalized the methodology to achieve the accurate urban feature extraction using remote Sensing data. Table 3 shows the comparison between satellite image, reference image and predicted output to visualize the accuracy of the semantic segmentation by our model. The predicted output depicts highly accurate segmentation results with features retaining their original shape. The edges of the features are segmented with high precision. The buildings and roads are accurately segmented into their respective classes which was the primary objective of the research.

Table 3. Predicted Output compared with Satellite image and reference image

Satellite image tiles	Reference image	Predicted output
		
		
		

Deep learning algorithms are preferred for higher accuracies since it works on millions of parameters, learn from the training data and predict the output. Theoretically, with deep learning algorithms highest accuracy for segmentation can be achieved but in practice, memory, computational speed, and dataset size determine the accuracy of the algorithm. The technology has been proved on a limited scale. It holds great potential for up-scaling for operational use.

6. CONCLUSION AND FUTURE WORK

Satisfactory results are achieved with Unet and Resnet 101 architecture combination for the available dataset. The accuracy of model can be achieved by increasing the size of the dataset. If the number of training, validation, and testing samples is increased, better accuracy can be expected. Using the transfer learning technique, computational time can be minimized. Computational time and segmentation accuracy trade-off can be decided as per the requirement of the experiment.

7. REFERENCES

Brice, Claude R., and Claude L. Fennema. *Scene Analysis Using Regions*. 1970, pp. 205–26.

Goodfellow, I., Bengio, Y. and Courville, A. (2016) 'Deep Learning', in. Massachusets, Cambridge, USA: MIT Press. Available at: http://www.deeplearningbook.org.

Guo, Y. *et al.* (2018) 'A review of semantic segmentation using deep neural networks', *International Journal of Multimedia Information Retrieval*. Springer London, 7(2), pp. 87–93. doi: 10.1007/s13735-017-0141-z.

Krizhevsky, A., Sutskever, I. and Hinton., G. E. (2012) 'Imagenet classification with deep convolutional neural networks', *Advances in neural information processing systems*. Available at: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networ.

Liu, X., Deng, Z. and Yang, Y. (2018) 'Recent progress in semantic image segmentation', *Artificial Intelligence Review*. Springer Netherlands. doi: 10.1007/s10462-018-9641-3.

Long, J., Shelhamer, E. and Darrell, T. (2015) 'Fully Convolutional Networks for Semantic Segmentation'.

Pal, N. R. and Pal, S. K. (1993) 'A review on image segmentation techniques', *Elsevier*, 26(9). Available at: https://www.sciencedirect.com/science/article/abs/pii/003132039390135J.

Ryan, T. W. (1985) 'Image Segmentation Algorithms', *Architectures and Algorithms for Digital Image Processing II*, 0534, p. 172. doi: 10.1117/12.946577.

Razeghi, O. (2015) *An investigation of a human in the loop approach to object recognition*. University of Nottingham. Available at: http://eprints.nottingham.ac.uk/29084/1/Thesis.pdf.

Ronneberger, O., Fischer, P. and Brox, T. (2015) 'U-net: Convolutional networks for biomedical image segmentation', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9351, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.

Saito, S. and Aoki, Y. (2015) 'Building and road detection from large aerial imagery', *Image Processing: Machine Vision Applications VIII*, 9405(February), p. 94050K. doi: 10.1117/12.2083273.

Thoma, M. (2016) 'A Survey of Semantic Segmentation', pp. 1–16.

Yann LeCun, et al. "Deep Learning." *Nature*, https://www.nature.com/articles/nature14539. Accessed 11 Dec. 2019.