International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-10 Issue-8, June 2021

Semantic Segmentation of Satellite Images using Deep Learning

Chandra Pal Kushwah, Kuruna Markam

***Abstract:** Bidirectional in recent years, Deep learning performance in natural scene image processing has improved its use in remote sensing image analysis. In this paper, we used the semantic segmentation of remote sensing images for deep neural networks (DNN). To make it ideal for multi-target semantic segmentation of remote sensing image systems, we boost the Seg Net encoder-decoder CNN structures with index pooling & U-net. The findings reveal that the segmentation of various objects has its benefits and drawbacks for both models. Furthermore, we provide an integrated algorithm that incorporates two models. The test results indicate that the integrated algorithm proposed will take advantage of all multi-target segmentation models and obtain improved segmentation relative to two models.*

Keywords: A Satellite Image, Deep Neural Network, U-net,SigNet.

I. INTRODUCTION

The emergence and growth of satellite imagery in high resolution have enhanced our planet's awareness. This useful information allows us to track variations in the surface of the Earth, keep maps up to date, and prepare the communities. One of the key tasks in satellite analysis is to classify land cover, which is semantic segmentation and can be described as dividing each image into meaningful areas and assigning one of the predefined Labels into individual region pixels. Because of a dynamic satellite imagery environment, manual segmentation is repetitive, time-consuming & vulnerable to human errors. Automatic segmentation is thus beneficial and in recent years has attracted significant attention.

Semantic image segmentation (SIS) is a method used for dividing images into many separate regions with their specific features& extracting interesting objects. SIS technologies will segment & label particular goals in remote sensing image analysis, to extract special knowledge. Semantic segmentation (SS) technology, for example, will segment & remove buildings or plants in images that help other analyses.SS of remote sensing image is a significant research study that can encourage the growth of military, agriculture, environmental protection &another major study Field [1]. When it comes to marine surveillance, it can accurately track the sea situation and alert in time.

Manuscript received on June 07, 2021.
Revised Manuscript received on June 10, 2021.
Manuscript published on June 30, 2021.

* Correspondence Author

Chandra Pal Kushwah*, Department of Electronics Engineering, Madhav Institute of Technology & Science, Gwalior (MP), India. E-mail: cpkushwah00@gmail.com

Kuruna Markam, Department of Electronics Engineering, Madhav Institute of Technology & Science, Gwalior (MP), India. E-mail: karunamarkam@gmail.com

The segmentation of structures, roads, and trees in urban planning will help the government plan the community efficiently. It will identify the goals of major and large emissions enterprises [2] in terms of environmental control, which can contribute to the efficient surveillance and regulation of emission behavior. With the advancement of remote sensing technologies, the remote sensing images will receive more and more information [3]. Remote sensing photographs are expanding and have become an essential part of human life.

CNN has obtained positive performance in the computer vision arena in recent years with the growth of ML and DL. It is well used in the identification of images, objects, and other fields. Deep learning is a recent area of study centered on machine learning. By building a neural network with several layers and using many training samples, it learns the key characteristics and eventually increases classification or prediction accuracy. [4]. because the convolutional layer is extremely successful in extracting images, several studies have attempted the deep convolutional neural network to the segmentation region of the image semantic [5, 6]. Investigators used the texture as well as the spectral characteristics of the images to separate before they were DL [7]. Nevertheless, the specifics of the information in the remote frame segmentation cannot be adequately detailed and the fulfilled segmentation cannot be provided for challenging tasks [8, 9].Current years have seen increasing benefits in image analysis for deep learning algorithms [10, 11]. In remote sensing image processing technology, this is very relevant. DL in the natural scene supports DL in remote sensing images. As the ambient noise interacts, an irregular texture distribution, uneven light transitions as well as other parameters, complex remote sensor images must be processed [12]. Thus the implementation of DL of remote sensing images needs greater specifications. There is an immediate need to explore how to more efficiently and reliably use deep learning for segments of remote sensing videos. In this analysis, we will use DL technologies to segment remote sensing imagery in multi-target semantic. The most interesting observation in remote sensing images is structures, highways, trees, and water. We develop the CNN SegNet & U-net models which may be applied to semantic multi-target image segmentation. To investigate the deep learning technologies used for semantic remote sensing image segmentation and to improve the emerging innovations and algorithms to achieve better accuracy.

The most important contributions are described below:

- This working model the encoder-decoder dependent on SegNet the CNN framework could be used to segment remote sensing images with the Pooling Index for multi-target semantic.

The Pooling Index will store the special pixel information that may be returned during the decoding process. As the pixel location can be restored the specifics of the segmentation of the edges can be given.

The study applies enhanced batch normalization to CNN for multi-objective semantic segmentation for remote sensing images to be made acceptable. The normalization of the load applied will adjust the distribution of the input data to the next stage, such that the current distribution cannot be learned, on the other hand, the parameter can be updated more effectively.

- This research contrasts the effects of these two approaches in particular the basic benefits and challenges in the remote sensing image segmentation for various classes of objects.
- This study provides a semantic segmentation algorithm of various CNN models. Each has its benefits and drawbacks in various implementations or real artifacts in deep learning models. The findings can be incorporated into the pixel segmentation with the algorithm suggested by this work. This allows multiple model benefits on various issues to be used and outcomes to be better.

The remaining of this paper is structured as occurs: Section II. Introduces summary of the associated study on Semantic Segmentation of Satellite Images. Section III. Proposed work is obtainable. In Section IV. gives & deliberates outcomes of the model simulation. Lastly in Section V. the conclusions & suggestions are given for future work.

II.REVIEW OF LITERATURE

Baghbaderani, R. et al. [13] suggested the inclusion of spectral unmixing approaches to get efficient spectral data representations for the satellite images' semantic segmentation. They demonstrate that the efficiency of land cover classifications can be improved by accurately extracting features as input in the deep learning model.

Saha, S., et al. [14] proposes a context-sensitive CD architecture with pre-trained CDN-based feature extraction in multi-temporal VHR images. Such a paradigm can effectively model spatial connections between nearby pixels in VHR images while not supervised.

Zhao, F., & Zhang, et al. [15] Work on implementing a building damage evaluation model based on the R-CNN mask with a two-stage practical training strategy. The initial stage motivates the training of ResNet 101 as a building feature extractor in the R-CNN mask. In stage 2 they draw on a model which was learned in stage 1 to further develop a more deep learning architecture, which can distinguish buildings with various degrees of damage from satellite images.

Gong, M., et al. [16] In this study they propose a new structure for remote sensing images that includes extraction of superpixel-based improvements and hierarchical interpretation of variations by NNs.

Guo, Z., et al [17] Suggested integration of SR innovations into the current architecture to increase the efficiency of segmentation. Using descriptive multi-source sample materials, the viability of the suggested approach was evaluated: Pan-chromatic satellite imaging HR &LR, such as training & research outcomes correspondingly.

Nivaggioli, A. et al. [18] suggested an Affinity Net adaptation which helps us to conduct SS directly. Their findings indicate that the labels produced resulted in the same outcomes when training multiple segmentation networks. In addition, the SS of Affinity Net and Random Walk is near the strongest completely supervised technique.

Zheng, Z., Zhong, et al. [19] suggested a new pyramid for multi-tasking end-to-end learning using the Encoder-Dual Decoder Framework on the Pyramid network (Pop-Net). The algorithm is a ResNet-101 backbone network that is deformable. The SS and height prediction are respectively due to two feature pyramid networks as decoders.

III.PROPOSED WORK

A. Problems Identification:

Image segmentation is the term we'd give if we'd asked to sum up all things that we did, in one word. Image segmentation is not an easy task to do it requires high computational power. The problem was to train the model with every label at once and using the basic approach to work. We identify the problem and tried to better out the existing ones and we did. Our approach was to tweak the model and lean it towards the new approach of segmentation. Our proposed approach has given a better result which outperforms the previous approaches.

B. Proposed Methodology:

The first model was designed for understanding and implementing a model of the NN. We tweak the U-net model for better performance. Our work consists of many operations like the image is being cropped into patches and provided with a label. We thought instead of giving the whole image and every label we designed it in a way so that it can train on each label one after another then combined the multiple outputs to segment the images.

We used many libraries as our work required them to go forward, loading the data and making the patches, selecting the class which you want to train first are as follows:

1:'Building',2:'Struct', 3:'Road', 4:'Track', 5:'Trees', 6:'Crops', 7:'Fast H20', 8:'Slow H20',9:'Truck', 10:'Car'.U-net needs very computational power to train and high computational power is equal more time. It took hours to train on only one label.

1) Image Classification

In the Image Classification issue, we assume that a computer would produce a discrete label, the key element in the image, as the most essential component in computer vision. In the classification of images, we conclude that the image contains only one (and not several) objects.

2) Semantic Segmentation

The purpose of SIS is to label every pixel in an image with its respective class. This function is widely referred to as dense prediction since we are forecasting for each pixel in an image.

Recognize that, during previous tasks, labels & bounding box parameters are not the expected performance in SS a high-resolution view itself (generally the same size as an image input) is an image in which each pixel is focused on a particular class. This is a pixel image classification.

Retrieval Number: 100.1/ijitee.H91860610821
DOI: 10.35940/ijitee.H9186.0610821
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication

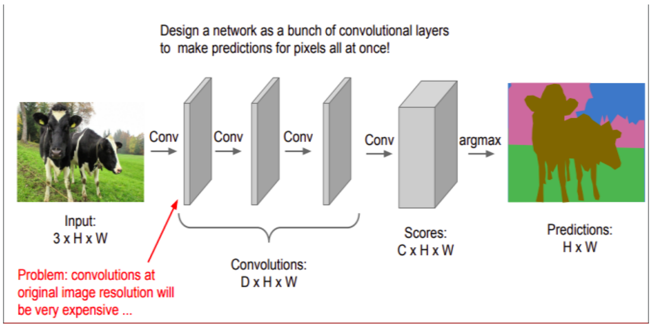
Fig.1. Segmented Image

3) U-net Model

U-Net [20], a network introduced for the medical imaging community with limited data training, is equivalent to SegNet. Functionally related to SegNet, their encoding blocks use the layer before maximum pooling and link them to their corresponding decoder layer. Up-sampling is done on the blocks at the expense of more memory by transmission convolutional. However, U-Net has shown its efficiency in data sets without training data in which it is necessary to use maximum users. There are many real-life applications where there is a lack of details. While satellite images can easily be accessed by various methods, data with accuracy is scarce mainly due to the lack of resources. The details were manually labeled for hours of power by men in most instances. For this cause, we use the U-Net framework as the basis of our research. This restriction drives our network.

U-Net Architecture

There are lots of architecture experimented and some of them gave good results too. Let's consider fully convolutional layers. This can learn very small features in the primary layers and more advanced features in the latter layers. At last with the help of the soft ax layer, we can make the network learn semantic segmentation. But with the convolutional neural network there comes an issue of computation complexity. For this, we have to keep the whole dimension throughout the network for learning which is very much complex.



Another approach to mention is encoded/decode architecture. Here we improve the computation complexity by reducing the spatial resolution, the network learns lower resolution features and it is appeared to be efficient. At last the architecture up samples the feature representations into a full-resolution segmentation map.

While going through different architecture most of the research studies end up in selecting U-Net architecture. It is understood that U-Net architecture is designed for semantic segmentation.

U-Net is a semantic segmentation structure. There is consisting of a contracting & an extensive path. The contracting path followed the traditional network design. This consists of 2-3x3 convolutions repeatedly applied (unpadded convolutions), including a rectified linear unit rectified (ReLU) as well as a max 2x2 max pooling process. At any down sampling point, we get many function channels. A 2x2 up sampling of the map, twice the number of feature channels, a concatenation with a track map that is consequently cropped, as well as two 3x3 convolutions, every accompanied by a ReLU, are included in all steps in the comprehensive path. Owing to a loss of pixels in each convolutional, cropping is required. A 1x1 convolutional on the output layer maps the required number of classes onto every 64-component feature vector. the system consisted of 23 convolutional layers.

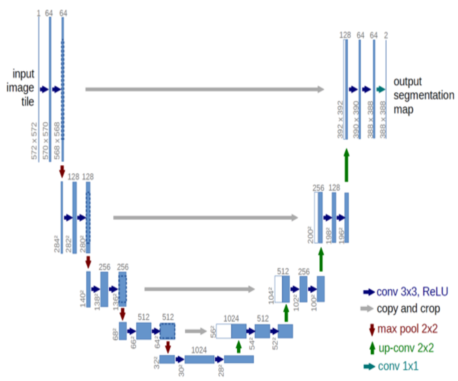


Fig.2. The architecture of U-net (sample for 32X32 pixels in lowest resolution). A multi-channel feature map relates to every other blue box. On top of the box is shown no. of channels. The X-y size is accessible at the bottom left edge of the case. Functional maps copied reveal white boxes. The arrows display the numerous operations.

The left-hand side of U-Net is the encoder network and the right-hand side is the decoder network. In between, it has links from an encoder to the decoder with a concatenation of feature maps that helps giving more localization information.

IV.RESULTS AND DISCUSSION

This work has implemented using python programming language and the platform used is Jupyter notebook (version 6.3.1) and done for the result at the proposed approach.

A. Dataset Description

In this competition, The DSTL (Defense Science and Technology Laboratory) is working on new ways to alleviate the pressure on their image analysts.

Hagglers are required to identify the features accurately in overall imagery at this competition. In all 3 band & 16 band formats, DSTL provides you with 1km x 1km of satellite images. Your objective is that the object forms contained in these regions be a detective and classified.

3- and 16-bands images

This competition contains two categories of spectral imagery data. 3-band images are real color images of the traditional RGB. 16-band images provide spatial data by the capture of wavelength channels. These multispectral (400-1040nm) and shortwave infrared (SWIR) (1195-2365nm) images are taken from this multi-band imagery. Both images are in GeoTiff format & can include viewing by GeoTiff (like QGIS). See our tutorial about how to display the images computationally.

B. Result of the Proposed Model

Input data for our model total 270 million pixels, around 14.79 percent of the total amount of pixels accessible & 85.21 percent of our target segmented (Buildings, Roads, Tracks, Truck, Car, Water, and Corps). Parameters are calculated are seen in the table below:

Table 1: The comparison of performance Parameter

Parameter	Base	Proposed
Precision	0.9023	0.9311
Recall	0.65	0.73

The comparison of performance Parameter shown in Table 1 the Precision, Recall, F1-score for Base and Proposed

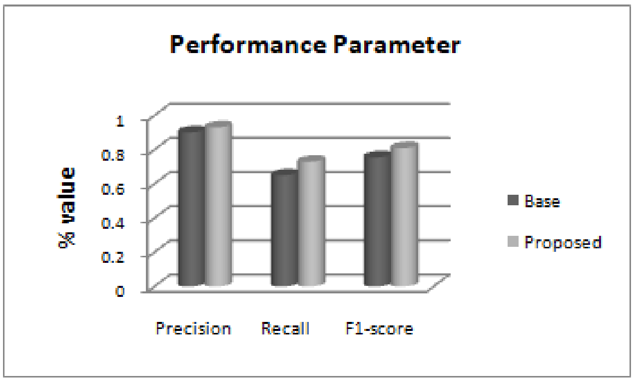


Fig. 3 Comparison graph Precision, Recall, F1-score for performance Parameter for both base and Proposed

Fig.3. Shows that proposed (U- Net Model) are used which gives that performance Parameter of are recorded and compared with each other and we can see that proposed paper gives us a better result than the base

Table 2: The comparison of performance Parameter

Parameter	Base	proposed
Accuracy	71	85

The Comparison of performance Parameters shown in Table 1 the Accuracy for a base and proposed.

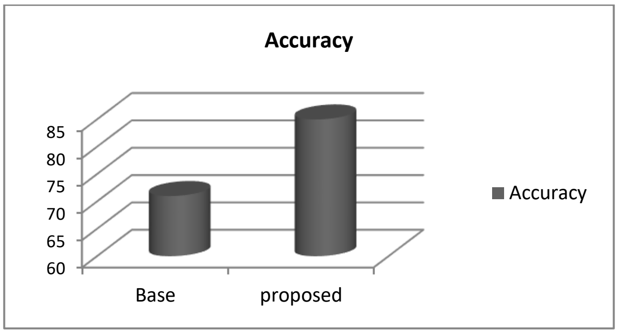


Fig. 4.The Comparison graph Accuracy for performance parameters for both base and proposed.

Fig.4. Shows that proposed (U-net Model)are used which gives the accuracy of are recorded and compared with each other and we can see that proposed paper gives us a better result than the base

V.CONCLUSION

DNN is considered to learn the newest developments & implements in the field and the possibility to help us out in this work. In recent research Segmentation and Semantic Segmentation are examined to identify our enhancements. For this reason, the available data sets were reviewed, & the best data set appropriate for these tasks was found & then the main framework of FCN is used to execute the studied analysis by using U Net as a pre-trained framework & python language. Both performance and outcomes are seen with the training data set at end of the performance.

FUTURE WORK

We will continue to work on our achievements both on the performance of the segmentation labels produced as well as on the direct semantic segmentation. Semantic segmentation of satellite images is a poorly controlled element for the transformation from delocalized text to complete semantic segmentation.

RREFERENCES

1. Maggiori E, Tarabalka Y, Charpiat G, Alliez P (2016) fully convolutional neural networks for remote sensing image classification. In: Geoscience remote sensing symposium
2. Bilgin G, Erturk S, Yildirim T (2011) Segmentation of hyper spectral images via subtractive clustering and cluster validation using one-class support vector machines. IEEE Trans Geosci Remote Sens. 49 (8):2936–2944 Article Google Scholar
3. Maggiori E, Tarabalka Y, Charpiat G, Alliez P (2017) Recurrent neural networks to enhance satellite image classification maps. IEEE Trans Geosci Remote Sens. PP (99):1–10 Google Scholar
4. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D,Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions:1–9
5. Paisitkriangkrai S, Sherrah J, Janney P, Hengel VD (2015) Effective semantic pixel labeling with convolutional networks and conditional random fields. In: Computer Vision Pattern Recognition Workshops
6. Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL (2014) Semantic image segmentation with deep convolutional nets and fully connected crfs, arXiv:1412.7062
7. Johnson M, Shotton J, Cipolla R (2008) Semantic texton forests for image categorization and segmentation. Proc IEEE Cvpr 5(7):1–8 Google Scholar

Retrieval Number: 100.1/ijitee.H91860610821
DOI: 10.35940/ijitee.H9186.0610821
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication

8. Fukushima K (1980) Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol Cybern 36(4):193–202
Article Google Scholar
9. Ohtsu N (2007) A threshold selection method from gray-level histograms. IEEE Trans Sys Man Cybern 9(1):62–66
Article Google Scholar
10. Krizhevsky A, Sutskever I, Hinton GE (2012) Image net classification with deep convolutional neural networks. In: International Conference on Neural Information Processing Systems
11. Bosch A, Zisserman A, Munoz X (2007) Image classification using random forests and ferns. In: IEEE International Conference on Computer Vision
12. Zhao W, Du S, Qiao W, Emery WJ (2017) contextually guided very-high-resolution imagery classification with semantic segments. ISPRS J Photogram Remote Sens. 132:48–60
13. Baghbaderani, R. K., & Qi, H. (2019). Incorporating Spectral Unmixing in Satellite Imagery Semantic Segmentation. 2019 IEEE International Conference on Image Processing (ICIP). doi:10.1109/icip.2019.8803372
14. Saha, S., Bovolo, F., & Brurzone, L. (2018). Unsupervised Multiple-Change Detection in VHR Optical Images Using Deep Features. IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium. doi:10.1109/igarss.2018.8519440
15. Zhao, F., & Zhang, C. (2020). Building Damage Evaluation from Satellite Imagery using Deep Learning. 2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI). doi:10.1109/iri49571.2020.00020
16. Guo, Z., Wu, G., Song, X., Yuan, W., Chen, Q., Zhang, H., Shao, X. (2019). Super-Resolution Integrated Building Semantic Segmentation for Multi-Source Remote Sensing Imagery. IEEE Access, 1–1. doi:10.1109/access.2019.2928646
17. Guo, Z., Wu, G., Song, X., Yuan, W., Chen, Q., Zhang, H. … Shao, X. (2019). Super-Resolution Integrated Building Semantic Segmentation for Multi-Source Remote Sensing Imagery. IEEE Access, 1–1. doi:10.1109/access.2019.2928646
18. Nivaggioli, A., & Randrianarivo, H. (2019). Weakly Supervised Semantic Segmentation of Satellite Images. 2019 Joint Urban Remote Sensing Event (JURSE). doi:10.1109/jurse.2019.8809060
19. Zheng, Z., Zhong, Y., & Wang, J. (2019). Pop-Net: Encoder-Dual Decoder for Semantic Segmentation and Single-View Height Estimation. IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium. doi:10.1109/igarss.2019.889792
20. O. Ronneberger, P.Fischer, and T. Brox,"U-net: Convolutional networks for biomedical image segmentation", In MICCAI, pp.234-241, Springer, 2015

AUTHORS PROFILE

Chandra Pal Kushwah, received the Bachelor of Engineering [B.E.] in Electronics & Instrumentation Engineering from the ShriRam College of Engineering & Management (SRCEM), Gwalior [M.P.] in University of Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal [M.P.], India in 2012, and I am doing M.Tech. Degree now in Electronics Engineering from Madhav Institute of Technology & Science, Gwalior [M.P.], India.

Karuna Markam, graduated Bachelor of Engineering [B.E.] in Electronics & Communication from GEC, Bhopal [M.P.], M.Tech in Digital Communication from MANIT, Bhopal [M.P.] and PhD from Barkatullah University, Bhopal [M.P.] India. She has more than 18 years of teaching experience. She has guided 35 M.Tech Dissertation. She published 40 papers in various national and international journals and conferences.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication