WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Satellite Image Segmentation using Convolution Neural Network - U-Net

K.Pritika

Assistant Professor, Department of Electronics and Communication Engineering, St. Martins Engineering College, Dhulapally, Secunderabad-500100, Telangana, INDIA

ABSTRACT

Nowadays, large amounts of high resolution remote-sensing images are acquired daily. However, the satellite image classification is requested for many applications such as modern city planning, agriculture and environmental monitoring. Many researchers introduce and discuss this domain but still, the sufficient and optimum degree has not been reached yet. Hence, this article focuses on evaluating the available and public remote-sensing datasets and common different techniques used for satellite image classification. In recent years, there has been an extensive popularity of supervised deep learning methods in various remote-sensing applications, such as geospatial object detection and land use scene classification. In this paper, carried out on one of the popular deep learning models, Convolution Neural Networks (CNNs). This study proposed an urban village mapping paradigm based on U-netdeep learning architecture. The Worldviewsatellite image with eight pan-sharpened bands spatial resolution and building boundaryvector file were used as research purposes. The deep neural network model was trained and tested based on theselected six and four sites of the urban village, respectively. Models for building segmentation andclassification were both trained and tested. The results indicated that the U-net model reached overallaccuracy over 86% for building segmentation and over 83% for the classification.

Keywords: Remote-sensing data, Convolution Neural Networks (CNNs), U-net, Deep learning models.

1 INTRODUCTION

A Satellite Image is an image of the whole or part of the earth taken using artificial satellites. It can either be visible light images, water vapor images or infrared images [1]. The different types of satellites produce (high spatial, spectral, and temporal) resolution images that cover the whole Earth in less than a day. The large-scale nature of these data sets introduces new challenges in image analysis.The analysis and classification of remote-sensing images is very important in many practical applications, such as natural hazards and geospatial object detection, precisionagriculture, urban planning, vegetation mapping, and military monitoring [2]. Despite decades of research, the degree of automation for remote-sensing images analysis still remains low.

Remote sensing-based mapping of urban buildings has a long history with substantial. With the increasing availability of high-resolution optical images, along with object-based image analysis has dominated the area of mapping urban buildings [3]. More often, most current techniques for image target detection are based on spectral and spatial features. In recent years, Random Forest as a machine learning method has become one of the most popular approaches in built-up areas mapping. Although numerous classification methods have been developed for urban land use mapping for some high-density built-up areas, the widely used pixelor object-based methods usually does not allow the extraction of individual buildings, but typically clusters several buildings into one segment. In a complex urban built-up area, the features (spectral, texture, shape, etc.) are uneasily described by conventional remote sensing methods, owing to the large variance among urban buildings [4]. Segmentation is the process of partitioning an image into segments by grouping neighboring pixels with similar feature values (brightness, texture, color, etc.). In high-density built-up areas, segmentation with OBIA is often impeded by difficulties such as the scale selection and rule definition. It is challenging to completely delineate the boundaries and preserve their shapes because the noise and textures on the building's edges usually degrade the performance of image segmentation [5]. Additionally, a generalized and straightforward methodology is often hard to obtain and has not been reported. Therefore, developing a reliable and accurate building segmentation method towards mapping the unplanned urban settlements is still challenging.

In some applications, such as emergency mapping, satellite images must be segmented in a short time in the aftermath of events, such as flood or

earthquake [6]. In similar scenarios, the tight time constraints prompt for solutions that allow reusing some algorithms previously trained over different images.

Image segmentation is the separation pf image into signification. The simplest approach to solve this problem is manual segmentation of images. Nevertheless, is a laborious and long process, which usually leads to make mistakes [7]. Currently the great interest of researchers in the field of machine learning is associated with development of automatic image segmentation system. This type of segmentation allows to process images immediately the necessary accuracy to be useful in practical applications.

In this paper presents the satellite segmentation. It describes as follows: section 1 general consists of satellite image segmentation. The section 2 literature survey in different authors presents their methods. The section 3 consists of the proposed segmentation method. Section 4 has results for the simulations followed by conclusions and references.

2 Literature Review

S. Ghassemiet al., (2019) [8]. We propose a convolutional encoder-decoder network able to learn visual representations of increasing semantic level as its depth increases, allowing it to generalize over a wider range of satellite images. Then, we propose two additional methods to improve the network performance over each specific image to be segmented. First, we observe that updating the batch normalization layers' statistics over the target image improves the network performance without human intervention. Second, we show that refining a trained network over a few samples of the image boosts the network performance with minimal human intervention. We evaluate our architecture over three data sets of satellite images, showing the state-of-the-art performance in binary segmentation of previously unseen images and competitive performance with respect to more complex techniques in a multiclass segmentation task.

Badrinarayanan et al., (2017) [9] presented a deep fully convolutional neural network. The network developed for semantic pixel-wise segmentation which has been termed SegNet. It has an encoder network and a corresponding decoder network. The encoder and decoder is followed by a pixelwise classification layer.

Ronneberger et al., (2015) [10] developed a method based on deep learning for biomedical image segmentation which has been termed as U-Net. It has been modified and extended to the fully convolutional network architecture such that the network works with very a smaller number of training images to give more precise segmentations.

Yoshihara et al., (2017) [11] proposed a semantic segmentation method for satellite images using fully convolutional network. The architecture of the network comprises of an encoder network followed by a corresponding decoder network. The input size to the network has been changed from the employed value to 256 × 256. The encoder network architecture was same with of a convolutional network.

Chaurasia et al., (2017) [12] proposed a deep neural network (LinkNet) which allows it to learn without any significant increase in number of parameters. Their proposed deep neural network architecture attempted to efficiently share the information learnt by the encoder with the decoder after each downsampling block.

Wu et al., (2019) [13] propose attention dilation-LinkNet (AD-LinkNet) neural network that adopts encoder–decoder structure, serial–parallel combination dilated convolution, channel-wise attention mechanism, and pretrained encoder for semantic segmentation. Serial–parallel combination dilated convolution enlarges receptive field as well as assemble multi-scale features for multiscale objects, such as long-span road and small pool. The channel-wise attention mechanism is designed to advantage the context information in the satellite image. The experimental results on road extraction and surface classification data sets prove that the AD-LinkNet shows a significant effect on improving the segmentation accuracy.

Awad, Mohamad(2010) [14] Image segmentation is an essential step in image processing. The goal of segmentation is to simplify and/or to change the representation of an image into a form easier to analyze. Many image segmentation methods are available but most of these methods are not suitable for satellite images and they require a priori knowledge. In order to overcome these obstacles, a new satellite image segmentation method is developed using an unsupervised artificial neural network method called Kohonen's self-organizing map and a threshold technique. Self-organizing map is used to organize pixels according to grey level values of multiple bands into groups then a threshold technique is used to cluster the image into disjoint regions, this new method is called TSOM.

Shafaeyet al., (2020) [15] The present classification methods for remote-sensing images are grouped according to the features they use into: manual feature-based methods, unsupervised feature learning methods, and supervised feature learning methods. In recent times, the supervised deep learning approaches are extensively introduced in various remote-sensing applications, such as object detection and land use scene classification. In this article, an experiment is conducted using one of the widespread deep learning models, Convolution Neural Networks (CNNs), specifically, *AlexNet* architecture on a standard sounded hyper spectral dataset, *Pavia University (PaviaU)*. A comparison with other different techniques is also introduced.

3 Proposed method

Deep Learning Segmentation Using U-Net**.** The u-net is convolutional network architecture for fast and precise segmentation of images. Up to now it has outperformed the prior best method (a sliding-window convolutional network) on the ISBI challenge for segmentation of neuronal structures in electron microscopic stacks.

Semantic segmentation is a process of taking an image and labeling each pixel in that image with a specific class. More often, these processes are completed by a deep convolutional neural network (CNN). Through assistance from Computer Vision, semantic segmentation segments the image automatically. To do that, researchers can either use a pre-trained CNN to perform segmentation, or create a CNN; by labeling the targeted objects, researchers use the labeled data to perform network model training.

In the procedure was summarized as follows: (1) A large number of building footprints from the labeled image were converted from a building vector file; (2) The training data was partitioned into training sets and validation sets, then the satellite image associated with the label data was inputted to train a network model, and generate a model training report; (3) The model was applied to given test data. After the model training was completed, the test data was input into the network model, and the model was able to segment the targets from the background; and (4) The results were evaluated. Our fully convolutional model was inspired by the family of U-Net architectures, where low-level feature maps are combined with higher-level ones, which enables precise localization. This type of network architecture was especially designed to effectively solve image segmentation problems. U-Net was the default choice for us and other competitors.

4 Network Architecture

In this study, semantic segmentation was developed based on U-net architecture. It is noteworthy that an encoder-decoder architecture becomes increasingly popular in semantic segmentation due to its high flexibility and performance. The U-net architecture uses the following types of layers and special operations: (1) Conv2D, Simple convolution layers with padding and 3 × 3 kernel; (2) MaxPooling2D, Simple max-pooling layers with 2 × 2 kernel; (3) A cropping 2D, cropping layer used to crop feature maps and concatenate; (4) A concatenate layer used to concatenate multiple feature maps from different stages of training; (5) An UpSampling2D layer used to increase the size of the feature map. Then, the decoder part is implemented for up-sampling; and (6) Finally, a softmax layer is added to generate a final segmentationmap. The structures of the encoder and decoder parts are symmetrical with skip connections between them, which proves to be effective to produce fine-grained segmentation results. More importantly, U-net can preserve the feature maps to the same size as the original image, where up-sampling layers are followed by several Conv layers to produce dense features with finer resolutions. This study modified the deep-learning architecture to handle the Worldview image with eight bands. The modified deep learning architecture can accept any size of images ranging from megabyte to gigabyte. The advantage of the modified U-net is an "end-to-end" procedure; for instance, segmentation. For high-density building segmentation using images, the modified method can assign a class to every pixel according to which class exactly belongs to the intricate.

Figure 1 gives an intuitive demonstration of the U-net structure. Like other commercial networks, the architecture of the U-net consists of a large number of different operations illustrated by small arrows. There are several details and basic concepts that are applied to the U-net: (1) the convolutional layers for feature extraction through multiple 3 × 3 convolution kernels (denoted by Convolution); (2) the batch normalization layer for accelerating convergence during the training (denoted by Batch Normalization); (3) the activation function layer for nonlinear transformation of the feature maps, in which we adopted the widely used rectified linear unit (ReLU) (denoted by Activation); (4) the max-pooling layer for down-sampling of the feature maps (denoted by Max-pooling); (5) the up-sampling layer for recovering the size of the feature maps that are down-sampled by the max-pooling layer (denoted by up-sampling); and (6) the concatenation layer for combining the up-sampled feature map in the deep layers with the corresponding feature map from the shallow layers (denoted by Concatenation).

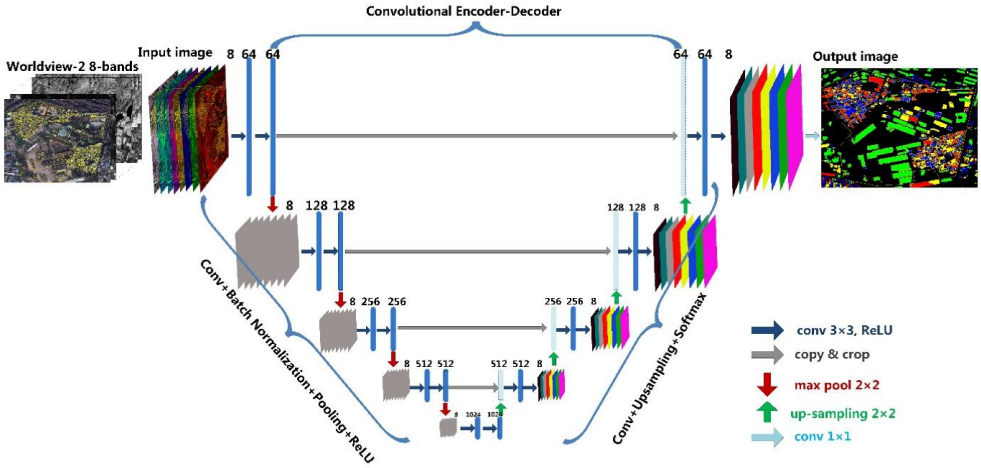


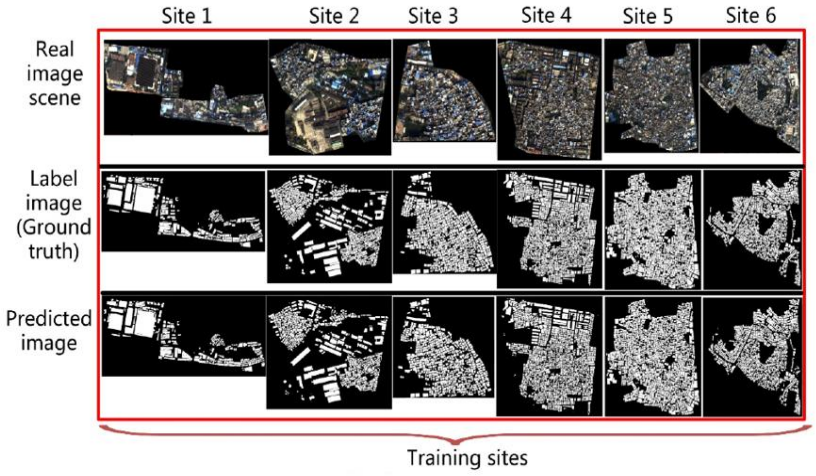
Figure 1. Illustration of the U-net Architecture for semantic segmentation, including the name and function of each layer.

In brief, the U-net is characterized by the convolutional max-pooling, cropping, concatenate, and up-sampling layers. It receives the input image and runs through two convolutional operations with ReLU activation. The image is then encoded into the pooling layer. This process happens a

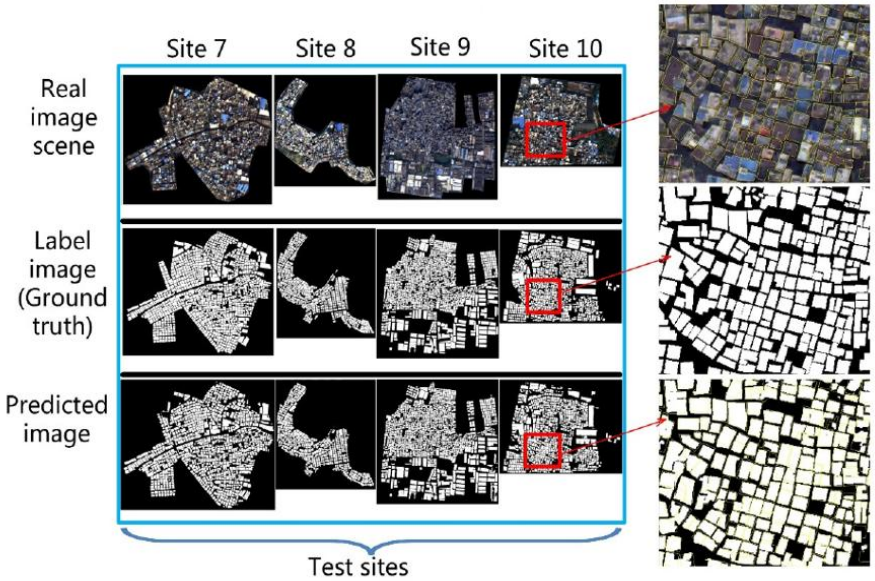
few times, reducing the size of the feature maps. Once 1024 sample maps are obtained, the model starts up-sampling. The layers of the feature maps are concatenated with the feature maps from the down-sampling process. The feature maps are up-sampled from the previous concatenated feature maps and the model finally outputs the segmentation map as the same size of the original image. From Figure 2, the data are propagated through the network by convolutional encoding and decoding with copy and crop steps to retain the image information. Along all the possible paths, at the end, the segmentation map is obtained.

5 Results

The results of building segmentation were presented in Figure 2, including the six training sites and four test sites within the red box and blue box, respectively. The predicted images in the training sites and test sites showed a good match with the label image (ground truth). For each high-density urban village with overcrowded, irregular-shaped buildings, the network model precisely delineated individual buildings. As presented in Figure 2, the data set used for training should get good prediction results; on the other hand, the test sites that also obtained good predicted maps visually looked like the label images. A snapshot of a magnified area also gave an in-depth inspection of the segmentation results; more importantly, it is shown that the boundaries of the adjacent buildings were correctly separated. Individual buildings are seldom missing, and boundary shapes are basically delineated and preserved.



A. training sites



B. Figure test sites

Figure 2. Performance of building segmentation by U-net.

The overall accuracy, All the training sites achieved overall accuracies of over 93%, and the test sites achieved an overall accuracy of over 86%. Moreover, both the training and test sites led to an Intersection over Union value over 90%, both of which indicate the excellent performance of the building's segmentation.

6 Conclusion

A deep learning paradigm based on U-net CNN waproposed for the semantic segmentation using a Worldview satellite image. The results indicated thatmost adjacent buildings were well separated; boundary shapes are basically delineated and preserved.The deep learning based on U-net significantly outperformed the widely used RF and OBIA, indicatingthat the deep learning provided greater potential to characterize the individual buildings in the complexurban villages. This study demonstrated the feasibility, efficiency, and potential of the deep learning indelineating the individual buildings in the high-density urban village. Therefore, it should contribute tothe knowledge gap-filling for remote sensing mapping of the unplanned urban settlements. This studyalso implies that integrating the deep learning based on U-net with the high spatial resolution satelliteimages can offer accuratebuilding information in the complex urban villages that is needed forurban redevelopment.

Reference

[1]. Wilkinson, Graeme G. "Results and implications of a study of fifteen years of satellite image classification experiments." *IEEE Transactions on Geoscience and remote sensing* 43, no. 3 (2005): 433-440.

[2]. Toutin, Thierry. "Geometric processing of remote sensing images: models, algorithms and methods." *International journal of remote sensing* 25, no. 10 (2004): 1893-1924.

[3]. Oikonomidis, D., S. Dimogianni, N. Kazakis, and K. Voudouris. "A GIS/remote sensing-based methodology for groundwater potentiality assessment in Tirnavos area, Greece." *Journal of Hydrology* 525 (2015): 197-208.

[4]. Waldhoff, Guido, Ulrike Lussem, and Georg Bareth. "Multi-Data Approach for remote sensing-based regional crop rotation mapping: A case study for the Rur catchment, Germany." *International Journal of Applied Earth Observation and Geoinformation* 61 (2017): 55-69.

[5]. Nogueira, Keiller, Mauro Dalla Mura, Jocelyn Chanussot, William Robson Schwartz, and Jefersson Alex dos Santos. "Dynamic multicontext segmentation of remote sensing images based on convolutional networks." *IEEE Transactions on Geoscience and Remote Sensing* 57, no. 10 (2019): 7503-7520.

[6]. Zanotta, Daniel Capella, MacielZortea, and MatheusPinheiro Ferreira. "A supervised approach for simultaneous segmentation and classification of remote sensing images." *ISPRS journal of photogrammetry and remote sensing* 142 (2018): 162-173.

[7]. Dong, Rongsheng, Xiaoquan Pan, and Fengying Li. "DenseU-net-based semantic segmentation of small objects in urban remote sensing images." *IEEE Access* 7 (2019): 65347-65356.

[8]. S. Ghassemi, A. Fiandrotti, G. Francini and E. Magli, "Learning and Adapting Robust Features for Satellite Image Segmentation on Heterogeneous Data Sets," in IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 9, pp. 6517-6529, Sept. 2019, doi: 10.1109/TGRS.2019.2906689.

[9]. 6. Badrinarayanan V, Kendall A, Cipolla R (2017) SegNet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell 39:2481–2495.

[10]. 18. Ronneberger O, Fischer P, Brox T (2015) U-Net: convolutional networks for biomedical image segmentation, computer science department and BIOSS centre for biological signalling studies, University of Freiburg, Germany

[11]. 19. Yoshihara A, Takiguchi T, Ariki Y (2017) Feature extraction and classification of multispectral imagery by using convolutional neural network. In: International workshop on frontiers of computer vision.

[12]. Chaurasia, Abhishek, and Eugenio Culurciello. "Linknet: Exploiting encoder representations for efficient semantic segmentation." In *2017 IEEE Visual Communications and Image Processing (VCIP)*, pp. 1-4. IEEE, 2017.

[13]. Wu, Ming, Chuang Zhang, Jiaming Liu, Lichen Zhou, and Xiaoqi Li. "Towards accurate high resolution satellite image semantic segmentation." *IEEE Access* 7 (2019): 55609-55619.

[14]. Awad, Mohamad. "An Unsupervised Artificial Neural Network Method for Satellite Image Segmentation." *Int. Arab J. Inf. Technol.* 7, no. 2 (2010): 199-205.

[15]. Shafaey, Mayar A., Mohammed A-M. Salem, Maryam N. Al-Berry, Hala M. Ebied, Elsayed A. El-Dahshan, and Mohammed F. Tolba. "Hyperspectral Image Classification Using Deep Learning Technique." In *Joint European-US Workshop on Applications of Invariance in Computer Vision*, pp. 334-342. Springer, Cham, 2020.