BUILDING DETECTION FROM SAR IMAGES USING UNET DEEP LEARNING METHOD

Recai Alper Emek1* and Nusret Demir²

 ¹ Akdeniz University, Institute of Science and Technology, Remote Sensing & GIS Graduate Program, 07058 Antalya, Turkey – recaialperemek@gmail.com
² Akdeniz University, Faculty of Science, Dept. of Space Science and Technologies, 07058 Antalya, Turkey – nusretdemir@akdeniz.edu.tr

KEY WORDS: SAR, U-Net, Deep Learning, Object Detection, Convolution Neural Network, Radar, Building

ABSTRACT:

SAR images are different from the optical images in terms of image properties with the values of scattering instead of reflectance. This makes SAR images difficult to apply the traditional object detection methodologies. In recent years, deep learning models are frequently used in segmentation and object detection purposes. In this study, we have investigated the potential of U-Net models for building detection from SAR and optical image fusion. The datasets used are Sentinel 1 SAR and Sentinel-2 multispectral images, provided from 'SpaceNet 6 Multi Sensor All- Weather Mapping' challenge. These images cover an area of 120 km² in Rotterdam, the Netherlands. As training datasets 20 pieces of 900 by 900 pixel sized HV polarized and optical image patches have been used together. The calculated loss value is 0.4 and the accuracy is 81%.

1. INTRODUCTION

Radar systems are one the types of active remote sensing with its capacity to operate in all weather conditions. The SAR images are created in a 2-dimensional plane by processing the scattering of the transmitted radar signals. Each pixel is an expression of the numerical code values of the retroreflection corresponding to its counterpart on the surface. The black and gray scales that make up the composition of the radar image show the strength of the reflected signals. SAR is a type of radar that provides higher resolution than the image normally obtained by a single larger unit by combining the radar images collected by more than one small radar unit electronically.

Different polarization of Radar signals provides various scattering from the objects with regard to their geometrical properties. Polarization simply refers to the orientation of the electric field, which is a resultant parameter of the electromagnetic wave. The most common polarization combinations are HH, VV, HV and VH. The first letter indicates the type of polarization sent, the second the type of the received back. When the SAR images were examined, the highest reflection value was seen at HV polarization in manmade objects. For this reason, only HV polarization band is used in the developed algorithm.

Building have strong backscattering with their suitability for the double-bouncing since they have many corners in their body. Therefore, buildings mostly appear brighter than the other objects on the Radar images.

On the other hand, deep learning methods are becoming popular in object detection, as will be seen in related works. One of the reasons is semantic segmentation gives better results (Long, et al., 2015). Accordingly, we used convolution neural network (CNN) which is a deep learning algorithm in this study.

Building detection is important for urban planning, traffic management, natural disasters and security etc. Deep learning

algorithms are frequently used at building detection research. Due to this reason building detection with applying deep learning is also hot topic in recent years.

Vakalopoulou el al., (2015) applied deep learning to detect the buildings from satellite images with applying. They trained some sample dataset and each of these patches is inserted into the ImageNet framework. Xu et al., (2018) first created new bands such as NDVI and included in the progress, they used Res-U-Net. Field boundaries were extracted from satellite imagery using ResUnet-a, which is an advanced convolutional neural network (CNN) model (Waldner et al., 2020). They formulated task as a multi-task semantic segmentation problem. Rice lodging was extracted using with U-net (Zhao et al., 2019) from high resolution aerial images. For road extraction Alshaikhli et al., (2019) applied CNN to aerial images. Proposed model's OAA is 98.35%.

Some of the studies on building detection from SAR images are listed below. The problem of object detection on SAR images is in the focus of some challenges and competition. One of them is building objects. Also, recent studies show that effective building detection is achieved from SAR images as well as optical images. Zhao et al., (2013) aimed extract building footprints on SAR images. Kaynarca and Demir (2018) detected buildings from SAR with fusing information from Gokturk multispectral images. Ferro et al., (2010) extracted low level features first then combined them to complete the building detection from high resolution SAR image. In the study (Marin, 2014) very high resolution SAR images were used by analyzing back scatter in the building change detection.

The use of Convolutional Neural Networks is also becoming common in SAR images. Xu et al., (2017) aimed to reduce the complexity, classified and extract building area from high resolution SAR images with using Res-U-Net deep learning method.

^{*} Corresponding author.

Not only building detection in SAR images with deep learning algorithms. Kang et al., (2017) focused ship detection on SAR images. In proposed method modified the Faster R-CNN algorithm for different size ships detection. Chen and Wang (2014) applied the CNN algorithm to the SAR images and achieved %87 classification accuracy. In this study they detected ground military targets.

2. DATA

2.1 Study area

The study area and dataset has been proposed by SpaceNet 6 Multi Sensor All- Weather Mapping challenge (SpaceNet, 2020). The test location is in Rotterdam, the Netherlands which is the largest port in Europe and where have thousands of buildings, vehicles, and boats of various sizes, which will make for an effective test bed for SAR and the fusion of these two types of data.



Figure 1. Study area

2.2 Data Type

In this work, the SAR images are from Capella Space with 50 cm ground sample distance. They are 20 pieces of 900 by 900 pixel sized HV polarized images. EO images taken from Maxar's WorldView 2 satellite.

2.2.1 Satellite Images

- SAR Images: 20 pieces of 900 by 900 pixel sized HV polarized images. Spatial resolution is 50 cm.
- EO Images: 20 pieces of 900 by 900 pixel. Spatial resolution is 50 cm. Include red, green, blue, near infrared bands.

- 2.2.2 Vector Data
 - Building footprints.



Figure 2. SAR Images with Building Footprints.

3. METHOD

In this study, semantic segmentation image processing models were selected. In this direction, we used the convolutional neural network (CNN) based Unet algorithm. CNN is the basic architecture of concept of deep learning for image processing. The size reduction in height and width that we apply along the convolutional neural network model is applied as a size increase in the second half of the model. Thanks to these layers, the resolution of the output is increased. For localization, high-resolution features and sampled output are combined throughout the model (Ronneberger, 2015).



Figure 3. Unet Algorithm

As can be seen in the figure, U-Net takes its name from its architecture similar to the letter U. At the exit from the input images, a segmented exit map is obtained. The most special aspect of its architect is its second half. The network has no fully connected layer. Only the convolution layer is used. Every standard convolution process is activated by ReLU. Pixels in the border region are added symmetrically around the image so that the images can be segmented continuously. Thanks to this strategy, the image is segmented completely.

Before to use Unet algorithm we developed preprocessing steps. Firstly generator created. The use of generators reduces processing time when working with large and multiple images. After that using red and near infrared bands calculated normalized difference vegetation index (NDVI). The purpose of this is to increase our overall accuracy by including the data we have produced. We applied KMeans which unsupervised classification method with red blue green and NDVI bands. Created building class mask from KMeans classes. From the building footprints data we have, created another binary mask for buildings. We masked the SAR images with this binary mask. HV band pixels inside the building mask are marked, averaged and recorded. HV band is re-masked with the building mask produced with KMeans. Pixels greater than the average SAR scatter value are also marked as buildings. The final mask is created from all pixel marked as buildings before Unet step. With these steps in the pixels that are not shown as a building in the building footprint data but appear as a building in the RGB image correctly classified. The final mask used as ground truth in Unet algorithm.



Figure 4. Processing Workflow

For samples with more than one tag, the sigmoid function is can be used as the activation function in the output layer. In this way, the class of the pixels is expressed as a percentage. In other words all result pixels take a value between 0 and 1. Sigmoid Function is :

$$S(x) = rac{1}{1+e^{-x}} = rac{e^x}{e^x+1}.$$
 (1)

Cross-entropy loss, or log loss, measures the performance of a classification model. Binary cross entropy loss function uses In case of classification with 2 classes (Goodfellow, 2016). Binary Cross Entropy is:

$$H_p(q) = -\frac{1}{N} \sum_{i=1}^{N} y_i \cdot \log(p(y_i)) + (1 - y_i) \cdot \log(1 - p(y_i))$$
(2)

4. RESULTS

The accuracy of the model produced on pilot images has been calculated to be 81 percent. Accuracy calculation is as follow:

$$Accuracy = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \quad (3)$$

Loss value is 0.4. Loss function is the function that measures the error rate of the designed model as well as its performance. The last layer of deep networks is the layer where the loss function is defined. The closer the loss value is to 0, the more compatible with the developed model data set. Loss value equal to 0 is undesirable. Because this means overfitting. We prevented overfitting by using validation dataset (Goodfellow, 2016).

After creating the final model, we applied it on the test SAR image.



Figure 5. Test SAR Image



Figure 6. Classified Image

As can see at figure 6 buildings are marked with high accuracy. But some of the high scattering pixels are marked as buildings although they do not belong to buildings. The main reason for this is the number of training data in the pilot application. Some of heavily wooded areas have high reflectance value. Therefore, classified as building some wooded area pixel.

5. CONCLUSIONS

In this work, the buildings are detected from high resolution SAR images with high degree of accuracy by applying deep learning methodology. The study shows the high potential of U-Net model not only for optical images as shown in the previous literature, but also for the SAR images.

To overcome the limitation of the proposed approach, the following recommendations are given as:

- Creation of training dataset for the entire dataset to develop the U-net model.
- Implementation of polarimetric decomposition to create and use volumetric scattering values

The future work includes also testing the approach in different resolution SAR images.

6. REFERENCES

Adamo. F., Brunner, D., Bruzzone, L., 2010. Building detection and radar footprint reconstruction from single VHR SAR images, IEEE International Geoscience and Remote Sensing Symposium, pp. 292-295, doi: 10.1109/IGARSS.2010.5652093.

Alshaikhli, T., Liu, W., Maruyama, Y., 2019. Automated method of road extraction from aerial images using a Deep convolutional neural network, applied science, 9(22), 4825. doi: https://doi.org/10.3390/app9224825

Chen, S., Wang, H., 2014. SAR target recognition based on deep learning, 2014 International Conference on Data Science and Advanced Analytics, doi: 10.1109/DSAA.2014.7058124

Goodfellow, I., Bengio Y., Courville, A., 2016, Deep Learning, MIT press, http://www.deeplearningbook.org/

Kang, M., Leng, X., Lin, Z., Ji, K., 2017. A modified faster R-CNN based on CFAR algorithm for SAR ship detection. 2017 International Workshop on Remote Sensing with Intelligent Processing, pp. 1-4, doi: 10.1109/RSIP.2017.7958815

Kaynarca, M., Demir, N.,2018. Detection of Urban Buildings by Using Multispectral Gokturk-2 and Sentinel 1A Synthetic Aperture Radar Images, MDPI Proceedings, vol.2, pp.360-360, doi: https://doi.org/10.3390/ecrs-2-05173

Long, J., Shelhamer, E., Darrell, T., 2015. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3431-3440

Marin, C., Bovolo, F., Bruzzone, L., 2014. Building Change Detection in Multitemporal Very High Resolution SAR Images, IEEE Transactions on Geoscience and Remote Sensing, Volume: 53, Issue: 5, doi: 10.1109/TGRS.2014.2363548

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation, Springer , Cham, pp 234-24, doi: https://doi.org/10.1007/978-3-319-24574-4_28

SpaceNet, 2020. SpaceNet 6 Multi Sensor All- Weather Mapping challenge, https://www.topcoder.com/challenges/ad4caf4a-f7c9-4330a5c2-56c7125b554f, February 2020.

Vakalopoulou, M., Karantzalos, K., Komodakis N., and Paragios, N., 2015. Building detection in very high resolution multispectral data with deep learning features, 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 1873-1876, doi: 10.1109/IGARSS.2015.7326158

Waldner, F., Diakogiannis, F., 2020. Deep learning on edge: extracting field boundaries from satellite images with a convolutional neural network, Remote Sensing of Environment vol. 245, 111741. doi: https://doi.org/10.1016/j.rse.2020.111741

Xu, Y., Wu, L., Xie, Z., Chen, Z., 2018. Building Extraction in Very High Resolution Remote Sensing Imagery Using Deep Learning and Guided Filters, Remote Sensing, 10, 144. doi: https://doi.org/10.3390/rs10010144

Xu, Z., Wang, R., Zhang, H., Li, Z., 2017, Building extraction from high-resolution SAR imagery based on deep neural networks, Remote Sensing Letters, vol. 8, issue 9. doi: https://doi.org/10.1080/2150704X.2017.1335906

Zhao, X., Yuan, Y., Song, M., Ding, Y., Lin, F., Liang, D., Zhang, D., 2019. Use of unmanned aerial vehicle imagery and Deep Learning Unet to extract rice lodging, sensors, 19(18), 3859. doi: https://doi.org/10.3390/s19183859