International Journal of Engineering and Geosciences (IJEG),
Vol; 5, Issue; 3, pp. 138-143, October, 2020, ISSN 2548-0960, Turkey,
DOI: 10.26833/ijeg.645426

FEATURE EXTRACTION FROM SATELLITE IMAGES USING SEGNET AND FULLY CONVOLUTIONAL NETWORKS (FCN)

Batuhan Sariturk[1]*, Bulent Bayram[2], Zaide Duran[1], Dursun Zafer Seker[1]

[1] Istanbul Technical University, Faculty of Civil Engineering, Department of Geomatics Engineering, Istanbul, Turkey (sariturkb/duranza/seker@itu.edu.tr); **ORCID 0000-0001-8777-4436, ORCID 0000-0002-1608-0119, ORCID 0000-0001-7498-1540**

[2] Yildiz Technical University, Faculty of Civil Engineering, Department of Geomatics Engineering, Istanbul, Turkey (bayram@yildiz.edu.tr); **ORCID 0000-0002-4248-116X**

*Corresponding Author, Received: 00/XX/201X, Accepted: 00/XX/201X

ABSTRACT: Object detection and classification are among the most popular topics in Photogrammetry and Remote Sensing studies. With technological developments, a large number of high-resolution satellite images have been obtained and it has become possible to distinguish many different objects. Despite all these developments, the need for human intervention in object detection and classification is seen as one of the major problems. Machine learning has been used as a priority option to this day to reduce this need. Although success has been achieved with this method, human intervention is still needed. Deep learning provides a great convenience by eliminating this problem. Deep learning methods carry out the learning process on raw data unlike traditional machine learning methods. Although deep learning has a long history, the main reasons for its increased popularity in recent years are; the availability of sufficient data for the training process and the availability of hardware to process the data. In this study, a performance comparison was made between two different convolutional neural network architectures (SegNet and Fully Convolutional Networks (FCN)) which are used for object segmentation and classification on images. These two different models were trained using the same training dataset and their performances have been evaluated using the same test dataset. The results show that, for building segmentation, there is not much significant difference between these two architectures in terms of accuracy, but FCN architecture is more successful than SegNet by 1%. However, this situation may vary according to the dataset used during the training of the system.

Keywords: *Photogrammetry, Deep Learning, Feature Extraction, SegNet, Fully Convolutional Networks*

1. INTRODUCTION

Building detection from remote sensing and photogrammetric images has been one of the most challenging tasks with important development and research efforts during recent years (Vakalopoulou et al., 2015). In remote sensing field, applications such as urban planning, land cover/use analysis and automatic generation or updating of the maps, along with the detection of buildings, are long-standing problems (Wu et al., 2018a).

Buildings, which are the most significant places for human life, are key elements in the mapping of urban areas (Chen et al., 2019). Due to the rapid changes in urban areas, it is important to create and update the location information of buildings (Wu et al., 2018b). Remote sensing has been an effective technology for accurate detection and mapping of buildings due to its capability for high-resolution imaging over large areas and advantages of fast and high accuracy data acquisition (Chen et al., 2019, Comert et al., 2019). Unfortunately, automatic building detection on aerial images is usually limited by the inadequate detection and segmentation accuracy (Chen et al., 2019). Most tasks still require great amounts of manual interventions by experts.

In recent years, as a consequence of the developments of imaging sensors and corresponding platforms, a rapid increase in the availability and accessibility of very high-resolution (VHR) remote sensing images has made this problem more and more urgent (Ma et al., 2017). In the literature, satellite images have been used widely for the classification of urban areas (Sevgen, 2019). Building extraction from satellite and aerial images is not an easy task because of complex backgrounds, different lightning conditions and external factors that reduce visibility or separability of buildings (Akbulut et al., 2018).

Recent progress in the field of computer vision (CV) indicates that, with the help of sufficient computing power and large training datasets (Cordts et al., 2016; Deng et al., 2009; Everingham et al., 2010; Lin et al., 2014), deep learning methods such as Convolutional Neural Networks (CNNs) (LeCun et al., 1989) can considerably improve the performance of object detection and segmentation tasks from high-resolution imagery (He et al., 2016; Krizhevsky et al., 2012). Neural networks can deal with complex problems to reach accurate solutions (Tasdemir & Ozkan, 2019). This situation strongly indicates that deep learning will play a critical role in promoting the accuracy of building segmentation toward practical applications of automatic mapping of features (Chen et al., 2019).

Since AlexNet overwhelmingly won the ImageNet Large-Scale Visual Recognition Challenge 2012 (LSVRC-2012) (URL-1), CNN-based algorithms have become the go-to standard in many computer vision tasks, such as image classification, object detection, and image segmentation (Wu et al., 2018a). In the beginning, researchers mainly applied patch-based CNN methods to detecting, classifying or segmenting buildings in aerial or satellite images and significantly improved the performances (Guo e al., 2016). However, as a result of extreme memory costs and low computational efficiency, Fully Convolutional Networks (FCNs) have eventually attracted more attention in this area (Wu et al., 2018a).

In this study, a comparison was made between SegNet and Fully Convolutional Networks (FCN) architectures. Inria Aerial Image Labeling Dataset which consists of 180 training images (with corresponding labels) and 180 test images was used. Two different models that use these architectures were trained using the prepared dataset and their performances have been evaluated. The creation of models and object segmentation processes were performed on the Python environment on Google Colab.

2. DATASET AND METHODOLOGY

2.1 Dataset

Dataset selected to be used is "Inria Aerial Image Labeling Dataset" (Maggiori et al., 2017). This dataset features:

- Coverage of 810 km² (405 km² for the training set and 405 km² for the testing set),
- Aerial (in color and orthorectified) imagery with a spatial resolution of 30 cm,
- Label images for two semantic classes: building and not building) (Maggiori et al., 2017).

The images from the dataset cover dissimilar urban settlements, differing from densely populated areas (e.g., Vienna) to less dense rural areas (e.g., Austrian Tyrol) (Fig 1) (Maggiori et al., 2017). The purpose of this is to improve the generalization power of the models (Maggiori et al., 2017). For example, while Chicago imagery may be used for training, the model should label images over other regions with varying conditions, urban landscape and time of the year (Maggiori et al., 2017).

Figure.1 Chicago - 5 sample image and corresponding label image (Maggiori et al., 2017)

In this study, only images from the training set were used. It is not possible to make comparisons between label images and predictions since there are no corresponding label images in the test set.

The training set contains 180 color images of size 5000×5000, covering a surface of 1500 m×1500 m each (Maggiori et al., 2017). There are 36 tiles for each of the following regions:

- Austin (TX, USA)
- Chicago (IL, USA)
- Kitsap County (WA, USA)
- Vienna (Austria)
- Western Tyrol (Austria) (Maggiori et al., 2017)

The format of the images is GeoTIFF. The pixels of label images have value 255 for building class and 0 for not building class (Maggiori et al., 2017).

To prepare the datasets for training and testing of the models, images from the training set and their corresponding label images were selected and divided into patches of size 224x224 pixels to reduce the computational cost and not lose resolution with resizing of images. Since the used architectures work with images in this size, images were prepared in size of 224x224.

To create the training dataset, 5 images and their corresponding label images were selected (Austin9, Chicago25, Kitsap18, Tyrol_w21 and Vienna15). For the test dataset, another 5 images were selected (Austin1, Chicago2, Kitsap30, Tyrol_w29 and Vienna9). During these selections, the distribution of rural and urban areas was considered. Images with no building or a low amount of buildings were removed from the datasets. Consequently, a total of 1500 images and label images for the training dataset and 300 images and label images for the test dataset were generated (Fig 2).

Figure.2 Sample image and corresponding label image from training dataset

2.2 Methodology

SegNet and FCN neural network architectures were used to train models using prepared training dataset.

2.2.1 SegNet

SegNet is a CNN architecture developed at Machine Intelligence Lab. of the University of Cambridge to design more suitable deep learning algorithms for image segmentation tasks (Badrinarayanan et al., 2017). SegNet has an encoder network and a decoder network that works according to this encoder, followed by a pixel-wise classification layer (Bozkurt, 2018).

Encoder network consists of 13 convolution layers, corresponding to the VGG16's first 13 convolution layers, which is a pre-trained network for object classification (Badrinarayanan et al., 2017). As mentioned in Badrinarayanan et al., 2017, at this network, convolutions and max-pooling are performed. At the deepest encoder output, fully connected layers are eliminated to protect higher resolution feature maps. This significantly reduces the number of parameters in the SegNet encoder network compared to other architectures.

Within the SegNet architecture, each encoder layer has its decoder layer (Badrinarayanan et al., 2017). Thus, the decoder network also has 13 layers (Badrinarayanan et al., 2017). The output of the last decoder layer produces probabilities of classes for each pixel, which feeds the classifier with probability values (Badrinarayanan et al., 2017). Illustration of the SegNet architecture is shown in Fig 3.

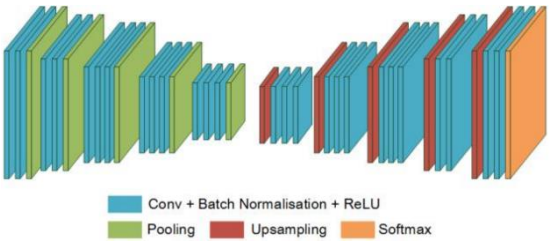


Figure.3 SegNet architecture (Du et al., 2018)

2.2.2 Fully Convolutional Networks (FCN)

Fully Convolutional Networks (FCNs) are being used for semantic segmentation of images, analysis of multi-modal medical images and classification and segmentation of high-resolution and multispectral satellite images (Long et al., 2015). In 2015, Long et al. adapted modern classification networks (AlexNet, VGGNet and GoogLeNet) into FCNs and transfer their learned representations by fine-tuning to the segmentation task. After that, they defined a novel architecture that combines semantic information from a deep, coarse layer with appearance information from a shallow, fine layer to produce accurate and detailed comprehensive (URL-2) (Fig 4).

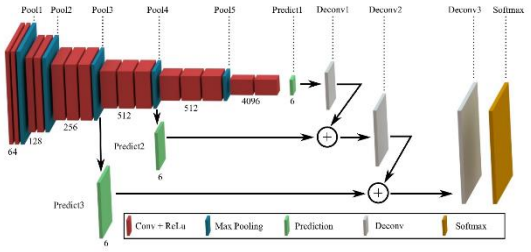


Figure.4 FCN architecture (De Souza, 2017)

FCNs built from locally connected convolutional, pooling and convolutional transpose layers (Long et al., 2015). No dense layer is used in this architecture (URL-3). The absence of dense layers makes it possible to feed the network in variable inputs (URL-3). An FCN has 2 parts:

- Downsampling path
- Upsampling path (URL-4)

As described in URL-4, downsampling path extract and interpret the context. The downsampling path consists of convolutional and max-pooling layers. Upsampling path enables precise localization of features. Upsampling path consists of convolutional, convolutional transpose and concatenate layers. Concatenation layers are used for skip connections. Skip connection is a type of connection that bypasses at least one layer. They are often used to transfer local information from the downsampling path to the upsampling path.

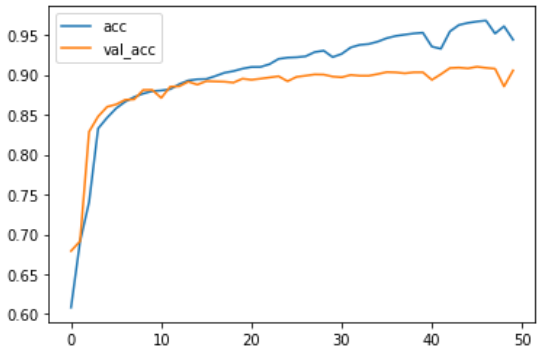
3. STUDY

In this study, all training and testing processes were conducted on Google Colab. Google Colab is a free Jupyter notebook environment that allows users to use free Tesla K80 GPU. It runs in the cloud and stores its notebooks and data on Google Drive.

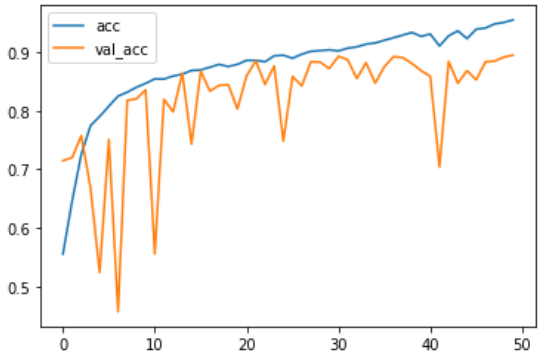
3.1 Training and Testing

To train the models, images loaded into the network. Thereafter, training dataset split according to an 85% / 15% training/validation ratio, 1275 images and 225 images respectively.

For training, the "Adam" optimizer was used to update model parameters with a fixed learning rate of 0.001. Both models were trained for 50 iterations with a batch size of 16 using the same hyperparameters. To calculate loss values, binary cross-entropy loss function was used. Changes in training accuracy and validation accuracy over 50 iterations are shown in Fig 5.



(a)



(b)

Figure.5 Accuracy values over 50 iterations (a) FCN (b) SegNet

To test the trained models, the test dataset that prepared separately from the training dataset was used.

3. RESULTS

The final accuracy results are shown in Fig 6. When the validation accuracy results examined it was seen that the model that uses FCN architecture has 94.39% training accuracy and 90.55% validation accuracy. On the other hand, the model that uses SegNet architecture has 95.49% training accuracy and 89.49% validation accuracy. FCN model is more accurate than the SegNet model by 1% according to validation accuracy results.

When training and validation accuracies of the models were compared, it was been seen that the FCN model has higher validation accuracy and the SegNet model has higher training accuracy.

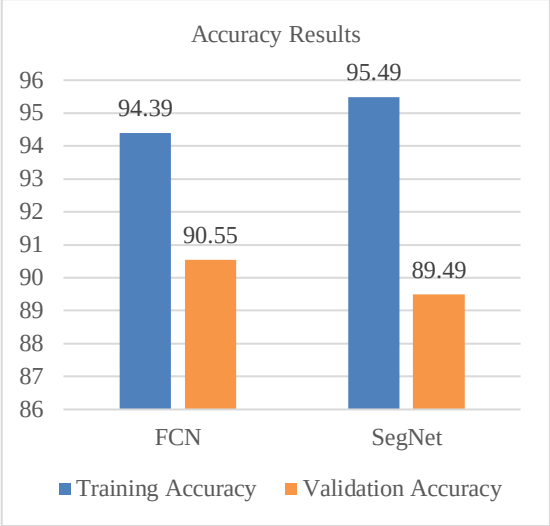


Figure.6 Training and validation accuracy results of models

When the differences between training and validation accuracies of the models examined, the model that uses SegNet architecture has a larger gap between them. This shows that the model's performance on training data is ahead of validation data. For the model that uses FCN architecture, this gap is smaller and it shows that this model is more accurate than the SegNet model.

Consequently, building segmentation was performed on the prepared test dataset using trained models. Examples from test, label and segmented images are shown in Fig 7 and 8.

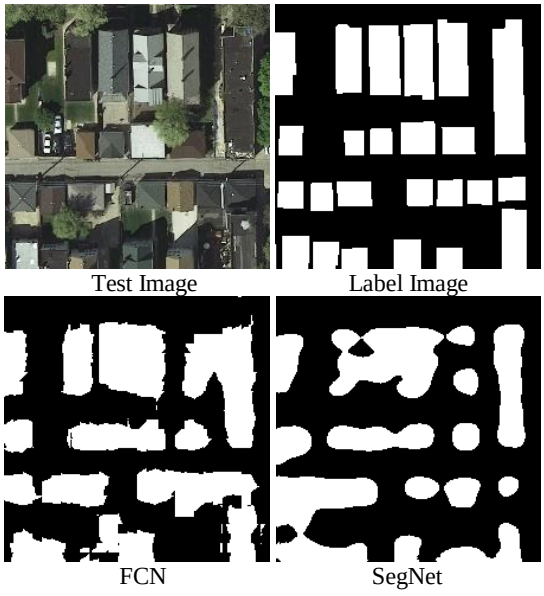


Figure.7 Segmentation results for test image 81

Figure.8 Segmentation results for test image 165

4. CONCLUSIONS

In this study, building segmentation from high-resolution images using SegNet and FCN neural network architectures were realized. Comparisons between these architectures were made. Models were trained and tested using datasets prepared from images from Inria Aerial Image Labeling Dataset.

It was observed that the model that uses FCN architecture gives more accurate results. It has higher accuracy and a smaller difference between training and validation accuracies. This can also be observed from the predicted segmentation results.

Further studies could include more datasets and different neural network architectures to make comparisons. Dataset could be augmented with unused images from Inria Dataset. More data to train the models would increase their performances. For this study, default settings were used for hyperparameters. Hyperparameter tuning could be done to improve the performances of the models. This is because hyperparameter optimization is crucial to achieve maximum performance.

REFERENCES

Akbulut, Z., Ozdemir, S., Acar, H., Dihkan, M., & Karsli, F. (2018). Automatic extraction of building boundaries from high resolution images with active contour segmentation. International Journal of Engineering and Geosciences, 3(1), 37-42.

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence, 39(12), 2481-2495.

Bozkurt S. (2018). Derin Ogrenme Algoritmalari Kullanilarak Cay Alanlarının Otomatik Segmentasyonu (Master's Thesis). YTU, Istanbul.

Chen, Q., Wang, L., Wu, Y., Wu, G., Guo, Z., & Waslander, S. L. (2019). Aerial imagery for roof segmentation: A large-scale dataset towards automatic mapping of buildings. ISPRS journal of photogrammetry and remote sensing, 147, 42-55.

Comert, R., Kucuk, D., & Avdan, U. (2019). Object Based Burned Area Mapping with Random Forest Algorithm. International Journal of Engineering and Geosciences, 4(2), 78-87.

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke U., Roth S. & Schiele, B. (2016). The cityscapes dataset for semantic urban scene understanding. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3213-3223).

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255). IEEE.

De Souza W. (2017). Semantic Segmentation using Fully Convolutional Neural Networks. Retrieved 19.03.2020, from https://medium.com/@wilburdes/semantic-segmentation-using-fully-convolutional-neural-networks-86e45336f99b

Du, Z., Yang, J., Huang, W., & Ou, C. (2018). Training SegNet for cropland classification of high resolution remote sensing images. In AGILE Conference.

Everingham, M., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2010). The pascal visual object classes (voc) challenge. International journal of computer vision, 88(2), 303-338.

Guo, Z., Shao, X., Xu, Y., Miyazaki, H., Ohira, W., & Shibasaki, R. (2016). Identification of village building via Google Earth images and supervised machine learning methods. Remote Sensing, 8(4), 271.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. Neural computation, 1(4), 541-551.

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollar P. & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In European conference on computer vision (pp. 740-755). Springer, Cham.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In

Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).

Ma, L., Li, M., Ma, X., Cheng, L., Du, P., & Liu, Y. (2017). A review of supervised object-based land-cover image classification. ISPRS Journal of Photogrammetry and Remote Sensing, 130, 277-293.

Maggiori, E., Tarabalka, Y., Charpiat, G., & Alliez, P. (2017). Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark. In 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) (pp. 3226-3229). IEEE.

Sevgen, S. C. (2019). Airborne lidar data classification in complex urban area using random forest: a case study of Bergama, Turkey. International Journal of Engineering and Geosciences, 4(1), 45-51.

Tasdemir, S., & Ozkan, I. A. (2019). Ann approach for estimation of cow weight depending on photogrammetric body dimensions. International Journal of Engineering and Geosciences, 4(1), 36-44.

URL-1, 2012, http://www.image-net.org/challenges/LSVRC/2012/results.html, [26.03.2020]

URL-2, 2017, https://meetshah1995.github.io/semantic-segmentation/deep-learning/pytorch/visdom/2017/06/01/semantic-segmentation-over-the-years.html, [19.03.2020].

URL-3, 2020, https://towardsdatascience.com/implementing-a-fully-convolutional-network-fcn-in-tensorflow-2-3c46fb61de3b, [19.03.2020].

URL-4, http://www.deeplearning.net/tutorial/fcn_2D_segm.html, [19.03.2020]

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

Wu, G., Guo, Z., Shi, X., Chen, Q., Xu, Y., Shibasaki, R., & Shao, X. (2018a). A boundary regulated network for accurate roof segmentation and outline extraction. Remote Sensing, 10(8), 1195.

Wu, G., Shao, X., Guo, Z., Chen, Q., Yuan, W., Shi, X., Xu Y. & Shibasaki, R. (2018b). Automatic building segmentation of aerial imagery using multi-constraint fully convolutional networks. Remote Sensing, 10(3), 407.