AUTOMATIC ROAD EXTRACTION USING MACHINE LEARNING APPROACH

B. Bayram [a], S. Sivri [b], F. Erdem [c, *]

[a] Yıldız Technical University, Dept. of Geomatics Engineering, 34220 Esenler, Istanbul, Turkey - bayram@yildiz.edu.tr
[b] Yıldız Technical University, Dept. of Geomatics Engineering, 34220 Esenler, Istanbul, Turkey - - sinan.sivri1@gmail.com
[c] Eskisehir Technical University, Institute of Earth and Space Sciences, 26555 Tepebasi, Eskişehir, Turkey - firaterdem@eskisehir.edu.tr

KEY WORDS: Deep Learning, Photogrammetry, Orthophoto, Image Processing

ABSTRACT:

Monitoring of highways has vital importance for logistic infrastructure and services provided by related government agencies. Remote sensing imagery data can be cost effective for mapping highways, which require up-to-date information about them. Many different image processing methods are used to extract highways from high resolution images. Although these techniques produce good results in data sets with specific colours and textures, however, they may not be effective when applied on different dataset. In recent years, deep learning architectures have been widely used in different object extraction applications to obtain accurate results from complex datasets.

The aim of this study is to extract highways from high resolution orthophotos by using deep learning techniques. Both U-Net and SegNet deep learning architectures have been evaluated on our dataset. In this study, 8-bit, 3-band (Red, Green, Blue), 10 cm ground sample distance orthophoto images of the province of Balıkesir the district of Savaştepe were used. For this purpose, the. The images were divided into small patches of size 512x512 pixel to train the architectures. 1098 samples were selected for training while 450 samples were used for testing. To evaluate the effective of both methods, DICE index was employed. Both methods were effective for road extraction and the calculated accuracies for SegNet and U-Net were %93.23, %95.95 respectively.

Transportation infrastructure is one of the most important economic components of the countries. A sustainable transportation infrastructure management can be carried out by analysing both the current situation and the temporal changes. In addition, during natural disasters (such as floods, earthquakes, etc.), to find usable access roads has crucial importance for crisis management. Traditionally, information on highways can be obtained in vector format by photogrammetric evaluation by operators. However, manual data collection can be time-consuming and prone to digitization and topological errors. This can cause data conversion difficulties to different formats and prevent to obtain well-structured data. The integration of extracted features with geographic information systems (GIS) from satellite and aerial images is indispensable for decision makers. Highway information is demanded for GIS platforms, civil and military purposes, navigation, location based systems and emergency plans and applications. Automatic road extraction will increase time and effort for generation of road data (Chaudhuri, et al, 2012).

Various approaches have been proposed for road extraction from remote sensing imageries. Active contour model (Anil and Natarajan, 2010), support vector machines (Simler, 2011), mathematical morphology (Ma et al, 2012), Mean-Shift (Miao et al, 2014) etc. In addition, deep learning algorithms, which have gained popularity in recent years, have also been used for automatic road segmentation (Zhang et al, 2018; Xu et al, 2018; Li et al, 2019).

Deep learning is a machine learning method based on learning computers by using the hierarchy of simple concepts in a similar way to people (Goodfellow et al, 2016). The main advantage of deep learning based approach compare to traditional machine learning algorithms is the ability to extract features automatically (Patterson & Gibson, 2017). In traditional machine learning algorithms, the features were hand-engineered which is very time-consuming. In addition, the user have to perform feature analysis on the dataset to obtain highly distinctive features.

The main objective of this study is to test the efficiency of deep learning based approach for automatic extraction of roads from high resolution orthophotos efficiently. We investigated the performance of both U-Net and SegNet architectures and the results were evaluated using DICE index. .

1. MATERIAL AND STUDY AREA

The orthophoto images include different type of land cover/use classes such as urban agricultural and forest areas. The technical specifications of the images are summarized in Table 1.

Number of Images	21
Number of Rows	5438
Number of column	6979
Number of Bands	3 (Red, Green, Blue)
Radiometric Resolution	8 bit
Ground Sample Distance	10 cm

Tabl1 1. Specifications of used orthophoto images.

The study area consisted of, Balıkesir province, Savaştepe district centre and its surrounding regions. Figure 1 shows the selected study area

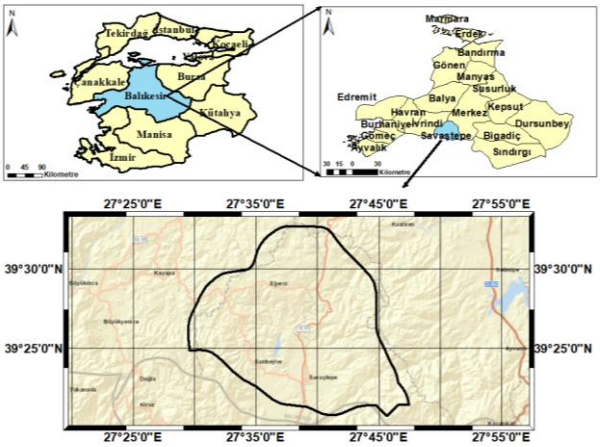

Figure 1. Study area.

2. METHODOLOGY

3.1 Preparation of training and reference dataset

The labelled images were prepared by digitizing and binarizing the images where white pixels represent road networks while black pixels represent background. (Figure 2).


Figure 2. Reference binarised images showing roads.

In order to train the deep learning models, the orthophotos and their reference images were subdivided into smaller patches of size 512x512 pixels. To avoid the effect of noise on the classification accuracy, the subset images having less than 1% area of roads shave been removed. 1098 randomly selected image patches were used for training while remaining 450 mages were used for testing. Figure 3 shows some samples images with respective reference binary images.


Figure 3. Sample dataset.

3.2 SegNet and U-Net architectures

The SegNet architecture consists of encoder, decoder layers followed by a pixel-based classification layer (Badrinarayanan et al., 2015). Encoding layers consist of layers of deep learning architecture VGG16 (Simonyan & Zisserman, 2014) developed for object detection.

3 x 3 convolution layers are applied for each scale. After each convolution layer, batch normalization and ReLU operations are applied. 2 x 2 max-pooling is utilized after this step. In the decoder layers, the matching of the low resolution feature maps produced in the encoder layers to the same resolution with the input image is performed. This is provided by transferring the indices calculated with maximum pooling in the encoding to the decoder layer with conjugate scale. After each upsampling, convolution, cluster normalization and ReLU operations are applied. The Softmax classification layer is applied to calculate the class probabilities for each pixel in the final layer. The SegNet architecture is given in Figure 4.

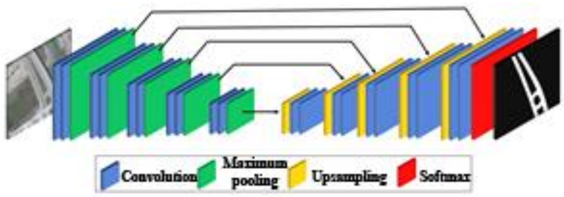

Figure 4. SegNet architecture.

The U-Net architecture developed for segmentation of biomedical images is also a fully convolutional network same as the SegNet architecture (Long et al., 2015). U-Net consists of two parts: a contracting path and an expansive path. In the contracting path section, for each scale, two 3 x 3 convolution layers after the convolution layers, the ReLU activation function followed by 2 x 2 maximum pooling layer. In the expansive path, the feature maps extracted in the contracting path are copied in to the scale they belong to in contracting path. This operation is called as concatenate. In the expansive path, 3x3 convolution is applied after each concatenate operation. Cluster normalization is followed after convolution. In the last layer, a convolution process is applied. The number of filters in the convolution is equal to the number of classes and their size is 1x1. The class probability for each pixel is calculated using the sigmoid function. The U-Net architecture used in this study is given in Figure 5.

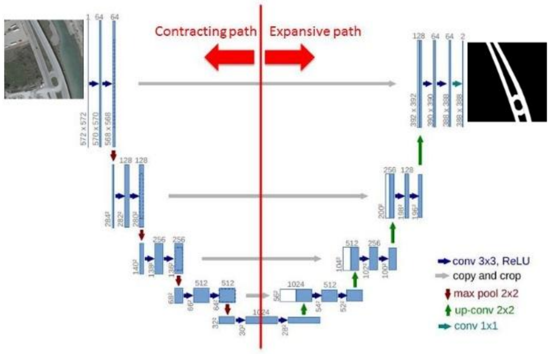

Figure 4. U-Net architecture (Ronneberger et al., 2015).

3. RESULTS AND DISCUSSION

SegNet and U-Net deep learning architectures are implemented in the Python Keras library (Chollet, 2019). The hyper parameters were tuned by searching on the training data for both SegNet and U-Net models. The empirically calculated values for hyper parameters are summarized in Table 2.

Hyper parameters	SegNet	U-Net
Optimization method	AdaDelta	Adam
Initial learning rate	1,00	0,0001
Number of epoch	50	131
Cluster size	4	4

Table 2. The used hyper parameters

Training and test results for SegNet and U-Net architectures are given in Figure 5 and Figure 6, respectively. The training results according to first 50 epochs show that SegNet architecture was less stable than U-Net. The calculated errors also support this situation. Especially in 5-22 epoch range, SegNet architecture created higher errors than U-Net.

450 subset images have been used for testing both SegNet and U-Net architectures. DICE index has been exploited for result binary images (road and others). Accuracy assessment results have been given in Table 3. The accuracy of SegNet and U-Net architectures have been calculated as %93.23 and %95.95, respectively. The obtained accuracy with U-Net architecture is %2.72 higher than SegNet architecture.

SegNet	U-Net
% 93.23	% 95.95

Table 3. Accuracy assessment results

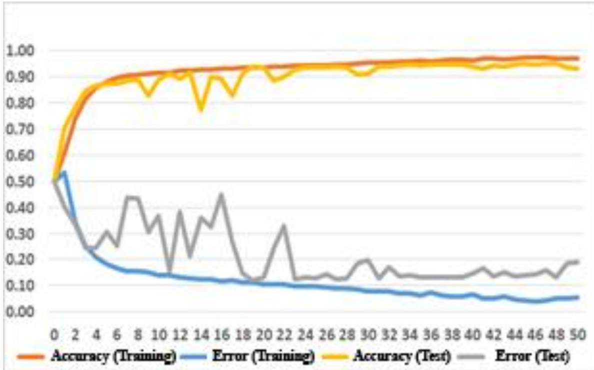

Figure 5. Accuracy assessment results of SegNet architecture

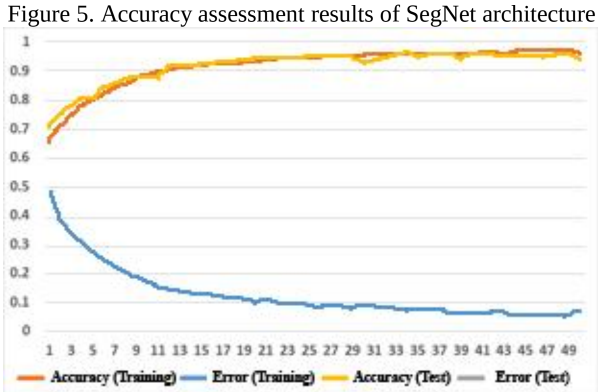

Figure 6. Accuracy assessment results of U-Net architecture

The sample segmented binary images obtained by applying the proposed method are shown in Figures 7, 8 and 9. In the figures, the white pixels correspond to the segmented roads, while the black pixels represent other objects. In Figure 7, it can be seen that road has not been segmented properly with SegNet model. In Figure 8, it is observed that the SegNet model combined nonadjacent roads. In Figure 9, ground object has been inaccurately segmented as road with SegNet model.

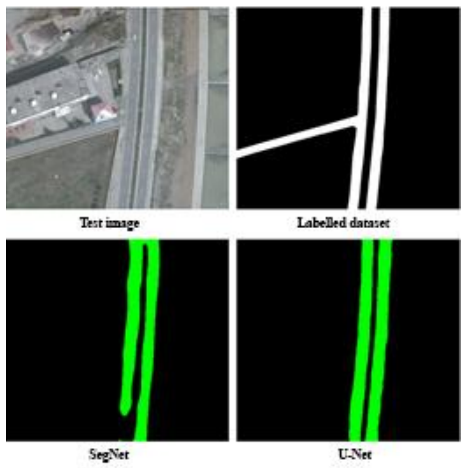

Figure 7. Segmentation results of first test image

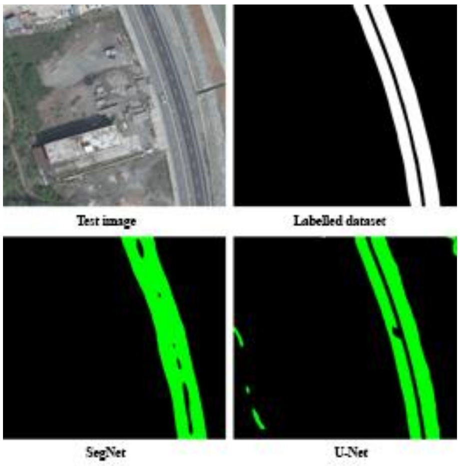

Figure 8. Segmentation results of first test image

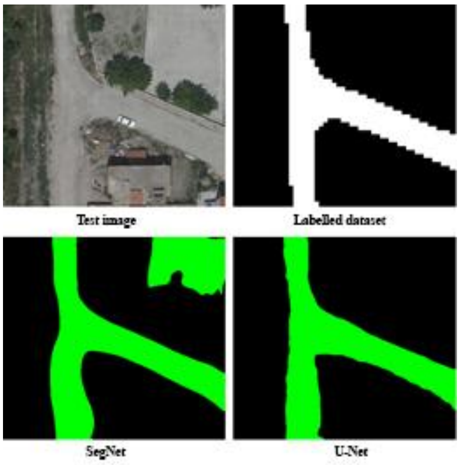

Figure 9. Segmentation results of first test image

4. CONCLUSION

This study investigated two deep learning approaches (SegNet and U-Net) for road extraction from high resolution orthophoto images. The results indicate that U-Net was more effective than SegNet. In SegNet, the upper sampling is performed using max-pooling, whereas in U-Net, the upper sampling process is performed by copying the feature maps. It is believed that this process causes data loss and SegNet creates smother borders compare to U-Net..

In future we would like to test other semantic segmentation models and a specific deep learning model will be designed for road extraction from orthophoto images.

ACKNOWLEDGMENT

This study was carried out according to corporate project between Yildiz Technical University, Department of Geomatics Engineering and General Directorate of Mapping (08 March 2017, TEK.HİZ.:35415333-0640-16484-17/P.P.D./Yurt İçi İlş.Ş. (340250)) which is entitled with "Three-Dimensional Extraction of Road Data by Digital Image Processing Techniques: The Case of Bergama / İzmir District"

REFERENCES

Anil, P. N., Natarajan, S., 2010. A novel approach using active contour model for semi-automatic road extraction from high resolution satellite imagery. In *2010 Second International Conference on Machine Learning and Computing*, pp. 263-266.

Badrinarayanan, V., Handa, A., Cipolla, R., 2015. Segnet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling, *arXiv preprint*, arXiv:1505.07293.

Chollet, F., https://github.com/fchollet/keras/ (Last Updated: March 27, 2019).

Chaudhuri, D., Kushwaha, N. K., Samal, A., 2012. Semi-automated road detection from high resolution satellite images by directional morphological enhancement and segmentation techniques. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 5(5), pp. 1538–1544.

Goodfellow, I., Bengio, Y., Courville, A.., Bengio, Y., 2016. *Deep Learning*. MIT Press, Cambridge.

Li, Y., Xu, L., Rao, J., Guo, L., Yan, Z., Jin, S., 2019. A Y-Net deep learning method for road segmentation using high-resolution visible remote sensing images. *Remote Sensing Letters*, 10(4), pp. 381-390.

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, R., Wang, W., Liu, S., 2012. Extracting roads based on Retinex and improved Canny operator with shape criteria in vague and unevenly illuminated aerial images. *Journal of Applied Remote Sensing*, 6(1), pp. 063610.

Miao, Z., Wang, B., Shi, W., Zhang, H., 2014. A semi-automatic method for road centerline extraction from VHR images. *IEEE Geoscience and Remote Sensing Letters*, 11(11), pp. 1856-1860.

Patterson, J., Gibson, A., 2017. *Deep Learning: A Practitioner's Approach*, California, O'Reilly Media.

Ronneberger O., Fischer P., Brox T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. *In: Navab N., Hornegger J., Wells W., Frangi A. (eds) Medical Image Computing and Computer-Assisted Intervention– MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Springer, Cham.*

Simler, C., 2011. An improved road and building detector on VHR images. In *2011 IEEE International Geoscience and Remote Sensing Symposium*, pp. 507-510.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition, *arXiv preprint*, arXiv:1409.1556.

Xu, Y., Xie, Z., Feng, Y., Chen, Z., 2018. Road Extraction from High-Resolution Remote Sensing Imagery Using Deep Learning. *Remote Sensing*, 10(9), pp. 1461.

Zhang, Z., Liu, Q., Wang, Y., 2018. Road extraction by deep residual U-net. *IEEE Geoscience and Remote Sensing Letters*, 15(5), pp. 749-753.