



International Journal of Environment and Geoinformatics (IJEGEO) is an international, multidisciplinary, peer reviewed, open access journal.

## **Comparison of Different U-Net Models for Building Extraction from High-Resolution Aerial Imagery**

**Firat ERDEM, Ugur AVDAN**

### **Chief in Editor**

Prof. Dr. Cem Gazioğlu

### **Co-Editors**

Prof. Dr. Dursun Zafer Şeker, Prof. Dr. Şinasi Kaya,

Prof. Dr. Ayşegül Tanık and Assist. Prof. Dr. Volkan Demir

### **Guest Editor**

Dr. Nedim Onur AYKUT

### **Editorial Committee (December 2020)**

Assos. Prof. Dr. Abdullah Aksu (TR), Assit. Prof. Dr. Uğur Algancı (TR), Prof. Dr. Bedri Alpar (TR), Prof. Dr. Levent Bat (TR), Prof. Dr. Paul Bates (UK), İrşad Bayırhan (TR), Prof. Dr. Bülent Bayram (TR), Prof. Dr. Luis M. Botana (ES), Prof. Dr. Nuray Çağlar (TR), Prof. Dr. Sukanta Dash (IN), Dr. Soofia T. Elias (UK), Prof. Dr. A. Evren Erginal (TR), Assoc. Prof. Dr. Cüneyt Erenoğlu (TR), Dr. Dieter Fritsch (DE), Prof. Dr. Çiğdem Göksel (TR), Prof. Dr. Lena Halounova (CZ), Prof. Dr. Manik Kalubarme (IN), Dr. Hakan Kaya (TR), Assist. Prof. Dr. Serkan Kükrer (TR), Assoc. Prof. Dr. Maged Marghany (MY), Prof. Dr. Michael Meadows (ZA), Prof. Dr. Nebiye Musaoğlu (TR), Prof. Dr. Masafumi Nakagawa (JP), Prof. Dr. Hasan Özdemir (TR), Prof. Dr. Chryssy Potsiou (GR), Prof. Dr. Erol Sarı (TR), Prof. Dr. Maria Paradiso (IT), Prof. Dr. Petros Patias (GR), Prof. Dr. Elif Sertel (TR), Prof. Dr. Nüket Sivri (TR), Prof. Dr. Füsün Balık Şanlı (TR), Prof. Dr. Uğur Şanlı (TR), Duygu Ülker (TR), Prof. Dr. Seyfettin Taş (TR), Assoc. Prof. Dr. Ömer Suat Taşkın (US), Assist. Prof. Dr. Tuba Ünsal (US), Dr. İnese Varna (LV), Dr. Petra Visser (NL), Prof. Dr. Selma Ünlü (TR), Prof. Dr. Murat Yakar (TR), Assit. Prof. Dr. Sibel Zeki (TR)



Dear colleagues and friends,

**International Symposium on Applied Geoinformatics (ISAG2019)** was held in Istanbul on 7-9 November 2019. The symposium is organized with the aim of promoting the advancements to explore the latest scientific and technological developments and opportunities in the field of **Geoinformatics**.

The symposium was jointly organized by the **Department of Geomatics Engineering, Yıldız Technical University, Istanbul, Turkey** and the **Institute of Geodesy and Geoinformatics, University of Latvia, Riga-Latvia**.

Our main aim was to bring researchers to share knowledge and their expertise about state-of-art developments in the field of **Geoinformatics**. We wish to discuss the latest developments, opportunities and challenges that can help the **Geoinformatics** community to solve many real-world challenges. Although this forum is initiated by two countries, Turkey and Latvia, it has a global perspective to promote technologies and advancements that would help us live in a better world.

**290 participants and scientists from 27 countries were attended to the ISAG2019. 118 oral and 16 poster presentations were presented by 45 international and 89 Turkish presenters in 29 sessions between 7-9 November 2019.**

We are much thankful to our supporting institutions Turkish General Directorate of Mapping, The Embassy of Latvia in Turkey, General Directorate of Geographical Information Systems/Turkey, Fatih Municipality.

The presentation "XXX" was presented at the **ISAG2019** and was proposed by our scientific committee for evaluation in the **International Journal of Environmental and Geoinformatics (IJECEO)**.

**The next ISAG symposium will be organized in Riga, Latvia on 16-17 November 2021. I do really hope to see you all in Latvia at the 2<sup>nd</sup> ISAG Symposium.**

**On behalf of ISAG-2019 Organization Committee**

**Conference Chair**

**Prof. Dr. Bülent Bayram**



# Comparison of Different U-Net Models for Building Extraction from High-Resolution Aerial Imagery

Firat Erdem\*,  Ugur Avdan 

Eskisehir Technical University, Earth and Space Sciences Institute, Eskisehir, Turkey

\* Corresponding author: F. Erdem  
E-mail: firaterdem@eskisehir.edu.tr

Received 05 Feb 2020  
Accepted 15 July 2020

**How to cite:** Erdem and Avdan (2020). Comparison of Different U-Net Models for Building Extraction from High-Resolution Aerial Imagery. *International Journal of Environment and Geoinformatics (IJECEO)*, 7(3): 221-227. DOI: 10.30897/ijegeo.684951

## Abstract

Building extraction from high-resolution aerial imagery plays an important role in geospatial applications such as urban planning, telecommunication, disaster monitoring, navigation, updating geographic databases, and urban dynamic monitoring. Automatic building extraction is a challenging task, as the buildings in different regions have different spectral and geometric properties. Therefore, the classical image processing techniques are not sufficient for automatic building extraction from high-resolution aerial imagery applications. Deep learning and semantic segmentation models, which have gained popularity in recent years, have been used for automatic object extraction from high-resolution images. U-Net model, which was originally developed for biomedical image processing, was used for building extraction. The encoder part of the U-Net model has been modified with Vgg16, InceptionResNetV2, and DenseNet121 convolutional neural networks. Therefore, building extraction was performed using Vgg16 U-Net, InceptionResNetV2 U-Net, and DenseNet121 U-Net models. In the fourth method, the results obtained from each U-Net model were combined in order to obtain the final result by maximum voting. This study aims to compare the performance of these four methods in building extraction from high-resolution aerial imagery. Images of Chicago from the Inria Aerial Image Labeling Dataset were used in the study. The images used have 0.3 m spatial resolution, 8-bit radiometric resolution, and 3-band (red, green, and blue bands). Images consist of 36 tiles and they were divided into image subsets of 512x512 pixels. Thus, a total of 2715 image subsets were formed. 80% of the image subsets (2172 image subset) were used as training and 20% (543 image subset) as testing. To evaluate the accuracy of methods, the F1 score of the building class was employed. The F1 scores for building class have been calculated as 0.866, 0.860, 0.856, and 0.877 on test images for U-Net Vgg16, U-Net InceptionResNetV2, U-Net DenseNet121, and majority voting method, respectively.

**Keywords:** building extraction, deep learning, remote sensing, semantic segmentation, U-Net

## Introduction

Automatic object extraction using remote sensing technology has become a popular research topic. Nowadays, the importance of automatic object extraction has increased with the availability of high spatial resolution images (Cheng and Han, 2016; Dervişoğlu et al., 2020). Building extraction from high-resolution aerial imagery plays an important role in geospatial applications such as urban planning, telecommunication, disaster monitoring, navigation, updating geographic databases, and urban dynamic monitoring (Ghanea, et al., 2016). Building extraction can be done by manual digitization techniques by experts. However, this process requires a lot of time and experience. Besides, manual digitization may lead to inattention errors.

Automatic building extraction is a challenging task, as the buildings in different regions have different spectral and geometric properties. Also, various complex factors like various scales, shadow, man-made non-building features, and heterogeneity of roof make automatic building extraction from high-resolution aerial imagery quite a challenging task (Yang, et al., 2018). Therefore, classical image processing techniques are not sufficient

for automatic building extraction applications. Deep learning and semantic segmentation models, which have gained popularity in recent years, have been used for automatic object extraction from high-resolution images.

Unlike other methods, deep learning methods are capable of extracting low-level and high-level features automatically (Patterson and Gibson, 2017; Esetlili et al., 2018; Çelik and Gazioğlu, 2020). The feature extraction process is performed manually by the data scientist in traditional machine learning algorithms, while this process is automatic in deep learning algorithms. Deep learning models like Convolutional Neural Networks (CNNs) use convolutions for automatic feature extraction. In 2014, pixel-based classification was made possible by adapting the CNN model to a fully convolutional neural network (Long, et al., 2015). Since then, deep learning models have been frequently used in semantic segmentation and object extraction studies (Lin, et al., 2019).

In recent years, studies using deep learning methods have been conducted in the fields of remote sensing such as image preprocessing (Huang, et al., 2015), target detection (Chen, et al., 2014), pixel-based classification (Hu, et al., 2015) and scene understanding (Zhang, et al.,

2016). Various studies have also been carried out in the field of automatic building extraction. Yang et al. (2018) proposed a novel network depend on DenseNets and the attention mechanism for utilizing different level features rationally. X. Li et al. (2018) designed a new deep adversarial network called Building-A-Nets which uses the adversarial structure as robust segmentation of building rooftops. L. Li et al. (2018) presented a novel CNN model called multiple-feature reuse network (MFRN) for reducing the GPU memory requirements. Lu et al. (2018) used a richer convolutional features (RCF) to detect building edges from high spatial resolution remote sensing imagery. Bittner et al. (2018) performed a Fully Convolutional Network (FCN) that effectively combines the high-resolution imagery with normalized DSMs and automatically generates building predictions. Xu et al. (2018) extracted buildings from high-resolution remote sensing imagery using Res-U-Net deep learning architecture and guided filters. Boonpook et al. (2018) applied SegNet deep learning architecture for building extraction from very high-resolution Unmanned Aerial Vehicle (UAV) images. H. Liu et al. (2019) proposed a fully convolutional network (DE-Net) which is created for information preservation through network computation, especially in downsampling, encoding, and upsampling procedures. Yi et al. (2019) compared the building extraction performance of proposed DeepResUnet with other semantic segmentation architectures which are FCN-8s, SegNet, DeconvNet, U-Net, ResUNet, and DeepUNet. Ji et al. (2019) presented a robust FCN which consists of Atrous convolutions and multi-scale aggregation to extract buildings from an open aerial and satellite dataset. Hui et al. (2019) designed a multitask driven deep neural network to extract unique features of buildings like shape and boundary. Pan et al. (2019) used a generative adversarial network model with spatial and channel attention mechanisms for selecting more useful features for building extraction. P. Liu et al. (2019) proposed SRI-Net which can capture and aggregate contexts from multi-scales. Schuegraf and Bittner (2019) designed a U-shaped neural network, which efficiently merges depth and spectral information within two parallel networks. Sun et al. (2019) used multiscale deep features, Support Vector Mechanism (SVM) based fusion strategy, and the superpixels refinement for building extraction. Lin et al. (2019) developed a deep learning architecture called ESFNet which consists of residual blocks and dilated convolutions for computational efficiency. Y. Liu et al. (2019) examined the loss of information caused by the use of pooling operations and suggested a light-weight deep learning model for its solution. Ji and Lu (2019) proposed a novel FCN based Siamese architecture and tested the architecture using their dataset which contains images from different sources. Zhang et al. (2020) improved the building extraction efficiency of well-known deep learning model Mask R-CNN using Sobel edge detection algorithm.

In this study, the results obtained from various U-Net models were compared and a new majority vote-based method was proposed. Accuracy analysis was performed with the F1 score method and the results were discussed.

## Materials and Methods

### Data Set

In the present study, Inria Aerial Image Labeling Dataset (Maggiori, et al., 2017) which is publicly available data set was used for comparing the performance of different deep learning models. Images in the dataset have 0.3 m spatial resolution and three spectral bands (red, green, and blue). Although the data set contains images from various cities, only the Chicago images were used in the study presented. These images consist of 36 tiles and cover an area of 81 km<sup>2</sup> in total. Each tile covers an area of 2.25 km<sup>2</sup> and tile sizes are 5000 x 5000.

The ground truth data are available for each image. Ground truth data is prepared in binary format and it shows the building and its building classes. A sample image in the dataset and its ground truth data is given in Figure 1. Labeled images cover dense urban areas and contain a variety of man-made or natural objects. Therefore, it is a suitable data set for testing algorithms.

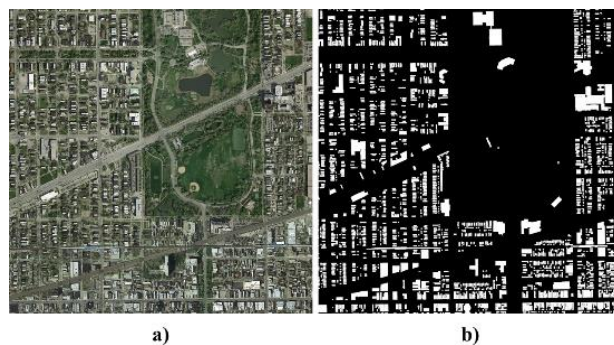


Figure 1. a) Sample image in the dataset. b) ground truth of the sample image.

### Method

Original U-Net model is a fully convolution network developed for biomedical image segmentation (Ronneberger, et al., 2016). The architecture consists of two parts: (1) contracting path, and (2) expansive path. The contracting path, which is similar to a standard CNN, has layers to extract low- and high-level features. In this study, various architectures are used for the contracting path. In the expansive path, the upsampling layers are used for pixel-based classification. The feature maps extracted in the encoder section are copied to the scale to which they belong in the upper scaling process, which is called concatenate. In the last layer, the probability of the class (building and nonbuilding) to which each pixel belongs is calculated using the sigmoid function. Illustration of the U-Net network structure is given in Figure 2.

In this study contracting path of the U-Net model was modified with different feature extraction deep learning algorithms. VGG16, InceptionResNetV2 and DenseNet121 models were used for contracting path. Finally, a new method is proposed by combining the results obtained from the U-Net VGG16, U-Net InceptionResNetV2, and U-Net DenseNet121 models with the majority voting principle.



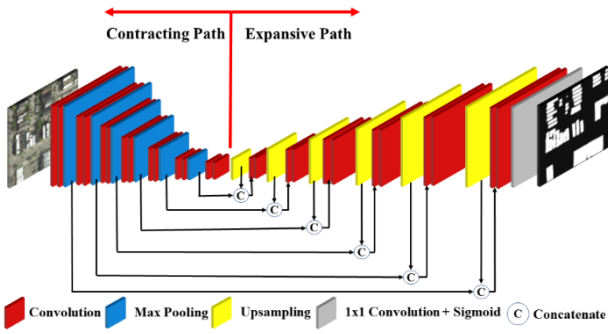


Figure 2. Illustration of U-Net network structure.

**U-Net VGG16**

U-Net VGG16 is a combination of VGG16 (Simonyan and Zisserman., 2014) and U-Net architectures. In the contracting path of the U-Net architecture, other layers of the VGG16 architecture are used except for the fully connected layer. After that, VGG-16 layers are concatenated with a basic U-Net model which is composed of convolutional and upsampling layers. The layers of VGG16 architecture used in the study are given in Table 1.

Table 1: Contracting path of U-Net VGG16 (Simonyan and Zisserman., 2014).

Layer	
<b>Convolution</b>	3x3, 64
<b>Convolution</b>	3x3, 64
<b>Pooling</b>	2 x2 Max Pooling, Stride 2
<b>Convolution</b>	3x3, 64
<b>Convolution</b>	3x3, 64
<b>Pooling</b>	2 x2 Max Pooling, Stride 2
<b>Convolution</b>	3x3, 256
<b>Convolution</b>	3x3, 256
<b>Convolution</b>	3x3, 256
<b>Pooling</b>	2 x2 Max Pooling, Stride 2
<b>Convolution</b>	3x3, 512
<b>Convolution</b>	3x3, 512
<b>Convolution</b>	3x3, 512
<b>Pooling</b>	2 x2 Max Pooling, Stride 2
<b>Convolution</b>	3x3, 512
<b>Convolution</b>	3x3, 512
<b>Convolution</b>	3x3, 512
<b>Pooling</b>	2 x2 Max Pooling, Stride 2

**U-Net InceptionResNetV2**

U-Net Inception-Resnet-v2 is a combination of Inception-Resnet-v2 and U-Net architectures. Inception-Resnet-v2 is developed based on a fusion of the Inception structure and the Residual connection (Szegedy, et al., 2017). In the Inception-Resnet block, multiple sized convolutions are merged by using residual connections. With the residual connections, the degradation problem was avoided, and the training time was reduced. In U-Net Inception-Resnet-v2 architecture, Inception-Resnet-v2 architecture was used as a contracting path of U-Net. The layers of the Inception-

Inception-Resnet-v2 architecture was used as a contracting path of U-Net. The layers of the Inception-

Table 2. Contracting path of Inception-Resnet-v2 (Szegedy, et al., 2017).

Layer
Stem block
5 x Inception-Resnet-A
Reduction-A
10 x Inception-Resnet-B
Reduction-B
5 x Inception-Resnet-C

**U-Net DenseNet121**

U-Net DenseNet121 is a combination of DenseNet121 and U-Net architectures. Block structures are used in DenseNet architectures. Within each block, there are convolution layers. Dense shortcut connections between these layers are used. Each layer in the block is connected to the previous layers. These shortcut connections are provided by transferring feature maps (Huang, et al., 2017). In the contracting path of the U-Net architecture, other layers of the DenseNet121 architecture are used, except for the fully connected layer. The layers of DenseNet121 architecture used in the study are given in Table 3.

Table 3. Contracting path of U-Net DenseNet121 (Huang, et al., 2017).

Layer	
<b>Convolution</b>	7 × 7, Stride 2
<b>Pooling</b>	3 × 3 Max Pooling, Stride 2
<b>Dense Block</b>	$\begin{bmatrix} 1 \times 1 \\ 3 \times 3 \end{bmatrix} \times 6$
<b>Transition Layer</b>	1 × 1 Convolution
	2 × 2 Average Pooling, Stride 2
<b>Dense Block</b>	$\begin{bmatrix} 1 \times 1 \\ 3 \times 3 \end{bmatrix} \times 12$
<b>Transition Layer</b>	1 × 1 Convolution
	2 × 2 Average Pooling, Stride 2
<b>Dense Block</b>	$\begin{bmatrix} 1 \times 1 \\ 3 \times 3 \end{bmatrix} \times 24$
<b>Transition Layer</b>	1 × 1 Convolution
	2 × 2 Average Pooling, Stride 2
<b>Dense Block</b>	$\begin{bmatrix} 1 \times 1 \\ 3 \times 3 \end{bmatrix} \times 16$

**Majority Voting Method**

The majority voting method was used to increase the accuracy of U-Net models. In this method, the results obtained by U-Net methods for each pixel were voted. In this way, the class which includes the majority of votes for the relevant pixel is selected. For example; if a pixel is classified as a building class by two models and classified as a non-building class by one model, then the building class is assigned to that pixel. The overall flow of the majority voting method for building extraction is given in Figure 3.

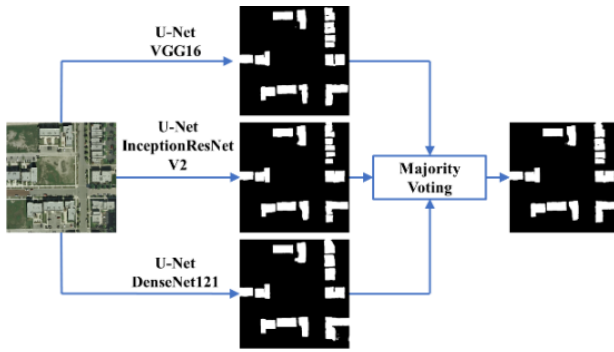


Figure 3. The overall flow of the majority voting method for building extraction.

**Accuracy Assessment**

In the present study, F1 scores of building class calculated used for the accuracy assessment of the architectures. The F1 score method was implemented pixel-based. F1 score is calculated as in formula (1):

$$F1 = 2 \times (precision \times recall) / (precision + recall) \quad (1)$$

where,

$$precision = t_p / (t_p + f_p), recall = t_p / (t_p + f_n) \quad (2)$$

where  $t_p$ ,  $f_p$ , and  $f_n$  are true positive, false positive, and false negative, respectively. Precision indicates what percentage of the pixels detected as buildings are actually buildings. On the other hand, recall; indicates the correct detection rate of the building class.

**Results**

The dimensions of the images are 5000 x 5000. These dimensions are quite high because deep learning architectures require a lot of processing power. Therefore, images and labels were cut, and 512 x 512 images were obtained. As a result of this process, a total of 2715 image subsets were obtained. 80% of the image subsets (2172 image subset) were used as training and 20% (543 image subset) as testing. Examples of images used for training and testing purposes are given in Figure 4.

In this study, deep learning models were created using an open-source deep learning framework called Keras (Chollet, 2020). In the training phase, the Adadelta algorithm (Zeiler, 2012) was used as the optimizer with the learning rate of 1. Binary Crossentropy function was used for loss function. Minibatch size and number of the epoch were chosen as 4 and 50, respectively. All models trained on the Google Colaboratory platform which uses NVIDIA Tesla K80 GPU computing processors.

After the training process, trained deep learning models were applied to test images. So, class probabilities for each pixel of test images were obtained. Then, binary images were obtained by applying the 0.5 threshold value to the class probabilities. Precision, recall, and F1 scores of building class for all methods are given in Table 4.

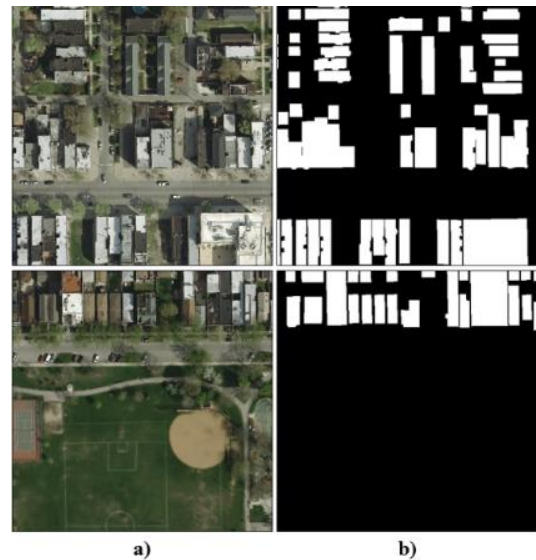


Figure 4. Examples of image subsets used for training and testing purposes. a) Sample image subsets b) ground truths of sample image subsets.

Table 4. Accuracy assessment results

Model	Precision (%)	Recall (%)	F1 Score (%)
U-Net VGG16	87.65	85.60	86.61
U-Net Inception-Resnet-v2	88.10	84.07	86.04
U-Net DenseNet121	84.27	86.97	85.60
Majority Voting	88.40	86.99	87.69

When the F1 Score values are analyzed, it is observed that the method with the highest value is the majority voting. Following the majority voting method, the model with the highest F1 Score was U-Net VGG16 with 86.61. The model with the least F1 Score was the U-Net DenseNet121.

As in the F1 Score, the majority voting method ranked first for both the precision and the recall metrics. It was concluded that the precision value of the U-Net Inception-Resnet-v2 model is high, but the recall value is low. On the other hand, in the U-Net-DenseNet121 model, it is concluded that the precision value is low, but the recall value is high. In the U-Net VGG16 model, there is a balance between precision and recall values. Examples of binary images obtained as a result of applying the models to the test data set are given in Figure 5 and Figure 6. U-Net VGG16 and majority voting methods are the methods that give the closest results to the label image. U-Net DenseNet121 and U-Net Inception-Resnet-v2 methods produced as a result of the binary images of the building's geometric structure could not be fully reflected.

All models can extract discrete building but, if the distance between the buildings is less than 5 pixels, it is found that the models make a merging error. It can be seen in the blue circle in Figure 6. For discrete buildings, U-Net VGG16 and majority voting methods are the methods that give the closest results to the label image. U-Net DenseNet121 and U-Net Inception-Resnet-v2

methods produced as a result of the binary images of the building's geometric structure could not be fully reflected.

**Discussion and Conclusion**

In this study, the effectiveness of various deep-learning-based U-Net models (U-Net VGG16, U-Net Inception-Resnet-v2, U-Net DenseNet121) in building extraction from high-resolution aerial imagery was investigated and a majority voting method was proposed. Results were analyzed using precision, recall, and F1 score metrics. When the results were examined, it was found that the most successful method was the majority voting with 87.69% F1 score. The majority voting method combines

prediction results from U-Net VGG16, U-Net Inception-Resnet-v2, and U-Net DenseNet121 models. Each of these models can have several advantages over each other. For example, the U-Net VGG16 model is a shallower model than other models and it has a basic network structure. On the other hand, dense connections and inception structures are used in the U-Net DenseNet121 and U-Net Inception-ResNet-v2 models, respectively. Thus, a wide variety of results can be obtained from various models. With the Majority voting method, the advantages of the models used are combined in a single method. Therefore, better prediction results can be obtained using this methodology compared to a single deep learning model.

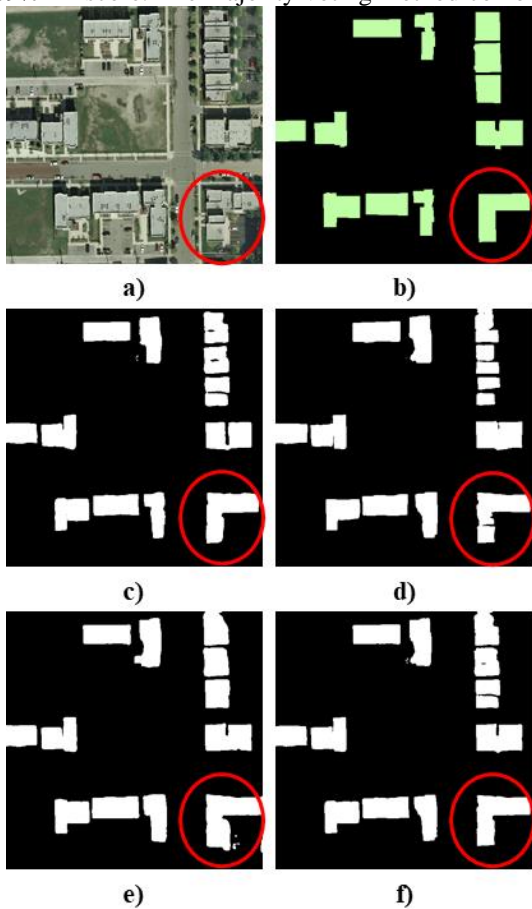


Figure 5. Binary classification results. a) Test image b) Ground truth c) U-Net VGG16 d) U-Net Inception-Resnet-v2 e) U-Net DenseNet121 f) Majority voting

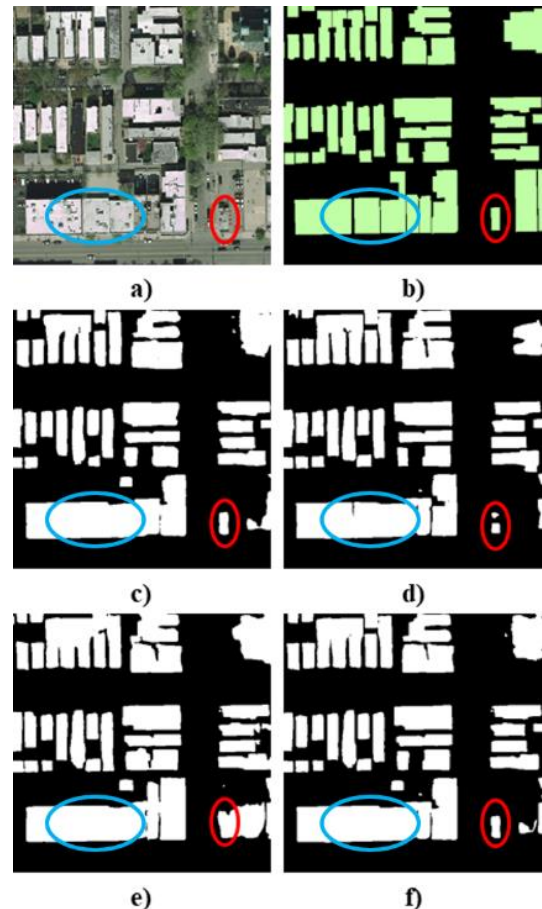


Figure 6. Binary classification results. a) Test image b) Ground truth c) U-Net VGG16 d) U-Net Inception-Resnet-v2 e) U-Net DenseNet121 f) Majority voting

Following the majority voting method, the model with the highest F1 Score was U-Net VGG16 with 86.61%. U-Net VGG16 model is very similar to original U-Net architecture which is a successful architecture when using a small number of data (Ronneberger et al., 2015). The models with the least F1 Score were the Inception-Resnet-v2 and U-Net DenseNet121. These models are deeper models than the VGG16 model. Therefore, more features are extracted in these models. However, considering the properties of the data set and the number of images in the training data set, more efficient results can be obtained with shallow models like U-Net VGG16.

In addition to the F1 score metric, the resulting binary images were analyzed visually. As a result, Majority Voting and U-Net VGG16 methods were the best methods that reflect the building geometry. Furthermore, if the distance between the buildings is less than 5 pixels, it is found that the models make a merging error.

The fact that the images have only RGB bands and the lack of Digital Elevation Model (DEM) makes the problem more difficult. In the next stage of this study, other deep learning models and different data sets which have DEM data are planned to be included in the study.

## References

- Algancı, U., Sertel, E., Kaya, Ş. (2018). Determination of the Olive Trees with Object Based Classification of Pleiades Satellite Image. *International Journal of Environment and Geoinformatics*, 5(2), 132-139. DOI: 10.30897/ijegeo.396713.
- Bittner, K., Adam, F., Cui, S., Körner, M., and Reinartz, P. (2018). Building footprint extraction from VHR remote sensing images combined with normalized DSMs using fused fully convolutional networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(8), 2615-2629.
- Boonpook, W., Tan, Y., Ye, Y., Torteeka, P., Torsri, K. and Dong, S. (2018). A Deep Learning Approach on Building Detection from Unmanned Aerial Vehicle-Based Images in Riverbank Monitoring. *Sensors*, 18(11), 3921.
- Çelik, O., Gazioğlu, C. (2020). Coastline Difference Measurement (CDM) Method. *International Journal of Environment and Geoinformatics*, 7(1), 1-5. DOI: 10.30897/ijegeo.706792.
- Chen, X., Xiang, S., Liu, C. L., Pan, C. H. (2014). Vehicle detection in satellite images by hybrid deep convolutional neural networks. *IEEE Geosci. Remote Sens. Lett.*, 11(10), 1797-1801.
- Cheng, G., Han, J. (2016). A survey on object detection in optical remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 117, 11-28.
- Chollet, F. (2020). Keras. Retrieved 27 April 2020 from <https://github.com/fchollet/keras/>
- Dervisoglu, A, Bilgilioğlu, B., Yağmur, N. (2020). Comparison of Pixel-Based and Object-Based Classification Methods in Determination of Wetland Coastline. *International Journal of Environment and Geoinformatics*, 7(2), 213-220. DOI: 10.30897/ijegeo.713307.
- Esetlili, M., Bektas Balcik, F., Balik Sanli, F., Kalkan, K., Ustuner, M., Goksel, Ç., Gazioğlu, C., Kurucu, Y. (2018). Comparison of Object and Pixel-Based Classifications for Mapping Crops Using Rapideye Imagery: A Case Study of Menemen Plain, Turkey. *International Journal of Environment and Geoinformatics*, 5(2), 231-243. DOI: 10.30897/ijegeo.442002.
- Ghanea, M., Moallem, P., Momeni, M. (2016). Building extraction from high-resolution satellite images in urban areas: recent methods and strategies against significant challenges. *International journal of remote sensing*, 37(21), 5234-5248.
- Hu, W., Huang, Y., Wei, L., Zhang, F., Li, H. (2015). Deep convolutional neural networks for hyperspectral image classification, *Journal of Sensors*, 2015, Article ID 258619, 1-12.
- Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4700-4708.
- Huang, W., Xiao, L., Wei, Z., Liu, H., Tang, S. (2015). A new pan sharpening method with deep neural networks. *IEEE Geosci. Remote Sens. Lett.*, 12(5), 1037-1041.
- Hui, J., Du, M., Ye, X., Qin, Q., and Sui, J. (2018). Effective building extraction from high-resolution remote sensing images with multitask driven deep neural network. *IEEE Geoscience and Remote Sensing Letters*, 16(5), 786-790.
- Ji, S., Wei, S., and Lu, M. (2018). Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. *IEEE Transactions on Geoscience and Remote Sensing*, 57(1), 574-586.
- Ji, S., Wei, S., and Lu, M. (2019). A scale robust convolutional neural network for automatic building extraction from aerial and satellite imagery. *International journal of remote sensing*, 40(9), 3308-3322.
- Li, L., Liang, J., Weng, M., and Zhu, H. (2018). A multiple-feature reuse network to extract buildings from remote sensing imagery. *Remote Sensing*, 10(9), 1350.
- Li, X., Yao, X., Fang, Y. (2018). Building-A-Nets: Robust Building Extraction From High-Resolution Remote Sensing Images With Adversarial Networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 99, 1-8.
- Lin, J., Jing, W., Song, H., Chen, G. (2019). ESFNet: Efficient Network for Building Extraction From High-Resolution Aerial Images. *IEEE Access*, 7, 54285-54294.
- Liu, H., Luo, J., Huang, B., Hu, X., Sun, Y., Yang, Y., ... and Zhou, N. (2019). DE-Net: Deep Encoding Network for Building Extraction from High-Resolution Remote Sensing Imagery. *Remote Sensing*, 11(20), 2380.
- Liu, P., Liu, X., Liu, M., Shi, Q., Yang, J., Xu, X., and Zhang, Y. (2019). Building footprint extraction from high-resolution images via spatial residual inception convolutional neural network. *Remote Sensing*, 11(7), 830.
- Liu, Y., Gross, L., Li, Z., Li, X., Fan, X., and Qi, W. (2019). Automatic building extraction on high-resolution remote sensing imagery using deep convolutional encoder-decoder with spatial pyramid pooling. *IEEE Access*, 7, 128774-128786.
- Long, J., Shelhamer, E., Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 3431-3440.
- Lu, T., Ming, D., Lin, X., Hong, Z., Bai, X., and Fang, J. (2018). Detecting building edges from high spatial resolution remote sensing imagery using richer convolution features network. *Remote Sensing*, 10(9), 1496.
- 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)*, 3226-3229.
- Pan, X., Yang, F., Gao, L., Chen, Z., Zhang, B., Fan, H., and Ren, J. (2019). Building extraction from high-resolution aerial imagery using a generative adversarial network with spatial and channel attention mechanisms. *Remote Sensing*, 11(8), 917.



- Ronneberger, O., Fischer, P., Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, 234-241.
- Schuegraf, P., and Bittner, K. (2019). Automatic Building Footprint Extraction from Multi-Resolution Remote Sensing Images Using a Hybrid FCN. *ISPRS International Journal of Geo-Information*, 8(4), 191.
- Simonyan, K., Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition, *arXiv preprint*, arXiv:1409.1556.
- Sun, G., Huang, H., Zhang, A., Li, F., Zhao, H., and Fu, H. (2019). Fusion of multiscale convolutional neural networks for building extraction in very high-resolution images. *Remote Sensing*, 11(3), 227.
- Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A. A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-First AAAI Conference on Artificial Intelligence*. 4278-4284.
- Xu, Y., Wu, L., Xie, Z., and Chen, Z. (2018). Building extraction in very high resolution remote sensing imagery using deep learning and guided filters. *Remote Sensing*, 10(1), 144.
- Yang, H., Wu, P., Yao, X., Wu, Y., Wang, B., Xu, Y. (2018). Building extraction in very high resolution imagery by dense-attention networks. *Remote Sensing*, 10(11), 1768.
- Yi, Y., Zhang, Z., Zhang, W., Zhang, C., Li, W., and Zhao, T. (2019). Semantic Segmentation of urban buildings from vhr remote sensing imagery using a deep convolutional neural network. *Remote Sensing*, 11(15), 1774.
- Zeiler, M. D. (2012). Adadelta: an adaptive learning rate method. *arXiv preprint*, arXiv:1212.5701.
- Zhang, F., Du, B., Zhang, L. (2016). Scene classification via a gradient boosting random convolutional network framework. *IEEE Trans. Geosci. Remote Sens.*, 54(3), 1793–1802.
- Zhang, L., Wu, J., Fan, Y., Gao, H., and Shao, Y. (2020). An Efficient Building Extraction Method from High Spatial Resolution Remote Sensing Images Based on Improved Mask R-CNN. *Sensors*, 20(5), 1465.