

Performance Evaluation of Jaccard-Dice Coefficient on Building Segmentation from High Resolution Satellite Images

İsa Ataş

Abstract— In remote sensing applications, segmentation of input satellite images according to semantic information and estimating the semantic category of each pixel from a given set of tags have of great importance for the automatic tracking task. It is important in situations such as building detection from high resolution satellite images, city planning, environmental preparation, disaster management. Buildings in metropolitan areas are crowded and messy, so high-resolution images from satellites need to be automated to detect buildings. Segmentation of remote sensing images with deep learning technology has been a widely considered area of research. The Fully Convolutional Network (FCN) model, a popular segmentation model, is used for building detection based on pixel-level satellite images. With the U-Net model, which was developed for biomedical image segmentation and modified in our study, performance analysis was performed for building segmentation from satellite images using customized loss functions such as Dice Coefficient and Jaccard Index measurements. Dice Coefficient score was obtained 84% and Jaccard Index score was obtained 70%. In addition, the Dice Coefficient score increased from 84% to 87% by using the Batch Normalization (BN) method instead of the Dropout method in the model.

Index Terms—Convolutional neural network, remote sensing imagery, semantic segmentation, U-Net.

I. INTRODUCTION

Buildings are identified as an important object class in remote sensing images. Identifying the spatial location and shape of buildings in these images takes an important role in assessment issues such as geographic planning, mapping, and post-disaster reconstruction. In particular, determining the number of buildings remaining in natural disasters such as earthquake events and subtract the degree of damage are

important indicators. Traditional remote sensing image classification methods use spectral, and spatial correlation information of the desired object to determine the class content of ground objects. Multi-scale feature aggregation etc. it is aimed to extract the spectral and shape properties of buildings by using techniques such as [1]. With the rapid development of deep learning recently, it has made great progress in the fields of agriculture [2-3], health [4], robotics and security [5], natural language processing [6], as well as in the fields of remote sensing, semantic segmentation and target recognition [1]. Deep learning methods are in the direction of producing more stable results for image recognition in remote sensing with high accuracy and high computational speeds [1]. With the development of remote sensing technology, many researchers were tried to develop, try and explore new methods to increase the accuracy and the speed of automatic classification algorithms from satellite images. In image processing applications, image segmentation and labeling are a very demanding task. Convolutional Neural Network (CNN) is a more effective approach than traditional image segmentation methods. CNN automatically learns multi-level representations that map the original input to specified binary or multiple labels or sequential vectors. CNN's powerful "representational learning" capability greatly simplifies the feature design upon building detection and has shown promising results [7]. U-Net is structurally one of the simplified CNN architectures used for image segmentation. With the semantic segmentation method, studies are carried out on many complex tasks. Autonomous vehicles [8], biomedical image diagnostics [9], geo-sensing [10], and precision agriculture [11] are a few of them.

In this study, building extraction was performed from high resolution images with Semantic Segmentation using FCN. The contribution of study to previous research is to investigate which resolution might be more useful by comparing object extraction and segmentation approaches from different resolution satellite images. In addition, the effect of the dropout layer and batch normalization layers on the U-Net architecture was analyzed. F-measure (Dice Coefficient) and Intersection Unit (IoU) / (Jaccard Coefficient) metrics were used to distinguish the building class from satellite images and to quantitatively evaluate the performance of the segmentation method. The remainder of the article was organized as follows. In the second part, references were made to previous studies in the literature. In the third chapter, the materials and methods used were explained. In the fourth section, the

İSA ATAŞ, is with Computer Technologies Department, Diyarbakır Vocational School of Technical Sciences, Dicle University, Diyarbakır, Turkey, (e-mail: isa_atas@dicle.edu.tr).

 <https://orcid.org/0000-0003-4094-9598>

Manuscript received December 1, 2022; accepted Jan 16, 2023
DOI: [10.17694/bajece.1212562](https://doi.org/10.17694/bajece.1212562)

analyzes and result graphics were shown. The existing study are summarized in the conclusion.

II. RELATED WORK

Wu et al. Firstly, it skillfully combined the spectral features and shape features of buildings to distinguish buildings from remote sensing images using the traditional remote sensing image feature extraction method [12]. Huang et al. used the multi-feature fusion method to extract features of buildings in remote sensing images and finally chose to use the SVM classifier to classify ground objects [13]. Xin et al. used the spatial position relationship between shadows and buildings in densely built areas and used the graph cutting algorithm to accurately draw the outlines of buildings [14]. Acar et al. analyzed Radarsat-2 satellite images to retrieve soil moisture using semi-empirical models and machine learning methods. By extracting the feature vector from preprocessed SAR image pixels, they stated that machine learning techniques outperform semi-empirical models [15]. Xu-dong et al. it has achieved good results using the pixel-based classification technique to process the image using the object-oriented method [16]. After machine learning techniques became widespread, Deep CNN (DCNN) started to be used in current image classification and extraction. Wu et al. performed end-to-end building segmentation from aerial images using multi-constraint FCN architecture [17]. Yuan, a simple FCN model architecture was proposed that combines several activation layers in pixel level estimation. In addition, the marked distance function of building boundaries with enhanced representativeness was introduced to present the outputs [18]. Chen et al. applied various cutting-edge deep U-Net frameworks such as pyramid scene decomposition network, feature pyramid network (FPN) and multi-scale feature fusion FPN to generate roof detection from large-scale comparative aerial images [19].

III. MATERIALS AND METHODS

A. Dataset

A dataset consisting of .tif files containing real aerial photographs collected from the satellite and masked images were used. The input images of the proposed model are in RGB format. The dataset consists of a total of 408 images and tag masks with a resolution of 512×512 (ranging from 0.3 to 2.5 m) received from the satellite, prepared by the Wuhan University Photogrammetry and Computer Vision (GPCV) Group [20].

B. Preprocessing

Images with few buildings, no buildings and poor resolution were removed from the dataset during the process. After the elimination, only a total of 360 images could be obtained. Since it would be insufficient to train and test this number of image models, data augmentation methods were applied to the images. To increase sample images, images and mask images were augmented using the data augmentation method rotation (90 degrees right) and (90 degrees left). With this technique, 2 additional variations of each original image were created. As a

result of this process, the dataset was tripled and 1080 images and related masked images were prepared. We resized the image and tag masks at 256×256 , 128×128 , and 64×64 pixels along with the existing 512×512 -pixel dataset to research on, train and test the optimal patch size on the FCN network. By comparing the performance of images and masks in four different pixels in the U-Net model, we observed that the 256×256 size was more suitable than the other sizes. From these prepared images, 972 images and corresponding labeled images ($\approx 80\%$) were randomly selected to be used as training set, and the remaining 108 images and corresponding labeled images ($\approx 20\%$) were used as test dataset.

Here, by applying random cropping from the image data, we cropped large images into small patches in the desired pixels and obtained more data for our model. Similarly, we applied the same for the masked dataset. In our preprocessing, we visualized both cropped patches (image and mask) and checked whether the newly formed image and tagged masks were aligned correctly. Since the tags we extracted from the satellite images consisted only of the "buildings" layer, we converted the tag images into binary masks of four different sizes. In the labels, we pixelated the buildings as 1 and the outside of the building as 0. In Fig. 1 shows the sample satellite image from the dataset, as well as the 256×256 patch image and mask that we referenced in our study and formed after cropping.

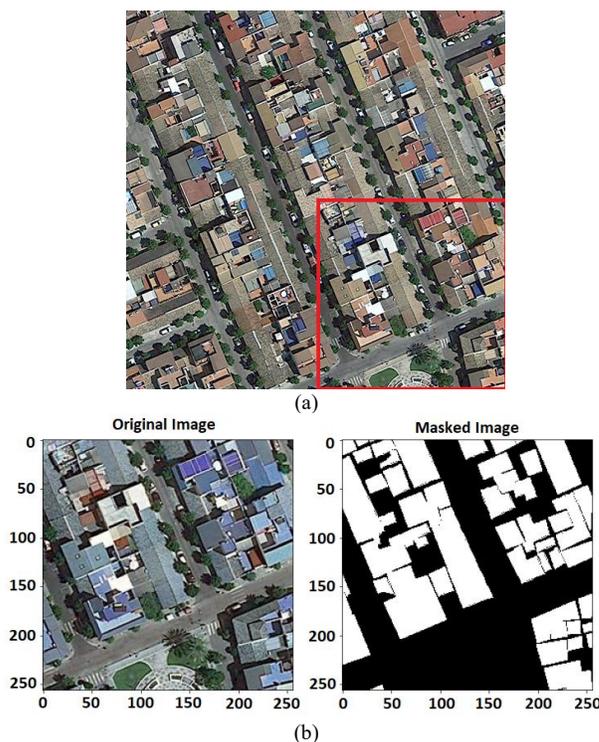


Fig.1. (a) Satellite image (b) Cropped Satellite image and corresponding mask with buildings identified in white.

C. Architecture of the model (U-Net)

The U-Net architecture was developed by Olaf Ronneberger et al. [21] for Biomedical Image Segmentation. It aims at image segmentation using a small training dataset. The general

structure of U-Net, which is a typical encoder-decoder architecture and modified in this study, is shown in Fig. 2.

U-Net Architecture includes two paths, convolution (Contraction path) and deconvolution (Expanding path) layers. The first way is the narrowing path, also called an encoder, which captures the input image at multiple different levels of feature representations. The encoder is just a traditional stack of convolution and max pooling layers. The second path is the symmetric expansion path, also called a decoder, which is to semantically project the distinctive features (lower resolution) into the pixel space (higher resolution). Basically, U-Net is built on FCN [22]. For this reason, it is also referred to as end-to-end FCN. It contains only Convolutional layers and no dense layers because it can accept images of any size. Feature maps created by convolution layers are clipped to input size. Clipping deals with the loss of edge pixels in the convolution process and subtracts the size of the convolution results, consistent with the deconvolution results. In the U-Net architecture, jump links (copy and crop) are added between the narrowing path and the widening path, allowing precise localization as well as features. Clipped feature maps are combined with deconvolution results via shortcut links. The network applies 1×1 convolution to the feature map to label the pixels and produces the segmentation result.

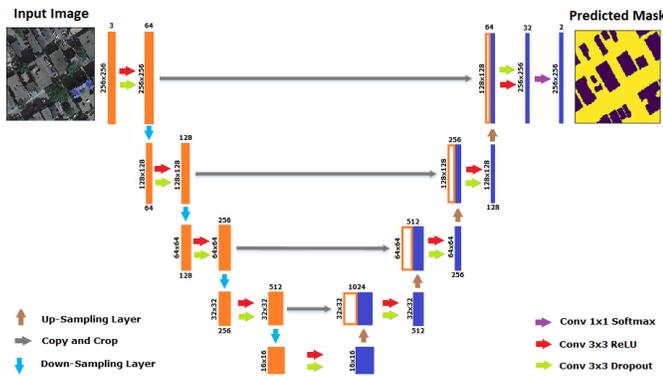


Fig. 2. Architecture of U-Net

In the U-Net architecture model, the orange boxes represent cross sections of the feature maps. The dimensions of each feature map are indicated on the left side, and the number of channels is indicated above. White boxes represent half of the channels with feature maps from which they were copied. Arrows represent actions indicated by the indicator; gray arrows represent copying (skipping links). The purpose of the encoder layer in the model is to extract feature information from the images until the model is at the bottom. The purpose of the decoder layer is to find the required area that the model should represent from the fragmentary patterns.

In our study, we modified the U-Net architecture for the building segmentation problem. We trained the model from scratch without using pre-trained weights with biomedical images. First, we replaced the stochastic gradient descent algorithm (SGD) [23] with the Adam [24] optimizing algorithm, which is known to converge faster during training. Next, we changed the dimensions of the input images to 256×256 , since the U-net architecture was designed for images with a size of 572×572 . In our experimentally applied model, the

optimum learning rate was chosen as 0.0001. To reduce the computational cost, we chose to simplify the model by changing the filter numbers of the convolutional layers to 16, 32, 64, and 128 from 64, 128, 256, and 512 in the U-Net architecture. In end-to-end training, we added “dropout” to avoid overfitting when training the model in the decoder section after rectified linear unit (ReLU) activation in all convolution layers.

Due to the limited amount of GPU memory, we set the training batch size to 16 and maximum number of periods according to Epoch 100 in this study. We got the output image from the model by adding CNN layer with SoftMax instead of sigmoid as activation at the end of the expanding path layer. As a result, our app is based on Keras [25] using a TensorFlow backend. All training and testing processes were performed in Google Colab environment [26] with 13,342 RAM - Tesla K80 GPU - NVIDIA T4 GPUs Card.

D. Metrics

We used precision, recall, F-measure (Dice Coefficient) and Intersection of Unity (IoU) / (Jaccard Coefficient) metrics to distinguish building class from satellite images and to quantitatively evaluate the performance of the segmentation method. Segmentation based on Dice and Jaccard coefficients is widely used for image segmentation as it allows to deal with class imbalance [27]. Precision and recall are defined using true positives (TP), false positives (FP), and false negatives (FN).

1) Precision

The precision refers to the ratio of positive data identified as true to all data identified as true. It is calculated by following equation [28].

$$\text{Precision} = TP / (TP + FP) \quad (1)$$

2) Recall

The recall refers to the ratio of positive data identified as true to the sum of positive data identified as true and negative data identified as false. It is calculated by following equation [28].

$$\text{Recall} = TP / (TP + FN) \quad (2)$$

3) Jaccard Index

The first of the metrics used to evaluate the education score in our study is the Jaccard Index, also known as IOU (Intersection over Union). The Jaccard index, also known as the Jaccard similarity coefficient, is a statistic used to measure similarity. It was explanation by Grove Karl Gilbert in 1884 as the verification rate [29]. It was later developed independently by Paul Jaccard [30]. The Jaccard coefficient measures similarity between finite sets of samples and is defined as the size of the intersection divided by the size of the union of sample sets. It is calculated by following equation.

$$0 \leq J(A, B) \leq 1$$

$$J(A,B)=|A\cap B|/(|A|+|B|-|A\cap B|)$$

$$J(A,B) = (\text{precision} \times \text{recall}) / (\text{precision} + \text{recall} - (\text{precision} \times \text{recall})) \quad (3)$$

4) Dice Coefficient

The second of the metrics used to evaluate the education score in our study is the Dice Coefficient, also known as the F1 score. Independently developed by Thorvald-Sørensen [31] and Lee-Raymond Dice [32], the Sørensen-Dice coefficient is a statistic used to measure the similarity of two samples. It is calculated by following equation.

$$0 \leq J(A, B) \leq 1$$

$$D(A,B) = 2 \times (|A \cap B|) / (|A| + |B|)$$

$$D(A,B) = 2 \times (\text{precision} \times \text{recall}) / (\text{precision} + \text{recall}) \quad (4)$$

IV. EXPERIMENTS/RESULTS/DISCUSSION

In order to obtain a successful U-Net training model in the segmentation task, some hyperparameters have been modified. In addition, analyzes were made to obtain loss and metric graph curves. Initially, training and validation analyzes of four different datasets with 64×64 , 128×128 , 256×256 , and 512×512 resolutions were made and the loss graphs are shown in Fig. 3. Considering the training and validation graph curves, convergences were observed differently in four different patch sizes. It is seen in the graphs that as the patch size increases, convergence tends to improve and then deteriorate.

Peak value was obtained with 256×256 pixels size providing the best IoU (Jaccard Coef.) 84.70%. The IoU plots of the datasets according to the four different patch sizes are shown in Fig. 4. Generally, in theory larger patches provide more information and accurate predictions. However, fine resolution images require a longer training time or added feature information may be unnecessary and affect the model training process [33]. However, cropped patches with a small window size may contain information that does not have sufficient properties for the model and may increase the risk of overfitting. Also, small size patches cannot complete all down sampling of the U-Net model [34].

The 64×64 satellite images failed to predict the building class with the modified U-Net model and the model was unable to complete the subsamples. It has followed an unstable path as it has not been able to extract adequate feature maps. It has been observed that the model building images from 128×128 and 512×512 sized satellite images have difficulty in separating from other objects. Therefore, in this study, a patch size of 256×256 pixels was chosen, which is suitable for the U-Net model. Intuitive comparisons of image segmentation results of input satellite images of 4 different sizes in model training are shown in Fig. 5.

The performances of the dice coefficient and jacquard index, which are among the missing metrics, on the model were analyzed.

The comparison of actual image, mask and predicted mask in Dice Coefficient and Jaccard-Index mode is shown in Fig. 6.

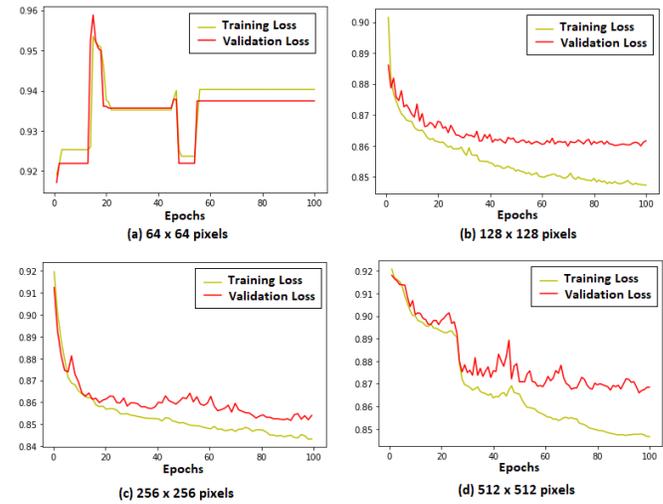


Fig. 3. Loss values of the proposed model over 100 epochs

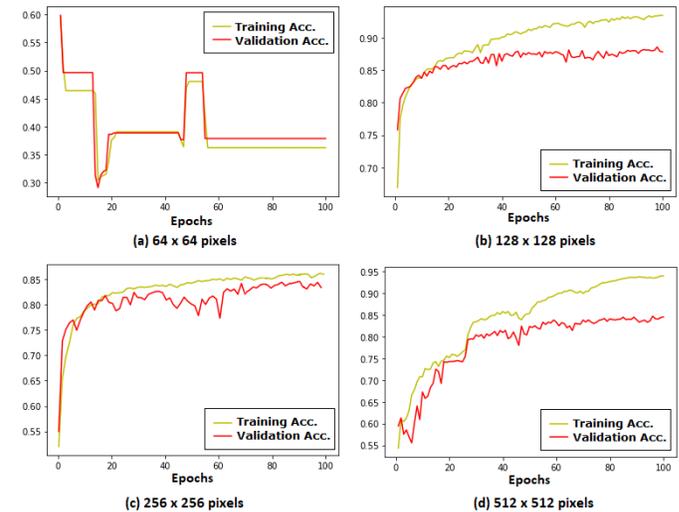


Fig. 4. Accuracy values of the proposed model over 100 epochs

Dice coefficient and Jaccard index loss curves on the training and validation dataset are shown in Fig. 7. The metric curves of the Dice coefficient and the Jaccard index are shown in Fig. 8 comparatively. When we look at the performances of both metrics, it is seen that the Dice coefficient is more consistent than the Jaccard index, and it provides superiority in loss and similarity rates. It also has realized less convergence in the training phase. In terms of improving the performance of the modified model, Dropout and BN methods were tried and analyzed as the threshold value before ReLU activation in the contraction path layer. The performances of the Dropout and BN index, which are among the missing metrics, on the model were analyzed. The comparison of actual image, mask and predicted mask in Dropout and BN mode is shown in Fig. 9. Dropout and BN loss curves on the training and validation dataset are shown in Fig. 10.

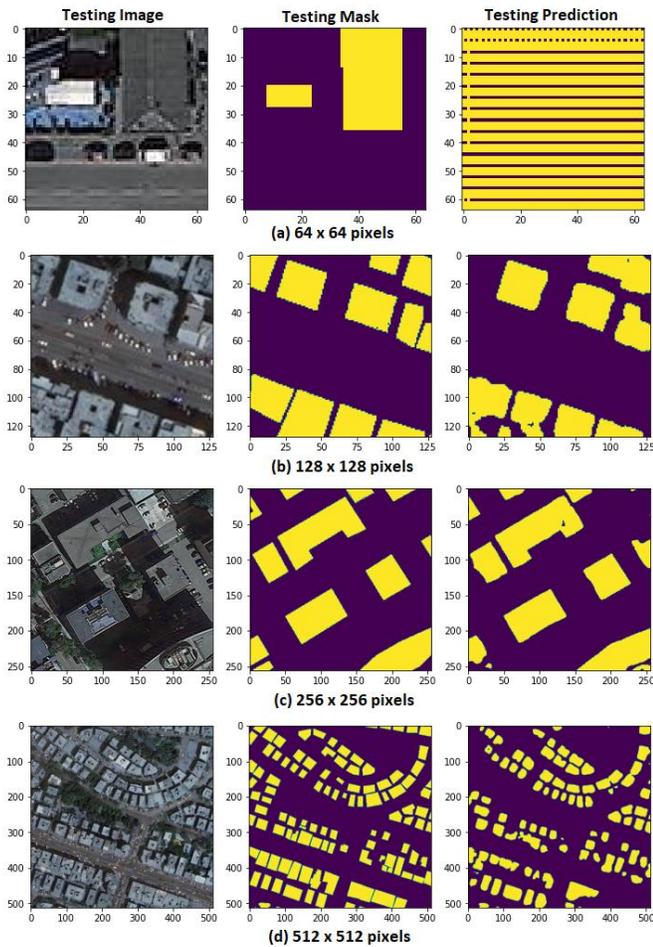


Fig. 5. Satellite images of four different dimensions, ground reality masks and masks predicted

Looking at the loss and metric graphs, it is seen that BN improves model performance and follows a stable path compared to the dropout method.

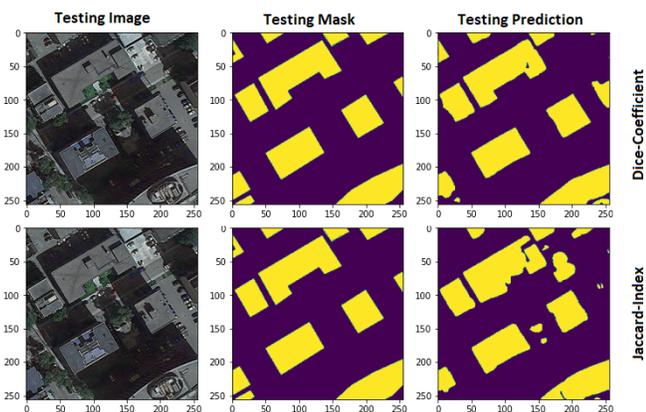


Fig. 6. Comparison of real image, mask and predicted mask in Dice-Coefficient and Jaccard-Index mode

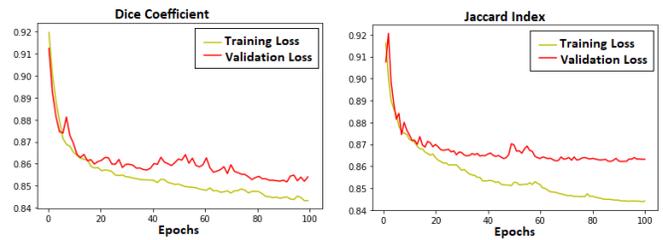


Fig. 7. Training and validation dataset: loss curves of Dice coefficient and Jaccard index

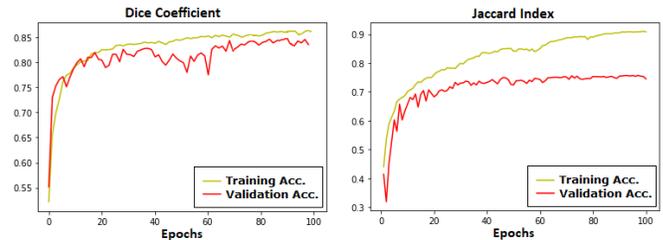


Fig. 8. Training and validation dataset: accuracy curves of Dice coefficient and Jaccard index

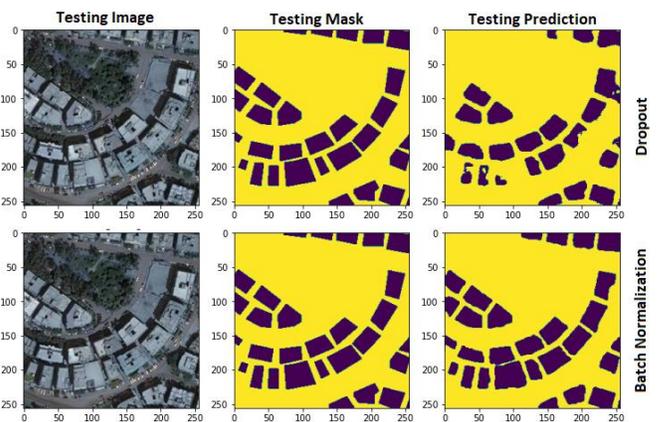


Fig. 9. Comparison of real image, mask and predicted mask in Dropout and Batch Normalization mode

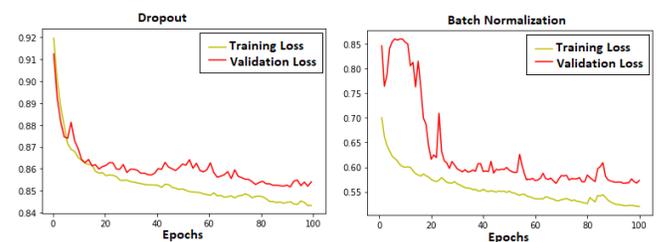


Fig. 10. Training and validation dataset: loss curves of Dropout and Batch Normalization

The metric curves of the Dropout and BN are shown in Fig. 11 comparatively. While the loss rate is 0.84 in the dropout method, it is seen that the loss rate decreases to 0.20 when the BN method is used. In addition, it was observed that the Dice-Coefficient metric value increased from 0.83 to 0.87 when the BN method was used instead of the dropout method.

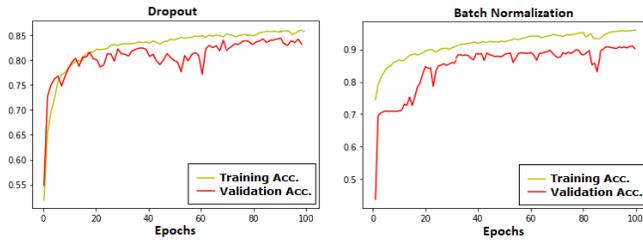


Fig. 11. Training and validation dataset: accuracy curves of Dropout and Batch Normalization

A rational comparison could not be made because experimental studies were conducted on different satellite image datasets and different metrics were used in the literature. However, a summary of the quantitative results based on similar studies is given in Table 1.

TABLE I
QUANTITATIVE COMPARISON OF SATELLITE IMAGE
SEGMENTATION STUDIES

Reference	Dataset	Model	Jaccard %	Dice %
[35]	Massachusetts	GMEDN	70.39	-
[36]	WHU	ESFNet	85.34	-
[37]	Carvana	U-Net	-	68.70
[38]	Inria	SegNet	70.14	-
[39]	MBRSC	U-Net	-	87.00
Our study	GPCV	U-Net	70.27	87.13

V. CONCLUSION

In this study, an end-to-end approach is adopted for binary mask classification using a small number of satellite image data. The prepared dataset was applied and evaluated in the CNN-based U-Net architecture developed by Ronneberger et al. The Keras framework and Python language were used to implement the U-Net model, facilitate data augmentation, and increase the robustness of the training. A patch size of 256×256 pixels was selected as an appropriate window size for the U-Net model in our apps. The model tested for the 64×64 pixel patch showed poor performance and resulted in weak compatibility. Since the test could not make meaningful predictions on the data, it was observed that high loss occurred with low accuracy. On the other hand, in the analysis of 128×128 and 512×512 pixel patches, overfitting occurred with poor performance. F-measure (Dice Coefficient) and Intersection Unit (IoU) / (Jaccard Coefficient) metrics were used to quantitatively evaluate and compare the performance of the proposed method. In comparison, it was seen that the dice coefficient was more consistent than the Jaccard index and was superior in loss and similarity rates. The BN process, in which the biased output distribution obtained from the previous layer is corrected, and the dropout process, in which some of the feature information obtained from the previous layer are randomly disabled are compared to observe the effect on the model. Dice Coefficient score was obtained 84% and Jaccard Index score was obtained 70%. In addition, the Dice Coefficient accuracy score increased from 84% to 87% by using the Batch Normalization (BN) method instead of the Dropout method in the model. As a result of the analysis, it

was determined that BN has a noticeable effect on performance compared to the Dropout layer.

In conclusion, considering the qualitative and quantitative results obtained in the analyzes performed in this study, the modified U-Net model performed well overall with its small dataset size and limited computational capacity.

REFERENCES

- [1] Q. Han, Q. Yin, X. Zheng, Z. Chen, "Remote sensing image building detection method based on Mask R-CNN." *Complex & Intelligent Systems*, 8(3), 1847-1855, 2022.
- [2] M. Ataş, "Fıstık sınıflandırma sistemi için Siirt fıstığı imgelerinden gürbüz özneliklerin çıkarılması." *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi* 7(1):93-102, 2016.
- [3] E. Acar, "Detection of unregistered electric distribution transformers in agricultural fields with the aid of Sentinel-1 SAR images by machine learning approaches." *Computers and Electronics in Agriculture*, 175, 105559, 2020.
- [4] A. D. Yetis, M. I. Yesilnacar, M. Atas, "A machine learning approach to dental fluorosis classification." *Arabian Journal of Geosciences*, 14(2):1-12, 2021.
- [5] M. Atas, Y. Dogan, İ. Atas, "Chess playing robotic arm." In 2014 22nd Signal Processing and Communications Applications Conference (SIU) (pp. 1171-1174). IEEE, 2014.
- [6] C. Özdemir, M. Ataş, A. B. Özer, "Classification of Turkish spam e-mails with artificial immune system." 21st Signal Processing and Communications Applications Conference (SIU). IEEE, 2013.
- [7] S. Ji, S. Wei, M. Lu, "Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery dataset." *IEEE Transactions on Geoscience and Remote Sensing*, 57(1), 574-586, 2018.
- [8] Ç. Kaymak, A. Uçar, "Semantic Image Segmentation for Autonomous Driving Using Fully Convolutional Networks." *International Artificial Intelligence and Data Processing Symposium (IDAP)*, 2019. DOI: 10.1109/IDAP.2019.8875923.
- [9] A. Valizadeh, M. Shariatee, "The Progress of Medical Image Semantic Segmentation Methods for Application in COVID-19 Detection." *Comput Intell Neurosci*. 2021, DOI: 10.1155/2021/7265644.
- [10] A. Mousavian, J. Kosecka, "Semantic Image Based Geolocation Given a Map." DOI: 10.48550/arXiv.1609.00278.
- [11] T. Anand, S. Sinha, M. Mandal, V. Chamola, F. R. Yu, "AgriSegNet: Deep aerial semantic segmentation framework for IoT-assisted precision agriculture." *IEEE Sensors Journal*, 21(16), 17581-17590, 2021.
- [12] W. Wu et al., "Building extraction from high resolution remote sensing imagery based on spatial-spectral method." *Geomat Inf Sci Wuhan Univ* 7:800-805, 2012.
- [13] X. Huang et al., "Classification of high spatial resolution remotely sensed imagery based upon fusion of multiscale features and SVM." *J Remote Sens* 11:48-54, 2007.
- [14] F. Xin, C. Shanxiong, "High-resolution remote sensing image building extraction in dense urban areas." *Bull Surv Mapp*, 2019.
- [15] H. Acar, M. S. Özerdem, E. Acar, "Soil moisture inversion via semiempirical and machine learning methods with full-polarization Radarsat-2 and polarimetric target decomposition data: A comparative study." *IEEE Access*, 8, 197896-197907, 2020.
- [16] W. Xu-dong, G. Jian-ming, J. Bai-jun et al., "Mixed-pixel classification of remote sensing images of cellular automata." *J Surv Mapp* 37(1):42-48, 2008.
- [17] G. Wu, X. Shao, Z. Guo, Q. Chen, W. Yuan, X. Shi, et al. "Automatic building segmentation of aerial imagery using multi-constraint fully convolutional networks", *Remote Sensing*, 10, p. 407, 2018.
- [18] J. Yuan, "Learning building extraction in aerial scenes with convolutional networks." *IEEE Transactions on Pattern Analysis Machine Intelligence*, 40, pp. 2793-2798, 2017.
- [19] Q. Chen, L. Wang, Y. Wu, G. Wu, Z. Guo, S. L. Waslander, "Aerial imagery for roof segmentation: A large-scale dataset towards automatic mapping of buildings." *ISPRS Journal of Photogrammetry and Remote Sensing*, 147, pp. 42-55, 2018.

- [20] <http://study.rsgis.whu.edu.cn/pages/download/>
- [21] O. Ronneberger, P. Fischer, T. Brox, "U-net: Convolutional networks for biomedical image segmentation." International conference on medical image computing and computer-assisted intervention, (pp. 234–241). Springer, 2015.
- [22] J. Long, E. Shelhamer, T. Darrell, "Fully Convolutional Networks for Semantic Segmentation", University of Berkeley, Proceedings of the IEEE, 2015.
- [23] N. Ketkar, "Stochastic gradient descent.", In Deep learning with Python (pp. 113-132). Apress, Berkeley, CA, 2017.
- [24] A. Radford, L. Metz, S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks.", arXiv preprint arXiv:1511.06434, 2015.
- [25] F. Chollet, "Keras: Deep learning library for theano and tensorflow", 2015, [online] Available: <https://github.com/fchollet/keras>.
- [26] Google Colab [Online] Access Link: <https://colab.research.google.com/>, on 21 November 2022.
- [27] G. Chhor, B. A. Cristian, B-L. Ianis, "Satellite image segmentation for building detection using U-Net." Web: <http://cs229.stanford.edu/proj2017/final-reports/5243715.pdf>, 2017.
- [28] İ. Ataş, "Human gender prediction based on deep transfer learning from panoramic dental radiograph images." *Traitement du Signal*, 39(5), 1585-1595, 2022. DOI:10.18280/ts.390515
- [29] A. H. Murphy, "The Finley Affair: A Signal Event in the History of Forecast Verification." *Weather and Forecasting*. 11 (1): 3, 1996.
- [30] Jaccard, Paul, "The Distribution of the Flora in the Alpine Zone.1". *New Phytologist*. 11 (2): 37–50, 1912. DOI:10.1111/j.1469-8137.1912.tb05611.x.
- [31] T. Sørensen, "A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons." *Kongelige Danske Videnskabernes Selskab*. 5 (4): 1–34, 1948.
- [32] L. R. Dice, "Measures of the Amount of Ecologic Association Between Species." *Ecology*. 26 (3): 297–302, 1945. DOI:10.2307/1932409.
- [33] F. Milletari, N. Navab, S. A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation." In Proceedings of the 14th 3D Vision, Stanford, CA, USA, 25–28, pp. 565–571, 2016.
- [34] J. Zhang, et al. "Segmenting purple rapeseed leaves in the field from UAV RGB imagery using deep learning as an auxiliary means for nitrogen stress detection." *Remote Sensing* 12.9, 1403, 2020.
- [35] J. Ma, et al., "Building Extraction of Aerial Images by a Global and Multi-Scale Encoder Decoder Network." *Remote Sens.*, 12, 2350, 2020.
- [36] J. Lin, W. Jing, H. Song, G. Chen, "ESFNet: Efficient Network for Building Extraction from High-Resolution Aerial Images." *IEEE Access*, 7, 54285–54294, 2019.
- [37] V. Iglovikov, A. Shvets, "Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation." arXiv preprint arXiv:1801.05746, 2018.
- [38] G. Chhor, C. B. Aramburu, I. Bougdal-Lambert, "Satellite image segmentation for building detection using U-Net." Web: <http://cs229.stanford.edu/proj2017/final-reports/5243715>, 2017.
- [39] D. Patil, K. Patil, R. Nale, S. Chaudhari, "Semantic Segmentation of Satellite Images using Modified U-Net," *IEEE 10. Regional Symposium (TENSYPMP)*, s.1-6, 2022.

BIOGRAPHIES

İSA ATAŞ has been an instructor with the Diyarbakır Vocational School of Technical Sciences, Dicle University. The research area is deep learning, computer vision, machine learning, and wireless networks.