6th Intercontinental Geoinformation Days

igd.mersin.edu.tr

Comparative analysis of semantic segmentation of terrestrial images using DeepLabv3+

Ahmet Verani[*1], **Muhammed Enes Atik** [1], **Zaide Duran** [1]

[1]*Istanbul Technical University, Faculty of Civil Engineering, Department of Geomatics Engineering, Istanbul, Türkiye*

Keywords
Deep Learning
Semantic Segmentation
Terrestrial
DeepLabv3+

Abstract

Feature extraction from images by semantic segmentation method with the help of deep learning algorithms is one of the methods in the sub-discipline called computer learning. Basically, the process is to give loaded data to deep artificial neural networks and train the artificial neural network with this data again and again until it creates correct predictions. In this study, DeepLab v3+ algorithm and two deep learning architectures such as Resnet18 and Resnet50 were chosen as backbone for feature extraction task from terrestrial images. The application was carried out on MATLAB. CamVid and Cityscapes datasets were used as datasets. Among the models applied, the one with the highest evaluation accuracy is, where the backbone is Resnet50 with 93.53% on Camvid and 89.29% on Cityscapes. The best results were applied to the images, which were taken for the study outside the data sets, and the results were evaluated visually.

1. Introduction

Deep learning is widely used for automated data processing as it requires less human intervention. (Atik and Ipbuker, 2021). Deep learning-based systems, both in the production and usage of products, is an issue on which studies are increasing (Sertkaya, 2022). Along with the developing hardware technologies (graphics cards, etc.), many new algorithms have been developed for the processing of big data with deep learning (Biyik *et al.*, 2023). Deep learning architectures are used for fast and automatic data extraction in large data sets (Atik *et al.*, 2022). Deep learning algorithms have a great importance in the field of image processing, which aims to obtain information from images by performing different operations on digital images (Ozgunluk *et al.*, 2022). Deep learning algorithms be trained with images, and it performs tasks such as object recognition and semantic segmentation in the field of image processing by using complex mathematical models containing multi-layer artificial neural networks. It has been shown in many studies that deep learning algorithms can detect complex patterns and show high performance in image processing thanks to its multi-layered structure.

In this study, deep learning models are trained using datasets made up of images prepared in accordance with semantic segmentation for the automatic recognition of objects in terrestrial images. DeepLab v3+ architecture (Chen et al. 2018) was used to capture the multi-scale context in images. Cambridge-driving Labeled Video Database (CamVid) is used for training and testing the algorithms. It is aimed to make high-accuracy predictions during the testing phase of the deep learning models used after the trainings.

2. Material and Method

In this section, DeepLabv3+ architecture used for feature extraction in semantic segmentation task and ResNet18 and ResNet50 models used as backbone structure will be discussed. In addition, information will be given about the data set and the metrics used in the evaluation.

2.1. Dataset

Cambridge-driving Labeled Video Database (CamVid) (Brostow *et al.*, 2009) was created by Cambridge University for autonomous driving and image processing studies. It contains 701 street images with a resolution of 720 x 960 with 32 classrooms. The data set is used in the training, testing and development of algorithms developed for autonomous driving and image processing. Label images are also available for each image, showing which class each pixel belongs to. The samples from the dataset are presented in Figure 1.

[*] **Corresponding Author**

[*](verani19@itu.edu.tr) ORCID ID 0000-0001-9532-1138
(atikm@itu.edu.tr) ORCID ID 0000-0003-2273-7751
(duranza@itu.edu.tr) ORCID ID 0000-0002-1608-0119

Cite this study

Verani, A., Atik, M. E., & Duran, Z. (2023). Comparative analysis of semantic segmentation of terrestrial images using DeepLabv3+. Intercontinental Geoinformation Days (IGD), 6, 324-326, Baku, Azerbaijan

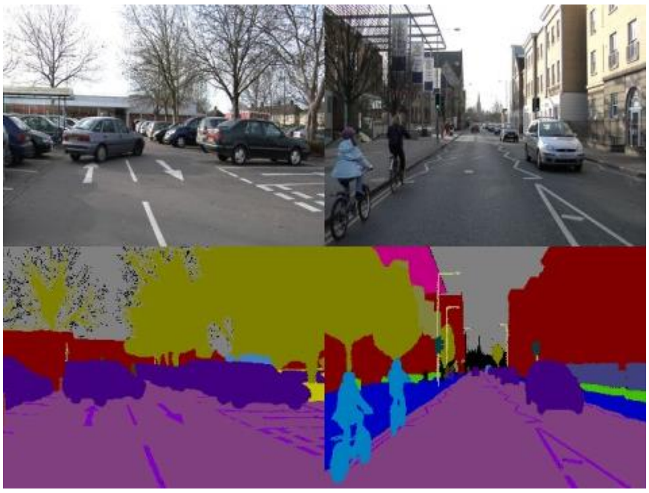

Figure 1. Image and ground truth samples from the CamVid dataset

2.2. DeepLab v3+

The algorithm used in the study uses the encoder-decoder module to capture semantic and spatial information from feature maps in the semantic segmentation task, and the spatial pyramid pooling module to capture object edge information (Chen et al., 2018). DeepLabv3 includes an ASPP module with batch normalization and image-level features. Atrous convolutions determine the density of features in convolutional networks.

In the DeepLabv3+ architecture, different deep learning networks are used as backbones. In this study, ResNet (He et al., 2015) versions were tested. Stacked residual units have been added to improve accuracy despite increased network depth in the ResNet architecture. Skipping connections solves the gradient disappearance problem by creating alternative shortcuts between connections (Atik et al., 2022). In other words, it implements the deep residual learning method, which performs identity matching with shortcut links that bypass one or more layers, and adds its outputs to the outputs of the other layer, to avoid deteriorating the training accuracy (He et al. 2015).

The ResNet architecture is named according to the number of layers. In this study, 18-layer and 50-layer models of the network are used. ResNet-18 has eighteen weight layers. ResNet-50 is designed in bottleneck structure to reduce training time. In ResNet-50, the bottleneck structure with three-layer blocks has been introduced instead of the two-layer blocks found in ResNet-34 (Atik et al., 2022).

2.3. Metrics

Two different metrics were used for the test set results of the data set. Overall accuracy is a metric calculated by taking the ratio of the total number of correct predictions obtained by a classification model to all predictions, and is designed so that classes are proportionally present and important. Unbalanced classes can prevent the metric from giving accurate results. Mean accuracy is obtained by calculating and averaging the accuracy for individual classes. It gives more effective results in cases where there is class imbalance.

2.4. Experiment

The images to be used are divided into 3 clusters as 60%, 20% and 20% as training, validation and test data, respectively. The training parameters are 15 epoch, sgdm optimizer, 0.001 learning rate were chosen. The optimum parameters were determined experimentally. For the experiments, HUAWEI D16 model computer with AMD Ryzen 5 4600H processor, 16 GB RAM, 512 SSD HD was used. All experiments are implemented in MATLAB environment.

3. Results

The test results obtained after the CamVid dataset training, in which ResNet18, ResNet50 networks that were used as the backbone infrastructure in the pre-trained DeepLab v3+ architecture, are presented in Table 1. Class-based accuracies have been found to differ according to the frequency, level of detail, and object sizes of the classes in the images. Given the uneven distribution of classes, the DeepLab v3+ResNet50 model produced more accurate prediction results than their average accuracy. In the test set, 85.65% accuracy was obtained with ResNet18 and 87.83% accuracy with ResNet50.

Table 1. CamVid Test Data Accuracies (%)

Model	Overall Acc.	Mean Acc.
DeepLabv3+ResNet18	91,20	85,65
DeepLabv3+ResNet50	92,49	87,83

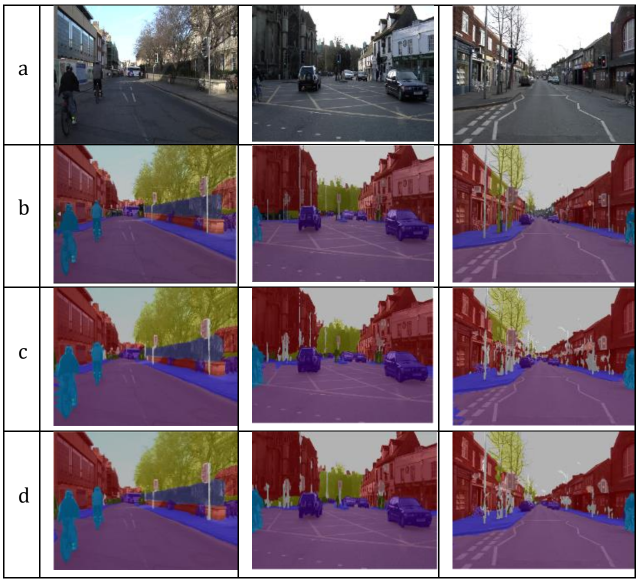

Figure 2. The result samples from the experiment. a) Image b) Ground truth c) Prediction of ResNet18 d) Prediction of ResNet50.

4. Discussion

In the study, ResNet18 and ResNet50 architectures were used in the backbone structure of the DeepLab v3+ network. As expected, ResNet-50 has higher accuracy.

However, the difference in accuracy between architectures is low. The model with ResNet 18 reduced the feature extraction time by 42.7%, while the average accuracy was reduced by 2.18%. Training times are 897 minutes for Resnet-50 and 514 minutes for ResNet-18. Since Resnet-50 is a deeper architecture, the training time is longer. Two different models were created that can generate predictive maps with semantically good accuracy.

5. Conclusion

In this study, research on semantic segmentation of terrestrial images with DeepLabv3+ architecture is presented. DeepLabv3+ which captures semantic information, spatial information and object edge information with its modules, is combined with ResNet which increases training accuracy with residual learning method. In future studies, the research can be expanded by including different datasets and architectures. Semantic segmentation of aerial images can also be studied.

References

Atik, M. E., Duran, Z., & Özgünlük, R. (2022). Comparison of YOLO versions for object detection from aerial images. International Journal of Environment and Geoinformatics, 9(2), 87-93.

Atik, S. O., & Ipbuker, C. (2021). Integrating convolutional neural network and multiresolution segmentation for land cover and land use mapping using satellite imagery. Applied Sciences, 11(12), 5551.

Atik, S. O., Atik, M. E., & Ipbuker, C. (2022). Comparative research on different backbone architectures of DeepLabV3+ for building segmentation. Journal of Applied Remote Sensing, 16(2), 024510-024510.

Biyik, M. Y., Atik, M. E., & Duran, Z. (2023). Deep learning-based vehicle detection from orthophoto and spatial accuracy analysis. International Journal of Engineering and Geosciences, 8(2), 138-145.

Brostow, G. J., Fauqueur, J., & Cipolla, R. (2009). Semantic object classes in video: A high-definition ground truth database. Pattern Recognition Letters, 30(2), 88-97.

Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European conference on computer vision (ECCV) (pp. 801-818).

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).

Sertkaya, C. (2022). Derin Öğrenme Tabanlı Nesne Tanıma Yeteneklerine Sahip Akıllı Robot Sisteminin Geliştirilmesi. European Journal of Science and Technology, (34), 211-216.