Intercontinental Geoinformation Days

http://igd.mersin.edu.tr/2020/

Effects of Orthophoto Band Combinations on Semantic Segmentation

Esra Altınoluk*[1], Özgün Akçay[2], Ahmet Cumhur Kınacı [3], Emin Özgür Avşar [2], Ahmet Batuhan Polat [2], Umut Aydar [2]

[1]*Çanakkale Onsekiz Mart University, Institute of Graduate Education, Geographic Information Technologies, Çanakkale, Turkey*
[2]*Çanakkale Onsekiz Mart University, Faculty of Engineering, Department of Geomatics Engineering, Çanakkale, Turkey*
[3]*Çanakkale Onsekiz Mart University, Faculty of Engineering, Department of Computer Engineering, Çanakkale, Turkey*

Keywords
Deep learning
Remote sensing
Classification
CNN
DeepLabv3

ABSTRACT

Recently classify of the high resolution orthophotos using the convolutional neural network (CNN), which is the popular architecture of image classification applications with deep learning. In this study, trainings were carried out using the DeepLabv3 architecture based on the CNN network named ResNet. Potsdam dataset was selected as the study region, which is presented as an open data set by the International Society for Photogrammetry and Remote Sensing (ISPRS).A total of 2112 images were used, 352 of this images used for verification and another 352 images used for test data. It has been trained with five different spectral band combinations: RG (red-green), RB (red-blue), GB (green-blue), RGB (red-green-blue) and IRRG (infrared-red-green). After the trainings,the classification success was compared on the test data. RG, RB, GB, RGB, IRRG band combinations produced, %91, %85, %91, %92, %91 training accuracy rates, respectively. Results demonstrate that, using different band combinations on trainings give us different accuracy.

1. INTRODUCTION

Remote sensing systems are an important data source that enables us to access up-to-date information about the earth. However, in line with the technological developments in recent years, interest in the production of high-resolution aerial images and analysis of these images is increasing. In the literature, land use / land cover detection from high-resolution aerial images (Castelluccio et al. 2015; Zhang and Zhu 2011) is one of the most common research topics and is carried out by classification process.

Classification in remote sensing means labeling each object in the image to its class. The traditional pixel-based classification method uses the spectral properties of pixels. However, the heterogeneous pixels of high-resolution aerial images negatively affect the classification results. For this reason, instead of using individual pixels, objects called segments were formed by grouping neighboring pixels with similar spectral properties, and an object-based classification method, which is used in spatial information as well as spectral content, was developed (Blaschke 2010; Veljanovski et al. 2011). Although these methods achieved important results in classification applications, the increasing resolution of the images and the more detailed information content could not achieve the desired success. Algorithms such as random forest (Breiman 2001), support vector machines (Cortes and Vapnik 1995), artificial neural networks (Foody 1996) etc. have been developed in order to overcome these weaknesses and increase classification accuracy. However, in order to use the mentioned methods, feature extraction that requires a lot of time and expertise should be done (Arel et al. 2010). On the other hand, deep learning, which emerged with the development of artificial neural networks, learned from the data itself without the need for feature extraction and overcome these problems and achieved significant success.

Deep learning is a hierarchical learning method that analyzes input data from simple to complex with the help of numerous nonlinear layers (Altınoluk et al. 2019; Lecun et al. 2015). For example; in image analysis, the first layers define the edges, while the next layers can define concepts such as numbers, letters or human faces.

*Corresponding Author

*(esraltnoluk@gmail.com) ORCID ID 0000-0002-4688-2366
(akcay@comu.edu.tr) ORCID ID 0000-0003-0474-7518
(cumhur.kinaci@comu.edu.tr.) ORCID ID 0000-0002-8832-5453
(ozguravsar@comu.edu.tr) ORCID ID 0000-0002-3804-1209
(abpolat@comu.edu.tr) ORCID ID 0000-0002-7495-1998
(umutaydar@comu.edu.tr) ORCID ID 0000-0002-3987-6435

Cite this study

Altınoluk E, Akçay Ö, Kınacı A C, Avşar Ö, Polat A B & Aydar U (2020). Effects of Orthophoto Band Combinations on Semantic Segmentation. Intercontinental Geoinformation Days (IGD), 9-12, Mersin, Turkey

Convolutional Neural Network (CNN), which is the most popular architecture of deep learning, is generally used in studies such as object recognition and image classification (Krizhevsky et al. 2012). Examples of these network structures that have achieved significant success in classification studies are LeNet(LeCun et al. 1989; LeCun et al. 1998), Alexnet (Krizhevsky et al. 2012), VGGNet (Simonyanand Zisserman 2014), GoogleNet (Szegedy et al. 2015) and ResNet (He et al. 2016).Recently, a series of DeepLab have been proposed based on the theory of atrous convolution(Yu and Koltun et al. 2015). The proposed scheme using DeepLabv3+ semantic segmentation algorithm can fully utilize the spatial–spectral and context information of images as well as recognize the spatial geometric relationship between intraclass and interclass of the ground objects(Zhang et al. 2020).

In this study, it is aimed to train ResNet-based DeepLabv3 deep learning network architecture with five different spectral band combinations, namely RG, RB, GB, RGB and IRRG, using the high resolution ISPRS-Potsdam data set for land use and land cover classification. In this context, the performance of the DeepLabv3 neural network was examined and the classification accuracies were analyzed. In addition to RGB input data, which is widely used in deep learning, the contribution of IRRG data with infrared band is use for network training is among the research questions.

2. METHOD

ResNet based DeepLabv3 architecture was trained with RG, RB, GB, RGB and IRRG combinations of ISPRS-Potsdam dataset as a solution to the classification problem of high-resolution remote sensing data. Using the obtained models, segmentation maps were produced from the images and success rates were calculated.

2.1. Dataset

The Potsdam data set produced by ISPRS for the 2D Semantic Labeling Contest and presented as an open data set was used in the study. There are 6 categories in the data set: impervious surfaces (white), building (blue), low vegetation (cyan), tree (green), car (yellow) andclutter (red). The Potsdam dataset contains 38 high resolution true orthophotos; the size of each photo is 6000x6000 pixels. Fig. 1 shows an example of ISPRS-Potsdam images.

Dataset preprocessing was carried out with Matlab, training and testing of DeepLabv3 was performed using Tesla K80 and T4 GPU, which is available via Google Colaboratory, using Tensorflow version 1.15.2.

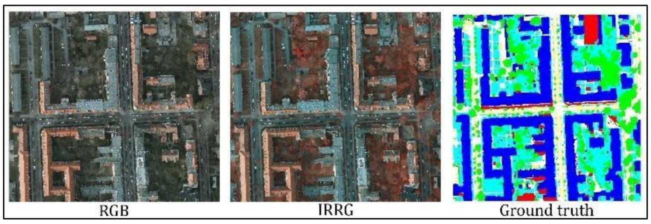

Figure 1. ISPRS-Potsdam orthophoto examples

In the training of the DeepLabv3neural network, RG, RB, GB and IRRG band combinations were used to investigate whether different band combinations contribute to the RGB data as well. Training data for each band combinations cover the same area.

Data slicing (Liu et al. 2018) was used in the preprocessing phase of the data. Since high resolution data will be difficult to transmit directly to the model due to GPU memory limit, 4 orthophotos selected for network training were sliced in 400x400 size and 30% overlap ratio.

Fig 2 shows sample data of 400x400 sizes RGB, IRRG, ground truth, RG, RB and GB used in training. DeepLabv3 neural network used in the study; it was trained with a total of 1760 images, including 1408 training (80%) and 352 verification data (20%) for five different spectral band combinations. The performance of the trained models was measured with 352 test data.

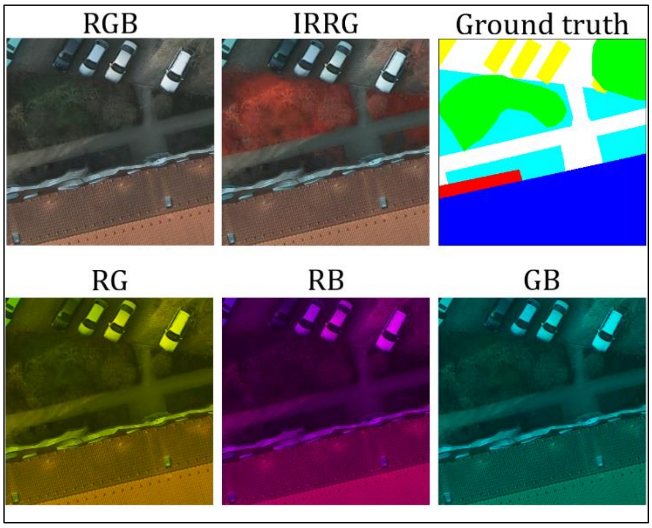

Figure 2. Sample RGB, IRRG, ground truth, RG, RB and GB data used in training

2.2. DeepLabv3

DeepLabv3 (Fig. 3) neural network is a CNN architecture developed for use in semantic segmentation applications (Chen et al. 2017b). This network, developed on the basis of the ResNet network structure, includes 4 ResNet blocks, atrous spatial pyramid pooling (ASPP) and global average pooling (GAP).

Atrous convolution, located in ResNet 4th block and ASPP, expands the field of view of the filter by a certain rate (r-1 number 0 is added) and provides more dense feature maps with fewer parameters than the classical convolution (Ratul et al. 2019; Wang et al. 2019).

GAP converts the feature map of size h (height) x w (width) x d (depth) into 1 x 1 x d. This process, which is used to reduce the tensor size as in the pooling layer, takes the average of each feature map and thus prevents the model from over learning (Lin et al. 2014).

ASPP, which first appeared in the DeepLabv2 (Chen et al. 2017a) architecture, this module applies different rates of atrous convolution (r = 6, 12, 18) operations in parallel, as well as global average pooling. In this way, feature maps with different levels of detail are produced. Classical convolution with 1x1 filter is applied by combining the feature maps produced as a result of GAP and each atrous convolution.

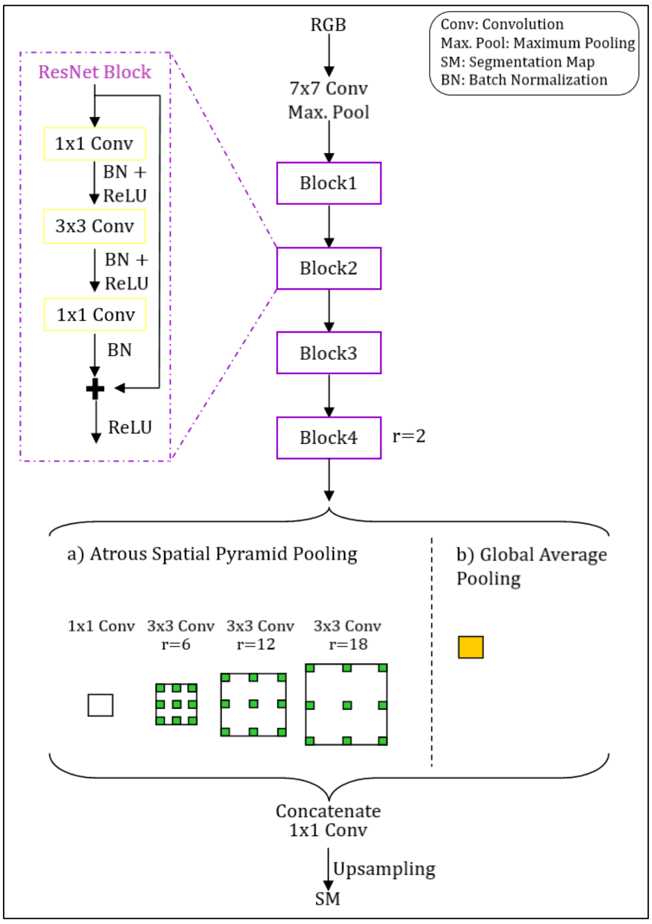

Figure 3. DeepLabv3 architecture

3. RESULTS

3.1. Training the DeepLabv3

DeepLabv3; for each spectral band combination, 40 epochs, 4 mini-batch sizes, Adam optimization algorithm and ResNet-50 weights were trained with using transfer learning. As a result of the trainings, the accuracy of the RG 0.91, RB 0.85, GB 0.91, RGB 0.92, IRRG 0.91 model was achieved. The training took 6 hours for RG and RB data, 6.5 hours for GB data, 9 hours for RGB data, and 9.5 hours for IRRG data.

3.2. Testing the DeepLabv3

DeepLabv3 models were tested with 352 test data which using RG images 30th epoch, RB images9th epoch, GB images32th epoch, RGB images38th epoch and IRRG images20th epoch weights. For the evaluation of segmentation maps produced from DeepLabv3 models, precision, precision and f-score were calculated (Table 1). The accuracy of the tested models are 0.77, 0.76, 0.83, 0.82, and 0.83 respectively. Segmentation maps are shown in Fig. 4.

4. DISCUSSION

According to the accuracy metrics, it has been observed that the orders of success between combinations are GB, IRRG, RGB, RG and RB. There is a difference of approximately 1% between the test accuracy of models trained with GB, IRRG and RGB data. In addition to having similar results, it has been observed that the GB combination is quite successful in impervious surface, building and car classes. It has been observed that RGB has low vegetation and IRRG is better in tree class than other combinations.

When the classification results were examined, it was concluded that the DeepLabv3 architecture trained with GB, RGB and IRRG data gave successful results in general. In addition, the red and infrared bands are thought to have an accuracy-enhancing effect on low vegetation.

Table 1. Accuracy metrics for each class

		İmp surf.	Building	Low veg.	Tree	Car	Clutter
Precision	RG	0.78	0.87	0.68	0.82	0.80	0.14
	RB	0.73	0.94	0.61	0.84	0.61	0.34
	GB	0.83	0.93	0.71	0.84	0.82	0.37
	RGB	0.82	0.94	0.70	0.83	0.72	0.26
	IRRG	0.83	0.95	0.71	0.81	0.81	0.32
Recall	RG	0.74	0.91	0.70	0.70	0.67	0.31
	RB	0.71	0.78	0.85	0.74	0.66	0.27
	GB	0.80	0.92	0.83	0.76	0.73	0.27
	RGB	0.78	0.90	0.84	0.78	0.73	0.22
	IRRG	0.78	0.90	0.83	0.82	0.70	0.27
F-Score	RG	0.76	0.89	0.69	0.76	0.73	0.19
	RB	0.72	0.85	0.71	0.78	0.63	0.30
	GB	0.81	0.93	0.76	0.80	0.78	0.31
	RGB	0.80	0.92	0.77	0.80	0.72	0.24
	IRRG	0.80	0.92	0.77	0.82	0.75	0.29

5. CONCLUSION

In the study, classification success of high resolution orthophotos using DeepLabv3 architecture was investigated and five different data structures of ISPRS-Potsdam data set were used in line with this goal.

As a result, the neural network trained with GB, IRRG and RGB data performed better than the RG and RB neural networks. When we look at the in-class performances, it is seen that GB gives better results in determining the impermeable surface, building, car and clutter, RGB in determining low vegetation class and IRRG in determining the tree class, albeit with slight differences. In general, the higher accuracy of building class compared to other classes are thought to be due to the fact that it is more intense in the data set. In particular, clutter class, which has the least density and is difficult to distinguish in the data set, has the lowest accuracy.

This study, conducted using Google Colaboratory, could not increase training data due to time and memory limitations. Therefore, it is predicted that it can perform a higher performance than the results obtained with more training data and a more powerful GPU. Generally; when all the obtained results were evaluated, it was concluded that using deep learning algorithms and GB, RGB and IRRG data structures, land use and land detection classification can be made.

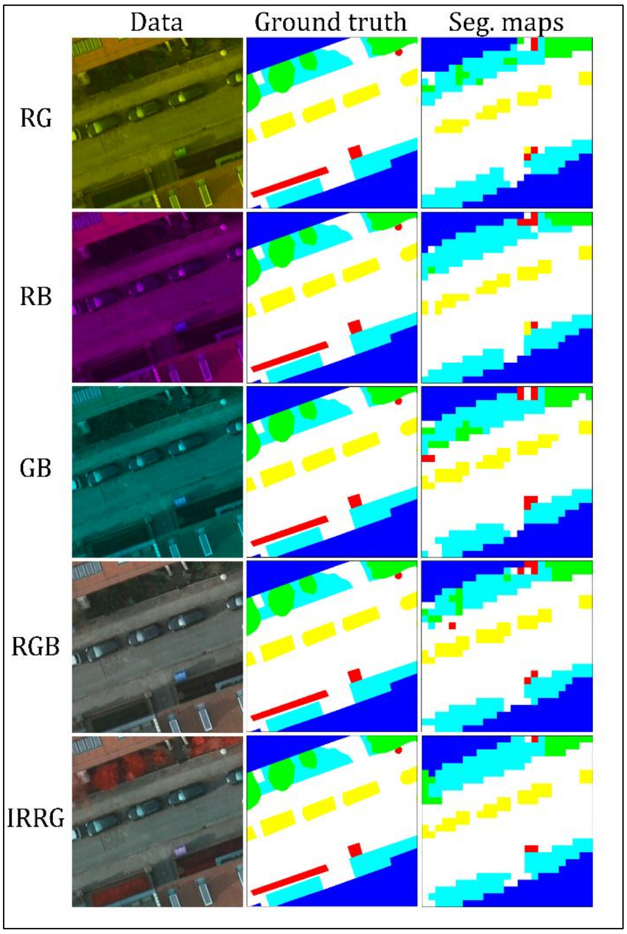

Figure 4. Segmentation maps examples

ACKNOWLEDGEMENT

This research was supported by The Scientific and Technological Research Council of Turkey (TÜBİTAK), Project No: 119Y363

REFERENCES

Altinoluk E, Akçay Ö, Avşar E Ö, Aydar U (2019). Recent Approaches in Segmentation Analysis with Deep Learning Methods. X. TUFUAB Teknik Sempozyumu, 20-25, Aksaray, Türkiye.

Arel I, Rose D C, & Karnowski T P (2010). Deep machine learning-a new frontier in artificial intelligence research [research frontier]. IEEE computational intelligence magazine, 5(4), 13-18.

Blaschke T (2010). Object based image analysis for remote sensing. ISPRS journal of photogrammetry and remote sensing, 65(1), 2-16.

Breiman L (2001). Random forests. Machine learning, 45(1), 5-32.

Castelluccio M, Poggi G, Sansone C & Verdoliva L (2015). Land use classification in remote sensing images by convolutional neural networks. arXiv preprint arXiv:1508.00092.

Chen L C, Papandreou G, Kokkinos I, Murphy K & Yuille A L (2017a). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence, 40(4), 834-848.

Chen L C, Papandreou G, Schroff F, Adam H (2017b). Rethinking Atrous Convolution for Semantic Image Segmentation. arXiv preprint arXiv:1706.05587.

Cortes C & Vapnik V (1995). Support-vector networks. Machine learning, 20(3), 273-297.

Foody G M (1996). Relating the land-cover composition of mixed pixels to artificial neural network classification output. Photogrammetric engineering and remote sensing, 62(5), 491-498.

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, 770-778.

Krizhevsky A, Sutskever I & Hinton G E (2012). ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems 25 (NIPS 2012), 1097-1105.

LeCun Y, Bengio Y & Hinton G (2015). Deep Learning. Nature, 521(7553), 436.

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

LeCun Y, Bottou L, Bengio Y, Haffner P (1998). Gradient-Based Learning Applied to Document Recognition. Proceedings of the IEEE, 86(11), 2278-2324.

Lin M, Chen Q, & Yan S (2014). Network In Network. arXiv preprint arXiv:1 312.4400.

Liu Y, Ren Q, Geng J, Ding M & Li J (2018). Efficient Patch-wise Semantic Segmentation for Large-scale Remote Sensing Images. Sensors, 18(10), 3232.

Ratul M A R, Mozaffari M H, Parimbelli E & Lee W (2019). Atrous Convolution with Transfer Learning for Skin Lesions Classification. *BioRxiv*, 746388.

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

Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D & Rabinovich A (2015). Going deeper with convolutions.In Proceedings of the IEEE conference on computer vision and pattern recognition.

Veljanovski T, Kanjir U & Ostir K (2011). Object-based image analysis of remote sensing data. Geodetski vestnik, 55(4), 678-688.

Wang Y, Liang B, Ding M, Li J (2019). Dense Semantic Labeling with Atrous Spatial Pyramid Pooling and Decoder for High-Resolution Remote Sensing Imagery. Remote Sensing, 11(1), 20.

Zhang R & Zhu D (2011). Study of land cover classification based on knowledge rules using high-resolution remote sensing images. Expert Systems with Applications, 38(4), 3647-3652.

Yu F & Koltun V (2015). Multi-scale context aggregation by dilated convolutions, arXiv: 1511.07122.

Zhang Z, Huang J, Jiang T, Sui B K & Pan X L (2020). Semantic segmentation of very high-resolution remote sensing image based on multiple band combinations and patchwise scene analysis. J. Appl. Remote Sens., 14, 016502.