5th Intercontinental Geoinformation Days

igd.mersin.edu.tr

Scene classification of Google Earth Images with different deep learning models

Şaziye Özge Atik*[1]

[1] *Gebze Technical University, Faculty of Engineering, Department of Geomatics Engineering, Kocaeli, Türkiye*

Keywords
Scene Classification
Algorithms
Automatic Labeling
CNN
Google Earth

Abstract

Deep Convolutional Neural Networks are widely used for the automatic labeling of satellite imagery. Four different CNN models were trained using the AID dataset, which is an openly shared dataset, and the performances of the trained models were tested using the general accuracy metric. Test performances of ESA models were compared with similar studies in the literature on the same data set. At the same time, the model transferability of the trained models with the domain-shift application was tested with the help of different images obtained from the Google Earth platform. In this way, it is aimed to expand the application of automatic labeling of land use and land cover classes with deep CNN models in different data. The test results also support that the use of semantic scene classification algorithms is becoming more and more promising with the developing technology and opportunities.

1. Introduction

A large number of research and studies are carried out for many purposes, such as monitoring the natural resources in the world, determining and monitoring the land use classes, and regional environmental monitoring programs. Remotely sensed scene classification applications are carried out with many different deep learning architectures in the literature (Chaib et al. 2017, Liu et al. 2018, Marmanis et al. 2016, Nogueira et al. 2017, Othman et al. 2016, Yu et al. 2018, Zhao et al. 2018). It is essential to use open shared and, free data in many studies carried out for environmental monitoring. The type and quality of the data used in the study are essential for these analyses carried out for large areas in the globalizing world and the applications are repeated at specific periods. For this purpose, it is aimed to use the free images provided by the Google Earth (GE) platform in the classification of aerial images. In this study, the performance of the models was tested on GE images of other regions with the help of ESA models trained on the openly shared data set. This study is also a proposal on determining the land use classes of the images of any region that need to be classified at any time. In the study, the AID dataset was used to train the algorithms. Densenet201, Resnet18, VGG16, and Alexnet were used as ESA models in the implementation phase. The classification results of these models were compared over the overall accuracy metric. At the same time, the results of this study were compared with other similar studies in the literature. Finally, the automatic classification performances of the models trained on GE images selected from different regions of İstanbul were also tested.

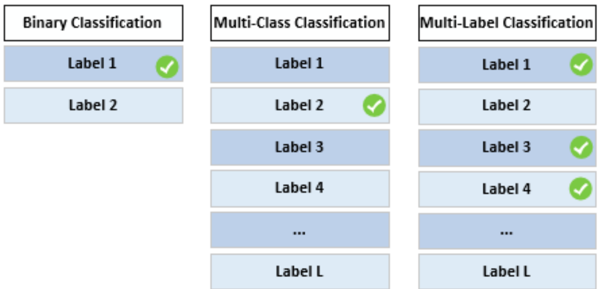

Figure 1. Types of scene classification

In Fig.1 the types of scene classification were illustrated. All classes are grouped under automatic labeling.

2. Data and Methodology

The AID dataset (Xia et al. 2017) contains 10000 GE images in 600 x 600 size, between 200-400 for 30 different classes, with larger capacity among similar datasets (UC Merced, WHU-RS19 and EuroSat). This is because GE images consist of a mosaic of images detected

* **Corresponding Author**

*(soatik@gtu.edu.tr) ORCID ID 0000-0003-2876-040X

Cite this study

Atik, S. O. (2022). Scene Classification of Google Earth Images with Different Deep Learning Models 5th Intercontinental Geoinformation Days (IGD), 153-155, Netra, India

by multiple sensors for different regions. For this reason, the AID data set consists of images with many other properties (resolution) detected by different sensors. This situation also significantly differs from datasets obtained with a single sensor, such as the UC Merced dataset. In this way, the study also considers the need to apply Geographic Information Systems (GIS) studies in different places and at other times.

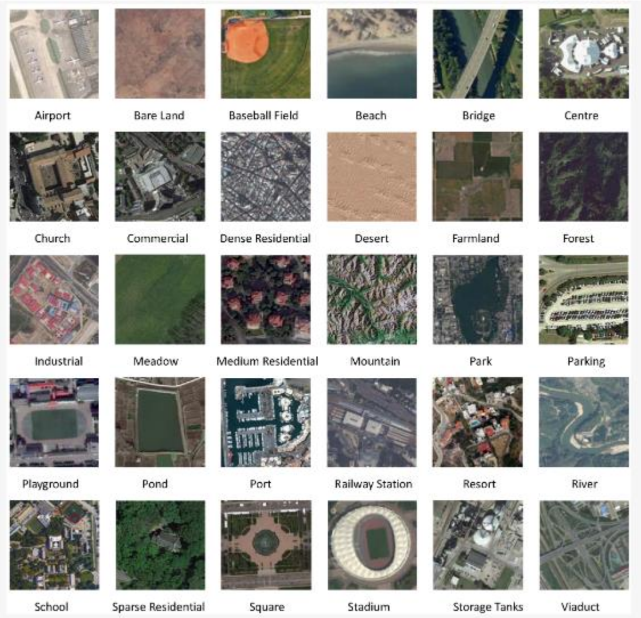

Figure 2. Selected sample images of the AID dataset (Zhang et al. 2020)

Convolutional Neural Networks (ESA) architectures have been used to classify images. AlexNet architecture (Krizhevsky e,t al. 2012) and validated in ImageNet and CIFAR da, is a widely used model. ResNet architecture (He et al. 2016) is another preferred algorithm with multiple convolution layers in many fields. The architecture takes names such as ResNet18 and 50, 101 according to the number of layers it contains. VGG was put there by Simonyan and Zisserman (Simonyan and Zisserman 2014) as ConvNet architecture with 1000 classes. VGG ,11,16 and 19 models are used according to their weight layers. DenseNet (Huang e,t al. 2017) is another architecture built as a Dense Convolutional Network and its performance was tested in CIFAR-10, CIFAR-100, SVHN, and ImageNet competitions.

$$Overall\ Accuracy = \left(\frac{TN+TP}{N}\right) \quad (1)$$

In the Eq.1, TP is True Positive, TN is True Negative, N is the total number of samples.

3. Experiment

The pre-trained weights of the ESA networks used were used as the starting weight. The number of epochs used in the training of ESA models was determined experimentally and selected as 20. The AID dataset is randomly divided into 50% training and 50% testing. However, after this segmentation, the performances of the ESA models were tested with the help of the same training and test images. A system equipped with i7-11800H, a 2.30 GHz processor, GTX 3070 graphics card, and 32 GB RAM was used in the applications.

Table 1. The overall accuracy of the CNN models

Model	Overall Accuracy
DenseNet-201	93,18
VGG-16	92,84
ResNet-18	89,72
AlexNet	87,04

The overall accuracy values of the models used are shown in Table 2. Accordingly, the highest accuracy value was obtained with the DenseNet201 model at 93.18%. The lowest accuracy value belongs to AlexNet model with, 87.04%. At the same time, the highest accuracy in the GE images selected for this study from different regions was again obtained with the DenseNet 201 model. The overall accuracy values are illustrated in Fig.3.

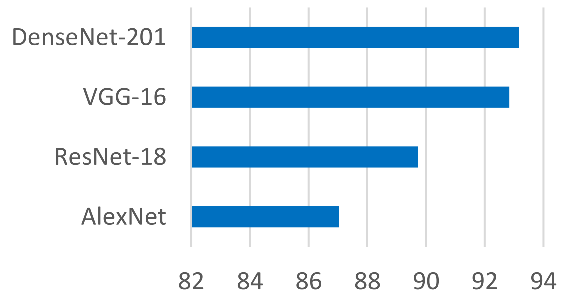

Figure 3. Overall accuracy values of CNN models

4. Results and Discussion

The test performances of the models used in this study were compared with similar studies in the literature. Accordingly, higher performance was obtained in the general accuracy criterion compared to other studies (Anwer et al. 2017, Xia et al. 2017). However, although similar accuracy values could not be reached on different images selected on Google Earth, the relevant land use class was determined correctly in many images. It is due to the model transfer process. Transferring the models to other data can decrease accuracy, as expected.

The models are used in the second test phase used for testing GE sample images. Classes of the AID dataset were collected on the GE platform from different regions of İstanbul. In Fig.3 and Fig.4, several samples are shown for false negative and true positive results. Alexnet model gave the worst test performances and DenseNen201 model provided the best performances for GE test images.

In future projections, this study can be enriched by testing models trained on a single dataset using images from different datasets. In this way, by measuring the model transfer capacity, semantic scene classification applications can be used by many experts in more expansive areas.

At the same time, then is that the models' data-based functionality areas, one of the main problems of deep learning, which is the data-based functionality of the models, will be able to go beyond.

Figure 3. False Positive Samples (Left: GT Pond and Predicted Park; Right: GT Park and Predicted: Bridge)

Figure 4. True Positive Samples (Left: GT and Predicted: Bareland, Right: GT and Predicted: Square)

5. Conclusion

Automatic classification has the potential to become very popular soon, as in many different applications, in the automatic detection of land use classes from aerial images. However, realizing such applications with images of free and continuous image-sharing platforms such as GE will be one of the main pillars of widespread use. From planning automatic multi-label applications to due diligence, it can be preferred in many subjects, from environmental analysis to future simulations.

References

Anwer, R.M.; Khan, F.S.; van deWeijer, J.; Monlinier, M.; Laaksonen, J. (2017) Binary patterns encoded convolutional neural networks for texture recognition and remote sensing scene classification. arXiv:1706.01171.

Chaib, S.; Liu, H.; Gu, Y.; Yao, H. (2017) Deep feature fusion for VHR remote sensing scene classification. IEEE Trans. Geosci. Remote Sens., 55, 4775–4784.

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).

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 (pp. 4700-4708).

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25, 1097-1105.

Liu, Y.; Zhong, Y.; Fei, F.; Zhu, Q.; Qin, Q. (2018) Scene Classification Based on a Deep Random-Scale Stretched Convolutional Neural Network. Remote Sens. 2018, 10, 444.

Marmanis, D.; Datcu, M.; Esch, T.; Stilla, U. (2016) Deep learning earth observation classification using ImageNet pretrained networks. IEEE Geosci. Remote Sens. Lett., 13, 105–109.

Nogueira, K.; Penatti, O.A.B.; dos Santos, J.A. (2017) Towards better exploiting convolutional neural networks for remote sensing scene classification. Pattern Recognit. 61, 539–556.

Othman, E.; Bazi, Y.; Alajlan, N.; Alhichri, H.; Melgani, H. (2016) Using convolutional features and a sparse autoencoder for land-use scene classification. Int. J. Remote Sens., 37, 2149–2167.

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

Xia, G. S., Hu, J., Hu, F., Shi, B., Bai, X., Zhong, Y., ... & Lu, X. (2017). AID: A benchmark data set for performance evaluation of aerial scene classification. IEEE Transactions on Geoscience and Remote Sensing, 55(7), 3965-3981.

Yu, Y.; Liu, F. (2018) A two-stream deep fusion framework for high-resolution aerial scene classification. Comput. Intell. Neurosci., 8639367.

Zhang, S., Wu, G., Gu, J., & Han, J. (2020). Pruning convolutional neural networks with an attention mechanism for remote sensing image classification. Electronics, 9(8), 1209.

Zhao, L.; Zhang, W.; Tang, P. (2018) Analysis of the inter-dataset representation ability of deep features for high spatial resolution remote sensing image scene classification. Multimed. Tools Appl.