SEGMENTATION OF SENTINEL SATELLITE IMAGES: COMPARISON OF MACHINE-LEARNING AND DEEP-LEARNING-BASED MODELS

Olga Tokareva, Ph.D. Associate Professor.
Cristian Machuca, Graduate Student.
Tomsk Polytechnic University
E-mail: kristianrodrigo1@tpu.ru

Introduction

Evaluating the impact of environmental dynamics and human activities on the landscape enable us to make better and informed decisions, but at the same time represent a challenge. Satellites can cover large and difficult access areas capturing multispectral images. Semantic segmentation of these images allows us to overcome such challenge. According to our needs and requirements there are several tools, methods and algorithms used for land cover mapping [1].

Therefore, the goal of this work is to compare the performance of Machine Learning (ML) and Deep Learning algorithms for land cover classification, and contrast them with the Sen2Cor processor for Sentinel-2 satellite images.

Algorithm Description

Three different classification algorithms were developed: K-Means clustering algorithm, Random Forest (RF) classifier and Convolutional Neural Networks (CNN). The algorithms were tested on a scene of 4 km^2 from a tropical dry forest in Ecuador. We created a balanced dataset for pixel-based classification. The dataset consists of 480000 observations (rows) and 13 variables (columns): 12 pixel spectral attributes and “label”. In turn, the “label” variable includes four land cover classes (water body, bare soil, clouds and vegetation). The pixels were obtained from the Sentinel-2 Hub Service which delivers atmospheric corrected images with 10 m spatial resolution. The dataset was splitted into a training and test dataset at a 70:30 ratio. The machine learning and deep learning algorithms were developed and evaluated using the Python programming language.

The Sen2Cor processor provided a Level-2A (L2A) output, which consisted of a Scene Classification (SCL) image with four main classes (Table 1). This image was used as reference for the comparison.

Twelve spectral bands were taken into account for all the algorithms. The Scikit-learn Python library was chosen for modelling the machine learning algorithms. The CNN was developed using the Keras and TensorFlow Python libraries.

Four cluster centers were defined for the K-Means algorithm. The random forest consisted of 100 estimators and depth tree equals three, while the CNN consisted of two convolution layers and a fully connected layer.

Evaluation

The RF classifier showed a classification accuracy of approximately 97,85%, whereas the CNN had an accuracy of 99.78%.

Effectiveness of semantic segmentation can be determined with the use of metrics such as Intersection over Union (IoU) [2]. This metric known also as the Jaccard index, evaluates how close the prediction is to the ground truth [3].

In this case, we have considered the Sen2Cor Scene Classification as ground truth to evaluate the predictions similarity. Each one of the classes was assessed individually. The mean Intersection over Union value (mIoU), and the Overall Accuracy (OA) were calculated. To determine the IoU per class the following formula was used:

$$IoU = TP / (TP + FN + FP)$$

Where TP is the true positives, FN is the false negatives and FP is the false positives. The mIoU value is the unweighted average of IoU over all the classes [4]. The OA value refers to the total true predicted results compared to all predictions, and was calculated as follows:

$$OA = (TP + TN) / (TP + TN + FN + FP)$$

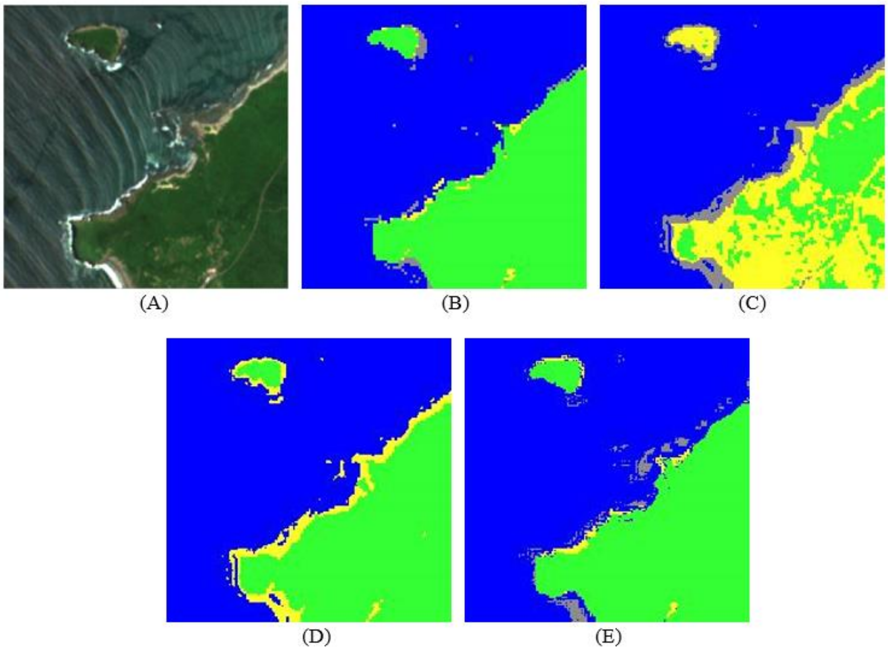

Fig. 1. A) RGB composite satellite image ; B) Sen2Cor Scene Classification and results of land cover classification using machine learning and deep learning algorithms; C) K-Means; D) Random Forest; E) CNN.

Table 1. IoU Score by class and algorithm, contrasted with the Sen2Cor Scene Classification. OIoU is the global IoU accuracy, while, mIoU is the unweighted average of IoU of each class.

Number	Class	Color	K-Means	RF	CNN
1	Water (dark and bright)		0.94831981	0.95693281	0.94820671
2	Bare soils / deserts		0.02849604	0.09298086	0.28343949
3	Vegetation		0.59282732	0.94181604	0.9822301
4	Clouds		0.17341329	0.00000000	0.05966162
mIoU			0.43576411	0.49793242	0.56838448
OA			0.70256235	0.89667844	0.92529842

Conclusion

Based on the Jaccard index, we can conclude that the CNN had the best performance, followed by the RF classifier and at last the K-Means algorithm. Nevertheless, if we contrast the satellite image with the classification algorithms (visual inspection), the RF algorithm is the most accurate, but due to the lack of real ground truth, we are not able to determine the IoU index. As the dataset continue to grow, more accurate and reliable segmentation is expected.

References

1. Horning N. Land cover classification methods. American Museum of Natural History. Center for Biodiversity and Conservation. – URL: http://biodiversityinformatics.amnh.org (accessed on February 19, 2020).
2. Siam M., Gamal M., Abdel-Razek M., Yogamani S., Jagersand M. RTSEG: Real-time semantic segmentation comparative study. In Proceedings of the IEEE International Conference on Image Processing. – 2018. – P. 1603-1607.
3. Rezatofighi H., Tsoi N., Gwak J., Sadeghian A., Reid I., Savarese S. Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2019. – P. 658-666.
4. Svensson T. Semantic Segmentation of Iron Ore Pellets with Neural Networks, Master's Thesis. – 2019. – P. 21-22.