A UNet-based solution for detecting deforestation and reduction of reservoirs and glaciers

Silvia Naro[1], Alice Bovio[2], Carmine Cisca[1] and Fabio Previtali[1,*]

[1]AlmavivA S.p.A., Via di Casal Boccone 188/190, Rome, 00137, Italy
[2]AlmavivA S.p.A., Via dei Missaglia, 97 - ed. B4, Milan, 20142, Italy

Abstract

Climate change is one of the biggest problem that humanity ever faced in their history. The causes are many and humans do have - unfortunately - a big responsibility about what is happening. Climate change is recognized as an issue that has negative effects on the ecosystem and it is mainly caused by the wild and uncontrolled deforestation of the main world forests. Another important negative effect of climate change is the sudden drying up of reservoirs and melting glaciers. Detecting deforestation and defining the causes of deforestation is an important process that could help monitor and prevent it to happen. Deforestation detection has been boosted by recent advances in geospatial technologies and applications, especially remote sensing technologies and machine learning techniques. This paper presents a land monitoring solution for deforestation, reduction of reservoirs and glaciers and the approach is based on a conceptual framework and it has been quantitatively validated on an open-source data set and qualitative evaluated on images of the Italian territory. The framework is based on machine learning and image processing techniques. It consists of three main steps, which are data preparation, model training and validation. The implementation of the proposed approach shows promising performance for detecting deforestation, reduction of reservoirs and glaciers.

1. Introduction

Image segmentation is one of the most addressed problems by scientific community in the field of Deep Learning (DL), that has the objective of assigning to every pixel of an image a pre-defined given class.

In the field of image analysis, deep neural networks have been proved to be effective in solving most of the problems such as for example related to smart surveillance (e.g., recognition of people, vehicles and other moving objects) and to medical diagnostics (e.g., recognition of injuries, diseases or tumors). In all these cases, the added value of solutions based on machine learning consists in the possibility of obtaining precise information much faster and with a lower cost with respect to traditional data analysis methods.

The combination of infrastructure cost reduction, computational power and effectiveness that characterizes DL-based image analysis solutions allowed to apply them on the field of remote sensing which is the discipline that deals with the analysis and extraction of information from collected data from remote instruments such as sensors, aircraft and, in particular, satellites. In recent decades, thanks to lowering of the price for acquiring high-quality satellites images and the availability of modern state-of-the-art machine learning frameworks, it was possible to acquire large amounts of data for training cutting-edge machine learning models for detecting interesting areas in an image.

Contributions. We propose a novel tool for automatic time-lapse land monitoring for a more effective and prompt control of the national territory for fighting against deforestation and for the reduction of reservoirs and glaciers.

2. Related Work

The feasibility of machine learning approaches in the remote sensing domain has been demonstrated in many applications such as for example earth observation [1], detecting changes on the earth's surface [2], semantic segmentation [3] and assessment of the sustainability of forest management [4, 5].

With regard to deforestation and reduction of reservoirs and glaciers, which are the general scope of this paper, machine learning approaches can be classified into two categories: 1) approaches detecting the location of areas at risk of deforestation and 2) approaches analyzing the variables that drive deforestation [6]. Chang et al. [7] proposed a machine learning model to enhance the estimates of forest land cover type and forest structural metrics. It is a multitask model that performs both classification and regression concurrently, thereby consolidating several independent tasks and models into one stream. Maeda et al. [8] applied a machine learning model to detect land use changes in the Amazon. Based on change interpretation, they could identify areas with high risk of being burned and improve current fire scar

Ital-IA 2023: 3rd National Conference on Artificial Intelligence, organized by CINI, May 29–31, 2023, Pisa, Italy
*Corresponding author.
s.naro@almaviva.it (S. Naro); a.bovio@almaviva.it (A. Bovio); c.cisca@almaviva.it (C. Cisca); f.previtali@almaviva.it (F. Previtali)

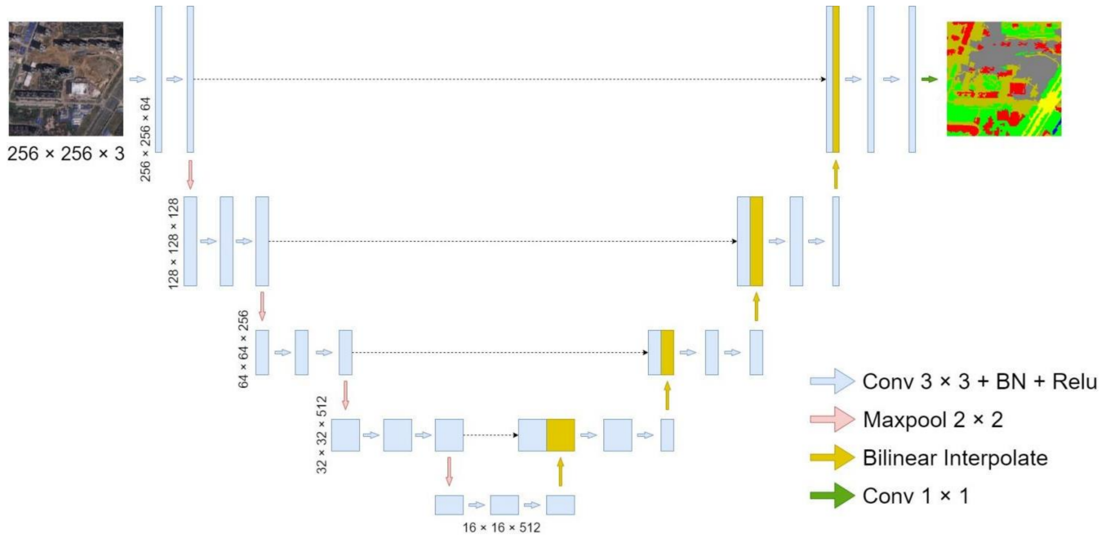


Figure 1: UNet architecture on which it has been designed the proposed solution.

mapping by enabling the distinction between fires in primary forests and fires in previously burned areas. Kehl et al. [9] proposed a study to detect daily deforestation in the Amazon rainforest. They developed an approach to train machine learning models on satellite images, and conducted a spectrum temporal analysis of the deforestation area. The approach aided in understanding the dynamics of the deforestation in the Amazon rainforest.

Class	Percentage
Reservoirs	6.00 %
Forest	33.30 %
Roads	1.60 %
Buildings	0.85 %
Other	58.25 %

Table 1
Summary of the LandCover.ai data set.

3. Modelling Approach

In the following next sections, we present the data set used for training the machine learning model and the architecture of the proposed solution.

3.1. Data set

The data set used for the development of the proposed machine learning model is the LandCover.ai[1] [10] (Land Cover from Aerial Imagery), that is composed of satellite images of the Polish territory. The data set consists of:

- RGB raster images in GeoTiff format with EPSG:2180 spatial reference system
- masks with a single channel in GeoTiff format with spatial reference system EPSG:2180

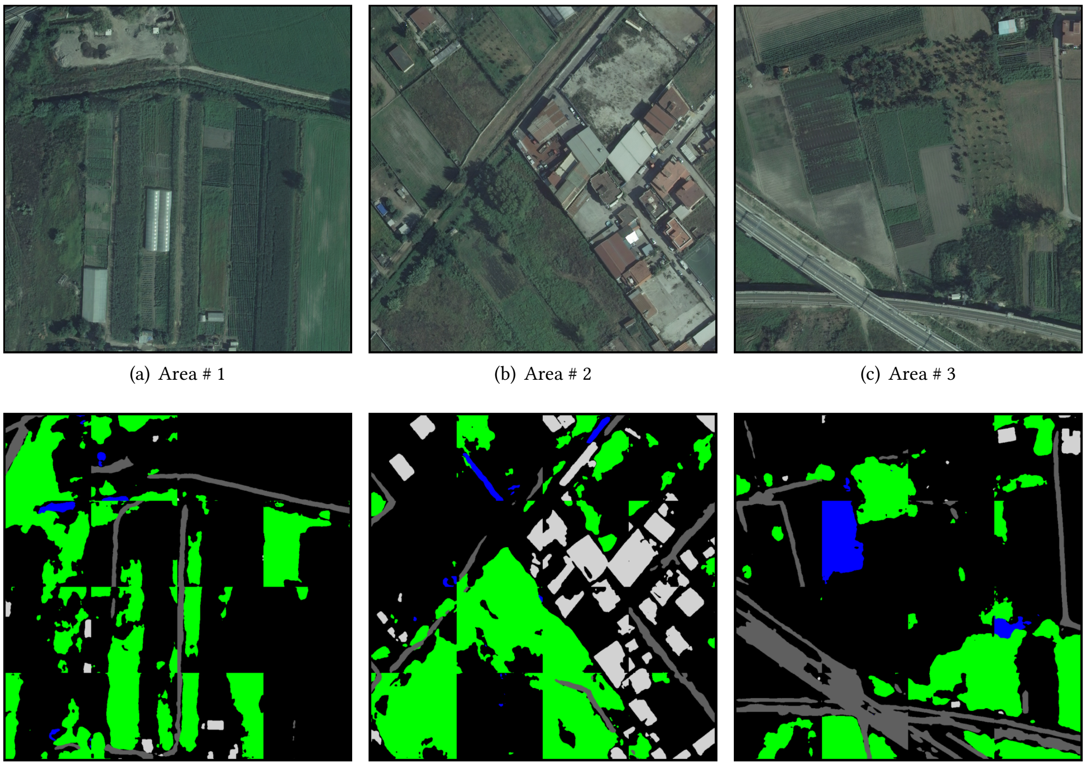
The data set provides a mapping of the areas represented in the images in five categories: reservoirs, forest, roads, buildings and other. The distribution of the classes within the data set is shown in Table 1.

[1]https://landcover.ai.linuxpolska.com

We processed images by dividing them into patches of size 256 × 256 with a resolution of 30 cm/pixel. Subsequently, all images that contain a total area of a target class (i.e., reservoirs, forest, roads and buildings) less than 5% were excluded. At the end of this phase, the training data set is composed of approximately 22.000 images representing all the main landscape scenarios of heterogeneous geographical areas. This data set was used for the development of territory segmentation model and it has been divided into train, validation and test for performance evaluation.

3.2. Proposed solution

We propose a segmentation-based approach which has been proved to be very effective for this kind of tasks. Image segmentation involves converting an image into a collection of pixel regions represented by a mask or labeled image. By dividing an image into segments, you can process only the important segments of the image instead of processing it entirely. We decide to use a UNet-based network (see Figure 1) that is a cutting-edge deep



(a) Area # 1 (b) Area # 2 (c) Area # 3

(d) Segmented Area # 1 (e) Segmented Area # 2 (f) Segmented Area # 3

Figure 2: In the top row are depicted three satellite images of Campania in 2018, whilst in the bottom row the results of the segmentation processing of the proposed solution.

convolutional neural network for segmentation tasks.

The UNet was designed to learn from few training samples and it does improve performance over existing Fully Convolutional Networks (FCNs).

The trained model consists of two components: the **backbone** model and the **final layer** of the UNet. The backbone model takes as input the image to be processed and it has the objective to extract key features from it. In fact, the goal is to extrapolate the semantic meaning from the image while preserving its spatial structure, even at a lower resolution. Typically the backbone is implemented as a standard architecture such as VGG, ResNet, Inception or MobileNet and by removing the last fully connected layer from them.

The feature extraction process is simplified and accelerated by starting from these architectures which are already pre-trained on specific data sets (e.g., ImageNet, COCO). In the proposed solution, the backbone selected is a ResNet-34 trained on ImageNet from which the final fully connected classification level (called backbone head) has been removed.

For training the model, we used a composite loss function that is specific for the segmentation task. In particular, it is a sum of two cost functions called *dice loss* (see Eq. 1) and *focal loss* (see Eq. 2). The dice loss measures the ratio between correctly classified pixels and the total number of predicted and real mask pixels.

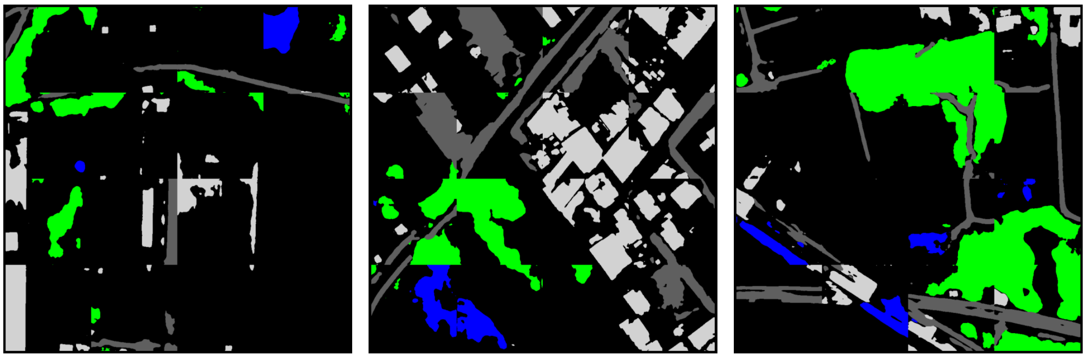
$$L_{dice} = \frac{2 * \sum p_{true} * p_{pred}}{\sum p_{true}^2 + p_{pred}^2 + \epsilon} \tag{1}$$

whilst the focal loss measures the degree of entropy - i.e. uncertainty - present in the classification process, scaling it by a factor γ capable of assigning, when estimating the value of the loss, greater weight to the more difficult examples to predict.

$$L_{focal} = -\sum_{i=1}^{N} (i - p_i)^{\gamma} \log_b(p_i) \tag{2}$$

where N represents the total number of classes in the data set.

(a) Area # 1 (b) Area # 2 (c) Area # 3



(d) Segmented Area # 1 (e) Segmented Area # 2 (f) Segmented Area # 3

Figure 3: In the top row are depicted the same three satellite images shown in Figure 2 but taken in 2020, whilst in the bottom row the results of the segmentation processing of the proposed solution.

4. Experimental Evaluation

The proposed solution has been evaluated by means of the following de-facto standard metrics.

4.1. Metrics

IoU (Intersection over Union). It is calculated by firstly computing the area of overlap between the predicted and ground-truth bounding boxes (i.e., the numerator) and then by computing the area of union which is the area encompassed by both the predicted and ground-truth bounding boxes (i.e., the denominator). The IoU (see Eq. 3) is then calculated by dividing the area of overlap by the area of union that yields our final score between 0 and 1. The higher the value is, the better is the model in segmenting the image.

$$IoU = \frac{P_{bbox} \cap GT_{bbox}}{P_{bbox} \cup GT_{bbox}} \tag{3}$$

F1-score. The F1-score (see Eq. 4.1) is the harmonic mean of the precision and recall. Precision refers to the number of true positives divided by the total number of positive predictions (i.e., the number of true positives plus the number of false positives). Recall, instead, refers to the number of true positives divided by the total objects of that class (i.e., the number of true positives plus the number of false negatives). The highest possible value of an F1-score is 1.0, indicating perfect precision and recall, and the lowest possible value is 0, if either precision or recall are zero.

$$F1\text{-}score = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

4.2. Experimental Setup

We performed the training on a workstation having the following characteristics: Precision 5820 Tower XCTO Base, Intel Xeon Processor W-2223 (4 Core @ 3.6 GHz with a Turbo of 3.9 GHz, HT, DDR42-63.25 @ 2.666 GHz), 64 GB RAM, Dual NVIDIA QUADRO RTX 6000 with 24 GB of memory, 256 GB PCIe NVMe Class 40 Solid State Drive and a 3.5 1 TB SATA HDD.

The training time needed for obtaining the final model was of six hours.

4.3. Performance Results

We computed the metrics presented in Section 4.1 on the final model obtaining the following performance results: **IoU score** of 0.654 and **F1-score** of 0.7433, which demonstrates promising results of the proposed solution on the LandCover.ai data set.

We then applied the final model on images coming from the Italian territory, more specifically the Campania region, for qualitative evaluating the time-lapse land monitoring performance of the proposed solution as shown in Figure 2 and 3. As it is shown in the pictures, the proposed solution correctly detects the reduction of vegetation between 2018 and 2020.

5. Conclusions

The aim of the proposed solution was to provide a framework for mapping the territory based on specific characteristics through the use of satellite images for detecting deforestation, reduction of reservoirs and glaciers.

The presented approach is based on a the image segmentation technique which is commonly used in digital image processing and analysis. The experimentation has been conducted on a publicly available data set for a quantitative evaluation and on images of Italian territory for a qualitative evaluation. The objective is to have a tool for:

- fighting against deforestation
- detecting the reduction of reservoirs
- detecting the reduction of glaciers

In this paper, we showed the promising performance of the solution on detecting the reduction of vegetation and reduction of reservoirs. A future work will be on testing the approach on a data set having also glaciers.

References

[1] S. Salcedo-Sanz, P. Ghamisi, M. Piles, M. Werner, L. Cuadra, A. Moreno-Martínez, E. Izquierdo-Verdiguier, J. Muñoz-Marí, A. Mosavi, G. Camps-Valls, Machine learning information fusion in earth observation: A comprehensive review of methods, applications and data sources, Information Fusion 63 (2020) 256–272.

[2] W. Shi, M. Zhang, R. Zhang, S. Chen, Z. Zhan, Change detection based on artificial intelligence: state-of-the-art and challenges, Remote Sensing 12 (2020) 1688.

[3] S. W. Chen, G. V. Nardari, E. S. Lee, C. Qu, X. Liu, R. A. F. Romero, V. Kumar, Sloam: Semantic lidar odometry and mapping for forest inventory, IEEE Robotics and Automation Letters 5 (2020) 612–619.

[4] X. X. Zhu, D. Tuia, L. Mou, G.-S. Xia, L. Zhang, F. Xu, F. Fraundorfer, Deep learning in remote sensing: a comprehensive review and list of resources, IEEE Geoscience and Remote Sensing Magazine 5 (2017) 8–36.

[5] T. Kattenborn, J. Leitloff, F. Schiefer, S. Hinz, Review on convolutional neural networks (cnn) in vegetation remote sensing, ISPRS Journal of Photogrammetry and Remote Sensing 173 (2021) 24–49.

[6] H. Mayfield, C. Smith, M. Gallagher, M. Hockings, Use of freely available data sets and machine learning methods in predicting deforestation, Environmental modelling & software 87 (2017) 17–28.

[7] T. Chang, B. P. Rasmussen, B. G. Dickson, L. J. Zachmann, Chimera: a multi-task recurrent convolutional neural network for forest classification and structural estimation, Remote Sensing 11 (2019) 768.

[8] E. E. Maeda, G. F. Arcoverde, P. K. Pellikka, Y. E. Shimabukuro, Fire risk assessment in the brazilian amazon using modis imagery and change vector analysis, Applied Geography 31 (2011) 76–84.

[9] T. N. Kehl, V. Todt, M. R. Veronez, S. C. Cazella, Real-time deforestation detection using ANN and satellite images: The Amazon Rainforest study case, Springer, 2015.

[10] A. Boguszewski, D. Batorski, N. Ziemba-Jankowska, T. Dziedzic, A. Zambrzycka, Landcover.ai: dataset for automatic mapping of buildings, woodlands, water and roads from aerial imagery, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2021, pp. 1102–1110.