A Dataset for Burned Area Delineation and Severity Estimation from Satellite Imagery

Original
A Dataset for Burned Area Delineation and Severity Estimation from Satellite Imagery / Colomba, Luca; Farasin, Alessandro; Monaco, Simone; Greco, Salvatore; Garza, Paolo; Apiletti, Daniele; Baralis, Elena; Cerquitelli, Tania. - ELETTRONICO. - (2022), pp. 3893-3897. (Intervento presentato al convegno International Conference on Information and Knowledge Management (CIKM) 2022 tenutosi a Atlanta (Georgia, USA) nel 17/10/2022 - 21/10/2022) [10.1145/3511808.3557528].

Availability:
This version is available at: 11583/2971026 since: 2022-10-27T13:32:44Z

Publisher:
ACM Press

Published
DOI:10.1145/3511808.3557528

Terms of use:
This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
ACM postprint/Author's Accepted Manuscript, con Copyr. autore

(Article begins on next page)
A Dataset for Burned Area Delineation and Severity Estimation from Satellite Imagery

Luca Colomba  
luca.colomba@polito.it  
Politecnico di Torino  
Torino, Italy

Alessandro Farasin  
alessandro.farasin@polito.it  
Politecnico di Torino  
Torino, Italy

Simone Monaco  
simone.monaco@polito.it  
Politecnico di Torino  
Torino, Italy

Salvatore Greco  
salvatore_greco@polito.it  
Politecnico di Torino  
Torino, Italy

Paolo Garza  
paolo.garza@polito.it  
Politecnico di Torino  
Torino, Italy

Daniele Apiletti  
daniele.apiletti@polito.it  
Politecnico di Torino  
Torino, Italy

Elena Baralis  
elena.baralis@polito.it  
Politecnico di Torino  
Torino, Italy

Tania Cerquitelli  
tania.cerquitelli@polito.it  
Politecnico di Torino  
Torino, Italy

ABSTRACT
The ability to correctly identify areas damaged by forest wildfires is essential to plan and monitor the restoration process and estimate the environmental damages after such catastrophic events. The wide availability of satellite data, combined with the recent development of machine learning and deep learning methodologies applied to the computer vision field, makes it extremely interesting to apply the aforementioned techniques to the field of automatic burned area detection. One of the main issues in such a context is the limited amount of labeled data, especially in the context of semantic segmentation. In this paper, we introduce a publicly available dataset for the burned area detection problem for semantic segmentation. The dataset contains 73 satellite images of different forests damaged by wildfires across Europe with a resolution of up to 10m per pixel. Data were collected from the Sentinel-2 L2A satellite mission and the target labels were generated from the Copernicus Emergency Management Service (EMS) annotations, with five different severity levels, ranging from undamaged to completely destroyed. Finally, we report the benchmark values obtained by applying a Convolutional Neural Network on the proposed dataset to address the burned area identification problem.

CCS CONCEPTS
• Computing methodologies → Image segmentation; Supervised learning; • Applied computing → Earth and atmospheric sciences.

KEYWORDS
earth observation, machine learning, deep learning

1 INTRODUCTION
Natural resources are limited and fundamental assets that must be preserved. Public entities and governments are constantly monitoring catastrophic events to preserve natural diversity and the environment, limiting as much as possible short- and long-term damages. Furthermore, once a catastrophic event such as a forest wildfire occurs, planning the post-event restoration process represents a fundamental activity to be held to contrast climate change. In this context, in-situ data acquisitions are often preceded by remote sensing, data gathering, and analyses performed jointly with computer vision techniques to help domain experts promptly identify areas of interest that require immediate support. Thanks to the great data availability and the development of deep learning algorithms, the Earth Observation (EO) domain represents an attractive research field in many different topics: disaster management and crisis response, agricultural monitoring, and change detection.

Natural hazards are a relevant cause of economic, humanitarian, and environmental losses. In the field of EO, change detection algorithms are largely adopted to identify land cover changes or natural calamities occurrences [3, 12, 18, 25, 26, 38], whereas other techniques and models are known to be well-suited for flood detection [9, 30, 37] and burned areas delineation [11, 23, 39].

Remote sensing enables domain experts to perform preliminary analyses and damage assessments immediately after the calamity ended, without the need for time-expensive in-situ analyses thanks to the availability of (open) data acquisitions with good resolutions, generally ranging from 10m to 30m [41]. Satellite images allow analyzing and characterizing extended areas for multiple purposes, such as aerosol characterization [5], air quality monitoring [42, 43], vegetation monitoring [8, 44].

The main difficulty, especially in the context of crisis management, is the limited availability of labeled data necessary to train deep learning models. In this paper, we introduce an open dataset that can be used for (i) the burned area delineation task and (ii) the damage severity estimation task. The first task consists in identifying the burned areas after a wildfire, while the second task focuses on also estimating the severity of the damage. In the dataset we collected, the severity level is an integer value ranging from 0 (undamaged area) to 4 (completely destroyed area). It was built leveraging on the information provided by Copernicus EMS [1] Rapid Mapping, Sentinel Hub service [2], and Sentinel satellite missions. The dataset spans over five different European countries, covering different soil types, with a total number of 73 regions at different points in time.
Information is gathered from Sentinel-1 and Sentinel-2 missions. To the best of our knowledge, no similar open dataset is available in the context of burned area delineation. The dataset is publicly available at https://doi.org/10.5281/zenodo.6597139 while the example code to load the data and apply a baseline deep learning model is available at https://github.com/lccol/burned-area-baseline.

Section 2 introduces the related work, Section 3 describes the dataset, the geographical distribution of the monitored areas, and the damage severity class distributions. Section 4 introduces some of the tasks that can benefit from the proposed dataset, while Section 5 provides a benchmark for the binary segmentation problem. Finally, Section 6 concludes the paper.

2 RELATED WORKS

In the Earth Observation domain, before the development of computer vision based on Convolutional Neural Networks, satellite imagery analysis was mainly based on spectral index computation and analysis. Considering the case of burned area delineation, different studies proposed an in-depth comparative analysis of different vegetation and burned area indexes [13, 32, 40] and automatic detection techniques to identify damaged regions based on thresholding [4, 14, 20, 27, 34]. Furthermore, beyond the burned area delineation problem, it is extremely important the ability to estimate severity levels to assess damage to vegetation in case of widespread wildfires. Such a problem is often tackled by analyzing the relatedness between different burned area indexes and the Composite Burned Area Index (CBI) [7, 19, 31, 36], which is computed by experts after in-situ data measurements and thus is extremely time-expensive to obtain. Additionally, the robustness of calibrated methodologies on a specific area to estimate the severity levels from remote sensing data is not assured to produce good results when applied to new regions [7]. Over the years, several methodologies have been proposed based on supervised learning algorithms [10, 11, 21, 33] to tackle both the delineation and severity estimation problem. The proposed supervised approaches have proven to be well-performing and require fewer data preprocessing. However, the main drawback of such approaches is the need for a great amount of data, possibly spanning over multiple regions and countries. Despite multiple methodologies being proposed in the literature, many researchers analyzed forest fires within a restricted number of countries [6, 15, 17], making it difficult to quantitatively compare different methodologies. Furthermore, the data used are rarely made public.

3 DATASET

Motivated by the lack of labeled datasets representative of different countries, we created a raster dataset consisting of different burned areas spread across Europe. For each of the monitored wildfires, data were retrieved from the Sentinel-2 L2A and Sentinel-1 GRD missions in a timespan of 2 months given the activation date of each wildfire provided by Copernicus EMS: 1 month prior and 1 month after the activation date. More specifically, Sentinel-1 data was acquired in IW mode with VV+VH polarisation. The images within the defined temporal range have been selected to (i) be available for at least the 90% of the desired AoI defined by the delineation, and (ii) cloud coverage must not exceed 10% of the acquisition.

![Figure 1: Sentinel-2 RGB acquisition with missing information (left) and its corresponding validity mask (right).](image)

The dataset was collected starting from the Rapid Mapping products publicly released by Copernicus EMS, from which the semi-automatic annotations are collected for each wildfire event and the ground truth labels for our semantic segmentation task are generated. In the following paragraphs, the data retrieval process is described, as well as the main characteristics of the dataset and the class distribution.

3.1 Data retrieval

The dataset was collected from the information provided by the products released by the Rapid Mapping team at Copernicus EMS, considering the forest wildfire type of events only. Such vector data completely describes the affected regions, providing useful information also in the land cover domain, such as the presence of roads, buildings, and agricultural areas. The vector data of interest for burned area delineation is the following: bounding box and delineation of the areas damaged by the wildfire with the associated severity level.

Given such coordinates and the activation date of a wildfire, we relied on SentinelHub [2] to download the Sentinel-2 L2A and Sentinel-1 GRD data available one month before and one month after the activation date. Each downloaded product available in the dataset is associated with a prefix, a date, and a validity mask (coverage), with the prefix being the product name itself (e.g., sentinel2), the date being the point in time in which the product was acquired from the satellite, and the validity mask (stored as the last channel of the product itself) being a binary mask indicating whether a pixel value is valid or not (1 for valid, 0 for invalid pixel). Invalid pixels are returned in case no valid data was found by the SentinelHub service at a certain date in the requested area of interest. An example of a Sentinel-2 product with missing information and corresponding coverage mask is shown in Figure 1.

We downloaded the satellite imagery with the highest possible resolution, up to a maximum of 5000x5000 pixels. We decided to download the data available one month before and one month after the activation date to have information about the pre- and post-wildfire events, which is needed if we are interested in applying change detection algorithms. Lastly, the downloaded products were manually analyzed to determine whether they contained useful information for the prediction task, i.e., images where the area of interest was covered with clouds or where data was absent were discarded.

To generate the ground truth mask necessary for the semantic segmentation task, we leveraged the vector data of Copernicus annotations: the coordinates of the area of interest and the bounding box. Such information was used to produce grayscale images
Figure 2: Forest wildfires event distribution in the dataset.

Figure 3: Pre-fire (top) and post-fire (bottom) samples with different morphological features and lighting conditions.

represents pixel-by-pixel damage severity levels. Ground truth masks have the same resolution as the downloaded products.

3.2 Dataset characteristics

The dataset is composed of 73 different forest wildfires, from 2017 to 2019, spread across Europe (Figure 2), for a total area of $\sim 19.000 \text{km}^2$ with different morphological features and terrain types. Domain experts may leverage Sentinel information to perform preliminary analyses, data cleaning, and lighting condition normalization to improve the robustness of the proposed solutions and assess the performance of machine learning-based methodologies. The areas of interest are mainly covered by forests. In some acquisitions, human settlements and buildings are also present. Some examples are shown in Figure 3.

Diversity in geographical locations and terrain types underlines the importance of developing new methodologies compared to the simplest threshold-based classification, which suffers from the difficulty of finding a unique threshold and a unique index valid for every region worldwide [13]. As a motivating example, we consider the NBR2 burned area index, defined as follows on Sentinel-2 data: $\text{NBR2} = \frac{B_{11} - B_{12}}{B_{11} + B_{12}}$, where $B_N$ stands for band $N$ of Sentinel-2. Figure 4 represents the NBR2 index distribution of two different burned areas: one in Sweden (depicted with blue curves), one in Italy, near the volcanic area of Vesuvio (depicted with red curves). In such cases, the definition of a single threshold to classify burned areas, which is valid for both cases of the NBR2 index is not a trivial task. This high-level analysis shows the diversity in the collected data and, hence, its potential usefulness to assess the quality of different burned area detection techniques under different conditions.

Figure 4: Burnt vs Unburnt pixel distribution for NBR2 index. Darker curves represent undamaged areas, lighter curves represent burnt areas instead.

3.3 Severity classes

Annotations released by Copernicus specify 5 different severity levels, which were encoded in a greyscale image in range $[0 – 255]$, from no damage to completely destroyed area.

Such classes are highly imbalanced: unburnt areas are the 91.9% of the whole images. Focusing on burnt areas only, class 1 covers around the 12%, while class 2 to 4 coverages stay around 30% (1%, 2.25%, 2.35%, and 2.5% of the total, respectively for the 4 classes).

4 TASKS

In the experimental section, we focus on the benchmarking of one specific segmentation problem: the burned area delineation task based on post-fire imagery only. However, our dataset can be a valuable resource for benchmarking several tasks.

Some of the tasks that can be performed on our dataset are:

- Burned area delineation based on post-wildfire imagery only and supervised techniques. In this scenario, a machine learning algorithm is trained on the post-wildfire images only and is used to predict for each pixel of the satellite acquisition if it represents a burned or an unburned area.

- Damage severity estimation based on post-wildfire imagery only and supervised techniques. This task is similar to the previous one. However, the damage severity level is predicted for each pixel of the analyzed post-wildfire images.

- Burned area delineation based on pre- and post-wildfire images and supervised or unsupervised techniques. For the unsupervised case, change detection techniques can be used to identify burned areas by comparing pre- and post-wildfire images. Labeled data can be used to either assess the performance of the proposed unsupervised approaches or train and assess the quality of supervised methods.

- Comparison of different spectral indexes and evaluation of separability by means of separability index [22, 28].
5 EXPERIMENTS

To assess the complexity and quality of the proposed dataset and to provide reference performance values, we report the prediction results for a baseline model. Considering the semantic segmentation task, we chose the UNet [35] architecture for our evaluation and compared it to the Otsu method [29] on NBR2. It demonstrated state-of-the-art performances in many segmentation tasks [16, 24], including burned area delineation [11]. The experiments were performed using only post-fire Sentinel-2 acquisitions.

**Data pre-processing** Being the satellite images in the proposed dataset of variable pixel resolution, we decided to generate 512x512 resolution tiles from the data and perform a random shuffle before training. The total number of tiles is 957. Partial overlap was admitted only when the original image was not a multiple of size 512 pixels. In this case, the second-to-last and last tiles overlap. Data were split into 7 folds according to geographical location, such that data belonging to the same fold maintains similar morphological features.

**Data augmentation** During the training phase, random data augmentation methodologies were applied to improve the performance and robustness of the model. We applied the following transformations: rotation (up to 50° on both sides), shear (up to 20°), horizontal flip, and vertical flip, all of them with a probability of 50%.

**Training configuration** We used a system with 128GB of RAM and an NVIDIA Tesla V100 with 16GB. The training was performed using 50 epochs, the Adam optimizer with a learning rate of 0.0001 with no weight decay. The Dice loss was used for training. To evaluate the performance, a 7-fold cross-validation approach was chosen: 5 folds were used as a training set, 1 fold as a validation set, and 1 fold as a test set. We implemented an early stopping mechanism with patience of 5 epochs and tolerance of 0.01 on validation loss. The final evaluation is performed on the test set.

**Experimental results** Table 1 shows the results achieved by the UNet architecture and the Otsu method in terms of accuracy, precision, recall, and F1-score on every test fold. The UNet achieves good results in the majority of the folds, being the lime fold the worst performing one. Such fold contains acquisitions from wildfires that took place in Southern Italy, including two volcanic areas (Etna and Vesuvius). The low performance is mainly due to the difficulty of correctly identifying burned areas in such aforementioned volcanic areas and by the presence of small burned regions. Three examples are shown in Figure 5. The first two examples show two areas near a volcano, with a high overestimation of burned areas. The last example represents an area with different small burned regions that are not correctly identified by the model. Otsu achieves low results, confirming the low quality of threshold-based techniques.

### Table 1: Performance for each test fold. Values are expressed in %.

<table>
<thead>
<tr>
<th>Test fold</th>
<th>Accuracy UNet</th>
<th>Otsu</th>
<th>Precision UNet</th>
<th>Otsu</th>
<th>Recall UNet</th>
<th>Otsu</th>
<th>F1-score UNet</th>
<th>Otsu</th>
</tr>
</thead>
<tbody>
<tr>
<td>coral</td>
<td>92.0</td>
<td>75.3</td>
<td>82.9</td>
<td>62.6</td>
<td>97.2</td>
<td>73.9</td>
<td>89.5</td>
<td>67.8</td>
</tr>
<tr>
<td>cyan</td>
<td>91.4</td>
<td>82.3</td>
<td>79.0</td>
<td>55.8</td>
<td>80.4</td>
<td>76.1</td>
<td>79.7</td>
<td>64.4</td>
</tr>
<tr>
<td>grey</td>
<td>95.0</td>
<td>73.6</td>
<td>70.4</td>
<td>27.5</td>
<td>97.3</td>
<td>80.3</td>
<td>81.7</td>
<td>41.0</td>
</tr>
<tr>
<td>lime</td>
<td>92.2</td>
<td>59.0</td>
<td>35.6</td>
<td>8.4</td>
<td>86.9</td>
<td>80.1</td>
<td>50.6</td>
<td>15.1</td>
</tr>
<tr>
<td>magenta</td>
<td>94.3</td>
<td>79.4</td>
<td>80.1</td>
<td>51.8</td>
<td>98.5</td>
<td>88.9</td>
<td>88.3</td>
<td>65.5</td>
</tr>
<tr>
<td>pink</td>
<td>96.4</td>
<td>79.9</td>
<td>84.8</td>
<td>45.3</td>
<td>97.4</td>
<td>65.3</td>
<td>90.7</td>
<td>53.5</td>
</tr>
<tr>
<td>purple</td>
<td>97.5</td>
<td>66.3</td>
<td>86.1</td>
<td>23.5</td>
<td>93.0</td>
<td>90.5</td>
<td>89.4</td>
<td>37.4</td>
</tr>
</tbody>
</table>

Mean 94.1 73.7 74.1 39.3 93.0 79.3 81.4 49.2
Median 94.3 75.3 80.1 45.3 97.2 80.1 88.3 53.5
Std dev 0.02 0.08 0.18 0.20 0.07 0.09 0.14 0.19

![Figure 5: Samples obtained from lime fold: RGB acquisition (top), ground truth (center) and prediction (bottom).](image)

We believe this dataset can be beneficial to researchers and public authorities for many tasks, such as recovery planning, constant monitoring of affected areas, and the development of deep learning models for severity estimation. The dataset is made publicly available to encourage future use and research activities.

### REFERENCES


