Magnifier: A Multigrained Neural Network-Based Architecture for Burned Area Delineation

Magnifier: A Multigrained Neural Network-Based Architecture for Burned Area Delineation

Daniele Rege Cambrin, Luca Colomba, and Paolo Garza

***Abstract*—In crisis management and remote sensing, image segmentation plays a crucial role, enabling tasks like disaster response and emergency planning by analyzing visual data. Neural networks are able to analyze satellite acquisitions and determine which areas were affected by a catastrophic event. The problem in their development in this context is the data scarcity and the lack of extensive benchmark datasets, limiting the capabilities of training large neural network models. In this article, we propose a novel methodology, namely Magnifier, to improve segmentation performance with limited data availability. The Magnifier methodology is applicable to any existing encoder–decoder architecture, as it extends a model by merging information at different contextual levels through a dual-encoder approach: a local and global encoder. Magnifier analyzes the input data twice using the dual-encoder approach. In particular, the local and global encoders extract information from the same input at different granularities. This allows Magnifier to extract more information than the other approaches given the same set of input images. Magnifier improves the quality of the results of $+2.65\%$ on average intersection over union while leading to a restrained increase in terms of the number of trainable parameters compared to the original model. We evaluated our proposed approach with state-of-the-art burned area segmentation models, demonstrating, on average, comparable or better performances in less than half of the giga floating point operations per second (GFLOPs).**

***Index Terms*—Deep learning, earth observation (EO), natural hazard management, postwildfire segmentation, semantic segmentation.**

I. Introduction

FORESTS represent a fundamental resource for the environment, society, and the entire global ecosystem. In recent years, an increase in the number of forest wildfires has been observed due to several different causes, such as human activities, climate change, and extreme weather conditions. It is estimated to grow in the next decades [1], [2], [3], endangering the ecosystem and leading to environmental damage and economic and humanitarian losses. Since the restoration process of damaged areas takes several years, losses are observed over the entire time the ecosystem requires to regenerate, leading to collateral damages such as a higher risk of landslides. Identifying the afflicted areas correctly and planning the postwildfire restoration process are relevant activities for natural hazard management. Such a process can be supported by data availability through high-resolution sensors mounted on aircrafts and satellites, which can acquire information at a continental scale. This, adopted in conjunction with the most recent development of machine learning and deep learning techniques, makes the Earth Observation (EO) field particularly interesting to researchers and domain experts to develop automatic monitoring techniques via remote sensing. Such methodologies support local authorities and could improve the handling of catastrophic events and natural hazards. Deep learning models have revolutionized computer vision and image analysis in different application domains, including remote sensing. A few examples of such applications in the remote sensing domain are wildfire severity estimation [4], [5] and delineation [6], [7], [8], [9], flooding detection [10], air quality estimation [11], land cover [12], [13], [14], and scene classification [15], [16].

In this article, we tackle the problem of burned area delineation (i.e., identifying areas previously affected by forest wildfires) using data collected by the Sentinel-2 and Landsat-8 missions' satellites and deep learning models. Such data and models enable the possibility to provide first postwildfire damage assessments quickly, compared to the time required by domain experts to identify damaged regions manually or semiautomatically.

The previously proposed deep learning-based approaches for burned area delineation (e.g., [6], [7], [8], [9]) apply general-purpose semantic segmentation models to solve this task. Different state-of-the-art (SOTA) deep learning architectures have been used, including convolutional neural networks (CNNs) (e.g., U-Net [17] or DeepLabV3+ [18]) and vision transformers (e.g., SegFormer [19]). Fig. 1 reports the average intersection over union (IoU) achieved by such standard semantic segmentation architectures and those of our proposed solution, called Magnifier, applied on three open-access datasets. The results of general-purpose SOTA deep-learning semantic segmentation models are quite satisfactory, particularly when CNN-based architectures are used. However, an uncommon behavior is highlighted in Fig. 1. Unlike in other semantic segmentation tasks [18], [19], [20], increasing the model size, considering the same architecture type, does not increase or negligibly improves the quality of the results (the results of two models of different sizes for each architecture type are reported in Fig. 1). This is probably because the available labeled datasets are not sufficiently large to train the larger models, which sometimes start overfitting, impacting the predictions (e.g., consider SegFormer B1 vs. SegFormer B0 in Fig. 1). Hence, further enhancing the results can be achieved only by increasing the amount of labeled

Received 29 August 2023; revised 22 October 2024 and 14 January 2025; accepted 19 April 2025. Date of publication 30 April 2025; date of current version 20 May 2025. *(Corresponding author: Daniele Rege Cambrin.)*

The authors are with Politecnico di Torino, 10129 Turin, Italy (e-mail: daniele.regecambrin@polito.it).

Digital Object Identifier 10.1109/JSTARS.2025.3565819

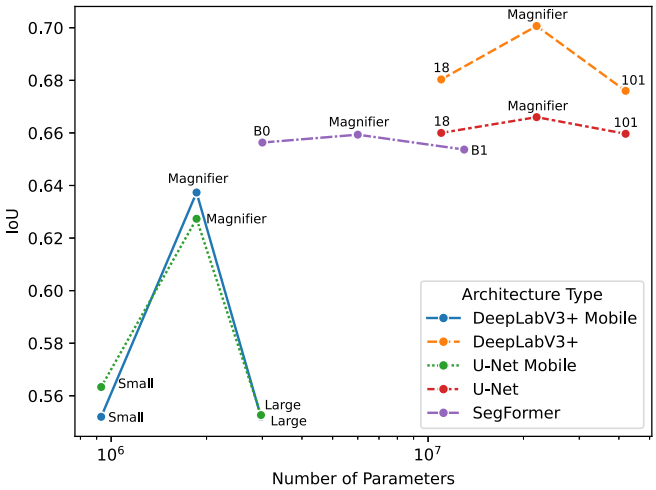

Fig. 1. Average mean IoU versus number of parameters. The Average Mean IoU has been computed considering all datasets. *Architecture Type* refers to the family of base networks employed for segmentation. On average, the *Magnifier* backbone achieves IoU improvements compared to MobileNetV3 Small and Large, ResNet-18 and 101, and MiT-B0 and B1 without increasing the number of parameters too much.

data, which is costly and frequently impracticable, or proposing a different approach to better use the already available labeled data. Also, we would like to point out that the largest models cannot be used in some scenarios due to resource constraints (e.g., on embedded devices). Hence, new architectures able to increase the quality of the results by limiting the size of the models without requiring more labeled data are helpful. To address such issues, we propose a multigrained neural network architecture, namely Magnifier, which combines information and features computed at different levels of detail (contextual levels) to perform better than SOTA architectures 1) using the *same amount of labeled data* and 2) *smaller neural networks*.

Fig. 1 shows how Magnifier, which can be applied on top of well-known deep learning networks, achieves better IoU than SOTA architectures without the need to use significantly larger models in terms of parameters.

The code of Magnifier is publicly available at https://github.com/DarthReca/magnifier-california.

Our contributions can be summarized as follows:

1) We propose a new technique to increase segmentation performance without collecting more labeled data.
2) We propose a versatile technique/architecture that can leverage many different encoder–decoder deep-learning models without much effort.
3) We compared the effectiveness of the proposed technique with previous well-known SOTA convolutional and transformer-based architectures for semantic segmentation.

II. Related Works

This section covers the works related to deep learning, semantic segmentation, and their application in EO and burned area delineation.

A. Deep Learning

Deep learning is a subset of machine learning that uses artificial neural networks with multiple layers to model and understand complex patterns in data [21]. The primary innovation of neural networks and deep learning is the ability to automatically learn hierarchical feature representations from raw data, eliminating the need for manual feature engineering. These models, especially CNNs [22], recurrent neural networks (RNNs) [23], [24], and Transformer [25] architectures, can process vast amounts of data and extract meaningful patterns at multiple levels of abstraction. CNNs, for example, are particularly well-suited for image processing tasks, while RNNs and Transformers are suited for sequential data like time series and language. More recently, the latter architecture demonstrated its superior performances also in the computer vision domain [26] in the presence of a large amount of training data. The hierarchical feature representation learning performed by neural networks is formulated as an optimization problem: given the predicted outcomes by the model and the ground truth information, a loss function is evaluated and minimized throughout the training process [21]. Deep learning has gained widespread attention because it successfully outperforms traditional machine learning algorithms in complex tasks. This success is largely attributed to the availability of large datasets, advances in computational power (particularly GPUs), and innovations in neural network architectures that enable these models to handle more diverse and intricate data types. It has become a foundational technology for researchers in different fields, such as computer vision [20], natural language processing [27], and reinforcement learning [28].

B. Semantic Segmentation

Semantic segmentation is a fundamental task in computer vision and image processing that involves classifying each pixel in an image based on its semantic meaning.

Classical methods in semantic segmentation have laid the foundation for developing more sophisticated techniques. For example, the usage of histogram analysis [29], graph partition techniques [30], and mean shift [31] provided promising results in segmenting with a certain number of limitations until the advent of deep learning.

The employment of CNNs revolutionized the computer vision field by solving various tasks [32] and providing outstanding results in dealing with complex segmentation problems. Networks such as the multiple versions of DeepLab [33], [34] and UNet [17] have proven to be quite capable in many different fields, ranging from medical to satellite imagery [35].

The last advancements in computer vision propose new vision transformer architectures such as Swin Transformer [36] and SegFormer [19], which provide new SOTA results in semantic segmentation benchmarks.

While well-known encoders like ResNets [37] and SegFormer [19] create representations at different resolution levels, they do not exploit different contextual levels. Some other works acted directly onto the input image resolution without changing also, in this case, the granularity of the context [38]. As described

in the following paragraphs, our work distinguishes itself against the latter due to the presence of multiple encoders to handle different information granularity, whereas the aforementioned paper adopts a Siamese U-Net model to analyze the original input image and a downsampled version (i.e., duplicated information is seen four times).

The Magnifier architecture we propose in the article is built on top of these established neural networks. It effectively combines models applied to images at different "granularities" to increase their performances without increasing the number of labeled input images, which would require time and resources. Since satellite imagery covers vast areas and contains coarse and fine details, combining global and local views leads to more informative embeddings without requiring more labeled data and larger models.

C. Earth Observation

Recently, researchers started investigating the applicability of SOTA machine learning models to hyperspectral data (e.g., satellite data) in the remote sensing domain. In fact, considering the field of EO, several tasks can be formulated as common machine learning problems, such as land cover classification [39], [40], crop classification [41], [42], [43], image segmentation [44], [45], [46], and visual question answering [47]. Over the years, the amount of labeled data has risen, enabling the possibility of training better and larger machine learning models. This latter point, combined with a greater availability of computational resources, led to the improvement of deep learning methodologies in many different fields, including computer vision and hyperspectral image analysis. Despite this, focusing on the context of resource and emergency management as well as disaster recovery, the data availability problem still persists due to the rarity of such events. As a consequence, this problem requires researchers to develop well-performing machine learning methodologies with a limited amount of data. Even with such a limit, machine learning methodologies demonstrated superior performances compared to traditional techniques. In [10], tree-based and UNet algorithms surpassed index-based analyses of flooded areas. Similarly, Montello et al. [48] tested several different architectures in flood segmentation and compared them against Otsu, whereas Farasin ety al. [5] evaluated the aforementioned neural network architecture against the dNBR index in the burned area delineation problem.

D. Burned Area Delineation

1) Burned Area Delineation Employing Spectral Indices: The burned area delineation problem is a well-known problem in the field of EO. Several works tackle the problem by analyzing the spectral signature of the affected areas (e.g., [49], [50], [51]). Domain experts developed spectral indices, also known as burned area indices, which are sensitive to vegetation and humidity to distinguish damaged and undamaged areas given a satellite acquisition. Such indices are computed by different combinations of spectral bands, obtaining a single-channel image that can highlight areas affected by the catastrophic event under consideration. This is possible because different bandwidths collect information strictly related to vegetation, making identifying the areas of interest feasible. Examples of such indices are BAIS2 [52], normalized burn ratio [53], and normalized burn ratio 2 [54]. The resulting index is consequently analyzed for solving image segmentation by applying threshold-based techniques. One of the main issues of the proposed approaches is that, given a burned area index, identifying a unique threshold valid worldwide, which adequately segments burned areas, is difficult due to different lighting conditions, vegetation, terrain characteristics, and morphological features [55]. Identifying the threshold for a single region is also a nontrivial task. For this reason, automatic deep-learning models are more effective and preferred.

2) Burned Area Delineation Employing Deep Learning Models: More recently, with the development of modern computer vision techniques based on the adoption of deep learning methodologies, researchers started the investigation and the application of neural networks also in the field of EO, demonstrating SOTA performances [56], [57]. Several deep models are currently adopted, demonstrating their effectiveness in analyzing satellite imagery and processing multispectral acquisitions. More specifically, different researchers successfully applied CNNs [58], [59] and transformer-based architectures [60] to the burned area delineation problem. From the latter work, UNet-based models demonstrated SOTA performances in the burned area identification domain [61]. Similarly, Seydi et al. [62] extended the U-Net architecture by enhancing the convolutional blocks in both the encoder and decoder part of the network with multipatch multilevel residual morphological blocks (MP-MRM), preserving the skip connections typical of such segmentation model. Internally, the MP-MRM blocks leverage skip-connections, erosion, and dilation operators with a quadratic structuring function, as well as traditional convolutional layers, to solve the burned area delineation problem. The architectures designed so far rely on a significant amount of data (collected from different sources, in some cases [6]) and use large general-purpose semantic segmentation models. Solutions based on many labeled data, multiple sources, or computationally expensive models could be challenging to scale to real applications. We propose a multigrained neural network architecture to reduce complexity and computational costs and improve the quality of the results without requiring more labeled data and using "small-size" models. Using the proposed architecture shows how smaller models can be competitive against more complex ones using the same amount of input data. Good performances combined with the necessity of fewer resources can boost the use of deep learning-based architectures in the EO domain.

III. Materials and Method

In this section, we formally present the problem of burned area delineation, the dataset used for the study, the employed architecture, and finally, the loss function for the neural network training and the metrics.

TABLE I
CHARACTERISTICS OF THE USED DATASETS

	CaBuAr [63]	Europe [5]	Indonesia [64]
Resolution	20m	10m	30m
Channels	12	12	8
Forest Fires	340	73	81
Start-End date	Jan, 2015 – Dec, 2022	July, 2017 – July, 2019	Jan, 2019 – Dec, 2021
Burned surface	$\sim$ 11000 km^2	$\sim$ 2000 km^2	$\sim$7000 km^2
Number of images	688	449	227

TABLE II
SENTINEL-2 SPECTRAL BANDS

Band	Resolution	Central wavelength	Description
B1	60 m	443 nm	Ultra Blue (Coastal and Aerosol)
B2	10 m	490 nm	Blue
B3	10 m	560 nm	Green
B4	10 m	665 nm	Red
B5	20 m	705 nm	Visible and Near Infrared (VNIR)
B6	20 m	740 nm	Visible and Near Infrared (VNIR)
B7	20 m	783 nm	Visible and Near Infrared (VNIR)
B8	10 m	842 nm	Visible and Near Infrared (VNIR)
B8a	20 m	865 nm	Visible and Near Infrared (VNIR)
B9	60 m	940 nm	Short Wave Infrared (SWIR)
B10	60 m	1375 nm	Short Wave Infrared (SWIR)
B11	20 m	1610 nm	Short Wave Infrared (SWIR)
B12	20 m	2190 nm	Short Wave Infrared (SWIR)

TABLE III
LANDSAT-8 SPECTRAL BANDS

Band	Resolution	Central wavelength	Description
B1	30 m	440 nm	Ultra Blue (Coastal and Aerosol)
B2	30 m	480 nm	Blue
B3	30 m	560 nm	Green
B4	30 m	650 nm	Red
B5	30 m	860 nm	Near Infrared (NIR)
B6	30 m	1610 nm	Short Wave Infrared (SWIR)
B7	30 m	2200 nm	Short Wave Infrared (SWIR)
B8	15 m	590 nm	Panchromatic
B9	30 m	1370 nm	Cirrus
B10	100 m	10900 nm	Thermal Infrared
B11	100 m	12000 nm	Thermal Infrared

A. Problem Statement

This work addresses the burned area segmentation problem based on postfire images. The task can be formalized as follows.

Let I be an arbitrary satellite image of size $W \times H \times D$, where W and H are the width and height of the images in pixels, respectively, while D is the depth of the images (i.e., the number of features per pixel). The objective is to automatically create a binary mask M of size $W \times H$ associated with I, where the value 1 indicates the associated pixel contains a burned area. In contrast, 0 is related to unburned pixels.

To solve the problem, we have at our disposal a set S_L of labeled images (with the same features of I) for which the associated masks (with the same characteristics of M) are known. The labeled set of images S_L is used to train a semantic segmentation model SM. After the training, SM can be used to predict the mask M of an arbitrary (new) image I for which the associated mask is unknown.

B. Datasets

We evaluated our methodology on three publicly available datasets spanning three different areas of the globe: California[1] [63], Europe[2] [5], and Indonesia[3] [64]. See Table I for an overview of their characteristics.

Both California and Europe datasets use the Sentinel-2 mission of the European Space Agency (ESA). It comprises a pair of identical satellites orbiting Sun-synchronously at an average height of 786 km. The mission was designed to contribute to Land Monitoring, Emergency Response, and Security services. Each satellite is equipped with the Multispectral Instrument (MSI) with 13 spectral bands spanning from the visible and the near-infrared to the short wave infrared with resolution from 10 to 60 m, as shown in Table II. Many of them are sensitive to vegetation, permitting the monitoring of green areas of the globe. The field of view is 290 km. They provide a mean global revisiting time smaller than five days [65].

ESA makes available to final users L1C and L2A products. These products comprise ortho-images in UTM/WGS84 projection, each covering a surface area of 110 $\times$ 110 km^2 . L2A products provide atmospherically corrected Surface Reflectance (SR) images derived from the associated Level-1C products. The atmospheric correction process involves several essential steps to account for various atmospheric effects. These steps include correcting for the scattering of air molecules, known as Rayleigh scattering, and compensating for the absorbing and scattering effects caused by atmospheric gases (e.g., ozone, oxygen, and water vapor) and aerosol particles [66]. This process results in the omission of band 10 from the final product, while L1C comprises all available spectral bands.

The Indonesia dataset uses the Landsat-8 mission of the National Aeronautics and Space Administration (NASA). It is comprised of a satellite equipped with the Operation Land Imager (OLI) and the Thermal Infrared Sensor (TIRS). OLI captures nine visible and infrared spectral bands with a resolution of 30 m, while TIRS has two thermal infrared bands at 100 m resolution as detailed in Table III. The revisit time is 16 days. It aims to develop a scientific understanding of the Earth system [67]. NASA makes L1GT, L1T, L1TP, and L2SR products available to final users. These products comprise ortho-images in UTM/WGS84 projection, each covering a surface area of 185 $\times$ 185 km^2. L2SR products provide atmospherically corrected SR images derived from the associated L1T or L1TP products. Hence, this dataset is composed of satellite images/input products with bands and characteristics different from those of the other two datasets. These differences do not impact the applicability and quality of our methodology.

1) CaBuAr Dataset: The CaBuAr (California Burned Areas) dataset [63] comprises L2A products of Sentinel-2 and ground truth masks provided by the California Department of Forestry and Fire Protection. In Fig. 3(a), it is possible to look at the geographical distribution of the analyzed wildfires over California. The dataset is composed of 340 wildfires that happened between January 2015 and December 2022. The total burned area surface is more than 11 000 km^2 . The post-fire acquisition

[1]Available at https://huggingface.co/datasets/DarthReca/california_burned_areas

[2]Available at https://zenodo.org/record/6597139

[3]Available at https://data.mendeley.com/datasets/fs7mtkg2wk

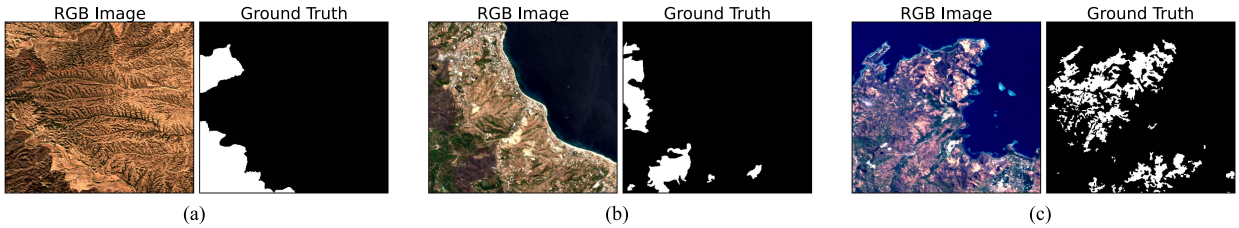

Fig. 2. RGB samples taken from the three datasets with the corresponding binary ground truth. (a) California. (b) Europe. (c) Indonesia.

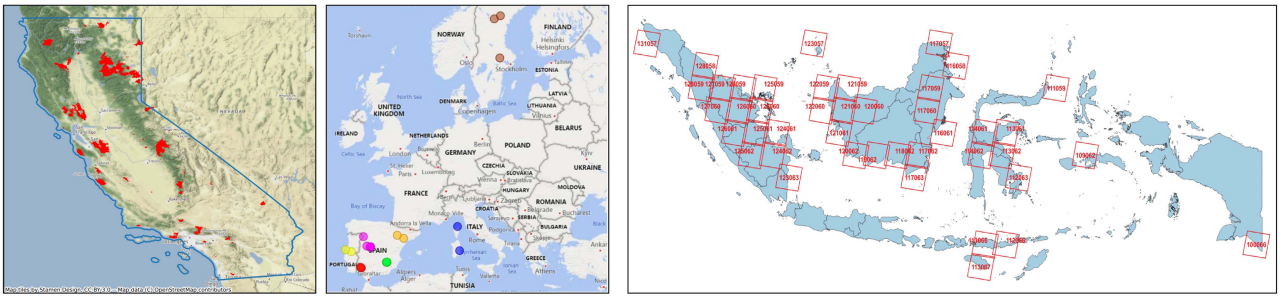

Fig. 3. Distribution of wildfires in the analyzed datasets. In (a) the areas covered by wildfires in *red*. In (b), the location of the wildfires, where each color represents the fold it belongs to. In (c), the areas covered by the dataset are in red squares.

is downloaded with at most a month delay after the wildfire containment date. If necessary, the bands were upsampled or downsampled to 20 m resolution to make the matrix tractable by a neural network. The dataset comprises 688 images of size $512 \times 512 \times 12$ divided into five folds. A complete overview of the characteristics can be seen in Table I and a sample can be seen in Fig. 2(a).

2) Europe Dataset: The Europe dataset [5] comprises L2A products of Sentinel-2 and Sentinel-1 and ground truth provided by the Copernicus Emergency Management System. In Fig. 3(b), it is possible to look at the geographical distribution of the considered wildfires. The dataset is composed of 73 wildfires that happened between July 2017 and July 2019. The total burned area is more than 2000 km^2 . The timeframe between the containment date and data acquisition is not specified, although the delineation maps were acquired after the wildfire event. All the bands are upsampled to 10 m resolution when necessary. The dataset comprises 449 images of size $512 \times 512 \times 12$, divided into seven folds. Its main characteristics are reported in Table I and a sample can be seen in Fig. 2(b).

3) Indonesia Dataset: The Indonesia dataset [64] consists of 227 manually annotated satellite images derived from Landsat-8 of size $512 \times 512 \times 8$ pixels. Four experts annotated the data, and three other experts evaluated the annotations. In Fig. 3(c), we show the area covered by the dataset. The dataset is composed of wildfires between 2019 and 2021. The total burned area is more than 7000 km^2. The timeframe between the containment date and data acquisition is not specified, although the delineation maps were acquired after the wildfire event. A complete characterization is reported in Table I and a sample can be seen in Fig. 2(c).

The three datasets cover different areas and allow us to analyze the general applicability of the proposed architecture in various contexts. Although CaBuAr focuses only one state (California), it covers the largest area ($\sim$ 11000 km^2 of the burned surface), followed by Indonesia ($\sim$ 7000 km^2) and Europe ($\sim$ 2000 km^2) (see Table I). Furthermore, California is characterized by various and complex geology, from high mountains to broad valleys.

C. Neural Network Architecture

The main research question addressed by this article is whether or not it is beneficial to a neural network-based architecture to fuse information extracted from different patch sizes extracted from SL, consequently providing merged contextual information at different granularity to the decoder of an encoder–decoder deep neural network, to perform per-pixel classification (i.e., semantic segmentation). For this reason, we propose an architecture based on two "paths" associated with patches of different sizes, extracted from the same image collection, that are fused before being fed to the decoder. The architecture is quite versatile since it can be applied to various well-known segmentation models in the literature. The main requirement is the model has to be based on an encoder–decoder structure (i.e., encoders and decoders of SOTA encoder–decoder semantic segmentation models can be used in Magnifier).

The Magnifier architecture is presented in Fig. 4. Here, we present its general architecture, which is independent of the encoders and decoders used to implement it. The details of the encoders and decoders used/considered to instantiate the Magnifier are reported in the following sections.

The input of the Magnifier architecture consists of size $W \times H \times C$ images. The encoder part comprises two branches, each associated with a different encoder and patch size. Finally, a single common decoder is trained to perform the predictions after a fusion step. One encoder receives as input the original

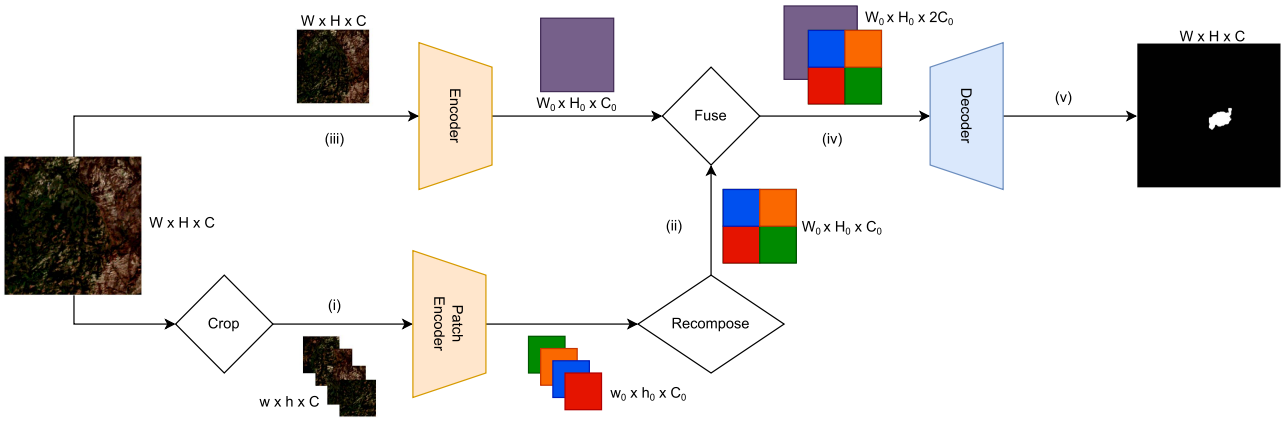

Fig. 4. Magnifier architecture. In the lower branch, (i) the image is cropped in smaller patches (as shown in Fig. 5), giving each patch to an encoder. (ii) The encodings are concatenated by putting each one in the original position in the image (as shown in Fig. 6). In the upper branch (iii), the entire image is given to an encoder. (iv) The two encodings are concatenated along the channel axis, and (v) they are given to the decoder to get the final prediction.

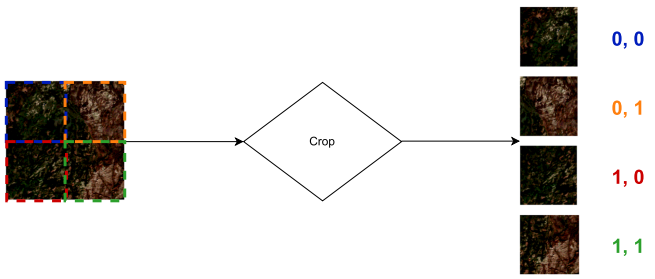

Fig. 5. Cropping procedure. The image is cropped in patches, and each of them keeps the original position associated.

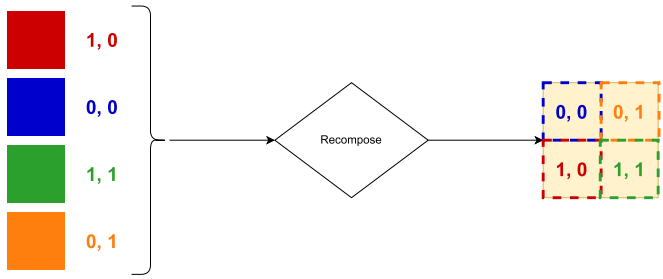

Fig. 6. Recompose procedure. The encodings of an image are merged into a single embedding matrix using the position information.

("coarse-grained") images of size $W \times H \times C$, while the other one (called *Patch Encoder*) accepts ("fine-grained") images of size $w \times h \times C$, where $w < W$ and $h < H$. We chose nonoverlapped patches for simplicity, with w and h sub-multiples of W and H, respectively. The two encoders share the same model (e.g., a ResNet encoder) but not the weights. The first path is used to learn "global" information, while the second learns "local" knowledge.

The input images are split into smaller patches to feed the second path. Specifically, given an input image I of size $W \times H \times C$, I is split/cropped into several smaller patches $P = \{P_0, P_1 \ldots\}$ of size $w \times h \times C$. Each patch is coupled with the information about its position in the original image as shown in Fig. 5. For this procedure, the original input image I can be seen as a grid of size $N \times M$ composed of $w \times h \times C$ patches. So, the information about the position of each patch in the original image I can be stored as a tuple $(\text{ROW}, \text{COLUMN})$, where $\text{ROW} \in \{0, \ldots, M-1\}$ and $\text{COLUMN} \in \{0, \ldots, N-1\}$. The position $(0, 0)$ was fixed at the top-left corner. ROW increases from left to right, and COLUMN from top to bottom.

The original input image I is embedded into a feature vector of size $W_0 \times H_0 \times C_0$ with the first encoder (top branch in Fig. 4). Parallelly, each one of the patches $P_i \in P$ is embedded into a feature vector of size $w_0 \times h_0 \times C_0$ with the *Patch Encoder* (bottom branch in Fig. 4).

The feature vectors associated with the patches P_i are then concatenated into a global feature vector as shown in Fig. 6, preserving the original order of the patches. For instance, a patch associated with position $(0, 0)$ has its encoding placed in position $(0, 0)$ of the embedding matrix. In this way, we obtain another embedding of size $W_0 \times H_0 \times C_0$ from the second branch of Magnifier.

Consequently, the two feature vectors of sizes $W_0 \times H_0 \times C_0$, extracted by the two encoders, are fused by concatenation along the channel axis (C_0). The resulting embedding is provided as input to the shared decoder to obtain the final prediction.

The Magnifier architecture is general and can be applied on top of different encoder–decoder deep learning models. For instance, the encoder and decoder of a CNN can be used, or those of a vision transformer. More details about which SOTA deep learning encoders and decoders have been used to instantiate the Magnifier and test it are reported in Section IV-A. The source code of Magnifier is publicly available at https://github.com/DarthReca/magnifier-california.

D. Loss Function

We employed the Asymmetric Unified Focal (AUF) loss [68], being the problem highly imbalanced. This loss combines cross-entropy-based and dice-based losses in a single formulation with few hyper-parameters. The asymmetric version of the loss

removes the focal suppression to the rare class, giving it more importance.

The formulation of AUF loss (L_{AUF}) is:

$$L_{\text{AUF}} = \lambda L_{\text{maF}} + (1-\lambda) L_{\text{maFT}} \tag{1}$$

where $\lambda \in [0,1]$ determine the weights of the two components of the loss, which are modified asymmetric Focal loss (L_{maF}) and modified asymmetric Focal Twersky loss (L_{maFT}).

The modified asymmetric Focal loss (L_{maF}) is defined as

$$L_{\text{maF}} = -\frac{\delta}{N} y_{i:r} \log p_{t,r} - \frac{1-\delta}{N} \sum_{c \neq r} (1-p_{t,c})^{\gamma} \log p_{t,r} \tag{2}$$

where N is the number of pixels, $p_{t,r}$ and $p_{t,c}$ are the probability the sample belongs to the rarest and most common classes, respectively, $y_{i:r}$ is the ground truth label for the sample belonging to the rare class, δ is the weighting factor of the components of the loss and γ is the suppression factor for the most common class.

The formulation of modified asymmetric Focal Twersky loss (L_{maFT}) is

$$L_{\text{maFT}} = \sum_{c \neq r} (1-\text{mTI}) + \sum_{c=r} (1-\text{mTI})^{1-\gamma} \tag{3}$$

where mTI is the modified Twersky Index, and γ is the enhancement factor applied to the rare class.

The modified Twersky Index is defined as

$$\text{mTI} = \frac{\sum_i^N p_{r,i} g_{r,i}}{\sum_i^N p_{r,i} g_{r,i} + \delta \sum_i^N p_{r,i} g_{c,i} + (1-\delta) \sum_i^N p_{c,i} g_{r,i}} \tag{4}$$

where $p_{r,i}$ and $p_{c,i}$ are the probability of a pixel belonging to the rarest and most common classes, respectively, and $g_{r,i}$ and $g_{c,i}$ are the ground truth values for the sample belonging to the rare and most common classes, respectively.

IV. Experiments

In our experiments, we consider as base model architectures, on top of which Magnifier is applied, two different families of CNN architectures (U-Net [17] and DeepLabV3+ [18]) and a vision transformer (SegFormer [19]). Using different base model architectures (i.e., using various types of encoders and decoders in Magnifier – Fig. 4), we show the effectiveness of the Magnifier approach independently of the models used to instantiate its encoders and decoder.

To evaluate the contribution of the Magnifier architecture, we compare the "single models" with the "magnified" versions. The single models are trained on the original single ("coarse") grained images in S_L, as commonly done in previous works [5], [6]. On the one hand, this allows comparing the magnifier with SOTA methods trained on single-grained Sentinel-2 images. On the other hand, this can be seen as an ablation study that analyzes the impact of the second branch, which is associated with the "fine-grained"/smaller patches, on the prediction quality. Section IV-A presents the base segmentation model architectures, and Section IV-B introduces the encoder used in each architecture as a feature extractor. Section IV-C analyzes the existing competitors, Section IV-D describes the metrics used for the quantitative comparison, and Section IV-E reports the experimental settings. Section IV-F provides the experimental results of the evaluation campaign. Finally, Section IV-G reports a transfer learning analysis between different areas.

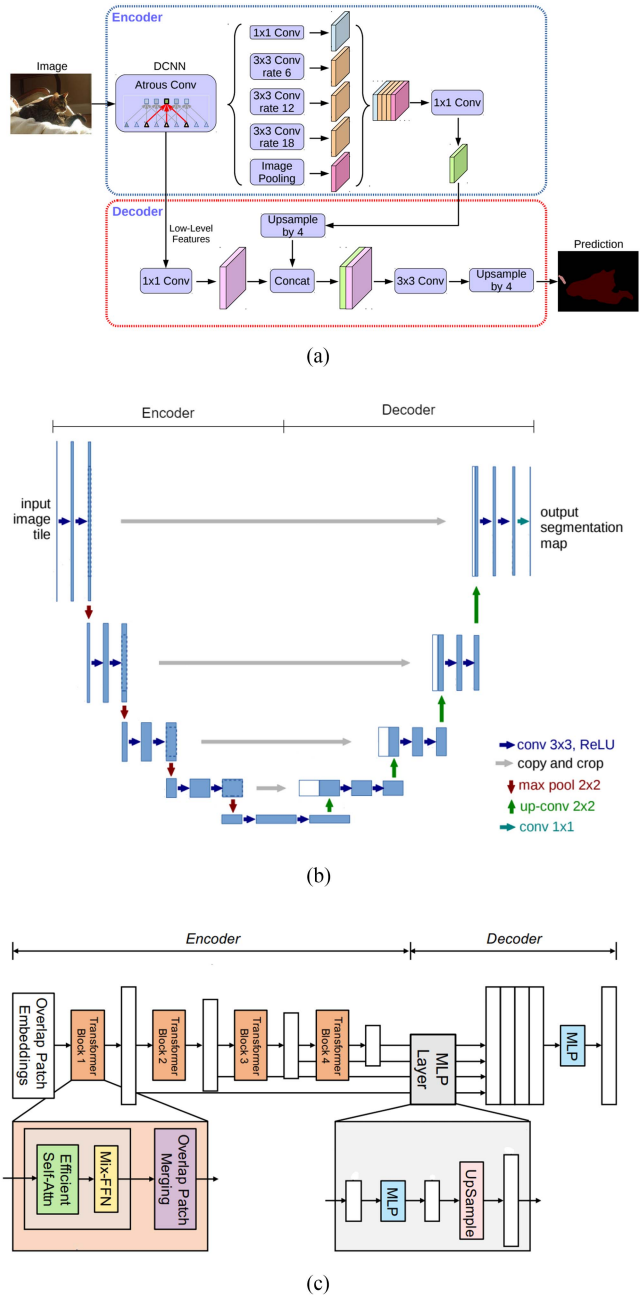

Fig. 7. Base architectures of DeepLabV3+, U-Net and SegFormer. (a)DeepLabV3+ [34]. (b) U-Net [17]. (c) SegFormer [19].

A. Base Model Architectures

The chosen base model architectures, as highlighted in Fig. 7, are different in their structure and technological choices. However, they are all based on two distinct, nonoverlapping parts:

an encoder and a decoder. We selected three of the most effective and well-known semantic segmentation architectures. Brief descriptions of these architectures follow.

1) DeepLabV3+: The encoder module of DeepLabV3+ [18] incorporates multiscale contextual information using the spatial pyramid pooling module and atrous separable convolution, a combination of atrous convolution and depthwise separable convolution [see Fig. 7(a)]. The decoder module is designed to recover lost spatial information during encoding.

2) U-Net: The encoder of U-Net [17], called the contracting path, reduces the dimensionality of the input. In contrast, the decoder, called the expansive path, gradually increases the dimensionality of the feature maps [see Fig. 7(b)]. Using skip connections, the encoder extracts features of different spatial resolutions concatenated with the embeddings at the same level in the decoder.

3) Segformer: The encoder of SegFormer [19] is positional-encoding-free, hierarchical, and composed of transformer blocks, which output multiscale features. The decoder comprises MLPs that aggregate local and global attention information to predict the final mask [see Fig. 7(c)].

These are three of the best-performing and most frequently used models for solving the semantic segmentation problem, including the burned area delineation/segmentation task.

For DeepLabV3+ and U-Net architectures, we evaluated two different backbones designed for devices with different resources (see Section IV-B). Instead, for the SegFormer model, we considered the standard encoder [19] proposed in the original paper, since it is already adaptable to different resource constraints.

B. Backbone Selection

In our experiments, we have chosen well-known backbones available in different "versions." This means all model versions share the same architecture and only differ in the number of parameters. We selected two versions for each backbone. They are referred to as "small" and "large" versions in the following. The Magnifier architecture has been applied only to the small models. The large models are used only to compare the Magnifier with larger single models.

In Table IV, it is possible to see the number of parameters for the various backbones. As expected, since Magnifier is based on two branches and hence two backbones, its encoders always have a number of parameters that are twice the parameters of the small single models on top of which it is applied. However, each Magnifier model is always characterized by fewer parameters than the "large" version of the corresponding single model (approximately 50% of the parameters). The head size is not reported because all models in a family share the same head.

We applied two well-known families of backbones to DeepLabV3+ and U-Net (the two CNN-based architectures we considered). Specifically, *MobileNetV3* [20], a SOTA backbone for low-resource devices, and *ResNet* [37], which is a SOTA solution for many tasks. We evaluated a single transformer-based backbone, *MiT*, because it is easy to adapt, has many different versions, and is the one proposed in the SOTA SegFormer transformer.

1) MobileNetV3: Mobile Nets are encoders designed to work in low-resource devices. They use Squeeze-and-Excite blocks, depthwise separable convolution, and hard-swish to reduce the computational cost without decreasing the global accuracy. The introduction of Squeeze-and-Excite permits the emulation of the attention layers at a lower cost by applying average pooling and two convolutions. The encoder is mainly composed of 3×3 and 5×5 bottleneck layers [20].

2) Resnet: ResNet is designed to overcome the issues regarding exploding and vanishing gradients. Adding shortcut connections enables computed features to reach deeper levels of the network. This permits the creation of very deep networks without losing training stability and increasing the model's accuracy. The network mainly comprises 3×3 convolutional layers, bottleneck layers, and shortcut connections [37].

3) Mit: Mix Transformer (MiT) is the hierarchical encoder proposed in SegFormer. It is composed of Transformer blocks, each one containing self-attention, mix-feed forward network (Mix-FNN), and overlap patch merging (used to generate CNN-like multilevel embeddings providing high-resolution coarse features and low-resolution fine-grained features). This encoder combines convolution and MLP of the Mix-FFN with the mechanism of self-attention, which is formulated with particular consideration to the efficiency [19].

C. Competitors

The current SOTA deep learning-based models for burned area delineation using satellite images are based on "single-grained" images associated with a traditional semantic segmentation model (e.g., [9], [35]). Hence, as competitors, we have considered single SOTA deep-learning models for semantic segmentation (both CNNs and vision transformers). We also included index-based methodologies with the Otsu segmentation method [29]. We selected a custom SOTA architecture (BurnNet) as the baseline algorithm to compare Magnifier's performances.

D. Evaluation Metrics

We report weighted mean F1 Score and IoU for the positive class (burned area) and the relative standard deviation computed across the folds since the number of samples per fold varies significantly according to the original settings ($\approx 64 \pm 32$ for Europe and $\approx 106 \pm 13$ for California). The F1-Score is more adequate for imbalanced classes, while IoU provides a good segmentation quality assessment. They are formulated as follows:

$$F1 = \frac{2TP}{2TP + FP + FN} \tag{5a}$$

$$\text{IoU} = \frac{TP}{TP + FP + FN} \tag{5b}$$

where TP, TN, FP, and FN are the number of True Positive, True Negative, False Positive, and False Negative predictions, respectively.

TABLE IV
RESULTS FOR THE DIFFERENT MODELS ON CALIFORNIA AND EUROPE DATASETS

				California		Europe		Indonesia		
Backbone type	Backbone	↓ GFLOPs	↓ BS	F1	IoU	F1	IoU	F1	IoU	MR
MobileNetV3	Small	6.0	0.93M	64.8 ± 7.9	48.4 ± 8.5	72.6 ± 15.2	59.1 ± 17.6	73.3 ± 6.1	58.1 ± 7.6	2.7
	Magnifier	8.6	1.86M	**69.7 ± 8.1**	**54.1 ± 9.2**	**79.7 ± 10.5**	**67.5 ± 13.9**	**80.2 ± 3.0**	**69.6 ± 4.3**	**1**
	Large	9.4	2.97M	60.5 ±1 2.2	44.4 ± 11.3	74.0 ± 13.1	60.3 ± 15.8	75.5 ± 4.9	60.9 ± 6.2	2.3
ResNet	18	40.2	11M	73.8 ± 10.2	59.4 ± 12.1	83.6 ± 8.2	**72.7 ± 11.8**	83.7 ± 2.8	72.0 ± 4.1	2.2
	Magnifier	76.9	22M	**77.8 ± 6.4**	**64.0 ± 8.0**	**83.7 ± 7.6**	**72.7 ± 11.0**	**84.7 ± 2.2**	**73.5 ± 3.4**	**1**
	101	115.6	42M	76.0 ±1 0.2	62.3 ± 12.0	81.9 ± 8 3	70.3 ± 12.0	82.4 ± 3.9	70.2 ± 5.5	2.7

(a) DeepLabV3+

				California		Europe		Indonesia		
Backbone type	Backbone	↓ GFLOPs	↓ BS	F1	IoU	F1	IoU	F1	IoU	MR
MobileNetV3	Small	20.7	0.93M	62.6 ± 14.4	47.0 ± 13.9	75.9 ±1 0.9	62.3 ± 13.6	74.6 ± 5.1	59.7 ± 6.4	2.3
	Magnifier	25.5	1.86M	**67.3 ± 10.9**	**51.6 ± 11.6**	**79.1 ± 9.6**	**66.4 ± 12.5**	**82.4 ± 4.0**	**70.2 ± 5.8**	**1**
	Large	24.7	2.97M	66.3 ± 5.8	49.9 ± 6.3	73.5 ± 17.2	60.8 ± 19.8	70.6 ± 8.5	55.1 ±1 0.4	2.7
ResNet	18	47.0	11M	73.2 ± 5.3	58.0 ± 6.3	**82.1 ± 7.6**	**70.4 ± 10.7**	82.0 ± 3.1	69.6 ± 4.4	2
	Magnifier	78.1	22M	**74.4 ± 5.7**	59.6 ± 6.9	81.0 ± 9.9	69.2 ± 13.3	**82.9 ± 4.0**	**71.0 ± 5.8**	**1.5**
	101	127.6	42M	73.5 ± 15.0	**60.0 ± 16.6**	80.8 ± 8.7	68.6 ± 12.1	81.8 ± 3.9	69.3 ± 5.6	2.5

(b) U-Net

				California		Europe		Indonesia		
Backbone type	Backbone	↓ GFLOPs	↓ BS	F1	IoU	F1	IoU	F1	IoU	MR
MiT	B0	16.0	3M	**71.7 ± 8.6**	56.5 ± 9.7	81.8 ± 10.5	70.4 ± 14.3	82.3 ± 3.0	70.0 ± 4.4	2.2
	Magnifier	21.3	6M	71.5 ± 12.0	**56.9 ± 13.2**	**82.5 ± 8.0**	71.0 ± 11.2	82.2 ± 2.4	69.9 ± 3.5	2
	B1	31.4	13M	69.0 ± 11.8	53.7 ± 12.4	82.4 ± 10.4	**71.4 ± 14.5**	**83.0 ± 2.4**	**71.0 ± 3.5**	**1.8**

(c) SegFormer

	California		Europe		Indonesia		
Index	F1	IoU	F1	IoU	F1	IoU	MR
NBR	15.0 ± 23.1	10.3 ± 18	44.0 ± 18.3	29.7 ± 14.9	18.8 ± 3.1	10.4 ± 1.9	2.3
NBR2	**22.6 ± 26.9**	**15.9 ± 20.9**	**49.2 ± 19.3**	**34.4 ± 16.1**	**30.1 ± 8.3**	**17.9 ± 6.0**	**1**
BAIS	4.0 ± 12.1	2.6 ± 8.6	17.6 ± 13.7	10.3 ± 9.0	8.8 ±1 .0	4.6 ± 0.6	3.7
BAIS2	19.4 ± 29.2	14.8 ± 25.2	28.5 ± 15.9	17.5 ± 10.8	-	-	3

(d) Indexes

		Number of parameters		California		Europe		Indonesia		
Model	↓ GFLOPs	↓ BS	↓ Size	F1	IoU	F1	IoU	F1	IoU	MR
NBR2	-	-	-	22.6 ± 26.9	15.9 ± 20.9	49.2± 1 9.3	34.4 ± 16.1	30.1 ± 8.3	17.9 ± 6.0	4.7
Magnifier DeepLabV3+-RN18	76.9	22M	24.2M	**77.8 ± 6.4**	**64.0 ± 8.0**	83.7 ± 7.6	72.7 ± 11.0	**84.7 ± 2.2**	**73.5 ± 3.4**	**1.3**
Magnifier UNet-RN18	78.1	22M	27.5M	74.4 ± 5.7	59.6 ± 6.9	81.0 ± 9.9	69.2 ± 13.3	82.9 ± 4.0	71.0 ± 5.8	2.8
Magnifier SegFormer-B0	21.3	6M	7.19M	71.5 ± 12.0	56.9 ± 13.2	82.5 ± 8.0	71.0 ± 11.2	82.2 ± 2.4	69.9 ± 3.5	3.3
BurntNet	219.0	14.5M	35M	71.6 ± 32.6	62.2 ± 29.6	**84.4 ± 8.2**	**73.7 ± 11.7**	-	-	2.8

(e) Comparisons

Each table refers to a specific architecture, and results are grouped by the backbone type and sorted by size (in terms of the number of parameters BS). We have not reported the head size since it is constant for each model. We reported the mean rank (MR) for each group.

We also report the Mean Rank (MR) for each algorithm (model) across the datasets to better evaluate the differences between the analyzed methods [69]. MR is formulated as follows:

$$\text{MR} = \frac{1}{md}\sum_{i=1}^{m}\sum_{j=1}^{d} r_{ji} \tag{6}$$

where r_{ji} is the rank (where 1 is assigned to the best) of the algorithm (model) among all algorithms for a given metric i on a dataset j, m is the total number of metrics, and d is the number of datasets. MR ranges from 1 to the number of compared algorithms/models. The smaller the MR value, the better the algorithm/model.

We evaluate the resource consumption of the networks expressed in GFLOPs. GFLOPs measure the number of mathematical operations a system is capable of performing per second and are often used to evaluate the computational demands of networks. A higher number of GFLOPs indicates a more complex model that requires more resources, such as CPU or GPU, to be processed and trained effectively.

E. Experimental Settings

Since the pretrains are not typical for such a type of image (with 12 or 8 channels), we randomly initialized the networks. We used the AdamW optimizer and a polynomial learning rate scheduler with a power of 1 for 55 iterations. SegFormers are trained with a starting learning rate of 0.001, UNets with 0.0001, ResNet-DeepLabV3+s with 0.01, and Mobile-DeepLabV3+s with 0.0001. According to the original formulation, the hyperparameters of AUF loss are $\lambda = 0.5, \delta = 0.6, \gamma = 0.1$. We kept the same configuration of the learning rate scheduler and AUF loss for BurntNet with a learning rate of 0.0001. A patch size of 64×64 was adopted for the smaller crops in Magnifier. All models were trained on a single Tesla V100 32GB GPU.

We applied five-fold cross-validation on the CaBuAr dataset and seven-fold cross-validation on the European one as in the original evaluation procedure [5], [63]. For the Indonesian dataset, we applied the five-fold cross-validation approach since no settings or train-test splitting were provided in the original paper. This means for K times, where K is the number of folds, we choose a fold for validation and another for testing, while the remaining ones are used for the model's training [70]. This statistical method provides a more robust evaluation of classifiers than the standard way of creating a single training–validation–testing split.

The solution presented in [6] combines many different data sources using an ensemble of U-Net models, each trained on one data source. Since we are interested in methods based only on a single source, we do not consider [6] in our experimental comparison.

F. Results

In Table IV, we present the results for the three considered architectures (see Section IV-A) and the different backbones (see Section IV-B) on the three datasets. We would like to remind you that we trained one Magnifier model for each combination (base model architecture, backbone type), considering the smallest backbone version for each backbone type (i.e.; ResNet-18, MobileNetV3-Small, and MiT-B0). Hence, five different models based on the Magnifier approach have been trained for each dataset (see Table IV).

1) Overall Magnifier Performances: We initially analyze the results reported in Table IV to identify the best-performing model independently of the used base model architecture and backbone. DeepLabV3+ with Magnifier-ResNet18 achieves the highest F1 score (77.8 on CaBuAr, 83.7 on Europe and 80.2 on Indonesia) and IoU (64.0 on CaBuAr, 72.7 on Europe, 73.5 on Indonesia) on all datasets [see Table IV(a)]. The improvement in terms of F1 score compared to the second-best-performing competitor is higher on the CaBuAr dataset (+1.8% with respect to DeepLabV3+ with ResNet101) and on the Indonesia one (+1.5% compared to DeepLabV3+ with ResNet18). In comparison, it is only +0.1% with respect to DeepLabV3+ with ResNet18 on the European dataset. Similar improvements are achieved in terms of IoU.

DeepLabV3+ with Magnifier-ResNet18 is always the best-performing model, independently of the dataset and evaluation metric. Conversely, the second-best-performing model among the competitors varies for the datasets (DeepLabV3+ with ResNet101 on CaBuAr, and DeepLabV3+ with ResNet18 on Europe and Indonesia). DeepLabV3+ equipped with ResNet18, the smallest version of considered ResNet, probably performs better than its larger version (ResNet101) in Europe and Indonesia due to the size of the datasets. They contain fewer samples than CaBuAr. Hence, more data may be needed to train the larger model.

Overall, the use of the combination (DeepLabV3+, ResNet), with or without the application of the magnifier approach, is the best choice. The MR of this model, when compared to other solutions [see Table IV(e)], confirms this claim.

Since the MobileNetV3 backbone is smaller than ResNet, MobileNetV3 performs worse than ResNet, independently of the model size and base model architecture.

The SegFormer-based models achieve intermediate results compared to the DeepLabeV3+-based and U-Net-based ones.

2) Neural Networks Versus Index-Based Segmentation: Otsu's automatic thresholding applied to several indexes performs poorly in the burned area delineation problem. We validated each index on a per-fold basis despite not requiring any training procedures, such that the final scores can be compared against the ML-based methodologies. BAIS2 was not applied to the Indonesia dataset, since it exploits spectral bands available in Sentinel-2. NBR2 index achieved the overall best results on all datasets with a MR of 1, followed by NBR [see Table IV(d)]. As can be seen from Table IV(a) and (b), MobileNetV3-based variants of DeepLab and U-Net are the worst-performing deep learning models on all datasets. However, those deep learning models also outperform by a large margin the traditional index-based methodology, motivating the use of deep neural networks.

3) Magnifier Versus Traditional Deep Models: We now analyze the improvement of applying the magnifier approach on top of traditional models to understand its impact. We recall that the smallest version of each conventional backbone (i.e.; ResNet-18, MobileNetv3-Small, and MiT-B0) is used as an encoder in the two branches of the magnifier.

When using DeepLabV3+ with the MobileNetV3-S backbone, magnifier boosts the F1 score and IoU performances on all datasets [see the first two lines of Table IV(a)]. The F1 score increases from 64.8 to 69.7 on CaBuAr (+4.9%), from 72.6 to 79.7 on Europe (+7.1%), and from 73.3 to 80.2 on Indonesia (+6.9%). Slightly higher improvements are achieved regarding the IoU metric (+5.7% on CaBuAr, +8.4% on Europe and 11.5% on Indonesia).

The improvements are similar when using the DeepLabV3+ with the ResNet18 backbone applied to the CaBuAr dataset (+4.0% in terms of F1 score and +4.6% in IoU). Conversely, the improvements given by Magnifier are close to zero on the European dataset. In Indonesia, we still get +1.0% in F1 and +1.5% in IoU.

Looking at the combination (U-Net, MobileNetV3-S) in Table IV(b), magnifier confirms its boost compared to the single model (U-Net with MobileNetV3-S) on all metrics and datasets (+4.7% in terms F1 score and +4.6% in IoU on CaBuAr, +3.2% in F1 score and +4.1% in IoU on Europe, and +7.8% in F1 and

+10.5% in IoU on Indonesia). In this case, similarly to the previous configuration, the improvement given by using Magnifier on the CaBuAr and Indonesia datasets is slightly less evident when considering the ResNet18 backbone and worsens on the European dataset. However, it is still the best solution according to the MR (1.5 compared to 2 of ResNet-18 alone).

Regarding the results obtained with SegFormer [see Table IV(c)], Magnifier performs slightly better than the single model (SegFormer with MiT-B0) on Europe (+0.7% in terms of F1 score and +0.6% in IoU). On the CaBuAr dataset, the Magnifier improves only the IoU metric (+0.4%). In Indonesia, Segformer-B1 remains the best-performing solution.

The use of the Magnifier approach proves effective, as summarized by the MRs. Magnifier gets the best MR in 4 out of 5 configurations (see Table IV). In the single failure case in terms of MR [the one based on SegFormer – Table IV(c)], all the configurations achieve similar ranks, probably due to the limitations of the transformer backbone itself.

The reported results confirm the positive impact of the magnifier approach compared to single models. Without increasing the number of labeled data, the combination of multigrained versions of the images introduced in magnifier boosts the F1 and IoU metrics on average. Moreover, in addition to the advantage in terms of segmentation performance, magnifier shows superiority also in terms of the amount of computational resources, with lower values of FLOPs compared to the enhanced, larger versions of each model. The only exception is observed for the U-Net model and MobileNetV3 backbone, where the magnifier model shows a slightly higher value of a number of operations. This is mainly due to the high optimization of the MobileNetV3 model and the low amount of layers being present: the duplication of the encoder component rivals the number of parameters of the larger version.

The experiments described in this section can also be seen as an ablation study in which only the first path of the magnifier is used. The results show that an additional branch with the same input images but at a different "granularity" can boost the model's performance.

4) Magnifier Versus Large Single Models: We finally performed experiments training single models based on larger versions of the three considered backbones (MobileNetV3-L, ResNet101, and MiT-B1). This experiment aims to understand if the improvements achieved by Magnifier compared to the smaller models (see Section IV-F3) are due to the increase in the number of parameters or to the use of images of different sizes and granularities (i.e., to the use of the magnifier approach). The large backbones considered in these experiments are characterized by approximately twice the parameters of the magnifier models and four times those of the small single models.

Considering all the combinations (base model architecture, large backbone) and all datasets, magnifier, trained on the small versions of the backbones, performs better than the large single models in terms of F1 score 14 times out of 15 cases. In comparison, it performs better than the large models in terms of IoU in 11 cases over 15 [see Table IV(a) to (c)]. The MRs highlight large models perform worse than smaller models in many cases: increasing the parameters is not the best solution with a limited amount of data due to overfitting. SegFormer-B1 is the only large model version that achieves a better MR compared to the small and magnifier MiT-based models, but the MRs are quite similar, probably due to the large amount of data that transformers need to learn relations [26]. However, we recall that the MiT-based models are not the best ones overall.

These results support the claim that the improvement provided by magnifier is related to its two paths and images analyzed at different levels and not to the number of parameters of the models. The innovative usage of the labeled data improves the quality of the trained models.

5) Magnifier Versus BurntNet: Table IV(e) summarizes the results achieved by the overall best models previously presented, as well as the performance achieved by the BurntNet segmentation model. Focusing on the California dataset and the machine learning-based approaches, it can be observed that BurntNet achieves the overall worst performance, surpassing only SegFormer B0 with magnifier by a low margin. Comparing the backbone and the total size, the SegFormer-based magnifier architecture has less than half the number of trainable parameters and almost a fifth of the total number of parameters compared to BurntNet, not justifying the higher complexity of the latter. The highest performance gap in F1 and IoU scores observed in this set of experiments is +6.2 and +1.8 points.

Instead, focusing on the Europe dataset, BurntNet showed higher performance compared to all magnifier-enhanced models, with the largest difference being +3.4 and +4.5 points in F1 and IoU scores, respectively. Considering the average results, Magnifier applied to DeepLabV3+ with ResNet18 backbone achieved the overall best result in both the evaluation metrics. The performance gap on the second dataset can be justified by 1) the larger overall dimension of the ResNet18 backbone, which might lead to overfitting issues, and 2) the higher difficulty due to transfer learning since the seven folds adopted in the cross-validation process are region-based. In fact, the SegFormer-based model is able to surpass the U-Net model despite the lower number of parameters. Moreover, the adoption of dilation and erosion mechanisms in the BurntNet architecture enables noise removal and better shape extraction in a context in which the input dataset is smaller and noisier compared to the California dataset.

When looking at Indonesia, BurntNet collapsed, probably due to the few samples in the dataset.

Overall, average scores demonstrate the better results of the Magnifier applied to DeepLabV3+ architecture, as well as better stability during the training procedure against BurntNet, which showed a model collapse in a few of the experimental runs, leading to invalid predictions.

Analyzing instead the number of floating point operations (GFLOPs), it can be seen that BurntNet is the least performing model, with 219 GFLOPs compared to the 76.9 GFLOPs of magnifier DeepLabV3+ with ResNet18 backbone, with a factor of 2.84 of difference. Such difference increases if the SegFormer model is being considered, with a difference factor of 10.28. Thus, all Magnifier models are associated with a higher throughput, which is important when analyzing a large number of images or images that cover huge areas, such as in the context of EO.

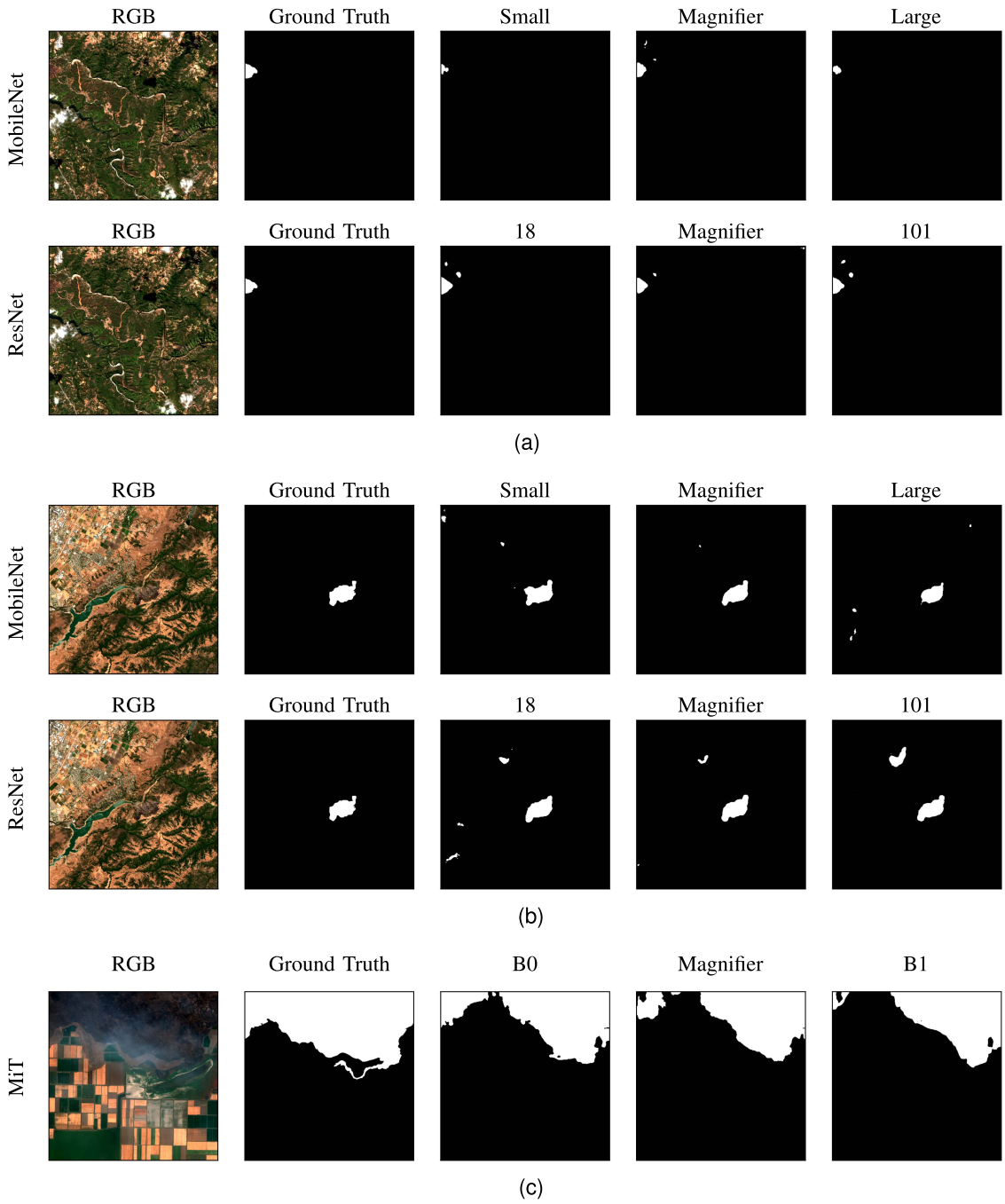

Fig. 8. Example RGB images and corresponding ground truth with predictions. Images are grouped by architecture and type of backbone. (a) DeepLabV3+. (b) U-Net. (c) SegFormer.

The erosion and dilation operations show a higher computational complexity.

6) Qualitative Analysis on Example Images: In Fig. 8, using the different models analyzed in this study. We selected three different representative images (one for each base model architecture).

The examples show that magnifier is more precise than the smallest model and generally more effective than the largest one, providing fewer false positive areas and shapes more similar to the ground truth. Magnifier reduce the noise that affects the non-Magnifier models.

Summarizing the results reported in Table IV, we can conclude the best IoU and F1 score on all datasets can be obtained using DeepLabV3+ with ResNet18 encoder in combination with the proposed magnifier architecture. Moreover, Magnifier boosts the average performance compared to single models.

G. Transfer Learning Between Continents

The datasets cover different areas of the globe with different morphological characteristics, however, we perform a cross-test of the models trained on each fold of California to the "purple"

fold of Europe, and we tested the counterparts of Europe on the "1" fold of California. The models show significant degradation in both cases: the ones trained in Europe get a mean IoU of 27.2, while the ones trained in California get 14.4. The usage of magnifier architecture does not provide any benefits in this case since it achieves a mean IoU of 22.0, while without it we can achieve 20.1. The main issue is, in both cases, the necessity of a more comprehensive collection of data to cover both areas to be capable of generalizing in both continents. We do not apply transfer learning to Indonesia since it contains data from a different satellite with different spectral bands.

V. Discussion

The introduction of the magnifier architecture marks an advancement in the field of remote sensing, particularly in how deep learning is applied to burned area delineation. Traditionally, the field has relied on general-purpose semantic segmentation models, which, while effective, often struggled to optimize the use of limited data. Our approach is a step in addressing these challenges.

A. Data Efficiency

Magnifier architecture has the ability to enhance performance without relying on large amounts of labeled data. Data scarcity has long been a barrier in remote sensing applications, as the manual annotation of satellite images is resource-intensive and time-consuming. While a multimodal approach can be appealing by providing more features per pixel, it increases the computation costs and requires extra data, which have to be harmonized together to provide consistent information. This approach is also difficult due to the different revolution times of different satellites. Magnifier's multigrained approach offers a solution by optimizing existing datasets and extracting fine and coarse-level features from the same images. This allows researchers and practitioners to achieve better results without the burden of large-scale data collection efforts.

B. Model Scalability

Magnifier achieves SOTA performance with small models, making it suitable for applications in resource-constrained environments. Many remote sensing operations, such as those deployed on satellites or drones, require models to be computationally efficient. The scalability of the magnifier architecture enables its deployment in these scenarios, opening up new possibilities for real-time environmental monitoring and disaster response. While many solutions focus only on improving performance, our proposal balances practicality, throughput, and model accuracy for a widespread application.

C. Multigranularity Feature Extraction

The introduction of multigrained feature extraction in the Magnifier architecture sets a new direction for future research in the remote sensing domain. Traditional models operate at a single granularity, while also extracting features at different resolutions. Magnifier has demonstrated the benefits of simultaneously capturing local and global contextual information. This highlights the importance of multiscale learning in burned areas delineation, particularly combining the network understanding at different contextual levels and not only at different resolutions. A broader exploration with different data types, such as optical, radar, and thermal imagery, and by leveraging multi-grained information across various spectral bands, future models could improve the precision and robustness of remote sensing applications in detecting and responding to environmental changes.

D. Transfer Learning

Due to the limited amount of data available and the different morphological and phenological characteristics of the areas under analysis caused by the different geographical distributions, we observed low segmentation performances when transferring the trained model from one region to another. In fact, all the trained architectures demonstrated the inability to generalize to different areas when switching from California to Europe and vice-versa. This observation raises the need to build a large- and global-scale dataset for emergency management.

VI. Conclusion

In this article, we presented a simple but effective way to improve the performance in burned area delineation, showing that an innovative combination of images at different levels and sizes can work better than big models. Semantic segmentation models are mainly designed to work with RGB images, while we have to deal with 12 or 8 channels. The combined usage of global (original images) and local (smaller patches) views has proved to grant better results with more features per pixel without increasing the number of needed labeled samples. The magnifier architecture was applied to well-known SOTA models, from classical CNNs to new Vision Transformers, proving its versatility.

In future research, we will investigate the design of simpler, more efficient, and effective ways to solve the burned area delineation task without the need for expensive hardware and a large amount of data. We will also investigate the effectiveness of magnifier in other domains and tasks in future works.

References

[1] J.-l. Dupuy et al., "Climate change impact on future wildfire danger and activity in southern Europe: A review," *Ann. Forest Sci.*, vol. 77, no. 2, pp. 1–24, 2020.

[2] J. Ruffault et al., "Increased likelihood of heat-induced large wildfires in the Mediterranean Basin," *Sci. Rep.*, vol. 10, no. 1, pp. 1–9, 2020.

[3] J. E. Halofsky, D. L. Peterson, and B. J. Harvey, "Changing wildfire, changing forests: The effects of climate change on fire regimes and vegetation in the Pacific Northwest, USA," *Fire Ecol.*, vol. 16, no. 1, pp. 1–26, 2020. [Online]. Available: https://doi.org/10.1186/s42408-019-0062-8

[4] R. K. Gibson, A. Mitchell, and H.-C. Chang, "Image texture analysis enhances classification of fire extent and severity using sentinel 1 and 2 satellite imagery," *Remote Sens.*, vol. 15, no. 14, 2023, Art. no. 3512.

[5] A. Farasin, L. Colomba, and P. Garza, "Double-step U-Net: A deep learning-based approach for the estimation of wildfire damage severity through Sentinel-2 satellite data," *Appl. Sci.*, vol. 10, no. 12, 2020, Art. no. 4332. [Online]. Available: https://www.mdpi.com/2076-3417/10/12/4332

[6] D. Rashkovetsky, F. Mauracher, M. Langer, and M. Schmitt, "Wildfire detection from multisensor satellite imagery using deep semantic segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 7001–7016, 2021.

[7] R. Ghali, M. A. Akhloufi, M. Jmal, W. S. Mseddi, and R. Attia, "Wildfire segmentation using deep vision transformers," *Remote Sens.*, vol. 13, no. 17, 2021, Art. no. 3527.

[8] L. Knopp, M. Wieland, M. Rättich, and S. Martinis, "A deep learning approach for burned area segmentation with Sentinel-2 data," *Remote Sens.*, vol. 12, no. 15, 2020, Art. no. 2422. [Online]. Available: https://www.mdpi.com/2072-4292/12/15/2422

[9] A. Farasin, L. Colomba, G. Palomba, G. Nini, and C. Rossi, "Supervised burned areas delineation by means of Sentinel-2 imagery and convolutional neural networks," in *Proc. 17th Int. Conf. Inf. Syst. Crisis Response Manage.*, 2020, pp. 24–27.

[10] G. Palomba, A. Farasin, and C. Rossi, "Sentinel-1 flood delineation with supervised machine learning," in *Proc. 17th Int. Conf. Inf. Syst. Crisis Response Manage.*, 2020, pp. 1072–1083.

[11] E. Arnaudo, A. Farasin, and C. Rossi, "A comparative analysis for air quality estimation from traffic and meteorological data," *Appl. Sci.*, vol. 10, no. 13, 2020, Art. no. 4587. [Online]. Available: https://www.mdpi.com/2076-3417/10/13/4587

[12] W. Zhang, P. Tang, and L. Zhao, "Fast and accurate land-cover classification on medium-resolution remote-sensing images using segmentation models," *Int. J. Remote Sens.*, vol. 42, no. 9, pp. 3277–3301, 2021. [Online]. Available: https://doi.org/10.1080/01431161.2020.1871094

[13] P. Zhang, Y. Ke, Z. Zhang, M. Wang, P. Li, and S. Zhang, "Urban land use and land cover classification using novel deep learning models based on high spatial resolution satellite imagery," *Sensors*, vol. 18, no. 11, 2018, Art. no. 3717. [Online]. Available: https://www.mdpi.com/1424-8220/18/11/3717

[14] L. Zhao and S. Ji, "CNN, RNN, or ViT? An evaluation of different deep learning architectures for spatio-temporal representation of sentinel time series," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 44–56, 2023.

[15] R. Datla, V. Chalavadi, and K. M. C, "Scene classification in remote sensing images using dynamic kernels," in *Proc. Int. Joint Conf. Neural Netw.*, IEEE, Jul. 2021, pp. 1–8. [Online]. Available: http://dx.doi.org/10.1109/IJCNN52387.2021.9533648

[16] R. Datla, N. Perveen, and K. M. C, "Learning scene-vectors for remote sensing image scene classification," *Neurocomputing*, vol. 587, Jun. 2024, Art. no. 127679. [Online]. Available: http://dx.doi.org/10.1016/j.neucom.2024.127679

[17] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Lecture Notes Comput. Sci.*, Berlin, Germany: Springer International Publishing, 2015, pp. 234–241. [Online]. Available: https://doi.org/10.1007/978-3-319-24574-4_28

[18] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis.*,, 2018, pp. 833–851.

[19] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer: Simple and efficient design for semantic segmentation with transformers," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, vol. 34, pp. 12077–12090. [Online]. Available: https://proceedings.neurips.cc/paper/2021/file/64f1f27bf1b4ec22924fd0acb550c235-Paper.pdf

[20] A. Howard et al., "Searching for MobileNetV3," in *Proc. 2019 IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2019, pp. 1314–1324.

[21] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015. [Online]. Available: http://dx.doi.org/10.1038/nature14539

[22] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[23] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997. [Online]. Available: https://doi.org/10.1162/neco.1997.9.8.1735

[24] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, Nov. 1997.

[25] A. Vaswani et al., "Attention is all you need," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, Red Hook, NY, USA, Curran Associates Inc., 2017, pp. 6000–6010.

[26] A. Dosovitskiy et al., "An image is worth 16 × 16 words: Transformers for image recognition at scale," 2020. [Online]. Available: https://arxiv.org/abs/2010.11929

[27] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Hum. Lang. Technol.*, Jun. 2019, vol. 1, pp. 4171–4186.

[28] V. Mnih et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015. [Online]. Available: http://dx.doi.org/10.1038/nature14236

[29] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. systems, man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.

[30] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.

[31] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.

[32] A. Koudounas, F. Giobergia, and E. Baralis, "Time-of-flight cameras in space: Pose estimation with deep learning methodologies," in *Proc. IEEE 16th Int. Conf. Application Inf. Commun. Technol.*, 2022, pp. 1–6.

[33] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[34] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017. [Online]. Available: https://arxiv.org/abs/1706.05587

[35] L. Colomba et al., "A dataset for burned area delineation and severity estimation from satellite imagery," in *Proc. 31st ACM Int. Conf. Inf. Knowl. Manage.*, 2022, pp. 3893–3897. [Online]. Available: https://doi.org/10.1145/3511808.3557528

[36] Z. Liu et al., "Swin Transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2021, pp. 10012–10022.

[37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[38] S. Ji, S. Wei, and M. Lu, "Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 574–586, Jan. 2019.

[39] K. Clasen, L. Hackel, T. Burgert, G. Sumbul, B. Demir, and V. Markl, "reBEN: Refined bigearthnet dataset for remote sensing image analysis," *IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, 2025.

[40] M. Wang, X. Zhang, X. Niu, F. Wang, and X. Zhang, "Scene classification of high-resolution remotely sensed image based on resnet," *J. Geovisualization Spatial Anal.*, vol. 3, no. 2, 2019, Art. no. 16.

[41] J. Zhang et al., "Winter wheat mapping method based on pseudo-labels and U-Net model for training sample shortage," *Remote Sens.*, vol. 16, no. 14, 2024, Art. no. 2553. [Online]. Available: https://www.mdpi.com/2072-4292/16/14/2553

[42] G. R. Faqe Ibrahim, A. Rasul, and H. Abdullah, "Improving crop classification accuracy with integrated sentinel-1 and Sentinel-2 data: A case study of barley and wheat," *J. Geovisualization Spatial Anal.*, vol. 7, no. 2, 2023, Art. no. 22.

[43] D. R. Cambrin, E. Poeta, E. Pastor, T. Cerquitelli, E. Baralis, and P. Garza, "Kan you see it? Kans and sentinel for effective and explainable crop field segmentation," 2024. [Online]. Available: https://arxiv.org/abs/2408.07040

[44] I. Demir et al., "DeepGlobe 2018: A challenge to parse the earth through satellite images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2018, pp. 172–181.

[45] G. Swetha, R. Datla, C. Vishnu, and K. M. C, "MS-VACSNet: A network for multi-scale volcanic ash cloud segmentation in remote sensing images," in *Proc. 18th Int. Conf. Mach. Vis. Appl.*, IEEE, Jul. 2023, pp. 1–6. [Online]. Available: http://dx.doi.org/10.23919/MVA57639.2023.10215928

[46] R. Datla, V. Chalavadi, and K. M. Chalavadi, "A multimodal semantic segmentation for airport runway delineation in panchromatic remote sensing images," in *Proc. 14th Int. Conf. Mach. Vis.*, Mar. 2022, Art. no. 21. [Online]. Available: http://dx.doi.org/10.1117/12.2622656

[47] X. Zheng, B. Wang, X. Du, and X. Lu, "Mutual attention inception network for remote sensing visual question answering," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5606514.

[48] F. Montello, E. Arnaudo, and C. Rossi, "MMFlood: A multimodal dataset for flood delineation from satellite imagery," *IEEE Access*, vol. 10, pp. 96774–96787, 2022.

[49] D. van Dijk, S. Shoaie, T. van Leeuwen, and S. Veraverbeke, "Spectral signature analysis of false positive burned area detection from agricultural harvests using Sentinel-2 data," *Int. J. Appl. Earth Observation Geoinformation*, vol. 97, 2021, Art. no. 102296. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0303243421000027
[50] D. Stroppiana, G. Bordogna, P. Carrara, M. Boschetti, L. Boschetti, and P. Brivio, "A method for extracting burned areas from landsat tm/etm images by soft aggregation of multiple spectral indices and a region growing algorithm," *ISPRS J. Photogrammetry Remote Sens.*, vol. 69, pp. 88–102, 2012.
[51] M. P. Martín, I. Gómez, and E. Chuvieco, "Burnt area index (BAIM) for burned area discrimination at regional scale using modis data," *Forest Ecol. Manage.*, vol. 234, 2006, Art. no. S221.
[52] F. Filipponi, "BAIS2: Burned area index for Sentinel-2," *Proceedings*, vol. 2, no. 7, 2018, Art. no. 364. [Online]. Available: https://www.mdpi.com/2504-3900/2/7/364
[53] D. P. Roy, L. Boschetti, and S. N. Trigg, "Remote sensing of fire severity: Assessing the performance of the normalized burn ratio," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 112–116, Jan. 2006.
[54] E. Roteta, A. Bastarrika, M. Padilla, T. Storm, and E. Chuvieco, "Development of a Sentinel-2 burned area algorithm: Generation of a small fire database for sub-saharan africa," *Remote Sens. Environ.*, vol. 222, pp. 1–17, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425718305649
[55] L. Saulino et al., "Detecting burn severity across mediterranean forest types by coupling medium-spatial resolution satellite imagery and field data," *Remote Sens.*, vol. 12, no. 4, 2020, Art. no. 741. [Online]. Available: https://www.mdpi.com/2072-4292/12/4/741
[56] N. Audebert, B. Le Saux, and S. Lefèvre, "Semantic segmentation of earth observation data using multimodal and multi-scale deep networks," in *Proc. Asian Comput. Vis.*, 2017, pp. 180–196.
[57] N. Lang, N. Kalischek, J. Armston, K. Schindler, R. Dubayah, and J. D. Wegner, "Global canopy height regression and uncertainty estimation from GEDI LiDAR waveforms with deep ensembles," *Remote Sens. Environ.*, vol. 268, 2022, Art. no. 112760. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425721004806
[58] A. Brand and A. Manandhar, "Semantic segmentation of burned areas in satellite images using a U-Net-based convolutional neural network," *Int. Arch. Photogrammetry Remote Sens. Spatial Inf. Sci.*, vol. 43, pp. 47–53, 2021.
[59] X. Hu, Y. Ban, and A. Nascetti, "Uni-temporal multispectral imagery for burned area mapping with deep learning," *Remote Sens.*, vol. 13, no. 8, 2021, Art. no. 1509. [Online]. Available: https://www.mdpi.com/2072-4292/13/8/1509
[60] X. Hu, P. Zhang, and Y. Ban, "Large-scale burn severity mapping in multispectral imagery using deep semantic segmentation models," *ISPRS J. Photogrammetry Remote Sens.*, vol. 196, pp. 228–240, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271622003410
[61] S. Monaco et al., "Attention to fires: Multi-channel deep learning models for wildfire severity prediction," *Appl. Sci.*, vol. 11, no. 22, 2021, Art. no. 11060. [Online]. Available: https://www.mdpi.com/2076-3417/11/22/11060
[62] S. T. Seydi, M. Hasanlou, and J. Chanussot, "Burnt-Net: Wildfire burned area mapping with single post-fire Sentinel-2 data and deep learning morphological neural network," *Ecological Indicators*, vol. 140, 2022, Art. no. 108999. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1470160X22004708
[63] D. R. Cambrin, L. Colomba, and P. Garza, "CaBuAr: California burned areas dataset for delineation [Software and Data Sets]," *IEEE Geosci. Remote Sens. Mag.*, vol. 11, no. 3, pp. 106–113, Sep. 2023.
[64] Y. Prabowo et al., "Deep learning dataset for estimating burned areas: Case study, Indonesia," *Data*, vol. 7, no. 6, 2022, Art. no. 78.
[65] M. Drusch et al., "Sentinel-2: ESAs optical high-resolution mission for GMES operational services," *Remote Sens. Environ.*, vol. 120, pp. 25–36, May 2012. [Online]. Available: https://doi.org/10.1016/j.rse.2011.11.026
[66] "Sentinel-2 l2a user guide," 2023. [Online]. Available: https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/product-types/level-2a
[67] D. P. Roy et al., "Landsat-8: Science and product vision for terrestrial global change research," *Remote Sens. Environ.*, vol. 145, pp. 154–172, 2014.
[68] M. Yeung, E. Sala, C.-B. Schönlieb, and L. Rundo, "Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation," *Computerized Med. Imag. Graph.*, vol. 95, 2022, Art. no. 102026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0895611121001750
[69] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, Dec. 2006.
[70] P. Refaeilzadeh, L. Tang, and H. Liu, *Cross-Validation*. New York, NY, USA: Springer, 2009. [Online]. Available: http://dx.doi.org/10.1007/978-0-387-39940-9_565

Daniele Rege Cambrin received the master's degree in computer engineering, in 2022, from Politecnico di Torino, Turin, Italy, in 2022, where he is working toward the Ph.D. degree in computer engineering with the Department of Control and Computer Engineering (DAUIN).

His main research interests are machine learning applied to geospatial data and scalable algorithms.

Luca Colomba received the master's degree in computer engineering, in 2019, from Politecnico di Torino, Turin, Italy, where he is currently working toward the Ph.D. degree in computer engineering with the Department of Control and Computer Engineering.

His major research interests include big data analytics, scalable algorithms, data mining, and machine learning applied to spatio-temporal data.

Paolo Garza received the master's and Ph.D. degrees in computer engineering from Politecnico di Torino, Turin, Italy, in 2001 and 2005, respectively.

Since December 2018, he has been an Associate Professor with the Dipartimento di Automatica e Informatica, Politecnico di Torino. He has worked on classification, clustering, itemset mining, and scalable algorithms. His research interests include data mining algorithms, big data analytics, and data science.