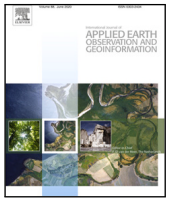
International Journal of Applied Earth Observations and Geoinformation 116 (2023) 103151

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Transformers for mapping burned areas in Brazilian Pantanal and Amazon with PlanetScope imagery

Diogo Nunes Gonçalves [g,c], José Marcato Junior [c], André Carceres Carrilho [c], Plabiany Rodrigo Acosta [a], Ana Paula Marques Ramos [d,e,*], Felipe David Georges Gomes [d], Lucas Prado Osco [f], Maxwell da Rosa Oliveira [h], José Augusto Correa Martins [c], Geraldo Alves Damasceno Júnior [i], Márcio Santos de Araújo [c], Jonathan Li [b], Fábio Roque [i], Leonardo de Faria Peres [g], Wesley Nunes Gonçalves [a,c], Renata Libonati [g]

[a] Faculty of Computer Science, Federal University of Mato Grosso do Sul, Av. Costa e Silva, s/n, Campo Grande, 79070-900, MS, Brazil
[b] Department of Geography and Environmental Management, University of Waterloo, Waterloo, N2L 3G1, ON, Canada
[c] Faculty of Engineering, Architecture, and Urbanism and Geography, Federal University of Mato Grosso do Sul, Av. Costa e Silva, s/n, Campo Grande, 79070-900, MS, Brazil
[d] Program of Environment and Regional Development, University of Western Sao Paulo, Rod Raposo Tavares, km 572, Limoeiro, Presidente Prudente, 19067-175, SP, Brazil
[e] Agronomy Program, University of Western Sao Paulo, Rod Raposo Tavares, km 572, Limoeiro, Presidente Prudente, 19067-175, SP, Brazil
[f] Faculty of Engineering and Architecture and Urbanism, University of Western Sao Paulo, Rod Raposo Tavares, km 572, Limoeiro, Presidente Prudente, 19067-175, SP, Brazil
[g] Department of Meteorology, Federal University of Rio de Janeiro, Av. Athos da Silveira Ramos, 274, Cidade Universitária, Rio de Janeiro, 21941-916, RJ, Brazil
[h] Department of Botany, Federal University de Minas Gerais, Av. Pres. Antônio Carlos, 6627 - Pampulha, Belo Horizonte, Belo Horizonte, 31270-901, BH, Brazil
[i] Department of Botany, Federal University of Mato Grosso do Sul, Av. Costa e Silva, s/n, Campo Grande, 79070-900, MS, Brazil

ARTICLE INFO

Keywords:
Multispectral imagery
Deep learning
Transfer learning
Wildfire

ABSTRACT

Pantanal is the largest continuous wetland in the world, but its biodiversity is currently endangered by catastrophic wildfires that occurred in the last three years. The information available for the area only refers to the location and the extent of the burned areas based on medium and low-spatial resolution imagery, ranging from 30 m up to 1 km. However, to improve measurements and assist in environmental actions, robust methods are required to provide a detailed mapping on a higher-spatial scale of the burned areas, such as PlanetScope imagery with 3–5 m spatial resolution. As state-of-the-art, Deep Learning (DL) segmentation methods, in specific Transformed-based networks, are one of the best emerging approaches to extract information from remote sensing imagery. Here we combine Transformers DL methods and high-resolution planet imagery to map burned areas in the Brazilian Pantanal wetland. We first compared the performances of multiple DL-based networks, namely Segformer and DTP Transformers methods with CNN-based networks like PSPNet, FCN, DeepLabV3+, OCRNet, and ISANet, applied in Planet imagery, considering RGB and near-infrared within a large dataset of 1282 image patches (512 × 512 pixels). We later verified the generalization capability of the model for segmenting burned areas in different areas, located in the Brazilian Amazon, which is also worldwide known due to its environmental relevance. As a result, the two transformers based-methods, SegFormer (F1-score equals 95.91%) and DTP (F1-score equals 95.15%), provided the most accurate results in mapping burned forest areas in Pantanal. Results show that the combination of SegFormer and RGB+NIR image with pre-trained weights is the best option (F1-score of 96.52%) to distinguish burned from not-burned areas. When applying the generated model in two Brazilian Amazon forest regions, we achieved F1-score averages of 95.88% for burned areas. We conclude that Transformer-based networks are fit to deal with burned areas in two of the most relevant environmental areas of the world using high-spatial-resolution imagery.

* Corresponding author at: Program of Environment and Regional Development, University of Western Sao Paulo, Rod Raposo Tavares, km 572, Limoeiro, Presidente Prudente, 19067-175, SP, Brazil.
E-mail address: anaramos@unoeste.br (A.P.M. Ramos).

https://doi.org/10.1016/j.jag.2022.103151
Received 23 July 2022; Received in revised form 27 November 2022; Accepted 8 December 2022

1. Introduction

Deep learning (DL) based methods have been a state-of-the-art approach to extracting information from remote sensing images (Osco et al., 2021). These methods have been used to attend scene classification, object detection, and semantic segmentation problems in several environmental applications (Ma et al., 2019). The semantic segmentation task performs a pixel-by-pixel classification to define the informational classes based on the spectral information of an image (Zhu et al., 2017). Several DL semantic segmentation architectures have been proposed by the computer vision community, and they have been assessed and adapted for different domains, including remote sensing image analysis (Yuan et al., 2021). Martins et al. (2021), for example, verified the performance of different semantic segmentation algorithms for tree mapping in urban areas with RGB images, and noted that the DeepLab v3+ approach achieved the best results. Another study, Torres et al. (2021), showed that ResU-Net was better for deforestation mapping in the Amazon forest, which presents an unbalanced class problem. DL approaches have been tested to deal with the labeling uncertainty problems and class imbalance aiming the vegetation mapping using remote sensing data as can see in Bressan et al. (2022).

The exploration of DL methods in remote sensing imagery has been noted in different environmental applications, and, in recent years, there are an increasing number of articles on deep learning for active fire detection and burned area (BA) mapping. A recent search on Web of Science (*'TS = ((deep learning) AND (wildfire OR burned area))'*) showed an increase of 104% and 80% of articles in this thematic in 2021 and 2020, respectively, compared to 2019. The majority of these approaches are based on orbital imagery, due to it global coverage. There are also assessments (Bushnaq et al., 2021; Bouguettaya et al., 2022) using UAV (unmanned aerial vehicle) imagery for early fire detection and mapping, but they are confined in small regions as UAV surveying is a cost and time-consuming task. For orbital imagery applications, several works (Hu et al., 2021; Arruda et al., 2021; Pinto et al., 2020; Rashkovetsky et al., 2021) applied DL methods in orbital images varying from the RGB spectral region to the short-wave infrared (SWIR) to map burned areas, like those offered freely, for example, by the Visible Infrared Imaging Radiometer Suite (VIIRS) systems, Landsat, and Sentinel 2A/B satellites. However, the images from these sensors present limitations in terms of temporal and spatial resolution. VIIRS imagery, for instance, is acquired daily; however, with ground sample distances (GSD) of 375 and 750 m. In contrast, Sentinel and Landsat imagery have higher spatial resolutions (10 and 30 m); however, with lower temporal resolutions of 5 and 16 days, respectively.

In terms of related works, Pinto et al. (2020) proposed a semantic segmentation algorithm named BA-Net for temporal image analysis of the VIIRS system to map burned areas that combines convolutional neural network (CNN) and long–short-term memory (LSTM). This approach was tested using data from five countries (Brazil, EUA, Portugal, Mozambique, and Australia). Another study (Hu et al., 2021) mapped burned areas in European countries like Portugal, Spain, Sweden, Greece, and Canada, using Landsat −8 and Sentinel-2 optical imagery processed by several DL methods (U-Net, HRNet, Fast-SCNN, and DeepLabv3+). The authors verified that DL methods provide higher accuracy when compared to traditional machine learning methods (random forest and support vector machine) and that HRNet outperforms other DL methods in terms of generalization of a data source. For mapping burned areas in a large area in Brazil (Savanna), Arruda et al. (2021) combined Google Earth Engine (GEE) and multi-layer perceptron (MLP), which does not compose the list of state-of-the-art DL methods. Considering the balance between spatial (10-20 m) and temporal resolutions (5 days), and also due to the availability of Synthetic Aperture Radar (SAR) data (less affected by clouds), Sentinel data have been frequently employed for burned area mapping. For example, Sentinel-2 data was applied for mapping burned areas in Portugal, southern France, and Greece using DL (Pinto et al., 2021). Belenguer-Plomer et al. (2021) combined Sentinel-1 SAR data and Sentinel-2 optical imagery with CNN for mapping burned areas. Also in this context, Zhang et al. (2021) proposed a deep learning multi-source-based method to combine SAR (Sentinel 1) and multispectral (Sentinel 2) data, using PlanetScope normalized difference vegetation index (NDVI) pre and post-fire data to generate the labeled dataset. These related works show that mapping burned areas with DL methods using high spatial–temporal resolution images like PlanetScope imagery is still little explored. This strategy constitutes a demand in areas similar to the Brazilian Pantanal and Amazon regions, characterized by intense wildfires every year.

The Brazilian Pantanal is the largest wetland region in the world, having as its main characteristic the flood pulse (Junk et al., 1989). The flooding in the Pantanal presents both temporal and spatial variations, presenting areas that never flood and areas permanently flooded (Moraes et al., 2013). These flooding variations associated with other factors make the Pantanal an extremely heterogeneous ecosystem, making it difficult to apply some remote sensing techniques. Therefore, identifying burned areas, be it on vegetation next to flood pulses or in dryer lands in the same biome, offers a potential challenge for traditional image segmentation approaches. Methods that provide information about burned areas using high-spatial-resolution images may return important information related to the quantification of emissions from fires, mainly from small and fragmented burned areas. Also can contribute to a better understanding of the causes, planning and impact analysis, restoration strategy definition, fire management assessment, etc.

PlanetScope daily imagery with a ground sample distance ranging from 3 to 5 meters are promising data to attend this demand. However, a literature analysis points out a lack of studies on mapping burned areas using the PlanetScope imagery. Even though Norway's International Climate /& Forests Initiative (NICFI) recently provided free access to Planet imagery for the world's tropics regions, which encompasses most of the Brazilian territory. Additionally, there is no information about the performance of novel semantic segmentation methods, such as SegFormer (Xie et al., 2021), DPT (Ranftl et al., 2021), ISANet (Huang et al., 2019), or OCRNet (Yuan et al., 2020), to map burnt areas using RGB and NIR images of high-spatial–temporal resolution like PlanetScope images. Among these DL algorithms, both SegFormer and DPT are characterized by using an encoder Vision Transformer-based. The use of the ViT as a backbone in image semantic segmentation task consist of a state-of-the-art approach (Xie et al., 2021; Ranftl et al., 2021). The use of architecture models ViT-based on revolutionized automatic translation and natural language processing, and they are now being investigated for classification and image segmentation (Zheng et al., 2021b). SegFormer, for instance, has advantages in relation to other ViT-based networks, mainly because it uses a hierarchically structured encoder that returns multiscale feature outputs, while also being constructed to avoid complex decoders (Xie et al., 2021). These characteristics help in combining local and global attention with its encoder, aggregating information from different layers of the network to render more powerful representations, thus improving its learning. It is also considered a lightweight type of network, which makes it suitable for multiple hardware.

In this paper, we mapped burned areas in the largest tropical wetland of the world, the Brazilian Pantanal, combining novel ViT-based deep learning methods and PlanetScope imagery. Pantanal experienced catastrophic wildfires in 2019, and 2020, and significant wildfires occurred in 2021 (Libonati et al., 2020; Leal Filho et al., 2021). We also verified the generalization capability of the model for segmenting burned areas in the Brazilian Amazon, which is also worldwide known due to its environmental relevance. In Brazil, an online platform based on the BA-Net, a deep learning method, was developed, providing daily information of burned area for Brazil, including the Pantanal and the Amazon regions, on an operational and near real-time basis,

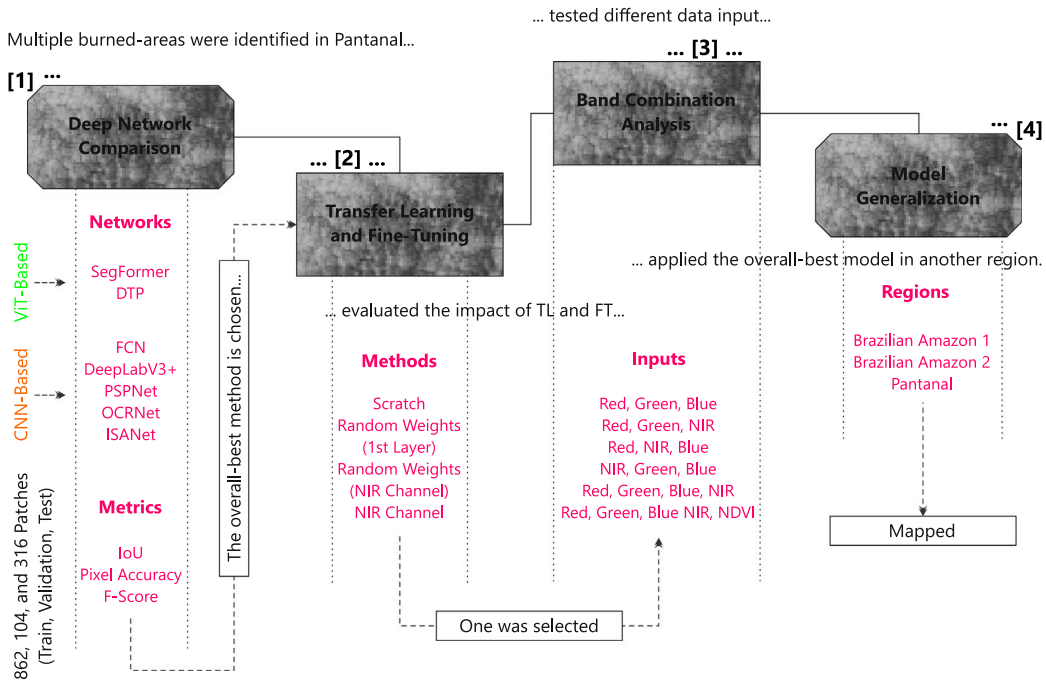


Fig. 1. A diagram simplifying the steps applied on the experiment.

the so called Alarmes Plaraform (https://alarmes.lasa.ufrj.br/) (Pinto et al., 2020). This platform uses VIIRS data, therefore, providing coarse burned area mapping. In this context, this work proposes a step toward an automated procedure to produce daily high-resolution burned areas over an extended region aiming for improving current early warning systems. This article brings three-fold contributions:

1. The most detailed mapping of burned areas for distinct Brazillian regions (Pantanal and Amazon) using PlanetScope imagery;
2. The assessment of different combinations of bands (B, G, R, NIR) to verify the spectral impact over the analysis;
3. The evaluation of state-of-the-art ViT-based deep learning methods in performing said task.

2. Materials and methods

The method organized for this study is separated into 4 phases. The first consists of a comparison between semantic segmentation deep networks, including two ViT-based methods and five CNNs. The second phase evaluates the effect of both transfer learning and fine-tuning techniques on the performance of the overall best method, identified in the previous phase. The third phase uses different band combinations and a vegetation index to verify their influences on the network's segmentation. The fourth and final phase applies the best possible model created for the Pantanal region to other tropical forest areas inside the Amazon forest and verifies its generative capabilities. Fig. 1 summarizes the process described in detail in the following sections.

2.1. Study area

The Pantanal has about 160.000 km^2 covering the countries of Bolivia, Paraguay, and Brazil (Damasceno-Junior and Pott, 2021). Brazil owns most of the Pantanal, which is more than 80% of the entire territory of the biome (Damasceno-Junior and Pott, 2021; Garcia et al., 2021). Elected biosphere reserve, it is one of the most conserved ecosystems, maintaining about 80% of its native vegetation (Roque et al., 2016). The most worrying factor for the conservation of the Pantanal today is wildfires (Garcia et al., 2021; Libonati et al., 2020). Around 8% of the Pantanal burns annually (de Oliveira-Junior et al., 2020; Libonati et al., 2022). In the year 2020, one of the worst in recent decades, fires in the Pantanal reached 43% of the entire territory, leading to the death of about 17 million vertebrates (Libonati et al., 2020; Garcia et al., 2021; Tomas et al., 2021; Libonati et al., 2022). In addition, considering the last two decades, the Pantanal has shown a tendency to increase the burned areas (Correa et al., 2022).

This scenario can be aggravated because it is predicted that the climate change in the Pantanal will present a reduction in rainfall and an increase in temperature (Silva et al., 2022), which may worsen the situation of wildfires in the region. The Pantanal has a high diversity of environments, the most representative being savanna environments, such as grasslands and open savannas, but it also has forest environments, such as dry forests and seasonal forests (dos Santos Vila da Silva et al., 2021; Pott and Pott, 2021). All these environments can present variations in their flood levels (dos Santos Vila da Silva et al., 2021; Pott and Pott, 2021). This variation allows the Pantanal to present highly heterogeneous landscapes, which can vary abruptly between completely different environments (Damasceno-Junior and Pott, 2021; Pott and Pott, 2021). For this reason, to generate more generic models we considered images acquired on several dates and three territories within its region (see Fig. 2).

2.2. Data

The images comprised PlanetScope multispectral imagery datasets (Blue—B, Green—G, Red—R, Near Infrared—NIR) with a ground sample distance (GSD) of 3.9 (±0.28) meters (PBC, 2021). PlanetScope images are acquired by a constellation of approximately 130 nanosatellites with a daily imaging coverage capacity of 200 million km^2/day. These images and freely accessible for research purposes, and are available orthorectified and in surface reflectance, that is, ready-to-use data. This eliminates the need for radiometric calibration and atmospheric correction of these scenes since images from different dates are used to map the burned areas.

To gather the reference data (i.e. burned and unburned areas) in PlanetScope imagery, manual labeling was performed by specialists with the assistance of the Geographical Information System (GIS) open-source software QGIS 3.22. Within the Pantanal, three areas containing

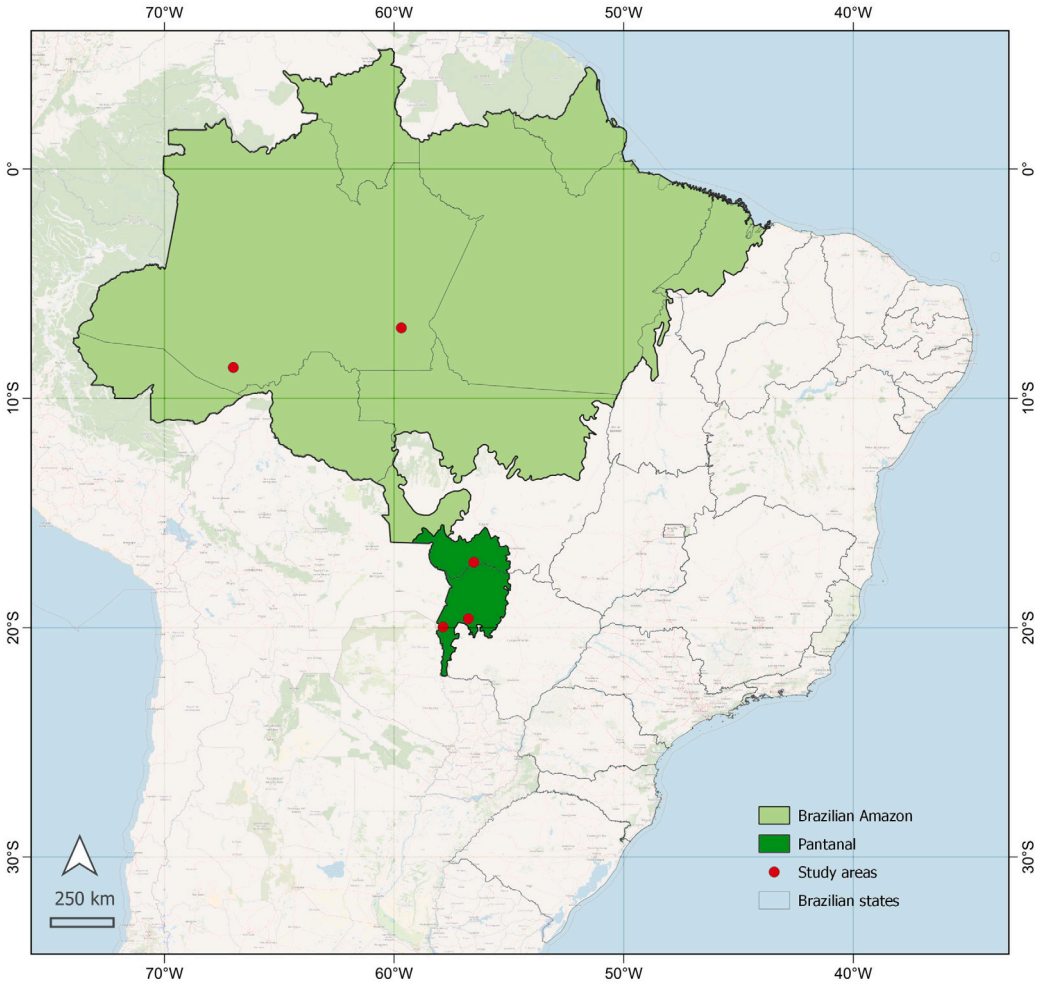


Fig. 2. Study areas in Brazilian Amazon and Pantanal.

burned regions were chosen to serve as ground truth to the comparison. In contrast, two areas containing burned regions in the Brazilian Amazon were selected. For the Pantanal region, the burned areas corresponded to 225,483.97 ha in total, which configures into 676,452 pixels marked as regions of interest. A total of 1502 fire or burning areas/events were recorded within this region. As for the Amazon area, a total of 11,906.88 ha of burned areas were identified, totalizing 35,720 pixels in 614 different areas. For the Pantanal region were registered between July and September of 2021, and different burning conditions were found, being from recently burned areas, areas that were burned but were already presenting early stages of regeneration, and areas partially covered by smoke from current active burning. To verify the impact of different band combinations in the overall best network, we used combinations among visible (Blue (B): 455–515 nm; Green (G): 500–590 nm; Red (R): 590–670 nm) and Near-Infrared (NIR) (780–860 nm), and the spectral index NDVI (Eq. (1)) as input for the DL method.

$$NDVI = (NIR - R)/(NIR + R) \tag{1}$$

We split the areas into patches of size 512 × 512 pixels without overlap due to the input dimension limitations of DL methods. A total of 1282 patches were obtained from the images. Each band was normalized between 0 and 1 according to Eq. (2). Normalization is important in this case so that the bands are on the same scale when training the networks.

$$\hat{b}(i,j) = \frac{b(i,j) - min(b)}{max(b) - min(b)} \tag{2}$$

where b is a band, $b(i,j)$ and $\hat{b}(i,j)$ are respectively the value of the band at position (i,j) and its normalized value.

2.3. Deep learning methods

To segment and map the burned areas, we used state-of-the-art semantic segmentation networks. We compared recent ViT-based methods, such as SegFormer (Xie et al., 2021) and DPT (Ranftl et al., 2021), with known CNN-based methods, like OCRNet (Yuan et al., 2020), FCN (Shelhamer et al., 2017), ISANet (Huang et al., 2019), PSPNet (Zhao et al., 2016) and DeepLabV3+ (Chen et al., 2018). In general, segmentation methods take an image as input and return a pixel-wise classification. In our case, the result of each method is an image with the class of each pixel that can be a background or a burned area. Traditional DL methods use convolution, pooling, and fully connected layers such as FCN, DeepLabV3+, PSPNet, ISANet, and OCRNet. As stated, Transformers have been used as a replacement for convolution layers to get global attention to the image. As traditional CNN methods are commonly explored in remote sensing, we did not describe them in detail. Below, we only describe the focused Transformers-based methods: SegFormer and DPT (Fig. 3.)

SegFormer (Xie et al., 2021) is an efficient semantic segmentation method that combines Transformers and multilayer perceptron decoders. SegFormer can be divided into two main modules, encoder, and decoder. In the encoder, multi-scale features are extracted from the image through hierarchically structured Transformers. Unlike the traditional Transformer, the position encoder on the encoder is implemented through convolutional layers, as it has superior performance

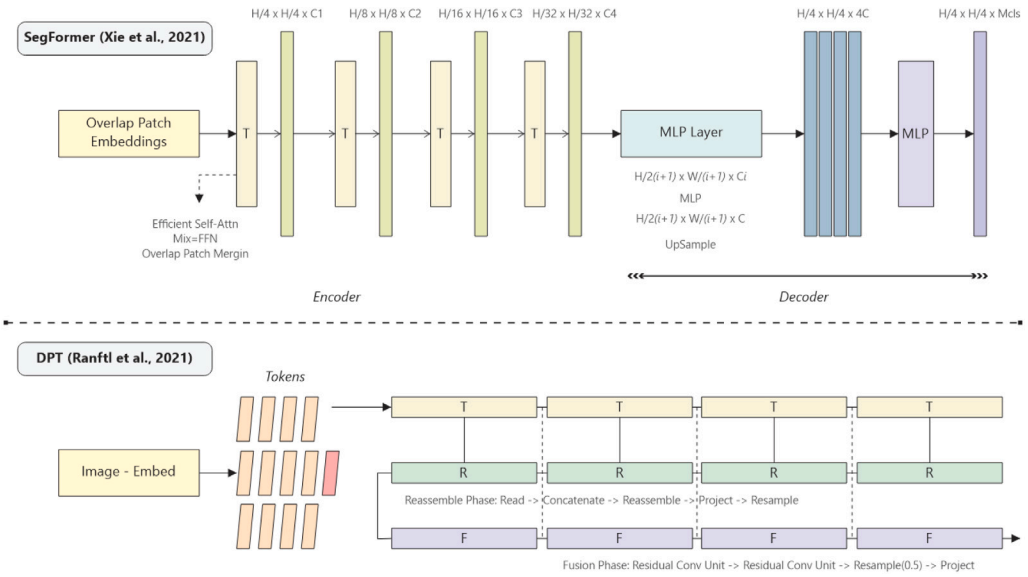


Fig. 3. Diagram summarizing the architectures of the SegFormer network (Xie et al., 2021) and the DPT (Ranftl et al., 2021).

at different image resolutions. In the decoder, the multi-scale features are aggregated to represent local and global information. Finally, the merged features are used to segment the input image. Despite the simple decoder, SegFormer provided superior results in traditional image datasets. We used the more powerful version, called SegFormer-B5 in the original work.

Dense Prediction Transformer (DPT) (Ranftl et al., 2021) is composed of an encoder–decoder structure. In the encoder, DPT uses vision transformers as a backbone to extract representations at various resolutions. The representations are composed of a set of tokens, i.e., image patches embedded in a feature space. Then the tokens are used in sequential multi-headed self-attention blocks to apply a global operation, as each token can attend to every other token. The decoder assembles the tokens in a two-dimensional (image-like) representation at various resolutions. These representations are progressively combined for a pixel-by-pixel prediction.

2.4. Experimental setup and protocol

As aforementioned, We initially performed a comparison between the state-of-the-art image segmentation methods applied in the burned area recognition task. The following methods were considered in the comparison: SegFormer (Xie et al., 2021), DPT (Ranftl et al., 2021), OCRNet (Yuan et al., 2020), FCN (Shelhamer et al., 2017), ISANet (Huang et al., 2019), PSPNet (Zhao et al., 2016) and DeepLabV3+ (Chen et al., 2018). For this comparison, the methods used four bands (R, G, B, NIR) as input and the pre-trained ImageNet weights. In general, the image segmentation methods are pre-trained on images with only 3 bands (R, G, and B). As the input in this experiment has four bands, the filters of the first layer of the backbone were randomly initialized and the others were initialized with the pre-trained weights.

A total of 862, 104, and 316 image patches (512 × 512 pixels) were used for training, validation, and testing the deep learning methods. In training each method, the encoder weights were initialized either with pre-trained weights or randomly, while the decoder weights were initialized randomly. Following the original Transformer papers, we used the AdamW optimizer for 80K iterations using a batch size of 2 for SegFormer and DPT. The initial learning rate was 0.00006 and updated by a Poly LR schedule with a factor of 1 by default. For CNN-based methods (FCN, DeepLabV3+, PSPNet, OCRNet, ISANet), we used the suggested parameters as SGD optimizer with a learning rate of 0.01, the momentum of 0.9, and weight decay of 0.0005. As with the other two methods, the training was performed for 80K iterations, but with a batch size of 4 due to the lower memory consumption of the methods based on CNNs.

We then explored the influence of transfer learning and fine-tuning procedures on the overall-best method selected from the previous comparison analysis. For this, we initialize the selected method backbone using several forms, a strategy known as transfer learning. The first strategy (scratch) consists of initializing the network's backbone weights at random. The second strategy (Random Weights - 1st Layer) was to randomly initialize only the weights of the first backbone layer, as this layer depends directly on the number of channels in the input image. The third and fourth strategies consist of initializing all backbone layers with pre-trained weights, including the first layer with the filter weights of R, G, and B band channels. The fourth channel of the filters of the first layer, which corresponds to the NIR channel of the input, was initialized randomly in the third strategy and with the weights of the Blue channel in the fourth strategy.

Finally, we evaluated the influence of the multispectral bands on burned area segmentation to determine the most important band channels. For that, we trained the overall-best method from the previous phase with ImageNet pre-trained weights and produced experiments with different configurations. Initially, we evaluated the use of only three bands, as well as most proposed DL methods. The first row of experimentation corresponds to the method using visible bands (R, G, and B). In the second, third, and fourth inputs, we use NIR in place of one of the visible spectrum bands. The idea is to understand how the NIR impacts the burned area segmentation and which band (R, G, or B) has pre-trained weights that can be used as NIR band weights. In addition, these experiments make it possible to understand the impact of each band. The organization of each input is also illustrated in Fig. 1.

To assess the generalizability of the generated model from the previous experiments, we used the selected network with four bands from the PlanetScope image collection to segment areas of the Brazilian Amazon. The Brazilian Amazon is one of the most important areas in the world, along with the Pantanal, as it represents a third of the world's tropical forests, and is home to the greatest biodiversity on the planet in plants, animals, and microorganisms. Within the Brazilian Amazon, two areas containing fire damage were chosen and manually labeled to serve as ground truth to the comparison.

The experiments were computed in a desktop computer with Intel(R) Xeon(R) CPU E3-1270@3.80 GHz, 64 GB memory, and NVIDIA

Table 1
Segmentation results of burned area using four bands (R, G, B, NIR).

Method	IoU		Pixel accuracy		F-score	
	Background	Burned area	Background	Burned area	Background	Burned area
FCN	89.48	90.35	92.9	96.4	94.45	94.93
DeepLabV3+	87.96	88.18	**95.51**	91.91	93.59	93.72
PSPNet	88.56	89.05	94.61	93.57	93.93	94.21
OCRNet	89.67	90.37	93.85	95.61	94.55	94.94
ISANet	89.15	89.82	93.87	95.01	94.26	94.64
SegFormer	**91.56**	**92.14**	94.91	**96.56**	**95.59**	**95.91**
DPT	90.04	90.75	93.85	96.01	94.76	95.15

Titan V Graphics Card (5120 Compute Unified Device Architecture—CUDA cores and 12 GB graphics memory). The methods were implemented using mmsegmentation toolbox (https://github.com/open-mmlab/mmsegmentation) on the Ubuntu 18.04 operating system. The performance of the models is evaluated using the metric F1-score (F1) (Eq. (5), pixel accuracy (Eq. (3)), and the Intersection over Union (IoU) (Eq. (4)), as they are currently used to assess semantic segmentation experiments (Xie et al., 2021; Yuan et al., 2020; Shelhamer et al., 2017; Chen et al., 2018).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$IoU = \frac{|GT \cap Prediction|}{|GT \cup Prediction|} \tag{4}$$

$$F1 - score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{5}$$

The F1-score metric is calculated based on the weighted average of Precision and Recall, where an F1-score reaches it's best value at 1 and the worst score at 0. The precision metric is defined as the number of True Positives (TP) divided by the number of true positives (TP) plus the number of False Positives (FP). The Recall metric is defined as the number of true positives (TP) over the number of true positives (TP) plus the number of False Negatives (FN). The IoU, also known as the Jaccard Index, is the ratio between the intersection and the union between the ground truth (GT) and the prediction masks.

3. Results

This section presents the segmentation results of the burned areas in the investigated regions of Pantanal using several DL based-methods for the semantic segmentation task. Later on, we present the observations of our best model to segment burned areas in two Amazon Forest regions.

3.1. Comparison of image segmentation methods

The results for the IoU, pixel accuracy, and f-score metrics are presented in Table 1. We report metrics separately for background and burned area pixels for a complete analysis of the results, as the occurrence of burned area pixels tends to be lower then the overall background. As we can see, SegFormer excelled in most metrics for the two classes, background and burned areas. Considering the IoU of the burned area, the Segformer obtained 92.14 against 90.75 for the DPT, the second-best method. This evidences the robustness of Transformers against convolutional layers, as both methods are based on this recent advance. Considering pixel accuracy, the best segmentations were from SegFormer, FCN, and DPT with metrics above 96%. For the F-score, the methods presented similar results for SegFormer, DPT, OCRNet, and FCN.

We performed a multi-fold test running the four best methods (Table 1) in two other splits of the dataset. In each split, the training, validation and test sets are randomly constructed. Table 2 presents the results for the three splits in addition to the average of each method. We can see that SegFormer continues with the best results in all metrics, further increasing the difference in its performance to the other methods. The second best method remains the DPT, which also uses transformers in its composition, which indicates that attention mechanisms may have positively influenced the results.

To compare the methods statistically, we applied the Friedman test followed by the Nemenyi post-hoc test using the IoU, pixel accuracy and F-score of the burned area. These metrics were calculated based on 316 images for the three repetitions. Friedman's test with $\alpha = 0.05$ rejected the null hypothesis that the methods have statistically similar performance. Then, the Nemenyi post-hoc test was applied to verify which pairs of methods present significantly different performance. Table 3 shows that, for $\alpha = 0.05$, SegFormer is superior to other methods. On the other hand, the other methods do not show statistical difference between them.

Finally, we performed an inference time experiment of all methods. Table 4 shows the mean time in seconds and the standard deviation of the methods. For this experiment we used all test images. We can see that the model with the lowest average inference time is ISANet with 0.053 s. SegFormer, which has the best results in segmentation metrics, maintains an acceptable time compared to the other methods, having the third best time.

For qualitative analysis, Fig. 4 presents examples of the segmentation performed by all tested methods. The first row of images corresponds to the RGB image while the second row to its ground-truth. The third, fourth, and fifth rows correspond to the Segformer, DPT, OCRNet, and FCN methods segmentation results, respectively. These methods were the best according to the quantitative analysis. Qualitatively, the best methods were SegFormer and FCN, achieving satisfactory results in these areas. The DPT and OCRNet performed worse, failing to segment significantly burned areas. The second example is partially covered by smoke, which is a common occurrence when dealing with active burning areas and visible range imagery. For this, three methods achieved good results (SegFormer, DPT, and FCN), but Segformer appears to have achieved a better definition at the edges of the burned areas.

The third example is also partially covered with smoke and SegFormer continues returning the best qualitative results, mainly achieving better definition at the edges and better dealing with the atmospheric pollution. The FCN, for instance, which had good qualitative results, was not able to segment burned areas that were under the smoke. Finally, the fourth example has a large burned area, occupying practically the entire image patch. In this case, we consider that all methods achieved good results. But, in general, SegFormer stands as the most consistent method, presenting satisfactory qualitative results for different situations, such as low-burned, large burned areas, and images partially covered, among others.

3.2. Influence of transfer learning and fine tuning

To ascertain the impact of transfer learning and fine-tuning processes, we chose the SegFormer, since it achieved satisfactory results, both quantitatively and qualitatively. The previous results showed the robustness of SegFormer against other methods using four bands (R, G, B, NIR). We then trained the SegFormer (fine-tuning) to evaluate the best initialization strategy, since the pre-trained ImageNet-1k weights are composed of only three bands (R, G, B). The results of this experiment were reported in Table 5. From this, it should be noted that the

Table 2
Segmentation results of burned area for three splits of the dataset. BA and BG stand for Burned area and Background, respectively.

Method	Splits	IoU		Pixel accuracy		F-score	
		BG	BA	BG	BA	BG	BA
SegFormer	Split 0	91.56	92.14	94.91	96.56	95.59	95.91
	Split 1	90.24	90.38	93.19	96.66	94.87	94.95
	Split 2	92.06	91.93	95.18	96.51	95.87	95.79
	Mean(std)	**91.28(±0.94)**	**91.48(±0.96)**	**94.42(±1.07)**	**96.57(±0.07)**	**95.44(±0.51)**	**95.55(±0.52)**
OCRNet	Split 0	89.67	90.37	93.85	95.61	94.55	94.94
	Split 1	85.77	86.11	90.18	94.74	92.34	92.54
	Split 2	89.59	89.34	94.2	94.69	94.51	94.37
	Mean(std)	88.34(±2.22)	88.60(±2.22)	92.74(±2,22)	95.01(±0.51)	93.8(±1.26)	93.95(±1.25)
DPT	Split 0	90.04	90.75	93.85	96.01	94.76	95.15
	Split 1	88.25	88.38	92.22	95.41	93.76	93.83
	Split 2	88.63	88.3	93.87	93.89	93.97	93.79
	Mean(std)	88.97(±0.94)	89.14(±1.39)	93.31(±0.94)	95.10(±1.09)	94.16(±0.52)	94.25(±0.77)
FCN	Split 0	89.48	90.35	92.9	96.4	94.45	94.93
	Split 1	86.44	87.22	88.86	97.14	92.73	93.17
	Split 2	89.21	89.15	93.1	95.5	94.3	94.26
	Mean(std)	88.37(±1.68)	88.90(±1.57)	91.62(±2.39)	96.34(±0.82)	93.82(±0.95)	94.12(±0.88)

Table 3
Nemenyi post-hoc test applied to IoU, pixel accuracy and F-score of the burned area for the three repetitions.

Methods	SegFormer	OCRNet	DPT	FCN
SegFormer	1	0.003	0.018	0.031
OCRNet	0.003	1	0.9	0.875
DPT	0.018	0.9	1	0.9
FCN	0.031	0.875	0.9	1

Table 4
Mean time and standard deviation of methods in all test images.

Method	Time per second
SegFormer	0.062(±0.051)
FCN	0.059(±0.098)
DPT	0.069(±0.059)
OCRNet	0.064(±0.074)
ISANet	0.053(±0.083)
DeepLabV3+	0.063(±0.091)
PSPNet	0.063(±0.090)

pre-trained weights are of critical importance for the proper training of segmentation methods since they returned better results.

For this study, we noticed that even with multispectral imagery, which is not the focus of the images on ImageNet, using pre-trained weights is better than using random weights. For example, there is an increment from 88.83 to 93.28 IoU for the burned area when using random and pre-trained weights, respectively. There is still a small increment in the IoU of the burned area when the first layer is also initialized, either with the random NIR channel or with replicated weights of the B channel (last two rows of Table 5).

3.3. Influence of the image bands input

To verify the impact of different data inputs on the SegFormer network, we tested specific groups of spectral bands (RGB + NIR) and a spectral index (NDVI) to determine if the network is capable of dealing with burned area segmentation tasks when mixing different information. Table 6 shows the results with the different band combinations. From the previous results, the B band weights are the best to initialize the NIR band. For comparison, the fifth row of the table presents the results using the four bands from the previous experiment, which can be seen as a baseline.

We observed that there is no significant impact between using all of the 4 bands (R, G, B, and NIR) of the sensor, and 3 bands with the NIR receiving the pre-trained weights of the B band. Ultimately, we evaluated the inclusion of NDVI as a fifth band in the input images according to the results in the last row of the table. The objective is to evaluate if a spectral index is known to be relevant to the network, though since is a combination of the other bands, it may help the learning process. However, the results showed that the DL method can learn band combinations as relevant as the NDVI since the results did not improve with its addition, deeming it irrelevant.

Lastly, regarding the Pantanal region, we observed that the SegFormer network was capable of dealing properly with different environmental conditions, such as the ones presented in the images (Fig. 5). SegFormer was capable of distinguishing burned areas in both old and new stages, as well as not confusing waterbodies with some of the darker burned portions. Since Pantanal is a wetland, it is often common for the presence of water at surface level. As for areas partially covered by smoke from the fires, SegFormer was still better than the other implemented methods, as previously presented (Fig. 4).

3.4. Generalization to other burned areas

The final experiment was conducted to establish the robustness and generalization of the model created in the previous steps with the SegFormer network. For that, we applied the SegFormer trained only on Pantanal images to segment burned areas in two Brazilian Amazon forest regions, which were also burned. The results are displayed in Table 7. The results in the Brazilian Amazon regions show that the method was able to generalize to other areas, obtaining results similar to those in the Pantanal regions. Regarding the IoU for the burned area, the method reached 92.15 and 92.03 for the two areas of the Brazilian Amazon, while 93.28 was obtained for the Pantanal. The other metrics were similar to the IoU.

The visual results of the segmentation of the Brazilian Amazon are organized in Figs. 6 and 7. The pixels in red represent the True Positives, i.e., pixels where the method and ground truth are both considered burned areas. The pixels in green and blue represent the errors in the prediction, respectively, the False Positives and False Negatives. It is possible to notice, from the visual results, that the method adequately predicts the vast majority of the burned areas. The main errors occurred in small portions that are difficult to label or define the class, such as the errors shown in Fig. 6.

4. Discussion

Segmenting burned areas in the largest wetland ecosystem on the planet is an important procedure that environmental and governmental institutions can use in decision-making tasks. As regarded previously, current information for the affected areas mapped in this study is

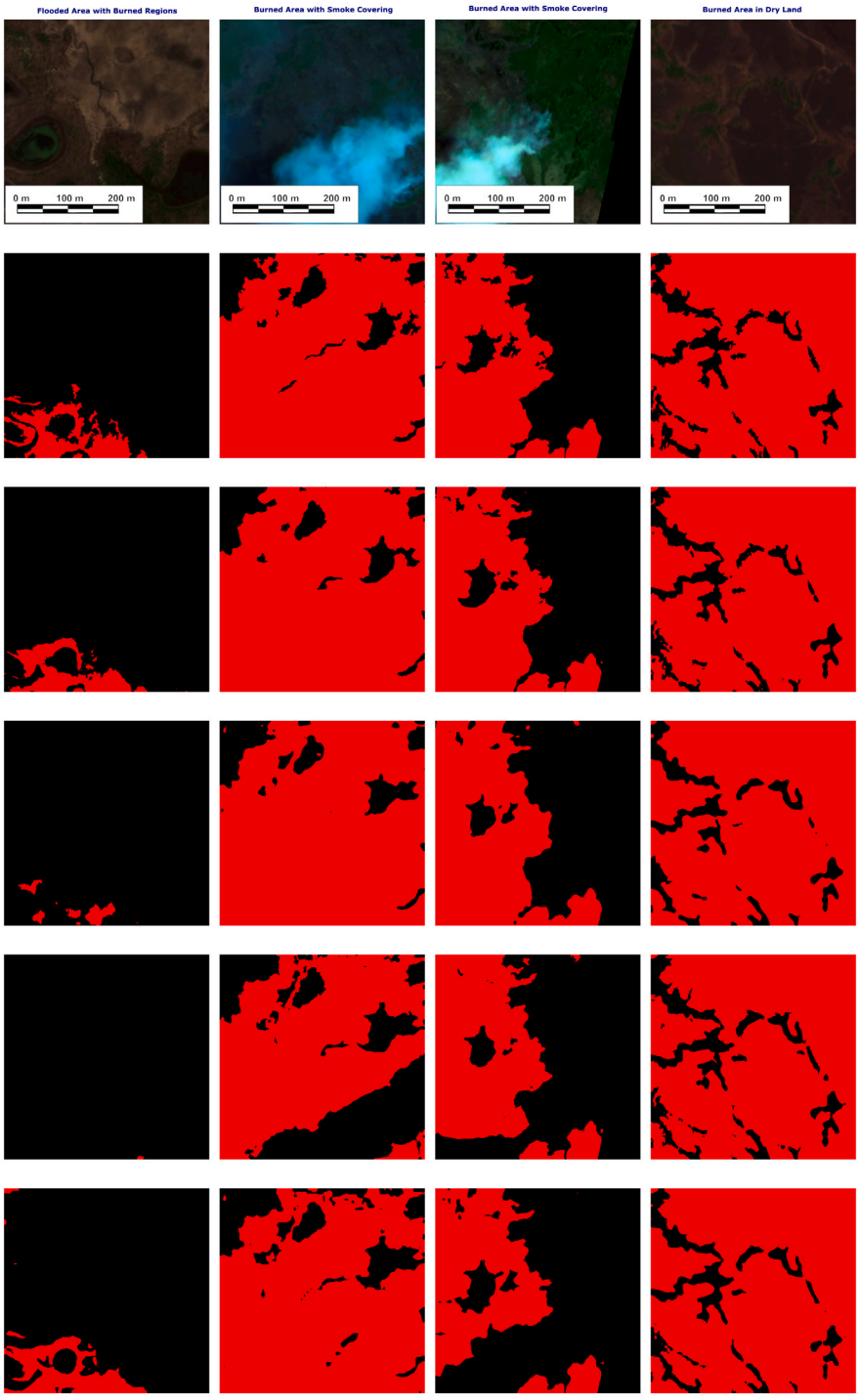


Fig. 4. Examples of segmentation for all methods. The first row corresponds to the RGB image while the second one corresponds to the groundtruth. The third, fourth and fifth rows corresponds to the segmentation of Segformer, DPT, OCRNet and FCN.

Table 5
Segmentation results of burned area with random weights (scratch) and pre-trained weights (Imagenet-1k).

Method	IoU		Pixel accuracy		F-score	
	Background	Burned area	Background	Burned area	Background	Burned area
Scratch	88.83	88.83	93.31	95.25	94.09	94.52
Random weights (1st Layer)	91.56	92.14	94.91	96.56	95.59	95.91
Random weights (NIR Channel)	92.77	93.26	95.57	**97.15**	96.25	96.51
NIR channel	**92.82**	**93.28**	**95.86**	96.92	**96.28**	**96.52**

Table 6
Segmentation results of burned area by combining different bands.

Method	IoU		Pixel accuracy		F-score	
	Background	Burned area	Background	Burned area	Background	Burned area
R, G, B	92.2	92.73	95.22	96.9	95.94	96.23
R, G, NIR	92.81	93.27	**95.87**	96.9	96.27	**96.52**
R, NIR, B	91.82	91.82	95.59	96.14	95.74	96.0
NIR, G, B	92.39	92.92	95.3	**97.04**	96.05	96.33
R, G, B, NIR	**92.82**	**93.28**	95.86	96.92	**96.28**	**96.52**
R, G, B, NIR, NDVI	92.75	93.20	95.85	96.85	96.24	96.48

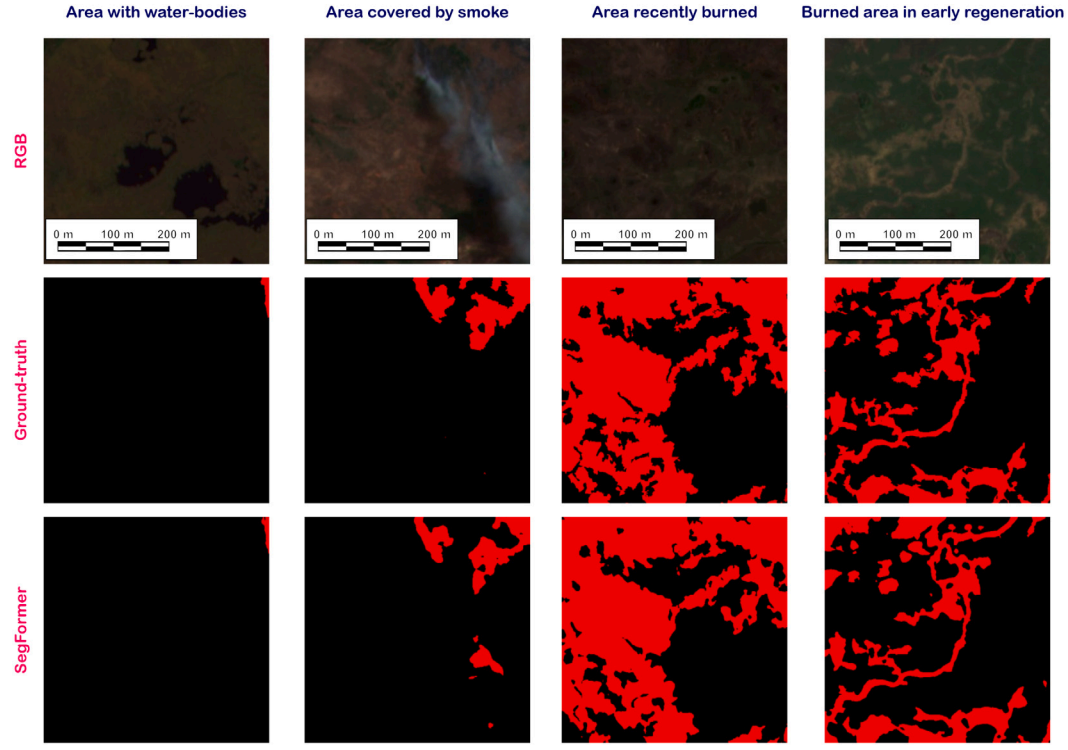


Fig. 5. Exemplification of different burned areas conditions as observed during the analysis and its segmentation result with the SegFormer network.

Table 7
Segmentation results for the Pantanal and two areas of the Brazilian Amazon.

Area	IoU		Pixel accuracy		F-score	
	Background	Burned area	Background	Burned area	Background	Burned area
Brazilian Amazon 1	99.56	92.15	99.57	99.76	99.78	95.91
Brazilian Amazon 2	99.31	92.03	99.61	96.42	99.66	95.85
Pantanal	92.82	93.28	95.86	96.92	96.28	96.52

produced by an online platform based on a DL method that uses VIIRS data (Pinto et al., 2020), which has coarse spatial resolution. For that, we demonstrated that the combination of deep learning methods and remote sensing imagery, such as the PlanetScope with RGB + NIR spectral bands and a spatial resolution of 3.9 (±0.28) meters, is suitable to map these areas in two of the most important environmental regions in Brazil, the Pantanal and the Amazon Forest. Not only does the method prove feasible to return high-detailed maps, but it also demonstrates the potential of using such data (PlanetScope), which revisits the areas daily. Since this constellation provides imagery data for each day, it is possible to increase the frequency in monitoring both active burning, as well as investigating previously burned areas, being useful for environmental planning in both controlling the current damages and restoring the destroyed areas.

As stated, we aimed to evaluate the performance of vision transformer (ViT-based) networks in dealing with the segmentation of burned and not-burned areas. ViT networks are capable of including both local and global information within their architecture (Dosovitskiy et al., 2020). This advantage over traditional CNNs based architectures should be evaluated in environmental studies, for example, but it was not yet tested in RGB + NIR high-spatially detail imagery with global coverage to perform the segmentation of burned area task for example. When comparing SegFormer and DTP performances with already known CNN-based methods (FCN, DeepLabV3+, PSPNet, OCRNet, and ISANet) its segmentation metrics (IoU, Pixel Accuracy, and F-Score) were quantitatively slight higher than these networks. Although this was emphasized by Table 1, when conducting a visual analysis, we were able to pinpoint problems within the CNN-based segmentation, especially when considering smaller burned areas, edges, and partially covered areas by smoke. This was ascertained by information exemplified in Fig. 4, demonstrating that both qualitative and quantitative analysis should verify the semantic segmentation results.

As previously stated, in recent literature, few studies investigated the capability of ViT-based networks to map fire-related issues. Dewangan et al. (2022) introduced what they called the Fire Ignition Library

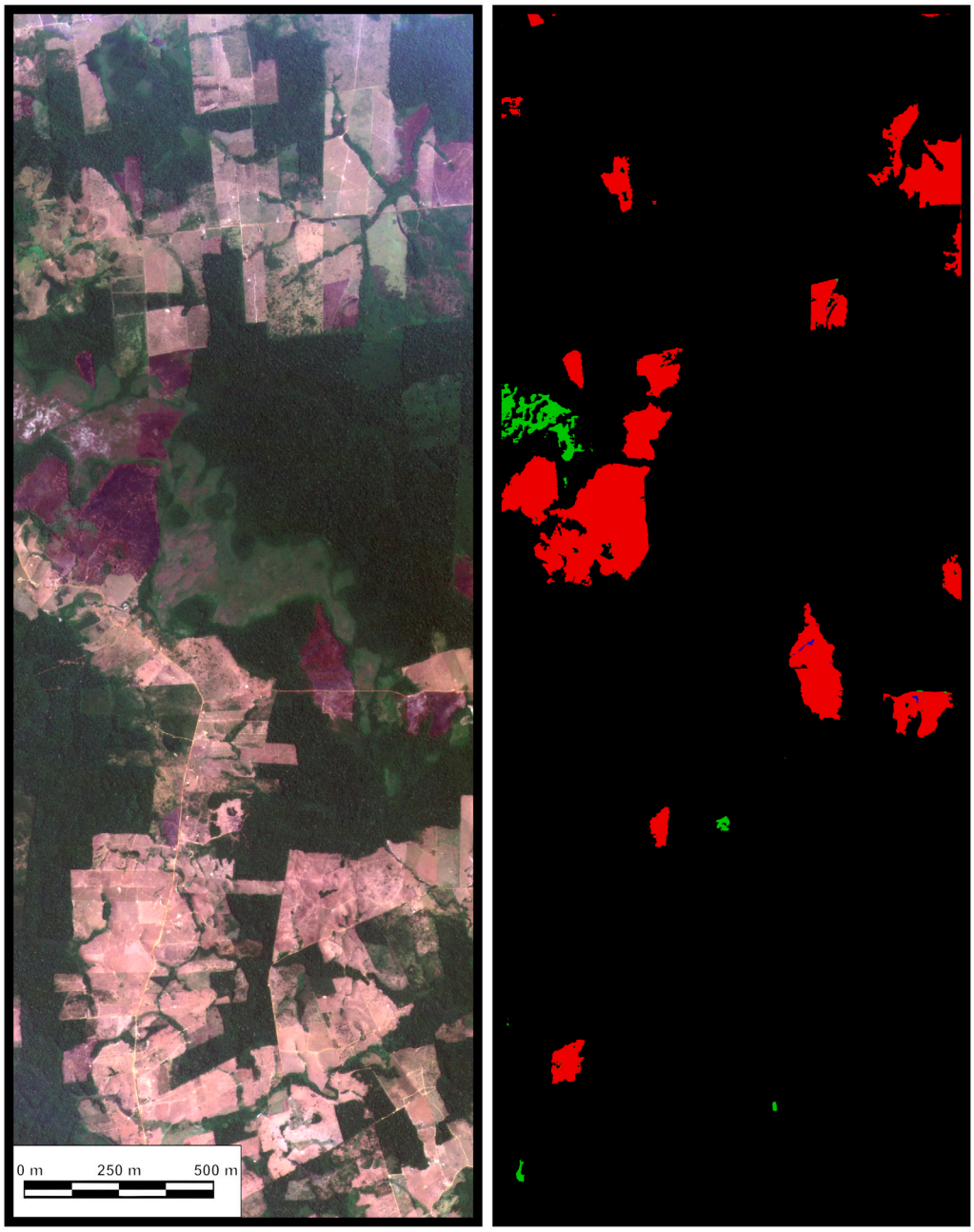


Fig. 6. Visual results of the segmentation of the burned area 1 in the Brazilian Amazon. (a) RGB image. (b) Segmentation with True Positives (red pixels), False Positives (green pixels), and False Negatives (blue pixels).

(FIgLib), which consists of a publicly available dataset containing approximately 25,000 labeled wildfire smoke images from fixed-view cameras. They also presented their network, SmokeNet, which uses ViT combined with convolutional layers and long–short-term memory cells. Ghali et al. (2022), on the other hand, presented a deep ensemble learning method combining the EfficientNet-B5 and DenseNet-201 models to classify wildfires in aerial images. Their approach also introduced a transformers comparison, achieving superior results, with F-scores higher than 99% for the ViT-based architectures implemented. Lastly, another paper from Ghali et al. (2021) addressed the problem related to the early detection of forest fires to predict their spread direction and investigated the performance of transformers in classifying imagery from publicly available datasets. Regardless, although these studies did not focus on the same aspects of remote sensing imagery as ours, they demonstrated the potential and tendency of ViT-based methods to promote higher accuracies than traditional deep learning networks, which we also observed here in this study with the network's comparison.

When considering a daily mapping approach, with active fires advancing in the area, orbital imagery is affected by atmospheric pollution resulting from the smoke, that by covering portions of the area, makes it difficult to determine the real damage at the time of the analysis. Additionally, it may also be difficult to determine the fire direction, which is important when considering animal and human rescue tasks, as well as promoving damage control actions in these areas. This may not be a hindrance when considering spectral data from the SWIR regions, mostly because of its capacity to penetrate some of the smoke particles in the atmosphere. However, for only RGB + NIR regions, our experiment demonstrated that CNN-based architectures have a hard time dealing with it, while the ViT-based networks were capable to circumvent this problem, which we assume, considering both the local and global information. Because of that, the ViT-based

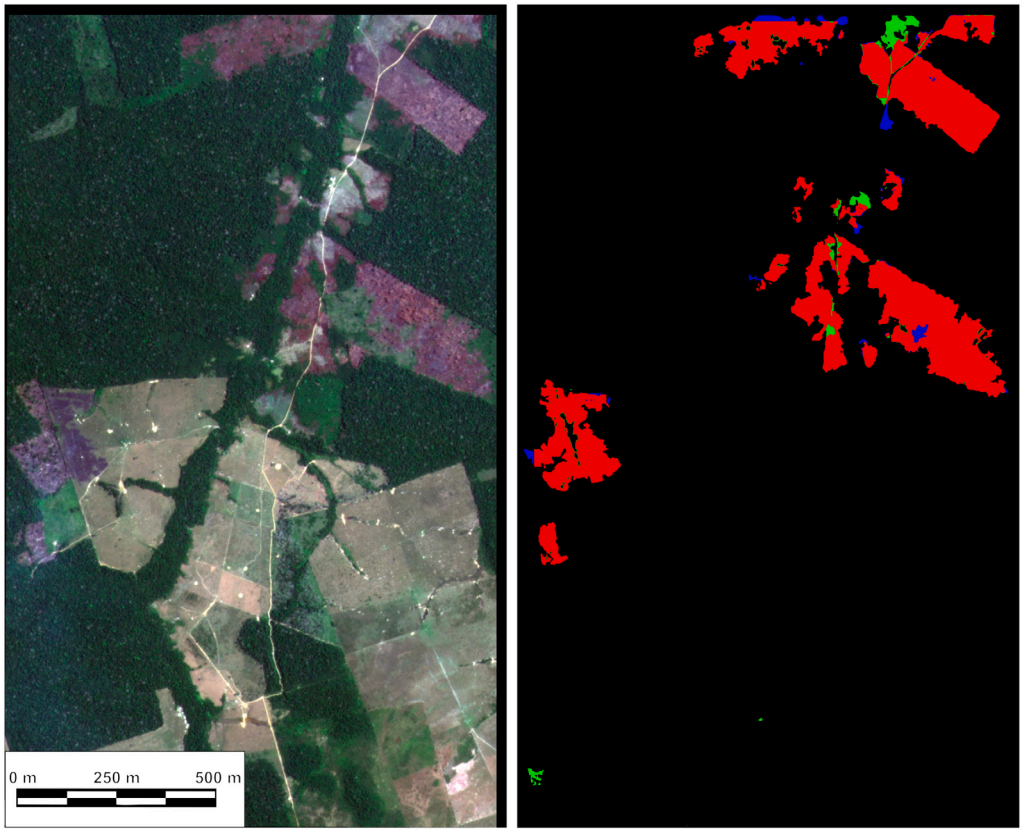


Fig. 7. Visual results of the segmentation of the burned area 2 in the Brazilian Amazon. (a) RGB image. (b) Segmentation with True Positives (red pixels), False Positives (green pixels), and False Negatives (blue pixels).

methods proved to be more suitable to resolve said problem with image of the VNIR (visible and near-infrared) regions. Because of this initial comparison, we chose to conduct additional tests with the overall-best method, SegFormer, and we were able to improve its accuracy further.

One important verification regarding our approach was investigating the transfer learning and fine-tuning conditions. Since most pre-trained networks are from RGB imagery datasets, like ImageNet-1k, our study compared the overall impact of pre-trained networks against the SegFormer method initialized with randomly generated weights and verified that even though the pre-trained models originate from RGB type of data, it returned in overall better results (Table 5). We also examined the influence of different band inputs into the network and noticed that it affected its performance, but still, the combination of RGB + NIR bands remains the overall best approach when dealing with segmenting burned areas. The spectral index NDVI was also added, but it did not improve the method's accuracies. This may be due to this index being a simple mathematical operation between the R and NIR bands and since they are also inserted as input variables into the network, one of the infinite possible combinations performed by the network could result in a similar value (Ramos et al., 2020). This is an important indicator since the introduction of spectral indexes in the analysis may result in redundant information, and as only the spectral bands appear to be sufficient, it reduces the amount of work necessary to prepare a dataset.

Additionally, it should be noted that mapping burned areas in wetlands are also a difficult task for humans mainly because of the amount of humid and water bodies surrounding the environment (Higa et al., 2022). When considering only RGB + NIR information, some of these regions tend to confuse manual labeling processes because it is difficult to distinguish between highly-burned areas (darker pixels) and some lakes or abandoned water courses throughout the wetlands areas. Regardless, the DL methods tested were quite capable of dealing with the wetland's natural characteristics. However, as an indicator of the model's generalization, the SegFormer method considering the pre-trained weights and the four spectral bands of the PlanetScope platform was used to map burned areas into two different Amazon Forest regions. Quantitatively (Table 7) it returned similar results when in comparison against the Pantanal region, and visually (Figs. 6 and 7) both areas were well detected. The model was capable of differentiating both natural water bodies, as well as agricultural regions of bare soil, being few regions confused with humid soils, presenting darker pixels.

Further studies should consider the combination of preliminary segmentation methods and DL networks, evaluating the impact of, for example, weakly-supervised methods and how well the methods are capable of improving the original segmentation. Another important piece of information to be evaluated is an analysis regarding multi-temporal imagery segmentation. Daily monitoring of wildfires is not only important to control an active burning, but also to detect and act on it, as soon as possible, minimizing the damage before it spreads into larger extensions. Lastly, techniques of domain adaptation to deal with multiple sensor data, as well as few-shot and sparse labeling investigations may be useful in novel approaches to improving the current method's generalization. These processes are considered state-of-the-art approaches (Qin and Liu, 2022; Zheng et al., 2021a) in computational vision tasks, and remote sensing imagery may greatly benefit from its integration with current ViT or CNN-based methods to investigate wildfires. Regardless, the current method proved satisfactory performance over difficult analysis situations, and it is indicative that visible to near-infrared regions and high-spatial detailed imagery is suitable to map burned areas in the wetlands.

5. Conclusion

We investigated the capabilities of deep learning methods, in specific Transformer-based networks, in mapping burned areas in the

Brazilian Pantanal wetlands. The results demonstrated that the networks based on vision transformers resulted in better accuracy than traditional CNNs architectures. The architecture SegFormer returned the best segmentation metrics, with an F1 score of 95.91%. We discovered that, when all layers are initialized with pre-trained weights from RGB imagery of ImageNet-1k, the segmentation results are better than randomly generated weights. Furthermore, the spectral band combinations affected the method's performance, but the addition of a spectral index like NDVI did not impact the segmentation task, mostly because the network is capable of achieving similar band combinations in its interactions. Still, the tests performed with SegFormer and various band combinations as input revealed that using an image of RGB+NIR is the best option (F1-score of 96.52 percent) for distinguishing burned from not-burned areas in multispectral high-spatial imagery. The experimental results in the Brazilian Amazon images also indicate that the model generated for Pantanal can be generalized to other areas (F1-Score of Brazilian Amazon areas equal 95.91% and 95.85%). We conclude that Transformer-based networks are fit to deal with burned forest areas in both Pantanal and Amazon forests, with high-spatial-resolution imagery mapping, and that future studies should focus on vision transformer's architectures to perform said task.

Funding

This research was funded by CNPq (p: 433783/2018-4, 310517/2020-6, 303559/2019-5, 304052/2019-1, 405997/2021-3, 445354/2020-8, and 311487/2021-1), FUNDECT (p: 71/009.436/2022, 427/2021), Project Rede Pantanal/FINEP (p: 01.20.0201.00), FAPERJ (26/202.174/2019) and CAPES PrInt (p:88881.311850/2018-01). Imasul TF (001/2022). The authors acknowledge the support of the UFMS (Federal University of Mato Grosso do Sul), Fundação Coppetec and CAPES (Finance Code 001).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors would like to acknowledge Nvidia Corporation for the donation of the Titan X graphics card. All authors approved the version of the manuscript to be published.

References

Arruda, V.L., Piontekowski, V.J., Alencar, A., Pereira, R.S., Matricardi, E.A., 2021. An alternative approach for mapping burn scars using Landsat imagery, Google Earth Engine, and Deep Learning in the Brazilian Savanna. Remote Sensing Applications: Society and Environment 22, 100472. http://dx.doi.org/10.1016/j.rsase.2021.100472.

Belenguer-Plomer, M.A., Tanase, M.A., Chuvieco, E., Bovolo, F., 2021. CNN-based burned area mapping using radar and optical data. Remote Sens. Environ. 260, 112468. http://dx.doi.org/10.1016/j.rse.2021.112468.

Bouguettaya, A., Zarzour, H., Taberkit, A.M., Kechida, A., 2022. A review on early wildfire detection from unmanned aerial vehicles using deep learning-based computer vision algorithms. Signal Process. 190 (16), http://dx.doi.org/10.1016/j.sigpro.2021.108309.

Bressan, P.O., Junior, J.M., Correa Martins, J.A., de Melo, M.J., Gonçalves, D.N., Freitas, D.M., Marques Ramos, A.P., Garcia Furuya, M.T., Osco, L.P., de Andrade Silva, J., Luo, Z., Garcia, R.C., Ma, L., Li, J., Gonçalves, W.N., 2022. Semantic segmentation with labeling uncertainty and class imbalance applied to vegetation mapping. Int. J. Appl. Earth Obs. Geoinf. 108, 102690. http://dx.doi.org/10.1016/j.jag.2022.102690.

Bushnaq, O.M., Chaaban, A., Al-Naffouri, T.Y., 2021. The role of UAV-IoT networks in future wildfire detection. IEEE Internet Things J. 8 (23), 16984–16999. http://dx.doi.org/10.1109/JIOT.2021.3077593.

Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European Conference on Computer Vision. ECCV, pp. 801–818.

Correa, D.B., Alcântara, E., Libonati, R., Massi, K.G., Park, E., 2022. Increased burned area in the Pantanal over the past two decades. Sci. Total Environ. 835, 155386.

Damasceno-Junior, G.A., Pott, A., 2021. General features of the pantanal wetland. In: Flora and Vegetation of the Pantanal Wetland. Springer, pp. 1–10.

de Oliveira-Junior, J.F., Teodoro, P.E., da Silva Junior, C.A., Baio, F.H.R., Gava, R., Capristo-Silva, G.F., de Gois, G., Correia Filho, W.L.F., Lima, M., de Barros Santiago, D., et al., 2020. Fire foci related to rainfall and biomes of the state of Mato Grosso do Sul, Brazil. Agricult. Forest Meteorol. 282, 107861.

Dewangan, A., Pande, Y., Braun, H.W., Vernon, F., Perez, I., Altintas, I., Cottrell, G.W., Nguyen, M.H., 2022. FIgLib &amp SmokeyNet: dataset and deep learning model for real-time wildland fire smoke detection. Remote Sens. 14 (4), 1007. http://dx.doi.org/10.3390/rs14041007.

dos Santos Vila da Silva, J., Pott, A., Chaves, J.V.B., 2021. Classification and mapping of the vegetation of the brazilian pantanal. In: Flora and Vegetation of the Pantanal Wetland. Springer, pp. 11–38.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2020. An image is worth 16x16 words: transformers for image recognition at scale. CoRR abs/2010.11929, URL: https://arxiv.org/abs/2010.11929, arXiv:2010.11929.

Garcia, L.C., Szabo, J.K., de Oliveira Roque, F., Pereira, A.d.M.M., da Cunha, C.N., Damasceno-Júnior, G.A., Morato, R.G., Tomas, W.M., Libonati, R., Ribeiro, D.B., 2021. Record-breaking wildfires in the world's largest continuous tropical wetland: integrative fire management is urgently needed for both biodiversity and humans. J. Environ. Manag. 293, 112870.

Ghali, R., Akhloufi, M.A., Jmal, M., Mseddi, W.S., Attia, R., 2021. Wildfire segmentation using deep vision transformers. Remote Sens. 13 (17), 3527. http://dx.doi.org/10.3390/rs13173527.

Ghali, R., Akhloufi, M.A., Mseddi, W.S., 2022. Deep learning and transformer approaches for UAV-based wildfire detection and segmentation. Sensors 22 (5), 1977. http://dx.doi.org/10.3390/s22051977.

Higa, L., Junior, J.M., Rodrigues, T., Zamboni, P., Silva, R., Almeida, L., Liesenberg, V., Roque, F., Libonati, R., Gonçalves, W.N., Silva, J., 2022. Active fire mapping on brazilian pantanal based on deep learning and CBERS 04A imagery. Remote Sens. 14 (3), 688. http://dx.doi.org/10.3390/rs14030688.

Hu, X., Ban, Y., Nascetti, A., 2021. Uni-temporal multispectral imagery for burned area mapping with deep learning. Remote Sens. 13 (8), http://dx.doi.org/10.3390/rs13081509.

Huang, L., Yuan, Y., Guo, J., Zhang, C., Chen, X., Wang, J., 2019. Interlaced sparse self-attention for semantic segmentation. arXiv preprint arXiv:1907.12273.

Junk, W.J., Bayley, P.B., Sparks, R.E., et al., 1989. The flood pulse concept in river-floodplain systems. Canad. Spec. Publ. Fish. Aquat. Sci. 106 (1), 110–127.

Leal Filho, W., Azeiteiro, U.M., Salvia, A.L., Fritzen, B., Libonati, R., 2021. Fire in paradise: why the pantanal is burning. Environ. Sci. Policy 123, 31–34. http://dx.doi.org/10.1016/j.envsci.2021.05.005.

Libonati, R., DaCamara, C.C., Peres, L.F., Sander de Carvalho, L.A., Garcia, L.C., 2020. Rescue Brazil's burning Pantanal wetlands. Nat. Publ. Group 588, URL: https://www.nature.com/articles/d41586-020-03464-1.

Libonati, R., Geirinhas, J.L., Silva, P.S., Russo, A., Rodrigues, J.A., Belém, L.B., Nogueira, J., Roque, F.O., DaCamara, C.C., Nunes, A.M., et al., 2022. Assessing the role of compound drought and heatwave events on unprecedented 2020 wildfires in the Pantanal. Environ. Res. Lett. 17 (1), 015005.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A., 2019. Deep learning in remote sensing applications: A meta-analysis and review. ISPRS J. Photogramm. Remote Sens. 152, 166–177. http://dx.doi.org/10.1016/j.isprsjprs.2019.04.015.

Martins, J.A.C., Nogueira, K., Osco, L.P., Gomes, F.D.G., Furuya, D.E.G., Gonçalves, W.N., Sant'Ana, D.A., Ramos, A.P.M., Liesenberg, V., dos Santos, J.A., de Oliveira, P.T.S., Junior, J.M., 2021. Semantic segmentation of tree-canopy in urban environment with pixel-wise deep learning. Remote Sens. 13 (16), http://dx.doi.org/10.3390/rs13163054.

Moraes, E.C., Pereira, G., da Silva Cardozo, F., 2013. Evaluation of reduction of Pantanal wetlands in 2012. Geografia 38, 81–93.

Osco, L.P., Marcato Junior, J., Marques Ramos, A.P., de Castro Jorge, L.A., Fatholahi, S.N., de Andrade Silva, J., Matsubara, E.T., Pistori, H., Gonçalves, W.N., Li, J., 2021. A review on deep learning in UAV remote sensing. Int. J. Appl. Earth Obs. Geoinf. 102, 102456. http://dx.doi.org/10.1016/j.jag.2021.102456.

PBC, P.L., 2021. Planet application program interface: in space for life on earth. URL: https://api.planet.com.

Pinto, M.M., Libonati, R., Trigo, R.M., Trigo, I.F., DaCamara, C.C., 2020. A deep learning approach for mapping and dating burned areas using temporal sequences of satellite images. ISPRS J. Photogramm. Remote Sens. 160, 260–274. http://dx.doi.org/10.1016/j.isprsjprs.2019.12.014.

Pinto, M.M., Trigo, R.M., Trigo, I.F., DaCamara, C.C., 2021. A practical method for high-resolution burned area monitoring using sentinel-2 and VIIRS. Remote Sens. 13 (9), http://dx.doi.org/10.3390/rs13091608.

Pott, A., Pott, V.J., 2021. Flora of the pantanal. In: Flora and Vegetation of the Pantanal Wetland. Springer, pp. 39–228.

Qin, R., Liu, T., 2022. A review of landcover classification with very-high resolution remotely sensed optical images—analysis unit, model scalability and transferability. Remote Sens. 14 (3), 646.

Ramos, A.P.M., Osco, L.P., Furuya, D.E.G., Gonçalves, W.N., Santana, D.C., Teodoro, L.P.R., da Silva Junior, C.A., Capristo-Silva, G.F., Li, J., Baio, F.H.R., Junior, J.M., Teodoro, P.E., Pistori, H., 2020. A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices. Comput. Electron. Agric. 178, 105791. http://dx.doi.org/10.1016/j.compag.2020.105791.

Ranftl, R., Bochkovskiy, A., Koltun, V., 2021. Vision transformers for dense prediction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12179–12188.

Rashkovetsky, D., Mauracher, F., Langer, M., Schmitt, M., 2021. Wildfire detection from multisensor satellite imagery using deep semantic segmentation. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 7001–7016. http://dx.doi.org/10.1109/JSTARS.2021.3093625.

Roque, F.O., Ochoa-Quintero, J., Ribeiro, D.B., Sugai, L.S., Costa-Pereira, R., Lourival, R., Bino, G., 2016. Upland habitat loss as a threat to Pantanal wetlands. Conservation Biology 30 (5), 1131–1134.

Shelhamer, E., Long, J., Darrell, T., 2017. Fully convolutional networks for semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39 (4), 640–651.

Silva, P.S., ao L. Geirinhas, J., Lapere, R., Laura, W., Cassain, D., Alegría, A., Campbell, J., 2022. Heatwaves and fire in pantanal: historical and future perspectives from CORDEX-CORE. Journal of Environmental Management 323, 116193. http://dx.doi.org/10.1016/j.jenvman.2022.116193, https://www.sciencedirect.com/science/article/pii/S0301479722017662.

Tomas, W.M., Berlinck, C.N., Chiaravalloti, R.M., Faggioni, G.P., Strüssmann, C., Libonati, R., Abrahão, C.R., do Valle Alvarenga, G., de Faria Bacellar, A.E., de Queiroz Batista, F.R., et al., 2021. Distance sampling surveys reveal 17 million vertebrates directly killed by the 2020's wildfires in the Pantanal, Brazil. Sci. Rep. 11 (1), 1–8.

Torres, D.L., Turnes, J.N., Soto Vega, P.J., Feitosa, R.Q., Silva, D.E., Marcato Junior, J., Almeida, C., 2021. Deforestation detection with fully convolutional networks in the amazon forest from landsat-8 and sentinel-2 images. Remote Sens. 13 (24), http://dx.doi.org/10.3390/rs13245084.

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P., 2021. SegFormer: Simple and efficient design for semantic segmentation with transformers. Adv. Neural Inf. Process. Syst. 34, 12077–12090.

Yuan, Y., Chen, X., Wang, J., 2020. Object-contextual representations for semantic segmentation. In: European Conference on Computer Vision. Springer, pp. 173–190.

Yuan, X., Shi, J., Gu, L., 2021. A review of deep learning methods for semantic segmentation of remote sensing imagery. Expert Syst. Appl. 169, 114417.

Zhang, Q., Ge, L., Zhang, R., Metternicht, G.I., Du, Z., Kuang, J., Xu, M., 2021. Deep-learning-based burned area mapping using the synergy of sentinel-1&2 data. Remote Sens. Environ. 264, 112575. http://dx.doi.org/10.1016/j.rse.2021.112575.

Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2016. Pyramid scene parsing network. CoRR abs/1612.01105, URL: http://arxiv.org/abs/1612.01105, arXiv:1612.01105.

Zheng, S., Lu, J., Zhao, H., Zhu, X., Luo, Z., Wang, Y., Fu, Y., Feng, J., Xiang, T., Torr, P.H., et al., 2021b. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6881–6890.

Zheng, J., Wu, W., Yuan, S., Zhao, Y., Li, W., Zhang, L., Dong, R., Fu, H., 2021a. A two-stage adaptation network (TSAN) for remote sensing scene classification in single-source-mixed-multiple-target domain adaptation (S^2M^2T DA) scenarios. IEEE Trans. Geosci. Remote Sens. 60, 1–13.

Zhu, X.X., Tuia, D., Mou, L., Xia, G.S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: A comprehensive review and list of resources. IEEE Geosci. Remote Sens. Mag. 5 (4), 8–36.