Green is the new Black: Multimodal Noisy Segmentation based fragmented burn scars identification in Amazon Rainforest

Satyam Mohla[iD][1,3*], **Sidharth Mohla**[2,3*][iD], **Anupam Guha**[1]

[1]IIT Bombay
[2]IIT Hyderabad
[3]Koloro AI Labs
✉satyammohla@gmail.com

Abstract

Detection of burn marks due to wildfires in inaccessible rain forests is important for various disaster management and ecological studies. The fragmented nature of arable landscapes and diverse cropping patterns often thwart the precise mapping of burn scars. Recent advances in remote-sensing and availability of multimodal data offer a viable solution to this mapping problem. However, the task to segment burn marks is difficult because of its indistinguishably with similar looking land patterns, severe fragmented nature of burn marks and partially labelled noisy datasets.

In this work we present AmazonNET – a convolutional based network that allows extracting of burn patters from multimodal remote sensing images. The network consists of UNet- a well-known encoder decoder type of architecture with skip connections. The proposed framework utilises stacked RGB-NIR channels to segment burn scars from the pastures by training on a new weakly labelled noisy dataset from Amazonia.

Our model illustrates superior performance by correctly identifying partially labelled burn scars and rejecting incorrectly labelled samples, demonstrating our approach as one of the first to effectively utilise deep learning based segmentation models in multimodal burn scar identification.

1 Introduction

In the Amazon forest, fire is associated with several land-practices. Slash-and-Burn is one of the most used practices in Brazilian agriculture (as part of a seasonal cycle called `"queimada"`[WinklerPrins, 2006]). Whether for opening and cleaning agricultural areas or renewing pastures, its importance in the agricultural chain is undeniable. Unfortunately, this is often the cause of wildfires in forests[Van Vliet *et al.*, 2012; Lewis *et al.*, 2015; Juárez-Orozco *et al.*, 2017].

Amazon rainforests are a major reservoir for flora and billions of tons of carbon, release of which can cause a major

---
*Work partially done while at Koloro AI Labs

Figure 1: Overview of burnings in the vicinity of BR-163 highway, Para, northern Brazil, in Amazon region. *Taken from [Basso, 2019]*

increase in temperatures. Recent news of wildfires in Amazon therefore, caused major uproar and concern (Fig. 1).

Uncontrollable fires, especially in dry season, have major local & regional impacts, leading to destruction of natural biomes, extinction of animal & plant species, pollution, erosion and an imbalance in the carbon cycle. Such disturbances affect agricultural production as well. Thus, many environmental studies & resources management activities require accurate identification of burned areas for monitoring the affected regions (the so-called scars from burning) spatially and temporally in order to understand and assess the vulnerability of these areas and to promote sustainable development.

Due to the large geographical extent of fires at regional and global scales and the limited accessibility of the areas affected by fire, remote sensing approaches have become cost effective alternatives in the last few years, capable of collecting burned area information at adequate spatial and temporal
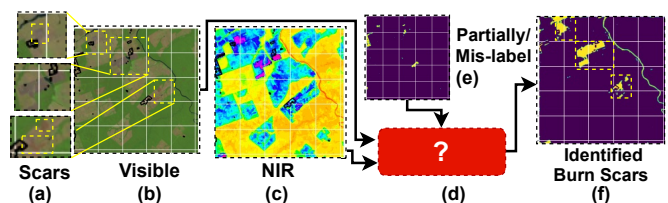

Figure 2: Generic schematic of a multi-modal noisy, weakly labelled burn scar identification. (a) Unlabelled/Correctly labelled burn scars (b) Visible Band (c) Near Infrared Band (d) Unknown Model (e) Partially/ Noisy training labels (f) Output burn scar map

resolutions. Remote sensing technologies can provide useful data for fire management, estimation & detection, fuel mapping, to post wildfire monitoring, including burn area and severity estimation[Lanorte *et al.*, 2013].

Current non-deep learning methods heavily rely on domain knowledge and manual input from the user and are unable to extract the abstract representations from the data. Deep learning attempts to resolve these problems however, they remain largely neglected in burn scar prediction due to general lack of any labelled data. In this work, we leverage the recent advances in sensing leading to ubiquitous availability of multimodal data and computer vision in remote sensing to utilise noisy, weakly labelled data to identify fragmented burn scars using UNet, making our approach one of the first to utilise deep learning based segmentation models in multimodal burn scar identification. The same has been illustrated in Fig. 2

2   Related Work

Semantic Segmentation   Semantic Segmentation is an important problem in computer vision. It involves the clustering of various pixels together if they belong to the same object class. Due to their ability to capture semantic context with precise localisation, they have been used for various applications in autonomous driving [Ess *et al.*, 2009; Geiger *et al.*, 2012; Cordts *et al.*, 2016], human-machine interaction [Oberweger *et al.*, 2015], diagnosis and detection [Shvets *et al.*, 2018; Tschandl *et al.*, 2019], remote sensing [Seferbekov *et al.*, 2018; Zhang and Wang, 2019] etc.

Before the advent of DNNs, variety of features were used for semantic segmentation, such as K-means[Hartigan, 1975], Histogram of oriented gradients[Dalal and Triggs, 2005], Scale-invariant feature transform[Lowe, 2004; Suga *et al.*, 2008] etc. Today, many encoder-decoder networks and their variants like SegNet [Badrinarayanan *et al.*, 2017], have been proposed. Specialized applications have led to novel improvements like UNet for Medical Image Segmentation[Ronneberger *et al.*, 2015], CRFs based networks for fine segmentation[Chen *et al.*, 2018]

Multimodal data in remote sensing   Multimodal segmentation in remote sensing involves utilising various strategies to efficiently combine information contained in multiple modalities to generate a single, rich, fused representation beneficial for accurate land use classification. Common methods involve concatenation channels at input stage [Iglovikov *et al.*, 2018], concatenation of features extracted from unimodal networks like CNNs, as in [Chen *et al.*, 2016; Chen *et al.*, 2017] to generate land mapping segmentation. Recent works involve more sophisticated ideas like 'cross-attention' [Mohla *et al.*, 2020] to fuse multiple modalities and generate attentive spectral and spatial representations.

Burn Scar Identification   Simultaneous availability of multimodal data has led to recent advances in locating fires and in quantifying the area burned. Each modality provides discriminating information about the same geographical region, helping in mapping amidst adverse conditions like spectral confusion (like due to cloud shadowing) & variability in burn scars making distinguishing between vegetation difficult. Majority of work done in this domain involves methods like autocorrelation [Lanorte *et al.*, 2013], self-organizing maps[Lasaponara *et al.*, 2019], linear spectral mixture model[Barbosa *et al.*, 2015], SVM [Pereira *et al.*, 2017], random forests [Liu *et al.*, 2020]. However, no recent works seem to utilise current deep learning methods like CNN or encoder-decoder models like SegNet or UNet, presumably due to lack of labelled data.

3   Proposed Method

The objective of this work is to perform semantic segmentation and identify burn scars by harnessing the spatio-spectral information constituted in visible and near infrared. To accomplish this task, we consider RGB and NIR samples $\mathcal{X} = \{\mathbf{x}_{RGB}^i, \mathbf{x}_{NIR}^i\}_{i=1}^n$ with the ground truth $\mathcal{Y} = \{y^i\}_{i=1}^n$.

Here, $\mathbf{x}_{RGB}^i \in \mathbb{R}^{M \times N \times B_1}$ and $\mathbf{x}_{NIR}^i \in \mathbb{R}^{M \times N \times B_2}$ where, B_1 and B_2 denote the number of channels, while n denotes the number of available samples. The ground-truth labels $y_i^n \in \{0, 1\}$, where 0 represents 'no burn scar' $\in \{$river, green pastures, brown pastures$\}$ and 1 represents 'burn scar'. The samples are sent to the proposed AmazonNet model which get processed as discussed ahead.
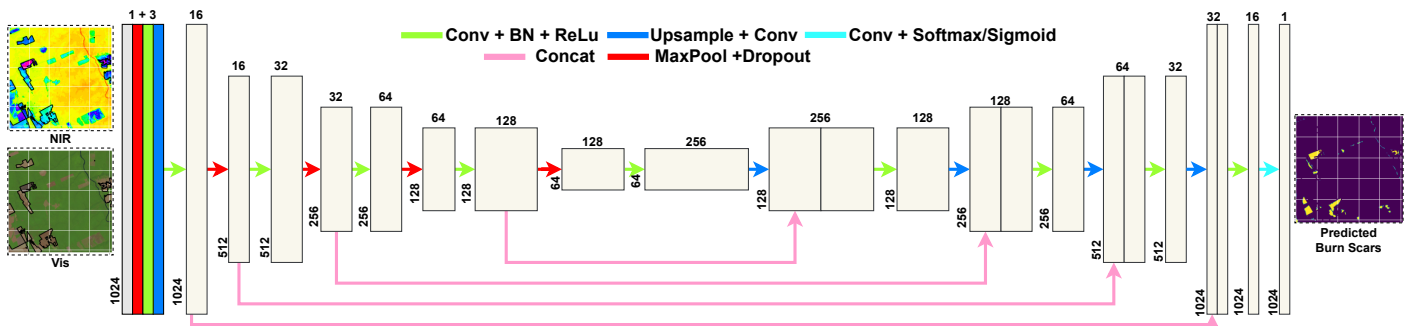

Figure 3: Architecture of U-net model: AmazonNet (presented on Amazon dataset). Input consists of 4 channel concatenated input corresponding to RGB and NIR (colormap: hsv). ReLu activation is used. 3x3 Convolution and 2x2 Max Pool are used thruout the network.

3.1 Architecture

In remote sensing, computer vision based methods are difficult to apply, due to lack of good labelled datasets, because the required data processing and labelling can only be done by field experts, making labelled data rare or unavailable. Similar problems arise in medical image segmentation, and so common approaches in remote sensing are sometimes inspired from medical segmentation domain.

For burn scar segmentation task, we base our network on the UNet[Ronneberger *et al.*, 2015] architecture, where the feature activations from encoder were stored and transferred to corresponding decoder layer for concatenation.

Encoder The encoder network consists of 3x3 convolution layers along with batch normalization layers, ReLU nonlinear activation layers, and 2x2 max-pooling layers.

Decoder The decoder network consists of UpSampling layers named as UpPool block, which performs 3x3 Conv2DTranspose, 3x3 Convolution along with batchnormalization layers and Dropout2D layers with a dropout value of 0.1. The output results are thresholded to obtain a binary output map denoting the burn scars.

3.2 Datasets

This dataset consists of a visible and near infrared satellite imagery from LANDSAT8 of the Amazon Rainforest. The dataset was acquired for 2017 and 2018 from over 4 states, namely *Tocantins, Maranhao, Mato Grosso, Para* covering over four terrestrial Brazilian biomes namely *Cerrado, Amazonia, Caatinga, Pantanal*. It consists of 299 samples VIS-NIR image pairs of size 1024x1024 with ground truth, which are binary images, in which 1 represent burn scars in the forest and 0 are areas that were not affected by the fire.

The dataset can be visualised in Fig 2. As can be seen the ground truth is noisy and also partially labelled, sometimes mislabelled as can be seen in Fig 2 (a). The dataset was curated by National Institute for Space Research (INPE) as part of the Quiemadas Project[INPE Brazil, 2018].

3.3 Inference and Training

The output map is subjected to a binary crossentropy loss which is backpropagated to train the AmazonNet model in an end-to-end fashion. Concatenated $[\mathbf{x}^i_{RGB}, \mathbf{x}^i_{NIR}]$ is input to the network. Adam's Optimizer with a starting learning rate of 0.0001 is used for minimizing the loss function. The model was fine-tuned for 50 epochs. The batch size of the training datasets was eight whereas, for validation datasets, the batch-size of 4 was chosen. Inbuilt callbacks functions, namely EarlyStopping, ReduceLROnPlateau, ModelCheckpoint were used for training our model.

Black contour indicates labelled data

(a) Visible   **(b) Near Infrared**   **(c) Labels**   **(d)Predicted Scars**   **(e)Binary Scar Map**
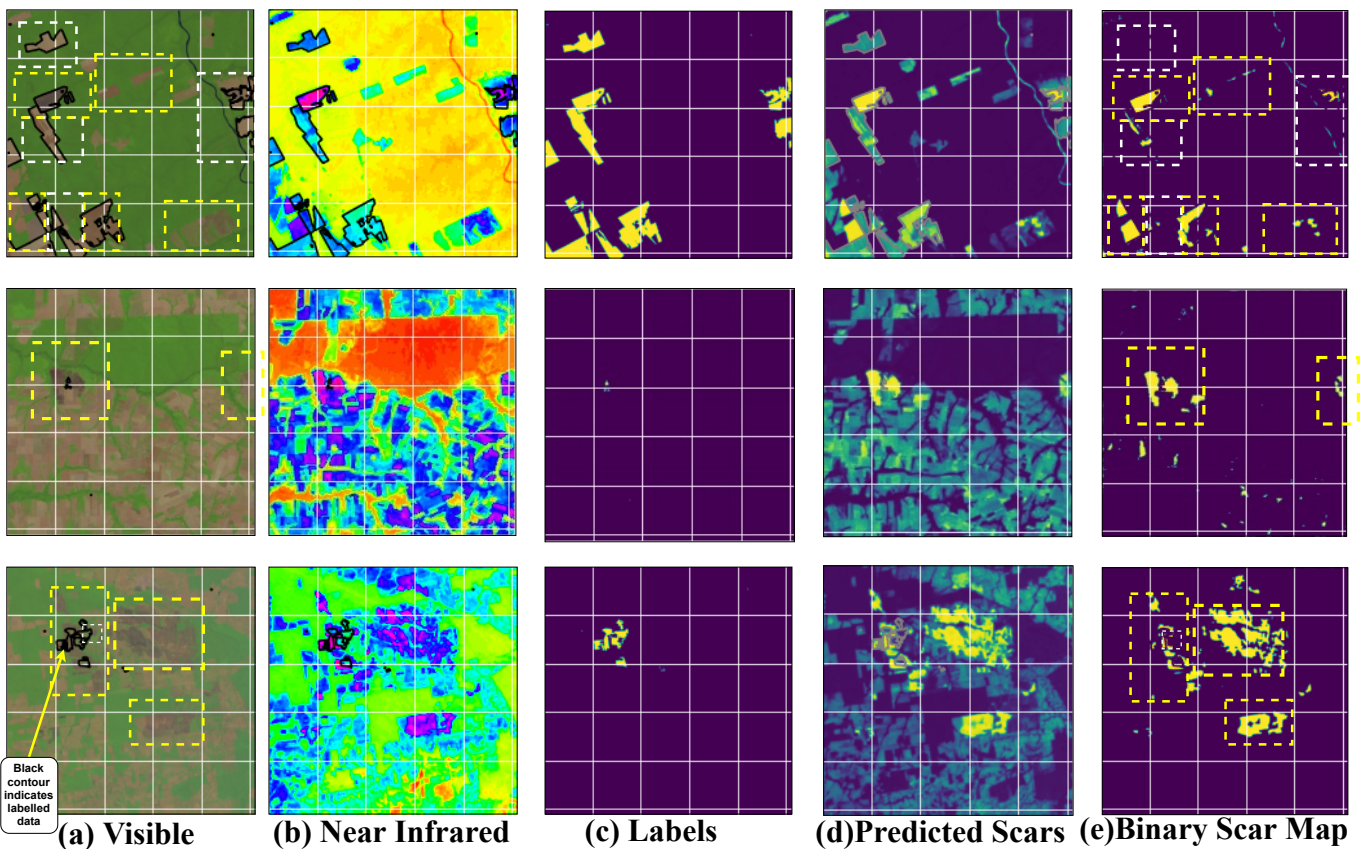
Figure 4: Results: The network correctly segments burn scars, rejecting incorrectly labelled spots and correctly identifying partially labelled or unlabelled samples. (a) Visible Channel (b) Near Infrared channel (false coloured hsv) (c) Available Labels (including partial/ incorrect labels) (d) Predicted Burn Scars (e) Binary Burn Scars. The black contour in (a) & (b) denote contour for labelled data(c) for easy visualisation. Yellow boxes denote burn scars which are correctly labelled or unlabelled. White boxes denote mislabelled/misidentified burn scars.

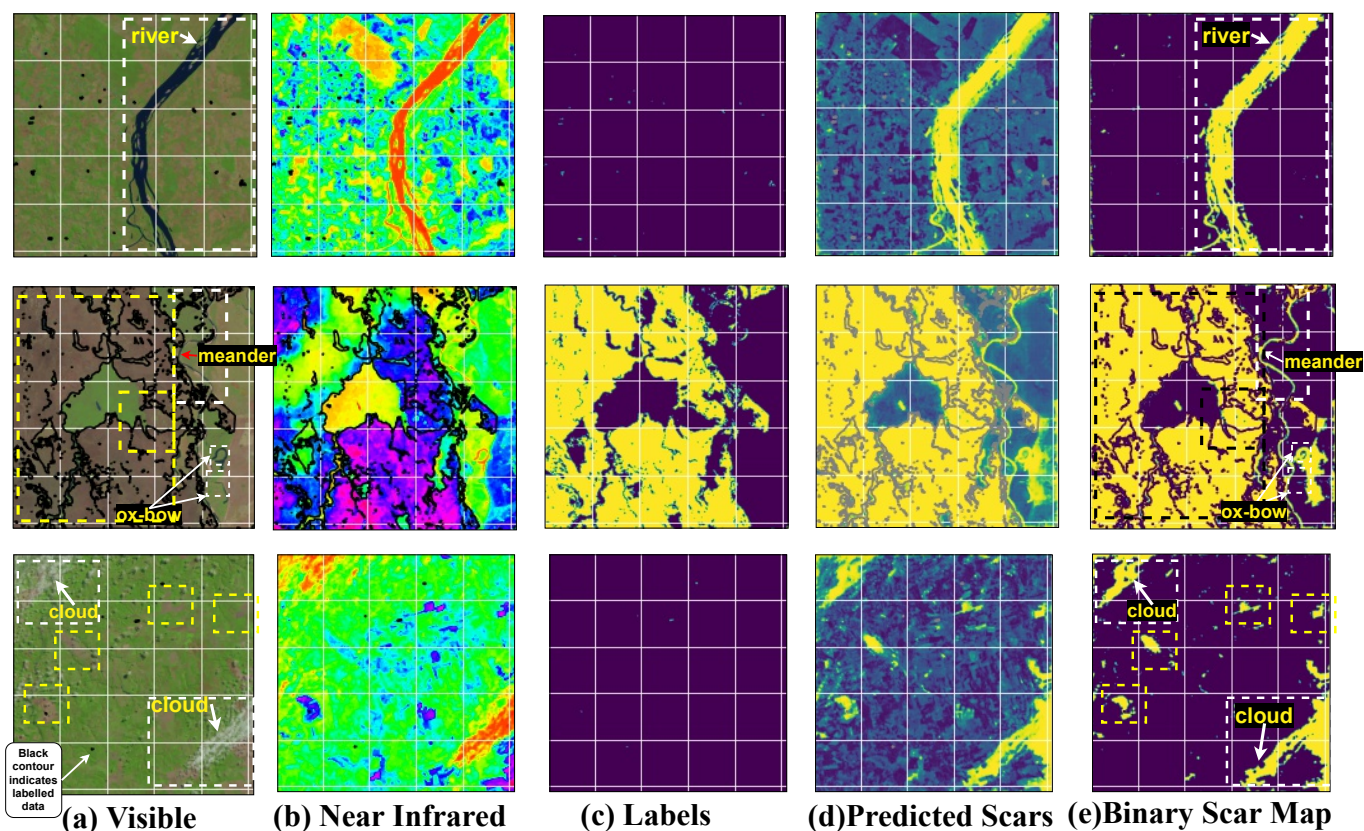(a) Visible   (b) Near Infrared   (c) Labels   (d)Predicted Scars   (e)Binary Scar Map

Figure 5: Minor defects in segmented burn scars: The network incorrectly segments rivers, meanders, ox-bow lakes and clouds as burn scars. The black contour in (a) & (b) denote contour for labelled data(c) for easy visualisation. Yellow boxes denote burn scars which are correctly labelled or unlabelled. White boxes denote mislabelled/misidentified burn scars. The defects are attributed to no segmented labels available and negligible instances of these features in the data-set.

4 Results

The model obtained a training accuracy of 69.51% & a validation accuracy of 63.33%. The results are presented in Fig 2 (f) & Fig 4 validating the efficacy of the utilising U-net based segmentation in burn scar identification. As can be seen in Fig 4, the network correctly identifies unlabelled fragmented burn scars (denoted as yellow-dash boxes) and deselects wrongly labelled areas (denoted as white-dash boxes) in the output binary map (correctly labelled outputs are highlighted as yellow and deselected labels as white).

Our network, however, fails to distinguish river and cloud patterns from burn scars as can be seen in Fig 5. Defects emerge when our network segments (a) river (b) meanders and ox-bow lake and (c) clouds as burn scar patterns.

It is interesting how in sample 2 and 3 in Fig 5, the network accurately segments the small fragmented burn scars but absolutely fails to reject these. This can be attributed primarily to (i) lack of any labelled examples and (ii) negligible samples containing the above geographical features in the dataset.

5 Conclusion and Future Work

We utilised a partially/mis-labelled dataset representing burn patterns in Amazon rainforest to propose U-net based segmentation network to correctly identify burn scars & reject

incorrect labels, demonstrating the effectiveness of AI in fragmented burn scar identification. We presented shortcomings & consider resolving these as future work.

Acknowledgements

References

[Badrinarayanan *et al.*, 2017] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017. 2

[Barbosa *et al.*, 2015] VL Barbosa, AD Jacon, PAC Moreira, A Lima, and CLE Oliveira. Detection of burned forest in amazonia using the normalized burn ratio (nbr) and linear spectral mixture model from landast 8 images. *Brasil. Simpósio Brasileiro de Sensoramiento Remoto*, 2015. 2

[Basso, 2019] Gustavo Basso. Overview of burnings in the vicinity of the br-163 highway in pará, northern brazil. *From Flickr with permission under Creative Commons Licence*, 2019. 1

[Chen *et al.*, 2016] Yushi Chen, Chunyang Li, Pedram Ghamisi, Chunyu Shi, and Yanfeng Gu. Deep fusion of hyperspectral and lidar data for thematic classification. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 3591–3594. IEEE, 2016. 2

[Chen *et al.*, 2017] Yushi Chen, Chunyang Li, Pedram Ghamisi, Xiuping Jia, and Yanfeng Gu. Deep fusion of remote sensing data for accurate classification. *IEEE Geoscience and Remote Sensing Letters*, 14(8):1253–1257, 2017. 2

[Chen *et al.*, 2018] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018. 2

[Cordts *et al.*, 2016] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016. 2

[Dalal and Triggs, 2005] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision & pattern recognition (CVPR'05)*, volume 1, pages 886–893. IEEE, 2005. 2

[Ess *et al.*, 2009] Andreas Ess, Tobias Müller, Helmut Grabner, and Luc J Van Gool. Segmentation-based urban traffic scene understanding. In *BMVC*, volume 1, page 2. Citeseer, 2009. 2

[Geiger *et al.*, 2012] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3354–3361. IEEE, 2012. 2

[Hartigan, 1975] John A Hartigan. *Clustering algorithms*. John Wiley & Sons, Inc., 1975. 2

[Iglovikov *et al.*, 2018] Vladimir Iglovikov, Selim Seferbekov, Alexander Buslaev, and Alexey Shvets. Ternausnetv2: Fully convolutional network for instance segmentation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018. 2

[INPE Brazil, 2018] National Institute for Space Research INPE Brazil. Quiemadas project. 2018. 3

[Juárez-Orozco *et al.*, 2017] SM Juárez-Orozco, C Siebe, and D Fernández y Fernández. Causes and effects of forest fires in tropical rainforests: a bibliometric approach. *Tropical Conservation Science*, 10:1940082917737207, 2017. 1

[Lanorte *et al.*, 2013] Antonio Lanorte, Maria Danese, Rosa Lasaponara, and Beniamino Murgante. Multiscale mapping of burn area and severity using multisensor satellite data and spatial autocorrelation analysis. *International Journal of Applied Earth Observation and Geoinformation*, 20:42–51, 2013. 2

[Lasaponara *et al.*, 2019] R Lasaponara, AM Proto, A Aromando, G Cardettini, V Varela, and M Danese. On the mapping of burned areas and burn severity using self organizing map and sentinel-2 data. *IEEE Geoscience and Remote Sensing Letters*, 2019. 2

[Lewis *et al.*, 2015] Simon L Lewis, David P Edwards, and David Galbraith. Increasing human dominance of tropical forests. *Science*, 349(6250):827–832, 2015. 1

[Liu *et al.*, 2020] Meng Liu, Sorin Popescu, and Lonesome Malambo. Feasibility of burned area mapping based on icesat-2 photon counting data. *Remote Sensing*, 12(1):24, 2020. 2

[Lowe, 2004] G Lowe. Sift-the scale invariant feature transform. *Int. J*, 2:91–110, 2004. 2

[Mohla *et al.*, 2020] Satyam Mohla, Shivam Pande, Biplab Banerjee, and Subhasis Chaudhuri. Fusatnet: Dual attention based spectrospatial multimodal fusion network for hyperspectral and lidar classification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020. 2

[Oberweger *et al.*, 2015] Markus Oberweger, Paul Wohlhart, and Vincent Lepetit. Hands deep in deep learning for hand pose estimation. *arXiv preprint arXiv:1502.06807*, 2015. 2

[Pereira *et al.*, 2017] Allan A Pereira, José Pereira, Renata Libonati, Duarte Oom, Alberto W Setzer, Fabiano Morelli, Fausto Machado-Silva, and Luis Marcelo Tavares De Carvalho. Burned area mapping in the brazilian savanna using a one-class support vector machine trained by active fires. *Remote Sensing*, 9(11):1161, 2017. 2

[Ronneberger *et al.*, 2015] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 2, 3

[Seferbekov *et al.*, 2018] Selim S Seferbekov, Vladimir Iglovikov, Alexander Buslaev, and Alexey Shvets. Feature pyramid network for multi-class land segmentation. In *CVPR Workshops*, pages 272–275, 2018. 2

[Shvets *et al.*, 2018] Alexey A Shvets, Alexander Rakhlin, Alexandr A Kalinin, and Vladimir I Iglovikov. Automatic instrument segmentation in robot-assisted surgery using deep learning. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 624–628. IEEE, 2018. 2

[Suga *et al.*, 2008] Akira Suga, Keita Fukuda, Tetsuya Takiguchi, and Yasuo Ariki. Object recognition and segmentation using sift and graph cuts. In *2008 19th International Conference on Pattern Recognition*, pages 1–4. IEEE, 2008. 2

[Tschandl *et al.*, 2019] Philipp Tschandl, Christoph Sinz, and Harald Kittler. Domain-specific classification-pretrained fully convolutional network encoders for skin lesion segmentation. *Computers in biology and medicine*, 104:111–116, 2019. 2

[Van Vliet *et al.*, 2012] Nathalie Van Vliet, Ole Mertz, Andreas Heinimann, Tobias Langanke, Unai Pascual, Birgit Schmook, Cristina Adams, Dietrich Schmidt-Vogt, Peter Messerli, Stephen Leisz, et al. Trends, drivers and impacts of changes in swidden cultivation in tropical forest-agriculture frontiers: a global assessment. *Global Environmental Change*, 22(2):418–429, 2012. 1

[WinklerPrins, 2006] AMGA WinklerPrins. Creating terra preta in home gardens: a preliminary assessment. *18th World Copngress of Soil Science*, 2006. 1

[Zhang and Wang, 2019] Zhengxin Zhang and Yunhong Wang. Jointnet: A common neural network for road and building extraction. *Remote Sensing*, 11(6):696, 2019. 2