Improving Wildfire Severity Classification of Deep Learning U-Nets from Satellite Images

(Article begins on next page)

Improving Wildfire Severity Classification of Deep Learning U-Nets from Satellite Images

Simone Monaco, Andrea Pasini, Daniele Apiletti, Luca Colomba, Paolo Garza, Elena Baralis
Department of Control and Computer Engineering, Politecnico di Torino, Italy
{name.surname}@polito.it

***Abstract*—Uncontrolled wildfires are dangerous events capable of harming people safety. To contrast their increasing impact in recent years, a key task is an accurate detection of the affected areas and their damage assessment from satellite images. Current state-of-the-art solutions address such problem through a double convolutional neural network able to automatically detect wildfires in satellite acquisitions and associate a damage index from a defined scale. However, such deep-learning model performance is strongly dependent on many factors. In this work, we specifically focus on a key parameter, i.e., the loss function, exploited in the underlying neural networks. Besides the state-of-the-art solutions based on the Dice-MSE, among the many loss functions proposed in literature, we focus on the Binary Cross-Entropy (BCE) and the Intersection over Union (IoU), as two representatives of the distribution-based and region-based categories, respectively. Experiments show that the BCE loss function coupled with a double-step U-Net architecture provides better results than current state-of-the-art solutions on a public labeled dataset of European wildfires.**

***Index Terms*—Burned Area delineation, Convolutional Neural Network, Deep learning, Supervised Learning, Semantic segmentation**

I. Introduction

European countries have been recently involved in an increasing number of wildfires. These events are causing large losses to people and the environment. The detection of the perimeter and the estimation of the severity level of the affected areas are fundamental for estimating the economical damage and planning the environment restoration.

Satellite images acquired by Sentinel2 can be used to automatically identify burned areas [1] and assess the damage severity without requiring human efforts. We can identify two different approaches to address this task: (i) assigning to each pixel of the satellite image a class label (i.e., burned or unharmed regions), or (ii) a numerical severity level measuring the damage. The former can be modeled with the well-known computer vision task called semantic segmentation, while the latter requires a regression methodology.

The current state-of-the-art approach proposes a convolutional neural network (CNN), called Double-Step U-Net [2], which involves both binary semantic segmentation and regression to obtain a damage-severity map. Specifically, each pixel is labeled with a numerical value representing the damage level (i.e., 0 - No damage, 1 - Negligible to slight damage, 2 - Moderately damaged, 3 - Highly Damaged, and 4 - Completely destroyed). The network is trained on manually labeled data gathered from the publicly available Copernicus Emergency Management Service dataset (Copernicus EMS) [3]

Previous works on semantic segmentation showed that the appropriate configuration of the CNN structure and the choice of loss functions have significant impacts on the final results [4], [5]. In this paper we focus on the inspection of different loss functions applied to the Double-Step U-Net. This neural network is composed of two different CNN blocks, each of them trained separately.

Our contribution consists of an in-depth analysis on the effects of different loss-functions applied to both modules, compared with the baseline results in [2]. We show that one of the analyzed loss-functions, the BCE-MSE, is capable of improving the state-of-the-art results in terms of RMSE.

Our paper is organized as follows. Section II presents the related works, while Section III discusses the neural network model and the proposed loss-functions. Finally, Section IV shows the experimental results and Section V draws conclusions.

II. Related work

In this section we firstly review previous works on wildfire prediction, then focus on state-of-the-art architectures for semantic segmentation and the adopted loss functions, highlighting the differences with the proposed technique.

Previous works typically monitor the evolution of wildfires during the event to support domain experts. Some of these techniques are implemented by means of deep learning models [6], [7]. Differently, in this paper we are more focused on automatic damage estimation after the event, by only exploiting post-event satellite images.

Most of the works in literature address damage severity estimation by means of specifically designed indexes, derived from remote sensors (UAV/satellites) or in-situ analyses. The Composite Burned Index (CBI) [8], the Normalized Burn Ratio (NBR) [9] and the delta Normalized Burnt Ratio (dNBR) [10] are some examples of these metrics. The estimation is made by analyzing such indexes, which are computed either on post-fire data or by comparing pre- and post-fire collected data. The weak point of these methodologies is that they significantly suffer from the different weather conditions at which satellite images are taken. Moreover, the usage of indexes to estimate the damage severity level typically requires the manual or semi-manual definition of predefined thresholds that are usually soil-dependent and cannot be easily set.

978-1-7281-6251-5/20/$31.00 ©2020 IEEE

The solution adopted in this work solves the previously mentioned issues by only requiring post-fire images and applying an automatic supervised prediction approach. Specifically, we apply a semantic segmentation model, combined with a regression one, to derive the final damage-severity maps. Many different semantic segmentation architectures have been proposed in literature [11]–[13], but the work in [2] shows that U-Net [14] is a valuable choice for addressing the wildfire damage-severity estimation task.

The Double-Step U-Net architecture [2] relies on the Dice loss function to learn predicting the boundaries of wildfires, and on the Mean Squared Error (MSE) function for estimating the final severity level. Many other different loss functions have been proposed in literature [15], and several works showed that a correct choice typically makes a real difference in the results [4]. In this work we focus on Binary Cross Entropy (BCE) and Intersection over Union (IoU) [15] aiming to obtain a more effective choice of the loss function for the Double-Step U-Net model.

III. Experimental Framework

In this section, we first describe the state-of-the-art neural network adopted in this work, then focus on the description of the proposed loss functions.

Damage severity prediction is modeled as a regression task, where each pixel in the satellite image must be associated with a real number in the severity range [0-4]. To this aim, the Double-Step U-Net architecture exploits the state-of-the-art U-Net model [14] as backbone, and combines two modules: (i) a Binary Classification U-Net, and (ii) a Regression U-Net. The Binary Classification U-Net distinguishes between burned and unburned areas, by assigning a binary label to each image pixel. Its output probability map is discretized with a step activation function, then provided to the Regression U-Net, which finally predicts the damage severity levels.

The two modules described so far are trained separately with different loss functions. In [2] the authors propose the usage of Dice for the Binary Classification U-Net and MSE (*Mean Square Error*) for the Regression U-Net. We denote this configuration with *Dice-MSE*.

In this work, we improve the performance of the deep learning model by analyzing the crucial loss-function choice [5]. Specifically, we intend to test the overall efficacy of the network both privileging either a *per-pixel* agreement with the ground-truth, or a *higher-order* correlation between pixels.

These two constraints can be enforced by means of the BCE and IoU loss functions, respectively. Indeed, BCE requires a *per-pixel* agreement by considering each pixel separately, and comparing its predicted likelihood of burned areas with respect to ground-truth. Instead, IoU, similarly to Dice, aims at obtaining a correct superposition of the predicted burned areas with respect to ground-truth (i.e., *higher-order* correlation), hence enforcing the CNN to be more accurate in recognizing the correct regions.

We inspect these two loss-functions by means of two representative configurations: (i) *BCE-MSE*, and (ii) *IoU-SoftIoU*. BCE-MSE applies the Binary cross-entropy for the Binary Classification U-Net, while MSE for the Regression one. Instead, IoU-SoftIoU exploits Intersection over Union for training both modules.

Since IoU is defined for classification tasks, we use a *SoftIoU* definition for the Regression neural network. Specifically, we generalize the IoU metric as follows.

Let Y_{GT}, Y_{PR} be the ground-truth severity map and the network estimation, respectively. Their values are first normalized in range [0, 1] by means of the sigmoid activation function. We denote the resulting matrices with $\widetilde{Y}_{GT}$ and $\widetilde{Y}_{PR}$. Afterwards, the SoftIoU can be computed with:

$$\textit{SoftIoU} = \frac{|\widetilde{Y}_{GT} \circ \widetilde{Y}_{PR}|}{|\widetilde{Y}_{GT} + \widetilde{Y}_{PR} - \widetilde{Y}_{GT} \circ \widetilde{Y}_{PR}|}\ , \tag{1}$$

where the symbol $\circ$ represents the element-wise product of the two matrices, and $|\cdot|$ is the sum of the matrix values. The numerator of Equation 1 represents the soft-intersection between predictions and ground truth, while the denominator represents the soft-union. The latter is computed by performing the sum of the two matrices and subtracting the value of the soft-intersection. The final loss function associated to this metric is then calculated as $L_{SoftIoU} = 1 - SoftIoU$.

IV. Experimental results

This section provides first the details of the dataset used to assess the proposed solution, then describes the evaluation process and the obtained results.

Dataset. Experiments have been conducted on the publicly available Copernicus Emergency Management Service dataset (Copernicus EMS) [3], which provides labeled satellite images with the areas hit by wildfires and the corresponding damage intensity, i.e. a severity level from 0 (no damage) to 4 (completely destroyed).

Satellite images acquired by Sentinel2 (L2A products) with 12 spectral bands can be used to identify burned areas [1] and classify the severity without requiring human efforts by means of an encoder-decoder convolutional neural network (Double-Step U-Net [2]). EMS data are collected from different European countries and separated into 7 folds, each of them containing elements that are geographically close to each other, hence possibly sharing similar morphology. These acquisitions have a size up to 5000×5000 pixels. To ease the application of the neural network, images are split into smaller tiles of 480×480 pixels, maintaining the spectral information along the 12 channels. Among these tiles, only the ones with at least one pixel with severity level greater than 0 are selected for cross-validation.

Evaluation methodology. In order to evaluate the performance of the different loss-function configurations, we performed a 7-fold cross validation among the previously defined dataset partitions. At each iteration, five folds are used as training set, one as validation set, and the remaining one as the test set. During the training phase, data augmentation techniques are performed on the training set by applying

TABLE I
PERFORMANCE FOR EACH WILDFIRE SEVERITY CLASS

	Overall Per-Class RMSE		
Severity	Dice-MSE	BCE-MSE	IoU-SoftIoU
0	0.34	**0.28**	1.22
1	1.05	**1.01**	1.32
2	1.07	0.84	**0.82**
3	1.01	0.75	**0.49**
4	1.28	**1.25**	1.72
tot	0.65	**0.54**	1.30

random transformations to the elements, in order to improve model generalization. The validation set is used to assess the model's performance at each training epoch for the early stopping regularization criteria.

Results. Results were obtained using CPUs on BigData@PoliTO Cluster [16] to perform the computation. Table I provides a comparison of the cross-validation results for the three configurations described in Section III. Each line corresponds to a specific severity level ([0-4]) and its scores are computed by considering only the pixels with that level as ground truth value. The Root Mean Square Error (RMSE) is then computed by comparing the ground-truth pixels with the predictions, averaging the results for each cross-validation fold.

The results show that BCE-MSE achieves higher performances w.r.t. the state of the art for all the severity levels. For class 2 and 3, the IoU-SoftIoU configuration achieves the best results. However, it also shows the worst scores for all the other three classes. On average (last line of Table I), the configuration with best scores is BCE-MSE (0.54), followed by DICE-MSE (0.65).

These results show that the metrics involving higher-order relationships between pixels (see Section III) are less capable of reducing the final RMSE.

V. Conclusion and Future Works

In this paper we addressed the task of predicting damage severity levels after wildfire events. We adopted a deep learning state-of-the-art technique to obtain the predictions and fine-tuned the choice of the loss functions. We learned from the results that the proposed configuration with BCE-MSE achieves better results with respect to the state-of-the-art.

Other loss-functions, together with the ones presented in this paper, will be considered in the future to inspect the efficacy of compound loss functions and reach higher performances.

Acknowledgment

The research leading to these results has been partially supported by the SmartData@PoliTO center for Big Data and Machine Learning technologies, and the HPC@PoliTO center for High Performance Computing. The authors are grateful to Moreno La Quatra for his help in exploiting the HPC resources.

References

[1] D. Stavrakoudis, T. Katagis, C. Minakou, and I. Z. Gitas, "Towards a fully automatic processing chain for operationally mapping burned areas countrywide exploiting sentinel-2 imagery," in *RSCy2019*, vol. 11174, p. 1117405, 2019.

[2] 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," *Applied Sciences*, vol. 10, no. 12, p. 4332, 2020.

[3] *European Union. Copernicus Sentinel-2 Mission. 2020.*, 2009 (accessed November 9, 2020).

[4] S. Jadon, O. P. Leary, I. Pan, T. J. Harder, D. W. Wright, L. H. Merck, and D. L. Merck, "A comparative study of 2d image segmentation algorithms for traumatic brain lesions using ct data from the protectiii multicenter clinical trial," in *Medical Imaging 2020: Imaging Informatics for Healthcare, Research, and Applications*, vol. 11318, p. 113180Q, International Society for Optics and Photonics, 2020.

[5] J. Ma, "Segmentation loss odyssey," *arXiv preprint arXiv:2005.13449*, 2020.

[6] Y. Ban, P. Zhang, A. Nascetti, A. R. Bevington, and M. A. Wulder, "Near real-time wildfire progression monitoring with sentinel-1 sar time series and deep learning," *Scientific Reports*, vol. 10, no. 1, pp. 1–15, 2020.

[7] M. M. Pinto, R. Libonati, R. M. Trigo, I. F. Trigo, and C. C. DaCamara, "A deep learning approach for mapping and dating burned areas using temporal sequences of satellite images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 160, pp. 260–274, 2020.

[8] C. H. Key and N. C. Benson, "Landscape assessment (la)," *In: Lutes, Duncan C.; Keane, Robert E.; Caratti, John F.; Key, Carl H.; Benson, Nathan C.; Sutherland, Steve; Gangi, Larry J. 2006. FIREMON: Fire effects monitoring and inventory system. Gen. Tech. Rep. RMRS-GTR-164-CD. Fort Collins, CO: US Department of Agriculture, Forest Service, Rocky Mountain Research Station. p. LA-1-55*, vol. 164, 2006.

[9] G. Navarro, I. Caballero, G. Silva, P.-C. Parra, Á. Vázquez, and R. Caldeira, "Evaluation of forest fire on madeira island using sentinel-2a msi imagery," *International Journal of Applied Earth Observation and Geoinformation*, vol. 58, pp. 97–106, 2017.

[10] J. D. Miller and A. E. Thode, "Quantifying burn severity in a heterogeneous landscape with a relative version of the delta normalized burn ratio (dnbr)," *Remote Sensing of Environment*, vol. 109, no. 1, pp. 66–80, 2007.

[11] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[12] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.

[13] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2881–2890, 2017.

[14] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*, pp. 234–241, Springer, 2015.

[15] S. Jadon, "A survey of loss functions for semantic segmentation," *arXiv preprint arXiv:2006.14822*, 2020.

[16] SmartData@PoliTO, "Bigdata cluster: Computing facilities." Available at "https://smartdata.polito.it/computing-facilities/".