Fire and Smoke Detection in Aerial Images

Bernardo Amaral[1]
bernardo.m.amaral@tecnico.ulisboa.pt
Alexandre Bernardino[1]
alex@isr.tecnico.ulisboa.pt
Catarina Barata[1]
ana.c.fidalgo.barata@tecnico.ulisboa.pt

[1] Institute for Systems and Robotics, ISR
Instituto Superior Tecnico, IST
Universidade de Lisboa,
Lisbon, PT

Abstract

Fire and smoke detection in images using image processing and deep learning techniques has proven to be a topic of high interest. Recent methods for object recognition and localization using deep learning achieve very reliable results. Methods that rely on large amounts of data with the respective annotations, which are both expensive and often subjective. To overcome this limitation, we study and implement a method to detect fire and smoke zones using only weakly supervised methods. We train a convolutional neural network based model (CNN) for object classification that relies only on image-level labels and still can learn to predict the location of fire and smoke. We demonstrate that the model is able to localize the discriminative image regions of fire and smoke despite not being trained for them.

1 Introduction

Forest fires are a scourge that every year destroy thousands of hectares of forest around the world. Forest fires have a series of effects on both the burned area and the underlying areas. The consequences go beyond the visible effects on nature and society such as the destruction of material assets and the effect on vegetation. The entire ecosystem is threatened, from fauna and flora to loss of biodiversity, soil degradation and erosion, to CO^2 emissions. For this reason it is extremely urgent to take measures to mitigate these dangers and reduce the risk of forest fires. According to [1], in Portugal on the last three years (2017-2019) about 630 thousand hectares were burnt due to forest fires.

A possible approach to create a wiser firefighting and minimize this threat is through the use of manned or unmanned aerial vehicles that collect real-time visual information from the fire site. Then the information can be feed to automatic systems that are able to locate regions of fire and smoke. This poses a considerable challenge, since neither fire nor smoke have a well-defined shape and a constant color.

The usual methods of locating/identifying objects using deep learning are trained on a large amount of fully annotated data, which means that in each image of the dataset there must be an annotation of where each class is present in the image, for example through the use of bounding boxes or pixel-level masking. Though, the creation of such annotations is very expensive and, especially in case of fire and smoke, it can be very subjective, since there are no rigid limits.

To overcome this situation we propose to use a weakly supervised method where the only needed annotation is at the image-level. Which means that each image in the dataset has only the information if each class is present in the image or not, there is no information of its location. With this approach, we can train a CNN model to classify the presence or absence of fire and smoke and still predict their correspondent localization.

2 Related work

The work done on fire and smoke detection based on computer vision presented a wide variety of methods. The big majority of them are based on color, motion, spatial and temporal features. This characteristics are very specific for fire compared to other objects. Most of them follow a common pipeline, first find moving pixels using background subtraction and then apply a color model to find fire color regions. The approach is to create a mathematical based model, defining a sub-space on a color space that represents all the fire-colored pixels in the image [2]. On this line, *Wang et al*. [3] proposed a method based on a Gaussian model learned in the YCbCr color space. *Uğur et al*. [4] added to the base pipeline a wavelet-based model of fire's frequency signature, with the idea that flames flicker with a characteristic frequency. Also, *Chine et al*. [5] added texture analysis to create *Bowfire* and prevent false positives resulted from the single use of color models. Methods using motion tracking are limited since they only work properly with fixed cameras, in surveillance scenarios [6]. They are therefore not compatible for use with aerial vehicles.

These methods depend heavily on the features delimited by the authors, which may make them too specific for a certain purpose. On the other hand, methods with deep learning, automatically learn which features are best for the given problem. This is why deep learning methods outperform them. On [7] the authors do a comparative analysis between color-model based methods versus deep learning methods. They use a logistic regression model, which is a very simple deep learning method and yet it is the one that obtains the best overall performance compared to all colour-based models. They also prove the robustness of these methods for colour changes and the presence of smoke.

Thus, considering the superiority of the deep learning methods compared to those based on colour, several authors presented methods using CNN to detect fire and/or smoke [8]. Still following the idea of using surveillance cameras, *K. Muhammad* [9] trained a SqueezeNet model for fire detection, localization, and semantic understanding of the scene of the fire. *Q Zhang et al*. [10] trained a Faster R-CNN to detect smoke in wildland forest fire by creating synthetic images with the addition of synthetic smoke to normal forest images. Also *Q. Zhang et al*. [11] propose a method to detect and localize fire using a CNN by using image patch division.

The lack of a good public accessible dataset for fire and smoke makes it hard to develop a good deep learning technique. For this reason some of the previous methods that were trained with small datasets, and even thought they used fully annotated images, might lack some robustness. In the method we propose, we can use the few datasets available [12] and complement them with as many images as we want without much effort to label them. Then we can build a robust and reliable method to detect and locate fire and smoke.

3 Methodology

Our approach is based on the powerful ability to locate objects from the convolutional layers in a CNN, using only a set of images that are annotated at image-level. We chosen a *VGG19* as our base model. We then removed the fully connected layers because they destroy the spatial integrity kept in the convolutional layers and added a *Global Average Pooling* (GAP) and *Sigmoid* layers. The *Sigmoid* layer makes the model behave as a common one-vs-all classifier for each class. Therefore, for each image we will always have the prediction if there is fire, smoke, both at the same time or none at all. This is very important since fire and smoke are highly correlated. *Figure 1* summarizes our proposal.

To obtain the location, we applied the methodology proposed in [13] to our specific case. The idea is to create a Class Activation Map (CAM) that can be used to localize the network's attention on the input images for fire and smoke even though the networks have only been trained on image-level labels. This is done by gathering the information from the features maps on the last convolutional layer. Thus, we had a *GAP* layer before the last convolutional layer so that we can weight the importance of each feature map for the predicted class(es). Then, we do a weighted sum of those feature maps according to the predicted class to produce the CAM. Hence, $H_c(x,y) = \sum_{i=1}^{n} w_i^c f_i(x,y)$ where H represents the CAM with the predicted location and w_i^c is the weight of the activation of the i^{th}

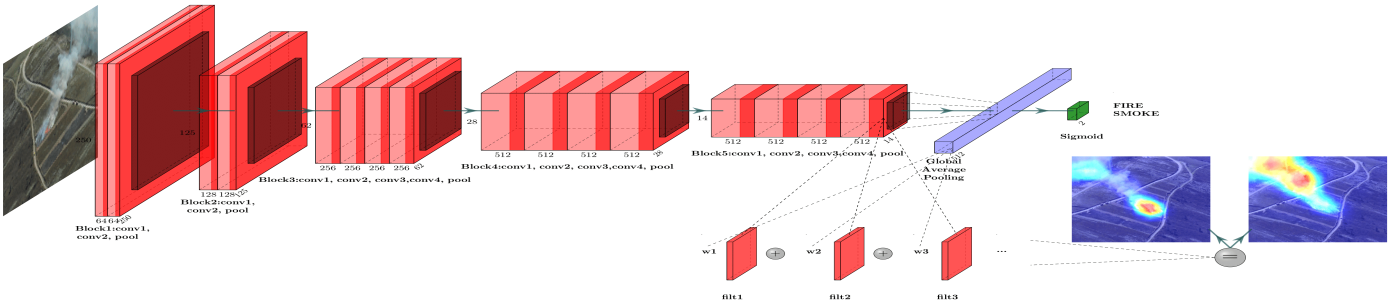

Figure 1: Global architecture

feature map $f(x,y)$ for the predicted class c.

By doing so, CAM behaves as a heatmap highlighting the areas in the image where it is more probable to be fire and/or smoke.

4 Experimental setup

To train the network, we build our own data set. As a starting point the dataset of [12] was used since it contains good examples of forest fires as well as controlled fire frames. In order to complement it, we added images from aerial views, images with only fire, images with only smoke and images of negative cases (sunsets, very orange scenes, clouds, etc.). Gathered from the web and from the Firefront project [14]. In the end, the data set is made up of 1770 images, of which 80% are for training, 12.5% are for validation and 7.5% are for testing.

Our model is exclusively trained for classification purposes and as it behaves as a one-vs-all classifier we used a binary cross-entropy loss. At any stage of the model training we use any location related loss.

5 Results

To evaluate our method, we tested our approach in terms of classification and segmentation.

For classification, we tested on our test set. The metrics presented are at the image level, the network predicts the presence of fire and smoke in the whole image. We achieved an accuracy of 92% for fire and 91% for smoke, which is a good result taking into account that we never said to the network what exactly is fire or smoke.

Regarding the segmentation, we tested only for the case of fire, using the images and their ground truth from [12]. To do this, we had to transform the CAMs into binary masks. The CAM behave like a heatmap with values between 0 and 1 in which areas with values close to 1 are more likely to belong to the class and vice versa. Thus, we have created a threshold where any value above is considered to belong to the class. Several values were tested taking into account the mean Intersection-Over-Union (IoU) . In the end we obtained a value of 0.35 for the threshold and a corresponding IoU of 0.575. To note that this a very good value taking into account that the ground truth masks are very detailed while our method, although accurate, is not so precise, is more rounded. If we apply a small dilation on the ground truth masks, we can do a more equitable comparison and achieve a mean IoU of 0.61.

6 Conclusions

Our method provides a solution for developing location capabilities when there is a lack of available data, in the specific case applied to fire and smoke. We have shown that even with only image level labels, we have been able to build a normal classification network to predict the location and have saved a lot of time on labeling.

We are aware that the heatmaps produced are not as sharp as an output of a segmentation network but are accurate in terms of location. Consequently, in a future work we could sharpen the heatmaps with the use of different available method's, for example Conditional Random Field (CRF).

Acknowledgments

This work was supported by FCT with the LARSyS - FCT Project UIDB/50009/2020 and project FIREFRONT (PCIF/SSI/0096/2017).

References

[1] PORDATA. Portugal - environment, energy and territory - territory and environmental protection - rural fires and burnt area, Accessed 10/08/2020. `www.pordata.pt/en/Portugal/Rural+fires+and+burnt+area+-+Mainland-1192`.

[2] Celik Turgay and Hasan Demirel. Fire detection in video sequences using a generic color model. *Fire safety journal*, pages 147–158, 2009.

[3] Wang De chang *et al*. Adaptive flame detection using randomness testing and robust features. *Fire safety journal 55*, pages 116–125, 2013.

[4] Töreyin B. Uğur *et al*. Computer vision based method for real-time fire and flame detection. *Pattern recognition letters*, pages 49–58, 2006.

[5] Daniel YT Chino *et al*. Bowfire: detection of fire in still images by integrating pixel color and texture analysis. *28th SIBGRAPI Conference on Graphics, Patterns and Images. IEEE*, pages 95–102, 2015.

[6] Poobalan Kumarguru and Siau-Chuin Liew. Fire detection algorithm using image processing techniques. *International Conference on Artificial Intelligence and Computer Science*, pages 160–168, 2015.

[7] Tom Toulouse *et al*. Automatic fire pixel detection using image processing: a comparative analysis of rule-based and machine learning-based methods. *Signal, Image and Video Processing*, pages 647–654, 2016.

[8] Panagiotis Barmpoutis *et al*. Fire detection from images using faster r-cnn and multidimensional texture analysis. *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8301–8305, 2019.

[9] Khan Muhammad *et al*. Efficient deep cnn-based fire detection and localization in video surveillance applications. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pages 1419–1434, 2018.

[10] Qi xing Zhang *et al*. Wildland forest fire smoke detection based on faster r-cnn using synthetic smoke images. *Procedia engineering*, pages 441–446, 2018.

[11] Qingjie Zhang *et al*. Deep convolutional neural networks for forest fire detection. *International Forum on Management, Education and Information Technology Application. Atlantis Press*, 2016.

[12] Tom Toulouse *et al*. Computer vision for wildfire research: An evolving image dataset for processing and analysis. *Fire Safety Journal 92*, pages 188–194, 2017.

[13] Bolei Zhou *et al*. Learning deep features for discriminative localization. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2921–2929, 2016.

[14] FIREFRONT. Real-time forest fire mapping and spread forecast using unmanned aerial vehicles. `www.firefront.pt/`.