Fire and Smoke Detection in Aerial Images

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

Introduction

Forest fires are a scourge that every year destroy thousands of hectares of forest around the world.

A possible approach to create a wiser firefighting is through the use of aerial vehicles that collect real-time visual information from the fire site and feed it to automatic systems that are able to locate regions of fire and smoke.

The usual methods of locating/identifying objects using deep learning rely on a large number of **expensive** and **subjective** fully supervised **annotations**.

Also, fire and smoke have a very **irregular shape**.

To overcome this situation we propose to use a **weakly supervised method** to **detect fire and smoke**, where the only needed annotation is at the **image-level**.

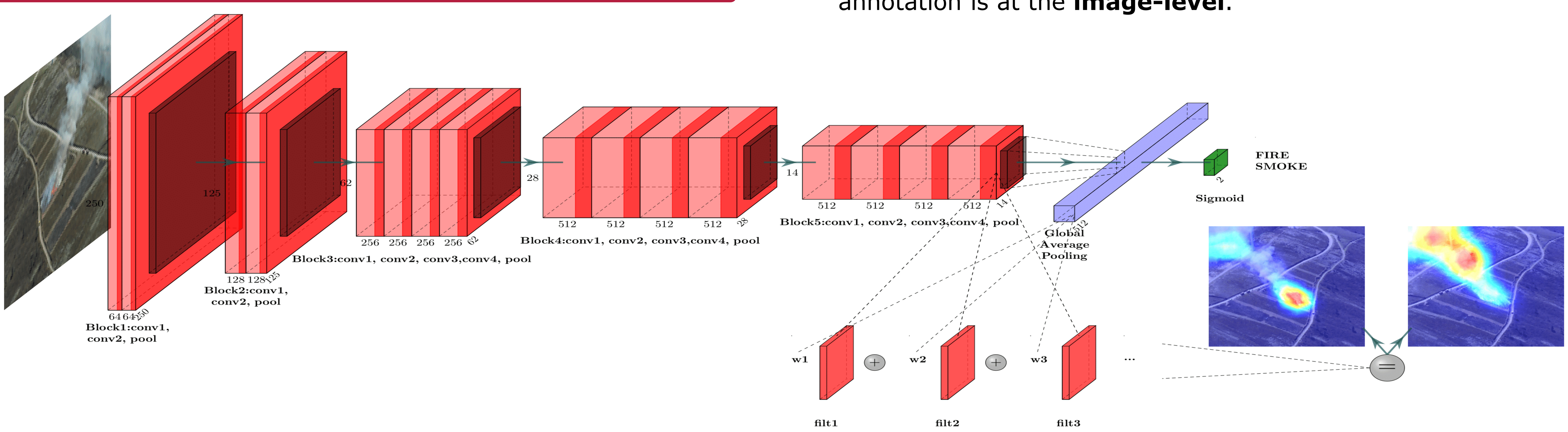


Figure 1 – Global architecture

Methodology

We train a classification CNN with **only image level labels** and then apply the CAM methodology proposed in [1] to our specific case.

We use a VGG19 with only convolution layers and added a Global Average Pooling (GAP) and a Sigmoid layers.

With the GAP layer we can **weight the importance** of each feature map on the last convolutional layer for the predicted class(es).

Then, we do **weighted sum of the feature maps** according to the predicted class to **produce the CAM**:

$$S_c = \sum_{i=1}^{n} w_i^c \, f_i(x,y).$$

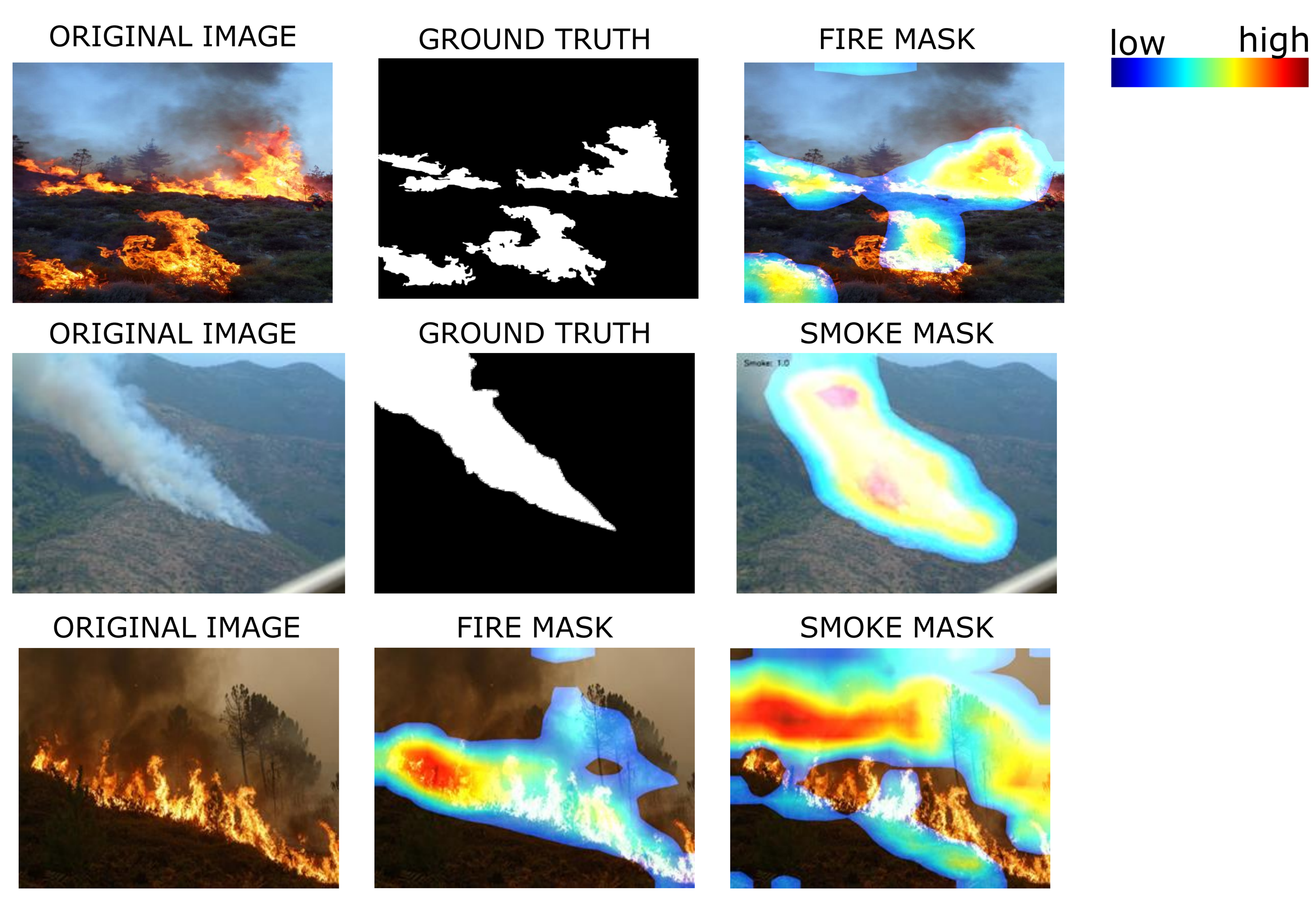
Where $f_i(x,y)$ represents the activation of the i^{th} feature map and w_i^c is the weight of that map to the desired output.

The CAM behaves as a **heatmap** highlighting the areas in the image where it is **more probable to be fire and/or smoke**.

Results

The classification results are at the image level. For segmentation, the mean IoU is computed by transforming the CAM into a binary mask, using a confidante threshold.

			Accuracy (%)
Classification	Fire		92
	Smoke		91
			Mean IoU
Segmentation	Fire	Ground truth	0.57
		Dilated Ground truth	0.61



[1] 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

Acknowledgments

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