Fire and Smoke recognition in crowdsourced images with YOLO networks

Ana Madeira
abreumadeira.ana@gmail.com
Catarina Silva
catarina@dei.uc.pt
Alberto Cardoso
alberto@dei.uc.pt
Bernardete Ribeiro
bribeiro@dei.uc.pt

University of Coimbra
CISUC - Centro de Informática e Sistemas
FCTUC-DEI - Departamento de Engenharia Informática
Coimbra, Portugal

Abstract

The early detection of a fire can largely mitigate harmful consequences. With the improvement in image quality, it is now possible to develop intelligent systems for visually detecting forest fires. An intelligent system for fire detection was implemented based on deep learning techniques for image object detection. As part of the fire detection approach development, different datasets are proposed to train and evaluate the YOLO models, specific to the fire and smoke recognition problem. The proposed Fire/Smoke annotated datasets can be used in future smoke, and fire detection research. Results show that a YOLOv4 one-stage detector can be used for image fire and smoke detection tasks, trained using manually annotated datasets and applied to a real application using crowdsourced data.

1 Introduction

As a way for people to report fires detected using their smartphones, the FireLoc project [1] of the Foundation of Science and Technology (FCT) [2] is set up as an alternative way of reporting fires. This project is based on voluntary contributions and aims to develop a system in which, through a smartphone application, users can send photos of fire taken with their smartphone camera. If present, smoke and fire are recognized in the images submitted, and the forest fire can be located on a map. This information is then sent to a server. The developed system will correspond to the submission validation module and validate whether there is fire or smoke in each contribution.

In this paper, the main focus is the development of an intelligent system for fire and smoke detection. integrated in the FireLoc application, using the proposed post-processing steps to obtain the image classification results, identifying whether user submissions are valid, i.e., whether they contain smoke or fire.

2 Related work

Recent studies show advantages in considering the localization as well as the classification of existing objects in an image as part of object detection problems [7]. The models used for object detection can be divided into two categories: Two-stage detection frameworks and One-stage (Unified) detection frameworks. In the first, the process is divided into two phases. First, there is a proposal for candidate regions of the image that may contain objects to be detected [6]. The classification is then made based on the first result, fine-tuning the regions, discarding false positives (for example, Faster R-CNN). One-stage detection frameworks perform the process at once, without the initial region proposal step and therefore allow a single model to be used, predicting the bounding boxes that contain the objects present, as well as the probabilities of these belonging to the classes considered [5] (for example the YOLO models). The YOLOv4 models analyze the image's features using different resolutions, maintaining the original image's height/width ratio. These models manage, by adapting the size of the initial anchors to the specific dataset, to detect objects of various scales in the images [6] and allow the correct detection of overlapping objects of different classes [3]. For this reason, they present advantages when used for this problem, since it is common to have the presence of smoke in the images where there is fire. However, these models have some disadvantages, such as the need to have a considerable amount of annotated images to obtain good results. As such, to solve the lack of annotated data, two datasets are proposed with the annotation of Fire and Smoke class objects for training and model evaluation. These datasets were used to optimize the results in the context of forest fires. In addition, the transfer learning technique was also used, with a pre-trained model with the Imagenet dataset. The initial weights resulting from the pre-training were kept, responsible for the extraction of more low-level features. The last layers of the model, responsible for the extraction of features specific to the problem, were retrained [6]. The use of transfer learning makes the training process less time-consuming and improves the model's learning capacity.

[1] Project PCIF / MPG / 0128/2017, FireLoc - Where's the Fire? - Identification, positioning, and monitoring forest fires with crowdsourced data

3 Proposed approach

For the development of the fire and smoke detection system, an object detection approach was adopted, using YOLOv4 [3] models. These models detect the specific location of fire and smoke in the images. Therefore, they require an indication of where the objects are present within the image, using bounding boxes. To perform the manual annotation of the training and testing datasets proposed, according to the YOLO annotation format, the tool LabelImg was used.

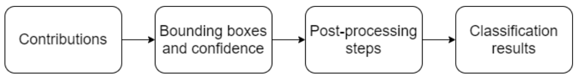

Figure 1: Detection system steps

Figure 1 shows the sequence of steps necessary to detect fire and smoke in the images submitted by the application users. The model first identifies the parts of the images that contain fire or smoke with bounding boxes, and the corresponding confidence score associated with each detected object. The classification results are then obtained with the post-processing step, which allows integration with the FireLoc system. For the classification results, the **Fire**, **Smoke**, and **Neutral** classes are considered. The images in which fire is detected belong to class **Fire**, and the images in which smoke is detected, and no fire is detected belong to class **Smoke**. The remaining images, in which no object is detected, belong to the Neutral class.

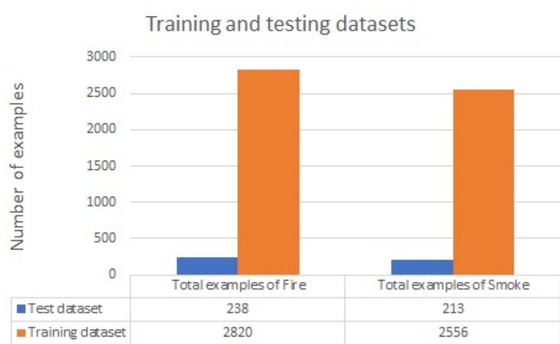

Figure 2: Number of object examples in each dataset

For the training of these models, an open-source dataset from [1] was used. This dataset contains an equal amount of images from each class: **Smoke**, **Fire**, and **Neutral**). It contains 1000 images of each category and is named *Fire-Smoke-Dataset*. To complement this training dataset