Classifying Fire and Non-fire images using CNN

Juliana Restrepo, Pedro Botero and Andrea Carvajal

May 29, 2022

1 Problem Statement

"Computer Vision has a dual goal. From the biological science point of view, computer vision aims to come up with computational models of the human visual system. From the engineering point of view, computer vision aims to build autonomous systems which could perform some of the tasks which the human visual system can perform (and even surpass it in many cases)"(Bhatt *et al.*, 2021). Computer vision is a field that has gained recognition among the scientific community over the past few years. Disregard being a considerably new area of study, it has presented major advances since Larry Roberts' contribution in 1963. Roberts is considered the father of computer vision thanks to his Ph.D. thesis "Machine Perception of Three-Dimensional Solids"(Roberts, 1963). This milestone encouraged other scientists such as David Marr to propose their point of view on related topics, for example, scene understanding.

There are many key tasks in computer vision. The present project will be focusing on image classification, which is the task of classifying images into a set of predefined classes based on a test set of sample images that have been classified previously (Bandyopadhyay, 2022). This task has numerous applications including object identification in satellite images, medical imaging, brake light detection and traffic control systems (Boesch, 2022). This particular project will be using image claasification with the aim of identifying a potential fire from images by classifying them into fire and non-fire categories.

1.1 Research question

Can computer vision be used to identify a potential fire with 85% of accuracy in frontal images of different scenarios taken with a camera?

2 Objectives

2.1 General Objetive

Propose a model that classifies images with fire and smoke from non-fire images.

2.2 Specific Objectives

- Perform a literature review in order to understand image classification.

- Classify images that contain fire and regular images.
- Obtain an accepatable level of accuracy in the classification.

3 State of the art

Bhatt *et al.* (2021) They present in their paper the primary taxonomy and deep CNN (convolution neural network) architectures. The authors propose eight categories in order to divide the multiple recent developments in CNN: spatial exploitation, multi-path, depth, breadth, dimension, channel boosting, feature-map exploitation, and attention-based CNN. The principal contribution of this work is the comparison between architectural evolutions in CNN, and their strengths and weaknesses. This research recognizes CNN as a structure that has demonstrated exceptional image classification and, as this is the objective of the project, this was the chosen architecture.

On the other hand, Muhammad *et al.* (2018) suggest an interesting application of CNN in fire detection systems. The authors assure this neural network will substantially improve detection accuracy, which will result in reducing fire disasters and the ecological and social impacts they have. This previous work was useful as a guide and inspiration to the development of the own project.

4 Methodology of the investigation

In this section, the methodology that will be implemented to explore and solve each step of the problem will be thoroughly described. The code and the README file can be found in `https://github.com/jrestrepot/Computer-vision-for-fire-detection`

4.1 Images Preprocessing

Each image is normalized and resized to be 300x300 pixels. In addition, 50% of the images are horizontally flipped. The dataset is then divided into two sub-groups: training and validation set (70%), and **test set** (30%). The training and validation set is further divided into a **training set** (75%) and a **validation set** (25%).

4.2 Convolutional Neural Network

A convolutional neural network is a type of artificial neural network that has been optimized to process pixel data. It is most commonly used for image analysis and image classification problems. CNNs are divided into three different layers: convolutional layers, which help convert the image into a feature map with the use of filters and kernels; pooling layers, which are usually found between convolutional layers and have the purpose to reduce the size of the feature map and finally, the fully connected layers, which connect every neuron in each layer with those in other layers.

The architecture of the neural network implemented in this project is the following:

- Convolutional layer, which receives 3 channels (the RGB channels of each image), returns 5 filters and has a kernel of 5. The resulting dimensions of the tensors are 296x296.

- Max pooling layer, which uses a sliding window of 2x2, reducing the dimensionality of the tensors to 148x148.
- Convolutional layer, which receives 5 channels, returns 10 filters and has a kernel of 5. The resulting dimensions of the tensors are 144x144.
- Max pooling layer, which uses a sliding window of 2x2, reducing the dimensionality of the tensors to 72x72.
- Linear layer with 72*72*10 features and returns 100 features.
- Linear layer with 100 features and returns 10 features.
- Linear layer with 10 features and returns 1 feature.

The activation function of the layers is the Relu function, except for the last linear layer which has a Sigmoid activation function. For this reason, the last layer returns the **probability** of an image belonging to the fire class.

Pytorch library was selected because it is more begginer-friendly and it is faster than Tensorflow Chirodea *et al.* (2021).

4.3 Validation and deployment

To train the model, the training set was used. Then, to test its performance, the validation set was used. The recall, precision, accuracy score, f1 score and AUC score were computed with the predictions of the validation set. The accuracy was calculated as it is a classic and commonly use metric. Nevertheless, the recall and precission were taken into account because both the false negatives and false positives are significant to evaluate the performance. The false negatives could endanger people by overlooking a potential fire, and a false positive could unnecessarily occupy resources and personnel of authorities such as firefighters. In addition, the f1 score is relevant due to the fact it is the harmonic mean between precision and recall. Finally, the AUC was computed to test the quality of the model as a classifier.
The model was tuned in order to improve all metrics, and then it was set for "deployment".
In the deployment stage, the tuned model was tested with the test set. This stage was thought to detect overfitting in the model.

5 Data analysis

During the NASA Space Apps Challenge in 2018, a team created this fire dataset, and its owner is Ahmed Saied. The images' source is Google images, and the collection method was regular image scraping and then an examination by team members.

The dataset was retrieved from Kaggle, and it consists of 755 outdoor fire images and 244 non-fire images. The images that contain fire can also have heavy smoke, the ones that don't consist mainly of forests, trees, grass, rivers, people, foggy forests, lakes, animals, roads, and waterfalls. It was last updated on 25/02/2020.

The classes in the data are imbalanced, with 24% of the images being in the non-fire class and the remaining 76% being in the fire class. The images have all a .png format, but their sizes are different. The mean height and width of the matrices derived from the images are 738.21 and 1153.18, respectively, while the median height and width are 534 and 860, respectively.

5.1 Exploratory data analysis

5.1.1 Raw comparison

First, three images from each class were randomly selected to have a first approach to the data.

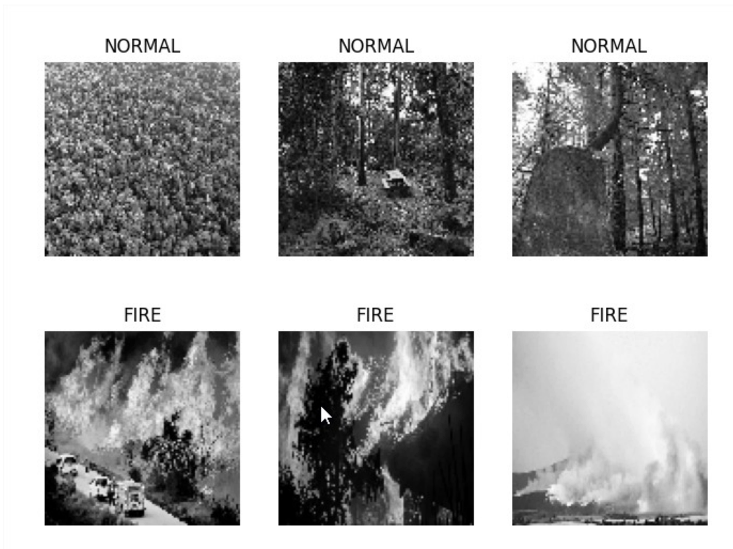

Figure 1: Example of images in both classes (fire and non-fire)

5.1.2 Average images

Each image was translated into its pixel values, then all the files are represented by a matrix of size $n \times m$, where n is the number of observations and m is the number of pixels. All this in order to find certain useful measures of the dataset.

Figure 2: Average images of class fire

Figure 3: Average images of class non-fire

From the average images, it can be concluded that the images with fire tend to have a darker tint around the borders but it is not significantly notorious.

5.1.3 Contrast between average images

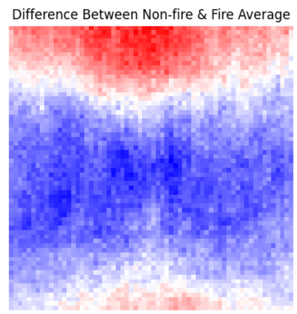

Figure 4: Contrast between fire and non-fire average images

5.1.4 Variability

Figure 5: Standard deviation of class fire

Figure 6: Standard deviation of class non-fire

In these images, the lighter area indicates higher variability. Therefore, the class fire has more variability even though there is not a remarkable difference.

5.1.5 Eigenimages

Principal component analysis was used as a dimension reduction technique to visualize the components that describe each class more accuretly.

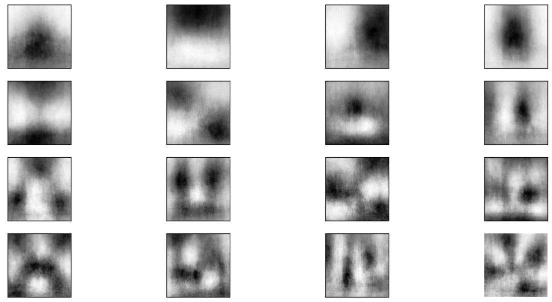

Figure 7: Eigenimages of class fire

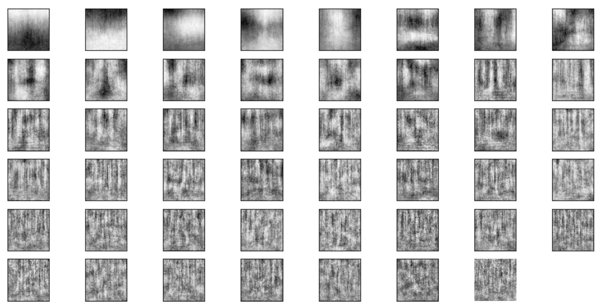

Figure 8: Eigenimages of class non-fire

The eigenimages of non-fire images show much more edge definition and sharpness in comparisson to the class fire.

6 Detailed plan

In order to show a detailed outline of the development of the project, a Gantt chart is shown below. Table 1 represents the proposed schedule and Table 2 represents the schedule that was actually followed for the development of the proyect.

Table 1: Initial schedule

Activity	*Weeks*																
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*
Problem statement and literature review	■	■	■	■	■	■											
Data mining and preprocessing						■	■	■	■								
Modelling										■	■	■	■				
Cross-validation and corrections														■	■	■	■
Writing process		■			■			■			■			■	■	■	■

Table 2: Real schedule

Activity	*Weeks*																
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*
Problem statement and literature review	■	■	■	■	■	■											
Data mining and preprocessing							■	■	■	■	■						
Modelling												■	■	■			
Cross-validation and corrections														■	■		
Writing process		■			■			■			■					■	■

As shown in the tables, the schedules differed because the data mining and processing took more time than expected. Nevertheless, since the model gave such good predictions, not many corrections had to be done, so at the end the times fitted perfectly.

7 Results and discussion

After performing the transformations previously described in section 4.1 the resulting images looked like this:

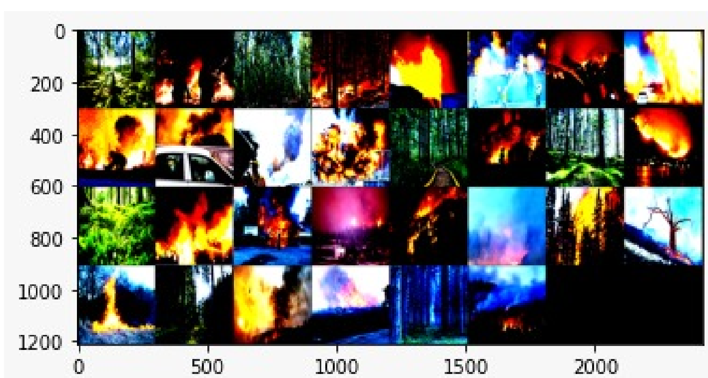

Figure 9: Transformed images

The results obtained with the validation and test set are presented in the attached table:

Set	Recall	Precision	f1 score	AUC score	Accuracy score
Validation set	0.985	0.957	0.971	0.983	0.954
Test set	0.987	0.953	0.969	0.989	0.950

Table 3: Metrics for validation and test sets

Table 3 shows no evident differences in the model's performances with the test and validation sets. The precision in the validation set is slightly better than the precision in the test set, and the recall in the validation is slightly worse than the recall in the test set. As per the AUC, the test set did better than the validation set, and as per the accuracy score, there is no significant difference between both. In general, the results are extremely good and does a great classification.

The confusion matriz and the ROC curve for both sets are exposed below.

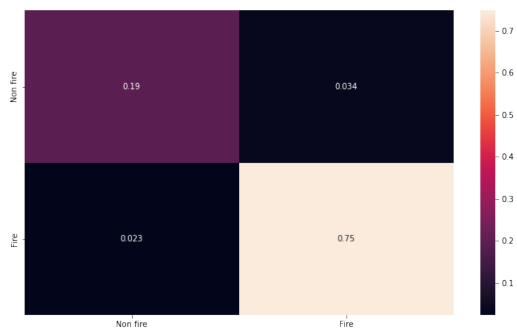

Figure 10: Confusion matrix for validation set

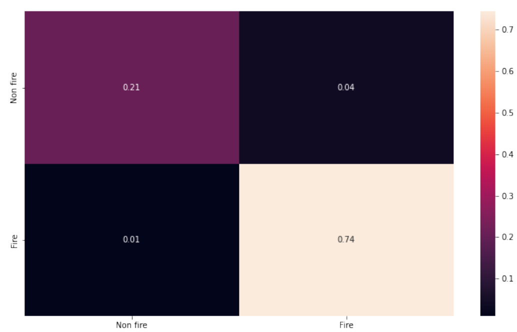

Figure 11: Confusion matrix for test set

The horizontal axis corresponds to the predicted values and the vertical axis are the actual values. With this in mind, there are more false positives than false negatives. This is preferable because eventhough resources might be wasted, lives are less likely to be risked. However, the results are significantly good.

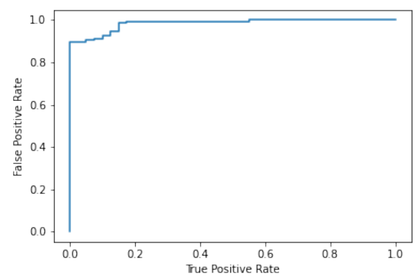

Figure 12: ROC curve for validation set

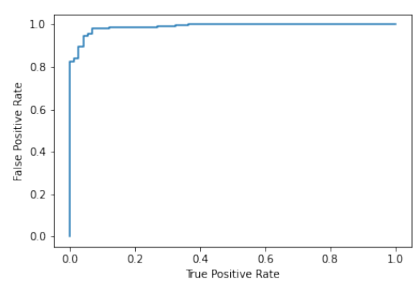

Figure 13: ROC curve for test set

It is evident that both curves are distant from the diagonal and are almost perpendicular, therefore the model works extremely well as a a classifier.

8 Ethical Implications

The main ethical concern of the project relies on misclassifying signals of fire or smoke. This can avoid an early detection of the catastrophe, causing a tardy reaction from the people surrounding and even the corresponding authorities. Taking into account time is key in emergency management, the above would have serious consequences such as a massive and uncontrollable spread of the fire attempting against the life of the nearby population. The damage of misclassifying data in real life is irreparable, such as the loss of a loved one or losing a home. On the other hand, the correct functioning of the present work will not replace any current worker since its main purpose is to alert an emergency playing a role similar to a fire alarm.

9 Comercial and legal aspects

Identifying fire and smoke can provide different commercial benefits as tis would allow enterprises, first aid institutions, or the State know when a fire is starting and act aproppiately to minimize the damage cuased. In legal terms, the project will be based on a public dataset found in Kaggle, meaning that a non-disclosure agreement is not actually needed in this case. Finally, there is a lack databases about fires and smoke, making it a bit more difficult to actually go ahead and find information that could prove useful.

10 Conclusions and future work

To conclude, the model's metrics are extremely good, even after an hypothetical deployment. In fact, the threshold defined in the objectives (85% accuracy) was surpassed. This could be due that the data is really clean, and it is suggested to monitor the performance of the model with other data.
As for the scope of the model,the metrics are good enough for it to be used as an alarm system for wildfires, since the forests cannot be monitored permanently by humans.
For future research an instance segmentation can be implemented to identify exactly in which sections of the image the fire or smoke is present. This task is useful to trace safe evacuation routes, provide more relevant information to the competent authorities and control the disaster with greater agility.

References

Bandyopadhyay, Hmrishav. 2022. Computer Vision: Everything You Need to Know. *V7.*

Bhatt, Dulari, Patel, Chirag, Talsania, Hardik, Patel, Jigar, Vaghela, Rasmika, Pandya, Sharnil, Modi, Kirit, & Ghayvat, Hemant. 2021. CNN Variants for Computer Vision: History, Architecture, Application, Challenges and Future Scope. *Electronics*, **10**(20), 2470.

Boesch, Gaudenz. 2022. A Complete Guide to Image Classification in 2022. *viso.ai.*

Chirodea, Mihai Cristian, Novac, Ovidiu Constantin, Novac, Cornelia Mihaela, Bizon, Nicu, Oproescu, Mihai, & Gordan, Cornelia Emilia. 2021. Comparison of tensorflow and pytorch in convolutional neural network-based applications. *Pages 1–6 of: 2021 13th International Conference on Electronics, Computers and Artificial Intelligence (ECAI).* IEEE.

Muhammad, Khan, Ahmad, Jamil, Lv, Zhihan, Bellavista, Paolo, Yang, Po, & Baik, Sung Wook. 2018. Efficient deep CNN-based fire detection and localization in video surveillance applications. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, **49**(7), 1419–1434.

Roberts, Lawrence G. 1963. *Machine perception of three-dimensional solids.* Ph.D. thesis, Massachusetts Institute of Technology.