MATH70103: Unstructured Data Analysis
Imperial College London
Autumn 2023
CCID: 00951537

Final Project

Plagiarism Statement: *The following assignment is a product of my own work.*

Abstract

In this report I explore the effectiveness of Convolutional Neural Networks in identifying images that contain fires in the wild, with the aim of producing an algorithm that could potentially be incorporated in early wildfire detection systems. I make use of the FLAME (Fire Luminoscity Airbourne-based Machine learning Evaluation) dataset [9] and build upon the *Xception* Deep Convolutional Neural network proposed by Google [4]. Key metrics such as binary accuracy, precision and recall are presented for training, validation, testing, and out of sample testing, the latter of which examines the performance of the model on images outside of the *FLAME* dataset. For this purpose, I use the Fire Classification dataset available on Kaggle [5].

All code and a 100MB subset of the FLAME Dataset have been uploaded to the following Github Repositry: `https://github.com/martinbatek/IC-UDA-Final-Project.git`

Contents

1 Introduction and Problem Statement

Wildfires are a pervasive issue across the globe today, causing millions of dollars in damage every year, loss in natuaral biodiversity in the affected areas, and significant loss and trauma to the affected communities. Recent examples of significant wildfires include the 2023 Canadian wildfire season [8], the 2019-20 Australian bushfires (nicknamed "The Black Summer") [6], and the fires in my home town of Cape Town this Christmas season [7]. It is widely known that natural occuring fires are an integral part of many ecosystems, as they clear out dead organic matter that prevents living orgnisms from accessing vital nutrients [10], alongside numerous other benefits [2]. However multiple sources cite an increasing rate of occurence and degree of

severity of global wildfires, past the point of biological sustainability [<**empty citation**>]. This phenomenan is widely attributed to human intervention such as ... .
The above justifies the need to develop novel methods for detecting wildfires, so that we may anticipate and manage the impact of wildfires in the best way possible. Upon reviewing multiple early-wildfires detection systems, Bouguettaya et al. found that the combined used of unmanned aerial vehicles (UAVs), mounted high resolution imagery and Deep Learning techniques provided the most promising results in terms navigability, versitility, speed and accuracy. Shamsoshoara et al. introduce a high-resulution image dataset composed of drone footage frames that were taken during a perscribed burning slash piles in Northern Arizona. In that same paper, the Google-Keras *Xception* network was applied to create a binary classifier model for fire image detection.
The problem statement for this report is therefore as follows:

Problem Statement:

- **Train a Convolutional Neural Network classifier for fire image classification.**
- **Investigate the effect of the preprocessing, hidden and output layers of the trained CNN model on an example image.**
- **Evaluate the Accuracy, Precision and Recall of the model on Validation, Testing and Out-of-Sample images.**

2 Image Dataset

2.1 Data Selection

Multiple datasets were considered for the purposes of this report, from sources such as Kaggle, and other publicaly available portals. Besides the stated criteria for this project (complexity and originality), I also aimed to use a dataset that is suitable given the findings in "A review on early wildfire detection from unmanned aerial vehicles using deep learning-based computer vision algorithms", wherein Bouguettaya et al. argue that UAV-sensor technology is more suitable for autonomous fire detection systems in terms of cost, accuracy and practicality than alternatives such as satellite imagery or stationery platforms. The ideal dataset would therefore include images taken from the perspective of UAVs.
To this end the FLAME Dataset produced by Shamsoshoara et al. addresses these criteria explicitly. The dataset includes drone imagery taken of the perscribed burning of slash piles in the forests of Arizona during January 2020, during which time the weather conditions were generally cloudy with an average temparature of -6 °C. The following Exploratory Data Anylisis includes visualizations of a few example images in the dataset. The full published dataset includes video footage, images, a variety of specturms used (including the normal RGB, fusion, white-hot and green-hot palettes), as well as images and masks for fire image segmentation (identifying the fire pixels in an image, thereby identifying where the fire is situated from the perspective of the UAV). For the purposes of this report, I use only the **Training, Validation and Test** classifiction images (all in the normal RGB spectrum), namely items 7) and 8) in the following IEEE-Dataport link: https://ieee-dataport.org/open-access/flame-dataset-aerial-imagery-pile-burn-detection-using-drones-uavs.
The classification dataset above is 1.46 GB in size. For the purpose of submission, I have produced a 100MB subset of this dataset using random sampling, and uploaded it to the the aforementioned Github repository, under `data/FLAME_dataset_subset`. However, note that the

model and results described in the following report were compiled on the basis of the entire set of images.

2.2 Exploratory Data Analysis

In this section, I aim to visualise the data downloaded from points 7) and 8) on the aforementioned IEEE-Dataport. As a first step, I visualise the directory heirarchy structure in `scripts/UDA_FinalProject_Batek.ipynb`:

```
-FLAME Dataset - Shamsoshoara
    -Test
        -Fire
            -5137 files
        -No_Fire
            -3480 files
    -Training
        -Fire
            -25018 files
        -No_Fire
            -14357 files
```

Figure 1: File structure of the Classification images in the FLAME Dataset

Figure 1 shows that the dataset contains 47992 images, 8617 (18%) of which are in the test set with the remaining 82% in the training/validation set. Examining the file counts clearly shows that there is a significant class imbalance between Fire and Non-Fire images, and that the Fire Class is overrepresented in the training set relative to the Test set. This imabalance will likely influence model specificity and recall later on.

3 Analysis

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

3.1 The Google-Keras Xception Network

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

3.2 Training and Validation Results

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

3.3 Testing on Unseen Data

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

3.4 Testing on Out of Sample Data

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

4 Summary and Conclusion

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

References

[1] Martin Batek. *Unstructured Data Analysis Final Project*. URL: https://github.com/martinbatek/IC-UDA-Final-Project.git.

[2] William John Bond and RE Keane. "Fires, ecological effects of". In: *Reference module in life sciences* (2017), pp. 1–11.

[3] Abdelmalek Bouguettaya et al. "A review on early wildfire detection from unmanned aerial vehicles using deep learning-based computer vision algorithms". In: *Signal Processing* 190 (2022), p. 108309. ISSN: 0165-1684. DOI: https://doi.org/10.1016/j.sigpro.2021.108309. URL: https://www.sciencedirect.com/science/article/pii/S0165168421003467.

[4] Francois Chollet. "Xception: Deep Learning With Depthwise Separable Convolutions". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* July 2017.

[5] Ahmed Gamaleldin et al. *FIRE Dataset.* https://www.kaggle.com/datasets/phylake1337/fire-dataset. Accessed: 2023-12-18. 2019.

[6] Megan Harwood-Baynes. "Australia bushfires: Devastating flames released more than twice the amount of CO2 than previously thought, says study". In: *Sky News* (Sept. 2021). URL: https://news.sky.com/story/australia-bushfires-devastating-flames-released-more-than-twice-the-amount-of-co2-than-previously-thought-says-study-12408680.

[7] Russel Hope. "Cape Town: Hundreds of firefighters battling blaze in South African city". In: *Sky News* (Dec. 2023). URL: https://news.sky.com/story/cape-town-hundreds-of-firefighters-battling-blaze-in-south-african-city-13034680.

[8] Oliver Milman. "After a record year of wildfires, will Canada ever be the same again?" In: *The Guardian* (Nov. 2023). URL: https://www.theguardian.com/world/2023/nov/09/canada-wildfire-record-climate-crisis.

[9] Alireza Shamsoshoara et al. "Aerial imagery pile burn detection using deep learning: The FLAME dataset". In: *Computer Networks* 193 (2021), p. 108001. ISSN: 1389-1286. DOI: https://doi.org/10.1016/j.comnet.2021.108001. URL: https://www.sciencedirect.com/science/article/pii/S1389128621001201.

[10] National Geographic Society. *The Ecological Benefits of Fire.* Accessed on December 20, 2023. 2023. URL: https://education.nationalgeographic.org/resource/ecological-benefits-fire/.