Classifying lightning events in low-quality webcam footage using a CNN-RNN algorithm

Kabelo Komape
School of Applied and Computer Science
University of the Witwatersrand
Johannesburg, South Africa
kabelk93@gmail.com

***Abstract*—Understanding characteristics of lightning events in scientific and engineering fields has implications for daily life. In recent years this research into lightning phenomena has been heavily influenced by the use of high-quality video cameras capable of capturing lightning events in great detail over the past 40 years. These cameras are often expensive and consume large amounts of data, and . In this paper, a single CNN-RNN hybrid algorithm is used to try an classify low-quality lightning footage into five categories.Model hyperparameters for the optimization function and learning rate are tested to find an optimal setup for classifying the data. The results are very consistent, albeit low in accuracy. Despite experimenting with multiple hyperparameters for the RNN and using multiple CNN architectures, the model produces an accuracy of 37%, which may point to a systematic error in the type of architecture used, or to the data being difficult to produce.**

I. Introduction

The advancement of lightning analysis has increased with time partly due to an increase in the development of the camera technology used to capture lightning events. A paper that has been widely referenced in lightning research using camera technology is [1] for the reason that it was one of the first, notable papers to use such methods. Through analytical methods data were mined from footage recorded by the technology available at the time. A breakthrough study came from [2] which used high-speed cameras for the first time to capture lightning events for research purposes. High-speed motion cameras make it possible for temporal high-resolution videos to be taken of lightning, to extract data metrics from video footage in order to investigate and understand the properties of lighting [3]. Since [2] many more studies like [4] and [5] have proved the importance of high speed cameras in the field. These types of cameras, if typically shooting at 18,000 frames-per-second and a resolution of 720p, consume more than 20 gigabytes of storage for 1 second of footage. So they require large amounts of hard data storage and are often expensive. Not only that, in order to capture a lightning event, many variables must be in a researcher's favour. The thunderstorm must be accurately predicted, scene must be empty and sprawling as to capture the event in it's entirety and without obstruction, camera must be positioned at a location that will capture strikes, and the operator must be alert enough to trigger the equipment to capture the event. Security cameras and webcams are widely installed all around the world for surveillance purposes and a natural thought is to try use them in lightning research. After all, if a thunderstorm were to occur in a city, it would be seen by many cameras at a variety of angles, so one single strike can have many points-of-view. The issue comes in that footage is of poor quality and captured at a low frames-per-second. More simplistic metrics may be obtained in the beginning, but with more research we may be able to gather more complicated metrics.

Machine learning has also helped in lightning research, however its use in this field to the best of this paper's knowledge is very limited. In [3] a machine learning system was developed that helped label lightning footage and automate the collection of metrics. Its first stage was a segmentation network that labelled (gave a feature to) each pixel in a cluster of pixels and in the second phase the output labels of the segmentation network were used as inputs to a temporal system that would separate these input labels into sub events, to then extract the desired metrics [3]. While the image segmentation network worked well with 87% accuracy at its highest, it incorrectly identified noise and buildings as lightning, while the temporal system did not gather metrics such as the number of events, beginning frame and duration of a flash of lightning accurately. These limitations open the door to more research on the topic. In recent years neural networks have gained prominence in image classification [6]. A study by [7] involving video classification and object tracking using low quality footage was done using a convolutional neural network algorithm, which holds strong relevance to the topic of this paper. Results showed an average accuracy of 69.35%, with the best hyperparameters being a learning rate of 0.00075 and using the Adam optimizer. Machine learning applied to lightning research could help with the automation of triggering systems. A manual triggering system activated by a human was used to start and stop the recording of an event [2], [5].

The data set to be used in this study is video footage overlooking a skyline in Johannesburg, South Africa, of thunderstorms. A large amount of this footage takes place at night [8]. Tracking any object in footage captured by a low resolution surveillance camera is a difficult task due to poor illumination and low contrast in the image [7]. Horizon detection in an image is important for many applications such as flight and navigation control and for the protection of sanctuaries [9].

Identify applicable funding agency here. If none, delete this.

These tasks all fall in the domain of image classification and in the past few years deep learning models have been used to overcome challenges faced with other classifying algorithms.

This paper will try to classify lightning events from video footage, using a hybrid CNN-RNN. To produce the best results different hyperparameters will be changed and tested against one another. A discussion will be conducted on the accuracy of results, and what limitations the experiments encountered. A list of possible improvements will also be explored.

II. Methodology

A. Data

The data used in this paper comes from [8] where the folder contained 3315 videos of lightning events captured from a tower overlooking a skyline in Johannesburg, South Africa. The videos are taken using three different cameras, each being able to give a different view of any given lightning strike. Thunderstorms happen at any time of the day, so lighting conditions vary. The videos are of .mp4 format, with a resolution of 640x360. Crucially, the videos have a frame rate of 5-30 frames per second. Videos are labelled as being:

- towerupward: upward flashes initiating from tall structures
- towerdownward: downward flashes initiating from clouds
- close: lightning events that do not attach to the tower are downward events, and the channel extends for 90% of the frame height
- far: lightning events that do not attach to the tower are downward events and the channel is less than 50% of the frame height.
- unclear: a bright flash, saturating at least 25% of the frame
- unclearbehind: lightning event occurs behind the camera so not captured directly, but the light is reflected in the clouds and that is captured.
- CC: Lightning discharges occur between clouds

The goal of this paper is to try classify the videos into one of these seven classes using a CNN-RNN architecture. There are some videos which are unlabelled, mostly coming from the camera designated ”Cam3”. As those are of no use to training nor validation the final amount of videos available for the experiment is 2095.

B. Architecture

A video is made up of a series of frames in a sequential order specific to the event captured. Each frame comprises spatial data, whereas the sequence of those frames contains temporal data. There are methods to model the spatial data, and different methods to model the temporal data. To model both, a hybrid architecture that includes convolutions (for spatial processing) and recurrent layers (for temporal processing) is used to simulate both of these elements. To this end a Convolutional Neural Network (CNN) and a Recurrent Neural Network (RNN) with GRU layers in particular is employed. A CNN-RNN is a popular name for this type of hybrid architecture.

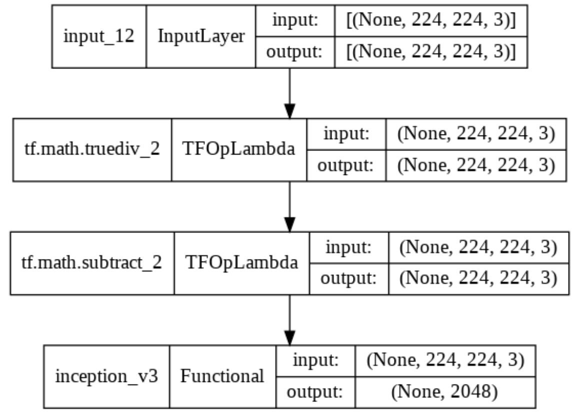

Fig. 1. Graphical representation of InceptionV3 model

The main objective of the CNN is to extract features from the videos to be used in the latter parts of the overall algorithm. As it is a feature extractor and the first part of the transfer learning process, a pre-trained CNN is used in this case. Inceptionv3, VGG-16, ResNet50 and Xception. These are popular, well-performing CNNs in the TensorFlow/Keras library and all models utilised ”imagenet” weights, pre-trained on the ImageNet data. Each video is split into its individual frames and each frame is given class label linked to the video it is derived from. Each image is reduced to a shape of (255, 255, 3). The CNN extracts features from these images to be used in the RNN.

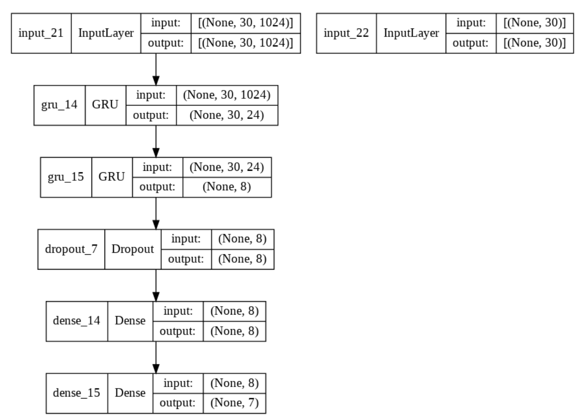

Fig. 2. Graphical representation of RNN phase

The RNN is comprised of an input layer, which connects to the first GRU layer of 24 units. Data flows into the next GRU layer of 8 units, and then into a Dropout layer with a frequency of 0.4. Next is the Dense layer. It is kept at 8 units, but experiments were done by changing the activation functions. Relu, sigmoid and softmax were used as hyperparameters. Finally, the data goes through one more Dense later, subjected to a softmax activation and then the output.

III. Results and Discussion

Acknowledgment

Thank you to Dr. Ritesh Ajoodha and Hugh Grant for supervision during the course of writing. All efforts, from video conferences to materials relevant to helping put this paper together, are greatly appreciated.

References

[1] R. D. Brantley, J. A. Tiller, and M. A. Uman, "Lightning properties in florida thunderstorms from video tape records," vol. 80, no. 24, August 1975, pp. 3402–3406. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/JC080i024p03402

[2] M. M. F. Saba, M. G. Ballarotti, and O. Pinto Jr., "Negative cloud-to-ground lightning properties from high-speed video observations," vol. 111, 2006.

[3] J. R. Smit, H. G. Hunt, T. Cross, C. Schumann, and T. A. Warner, "Generation of metrics by semantic segmentation of high speed lightning footage using machine learning," in *2020 International SAUPEC/RobMech/PRASA Conference*, January 2020, pp. 1–6.

[4] Q. Li, P. Yuan, J. Cen, and X. Wang, "The luminescence characteristics and propagation speed of lightning leaders," vol. 173, 2018, pp. 128–139.

[5] L. Z. Campos, M. M. Saba, T. A. Warner, O. Pinto, E. P. Krider, and R. E. Orville, "High-speed video observations of natural cloud-to-ground lightning leaders – a statistical analysis," vol. 135-136, 2014, pp. 285–305. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169809513000057

[6] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," vol. 29, 06 2017, pp. 1–98.

[7] Z. Kadim, M. A. Zulkifley, and N. Hamzah, "Deep-learning based single object tracker for night surveillance," vol. 10, 2020, pp. 3576–3587.

[8] H. G. Hunt, "Dataset of photographed lightning events attaching to and around the brixton tower, johannesburg, south africa for the 2015-2016 thunderstorm season." vol. 30, 2020, p. 105630. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352340920305242

[9] S. Fefilatyev, V. Smarodzinava, L. O. Hall, and D. B. Goldgof, "Horizon detection using machine learning techniques," in *2006 5th International Conference on Machine Learning and Applications (ICMLA'06)*, December 2006, pp. 17–21.

[10] W. L. Boeck, O. H. Vaughan Jr., R. J. Blakeslee, B. Vonnegut, M. Brook, and J. McKune, "Observations of lightning in the stratosphere," vol. 100, no. D1, 1995, pp. 1465–1475. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/94JD02432