Semantic segmentation of high speed lightning footage using machine learning

Tyson Cross and Jason Smit
Supervisor: Dr. Hugh Hunt

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG
SCHOOL OF ELECTRICAL AND INFORMATION ENGINEERING

Introduction

High-speed lighting footage analysis currently requires large amounts of manual processing to extract useful metrics required for research.

At tens of thousands of frames per second, many details of lightning events can be captured, at the cost of image resolution. The duration of ionised channels in which lightning can form, the direction of the lighting, and the number of strokes are of particular interest to many physicists and engineers.

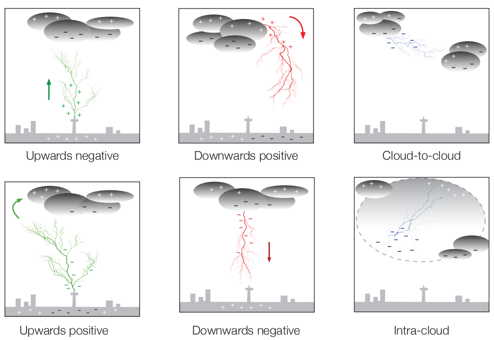
We investiged whether machine learning techniques could be used to provide a method to extracting useful metrics quickly and robustly, to assist researchers in counting and classifying events from captured footage.

Lightning

Lightning discharge events occur when there is a large charge difference between a cloud and the ground or between cloud formations. When a sufficiently large charge difference builds up, a break-down event can be triggered, leading to a flow of charge carriers (usually electrons). The initial visible discharge resulting from the movement of the charge as the lightning tries to form a channel is referred to as a *leader*.

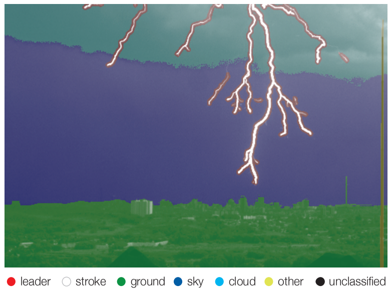
The direction of the movement of the leader is classified as:

Cloud to ground (Downwards event)
Ground to cloud (Upwards event)
Cloud to Cloud (Horizontal event)



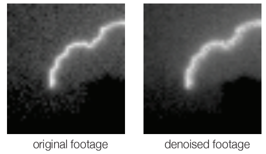
If a leader connects two potentials of opposite polarity, a large current flows, leading to the characteristic flash of lightning from the discharge, this is termed an attachment event.

Data Preperation



The input sequences were labelled with the aid of compositing software, to create binary *mattes* indicating the features to be learned for image segmentation. The mattes were created using luminance keys, and by applying a Laplacian filtering operation to reveal edges and other details in the image structure. This allowed precise per-pixel definition of ground truth across sequences, without having to process and label each frame individually.

The source data had high variation, and often was extremely noisy and dark. Particulate matter, rain and various forms of occlusion from stormy weather added additional sources of noise, along with digital sensor noise to the footage. To improve input data quality, a time-averaging denoising operation is applied to the images before they are input to the network.

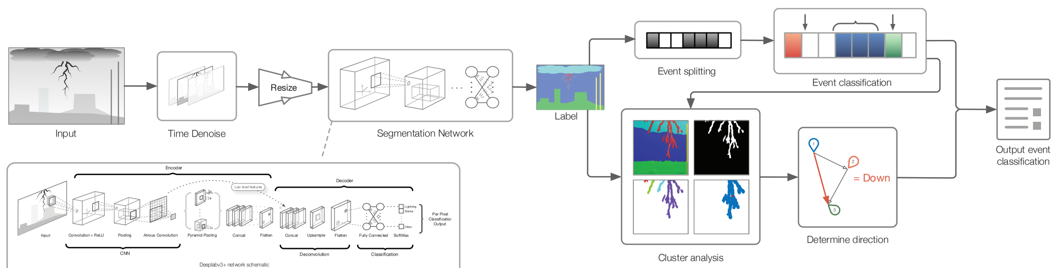


Methodology

Transfer learning was used with five pre-trained convolutional neural networks (CNNs) implemented in MATLAB. These networks had previously been trained on millions of images from ImageNet, and are known to be performant with high accuracy at image segmentation. The networks were trained with 30 000 labelled images from sequences extracted from the high-speed footage, resized to fit the network inputs.

Given the highly varied luminance values between individual videos (some filmed at night, others in the afternoon) we augmented the training set with versions of the sequences that had been exposed up and down in brightness. Batch normalization was used to rescale each mini-batch of 16-bit image pixel values to be between -1 and 1. Additionally, the input image set was augmented with random translation, scaling and horizontal flipping of the images between training epochs.

The input images were divided into train/validation/test sets with a 70/15/15 split. The networks were all trained using stochastic gradient descent with momentum (SGDM) method, with a validation criteria to prevent overfitting. The initial learning rates and schedule for adjusting the learning rates was altered for each network depending on the batch size for each network. Due to different memory requirements for individual networks on the available hardware, the batch size varied between 20 and 120. Deeper networks with larger memory requirements and smaller batch sizes were given smaller learning rates, and longer between learning rate adjustments.



Overall System Diagram

The output labels consist of images with each pixel corresponding to a particular feature (integers 1 to 6). The output classified frames are then split into events, determined by the length of the gaps between frames with lightning. The number of frames containing pixels labelled as *stroke* are then classified as attachment events, and the subdivided sequences that contain lightning but no stroke labels are marked as *attempted leader*. The number of strokes in each event are counted, along with the frames that the stroke labels begin on. The starting frames of each event are then analysed using a clustering technique, to classify the overall direction of the lightning event.

Direction Finding

The labelled image output from the CNN is transformed from pixels into points. A clustering algorithm (DBSCAN) identifies connected sections of the *leader* shapes, and excludes outlier points. The centroids of the clusters are then averaged, weighted by the cluster size in the frame. This results in a centre of mass for the overall lighting shape.

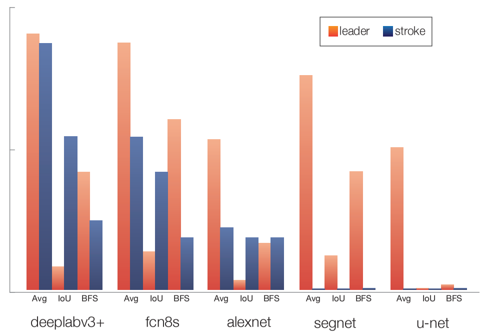
This centroid is tracked spacially between frames, creating a directional vector. The vectors are summed across the candidate frames determined by the event-splitter. The resultant vector is then considered. If the final vector points down or up, then the event is classified as that direction.

We thank Tom Warner and the University of the Witwatersrand, Johannesburg for the use of the video footage, and Dr. Albertyn for generously providing the computational resources used on this project.

Results

- 23 000 frames per second Phantom Camera footage
- 187 CINE videos
- 246 423 exported images source data
- 48 381 images labelled
- 422 sequences of lightning events
- 111 labelled sequences

Network	Layers	Total Learnables	Size(mB)
alexnet	27	57 040 269	212
deeplabv3+	101	20 611 166	61,6
fcn8	51	134 300 132	501,2
segnet	91	29 446 482	110,7
u-net	58	31 032 068	116



Performance metrics for implemented networks

The system is able to detect and classify lightning events from high-speed footage, suggesting a direction and counting the number of strokes.