Searching for Long-Period Comets with Deep Learning Tools

Susana Zoghbi†1, **Marcelo De Cicco**†1,2, **Antonio J.Ordonez**†1, **Andres Plata Stapper**†1, **Jack Collison**1,6, **Peter S. Gural**3, **Siddha Ganju**1,5, **Jose-Luis Galache**1, **Peter Jenniskens**1
1 SETI Institute 189 Bernardo Ave, Mountain View, CA 94043, USA.
2 INMETRO, Divisão de Metrologia de Materiais, Av. Nossa Senhora das Graças 50, Duque de Caxias, 25250-020, Brazil.
3 Gural Software Development, 351 Samantha Drive, Sterling, VA 20164, USA.
4 Stanford University, 450 Serra Mall, Stanford, CA 94305, USA
5 Deep Vision Inc, 2627 Hanover Street, Palo Alto CA 94304
susana.zoghbi@gmail.com, decicco10@gmail.com, antonio.j.ordonez@gmail.com, andresplatastapper@gmail.com, siddhaganju@gmail.com, jlgalache@gmail.com, petrus.m.jenniskens@nasa.gov
† Equal contribution

Abstract

The aim of this paper is to provide Deep Learning tools to aid the search for debris of long-period comets. We describe an effort to automate the CAMS (Cameras for Allsky Meteor Surveillance) data reduction pipeline to discriminate meteors from other types of detection. The effort helps to process the data collected every night from low-light video observations, and makes results available to the observer the following day. Detections are classified as meteors or non-meteors using two methods, a convolutional neural network (CNN), and a long short term memory network (LSTM). Our achieved performance makes our models suitable to be deployed on site and help alleviate the cognitive load on astronomers to classify night detections.

1 Introduction

Long-period comets[1] (LPC), due to their potentially large size and fast impact speeds of up to 72 km/s, contribute to the impact hazard on planet Earth along with short-period comets and asteroids. These impacts can severely disrupt the ecosphere and entire human populations. Evidence indicates that the impact of a comet or asteroid, having a diameter of about 10 km, was responsible for the mass extinction of most species of dinosaurs. Impacts of future threatening asteroids could be mitigated if given sufficient warning time. However, any new long-period comet type on an impact trajectory with Earth would likely only be discovered 6-12 months before impact, when it becomes visible as the Sun's heat and wind start sublimating its icy surface and ejecting rocky debris.

To provide extra warning time, the orbits of the comet's debris could be used by researches to guide the search for comets while they are still far out, providing us years of extra warning time in case they can be detected along that orbit. Most suitable is debris that was ejected during the previous return of the comet to the inner Solar System, which will cause rare aperiodic meteor showers (outbursts). Therefore, detecting those showers requires a continuous and global search [1]

Figure 1 shows how the debris evolves from a meteoroid coma into a meteoroid stream in one revolution as a result of differences in orbital period between grains. The intersection point at Earth's

[1] Comets that take at least 200 years to go around the Sun.

Workshop on Deep Learning for Physical Sciences (DLPS 2017), NIPS 2017, Long Beach, CA, USA.

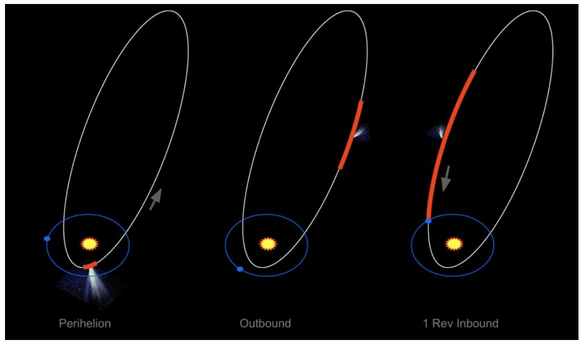

Figure 1: A diagram showing debris trails formed following a previous LPC travel to the inner Solar System. An LPC outburst happens when Earth (the blue point) cross those debris.

orbit (the node) of those orbits is not constant. As described by Jenniskens (1997), LPC outbursts are due to gravitational perturbations on the individual meteoroid orbits, causing a periodic displacement of the stream relative to Earth's orbit, which follows the Sun's reflex motion around the barycenter, with dominating contributions by Jupiter (12-year period) and Saturn (30-year).

Given that the long term hazard posed by long-period comets is statistically comparable to other naturally occurring events [2], it is important to design detection and prevention strategies in order to mitigate potential impacts.

To do so, the night sky needs to be monitored for an extended period of time ($\sim$ 60 years) and from locations around the globe [1, 3]. The Cameras for Allsky Meteor Surveillance (or CAMS) is a network of low-light video cameras, established by the SETI[2] Institute in different locations across the globe, that monitors the sky to detect meteors. CAMS has demonstrated to produce meteoroid orbits that can be used to identify where a parent long-period comet may reside, when it is still many years out.

Until now, processing the images collected by CAMS has required time-consuming human input. On an average night, an astronomer receives -per camera- around 500 detections consisting of images and light intensity curves (a sequence of measurements of how light intensity changes as detected objects move in the sky). A total of 8,000 observations with 16 cameras per site.

Figure 2 presents examples of images captured by CAMS. Most of these turn out to be false detections, such as planes, birds, clouds, etc. Sorting through these every night is not scalable.

To alleviate this, our goal is to improve and automate the classification of meteors from non-meteors using deep learning tools. Specifically, we focus on processing images and light curves captured by CAMS. To the best of our knowledge, this is the first time that deep learning techniques have been applied to this endeavour.

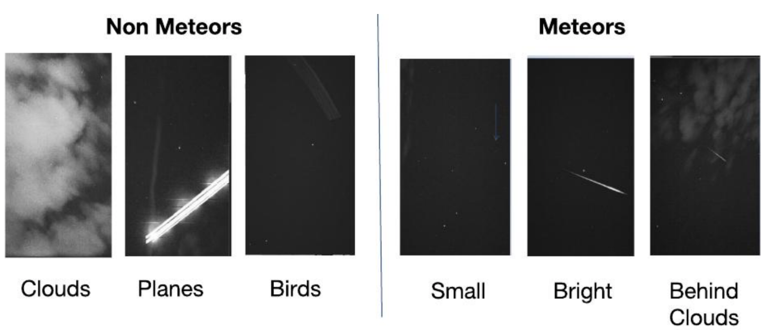

Figure 2: Examples of images detected by CAMS.

[2]Search for Extraterrestial Intelligence

Table 1: Performance of our methods in the heldout test set.

Method	Input	Precision	Recall	F1
CNN	Images	88.3	90.3	89.5
Random Forest	Tracklets	90.0	80.6	84.9
LSTM	Tracklets	90.0	89.1	89.6

2 Methodology

As a baseline, we trained a random forest [4] to classify light curve tracklets. For this classifier, we focused on the time-series spatial and photometric data for each object. Given the heterogeneity of the data (i.e. different numbers of data points on each tracklet - the duration of the detection is different), we opted to mine metrics from these time-series data that describe the trajectories and light curves of each object.

We calculated the coefficient of determination (R2) and residual standard deviation of a best-fit line to the XY paths in the images, as well as the total distance traveled. Regarding the photometric variations, we extracted several statistical measures of the shape of their light curves including the mean intensity, median absolute deviation (MAD), skew, and kurtosis. We also included in this model two measures of the timescale of the event: the total time observed and the optimal period from a fast Fourier transform on the light curves. We used the random forest implementation from `http://scikit-learn.org` [5].

As an alternative to hand crafting features from the light time series, we fed the raw measurements of light intensity, XY position and time stamps to a Long-Short Term Memory network (LSTM) [6]. The LSTM encodes the light curve tracklets into a latent space, and learns to predict whether or not the tracklet corresponds to a meteor. We used one hidden layer and a softmax classifier. We used cross-validation to determine the number of hidden units, and found that 1,000 units worked well. We used a a batch size of 128 tracklets. We set the maximum number of steps for each tracklet to 32, and zero padded or discarded extra steps, as necessary. We had 200,000 tracklets of light curves, where $\sim 3\%$ belonged to the meteor class. We randomly split the data into training (60%), validation (20%), and a heldout test (20%) set.

In addition, we trained a Convolutional Neural Network (CNN) [7] that discerns images of meteors vs. other objects in the sky. As in common practice, our convolutional layer consists of small, learnable filters. During the forward pass, we convolve each filter with the input by successively computing dot products between small windows of the input and the filter. As we slide the filter across the input, we produce an activation map that gives the responses of that filter at every spatial position. We use backpropagation with stochastic gradient descent to learn how to update the weights to minimize a cross-entropy loss function.

We used five convolutional layers followed by two fully connected layers and a binary softmax classifier. Dropout [8] and max-pooling layers were used. The software library Keras [9] was used to train both the LSTM and CNN.

The image dataset consisted of about 34,110 gray scale images, where $\sim 23\%$ were meteors and the rest corresponded to other objects. We randomly split the data into a training (80%), validation (10%) and heldout test (10%) set. We performed standard data augmentation techniques on instances of the positive class, such as rotation and flipping. The size of each image is 480x640. We decided to keep the original image size since reducing the image size led to lower performance.

3 Results and Discussion

Table 1 presents our results on the heldout test set. Our CNN achieves precision and recall scores of 88.3% and 90.3%, respectively[3]. The LSTM achieves a precision of 90.0% and a recall of 89.1%. Whereas the random forest classifier achieves 90.0% precision and 80.6% recall. One key advantage

[3]Precision indicates what proportion of instances we classified as meteors were actually a meteor. Recall is what proportion of actual meteors were classified as meteors. This means that the CNN would make 11.7% false detections and miss 9.7% of all meteors in the data.

of using Deep Learning is that we did not have to hand-engineer the meaningful features from both images and light curves. The CNN and LSTM networks can learn these on their own.

We did a qualitative evaluation by visually inspecting the network's predictions. Figure 3 presents some examples. False negatives often happen when meteors are very faint and hard to see. False positives tend to occur when there is an object (like a satellite) that looks very similar to meteors.

It takes an average of 1.8s. to perform one forward pass of one image on an off-the shelf laptop[4] with no GPU. This makes it very suitable to be deployed on site, where the cameras are located.

Compared to human performance, an expert at peak productivity can annotate about one image per second and achieve 99% precision and recall. While we would desire to achieve or surpass human performance, it is not a requirement for the automation success. It is not sustainable for an astronomer to perform this process every day. We can tolerate inaccuracies of up to about 10% given that we can 1) remove the drudgery of human annotation and save time through an automated annotation process, and 2) provide a cleaner data set that feeds the downstream process of calculating meteoroids' orbits. Ultimately our tools free the astronomer from this low-cognitive task and help the search for debris of potentially dangerous long-period comets.

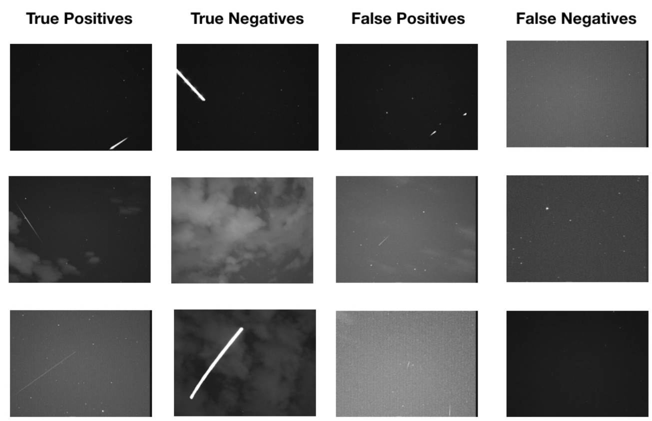

Figure 3: Examples of classification results.

4 Conclusions

In order to fully automate meteor shower detection and classification from video observations, we applied Deep Learning methods: convolutional neural networks and long-short term memory networks. Both methods enabled a meteor/non-meteor classifier that output a probability score. The CNN took as input a set of max-pixel image frames, whereas the LSTM took as an input the values of time-series light intensity observations.

Performance of both methods were similar. Both performed well enough to be deployed on site and enable an automated data reduction pipeline for meteor triangulations.

Our methods allow us to provide cleaner data to the subsequent process of identifying meteoroid orbits potentially associated with long-period comets that pass close to Earth's orbit. When they were ejected in the previous return, the orbits of the meteoroids should directly trace that of their parent comet. Taking into consideration uncertainties in the orbital parameters of these meteoroids, the search space for these hazardous comets can be narrowed down. These regions can be then probed with dedicated, deep sky surveys searches to attempt to locate these long-period comets.

Acknowledgments

This effort was executed as part of NASA's Frontier Development Lab, a research accelerator led by James Parr and Jason Kessler and by the SETI Institute's Bill Diamond. We acknowledge support of the NASA Portal (Robert B. Pittman and Daniel J. Rasky). The effort was sponsored by

[4] 2.5Ghz Intel Core i7 processor, 16Gb of RAM.

NASA (Victoria Friedensen and Darlene Weidemann) and by private parties, in particular IBM (Troy Hernandez and Graham Mackintosh) and NVIDIA (Alison Lowndes).

References

[1] P. Jenniskens. Meteor stream activity. IV. Meteor outbursts and the reflex motion of the Sun. *Astronomy Astrophysics*, 317:953–961, February 1997.

[2] C. R. Chapman. Impacts on the Earth by asteroids and comets: assessing the hazard. *Nature*, 367:33–40, January 1994.

[3] E. Lyytinen and P. Jenniskens. Meteor outbursts from long-period comet dust trails. *Icarus*, 162:443–452, April 2003.

[4] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, Oct 2001.

[5] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[6] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9:1735–80, 12 1997.

[7] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems 25*, pages 1097–1105, 2012.

[8] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014.

[9] François Chollet et al. Keras. `https://github.com/fchollet/keras`, 2015.