It's a Bird it's a Plane it's a Meteor

Surya Ambardar\textsuperscript{1}, Siddha Ganju\textsuperscript{2}, and Peter Jenniskens\textsuperscript{3}

\textsuperscript{1}University of Virginia, Department of Computer Science, United States of America; SpaceML
\textsuperscript{2}NVIDIA; SpaceML; Frontier Development Lab
\textsuperscript{3}SETI Institute, Mountain View, CA, United States of America

Meteor showers are some of the most dazzling and memorable events occurring in the night sky. Caused by bits of celestial debris from comets and asteroids entering Earth's atmosphere at astronomical speeds, meteors are bright streaks of light in the night sky, sometimes called shooting stars. Those meteors are recorded, tracked and triangulated by low-light surveillance cameras in a project called CAMS: Cameras for Allsky Meteor Surveillance. CAMS offers insights into a universe of otherwise invisible solar system bodies, but that task has proven difficult due to the lack of automated supervision. Until recently, much of the data control was done by hand. Necessary to build supervised classification models, labeled training data is essential because other man-made objects such as airplanes and satellites can be mistaken for meteors. To address this issue, we leverage one year's worth of meteor activity data from CAMS to provide weak supervision for over a decade of collected data, drastically reducing the amount of manual annotation necessary and expanding available labeled meteor training data.

Founded in 2010, CAMS aims to automate video surveillance of the night sky to validate the International Astronomical Union's Working List of Meteor Showers, discover new meteor showers, and predict future meteor showers. Since 2010, CAMS has collected a decade's worth of night sky activity data in the form of astrometric tracks and brightness profiles, a year of which has been manually annotated. We utilize this one year's labelled data to train a high confidence LSTM meteor classifier to generate low confidence labels for the remaining decade's worth of meteor data. Our classifier yields confidence levels for each prediction, and when the confidence lies above a statistically significant threshold, predicted labels can be treated as weak supervision for future training runs. Remaining predictions below the threshold can be manually annotated. Using a high threshold minimizes label noise and ensures instances are correctly labeled while considerably reducing the amount of data that needs to be annotated. Weak supervision can be confirmed by checking date ranges and data distributions for known meteor showers to verify predicted labels.

To encourage discovery and distribution of training data and models, we additionally provide
scripts to automate data ingestion and model training from raw camera data files. The data scripts handle processing of CAMS data, providing a pipeline to encourage open sharing and reproduction of our research. Additionally, we provide code for a LSTM classifier baseline model which can identify probable meteors. This baseline model script allows further exploration of CAMS data and an opportunity to experiment with other model types.

In conclusion, our contributions are (1) a weak supervision method utilizing a year’s worth of labelled CAMS data to generate labels for a decade’s worth of data, along with (2) baseline data processing and model scripts to encourage open discovery and distribution. Our unique contributions expand access to labeled training meteor data and make the data globally and publicly accessible thorough daily generated maps of meteor shower activity posted at http://cams.seti.org/FDL/.