Data Engineering With Satellite Imagery

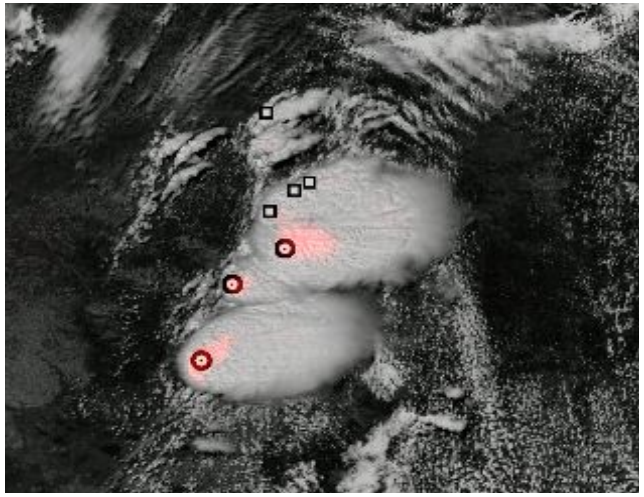
Arash Hosseini Jafari M.S.
NIFS Intern, Data Science
NASA Langley Research Center

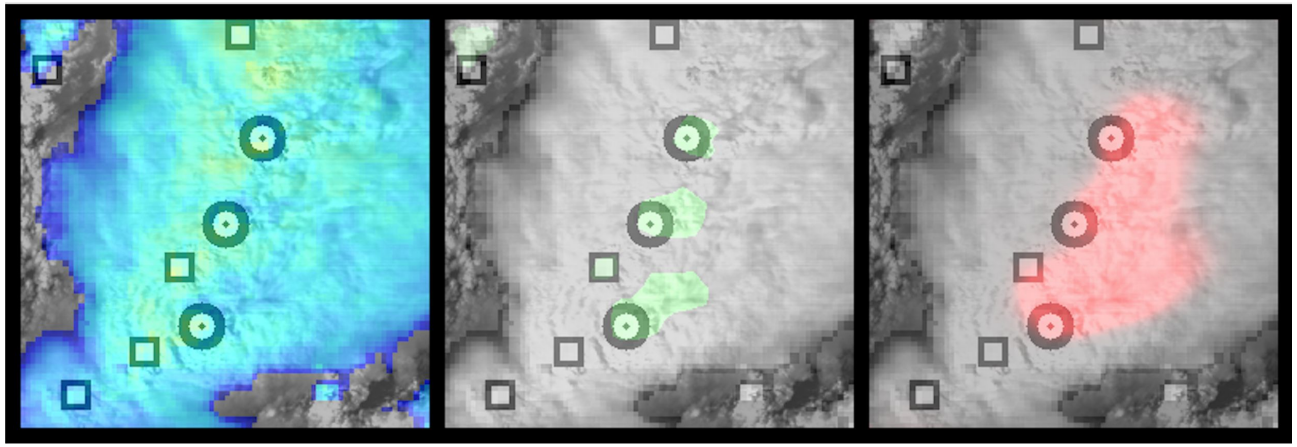
Fall 2019 Internship Presentation

Mentor: Charles Liles, OCIO Data Science Team
In collaboration with Kris Bedka, Science Directorate

The AACP Project Overview

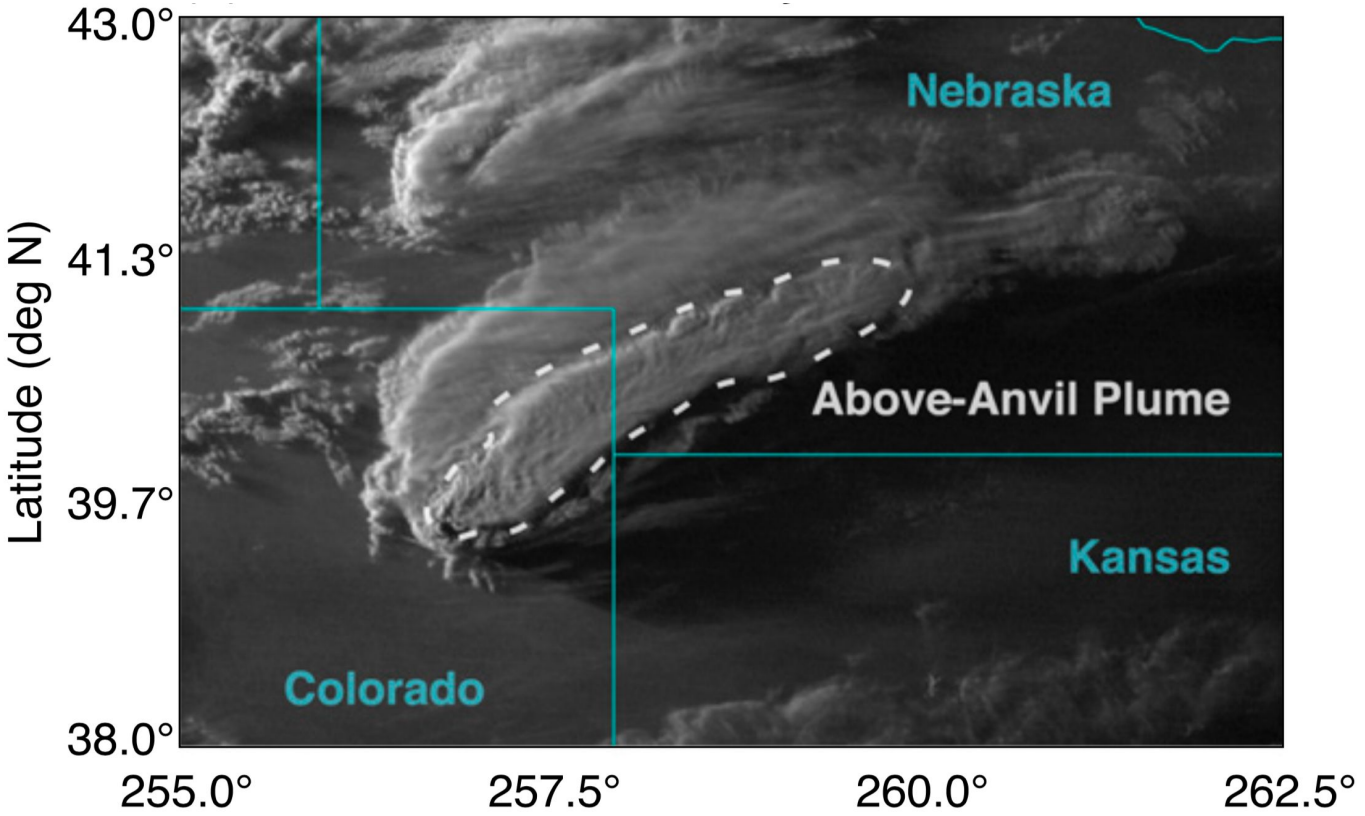
- Above Anvil Cirrus Plume (AACP), when seen on satellite imagery, is the most powerful predictor of severe weather events[1].
    - Explore the use of deep learning models to automate the detection of AACP's





- Left image is an example of model prediction.
- Right is sample of the input image, the hand labeled mask, and the model prediction (left to right)

Data Engineering With Satellite Imagery



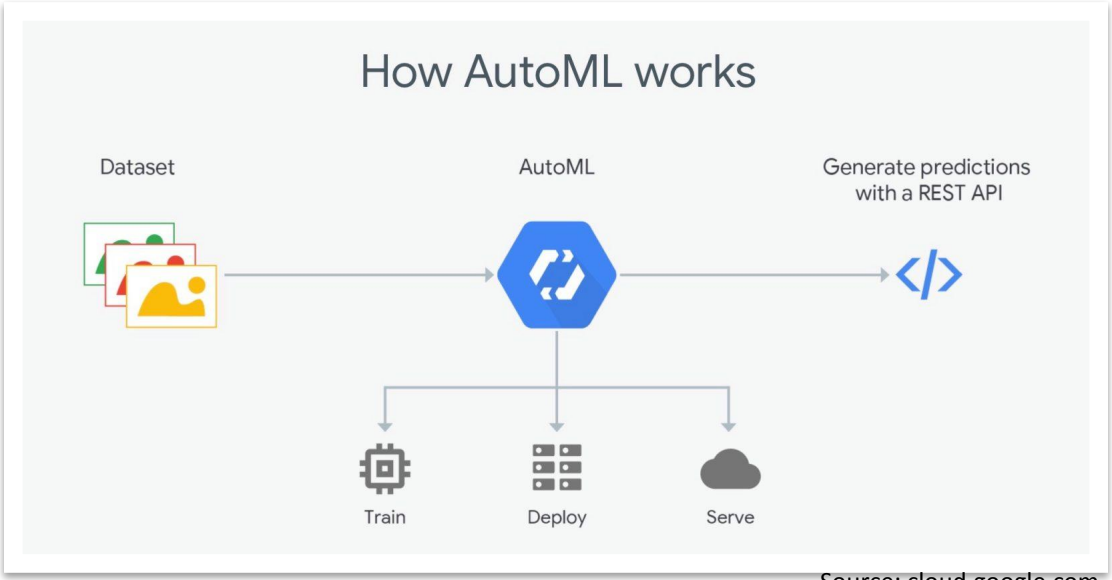
Map of GOES 1-km visible image for an above anvil cirrus plume case valid at 0115 UTC 21 May 2014. This case illustrates the extent of the observed above anvil cirrus plume [2].

Data Engineering With Satellite Imagery

- My Role:
  - Work on designing and building an automated data engineering pipeline with satellite infrared, visual, and timeseries data to enable input into GCP AutoML model (or other object tracking model that takes advantage of bounding boxes).
- GCP Input Data Requirements:
  - Framerate must be over 6fps we went with10fps
  - ~~Most popular video formats accepted (AVI, MP4, MOV, MPEG4)~~
  - Using H.264 AVI encoding
  - ~~Bounding box min size: 10x10 pix~~
  - 25 x 25 being used for AACP
  - Max input resolution w/o loss of quality: 1024x1024

Data Engineering With Satellite Imagery

- Google Cloud Platform AutoML Video Intelligence Object Tracking
  - Allows training for detection and tracking of multiple objects that appear in video segments.



Source: cloud.google.com

Data Engineering With Satellite Imagery

- Extensive data engineering is required prior to training models
  - Image engineering and computer vision techniques
    - IR + VIS images
    - Cropping region of interest (re-indexing)
    - Video production
      - Stitching together images based on timeseries data.
    - Annotations
      - Labeling
      - Bounding Box Calculation

Data Engineering With Satellite Imagery

- Approach: Wrote script to perform the following
  - Finding gaps and other anomalies in satellite feed
  - Downsizing images
  - Recalculating storm coordinates
  - Generate bounding boxes
  - Building videos from the images
  - Automatically produce CSV files required to train AutoML

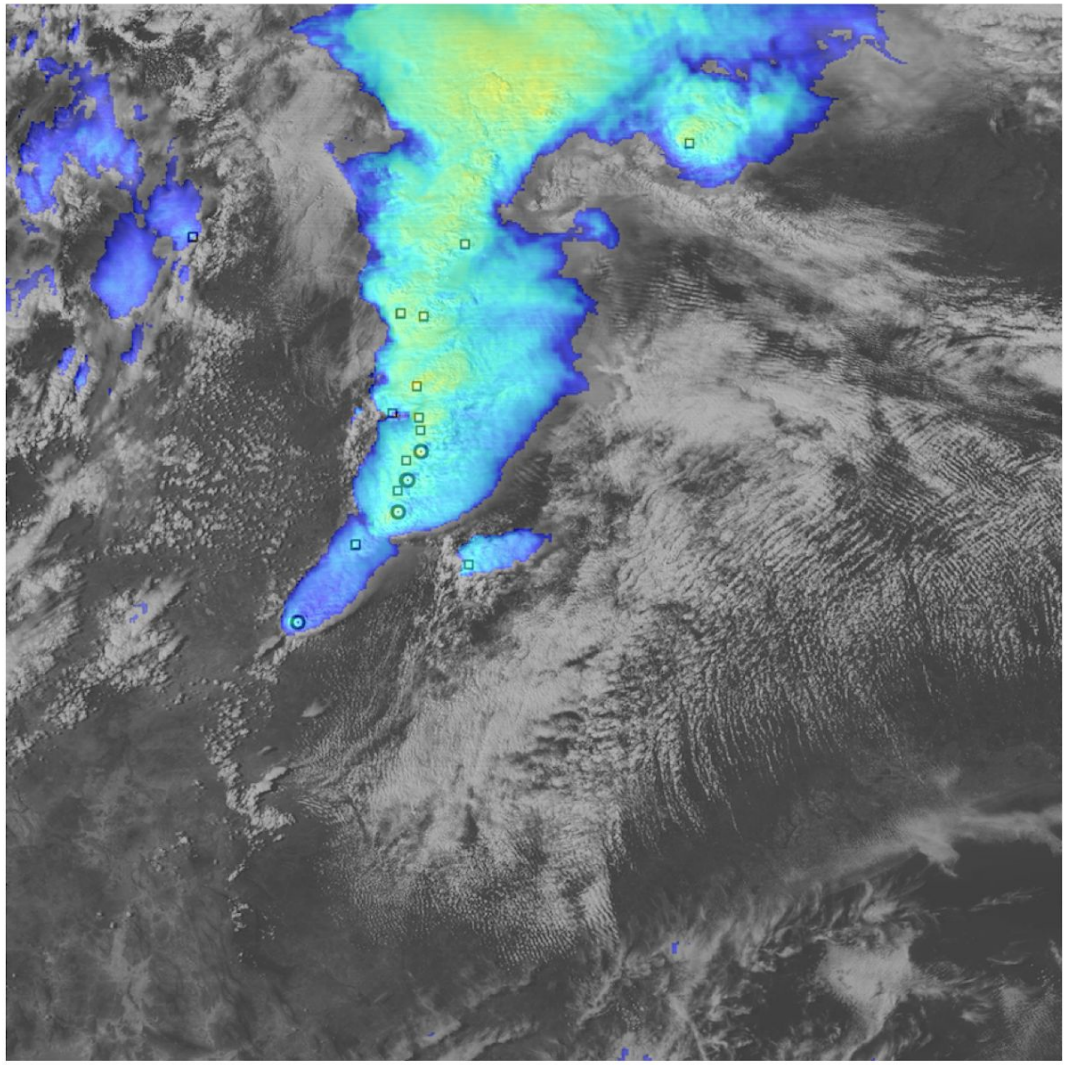
Data Engineering With Satellite Imagery

- Method:
  - Python
  - Pandas, OpenCV, Numpy, Matplotlib
  - Main Scripts:
    - csvmaker.py,
      - Finds gaps in satellite feed
      - Cleans data
        - Iterates through all available days of satellite data
        - SZA test
        - Recalculates all storm coordinates after 1024 by 1024 crop
          - Removes storms that are out of range after crop
        - Filters out missing data IR or Visual data
        - Calculates timestamp for each image based on framerate
        - Creates bounding boxes based on label coordinates
    - videomaker.py
      - Iterates through all satellite images over multiple days
      - Crops images based on the region with the most amount of storm activity (region of interest)
      - Produces video at a chosen frame rate and chosen video format

Data Engineering With Satellite Imagery

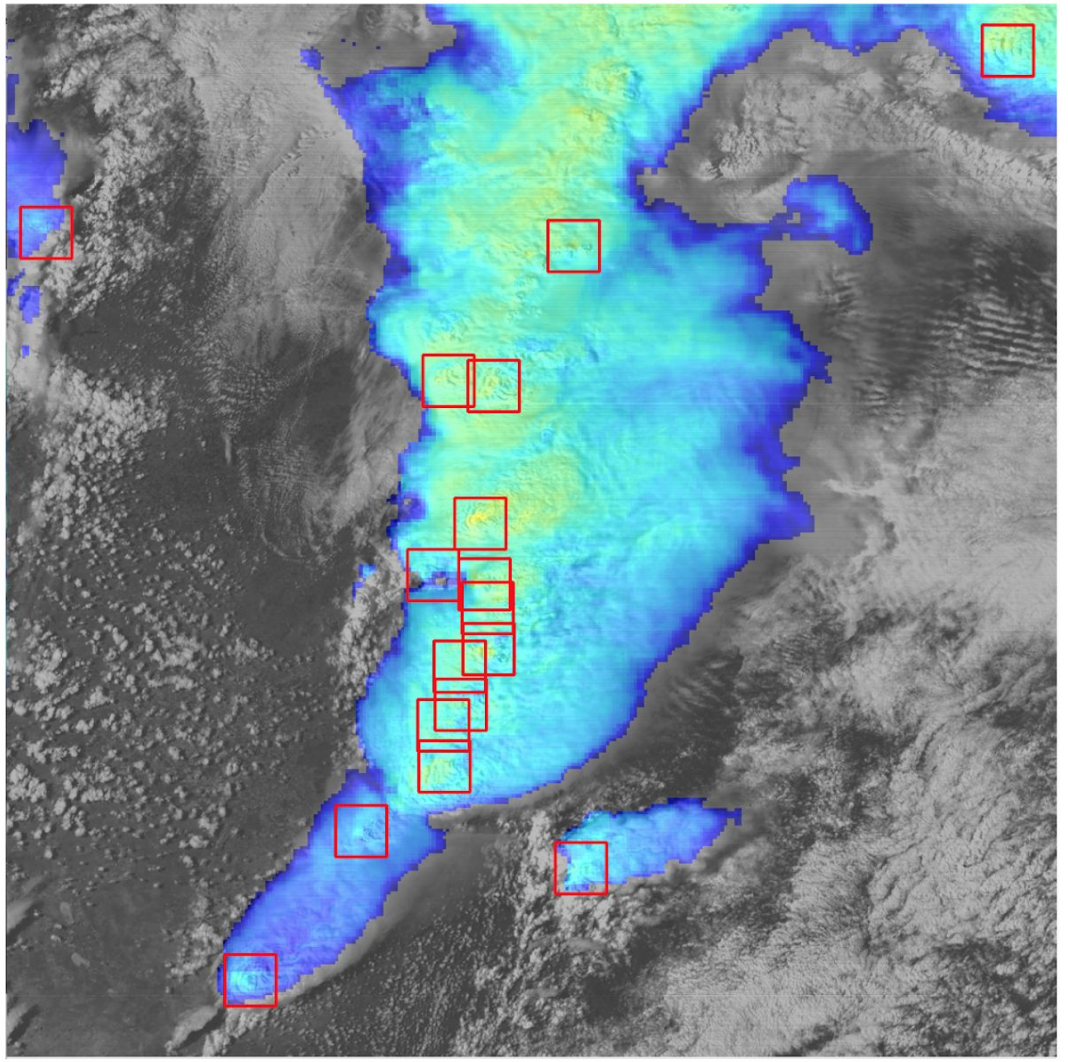
Original image

- 2000 x 2000



GCP Input Data

- 1024 x 1024 ROI with bounding boxes



Data Engineering With Satellite Imagery

- Other related tasks
  - Cost analysis
  - Troubleshooting GCP
  - Regularly scheduled meetings with Google's AutoML product team to solve issues and give feedback
  - Producing different iterations of the input data for AutoML
  - Optimizing and documenting data pipeline code
  - Experimentation with other neural networks and ML techniques
  - Giving weekly project updates to data science team
  - Putting together the AACP project poster presentation for the American Meteorological Society Conference

Data Engineering With Satellite Imagery

References:

1) Kristopher Bedka et al. 2018. The Above-Anvil Cirrus Plume Signature: The Most Definitive Indicator of a Severe Deep Convective Storm within Visible and Infrared Satellite Imagery (American Mariological Society Conference Presentation)

2) Homeyer CR, McAuliffe JD, Bedka KM. 2017. On the development of above-anvil cirrus plumes in extratropical convection. *J. Atmos. Sci.* **74**: 1617– 1633.

Data Engineering With Satellite Imagery

Thank You

Contact Information:

Arash Hosseini Jafari

arashh@uci.edu

ahoss001@ucr.edu

Linkedin: https://www.linkedin.com/in/arash-jafari-805789117/