Automated Detection of the Above Anvil Cirrus Plume Severe Storm Signature with Deep Learning

Charles Liles
Office of Chief Information Officer, NASA Langley Research Center, Hampton, Virginia

Kristopher Bedka
Science Directorate, NASA Langley Research Center, Hampton, Virginia

Edward Xia, YuXuan Huang, Roshni Biswas, Connor Dolan, and Arash Hosseini Jafari
NASA Internships, Fellowships & Scholarships, NASA Langley Research Center, Hampton, Virginia

Travis Smith
Science Applications International Corporation, NASA Langley Research Center, Hampton, Virginia

Submitted as an Article for Publication for the
19th Conference on Artificial and Computational Intelligence and its Applications to the Environmental Sciences

Jan 13, 2020

Corresponding Author Contact Information:
Charles Liles
NASA Langley Research Center
Mail Stop 50
Hampton, VA 23681-0001
Email: charles.a.liles@nasa.gov

ABSTRACT

The above anvil cirrus plume (AACP) is a weather phenomenon that signifies an intense tropopause-penetrating updraft which can inject cirrus clouds several kilometers into the stratosphere. Storms that have such intense updrafts are often supercells which generate severe weather such as tornadoes, high winds, and hail. In addition, AACPs moisten the stratosphere and influence the Earth's radiative balance. Though an AACP can be identified by the human eye, no automated AACP detection methods currently exist. Lack of detection inhibits understanding of where and how often AACPs occur, and how these storms influence stratospheric air composition. Previous work involved synthesis of multiple remote sensing and severe weather report/warning data sources to identify AACPs in Geostationary Operational Environmental Satellite system (GOES) satellite imagery and better understand their weather impacts. This current study demonstrates an automated AACP identification method based on the application of a deep learning segmentation model known as a U-net. This study documents the development of a U-net model capable of identifying emergent AACPs using only satellite infrared (IR) and visible reflected sunlight imagery. The performance of a U-net is quantitatively benchmarked with human AACP identifications and qualitatively assessed through animations of detections generated from GOES-16 1-minute temporal resolution imagery.

1. Introduction and Background

AACPs have been studied during the last 35 years and have been shown to be a precursor to severe weather patterns. Previous research has shown that "1) AACP storms are significantly

more likely to be severe than non-AACP storms, 2) AACPs regularly occur well in advance of severe weather, 3) severe weather occurs most often while an AACP is actively produced, and 4) the majority of severe weather reports, especially EF-2+ tornadoes and 2+ inch hail, are produced by storms with AACPs, and 5) early recognition of an AACP can provide comparable severe weather warning lead time to an expert NOAA National Weather Service forecaster" (Bedka et al. 2018).

Previous AACP research has successfully temporally and spatially fused data sources from GOES geostationary satellites (GEOsats) and U.S. Next Generation Weather Radar (NEXRAD) information. Storm updrafts identified in NEXRAD data were manually identified in GOES satellite IR and visible imagery and labeled as an active AACP or just a regular storm updraft (Bedka et al. 2018).

Although NEXRAD Doppler radar data has been successfully utilized in the U.S. to forecasts severe weather patterns, these radar systems are not available in many less well-developed nations where AACPs occur. However, GOES satellites are capable of collecting IR and visible imagery over these regions of the world with continuous observations collected at 10 minute intervals, and up to 30 second intervals during field campaigns or especially high impact weather events. The tasks of manually identifying AACPs in satellite imagery is time-intensive though, thus an automated AACP detection method is highly desired.

This study will establish an automated means of detecting AACPs using a machine learning model known as a U-net. U-nets were originally developed as a deep learning approach for the segmentation of neuronal structures in biomedical light microscopy images (Ronneberger et al.,

2015). This study will show the application of this same model to GOES satellite IR and visible imagery for the semantic segmentation of active AACPs.

2. Dataset and Exploratory Data Analysis

The dataset used for this study consisted of six days of GOES imagery and NEXRAD radar storm tracks taken over both the Great Plains and the southeast regions of the U.S. during 2017. The NEXRAD storm tracking method is described by Bedka et al. (2018) and Sandmael et al. (2019). GOES Imagery consisted of both Visible and IR imagery stored in Network Common Data Form (NetCDF) format, acquired from the University of Wisconsin Space Science and Engineering Center. IR imagery from GOES-16 Mesoscale Domain Sectors had a resolution of 500 x 500 pixels, whereas visible imagery had a resolution of 2,000 x 2,000 pixels. IR image pixels were replicated to match the dimensions of the visible image. Visible imagery reflectance varies throughout a day as clouds are illuminated differently as a function of solar zenith angle. Figures 2.a, 2.b, and 2.c below demonstrate the gradual decrease in reflectance from midday to sunset over the Great Plains region which can impact AACP detection capability.

The NEXRAD information consisted of radar tracks of storm updrafts which were sampled at five minute intervals, but interpolated to one minute intervals to match the GOES-16 imaging frequency. Each minute of storm lifetime had been labeled as either an active AACP or a non-AACP storm cell based on analysis by a human expert for a previous research effort (Bedka et al., 2018). Both AACP and non-AACP updrafts were observed to be simultaneously present in GOES imagery and often geographically close to one another as displayed in Figure 1 below.

Although these GOES and NEXRAD were not synchronized to acquire data at exactly the same time, the relatively high sampling rate ensured that satellite imagery could be temporally matched to radar storm tracks with only a maximum of 30 seconds of potential time-shift error between the two. The NetCDF files also contained latitude and longitude information for each pixel in an IR or visible satellite image. After applying corrections to the satellite latitudes and longitudes due to parallax, each pixel of satellite imagery could be spatially matched to storm track information in the NEXRAD dataset.

3. Data Engineering and Labeling

In this study, the data consisted of radar-observed storm cells which had been manually expert-labeled as either an AACP or a non-AACP. This data could be aligned temporally and spatially to find point labels of storm cell updrafts on both IR and visible GOES imagery. However, the process of originally labeling a storm cell as an AACP or non-AACP depended not only on visually identifying a storm updraft but also of observing a cirrus plume being ejected from the updraft. These cirrus plumes were not localized to a single pixel, but could be observed in several storms as flowing outward from the updraft up to hundreds of kilometers downwind in the stratosphere.

Because the manual process of identifying AACPs depended on the observation of an updraft with an active plume, the model selected for AACP identification was a U-net for semantic segmentation. Segmentation models rely on hand-labelled imagery with masks identifying separate classes of prediction. Utilizing a U-net for this study required methodical image preprocessing prior to model training. It also necessitated hand-labeling of both AACP updrafts and associated plumes in GOES images.

Satellite image “sandwiching” was selected as the method for combining both IR and visible satellite channels for input into the U-Net. This method has been used extensively by weather forecasters to view visible data in grayscale with an IR colormap overlay on top as shown in Figure 3 below (Valachová and Setvák, 2017). Sandwiching was selected as a means of preserving both cloud temperature and visible light texture information in both time and space for input to the machine learning model.

Prior to combining IR and visible data, each channel required separate preprocessing. IR pixel values greater than 233.15K were not used to build the colormap overlay. All remaining values less than or equal to 233.15K were projected into a dedicated colormap ranging in scale from 190K to 233.15K. Any value less than 190K was set to 190K. The main goal of this process was to project temperature pixel values to the same color for the colormap regardless of the range of temperature values present in a single image. This would ensure that the model could learn the significance of cloud temperatures across multiple images, the colormap range would not be constantly shifting between relative temperature ranges present for only a single image.

Visible image processing was designed to be robust to changes in light reflectance throughout the period of visible daylight. Unlike the IR imagery though, reflectance visible pixel values had a wide degree of variability throughout the period of visible sunlight as shown in Figures 2.a, 2.b, and 2.c below. Also, visible data was not available during night local time of observed areas.

The dataset was first subsetted to exclude periods where solar zenith angle (SZA) exceeded 82 degrees. Each visible image was normalized based on the minimum and maximum pixel values in the image. Contrast limited adaptive histogram equalization (CLAHE) was then

performed across the visible imagery in order to improve contrast in textured storm cloud images (Pizer et al. 2018). A comparison of the original visible imagery pre and post-CLAHE can be viewed in Figures 4.a and 4.b below.

Labeling data presented a challenge given the size and number of storms in the dataset. Previous efforts at labeling AACP and non-AACP updrafts were point labels centered on the storm updraft (Bedka et al. 2018). The U-net model required segmented images where an entire region of pixels comprising the AACP would be marked as an AACP as seen in Figure 5.b below. This labeling process involved drawing polygons across imagery which would encapsulate both an AACP's storm updraft as well as the cirrus plume produced by the updraft.

Fuzzy labeling was also utilized for drawing AACP polygons in the segmentation masks which would be used for model prediction. Storm updrafts which were not AACPs were not marked with a polygon and had a corresponding value of 0.0 in the labeled mask representing a probability of 0.0 for the existence of an AACP. Storm cells which had developed into easily distinguishable AACP updrafts with an associated ejected plume were segmented in mask polygons with a value of 1.0 for a 1.0 probability of being an AACP. However, storms which were transitioning between a state of non-AACP to AACP (or vice-versa) were often less readily-distinguishable. Polygons with a mask value of 0.5 were drawn around these storms area to represent a reduced degree of uncertainty. Fuzzy labeling was utilized to provide a smoother gradient for model training over storm draft areas in a transitory, uncertain state (Zadeh, 1965).

The six days of previously labeled data included 11,470 minutes of storm cell lifetime captured across 3,179 GOES images. Due to time and manpower constraints, only a subset of these images were labeled with polygons denoting an AACP cloud. Because the U-net image

input size was set at 256 x 256 image pixels, a reduced subset of the original 2,000 x 2,000 pixel imagery could be labeled without requiring complete labeling of all images in the dataset.

4. Model Training, Cross-Validation, and Evaluation of Results

Once the data had been prepared and properly labeled, the U-net model was built and trained to provide segmentation on a 256 x 256 pixel subset of sandwiched IR and visible imagery. The U-net architecture consisted of 4 blocks of convolutional and max pooling downsampling layers. These blocks were then upsampled in follow-on convolutional and upsampling blocks that were concatenated with the original downsampling block layers as per a standard U-net design (Hao, 2019). Dropout layers were also included within the model for regularization. The model architecture is depicted in Figure 6 below.

The model was trained with six-fold cross-validation to test goodness of fit across the entire dataset. Binary cross-entropy was utilized as the loss function with the Adam algorithm as an optimizer (Kingma and Ba, 2014). Details on cross-validation are provided below. Model overfitting was also prevented by monitoring a holdout validation set and stopping learning once improvement in validation performance began to deteriorate.

As mentioned above, the dataset consisted of AACP-producing storm systems which had been collected and labeled for six days of data. 6-fold cross-validation was utilized with each separate day of data representing a single fold. Fold stratification by day vice randomly sampling fold membership across the entire dataset ensured avoidance of inadvertent data leaks between training and test folds and consequent inflated model performance results. This cross-validation strategy was inspired by psychophysiology classification where non-stationarity in physiological

time-series signals and random sampling training and test data may lead to model overfitting on noise and poor model generalization (Brouwer et al. 2015). For each fold of cross-validation, four folds (days) were used for training, 1 fold for validation to prevent overfitting, and one fold for testing. Training, validation, and test fold membership were rotated as shown in Table 1 below.

Objective model evaluation proved to be challenging for this dataset. This was primarily due to the subjectivity of hand-labeling of AACP cloud boundaries, with bounds that may vary depending on the interpretation of the human analyst. Therefore we assess the quality of results based on both qualitative visual analysis and quantitative metrics. A naïve benchmark was utilized to determine if the model was subject to high bias. As mentioned above, the U-net model was trained using binary cross-entropy. This function is shown below:

$$L = -\frac{1}{n}\sum y * \ln(\hat{y}) + (1 - y) * \ln(1 - \hat{y})$$

Where *n* is the number of pixels, *y* is the "true" labeled value associated with a pixel, and $\hat{y}$ is the predicted probability from the model. High bias machine learning models often simply learn the underlying probability distribution within a training set's labels. A naïve benchmark was created simply by calculating binary cross entropy using the arithmetic mean of all pixels in each cross-validation folds' training set labels against the test folds' true label values. These were then compared against model prediction to ensure the model's binary cross entropy calculation was consistently less than the naïve benchmark for each fold of cross-validation. Results are shown in Table 2 below.

Intersection over union (IoU) was also utilized to evaluate model performance. IoU is a standard metric for measuring semantic segmentation model goodness of fit and is calculated by dividing true positives by the sum of false positives, true positives, and false negatives as shown

in Figure 7 below. IoU results across all folds of cross-validation is shown in the third column of Table 2 below. IoU averaged across all cross-validation folds was 0.33126. The IoU metric is typically used for the semantic segmentation of discrete objects in imagery (i.e. vehicles, pedestrians, roads, etc.). When applied for evaluation on semantic segmentation of a dynamic fluid like an AACP updraft and plume, the IoU metric can be penalized due to inherent uncertainty in boundary labeling and a tendency for humans to label areas in which they are most confident, thereby excluding diffuse plume edges. Therefore, qualitative analysis of model prediction was also performed post model training. As shown in Figure 8.c below, the U-net model has identified active AACP updrafts and downwind ejection of plumes. However, the area of AACP prediction does not exactly correspond with the hand-labeled AACP masks displayed in Figure 8.b against which model prediction was evaluated. For this image, the U-net model's IoU was 0.225. Purely quantitatively, the model appears to have performed poorly in identifying active AACPs. However, from a qualitative perspective, the model's performance appears stronger. It has identified active AACP updrafts as well as regions downwind associated with the ejected plume.

5. Summary

Above anvil cirrus plumes (AACP) often precede severe weather events such as tornadoes and severe hail. Automated detection of AACPs using satellite imagery alone would aid in the early detection of such events even in regions of the world where weather radar coverage is lacking. An automated detection mechanism would also improve understanding of the climatology of AACP storms and their impact on stratospheric composition. Using visible and IR satellite imagery

as input into a U-net deep convolutional neural network with radar-derived storm cell tracks and human AACP identifications as training output, we were able to achieve promising automated detection of AACPs.

6. Future Work

Future work for automated AACP identification will focus on incorporating temporal data into model training by leveraging deep learning models with memory. One particularly promising model for future exploration, implementation, and evaluation will be a fully connected network long short-term memory (FCN-LSTM) model which has been previously utilized for spatio-temporal identification of specific crops in satellite imagery (Teimouri et al. 2019).

Additionally, separate machine learning models for both updraft identification and plume detection will be explored. Lightning flash extent density from ground-based networks like the Earth Networks Total Lightning Network (Rudlosky 2015) or from the space-borne Geostationary Lightning Mapper instrument (Goodman et al. 2013) will also be incorporated as a third channel for updraft identification of active AACPs. The combined output of these two models may then be synthesized utilizing meta-modeling machine learning techniques. Results for model prediction will also be improved using probability threshold tuning.

Cloud-hosted machine learning technologies will also be investigated for automated AACP pattern identification. For example, the Google Cloud Platform (GCP) AutoML Video Intelligence service is specifically designed for pattern recognition in video. This off-the-shelf capability may provide high-fidelity predictions and/or a separate input for improving meta-model predictions.

7. Acknowdelgments

This work was supported by the NASA Langley Science Directorate and the NASA Langley Office of the Chief Information Officer (OCIO). It benefitted heavily from research conducted as part of the NASA ROSES 2015 "Severe Storms Research Program" (NNX15AV81G) as well as from Cameron Homeyer, Elisa Murillo, and Thea Sandmael from the University of Oklahoma for providing GridRad and storm cell track data.

8. References

Bedka, K. M., E.M. Murillo, C.R. Homeyer, B. Scarino, and H. Mersiovsky, 2018: The Above-Anvil Cirrus Plume: an Important Severe Weather Indicator in Visible and Infrared Satellite Imagery. Wea. Forecasting, 33, 1159–1181, https://doi.org/10.1175/WAF-D-18-0040.1

Brouwer, A.M., T.O. Thorsten, J.B.F. van Erp, J.E. Korteling, and A.W. Bronkhorst, 2015: Using Neurophysiological Signals that Reflect Cognitive or Affective State: Six Recommendations to Avoid Common Pitfalls. *J. Front. Neurosci.*, 30 April 2015. https://www.frontiersin.org/articles/10.3389/fnins.2015.00136/full

Goodman, S. J., and Coauthors, 2013: The GOES-R Geostationary Lightning Mapper (GLM). *Atmos. Res.*, **125**–**126**, 34–49, doi:10.1016/j.atmosres.2013.01.006.

Hao, Z.X., 2019: Implementation of Deep Learning Framework -- Unet, using Keras. https://github.com/zhixuhao/unet/

Kingma, D. and J. Ba, 2014: Adam: A Method for Stochastic Optimization, https://arxiv.org/abs/1412.6980v8

Pizer, S.M., E.P. Amburn, J.D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B.T.H. Romeny, J.B. Zimmerman, and K. Zuiderveld, 1987: Adaptive Histogram Equalization and its Variations, *Computer Vision, Graphics, and Image Processing,* **39**, 355-368, http://www.cs.unc.edu/Research/Image/MIDAG/pubs/papers/Adaptive%20Histogram%20Equalization%20and%20Its%20Variations.pdf

Ronneberger, O., P. Fischer, and T. Brox, 2015: U-Net: Convolutional Networks for Biomedical Image Segmentation, https://arxiv.org/abs/1505.04597

Rudlosky, S. D., 2015: Evaluating ENTLN Performance Relative to TRMM/LIS. *J. Operational Meteor.*, **3** (2), 11–20, doi: http://dx.doi.org/10.15191/nwajom.2015.0302.

Sandmæl, T.N., C.R. Homeyer, K.M. Bedka, J.M. Apke, J.R. Mecikalski, and K. Khlopenkov, 2019: Evaluating the Ability of Remote Sensing Observations to Identify Significantly Severe and Potentially Tornadic Storms. J. Appl. Meteor. Climatol., https://doi.org/10.1175/JAMC-D-18-0241.1

Teimouri, N., M. Dyrmannm, and R. N. Jørgensen, 2019: A Novel Spatio-Temporal FCN-LSTM Network for Recognizing Various Crop Types Using Multi-Temporal Radar Images. Remote Sens. 2019, 11, 990. https://www.mdpi.com/2072-4292/11/8/990

Valachová, M. and M. Setvák, 2017: Satellite Monitoring of the Convective Storms. EUMETSAT Workshop at ECMWF User Meeting: https://www.ecmwf.int/sites/default/files/elibrary/2017/17282-satellite-monitoring-convective-storms-forecasters-point-view.pdf

Zadeh, L.A., 1965: Fuzzy Sets. *Information and Control*, 8, 3. https://www.sciencedirect.com/science/article/pii/S001999586590241X

Fold	Training Data Dates	Validation Data Date	Test Data Date
1	20170516, 20170518, 20170628, 20170629	20170405	20170328
2	20170328, 20170518, 20170628, 20170629	20170516	20170405
3	20170328, 20170405, 20170628, 20170629	20170518	20170516
4	20170328, 20170405, 20170516, 20170629	20170628	20170518
5	20170328, 20170405, 20170516, 20170518	20170629	20170628
6	20170405, 20170516, 20170518, 20170628	20170328	20170629

Table 1: Cross-validation fold membership by date. Dates are listed in 'yyyymmdd' format.

Fold	Naïve Benchmark Loss	Model Test Set Loss	Model Test Set IoU
1	0.17614	0.11920	0.28538
2	0.25143	0.14934	0.40495
3	0.18237	0.11001	0.37995
4	0.21927	0.18067	0.32163
5	0.15833	0.11326	0.26869
6	0.13114	0.08464	0.32694

Table 2: Naïve binary cross entropy benchmark and model performance on the test set by cross-validation fold.

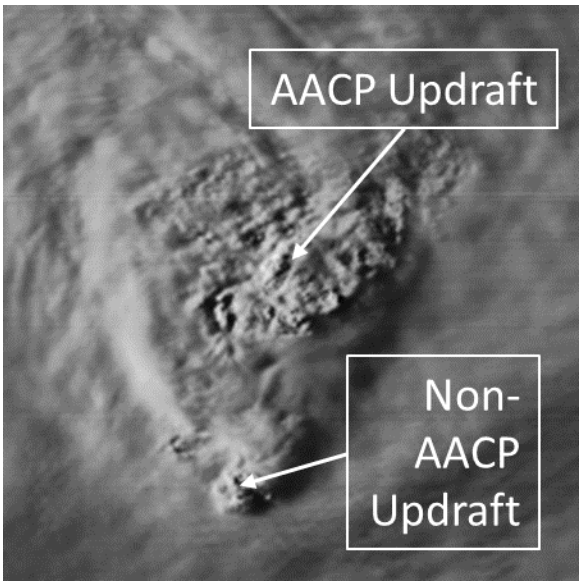

Figure 1: GOES-16 visible imagery of two closely located storm updrafts. The updraft in the center of the image is a fully developed AACP. The updraft to the south is a storm that has not developed into an AACP.

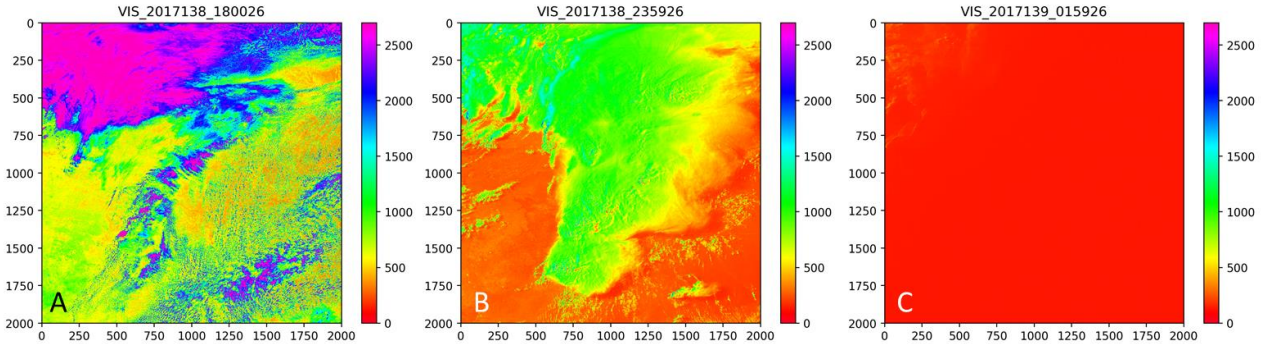

Figure 2: GOES visible images taken over the Great Plains region on May 18 and 19, 2017 at times (A) 18:00:26 UTC, (B) 23:59:26 UTC, and (C) 01:59:26 UTC. Visible reflectance is shown with a constant colormap across the imagery taken with varying levels of sunlight.

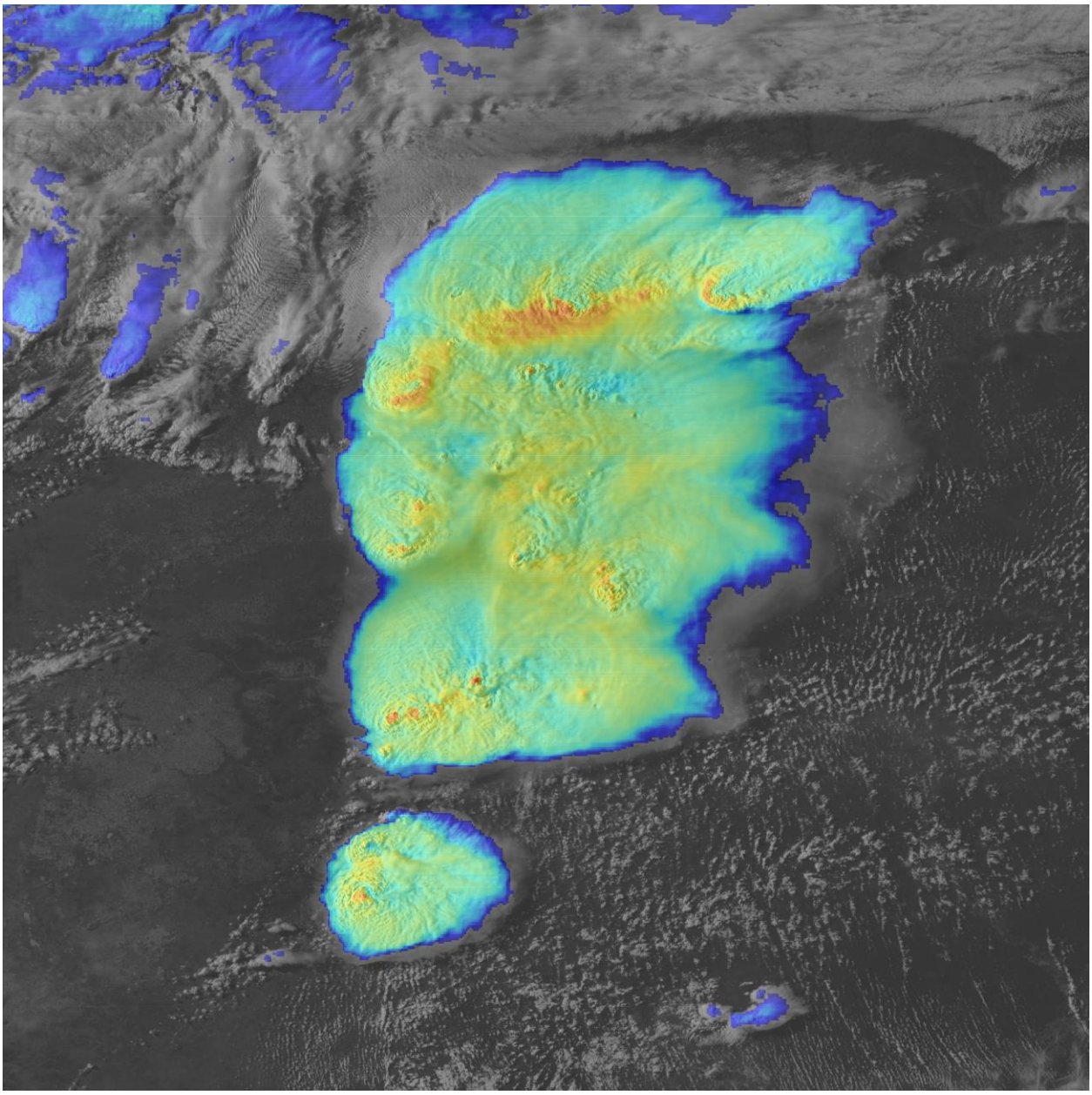

Figure 3: GOES-16 visible image with an IR colormap overlay “sandwich”.

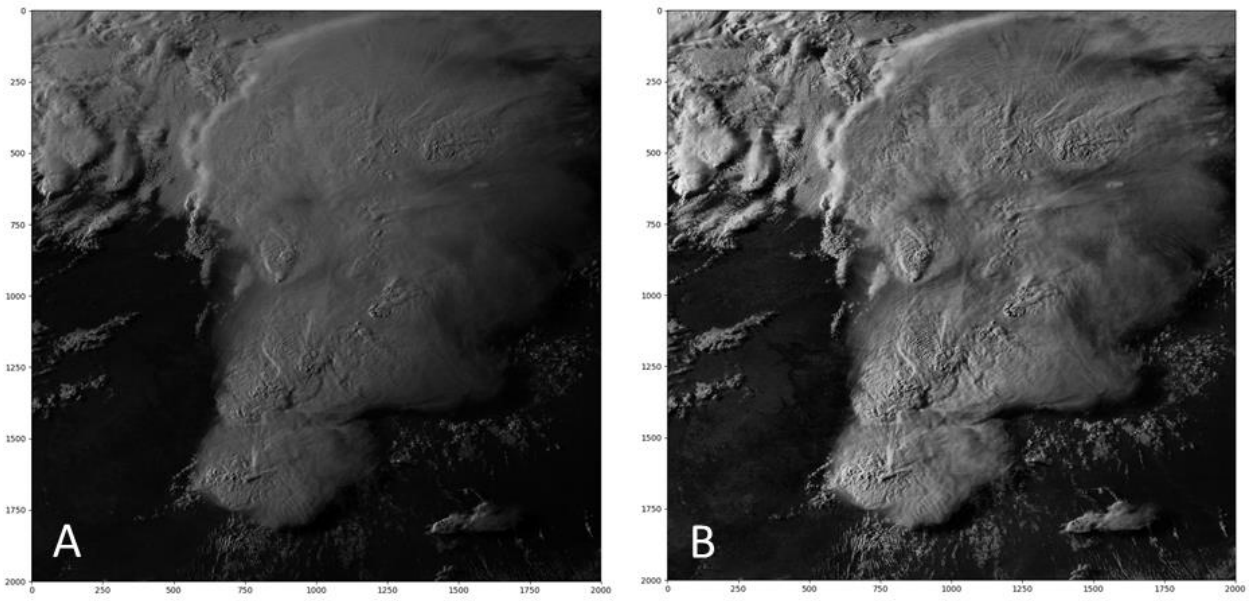

Figure 4: GOES 16 (A) original visible imagery and (B) image after CLAHE preprocessing

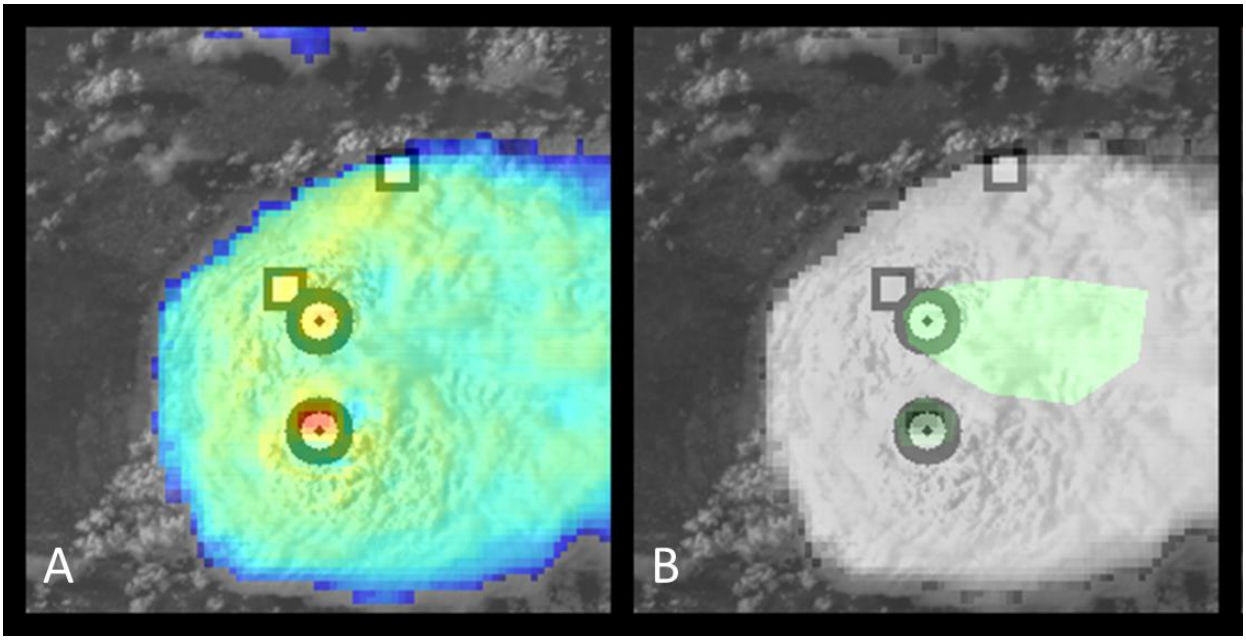

Figure 5: GOES 16 (A) visible-IR sandwiched imagery and (B) hand-labeled segmentation of active AACPs (green overlay). The original expert-labeled storm cells are also overlaid as black "bullseye" circles and black squares for AACP and non-AACP updrafts respectively.

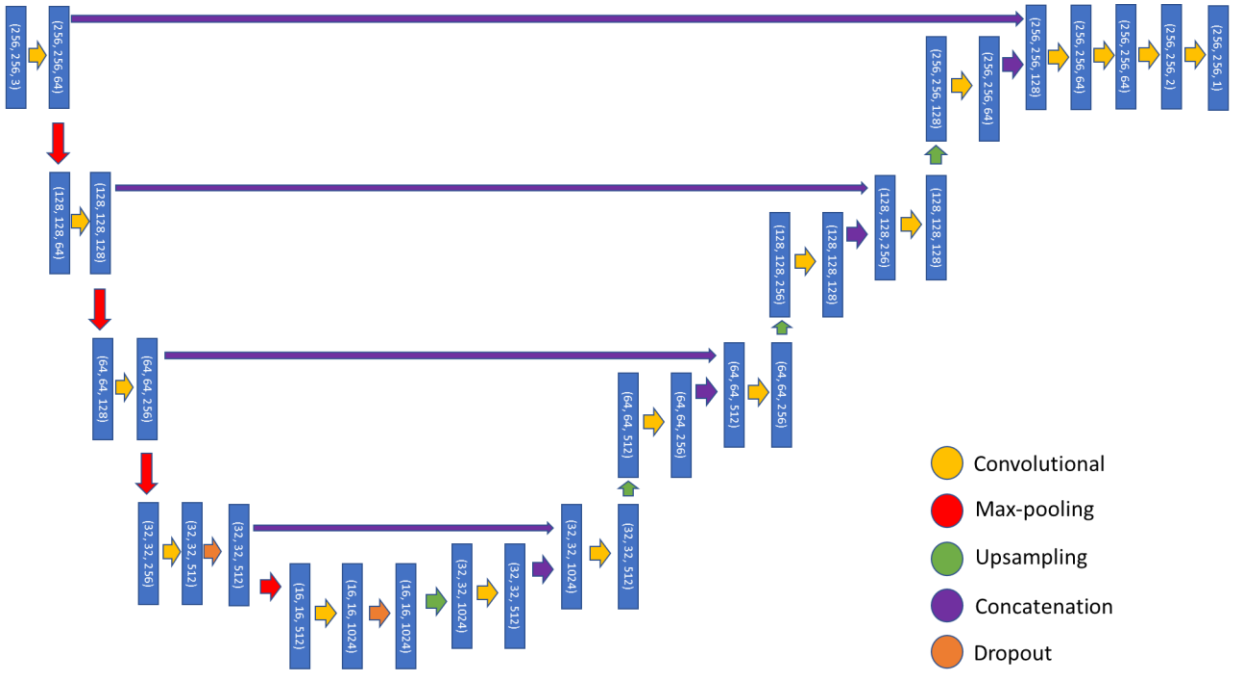

Figure 6: U-net model architecture.

$$IoU = \frac{TP}{FP + TP + FN} = \frac{\text{area of intersection}}{\text{area of union}}$$

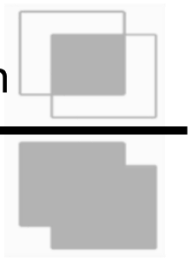

Figure 7: Intersection over union calculation.

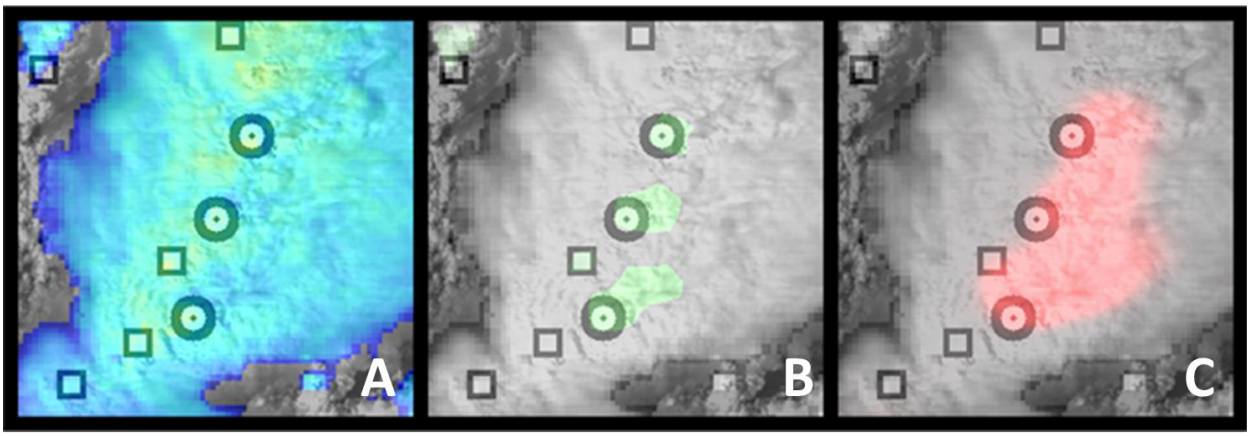

Figure 8: GOES 16 (A) visible-IR sandwiched imagery, (B) hand-labeled masks of AACP updrafts and plumes, and (C) U-net model prediction for AACP updrafts and plumes. The original expert-labeled storm cells are also overlaid as black "bullseye" circles and black squares for AACP and non-AACP updrafts respectively.