Leveraging Lightning with Convolutional Recurrent AutoEncoder and ROCKET for Severe Weather Detection

Nadia Ahmed
University of California, Irvine
ahmedn@uci.edu

Marek Marek Slipski
NASA Jet Propulsion Laboratory
marek.slipski@gmail.com

Iván Venzor-Cárdenas
Universidad Autónoma de Nuevo León
ivenzor@gmail.com

Maria J. Molina
National Center for Atmospheric Research
molina@ucar.edu

Greg Sénay
Augustus/xBrain
gregory.senay@gmail.com

Mark Cheung
Lockheed Martin
mark.cheung@gmail.com

Clem Tillier
Lockheed Martin
clem.tillier@lmco.com

Samantha Edgington
Lockheed Martin
samantha.f.edgington@lmco.com

Greg Renard
Augustus/xBrain
gregory.renard@xbrain.io

Abstract

Previous studies have shown that increases in flash rates detected in ground-based lightning data can be a precursor to severe weather hazards. Lightning data from the Geostationary Lightning Mapper (GLM) aboard the GOES-R satellite is not part of an operational model used by forecasters and is underutilized in severe storm research. The Advanced Baseline Imager's (ABI) visible imagery also shows cloud features, such as overshooting tops and above-anvil cirrus plumes, which have been associated with severe weather hazards. We introduce a generative video frame prediction methodology using a convolutional recurrent autoencoder, to leverage these spatio-temporal patterns in GLM and ABI, along with ground-based severe weather data. An initial case study is presented and contrasted with a time series classification of GLM data. Through this study, we seek to highlight the value of GLM data to assist meteorologists in time-constrained nowcasting (15-30 minute lead time) of severe hazards.

1 Introduction

Operational meteorologists are responsible for issuing alerts for impending severe weather with sufficient lead time to keep the public safe while reducing the number of false alarms. However, generating useful forecasts of severe thunderstorm hazards is a challenging task due to the variety of data sources, the complexity and scale of data, and the chaotic nature of cloud behaviors. Often, high computational costs of resolving convection in Numerical Weather Prediction (NWP) and physical models constrain the scale and precision of any modelling (e.g. Alexander et al. 2010). Scientific understanding of the microscale cloud features that can affect a thunderstorm's ability to produce severe hazards is limited in areas. Finally, the scale of available observations and model outputs (e.g. High-Resolution Rapid Refresh model) is beyond conventional analysis. While satellite data presents

AI for Earth Sciences Workshop at NeurIPS 2020.

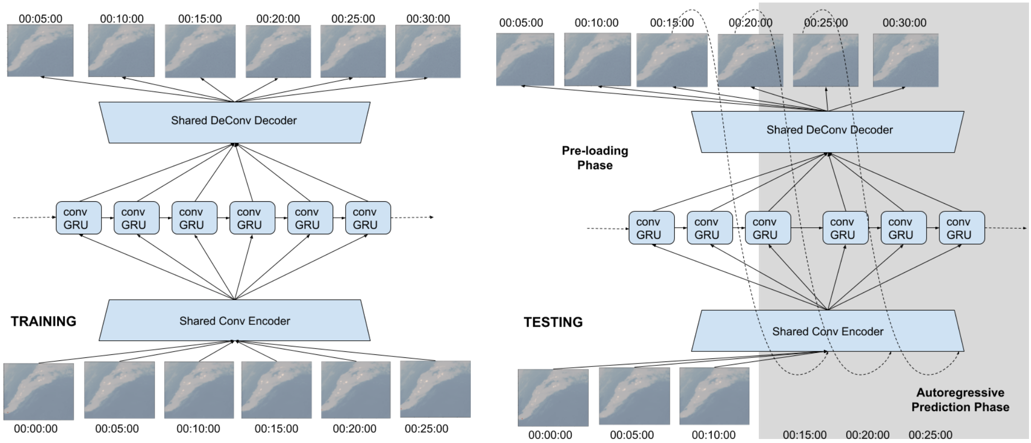

Figure 1: On the left, we train the model using Teacher Forcing. On the right, testing of the recurrent model follows a pre-loading phase during which the network acquires context to generate its first prediction. Following generation, the predictions are used as input for the sequence of frames corresponding to nowcasting constraints of 15 minutes.

an opportunity to explore trends in severe thunderstorm frequency and behavior, the application of machine learning across entire datasets can positively impact the field of weather prediction through the automation of data ingestion and extraction of salient features, accelerating improvements in the prediction and analysis of severe weather events.

1.1 Current Art

Various studies have been conducted that use machine learning to improve the lead time and accuracy of severe hazard predictions (e.g., Cintineo et al. 2018; 2020; Gagne II et al. 2019; Lagerquist et al. 2020). For example, convolutional neural networks (CNNs) have been used to create one-hour lead time forecasts of large hail from convection-permitting model output (Gagne II et al. 2019) and of tornadoes using radar (Lagerquist et al. 2020). Arguably one of the most comprehensive machine learning models for the prediction of severe hazards is the empirical Probability of Severe (ProbSevere) model (Cintineo et al. 2018). ProbSevere is an operational short-term forecasting subsystem within the Multi-Radar Multi-Sensor (MRMS) system for prediction of severe hazards developed by the National Oceanic and Atmospheric Administration (NOAA) and the Cooperative Institute for Meteorological Satellite Studies (CIMSS) at the University of Wisconsin (Cintineo et al. 2018). More recently, ProbSevere version 2 (Cintineo et al. 2020) contains forecast probabilities for specific hazards, such as tornadoes (ProbTor), large hail (ProbHail), and damaging winds (ProbWind). Both versions utilize several data sources, including ground-based lightning data and some geostationary satellite data, but these models do not incorporate data from the GLM. The large geographic scope and high temporal frequency of the underutilized GLM presents an opportunity to improve upon forecasts using lightning data from geostationary satellites.

2 Methodology

Lightning observations from the GLM onboard the GOES-R spacecraft (Goodman et al., 2013) are detected at 777.4 nm in a narrow band at 500 frames per second using a charge-coupled device (CCD) camera (1372 x 1300 pixels). The spatial resolution is approximately 8 km at nadir and 14 km at the edge of the disk. GLM data is publicly available as GLM Level 2 (L2) products, the output of the Lightning Cluster Filter Algorithm (Goodman et al, 2012). The L2 data contains information on individual lightning events, groups, and flashes. The colocated ABI provides 1500 x 2500 pixel multispectral images in the infrared, near-infrared, and the visible spectrum transmiting every 5 minutes. Additionally, NOAA National Center for Environmental Information (NCEI) confirmed severe event reports from the Storm Events dataset (and preliminary reports when confirmed unavailable) provide ground-based observations. These data can be represented as a point in a two dimensional space. To augment the data a Gaussian blur dilates the single point to an area two standard deviation from the center pixel.

To examine spatio-temporal patterns in the GLM flash extent density, ABI infrared channel 13, and the ground based tornado and hail report data, we utilize a convolutional recurrent autoencoder network architecture as shown in Figure 1. We further propose a strong baseline model: time series classification. Analysis of the GLM in relation to ground hail and tornado reports is performed using the RandOm Convolutional KErnel Transform (ROCKET) (Dempster et al. 2019), a technique that circumvents the computational overhead of traditional deep learning techniques by leveraging fast Fourier transforms to process the time series signal which is fed as input to a linear Ridge classifier. The skill of the model is then assessed using two widely used metrics in atmospheric science, the Critical Success Index (CSI) and the False Alarm Ratio (FAR) which rely on confusion matrix quantities true positives(TP), false negatives(FN), and false positives(FP) as described in Section 6.

3 Case Study

In this section we perform a case study using GLM data transmitted from the GOES-16 with a restricted geographic extent of 800 km x 1000 km region over the central United States (centered at 37°N, 97.5° W) isolated to spring of 2019. GLM level 2 quantities are transformed to 2 km x 2 km satellite relative fixed grid at a 1 minute cadence using the GLM tools package (Bruning et al., 2019). To reduce noise caused from daytime solar reflections, channel 13 (infrared at 10.3 microns), identified as indicator of cloud top temperatures, is selected to serve as a proxy for storm intensity. NOAA ground reports representing point space observations are isolated to hail and tornado reports for groups of frames corresponding to the time duration of the severe event.

Video Frame Prediction With the GLM, ABI, and Gaussian smoothed ground report data comprising each channel of a single frame respectively, a 3 channel, frame sequence was defined over 6 frames corresponding 30 minutes prior to, 5 minutes including, and 30 minutes following a severe event, producing volumes with a depth of 13 frames at a 5 minute cadence based on the available ABI data. The model constructed is a convolutional recurrent autoencoder (Figure 1) where the convolutional encoder is a two layer 2D convnet architecture with maxpooling and zero padding designed to decompose the frame sequence into feature maps summarizing the statistics of the original data while compressing the data to smaller dimensions. These 2D frame sequences serve as volume input to the recurrent unit comprised of three chained convGRU units mapping to a generated prediction of the next frame in the sequence. The prediction is then fed into a two layer shared deconvolutional decoder which upsamples the generated image to the original image dimensions for comparison with actual data from the previous timestep during training using Teacher Forcing. To train, Adam was used as a gradient optimizer to minimize the mean squared error loss. Training is done using truncated backpropagation through time (5 time steps) to avoid backpropagation through the entire dataset. In testing a pre-loading phase of 4 frames corresponding to 15 minutes was used to provide context followed by the autoregressive prediction phase whereby predicted frames were fed as input on the the next time step. The training and test sets contain 100 and 15 shuffled video sequences respectively for the month of May 2019.

Time Series Classification Using the 8 derived quantities from GLM gridded data, a time series in a 60 km x 60 km area centered on 1,161 severe hail events, 184 tornadoes and 740 null cases spanning March 2019 to June 2020 were constructed. 184 tornadoes, 184 hail events and 368 null events were randomly sampled. The tornadoes and hail formed the positive severe weather class. Sampled events were randomly split in a 70/30 proportion for training and testing respectively. Selected parameters of ROCKET relied on 10,000 convolutional kernels to effectively deconstruct the GLM time signal as a combination of 20,000 features. The hyper-parameter for the regularization of the classifier was cross-validated using 5 folds. We normalized the set to true and balanced class weights with recall as the score metric. These steps were repeated 100 times (for different sampling, split, and ROCKET parameters) to effectively create 100 different models to compute the confusion matrix metrics for each model. CSI and FAR metrics were assessed.

Performance Comparison for Spring 2019		
	ROCKET	State of Art
lead	15 min	> 15 min
coverage	Central US	CONUS
period	Spring 2019	4/14-7/14, 3/16-12/16
CSI	0.49	≈ 0.35
FAR	0.41	≈ 0.55

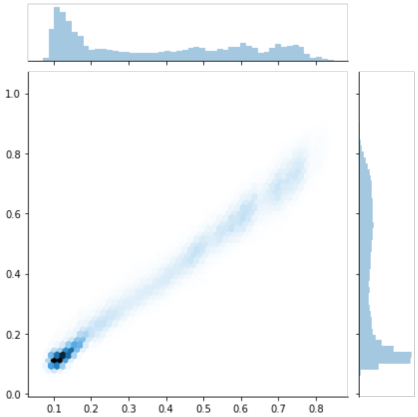

Figure 2: Averaged time series performance of 100 models generated via ROCKET of derived CSI and FAR metrics on left. Joint distribution and marginals of the ground truth (x-axis) and the prediction(y-axis) of the frame prediction model on right.

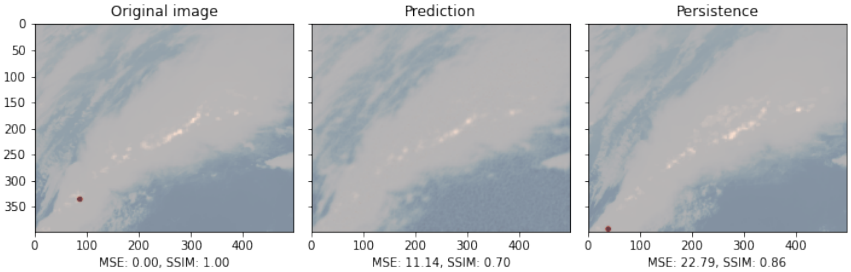

Figure 3: Sample generated frame prediction using encoder-convGRU-decoder architecture with respect to actual image and the frame 5 minutes prior. Red point indicates tornado location. Generated frame, middle, shows improvements in MSE loss from persistence model but needs further improvement for SSIM metric.

4 Results and Future Work

Predictions from the convolutional recurrent autoencoder show that generated frames have a smaller mean squared error (MSE) from ground truth than the persistence model for estimations within nowcast constraints of 15 minutes, 3 frames, for the GLM channel. For the ABI channel, the MSE of prediction is greater than persistence under the same constraints. When the ABI and GLM channels are averaged, the resulting MSE is improved for the very first prediction following the pre-loading phase for the recurrent autoencoder as seen in Figure 3. However, in terms of image quality metrics, the structural similarity index measure (SSIM) of the model prediction is consistently less accurate than the persistence model for all channels which may be a result of noise from upsampling. For predictions over fifteen minutes, the MSE and SSIM deteriorate for both persistence and the model, though the rate of degradation for the convolutional recurrent autoencoder is greater. This may depend on the length of the pre-loading phase, where longer phases allow the network to have a broader temporal context. Upon examination of the third severe weather channel, neither the recurrent autoencoder nor the persistence model is able to localize areas of potential tornado/hail events. Such outcome necessitates the use of additional information such as HRRR updraft helicity and wind shear. Analysis of the joint distribution of ABI and GLM data between the predicted frame and the actual frame show that the marginal distribution of the prediction follows the actual frame. Varying the type of severe weather event, or redefining the classification problem in terms of whether a storm core may result in tornadic or hail producing behavior may broaden the scope of the problem while still providing benefit to traditional practices.

The time series classification study of GLM in the context of severe weather events showed promising results. For the 100 trained time series models, results with respect to CSI and FAR (Figure 2) show ROCKET performs better than the state of the art (Cintineo et al. 2018) within the context of a fifteen minute nowcast constraint over the central US for the spring of 2019. However, caution is needed when comparing these metrics, as the state of the art provides greater lead time, larger geographic coverage, and a longer temporal range that includes cool season months as well as output from large physical models.

5 Conclusion

The study conducted supports the potential value of machine learning in the analysis of GLM data for severe thunderstorm prediction. Specifically, recurrent neural networks have been demonstrated to achieve great boosts in performance compared to linear models. In this paper we laid the foundations for such a recurrent neural network approach, and presented a time series classification. Our results suggest that the recurrent neural network approach still has significant potential which remains to be unlocked on this challenging problem.

References

[1] Alexander, Curtis R., S. S. Weygandt, T. G. Smirnova, S. Benjamin, P. Hofmann, E. P. James, and D. A. Koch. High Resolution Rapid Refresh (HRRR): Recent enhancements and evaluation during the 2010 convective season. In Preprints, *25th Conf. on Severe Local Storms*, Denver, CO, Amer. Meteor. Soc, vol. 9. 2010.

[2] Ashley, W. S., & Strader, S. M. (2016). Recipe for disaster: How the dynamic ingredients of risk and exposure are changing the tornado disaster landscape. *Bulletin of the American Meteorological Society*, 97(5), 767-786.

[3] Bedka, K., Brunner, J., Dworak, R., Feltz, W., Otkin, J., & Greenwald, T. (2010). Objective satellite-based detection of overshooting tops using infrared window channel brightness temperature gradients.*Journal of applied meteorology and climatology*, 49(2), 181-202.

[4] Bedka, K., Murillo, E. M., Homeyer, C. R., Scarino, B., & Mersiovsky, H. (2018). The above-anvil cirrus plume: An important severe weather indicator in visible and infrared satellite imagery.*Weather and Forecasting*, 33(5), 1159-1181.

[5] Brooks, H. E., & Correia Jr, J. (2018). Long-term performance metrics for National Weather Service tornado warnings.*Weather and Forecasting*, 33(6), 1501-1511.

[6] Bruning, E. C., Tillier, C. E., Edgington, S. F., Rudlosky, S. D., Zajic, J., Gravelle, C.,& Schultz, C. J. (2019). Meteorological Imagery for the Geostationary Lightning Mapper.*Journal of Geophysical Research: Atmospheres*, 124(24), 14285-14309.

[7] Cintineo, John L., Michael J. Pavolonis, Justin M. Sieglaff, Daniel T. Lindsey, Lee Cronce, Jordan Gerth, Benjamin Rodenkirch, Jason Brunner, & Chad Gravelle. The NOAA/CIMSS ProbSevere model: Incorporation of total lightning and validation.*Weather and Forecasting*, 33, no. 1 (2018): 331-345.

[8] Cintineo, J. L., Pavolonis, M. J., Sieglaff, J. M., Cronce, L., & Brunner, J. (2020). NOAA ProbSevere v2. 0—ProbHail, ProbWind, and ProbTor.*Weather and Forecasting*, 35(4), 1523-1543.

[9] Darden, C. B., et al. "Utilizing total lightning information to diagnose convective trends."*Bulletin of the American Meteorological Society* 91.2 (2010): 167-176.

[10] Dempster, Angus, François Petitjean, and Geoffrey I. Webb. "ROCKET: Exceptionally fast and accurate time series classification using random convolutional kernels."*Data Mining and Knowledge Discovery* (2020): 1-42.

[11] Doswell, C. A. (2001). Severe convective storms—An overview. In Severe convective storms (pp. 1-26).*American Meteorological Society*, Boston, MA.

[12] Fawaz, Hassan Ismail, et al. "Inceptiontime: Finding alexnet for time series classification."*arXiv preprint* arXiv:1909.04939 (2019).

[13] Gagne II, D. J., Haupt, S. E., Nychka, D. W., & Thompson, G. (2019). Interpretable deep learning for spatial analysis of severe hailstorms.*Monthly Weather Review*, 147(8), 2827-2845.

[14] Gatlin, P. N., and S. J. Goodman. A total lightning trending algorithm to identify severe thunderstorms.*Journal of atmospheric and oceanic technology* 27.1 (2010): 3-22.

[15] Goodfellow, I., Bengio, Y., & Courville, A. Deep Learning.*MIT Press*, 2012.

[16] Goodman, S. J., Mach, D., Koshak, W., & Blakeslee, R. (2012). GLM lightning cluster-filter algorithm.*Algorithm Theoretical Basis Document*, Ver, 3.

[17] Goodman, S. J., Blakeslee, R. J., Koshak, W. J., Mach, D., Bailey, J., Buechler, D., & Stano, G. (2013). The GOES-R geostationary lightning mapper (GLM).*Atmospheric research*, 125, 34-49.

[18] Lagerquist, R., McGovern, A., Homeyer, C. R., Gagne, D. J., & Smith, T. (2020). Deep Learning on Three-dimensional Multiscale Data for Next-hour Tornado Prediction. *Monthly Weather Review*.

[19] Schultz, C. J., Petersen, W. A., & Carey, L. D. (2011). Lightning and severe weather: A comparison between total and cloud-to-ground lightning trends.*Weather and forecasting*, 26(5), 744-755.

[20] Smith, A. B., & Matthews, J. L. (2015). Quantifying uncertainty and variable sensitivity within the US billion-dollar weather and climate disaster cost estimates. *Natural Hazards*, 77(3), 1829-1851.

[21] Verbout, S. M., Brooks, H. E., Leslie, L. M., & Schultz, D. M. (2006). Evolution of the US tornado database: 1954–2003.*Weather and Forecasting*, 21(1), 86-93.

6 Appendix

Model Architecture In the convolutional recurrent autoencoder architecture, the shared convolutional encoder extracts dominant spatial features leading up and following severe weather events treating each frame as a two-dimensional image. This is achieved by taking the volume, represented as a sequence of two dimensional image frames, and multiplying the time dimension with the batch size. The frames are then recombined as a volume by extracting the sequence from the batch size to the convGRU unit. This recurrent unit provides the temporal context for the evolution of the spatial data with respect to frames corresponding to 30 minutes prior to and following the severe weather event. The frames are then fed individually to a shared deconvolutional decoder to transform the compressed image produced by the shared convolutional encoder-convGRU back to the original dimensions to compare the generated image with the actual image from a prior time step using the mean squared error loss. This technique, formally known as Teacher Forcing (Goodfellow et al. 2016) utilizes the ground truth output from the training data as the input to the hidden recurrent unit on the next time step. Since hidden units are functions of earlier timesteps, we train with both Teacher Forcing along the network.

While Teacher Forcing is used for training, in testing, an adapted vanilla recurrent neural network evaluation protocol is utilized. During testing, the network is pre-loaded with testing data to gain context over a sequence of frames prior to making its first generated prediction. This generated prediction is then passed as the input frame to the network at the next time step as in Figure 1. This process continues corresponding to the number of frames required to nowcast, predict 15 minutes in advance, severe weather phenomenon.

ROCKET Evaluation Protocol The definition of the CSI and the FAR are based on the 2x2 contingency table (confusion matrix) quantities. True positive (TP) events are defined as forecasted events that are observed (‘hits’), and false negative (FN) events are defined as non-forecasted events that did occur (‘misses’). True negative (TN) events are defined as non-forecasted events that are not observed (‘correct rejections’), and false positive (FP) events are defined as forecasted events that are not observed (‘false alarms’). The skill of the model in terms of the CSI and FAR metric is defined as

$$CSI = \frac{TP}{TP + FN + FP}, \qquad FAR = \frac{FP}{FP + TP}, \tag{1}$$

where CSI represents the ratio of hits to the sum of hits, misses, and false alarms (higher values are better) and the FAR is defined as the ratio of false alarms to the sum of false alarms and hits (lower values are better).