Severe Weather Forecasting via the Sole Use of Satellite Data

Brianna D'Urso[1], Sheikh Rabiul Islam[2], Kamruzzaman Sarker[3], Ingrid Russell[1]

[1] Department of Computing Sciences, University of Hartford
[2] Department of Computer Science, Rutgers University-Camden
[3] Department of Computer Science, Bowie State University
{jdurso, irussell}@hartford.edu, sheikh.islam@rutgers.edu, ksarker@bowiestate.edu

Abstract

The forecasting of (severe) weather/climate systems using satellite telemetry and Machine Learning (ML) is generally held back by the size and availability of the pertaining datasets. This research outlines a newly devised pipeline for the automated construction of concise datasets designed to convert computationally expensive raw data from a netCDF4 database into a simpler format, with the end goal of future use in severe weather forecasting via the sole use of satellite data as an alternative to more conventional, expensive and localized means. By representing components of the dataset as int8 RGB(A) values of PNG images, data can be spatially related in a concise, consistent and visualizable manner that significantly reduces dataset size relative to the size of the raw dataset. This method is used on Atmospheric Motion Vectors (AMVs) derived from multispectral satellite telemetry via Optical flow Code for Tracking, Atmospheric motion vector, and Nowcasting Experiments (OCTANE) in the construction of a dataset capable of use in prediction of future movements of clouds.

Introduction[1]

Two common issues with the application of ML to Satellite Imagery Time Series (SITS) data are the sheer size of the raw datasets and the relatively few number of datasets available for training (Chen et al. 2023). The development of a framework/tool capable of the dynamic production of usable datasets from raw SITS data may be a useful step forwards. This research outlines and demonstrates a newly devised modular pipeline for the automated production of image-based SITS datasets from raw SITS data that are significantly smaller than the raw datasets.

Data Preparation and Methodology

Proper configuration and construction of the dataset used for training is paramount to the successful application of a ML model. ML models are optimization algorithms at their core. If something is not consistent, it cannot be optimized, hence consistency is of utmost importance.

Any image-based dataset can be used by anything else that can use images; such datasets can be natively utilized by any program or process which can use images, from image viewers to computer vision frameworks, allowing easier comprehension and analysis of data and expediting processes involved with working with data. Any vector field with up to four degrees of freedom can be simply and concisely depicted as different components of an RGB(A) image. Representation of vector components as colors is a simple and intuitive method for depicting data, particularly for data with localized relationships such as AMVs.

The PERiLS_2022 GOES-16 ABI Mesoscale Sector Data database (UCAR/NCAR 2022) is the source of all raw data used, composed of netCDF4 (Unidata 2012) data sourced from the Advanced Baseline Imager (ABI) (NASA/NOAA 2018) of the GOES-16 satellite. The data used pertains to all 24 hours of 4-03-22 to 4-07-22 (inclusive), chosen for the presence of tornadic activity. Only five days were used due to Computational Limitations (CL). OCTANE (Apke et al. 2023) is fed 3 ABI channels, each chosen for what they represent and how they interact; 08 (upper-level cloud water vapor, λ = 6,200 nm), 09 (mid-level cloud water vapor, λ = 6,900 nm), and 11 (cloud top phase, λ = 8,400 nm)[1] and outputs vector fields *U* (zonal flow) and *V* (meridional flow).

Custom Dataset Construction Pipeline

The image composition process allows AMV data to be represented and parsed by a model very consistently; any wind speed at any point in any timestep is mapped to the same data value for any point in any other timestep.

Data is processed considering the range that int8 data of RGB(A) PNGs can represent, 0-255. To normalize AMVs, 0 m/s *U* and 0 m/s *V* are set to 127, effectively setting 0 m/s to approximately the center of the range 0 to 255, thus allowing effective representation of both positive and negative wind velocities. To standardize AMVs, ±50 m/s lies on the borderline between what would be classified as an EF1 or an EF2 tornado. Data values higher than ±50 m/s for *U* or *V* are set to ±50 m/s; thus velocities from -50 m/s to +50 m/s are mapped to 0 to 255. *Rad* (radiance, i.e. light) values are normalized by mapping 0-5 to the range

0-255, since most *Rad* data lies below 5 with room to spare. The *Rad* data for this iteration of the dataset was taken from channel 08 of the ABI data. Any lost or broken data is manually picked out prior to use. A custom data loader made for use with MS-RNN prepares the data by decomposing the images into their component colors and stacking them vertically, making each image three times as tall while allowing them to be as monochrome images, and standardizes the range 0-255 (i.e. int8) to the range 0-1 prior to use by the model.

Preliminary Experimentation and Results

Extracting and combining only the data most necessary to the application (i.e. AMV prediction) saves computational resources in large amounts. The final size of the dataset generated for this demonstration is ~212 MB (MegaBytes), ~0.02% the raw dataset size (~112.1 GB (GigaBytes) compressed), and ~33% the size of the subset of raw data used to make it (~650 MB compressed); thus the dataset is much more concise, and the model need not consider relationships between data of focus (i.e. AMVs) and potentially contaminant data (i.e. unused ABI channels).

Training, validation, and testing were carried out via a modified version of the MultiScale Recurrent Neural Network (MS-RNN) (Ma et al. 2022a) framework using the same method employed by the MS-RNN team for the KTH action dataset. Similarly to their method, the dataset used for this demonstration was divided into 11 mutually exclusive subsets, seven of which were used for training, and four of which were used for testing. These subsets were put into SITS batches of 20 timesteps, where the first 10 timesteps are used to predict the next 10 timesteps. Further details on training and testing can be found in the MS-RNN paper.

The results show prediction of 10 future timesteps with an average Mean Square Error (MSE) increase of ~9.947 per frame (Figure 2) and an average decrease in SSIM (Structural Similarity Index Measure) of ~0.012 per frame (Figure 1) with MS-ConvLSTM. Table 1 compares results of tested models. Differences in settings between models were necessary due to CL derived from having no funding.

Table 1: Compared average results between models trained on the custom dataset via the MS-RNN framework/repository

Model	Time	Params	Loss	SSIM	MSE	MAE	PSNR	Settings
MS-ConvLSTM	2.81 H	1.773M	715.736	0.92	37.78	677.96	32.27	120 px 10 in, 10 out
ConvLSTM	3.49 H	1.773M	734.741	0.918	36.90	697.83	31.62	120 px 6 in, 6 out
MK-LSTM	6.33 H	7.577M	293.978	0.926	13.48	280.05	32.52	80 px 6 in, 6 out
MS-TrajGRU	3.17 H	1.916M	556.17	0.939	25.74	529.91	33.90	120 px 5 in, 5 out
TrajGRU	1.68 H	1.916M	341.313	0.921	19.61	320.76	31.20	80 px 4 in, 4 out

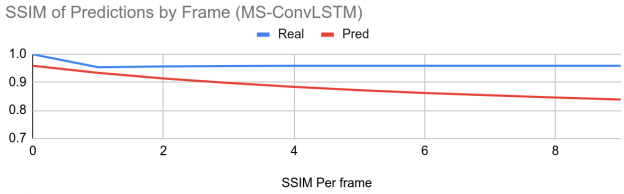

Figure 1: The blue line depicts the *Real* row of the above *SSIM per frame* table, and the red line depicts the *Pred* row.

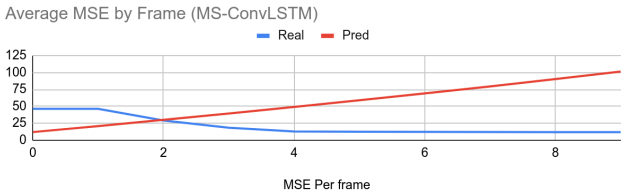

Figure 2: The blue line depicts the *Real* row of the *MSE Per Frame* table, and the red line depicts the *Pred* row.

Conclusions and Future Research

The dataset production pipeline, alongside the modified MS-RNN framework, shows potential for future work in prediction and analysis of SITS data and AMVs in a concise, simple, scalable, robust and generalizable manner, and may prove to be an accessible method for weather forecasting/nowcasting in a more cost-effective manner than more conventional localized, expensive and specialized ground equipment, and without the need/use of an extensive backlog of data. As climate change continues to progress, the weather patterns of our world are likely to become increasingly troublesome to analyze and account for. It is hoped that this work may be found useful as a tool, reference and/or building block in the future development of tools and techniques to help combat climate change as its prevalence continues to grow.

Potential in Future Research

Atmospheric phenomena are multiscale in nature. Use of other spatiotemporal scales has shown potential in testing. Conjoined use of multiple scales is also a potential avenue for future work. Lightning flash rates and size from the Geostationary Lightning Mapper (GLM) (NASA 2018) have been found to be strongly tied to “convective intensity and stratiform precipitation.” (Thiel et al. 2020). Tornadic systems can be characterized by vortices and convection currents, thus are capable of characterization by harmonics and oscillations. Signature detection on satellite data has shown viable for tornado detection (Apke et al. 2021) (Bruning et al. 2019); several forms of Proper Orthogonal Decomposition (POD) seem viable for this, such as MODULO/mPOD (Ninni and Mendez 2020), and SPOD (Mengaldo and Maulik 2021). Assimilating AMVs into analytical weather prediction models is known to improve predictions (Zhao et al. 2021).

References

NASA/NOAA. 2018. ABI Bands Quick Information Guides. www.goes-r.gov/mission/ABI-bands-quick-info.html.

Apke, J. M.; Daniels, J.; Hilburn, K.; Ebert-Uphoff, I.; Miller, S. D.; and Line, B. 2023. Advances in Optical Flow Retrieval Methods for Inferring Atmospheric Winds and Motions. Paper presented at the 16th International Winds Workshop Group Meeting. Montréal, Québec, May 8-12.

Schmit, T. J.; Lindstrom, S. B.; Gerth, J. J.; Gunshor, M. M. 2018. Applications of the 16 spectral bands on the Advanced Baseline Imager (ABI). *Journal of Operational Meteorology, 06*(04), 33-46. doi.org/10.15191/nwajom.2018.0604.

Harris, C. R.; Millman, K. J.; Van Der Walt, S. J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J.; Kern, R.; Picus, M.; Hoyer, S.; Van Kerkwijk, M.H.; Brett, M.; Haldane, A.; Del Río, J. F.; Wiebe, M.; Peterson, P.; Oliphant, T. E.; et al. 2020. Array programming with NumPy. *Nature, 585*(7825), 357-362.doi.org/10.1038/s41586-020-2649-2.

Shi, X.; Gao, Z.; Lausen, L.; Wang, H.; Yeung, D. Y.; Wong, W. K.; and Woo, W. C. 2017. Deep Learning for Precipitation Nowcasting: a benchmark and a new model. *arXiv preprint arXiv:1706.03478.*

NASA. 2018. Geostationary Lightning Mapper (GLM). *GHRC Lightning Research.* ghrc.nsstc.nasa.gov/lightning/overview_glm.html

Thiel, K.; Calhoun, K. M.; Reinhard, A. E.; and MacGorman, D. R. 2020. GLM and ABI Characteristics of Severe and Convective Storms. *Journal of Geophysical Research: Atmospheres, 125*(17). doi.org/10.1029/2020jd032858.

Marchetti, G. L.; Hillar, C. J.; Kragić, D.; and Sanborn, S. 2023. Harmonics of Learning: universal fourier features emerge in invariant networks. *arXiv preprint arXiv:2312.08550.*

Zhao, J.; Gao, J.; Jones, T. A.; and Hu, J. 2021. Impact of Assimilating High-Resolution Atmospheric Motion vectors on convective Scale Short-Term Forecasts: 1. Observing Simulation Experiment (OSSE). *Journal of Advanced in Modeling Earth Systems, 13*(10). doi.org/10.1029/2021ms002484.

Wang, X.; Cai, Z; Luo, Y.; Wen, Z.; and Ying, S. 2022. Long Time Series Deep Forecasting with Multiscale Feature Extraction and Seq2seq Attention Mechanism. *Neural Processing Letters, 54*(4), 3443-3466. doi.org/10.1007/s11063-022-10774-0.

Chen, L.; Han, B.; Wang, X.; Zhao, J.; Yang, W.; and Yang, Z. 2023. Machine Learning Methods in Weather and Climate Applications: a survey. *Applied Sciences, 13*(21), 12019. doi.org/10.3390/app132112019.

Hunter, J. D. 2007. MatPlotLib: a 2D Graphics environment. *Computing in Science and Engineering, 9*(3), 90-95. doi.org/10.1109/mcse.2007.55.

Bruning, E. C.; Tillier, C.; Edgington, S.; Rudlosky, S. D.; Zajic, J.; Gravelle, C. M.; Foster, M.; Calhoun, K. M.; Campbell, P. A.; Stano, G. T.; Schultz, C. J.; and Meyer, T. C. 2019. Meteorological imagery for the Geostationary Lightning Mapper. *Journal of Geophysical Research: Atmospheres, 124*(24), 14285-14309. doi.org/10.1029/2019jd030874

Ninni, D., and Mendez, M. A. 2020. MODULO: A software for Multiscale Proper Orthogonal Decomposition of data. *SoftwareX, 12,* 100622. doi.org/10.1016/j.softx.2020.100622.

Ma, Z.; Zhang, H.; and Liu, J. 2023. MS-LSTM: Exploring spatiotemporal multiscale representations in video prediction domain. *Applied Soft Computing, 147,* 110731. doi.org/10.1016/j.asoc.2023.110731.

Ma, Z.; Zhang, H.; and Liu, J. 2022a. MS-RNN: A Flexible Multi-Scale Framework for Spatiotemporal Predictive Learning. *arXiv preprint arXiv:2206.03010.*

Unidata. 2012. Network Common Data Form (netCDF) version 4.9.2. *UCAR/Unidata*. doi.org/10.5065/D6H70CW6.

Apke, J. M., and Mecikalski, J. R. 2021. On the Origin of Rotation Derived from Super Rapid Scan Satellite Imagery at the Cloud-Tops of Severe Deep Convection. *Monthly Weather Review*. doi.org/10.1175/mwr-d-20-0209.1.

UCAR/NCAR - Earth Observing Laboratory. 2022. PERiLS 2022: GOES-16 Advanced Baseline Imager (ABI) Mesoscale Sector Data. Version 1.0. *UCAR/NCAR - Earth Observing Laboratory.* doi.org/10.26023/GQMQ-Q2T1-TB0B

Ma, Z.; Zhang, H.; and Liu, J. 2022b. PrecipLSTM: A Meteorological Spatiotemporal LSTM for precipitation nowcasting. *IEEE Transactions on Geoscience and Remote Sensing, 60,* 1-8. doi.org/10.1109/tgrs.2022.3198222.

Mengaldo, G., Maulik, R. 2021. PySPOD: A Python package for Spectral Proper Orthogonal Decomposition (SPOD). *Journal of Open Source Software, 6*(60), 2862. doi.org/10.21105/joss.02862.

Schüldt, C.; Laptev, I.; and Caputo, B. 2004. Recognizing human actions: a local SVM approach. In Proceedings of the 17th International Conference on Pattern Recognition. doi.org/10.1109/ICPR.2004.1334462.

Han, J., and Wang, C. 2022. TSR-VFD: Generating temporal super-resolution for Unsteady Vector Field Data. *Computers & Graphics, 103*, 168-179. doi.org/10.1016/j.cag.2022.02.001.

Hoyer, S., and Hamman, J. 2017. xarray: N-D labeled Arrays and Datasets in Python. *Journal of Open Research Software, 5*(1), 10. doi.org/10.5334/jors.148.

Shi, X.; Chen, Z.; Wang, H.; Yeung, D. Y.; Wong, W. K.; and Woo, W. C. 2015. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. *arXiv preprint arXiv:1506.04214v2.*