Application of Deep Convolutional Neural Networks for Detecting Extreme Weather in Climate Datasets

Yunjie Liu[1]**,Evan Racah**[1]**,Prabhat**[1]**,Joaquin Correa**[1]**,Amir Khosrowshahi**[2]**,**
David Lavers[3]**,Kenneth Kunkel**[4]**,Michael Wehner**[1]**,William Collins**[1]

[1]Lawrence Berkeley Lab, Berkeley, CA, US
[2]Nervana Systems, San Diego, CA, US
[3]Scripps Institution of Oceanography, San Diego, CA, US
[4]National Oceanic and Atmospheric Administration, Asheville, NC, US

Abstract— *Detecting extreme events in large datasets is a major challenge in climate science research. Current algorithms for extreme event detection are build upon human expertise in defining events based on subjective thresholds of relevant physical variables. Often, multiple competing methods produce vastly different results on the same dataset. Accurate characterization of extreme events in climate simulations and observational data archives is critical for understanding the trends and potential impacts of such events in a climate change content. This study presents an application of Deep Learning techniques as alternative methodology for climate extreme events detection. Deep neural networks are able to learn high-level representations of a broad class of patterns from labeled data. In this work, we developed deep Convolutional Neural Network (CNN) classification system and demonstrated the usefulness of Deep Learning technique for tackling climate pattern detection problems. Coupled with Bayesian based hyper-parameter optimization scheme, our deep CNN system achieves 89%-99% of accuracy in detecting extreme events (Tropical Cyclones, Atmospheric Rivers and Weather Fronts).*

Keywords: Pattern Recognition, Deep Learning; Convolutional Neural Network; Climate Analytics; Extreme Events

1. Introduction

Extreme climate events (such as hurricanes and heat waves) pose great potential risk on infrastructure and human health. Hurricane Joaquin, for example, hit Carolina in early October 2015, and dropped over 2 feet of precipitation in days, resulted in severe flooding and economic loss. An important scientific goal in climate science research is to characterize extreme events in current day and future climate projections. However, understanding the developing mechanism and life cycle of these events as well as future trend requires accurately identifying them in space and time. Satellites acquire 10s of TBs of global data every year to provide us with insights into the evolution of the climate system. High resolution climate models produces 100s of TBs of data from multi-decadal run to enable us to explore future climate scenarios under global warming. Detecting extreme climate events in terabytes of data presents an unprecedented challenge for climate science.

Existing extreme climate events (e.g. hurricane) detection methods all build upon human expertise in defining relevant events based on evaluating of relevant spatial and temporal variables on hard and subjective thresholds. For instance, tropical cyclones are strong rotating weather systems that are characterized by low pressure and warm temperature core structures with high wind. However, there is no universally accepted sets of criteria for what defines a tropical cyclone [1]. The "Low" pressure and "Warm" temperature are interpreted differently among climate scientists, therefore different thresholds are used to characterize them. Researchers [2], [3], [4], [5], [6], [7] have developed various algorithms to detect tropical cyclones in large climate dataset based on subjective thresholding of several relevant variables (e.g. sea level pressure, temperature, wind etc.). One of the general and promising extreme climate event detecting software, Toolkit for Extreme Climate Analysis (TECA) [6], [7], is able to detect tropical cyclones, extra-tropical cyclones and atmospheric rivers. TECA utilizes the MapReduce paradigm to find pattern in Terabytes of climate data with in hours. However, many other climate extreme events do not have a clear empirical definition (e.g. extra-tropical cyclone and mesoscale convective system), which precludes the development and application of algorithms for detection and tracking. This study attempts to search for an alternative methodology for extreme events detection by designing a neural network based system that is capable of learning a broad class of patterns from complex multi-variable climate data, thus avoiding subjective threshold.

Recent advances in deep learning have demonstrated exciting and promising results on pattern recognition tasks, such as ImageNet Large Scale Visual Recognition Challenge [8], [9], [10] and speech recognition [11], [12], [13], [14]. Many of the state-of-art deep learning architectures for visual pattern recognition are based on the hierarchical feature learning convolutional neural network (CNN). Modern CNN systems tend to be deep and large with many hidden layers and millions of neurons, making them flexible in learning a broad class of patterns simultaneously from data. AlexNet (7

layers with 5 convolutonal layer and 2 fully connected layer) developed by [8] provides the first end to end trainable deep learning system on objective classification, which achieved 15.3% top-5 classification error rate on ILSVRC-2012 data set. On the contrary, previous best performed non-neural network based systems achieved only 25.7% top-5 classification error on the same data set. Shortly after that, Simonyan and Zisserman [9] further developed AlexNet and introduced an even deeper CNN (19 layers with 16 convolutional layer and 3 fully connected layer) with smaller kernel (filter) and achieved an impressively 6.8% top-5 classification error rate on ILSVRC-2014 data set. Szegedy et al.[10] introduced the "inception" neural network concept (network includes sub-network) and developed an even deeper CNN (22 layers) that achieved comparable classification results on ImageNet benchmark. Build on deep CNN, Sermanet et al. [15] introduced an integrated system of classification and detection, in which features learned by convolutional layers are shared among classification and localization tasks and both tasks are performed simultaneously in a single network. Girshick et al. [16] took a completely different approach by combining a region proposal framework [17] with deep CNN and designed the state of art R-CNN object detection system.

In this paper, we formulate the problem of detecting extreme climate events as classic visual pattern recognition problem. We then build end to end trainable deep CNN systems, following the architecture introduced by [8]. The model was trained to classify tropical cyclone, weather front and atmospheric river. Unlike the ImageNet challenge, where the training data are labeled natural images, our training data consist of several continuous spatial variables(e.g. pressure, temperature, precipitation) and are stacked together into image-like patches.

2. Related Work

Climate data analysis requires an array of advanced methodology. Neural network based machine learning approach, as a generative analysis technique, has received much attention and been applied to tackle several climate problems in recent year. Chattopadhyay et al. [18] developed a nonlinear clustering method based on Self Organizational Map (SOM) to study the structure evolution of Madden-Julian oscillation (MJO). Their method does not require selecting leading modes or intraseasonal bandpass filtering in time and space like other methods do. The results show SOM based method is not only able to capture the gross feature in MJO structure and development but also reveals insights that other methods are not able to discover such as the dipole and tripole structure of outgoing long wave radiation and diabatic heating in MJO. Gorricha and Costa [19] used a three dimensional SOM on categorizing and visualizing extreme precipitation patterns over an island in Spain. They found spatial precipitation patterns that traditional precipitation index approach is not able to discover,

and concluded that three dimensional SOM is very useful tool on exploratory spatial pattern analysis. More recently, Shi et al. [20] implemented a newly developed convolutional long short term memory (LSTM) deep neural network for precipitation nowcasting. Trained on two dimensional radar map time series, their system is able to outperform the current state-of-art precipitation nowcasting system on various evaluation metrics. Iglesias et al. [21] developed a multitask deep fully connected neural network on prediction heat waves trained on historical time series data. They demonstrate that neural network approach is significantly better than linear and logistic regression. And potentially can improve the performance of forecasting extreme heat waves. These studies show that neural network as a generative method and can be applied on various climate problems. In this study, we explore deep CNN on solving climate pattern detection problem.

3. Methods

3.1 Convolutional Neural Network

A Deep CNN is typically comprised of several convolutional layers followed by a small amount of fully connected layers. In between two successive convolutional layers, subsampling operation (e.g. max pooling, mean pooling) is performed typically. Researchers have questioned about the necessity of pooling layers, and argue that they can be simply replaced by convolutional layer with increased strides, thus simplify the network structure [22]. In either case, the inputs of a CNN is (m,n,p) images, where m and n is the width and height of an image in pixel, p is the number of color channel of each pixel. The output of a CNN is a vector of q probability units (class scores), corresponding to the number of categories to be classified (e.g. for binary classifier q=2).

The convolutional layers perform convolution operation between kernels and the input images (or feature maps from previous layer). Typically, a convolutional layer contains k filters (kernels) with the size (i,j,p). Where i,j is the width and height of the filter. The filters are usually smaller than the width m and height n of input image. p always equal to the number of color channel of input image (e.g. a color image has three channels: red, green, and blue). Each of the filters is independently convolved with the input images (or feature maps from previous layer) followed by non-linear transformation and generates k feature maps, which serve as inputs for the next layer. In the process of convolution, a dot product is computed between the entry of filter and the local region that it is connected to in the input image (or feature map from previous layer). The parameters of convolutional layers are these learnable filters. Sliding convolutional kernels across all the input will produce larger outputs for certain sub-regions than for others. This allows features to be extracted from inputs and preserved in the feature maps regardless of where the feature is located in

the input. The pooling layer subsamples the feature maps generated from convolutional layer over a *(s,t)* contiguous region, where *s,t* is the width and height of the subsampling window. This operation reduces the resolution of feature maps with the depth of CNN. All feature maps are high-level representations of the input data. The fully connected layer has connections to all hidden units in previous layer. If it is the last layer within CNN architecture, the fully connected layer also does the high level reasoning based on the feature vectors from previous layer and produce final class scores for image objects.

3.2 Hyper-parameter Optimization

Training deep neural network is known to be hard [23], [24]. Effectively and efficiently train deep neural network not only requires large amount of training data, but also requires carefully tuning model hyper-parameters (e.g. learning parameters, regularization parameters) [25]. The parameter tuning process, however, can be tedious and non-intuitive. Hyper-parameter optimization can be reduced to find a set of parameters for a network that produces the best possible validation performance. As such, this process can be thought of as a typical optimization problem of finding a set, x, of parameter values from a bounded set X that minimize an objective function $f(x)$, where x is a particular setting of the hyper-parameters and $f(x)$ is the loss for a deep neural network with a particular set of training and testing data as function of the hyper-parameter inputs. Training a deep neural network is not only a costly (with respect to time) procedure, but a rather opaque process regarding to how the network performance varies with respect to its hyper-parameter inputs. Because training and validating a deep neural network is very complicated and expensive, Bayesian Optimization (which assumes $f(x)$ is not known, is non-convex and is expensive to evaluate) is a well-suited algorithm for hyper-parameter optimization for our task at hand. Bayesian Optimization attempts to optimize $f(x)$ by constructing two things: a probabilistic model of $f(x)$ and an acquistion function that picks which point x in X to evaluate next. The probabilistic model is updated with Bayesian rule with a Gaussian prior. The acquisition function suggests hyper-parameter settings or points to evaluate by trying to balance evaluating parameter settings in regions, where $f(x)$ is low and points in regions where the uncertainty in the probabilistic model is high. As a result the optimization procedure attempts to evaluate as few points as possible [26], [25]. In this study, we use spearmint (https://github.com/JasperSnoek/spearmint) for performing network hyper-parameter optimization.

3.3 CNN Configuration

Following AlexNet [8], we developed a deep CNN which has totally 4 learnable layers, including 2 convolutional layers and 2 fully connected layers. Each convolutional layer

is followed by a max pooling layer. The model is constructed based on the open source python deep learning library NOEN (https://github.com/NervanaSystems/neon). The configuration of our best performed architectures are shown in Table 1.

The networks are shallower and smaller comparing to the state-of-art architecture developed by [9], [10].The major limitations for exploring deeper and larger CNNs is the limited amount of labeled training data that we can obtain. However, a small network has the advantage of avoiding over-fitting, especially when the amount of training data is small. We also chose comparatively large kernels (filters) in the convolutional layer based on input data size, even though [9] suggests that deep architecture with small kernel (filter) is essential for state of art performance. This is because climate patterns are comparatively simpler and larger in size as compared to objects in ImageNet dataset.

One key feature of deep learning architectures is that it is able to learn complex non-linear functions. The convolutional layers and first fully connected layer in our deep CNNs all have Rectified Linear Unit (ReLU) activation functions [27] as characteristic. ReLU is chosen due to its faster learning/training character [8] as compared to other activation functions like Tanh.

$$f(x) = max(0, x) \qquad (1)$$

Final fully connected layer has Logistic activation function as non-linearity, which also serves as classifier and outputs a probability distribution over class labels.

$$f(x) = \frac{1}{1 + e^{-x}} \qquad (2)$$

3.4 Computational Platform

We performed our data processing, model training and testing on Edison, a Cray XC30 and Cori, a Cray XC40 supercomputing systems at the National Energy Research Scientific Computing Center (NERSC). Each of Edison computing node has 24 2.4 GHz Intel Xeon processors. Each of Cori computing node has 32 2.3 GHz Intel Haswell processors. In our work, we mainly used single node CPU backend of NEON. The hyper-parameter optimization was performed on a single node on Cori with tasks fully parallel on 32 cores.

4. Data

In this study, we use both climate simulations and reanalysis products. The reanalysis products are produced by assimilating observations into a climate model. The spatial scale of both climate model simulation and reanalysis products covers the entire global. A summary of the data source and its temporal and spatial resolution is listed in Table 2. Ground truth labeling of various events is obtained via multivariate threshold based criteria implemented in TECA

Table 1: Deep CNN architecture and layer parameters. The convolutional layer parameters are denoted as <filter size>-<number of feature maps> (e.g. 5x5-8). The pooling layer parameters are denoted as <pooling window> (e.g. 2x2). The fully connected layer parameter are denoted as <number of units> (e.g. 2).

	Conv1	Pooling	Conv2	Pooling	Fully	Fully
Tropical Cyclone	5x5-8	2x2	5x5-16	2x2	50	2
Weather Fronts	5x5-8	2x2	5x5-16	2x2	50	2
Atmospheric River	12x12-8	3x3	12x12-16	2x2	200	2

Table 2: Data Sources

Climate Dataset	Time Frame	Temporal Resolution	Spatial Resolution (lat x lon degree)
CAM5.1 historical run	1979-2005	3 hourly	0.23x0.31
ERA-Interim reanalysis	1979-2011	3 hourly	0.25x0.25
20 century reanalysis	1908-1948	Daily	1x1
NCEP-NCAR reanalysis	1949-2009	Daily	1x1

Table 3: Dimension of image, diagnostic variables (channels) and labeled dataset size for extreme events considered in this study (PSL: sea surface pressure, U: zonal wind, V: meridional wind, T: temperature, TMQ: vertical integrated water vapor, Pr: precipitation)

Events	Image Dimension	Variables	Total Examples
Tropical Cyclone	32x32	PSL,V-BOT,U-BOT, T-200,T-500,TMQ, V-850,U-850	10,000 +ve 10,000 -ve
Atmospheric River	148 x 224	TMQ, Land Sea Mask	6,500 +ve 6,800 -ve
Weather Front	27 x 60	T-2m, Pr, PSL	5,600 +ve 6,500 -ve

[6], [7], and manual labeling by experts [28], [29]. Training data comprise of image patterns, where several relevant spatial variables are stacked together over a prescribed region that bounds an event. The dimension of the bounding box is based on domain knowledge of events spatial extent in real word. For instance, tropical cyclone radius are typically with in range of 100 kilometers to 500 kilometers, thus bounding box size of 500 kilometers by 500 kilometers is likely to capture most of tropical cyclones. The chosen physical variables are also based on domain expertise. The prescribed bounding box is placed over an event. Relevant variables are extracted within the bounding box from the climate model simulations or reanalysis products and stacked together. To facilitate model training, bounding box location is adjusted slightly such that all of events are located approximately at the center. Image patches are cropped and centered correspondingly. Because of the spatial dimension of climate events vary quite a lot and the spatial resolution of source data is non-uniform, final training images prepared differ in their size among the three types of events. The class labels of images are "containing events" and "not containing events", in other words, we formulate the problem as binary classification task. A summary of the attributes of training images is listed in Table 3.

5. Results and Discussion

Table 4 summarizes the performance of our deep CNN architecture on classifying tropical cyclones, atmospheric rivers and weather fronts. We obtained fairly high accuracy (89%-99%) on extreme event classification. In addition, the systems do not suffer from over-fitting. We believe this is mostly because of the shallow and small size of the architecture (4 learnable layers) and the weight decay regularization. Deeper and larger architecture would be inappropriate for this study due to the limited amount of training data. Fairly good train and test classification results also suggest that the deep CNNs we developed are able to efficiently learn representations of climate pattern from labeled data and make predictions based on feature learned. Traditional threshold based detection method requires human expert carefully examine the extreme event and its environment, thus come up with thresholds for defining the events. In contrast, as shown in this study, deep CNNs are able to

learn climate pattern just from the labeled data, thus avoiding subjective thresholds.

Table 4: Overall Classification Accuracy

Event Type	Train	Test	Train time
Tropical Cyclone	99%	99%	$\approx$ 30 min
Atmospheric River	90.5%	90%	6-7 hour
Weather Front	88.7%	89.4%	$\approx$ 30 min

5.1 Classification Results for Tropical Cyclones

Tropical cyclones are rapid rotating weather systems that are characterized by low pressure center with strong wind circulating the center and warm temperature core in upper troposphere. Figure 1 shows examples of tropical cyclones simulated in climate models, that are correctly classified by deep CNN (warm core structure is not shown in this figure). Tropical cyclone features are rather well defined, as can be seen from the distinct low pressure center and spiral flow of wind vectors around the center. These clear and distinct characteristics make tropical cyclone pattern relatively easy to learn and represent within CNN. Our deep CNNs achieved nearly perfect (99%) classification accuracy.

Figure 2 shows examples of tropical cyclones that are mis-classified. After carefully examining these events, we believe they are weak systems (e.g. tropical depression), whose low pressure center and spiral structure of wind have not fully developed. The pressure distribution shows a large low pressure area without a clear minimum. Therefore, our deep CNN does not label them as tropical cyclones.

Table 5: Confusion matrix for tropical cyclone classification

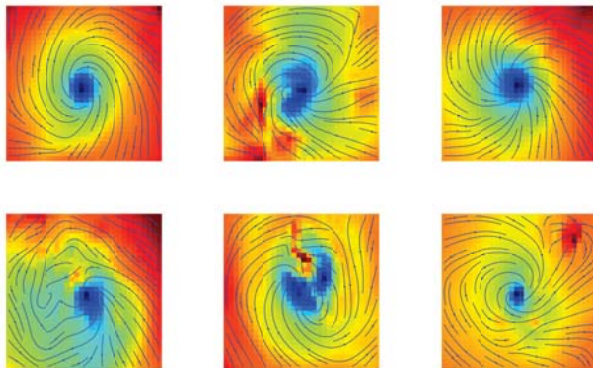|	Label TC	Label Non_TC
Predict TC	0.989	0.003
Predict Non_TC	0.011	0.997

Fig. 1: Sample images of tropical cyclones correctly classified (true positive) by our deep CNN model. Figure shows sea level pressure (color map) and near surface wind distribution (vector solid line).
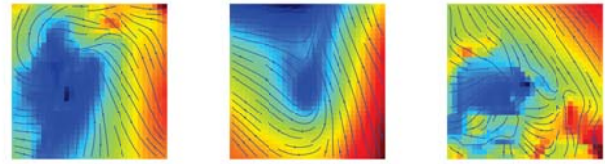

Fig. 2: Sample images of tropical cyclones mis-classified (false negative) by our deep CNN model. Figure shows sea level pressure (color map) and near surface wind distribution (vector solid line).

5.2 Classification Results for Atmospheric Rivers

In contrast to tropical cyclones, atmospheric rivers are distinctively different events. They are narrow corridors of concentrated moisture in atmosphere. They usually originate in tropical oceans and move pole-ward. Figure 3 shows examples of correctly classified land falling atmospheric rivers that occur on the western Pacific Ocean and north Atlantic Ocean. The characteristics of narrow water vapor corridor is well defined and clearly observable in these images.

Figure 4 are examples of mis-classified atmospheric rivers. Upon further investigation, we believe there are two main factors leading to mis-classification. Firstly, presence of weak atmospheric river systems. For instance, the left column of Figure 4 shows comparatively weak atmospheric rivers. The water vapor distribution clearly show a band of concentrated moisture cross mid-latitude ocean, but the signal is much weaker comparing to Figure 3. Thus, deep CNN does not predict them correctly. Secondly, the presence of other climate event may also affect deep CNN representation of atmospheric rivers. In reality, the location and shape of atmospheric river are affected by jet streams and extra-tropical cyclones. For example, Figure 4 right column shows rotating systems (likely extra-tropical cyclone) adjacent to the atmospheric river. This phenomenon presents challenge for deep CNN on representing atmospheric river.

Table 6: Confusion matrix for atmospheric river classification

	Label AR	Label Non_AR
Predict AR	0.93	0.107
Predict Non_AR	0.07	0.893

5.3 Classification Results for Weather Fronts

Among the three types of climate events we are looking at, weather fronts have the most complex spatial pattern. Weather fronts typically form at the interface of warm air and cold air, and usually associated with heavy precipitation due moisture condensation of warm air up-lifting. In satellite
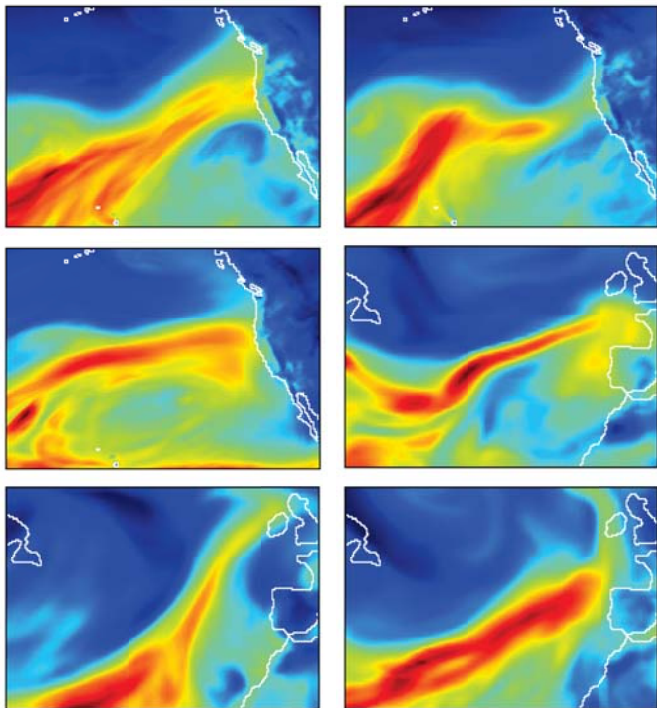
Fig. 3: Sample images of atmospheric rivers correctly classified (true positive) by our deep CNN model. Figure shows total column water vapor (color map) and land sea boundary (solid line).
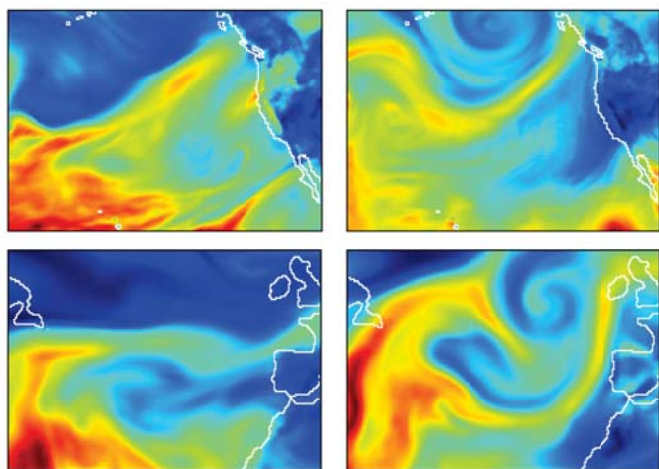

Fig. 4: Sample images of atmospheric rivers mis-classified (false negative) by our deep CNN model. Figure shows total column water vapor (color map) and land sea boundary (solid line).

images,a weather front is observable as a strip of clouds, but it is hardly visible on two dimensional fields such as temperature and pressure. In middle latitude (e.g. most U.S.), a portion of weather front are associated with extra-tropical

cyclones. Figure 5 shows examples of correctly classified weather front by our deep CNN system. Visually, the narrow long regions of high precipitation line up approximately parallel to the temperature contour. This is a clear characteristics and comparatively easy for deep CNNs to learn.

Because patterns of weather fronts is rather complex and hardly show up in two dimensional fields, we decided to further investigate it in later work.

Table 7: Confusion matrix for weather front classification

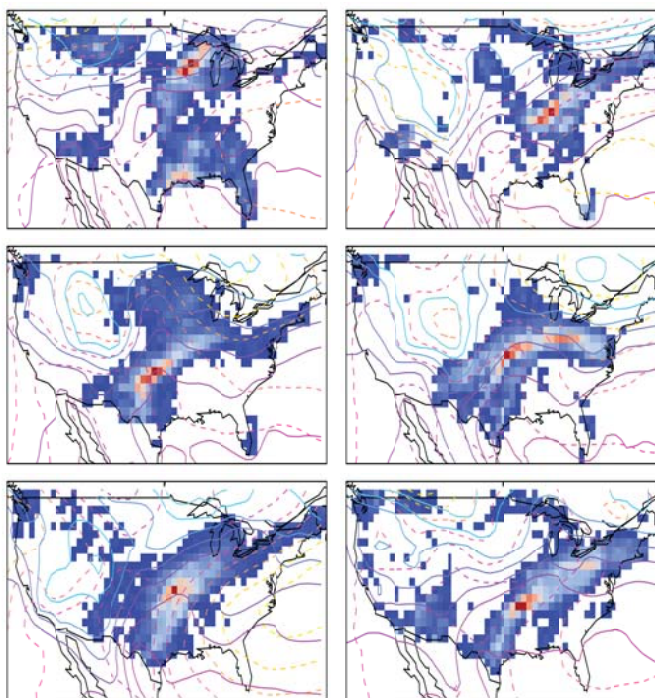|	Label WF	Label Non_WF
Predict WF	0.876	0.18
Predict Non_WF	0.124	0.82

Fig. 5: Sample images of weather front correctly classified by our deep CNN model. Figure shows precipitation with daily precipitation less than 5 millimeters filtered out (color map), near surface air temperature (solid contour line) and sea level pressure (dashed contour line)

6. Future Work

In the present study, we trained deep CNNs separately for classifying tropical cyclones, atmospheric rivers and weather fronts. Ideally, we would like to train a **single** neural network for classifying all three types of events. Unlike object recognition in natural images, climate patterns detection have unique challenges. Firstly, climate events happen at vastly different spatial scales. For example, a tropical cyclone typically extends over less than 500 kilometers in radius, while an atmospheric river can be several

thousand kilometers long. Secondly, different climate events are characterized by different sets of physical variables. For example, atmospheric rivers correlate strongly with the vertical integration of water vapor, while tropical cyclones has a more complex multi-variable pattern involving sea level pressure, near surface wind and upper troposphere temperature. Future work will need to develop generative CNN architectures that are capable of discriminating between different variables based on the event type and capable of handling events at various spatial scale. Note that we have primarily addressed **detection** of extreme weather patterns, but not their **localization**. We will work on architectures for spatially localizing weather pattern in the future.

Several researchers have pointed out that deeper and larger CNNs perform better for classification and detection tasks[9], [10] compared to shallow networks. However, deep networks require huge amount of data to be effectively trained, and to prevent model over fitting. Datasets, such as ImageNet, provide millions of labeled images for training and testing deep and large CNNs. In contrast, we can only obtain a small amount of labeled training data, hence we are constrained on the class of deep CNNs that we can explore without suffering from over-fitting. This limitation also points us to the need for developing unsupervised approaches for climate pattern detection. We believe that this will be critical for the majority of scientific disciplines that typically lack labeled data.

7. Conclusion

In this study, we explored deep learning as a methodology for detecting extreme weather patterns in climate data. We developed deep CNN architecture for classifying tropical cyclones, atmospheric rivers and weather fronts. The system achieves fairly high classification accuracy, range from 89% to 99%. To the best of our knowledge, this is the first time that deep CNN has been applied to tackle climate pattern recognition problems. This successful application could be a precursor for tackling a broad class of pattern detection problem in climate science. Deep neural network learns high-level representations from data directly, therefore potentially avoiding traditional subjective thresholding based criteria of climate variables for event detection. Results from this study will be used for quantifying climate extreme events trend in current day and future climate scenarios, as well as investigating the changes in dynamics and thermodynamics of extreme events in global warming contend. This information is critical for climate change adaptation, hazard risk prediction and climate change policy making.

8. Acknowledgments

References

[1] D. S. Nolan and M. G. McGauley, "Tropical cyclogenesis in wind shear: Climatological relationships and physical processes," in *Cyclones: Formation, Triggers, and Control*, 2012, pp. 1–36.

[2] F. Vitart, J. Anderson, and W. Stern, "Simulation of interannual variability of tropical storm frequency in an ensemble of gcm integrations," *Journal of Climate*, vol. 10, no. 4, pp. 745–760, 1997.

[3] ——, "Impact of large-scale circulation on tropical storm frequency, intensity, and location, simulated by an ensemble of gcm integrations," *Journal of Climate*, vol. 12, no. 11, pp. 3237–3254, 1999.

[4] K. Walsh and I. G. Watterson, "Tropical cyclone-like vortices in a limited area model: comparison with observed climatology," *Journal of Climate*, vol. 10, no. 9, pp. 2240–2259, 1997.

[5] K. Walsh, M. Fiorino, C. Landsea, and K. McInnes, "Objectively determined resolution-dependent threshold criteria for the detection of tropical cyclones in climate models and reanalyses," *Journal of Climate*, vol. 20, no. 10, pp. 2307–2314, 2007.

[6] Prabhat, O. Rübel, S. Byna, K. Wu, F. Li, M. Wehner, W. Bethel, *et al.*, "Teca: A parallel toolkit for extreme climate analysis," in *Third Worskhop on Data Mining in Earth System Science (DMESS) at the International Conference on Computational Science (ICCS)*, 2012.

[7] Prabhat, S. Byna, V. Vishwanath, E. Dart, M. Wehner, W. D. Collins, *et al.*, "Teca: Petascale pattern recognition for climate science," in *Computer Analysis of Images and Patterns*. Springer, 2015, pp. 426–436.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems (NIPS)*, 2012, pp. 1097–1105.

[9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Internaltional Conference on Learning Representation (ICLR)*, 2015.

[10] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9.

[11] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *Signal Processing Magazine, IEEE*, vol. 29, no. 6, pp. 82–97, 2012.

[12] G. E. Dahl, D. Yu, L. Deng, and A. Acero, "Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 20, no. 1, pp. 30–42, 2012.

[13] A. Graves, A.-r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 6645–6649.

[14] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Advances in neural information processing systems*, 2014, pp. 3104–3112.

[15] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," in *International Conference on Learning Representations (ICLR)*, 2014.

[16] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 580–587.

[17] J. R. Uijlings, K. E. van de Sande, T. Gevers, and A. W. Smeulders, "Selective search for object recognition," *International Journal of Computer Vision*, vol. 104, no. 2, pp. 154–171, 2013.

[18] R. Chattopadhyay, A. Vintzileos, and C. Zhang, "A description of the madden–julian oscillation based on a self-organizing map," *Journal of Climate*, vol. 26, no. 5, pp. 1716–1732, 2013.

[19] J. Gorricha, V. Lobo, and A. C. Costa, "A framework for exploratory analysis of extreme weather events using geostatistical procedures and 3d self-organizing maps," *International Journal on Advances in Intelligent Systems*, vol. 6, no. 1, 2013.

[20] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, "Convolutional lstm network: A machine learning approach for precipitation nowcasting," in *Advances in Neural Information Processing Systems: Twenty-Ninth Annual Conference on Neural Information Processing Systems (NIPS)*, 2015.

[21] G. Iglesias, D. C. Kale, and Y. Liu, "An examination of deep learning for extreme climate pattern analysis," in *The 5th International Workshop on Climate Informatics*, 2015.

[22] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," in *International Conference on Learning Representation (ICLR)*, 2015.

[23] H. Larochelle, Y. Bengio, J. Louradour, and P. Lamblin, "Exploring strategies for training deep neural networks," *The Journal of Machine Learning Research*, vol. 10, pp. 1–40, 2009.

[24] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *International conference on artificial intelligence and statistics*, 2010, pp. 249–256.

[25] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," in *Advances in neural information processing systems*, 2012, pp. 2951–2959.

[26] E. Brochu, V. M. Cora, and N. De Freitas, "A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning," *arXiv preprint arXiv:1012.2599*, 2010.

[27] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th International Conference on Machine Learning (ICML)*, 2010, pp. 807–814.

[28] K. E. Kunkel, D. R. Easterling, D. A. Kristovich, B. Gleason, L. Stoecker, and R. Smith, "Meteorological causes of the secular variations in observed extreme precipitation events for the conterminous united states," *Journal of Hydrometeorology*, vol. 13, no. 3, pp. 1131–1141, 2012.

[29] D. A. Lavers, G. Villarini, R. P. Allan, E. F. Wood, and A. J. Wade, "The detection of atmospheric rivers in atmospheric reanalyses and their links to british winter floods and the large-scale climatic circulation," *Journal of Geophysical Research: Atmospheres*, vol. 117, no. D20, 2012.