Spatio-temporal segmentation and tracking of weather patterns with light-weight Neural Networks

Lukas Kapp-Schwoerer*
ETH Zurich
lukaska@ethz.ch

Andre Graubner*
ETH Zurich
andregr@ethz.ch

Sol Kim
University of California, Berkeley
solkim@lbl.gov

Karthik Kashinath
Lawrence Berkeley National Laboratory
kkashinath@lbl.gov

Abstract

The reliable detection and tracking of weather patterns is a necessary first step towards characterizing extreme weather events in a warming world. Recent work [Prabhat *et al.*(2020)] has shown that weather pattern recognition by deep neural networks can work remarkably better than feature engineering, such as hand-crafted heuristics, used traditionally in climate science. As an extension of this work, we perform Deep Learning - based semantic segmentation of atmospheric rivers and tropical cyclones on the expert-annotated ClimateNet data set, and track individual events using a spatio-temporal overlapping approach. Our approach is fast and scalable to more data modalities and event types, motivating expansion of the ClimateNet dataset and development of novel deep learning architectures. Furthermore, we show that the spatio-temporal tracking capability enables investigating a host of important climate science research questions pertaining to the behavior of extreme weather events in a warming world.

1 Introduction

Climate change is arguably one of the most pressing challenges facing humanity in the 21st century. Identifying weather patterns that frequently lead to extreme weather events is a crucial first step in understanding how they may vary under different climate change scenarios. Until recently climate scientists have largely relied on custom *heuristics* for the identification of these events [Hodges(1995), Neu *et al.*(2013), Prabhat *et al.*(2015), Ullrich and Zarzycki(2017), Shields *et al.*(2018), Guan and Waliser(2019)]. However, recent work showed conclusively that deep learning trained on the expert-annotated ClimateNet data set can work as well and extremely efficiently [Prabhat *et al.*(2020)].

Two limitations of the above-mentioned work are: (i) the DL segmentation model used, DeepLabv3+, is computationally expensive and hard to train; (ii) the method cannot spatio-temporally track individual unique events. In this paper we address these limitations and make the following key contributions: (i) we implement and train a light-weight neural network for detecting atmospheric rivers (ARs) and tropical cyclones (TCs) in simulated climate data; and (ii) we propose a spatio-temporal tracking algorithm to track individual events and show a host of climate-scientific downstream analyses that can be performed with this capability. We open-source our code on GitHub[1] and provide an easy-to-use tool to apply our tracking algorithm, enabling simple adoption.

*equal contribution
[1]https://github.com/andregraubner/ClimateNet

AI for Earth Sciences Workshop at NeurIPS 2020.

2 Methods

2.1 Dataset

The ClimateNet dataset consists of 459 feature-label pairs (x_i, y_i), where x_i is a 16-channel, $(1152, 768)$ pixel image that is the output of a historical 25-km CAM5.1 climate model output [Wehner *et al.*(2014)] and y_i is the corresponding pixel-wise labeling (background, atmospheric river, tropical cyclone). Details about the data set, including the expert-annotated labels, are in [Prabhat *et al.*(2020)]. We choose 4 channels that are most relevant to detecting TCs and ARs: TMQ (Total vertically integrated water vapor), U850 (Zonal wind at 850 mbar), V850 (Meridional wind at 850 mbar) and PSL (pressure at sea level).

2.2 Model

We use a segmentation architecture based on CGNet [Wu *et al.*(2019)]. Besides several standard convolutional layers, the core building block of the network is the CGBlock, which first integrates local as well as surrounding context by concatenating the output of normal and dilated convolutions, then refines the representation with global context through average pooling. In order to capture the topology of the input domain (connecting opposing sides in a rectangular mapping of the earth) we perform circular padding. At 494232 parameters, this architecture is light-weight, allowing for a very low memory footprint and high training and inference speeds, making it a viable solution for detecting and tracking extreme weather events on large climate data sets.

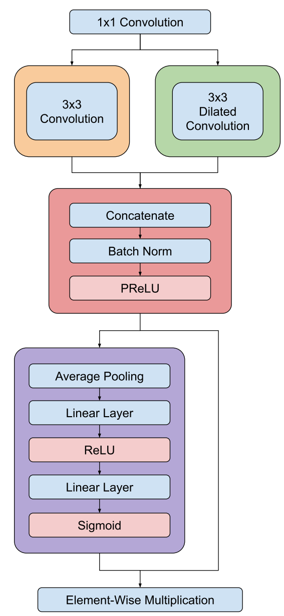

Figure 1: A CGBlock integrates local features and surrounding features into a joint feature. It then refines the representation using global context.

2.3 Training

As common in semantic segmentation, our problem suffers from strong data-imbalance, with the overwhelming majority of pixels belonging to the background class. Instead of using the commonly used weighted cross-entropy loss for addressing this issue, we decide to optimize a loss more directly related to the Jaccard Index (Intersection over Union), a standard metric for evaluating segmentation results. For one sample, the Jaccard Loss [Rahman and Wang(2016)] is defined as:

$$\mathcal{L}(P, Y) = 1 - \frac{\sum_{ijk} (p_{ijk} \cdot y_{ijk})}{\sum_{ijk} (p_{ijk} + y_{ijk} - p_{ijk} \cdot y_{ijk}) + \epsilon}$$

Here, P is the output of the network given x, softmaxed along the class axis such that the scores for each class are positive and sum up to 1 per pixel. Y is the one-hot encoded corresponding labeling, such that y_{ijk} is 1 if and only if pixel (i, j) has been labeled as class k for the corresponding sample. We add a small $\epsilon = 1e^{-7}$ to the denominator for numerical stability. We minimize this loss for 15 epochs over our training set using the Adam optimizer [Kingma and Ba(2014)]. The model reaches 59.3% training IoU and 56.1% test IoU on the held out test set (13% of the entire data set).

2.4 Spatio-temporal event tracking

Many important scientific questions about extreme weather events require an architecture that can track events during time, such as counting number of events per year or season, identifying genesis and termination locations, calculating lifetimes, speed, distance traveled, etc. Since our network does not perform spatio-temporal segmentation and only makes its predictions per-sample (timestep), we apply a post-processing scheme to the network predictions to enable the tracking of individual events over time. For each of the event classes (AR or TC) we first identify connected components in the labelling output per sample. We then iterate through the samples in chronological order and check if a connected component overlaps with a connected component of the previous timestep. If it does, we assign it the ID of the connected component from the previous timestep, otherwise we assign it a new

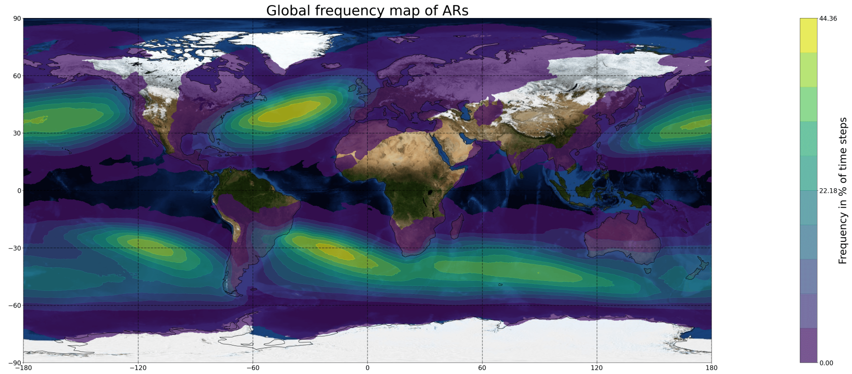

Figure 2: Global frequency map of ARs

unique ID. Additionally, in order reduce noise we omit connected components that are smaller than 250 pixels and events that only appear briefly (less than 12 hours).

3 Results and Discussion

We provide a highlight of our detection and tracking method using one year of CAM5 output at https://youtu.be/_tNuDPsr1Uc. All following analysis is based on five years of output. Due to space constraints, we limit our discussion to results pertaining to ARs. A similar discussion for TCs is reserved for future work.

3.1 Atmospheric rivers

For ARs, although there exist over a dozen detection and tracking heuristics which are extensively compared in [Shields *et al.*(2018)], we compare our results with a state-of-the-art detection and tracking heuristic - Guan and Waliser's tARget v3 (GW hereinafter) [Guan and Waliser(2019)]. This heuristic uses a location and season specific IVT threshold (85th percentile). A notable feature of this heuristic is that compared to other existing heuristics, it produces higher frequencies of ARs over land and polar regions due to its use of a relative threshold [Rutz *et al.*(2019)].

3.1.1 Global frequency of events

The global frequency of atmospheric rivers using our method are shown in Fig. 2. The structure of the frequencies shown matches GW [Guan and Waliser(2015)] with high values over the expected regions of AR activity (e.g. over the Pacific and Atlantic storm track). However, the frequencies tend to be approximately x3 larger in our approach. While this discrepancy seems quite large, in [Shields *et al.*(2018)], it was found that human labelers largely exceeded the frequency of AR events compared to a range of heuristics. Qualitatively, the ARs detected with our method also tend to be of larger sizes compared to GW, which inflates the frequencies. Further, our method produces very low frequencies over land and in higher latitudes. As GW uses relative thresholds, the IVT values can be quite low in higher latitudes but the heuristic would still identify ARs. The DL method, which is trained on human labels, relies on labelers to use absolute values of fields such as IVT and IWV.

3.1.2 Genesis frequency of events

The genesis frequency is shown in Fig. 3(a). Areas of higher genesis activity match GW [Guan and Waliser(2019)] but there are some significant differences in location and extent. The frequencies also tend to be lower than GW. To note, their frequency plots for genesis and termination use any ARs events that exist for 6 hours or longer (two or more timesteps since any less would make genesis and termination ambiguous), while the frequencies plotted here are events that exist 12 or more hours. The lower genesis frequencies in our method again could also be a result of the high land and polar bias in GW. For example, our method produces a region of high AR genesis activity near the east coast of the U.S. GW also produce high frequencies here but also extends over much of CONUS; while our method produces zero frequencies over the western portion of the U.S.

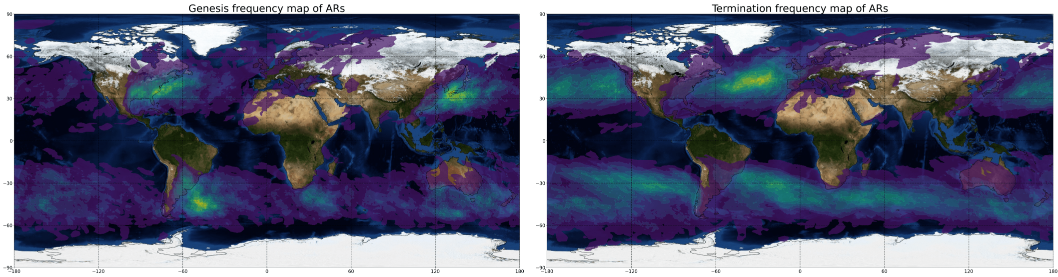

Figure 3: AR genesis (left, max frequency in events per month = 0.73) and AR termination (right, max frequency in events per month = 1.1).

3.1.3 Termination frequency of events

The termination frequency is shown in Fig. 3(b). There are significant differences in termination between our method and GW. GW show the highest termination frequencies around 60°s in either hemisphere. With our method, ARs tend to terminate in the midlatitudes. Again, due to the location and season specific thresholds used by GW, their method will identify ARs in the polar regions at far higher frequencies vs human labelers.

3.2 Lifecycle Characteristics

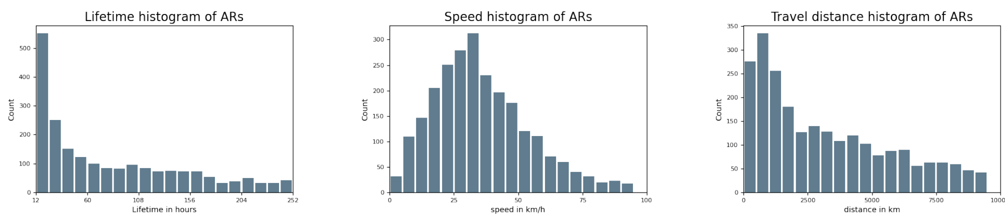

Figure 4: Histograms of AR lifecycle characteristics (left to right: lifetime duration, speed, and distance for columns).

We next examine basic lifecycle characteristics of ARs tracked with our method. The lifecycle duration, speed, and travel distance are shown in Fig. 4. AR lifetime duration peaks at 12 hours (the minimum duration requirement in post-processing) and generally decreases as is expected with ARs ([Guan and Waliser(2019)],[Zhou and Kim(2019)]). The distribution of AR speeds matches well with GW with most ARs having speeds between 20-60 km/hr with a peak at 25-35 km/hr. Travel distance distribution also generally match GW with a few differences. For our tracking, AR travel distance peaks at 500-1000km while the GW tracking has distance peak at 0-500km. Their tracking also monotonically decreases whereas our shows slight variation.

4 Conclusions and Future Work

The light-weight neural network and spatio-temporal event tracking scheme proposed in this paper address the main limitations of state-of-the-art Deep Learning models trained on the curated expert-labeled climate data – ClimateNet. The new and improved methods proposed here enable ultrafast spatio-temporal segmentation and tracking on large climate data sets. Further, we show many useful downstream scientific analyses now possible using the event tracking scheme.

Opportunities for future work include: (i) applications to reanalyses products or observational data using transfer learning; (ii) extensions to other event types as the approach is inherently event-type agnostic; (iii) probabilistic segmentation and uncertainty quantification; (iv) end-to-end learning based approaches to the tracking of events that do not require post-processing the spatial segmentation output; and (v) rapidly exploring hypotheses related to dynamical mechanisms of extreme events, including interactions between different types of events.

References

[Guan and Waliser(2015)] Guan, B. and Waliser, D. E. (2015). Detection of atmospheric rivers: Evaluation and application of an algorithm for global studies. *Journal of Geophysical Research: Atmospheres*, **120**(24), 12514–12535.

[Guan and Waliser(2019)] Guan, B. and Waliser, D. E. (2019). Tracking atmospheric rivers globally: spatial distributions and temporal evolution of life cycle characteristics. *Journal of Geophysical Research: Atmospheres*, **124**(23), 12523–12552.

[Hodges(1995)] Hodges, K. I. (1995). Feature tracking on the unit sphere. *Monthly Weather Review*, **123**(12), 3458–3465.

[Kingma and Ba(2014)] Kingma, D. P. and Ba, J. (2014). Adam: A Method for Stochastic Optimization. *ArXiv e-prints*.

[Neu *et al.*(2013)] Neu, U., Akperov, M. G., Bellenbaum, N., Benestad, R., Blender, R., Caballero, R., Cocozza, A., Dacre, H. F., Feng, Y., Fraedrich, K., Grieger, J., Gulev, S., Hanley, J., Hewson, T., Inatsu, M., Keay, K., Kew, S. F., Kindem, I., Leckebusch, G. C., Liberato, M. L. R., Lionello, P., Mokhov, I. I., Pinto, J. G., Raible, C. C., Reale, M., Rudeva, I., Schuster, M., Simmonds, I., Sinclair, M., Sprenger, M., Tilinina, N. D., Trigo, I. F., Ulbrich, S., Ulbrich, U., Wang, X. L., and Wernli, H. (2013). Imilast: A community effort to intercompare extratropical cyclone detection and tracking algorithms. *Bulletin of the American Meteorological Society*, **94**(4), 529–547.

[Prabhat *et al.*(2020)] Prabhat, Kashinath, K., Mudigonda, M., Kim, S., Kapp-Schwoerer, L., Graubner, A., Karaismailoglu, E., von Kleist, L., Kurth, T., Greiner, A., Yang, K., Lewis, C., Chen, J., Lou, A., Chandran, S., Toms, B., Chapman, W., Dagon, K., Shields, C. A., O'Brien, T., Wehner, M., and Collins, W. (2020). Climatenet: an expert-labelled open dataset and deep learning architecture for enabling high-precision analyses of extreme weather. *Geoscientific Model Development Discussions*, **2020**, 1–28.

[Prabhat *et al.*(2015)] Prabhat, Byna, S., Vishwanath, V., Dart, E., Wehner, M., Collins, W. D., *et al.* (2015). Teca: Petascale pattern recognition for climate science. In *International Conference on Computer Analysis of Images and Patterns*, pages 426–436. Springer.

[Rahman and Wang(2016)] Rahman, M. A. and Wang, Y. (2016). Optimizing intersection-over-union in deep neural networks for image segmentation. In *ISVC*.

[Rutz *et al.*(2019)] Rutz, J. J., Shields, C. A., Lora, J. M., Payne, A. E., Guan, B., Ullrich, P., O'Brien, T., Leung, L. R., Ralph, F. M., Wehner, M., *et al.* (2019). The atmospheric river tracking method intercomparison project (artmip): quantifying uncertainties in atmospheric river climatology. *Journal of Geophysical Research: Atmospheres*, **124**(24), 13777–13802.

[Shields *et al.*(2018)] Shields, C. A., Rutz, J. J., Leung, L.-Y., Ralph, F. M., Wehner, M., Kawzenuk, B., Lora, J. M., McClenny, E., Osborne, T., Payne, A. E., Ullrich, P., Gershunov, A., Goldenson, N., Guan, B., Qian, Y., Ramos, A. M., Sarangi, C., Sellars, S., Gorodetskaya, I., Kashinath, K., Kurlin, V., Mahoney, K., Muszynski, G., Pierce, R., Subramanian, A. C., Tome, R., Waliser, D., Walton, D., Wick, G., Wilson, A., Lavers, D., Prabhat, Collow, A., Krishnan, H., Magnusdottir, G., and Nguyen, P. (2018). Atmospheric river tracking method intercomparison project (artmip): Project goals and experimental design. *Geoscientific Model Development Discussions*, **2018**, 1–55.

[Ullrich and Zarzycki(2017)] Ullrich, P. A. and Zarzycki, C. M. (2017). Tempestextremes: a framework for scale-insensitive pointwise feature tracking on unstructured grids. *Geoscientific Model Development*, **10**(3), 1069.

[Wehner *et al.*(2014)] Wehner, M. F., Reed, K. A., Li, F., Bacmeister, J., Chen, C.-T., Paciorek, C., Gleckler, P. J., Sperber, K. R., Collins, W. D., Gettelman, A., *et al.* (2014). The effect of horizontal resolution on simulation quality in the community atmospheric model, cam5. 1. *Journal of Advances in Modeling Earth Systems*, **6**(4), 980–997.

[Wu *et al.*(2019)] Wu, T., Tang, S., Zhang, R., and Zhang, Y. (2019). CGNet: A Light-weight Context Guided Network for Semantic Segmentation.

[Zhou and Kim(2019)] Zhou, Y. and Kim, H. (2019). Impact of distinct origin locations on the life cycles of landfalling atmospheric rivers over the us west coast. *Journal of Geophysical Research: Atmospheres*, **124**(22), 11897–11909.