Atmospheric River Identification in Climate Datasets using Deep Learning

Deveshi Buch
deveshi@stanford.edu

Abstract

Atmospheric Rivers (ARs) are long, narrow plumes of water vapor occurring globally and produce beneficial water supply but also destructive storms in mid-latitudinal regions around the world. This study is an exploratory effort in the nascent field of atmospheric river detection using deep learning. Building on the recent ClimateNet [7] model and expert-labeled dataset, this study presents the ClimateNet-Regional dataset and explores the capabilities of the CGNet context-guided segmentation model for regional perspectives. Regional-trained performance reaches a mean IOU of 0.5732, slightly exceeding the globally-trained accuracy, on regional datasets. Moving forward, regionally-trained deep learning models have the potential to improve the efficiency of these models and their localized accuracy, with significant implications for extreme weather detection in climate datasets.

1 Introduction

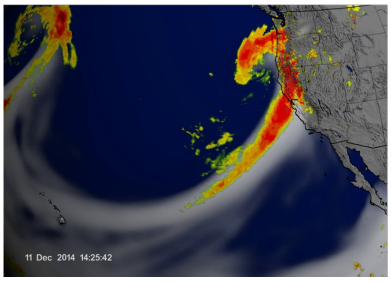

Figure 1: Visualization of Atmospheric River event in North-Eastern Pacific impacting U.S. West Coast on 11 Dec. 2014 from NASA.gov.

Atmospheric Rivers (ARs) are among a diverse set of extreme weather phenomena, including tropical cyclones (TCs) and tornadoes, that impact regions across the globe. ARs are long, filamentary corridors of moisture and winds in the troposphere that can carry more than double the flow of the Amazon River. AR events are often responsible for significant precipitation across the world and are uniquely capable of both beneficial and destructive impacts depending on intensity—replenishing much-needed water resources during drought, but also producing catastrophic floods that can result in $1 billion in damage [1]. In particular, California's water supply depends greatly upon ARs, which provide up to 50% of annual precipitation [2]. Although AR research has emerged relatively recently, within the last couple of decades, identifying and analyzing AR events has been crucial to water resource management, especially as larger and more intense ARs are predicted as our climate changes [3]. Because AR identification can vary depending on algorithm or human skill [4] (also briefly discussed in [5]), deep learning has great potential to efficiently identify ARs in large climate datasets.

More recently, many studies have been conducted to analyze AR regional impacts (such as North-Eastern Pacific / U.S. West Coast in [5], South-Eastern Pacific / South American West Coast in [11]). This study explores applying a deep convolutional neural network (CNN) to AR identification and

CS230: Deep Learning, Winter 2021, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

segmentation on both global and regional scales. Inputs to the context-guided segmentation model, derived from CGNet/ClimateNet, are "images" with 4 channels—vertically-integrated precipitable water (TMQ), zonal (U850) and meridional (V850) winds at 850mb, and sea-level pressure (PSL)—along with labeled masks—0, 1, 2 for Background (BG), TC, and AR classes respectively. Outputs are predicted segmentation masks with class labels and event counts.

2 Related Work

Although deep learning applied to AR identification is in its early stages, there has been some progress in recent years in experimenting with applications of deep learning to this space. AR identification is often performed manually on small scales and by heuristics-based algorithms on larger scales. The Atmospheric River Tracking Method Intercomparison Project (ARTMIP) has been launched to illustrate the vast variation in the different models and strategies for identifying these features in large datasets and reports that different heuristics-based AR identification algorithms can differ by up to an order of magnitude [4]. Thus, deep learning has gained momentum in the atmospheric science community as supervised learning models can combine expert-level understanding of ARs and other extreme weather events with deep networks' abilities to learn and predict features of them.

One study [8] created a deep CNN based on AlexNet to classify extreme weather events. This model contained just 4 learnable layers—2 convolutional, each followed by max pooling, and 2 fully-connected—and cited the lack of large labeled datasets. The model classified input images of ARs, TCs, and weather fronts with a reasonably high accuracy (89-99%) but a notable training time for AR classification of 6-7 hours for ~13,000 148×224 images. Climate data can be particularly complex, so for more computationally-intensive tasks such as segmentation (this project, based on [7]), it is important to keep this smaller-scale regional perspective in mind, even with a limited number of training examples.

The lack of expert-labeled datasets, as well as the observation that different weather extremes can have widely different characteristics, prompted the recent development of ClimateNet [7] based on DeepLabv3+ [9], which has depth-wise separable convolutional layers, a pooling module with 2-D pooling, Conv, batch norm, and ReLU layers, and an encoder to segment objects along with a decoder to create the higher-resolution segmentation boundaries. [7] also compiled an expert-labeled global dataset with a few hundred labeled examples—much fewer than may be in a typical deep learning or computer vision dataset, but regardless posing a new opportunity for more precise AR localization via segmentation. While the authors reported on a DeepLabv3+ implementation of ClimateNet in [7], a version using the CGNet architecture [10] in PyTorch was released to the scientific community for further exploration. One of the primary differences between CGNet and DeepLabv3+, as suggested in [10], is the "lightweight" characteristic of the former, and CGNet has relatively few layers (51) and parameters compared to other segmentation models described in [9] and [10]. With a limited memory and computational resources, this study leverages the CGNet/ClimateNet implementation in order to study the application to region-localized AR events. (See Section 5 for additional CGNet/ClimateNet details.) To build on ClimateNet, this study aims to create a new regional dataset based on the ClimateNet set, understand the potential of a context-based segmentation model for AR identification, explore the extension of these applications to both global and regional scales.

3 Dataset

While climate data has become increasingly available in recent years, challenges associated with deep learning for extreme weather include the limited amount of expert-labeled data, the complexity of atmospheric features, and the notably low subject-to-background ratio (ARs makes up a small percentage of the overall image) [7]. Furthermore, a model for AR identification can take several hours to train, compared to less than an hour for other weather events [8]. The ClimateNet dataset was designed to address some of these issues; however, while high-resolution global climate data provide an abundance of information for a model, it also requires considerable storage, memory, and computational time (loading the 398 labeled global images for analysis from a local machine required ~12 hours, and training attempts exceeded the GPU computational budget for this course allocation). By contrast, regional data can be more time- and space-efficient for training and evaluating models, and many deep learning models are generally known to perform well even with lower-resolution input images. Further, regional analyses provide in-depth views of AR impacts, so localizing ARs on

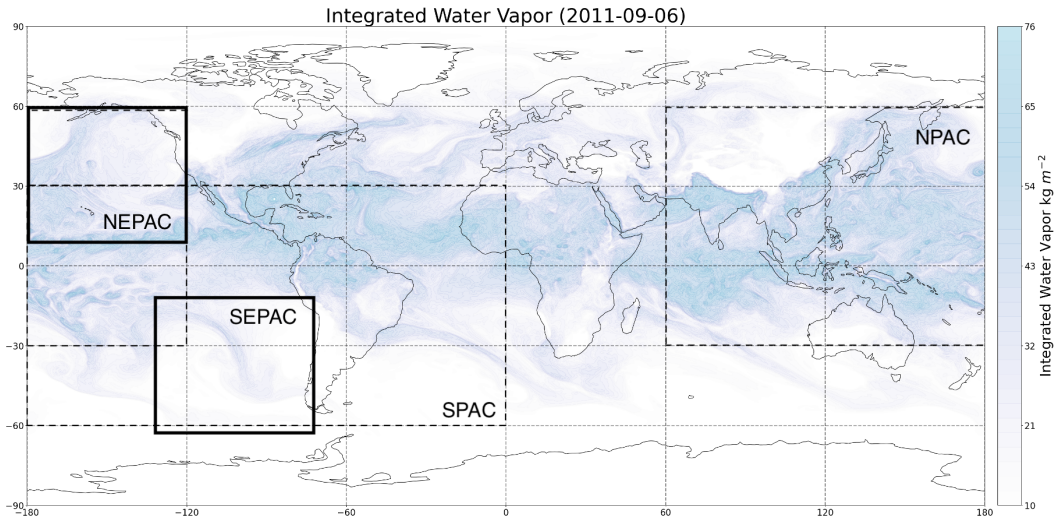

Figure 2: Global water vapor input showing regions of NEPAC, SEPAC (solid), NPAC, SPAC (dotted) for new ClimateNet-Regional dataset. Visualization created based on code from my previous research [5] and ClimateNet analysis visualization code. NEPAC and SEPAC are 5350 $\times$ 4800 km, selected for their AR-prone location in the mid-latitudes and ongoing research interests in these areas. Since ARs are typically $\leq$1000 km wide and $\geq$2000 km long, these regions can be suitable for containing most or the entirety of an AR.

a regional scale can be additionally beneficial as it is of particular interest to the broader scientific community (see Section 1; [4], [11]).

Thus, this work presents ClimateNet-Regional, a regional dataset spanning four global regions (see Fig. 2), based on the released ClimateNet dataset containing 459 labeled examples from the CAM5.1 25-km resolution climate model. The data are formatted as NetCDF files, commonly used in climate science, containing data for four atmospheric channels (described in Section 1), latitude and longitude coordinates, and input labels, and outputs labels and segmentation masks associated with climate events. The North-Eastern Pacific (NEPAC) and South-Eastern Pacific (SEPAC) images are 215×193, and images centered around the North-Pacific (NPAC) and South-Pacific (SPAC) are 384×576. Due to the limited amount of data, only training and validation sets (size 398 and 61 respectively) are used in this study. These data are normalized with ClimateNet's global mean and standard deviation for the four channels when loaded into the model, although how these values were calculated by [7] is a bit unclear. The new ClimateNet-Regional dataset is in the process of being made available.

4 Methods

The methodology in this study consists of three components:

1. Develop ClimateNet-Regional, a dataset derived from global CAM 5.1 output (as in [7]) that focuses on four carefully-curated regional snapshots of relevance to the AR research community (detailed in Section 3; [4], [11]);
2. Apply the globally-trained CGNet/ClimateNet context-guided segmentation model to global, NEPAC, and SEPAC data;
3. Explore possibilities for training model on regional NEPAC data, and fine-tuning and evaluating performance on global, NEPAC, and SEPAC data.

Following the creation of ClimateNet-Regional dataset, this project validates the ClimateNet/CGNet model on provided global datasets, and regional applications are explored thereafter. The CGNet/ClimateNet model, as per [10], considers surrounding and global context in carrying out semantic segmentation of images.

Some notes about this study compared to the methodology in [7]: precipitation rate was used for [7], but sea-level pressure (PSL) was included in the ClimateNet released dataset and is therefore used in this project; the ClimateNet released dataset has a slightly different train-dev split (398-61) than

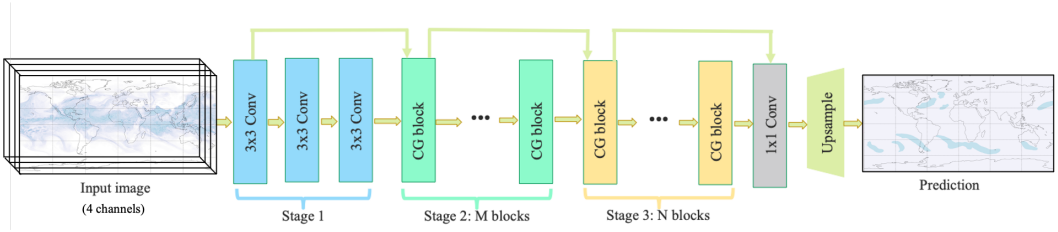

Figure 3: CGNet/ClimateNet consists of 51 layers: 3x3 Conv (convolution followed by Batch Normalization (BN) and Parametric ReLU (PReLU), a type of leaky ReLU) layers in Stage 1, M=3 CG blocks in Stage 2, and N=3 CG blocks in Stage 3. CG blocks contain a local feature extractor, surrounding context extractor, joint feature extractor, and a global context extractor; the local (stanford Conv) and surrounding (dilated/atrous Conv, similar to DeepLabv3+) extractors learn nearby context, feed into the joint extractor via concatenation, BN, and PReLU, and is improve via global context extraction. After Stage 3, the result is passed through a 1x1 Conv layer and upsampled to produce the segmentation mask prediction. Image from [10], modified for this study.

the train-dev-test split in [7] (422-18-19); finally, ClimateNet by design generates masks for both ARs and TCs, and while ARs are the focus of this current study, TCs were also included in the model training and runs for sake of completeness.

During training, which was performed using the NVIDIA Tesla K80 GPU on AWS, Jaccard (Intersection-Over-Union, or IOU) Loss is calculated in order to learn image features and to yield the IOU between predicted mask P and ground-truth mask G, and the model is trained until convergence with early stopping to help combat overfitting due to the relatively small dataset size. The equation is as follows:

$$J(P,G) = \frac{|P \cap G|}{|P \cup G|} = \frac{|P \cap G|}{|P| + |G| - |P \cap G|}$$

Note $J(P,G)$ measures IOU, so $L(J) = 1 - J(A,B)$ is minimized. Adam optimization was used at each step, as it intersects the benefits of RMSProp and momentum and is generally regarded to be computationally efficient. Finally, model hyperparameters including number of training epochs, learning rate, and regularization, are tuned to suit regional and global AR identification for the NEPAC box, with future goals of applications to other global regions included in the created ClimateNet-Regional dataset. Training batch size (4) and validation batch size (8) were kept constant during these hyperparameter changes.

5 Results and Discussion

The CGNet/ClimateNet model, pre-trained on global data, was first evaluated on the global, NEPAC, and SEPAC sets. It was found that the CGNet/ClimateNet results were comparable to those reported in [7] for the DeepLabv3+/ClimateNet model, and the globally-trained model performs relatively well on the NEPAC data (see Appendix Table 2).

	mIOU
NEPAC validation on NEPAC-trained	0.5732
NEPAC validation on Global-trained	0.5476

Table 1: Comparison of NEPAC mIOU after training on global vs. regional data. Trained using 1 epoch with learning rate $1e-3$.

Next, the model was trained on the regional NEPAC data. When randomly-initialized, the starting (learning rate of 0, 1 epoch) model training mIOU was 0.1875, and when initialized from the global weights, increased to 0.4841 (along with a lower starting loss). Thus, remaining experiments were conducted using this global weight initialization for training.

The mIOU was used as the evaluation metric for global, NEPAC, and SEPAC validation set performance on the NEPAC training set after 0 (initial), 1, 5, and 10 epochs, with a learning rate of $1e-3$ (chosen due to relatively small dataset). Table 1 shows the best mIOU (slightly higher than the globally-trained regional accuracy) for NEPAC trained on NEPAC data.

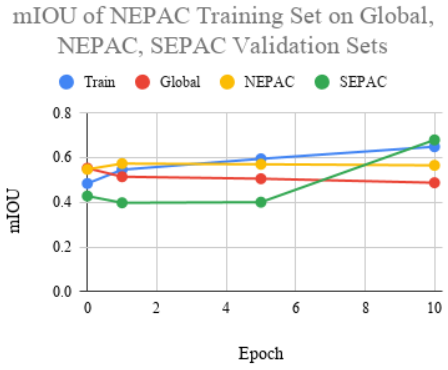

Figure 4: Training and validation mIOU for 0, 1, 5, and 10 epochs. Epoch "0" here refers to the state prior to training on the new NEPAC dataset.

Fig. 3 suggests two main points: (1) the number of epochs does not greatly affect the global and NEPAC accuracies, although (2) there do appear to be increases in training performance and slight decreases in validation performance, particularly for global and NEPAC validation sets, which suggests that the model may be slightly overfitting to the NEPAC training data; this would make sense because the model is being trained on snapshots from the NEPAC region.

Additional experiments were conducted to determine effects of epoch number, learning rate, and L2 regularization on mIOU for global and regional sets: 2 additional learning rates ($1e-4$ and $1e-5$) along with three levels of weight decay (0, $1e-3$ and $1e-4$) and 1 and 5 epochs were tested (please see a subset of hyperparameter experiments in Appendix Table 3). Results do not appear to show significant improvement with the weight decay term, although a learning rate of $1e-5$ may be too small to improve performance, and mIOU results are affected by lower TC accuracy; additional ranges of hyperparameters can be tested in future with more time.

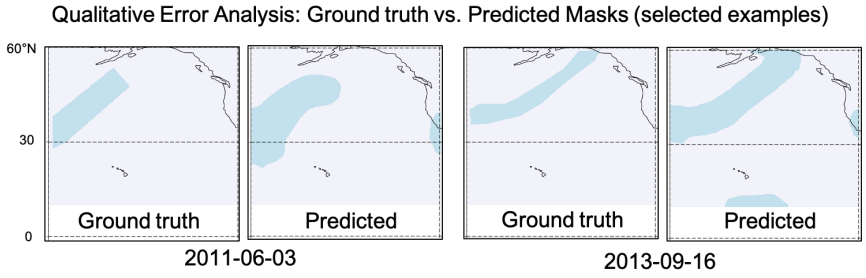

Figure 5: Segmentation mask comparisons, ground-truth and predicted, for NEPAC region. Trained on NEPAC in 5 epochs using learning rate $1e-4$ and weight decay $1e-3$, which had the highest NEPAC validation performance (0.5498) for the Appendix Table 3 experiments. Even so: the 2013 prediction successfully matches the AR trajectory—a crucial but difficult aspect of AR forecasting—but is ~2x wider than the ground-truth. The 2011 prediction is also wider, but both its shape and its trajectory differs from the ground-truth.

Fig, 5 illustrates why assessing predictions on regional datasets is particularly challenging. The new ClimateNet-Regional dataset uses segmentation masks from the original global set, which can vary (coarse in 2011 example, detailed in 2013); it is possible that with more precise expert labels on regional images, accuracy may further improve.

6 Conclusion and Future Work

This work produces the ClimateNet-Regional dataset and modifies CGNet/ClimateNet to conduct experiments to investigate regionally-trained model performance on global and regional datasets; NEPAC-trained NEPAC performance peaked at mIOU$= 0.5732$, exceeding the globally-trained accuracy on the regional set. Given more time and computational resources, I hope to build on this project by exploring the following avenues: continue ClimateNet-Regional experiments on additonal regions; regional dataset can allow more computational resources to be used for learning AR "movies"; expand AR classification to include different types of ARs based AR scale [6]. It is hoped that this work, along with the ClimateNet-Regional dataset, furthers the space of deep learning applied to AR and extreme weather identification.

7 Contributions and Acknowledgements

This study was performed by Deveshi Buch. Many thanks to project mentor Avoy Datta for guidance and helpful conversations; Professors Andrew Ng and Kian Katanforoosh for instruction on deep learning; and all CS230 teaching staff for support throughout the course. Computational resources provided by Amazon Web Services (AWS), using compute credits granted by CS230. Packages used include: NumPy, PyTorch, NetCDF4, Cartopy, NetCDF Operators (NCO). ClimateNet available under MIT License.

References

[1] Corringham, T. W., et al. (2019). Atmospheric rivers drive flood damages in the western United States. *Science advances*, 5(12), eaax4631.

[2] Dettinger, M. D., et al. (2011). Atmospheric rivers, floods and the water resources of California, *Water*, 3, 445–478, 2011.

[3] Espinoza, V., et al. (2018). Global analysis of climate change projection effects on atmospheric rivers. *Geophys. Res. Lett.*, 45, 4299–4308.

[4] Shields, C. A., et al. (2018). Atmospheric River Tracking Method Intercomparison Project (ARTMIP): project goals and experimental design. *Geoscientific Model Development*, 11, 2455–2474.

[5] Buch, D. (2018 June 25-28). Climatological Analysis of Atmospheric Rivers in the Eastern Pacific: A Comparative Study. 2nd International Atmospheric Rivers Conference, La Jolla, CA, United States.

[6] Ralph, F. M., et al. (2019). A Scale to Characterize the Strength and Impacts of Atmospheric Rivers. *Bulletin of the American Meteorological Society*, 100(2), 269-289.

[7] Kashinath, K., et al. (2021). ClimateNet: an expert-labeled open dataset and deep learning architecture for enabling high-precision analyses of extreme weather. *Geoscientific Model Development*, 14(1), 107-124.

[8] Liu, Y., et al. (2016). Application of deep convolutional neural networks for detecting extreme weather in climate datasets. arXiv preprint arXiv:1605.01156.

[9] Chen, L, et al. (2018). Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. *ArXiv*, abs/1802.02611.

[10] Wu, T., et al. (2020). CGNet: A Light-weight Context Guided Network for Semantic Segmentation. *IEEE Transactions on Image Processing* 30, 1169-1179.

[11] Viale, M, et al. (2018). Impacts of Atmospheric Rivers on Precipitation in Southern South America. *Journal of Hydrometeorology*, 19, 10; 10.1175/JHM-D-18-0006.1

Please note that any other citations are included within the body of the report, marked by an underline and hyperlink to the source.

Please see Appendix on next page.

Appendix

	BG IOU	TC IOU	AR IOU	mIOU
CN-Global	0.9389	0.2441	0.3910	0.5247
Global	0.9306	0.3344	0.3931	0.5527
NEPAC	0.9112	0.3978	0.3337	0.5476
SEPAC	0.8398	0	0.4445	0.4281

Table 2: IOU for global, NEPAC, SEPAC validation sets run on baseline model pre-trained on ClimateNet global data. Results from ClimateNet model (run on DeepLabv3+) shown for comparison (CN-Global). Mean human expert IOU is ~0.5120 (according to [7]), so performance already exceeds this for Global and NEPAC sets. mIOU performance on the SEPAC data is not as comparable to those of the other three—no TCs were found in the SEPAC set—but this is likely because it is smaller-scale regional data.

Num Epochs	LR	Weight Decay	Train mIOU	Global mIOU	NEPAC mIOU	SEPAC mIOU
1	1e-4	0	0.4336	0.5249	0.5425	0.4445
1	1e-5	0	0.4164	0.4991	0.5239	0.4298
1	1e-4	1e-3	0.4285	0.5226	0.5497	0.4330
1	1e-5	1e-3	0.4147	0.4967	0.5230	0.4273
1	1e-4	1e-4	0.4332	0.5256	0.5448	0.4425
1	1e-5	1e-4	0.4162	0.4989	0.5239	0.4295
5	1e-4	0	0.4634	0.5248	0.5396	0.4462
5	1e-5	0	0.4415	0.5162	0.5397	0.4439
5	1e-4	1e-3	0.4636	0.5145	0.5498	0.4409
5	1e-5	1e-3	0.4338	0.5099	0.5383	0.4395
5	1e-4	1e-4	0.4631	0.5239	0.5447	0.4451
5	1e-5	1e-4	0.4406	0.5154	0.5397	0.4433

Table 3: Subset of hyperparameter experiments with different learning rates, weight decay amounts, and epoch sizes. NEPAC-trained model with initialization from globally-pretrained weights.