Focus and track: pixel-wise spatio-temporal hurricane tracking

Anonymous Authors[1]

Abstract

We proposes a pixel-wise extreme climate event tracking framework to track a target in the multiple moving objects scenario. We applied our model to tackle the challenging hurricane tracking problem. The proposed framework consists of two sub-models based on multi-layered ConvLSTM: a focus learning and a tracking model. Focus learning model learns location and appearance of target at first frame of video with one-shot auto-encoding fashion, and then, learned feature is fed into tracking model to follows the target in consecutive time frames. Extensive experiments show that the proposes tracking framework significantly outperforms against state-of-the-art tracking algorithms.

1. Introduction

Tracking climate events are pressing and challenging problems that humanity has faced for a long time. Traditionally, most conventional approaches have been built upon human expertise based on scientific intuition and related physics variables (Lorenc, 1986). Recently, computer vision community has made significant progress by applying various pattern recognition techniques in visual object tracking, a task to locate a target object in a video, maintaining its identity and yielding its individual trajectory, given its initial location in the first frame. **Extreme climate event tracking** is similar to visual object tracking, but it has unique and challenging aspects:

1. Climate events may be dependent on *longer-term* and *wider-range* spatio-temporal dynamics (known as 'butterfly effect') between multiple scientific variables than the targets in visual object tracking do on RGB pixels.
2. The target events are not often defined as rigid bodies, *flexibly changing their shape with no clear boundary*, and are *harder to visually distinguish* from each other. This fact makes it difficult to associate an object of interest with the correct one in consecutive frames.

Because of these unique properties, a conventional tracking-by-detection method, which detects the object mainly by its appearance but relatively neglects spatio-temporal dynamics, is less suitable to this problem. An ideal climate event tracker needs to effectively take long-term and wide-range dynamics into account, capturing subtle differences among events from sparsely collected training data. In this work, we propose a simple but robust end-to-end model, suitable for the climate event tracking problem utilizing ConvLSTM. Specifically, the proposed model consists of two sub-modules, **(1) the focus learning module** to learn where and what to focus, and **(2) the tracking module** to track what we focused on. The focus learning module is designed to extract the latent feature of the target event from the first frame, given spatio-temporal data and the initial location of the target. Given the representation of the target event, the tracking module localizes the learned feature of a target object in the subsequent frames, predicting its location.

2. Related Work

Conventional extreme climate event detection and tracking methods rely on numerical simulation-based methods, including an ensemble of multiple prediction models or multi-scale prediction systems (Weber, 2003; Elsberry et al., 2008; Sippel & Zhang, 2008; Poroseva et al., 2010; Majumdar & Finocchio, 2010; Snyder et al., 2010; Tien et al., 2012; Qi et al., 2014; Thanh et al., 2016; Dong & Zhang, 2016). Recently, climate research communities have started to leverage various deep learning techniques. Extreme climate event detection and localization problem was tackled with RNN (Alemany et al., 2018) and spatio-temporal CNN (Racah et al., 2017). Also, ConvLSTM (Kim et al., 2019a) and incremental neural network (Kim & Hasegawa, 2018) were proposed to predict future trajectory of hurricanes and cyclones. R-CNN was applied to classify different types of extreme climate events (Liu et al., 2016; Kim et al., 2017). Kim et al. (Kim et al., 2019b) predicted the concentration of air pollutants using LSTM. Most existing works, however, have not addressed unique challenges to deal with

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

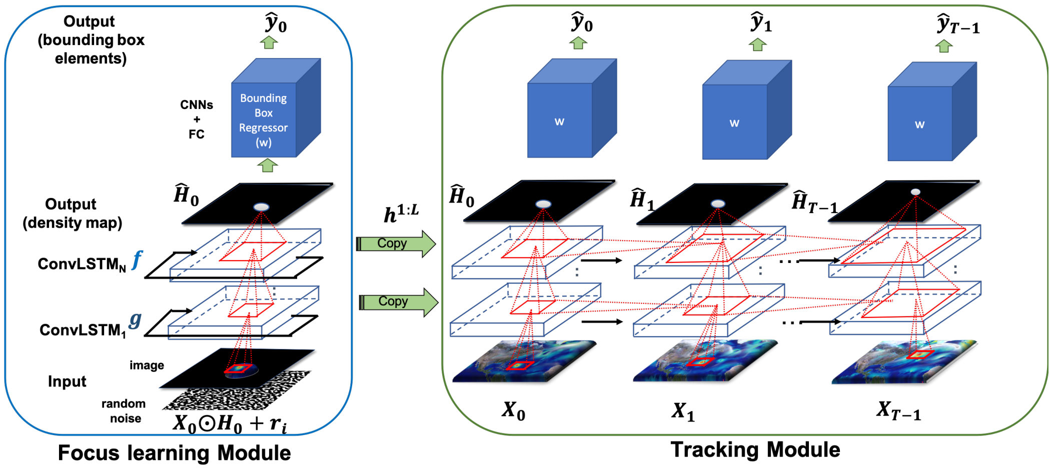

Figure 1. Overview of our proposed focus (*left*) and tracking(*right*) approach.

sparse climate data covering wide geographic range for an extended period. In this paper, we tackle the unique challenges of the climate event tracking problem with a ConvLSTM-variant model, which is specially designed to capture wide-range of spatio-temporal dynamics.

3. Method

3.1. Problem Formulation

Notations. We denote $\mathbf{X} = \{\mathbf{X}_0, \mathbf{X}_1, ..., \mathbf{X}_{T-1}\}$ be a climate video of length T, and each $\mathbf{X}_i \in \mathbb{R}^{m \times n \times c}$ is a 2-D climate image of size $m \times n$ with c climate channels (e.g., surface-level pressure or wind speed). $\mathbf{X}$ may contain multiple trajectories of target events (e.g., hurricanes), starting or ending at any time, but we track one target event at a time. The ground truth $\mathbf{y} = \{\mathbf{y}_0, \mathbf{y}_1, ..., \mathbf{y}_{T-1}\}$ is the location of the target climate event. Each $\mathbf{y}_i \equiv \{x, y, w, h\}$ is given as a bounding box of the target event, where (x, y) is the top-left point, w is the width, and h is the height of it.

Extreme Climate Event Tracking Problem. Given a climate video $\mathbf{X}$ and the initial location of the target event $\mathbf{y}_0$, the goal is to estimate its locations $\hat{\mathbf{y}}_i$ in subsequent image frames closely to the ground truth $\mathbf{y}_i$, for $i = 1, ..., T-1$.

3.2. Proposed Framework

Given a climate video $\mathbf{X}$ and an initial location of the target object $\mathbf{y}_0$, our framework aims to predict the trajectory of the target object. It is tempting to regress bounding box elements directly from the input image $\mathbf{X}_i$, but we have a couple of issues. First, as the boundary between the target event and the background is often blurry, a direct bounding box ($\hat{\mathbf{y}}_i$) regression from pixels $\mathbf{X}_i$ is challenging. Second, when there exist multiple events in the frame, data association is difficult as their appearance is often too similar to distinguish visually. To address these challenges, we represent both ground truth and prediction as density maps. That is, each ground truth label $\mathbf{y}_i$ is transformed to a density map $\mathbf{H}_i \in \mathbb{R}^{m \times n}$ with Gaussian mixtures $\mathcal{N}(\mathbf{y}_i, \sigma^2 \mathbf{I})$, where the variance σ^2 is determined by the hurricane radius.[1] Given inputs (the climate video $\mathbf{X}$, the initial bounding box $\mathbf{y}_0$ of the target, and a density map $\mathbf{H}_0$ created from $\mathbf{y}_0$), we model the tracking problem as a pixel-wise regression problem at each time step, minimizing the pixel-wise mean squared error between the ground truth $\mathbf{H}_i$ and our prediction $\hat{\mathbf{H}}_i \in \mathbb{R}^{m \times n}$, with the probability of an object observed in each pixel. Once $\hat{\mathbf{H}}$ is obtained, we regress it to the original bounding box $\hat{\mathbf{y}}$.

Model Overview. Our framework in Figure 1 consists of two modules: (1) the *focus learning module* that learns to extract the latent features of the target object at the initial time step, and (2) the *tracking module* to estimate the bounding box information of the target trajectory in subsequent time frames. At the initial frame of the video, the focus learning module takes the input climate image $\mathbf{X}_0$ multiplied in a pixel-wise manner with the density map of the target $\mathbf{H}_0$, and estimate the density map of the target $\hat{\mathbf{H}}_0$ By feeding slightly perturbed input and output pairs multiple times using recurrent convolutional networks, the focus learning module imprints appearance and location of the target object in a one-shot learning fashion in its hidden state. In this manner, the focus learning module learns where and what to focus at the beginning of the tracking procedure. Once the focus learning module learns the feature of the target object, the tracking module takes the hidden state from the focus learning module and detect location of the target object in subsequent frames by updating the hidden

[1] We chose diagonal one because most hurricanes are in circular shape. For other types of extreme climate events, e.g., atmospheric river, we may use more general covariance matrix.

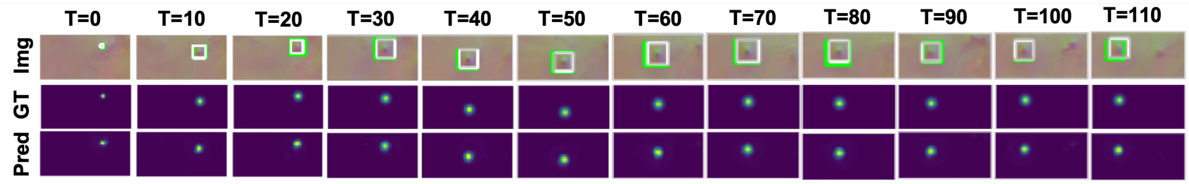

Figure 2. Hurricane tracking results. From the top, input climate image(channeled with PSL, U850, V850) overlaid with bounding box results (White: ground truth, Green: prediction), ground truth density map, and output density map.

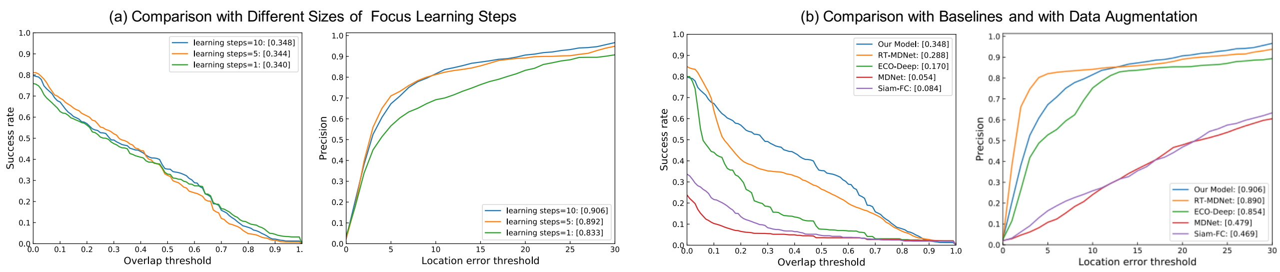

Figure 3. Success plot and ROC plot with AUC score. (a) Comparison by changing focus learning steps *(left)* (b) Comparison with baselines and comparison with data augmentation *(right)*

states. We use many-to-many RNN architecture with ConvLSTM cells. The weight sharing is applied between the focus learning and the tracking modules to update dynamical spatio-temporal changes of the target object. We regress from the produced density map $\hat{\mathbf{H}}_0$ to the original bounding-box ground truth, $\mathbf{y}_0$, with a *bounding box regressor* w, where its output consists of the four bounding box elements, $\{x, y, w, h\}$. We use a multi-layered CNN followed by fully connected layers for *w*. To train the model, we minimize the pixel-wise mean squared loss between the estimated density-map $\hat{\mathbf{H}}_0$ and the ground-truth density-map $\mathbf{H}_0$, averaged over each perturbed images. Similarly, for the model *w*, we minimize the squared loss between the estimated bounding box elements $\hat{\mathbf{y}}_\mathbf{0}$ and ground truth $\mathbf{y}_\mathbf{0}$.

4. Experiment and Results

We evaluate our proposed tracking framework on CAM5 hurricane dataset with 3 channels of surface level pressure(PSL), eastward wind(U850), zonal wind(V850). The main contribution of our tracking framework is the capability of the focus learning model to rigorously learn the feature of the target object by repeating learning steps. To show the effectiveness of repeated feature learning, we conduct comparison among variations of focus learning steps. Figure 3 *(left)* shows the success plots and the ROC plots of hurricane tracking trained and tested with CAM5 climate data. Increasing focus learning steps slightly increases the tracking performance until the learning steps of 10. We discuss the larger the learning step is, the stronger the model imprint the feature of target in its hidden state. Depending on the data and target, the size of learning step can be tuned accordingly. Figure 2 shows the qualitative results of our method in its best performing case(learning step of 10) on challenging hurricane tracking scenario where the new hurricane starts to emerge from the left side of the image frame at about time step of 90. Both from the bounding box results and density map results, we see our model robustly tracks the target hurricane from start to end for a long period of time (110 time steps=330 hours) without being confused by new hurricane. We presents a comparison study of our tracking framework with the state-of-art baselines, including Real-time MDNet, ECO-Deep, MDNET, Siam-FC. We compares both ROC and success plots including AUC scores for all baselines with our framework. As shown in Figure 3 *(right)*, our tracking framework significantly outperforms all baseliness. Comparison between the best performance of our framework with the best performing baseline (RT-MDNET) shows about **1.8% performance gain in its AUC of precision and 20.8% performance gain in its AUC of success rate**.

5. Conclusion

We proposed a novel pixel-wise tracking framework based on ConvLSTM specifically tackling extreme climate event. Our tracking framework consists of focus learning module to learn feature of target at first frame and tracking module to follow target at consecutive frames. We achieved outstanding performance in hurricane tracking, compared to the state-of-the-art tracking algorithms.

References

Alemany, S., Beltran, J., Perez, A., and Ganzfried, S. Predicting hurricane trajectories using a recurrent neural network. In *arXiv:1802.02548v2*, 2018.

Dong, L. and Zhang, F. Obest: An observation-based ensemble subsetting technique for tropical cyclone track prediction. *Weather and Forecasting*, 31(1):57–70, 2016.

Elsberry, R. L., Hughes, J. R., and Boothe, M. A. Weighted position and motion vector consensus of tropical cyclone track prediction in the western north pacific. *Monthly Weather Review*, 136(7):2478–2487, 2008.

Kim, S., Ames, S., Lee, J., Zhang, C., Wilson, A. C., and Williams, D. Massive scale deep learning for detecting extreme climate events. In *Climate Informatics*, 2017.

Kim, S., Kim, H., Lee, J., Yoon, S., Kahou, S. E., Kashinath, K., and Prabhat, M. Deep-hurricane-tracker: Tracking and forecasting extreme climate events. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2019a.

Kim, S., Lee, J. M., Lee, J., and Seo, J. Deep-dust: Predicting concentrations of fine dust in seoul using lstm. In *Climate Informatics*, 2019b.

Kim, W. and Hasegawa, O. Time series prediction of tropical storm trajectory using self-organizing incremental neural networks and error evaluation. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 22(4):465–474, 2018.

Liu, Y. et al. Application of deep convolutional neural networks for detecting extreme weather in climate datasets. *ArXiv:1605.01156*, 2016.

Lorenc, A. C. Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 112(474):1177–1194, 1986.

Majumdar, S. J. and Finocchio, P. M. On the ability of global ensemble prediction systems to predict tropical cyclone track probabilities. *Weather and Forecasting*, 25 (2):659–680, 2010.

Poroseva, S. V., Lay, N., and Hussaini, M. Y. Multimodel approach based on evidence theory for forecasting tropical cyclone tracks. *Monthly Weather Review*, 138(2): 405–420, 2010.

Qi, L., Yu, H., and Chen, P. Selective ensemble-mean technique for tropical cyclone track forecast by using ensemble prediction systems. *Quarterly Journal of the Royal Meteorological Society*, 140(680):805–813, 2014.

Racah, E., Beckham, C., Maharaj, T., Kahou, S. E., Prabhat, and Pal, C. ExtremeWeather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events. In *Advances in Neural Information Processing Systems (NIPS)*, 2017.

Sippel, J. A. and Zhang, F. A probabilistic analysis of the dynamics and predictability of tropical cyclogenesis. *Journal of the Atmospheric Sciences*, 65(11):3440–3459, 2008.

Snyder, A. D., Pu, Z., and Zhu, Y. Tracking and verification of east atlantic tropical cyclone genesis in the ncep global ensemble: Case studies during the NASA African Monsoon multidisciplinary analyses. *Weather and Forecasting*, 25(5):1397–1411, 2010.

Thanh, C., Tien, T. T., and Chanh, K. Q. Application of breeding ensemble to tropical cyclone track forecasts using the regional atmospheric modeling system (rams) model. *Applied Mathematical Modelling*, 40(19–20): 8309–8325, 2016.

Tien, T. T., Thanh, C., Van, H. T., and Chanh, K. Q. Two-dimensional retrieval of typhoon tracks from an ensemble of multimodel outputs. *Weather and Forecasting*, 27(2): 451–461, 2012.

Weber, H. C. Hurricane track prediction using a statistical ensemble of numerical models. *Monthly Weather Review*, 131(5):749, 2003.