Tropical and Extratropical Cyclone Detection Using Deep Learning

Christina Kumler-Bonfanti,[a,b] Jebb Stewart,[b] David Hall,[c] and Mark Govett[b]

[a] Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, Colorado
[b] NOAA/OAR/ESRL/Global Systems Laboratory, Boulder, Colorado
[c] NVIDIA Corporation, Boulder Colorado

ABSTRACT: Extracting valuable information from large sets of diverse meteorological data is a time-intensive process. Machine-learning methods can help to improve both speed and accuracy of this process. Specifically, deep-learning image-segmentation models using the U-Net structure perform faster and can identify areas that are missed by more restrictive approaches, such as expert hand-labeling and a priori heuristic methods. This paper discusses four different state-of-the-art U-Net models designed for detection of tropical and extratropical cyclone regions of interest (ROI) from two separate input sources: total precipitable water output from the Global Forecast System (GFS) model and water vapor radiance images from the Geostationary Operational Environmental Satellite (GOES). These models are referred to as International Best Track Archive for Climate Stewardship (IBTrACS)-GFS, Heuristic-GFS, IBTrACS-GOES, and Heuristic-GOES. All four U-Nets are fast information extraction tools and perform with an ROI detection accuracy ranging from 80% to 99%. These are additionally evaluated with the Dice and Tversky intersection-over-union (IoU) metrics, having Dice coefficient scores ranging from 0.51 to 0.76 and Tversky coefficients ranging from 0.56 to 0.74. The extratropical cyclone U-Net model performed 3 times as fast as the comparable heuristic model used to detect the same ROI. The U-Nets were specifically selected for their capabilities in detecting cyclone ROI beyond the scope of the training labels. These machine-learning models identified more ambiguous and active ROI missed by the heuristic model and hand-labeling methods that are commonly used in generating real-time weather alerts, having a potentially direct impact on public safety.

KEYWORDS: Extratropical cyclones; Subtropical cyclones; Tropical cyclones; Pattern detection; Machine learning; Neural networks

1. Introduction

Enormous quantities of data are being generated by increasingly higher-resolution models, next-generation satellites, radar, and a myriad of ground-based in situ observation data. Only a small percentage of these data are processed and utilized in real-time applications, such as severe weather alerts, numerical weather prediction (NWP) models, and data assimilation (DA) because of the time required to process them. In addition, detection of severe weather in model output or satellite images relies upon trained experts to identify significant features such as cyclones or weather fronts. However, this process takes time and is subject to human bias and error.

Historically, heuristic rule-based models have been used to automatically identify strong storms by using a set of well-defined rules. However, complex weather phenomena are difficult to quantify and heuristic models tend to be brittle. A community effort to intercompare extratropical cyclone detection and tracking algorithms, known as Intercomparison of Mid Latitude Storm Diagnostics (IMILAST), compiles an analysis of extratropical cyclone trackers that both qualitatively and quantitatively compare the performance of these types of models, detailing the limitations of each (Neu et al. 2013). For example, some models are limited to the ocean and do not function over land. These methods can be accurate when the feature of interest is well defined, but struggle to identify more

ambiguous regions. Tropical cyclones are well defined as any closed, low-level rotating system with thunderstorms and high winds that originate over the waters of the tropics (Holland 1993; Merrill 1993). However, there is no universally agreed-upon definition for extratropical cyclones. NOAA defines extratropical cyclones as low-pressure systems that could be associated with a front (Atlantic Oceanographic and Meteorological Laboratory 2020), whereas the American Meteorological Society defines them as any cyclonic storm that is not a tropical cyclone (American Meteorological Society 2019). Heuristic-based models typically require a specific set of meteorological variables provided by weather model outputs and often cannot be run directly with observations from sources such as satellites without further information or routines.

Machine-learning (ML) techniques can be used effectively to identify a range of different meteorological features that may be broadly classified as regions of interest (ROI). For instance, ML methods have successfully been used to classify certain types of thunderstorms (Jergensen et al. 2020). In contrast with explicit heuristic approaches, ML techniques learn to define features implicitly by working backward from a given set of examples. Neural networks (NNs) are a type of ML techniques that take inspiration from human cognition and are able to perform many tasks previously beyond the reach of explicit rule-based approaches. An advantage of this approach is that NNs can more easily identify complex and ambiguous phenomena that are not easily described by simple heuristics. Additionally, a trained NN is often much faster than a hand-coded heuristic model making it more practical for real-time identification.

Corresponding author: Christina Kumler-Bonfanti, christina.e.kumler@noaa.gov

There are three main reasons NNs have become increasingly popular for image recognition: increasing quantities of data, the exponential growth of computational resources, and the ready availability of modern ML tools, such as Python libraries, that simplify the process of model development. NNs have been trained to identify or classify both clear or ambiguous objects on large datasets, such as Modified National Institute of Standards and Technology database (MNIST; http://yann.lecun.com/exdb/mnist/), which classifies written digits as numbers from zero to nine (Deng 2012), and ImageNet, which classifies objects contained in a wide variety of images (Deng et al. 2009). In some cases, NNs have been trained to perform object recognition tasks at superhuman levels. Such comparisons demonstrate enormous potential benefits in using NNs for classification in many Earth science fields (Ciresan et al. 2012).

Convolutional neural networks (CNNs) are a class of NNs that have proven to be particularly effective for image analysis (Patterson and Gibson 2017). They rely upon one or more convolutional layers that are able to recognize important features regardless of where that feature appears within the image. These NNs are particularly effective for image problems because they use small segments or squares of input data and retain spatial information by surrounding pixels needed for analysis. CNNs are used in a variety of science applications including medical scan analysis, signal processing, short-term prediction, image recognition, identification, and segmentation tasks (Cheng et al. 1996; Dia 2001; Cichoki and Unbehauen 1996; Lawrence et al. 1997; Liang and Hu 2015; Giacinto and Roli 2001; Long et al. 2015). AlexNet and VGGNet, two of the most popular and widely used CNNs, were trained with millions of images and are highly accurate (Krizhevsky et al. 2017; Simonyan and Zisserman 2015).

Deep learning (DL) is a term used to describe the optimization of NNs with multiple layers. Deep networks tend to be large relative to other ML models and their training typically requires powerful graphics processing unit (GPU) acceleration in order to make training practical. In comparison with other approaches, deep CNNs are particularly effective for image classification and segmentation (LeCun et al. 2015; He et al. 2016; Liu and Deng 2015). DL models are increasingly being applied to complex problems in Earth science and meteorology, including probabilistic hail forecasting, cloud classification, predicting algal blooms, tropical-cyclone-track forecasts, and severe weather detection (Gagne et al. 2017; Giffard-Roisin et al. 2020; Lagerquist et al. 2020; Lee et al. 2004; Recknagel et al. 1997; McGovern et al. 2017).

This paper describes work using DL to perform image segmentation for the detection of tropical and extratropical cyclone ROI. The automated detection of significant ROIs is potentially valuable for improving data assimilation and model initialization for numerical weather prediction models, among many other possible applications. For example, Fig. 1 shows a water vapor image produced by an older generation NOAA Geostationary Operational Environmental Satellites (GOES). In this figure, red boxes indicate ROIs for Hurricane Harvey and other cyclones. DL segmentation enables the rapid, automatic identification of specific regions of interest from high volume datasets such as GOES satellite observations. This type of automation is critical due to the enormous volume of data waiting to be processed. In particular, NOAA's latest-generation *GOES-16* and *GOES-17* have 16 bands that take data samples every half second to five minutes at $1/2$ to 2 km resolution (Schmit et al. 2018). Only an estimated 3%–7% of observational data are selected to be used in NWP models, and an even smaller fraction of that data are actually assimilated into the models (Weingroff 2014). *GOES-16* and *GOES-17* produce over 100 times as much data as the previous GOES missions, with high potential value for NWP. With the current system, the amount of data far exceeds the computing time available to process it and instead, simple data thinning techniques are applied and the majority of the data are discarded. In contrast, targeted selection of data guided by DL can intelligently select highly relevant observations, providing additional data in regions that are active or soon to be active.

There are a number of DL models that may be used to solve problems in the Earth sciences. The selection of the right model depends on the type of problem being solved. For image-segmentation tasks, the U-Net deep convolutional neural network has become a de facto standard choice. The U-Net structure is specifically designed for image-segmentation tasks. It employs an encoder–decoder structure with skip connections that is able to identify important features at multiple spatial scales. Although originally designed for medical segmentation tasks, the U-Net has proven successful at extracting both the coarse and finescale features important for determining events of interest in meteorological data (Ronneberger et al. 2015; Ronneberger 2017). Deep convolutional networks, such as U-Net, tend to identify simple features in the first few layers, which are combined to form high-level features in subsequent layers (LeCun and Bengio 1995). Both low-level and high-level features are learned simultaneously in an iterative fashion as the model learns to fit the training data.

The paper is organized as follows. In section 2, the U-Net architecture is described, as are the numerical metrics used to evaluate its success. Section 3 describes both qualitatively and quantitatively the design and performance of the best U-Net model obtained for identifying tropical cyclones using the Global Forecast System (GFS) total precipitable water field as inputs. In section 4, three additional U-Net models are introduced that identify both tropical and extratropical cyclones from GFS total precipitable water output data as well as GOES water vapor data inputs. Section 5 provides a summary of the U-Net approach and offers further discussion on potential applications in NWP.

2. Deep-learning architecture: The U-Net

The following four sections describe the programming environment, the model design, performance metrics, and the central processing unit (CPU) and GPU performance measurements for the training and testing environment. Performance and times will change depending on the system that is training and testing the model. All code and data are publicly available through request to NOAA's Global Systems Laboratory.

a. Programming environment

The U-Net models were written with the Keras NN application programming interface (API) in the Python coding language because of its user-friendly interface (Gulli and Pal 2017).
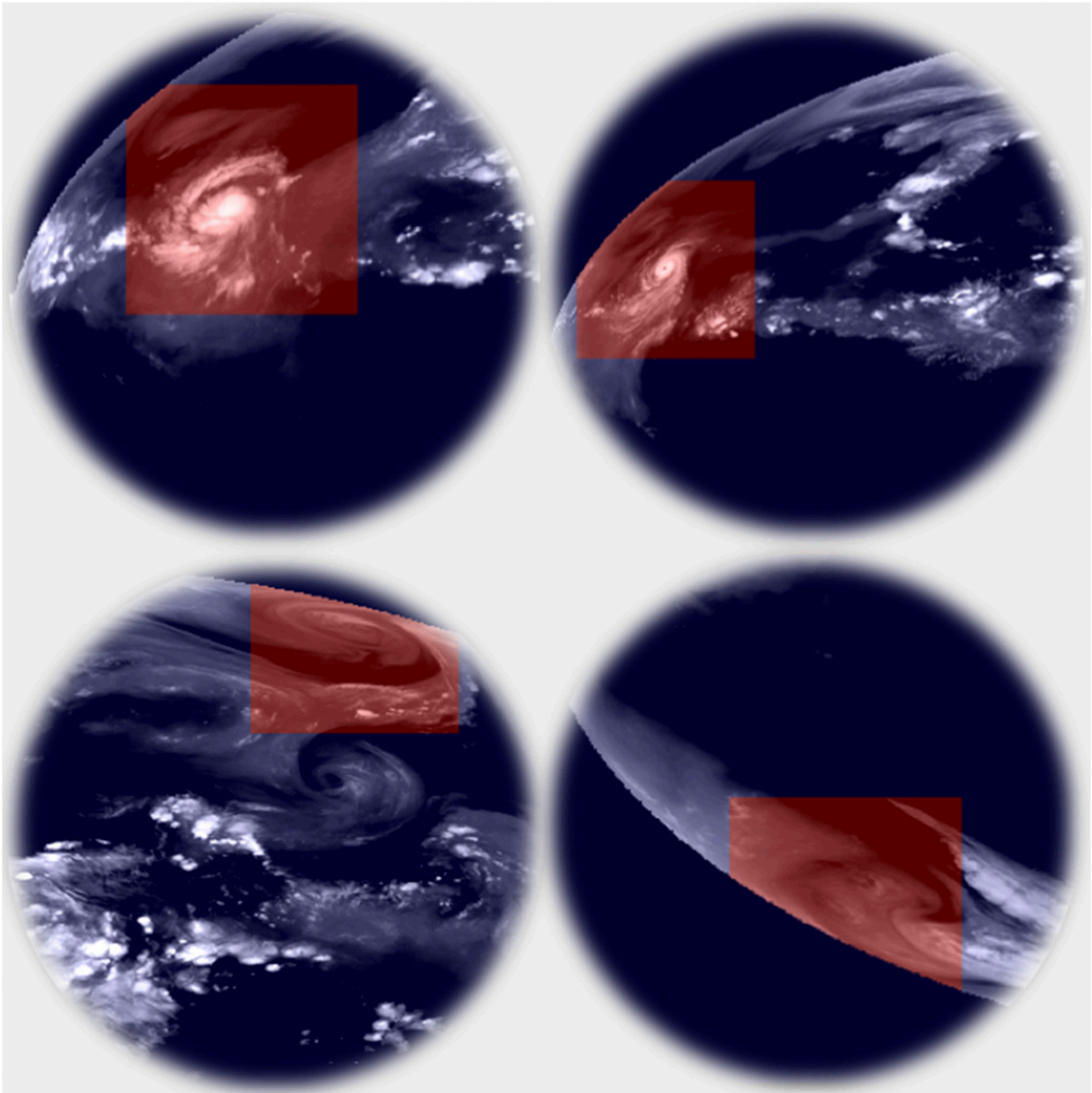
FIG. 1. Four different examples of cyclone appearance in the *GOES-13* water vapor are highlighted in a red ROI. Shown are Hurricane Harvey on (top right) 19 Aug and (top left) 21 Aug (left) 2017 and Extratropical cyclones on (bottom right) 19 Aug and (bottom left) 20 Aug 2017 in the Northern and Southern Hemispheres.

Keras is simple to use and well suited both for model prototyping and deployment. Keras offers several framework options "under the hood" and Google's TensorFlow framework was selected for its ease of use and its performance on multiple GPU machines. This choice of framework enables training, analysis, and visualization to be readily combined in a single script.

b. U-Net design

The cyclone ROI U-Net models have several, dynamic hyperparameters common to DL models that were fine-tuned.

Some of which include but not limited to number of convolution layers, pooling layers, rectifier activation functions, and loss functions. Unique to the U-Net structure, there are additional up convolutions and skip connections, which connect relevant information that might be lost in the pooling layer to the expanding path. The structure depth can vary depending on the number of pooling and up-convolution layers, where each pooling layer reduces the size of the feature map. The maximum depth is restricted by the initial input size. The cyclone ROI U-Net models each had unique depths for optimal model
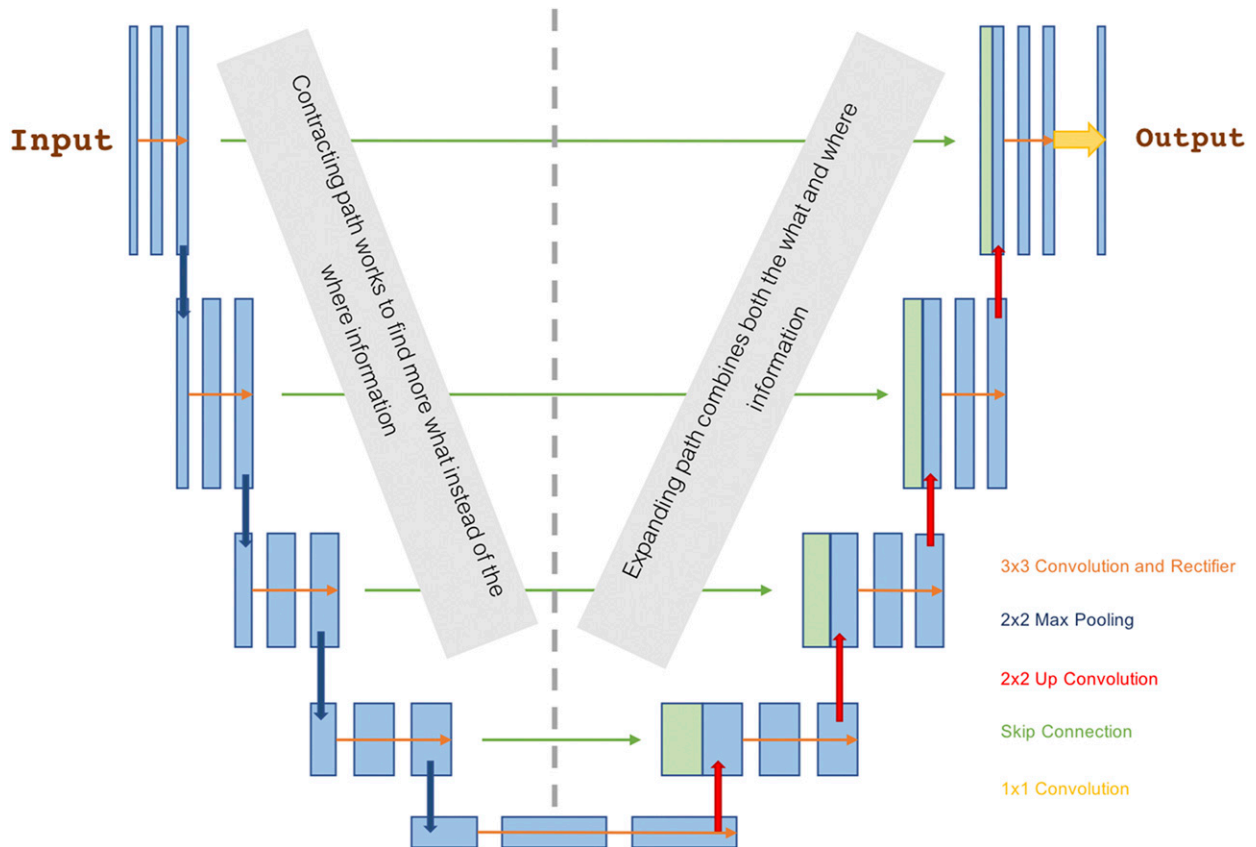
FIG. 2. The structure of the U-Net model for all four models. Depth of the model, or how many layers of convolution and maximum pooling for the model, is dynamic and several depths were experimented with to obtain the best architecture for each model. Orange arrows indicate each convolution, dark blue represents the maximum-pooling layer, green arrows indicate the skip connections, and red arrows indicate the up convolutions.

performance; Fig. 2 illustrates a general U-Net structure that was used as a base for all four of the U-Net models. While the depth was experimented with each different model, the convolutions and pooling layer design stayed the constant.

As the U-Net contracts, or encodes, it extracts generalized information from the input image. During the U-Net expansion, or decoder part, skip connections join generalized information with more localized information, often represented graphically as a ''U'' shape. Each cyclone ROI U-Net model has batch normalization after the convolution, which normalizes the inputs into the next layer, to increase training performance and time.

Additional hyperparameters that were experimented with for the cyclone ROI U-Net models were the activation function, dropout or noise, and loss functions. The activation function is used at each convolution and after tuning experiments the rectified linear unit (ReLu) performed best across the designed U-NETs. Other functions were used in experiments; however, the ReLu consistently provided the best results. ReLu will take the convolution field as inputs and if the value is positive, keep it. If the value is negative, it sets it equal to zero (Glorot et al. 2011). The model can avoid overfitting with either noise or dropout additions to the U-Net. Adding

''noise'' is when random values are taken from something such as a Gaussian curve and applied to the output at the end of a layer. Dropout is where randomly selected neuron units are turned off during the gradient process, essentially preventing the model from learning from certain neurons for that iteration (Srivastava et al. 2014). A loss function is how a DL model learns and different loss functions perform better depending on the type of dataset distribution.

The datasets used to train the U-Net models are divided into training, validation, and testing groups. The training and validation data are used to build the model with its weights and features. The test data are kept hidden from the training and validating stages and treated as new, unseen inputs that evaluate the robustness of a model on completely separate input data. To avoid memorization and overfitting by the U-Net, the training, validation, and test data must be chosen carefully. For instance, the model can identify subsequent time steps from long-lasting cyclones that persist through three or more image inputs and will overfit the model due to memorization. The cyclone ROI U-Net models accounted for this issue by having training, validation, and testing datasets that were separated by yearly chunks. The yearly breakup was determined based on available input data and labels.

Another parameter that contributes to either overfitting or poor performing models comes from selecting too high or low of a number of epochs used to train a model, respectively. A single completed epoch is when all of the data in the whole training set have been iterated through training and validation; this is done within batches. A batch is a random subgroup of data selected from whole training dataset that is used in training and validation stages against each other. As the training process works through more epochs, each refines the previously designed weights of the model. Selecting an optimized combination of the batch size of the inputs into the U-Net as well as the number of epoch iterations that are allowed to run produces the U-Net that converges on an optimized model. A model is determined to converge on the best model when the loss function, or training errors, is minimized and the training accuracy is high.

U-Net models can take as many input channels of data as memory allows so long as these inputs are all of the same dimension. To provide additional information to the cyclone U-Net models, all the cyclone ROI U-Net models were run with three input channels that were selected from the same data input field. The three channels represented data at the current t, $t - 1$, and $t - 2$ time steps for the starting t time step. This method gave the U-Nets a sense of time, cyclone rotation, and performance improvements.

A segmentation DL model means that the cyclone ROI U-Net models have input image pixel values defined to be either 0 for no cyclone or 1 for yes there is a cyclone. The model produced image pixel probability values that ranged from 0 to 1, where 1 is 100% likely or confident that the pixel is a cyclone pixel and 0% means no cyclone. A threshold can be applied to these outputs to obtain higher-probability outputs only. There is one last parameter that was changed based on the size of the cyclone that was being segmented. Each cyclone center, represented by the labeled data latitude and longitude center point, is additionally labeled by a pixel bounding box that contains all 1s. This is done to offset any errors in cyclone center value as well as to encompass a larger area of the storm. It also provides more "yes cyclone" pixels in an image that contains mostly "no cyclone" pixels.

c. Evaluating U-Net accuracy and performance

The cyclone ROI U-Nets are a classic imbalanced dataset problem since both tropical and extratropical cyclone labels are rare-event labels in any given input image, which presents a challenge in measuring the U-Net model performance. Presenting results only in terms of accuracy is not sufficient. High accuracy for these cyclone U-Nets indicates that while they are good at detecting the noncyclone events, it is unclear how good the models are for the yes-cyclone event detection. To understand the similarity between the truth labels and the U-Net model identified ROI, the coefficient values from the loss functions were evaluated, where the values range over [0, 1] and 1 represents a perfect match. The best ML models are designed to minimize the loss function. Four loss functions known to work better for imbalanced datasets were chosen: binary cross entropy (BCE), Dice, Tversky, and focal loss.

The first loss function, BCE, is defined in Eqs. (1) and (2) where p is the model probability over [0, 1] and y is the classification value from the truth-labeled point (Lin et al. 2017). This loss function is logarithm based in nature and penalizes instances in which the model predicts the wrong result:

$$\text{BCE}(p, y) = -\log(p) \quad \text{if} \quad y = 1 \quad \text{and} \tag{1}$$

$$\text{BCE}(p, y) = -\log(1 - p) \quad \text{otherwise.} \tag{2}$$

The second and third (Dice and Tversky, respectively) loss functions are intersection-over-union (IoU) methods that overlap regions of "yes cyclone" and no cyclone (Sudre et al. 2017). The Tversky coefficient (TC) defined in Eq. (3) assigns a weight to false positive (α) and false negative (β) results:

$$\text{TC} = \frac{|X \cap Y|}{|X \cap Y| + \alpha|X - Y| + \beta|Y - X|}. \tag{3}$$

In the instance of all four ROI models, it was found that setting $\alpha = 0.3$ and $\beta = 0.7$ gave better model performance since the model gave more weight to the instances in which a cyclone detection was missed. The Dice coefficient shown in Eq. (4) is simply the TC equation reduced when $\alpha = \beta = 0.5$:

$$\text{DC} = \frac{2|X \cap Y|}{|X| + |Y|}. \tag{4}$$

The fourth, focal loss, function is designed to reduce the weight of the predominant no-cyclone pixels so that the CNN focuses on the harder and fewer, yes-cyclone pixels. For a two-class problem such as the cyclone ROI, focal loss is defined as

$$\text{FL}(p) = -(1 - p)^{\gamma} \log(p). \tag{5}$$

Notice that if $\gamma = 0$, the focal loss is equal to the cross-entropy loss function. In setting $\gamma > 0$, the loss of a nearly correctly labeled point becomes orders of magnitude less than if $\gamma = 0$. For the cyclone ROI models, $\gamma = 2$. This is a standard number for most implementations of this loss function.

d. CPU and GPU usage

The training period of U-Net and DL models can take long periods of time on CPUs. Training time improves substantially when the U-Net model uses GPUs (Bahrampour et al. 2015). Table 1 shows the difference between model training times from an experimental cyclone ROI U-Net model on a CPU and NOAA's GPUs. The dramatic difference in training times proved the need for GPU use in order to train models in a practical time frame. Both the data processing for U-Net inputs and U-Net model training processes were run on NOAA's GPU supercomputer system. The Horovod DL software-enabled running on multiple GPUs simultaneously and further reduced the U-Net training time. Horovod was designed by Uber to support both startup and speedup of DL using TensorFlow distributed on large-scale systems (Sergeev and Del Balso 2018). Once a model has been trained, the inference stage, where a trained model identifies ROI from unseen data input, is completed in a fraction of time.

TABLE 1. Comparison between an IBTrACS U-Net model training and inference times on a CPU system as compared with running on GPU.

	CPU	GPU
Hardware	Two 10 Core Haswell Intel	8 Tesla (P100)
Training per epoch	11.5 h	3 min
Complete training (~70 epochs)	~5 weeks	~3 h
Inference	1 s	40 m s

3. Tropical cyclone ROI U-Net model results

The U-Net is a type of supervised DL model that performs best when trained with substantial quantities of broadly representative, labeled data. The labeled data provide the "truth" labels for the input training data, indicating which areas of an input image encompass a cyclone. This section will discuss the tropical cyclone ROI U-Net model, which will be referred to as the IBTrACS-GFS U-Net. It was trained using truth labels from the International Best Track Archive for Climate Stewardship (IBTrACS) tropical cyclone database and detected the locations of the tropical cyclones in the GFS total precipitable water output data field (Knapp et al. 2010). To remain consistent with the Saffir–Simpson scale, a wind threshold of 34 kt ($1 \, \mathrm{kt} \approx 0.51 \, \mathrm{m \, s}^{-1}$) or higher was applied to the IBTrACS database (National Hurricane Center 2013). Since the GFS total precipitable water output data field is $1/2°$ resolution in both latitude and longitude, there is noticeable data skewing in projection near the poles. GFS data are blended short-term forecasts with the latest observations, through a DA process, and are produced every 6 h (NCEI 2019). Each file contains the best approximation of the state of the atmosphere at the 0-h time. Additional GFS forecasts are provided within the same file. The IBTrACS-GFS U-Net used input data from both the 0-h (analysis) and 3-h forecast within each GFS file. The 3-h forecast was treated like the following time step's 0-h analysis to provide more input training data. Therefore, there were total precipitable water outputs available at 0000, 0300, 0600, 0900, 1200, 1500, 1800, and 2100 UTC per day. There is a small difference in the total precipitable water field's forecast and actual observed state, resulting in a cyclone center label that might not be over the correct spot in the image. However, this difference in location is acceptable because it still remains within the bounded region for a "yes cyclone." Using the 3-h GFS forecast is appropriate for slower-moving, longer-living cyclone systems because the additional difference in storm center location is also within the bounding region.

There are many hyperparameters and metrics that can be tuned to optimize a U-Net, and it is not feasible to tune every single one for every possible value. Those addressed in section 2, the loss function, activation function, number of epochs, using dropout or noise, U-Net depth, and pixel bounding size, were manually changed on the basis of experiments to derive the models that performed best. Refining even each of these listed parameters required training of multiples of U-Nets and in turn, extensive computational time, and were compared to previous U-Net model performances. It was determined that model converged on a best-performing model when the accuracy, Tversky, Dice, and loss values reached their best values with minimal change based on the parameter tuning. Row 1 in Table 2 shows the combination of tuning parameters that produced the best-performing IBTrACS-GFS U-Net model (the subsequent rows show the best combinations of tuning parameters for U-Nets that used other combinations of labeled data and input data). The pixel bounding region was 25×25 pixels from the GFS (approximately 300 km square) and the input image size was the full GFS resolution of 720×361. The Tversky loss was the best-performing loss function for this U-Net and the model structure was six-block layers deep. There were 8622 training samples over the 4-yr period, and it took the U-Net roughly 36 min to train on GPUs over the course of 37 epochs. Once the IBTrACS-GFS U-Net was trained, it quickly identified tropical cyclone ROI from a test image in roughly 0.03 s. Table 3 shows the statistical testing performance comparing the truth, based on the labeled inputs, with the ROI produced by U-Net. The IBTrACS-GFS U-Net (row two) achieved an accuracy of 99% and high Dice and Tversky coefficients of 0.75–0.76 (Table 3, row 1).

In Fig. 3, there are four images that progress forward 3 h in time each panel. Within each panel, the truth-labeled image is above the IBTrACS-GFS U-Net model ROI labeled image. The IBTrACS-GFS U-Net tended to identify regions as boxes and this is likely due to the square areas of interest that trained the model. The model captures nearly all tropical cyclones within all ocean basins and has no issues with detection near the edges of the domain. Since these cyclones are circular, very bright (wet) in their signature, and fairly uniform in appearance, the model rarely identifies a region that does not appear somewhat as a tropical cyclone. Both well-developed tropical cyclones with identifiable eye features as well as weaker or newly forming/dissipating tropical cyclones are detected by the U-Net. The U-Net does not always detect ROI consistently in time. There are instances when an ROI would go away prematurely for a time step and then might come back and continue detection. Other times, the truth label terminated prematurely, which might explain these instabilities in how the model learned the behavior. One hypothesis for this behavior is if the IBTrACS labels start late or terminate early because of the applied wind threshold, then the U-Net has shown to learn weaker tropical cyclone structure and can continue to track the storm. This is because the U-Net was trained without the wind threshold knowledge explicitly and likely learns more from cyclone structure. Another hypothesis is that if a tropical cyclone weakens and its structure breaks up, then it might still have the wind speed required to keep the truth label but be more ambiguous of an object for the U-Net to detect at each time step. With these additional tropical cyclone ROI labels, this is beneficial for early detection of storms that might become tropical cyclones or regions of cyclonic potential.

Figure 3 is plotted to show the total likelihood of cyclone regions. Lower-confidence areas are those with a numerical value closer to 0 and are indicated by the lighter-blue-colored boxes. Higher-confidence values are those closer to 1 and are

shown in more red-colored boxes. Occasionally, very bright signatures in the total precipitable water field that are non-tropical cyclone regions along the intertropical convergence zone (ITCZ) are labeled as ROI by the U-Net. While rarely these ROI instances had a high confidence (red), the majority of these cases were labeled with a lower confidence (blue) and could be eliminated if a threshold were applied to the U-Net ROI labels. This figure shows the U-Net model tropical cyclone ROI that are labeled outside of the tropics. These regions have a lower threshold confidence but since the model was not trained to be constrained in latitude, it has the potential to track tropical cyclones as through their transition as they become extratropical.

4. Variances of U-Net model for extratropical cyclone detection

On the basis of the success of the U-Net model in section 3, the same approach to investigate the DL detection of tropical cyclones was extended to the detection of extratropical cyclones with either GFS total precipitable water data as inputs or GOES water vapor satellite data as inputs. The extratropical cyclone data labels were collected from prior work by Bonfanti et al. (2018). The three new U-Net variant models created will be referred to as Heuristics-GFS U-Net, IBTrACS-GOES U-Net, and Heuristics-GOES U-Net. In addition, other fine-tuning components in the U-Net, such as model depth or the loss functions, were modified to optimize each individual U-Net's performance. The following three sections describe each U-Net model.

a. Heuristic cyclone labels with GFS inputs

Defining extratropical cyclones is a harder task than identifying tropical cyclones due to the diversity in water vapor signatures and varying definitions for extratropical cyclones. To avoid a time-consuming and subjective process of hand labeling extratropical cyclones, the table outputs from the heuristic cyclone labeler by Bonfanti et al. (2018) was used to provide the truth labels. That heuristic model was designed for machine-learning applications and did not label cyclones in polar regions. The work identified storms over landmasses, providing the U-Net model with more training examples than other heuristic model options. However, a heuristic model introduces its own set of biases and missed cyclones and heuristics datasets are not fully inclusive of all events. Given the nature of ambiguity in the definition of extratropical cyclones, there are discrepancies between truth-labeled datasets on correctly labeled cyclone storms. This made it difficult to evaluate the numerical performance of the U-Net. This is not directly quantifiable and therefore difficult to extract from numerical results, but differences between truth and U-Net labeled ROI will lower the coefficients and accuracy. For instance, in training, the model might learn that an extratropical cyclone is not a cyclone because the truth label is "no cyclone." Therefore, the U-Net will identify similar cyclone events as a "not cyclone" when it could be a "yes cyclone" depending had a different truth definition been selected. This problem can be flipped, meaning that a U-Net learns to label certain events

TABLE 2. A summary of the architecture of the four selected U-Net models, with the specifics of each model. Details include: "truth" label source, input pixel size, number of input images for training and validation, number of images per batch, model activation and loss function, number of epochs to convergence, time for training, and time for inference both on a single input and for a month of inputs.

Model labels	Model input	Input image size	Training size	Validation size	Batch size	ROI pixel size	Dropout or noise	Loss function	U-Net model depth	Epochs used	Training time (s)	Inference time per single run (s)	Inference time per month of runs (s)
IBTrACS	GFS	720 × 361	8622	3020	256	25	Noise 0.2	Focal	6	37	2130	0.03	8.16
Heuristic	GFS	720 × 361	15574	2902	1520	30	Dropout 0.1	Dice	5	200	2200	0.03	6.48
IBTrACS	GOES	1024 × 560	5638	2214	128	25	Dropout 0.2	BCE	5	70	3540	0.15	36
Heuristic	GOES	1024 × 560	25288	2735	720	60	Dropout 0.1	Tversky	4	150	4650	0.06	14.4

TABLE 3. Results from the best four U-Net models with either IBTrACS or Heuristic "truth" labels for both GFS and GOES image inputs. Model performance is measured by Dice and Tversky coefficients as well as accuracy.

Model labels	Model input	Accuracy	Loss score	Dice coef	Tversky coef	Optimizer	Dropout or noise	Batch normalization
IBTrACS	GFS	0.991	0.237	0.763	0.75	RMS 0.00008	Noise 0.2	Yes
Heuristic	GFS	0.807	0.351	0.58	0.649	RMS 0.00001	Dropout 0.1	Yes
IBTrACS	GOES	0.996	0.311	0.689	0.68	RMS 0.0001	Dropout 0.2	Yes
Heuristic	GOES	0.901	0.442	0.511	0.558	RMS 0.00001	Dropout 0.1	Yes

as cyclones when, had a different truth-label set been picked, it would not have been a cyclone. These issues are addressed in the qualitative analysis. Given those challenges, the U-Net trained from these labels still identify ROI that are interesting, valuable, and potentially correct even if the truth label is missing. The goal of this research was to identify regions of cyclonic interest, and therefore regions that are missed in a truth label but identified in the U-Net are of good value.

The Heuristic-GFS U-Net performed slightly poorer than the IBTrACS-GFS U-Net model because of the inclusion of the extratropical cyclones. Table 2 shows that both the loss function and depth that produced the optimal Heuristic-GFS U-Net were different than the IBTrACS-GFS U-Net, having a depth of only 5 and using the Dice loss. The shallower the U-Net, the fewer hyperparameters the model has to tune. One potential reason that the Heuristic-GFS U-Net has a shallower optimized U-Net is that it has a larger training size. This may mean that fewer hyperparameters are needed to converge on the better performing model. One of the biggest differences between the IBTrACS-GFS U-Net and this Heuristic-GFS U-Net was the bounding-box size, which was increased from 25×25 to 30×30. Since extratropical cyclones are bigger than tropical cyclones, increasing the size from 25 to 30 improved the U-Net performance by encompassing the larger areas of extratropical cyclones. Row 2 of Table 3 shows that the accuracy of this model remains high, at 80%, but the Dice and Tversky coefficients were lower than the IBTrACS-GFS U-Net at 0.5 and 0.6, respectively.

This U-Net model has a fast inference stage on GPU, identifying extratropical and tropical cyclone ROI from an input in 0.03 s. It identified these ROI much faster than the heuristic model used to create the data labels on the same input source. The inference time for the heuristic model (used to create the labels) for a month's worth of cyclone ROI took 18.67 s, whereas the Heuristic-GFS U-Net ran in 6.48 s. This shows that the U-Net model identified cyclone ROI 2.88 times as fast as the heuristic method.

The interest in identifying extratropical cyclone ROI extends beyond matching truth labels to U-Net labels because there is much value, such as early detection, in the regions uniquely labeled by the U-Net. To better understand the behavior of which types of cyclones were identified as ROI in the Heuristic-GFS U-Net, a qualitative analysis was completed. This was done on images like Fig. 4 and compared the labeled truth-labeled GFS data (top) with the U-Net identified cyclone ROI (bottom). Figure 4 has a threshold applied to the U-Net

ROI of 70% confidence, meaning that ROI are only colored if there is a chance of at least 70% cyclone likely. This was an arbitrary threshold selection. Plots could have no threshold, such as Fig. 3, or could be plotted with a higher or lower threshold value, depending on the desired output for cyclone labels. A lower threshold corresponds to a less selective cutoff value for labeled cyclone pixels and vis versa. The figure shows areas of missed ROI detection as well as an area of false ROI detection. An explanation for this behavior is that GFS total precipitable water output data alone might not provide enough information to the U-Net in the high-latitude regions to distinguish between different types of extratropical cyclones as well as noncyclone events. A different reason is that the extratropical cyclones heuristic truth labels might have incorrectly missed a yes cyclone label for certain cyclones due to the heuristic set of rules. This would incorrectly train the U-Net to miss extratropical cyclones by assigning an error score to what should have been a correctly identified cyclone but instead the truth label said it was not a cyclone. This case is seen in Fig. 4 with the missed identification of a tropical cyclone in the heuristic truth label but where the U-Net had labeled it as an ROI. This makes quantitative analysis of U-Net performance alone incomplete. The qualitative analysis proves that the U-Net provided an important label on ROI that was missed by a different method.

On the basis of the analysis concluded by IMILAST, there is a climatological pattern of more frequent labels in the Southern Hemisphere than the Northern Hemisphere. The Southern Ocean has more extratropical cyclones and the Heuristic-GFS U-Net identifies the more ambiguous cyclone events that occur in that ocean basin because of that learned trait. Since there is no distinction between tropical and extratropical cyclones, this model also will track cyclones from tropics through the extratropics.

b. IBTrACS tropical cyclones labels with GOES inputs

This U-Net model uses the IBTrACS labels with GOES water vapor imagery. The most recent GOES satellite imagery ranges from 75° to 135°W longitude, so unlike the previous two U-Net models, this does not cover the whole globe (Jenner 2015). Only the water vapor channel (6.48 μm), with a resolution of 4–8 km, was used as input. Tropical cyclone signatures are detectable by the trained human eye in water vapor from a single time step or in a series of sequential images as distinct, bright, small circles. If the cyclone is well developed, it sometimes has the prominent "eye" feature. That signature persists at any time or lighting condition during an Earth day, making this IBTrACS-GOES U-Net
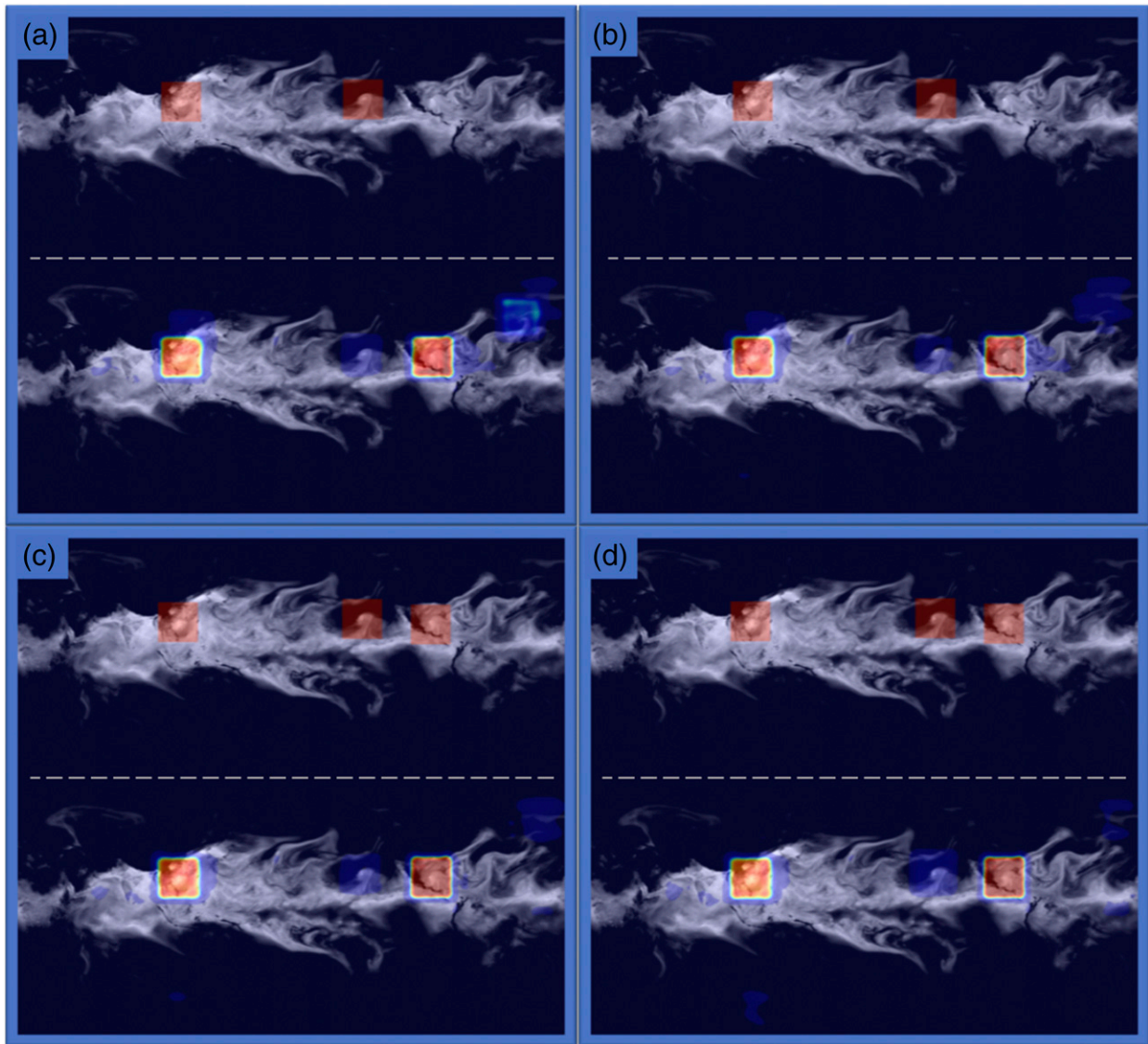
FIG. 3. GFS total precipitable water image on 23 Aug 2017. Time progresses in 3-h increments from (a) 1200 UTC, to (b) 1500 UTC, to (c) 1800 UTC, to (d) 2100 UTC. Images above the dotted line are the IBTrACS-labeled cyclone ROI used as "truth" with solid red boxes to indicate the labeled cyclone region. The corresponding images directly below the dotted line are the U-Net labeled ROI. In the U-Net results, the model identified a potential ROI if the area in the image is shaded. The shading ranges from dull, almost invisible, blue to a bright red color to indicate the ROI confidence value. Those values closer to 1 are in bright red.

model useful in identifying potential or current tropical cyclones in real-time applications.

Table 2 row 3 shows there are 5638 test samples for the IBTrACS-GOES U-Net and that the U-Net is five layers deep. The BCE loss function gave the best-performing model for these inputs and labels. Since GOES satellite imagery is very large and with the complexity of the U-Net model structure, the data had to be resized to 1024×512 to fit in GPU memory. In addition to help with memory, the batch size was decreased to fit within constraints. Similar to the IBTrACS-GFS U-Net, the IBTrACS-GOES U-Net had a high accuracy of 99%. Table 3 shows a lower Dice and Tversky coefficient of about

0.7, indicating a slightly poorer ROI detection performance for GOES data inputs than GFS.

This U-Net might have had a harder time identifying ROI because of scanning gaps or projection skewing in the satellite imagery because the angle of observation of Earth is more oblique as it approaches the outer edges of the imagery. There were small horizontal data gaps in some of the data inputs due to satellite scanning issues. Figure 5 indicates the potential for errors near the boundaries of the satellite images due to curvature and skewing of the data. Despite skewing, the IBTrACS-GOES U-Net performed generally well, but when it missed ROI, it missed them near the boundaries more than
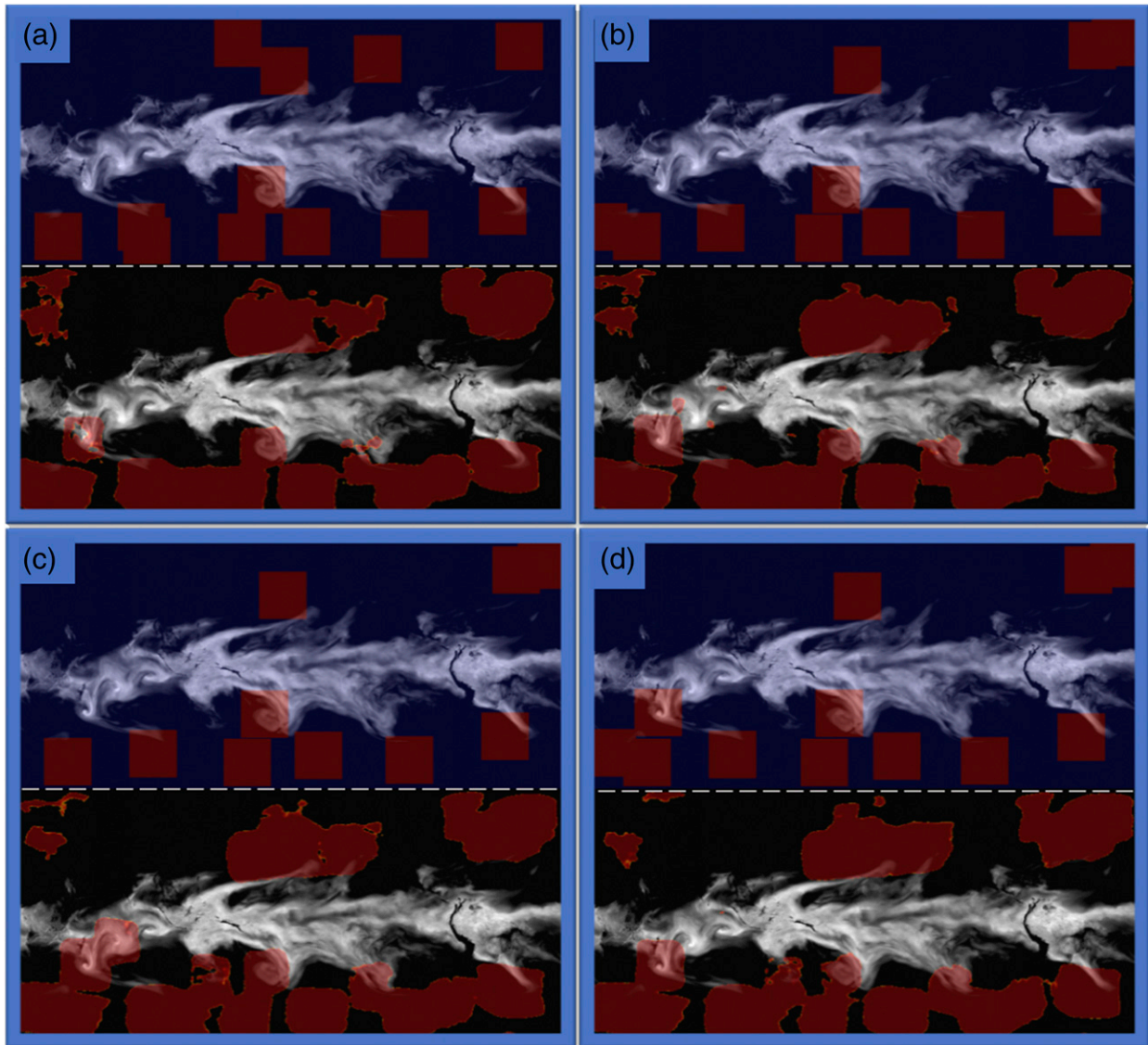
FIG. 4. GFS total precipitable water image on 10 Mar 2017. Time progresses in 3-h increments from (a) 1500 UTC, to (b) 1800 UTC, to (c) 2100 UTC, to (d) 0000 UTC 11 Mar. Images above the dotted line are the heuristics-labeled cyclone ROI used as "truth" with solid red boxes to indicate the labeled cyclone region. The corresponding images directly below the dotted line are the U-Net labeled ROI. The U-Net model output shown here has a confidence threshold applied of 70%, meaning that all red-segmented regions have a value of at least 0.7 and indicate a higher confidence of cyclone ROI event.

for storms centered in the image. The greatest impact on the U-Net results may have been the smaller sample size available for training. The IBTrACS-GOES U-Net had 3000 fewer training samples largely because of the smaller coverage area.

Relying on qualitative metrics alone is not sufficient. The broader definition of cyclones in the data suggests early detection of cyclones that might not yet be identified in the IBTrACS database or with heuristic truth labels. Identification of such cases would be beneficial for real-time alert systems but are scored negatively in the quantitative estimates. For example, not shown in the figure are instances when the U-Net incorrectly identified a noncyclone weak tropical storm or area

of tropical convection in the ITCZ. This is accurately scored in the numerical quantitative results. However, Fig. 5 (plotted with a 70% yes-cyclone pixel confidence threshold) indicates that there are instances of early detection as well as false positives. The U-net is said to provide early detection when the ROI segmentation appears in images in time over a region that the truth label later identifies. Figures 5a–c show that the U-net identified an ROI before the truth label appeared in the final Fig. 5d. The IBTrACS-GOES U-Net correctly identifies the two truth-labeled ROI cyclones. It further detects an additional region in the tropics that has a very similar signature to a hurricane that persists through all four time frames. The U-Net
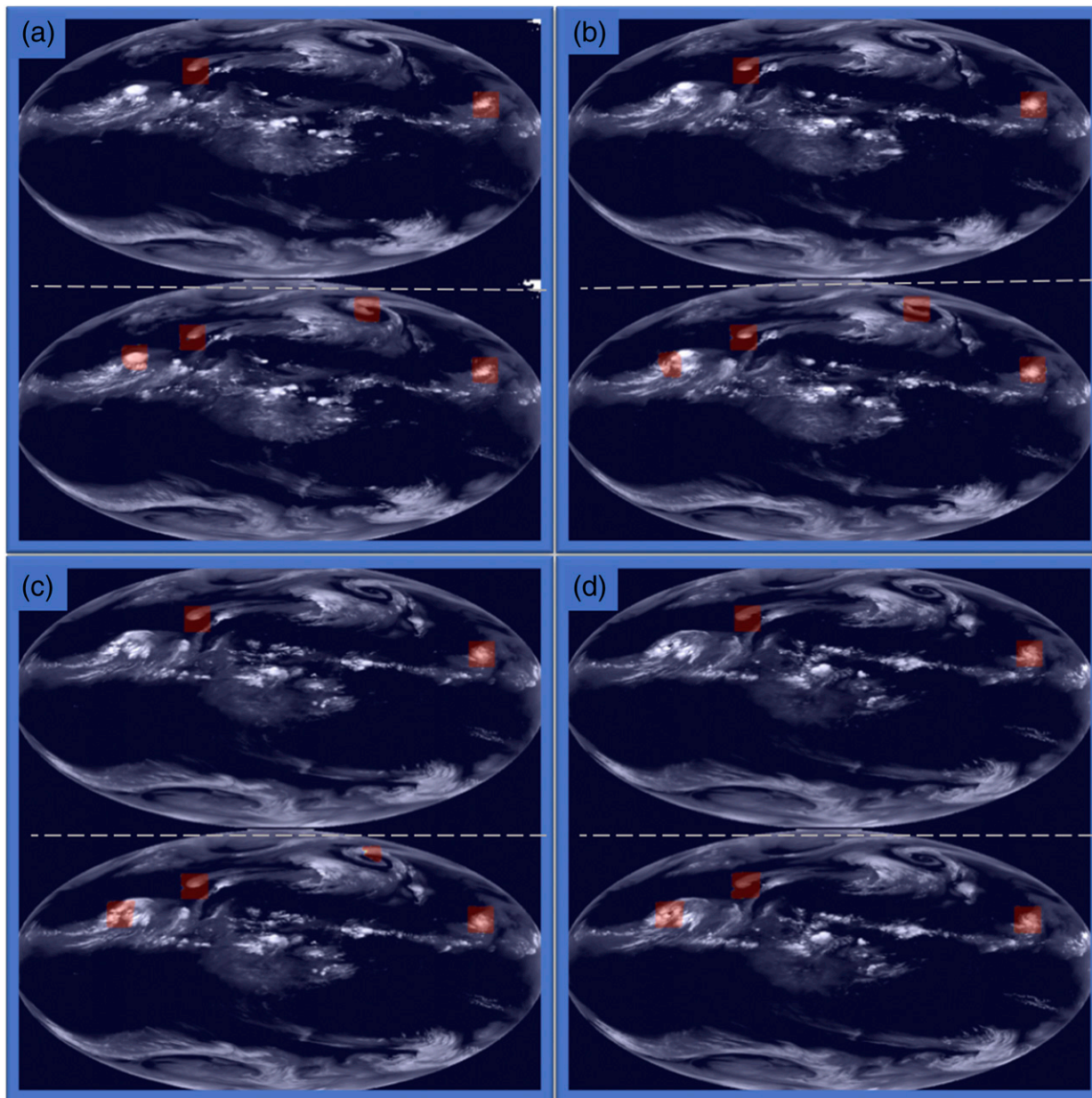
FIG. 5. *GOES-13* image on 30 Aug 2017. Time progresses in 3-h increments from (a) 0900 UTC, to (b) 1200 UTC, to (c) 1500 UTC, to (d) 1800 UTC. Images above the dotted line are the IBTrACS-labeled cyclone ROI used as "truth" with solid red boxes to indicate the labeled cyclone region. The corresponding images directly below the dotted line are the U-Net labeled ROI. The U-Net model output shown here has a confidence threshold applied of 70%, meaning that all red-segmented regions have a value of at least 0.7 and indicate a higher confidence of a cyclone ROI event.

additionally identifies a cyclone ROI transitioning into the extratropics.

c. Heuristic cyclone labels with GOES inputs

The Heuristic-GOES U-Net used the heuristic labels from Bonfanti et al. (2018) and inputs from the GOES water vapor data. This U-Net had the most training samples with 25 288 and the largest bounding box (60 pixels). This larger bounding box

is the biggest difference between this U-Net model and the other models listed in Table 2. Extratropical cyclones consume more pixel area and therefore warrant a larger bounding size. The best configuration of the model was the shallowest U-Net, with four layers giving the optimal model, and used the Tversky loss function. One guess as to why the shallower U-Net model gave the optimized model for these labels and input source is because it generalized better to differences in satellite extratropical cyclone
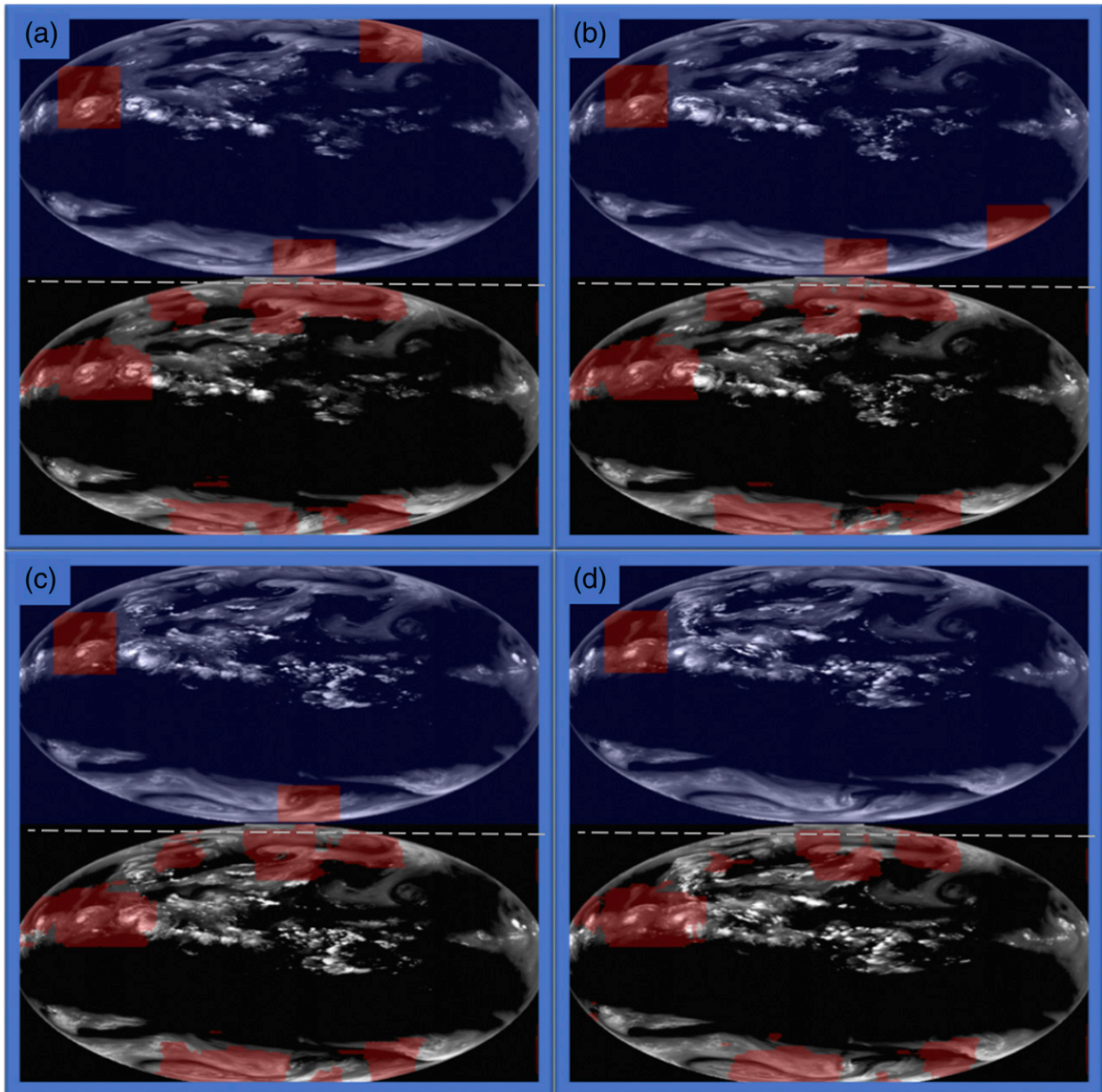
FIG. 6. *GOES-13* image on 24 Jul 2017. Time progresses in 3-h increments from (a) 1500 UTC, to (b) 1800 UTC, to (c) 2100 UTC, to (d) 0000 UTC 25 Jul. Images above the dotted line are the heuristics-labeled cyclone ROI used as "truth" with solid red boxes to indicate the labeled cyclone region. The corresponding images directly below the dotted line are the U-Net labeled ROI. The U-Net model output shown here has a confidence threshold applied of 70%, meaning that all red-segmented regions have a value of at least 0.7 and indicate a higher confidence of a cyclone ROI event.

appearance. The Heuristic-GOES U-Net also had an ambiguous interpretation of both quantitative and qualitative results similar to the Heuristic-GFS U-Net due to both a broader definition of cyclones and early detection of cyclones in the GOES input data. The quantitative results can be misleading. Table 3 shows that the Heuristic-GOES U-Net had a higher accuracy at 90% than its GFS counterpart, but that it had the lowest Dice and Tversky coefficient values at about 0.5. While that value is lower than the other models, visually there is

still much agreement between the truth labels and Heuristic-GOES U-Net identified ROI.

Similar to the Heuristic-GFS U-Net, the Heuristic-GOES U-Net had a pattern of more frequent cyclone ROI detections in the Southern Hemisphere than the Northern Hemisphere. Figure 6 compares truth labels from the heuristic model with the Heuristic-GOES U-Net detected ROI. The plots have a threshold confidence of at least 70% chance that the pixel contains a cyclone. The U-Net had a tendency to correctly
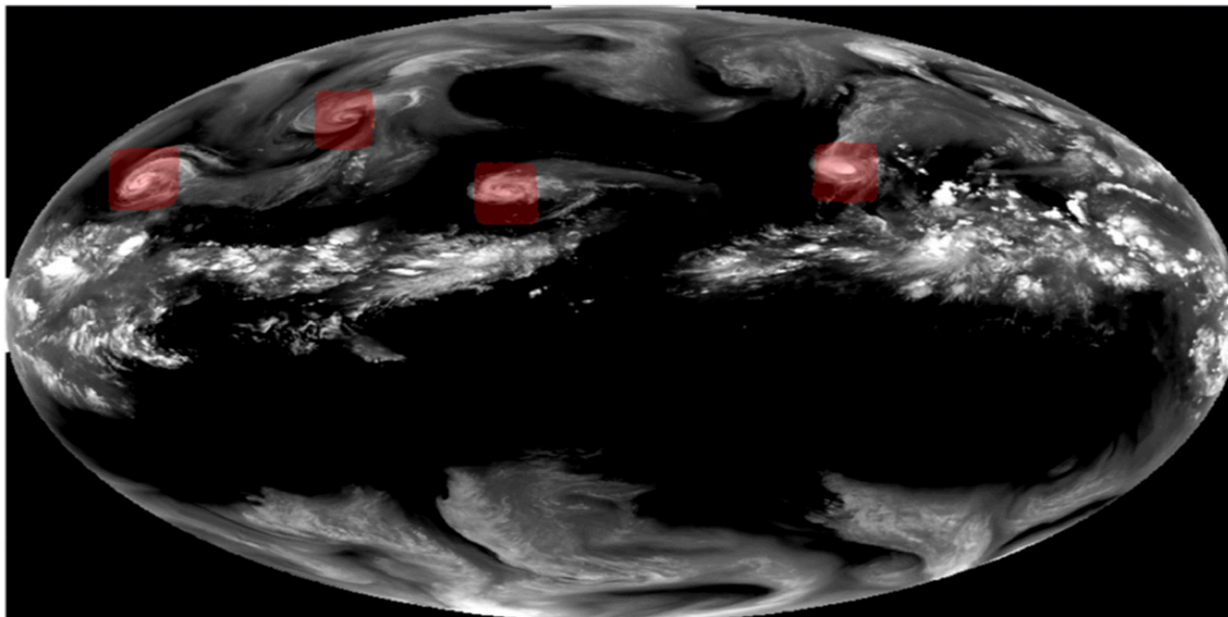
FIG. 7. Output from the U-Net trained with IBTrACS tropical cyclone labels on GOES water vapor brightness temperature satellite imagery. The regions in red are U-Net identified ROI from 0000 UTC 5 Sep 2015. The U-Net model output shown here has a confidence threshold applied of 70%, meaning that all red-segmented regions have a value of at least 0.7 and indicate a higher confidence of a tropical cyclone ROI event.

label fewer ROI along the boundary edges than in other regions in the satellite image and had occasional noise in the detection along the boundaries. There was a higher quantity of Heuristic-GOES U-Net detected ROI than truth labels in general. This figure was selected to show an example of when the Heuristic-GOES U-Net correctly labeled all three tropical cyclones while the truth labels had not yet identified them all. False detection by the U-Net occurred when it incorrectly labeled a very small ROI in the ITCZ as a cyclone when they are not. Similar cases were observed in the other U-Nets and it has been discussed how this behavior degrades the quantitative performance of the model. It was more difficult to qualitatively analyze extratropical cyclone U-Net detected ROI because of the diversity in extratropical cyclone appearance in the water vapor data.

Given the variety of types and appearances in the water vapor channel of extratropical cyclones, the Heuristic-GOES U-Net model impressively identified diverse looking patterns for extratropical cyclones at the same time as detecting brighter, smaller tropical cyclones. The most common extratropical ROIs detected in the U-Net have clear comma-shaped wet–dry signatures, indicating rotation. This is expected since it is also an easier signature for humans to identify as well and indicates a well-developed cyclone. In general, there were more false labels and noise in the Heuristic-GOES U-Net than in both IBTrACS U-Nets, but as is the case in all other models, almost all noncyclone events for the Heuristic-GOES U-Net remained correctly unlabeled. This indicates that the model can be used for fast, early cyclone detection as well as detection of ROI.

5. Summary and discussion

Four individual U-Net models were created to detect cyclone ROI from two different data sources and two different labeling sources: IBTrACS-GFS, Heuristic-GFS, IBTrACS-GOES, and Heuristic-GOES. A multi-GPU system was used to significantly decrease the data processing and U-Net model training times. Once trained, the inference time for the Heuristic-GFS U-Net ran 3 times as fast as the heuristic model used to create the labels from the same GFS input. This comparison shows that deep-learning models can extract cyclone location information from GFS data faster than heuristic methods can. All of the models achieved a relatively high level of accuracy, ranging from 81% (Heuristic-GFS) to 99% (both IBTrACS U-Nets). IoU metrics were also used for evaluation. The Tversky coefficients for the Heuristic-GOES U-Net, Heuristic-IBTrACS U-Net, IBTrACS-GOES U-Net, and IBTrACS-GFS U-Net models, respectively, were 0.558, 0.649, 0.680, and 0.750. These results show the U-Nets were optimized without overfitting and gave good results for diverse cyclone event detection. Further improvements to the U-Nets that could be explored in the future include newer activation functions, types of convolutions, increased data input. In particular, having additional satellite channel inputs or adding sensor information may further improve the satellite U-Nets.

The performance of these U-Net models proves that for image-segmentation tasks in meteorology and climatology related fields, ML and DL models provide unique and faster alternatives to existing methods. Aside from numerical metrics, the U-Nets identified unique ROI that were not included within the truth-labeled dataset. The U-Net models have

versatile applications, such as cyclone ROI extraction in large datasets of high resolution when it is expensive to hand label extreme events of interest or impractical to identify ROI in real-time applications. These models show that, when compared with the truth labels, they identify a high number of the same ROI area as heuristic or hand-labeling methods, but also uniquely identify additional regions that are missed by more traditional methods. This addresses the issue that labeled datasets are not always inclusive of all events and will impact how the U-Nets are evaluated numerically. It shows the benefit of integrating DL methods for diversifying an event database and for real-time applications, such as actively identifying ROI for potential cyclogenesis. Likewise, U-Nets provide fast detection of regions with high-impact weather in a time-sensitive scenario. They label the location information for areas with active or high-potential weather that benefits numerous weather products, such as aviation warnings and DA. Figure 7 highlights the tropical cyclone information in GOES water vapor data extracted from the IBTrACS-GOES U-Net as an example of what can be output at subsecond speeds to quickly locate ROI. Such information provides more data points for better understanding and forecasting the cyclone system.

The U-Net architecture used in these models is extensible to training new models for the detection and classification of other types of weather events. U-Nets can identify patterns of noncyclone ROI, such as convection or convection initiation. The application of the U-Net can produce real-time information on active or high-potential weather locational information. It can scale to climatological settings by providing information on past weather events in climate data. These types of DL models are underutilized. They show increasing success for ROI detection and their performance encourages the development of more DL models for future ROI detection schemes in a multitude of weather and climate applications.

Data availability statement. All code and data are publicly available through request to NOAA's Global Systems Laboratory.

REFERENCES

American Meteorological Society, 2019: Extratropical cyclone. Glossary of Meteorology, http://glossary.ametsoc.org/wiki/Extratropical_cyclone.

Atlantic Oceanographic and Meteorological Laboratory, 2020: What is an extra-tropical cyclone? NOAA, accessed 4 May 2020, http://www.aoml.noaa.gov/hrd/tcfaq/A7.html.

Bahrampour, S., N. Ramakrishnan, L. Schott, and M. Shah, 2015: Comparative study of deep learning software frameworks. arXiv 1511.06435, 9 pp., https://arxiv.org/pdf/1511.06435.pdf.

Bonfanti, C., L. Trailovic, J. Stewart, and M. Govett, 2018: Machine learning: Defining Worldwide cyclone labels for training. *2018 21st Int. Conf. on Information Fusion (FUSION)*, Cambridge, United Kingdom, IEEE, 753–760, https://doi.org/10.23919/icif.2018.8455276.

Cheng, K.-S., J.-S. Lin, and C.-W. Mao, 1996: The application of competitive Hopfield neural network to medical image segmentation. *IEEE Trans. Med. Imaging*, **15**, 560–567, https://doi.org/10.1109/42.511759.

Cichoki, A., and R. Unbehauen, 1996: *Neural Networks for Optimization and Signal Processing.* John Wiley and Sons, 544 pp.

Ciresan, D., U. Meier, and J. Schmidhuber, 2012: Multi-column deep neural networks for image classification. *2012 IEEE Conf. on Computer Vision and Pattern Recognition*, Providence, RI, IEEE, 3642–3649, https://doi.org/10.1109/cvpr.2012.6248110.

Deng, J., W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, 2009: ImageNet: A large-scale hierarchical image database. *2009 IEEE Conf. on Computer Vision and Pattern Recognition*, Miami, FL, IEEE, 248–255, https://doi.org/10.1109/cvpr.2009.5206848.

Deng, L., 2012: The MNIST database of handwritten digit images for machine learning research. *IEEE Signal Process. Mag.*, **29**, 141–142, https://doi.org/10.1109/MSP.2012.2211477.

Dia, H., 2001: An object-oriented neural network approach to short-term traffic forecasting. *Eur. J. Oper. Res.*, **131**, 253–261, https://doi.org/10.1016/S0377-2217(00)00125-9.

Gagne, D. J., A. Mcgovern, S. E. Haupt, R. A. Sobash, J. K. Williams, and M. Xue, 2017: Storm-based probabilistic hail forecasting with machine learning applied to convection-allowing ensembles. *Wea. Forecasting*, **32**, 1819–1840, https://doi.org/10.1175/WAF-D-17-0010.1.

Giacinto, G., and F. Roli, 2001: Design of effective neural network ensembles for image classification purposes. *Image Vis. Comput.*, **19**, 699–707, https://doi.org/10.1016/S0262-8856(01)00045-2.

Giffard-Roisin, S., M. Yang, G. Charpiat, C. K. Bonfanti, B. Kégl, and C. Monteleoni, 2020: Tropical cyclone track forecasting using fused deep learning from aligned reanalysis data. *Front. Big Data*, **3**, 1, https://doi.org/10.3389/fdata.2020.00001.

Glorot, X., A. Bordes, and Y. Bengio, 2011: Deep sparse rectifier neural networks. *14th Int. Conf. on Artificial Intelligence and Statistics (AISTATS)2011*, Fort Lauderdale, FL, JMLR, 315–323.

Gulli, A., and S. Pal, 2017: *Deep Learning with Keras: Implement Neural Networks with Keras on Theano and Tensorflow.* Packt Publishing, 318 pp.

He, K., X. Zhang, S. Ren, and J. Sun, 2016: Deep residual learning for image recognition. *2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, IEEE, 770–778, https://doi.org/10.1109/cvpr.2016.90.

Holland, G. J., 1993: Ready reckoner. Global Guide to Tropical Cyclone Forecasting, World Meteorological Organization Doc. WMO/TC-560, Rep. TCP-31, 9.1–9.32, https://library.wmo.int/doc_num.php?explnum_id=9598.

Jenner, L., 2015: GOES overview and history. NASA, accessed 25 September 2018, http://www.nasa.gov/content/goes-overview/index.html.

Jergensen, G. E., A. McGovern, R. Lagerquist, and T. Smith, 2020: Classifying convective storms using machine learning. *Wea. Forecasting*, **35**, 537–559, https://doi.org/10.1175/WAF-D-19-0170.1.

Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann, 2010: The International Best Track Archive for Climate Stewardship (IBTrACS). *Bull. Amer. Meteor. Soc.*, **91**, 363–376, https://doi.org/10.1175/2009BAMS2755.1.

Krizhevsky, A., I. Sutskever, and G. E. Hinton, 2017: ImageNet classification with deep convolutional neural networks. *Commun. ACM*, **60**, 84–90, https://doi.org/10.1145/3065386.

Lagerquist, R., A. McGovern, C. R. Homeyer, D. J. Gagne II, and T. Smith, 2020: Deep learning on three-dimensional multiscale data for next-hour tornado prediction. *Mon. Wea. Rev.*, **148**, 2837–2861, https://doi.org/10.1175/MWR-D-19-0372.1.

Lawrence, S., C. Giles, A. C. Tsoi, and A. Back, 1997: Face recognition: A convolutional neural-network approach. *IEEE Trans. Neural Networks.*, **8**, 98–113, https://doi.org/10.1109/72.554195.

LeCun, Y., and Y. Bengio, 1995: Convolutional networks for images, speech, and time series. *The Handbook of Brain Theory and Neural Networks*, MIT Press, 276–279.

——, ——, and G. Hinton, 2015: Deep learning. *Nature*, **521**, 436–444, https://doi.org/10.1038/nature14539.

Lee, Y., G. Wahba, and S. A. Ackerman, 2004: Cloud classification of satellite radiance data by multicategory support vector machines. *J. Atmos. Oceanic Technol.*, **21**, 159–169, https://doi.org/10.1175/1520-0426(2004)021<0159:CCOSRD>2.0.CO;2.

Liang, M., and X. Hu, 2015: Recurrent convolutional neural network for object recognition. *2015 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, IEEE, 3367–3375, https://doi.org/10.1109/cvpr.2015.7298958.

Lin, T.-Y., P. Goyal, R. Girshick, K. He, and P. Dollar, 2017: Focal loss for dense object detection. *2017 IEEE Int. Conf. on Computer Vision (ICCV)*, Venice, Italy, IEEE, 2999–3007, https://doi.org/10.1109/iccv.2017.324.

Liu, S., and W. Deng, 2015: Very deep convolutional neural network based image classification using small training sample size. *2015 Third IAPR Asian Conf. on Pattern Recognition (ACPR)*, Kuala Lumpur, Malaysia, IEEE, 730–734, https://doi.org/10.1109/acpr.2015.7486599.

Long, J., E. Shelhamer, and T. Darrell, 2015: Fully convolutional networks for semantic segmentation. *2015 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, IEEE, 3431–3440, https://doi.org/10.1109/cvpr.2015.7298965.

McGovern, A., K. L. Elmore, D. J. Gagne, S. E. Haupt, C. D. Karstens, R. Lagerquist, T. Smith, and J. K. Williams, 2017: Using artificial intelligence to improve real-time decision-making for high-impact weather. *Bull. Amer. Meteor. Soc.*, **98**, 2073–2090, https://doi.org/10.1175/BAMS-D-16-0123.1.

Merrill, R. T., 1993: Tropical cyclone structure. Global Guide to Tropical Cyclone Forecasting, World Meteorological Organization Doc. WMO/TC-560, Rep. TCP-31, 2.1–2.60, https://library.wmo.int/doc_num.php?explnum_id=9598.

National Hurricane Center, 2013: Saffir-Simpson Hurricane Wind Scale. NOAA, https://www.nhc.noaa.gov/aboutsshws.php.

NCEI, 2019: Global Forecast System (GFS). NCEP National Climatic Data Center, accessed 15 December 2019, http://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs.

Neu, U., and Coauthors, 2013: IMILAST: A community effort to intercompare extratropical cyclone detection and tracking algorithms. *Bull. Amer. Meteor. Soc.*, **94**, 529–547, https://doi.org/10.1175/BAMS-D-11-00154.1.

Patterson, J., and A. Gibson, 2017: Major architectures of deep networks. *Deep Learning: A Practitioner's Approach*. O'Reilly, 117–164.

Recknagel, F., M. French, P. Harkonen, and K.-I. Yabunaka, 1997: Artificial neural network approach for modelling and prediction of algal blooms. *Ecol. Modell.*, **96**, 11–28, https://doi.org/10.1016/S0304-3800(96)00049-X.

Ronneberger, O., 2017: Invited Talk: U-Net Convolutional Networks for biomedical image segmentation. *Bildverarbeitung für die Medizin 2017*, *Informatik aktuell*, Springer, 3, https://doi.org/10.1007/978-3-662-54345-0_3.

——, P. Fischer, and T. Brox, 2015: U-Net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, Lecture Notes in Computer Science, Springer, 234–241, https://doi.org/10.1007/978-3-319-24574-4_28.

Schmit, T. J., S. S. Lindstrom, J. J. Gerth, and M. M. Gunshor, 2018: Applications of the 16 spectral bands on the Advanced Baseline Imager (ABI). *J. Oper. Meteor.*, **06**, 33–46, https://doi.org/10.15191/nwajom.2018.0604.

Sergeev, A., and M. Del Balso, 2018: Horovod: Fast and easy distributed deep learning in TensorFlow. arXiv 1802.05799, 10 pp., https://arxiv.org/pdf/1802.05799.pdf.

Simonyan, K., and A. Zisserman, 2015: Very deep convolutional networks for large-scale image recognition. arXiv 1409.1556, 14 pp., https://arxiv.org/pdf/1409.1556.pdf.

Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, 2014: Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, **15**, 1929–1958.

Sudre, C. H., W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, 2017: Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, M. Cardoso et al., Eds., Lecture Notes in Computer Science, Springer, 240–248, https://doi.org/10.1007/978-3-319-67558-9_28.

Weingroff, M., 2014: How satellite observations impact NWP. UCAR, accessed 15 March 2019, http://kejian1.cmatc.cn/vod/comet/nwp/sat_nwp/print.php.htm.