

Physics-Augmented Deep Learning to Improve Tropical Cyclone Intensity and Size Estimation from Satellite Imagery

JING-YI ZHUO^{a,b} AND ZHE-MIN TAN^{a,b}

^a Key Laboratory of Mesoscale Severe Weather/Ministry of Education, Nanjing University, Nanjing, China

^b School of Atmospheric Sciences, Nanjing University, Nanjing, China

(Manuscript received 12 October 2020, in final form 7 April 2021)

ABSTRACT: A deep learning-based method augmented by prior knowledge of tropical cyclones (TCs), called DeepTCNet, is introduced to estimate TC intensity and wind radii from infrared (IR) imagery over the North Atlantic Ocean. While standard deep learning practices have many advantages over conventional analysis approaches and can produce reliable estimates of TCs, the data-driven models informed by machine-readable physical knowledge of TCs could achieve higher performance. To this end, two approaches are explored to develop the physics-augmented DeepTCNet: (i) infusing the auxiliary physical information of TCs into models for single-task learning and (ii) learning auxiliary physical tasks for multitask learning. More specifically, augmented by auxiliary information of TC fullness (a measure of the radial decay of the TC wind field), the DeepTCNet yields a 12% improvement in estimating TC intensity over the nonaugmented one. By learning TC wind radii and auxiliary TC intensity task simultaneously, the model's wind radii estimation skill is improved by 6% over only learning four wind radii tasks and by 9% over separately learning a single wind radii task. The evaluation results showed that the DeepTCNet is in-line with the Satellite Consensus technique (SATCON) but systematically outperforms the advanced Dvorak technique (ADT) at all intensity scales with an averaged 39% enhancement in TC intensity estimation. The DeepTCNet also surpasses the Multiplatform Tropical Cyclone Surface Wind Analysis technique (MTCSWA) with an average improvement of 32% in wind radii estimation.

KEYWORDS: Tropical cyclones; Satellite observations; Neural networks; Deep learning

1. Introduction

Estimating the intensity and wind radii of tropical cyclones (TCs) is the first step in the process of monitoring and forecasting these destructive systems. However, making accurate estimates of TC intensity and size has been a long-standing challenge in tropical meteorology, and is limited by the available observations and technologies (e.g., Landsea and Franklin 2013; Knaff and Sampson 2015). Besides, global TC warning centers routinely require multiple measures of TC characteristics. These include two measures of TC intensity, the maximum sustained surface wind (MSW) and minimum sea level pressure (MSLP), together with different TC size measures, generally provided by forecast centers as “wind radii,” including the gale-force (34 kt; $1 \text{ kt} = 0.51 \text{ m s}^{-1}$; R34), storm-force (50 kt; R50), and hurricane-force (64 kt; R64) wind radii in geographic quadrants (i.e., in the northeast, southeast, southwest, and northwest directions), and the radius of maximum wind (RMW). The production of such a large number of estimates in real time may be difficult and time-consuming (Sampson et al. 2018).

Operational TC intensity estimation still relies primarily on the Dvorak technique (Velden et al. 2006), which uses the cloud features in infrared (IR) and/or visible satellite images to assign TC intensity. As the complex TC dynamics have been comprehensively abstracted as cloud features, the Dvorak technique can provide relatively reliable intensity estimates, and has been one of the most successful satellite applications

for TC analysis over more than four decades (Velden et al. 2006). Nonetheless, this technique is subjective, so its estimation accuracy depends on the skills of analysts. To increase the objectivity and the automatization of IR-based intensity analysis, advanced versions of the Dvorak technique (e.g., Olander and Velden 2007; 2019) and many other algorithms have been proposed (e.g., Kossin et al. 2007; Ritchie et al. 2012; Fetanat et al. 2013). However, most of these algorithms have proved to be less reliable than the Dvorak technique, mainly because they are based on conventional algorithms, such as principal component analysis, which extract only very limited features from the satellite data. The situation is similar in other satellite-based TC intensity estimation methods (e.g., Demuth et al. 2004, 2006; Jiang et al. 2019). The method that has the best accuracy in TC intensity estimation at present is the Satellite Consensus technique (SATCON), which is a weighted consensus algorithm designed to optimize the strengths and minimize the weaknesses of multiple IR-based and microwave-based techniques (Velden and Herndon 2020). The typical qualitative uncertainty level of the Dvorak technique, ADT and SATCON is around 10 kt for intensity in wind (Knaff et al. 2010; Olander and Velden 2019; Velden and Herndon 2020).

The satellite-based estimation of TC wind radii is more difficult. There is no such widely applied technique as the Dvorak method for wind radii estimation, given that the physical relationship between the convection and the wind field is less clear (e.g., Lajoie and Walsh 2008). The outer-core wind radii (e.g., R34) are among the more tractable TC wind radii and the scatterometer is one of the best spaceborne instruments for estimating outer-core wind radii (e.g., Brennan et al. 2009).

Corresponding author: Zhe-Min Tan, zmtan@nju.edu.cn

DOI: 10.1175/MWR-D-20-0333.1

© 2021 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

Scatterometer measurements operating at Ku-band have an issue to be affected by heavy rain in TCs and saturate for wind speeds near 50 kt (Brennan et al. 2009), but the lower C-band frequencies are less sensitive to rain and saturate at slightly higher wind speeds, which is close to 70 kt (Stoffelen 1998; Stoffelen et al. 2017). The L-band radiometers, which do not suffer from signal saturation at high winds and are minimally affected by rain, are important tools to determining the R34, R50, and R64 in TCs (e.g., Reul et al. 2017). Other methods that show skill in deriving TC wind radii include IR representations (e.g., Mueller et al. 2006; Kossin et al. 2007; Knaff et al. 2016; Dolling et al. 2016), microwave sounder proxies (e.g., Demuth et al. 2006), multi-satellite-platform analysis (Knaff et al. 2011), and the reflectometry-based technique (Morris and Ruf 2017). The estimation of the radius of maximum wind, which poses difficulties for observation because of the turbulent nature of the inner core of a TC (e.g., Chavas et al. 2015), however, remains a challenge.

Errors in the best track intensity have been estimated to be ~10%–20%, and the errors for best track wind radii could be as high as 40%, depending on the quality and quantity of the available observations (Landsea and Franklin 2013; Knaff and Sampson 2015). However, accurate TC intensity and wind radii estimates are crucial to making forecasts and mitigate the losses due to TC-related hazards (e.g., Sampson et al. 2010; Torn and Snyder 2012; DeMaria et al. 2014; Bender et al. 2017), and translate to the postprocessed guidance such as wind speed probabilities (e.g., DeMaria et al. 2013). Therefore, more accurate estimations of TC intensity and wind radii are still badly needed. Moreover, with the recent advances in the satellite observations of TCs, updating the analysis technique, especially objective algorithms that can interpret complex TC dynamics from the satellite observations, is of vital importance.

Deep learning is a type of artificial intelligence algorithm that has revolutionized computer vision, language recognition, game strategy and many research fields (e.g., Hinton et al. 2012; LeCun et al. 2015; Silver et al. 2016; Krizhevsky et al. 2017). More recently, deep learning is also providing insights into atmospheric science with broad applications, which cover pattern detection, physical parameterization, and state prediction (e.g., Rasp et al. 2018; Ham et al. 2019; Reichstein et al. 2019). Recent works have also begun to apply deep learning to TC intensity estimation. Pradhan et al. (2018) is among the earliest applications of deep learning to estimate TC intensity from IR imagery. However, this work is somewhat incomplete because they did not use an independent dataset for evaluation. Chen et al. (2019) used a larger amount of dataset (global TC cases) than Pradhan et al. (2018) and utilized IR imagery and passive microwave-retrieved precipitation to train deep learning models. However, the optimal estimates of Chen et al. (2019) are not available in real time due to the intermittent microwave rain-rate data and postanalysis smoothing required. Wimmers et al. (2019) estimated TC intensity from passive microwave imagery, which has a major advantage in showing convective features of TCs. However, at present, there is no deep learning application for TC size estimation. Moreover, despite the challenges of deep learning, such as the high cost of data acquisition and making mistakes on what would appear

trivial to a human for lacking any general knowledge of the world, little work has successfully capitalized on physical or prior knowledge to improve deep learning for TC applications. The main target of this study is to harness deep learning to produce more accurate estimations of both TC intensity and wind radii from IR imagery. Moreover, we endeavor to augment the deep learning models by incorporating prior knowledge of TCs into the model.

The rest of this paper is organized as follows. In the next section, we present a description of the data and the deep learning backbone we selected to build our method. We then describe the approaches to augment deep learning models for intensity estimation and evaluate model performance in section 3. Section 4 provides the augmented deep learning for wind radii estimation as well as the evaluation. The discussion on our method and the summaries of the findings are presented in sections 5 and 6.

2. Data and method

a. Data

Necessary information is required to set up the input–output pairs to train the deep learning models. In this paper, the outputs/labels given to the models were the TC intensity (MSW and MSLP) and wind radii (R34, R50, R64, and RMW) from the postseason/final best track dataset. These data were obtained from the IBTrACS database (Knapp et al. 2010). The primary inputs of the models are the TC-centered IR images, which were from the Hurricane Satellite dataset (HURSAT-B1; 3-h interval; 8-km resolution; Knapp and Kossin 2007). As the HURSAT-B1 images are only available up to 2016, we used the IR images from the GridSat-B1 archive (Knapp et al. 2011) from then onward. GridSat-B1 has the same temporal and spatial resolution as the HURSAT-B1 but at a global scale. Geostationary satellite IR imagery was used for its real-time availability and high spatial and temporal resolution. We also use auxiliary storm information as inputs. These include the TC fullness (Guo and Tan 2017), location and motion of a storm. The TC fullness is a new concept of TC structure, which is defined as the normalized extent of the outer-core wind skirt, i.e., $(R34 - RMW)/R34$. The auxiliary information was obtained from the operational Tropical Cyclone Vitals Database (TC Vitals¹) used for initializing numerical weather prediction model guidance (e.g., Tallapragada et al. 2014; Bender et al. 2017). For the cases with invalid wind radii estimates from TC Vitals, the corresponding TC fullness is set as zero. Only North Atlantic TCs with a lifetime-maximum intensity of at least 34 kt (i.e., tropical storms) are considered. Extratropical systems and tropical waves were removed.

In the data preprocessing, sample-specific normalization was applied to IR images by subtracting the average and dividing

¹ In the Automated Tropical Cyclone Forecast System (Sampson and Schrader 2000), these are the CARQ entries in the aid deck (adeck). The TC Vitals has also been referred to as “the bogus” in past literature.

by the standard deviation. The IR images were then cropped only to include the center area. For intensity estimation, images of 58×58 pixels were used, which is found to work well and effectively in our practice. For size estimation, larger images of 156×156 pixels were used to include the outer-range cloud information, which is shown to be necessary to produce quality outer-core wind radii estimates. Given the R34, R50, and R64 are recorded for each quadrant (NE, SE, SW, and NW) in the IBTrACS database, the nonzero-azimuthal average was applied to get the wind radii labels. Note that the R34, R50, and R64 denote the azimuthally averaged wind radii data hereafter. As there is no guarantee that all of the wind radii labels from IBTrACS are physically consistent; for example, the R50 may be smaller than the R64, even though this arrangement is not physically possible. Therefore, wind radii labels that did not satisfy the relationship $R34 > R50 > R64$ and $R34 \geq \text{RMW}$, $R50 \geq \text{RMW}$, $R64 \geq \text{RMW}$ (Kossin et al. 2007) were removed.

The data were then split into three blocks (Table 1): the training set including the 2005–15 cases; the validation set consisting of 2016 and 2018 cases; and the test set made up of 2017 and 2019 cases. 2005, when the RMW in the IBTrACS became available, is chosen as the accordant start season for the training data. The MSW and MSLP labels from 1990 to 2004, which are available in IBTrACS, are used to test the influence of training data amounts and to build a final intensity estimation model (sections 3c and 3d). The best track fixes have aircraft reconnaissance observations within 3 h (recon-aided hereafter) in the test set, which have higher quality, therefore, are used as “ground truth” to evaluate the performance of the DeepTCNet in sections 3d and 4b.

b. Selection of the DeepTCNet's architecture

In this study, convolutional neural networks (CNNs) (LeCun et al. 2015) are selected as the deep learning backbone of DeepTCNet to estimate TC intensity and wind radii from IR imagery. CNNs are a particular class of artificial neural networks considered to be the state-of-the-art tool for analyzing image data (Simonyan and Zisserman 2014; Krizhevsky et al. 2017). CNNs assign importance to and transfer various aspects in the input image through stacks of convolutional, pooling, and fully connected layers. Overall, the whole network approximates a nonlinear differentiable function between the input and the outcome of interest. Taking advantage of the convolutional filters, the hierarchical cascade, and other specific designs in the architecture, CNNs are very good at capturing the spatial dependencies and extracting intricate patterns in an image. They are also useful in uncovering new properties in the data (LeCun et al. 2015). Moreover, they also have the great advantage of requiring little manual feature design/extraction compared to other conventional machine learning algorithms (LeCun et al. 2015; Goodfellow et al. 2016). CNNs now have growing applications in atmospheric science, for example, to detect various weather phenomena in observations and climate models (e.g., Liu et al. 2016; Lagerquist et al. 2019; McGovern et al. 2019), represent subgrid-scale processes (e.g., Bolton and Zanna 2019; Han et al. 2020) and improve the weather predictions and climate projections

TABLE 1. The data size of the labels of training, validation, and test datasets for 1-min maximum sustained surface winds (MSW), minimum sea level pressure (MSLP), radius of maximum wind (RMW), and radii of 64-, 50-, and 34-kt winds (R64, R50, and R34). The units for MSW and MSLP are kt and hPa, respectively. The units for RMW, R64, R50, and R34 are n mi (1 n mi \approx 1.85 km).

TC parameter	Training dataset: 2005–15 (1990–2015)	Validation dataset: 2016, 2018	Test dataset: 2017, 2019
MSW	6695 (16 136)	1435	1463
MSLP	6695 (16 136)	1435	1463
RMW	4841	1085	1168
R64	1707	380	567
R50	2897	622	775
R34	4841	1085	1168

(e.g., Ham et al. 2019; S nderby et al. 2020; Weyn et al. 2020). Considering the efficacies and advantages of CNNs, we decided to use them for modeling TC intensity and wind radii from IR imagery.

Different CNNs have different architectures, in terms of the types of layers (e.g., convolutional, pooling, and fully connected layers), depth (the number of layers which have trainable parameters), and the number and kernel size of convolutional filters (the kernel size refers to the width \times height of the filter mask) of these layers (Goodfellow et al. 2016; Ebert-Uphoff and Hilburn 2020). Many heuristics of the CNNs have been developed, for example, as the networks go deeper (have more trainable layers), they could learn more abstract representations that disentangle and hide explanatory factors of variation behind data (e.g., Goodfellow et al. 2016). A working architecture, however, is usually found by trial and error. To select a principal architecture from the wide choices of network configurations for our application, we tested five modern architectures at or near the state-of-the-art (Table 2) (Simonyan and Zisserman 2014; Szegedy et al. 2016; He et al. 2016; Krizhevsky et al. 2017; Chollet 2017), and assessed the influences of two important aspects of CNN design, i.e., the depth and the kernel size in the first convolutional layer. Because the performance of the model varies slightly even with the same training scenario, we trained the model five times for each scenario with the 2005–15 training data. The averaged performance with respect to the test dataset was then used to represent the grouping. The results (Tables 2 and 3) show that the configuration that yielded the best performance was VGGNet with 13 layers and small (3×3) convolutional filters, which therefore was selected as the backbone of the DeepTCNet. Given the infinite number of CNN hyperparameters combined, our proposed architecture could still be suboptimal, but it works well in practice.

As shown in Fig. 1, the architecture of DeepTCNet consists of 10 convolutional layers (CONV1–10) and 3 fully connected layers (FC11–13). The channel depths of CONV1–10 are 64, 64, 128, 128, 256, 256, 512, 512, 512, and 512, respectively, and the sizes of FC11–13 are 128, 64, and 1, respectively. There is a 3×3 patched maxpooling layer following every two CONV. Flattening is applied before

TABLE 2. Performance and key configuration of different convolutional neural network architectures for the DeepTCNet model. The second and third columns are the network depth and the kernel size in the first convolutional layer, respectively. MAE is the mean absolute error for TC intensity (maximum surface wind; kt) estimation counted with the test dataset. AlexNet is a landmark CNN architecture that almost halves the object recognition error rate (Krizhevsky et al. 2017). VGGNet is a deeper network than AlexNet and uses more efficient 3×3 convolutional kernels (Simonyan and Zisserman 2014). GoogLeNet introduces parallel towers of convolutional kernels with different sizes (Szegedy et al. 2016). ResNet enables CNNs to go very deep by using skip connection (He et al. 2016). Xception is an adaption of GoogLeNet but depthwise separable convolutions are included and it has much lighter parameters (Chollet 2017). The lowest error is shown in boldface.

CNN architecture	Depth	Kernel size	MAE (kt)
AlexNet	8	12	12.0
VGGNet	13	3	10.3
GoogLeNet	22	7	11.4
ResNet	34	7	12.2
Xception	37	3	10.6

the fully connected layers and ReLu activation functions are used.

c. Development and evaluation of the DeepTCNet

As represented in Fig. 1, DeepTCNet is built up in the single-task (STL) and multitask learning (MTL) frameworks. STL is a standard methodology that trains deep learning models to be optimized for one desired task and usually performs well given abundant training data. MTL, however, learns multiple tasks simultaneously. The unique advantage of MTL is that it can leverage shared information and/or the relationship contained in related tasks. Therefore, MTL could improve the accuracy and generality of the model, especially when the training data are scarce (e.g., Zhang and Yang 2017). In this study, STL is applied for intensity (MSW and MSLP) estimation and MTL for critical wind radii (RMW, R64, R50, and R34) estimation, for larger number of high-quality TC intensity samples is available for training than the TC wind radii (Table 1). Besides, the use of MTL to train the model for learning the wind radii simultaneously could also benefit from the underlying physical relationship of the integral TC wind field structure. In MTL, the fully connected layers were separated into multiple flows to learn task-specific model parameters, which could be essential in considering the independent features among inner- and outer-core wind radii tasks (e.g., Weatherford and Gray 1988; Chavas et al. 2015). However, the subnets before the fully connected layers in MTL share the same parameters among tasks to learn transferable features across different tasks. As also displayed in Fig. 1, auxiliary input/information, if present, was concatenated to model in the first fully connected layer.

One challenge in the applications of CNNs to many TC-related research problems is the large amount of training data that required may not be available. The fact of being purely data-driven and black box also makes the CNNs not easy to

TABLE 3. The mean absolute errors on the test dataset for TC intensity estimation (MAE; unit: kt) for different depths and kernel sizes in the first convolutional layer of the VGGNet. The architectures of networks tested are the same as VGGNet in Table 2, but for the depth or kernel size. The MAEs are counted with the validation dataset. The depth and kernel size settings with the lowest estimation error are shown in boldface.

Depth	MAE (kt)	Kernel size	MAE (kt)
8	10.6	3	10.3
13	10.3	5	10.4
22	11.2	7	10.5
34	11.0	9	10.8
37	11.3	11	10.9

obtain predictions with high accuracy and reasonableness. To alleviate these problems, research efforts have been made by integrating prior knowledge/physics into the model development (e.g., Raissi et al. 2019; Snaiki and Wu 2019). Instead of leveraging physics-based formulas as constraints, this study developed a more general physics-augmented deep learning by expanding the input/output learning space. More specifically, auxiliary physical information was used to augment the intensity estimation model (section 3), while learning an auxiliary physical task in MTL was applied to augment the size estimation model (section 4). Since the DeepTCNet approximates a function between the input data and output targets, extending either the input, i.e., incorporating auxiliary physical information or output space, i.e., introducing auxiliary physical tasks based on prior knowledge of TCs could augment the model to learn a more physically generalized function.

All the models were trained by minimizing the loss function, i.e., the mean absolute error (MAE) between the network's outputs and the best track labels. We used MAE as the loss function for it is robust to abnormal data. For MTL-based networks, a weighted sum of the losses of multiple tasks was used, whose weights were trainable and self-updated during training (Cipolla et al. 2018). The advantage of this loss is that it can bring all losses to the same scale to avoid one task dominating the overall loss. Besides, the multitask learning model learns to output inner- to outer-core wind radii, which means that a multitask label is required for each data case. Therefore, when a storm has not reached a specific wind threshold, the missing radii label was set as the RMW. For example, if the case has an intensity of 60 kt, the corresponding multitask label is set as [RMW, RMW, R50, R34]. For such a case, the multitask learning model still learns to search for proper features that can produce reliable inner- to outer-wind radii estimation.

Other hyperparameters and evaluations for the model development are as follows. As also noted in section 2a, each model in this paper was trained five times in each scenario and the averaged performance (in terms of MAE) with respect to the test dataset is selected to represent the model. The model was trained for 200 epochs, and only the learned parameters with the lowest MAE on the validation dataset were saved. In sections 3d and 4b, the mean estimate of intensity or size for each TC case, which was obtained by the five-times-repeated

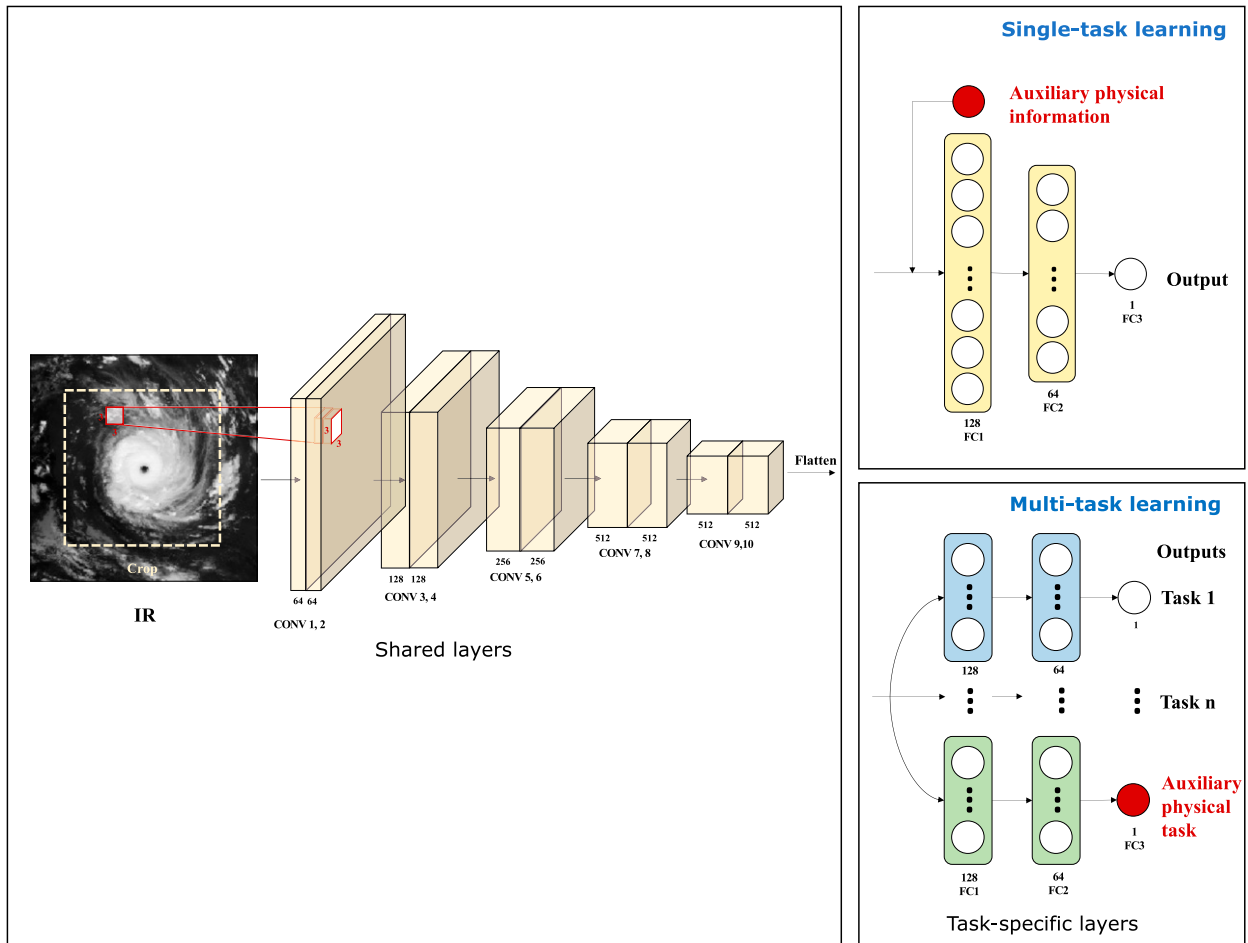


FIG. 1. The configuration of DeepTCNet in both single-task and multitask learning frameworks.

training, was used to present the final estimate of the optimal model. The Adam optimization scheme (Kingma and Ba 2014) with default configuration parameters was used as the learning strategy. The learning rate was fixed as 1×10^{-3} . The batch size (meaning that the data are randomly packed into small groups to be fed to the network) was empirically set 48. The algorithms were built using the open-source Python library Keras (<https://keras.io>) with Tensorflow (<http://www.tensorflow.org>) as the backend and all experiments were run on a graphics processing unit (Nvidia Tesla V100).

3. DeepTCNet for intensity estimation

The sensitivity test applied in the previous section to select the architecture of DeepTCNet has demonstrated a baseline configuration, i.e., the VGGNet with 13 layers and very small 3×3 convolutional kernels, can obtain reliable intensity estimates of TCs from IR imagery. However, such a model could be data inefficient, which means that they are essentially dependent on the quantity and quality of training data yet lack physical knowledge of the natural world. To improve the model for estimating TC intensity, we first address a method to

incorporate auxiliary physical information of TCs for augmentation in this section. Sequential IR images, which represent the continuous TC development, are then introduced to improve intensity estimation. In addition, the influence of the amount of training data is discussed. Finally, we will finally integrate all the positive settings to build up an optimal DeepTCNet model for TC intensity estimation.

a. Auxiliary physical information augments learning

An effective and practical approach to capitalizing on prior physical knowledge of TCs is explored to alleviate the risks of the deep learning model for naively memorizing training data, such as the pseudo features and data bias in the IR imagery and the intensity labels. The hypothesis is that the auxiliary physical information of TCs introduces an additional physical relationship between the input space and the output TC intensity target; therefore, it could constrain the model to produce better TC intensity estimates.

The auxiliary physical information used in this study include the surface information (i.e., a factor valued 0 or 1 to indicate whether the TC is over land or not), the storm age (the period since the TC become a named storm in the unit of the hour;

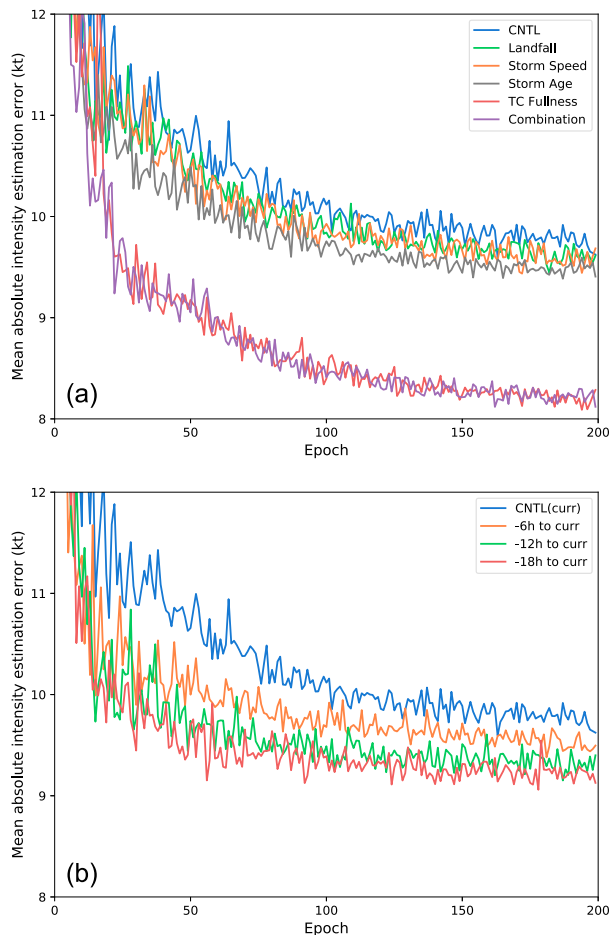


FIG. 2. Learning curves, i.e., mean absolute errors for the intensity estimation on the validation set during each training epoch of DeepTCNet developed with (a) different auxiliary information and (b) successive IR sequences of different time span.

e.g., Kossin et al. 2007), the storm's translation speed (Mei et al. 2012) and the TC fullness (Guo and Tan 2017). The TC fullness is defined as the ratio of the extent of the outer-core wind skirt to the outer-core size of the TC, i.e., $(R_{34} - RMW)/R_{34}$, which is a new concept for the representation of the storm wind structure and has a good physical relationship with TC intensity and intensification (Guo and Tan 2017).

To compare the effect of the auxiliary physical information, the model (VGGNet in Table 2) that only takes in the current IR imagery and output the MSW intensity is used as the control run (CNTL) model. Its estimation error on the test dataset ($MAE = 10.3$ kt) thus serves as the error baseline. As illustrated by the learning curves in Fig. 2a, integrating the auxiliary physical information made the model achieve better performance than the CNTL on the validation dataset. Moreover, the performance with respect to the test set for models augmented by the landfall information, storm speed, storm age, the TC fullness, and the combination of all the auxiliary information is 10.2, 10.3, 9.6, 9.1, and 8.6 kt, respectively. The results show that incorporating TC wind radii structure description, i.e., the

TC fullness, caused the largest error reduction (12%) in estimating TC intensity over the CNTL. Although it is impossible to derive TC intensity from the TC fullness directly, the TC fullness has shown to be a very useful auxiliary information to augment the model for intensity estimation. Since there exists a good physical relationship between the intensity and the radial wind structure of TCs, i.e., the TC fullness (Guo and Tan 2017), adding the TC fullness as an auxiliary input factor is equivalent to imposing a constraint between the TC wind field structure and the targeted output intensity, which could be the cause of the improvement. The storm age also contributed to a 7% error reduction in estimating MSW compared to the CNTL. A general characteristic of TC intensity evolution conveyed by the storm age could be the reason. That is, the model could be constrained by the early storm age to carefully give low-value estimates of intensity instead of starting with larger ones. The storm speed and surface information have a small contribution in improving the TC intensity. Overall, the model augmented by the auxiliary information combined achieved the best performance on the test set (17% improved than the CNTL). Besides, Fig. 2a illustrates that adding auxiliary physical information resulted in the estimation errors decreasing at an early stage of the learning. This suggests that the physical information may also help discard some implausible features in the data, hence accelerating the learning.

b. Sequential IR imagery helps to improve

Meteorologists track cloud features through successive periods of TC development to help determine the intensity (Velden et al. 2006). Therefore, it is natural to ask whether the deep learning model can improve accuracy by including recent images of a TC. In this study, we used a sequence of IR images at 6-h intervals from the previous 18, 12, or 6 h before the current image.

As shown in Fig. 2b, MSW estimation errors consistently decrease when previous IR images were included, and the result holds on the test set. The highest performance was achieved by using a sequence of images covering an 18-h period, and this has led to an error reduction of 6% lower than CNTL (with respect to the test set). This confirms that physical information on how the TC changes with time in the sequential images can benefit the intensity estimation. Likewise, the risk of important features disappearing in one IR image is also alleviated. However, note that using images that are too old (e.g., 24 h or earlier) does not improve the estimation and may even harm the estimation skill as the systems undergo rapid intensity change. Better results may be achieved by using hourly or even higher-frequency IR data (e.g., Himawari-8, GOES-16/-17 with 10-min global coverage) and including a channel-wise attention module to improve the representation power of the model by focusing information of interest (e.g., Woo et al. 2018).

c. Sensitivity to the amount of training data

The backbone of DeepTCNet is CNN of complex architecture and heavy parameters (~ 10 million trainable parameters), which usually rely on a large-sample-size dataset and many strategies to perform well (e.g., Goodfellow et al. 2016). However, the large labeled datasets used in computer vision, such as the ImageNet could not be available in Earth science.

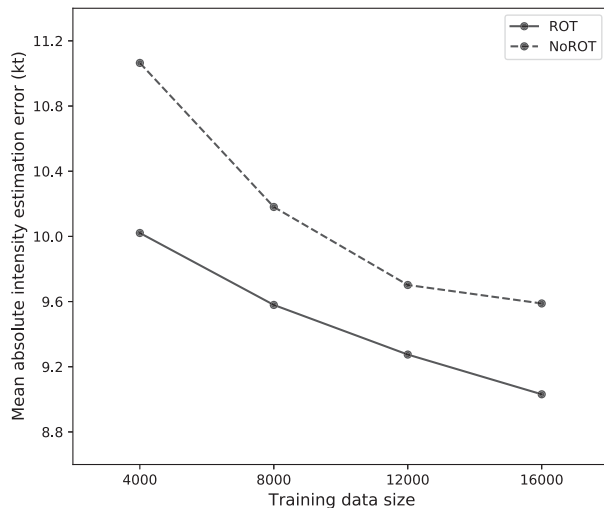


FIG. 3. Mean absolute errors of MSW (kt) on the test dataset produced by the DeepTCNet obtained by different training data sizes. The model trained with rotation augmentation is denoted as ROT (dashed line).

Therefore, it is important to ask how the training data size could impact the intensity estimation of TCs.

To test the impact of training data size, more intensity records from the IBTrACS database from 1990 to 2004 were also included as training data (Table 1). The 4000, 8000, 12 000, and 16 000 instances were then sampled from the whole 1990–2015 training data. These samplings follow an identical distribution in which the ratio of the number of tropical depressions and tropical storms ($MSW < 64$ kt), minor hurricanes ($64 \leq MSW < 96$ kt), and major hurricanes ($MSW \geq 96$ kt) is 0.7:0.2:0.1. Moreover, the influence of the rotation augmentation, which is a commonly applied data augmentation technique in the context of deep learning, is also explored.

As illustrated in Fig. 3, more training data leads the model to achieve lower intensity estimation errors, as expected. Besides, the model trained with the rotation augmentation had a further enhancement in estimating the TC intensity. The model trained with 16 000 samples together with the rotation augmentation has the lowest estimation error on the test dataset (9.0 kt), which is in-line with the model trained with the 2005–15 training data and augmented by the TC fullness. Figure 3 also shows that the model's improvement gained by including more observational samples is more evident than by rotation augmentation. This result suggests that enlarging the training dataset by observational cases is key to improving the deep learning-based model. Likewise, this sensitivity test was based on the intensity, which has more quality-consistent data than wind radii. We found that including more training samples of the wind radii labels obtained from the TC Vitals database, which is not postanalyzed thus could be of poor quality, does not help to improve model performance, meaning that data quality is also important.

d. Performance evaluation

A final intensity estimation model named DeepTCNet-I is developed by combining all the beneficial training settings. The

settings include using the auxiliary information combined (storm age, storm speed, surface information and the TC fullness), multichannel images of IR sequence (previous 18, 12, 6 h, and current IR), more training data (the best track intensity labels from IBTrACS database for the TCs in 1990–2015) and the rotation augmentation.

The promising accuracy of the DeepTCNet-I across nearly all TC intensity scales is illustrated in Fig. 4. The Pearson correlation coefficient (R) between the DeepTCNet-based MSW (MSLP) and the recon-aided best track “ground truth” in the test set is 0.97 (0.97). The accuracy also holds with respect to the test set of 1463 samples ($R = 0.97$). To better evaluate the intensity estimation accuracy of the DeepTCNet method, we compared the DeepTCNet-I with two benchmark intensity estimation methods, the IR-based ADT method (Olander and Velden 2019) and the multi-satellite-based SATCON method (Velden and Herndon 2020). As revealed in Fig. 5, the DeepTCNet systematically outperforms the ADT method for all intensity categories. The overall improvement of the DeepTCNet over the ADT is 39% for MSW and 33% for MSLP estimation (Table 4). These improvements are significantly different from zeros with a 99% confidence level in the Kolmogorov–Smirnov test. Since both the DeepTCNet and ADT are primarily based on IR imageries, the result implies that the DeepTCNet method can dramatically improve the efficacy of information in the IR observations. The DeepTCNet overall is in comparison with the SATCON. However, it tends to handle weaker systems (TD/TS and minor hurricanes) better than the SATCON. The negative bias for the MSW estimation for category-5 hurricanes is evident for DeepTCNet. The high percentage of tiny unresolved eyes at the strong intensities could be a cause for this underestimation. Given the microwave sensors can detect clear eye structure in strong intensities, SATCON would have better estimates of the intense TCs and thus outperforms the DeepTCNet and ADT for these cases. The underestimation of intense hurricanes could also reveal a real bias of the best track intensity dataset used to train the DeepTCNet. That is, we are detecting higher intensities than before as observations increase. Since the DeepTCNet-I was trained on historical data of TCs in the 1990–2015 seasons, it would also lack the ability to produce newly observed large intensity estimates based on higher-temporal/spatial-resolution IR data and improved aircraft reconnaissance.

As the DeepTCNet-I is trained with the best track intensity data of Atlantic TCs, and 70% of the data were solely supported by satellite observations (Landsea and Franklin 2013; Torn and Snyder 2012), it is natural to speculate that the best DeepTCNet-I can do is to reproduce the Dvorak method (Velden et al. 2006; Knaff et al. 2010). However, Table 5 shows that the DeepTCNet-I also performs better than the subjective Dvorak² method based on 2 years of homogeneous testing data. Furthermore, the relative error levels of DeepTCNet-I

² The subjective Dvorak intensity estimates were obtained from the ATCF fixes files in which the intensities are indicated as DVTS in the fifth column.

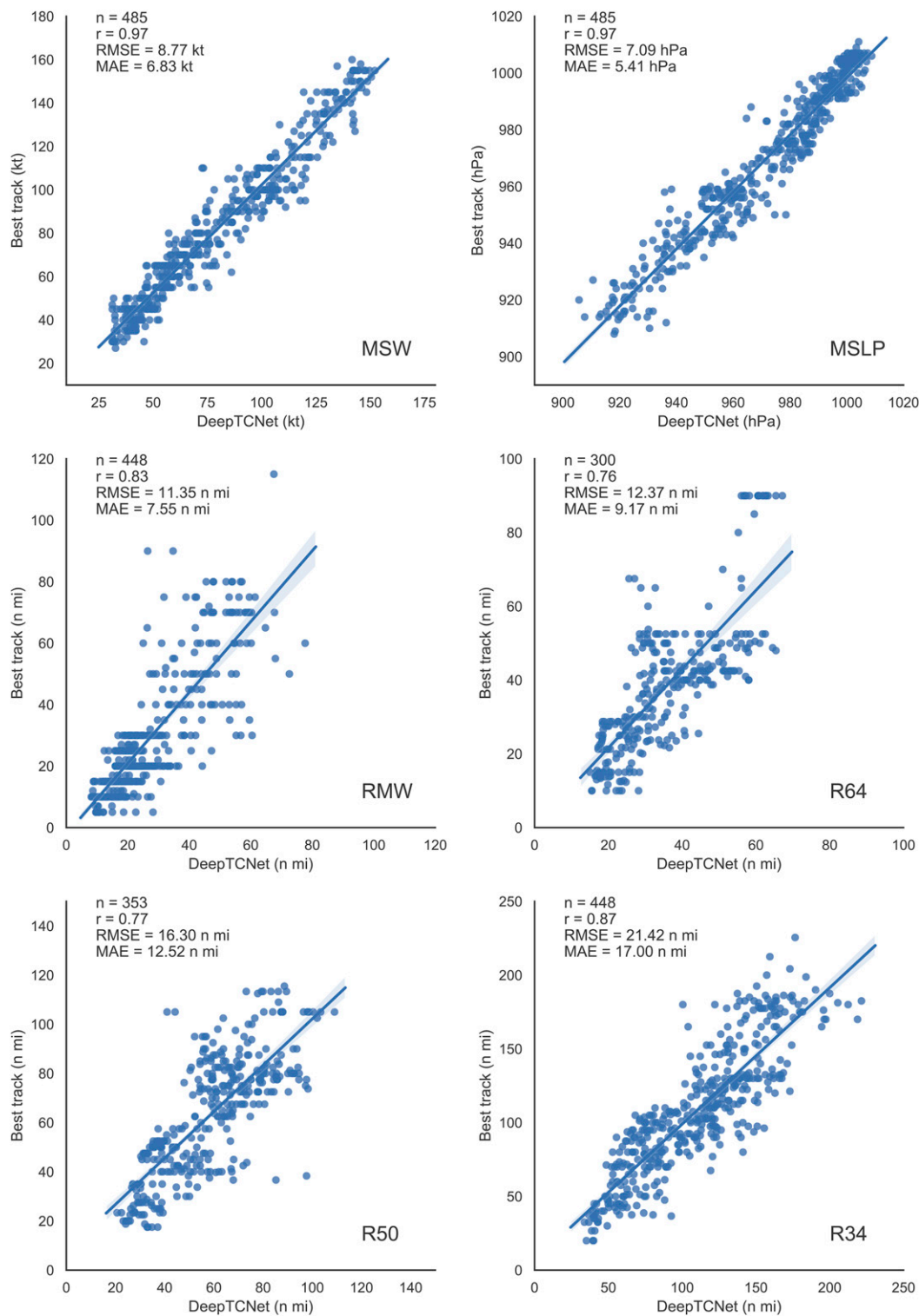


FIG. 4. Scatterplots of the DeepTCNet-estimated intensity (MSW, MSLP) and wind radii (RMW, R64, R50, and R34) vs the ± 3 -h recon-aided best track records for the Atlantic TCs in 2017 and 2019.

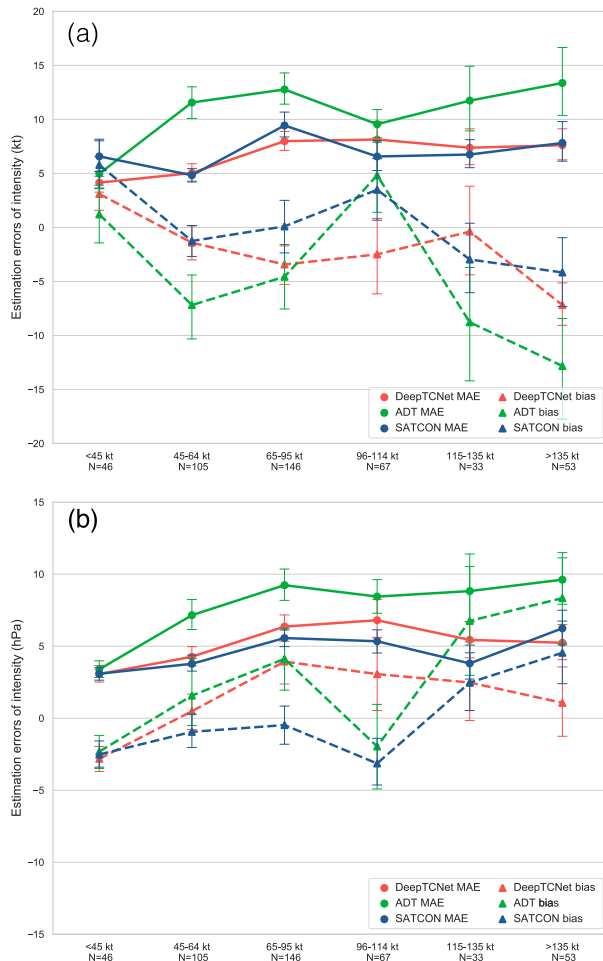


FIG. 5. Performance of (a) DeepTCNet MSW estimates (kt) and (b) MSLP estimates (hPa) against ADT and SATCON as a function of TC intensity (MSW) bins (kt; x axis). The metrics are counted using the ± 3 -h recon-aided best track records for the Atlantic TCs in 2017 and 2019, with 95% confidence intervals obtained by bootstrapping.

are counted and compared with the intrinsic data uncertainty in the best track (Fig. 6). For aircraft/satellite monitoring and for U.S. landfalling cyclones, the relative uncertainty is about 15% for tropical storms, $\sim 10\%$ for category-1 and -2 hurricanes, $\sim 8\%$ for major hurricanes, and $\sim 10\%$ for all categories (Landsea and Franklin 2013). In comparison, the relative error of the DeepTCNet-based MSW for the recon-aided testing cases ($N = 485$) is $\sim 10\%$ for tropical storms, $\sim 10\%$ for category-1 and -2 hurricanes, $\sim 7\%$ for major hurricanes, and $\sim 8\%$ for all categories, which is near the uncertainty in the recon-aided best track intensity data. For the best track cases without aircraft reconnaissance fixes within ± 3 h in the testing dataset ($N = 978$), the relative error of the DeepTCNet-based MSW is 10%, which is lower than the inherent uncertainty of the satellite-only best track estimates (Fig. 6). This suggests that the DeepTCNet has the potential to improve the accuracy of intensity estimates when it is incorporated in the poststorm

TABLE 4. Homogeneous comparison between the ADT, SATCON, and DeepTCNet estimates of Atlantic TC intensity in 2017 and 2019 against the recon-aided best track for MSW (kt) and MSLP (hPa) ($N = 450$).

	Bias	MAE	RMSE
ADT – MSW	–4.5	11.2	14.2
SATCON – MSW	0.1	7.3	9.4
DeepTCNet – MSW	–2.3	6.8	8.7
ADT – MSLP	2.6	8.0	10.2
SATCON – MSLP	–0.4	4.8	6.0
DeepTCNet – MSLP	1.8	5.4	7.0

intensity analysis. In addition, a more accurate version of DeepTCNet-I could be obtained by fine-tuning/transfer learning with the recon-aided best track intensity labels in future work. To note, however, the data uncertainties of best track were obtained subjectively (Landsea and Franklin 2013).

The error metrics for the MSW estimation of DeepTCNet-I also suggest a comparatively better performance over the previous deep learning-based intensity estimation methods. The root-mean-square estimation errors (RMSE) of MSW in Pradhan et al. (2018) is 10.2 kt (11.4 kt for a recon-aided dataset). Chen et al. (2019) conducted post smoothing procedure to improve their intensity estimation accuracy. That is, the RMSE decreased from 10.4 to 8.7 kt (20% improvement) by applying a five-point weighted average. Given such post-analysis smoothing requires the past and future intensity estimates, which is not valid for near-real-time estimation, and the other deep learning applications do not apply postanalysis, we used the 10.4 kt to represent Chen et al. (2019). And note that Chen et al. (2019) reported a 32% improvement than the ADT technique by a homogeneous comparison based on 144 samples of TCs in 2017. Wimmers et al. (2019)'s RMSE of MSW as 14.3 kt (10.6 kt on a recon-influenced dataset). In comparison, the RMSE of DeepTCNet-I on the validation dataset is 7.7 kt (8.7 kt on the recon-aided test dataset), which is lower than the other deep learning methods. The superiority of DeepTCNet-I in terms of the lowest RMSE could be because of incorporating some physics of TC into algorithms. However, an objective comparison of all the deep learning-based models using homogeneous testing data is required in the future.

4. DeepTCNet for wind radii estimation

This section sets up the DeepTCNet to estimate critical wind radii (R34, R50, R64, and RMW) of TCs from IR imagery.

TABLE 5. Homogeneous comparison between the subjective Dvorak and DeepTCNet estimates of Atlantic TC intensity in 2017 and 2019 against the recon-aided best track for MSW (kt). The subjective Dvorak intensity estimates were obtained from the ATCF fixes files in which the intensities are indicated as DVTS in the fifth column ($N = 333$).

	Bias	MAE	RMSE
Dvorak – MSW	–4.9	7.6	9.4
DeepTCNet – MSW	–1.4	6.3	8.1

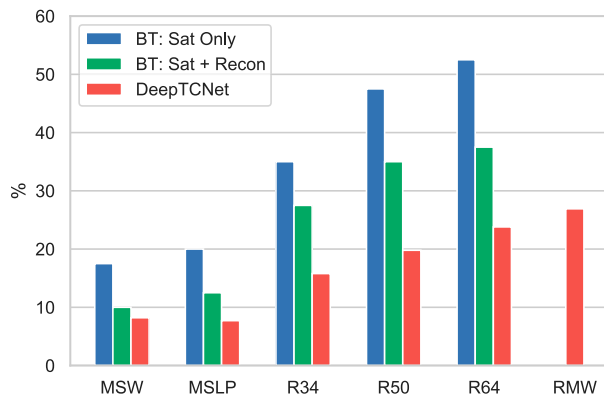


FIG. 6. Relative uncertainty in the best tracks (blue and green bars) and the percentage relative estimation errors of the DeepTCNet (red bars) for MSW, MSLP, R34, R50, R64, and RMW. The statistics of best track uncertainty (if available) were taken from Landsea and Franklin (2013) as references. The verification of DeepTCNet was based on 2017 and 2019 cases in the Atlantic basin with ± 3 -h aircraft reconnaissance fixes.

Though the four wind radii estimates could be separately obtained by training four independent networks as we applied for TC intensity estimation and augmented information could also be incorporated to enhance the model, we explored another approach based on multitask learning (MTL). MTL is selected for its advantages in leveraging useful information contained in multiple related learning tasks (Zhang and Yang 2017), which would help alleviate the data sparsity problem in wind radii estimation. In the rest of this section, the efficacy of MTL is first explored, after which the performance of DeepTCNet for wind radii estimation is then evaluated. Note that rotation data augmentation is already applied to training all the wind radii estimation models.

a. Multiple physical-related tasks augment learning

The virtue of jointly learning from multiple tasks was shown by comparing MTL-based models with four STL-based models that separately estimate RMW, R64, R50, and R34. As illustrated in Fig. 7, the proposed MTL4 that solves four wind radii tasks (RMW, R64, R50, and R34) simultaneously outperforms each STL that only models a single wind radius task. Since the four critical wind radii approximate the integrated wind field of TCs, the result implies that the DeepTCNet could benefit from learning physically related tasks. One possible explanation is that multiple physically related wind radii tasks might impose the model to learn more useful and generalized features from data. Therefore, the model can produce more reliable size estimations from the IR imagery.

Moreover, we added TC intensity, i.e., MSW and MSLP, as auxiliary tasks to explore if learning with TC intensity could enhance the wind radii estimation. As displayed in Fig. 7, the estimation errors on the test dataset for four wind radii parameters all decreased by learning MSW as the auxiliary task (the MTL5). Besides, it is interesting to observe that using MSW as the auxiliary task contributes more improvement in the inner-core wind radii estimations (RMW and R64), whereas using

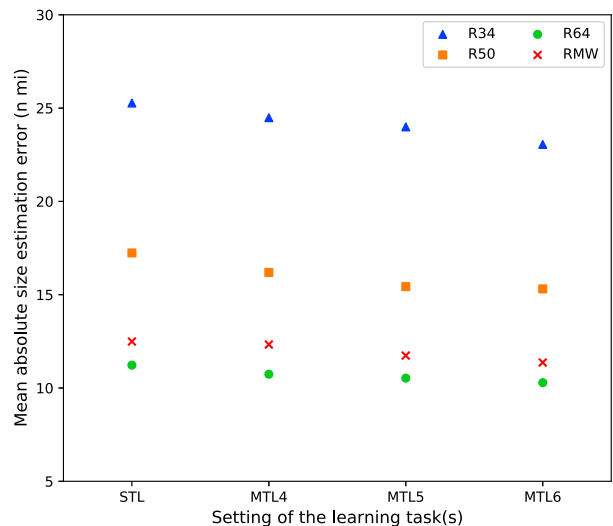


FIG. 7. Sensitivity of wind radii estimation absolute errors (n mi; 1 n mi \approx 1.85 km) to single-task learning (STL) and multitask learning (MTL). The MTL4 denotes a model jointly learns four wind radii tasks: RMW, R64, R50, and R34. The MTL5 jointly learns five tasks for wind radii and intensity estimation: RMW, R64, R50, R34, and MSW. The MTL6 model jointly learns six tasks for wind radii and intensity estimation: RMW, R64, R50, R34, MSW, and MSLP. The metrics were counted with the test dataset listed in Table 1.

MSLP as the auxiliary task (not shown) mainly boosted the estimations of the outer-core wind radii (R34 and R50). The improvement on RMW and R64 by the auxiliary MSW learning task could because they are both primarily determined by the inner-core dynamics of TCs (Emanuel and Rotunno 2011; Weatherford and Gray 1988). Whereas, since the MSLP also depends on the radial integral of the wind field (Knaff and Zehr 2007; Courtney and Knaff 2009), which could be approximated from radial momentum equation by assuming cylindrical gradient wind balance in TCs (e.g., Chavas et al. 2017), the model could have capitalized on the physical relationship between MSLP and R34 to support learning the outer-core wind radii (R34). Hence, both MSW and MSLP are included as auxiliary learning tasks to augment modeling critical wind radii, RMW, R64, R50, and R34, which is hereafter referred to as MTL6. As also shown in Fig. 7, the MTL6 model has the overall best performance. Overall, the MTL6 has reduced the wind radii estimation error by 6% than the MTL4 and by 12% over the STL. This result suggests that the MTL-based DeepTCNet could take advantage of the physical relationship(s) between the TC intensity and wind structure described by critical wind radii to yield improvement. This augmentation is similar to the significant improvement in MSW estimation obtained by introducing the auxiliary physical information of TC structure, i.e., TC fullness (Fig. 2a), to the intensity estimation model. Moreover, since the TC intensity (MSW and MSLP) and the wind fields expressed by critical wind radii (R34, R50, R64, and RMW) together empirically approximate the wind–pressure relationship of TCs (Knaff and Zehr 2007; Courtney and Knaff 2009; Chavas et al. 2017), the wind–pressure relationship could be the reason that

TABLE 6. Homogeneous comparison between the MTCSWA and DeepTCNet estimates of Atlantic TC wind radii in 2017 and 2019 against the recon-aided best track for RMW, R64, R50 and R34 (n mi; 1 n mi \approx 1.85 km).

Parameter	Technique	Bias	MAE	RMSE
RMW ($N = 381$)	MTCSWA	4.6	11.7	18.2
	DeepTCNet	−1.3	7.3	11.0
R64 ($N = 216$)	MTCSWA	4.5	10.2	13.1
	DeepTCNet	−0.8	7.3	10.0
R50 ($N = 313$)	MTCSWA	−0.8	15.8	20.7
	DeepTCNet	−3.1	11.8	15.5
R34 ($N = 382$)	MTCSWA	−7.4	26.4	33.0
	DeepTCNet	2.7	17.2	21.8

MTL6 performs best among the size estimation models. Note that although the MTL6 also estimates TC intensity, in this study it can only apply to the storms which have intensities larger than 34 kt.

b. Performance evaluation

Given the impressive performance of the MTL6, it is selected as the final wind radii estimation model named DeepTCNet-R. Scatterplots of the wind radii estimates produced by the DeepTCNet-R against the recon-aided best track “ground truth” are also shown in Fig. 4. The overall performance of the DeepTCNet-R is quite impressive. Especially, the R34 estimations given by the DeepTCNet-R compare well to the best track data ($R = 0.87$). However, it is worth noting that the scatter of the size estimation points shown in Fig. 4 is much larger than the intensity estimations. The smaller sample size training dataset used to train the DeepTCNet-R and the more considerable wind radii data uncertainty could be the causes.

As shown in Table 6, the DeepTCNet-R outperforms the coincident operational wind radii estimates from the Multiplatform Tropical Cyclone Surface Wind Analysis technique (MTCSWA; Knaff et al. 2011) with an average improvement of 32% (32%, 25%, 28%, and 38% for RMW, R64, R50, and R34, respectively). These differences also passed the Kolmogorov-Smirnov test with a 99% confidence level. Given the DeepTCNet-R only relies on IR imagery to estimate wind radii while the MTCSWA is based on multiple satellite channels, including more satellite channels into DeepTCNet-R is likely to further increase wind radii estimations’ accuracy. This is very important because there still is large uncertainty in the satellite-based wind radii estimations (Knaff and Sampson 2015). As shown in Fig. 6, the relative wind radii estimation errors provided by the DeepTCNet-R is much lower than relative uncertainty of the satellite-only-based wind radii observations. The DeepTCNet-R gives the most accurate estimation for R34, then are R50 and R64. However, the DeepTCNet-derived RMW estimates have the highest relative errors among the wind radii estimates, suggesting that endeavors are still required to estimate this most difficult-to-measure TC inner-core feature. To the authors’ knowledge, DeepTCNet is the first deep learning application for the real-time estimation of TC wind radii. Moreover, the unique point of DeepTCNet is that it can produce multiple wind radii simultaneously. Although this paper only discussed the symmetric wind radii estimation, the DeepTCNet can also be applied to estimate the TC wind field’s

asymmetries when asymmetric features (such as storm motion and vertical wind shear) are included as the auxiliary information of the model. A subjective comparison suggests that the DeepTCNet also outperforms or in line with the previous techniques (Table 7) though a homogeneous comparison is needed in the future.

5. Discussion

a. Interpretation and visualization of DeepTCNet

The backbone algorithm of DeepTCNet is the deep convolutional neural networks (CNNs), which are often treated as black boxes. That is, CNNs pick and decide features that they think essential from input data, thus usually left little ideas to the humans. However, the interpretation of CNNs is useful for establishing trust and confidence in users and suggesting the models’ signs of disabilities. To provide the readers some information about how the DeepTCNet makes the decision, we implemented saliency maps (Smilkov et al. 2017) and Layer-wise relevance propagation (LRP; Montavon et al. 2017, 2018) to attribute what features of an input are responsible for the model’s output. The saliency maps method uses the gradient of one output with respect to individual pixels of the IR image to reflect their influence on the final intensity or size estimation. The saliency value can be either positive or negative. And in this study, the SmoothGrad (Smilkov et al. 2017) was used to help visually sharpen the gradient-based saliency maps. One disadvantage of the saliency method is that it only looks at local gradients in the input space, which could be limiting (Montavon et al. 2017). The LRP explains the decision of a model by propagating the prediction from the output to the input using local redistribution rules. Since LRP does not rely on gradients, it does not suffer from problems such as gradient shattering and explanation discontinuities (Montavon et al. 2018), which also appear when plotting the saliency maps for DeepTCNet. One should note that LRP was designed for classification networks by backpropagating the activation scores of the target class to the input space. In the extensible regression setting (for the DeepTCNet), a larger LRP value represents that one feature has higher relevance to increasing the model’s output intensity or size.

Consider the example in Fig. 8, the saliency (SmoothGrad) and the LRP reveal that the model (DeepTCNet-R in this case) has learned different features for estimating diverse TC parameters. The saliency results show that the increase of bright temperature in the eye of Irma (2017) can cause the model to produce a higher intensity (MSW). The decrease of temperature around the storm’s eye, which means the deeper convections in the eyewall region, can also increase the value of intensity (MSW) estimation. The saliency map for the MSLP tells the same story, i.e., the MSLP drops (i.e., intensity raises) with a warmer eye and colder eyewall. The saliency features in the outer region seem fuzzy and not physically meaningful. In addition, the estimated RMW of DeepTCNet-R (denoted by the blue ring in Fig. A1c) locates near the center region where saliency values transfer from negative to positive, meaning that the model might have learned that the concentration and expansion of the eye/eyewall directly determine the TC’s RMW.

TABLE 7. Mean absolute errors (unit: n mi; 1 n mi \approx 1.85 km) of wind radii estimates derived by DeepTCNet and other methods, as well as the satellite technique used for each method. The metrics of DeepTCNet are counted with Atlantic TC wind radii in 2017 and 2019 against the recon-aided best track for R34, R50, R64, and RMW. Note that the statistics for the existing size estimation models are taken from the original papers and their evaluation samples are inconsistent.

Method	Satellite technique	RMW	R64	R50	R34
Demuth et al. (2006)	AMSU	—	6.8	13.3	16.9
Kossin et al. (2007)	IR	11.4	14.5	19.8	24.2
Knaff et al. (2011)	IR, Scatterometer, AMSU	—	13.0	17.8	36.5
Knaff et al. (2016)	IR	—	12.0	20.0	37.0
Dolling et al. (2016)	IR	—	7.3	12.5	20.8
DeepTCNet	IR	7.6	9.2	12.5	17.0

The estimated R34 is located at a ring region where the bubble-like saliency features change into roll-like patterns. It is also interesting to observe from the saliency maps near the center that the model assigns more importance to axisymmetric cloud features for estimating the inner-core parameters (MSW, MSLP, RMW, and R64), while the important features for outer-core wind radii (R50 and R34) are asymmetric. As the supplement, the LRP reveals that the intensity is largely

determined by the IR properties near the center (e.g., eye and the coldest clouds), wind radii of R34 is dependent on diverse clouds, convections and rainbands features in the inner- to outer-core regions, while the MSLP is determined by a combination of that information. Moreover, we further compare the LRP patterns of R34 for the smallest (25% quantile) and the largest (75% quantile) TCs in the testing dataset. As shown in Fig. 9, there is an obvious difference in the R34's LRP results for the large and small TCs, that is the larger TCs have wider spread LRP relevance values. This result gives us more confidence that the DeepTCNet might have learned physically important features from the IR images. The interpretation maps of the DeepTCNet-I, which uses the small inner-core-region image for estimating intensity, are also shown (inserted graphs in Figs. 8a,b,g,h), and showing similar descriptions with the interpretation maps of the DeepTCNet-R. These results suggest that the CNNs applied with proper interpretation methods could be useful for exploring more complex TC mechanisms, while future studies and more advanced and suitable interpretation techniques are warranted.

b. Broader applications and limitations of DeepTCNet

Section 3a shows that including auxiliary information, especially the TC fullness, can help the model improve estimating TC intensity from IR imagery, the TC fullness hence was used to develop a final intensity estimation model (DeepTCNet-I). Here we further discuss how the auxiliary information of TC fullness could influence the accuracy of TC

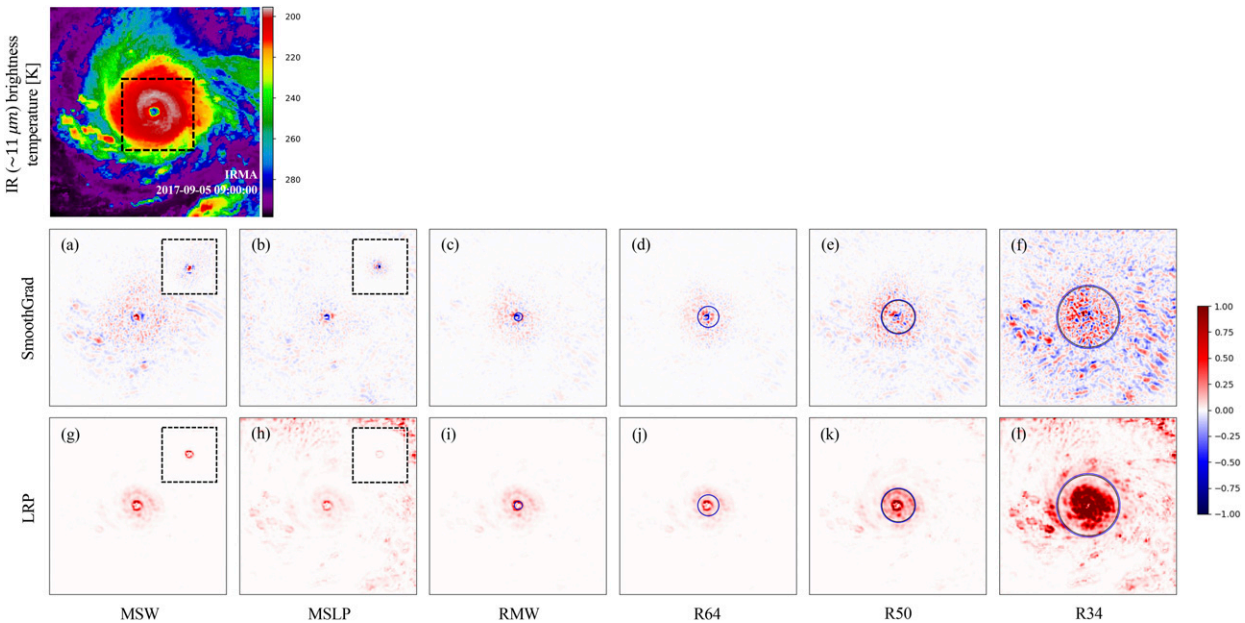


FIG. 8. The (top left) input imagery and (a)–(l) interpretation heatmaps of DeepTCNet for estimating intensity and wind radii. The heat maps were obtained by (a)–(f) SmoothGrad (Smilkov et al. 2017) and (g)–(l) LRP (Montavon et al. 2018) for Hurricane Irma at 0900 UTC 5 Sep 2017 with an intensity of 142 kt. The inserted boxes denote the small IR image and the corresponding heat maps [(a),(b),(g),(h)] for the DeepTCNet-I, while the larger ones are for the DeepTCNet-R. The black rings and blue rings in (c)–(f) and (i)–(l) represent the best track wind radii and the DeepTCNet-estimated wind radii, respectively. The relative error for each estimated parameter was $<1\%$ compared with the best track observation. The LRP and saliency maps are calculated using the iNNvestigate package (Alber et al. 2019), and LRP is applied with the $\alpha = 1$ and $\beta = 0$.

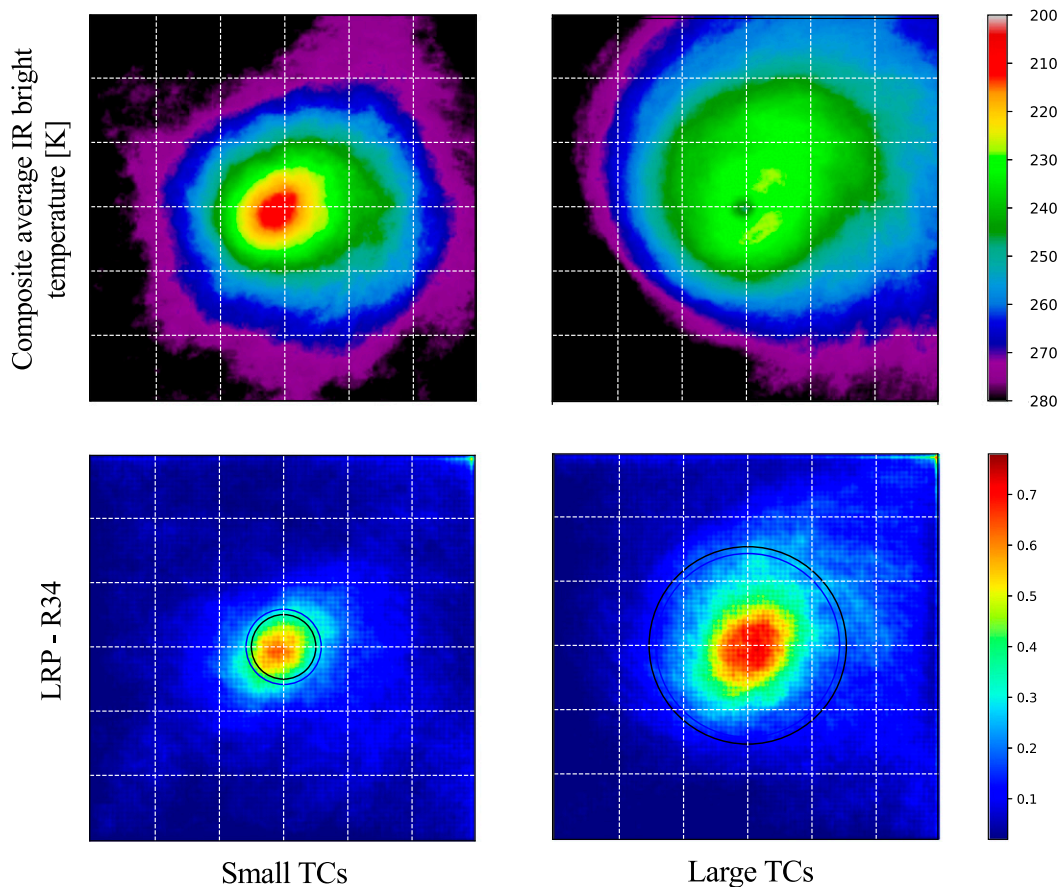


FIG. 9. Composite average IR bright temperature and LRP heat maps for DeepTCNet-R to estimate the R34 for the (left) small TCs and (right) large TCs. The 25% quantile samples of the R34 in the recon-aided testing dataset were used to compose small TCs ($N = 110$, $\mu = 54$ n mi), while the 75% quantile samples composed the large TCs ($N = 112$, $\mu = 165$ n mi). The black rings and blue rings in LRP represent the averaged best track R34 and the averaged DeepTCNet-estimated R34, respectively.

intensity estimates in an operational setting when the real-time TC Vitals data are unavailable. As displayed in Table 8, the intensity errors increase when the TC fullness from the real-time MTCSWA database is used. This is expected since the model was trained with the TC Vitals-based TC fullness but used the MTCSWA-based TC fullness for application. From a more convenient application perspective of the DeepTCNet method, it is also a nice choice to use TC fullness derived from MTL4 and MTL6 as the auxiliary augmentation information. In this case, the DeepTCNet could rely on frequently available IR observations to produce TC intensity estimates of high temporal resolution. Table 8 shows that using the MTL6-based TC fullness as auxiliary information can produce better intensity estimation than the MTL4- and MTCSWA-based TC fullness, yet the differences are small and not statistically significant. As a comparison, we also used the best tracked TC fullness obtained from the IBTrACS to make TC intensity estimation. As expected, the intensity estimation errors further decreased to 6.7 kt when the more accurate TC fullness was used in place (Table 8). These error metrics obtained with best tracked TC fullness could be

regarded as a potential performance of DeepTCNet-I when more accurate physical auxiliary information of TC fullness was available in real time. Besides, the model augmented with the real-time-available TC fullness data still yields better performance over without using this auxiliary information (Table 8).

While the DeepTCNet was developed and evaluated with the TC cases in the Atlantic basin, where reconnaissance aircraft data are available, the methods discussed could extend to other ocean basins. Our preliminary testing results show that DeepTCNet developed with the Atlantic data can be applied to the western North Pacific basin with reliable intensity and size estimation performances and retraining the model with the Northern Hemisphere data can result in better results. The ability of DeepTCNet to generalize to other basins is essential in that these regions rely extensively on satellites for monitoring and forecasting the TCs. In another important way, DeepTCNet can contribute to TC intensity and size analysis in poststorm assessments and the best track procedures. This could be particularly important for the TC size analysis because most operational agencies just started to produce

TABLE 8. MAE and RMSE of the DeepTCNet-I estimates given different data sources of TC fullness. “No TC fullness” denotes the comparison model that was built up with the identical model settings of the DeepTCNet-I but without using TC fullness as auxiliary information. The statistics are calculated with the recon-aided best track for MSW (kt) for Atlantic TCs in 2017 and 2019 ($N = 485$).

Data source of TC fullness	MAE (kt)	RMSE (kt)
TC Vitals	6.8	8.8
MTCSWA	7.5	9.6
MTL4	7.4	9.4
MTL6	7.2	9.2
Best track	6.7	8.7
No TC fullness	8.0	10.5

quality-controlled estimates of wind radii since 2004 (e.g., National Hurricane Center) or even more recent since 2016 at Joint Typhoon Warning Center (Sampson et al. 2018). Moreover, there are no quality-controlled estimates of RMW provided by these operational agencies (Sampson et al. 2018). All of these are primarily caused by the challenges faced with observing surface wind fields from space, and no reliable method like the Dvorak is widely applied for size estimation. Given the best track data are widely used to build or validate TC analysis algorithms and forecasting, more efforts to produce higher-quality best tracks of intensity and wind radii are still of vital concern.

Although the evaluation results show that DeepTCNet has outstanding performance for estimating both intensity and wind radii of TCs, there are some limitations of the DeepTCNet. The first is that the model built up for intensity estimation underperforms for very intense TCs. The high percentage of unresolved tiny eyes in the IR imagery for strong intensities as well as the relatively low number of these cases for the model to learn from could be the causes. Whereas, this underestimation bias at intense TCs could reveal an observational bias in the best track. That is, we are observing more intense hurricanes than before. Furthermore, since the DeepTCNet is trained on historical data of TCs before 2016, it would lack the ability to produce large intensity estimates as recently observed. The utility of microwave-based satellite observations and more accurate recon-aided data could boost the estimation of TC intensity and size.

6. Summary

DeepTCNet is a deep convolutional neural network as a backend method designed to capitalize on physical knowledge of TCs to produce accurate intensity (MSW and MSLP) and wind radii estimates (R34, R50, R64, and RMW) of TCs from IR imagery. Introducing auxiliary physical information is an effective approach to improve the model for intensity estimation. It is demonstrated that incorporating TC fullness, which describes the critical structure of TCs, to the model can significantly enhance its performance for intensity estimation. Multitask learning, which has the advantage of sharing different and more general features among tasks was applied to improve the model for wind radii estimation. Our results show that learning multiple wind radii tasks simultaneously can yield

more accurate wind radii estimation than separately learn a single wind radii task. Moreover, we also found that the model trained to estimate two intensity measures (MSW and MSLP) and four critical wind field measures (R34, R50, R64, and RMW) simultaneously yields the best wind radii estimates. The inherent wind–pressure relationship approximated by these tasks could be a cause. Besides, results show that including continuous IR images from near-previous times benefited the model for intensity estimation.

The evaluation results based on homogeneous testing samples showed that the DeepTCNet is in-line with SATCON, but systematically outperforms the ADT by 39% improvement for MSW and 33% improvement for MSLP estimation. The DeepTCNet also exceeds the Dvorak method for TC intensity estimation. Besides, DeepTCNet also surpasses MTCSWA in estimating critical wind radii of TCs, by 32%, 25%, 28%, and 38% for RMW, R64, R50, and R34, respectively. These results demonstrate that DeepTCNet can fundamentally improve the mining of IR imagery for TC intensity and wind radii estimation. Therefore, there is reason to believe the DeepTCNet method can be applied to enhancing the utilization of other observations of TCs.

More training samples with high-quality to build up deep learning models, such as aircraft observations, are particularly crucial for intensity and size estimates. However, the dilemma of data limitation for deep learning methods would be tough in the short term, especially in the research area of extreme weather systems like TCs. Therefore, exploring and developing appropriate approaches to incorporating physical knowledge of TCs into deep learning algorithms is the essential further direction. Besides, introducing the microwave sensor-based observations into the DeepTCNet could be crucial for producing more accurate estimates.

Acknowledgments. We acknowledge John Knaff and Haiyan Jiang for valuable feedback that greatly improved this work. We also thank Clark Evans for helping to improve this manuscript. One anonymous reviewer also enhanced this manuscript. This study is jointly supported by the National Key Research and Development Program of China under Grant 2017YFC1501604 and the National Natural Science Foundation of China under Grant 61827901.

Data availability statement. The intensity estimates of SATCON and ADT are downloaded from <http://tropic.ssec.wisc.edu>. Both the subjective Dvorak estimates and the TC Vitals data can be downloaded from <https://ftp.nhc.noaa.gov/atcf/archive/>. The MTCSWA data are obtained from https://rammb-data.cira.colostate.edu/tc_realtime. Interested readers can contact Jing-Yi Zhuo (jyzyhuo@smail.nju.edu.cn) to obtain the data underlying the findings of this paper.

REFERENCES

- Alber, M., and Coauthors, 2019: iNNvestigate neural networks! *J. Mach. Learn. Res.*, **20**, 93, www.jmlr.org/papers/volume20/18-540/18-540.pdf.
- Bender, M. A., T. P. Marchok, C. R. Sampson, J. A. Knaff, and M. J. Morin, 2017: Impact of storm size on prediction of storm track and intensity using the 2016 operational GFDL hurricane

- model. *Wea. Forecasting*, **32**, 1491–1508, <https://doi.org/10.1175/WAF-D-16-0220.1>.
- Bolton, T., and L. Zanna, 2019: Applications of deep learning to ocean data inference and subgrid parameterization. *J. Adv. Model. Earth Syst.*, **11**, 376–399, <https://doi.org/10.1029/2018MS001472>.
- Brennan, M. J., C. C. Hennon, and R. D. Knabb, 2009: The operational use of QuikSCAT ocean surface vector winds at the National Hurricane Center. *Wea. Forecasting*, **24**, 621–645, <https://doi.org/10.1175/2008WAF2222188.1>.
- Chavas, D. R., N. Lin, and K. Emanuel, 2015: A model for the complete radial structure of the tropical cyclone wind field. Part I: Comparison with observed structure. *J. Atmos. Sci.*, **72**, 3647–3662, <https://doi.org/10.1175/JAS-D-15-0014.1>.
- , K. A. Reed, and J. A. Knaff, 2017: Physical understanding of the tropical cyclone wind–pressure relationship. *Nat. Commun.*, **8**, 1360, <https://doi.org/10.1038/s41467-017-01546-9>.
- Chen, B. F., B. Chen, H. T. Lin, and R. L. Elsberry, 2019: Estimating tropical cyclone intensity by satellite imagery utilizing convolutional neural networks. *Wea. Forecasting*, **34**, 447–465, <https://doi.org/10.1175/WAF-D-18-0136.1>.
- Chollet, F., 2017: Xception: Deep learning with depthwise separable convolutions. *2017 IEEE Conf. on Computer Vision and Pattern Recognition*, Honolulu, HI, IEEE, 1800–1807, <https://doi.org/10.1109/CVPR.2017.195>.
- Cipolla, R., Y. Gal, and A. Kendall, 2018: Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, 7482–7491, <https://doi.org/10.1109/CVPR.2018.00781>.
- Courtney, J., and J. A. Knaff, 2009: Adapting the Knaff and Zehr wind–pressure relationship for operational use in tropical cyclone warning centres. *Aust. Meteor. Oceanogr. J.*, **58**, 167–179, <https://doi.org/10.22499/2.5803.002>.
- DeMaria, M., and Coauthors, 2013: Improvements to the operational tropical cyclone wind speed probability model. *Wea. Forecasting*, **28**, 586–602, <https://doi.org/10.1175/WAF-D-12-00116.1>.
- , C. R. Sampson, J. A. Knaff, and K. D. Musgrave, 2014: Is tropical cyclone intensity guidance improving? *Bull. Amer. Meteor. Soc.*, **95**, 387–398, <https://doi.org/10.1175/BAMS-D-12-00240.1>.
- Demuth, J. L., M. DeMaria, J. A. Knaff, and T. H. Vonder Haar, 2004: Evaluation of Advanced Microwave Sounding Unit tropical-cyclone intensity and size estimation algorithms. *J. Appl. Meteor.*, **43**, 282–296, [https://doi.org/10.1175/1520-0450\(2004\)043<0282:EOAMSU>2.0.CO;2](https://doi.org/10.1175/1520-0450(2004)043<0282:EOAMSU>2.0.CO;2).
- , —, and —, 2006: Improvement of Advanced Microwave Sounding Unit tropical cyclone intensity and size estimation algorithms. *J. Appl. Meteor. Climatol.*, **45**, 1573–1581, <https://doi.org/10.1175/JAM2429.1>.
- Dolling, K., E. A. Ritchie, and J. S. Tyo, 2016: The use of the deviation angle variance technique on geostationary satellite imagery to estimate tropical cyclone size parameters. *Wea. Forecasting*, **31**, 1625–1642, <https://doi.org/10.1175/WAF-D-16-0056.1>.
- Ebert-Uphoff, I., and K. Hilburn, 2020: Evaluation, tuning, and interpretation of neural networks for working with images in meteorological applications. *Bull. Amer. Meteor. Soc.*, **101**, E2149–E2170, <https://doi.org/10.1175/BAMS-D-20-0097.1>.
- Emanuel, K., and R. Rotunno, 2011: Self-stratification of tropical cyclone outflow. Part I: Implications for storm structure. *J. Atmos. Sci.*, **68**, 2236–2249, <https://doi.org/10.1175/JAS-D-10-05024.1>.
- Fetnat, G., A. Homaifar, and K. R. Knapp, 2013: Objective tropical cyclone intensity estimation using analogs of spatial features in satellite data. *Wea. Forecasting*, **28**, 1446–1459, <https://doi.org/10.1175/WAF-D-13-00006.1>.
- Goodfellow, I., Y. Bengio, and A. Courville, 2016: *Deep Learning*. MIT Press, 800 pp.
- Guo, X., and Z.-M. Tan, 2017: Tropical cyclone fullness: A new concept for interpreting storm intensity. *Geophys. Res. Lett.*, **44**, 4324–4331, <https://doi.org/10.1002/2017GL073680>.
- Ham, Y. G., J. H. Kim, and J. J. Luo, 2019: Deep learning for multi-year ENSO forecasts. *Nature*, **573**, 568–572, <https://doi.org/10.1038/s41586-019-1559-7>.
- Han, Y., G. J. Zhang, X. Huang and Y. Wang, 2020: A moist physics parameterization based on deep learning. *J. Adv. Model. Earth Syst.*, **12**, e2020MS002076, <https://doi.org/10.1029/2020MS002076>.
- He, K., X. Zhang, S. Ren, and J. Sun, 2016: Deep residual learning for image recognition. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, IEEE, 770–778, <https://doi.org/10.1109/CVPR.2016.90>.
- Hinton, G., and Coauthors, 2012: Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal Process. Mag.*, **29**, 82–97, <https://doi.org/10.1109/MSP.2012.2205597>.
- Jiang, H., C. Tao, and Y. Pei, 2019: Estimation of tropical cyclone intensity in the North Atlantic and northeastern Pacific Basins using TRMM satellite passive microwave observations. *J. Appl. Meteor. Climatol.*, **58**, 185–197, <https://doi.org/10.1175/JAMC-D-18-0094.1>.
- Kingma, D. P., and J. Ba, 2014: Adam: A method for stochastic optimization. <https://arxiv.org/abs/1412.6980>.
- Knaff, J. A., and R. M. Zehr, 2007: Reexamination of tropical cyclone wind–pressure relationships. *Wea. Forecasting*, **22**, 71–88, <https://doi.org/10.1175/WAF965.1>.
- , and C. R. Sampson, 2015: After a decade are Atlantic tropical cyclone gale force wind radii forecasts now skillful? *Wea. Forecasting*, **30**, 702–709, <https://doi.org/10.1175/WAF-D-14-00149.1>.
- , D. P. Brown, J. Courtney, G. M. Gallina, and J. L. Beven, 2010: An evaluation of Dvorak technique–based tropical cyclone intensity estimates. *Wea. Forecasting*, **25**, 1362–1379, <https://doi.org/10.1175/2010WAF2222375.1>.
- , M. DeMaria, D. A. Molenaar, C. R. Sampson, and M. G. Seybold, 2011: An automated, objective, multiple-satellite-platform tropical cyclone surface wind analysis. *J. Appl. Meteor. Climatol.*, **50**, 2149–2166, <https://doi.org/10.1175/2011JAMC2673.1>.
- , C. J. Slocum, K. D. Musgrave, C. R. Sampson, and B. R. Strahl, 2016: Using routinely available information to estimate tropical cyclone wind structure. *Mon. Wea. Rev.*, **144**, 1233–1247, <https://doi.org/10.1175/MWR-D-15-0267.1>.
- Knapp, K. R., and J. P. Kossin, 2007: New global tropical cyclone data set from ISCCP B1 geostationary satellite observations. *J. Appl. Remote Sens.*, **1**, 013505, <https://doi.org/10.1117/1.2712816>.
- , 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>.
- , and Coauthors, 2011: Globally gridded satellite observations for climate studies. *Bull. Amer. Meteor. Soc.*, **92**, 893–907, <https://doi.org/10.1175/2011BAMS3039.1>.
- Kossin, J. P., and Coauthors, 2007: Estimating hurricane wind structure in the absence of aircraft reconnaissance. *Wea. Forecasting*, **22**, 89–101, <https://doi.org/10.1175/WAF985.1>.

- 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, and D. J. Gagne II, 2019: Deep learning for spatially explicit prediction of synoptic-scale fronts. *Wea. Forecasting*, **34**, 1137–1160, <https://doi.org/10.1175/WAF-D-18-0183.1>.
- Lajoie, F., and K. Walsh, 2008: A technique to determine the radius of maximum wind of a tropical cyclone. *Wea. Forecasting*, **23**, 1007–1015, <https://doi.org/10.1175/2008WAF2007077.1>.
- Landsea, C. W., and J. L. Franklin, 2013: Atlantic hurricane database uncertainty and presentation of a new database format. *Mon. Wea. Rev.*, **141**, 3576–3592, <https://doi.org/10.1175/MWR-D-12-00254.1>.
- LeCun, Y., Y. Bengio, and G. Hinton, 2015: Deep learning. *Nature*, **521**, 436–444, <https://doi.org/10.1038/nature14539>.
- Liu, Y., and Coauthors, 2016: Application of deep convolutional neural networks for detecting extreme weather in climate datasets. <https://arxiv.org/abs/1605.01156>.
- McGovern, A., R. A. Lagerquist, D. J. Gagne, G. E. Jergensen, K. L. Elmore, C. R. Homeyer, and T. Smith, 2019: Making the black box more transparent: Understanding the physical implications of machine learning. *Bull. Amer. Meteor. Soc.*, **100**, 2175–2199, <https://doi.org/10.1175/BAMS-D-18-0195.1>.
- Mei, W., C. Pasquero, and F. Primeau, 2012: The effect of translation speed upon the intensity of tropical cyclones over the tropical ocean. *Geophys. Res. Lett.*, **39**, L07801, <https://doi.org/10.1029/2011GL050765>.
- Montavon, G., S. Lapuschkin, A. Binder, W. Samek, and K.-R. Müller, 2017: Explaining nonlinear classification decisions with deep Taylor decomposition. *Pattern Recognit.*, **65**, 211–222, <https://doi.org/10.1016/j.patcog.2016.11.008>.
- , W. Samek, and K.-R. Müller, 2018: Methods for interpreting and understanding deep neural networks. *Digit. Signal Process.*, **73**, 1–15, <https://doi.org/10.1016/j.dsp.2017.10.011>.
- Morris, M., and C. S. Ruf, 2017: Determining tropical cyclone surface wind speed structure and intensity with the CYGNSS satellite constellation. *J. Appl. Meteor. Climatol.*, **56**, 1847–1865, <https://doi.org/10.1175/JAMC-D-16-0375.1>.
- Mueller, K. J., M. DeMaria, J. Knaff, J. P. Kossin, and T. H. Vonder Haar, 2006: Objective estimation of tropical cyclone wind structure from infrared satellite data. *Wea. Forecasting*, **21**, 990–1005, <https://doi.org/10.1175/WAF955.1>.
- Olander, T. L., and C. S. Velden, 2007: The advanced Dvorak technique: Continued development of an objective scheme to estimate tropical cyclone intensity using geostationary infrared satellite imagery. *Wea. Forecasting*, **22**, 287–298, <https://doi.org/10.1175/WAF975.1>.
- , and —, 2019: The advanced Dvorak technique (ADT) for estimating tropical cyclone intensity: Update and new capabilities. *Wea. Forecasting*, **34**, 905–922, <https://doi.org/10.1175/WAF-D-19-0007.1>.
- Pradhan, R., R. S. Ayygun, M. Maskey, R. Ramachandran, and D. J. Cecil, 2018: Tropical cyclone intensity estimation using a deep convolutional neural network. *IEEE Trans. Image Process.*, **27**, 692–702, <https://doi.org/10.1109/TIP.2017.2766358>.
- Raissi, M., P. Perdikaris, and G. E. Karniadakis, 2019: Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.*, **378**, 686–707, <https://doi.org/10.1016/j.jcp.2018.10.045>.
- Rasp, S., M. S. Pritchard, and P. Gentine, 2018: Deep learning to represent subgrid processes in climate models. *Proc. Natl. Acad. Sci. USA*, **115**, 9684–9689, <https://doi.org/10.1073/pnas.1810286115>.
- Reichstein, M., G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, and N. Carvalhais, 2019: Deep learning and process understanding for data-driven earth system science. *Nature*, **566**, 195–204, <https://doi.org/10.1038/s41586-019-0912-1>.
- Reul, N., and Coauthors, 2017: A new generation of tropical cyclone size measurements from space. *Bull. Amer. Meteor. Soc.*, **98**, 2367–2385, <https://doi.org/10.1175/BAMS-D-15-00291.1>.
- Ritchie, E. A., G. Valliere-Kelley, M. F. Piñeros, and J. S. Tyo, 2012: Tropical cyclone intensity estimation in the North Atlantic basin using an improved deviation angle variance technique. *Wea. Forecasting*, **27**, 1264–1277, <https://doi.org/10.1175/WAF-D-11-00156.1>.
- Sampson, C. R., and A. J. Schrader, 2000: The Automated Tropical Cyclone Forecasting System (version 3.2). *Bull. Amer. Meteor. Soc.*, **81**, 1231–1240, [https://doi.org/10.1175/1520-0477\(2000\)081<1231:TATCFS>2.3.CO;2](https://doi.org/10.1175/1520-0477(2000)081<1231:TATCFS>2.3.CO;2).
- , P. A. Wittmann, and H. L. Tolman, 2010: Consistent tropical cyclone wind and wave forecasts for the U.S. Navy. *Wea. Forecasting*, **25**, 1293–1306, <https://doi.org/10.1175/2010WAF2222376.1>.
- , J. S. Goerss, J. A. Knaff, B. R. Strahl, E. M. Fukada, and E. A. Serra, 2018: Tropical cyclone gale wind radii estimates, forecasts, and error forecasts for the western North Pacific. *Wea. Forecasting*, **33**, 1081–1092, <https://doi.org/10.1175/WAF-D-17-0153.1>.
- Silver, D., and Coauthors, 2016: Mastering the game of Go with deep neural networks and tree search. *Nature*, **529**, 484–489, <https://doi.org/10.1038/nature16961>.
- Simonyan, K., and A. Zisserman, 2014: Very deep convolutional networks for large-scale image recognition. <https://arxiv.org/abs/1409.1556>.
- Smilkov, D., N. Thorat, B. Kim, F. Viégas, and M. Wattenberg, 2017: SmoothGrad: Removing noise by adding noise. <https://arxiv.org/abs/1706.03825>.
- Snaiki, R., and T. Wu, 2019: Knowledge-enhanced deep learning for simulation of tropical cyclone boundary-layer winds. *J. Wind Eng. Ind. Aerodyn.*, **194**, 103983, <https://doi.org/10.1016/j.jweia.2019.103983>.
- Sønderby, C. K., and Coauthors, 2020: MetNet: A neural weather model for precipitation forecasting. <https://arxiv.org/abs/2003.12140>.
- Stoffelen, A., 1998: Scatterometry. Ph.D. thesis, University of Utrecht, 195 pp., <https://dspace.library.uu.nl/bitstream/handle/1874/636/full.pdf>.
- , A. J. Verspeek, J. Vogelzang, and A. Verhoef, 2017: The CMOD7 geophysical model function for ASCAT and ERS wind retrievals. *IEEE J. Sel. Top. Appl. Earth Obs.*, **10**, 2123–2134, <https://doi.org/10.1109/JSTARS.2017.2681806>.
- Szegedy, C., V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, 2016: Rethinking the inception architecture for computer vision. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, IEEE, 2818–2826, <https://doi.org/10.1109/CVPR.2016.308>.
- Tallapragada, V., and Coauthors, 2014: Hurricane Weather Research and Forecasting (HWRF) model: 2014 scientific documentation. Developmental Testbed Center, 105 pp., https://www.dtcenter.org/HurrWRF/users/docs/scientific_documents/HWRFv3.6a_ScientificDoc.pdf.
- Torn, R. D., and C. Snyder, 2012: Uncertainty of tropical cyclone best-track information. *Wea. Forecasting*, **27**, 715–729, <https://doi.org/10.1175/WAF-D-11-00085.1>.

- Velden, C., and D. Herndon, 2020: A consensus approach for estimating tropical cyclone intensity from meteorological satellites: SATCON. *Wea. Forecasting*, **35**, 1645–1662, <https://doi.org/10.1175/WAF-D-20-0015.1>.
- , and Coauthors, 2006: The Dvorak tropical cyclone intensity estimation technique: A satellite-based method that has endured for over 30 years. *Bull. Amer. Meteor. Soc.*, **87**, 1195–1210, <https://doi.org/10.1175/BAMS-87-9-1195>.
- Weatherford, C., and W. Gray, 1988: Typhoon structure as revealed by aircraft reconnaissance. Part I: Data analysis and climatology. *Mon. Wea. Rev.*, **116**, 1032–1043, [https://doi.org/10.1175/1520-0493\(1988\)116<1032:TSARBA>2.0.CO;2](https://doi.org/10.1175/1520-0493(1988)116<1032:TSARBA>2.0.CO;2).
- Weyn, J. A., D. R. Durran and R. Caruana, 2020: Improving data-driven global weather prediction using deep convolutional neural networks on a cubed sphere. *J. Adv. Model. Earth Syst.*, **12**, e2020MS002109, <https://doi.org/10.1029/2020MS002109>.
- Wimmers, A., C. Velden, and J. H. Cossuth, 2019: Using deep learning to estimate tropical cyclone intensity from satellite passive microwave imagery. *Mon. Wea. Rev.*, **147**, 2261–2282, <https://doi.org/10.1175/MWR-D-18-0391.1>.
- Woo, S., J. Park, J. Y. Lee, and I. S. Kweon, 2018: CBAM: Convolutional block attention module. *European Conf. on Computer Vision (ECCV)*, Munich, Germany, ECCV, 3–19, https://doi.org/10.1007/978-3-030-01234-2_1.
- Zhang, Y., and Q. Yang, 2017: An overview of multi-task learning. *Natl. Sci. Rev.*, **5**, 30–43, <https://doi.org/10.1093/nsr/nwx105>.