A Deep Learning Model for Estimating Tropical Cyclone Wind Radius from Geostationary Satellite Infrared Imagery

Chong Wang[a,b] and Xiaofeng Li[a]

[a] *Key Laboratory of Ocean Circulation and Waves, Institute of Oceanology, Qingdao, China*
[b] *University of Chinese Academy of Sciences, Beijing, China*

ABSTRACT: This article developed a deep learning (DL) model for estimating tropical cyclone (TC) 34-, 50-, and 64-kt (1 kt ≈ 0.51 m s^{-1}) wind radii in four quadrants from infrared images in the global ocean. We collected 63 675 TC images from 2004 to 2016 and divided them into three periods (2004–12, 2013–14, and 2015–16) for model training, validation, and testing. First, four DL-based radius estimation models were developed to estimate the TC wind radius for each of the four quadrants. Then, the entire original images and the one-quarter-quadrant subimages were included in the model training for each quadrant. Last, we modified the mean absolute error (MAE) loss function in these DL-based models to reduce the side effect of an unbalanced distribution of wind radii and developed an asymmetric TC wind radius estimation model globally. The comparison of model results with the best-track data of TCs shows that the MAEs of 34-kt wind radius are 18.8, 19.5, 18.6, and 18.8 n mi (1 n mi = 1.852 km) for the northeast, southeast, southwest, and northwest quadrants, respectively. The MAEs of 50-kt wind radius are 11.3, 11.3, 11.1, and 10.8 n mi, respectively, and the MAEs of 64-kt wind radius are 8.9, 9.9, 9.2, and 8.7 n mi, respectively. These results represent a 12.1%–35.5% improvement over existing methods in the literature. In addition, the DL-based models were interpreted with two deep visualization toolboxes. The results indicate that the TC eye, cloud, and TC spiral structure are the main factors that affect the model performance.

KEYWORDS: Tropical cyclones; Remote sensing; Deep learning

1. Introduction

Tropical cyclones (TCs) are intense weather processes generated over tropical oceans and could cause enormous damage to human lives and properties in the coastal regions owing to severe flooding, destructive wind, and coastal inundation from storm surges. Therefore, TC monitoring is critical to mitigating damage in coastal areas. TC wind radius represents the size of the TC influence area, and an accurate estimate of the TC wind radius is critical for emergency responders to provide TC warning information (DeMaria et al. 2013; Bender et al. 2017). TC wind radius is defined into three categories according to wind speed, including the gale-force winds (R34; 34 kt, where 1 kt = 0.51 m s^{-1}), destructive winds (R50; 50 kt), and hurricane-force winds (R64; 64 kt) as measured radially in four geographic quadrants [i.e., in the northeast (NE), southeast (SE), southwest (SW), and northwest (NW) directions].

Satellite observations are widely used in TC wind radii estimation (Landsea and Franklin 2013). According to the type of satellite, satellite-based TC wind radius estimation methods are classified as scatterometers-based (Halpern 1993; Mears et al. 2001; Ebuchi et al. 2002; Chavas and Emanuel 2010; Chan and Chan 2012; Landsea and Franklin 2013), spaceborne-synthetic-aperture-radars-based (SAR-based) (Zhang and Perrie 2012; Zhang et al. 2014; Horstmann et al. 2015; Mouche et al. 2017; Fan et al. 2020; Mouche et al. 2019; Zhang et al. 2021), and satellite-infrared-image-based (Mueller et al. 2006; Kossin et al. 2007; Knaff et al. 2016; Dolling et al. 2016; Zhuo and Tan 2021) methods.

The scatterometer-based (e.g., ASCAT, QuikSCAT, and OSCAT scatterometers) and SAR-based methods are similar. These methods first measure the surface wind and then extract the TC wind radius from the surface wind (Reul et al. 2012; Yueh et al. 2016; Reul et al. 2017; Fore et al. 2018; Sun et al. 2019; Mouche et al. 2019; Zhang et al. 2021). However, the scatterometer-based and SAR-based methods are not a static observation of a region, frequently only observe a part of TC, and are affected by rainfall, have certain limitations.

The satellite-infrared-image-based methods are real-time and objective. Mueller et al. (2006) and Kossin et al. (2007) used infrared images, the maximum sustained wind speed, and the radius of maximum wind to establish a multiple linear regression model to estimate TC wind radius. Dolling et al. (2016) used deviation angle variance calculated from infrared images, TC age, and SST (sea surface temperature) to establish the TC wind radii estimation with a multiple linear regression model. They found that the unbalanced distribution of samples with different TC wind radii may result in large errors for large TC wind radii with small sample sizes (Dolling et al. 2016). The satellite-infrared-image-based methods have not yet determined the physical relationships between the structure of the TC wind and the top of the cloud field. Therefore, developing objective methods to interpret complex TC dynamics from satellite observations is critical.

Several satellite-based TC wind radii estimation methods are available, each with advantages and disadvantages. The errors of TC wind radius from satellite-based methods have been estimated to be 10%–60%, depending on the quality and quantity of the available observations (Landsea and Franklin 2013). Therefore, it is also a great challenge to

---

Corresponding author: Xiaofeng Li, xiaofeng.li@ieee.org

develop an accurate satellite-based method to estimate TC wind radii.

Deep learning (DL) is a type of artificial intelligence algorithm, including deep neural networks, recurrent neural networks, and convolutional neural networks (CNN), among others (Goodfellow et al. 2016). CNNs are widely used in the field of image processing. A CNN is a feed-forward deep neural network inspired by the biological natural visual cognition mechanism (Goodfellow et al. 2016) and has achieved great success in image classification, target recognition, and other research fields (LeCun et al. 1995; Sun et al. 2014; Li et al. 2020, 2022). Recently, Zhuo and Tan (2021) developed a CNN model to estimate TC wind radii and obtain good accuracy. However, the asymmetric TC wind structure was not considered, and they used the nonzero-azimuthal average of data from four quadrants and replaced wind radii for four quadrants.

This study addresses three key issues in TC radii estimate from geostationary infrared imagery. First, there is no deep learning model for estimating asymmetric TC wind radius. Second, the unbalanced distribution of samples with different TC wind radii could cause large errors in TCs with large wind radii. Third, TCs in different oceans have different structural characteristics. The first goal is to overcome the difficulties mentioned above and develop an objective and efficient TC wind radii estimation method for estimating TC wind radius in four quadrants, and the second goal is to explore the features in infrared images that are more related to estimating TC wind radius. Therefore, we developed a group of deep learning models (DL-TCR) and selected the best to estimate the TC wind radius over the global ocean using infrared images.

Sections 2 and 3 introduce the dataset and the DL-based models. Then, the model results analysis and discussion are given in section 4. The summaries are presented in section 5.

2. Data and method

a. Geostationary satellite infrared TC images

The TC infrared images used in this study were from the Hurricane Satellite dataset (HURSAT-B1), which was compiled from the global constellation of geostationary satellites (GOES, GMS, Meteosat series, etc.) between 1979 and 2016 (Knapp and Kossin 2007). HURSAT-B1 has about 380 000 satellite images from the infrared ($\sim$11 μm), near-infrared ($\sim$3.9 μm), split window ($\sim$12 μm), and water vapor ($\sim$6.7 μm) bands, and so on. HURSAT-B1 includes TC images in the Atlantic, Pacific, and Indian Ocean basins, with the most data for the Pacific. The time interval of HURSAT-B1 is 3 h. The infrared images were gridded to 8 km, with grid centers fixed on the TC center, and the size of each infrared image is 301×301 pixels. About 0.5% of images were removed because of poor imaging quality (strip loss, etc.).

b. Best-track dataset of TCs

The International Best Track Archive for Climate Stewardship (IBTrACS) provided by the National Oceanic and Atmospheric Administration was used to label infrared images (Knapp et al. 2010). IBTrACS data combine R34 (34-kt wind
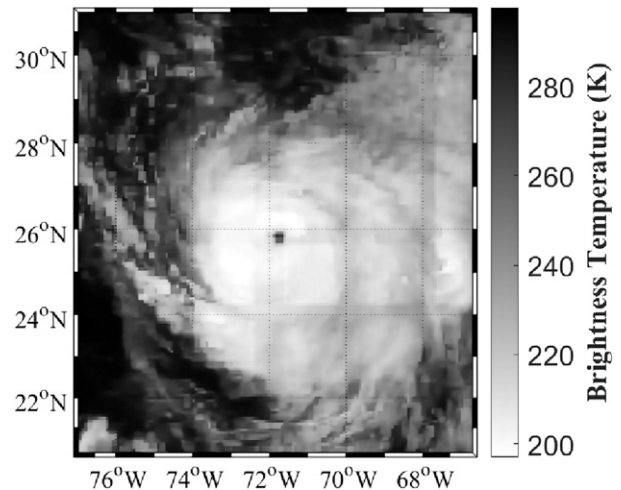

FIG. 1. Brightness temperature (K) image of TC Joaquin at 1500 UTC 3 Oct 2015, with the asymmetric TC wind radius in the NE, SE, SW, and NW quadrants of 155, 180, 130, and 110 n mi, respectively.

radii), R50 (50-kt wind radii), and R64 (64-kt wind radii) for each quadrant (NE, SE, SW, and NW). The best-track data have been widely used to validate the model results of wind radius (Zhuo and Tan 2021; Zhang et al. 2021). We also ran the quality control of the dataset. When wind radii labels did not meet the R34 > R50 > R64, R34 $\geq$ radius of maximum wind (RMW), R50 $\geq$ RMW, and R64 $\geq$ RMW criteria, they were considered to be bad data and were eliminated (Kossin et al. 2007; Zhuo and Tan 2021).

c. Data preprocessing

Data normalization is crucial before neural network training to speed up calculations and get good results (Sola and Sevilla 1997). In this article, infrared images were linearly transformed to the interval [0, 1] by

$$y = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \quad (1)$$

where $x_{\min}$ and $x_{\max}$ are the minimum and maximum values of the brightness temperature from all images, and y is the normalized pixel value limited to [0, 1]. Different ocean basins (the North Atlantic, the Northwest Pacific, the Northeast Pacific, and the Southern Hemisphere) will be normalized separately.

For each image, we extracted a 151×151 pixel image centered on the TC center from IBTrACS. An example is shown in Fig. 1. Continuous-time images proved useful for TC wind radii estimation (Zhuo and Tan 2021). The current image and the image 3 h prior were used as inputs to our models.

We obtained 63 076, 27 210, and 19 937 sample images for R34, R50, and R64 wind radii, respectively, between 2004 and 2016. In artificial intelligence (AI) model development, the dataset is usually divided into three groups: training, validation, and test. In the model training process, the training

TABLE 1. The training, validation, and test datasets for TC wind radii of R34, R50, and R64.

Name	Training data (2004–12)	Validation data (2013–14)	Test data (2015–16)
R34	43 466	11 231	8379
R50	18 192	4320	4698
R64	13 839	2719	3379

data are used to update the model's internal parameters; the validation data are used to assist model training and prevent model overfitting; the test data are used for model evaluation. The data were divided into training data from 2004 to 2012, validation data from 2013 to 2014, and test data from 2015 to 2016 (Table 1).

We developed different models for different ocean basins. The models for each ocean are separately trained, validated, and tested using the TC data of that ocean. The north Indian Ocean was not considered independently because of the small amount of data. The number of training, validation, and test data are shown in Table 2.

In the IBTrACS data, the relative uncertainty of data using only satellite observation is between 35% and 53%, and the same data using aircraft reconnaissance is between 28% and 38% (Landsea and Franklin 2013). Therefore, more aircraft observations are available for the IBTrACS data in the North Atlantic. On the other hand, in DL, more stable data allow for more accurate optimization of model parameters (Goodfellow et al. 2016), and more aircraft observations are available in the North Atlantic with minimal uncertainty in the IBTrACS data. Thus, the data over the North Atlantic between 2004 and 2016 were used for training, validation, and testing to obtain the optimal DL-TCR model parameter. Once we found the optimal DL-TCR model parameter, the global TC wind radius estimation models were further developed.

3. TC wind radius estimation model

a. Development of the DL-TCR model

The DL model is based on a CNN architecture that contains convolutional, pooling, and fully connected layers (Srivastava et al. 2014; Goodfellow et al. 2016). The input images were extracted image features at the convolutional layer and filtered the maximum value of these features at the max-pooling layer. The fully connected layer learns all the relationships between features and wind radii. The CNN is designed as a feedforward network and trained with the backpropagation algorithm. The errors back propagate in a network; the CNN is optimized by updating the weights and biases to minimize the loss function (Srivastava et al. 2014; Goodfellow et al. 2016). CNN has been widely used in atmospheric and ocean science, e.g., estimate TC intensity (Chen et al. 2019; Wimmers et al. 2019; Wang et al. 2022), estimate TC wind radii (Zhuo and Tan 2021), detect and forecast ocean phenomena, sea ice, and TC rapid intensification (Zhang and Li 2021; Liu et al. 2021; S. Zhang et al. 2022; Ren et al. 2022; X. Zhang et al. 2022; Griffin et al. 2022). Widely used CNN models include VGGNet, ResNet, and GoogLeNet, etc. (Krizhevsky et al. 2012; Simonyan and Zisserman 2014; Szegedy et al. 2016; He et al. 2016; Chollet 2017). Different CNN models have different advantages and drawbacks. Zhuo and Tan (2021) have proven that the VGGNet is suitable for estimating TC wind radii. Therefore, we selected the VGGNet as the basic model.

Since the structure of TC wind radii is asymmetrical, making it hard to train the CNN model using the single original image, we need more useful information or features to help improve model performance. Therefore, four improved CNN models (DL-TCR) were developed to estimate R34, R50, and R64 in the NE, NW, SE, and SW quadrants. As shown in Fig. 2, the DL-TCR model contains the standard CNN architecture (Part A, labeled "A" in Fig. 2) and an additional Part B (labeled "B" in Fig. 2), which is used to enhance the learning of the features of the corresponding quadrant TC wind radii. It means that four images are input to the DL-TCR model, two being the original images at the current time and 3 h prior, and two corresponding quadrant images (Fig. 2). For example, the subimage of the NE quadrant is the NE quadrant wind radii DL-TCR model input and is used to estimate the TC wind radius at the NE quadrant. The features of the two images were extracted at the convolution layer and connected at the full connection layer. Last, the TC wind radius at the NE quadrant is obtained.

The channel size of DL-TCR convolutional layers and fully connected layers are shown in Fig. 2. The spatial and channel

TABLE 2. The number of training, validation, and test datasets for TC wind radii of R34, R50, and R64 in the North Atlantic, Northwest Pacific, Northeast Pacific, and Southern Hemisphere.

| Name	Training data (2004–12)	Validation data (2013–14)	Test data (2015–16)
R34—North Atlantic	8502	860	940
R34—Northwest Pacific	17 094	2599	2769
R34—Northeast Pacific	6882	1875	2301
R34—The Southern Hemisphere	14 012	2870	2369
R50—North Atlantic	3677	304	483
R50—Northwest Pacific	8553	1815	1608
R50—Northeast Pacific	3168	979	1435
R50—The Southern Hemisphere	6653	1222	1172
R64—North Atlantic	2401	160	329
R64—Northwest Pacific	5726	1358	1326
R64—Northeast Pacific	1877	432	961
R64—The Southern Hemisphere	3835	769	763
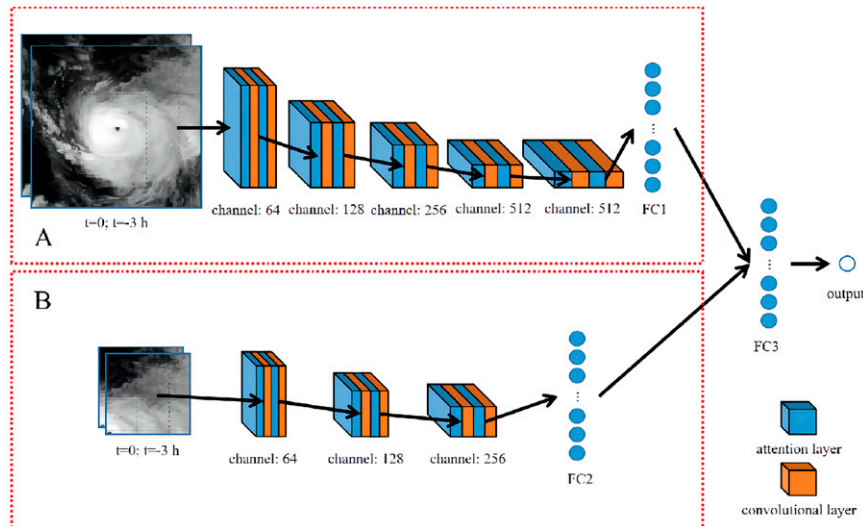
FIG. 2. The architecture of the DL-TCR model for estimating TC wind radii (FC means the fully connected layer), showing Parts A and B.

attention layers were added before each convolutional layer. These spatial and channel attention layers add weight to the important parts of the images. The areas in images DL-TCR models considered important for TC wind radii estimation were found by analyzing the attention layer, as discussed in section 4d.

b. Optimize the DL-TCR model architecture

The model input and parameters affect the DL-TCR model's performance. In section 4a, several DL-TCR models with different inputs or model parameters are designed to obtain the optimal model for TC wind radius estimation from IR images based on the North Atlantic validation data. The nonlinear Relu was adopted to activate the convolution and the full connection layers. The optimization function is Adam (Kingma and Ba 2014). The loss function is the mean absolute error (MAE). The dropout layer was used for each full connection layer (Srivastava et al. 2014). Each DL-TCR model is composed of 12 submodels for the estimation of tropical cyclone wind radius in four quadrants of R34, R50, and R64.

We applied the optimal model to the global TC wind radii estimation and discussed the results based on the global ocean test data in section 4b. A modified loss function is used to decrease the side effect of the unbalanced TC wind radii data with different sizes in section 4c. Note that each model described

in section 4 was trained three times. To evaluate the model performance, we take the 3-times-averaged model results.

4. Results and discussion

a. DL-TCR model performance with different model architecture

This section compares the results of the models with different inputs and model configurations. For example, the different architectures (Parts A and B) of the DL-TCR model are shown in Fig. 2.

Table 3 and Fig. 3 show the TC wind radii estimation results for the DL-TCR models using validation data from different model configurations and input. The MAE of the DL-TCR 1 model for NE, SE, SW, and NW quadrants is 20.9, 22.0, 18.5, and 19.8 n mi (1 n mi = 1.852 km), respectively. The results show that adding the channel and spatial attention layers improves model performance by enhancing the model's focus on key regions. DL-TCR 2 has achieved 1.4%, 2.3%, 3.2%, and 2.5% improvements in the NE, SE, SW, and NW quadrants, respectively. We then investigated the influence of various input data on model performance. Combining the original image and corresponding quadrant subimage as DL-TCR 3 input data obtain the best results in Table 3. The MAEs for NE, SE, SW, and NW quadrants are 19.5, 21.0, 17.0, and 18.9 n mi, respectively.

TABLE 3. The mean absolute errors (n mi) over the North Atlantic validation datasets of R34 TC wind radii estimation with different model configurations and input. The VGGNet is shown in Fig. 2 (labeled "A").

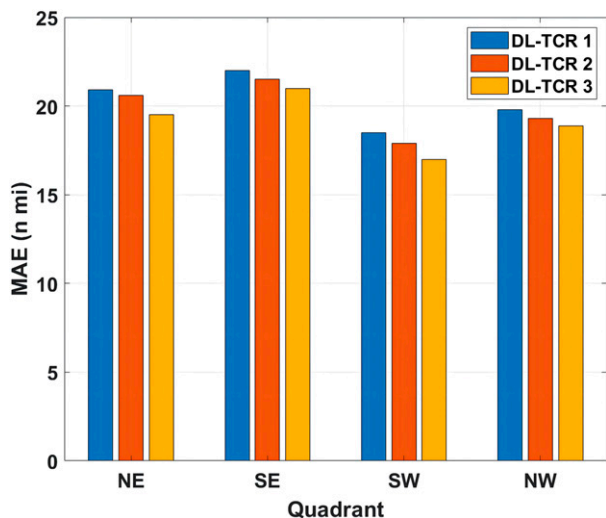| Model identifier	Input	Configurations	NE	SE	SW	NW
DL-TCR 1	Original image	VGGNet	20.9	22.0	18.5	19.8
DL-TCR 2	Original image	VGGNet + attention layer	20.6 (1.4%)	21.5 (2.3%)	17.9 (3.2%)	19.3 (2.5%)
DL-TCR 3	Original image + corresponding quadrant image	VGGNet + attention layer	19.5 (6.7%)	21.0 (4.5%)	17.0 (8.1%)	18.9 (4.5%)

FIG. 3. The mean absolute errors for each quadrant R34 TC wind radii estimation of the DL-TCR 1 (blue), DL-TCR 2 (red), and DL-TCR 3 (yellow) models on the North Atlantic validation datasets.

Figure 4 shows the R34 TC wind radius scatterplots from the best-track data and the DL-TCR 3 on the North Atlantic validation datasets, illustrating the R34 TC wind radius asymmetry. The maximal and minimal sizes are different in the four quadrants. The NE, SE, SW, and NW quadrants' correlation coefficients are about 0.80. As shown in Fig. 4, for the large size TCs (TC wind radii > 140 n mi), the DL-TCR underestimates the size of R34. The phenomenon is due to the unbalanced distribution of TC wind radius samples, which is also a common problem in other DL methods (Pradhan et al. 2018; Chen et al. 2019). i.e., it is difficult to obtain good results for the unbalanced data distribution. And we will discuss the problem in section 4c.

Once we found the best DL-TCR 3 model configurations, R50 and R64 TC wind radii estimation models were further developed. The comparisons of R50 and R64 between the best-track data and the DL-TCR 3 on the North Atlantic validation datasets are shown in Table 4 and Figs. 5 and 6. The R50 TC wind radius has MAEs of 11.0, 11.9, 9.8, and 10.1 n mi for the NE, SE, SW, and NW quadrants, respectively. The R64 TC wind radius has MAEs of 8.6, 9.4, 9.3, and 8.3 n mi for the same quadrants. The MAEs of R34 are greater than the error of R54 and greater than the error of R64. The R50 has the correlation coefficient of 0.75, 0.74, 0.78, and 0.72 for the NE, SE, SW, and NW quadrants, respectively, and the R64 TC wind radius has the correlation coefficient of 0.72, 0.76, 0.75, and 0.71 for the same quadrants. Although DL-TCR 3 obtains the lowest MAE of TC wind radius, it also underestimates the large wind radius data.

b. The global TC wind radii estimation models

In this section, we divide the globe into four regions: the North Atlantic, the Northwest Pacific, the Northeast Pacific, and the Southern Hemisphere. The Northern Indian Ocean was not considered in this paper because of the small amount of data. We trained different models with DL-TCR 3 model configurations for different regions, and the models for the North Atlantic have been described in section 4a. The models for each region are separately trained, validated, and tested using the TC data of that region. The four regions' TC wind radii estimation models form the global TC wind radii estimation models. The results were evaluated by comparing the model result with the best-track data on test datasets described in section 2.

Table 5 and Fig. 7 show the MAEs of the TC wind radius estimation model in different oceans. In the Northwest Pacific, the MAEs of the R34 TC wind radius are 18.6, 17.6, 17.3, and 19.0 n mi for the NE, SE, SW, and NW, respectively. The MAEs of R50 TC wind radius are 11.2, 11.3, 11.4, and 10.7 n mi, while the MAEs of R64 TC wind radius are 8.8, 9.6, 8.9, and 8.5 n mi for the same quadrants. In the Northeast Pacific, the MAEs of the R34 TC wind radius are 17.5, 17.3, 16.9, and 16.9 n mi for the NE, SE, SW, and NW, respectively. The MAEs of R50 TC wind radius are 10.5, 10.9, 11.3, and 10.2 n mi, and the MAEs of R64 TC wind radius are 8.7, 9.3, 8.7, and 8.4 n mi for the same quadrants. In the Southern Hemisphere, the MAEs of R34 TC wind radius are 18.9, 21.8, 20.7, and 18.4 n mi for the NE, SE, SW, and NW, respectively. The MAEs of R50 TC wind radius are 11.6, 12.1, 10.2, and 10.9 n mi, and the MAEs of R64 TC wind radius are 9.2, 10.4, 9.7, and 9.1 n mi for the same quadrants. The MAEs in the Southern Hemispheres are greater than in the North Atlantic, greater than in the Northwest Pacific, and greater than in the Northeast Pacific. The results indicate that the TC characteristics of different oceans are different, and the model has different learning abilities.

c. Unbalanced distribution of TC wind radii size

Figures 4–6 show that the unbalanced distribution of TC wind radius causes the models to underestimate the large TC wind radius. In this section, R34 TC wind radii will be used as an example to discuss the solution.

As Table 6 shows, the MAEs are not satisfactory for estimating large TC wind radii with unbalanced training samples. The biases of the TC wind radii larger than 210 n mi are −34.0, −41.3, −49.4, and −48.0 n mi for the NE, SE, SW, and NW quadrants, respectively. To see why, assume 100 training samples for which 98 have a TC wind radius of 100 n mi (hereinafter referred to as Part I) and 2 have a wind radius of 200 n mi (hereinafter referred to as Part II). Using the traditional MAE loss function (function 2.1), if the model estimates Part I as 110 n mi, the loss value of Part I is 9.8. If the model estimates Part II as 190 n mi, the loss value of Part II is 0.2. As a result, the combined loss is 10. Obviously, improving the accuracy of Part I is more helpful in reducing the loss value. Therefore, the model tends to learn more from Part I first. However, the model does not estimate perfectly for all samples, and when much attention is paid to Part I, Part II will be neglected.
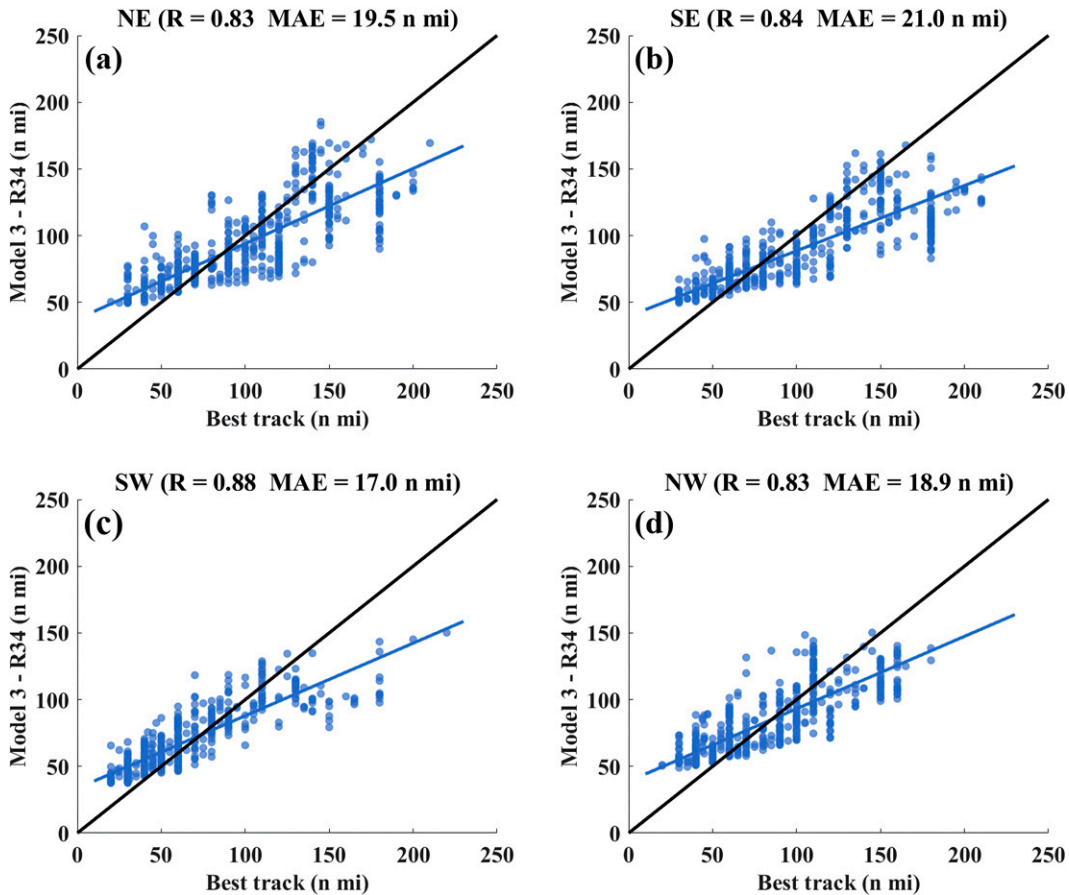
FIG. 4. Comparison of the R34 TC wind radii from the best-track data with the results of the DL-TCR 3 models on the North Atlantic validation datasets: (a) NE quadrant, (b) SE quadrant, (c) SW quadrant, and (d) NW quadrant.

To reduce the side effect of an unbalanced distribution of TC wind radius, we can add or remove parts of the data to keep the number of TC wind radii data in each size (Pradhan et al. 2018). However, this reduction would dramatically reduce the number of training data. Another method is to increase the weight of the sample with fewer data in the loss, that is, focal loss, which shows its superiority in the field of target recognition (Lin et al. 2017). Therefore, we modified the MAE loss function, named MAE-weight, which is a combination of the traditional MAE loss function (function 2.1) and the modified MAE loss function (function 2.2):

$$
\text{loss} = \begin{cases} \dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\text{abs}(y_{\text{true}} - y_{\text{pred}}) & (\text{if } y_{\text{true}} < 140; \text{ function 2.1}) \\[2ex] \dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\text{abs}(y_{\text{true}} - y_{\text{pred}}) \times a_t \times \dfrac{y_{\text{true}} - y_{\text{true}}^{\min}}{y_{\text{true}}^{\max} - y_{\text{true}}^{\min}} & (\text{if } y_{\text{true}} \geq 140; \text{ function 2.2}) \end{cases}
\tag{2}
$$

TABLE 4. The mean absolute errors (n mi) on the North Atlantic validation datasets of R34, R50, and R64 TC wind radii estimation using DL-TCR 3 models.

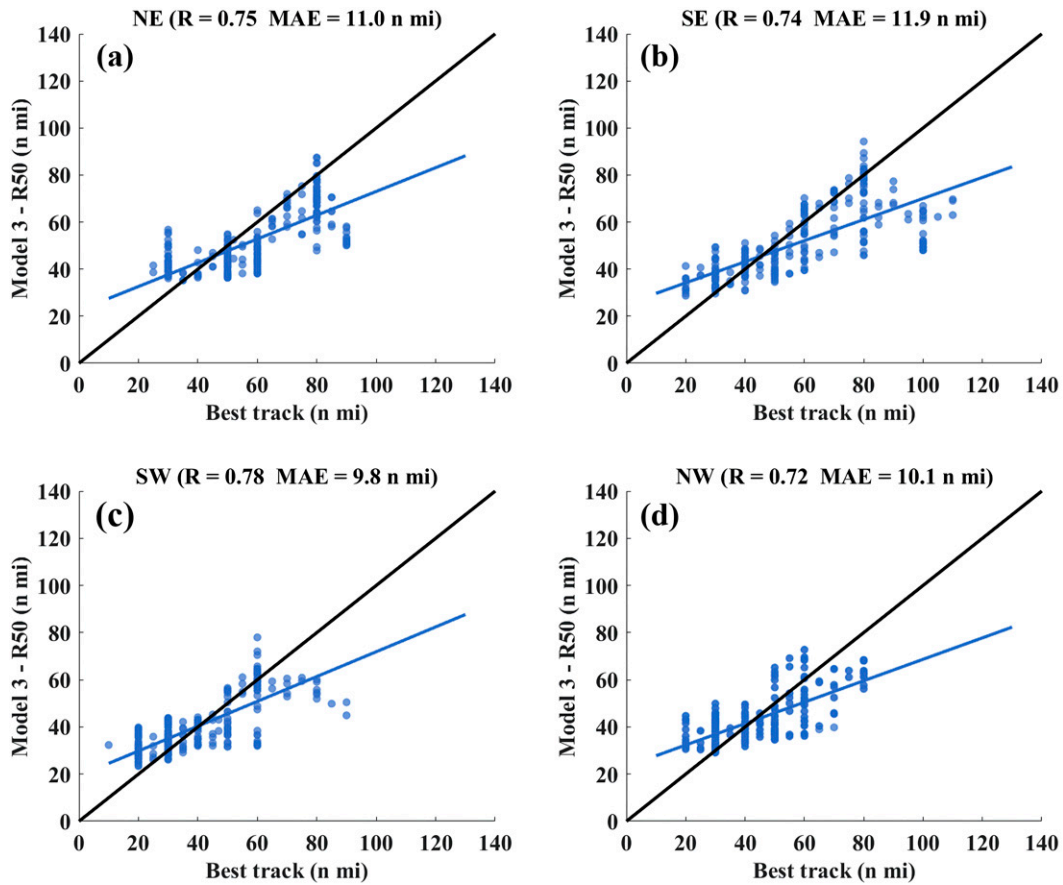|	Input	Configurations	NE	SE	SW	NW
R34	Original image + corresponding quadrant image	VGGNet + attention layer	19.5	21.0	17.0	18.9
R50	Original image + corresponding quadrant image	VGGNet + attention layer	11.0	11.9	9.8	10.1
R64	Original image + corresponding quadrant image	VGGNet + attention layer	8.6	9.4	9.3	8.3

FIG. 5. As in Fig. 4, but for R50.

where N is the number of samples; abs() calculates the absolute value; y_{true} is the true value from the best-track data; y_{pred} is the model output; a_t is the weight coefficient (in this article, $a_t = 5$); and $y_{\text{true}}^{\min}$ and $y_{\text{true}}^{\max}$ are the minimum and maximum of the samples, respectively.

As shown in function 2.2, we increase the weight of the samples in which the TC wind radii are larger than 140 n mi. When the TC wind radius is greater than 140 n mi, $a_t \times (y_{\text{true}} - y_{\text{true}}^{\min})/(y_{\text{true}}^{\max} - y_{\text{true}}^{\min})$ increases with the increase of the TC wind radius. Therefore, the cost (loss value) of underestimating the large TC wind radius samples is increased so that the DL model will enhance the learning of the large TC wind radius samples. Table 6 and Fig. 8 show that the underestimation at the large wind radii becomes lower using MAE-weight in DL-TCR 3. The bias of R34 greater than 140 n mi decreased for the NE, SE, SW, and NW quadrants. On the contrary, the bias of R34 between 70 and 140 n mi increased from -6.7, -6.8, -7.2, and -6.7 n mi to 3.7, 3.0, 1.9, and 3.7 n mi for the same quadrants. The bias of R34 less than 70 n mi increased. For R34 greater than 70 n mi, the model's bias using the MAE-weight loss function is closer to 0 than the MAE loss function. There is little variation in the MAEs of the R34 TC wind radius. The results of R34 less than 70 n mi indicate that the increasing weights can improve the estimation of

the model for a specific sample but at the expense of performance for other samples. A balance between enhancement and sacrifice can be found by adjusting the weighting factor.

The results show that the side effects of the unbalanced distribution of TC wind radius can be reduced by modifying the MAE loss function in the DL-TCR model. This MAE-weight loss function is also used for R50 and R64 TC wind radius estimation. The DL-TCR 3 model with the corresponding quadrant image, attention layer, and MAE-weight loss function is considered the best global TC wind radii estimation model. As shown in Table 7, the global TC wind radii estimation models obtain MAEs of R34 of 18.8, 19.5, 18.6, and 18.8 n mi for the NE, SE, SW, and NW quadrants, MAEs of R50 of 11.3, 11.3, 11.1, and 10.8 n mi for the same quadrants, and MAEs of R64 of 8.9, 9.9, 9.2, and 8.7 n mi for the same quadrants. To the authors' knowledge, DL-TCR is the first DL model for estimating asymmetries TC wind radius from infrared satellite images. We compared the TC wind radii estimation model's performance in estimating TC wind radii with the techniques using infrared images listed in Table 7.

In comparison with Kossin's and Zhuo's methods (Kossin et al. 2007; Zhuo and Tan 2021), our method can estimate the asymmetric TC wind radius. Using data from the same year as
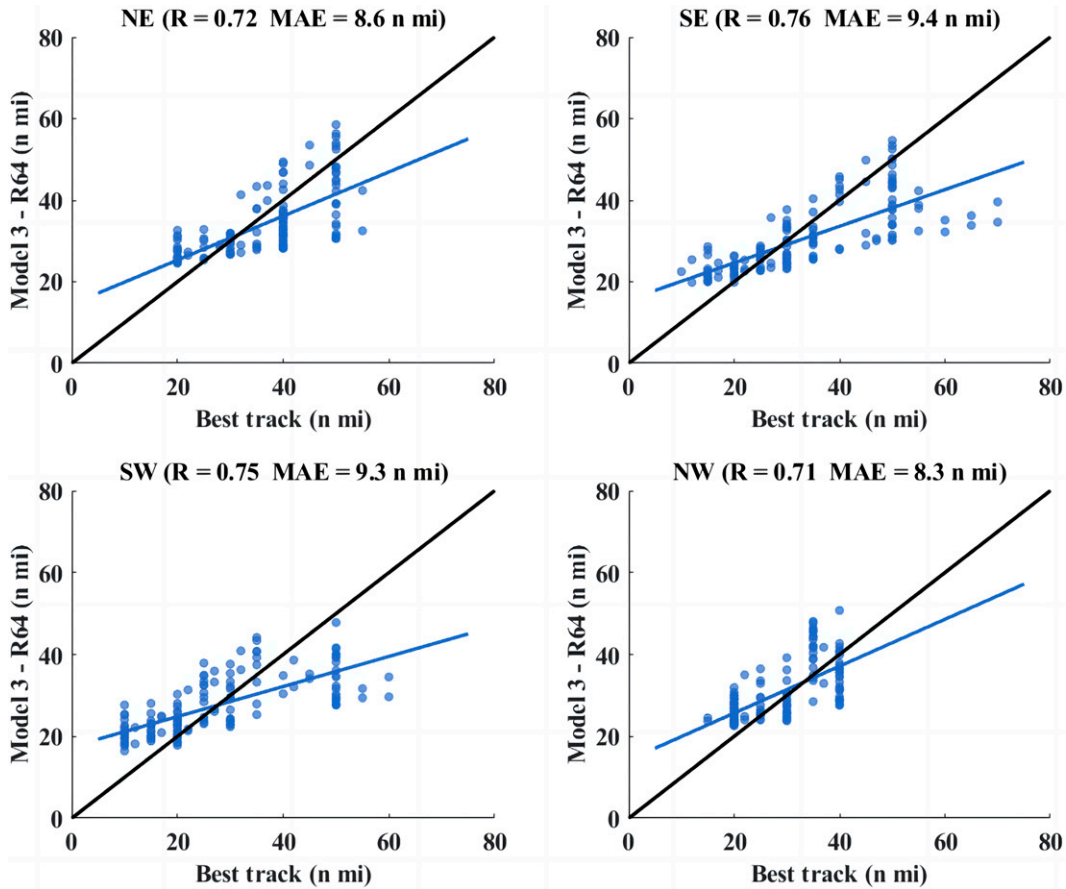
FIG. 6. As in Fig. 4, but for R64.

test data and other data as training and validation data, we compare the performance of our and other asymmetric TC wind radius estimation methods (Knaff et al. 2016; Dolling et al. 2016). The results show that our method exhibits smaller MAEs than the traditional method for estimating all four quadrants of the asymmetric TC wind radii globally (Knaff et al. 2016; Dolling et al. 2016).

d. Visualization of DL-TCR

DL is commonly recognized as a black box, extracting many features from infrared images and establishing relationships between these features and the output. However, the black box does not explain how to understand the DL-based

model and TCs. The interpretability of DL models is a challenge and a hot research topic in DL. To our knowledge, the weight and heat maps are currently the most common and effective methods to explain the DL model. The weight map shows the weights calculated by the attention layer. The heat map is the effect of the input image on the model output. Since the DL-TCR model extracts different features from different samples, we averaged the weights of the attention layer to analyze all the TC samples in test data and used the heat map of the input layer to analyze the individual samples. The toolboxes developed by Kotikalapudi et al. (2017), and Lundberg and Lee (2017) were used for the model interpretation work. However, note that most publicly AI interpretable

TABLE 5. The mean absolute errors (n mi) on test datasets of R34, R50, and R64 TC wind radii estimation in different oceans.

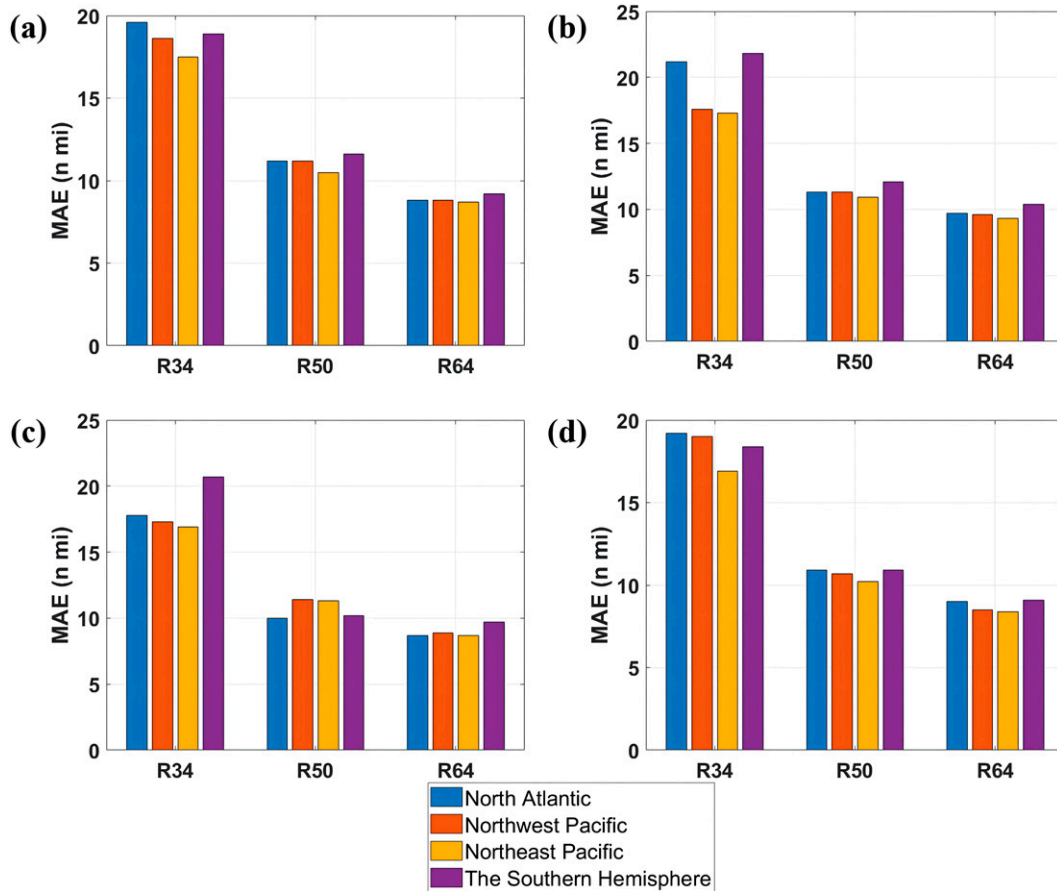| Test data	R34				R50				R64			
	NE	SE	SW	NW	NE	SE	SW	NW	NE	SE	SW	NW
North Atlantic	19.6	21.2	17.8	19.2	11.2	11.3	10.0	10.9	8.8	9.7	8.7	9.0
Northwest Pacific	18.6	17.6	17.3	19.0	11.2	11.3	11.4	10.7	8.8	9.6	8.9	8.5
Northeast Pacific	17.5	17.3	16.9	16.9	10.5	10.9	11.3	10.2	8.7	9.3	8.7	8.4
Southern Hemisphere	18.9	21.8	20.7	18.4	11.6	12.1	10.2	10.9	9.2	10.4	9.7	9.1
Global	18.5	19.1	18.2	18.3	11.1	10.8	10.9	10.6	8.9	9.7	9.0	8.6

FIG. 7. The mean absolute errors on test datasets of R34, R50, and R64 TC wind radii estimation in different oceans.

methods are for classification, not regression problems. Therefore, we use the Shap values and saliency maps methods, which can be extended to regression applications, to calculate the heat map (Lundberg and Lee 2017; Simonyan et al. 2013). In the weight and heat maps, a larger value represents the feature with higher relevance to the TC wind radii.

Figure 9 shows the weights and differences of the first attention layer behind the input layer of the DL-TCR model. Figures 9a, 9d, 9g, and 9j are the weights of the NE, SE, SW, and NW quadrants TC wind radii estimation models on the

original image. The weight distributions of the four models for the original images are very similar, differing only in value. The distribution of weights in the four quadrants is asymmetric, indicating that our models can learn the asymmetric structure of the TC. The NE quadrant has more weight than the other quadrants, with the largest weights near the center of the TC and decreasing toward the periphery. The results indicate that our model considers the area closer to the TC center more important for estimating the TC wind radius.

TABLE 6. The bias (n mi) of R34 TC wind radii estimation with different loss functions in the global ocean (including the Northwest Pacific, Northeast Pacific, and the Southern Hemisphere) on test datasets. The value in the parentheses represents the number of samples. Note that a TC may have observation records from multiple satellite observations simultaneously, and only one record will be kept for these samples.

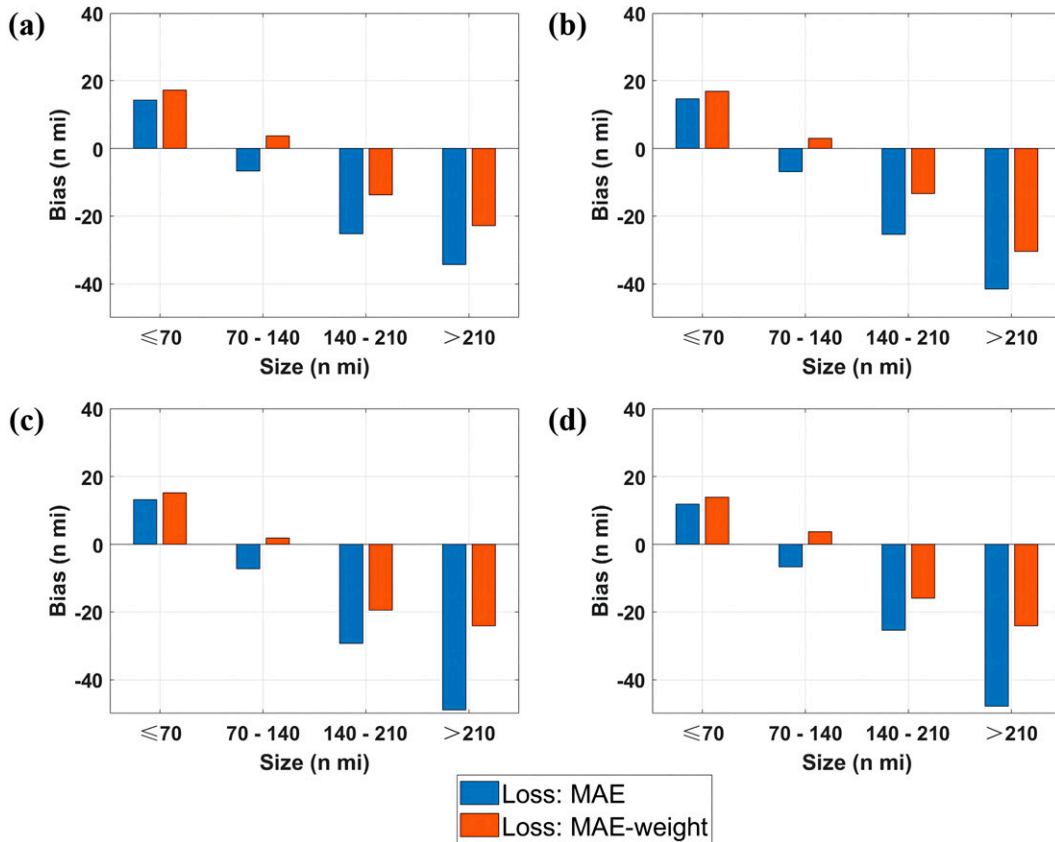| Size (n mi)	Bias for loss function = MAE				Bias for loss function = MAE-weight			
	NE	SE	SW	NW	NE	SE	SW	NW
≤70	14.5 (1327)	14.8 (1674)	13.0 (2246)	11.7 (1595)	17.3	16.9	15.5	14.0
70–140	−6.7 (2663)	−6.8 (2420)	−7.2 (2064)	−6.7 (2643)	3.7	3.0	1.9	3.7
140–210	−25.3 (550)	−25.6 (474)	−29.3 (259)	−25.8 (331)	−13.7	−13.6	−19.5	−16.0
>210	−34.0 (89)	−41.3 (61)	−49.4 (60)	−48.0 (60)	−22.8	−30.7	−24.1	−24.1
MAE	18.5	19.1	18.2	18.3	18.8	19.5	18.6	18.8

FIG. 8. The bias (n mi) for each quadrant at different TC wind radii of the DL-TCR models with a different loss function in the global ocean (including the Northwest Pacific, Northeast Pacific, and the Southern Hemisphere) on test datasets.

Figures 9b, 9e, 9h, and 9k are the weights of the NE, SE, SW, and NW quadrants TC wind radii estimation models on the subimage. Same to the weight distribution in the original image, the weights near the TC center are larger than in the other area. The weight distribution in the NE and SW quadrants is close to a one-quarter circle, and the weight distribution in the SE and NW quadrants is closer to a one-quarter ellipse. The results illustrate that the features learned in different quadrants are different.

Figures 9c, 9f, 9i, and 9l are the difference between the weights of the corresponding quadrant on the original image and subimage. Since their weights have different value ranges, the left and center columns of Fig. 9 were transformed to the interval [0, 1], and then the right column of Fig. 9 is the absolute value of their difference. Figures 9c, 9f, 9i, and 9l show that the whole images and the corresponding subimages provide different feature information to the DL model, both of which are helpful to the DL model.

TABLE 7. The comparison of the MAEs (n mi) of DL-TCR models and other methods in estimating TC wind radii using infrared satellite image.

| Method	Test data	R34				R50				R64			
		NE	SE	SW	NW	NE	SE	SW	NW	NE	SE	SW	NW
Kossin et al. (2007)	2005		24.2				19.8				14.5		
Knaff et al. (2016)	2011–13		37				20				12		
Knaff et al. (2016)	2011–13	44	37	36	41	21	21	23	23	13	13	13	14
DL-TCR	Same year data as in Knaff et al. (2016)	19.8	19.3	17.9	20.0	11.7	11.4	10.8	11.1	9.1	9.8	9.7	8.9
Zhuo and Tan (2021)	2017 and 2019		17.0				12.5				9.2		
Dolling et al. (2016)	21 TCs between 2004 and 2010		20.8				12.5				7.3		
Dolling et al. (2016)	21 TCs between 2004 and 2010	27.7	25.2	19.9	30.1	17.4	16.7	12.3	18.1	10.3	8.9	6.9	9.1
DL-TCR	Same TC data as in Dolling et al. (2016)	18.8	17.8	17.5	19.4	11.5	11.2	10.9	10.9	8.9	9.5	9.0	8.8
DL-TCR	2015–16	18.8	19.5	18.6	18.8	11.3	11.3	11.1	10.8	8.9	9.9	9.2	8.7
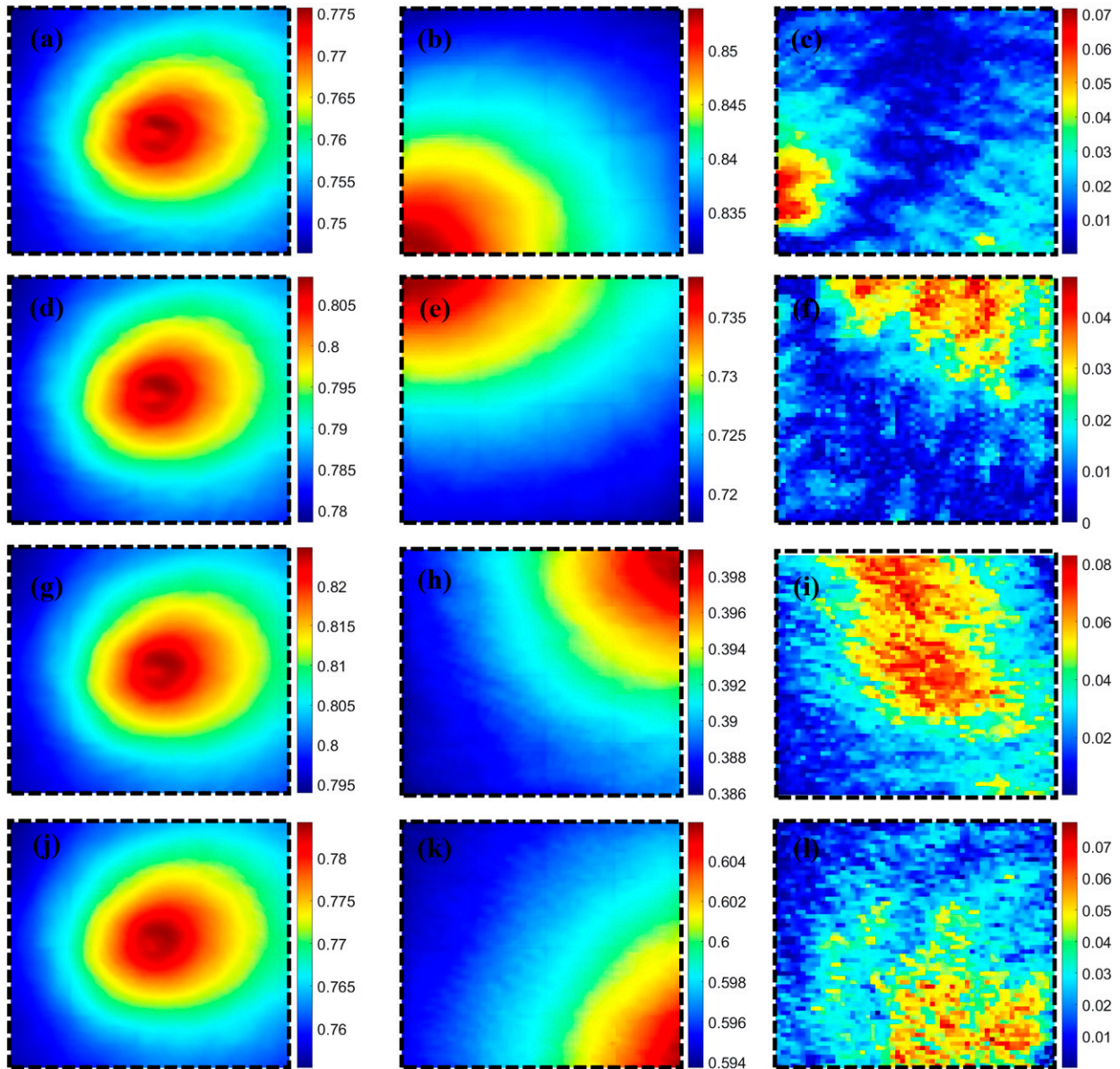
FIG. 9. The weights and differences of the first attention layer behind the input layer at the (a)–(c) NE, (d)–(f) SE, (g)–(i) SW, and (j)–(l) NW quadrants in DL-TCR models: (left) the weights for original images from Part A of the DL-TCR model, (center) the weights of the same quadrants for subimages from Part B of the DL-TCR model, and (right) the difference between the corresponding quadrants. Note that the left and center columns were transformed to the interval [0, 1] and then the right column is the absolute value of their difference.

For different samples, the DL-TCR model learns different features. For example, Figs. 10a–e are the samples of tropical storm (34–63 kt), category 1 (64–82 kt), category 2 (83–95 kt), category 3 (96–112 kt), and category 4 (113–136 kt). The left and left-center columns are the input images at the current time and 3 h ago, respectively. The right-center and right columns are the heat maps of the input layer calculated by the Shap values (called Shap heat map) and saliency maps methods (called saliency heat map), respectively.

The Shap method uses the classical Shapley values from game theory and their associated extensions to explain the output of any DL model. The Shap heat map shows that the model focuses on the cloud regions, spiral structures in the TC eye, and low brightness temperature. In particular, the model focuses on the edges of the TC cloud region. The spiral structure of the TC can be seen in the right middle column of Figs. 10a and 10c. This result indicates that the model considers the spiral, which represents convection to some extent, important for estimating TC wind radius.

The saliency maps method focuses on the relevance between individual pixels in the input and output images. So there is a difference between the heat map calculated by these
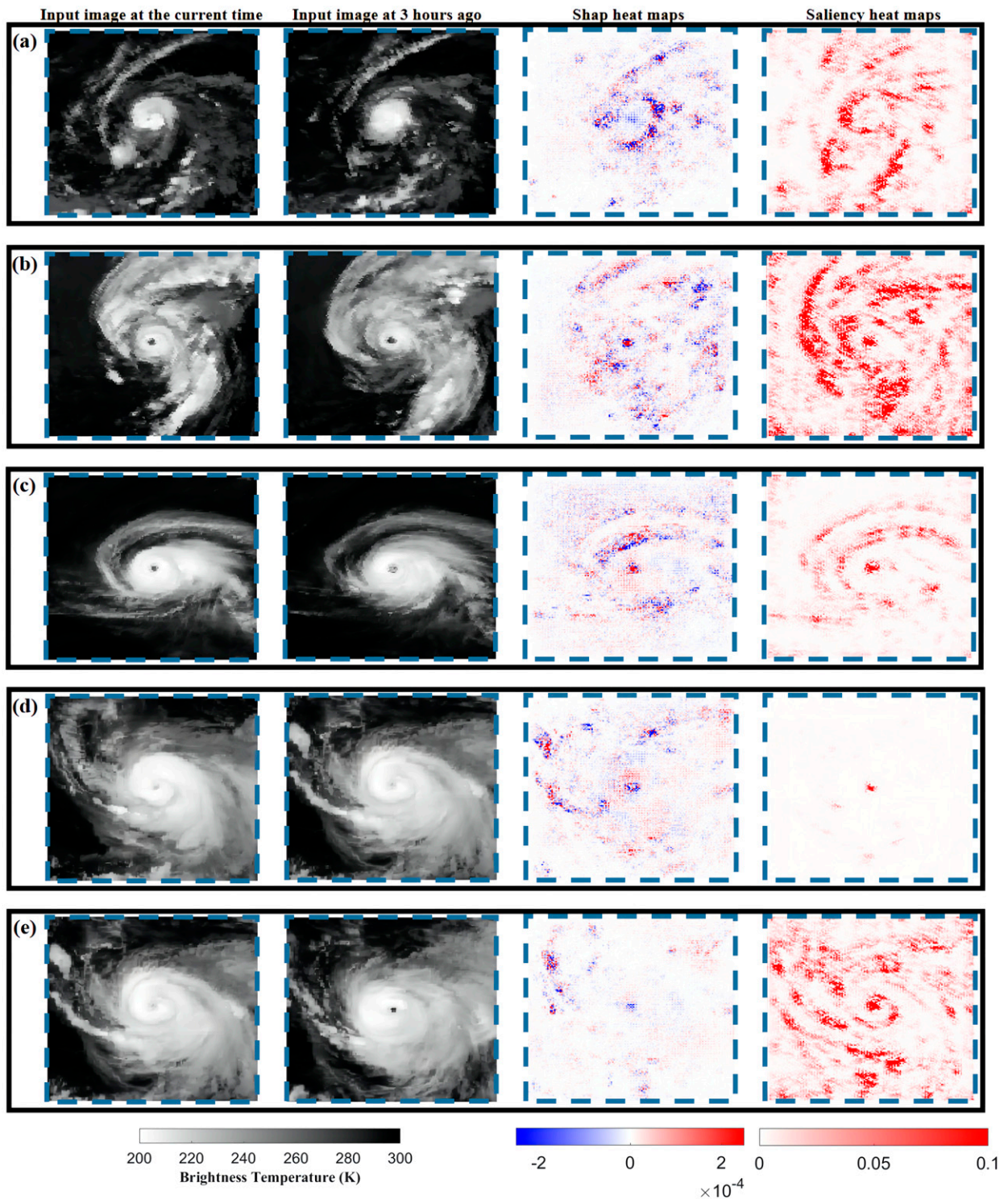
FIG. 10. Original images and heat maps of (a) tropical storm (34–63 kt), (b) category 1 (64–82 kt), (c) category 2 (83–95 kt), (d) category 3 (96–112 kt), and (e) category 4 (113–136 kt), showing (left) the input image at the current time, (left center) the input image 3 h prior, (right center) Shap heat maps, and (right) saliency heat maps.

two methods. The saliency heat maps show that the model considers the TC eye and the TC spiral structure most relevant to the TC wind radius. The heat maps calculated by both methods show that the TC center, cloud, and TC spiral structure are the main factors that affect the model performance. The result is the same as the traditional perception and proves that the DL-TCR model can accurately extract features related to TC wind radius from the images.

Traditional techniques Kossin et al. (2007), Knaff et al. (2016), Dolling et al. (2016) are mainly based on human understanding of TCs and manual analysis of specific features (e.g., spiraling, central cold clouds, asymmetries or symmetries, deviation angle variance, etc.). TC wind radii are a function of these features. However, DL-TCRs can extract more abstract features from TC images and establish a nonlinear relationship between these features and TC wind radius. Therefore, it is important to visualize DL-TCRs using interpretation methods that can help people understand TCs, which can be helpful for future TC research.

5. Summary

TCs are one of the most serious disasters. It is important to estimate their wind radius. A large TC wind radius represents a large TC impact area, which can cause more damage. However, in existing satellite-infrared-image-based methods, the large TC wind radius is often underestimated because of the unbalanced samples, and there are no DL methods for estimating global asymmetric TC wind radii. This article developed objective DL-TCR models to estimate asymmetric TC wind radii from infrared images. Results indicate that adding the attention layers in the VGGNet and inputting the corresponding quadrant images are beneficial for improving the accuracy of asymmetric TC wind radius estimation. TC wind radius in different oceans has different characteristics with different difficulty levels to learn. A DL model for estimating global TC wind radius was developed by combining separately trained DL models in the North Atlantic, the Northwest Pacific, the Northeast Pacific, and the Southern Hemispheres. The large TC wind radius underestimation was reduced by using the modified MAE-weight loss function. In particular, the bias of R34 greater than 210 n mi at the SW quadrant is significantly improved by 51.2%. The visualization of DL-TCR shows that our models can learn the asymmetric structure of the TC, and the area closer to the TC center, cloud with low brightness temperature, and TC spiral structure are the main factors possibly influencing the model performance.

The global TC wind radii estimation models obtain MAEs of R34 of 18.5, 19.1, 18.2, and 18.3 n mi for the NE, SE, SW, and NW quadrants, respectively, MAEs of R50 of 11.1, 10.8, 10.9, and 10.6 n mi for the same quadrants and MAEs of R64 of 8.9, 9.7, 9.0, and 8.6 n mi for the same quadrants. The evaluation results based on homogeneous test samples showed that the global asymmetric TC wind radii estimation models surpass existing satellite-infrared-image-based methods by 12.1%–35.5% for asymmetric TC wind radii.

The structure of TCs is asymmetric, and studying TC asymmetry may enhance the estimation of TC information, such as TC spiral cloud rainbands, TC intensity, RMW, TC minimum pressure, and so on. Our results demonstrate that DL-TCR 3 model can mine the asymmetric TC information from satellite infrared images by introducing corresponding quadrant subimages. It provides an idea for the extraction of TC asymmetric information.

Unfortunately, the HURSAT-B1 dataset stopped updating in 2016, and GridSat-B1 can supplement TC data in recent years (Knapp et al. 2011). Combining scatterometer, SAR, and infrared satellite images may greatly enhance the accuracy of TC wind radius estimation. More TC data are helpful to reduce the error of fewer sample data, such as the large TC wind radius. In addition, continuing to explore the interpretation of deep learning and the integration of deep learning and atmospheric or ocean science is essential. Adding physical constraints or a priori knowledge to the DL model may be the future direction of this research area.

Data availability statement. The infrared images were obtained online (https://www.ncdc.noaa.gov/hursat/), as were the IBTrACS data (https://www.ncdc.noaa.gov/ibtracs/).

REFERENCES

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.

Chan, K. T. F., and J. C. L. Chan, 2012: Size and strength of tropical cyclones as inferred from QuikSCAT data. *Mon. Wea. Rev.*, **140**, 811–824, https://doi.org/10.1175/MWR-D-10-05062.1.

Chavas, D. R., and K. A. Emanuel, 2010: A QuikSCAT climatology of tropical cyclone size. *Geophys. Res. Lett.*, **37**, L18816, https://doi.org/10.1029/2010GL044558.

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. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Honolulu, HI, Institute of Electrical and Electronics Engineers, 1800–1807, https://doi.org/10.1109/CVPR.2017.195.

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.

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.

Ebuchi, N., H. C. Graber, and M. J. Caruso, 2002: Evaluation of wind vectors observed by QuikSCAT/SeaWinds using ocean buoy data. *J. Atmos. Oceanic Technol.*, **19**, 2049–2062, https://doi.org/10.1175/1520-0426(2002)019<2049:EOWVOB>2.0.CO;2.

Fan, S., B. Zhang, A. A. Mouche, W. Perrie, J. A. Zhang, and G. Zhang, 2020: Estimation of wind direction in tropical cyclones using C-band dual-polarization synthetic aperture radar. *IEEE Trans. Geosci. Remote Sens.*, **58**, 1450–1462, https://doi.org/10.1109/TGRS.2019.2946885.

Fore, A. G., S. H. Yueh, B. W. Stiles, W. Tang, and A. K. Hayashi, 2018: SMAP radiometer-only tropical cyclone intensity and size validation. *IEEE Geosci. Remote Sens. Lett.*, **15**, 1480–1484, https://doi.org/10.1109/LGRS.2018.2849649.

Goodfellow, I., Y. Bengio, and A. Courville, 2016: *Deep Learning*. Vol. 1. MIT Press, 800 pp.

Griffin, S. M., A. Wimmers, and C. S. Velden, 2022: Predicting rapid intensification in North Atlantic and eastern North Pacific tropical cyclones using a convolutional neural network. *Wea. Forecasting*, **37**, 1333–1355, https://doi.org/10.1175/WAF-D-21-0194.1.

Halpern, D., 1993: Validation of special sensor microwave imager monthly-mean wind speed from July 1987 to December 1989. *IEEE Trans. Geosci. Remote Sens.*, **31**, 692–699, https://doi.org/10.1109/36.225535.

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*, Las Vegas, NV, Institute of Electrical and Electronics Engineers, 770–778, https://doi.org/10.1109/CVPR.2016.90.

Horstmann, J., S. Falchetti, C. Wackerman, S. Maresca, M. J. Caruso, and H. C. Graber, 2015: Tropical cyclone winds retrieved from C-band cross-polarized synthetic aperture radar. *IEEE Trans. Geosci. Remote Sens.*, **53**, 2887–2898, https://doi.org/10.1109/TGRS.2014.2366433.

Kingma, D. P., and J. Ba, 2014: Adam: A method for stochastic optimization. arXiv, 1412.6980v9, https://doi.org/10.48550/arXiv.1412.6980.

Knaff, J. A., 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) unifying tropical cyclone data. *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., J. A. Knaff, H. I. Berger, D. C. Herndon, T. A. Cram, C. S. Velden, R. J. Murnane, and J. D. Hawkins, 2007: Estimating hurricane wind structure in the absence of aircraft reconnaissance. *Wea. Forecasting*, **22**, 89–101, https://doi.org/10.1175/WAF985.1.

Kotikalapudi, R., and Coauthors, 2017: keras-vis. GitHub, accessed 1 June 2022, https://github.com/raghakot/keras-vis.

Krizhevsky, A., I. Sutskever, and G. E. Hinton, 2012: ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, F. Pereira et al., Eds., Vol. 25, Curran Associates, 1097–1105.

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., and Coauthors, 1995: Convolutional networks for images, speech, and time series. *The Handbook of Brain Theory and Neural Networks*, M. A. Arbib, Ed., MIT Press, 276–279.

Li, X., and Coauthors, 2020: Deep-learning-based information mining from ocean remote-sensing imagery. *Natl. Sci. Rev.*, **7**, 1584–1605, https://doi.org/10.1093/nsr/nwaa047.

——, Y. Zhou, and F. Wang, 2022: Advanced information mining from ocean remote sensing imagery with deep learning. *J. Remote Sens.*, **2022**, 9849645, https://doi.org/10.34133/2022/9849645.

Lin, T.-Y., P. Goyal, R. Girshick, K. He, and P. Dollár, 2017: Focal loss for dense object detection. *Proc. IEEE Int. Conf. on Computer Vision*, Venice, Italy, Institute of Electrical and Electronics Engineers, 2999–3007, https://doi.org/10.1109/TPAMI.2018.2858826.

Liu, Y., Q. Zheng, and X. Li, 2021: Characteristics of global ocean abnormal mesoscale eddies derived from the fusion of sea surface height and temperature data by deep learning. *Geophys. Res. Lett.*, **48**, e2021GL094772, https://doi.org/10.1029/2021GL094772.

Lundberg, S. M., and S.-I. Lee, 2017: A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems 30*, I. Guyon et al., Eds., Curran Associates, 4765–4774.

Mears, C. A., D. K. Smith, and F. J. Wentz, 2001: Comparison of special sensor microwave imager and buoy-measured wind speeds from 1987 to 1997. *J. Geophys. Res.*, **106**, 11 719–11 729, https://doi.org/10.1029/1999JC000097.

Mouche, A., B. Chapron, B. Zhang, and R. Husson, 2017: Combined co-and cross-polarized SAR measurements under extreme wind conditions. *IEEE Trans. Geosci. Remote Sens.*, **55**, 6746–6755, https://doi.org/10.1109/TGRS.2017.2732508.

——, ——, J. Knaff, Y. Zhao, B. Zhang, and C. Combot, 2019: Copolarized and cross-polarized SAR measurements for high-resolution description of major hurricane wind structures: Application to Irma category 5 hurricane. *J. Geophys. Res. Oceans*, **124**, 3905–3922, https://doi.org/10.1029/2019JC015056.

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.

Pradhan, R., R. S. Aygun, 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.

Ren, Y., X. Li, and W. Zhang, 2022: A data-driven deep learning model for weekly sea ice concentration prediction of the pan-arctic during the melting season. *IEEE Trans. Geosci. Remote Sens.*, **60**, 1–19, https://doi.org/10.1109/TGRS.2022.3177600.

Reul, N., J. Tenerelli, B. Chapron, D. Vandemark, Y. Quilfen, and Y. Kerr, 2012: SMOS satellite L-band radiometer: A new capability for ocean surface remote sensing in hurricanes. *J. Geophys. Res.*, **117**, C02006, https://doi.org/10.1029/2011JC007474.

——, 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.

Simonyan, K., and A. Zisserman, 2014: Very deep convolutional networks for large-scale image recognition. arXiv, 1409.1556v6, https://doi.org/10.48550/arXiv.1409.1556.

——, A. Vedaldi, and A. Zisserman, 2013: Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv, 1312.6034v2, https://doi.org/10.48550/arXiv.1312.6034.

Sola, J., and J. Sevilla, 1997: Importance of input data normalization for the application of neural networks to complex industrial problems. *IEEE Trans. Nucl. Sci.*, **44**, 1464–1468, https://doi.org/10.1109/23.589532.

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, https://dl.acm.org/doi/10.5555/2627435.2670313.

Sun, Y., X. Wang, and X. Tang, 2014: Deep learning face representation from predicting 10,000 classes. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Columbus, OH, Institute of Electrical and Electronics Engineers, 1891–1898, https://doi.org/10.1109/CVPR.2014.244.

Sun, Z., B. Zhang, J. A. Zhang, and W. Perrie, 2019: Examination of surface wind asymmetry in tropical cyclones over the northwest Pacific Ocean using SMAP observations. *Remote Sens.*, **11**, 2604, https://doi.org/10.3390/rs11222604.

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*, Las Vegas, NV, Institute of Electrical and Electronics Engineers, 2818–2826, https://doi.org/10.1109/CVPR.2016.308.

Wang, C., G. Zheng, X. Li, Q. Xu, B. Liu, and J. Zhang, 2022: Tropical cyclone intensity estimation from geostationary satellite imagery using deep convolutional neural networks.

IEEE Trans. Geosci. Remote Sens., **60**, 1–16, https://doi.org/10.1109/TGRS.2021.3066299.

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.

Yueh, S. H., A. G. Fore, W. Tang, A. Hayashi, B. Stiles, N. Reul, Y. Weng, and F. Zhang, 2016: SMAP L-band passive microwave observations of ocean surface wind during severe storms. *IEEE Trans. Geosci. Remote Sens.*, **54**, 7339–7350, https://doi.org/10.1109/TGRS.2016.2600239.

Zhang, B., and W. Perrie, 2012: Cross-polarized synthetic aperture radar: A new potential measurement technique for hurricanes. *Bull. Amer. Meteor. Soc.*, **93**, 531–541, https://doi.org/10.1175/BAMS-D-11-00001.1.

——, ——, J. A. Zhang, E. W. Uhlhorn, and Y. He, 2014: High-resolution hurricane vector winds from C-band dual-polarization SAR observations. *J. Atmos. Oceanic Technol.*, **31**, 272–286, https://doi.org/10.1175/JTECH-D-13-00006.1.

——, Z. Zhu, W. Perrie, J. Tang, and J. A. Zhang, 2021: Estimating tropical cyclone wind structure and intensity from spaceborne radiometer and synthetic aperture radar. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, **14**, 4043–4050, https://doi.org/10.1109/JSTARS.2021.3065866.

Zhang, S., Q. Xu, H. Wang, Y. Kang, and X. Li, 2022: Automatic waterline extraction and topographic mapping of tidal flats from SAR images based on deep learning. *Geophys. Res. Lett.*, **49**, e2021GL096007, https://doi.org/10.1029/2021GL096007.

Zhang, X., and X. Li, 2021: Combination of satellite observations and machine learning method for internal wave forecast in the Sulu and Celebes Seas. *IEEE Trans. Geosci. Remote Sens.*, **59**, 2822–2832, https://doi.org/10.1109/TGRS.2020.3008067.

——, H. Wang, S. Wang, Y. Liu, W. Yu, J. Wang, Q. Xu, and X. Li, 2022: Oceanic internal wave amplitude retrieval from satellite images based on a data-driven transfer learning model. *Remote Sens. Environ.*, **272**, 112940, https://doi.org/10.1016/j.rse.2022.112940.

Zhuo, J.-Y., and Z.-M. Tan, 2021: Physics-augmented deep learning to improve tropical cyclone intensity and size estimation from satellite imagery. *Mon. Wea. Rev.*, **149**, 2097–2113, https://doi.org/10.1175/MWR-D-20-0333.1.

Developing a Data-Driven Transfer Learning Model to Locate Tropical Cyclone Centers on Satellite Infrared Imagery

CHONG WANG[a,b] AND XIAOFENG LI[a]

[a] *CAS Key Laboratory of Ocean Circulation and Waves, Institute of Oceanography, Chinese Academy of Sciences, Qingdao, China*
[b] *University of Chinese Academy of Sciences, Beijing, China*

ABSTRACT: In this paper, a data-driven transfer learning (TL) model for locating tropical cyclone (TC) centers from satellite infrared images in the northwest Pacific is developed. A total of 2450 satellite infrared TC images derived from 97 TCs between 2015 and 2018 were used for this paper. The TC center location model (ResNet-TCL) with added residual fully connected modules is built for the TC center location. The MAE of the ResNet-TCL model is 34.8 km. Then TL is used to improve the model performance, including obtaining a pretrained model based on the ImageNet dataset, transferring the pretrained model parameters to the ResNet-TCL model, and using TC satellite infrared imagery to fine-train the ResNet-TCL model. The results show that the TL-based model improves the location accuracy by 14.1% (29.3 km) over the no-TL model. The model performance increases logarithmically with the amount of training data. When the training data are large, the benefit of increasing the training samples is smaller than the benefit of using TL. The comparison of model results with the best track data of TCs shows that the MAEs of TCs center is 29.3 km for all samples and less than 20 km for H2–H5 TCs. In addition, the visualization of the TL-based TC center location model shows that the TL model can accurately extract the most important features related to TC center location, including TC eye, TC texture, and contour. On the other hand, the no-TL model does not accurately extract these features.

KEYWORDS: Tropical cyclones; Remote sensing; Deep learning

1. Introduction

Tropical cyclones (TCs) are intense weather processes generated over tropical oceans. After making landfall, TCs can cause mudslides, flash floods, and other disasters, which can cause great damage to people (Zhang and Li 2017; Fernandez et al. 2006). Therefore, the monitoring and forecasting of TCs is very important. In addition, TC intensity estimation, TC tracking, and TC forecasting require accurate TC center location (Olander and Velden 2007, 2019). Therefore, accurate TC center location is crucial for forecasters and emergency responders (Jaiswal and Kishtawal 2013; Hu et al. 2017; Lu et al. 2017).

Satellite remote sensing is widely used to locate TC centers because they have wide spatial and temporal coverage (Zheng et al. 2019). Based on the sensor type, existing TC center location methods are mainly classified as 1) infrared based (IR based) (Velden and Hawkins 2002), 2) synthetic aperture radar based (SAR based), and 3) microwave based (MIC based).

Existing IR-based and SAR-based methods include 1) the subjective empirical judgment method (Olander and Velden 2007; Dvorak 1975, 1984), 2) the threshold method (Fett and Brand 1975; Chaurasia et al. 2010; Jin et al. 2014; You et al. 2022), 3) the spiral curve method (Jaiswal and Kishtawal 2011; Lu et al. 2019; Shin et al. 2022), and 4) the cloud-derived wind method (Wood 1994; Zheng et al. 2019, 2016; Cecil and Biswas 2017; Hu et al. 2019; Liu and Zhang 2022; Y. Wang et al. 2021).

Below, a concise overview of these methods is provided, with their limitations highlighted.

The Dvorak (Dvorak 1975) method is the most popular technique worldwide, which divides TCs into several classes according to their intensity, each corresponding to several special TC pattern maps. The forecaster obtains the location of the TC center concerning the pattern map. Although the method is widely used, it is highly subjective. Dvorak (Dvorak 1984) and Olander et al. (Olander and Velden 2007, 2019) have optimized and automated the Dvorak method.

The threshold method is based on work by Fett and Brand (1975) and Chaurasia et al. (2010) that noted the structure of the central dense overcast of a TC is approximately elliptical. After the TC intensity reaches a certain level, the central dense overcast appears circular. The location of the TC center can be judged by observing the morphological features of the central dense overcast. The threshold method segments and identifies the central dense overcast and defines the morphological center of the central dense overcast as the TC center. This method requires complex preprocessing of the image and does not apply to TCs without an obvious central dense overcast.

The spiral curve method makes use of the vortex structure of a TC, from which the TC center can be determined by extracting the spiral curve of the TC cloud system. This method requires thresholding, contrast enhancement, histogram equalization, Gaussian smoothing, and filtering satellite infrared images before matching the spiral curve and identifying the spiral center as the TC center (Jaiswal and Kishtawal 2011; Yurchak 2007). This method takes longer to calculate and only applies to TCs with distinct spiral cloud bands.

Corresponding author: Xiaofeng Li, lixf@qdio.ac.cn

The cloud-derived wind method (Zheng et al. 2019, 2016; Hu et al. 2019; Liu and Zhang 2022; Y. Wang et al. 2021) extracts vector wind information from cloud positions retrieved from a time series of IR imagery. The wind field is then used to determine the TC center.

SAR imagery (Jin et al. 2014) and microwave data (Zhang et al. 2014; Hu et al. 2019) have been used for TC structural analysis work. However, SAR systems are carried on polar-orbiting satellites, which often can only observe a part of the TC. Additionally, the spatial resolution of microwave data is low, which poses challenges for TC center localization. IR instruments are found on geostationary satellites with a broad look angle and high repeat coverage, making this format more convenient for TC analysis and forecast work, despite a lower spatial resolution than SAR.

In addition to some of the challenges identified above, the IR methods also rely on computational processing that converts satellite images into high-level artificial features (e.g., texture features, thresholding, edge detection) and requires complex image preprocessing with long computation time. IR imagery can return large errors for locating TC centers when TCs are low intensity or have poorly defined structure. Therefore, developing accurate and efficient methods for TC center locations remains challenging.

A more recent computational processing development is deep learning (DL). It has powerful data mining capabilities (Li et al. 2020, 2022) and has been widely used for a variety of applications in remote sensing in recent years, such as TC intensity estimation (Chen et al. 2019; C. Wang et al. 2021; Zheng et al. 2022), TC wind radius estimation (Zhuo and Tan 2021; Wang et al. 2022), identification and forecasting of ocean phenomena, sea ice (Zhang and Li 2020; S. Zhang et al. 2022; X. Zhang et al. 2022), and other (Wang et al. 2022; Wu et al. 2023). Yang et al. (2019) applied DL techniques to study TC center locations. Wang et al. (2019) used the CNN-L model to locate the TC center from infrared satellite images and further improved the accuracy of the TC center location by combining the TC identification model and the TC center location model (Wang et al. 2020).

DL originated in computer vision, and with the increased computational power, DL models have evolved toward deeper and more complex directions (He et al. 2016). Along with increased depth and complexity, the accuracy of DL models utilized in various applications has also improved. However, it is important to note that DL models have many parameters, and the models can be easily overfitted if not supported by sufficient training data (Goodfellow et al. 2017).

Regarding the application of DL in TC research, for instance, for tasks like TC intensity or wind field estimation, it is noteworthy that the latest DL models like ResNet or GoogLeNet (Zhuo and Tan 2021) exhibit lower performance compared to older models like VGG-19 (Simonyan and Zisserman 2014). This result is primarily due to the challenge that more complex DL approaches face when dealing with limited training data.

To address this constraint in utilizing DL for TC analysis, one strategy is to employ a transfer learning (TL) approach. TL operates on the premise that the fundamental features (e.g., edges and textures) extracted by DL models for diverse tasks are shared or similar. DL models trained on extensive datasets possess more robust capabilities for extracting these fundamental features than those trained on smaller datasets. TL serves as an efficient remedy for the issue of limited training data by transferring the robust feature extraction capabilities from DL models trained on large datasets to other DL models. Consequently, TL is an effective solution for addressing the challenge of working with small training datasets in computational tasks. Li et al. (2019) proposed a visibility detection method based on TL. Jeon et al. (2020) achieved high-precision sea fog detection from GOCI images using TL. Han et al. (2022) improved the accuracy of radar-based rainfall nowcasting using TL. X. Zhang et al. (2022) used a TL approach for predicting internal wave amplitude that linked a laboratory dataset (888 samples) to a smaller observational dataset (121 samples) and improved prediction accuracy by 21% over the observational dataset alone. This result indicates that TL approaches are promising for oceanographic applications where observational data are often limited.

The ImageNet dataset (Yang et al. 2020) (https://image-net.org/about.php), which is commonly used in computer vision, has more than 14 million training samples. Complex models trained with the ImageNet dataset have powerful feature extraction capabilities. In many studies, the fundamental features may be used in other applications. For example, the contour and texture feature extraction capability learned from the ImageNet dataset can be used for TC contour and texture extraction. More accurate features such as contours and textures help locate the TC center. The approach proposed in this paper will improve locational accuracy of TC center identification by combining the greater effectiveness, albeit lower accuracy, of simpler DL models in a limited observational data context with enhanced feature recognition developed for other applications and integrated in using TL. The approach proposed in this paper is the first time the TL approach has been combined with DL for TC center identification.

To solve the small number of data point issues, we first obtain an ImageNet-based pretraining model. Then the model parameters of the convolutional layer of the pretraining model are transferred to the TC center location model. Finally, satellite infrared TC images are used for fine-tuning the TC center location model parameters. Thus, the reuse of the fundamental features is achieved, and the location accuracy of the TC center is greatly improved.

Our contributions can be summarized as follows:

1) Improved ResNet model adds the residual fully connected modules to better cope with the TC center location problems.
2) The proposed idea of TL across disciplinary domains transfers model knowledge from the computer vision domain to TC center location research, realizing the reuse of the fundamental feature extraction capability and greatly improving the model performance, and improving the accuracy of weak-intensity TC center location.
3) The proposed TL idea enables the small sample TC center location data to be applied to the large parameter DL

model as well, and can better exploit the upper-performance limit of the large parameter DL model.

The dataset and the DL-based models are introduced in sections 2 and 3. After those, section 4 provides an analysis and discussion of the model results. Finally, section 5 presents the summaries.

2. Data and preprocessing

a. Geostationary satellite infrared imagery of TCs

The *Himawari-8* (*H-8*) satellite, which was launched by the Japan Meteorological Agency (JMA) in October 2014, provided the TC satellite infrared imagery used in this investigation. Three visible (0.47–0.64 μm), 3 near-infrared (0.86–2.3 μm), and 10 thermal infrared (3.9–13.3 μm) bands were among the 16 bands of data provided by the Advanced Himawari Imager (AHI) on board *H-8*. With a temporal resolution of 10 min and a spatial resolution of 0.5–5 km, the *H-8* imaging range covers the Pacific Ocean (Bessho et al. 2016). Lu et al. (2019) found that multichannel image fusion could improve the TC center location accuracy. Therefore, in this paper, we select channels 8 (6.2 μm), 13 (10.4 μm), and 15 (12.3 μm) with high transmittance near the atmospheric window for the TC center location study. As shown in Fig. 1, with the same color bar, different channels show different information. For example, compared to channels 13 and 15, channel 8 (water vapor channel) image has a lower brightness temperature in the areas with a large water vapor content. A total of 2450 images with 5 km spatial resolution were utilized to create a dataset of 97 TCs over the northwest Pacific Ocean from 2015 to 2018.

b. Best track dataset of TCs

The best track dataset for TCs provided by the China Meteorological Administration (CMA; https://tcdata.typhoon.org.cn/) was used to extract TC images with their corresponding labels (Ying et al. 2014). The location and intensity of TCs in the North Pacific are specified at 6-h intervals. After 2017, for TCs making landfall in China, the temporal resolution is improved to 3 h during the 24 h before its landfall.

c. Data preprocessing

Data normalization can prevent model gradient explosion and speed up model computation (Sola and Sevilla 1997). In this paper, satellite infrared images are linearly transformed to the interval [0, 1] by

$$y = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \tag{1}$$

where $x_{\min}$ and $x_{\max}$ are the minima and maximum values of the brightness temperature from bands 8, 13, and 15; y is the normalized value limited to [0, 1].

After normalization, the TC images are randomly split 3:1:1 into training, validation, and test data (Table 1).

A 321 × 321 size image centered at the TC center locations provided by the CMA best track dataset for each TC image
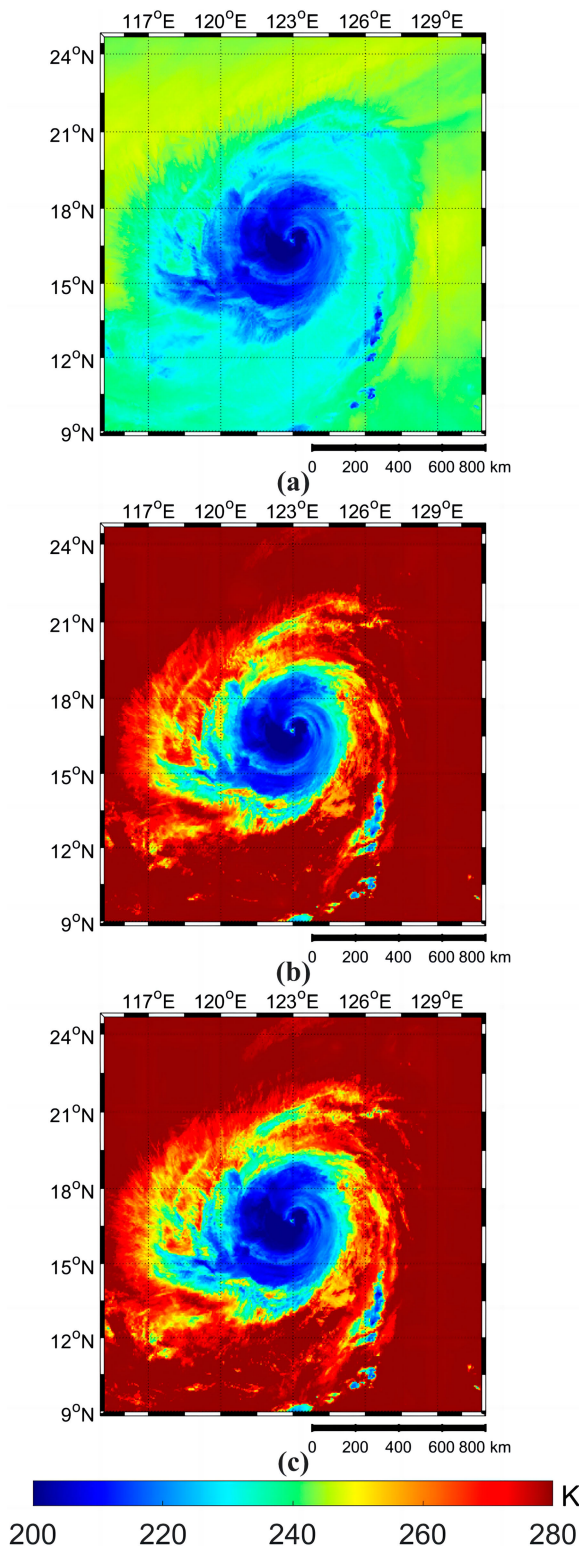

FIG. 1. Brightness temperature (K) images from different channels: (a) channel 8, (b) channel 13, and (c) channel 15, with a spatial coverage of 1600 km × 1600 km.

TABLE 1. Number of training, validation, and test data for the DL-based TC center location model.

	Train	Validation	Test
TS (17.1–32.4 m s^{-1})	1227	296	277
H1 (32.5–42.2 m s^{-1})	228	89	103
H2 (42.3–48.9 m s^{-1})	126	35	39
H3 (49.0–57.6 m s^{-1})	124	46	45
H4 (57.7–70.0 m s^{-1})	58	22	24
H5 (>70.0 m s^{-1})	7	2	2
Total	1470	490	490

was extracted. Each training or validation image was reduced from 321 × 321 to 224 × 224 pixels, randomly shifted by 0–30 pixels up, down, left, and right three times (Fig. 2). The shift range was 0–30. Finally, we label the subimages with the number and orientation of the shifted pixels. For example, if an imaging center were shifted 5 pixels to the left and 10 pixels up, the image would be labeled as (−5, 10). It should be noted that the test image was also cropped only once. These manipulations resulted in an expanded dataset of 4410 training images, 1470 validation images, and 490 test images with a reduced image size of 224 × 224.

3. Deep learning model development

a. TC center location model configuration

Among DL models, the CNN model (LeCun et al. 2015) is good at capturing spatial correlation in images and can extract complex image features. CNN can accurately and quickly extract features such as TC contours, TC textures, and TC eyes. Therefore, CNN models are often used to extract TC information from satellite images. The CNN framework is often designed as a feed-forward network that updates the model weights by a back-propagation algorithm (LeCun et al. 2015). It is made up of a fully connected layer, a pooling layer, and a convolutional layer. The fully connected layer learns the intricate nonlinear relationship between features and outputs, while the convolutional layer extracts imagery features, and the pooling layer smooths these features using filtering techniques. As a result, CNN avoids complex image preprocessing and feature engineering (feature engineering refers to transforming, selecting,

creating, or preparing data features) (LeCun et al. 2015). The Alexnet (Krizhevsky et al. 2017) was an improved model based on the CNN framework that halved the error rate of target recognition. After that, VGG-19 (Simonyan and Zisserman 2014) deepened the model depth using small convolutional kernels; ResNet (He et al. 2016) further refined this approach by introducing a residual module that significantly deepened the model complexity, allowing the model to generalize and handle more complex tasks. Therefore, the ResNet model is selected as the base model for the TC center location model developed in this paper.

The ResNet model was originally designed for image classification work in computer vision. The convolutional layer structure of the ResNet model is kept unchanged so that the model can inherit the feature extraction capability of the pretrained model using TL. Two modifications (ResNet-TCL-A and ResNet-TCL-B) are introduced here to make it more suitable for the TC center location (Fig. 3).

These are expanded upon below. First, the attention mechanism originated from the study of human vision and has helped to improve model performance in many studies (C. Wang et al. 2021; Wang et al. 2022). Therefore, ResNet-TCL-A adds the spatial and temporal attention layers between the input layer and the first convolutional layer. Second, each node of the fully connected layer connects all nodes of the previous layer, giving the model have stronger learning ability. Compared with the commonly used DL models (i.e., VGGNet), the ResNet model is not designed with a fully connected layer. Therefore, ResNet-TCL-B adds 2 residual fully connected modules between the last convolutional layer and the output layer.

Other than model architecture, the performance of the DL model is positively correlated with the number of samples required for training. The amount of modeling data determines the lower limit of model performance (Pan and Yang 2010). The problem of an insufficient amount of training data has been observed in many fields. This issue spurred the development of TL techniques.

TL is an important tool to solve the problem of insufficient training data and consists of two concepts: source and target domains. The target domain is the knowledge that has to be learned, and the source domain is the knowledge that already exists. TL is the application of knowledge or patterns learned in the source domain to the target domain (Pan and Yang 2010).
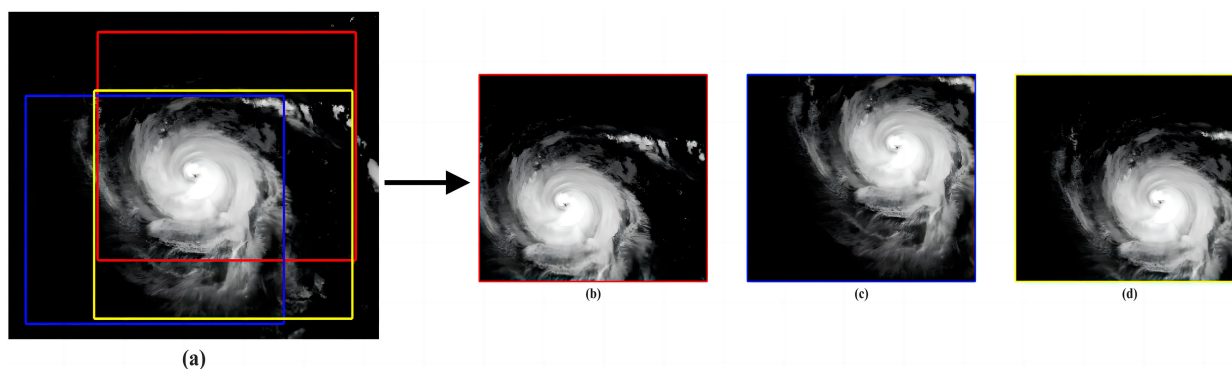

FIG. 2. Random cut images: (a) original image (size: 321 × 321) and (b)–(d) randomly cropped images (size: 224 × 224).
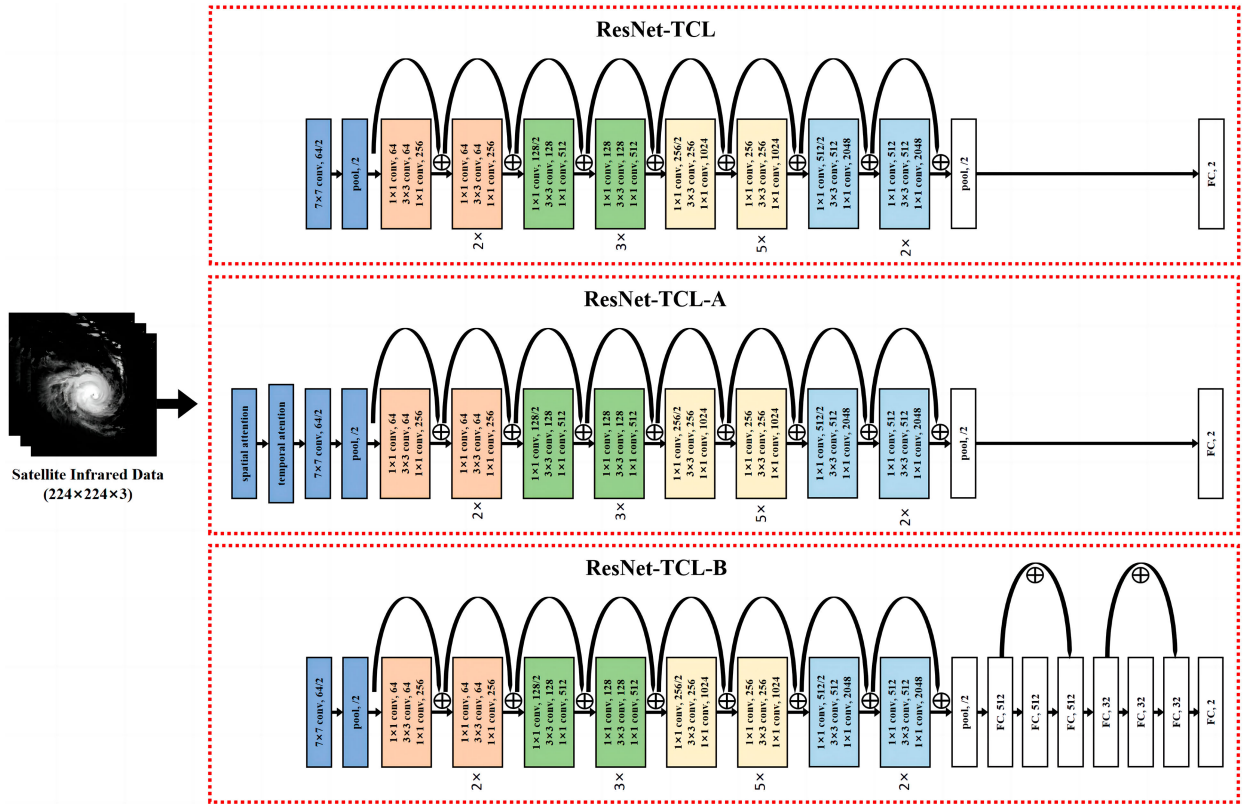
FIG. 3. Architecture of ResNet-TCL, ResNet-TCL-A, and ResNet-TCL-B model for locating TC center (FC means the fully connected layer).

The requirements for the data needed for DL are relaxed by TL. TL enables the reuse of the fundamental features, greatly reducing the data and training time required for the target task (Pan and Yang 2010; Hu et al. 2018).

In summary, once the model architecture has been established, the next major step is to train the TC center location model using the domain-adversarial TL (Fig. 4). The training steps of the TC center location model (ResNet-TCL, ResNet-TCL-A, ResNet-TCL-B) proposed in this paper include five steps:

First, the pretrained models (ResNet-pre) trained on the ImageNet dataset, which consists of more than 14 million training samples and is a well-known dataset containing natural images for various object recognition tasks (Yang et al. 2020), are obtained (https://image-net.org/about.php). ResNet-pre follows the standard ResNet-50 structure, which includes 16 residual blocks, 49 convolutional layers, and 1 fully connected layer. ReLU is the activation function for all layers except the fully connected one. It is constructed using the Python TensorFlow-Keras programming language. Pretrained models based on the ImageNet dataset are offered by Keras, which can be directly downloaded without the need for retraining on the ImageNet dataset (for details, see https://keras.io/api/applications/resnet/).

Second, the TC center location model is built to locate the TC center. It is important to note that the TC center location model is used for all TCs, and the only similarity with the pretrained model (model-pre) is in the convolutional layer shown in Fig. 4.

Third, the weights of the convolutional layer in model-pre are transferred to TC center location model, and the weights of the TC center location model fully connected layer are initialized randomly.

Fourth, the TC center location dataset is used to fine-train TC center location model. It includes fine-tuning the weights of the convolutional layer and training the weights of the fully connected layer. It means that TC center location model inherits the pretrain model's powerful fundamental feature extraction capability and improves the natural image feature extraction toward TC feature extraction during the fine-tuning process.

Fifth, the TL-based TC center location models are obtained.

b. Setup of experiments

First, the performance of the three model architectures for the TC center location is compared in Table 2 to determine the optimal model architecture suitable for the TC centers location. Then, three sets of experiments (Table 2) are set up in this paper to compare the model performance with and without TL to explore the role of TL on the ResNet-TCL, ResNet-TCL-A, and ResNet-TCL-B model.

Once the optimal model architecture (ResNet-TCL-B, Table 2) has been established, the results in sections 4b–4d
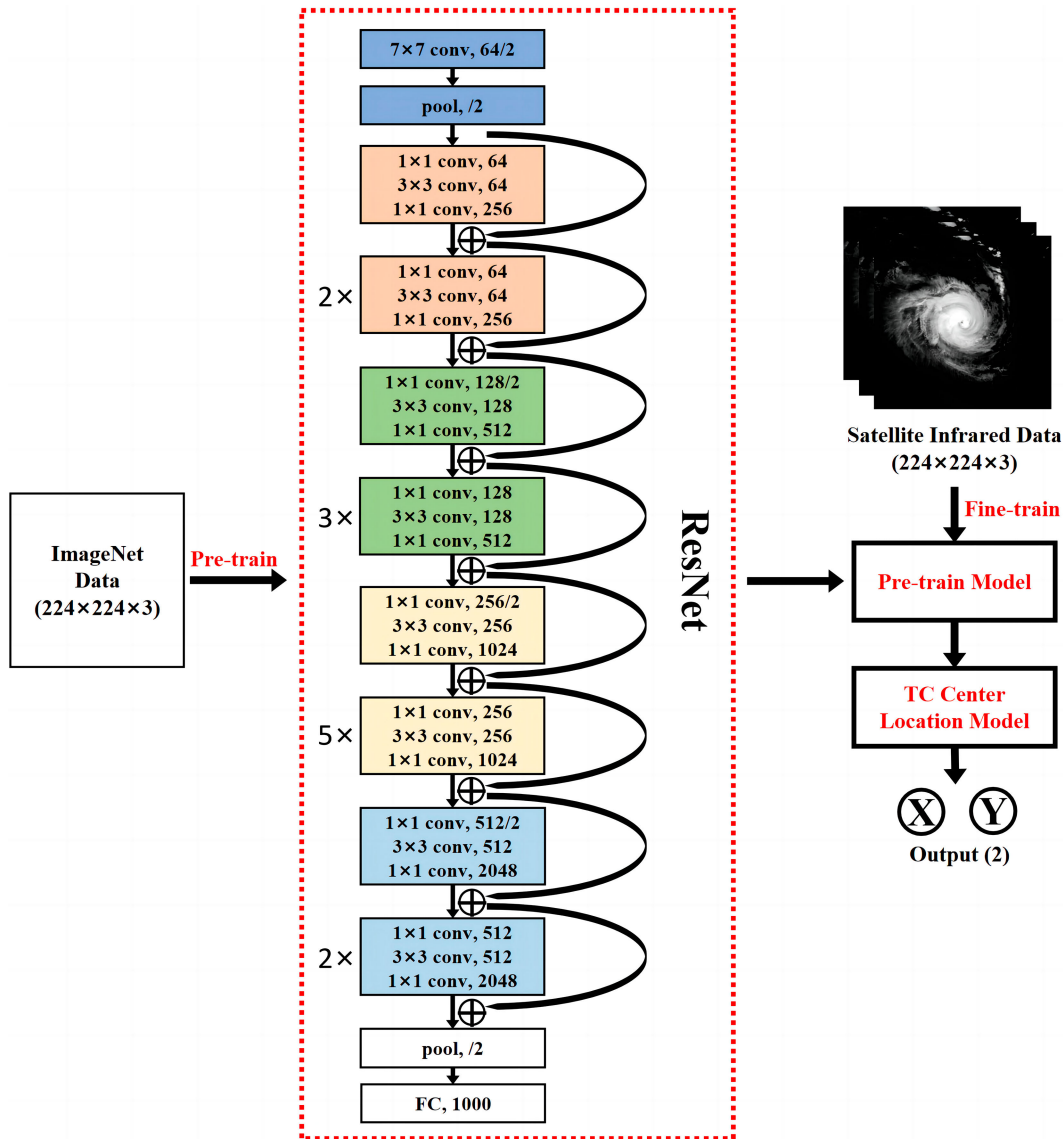
FIG. 4. Modeling process of the ResNet model based on TL.

were all based on ResNet-TCL-B. In section 4b, the training data were divided into four equal parts. Four experiments were set up without TL, sequentially increasing the amount of training data; meanwhile, another set of four experiments was set up using TL with the same training data. How the size of

TABLE 2. Mean TC center location error (MAE) of ResNet-TCL, ResNet-TCL-A, and ResNet-TCL-B models with and without TL.

ID	Method	No-TL model (km)	TL model (km)
ResNet-TCL	ResNet	34.5	29.8
ResNet-TCL-A	ResNet + attention	34.5	29.8
ResNet-TCL-B	ResNet + residual FC	34.1	29.3

the training dataset and the utilization of TL impact the accuracy of determining the center location was evaluated.

Note that the hyperparameter selection for the ResNet-TCL, ResNet-TCL-A, and ResNet-TCL-B models mentioned above is as follows, aiming to achieve fast convergence and high performance: 1) ReLU activation functions are used for the convolutional and fully connected layers to expedite model convergence. 2) The output layer employs a sigmoid activation function, the most commonly used activation function in regression model output layers. 3) The loss function chosen is mean-squared error (MSE), which closely aligns with the calculation method for distance $(x^2 + y^2)$. 4) The optimization function is based on the adaptive moment estimation (Adam) algorithm. 5) The initial learning rate for the optimization function is set to 0.0005, and if the validation
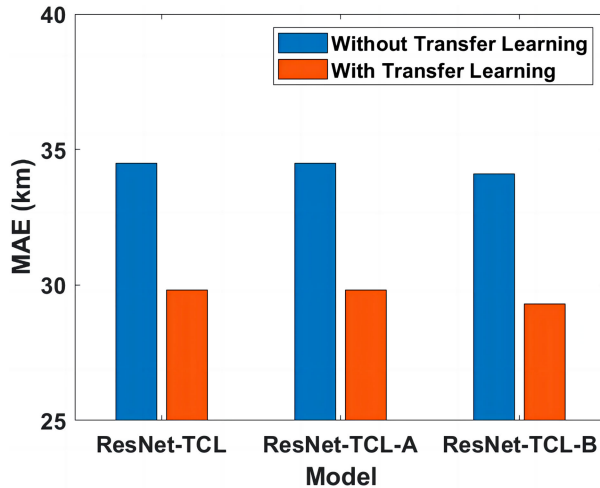
FIG. 5. MAE of ResNet-TCL, ResNet-TCL-A, and ResNet-TCL-B model with and without TL.

FIG. 6. Loss curve during the training of the ResNet-TCL model with and without TL.

group's loss value does not improve for eight consecutive epochs, the learning rate is increased by a factor of 0.5. Specifically, the initial learning rate of 0.0005 was selected to provide stable convergence, and the learning rate schedule helps prevent overfitting by adjusting the learning rate during training. This strategy was found effective during our experimentation. 6) The early stop function is applied during model training. If the validation data loss value does not decrease for over 10 epochs, the model's training is stopped.

In addition, DL is often seen as a black box that does not explain the mechanism of how it works. Therefore, DL model interpretability is a key concern for scholars. The differences between TL and no-TL models using interpretability tools is discussed in section 4d.

4. Results and discussion

a. Modeling based on TL or no-TL

In this section, the results of the ResNet-TCL-based model with/without TL are compared. The input satellite infrared images and model output was the same for the model described in this section. The difference is whether TL was being used or not. Model performance is evaluated using the mean absolute error (MAE):

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\sqrt{(x_{\text{true}} - x_{\text{model}})^2 - (y_{\text{true}} - y_{\text{model}})^2}, \quad (2)$$

where x_{true} and y_{true} are the real locations of the TC center, x_{model} and y_{model} are the location of the TC center located by the DL model, and N is the number of test data.

Table 2 and Fig. 5 show that the MAE of the ResNet-TCL, ResNet-TCL-A, and ResNet-TCL-B models are 34.5, 34.5, and 34.1 km. The results of the ResNet-TCL and ResNet-TCL-A models are comparable. It indicates that the attention layer does not help in the TC center location. The performance of the ResNet-TCL-B model without TL is 1.2% higher
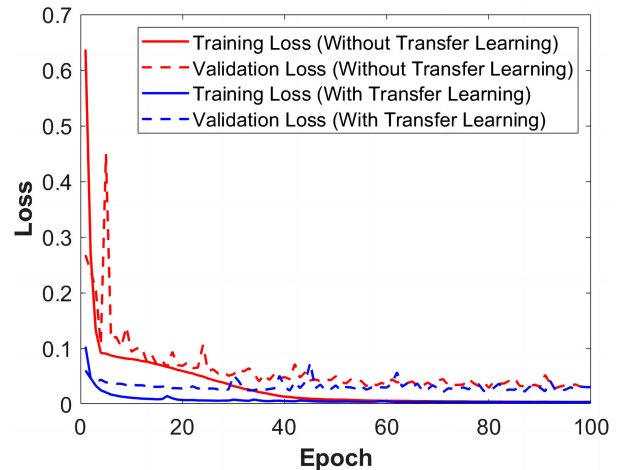
than that of ResNet-TCL. This is because the ResNet-TCL-B model has additional residual fully connected modules compared to the ResNet-TCL model. The residual fully connected modules allow the model to simulate complex tasks better.

Figure 6 shows the loss curves of the ResNet-TCL model for the training and validation data during the training process. The solid and dashed lines are the variation of the training and validation data loss values with the number of training epochs, respectively. Red and blue represents the ResNet-TCL model without and with TL, respectively. As shown in Fig. 6, the ResNet-TCL model with TL converged after 20 training epochs, while the ResNet-TCL model without TL started to converge only after 40 training epochs. Moreover, the loss values of the ResNet-TCL model with TL for both training and validation data are consistently lower than those of the ResNet-TCL model without TL. The results show that TL can accelerate model convergence and improve model performance.

ResNet-TCL, ResNet-TCL-A, and ResNet-TCL-B models improved by 13.6% (29.8 km), 13.6% (29.8 km), and 14.1% (29.3 km) after using TL. The ResNet-TCL-B model has the lowest MAE and the largest performance improvement compared to the model without TL. At the same time, the residual fully connected modules of the ResNet-TCL-B model greatly increase the model parameters. More trainable parameters mean that the model requires more training data. TL can effectively address model data requirements and enable it to realize its upper-performance limit, especially for the ResNet-TCL-B model, fully.

b. The effect of training data on the model

The effect that size of the training dataset has on TL model performance is examined in this section. The best performing ResNet-TCL-B model in section 4a is chosen for the TL-based model in this section. The validation data and test data are consistent with section 4a. First, the training data

TABLE 3. TC center location results are based on different amounts of training data.

ID	The amount of training data	No-TL model (km)	TL model (km)
ResNet-TCL-1	367 × 3 (25%)	56.1	47.2
ResNet-TCL-2	734 × 3 (50%)	43.8	37.1
ResNet-TCL-3	1101 × 3 (75%)	37.0	31.6
ResNet-TCL-4 (ResNet-TCL-B)	1470 × 3 (100%)	33.9	29.3

(not randomly cropped) were divided into four equal parts, and then each part was randomly cropped three times (as described in section 3c) to expand the data. As shown in Table 3, ResNet-TCL-1 to ResNet-TCL-1 were trained using 25%, 50%, 75%, and 100% of the data.

Table 3 and Fig. 7 show the effect of the amount of training data on the TL and no-TL models. The ResNet-TCL-1 to ResNet-TCL-4 mean location errors are 56.1, 43.8, 37.0, and 33.9 km when TL is not used. The model performance is improved as the training data increase, at a decreasing rate. When the training data were increased from 25% to 50%, the model performance was improved by 21.9%. When the training data were increased from 75% to 100%, the model performance improved by only 8.4%. The results show that the model performance tends to increase logarithmically with the number of training data.

The model performance of ResNet-TCL-1 to ResNet-TCL-4 with TL was improved by 15.8%, 15.3%, 14.6%, and 13.6% compared to that without TL. The results show that TL can effectively improve the model performance. In particular, the highest improvement is achieved when the number of training data is small, and model performance improvement only slightly decreases with the increase in training data. When the amount of training data is large, such as ResNet-TCL-3, adding 25% of data improves the model performance by 8.4%, and using TL improves the model performance by 14.6%. The results show that the improvement of model performance by strong fundamental feature extraction capability is greater than that by increasing the amount of data.
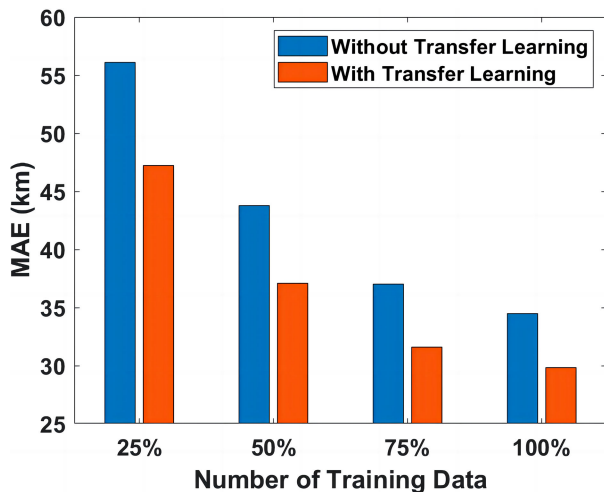
Figure 8 shows the location results of the ResNet-TCL-4 with/without the TL model. Weak-intensity TCs have inconspicuous structures that are difficult to locate. Strong TCs have obvious TC eyes and structure and are more easily located. As a result, the location error gradually decreases with the increase in TC intensity. The ResNet-TCL-4 model with TL has less MAE than the ResNet-TCL-4 model without TL at each intensity. For example, the MAE for H1–H5 is 20.0 km, with less than 20 km for H2–H5.

c. Comparison with the latest technology

The performance of the ResNet-TCL-B model with TL in locating the TC center was compared with the latest methods listed in Table 4 in this section. The location accuracy of DL models is generally comparable to or surpasses that of these methods. However, the methods proposed by Zheng et al. (2019), Lu et al. (2019), Wang et al. (2020), Shin et al. (2022), and Liu and Zhang (2022) lack publicly available source code, making it impossible to compare them using the same dataset. Although the comparison in Table 4 is not equitable, the results show that our model has an overall good accuracy by the existing standard. On the other hand, the DL models from Wang et al. (2019) were retrained using the same dataset as this paper (Table 4). Compared with the DL model of Wang et al. (2019), the ResNet-TCL-B model with TL performs more accuracy. The VGG-19 model has also been tested with TL. The MAE of VGG-19 without/with TL is 34.8 and 31.4 km, respectively, which is higher than the ResNet model. The results show the effectiveness of the TL-based DL model proposed in this paper.

FIG. 7. TC center location results based on different amounts of training data.
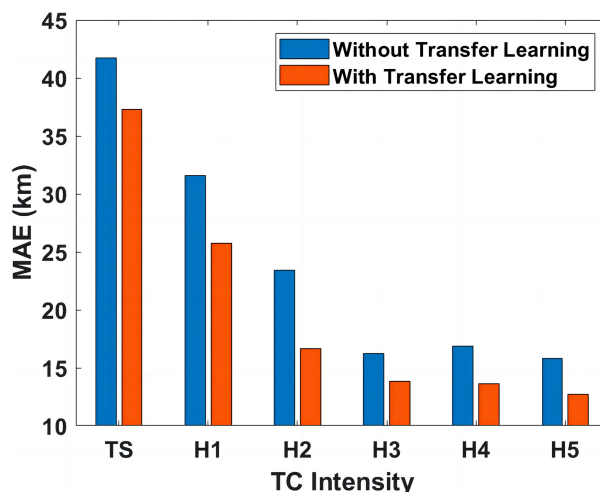

FIG. 8. MAE of the ResNet-TCL-4 model with TL.

TABLE 4. Comparison of the performances of our model with TL and other techniques in locating TC center from satellite infrared image.

Literature	Method	Dataset	MAE
Zheng et al. (2019)	Cloud-derived wind	Cloud-derived wind	—
Lu et al. (2019)	Fitting TC spiral	TCs from 2012 to 2016	54 km
Wang et al. (2019)	Deep learning	TCs from 2015 to 2018	54 km
Wang et al. (2020)	Deep learning	TCs from 2008 to 2011 and 2017 to 2019	Longitude 0.237°, latitude 0.237°
Liu and Zhang (2022)	Cloud-derived wind	3 TC cases	41 km
Shin et al. (2022)	Fitting TC spiral	TCs from 2011 to 2019	0.38°
Wang et al. (2019)	Deep learning	Same test TC data from 2015 to 2018	39.1 km
VGG-19 model			34.8 km (without TL), 31.4 km (with TL)
Our model			29.3 km

d. Visualization and interpretation of the ResNet-TCL model

A DL model is often thought of as a black box. In this black box, DL models extract features from images and learn the laws from the features to the target task. However, DL/TL results ultimately must be related to physical processes. Therefore, the interpretability of DL models has become a popular research topic. However, the existing DL model for the TC center location study (Yang et al. 2019; Wang et al. 2019; Wang et al. 2020) neglects the interpretation of the model. Therefore, the ResNet-TCL-B model will be analyzed using DL interpretability methods. In this section, the aim is to 1) emphasize why TL-based deep learning models excel in TC center location and 2) identify potential error sources.

To our knowledge, the most popular and efficient techniques for understanding DL models are feature maps and heat maps. The feature map shows the features extracted from the input image by the convolutional layer. The input image impact on the model output can be seen in the heat map. Heat map techniques such as activation maximization analysis (Toms et al. 2020), network layer correlation propagation (Andersson et al. 2021), and sensitivity analysis (Espeholt et al. 2022) have been widely used in the field of geomatics to provide insight into the internal mechanisms of DL models. For example, Toms et al. (2020), based on the heat map, found that the most relevant region of the ENSO phase category identified by the DL model is consistent with the Niño-3.4 region. This result validates that the working mechanism of DL models is consistent and complementary with existing knowledge, offering the possibility of DL feeding atmospheric ocean science. The toolboxes developed by Kotikalapudi (Kotikalapudi et al. 2017) and Lundberg (Lundberg and Lee 2017) bring together a variety of model interpretable methods. Such toolboxes can be used to invoke interpretable methods to obtain feature maps and heat maps. It is important to note that most of the existing DL interpretable methods target classification problems, not regression problems. The saliency map method, which calculates the effect of pixel changes in the input image on the results, is used in this paper to obtain the heat map.

The ResNet-TCL-B model has 49 convolutional layers, and the deeper the convolutional layer, the more abstract the features extracted. Therefore, this section shows the feature maps extracted from the third and thirteenth convolutional layers. The third and thirteenth convolutional layers have 64 and 128 convolutional kernels, respectively, so 64 and 128 feature maps are extracted, respectively.

Figures 9a and 9b show the feature maps extracted from the third convolutional layer in the ResNet-TCL-B model; Figs. 9c and 9d show the feature maps extracted from the thirteenth convolutional layer in the ResNet-TCL-B model. The feature map shown in Fig. 9a blurs the difference between TC clouds and sea in the input image and does not extract important fundamental features for the TC center location. The feature maps in Fig. 9b show the texture and contour features of the TC. It is seen in Fig. 9b that the area close to the TC center and the TC contour features are extracted. Figure 9c shows the texture features and contour features of the TC. The feature maps in Fig. 9d become abstract, and the features of the TC eye can be seen in some of the feature maps.

The features extracted in Figs. 9b and 9c are more similar. It shows that the no-TL model cannot learn the fundamental features of the TC at the shallow convolutional layers and only learns the important fundamental features of TC (texture, contour, etc.) at the deeper layer. On the other hand, the TL model can learn the important fundamental features of the TC at a shallow convolutional layer so that more convolutional layers can be used to extract further and learn the relationship between the fundamental features and the TC center.

It is as if three steps are needed to solve a problem. The transferred learning model already has the empirical knowledge from step 1 to step 2 and only needs to go from step 2 to step 3. On the other hand, a no-TL model needs to go from step 1 to step 2 to step 3, and errors from step 1 to step 2 affect the result of step 3. TL enables the migration of the fundamental feature extraction capabilities, thus enabling faster and better problem-solving.

Figure 10 shows the location error and heat maps for different intensity images. The errors on the saliency heat maps in the middle and right columns of Fig. 10 are the location errors of the TL and no-TL models. The model location error is small for H1–H5 TC with obvious TC eyes. For the lower-intensity TS TC, the model location error is larger. It is mainly because low-intensity TCs do not have significant circulation characteristics. Therefore, most TC center location methods cannot accurately locate the center with low intensity. For

No-TL   TL

Conv 3
(56×56)

(a)   (b)

Conv 13
(28×28)

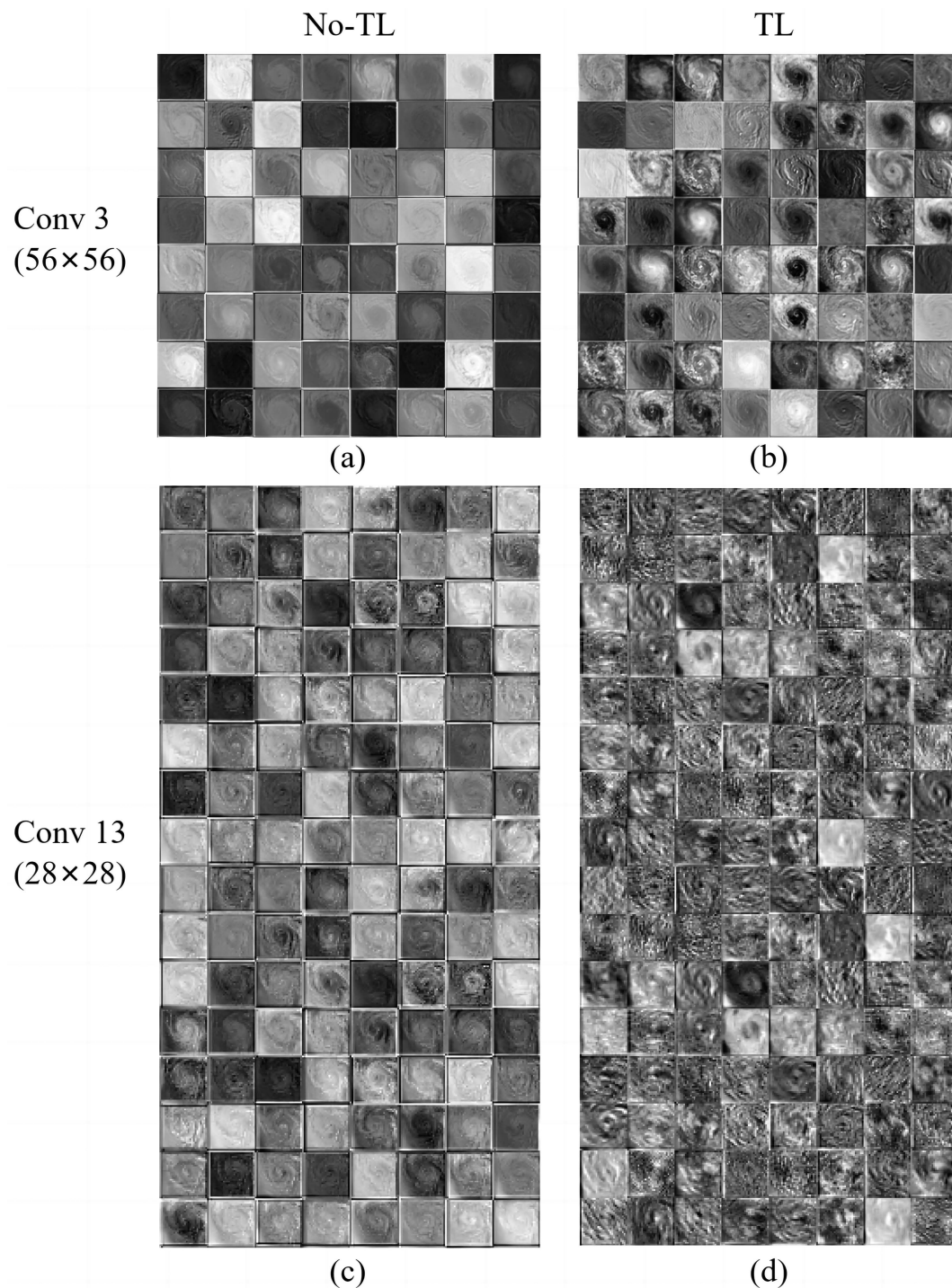(c)   (d)

FIG. 9. Feature maps are generated from different layers in the ResNet-TCL-B model (the latitude–longitude scale of input TC case is 8.80°–24.80°N, 115.00°–131.00°E). (a) Third convolutional layer in ResNet-TCL-B model with TL. (b) Third convolutional layer in ResNet-TCL-B model without TL. (c) Thirteenth convolutional layer in ResNet-TCL-B model with TL. (d) Thirteenth convolutional layer in ResNet-TCL-B model without TL. The input data are three-channel (channels 8, 13, and 15) *H-8* images (Fig. 1).
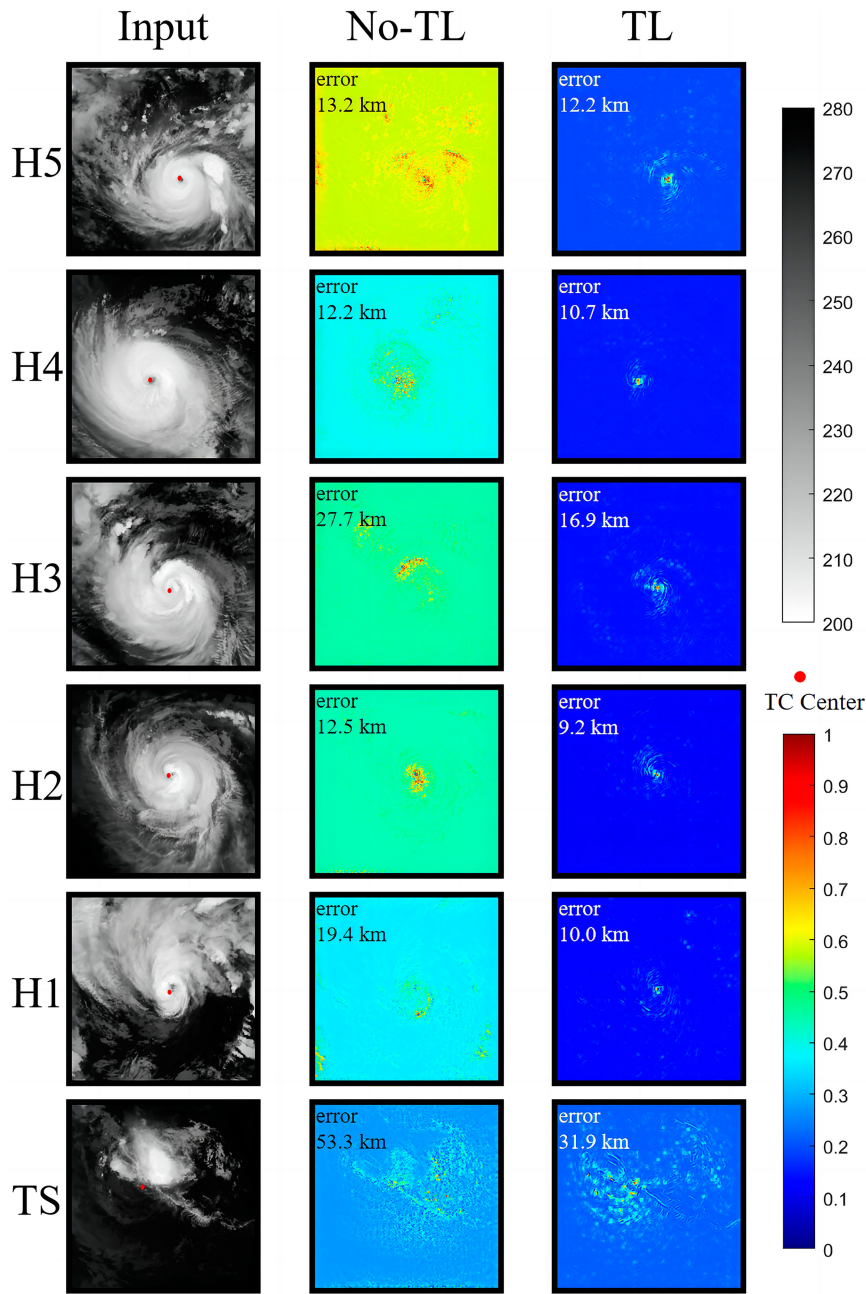
FIG. 10. Original images and saliency heat maps of TS (15.25°–26.40°N, 125.70°–136.85°E), H1 (25.10°–36.25°N, 122.70°–133.85°E), H2 (11.65°–22.80°N, 117.85°–129.00°E), H3 (9.80°–20.95°N, 137.20°–148.35°E), H4 (10.90°–22.05°N, 131.45°–142.60°E), and H5 (15.20°–26.35°N, 118.20°–129.35°E) of the TL and no-TL models. (top to bottom) The samples of H5 to TS. (left) Input image, (center) saliency heat maps of no-TL model, and (right) saliency heat maps of TL model.

operational activities, the wind field simulated by the numerical model is usually combined to assist in locating the TC center.

In the following, the reasons for the poor location results of the DL model will be analyzed using saliency heat maps. In Fig. 10, the left column is the input image, and the red dots are the TC center locations. The center column is the heat map of the no-TL model. The right column is the heat map of the TL model. The values of the heat map range from 0 to 1. The larger the value, the greater the effect of the change in that point on the result.

In the H1–H5 images with obvious TC eyes, the region of attention of the TL model is concentrated near the TC eyes, while the region outside the TC eyes is almost 0. It indicates

that the TL model learns the feature of TC eyes, which is the most important for the TC center location. However, the no-TL model focuses on more scattered regions, including image edges and TC peripheral cloud areas, in addition to the TC eye, which interferes with the TC center location. Especially for the H3 image, the no-TL model does not extract the features of the TC eye, which leads to a large location error. It is noted that for strong-intensity TCs (H1–H5) with clear TC eye, the ability to accurately capture the TC eye is considered a potential error source.

The DL model relies on recognizing the TC contour and morphology to locate the TC center in TS images where the TC eye is not obvious. The attention of the TL model is focused on the left side of the image where the TC eye is located, while the opposite is true for the no-TL model. The results show that for weaker-intensity TCs without distinct eye features, the extraction of features related to TC structure and spiral bands becomes a potential error source. The TL model has better feature extraction ability and feature learning ability and can accurately extract the most important features associated with the TC center location.

In summary, the interpretability analysis of DL models is essential. It can assist us in identifying potential error sources in TC center location and enhance our understanding of TCs. In the images with TC eyes, the TL model focuses on the region of TC eyes, which is consistent with the perception of the threshold method to locate the TC center. In images without TC eyes, ResNet-TCL-B can also locate the TC center more accurately by relying on TC morphological features, solving the drawback that traditional methods cannot be used for images without obvious TC eyes. The results show that the fundamental feature extraction capability from other fields can be transferred to the study of the TC center location, and the features related to the TC center can be extracted more accurately to achieve high-accuracy TC center location.

5. Conclusions

A limited availability of training data makes application of newer, large-parameter DL models to the TC location problem difficult. Instead, simpler DL models must be used, which imposes a minimum MAE on the accuracy of TC centers located with such methods. In this paper, a novel approach has been presented, which uses TL to transfer into the DL model information about structures and features in the image, which enhance rate of DL model convergence and improve location accuracy.

Adding the residual full connectivity modules in front of the output layer of the ResNet-TCL model improves the model performance. Compared with the ResNet-TCL model, the ResNet-TCL-B model can locate the TC center more accurately. The ResNet-TCL-B model improves the location accuracy by 1.2% over the ResNet-TCL model. The TL-based ResNet-TCL-B model improves the location accuracy by 14.1% (29.3 km) over the no-TL model. The performance of the ResNet-TCL model based on the domain knowledge of TC center location increases logarithmically with the amount of training data. When the number of samples is relatively

small, increasing the number of training samples can greatly improve the model location accuracy. When the training samples are large, the benefit of increasing the training samples is smaller than the benefit of using TL. The visualization of the ResNet-TCL model shows that the TL model can accurately extract the most important features related to TC center location, including TC eye, TC texture, and contour.

The MAE of the ResNet-TCL-B model with TL for all TC cases in the test group is 29.3 km, 20 km for H1–H5 TCs, and less than 20 km for H2–H5 TCs. The location accuracy of the ResNet-TCL-B model with TL has improved by 15%–45% compared with the latest TC center location methods based on satellite infrared images. It can provide data support for TC monitoring.

The data available for TC research are much less than that for computer vision and other fields. Therefore, the cross-domain TL idea developed in this paper provides ideas for small-sample TC information extraction and modeling and can be used in other TC monitoring research, such as TC intensity estimation and TC wind radius estimation.

This paper uses the northwest Pacific Ocean for piloting and shows the potential of TL modeling. However, TCs in different seas have different characteristics, and a study examining a global TC dataset should be considered in the future. Further investigation into the interpretability of DL and its integration with TC research is crucial, and future studies in this field may focus on adding physical limitations or prior knowledge to DL models.

REFERENCES

Andersson, T. R., and Coauthors, 2021: Seasonal Arctic sea ice forecasting with probabilistic deep learning. *Nat. Commun.*, **12**, 5124, https://doi.org/10.1038/s41467-021-25257-4.

Bessho, K., and Coauthors, 2016: An introduction to Himawari-8/9—Japan's new-generation geostationary meteorological satellites. *J. Meteor. Soc. Japan*, **94**, 151–183, https://doi.org/10.2151/jmsj.2016-009.

Cecil, D. J., and S. K. Biswas, 2017: Hurricane Imaging Radiometer (HIRAD) wind speed retrievals and validation using

dropsondes. *J. Atmos. Oceanic Technol.*, **34**, 1837–1851, https://doi.org/10.1175/JTECH-D-17-0031.1.

Chaurasia, S., C. M. Kishtawal, and P. K. Pal, 2010: An objective method of cyclone centre determination from geostationary satellite observations. *Int. J. Remote Sens.*, **31**, 2429–2440, https://doi.org/10.1080/01431160903012457.

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.

Dvorak, V. F., 1975: Tropical cyclone intensity analysis and forecasting from satellite imagery. *Mon. Wea. Rev.*, **103**, 420–430, https://doi.org/10.1175/1520-0493(1975)103<0420:TCIAAF>2.0.CO;2.

——, 1984: Tropical cyclone intensity analysis using satellite data. NOAA Tech. Rep. NESDIS 11, 47 pp., http://satepsanone.nesdis.noaa.gov/pub/Publications/Tropical/Dvorak_1984.pdf.

Espeholt, L., and Coauthors, 2022: Deep learning for twelve hour precipitation forecasts. *Nat. Commun.*, **13**, 5145, https://doi.org/10.1038/s41467-022-32483-x.

Fernandez, D. E., J. R. Carswell, S. Frasier, P. S. Chang, P. G. Black, and F. D. Marks, 2006: Dual-polarized C- and Ku-band ocean backscatter response to hurricane-force winds. *J. Geophys. Res.*, **111**, C08013, https://doi.org/10.1029/2005JC003048.

Fett, R. W., and S. Brand, 1975: Tropical cyclone movement forecasts based on observations from satellites. *J. Appl. Meteor.*, **14**, 452–465, https://doi.org/10.1175/1520-0450(1975)014%3C0452:TCMFBO%3E2.0.CO;2.

Goodfellow, I., Y. Bengio, and A. Courville, 2017: *Deep Learning.* Vol. 1. MIT Press, 800 pp.

Han, L., Y. Zhao, H. Chen, and V. Chandrasekar, 2022: Advancing radar nowcasting through deep transfer learning. *IEEE Trans. Geosci. Remote Sens.*, **60**, 4100609, https://doi.org/10.1109/TGRS.2021.3056470.

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*, Las Vegas, NV, IEEE, 770–778, https://doi.org/10.1109/CVPR.2016.90.

Hu, J., L. Shen, and G. Sun, 2018: Squeeze-and-excitation networks. *2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, UT, IEEE, 7132–7141, https://doi.org/10.1109/CVPR.2018.00745.

Hu, T., X. Wang, D. Zhang, G. Zheng, Y. Zhang, Y. Wu, and B. Xie, 2017: Study on typhoon center monitoring based on HY-2 and FY-2 data. *IEEE Geosci. Remote Sens. Lett.*, **14**, 2350–2354, https://doi.org/10.1109/LGRS.2017.2764620.

——, Y. Wu, G. Zheng, D. Zhang, Y. Zhang, and Y. Li, 2019: Tropical cyclone center automatic determination model based on HY-2 and QuikSCAT wind vector products. *IEEE Trans. Geosci. Remote Sens.*, **57**, 709–721, https://doi.org/10.1109/TGRS.2018.2859819.

Jaiswal, N., and C. M. Kishtawal, 2011: Automatic determination of center of tropical cyclone in satellite-generated IR images. *IEEE Geosci. Remote Sens. Lett.*, **8**, 460–463, https://doi.org/10.1109/LGRS.2010.2085418.

——, and ——, 2013: Objective detection of center of tropical cyclone in remotely sensed infrared images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, **6**, 1031–1035, https://doi.org/10.1109/JSTARS.2012.2215016.

Jeon, H.-K., S. Kim, J. Edwin, and C.-S. Yang, 2020: Sea fog identification from GOCI images using CNN transfer learning models. *Electronics*, **9**, 311, https://doi.org/10.3390/electronics9020311.

Jin, S., S. Wang, and X. Li, 2014: Typhoon eye extraction with an automatic SAR image segmentation method. *Int. J. Remote Sens.*, **35**, 3978–3993, https://doi.org/10.1080/01431161.2014.916447.

Kotikalapudi, R., and Coauthors, 2017: Keras-vis. GitHub, https://github.com/raghakot/keras-vis.

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.

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

Li, Q., S. Tang, X. Peng, and Q. Ma, 2019: A method of visibility detection based on the transfer learning. *J. Atmos. Oceanic Technol.*, **36**, 1945–1956, https://doi.org/10.1175/JTECH-D-19-0025.1.

Li, X., and Coauthors, 2020: Deep-learning-based information mining from ocean remote-sensing imagery. *Natl. Sci. Rev.*, **7**, 1584–1605, https://doi.org/10.1093/nsr/nwaa047.

——, Y. Zhou, and F. Wang, 2022: Advanced information mining from ocean remote sensing imagery with deep learning. *J. Remote Sens.*, **2022**, 9849645, https://doi.org/10.34133/2022/9849645.

Liu, J., and Q. Zhang, 2022: Objective detection of a tropical cyclone's center using satellite image sequences in the northwest Pacific. *Atmosphere*, **13**, 381, https://doi.org/10.3390/atmos13030381.

Lu, X., H. Yu, X. Yang, and X. Li, 2017: Estimating tropical cyclone size in the northwestern Pacific from geostationary satellite infrared images. *Remote Sens.*, **9**, 728, https://doi.org/10.3390/rs9070728.

——, ——, ——, ——, and J. Tang, 2019: A new technique for automatically locating the center of tropical cyclones with multi-band cloud imagery. *Front. Earth Sci.*, **13**, 836–847, https://doi.org/10.1007/s11707-019-0784-6.

Lundberg, S. M., and S.-I. Lee, 2017: A unified approach to interpreting model predictions. *Proc. 31st Int. Conf. on Neural Information Processing Systems*, Long Beach, CA, Association for Computing Machinery, https://dl.acm.org/doi/10.5555/3295222.3295230.

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.

Pan, S. J., and Q. Yang, 2010: A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.*, **22**, 1345–1359, https://doi.org/10.1109/TKDE.2009.191.

Shin, Y., J. Lee, J. Im, and S. Sim, 2022: An advanced operational approach for tropical cyclone center estimation using geostationary-satellite-based water vapor and infrared channels. *Remote Sens.*, **14**, 4800, https://doi.org/10.3390/rs14194800.

Simonyan, K., and A. Zisserman, 2014: Very deep convolutional networks for large-scale image recognition. arXiv, 1409.1556v6, https://doi.org/10.48550/arXiv.1409.1556.

Sola, J., and J. Sevilla, 1997: Importance of input data normalization for the application of neural networks to complex

industrial problems. *IEEE Trans. Nucl. Sci.*, **44**, 1464–1468, https://doi.org/10.1109/23.589532.

Toms, B. A., E. A. Barnes, and I. Ebert-Uphoff, 2020: Physically interpretable neural networks for the geosciences: Applications to Earth system variability. *J. Adv. Model. Earth Syst.*, **12**, e2019MS002002, https://doi.org/10.1029/2019MS002002.

Velden, C., and J. Hawkins, 2002: The increasing role of weather satellites in tropical cyclone analysis and forecasting. *Fifth Int. Workshop on Tropical Cyclones*, Cairns, Australia, WMO.

Wang, C., Q. Xu, X. Li, G. Zheng, B. Liu, and Y. Cheng, 2019: An objective technique for typhoon monitoring with satellite infrared imagery. *2019 Photonics and Electromagnetics Research Symp.—Fall*, Xiamen, China, IEEE, 3218–3221, https://doi.org/10.1109/PIERS-Fall48861.2019.9021497.

——, G. Zheng, X. Li, Q. Xu, B. Liu, and J. Zhang, 2021: Tropical cyclone intensity estimation from geostationary satellite imagery using deep convolutional neural networks. *IEEE Trans. Geosci. Remote Sens.*, **60**, 4101416, https://doi.org/10.1109/TGRS.2021.3066299.

Wang, G., X. Wang, X. Wu, K. Liu, Y. Qi, C. Sun, and H. Fu, 2022: A hybrid multivariate deep learning network for multistep ahead sea level anomaly forecasting. *J. Atmos. Oceanic Technol.*, **39**, 285–301, https://doi.org/10.1175/JTECH-D-21-0043.1.

Wang, P., P. Wang, C. Wang, Y. Yuan, and D. Wang, 2020: A center location algorithm for tropical cyclone in satellite infrared images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, **13**, 2161–2172, https://doi.org/10.1109/JSTARS.2020.2995158.

Wang, Y., G. Zheng, X. Li, L. Zhou, B. Liu, P. Chen, L. Ren, and X. Li, 2021: An automatic algorithm for estimating tropical cyclone centers in synthetic aperture radar imagery. *IEEE Trans. Geosci. Remote Sens.*, **60**, 4203716, https://doi.org/10.1109/TGRS.2021.3105705.

Wood, V. T., 1994: A technique for detecting a tropical cyclone center using a Doppler radar. *J. Atmos. Oceanic Technol.*, **11**, 1207–1216, https://doi.org/10.1175/1520-0426(1994)011<1207:ATFDAT>2.0.CO;2.

Wu, X., G. Han, W. Li, Q. Shao, and L. Cao, 2023: Deep learning–based prediction of Kuroshio path south of Japan. *J. Atmos. Oceanic Technol.*, **40**, 175–190, https://doi.org/10.1175/JTECH-D-22-0043.1.

Yang, K., K. Qinami, F.-F. Li, J. Deng, and O. Russakovsky, 2020: Towards fairer datasets: Filtering and balancing the distribution of the people subtree in the ImageNet hierarchy. *Proc. 2020 Conf. on Fairness, Accountability, and Transparency*, Barcelona, Spain, Association for Computing Machinery, 547–558, https://doi.org/10.1145/3351095.3375709.

Yang, X., Z. Zhan, and J. Shen, 2019: A deep learning based method for typhoon recognition and typhoon center location. *2019 IEEE Int. Geoscience and Remote Sensing Symp.*, Yokohama, Japan, IEEE, 9871–9874, https://doi.org/10.1109/IGARSS.2019.8899322.

Ying, M., W. Zhang, H. Yu, X. Lu, J. Feng, Y. Fan, Y. Zhu, and D. Chen, 2014: An overview of the China Meteorological Administration tropical cyclone database. *J. Atmos. Oceanic Technol.*, **31**, 287–301, https://doi.org/10.1175/JTECH-D-12-00119.1.

You, Q., Z. Li, C. Qian, and T. Wang, 2022: A tropical cyclone center location method based on satellite image. *Comput. Intell. Neurosci.*, **2022**, 3747619, https://doi.org/10.1155/2022/3747619.

Yurchak, B. S., 2007: Description of cloud-rain bands in a tropical cyclone by a hyperbolic-logarithmic spiral. *Russ. Meteor. Hydrol.*, **32**, 8–18, https://doi.org/10.3103/S1068373907010025.

Zhang, D., Y. Zhang, T. Hu, B. Xie, and J. Xu, 2014: A comparison of HY-2 and QuikSCAT vector wind products for tropical cyclone track and intensity development monitoring. *IEEE Geosci. Remote Sens. Lett.*, **11**, 1365–1369, https://doi.org/10.1109/LGRS.2013.2293496.

Zhang, J. A., and X. Li, 2017: Tropical cyclone multiscale wind features from spaceborne synthetic aperture radar. *Hurricane Monitoring with Spaceborne Synthetic Aperture Radar*, Springer, 25–39, https://doi.org/10.1007/978-981-10-2893-9_2.

Zhang, S., Q. Xu, H. Wang, Y. Kang, and X. Li, 2022: Automatic waterline extraction and topographic mapping of tidal flats from SAR images based on deep learning. *Geophys. Res. Lett.*, **49**, e2021GL096007, https://doi.org/10.1029/2021GL096007.

Zhang, X., and X. Li, 2020: Combination of satellite observations and machine learning method for internal wave forecast in the Sulu and Celebes Seas. *IEEE Trans. Geosci. Remote Sens.*, **59**, 2822–2832, https://doi.org/10.1109/TGRS.2020.3008067.

——, H. Wang, S. Wang, Y. Liu, W. Yu, J. Wang, Q. Xu, and X. Li, 2022: Oceanic internal wave amplitude retrieval from satellite images based on a data-driven transfer learning model. *Remote Sens. Environ.*, **272**, 112940, https://doi.org/10.1016/j.rse.2022.112940.

Zheng, G., J. Yang, A. K. Liu, X. Li, W. G. Pichel, and S. He, 2016: Comparison of typhoon centers from SAR and IR images and those from best track data sets. *IEEE Trans. Geosci. Remote Sens.*, **54**, 1000–1012, https://doi.org/10.1109/TGRS.2015.2472282.

——, J. Liu, J. Yang, and X. Li, 2019: Automatically locate tropical cyclone centers using top cloud motion data derived from geostationary satellite images. *IEEE Trans. Geosci. Remote Sens.*, **57**, 10175–10190, https://doi.org/10.1109/TGRS.2019.2931795.

Zheng, Z., C. Hu, Z. Liu, J. Hao, Q. Hou, and X. Jiang, 2022: Deep learning for typhoon intensity classification using satellite cloud images. *J. Atmos. Oceanic Technol.*, **39**, 55–69, https://doi.org/10.1175/JTECH-D-19-0207.1.

Zhuo, J.-Y., and Z.-M. Tan, 2021: Physics-augmented deep learning to improve tropical cyclone intensity and size estimation from satellite imagery. *Mon. Wea. Rev.*, **149**, 2097–2113, https://doi.org/10.1175/MWR-D-20-0333.1.
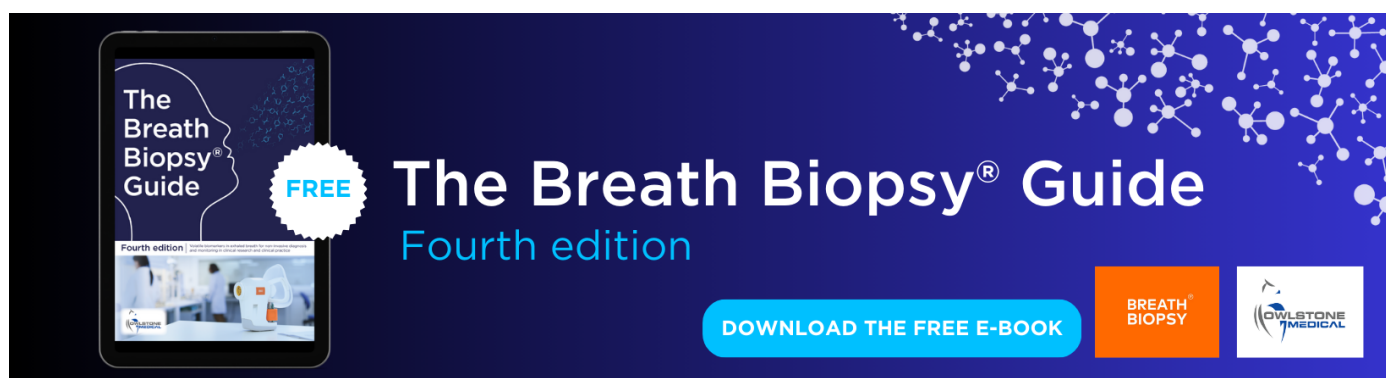
LETTER • OPEN ACCESS

Tropical cyclone intensity forecasting using model knowledge guided deep learning model

To cite this article: Chong Wang *et al* 2024 *Environ. Res. Lett.* **19** 024006

View the article online for updates and enhancements.

ENVIRONMENTAL RESEARCH
LETTERS

LETTER

Tropical cyclone intensity forecasting using model knowledge guided deep learning model

Chong Wang[1,2] , Xiaofeng Li[1,*] and Gang Zheng[3,*]

[1] Key Laboratory of Ocean Circulation and Waves, Institute of Oceanology, Qingdao, People's Republic of China
[2] University of the Chinese Academy of Sciences, Beijing, People's Republic of China
[3] State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou, People's Republic of China
[*] Authors to whom any correspondence should be addressed.

E-mail: lixf@qdio.ac.cn and zhenggang@sio.org.cn

Abstract

This paper developed a deep learning (DL) model for forecasting tropical cyclone (TC) intensity in the Northwest Pacific. A dataset containing 20 533 synchronized and collocated samples was assembled, which included ERA5 reanalysis data as well as satellite infrared (IR) imagery, covering the period from 1979 to 2021. The *u*-, *v*- and *w*-components of wind, sea surface temperature, IR satellite imagery, and historical TC information were selected as the model inputs. Then, a TC-intensity-forecast-fusion (TCIF-fusion) model was developed, in which two special branches were designed to learn multi-factor information to forecast 24 h TC intensity. Finally, heatmaps capturing the model's insights are generated and applied to the original input data, creating an enhanced input set that results in more accurate forecasting. Employing this refined input, the heatmaps (model knowledge) were used to guide TCIF-fusion model modeling, and the model-knowledge-guided TCIF-fusion model achieved a 24 h forecast error of $3.56 \, \mathrm{m \, s^{-1}}$ for Northwest Pacific TCs spanning 2020–2021. The results show that the performance of our method is significantly better than the official subjective prediction and advanced DL methods in forecasting TC intensity by 4% to 22%. Additionally, compared to operational approaches, model-guided knowledge methods can better forecast the intensity of landfalling TCs.

1. Introduction

Tropical cyclones (TCs) are extremely powerful weather phenomena originating in tropical oceans, potentially unleashing catastrophic devastation upon coastal regions (DeMaria 2009, DeMaria *et al* 2014, Cangialosi *et al* 2020). The significant hazards TCs pose include severe flooding, destructive winds, and coastal inundation caused by sorm surges, all of which substantially threaten human lives and property (Zheng *et al* 2015, 2019, Klotzbach *et al* 2018, Wang and Toumi 2021, Bhatia *et al* 2022, Li *et al* 2023). Hence, precise TC intensity forecasts are pivotal in helping individuals take proactive precautions and mitigate potential losses (Woodruff *et al* 2013, Landsea and Cangialosi 2018, Yu *et al* 2020, Wang and Toumi 2022, Zhang *et al* 2023, Wang and Li 2023a).

The traditional methods for forecasting TC intensity can be classified as (1) dynamical forecast models (Ma 2014), (2) statistical models (Emanuel 1986, DeMaria and Kaplan 1994, 1999, Knaff *et al* 2003, DeMaria *et al* 2005, DeMaria 2009, Chen *et al* 2011).

The dynamical model method is the primary approach for TC intensity forecasting (Saha *et al* 2014, Bao *et al* 2022). This method utilizes a set of mathematical equations to represent the fundamental principles governing atmospheric motion and thermodynamics (Weber 2003, Zheng *et al* 2017, 2021, 2023, Shen *et al* 2022). By assimilating extensive observational data into the forecasting models, a dynamical model can be initialized to forecast the evolution of TC intensity. The dynamical models consider a range of atmospheric and oceanic parameters, including sea surface temperatures (SSTs), humidity,

wind patterns, and pressure systems, which influence the behavior of TCs. Nonetheless, recent advances in computing technology and data assimilation techniques have significantly improved the accuracy of dynamical models, resulting in a considerable reduction in the forecast error for TC tracks, while the progress in intensity forecasting has been relatively slow (Landsea and Cangialosi 2018).

Statistical methods for TC intensity forecasting rely on historical observational data to develop mathematical models that capture the relationships between various meteorological parameters and the evolution of TC intensity (DeMaria and Kaplan 1994, 1999, DeMaria *et al* 2005, DeMaria 2009). These models often involve regression analysis, time series analysis, and other statistical techniques (Knaff *et al* 2003). The advantage of statistical methods lies in their simplicity and ease of implementation, particularly when real-time data is limited. However, these models are constrained by the assumption that historical patterns will continue to hold, limiting their effectiveness in forecasting extreme events, such as rapid intensification or sudden weakening of TCs (Lin *et al* 2009, Sandery *et al* 2010, DeMaria *et al* 2014). They are often combined with other methods to enhance overall forecast accuracy.

However, the intensity changes of TCs are influenced by many factors, such as the intricate interactions between the atmosphere and the ocean and the broader atmospheric conditions at play. These factors are complex and difficult to explain. Traditional methods cannot effectively capture the non-linear processes of TCs. Deep learning (DL) has emerged in response to this challenge. In recent years, there have been significant advancements in DL technology (Li *et al* 2020, 2022, Wang and Li 2023b), which has been successfully applied to various forecasting tasks (Lagerquist *et al* 2020, Zheng *et al* 2020, Ravuri *et al* 2021, Wang *et al* 2022, 2023, Zhang and Li 2022, Ren and Li 2023) and is currently being explored for TC intensity forecasting (Baik and Paek 2000, Pan *et al* 2019, Xu *et al* 2021, 2022, Yuan *et al* 2021, Zhang *et al* 2022, Ma *et al* 2023, Meng *et al* 2023b). DL algorithms, such as convolutional neural networks, can process vast amounts of data and identify complex patterns within meteorological datasets. This makes them particularly well-suited for handling the intricate and nonlinear characteristics of TC intensity forecasts (TCIFs). By learning from historical TC data, DL models can capture subtle relationships and nonlinear factors that traditional forecasting methods may overlook. With continuous learning and improvement from new data, DL methods demonstrate adaptability and flexibility in handling changing atmospheric conditions.

In recent years, DL has shown high accuracy and efficiency in the field of TCIFs. Baik and Paek (2000) designed a TC intensity forecasting model based on a multi-layer perceptron for the 12–72 h

period, which resulted in a 7%–16% reduction in forecast errors compared to statistical methods. Pan *et al* (2019) and Yuan *et al* (2021) considered the time series dependency in intensity forecasting based on long short-term memory (LSTM) models (Graves 2012), leading to significant improvements in forecast accuracy. Xu *et al* (2021) utilized inputs from Statistical Hurricane Intensity Prediction Scheme(SHIPS) and Dynamical Statistical Hurricane Prediction (DSHP) statistical methods, combining them with a multi-layer perceptron to reduce forecast errors by 5%–22%. Furthermore, Xu *et al* (2022) and Meng *et al* (2023b) introduced the three-dimensional structure of the TC wind field and constructed the spatial attention fusing network (SAF-NET) TC intensity forecasting model, resulting in enhanced forecasting performance. Zhang *et al* (2022) incorporated the two-dimensional sea surface field into the three-dimensional atmospheric field, augmenting the model with LSTM modules to improve its ability to extract temporal information. Ma *et al* (2023) introduced Gated Recurrent Unit (GRU) modules during the modeling process to further enhance the model's capability to extract temporal information, resulting in improved forecast accuracy.

Although these DL methods have achieved remarkable accuracy in TC intensity forecasting, three key issues remain to be addressed. Firstly, satellite infrared (IR) imagery is commonly utilized for rapid intensification forecasts due to its strong correlation with TC intensity changes (Adler and Rodgers 1977, Steranka *et al* 1986, DeMaria and Kaplan 1994, 1999, DeMaria *et al* 2005, Su *et al* 2020), yet existing DL methods only have to consider atmosphere and ocean factors from reanalysis data and ignore satellite IR imagery. Secondly, the TC process is inherently complex, and the interactions among multiple input factors are intricate. Available networks often extract features individually for each factor or use simple concatenation, leading to limited capability in learning the interplay between these factors. Lastly, in the three-dimensional atmospheric field data, besides the signals strongly correlated with TC intensity, numerous factors interfere with the model's forecasts, which have been previously overlooked during the modeling process.

In this paper, the goal is to achieve accurate 24 h TCIFs, which is defined by the 2 min maximum sustained wind speed. The contributions of this study are threefold: firstly, it demonstrates the positive impact of utilizing satellite IR imagery on TC intensity forecasting performance. Secondly, it designs feature fusion modules for different factors, enhancing the DL model's capacity to learn and represent interactions between them. Lastly, by leveraging DL model knowledge (MK) to guide the modeling process, the study focuses the model more on signals closely associated with TC intensity evolution,

resulting in improved model performance and training efficiency.

Section 2 introduces the data and methods. The performance of 24 h TCIF models analysis and discussion are given in section 3. The conclusion is given in section 4.

2. Data and methods

2.1. Data

The TC IR images used in this study were from the Gridded Satellite Data (GridSat-B1). GridSat-B1 data were created to facilitate to use of geostationary data (Knapp *et al* 2011). GridSat-B1 data are gridded International Satellite Cloud Climatology Project B1 data on a 0.07° latitude equal-angle grid. Satellites are merged by selecting the nadir-most observations for each grid point. GridSat-B1 offers IR satellite imagery with a temporal resolution of 3 h, covering the period from 1981 to the present. The imagery includes wavelengths of 11, 0.6, and 6.7 μm. However, due to the absence of 0.6 and 6.7 μm band images before the year 2000, this study only utilized the 11 μm band images.

The model inputs included environmental variables provided by ERA5 reanalysis data (Hersbach *et al* 2023), comprising the *u*- (*U*), *v*- (*V*), and *w*- (*W*) components of wind and SST. These variables strongly correlate with TC intensity (Baik and Paek 2000, Vecchi and Soden 2007, Tang and Emanuel 2010). To represent the vertical structure of TCs, four isobaric levels at 200, 500, 850, and 1000 hPa were chosen. The data collected spans from 1979 to 2021, with a spatial resolution of 1° and a temporal resolution of 6 h.

The Best Track dataset for TCs provided by the Shanghai Typhoon Institute, Chinese Meteorological Administration (STI/CMA) was used to label and extract TC samples (Ying *et al* 2014). The track and intensity of TC are recorded every 3 or 6 h in this dataset. The TC data were collected from 1979 to 2021. The data from 1979 to 2019 were partitioned, with 90% allocated for training and 10% reserved for validation. The data from 2020 and 2021 were utilized as independent test data.

For details on data preprocessing, please refer to the supporting information section S1.

2.2. Design of the TCIF model input data

The input factors of the TCIF model can directly impact the results of TCIFs. Dynamical model usually consider atmospheric and oceanic elements. However, most existing DL-based methods mainly consider atmospheric factors (such as historical TC information (HIS), *U*, *V*, *W*), while Zhang *et al* (2022) and Ma *et al* (2023) have included oceanic factors (SST). The studies indicate that IR data can depict TC morphology and convective activities and is commonly used for rapid intensification forecasting

(DeMaria and Kaplan 1994, 1999, DeMaria *et al* 2005, Su *et al* 2020). However, no DL-based TC intensity forecasting method has incorporated IR. *U*, *V*, *W*, SST, and IR into the input factors of the TCIF model.

The model architecture for the input factor selection experiments is depicted in figure 1. For example, when the input factors are *U*, *V*, *W*, and HIS (supplementary table 1), the model exclusively comprises the *U*, *V*, *W*, and HIS branches (supplementary figure 4, consists of gray arrows). However, with the addition of the input IR, the model integrates the IR branch (supplementary figure 4, consists of gray and blue arrows).

In addition, the arrangement of data sequences plays a pivotal role in influencing the computational methodology of convolutional kernels. For details on the arrangement of data sequences, please refer to the supporting information in section S2. As shown in supplementary table 2, compared with the '*x*–*y*–*z*–*t*' arrangement, the '*x*–*y*–*t*–*z*' arrangement leads to a 2.5% reduction in error. Therefore, all experiments in this paper use the '*x*–*y*–*t*–*z*' data arrangement.

2.3. Design of the TCIF model architecture

In previous studies, a common approach involved using a multi-branch network structure (24 h TCIF-basic consists of blue arrows and blocks in figure 1) to independently extract each factor's features and concatenate them at the fully connected layer. The structure of the block is illustrated in supplementary figure 2.

Two improvements have been made to the basic framework shown in figure 1. The first improvement involves adding a feature fusion branch to the 24 h TCIF-merge model (consisting of blue and green arrows and blocks in the main paper figure 1). The second improvement is the addition of an extra input branch to the 24 h TCIF-fusion model (consisting of blue, green, and gray arrows and blocks in the main paper figure 1). For details on model architecture design concepts and performances of different model architectures, please refer to the supporting information section S3. Finally, the TCIF-fusion model was confirmed for 24 h TC intensity forecasting.

In these models, the output of the TCIF model is the 24 h TC intensity, the linear activation function is employed in the output layer, while ReLU is used for the other layers. The optimization function selected is Adam, while the loss function employed is mean absolute error (MAE).

2.4. Design of MK-guided training strategy

Besides the signals strongly correlated with TC intensity in the three-dimensional atmospheric data, numerous spurious signals exist that can potentially interfere with the model's forecasts, previously overlooked during the modeling process.

In DL model interpretability has been a focal point of research. Among various interpretability

Figure 1. The architecture of TCIF-basic, TCIF-merge, TCIF-fusion model.
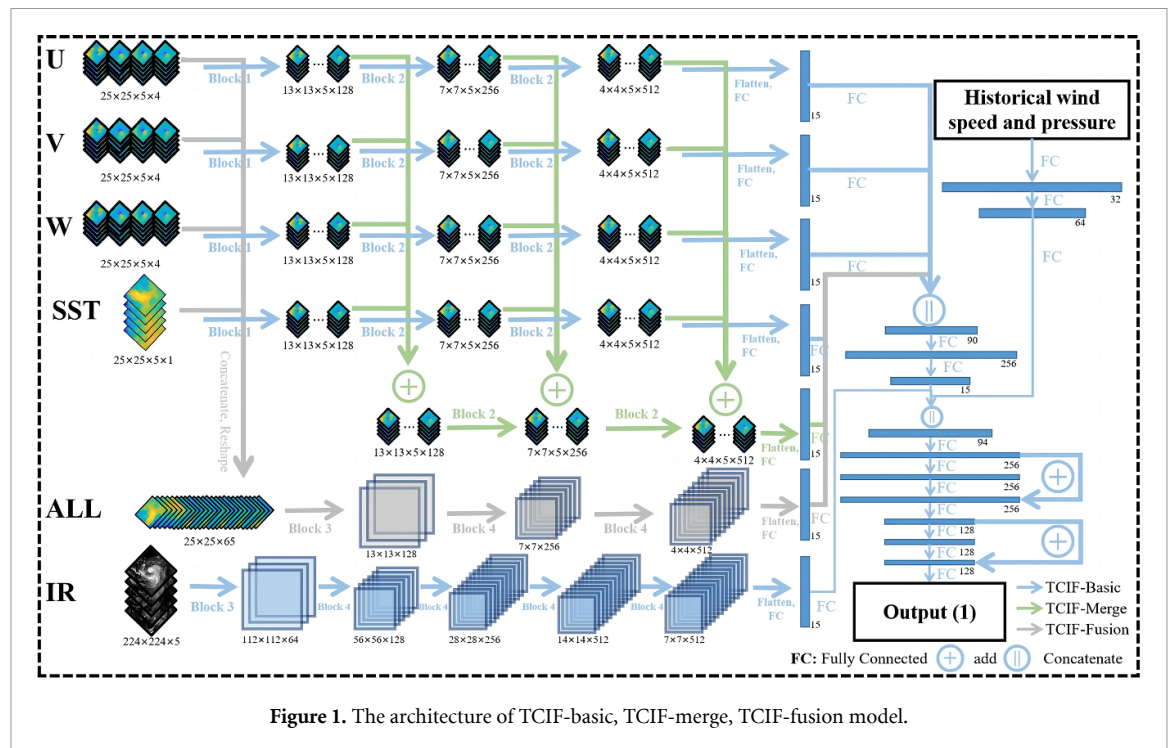
methods, gradient-weighted class activation mapping (Grad-CAM) (Selvaraju *et al* 2020) allows the computation of gradients from the model's output back to its input, thus generating heatmaps. The Grad-CAM method has been employed to interpret DL models for TC study (Baek *et al* 2022, Tian *et al* 2022). These heatmaps signify the relevance and importance of input features to the model's output, with regions exhibiting higher values deemed crucial by the DL model for TC intensity forecasting. It is referred to as 'MK'.

As shown in figure 2, the heatmap assigns higher weights to regions with high wind speeds, while regions characterized by lower wind speeds and areas deemed insignificant by other models receive comparatively lower weights. Specifically, the areas with higher weights in the *U* and *V* heatmaps are situated near the TC center, where strong wind speeds are predominant. Notably significant weight regions in the *W* heatmap are observed both within the TC eye and in its peripheral areas. The SST heatmap closely mirrors the genuine SST distribution. Furthermore, regions exhibiting higher weights in the IR heatmap correspond to the TC's eye and its spiral bands. A new dataset is generated through an element-wise multiplication of the original data with the absolute values of the heatmap (figure 2). This novel dataset empowers the model to strengthen the extraction of features in regions exhibiting high wind speeds.

In light of this, a method is proposed that effectively utilizes 'MK' to guide the modeling process. The proposed approach involves the following steps (supplementary figure 3): (1) training a TCIF-fusion model, denoted as the TCIF-initial model; (2) utilizing Grad-CAM to produce heatmaps for the training, validation, and test datasets; (3) creating a new input data by element-wise multiplication of the absolute value of the heatmaps with the corresponding original input data points (figure 2); (4) training a new TCIF-fusion model using the augmented dataset, thereby obtaining a TCIF-fusion model (the MK-guided TCIF-fusion model) that is guided by the acquired MK. Note that both the TCIF-initial Model and the MK-guided TCIF-fusion model employ the same training, validation, and test datasets, ensuring the absence of any data leakage issue.

3. Results and discussion

3.1. The roles of SST and IR images in TCIFs

TCs are complex weather phenomena influenced by multiple factors. Most existing DL-based methods have mainly considered atmospheric factors (such as HIS, *U*, *V*, *W*), while Zhang *et al* (2022) and Ma *et al* (2023) have included oceanic factors (SST). The studies indicate that IR data can depict TC morphology and convective activities and is commonly used for rapid intensification forecasting. However, no DL-based TC intensity forecasting method has incorporated IR. This section compares the performance of the MK-guided TCIF-fusion model under varying inputs. As shown in table 1 and figure 3(a), the basic model employs HIS, *U*, *V*, and *W* as inputs while introducing SST and IR inputs individually. Table 1 and figure 3(a) depicts the average absolute errors of
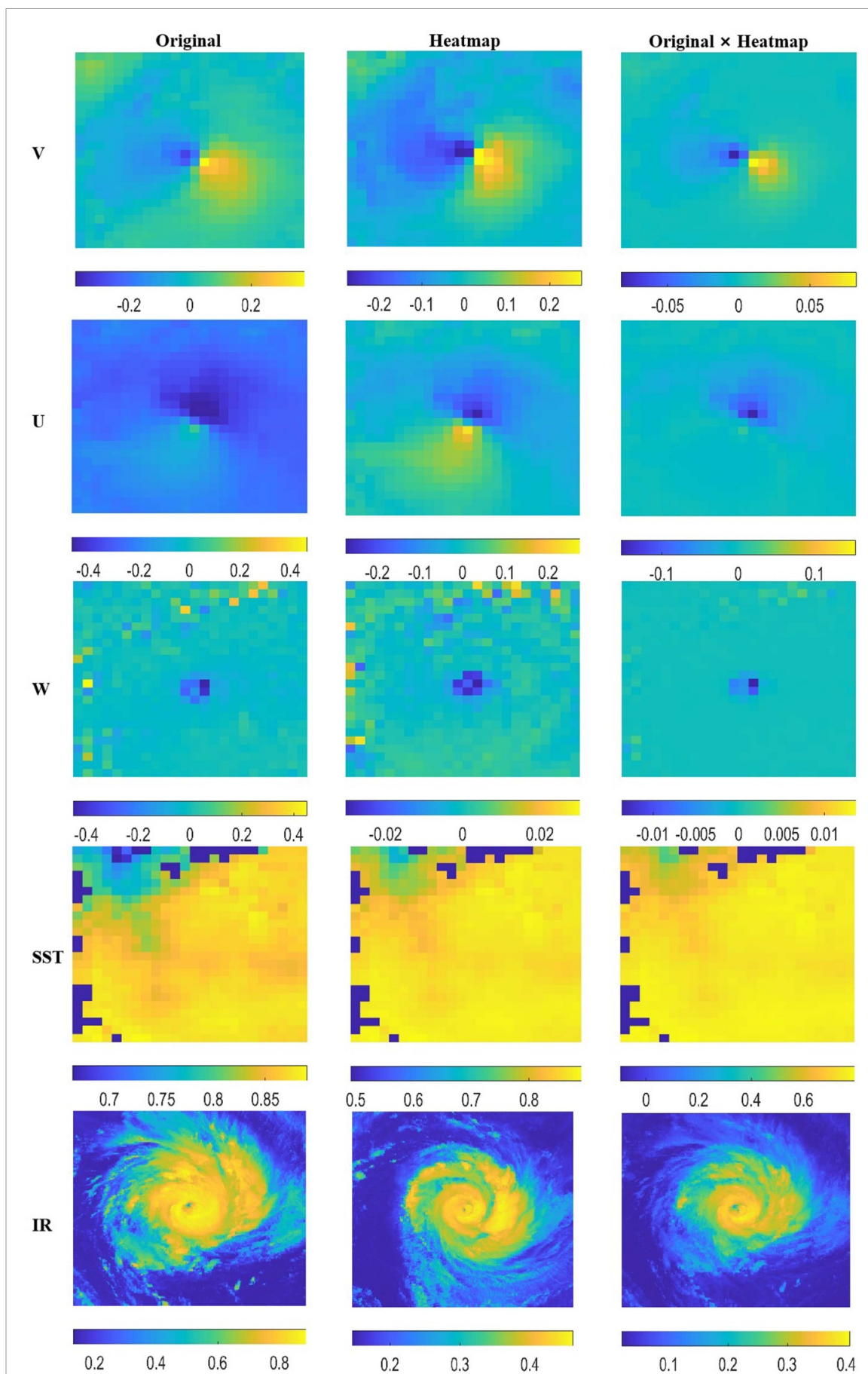
Figure 2. The process of MK-guided generation of new input data. The deep blue dots in the SST represent processed NaN values, please refer to the supplementary information section S1 for details.

Table 1. The mean absolute errors (MAE, m s^{-1}) over the Northwest Pacific test data (2020–2021) of 24 h TC intensity forecasting with different model inputs.

ID	Input	Model (data structure)	MAE
TCIF-1	HIS, *U*, *V*, *W*	MK-guided TCIF-fusion	3.87
TCIF-2	HIS, *U*, *V*, *W*, SST	25 × 25 × 5×	3.74
TCIF-3	HIS, *U*, *V*, *W*, IR		3.69
TCIF-4	HIS, *U*, *V*, *W*, IR, SST		3.56

the model's 24 h forecasts for test data spanning 2020 to 2021.

When input into HIS, *U*, *V*, and *W*, the MK-guided TCIF-1 model registers an error of 3.8 m s^{-1}. However, with the integration of SST and IR inputs, the model's forecast errors are notably reduced to 3.74 m s^{-1} (MK-guided TCIF-2 model) and 3.69 m s^{-1} (MK-guided TCIF-3 model), respectively. The MK-guided TCIF-4 model, incorporating HIS, *U*, *V*, *W*, IR, and SST inputs in concert, achieves even further error reduction, yielding an error of 3.56 m s^{-1} and a performance improvement of 8.0% (compared with MK-guided TCIF-1).

These findings underscore a significant observation: apart from wind speed components directly correlating with intensity, incorporating SST and IR inputs greatly enhances the model's performance in forecasting TC intensity.

3.2. The roles of MK in TCIFs

The original input data contains extraneous 'noise' beyond the information relevant to TC intensity, which could impede the model's learning process. Introducing MK can effectively reduce the impact of 'noise' on the model. For details on the method of introducing model knowledge, please refer to the Method section.

The progression of model training is represented by the loss curve presented in figure 3(b). As shown in figure 3(b), the solid and dashed red lines denote the model's training and validation loss values with the original input data. Correspondingly, the solid and dashed blue lines represent the training and validation loss values of the MK-guided TCIF-fusion model. The results underscore that the TCIF-fusion model without MK-guided converges after 40 training epochs, while the MK-guided TCIF-fusion model achieves convergence within just 20 training epochs, displaying even lower loss values.

Figures 3(c) and (d) illustrate the forecasts of the TCIF-fusion model without MK-guided and the MK-guided TCIF-fusion model against the test dataset. Notably, the MK-guided TCIF-fusion model exhibits a higher correlation (0.92) and lower MAE (3.56 m s^{-1}). Figures 3(c) and (d) demonstrate that incorporating MK mitigates the model's tendency to

underestimate high wind speed samples. The results strongly indicate that incorporating MK contributes to enhanced model performance.

3.3. Results analysis

Forecasting results are compared between the MK-guided TCIF-fusion model, official subjective forecasts (table 2), dynamical models, and DL-based methods. As a result, superior or equivalent forecasting performance is achieved. The STI/CMA annually evaluates various operational dynamical models for forecasting Northwest Pacific TCs (Chen *et al* 2022, Yang *et al* 2023). Official subjective forecasts from CMA, Japan Meteorological Agency (JMA), and Joint Typhoon Warning Center (JTWC), as well as dynamical model forecasts from European Centre for Medium-Range Weather Forecasts (ECMWF) and National Centers for Environmental Prediction / Global Forecast System (NCEP/GFS), are listed in table 2 (Chen *et al* 2022, Yang *et al* 2023). The results reveal that both the DL-based methods and our proposed approach outperform traditional methods by 4%–22%, underscoring the substantial potential of DL in the TCIF. In contrast to other DL-based methods, the approach proposed in this paper (1) adds satellite imagery as input, (2) prioritizes the learning of interactions among factors, and (3) employs model-guided knowledge for modeling. The performance of the MK-guided TCIF-fusion model is more than 4% higher than that of other deep learning methods, proving that the method is advanced.

Additionally, the forecasting results of the MK-guided TCIF-fusion model have been compared across different TC intensities (figure 4). Figure 4(a) depicts a bar chart illustrating the forecast errors for various intensity levels. The forecast error of our model increases with higher TC intensity levels. In figure 4(b), different TC intensity occurrence frequencies are showcased in both observations and model forecasts. Notably, the occurrence frequencies of distinct intensity of TCs are quite comparable between our model and the observations. Our model tends to make fewer forecasts for TCs of tropical storm (TS) and SuperTY intensity and more forecasts for tropical depression (TD) (10.8–17.1 m s^{-1}), severe tropical storm (STS) (24.5–32.6 m s^{-1}), and typhoon (TY) (32.7–41.4 m s^{-1}) intensity.

The frequency distribution illustrated in figure 4(b) implies that our model might misclassify TS (17.2–24.4 m s^{-1}) TCs as TD and SuperTY (>51.0 m s^{-1}) TCs as STS (41.5–50.9 m s^{-1}) or TY. This tendency could stem from the varying sample sizes of TCs across different intensities, highlighting a limitation of DL methods. This issue poses a challenge to address in the future.

When a TC makes landfall, the transition from ocean to land alters the TC's energy source, encountering friction and changes in terrain, thereby
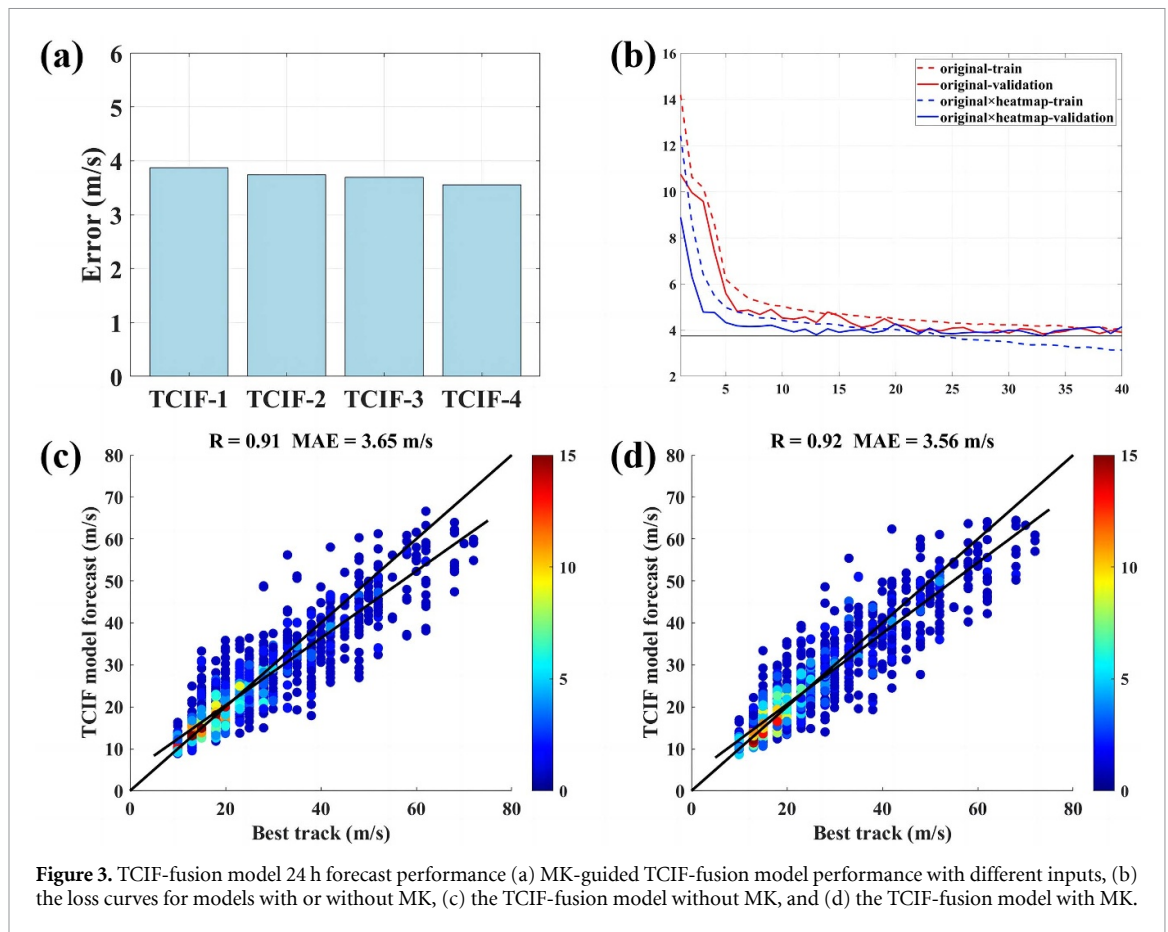
Figure 3. TCIF-fusion model 24 h forecast performance (a) MK-guided TCIF-fusion model performance with different inputs, (b) the loss curves for models with or without MK, (c) the TCIF-fusion model without MK, and (d) the TCIF-fusion model with MK.
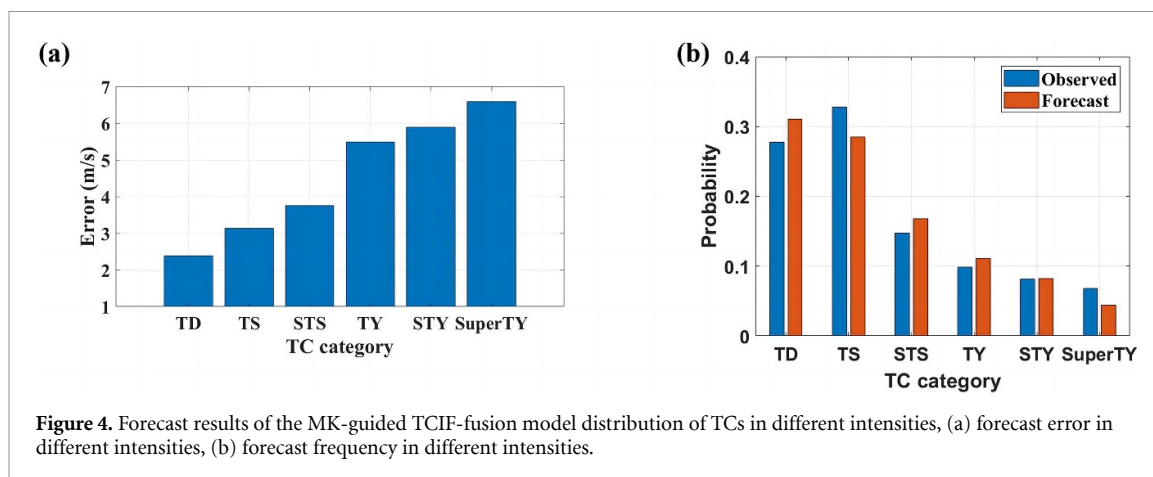
Table 2. The comparison of the MAEs of the MK-guided TCIF-fusion model and other methods in forecasting TC intensity.

	Method	Test data	Region	MAE (2020/2021)
DL	SAF-Net (Xu *et al* 2022)	2015–2018	Northwest Pacific	4.30
	PTCIF (Meng *et al* 2023b)	2015–2018	Northwest Pacific	4.60
	TC-Pred (Zhang *et al* 2022)	2019–2020	Northwest Pacific	3.98
	Pre_3D (Ma *et al* 2023)	2014–2019	Northwest Pacific	3.72
	MLP (Xu *et al* 2021)	2019–2020	Atlantic	4.22
	TCP-NGBoost (Meng *et al* 2023a)	2019–2020	Atlantic	4.49
Subjective Forecasts	CMA	2020–2021	Northwest Pacific	4.50/4.30
	JMA	2020–2021	Northwest Pacific	4.50/4.30
	JTWC	2020–2021	Northwest Pacific	4.90/4.60
Dynamical Model	ECMWF-IFS	2020–2021	Northwest Pacific	6.70/8.00
	NCEP-GFS	2020–2021	Northwest Pacific	4.70/5.90
Our	MK-guided TCIF-fusion	2020–2021	Northwest Pacific	3.56 (3.64/3.51)

posing difficult challenges for accurate forecasting. In 2020 and 2021, there were 33 instances of landfalling TCs, with 19 in 2020 and 14 in 2021. For the MK-guided TCIF-fusion model, CMA, JMA, and JTWC, the TCIF errors for landfalling TCs in 2020 were $4.55\,\mathrm{m\,s^{-1}}$, $4.95\,\mathrm{m\,s^{-1}}$, $5.31\,\mathrm{m\,s^{-1}}$, and $4.87\,\mathrm{m\,s^{-1}}$. In the case of landfalling TCs in 2021, the intensity forecast errors were $4.46\,\mathrm{m\,s^{-1}}$, $6.42\,\mathrm{m\,s^{-1}}$,

$5.89\,\mathrm{m\,s^{-1}}$, and $5.43\,\mathrm{m\,s^{-1}}$. The results show that the MK-guided TCIF-fusion model has greater stability and accuracy in forecasting the intensity of landfalling TCs, but further improvement is needed.

These results show that the MK-guided TCIF-fusion model can provide accurate 24 h TCIFs, delivering crucial decision support for forecasters.

Figure 4. Forecast results of the MK-guided TCIF-fusion model distribution of TCs in different intensities, (a) forecast error in different intensities, (b) forecast frequency in different intensities.

4. Conclusions

TCs represent some of the most formidable natural disasters, underscoring the pivotal significance of precise TCIFs. Despite achieving notable forecasting accuracy levels, DL-based methods continue to grapple with certain challenges. This paper verifies the advantageous role of SST and IR imagery in enhancing TCIF. By merging SST and IR data with atmospheric factors, the performance of the DL model experiences an 8.0% enhancement. Given the highly intricate interactions governing the nonlinear dynamics of TCs involving multiple factors, this research introduces a model design that explicitly accounts for inter-factor dependencies. This design augments the model's capacity to capture the detailed evolution of TC intensity, effectively reducing intensity forecast errors. Furthermore, integrating model-guided knowledge during the modeling process mitigates the interference from environmental 'noise', subsequently amplifying both learning speed and model performance.

The MK-guided TCIF-fusion model delivers a 24 h forecast error of 3.56 m s^{-1} for Northwest Pacific TCs spanning 2020–2021. This method is comparable to, or surpasses, traditional and DL-based TC intensity prediction methods.

Given the complexity of multi-factor interactions inherent in the TC phenomenon, the tailored MK-guided TCIF-fusion model augments the comprehension of inter-factor relationships and effectively mitigates environmental noise interference in forecasting. This model framework holds relevance for TCIFs and TC track forecasts, rainfall forecasting, etc.

Regrettably, similar to existing DL-based methods, this study relies on reanalysis data that lacks accessibility for operational forecasting. Pioneering the development of purely satellite image-based TC forecasting models is an impending challenge. Furthermore, considering TC processes' intricate and dynamic essence, integrating physical constraints or prior knowledge into DL models is a promising way for future research within this domain.

Data availability statement

The IR images are downloaded from www.ncei.noaa. gov/products/gridded-geostationary-brightness-temperature. The CMA best track data can be downloaded from https://tcdata.typhoon.org.cn/zjljsjj. html. The ERA5 reanalysis data are downloaded from https://cds.climate.copernicus.eu/cdsapp#!/ dataset/reanalysis-era5-single-levels?tab=form and https://cds.climate.copernicus.eu/cdsapp#!/dataset/ reanalysis-era5-pressure-levels?tab=form.

The code is available in GitHub: https://github. com/wangchong96/TCIF-fusion.

ORCID iDs

Chong Wang ⬤ https://orcid.org/0000-0002-8275-8450
Xiaofeng Li ⬤ https://orcid.org/0000-0001-7038-5119

References

Adler R F and Rodgers E B 1977 Satellite-observed latent heat release in a tropical cyclone *Mon. Weather Rev.* **105** 956–63
Baek Y-H, Moon I-J, Im J and Lee J 2022 A novel tropical cyclone size estimation model based on a convolutional neural network using geostationary satellite imagery *Remote Sens.* **14** 426

Baik J-J and Paek J-S 2000 A neural network model for predicting typhoon intensity *J. Meteorol. Soc. Japan* II **78** 857–69

Bao S, Zhang Z, Kalina E and Liu B 2022 The use of composite GOES-R satellite imagery to evaluate a TC intensity and vortex structure forecast by an FV3GFS-based hurricane forecast model *Atmosphere* **13** 126

Bhatia K *et al* 2022 A potential explanation for the global increase in tropical cyclone rapid intensification *Nat. Commun.* **13** 6626

Cangialosi J P, Blake E, DeMaria M, Penny A, Latto A, Rappaport E and Tallapragada V 2020 Recent progress in tropical cyclone intensity forecasting at the national hurricane center *Weather Forecast.* **35** 1913–22

Chen G, Yang M, Zhang X, Bai L, Wan R and Cao Q 2022 Verification on forecasts of typhoons over Western North Pacific and South China sea in 2020 *Meteorol. Mon.* **48** 516–25

Chen P, Yu H and Chan J C 2011 A western North Pacific tropical cyclone intensity prediction scheme *Acta Meteorol. Sin.* **25** 611–24

DeMaria M 2009 A simplified dynamical system for tropical cyclone intensity prediction *Mon. Weather Rev.* **137** 68–82

DeMaria M and Kaplan J 1994 A statistical hurricane intensity prediction scheme (SHIPS) for the atlantic basin *Weather Forecast.* **9** 209–20

DeMaria M and Kaplan J 1999 An updated statistical hurricane intensity prediction scheme (SHIPS) for the atlantic and eastern North Pacific basins *Weather Forecast.* **14** 326–37

DeMaria M, Mainelli M, Shay L K, Knaff J A and Kaplan J 2005 Further improvements to the statistical hurricane intensity prediction scheme (SHIPS) *Weather Forecast.* **20** 531–43

DeMaria M, Sampson C R, Knaff J A and Musgrave K D 2014 Is tropical cyclone intensity guidance improving? *Bull. Am. Meteorol. Soc.* **95** 387–98

Emanuel K A 1986 An air-sea interaction theory for tropical cyclones. Part I: steady-state maintenance *J. Atmos. Sci.* **43** 585–605

Graves A 2012 *Long Short-Term Memory* (Springer) pp 37–45

Hersbach H *et al* 2023 ERA5 hourly data on pressure levels from 1940 to present. Copernicus climate change service (C3S) climate data store (CDS) (accessed 01 June 2022)

Klotzbach P J, Bowen S G, Pielke R and Bell M 2018 Continental US hurricane landfall frequency and associated damage: observations and future risks *Bull. Am. Meteorol. Soc.* **99** 1359–76

Knaff J A, DeMaria M, Sampson C R and Gross J M 2003 Statistical, 5-day tropical cyclone intensity forecasts derived from climatology and persistence *Weather Forecast.* **18** 80–92

Knapp K R *et al* 2011 Globally gridded satellite (GridSat) observations for climate studies 2 *Bull. Am. Meteorol. Soc.* **92** 893–907

Lagerquist R, McGovern A, Homeyer C R, Gagne D J I I and Smith T 2020 Deep learning on three-dimensional multiscale data for next-hour tornado prediction *Mon. Weather Rev.* **148** 2837–61

Landsea C W and Cangialosi J P 2018 Have we reached the limits of predictability for tropical cyclone track forecasting? *Bull. Am. Meteorol. Soc.* **99** 2237–43

Li X, Liu B, Zheng G, Ren Y, Zhang S, Liu Y, Gao L, Liu Y, Zhang B and Wang F 2020 Deep-learning-based information mining from ocean remote-sensing imagery *Natl Sci. Rev.* **7** 1584–605

Li X, Zhou Y and Wang F 2022 Advanced information mining from ocean remote sensing imagery with deep learning *J. Remote Sens.* **2022** 9849645

Li Y, Tang Y, Wang S, Toumi R, Song X and Wang Q 2023 Recent increases in tropical cyclone rapid intensification events in global offshore regions *Nat. Commun.* **14** 5167

Lin I-I, Chen C-H, Pun I-F, Liu W T and Wu C-C 2009 Warm ocean anomaly, air sea fluxes and the rapid intensification of tropical cyclone Nargis (2008) *Geophys. Res. Lett.* **36**

Ma D, Wang L, Fang S and Lin J 2023 Tropical cyclone intensity prediction by inter-and intra-pattern fusion based on multi-source data *Environ. Res. Lett.* **18** 014020

Ma L-M 2014 Research progress on China typhoon numerical prediction models and associated major techniques *Prog. Geophys.* **29** 1013–22

Meng F, Yang K, Yao Y, Wang Z and Song T 2023b Tropical cyclone intensity probabilistic forecasting system based on deep learning *Int. J. Intell. Syst.* **2023** 1–17

Meng F, Yao Y, Wang Z, Peng S, Xu D and Song T 2023a Probabilistic forecasting of tropical cyclones intensity using machine learning model *Environ. Res. Lett.* **18** 044042

Pan B, Xu X and Shi Z 2019 Tropical cyclone intensity prediction based on recurrent neural networks *Electron. Lett.* **55** 413–5

Ravuri S *et al* 2021 Skilful precipitation nowcasting using deep generative models of radar *Nature* **597** 672–7

Ren Y and Li X 2023 Predicting the daily sea ice concentration on a sub-seasonal scale of the pan-arctic during the melting season by a deep learning model *IEEE Trans. Geosci. Remote Sens.* **61** 1–15

Saha S *et al* 2014 The NCEP climate forecast system version 2 *J. Clim.* **27** 2185–208

Sandery P, Brassington G, Craig A and Pugh T 2010 Impacts of ocean–atmosphere coupling on tropical cyclone intensity change and ocean prediction in the australian region *Mon. Weather Rev.* **138** 2074–91

Selvaraju R R, Cogswell M, Das A, Vedantam R, Parikh D and Batra D 2020 Continental us hurricane landfall frequency and associated damage: observations and future risks *Int. J. Comput. Vis.* **128** 336–59

Shen D, Bao S, Pietrafesa L J and Gayes P 2022 Improving numerical model predicted float trajectories by deep learning *Earth Space Sci.* **9** e2022EA002362

Steranka J, Rodgers E B and Gentry R C 1986 The relationship between satellite measured convective bursts and tropical cyclone intensification *Mon. Weather Rev.* **114** 1539–46

Su H, Wu L, Jiang J H, Pai R, Liu A, Zhai A J, Tavallali P and DeMaria M 2020 Applying satellite observations of tropical cyclone internal structures to rapid intensification forecast with machine learning *Geophys. Res. Lett.* **47** e2020GL089102

Tang B and Emanuel K 2010 Midlevel ventilation's constraint on tropical cyclone intensity *J. Atmos. Sci.* **67** 1817–30

Tian W, Zhou X, Niu X, Lai L, Zhang Y and Sian K T C L K 2022 A lightweight multitask learning model with adaptive loss balance for tropical cyclone intensity and size estimation *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **16** 1057–71

Vecchi G A and Soden B J 2007 Effect of remote sea surface temperature change on tropical cyclone potential intensity *Nature* **450** 1066–70

Wang C and Li X 2023a A deep learning model for estimating tropical cyclone wind radius from geostationary satellite infrared imagery *Mon. Weather Rev.* **151** 403–17

Wang C, Xu Q, Cheng Y, Pan Y and Li H 2022 Ensemble forecast of tropical cyclone tracks based on deep neural networks *Front. Earth Sci.* **16** 671–7

Wang H, Hu S and Li X 2023 An interpretable deep learning enso forecasting model *Ocean Land Atmos. Res.* **2** 0012

Wang H and Li X 2023b Deepblue: advanced cnn applications for ocean remote sensing *IEEE Geosci. Remote Sens. Mag.* **2**–25

Wang S and Toumi R 2021 Recent migration of tropical cyclones toward coasts *Science* **371** 514–7

Wang S and Toumi R 2022 An analytic model of the tropical cyclone outer size *npj Clim. Atmos. Sci.* **5** 46

Weber H C 2003 Hurricane track prediction using a statistical ensemble of numerical models *Mon. Weather Rev.* **131** 749–70

Woodruff J D, Irish J L and Camargo S J 2013 Coastal flooding by tropical cyclones and sea-level rise *Nature* **504** 44–52

Xu G, Lin K, Li X and Ye Y 2022 SAF-Net: a spatio-temporal deep learning method for typhoon intensity prediction *Pattern Recognit. Lett.* **155** 121–7

Xu W, Balaguru K, August A, Lalo N, Hodas N, DeMaria M and Judi D 2021 Deep learning experiments for tropical cyclone intensity forecasts *Weather Forecast.* **36** 1453–70

Yang M, Chen G, Zhang X, Bai L, Wan R and Cao Q 2023 Verification on typhoons forecasts over the western North Pacific and the South China sea in 2021 *Meteorol. Mon.* **49** 1005–19

Ying M, Zhang W, Yu H, Lu X, Feng J, Fan Y, Zhu Y and Chen D 2014 An overview of the China meteorological administration tropical cyclone database *J. Atmos. Ocean. Technol.* **31** 287–301

Yu Z, Chen Y J, Ebert B, Davidson N E, Xiao Y, Yu H and Duan Y 2020 Benchmark rainfall verification of landfall tropical cyclone forecasts by operational ACCESS-TC over China *Meteorol. Appl.* **27** e1842

Yuan S, Wang C, Mu B, Zhou F and Duan W 2021 Typhoon intensity forecasting based on LSTM using the rolling forecast method *Algorithms* **14** 83

Zhang X, Fang J and Yu Z 2023 The forecast skill of tropical cyclone genesis in two global ensembles *Weather Forecast.* **38** 83–97

Zhang X and Li X 2022 Satellite data-driven and knowledge-informed machine learning model for estimating global internal solitary wave speed *Remote Sens. Environ.* **283** 113328

Zhang Z, Yang X, Shi L, Wang B, Du Z, Zhang F and Liu R 2022 A neural network framework for fine-grained tropical cyclone intensity prediction *Knowl.-Based Syst.* **241** 108195

Zheng F, Sun Y, Yang Q and Mu L 2021 Evaluation of arctic sea-ice cover and thickness simulated by MITgcm *Adv. Atmos. Sci.* **38** 29–48

Zheng G, Li X, Zhang R-H and Liu B 2020 Purely satellite data–driven deep learning forecast of complicated tropical instability waves *Sci. Adv.* **6** eaba1482

Zheng G, Liu J, Yang J and Li X 2019 Automatically locate tropical cyclone centers using top cloud motion data derived from geostationary satellite images *IEEE Trans. Geosci. Remote Sens.* **57** 10175–90

Zheng G, Yang J, Liu A K, Li X, Pichel W G and He S 2015 Comparison of typhoon centers from SAR and IR images and those from best track data sets *IEEE Trans. Geosci. Remote Sens.* **54** 1000–12

Zheng W, Ek M, Mitchell K, Wei H and Meng J 2017 Improving the stable surface layer in the NCEP global forecast system *Mon. Weather Rev.* **145** 3969–87

Zheng Z-W, Lin J-Y, Gopalakrishnan G, Chen Y-R, Doong D-J, Ho C-R, Zheng Q, Wu C-R and Huang C-F 2023 Extreme cooling of 12.5° C triggered by typhoon Fungwong (2008) *Ocean Model.* **182** 102176