

MSFF-CAA-ResNet: multi-scale and attention-enhanced residual network for tropical cyclone center location*

Zhengtang CHEN^{1,2}, Ye TAO^{1,**}, Chong WANG^{2,**}, Yanli WU¹

¹Qingdao University of Science and Technology, Qingdao 266061, China

²Key Laboratory of Ocean Circulation and Waves, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China

Received Jul. 7, 2025; accepted in principle Oct. 13, 2025; accepted for publication Nov. 22, 2025

© Chinese Society for Oceanology and Limnology, Science Press and Springer-Verlag GmbH Germany, part of Springer Nature 2026

Abstract To locate the tropical cyclone (TC) center is a critical task for reducing the impact of TC. However, traditional deep-learning methods struggle to extract high-dimensional features from the cloud structure of TC, leading to significant errors in the center location. We collected ~16 000 infrared images from the HURSAT-B1 Global Tropical Cyclone Satellite Dataset, covering TC samples from multiple satellite remote sensing platforms and ocean regions, and designed a TC center location model MSFF-CAA-ResNet that is built upon the classic residual structure by introducing the multi-scale feature fusion (MSFF) module and the center-aware attention (CAA) module, which works synergistically to enhance the model's ability to locate the TC center position. Experimental results show that the MSFF-CAA-ResNet model could not only adapt to multi-source remote sensing platforms but also achieve high-precision TC center location worldwide. Ablation experiments indicate that both MSFF and CAA contribute to TC center location. The improved MSFF-CAA-ResNet-F model achieves an average location error of 23.9 km, representing a 12.7% reduction compared to the baseline model. Interpretability analysis shows that the MSFF-CAA-ResNet-F model focuses more on the central features of TCs compared to the baseline model. This indicates that the MSFF-CAA-ResNet-F model holds promising application prospects.

Keyword: tropical cyclone (TC); deep learning; multisource data

1 INTRODUCTION

Tropical cyclones (TCs) are intense weather systems that form over tropical oceans and pose a serious threat to human life and property (Wang and Li, 2023b; Wang et al., 2024a, 2025). Accurate determination of the TC center's location plays a crucial role in monitoring TCs. Additionally, TC intensity inversion (Pradhan et al., 2018; Wang et al., 2021a), TC trajectory tracking (Horn et al., 2014), and TC trajectory forecasting (Zhou and Toth, 2020; Wang et al., 2023) also require accurate TC center locations, and these tasks help reduce the losses and impacts caused by TCs (Zhang et al., 2009; Xiao and Xiao, 2010; Giffard-Roisin et al., 2020).

In the subjective experience-based method, the

center of the TC is determined by meteorologists based on their subjective experience. The most representative method is the Dvorak technique (Lee and Lin, 2001; Velden et al., 2006; Zhang et al., 2020). Specifically, meteorologists determine the center location by combining the cloud system structure in satellite images with the development status of the TC. This method relies heavily on subjective judgment and lacks uniform criteria for

* Supported by the National Natural Science Foundation of China (No. 4250060080), the TS Scholar Young Expert Program in 2026 (No. tsqn202507276), the Qingdao Science and Technology for People's Livelihood Demonstration Special Project (No. 25-1-5-cspz-18-nsh), the Postdoctoral Fellowship Program of CPSF (No. GZC20241741), and the Postdoctoral Innovation Project of Shandong Province (No. SDCX-ZG-202501036)

** Corresponding authors: ye.tao@qust.edu.cn; wangchong1@qdio.ac.cn

determination. Despite the automation of the Dvorak method in recent years (Olander and Velden, 2019), errors remain significant when the TC structure is unclear and the eye is not distinct.

Regression methods in physical modeling, such as the spiral curve method, determine the center location of a TC by extracting the spiral lines of the TC cloud system. Yurchak (2007) used a logarithmic spiral model to model the spiral structure in infrared images, achieving high-precision location of the TC center location. However, this method is computationally intensive and time-consuming, and is only applicable to TCs with distinct spiral structures.

Regression methods for physical modeling also include the threshold method (Jin et al., 2019). The threshold method locates the TC center by finding the dense cloud area at the TC center. For example, Fett and Brand (1975) proposed that the dense cloud cluster area at the TC center is approximately elliptical, and the center of the ellipse fitted with the least error is the TC center. Although the threshold method does not require complex data preprocessing, it is difficult to locate weak TCs.

In recent years, deep-learning technology has been widely applied in the field of geosciences (Li et al., 2022; Wang et al., 2024a, b; Wang and Li, 2024a, b; Lee et al., 2025). Researchers have successfully applied it to TC intensity inversion (Pradhan et al., 2018; Wang et al., 2021a), wind radius inversion (Chavas et al., 2022; Wang and Li, 2023c), sea ice identification (Dumitru et al., 2019; Zhang and Li, 2022), and ocean internal wave identification (Zheng et al., 2022; Zhang and Li, 2024). Deep-learning methods have significantly improved the computational efficiency and accuracy of geoscience tasks.

In TC center location studies, the development of deep learning technology has also been thriving (Wang et al., 2020, 2024b). Yang et al. (2019) proposed a twin neural network, which first uses an upstream classification model to determine whether an eye exists in satellite images, and then uses a downstream model to achieve precise location of the TC center, with an average location error of 28.6 km. Wang et al. (2019) designed a parallel deep-learning method for TC intensity inversion and TC center location, with a location error of 54 km. Subsequently, Wang et al. (2021b) introduced reinforcement learning methods, modeling the process of searching for the TC center as a Markov decision process, and dynamically adjusting the

search window to reduce the location error to 29.4 km. Additionally, Tan (2021) developed an end-to-end U-Net model that locates the TC center location by outputting a heatmap. Hu and Zou (2021) employed azimuthal spectral analysis as the center location algorithm, achieving a location error of 33.8 km for tropical storm (TS). Liu and Zhang (2022) developed the Cloud Motion Wind algorithm, achieving a location error of 41 km. Wang and Li (2023a) utilized a ResNet50 model constructed with ImageNet pre-trained parameters to achieve a location error of 29.3 km. Zhou et al. (2024b) employed a Chinese geostationary satellite location algorithm to achieve an average absolute error of 29.4 km in the western Pacific region.

Although deep learning has made significant progress in TC center location research, existing models still face many technical challenges. First, existing models often struggle to extract deep features of TC cloud systems. Secondly, existing models are unable to effectively capture the correlation between local details and global structural information of TCs. Thirdly, most models are based on a single satellite platform or focus on specific ocean basins, and are unable to achieve cross-satellite platform and cross-ocean TC center location.

To address the above issues, we proposed the MSFF-CAA-ResNet model, which combines the multi-scale feature fusion (MSFF) and center-aware attention (CAA) modules. The MSFF module extracts multi-scale TC features, while the CAA module dynamically enhances and adjusts the network's attention to the center region of the TC, thereby further improving the accuracy and robustness of TC center location.

We developed the MSFF-CAA-ResNet-F model to locate TC center using high temporal and spatial coverage of satellite images. Our goal is to develop an objective and efficient TC center location method. We organize the paper as follows. Section 2 introduces the data and data preprocessing. The MSFF-CAA-ResNet model and experimental schemes are presented in Section 3. The performance analysis and discussion are given in Section 4. The conclusions are given in Section 5.

2 DATA

2.1 IR satellite image of TCs

The data used in this study are derived from the Hurricane Satellite Image Dataset (HURSAT-B1).

The HURSAT-B1 dataset originates from the International Satellite Cloud Climatology Program (ISCCP), which provides global TC infrared (IR) images in NetCDF format. HURSAT-B1 data are obtained from observations by geostationary satellite constellations (such as the GOES, Meteosat, GMS, and FY2 series). It is worth noting that the HURSAT-B1 dataset takes the TC center coordinates provided by the IBTrACS best track data as the reference center of each image. Satellite images with a fixed size of 301×301 pixels are sampled and stored at an interval of three hours. In this paper, the TC center is defined as the latitude and longitude of the surface (or near-surface) circulation center.

We selected three channels—IRWIN (wavelength $10.6\text{--}12.5\ \mu\text{m}$), IRWVP ($6.47\text{--}7.04\ \mu\text{m}$), and IRNIR ($3.78\text{--}3.98\ \mu\text{m}$)—for TC center location (Fig.1). The spatial and temporal resolutions of HURSAT-B1 data are 7 km and 3 h, respectively. We divided the HURSAT-B1 data into two time periods: 2010 to 2015 for training data (approximately 85%) and 2016 for validation and testing (approximately 15%). Among them, the test set and validation set are evenly split by time, with the splitting time point set in August 2016.

As shown in Table 1, all data are categorized into different ocean regions. Data are categorized into different ocean regions. Among them, the Northeast Pacific region contains the largest number of samples (7 241 samples), followed by the North Atlantic (4 151 samples), the Southern Hemisphere

(2 100 samples), and the Western North Pacific (1 912 samples). In contrast, the Indian Ocean has the smallest number of samples, with only 396 samples. Overall, the dataset exhibits a pronounced regional imbalance, with many samples concentrated in the northeast Pacific and North Atlantic basins.

2.2 Data preprocessing

We used satellite images from three bands of the HURSAT-B1 dataset, with an image size of 301×301 pixels and a coverage area of $2\ 100\ \text{km} \times 2\ 100\ \text{km}$. To enhance the diversity of training samples and facilitate the labeling of ground truth labels such as location information, we employed a random cropping method to crop the original image (301×301 pixels) into sub-images of 224×224 pixels (Fig.2) for model training.

Specifically, in each original image of 301×301

Table 1 Classification of TC center locating dataset (Saffir-Simpson Hurricane Scale)

Category and the value range (m/s)	Train	Validation	Test
TS (17.1–32.4)	5 369	416	402
H1 (32.5–42.2)	3 596	308	288
H2 (42.3–48.9)	1 889	179	195
H3 (49.0–57.6)	1 265	136	145
H4 (57.7–70.0)	801	96	110
H5 (>70.0)	480	65	60
Total	13 400	1 200	1 200

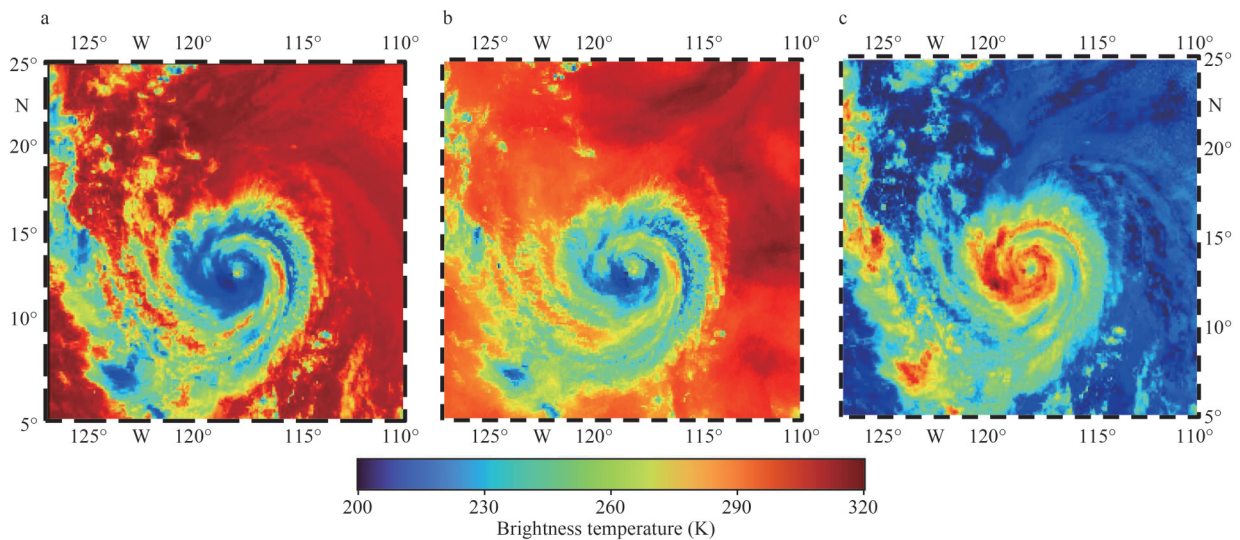


Fig.1 The IR images from different channels

a. IRWIN ($10.6\text{--}12.5\ \mu\text{m}$); b. IRWVP ($6.47\text{--}7.04\ \mu\text{m}$); c. IRNIR ($3.78\text{--}3.98\ \mu\text{m}$). In spatial coverage of $2\ 100\ \text{km} \times 2\ 100\ \text{km}$ (301×301 pixels). The images of TC “DARBY” were acquired at UTC 00:00 on 15 July 2016.

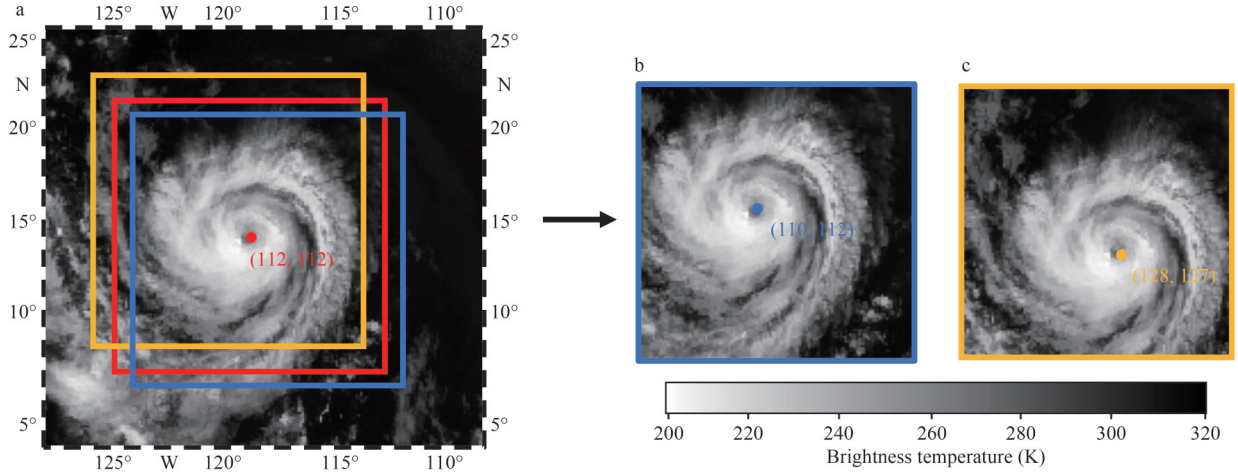


Fig.2 The random cropping strategy

a. the red box indicates a 224×224-pixel region centered on the TC center, with its center point set at coordinates (112, 112). By randomly shifting this red box within the original image, with a maximum displacement of 20 pixels in any direction (up, down, left, or right), regions such as those marked by the yellow and blue boxes are formed; b–c. present examples of images after undergoing the random cropping process.

pixels, a cropping window of 224×224 pixels is first set with the center of the original image as the center. Subsequently, the center of the cropping window is randomly moved 0–20 pixels in the horizontal and vertical directions, respectively, and the final sub-image for TC center location is cropped. For example, when the center of the cropping window is shifted 15 pixels to the left and 5 pixels downward, the center coordinates of the sub-image correspondingly become (112+15=127, 112–5=107). Record this offset value as the label information for each sub-image, i.e., the true coordinates of the TC center. This strategy not only maintains the spatial consistency of the overall TC structure but also enables the deep learning model to learn image spatial displacement features, thereby enhancing the model’s robustness and generalization capabilities in TC location tasks.

3 METHODOLOGY

Given the difficulty of traditional deep learning methods in extracting high-dimensional features from the cloud structure of TC, we replaced the first two stages of the traditional ResNet50 model with the MSFF-CAA module. Figure 3 illustrates the complete workflow of the model for TC center location. First, three-channel satellite images are input, followed by a 7×7 convolutional layer to capture the overall contour of the TC. This is then compressed through a maximum pooling layer to reduce spatial dimensions while retaining key structural features of the TC. Four subsequent stages

follow: the first two stages incorporate MSFF and CAA to effectively extract multi-scale features and enhance feature weights in the TC’s central region. The final two stages utilize traditional bottleneck layers to further reduce computational complexity and strengthen deep features. Finally, the feature map is compressed via a global average pooling layer and mapped to two-dimensional coordinates via a fully connected layer.

3.1 Center-aware attention

To more effectively capture regional features related to the TC center, the MSFF-CAA-ResNet model designed the CAA module. Unlike traditional attention mechanisms, the CAA module has learnable spatial offset parameters that can be combined with Gaussian distributions to generate spatial attention weights, enabling the model to adaptively focus on the central region of the TC and thereby improve location accuracy.

Specifically, let the input feature map be $X \in \mathbb{R}^{B \times C \times H \times W}$, where B is the batch size, C is the number of channels, and H and W represent the spatial dimensions. A normalized coordinate grid $(x, y) \in [0, 1]$ is first constructed over the spatial domain. The learnable attention center is then defined as:

$$c_x = 0.5 + \Delta x, c_y = 0.5 + \Delta y, \quad (1)$$

where Δx and Δy are trainable parameters initialized to zero, representing the deviation from the geometric center.

Using the defined attention center, we construct a

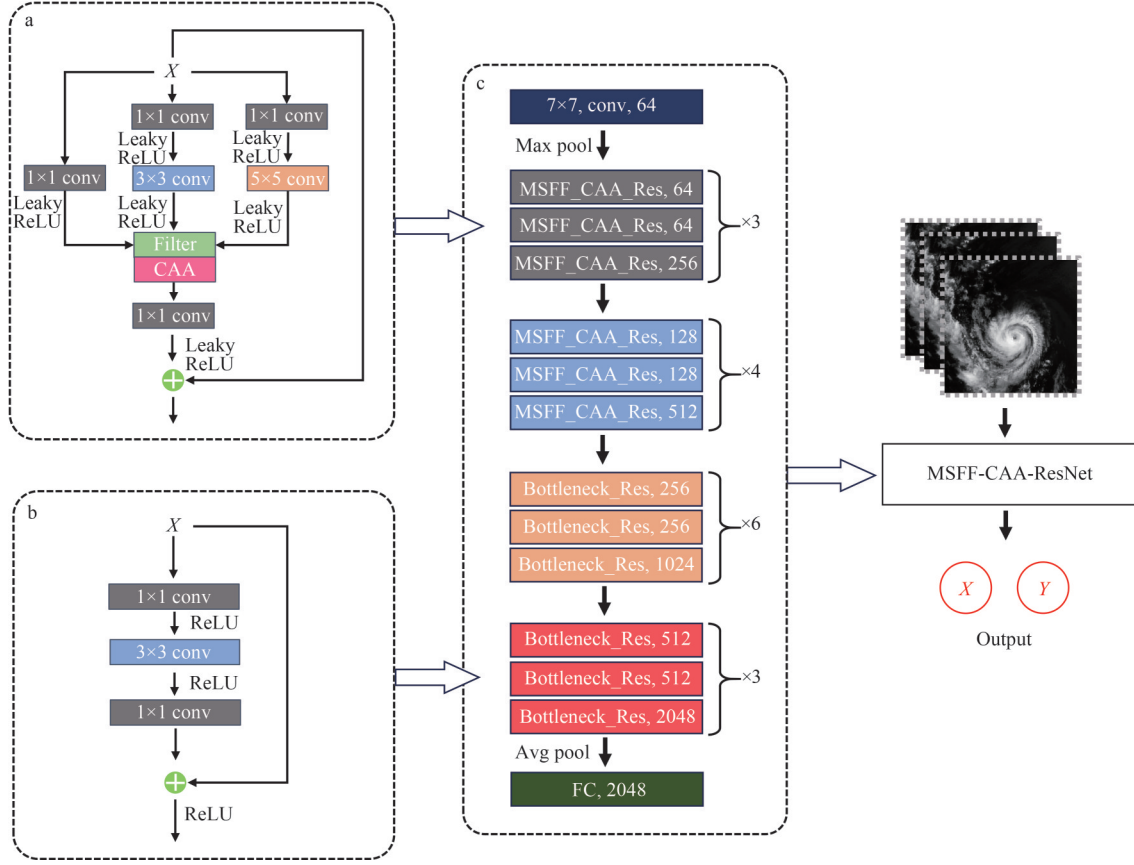


Fig.3 The MSFF-CAA-ResNet model architecture diagram

a. MSFF-CAA module; b. comparison of the original residual block structure with the innovative module in (a); c. the overall architecture of MSFF-CAA-ResNet.

spatial attention map $A(x, y) \in \mathbf{R}^{B \times 1 \times H \times W}$ using a two-dimensional Gaussian function:

$$A(x, y) = \exp\left\{-\frac{[(x-c_x)^2 + (y-c_y)^2]}{(2\sigma^2)}\right\}, \quad (2)$$

where σ is a learnable standard deviation that determines the spread of the attention—smaller values lead to a sharper focus around the center.

The generated attention map is subsequently applied to the input feature map through element-wise (Hadamard) multiplication:

$$X_{\text{out}} = X \odot A, \quad (3)$$

where $\odot$ denotes the Hadamard product (element-wise multiplication), and X_{out} represents the output feature map modulated by spatial attention.

3.2 Multi-scale feature fusion module

TCs of different intensity levels exhibit unique characteristics. Strong TCs have distinct TC eyes, and accurate location can be achieved using only small-scale features near the TC eye. Weak TCs are structurally asymmetric and lack TC eyes, making it impossible to achieve accurate location using only

small-scale features. Instead, large-scale features must be utilized to locate the TC center. However, previous studies have largely ignored the impact of multi-scale features on TC center location. Traditional ResNet architectures typically use fixed-size convolutional kernels within a single convolutional layer, such as a 3×3 kernel (Fig.3b). Inspired by the parallel convolutional concept in the Inception architecture of the GoogleNet model (Szegedy et al., 2015), the MSFF module (Fig.3a) is designed as a structure composed of multiple parallel convolutional branches, each using three different-sized convolutional kernels: 1×1 , 3×3 , and 5×5 . These three different-sized convolutional kernels allow the convolutional layer to extract features from different receptive fields, with larger kernels focusing on global features and smaller kernels focusing on local features. This design not only enriches the expression of multi-scale features but also helps accelerate model convergence and enhance its generalization ability. Then, the features extracted by each convolutional branch are

combined and input into the “filter” and CAA modules. Finally, feature fusion and attention enhancement are achieved through 1×1 convolution operations to improve model performance.

The fusion process can be formally expressed as:

$$F_{\text{fused}} = F(\text{Concat}[F_1, F_2, F_3]), \quad (4)$$

where $F(\cdot)$ denotes the feature fusion operation.

In addition, we chose Leaky ReLU (Xu et al., 2020) instead of ReLU as the activation function. This is because the commonly used ReLU activation function tends to cause permanent neuron inactivation during training, and these inactivated neurons will occupy computing resources, thereby affecting computing efficiency. The Leaky ReLU activation function applies a small slope to input values when the input is negative, addressing the zero-gradient limitation in negative input scenarios and effectively mitigating the issue of neuron death. Additionally, to address the stability issues of standard batch normalization (BN) during model training, the MSFF-CAA-ResNet model replaces BN with Group Normalization (GN), as proposed by Wu and He (2018) during the feature transformation stage. GN independently calculates the mean and variance within each group of channels, avoiding dependence on batch size. This not only better preserves the diversity of features between channels but also demonstrates improved stability during small-batch training.

3.3 Experimental design

All experiments involving the MSFF-CAA-ResNet model were conducted on a Linux Ubuntu system equipped with four NVIDIA Tesla V100 GPUs (each with 32 GB of VRAM). To validate the effectiveness of the MSFF and CAA modules, three sets of experiments were designed, including: (1) ablation experiments on infrared channel combinations: testing the impact of different infrared channel combinations on TC location accuracy; (2) CAA addition location ablation experiments: testing the impact of adding CAA alone to the baseline model, adding MSFF-CAA to different location in the baseline model, and removing the CAA module from the model with the best location accuracy on location accuracy; (3) comparison experiments between the MSFF-CAA-ResNet model and mainstream models.

The MSFF-CAA-ResNet model uses the AdamW optimizer (Zhou et al., 2024a), with an initial learning rate of 5×10^{-4} , a batch size of 4, and 50

training epochs. During training, the learning rate is adjusted using an early stopping strategy. When the validation set loss does not improve for five consecutive epochs, the learning rate is reduced to 0.2 times the original value, with a minimum of 1×10^{-6} . The accuracy of the TC center location model is evaluated using the mean distance error (MDE) between the center identified by the model and the center recorded in the randomly cropped image. The MDE is defined as:

$$\text{MDE} = \frac{1}{N} \sum_{i=1}^N \sqrt{(x_i^{\text{pred}} - x_i^{\text{gt}})^2 + (y_i^{\text{pred}} - y_i^{\text{gt}})^2}, \quad (5)$$

where $(x_i^{\text{pred}}, y_i^{\text{pred}})$ represents the model results of TC center coordinates, $(x_i^{\text{gt}}, y_i^{\text{gt}})$ denotes the ground truth coordinates, and N is the total number of test samples.

4 RESULT AND DISCUSSION

4.1 The impact of infrared channels on TC center location

Table 2 shows the impact of different infrared channel combinations on the overall performance of the TC center location model. All models in Table 2 use the MSFF-CAA-ResNet-F architecture (Fig.4) to ensure that all experimental conditions except for the channel combinations remain consistent. It is worth noting that in this study, TCs with intensities lower than or equal to TS are classified as weak TCs, while those with intensities ranging from H1 to H5 are defined as strong TCs. Moreover, the “Total” category in the table includes both strong and weak TCs.

Table 2 Effects of different input channel combinations based on MSFF-CAA-ResNet-F on the MDE of TC center location

IRWIN	IRWVP	IRNIR	MDE (km)	
			Weak TCs category (TS)	Total category
✓			30.1	25.8
	✓		44.8	34.6
		✓	33.3	28.8
✓	✓		30.8	26.0
✓		✓	27.9	24.1
	✓	✓	49.7	43.5
✓	✓	✓	27.8	23.9

The total category in the table includes both strong and weak TCs. Bold face indicates the best-performing combination.

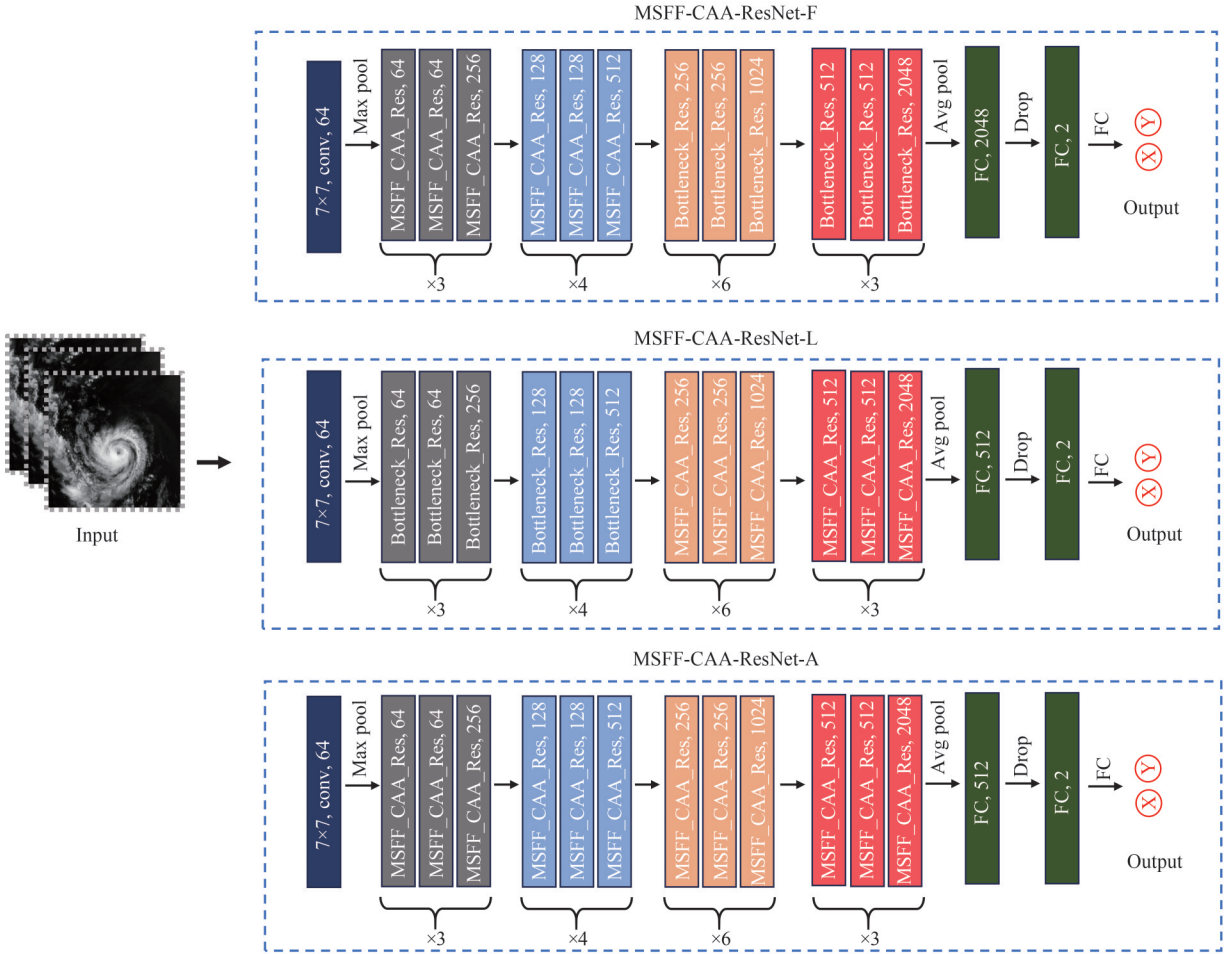


Fig.4 Three types of network structures based on ResNet and MSFF-CAA module, including MSFF-CAA-ResNet-F, MSFF-CAA-ResNet-L, and MSFF-CAA-ResNet-A

In the MSFF-CAA-ResNet-F, the MSFF-CAA modules are located at the front of the network and are involved in the early feature extraction phase, while in the MSFF-CAA-ResNet-L, the MSFF-CAA modules are located at the back of the network and are used for multiscale attentional fusion when acquiring more advanced features. In the MSFF-CAA-ResNet-A, the MSFF-CAA modules are located all over the network and continue to play a role throughout the feature extraction process.

As shown in Table 2, when only the IRWIN, IRWVP, and IRNIR channels are input, the models' MDE for TS intensity category TC are 30.1, 44.8, and 33.3 km, with overall MDEs of 25.8, 34.6, and 28.8 km, respectively. The TC center location model using only IRWIN input performs better than models using only IRWVP and IRNIR channels.

The results indicate that the IRWIN channel performs optimally in the task of locating the center of a TC. At around 11 μm , it can clearly depict the structure of the TC cloud system and the spatial distribution of the cloud top brightness temperature, which are closely related to the structure of the TC and help the model accurately identify and locate the TC center. In contrast, although the IRWVP and IRNIR channels provide supplementary information on water vapor distribution and cloud top

reflectance, respectively, these data contribute limited value to TC center location, resulting in poorer model performance when only these two channels are input.

We further investigated the impact of dual-channel combinations on TC center location. Experimental results showed that compared to using a single channel, most models achieved some performance improvements. For example, the combination of the IRWIN and IRNIR channels reduced the overall MDE to 24.1 km. This indicates that the IRNIR and IRWIN channel information complement each other, enabling the model to capture more detailed features and thereby improving location accuracy. It is worth noting that the fusion of IRWVP and IRNIR channels resulted in a performance decline, with the overall MDE

increasing to 43.5 km and the MDE for TS intensity levels rising to 49.7 km. This suggests that the information provided by the two channels may be redundant or conflicting, making it difficult for the model to learn effectively.

Finally, we investigated the impact of the combination of IRWIN, IRWVP, and IRNIR channels on model performance. The results showed that the three-channel combination input achieved the best TC center location accuracy, with the overall MDE reduced to 23.9 km and the MDE of the TS intensity level reduced to 23.9 km. This reflects that the complementary information between multiple channels enables the model to comprehensively understand the characteristics of the TC center location, thereby enhancing the model's robustness.

4.2 Comprehensive ablation analysis of MSFF modules and CAA mechanisms

This section discusses ablation experiments for the CAA module and MSFF module. The models shown in Table 3 differ only in their model structures, with inputs consisting of combinations of IRWIN, IRWVP, and IRNIR channels.

To evaluate the impact of the CAA module on model performance, Table 3 first compares the baseline model (standard ResNet50) with the ResNet50 model incorporating the CAA module. The results show that adding the CAA module improves the model's accuracy in locating the center of TCs across all intensity levels, indicating that the CAA module effectively guides the model to focus on key regions relevant to TC center location through spatial attention based on Gaussian distribution mapping.

Building on this foundation, the study further

focuses on the role of the MSFF-CAA module. Figure 4 illustrates three variants of the MSFF-CAA module embedded at different location within the model: the prefix "F" (Front) indicates replacing the front part of the feature extraction with the MSFF-CAA module; "L" (Last) indicates replacing the rear module section; and "A" (All) indicates a complete replacement. MSFF-CAA-ResNet-L uses the MSFF-CAA module in the rear part of the model, achieving an MDE of 26.3 km. The MSFF-CAA-ResNet-A model reduces the MDE to 24.6 km through full-process multi-scale feature extraction, outperforming the MSFF-CAA-ResNet-L scheme. MSFF-CAA-ResNet-F achieves an MDE of 23.9 km, the best performance, indicating that introducing multi-scale information early in feature extraction helps the model obtain richer and more detailed structural features. Furthermore, removing the CAA module from the MSFF-CAA-ResNet-F model increases the overall location error, with a particularly noticeable increase in MDE for low-intensity TCs, further validating the positive role of the CAA module in locating weak TCs.

Figure 5 demonstrates the effectiveness of the MSFF/CAA module through visual analysis. Feature maps extracted from the intermediate convolutional layers of the MSFF-CAA-ResNet-F model and the baseline model. The results show that the highly focused regions of the MSFF-CAA-ResNet-F model are closer to the TC center and can effectively suppress the influence of irrelevant cloud systems on TC location. In contrast, the features focused on by ResNet50 are more dispersed and do not eliminate the interference from irrelevant cloud layers.

Figure 6 shows the heat maps generated by the Grad-CAM interpretability method, where higher values indicate that the model considers the location

Table 3 Evaluation of MSFF and CAA module ablation experiments based on three channel inputs (IRWIN, IRWVP, IRNIR)

Model	MDE (km)						
	TS ($n=402$)	H1 ($n=288$)	H2 ($n=195$)	H3 ($n=145$)	H4 ($n=110$)	H5 ($n=60$)	Total ($n=1\ 200$)
ResNet50 (baseline)	32.8	27.0	25.6	23.2	21.9	19.4	27.4
ResNet50 (with CAA)	31.1	29.0	24.9	23.1	19.5	17.8	26.9
MSFF-CAA-ResNet-A	28.5	26.4	23.2	20.4	18.1	16.6	24.6
MSFF-CAA-ResNet-L	31.5	26.4	25.5	22.0	19.2	17.0	26.3
MSFF-CAA-ResNet-F	27.8	25.1	23.2	20.0	17.1	17.0	23.9
MSFF-CAA-ResNet-F (without CAA)	28.0	25.2	23.8	19.3	17.3	14.5	24.0

Bold values in category columns indicate the best performance for each category, while bold text in the Model column denotes the overall best-performing model.

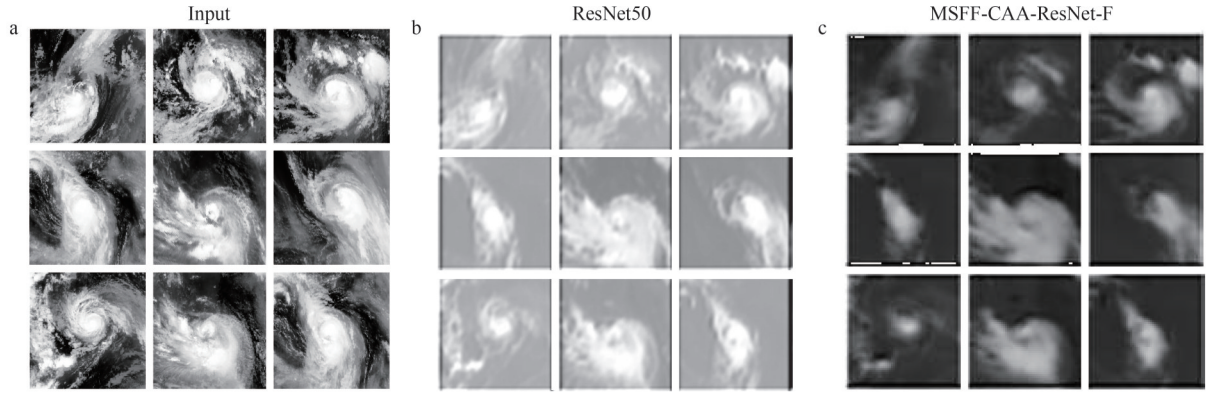


Fig.5 Comparison of feature maps across different models

a. original input image; b. feature map from the 12th convolutional layer of the baseline model (ResNet50); c. feature map from the 12th convolutional layer of the proposed MSFF-CAA-ResNet-F model.

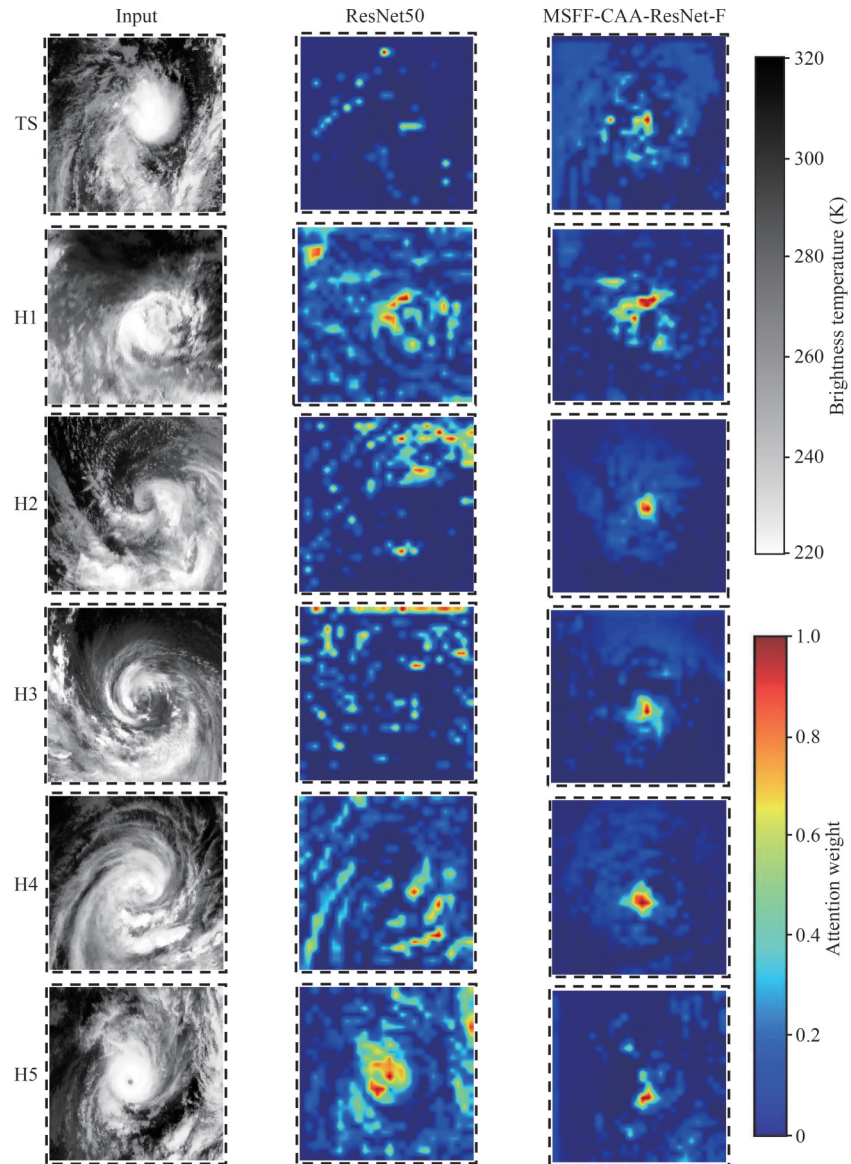


Fig.6 Comparison of TC imagery (input) and Grad-CAM heatmaps across intensities

of the TC to be more important. Compared to the baseline model, MSFF-CAA-ResNet-F accurately focuses on the center of the TC at all TC intensity levels, while the heat map of ResNet50 is more scattered. The results demonstrate the effectiveness and reliability of the MSFF-CAA module under complex meteorological conditions.

4.3 Comparative analysis of MSFF-CAA-ResNet-F and other models

Table 4 lists the advanced TC center location methods proposed in recent years. Since their research does not provide open-source code and data, we are unable to conduct a fair comparison using the same dataset. Therefore, the comparison in Table 4 is for reference only. In addition, to verify the advanced nature of the method, we selected two representative deep learning architectures, ResNet50 and ConvNeXt, for training and evaluation using the same dataset.

The experimental results show that MSFF-CAA-ResNet-F achieves smaller location error than ResNet50 and ConvNeXt on the same dataset. Compared with ResNet50 and ConvNeXt, the MDE of MSFF-CAA-ResNet-F is reduced by 12.8% and 11.8%, respectively, indicating that the MSFF-CAA-ResNet-F model has superior location accuracy.

It is important to note that the methods listed in

Table 4 are based on modeling using data from a single satellite platform or specific sea areas, whereas the MSFF-CAA-ResNet-F model is based on modeling using observational data from multiple satellite platforms covering multiple sea areas. The MSFF-CAA-ResNet-F model can simultaneously achieve TC center location in the Northeast Pacific, North Atlantic, Northwest Pacific, Indian Ocean, and Southern Hemisphere regions. As shown in Table 4, the MSFF-CAA-ResNet-F model achieved an MDE of 21.9 km in the Northeast Pacific and North Atlantic regions, respectively. The MDE is highest in the Indian Ocean region at 31.4 km, which is due to the limited availability of TC cloud maps in this region, resulting in insufficient model training in this area. However, the MSFF-CAA-ResNet-F model achieves an MDE of 23.9 km globally, demonstrating its ability to adapt to remote sensing images from different ocean regions and satellite platforms, showcasing exceptional robustness and generalization capabilities.

5 CONCLUSION

TC are natural disasters that pose a serious threat to human life and property. This paper addresses the most important issue in TC monitoring, namely center location, and develops a new deep learning

Table 4 Comparison of TC center location performance between MSFF-CAA-ResNet-F model and other methods

Literature	Model	Region	Input data type	Dataset	Mean distance error (km)
Yang et al., 2019	ClaNet + LocNet	Northwest Pacific Ocean	Infrared image	\	28.6
Wang et al., 2019	CNN-Location	Northwestern Pacific Ocean	Infrared image	2015–2018	54.0
Hu and Zou, 2021	Azimuth Spectrum Analysis	Western Pacific Ocean	Microwave data	2019	26.2
Liu and Zhang, 2022	Cloud Motion Wind Model	Pacific Ocean	Infrared image	\	41.0
Wang and Li, 2023a	Transfer Learning Algorithm	Northwest Pacific Ocean	Infrared image	2015–2018	29.3
Zhou et al., 2024b	Improved-TCLNet	Northwest Pacific Ocean	Infrared image	2017–2021	29.4
ResNet50	ResNet50	Global	Infrared image	2010–2016	27.4
ConvNeXt	ConvNeXt				27.1
Ours	MSFF-CAA-ResNet-F	Northeastern Pacific Ocean	Infrared image	2010–2016	21.9
		Southern Hemisphere			27.3
		North Atlantic Ocean			22.0
		Northwestern Pacific Ocean			26.6
		Indian Ocean			31.4
		Global			23.9

Models in bold denote those with the optimal performance, and bold numerical values indicate the localization errors achieved by the best-performing model in different ocean basins.

framework called MSFF-CAA-ResNet-F. The MSFF and CAA modules in the model work together to improve the accuracy of TC center location.

The MSFF module uses multi-scale parallel convolution kernels to extract TC features from multiple scales, with large convolution kernels responsible for global receptive fields and small convolution kernels responsible for local receptive fields, achieving efficient feature extraction. The CAA module uses learnable spatial offset parameters to dynamically learn the TC center. The TC center location error of the MSFF-CAA-ResNet-F model is 23.9 km, representing improvements of 12.8% and 11.8% compared to the state-of-the-art ResNet50 and ConvNeXt models, respectively. Visualization results of the model's feature maps and heatmaps indicate that the high-attention regions of the MSFF-CAA-ResNet-F model are more concentrated around the TC center, effectively suppressing the influence of irrelevant cloud layers on TC center location. MSFF-CAA-ResNet-F accurately captures the features of the TC center across all TC intensity levels, while the baseline model's attention is blurred and dispersed.

This paper proposes the MSFF-CAA-ResNet-F model framework design concept. The MSFF and CAA modules can be used not only for TC center location tasks but also extended to other natural disaster monitoring. For example, in tasks such as the location and tracking of extreme events like TCs, sandstorms, and thunderstorms, the MSFF and CAA modules of the MSFF-CAA-ResNet framework also demonstrate theoretically strong adaptability.

Although the MSFF-CAA-ResNet-F model achieves high TC center location accuracy, the model is mainly trained on medium- and low-resolution infrared images, and it has some generalization bias when dealing with high-resolution remote sensing images. Future developments will focus on combining other high-resolution sensors, such as microwave and SAR, for collaborative monitoring of TC information.

6 DATA AVAILABILITY STATEMENT

The datasets generated during or analyzed during the current study are available from the corresponding author on reasonable request.

7 ACKNOWLEDGMENT

This work was supported by the Oceanographic Data Center, Chinese Academy of Sciences.

References

- Chavas D R, Knaff J A. 2022. A simple model for predicting the tropical cyclone radius of maximum wind from outer size. *Weather and Forecasting*, **37**(5): 563-579, <https://doi.org/10.1175/WAF-D-21-0103.1>.
- Dumitru C O, Andrei V, Schwarz G et al. 2019. Machine learning for sea ice monitoring from satellites. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, **XLII-2/W16**: 83-89, <https://doi.org/10.5194/isprs-archives-XLII-2-W16-83-2019>.
- Fett R W, Brand S. 1975. Tropical cyclone movement forecasts based on observations from satellites. *Journal of Applied Meteorology*, **14**(4): 452-465, [https://doi.org/10.1175/1520-0450\(1975\)014%3C0452:TCMFBO%3E2.0.CO;2](https://doi.org/10.1175/1520-0450(1975)014%3C0452:TCMFBO%3E2.0.CO;2).
- Giffard-Roisin S, Yang M, Charpiat G et al. 2020. Tropical cyclone track forecasting using fused deep learning from aligned reanalysis data. *Frontiers in Big Data*, **3**: 1, <https://doi.org/10.3389/fdata.2020.00001>.
- Horn M, Walsh K, Zhao M et al. 2014. Tracking scheme dependence of simulated tropical cyclone response to idealized climate simulations. *Journal of Climate*, **27**(24): 9197-9213, <https://doi.org/10.1175/JCLI-D-14-00200.1>.
- Hu Y Y, Zou X L. 2021. Tropical cyclone center positioning using single channel microwave satellite observations of brightness temperature. *Remote Sensing*, **13**(13): 2466, <https://doi.org/10.3390/rs13132466>.
- Jin S H, Li X F, Yang X F et al. 2019. Identification of tropical cyclone centers in SAR imagery based on template matching and particle swarm optimization algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, **57**(1): 598-608, <https://doi.org/10.1109/TGRS.2018.2863259>.
- Lee J, Moon I J, Im J et al. 2025. Multi-task feature transfer deep learning-based tropical cyclone center estimation (MFT-TC) using geostationary satellite observations. *Weather and Climate Extremes*, **49**: 100796, <https://doi.org/10.1016/j.wace.2025.100796>.
- Lee R S T, Lin J N K. 2001. An elastic contour matching model for tropical cyclone pattern recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, **31**(3): 413-417, <https://doi.org/10.1109/3477.931532>.
- Li X F, Zhou Y, Wang F. 2022. Advanced information mining from ocean remote sensing imagery with deep learning. *Journal of Remote Sensing*, **2022**: 9849645, <https://doi.org/10.34133/2022/9849645>.
- Liu J, Zhang Q. 2022. Objective detection of a tropical cyclone's center using satellite image sequences in the

- northwest Pacific. *Atmosphere*, **13**(3): 381, <https://doi.org/10.3390/atmos13030381>.
- Olander T L, Velden C S. 2019. The advanced Dvorak technique (ADT) for estimating tropical cyclone intensity: update and new capabilities. *Weather and Forecasting*, **34**(4): 905-922, <https://doi.org/10.1175/WAF-D-19-0007.1>.
- Pradhan R, Aygun R S, Maskey M et al. 2018. Tropical cyclone intensity estimation using a deep convolutional neural network. *IEEE Transactions on Image Processing*, **27**(2): 692-702, <https://doi.org/10.1109/TIP.2017.2766358>.
- Szegedy C, Liu W, Jia Y Q et al. 2015. Going deeper with convolutions. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Boston, USA. p.1-9, <https://doi.org/10.1109/CVPR.2015.7298594>.
- Tan C. 2021. TCLNet: learning to locate typhoon center using deep neural network. In: Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, Brussels, Belgium. p.4600-4603, <https://doi.org/10.1109/IGARSS47720.2021.9554524>.
- Velden C, Harper B, Wells F et al. 2006. The Dvorak tropical cyclone intensity estimation technique: a satellite-based method that has endured for over 30 years. *Bulletin of the American Meteorological Society*, **87**(9): 1195-1210, <https://doi.org/10.1175/BAMS-87-9-1195>.
- Wang C, Li X F, Zheng G. 2024a. Tropical cyclone intensity forecasting using model knowledge guided deep learning model. *Environmental Research Letters*, **19**(2): 024006, <https://doi.org/10.1088/1748-9326/ad1bde>.
- Wang C, Li X F. 2023a. Developing a data-driven transfer learning model to locate tropical cyclone centers on satellite infrared imagery. *Journal of Atmospheric and Oceanic Technology*, **40**(12): 1605-1618, <https://doi.org/10.1175/JTECH-D-23-0026.1>.
- Wang C, Li X F. 2023b. Deep learning in extracting tropical cyclone intensity and wind radius information from satellite infrared images—a review. *Atmospheric and Oceanic Science Letters*, **16**(4): 100373, <https://doi.org/10.1016/j.aosl.2023.100373>.
- Wang C, Li X F. 2023c. A deep learning model for estimating tropical cyclone wind radius from geostationary satellite infrared imagery. *Monthly Weather Review*, **151**(2): 403-417, <https://doi.org/10.1175/MWR-D-22-0166.1>.
- Wang C, Xu Q, Li X F et al. 2019. An objective technique for typhoon monitoring with satellite infrared imagery. In: Proceedings of the 2019 Photonics & Electromagnetics Research Symposium-Fall (PIERS-Fall). IEEE, Xiamen, China. p.3218-3221, <https://doi.org/10.1109/PIERS-Fall48861.2019.9021497>.
- Wang C, Yang N, Li X F. 2025. Advancing forecasting capabilities: a contrastive learning model for forecasting tropical cyclone rapid intensification. *Proceedings of the National Academy of Sciences of the United States of America*, **122**(4): e2415501122, <https://doi.org/10.1073/pnas.2415501122>.
- Wang C, Zheng G, Li X F et al. 2021a. Tropical cyclone intensity estimation from geostationary satellite imagery using deep convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, **60**: 4101416, <https://doi.org/10.1109/TGRS.2021.3066299>.
- Wang H Y, Li X F. 2024a. Expanding horizons: U-net enhancements for semantic segmentation, forecasting, and super-resolution in ocean remote sensing. *Journal of Remote Sensing*, **4**: 0196, <https://doi.org/10.34133/remotesensing.0196>.
- Wang H Y, Li X F. 2024b. DeepBlue: advanced convolutional neural network applications for ocean remote sensing. *IEEE Geoscience and Remote Sensing Magazine*, **12**(1): 138-161, <https://doi.org/10.1109/MGRS.2023.3343623>.
- Wang H, Xu Q, Yin X B et al. 2024b. Determination of low-intensity tropical cyclone centers in geostationary satellite images using a physics-enhanced deep-learning model. *IEEE Transactions on Geoscience and Remote Sensing*, **62**: 4202810, <https://doi.org/10.1109/TGRS.2024.3363842>.
- Wang L, Wan B C, Zhou S H et al. 2023. Forecasting tropical cyclone tracks in the northwestern Pacific based on a deep-learning model. *Geoscientific Model Development*, **16**(8): 2167-2179, <https://doi.org/10.5194/gmd-16-2167-2023>.
- Wang P P, Wang P, Wang C et al. 2020. A center location algorithm for tropical cyclone in satellite infrared images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **13**: 2161-2172, <https://doi.org/10.1109/JSTARS.2020.2995158>.
- Wang P, Yang X, Ji Z et al. 2021b. A typhoon center location method on satellite images based on deep reinforcement learning. In: Proceedings of the 40th Chinese Control Conference (CCC). IEEE, Shanghai, China. p.7046-7053, <https://doi.org/10.23919/CCC52363.2021.9550197>.
- Wu Y X, He K M. 2018. Group Normalization. arXiv e-prints, 2018: arXiv: 1803.08494. <https://doi.org/10.48550/arXiv.1803.08494>.
- Xiao F J, Xiao Z N. 2010. Characteristics of tropical cyclones in China and their impacts analysis. *Natural Hazards*, **54**(3): 827-837, <https://doi.org/10.1007/s11069-010-9508-7>.
- Xu J, Li Z, Du B et al. 2020. Reluplex made more practical: Leaky ReLU. In 2020 IEEE Symposium on Computers and communications (ISCC). IEEE, Rennes, France, p.1-7, <https://doi.org/10.1109/ISCC50000.2020.9219587>.
- Yang X, Zhan Z Q, Shen J P et al. 2019. A deep learning based method for typhoon recognition and typhoon center location. In: Proceedings of the 2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, Yokohama, Japan. p.3218-3221, <https://doi.org/10.1109/IGARSS.2019.8899322>.

- Yurchak B S. 2007. Description of cloud-rain bands in a tropical cyclone by a hyperbolic-logarithmic spiral. *Russian Meteorology and Hydrology*, **32**(1): 8-18, <https://doi.org/10.3103/S1068373907010025>.
- Zhang G S, Perrie W, Zhang B et al. 2020. Monitoring of tropical cyclone structures in ten years of RADARSAT-2 SAR images. *Remote Sensing of Environment*, **236**: 111449, <https://doi.org/10.1016/j.rse.2019.111449>.
- Zhang Q, Wu L G, Liu Q F. 2009. Tropical cyclone damages in China 1983-2006. *Bulletin of the American Meteorological Society*, **90**(4): 489-496, <https://doi.org/10.1175/2008BAMS2631.1>.
- Zhang X D, Li X F. 2022. Satellite data-driven and knowledge-informed machine learning model for estimating global internal solitary wave speed. *Remote Sensing of Environment*, **283**: 113328, <https://doi.org/10.1016/j.rse.2022.113328>.
- Zhang X D, Li X F. 2024. Constructing a 22-year internal wave dataset for the northern South China Sea: spatiotemporal analysis using MODIS imagery and deep learning. *Earth System Science Data*, **16**(11): 5131-5144, <https://doi.org/10.5194/essd-16-5131-2024>.
- Zheng Y G, Zhang H S, Qi K T et al. 2022. Stripe segmentation of oceanic internal waves in SAR images based on SegNet. *Geocarto International*, **37**(25): 8567-8578, <https://doi.org/10.1080/10106049.2021.2002430>.
- Zhou F F, Toth Z. 2020. On the prospects for improved tropical cyclone track forecasts. *Bulletin of the American Meteorological Society*, **101**(12): E2058-E2077, <https://doi.org/10.1175/BAMS-D-19-0166.1>.
- Zhou P, Xie X Y, Lin Z C et al. 2024a. Towards understanding convergence and generalization of AdamW. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **46**(9): 6486-6493, <https://doi.org/10.1109/TPAMI.2024.3382294>.
- Zhou Y X, Min M, Li J et al. 2024b. Enhanced typhoon center localization using geostationary satellite imagery. *Geophysical Research Letters*, **51**(22): e2024GL112252, <https://doi.org/10.1029/2024GL112252>.