A Lightweight Multitask Learning Model With Adaptive Loss Balance for Tropical Cyclone Intensity and Size Estimation

Wei Tian [ID], Xinxin Zhou [ID], Xianhua Niu, Linhong Lai, Yonghong Zhang, and Kenny Thiam Choy Lim Kam Sian [ID]

Abstract—**Accurate tropical cyclone (TC) intensity and size estimation are key in disaster management and prevention. While great breakthroughs have been made in TC intensity estimation research, there is currently a lack of research on TC size reflecting TC influence radius. Therefore, we propose a lightweight multi-task learning model (TC-MTLNet) with adaptive loss balance to simultaneously estimate TC intensity and size. Adaptive loss balance is utilized to solve the problem of inconsistent convergence speed of TC intensity and size estimation tasks. The model based on four 2-D convolutions, four 3-D convolutions and three fully connected layers takes up less computational and storage space and improves the accuracy of TC intensity and size estimation by sharing knowledge among multiple tasks. In addition, due to the imbalanced distribution of TC samples, with significantly few low-intensity and high-intensity TC satellite data, this phenomenon poses a great challenge to TC intensity and size estimation. So, we utilize the influence of nearby samples to calibrate the sample density to weight the loss function to enable the model to be generalized to all samples. The result shows that the root-mean-square error (RMSE) of TC intensity estimation is 8.40 kts, which is 33.5% lower than that of the Advanced Dvorak Technique (ADT) and 11.4% lower than that of the deep learning method (3DAttentionTCNet). The mean absolute error (MAE) of the TC size estimation is 20.89 nmi, which is a 16% reduction compared to the Multi-Platform Tropical Cyclone Surface Winds Analysis (MTCSWA).**

Index Terms—**Balanced data distribution, dual attention, lightweight multitask learning model, tropical cyclone (TC) intensity, TC size.**

I. INTRODUCTION

TROPICAL cyclones (TCs) are phenomena with extremely low central pressure, forming over the tropical ocean and eventually dissipating at sea or after moving over land. Landfalling TCs are accompanied by severe weather such as strong wind, rainstorms, and storm surges, which can cause significant loss of life and property [1], [2]. TC intensity is defined as the maximum average wind speed near the TC center. It is an important parameter that measures the destructive power of TC and is used in TC warning, prevention, and management. Accurate TC intensity estimation also helps to predict the rapid intensification of TC intensity. TC size indicates the radius of TC influence. TC size is usually measured by several wind radii provided by the forecast centers, including the gale-radius (35 kts, R35), storm-radius (50 kts, R50), hurricane-radius (64 kts, R64) and the radius of maximum wind (RMW) [3], [4], [5], [6], [7]. R35 represents the potential impact area of a TC and is one of the most widely used parameters to predict TC influence and mitigate TC impact. Thus, the present study conducts TC size estimations based on R35.

Obtaining TC observations is difficult because TCs spend most of their lifetime over the ocean, where deploying observation equipment is challenging. Therefore, aircraft and ships are used to obtain TC observations at sea. However, these observational methods are very expensive. With the development of artificial intelligence, satellite imagery has become the main source of information for TC intensity estimation. Although satellite imagery cannot directly measure TC intensity, it can be estimated indirectly through the captured cloud structure [8], [9]. For example, infrared satellite imagery provides the temperature distribution of radiation surfaces, water vapor satellite imagery provides information on the water vapor in the clouds, and microwave satellite imagery provides information such as the TC eye, eyewall and spiral rainbands. TC intensity estimation using satellite imagery is based on the fact that TCs of similar intensities have similar cloud structures. In meteorology, the main TC intensity estimation methods include the Dvorak technique [10], the advanced Dvorak technique (ADT) [11], the deviation angle variance technique (DAV) [12], and the satellite consensus technique (SATCON) [13], [14], [15]. These methods rely on artificial experience or various algorithms to obtain features related to TC intensity and then use regression models to obtain TC intensity. However, the cloud structure features related to TC intensity determined by humans are subjective. In addition, the design of feature extraction algorithms requires expertise, which greatly limits TC intensity estimation. With the development of deep learning technology, intensity estimation

Wei Tian, Xinxin Zhou, and Linhong Lai are with the School of Computer and Software, Engineering Research Center of Digital Forensics, Ministry of Education, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: tw@nuist.edu.cn; 17746377535@163.com; 13602537214@163.com).

Xianhua Niu is with the State Key Laboratory of Geo-Information Engineering, Xi'an 71004, China (e-mail: nxh0322@126.com).

Yonghong Zhang is with the School of Automation, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: zyh@nuist.edu.cn).

Kenny Thiam Choy Lim Kam Sian is with the Wuxi University, Wuxi 214063, China (e-mail: kennlks@gmail.com).

Digital Object Identifier 10.1109/JSTARS.2022.3225154

technology has gradually improved. However, there are still challenges in TC intensity estimation based on convolutional neural network (CNN) [16], [17], [18], [19], [20], [21], [22]. For example, the dependence of spatial information and the correlation of channel information fail to take into account. Due to the lack of low-intensity and high-intensity TC observation data, low-intensity TCs are overestimated and high-intensity TCs are underestimated. At present, TC intensity estimation methods focus more on improving accuracy and ignoring the study of model interpretability.

Obtaining *in-situ* measurements during TC events is difficult. Therefore, other observation methods are used for TC size estimation, such as satellite [23], [24], [25], [26], [27], scatterometer [28] and microwave sounder [29]. There is no technology as widely used as the Dvorak method for TC size estimation because the physical environment information involved in TC size estimation is very complex. Recently, more and more researchers have started applying deep learning to estimate TC size. However, obtaining accurate estimations is still a challenge due to the complexity of the physical relationship between convection and wind field.

Currently, there is a lack of studies that simultaneously estimate TC intensity and size. We design a lightweight multitask learning model (TC-MTLNet) with parallel dual attention to estimate TC intensity and size intensity simultaneously. We employ 2-D convolution and 3-D convolution, respectively, to extract features from the infrared channel and multichannel satellite imagery. Before the fusion of spatial features and channel features, the position attention module (PAM) and channel attention module (CAM) are applied to the spatial dimension and channel dimension, respectively. This dual attention considers the dependency between the feature map pixels and the correlation between the adjacent channels and allows local features and global dependencies to be effectively integrated to improve the accuracy of TC intensity and size estimation. Then, we incorporate environmental factors into the fully connected layer after the spatial and channel features are fused and flattened into one dimension, which effectively provides features related to TC intensity and size. Subsequently, we use two branches to learn TC intensity and size separately, while using minimum sea-level pressure (MSLP) as an auxiliary task to provide additional feature information for TC intensity and size estimation. Our main contributions are as follows.

1) Before the two branches are combined, we apply dual attention to improve feature representation, thus improving the model's performance. The PAM represents the interdependence between spatial features on the feature maps, and the CAM focuses on the correlation of channels.

2) Adaptive loss balance method is utilized to solve inconsistent TC intensity and size convergence speed in multi-task learning. The method automatically balances the training speed of multiple tasks by dynamically adjusting the gradients. This method enables multiple tasks to learn at a similar speed to avoid multi-task learning being dominated by one task.

3) Concerning the overestimation and underestimation of TC intensity and size due to unevenly distributed TC samples, we employ the label distribution smoothing method, which

convolves the symmetric kernel with the empirical density distribution to acquire the effective label density distribution that reflects the real sample unevenness. Finally, the loss function is designed based on the effective label density to alleviate the problem of overestimation and underestimation.

4) We explore model interpretability by deep learning visualization techniques. First of all, by visualizing the feature maps, we can understand the features learned by the convolution kernels. Then, we use the deconvolution technique to visualize the feature maps generated by the filters to understand the role of the filter. Finally, by visualizing the high-dimensional features through Grad-CAM++ technology, we can understand the contribution of each part in the satellite imagery to the intensity and size estimation.

The rest of this paper is organized as follows. The next section presents the research status of TC intensity and size estimation based on traditional and deep learning methods. Section III introduces the data, data preprocessing and the architectural details of the model used. The experimental results are discussed in Sections IV and V summarizes the research.

II. RELATED WORK

A. TC Intensity Estimation Based on Traditional Meteorological Methods

The Dvorak technique is based on infrared satellite imagery to estimate TC intensity. This approach correlates TC intensity with its inner core, vertical motion patterns and outer cloud structures, assuming that TCs with similar cloud structures have the same intensity. However, the Dvorak technique is highly subjective as meteorologists estimate TC intensity by manually analyzing the TC cloud structure from satellite imagery. The ADT has been developed to overcome the limitations of the Dvorak technique. This method eliminates subjectivity in intensity estimation by using the objective storm center determination scheme and cloud pattern determination logic. The DAV technique takes the organizational level of infrared TC cloud features as an indirect measurement of TC intensity. This method applies the DAV to quantify the degree of symmetry of cloud clusters and considers that similar intensity TCs have a similar degree of cloud clusters symmetry. However, this method requires a prior TC center, and errors in the location may lead to inaccurate TC intensity estimation. The Advanced Microwave Sounding Unit (AMSU) is based on the relationship between the brightness temperature of each channel with TC intensity. The SATCON technique is a weighted estimation scheme that sums up the ADT and several TC intensity estimation methods to minimize the weaknesses of each method.

B. TC Intensity Estimation Based on Deep Learning Methods

With the development of artificial intelligence, TC intensity estimation based on deep learning has been increasingly applied. Pradhan et al. [30] were the first to apply deep learning to estimate TC intensity. However, they manually deleted the satellite imagery of poor quality, so the designed network

lacked applicability. Chen et al. [31] designed a novel CNN model, which proposed and verified the rotational invariance of TC. Chen et al. [32] proposed TC intensity estimation model based on multi-channel satellite imagery, and integrated basin information and location information into the network. Lee et al. [33] designed a multi-dimensional CNN to estimate TC intensity from geostationary satellite data and analyzed extracted intensity features from multispectral infrared imagery through heat maps. Zhuo et al. [34] proposed a network for TC intensity and size estimation based on physical enhancement. With the augmentation of auxiliary physical information and tasks, the network achieved good performance. Tian et al. [35] designed a 3D CNN to extract TC features from multi-channel satellite imagery, which improved the model's ability to focus on features related to TC intensity by CBAM. Sun et al. [36], [37] proposed novel image retrieval methods, which can mine similar TCs to further improve the accuracy of TC intensity estimation. Duan et al. [38], [39] proposed a self-supervised learning method to learn the deep neural network from unlabelled hyperspectral data and a novel hyperspectral image classification framework using the fusion of dual spatial information. These methods he proposed can enhance the extraction of TC intensity and size features, which can further improve the accuracy of TC intensity and size estimation. Tan et al. [40] designed a model incorporating a residual learning mechanism and an attention mechanism. However, these methods for TC intensity estimation still suffer shortcomings. For example, the dependence of spatial feature information and the correlation of channel feature information is not considered. Due to the lack of observation data of low- and high-intensity TC, low-intensity TC is overestimated and high-intensity TC is underestimated. In addition, intensity estimation methods focus more on improving accuracy but ignore the study of interpretability of models.

C. TC Size Estimation Based on Traditional Meteorological Methods

Demuth et al. [41] employed parameters derived from the AMSU data to measure TC intensity (maximum sustained winds and MSLP) and size (34 kts, 50 kts and 64 kts wind radius). However, the resolution of these data is too low to adequately reflect the TC structure. In order to overcome the limitations of this technique, many methods (e.g. [6], [42]) based on satellite imagery are designed to estimate TC size. Knaff et al. [3] developed an automated, objective, MultipleSatellite-Platform Tropical Cyclone Surface Wind Analysis (MTCSWA) [3] method for TC size estimation that allows variable data weights to be applied to the input data. The combination of overall quality control and weighted variational analysis yielded smaller errors in TC size estimation.

D. TC Size Estimation Based on Deep Learning Methods

At present, deep learning has entered many research fields. Motivated by the positive progress of deep learning in TC intensity estimation [30], [31], [32], [33], [34], [35], [40], recent studies have begun to explore the application of deep learning to TC size estimation. Zhuo et al. [34] were the first to estimate TC size using deep learning. They designed a physics-augmented

TABLE I
DETAILS OF THE IR, WV, PMW CHANNEL SATELLITE IMAGERY

Channel	Temporal resolution	Spatial resolution	Date source	Time range
IR	3h	0.07	GridSat	2003–2017
WV	3h	0.07	GridSat	2003–2017
PMW	3h	0.25	CMORPH	2003–2017

multi-task learning model to estimate TC size and found that learning multiple wind radius tasks and auxiliary intensity estimation tasks simultaneously yielded more accurate TC size estimation. Baek et al. [43] developed a novel multi-task learning model (tc-sem) in the western north pacific for TC size estimation, which improved the accuracy of TC size estimation through knowledge sharing among multiple related tasks.

III. DATA AND METHODOLOGY

A. Data and Data Preprocessing

1) Data: The TCIR [31] dataset from 2003 to 2017 is used in our study. This dataset provides infrared (IR), water vapor (WV), visible (VIS) and passive microwave rain rate (PMW) channel satellite imagery. Since the visible channel is very noisy at night, this study is based on IR, WV and PMW channels. The details of the satellite imagery are shown in Table I. The spatial resolution of the IR and WV channels is $0.07° \times 0.07°$, and the spatial resolution of the PMW channel is $0.25° \times 0.25°$. Therefore, the PMW channel is enlarged by about 4 times using linear interpolation to have uniform channel sizes. All imageries are 201×201 pixels, and the distance between the two data points is 4 km. In addition, the TCIR dataset provides TC center location, intensity size and minimum sea-level pressure (MSLP). The data from 2003 to 2016 is used for training, testing, and validation, and the data from 2017 is used to evaluate the performance of the best-optimized model. Due to the rotation characteristic of TC, we rotate satellite imagery 10 times ($0°$, $36°$, $72°$, $108°$, etc.) during testing and validation, and then take the average estimated value of 10 satellite imageries as the TC estimated value to achieve the more stable performance of the model.

2) Data Preprocessing: In order to improve generalization, we design three data enhancement methods.

1) To retain the TC wind eye and the TC cloud structure and omit irrelevant features in the outer region, we cut the satellite imagery in the center area from 201×201 to 141×141, which has been proved to be the most effective area for TC intensity and size estimation.

2) We apply the *z*-score standardized method to TC data. The data are uniformly converted to the same magnitude by subtracting the mean and dividing by the standard deviation to improve the comparability of the data.

3) Given the rotation invariance of TC, the satellite imagery is randomly rotated, which is helpful to improve the generalization ability of our model.

B. Methodology

In this section, we first give an overview of the model and then give a detailed introduction to each module.
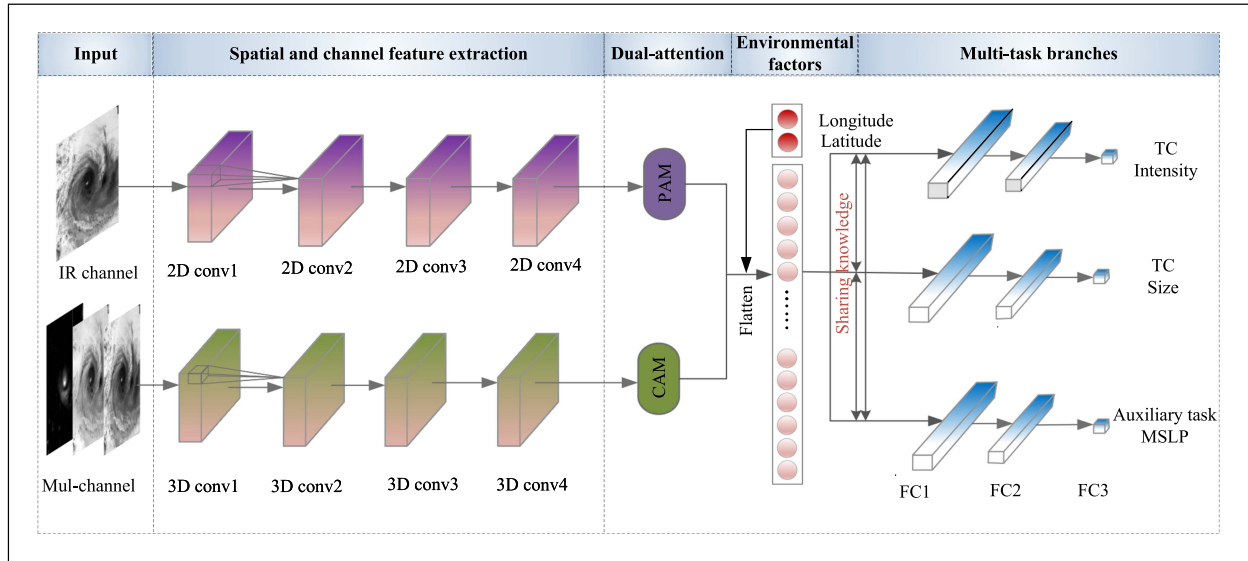
Fig. 1. Network architecture of TC-MTLNet for TC intensity and size estimation. The TC-MTLNet contains 5 modules: spatial feature extraction module, channel feature extraction module, PAM, CAM, and multi-task branch module.

We design a parallel dual-attention TC intensity and size estimation network (TC-MTLNet) based on multitask learning. The overall network architecture is shown in Fig. 1. The IR channel and IR-WV-PMW three-channel combination are entered into the spatial and channel feature extraction module, respectively. The spatial feature extraction module uses four 2-D convolutions, while the channel feature extraction module uses four 3-D convolutions. The features obtained from the spatial feature extraction module are then fed into the PAM, and the features from the CAM are fed into the CAM. After the extracted spatial and channel features are fused and flattened, environmental factors are introduced into the fully connected layer. Subsequently, we employ two branches to study features related to TC intensity and size, and an auxiliary task branch to provide additional useful information for TC intensity and size estimation.

1) Spatial Feature Extraction: The spatial feature extraction module uses four 2-D convolutions. The size of the input IR channel imagery is 141×141, the convolution kernel of the first convolution block (2conv1) is 4×4, and the convolution kernels of 2conv2, 2conv3, and 2conv4 are 3×3. The step is 2 and the fill is 0. Rectified linear unit (ReLU) activation functions are introduced into the model after each convolution. The detailed network structure is shown in Fig. 1.

2) Channel Feature Extraction: Channel feature extraction: 2-D convolution slides only spatially, and its output is a cube made up of many faces, where the channel information completely overlaps. Unlike the 2-D convolution, 3-D convolution slides both in the spatial and channel dimensions. It outputs a large cube stacked with many cubes and retains channel information well. Therefore, for the channel feature extraction module, we employ four 3-D convolutions with the following convolution kernels sizes: $1 \times 4 \times 4$ for 3conv1, $1 \times 3 \times 3$ for 3conv2, $2 \times 3 \times 3$ for 3conv3 and 3conv4. The number of convolution kernels are 16, 32, 64, and 128, respectively. Strides
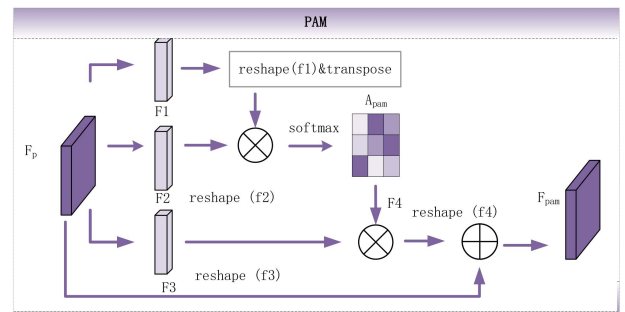

Fig. 2. Detailed structure of the PAM.

are set to $1 \times 2 \times 2$, $1 \times 3 \times 3$, $2 \times 2 \times 2$ and $2 \times 2 \times 2$, respectively, and the padding is 0.

3) Positional Attention Module: In the spatial feature extraction module, we adopt 3×3 and 4×4 convolution kernels, which only consider local receptive fields, while TC intensity and size estimation should take more global features into account. In order to capture the global dependence between different locations in the feature maps, we introduce the location attention module. Next, we elaborate the process of learning global features in the location attention module. The detailed structures are shown in Fig. 2. Extracted spatial features $F_p \in R^{C \times H \times W}$ is first fed into three convolution layers to generate three new feature maps $F1 \in R^{C1 \times H \times W}$, $F2 \in R^{C1 \times H \times W}$ and $F3 \in R^{C \times H \times W}$. Next, F1 and F2 then are reshaped to $f1 \in R^{C1 \times N}$ and $f2 \in R^{C1 \times N}$, where N is equal to $H \times W$. Then, we feed the result of multiplying the transpose of f1 by f2 into the softmax function. The softmax function normalizes the result to obtain the location attention map $A_{pam} \in R^{(H \times W) \times (H \times W)}$. Therefore, the a_{ij} on the location attention map A_{pam} represents the relationship between the ith and jth location. F3 is reshaped to f3 and $f3 \in R^{C \times N}$ is multiplied by the spatial attention map
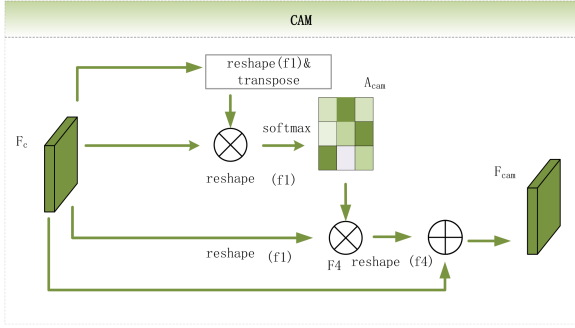
Fig. 3.   Detailed structure of the CAM.

A_{pam} to obtain $F4 \in R^{C \times N}$. The element f_{11} on F4 represents the relationship between position 1 and all positions in channel 1. Finally, F4 is reshaped to $f4 \in R^{C \times H \times W}$. The final output $F_{\text{pam}} \in R^{C \times H \times W}$ is obtained by adding f4 and the original feature F_p.

$$a_{ji} = \frac{\exp(f1_i \cdot f2_j)}{\sum_{i=1}^{N} \exp(f1_i \cdot f2_j)}. \tag{1}$$

4) Channel Attention Module: The IR channel provides information about water vapor content in the upper atmosphere, the WV channel provides information about water vapor composition in the middle atmosphere, and the PMW channel can penetrate clouds to detect rain. It is found that three channels are correlated. For example, previous studies have shown that features extracted from the IR and WV channels are similar, and features extracted from IR and PMW channels are complementary. In order to better extract the correlation between the channels more fully, we directly process the feature maps $F_c \in R^{C \times H \times W}$ extracted by the channel feature extraction module to get the channel attention maps $A_{\text{cam}} \in R^{C \times C}$. As shown in Fig. 3, we first multiply the reshape of F_c with the transpose of the reshape of F_c, and then normalize the results by the softmax function to obtain the channel attention map A_{cam}. Therefore, the element a_{ij} on the channel attention map A_{cam} represents the relationship between the ith and jth channel. We then multiply A_{cam} by the reshape of F_c to get $F4 \in R^{C \times N}$. The element f_{11} on F4 represents the relationship between position 1 on channel 1 and position 1 on the other channels. Finally, F4 is reshaped to $f4 \in R^{C \times H \times W}$, and f4 is added to the original feature map F_c to obtain the final output $F_{\text{cam}} \in R^{C \times H \times W}$.

5) Multitask Learning: In order to simultaneously estimate TC intensity and size, we design a multitask learning model (see Fig. 1). The convolution layer parameters of the model are shared, and the parameters of the full connection layer are learned independently by three tasks. In the fully connected layers, we utilize two branches independently to learn features associated with TC intensity and size from satellite imagery. In addition, a branch of auxiliary tasks helps to learn features related to TC intensity and size by providing additional useful information. The underlying shared parameters of multitask learning are equivalent to data augmentation. For example, when a TC intensity estimation task encounters features that are difficult to learn, the TC size task or auxiliary task can easily

learn these features. The TC intensity estimation task learns these hard-to-learn features because the underlying parameters of multiple tasks are shared.

IV. MODEL TRAINING

Our model is trained on GeForce RTX 2080ti 11 GB GPU based on the pytorch framework. The CPU environment is Inter Core i9-9900 K. During the training, we adopt the adam optimizer with a learning rate of 0.0005. We define the multitask loss function as $L(t)$, which is achieved through a weighted sum of losses for different tasks. The detailed expression is as follows (2). $L_i(t)$ is the loss function of task i, and $w_i(t)$ is the weight of the loss function $L_i(t)$. $w_i(t)$ and $L_i(t)$ are dynamically updated with the number of iteration steps t

$$L(t) = \sum_i w_i(t) x L_i(t). \tag{2}$$

A. Update $w_i(t)$ by Optimizing $L_{grad}(t; w_i(t))$

Since the gradient magnitude and convergence speed of different tasks are different, it is necessary to prevent multitask learning from being dominated by one task, which negatively affecting the effect of other tasks. $L_{\text{grad}}(t; w_i(t))$ is a function of the weight $w_i(t)$, which is shown in (3). We update the weight parameters $w_i(t)$ of multiple task loss functions $L_i(t)$ through $L_{\text{grad}}(t; w_i(t))$

$$L_{\text{grad}}(t; w_i(t)) = \sum_i \left| G_w^i(t) - \overline{G}_w(t) \right| \times [r_i(t)]^\alpha \tag{3}$$

The following formulas are the core of $L_{\text{grad}}(t; W_i(t))$. First, we define a formula to measure the loss magnitude of a task, which is expressed in detail as follows:

$$G_w^i(t) = \| \nabla W w_i(t) \times L_i(t) \|_2 \tag{4}$$

The W is the multitask learning network parameter of the share section's last layer. $G_w^i(t)$ is the L2 norm of the gradient for parameter w based on the result of multiplying the loss $L_i(t)$ of task i by the weight $w_i(t)$ of the loss $L_i(t)$. $G_w^i(t)$ can measure the magnitude of the loss of a task. The larger the $G_w^i(t)$, the larger the magnitude of the loss. $\overline{G}_w(t)$ is the average of $G_w^i(t)$ of all tasks. The formula is as follows:

$$\overline{G}_w(t) = E_{\text{task}} \left[G_w^i(t) \right]. \tag{5}$$

Then, we adopt two formulas to measure the learning speed for multiple tasks, which are expressed in detail as follows:

$$\tilde{L}_i(t) = \frac{L_i(t)}{L_i(0)} \tag{6}$$

$$r_i(t) = \frac{\tilde{L}_i(t)}{E_{\text{task}} \left[\tilde{L}_i(t) \right]}. \tag{7}$$

$L_i(t)$ and $L_i(0)$ are the losses of steps t+1 and steps 1 of task i. $\tilde{L}_i(t)$ reflects the reverse training speed of task i. The larger the $\tilde{L}_i(t)$, the slower the loss decreases. $E_{\text{task}} [\tilde{L}_i(t)]$ is the average of the $\tilde{L}_i(t)$ of all tasks. $r_i(t)$ represents the relative
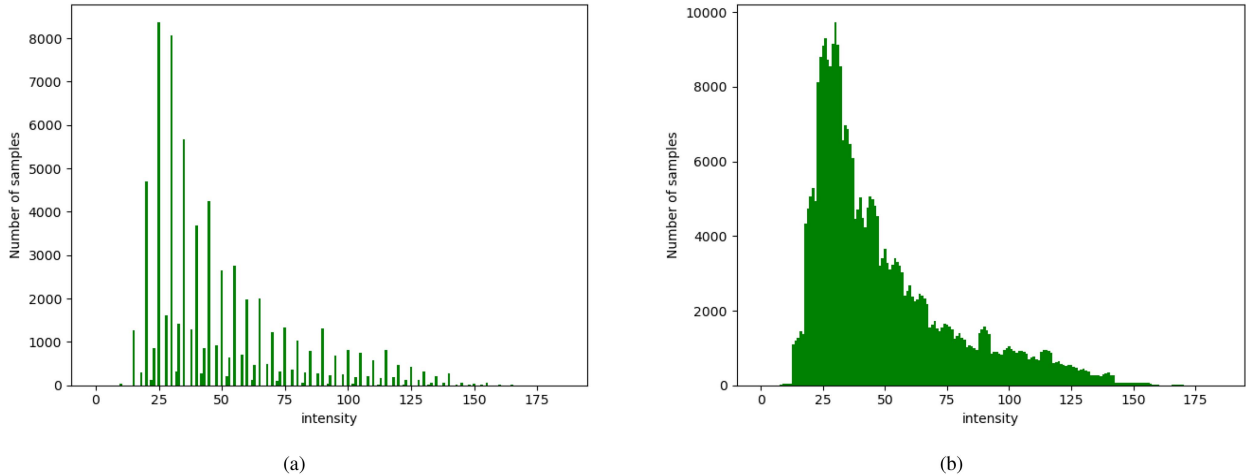
Fig. 4. (a) Empirical label density distribution. (b) Effective label density distribution considering label continuity is obtained by convoluting the symmetric kernel with the empirical label density distribution.

reverse training speed of task i. The larger $r_i(t)$, the slower task i is trained in all tasks.

It can be seen from $L_{grad}(t; W_i(t))$ that when the loss of a task is too large or too small, $|G_w^i(t) - \overline{G}_W(t)|$ becomes so large that $L_{grad}(t; W_i(t))$ increases. When the training speed of a task is too slow, $r_i(t)$ becomes larger, so $L_{grad}(t; W_i(t))$ also increases. The process of optimizing $L_{grad}(t; W_i(t))$ is to force the model to select an appropriate loss function weight $w_i(t)$ so that the gradient magnitude and convergence speed remain roughly the same for multiple tasks.

B. Design $L_i(t)$ Based on Label Distribution Smoothing

The TC observation data displays a long-tail distribution with significantly less low-intensity and high-intensity TC data. Most current methods are based on empirical label density to solve overestimation and underestimation (Density is the number of samples.). However, for continuous labels, the empirical label density fails to reflect the real imbalance seen by the network. For example, labels t1 and t2 both have a small amount of data [see Fig. 4(a)]. t1 is in the neighborhood of the high-density samples (i.e., there are many samples in the range $[t1 - \Delta, t1 + \Delta]$, while t2 is in the neighborhood of the low-density samples (i.e., there are few samples in the range $[t2 - \Delta, t2 + \Delta]$). In this case, t1 does not have the same degree of unevenness as t2 for continuous labels because of the dependence between data samples on nearby labels. In order to alleviate the problem of overestimation and underestimation due to the unevenly distributed data, we adopt a label distribution smoothing approach. The idea of label distribution smoothing is to convolve the symmetric kernel with the empirical label density to obtain the effective label density distribution reflecting the real imbalance by using the similarity between nearby objects. Fig. 4(b) shows the effective label density distribution obtained by smoothing the label distribution, which effectively reflects the imbalance observed by the neural network. l_k is the density of samples labeled k. E_{den} is the average of the reciprocal of the effective label density of all labels. The detailed expression is shown as

follows:

$$E_{den} = \frac{1}{m} \sum_{k=1}^{m} \frac{1}{l_k}. \qquad (8)$$

We define the weight parameter of each sample loss as p_j, which is achieved through dividing $\frac{1}{l_k}$ by E_{den}

$$p_j = \frac{1}{l_k E_{den}} \qquad (9)$$

We obtain $L_i(t)$ by multiplying the loss $(\hat{y}_j - y_j)^2$ for each sample by the weight parameter p_j to address the problem caused by imbalanced TC samples. (see (9))

$$L_i(t) = \sqrt{\frac{1}{m} \sum_{j=1}^{m} p_j(\hat{y}_j - y_j)^2}. \qquad (10)$$

V. EXPERIMENT AND RESULTS

In this section, we first conduct module ablation experiments to verify the effectiveness of our model. Second, we verify the superiority of multitask learning compared to single-task learning, and the superiority of the automatic weighting method of loss function based on gradient normalization in multitask learning. Third, we verify the effect of label distribution smoothing method on the TC intensity overestimation and underestimation. We then apply visualization techniques to understand the features learning process and the important features for TC intensity and size estimation. Finally, we validate the effectiveness of our optimal model by scatter diagrams and lifecycle sequence diagrams, and compare the TC-MTLNet with traditional and deep learning methods.

A. Validation of Model Structure Design

In order to verify the validity of the model structure, we design 5 model experiments (M1, M2, M3, M4, M5) by gradually increasing the structure of the model. The evaluation and analysis of the experimental results are shown in Table II and Fig. 5. Considering that IR satellite imagery provides the most

TABLE II
NETWORK CONFIGURATION AND EXPERIMENTAL RESULTS OF M1, M2, M3, M4, AND M5

| Model	Input	Module	Configurations	RMSE (kts)
M1	IR channel	2conv1,2con2, 2conv3,2conv4, FC1,FC2,FC3	(16@4×4,s=2,p=0),(32@3×3,s=2,p=0), (64@3×3,s=2,p=0),(128@3×3,s=2,p=0), FC1(256), FC2(128), FC3(1)	9.33
M2	Multi-channel	3conv1,3con2, 3conv3,3conv4, FC1,FC2,FC3	(16@1×4×4,s=(1,2,2),p=0),(32@1×3×3,s=(1,3,3),p=0), (64@2×3×3,s=(2,2,2),p=0),(128@2×3×3,s=(2,2,2),p=0), FC1(256),FC2(128),FC3(1)	8.88
M3	IR channel & Multi-channel	(2conv1,2con2, 2conv3,2conv4)& (3conv1,3con2, 3conv3,3conv4) FC1,FC2,FC3	((16@4×4,s=2,p=0),(32@3×3,s=2,p=0), (64@3×3,s=2,p=0),(128@3×3,s=2,p=0))& ((16@1×4×4,s=(1,2,2),p=0),(32@1×3×3,s=(1,3,3),p=0), (64@2×3×3,s=(2,2,2),p=0),(128@2×3×3,s=(2,2,2),p=0)) FC1(256),FC2(128),FC3(1)	8.76
M4	IR channel & Multi-channel	(2conv1,2con2, 2conv3,2conv4 PAM)& (3conv1,3con2, 3conv3,3conv4 CAM) FC1,FC2,FC3	((16@4×4,s=2,p=0),(32@3×3,s=2,p=0), (64@3×3,s=2,p=0),(128@3×3,s=2,p=0), PAM)& ((16@1×4×4,s=(1,2,2),p=0),(32@1×3×3,s=(1,3,3),p=0), (64@2×3×3,s=(2,2,2),p=0),(128@2×3×3,s=(2,2,2),p=0), CAM) FC1(256),FC2(128),FC3(1)	8.69
M5	IR channel & Multi-channel	(2conv1,2con2, 2conv3,2conv4 PAM)& (3conv1,3con2, 3conv3,3conv4 CAM)) Environmental factors, FC1,FC2,FC3	((16@4×4,s=2,p=0),(32@3×3,s=2,p=0), (64@3×3,s=2,p=0),(128@3×3,s=2,p=0), PAM))& ((16@1×4×4,s=(1,2,2),p=0),(32@1×3×3,s=(1,3,3),p=0), (64@2×3×3,s=(2,2,2),p=0),(128@2×3×3,s=(2,2,2),p=0), CAM)) Environmental factors, FC1(256),FC2(128),FC3(1)	**8.57**
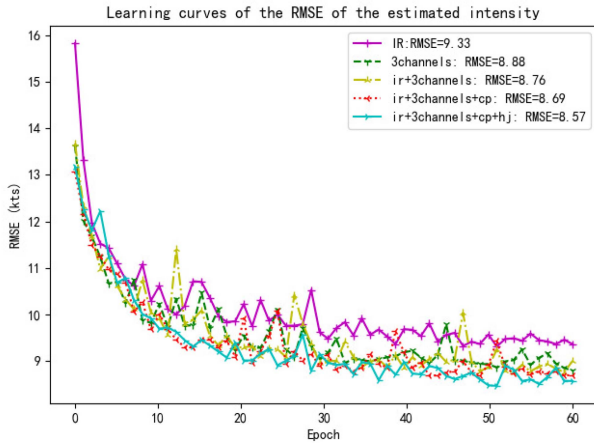
Fig. 5. Learning curve of the RMSE of TC intensity estimation for model M1, M2, M3, M4, and M5.

useful features for TC intensity estimation, we design a 2-D convolution model (M1) based on IR satellite imagery to extract spatial features. The details of the model parameters are shown in Table II. Then, we design a 3-D convolution model (M2) to extract the 3-D features of TC from the IR, WV, and PMW channels. M3 is a fusion of M1 and M2. It designs two branches to extract features from the IR channel and the multichannel satellite imagery using 2-D and 3-D convolutions, respectively. The root-mean-square error (RMSE) of M1, M2, and M3 are 9.33 kts, 8.88 kts, and 8.76 kts, respectively. The experimental results show that the combination of the two branches produces a lower error, which may be because the combination of spatial

and channel features extracted by M3 obtains richer TC features. Next, M4 adds PAM and CAM to the spatial feature extraction module and the channel feature extraction module based on M3. The RMSE generated by this model is 8.69 kts. The experimental results show that the dual attention module pays better attention to the dependence of spatial features and the correlation of channel features. The last model M5 introduces location environment information (longitude, latitude) before the fully connected layer. The intensity estimation error is further reduced by model M5, which proves the validity of environmental information. Therefore, M5 is selected as the backbone of the TC-MTLNet.

B. Single-Task and Multitask Comparative Experiments

Multitask learning can learn useful information from related tasks, which can help improve individual performances. Fig. 6(a) shows that there is a significant positive correlation between TC intensity and size ($r = 0.66$). With the increase of TC intensity, the TC size increases. Fig. 6(b) shows a significant negative correlation between TC intensity and MSLP $r = -0.95$), which means that MSLP decreases with the increase in TC intensity. Fig. 6(c) shows a negative correlation between TC size and MSLP ($r = -0.72$). Therefore, we design a multitask learning model which shares parameters in the convolution layers. We utilize two branches to independently learn TC intensity and size features in the fully connected layers and adopt an auxiliary task (MSLP) branch to learn other useful information for TC intensity and size estimation.

In order to verify the effectiveness of the multitask learning model, we conduct two single-task experiments and four
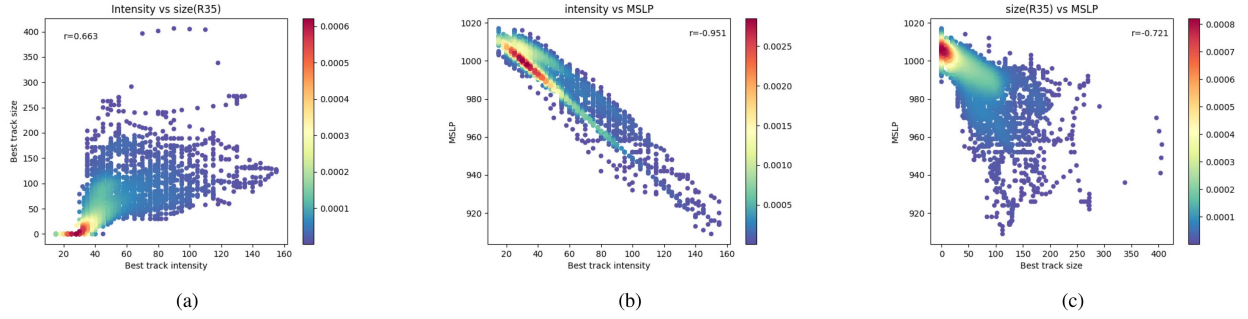
Fig. 6.   (a) Correlation between best-track intensity and size. (b) Correlation between best-track intensity and MSLP. (c) Correlation between best-track size and MSLP. The colors indicate different data densities. The correlation coefficients are shown in the figure.

TABLE III
RMSE, MAE, AND BIAS OF TC INTENSITY AND SIZE ESTIMATION FOR SINGLE-TASK AND MULTITASK LEARNING

Model	Loss function	Intensity			Size(R35)		
		RMSE	MAE	Bias	RMSE	MAE	Bias
Single_intensity	L1	8.57	7.04	−1.01			
Single_size	L2				23.93	21.83	−4.7
Multi-task1	(L1+L2+L3)/3	8.85↑	7.41↑	−1.3	23.94↑	21.58↑	−4.53
Multi-task2	0.75L1+0.2L2+0.05L3	8.50↓	7.00↓	−1.93	23.83↓	21.70↓	0.80
Multi-task3	0.85L1+0.1L2+0.05L3	8.45↓	6.98↓	0.16	23.62↓	21.65↓	−5.27
Multi-task4	Adaptive loss balance function	**8.40↓**	**6.90↓**	−0.44	**23.01↓**	**20.89↓**	−2.3
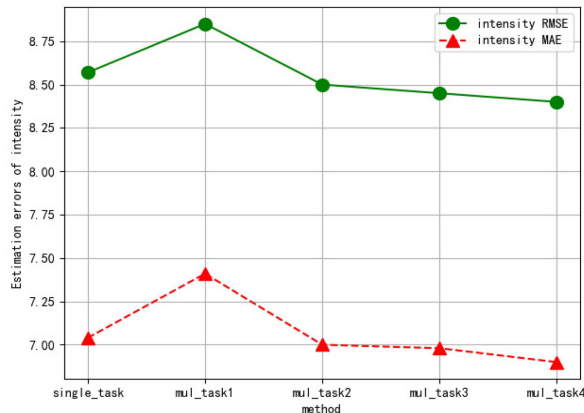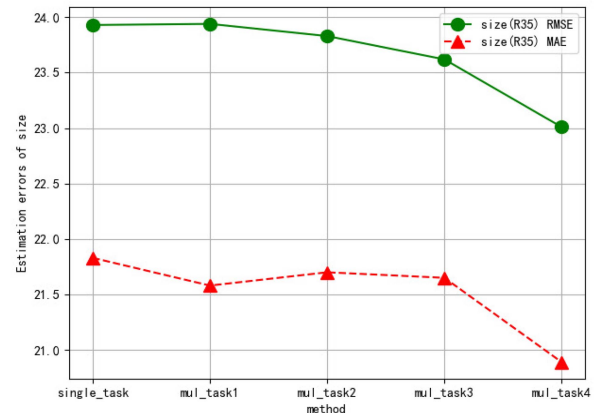
Fig. 7.   (a) The RMSE and MAE of TC intensity estimation in single-task and multi-task learning. (b) The RMSE and MAE of TC size estimation in single-task and multitask learning.

multitask experiments. The experimental results are shown in Table III and Fig. 7. The TC intensity estimation RMSE, mean absolute error (MAE) and bias in the single-task experiment are 8.57 kts, 7.04 kts, and −1.01 kts, respectively. The single-task experiment for TC size estimation has an RMSE of 23.93 nmi, a MAE of 21.83 nmi, and a bias of −4.7 nmi. In multitask learning, the multitask1 directly adds three tasks and averages them. The experimental results are shown in Table III. The intensity estimation RMSE, MAE, and bias are 8.85 kts, 7.41 kts, and −1.3 kts, respectively. The size estimation RMSE, MAE, and bias of the size estimation are 23.94 nmi, 21.58 nmi

and −4.53 nmi, respectively. It is obvious that the loss of TC intensity is higher than that of single-task learning. Because the magnitude of loss for different tasks may be different. The way in which the losses are directly added may lead to the learning of multiple tasks being dominated by one task. In other words, the multitask1 model tends to fit the TC size estimation task, which leads to a negative impact on the performance of the TC intensity estimation task. Therefore, the effect of the TC intensity estimation is worse. Compared with single-task learning, the accuracy of TC intensity and size estimation is improved in multitask2 and multitask3. However, fixed weights do not

TABLE IV
COMPARISON OF RMSE, MAE, AND BIAS OF 8 TC CATEGORIES AFTER LABEL SMOOTH DISTRIBUTION PROCESSING AND UNPROCESSED DATA

| Category	Intensity				Size(R35)			
	Imbalanced data		Balanced data		Imbalanced data		Balanced data	
	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
NC	7.12	7.11	5.3↓	4.6↓	7.2	7.1	6.1↓	4.4↓
TD	3.84	2.16	5.2	1.1	13.8	12.2	13.6	11.3
TS	6.64	−2.78	7.2	−1.7	33.3	−24.1	35.7	−26.7
H1	11.22	−6.30	10.7↓	−6.8	25.0	−15.8	25.7	−18.5
H2	11.98	−6.27	11.9↓	−6.3	23.3	−14.4	23.1↓	−14.0↓
H3	11.32	−3.9	11.6	−4.8	20.9	−11.5	20.6↓	−11.4↓
H4	7.59	−5.36	7.3↓	−5.2↓	34.6	−25.8	33.1↓	−24.0↓
H5	10.34	−10.22	8.0↓	−6.8↓	18.2	−10.4	16.4↓	−9.5↓
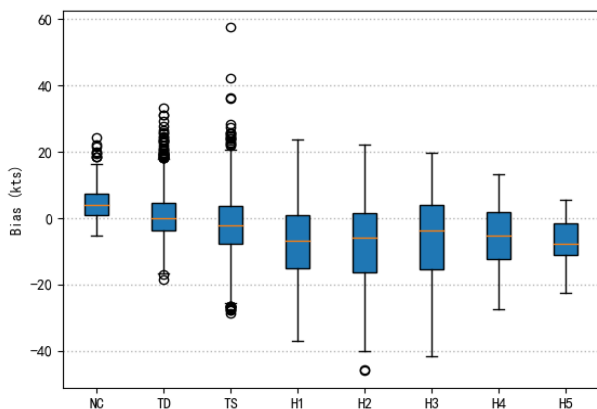
Fig. 8. Box plot of the bias between the estimated intensity and best-track intensity of TC for different categories.

change during training. If the task of size estimation converges, but the task of intensity estimation does not converge. If the training continues until the intensity estimation task converges, the size estimation task will be overfitted. If the size estimation task is not trained at this time, the task of intensity estimation will not converge. Finally, we design a model multitask4 based on gradient normalization to dynamically update loss function weight. The experimental results show that multitask4 yields minimal TC intensity and size estimation errors. As shown in Fig. 7, the RMSE and MAE of the intensity estimation and size estimation in multitask4 are the smallest (RMSE: 8.40 kts and 23.01 nmi, MAE : 6.90 kts and 20.89 nmi, respectively).

C. Optimized Versions of Overestimated and Underestimated Model

In this study, TCs are classified into eight categories based on the Saffir-Simpson Hurricane Wind Scale (SSHWS) (T1-T5) [30], Tropical Storm (TS), Tropical Depression (TD), and No Category (NC).

TC observation data present a long-tailed distribution, in which TD and TS samples account for 29.3% and 37.7%, whereas NC, H1, H2, H3, H4, and H5 samples account for 8.8%, 9.9%, 5.1%, 3.9%, 4%, and 0.8%. We validate the model's

performance using independent data in 2017 based on the optimal Mul_task4 configuration introduced in the previous section. As shown in Fig. 8, for NC (intensity < 20 kts) category, the bias of most samples is greater than 0, that is, these samples are significantly overestimated. For H1, H2, H3, H4, and H5 categories (intensity > 63 kts), approximately three-quarters of the samples are underestimated.

We utilize the label distribution smoothing method based on the similarity of nearby objects to deal with the overestimation and underestimation problems. The detailed experimental results are shown in Table IV. For the NC category, the RMSE of TC intensity estimation without data balancing is 7.12 kts and the bias is 7.11 kts. The RMSE from data balance experiment is 5.3 kts and the bias is 4.6 kts. Obviously, the label distribution smoothing method effectively alleviates the overestimation of TC intensity. For the H5 category, the RMSE of TC intensity estimation decreases from 10.34 kts to 8.0 kts after the balancing operation and tne bias changes from −10.22 kts to −6.8 kts. Obviously, the bias between the estimated and best-track intensity is reduced. For the TC size estimation, the RMSE and bias without data balancing are 7.21 nmi and 7.1 nmi for the NC category, and the RMSE from data balance experiment is 6.1 nmi and the bias is 4.4 nmi. For the H4 category, the RMSE of TC size changes from 34.6 nmi to 33.1 nmi after the balancing operation, and the bias change from −25.8 nmi to −24.0 nmi. The experimental results show that the label distribution smoothing method can effectively alleviate the problem of overestimation and underestimation for TC intensity and size estimation caused by unevenly distributed data.

D. Visualization

The TC-MTLNet model achieves excellent performance. However, we cannot accurately understand the CNN model's internal knowledge and the underlying reasons that drive it to make specific decisions. We visualize the output feature maps of our optimal model to understand the features learned by the network through different convolutional layers. What's more, we understand the role of the convolution kernel by the deconvolution visualization technique. Finally, we understand
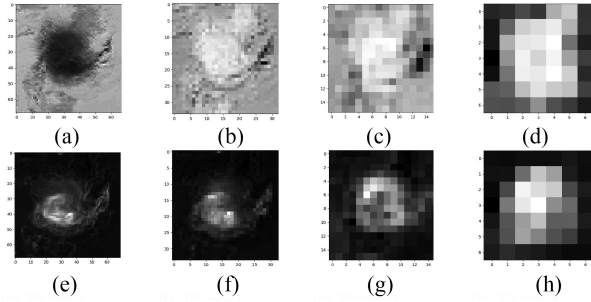
Fig. 9. Fused feature maps of 2D conv1, 2D conv2, 2D conv3, 2D conv4 convolutional layers and the fused feature maps of the 3D conv1, 3D conv2, 3conv3, and 3con4 convolutional layers. (a) 2D conv1. (b) 2D conv2. (c) 2D conv3. (d) 2D conv4. (e) 3D conv1. (f) 3D conv2. (g) 3D conv3. (h) 3D conv4.

Fig. 10. Features learned by the 16 convolution kernels of 2conv2. Features learned by the 64 convolution kernels of 2conv4.

the contribution of each part of the imagery to TC intensity and size estimation through heatmaps.

1) The Visualization of Feature Maps: By visualizing the feature maps generated by each convolution layer, we can understand the learning process of the TC-MTLNet. Based on the optimal model, the fused output feature maps of the 2D conv1, 2D conv2, 2D conv3, 2D conv4, 3D conv1, 3D conv2, 3D conv3, and 3D con4 convolutional layers are visualized, which are shown in Fig. 9. Fig. 9(a) and (e) is the fused feature maps of the first 2D and 3D convolutional layer. We find that the network close to the bottom layer extracts easy-to-understand and general visual features such as TC contour edge features. Fig. 9(d) and (h) are fused feature maps of the fourth 2D and 3D convolutional layer, and the extracted features are mainly the TC structural features. These features synthesize the underlying visual features, which are more conducive to making TC intensity and size estimation decisions. Obviously, as the number of convolutions deepens, the learned features become more abstract and detailed. Abstract features represent high-level visual features, which are more helpful to accurate TC intensity and size estimation. Compared with the fused feature maps of 2D conv1, 2D conv2, 2D conv3, and 2D conv4, the fused feature maps of 3D conv1, 3D conv2, 3D conv3, and 3D conv4 extract more abundant features related to TC intensity and size. It is demonstrated that 3-D convolution extracts the 3-D features of TC by sliding in the channel dimension.

2) Understanding the Role of the Convolution Kernel by Deconvolution Visualization: Because different filters extract different features, we can understand the role of filters by visualizing the convolution kernels. However, directly visualized convolution kernels are abstract and which do not reflect much information. In order to understand the role of the filter, we use the deconvolution technique to visualize the feature maps generated by the filters. The specific approach is to multiply the feature maps learned by the our model by the transpose matrix of the convolution kernel corresponding to these feature maps to find the pixels activated by specific feature maps. The visualized imageries obtained by deconvolution technology are shown in Fig. 10. Fig. 10(a) shows simple spiral edge features learned by the 16 convolution kernels of 2D conv2. For example, it can be seen that the 11th convolution kernel of 2D conv2 mainly
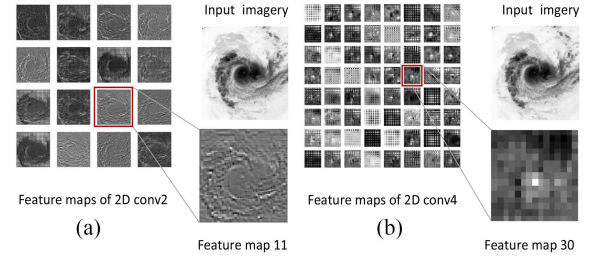
extracts edge features. Fig. 10(b) shows abstract detail features learned by the 64 convolution kernels of 2D conv4. For example, the 30th convolution kernel of 2D conv4 extracts the TC eye features.

3) Visualization of High Dimensional Features: There is a certain spatial correspondence between the output feature maps by the convolutional layer and the original imagery. We employ the Grad-Cam++ method to visualize the high-dimensional features. The method utilizes the feature maps of the last convolutional layer to generate a heatmap. We judge the contribution of each part of the satellite imagery to TC intensity and size estimation results by heatmaps, and reveal the critical factors that affect the intensity and size estimation results. The heatmaps and original imageries based on eight categories of satellite imageries are shown in Fig. 11. In the heatmap, the red areas represent the most important areas. Obviously, the higher the TC intensity, the higher the importance of the central region. It can be found that H1, H2, H3, H4, and H5 categories pay more attention to the inner core area. That is, high-intensity TCs pay more attention to the TC eye area. Because weak TCs do not focus on convective cloud centers, NC, TD, and TS categories pay more attention to the outer rainbands and cloud areas.

E. Performance Evaluation

In this section, we select independent data from 2017 to evaluate the performance of our best-optimized model.

To verify the applicability of our model, we select TC No. 11 and No. 15 in the Atlantic and TC No. 6 and No. 15 in the East Pacific to evaluate the ability of the model to estimate TC intensity. The TC life cycle sequence diagrams are shown in Fig. 12. The abscissa is time, and the ordinate is TC intensity. Compared with the most widely used ADT and SATCON techniques, our experimental results show that our model generalizes well for full-lifetime TCs. We employ TC No. 9, No. 6, No. 13, No. 10, and No. 05 in the East Pacific in 2017 and TC No. 12, No. 17, No. 09, No. 11, No. 08, and No. 13 in the Atlantic to evaluate the ability of the model to estimate TC size. The time series graphs of life cycle are shown in Fig. 13. The red dotted line represents the best-track size and the blue dotted line represents the model-estimated TC size. We can observe that the model performance is lower when TC size increases, which may be because more complex meteorological factors are involved in the outer wind radius, which is a great challenge for our model and is worth studying in the future.

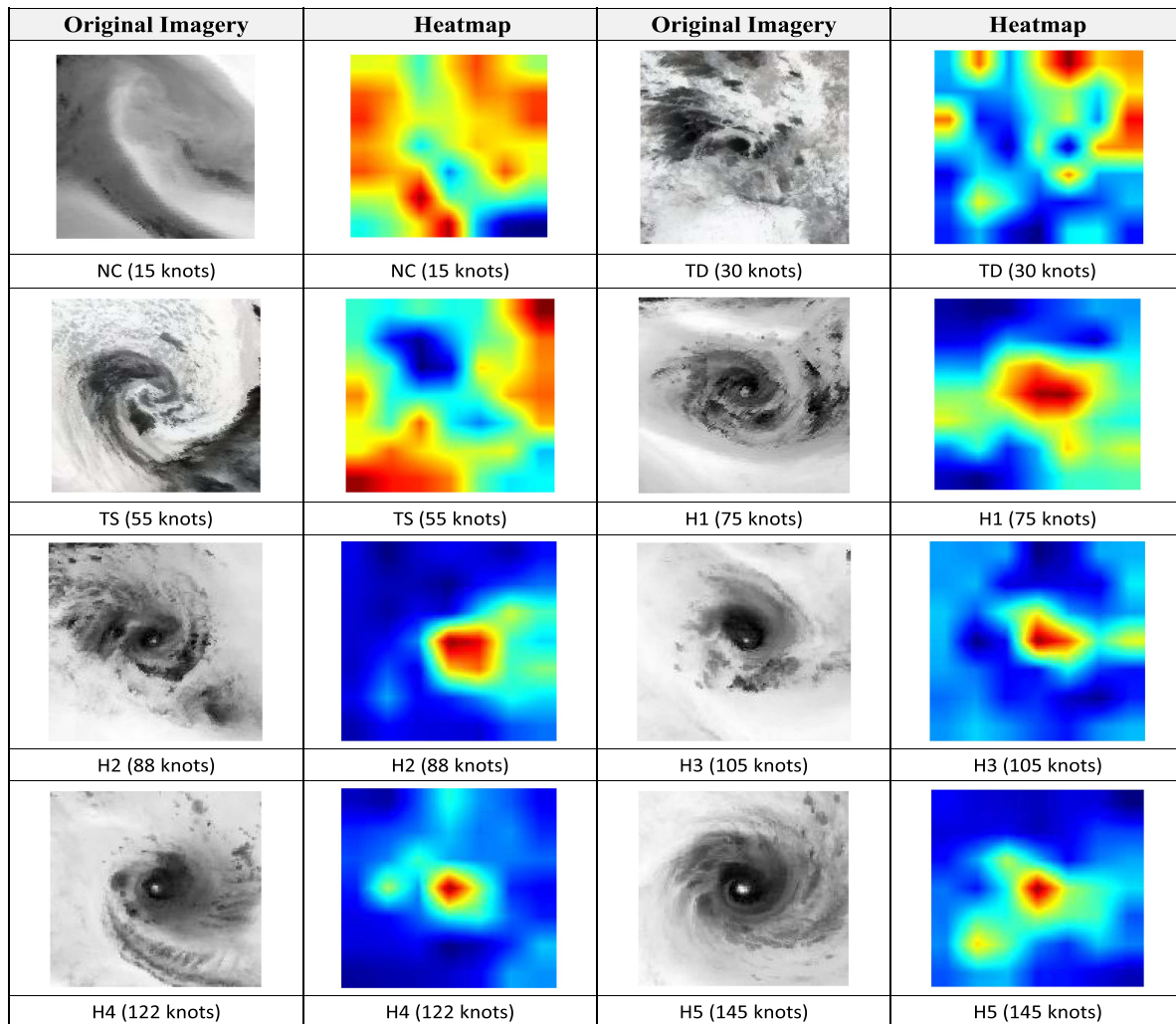| Original Imagery	Heatmap	Original Imagery	Heatmap
NC (15 knots)	NC (15 knots)	TD (30 knots)	TD (30 knots)
TS (55 knots)	TS (55 knots)	H1 (75 knots)	H1 (75 knots)
H2 (88 knots)	H2 (88 knots)	H3 (105 knots)	H3 (105 knots)
H4 (122 knots)	H4 (122 knots)	H5 (145 knots)	H5 (145 knots)

Fig. 11. Visualization of high dimensional features by Grad-Cam++ method. The heatmaps and original satellite imageries based on 8 TC categories of satellite imageries.

TABLE V
COMPARISON OF OUR METHOD AND EXISTING METHODS FOR TC INTENSITY ESTIMATION

	Method	Input	Region	RMSE (kts)
Dvorak [10]	Empirical	IR,VIS	Global	10–15
DAVT [12]	Statistical	IR	North east/West pacific	12.9–13.4
SATCON [15]	Statistical ensemble	IR,PWM	Global	9.0
ADT [11]	Statistical empirical	IR,PWM	Global	12.65
Ritchie et al. [12]	Static Analysis	IR	Atlantic, Pacific	12.7
Pradhan et al. [30]	2D CNN	IR	Atlantic, Pacific	10.18
Combinid et al. [44]	2D CNN	IR	Western North Pacific	13.23
Winmmer et al. [45]	2D CNN	37GHz,85-92GHz	Atlantic, Pacific	14.3
Tian et al. [35]	3D CNN	IR,WV,PWM	Atlantic, Eastern North Pacific and Western North Pacific	9.48
TC-MTLNet	2D and 3D CNN	IR,WV,PWM	Global	**8.40**

In order to verify the validity of our model, we first display a scatter plot to examine the model's ability to fit the best-track intensity and best-track size (see Fig. 14). In Fig. 14(a), the abscissa represents the best-track intensity, and the ordinate represents the estimated TC intensity. The scatter plot shows a significant positive correlation between the best-track intensity and the estimated intensity ($r = 0.935$), reflecting our model's excellent performance. In Fig. 14(b), the abscissa represents the best-track size, and the ordinate represents the estimated TC size. It is seen that although there are few outliers, the estimated size is significantly correlated with the best-track size. In general, the model fits the best-track size well.

As shown in Table V, the quantitative comparisons are impossible since previous methods employed satellite data from
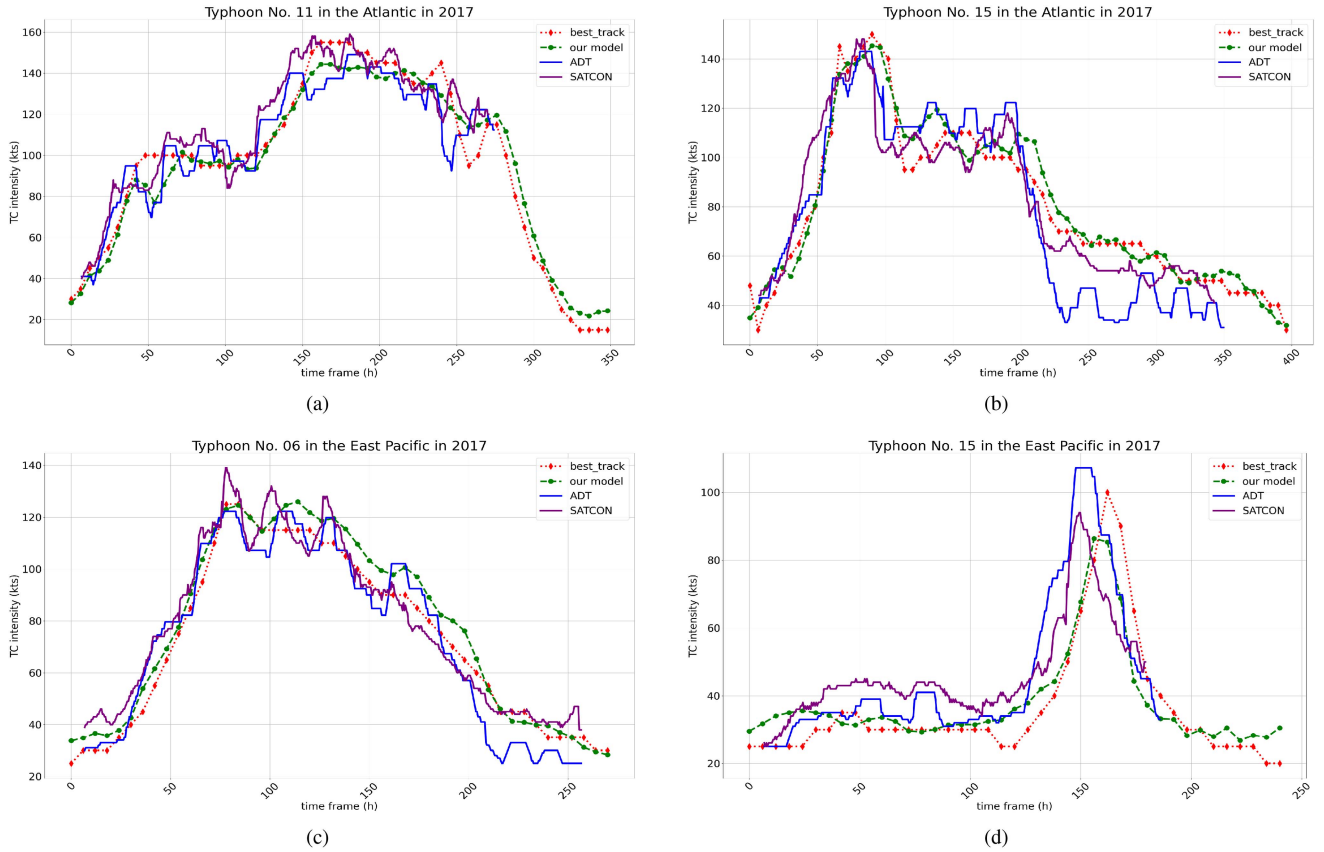
Fig. 12.   Time series diagram of life cycle of TC intensity from TC No. 11 and No. 15 in the Atlantic and TC No. 06 and No. 15 in the East Pacific in 2017, to evaluate the ability of the model to estimate TC intensity.

TABLE VI
COMPARISON OF OUR METHOD AND EXISTING METHODS FOR TC
SIZE ESTIMATION

	Input	Region	MAE (nmi)
MTCSWA [3]	IR		26.5
Meng et al. [46]	IR	Northwest Pacific, Northeast Pacific, Atlantic Ocean	24.4
TC-MTLNet	IR,WV,PMW	Global	**20.89**

TABLE VII
PARAMETERS, SPACE CONSUMPTION AND TIME CONSUMPTION OF
MODEL TRAINING

Layer	Parameter	Memory	Training time
Input	3×141×141		
2D conv1	256		
2D conv2	4608		
2D conv3	18432		
2D conv4	115774		
PAM	2048		
	2048		
	16384		
3D conv1	256	6695 MB	8 h
3D conv2	4608		
3D conv3	36864		
3D conv4	147456		
CAM			
Environment	2		
3×FC1	803072		
3×FC2	8192		
3×FC3	64		
Sum	1971394		

different regions and channels. However, our study is based on satellite data from three channels around the globe, and our model has wider applicability than other methods. In Table V, we compare the results of our model with existing methods. The TC intensity RMSE of of our model is 8.40 kts, which is lower than that of previous study. Thus, the model we designed has a higher accuracy and wider applicability.

In Table VI, we compare the TC-MTLNet model with existing deep learning models and traditional methods. It can be found that the performance of the TC-MTLNet model exceeds that of MTCSWA and Meng et al.

The parameters, space consumption, and time consumption of model training are shown in Table VII, which clearly shows the number of parameters in each layer. To reduce the number of parameters, we mainly use 3×3 and 1×1 convolution. We can find that a t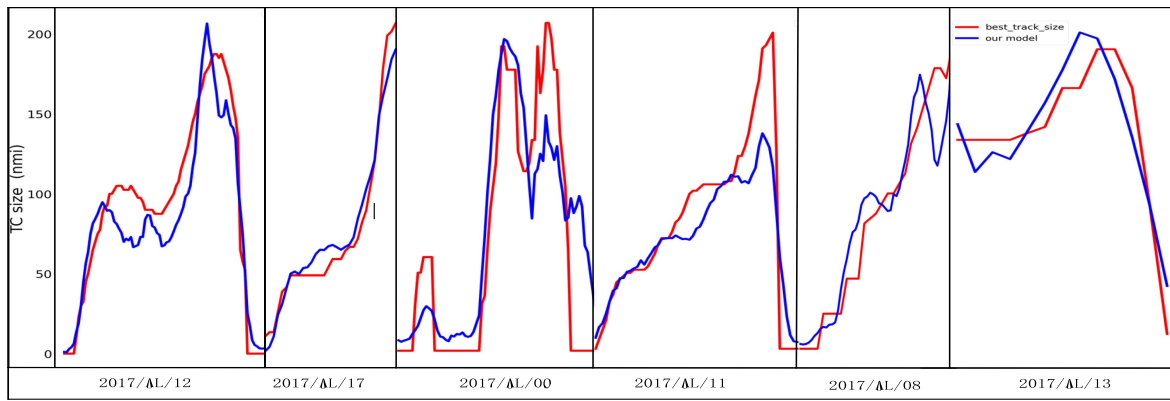otal of 1 971 394 parameters are used in the process of model learning, occupying 6695 MB of space resources and about 8 h of time resources. Under the premise of ensuring the accuracy of TC intensity and size estimation, our model also consumes less time, space, and computational resources.
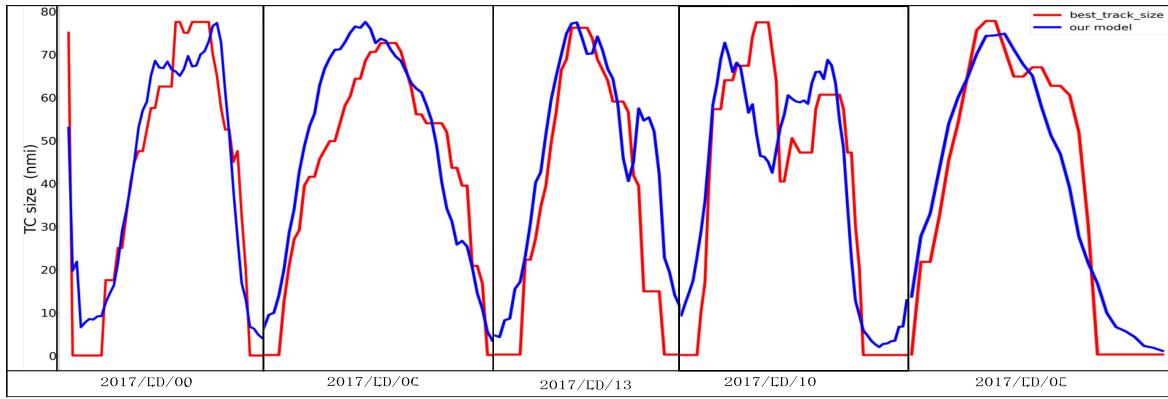
(a)

(b)

Fig. 13. Time series diagram of life cycle of TC size from TC No. 09, No. 06, No. 13, No. 10, and No. 05 in the East Pacific (EP) in 2017. Time series diagram of life cycle of TC size from TC No. 12, No. 17, No. 09, No. 11, No. 08, and No. 13 in the Atlantic (AL) in 2017.
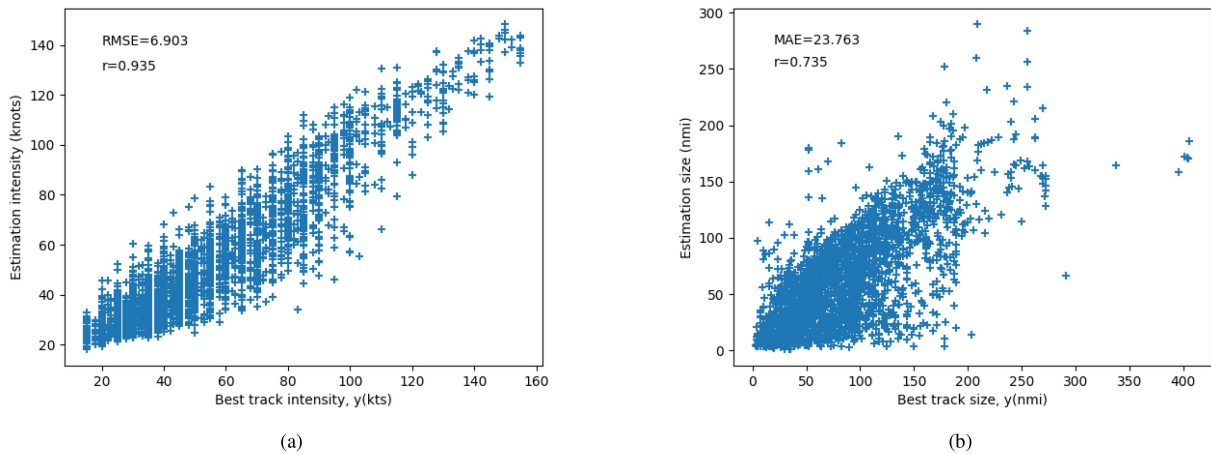

(a)

(b)

Fig. 14. Scatter plots of correlation between estimated intensity and best-track intensity and estimated size and best-track size.

VI. CONCLUSION

In this study, we proposed a parallel dual-attention TC intensity and size estimation model (TC-MTLNet) based on multitask learning. The model takes three measures to improve the ability of TC intensity and size estimation. First, the model employs two parallel branches to extract spatial and channel features. Before the fusion of the two branches, the spatial and the CAMs are applied to the spatial and channel dimensions to obtain the interdependence of features on the feature maps

and the correlation between channels. This approach improves the capability of feature representation, thereby improving the performance of our model. Compared with single-task learning, multitask learning can learn additional useful information from related tasks, positively affecting TC intensity and size estimation. However, the multiple tasks learning speed and gradient magnitude are different. Therefore, we adopt the gradient regularization method to prevent the learning of multiple tasks from being dominated by one task. In addition, we propose a label distribution smoothing based on kernel distribution by taking advantage of the similarity between nearby targets in order to solve the overestimation and underestimation caused by the uneven distribution in TC observation data. This method reduces the bias in TC intensity and size estimation, and further improves the accuracy of model estimation. Finally, we visualize feature maps, convolution kernels, and high-dimensional features to understand how the TC-MTLNet model learns intensity and size features and identify the essential factors for intensity and size estimation from TC satellite imagery. We evaluated the performance of the best-optimized model using independent data from 2017. The experiments show that our model generates an intensity estimation RMSE of 8.40 kts, which is lower by 33.5% compared to ADT and 11.4% compared to 3DAttentionTCNet. Furthermore, a size estimation MAE of 23.76 nmi shows that the performance of our model surpasses MTCSWA by 10%. In this article, we design a model that alleviates the overestimation and underestimation problems caused by unevenly distributed TC data. In future work, this is a research direction. In addition, due to the lack of deep learning research on TC size, a future work will focus on systematically studying TC size estimation to provide more convenience for TC disaster prevention.

REFERENCES

[1] Q. Zhang, L. Wu, and Q. Liu, "Tropical cyclone damages in china 1983–2006," *Bull. Amer. Meteorological Soc.*, vol. 90, no. 4, pp. 489–496, 2009.

[2] S. Seneviratne et al., "Changes in climate extremes and their impacts on the natural physical environment," pp. 109–230, 2012.

[3] J. A. Knaff, M. DeMaria, D. A. Molenar, C. R. Sampson, and M. G. Seybold, "An automated, objective, multiple-satellite-platform tropical cyclone surface wind analysis," *J. Appl. Meteorol. Climatol.*, vol. 50, no. 10, pp. 2149–2166, 2011.

[4] M. DeMaria, C. R. Sampson, J. A. Knaff, and K. D. Musgrave, "Is tropical cyclone intensity guidance improving?," *Bull. Amer. Meteorological Soc.*, vol. 95, no. 3, pp. 387–398, 2014.

[5] J. A. Knaff and C. R. Sampson, "After a decade are atlantic tropical cyclone gale force wind radii forecasts now skillful?," *Weather Forecasting*, vol. 30, no. 3, pp. 702–709, 2015.

[6] J. A. Knaff, C. J. Slocum, K. D. Musgrave, C. R. Sampson, and B. R. Strahl, "Using routinely available information to estimate tropical cyclone wind structure," *Monthly Weather Rev.*, vol. 144, no. 4, pp. 1233–1247, 2016.

[7] M. A. Bender, T. P. Marchok, C. R. Sampson, J. A. Knaff, and M. J. Morin, "Impact of storm size on prediction of storm track and intensity using the 2016 operational GFDL hurricane model," *Weather Forecasting*, vol. 32, no. 4, pp. 1491–1508, 2017.

[8] D. J. Cecil and E. J. Zipser, "Relationships between tropical cyclone intensity and satellite-based indicators of inner core convection: 85-GHZ ice-scattering signature and lightning," *Monthly Weather Rev.*, vol. 127, no. 1, pp. 103–123, 1999.

[9] D. Herndon and C. Velden, "An update on tropical cyclone intensity estimation from satellite microwave sounders," in *Proc. 31st Conf. Hurricanes Trop. Meteorol.*, 2014, p. 34.

[10] V. F. Dvorak, "Tropical cyclone intensity analysis and forecasting from satellite imagery," *Monthly Weather Rev.*, vol. 103, no. 5, pp. 420–430, 1975.

[11] C. Velden, T. Olander, D. Herndon, and J. P. Kossin, "Reprocessing the most intense historical tropical cyclones in the satellite era using the advanced dvorak technique," *Monthly Weather Rev.*, vol. 145, no. 3, pp. 971–983, 2017.

[12] E. A. Ritchie, K. M. Wood, O. G. Rodríguez-Herrera, M. F. Piñeros, and J. S. Tyo, "Satellite-derived tropical cyclone intensity in the north pacific ocean using the deviation-angle variance technique," *Weather Forecasting*, vol. 29, no. 3, pp. 505–516, 2014.

[13] D. Herndon, C. Velden, J. Hawkins, T. Olander, and A. Wimmers, "The cimss satellite consensus (SATCON) tropical cyclone intensity algorithm," in *Proc. 29th Conf. Hurricanes Trop. Meteorol. D*, 2010.

[14] C. Velden and D. Herndon, "Update on the satellite consensus (SATCON) algorithm for estimating TC intensity," in *Proc. 31st Conf. Hurricanes Trop. Meteorol.*, 2014.

[15] C. Velden and D. Herndon, "An update on the CIMSS satellite consensus (SATCON) tropical cyclone intensity algorithm," in *Proc. Joint Satell. Conf.*, 2019.

[16] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[18] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 818–833.

[19] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[20] C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[22] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.

[23] S. Hsu and A. Babin, "Estimating the radius of maximum wind via satellite during hurricane lili (2002) over the gulf of Mexico," *Nat. Weather Assoc. Electron. J*, vol. 6, no. 3, pp. 1–6, 2005.

[24] K. J. Mueller, M. DeMaria, J. Knaff, J. P. Kossin, and T. H. Vonder Haar, "Objective estimation of tropical cyclone wind structure from infrared satellite data," *Weather Forecasting*, vol. 21, no. 6, pp. 990–1005, 2006.

[25] M. Kwon, "Estimation and statistical characteristics of the radius of maximum wind of tropical cyclones using coms IR imagery," *Atmosphere*, vol. 22, no. 4, pp. 473–481, 2012.

[26] Y.-K. Lee and M. Kwon, "An estimation of the of tropical cyclone size using coms infrared imagery," *Atmosphere*, vol. 25, no. 3, pp. 569–573, 2015.

[27] J. A. Knaff, S. P. Longmore, R. T. DeMaria, and D. A. Molenar, "Improved tropical-cyclone flight-level wind estimates using routine infrared satellite reconnaissance," *J. Appl. Meteorol. Climatol.*, vol. 54, no. 2, pp. 463–478, 2015.

[28] M. J. Brennan, C. C. Hennon, and R. D. Knabb, "The operational use of quikscat ocean surface vector winds at the national hurricane center," *Weather Forecasting*, vol. 24, no. 3, pp. 621–645, 2009.

[29] J. L. Demuth, M. DeMaria, J. A. Knaff, and T. H. Vonder Haar, "Evaluation of advanced microwave sounding unit tropical-cyclone intensity and size estimation algorithms," *J. Appl. Meteorol.*, vol. 43, no. 2, pp. 282–296, 2004.

[30] R. Pradhan, R. S. Aygun, M. Maskey, R. Ramachandran, and D. J. Cecil, "Tropical cyclone intensity estimation using a deep convolutional neural network," *IEEE Trans. Image Process.*, vol. 27, no. 2, pp. 692–702, Feb. 2017.

[31] B. Chen, B.-F. Chen, and H.-T. Lin, "Rotation-blended CNNs on a new open dataset for tropical cyclone image-to-intensity regression," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2018, pp. 90–99.

[32] B.-F. Chen, B. Chen, H.-T. Lin, and R. L. Elsberry, "Estimating tropical cyclone intensity by satellite imagery utilizing convolutional neural networks," *Weather Forecasting*, vol. 34, no. 2, pp. 447–465, 2019.

[33] J. Lee, J. Im, D.-H. Cha, H. Park, and S. Sim, "Tropical cyclone intensity estimation using multi-dimensional convolutional neural networks from geostationary satellite data," *Remote Sens.*, vol. 12, no. 1, pp. 108–131, 2019.

[34] J.-Y. Zhuo and Z.-M. Tan, "Physics-augmented deep learning to improve tropical cyclone intensity and size estimation from satellite imagery," *Monthly Weather Rev.*, vol. 149, no. 7, pp. 2097–2113, 2021.

[35] W. Tian, X. Zhou, W. Huang, Y. Zhang, P. Zhang, and S. Hao, "Tropical cyclone intensity estimation using multi-dimensional convolutional neural network from multi-channel satellite imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, no. 12, pp. 1–5, 2021.

[36] Y. Sun et al., "Multisensor fusion and explicit semantic preserving-based deep hashing for cross-modal remote sensing image retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, no. 12, pp. 1–14, 2021.

[37] Y. Sun et al., "Unsupervised deep hashing through learning soft pseudo label for remote sensing image retrieval," *Knowl. Based Syst.*, vol. 239, 2022, Art. no. 107807.

[38] P. Duan, P. Ghamisi, X. Kang, B. Rasti, S. Li, and R. Gloaguen, "Fusion of dual spatial information for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7726–7738, Sep. 2020.

[39] P. Duan, Z. Xie, X. Kang, and S. Li, "Self-supervised learning-based oil spill detection of hyperspectral images," *Sci. China Technological Sci.*, vol. 65, no. 4, pp. 793–801, 2022.

[40] J. Tan, Q. Yang, J. Hu, Q. Huang, and S. Chen, "Tropical cyclone intensity estimation using himawari-8 satellite cloud products and deep learning," *Remote Sens.*, vol. 14, no. 4, pp. 812–829, 2022.

[41] J. L. Demuth, M. DeMaria, and J. A. Knaff, "Improvement of advanced microwave sounding unit tropical cyclone intensity and size estimation algorithms," *J. Appl. Meteorol. Climatol.*, vol. 45, no. 11, pp. 1573–1581, 2006.

[42] J. P. Kossin et al., "Estimating hurricane wind structure in the absence of aircraft reconnaissance," *Weather Forecasting*, vol. 22, no. 1, pp. 89–101, 2007.

[43] Y.-H. Baek, I.-J. Moon, J. Im, and J. Lee, "A novel tropical cyclone size estimation model based on a convolutional neural network using geostationary satellite imagery," *Remote Sens.*, vol. 14, no. 2, pp. 426–444, 2022.

[44] J. S. Combinido, J. R. Mendoza, and J. Aborot, "A convolutional neural network approach for estimating tropical cyclone intensity using satellite-based infrared images," in *Proc. 24th Int. Conf. Pattern Recognit.*, 2018, pp. 1474–1480.

[45] A. Wimmers, C. Velden, and J. H. Cossuth, "Using deep learning to estimate tropical cyclone intensity from satellite passive microwave imagery," *Monthly Weather Rev.*, vol. 147, no. 6, pp. 2261–2282, 2019.

[46] F. Meng, P. Xie, Y. Li, H. Sun, D. Xu, and T. Song, "Tropical cyclone size estimation using deep convolutional neural network," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 8472–8475.

Wei Tian received the M.S. degree in systems analysis and integration and the Ph.D. degree in meteorology from Nanjing University of Information Science and Technology, Nanjing, China, in 2007 and 2016, respectively.

He worked as a Research Scholar with Michigan State University, East Lansing, MI, USA, from Apr. 2017 to May 2018. He is currently an Associate Professor with the school of computer and software, Nanjing University of Information Science and Technology. His research interest covers computer software, tropical cyclone, remote sensing image processing, and deep learning.

Xinxin Zhou is currently working toward the M.S. degree in software engineering with the Nanjing University of Information Science and Technology, Nanjing, China.

Her research interests include tropical cyclone and deep learning.

Xianhua Niu received the M.S. degree in information engineering from the Nanjing University of Information Science and Technology, Nanjing, China, in 2006.

Her research interests include meteorological ocean information and deep learning.

Linhong Lai is currently working toward the M.S. degree in software engineering with the Nanjing University of Information Science and Technology, Nanjing, China.

His research interests include tropical cyclone and deep learning.

Yonghong Zhang received the M.S. degree in mechanical design and theory from Nanjing Forestry University, Nanjing, China, in 2002, and the Ph.D. degree in mechanical manufacturing and its automation from Shanghai Jiao Tong University, Shanghai, China, in 2005.

He was a Postdoctoral Researcher with the University of Alberta, Edmonton, AB, Canada from April 2007 to September 2008. From July 2009 to June 2017, he successively served as the Deputy Dean of the School of Automation, Nanjing University of Information Science and Technology, the Deputy Director of the Academic Affairs Office (in charge of the work), and the Director of the Academic Affairs Office (from January 2017 to June 2017, he temporarily served as the Standing Committee Member and Deputy Director of Xishan District Committee of Wuxi City); From July 2017 to June 2021, he was a member of the Standing Committee of the Party Committee and Vice President of Nanjing University of Information Science and Technology. In July 2021, he served as Deputy Secretary of CPC Committee and President of Wuxi University.

Kenny Thiam Choy Lim Kam Sian received the M.Sc. degree in meteorology from the Nanjing University of Information Science and Technology (NUIST), Nanjing, China, in 2017, and the Ph.D. degree in meteorology from NUIST, in 2020.

He is currently a Lecturer with Wuxi University, Wuxi, China. His research interest includes air-sea interaction, coupled numerical modeling of tropical cyclones, and weather forecasting.