Tropical Cyclone Intensity Classification and Estimation Using Infrared Satellite Images With Deep Learning

Chang-Jiang Zhang, Xiao-Jie Wang, Lei-Ming Ma, and Xiao-Qin Lu

***Abstract*—A novel tropical cyclone (TC) intensity classification and estimation model (TCICENet) is proposed using infrared geostationary satellite images from the northwest Pacific Ocean basin in combination with a cascading deep convolutional neural network (CNN). The proposed model consists of two CNN network modules: a TC intensity classification (TCIC) module and a TC intensity estimation (TCIE) module. First, the TCIC module is utilized to divide TC intensity into three categories using infrared satellite images. Next, three TCIE models based on the CNN regression network that combine different intensity types of infrared satellite images with the TC best track data are presented. The three TCIE models consider classification error with the TCIC module in order to improve TCIE accuracy. A total of 1001 TCs from 1981–2019 were used to verify the proposed TCICENet model, with 844 TCs from 1981–2013 employed as training samples, 76 TCs from 2014–2016 used as validation samples, and 81 TCs from 2017–2019 used as testing samples. In order to reduce the computation burden of training the TCICENet model, various input image sizes were explored. An image size of 170 × 170 pixels achieved the best performance, with an overall root mean square error of 8.60 kt and a mean absolute error of 6.67 kt compared to the best track.**

***Index Terms*—Deep convolutional neural network (CNN), intensity estimation, intensity grade classification, tropical cyclone (TC).**

I. Introduction

THE northwest Pacific Ocean basin is one of the most active tropical cyclone (TC) areas in the world, generating approximately 27 TCs per year. According to the TC standards (http://agora.ex.nii.ac.jp/digital-typhoon/help/unit.html.ja#id2) released by the Japan Meteorological Agency (JMA), TCs can be divided into six grades: tropical depression (TD; ∼33 kt), tropical storm (TS; 34–47 kt), severe tropical storm (STS; 48–63 kt), strong typhoon (STY; 64–84 kt), very strong typhoon (VSTY; 85–104 kt), and violent typhoon (ViolentTY; ≥ 105 kt). In the 1940s, aircraft detection was first utilized as an effective means of monitoring the current position, intensity, and development status of TCs. However, aircraft detection primarily focuses on the Atlantic Ocean basin. In recent years, the development of various means of observation and numerical forecasting has improved storm track prediction, but improving intensity estimation has been very slow.

Current TC intensity estimation primarily depends on microwave data from polar-orbiting [1]–[6] and geostationary satellite imagery. The microwave data are easily disturbed by harsh environments, such as heavy rainfall. Compared with polar-orbiting satellite data, the geostationary satellites are unable to obtain near-surface structure of TC but have a higher temporal resolution and the stable image quality, thus are currently more commonly used for TC intensity estimation. Dvorak proposed and developed methods [7]–[10] for TC intensity estimation using infrared satellite images, which rely upon the experience of the forecaster [11]. Although some improved Dvorak techniques [12], [13] reduce the subjective steps, most of them are not suitable for estimating the intensity of weak TCs [14]. Olander and Velden [15] further improved the Dvorak technique, creating a method known as the advanced Dvorak technique (ADT) version 9 to improve the intensity estimation for weak TCs. The deviation-angle variance technique (DAT) introduced by Pineros *et al.* [16] is a method that quantifies the axisymmetry of a TC from infrared satellite imagery in order to investigate structural and intensity changes of TCs [17]. The DAT and its improved versions produce good TC intensity estimation results [18], [19].

Researchers have utilized traditional machine learning with infrared satellite images for TC intensity estimation. The multivariate linear regression models [14]–[20], the K-nearest neighbors algorithm [21], [22], the multilayer perceptron [23], support vector machine (SVM) [24], and relevance vector machine (RVM) [25]–[27] have been successfully used to estimate TC intensity. These methods mainly focus on the manual extraction of statistical features [1], [2], [14], [23], [25]–[27] or structural features [7]–[10], [12], [13], [15]–[17], [42] of TC. These characteristics are mainly as follows: the TC center, rain band features, brightness temperature gradient of cloud tops, statistics of radial cloud top brightness temperature, slope of TC inner-core cloud top, the deviation angle, spiral rain band fitting,

Manuscript received June 20, 2020; revised October 30, 2020, November 28, 2020, and December 20, 2020; accepted December 29, 2020. Date of publication January 12, 2021; date of current version February 3, 2021. This work was supported in part by National Natural Science Foundation of China (42075140 and 41575046), and in part by the Project of Commonwealth Techniques and Application Research of Zhejiang Province, China under Grant LGF20D050004. *(Chang-Jiang Zhang and Xiao-Jie Wang contributed equally to this work.) (Corresponding author: Chang-Jiang Zhang.)*

Chang-Jiang Zhang and Xiao-Jie Wang are with the College of Physics and Electronic Information Engineering, Zhejiang Normal University, Jinhua 321004, China (e-mail: zcj74922@zjnu.edu.cn; wang_xiao_jayne@163.com).

Lei-Ming Ma is with the Shanghai Meteorological Center, Shanghai 200030, China (e-mail: malm@typhoon.org.cn).

Xiao-Qin Lu is with the Shanghai Typhoon Institute of the China Meteorological Administration, Shanghai 200030, China (e-mail: luxq@typhoon.org.cn).

Digital Object Identifier 10.1109/JSTARS.2021.3050767

and the symmetry of inner core. Furthermore, these features (usually <10) [36] are largely dependent on human subjectivity and experience, as well as prior information concerning satellite data. It is difficult for a meteorologist to determine whether one feature is suitable for the intensity regression of all of the various TCs in different basins and developing stages.

The convolutional neural network (CNN) of deep learning is very suitable for extracting the deep features (usually more than 1000) from an image using convolutional, pooling, and fully connected layers. It is unnecessary to extract the TC intensity indicating factors from infrared satellite images in advance for CNNs. Pradhan *et al.* [35], for example, used the CNN classification output vector to directly weight the average the intensity of the two highest possible categories in order to obtain the estimated TC intensity. This is straightforward, but it is easy to misclassify TCs. Recent studies [38]–[41] used CNN-based models to estimate the TC intensity as a regression task. The regression method is relatively simple and straightforward, but TCs exhibit rich and varied forms as well as a wide range of TC wind speed fluctuations, making it difficult for these single models to fully cover TC changes, especially when the TC samples of different intensities are not balanced. Existing TC intensity estimation models based on CNNs have only focused on basic and relatively simple deep learning network structures, such as AlexNet or VGGNet [34]–[41].

Unlike the methods above, we propose a cascade-double CNN model for estimating TC intensity (TCICENet) from infrared geostationary satellite images. The infrared satellite images are adopted as the input of the TCICENet model. The TCICENet model consists of two modules: the TC intensity grade classification (TCIC) module and the TC intensity estimation (TCIE) module from three TC categories as a regression task. Compared with the method of directly weighting the average TC intensity of the two highest possible categories [35], TCIE takes the characteristics of the infrared satellite image into account again, which is more reliable because of rough classification and fine-tuning.

TC intensity grade classifications are different from the conventional rigid targets (cars, animals, and etc.). TCs are not rigid bodies, and the structure of it will change greatly in different development stages with continuous rotation and translation. The characteristics of adjacent intensity categories of TC are quite similar, leading to a high probability of misclassification between adjacent intensity categories. Also, TC images with very different cloud structures may belong to the same intensity category. Conversely, cloud structure images that look similar may also belong to two different intensity categories, such as in the life cycle of the same TC. For these reasons, LeNet and other CNN networks with simple structure have difficulty achieving high-precision classification of TC intensity grade. A deeper and wider CNN network is required to extract more deep semantic information. The Inception module of GoogLeNet performs multiscale feature extraction on the input feature map. This model is wider, but the problem of gradient dispersion caused by network deepening still exists. ResNet learns the difference between the input and output of the network, which greatly alleviates performance degradation when the network is deeper. Inception-Resnet-v2 combined inception and residual structure, and achieved the best classification performance on the ImageNet dataset in 2016 [33]. Some deeper and broader CNN models have also been proposed recently, although no significant improvement in image classification has been obtained. Therefore, Inception-Resnet-v2 is suitable for the challenging task for the challenging task of TCIC.

After obtaining more accurate classification results of a TCIC module, the TCIE module need not be as deep and wide as that of the TCIC module considering both the calculation cost and intensity estimation accuracy of the TCIE module. Now, simple CNN models mainly include LeNet [28], AlexNet [29], and VGGNet [30]. Compared with LeNet, AlexNet has deeper layers, and its activation function is Relu, which can accelerate the convergence speed of stochastic gradient descent and alleviate the problem of gradient dispersion. Finally, our TCIE model combines the lightweight characteristics of AlexNet and the advantages of the small convolution kernel of VGGNet to increase the nonlinear expression ability of the network model, which can meet the requirements of TC intensity estimation.

Our contributions can be summarized as follows.

1) We propose a cascading CNN model to objectively estimate the TC intensity.
2) We improve Inception-ResNet-v2 using the channel and spatial attention modules to implement the challenging task of TC intensity grade classification.
3) We improve the AlexNet model and respectively build TC intensity estimation models for different categories of TCs. The misclassification ratio of the validation data of TCIC is used to guide the training of the TCIE module.

II. TCICENet Model

A. Overall Model Architecture

The TCICENet model is used to implement the TC-image-to-category-to-intensity estimation task. The TCICENet evolved from an improvement of the Inception-ResNet-v2 [33] and AlexNet [29] and is well suited to learning features of TC representation and distinguishing different TC categories from satellite images. The framework of the proposed TCICENet model architecture is shown in Fig. 1. The TCICENet consists of two modules: TC intensity classification (TCIC) and TC intensity estimation (TCIE). First, infrared satellite images are divided into three categories (category1, category2, and category3) using the TCIC module, where category1 is comprised of TS+STS, category2 consists of STY, and category3 comprises VSTY+ViolentTY. Second, three TC intensity estimation models corresponding to the three different TC intensity categories are respectively built using the TCIE module.

B. TCIC Module Architecture

Existing TC intensity estimation models rarely consider the intensity grade classification before intensity estimation, although some researchers have recently divided the TC intensity grades into many categories using CNNs [34], [35]. Due to the high misclassification probability when directly dividing

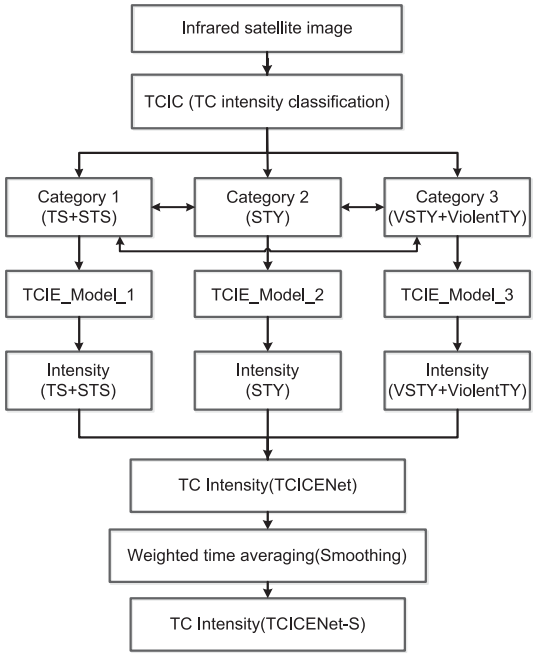


Fig. 1. Framework of the proposed TCICENet model architecture.

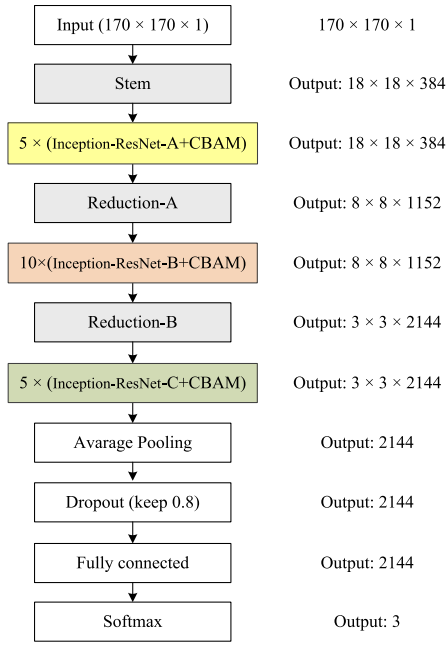


Fig. 2. Flowchart of the TCIC internal network architecture (the numbers 5, 10, and 5 indicate that there are 5, 10, and 5 duplicate blocks.).

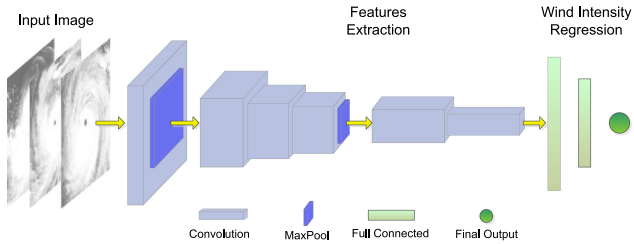


Fig. 3. Schematic block diagram of the TCIE network architecture.

TC intensity grades into too many categories, the accuracy of the intensity estimation is bound to be affected. We classified the TC intensity grades into three categories (TS+STS, STY, and VSTY+ViolentTY) in order to improve the classification accuracy. The Inception-ResNet-v2 was adopted as the basic network of the TCIC module, which is a very deep and wide CNN that has been successfully applied to different image feature extraction tasks. In addition, the convolutional block attention module (CBAM) [43] was integrated into the network structure of the Inception-ResNet-v2. The CBAM includes the channel attention module and the spatial attention module in order to make the network focus on the important features of the TC categories and suppress the unnecessary ones. The flowchart of the TCIC internal network architecture is presented in Fig. 2.

First, the infrared satellite image is input through the Stem module to initially extract the low-level features, and the output size is $18 \times 18 \times 384$ pixels. Next, five linearly stacked Inception-ResNet-A+CBAM modules are used to extract richer features, and then the Reduction-A module is used to reduce the feature map size and increase the number of channels so that the output size is $8 \times 8 \times 1152$ pixels. After ten linearly stacked Inception-ResNet-B+CBAM modules, the model further extracts richer features and semantic features. The Reduction-B module then continues to reduce the size of the feature map, while increasing the number of channels to improve the model performance so that the output size is $3 \times 3 \times 2144$ pixels. After five linearly stacked Inception-ResNet-C+CBAM modules, feature extraction is complete. Average pooling is used to convert the feature map into a one-dimensional vector. A dropout layer is used to prevent overfitting, followed by a fully connected layer. Finally, softmax outputs a one-dimensional vector with a size of three, which represents the possibility of each category. The category with the highest possibility is the final TC category.

C. TCIE Module Architecture

The TCIE module is based on AlexNet. Fig. 3 presents the schematic diagram of the TCIE network structure. The network structure parameters of the TCIE module are listed in Table I. The network consists of six convolutional layers, two pooling layers, and three fully connected layers. We discarded the large 11×11 convolutional layer in AlexNet layer 1 since large filters tend to increase computation time and will lose too many details in infrared (IR) images. The convolutional layer with a convolution kernel size of 5×5 in the second layer of AlexNet is directly used as the convolutional layer of the first layer of the TCIE, which is used to extract low-level features (contours and textures) from the input satellite image. We use a convolution kernel size of 3×3 from the second layer to the sixth layer. After the first pooling layer, three consecutive convolutional layers are used to extract the features of the satellite image, due to the complex relationship between satellite image information and TC intensity. After the pooling layer, two more successive

TABLE I
NETWORK STRUCTURE OF TCIE MODEL[a]

Layer	Shape[b]	Activation	Output size
input	1@170×170	—	170×170
Conv1	8@5×5, s=1, p=0	Leaky Relu	166×166
Pool1	4×4, s=4, p=0	—	41×41
Conv2	32@3×3, s=1, p=0	Leaky Relu	39×39
Conv3	32@3×3, s=1, p=0	Leaky Relu	37×37
Conv4	32@3×3, s=1, p=0	Leaky Relu	35×35
Pool2	2×2, s=2, p=0	—	17×17
Conv5	64@3×3, s=1, p=0	Leaky Relu	15×15
Conv6	64@3×3, s=1, p=0	—	13×13
Fc6	10816	Leaky Relu	512
Fc7	512	Leaky Relu	16
Fc8(Output)	1	—	1

[a]Terms—Shape: Size of the convolutional filter; Activation: Nonlinear "signal gate" processing method following the layer process; Leaky ReLU: "Leaky rectified linear unit" (a type of activation function that prevents neuron saturation in deep networks); Output size: Row×column size of the product of convolution/pooling/fully connected layer.

[b]Terms—*s*: Stride (Spacing of the convolutional/pooling filter application); p: Padding (Padding pixel to the edge of the input feature map)

convolutional layers are added in the TCIE. The first and fourth convolutional layers are connected with the pooling layers, with the goal of cutting down the size of the input image. The last convolutional layer has no pooling layer. We also use max pooling [29] to reduce the number of features by only keeping the maximum value of features in a sliding window. After max pooling the layers, the features are reduced sharply. "Leaky ReLU" [44] activation is used for nonlinearity, which is a fairly rapid and streamlined activation function.

Compared to AlexNet, our model has the following differences. The TCIE model discards the large 11×11 convolutional layer in the first layer and adds the last two consecutive convolutional layers Conv5 and Conv6. The number of channels in Conv1 is 8, that in Conv2–Conv4 is 32, and that in Conv5 and Conv6 is 64. The numbers of neurons in the full connection layers are 512, 16, and 1, respectively.

D. Loss Function

The cross-entropy is used as the loss function of the TCIC, as shown in the following equation:

$$H_{y'}\ (y) = \ -\sum_i y_i' \log\left(y_i\right) \tag{1}$$

where y_i' is the actual tag value and y_i is the predict label value of the output.

The L_2 regularization term Ω is added to the loss function, as shown in the following equation:

$$\Omega \ = \ \alpha \left\|\omega\right\|_2^2 \tag{2}$$

where ω is the total weight of the network model and α is the weighted decay.

The weight decay of the regularization term Ω is expressed as α, and has a value of 0.0005. By reducing the numerical value of the weight, the complexity of the model is reduced and overfitting is prevented. The total loss function is expressed as follows:

$$L \ = H_{y'}\ (y) + \Omega. \tag{3}$$

The loss function of the TCIE training process is $\text{smooth}_{\text{L1}}$ loss [45]

$$\text{smooth}_{L_1}\ (x) = \begin{cases} 0.5x^2, \text{ if } |x| < 1 \\ |x| - 0.5, \text{otherwise} \end{cases}. \tag{4}$$

L_2 loss converges faster and has a derivative at 0 to facilitate convergence. L_1 loss makes the network more robust to outliers. $\text{Smooth}_{\text{L1}}$ loss combines the advantages of both L_1 loss and L_2 loss.

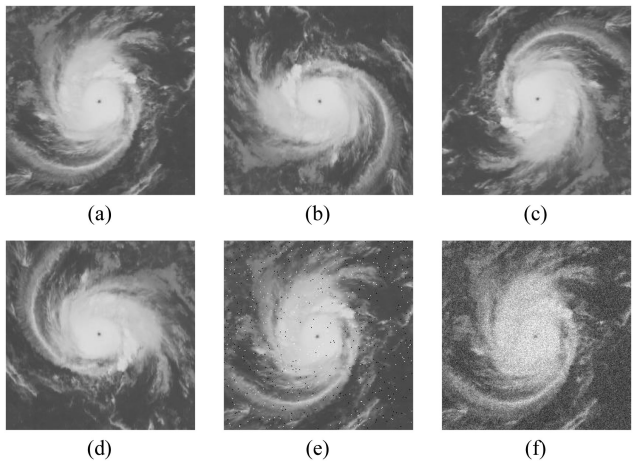


Fig. 4. Expansion of sample data set by different methods (original image from 0000 UTC December 22, 2001). (a) Original. (b) Rotated 90°. (c) Rotated 180°. (d) Rotated 270°. (e) With added salt-and-pepper noise. (f) With added Gaussian noise.

III. DATA

The data of 1001 TCs from 1981 to 2019, comprising five categories—TS, STS, STY, VSTY, and ViolentTY—were obtained from the National Institute of Informatics of Japan (http://agora.ex.nii.ac.jp/digital-typhoon/year/wnp/), and were then used to verify the performance of the proposed model. The infrared satellite images used in this study derive from the GMS1–5, GEO 9, MTSAT-1R, MTSAT-2, and Himawari 8 satellites in the northwest Pacific Ocean basin. The wavelength of the images is 10.2–12.5 μm. A total of 19 451 satellite images of 844 TCs from 1981 to 2013 were selected to train the proposed TCICENet model, 1907 images of 76 TCs from 2014 to 2016 were selected for validation, and 1794 images of 81 TCs from 2017 to 2019 were selected for testing. Please refer to http://www.jmanet.go.jp/sat/himawari/enkaku.html#nheader for more information pertaining to our data. Since the deep learning training process requires a large dataset of balanced types, the existing data did not meet the requirements of the training samples. Therefore, the original 19 451 infrared satellite images were expanded by rotation and were provided with additional noise. Fig. 4 shows an example of the sample expansion. The

TABLE II
NUMBER OF TRAINING AND TEST SAMPLES IN EACH CATEGORY

Data Category	Training			Validation	Test
Original/ Expanding	Original (%)	Expanding_5(%)	Expanding_3(%)	Original (%)	Original (%)
TS	6484(33.33)	6484(22.18)	6484(19.50)	647(33.93)	730(40.69)
STS	5090(20.74)	5090(17.42)	5090(15.31)	364(19.09)	383(21.35)
STY	5036(20.52)	6043(20.68)	10072(30.29)	468(24.54)	390(21.74)
VSTY	2498(11.18)	7494(25.64)	7494(22.53)	347(18.20)	226(12.60)
ViolentTY	343(1.04)	4116(14.08)	4116(12.38)	81(4.25)	65(3.62)
Overall	19451(100)	29227(100)	33256(100)	1907(100)	1794(100)

TABLE III
NUMBER OF TRAINING, VALIDATION, AND TEST SAMPLES IN TCIC MODULE

Data Category	Training		Validation	Test
Original/ Expanding	Original(%)	Expanding(%)	Original(%)	Original(%)
TS+STS	11574(54.07)	11574(34.80)	1011(53.02)	1113(62.04)
STY	5036(20.74)	10072(30.29)	468(24.54)	390(21.74)
VSTY+ ViolentTY	2841(12.22)	11610(34.91)	428(22.44)	291(16.22)
Overall	19451(100)	33256(100)	1907(100)	1794(100)

numbers of training and test samples in each intensity category are listed in Table II.

In this investigation, we also explored the direct classification of the five TC intensity grades. The original STY data were expanded from 5036 to 6043 images after being rotated 90° in order to balance the data proportions among the five TC categories, as listed in Expanding_5 of Table II.

Furthermore, all of the imagery, including duplicate imagery, was varied by rotating and adding noise, as delineated in Table II. This approach is necessary for the model to assign equal importance to all of the TC intensities during training, and is common practice in CNN training in order to prevent the model from inherently increasing its skill with the greater sampled image types at the expense of the lesser sampled image types. The testing dataset did not need to be balanced in this way.

As previously mentioned, the TCICENet model proposed in this study contains two modules: the TC intensity classification module (TCIC) and the TC intensity estimation module (TCIE). For the TCIC module, we divided the TC intensity level into three categories according to the input infrared satellite images: TS+STS, STY, and VSTY+ViolentTY. We then attempted to balance the proportions of each of the three types of typhoon data. Therefore, in the TCIC module experiment, the TS and STS data were directly combined, so that the number of TS+STS sample images was 11574. The original STY data were expanded from 5036 to 10 072 after being rotated 90°. The VSTY data were expanded from 2498 samples to 7494 by rotating 90° and 180°. After the ViolentTY data were rotated by 90°, 180°, and 270°, Gaussian noise and salt-and-pepper noise were added to the rotated ViolentTY data to expand the original 343 samples to 4116. The data before and after the expansion are listed in Table III. After the above expansion, the proportions of the expanded sample data of the three categories were basically balanced, as shown in Table III.

For the intensity estimation stage (i.e., the TCIE module), our training data are still the data after the classification expansion. In the TCIE module, three intensity estimation models are established for TS+STS, STY, and VSTY+ViolentTY, designated TCIE_Model_1, TCIE_Model_2, and TCIE_Model_3, respectively. In order to enhance the diversity of the training samples in the TS+STS category and improve the robustness of its intensity estimation performance, we expanded the training samples from the original 11 574 to 23 148 by rotating them 90°. The training samples of the three categories without considering the misclassification percent of the TCIC are listed in Table IV by no wrong proportion (NWP), and comprise 23 148, 10 072, and 11 610 images, respectively. Since the training samples of the STY and VSTY+ViolentTY intensity levels were already expanded in the TCIC module, their training samples were not expanded in the TCIE module.

In addition, in order to improve the accuracy of the TCIE intensity estimation, we considered the percent of intensity level misclassification in the three categories of the TCIC validation set in the TCIE module. Finally, the three training datasets of the intensity estimation model were expanded. In total, 3443 samples were selected from the original STY training samples in order of wind speed from low to high, along with 101 samples from the original VSTY+ViolentTY training samples in order of wind speed from low to high to join the training set of the intensity estimation model TCIE_Model_1. From the original TS+STS training samples, 3046 samples were selected in order of wind speed from high to low, and from the original VSTY+ViolentTY training samples, 2977 samples were selected in order of wind speed from low to high to join the training set of TCIE_Model_2.

TABLE IV
NUMBER OF TRAINING, VALIDATION, AND TEST SAMPLES FOR THE THREE TCIE MODELS

Data	Training		Validation	Test
NWP/WP	NWP	WP	—	—
TCIE_Model_1(TS+STS)	23148	26692	1011	1113
TCIE_Model_2(STY)	10072	16095	468	390
TCIE_Model_3(VSTY+ ViolentTY)	11610	13053	428	291
Overall	44830	55840	1907	1794

From the original TS+STS training samples, 69 samples were selected in order of wind speed from high to low, and from the original STY training samples, 1374 samples were selected in the order of wind speed from high to low to join the training set of TCIE_Model_3. When adding the wrong-proportion samples, if the raw samples were insufficient (mainly VSTY+ViolentTY), the selection was continued in the same way to expand the dataset. The training data (WP) for the three intensity estimation models in the TCICENet model were expanded to 26 692, 16 095, and 13 053 sample images, respectively (listed in Table IV). In Table IV, NWP indicates that the wrong-proportion samples were not added, whereas WP indicates that the wrong-proportion samples were added.

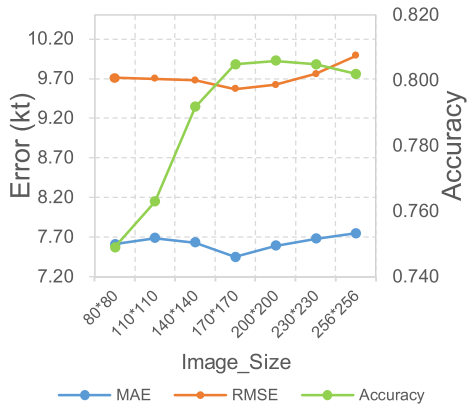


Fig. 5. Classification accuracy in the TCIC module and the intensity estimation of the MAE and RMSE in the TCIE module of the input image size.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

All of the models were trained using the TensorFlow framework (in Python 3.6), running on an NVIDIA 2080 Ti. The TCICENet model contains two modules: the TCIC and the TCIE. The experiment was divided into two stages: the training stage and the testing stage. Each stage employed the same data preprocessing method. In the data preprocessing stage, the mean value and the normalized satellite image data ranged from [-1,1] in order to improve the training speed and accuracy. Both the TCIC and TCIE were trained from scratch. The batch size of the TCIC training was 8, we used the Adadelta Optimizer (Zeiler) [46], the learning rate was 0.1, and the learning rate was reduced 0.1 times every 50 epochs. The batch size of the TCIE training was 64, we used the Adam optimizer [47] with the default parameters in the experiment, and the learning rate was 0.001.

A. Effect of Input Image Size on TCIC and TCIE Performance

We designed a two-stage experiment to investigate the effect of input image size on TC intensity classification and TC intensity estimation. In both stages, we compared the precision input images of seven sizes, including the 256 × 256 pixel input images cropped from the original satellite images, which were then reduced in size to between 230 × 230 and 80 × 80. The impacts of scaling the image size on the performance of the TCIC and TCIE modules are presented in Fig. 5.

The accuracy of TCIC classification is low when the input image size is 110 × 110 or 140 × 140, whereas the classification accuracy can reach approximately 80% for an image size of 170 × 170. The accuracy of TC intensity classification is basically the same for input image sizes of 170 × 170 and 200 × 200. On the premise of ensuring classification accuracy while reducing computation, we selected 170 × 170 as the input image size of the TCIC. During the reduction of the input image size, the MAE and RMSE of the intensity estimated by the TCIE decrease initially and then rise, reaching minimum values at 170 × 170. Therefore, input images with a size of 170 × 170 pixels were used for the TCIC and TCIE modules.

B. TC Intensity Grade Classification

As mentioned above, the existing TC intensity classification models based on deep learning are divided into multiple categories based on the TC intensity grade classification standard. According to the JMA's TC intensity classification standard, TC intensity grade is divided into a total of six categories. In this study, we only focused on TC intensity estimation for systems with an intensity level of TS and above, i.e., TC intensity estimation with five intensity levels. In order to explain the reason for our three-category classification, we designed direct five-category classification and three-category classification experiments in the TCIC module. We used the Expanding_5 training, validation, and test data presented in Table II to verify the performance of the TCIC module. Precision (P), recall (R, probability of detection), and F1-score (F1) were utilized to evaluate the performance of our model.

After the iteration of 35 epochs for the TCIC module, the validation data loss reached its minimum and the accuracy levels of the five classifications reached their maximum. Table V lists the results of the five classifications. For these classifications, the average precision, recall, and F1-score ranged from approximately

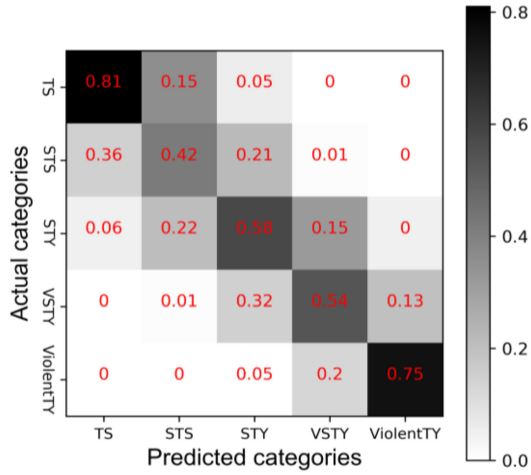
TABLE V
PRECISION, RECALL, AND F1-SCORE FOR THE FIVE TC CATEGORIES, AND THE TOTAL NUMBER OF SAMPLE TEST TC IMAGES FOR THE TCIC MODULE

Category	P	R	F1	Total Samples
TS	0.79	0.81	0.80	730
STS	0.45	0.42	0.44	383
STY	0.54	0.58	0.56	390
VSTY	0.62	0.54	0.57	226
ViolentTY	0.61	0.75	0.68	65
Avg	0.60	0.62	0.61	1794

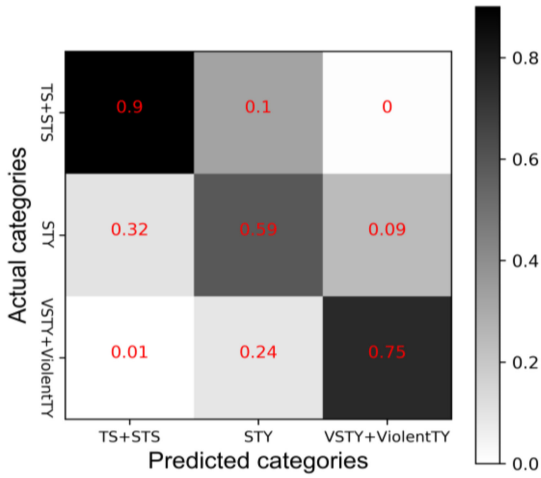
0.60 to 0.62. Thus, if the five classification results were directly applied to the subsequent classification intensity estimation, this would lead to a large intensity estimation error. Fig. 6(a) is the confusion matrix of the five classifications for the test samples. The recall levels from weak to strong intensity of the five TC categories were 0.81, 0.42, 0.58, 0.54, and 0.75, respectively. As can be seen from Fig. 6(a), TCs of all five intensity levels were misclassified. Misclassification to an adjacent intensity level was the most likely occurrence. Therefore, we combined TS and STS into one category. Although the misclassification ratio between STY and VSTY was also high, VSTY and ViolentTY also tended to be misclassified for one another. Considering that the number of ViolentTY samples is small, VSTY and ViolentTY were combined into one category, and STY was taken as a separate category. Hence, we divided the TC intensity levels into three categories: TS+STS, STY, and VSTY+ViolentTY.

We used the training, validation, and test data in Table III to verify the performance of the TCIC module for the three classifications of TC intensity grades. The iterative curves for recognition rate and loss function are presented in Fig. 7. As shown in Fig. 7(a) and (b), the total loss function value of the training set gradually decreased with the increasing number of epochs, stabilizing after 80 epochs. The accuracy of the training set gradually increased with the increasing number of epochs, increasing rapidly during the early stage with high accuracy, which is related to the fact that our TCIC module is good at extracting complex features from infrared images. After 50 epochs, the classification accuracy of the TCIC module had basically stabilized. As can be seen from Fig. 7(c) and (d), the loss function value of the validation set initially exhibited a decreasing trend and then rose slowly after 50 epochs due to overfitting. In the 31st epoch, the loss of validation data reached its minimum and the classification accuracy reached its maximum, which are denoted by red dots in Fig. 7(c) and (d). We saved the model of the 31st epoch, and the validation results of this epoch were then employed to guide the subsequent three TCIE models in order to produce the proper training samples according to the misclassification percentages among the three intensity categories.

We used the test samples listed in Table III to analyze the classification performance of the trained TCIC module. The average accuracy of the three classifications for the test set was



(a)



(b)

Fig. 6. (a) Confusion matrix of the test samples of five intensity categories for the TCIC. (b) Confusion matrix of the test samples of three intensity categories for the TCIC.

TABLE VI
PRECISION, RECALL, AND F1-SCORE FOR THE THREE TC CATEGORIES, AND THE TOTAL NUMBER OF SAMPLE TEST TC IMAGES FOR THE TCIC MODULE

Category	P	R	F1	Total Samples
TS+STS	0.89	0.90	0.89	1113
STY	0.55	0.59	0.57	390
VSTY+ViolentTY	0.86	0.75	0.80	291
Avg	0.77	0.74	0.75	1794

80.49%, which was a substantial improvement over the accuracy of the previously derived direct division into five categories.

It can be seen from Table VI that the classification indices of the TS+STS category were the highest, followed by the VSTY+ViolentTY category. The precision, recall, and F1-score of the STY category were significantly lower, indicating that it had more classification errors. The average precision, recall, and

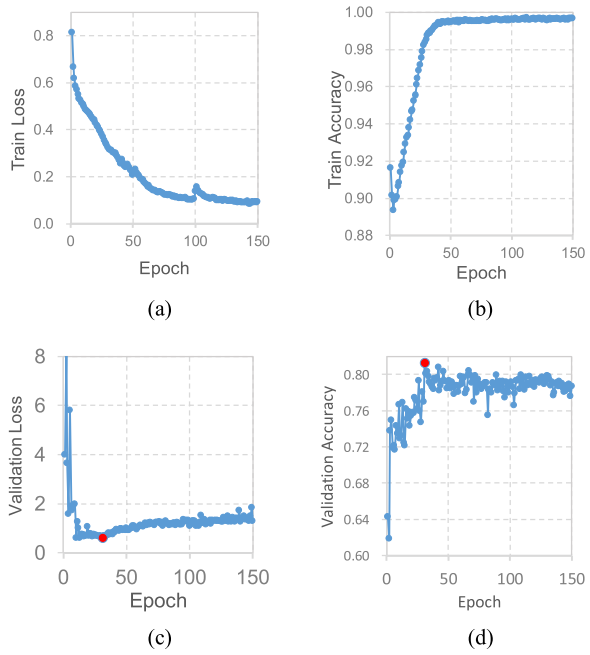


Fig. 7. Accuracy curve and loss function curve for the TCIC module. (a) Training loss function curve. (b) Training accuracy curve. (c) Validation loss function curve. (d) Validation accuracy curve.

F1-score ranged from approximately 0.74 to 0.77. Fig. 7(a) and (b) reveal that the misclassifications of the TCIC module for the three classifications of TC intensity grades were smaller. Therefore, in this study, we divided the TC intensity grades into three categories based on infrared satellite images.

C. TC Intensity Estimation

In order to illustrate the advantages and performance of the proposed TCICENet model, three other models were explored: the TCIDENet, TCICENet-NWPS, and TCICENet-S. The TCIDENet is a method of direct intensity estimation by the TCIE module without classification. The TCICENet-NWPS uses the TCIC module for classification, in which the TCIE module for intensity estimation does not include the misclassified proportion of the TCIC module. The TCICENet-S represents the TCICENet model whose intensity estimation results are smoothed with the results of adjacent satellite images prior to the current time. Among these models, the training data of the TCIDENet and TCICENet-NWPS consisted of the data in Table IV without the addition of the misclassification samples, whereas the training data of TCICENet and TCICENet-S consisted the data in Table IV with the misclassification samples added. In order to verify the performances of the different models, the RMSE, MAE, bias (the difference between the best-track data and the estimated intensity), and absolute error were employed as evaluation indicators on the test dataset. RMSE, MAE, and absolute error measure the differences between two variables, whereas bias indicates the absolute degree of overestimation and underestimation of each variable.

1) TCIDENet: All three TC intensity categories were combined to create an intensity estimation model known as the TC direct intensity estimation model (TCIDENet). The intensity estimation error of the STY category was the largest, with an MAE of 10.2 kt and an RMSE of 12.6 kt. The intensity estimations of the other four categories were not appreciably different. The TCIDENet achieved an overall MAE of 7.89 kt and an RMSE of 10.25 kt, exhibiting an intensity estimation accuracy similar to that of the models in Chen *et al.* [36] and Zhang *et al.* [26].

2) TCICENet-NWPS: The STY category was the largest, with an MAE of 11.42 kt and an RMSE of 14.15 kt. The TCICENet-NWPS achieved an overall MAE of 7.85 kt and an RMSE of 10.18 kt. The MAE and RMSE values of TS, STS, and ViolentTY for the TCICENet-NWPS were less than those of the TCIDENet, although the TCIDENet was better at estimating STY and VSTY.

3) TCICENet: In this study, we separately built three TC intensity estimation models to correspond to the three TC intensity categories, referred to as the separate models of the TCICENet. Fig. 8 shows the loss function curves of the training and validation data of the three intensity estimation models, revealing that the loss of the training set gradually converged. The validation set loss of the three TCIE models reached its minimum value at 19, 82, and 71 epochs, respectively. The TC intensity estimation results of the proposed TCICENet are listed in Table VII.

Compared with the TCIDENet and TCICENet-NWPS, the final TC intensity estimation performance of the TCICENet achieved an MAE of 7.45 kt and an RMSE of 9.59 kt, due to the addition of misclassification samples during the training of the three intensity estimation models. With the exception of no obvious improvement to STS, the intensity estimation accuracy of all the other four categories improved.

4) TCICENet-S: In order to reduce the influence of model time evolution and TC intensity noise over time, we used the weighted average of the estimated TC intensity values of the TCICENet at the current time and the previous two times as the TCICENet-S intensity estimation at the current time. Since the data used in our model were not interpolated, the previous two times were the previous 6 h and the previous 12 h. This smoothing scheme is similar or equivalent to those of some existing TC intensity estimation models [12]–[21]. In this study, the TC intensity estimation at time t was smoothed by the following formula:

$$p\ (x_t) = 0.49 f\ (x_t) + 0.29 f\ (x_{t-1}) + 0.22\ f\ (x_{t-2})\ . \quad (5)$$

5) Comparison of the Four Models: In order to facilitate the comparison of the performances of the above four models, the structure and intensity estimation evaluation indices of the models are listed in Table VIII. It can be seen from this table that the intensity estimation capabilities of the TCIDENet and TCICENet-NWPS models were basically the same. This indicates that the effect of direct TCIE intensity estimation after TCIC classification is not significantly improved compared with the direct TCIE module. The TCIDENet model is also the current model used for most TC intensity estimation based on deep

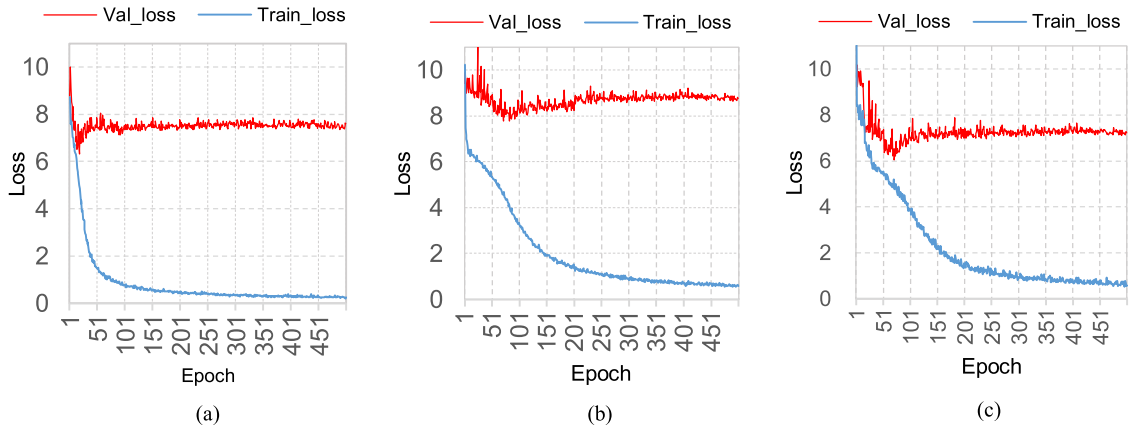


Fig. 8. Loss function curves of training and validation for each TC category using the TCICENet model. (a) TS+STS. (b) STY. (c) VSTY+ViolentTY.

TABLE VII
RMSE, MAE, AND BIAS (KT) FOR EACH TC CATEGORY, AND THE NUMBER OF TEST TC IMAGES OVERESTIMATED AND UNDERESTIMATED IN EACH CATEGORY USING THE TCICENET MODEL

Model	Category	Test samples	MAE	RMSE	Bias	Overestimated samples(%)	Underestimated samples(%)
TCICENet	TS	730	6.35	8.25	5.63	625(85.62)	105(14.38)
	STS	383	6.74	8.45	-0.05	174(45.43)	209(54.57)
	STY	390	10.35	12.83	-6.20	113(28.97)	277(71.03)
	VSTY	226	7.41	9.33	-4.66	74(32.74)	152(67.26)
	ViolentTY	65	6.85	7.87	-6.00	7(10.77)	58(89.23)
	Avg	1794	7.45	9.59	0.13	993(55.35)	801(44.65)

TABLE VIII
FOUR TC INTENSITY ESTIMATION MODELS AND THEIR COMPREHENSIVE EVALUATION INDICES

Models	TCIC	TCIE	WP	Smooth	MAE	RMSE	Overestimated samples(%)	Underestimated samples(%)
TCIDENet	no	yes	no	no	7.89	10.25	1010（56.30）	784（43.70）
TCICENet-NWPS	yes	yes	no	no	7.85	10.18	1021（56.91）	773（43.09）
TCICENet	**yes**	**yes**	**yes**	**no**	**7.45**	**9.59**	993（55.35）	801（44.65）
TCICENet-S	**yes**	**yes**	**yes**	**yes**	**6.67**	**8.60**	981（54.68）	813（45.32）

The bold entities in tables represent our proposed method.

learning [36], [38], [41]. Compared with the TCIDENet and TCICENet-NWPS, the TCICENet exhibited the best intensity estimation performance, indicating that the misclassification percent of TC intensity classification can be effectively used to improve the intensity estimation performance. As mentioned above, utilizing three categories of TCs will produce different degrees of misclassification. This phenomenon is particularly obvious for the STY category, because this category is very similar in terms of cloud structure and inner core features, such as spiral rain bands, appearing on satellite images during certain development stages of the TS+STS and VSTY+ViolentTY categories. Compared with the VSTY+ViolentTY category, the possibility of misclassifying the STY category as the TS+STS category is high. The probability of the TS category being misclassified as VSTY+ViolentTY is very small because the TS+STS and VSTY+ViolentTY categories differ greatly in intensity level, cloud pattern, and inner core characteristics. The same is true for the VSTY+ViolentTY category. After classification, the TCICENet takes into account the misclassification of the intensity classification, expands the training samples of the TCIE model according to the misclassification percent, and reduces the subsequent intensity estimation error caused by the classification error to a certain extent. Therefore, the comprehensive performance of the TCICENet intensity estimation is better than the performances of the TCIDENet and TCICENet-NWPS models.

The TC is a dynamic system, exhibiting continuous development and change over time. The current intensity of a TC is not only closely related to the cloud pattern and inner core characteristics it displays in the most recent satellite image, but is also related to the development and change of the cloud and inner core characteristics of the previous several temporally adjacent satellite images, a relationship that is not considered in the first three models. Therefore, the performance of the

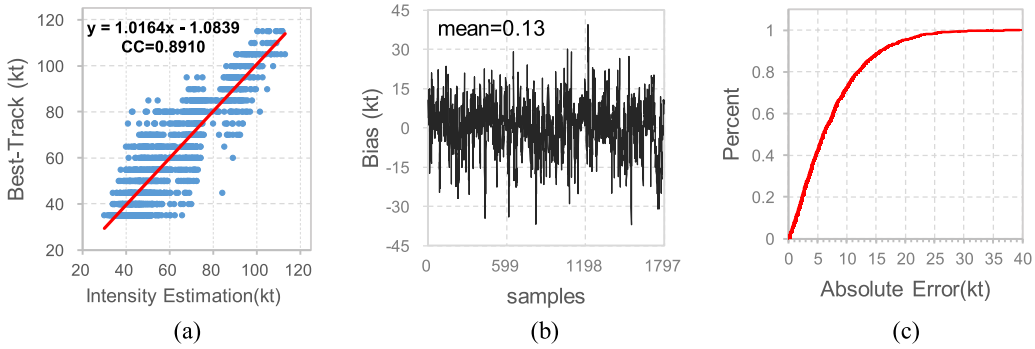


Fig. 9. (a) Scatter plot of best-track intensities versus estimated intensities using the TCICENet model. (b) Bias curve between the best track intensities and the estimated intensities using the TCICENet model for 2017–2019 TC samples. (c) Absolute error curve related to the percentage of absolute errors using the TCICENet model.

TCICENet-S model can be significantly improved when the sequential characteristics of intensity changes are considered.

V. Analysis and Discussion

A. Analysis of Error Distribution

Fig. 9(a) shows the scatter plot of the best track intensities versus the estimated intensities. The red straight line is a linear fitting line of the estimated intensities versus the best track intensities. The linear fitting equation can be expressed as $y = 1.0164x - 1.0839$. Its correlation coefficient (CC) is 0.891. The estimated intensities ranged from 30.11 to 113.15 kt. Fig. 9(b) shows the bias curve of the test samples. Based on Fig. 9(b), the difference between overestimation and underestimation was not very obvious. The average bias of the test samples was 0.13 kt. Fig. 9(c) represents the probability distribution curve of the absolute error. The percentages of absolute error between the best track intensity and the estimated intensity less than 3, 5, 10, and 15 kt were 27%, 43%, 73%, and 89%, respectively.

B. Intensity Estimation for Different TC Categories

The box plot of the bias in each TC category is presented in Fig. 10(a). The green line represents the position of the median, the black lines at both ends represent the upper and lower limits of the deviation, and the blue boxes represent the first and third quartiles of the deviation. Consistent with the MAE of each TC category, the TS category exhibited the lowest MAE, with a median of 4.82 kt (except for outliers; the same below). The lower and upper limits of the TS category bias were −8.08 and 18.57 kt, respectively. The maximum absolute error in the TS category was 39.51 kt, which was the largest of all of the test TC samples. Meanwhile, the bias of the STY category displayed the highest MAE, with a median of −6.50 kt and lower and upper limits of −37.09 and 22.66 kt, respectively. The bias of the STS category exhibited a distribution ranging from −18.72 to 21.38 kt, with a median of 0.89 kt. The bias of the VSTY category showed a distribution ranging from −28.42 to 15.53 kt, with a median of −4.88 kt. The bias of the ViolentTY category had the lowest RMSE, with a median of –6.97 kt and lower and upper limits of −14.55 and 8.79 kt, respectively. The overall bias

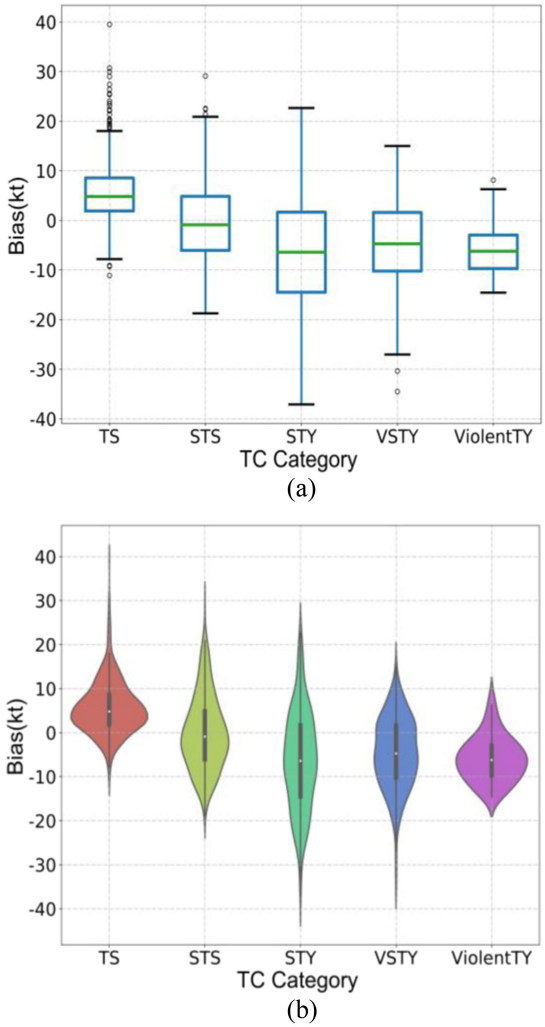


Fig. 10. Box plot and violin plot of bias in each TC category using the TCICENet model.

fluctuated around 0. The upper and lower limits of the deviation distributions of STS, STY, and VSTY were large, whereas the upper and lower limits of the deviation distributions of TS and ViolentTY were small. The median deviation of the TS category

was positive, whereas the median prediction deviations of the other categories were negative.

The violin plot of the bias of each TC category is shown in Fig. 10(b). The violin plot can help us to understand the predicted bias distribution more intuitively. The deviations of TS and ViolentTY were concentrated, with the TS deviation concentrated between 0–10 kt, and the ViolentTY deviation concentrated from 0 to –15 kt. The STS and VSTY deviations were relatively dispersed, generally ranging from –20 to 10 kt. STY was the most dispersed, with most of deviations ranging from –20 to 10 kt.

The TC intensity estimation results of the TCICENet, TCICENet-S, and the best track intensities for the 81 TCs from 2017 to 2019 are presented in Fig. 11. TCs in 2017 were estimated by the TCICENet with an MAE of 6.86 kt, an RMSE of 8.77 kt, and a bias of 0.95 kt. TCs in 2018 were estimated by the TCICENet with an MAE of 7.62 kt, an RMSE of 9.73 kt, and a bias of 0.07 kt. TCs in 2019 were estimated by the TCICENet with an MAE of 7.63 kt, an RMSE of 9.89 kt, and a bias of 0.16 kt. Overall, we can see that the intensity curves estimated by the TCICENet and TCICENet-S were basically consistent with the intensity of JMA's best track data, and the intensity estimation results of the TCICENet-S were superior to those of the TCICENet. The satellite image of the TS category was overestimated and the ViolentTY category was clearly underestimated. The estimated bias was not large, so the estimated MAE of these two categories was small. For the satellite images of STS, STY, and VSTY, the bias distributions of the intensity estimation ranges were larger, and the overestimation and underestimation were almost balanced. The TCICENet-S estimated results were slightly lower than those of the TCICENet during the TC formation stage, i.e., the stage in which the TC intensity is increasing, whereas the TCICENet-S results were slightly higher than those of the TCICENet during the TC dissipation stage. This is because the TCICENet-S estimation results are those of the smoothed TCICENet using the previous two adjacent intensities.

C. Comparison to Other Satellite Estimation Methods

In order to verify the performance of our model, it was compared with the existing TC intensity estimation models listed in Table IX. Using our IR satellite image dataset, the performance between our model and nine other models—the ETCI [22], DAVT [18], PHURIE-SVR [24], MLR [20], RVM+DADI [27], RVM+DAGCOM [26], DeepCNN [35], VGG19 [41], and Deep-PHURIE [38]—was compared.

It can be seen from Table IX that the intensity estimation results based on deep learning were generally better than the results of traditional machine learning, such as ETCI [22], DAVT [18], PHURIE-SVR[24], MLR[20], RVM+DADI [27], and RVM+DAGCOM [26]. Most TC intensity estimation models based on traditional machine learning mainly focus on extracting some statistical features from infrared satellite images, such as infrared brightness temperature gradient, deviation angle variance, or probability density of mean deviation angle (P_MDA). Although linear regression, support vector regression (SVR), and RVM exhibited good performance in TC intensity estimation, compared with the CNN models, the performance of these models still have a large amount of room for improvement. Some CNN models have been applied to TC intensity estimation. For example, a CNN-based hurricane intensity estimation network known as Deep-PHURIE [38] from IR satellite imagery can reduce the dependence on the artificial labeling of the TC center. The estimated intensity can be directly determined as the weighted average of the two highest possible categories with respect to their probabilities, a method known as the DeepCNN model [35]. In contrast, we can obtain more accurate results using our three TCIE models.

Our TC intensity model only uses infrared satellite images as the input data of the model. The datasets of the other nine models in Table IX mainly focus on microwave datasets [4], [5], [40] or combinations of various datasets (e.g., infrared, water vapor, and microwave data) [6], [15], [34], [36], [39], [48]. For example, the Hong WS algorithm used a combination of satellite-observed microwave data to estimate TC intensity [4]. The PMW-IE combined model employed both Tropical Rainfall Measuring Mission Microwave Imager (TMI) 85-GHz brightness temperatures and near-surface rain-rate retrievals to estimate TC intensity [5]. The MLR-TB-SS model used both sea surface wind speed and SSMIS microwave radiometer data to estimate TC intensity [6]. The CNN-TC model used both infrared and microwave images to estimate TC intensity [36]. The TCIENet model [34] and WIRa-based model [39] used both the water vapor channel and infrared satellite images for TC estimation.

Dvorak is the most mature TC intensity estimation technique in the world. Olander and Velden [15] further improved the ADT in combination with aircraft observation data, microwave data, and multispectral satellite images, an approach that is known as the ADT version 9. This method is used to analyze many typhoon-intensive regions in the world. Our model is only for the northwest Pacific Ocean basin, and the results are slightly better. The state-of-the-art objective method from several different satellite sources is known as Satellite Consensus (SATCON), which is used operationally at TC analysis centers worldwide (Herndon and Velden [48]). Wimmers *et al.* [40] used a CNN model known as "DeepMicroNet" to explore the possibilities of estimating TC intensity from satellite imagery of the 37 and 85–92 GHz bands. The uncertainty of TC best track information is considered when making TC classifications and calculating classification losses. Microwave images are often obscured and not very clear, so our method based on infrared satellite images can be used as a supplement to this method.

Strictly speaking, it is not fair to directly compare our model with these nine models [4]–[6], [15], [34], [36], [39], [40], [48], due to the use of different datasets. However, in order to illustrate the performance and advantages of our model, we still listed the performance indices of these nine models in Table IX (these indices were taken directly from the corresponding literature). The objective was to demonstrate that relatively satisfactory results could still be obtained using only infrared channel data combined with the CNN model to estimate TC intensity.

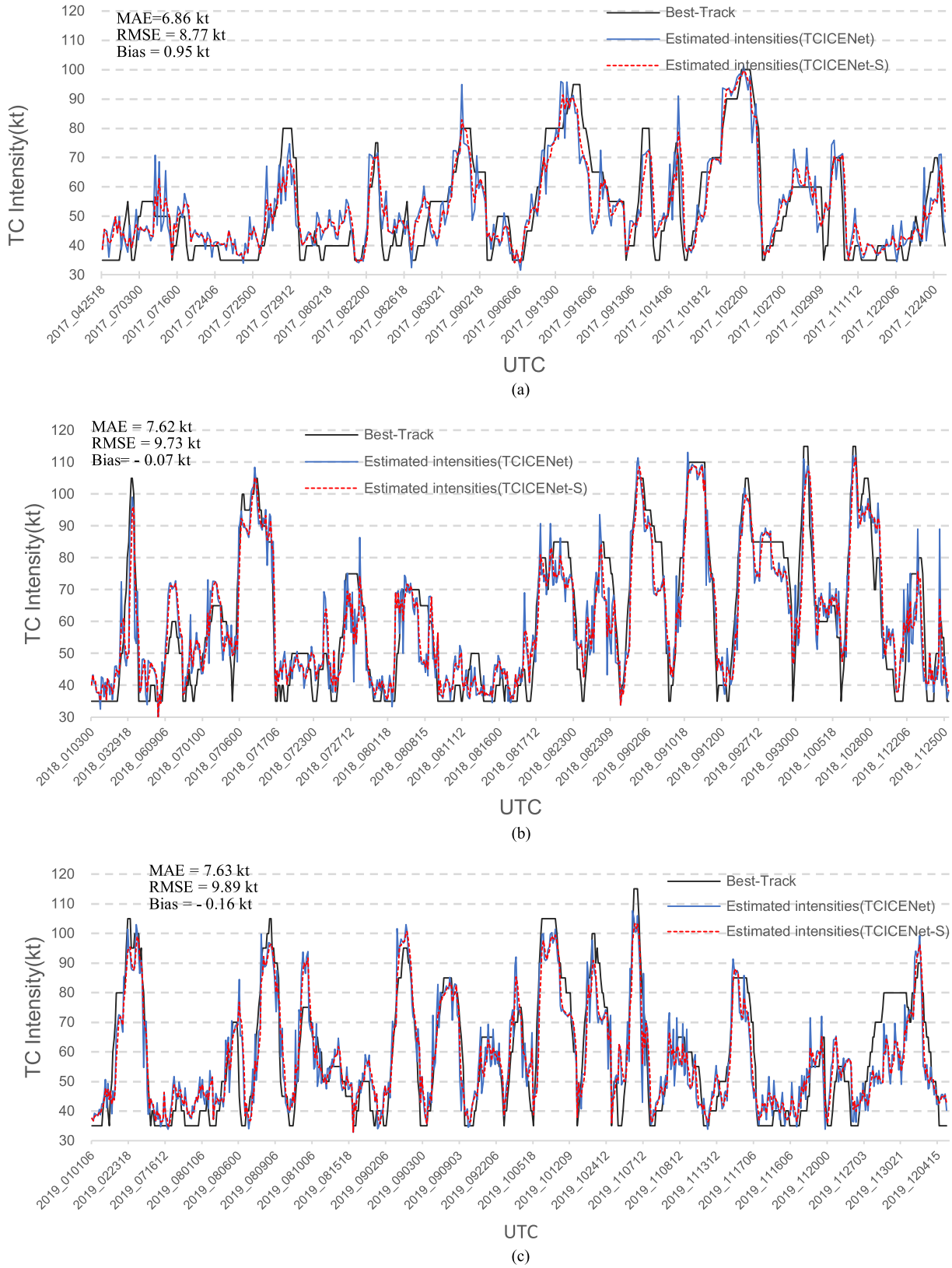


Fig. 11. TC intensity estimation results of the TCICENet, TCICENet-S, and the best-track intensities for 81 TCs from 2017 to 2019. The black solid line represents best track, the blue solid line represents the intensity curve estimated by the TCICENet, and the red dashed line represents the intensity curve estimated by the TCICENet-S.

TABLE IX
COMPARISONS OF RMSE AND MAE OF TC INTENSITY ESTIMATION OF THE TCICENET MODEL AND OTHER SATELLITE-BASED MODELS

Models	Data	RMSE (kt)	MAE(kt)	Reference
Hong Ws algorithm(relative to best track of JMA)	PMW	18.44	-	Hong et al. [4]
WIRa-based	IR, WV	14.7hPa(~17.51)	10.45hPa(~12.45)	Zhuge et al. [39]
PMW-IE combined model	TMI	12.06	8.95	Jiang et al. [5]
MLR-TB-SSW	SSMIS, SSW	11.56	8.99	Yang et al. [6]
Consensus approach (SATCON)	ADT, AMSU, SSMIS, ATMS	8.90	-	Herndon and Velden [48]
ADT-version-9(WNP)	IR, VIS, PMW	11.24	8.67	Olander et al. [15]
DeepMicroNet	MINT	10.60	-	Wimmers et al.[40]
CNN-TC(nosmoothed)	IR, PMW	10.38	-	Chen et al.[36]
CNN-TC(smoothed)	IR, PMW	8.74	-	Chen et al.[36]
TCIENet model	IR, WV	9.98	7.84	Zhang R et al. [34]
ETCI	IR	16.34	14.03	Liu et al. [22]
DAVT	IR	19.01	16.64	Ritchie et al. [18]
PHURIE-SVR	IR	18.09	15.62	Asif et al. [24]
PHURIE-SVR(smoothed)	IR	16.89	14.14	Asif et al. [24]
MLR	IR	16.21	14.88	Zhao et al. [20]
RVM+DADI	IR	18.84	16.29	Zhang et al. [27]
RVM+DAGCOM(optimal radial scale=250km)	IR	16.76	14.45	Zhang et al. [26]
Deep CNN	IR	12.95	10.12	Pradhhan et al. [35]
CNN（VGG19）	IR	10.49	8.18	Combinido et al. [41]
Deep-PHURIE	IR	10.55	8.37	Dawood and Asif. [38]
Deep-PHURIE(smoothed)	IR	8.94	7.96	Dawood and Asif. [38]
TCICENet	IR	**9.59**	**7.45**	**This study**
TCICENet-S	IR	**8.60**	**6.67**	**This study**

In summary, the RMSE and MAE values of the intensities estimated by the TCICENet model were both smaller than those of the machine learning model [4]–[6], [20], [22], [24], [27], WIRa model [39], deviation angle variance model [18], SATCON model [48], Dvorak model [15], and deep learning model [34]–[36], [38], [40], [41]. The overall performance of our model is roughly comparable to the performances of the SATCON, CNN-TC, TCIENet, and Deep-PHURIE (smoothed) models. Our model is a promising TC intensity estimation technique compared to other similar models using only infrared satellite images.

D. Error Analysis of Individual Cases

We selected three typical TCs to verify the performance of the TCICENet model.

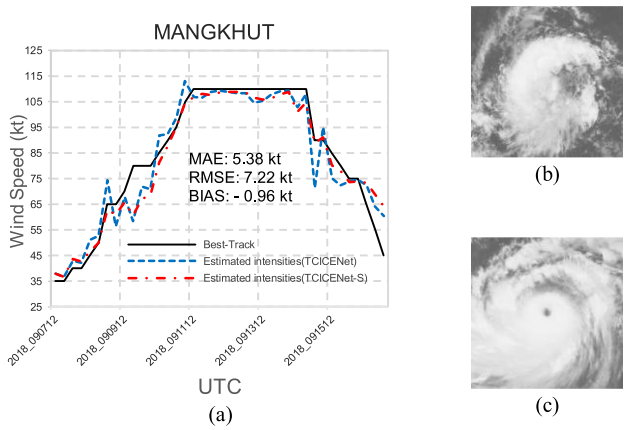


Fig. 12. (a) Time evolution of the estimated intensities and best track intensities for ViolentTY Mangkhut. (b) Mangkhut satellite image at 0600 UTC 8 September 2018. (c) Mangkhut satellite image at 1800 UTC 13 September 2018.

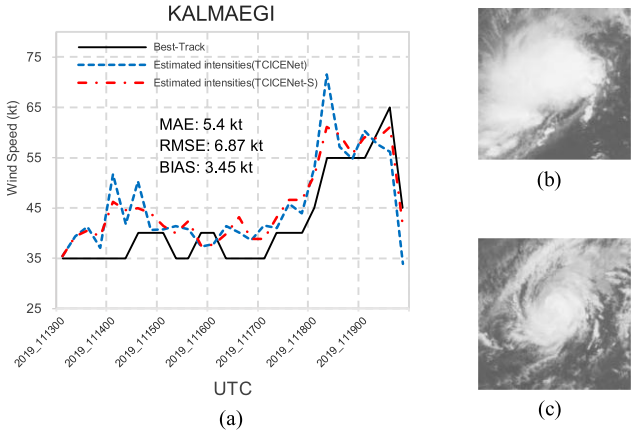


Fig. 13. (a) Time evolution of the estimated intensities and best track intensities for STY Kalmaegi. (b) Kalmaegi satellite image at 1200 UTC 14 November 2019. (c) Kalmaegi satellite image at 0600 UTC 18 November 2019.

1) Well-Performing Case: Typhoon Mangkhut in 2018 was chosen as a case to verify the performance of the TCICENet model. Its intensities estimated by the TCICENet had MAE, RMSE, and bias values of 5.38, 7.22, and –0.96 kt, respectively, as shown in Fig. 12(a). During the development period, Mangkhut's estimated intensities were near 35 kt. At this point, it was slightly overestimated, after which the wind speed increased. The period 0600 UTC 9 September–0600 UTC 10 September was continuously underestimated. From 1800 UTC 10 September–0600 UTC 14 September, the estimated intensities were almost the same as those of the best-track. Fig. 12(b) and (c) shows two infrared satellite images of Mangkhut. In Fig. 12(b) and (c), the absolute error of the wind speed estimation was $<$2.3 kt. The absolute error in Fig. 12(b) was smaller than that of Fig. 13(b). By comparison, the weaker Typhoon Kalmaegi was more likely to be overestimated. In Fig. 12(b), Mangkhut's cloud system was more scattered and its spiral structure was not obvious, which is the reason it was not easily overestimated at this stage. Fig. 12(c) shows the cloud image of Mangkhut when its best track intensity was 110 kt. At this time, the wind speed

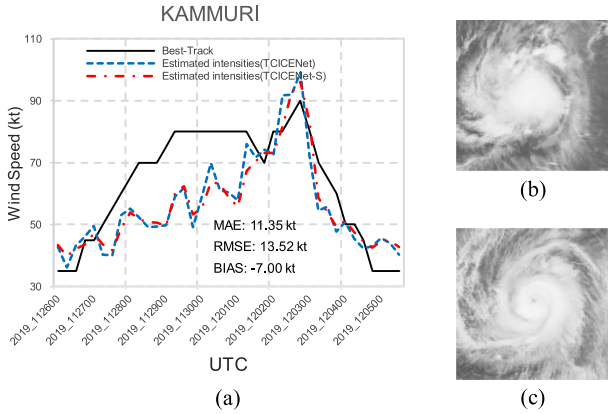


Fig. 14. (a) Time evolution of the estimated intensities and best-track intensities for VSTY Kammuri. (b) Kammuri satellite image at 0000 UTC 29 November 2019. (c) Kammuri satellite image at 0600 UTC 02 December 2019.

was high, the cloud system size was large, and the cloud structure was relatively sprawling and uniform. Our model could obtain a good intensity estimation result for the development stage shown in Fig. 12(c).

2) Overestimation Case: Typhoon Kalmaegi in 2018 was estimated by the TCICENet with an MAE of 5.40 kt, an RMSE of 6.87 kt, and a bias of 3.45 kt, as shown in Fig. 13(a). During the entire life cycle of Kalmaegi, there were a total of 28 satellite images, and 79% of the TC intensities were overestimated. In particular, the best tracks corresponding to the satellite images with wind speeds of 35 kt constituted a high proportion, totaling 12 images, and all of them were overestimated. Fig. 13(b) and (c) is the infrared satellite images at 1200 UTC November 14, 2019 and 0600 UTC November 18, 2019, respectively. Although the spiral cloud structure is not obvious, the cloud structure is compact and the brightness is high, resulting in intensity overestimation. In Fig. 13(c), the TC is more compact and the spiral structure is obvious, resulting in intensity overestimation. In addition, as can be seen from Fig. 13, our method failed to estimate the rapid intensity changes that occurred in Kalmaegi, such as a sudden weakening, resulting in overestimation.

3) Underestimation Case: It can be seen from Table VII that our TCICENet model has estimated bias of -6.20 kt and -4.66 kt for the intensity of the satellite images in STY and VSTY, respectively, indicating that the model underestimates the satellite images at these two stages. The main reason can be seen from Fig. 6 that 32% of the STYs in the test set were misclassified to a weaker category (TS+STS).

As show in Fig. 14(a), 31 satellite images of Typhoon Kammuri belonging to the STY category were misclassified as the TS+STS category. This was the main reason for the underestimation of Kammuri. For the complete life cycle of it, the RMSE was 13.52 kt, and approximately 64% of the samples were underestimated. During the TC formation stage, the wind speed gradually increased, and the wind speed was continuously underestimated from 0600 UTC November 27–1200 UTC December 1, 2019. In particular, during Kammuri's life cycle, 31% of the samples had best track wind speeds of 80 kt, of which 92%

were underestimated. In Fig. 14(b), the satellite image appears as a cloud cluster with no obvious spiral structure and no eye, which may also have caused the underestimation. In Fig. 14(c), the TC image is more compact, the spiral rain band is obvious, and the eye has begun to appear. The TC wind speed estimated by the TCICENet model also increased rapidly, as shown in Fig. 14(a), approaching that of the best track intensity.

VI. Conclusion

In this study, TCIC and TCIE cascaded CNN models were used to construct TCICENet models for TC intensity estimation. Based on the analysis of the intensity estimation error, the following conclusions were drawn.

1) There is essentially no significant difference between direct CNN intensity estimation and CNN intensity estimation after intensity classification without considering the misclassification percent. The classification accuracy of TC intensity categories had a greater impact on the final intensity estimation results. In addition, in terms of the TCICENet model, the intensity estimation error of the TCICENet-S model was smaller after applying time-weighted smoothing to the estimated intensities of the TCICENet model. This indicates that the current TC intensity is strongly correlated with previous adjacent intensities.
2) The RMSE and MAE values of the TCICENet model for the test samples were 9.59 kt and 7.45 kt, respectively. Compared with the TCICENet model, the intensity estimation results of the TCICENet-S model were more consistent with the best track data of the JMA, as indicated by the CC of 0.891. The RMSE and MAE of the test samples were 8.60 kt and 6.67 kt, respectively.

Although the proposed TCICENet model has achieved good accuracy in the estimation of TC intensity, there is still additional work to be done in order to further improve its TC intensity estimation accuracy.

1) We will further expand the training dataset in order to enhance the generalization ability and robustness of the TC intensity estimation model. All of the data used in our intensity estimation model have been from the northwest Pacific Ocean basin, and best track data have been used as the reference intensities. Subsequently, we can extend our model to other basins, such as the Atlantic Ocean, and use aircraft observations to guide, verify, and improve our model performance.
2) Next, we will attempt to combine a CNN with along short-term memory model [49], [50], which can efficiently extract the temporal features from the adjacent TC satellite images, in order to further improve the TC intensity estimation accuracy.
3) At present, we only use infrared satellite images for intensity estimation. In the future, we will consider combining other data, such as water vapor channel data and microwave data, and make full use of the advantages of various data to improve the performance of the model.
4) Currently, our intensity estimation model does not consider some physical factors, such as vertical wind shear [51], [52], ocean surface temperature [53], latitude and longitude of the TC center [36], and TC size [54], even though they may be important influencing factors for the rapid change of TC intensity. In the future, the above physical characteristics will be added to our model in order to improve its TC intensity estimation performance, and especially to provide new ideas for investigating sudden changes of TC intensity.

Acknowledgment

Part of the TCIC code refers to the code written by Kobiso (https://github.com/kobiso/CBAM-tensorflow). Code is available on the website (https://github.com/cherry-and-leaves/TCICENet). The authors would like to thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

References

[1] S. Q. Kidder *et al.*, "Satellite analysis of tropical cyclones using the advanced microwave sounding unit (ASMU)," *Bull. Amer. Meteor. Soc.*, vol. 81, no. 6, pp. 1241–1259, Jun. 2000.
[2] C. M. Kishtawal, F. Patadia, R. Singh, S. Basu, and M. S. Narayanan, "Automatic estimation of tropical cyclone intensity using multi-channel TMI data: A genetic algorithm approach," *Geophys. Res. Lett.*, vol. 32, no. 11, Jun. 2005, Art. no. L11804.
[3] N. Jaiswal and C. M. Kishtawal, "Prediction of tropical cyclogenesis using scatterometer data," *IEEE Geosci. Remote Sens. Lett.*, vol. 49, no. 12, pp. 4904–4909, Dec. 2011.
[4] S. Hong, H. J. Seo, and Y. J. Kwon, "A unique satellite-based sea surface wind speed algorithm and its application in tropical cyclone intensity analysis," *J. Atmos. Ocean. Technol.*, vol. 33, no. 7, pp. 1363–1375, Jul. 2016.
[5] H. Y. Jiang, C. Tao, and Y. X. Pei, "Estimation of tropical cyclone intensity in the north Atlantic and northeastern Pacific basins using TRMM satellite passive microwave observations," *J. Appl. Meteor.*, vol. 58, no. 2, pp. 185–197, Feb. 2019.
[6] X. Yang and K. Xiang, "Synergistic use of satellite active and passive microwave observations to estimate typhoon intensity," in *Proc. Photon. Electromagn. Res. Symp.-Spring*, Jun. 2019, pp. 1612–1617.
[7] V. F. Dvorak, "A technique for the analysis and forecasting of tropical-cyclone intensities from satellite pictures," NOAA Tech. Memorandum NESS, vol. 45, p. 19, 1973.
[8] C. A. Dvorak, "Tropical cyclone intensity analysis and forecasting from satellite imagery," *Monthly Weather Rev.*, vol. 103, no. 5, pp. 420–430, May 1975.
[9] V. F. Dvorak, "Tropical cyclone intensity analysis using satellite data," NOAA Tech. Rep., NOAA/NESDIS, Washington, DC, USA, vol. 11, p. 47, 1984.
[10] V. Dvorak and S. Wright, "Tropical cyclone intensity analysis using enhanced infrared satellite data," in *Proc. 11th Tech. Conf. Hurricanes Trop. Meteorol.*, 1977, pp. 268–273.
[11] G. Engel, "Satellite applications at the joint typhoon warning center. Rapporteur Report, Topic 0.1 e," in *Proc. 5th WMO Int. Workshop Tropical Cyclones*, 2002, vol. 1136.
[12] T. L. Olander and C. S. Velden, "The advanced Dvorak technique: Continued development of an objective scheme to estimate tropical cyclone intensity using geostationary infrared satellite imagery," *Wea. Forecasting*, vol. 22, no. 2, pp. 287–298, Apr. 2007.
[13] J. P. Kossin *et al.*, "Estimating hurricane wind structure in the absence of aircraft reconnaissance," *Weather Forecasting*, vol. 22, no. 1, pp. 89–101, Feb. 2007.
[14] X. Q. Lu, X. T. Lei, H. Yu, and B. K. Zhao, "Objective estimation of tropical cyclone intensity based on satellite data," *J. Appl. Meteorol.*, vol. 25, no. 1, pp. 52–58, Jan. 2014.
[15] T. L. Olander and C. S. Velden, "The advanced Dvorak technique (ADT) for estimating tropical cyclone intensity: Update and new capabilities," *Weather Forecasting*, vol. 34, no. 4, pp. 905–922, 2019.

[16] M. F. Pineros, E. A. Ritchie, and J. S. Tyo, "Objective measures of tropical cyclone structure and intensity change from remotely sensed infrared image data," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3574–3580, Nov. 2008.
[17] M. F. Pineros, E. A. Ritchie, and J. S. Tyo, "Estimating tropical cyclone intensity from infrared image data," *Weather Forecasting*, vol. 26, no. 5, pp. 690–698, Oct. 2011.
[18] E. A. Ritchie, K. M. Wood, O. G. Rodriguez-Herrera, M. F. Pineros, 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, Jun. 2014.
[19] C.-C. Liu, C.-Y. Liu, T.-H. Lin, and L.-D. Chen, "A satellite-derived typhoon intensity index using a deviation angle technique," *Int. J. Remote. Sens.*, vol. 36, no. 4, pp. 1216–1234, Feb. 2015.
[20] Y. Zhao, C. Zhao, R. Sun, and Z. Wang, "A multiple linear regression model for tropical cyclone intensity estimation from satellite infrared images," *Atmosphere*, vol. 7, no. 3, Mar. 2016, Art. no. 40.
[21] G. Fetanat and A. Homaifar, "Objective tropical cyclone intensity estimation using analogs of spatial features in satellite data," *Weather Forecasting*, vol. 28, no. 6, pp. 1446–1459, Dec. 2013.
[22] J. Liu, X. Xu, and X. Luo, "Estimation of tropical cyclone intensity using infrared data from a geostationary satellite," *Sci. Online Lett. Atmos.*, vol. 15, pp. 189–192, 2019.
[23] C. Kar, A. Kumar, and S. Banerjee, "Tropical cyclone intensity detection by geometric features of cyclone images and multilayer perceptron," *SN Appl. Sci.*, vol. 1, no. 9, Aug. 2019, Art. no. 1099.
[24] A. Asif *et al.*, "Phurie: Hurricane intensity estimation from infrared satellite imagery using machine learning," *Neural Comput. Appl.*, vol. 32, no. 9, pp. 4821–4834, 2020.
[25] L. J. Dai, C. J. Zhang, L. C. Xue, L. M. Ma, and X. Q. Lu, "Eyed tropical cyclone intensity objective estimation model based on infrared satellite image and relevance vector machine," *J. Remote Sens.*, vol. 22, no. 4, pp. 581–590, Jan. 2018.
[26] C. J. Zhang, Q. Luo, L. J. Dai, L. M. Ma, and X. Q. Lu, "Intensity estimation of tropical cyclones using the relevance vector machine from infrared satellite image data," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 12, no. 3, pp. 763–773, Mar. 2019.
[27] C. J. Zhang, J. F. Qian, L. M. Ma, and X. Q. Lu, "Tropical cyclone intensity estimation using RVM and DADI based on infrared brightness temperature," *Weather Forecasting*, vol. 31, no. 5, pp. 1643–1654, Oct. 2016.
[28] 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.
[29] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
[31] C. Szegedy *et al.*, ''Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Jun. 2015, pp. 1–9.
[32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2016, pp. 770–778.
[33] C. Szegedy, S. Iofffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *31st AAAI Conf. Artif. Intell.*, 2017, pp. 4278–4284.
[34] R. Zhang, Q. Liu, and R. Hang, "Tropical cyclone intensity estimation using two-branch convolutional neural network from infrared and water vapor images," *IEEE Trans. Geosci. Remote. Sens*, vol. 58, no. 1, pp. 586–597, Jan. 2019.
[35] 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. 2018.
[36] 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.
[37] G. Chen, Z. Chen, F. Zhou, X. Yu, H. Zhang, and L. Zhu, "A semisupervised deep learning framework for tropical cyclone intensity estimation," in *Proc. 10th Int. Workshop Anal. Multitemporal Remote Sens. Images*, Aug. 2019, pp. 1–4.
[38] M. Dawood *et al.*, "Deep-PHURIE: Deep learning based hurricane intensity estimation from infrared satellite imagery," *Neural Comput. Appl.*, vol. 32, no. 5, pp. 1–9, Jan. 2019.
[39] X.-Y. Zhuge, J. Guan, F. Yu, and Y. Wang, "A new satellite-based indicator for estimation of the western north Pacific tropical cyclone current intensity," *IEEE Trans. Geosci. Remote. Sens*, vol. 53, no. 10, pp. 5661–5676, Oct. 2015.
[40] 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, Jun. 2019.
[41] 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.*, Aug. 2018, pp. 1474–1480.
[42] E. A. Ritchie and K. G. Valliere, "Tropical cyclone intensity estimation in the north Atlantic basin using an improved deviation angle variance technique," *Weather Forecasting*, vol. 27, no. 5, pp. 1264–1277, Oct. 2012.
[43] S. Woo, J. Park, J.-Y. Lee, and I. So Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.
[44] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. Int. Conf. Mach. Learn.*, Jun. 2013, vol. 30, p. 3.
[45] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vision*, 2015, pp. 1440–1448.
[46] M. D. Zeiler, "ADADELTA: An adaptive learning rate method," 2012, *arXiv:1212.5701*.
[47] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[48] C. Velden and D. Herndon, "An update on the CIMSS SATellite CONsensus (SATCON) tropical cyclone intensity algorithm," in *Proc. Joint Satell. Conf.*, 2019.
[49] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
[50] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," *Adv. Neural Inf. Process. Syst.*, vol. 28, pp. 802–810, 2015.
[51] S. Bhalachandran, Z. S. Haddad, S. M. Hristova-Veleva, and F. Marks, Jr., "The relative importance of factors influencing tropical cyclone rapid intensity changes," *Geophys. Res. Lett.*, vol. 46, no. 4, pp. 2282–2292, 2019.
[52] A. Colomb, T. Kriat, and M.-D. Leroux, "On the rapid weakening of very intense Tropical Cyclone Hellen (2014)," *Monthly Weather Rev.*, vol. 147, no. 8, pp. 2717–2737, 2019.
[53] N. T. Thanh, H. D. Cuong, N. X. Hien, and C. Kieu, "Relationship between sea surface temperature and the maximum intensity of tropical cyclones affecting Vietnam's coastline," *Int. J. Climatol.*, vol. 40, no. 5, pp. 2527–2538, 2020.
[54] M. A. Bender, T. P. Marchok, C. R. Sampson, J. A. Knaffff, 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.

Chang-Jiang Zhang was born in Heilongjiang Province, China, in 1974. He received the Ph.D. degree in control theory and control engineering from Beijing Institute of Technology, Beijing, China, in 2004.

He was the Middle-Aged Discipline Leader of Zhejiang Normal University, Jinhua, China, in 2010. From 2012 to 2013, he was a visiting scholar with Cornell University, Ithaca, NY, USA. He is currently a Professor with the Electronic and Information Engineering Department, College of Physics and Electronic Information Engineering, Zhejiang Normal University. He is the Head of Research Institution of Intelligent Information Processing, Zhejiang Normal University. He is the author of three books, more than 80 articles, and more than 20 inventions. He is the Principal Investigator of more than 20 foundations and projects. His research interests include image processing, machine learning, multiscale geometry analysis, and applied meteorology.

Dr. Zhang was awarded first prize, second prize, and third prize by Education Department of Zhejiang Province, China. He was awarded the Second Level of New Century 151 Talent Project of Zhejiang Province, China.

Xiao-Jie Wang was born in China in 1996. She is currently working toward the M.S. degree in physics and electronics with Zhejiang Normal University, Jinhua, China.

Her research interests include image processing and deep learning.

Lei-Ming Ma was born in China in 1975. He received the Ph.D. degree in meteorology from Nanjing University, Nanjing, China, in 2009.

He was the Director of Numerical Forecasting Center, Shanghai Typhoon Institute of the China Meteorological Administration, Shanghai, China. He is currently a Researcher, Committee Member of Numerical Forecasting Committee, China Meteorological Society, and Vice-Manager of Shanghai Meteorological Center, China. He is the author of more than 20 articles. He is the Principal Investigator of more than ten foundations and projects. His research interests include numerical forecasting theory and methods for tropical cyclone.

Dr. Ma's awards and honors include the Science and Technology Progress Award of Shanghai City, the National Ocean Innovation Award of China, and the National Outstanding Young Meteorologist of China.

Xiao-Qin Lu was born in China in 1975. She received the M.S. degree in geographical information system from East China Normal University, Shanghai, China, in 2003.

She is currently an Associate Researcher with Shanghai Typhoon Institute of the China Meteorological Administration, Shanghai, China. Her research interests include data and information processing for tropical cyclone and geographical information systems.