

DMANet_KF: Tropical Cyclone Intensity Estimation Based on Deep Learning and Kalman Filter From Multispectral Infrared Images

Wenjun Jiang , Gang Hu , Tiantian Wu, Lingbo Liu, Bubryur Kim, Yiqing Xiao, and Zhongdong Duan

Abstract—It is very crucial to identify the intensity of tropical cyclone (TC) accurately. In this article, a novel TC intensity estimation method is proposed to estimate the TC intensity from multispectral infrared images in the Northwest Pacific Basin. A deep multisource attention network (DMANet) is proposed to model the dynamics of multispectral infrared images along the spatial dimension. We first introduce a message-passing enhancement module based on the conditional random fields to process multispectral infrared images. Multispectral data transfer the complementary information to refine the features of TC. Second, we utilize a local global attention module to make the model focus on local key features (i.e., the typhoon eye) and obtain deeper global semantic information of TC. The ablation experiment is set up in the same dataset and computing environment to verify the effectiveness of each module. Finally, we use a Kalman filter to correct the error of TC intensity during its lifetime estimated by the DMANet model. After using Kalman filter, the evolution of TC intensity becomes smooth and corresponding root-mean-square error (RMSE) decreases from 9.79 to 7.82 knots. Compared with the best result of the existing TC intensity estimation method, the RMSE of our method is reduced by 9.07%. Therefore, the proposed TC intensity estimation method shows a great potential for accurately estimating the TC intensity.

Index Terms—Attention mechanism, deep learning, intensity estimation, Kalman filter, tropical cyclone (TC).

Manuscript received 22 October 2022; revised 8 January 2023, 24 February 2023, and 24 March 2023; accepted 27 April 2023. Date of publication 8 May 2023; date of current version 17 May 2023. This work was supported in part by the National Key R&D Program of China under Grant 2021YFC3100702, in part by the National Natural Science Foundation of China under Grant 52108451, in part by Shenzhen Science and Technology Innovation Commission under Grant SGDX20210823103202018 and Grant GXWD2020123015542700320200823230021001, in part by Shenzhen Key Laboratory Launching Project under Grant ZDSYS20200810113601005, and in part by the Guangdong-Hong Kong-Macao Joint Laboratory for Data-Driven Fluid Mechanics and Engineering Applications under Grant 2020B1212030001. (Corresponding author: Gang Hu.)

Wenjun Jiang, Gang Hu, Tiantian Wu, Yiqing Xiao, and Zhongdong Duan are with the School of Civil and Environmental Engineering, Harbin Institute of Technology, Shenzhen 518055, China (e-mail: 21s154182@stu.hit.edu.cn; hugang@hit.edu.cn; wutt0824@outlook.com; xiaoyq@hit.edu.cn; duanzd@hit.edu.cn).

Lingbo Liu is with the Department of Land Surveying and Geoinformatics, The Hong Kong Polytechnic University, Hong Kong (e-mail: lingbo.liu@polyu.edu.hk).

Bubryur Kim is with the Department of Robot and Smart System Engineering, Kyungpook National University, Daegu 41566, South Korea (e-mail: brkim@knu.ac.kr).

Digital Object Identifier 10.1109/JSTARS.2023.3273232

I. INTRODUCTION

GLOBAL warming did not significantly increase the occurrence frequency of tropical cyclones (TCs) in the past few decades but had made TCs stronger [1]. The Northwest Pacific Basin is one of the most active areas of TCs. In total around, 23 TCs occur over the Northwest Pacific Basin annually, causing serious casualties and economic losses to the coastal areas [2], [3]. In order to mitigate TC-induced disasters in the coastal areas, it is important to estimate the TC intensity fast and accurately.

TCs usually form and mature on the warm ocean far away from the land. Due to the limited detection distance, most onshore weather radars cannot measure wind speed of TCs. In contrast, the meteorological satellites can stably observe TCs and obtain the satellite cloud images (SCIs) containing abundant TC feature information. Although, SCIs cannot directly reflect the TC intensity, it is very useful for estimating the TC intensity [4]. Dvorak [5] proposed a Dvorak technique to estimate the TC intensity only based on TC cloud features observed in visible light satellite images. This technique focuses on the typhoon eye area, the cloud type features of the typhoon eye wall, and the spiral rain belt features of the periphery. However, the Dvorak technique largely relies on the experience and intuition of meteorological experts. With the development of infrared imaging technology, Dvorak [6] introduced infrared satellite images to obtain the cloud top brightness temperature of TC at night, which promoted the development of Dvorak technique. In 1984, Dvorak [7] further improved the objectivity of Dvorak technique. After that, Velden et al. [8], [9], [10] and others continuously optimized the Dvorak technique and successively proposed ODT, AODT, and ADT algorithms, which further improved the accuracy of Dvorak technique and reduced its subjectivity, and achieved the automatic determination of TC intensity.

The previous studies also used machine learning to estimate the TC intensity. Statistical features and structural features were extracted from TC infrared images, which can be used as the input of machine learning models. Zhang et al. [11] extracted 15 statistical parameters from infrared images and proposed an objective technique to estimate the TC intensity using a correlation vector machine. Zhao et al. [12] extracted deviation angles and radial profiles from infrared images, then these features were used to estimate the TC intensity based on a multiple linear regression model. Dai et al. [13] established a relevance vector

machine (RVM) using the mean brightness temperature gradient of the TC eyewall with a probability of 95%. Zhang et al. [14] constructed the deviation angle co-occurrence matrix based on TC infrared geostationary satellite images, which was used to estimate the TC intensity using RVM. Xiang et al. [15] developed an intensity estimation method based on the multivariate linear regression algorithm by learning the relationship among the TC maximum sustained wind speed, the microwave brightness temperature, and the sea surface wind speed. Liu et al. [16] used the two-dimensional (2D) PCA algorithm to extract features from satellite bright temperature images, then the features were used in the k-nearest neighbor algorithm to estimate the TC intensity. Asif et al. [17] used a kernelized support vector regression to estimate the TC intensity. Lee et al. [18] developed a machine learning intensity estimation system based on the spatial and temporal features of TC satellite images.

However, the previous methods for TC intensity estimation, including the Dvorak technique and machine learning methods, usually extract TC features manually from the infrared satellite images depending on the experience and intuition of experts, leading to a certain degree of subjectivity [19], [20]. The features of TC are extracted by the deep learning model, which can avoid the subjectivity produced by extracting features manually. In recent years, deep learning has been developed rapidly [21], e.g., convolutional neural networks (CNNs) [22], recurrent neural networks [23], and generative adversarial networks [24], [25]. Due to the strong capability of deep learning, some researchers tried to apply the deep learning models to the field of remote sensing [26], [27], [28]. Combinido et al. [29] used a VGG-19 model to achieve a root-mean-square error (RMSE) of 13.23 knots, an accuracy comparable to the current feature-based intensity estimation technologies (i.e., [12], [30], [31], [32]), and pointed out that a clear typhoon eye is a sign of strong TC. Chen et al. [33] used the CNN model to estimate the TC intensity, which is trained by the satellite infrared brightness temperature and microwave rain-rate data of TCs in all basins around the world. But the model did not use the max-pooling layer to prevent the typhoon eye feature from being ignored during the learning process. Wimmers et al. [34] used 37, 85–92 GHz channels to extract TC images as the input data of the proposed model, and the corresponding RMSE result is 10.60 knots. Compared with aircraft reconnaissance observations, the RMSE result is accurate enough. Kar et al. [35] extracted the geometric features of TC images, which were used for the classification based on a multilayer perceptron. Lee et al. [36] took the superposition of infrared satellite images of multiple different channels as the input of CNN model and indicated that infrared images of different wavelengths can obtain the cloud information of different heights. The simple superposition input of TC images of different wavelengths cannot effectively take the advantage of remote sensing data. Zhang et al. [37] proposed a two-branch CNN model to estimate the TC intensity, which is trained by the infrared and water vapor images of the Northwest Pacific Basin. They used the temporal information, i.e., there is no obvious change in the TC intensity at adjacent instants. This strategy was introduced into the loss function of CNN model. Dawood et al. [38] used the publicly available dataset HURSAT-B1 to input

CNN model for the TC intensity estimation. Higa et al. [20] performed the fisheye distortion preprocessing on the satellite images to enhance the features of TCs (i.e., typhoon eye, eye wall, and cloud distribution). Combined with the knowledge in this field, the VGG-16 model is used to estimate the TC intensity category. Wang et al. [39] proposed a CNN model using the attention mechanism to achieve great results of the TC intensity estimation. Apparently, deep learning has been successfully used to estimate the TC intensity [40].

Although deep learning has made significant achievements in the TC intensity estimation, it is still a very challenging problem. First, with the rapid development of remote sensing technique, the resolution and number of remote sensing satellite data are increasing dramatically. For example, BlackSky constellation includes a total of seven remote sensing satellites with a resolution of 0.85–1.3 m per pixel, which can get submeter-level remote sensing images [41]. However, some of the above scholars only use single-channel spectral data, while others use multispectral data, but only stack different channel spectral data as the input of deep learning models. Remote sensing technology conducts large-scale detection and observation from different heights, which contains a lots of spatial information, but the advantage of remote sensing data is not exploited in the above TC estimation methods. Therefore, how to effectively use multispectral data is still a challenge. Second, most of the deep learning models repeatedly stacked convolution layers and max-pooling layers, so the feature of the typhoon eye is weakened. However, the typhoon eye is one of the most important features to characterize the TC intensity, and a full exploitation of this feature is crucial for estimating the TC intensity accurately by using deep learning model. Third, in the research of deep learning for the TC intensity estimation, many scholars smoothed the results estimated by the deep learning models to improve the accuracy of intensity estimation [19], [37]. However, the method of directly smoothing the full cycle intensity curve is inappropriate. The reason is that the smoothing operation of current TC intensity should not consider the future TC intensity.

In this study, in order to solve the above problems, a novel TC intensity estimation method is proposed, and the flowchart is shown in Fig. 1. First, a data preprocessing method is proposed for the unbalanced dataset. Second, a novel deep multisource attention network (DMANet) is proposed for the TC intensity estimation, which makes full use of advantages from two aspects: the beneficial fusion of the multispectral data and the focus on the important feature of TC. A message-passing enhancement module (MPEM) is used to capture and transmit complementary information between the multispectral data and refine the multispectral features from different subnetworks. Meanwhile, the attention mechanism is applied in a local global attention module (LGAM), so higher altitude is given to the area of typhoon eye by the local attention mechanism. The global attention mechanism helps the DMANet model to improve its overall performance and learn deeper global semantic information. Finally, each subnetwork separately estimates the TC intensity from different spectral data and adds the weights to obtain the finally TC intensity estimation of the DMANet model. Third, we use a Kalman filter for better estimation results.

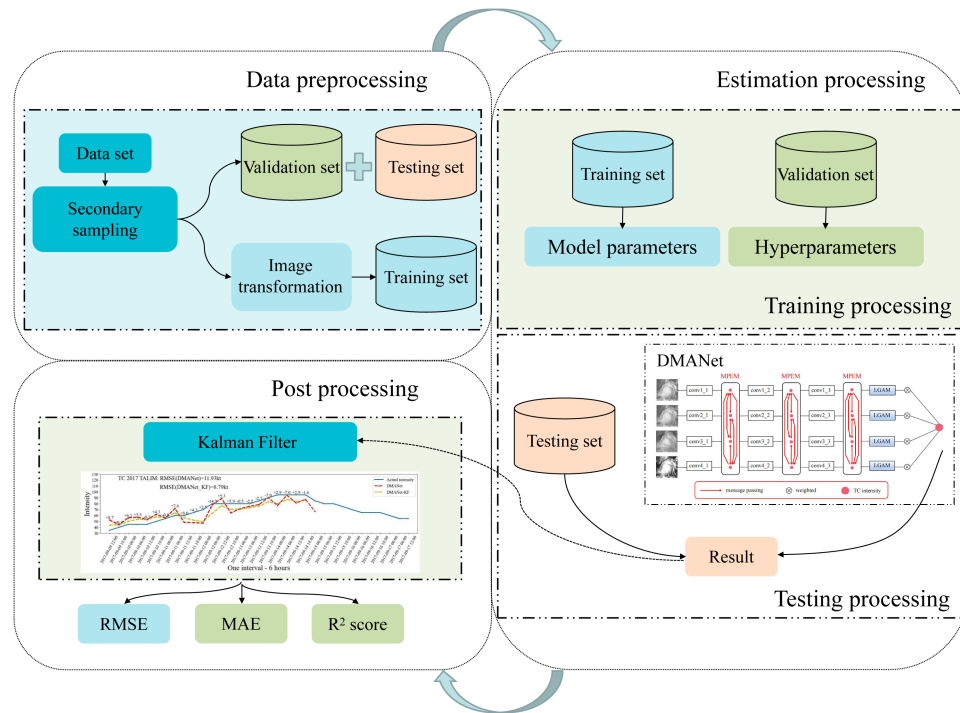


Fig. 1. Flowchart of the proposed method.

Kalman filter uses the dynamic information of TC intensity change. The state information and mean squared error information at time step of $n - 1$ can be used to recursively obtain the TC intensity estimation at time step of n , then the estimated value at time step of n can be used to correct the TC intensity estimated by the DMANet model, and better results can be obtained.

In summary, the contributions of the present study are three-fold.

- 1) We introduce an MPEM in the CNN model. The message-passing operation can be used to enhance the expressiveness of important features of TCs so as to improve the accuracy of TC intensity estimated by the model.
- 2) We propose an LGAM in the CNN model. It can make the model focus on the area with a large amount of information and enhance the ability of grasping the global feature. The robustness of the model to object variability and spatial layout can also be improved.
- 3) We use a Kalman filter to correct the error of intensity estimation by the DMANet model. To our best knowledge, Kalman filter is first used for TC intensity estimation correction, and we have verified its applicability in this study.

II. DATA OF TC

A. Data Source

The input data used for training the DMANet model come from the high-resolution infrared satellite images of four infrared channels captured by the Japanese Meteorological Satellites (i.e., MTSAT-1R, MTSAT-2, and HIMAWARI-8 Satellites) in

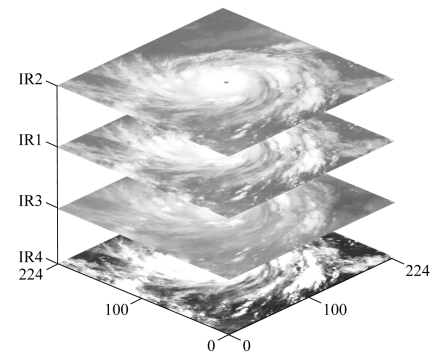


Fig. 2. Different wavelength channels contain TC information at different heights.

the Northwest Pacific Basin. The data were obtained from the National Institute of Informatics of Japan (<http://agora.ex.nii.ac.jp/digital-typhoon/>). The TC intensity data are provided in the format of an integral multiple of five knots. When the TC intensity is lower than 35 knots, it is marked as 0 knots. Each original satellite image contains 512×512 flat pixels and covers about $20^\circ \times 20^\circ$ geographical area.

The infrared channels with different wavelengths (see Table I and Fig. 2) are used by meteorological satellites to detect stratus information and convective patterns at different heights of atmosphere [36]. Infrared 1 (IR1, 10.3–11.3 μm) and infrared 2 (IR2, 11.5–12.5 μm) are widely used to detect high-level cloud information, e.g., water vapor information [42], [43]. The spectral response function of the middle wavelength shows that the troposphere contributes the most energy, so infrared 3 (IR3,

TABLE I
INSTRUMENT SPECIFICATIONS FOR WEATHER SATELLITES PROVIDING DATA

	Himawari-6 (MTSAT-1R) / Himawari-7 (MTSAT-2)	Himawari-8
Infrared 1(IR1)	10.3–11.3 μm	10.4 μm
Infrared 2(IR2)	11.5–12.5 μm	12.4 μm
Infrared 3(IR3)	6.5–7.0 μm	6.2 μm
Infrared 4(IR4)	3.5–4.0 μm	3.9 μm

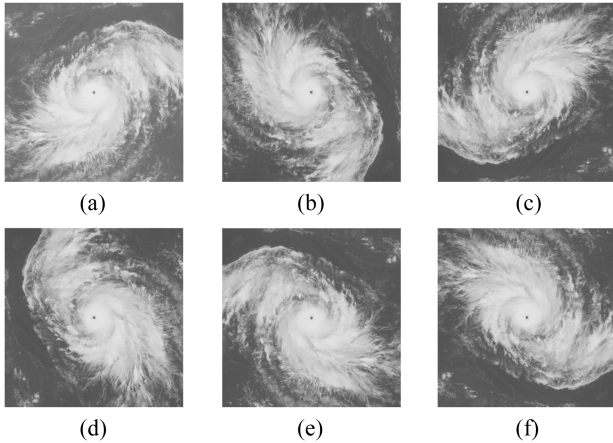


Fig. 3. Image transformation. (a) Original image. (b) Rotated 90° clockwise. (c) Rotated 180° clockwise. (d) Rotated 270° clockwise. (e) Horizontal flip. (f) Vertical flip.

6.5–7.0 μm) is more sensitive to the atmospheric composition in the middle layer [44]. Infrared 4 (IR4, 3.5–4.0 μm) is more sensitive to the droplet size change than the long wave channel in the lower altitude atmosphere. Infrared 4 is particularly useful for low cloud identification and is widely used to detect low clouds [45], [46].

B. Data Preprocessing

In order to train the deep learning network, we collected the data of four infrared channels as the input of the model. A total of 343 TCs were recorded from 2007 to 2021. The sampling frequency of satellite images is 6 h, but only about 3.3% of the TC images represent TC intensity exceed 100 knots, so the dataset is unbalanced. Unbalanced data make the model to fit primary data first, eventually leading to model overfitting [47]. In order to solve the problem of dataset imbalance, the dataset is balanced mainly by the secondary sampling and image transformation. When the TC intensity reaches 100 knots, the sampling frequency is 1 h. Even after the secondary sampling, the dataset is still unbalanced. As shown in Fig. 3, we continue to balance the dataset by the image transformation. Rotating TC infrared satellite images at different angles to expand the high-intensity data samples, we only select 90°, 180°, and 270° as the rotation angles because other rotation angles may cause the TC satellite images to lose important features of TCs, such as the spiral rain belt. TCs in the northern hemisphere rotate counterclockwise in SCIs. After the images are transformed by the horizontal

TABLE II
NUMBER OF SAMPLES BEFORE AND AFTER IMAGE TRANSFORMATION

	Original	Image transformation
Training set	6706	17 979
Validation set	1397	1397
Testing set	832	832
Total	8935	20 208

and vertical flip, the TC rotation direction becomes clockwise, which can enhance the generalization ability of the network. For samples with intensity of 120 knots and 125 knots, not only five image transformation methods will be used but also the number of samples will be increased by copying.

After the second sampling operation, the dataset is divided into the training set, validation set, and testing set. Under the same TC, the intensity of TC and the cloud distribution of SCIs in the adjacent time have insignificant change. If the dataset of SCIs for all TCs is mixed and simply divided by percentage, the data information will be leaked out. Specifically, the training set, validation set, and testing set may all include SCIs of one TC. In order to prevent the data information leakage, the SCI samples of 36 TCs were randomly selected as the validation set and the other 36 TCs as the testing set. More details of dividing the dataset are given in Table II. To eliminate the influence of unnecessary features (e.g., disordered clouds) in the SCIs on the performance of the network, the original image with a size of 512×512 is cropped to 400×400 with remaining the center of the image, and then is compressed to 224×224 . The selection of cropping size is illustrated in Section V-A1. Meanwhile, all pixel values are normalized as the input of the deep learning model in this study.

III. METHODOLOGY

A. DMANet Overview

In the TC intensity estimation, the CNN model is usually used to estimate the TC intensity. Lee et al. [36] used four infrared satellite images with different wavelengths, which are stacked into four channel images as the input of CNN model. In this way, the complementary information between different spectral channels cannot be fully captured. To overcome this weakness, an MPEM is introduced in the present study. The module fuses multispectral data and passes the complementary information between different spectral data. The MPEM is described in detail in Section III-B. In addition, inspired by Higa et al. [20], we

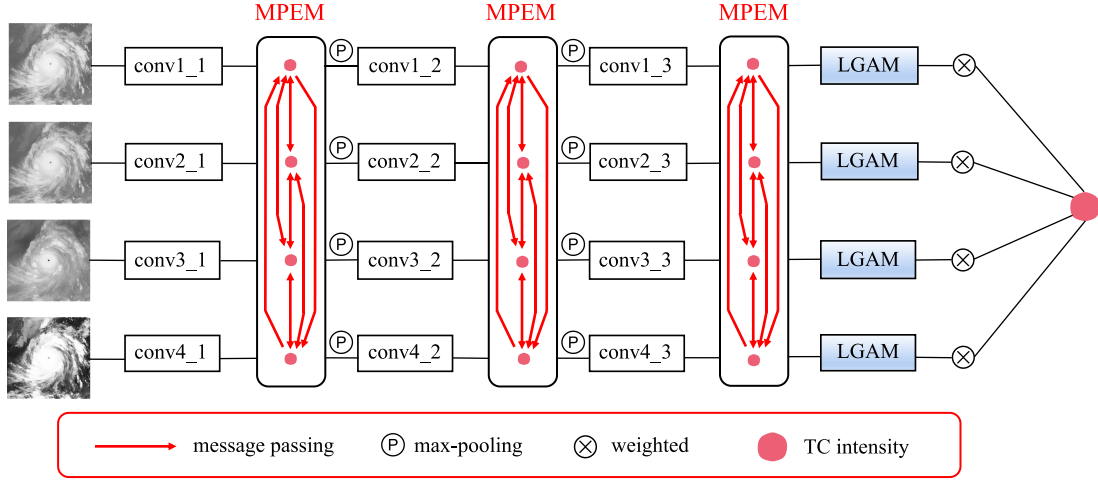


Fig. 4. Overall framework of the proposed DMANet.

realize that the clarity of the typhoon eye and the overall shape of the TC are two important features to estimate the TC intensity. We propose the LGAM to enable model to focus on the feature of the typhoon eye, surrounding clouds and overall shape of TC. The current TC intensity is assessed by the network from these above features. The LGAM is described in detail in Section III-C. The proposed DMANet) is mainly composed of the MPEM and LGAM.

To ensure a fair comparison with different models that use RGB images, we convert grayscale images to RGB during the preprocessing stage. As shown in Fig. 4, in the preceding steps of LGAM, each subnetwork is repeatedly stacked by convolution layers and max-pooling layers. The corresponding dimensions are $\text{conv}i_1$ [32, 3, 5, 0] (input 3 dimensions, output 32 dimensions, 5×5 convolution kernel, 0 padding, i corresponds to the subnetwork), $\text{max-pooling}i_1$ [4, 2] (4×4 window size, 2 stripes), $\text{conv}i_2$ [64, 32, 3, 0], $\text{max-pooling}i_2$ [3, 2], and $\text{conv}i_3$ [64, 64, 3, 0]. The feature map after each convolution layer corresponds to different dimensions ($32 \times 220 \times 220$, $64 \times 107 \times 107$, $64 \times 51 \times 51$), so three groups of different scale features are formed. Different feature maps in each group use the MPEM for message passing. The feature map generated by the last MPEM of each subnetwork is input to the LGAM, respectively. Since the feature maps of each group correspond to different dimensions, the enhanced feature maps contain deeper semantic information about the features of TCs.

In each subnetwork, the final output value P_i of feature map is multiplied by a certain weight α_i , and their summation is used to estimate the TC intensity. This process can be expressed as the following formula:

$$kt = \sum_{i=1}^4 P_i \alpha_i \quad (1)$$

where kt is the TC intensity estimation of the DMANet model; P_i is the final output value in each subnetwork; and α_i is the weight corresponding to P_i .

B. MPEM

The advanced meteorological satellite technology provides multispectral infrared satellite images. How to make full use of multispectral data is a challenge, so an MPEM is proposed. The MPEM refines features on the feature map generated by infrared images of different wavelengths by adequately exploring the complementarity of multispectral data based on conditional random fields (CRFs) [48]. Each feature map can transfer its own useful messages to other feature maps and receive useful messages from other feature maps simultaneously.

Different spectral data can represent the cloud information at different heights of TCs. The MPEM is used to dynamically fuse the multispectral data to improve the accuracy of TC intensity estimated by the DMANet model. The feature maps generated by different subnetwork are denoted as $X = \{x_1, x_2, \dots, x_i\}$, while the output feature maps obtained after message passing are represented by $\hat{X} = \{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_i\}$. CRFs are used to model the conditional distribution of X and \hat{X} , and the condition distribution is defined as follows:

$$P(\hat{X}|X, \theta) = \frac{1}{Z(X, \theta)} \exp(\Phi(\hat{X}, X, \theta)) \quad (2)$$

where $Z(X, \theta)$ is the normalization constant; $\Phi(\hat{X}, X, \theta)$ is the energy function; θ is the set of parameters; $Z(X, \theta)$ is defined as follows:

$$Z(X, \theta) = \int_{\hat{X}} \exp(\Phi(\hat{X}, X, \theta)) d\hat{X} \quad (3)$$

$\Phi(\hat{X}, X, \theta)$ is consist of a unary potential $\phi(\hat{x}_i, x_i)$ and a pairwise potential $\psi(\hat{x}_i, \hat{x}_j)$. It is defined as follows:

$$\Phi(\hat{X}, X, \theta) = \sum_i \phi(\hat{x}_i, x_i) + \sum_{i,j} \psi(\hat{x}_i, \hat{x}_j). \quad (4)$$

The unary potential $\phi(\hat{X}_i, X_i)$ is used to describe the similarity between the original feature maps and the feature maps after receiving complementary information, which is defined as follows:

$$\phi(\hat{x}_i, x_i) = -\frac{1}{2} \|\hat{x}_i - x_i\|^2. \quad (5)$$

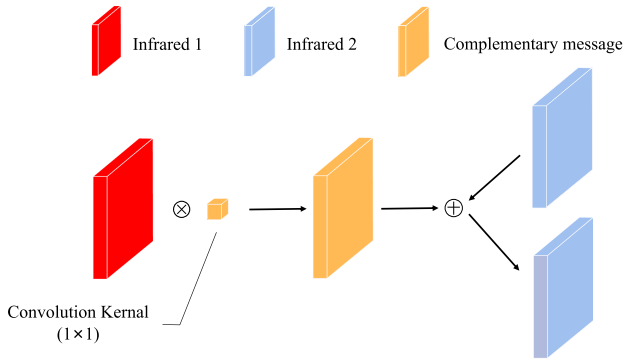


Fig. 5. Overall framework of the proposed MPEM. Taking the information transmitted from infrared 1 to infrared 2 as an example. The implementation of MPEM is briefly described. The feature map of infrared 1 passes the complementary information to another through a 1×1 convolution layer.

The pairwise potential $\psi(\hat{x}_i, \hat{x}_j)$ is used to describe the correlation between enhanced features, which is defined as follows:

$$\psi(\hat{x}_i, \hat{x}_j) = (\hat{x}_i)^T w_j^i \hat{x}_j \quad (6)$$

where w_j^i is the learned parameter.

The basic formula of CRFs is given above. The message-passing formula is given as follows:

$$\hat{x}_i = x_i + \sum_{i \neq j} w_{ij} \hat{x}_j \quad (7)$$

where \hat{x}_i represents x_i receiving feature maps from other spectral data to form enhancement feature maps, and it continues to propagate along the subnetwork; w_{ij} is a weighting factor that controls the complementary information passed from x_j to x_i . Due to the interdependence between \hat{x}_i and \hat{x}_j , the refined \hat{x}_i is obtained by the following formula:

$$\begin{cases} h_i^0 = x_i \\ h_i^t = x_i + \sum_{i \neq j} w_{ij} \hat{h}_j^{t-1}, t = 1 \text{ to } n \\ \hat{x}_i = h_i^n \end{cases} \quad (8)$$

where n is the number of iterations. Equation (8) is very simple to be implemented in a CNN model. The passing of message from x_j to x_i can be achieved by applying a 1×1 convolution kernel (as shown in Fig. 5). w_{ij} is the learning parameter of the convolution layer.

As shown in Fig. 4, we set up the MPEM after each of the first three convolution layers, but simply input the multispectral data into the model, which cannot fully capture the complementary information between the multispectral data. The MPEM can be used to fuse multispectral data and then fully mine the connection between the data and the TC intensity.

C. Local Global Attention Module

The typhoon eye is an important feature for the estimation of TC intensity by using the Dvorak technique [5]. Chen et al. [33] did not set the max-pooling layer in the CNN because the repeated pooling operation will reduce the resolution of typhoon eye. Higa et al. [20] enhanced the expressiveness of typhoon

eye through a preprocessing. Apparently, the typhoon eye is an important feature for making a correct estimation of the TC intensity. Inspired by Li et al. [49], [50], the attention mechanism is used to make the model to focus on the discriminant features. Meanwhile, the global information is also important. So, the LGAM includes the local attention module (LAM) and global attention module (GAM) to learn the depth and global representation and capture complementary information between the local and global. The proposed LGAM architecture is illustrated in Fig. 6.

1) *Patch Generation*: The default location of typhoon eye on the SCIs is not far from the image center or in the center. The dimension of the enhanced feature map passed by each subnetwork through MPEM is $64 \times 51 \times 51$. The central clipping operation is carried out on this feature map. The clipping dimension is $64 \times 5 \times 5$, which basically covers the size of the typhoon eye. The center cutting operation is shown in Fig. 7.

2) *LAM*: As shown in Fig. 6, the patch generation is the beginning of LAM, and the weighting vector \hat{v}_l is formed in the end of LAM. The LAM aims to focus on the local representative patch. The patch we get is operated on the feature map, so we can get a larger receptive field. The LAM focuses on the identification of local feature, i.e., the typhoon eye patch. In the LAM, we do not add the max-pooling layer because the max-pooling layer will cause the loss of high gradient information at the area of typhoon eye, and the effectiveness of attention mechanism will be reduced. For the clipped local patch, we only use the convolution layer.

Although the max-pooling layer has been shown to have a negative impact on the accuracy of TC intensity estimation [51], DMANet without this layer results in an excessive consumption of computing resources. As a result, the max-pooling layer is only removed in LAM. The dimension of input feature map f_l is $64 \times 5 \times 5$. After the feature map passes through a convolution layer, the dimension of feature map becomes $64 \times 3 \times 3$, then a double branch is connected behind the feature map f_l . The first one is directly paved into a 1-D vector v_l , and the other is connected to the attention model. The main function of attention model is to generate a weight associated with v_l . The main process of attention model is described as follows: after the feature map passes through a convolution layer, the output dimension is $64 \times 1 \times 1$, then the feature map is flattened into a 64-D vector. The vector passes through the multiple full connection layers, then a weight α_l is output using the Sigmoid function. The output is limited to a value between 0 and 1 by the Sigmoid function, which represents the importance of this local patch for TC intensity estimation. Specifically, 0 represents an unrelated patch and 1 represents a particularly important patch

$$\alpha_l = w_l(f_l) \quad (9)$$

where α_l is the weight generated by the attention model; w_l is the operations of attention model; f_l is the input vector. The output of LAM follows the formula:

$$\hat{v}_l = v_l \alpha_l \quad (10)$$

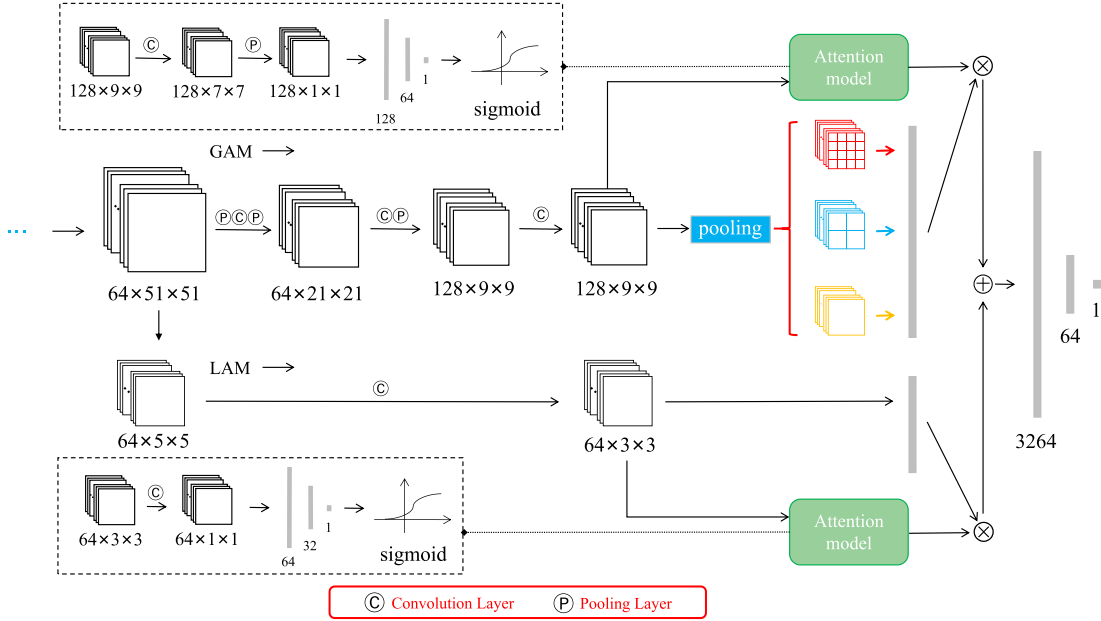


Fig. 6. Detailed illustration of the proposed LGAM.

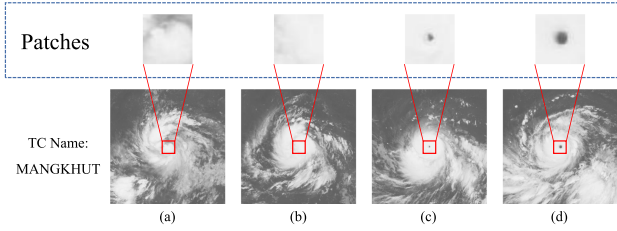


Fig. 7. Interception of the patches (typhoon eye). (a) 35 knots. (b) 65 knots. (c) 95 knots. (d) 110 knots. Taking the TC MANGKHUT as an example. The infrared satellite images (Infrared 1) of TCs with different intensities were selected in turn from the development process of the TC. During the development process of the TC, the cloud amount continued to accumulate, and then the typhoon eye appeared and became clearer as the intensity increased.

where v_l is a vector that represents the unweighted feature of local patch; α_l is the weight of v_l ; \hat{v}_l is the vector that represents the weighted feature of local patch.

3) *GAM*: This section is to describe the GAM. As mentioned above, the LAM can automatically learn the important features of TCs through the attention model. However, the typhoon eye is only a part of the important features of TCs, and the overall TCs also contain many important TC features, e.g., spiral rain belt and global semantic information. The GAM is introduced to improve the overall network performance and help the network to learn deep global semantic information. As shown in Fig. 6, the GAM is mainly composed of two branches, repeated convolution layers and max-pooling layers.

The input feature map is defined as f_g , the repeated convolution layers and max-pooling layers before input to the two branches are defined as w_g , then the dimension of feature map \hat{f}_g becomes $128 \times 9 \times 9$. This process is described as follows:

$$\hat{f}_g = w_g f_g. \quad (11)$$

The feature map \hat{f}_g is feed into the two branches. The first branch is connected to the spatial pyramid pooling (SPP). The SPP is mainly used to solve the problem of inconsistent input data images and improve the robustness of CNN model. After the SPP, the same number of vectors will be formed into the linear layer. In the DMANet model, we use the SPP to enable this method to capture spatial feature information of different sizes, which can improve the robustness of the model to object variability and spatial layout, and prevent the model overfitting. The feature map becomes a 1-D global vector through the SPP. The 1-D unweighted vector is defined as v_g .

The second branch is the attention model. The attention model is composed of a convolution layer, a max-pooling layer, three full connection layers, and a Sigmoid function. The weight of output is the global attention weight α_g . The last output vector of GAM is represented by the following formula:

$$\hat{v}_g = \alpha_g v_g \quad (12)$$

where \hat{v}_g is the vector that represents the weighted feature of global; v_g is the vector that represents the unweighted feature of global; α_g is the weight of v_g .

D. Evaluation Indicators

The mean absolute error (MAE), RMSE, and R^2 are commonly used as evaluation indicators to evaluate the effect of regression models. Their formulae are shown as follows:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^n |Y - \hat{Y}| \quad (13)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^n (Y - \hat{Y})^2} \quad (14)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y - \hat{Y})^2}{\sum_{i=1}^n (Y - \bar{Y})^2} \quad (15)$$

where N is the total number of samples; Y is the real TC intensity; \hat{Y} is the intensity of the TC estimated by the model; \bar{Y} is the mathematical expectation of Y ; MAE is the mean of absolute errors between the predicted and observed values; RMSE is the sample standard deviation of the difference between the predicted and observed values (i.e., residuals.)

E. Experimental Designs

The DMANet model contains several components, including the MPEM and LGAM. In order to demonstrate the superiority of DMANet model of the testing set, we set up an ablation experiment to verify the effectiveness of each module. The compared networks include: First, DANet-4: in the DMANet, the MPEM is removed. Second, DANet-1: the model only uses the LGAM and infrared 1 satellite images as data sources, i.e., only one subnetwork in DMANet is taken. Third, DNet: in the DANet-1, the LGAM is removed. All networks operate in the same environment, and the parameter sets are the same. In addition, we set an experiment to compare the accuracy of different cropping sizes of input images, including the 450×450 image, the 400×400 image, and the 350×350 image.

Our models are implemented with PyTorch. The training process runs on a GeForce RTX 3090 24 G GPU. We choose MSE as the loss function of the network, optimizing our network with Adam and a learning rate of $2e-4$ by minimizing the MSE loss, and the batch-size is 128. We use training set and validation set to determine the parameters and hyperparameters of our model. After determining the parameters and hyperparameters, we mix the training set and the validation set together for training. The epoch is set as 300, and the last epoch parameter is used in the testing set. The network is trained three times for testing, and the lowest loss is selected as the testing data result.

IV. KALMAN FILTER

A. Background of Kalman Filter

Kalman filter is an algorithm that uses the linear system state equation to estimate the system optimally through the system input and output data with Gaussian noise [52]. Because the data include the influence of noise and interference, the optimal estimation can be regarded as a filtering process. Even if the noise is non-Gaussian, Kalman filter still has more advantages than the other filters. Kalman filter is a data processing method to remove noise and restore real data, which is suitable to linear, discrete, and finite dimensional systems. It can estimate the state of dynamic systems in a series of data with errors [53]. Because the Kalman filter is easy to program and has strong applicability, it is the most widely used filter at present and has been widely used in communication, data assimilation, and other industries [54], [55].

B. Kalman Filter

In the present study, the Kalman filter is used to deal with the time-series problem. To avoid disclosing the future TC intensity information, when Kalman filter is adopted to determine the current TC intensity information, we use the current and previous TC intensity estimated by the DMANet model. The TC intensity change is continuous. Assuming that the intensity change process is a linear process, we use the original Kalman filter to smooth the TC intensity change process. An important point for the Kalman filter is to determine the initial value, which is sampled by a normal distribution. Considering that the TC intensity is a development process and slowly absorbs energy from the marine environment to enhance its intensity, the TC cannot suddenly change from a tropical depression to a TC with a certain intensity. Meanwhile, the minimum wind speed of TCs is 35 knots. Therefore, the mean value and variance of initial value are set to 35 knots and 0.2 knots, respectively.

Kalman filter is mainly proposed for linear systems. Here, we will introduce the basic principle of Kalman filter. The state and observation equations are to infer the state of the current moment based on the state and control variables of the previous moment, and the equations have the following expression:

$$\begin{cases} x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1} \\ z_k = Hx_k + v_k \end{cases} \quad (16)$$

where x_k is the state vector at the current moment; x_{k-1} is the transition vector at the previous moment; z_k is the measured value at time k , this study considers the output value of the model at time k as the measured value; A represents the system parameters; B is the matrix that transforms the input into states, and set to 0; H represents the parameters of the measuring system, and set to 1; u_{k-1} is the amount of control over the system at time $k-1$; w_{k-1} and v_k are the Gaussian noise of the prediction process, they are white Gaussian noise with expectation 0 and covariance Q and R as follows:

$$\begin{cases} w_{k-1} \sim N(0, Q) \\ v_k \sim N(0, R). \end{cases} \quad (17)$$

Kalman filter process and five basic formulae are given as follows:

$$\begin{cases} \hat{x}_k^- = A\hat{x}_{k-1} + Bu_{k-1} \\ P_k^- = AP_{k-1}A^T + Q \\ \hat{x}_k = \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-) \\ K_k = P_k^- H^T (HP_k^- H^T + R)^{-1} \\ P_k = (I - K_k H)P_k^- \end{cases} \quad (18)$$

where \hat{x}_k^- is the prior state estimate at time k ; \hat{x}_{k-1} is the optimal estimate at time $k-1$; P_k^- is the *a priori* estimated covariance at time k ; P_{k-1} is the posterior estimated covariance at time $k-1$; \hat{x}_k is also called the optimal estimate; K_k is the filter gain matrix, which is the intermediate calculation result of filtering, also called the Kalman gain or Kalman coefficient; P_k is the optimal estimate.

TABLE III
PERFORMANCE COMPARISON ON DIFFERENT MODELS

Model	RMSE (knots)	MAE (knots)	Parameters (MB)
DNet	11.98	9.28	2.11
DANet-1	10.60	8.25	3.18
DANet-4	10.50	8.16	12.72
DMANet	9.79	7.52	12.74

C. Kalman Filter Parameter Settings

In this article, the Kalman filter method is used to deal with the 1-D time series of TC intensity. Different parameter settings in Kalman filter will play a decisive role in the correction of the model. The observation covariance is set to 0.2, and the transition covariance is set to 0.1, corresponding to R and Q in (17), respectively. The transition matrix is set to 1, it is A in (16).

V. RESULTS

A. Model Performance

1) *Ablation Experiment in Testing Set*: First, the effectiveness of MPEM and LGAM is discussed. The evaluation indicators, such as RMSE and MAE, are used to evaluate the performance of these models. As shown in Table III, the DNet exhibits the worst performance on the testing set due to the lack of MPEM and LGAM. Compared with the DNet, the DANet-1 applies the LGAM to estimate the TC intensity. The reason is that the DANet-1 relies on local important feature and captures deeper global semantic information. As a result, the DANet-1 achieves better performance. Specifically, the RMSE is reduced by 10.6%, and the MAE is reduced by 8.8%. Compared with the DANet-1, the DANet-4 includes multispectral data, which leads to a reduction of 0.94% in the RMSE and 1.09% in the MAE. Apparently, the introduction of multispectral data can indeed enhance the network to extract more features related to TC intensity. Meanwhile, the MPEM in DMANet fully explores the complementarity between multispectral data and enhances the features with CRFs; thus, the DMANet has a significant improvement in performance. Compared with the above three models, the DMANet is the optimal one. Compared with the DNet, the RMSE and MAE of DMANet decrease by 18.3% and 19.0%, respectively. These comparisons demonstrate the effectiveness of each module of the DMANet.

Fig. 8 shows scatter plots of the actual value of TC intensity versus the estimation of network using different models. The estimated results of all models share a common feature, i.e., low-intensity samples are more overestimated and high intensity are more underestimated. In contrast, each module can reduce the bias and make the estimated intensity closer to the black dotted line.

Table IV presents the performance comparison on different cropping sizes. The only difference is the cropping sizes of input images, and the performances of different cropping sizes are compared to find a suitable size. When cropping size is equal to 350×350 , the RMSE and MAE are lower than that of 400×400 .

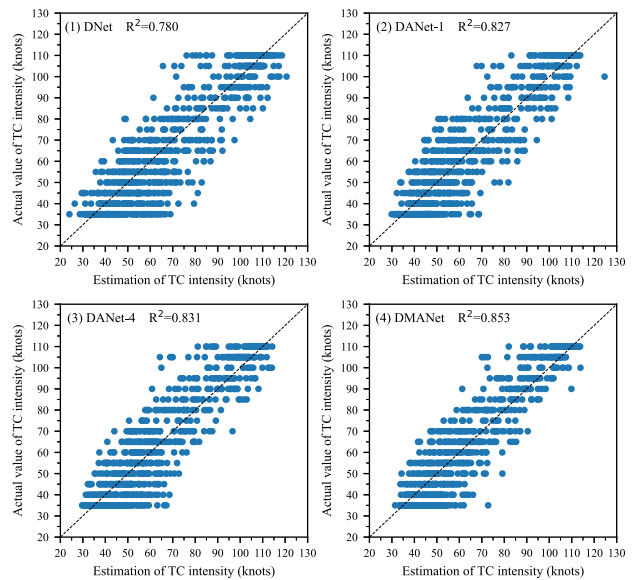


Fig. 8. Scatter plots of best-track intensity versus estimated intensity using different models.

TABLE IV
PERFORMANCE COMPARISON ON DIFFERENT CROPPING SIZES

Model	RMSE (knots)	MAE (knots)
350	10.34	8.02
400	9.79	7.52
450	10.15	7.81

The reason is that this cropping size crops out some important features of TC, resulting in reduced performance. In addition, the RMSE and MAE of 450×450 are also lower than that of 400×400 , the reason is that the surrounding unclipped clutter of clouds affects the power of DMANet model.

2) *Ablation Experiment of Different TC Categories*: Table V presents the comparison of different models under different TC categories. The definition of maximum sustained wind speed in Table V comes from Japan Meteorological Agency (<https://www.jma.go.jp/jma/kishou/now/typhoon/1-3.html>). Except for the RMSE in violent category, the results of ablation experiment show that the DMANet achieves the best performance compared with other models. The MPEM and LGAM of DMANet can complement each other to achieve a significant improvement. For example, in the very strong category, the RMSE and MAE of DANet-4 (MPEM) and DANet-1 (LGAM) are close to the performance of DNet, so the effect of MPEM and LGAM has not been fully exerted. Compared with the DNet, the DMANet mixes the two modules to obtain significant improvement of 15.1% in RMSE and 19.2% in MAE. Section II-A has pointed out that multispectral data represent the TC cloud information and convective patterns at different heights. The MPEM transmits the complementary message between multispectral data to enhance feature maps. Meanwhile, the LGAM directly performs the attention operation on the feature maps outputted by the

TABLE V
INTENSITY AND THE NUMBER OF SAMPLES IN THE TESTING SET FOR EACH TC CATEGORY AND THE PERFORMANCE OF ALL MODELS IN EACH CATEGORY

Category	Wind speed	Sample size	DMANet		DMNet-4		DANet-1		DNet	
			RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Normal (N)	35-64	439	9.41	7.32	9.77	7.70	9.97	7.80	12.46	9.83
Strong (S)	65-84	154	10.99	9.16	11.73	9.95	13.03	10.75	12.42	10.23
Very strong (VS)	85-104	98	9.36	7.10	11.24	8.90	10.32	8.29	11.02	8.79
Violent (V)	105-	132	9.85	6.61	10.76	7.02	9.68	6.85	10.37	6.69

The bold entities indicate the best performance of the comparison methods.

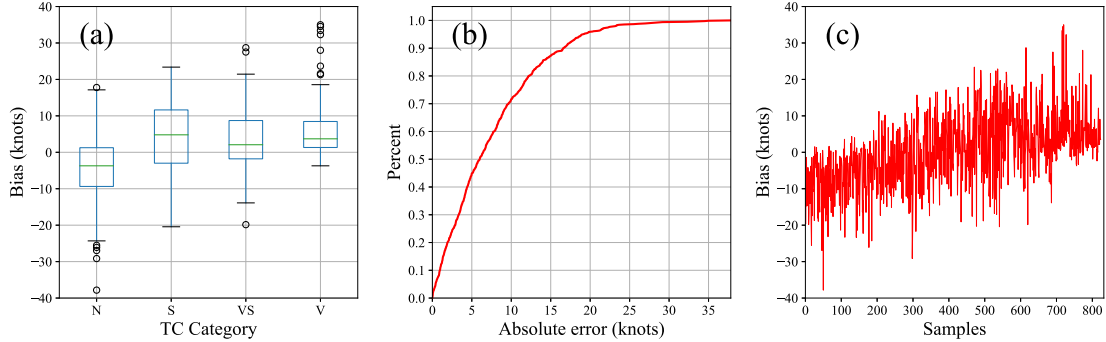


Fig. 9. (a) Box plot of bias in each category. (b) Absolute error curve related to the percentage using the DMANet. (c) Bias curve between the actual intensity and the estimated intensity in validation set samples.

MPEM. The LGAM can estimate the TC intensity based on the enhanced 3-D TC features, so the DMANet shows reliable estimation accuracy in both high intensity and low intensity and the DMANet errors are reduced.

3) *Analysis of DMANet Error Distribution*: Fig. 9(a) shows the box plot of bias in each category using the DMANet. The green line in the middle of box is the median of bias in different categories. The boundary lines of blue box and black line represent the first and third quartiles of the bias and the upper and lower limits of the bias, respectively. If the TC intensity bias has outliers that exceed the maximum or minimum observed values, these outliers are expressed in dots. Regardless of outliers, the MAE of violent category is the lowest. The median, lower limit, and upper limit are 3.70 knots, -3.71 knots, and 18.56 knots, respectively. When TC intensity is violent, it will bring greater disasters to the coast, so the violent TC is paid more attention. Obviously, the DMANet has completed this task well. Meanwhile, the above results show that the estimation of DMANet is conservative. That is to say, when the TC intensity category is violent, the model tends to underestimate the TC intensity. Because for a dataset with continuous numeral labels, the estimated intensity tends to be biased toward the middle of labels, which can bring loss function (MSE) greater benefits (the normal category is the same result). The highest error of MAE is the strong category, with a median of 4.79 knots and lower limit and upper limit of -20.44 knots and 23.37 knots, and the highest upper limit is from strong category. The smallest lower limit is root in normal category, with a median of -3.71 knots and lower limit and upper limit of -24.31 knots and 17.14 knots. In addition, the bias of very strong category has a median of 2.07 knots, and the lower and upper limits are -13.88 knots and 21.44 knots, respectively.

The absolute error curve related to the percentage is shown in Fig. 9(b). The percentages corresponding to the absolute error of 5 knots, 10 knots, 15 knots, and 20 knots are 45.0%, 71.7%, 87.5%, and 96.0%, respectively. Fig. 9(c) shows the bias curve between the actual intensity of TC and the estimated intensity in the testing set. The intensities of testing samples are arranged from low to high. The bias is obtained by subtracting the estimated value from the actual value. Obviously, the overall trend of bias gradually changes from negative value to positive value with the increase of TC intensity in Fig. 9(c).

B. Comparison With General Models

Due to the development of deep learning, a lot of general models (e.g., VGG-16 or ResNet-50) have been proved to perform well in regression tasks [56]. Meanwhile, the general models also show great accuracy in the TC intensity regression tasks [20], [29]. In order to contrast the performance of DMANet with the general models, the DMANet is compared with the general models (i.e., AlexNet, VGG-16, and ResNet-50) with the same tuning parameters and dataset. The results are listed in Table VI. The performances of DMANet and other general models are compared mainly based on the RMSE and MAE. Table VI presents that the DMANet model can obtain lower error and better accuracy of TC intensity estimation than the other general models.

C. Kalman Filter Performance

The results predicted by the DMANet and DMANet with a Kalman filter (DMANet_KF) are shown in Table VII. After Kalman filter correction, the RMSE error and MAE error of

TABLE VI
PERFORMANCE OF DMANET AND OTHER GENERAL MODELS

Model	RMSE (knots)	MAE (knots)	Parameters (MB)
AlexNet	12.94	10.15	11.67
ResNet-50	11.13	8.29	89.68
VGG-16	11.05	8.54	61.26
DMANet	9.79	7.52	12.74

TABLE VII
PERFORMANCE AFTER USING KALMAN FILTER

Model	RMSE (knots)	MAE (knots)	R^2
DMANet	9.79	7.52	0.853
DMANet_KF	7.82	6.19	0.907

network are reduced by 20.1% and 21.5%, respectively. The reason why Kalman filter works is explained in Section IV.

As shown in Fig. 10, four TC sequences are randomly selected from the testing set. As TC system absorbs energy from the surroundings to enhance the TC intensity, the TC intensity is a continuous development process. From the actual intensity, we can see that the intensity of TCs rarely changes in a short time, and most of them increase or decrease slowly. However, although the SCIs are from adjacent instants, the cloud cover and shape also show a different state, and the intensity estimated by model is also different. After correction with Kalman filter, the intensity changes more slowly. The intensity at most instants is closer to the actual intensity best track than the uncorrected intensity. In the case of TC 2017 TALIM, the RMSE and the maximum bias estimated by the DMANet from 2017/09/15/18:00 UTC to 2017/09/17/00:00 UTC are 14.78 knots and 22.50 knots, respectively. However, the estimated intensity of DMANet is unreliable. Then, Kalman filter is added to correct it, and the RMSE and the maximum bias are reduced to 6.02 knots and 11.6 knots, respectively.

Fig. 11 shows the scatter plots of best-track intensity versus estimated intensity using the DMANet model and DMANet_KF model. After the correction using Kalman filter, the performance of DMANet_KF model is better compared with that of DMANet model, which demonstrates that Kalman filter can indeed use information of intensity series to eliminate the error of the model to a certain extent. To our best knowledge, Kalman filter is first used for TC intensity estimation correction and the time-series estimation results are smoother after using Kalman filter. The results of modifying TC intensity time series well demonstrate the effectiveness of Kalman filter.

D. Comparison With Other Satellite Estimation Methods

The performance results of the existing deep learning models and DMANet_KF are shown in Table VIII. Strictly speaking, it is not fair to directly compare our model with these models because of different datasets used in these models. In

addition, it is difficult to reproduce the data and model at the same time. However, in order to show the performance of our method, the intensity estimation methods of other scholars are still used to compare. Compared with the intensity estimation methods of other scholars, the DMANet_KF achieves better performance. The best RMSE of the existing deep learning models is 8.60 knots. Compared with the best result, the RMSE of DMANet_KF model is reduced by 9.07%.

VI. DISCUSSION

With the rapid development of remote sensing technology, many multimodal data with complex and heterogeneous observations can be obtained. Multimodal remote sensing data on TCs are used to estimate TC intensity. Deep learning has been successfully applied to multimodal remote sensing data processing due to its ability to mine deep features and powerful processing capabilities. The main challenges faced by the TC intensity estimation are mainly attributed to the typhoon's variable cloud system features and the multimodal data fusion. On the one hand, this complex cloud distribution makes it difficult for models to determine typhoon intensity. And on the other hand, due to the heterogeneity of multimodal data, the simple superposition of data cannot take the advantage of multimodal data.

The study results show the potential of DMANet to integrate multimodal remote sensing data to estimate TC intensity. For the remote sensing images with different channels, the information is different. The message-passing enhancement mechanism of the DMANet model can capture and transfer complementary information between multimodal remote sensing data. Furthermore, the message-passing mechanism makes it suitable for remote sensing data of various spectra. Hence, their feature expression ability can be enhanced by fusing the advantageous features of multiple channels. These improved features help the DMANet model to effectively identify and extract deep features from the cluttered cloud information. Quantitative experiments are proven effective.

The cloud information of the multimodal remote sensing data contains many chaotic clouds. In order to avoid model confusion, the typhoon eye is given greater weight to help the DMANet model make the accurate judgment. However, this given weight is automatically calculated by the attention model of the DMANet model. The DMANet model judges the importance of the typhoon eye in the input remote sensing image to the task and automatically calculates the corresponding weight. Additionally, experts in related fields assess the intensity of TCs through the typhoon eye, and this knowledge is introduced into the structure of the DMANet model through the attention mechanism.

Furthermore, adding a time factor to TC intensity estimation can provide a more reasonable input for disaster prevention and mitigation models. This article uses the Kalman filter to correct the TC intensity for the first time. As shown in Fig. 10, the performances of correcting TC intensity time series well show the effectiveness of Kalman filter.

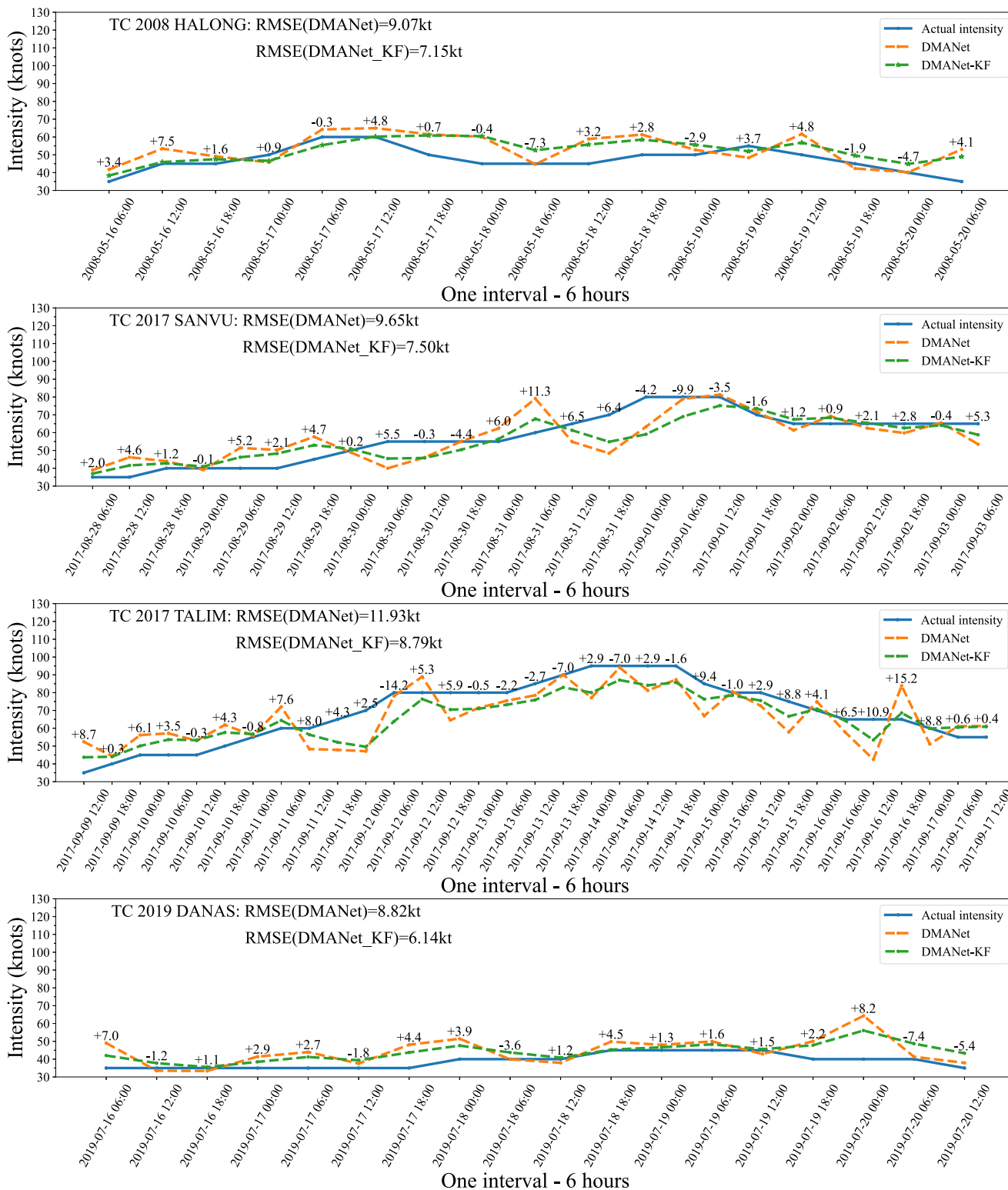


Fig. 10. Estimation of TC intensity over time series. The number above each time point represents the difference between the error estimated by the model and the error after Kalman filter. If the value is positive, it means that the estimated value after Kalman filter is closer to the estimated value of the network, and the closer value is the positive value. If it is negative, it is the opposite.

TABLE VIII
PERFORMANCE COMPARISON WITH OTHER SATELLITE ESTIMATION METHODS

References	data	models	RMSE (knots)	MAE (knots)
[34]	MINT	DeepMicroNet	10.60	-
[37]	IR, WV	TCIENet	9.97	7.83
[38]	IR	Deep-PHURIE (smoothed)	8.94	7.96
[20]	IR	VGG-16	11.68	-
[19]	IR	TCICENet (smoothed)	8.60	6.67
[39]	IR	CNN model	8.98	-
Our model	IR1, IR2, IR3, IR4	DMANet_KF	7.82	6.19

The bold entities indicate the best performance of the comparison methods.

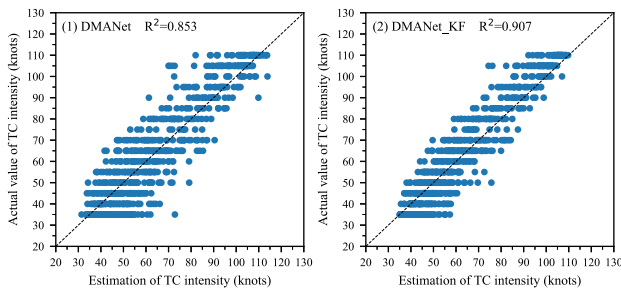


Fig. 11. Scatter plots of best-track intensity versus estimated intensity using the DMANet model and DMANet_KF model.

VII. CONCLUSION

In this article, we propose a novel DMANet for the TC intensity estimation, which provides practical guidance of using advanced computer vision technique to estimate the TC intensity. The multispectral satellite images in the Northwest Pacific Basin for the period 2007–2021 are downloaded from the National Institute of Informatics of Japan. A preprocessing method is proposed for solving the imbalance of dataset. Then, the DMANet model is proposed to estimate the TC intensity using multispectral satellite images. The model includes two aspects: first, an MPEM based on CRFs is used to produce the enhanced feature map by multispectral data, which is more conducive to mining different features of multiple wavelength channel images. Second, an LGAM is proposed to pay attention to the important feature and obtain the deeper global semantic information.

The ablation experiment shows that the DMANet model achieves excellent performance using SCIs and verifies the effectiveness of each module. Compared with the general models, the RMSE of DMANet is reduced by 11.40%–24.34%, and the MAE of DMANet is reduced by 7.59%–25.91%. Meanwhile, the DMANet model has the best performance when the TC category is violent, which indicates that the DMANet can estimate the intensity of violent TC more accurately. Finally, Kalman filter is used to modify the time-series estimation results. The RMSE of the testing set is decreased from 9.79 to 7.82 knots, and the MAE is decreased from 7.52 to 6.19 knots. Kalman filter verifies its applicability in the estimation of TC intensity.

REFERENCES

- [1] S. Wang and R. Toumi, “Recent migration of tropical cyclones toward coasts,” *Science*, vol. 371, no. 6528, pp. 514–517, 2021.
- [2] P. J. Webster, “Myanmar’s deadly daffodil,” *Nature Geosci.*, vol. 1, no. 8, pp. 488–490, 2008.
- [3] J. Z. Sippo, C. E. Lovelock, I. R. Santos, C. J. Sanders, and D. T. Maher, “Mangrove mortality in a changing climate: An overview,” *Estuarine, Coastal Shelf Sci.*, vol. 215, pp. 241–249, 2018.
- [4] S. Q. Kidder et al., “Satellite analysis of tropical cyclones using the advanced microwave sounding unit (AMSU),” *Bull. Amer. Meteorol. Soc.*, vol. 81, no. 6, pp. 1241–1260, 2000.
- [5] V. F. Dvorak, “A technique for the analysis and forecasting of tropical cyclone intensities from satellite pictures,” Nat. Oceanic Atmos. Admin., Nat. Environ. Satellite Serv., Washington, DC, USA, Tech. Memorandum NESS 45, 1973.
- [6] V. F. Dvorak, “Tropical cyclone intensity analysis and forecasting from satellite imagery,” *Monthly Weather Rev.*, vol. 103, no. 5, pp. 420–430, 1975.
- [7] V. F. Dvorak, “Tropical cyclone intensity analysis using satellite data,” U.S. Dept. Commerce, Nat. Oceanic Atmos. Admin., Washington, DC, USA, NOAA Tech. Rep. NESDIS 11, 1984.
- [8] C. S. Velden, T. L. Olander, and R. M. Zehr, “Development of an objective scheme to estimate tropical cyclone intensity from digital geostationary satellite infrared imagery,” *Weather Forecasting*, vol. 13, no. 1, pp. 172–186, 1998.
- [9] C. Velden et al., “The Dvorak tropical cyclone intensity estimation technique: A satellite-based method that has endured for over 30 years,” *Bull. Amer. Meteorol. Soc.*, vol. 87, no. 9, pp. 1195–1210, 2006.
- [10] 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.
- [11] 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, 2016.
- [12] 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, 2016, Art. no. 40.
- [13] L. Dai, C. Zhang, L. Xue, L. Ma, and X. 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, 2018.
- [14] 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. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 3, pp. 763–773, Mar. 2019.
- [15] K. Xiang, X. Yang, M. Zhang, Z. Li, and F. Kong, “Objective estimation of tropical cyclone intensity from active and passive microwave remote sensing observations in the northwestern Pacific ocean,” *Remote Sens.*, vol. 11, no. 6, 2019, Art. no. 627.
- [16] J. Liu, X. Xu, and X. Luo, “Estimation of tropical cyclone intensity using infrared data from a geostationary satellite,” *SOLA*, vol. 15, pp. 189–192, 2019.
- [17] A. Asif, M. Dawood, B. Jan, J. Khurshid, M. DeMaria, and F. ul Amir Afsar Minhas, “Phurie: Hurricane intensity estimation from infrared satellite imagery using machine learning,” *Neural Comput. Appl.*, vol. 32, no. 9, pp. 4821–4834, 2020.

- [18] Y.-J. Lee, D. Hall, Q. Liu, W.-W. Liao, and M.-C. Huang, "Interpretable tropical cyclone intensity estimation using Dvorak-inspired machine learning techniques," *Eng. Appl. Artif. Intell.*, vol. 101, 2021, Art. no. 104233.
- [19] C.-J. Zhang, X.-J. Wang, L.-M. Ma, and X.-Q. Lu, "Tropical cyclone intensity classification and estimation using infrared satellite images with deep learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2070–2086, Jan. 2021.
- [20] M. Higa et al., "Domain knowledge integration into deep learning for typhoon intensity classification," *Sci. Rep.*, vol. 11, no. 1, 2021, Art. no. 12972.
- [21] S. Dong, P. Wang, and K. Abbas, "A survey on deep learning and its applications," *Comput. Sci. Rev.*, vol. 40, 2021, Art. no. 100379.
- [22] A. Khan, A. Sohail, U. Zahoor, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artif. Intell. Rev.*, vol. 53, no. 8, pp. 5455–5516, 2020.
- [23] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Physica D, Nonlinear Phenomena*, vol. 404, 2020, Art. no. 132306.
- [24] F. Zhou, S. Yang, H. Fujita, D. Chen, and C. Wen, "Deep learning fault diagnosis method based on global optimization GAN for unbalanced data," *Knowl.-Based Syst.*, vol. 187, 2020, Art. no. 104837.
- [25] G. Hu, L. Liu, D. Tao, J. Song, K. T. Tse, and K. C. S. Kwok, "Deep learning-based investigation of wind pressures on tall building under interference effects," *J. Wind Eng. Ind. Aerodyn.*, vol. 201, 2020, Art. no. 104138.
- [26] X. Huang, D. Zhu, F. Zhang, T. Liu, X. Li, and L. Zou, "Sensing population distribution from satellite imagery via deep learning: Model selection, neighboring effects, and systematic biases," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 5137–5151, Apr. 2021.
- [27] F. Matrone, M. Paolanti, A. Felicetti, M. Martini, and R. Pierdicca, "Bubblx: An explainable deep learning framework for point-cloud classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 6571–6587, Aug. 2022.
- [28] L. Yue, F. Zan, X. Liu, Q. Yuan, and H. Shen, "The spatio-temporal reconstruction of lake water levels using deep learning models: A case study on Altai mountains," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4919–4940, Jun. 2022.
- [29] 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.
- [30] R. L. Bankert and P. M. Tag, "An automated method to estimate tropical cyclone intensity using SSM/I imagery," *J. Appl. Meteorol.*, vol. 41, no. 5, pp. 461–472, 2002.
- [31] M. F. Piñeros, E. A. Ritchie, and J. S. Tyo, "Estimating tropical cyclone intensity from infrared image data," *Weather Forecasting*, vol. 26, no. 5, pp. 690–698, 2011.
- [32] A. Kulkarni, R. Bankert, and M. Hadjimichael, "Tropical cyclone intensity estimation using neural networks," in *Proc. ASPRS Annu. Conf.*, 2010.
- [33] 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.
- [34] 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.
- [35] 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, 2019, Art. no. 1099.
- [36] 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, 2020, Art. no. 108.
- [37] 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. 2020.
- [38] M. Dawood, A. Asif, and F. ul Amir Afsar Minhas, "Deep-Phurie: Deep learning based hurricane intensity estimation from infrared satellite imagery," *Neural Comput. Appl.*, vol. 32, no. 13, pp. 9009–9017, 2020.
- [39] C. Wang, G. Zheng, X. Li, Q. Xu, B. Liu, and J. Zhang, "Tropical cyclone intensity estimation from geostationary satellite imagery using deep convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Mar. 2021, Art. no. 4101416.
- [40] R. Chen, W. Zhang, and X. Wang, "Machine learning in tropical cyclone forecast modeling: A review," *Atmosphere*, vol. 11, no. 7, 2020, Art. no. 676.
- [41] B. Zhang et al., "Progress and challenges in intelligent remote sensing satellite systems," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1814–1822, Feb. 2022.
- [42] J. Schmetz, S. Tjemkes, M. Gube, and L. van de Berg, "Monitoring deep convection and convective overshooting with MeteoSAT," *Adv. Space Res.*, vol. 19, no. 3, pp. 433–441, 1997.
- [43] F. M. Ralph, P. J. Neiman, and G. A. Wick, "Satellite and Caljet aircraft observations of atmospheric rivers over the Eastern North Pacific Ocean during the winter of 1997/98," *Monthly Weather Rev.*, vol. 132, no. 7, pp. 1721–1745, 2004.
- [44] C. S. Velden, C. M. Hayden, S. J. W. Nieman, W. P. Menzel, S. Wanzong, and J. S. Goerss, "Upper-tropospheric winds derived from geostationary satellite water vapor observations," *Bull. Amer. Meteorol. Soc.*, vol. 78, no. 2, pp. 173–196, 1997.
- [45] T. F. Lee, F. Joseph Turk, and K. Richardson, "Stratus and fog products using goes-8-9 3.9- μm data," *Weather Forecasting*, vol. 12, no. 3, pp. 664–677, 1997.
- [46] D. Rosenfeld, W. L. Woodley, A. Lerner, G. Kelman, and D. T. Lindsey, "Satellite detection of severe convective storms by their retrieved vertical profiles of cloud particle effective radius and thermodynamic phase," *J. Geophys. Res., Atmos.*, vol. 113, no. D4, 2008, Art. no. D04208.
- [47] H.-J. Ye, D.-C. Zhan, and W.-L. Chao, "Procrustean training for imbalanced deep learning," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 92–102.
- [48] L. Liu, Z. Qiu, G. Li, S. Liu, W. Ouyang, and L. Lin, "Crowd counting with deep structured scale integration network," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 1774–1783.
- [49] Y. Li, J. Zeng, S. Shan, and X. Chen, "Occlusion aware facial expression recognition using CNN with attention mechanism," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2439–2450, May 2019.
- [50] L. He, J. C.-W. Chan, and Z. Wang, "Automatic depression recognition using CNN with attention mechanism from videos," *Neurocomputing*, vol. 422, pp. 165–175, 2021.
- [51] 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.
- [52] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng.*, vol. 82, pp. 35–45, 1960.
- [53] G. Welch and G. Bishop, *An Introduction to the Kalman Filter*, Univ. North Carolina at Chapel Hill, Chapel Hill, NC, USA, 1995.
- [54] F. Auger, M. Hilaret, J. M. Guerrero, E. Monmasson, T. Orłowska-Kowalska, and S. Katsura, "Industrial applications of the Kalman filter: A review," *IEEE Trans. Ind. Electron.*, vol. 60, no. 12, pp. 5458–5471, Dec. 2013.
- [55] T.-C. Wu, C. S. Velden, S. J. Majumdar, H. Liu, and J. L. Anderson, "Understanding the influence of assimilating subsets of enhanced atmospheric motion vectors on numerical analyses and forecasts of tropical cyclone track and intensity with an ensemble Kalman filter," *Monthly Weather Rev.*, vol. 143, no. 7, pp. 2506–2531, 2015.
- [56] S. Lathuilière, P. Mesejo, X. Alameda-Pineda, and R. Horaud, "A comprehensive analysis of deep regression," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 9, pp. 2065–2081, Sep. 2020.

Wenjun Jiang was born in China in 2000. He is currently working toward the M.S. degree in civil engineering with the Harbin Institute of Technology, Shenzhen, China.

His research interests include deep learning and data and information processing for tropical cyclone.

Gang Hu received the M.S. degree in structure engineering from the Harbin Institute of Technology, Shenzhen, China, in 2011, and the Ph.D. degree in civil engineering from the Hong Kong University of Science and Technology, Hong Kong, in 2015.

He is currently a Professor with the School of Civil and Environmental Engineering, Harbin Institute of Technology, Shenzhen, China. His research interests include application of machine learning and artificial intelligence technology in civil engineering and evaluation and optimization of urban wind environment.

Tiantian Wu is currently working toward the Ph.D. degree in civil engineering with the Harbin Institute of Technology, Shenzhen, China.

Her research interests include data and information processing for tropical cyclone and wind engineering.

Lingbo Liu received the Ph.D. degree in computer science from Sun Yat-Sen University, Guangzhou, China, in 2021.

He was a Visiting Scholar with the University of Sydney. He is currently a Research Assistant Professor with Hong Kong Polytechnic University, Hong Kong. His research interests include urban computing and machine learning.

Bubryur Kim received the Ph.D. degree in civil engineering from The Hong Kong University of Science and Technology, Hong Kong, China, in 2019.

Yiqing Xiao received the M.S. degree in structural mechanics and the Ph.D. degree in engineering mechanics from the Harbin Institute of Technology, Shenzhen, China, in 1998 and 2001, respectively.

He is currently a Professor with the School of Civil and Environmental Engineering, Harbin Institute of Technology, Shenzhen, China. His research interests include wind engineering, structural health monitoring, and safety assessment research.

Zhongdong Duan received the M.S. and Ph.D. degrees in engineering mechanics from the Harbin Institute of Technology, Shenzhen, China. He is currently a Professor with the School of Civil and Environmental Engineering, Harbin Institute of Technology, Shenzhen, China. His research interests include wind engineering, ocean engineering, natural disaster risk analysis, and simulation and structural random vibration and reliability analysis.