

Article

Tropical Cyclone Intensity Estimation Using Multi-Dimensional Convolutional Neural Networks from Geostationary Satellite Data

Juhyun Lee ¹, Junggho Im ^{1,*} , Dong-Hyun Cha ¹, Haemi Park ²  and Seongmun Sim ¹

¹ School of Urban & Environmental Engineering in Ulsan National Institute of Science and Technology, Ulsan 44919, Korea; wngus_0225@unist.ac.kr (J.L.); dhcha@unist.ac.kr (D.-H.C.); smsim@unist.ac.kr (S.S.)

² Institute of Industrial Science in the University of Tokyo, A building, 4 Chome-6-1 Komaba, Meguro City, Tokyo 153-8505, Japan; uiishmpark@g.ecc.u-tokyo.ac.jp

* Correspondence: ersgis@unist.ac.kr; Tel.: +82-52-217-2824

Received: 25 November 2019; Accepted: 25 December 2019; Published: 28 December 2019



Abstract: For a long time, researchers have tried to find a way to analyze tropical cyclone (TC) intensity in real-time. Since there is no standardized method for estimating TC intensity and the most widely used method is a manual algorithm using satellite-based cloud images, there is a bias that varies depending on the TC center and shape. In this study, we adopted convolutional neural networks (CNNs) which are part of a state-of-art approach that analyzes image patterns to estimate TC intensity by mimicking human cloud pattern recognition. Both two dimensional-CNN (2D-CNN) and three-dimensional-CNN (3D-CNN) were used to analyze the relationship between multi-spectral geostationary satellite images and TC intensity. Our best-optimized model produced a root mean squared error (RMSE) of 8.32 kts, resulting in better performance (~35%) than the existing model using the CNN-based approach with a single channel image. Moreover, we analyzed the characteristics of multi-spectral satellite-based TC images according to intensity using a heat map, which is one of the visualization means of CNNs. It shows that the stronger the intensity of the TC, the greater the influence of the TC center in the lower atmosphere. This is consistent with the results from the existing TC initialization method with numerical simulations based on dynamical TC models. Our study suggests the possibility that a deep learning approach can be used to interpret the behavior characteristics of TCs.

Keywords: tropical cyclones; multispectral imaging; 2D/3D convolutional neural networks

1. Introduction

On-going climate change makes natural disasters unpredictable. The Intergovernmental Panel on Climate Change (IPCC) special report said that the global warming over 2 °C leads to an increase of heavy rainfall frequency and a decrease in the occurrence of tropical cyclones (TCs), but an increase of the number of strong TCs. It makes people hard to prepare natural disasters such as TCs, which cause huge damage to human beings and infrastructure [1,2]. According to World bank annual report 2012, behavioral changes in TCs could cause a direct increase in the economic damage they create, from \$28 billion in 2010 to \$68 billion by 2100 [3,4]. East Asia is one of the regions vulnerable to natural disasters, where approximately 30% of the global economic damage caused by TCs occurs [5]. In order to better prepare for and respond to TC disasters, the quick identification of TC intensity is crucial, as well as accurate forecasting.

However, it is difficult to monitor TCs using ground-based observations because TCs generally occur and develop in the middle of the ocean. The development of meteorological satellite sensor

systems has opened a new era in climate forecasting and meteorological observations. High temporal resolution geostationary satellite data are considered to be one of the most reliable means of observing TCs in real-time, producing information on the various characteristics of TCs, such as intensity and center location [6–8]. A widely used method to determine TC intensity is the Dvorak technique [9]. It is a manual algorithm, which determines the intensity of TCs based on empirical satellite image analysis. The Dvorak technique has been used for extracting the scale of TCs and establishing relevant damage recovery policies. However, due to its subjectivity, the reliability of the real-time intensity readings based on the algorithm is inevitably low [10–13].

To overcome this limitation, many researchers have proposed objective TC intensity estimation algorithms. Velden et al. [14] proposed the objective Dvorak technique (ODT), which is a computer-based algorithm that uses satellite images. The coldest ring of a cyclone, which has the lowest brightness temperature within a certain radius around the eye of the cyclone, is used for estimating the TC's intensity. It is verified using the minimum sea level pressure (MSLP) to represent the intensity of the TC, resulting in a root mean square error (RMSE) of 11.45 hPa. Olander et al. [10] proposed an advanced Dvorak technique (ADT), which is an improved version of ODT that uses specific physical conditions for not only TCs with eyes but also non-eyed weak TCs. ADT performed about 20% better than the ODT method. Pineros et al. [15] introduced a systematic interpretation method using satellite infrared images of TCs and Ritchie et al. [16] proposed the deviation-angle variance technique (DAV-T) based on the structural analysis of TCs. DAV-T quantifies the trend of pixel-based brightness temperatures toward the center of a TC and determines its intensity using a degree of concentration in the center location. Their estimation model was tested based on mean sustained wind speed (MSW) as the reference intensity of TCs with validation using the best track data issued by the Joint Typhoon Warning Center (JTWC). The DAV-T showed an RMSE of 12.7 kts for TCs that occurred in the northwestern Pacific from 2005 to 2011.

More recently, convolutional neural networks (CNNs), which are one of the deep learning techniques in artificial intelligence, have been used to analyze satellite-based TC images to estimate their intensity. Pradhan et al. [17] estimated the intensity of TCs using single IR images based on CNNs, resulting in an RMSE of 10.18 kts. Combinido et al. [18] adopted the Visual Geometry Group 19-layer (VGG19) model for estimating TC intensity, which is a well-performing 2D-CNN architecture for image analysis proposed by Simonyan et al. [19]. They used single IR TC images from multiple geostationary satellite sensors from 1996 to 2016 over the Western North Pacific to develop the model, resulting in an RMSE of 13.23 kts. Wimmers et al. [20] used satellite-based passive microwave sensors to estimate TC intensity using the 2D-CNN approach. They used 37, 85–92 GHz channels to extract TC images as input data. The model shows a validation RMSE of 14.3 kts.

In this study, we benchmarked the existing TC intensity estimation methods based on CNNs and proposed improved CNN models using geostationary satellite-based multi-spectral images, which adopt a multi-dimensional approach, considering the vertical structure of TCs. Whereas existing methods using geostationary satellites consider single infrared channels for extracting cloud top pattern of TCs, we attempted to incorporate the three-dimensional asymmetrical structure of TCs caused by vertical wind shear which affects the intensity of TCs [21]. The horizontal and vertical TC patterns were analyzed by multi-dimensional CNNs which have shown marvelous performance in image pattern recognition and remote sensing [22–26]. Multi-channels input data were used in the proposed CNN models for estimating TC intensity. The proposed approach uses multi-sensor based satellite images to show the correlation between the shape and intensity of TCs with consideration for the water vapor diameters constituting TCs. The objectives of this research were to 1) propose an objective TC intensity estimation algorithm based on CNN approaches using geostationary satellite data, 2) identify the best CNN model with optimized parameters adapted to estimate TC intensity using satellite-based multi-spectral images, and 3) examine the significant impact of the vertical relative distribution of water vapor on TC intensity estimation using heat maps, which is one of the visualization methods for interpreting CNN-based deep learning models.

2. Data

2.1. Geostationary Meteorological Satellite Sensor Data

We used Communication, Ocean, and Meteorological Satellite (COMS) Meteorological Imager (MI) sensor data to estimate the intensity of TCs. COMS is the first Korean geostationary meteorological satellite and was launched in 2010. It is stationed at the longitude of 128.2°E and 36,000 km above the earth equator [27]. The COMS MI sensor observes every 15 min over one side of the Earth with a horizontal spatial resolution of 1 km to 4 km. The sensor consists of five spectral channels—one visible channel and four infrared channels (Table 1).

Table 1. Instrument specification of Communication, Ocean and Meteorological Satellite (COMS) Meteorological Imager (MI) sensor. The infrared bands (SWIR, WV, IR1 and IR2) were used in this study.

Channel	Wavelength Range (μm)	Central Wavelength (μm)	Spatial Resolution (km)	Temporal Resolution (min)
Visible (VIS)	0.55-0.8	0.67	1	
Shortwave Infrared (SWIR)	3.5-4.0	3.7	4	
Water vapor (WV)	6.5-7.0	6.7	4	15
Infrared 1 (IR1)	10.3-11.3	10.8	4	
Infrared 2 (IR2)	11.5-12.5	12.0	4	

The infrared channels are widely used for deriving cloud information such as water vapor content of atmospheric layers. The MI sensor has multi-spectral channels from a short-wave infrared channel of 3.7 μm to a long wavelength channel of 12.0 μm . The long-wavelength channels with infrared 1 (IR1, 10.8 μm) and infrared 2 (IR2, 12.0 μm) are sensitive to the water vapor contents in the upper atmosphere [6,14,28]. The water vapor channel (WV, 6.7 μm) provides middle atmospheric components [29]. The shortwave infrared (SWIR, 3.7 μm) is widely used for detecting low clouds. Because it has lower brightness temperature as the diameter of the droplet in the atmosphere increases, the value variation is larger than that of the longwave channel [30,31]. Rosenfeld et al. [32] proposed the vertical profile of the cloud drop effective radius in a severe convective storm through satellite observation-based simulation. In severe storms, the higher the altitude of the atmosphere, the larger effective radius (i.e., effective radius of 1–7 μm at the atmospheric altitude < 2 km to 10–30 μm at the atmospheric height > 6 km). In addition, the water droplet-content-rate goes higher with a higher atmospheric altitude in severe storms. Since the WV and SWIR channel signals were related to the lower atmosphere, they could be considered as the cloud droplet distribution with a smaller effective radius in a lower altitude of the atmosphere [32,33].

2.2. Best Track Data

JTWC produces tropical cyclone data known as “best track” data which is from the International Best Track Archives for Climate Stewardship (IBTrACS). It includes the location of the tropical cyclone center (degrees), maximum sustained wind speed (kts), minimum sea level pressure (hPa) and tropical cyclone radius for the Southern Hemisphere (SH), the Northern Indian Ocean (NIO) and the Western North Pacific (WNP) regions. The annually-organized best track data with a 6-hour interval are officially provided 6 months to 1 year after the previous year due to their post-processing using corrected observational data and numerical model results [34–36]. Our research was conducted using the TCs generated in the WNP region within the observation range of the COMS MI sensor. There are the most frequent occurrences and the largest lifetime maximum intensity (LMI) variations of TCs over this region.

3. Methodology

3.1. Input Data Preparation

TC images used to develop the estimation models were extracted from four infrared channels of COMS MI. Since tropical cyclone eyewalls, the shape of spiral rain-bands formed by the cirrus outflow, and vertical wind shear are all crucial structural factors for estimating intensity [7], it is necessary to use an image that covers the whole shape of a TC as an input for training the patterns of TCs. We delineated one 301×301 -pixel image (i.e., $1204 \text{ km} \times 1204 \text{ km}$) per TC based on the grid of each TC center location from JTWC best track data. This is illustrated using COMS in Figure 1. The delineated input images were scaled up to 101×101 pixel images based on bilinear interpolation using the image resize tool available in MATLAB 2018a for computational efficiency so that the input images had a horizontal spatial resolution of about 12 km.

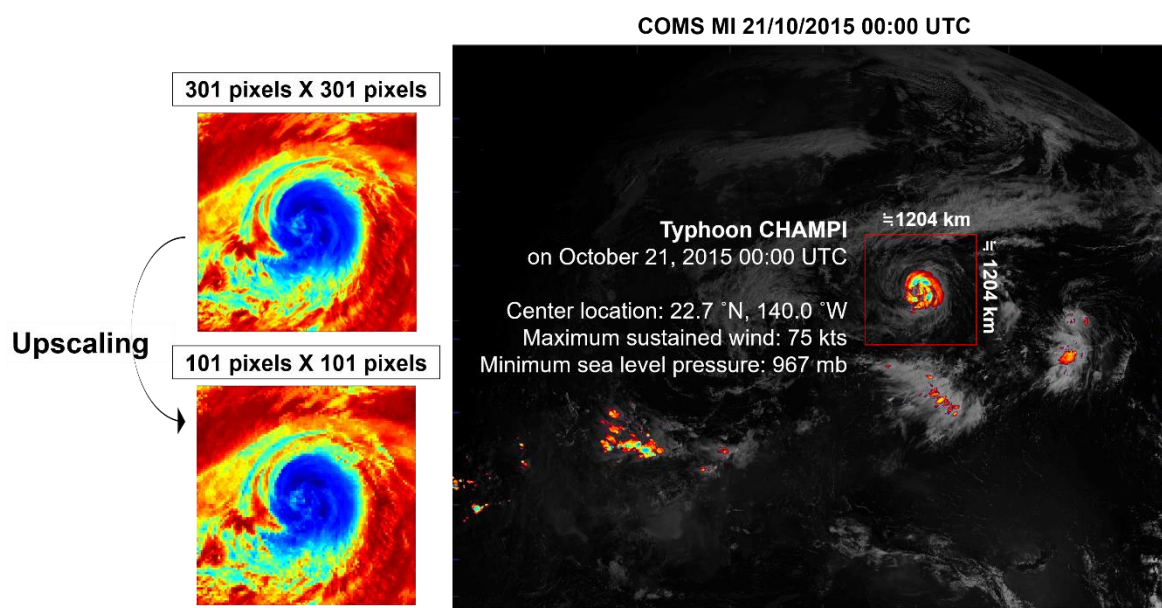


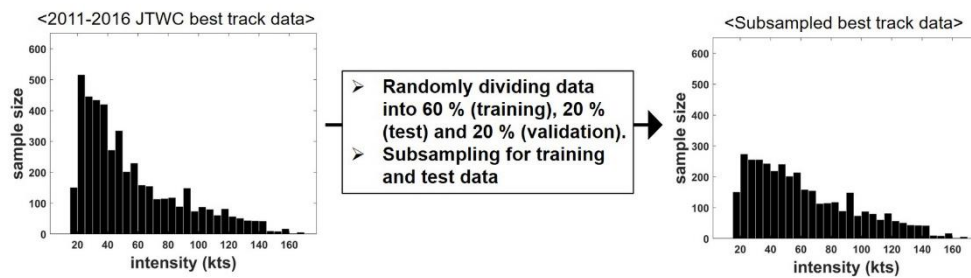
Figure 1. COMS MI data-based tropical cyclone image extraction. The images are extracted with a 301×301 pixel rectangle corresponding to $1204 \text{ km} \times 1204 \text{ km}$ based on the tropical cyclone center location. The images are upscaled to 101×101 pixels to reduce computational demand. Since tropical cyclones are larger scaled phenomena than cloud clusters, tropical cyclone patterns, such as spiral rain bands or eyewall strength pattern, clearly appear in the upscaled images.

To train CNN-based intensity estimation models, we used COMS MI images of TCs which developed over the WNP from 2011 to 2016 based on the best track data. Since the track data consists of accumulating the intensity of the TC lifetime, it causes an imbalance problem in training data in terms of intensity. Approximately 70% of the TCs have an intensity fewer than 63 kts, whereas only 12% show intensity of 96 kts or more, which is likely to cause devastating damage. The data imbalance problem leads to overfitting major samples and poorly estimates minority samples [37]. To overcome that, we balanced our dataset through subsampling and oversampling processes.

Prior to these processes, the best track data between 2011 and 2016 were randomly divided into 6:2:2 for training, test, and validation data, respectively, and only training and test data were balanced according to intensity for unbiased training and parameterization. To balance our dataset, we removed the samples with a high frequency of intensity. Using the 10 kts interval-based histogram, we randomly removed the samples on some bins which had more than 25% frequency. Then, the data of the other bins were augmented through two oversampling processes: hourly interpolation and image rotation. The 6-hour-interval TC tracks were interpolated into hourly data during the subsampling

process, except for the randomly removed ones (Figure 2). The distribution of temporally interpolated hourly data was still imbalanced due to a minority of high-intensity samples. To balance the dataset, the extracted images were augmented by rotating them to various angles. The smaller the number of binned data, the more images were augmented with smaller angles. In addition, four major TCs developed in 2017 (i.e., 2017 Typhoon HATO, KAHNUN, LAN, and DAMREY) were used for additional hindcast validation of our proposed models. Finally, a total of 49,420 balanced samples were prepared. Table 2 shows the difference in the number of samples before and after data adjustment.

1) Subsampling of reference data by TC intensity



2) Oversampling of reference data through temporal interpolation and image rotation

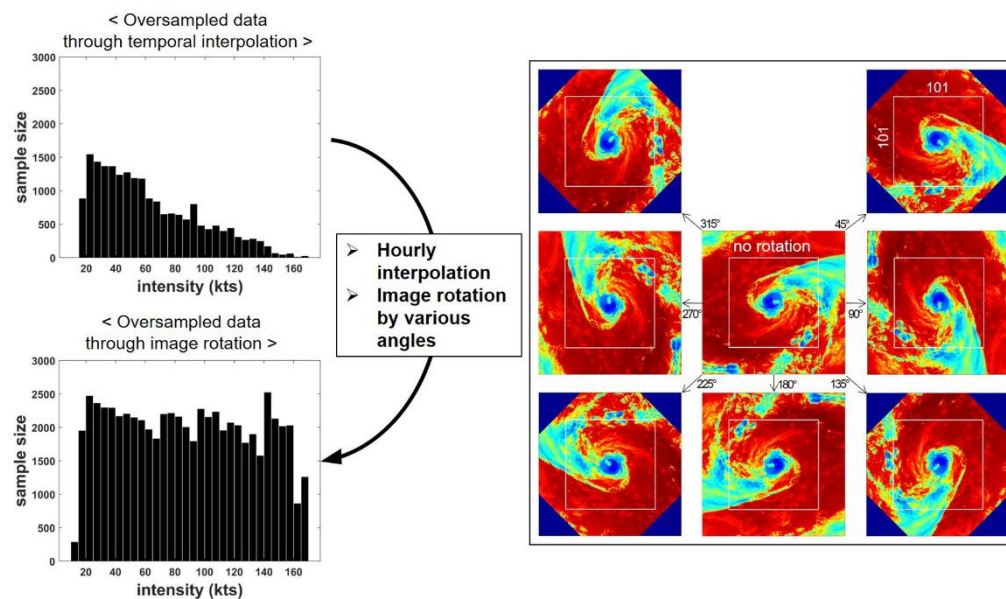


Figure 2. Overview of subsampling and oversampling processes for balancing tropical cyclone samples.

Table 2. The number of samples before and after data balancing processes.

	Original	Balanced
Training	2742	34,802
Test	914	13,632
Validation	915	915
Hindcast validation for 2017	71	71
Sum	4642	49,420

Based on the preprocessed data, four infrared images of TCs were extracted and stacked to construct the input data. Each channel of input data was normalized from 0 to 1 in order to focus on the pattern of clouds [38]. Figure 3 shows an example of the difference in the convective pattern distribution

by wavelength. Each wavelength depicts a pattern that is affected by the wavelength-corresponding particle size. That is, the stacking of the multi-infrared images shows the relative difference in droplet particle size. Therefore, we used the stacked dataset as an indicator of relative TC formation.

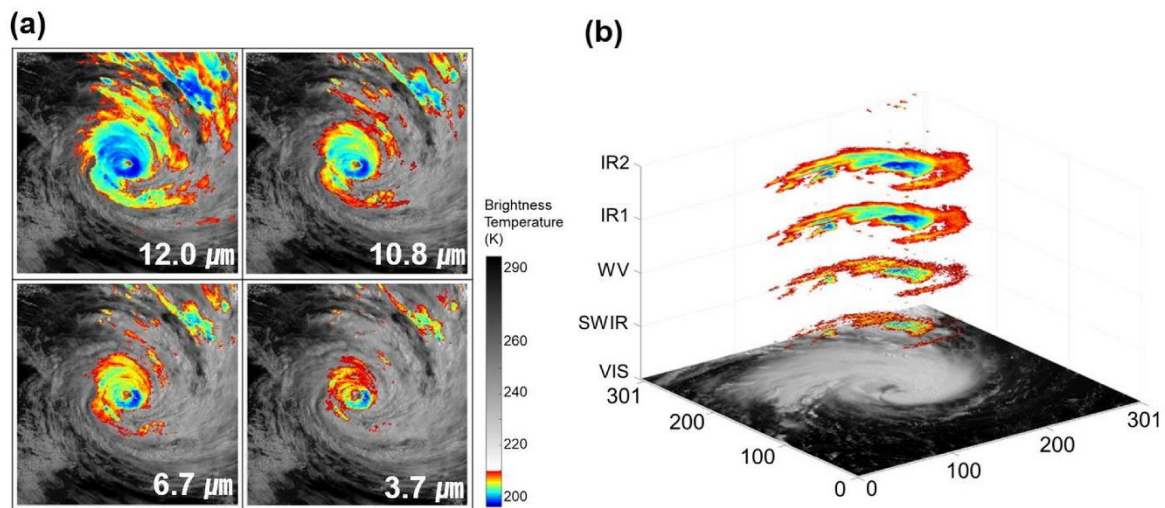


Figure 3. Deep convective areas of four channels. They show the spiral patterns by channel—the lower 20% of the brightness temperature in each channel is indicated as relatively strong convection. (a) Longer-wave infrared channels show wider convective areas than shorter-wave infrared channels for a developed tropical cyclone. (b) Spiral convective area distribution for each channel image. This can be represented by the size of the droplets corresponding to each wavelength.

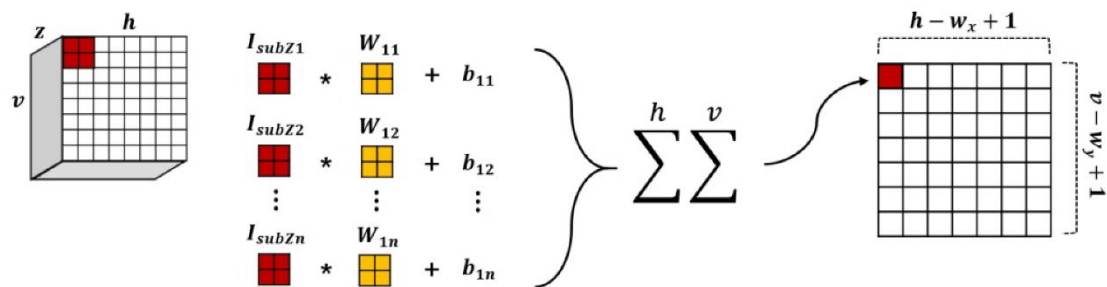
3.2. Convolutional Neural Networks (CNNs)

CNNs are one of the hierarchical neural network systems which use feature vector extraction to make multi-dimensional complex data analyzable through dimensional reduction [39–41]. Thus, it has been widely adopted for recognizing visualization data such as handwriting, photos, and medical images [27,42–47]. Recently, it has been used in climate-related research to recognize the pattern of numerical model results, reanalysis data, and satellite-based observations [29,48–51]. CNNs consist of three major parts: the convolutional layers, pooling layers, and the fully connected layer. The information from the input data pattern is extracted in the convolutional layers and the data dimension decreases through the pooling layers. After that, the extracted feature determines the output based on the extracted values in the fully connected layers. In the convolutional layers, weights and biases are compared in each layer and any that are shared are optimized based on a back-propagation process. Subsequently, the last convolutional layer is flattened so that it can be applied to the fully connected layer to train the features extracted from the convolutional layers [22,42,52].

In this study, we used two types of CNNs for estimating TC intensity. One is two-dimensional (2D)-CNNs and the other is three-dimensional (3D)-CNNs. Figure 4 summarizes how the calculations in the convolutional layers are conducted showing the difference between the two methods (i.e., 2D- and 3D-CNNs). The major difference between the two is the dimension of the kernel applied to each convolutional layer. In 2D-CNNs, an activation map, which is the output of each convolutional layer, is calculated as the sum of the layers applied with the kernel and bias of each layer. 3D-CNNs, however, use one-dimensional level higher than 2D-CNNs in terms of hyper-parameters, such as kernels and pooling layers, which are effective in multi-dimensional data [19,44]. The three-dimensional kernel and bias are applied to the segmented inputs. Since kernels are convoluted three-dimensionally in 3D-CNNs, the output can have a multi-channel, which enables the information from each input channel to be preserved. If the user defines the kernel depth of a 3D-CNN model as 1, the number of activation map bundles corresponding to the number of the input data depth is generated. On the other hand, 2D-CNNs generate as many activation maps as kernels in each convolutional layer. As one

convolutional layer passes, a 2D-CNN model has a set of as many activation maps as the number of kernels. However, a 3D-CNN can generate multiple times as many activation maps as the number of kernels, which results in a significant increase in computational demand. Whereas a low-dimensional CNN is more adept at simplifying and effectively training input data through the aggregation of the information as a convolutional layer passes by, a high-dimensional CNN could train the model while preserving more information from each channel. However, a high-dimensional CNN has a disadvantage in terms of the complexity of the model including parameter optimization.

(a) Calculation in 2D convolutional layer with a kernel (w_x, w_y)



(b) Calculation in 3D convolutional layer with a kernel ($1, w_x, w_y$)

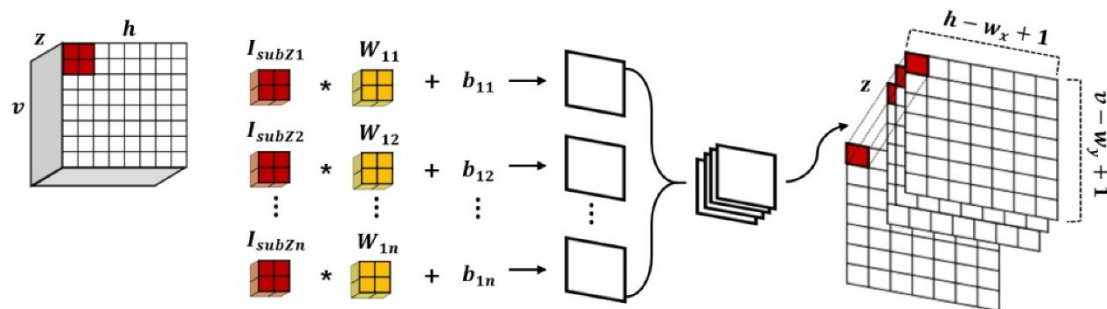
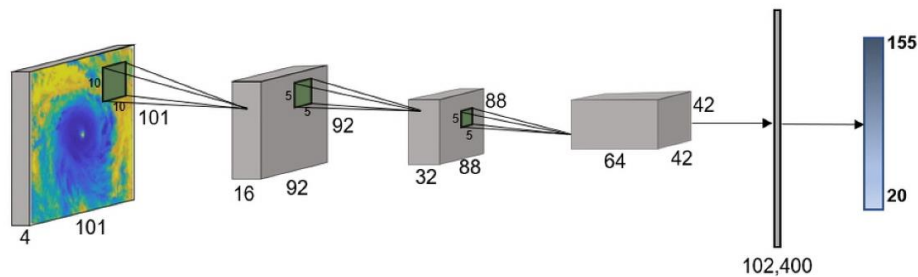


Figure 4. Summary of how calculations in 2D-CNNs and 3D-CNNs are conducted. The h, v and z mean horizontal, vertical, and depth scale of input data, respectively. The set of I_{subZn} is the segmented features of input data through convolution. W and b are the weight and bias which are applied to I_{subZn} . The size of the output is determined by the filter and pooling layer size.

Figure 5 shows the basic architectures of 2D-CNN and 3D-CNN models proposed in this study. Due to the high dimensional level of the 3D-CNN model, it takes much more processing time with more parameters to be optimized when compared to the 2D-CNN with similar hyper-parameters: horizontal kernel size and pooling layers. A large number of parameters causes the model to become complicated, which often results in an overfitting problem [53]. Nevertheless, the advantage of 3D-CNN is that it can keep the characteristics of each channel according to the user-defined filters, while 2D-CNN combines the information that passes on a convolutional layer. Previous studies on estimating TC intensity using satellite data have used long-wavelength infrared images at about $11 \mu\text{m}$, which can be used to observe the cloud top pattern. The studies have typically used 2D-CNN models to analyze single-spectral TC images [17,18]. In this study, we used multi-spectral infrared images from short-wavelength at $3.7 \mu\text{m}$ to long-wavelength at $12.0 \mu\text{m}$ that were derived from a meteorological satellite sensor in a multi-dimensional CNN framework. The multi-dimensional CNN such as 3D-CNN has been effectively used for image analysis considering the three-dimensional shape of the object under investigation or the time-series of scenes [44,54,55]. The multi-layered input data were analyzed using both 2D-CNN

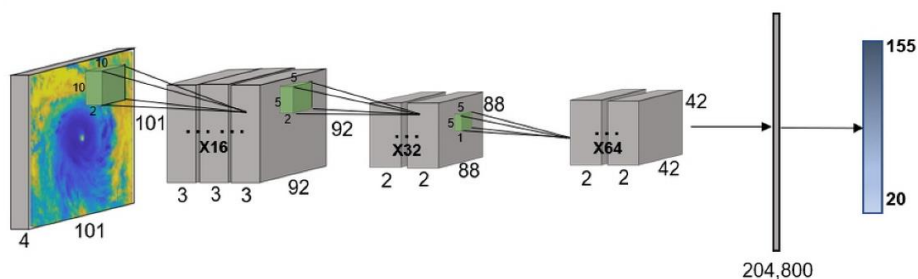
and 3D-CNN to consider the characteristics of each channel. The CNN experiments were conducted using Tensorflow of Keras deep learning framework in Python. The proposed networks were trained in a GPU environment which provided more cost-effective calculations than a CPU [56–59]. We used a Volta GPU (NVIDIA Tesla V100) which has 5,376 CUDA cores and 16 GB of memory.

(a) 2D CNN architecture



	*ConvLayer1	ConvLayer2	ConvLayer3	**FC1
Input	(* ,4,101,101)	(* ,16,92,92)	(* ,32,82,82)	(* ,102400)
Kernel	(16,10,10)	(32,5,5)	(64,5,5)	
Pooling	(1,1)	(1,1)	(2,2)	
Output	(* ,16,92,92)	(* ,32,82,82)	(* ,64,40,40)	(* ,1)

(b) 3D CNN architecture



	*ConvLayer1	ConvLayer2	ConvLayer3	**FC1
Input	(* ,4,1,101,101)	(* ,16,3,92,92)	(* ,32,2,82,82)	(* ,204800)
Kernel	(16,2,10,10)	(32,2,5,5)	(64,1,5,5)	
Pooling	(1,1,1)	(1,1,1)	(1,2,2)	
Output	(* ,16,3,92,92)	(* ,32,2,82,82)	(* ,64,2,40,40)	(* ,1)

*ConvLayer = Convolutional Layer
 **FC = Fully Connected Layer

Figure 5. (a) 2D-CNN model architecture and (b) 3D-CNN model architecture used in this study—the major distinction between the two is that the three-dimensional kernels are applied in the 3D-CNN model. Whereas three-dimensional kernels enable the preservation of the difference between channels, the computational load is much more significant than the 2D-CNN model for training the parameters.

3.3. Optimization and Schemes

A CNN model is optimized by a set of hyper-parameters, such as the depth of convolutional layers, size and number of filters, and scale of the pooling layer. In particular, the convolutional depth and filter size of each layer is sensitive to the characteristics of the input data. The smaller the filter size, the more the model is able to catch the local characteristics of the input images. On the

other hand, large filter size is suitable for getting the general pattern of the input images. Whereas a small filter extracts lots of information from the input data, it slows down the dimensional reduction, which may require training through a deeper convolutional layer [22]. Therefore, it is important to find an optimum filter size and convolutional layer depth appropriate for the characteristics of the satellite-based TC images. Several schemes were tested with hyper-parameters adjustments to find an optimum model.

Table 3 shows the schemes with model architectures that were evaluated in this study. The “Control” model mimics the model proposed by Pradhan et al. [17] (i.e., CNN based TC intensity estimation algorithm with a single-spectral channel (IR1)-based TC images), which was tested with our dataset. The “Control4channels” model has the exact same architecture as the “Control” model except for the depth of the input data with multi-spectral channels (i.e., IR2, IR1, WV, and SWIR). A comparison of the two models shows the difference in model performance according to the type of input dataset. The names of the other models (i.e., IDs) consist of the CNN type and number. All the models except the “Control” model use a four-layered TC dataset.

Table 3. Details of CNN model architectures. C, P and FC means a convolutional layer, pooling layer and fully connected layer, respectively. The meaning of the numbers in each abbreviation is as follows; C(# of filters)@(size of filter(horizontal size * vertical size)), P(size of pooling layer (horizontal size * vertical size)). Each convolutional layer has a ReLu activation function and all the models are optimized using Adam optimizer with $\beta = 0.999$, $\epsilon = 1 \times 10^{-6}$.

Model ID	Input Channel	CNN Type	Conv Layer	Parameters
Control	IR1	2D	3	C64@10, P2, C256@5, P3, C288@3, P3, FC256, dropout = 0.5, stride = 1, $\beta = 0.999$, $\epsilon = 1 \times 10^{-6}$
Control 4channels		2D	3	C64@10, P2, C256@5, P3, C288@3, P3, FC256, dropout = 0.5, stride = 1, $\beta = 0.999$, $\epsilon = 1 \times 10^{-6}$
2d1		2D	5	C16@10, P1, C32@5, P2, C32@5, P2, C128@5, C128@5, FC512, dropout = 0.5, stride = 1, $\beta = 0.999$, $\epsilon = 1 \times 10^{-6}$
2d2		2D	6	C32@3, P2, C64@3, P3, C128@3, P1, C256@3, P1, C512@3, P1, C128@3, dropout = 0.25, FC512, stride = 1, $\beta = 0.999$, $\epsilon = 1 \times 10^{-6}$
2d3	IR2, IR1, WV, SWIR	2D	6	C32@7, P2, C64@7, P3, C128@7, P1, C256@7, P1, C512@7, P1, C128@7, P1, dropout = 0.25, FC512, stride = 1, $\beta = 0.999$, $\epsilon = 1 \times 10^{-6}$
2d4		2D	6	C32@10, P2, C64@10, P3, C128@10, P1, C256@10, P1, C512@10, P1, C128@10, P1, dropout = 0.25, FC512, stride = 1, $\beta = 0.999$, $\epsilon = 1 \times 10^{-6}$
3d1		3D	4	C16@10*2, P1*1, C32@5*2, P2*1, C32@5*1, C128@5*1, FC51200, dropout = 0.5, stride = 1, $\beta = 0.999$, $\epsilon = 1 \times 10^{-6}$
3d2		3D	6	C32@3*1, P2*1, C64@3*1, P2*1, C128@3*1, P1*1, C256@3*1, P1*1, C512@3*1, P1*1, C128@3*1, P1*1, dropout = 0.25, FC512, stride = 1, $\beta = 0.999$, $\epsilon = 1 \times 10^{-6}$
3d3		3D	6	C32@5*1, P2*1, C64@5*1, P3*1, C128@5*1, P1*1, C256@5*1, P1*1, C512@5*1, P1*1, C128@5*1, P1*1, dropout = 0.25, FC512, stride = 1, $\beta = 0.999$, $\epsilon = 1 \times 10^{-6}$

3.4. Accuracy Assessment

Model performances were evaluated and compared using various statistical indices: mean absolute error (MAE), root mean squared error (RMSE), relative root mean squared error (rRMSE), mean error (ME), mean percentage error (MPE), and Nash–Sutcliffe model efficiency coefficient (NSE).

$$ME = \frac{1}{n} * \sum (f_x - a_x) \quad (1)$$

$$MAE = \frac{1}{n} * \sum \text{abs}(f_x - a_x) \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n} * \sum (f_x - a_x)^2} \quad (3)$$

$$rRMSE = 100 * \frac{\sqrt{\frac{\sum (f_x - a_x)^2}{n-1}}}{f_x} \quad (4)$$

$$MPE = 100 * \frac{1}{n} * \sum \frac{(f_x - a_x)}{a_x} \quad (5)$$

$$NSE = 1 - \left(\frac{\sum (f_x - a_x)^2}{\sum (a_x - \bar{a})^2} \right) \quad (6)$$

where n is the number of samples, and f_x and a_x mean the estimated and actual intensity values for each sample, respectively. Whereas RMSE and MAE were used to document the degree of absolute errors in the modeling results, ME and MPE were used as indicators of overestimation (positive) and underestimation (negative) in the models. rRMSE is calculated with RMSE divided by the average of the references. According to Li et al. [60], an rRMSE less than 30% means that the model has fair performance. NSE shows how a model fits observations. Whereas a correlation coefficient assumes that data are unbiased, NSE proposed by Nash et al. [61] can be adopted for various models including nonlinear models [62]. Moriassi et al. [63] suggested four performance levels according to NSE values: unsatisfactory ($NSE \leq 0.5$); satisfactory ($0.5 < NSE \leq 0.65$); good ($0.65 < NSE \leq 0.75$); and very good ($0.75 < NSE \leq 1.0$).

These statistical values were used to evaluate our CNN-based models. RMSE and MAE show the overall accuracy of the models, which were used to select the most optimized version of each CNN-based model. The selected models were validated using the validation dataset according to the Saffir–Simpson scale which is the disaster-potential scale of a TC proposed by Simpson et al. [64] and widely used for defining the experimental danger stage based on wind speed. Their study showed the estimation performance according to the TC development stage.

4. Results

4.1. Modeling Performance

Table 4 shows the overall errors in terms of five accuracy metrics by model and Figure 6 shows the categorical performance differences between “Control” and “Control4channels” on the Saffir–Simpson scale. Whereas the “Control” and “Control4channels” models showed similar performances in terms of overall accuracy metrics, they yielded a considerable different MPE. The one-channel-based model (“Control”) showed a high MPE in weak TC stages and a low MPE in strong TC stages when compared to the multi-channel-based model. The IR-only model tended to overestimate the intensity of weak TCs and to underestimate that of strong TCs. On the other hand, the multi-channel-based model resulted in a relatively stable performance over all stages of TCs. This implies that the multi-infrared channel-based TC intensity estimation approach is a more reasonable method than the single-channel model. Because all the models tested in this study showed reasonable performances in terms of the ME, rRMSE, and NSE at around ± 1 kts, under 25% and over 0.75, respectively, MAE and RMSE were mainly used to determine the best models. In this study, the “2d3” model resulted in the highest performance with an MAE of 6.09 kts and an RMSE of 8.32 kts. The 2D-CNNs showed slightly better performance than the 3D-CNNs when using the same hyper-parameter conditions. Due to the large number of parameters to be optimized in the 3D-CNN models, the 3D-CNN models could only produce relatively

low performance when the same conditions of convolutional layer depth and epochs as the 2D-CNN models were used.

Table 4. Training (through parameterization) and validation results based on the test and validation datasets, respectively, for the nine CNN-based TC intensity estimation models. The model which resulted in the best validation accuracy for each of the 2D-CNN and 3D-CNN approaches is in bold. (Index (unit): MAE (kts), RMSE (kts), rRMSE (%), MPE (%) and NSE (0-1, unitless)). The best performing 2D-CNN and 3D-CNN based models are shown in bold.

Model ID	Training through Parameterization						Validation					
	MAE	RMSE	rRMSE	ME	MPE	NSE	MAE	RMSE	rRMSE	ME	MPE	NSE
Control	9.15	12.25	13.12	0.391	5.72	0.87	9.70	12.97	24.35	3.38	14.98	0.84
Control 4channels	8.96	12.28	13.15	-0.07	4.10	0.93	9.13	11.98	22.49	1.30	7.90	0.86
2d1	6.89	9.72	10.41	0.03	1.93	0.95	6.48	8.86	16.63	-0.06	2.02	0.93
2d2	7.51	10.19	10.92	-1.31	1.63	0.95	7.40	9.91	18.06	0.30	4.45	0.91
2d3	6.34	9.09	9.73	1.15	4.62	0.96	6.09	8.32	15.45	1.74	6.33	0.93
2d4	6.78	9.57	10.25	-0.49	2.63	0.96	6.11	8.74	15.94	1.23	6.19	0.93
3d1	9.11	11.97	12.81	1.19	5.18	0.93	9.16	11.79	22.13	2.16	9.69	0.87
3d2	8.96	11.98	12.82	-0.21	2.81	0.93	8.65	11.34	21.29	1.04	6.89	0.88
3d3	8.99	12.09	12.96	-0.05	4.37	0.93	8.93	11.72	22.01	1.30	8.76	0.87

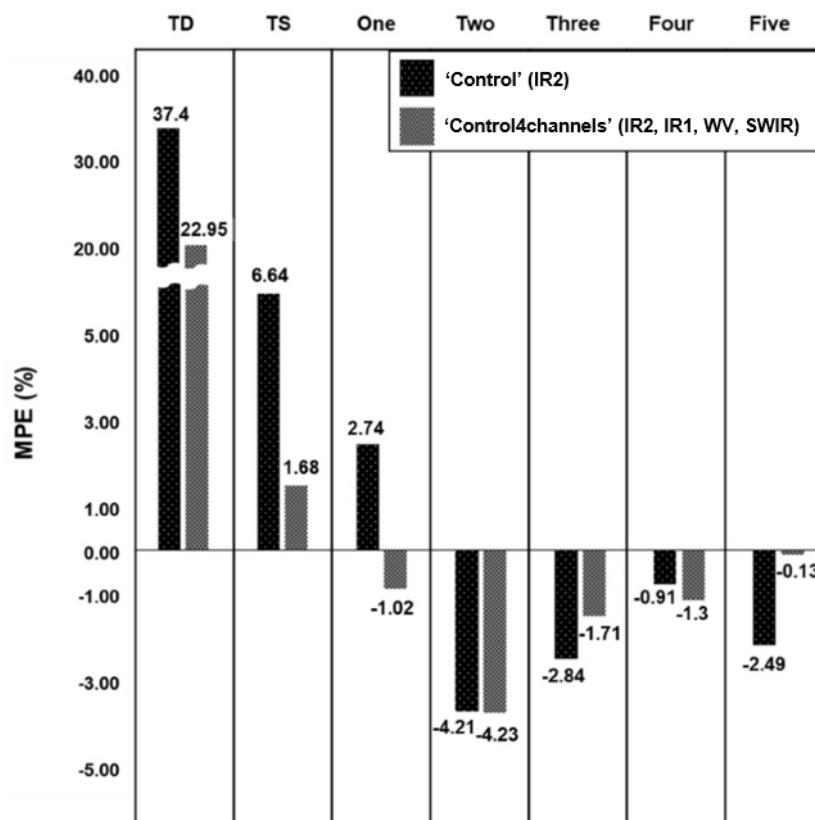


Figure 6. Mean percentage error (MPE) of 'Control' and 'Control4channels' on the Saffir–Simpson scale. Both models have the same CNN architecture except for input data. The 'Control' model overestimated the weak TCs whereas underestimating the strong TCs. 'Control4channels' model showed relatively stable performance when estimating TC intensity by scale.

The best 2D-CNN and 3D-CNN-based models showing the highest training accuracy with optimization were the 2d3 and 3d2 models, respectively. They were compared in order to evaluate the model performance using validation data. Figure 7 shows the overall validation results of the two selected models. Whereas the 3D-CNN-based model showed higher MAE and RMSE than the 2D-CNN model overall, there were different trends in the estimation results according to the development stage of the TCs. Table 5 shows the Saffir–Simpson typhoon scale-based categorical performance of the two models using ME, MPE, RMSE, and rRMSE metrics. Both models yielded the best performance with the lowest RMSE and rRMSE in Phase Five. However, the 2D-CNN model showed the largest RMSE with 10.11 kts at Phase One, whereas the 3D-CNN model yielded the largest RMSE of 17.22 kts at Phase Three. Whereas the 2D-CNN model showed almost positive values of MPE over strong TCs after Phase One, the 3D-CNN estimations resulted in almost negative values after the phase. This implies that the 3D-CNN model generally underestimated high-intensity TC when compared to the 2D-CNN model, possibly due to the limited optimization of the hyper-parameters in 3D-CNN.

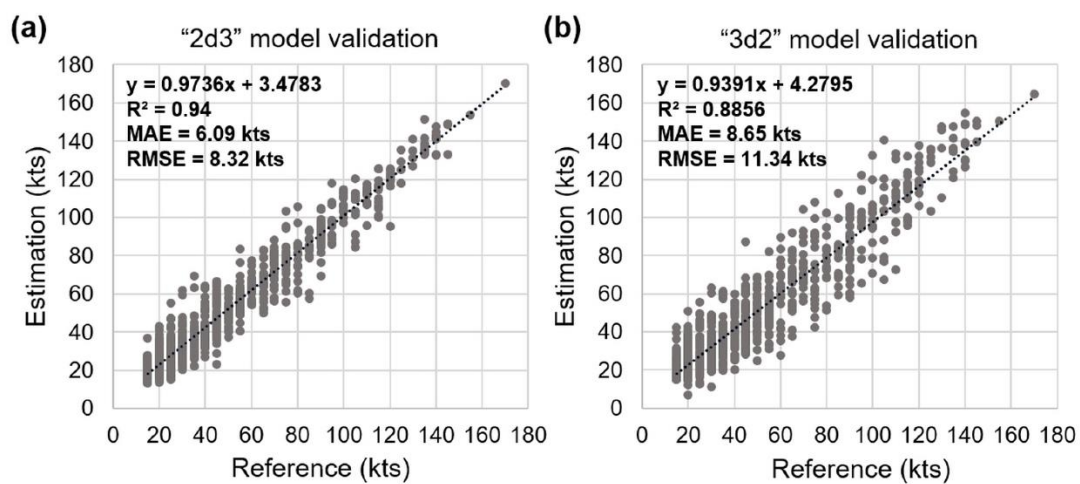


Figure 7. Validation results of the tropical cyclone intensity estimation based on the selected 2D-/3D-CNN models.

Table 5. Validation results according to Saffir–Simpson typhoon scale (Index (unit): ME (kts), MPE (%), RMSE (kts) and rRMSE (%)).

Category	Wind Speed (kts)	Samples	2D-CNN				3D-CNN			
			ME	MPE	RMSE	rRMSE	ME	MPE	RMSE	rRMSE
Tropical depression	≤33	288	3.38	16.35	7.88	33.69	4.59	22.41	9.72	41.56
Tropical storm	34–63	325	1.58	3.50	8.13	17.76	−1.05	−2.01	11.22	24.52
One	64–82	97	1.54	2.21	10.11	14.14	−0.28	−0.20	14.29	19.98
Two	83–95	78	0.53	0.50	9.17	10.21	−3.84	−4.37	15.14	16.86
Three	96–112	58	2.47	2.46	8.37	7.96	−2.59	−2.45	17.22	16.36
Four	113–136	53	0.11	−0.01	7.62	6.20	0.16	0.07	10.72	8.72
Five	≥137	16	0.11	0.09	5.41	3.74	−0.94	−0.53	8.26	5.71

It is not possible to directly compare the results among different studies since different areas and data were used. Nonetheless, it is meaningful to qualitatively compare the results with similar studies. Table 6 shows the TC intensity estimation performance of the existing models from the literature and the models proposed in this study. The multi-spectral infrared image-based models, which were proposed in this study, showed comparable and even better results when compared to the existing models.

Table 6. Comparison of model performances with the exiting satellite-based TC intensity estimation approaches.

Model.	Approach	Data Source	Inputs	Region	Covered Duration	RMSE (kts)
Ritchie et al. [16]	Statistical analysis	GOES-series	IR (10.7 μm)	Western North Pacific	2005–2011	12.7
Pradhan et al. [17]	2D-CNN	GOES-series	IR (10.7 μm)	Atlantic and Pacific	1999–2014	10.18
Combinido et al. [18]	2D-CNN	GMS-5, GOES-9, MTSAT-1R, MTSAT-2, Himawari-8	IR (11.0 μm)	Western North Pacific	1996–2016	13.23
Wimmers et al. [20]	2D-CNN	TRMM, Aqua, DMSP F8-F15, DMSP F16-F18	37 GHz, 85–92 GHz	Atlantic and Pacific	2007, 2010, 2012	14.3
2d3 (this study)	2D-CNN	COMS MI	IR2 (12.0 μm) IR1 (10.8 μm)	Western North Pacific	2011–2016	8.32
3d2 (this study)	3D-CNN		WV (6.7 μm) SWIR (3.7 μm)			11.34

Figure 8 shows the time series estimation results of the four models (i.e., “2d3”, “3d2”, “Control”, and “Control4channels” models) using validation samples from typhoons MUIFA in 2011, BOLAVEN in 2012, NOUL in 2015, and LIONROCK in 2016. The development phases of MUIFA in 2011 were overestimated in the 3d2 model and underestimated in the 2d3 model. The 3d2 model resulted in a higher variation in the extinction phase of the TC. In the cases of BOLAVEN in 2012 and LIONROCK in 2016, both models gave more stable performances compared to the other typhoon cases. In the case of typhoon NOUL in 2015, the 3D-CNN-based estimation results yielded high variation and bias, especially in the developed stage, whereas the 2D-CNN-based estimation results performed better than the 3D-CNN model in general. In particular, post-development TCs after 09/05/2015 18:00 UTC were underestimated by the 3D-CNN-based model in this case. The Control model trained with single 10.8 μm channel images underestimated the intensity of strong TCs from 01/08/2011 00:00 UTC to 06/08/2011 00:00 UTC of typhoon MUIFA in 2011. They underestimated about 5 to 20 kts when compared to the Control4channels model for strong TCs (>96 kts), which corresponds to the results in Figure 6.

In addition, we tested our models with typhoon HATO, KHANUN, LAN, and DAMREY in 2017 in order to evaluate the general estimation performance of our models to unseen TC cases. Figure 9 shows the time series estimation results of both models on the four main TCs in 2017. In the case of typhoon HATO in 2017, the 3d2 model has higher biases compare to the 2d3 model results. The intensities of the TCs in the developing phase from 21/08/2017/18:00 UTC to 22/08/2017 06:00 UTC was overestimated with the 3d2 model while the 2d3 model has relatively stable performance. The results for HATO are similar to typhoon MUIFA in 2011. On the other hand, in the case of typhoon KHANUN, LAN and DAMREY, both models have good estimation performances from developing to extinction phase of the typhoon. While the TCs from 18/10/2017 12:00 UTC to 20/10/2017 18:00 UTC of typhoon LAN in 2017 has a large fluctuation in intensity, both models have stably estimated the intensity. Similar to the MUIFA case, the Control model underestimated the intensity of TCs when compared to the Control4channels model especially for strong TCs over 64 kts (i.e., 20/10/2017 06:00 UTC to 22/10/2017 06:00 UTC of typhoon LAN and 03/11/2017 00:00 UTC to 03/11/2017 12:00 UTC of typhoon DAMREY). The ME, MPE, MAE, RMSE, and rRMSE of the hindcast validation of the 2d3

model for 2017 cases were 2.15 kts, 4.92%, 6.83 kts, 8.31 kts, and 13.63%, respectively, which were comparable to the validation results. Similarly, the metrics of the 3d3 model for 2017 were 1.34 kts, 4.18%, 8.6 kts, 11.31 kts, 18.54%, respectively. These results confirm that the models proposed in this study successfully learned various TC convective patterns by intensity from the past data, which can be generalized to estimate the intensity for new TCs in the future.

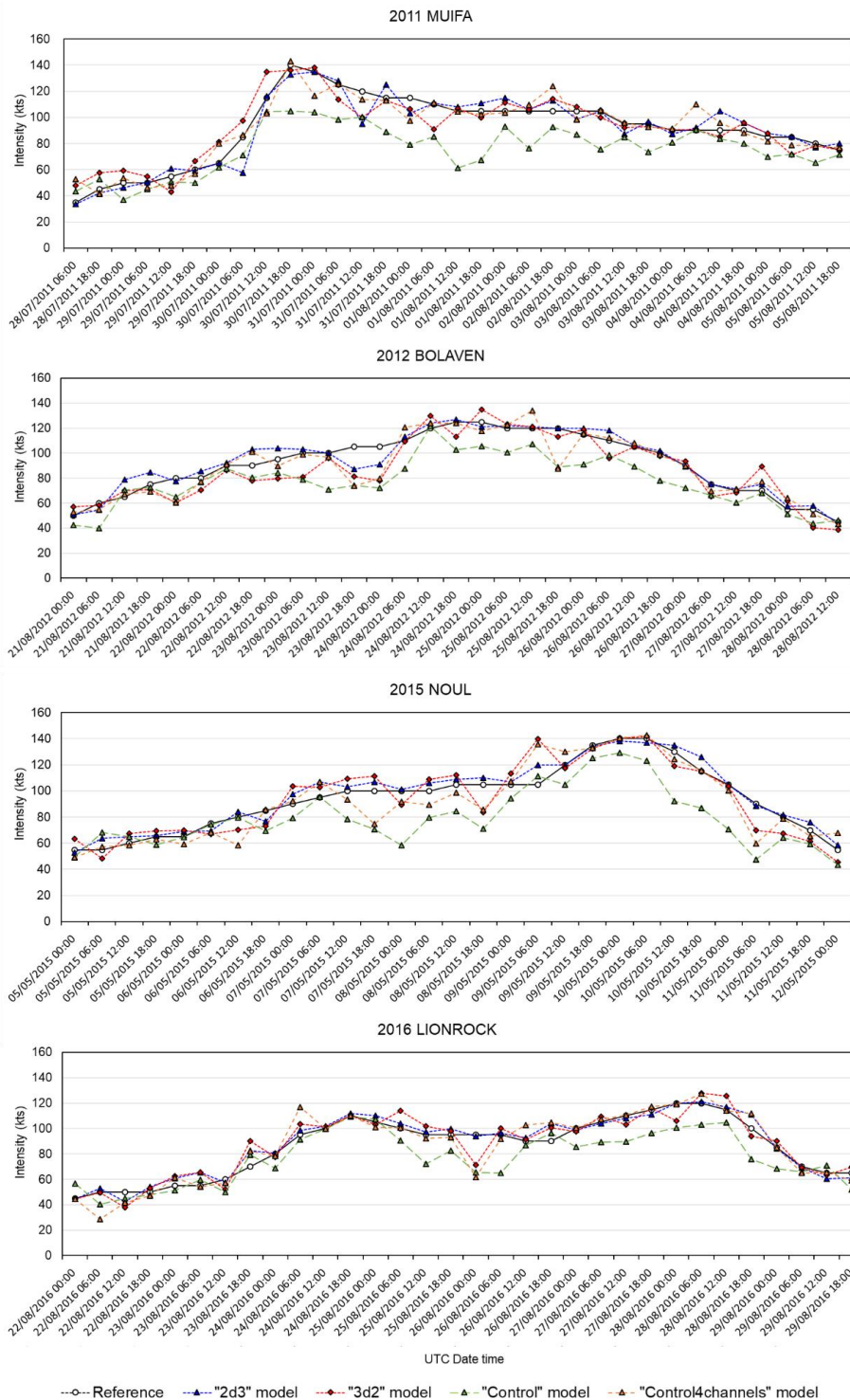


Figure 8. Time-series comparison of the validation results of four models (‘2d3’ and ‘3d2’ models which are proposed by this study and ‘Control’ and ‘Control4channels’ model, respectively) with JTWC best track data of typhoon MUIFA in 2011, BOLAVEN in 2012, NOUL in 2015, and LIONROCK in 2016.

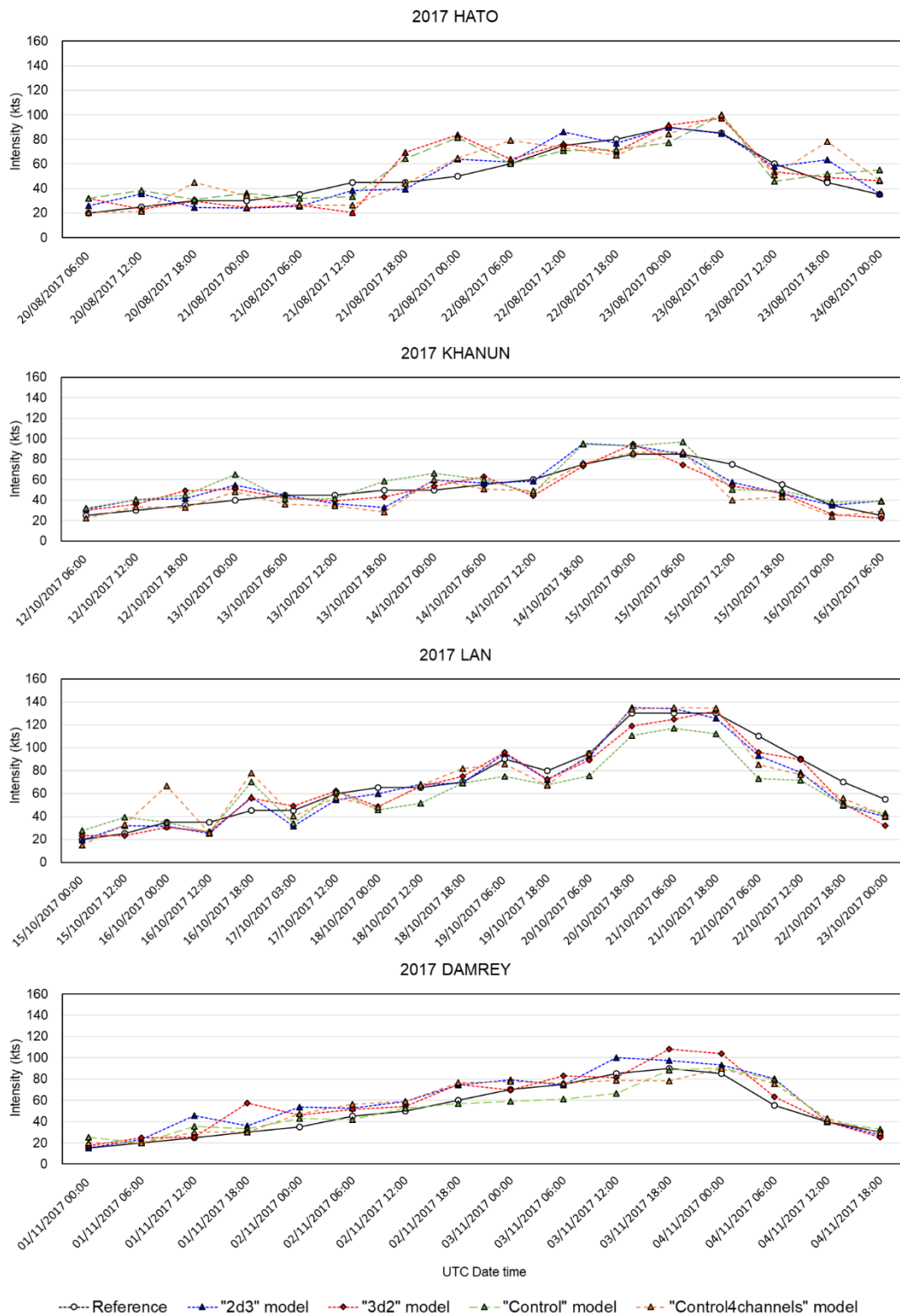


Figure 9. Time-series comparison of the additional hindcast validation results of four models (‘2d3’ and ‘3d2’ models which are proposed by this study and ‘Control’ and ‘Control4channels’ model, respectively) with JTWC best track data of typhoon HATO, KHANUN, LAN and DAMREY in 2017.

5. Discussion

5.1. Visualization

Deep learning methods, including CNN, are widely called “Black box” because they struggle to identify the causal relationship among the variables and the model parameters. Zeiler et al. [65] proposed an innovative visualization method called a “heat map”. It is extracted based on the sum of the activation maps in the last convolutional layer [65,66]. In this paper, we resize the heat map to the size of the raw input data to intuitively interpret the images. Some TC cases of typhoon BOLAVEN in 2012 and corresponding heat maps are shown in Figure 10. Due to the difference in the kernel processes between 2D-CNN and 3D-CNN, 2D-CNN has only one heat map for the multi-channel input data, whereas 3D-CNN has more than two heat maps (i.e., corresponding to the size of input channels). In this study, we designed the 3D-CNN model to extract four heat maps to understand and interpret the effect of each layer. Whereas most high-intensity TCs showed a clear whirling pattern in the center of the TCs, the TC pattern 23/08/2012 18:00 UTC in Figure 10 looked like a cluster of clouds rather than a spiral pattern. Because most high-intensity TCs have a clear spiral pattern, this anomalous pattern might result in the models underestimating the intensity.

It is obvious that the higher the intensity of a TC, the higher the importance around the center of the TC. Weak TCs such as the 19/08/2012 06:00 UTC case in Figure 10 do not have a concentrated convection cloud center, but the heat map showed patterns of clusters with small convective clouds around the edge of TCs. Strong TCs such as the 24/08/2012 06:00 UTC case in Figure 10 typically have a whirlpool shape around the center. The areas of high values in the heat map (i.e., important part of a TC image) have a similar shape to the main pattern of the TC proposed in Dvorak technique’s figures which were categorically suggested according to the T-number and TC intensity [67]. Figure 11 depicts the “3d2” model-based significant regions of each TC according to each development stage and the corresponding Dvorak’s pattern. The red area indicates the most important region, which has high values in the heat map. The red regions are like the TC patterns used in the Dvorak technique-based algorithm [6,7,67]. This implies that our CNN-based model has a ability to objectively replicate the Dvorak technique.

5.2. Interpretation of Relationship between Multi-Spectral TC Images and Intensity

In this study, we proposed several 2D-/3D-CNN based TC intensity estimation models that use multi-spectral satellite images. Whereas long-wave infrared images have been mainly used to estimate TC intensity, the lower atmosphere has a significant effect in estimating TC intensity, especially for high-intensity TCs [68]. Each infrared channel does not represent the exact altitude of the atmosphere. However, thanks to the differences in the wavelengths of channels, different convective patterns can be yielded from multiple channels. Through the results of ‘Control’ and ‘Control4channels’ models in Section 4.1, it is considered that the multi-spectral image-based CNN models showed better or comparable performance compared to the existing methods when using only a single long-wavelength infrared image.

The stronger the intensity of a TC, the stronger the vortex around the TC center of the lower atmosphere. Whereas the Dvorak technique has been widely used to estimate the intensity of TCs with cloud top images (i.e., long-wavelength infrared images), actual TC intensity is significantly influenced by lower layers of the atmosphere [68–70]. Cha et al. [68] showed that an improved initialization method of TC vortex with consecutive cycle simulations of a dynamical model could realistically enhance TC intensity by improving the initial three-dimensional structure of TC, in particular, stronger tangential and radial winds at the lower level. Thus, multi-levels infrared images including low and mid-levels could contribute to the realistic intensity estimation in operational TC centers. The 3D-CNN-based multi-layered heat maps can reasonably represent the vertically coupled TC structure between lower and upper levels, which was also confirmed by numerical simulations with dynamical TC models.

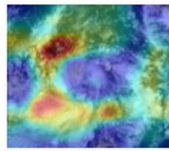
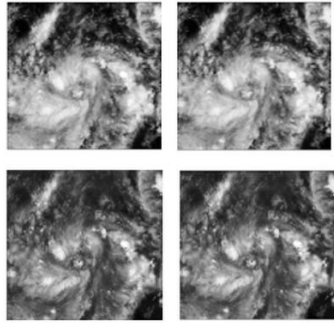
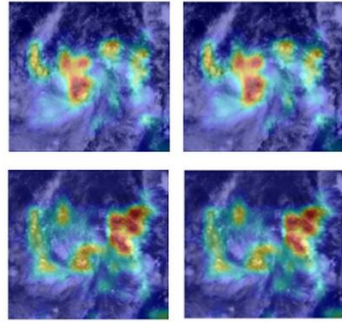
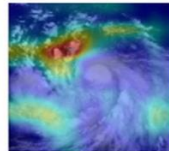
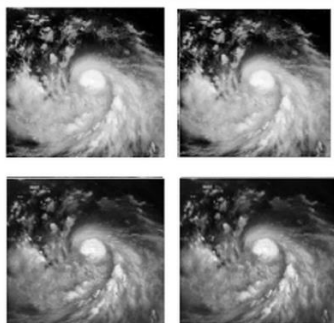
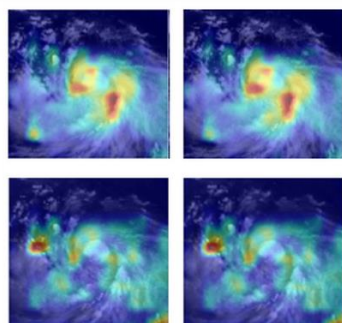
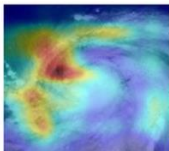
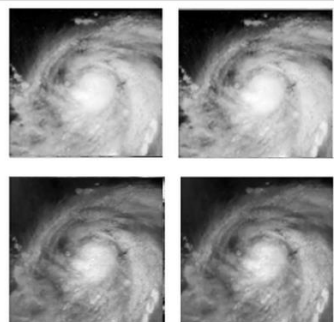
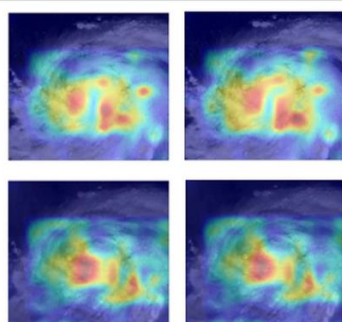
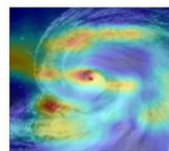
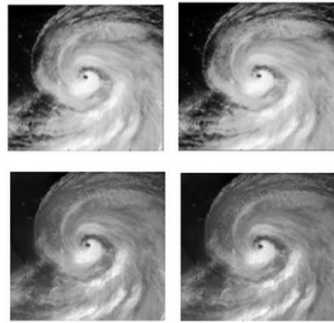
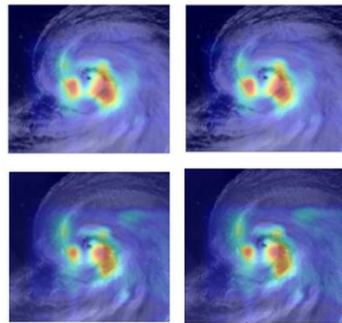
	"2d3" model	RAW	"3d2" model
Intensity	17.56 kts	20 kts	24.02 kts
19/08/2012 06:00 UTC			
Intensity	48.75 kts	60 kts	54.48 kts
21/08/2012 06:00 UTC			
Intensity	77.05 kts	105 kts	87.72 kts
23/08/2012 18:00 UTC			
Intensity	117.77 kts	110 kts	112.85 kts
24/08/2012 06:00 UTC			

Figure 10. Validation results and heat maps of the 2D-CNN based "2d3" model and 3D-CNN based "3d2" model in case of BOLAVEN in 2012. The more reddish the shade on the heat map, the greater its importance in the model estimation process. Upper left: IR2 (channel 1), Upper right: IR1 (channel 2), Lower left: wv (channel 3), Lower right: swir (channel 4) in Raw and "3d2" column.

Stage (#T)	Intensity (kts)	Model based activated region		Dvorak pattern	Stage (#T)	Intensity (kts)	Model based activated region		Dvorak pattern
		Heat map + IR2	Activated region				Heat map + IR2	Activated region	
Weak TD (T1.0-1.5)	20				One (T4.0-4.5)	65			
	20					70			
	25					90			
TD (T2)	30				Three (T5.5)	105			
TS (T2.5)	35				Four (T6.0-6.5)	115			
TS (T3)	45					120			
					140				
TS (T3.5)	55				Five (T7.0-8.5)	145			

Figure 11. Matching between the Dvorak technique-based TC patterns [57] and the CNN model-based activated regions according to TC categories. The “3d2” model-based activated region in the IR2 channel is marked in red. The 1st column shows heat maps with the IR2 channel and the most significant region is extracted using the upper 3% of the heat map values. It is marked as the red region in the 2nd column. The 3rd column shows the TC patterns of each categorical TC used in the Dvorak technique.

The integrated map by distance from the center of Typhoon MUIFA using the multi-layered heat maps is shown in Figure 12. Heat map values that had the same pixel-based distance from the center were summed in the integrated heat map with the distance measurements in the x-axis. The resultant integrated map was originally a 4 × 71 sized matrix, which was then resized to a 28 × 71 sized matrix using linear interpolation for a more intuitive interpretation of the vertical trends of a TC. The integrated heat map shows the difference in the region of importance according to the distance from the center and the relative atmospheric height. Typhoon MUIFA in 2011 and Typhoon BOLAVEN in 2012 were tested when identifying the vertical structure of TCs according to their intensity. Figure 13 shows the integrated heat maps according to intensity. With weak TCs, the heat map importance was concentrated far from the center of the TCs. This confirms the typical patterns of weak TCs, which have scattered partial convective clouds [7,9,15,67]. On the other hand, strong TC

based heat maps showed a tendency for the heat map importance to focus on the center of TCs in the lower layer. This also corresponds to the vertical behavior of strong TCs (i.e., central concentrated convection) [68–70]. These results verified that the 3D-CNN-based estimation models considered the geophysical characteristics of TCs, which in turn affected the estimation results.

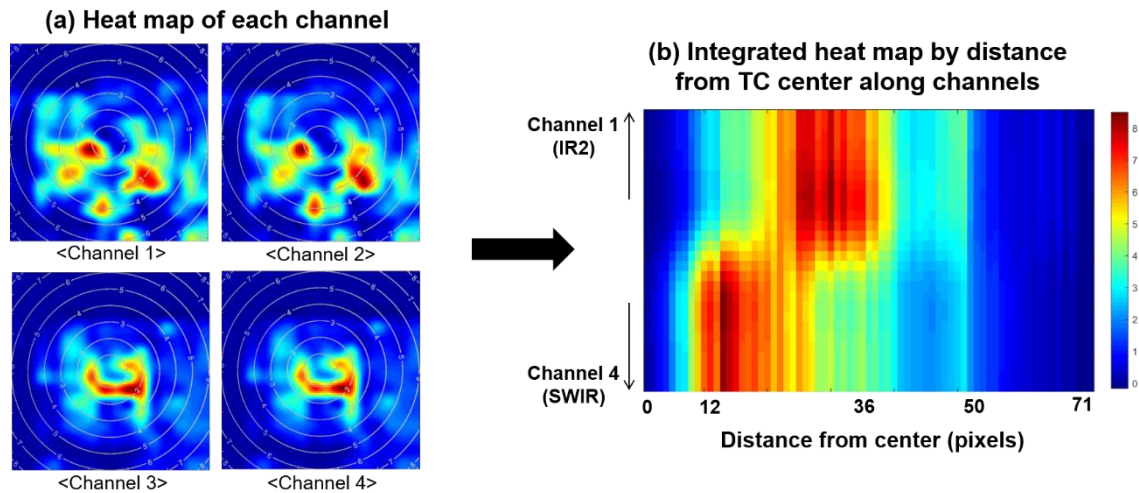


Figure 12. Four-layer heat maps and the integrated heat map using pixel distance from the TC center using Typhoon MUIFA on 05/06/2011 06:00 UTC (85 kts). (a) Four-layer heat maps, where the white contour lines show the distance (pixel) from the TC center. The channels 1 to 4 correspond to IR2, IR1, WV, and SWIR, respectively. (b) Integrated multi-dimensional heat map. The values at the same distance from the center in all layers are added in the integrated heat map (1 pixel in (b) corresponds to about 12 km).

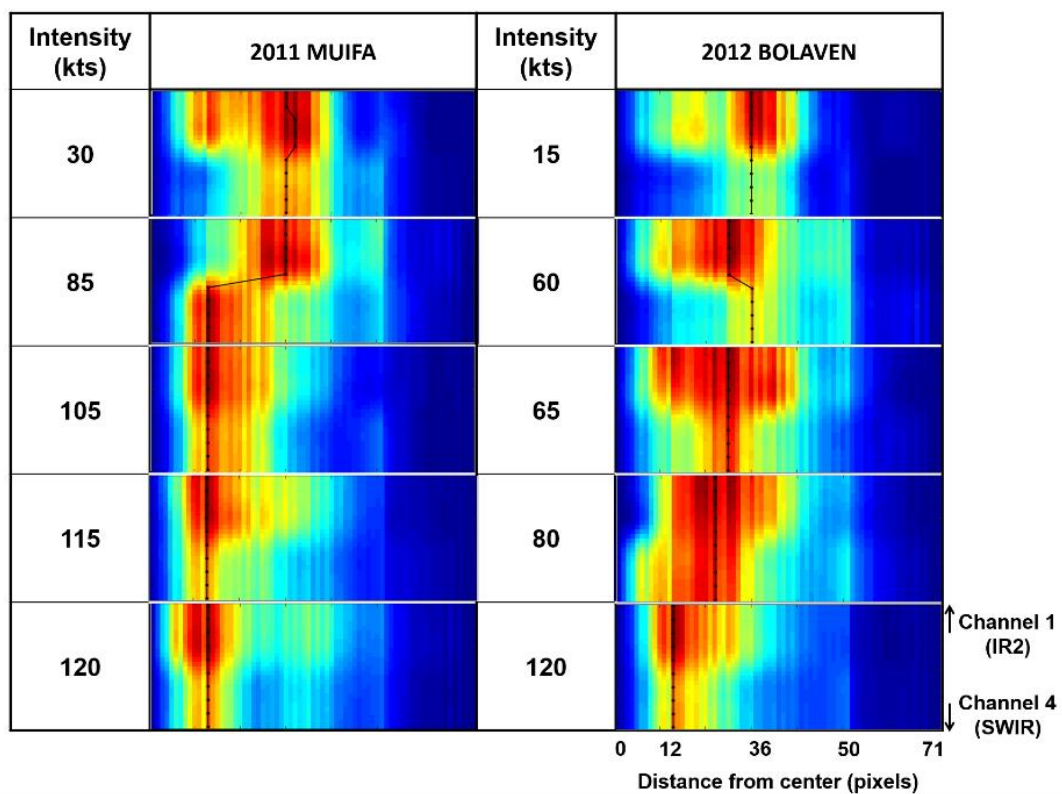


Figure 13. The integrated heat maps of validation typhoon cases—typhoon MUIFA in 2011 and typhoon BOLAVEN in 2012. The highest heat map values of each row are marked and connected using black lines. The stronger TC has a high heat map value near the center in the lower layer.

5.3. Novelty and Limitation

In this study, we proposed a CNN-based TC intensity estimation approach using multi-spectral satellite images. We found out that CNN-based models successfully mimicked the manual TC intensity estimation algorithms, and the proposed multi-spectral approach made a significant contribution to TC intensity estimation. In particular, multi-channel infrared data-based TC intensity estimation models showed more stable performance when compared to the one-channel-based model, which has been widely used for estimating TC intensity up to now. It was verified that the convective cloud patterns of the middle and lower layer, as well as the upper convective cloud pattern, has considerable effects on reliable estimations of TC intensity. The significant regions of each vertical layer using heat maps were also identified using the 3D-CNN-based model.

However, there are still some limitations in the approach proposed in this research: 1) since CNNs, in particular, 3D-CNNs, require significant computational demand in terms of memory and running time, it is often difficult to fully optimize the hyper-parameters of the CNN models, and 2) it is hard to clearly understand how the models consider input data through the neural networks to produce reasonable results. Whereas various combinations of hyper-parameters were tested to identify an optimum model, it was not possible to test all available ones for our dataset. There is still a possibility that there may be a better performing model, especially one using 3D-CNN, which was not tested. Significant computational demand is one of the main problems in deep learning-based research. Whereas deep learning often has high uncertainty in the modeling process due to its dependency on training data, it can be helpful when we need to examine a huge amount of unknown information [71,72]. However, it is also difficult to identify and quantify the effects of training data on CNN models. Whereas several methods, such as heat maps and occlusion maps, have been proposed for the interpretation of CNN results, it is still not clear how CNN models recognize the pattern of input images [73–76].

The research findings from this study deserve further investigation. In future work, hyper-parameter-optimization of the 3D-CNN model using cost-effective approaches (e.g., auto-parameterization tools such as AutoKeras and Keras-tuner) should be conducted. A fully optimized 3D-CNN based model may provide a more stable and robust performance than the present model. In addition, numerical model-derived multi-dimensional TC variables can be examined in conjunction with 3D-CNN models, providing an in-depth understanding of the relationship between three-dimensional TC structure and intensity. While this study is focused on estimating TC intensity, deep learning can be adopted for forecasting TC intensity (e.g., 12, 24, and 36 h), which needs further investigation.

6. Conclusions

Since the widely-used TC intensity estimation method, the Dvorak technique, is a manual algorithm, there is a need to have a standardized and objective way to quantify TC intensity for end-users such as Typhoon centers and forecasters. In this research, both 2D-CNN and 3D-CNN approaches were carefully evaluated to analyze the interrelationships between multi-channel-TC images and their intensity. The 2D-CNN-based approach resulted in a very good performance with an RMSE of 8.32 kts, which is very competitive when compared to the existing approaches. Although the 3D-CNN based model yielded an RMSE of 11.34 kts, which is higher than that of the 2D-CNN based model, its performance was still comparable with the existing approaches. For the hindcast validation, both 2D- and 3D-CNN models produced similar results (i.e., RMSE of 8.31 and 11.31 kts, respectively), which proved the robustness of the proposed models. Our TC intensity estimation model based on multi-spectral channels showed better performance (~35.9%) when compared to the existing approach with a single-spectral channel [17] based on the same datasets.

Furthermore, how the 3D-CNN model regressed the TC intensity using satellite-based multiple infrared images was examined using heat maps. Through this experiment, it was found that the pattern of the inner core part of a TC was closely related to TC intensity. In particular, the proposed

3D-CNN model enabled the identification of the effective region of each channel and the different vertical pattern of the significant regions, as well as a horizontal pattern according to TC intensity.

Author Contributions: J.L. led manuscript writing and contributed to data analysis and research design. J.I. supervised this study, contributed to the research design, manuscript writing and discussion of the results, and served as the corresponding author. D.-H.C. contributed to the research design and discussion of the results. H.P. and S.S. contributed to data analysis and model design. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT (NRF-2016M3C4A7952637), by the National Research Foundation of Korea (NRF- 2017M1A3A3A02015981), by Ministry of Interior and Safety (MOIS), Korea (2019-MOIS32-015), and by the Ministry of Science and ICT (MSIT), Korea (IITP-2019-2018-0-01424).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Seneviratne, S.I.; Nicholls, N.; Easterling, D.; Goodess, C.M.; Kanae, S.; Kossin, J.; Luo, Y.; Marengo, J.; McInnes, K.; Rahimi, M.; et al. Changes in climate extremes and their impacts on the natural physical environment. In *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Special Report of the Intergovernmental Panel on Climate Change*; Cambridge University Press: Cambridge, UK, 2012; pp. 109–230.
2. Hoegh-Guldberg, O.; Jacob, D.; Taylor, M. Impacts of 1.5 C Global Warming on Natural and Human Systems. In *Global Warming of 1.5 C: An IPCC Special Report on the Impacts of Global Warming of 1.5 C above Pre-Industrial Levels and Related Global Greenhouse Gas Emission Pathways, in the Context of Strengthening the Global Response to the Threat of Climate Change, Sustainable Development, and Efforts to Eradicate Poverty*; MassonDelmotte, V., Zhai, P., Pörtner, H.O., Roberts, D., Skea, J., Shukla, P.R., Pirani, A., Moufouma-Okia, W., Péan, C., Pidcock, R., et al., Eds.; 2018; pp. 175–311, in press.
3. Knutson, T.R.; McBride, J.L.; Chan, J.; Emanuel, K.; Holland, G.; Landsea, C.; Held, I.; Kossin, J.P.; Srivastava, A.K.; Sugi, M. Tropical cyclones and climate change. *Nat. Geosci.* **2010**, *3*, 157–163. [[CrossRef](#)]
4. World Bank. Information, Communication Technologies, and infoDev (Program). In *Information and Communications for Development 2012: Maximizing Mobile*; World Bank Publications: Washington, DC, USA, 2012.
5. Mendelsohn, R.; Emanuel, K.; Chonabayashi, S.; Bakkensen, L. The impact of climate change on global tropical cyclone damage. *Nat. Clim. Chang.* **2012**, *2*, 205–209. [[CrossRef](#)]
6. Schmetz, J.; Tjemkes, S.A.; Gube, M.; Van de Berg, L. Monitoring deep convection and convective overshooting with METEOSAT. *Adv. Space Res.* **1997**, *19*, 433–441. [[CrossRef](#)]
7. Dvorak, V.F. Tropical cyclone intensity analysis using satellite data. In *NOAA Technical Report NESDIS, 11*; US Department of Commerce, National Oceanic and Atmospheric Administration, National Environmental Satellite, Data, and Information Service: Washington, DC, USA, 1984; pp. 1–47.
8. Menzel, W.P.; Purdom, J.F. Introducing GOES-I: The First of a New Generation of Geostationary Operational Environmental Satellites. *Bull. Am. Meteorol. Soc.* **1994**, *75*, 757–781. [[CrossRef](#)]
9. Dvorak, V.F. Tropical Cyclone Intensity Analysis and Forecasting from Satellite Imagery. *Mon. Weather Rev.* **1975**, *103*, 420–430. [[CrossRef](#)]
10. Olander, T.L.; Velden, C.S. The advanced Dvorak technique: Continued development of an objective scheme to estimate tropical cyclone intensity using geostationary infrared satellite imagery. *Weather Forecast.* **2007**, *22*, 287–298. [[CrossRef](#)]
11. Olander, T.L.; Velden, C.S. The Advanced Dvorak Technique (ADT) for Estimating Tropical Cyclone Intensity: Update and New Capabilities. *Weather Forecast.* **2019**, *34*, 905–922. [[CrossRef](#)]
12. Velden, C.S.; Olander, T.L.; Zehr, R.M. Development of an Objective Scheme to Estimate Tropical Cyclone Intensity from Digital Geostationary Satellite Infrared Imagery. *Weather Forecast.* **1998**, *13*, 172–186. [[CrossRef](#)]
13. Velden, C.; Harper, B.; Wells, F.; Beven, J.L.; Zehr, R.; Olander, T.; Mayfield, M.; Guard, C.C.; Lander, M.; Edson, R.; et al. The Dvorak Tropical Cyclone Intensity Estimation Technique: A Satellite-Based Method that Has Endured for over 30 Years. *Bull. Am. Meteorol. Soc.* **2006**, *87*, 1195–1210. [[CrossRef](#)]

14. Velden, C.S.; Hayden, C.M.; Nieman, S.J.W.; Paul Menzel, W.; Wanzong, S.; Goerss, J.S. Upper-tropospheric winds derived from geostationary satellite water vapor observations. *Bull. Am. Meteorol. Soc.* **1997**, *78*, 173–195. [[CrossRef](#)]
15. Piñeros, M.F.; Ritchie, E.A.; Tyo, J.S. Objective measures of tropical cyclone structure and intensity change from remotely sensed infrared image data. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 3574–3580. [[CrossRef](#)]
16. Ritchie, E.A.; Wood, K.M.; Rodríguez-Herrera, O.G.; Piñeros, M.F.; Tyo, J.S. Satellite-derived tropical cyclone intensity in the north pacific ocean using the deviation-angle variance technique. *Weather Forecast.* **2014**, *29*, 505–516. [[CrossRef](#)]
17. Pradhan, R.; Aygun, R.S.; Maskey, M.; Ramachandran, R.; Cecil, D.J. Tropical cyclone intensity estimation using a deep convolutional neural network. *IEEE Trans. Image Process.* **2018**, *27*, 692–702. [[CrossRef](#)] [[PubMed](#)]
18. Combinido, J.S.; Mendoza, J.R.; Aborot, J. A Convolutional Neural Network Approach for Estimating Tropical Cyclone Intensity Using Satellite-based Infrared Images. In Proceedings of the 2018 24th ICPR, Beijing, China, 20–24 August 2018.
19. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the ICLR 2015, San Diego, CA, USA, 7–9 May 2015.
20. Wimmers, A.; Velden, C.; Cossuth, J.H. Using deep learning to estimate tropical cyclone intensity from satellite passive microwave imagery. *Mon. Weather Rev.* **2019**, *147*, 2261–2282. [[CrossRef](#)]
21. Elsberry, R.L.; Jeffries, R.A. Vertical wind shear influences on tropical cyclone formation and intensification during TCM-92 and TCM-93. *Mon. Weather Rev.* **1996**, *124*, 1374–1387. [[CrossRef](#)]
22. LeCun, Y.; Boser, B.E.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.E.; Jackel, L.D. Handwritten digit recognition with a back-propagation network. In *Advances in Neural Information Processing Systems*; MITpress: Cambridge, MA, US, 1990; pp. 396–404.
23. Yang, H.; Yu, B.; Luo, J.; Chen, F. Semantic segmentation of high spatial resolution images with deep neural networks. *GISci. Remote Sens.* **2019**, *56*, 749–768. [[CrossRef](#)]
24. Liu, T.; Abd-Elrahman, A.; Morton, J.; Wilhelm, V.L. Comparing fully convolutional networks, random forest, support vector machine, and patch-based deep convolutional neural networks for object-based wetland mapping using images from small unmanned aircraft system. *GISci. Remote Sens.* **2018**, *55*, 243–264. [[CrossRef](#)]
25. Kim, M.; Lee, J.; Im, J. Deep learning-based monitoring of overshooting cloud tops from geostationary satellite data. *GISci. Remote Sens.* **2018**, *55*, 763–792. [[CrossRef](#)]
26. Yu, X.; Wu, X.; Luo, C.; Ren, P. Deep learning in remote sensing scene classification: A data augmentation enhanced convolutional neural network framework. *GISci. Remote Sens.* **2017**, *54*, 741–758. [[CrossRef](#)]
27. Ou, M.L.; Jae-Gwang-Won, S.R.C. Introduction to the COMS Program and its application to meteorological services of Korea. In Proceedings of the 2005 EUMETSAT Meteorological Satellite Conference, Dubrovnik, Croatia, 19 September 2005; pp. 19–23.
28. Ralph, F.M.; Neiman, P.J.; Wick, G.A. Satellite and CALJET aircraft observations of atmospheric rivers over the eastern North Pacific Ocean during the winter of 1997/98. *Mon. Weather Rev.* **2004**, *132*, 1721–1745. [[CrossRef](#)]
29. Durrý, G.; Amarouche, N.; Zéninari, V.; Parvitte, B.; Lebarbu, T.; Ovarlez, J. In situ sensing of the middle atmosphere with balloonborne near-infrared laser diodes. *Spectrochim. Acta Part A* **2004**, *60*, 3371–3379. [[CrossRef](#)] [[PubMed](#)]
30. Lee, T.F.; Turk, F.J.; Richardson, K. Stratus and fog products using GOES-8–9 3.9- μ m data. *Weather Forecast.* **1997**, *12*, 664–677. [[CrossRef](#)]
31. Nakajima, T.; King, M.D. Determination of the Optical Thickness and Effective Particle Radius of Clouds from Reflected Solar Radiation Measurements. Part I: Theory. *J. Atmos. Sci.* **1990**, *47*, 1878–1893. [[CrossRef](#)]
32. Rosenfeld, D.; Woodley, W.L.; Lerner, A.; Kelman, G.; Lindsey, D.T. Satellite detection of severe convective storms by their retrieved vertical profiles of cloud particle effective radius and thermodynamic phase. *J. Geophys. Res. Atmos.* **2008**, *113*, D04208. [[CrossRef](#)]
33. Martins, J.V.; Marshak, A.; Remer, L.A.; Rosenfeld, D.; Kaufman, Y.J.; Fernandez-Borda, R.; Koren, I.; Correia, A.L.; Artaxo, P. Remote sensing the vertical profile of cloud droplet effective radius, thermodynamic phase, and temperature. *Atmos. Chem. Phys.* **2011**, *11*, 9485–9501. [[CrossRef](#)]

34. Lowry, M.R. Developing a Unified Superset in Quantifying Ambiguities among Tropical Cyclone Best Track Data for the Western North Pacific. Master's Thesis, Dept. Meteorology, Florida State University, Tallahassee, FL, USA, 2008.
35. Guard, C.P.; Carr, L.E.; Wells, F.H.; Jeffries, R.A.; Gural, N.D.; Edson, D.K. Joint Typhoon Warning Center and the Challenges of Multibasin Tropical Cyclone Forecasting. *Weather Forecast.* **1992**, *7*, 328–352. [[CrossRef](#)]
36. Knapp, K.R.; Kruk, M.C.; Levinson, D.H.; Diamond, H.J.; Neumann, C.J. The International Best Track Archive for Climate Stewardship (IBTrACS). *Bull. Am. Meteorol. Soc.* **2010**, *91*, 363–376. [[CrossRef](#)]
37. Longadge, R.; Dongre, S. Class Imbalance Problem in Data Mining Review. *arXiv* **2013**, arXiv:1305.1707.
38. Romero, A.; Gatta, C.; Camps-Valls, G. Unsupervised deep feature extraction for remote sensing image classification. *IEEE Trans. Image Process.* **2015**, *54*, 1349–1362. [[CrossRef](#)]
39. LeCun, Y.; Bengio, Y.; Hinton, H. Deep Learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)]
40. Hopfield, J.J.; Tank, D.W. “Neural” computation of decisions in optimization problems. *Biol. Cybern.* **1985**, *52*, 141–152. [[PubMed](#)]
41. Amit, D.J. Modeling Simplified Neurophysiological Information. In *Modeling Brain Function: The World of Attractor Neural Networks*, 1st ed.; University Cambridge Press: Cambridge, UK, 1992.
42. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [[CrossRef](#)]
43. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the NIPS 2012, Lake Tahoe, NV, USA, 3–8 December 2012; pp. 1097–1105.
44. Kamnitsas, K.; Ledig, C.; Newcombe, V.F.; Simpson, J.P.; Kane, A.D.; Menon, D.K.; Rueckert, D.; Glocker, B. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* **2017**, *36*, 61–78. [[CrossRef](#)] [[PubMed](#)]
45. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; Jeroen, A.W.M.L.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [[CrossRef](#)]
46. Key, J.R.; Santek, D.; Velden, C.S.; Bormann, N.; Thepaut, J.N.; Riishojgaard, L.P.; Zhu, Y.; Menzel, W.P. Cloud-drift and water vapor winds in the polar regions from MODIS. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 482–492. [[CrossRef](#)]
47. Sharif Razavian, A.; Azizpour, H.; Sullivan, J.; Carlsson, S. CNN features off-the-shelf: An astounding baseline for recognition. In Proceedings of the IEEE Conference on CVPR Workshop, Columbia, OH, USA, 23 June 2014; pp. 806–813.
48. Liu, Y.; Racah, E.; Correa, J.; Khosrowshahi, A.; Lavers, D.; Kunkel, K.; Wehner, M.; Collins, W. Application of Deep Convolutional Neural Networks for Detecting Extreme Weather in Climate Datasets. *arXiv* **2016**, arXiv:1605.01156.
49. Toms, B.A.; Kashinath, K.; Yang, D. Deep Learning for Scientific Inference from Geophysical Data: The Madden-Julian Oscillation as a Test Case. *arXiv* **2019**, arXiv:1902.04621.
50. Zhou, Y.; Luo, J.; Yang, Y.; Chen, Y.; Wu, W. Long-short-term-memory-based crop classification using high-resolution optical images and multi-temporal SAR data. *GISci. Remote Sens.* **2019**, *56*, 1170–1191. [[CrossRef](#)]
51. Lee, C.; Sohn, E.; Park, J.; Jang, J. Estimation of soil moisture using deep learning based on satellite data: A case study of South Korea. *GISci. Remote Sens.* **2019**, *56*, 43–67. [[CrossRef](#)]
52. LeCun, Y.; Bengio, Y. Convolutional networks for images, speech, and time series. In *The Handbook of Brain Theory and Neural Networks*; MIT Press: Cambridge, MA, USA, 1995; Volume 3361.
53. Lee, H.; Kwon, H. Going Deeper With Contextual CNN for Hyperspectral Image Classification. *IEEE Trans. Image Process.* **2017**, *26*, 4843–4855. [[CrossRef](#)]
54. Su, H.; Maji, S.; Kalogerakis, E.; Learned-Miller, E. Multi-View Convolutional Neural Networks for 3D Shape Recognition. In Proceedings of the IEEE ICCV 2015, Santiago, Chile, 7–13 December 2015; pp. 945–953.
55. Ji, S.; Xu, W.; Yang, M.; Yu, K. 3D convolutional neural networks for human action recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 221–231. [[CrossRef](#)] [[PubMed](#)]
56. Soós, B.G.; Rák, Á.; Veres, J.; Cserey, G. GPU Boosted CNN Simulator Library for Graphical Flow-Based Programmability. *EURASIP J. Adv. Signal Process.* **2009**, 1–11.
57. Potluri, S.; Fasih, A.; Vutukuru, L.K.; Al Machot, F.; Kyamakya, K. CNN based high performance computing for real time image processing on GPU. In Proceedings of the Joint INDS'11 & ISTET'11, Klagenfurt, Austria, 25–27 July 2011.

58. Li, D.; Chen, X.; Becchi, M.; Zong, Z. Evaluating the Energy Efficiency of Deep Convolutional Neural Networks on CPUs and GPUs. In Proceedings of the 2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom), Atlanta, GA, USA, 8–10 October 2016.
59. Hadjis, S.; Zhang, C.; Mitliagkas, I.; Iyer, D.; Ré, C. Omnivore: An Optimizer for Multi-Device Deep Learning on Cpus and Gpus. *arXiv* **2016**, arXiv:1606.04487.
60. Li, M.F.; Tang, X.P.; Wu, W.; Liu, H.B. General models for estimating daily global solar radiation for different solar radiation zones in mainland China. *Energy Convers. Manag.* **2013**, *70*, 139–148. [[CrossRef](#)]
61. Nash, J.E.; Sutcliffe, J.V. River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.* **1970**, *10*, 282–290. [[CrossRef](#)]
62. McCuen, R.H.; Knight, Z.; Cutter, A.G. Evaluation of the Nash–Sutcliffe Efficiency Index. *J. Hydrol. Eng.* **2006**, *11*, 597–602. [[CrossRef](#)]
63. Moriasi, D.N.; Arnold, J.G.; Van Liew, M.W.; Bingner, R.L.; Harmel, R.D.; Veith, T.L. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Trans. ASABE.* **2007**, *50*, 885–900. [[CrossRef](#)]
64. Simpson, R.H.; Saffir, H. The Hurricane Disaster Potential Scale. *Weatherwise* **1974**, *27*, 169–186.
65. Zeiler, M.D.; Fergus, R. Visualizing and Understanding Convolutional Networks. In *Computer Vision—ECCV 2014*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2014; Chapter 53; pp. 818–833.
66. Zech, J.R.; Badgeley, M.A.; Liu, M.; Costa, A.B.; Titano, J.J.; Oermann, E.K. Confounding Variables Can Degrade Generalization Performance of Radiological Deep Learning Models. *arXiv* **2018**, arXiv:1807.00431.
67. Dvorak, V.F. A Technique for the Analysis and Forecasting of Tropical Cyclone Intensities from Satellite Pictures. Technical Memorandum. 1973. Available online: <https://repository.library.noaa.gov/view/noaa/18546> (accessed on 27 December 2019).
68. Cha, D.H.; Wang, Y. A Dynamical Initialization Scheme for Real-Time Forecasts of Tropical Cyclones Using the WRF Model. *Mon. Weather Rev.* **2013**, *141*, 964–986. [[CrossRef](#)]
69. Wang, Y. Structure and formation of an annular hurricane simulated in a fully compressible, nonhydrostatic model—TCM4. *J. Atmos. Sci.* **2008**, *65*, 1505–1527. [[CrossRef](#)]
70. Moon, Y.; Nolan, D.S. Spiral rainbands in a numerical simulation of Hurricane Bill (2009). Part I: Structures and comparisons to observations. *J. Atmos. Sci.* **2009**, *72*, 164–190. [[CrossRef](#)]
71. Gal, Y.; Ghahramani, Z. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In Proceedings of the ICML 2016, New York, NY, USA, 19–24 June 2016; pp. 1050–1059.
72. Sünderhauf, N.; Brock, O.; Scheirer, W.; Hadsell, R.; Fox, D.; Leitner, J.; Upcroft, B.; Abbeel, P.; Burgard, W.; Milford, M.; et al. The limits and potentials of deep learning for robotics. *Int. J. Robot. Res.* **2018**, *37*, 405–420. [[CrossRef](#)]
73. Yosinski, J.; Clune, J.; Nguyen, A.; Fuchs, T.; Lipson, H. Understanding Neural Networks through Deep Visualization. *arXiv* **2015**, arXiv:1506.06579.
74. Li, J.; Chen, X.; Hovy, E.; Jurafsky, D. Visualizing and Understanding Neural Models in NLP. *arXiv* **2016**, arXiv:1506.01066.
75. Samek, W.; Wiegand, T.; Müller, K.R. Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models. *arXiv* **2017**, arXiv:1708.08296.
76. Zhang, Q.S.; Zhu, S.C. Visual interpretability for deep learning: A survey. *Front. Inf. Technol. Electron. Eng.* **2018**, *19*, 27–39. [[CrossRef](#)]

