Research Centre for Wind Engineering and
Engineering Vibration, Guangzhou University
Guangzhou, China

11 February 2022

In reference to amt-2021-405 "Identification of tropical cyclones via deep convolutional neural network based on satellite cloud images":

The authors appreciate the referee for his/her valuable comments and suggestions. We will address these concerns below by first quoting the comments.

Comments from Referee #2

General comments:

1. In this study, deep convolutional neural network (DCNN) is adopted to identify TC satellite images. Efforts are also made to explore how the DCNN models work internally. Overall, the work is interesting, and the manuscript is well written, with analyses/discussions presented comprehensively and reasonably. I think this is a good piece of work which can contribute to existing literature. I have only some minor comments. It is suggested the article be accepted after minor revision.

Response: Thanks for the reviewer's constructive suggestions as well as the encouraging comments. According to your comments, we have revised the manuscript. Detailed responses are stated as below.

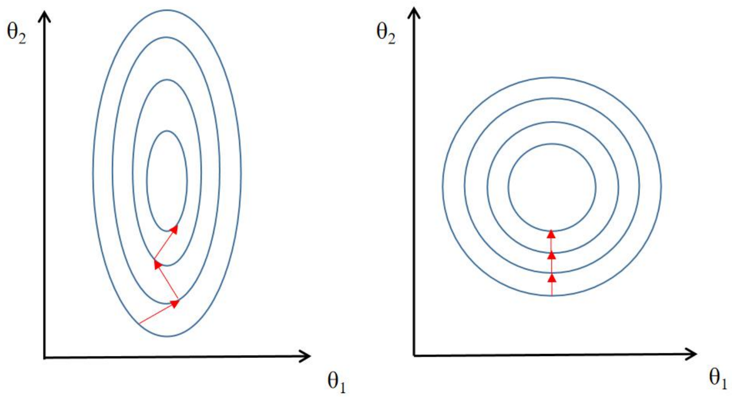
Specific Comments:

2. Line 24: ever--every

Response: Revised accordingly.

3. The authors claim that the normalization of image pixel values can accelerate the model convergence. Why?

Response: To answer the above question more explicitly, we may take the following regression issue as an example. For equation $f_\theta(x) = \theta_1 x_1 + \theta_2 x_2 + b$, given two sets of values for variables x_1 and x_2 which are respectively in the range of [0,100] and [0,1], our aim is to determine the optimal values of coefficients θ_1 and θ_2. Let's assume that the influence of x_1-related item and x_2-related item in the equation on estimation of θ_1 and θ_2 is equal, which is usually the case in the field of image identification, then θ_2 should be larger than θ_1. By convention, the gradient descent method is adopted to estimate the optimal values of the two coefficients. To minimize the difference between training results and true values, one has to compute the derivative of θ_1 and θ_2. Since x_1 is greater than x_2, it can be easily deduced from the derivation formula that the descent speed of θ_1 is much larger than that of θ_2.



(a) without normalization (b) with normalization

Figure 1. Schematic diagram of optimization process via gradient descent method

Figure 1 shows a schematic diagram of the optimization process via the gradient decent method both without (a) and with (b) normalization technique. For Fig. 1(a), as the value ranges of θ_1 and θ_2 vary significantly, the gradient vector (marked as red arrow line) computed based on variable records at one step may not be in parallel with the one computed based on those at neighboring steps (i.e., demonstrating a zigzag pattern), which makes the optimization process to be comparatively longer and more time-consuming. By contrast, for Fig. 1(b), because all variable records are normalized to be in the range of [0 1], the values of θ_1 and θ_2 become in a similar range, so does their derivatives. As a result, the gradient vector computed based on variable records at one step tends to be in parallel with the one computed based on those at neighboring steps, which makes the optimization process to be shorter and more time-saving. Overall, the normalization technique is beneficial for speeding-up the model convergence.

4. Please explain a little more about the Dropout layer.

Response: The authors have added one more statement about how a Dropout layer works in the updated manuscript (lines 161-162): "During training, the dropout layer can randomly drop neural units from the neural network."

In a neural network, one layer is called the dropout layer because some of the neurons are removed from the neural network. To explain the above point more clearly, Figure 2 depicts a schematic diagram of the internal structure of a dropout layer. As can been seen, some neurons are disconnected with others, as a result, they seem to be dropped out from the network system. The Dropout layer is usually involved in the following steps during the training process:

a) Set the dropout rate of each Dropout layer;
b) Remove part of the neurons according to the corresponding rate before the training, and update online neurons / weight parameters during the training process;
c) After all parameters are updated, some neurons are removed again according to the corresponding rate, and then the training begins;
d) Repeat the above process until an acceptable fitting result is achieved.

In general, the larger and deeper the neural network is, the more likely it tends to suffer from over-fitting problems. In this regard, owing to the operations of dropout layer, some neurons can be randomly removed from the network, which is pretty useful for preventing over-fitting problem and for improving the universality of fitting results.

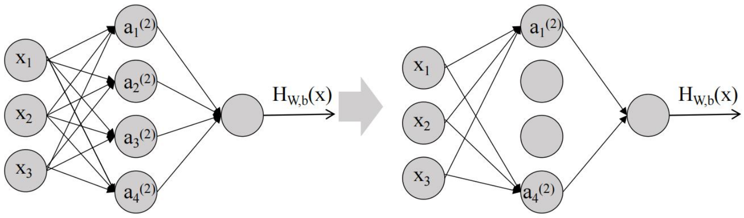


Figure 2. Schematic diagram of the structure of a Dropout layer

5. Results of the evolutional curve in Figure 5(b) suggest that the training accuracy is not improved consistently with the increase of training epochs. When should you finish the training process?

Response: Thanks for the meaningful comments. Based on our practical experience, it is not always the case that training accuracy is improved in trend with the increase of training epochs. Sometimes, training accuracy may decrease slightly and slowly with the increase of training epochs, as the one shown in Figure 5(b) in the manuscript. Generally, when the values of loss function become small enough (i.e., lower than a certain value) and change insignificantly with the increase of training epochs, the training process can be stopped. Our experience suggests that the best training results can be achieved within 80-100 epochs. Thus, in the presented manuscript, the model has been trained for 100 epochs. There are three points to be stressed.

(1) In practice, the training process would be stopped automatically according to the results of loss function. Although there are 100 epochs during the training process for the example show in Figure 5(b), the training process would have finished within the first 40-50 epochs in practice. The training process lasted for 100 epochs only because we forced it to do so.

(2) In this study, a method is adopted to retain the training information with the best validation accuracy, i.e., even that the training process lasted for 100 epochs in the presented example, only the parameterization information associated with the best training epoch has been retained, or more explicitly, the model after training for 100 epochs would be the same with the one trained for 50 epochs in this example.

(3) There are other embedded methods to determine when the training process would be stopped. For example, the training can be stopped in advance if the validation accuracy or loss is not improved within 10 epochs.

6. The authors present many heat maps of TC images. How about those of non-TC image? Are there any typical differences between these two types of heat maps?

Response: Thank you for your useful comments. Non-TC samples also have heat maps, which may differ from those of TC images significantly. An obvious difference between these two types of heat maps lies that there is a lack of massive concerning (or weighted) area(s) or the concerning areas are dispersedly distributed in the heat maps for non-TC samples, as demonstrated in Figure 3. In addition, for some heat maps corresponding to NWPO images with non TCs, there are relatively more concerns with onshore clouds, as shown in Figure 3 (c, d). However, sometimes, the two types of heat maps may demonstrate similar characteristics, as those discussed for Figure 11 and Figure 16 in the manuscript. As this article focuses on the identification of TC images, we have not discussed too much about the typical characteristics of heat maps associated with non-TC images.

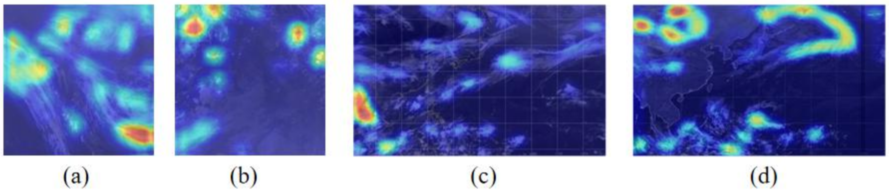


Figure 3. Heat maps of non-TC images: (a) L image I; (b) L image II; (c) NWPO image I; (d) NWPO image II

The author sincerely thanks the reviewer for his kind advice and meaningful comments, which are valuable in improving the quality of our manuscript.

Sincerely yours,

Dr. Y.C. He (Associate Prof.)
Guangzhou University
E-mail: yuncheng@gzhu.edu.cn