

# Quantitative Estimation of Rainfall Rate Intensity Based on Deep Convolutional Neural Network and Radar Reflectivity Factor

Haochen Yang<sup>1</sup>, Tieqiao Wang<sup>1</sup>, Xuesong Zhou<sup>2</sup>, Jiwen Dong<sup>1</sup>, Xizhan Gao<sup>1</sup>, Sijie Niu<sup>1,\*</sup>

<sup>1</sup> School of Information Science and Engineering, University of Jinan, Jinan, Shandong, 250022

<sup>2</sup> Information Service Centre, Shandong Meteorological Bureau, Jinan, Shandong, 250031

Yanghc\_Dawns@163.com; quasre@126.com; cedarzhou2005@163.com; ise\_dongjw@hotmail.com; Gaoxizhan123@126.com; sjniu@hotmail.com

## ABSTRACT

The Convolutional Neural Network Model (CNN) has shown excellent performance in many tasks in recent years, such as the application in image recognition and classification. In this paper we propose a method based on the deep convolutional neural network model VGG and radar reflectivity factor to quantitatively estimation the rainfall rate intensity, which improves the shortcomings of the traditional method Z-R relationship with high error. We selected the shallow convolutional neural network LeNet for comparison. Then, we studied the effect of selecting different sizes of radar reflectivity factor images on the quantitative estimation of rainfall rate intensity. The results show that the depth convolutional neural network VGG and the relatively large size of radar reflectivity factor image are better than the traditional Z-R relationship and shallow convolutional neural network LeNet.

## CCS Concepts

• Computing methodologies→Neural networks • Information systems→Geographic information systems

## Keywords

Convolutional neural network, Z-R relationship, Radar reflectivity factor, Rainfall rate intensity

## 1. INTRODUCTION

Quantitative estimation of rainfall based on radar reflectivity factor has the advantages of wide measurement range and high spatial and temporal resolution. The ground-measured rainfall stations often show difficulties in accurately calculating the rainfall in a certain area because of the sparse distribution [1]. In recent years, quantitative estimation of precipitation based on radar reflectivity factors has been widely used in meteorology, flood prevention, agricultural production and other fields.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).

ICBDT2019, August 28–30, 2019, Jinan, China

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-7192-6/19/08...\$15.00

<https://doi.org/10.1145/3358528.3358582>

At present, the Z-R relationship is generally applied to quantitatively estimate the rainfall rate intensity. The formula of Z-R relationship is  $Z = AR^b$ , where  $Z$  is the radar reflectivity factor, and the magnitude of the radar reflectivity factor depends on the unit of the echo intensity indicating the precipitation target. It is related to the size, quantity and phase state of the precipitation particles per unit volume of the rainfall target.  $R$  is the rainfall rate intensity. Currently, the rain series algorithm of the new generation weather radar uses the formula of Z-R relationship set in WSR-88D, that is,  $Z = 300R^{1.4}$  [2]. Because different climates, seasons and regions have great influence on the error of Z-R relationship, many scholars have proposed different optimization algorithms to optimize Z-R relationship. Liu *et al* [3] proposed a grouped Z-R relationship based on the optimization algorithm and applied it to the radar quantitative estimation of rainfall in the Huaihe River Basin. Based on probability matching method (PMM), Zhang *et al* [4] established a climatic Z-R relationship model for convective, thunderstorm, and shower-mixed rain patterns with ranging. Xu *et al* [5] proposed a method for optimizing the coefficient of Z-R relationship using genetic algorithm. However, since the Z-R relationship is estimated by a fixed coefficient, there is a real-time inefficiency problem in principle. The subtle changes of parameters  $A$  and  $b$  lead to the disparity of results. The non-parametric and adaptive learning method of neural network provide a more effective way to correct the error, and effectively solve the real-time inefficient problem of Z-R relationship.

Convolutional neural network is a kind of feedforward neural network with convolutional computation and deep structure. It is modeled on the visual perception mechanism of the organism, which can extract the discriminative features from the whole of image. It is widely used in image recognition and classification [9-12]. VGG model is a kind of deep convolutional neural network. It was proposed by the Visual Geometry Group of the University of Oxford and won the second place of ILSVRC in 2014 [6].

In this paper, we applied the deep convolutional neural network VGG model to this problem. At the same time, we compare the traditional Z-R relational model and the shallow convolutional neural network LeNet model [7] in this subject. In additional, we verify the influences of radar reflectivity factor range based on different sizes on the experimental results.

---

\* Corresponding author: Sijie Niu([sjniu@hotmail.com](mailto:sjniu@hotmail.com))

## 2. METHOD

### 2.1 VGG Model Structure

The VGG model is characterized by a deep network of small-sized convolutional kernels. It consists of 13 convolutional layers, 5 pooling layers, and 3 fully connected layers. The model is showing in Figure 1.

The structure of the VGG model consists of convolutional layer, pooling layer, and fully connected layer. As shown in Figure 1, where “Conv” represents the convolutional layer, “Pool” represents the pooling layer, and “FC” represents the fully connected layer.

#### 2.1.1 The Convolutional Layer

Each convolutional layer contains multiple convolutional kernels. A convolutional kernel corresponds to a weight coefficient matrix. The convolutional kernels slide through the input image to extract the high-dimensional feature map by doing convolutional operation between the weight coefficient matrix and the original image. The convolutional operation is defined as equation (1).

$$y_{ij}^l = \sum_{a=0}^{d-1} \sum_{b=0}^{d-1} k_{ab}^l * x_{(i+a)(j+a)}^l \quad (0 \leq i, j \leq n-d) \quad (1)$$

In the above formula,  $x^l$  represents the input feature map of the  $l$ th layer,  $y^l$  represents the output feature map corresponding to the  $l$ th layer,  $d$  represents the size of the convolutional kernel,  $n$  represents the size of the input feature map, and  $k$  represents the weight coefficient matrix of the convolutional kernel. The output feature map  $y^l$  is then passed through the activation function of equation (2) as the input feature map  $x^{l+1}$  of the  $(l+1)$ th layer.

$$x^{l+1} = f(y^l + b^l) \quad (2)$$

In the above formula,  $b^l$  is the offset of the  $l$ th convolutional layer, and  $f(\cdot)$  is the activation function.

#### 2.1.2 The Pooling Layer

The pooling layer is also referred to the sampling layer, which is a sampling operation on the input image. The dimensions of the feature map become too high because many convolutional kernels in the convolutional layer extract a large number of features on

the target image. The pooling layer is mainly used for feature dimension reduction, compressing the number of data and parameters, reducing over-fitting, and improving the fault tolerance of the model. Currently, the maximum pooling, average pooling, and random pooling are generally used. By using the pooling matrix with  $d * d$  size, the formula of the maximum pooling is as equation (3).

$$x_{ij}^{l+1} = \max_{0 \leq a, b < d} x_{i+a, j+b}^l \quad (0 \leq i, j \leq n-d) \quad (3)$$

#### 2.1.3 The Fully Connected Layer

The fully connected layer can combine these local features into a complete feature map. Generally, in the classification problem, the fully connected layer acts as a classifier throughout the neural network. In this paper, quantitative estimation of rainfall rate intensity is a regression problem, there the function of the fully connected layer combines all local features.

## 2.2 Gradient Descent With Momentum

In this paper, the neural network optimization algorithm is the momentum gradient descent method [8], which is generally faster than the standard gradient descent algorithm. The momentum gradient descent algorithm updates the weight of the neural network by calculating the exponential weighted average of the gradient.

## 3. DATA SET PROCESSING

Our data are collected from the Shandong Meteorological Bureau, including radar reflectivity factor data and rainfall rate intensity data during August 17, 2018 to August 21, 2018.

The radar reflectivity factor data are scanned and collected by the meteorological radar each ten minutes. The longitude range is between  $113.20^\circ$  and  $124.30^\circ$ , and the latitude range is between  $32.80^\circ$  and  $39.80^\circ$ .

Figure 2 shows an image of the radar reflectivity factor at 12 o'clock on August 17, 2018, GMT.

The rainfall data is measured by 1,536 measurement stations in Shandong Province. The site records the cumulative rainfall rate intensity each hour from 14:00 on August 17, 2018 to 17:00 on August 20, 2018. There are 112,128 rainfall intensity data points.

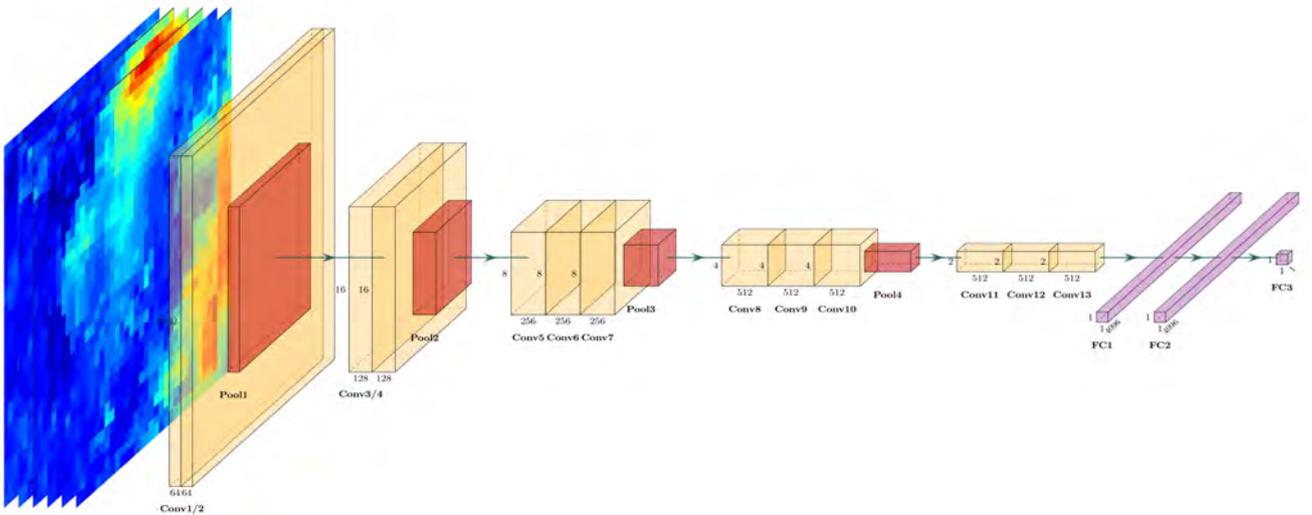
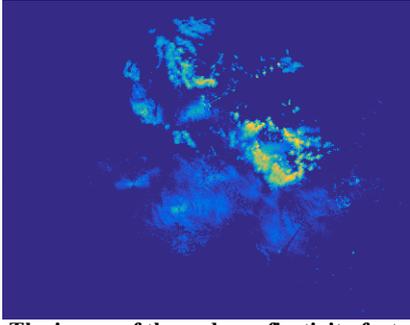
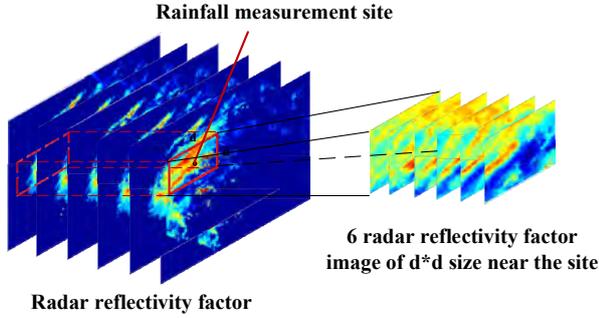


Figure 1. VGG Convolutional Neural Network



**Figure 2. The image of the radar reflectivity factor at 12 o'clock on August 17, 2018, GMT**

This paper constructs a set of data through 112,128 rainfall rate intensity data points. By using the site-wide acquisition method, the radar reflectivity factor images of 112,128 rainfall rate intensity data points corresponding to the  $d * d$  range near the site was extracted. Since the radar reflectance factor image is scanned each ten minutes, the rain rate intensity data are collected each hour. That is, each rainfall rate intensity data point corresponds to 6 radar reflectivity factor images, as shown in Figure 3.



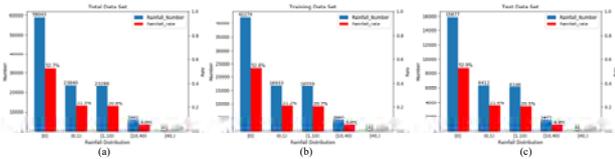
**Figure 3. Radar reflectivity factor image of  $d * d$  range near a rainfall measurement site from 4 am to 5 am on August 18, 2018**

Then we constructed 112,128 pairs of data samples. The samples are independent of each other. After randomly scrambling the data points, 80,000 pairs of data samples are selected as the training set, and 30,000 pairs of data samples are used as the test set, as shown in Table 1.

**Table 1. Data format of training data set and test data set**

	Number of Data	Image Format	Label Format
Training set	80000	[d, d, 6]	[1]
Test set	30000	[d, d, 6]	[1]

In order to make the data distribution of the training set and the test set reasonable, we separately calculated the distribution of the rainfall rate intensity of the training data set and the test data set, as shown in Figure 4.



**Figure 4. Distribution of data sets**

In Figure 4, Figure (a) shows the distribution of rainfall rate intensity for all data, Figure (b) shows the distribution of rainfall rate intensity in the training data set, and Figure (c) shows the distribution of rainfall rate intensity in the test data set.

By comparing the distribution of the training data set, test data set and overall data set. It is found that both the training data set and the test data set obey the distribution of the entire data set, which proves the division of the data set is reasonable.

## 4. RESULTS

### 4.1 Evaluation Parameters

In this paper, the performance of the models is quantitatively analyzed by means of mean square error ( $MSE$ ), root mean square error ( $RMSE$ ), mean relative error ( $MRE$ ), correlation coefficient ( $CC$ ) and decision coefficient ( $R^2$ ). The formulas for  $MSE$ ,  $RMSE$ ,  $MRE$ ,  $CC$ , and  $R^2$  are showing as below:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - p_i)^2 \quad (4)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - p_i)^2} \quad (5)$$

$$MRE = \frac{1}{N} \sum_{i=1}^N |y_i - p_i| / \bar{y} \quad (6)$$

$$CC = \frac{Cov(y, p)}{\sigma_y \sigma_p} \quad (7)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - p_i)^2}{\sum_{i=1}^N (y_i - \frac{1}{N} \sum_{i=1}^N y_i)^2} \quad (8)$$

In the above formulas,  $y$  represents the sequence of rainfall station measurements,  $p$  represents the model prediction value, and  $\bar{y}$  represents the average value of the rainfall station measurements. With the smaller values of  $MSE$ ,  $RMSE$  and  $MRE$  and the higher coefficients of  $CC$  and  $R^2$ , the performance of the model will be better.

### 4.2 Comparison Results

In this experiment, the paper compares the accuracy of the VGG model with the traditional method Z-R relationship model and the shallow convolutional neural network LeNet model. The Z-R relationship model selects the formula  $Z = 300R^{1.4}$  which is the most commonly used.

In order to compare the performance of different models finely, the experimental selection of  $d$  which is the size of radar reflectivity factor image equals 32.  $d$  is defined in Figure 3. The experiment was performed on the training set and the test set using different models. The scatter plot of the estimated value of the model and the measured value of the sites are shown in Figure 5. From the distribution of scatter points in the figure, the scatter distribution of the VGG model is relatively more concentrated and has less dispersion.

Table 2 and Figure 6 gives the evaluation parameters results of the Z-R relationship, LeNet and VGG model. It can be clearly found that the evaluation parameters of the VGG model are superior to the Z-R relationship model and the LeNet model.

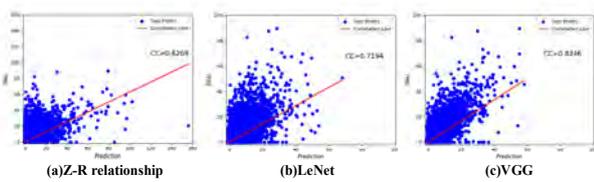


Figure 5. The scatter points of estimated and measured

Table 2. Evaluation parameters of different models

	MSE	RMSE	MRE	CC	R <sup>2</sup>
Z-R	18.66	4.32	0.78	0.62	0.29
LeNet	12.79	3.57	0.74	0.71	0.51
VGG	8.35	2.88	0.58	0.83	0.68

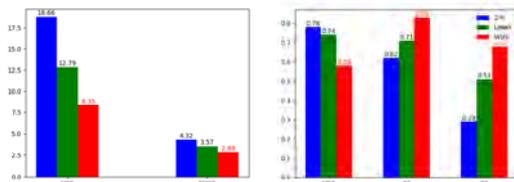


Figure 6. The histogram of model comparison experiment

### 4.3 Analysis of the Sample Size

By conducting this experiment of comparing different models above, it is proved that the performance of the VGG model is significantly better than the Z-R relation model and the LeNet model.

The basic premise of this experiment was to take  $d = 32$ , but different  $d$  will also affect the experimental results. Thus, we have taken four different  $d$  for the VGG model in this experiment. The experimental results are shown in Table 3 and Figure 7. It can be clearly seen that choosing a relatively large radar reflectivity factor image has better performance.

Table 3. Evaluation parameters of different  $d$  in VGG model

	MSE	RMSE	MRE	CC	R <sup>2</sup>
d=16	10.93	3.30	0.72	0.78	0.58
d=32	8.35	2.88	0.58	0.83	0.68
d=48	6.76	2.60	0.52	0.86	0.74
d=64	6.16	2.48	0.49	0.87	0.76

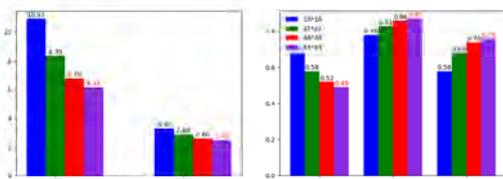


Figure 7. The histogram of size comparison experiment

## 5. CONCLUSION

In this paper, we innovatively applied the deep convolutional neural network VGG model to solve the problem of using the radar reflectivity factor to estimate rainfall rate intensity. It was proved by the traditional Z-R relationship model and shallow convolutional neural network LeNet model. The VGG model performance is superior to the above estimation methods.

Experiments show that relatively large-scale radar reflectivity factor image and deeper neural network model have better performance in this problem.

## 6. ACKNOWLEDGMENT

The work is supported by the National Natural Science Foundation of China under Grant No. 61701192, No.61873324, No.61673405, and No.61573166. the Natural Science Foundation of Shandong Province, China, under Grant No. ZR2017QF004, China Postdoctoral Science Foundation under Grants No. 2017M612178, the Shandong Province Key Research and Development Program under Grant with No.2018GGX101048, and the Project of Shandong Province Higher Educational Science and Technology Program under Grant with No.J16LN07.

## 7. REFERENCES

- [1] Xiaoqing, Y., Guangheng, N., Anjun, P., and Lichuan, W. 2010. Z-R Power Relationship in Beijing Based on Optimization Algorithm. *Hydrology*. 1, 1-6.
- [2] Richard, A. Fulton., Jay P. Breidenbach., Dong-Jun Seo., Dennis A. Miller., and Timothy O'Bannon. 1997. The wrs-88d rainfall algorithm. *Weather & Forecasting*. 13, 2, 377-395.
- [3] Juan, L. and Sheng, X. 1999. Classified Z-I Relationship and Its Application to the Measurement of Rainfall by Weather Radar over the Huaihe River Basin. *Meteorological Science*. 2, 213-220.
- [4] Jiaguo, Z. and Hong, C. 1999. The Climatic Reflectivity-Rain Rate Relations in Sanxia Region. *Meteorology*. 3, 44-48.
- [5] Zhifang, X., Jun, X., and Wenzhong, G. 2006. Application of Genetic Algorithm in Optimizing the Z-R Parameter to Radar Rainfall Estimation. *Plateau Meteorology*. 25, 4, 710-715.
- [6] Simonyan, K. and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556*.
- [7] Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*. 86, 11, 2278-2324.
- [8] Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A., and Jaitly, N. 2012. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. *IEEE Signal Processing Magazine*. 29, 6, 82-97.
- [9] Zhang, M., Li, W., and Du, Q. 2018. Diverse region-based cnn for hyperspectral image classification. *IEEE Transactions on Image Processing*. 1-1.
- [10] Hansen, M. F., Smith, M. L., Smith, L. N., Salter, M. G., Baxter, E. M., and Farish, M. 2018. Towards on-farm pig face recognition using convolutional neural networks. *Computers in Industry*. 98, 145-152.
- [11] Ren, S., He, K., Girshick, R., and Sun, J. 2015. Faster r-cnn: towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence*. 39, 6, 1137-1149.
- [12] Mehdipour Ghazi, M., Yanikoglu, B., and Aptoula, E. 2017. Plant identification using deep neural networks via optimization of transfer learning parameters. *Neurocomputing*. 235, 228-235.