A Novel Computer-Aided Cloud Type Classification Method Based on Convolutional Neural Network with Squeeze-And-Excitation

Guanshen Li
Xavier High School, 181 Randolph Rd, Middletown, CT 06457, America

*Corresponding author: GuLi@xavierhighschool.org

Abstract. Clouds have a huge impact on the energy balance, climate and weather of the earth. Cloud types have different cloud radiation effects, which is an important indicator of cloud radiation effects. Therefore, determining the type of cloud is of great significance in meteorology. In this paper, the Convolutional neural network with Squeeze & Excitation Networks (SENet) are mainly used to solve this problem. CNN can automatically learn the filters that need to be manually set before, and can learn complex edge, spatial and texture information in the image which are difficult for traditional methods to learn and extract. Moreover, a website and a deep learning framework are established to showcase the results of this article and to further develop our models and methods through open source methods.

Keywords: Cloud Type Classification, Convolutional Neural Network, Squeeze & Excitation Networks, Artificial Intelligence.

1. Introduction
Automatic and accurate cloud classification is one of the important applications of meteorological research. Clouds cover most of the world's surface. Therefore, through short wave cooling and long wave heating effects, they play an important role in regulating the earth's radiation budget [1]. At the same time, changes in cloud types will affect short wave and long wave radiation. In addition, the distribution of various cloud types in different time and space may change with climate change [2-3]. Computer based automatic classification systems will help forecasters in several ways. The importance of this is that a large number of satellite images are generated every day, for example, each geostationary environment satellite (goes) generates 100GB of data per day. However, the current classifier effect is not very good, many times the error is very large. This is because the impact of orbit on global climate cannot be ignored. If the ground cloud database considers orbit, it will make it more valuable than the existing ground cloud database, which will help to promote the research on the impact of orbit on global warming. In addition, high-precision automatic cloud classification methods, especially convective cloud identification, are very important for identifying dangerous weather processes. The accuracy of weather forecast and rainfall estimation system is greatly improved by obtaining cloud cover distribution information and tracking the changes of meteorological conditions [4]. For example, automatic and immediately available interpretation of cloud images will be a powerful help to US Navy forecasters, who need to issue specialized analysis, real-time forecasts and sea forecasts [5]. It is a...
tedious and unreliable task to extract cloud field information from these images by visual / manual interpretation, and the result depends on the operator to some extent. Therefore, an efficient and robust cloud classification scheme is needed to automatically process satellite cloud images in meteorological applications.

Traditional cloud classification methods still rely on the experience of experts, which makes it need a lot of manpower. In addition, the task is unreliable, time-consuming, and to some extent depends on the experience of the operator. In order to extract cloud information from these images by visual / manual interpretation, the classification results usually bring uncertainty and bias. On the contrary, the detailed information of local cloud cover and cloud cover type can be easily obtained by using ground observation instruments. Therefore, more and more cloud texture and structure feature extraction algorithms are proposed for cloud recognition. Generally speaking, there are two kinds of cloud features in cloud classification: spectral features and texture features. The first feature extracts the cloud radiation information of different bands, which plays a more important role in cloud classification. The most commonly used methods in this category include threshold-based scheme [6], histogram Scheme [7] and multispectral method [8]. Previous texture-based cloud classification methods mostly use statistical measures based on gray level co-occurrence matrix (GLCM) [9] and its variants, such as gray difference vector (gldv), gray difference matrix (GLDM) and difference histogram (SADH) [10]. Although these efforts contribute to cloud classification, accurate cloud classification has not been achieved satisfactorily. Therefore, a high-precision automatic cloud image classification algorithm is still under development, and it is necessary to develop efficient cloud data feature extraction scheme.

In recent years, convolutional neural networks (CNNs), as a deep learning structure, have made remarkable achievements in the field of computer vision and pattern recognition [11]. CNN can learn increasingly complex patterns and distinguish textures from a large number of pre trained and labeled datasets. However, due to the complexity of cloud texture and pattern, traditional methods are difficult to define and extract cloud features. In addition, CNN is usually a hierarchical feature extraction framework. Generally speaking, the shallow CNN captures the fine texture, while the deep layer reflects the high-level semantic information. Previous studies have shown that the texture and semantic features of cloud are indispensable for cloud feature description.

In this paper, we mainly used the Convolutional neural network with Squeeze & Excitation Module (SENet). CNN can automatically learn the filters that need to be manually set before, and can learn complex edge, spatial and texture information in the image which are difficult for traditional methods to learn and extract. By providing the forecast in a small area(5*5km), it can help people to get the weather information they need, such as the up-coming weather of a golf court, a ski ranch or a race track. In the second section, we will illustrate the foundation of our model and then describe our model in the third section. Finally, the experiment result will be shown in the fourth section.

2. Methods
In this paper, Convolutional neural network and Squeeze-and-Excitation Module will be used to comprise the whole network model. So, it is necessary for us to demonstrate these models and techniques in this section.

2.1. Convolutional neural networks (CNN)
Convolutional neural networks (CNN) are a kind of feedforward neural networks with deep structure including convolution computation. It is one of the representative algorithms of deep learning. Convolutional neural network has the ability of representation learning and can shift invariant classification of input information according to its hierarchical structure.

Convolutional neural network, which imitates the biological visual perception mechanism, can be used for supervised and unsupervised learning. The convolution kernel parameter sharing in the hidden layer and the sparsity of the connection between layers make the convolution neural network grid like with less computation Topology) features, such as pixels and audio, have stable effects and have no additional feature engineering requirements for data.
2.2. Squeeze-and-Excitation Module
The convolution kernel, which is the core of the Convolutional Neural Network, is generally understood as an information extractor that aggregates spatial information and channel-wise information in the local receptive field. Convolutional Neural Networks consist of a series of convolutional layers, nonlinear layers, and downsampling layers so that they can capture the features of the image from the global receptive field to describe the image.

However, it is quite difficult to learn a very powerful network. The difficulties come from many aspects. Recently, a lot of work has been proposed to improve the performance of the network from the spatial dimension level, such as embedding multi-scale information in the Inception structure, aggregating features on different receptive fields to obtain performance gains; considering contextual information in space in the Inside-Outside network; and bringing the attention mechanism to the spatial dimension and so on. All of this work has yielded pretty good results.

There has been a lot of work focused on the spatial dimension to improve the performance of the network. Therefore, some scholars proposed Squeeze-and-Excitation Networks (SENet) considering the relationship between feature channels. The core idea of SENet is to learn the feature weights according to the loss through the network, in order to train the model in a way that the weight of effective feature map is significant and the weight of invalid feature map is small for better results. Apparently, the SE block embedded in some of the original classification networks inevitably adds some parameters and computation, but it is acceptable for effect.

As shown in Fig.3, Squeeze-and-Excitation block contains three operations. Given an input matrix X with the number of a feature channel as $c_1$, a feature with the number of feature channels as $c_2$ is obtained after a series of convolution and other general transformations. Different from traditional CNN, SE block re-calibrate the previously obtained features through three operations.
SE block first utilizes the Squeeze operation to compress the feature map along the spatial dimension, transforming each two-dimensional feature channel into a real number. To some extent, the real number has a global receptive field, it characterizes the global distribution of responses on feature channels and allows layers close to the input to obtain global receptive fields, which is very useful in many tasks. The formulas with respect to convolution operation on input and Squeeze operation are shown below.

\[
u_c = v_c \ast X = \sum_{s=1}^{C'} V_c^s \ast x^s
\]

(1)

\[
Z_c = F_{sq}(u_c) = \frac{1}{w \times h} \sum_{i=1}^{W} \sum_{j=1}^{H} u_c(i, j)
\]

(2)

Then Excitation operation, which is a mechanism similar to the gate in recurrent neural network, learns weights from each feature channel to explicitly model the correlation between feature channels. The formula on Excitation is shown below.

\[
s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z))
\]

(3)

The output of the Excitation is regarded as weight of channel. Finally, reweight operation multiplies weight and previous input channel by channel to complete recalibration of the original feature [17]. The formula on Reweight is shown below.

\[
\tilde{x}_c = F_{scale}(u_c, s_c) = s_c \ast u_c
\]

(4)

3. The Framework and Architecture of the Model

The proposed model is a Convolutional Neural Network with an added squeeze and excitation (SE) architectural element for sequence-to-sequence learning task. The model was motivated by the CNN-SENet for biometric fish Classification of Temperate species [12]. The architecture of the model is depicted in Fig.4. The model takes a batch of fixed-size of RGB cloud images and outputs a label which represents the cloud’s category. The notations of the model architecture are as follow. Batch size(B), Image size in width(W), height(H) and depth, Filter size(S), the amount of Filter(F), Units after flatten(D), the number of classes(C) and Reduction ratio(r) in Fig.5.

![Figure 3. CNN-SENet Architecture](image-url)
The core element of the proposed model is squeeze-and-excitation block, depicted in Fig. 3. Squeeze operation is achieved by global average pooling which turns two-dimensional feature map to a real number. The following flatten layer reshape previous output to 1*1*F. In order to reduce computation, SE block utilizes two full connected (FC) layers to first scale the number of channels and then recover. It has been proved by experiments that the performance is best when the reduction ratio is set to 16. The ReLu layer makes the block more nonlinear and can better fit the complex correlation between channels. And finally, a sigmoid activation function transforms the previous outputs to weights number between 0 and 1. The input layer of the network takes 200*200-size images with 3 channels. Convolution 2D layer uses different filters to extract features, which reduce the input’s width and height, and increases the depth. Before entering the SE-block, the output has to be batch normalized. The ReLu layer keeps the dimensions of the output from SE-block unchanged. Finally, the output is sent to Max Pooling 2D with strides 2*2, rendering an output of 98*98*32. This iteration which is seen as crucial portion of the model is accumulated for 5 times. The first iteration has a convolution layer of 32 filters which sizes are 5*5. The second and third iteration both have a convolution layer of 64 filters in 3*3. The fourth iteration’s convolution layer has 128 filters in 2*2. The last iteration’s convolution layer has 256 filters in 2*2. The output matrix in 5*5*256 is then sent to flatten layer, changing the dimension to 6400.

In addition to five kernels, there are 3 FC layers in the architecture. With 256 neurons, the first FC layer first extract features, and then batch normalize the output, finally activate output with ReLu function. The second FC layer with 128 neurons is close to first FC layer except adds Dropout at the end. The Dropout which percentage is set to 50%, inactivates half of the neurons in the layer to prevent overfitting. The last FC layer, which number of neurons are the same as the number of classes, applies softmax function to predict the probability distribution per class per input image. The paper compiles the model with Adam as the optimizer and categorical cross entropy as loss.

Figure 4. Squeeze-and-Excitation block

4. Experiment Results
Currently we use the dataset from [13], the Cirrus Cumulus Stratus Nimbus (CCSN) dataset which has 2,543 cloud images in jpeg format with a resolution of 256*256. It is shown in fig.6.

Figure 5. Cirrus Cumulus Stratus Nimbus (CCSN) dataset
At the same time, these pictures are divided into ten categories according to the classification recommendation proposed by the World Metrological Organization. In addition, contrails that may have impacts on the weather has also been included. Eleven cloud types added in CCSN dataset in total can effectively classify cloud types to achieve the purpose of forecasting the weather.

In the experiment, we mainly use two kinds of networks, one is Alexnet, the other is Resnet-50. Alexanet was designed by Hinton, the 2012 Imagenet champion, and his student, Alex krizhevsky. Also after that year, more and deeper neural networks were proposed, such as the excellent VGG and googlenet. Alexnet contains several relatively new technical points, and has successfully applied such tricks as relu, dropout and LRN in CNN for the first time. At the same time, alexnet also uses GPU to accelerate the operation. Alexnet develops lenet's ideas and applies CNN's basic principles to a very deep and wide network.

Resnet is a convolutional neural network proposed by four scholars from Microsoft Research. It won the image classification and object recognition in the Imagenet large scale visual recognition challenge (ilsvrc) in 2015. The characteristic of Resnet is that it is easy to optimize and can improve the accuracy by increasing the depth. The internal residual block uses jump connection to alleviate the gradient disappearance problem caused by increasing depth in the depth neural network.

The results are shown in Fig.7 and Fig.8. They show that overfitting occurs in training due to the complexity of network. The curve for AlexNet shows that it could learn texture and spatial features in
cloud images, but not good enough. The proposed model learns the features rapidly and eventually achieve great results.

The max validation accuracy during training is kept and the time per epoch is recorded. The table as follows shows that the proposed model is robust and suitable for cloud image classification. The max validation accuracy also exceeds the result of CloudNet in [24] which achieves 91% accuracy.

| Network      | Validation Accuracy | Time One Epoch |
|--------------|---------------------|----------------|
| ResNet-50    | 14.58%              | 25.30          |
| AlexNet      | 74.97%              | 39.06          |
| Proposed Model | 94.50%            | 21.37          |

5. Conclusions
In this paper, we mainly use the convolutional neural network (senet) with compressed excitation network to solve the cloud shape classification problem. And finally achieved a better result. In the future, we will improve the way of network, such as introducing object detection, such as yolo-v5 or mask RCNN, to detect clouds in the whole image, so as to achieve better classification effect.

References
[1] J. Duda, P., Minnis, P., Khlopenkov, K., Chee, T. L., & Boeke, R. (2013). Estimation of 2006 Northern Hemisphere contrail coverage using MODIS data. Geophysical Research Letters, 40, 612–617.
[2] Chen, T., Rossow, W. B., & Zhang, Y. (2000). Radiative effects of cloud-type variations. Journal of Climate, 13(1), 264–286.
[3] Stephens, G. (2005). Cloud feedbacks in the climate system: A critical review. Journal of Climate, 18(2), 273–273.
[4] Torsum, I. S., & Kwiatkowska, E. (1999). Neural network system for cloud classification from satellite images. Proc. SPIE, 6(6), 3785–3790.
[5] Bankert, L., & Richard (1994). Cloud classification of AVHRR imagery in maritime regions using a probabilistic neural network. Journal of Applied Meteorology, 33, 909–918.
[6] Liu, S., & Zhang, Z. (2016). Learning group patterns for ground-based cloud classification in wireless sensor networks. Eurasip Journal on Wireless Communications and Networking, 1, 1–6.
[7] Slater, D. W., Long, C. N., & Tooman, T. P. (2001). Total sky imager/whole sky imager cloud fraction comparison.
[8] W. E. Shenk and R. J. Holub, “A multispectral cloud type identification method developed for nimbuss 3 mrir measurements,” Mon. Weather Rev., vol. 104, pp. 284–291, Mar. 1976.
[9] Desbois and G. Szejwach, “Automatic classification of clouds on meteosat imagery: Application to high-level clouds,” J. Clim. Appl. Meteorol., vol. 21, pp. 401–412, 1982.
[10] D. W. Reynolds and T. H. Vonder-Haar, “Bispectral method for cloud parameter determination,” Mon. Weather Rev., vol. 105, pp. 446–457, Mar. 1977.
[11] J. Parikh and A. Rosenfeld, “Automated segmentation and classification of infrared meteorological satellite data,” IEEE Trans. Syst. Man. Cybern, vol. 8, pp. 736–743, 1978.
[12] Zhuo, W., Cao, Z., & Xiao, Y. (2014). Cloud classification of ground-based images using texture-structure features. Journal of Atmospheric and Oceanic Technology, 31(1), 79–92.
[13] zhang, J., liu, P., Zhang, F., & Song, Q. (2018). CloudNet: Ground-based Cloud Classification with Deep Convolutional Neural Network. Geophysical Research Letters.