A NEPHOGRAM PREDICTION METHOD BASED ON GENERATIVE ADVERSARIAL NETWORK

Xun Yu [1,*], Yu Wenqing [1]

[1] Shandong University of Science and Technology, Qingdao, China - mottoxy@163.com

KEY WORDS: Nephogram Prediction, Unsupervised Learning, Generative Adversarial Network, Multi-scale Network

ABSTRACT:

As one of the important sources of meteorological information, satellite nephogram is playing an increasingly important role in the detection and forecast of disastrous weather. The predictions about the movement and transformation of cloud with certain timeliness can enhance the practicability of satellite nephogram. Based on the generative adversarial network in unsupervised learning, we propose a prediction model of time series nephogram, which construct the internal representation of cloud evolution accurately and realize nephogram prediction for the next several hours. We improve the traditional generative adversarial network by constructing the generator and discriminator used the multi-scale convolution network. After the scale transform process, different scales operate convolutions in parallel and then merge the features. This structure can solve the problem of long-term dependence in the traditional network, and both global and detailed features are considered. Then according to the network structure and practical application, we define a new loss function combined with adversarial loss function to accelerate the convergence of model and sharpen predictions which keeps the effectivity of predictions further. Our method has no need to carry out the stack mathematics calculation and the manual operations, has greatly enhanced the feasibility and the efficiency. The results show that this model can reasonably describe the basic characteristics and evolution trend of cloud cluster, the prediction nephogram has very high similarity to the ground-truth nephogram.

1. INTRODUCTION

Satellite observations of the earth have become an indispensable source of information in today's world. Satellite nephogram is obtained from weather satellites that describes the cloud coverage and earth's surface features. Satellite nephogram can be used to identify different weather systems, determine their locations, estimate their intensity and development trend, and provide basis for weather analysis and forecast. In regions lacking meteorological observation stations, such as oceans, deserts and plateaus, the information provided by satellite nephogram makes up for the deficiency of conventional observation data and play an important role in improving the accuracy of prediction. With the rapid development of the aerospace industry, satellite nephogram is becoming more and more abundant. In addition to monitoring and analyzing the real-time cloud information, the movement and evolution of cloud clusters on the nephogram has also attracted much attention. The remote sensing satellite cloud images is used to predict the movement and change of convective cloud clusters that is helpful to predict the small and medium-sized severe weather such as squall line, rainstorm and thunderstorm gale accurately and timely. However, due to the extremely unstable convective cloud cluster, there is still no effective dynamic tracking and prediction method. How to predict the movement trend of convective cloud cluster is also one of the urgent problems to be solved in meteorological forecasting.

Big data technology can quickly obtain valuable information from various types of massive data. The ubiquitous information perception and acquisition terminal has collected a large amount of data for us. The continuous progress of cloud computing technology has provided us with powerful computing ability. Big data is showing its advantages more and more, and its application fields are becoming more and more extensive. As a very effective means to deal with big data, deep learning has

been widely studied by scholars at home and abroad. The real big data itself is bound to be weakly labeled, or even unlabeled. In the current situation that there is huge share gap between unlabeled data and labeled data, unsupervised learning has greater potential than supervised learning relying on data labeling (Le, 2013).

Forecasting the evolution of cloud is challenging due to the following aspects：
(1) Most observed cloud changes are nonstationary, nonlinear and irregular.
(2) Different cloud systems in the same area vary in degree, which is difficult to describe accurately by a single model.

At present, the work done in the field of cloud map prediction at home and abroad includes: The initial research on nephogram prediction is mainly based on the premise that the system remains stable (Taravat A, 2015). Cross correlation method, time mutation method and others are used for linear extrapolation based on the local feature of cloud image and the vector relationship between previous and feature cloud movements (Stowe L L, 2010). These methods have some shortcomings in the prediction of cloud development trend and forecast time, which greatly limit the availability of cloud forecast. These methods have poor performance in predicting cloud trends and the validity of prediction is not satisfied, which greatly limit the availability of cloud prediction. MOS method is adopted for cloud image prediction. The influence factors of atmospheric physical quantities and cloud physical processes are considered in the numerical prediction model, which represents the development trend and advanced research direction of this field (Mackie S, 2010). However, this method involves a huge amount of computation and complex technical requirements, and lacks of follow-up research and application. The CA Cellular Automata model was established by using

---

* Corresponding author

parameters representing the dynamic and thermal characteristics of clouds to predict the development and changes of satellite cloud images in the next 12 h, and to explore the applicability of the new method for short-term satellite cloud image prediction. This method can effectively describe the distribution and movement of uniformly distributed cloud clusters, but the prediction of scattered cloud clusters and the reduction of detailed features are not ideal.

In this paper, a cloud image prediction model is constructed by the research of the generative counter network (Goodfellow, 2014) in unsupervised learning. A large number of continuous satellite cloud images are taken as samples for training, so as to accurately construct the intrinsic characteristics of cloud image evolution and predict the cloud image at several future moments (Ranzato, 2014). The basic framework of the model is generative adversarial network, which consists of generator and discriminator. A generative model G that captures the data distribution, and a discriminative model D that estimates the probability that a sample came from the training data rather than G. The training procedure for G is to maximize the probability of D making a mistake. This framework corresponds to a minimax two-player game. (5) To preserving the long-range dependencies of the network and addressing the problem of lack of sharpness in the predictions, the original convolutional network is improved by multi-scale structure. The original samples are sampled in different degrees S1...Sn by multi-scale transformation，Batch samples with the lowest resolution S1 participate in the training to get a set of prediction samples first，After ascending sampling, the S2 batch sample has the same resolution as the S2 batch sample, which is used together as the input of the second training. The discriminative model D takes a sequence of frames, and is trained to predict the probability that the last frames of sequence are generated by G (Mathieu M, 2015). This allows the discriminative model to make use of temporal information, so that G learns to produce sequences that are temporally coherent with its input. In order to further optimize the prediction results, the model adopts a combination of L_p loss and adversarial loss.

The main contributions of this study include：
(1) Put forward the new way to predict the cloud nephogram in future time. Using the existing nephogram as the samples for feature learning, in a short time, with a small number of manual operations, it is reliable to predict the cloud movement and shape change in several future time.
(2) It can be used to assist existing meteorological forecasting methods, and the model is also applicable to the prediction of the spread of some natural disasters and the moving track of objects.

2. METHOD

2.1 Overview

A classic Generative Adversarial Network (GAN) consists of two neural networks trained in opposition to each other (Goodfellow, 2014). The two adversarially trained models are: a generative model G that captures the data distribution, and a discriminative model D that estimates the probability that a sample came from the training data rather than G. The generator G takes a latent variable z as input, and outputs sample $G(z)$. The discriminator D takes a sample x as input and outputs $D(x)$

which represents the probability that it is real (Figure.1). The training procedure is similar to a two-player min-max game with the following objective function:

$$\min_{G} \max_{D} V(D,G) =$$
$$E_{x \sim p_{data}(x)} \left[log D(x) \right] + E_{z \sim p_z(z)} \left[log \left(1 - D\left(G(z) \right) \right) \right] \quad (1)$$

Where z is a noise sample from a prior probability distribution p_z, and x denotes a real image following a certain distribution p_{data}.
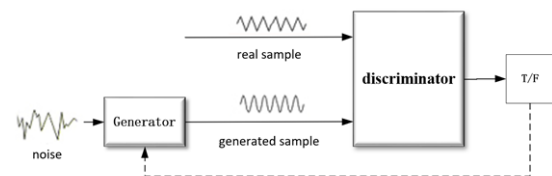

Figure 1. Generative adversarial networks

The classic model use the convolutional network，alternating convolutions and Rectified Linear Units (ReLU). However, convolutions only account for short-range dependencies, limited by the size of kernels. So we combine multiple scales linearly as in the reconstruction process of a Laplacian pyramid (Denton, 2015). The inputs of the discriminator are a sequence of frames and only the last frames are either real of generated by G, the rest of the sequence is always from the dataset. Since the discriminator can take use of the temporal information, so that G learns to produce sequences that are temporally coherent with its input. The discriminator is a multi-scale convolutional network with a single scalar output.

2.2 Generator and discriminator

To solve the short-term dependency problem, the generator is improved to multi-scale (K. Gregor, 2015). Define a series of generated network G_k recursively，S_k is the size of input of G_k, in our experiments, we set $S_1 = 8X8$, $S_2 = 16X16$, $S_3 = 32X32$, $S_4 = 64X64$. Let u_k be the upscaling operator toward size S_k. Let X_k^i, Y_k^i denote the downscaled versions of X^i and Y^i of S_k, and G_k' be a network that learns to predict from X_k and a coarse guess of Y_k. The prediction is calculated by

$$\hat{Y}_k = G_k(X) = u_k\left(\hat{Y}_{k-1}\right) + G_k'\left(X_k, u_k\left(\hat{Y}_{k-1}\right)\right) \quad (2)$$

The architecture of generator is illustrated in Figure 2. Samples are drawn in a coarse-to-fine fashion, commencing with a low-frequency residual image. The second stage samples the band-pass structure at the next level, conditioned on the sampled residual. Subsequent levels continue this process, always conditioning on the output from the previous scale, until the final level is reached. With this structure the generator is able to produce high-quality sample images that are both qualitatively and quantitatively better than other deep generative modeling approaches.

The discriminator is a multi-scale convolutional network with a single scalar output which representing the probability that the

sample comes from the true dataset (Mathieu M, 2015). The input of discriminative model D is a sequence of frames which makes the discriminative model use the temporal information, so that G learns to produce sequences that are temporally coherent with its input.

2.3 Details of adversarial training

We train G and D alternately until optimality, and finally G learns the desired cloud transformation and D becomes a reliable estimator (Karpathy, 2014).

$$L_b\left(Y,\hat{Y}\right) = -\sum_i \hat{Y}_i \log\left(Y_i\right) + \left(1-\hat{Y}_i\right)\log\left(1-Y_i\right) \quad (3)$$

The real input is (X, Y), the corresponding output of the discriminant model is 1; Generated input as $(X, G(X))$, the corresponding output of the discriminant model is 0.

Therefore, the loss function we used to train D is:

$$L_{adv}^D\left(X,Y\right) = \sum_{k=1}^{Nscale} L_b\left(D_k\left(X_k,Y_k\right),1\right) + L_b\left(D_k\left(X_k,G_k\left(X\right)\right),0\right) \quad (4)$$

In order to make the output of the generated model as similar as possible to the real input of the discriminating model, we define the adversarial loss function as:

$$L_{adv}^G\left(X,Y\right) = \sum_{k=1}^{Nscale} L_b\left(D_k\left(X_k,G_k\left(X\right)\right),1\right) \quad (5)$$

In order to reduce the difference between predicted results and ground truth, combined use Lp loss, the generator is therefore trained to minimize $\lambda_{adv}L_{adv} + \lambda_p L_p$.
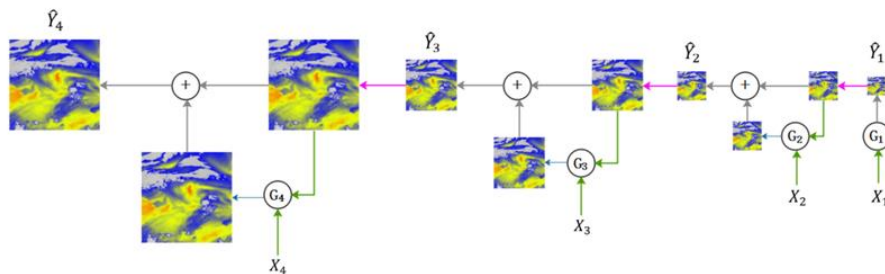

Figure 2. Multi-scale architecture. The generator starts with a noise sample X_1 (right side) and use a generative model G_1 to generate $\hat{Y}_1$. Then the images are upsampled (pink arrow) and used as the conditioning variable for the generative model at the next level G_2 Together with another noise sample X_2, G_2 generates a difference image which is added to up sample $\hat{Y}_1$ to create $\hat{Y}_2$. This process repeats across two subsequent levels to yield a final full resolution sample $\hat{Y}_1$.

3. EXPERIMENT

3.1 Datasets

The dataset used in this paper is based on the water vapor satellite nephogram downloaded from the National Satellite Meteorological Centre Website. The time series images dataset has a duration of one month and an hour interval, and contains a total of 720 images. Each image is divided into 80x80 pixels samples. Samples at the same geographical location form a group of sequential samples. Training data and test data is distributed in a ratio of 5:1.

3.2 Implementation detail

The generative model G and discriminative model D architecture are presented in Table 1.The depth of each scale network and the size of the convolution kernel are determined after continuous adjustment, which can make the whole network converge as soon as possible, extract the feature information accurately, and ensure the stability of the system.

On the basis of the traditional GAN loss function, we add a regularization to restrict the parameters of the model. We use the L_2 regularization in this paper, which refers to the square

root of the sum of squares of the various parameters of the model. Its main function is to prevent overfitting.

G	G1	G2	G3	G4
Channels	32 64 32	32 64 32	64 128 256 128 64	128 256 512 256 128
k-size	3 3 3 3	3 3 3 3	5 3 3 3 3 5	5 5 5 5 5 5

D	D1	D2	D3	D4
Channels	32	32 64 64	64 128 128	64 128 256 64
k-size	3			5 5 5 5
FC	256 128	512 256	512 256	512 256

Table 1. Network architecture

3.3 Quantitative evaluation

We comprehensively analysed the basic network and improved network performance from the loss value, PSNR and SSIM value changes, and the similarity on visual effect between the generated predicted images and the ground truth images.

3.3.1 Loss value curves and box plots.
The loss value curves reflect the change of the loss value and the trend of the gradual convergence of the model. The box plots illustrate the extremum and median of loss value.

Figure 3 illustrate that GAN network can converge stably during training, and the final loss value of the generator and discriminator oscillates above and below a stable horizontal line. After the traditional loss function is improved and the L_p regularization is added, the convergence value of the network will be lower, which indicates that the predicted value and the real value are more similar than before the improvement. In the fitting process, it is generally preferred to make the weight as small as possible and finally construct a model with all the parameters relatively small. The model with small parameter value has simple structure and can adapt to different data sets. For a linear regression equation. If the parameters are small enough, the data migration will not affect the results, and the whole model has a strong anti-disturbance ability.
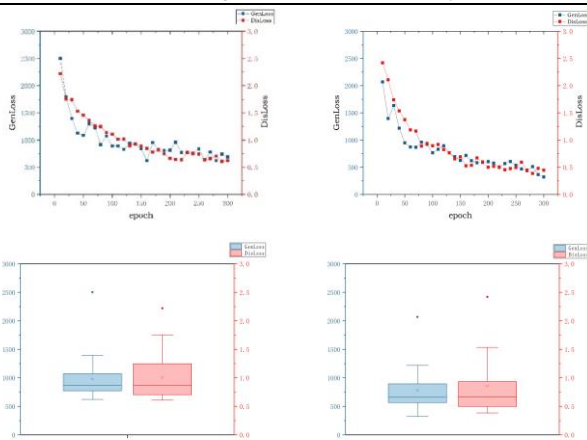

Figure 3. The curves and box plots in each experiment

3.3.2 At the view of classification

This task can be approximated as a classification task due to the small variation amplitude between the graphs at each adjacent moment. We tested the performance of the model using two metrics that are popular for classification accuracy evaluation, overall accuracy (OA) and kappa coefficient (kappa). Both of them are calculated based on confusion matrix. We set up two groups of experiments according to whether L_p regularization is used or not, and calculated OA and kappa between the predicted results and the ground truth of three groups of samples respectively. The statistical result is displayed in Figure.4.
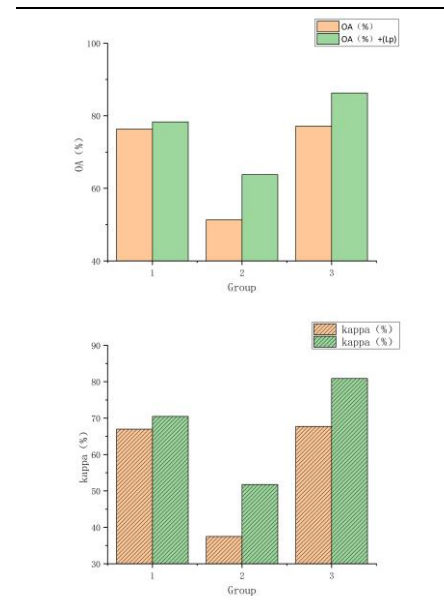

Figure 4. The OA and kappa with and without L_p.

Among the three groups of results, kappa coefficient of two groups is greater than 70% or even as high as 90%, which proves that the predicted results have a high similarity with the real data, indicating that the model can generate cloud maps of the subsequent moments more accurately. The performance of the improved model with L_p regularization is obviously better than that of the unimproved model.

3.3.3 PSNR/SSIM

To evaluate the quality of the image predictions resulting from the different tested system, we compute the Peak Signal to Noise Ratio (PSNR) between the true frame Y_k and the prediction $\hat{Y}_k$:

$$\mathrm{PSNR}\left(Y, \hat{Y}\right) = 10 \log_{10} \frac{max_{\hat{Y}}^2}{\frac{1}{N}\sum_{i=0}^{N}\left(Y_i - \hat{Y}_i\right)} \tag{6}$$

Where $max_{\hat{Y}}$ is the maximum possible value of the image intensities. The Structural Similarity Index Measure (SSIM) is also calculated which ranges between -1 and 1, a larger score meaning a greater similarity between the two images (Wang, 2004).

	Groundtruth1/Output1	Groundtruth2/Output2	Groundtruth3/Output3
without L_p	PSNR : 15.313243 SSIM : 0.44188363	PSNR : 17.666683 SSIM : 0.60298071	PSNR : 15.420183 SSIM : 0.40908234
with L_p	PSNR :16.096272 SSIM :0.46346072	PSNR :19.77556 SSIM :0.63407075	PSNR :16.680063 SSIM :0.4558526

Table 2. The ground truth images、 output images and the value of PSNR and SSIM with and without L_p

We selected 3 groups of samples to test the trained network, predicted cloud images at 5 future moments, and calculated the SSIM value and PSNR value between the ground truth images and the predicted images, and use them to draw line graphs to study the predicted performance trends. The results are listed in table 3 and figure 5. We observed the predicted pictures and found that the predicted results obtained by different input samples were different in quality, but all of them could accurately describe the attribute information such as colour, position and shape, and with a high consistency in the overall structure. Through comparison, it was found that when the loss function was added with a regularization, the details in the picture were retained more completely, and the sharpness between the true frame and the prediction was higher.

We input 4 samples to generate 5 predicted images of future moments. With the increase of prediction sequence number, the prediction effect gradually becomes worse, PSNR value and SSIM value both gradually decrease, which are closest to the first prediction result input and have the highest similarity with the real image. In general, when the loss function is added to the regularization, the network has a better performance, which is displayed in Figure3. It is worth noting that the counter loss and $L_{\mathcal{D}_p}$ loss are not simply added together, but are reconciled by adding coefficient terms respectively. After continuous experimental analysis, an optimal combination loss function is finally obtained.
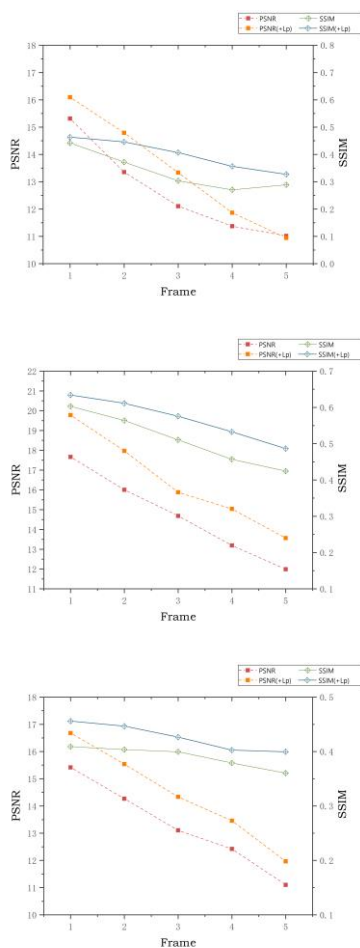

Figure 5. Subsequent PSNR and SSIM line charts at multiple moments.

4. CONCLUSION

The theory and technology of deep learning have greatly promoted the development of many fields of computer vision. GAN, however, is not restricted by labeled data and does not need too much manual processing, which has become a promising research direction in unsupervised learning. We only need to collect enough data and input the network structure which can adjusted casually, the network can gradually learn the internal representation of input data, and generate the similar images by training. If the network combines time series information, it can also study the link characteristics between each frame of the input data, and generate an image similar to the subsequent development. Cloud change has the characteristics of nonstationary and nonlinear properties, irregular, if on the basis of mathematical methods to study changing regularity, not only need a lot of calculation, but also is difficult to establish universal model, thus the change of the cloud of prediction has been the lack of corresponding research. But the rapid development of aerospace enterprise, bring us a lot of satellite images, for the application of deep learning on the cloud prediction and research. At the same time, the deep learning framework is flexible, and even when training the same network, the selection of the optimal parameters is also effected by the sample. Therefore, we can constantly improve the network structure, adjust parameters according to the results, and finally get an optimized adaptive network.

REFERENCES

Denton, Emily, Chintala, Soumith, Szlam, Arthur, and Fergus, Rob. Deep generative image models using a Laplacian pyramid of adversarial networks. In NIPS, 2015.

Goodfellow, Ian J., Pouget-Abadie, Jean, Mirza, Mehdi, Xu, Bing, Warde-Farley, David, Ozair, Sherjil, Courville, Aaron C., and Bengio, Yoshua. Generative adversarial networks. NIPS, 2014.

Karpathy, Andrej, Toderici, George, Shetty, Sanketh, Leung, Thomas, Sukthankar, Rahul, and FeiFei, Li. Large-scale video classification with convolutional neural networks. In CVPR, 2014.

K. Gregor, I. Danihelka, A. Graves, and D. Wierstra. DRAW: A recurrent neural network for image generation. CoRR, abs/1502.04623, 2015.

Le, Quoc V. Building high-level features using large scale unsupervised learning. In ICASSP, pp. 8595–8598, 2013.

Mackie S, Embury O, Old C. Generalized Bayesian cloud detection for satellite imagery. Part 1: Technique and validation for night-time imagery over land and sea[J]. International Journal of Remote Sensing (S0143-1161), 2010, 31(10): 2573-2594

Mathieu M, Couprie C, Lecun Y. Deep multi-scale video prediction beyond mean square error[J]. 2015.

Ranzato, Marc'Aurelio, Szlam, Arthur, Bruna, Joan, Mathieu, Micha¨el, Collobert, Ronan, and Chopra, Sumit. Video (language) modeling: a baseline for generative models of natural videos. CoRR, abs/1412.6604, 2014.

Stowe L L, Davis P A, Mcclain E P. Scientific Basis and Initial Evaluation of the CLAVR-1 Global Clear/Cloud Classification Algorithm for the Advanced Very High Resolution Radiometer[J]. Journal of Atmospheric & Oceanic Technology (S0739-0572), 2010, 16(6): 656-681.

Taravat A, Del Frate F, Cornaro C, et al. Neural Networks and Support Vector Machine Algorithms for Automatic Cloud Classification of Whole-Sky Ground-Based Images[J]. IEEE Geoscience & Remote Sensing Letters (S1545-598X), 2015, 12(3): 666-670.

Wang, Zhou, Bovik, Alan C., Sheikh, Hamid R., and Simoncelli, Eero P. Image quality assessment: From error visibility to structural similarity. IEEE Trans. on Im. Proc., 13(4):600–612, 2004.