A Survey of Deep Learning Radar Echo Extrapolation Networks

Yanping Li[1], Zhiyun Lai[1,*], Qingjun Yang[1], Junye Wei[2], Yu Qian[1]
[1] Qinghai Meteorological Information Center, Xining, Qinghai, China
[2] Geermu Meteorological Bureau, Haixi, Qinghai, China
*Corresponding author: Zhiyun Lai (Email: 26766913@qq.com)

Abstract

Radar echo extrapolation has a great significance for strong convective weather warning and short-term weather approaching now-casting. In this paper we reviewed deep learning radar echo extrapolation algorithms, and categorized them into three types according to their basic architectures - extrapolation networks based on Recurrent Neural Network (RNN), extrapolation networks based on Convolutional Neural Network (CNN) and extrapolation networks based on Generative Adversarial Network (GAN). We introduced the structure of three basic networks, and the evolution of networks developed on these three architectures, then provided a brief overview of their characteristics. This paper aims to find some ways for optimizing deep learning radar echo extrapolation algorithms. The paper also introduced some commonly used open-source datasets for radar echo extrapolation algorithms of their scale and features.

Keywords

Radar Echo Extrapolation; CNN; RNN; GAN.

1. Introduction

Doppler weather radar can detect precipitation particles in the atmosphere by emitting radio waves and receiving the reflected signals. It can monitor and analyze the distribution, intensity, and movement trend of precipitation. It has all-weather observation capabilities with wide coverage, high temporal and spatial resolution. It plays a significant role in warning disastrous weather, and short-term weather approaching now-casting. Weather forecasting model driven by radar, satellite, and other multi-source data has shown performance superior to numerical model forecasts in precipitation forecasting. The accuracy of radar echo extrapolation constrains the accuracy of weather forecasting. Traditional optical flow methods have strict assumptions and usually produce significant errors when facing rapidly developing and changing severe convective systems. With the rapid development of deep learning in computer vision, radar echo extrapolation algorithms based on deep learning are increasingly valued in monitoring and identification severe convective weather. Deep learning can not only mine the spatial characteristics of the generation and dissipation of radar echoes, but also effectively utilize historical echo data for analyzing temporal dimension. It also has a extremely fast speed in computation. Radar echo extrapolation is essentially considered as a time series prediction problem in deep learning. In this paper we reviews the principles and performance of deep learning radar echo extrapolation algorithms, hoping to provide a reference for improving them.

2. Related Theories

The depth of artificial neural networks began to extend to hundreds of layers after gradient explosion problem was solved. Deeper networks allows extracting high-dimensional features, then deep learning starts to flourish. All the deep learning networks used for radar echo extrapolation can

ISSN: 2414-1895 DOI: 10.6919/ICJE.202502_11(2).0007

generally be divided into three architectures: those based on RNN, those based on CNN, and those based on GAN. RNN [1] was initially designed for natural language processing to understand the context of language. Long Short-Term Memory (LSTM) network is a representative network of RNN, its structure has shown in Figure 1. While training LSTM, the input of node t will use the output of node (t-1), thus gives LSTM a certain memory capability. Time series problems usually comes with dependencies between the preceding and following node. The memory capability makes LSTM suitable for processing time series researches.

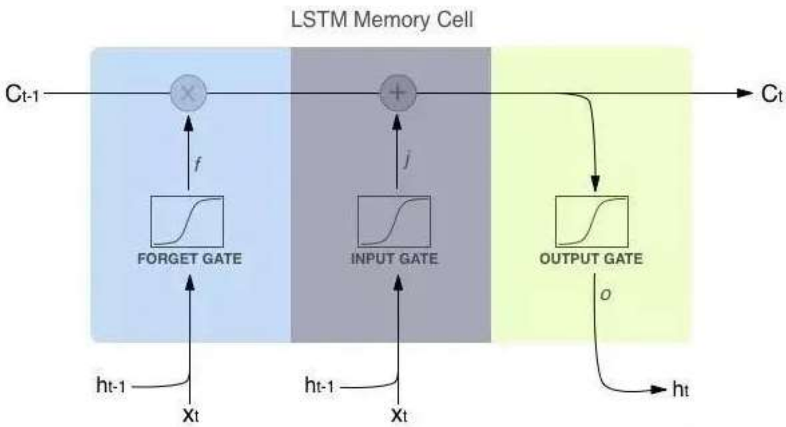

Figure 1. The structure of LSTM.

The basic unit of CNN [2] network consists of several convolutional layer, pooling layer, and activation layer. UNet is a representative example of CNN, its structure has shown in Figure 2. Convolutional layer is used to extract feature, pooling layer to reduce the number of parameters, and activation layer to achieve nonlinear mapping. Typically, CNN uses a fully connected layer to output prediction. CNN has laid the foundation for deep learning with simple structure and less parameters. It has been the mainstream architecture of deep learning networks since its appearance.

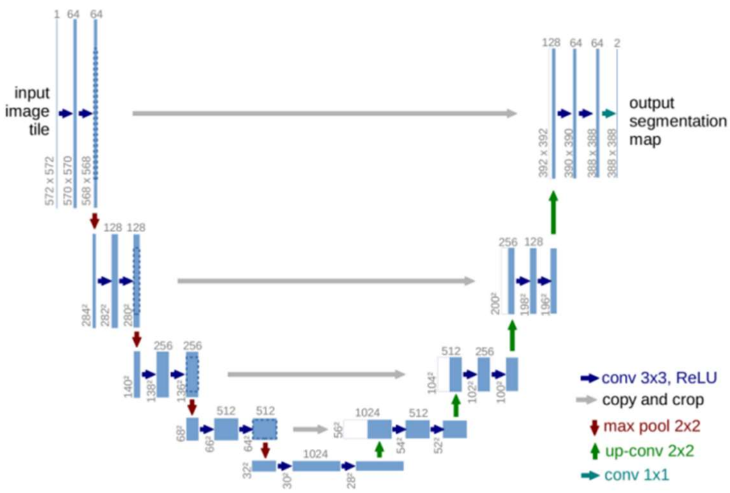

Figure 2. The structure of UNet.

GAN [3] always contains a generator and a discriminator, its structure has shown in Figure 3. The generator is used for producing fake images from random noise, while the discriminator is used for distinguishing between label and fake image generated by the generator. The generator and discriminator are trained in a sort of adversarial game until the generator's output is capable of "fooling" the discriminator.

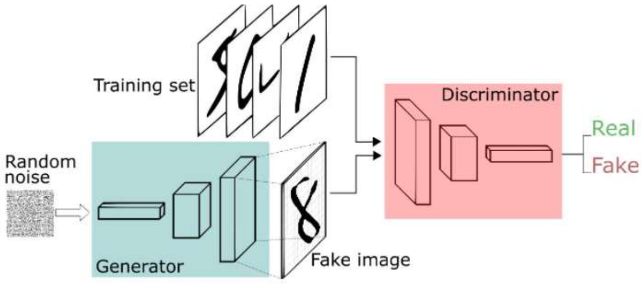

Figure 3. The structure of GAN.

3. Datasets

Radar observations after quality control can serves as dataset for radar echo extrapolation. In addition, some open-source datasets are also available for radar echo extrapolation research. HKO-7 [4] dataset was provided by the Hong Kong Observatory. It contains radar images observed from 2009 to 2015. All the images in HKO-7 are grayscale images with a size 480 × 480. Among them, 812 days of HKO-7 are designed for training, 50 days for validation, and 131 days for testing. Radar echo extrapolation belongs to video prediction in computer vision, thus datasets for video prediction can also be used in radar echo extrapolation. Moving MNIST [5] consists of 10,000 video sequences. Each sequence contains 20 frames with a resolution of 64 × 64. In each video sequence, two digits move independently around the frame. They often crossing each other and bouncing off the edges of a frame. KTH [6] was an action recognition dataset created by the Royal Institute of Technology in Sweden since 2004. KTH includes 6 types of actions - walking, jogging, running, boxing, waving, and clapping in 4 different scenarios performed by 25 peoples. KTH has a total number of 2,391 video samples. The scale variations, clothing changes, and lighting changes of these samples make it the largest human action dataset at the time. KTH makes it possible to evaluate different algorithms on the same input data.

4. Networks based on RNN

Usually, CNNs are used to extract spatial features, and RNNs to provide temporal features. ShiXingjian proposed ConvLSTM [4] in 2017 to integrate the advantages of both networks to mine temporal and spatial features at the same time. LSTM [7] can only mine features in the time dimension, while ConvLSTM extends it to the spatial dimension. ConvLSTM transformed two-dimensional image data into three-dimensional tensor data, and achieved better performance than optical flow methods in radar echo extrapolation tasks. ConvLSTM was a pioneering work in temporal and spatial data mining, and many deep learning radar echo extrapolation methods were inspired by it.

ConvLSTM has a issue of too many parameters and complex construction. ShiXingjian then proposed ConvGRU to solve this problem. ConvGRU [8] could be considered as introducing convolutional operatios into GRU structure. It simplified ConvLSTM's structure and achieved comparable performance to ConvLSTM. ConvLSTM applies the same convolutional operation to different locations of radar echo images. However, spatial features at different locations are actually different, so TrajGRU [9] actively uses recurrent connections to learn structures changing with position. TrajGRU performs slightly better than ConvGRU on the HKO-7 dataset, but the training cost of TrajGRU is higher.

PredRNN [10] consists of ST-LSTM units, and introduces a hidden state M to ConvLSTM. This state flowed temporal and spatial information from the top layer of the previous moment into the bottom layer of the current moment. PredRNN establishes a temporal and spatial information connection between layers. PredRNN can generate clearer echo images and provide more accurate echo center compared with ConvLSTM. PredRNN++ [11] replaced ST-LSTM units in PredRNN with Casual-LSTM units. Casual-LSTM alleviated gradient disappearance problem in PredRNN. PredRNN++ has shown a better performance on KTH and Moving-Mnist datasets than TrajGRU and PredRNN.

ISSN: 2414-1895
DOI: 10.6919/ICJE.202502_11(2).0007

5. Networks based on CNN

UNet [19] was initially designed for medical image segmentation. Then some researchers applied it to radar echo extrapolation and short-term weather approaching now-casting and achieved promising results. UNet is a symmetric "U" shaped architecture. It consists of a downsampling branch for extracting feature, and an upsampling branch for restoring feature resolution. The skip connections between two branches are designed for fusing multi-scale features. UNet is a simple yet effective architecture. Because it is ease to extend or adjust, there has been a series of variant networks developed based on UNet.

SmaAt-UNet [20] incorporated the Convolutional Block Attention Module (CBAM) into encoder to identify important features across channels and image spatial regions. Besides, it replaced regular convolution operation with depthwise separable convolutions to reduce parameters. CBAM takes input image and applies attention sequentially to the channel and spatial dimension. SmaAt-UNet has significantly improved common baselines on precipitation datasets. RainNet [21] followed the encoder-decoder structure, but adjusted network parameters. It has outperformed the baseline Rainymotion in predicting 5-minute precipitation of weather radar echo. Broad-UNet [22] used a asymmetric parallel convolution and Atrous Spatial Pyramid Pooling (ASPP) module to learn more complex patterns. Broad-UNet also has fewer parameters than UNet.

SimVP [23] consists of an encoder, a translator, and a decoder. The encoder is used to extract spatial features, translator to learn temporal evolution, and decoder to integrate temporal information for predicting future frames. SimVP is trained end-to-end with mean squared error (MSE) loss. It achieved state-of-the-art performance on Moving MNIST dataset without introducing any additional tricks or complex strategies. FURENet [24] used a high-dimension fusion strategy to capture high-dimension interactions between multiple variables. It also introduced compression activation module to add a channel-wise attention mechanism. FURENet improved the precision of short-term weather approaching now-casting. There are also some other networks like SE-ResUNet [25] have shown impressive performance in time series prediction.

6. Networks based on GAN

Networks based on GAN consist of a generator and a discriminator. The difference between these networks lies in the type of generator and discriminator. TimeGAN [26] has two parts: an autoencoder component and an adversarial component. The autoencoder consists of an embedding function and a recovery function, and the adversarial component consists of a sequence generator and a sequence discriminator. The key point of TimeGAN is that the two components are trained jointly. That allows simultaneous learn encoding features, generating representations, and iterating across time dimension. Echo images extrapolated by ConvGRU have issues like blurriness, lack of multimodality and skewed distribution. The Generative Adversarial Convolutional Gated Recurrent Unit (GA-ConvGRU) network can overcome these limitations. GA-ConvGRU [27] maintained the overall architecture of GAN and used ConvGRU as generator and CNN as discriminator. Experiments on radar echo extrapolation images collected by the Shenzhen Meteorological Bureau showed that GA-ConvGRU's extrapolation images are more realistic and accurate than ConvGRU.

MoCoGAN [28] generated videos by mapping a series of random vectors to a sequence of video frames. Each random vector consists of a content part and a motion part. While the content part remains fixed, the motion part is realized as a random process. MoCoGAN used Gaussian distribution to model the content space, and the same implementation to generate each frame of the video. It used a RNN to sample from the motion space. Spatiotemporal Residual Predictive Model [29] (STRPM) used an encoding-decoding structure to preserve more temporal and spatial information for high-resolution video. It designed a Residual Predictive Memory (RPM) model to model the temporal and spatial residual features between previous frame and future frame. RPM can also supervise spatial encoder and temporal encoder to respectively extract different features in spatial and temporal domain.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202502_11(2).0007

STRPM is trained by a generative adversarial network with a learned perceptual loss to improve the perceptual quality of predictions.

DGMR's [30] generator consists of a conditional stack and a sampler. The conditional stack is a feedforward convolutional neural network designed to generate conditional representation from the past four frames, and the sampler is a ConvGRU network. DGMR has two discriminators, one for spatial adversarial learning and the other for temporal adversarial learning. The loss functions defined by the temporal and spatial discriminators respectively. NowcastNet [31] is a short-term precipitation forecasting network proposed by a team from Tsinghua University and Central Meteorological Observatory. This network is designed for forecasting extreme precipitation. It integrates physical evolution equation and generative model into a single framework. NowcastNet is divided into two parts: an evolution network and a generative network. The evolution network based on physical mechanism constraints is used to forecasts deterministic changes at 20km scale from past radar images. The generative network uses historical radar echoes and the prediction of the evolution network to generate convective details between 1km and 2km scale.

7. Summary and Outlook

In this paper, we use radar echo extrapolation algorithms as an example to introduce the development of deep learning time series prediction networks in computer vision. We provide a brief overview of the optimization directions to various networks evolved from RNN, CNN, and GAN. We hope to point out several ideas to improve deep learning radar echo extrapolation method. With the rapid development of deep learning, researches related to video frame prediction is increasing. The prediction accuracy is getting better, and the researches develop towards semi-supervised and unsupervised methods. That will further reduce the cost of labeling datasets. As China's radar observation becomes increasingly sophisticated, radar is becoming the most important high temporal and spatial resolution observational equipment. We believes the further researches on deep learning radar echo extrapolation will help Doppler weather radar play a greater role in short-term weather now-casting and disastrous weather warning.

Acknowledgments

This research was supported by Qinghai provincial meteorological bureau's key project "Construction of Artificial Intelligence Application Training Dataset for Short-duration Heavy Precipitation in Qinghai" (QXZD2024-03), the National Meteorological Information Center's open competition mechanism project "Construction of Artificial Intelligence Application Training Dataset for Severe Convective Weather on Qinghai-Tibet Plateau" (09), and Global Atmospheric Background and Qinghai-Tibet Plateau Big Data Application Center Science and Technology Innovation Platform (2023-SF-10).

References

[1] Lipton Z C. A Critical Review of Recurrent Neural Networks for Sequence Learning[J]. arXiv Preprint, CoRR, abs/1506.00019, 2015.

[2] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[3] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.

[4] Shi X, Chen Z, Wang H, et al. Convolutional LSTM network: A machine learning approach for precipitation nowcasting[J]. Advances in neural information processing systems, 2015, 28.

[5] Srivastava N, Mansimov E, Salakhudinov R. Unsupervised learning of video representations using lstms[C]//International conference on machine learning. PMLR, 2015: 843-852.

[6] Schuldt C, Laptev I, Caputo B. Recognizing human actions: a local SVM approach[C]//Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004. IEEE, 2004, 3: 32-36.

[7] Gers F A, Schmidhuber J, Cummins F. Learning to forget: Continual prediction with LSTM[J]. Neural computation, 2000, 12(10): 2451-2471.

[8] Ballas N, Yao L, Pal C, et al. Delving deeper into convolutional networks for learning video representations[J]. arXiv preprint arXiv:1511.06432, 2015.

[9] Shi X, Gao Z, Lausen L, et al. Deep learning for precipitation nowcasting: A benchmark and a new model[J]. Advances in neural information processing systems, 2017, 30.

[10] Wang Y, Long M, Wang J, et al. Predrnn: Recurrent neural networks for predictive learning using spatiotemporal lstms[J]. Advances in neural information processing systems, 2017, 30.

[11] Wang Y, Gao Z, Long M, et al. Predrnn++: Towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning[C]//International conference on machine learning. PMLR, 2018: 5123-5132.

[12] Wu H, Yao Z, Wang J, et al. MotionRNN: A flexible model for video prediction with spacetime-varying motions[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 15435-15444.

[13] Lin Z, Li M, Zheng Z, et al. Self-attention convlstm for spatiotemporal prediction[C]//Proceedings of the AAAI conference on artificial intelligence. 2020, 34(07): 11531-11538.

[14] Guen V L, Thome N. Disentangling physical dynamics from unknown factors for unsupervised video prediction[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 11474-11484.

[15] Wang Y, Zhang J, Zhu H, et al. Memory in memory: A predictive neural network for learning higher-order non-stationarity from spatiotemporal dynamics[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 9154-9162.

[16] Yao Z, Wang Y, Long M, et al. Unsupervised transfer learning for spatiotemporal predictive networks[C]//International Conference on Machine Learning. PMLR, 2020: 10778-10788.

[17] Sønderby C K, Espeholt L, Heek J, et al. Metnet: A neural weather model for precipitation forecasting. arXiv 2020[J]. arXiv preprint arXiv:2003.12140.

[18] Chang Z, Zhang X, Wang S, et al. Mau: A motion-aware unit for video prediction and beyond[J]. Advances in Neural Information Processing Systems, 2021, 34: 26950-26962.

[19] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. Springer International Publishing, 2015: 234-241.

[20] Trebing K, Staṅczyk T, Mehrkanoon S. SmaAt-UNet: Precipitation nowcasting using a small attention-UNet architecture[J]. Pattern Recognition Letters, 2021, 145: 178-186.

[21] Ayzel G, Scheffer T, Heistermann M. RainNet v1. 0: a convolutional neural network for radar-based precipitation nowcasting[J]. Geoscientific Model Development, 2020, 13(6): 2631-2644.

[22] Fernandez J G, Broad-UNet S M. Multi-scale feature learning for nowcasting tasks., 2021, 144[J]. DOI: https://doi. org/10.1016/j. neunet, 2021, 36: 419-427.

[23] Gao Z, Tan C, Wu L, et al. Simvp: Simpler yet better video prediction[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 3170-3180.

[24] Pan X, Lu Y, Zhao K, et al. Improving nowcasting of convective development by incorporating polarimetric radar variables into a deep-learning model[J]. Geophysical Research Letters, 2021, 48(21): e2021GL095302.

[25] Song K, Yang G, Wang Q, et al. Deep learning prediction of incoming rainfalls: An operational service for the city of Beijing China[C]//2019 International Conference on Data Mining Workshops (ICDMW). IEEE, 2019: 180-185.

[26] Yoon J, Jarrett D, Van der Schaar M. Time-series generative adversarial networks[J]. Advances in neural information processing systems, 2019, 32.

[27] Tian L, Li X, Ye Y, et al. A generative adversarial gated recurrent unit model for precipitation nowcasting[J]. IEEE Geoscience and Remote Sensing Letters, 2019, 17(4): 601-605.

[28] Tulyakov S, Liu M Y, Yang X, et al. Mocogan: Decomposing motion and content for video generation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 1526-1535.

[29] Chang Z, Zhang X, Wang S, et al. Strpm: A spatiotemporal residual predictive model for high-resolution video prediction[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 13946-13955.

[30] Ravuri S, Lenc K, Willson M, et al. Skilful precipitation nowcasting using deep generative models of radar[J]. Nature, 2021, 597(7878): 672-677.

[31] Zhang Y, Long M, Chen K, et al. Skilful nowcasting of extreme precipitation with NowcastNet[J]. Nature, 2023, 619(7970): 526-532.