Learnable Optical Flow Network for Radar Echo Extrapolation

Chengwei Zhang, Xudong Zhou, Xiaoyong Zhuge, and Meng Xu

Abstract—**The impact of extreme weather on maintaining flight schedules is becoming more pronounced. Currently, radar echo extrapolation technology is widely used in the nowcasting of severe convection, in which the optical flow method is a representative example of traditional extrapolation algorithms. By training a large number of known samples to find the optimal solution, the deep extrapolation models have gradually become better than the traditional algorithms in recent years. In this study, after examining the optical flow method and other deep learning models, a learnable optical flow deep model with a fully convolutional structure is proposed. Using the convolutional deep learning of optical flow information, this new model can overcome the kernel size limitation of traditional convolutional neural networks, and it can correlate the data history further in time and space. The six-year radar mosaics of Guangdong Province, China, were used as the data set to independently train and verify the new model. The results reveal that the new model outperformed the traditional optical flow method and it is also better than other deep learning models.**

Index Terms—**Deep learning (DL), nowcasting, optical flow, radar echo extrapolation.**

I. Introduction

SEVERE convective weather is one of the most important factors affecting aviation safety and efficiency. For the Pearl River Delta region of Guangdong, China, where flights are numerous and airspace resources are tight, it is particularly important to study nowcasting for severe convection. The objective short-term forecasting technique is mainly classified into two types: 1) numerical prediction modeling and 2) radar extrapolation. The numerical modeling based on atmospheric dynamic equations can provide weather forecasts for as short as a few hours to as long as approximately two weeks. Due to the uncertainty and imperfection of the cloud-related subgrid-scale parameterization scheme, numerical modeling is unable to provide satisfactory and refined convective weather nowcasting at the current technology level. Thus, it is noticeably deficient in aviation meteorological applications where prediction "in the right location at the right time with high accuracy" is highly desired.

Although the radar echo extrapolation method is a phenomenological technique, it is often able to achieve better results in weather nowcasting. Radar echo extrapolation utilizes the fundamental equations of fluid mechanics, and local weather (e.g., precipitation) is constrained by the formation and dissipation terms, as well as the advection term. The advection term is mainly approximated by the radar-retrieved wind field, so the key to extrapolation is to reasonably estimate the sizes of the formation and dissipation terms. The commonly used radar echo extrapolation methods in practice include the centroid tracking method and the cross-correlation method [1]. The centroid tracking method treats thunderstorms as three-dimensional entities for identification and analysis and performs extrapolation by fitting the location of the centroid of the predicted echoes, which is only suitable for the tracking of convective systems [2]. The cross-correlation method determines the motion vector by searching the position with maximum correlation coefficient (CC) on two consecutive radar images. The calculation is simple and can be used for stratiform precipitation and is often used by meteorological agencies [3]. For locally formed or rapidly changing radar echoes, the tracking failures by the cross-correlation method may increase significantly. The optical flow (OptFlow) [4] method proposed by Gibson in the field of computer vision can overcome the above shortcomings. Among these approaches, multiscale optical-flow by variational analysis (MOVA) [5] has currently achieved the best results. Even so, all these methods traditionally must divide the radar echo extrapolation into two stages, estimating the wind field then extrapolating the radar echo. Since they can hardly explain the complex relationship between the formation and dissipation terms and the existing flow fields, they are generally only able to predict the movement of radar echoes over a short time.

In recent years, deep learning (DL) algorithms emerging from the field of computer vision have acquired the ability to hierarchically learn representative and discriminative semantic features from the data [6] and have achieved amazing success in fields such as image recognition, signal processing, and natural language processing [7]. The DL-based radar echo extrapolation model is also inspired by the video prediction algorithm

Manuscript received June 30, 2020; revised September 21, 2020; accepted October 11, 2020. Date of publication October 21, 2020; date of current version January 6, 2021. This work was supported in part by the National Key Research and Development Programs of China under Grant 2018YFC1507302 and Grant 2017YFC1501603, and in part by the National Natural Science Foundation of China under Grant 41505086. *(Corresponding authors: Xiaoyong Zhuge; Meng Xu.)*

Chengwei Zhang is with Shenzhen Air Traffic Management Station, Civil Aviation Administration of China, Shenzhen 518000, China (e-mail: caaczcw@gmail.com).

Xudong Zhou is with Beijing Metled Information Technology Company Ltd., Beijing 100076, China.

Xiaoyong Zhuge is with Nanjing Joint Institute for Atmospheric Sciences, Chinese Academy of Meteorological Sciences, Nanjing 210041, China, and also with the School of Atmospheric Sciences, Nanjing University, Nanjing 210023, China (e-mail: xyzhuge@yeah.net).

Meng Xu is with Jiangsu Climate Center, Jiangsu Provincial Meteorological Bureau, Nanjing 210041, China (e-mail: xumeng_123@126.com).

Digital Object Identifier 10.1109/JSTARS.2020.3031244

[19]–[22], with the difference that the proportion of interest samples on radar images is seriously imbalanced compared to their importance. Those pixels only constitute a small portion of the image but contribute the majority of the severe weather. Therefore, the direct migration of the computer vision method does not perform well, so special treatment must be applied to archive practical results. At present, the state-of-the-art models for radar echo extrapolation in DL are mainly the convolutional long short-term memory network (ConvLSTM) [8] and the improved trajectory gate recurrent unit (TrajGRU) [9]. The evaluation of these methods has indicated that their application has great potential compared with the traditional variational OptFlow method.

The rest of this article is organized as follows. Section II proposes a new DL model, featuring a fully convolutional structure based on the analysis of the ConvLSTM and TrajGRU models. Section III introduces the training and test sets and a comparative experiment design. Section IV compares and analyzes the results. Section V presents the conclusion.

II. Model

ConvLSTM replaces the matrix multiplication in the traditional long short-term memory network (LSTM) [10] model with a convolution operation. Using the other spatiotemporal correlations in the convolution kernel-associated sample space at particular locations, it comprehensively takes advantage of the strength of the convolutional neural network (CNN) [11], [14] and the LSTM. The issue is that in order to detect motions with larger incremental steps, a larger convolution kernel is needed [12], and the learning and calculation of a larger convolution kernel is expensive [13], [14]. A more serious problem is that in the CNN model, which was mainly developed from image recognition and classification, the training process and goal include learning a fixed set of location-independent convolution kernel parameters [14], [16]. For radar echo extrapolation, it is equivalent to forcing the *a priori* constraint that "the motion effects at different points on the sample are equal everywhere," which is problematic (for example, different underlying surfaces clearly have different effects on an echo).

To solve ConvLSTM's fixed convolution kernel problem and shift-invariant problem [14], [15], [17] which are inherently characteristic of CNN networks, TrajGRU uses a shallow CNN model to calculate the radar echo optical flow [23], [24] and then selectively samples the past spatio-temporal information based on the OptFlow approaches. Due to the adoption of the recurrent neural network (RNN), no matter whether it is for forward propagation or backward propagation, the computational cost of the training process and the storage cost of the intermediate process are both very expensive. The TrajGRU model often cannot be stacked to a suitable depth [25], and, thus, the advantages of the deep model cannot be better utilized (more computing power is used in the RNN state transition).

For the above considerations, we proposed the learnable optical flow network (LofNet) model, which predicts the optical flow for each pixel in a sample and the kernel at that location using a fully convolutional network (FCN) [18], and then utilizes this information to obtain the predicted result for the last frame of the input image sequence. FCNs can achieve pixel-level classification of images. Unlike CNNs, which perform the classification by employing fully connected layers to obtain fixed-length feature vectors after the convolution layer, FCNs can accept input images of any size and use the deconvolution layer to upsample the feature vectors of the last convolution layer, thus restoring the size of this layer to the size of the input image. In this way, a prediction can be generated for each pixel while retaining the spatial information in the original input image [27], and the pixelwise classification can then be conducted on the upsampled feature maps. LofNet considers the optical flow at each point [23], [24] to overcome the issue of ConvLSTM's fixed convolution kernel, and the use of a deep network for calculating the optical flow to overcome the problem of the stacking defects caused by RNN in TrajGRU.

The LofNet model is defined as: $\mathcal{G}(I^{t-n}, \ldots, I^{t-1}, I^t; \theta)$. I^t is the radar echo at time t, while $\mathcal{G}$ is an FCN and $(I^{t-n}, \ldots, I^{t-1}, I^t)$ is the sequence of model input. In the radar echo extrapolation task, the radar reflectivity is stacked in the channel direction into the sequence to obtain the input $(I^{t-n-1}, \ldots, I^{t-1}, I^t) \in \mathbb{R}^{N\times H\times W}$. LofNet needs to train and learn the parameter θ

$$\theta = \begin{bmatrix} U \\ V \\ K_h \\ K_v \\ \text{bias} \end{bmatrix}$$

where U and V are the wind field components obtained as the optical flow, while K_h and K_v describe the locally connected kernel; thus, we have $U,\ V \in \mathbb{R}^{H\times W}$, $K_h \in \mathbb{R}^{1\times S\times H\times W}$, and $K_v \in \mathbb{R}^{S\times 1\times H\times W}$. Using separable convolution to divide the local connection weights of size $S\times S\times H\times W$ into two convolutions, K_h and K_v, we can reduce the computational and spatial complexity from $\mathrm{O}(S^2)$ to $2S$ [26]. Finally, the extrapolation is achieved by

$$I^{t+1} = \mathcal{H}\left(\omega\left(I^t,\ U,\ V\right),\ K_h \otimes_{\text{outer}} K_v\right) + \text{bias}.$$

Here, ω denotes a grid-sample operation, and $\mathcal{H}$ is defined as the locally connected layer.

This model is inspired by the spatial transformer network [27]. Unlike TrajGRU and ConvLSTM, LofNet explicitly demands the model to generate a specific output on the corresponding channel by comparing the parameters of the network $\mathcal{G}$ to get the significative result. The innovation of the physical structure of the model is that clear prior knowledge has been instructed for the model to learn to generate reasonable optical flow predictions and the relevant information about the neighborhood of each pixel through K_h and K_v. By reducing the ambiguity of learning, model parameters can be used more effectively; therefore, LofNet can theoretically produce more accurate prediction results.

For severe weather nowcasting, calculation efficiency must be considered if rapid results are required. Easier stacking depth and higher efficiency than RNN are the benefits of CNN which we can take advantage of. In comparison with ConvLSTM and

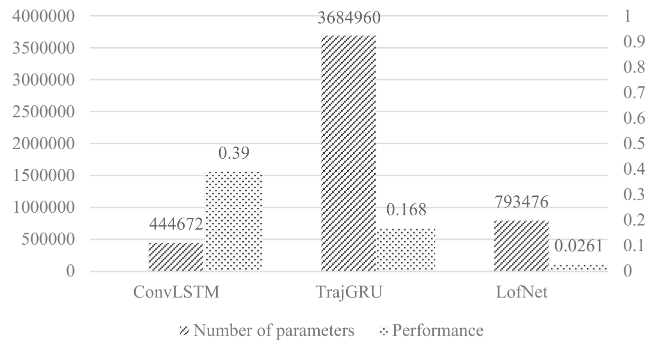


Fig. 1. Comparison of ConvLSTM, TrajGRU, and LofNet in terms of the number of parameters and prediction performance, where the batch size of models is set to 1 to simulate the real-time scenarios. The experiments were run on a computer with a single NVIDIA GTX 1080Ti GPU.

TrajGRU, our model can efficiently utilize the more abstract features. Evaluation shows that we have achieved significant performance improvements which are about 10 times the efficiency improvement while generating reasonable predictions. Fig. 1 compares ConvLSTM, TrajGRU, and LofNet in terms of the number of parameters and prediction performance.

III. Data and Comparative Design

The radar images used were the mosaic constant-altitude plan position indicator of reflectivity factor at 3-km altitude from 9 Doppler weather radars of the Guangdong Provincial Meteorological Bureau between 2013 and 2018. The spatial and temporal resolution for the radar images are 1 km and 6 min, respectively. The number of pixels in each image is 256×256. In this study, the ground clutter and sun spikes were removed [9] and only the images with a maximum echo intensity $>$17 dBZ were retained. After this, about 55 000 samples were taken using a sliding window with a step size of 6 (half an hour) and a window size of 40 (2 h before and after each slide). After the images were sorted by time, the first 50 000 samples were used as the training set, and the last 5000 as the test set (there was no overlap between the training set and the test set). To ensure that the results represent the strong echo situations, all of the samples in the test set were sorted by the area with echo intensity $>$17 dBZ, and 500 cases were then randomly selected from the top 25% of the sorted test set. For these 500 cases, the 6-min interval radar images from the current time through the future 150 min were restored.

In the next section, we will compare the prediction skills of OptFlow, ConvLSTM, TrajGRU, and LofNet models. The OptFlow method does not need a training process. The MOVA algorithm in [5] was used to directly extrapolate the radar echo. The specific configures for the latter three DL models were

ConvLSTM: 2 layers, hidden_size = 64, kernel_size = 3, stride = 1, padding = 1;

TrajGRU: 3 layers, $L = (13, 13, 9)$, hidden_size = (64, 192, 192), kernel_size = (7, 3, 1), stride = (5, 1, 1);

LofNet: 6 layers, num_kernel = (8, 16, 32, 64, 128, 256), kernel_size = 3, stride = 1, padding = 1.

The training set was used to generate the ConvLSTM, TrajGRU, and LofNet models until they converged. Then, the above four models are used to predict the radar image for 500 cases. The maximum prediction lead time is 150 min. Finally, an objective evaluation was carried out between the predicted radar images and the associated "truth" for the four models.

IV. Results

In the evaluation, the radar echo is converted to hourly rainfall intensity by using a $Z{-}R$ relationship in [9]: $\mathrm{dBZ} = 10\log 58.53 + 15.6\log R$, where R is the rainfall intensity with the unit of mm. The evaluation is conducted at the pixel level from the dichotomous and continuous views, respectively.

First, the rainfall intensity was divided into the thresholds of $>$10 mm and $>$1 mm. If the predicted rainfall intensity and the observed value were both greater than a given threshold, a hit occurred. If the observed value exceeded a threshold and the predicted one was less than the threshold, a report was missed. If the predicted value exceeded a threshold, but the observed value did not, a false alarm was made. This study used the following three scores to evaluate the results: the probability of detection (POD), the false alarm ratio (FAR), and the critical success index (CSI):

$$\mathrm{POD} = \frac{\mathrm{Hit}}{\mathrm{Hit} + \mathrm{Miss}}$$

$$\mathrm{FAR} = \frac{\mathrm{False}}{\mathrm{Hit} + \mathrm{False}}$$

$$\mathrm{CSI} = \frac{\mathrm{Hit}}{\mathrm{Hit} + \mathrm{Miss} + \mathrm{False}}$$

where Hit represents the number of hits, Miss represents the number of missed reports, and False represents the number of false alarms.

In addition, we calculated the root-mean-square error (RMSE), and the CC between the predicted value and the actual value, using the following formula:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(R_{p,i} - R_{o,i}\right)^2}{n}}$$

$$\mathrm{CC} = \frac{\sum_{i=1}^{n}\left(R_{p,i} - \overline{R_p}\right)\left(R_{o,i} - \overline{R_o}\right)}{\sqrt{\sum_{i=1}^{n}\left(R_{p,i} - \overline{R_p}\right)^2}\sqrt{\sum_{i=1}^{n}\left(R_{o,i} - \overline{R_o}\right)^2}}$$

where R_p represents the predicted rainfall intensity, $\overline{R_p}$ is its average value, R_o represents the actual rainfall intensity, $\overline{R_o}$ is its average value, and n is the total number of pixels. In these scores, since POD, FAR, and CSI are counted using thresholds, while RMSE and CC calculate all pixels individually, the latter scores are more objective and rigorous.

A. Overall Performance

Fig. 2 presents all of the evaluation scores for the 500 cases. Because each case contained 256×256 pixels, the total number of pixels in the statistics was 32 768 000.

Since the OptFlow method cannot learn the statistical rules of the historical cases, it can be seen from the line graphs

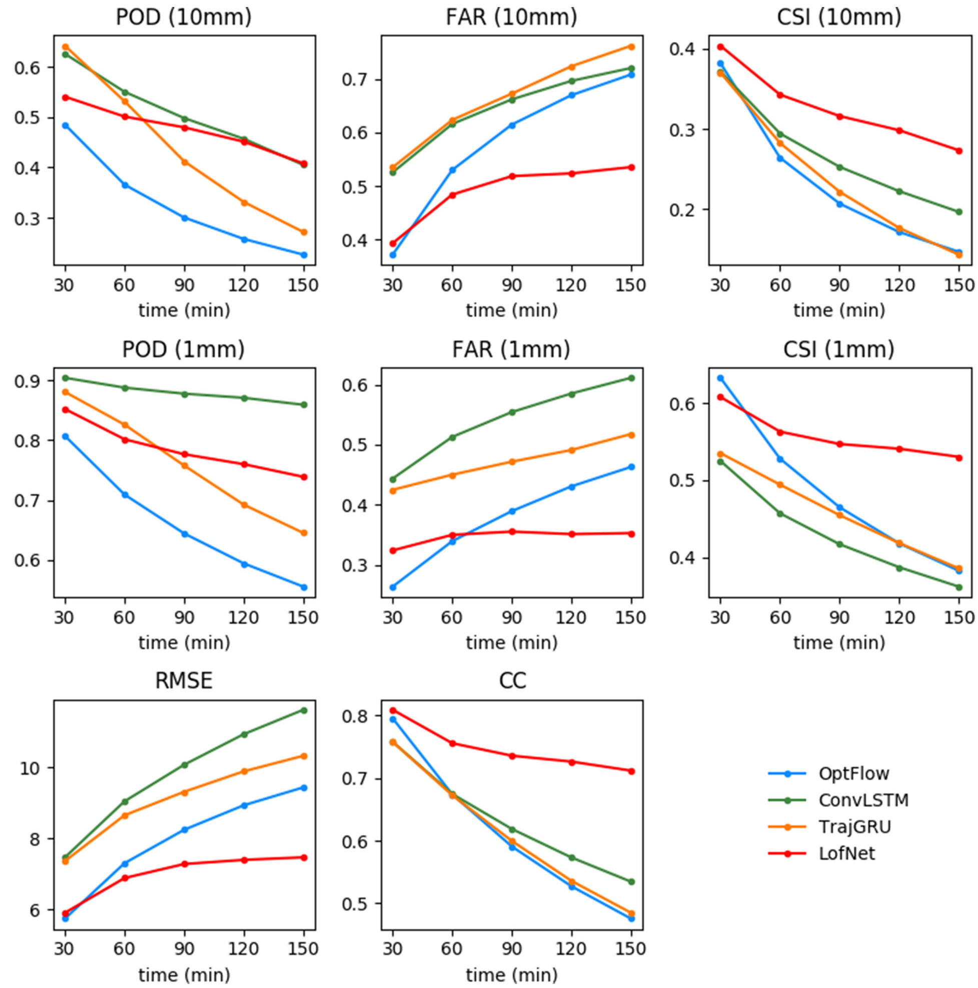


Fig. 2. Evaluation scores of the extrapolation performance of test samples. The first and second rows are scores for 10 and 1 mm rainfall intensity, respectively.

that the three DL models essentially outperformed OptFlow. In particular, as the prediction lead time increased, their advantages became greater. It can also be seen, however, that the RMSE values of the ConvLSTM and TrajGRU models were not as good as those of OptFlow. In terms of POD, ConvLSTM performed better overall (especially for the rainfall intensity of 1 mm), yet, at the same time, its FAR was at a high level, indicating that ConvLSTM tended to overpredict the echo coverage area. In terms of the more rigorous scores, CSI and RMSE, the TrajGRU model was an improvement over ConvLSTM (except for precipitation $>$10 mm, for which its CSI was somewhat less). A noticeable fact is that LofNet performed better than the other models by a large margin (except for the 30-min RMSE, which was slightly inferior to that of OptFlow), which was the closest to the observations, and thus represents the best solution among all of these extrapolation technologies. Furthermore, as the prediction lead time increased, LofNet exhibited increasingly better performance. Compared to OptFlow, LofNet's CSI increased by 85.9%, its CC increased by 49.6%, and its RMSE decreased by 20.9%. Compared to the current best DL models, i.e., the ConvLSTM and TrajGRU (using their best values for comparison), LofNet's CSI increased by 38.9%, its CC increased by 33.1%, and its RMSE decreased by 27.7%.

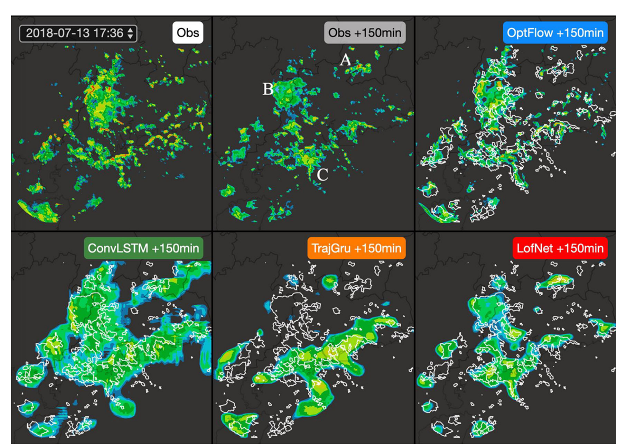


Fig. 3. Comparison of extrapolation results from the case at 17:36 BJT, July 13, 2018. "Obs" represents the observed value. The white lines in the predicted graphs depict the contours of the observed echoes after 150 min. "A" was the newly generated convection and "B" was the weakening echo. The echo at "C" moved and gradually disappeared.

B. Case Analysis

Two cases, one in which the motion of radar echoes was the dominant characteristic, and the other in which the formation

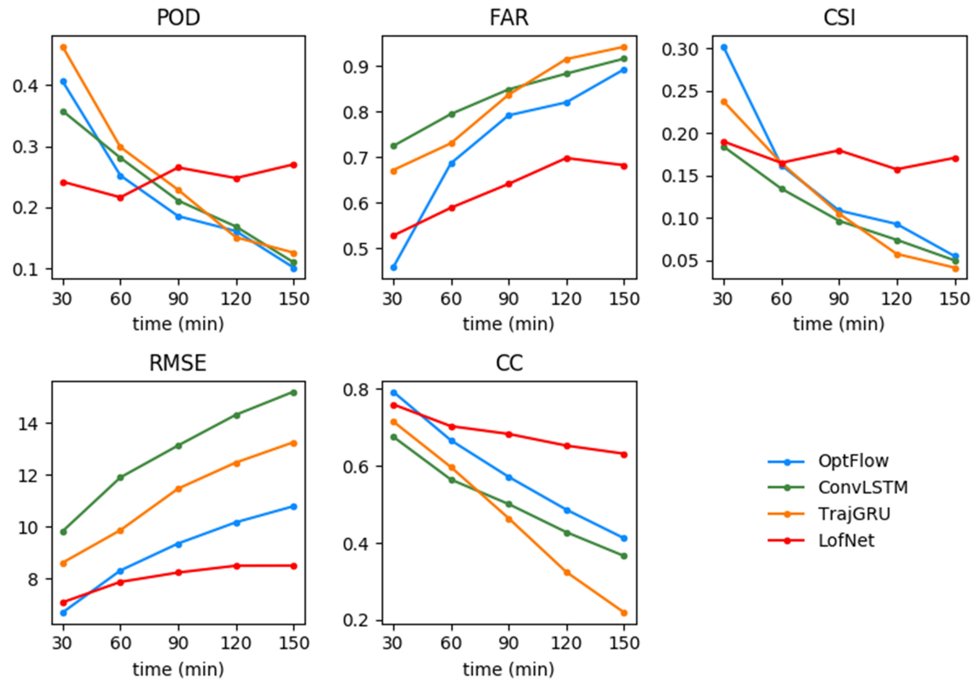


Fig. 4. Evaluation scores of the extrapolation performance of case at 17:36 BJT, July 13, 2018. POD, FAR, and CSI are for 1 mm rainfall intensity.

and dissipation of radar echoes predominated, were selected for comparative analysis. (More details concerning the cross-comparisons of all test set processes as well as their score curves can be found at https://up.metled.com.cn/rc/.)

1) Motion-Dominant Mode: Convective cells are continually forming, developing, and dissipating. The "motion-dominant mode" refers to the scenario in which, under the influence of an essentially consistent environmental wind field, the radar echoes exhibit a tendency of monolithic translation. Figs. 3 and 4 show the comparisons for the extrapolation results reported at 17:36 Beijing Time (BJT; equals to UTC+8), July 13, 2018.

Because the "motion-dominant mode" process happens to follow the traditional extrapolation algorithm, in which the extrapolation is performed by calculating the initial wind field, both the RMSE and CC values of OptFlow exceeded those of ConvLSTM and TrajGRU, and even exceeded LofNet at 30 min. However, as the extrapolation time increased, the performance of OptFlow declined. The new convection marked "A" in Fig. 3 was not predicted, and there were certain misaligned precipitation areas. After all, the environmental wind field cannot stay unchanged. ConvLSTM exhibited areas of false-alarmed precipitation that were too large, as mentioned above. Even though TrajGRU improved on the problem of overly large precipitation areas, it failed to predict the precipitation at "A" and "B" and produced a false alarm at "C," where the precipitation had dissipated. Because LofNet can obtain the optical flow of each pixel through DL, it has pronounced advantages over the other methods, and its prediction at 150 min was closest to the actual observation.

2) Formation and Dissipation-dominant Mode: Figs. 5 and 6 show the comparisons for the extrapolation results reported at 14:48 BJT, June 23, 2018. Obviously, the traditional extrapolation algorithms appeared unable to deal with the scenario of frequent formation and dissipation. ConvLSTM still overpredicted the precipitation areas, and it failed to predicate the dissipation at "F." TrajGRU performed significantly better than ConvLSTM. It successfully predicted the echo formation at "D," and part of the dissipation at "F," although it produced false alarms for a large number of echoes near "E." The RMSE and CC of LofNet were the best, and its results were closest to the observations. LofNet accurately predicted not only the formation and dissipation of echoes but also areas and shapes that were consistent overall with the observations.

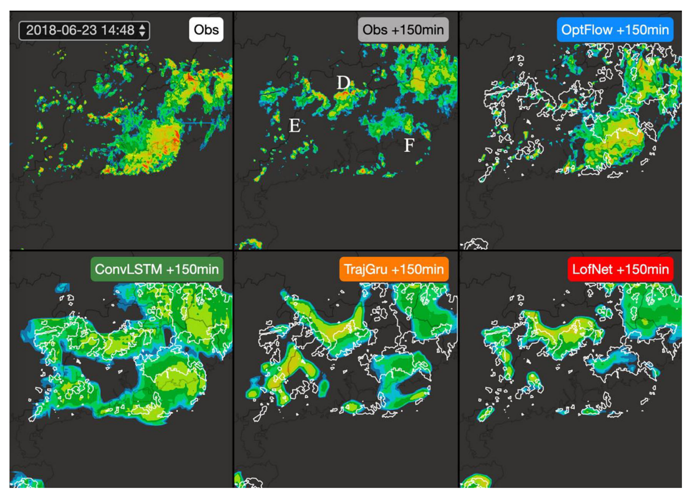


Fig. 5. Comparisons of extrapolated results from the case at 14:48 BJT, June 23, 2018. "Obs" represents the observed value. The white lines in the predicted graphs depict the contours of the observed echoes after 150 min. "D" was the newly generated convection and "E" was free of echo. The large echo area at "F" disappeared.

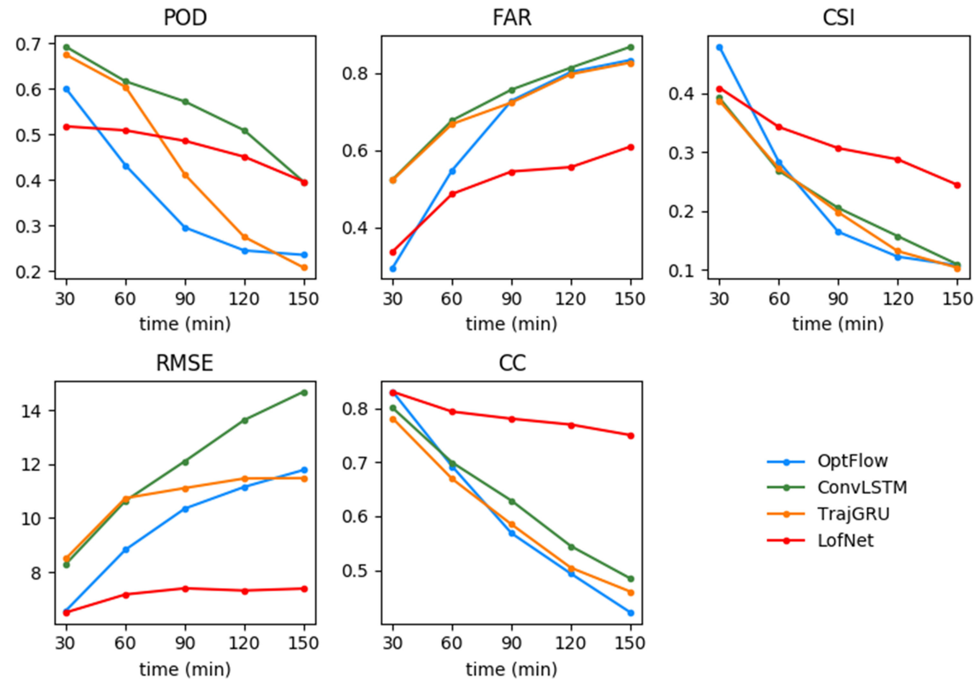


Fig. 6. Evaluation scores of the extrapolation performance of case at 17:36 BJT, July 13, 2018. POD, FAR, and CSI are for 1 mm rainfall intensity.

V. Conclusion

This study described the proposed LofNet model and compared it with the OptFlow method, ConvLSTM, and TrajGRU models for radar nowcasting experiments. The following conclusions are drawn. 1) Traditional algorithms are more suitable only for the situation in which the environmental wind field is consistent overall, i.e., the "motion-dominant" scenario, yet exhibit a poor ability to predict the formation and dissipation of radar echoes, and their ability diminishes rapidly as the prediction lead time increases. 2) The DL model generally performs better than the traditional algorithms, and its advantage becomes more pronounced as the prediction lead time increases. It displays a particular ability to predict echo formation and dissipation, which is worthy of further study and application. ConvLSTM tends to overpredict the precipitation areas. TrajGRU has improved on the overprediction issue to some extent, which indicates that the introduction of optical flow information can improve the prediction ability of DL models. LofNet combines the advantages of ConvLSTM and TrajGRU and learns the optical flow of each pixel through a FCN. It features the best ability and efficiency for spatial feature extraction, i.e., predictions by the LofNet model are closest to the observations among these algorithms.

At the same time, as a nonlinear statistical model, DL has the following limitations when applied to radar echo extrapolation: 1) the training of the model has certain requirements for the amount of historical data, and not many historical cases of severe convective weather are available; and 2) since weak-echo pixels predominate, the prediction results of the model tend to weaken the echoes. In the future, it will be necessary to continue examining the imbalance problem mentioned above regarding the number of radar echo samples and to further optimize the prediction network model.

References

[1] R. E. Rinehart and E. T. Garvey, "Three-dimensional storm motion detection by conventional weather radar," *Nature*, vol. 273, no. 5660, pp. 287–289, 1978.

[2] M. Dixon and G. Wiener, "TITAN: Thunderstorm identification, tracking, analysis, and nowcasting—A radar-based methodology," *J. Atmos. Oceanic Technol.*, vol. 10, no. 6, pp. 785–797, 1993.

[3] Q. Q. Liang *et al.*, "A composite approach of radar echo extrapolation based on TREC vectors in combination with model-predicted winds," *Adv. Atmos. Sci.*, vol. 27, no. 5, pp. 1119–1130, 2010.

[4] J. J. Gibson, *The Perception of the Visual World*. Boston, MA, USA: Houghton Mifflin, 1950.

[5] P. W. Li and E. S. T. Lai, "Applications of radar-based nowcasting techniques for mesoscale weather forecasting in Hong Kong," *Meteor. Appl.*, vol. 11, pp. 253–264, 2004.

[6] L. P. Zhang, L. F. Zhang, and B. Du, "Deep learning for remote sensing data a technical tutorial on the state-of-the-art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, Jun. 2016.

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.

[8] X. Shi, Z. Chen, H. Wang, D. Y. Yeung, W. K. Wong, and W. C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Advances in Neural Information Processing Systems*, Montréal, Canada, 2015, pp. 802–810.

[9] X. Shi *et al.*, "Deep learning for precipitation nowcasting: A benchmark and a new model," in *Advances in Neural Information Processing Systems*, Long Beach, CA, USA, 2017, pp. 5617–5627.

[10] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, pp. 1735–1780, 1997.

[11] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," *Handbook Brain Theory Neural Netw.*, vol. 3361, no. 10, 1995, Art. no. 1995.

[12] M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury, and L. S. Davis, "Learning temporal regularity in video sequences," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogit.*, 2016, pp. 733–742.

[13] A. G. Howard *et al.*, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.

[14] Y. LeCun *et al.*, "Handwritten digit recognition with a back-propagation network," in *Advances in Neural Information Processing Systems.* San Francisco, CA, USA: Morgan Kaufmann Publishers, 1990.

[15] A. Kanazawa, A. Sharma, and D. Jacobs, "Locally scale in variant convolutional neural networks, 2014, *arXiv:1412.5104*.

[16] A. Krizhevsky, "Learning multiple layers of features from tiny images," Master's thesis, Dep. Comput. Sci., Uni. Toronto, 2009.

[17] J. Mairal, P. Koniusz, Z. Harchaoui, and C. Schmid, "Convolutional kernel networks," 2014, *arXiv:1406.3332*.

[18] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.

[19] M. Babaeizadeh, C. Finn, D. Erhan, R. H. Campbell, and S. Levine, "Stochastic variational video prediction," 2017, *arXiv:1710.11252*.

[20] W. Byeon, Q. Wang, R. K. Srivastava, and P. Koumoutsakos, "Fully context-aware video prediction," 2017, *arXiv:1710.08518*.

[21] X. Liang, L. Lee, W. Dai, and E. P Xing, "Dual motion GAN for future-flow embedded video prediction," 2017, *arXiv:1708.00284*.

[22] M. Mathieu, C. Couprie, and Y. LeCun, "Deep multi-scale video prediction beyond mean square error," 2015, *arXiv:1511.05440*.

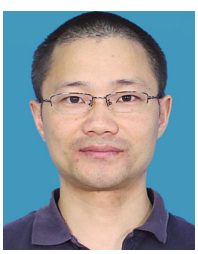
[23] A. Dosovitskiy *et al.*, "FlowNet: Learning optical flow with convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 2758–2766.

[24] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, and T. Brox, "FlowNet 2.0: Evolution of optical flow estimation with deep networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2, 2017, pp. 2462–2470.

[25] R. Eldan and O. Shamir, "The power of depth for feedforward neural networks," 2016, *arXiv:1512.03965*.

[26] C. Szegedy *et al.*, "Going deeper with convolutions," 2015, *arXiv:1409.4842*.

[27] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," 2015, *arXiv:1505.04597*.



Chengwei Zhang received the M.S. and Ph.D. degrees in atmospheric science from Nanjing University, Nanjing, China, in 2008 and 2019, respectively.

He has been a Forecaster with Shenzhen Air Traffic Management Station, Civil Aviation Administration of China, Shenzhen, China, since 2008. His current research interests include remote sensing, machine learning, and aviation products.

Xudong Zhou received the B.S. degree in atmospheric science from Nanjing University of Information Science & Technology, Nanjing, China, in 2015.

Since 2017, he has been a Research Engineer with Beijing Metled Information Technology Company Ltd., Beijing, China. His research interests mainly focus on big data and deep learning.

Xiaoyong Zhuge received the B.S. and Ph.D. degrees in atmospheric science from Nanjing University, Nanjing, China, in 2009 and 2014, respectively.

From 2014 to 2020, he was an Assistant Professor with the School of Atmospheric Sciences, Nanjing University. Since 2020, he has been an Associate Professor with Nanjing Joint Institute for Atmospheric Sciences, Chinese Academy of Meteorological Sciences, Nanjing, China. His current research interests include satellite remote sensing and data assimilation.

Meng Xu received the M.S. degree in atmospheric science from Nanjing University, Nanjing, China, in 2006.

He is currently working with the Jiangsu Climate Center, Nanjing, China. He has been committed to the application of multisource satellite remote sensing data in weather and ecology, as well as in the construction of comprehensive meteorological observation systems.