

MSF: One Lightweight Deep Learning Nowcasting Method with Attention Mechanism using Dual-Polarization Radar Observations

Kexin Zhu¹, Haonan Chen¹, and Lei Han²

¹Colorado State University, Fort Collins, CO

² College of Information Science and Engineering, Ocean University of China, Qingdao, China

Abstract—Research on nowcasting through dual-polarization weather radar data using deep learning approach is rare but worth exploring. This paper lightens a previous work, the MCT (Multivariate Channel Transformer) model, which leads to the design of the MSF (Multivariate Swin Fusion) model. The commonalities between the two are as follows: on one hand, both fuses several dual-polarization observables including reflectivity (Z), specific differential phase (K_{dp}), and differential reflectivity (Z_{dr}) to more comprehensively consider meteorological particle features; on the other hand, they introduces the attention mechanism to more fully fuse multi-frame, multi-variate, and multi-scale features. In the experimental evaluation, this study first selects observation data from KMLB radar in FL, USA, and uses traditional optical flow method, deep learning TrajGRU method, etc. as controls. The results show that both MCT and MSF perform better than the control, and the 60min forecast scores of both are 8.78/9.31 for RMSE and 0.46/0.18/0.07 for CSI (20/35/45dBZ), and this conclusion is verified by case study. Further, the role of the attention mechanism is verified by ablation experiments.

Index Terms—Nowcasting, dual-polarization radar, self-attention, deep learning

I. INTRODUCTION

Strong convective weather often brings great harm to the production and life of the society, and nowcasting based on radar data is an effective way to reduce its harm [1]. In addition to radar reflectivity factor (Z_h), polarization radar can additionally provide observation variables such as differential reflectivity factor (Z_{dr}) and specific differential phase shift (K_{dp}), which enriches the monitoring capability of strong convective weather [2], [3]. Deep learning continues to evolve in the field of nowcasting, where the attention mechanism of Transformer excels in capturing correlations in sequences. Previous work [4] proposed the Multivariate Channel Transformer (MCT), which feeds the encoding results of different semantic layers to the Channel Transformer, thus generating feature maps of different semantic layers containing global information, which are delivered to the decoder of the corresponding layer. Applying the MCT model to the nowcasting, the information from moments $T-J$ to T can be input to obtain the predictions from $T+1$ to $T+K$. The inputs include Z , Z_{dr} , and K_{dp} ; the outputs are the combination reflectivity.

However, a smaller number of network parameters with guaranteed forecast performance bring more flexibility to the

task, such as wider geographic coverage. Hence, lightweight networks are also worth investigating. By analyzing the difference between computer vision and natural language processing, the authors of swin Transformer propose two concepts for the application of Transformer to images [5]: hierarchical features and sliding windows, where the concept of windows allows the computational cost to be reduced from the square of the feature map to the square of the window.

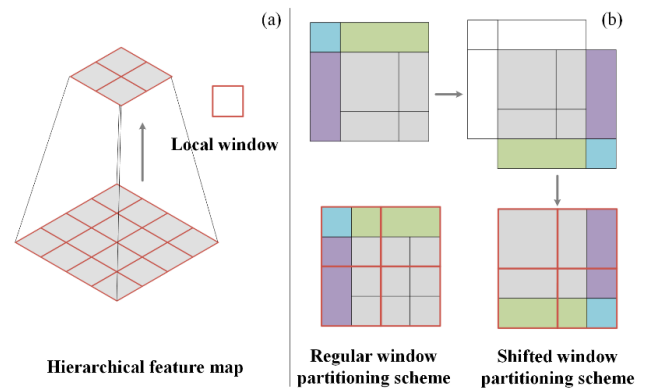


Fig. 1. Two main contributions in Swin Transformer. (a) hierarchical architecture. (b) shifted windows.

In this study, a lightweight improvement is made to design the Multivariate Swin Fusion (MSF) model. Based on the commonality, this concept is called Multivariate Transformer Nowcasting framework (MTN).

The remainder of this paper is organized as follows. Section II details the methodology, and Section III analyzes the experimental results. Summary is presented in Section IV.

II. METHODOLOGY

A. Background: overview of MTN

The multiscale encoder-decoder (ED) structure is chosen as the basic network. Based on the ED framework, MTN comes up to improve the quality of the nowcasting. Specifically, MTN refers to a concept of nowcasting that makes use of the attention mechanism in Transformer, where multiple variables are used (see Fig. 2(a)).

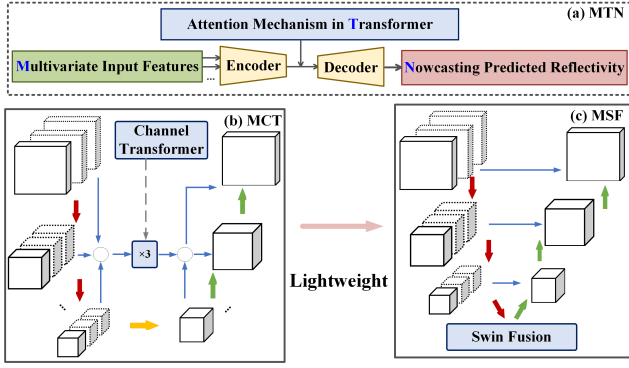


Fig. 2. MTN framework. (a) Core ideas of MTN framework. (b) Outline of MCT model. (c) Outline of MSF model.

Based on this framework, this paper lightens the MCT and designs the MSF model. Compared to MCT, MSF only performs the feature fusion of the map at the deeper semantic layer and reduces the network parameters by swin Transformer technique. More details are shown in TABLE I.

TABLE I
DETAILS OF THE TWO MODELS

Model	attention mechanism is used in	Params	Transformer parameters			
			patch size	block num	head num	MLP ratio
MCT	connector	170.09M	16,8,4,2	4	3	3
MSF	layer 4-5	54.77M	4,2	2,4	2,4	3

B. MSF: Implementation model

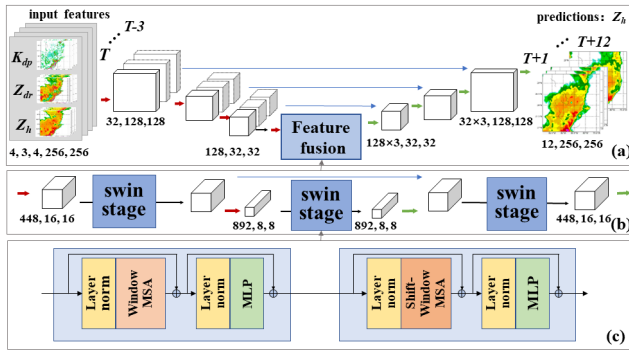


Fig. 3. MSF model (a) The way the MSF model is introduced in the multi-scale Encoder-Decoder. (b) The way the swin Transformer fuses features. (c) The specific implementation principle of the swin Transformer.

Long-range correlations of futures are expected when fusing the feature, so we turn to Swin Transformer. Due to the expensive cost of attention operations, MSF model has been improved in two aspects:

- i Swin Transformer is only used for deeper layers with smaller spatial size (Fig 3. (a) and (b)).
- ii Swin Transformer itself limits the computation scope to every window (Fig 3. (c)).

Specifically, layer 4 has two swin Transformer stages, layer 5 has one. Before entering the fusion structure, the feature map will get through patch embedded (patch size = 2) and convolution. In the first block operation at layer 4, the size of feature map entered into Swin blocks is 16×16 , which is evenly partitioned into 4×4 windows of size 4×4 ($M=4$). Then, the next module adopts a windowing configuration by displacing the windows by $(\lfloor \frac{M}{2} \rfloor, \lfloor \frac{M}{2} \rfloor)$, i.e., 2×2 pixels from the regularly partitioned windows. By down-sampling and patch merging, size of feature map becomes 8×8 , it is sent to blocks in layer 5 and repeats the block operations.

C. Comparison Methods

There are three comparison methods, covering both traditional and deep learning methods. The details are as follows:

1) *ED*: Encoder-Decoder structure. It is the base model for MCT/MSF, and uses the same model parameters, but no attention mechanism is introduced.

2) *TGRU*: TrajGRU proposed by [6]. It uses a three-layer encoding-forecasting structure with the number of filters for the RNNs set to 64, 128, and 128. The kernel size is set to 3×3 . It is a benchmark deep learning method for comparison. The input is 4-frame, composite reflectivity.

3) *OF*: Optical flow. Traditional method for comparison. It uses Lucas Kanade method for Lagrange extrapolation which is implemented by PySteps. The input is 4-frame, composite reflectivity.

D. Evaluation Metrics

Probability of detection (POD), false alarm ratio (FAR), critical success index (CSI) and root mean square error (RMSE) are used as the nowcasting evaluation metrics, which are commonly used in weather predictions.

$$\begin{aligned}
 POD &= \frac{TP}{TP + FN} \\
 FAR &= \frac{FP}{TP + FP} \\
 CSI &= \frac{TP}{TP + FN + FP}
 \end{aligned} \tag{1}$$

The reflectivity factor above a certain threshold is considered a positive sample. TP means that the predicted positive one is actual, FN means samples that are not predicted, but are positive indeed, and FP means that a negative sample is wrongly predicted as a positive one.

RMSE can measure pixel-level error, it is defined as:

$$RMSE = \sqrt{\frac{1}{HW} \sum_{i=1}^W \sum_{j=1}^H (f_{i,j} - \bar{f}_{i,j})^2} \tag{2}$$

For good results, we expect higher POD, lower FAR, higher CSI, and lower RMSE.

TABLE II
QUANTITATIVE EVALUATION OF DIFFERENT METHODS FOR
THE 30 MIN NOWCASTS

	Z (dBZ)	POD	FAR	CSI	RMSE
MCT	20	0.79	0.31	0.59	6.57
	35	0.61	0.62	0.31	
	45	0.29	0.75	0.16	
MSF	20	0.81	0.32	0.59	6.83
	35	0.7	0.65	0.3	
	45	0.44	0.79	0.16	
ED	20	0.8	0.47	0.47	12.67
	35	0.63	0.78	0.2	
	45	0.3	0.91	0.07	
TGRU	20	0.78	0.35	0.55	7.57
	35	0.54	0.69	0.25	
	45	0.19	0.81	0.1	
OF	20	0.63	0.29	0.5	7.8
	35	0.31	0.65	0.2	
	45	0.11	0.87	0.06	

TABLE III
QUANTITATIVE EVALUATION OF DIFFERENT METHODS FOR
THE 60 MIN NOWCASTS

	Z (dBZ)	POD	FAR	CSI	RMSE
MCT	20	0.7	0.42	0.46	8.78
	35	0.41	0.76	0.18	
	45	0.13	0.88	0.07	
MSF	20	0.73	0.45	0.46	9.323
	35	0.49	0.78	0.18	
	45	0.17	0.89	0.07	
ED	20	0.73	0.57	0.37	15.782
	35	0.44	0.87	0.12	
	45	0.15	0.96	0.03	
TGRU	20	0.68	0.47	0.42	9.93
	35	0.35	0.81	0.14	
	45	0.08	0.88	0.05	
OF	20	0.46	0.43	0.34	10.2
	35	0.12	0.86	0.07	
	45	0.02	0.98	0.01	

III. RESULTS AND EVALUATION

A. Quantitative Results

TABLES II to III show results for the 30- and 60-min nowcasts of different methods. Specific to the performance of each prediction time, firstly, at 30 min, the performance of MSF is gradually approaching that of MCT. Both are superior to other methods. When assessing the 60 min forecast performance, the CSI for both at 35 dBZ are 0.18, while ED/TGRU/OF are only 0.12/0.14/0.07, respectively. The overall POD of MSF is higher, indicating that this model is biased towards hits and more likely to forecast high reflectivity compared to MCT. Combined with CSI, MSF achieves similar results to MCT with fewer parameters in weather capture capability. However, in terms of the overall error RMSE, the MCT is still better.

B. Case Study

Fig. 4-5 show the information on June 24, 2022 at 23:05 UTC. It can be seen from Fig. 4 that both MSF and MCT achieve more favorable CSI. Fig. 5 verifies the performance of MCT and MSF, as evidenced by the fact that MCT predicts the shape of the upper and lower clouds most closely, and that reflectivity larger than 40 dBZ are predicted only by MCT

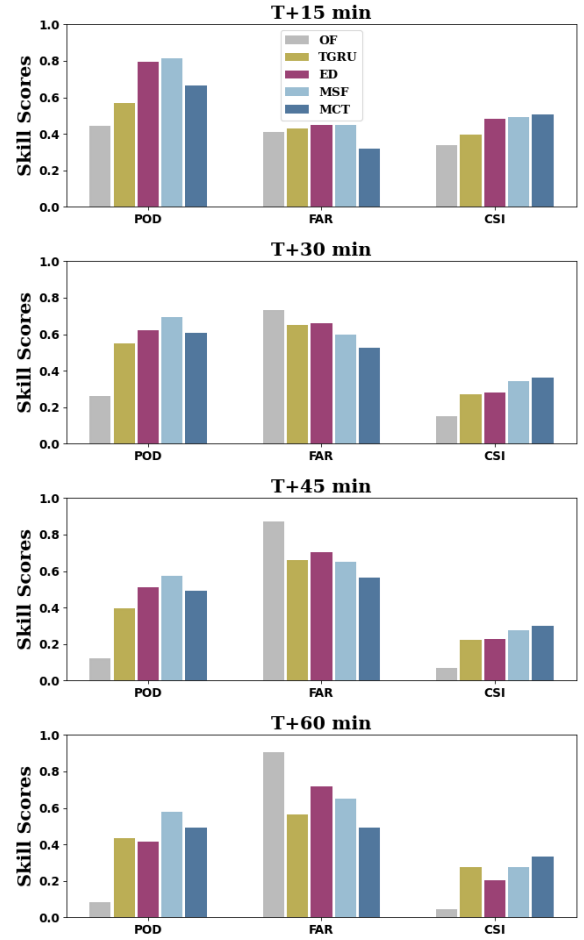


Fig. 4. Skill scores including POD/FAR/CSI where T is 2305 UTC 24 Jun. 2022.

and MSF, and reflectivity larger than 45 dBZ are captured only by MSF. For the case with smaller motion velocity, the prediction of motion by OF is also conservative, and 60min also makes a slight reversal in the prediction of the original figure, and cannot predict the information of growth, merging, and deformation. The prediction of the overall shape of the cloud mass and the position of the peak energy by TGRU and ED have large deviation.

C. Ablation Experiment

In order to verify the contribution of the attention mechanism to the nowcasts, ED-3 model is designed in this paper for ablation experiment. It is Encoder-Decoder structure and the base model for MCT/MSF, using the same model parameters, but no attention mechanism introduced. The input is Z , Z_{dr} , K_{dp} . In ablation experiment, POD/FAR/CSI (thresholds = 35 dBZ) and RMSE are chosen as evaluation metrics to assess the 12-frame nowcast results of the test set. Fig. 6 shows the model with the attention mechanism removed has a substantial decrease in the accuracy of the forecasts. In this way, the effect of attention mechanisms is demonstrated.

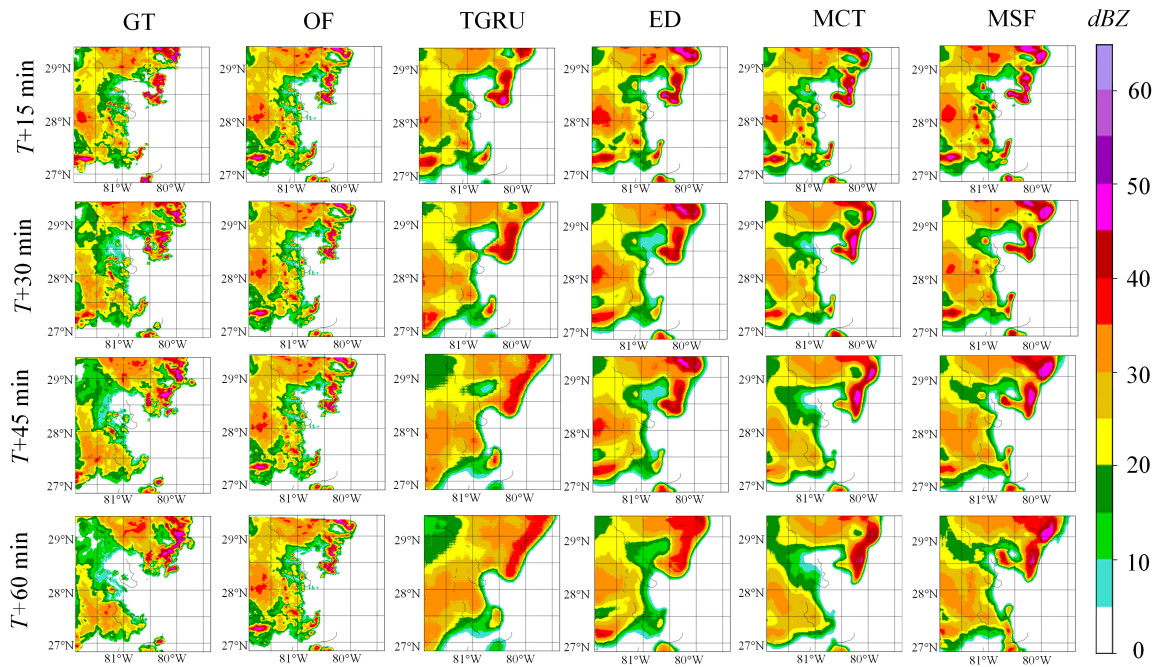


Fig. 5. The 60 min nowcasting results in the study domain with five different methods, where T is 0216 UTC on June 24, 2022.

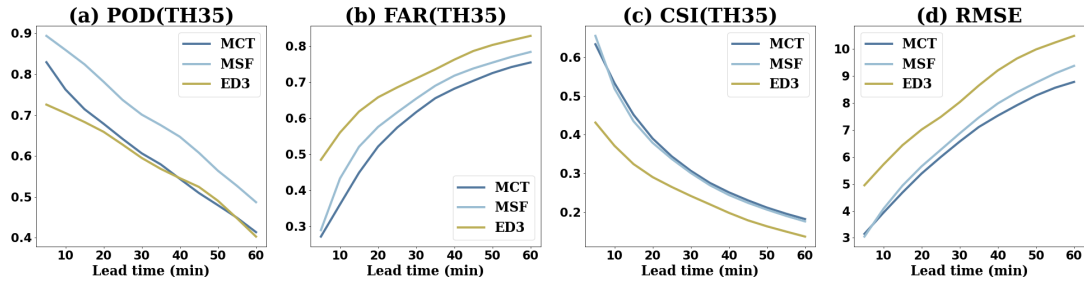


Fig. 6. Trend of score metrics over time for MCT, MSF, ED3. (a) POD; (b) FAR; (c) CSI; The thresholds of the above three are 35 dBZ; (d) RMSE.

IV. SUMMARY

This article articulates the MSF model and its performance on nowcasting. The MSF model is refined based on the MCT model with reduced model parameters. The advantages of both models are shown in the quantitative results and are validated in the case study. In addition, the potential of the application of the attention mechanism is demonstrated in the results of the ablation experiments. From the prediction of reflectivity to the precipitation will be the future research prospect of this work. To explore the impact of polarization parameters on deep-learning networks for nowcasting is another worthy idea to try for future study.

ACKNOWLEDGMENT

The research was supported by the U.S. National Science Foundation (NSF) CAREER program. The work of Lei Han was supported by the National Natural Science Foundation of China (grant number 42275003). The authors gratefully acknowledge the support of NVIDIA Corporation for the donation of GPU used for this research.

REFERENCES

- [1] L. Han, J. Sun, and W. Zhang, "Convolutional neural network for convective storm nowcasting using 3-d doppler weather radar data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 2, pp. 1487–1495, 2019.
- [2] H. Chen, V. Chandrasekar, and R. Bechini, "An improved dual-polarization radar rainfall algorithm (DROPS2. 0): Application in NASA IFloodS field campaign," *Journal of Hydrometeorology*, vol. 18, no. 4, pp. 917–937, 2017.
- [3] Y. Gou, H. Chen, H. Zhu, and L. Xue, "Microphysical processes of super typhoon lekima (2019) and their impacts on polarimetric radar remote sensing of precipitation," *Atmospheric Chemistry and Physics*, vol. 23, no. 4, pp. 2439–2463, 2023. [Online]. Available: <https://acp.copernicus.org/articles/23/2439/2023/>
- [4] K. Zhu, H. Chen, and L. Han, "Mct u-net: A deep learning nowcasting method using dual-polarization radar observations," in *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, 2022, pp. 4665–4668.
- [5] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 10012–10022.
- [6] X. Shi, Z. Gao, L. Lausen, H. Wang, D.-Y. Yeung, W. kin Wong, and W. chun Woo, "Deep learning for precipitation nowcasting: A benchmark and a new model," 2017.