Nowformer : A Locally Enhanced Temporal Learner for Precipitation Nowcasting

Jinyoung Park
KAIST
jinyoungpark@kaist.ac.kr

Inyoung Lee
KAIST
inzero24@kaist.ac.kr

Minseok Son
KAIST
ksos104@kaist.ac.kr

Seungju Cho
KAIST
joyga@kaist.ac.kr

Changick Kim
KAIST
changick@kaist.ac.kr

Abstract

The precipitation video datasets have distinctive meteorological patterns where a mass of fluid moves in a particular direction across the entire frames, and each local area of the fluid has an individual life cycle from initiation to maturation to decay. This paper proposes a novel transformer-based model for precipitation nowcasting that can extract global and local dynamics within meteorological characteristics. The experimental results show our model achieves state-of-the-art performances on the precipitation nowcasting benchmark.

1 Introduction

Climate change has induced heavy downpours in many parts of the globe, causing significant damage to society [1, 2, 3, 4, 5, 6]. Therefore, predicting short-term precipitation changes in advance is becoming important and has received increasing interest from researchers [7, 8, 9, 10]. Precipitation nowcasting predicts precipitation changes within 6 hours, predicting and responding to rapid changes in real time [2, 6, 11, 12].

Deep learning-based precipitation nowcasting tasks are defined to predict future precipitation conditions using satellite videos or radar measurements. Many previous models leverage the architecture of video prediction, a similar task to nowcasting, focusing on exploiting spatial and temporal information by applying convolutional layers and transformer blocks [5, 9, 10, 11, 13, 14]. However, precipitation data have unique characteristics that differ from common video prediction benchmarks. The general video has a moving specific rigid body with a stationary background [15, 16, 17], whereas the precipitation video features a fluid mass spreading in a particular direction. Furthermore, each area of the fluid has an individual life cycle from initiation to maturation to decay; that is, the intensity of the moving fluid changes continuously [18]. Therefore, a precipitation nowcasting model reflecting meteorological patterns is required.

In this paper, we propose a novel transformer-based model for precipitation nowcasting, which can extract global and local dynamics within meteorological characteristics. The global dynamic attention module can extract global spatial features from the frames so that the model can predict long-term future frames with high accuracy, while locally extracted temporal relationships from the local dynamic attention module help the model learn how the intensity of fluids changes at each point in the frames. Our proposed model shows the

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2021.

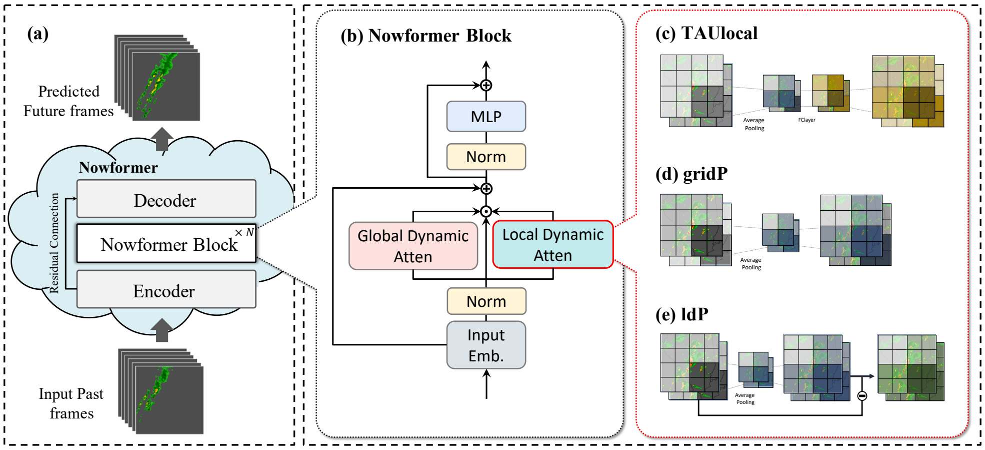

Figure 1: (a) Overall framework of Nowformer. (b) Nowformer block. (c-e) The variants of the local dynamic strategies, which are named TAUlocal, gridP, and ldP.

best performance compared with other state-of-the-art video prediction and precipitation nowcasting models in the SEVIR precipitation nowcasting benchmark [2].

2 Methodology

2.1 Overall Pipeline

We formulate precipitation nowcasting as a spatiotemporal predictive learning problem. Given past T frames of precipitation, the model predicts the future T' frames. As shown in Fig. 1 (a), the overall structure of the nowformer contains an encoder, nowformer blocks, and a decoder. The encoder and decoder comprise convolutional blocks that extract features and map features to predictions. The nowformer block captures spatiotemporal patterns from frame features, as detailed in Section 2.2. After the nowformer block, an additional residual connection from the encoder to the decoder layer is applied to directly use the past frame information. We train the nowformer using the mean squared error (MSE) loss between the predicted precipitation frames and its target.

2.2 Nowformer Block

Each nowformer block has two subblocks of spatiotemporal attention and multi-layer perceptron (MLP) [19, 20]. As shown in Fig. 1, the nowformer block module is formulated as follows:

$$Z_l = \texttt{MLP}(BN(Z'_l)) + Z'_l, \quad Z'_l = \texttt{Attention}(BN(Z_{l-1})) + Z_{l-1}, \tag{1}$$

where, given the previous block features Z_{l-1} as an input of the $l-th$ block, Z'_l and Z_l are the outputs of the `Attention` module and `MLP` module, respectively. Through the attention module, the model can capture the fluid's average flow direction and variations in intensity based on the fluid's life cycle in each area. Batch normalization (BN) is used before each subblock, and a residual connection is used after each subblock.

Attention. To capture the spatiotemporal pattern of precipitation features, we build the attention block with two core designs: (1) global dynamic attention (`GlobalDynamicAttn` and (2) local dynamic attention (`LocalDynamicAttn`), which is

$$\texttt{Attention} = (\texttt{GlobalDynamicAttn} \otimes \texttt{LocalDynamicAttn}) \otimes F, \tag{2}$$

where $F \in \mathbb{R}^{C\times H\times W}$ is an input feature and $\texttt{GlobalDynamicAttn}, \texttt{LocalDynamicAttn} \in \mathbb{R}^{C\times H\times W}$ denote an attention map, respectively. The score on the attention map represents the importance of each feature, and $\otimes$ denotes the Hadamard product.

Global Dynamic Attention. For long-term future predictions and rapidly changing precipitation events such as storms, it is important to capture global spatial relationships to guide the model on where fluid moves [5]. Inspired by previous studies [21, 13], we also apply large kernel convolutions in global dynamic attention to generate a broad receptive field and mine global spatial relationships with long-range dependencies The global dynamic attention calculation is formulated as follows:

$$\texttt{GlobalDynamicAttn} = Conv_{1\times 1}(DWDConv(DWConv(F)). \tag{3}$$

We apply 3×3 depthwise convolution ($DWConv$) to extract local spatial information, depthwise dilated separable convolution ($DWDConv$) to generate long-range spatial features, and 1×1 convolution ($Conv_{1\times 1}$) to model the temporal-wise convolution at a pixel level. Our global dynamic attention allows the model to learn the movement of fluids captured from the global view while using only at a minimum computational cost.

Local Dynamic Attention. When performing precipitation nowcasting focusing on global temporal evolution, a key issue is that valuable precipitation life-cycle details in local areas are removed, making the prediction results less accurate. Since neighboring pixels are crucial references that commonly share a life-cycle, we suggest temporal dynamics modeling across the local area. Specifically, as shown in Fig. 1(c-e), we propose three different local dynamic attention strategies to fully capture local temporal evolution: (1) `TAUlocal`, (2) grid pooler (`gridP`), and (3) local dynamic pooler (`ldP`).

`TAUlocal` learns the local dynamics in a squeeze-and-excitation manner through the grid on the feature maps. We first split the features into grids and apply an average pooling layer in each grid to capture the average movement of the local area. Finally, MLP is applied within the common grid area to capture local temporal evolution.

`gridP` is a simple but powerful module. We split the features into grids and apply an average pooling layer in each grid to capture the average movements of the local area at each time point; we then use the normalized local information directly as attention weights.

`ldP` focuses on temporal gradients. We first average `gridP` weights across time to capture average temporal movement and calculate the temporal gradient of `gridP` by simply subtracting `gridP` weights from average temporal movement.

3 Experiments

3.1 Experimental Setup

Dataset. A storm event imagery dataset (SEVIR) [2] offers radar and satellite images for various weather conditions. We adopt the vertically integrated liquid (VIL) data, comprising a sequence of images taken at intervals of 5 minutes, each one spanning a 384 km $\times$ 384 km region. The goal of the nowcasting task is to predict the precipitation for the following 12 frames (1 hour) using the previous 13 frames (about 1 hour) as input.

Evaluation metrics. We use the two commonly-used precipitation nowcasting evaluation metrics and two image quality metrics to evaluate the performance: the critical success index (CSI = $\frac{\#Hits}{\#Hits+\#Misses+\#F.Alarms}$) [2], probability of detection (POD = $\frac{\#Hits}{\#Hits+\#Misses}$) [2], mean absolute error (MAE), and MSE. When the pixel values of the target and prediction are binarized as a threshold value, the number of pixels with target=prediction=1 is "$\#Hits$", and target=1, prediction=0 is "$\#Misses$", conversely, target=0, prediction=1 is "$\#F.Alarms$". CSI-N and POD-N indicate that the threshold value of each metric is N.

3.2 Experimental Results

The experimental results are summarized in Table 1. All three modules, TAUlocal, gridP, and ldP showed better performances than the baselines in CSI, POD, MAE, and MSE. Moreover, when considering the number of parameters and GFLOPs, our models are comparable with other methods in terms of complexity, showing efficiency and scalability. The performance measured per frame is shown in Fig. 2, and we can verify that our methods outperformed baseline models for all frames as well as average scores. Better performances at all frames

Table 1: Comparison of our methods with baselines on several metrics.

Model	Param. (M)	Metrics						
		GFLOPs	CSI-74	CSI-133	POD-74	POD-133	MAE↓	MSE(10^{-2}) ↓
Persistence	–	–	0.4766	0.4500	0.6072	0.5814	23.60	35.81
UNet [22]	4.14	4.79	0.6201	0.5820	0.7013	0.6547	21.67	24.48
Rainformer [5]	212.40	50.89	0.6340	0.6055	0.7559	0.7209	20.57	23.79
SimVP [14]	14.03	74.01	0.6507	0.6207	0.7583	0.7231	19.65	23.67
TAU [13]	11.25	55.55	0.6453	0.6152	0.7543	0.7180	19.32	22.20
Nowformer w/ TAUlocal	10.78	55.56	0.6457	0.6183	0.7861	0.7461	19.95	21.30
Nowformer w/ gridP.	10.53	55.56	0.6551	0.6309	**0.8091**	**0.7786**	20.37	22.51
Nowformer w/ ldP.	10.53	55.56	**0.6592**	**0.6331**	0.7881	0.7567	**18.91**	**21.22**

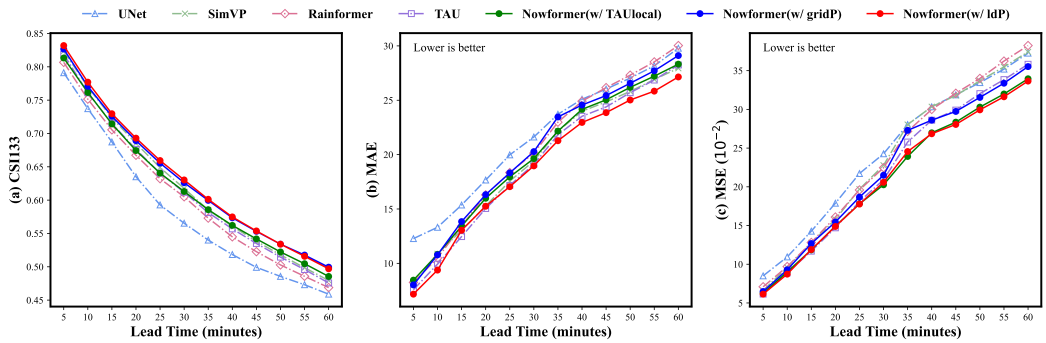

Figure 2: Performance over lead times.

indicate that it is important to observe locality simultaneously when considering the temporal relationships. Particularly, the performance gap is widening when predicting frames in the further future, indicating that the nowformer is stabler than other models. More detailed results, including visualization, are denoted in the appendix.

4 Application Context

Our methods are easy to apply wherever data exist since the proposed methods are data-based approaches that do not require complex physical formulas or dynamic models. Thus, our studies are applicable without domain expert knowledge or expensive computation. The performance can be further improved or adapted in the future by leveraging data that will continue to accumulate. The proposed methodology does not work only for specific data, although our methods used only VIL images from ground radar For instance, our methods can use data from satellites, various sensors, and weather stations. Thus, our methods work universally without relying only on specific data, suggesting the possibility that they could solve more varied tasks related to climate change, such as cyclone intensity prediction or hail storms [2, 23, 24]

5 Conclusion

We have proposed the nowformer, a locally enhanced temporal learner for precipitation nowcasting. We devise methods to provide an appropriate attention technique for precipitation data by using global dynamic attention and local dynamic attention. We experimentally demonstrate the superiority of the nowformer on the SEVIR dataset, one of the benchmark datasets for precipitation nowcasting, compared with other methods for video prediction and precipitation nowcasting. Since the nowformer is designed to consider the spatiotemporal pattern of precipitation features, we believe our methods are applicable to various data appearing in weather-related data, leading to preparing or predicting extreme climate changes, including precipitation predictions that we mainly target.

References

[1] Lorenzo Alfieri, Berny Bisselink, Francesco Dottori, Gustavo Naumann, Ad de Roo, Peter Salamon, Klaus Wyser, and Luc Feyen. Global projections of river flood risk in a warmer world. *Earth's Future*, 5(2):171–182, 2017.

[2] Mark Veillette, Siddharth Samsi, and Chris Mattioli. Sevir: A storm event imagery dataset for deep learning applications in radar and satellite meteorology. *Advances in Neural Information Processing Systems*, 33:22009–22019, 2020.

[3] Sylwester Klocek, Haiyu Dong, Matthew Dixon, Panashe Kanengoni, Najeeb Kazmi, Pete Luferenko, Zhongjian Lv, Shikhar Sharma, Jonathan Weyn, and Siqi Xiang. Ms-nowcasting: Operational precipitation nowcasting with convolutional lstms at microsoft weather. *arXiv preprint arXiv:2111.09954*, 2021.

[4] Gabriela Czibula, Andrei Mihai, Alexandra-Ioana Albu, Istvan-Gergely Czibula, Sorin Burcea, and Abdelkader Mezghani. Autonowp: An approach using deep autoencoders for precipitation nowcasting based on weather radar reflectivity prediction. *Mathematics*, 9(14):1653, 2021.

[5] Cong Bai, Feng Sun, Jinglin Zhang, Yi Song, and Shengyong Chen. Rainformer: Features extraction balanced network for radar-based precipitation nowcasting. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2022.

[6] Irina Malkin Ondík, Lukáš Ivica, Peter Šišan, Ivan Martynovskyi, David Šaur, and Ladislav Gaál. A concept of nowcasting of convective precipitation using an x-band radar for the territory of the zlín region (czech republic). In *Computer Science On-line Conference*, pages 499–514. Springer, 2022.

[7] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. *Advances in neural information processing systems*, 28, 2015.

[8] Kevin Trebing, Tomasz Sta czyk, and Siamak Mehrkanoon. Smaat-unet: Precipitation nowcasting using a small attention-unet architecture. *Pattern Recognition Letters*, 145:178–186, 2021.

[9] Chuyao Luo, Xinyue Zhao, Yuxi Sun, Xutao Li, and Yunming Ye. Predrann: The spatiotemporal attention convolution recurrent neural network for precipitation nowcasting. *Knowledge-Based Systems*, 239:107900, 2022.

[10] Jie Liu, Lei Xu, and Nengcheng Chen. A spatiotemporal deep learning model st-lstm-sa for hourly rainfall forecasting using radar echo images. *Journal of Hydrology*, 609:127748, 2022.

[11] Yimin Yang and Siamak Mehrkanoon. Aa-transunet: Attention augmented transunet for nowcasting tasks. *arXiv preprint arXiv:2202.04996*, 2022.

[12] Rachel Prudden, Samantha Adams, Dmitry Kangin, Niall Robinson, Suman Ravuri, Shakir Mohamed, and Alberto Arribas. A review of radar-based nowcasting of precipitation and applicable machine learning techniques. *arXiv preprint arXiv:2005.04988*, 2020.

[13] Cheng Tan, Zhangyang Gao, Siyuan Li, Yongjie Xu, and Stan Z Li. Temporal attention unit: Towards efficient spatiotemporal predictive learning. *arXiv preprint arXiv:2206.12126*, 2022.

[14] Zhangyang Gao, Cheng Tan, Lirong Wu, and Stan Z Li. Simvp: Simpler yet better video prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3170–3180, 2022.

[15] Sergiu Oprea, Pablo Martinez-Gonzalez, Alberto Garcia-Garcia, John Alejandro Castro-Vargas, Sergio Orts-Escolano, Jose Garcia-Rodriguez, and Antonis Argyros. A review on deep learning techniques for video prediction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[16] Arunkumar Byravan and Dieter Fox. Se3-nets: Learning rigid body motion using deep neural networks. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 173–180. IEEE, 2017.

[17] Yue Wu, Rongrong Gao, Jaesik Park, and Qifeng Chen. Future video synthesis with object motion prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5539–5548, 2020.

[18] Robert G Fovell and Pei-Hua Tan. The temporal behavior of numerically simulated multicell-type storms. part ii: The convective cell life cycle and cell regeneration. *Monthly Weather Review*, 126(3):551–577, 1998.

[19] Wenhai Wang, Enze Xie, Xiang Li, Deng-Ping Fan, Kaitao Song, Ding Liang, Tong Lu, Ping Luo, and Ling Shao. Pvt v2: Improved baselines with pyramid vision transformer. *Computational Visual Media*, 8(3):415–424, 2022.

[20] Yawei Li, Kai Zhang, Jiezhang Cao, Radu Timofte, and Luc Van Gool. Localvit: Bringing locality to vision transformers. *arXiv preprint arXiv:2104.05707*, 2021.

[21] Meng-Hao Guo, Cheng-Ze Lu, Zheng-Ning Liu, Ming-Ming Cheng, and Shi-Min Hu. Visual attention network. *arXiv preprint arXiv:2202.09741*, 2022.

[22] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[23] Nicholas WS Demetriades and RL Holle. Long range lightning nowcasting applications for tropical cyclones. In *Proceedings of 2nd Conference on the Meteorology Application of Lightning Data, Atlanta, USA*, volume 29, 2006.

[24] Jinkai Tan, Qidong Yang, Junjun Hu, Qiqiao Huang, and Sheng Chen. Tropical cyclone intensity estimation using himawari-8 satellite cloud products and deep learning. *Remote Sensing*, 14(4):812, 2022.

A Appendix

A.1 Implementation details

We use an AdamW optimizer with momentum terms of (0.9,0.999), 0.001 as the initial learning rate, and 0.0001 as the weight decay. All models are trained with a batch size of 12 precipitation sequences for 30 epochs on a single NVIDIA-A100 GPU.

A.2 Illustration of three local dynamic strategies

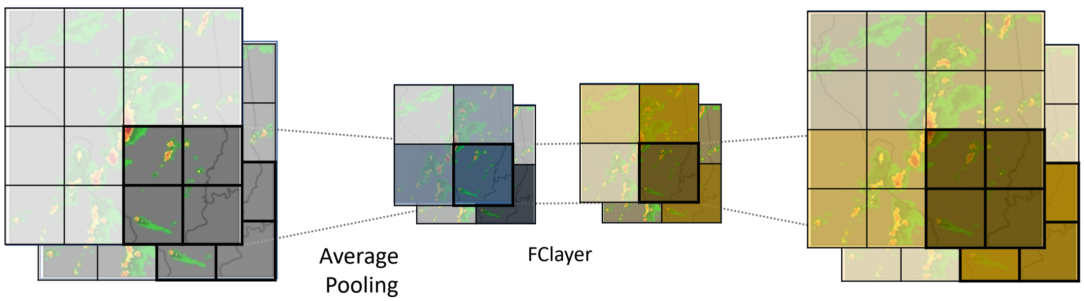

Figure 3: Illustration of TAUlocal.

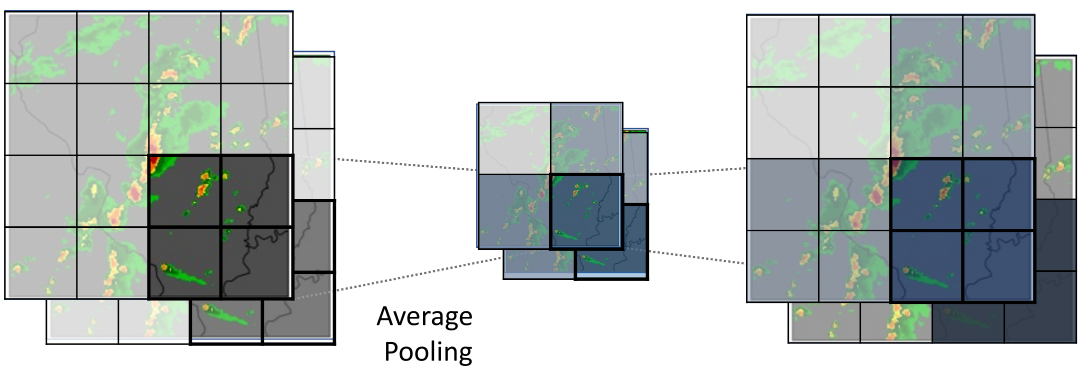

Figure 4: Illustration of gridP.

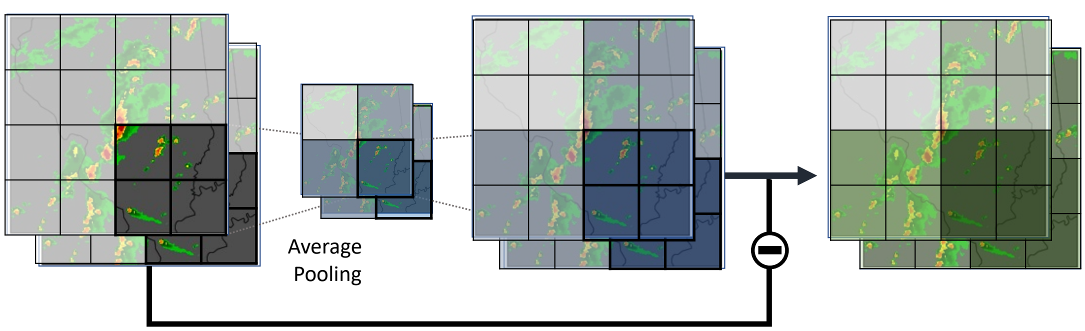

Figure 5: Illustration of ldP.

A.3 Visualization

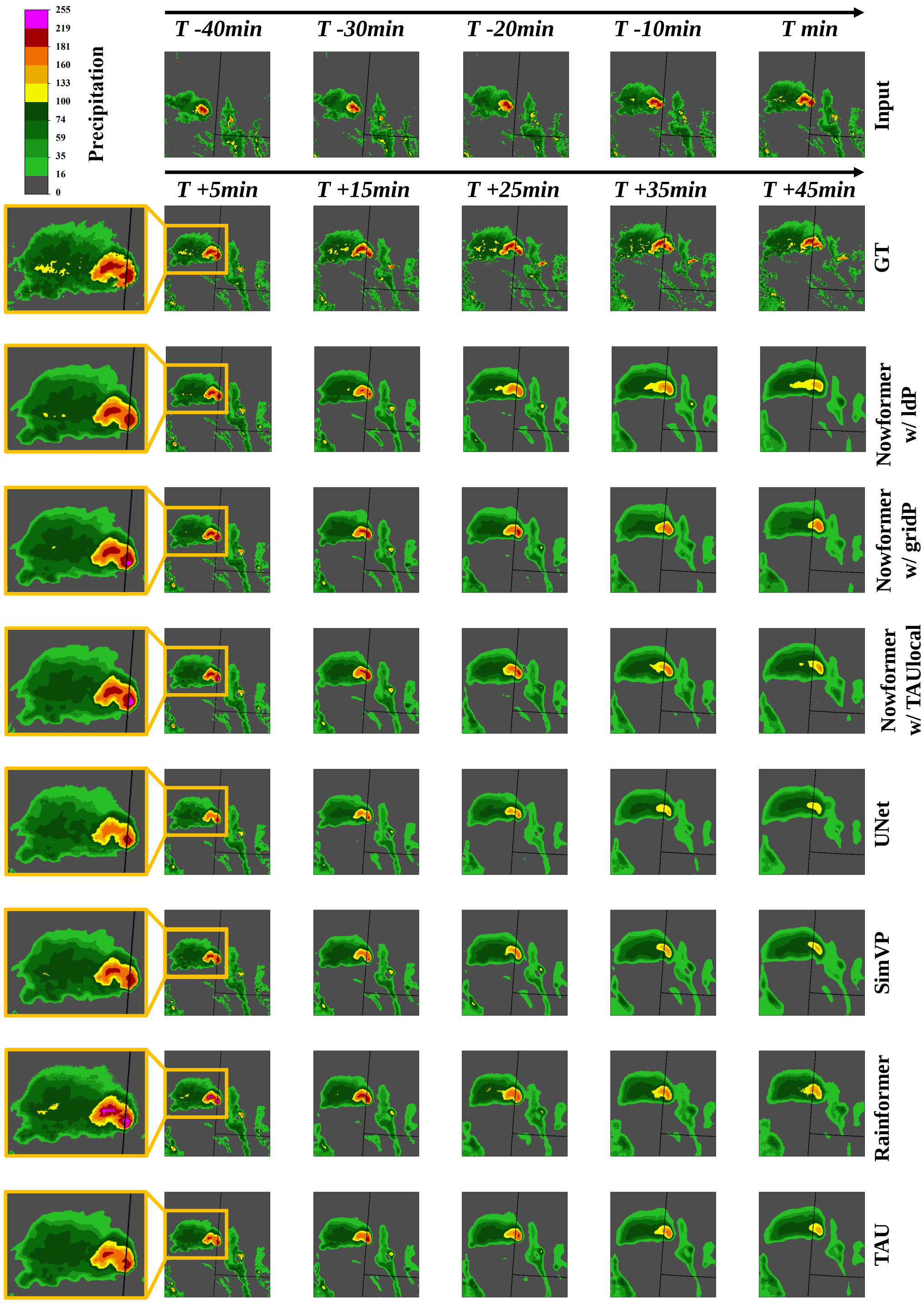

Figure 6: Comparison of nowcasting results with other methods