

A Spatial-temporal Deep Probabilistic Diffusion Model for Reliable Hail Nowcasting with Radar Echo Extrapolation

Haonan Shi †

School of Computer Science and Technology, Xidian University
Chang'an Qu, Xi'an Shi, China

Long Tian †

School of Computer Science and Technology, Xidian University
Chang'an Qu, Xi'an Shi, China
tianlong@xidian.edu.cn

Jie Tao

Hangzhou Institute of Technology, Xidian University
Xiao Shan Qu, Hang Zhou Shi, China

Yufei Li

School of Computer Science and Technology, Xidian University
Chang'an Qu, Xi'an Shi, China

Liming Wang ✉

School of Computer Science and Technology, Xidian University
Chang'an Qu, Xi'an Shi, China
wanglm@mail.xidian.edu.cn

Xiyang Liu

School of Computer Science and Technology, Xidian University
Chang'an Qu, Xi'an Shi, China

ABSTRACT

Hail nowcasting is a considerable contributor to meteorological disasters and there is a great need to mitigate its socioeconomic effects through precise forecast that has high resolution, long lead times and local details with large landscapes. Existing medium-range weather forecasting methods primarily rely on changes in upper air currents and cloud layers to predict precipitation events, such as heavy rainfall, which are unsuitable for hail nowcasting since it is mainly caused by low-altitude local strong convection associated with terrains. Additionally, radar captures the status of low cloud layers, such as water vapor, droplets, and ice crystals, providing rich signals suitable for hail nowcasting. To this end, we introduce a Spatial-Temporal gEnerAtive Model called SteamCast for hail nowcasting with radar echo extrapolation, it is a deep probabilistic diffusion model based on spatial-temporal representations including radar echoes as well as their position/time embeddings, which we trained on historical reanalysis archive from Yan'an Meteorological Bureau in China, where the crop yield like apple suffers greatly from hail damage. Considering the short-term nature of hail, SteamCast provides 30-minute nowcasts at 6-minute intervals for a single radar reflectivity variable, across 9 different vertical angles, on a latitude-longitude grid with approximately 1 km × 1 km resolution per pixel in Yan'an City, China. By successfully fusing the spatial-temporal features of radar echoes, SteamCast delivers competitive, and in some cases superior, results compared to other deep learning-based models such as PredRNN and VMRNN.

CCS CONCEPTS

• **Computing methodologies** → *Spatial and physical reasoning.*

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

WebST-25, April 28–29, 2025, Sydney, Australia.

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
<https://doi.org/XXXXXXXX.XXXXXXX>

KEYWORDS

Radar Extrapolation, Hail Nowcasting, Deep Probabilistic Diffusion Model, Spatial-temporal Feature Representation

ACM Reference Format:

Haonan Shi †, Long Tian †, Jie Tao, Yufei Li, Liming Wang ✉, and Xiyang Liu . 2025. A Spatial-temporal Deep Probabilistic Diffusion Model for Reliable Hail Nowcasting with Radar Echo Extrapolation. In *Proceedings of The 1st International Workshop on Spatio-Temporal Data Mining from the Web (WebST-25)*. ACM, New York, NY, USA, 5 pages. <https://doi.org/XXXXXXXX.XXXXXXX>

1 INTRODUCTION

Hail nowcasting is the task of predicting the occurrence, intensity, and location of hail events. This involves analyzing a variety of meteorological data, such as temperature profiles and radar reflectivity, to identify atmospheric conditions favorable for hail formation. The goal is to provide timely warnings to the public and relevant authorities so that they can take necessary precautions to mitigate the impacts of hail on agriculture, infrastructure, and public safety. Accurate hail forecasting is challenging due to the complex and rapidly changing nature of the atmospheric conditions that lead to hail, and it often necessitates advanced extrapolation techniques, encompassing numerical weather prediction systems [1, 2] and data-driven machine learning algorithms [3, 4].

Ensemble numerical weather prediction (NWP) systems often struggle to deliver accurate precipitation forecasts with a lead time of zero to two hours, primarily due to challenges in assimilating non-Gaussian data [5]. As a result, alternative methods using radar data, such as PySTEPS [2] and DARTS [1], have been developed based on advection schemes, which are fundamentally rooted in the continuity equation. However, current implementations of advection schemes fail to incorporate nonlinear evolution processes, leading to reduced accuracy in precipitation nowcasts, such as heavy rain reported in [4] and hailstorms analyzed in our study.

Deep-learning methods have been applied to precipitation nowcasting in recent years [6–9]. These methods aim to better capture traditionally challenging non-linear precipitation phenomena, such as convective initialization and extreme precipitation, by leveraging large datasets of radar echoes to train neural networks, instead

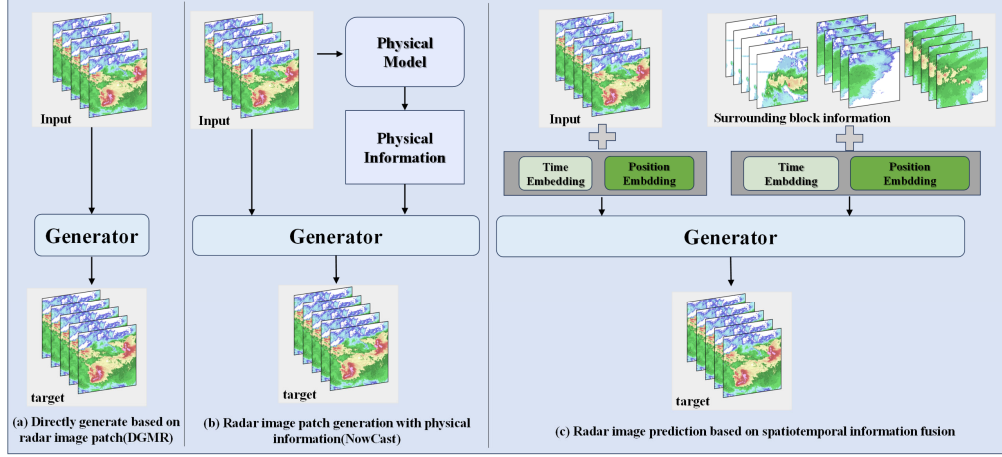


Figure 1: Comparison of our SteamCast in (c) with the DGM-based method in (a) and the physics-informed method in (b). Enhancing the capability of capturing spatiotemporal features is one of the most important aspects of reliable precipitation nowcasting. Methods (b) and (c) utilize a physical model and a well-designed spatiotemporal DGM to enhance the capability of capturing spatiotemporal features.

of relying on physical assumptions. This end-to-end data-driven optimization effectively reduces inductive biases. Consequently, the nowcast quality of these methods improves, as reflected in metrics like the per-grid-cell Critical Success Index (CSI). However, these methods encounter increasing uncertainty at longer lead times, resulting in blurrier precipitation fields, which are critical for accurate nowcasts of small-scale radar echo patterns. We speculate that poor spatial-temporal representations limit the models’ operational utility, preventing them from simultaneously providing consistent predictions across multiple spatial and temporal scales. A large step forward in enhancing spatial and temporal representations has been the Deep Generative Model (DGM) of Radar called DGMR [3] developed by DeepMind and the UK Met Office. DGMR is also the first work employing DGM for precipitation nowcasting. Long et al. [4] propose NowcastNet, a reliable nowcasting framework that aggregates spatial-temporal features by integrating DGM with physical knowledge of precipitation processes, including the conservation law of cloud transport [1] and the rain rate distribution [10]. In contrast to physically informed spatial-temporal aggregation, we propose an alternative approach by extracting spatial-temporal features of radar echoes using a well-designed Spatial-Temporal gEnerAtive Model called SteamCast.

We adopt a stable diffusion architecture [11] with slight yet critical modifications to represent effective and flexible spatiotemporal radar echoes. We employ a lightweight condition encoder to extract spatiotemporal features from a target patch and its neighboring regions in eight directions for hail nowcasting. A stable diffusion U-Net layer then merges the randomly initialized patch with these features to generate 30-minute nowcasting results. Leveraging cross-attention, our model effectively captures radar echoes and adapts to forecasting tasks with flexible input (N) and output (M) lengths. Additionally, to enhance efficiency under limited computing resources, we process radar images in smaller sub-problems instead of full $H \times W$ resolution, seamlessly aggregating the results. In summary, SteamCast introduces a stable diffusion-based architecture with an efficient spatial-temporal feature extraction

mechanism, enabling flexible and high-resolution hail nowcasting. By improving forecasting adaptability, reducing computational overhead, and enhancing multi-scale consistency, SteamCast provides a practical and scalable solution for operational precipitation nowcasting.

2 RELATED WORK AND KNOWLEDGE

2.1 Related Work

One of the earliest deep learning-based methods is the ConvLSTM Network for precipitation nowcasting [9], in which the authors framed nowcasting as a spatiotemporal sequence-to-sequence problem and proposed a cascaded Convolutional LSTM for its solution. Similarly, PredRNN [6] introduced a spatiotemporal memory flow mechanism for complex pattern nowcasting. More recently, DGM-based methods have gained significant attention for addressing this challenge task [4, 8], as previously discussed. In this work, we focus on hail nowcasting, which represents a novel and significant task. Unlike heavy rainfall, hail formation is strongly associated with terrains, making precise modeling of spatiotemporal correlations crucial for hail nowcasting, a topic that has been rarely studied.

2.2 Diffusion Models

Diffusion models are a class of generative models that generate new data by iteratively denoising a noisy sample drawn from a simple prior distribution. The generation process is formulated as a reverse Markov chain, where each step progressively removes noise to recover the original data distribution. These models have achieved state-of-the-art performance in various generative tasks, including image, text, and audio synthesis.

3 METHODOLOGY

3.1 Task Definition

We consider reference patches of radar echoes denoted by $\{x_i^r\}_{i=1}^T$, where $x_i^r \in \mathbb{R}^{T \times H \times W \times N}$. Here T , H , W , and N separately denote the number of reference patch, patch height, patch width, and

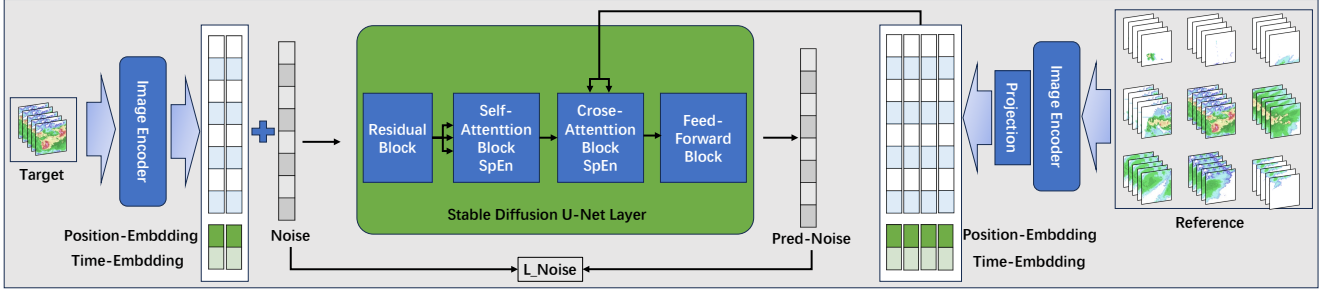


Figure 2: SteamCast architecture details. SteamCast adopts a SD architecture with slight yet critical modifications to effectively and flexibly represent spatiotemporal radar echoes. Left: The pre-trained encoder in SD captures features from the target patch of radar echoes, the target patch consists of M nowcasted time steps. Middle: In U-Net, we apply self-attention within the target patch to encourage target-to-target consistency. We also apply a cross-attention block between the target patch and T reference patches to enhance reference-to-target consistency. In each attention block, SpEn is employed for the key and query, enabling the attention map to capture relative positions and time steps. Right: The pre-trained encoder in SD and projection layer capture features from T reference patches, each patch contains N historical time steps.

historical time steps. Similarly, the target patch is denoted by $x_j^t \in \mathbb{R}^{1 \times H \times W \times M}$ corresponding to the j -th reference patch. Here $j = \{1, \dots, T\}$ and M is the number of nowcasted time steps. Our goal is to train a nowcasting model $\hat{x}_j^t = f_{\theta}(\{x_i^r\}_{i=1}^T)$ such that $\|x_j^t - \hat{x}_j^t\| \leq \epsilon$, where ϵ is a small positive value. We randomly select a target patch from the reference patch set to train the model.

3.2 SteamCast Architecture

The architecture details of SteamCast, as shown in Fig. 2, are designed based on two key principles: 1) It leverages existing image-based diffusion models to inherit their powerful generalization capabilities. 2) It encodes spatiotemporal information for each tokenized input patch, enabling the model to process an arbitrary number of patches and time steps for flexible nowcasting.

Target Patch Generation : SteamCast is designed by adopting the architecture of the Stable Diffusion v1.5 (SD) [11]. To adapt SD, which was originally designed for text-to-image generation, for hail nowcasting using radar extrapolation, several key modifications are implemented. First, we reimplement the self-attention block in the original SD to calculate interactions within the target patch across M different nowcasted time steps, thereby enhancing target-to-target consistency. For the cross-attention block in the original SD, we modify it to calculate interactions from T reference patches to the target patch, ensuring reference-to-target consistency. The features extracted by the SD encoder are $h_j^t \in \mathbb{R}^{b \times d}$, where $b = (h \times w) \times M$, with $h \times w$ is the total number of tokenized features.

Reference Patches Condition: In hail nowcasting, it is crucial that the conditioning features accurately capture the texture details present in the reference patches, which has often been overlooked in previous studies. In SteamCast, we treat reference patches as sets of tokens and use an image encoder to extract such conditioning signals. This approach enables us to maintain flexibility in handling a variable number of patches and time steps. We choose to compress the reference patches to smaller resolution features with ConvNeXtv2-Tiny [12], a lightweight and highly efficient CNN architecture. The features of the reference patches after CNN are expressed as $h_i^r \in \mathbb{R}^{b' \times d}$ for $i = 1, \dots, T$, where $b' = (h' \times w') \times N$, with $h' \times w'$ the total number of tokenized features. The features

after the self-/cross-attention block with SpEn can be written as $\hat{h}_j^t, \hat{h}_j^t \in \mathbb{R}^{b \times d}$. Details on SpEn are provided in Sec. 3.3.

Model Training and Inference: We use the original SD objective, which is to reconstruct the error, to fine-tune the parameters of SteamCast, which is initialized with pre-trained SD v1.5 and ConvNeXtv2-Tiny. During inference, SteamCast takes a set of reference patches with N historical time steps and a randomly initialized target patch with M nowcasted time steps. Then it provides users with the corresponding nowcasting for the target patch of interest.

3.3 Spatiotemporal Encoding (SpEn)

To efficiently encode spatiotemporal information into the reference and target patches in the U-Net layer of SD, we propose SpEn, inspired by advancements in the language domain. For illustration, consider the i -th reference patch at the n -th time step, denoted as $h_i^r(:, n)$. Suppose there are $4 \times 4 = 16$ reference patches and $5 + 5 = 10$ time steps, with 5 for historical time steps and 5 for nowcasting. Time and position indexes can be separately represented by 4-D binary vectors ($\lceil \log_2 K \rceil$). We employ the standard Sin/Cos embedding scheme used in Transformers [13] to encode each index (time or position), resulting in an encoded vector $\rho \in \mathbb{R}^{1 \times 8}$. The final spatiotemporal embedding is obtained by repeating ρ to match the length of $h_i^r(:, n)$ along the last dimension. Hence, the tokenized feature after SpEn can be expressed as $\hat{h}_i^r(:, n) = \rho \odot h_i^r(:, n)$, where \odot represents element-wise multiplication.

4 EXPERIMENTS

4.1 Experimental Setups

Dataset: We use the 2-year historical reanalysis archive of radar reflectivity measurements with a resolution of $1024 \times 1024 \times 3$ from Yan'an Meteorological Bureau in China to validate the performance of SteamCast. The interval between two adjacent measurements is 6-minute. We collect a total of 10,000 radar reflectivity patches using a sliding window with size of $10 \times 256 \times 256 \times 3$ without overlapping across the entire historical dataset. Then, we randomly split the entire set of patches into 6,400, 1,600, and 2,000 patches for training, validation, and testing, respectively.

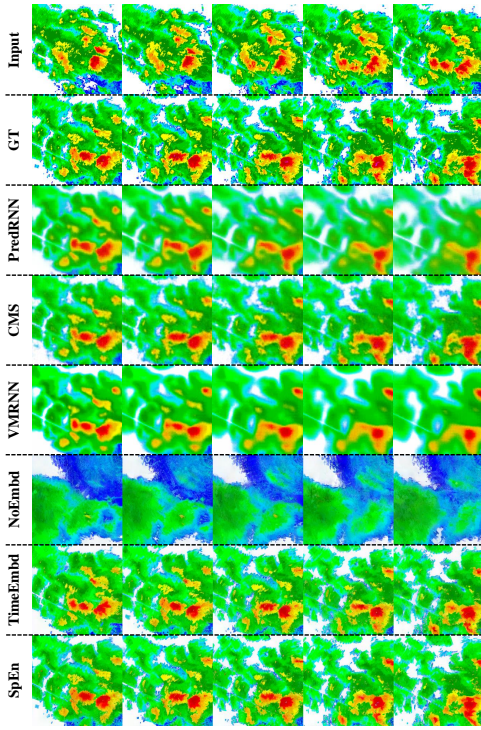


Figure 3: Visualization results of hail nowcasting. Input and GT denote target patches of 5 historical and 5 future time steps, respectively. PredRNN, CMS, and VMRNN represent nowcasting results from the respective competitors. SpEn represents the hail nowcast of SteamCast. NoEmb and TimeEmb represent hail nowcasts from our two variants, as discussed in Sec. 4.3 of the ablation studies.

Competing Methods: We compare our SteamCast with some of the most recently proposed spatiotemporal prediction methods, including: 1) PredRNN [6]: A spatiotemporal recurrent network. 2) CMS [14]: A memory-based convolutional network. 3) VMRNN [15]: A memory-based recurrent network with variational autoencoding. **Evaluation Protocols:** We use the following metrics to evaluate performance, including: 1) MSE (Mean Squared Error): Pixel-wise difference between the ground truth and predicted target patches. 2) PSNR (Peak Signal-to-Noise Ratio): Reflects the quality of the predicted target patches, with higher values indicating better quality. 3) SSIM (Structural Similarity Index): Assesses the structural similarity between the ground truth and predicted target patches, considering luminance, contrast, and texture. 4) ETS (Equitable Threat Score): Measures the prediction accuracy for rare event predictions. 5) ACC (Accuracy): Measures the overall prediction accuracy.

Implementation Details: We use the encoder and decoder of SD v1.5 [11] to separately extract features of the target patch and generate the nowcasted target patch during training. We use ConvNeXtv2-Tiny [12] to extract features of the set of reference patches. The U-Net layer is a modified version from SD v1.5. We use the pre-trained parameters of SD v1.5 and ConvNeXtv2-Tiny to initialize our SteamCast. Our model is trained with the follows hyper-parameters: batch size 4, learning rate 1×10^{-4} , AdamW optimizer [16], training steps 100,000. Additionally, we use mixed

Table 1: Nowcasting performance on various metrics. The best results are in bold.

| Model | MSE ↓ | PSNR ↑ | SSIM ↑ | ETS ↑ | ACC ↑ |
|--------------------|-------------|--------------|-------------|-------------|-------------|
| PredRNN (TPAMI’22) | 0.05 | 13.86 | 0.62 | 0.08 | 0.96 |
| CMS (ICME’22) | 0.06 | 13.09 | 0.60 | 0.07 | 0.96 |
| VMRNN (CVPR’24) | 0.03 | 17.02 | 0.79 | 0.10 | 0.97 |
| SteamCast (Ours) | 0.02 | 23.15 | 0.81 | 0.18 | 0.99 |

Table 2: Ablation studies on spatial and temporal embeddings with various metrics. The best results are in bold.

| Position | Time | MSE ↓ | PSNR ↑ | SSIM ↑ | ETS ↑ | ACC ↑ |
|----------|------|-------------|--------------|-------------|-------------|-------------|
| - | - | 0.40 | 10.00 | 0.43 | 0.01 | 0.97 |
| - | ✓ | 0.05 | 23.21 | 0.74 | 0.16 | 0.98 |
| ✓ | ✓ | 0.02 | 23.15 | 0.81 | 0.18 | 0.99 |

precision training for acceleration. The experiments are conducted with two NVIDIA GTX 4090 GPUs.

4.2 Enumerated Results

Quantitative results: We compare SteamCast with recent state-of-the-art methods, and the results are presented in Table 1. All competing models are designed for spatiotemporal prediction, aiming to nowcast the target patch for the next 30 minutes based on historical observations. SteamCast consistently outperforms other methods across various metrics, demonstrating its effectiveness in hail nowcasting. Given the rarity of hail events, ACC is not a reliable metric for evaluating predictive improvements. Instead, the Equitable Threat Score (ETS), which accounts for both true positives and random chance, provides a more meaningful assessment. SteamCast achieves a notable ETS increase from 0.10 to 0.18, demonstrating a substantial enhancement in hail detection capability.

Qualitative results: We report visualized nowcasting results of a hail event in Fig. 3. As we can see, our model achieves promising performance compared with others not only in correct trends but also in richer details.

4.3 Ablative Analysis

We emphasize the importance of properly aggregating spatiotemporal information of the target patch. To achieve this, we designed two additional variants of SteamCast: 1) NoEmb: without time and position embeddings. 2) TimeEmb: employing only time embedding. According to the results in Table 2, we find that leveraging time and position embeddings contributes significantly to improved performance. Additionally, qualitative results in Fig. 3 further support, in a more intuitive way, that SpEn is indispensable.

5 CONCLUSION

Our study demonstrates that SteamCast, a spatial-temporal deep probabilistic diffusion model, enhances the accuracy and adaptability of hail nowcasting. Its potential applications span precision agriculture, disaster prevention, and urban resilience planning, where timely and accurate hail forecasts can help mitigate economic losses and societal disruptions. Future work will focus on integrating multi-source meteorological data and enhancing long-term forecasting capabilities to further improve predictive performance.

ACKNOWLEDGMENTS

This work was supported by the Foundation of Aerospace SAST2021-012.

REFERENCES

- [1] Evan Ruzanski, V Chandrasekar, and Yanting Wang. 2011. The CASA nowcasting system. *Journal of Atmospheric and Oceanic Technology* 28, 5 (2011), 640–655.
- [2] Seppo Pulkkinen, Daniele Nerini, Andrés A Pérez Hortal, Carlos Velasco-Forero, Alan Seed, Urs Germann, and Loris Foresti. 2019. Pysteps: An open-source Python library for probabilistic precipitation nowcasting (v1. 0). *Geoscientific Model Development* 12, 10 (2019), 4185–4219.
- [3] Suman Ravuri, Karel Lenc, Matthew Willson, Dmitry Kangin, Remi Lam, Piotr Mirowski, Megan Fitzsimons, Maria Athanassiadou, Sheleem Kashem, Sam Madge, et al. 2021. Skillful precipitation nowcasting using deep generative models of radar. *Nature* 597, 7878 (2021), 672–677.
- [4] Yuchen Zhang, Mingsheng Long, Kaiyuan Chen, Lanxiang Xing, Ronghua Jin, Michael I Jordan, and Jianmin Wang. 2023. Skillful nowcasting of extreme precipitation with NowcastNet. *Nature* 619, 7970 (2023), 526–532.
- [5] Mark Buehner and Dominik Jacques. 2020. Non-Gaussian deterministic assimilation of radar-derived precipitation accumulations. *Monthly Weather Review* 148, 2 (2020), 783–808.
- [6] Yunbo Wang, Haixu Wu, Jianjin Zhang, Zhifeng Gao, Jianmin Wang, S Yu Philip, and Mingsheng Long. 2022. Predrnn: A recurrent neural network for spatiotemporal predictive learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 2 (2022), 2208–2225.
- [7] Lasse Espeholt, Shreya Agrawal, Casper Sønderby, Manoj Kumar, Jonathan Heek, Carla Bromberg, Cenk Gazen, Rob Carver, Marcin Andrychowicz, Jason Hickey, Aaron Bell, and Nal Kalchbrenner. 2022. Deep learning for twelve hour precipitation forecasts. *Nature Communications* 13, 1 (2022). <https://doi.org/10.1038/s41467-022-32483-x>
- [8] Rui Wang, Lin Su, Wai Kin Wong, Alexis K. H. Lau, and Jimmy C. H. Fung. 2023. Skillful Radar-Based Heavy Rainfall Nowcasting Using Task-Segmented Generative Adversarial Network. *IEEE Transactions on Geoscience and Remote Sensing* 61 (2023). <https://doi.org/10.1109/TGRS.2023.3295211>
- [9] Xingjian Shi, Hourong Chen, Hao Wang, Dit-Yan Yeung, Wai-kin Wong, and Wang-chun Woo. 2015. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. In *Advances in Neural Information Processing Systems*, Vol. 28. 802–810.
- [10] R.K. Crane. 1990. Space–time structure of rain rate fields. *Journal of Geophysical Research* 95, D12 (1990), 20455–20462.
- [11] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis with Latent Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10684–10695.
- [12] Sanghyun Woo, Shoubhik Debnath, Ronghang Hu, Xinlei Chen, Zhuang Liu, In So Kweon, and Saining Xie. 2023. ConvNeXt V2: Co-designing and Scaling ConvNets with Masked Autoencoders. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 16000–16009. https://openaccess.thecvf.com/content/CVPR2023/papers/Woo_ConvNeXt_V2_Co-Designing_and_Scaling_ConvNets_With_Masked_Autoencoders_CVPR_2023_paper.pdf
- [13] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need. In *Advances in Neural Information Processing Systems (NeurIPS)*. 5998–6008.
- [14] Zenghao Chai, Zhengzhuo Xu, Yunpeng Bail, Zhihui Lin, and Chun Yuan. 2022. CMS-LSTM: Context embedding and multi-scale spatiotemporal expression LSTM for predictive learning. In *2022 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 01–06.
- [15] Yujin Tang, Peijie Dong, Zhenheng Tang, Xiaowen Chu, and Junwei Liang. 2024. Vmrrn: Integrating vision mamba and lstm for efficient and accurate spatiotemporal forecasting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5663–5673.
- [16] Ilya Loshchilov and Frank Hutter. 2019. Decoupled Weight Decay Regularization. In *International Conference on Learning Representations (ICLR)*. <https://arxiv.org/abs/1711.05101>