

基于双水平路由注意力机制的时序Transformer临近预报研究

卢 姁¹, 王也英¹, 李一玲¹, 董先飞¹, 初奕琦², 王 曦², 孙国栋²

¹中国人民解放军32021部队, 北京

²北京无线电测量研究所, 北京

收稿日期: 2024年10月25日; 录用日期: 2024年11月22日; 发布日期: 2024年11月29日

摘 要

0~2小时的临近预报是预报强对流天气的重要手段, 准确的临近预报是气象防灾的重要屏障。临近预报是气象预报的一个重要业务。本文提出了一个基于时序Transformer的全局特征提取的Encoder-Forecasting临近预报模型结构。该模型的Transformer利用双级路由注意力机制定位特征中的几个最相关的键值对来提高计算效率。模型同时采用MaxPool和AvgPool提取目标对象的局部特征, 并将局部特征和全局特征进行有效融合, 进而充分提取目标对象的特征用于临近预报。提出的模型与几个模型在一个公共的临近预报数据集上进行实验验证。实验结果充分证明了我们提出的模型的有效性和准确性。

关键词

临近预报, 深度学习, 双水平路由注意力, 视觉Transformer

A Study of Temporal Transformer Nowcast Based on Bilevel Routing Attention Mechanism

Xu Lu¹, Yeying Wang¹, Yiling Li¹, Xianfei Dong¹, Yiqi Chu², Xi Wang², Guodong Sun²

¹The People's Liberation Army Unit 32021, Beijing

²Beijing Institute of Radio Measurement, Beijing

Received: Oct. 25th, 2024; accepted: Nov. 22nd, 2024; published: Nov. 29th, 2024

Abstract

The nowcast of 0~2 hour is an important toolth for forecasting severe convective weather. Accurate nowcast is an important barrier to prevent from weather disaster. Therefore, nowcast is a very

文章引用: 卢姁, 王也英, 李一玲, 董先飞, 初奕琦, 王曦, 孙国栋. 基于双水平路由注意力机制的时序Transformer临近预报研究[J]. 计算机科学与应用, 2024, 14(11): 208-217. DOI: 10.12677/csa.2024.1411230

important field in the meteorological society. This paper proposes a nowcast model which based on Encoder-Forecasting architecture and temporal Transformer. The transformer utilizes bilevel routing attention to improve the computational efficiency. At the same time, the proposed model also use the operations of Maxpool and AvgPool to extract the local features. The global features and local features are effectively fused to represent the objects' features. The comprehensive experiments based on the proposed model and four state-of-art models are conducted on a public datasets. The experimental results effectively show the correctness and effectiveness of the proposed model.

Keywords

Nowcast, Deep Learning, Bilevel Routing Attention, Vision Transformer

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

强对流天气由于破坏性强, 给人们的生命财产安全造成比较大的损失。同时由于它形成的突发性强, 持续发生时间比较短, 并且影响范围相对比较小等特点, 因此对强对流天气的预报变得比较困难[1]。强对流天气是中小尺度天气系统的一部分。准确的预报强对流天气的发生又是气象防灾减灾及气象预警工作的重要步骤。气象雷达主要用于大气探测, 它是探测中小尺度天气的重要工具, 雷达回波能够比较准确的反映实际天气状况。因此准确的外推雷达回波是预报强对流天气的重要手段。0~2 小时的雷达回波外推得到气象雷达探测目标未来 2 小时的气象信息, 从而能够比较准确的预测未来天气的情况。

2. 相关工作

2.1. 传统外推方法

传统的临近预报技术主要包括雷达回波外推[2]和数值预报[3]。然而在 0~2 小时的临近预报应用中, 数值预报由于模拟复杂的天气系统需要较多的计算资源及太多的假设初始条件而得到的结果次于雷达回波外推得到的结果[4]。因此对临近预报的研究主要集中于雷达回波外推。传统的雷达回波外推技术主要采用交叉相关法[5]和光流法[6]。但由于雷达回波运动的轨迹受多种气象要素的约束变化无常, 很难用数学公式进行描述表示。传统的光流法和交叉相关法主要利用相邻的几张雷达回波进行回波轨迹的预测, 由于使用的回波数量少, 很难实现对回波运动轨迹规律的描述, 同时随着外推时间的变长, 对雷达回波的生消演变结果的准确度急剧降低, 因此外推效果需要进行极大的改进以满足人们对气象预报准确性的要求。

2.2. 基于深度学习的临近预报

近年来, 随着图形处理器(Graphic Process Unit, GPU)计算力的快速增长及各种大数据的出现, 深度学习[7]的兴起及其在许多领域得到了广泛的应用并取得了重要的成功。Shi 等人基于传统的长短期记忆网络(LSTM)并结合雷达回波的空间特征属性提出了卷积长短期记忆网络(Convolutional LSTM, ConvLSTM)[8]用于雷达回波外推并取得了比传统光流法更好的外推效果。针对卷积操作的局部不变性特点和雷达回波运动是局部变化的两种状况, Shi 等人又改进提出了 TrajGRU [9]用于主动学习回波序列中的局部变化结构用于临近预报。Wang 等人提出的 PredRNN (predictive recurrent neural network) [10]模型使用一个统

一的存储池学习预测目标的空间外形和时序变化特征。Trebing 等人[11]基于 UNet 框架使用深度可分离卷积减少模型的参数,同时在编码器中加入 CBAM (Convolutional Block Attention Module)注意力机制, SmartAt-UNet 在简化网络模型参数、提高模型推理的情况下仍然得到了较好的预测效果。Wang 等人在论文[12]中通过时空记忆流的锯齿状循环转换机制和一对解耦记忆单元的新卷积循环单元改进了 PredRNN。DeepMind 团队[13]基于生成对抗网络开发了一套降水概率预报模型。

以上这些预报模型都是基于卷积操作进行临近预报,然而受限于卷积操作固有的平均属性,这些模型在预测高强度回波值时的准确度就比较低。同时卷积操作的效果受其卷积核的限制,它只能提取目标对象的局部特征。然而随着 Transformer 在自然语言[14]、图像处理[15]等多种应用中取得了令人瞩目的成功,与传统的卷积操作相对比,Transformer 在提取目标对象的全局及长时序特征方面表现更加出色。基于 Transformer 的深度学习模型也逐渐应用于临近预报中。论文[16]提出了基于窗口的多头自注意力机制的 Rainformer 用于提升高强度降水预报的准确性。Li 等人[17]提出了一种采用 LPT-QPN 模型用于临近降水预报。LPT-QPN 模型采用新颖的 Transformer 模块来学习提取降水的长期演变规律特征,它同时采用多头平方注意力(multihead squared attention, MHSA)构建降水的非线性关系,使用多头平方注意力同时能够减少 LPT-QPN 模型的计算复杂度。实验验证表明 LPT-QPN 模型能够提高对某些特定降水时间的预测精度。为了减少临近预报外推中的累计误差,一个时序空间并行的 Transformer [18]被提出用于临近预报并取得了很好的效果。

虽然深度学习在临近预报应用中取得了重要的成功,但其在临近预报中的研究也处于发展阶段。目前研究主要是利用大量的回波数据让深度学习模型进行自动学习其运行规律,同时深度学习模型也在快速的发展中,本文根据目前发展的趋势及深度学习最新技术,提出并改进了一种基于 Transformer 的深度学习模型并将其用于临近预报研究,努力尝试提高模型在预报高强度回波的预报精度。

3. 研究方法

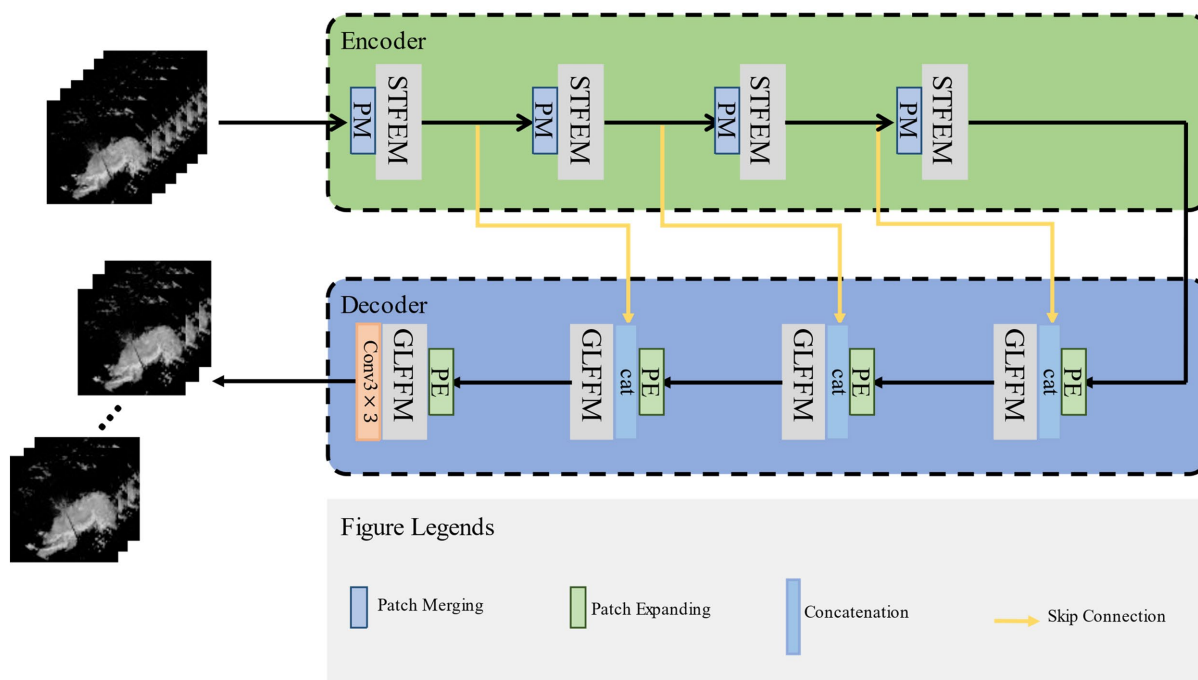


Figure 1. Network architecture of the proposed nowcast model

图 1. 提出的临近预报模型网络结构图

本文中,我们充分考虑雷达回波序列的时序特征和空间特征及 Vision Transformer 的全局特征提取属性,采用 Encoder-Forecaster 架构用于临近预报。整个模型架构如图 1 所示。Encoder 和 Forecaster 都是由四个模块组成,Encoder 中的每个模块最开始执行 PatchMerging 操作将目标对象的尺寸降低 1/2,Forecaster 中的四个模块开始则执行 PatchExpanding 操作将目标对象的尺寸放大 2 倍,这样最终恢复目标到原始的尺寸大小。其中 Encoder 和 Forecaster 模型的组成构件都是一样的,但它们处理回波数据的方式不同。Encoder 中处理输入数据包含了回波序列的时序性和空间特性,因此我们将其称为时序空间 Encoder,提取回波空间时序特征的模块称为空间时序特征提取模块(Spatialtemporal Feature Extract Module, STFEM)。Forecaster 处理回波的空间特性,我们将其称其为空间 Forecaster。Forecaster 的 4 个模块中都分别包含 Transformer 组件和卷积组件用于提取全局特征与局部特征并将其进行融合,我们称这 4 个模块为全局与局部特征融合模块(Global and Local Feature Fusion Module)。我们称提出的模型为 RainbiFormer。我们在随后的两个小节部分对时序空间 Encoder 和空间 Forecaster 进行详细的介绍。

3.1. 时序空间 Encoder

在提出模型的时序空间 Encoder 部分,我们引入双级路由注意力机制的 Transformer 模块对输入的每个回波进行全局特征提取,这样不但学习到回波的空间特征、同时也能够提取到整个回波序列的时序特征。双级路由注意力的结构如图 2 中的 Global Module 所示。双级路由注意力主要是为了降低 Transformer 中 multihead self-attention (MHSA)计算的复杂度[19]。双级路由注意力首先将输入的 2D 特征 $X \in R^{H \times W \times C}$ 划分为 $S \times S$ 的非重叠区域,这些区域包含 $\frac{HW}{S^2}$ 个特征向量。通过这种划分方式得到的 query, key, value 向量为 $Q, K, V \in R^{S^2 \times \frac{HW}{S^2} \times C}$ 。这些向量执行的线性投影操作如公式 1 所示。

$$Q = X^r W^q, K = X^k W^k, V = X^v W^v \quad (1)$$

其中 W^q, W^k 和 $W^v \in R^{C \times C}$ 分别为 query, key, value 的投影权值矩阵。然后通过有向图在区域之间进行交互。其中区域级的 queries 和 keys 为 $Q^r, K^r \in R^{S^2 \times C}$, 区域之间的仿射图通过 Q^r 和 K^r 的转置矩阵乘得到邻接矩阵 $A^r \in R^{S^2 \times S^2}$, 操作定义如公式 2。

$$A^r = Q^r (Q^r)^T \quad (2)$$

然后仅保留每个区域的仿射图中前 k 个连接,从而使用行相关操作得到路由索引矩阵 $I^r \in N^{S^2 \times k}$ 。根据现代 GPU 合并操作原理,将 key 和 value 向量首先聚合起来,聚合的 key 和 value 向量表示为公式 3 所示。

$$K^g = \text{gather}(K, I^r), V^g = \text{gather}(V, I^r) \quad (3)$$

聚合的 key 和 value 向量在聚集的 key 和 value 值对上使用 Attention 操作,同时使用了一个局部场景增强项。整个双级路由注意力操作如公式 4 所定义

$$O = \text{Attention}(Q, K^g, V^g) + LCE \quad (4)$$

局部特征提取单元则是将整个雷达回波序列作为一个时空数据整体进行处理和学习,它的结构如图 2 所示。局部特征提取单元首先得到特征图的最大值和平均值,然后将其对应乘积到特征图,得到转换后的两个特征图通过卷积核为 3 的卷积操作进行降维融合操作。基于全局特征提取的特征与基于局部特征提取单元的特征再通过特征融合单位进行信息的融合处理,这样 Encoder 模块能够高效的提取雷达回波序列的时空特征,学习得到雷达回波的运动轨迹。

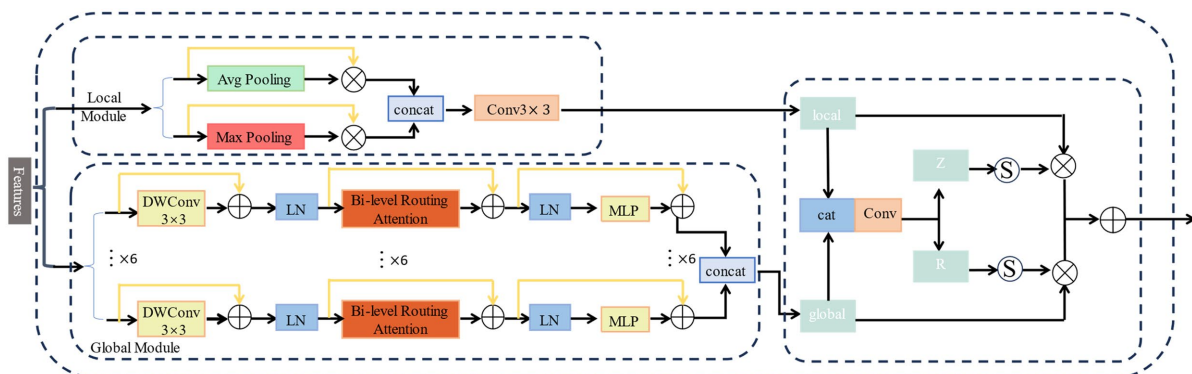


Figure 2. Spatial timing Encoder module structure framework diagram

图 2. 空间时序 Encoder 模块结构框架图

在本文中采用 Batch 为 12，输入 6 个回波，预测未来 12 个回波的方式。也就是根据当前的半小时回波情况预测未来 1 小时的回波运行状态。时序空间 Encoder 中的四个模块中的全局特征提取模块都是采用独立的处理方式对每个回波进行单独提取。Local Module 采用 Max Pooling 与 Avg Pooling 两种操作方式分别提取回波的局部特征，然后再将得到特征图进行 concat 操作。而 Local Model 对回波进行整体操作，不再使用时序方式进行。Global Module 对 6 个时序回波进行特征提取后再在 channel 上进行 concat 操作(Batch channel Height, width)的方式，这样与 Local Module 得到的结果形式一样，再进行融合操作。

3.2. 空间 Forecaster

空间 Forecaster 同样是由基于全局特征提取的空间 Transformer 组件和局部特征提取组件所构成，他们的输入首先执行 PatchExpanding 操作放大特征图的尺寸。其输入数据格式为 $B \times C \times H \times W$ ，表示为 Batch、channel、height 和 width。Forecaster 中模块的结构如图 3 所示。Forecaster 中模块使用的组件与 Encoder 模块中的组件是一样的。但在 Forecaster 模块中只考虑回波序列的空间特征，因此他们的处理是将输入的回波数据作为一个整体进行提取操作，但不考虑输入数据的时序性。Forecaster 输出的特征图的通道数为 96，然后再加入一个卷积核为 1 的卷积操作将其维度降为所需要预测的回波数目 12。

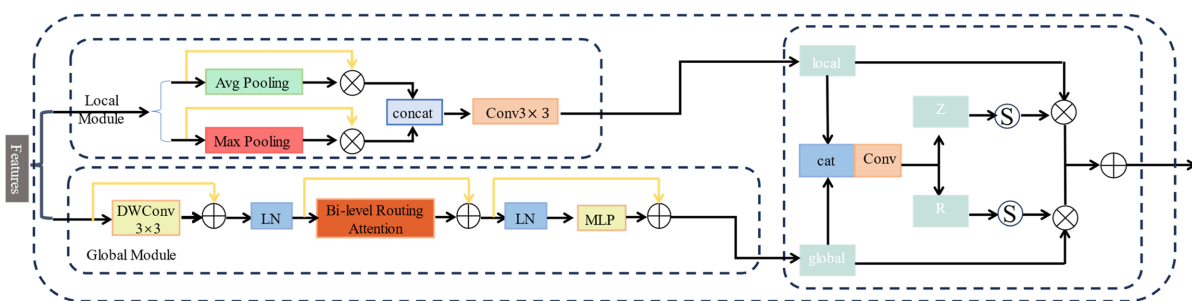


Figure 3. Structural framework diagram of the space Forecaster module

图 3. 空间 Forecaster 模块结构框架图

4. 实验结果与分析

为了验证提出算法模型的有效性，我们选择 Storm Event Imagery (SEVIR) [20]数据集中的 NEXRAD 天气雷达数据，该数据的空间分辨率是 1 公里，时间分辨率是 5 分钟，图像大小是 384*384。同时该数据集包含了 4 个小时为一个时间段的 20,393 个事件。在实际验证过程中根据 GPU 和我们将这个数据集按

照论文[21]中的划分方式将其划分为训练集、验证集和测试集三部分。模型输入 6 帧, 预测未来的 12 帧, 即根据过去的半小时回波情况, 外推未来的 1 小时回波信息。在 4 小时的数据集中, 采用时间跨度为 12, 训练集、验证集和测试集的个数分别为 35,718, 9060 和 12,159。为了有效验证我们提出模型, 我们采用 ConvLSTM, SmaAt-UNet, LPT-QPN 和 Rainformer 等四个模型与我们提出的模型进行量化和可视化两种对比方式。所有模型都使用深度学习平台框架 Pytorch 进行实现。实验过程中采用了 AdamW 优化器, 该优化器是在 Adam 基础上增加了权值衰减正则化用于提升训练模型的收敛速度。学习率的初始值设置为 0.001, 我们同时选择余弦退火调整学习率, 模型训练结束的迭代次数设置为最大 60 次, 在训练过程中我们取验证集中的 CSI 值最大的模型用于测试集。为了量化模型的预报效果, 我们采用命中率(probability of detection, POD)和临界成功子数(Critical Success Index, CSI)两个准则来量化预报结果, 这两个指标的值越大表示预测的效果越好。我们使用 $HITS = (真实 = 1, 预测 = 1)$, $FALSE = (真实 = 0, 预测 = 1)$, $MISSES = (真实 = 1, 预测 = 0)$ 来表示真实情况和预测情况。其中 POD 和 CSI 的定义分别如公式 5 和 6 所示。

$$POD = \frac{HITS}{HITS + FALSE} \quad (5)$$

$$CSI = \frac{HITS}{HITS + FALSE + MISSES} \quad (6)$$

4.1. 实验结果量化对比

在实际的气象预报业务应用中对不同强度的回波值的关注度不一样, 回波值越高表示发生的强对流天气的强度越大, 因此为了更好的验证提出模型在预报强对流天气方面的作用及预报效果, 我们也采用论文中将回波值划分为 16, 74, 133, 160, 181, 219 等 6 个区间进行量化对比验证。我们对所给对比数据取得最好结果的记录用黑体标注, 排在次好的记录用下划线表示。表 1 给出的是各个模型的平均 POD 值及 6 个像素值区间的 POD 值。我们从表 1 可以看出提出的模型在得到的平均 POD 值和多个阶段的 POD 值都远远高于排在第二位模型的数据。我们提出的模型得到的 POD 平均值比排在第二的 LPT-QPN 高接近 2.7 个百分点, 同时提出模型得到的 POD 值在几个高强度值区间都高于其他模型, 其中最多高的在 POD-133 中高近 9 个百分点, 这表明提出模型在预报强对流天气时具有更好的优势。在 POD 值方面只是在 219 时比最好模型 LPT-QPN 的值低接近 1.5。量化实验结果表明提出的模型具有非常好的预报能力, 这表明提出模型的有效性。

Table 1. Quantitative comparison of the POD for various levels of all models

表 1. 模型对比的各级 POD 值

	POD-M	POD-16	POD-74	POD-133	POD-160	POD-181	POD-219
ConvLSTM	0.4694	0.9006	0.7711	0.4579	0.3076	0.2570	0.1219
SmaAt-UNet	0.4437	0.9139	0.7778	0.4071	0.2677	0.2051	0.0905
LPT-QPN	0.5063	<u>0.9138</u>	<u>0.7975</u>	<u>0.5033</u>	<u>0.3466</u>	0.2973	0.1793
RainFormer	0.4982	0.9125	0.7831	0.4926	0.3432	0.2890	<u>0.1687</u>
RainbiFormer	0.5319	0.9179	0.8281	0.5925	0.3819	0.3051	0.1651

表 2 给出的是各个模型在测试集中取得的 CSI 平均值及各个区间的 CSI 值, 从表中的数据可以看出, 在量化指标 CSI 值方面没有哪个模型得到非常大的优势结果。SmarAt-UNet 得到的效果最低, 这可能是 SmarAt-UNet 模型只考虑了回波序列的空间特征而完全没有考虑时序特征的结果。ConvLSTM 模型因为

将回波序列的时空特征都进行了考虑并进行时空特征提取,因此得到了比 SmarAt-UNet 模型更好的结果,ConvLSTM 模型得到的 CSI 平均值比 SmarAt-UNet 模型高 2.2 个百分点。三个基于 Transformer 的模型得到比 ConvLSTM 更好的 CSI 结果,这表明 Transformer 相对于 CNN 能够更加有效提取回波目标的全局特征信息,因此模型能够取得更好的外推结果。CSI 平均值中 LPT-QPN, RainFormer 和我们提出的模型都非常接近,但我们提出的模型在高强度值 133,160 和 181 中都取得最好的结果,同时在最高强度值上取得的 CSI 值仅比最好的 RainFormer 仅仅相差 0.23 个百分点。这表明相对于其他模型,提出的模型对强对流预测结果具有更好和更综合的预测结果。

Table 2. CSI values for each level of model comparison
表 2. 模型对比的各级 CSI 值

	CSI-M	CSI-16	CSI-74	CSI-133	CSI-160	CSI-181	CSI-219
ConvLSTM	0.4013	0.7546	0.6704	0.3772	0.2680	0.2258	0.1116
SmaAt-UNet	0.3784	0.7394	0.6654	0.3476	0.2418	0.1892	0.0867
LPT-QPN	0.4135	0.7408	<u>0.6711</u>	0.3890	0.2867	0.2478	<u>0.1455</u>
RainFormer	0.4175	<u>0.7483</u>	0.6758	<u>0.3944</u>	<u>0.2917</u>	<u>0.2489</u>	0.1457
RainbiFormer	<u>0.4165</u>	0.7387	0.6676	0.4015	0.2938	<u>0.2540</u>	0.1434

4.2. 实验结果视觉对比

为了量化图 4 中给出的预报结果,我们在表 3 和表 4 中分别给出预报的平均 CSI 和 POD 值,第 10、20、30、40、50 和 60 分钟的 CSI 及 POD 值。从表 3 中我们可以看出,提出模型在预报的 CSI 平均值及 30 分钟和其后的时间里的 CSI 值都是最高的,特别是在第 60 分钟进行的预报结果高于排在第二位模型接近 4 个百分点。而对于 POD 评价标准,提出模型在均值和所有时间段的结果都优于其他模型。通过这两个表给出的量化标准表明提出的模型 RainbiFormer 在较长预报时间时的结果优于其他模型,这表明我们提出模型能够更好的预测更长时间的结果,这充分证明了提出模型的有效性。

Table 3. Comparative results of CSI values at different time intervals for visual contrasts
表 3. 视觉对比的不同时间间隔的 CSI 值对比结果

	CSI-M	CSI-10	CSI-20	CSI-30	CSI-40	CSI-50	CSI-60
ConvLSTM	0.5826	0.7073	0.6205	0.5943	<u>0.5500</u>	0.4839	0.4421
SmaAt-UNet	0.5737	0.7278	0.6025	0.5619	0.5296	0.4788	0.4426
LPT-QPN	0.5836	0.7216	0.6260	0.5808	0.5379	<u>0.4898</u>	<u>0.4505</u>
RainFormer	<u>0.5870</u>	0.7359	0.6445	<u>0.5956</u>	0.5451	0.4739	0.4365
RainbiFormer	0.6065	<u>0.7310</u>	<u>0.6399</u>	0.5971	0.5645	0.5310	0.4891

Table 4. Comparative results of POD values at different time intervals for visual comparison
表 4. 视觉对比的不同时间间隔的 POD 值对比结果

	POD-M	POD-10	POD-20	POD-30	POD-40	POD-50	POD-60
ConvLSTM	0.6983	0.7963	0.7285	0.7114	0.6803	0.6098	0.5679
SmaAt-UNet	0.6540	0.8072	0.6834	0.6421	0.6126	0.5586	0.5185
LPT-QPN	0.7193	0.8342	0.7654	<u>0.7375</u>	<u>0.6842</u>	<u>0.6269</u>	<u>0.5880</u>
RainFormer	0.7094	<u>0.8464</u>	0.7608	0.7217	0.6804	0.6025	0.5680
RainbiFormer	0.7695	0.8602	0.7925	0.7638	0.7444	0.7097	0.6611

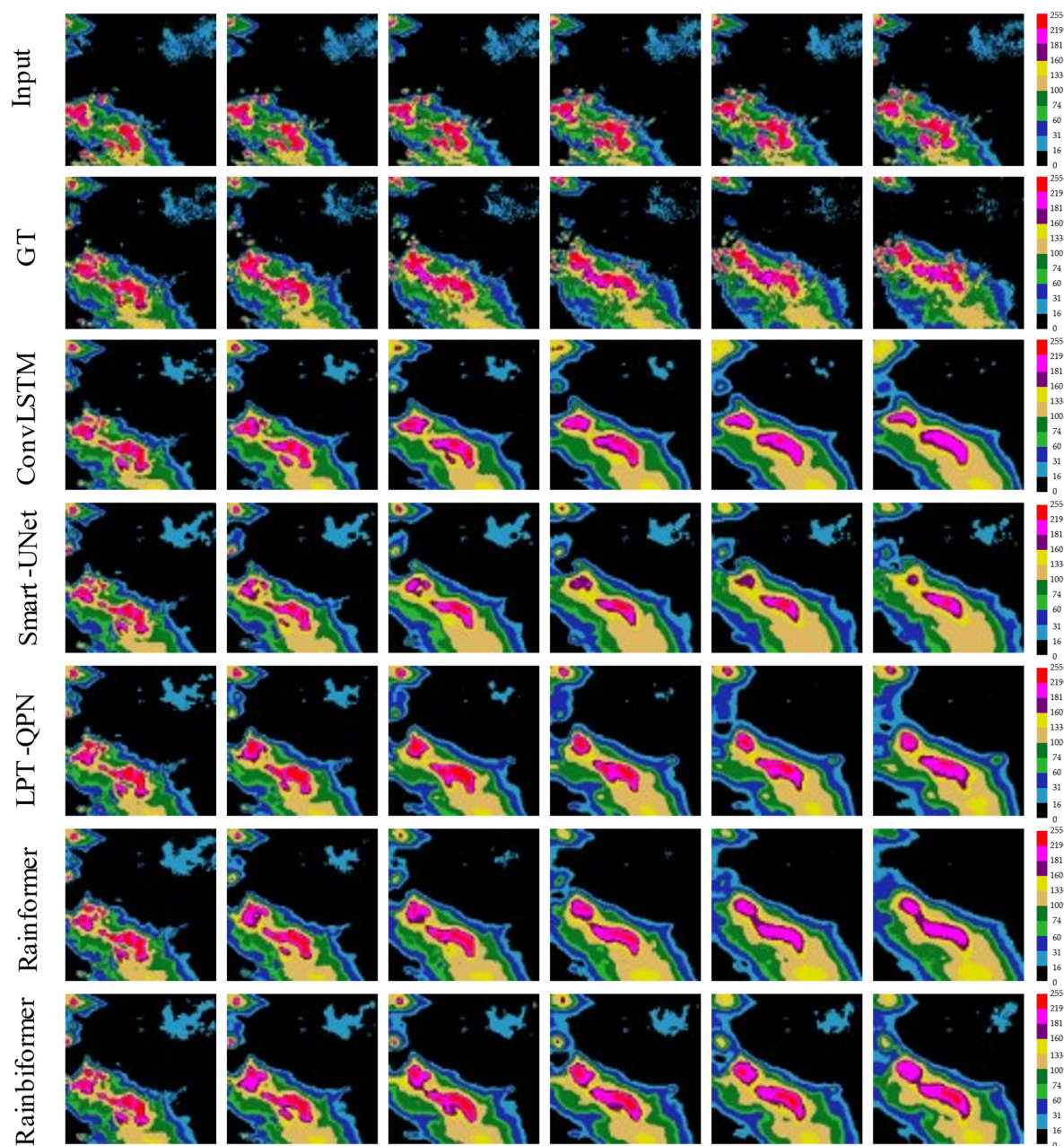


Figure 4. Visual comparison of all models

图 4. 所有模型的视觉对比

5. 结论与展望

本研究针对临近预报应用中对强对流天气预报准确度不高的情况，提出了一种基于 Transformer 的时空预报模型，该模型采用 Encoder-Forecaster 架构，每个 Encoder 和 Forecaster 包含四个模块。Encoder 中的模块采用时空处理方式对每个回波序列按时序分别进行单独和整体两种方式的处理，这样能够有效提取回波序列的时序和空间特征，从而实现对回波轨迹运动规律的高效建模。Forecaster 端包含的四个模块的输入包括前一模块的输出和 Encoder 端对应模块的输出，这两个输出进行拼接后进行融合，这样能够有效组合回波特征。所有模块中使用了提取目标对象的局部特征的最大池化操作和均值池化操作

及提取全局特征的双级路由注意力机制,这两种特征的结合实现了模型对高强度值回波的高准确度预报,实验结果验证了这一假设。

本研究提出的模型主要还是专注于从大量的气象数据中学习回波的运动轨迹,但没有考虑回波运动轨迹的气象物理规律及天气预报的不确定性规律。因此我们未来的研究将着重从深度学习模型提取回波运动的物理规律入手,同时考虑结合生成模型[22]建模天气预报的不确定性规律。通过这些方法研究的模型可能提高对高强度回波值的预测,从而能够提升对强对流天气预报结果的准确性。

参考文献

- [1] Han, L., Zhang, J., Chen, H., Zhang, W. and Yao, S. (2022) Toward the Predictability of a Radar-Based Nowcasting System for Different Precipitation Systems. *IEEE Geoscience and Remote Sensing Letters*, **19**, 1-5. <https://doi.org/10.1109/lgrs.2022.3185031>
- [2] Reyniers, M. (2008) Quantitative Precipitation Forecasts Based on Radar Observations: Principles, Algorithms and Operational Systems. Institut Royal Météorologique de Belgique.
- [3] Sun, J., Xue, M., Wilson, J.W., Zawadzki, I., Ballard, S.P., Onvlee-Hoimeyer, J., et al. (2014) Use of NWP for Nowcasting Convective Precipitation: Recent Progress and Challenges. *Bulletin of the American Meteorological Society*, **95**, 409-426. <https://doi.org/10.1175/bams-d-11-00263.1>
- [4] Cheung, P. and Yeung, H.Y. (2012) Application of Optical-flow Technique to Significant Convection Nowcast for Terminal Areas in Hong Kong. *The 3rd WMO International Symposium on Nowcasting and Very Short Range Forecasting (WSN12)*, Rio de Janeiro, 6-10 August 2012, 6-10.
- [5] Rinehart, R.E. (1981) A Pattern Recognition Technique for Use with Conventional Weather Radar to Determine Internal Storm Motions. *Atmos Technologies*, **13**, 119-134.
- [6] 曹春燕, 陈元昭, 刘东华. 光流法及其在临近预报中的应用[J]. 气象学报, 2015, 73(3): 471-480.
- [7] LeCun, Y., Bengio, Y. and Hinton, G. (2015) Deep learning. *Nature*, **521**, 436-444. <https://doi.org/10.1038/nature14539>
- [8] Shi, X., Chen, Z. and Wang, H. (2015) Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. *Advances in Neural Information Processing Systems 28 (NIPS 2015)*, Montreal, 7-12 December 2015, 802-810. <https://proceedings.neurips.cc/paper/2015/hash/07563a3fe3bbe7e3ba84431ad9d055af-Abstract.html>
- [9] Shi, X., Gao, Z. and Lausen, L. (2017) Deep Learning for Precipitation Nowcasting: A Benchmark and a New Model. *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, California, 4-9 December 2017, 5617-5627. <https://proceedings.neurips.cc/paper/2017/hash/a6db4ed04f1621a119799fd3d7545d3d-Abstract.html>
- [10] Wang, Y., Gao, Z., Long, M., et al. (2018) PredRNN++: Towards A Resolution of the Deep-in-Time Dilemma in Spatiotemporal Predictive Learning. arXiv: 1804.06300.
- [11] Trebing, K., Stańczyk, T. and Mehrkanon, S. (2021) SmaAT-UNet: Precipitation Nowcasting Using a Small Attention-UNet Architecture. *Pattern Recognition Letters*, **145**, 178-186. <https://doi.org/10.1016/j.patrec.2021.01.036>
- [12] Wang, Y., Wu, H., Zhang, J., Gao, Z., Wang, J., Yu, P.S., et al. (2023) PredRNN: A Recurrent Neural Network for Spatiotemporal Predictive Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **45**, 2208-2225. <https://doi.org/10.1109/tpami.2022.3165153>
- [13] Ravuri, S., Lenc, K., Willson, M., Kangin, D., Lam, R., Mirowski, P., et al. (2021) Skilful Precipitation Nowcasting Using Deep Generative Models of Radar. *Nature*, **597**, 672-677. <https://doi.org/10.1038/s41586-021-03854-z>
- [14] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N. and Polosukhin, I. (2017) Attention Is All You Need. *Advances in Neural Information Processing Systems*, **30**, 6000-6010.
- [15] Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. (2021) An Image Is Worth 16 x 16 Words: Transformers for Image Recognition at Scale. *Proceedings of the International Conference on Learning Representations*, Vienna, 4 May 2021, 11929.
- [16] Bai, C., Sun, F., Zhang, J., Song, Y. and Chen, S. (2022) Rainformer: Features Extraction Balanced Network for Radar-Based Precipitation Nowcasting. *IEEE Geoscience and Remote Sensing Letters*, **19**, 1-5. <https://doi.org/10.1109/lgrs.2022.3162882>
- [17] Li, D., Deng, K., Zhang, D., Liu, Y., Leng, H., Yin, F., et al. (2023) LPT-QPN: A Lightweight Physics-Informed Transformer for Quantitative Precipitation Nowcasting. *IEEE Transactions on Geoscience and Remote Sensing*, **61**, 1-19. <https://doi.org/10.1109/tgrs.2023.3328945>
- [18] Chen, S., Shu, T., Zhao, H., Zhong, G. and Chen, X. (2023) TempEE: Temporal-Spatial Parallel Transformer for Radar Echo Extrapolation Beyond Autoregression. *IEEE Transactions on Geoscience and Remote Sensing*, **61**, 1-14.

<https://doi.org/10.1109/tgrs.2023.3311510>

- [19] Zhu, L., Wang, X., Ke, Z., Zhang, W. and Lau, R. (2023) BiFormer: Vision Transformer with Bi-Level Routing Attention. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, 17-24 June 2023, 10323-10333. <https://doi.org/10.1109/cvpr52729.2023.00995>
- [20] Veillette, M., Samsi, S. and Mattioli, C. (2020) SEVIR: A Storm Event Imagery Dataset for Deep Learning Applications in Radar and Satellite Meteorology. *Advances in Neural Information Processing Systems*, **33**, 22009-22019.
- [21] Gao, Z., Shi, X., Wang, H., *et al.* (2023) Earthformer: Exploring Space-Time Transformers for Earth System Forecasting. *Proceedings of the 36th International Conference on Neural Information Processing Systems*, New Orleans, 28 November-9 December 2022, 2539-2540.
- [22] Yu, D., Li, X., Ye, Y., Zhang, B., Luo, C., Dai, K., *et al.* (2024) DiffCast: A Unified Framework via Residual Diffusion for Precipitation Nowcasting. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 16-22 June 2024, 27758-27767. <https://doi.org/10.1109/cvpr52733.2024.02622>