

基于交叉型窗口自注意力机制的Transformer 临近预报

张坤林, 张福贵, 雷勇平, 熊太松*

成都信息工程大学电子工程学院, 四川 成都

收稿日期: 2025年3月3日; 录用日期: 2025年4月1日; 发布日期: 2025年4月9日

摘 要

临近预报旨在预测未来0~2小时的天气情况, 这对于减轻强对流天气的影响并保障人们的日常活动至关重要。作为当前天气预报中最具挑战性的任务之一, 临近预报需要同时具备高精度和高时效性。为了应对这一挑战, 本文提出了一种基于Transformer的编码器-解码器临近预报模型结构。该模型使用基于交叉型窗口的自注意力机制, 从而高效捕捉天气雷达回波的全局特征; 同时模型结合卷积块和高效的多尺度注意力机制, 实现对多尺度局部特征的有效提取。模型通过引入门控机制有效地融合全局与局部特征, 从而进一步提升模型预报性能。提出的模型在公开的上海雷达回波数据集上的实验结果有效验证了它的有效性和实用性。

关键词

临近预报, 强对流天气, 基于交叉型窗口的自注意力机制, Transformer, 特征提取与融合

Transformer Nowcasting Based on Cross-Window Self-Attention Mechanism

Kunlin Zhang, Fugui Zhang, Yongping Lei, Taisong Xiong*

College of Electronic Engineering, Chengdu University of Information Technology, Chengdu Sichuan

Received: Mar. 3rd, 2025; accepted: Apr. 1st, 2025; published: Apr. 9th, 2025

Abstract

Nowcasting aims to predict weather condition within the next 0~2 hours, which is critical for

*通讯作者。

文章引用: 张坤林, 张福贵, 雷勇平, 熊太松. 基于交叉型窗口自注意力机制的 Transformer 临近预报[J]. 计算机科学与应用, 2025, 15(4): 95-105. DOI: 10.12677/csa.2025.154082

mitigating the impacts of severe convective weather and ensuring the safety of daily activities. As one of the most challenging tasks in modern weather forecasting, nowcasting demands both high accuracy and timeliness. To alleviate this challenge, this paper proposes a Transformer based on encoder-decoder architecture for nowcasting. The proposed model employs a cross-shaped window self-attention mechanism to efficiently capture the global features of radar echo maps. Additionally, it uses convolutional blocks and a multi-scale attention mechanism to effectively extract multi-scale local features. By incorporating a gating mechanism, the model effectively fuses global and local features, further enhancing forecasting performance. Experimental results demonstrate that the proposed model achieves superior performance on the public Shanghai radar echo dataset.

Keywords

Nowcasting, Severe Convective Weather, Cross-Window Self-Attention Mechanism, Transformer, Feature Extraction and Fusion

Copyright © 2025 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

近年来,随着社会经济的快速发展,灾害性天气对社会经济和生命财产的威胁日益加剧。强对流天气作为我国主要的灾害性天气之一,具有突发性强、破坏力大、空间范围小、持续时间短等特点[1],这些特点给预报结果的准确性带来了巨大的挑战。强对流天气的精准预报给防灾减灾措施的制定和实施提供了重要的气象指导。0~2 小时的临近预报旨在利用历史气象数据预测未来 2 小时内的天气情况,以便有效预测并应对潜在的强对流天气事件。尽管在过去几十年间,研究学者在临近预报领域取得了一定成果,但预报的准确性还需要进一步提高。

2. 相关工作

2.1. 传统临近预报方法

传统临近预报方法主要包括数值天气预报[2]和基于雷达外推[3]的方法。数值天气预报是基于物理方程和气象观测数据来预测天气变化的一种方法。该方法通过建立大气运动的数学模型,利用高性能计算机进行数值模拟,从而得到未来天气的预报结果。虽然数值天气预报能有效地预测天气的长期运动趋势,但针对于临近预报任务而言,该方法存在一定的局限性,比如难以精确捕捉到小尺度的降水等[4]。因此,性能更好地基于雷达外推的方法成为了临近预报系统中的主流方法。这种方法具有较高的时空分辨率,能够较好地捕捉到小尺度降水系统的变化。

交叉相关法[5]和光流法[6]是两种代表性的传统外推方法,它们主要依赖于雷达观测数据,通过分析天气系统的运动趋势,将当前的雷达回波分布外推到未来时刻,从而实现临近预报。然而,这些方法也存在一些局限性。首先,它们假设雷达回波的运动是均匀的,即回波的形状和强度在短时间内保持不变,这通常与实际降水的演变趋势不符[7]。其次,这些方法仅考虑了相邻两帧雷达回波图像之间的相关性,而忽略了连续序列中其他帧之间的时空依赖关系。因此在预报降水的强度变化和持续时间方面存在一定的困难。基于雷达外推的方法还需要进一步的改进,才能满足当今社会对临近预报的需求。

2.2. 基于深度学习的临近预报方法

近年来,深度学习[8]技术得到了广泛的关注和研究。该技术具有强大的特征学习能力和模型表达能

力,能够从大量的历史数据中学习到复杂的非线性关系和时空模式,这为解决临近预报此类密集型预测任务提供了新的研究方向。

基于深度学习的时间序列预测技术,在降水预测中起到了重要的里程碑作用。Shi 等人将临近预报任务视为时间序列预测任务,并提出了卷积长短期记忆神经网络(Convolutional Long Short-Term Memory, ConvLSTM) [9]。该模型通过在传统的长短期记忆网络(Long Short-Term Memory, LSTM) [10]中引入卷积操作,有效地捕捉到雷达回波数据中的空间结构信息,从而更好地预测未来的天气状况。为了解决 ConvLSTM 在处理复杂降水系统进行运动时难以准确捕捉其运动趋势的不足,Shi 等人进一步提出了轨迹门控循环单元(Trajectory Gated Recurrent Unit, TrajGRU) [11],它可以根据输入数据学习位置变化的连接结构,从而更好地捕捉运动模式的动态变化。为了解决时空预测学习中长期依赖建模不足和信息流动受限的问题,Wang 等人提出了预测递归神经网络(Predictive Recurrent Neural Network, PredRNN) [12]。PredRNN 的锯齿型记忆流机制允许记忆单元跨层通信,有效提升了模型对降水系统短期形变特征与长期演变规律的多尺度动态建模能力。在 PredRNN 的基础上,PredRNN-V2 [13]结合了记忆去耦损失和反向调度采样策略,进一步提高了模型对复杂的时空变化进行捕捉的能力和模型生成图像的质量。

除了上述基于循环神经网络(Recurrent Neural Network, RNN)的工作,卷积神经网络(Convolutional Neural Network, CNN)凭借出色的图像特征提取能力,在临近预报领域中也得到了广泛应用。Agrawal 等人在时间维度上连续输入多帧图像,并使用 U-Net [14]架构进行预测,其预测性能优于传统的外推方法 [15]。Trebing 等人以 U-Net [14]架构为基础,通过引入卷积注意力模块和深度可分离卷积,从而提出了 SmaAt-UNet [16]。它仅需原始 U-Net 四分之一的参数量,即可达到与其相当的性能表现。

虽然 CNN 在捕捉空间依赖关系方面表现优异,但其固有的归纳偏差(如平移不变性和局部感受野)限制了其捕捉空间变换和长期依赖关系的能力。相比之下,RNN 虽然擅长处理时间依赖关系,但其序列化特性使其无法实现并行化,导致反向传播效率低下。

Transformer [17]最初应用于自然语言处理(Natural Language Processing, NLP)领域,但其出色的长期依赖性建模能力使其在短时间内迅速扩展到计算机视觉[18]的多个领域,包括语义分割[19]、自动驾驶[20]以及医学图像分割[21]等。Transformer 特有的多头注意力机制能够高效捕获长期依赖关系,同时支持并行计算。Rainformer [22]应用基于窗口的多头注意力(Window-based Multi-head Self-Attention, W-MSA) [23]机制和全卷积网络学习全局特征,提升了高强度降雨预测的性能。LPTQPN [24]提出了一个带物理信息的轻量级 Transformer 框架,采用多头平方注意力(Multi-Head Squared Attention, MHSA)来模拟降水的高度非线性关系,同时降低了计算复杂度。TempEE [25]提出了一个多层次的时空注意机制,并结合一步前推策略来有效防止累积误差的扩散。这种机制和策略能有效捕捉降水数据中的时空动态变化,从而使 TempEE 能以较低的计算成本实现先进的预报性能。

尽管深度学习在临近预报领域取得了显著进展,但该领域的研究仍在不断进步之中。目前,研究工作主要集中在利用丰富的雷达回波数据,通过训练深度学习模型来学习雷达回波演变规律。随着深度学习技术的不断进步,本文紧跟领域前沿,提出了一种改进的基于 Transformer 的深度学习模型,并将其应用于临近预报任务,旨在提升模型对高强度回波的预测精度。通过这一研究,我们期望为深度学习在临近预报领域的应用提供新的思路,并推动该领域的进一步发展与优化。

3. 研究方法

3.1. 网络结构

如图 1 所示,我们提出的 CSTN-PN (CSwinTransformer Network for Precipitation Nowcasting)网络架构

由编码器和解码器两部分组成，每个部分均包含四个阶段，分别标记为 Stage 1 至 Stage 4。该结构通过逐层提取全局与局部特征而后进行特征融合，实现了对未来降水的预测。编码器的每个阶段均包含全局特征提取模块(Global Feature Extraction Module, GFEM)、局部特征提取模块(Local Feature Extraction Module, LFEM)以及特征融合模块(Feature Fusion Module, FFM)。在每个阶段的初始部分，模型通过 Patch Merging 操作减少特征图的尺寸并增加通道数，从而实现更高效的特征提取。随后，GFEM 和 LFEM 同步提取特征，并通过 FFM 对提取到的特征进行平衡与融合。

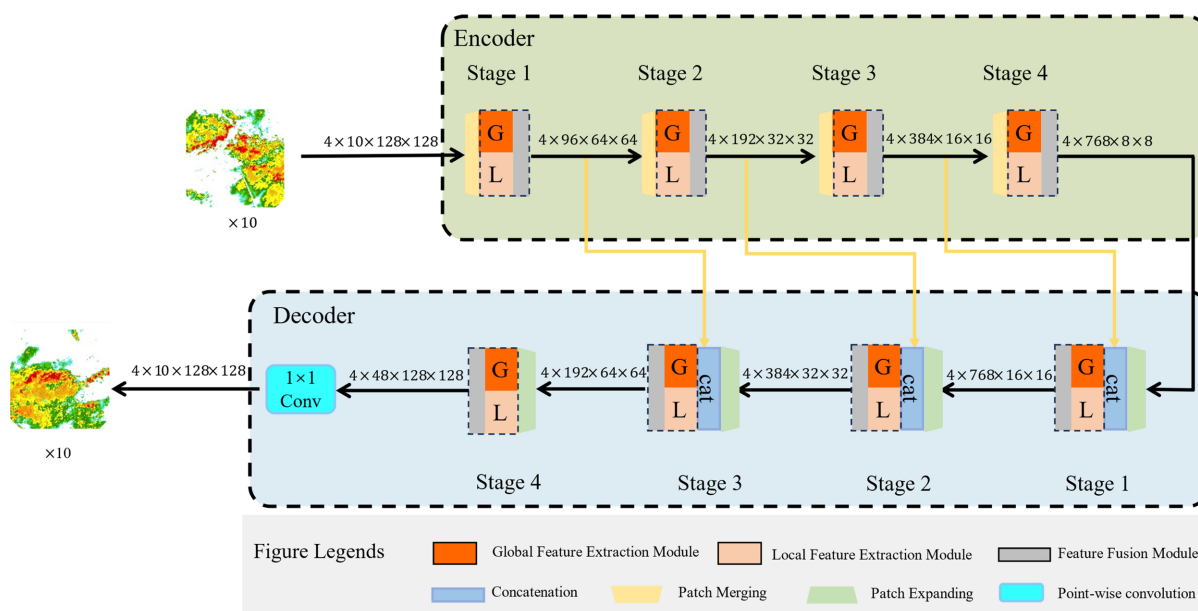


Figure 1. Diagram of the CSTN-PN

图 1. CSTN-PN 结构图

解码器的结构与编码器类似，但在每个阶段的末尾通过 Patch Expanding 操作逐步恢复特征图的尺寸并减少通道数。在解码器的 Stage 1 至 Stage 3 中，模型通过拼接(Concatenation, cat)操作将输入特征图与编码器对应阶段的输出特征图相结合，从而实现不同层次特征的连接与多尺度融合。此外，在 Stage 4 到模型输出之间，我们引入了一个卷积层来逐步降低通道维度，以减少信息丢失。具体而言，模型首先将特征图通过卷积操作过渡到一个中间缓冲层，最后将通道数减少到预测帧数 10，以缓解直接减少通道数时可能造成的信息丢失问题。解码器的设计有助于在特征提取过程中保留关键信息，并充分利用多层次特征，从而提升模型在降水预测任务中的整体性能。

3.2. 全局特征提取模块

全局特征提取模块(Global Feature Extraction Module, GFEM)是 CSTN-PN 的核心组件之一，其主要作用是增强模型对高强度降水特征的捕获能力。如图 2 左上方所示，我们引入了基于交叉型窗口的注意力机制的 CSwinTransformer [26]模块，用于提取降水系统的全局特征。

临近预报任务通常需要模型高效处理大规模数据，以实现高分辨率降水模式的密集预测。然而，传统的注意力机制由于计算复杂度与标记数量的平方成正比，因此，这类机制在处理高分辨率图像或大规模标记集时效率较低。为了缓解这种问题，我们采用了基于交叉型窗口的注意力机制，实现了对整个模型计算复杂度的有效降低，从而提升了模型的效率。

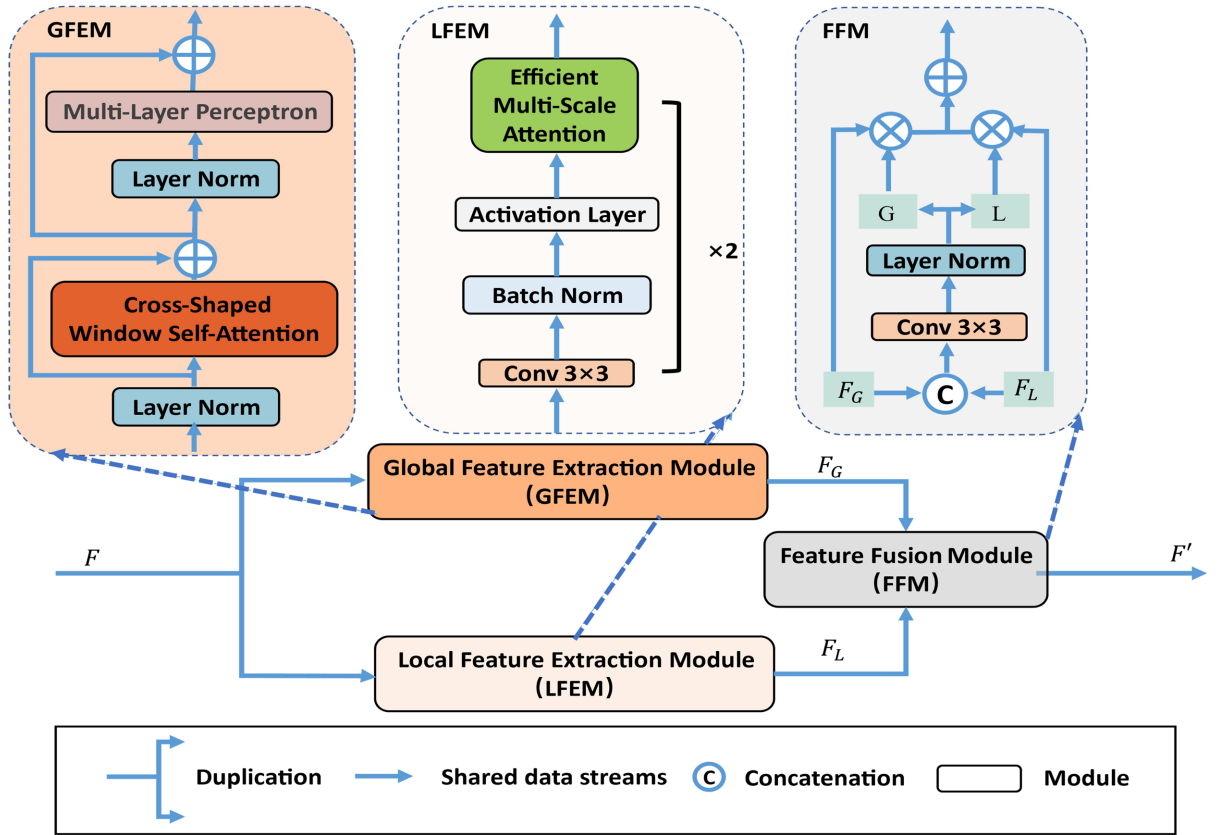


Figure 2. Architecture of the GFEM, LFEM and FFM

图 2. GFEM、LFEM、FFM 结构图

具体而言，基于交叉型窗口的注意力机制首先将输入特征图 $X \in R^{H \times W \times C}$ 沿着通道维度分割为 K 个头，每个头的维度为 C/K 。对于每个头 k ($k = 1, 2, \dots, K$)，自注意力计算分为水平和垂直两个方向。以水平方向为例，特征图 X 沿宽度方向被分割成 M 个非重叠的水平条纹 $[X_1, X_2, \dots, X_M]$ ，每个条纹的宽度为 sw 。对于每个条纹 X_i ，自注意力计算如下：

$$Y_i^k = \text{Attention}(X_i W_Q^k, X_i W_K^k, X_i W_V^k) \quad (1)$$

其中， $X_i \in R^{(sw \times W) \times C}$ ， $M = \frac{H}{sw}$ ， $i = 1, \dots, M$ 。 $W_Q^k \in R^{C \times d_k}$ ， $W_K^k \in R^{C \times d_k}$ 和 $W_V^k \in R^{C \times d_k}$ 分别是第 k 个头的查询 (Query)、键 (Key) 和值 (Value) 的线性变换矩阵。

接下来，将 K 个头分为两组，每组 $K/2$ 个头，分别执行水平和垂直方向自注意力。这样，每个头都能在各自的方向上捕捉长距离依赖。

$$\text{head}_k = \begin{cases} H - \text{Attention}_k(X) & k = 1, \dots, \frac{K}{2} \\ V - \text{Attention}_k(X) & k = \frac{K}{2} + 1, \dots, K \end{cases} \quad (2)$$

最后，将水平和垂直方向的自注意力输出在通道维度上进行拼接，并通过一个线性变换 W_o 进行融合，得到完整的交叉型窗口注意力：

$$\text{CSWin-Attention}(X) = \text{Concat}(\text{head}_1, \dots, \text{head}_K) W_o \quad (3)$$

其中, $W_o \in R^{C \times C}$ 是线性变换矩阵。最终的输出是经过自注意力加权的特征图, 其中包含了丰富的全局信息, 为后续的阶段提供了强有力的特征表示。这种设计不仅能提高模型的计算效率, 还能更有效地提取降水信息的全局特征, 从而增强模型对高强度降水系统的建模能力。

3.3. 局部特征提取模块

在临近预报任务中, 除全局特征具有重要的价值外, 局部特征同样发挥着不可或缺的作用。如图 2 正上方所示, 我们提出的局部特征提取模块(Local Feature Extraction Module, LFEM)专门用于从降水系统中提取关键局部特征, 以增强模型对中低强度降水的预测能力。该模块主要由卷积块和高效多尺度注意力机制(Efficient Multi-scale Attention, EMA) [27]组成。具体而言, EMA 通过将部分通道重构为批处理维度, 并将通道维度分组为多个子特征图, 这种做法既保留了每个通道上的特征信息, 又显著降低了计算开销。通过 LFEM, 模型能够有效捕捉局部细节信息, 从而进一步提升降水预报的性能。

3.4. 特征融合单元

特征融合模块(Feature Fusion Module, FFM)旨在实现全局与局部特征的有效融合。研究发现, 将全局与局部特征直接进行相加或相乘的方式并不能取得很好的融合效果。为此, 我们提出了 FFM, 其结构如图 2 右上方所示。FFM 主要由权重调整和特征融合两个步骤组成, 其计算过程如下:

$$G, L = \text{split}\left(\sigma\left(\text{Conv}\left(\text{Concat}\left(F_G, F_L\right)\right)\right)\right) \quad (4)$$

$$\text{out} = F_G * G + F_L * L \quad (5)$$

其中, F_L 和 F_G 分别表示局部特征和全局特征, L 和 G 为对应的遗忘矩阵。 L 和 G 的值介于 0 和 1 之间, 其功能类似于门控循环单元(Gated Recurrent Unit, GRU) [28]中的门控机制, 这种机制能够选择性地丢弃不相关信息并保留重要特征。FFM 将这种门控机制应用于全局和局部特征的融合过程中, 从而有效减少了特征之间的数值差异。通过这种设计, FFM 能改善全局和局部特征之间融合不平衡的问题。

4. 实验结果与分析

4.1. 实验设置

本文选择使用文献[29]提供的上海雷达数据, 该数据集涵盖了 2015 年 10 月至 2018 年 7 月期间生成的 17 万张雷达回波图像, 时间分辨率为 6 分钟, 空间分辨率为 1 公里。在数据预处理阶段, 我们剔除了因雷达故障或维护导致的非连续雷达回波, 并通过步长为 1 的滑动窗口生成了 76,779 个长度为 20 的序列样本。最终, 数据集被划分为训练集、测试集和验证集, 分别包含 44,099、20,440 和 12,260 个序列样本。模型输入 10 帧, 预测未来的 10 帧, 即根据过去的 1 小时回波情况, 外推未来的 1 小时回波信息。在预处理过程中, 我们首先对雷达图像进行中心裁剪, 保留中心 384×384 区域, 随后将图像尺寸调整为 128×128 以降低计算成本。

实验基于 PyTorch 框架(版本 1.8.2)构建, 并在 NVIDIA A40 图形处理器(Graphic Process Unit, GPU)上进行训练和优化。我们采用 Adam 优化器, 初始学习率设置为 0.0001, 学习率衰减步长为 20,000, 衰减因子为 0.7, 并使用 StepLR 作为学习率调度器。为防止模型过拟合, 我们在训练过程中持续监测验证集上的临界成功指数(Critical Success Index, CSI), 最终选择验证 CSI 最高的模型作为预测模型。此外, 我们对所有模型应用了梯度裁剪, 裁剪阈值设置为 50, 以稳定训练过程。为有效验证所提出模型的性能, 我们选择了 ConvLSTM、SmaAt-UNet、LPT-QPN 和 Rainformer 四个基准模型进行对比, 采用量化和可视化两种评估方式。

4.2. 评价指标

我们使用气象评价指标、图像质量评价指标来评估模型。首先，对于气象评价指标，我们使用气象预报领域中广泛采用的指标来评估模型性能，包括临界成功指数(Critical Success Index, CSI)和海德克技能分数(Heidke Skill Score, HSS)。这两个指标的计算过程如下：首先，设定降水阈值，将预测图像和真实图像中超过阈值的网格点分类为雨点(标记为 1)，低于阈值的网格点分类为非雨点(标记为 0)。随后，逐像素遍历预测图像和真实图像，统计以下四种情况的数量：TP=(真实 =1, 预测 =1)，FP=(真实 =0, 预测 =1)，TN=(真实 =0, 预测 =0)，FN=(真实 =1, 预测 =0)。基于上述统计结果，CSI 和 HSS 的计算公式如下：

$$CSI = \frac{TP}{TP + FN + FP} \quad (6)$$

$$HSS = \frac{2 \times (TP \times TN - FP \times FN)}{(TP + FN)(FN + TN) + (TP + FP)(FP + TN)} \quad (7)$$

评估以上的两个指标时，我们参考[9]的研究，将 0.5、2、5、10 和 30 mm/h 定义为具体阈值。

此外，我们采用了均方误差(Mean Squared Error, MSE)、峰值信噪比(Peak Signal-to-Noise Ratio, PSNR)和结构相似性指数(Structural Similarity Index, SSIM)来评估图像质量。MSE 用于量化预测值与真实值之间的像素级误差，PSNR 反映了预测图像的质量，而 SSIM 则从亮度、对比度和结构三个方面评估预测图像与真实图像之间的相似性。这些指标的结合使用能够全面评估模型在临近预报任务中的性能表现。

4.3. 实验结果量化对比

在对比分析中，我们将取得最佳结果的记录以黑体标注，次佳结果以下划线表示。在上海数据集上，我们将 CSTN-PN 与现有的先进模型进行了对比，结果如表 1 所示。在图像质量评价指标上，各模型表现相近，但 CSTN-PN 仍展现出一定的优势。尽管在 MSE 指标上，SmaAt-UNet 表现最佳，但 CSTN-PN 以 0.4% 的微小差距位居第二，显示出其较强的竞争力。

在气象评价指标方面，CSTN-PN 在 CSI 和 HSS 上均表现优异，分别比第二名的模型高出 2.1% 和 2.9%。这一结果表明，CSTN-PN 在预测降水事件的成功率以及区分降水与非降水区域的能力上具有显著优势，这进一步验证了其在临近预报任务中的有效性。

Table 1. Quantitative comparison of results obtained by all models

表 1. 模型结果各量化对比评价指标

	CSI↑	HSS↑	MSE↓	PSNR↑	SSIM↑
ConvLSTM	0.3464	0.4559	7.894	28.02	<u>0.8353</u>
SmaAt-UNet	0.3480	0.4600	7.774	27.82	0.8260
Rainformer	<u>0.3644</u>	0.4821	8.108	<u>28.07</u>	0.8342
LPT-QPN	0.3643	<u>0.4846</u>	8.180	27.79	0.8284
CSTN-PN	0.3748	0.4948	<u>7.810</u>	28.24	0.8368

表 2 展示了不同预测区间下各模型的 CSI 和 HSS 的得分情况。从结果可以看出，ConvLSTM 在预测中低强度降水时表现较好，但在处理高强度降水时表现欠佳。这可能是由于其卷积结构在捕捉全局信息方面存在缺陷，导致其对高强度降水的预测能力不足。SmaAt-UNet 的整体预报效果最差，这可能是由于

该模型仅依赖卷积结构提取空间特征，而缺乏对时序特征的充分建模，从而降低了其对全局信息的捕捉能力。基于 Transformer 的模型 Rainformer 和 LPT-QPN 在低强度降水预测中的表现略逊于 ConvLSTM，但在高强度降水预测中表现更为优异。这表明 Transformer 结构相较于传统的卷积结构，能够更有效地提取回波的全局特征信息，从而更好地预测强对流天气。相比之下，我们提出的 CSTN-PN 在大多数预测区间内均取得了最佳结果，尤其是在高强度($r \geq 10, 30$)降水预测中表现尤为突出。与其他方法相比，CSTN-PN 在高强度降水预测中的优势体现了其对强对流天气的预测能力。

Table 2. Quantitative comparison of the CSI and HSS for various levels obtained by all models
表 2. 模型对比的各级 CSI 和 HSS 值

Threshold	CSI↑					HSS↑				
	$r \geq 0.5$	$r \geq 2.0$	$r \geq 5.0$	$r \geq 10.0$	$r \geq 30.0$	$r \geq 0.5$	$r \geq 2.0$	$r \geq 5.0$	$r \geq 10.0$	$r \geq 30.0$
ConvLSTM	0.6994	<u>0.4752</u>	0.2739	0.2456	0.0378	0.7962	<u>0.6224</u>	<u>0.4179</u>	0.3797	0.0632
SmaAt-UNet	0.6815	0.4740	0.2644	0.2382	0.0821	0.7838	0.6134	0.3997	0.3662	0.1367
Rainformer	0.6879	0.4790	0.2745	<u>0.2560</u>	0.1245	0.7867	0.6179	0.4268	0.3795	0.1995
LPT-QPN	0.6875	0.4706	<u>0.2750</u>	0.2554	<u>0.1328</u>	0.7869	0.6176	0.4271	<u>0.3813</u>	<u>0.2100</u>
CSTN-PN	<u>0.6987</u>	0.4814	0.2802	0.2670	0.1467	<u>0.7958</u>	0.6275	0.4329	0.3950	0.2226

量化实验结果表明，我们提出的模型具有出色的预报能力，证明了模型的实用性和可靠性。

4.4. 实验结果视觉对比

为了进一步体现 CSTN-PN 对高强度降水的预测能力，图 3 显示了一个随机选择的可视化示例。从这个图中我们可以看出，所有模型在初始阶段均能生成与真实回波图像较为接近的预测结果。然而，随着时间的推移，ConvLSTM 和 SmaAt-UNet 的预测能力明显下降，生成结果开始模糊，尤其是高强度降水区域，其范围和边界细节表现出显著偏离。Rainformer 虽然在边界平滑性上有所改善，但其对高强度区域的预测能力仍有不足，存在一定的漏报现象。LPT-QPN 能够较好地捕提高强度区域，但其预测结果中存在较多的误报。此外，随着时间的增加，这些模型的预测结果在低强度区域存在一定的扩散现象，虚假回波增多。

Table 3. CSI and HSS values at different time intervals for visual comparison
表 3. 视觉对比的不同时间间隔的 CSI 和 HSS 值

TimeStamp	CSI↑						HSS↑					
	t = 12	t = 24	t = 36	t = 48	t = 60	mean	t = 12	t = 24	t = 36	t = 48	t = 60	mean
ConvLSTM	0.6548	0.5514	0.4882	0.4509	0.4278	0.5146	0.7440	0.6699	0.6023	0.5627	0.5182	0.6194
SmaAt-UNet	0.6583	0.5382	0.4787	0.4435	0.4321	0.5102	0.7409	0.6585	0.5841	0.5582	0.5167	0.6117
Rainformer	<u>0.6684</u>	<u>0.5611</u>	0.4888	<u>0.4613</u>	0.4445	0.5248	<u>0.7875</u>	<u>0.6815</u>	0.6112	0.5625	0.5427	0.6371
LPT-QPN	0.6682	0.5555	<u>0.4893</u>	0.4607	<u>0.4528</u>	<u>0.5253</u>	0.7841	0.6776	<u>0.6115</u>	<u>0.5643</u>	<u>0.5532</u>	<u>0.6381</u>
CSTN-PN	0.6778	0.5698	0.4945	0.4723	0.4667	0.5362	0.7954	0.6937	0.6173	0.5780	0.5658	0.6500

相比之下，CSTN-PN 在高强度降水区域的捕捉上表现优异，其预测结果清晰且接近真实降水分布。

尤其是在第 60 分钟的时候, CSTN-PN 生成的回波图虽然和真实值存在一定差异, 但较其他模型能更准确地捕捉高强度区域降水, 且降水边界清晰, 低强度虚假回波较少, 这表明我们所提出的模型能够更好地预测更长时间的结果。

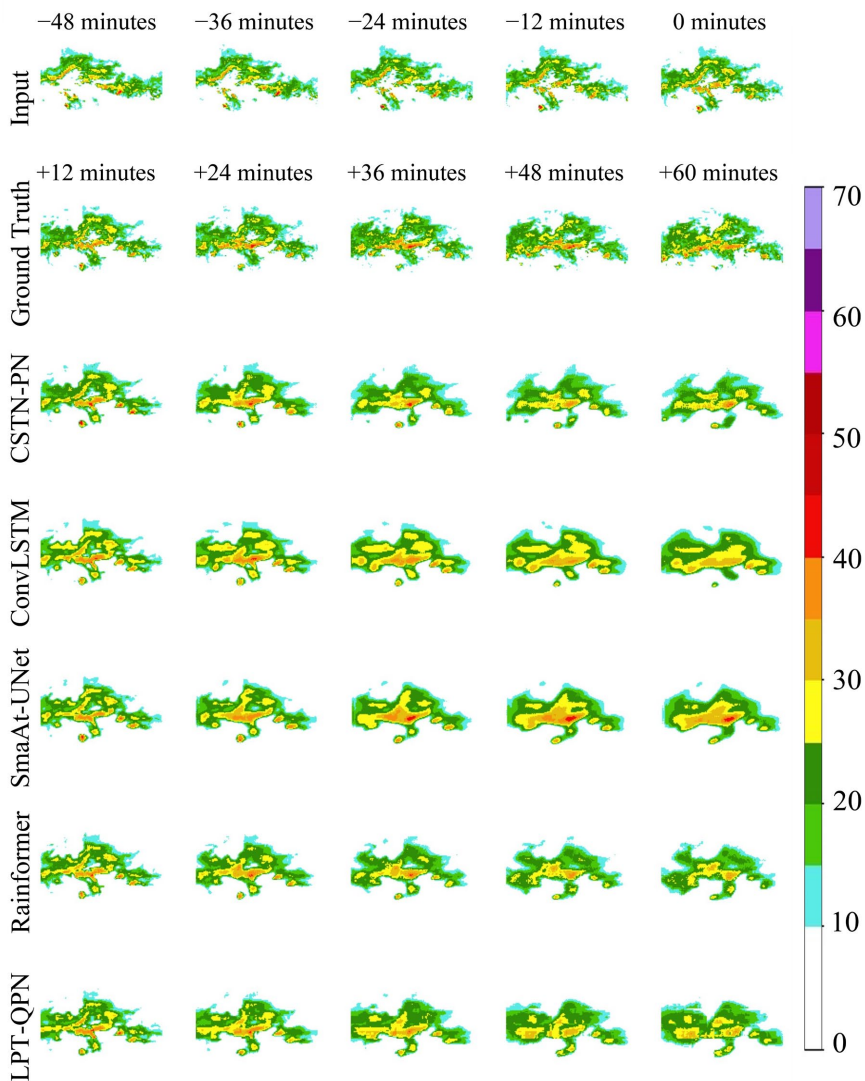


Figure 3. Visual comparison of all models

图 3. 所有模型的视觉对比

为了量化图 3 中的预测结果, 表 3 给出了本样例在第 12、24、36、48 和 60 分钟的 CSI 和 HSS 值。从表 3 中可以看出, CSTN-PN 在各个时间段的 CSI 和 HSS 值均优于其他模型, 尤其是在第 60 分钟的预测中, 其 CSI 和 HSS 值分别比排名第二的模型高出 3.1% 和 2.3%。这一量化指标进一步证明了 CSTN-PN 在长时间步的预测中表现优异, 同时验证了模型的稳定性和有效性。

5. 结论与展望

在本研究中, 我们提出了一种名为 CSTN-PN 的模型用于提高临近预报的准确性。该模型采用编码器-解码器结构, 编码器和解码器均包含四个阶段。编码器通过逐步减少特征图的尺寸并增加通道数, 实

现了高效的特征提取；解码器则通过多层次特征的拼接与逐步还原操作，充分保留并利用了关键特征信息，使模型在雷达回波预测任务中表现出色。每个阶段均采用基于交叉型窗口的注意力机制捕获全局特征，并通过 EMA 挖掘局部多尺度信息，最后通过门控机制实现特征的深度融合。在上海数据集中的实验表明，这种全局与局部特征的科学结合显著提升了模型的降水预测能力。

虽然 CSTN-PN 这种基于数据驱动的方法能够学习到复杂的非线性关系，但模型未能充分考虑到气象数据中的不确定性和极端降水情况，同时也缺乏物理约束和(模型)可解释性。未来，我们计划引入随机时空模型和变分推断方法[30]，捕捉天气雷达回波分布和极端降水情况，从而提升预测的全面性和可靠性。此外，我们还将结合理论驱动数值方法，使模型兼具物理严谨性和数据驱动的灵活性，进一步增强其可解释性和适应性，为临近预报任务提供更准确、可靠的解决方案。

基金项目

四川省自然科学基金项目(No. 2023NSFSC0245)；成都信息工程大学大学生创新训练项目(No. 202310621073)。

参考文献

- [1] 俞小鼎, 郑永光. 中国当代强对流天气研究与业务进展[J]. 气象学报, 2020, 78(3): 391-418.
- [2] Kimura, R. (2002) Numerical Weather Prediction. *Journal of Wind Engineering and Industrial Aerodynamics*, **90**, 1403-1414. [https://doi.org/10.1016/s0167-6105\(02\)00261-1](https://doi.org/10.1016/s0167-6105(02)00261-1)
- [3] Li, P.W. and Lai, E.S.T. (2004) Applications of Radar-Based Nowcasting Techniques for Mesoscale Weather Forecasting in Hong Kong. *Meteorological Applications*, **11**, 253-264. <https://doi.org/10.1017/s1350482704001331>
- [4] Sun, J., Xue, M., Wilson, J.W., Zawadzki, I., Ballard, S.P., Onvlee-Hooimeyer, 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>
- [5] Wilson, J.W., Crook, N.A., Mueller, C.K., Sun, J. and Dixon, M. (1998) Nowcasting Thunderstorms: A Status Report. *Bulletin of the American Meteorological Society*, **79**, 2079-2099. [https://doi.org/10.1175/1520-0477\(1998\)079<2079:ntasr>2.0.co;2](https://doi.org/10.1175/1520-0477(1998)079<2079:ntasr>2.0.co;2)
- [6] Ayzel, G., Heistermann, M. and Winterrath, T. (2019) Optical Flow Models as an Open Benchmark for Radar-Based Precipitation Nowcasting (Rainymotion V0.1). *Geoscientific Model Development*, **12**, 1387-1402. <https://doi.org/10.5194/gmd-12-1387-2019>
- [7] Prudden, R., Adams, S., Kangin, D., Robinson, N., et al. (2020) A Review of Radar-Based Nowcasting of Precipitation and Applicable Machine Learning Techniques. arXiv: 2005.04988. <https://doi.org/10.48550/arXiv.2005.04988>
- [8] LeCun, Y., Bengio, Y. and Hinton, G. (2015) Deep Learning. *Nature*, **521**, 436-444. <https://doi.org/10.1038/nature14539>
- [9] Shi, X., Chen, Z., Wang, H., Yeung, D.Y., et al. (Year). Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. *Annual Conference on Neural Information Processing Systems (NIPS 2015)*, Montreal, 7-12 December 2015, 802-810. <https://proceedings.neurips.cc/paper/2015/hash/07563a3fe3bbe7e3ba84431ad9d055af-Abstract.html>
- [10] Graves, A. (2012) Long Short-Term Memory. In: Graves, A., Ed., *Supervised Sequence Labelling with Recurrent Neural Networks*, Springer, 37-45. https://doi.org/10.1007/978-3-642-24797-2_4
- [11] Shi, X., Gao, Z., Lausen, L., Wang, H., et al. (Year) Deep Learning for Precipitation Nowcasting: A Benchmark and a New Model. *Annual Conference on Neural Information Processing Systems (NIPS 2017)*, California, 4-9 December 2017, 5618-5628. <https://proceedings.neurips.cc/paper/2017/hash/a6db4ed04f1621a119799fd3d7545d3d-Abstract.html>
- [12] Wang, Y., Long, M., Wang, J., Gao, Z., et al. (Year) PredRNN: Recurrent Neural Networks for Predictive Learning Using Spatiotemporal LSTMs. *Annual Conference on Neural Information Processing Systems (NIPS 2017)*, California, 4-9 December 2017, 880-889. <https://proceedings.neurips.cc/paper/2017/hash/e5f6ad6ce374177eef023bf5d0c018b6-Abstract.html>
- [13] 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>

- [14] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W. and Frangi, A., Eds., *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, Springer, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [15] Agrawal, S., Barrington, L., Bromberg, C., Burge, J., *et al.* (2019). Machine Learning for Precipitation Nowcasting from Radar Images. arXiv: 1912.12132. <https://doi.org/10.48550/arXiv.1912.12132>
- [16] Trebing, K., Stańczyk, T. and Mehrkanoon, 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>
- [17] 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.
- [18] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., *et al.* (2021). An Image Is Worth 16 x 16 Words: Transformers for Image Recognition at Scale. *International Conference on Learning Representations (ICLR)*, Vienna, 4 May 2021.
- [19] Xie, E., Wang, W., Yu, Z., Anandkumar, A., *et al.* (Year) SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. *Conference on Neural Information Processing Systems (NeurIPS 2021)*, 6-14 December 2021, 12077-12090. <https://proceedings.neurips.cc/paper/2021/hash/64f1f27bfb4ec22924fd0acb550c235-Abstract.html>
- [20] Chitta, K., Prakash, A., Jaeger, B., Yu, Z., Renz, K. and Geiger, A. (2023) Transfuser: Imitation with Transformer-Based Sensor Fusion for Autonomous Driving. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **45**, 12878-12895. <https://doi.org/10.1109/tpami.2022.3200245>
- [21] Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., *et al.* (2022) UNETR: Transformers for 3D Medical Image Segmentation. 2022 *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, 4-8 January 2022, 1748-1758. <https://doi.org/10.1109/wacv51458.2022.00181>
- [22] 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>
- [23] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., *et al.* (2021) Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 9992-10002. <https://doi.org/10.1109/iccv48922.2021.00986>
- [24] 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>
- [25] 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>
- [26] Dong, X., Bao, J., Chen, D., Zhang, W., Yu, N., Yuan, L., *et al.* (2022) CSWin Transformer: A General Vision Transformer Backbone with Cross-Shaped Windows. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 12114-12124. <https://doi.org/10.1109/cvpr52688.2022.01181>
- [27] Ouyang, D., He, S., Zhang, G., Luo, M., Guo, H., Zhan, J., *et al.* (2023) Efficient Multi-Scale Attention Module with Cross-Spatial Learning. *ICASSP 2023—2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes Island, 4-10 June 2023, 1-5. <https://doi.org/10.1109/icassp49357.2023.10096516>
- [28] Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., *et al.* (2014) Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, 25-29 October 2014, 1724-1734. <https://doi.org/10.3115/v1/d14-1179>
- [29] Chen, L., Cao, Y., Ma, L. and Zhang, J. (2020) A Deep Learning-based Methodology for Precipitation Nowcasting with Radar. *Earth and Space Science*, **7**, e2019EA000812. <https://doi.org/10.1029/2019ea000812>
- [30] Denton, E. and Fergus, R. (2018) Stochastic Video Generation with a Learned Prior. 35th *International Conference on Machine Learning (ICML 2018)*, Stockholm, 10-15 July 2018, 1906-1919. <https://proceedings.mlr.press/v80/denton18a.html>