

# 基于隐式自回归时空通道聚合策略的多输入单输出云图预测算法<sup>\*</sup>

吴禹乾<sup>1</sup> 张秀再<sup>2,3</sup> 李景轩<sup>3</sup>

(1. 南京信息工程大学长望学院 南京 210044; 2. 南京信息工程大学江苏省大气环境与装备技术协同创新中心 南京 210044; 3. 南京信息工程大学电子与信息工程学院 南京 210044)

**摘 要:** 云图预测的本质是时空序列预测。基于深度学习的时空序列预测算法可以归纳为三种框架,即 SISO、MIMO 和 MISO。针对云图运动特点,在 MISO 框架下结合 MIMO 与 SISO 两种框架特性设计了一种基于隐式自回归时空通道聚合策略的云图预测算法(IASCACP)。针对 MIMO 类模型存在的图像关联性缺失以及 SISO 类模型存在的误差累积过量的问题,引入一种隐式自回归编解码器,利用隐式自回归结构为模型赋予一定的递归特性以捕获序列中的时空关联信息,再用掩码与真实映射模块减少递归带来的误差累积,增强模型鲁棒性;针对云运动不稳定与非线性等问题,设计了一种时空通道聚合预测器,该模块可以有效聚合多阶时空相互作用信息,并对其进行自适应通道再分配以减少特征冗余。将该算法分别在 MovingMNIST 数据集及 FY-4A 卫星云图数据集上进行实验,实验结果表明,该算法有效改进了 MIMO 与 SISO 类模型的缺点,预测精度高于其他模型,体现了该模型在云图预测领域的可靠性与有效性。

**关键词:** 卫星云图预测;风云四号 A 星;隐式自回归;多输入单输出框架  
**中图分类号:** TP751.1;TN0 **文献标识码:** A **国家标准学科分类代码:** 520.2

## Multiple-in-single-out cloud image prediction algorithm based on implicit autoregressive spatiotemporal channel aggregation strategy

Wu Yuqian<sup>1</sup> Zhang Xiuzai<sup>2,3</sup> Li Jingxuan<sup>3</sup>

(1. Changwang School of Honors, Nanjing University of Information Science & Technology, Nanjing 210044, China; 2. Jiangsu Province Atmospheric Environment and Equipment Technology Collaborative Innovation Center, Nanjing University of Information Science & Technology, Nanjing 210044, China; 3. School of Electronic and Information Engineering, Nanjing University of Information Science & Technology, Nanjing 210044, China)

**Abstract:** The essence of cloud image prediction is spatiotemporal sequence prediction. Deep learning-based spatiotemporal sequence prediction algorithms can be categorized into three frameworks: SISO, MIMO and MISO. Based on the characteristics of cloud image movement, designed a cloud image prediction algorithm under the MISO framework, combining the features of both MIMO and SISO frameworks, called the implicit autoregressive spatiotemporal channel aggregation prediction (IASCACP) algorithm. To address the issue of image correlation loss in MIMO models and excessive error accumulation in SISO models, introduce an implicit autoregressive encoder-decoder. This encoder-decoder endows the model with certain recursive properties through an implicit autoregressive structure to capture spatiotemporal correlation information in sequences. Additionally, a masking and true mapping module is used to reduce error accumulation from recursion and enhance model robustness. To tackle issues such as instability and nonlinearity in cloud movement, we designed a spatiotemporal channel aggregation predictor. This module effectively aggregates multi-order spatiotemporal interaction information and performs adaptive channel reallocation to reduce feature redundancy. The algorithm was tested on the MovingMNIST dataset and the FY-4A satellite cloud image dataset. Experimental results show that this algorithm effectively improves the shortcomings of MIMO and SISO models and achieves higher prediction accuracy compared to other models, demonstrating its reliability and effectiveness in the field of cloud image prediction.

**Keywords:** satellite cloud image prediction; Fengyun-4A satellite; implicit autoregression; multiple-in-single-out framework

## 0 引 言

卫星云图预测是一种利用卫星获取的云图数据来预测

天气和气候变化的方法。云图预测的本质是时空序列预测,当前基于深度学习的时空预测算法可以分成 3 类框架,即单输入单输出框架(single in single out, SISO)、多输入

收稿日期:2024-11-24

<sup>\*</sup> 基金项目:第二次青藏高原综合科学考察研究项目(2019QZKK0105)、国家社会科学基金一般项目(22BZZ080)、国家级大学生创新创业训练计划支持项目(202410300088Z)资助

多输出框架(multiple in multiple out, MIMO)和多输入单输出框架(multiple in single out, MISO)。

SISO 框架类预测模型主要利用循环神经网络及其变体长短期记忆网络(long short term memory, LSTM)<sup>[1]</sup>、门控循环单元(gated recurrent unit, GRU)<sup>[2]</sup>为模型赋予递归特性,实现序列预测。2014 年 Ranzato 等<sup>[3]</sup>将循环神经网络(recurrent neural networks, RNN)应用到时空序列预测领域并提出一种利用视频数据进行无监督特征学习的强基线预测模型。Srivastava 等<sup>[4]</sup>将 LSTM 引入时空序列预测领域,使用 LSTM 分别作为预测模型的编码器与解码器实现序列预测,并点明了卷积神经网络(convolutional neural network, CNN)与 LSTM 结合以提升预测精准度的可行性。Shi 等<sup>[5]</sup>参考 Srivastava 的思想将 CNN 与 LSTM 结合提出了适用于图像序列预测的 ConvLSTM,根据网络中信息流的传输设计了一种编解码预测结构(encoding-forecasting, EF),将该预测模型用于降水预测取得了不错的效果。之后 Shi 等<sup>[6]</sup>进一步优化了 EF 预测结构,提出了较 ConvLSTM 更加轻量化的 ConvGRU,并引入可学习卷积提出了 TrajectoryGRU,该算法理论上更适于预测降水等具有无规则运动特性的数据。2017~2019 年王云波团队在 ConvLSTM 的基础上融入了 3D 卷积、时空记忆单元、梯度高速公路等,先后提出了 PredRNN<sup>[7]</sup>、PredRNN++<sup>[8]</sup>以及 E3D-LSTM<sup>[9]</sup>为时空序列预测技术的发展做出了重大贡献。为解决序列预测时内存占用过大,且序列特征信息易流失的问题。2020 年, Yu 等<sup>[10]</sup>设计了一种条件可逆预测算法 CrevNet,该算法在编解码及预测阶段均使用可逆循环框架,能够做到减少特征提取阶段信息流失的同时提高模型计算效率及预测精度。2023 年,基于 CrevNet 的 SmartCrevNet 被提出并应用于云图预测,表现出更为有效的预测效果<sup>[11]</sup>。以上 SISO 类预测模型主要以当前帧为基础预测下一帧,存在着无法避免的递归误差累积的问题,为解决该问题,研究人员也在探索 CNN 在时空序列预测领域的运用,并逐渐形成了 MIMO 预测框架。

MIMO 类预测算法将整个序列帧全部输入进编码器中,经过预测模块转换后,同时解码所有帧,打破了误差累积链。如 Liu 等<sup>[12]</sup>设计 DVF 模型,通过 CNN 自编码器学习体素流,通过从附近帧借用体素来重建预测帧。2018 年 Xu 等<sup>[13]</sup>提出了 PredCNN,将级联乘法单元与卷积神经网络结合以捕获帧间相关性。MIMO 类模型不依赖前一帧的预测方式,解决了误差累积的问题,但是此类模型不含有自回归特性,预测的每帧之间缺乏时空关联性。2022 年 Gao 等<sup>[14]</sup>在 MIMO 框架下设计了一种纯卷积预测网络 SimVP,其预测器中利用组卷积的并联来获取不同尺度下的时空特征以此来加强预测图像间的关联性,一定程度上弥补了 MIMO 类模型的缺陷。

SISO 类预测算法存在误差堆积问题,而 MIMO 类预测算法存在预测图像缺乏关联性的问题。为了解决以上问

题, Seo 等<sup>[15]</sup>综合两种框架特点,在 2023 年介绍了 MISO 预测框架的两种预测模式,并提出了基于 MISO 框架的预测模型 IAM4VP,该算法将自回归的思想运用到卷积网络中以解决图像关联性缺失的问题。

MISO 预测框架结合了 SISO 与 MIMO 框架的优点,利用多帧图像预测单帧图像,并将单帧预测图像堆叠至输入端参与下一帧的预测。该框架在一定程度上可以解决 SISO 框架存在的误差累积问题以及 MIMO 框架存在的图像关联性缺失问题。针对云图运动特点以及 MISO 框架特性,以 MISO 为基础框架,基于 IAM4VP 提出了一种基于隐式自回归时空通道聚合策略的卫星云图预测算法(implicit autoregressive spatiotemporal channel aggregation algorithm for cloud image prediction, IASCACP)。具体工作如下:

1) 自回归算法有时空关联性但是存在误差累积,非自回归算法没有误差累积,但是缺乏时空关联性。针对该问题引入一种隐式自回归编解码器,将预测的单帧图像堆叠至输入队列参与下一帧的预测,为模型赋予自回归特性,同时在编码阶段加入掩码模块,通过学习,屏蔽未来帧递归带来的误差信息,增加模型鲁棒性。

2) 对于云图运动预测这种特征复杂度高且运动趋势非线性的时空序列预测任务,在预测阶段应当更加关注全局时空上下文信息而非局部信息。针对这一问题在预测模块中引入了时空通道聚合器(spatiotemporal channel aggregator, SCA),其中时空聚合器(spatiotemporal aggregator, SA)拥有平衡多阶特征相互作用的功能,可以有有效的聚集低阶局部特征及中高阶上下文聚合特征。通道聚合器(channel aggregator, CA)用来对时空聚合器输出的多阶特征进行自适应信道特征再分配以减少信道特征冗余,增加多阶特征的有效性,继而增加预测精度。

## 1 相关理论基础

### 1.1 卫星云图预测基本原理

在基于深度学习的云图预测领域中,一般利用神经网络根据之前观测的几帧时序云图来预测后面几帧时序云图。若观测的每帧时序图都被记录在  $H \times M$  大小的网格中,则每帧时序云图可以用一个张量  $x \in R^{H \times W}$  来表示。将观测的时序图像帧表示为  $x_1, x_2, x_3, \dots$ , 时空序列预测问题就是利用时间步上连续观察到的  $m$  帧时序图来预测后  $n$  帧时序图,表示为:

$$\begin{aligned} & \widetilde{x}_{t+1}, \dots, \widetilde{x}_{t+n} = \\ & \underset{x_{t+1}, \dots, x_{t+n}}{\operatorname{argmax}} p(x_{t+1}, \dots, x_{t+n} \mid x_{t-m+1}, \dots, x_t) \end{aligned} \quad (1)$$

### 1.2 三种预测框架

#### 1) 单输入单输出预测框架

基于 SISO 框架的预测模型,其实是递归模型,即通过学习历史序列信息的隐藏状态,以当前帧为主体来生成

下一帧。因此,当序列前期的预测帧产生细小错误时,就会产生蝴蝶效应,随着生成帧的延长,其图像帧的质量和准确性将快速下降。SISO 框架如图 1 所示。

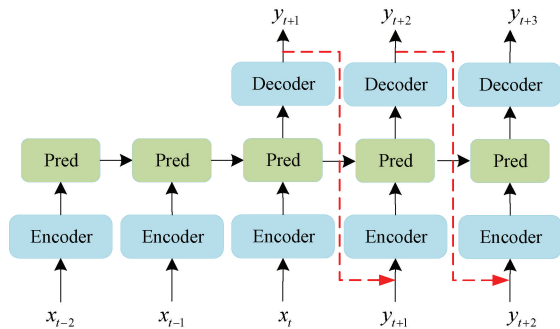


图 1 SISO 预测框架

Fig. 1 SISO prediction framework

### 2) 多输入多输出预测框架

基于 MIMO 框架的模型将所有序列帧都输入编码器中,经过预测模块学习时空信息后,再使用解码器同时解码所有帧。与 SISO 框架相比,该类框架模型的预测结果不依赖于前一次的预测结果,并不会出现误差累积。但需要注意 MIMO 模型并不具有递归性,一次性解码输出所有帧的方法使得此类模型预测的帧与帧之间缺乏相关性。MIMO 预测框架如图 2 所示。

### 3) 多输入单输出预测框架

SISO 类预测模型有着严重的误差累积问题,MIMO 类预测模型预测图像间缺少时空关联性,而多输入单输出预测框架 MISO 综合了以上两种框架的优点。

MISO 预测框架如图 3 所示,输入的多帧序列  $x_{t-2}, x_{t-1}, x_t$  同时进入编解码预测模型并只预测一帧图像  $y_{t+1}$ ,再将预测得到的  $y_{t+1}$  与  $x_{t-2}, x_{t-1}, x_t$  组合输入进编解码预测模型中得到  $y_{t+2}$ 。此类框架综合了 SISO 与 MIMO 的优点,用输入多帧生成单帧的思想缓解了 SISO 框架下的误差累积问题,而输入帧与预测帧组合输入预测器再次预测

的手段,为模型增加了递归的特性,弥补了 MIMO 类模型所缺乏的时空关联性。

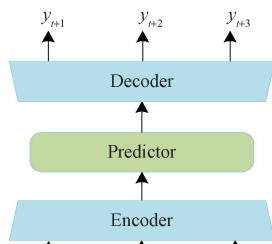


图 2 MIMO 预测框架

Fig. 2 MIMO prediction framework

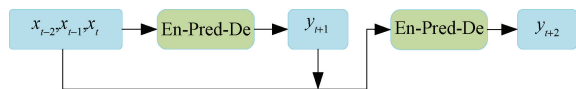


图 3 MISO 预测框架

Fig. 3 MISO prediction framework

## 2 模型设计

### 2.1 IASCACP 网络模型

IASCACP 网络模型基于 MISO 框架采用隐式自回归策略,存在一定的递归特性。模型为了削弱递归带来的误差累积,在编码器后加入掩码模块并在预测输出端加入真实映射模块(true mapping, TM)。同时还设计了一种时空通道聚合预测器,使预测模块选择性地关注云图序列中高阶上下文时空特征并与低阶局部特征进行聚合,以全面学习时空运动趋势。IASCACP 框架如图 4 所示。

IASCACP 模型采用编码器-预测器-解码器结构。设输入编码器中长度为  $n$  观测序列图像为  $x_{t-n+1:t} \in \mathbb{R}^{T \times C \times H \times W}$ ,其中  $T, C, H$  和  $W$  分别表示序列图像的帧长、通道数、高和宽。若设当前回环帧序列为  $\tilde{x}_{t+1:t+k}$ ,则预测

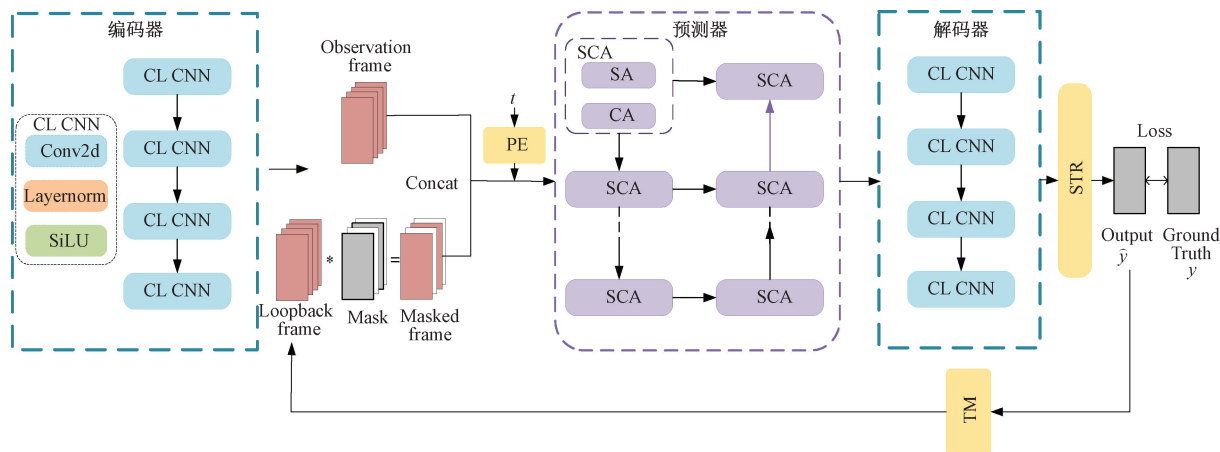


图 4 IASCACP 框架

Fig. 4 IASCACP framework

帧为  $\tilde{x}_{t+k+1}$ 。IASCACP 预测流程表示为:

$$\tilde{x}_{t+k+1} = D \left( P \left( S(t), \text{Concat}(E(x_{t-n+1:t}), \text{Mask}(\tilde{x}_{t+1:t+k})) \right) \right) \quad (2)$$

其中,  $E$ 、 $P$  和  $D$  分别表示编码器, 预测器和解码器。 $S$  为位置编码函数,  $\text{Mask}$  为掩码器,  $\text{Concat}$  为堆叠操作。

## 2.2 隐式自回归编解码器

结合 MIMO 与 SISO 框架特点, 引入一种隐式自回归编解码器。隐式自回归编解码器结构如图 5 所示。

由图 5 可知, 策略中的帧序列分为观察帧序列(observation frames)、回环帧序列(loopback frames)与掩码帧序列(masked frames), 观察帧序列即为输入序列经过编码器得到的结果, 其步长随时间的增加而保持不变, 始

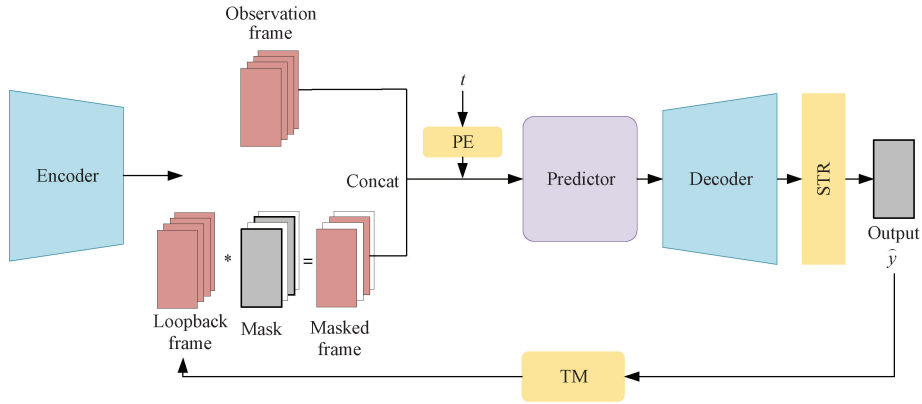


图 5 隐式自回归编解码器

Fig. 5 Implicit autoregressive encoder-decoder

隐式自回归编解码器的核心是一种掩码模块<sup>[16]</sup>, 该模块可以在训练阶段随机屏蔽回环帧序列中的冗余信息, 提供不完整的上下文时空序列信息, 继而增加模型的鲁棒性, 减少误差堆叠对模型的影响。模型训练阶段的操作流程如算法 1 所示。

### 算法 1: 隐式自回归训练策略

输入: 输入帧  $x$ , 真实未来帧  $y$ , 目标预测帧长度  $\tilde{T}$ , 编码器  $E(\cdot)$ , 预测器  $P(\cdot)$ , 解码器  $D(\cdot)$ , 观测帧序列  $Q_o$ , 回环帧序列  $Q_f$ , 位置编码器  $S(\cdot)$ 。

输出: 预测值  $\hat{y}_t$

For  $x_i, y_i$  in DataSet;

$t = \text{random}(\tilde{T}) t \in \tilde{T}$

$Q_o, Q_f = E(x_i), E(y_i)$  生成观测帧与回环帧序列

$\text{Mask} = M_g(t)$  根据  $t$  生成掩码器

$Q_m = \text{Mask} \times Q_f$  生成掩码序列

$\hat{y}_t = D(P(\text{Concat}(Q_o, Q_m), S(t)))$  得到预测值

End for

终具有相同的编码值。回环帧初始全填充为零张量, 随着时间步长的推进, 预测帧逐渐输出, 经过 TM 后填充到回环帧序列中。回环帧序列再经过掩码器得到掩码帧序列。位置编码模块(positional encoding, PE)用来建模特征中的位置关系并增强序列间的时空相关性。掩码帧序列与观测帧序列经过堆叠后与 PE 生成的数据组合输出到预测模块学习时空趋势变化, 再利用解码器解码得到当前预测帧。解码器后引入一个时空细化模块(spatio-temporal refinement, STR)<sup>[15]</sup>对解码帧进行细化。预测帧与标签帧计算损失的同时, 通过 TM 模块转换为更符合真实特征的回环帧序列, 回环帧序列再经过掩码后与输入帧一起来预测下一帧。模型训练时为了减少训练成本, 在训练阶段将真实的未来帧填充进回环帧序列, 在验证阶段将预测帧填充进回环帧序列。

掩码器函数表示为:

$$M_g(t) = \begin{cases} \text{Random}(t), & t > \text{Index}(Q_f) \\ 0, & t \leq \text{Index}(Q_f) \end{cases} \quad (3)$$

虽然掩码模块对回环帧序列的随机屏蔽可以增强网络的鲁棒性, 但是直接从预测帧中提取的特征向量并不能反应真实未来帧的特征向量, 在模型验证及测试阶段预测误差可能会随着预测的深入而恶化。为了解决该问题, 在预测帧输出端口引入了真实映射模块 TM, 该模块可以对预测帧投影特征向量, 使其更接近实际未来帧中的特征向量。在实际代码中 TM 模块与编码模块的结构相同。

## 2.3 时空通道聚合预测器

预测器使用  $N_t$  个 SCA 堆叠学习云图序列的动态变化趋势。SCA 由时空聚合器 SA 与通道聚合器 CA 构成。具有局部感知或者上下文聚合特性的传统神经网络倾向于极序交互, 即仅处理低阶局部信息或者高阶全局信息, 而忽视了最具差异性的中阶交互信息<sup>[17-19]</sup>。为了有效捕获具有多阶平衡性交互的上下文信息, 引入了一个时空聚合模块 SA, 其结构如图 6 所示。

图 6 中 GAP 为全局平均池化(global average pooling,



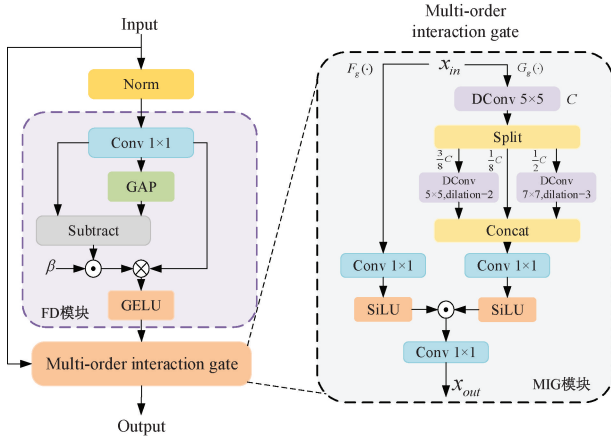


图 6 时空聚合器结构

Fig. 6 Implicit autoregressive encoder-decoder

GAP), Subtract 表示两值相减。SA 主要由特征分解模块 (feature decomposition, FD) 与多阶交互门 (multi-order interaction gate, MIG) 构成。SA 作为一个纯卷积结构, 能够提取多阶特征, 使网络同时具有静态及自适应局部感知, 其公式化表示为:

$$SA(X) = X + MIG(FD(Norm(X))) \quad (4)$$

FD 模块可以动态地剔除琐碎的特征交互, 利用  $Conv_{1 \times 1}(\cdot)$  与  $GAP(\cdot)$  分别获得普通局部纹理的低阶相互作用以及覆盖全局复杂形状的高阶相互作用。FD 模块公式化表示为:

$$Y = Conv_{1 \times 1}(X) \quad (5)$$

$$Z = GELU(Y + \beta \odot (Y - GAP(Y)))$$

其中,  $\beta \in R^{C \times 1}$  为初始值为零的缩放因子, 用来对  $Y - GAP(Y)$  重新加权互补相互作用分量。

此外, MIG 模块包含两个分支, 分别为选通门分支  $F_g(\cdot)$  和上下文分支  $G_g(\cdot)$ 。在 MIG 模块的上下文分支中集成了多个深度卷积 DConv 来编码多阶时空特征。使用 3 个尺寸的 DConv, 其膨胀比  $d \in \{1, 2, 3\}$ , 并行捕获低、中、高阶相互作用特征。给定 MIG 模块输入特征  $X_{in} \in R^{C \times HW}$ , 经过  $DConv_{5 \times 5, d=1}$  获取低阶相互作用信息后, 输出特征沿着通道维度分解为 3 组输出  $X_l \in R^{C_l \times HW}$ ,  $X_m \in R^{C_m \times HW}$ ,  $X_h \in R^{C_h \times HW}$ , 其中  $C_l + C_m + C_h = C$ 。  $X_m, X_h$  分别输入给  $DConv_{5 \times 5, d=2}, DConv_{7 \times 7, d=3}$  编码中阶及高阶上下文信息得到  $Y_m, Y_h$ 。最后  $X_l, Y_m, Y_h$  拼接形成多阶上下文特征  $Y_c = Concat(X_l, Y_m, Y_h)$ 。为了聚合多阶上下文特征, 分别在选通门分支与上下文分支的输出端加入 SiLU 激活函数, 该函数具有 Sigmoid 函数的门控特性, 同时训练较为稳定。

综上, MIG 模块公式化如下:

$$MIG(X) = Conv_{1 \times 1}(SiLU(Concat(Conv_{1 \times 1}(X), SiLU(Concat(Conv_{1 \times 1}(Y_c)))))) \quad (6)$$

在 SA 模块输出端引入了一个通道聚合器 CA, 用于在

高维隐藏空间中进行自适应通道再分配, 消除信道中的信息冗余。CA 结构如图 7 所示。

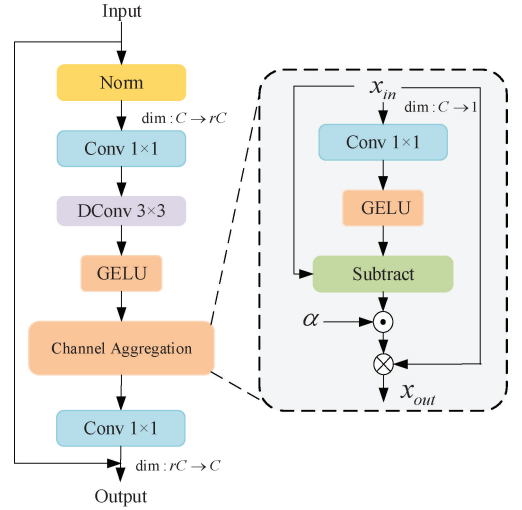


图 7 通道聚合器结构

Fig. 7 Channel aggregator structure

整体结构公式为:

$$Y = GELU(DConv_{3 \times 3}(Conv_{1 \times 1}(Norm(X)))) \quad (7)$$

$$Z = Conv_{1 \times 1}(CA(Y)) + X \quad (8)$$

对其中的核心  $CA(\cdot)$ , 利用一条分支来减少通道投影  $X_{in} \in R^{C \times HW} \rightarrow X_r \in R^{1 \times HW}$  以收集和重新分配通道信息。  $CA(\cdot)$  公式为:

$$CA(X) = X + \alpha \odot (X - GELU(X_r)) \quad (9)$$

其中,  $\alpha$  为初始值为零的通道尺度因子, 用于对  $X - GELU(X_r)$  再分配互补通道相互作用信号。

云图的运动具有复杂性, SA 模块的引入可以使预测模型充分获取时序云图的低、中、高阶特征交互, 多维度的学习云图时空运动趋势。CA 模块的引入对于 SA 模块学习到的多阶时空交互信息起到了筛选作用, 消除了信道中的冗余信息, 增强了预测模块的学习效率, 提高了最终的预测精度。

### 3 实验与分析

#### 3.1 实验环境及细节

将 IASCACP 在 FY-4A 卫星云图数据集和 MovingMNIST 数据集进行实验。IASCACP 中编解码模块 CL-CNN 层数设置为 4, 译码器中 SCA 堆叠层数为 5, 初始学习率为 0.000 1。所有模型均使用 L1+L2 损失函数以及 Adam<sup>[20]</sup> 优化器。实验硬件条件为 NVIDIA GeForce RTX 3060 12 G GPU, Intel Core i5-12400F CPU, 32 G 内存。

#### 3.2 FY-4A 云图数据集实验

长三角地区是中国经济最有活力的区域之一, 同时也是气象灾害最频繁的区域之一, 长三角地区的夏季受太阳辐射强, 地面温度高, 大气不稳定, 容易形成对流性云, 而

随着温度的升高,对流层中上升气流的增长,容易形成雷暴、龙卷风等强对流天气。相比于其他季节,长三角地区夏季的云图运动速度更快,云图的不稳定、非线性的特点更显著,所以,针对夏季长三角地区上空云图开展预测研究具有重要代表性、挑战性。FY-4A 是中国第二代静止轨道气象卫星,其搭载了 14 个辐射成像通道,能够覆盖可见光、近红外、中红外及远红外等多个波段。为获取云图的全天可观测图像,选取 FY-4A 多通道成像仪 L1 级数据中的长波红外通道的数据进行云图预测实验。

选取 2021 年夏季 6~10 月长三角地区云图样本,时间间隔为 1 h,在筛选掉坏图与错图后得到 3 488 张云图,按照序列长度为 8 的滑动取值,共获得 3 481 个云图序列,随机选择 2 780 个序列作为训练集,701 个序列作为验证集。云图样本的经纬度范围为东经 115°2′E ~123°5′E,北纬 26°6′N~35°2′N,像素大小为 256 pixel×256 pixel,空间分辨率为 4 km,STA-GRU 中的  $\tau$  设置为 4。经实验评估,兼顾实用性和有效性,该算法对 4 h 的云图预测综合性能最优。更长时间跨度的云图预测将会使预测精度急剧下降,预测结果将不具有实用价值。

另外,为全面分析云图预测结果,实验使用均方误差(MSE)、平均绝对误差(MAE)、结构相似性(SSIM)以及峰值信噪比(PSNR)<sup>[21]</sup>对预测图像进行评估。SSIM 是衡量两张图片相似度的指标,范围在-1~1,值越接近 1 说明两张图片越相似,PSNR 是一种衡量图像质量的指标(常用单位为 dB),其值越高预测云图质量越好。

1)云图数据集消融实验结果分析

为了评估 IASCACP 内部各模块的有效性,将 IASCACP 拆分进行消融实验,并使用 MSE 与 SSIM 作为

评价指标,实验结果如表 1 所示。

表 1 IASCACP 模块消融实验  
Table 1 Ablation experiments of IASCACP modules

MASK	PE	STR	TM	SCA	MSE( $10^{-3}$ ) ↓	SSIM ↑
	✓	✓	✓	✓	19.12	0.563
✓		✓	✓	✓	18.46	0.570
✓	✓		✓	✓	18.58	0.567
✓	✓	✓		✓	18.45	0.569
✓	✓	✓	✓	✓	<b>18.15</b>	<b>0.573</b>

首先将 MASK 模块去除,此时回环帧序列将不经过掩码而直接与观测帧序列进行堆叠,然后与 PE 的输出一起送入预测模块,此种状态下模型递归带来的误差累积将会放大,模型的鲁棒性也会减弱,模型的预测精度显著降低。PE 模块作为位置编码模块能够增强特征间的位置关系,弥补 CNN 网络的绝对位置建模能力,当 PE 模块单独剔除后输入到预测模块的时空特征缺乏时空相关性,继而导致预测质量的降低。STR 与 TM 模块分别对网络的输出进行细化以及真实化映射,如表中所示,当去除两种模块后,模型的预测精度也均有不同程度的衰减。以上消融实验的结果证明了模型各个模块的有效性与创新性。

2)云图数据集对比实验结果分析

将 MISO 类模型 IASCACP 分别与 SISO 类新模型 CrevNet、SmartCrevNet 以及 MIMO 类新模型 SimVP、FACP 在 FY-4A 卫星云图数据集进行对比实验以深入探索不同预测框架在云图预测领域的运用,定量评价如表 2 所示。

表 2 多框架预测模型在云图数据集上的定量评价

Table 2 Quantitative evaluation of multi-framework prediction models on cloud image dataset

框架	模型(4→4)	MSE( $10^{-3}$ ) ↓	MAE( $10^{-3}$ ) ↓	SSIM ↑	PSNR ↑
SISO	CrevNet	19.64	72.41	0.482	17.07
	SmartCrevNet	18.21	67.34	0.520	17.40
MIMO	SimVP	20.63	80.37	0.561	16.84
	FACP	19.82	78.55	0.573	17.03
MISO	IASCACP	<b>18.15</b>	<b>66.01</b>	<b>0.573</b>	<b>17.41</b>

表 2 显示了 3 种框架下的 5 种不同预测算法在 FY-4A 卫星数据集上的性能对比。从 MSE、MAE、PSNR 3 项指标来看,SISO 类模型整体优于 MIMO 类模型,而 MIMO 类模型的 SSIM 指标均优于 SISO 类模型,这是因为 SSIM 是从图像的亮度、对比度和结构来判断图片优劣,SISO 类模型的误差堆叠会导致预测图像中噪声信息增多,改变了预测图像的亮度、纹理等信息,进而导致 SSIM 的值相对较低。IASCACP 作为 MISO 类预测模型,结合

了两种框架的特点,相比 SISO 类模型减少了误差堆叠对预测的影响,相比 MIMO 模型引入了递归特性增强了预测数据的时空关联性。IASCACP 在 MSE、MAE、PSNR 指标下均优于其他框架算法,SSIM 指标与 FACP 持平。综合来看,MISO 框架的 IASCACP 模型整体优于 SISO 与 MIMO 框架的预测模型。

IASCACP 与 CrevNet、SmartCrevNet、SimVP、FACP 模型在云图预测方面的定性比较如图 8 所示。

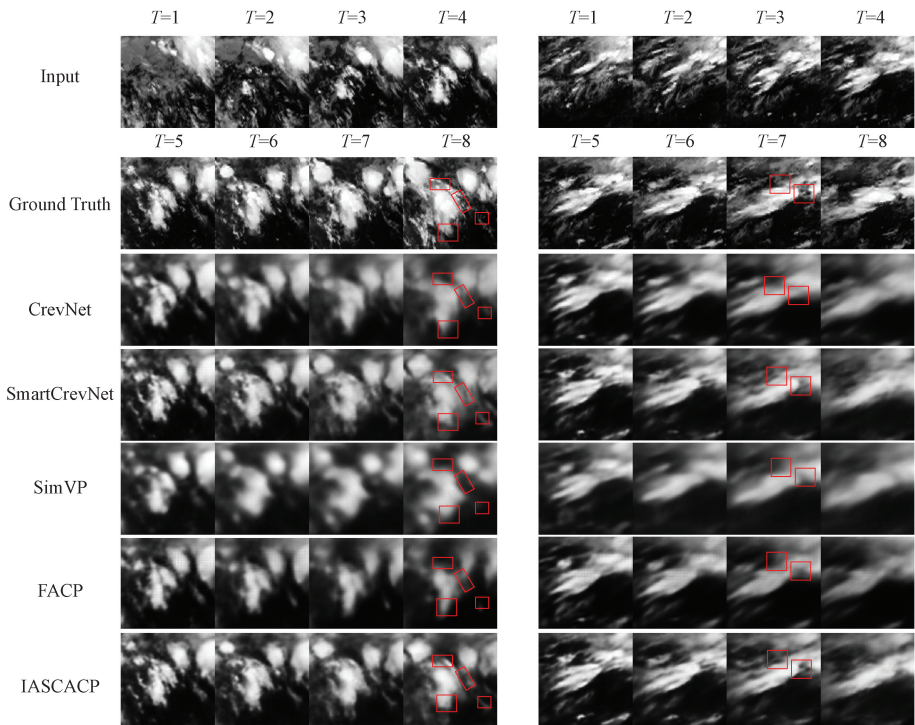


图 8 多框架预测模型云图预测定性对比

Fig. 8 Qualitative comparison of cloud image prediction among multi-framework prediction models

由图 8 可知, SISO 框架下的 CrevNet 与 SmartCrevNet 均较为准确的预测到了云图轨迹。示例中 SimVP 的预测精度在第 2 帧便急剧下降, 之后的运动趋势与实际趋势相差较大, FACP 其双路特征提取模块的存在增强了序列中的局部依赖与全局依赖, 对 MIMO 框架存在的问题进行了针对性的改进, 相比 SimVP, FACP 在第 2 帧之后预测精度的衰减较为缓慢。整体来看

SmartCrevNet 与 IASCACP 在示例中的预测效果更好。而 IASCACP 预测后两帧中对不同碎云的解耦化处理相较 SmartCrevNet 更为精确。

3 种框架模型在云图数据集上预测的逐帧对比如图 9 所示, 可以看到 MSE 评价标准下 SmartCrevNet 与 IASCACP 是起伏最小的, 而 SSIM 标准下 FACP 与 IASCACP 是衰减最为缓慢的。

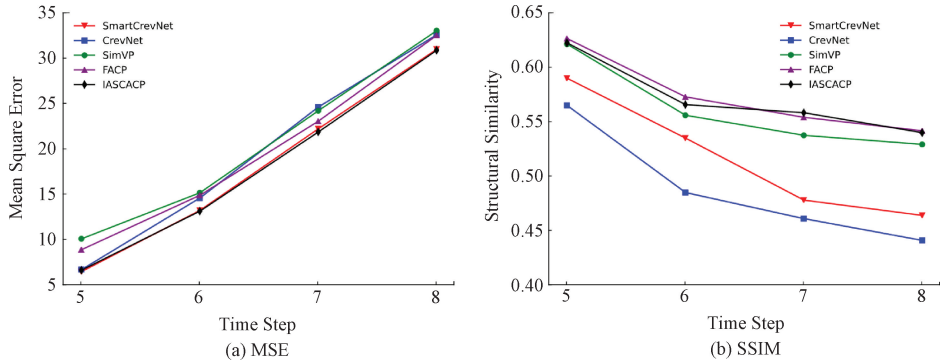


图 9 多框架预测模型 MSE 与 SSIM 的预测逐帧比较

Fig. 9 Frame-by-frame comparison of MSE and SSIM for multi-framework prediction models

3.3 MovingMNIST 数据集实验

MovingMNIST 移动手写数字数据集是时空序列预测中使用最广泛的数据集。该数据集可以随机生成数字, 并在  $64 \times 64$  的网格中做无规则运动。数据集生成数字的数

量, 移动的速度, 生成的序列长度都是可以自定义的, 理论上该数据集包含无穷多个不重复的序列, 非常符合实验要求。实验使用 MovingMNIST 随机创建的数据集, 数据集设置为 10 000 个序列, 其中训练集由 8 000 个序列组成, 验证集由 2 000 个序列组成, 序列长度为 10, batchsize 设置



为 16。

1)MovingMNIST 数据集对比实验结果分析

为了进一步研究预测模型的预测性能,将 3 个框架下的 5 种模型在 MovingMNIST 数据集上进行对比实验。定量评价如表 3 所示。

表 3 多框架预测算法在 MovingMNIST 数据集下的定量评价

Table 3 Quantitative evaluation of multi-framework prediction algorithms on the MovingMNIST dataset

框架	模型	MSE( $10^{-3}$ ) ↓	SSIM ↑
SISO	CrevNet	5.603	0.925
	SmartCrevNet	5.261	0.936
MIMO	SimVP	5.024	0.944
	FACP	4.608	0.967
MISO	IASCACP	<b>4.472</b>	<b>0.969</b>

由表 3 可知,与云图数据集结果不同,MIMO 框架下的 SimVP 与 FACP 在 MSE、SSIM 两种评价指标下精度优于 SISO 框架下的 CrevNet 与 SmartCrevNet。这是因为 MIMO 类预测模型更适用于特征较为简单的小尺度时空序列预测任务。作为结合了 SISO、MISO 两种框架优点的

MISO 类模型 IASCACP 在 MovingMNIST 数据集上有了更为亮眼的展现,两种指标均优于其他模型。

图 10 展示了多框架预测模型在 MovingMNIST 数据集下的 3 组预测示例。第 1 组“73”示例运动趋势较为简单,数字运动时不存在重合状态,5 个模型均实现了准确的预测,仔细来看 CrevNet、SmartCrevNet、SimVP 预测的最后一帧“3”的开口存在黏连,而 FACP 与 IASCACP 对该区域预测的更为精准。“36”示例在输入阶段一直处于融合状态,且标签中的“3”存在一定的噪声干扰。总体来看在“36”示例下,5 个模型在预测时均自动忽略掉了“3”中的噪声信息,且准确的预测到了运动趋势,其中 CrevNet 预测的“6”存在较大形变,SmartCrevNet 预测结果中两个数字存在较多黏连,IASCACP 预测的结果最为精准,数字状态最为清晰。“22”示例中两个数字在预测阶段均处于黏连或重合状态,为预测模型的性能提出了进一步的要求。具体的,在“22”示例下,5 个模型准确的预测到了前两帧数字,第 3 帧之后两个数字完全重合在了一起,CrevNet 与 SmartCrevNet 预测的最后一帧数字字叠处较为模糊,SimVP 与 FACP 在预测阶段学到了错误的趋势信息,导致后两帧的顶部出现异常。综合以上 3 组预测示例,IASCACP 的预测结果最为精确,随着预测的深入其对预测目标状态的维持能力最佳。

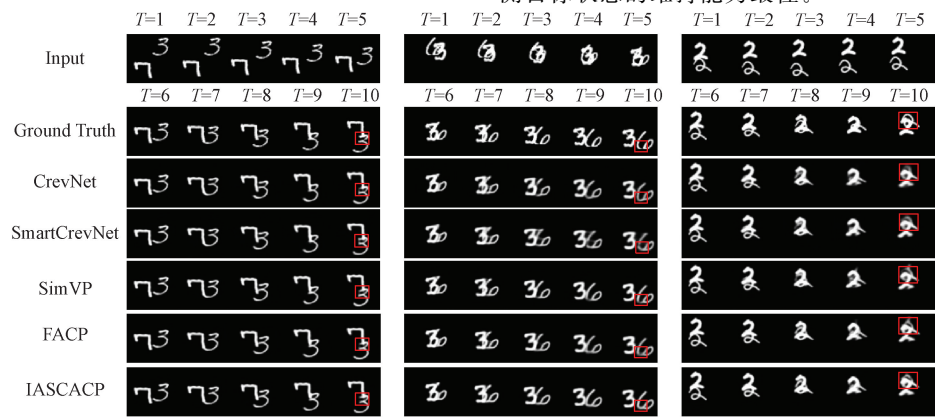


图 10 多框架预测模型 MovingMNIST 预测定性对比

Fig. 10 Qualitative comparison of multi-framework prediction models on MovingMNIST

4 结 论

本文以 MISO 为基础预测框架提出了一种基于隐式自回归时空通道聚合策略的卫星云图预测算法 IASCACP。该算法融合了 SISO 与 MIMO 两种预测框架的优点。具体的,引入了一种隐式自回归编解码器,为模型增加自回归特性以缓解 MIMO 框架下存在的时空关联性缺失问题,在编码阶段与解码阶段分别加入掩码模块与真实映射模块以消除自回归特性带来的误差累积。另外针对运动趋势不稳定、非线性的云图预测问题,在预测模块中引入了一种时空通道聚合器 SCA,其中时空聚合器

SA 用于获取序列云图的低、中、高阶特征相互作用信息,通道聚合器 CA 用于对多阶相互作用信息进行自适应再分配,减少了信道特征冗余,提升了多阶时空信息的有效性继而增加预测精度。将 IASCACP 与 SISO、MIMO 框架下的前沿模型在 FY-4A、MovingMNIST 数据集上进行了实验对比,结果表明 IASCACP 在各项指标下均优于其他模型,证明了该模型在云图预测领域相比 SISO、MIMO 框架模型更具备可靠性与有效性。

本文实验仍然具有一定的局限性,实验仅选取了 FY-4A 卫星夏季长三角地区的云图样本制作数据集,后续为了增加模型的泛用性和鲁棒性可以将数据集扩展到不同



季节全国的云图样本进行实验。

## 参考文献

- [1] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. *Neural Computation*, 1997, 9(8): 1735-1780.
- [2] CHO K, VAN MERRIËNBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[J]. *ArXiv preprint arXiv:1406.1078*, 2014.
- [3] RANZATO M A, SZLAM A, BRUNA J, et al. Video(language) modeling: A baseline for generative models of natural videos[J]. *ArXiv preprint arXiv:1412.6604*, 2014.
- [4] SRIVASTAVA N, MANSIMOV E, SALAKHUDINOV R. Unsupervised learning of video representations using LSTMs [C]. *International Conference on Machine Learning*. PMLR, 2015: 843-852.
- [5] SHI X J, CHEN ZH R, WANG H, et al. Convolutional LSTM network: A machine learning approach for precipitation nowcasting [C]. *29th International Conference on Neural Information Processing Systems*, 2015, 1: 802-810.
- [6] SHI X J, GAO ZH H, LAUSEN L, et al. Deep learning for precipitation nowcasting: A benchmark and a new model [C]. *31st International Conference on Neural Information Processing Systems*, 2017: 5622-5632.
- [7] WANG Y B, WU H X, ZHANG J J, et al. PredRNN: Recurrent neural networks for predictive learning using spatiotemporal LSTMs [C]. *31st International Conference on Neural Information Processing Systems*, 2017: 879-888.
- [8] WANG Y B, GAO ZH F, LONG M SH, et al. PredRNN++: Towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning [C]. *International Conference on Machine Learning*. PMLR, 2018: 5123-5132.
- [9] WANG Y B, JIANG L, YANG M H, et al. Eidetic 3D LSTM: A model for video prediction and beyond[C]. *International Conference on Learning Representations*, 2019.
- [10] YU W, LU Y CH, EASTERBROOK S, et al. Efficient and information-preserving future frame prediction and beyond[C]. *International Conference on Learning Representations*, 2020: 1-14.
- [11] 张秀再, 李景轩, 杨昌军, 等. 基于时空注意力门控循环单元的特征增强云图预测算法[J]. *激光与光电子学进展*, 2023, 60(24): 77-86.  
ZHANG X Z, LI J X, YANG CH J, et al. Feature-enhanced cloud image prediction algorithm based on spatio-temporal attention gated recurrent unit [J]. *Laser & Optoelectronics Progress*, 2023, 60 (24): 77-86.
- [12] LIU Z W, YEH R A, TANG X OU, et al. Video frame synthesis using deep voxel flow [C]. *IEEE International Conference on Computer Vision*, 2017: 4463-4471.
- [13] XU Z R, WANG Y B, LONG M SH, et al. PredCNN: Predictive learning with cascade convolutions [C]. *International Joint Conference on Artificial Intelligence*, 2018: 2940-2947.
- [14] GAO ZH Y, TAN CH, WU L R, et al. Simvp: Simpler yet better video prediction [C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022: 3170-3180.
- [15] SEO M, KIM D, SHIN S, et al. Simple baseline for weather forecasting using spatiotemporal context aggregation network [J]. *ArXiv preprint arXiv:2212.02952*, 2022.
- [16] VOLETI V, JOLICOEUR-MARTINEAU A, PAL C. MCVD-masked conditional video diffusion for prediction, generation, and interpolation [J]. *Advances in Neural Information Processing Systems*, 2022, 35: 23371-23385.
- [17] DENG H Q, REN Q H, ZHANG H, et al. Discovering and explaining the representation bottleneck of DNNs [C]. *International Conference on Learning Representations*, 2021.
- [18] LI K CH, WANG Y L, ZHANG J H, et al. Uniformer: Unifying convolution and self-attention for visual recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45 (10): 12581-12600.
- [19] PINTO F, TORR P H S, DOKANIA P. An impartial take to the cnn vs transformer robustness contest [C]. *European Conference on Computer Vision*. Cham: Springer Nature Switzerland, 2022: 466-480.
- [20] KINGA D, ADAM J B. A method for stochastic optimization [C]. *International Conference on Learning Representations*, 2015.
- [21] HORE A, ZIOU D. Image quality metrics: PSNR vs. SSIM [C]. *2010 20th International Conference on Pattern Recognition*. IEEE, 2010: 2366-2369.

## 作者简介

吴禹乾, 本科, 主要研究方向为深度学习。

E-mail: 2531627147@qq.com

张秀再(通信作者), 博士, 教授, 硕士生导师, 主要研究方向为气象通信技术与安全、机器学习。

E-mail: zxzhering@163.com

李景轩, 硕士研究生, 主要研究方向为深度学习。

E-mail: 1220947014@qq.com