

DOI:10.13232/j.cnki.jnju.2023.01.015

基于 3D 卷积和自注意力机制的卫星云图预测研究

方 巍^{1,2,3*}, 李佳欣¹, 陆文赫¹

(1. 数字取证教育部工程研究中心, 南京信息工程大学计算机学院, 南京, 210044; 2. 气象灾害国家重点实验室, 北京, 100081; 3. 江苏省大气环境与装备技术协同创新中心, 南京信息工程大学, 南京, 210044)

摘要: 卫星云图是气象预报的重要资源之一, 可以显示云层的生消变化, 对气象分析和预报工作有极大的作用. 对云图进行一定时间段的预测有助于及时掌握云层的移动轨迹和变化情况, 提高卫星云图资料的实用性. 然而, 当前卫星云图的预测面临诸多困难, 例如, 云团的变化大多是非平稳、非线性的; 云图数据量小, 实时性差等. 因此, 从时空序列的角度出发, 提出一种基于 3D 卷积和自注意力机制的卫星云图预测模型, 该模型在 ST-LSTM (Spatiotemporal Long Short-Term Memory) 的基础上, 在其单元内部引入 3D 卷积和自注意力机制, 使模型能同时提取时间信息和空间特征, 进一步增强云层短期趋势和长期依赖的联系; 同时, 在其外部框架使用空间和通道注意力机制, 促进对云图空间特征的提取. 在风云四号的卫星云图上进行评估, 实验结果证明, 该模型能够较准确地预测云层的形态变化和运动轨迹, 各项评价指标均优于现有模型.

关键词: 风云四号, 云图预测, 3D 卷积, 注意力机制

中图分类号: TP301

文献标志码: A

Research on satellite cloud image prediction based on 3D convolution and self-attention

Fang Wei^{1,2,3*}, Li Jiaxin¹, Lu Wenhe¹

(1. Engineering Research Center of Digital Forensics, Ministry of Education, School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, 210044, China; 2. State Key Laboratory of Meteorological Disasters, Beijing, 100081, China; 3. Jiangsu Provincial Collaborative Innovation Center for Atmospheric Environment and Equipment Technology, Nanjing University of Information Science and Technology, Nanjing, 210044, China)

Abstract: Satellite cloud image is one of the important resources of meteorological forecast. It plays a great role in meteorological analysis and forecasting by showing the generation and disappearance of clouds. Predicting cloud image in a certain period of time is helpful to grasp the movement trajectory and changes of cloud layers in time, and improve the practicability of satellite cloud image data. However, the prediction of the satellite cloud images is facing many difficulties, such as most changes in cloud clusters are non-stationary and nonlinear. There are many problems such as small cloud map data and poor real-time performance. Therefore, we propose a satellite cloud image prediction model based on 3D convolution and self-attention mechanism from the perspective of spatiotemporal sequence. On the basis of ST-LSTM (Spatiotemporal Long Short-Term Memory), this model introduces 3D convolution and self attention mechanism into its unit, which enables the model to extract temporal information and spatial features at the same time, furtherly enhance the relationship between short-term trend and long-term dependence on clouds; At the same time, space and channel attention mechanisms are used in

基金项目: 国家自然科学基金(42075007), 灾害性天气国家重点实验室开放基金(2021LASW-B19), 江苏省研究生科研与实践创新计划(KYCX22_1218)

收稿日期: 2022-10-06

* 通讯联系人, E-mail: hsfangwei@sina.com

its external framework to promote the extraction of spatial features of cloud images. In this paper, the evaluation is carried out on the Fengyun-4 satellite cloud image. Experimental results show that the model more accurately predicts the morphological changes and movement trajectories of clouds, and is superior to the existing models in various evaluation indicators.

Key words: Fengyun-4, cloud image prediction, 3D convolution, attention mechanism

随着中国科技的快速发展,国家气象卫星技术也在不断提高,获取更多的气象卫星数据是进行气象预测时不可或缺的资源之一,获取实时准确的气象卫星数据对保障民众日常生活有至关重要的作用.通过卫星云图可以看到地面云层的分布状态,从而进一步地分析和预测云团的生消演变.同时,不同形状和亮度的云层代表不同的天气状态,可以成为是否降水的依据,从而对强对流天气以及极端天气作出更准确的预报.但在现实工作中,由于气象卫星传输方式的限制,地面工作人员在接收气象卫星传过来的云图时存在较大的时延^[1],还存在云图丢失等情况,大大降低了预测的时效性.当前,人们对云图中的云层变化也只是依靠简单的线性外推,这种方式限制了气象预报的发展.图像序列预测为解决上述问题提供了一种十分有效的方法,因此,本文对卫星云图的时序预测展开了研究.

卫星云图预测实际是一个时空序列预测问题,将过去的卫星云图序列作为输入来对未来若干时刻的卫星云图进行预测,输出预测结果^[2].然而,由于输入的卫星云图观测的区域较大,生成的图片也较大,导致计算量十分庞大.同时,大气混沌效应使图像序列的变化规律非常复杂,这给建立有效的卫星云图预测模型带来了极大的挑战.

随着深度学习的快速发展,计算机视觉在其影响下得到快速进步,视频插值^[3]、自动驾驶^[4]等领域因其得到了快速发展,气象学相关预报和应用在深度学习的影响下也逐步成为热门的研究方向.许多研究将神经网络应用于卫星图像^[5-6],证明了深度神经网络在卫星图像上的有效性.

本文对卫星云图预测问题进行研究,提出一个新的预测模型,将卫星云图与计算机视觉、视频预测等技术相结合,对一定时间内的卫星云图进行准确、及时的预测.本文提出3D自注意力时空LSTM(Long Short-Term Memory),将3D卷积

和自注意力机制融入LSTM,可以更好地提取卫星云图的时间和空间信息.

本文的主要贡献:

(1)为了更好地预测卫星云图,使用卷积注意力模块(Convolutional Block Attention Module, CBAM)来对卫星云图进行特征提取,它可以有效地融合云图的通道和空间信息.

(2)设计了一种注意力记忆模块,将该模块无缝集成到记忆单元内部,使记忆单元在记住更早的历史信息的同时更注意序列中的关键细节.

(3)考虑卫星云图预测场景的特殊性,采用3D卷积,不仅提取当前卫星云图的轮廓、纹理等空间特征,还能提取相邻云图之间的时变信息.

1 相关工作

传统的卫星云图预测方法主要通过云团的匹配和追踪来进行研究^[7].研究人员肉眼观看某个云图后,记录观察到的云团,将其结构形状用特征量表示出来,再查找后面时刻的卫星云图是否存在相似的云团,如果存在则记录该云团的空间信息,进而确定云团的移动轨迹^[8].最后,根据大气运动的时空延续性,对目标云团进行线性外推,实现卫星云图的预测.2000年龚克等^[9]利用基于MPEG-2中运动矢量的方法预测了云团运动轨迹.2004年Lorenz et al^[10]使用Heliosat2方法获取云图的相关指标,结合运动矢量方法实现云团运动预测.2010年杨俊等^[11]利用局部阈值的思想对地基云图中的云团进行了预测.传统方法对卫星云图中云团的匹配和追踪有一定难度,采用的预测方法大多是线性的.然而,大气运动过程中存在大量非线性变化,如云团自身的形变和翻转等非平稳变化,因此,传统卫星云图预测方法很难得到准确的预测结果.

卫星云图预测任务本质上是一个时空预测任务,但与传统的时间序列不一样.时空序列预测

同时包含空间和时间信息,是一种无监督深度学习算法,利用大量的无标签数据作为输入,采用网络模型去学习连续的图像帧之间的联系与特征信息,通过图像之间的联系性准确地输出后面的图像,具有极大的研究潜力与应用价值. 这个算法的核心思路是向网络中输入 X_1, X_2, \dots, X_n , 准确生成后续的帧 $X_{n+1}, X_{n+2}, \dots, X_{n+t}$. 时序图像预测模型大多基于自编码模型和递归神经网络.

自编码器能实现对图像的压缩,其预测模块采用自编码器对图像进行压缩和预测. 基于自编码器的预测模型结构图如图1所示.

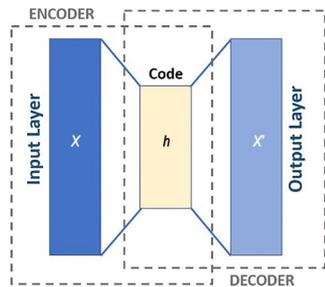


图1 自编码器的结构

Fig. 1 Autoencoder structure

Xing et al^[12]提出一种动态编码模型. 首先,编码器对像素图像进行编码,得到一个变量,这个变量与其状态相关,再对这个变量进一步编码,反复操作. 这就是采用逐层训练的方法构建一个多层编码器. Udrescu and Tegmark^[13]搭建一个基于卷积自编码器的预测模型,模型主要由两部分组成,第一部分用于获取输入的图像和时间差,将这两个部分进行编码之后传递给第二部分,也就是解码器,将输出的矢量转换为对应的图像,但该模型无法准确地预测动作变化. Liu et al^[14]构建了一个光流模型,它由三个卷积层和三个反卷积层组成,卷积层和反卷积层用跳跃连接,能够更好地保留空间信息. Xue et al^[15]搭建基于自编码器和由卷积网络改进的预测模型,创建了条件变分自编码器,能获取未来图像的条件分布,实验证明该模型取得了不错的效果.

循环神经网络(Recurrent Neural Network, RNN)^[16]由于其独特的性质在时间序列预测任务上取得了非常好的效果. 循环神经网络的架构如图2所示.

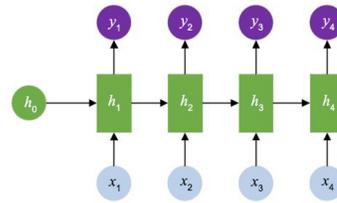


图2 循环神经网络的架构

Fig. 2 The architecture of RNN

RNN有权值共享的优势,在时间序列预测任务中可以根据当前时刻的输入数据和上一个时刻的隐藏状态得到当前时刻的预测数据,并对隐藏状态进行更新和记录,因此具有一定的记忆功能. 但是对于长时间序列问题,RNN的处理效果较差,还容易出现梯度消失等问题. Hochreite and Schmidhuber^[17]提出 LSTM,解决了这个问题. LSTM通过状态单元记录从长时间序列中提取有用的信息并记录下来,同时采用输入门、遗忘门和输出门控制输入信息的提取、状态单元的更新以及输出信息的选择. 尽管如此,LSTM也只能处理时序数据,对于卫星云图预测和雷达回波等时空序列问题,无法捕捉图像的空间信息.

Shi et al^[18]在 LSTM 的基础上提出卷积长短期记忆神经网络(Convolutional LSTM, Conv-LSTM)用于预测雷达回波图,取得了良好的预测结果,开启了利用RNN进行时空预测研究的新纪元. ConvLSTM通过卷积学习之前状态的空间信息,因此具备了同时提取空间特征和时间特征的能力. 该方法成功地对短时雷达回波图像进行了预测,其结果比实时光流法更加准确. 此后,在 ConvLSTM 的基础上发展了很多改进的变体结构,如 Shi et al^[19]引入光流轨迹的思想提出轨迹门控循环单元(Trajectory GRU, Traj GRU),可以主动学习用于重复连接的位置变化结构,进一步提高了预测的准确率.

为了提高网络对时空特征的非线性拟合能力,LSTM常被堆叠使用,但对于简单的堆叠结构,时空信息分别沿时间步横向传递和沿堆叠层纵向传递,纵向传递的时空信息没有得到有效利用. 因此,Wang et al^[20]提出时空长短期记忆(Spatio-temporal LSTM, ST-LSTM),将纵向传递的信息连接到下一时间步,并使用额外的时空记忆单元

对其进行更新和传递. Requena-Mesa et al^[21]将生成对抗网络和 LSTM 网络结合起来预测卫星云图,效果不错. 然而,标准 LSTM 单元仅使用一个输入门和一个遗忘门来完成输入信息的提取以及状态单元的遗忘和更新,往往只能提取输入信息中的部分信息对状态单元进行更新,因此输入信息中的部分有效信息可能会被丢失. Chang et al^[22]提出运动感知单元 (Motion-Aware Unit, MAU),通过有效拓宽时间感受野来提高模型在捕捉运动信息方面的表达能力. 其包含两个模块,一个关注模块和一个融合模块,实验结果证明,两个模块取得了不错的效果. Gao et al^[23]提出 SimVP,不使用复杂的模块,如 RNN, LSTM 和 Transformer,也不引入复杂的训练策略,如对抗性训练和课程学习. SimVP 仅需 CNN、跳连和 MSE 损失,为未来的研究提供了一种新思路.

2 方法

2.1 整体框架 图3是本文使用的网络架构图. 该网络主要由 CBAM 注意力模块、双向自编码器模块^[24]和 3D 自注意力时空 LSTM 预测模块组成. 首先,将多时段的卫星云图作为输入,利用 CBAM 充分提取特征后,将其输入双向自编码器模块,最后再用学习融合后的时空特征来预测未来时刻的云图序列.

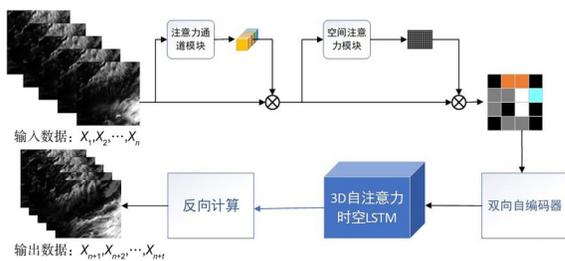


图3 本文的网络架构图

Fig. 3 Network architecture diagram of our algorithm

2.2 卷积注意力模块 为了更好地预测卫星云图,本文通过利用卷积注意力^[25]模块对卫星云图进行特征提取. 卷积注意力模块主要由两部分组成,即通道注意力机制和空间注意力机制,它的整体架构如图4所示. 通道注意力模块的作用是关注不同通道之间的相关性,采用计算获取不同通道的权重信息,并将获取的权重信息重新应用到

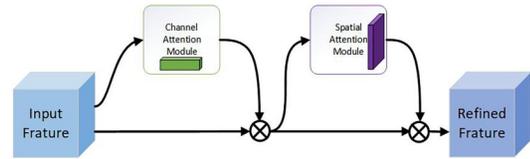


图4 卷积注意力模块的整体架构

Fig. 4 Convolutional attention module

提取的通道中,以此学习不同通道的特征. 空间注意力模块的作用是捕获特征图中不同像素位置之间的空间相关性,不同位置的像素对网络学习的重要程度不同,能得到重要区域上的相关信息,这是对通道注意力的有效补充.

通道注意力模块的具体结构如图5所示. 首先,将输入的特征 f_{ch} 进行降维,分别进行一次最大池化和一次平均池化,得到两个 $1 \times 1 \times C$ 的特征向量,这两个向量都包含输入特征在通道维度上的全局分布. 并且,为了减少计算量,在完成池化操作后,再用一个卷积对这两个特征向量进行降维,使模型的通道数减少到之前的 $1/16$. 其次,将两个降维后的特征向量进行叠加,通过一个 1×1 的卷积对两者进行融合,并将通道数恢复至原来的数量 C . 最后,经过一个 sigmoid 函数,得到通道注意力矩阵 CA_{ch} ,将其与原始输入逐元素相乘,实现在通道维度上对原输入特征的自适应调整. 通道注意力模块的数学表达如式(1)所示:

$$F_{channel} = CA_{ch} \cdot f_{channel} = \sigma \left(\delta \left(w_1 \cdot (w_0 \cdot v_{max} + w_0 \cdot v_{avg}) \right) \right) \cdot f_{channel} \quad (1)$$

其中, σ 表示 sigmoid, $w_0 \in R^{C/r \times C}$, $w_1 \in R^{C \times C/r}$ 分别代表两个 1×1 卷积核的权重, $\delta(\cdot)$ 代表 relu 函数, v_{max} 和 v_{avg} 分别代表经过最大池化和平均池化后的特征向量.

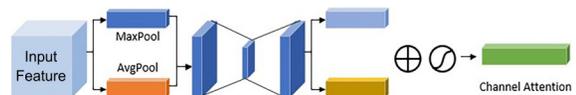


图5 通道注意力模块

Fig. 5 Channel attention module

空间注意力模块的流程与通道注意力模块相似,具体结构如图6所示. 首先,对通道注意力提取的特征沿通道分别进行一次最大池化和一次平均池化操作,将尺寸为 $H \times W \times C$ 的输入特征压缩为两张 $H \times W \times 1$ 的单通道特征图,展示输入

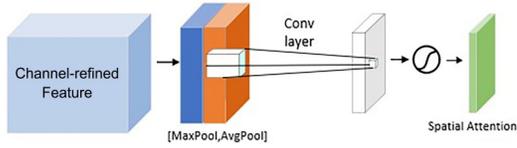


图 6 空间注意力模块

Fig. 6 Spatial attention module

在空间维度上的分布. 然后, 将这两个单通道特征图在通道维度上进行拼接, 再使用一个卷积对其进行融合和参数学习, 得到一张 $H \times W \times 1$ 的特征图. 最后, 经过一个 sigmoid 函数, 得到空间注意力矩阵 SA_{ch} , 将其与输入的原始特征逐元素相乘, 得到经过双重注意力细化的特征表达. 空间注意力模块的数学表达如式(2)所示:

$$F_{\text{spatial}} = SA_{ch} \cdot f_{\text{channel}} = \sigma\left(\delta\left(\omega_2 \cdot (f_{\text{max}} + f_{\text{avg}})\right)\right) \cdot f_{\text{channel}} \quad (2)$$

其中, F_{spatial} 代表经过注意力模块提取后的特征, ω_2 代表 1×1 卷积核的权重, f_{max} 和 f_{avg} 分别代表经过最大池化和平均池化后的特征图.

2.3 3D 自注意力时空 LSTM (3D SA-STLSTM)

2.3.1 3D 自注意力机制

2015年 Chorowski et al^[26]提出注意力机制来获取机器翻译中原始序列和目标序列之间的联系, 让模型能获取更长时间的信息. 注意力机制实质上是人类观察事物的方式, 也就是人脑会特意注重某些特殊的事物, 不自觉地忽视该事物之外的其他部分. 注意力机制可以给输入特征的不同部分赋予不同的权重, 经过权重重新表达的输入特征更容易被提取到有效的信息. 此处的权重可以看作人眼的敏锐度, 即视网膜的不同部位有不同程度的信息处理能力. 近几年, 注意力机制被广泛应用于图像处理、自然语言处理和情感分类等多样化任务^[27-28], 均获得了可观的性能提升. 传统 ConvLSTM 虽然能学习图像序列的有效时空信息并记忆特征, 却无法掌握图像序列的重要部分来提取某些细节. 为了解决这一问题, 一些工作^[29-30]尝试将注意力机制应用于 LSTM 或 ConvLSTM 以提高其序列学习性能, 然而, 大多数模型都将注意力机制放入外部来提取特征, 将注意力机制融合进预测细胞内部的研究还不多, 所以没能完全消除 ConvLSTM 缺少掌握细节能力的缺点. 受此启发, 本文设计了一种新的自注意力记忆模块, 将该模块与预测单元

融合, 下文中称其为自注意力门. 预测单元加入自注意力门后, 不仅能记录历史信息, 还能把控关键细节, 以此来计算云图序列的哪一部分更重要, 并赋予更大的权重. 其具体结构如图 7 所示.

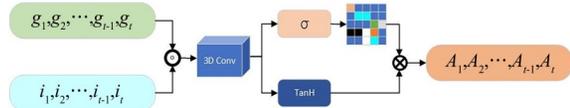


图 7 自注意力门的结构图

Fig. 7 Structure diagram of self-attention gate

自注意力门 A_t 由记忆门 g_t 和输入门 i_t 共同决定. 首先, 将记忆门与输入门拼接起来, 再分别通过卷积赋予权重. 其中一个使用 sigmoid 函数生成注意力矩阵, 注意力矩阵表示当前时刻的输入占据的不同比重, 同时还需要 tanh 函数来进行另一操作, 获取候选特征信息. 最后, 将两个函数输出的特征信息进行逐元素相乘, 得到自注意力门 A_t 的输出 A'_t , 并将该输出加入时间记忆细胞 C'_t 的计算, 为 C'_t 补充更多长远程距离的时变信息. 最后, 再与空间记忆细胞 M'_t 相加, 实现对当前时间步隐藏变量 M'_t 更精准的编码, 进而提升模型的预测精度. 自注意力门 A_t 的更新如式(3)所示:

$$A_t = \sigma(\omega_a \times [g_t, i_t] + b_a) \cdot \tanh(\omega_a \times [g_t, i_t] + b_a) \quad (3)$$

2.3.2 时空 LSTM

3D 自注意力时空 LSTM 是预测框架的基本组成单元, 其具体的内部结构如图 8 所示. 该记忆单元能同时对时间信息和空间特征进行提取, 进一步增强了短期趋势和长期依赖的特征提取和记忆, 适合解析卫星云图这一类演化规律复杂的预测任务. 该记忆单元主要由 ConvLSTM 的基本结构、空间记忆细胞 M_t 和自注意力模块三部分组成.

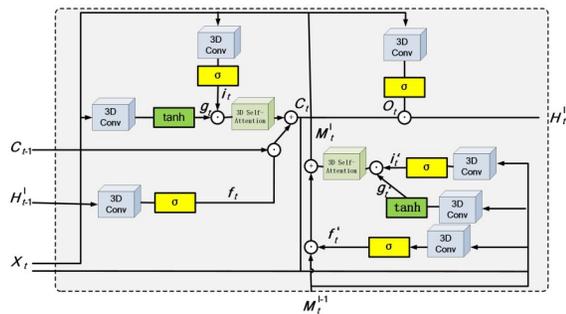


图 8 3D 自注意力时空 LSTM 结构图

Fig. 8 3D self-attention spatiotemporal LSTM structure diagram

ST-LSTM 是基于 ConvLSTM 改进的一种新的记忆单元,也是通过门控机制对特征信息进行筛选和传递,保留遗忘门、输入门、调制门、输出门、时间记忆细胞以及隐藏状态,分别对应图中的 f_t, i_t, g_t, o_t, C_t 和 H_t . 与 ConvLSTM 不同,ST-LSTM 在其基础上引入额外的空间记忆细胞 M_t ,在不同层之间垂直地提取和传递空间结构特征. 时间记忆细胞 C_t 用来捕捉当前时间步到下一时间步的时间依赖性,空间记忆细胞 M_t 用来捕捉同一时间步底层到顶层的空间依赖性.

综上所述,本文提出的 3D 自注意力时空 LSTM 的整体方程如式(4)所示:

$$\begin{aligned}
 g_t &= \tanh(W_g \times [X_t, H_{t-1}^t, C_{t-1}^t] + b_g) \\
 i_t &= \sigma(W_i \times [X_t, H_{t-1}^t, C_{t-1}^t] + b_i) \\
 f_t &= \sigma(W_f \times [X_t, H_{t-1}^t, C_{t-1}^t] + b_f) \\
 A_t &= \sigma(w_a \times [g_t, i_t] + b_a) \cdot \tanh(w_a \times [g_t, i_t] + b_a) \\
 C_t^t &= C_{t-1}^t \cdot f_t + A_t \\
 g_t^t &= \tanh(W_g^t \times [X_t, M_{t-1}^t, C_t^t] + b_g^t) \\
 i_t^t &= \sigma(W_i^t \times [X_t, M_{t-1}^t, C_t^t] + b_i^t) \\
 f_t^t &= \sigma(W_f^t \times [X_t, M_{t-1}^t, C_t^t] + b_f^t) \\
 A_t^t &= \sigma(w_a^t \times [g_t^t, i_t^t] + b_a^t) \cdot \tanh(w_a^t \times [g_t^t, i_t^t] + b_a^t) \\
 M_t^t &= M_{t-1}^t \cdot f_t^t + A_t^t \\
 o_t &= \sigma(W_{o0} \times X_t + W_{ho} \times H_{t-1}^t + W_{co} \times C_t^t + \\
 &\quad W_{mo} \times M_t^t + b_o) \\
 H_t &= o_t \cdot \tanh(W_{1 \times 1} \times [C_t^t, M_t^t])
 \end{aligned} \tag{4}$$

其中, W 和 b 分别代表权重和偏置项, $W_{1 \times 1}$ 代表 1×1 卷积滤波器的权重矩阵,用来控制特征的通道数. 由式(4)可见,时间记忆细胞 C_t^t 由遗忘门 f_t 、输入门 i_t 、记忆门 g_t 、同层的上一时间步时间记忆细胞 C_{t-1}^t 以及自注意力门 A_t 共同控制,输出的隐藏状态 H_t 由时间记忆细胞 C_t^t 、空间记忆细胞 M_t^t 和输出门 o_t 共同决定.

由此,时序信息和空间特征实现了无缝融合,能准确地对云图序列中的外观形变和时间轨迹等时空特征进行统一建模.

3 实验

3.1 卫星云图数据集 实验使用采集自风云四号的地球同步卫星数据. 风云四号是我国近几年研制的静止气象卫星,装载了多个气象探测传感

器,如干涉式大气垂直探测仪、空间环境监测仪器、闪电成像仪等,其中最重要的是多通道扫描成像辐射计 AGRI,其提供的多通道卫星云图也是本文主要的数据来源. 与上一代国产静止卫星风云二号(FY-2)的五个观测通道相比,FY-4A 装载的 AGRI 可以获取 14 个通道的卫星云图,还可提供彩色卫星云图,最快 1 min 即可生成一次区域观测图像. AGRI 的多通道观测数据使 FY-4A 不仅拥有观测云、水汽、植被、地表及捕捉气溶胶和雪的能力,还能清晰区分云的不同相态和高、中层的水汽^[31]. 本文主要使用 AGRI 的 L1 级数据,该数据共有 14 个通道,每个通道的序号、中心波长、空间分辨率和主要用途^[32]如表 1 所示.

本文采用 11 号水汽通道序列云图数据,区域范围为北纬 $26.6^\circ \sim 35.2^\circ$,东经 $115.2^\circ \sim 123.5^\circ$,视野从东到西为安庆和上海西部,从北到南为盐城和台州. 该数据集是 2021 年七月和八月的卫星云图,每张图片的分辨率被处理为 256×256 的 PNG 图片,帧与帧之间的采样间隔为 1 h,共 1488 张原始图片. 实验过程中,训练数据集和验证数

表 1 风云四号各通道的相关信息

波段	通道名称	中心波长 (μm)	空间分辨率 (km)	主要用途
可见光	1	0.46	1	植被
	2	0.64	0.5	植被,雾,恒星观测
近红外	3	0.86	1	植被,水面上气溶胶
	4	1.38	2	卷云
短波红外	5	1.61	2	低云和雪识别,水云和冰
	6	2.25	2~4	卷云,气溶胶,恒星观测
中波红外	7	3.38 (H)	2~4	火情
	8	3.80 (L)	4	水汽,雾
水汽	9	6.50	4	高层水汽
	10	7.20	4	中层水汽
	11	8.50	4	云
长波红外	12	10.80	4	地表温度
	13	12.00	4	地表温度
	14	13.30	4	水汽

据集采用 8:1 的比例分配,最终训练数据集有 596 组,验证数据集有 75 组。

3.2 实验设备 所有实验均在 Pytorch-GPU-1.10.0 和 python-3.7 上实现,并在单张 NVIDIA GeForce RTX 3060(显存 12 G)上运行。

3.3 评价指标 选用 *MSE* (Mean Square Error), *MAE* (Mean Absolute Error), *SSIM* (Structural Similarity Index Measure)^[33] 和 *PSNR* (Peak Signal to Noise Ratio)^[33] 来评估模型的预测性能。*MSE* 和 *MAE* 是计算预测图像与真实图像之间差距的指标,也是生成图像变化程度的指标。*SSIM* 和 *PSNR* 是计算机视觉领域广泛使用的两种图像级评价指标,*SSIM* 是衡量两幅图像结构相似性的主观指标,*PSNR* 是评价图像质量的客观指标。*MSE* 和 *MAE* 越小或 *SSIM* 和 *PSNR* 越大,说明模型的准确性更好,预测结果也更好。

3.4 模型评估

3.4.1 数据预处理 从国家卫星中心下载数据后需要进行预处理,主要包含两部分:(1)针对某个时间点的卫星云图丢失,采用最近的相邻时间点云图进行代替,同时将原始文件的 HDF 格式转换为 PNG 格式,方便后续文件的读取、处理和可视化;(2)由于实验设备的限制,将所有卫星云图统一缩小到 256×256 的尺寸,同时将图片像素归一化到 [0, 1](如图 9 所示),在可控范围内减少计算资源的消耗,同时最大化利用服务器的算力。



图 9 处理后的卫星云图

Fig. 9 The processed satellite image

3.4.2 数值分析 使用 *MAE*, *MSE*, *SSIM*, *PSNR* 对四种现有模型 ConvLSTM, PredRNN, PredRNN++, Crevnet 和本文的 3D SA-STLSTM 在风云四号卫星云图数据集上进行对比实验,表 2 展示了每个模型在不同评价指标上的测试结果,表中数值是模型预测的后五帧预测云图的平

均值。由表可见,本文模型的四个评价指标都优于其他模型。与 ConvLSTM 这种仅具有单一细胞的模型相比,各项指标都有大幅的提高,*MAE*, *MSE*, *PSNR* 和 *SSIM* 分别提升 0.0442, 0.0001, 21.234, 0.301, 这是因为后续模型引入了记忆细胞 M_t 。与 PredRNN, PredRNN++ 和 Crevnet 这类具有双重记忆细胞的模型相比,3D SA-STLSTM 的 *MAE*, *MSE*, *PSNR* 和 *SSIM* 最多提升 0.002, 0.003, 27.891 和 0.034, 说明本文模型和其他模型相比,预测的卫星云图在图像保真度上更趋于真实值,可以更加准确地预测云团的移动轨迹和形状变化。这是因为引入 3D 卷积和自注意力后,模型加强了对云层短期变化趋势的学习,也能更好地捕捉长期的时空特征,使预测性能在一定程度上得到提高。

表 2 3D SA-STLSTM 与其他模型在卫星云图数据集上的测试结果

Table 2 Testing results of 3D SA-STLSTM and other models on satellite cloud image datasets

Method	<i>MAE</i> ↓	<i>MSE</i> ↓	<i>PSNR</i> ↑	<i>SSIM</i> ↑
ConvLSTM ^[18]	0.0528	0.0138	7.561	0.194
PredRNN ^[20]	0.0109	0.0127	14.730	0.259
PredRNN++ ^[34]	0.0135	0.0037	20.757	0.359
Crevnet ^[24]	0.0144	0.0008	27.891	0.461
3D SA-STLSTM	0.0086	0.0007	28.795	0.495

为了更全面地对比不同模型的预测性能,对每一轮的预测结果也进行了对比,同时绘制了四项评价指标的变化曲线,如图 10 所示。

由图可见,随着轮数不断增加,每个模型的损失都是整体呈下降趋势,图像精度呈上升趋势。本文模型和其他模型相比,表现了更好的整体性能,证明本文模型在捕捉时空特征方面有更好的优势,能更好地预测云图。

3.4.3 可视化 为了更直观地展示不同算法在卫星云图数据集上的预测效果,图 11 展示了 ConvLSTM, PredRNN, PredRNN++, Crevnet 和 3D SA-STLSTM 的预测结果。其中,Input 代表输入的前五帧图像,Ground Truth 代表真实的后五帧图像;白色区域代表云团,黑色区域代表陆地或海洋。白色区域越亮代表云团越多,降水的可能性

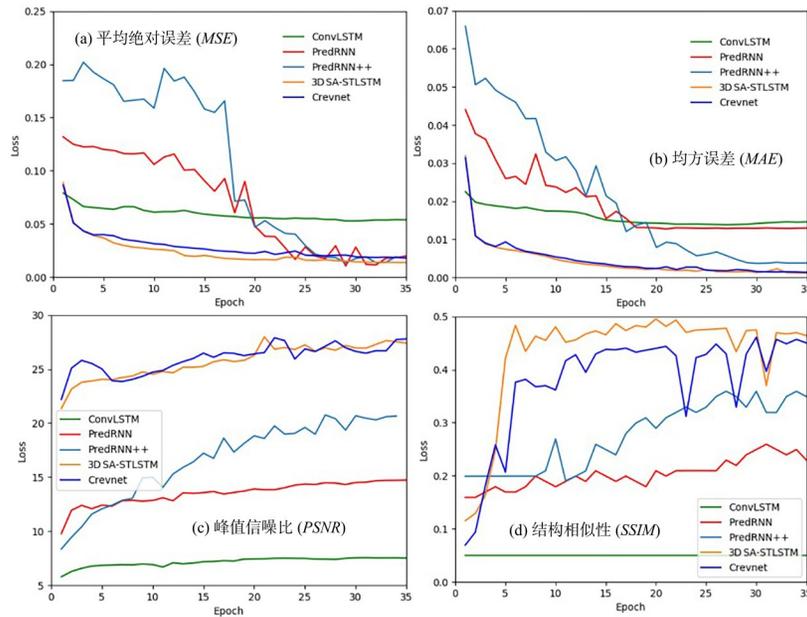


图 10 模型在不同指标中每轮显示

Fig.10 The models are shown each round in different metrics

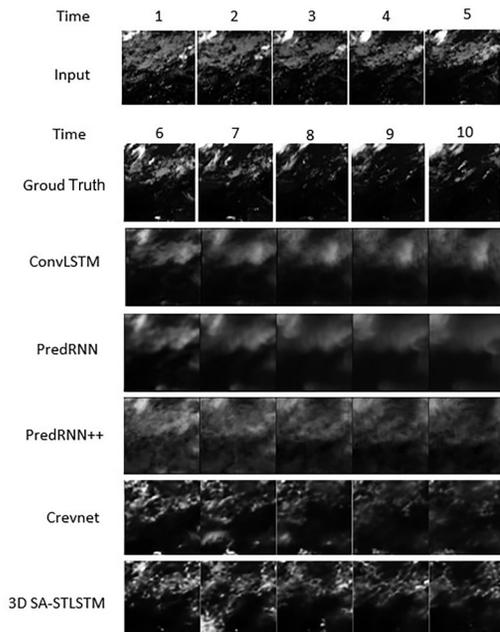


图 11 模型在卫星云图数据集上的预测样本展示

Fig. 11 Prediction sample display of the models in the satellite cloud image dataset

越大;白色区域的明暗,代表不同厚度的云团。

由图可见, ConvLSTM, PredRNN 和 PredRNN++ 预测的云团与 Ground Truth 的云图消散趋势大致一致, 虽然能学习到云团大致的演化规律和运动趋势, 但生成的是模糊的外观, 无法达到

气象预测业务的需求. 与上述模型相比, Crevnet 和 3D SA-STLSTM 不仅能准确地预测云团的演变趋势, 而且, 随着时间步的推移, 仍然能保持较清晰的轮廓. 相比之下, 3D SA-STLSTM 不仅能保持完整轮廓, 还能保留局部细节特征. 通过上述对比实验发现, 3D SA-STLSTM 比其他模型表现更好, 在跨度 5 h 的卫星云图预测中可以达到较高的精度, 并能保持准确和清晰的云团, 提高了卫星云图预测的质量.

4 结论

通过对国内外卫星云图预测和深度学习相关资料的整理, 分析和总结传统方法的优缺点, 针对 ConvLSTM 模型误差大、泛化性差及卫星云图的特殊性等问题, 提出一个基于 3D 自注意力机制的卫星云图预测模型. 通过引入 3D 卷积和注意力机制, 改进了模型的内部单元和外部架构, 有助于增强对时间信息和空间特征的获取能力, 提升了模型对短期信息和长远信息的预测能力.

在风云四号卫星云图数据集上的对比实验, 验证了模型的有效性和优越性. 实验结果证明, 提出的模型能更准确地预测卫星云图上云图的移动轨迹和形态变化, 并在各项评估指标上均优于现

有的模型,展现了一定的优越性和有效性,实现了更加精准化和智能化的卫星云图预测。

卫星云图预测始终是气象预报中的难点与重点,本文利用深度学习,结合该方向进行了初步测试,效果较好,但离真实图像还相差甚远。未来的研究需要增大研究样本,并将更多的气象数据(其他通道的卫星云图、大气矢量图)纳入模型,提升模型的预测准确性。

参考文献

- [1] Wang F, Jiang D D, Qi S, et al. A dynamic resource scheduling scheme in edge computing satellite networks. *Mobile Networks and Applications*, 2021, 26(2):597–608.
- [2] Jing L L, Tian Y L. Self-supervised visual feature learning with deep neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(11):4037–4058.
- [3] Oprea S, Martinez-Gonzalez P, Garcia-Garcia A, et al. A review on deep learning techniques for video prediction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 44(6):2806–2826.
- [4] Bhattacharyya A, Fritz M, Schiele B. Long-term on-board prediction of people in traffic scenes under uncertainty//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018:4194–4202.
- [5] Kong T, Sun F C, Liu H P, et al. Foveabox: Beyond anchor-based object detection. *IEEE Transactions on Image Processing*, 2020(29):7389–7398.
- [6] Shi Q, Liu M X, Li S C, et al. A deeply supervised attention metric-based network and an open aerial image dataset for remote sensing change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 2021(60):5604816.
- [7] 白洁,王洪庆,陶祖钰. GMS 卫星红外云图强对流云团的识别与追踪. *热带气象学报*, 1997, 13(2): 63–72. (Bai J, Wang H Q, Tao Z Y. Recognition and tracing of severe convective cloud from IR images of GMS. *Journal of Tropical Meteorology*, 1997, 13(2):63–72.)
- [8] 陈靖. 地基云图中云团的识别和短时外推方法研究. 硕士学位论文. 天津:天津大学, 2015. (Chen J. Research on clouds recognition and short-time extrapolation method in ground-based cloud image. Master Dissertation. Tianjin: Tianjin University, 2015.)
- [9] 龚克,叶大鲁,葛成辉. 卫星云图预测的运动矢量方法. *中国图象图形学报*, 2000, 5(4): 349–352. (Gong K, Ye D L, Ge C H. A method for geostationary meteorological satellite cloud image prediction based on motion vector. *Journal of Image and Graphics*, 2000, 5(4):349–352.)
- [10] Lorenz E, Hammer A, Heinemann D. Short term forecasting of solar radiation based on satellite data// ISES Europe Solar Congress. Freiburg, Germany: IEEE Press, 2004:841–848.
- [11] 杨俊,吕伟涛,马颖,等. 基于局部阈值插值的地基云自动检测方法. *气象学报*, 2010, 68(6): 1007–1017. (Yang J, Lu W T, Ma Y, et al. An automatic ground-based cloud detection method based on the local threshold interpolation. *Acta Meteorologica Sinica*, 2010, 68(6):1007–1017.)
- [12] Xing Y, Hong C, Shan S G, et al. Modeling video dynamics with deep dynencoder//The 13th European Conference on Computer Vision. Springer Berlin Heidelberg, 2014:215–230.
- [13] Udrescu S M, Tegmark M. Symbolic regression: Discovering physical laws from distorted video. *Physical Review E*, 2021, 103(4):043307.
- [14] Liu Z W, Yeh R A, Tang X O, et al. Video frame synthesis using deep voxel flow//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017:4473–4481.
- [15] Xue T F, Wu J J, Bouman K L, et al. Visual dynamics: Probabilistic future frame synthesis via cross convolutional networks//Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain: Curran Associates Inc., 2016:91–99.
- [16] Mikolov T, Karafiát M, Burget L, et al. Recurrent neural network based language model// INTERSPEECH 2010. Chiba, Japan: ISCA, 2015: 1045–1048.
- [17] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Computation*, 1997, 9(8): 1735–1780.
- [18] Shi X J, Chen Z R, Wang H, et al. Convolutional LSTM network: A machine learning approach for precipitation nowcasting//Proceedings of the 28th

- International Conference on Neural Information Processing Systems. Montreal, Canada: MIT Press, 2015:802–810.
- [19] Shi X J, Gao Z H, Lausen L, et al. Deep learning for precipitation nowcasting: A benchmark and a new model//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, CA, USA: Curran Associates Inc., 2017:5622–5632.
- [20] Wang Y B, Long M S, Wang J M, et al. PredRNN: Recurrent neural networks for predictive learning using spatiotemporal LSTMs//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, CA, USA: Curran Associates Inc., 2017:879–888.
- [21] Requena - Mesa C, Benson V, Reichstein M, et al. EarthNet2021: A large - scale dataset and challenge for Earth surface forecasting as a guided video prediction task//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, TN, USA: IEEE, 2021: 1132–1142.
- [22] Chang Z, Zhang X, Wang S, et al. Mau: A motion - aware unit for video prediction and beyond. Advances in Neural Information Processing Systems, 2021(34): 26950–26962.
- [23] Gao Z Y, Tan C, Wu L R, et al. SimVP: Simpler yet better video prediction//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, LA, USA: IEEE, 2022: 3170–3180.
- [24] Yu W, Lu Y C, Easterbrook S, et al. Efficient and information - preserving future frame prediction and beyond//International Conference on Learning Representations. Addis Ababa, Ethiopia: ICLR, 2020:12.
- [25] Woo S, Park J, Lee J Y, et al. CBAM: Convolutional block attention module//Proceedings of the 15th European conference on computer vision. Springer Berlin Heidelberg, 2018:3–19.
- [26] Chorowski J, Bahdanau D, Serdyuk D, et al. Attention - based models for speech recognition// Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, Canada: MIT Press, 2015:577–585.
- [27] Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature, 2021, 596(7873):583–589.
- [28] Lin X, Zhong G Q, Chen K, et al. Attention - augmented machine memory. Cognitive Computation, 2021, 13(3):751–760.
- [29] 王红, 史金钊, 张志伟. 基于注意力机制的 LSTM 的语义关系抽取. 计算机应用研究, 2018, 35(5): 1417–1420, 1440. (Wang H, Shi J C, Zhang Z W. Semantic relation extraction of LSTM based on attention mechanism. Application Research of Computers, 2018, 35(5): 1417–1420, 1440.)
- [30] Solgi R, Loáiciga H A, Kram M. Long short - term memory neural network (LSTM - NN) for aquifer level time series forecasting using in - situ piezometric observations. Journal of Hydrology, 2021(601): 126800.
- [31] 陆风, 张晓虎, 陈博洋, 等. 风云四号气象卫星成像特性及其应用前景. 海洋气象学报, 2017, 37(2):1–12. (Lu F, Zhang X H, Chen B Y, et al. FY - 4 geostationary meteorological satellite imaging characteristics and its application prospects. Journal of Marine Meteorology, 2017, 37(2):1–12.)
- [32] 陶润喆. 基于风云 4 号卫星图像的西藏地区云检测和降水外推预报研究. 博士学位论文. 南京: 南京信息工程大学, 2021. (Tao R Z. Research on cloud detection and precipitation extrapolation forecast in Tibet based on Fengyun 4 satellite image. Ph.D. Dissertation. Nanjing: Nanjing University of Information Science and Technology, 2021.)
- [33] Setiadi D R I M. PSNR vs SSIM: Imperceptibility quality assessment for image steganography. Multimedia Tools and Applications, 2021, 80(6): 8423–8444.
- [34] Wang Y B, Gao Z F, Long M L, et al. Predrnn++ : Towards a resolution of the deep - in - time dilemma in spatiotemporal predictive learning//Proceedings of the 35th International Conference on Machine Learning. Stockholm, Sweden: PMLR Press, 2018: 5110–5119.

(责任编辑 杨可盛)