

基于时空注意力门控循环单元的特征增强云图预测算法

张秀再^{1,2}, 李景轩², 杨昌军^{3,4*}, 冯绚⁵

¹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, 江苏 南京 210044;

²南京信息工程大学电子与信息工程学院, 江苏 南京 210044;

³中国气象局中国遥感卫星辐射测量和定标重点开放实验室/
国家卫星气象中心(国家空间天气监测预警中心), 北京 100081;

⁴许健民气象卫星创新中心, 北京 100081;

⁵中国科学院空间利用技术与工程中心, 北京 100094

摘要 各种天气现象的变化都伴随着云的运动, 气象卫星获取的连续卫星云图含大量的时空序列信息, 即连续卫星云图具有显著的时序特征, 可作为云图预测的基本信息。云图预测本质上是一种处理云图时空序列特性的视频预测问题, 为了准确预测云的位置变化, 针对云的不稳定、非线性运动特性, 基于CrevNet视频预测模型提出SmartCrevNet云图预测算法。在该算法中设计了一种时空注意力门控循环预测单元(STA-GRU), 同时在CrevNet原有的双向自编码模块中引入轻量型注意力模块(SGE), 可在不增加计算量的情况下增强云图语义信息, 提高特征提取能力。将该算法分别在公共数据集 MovingMNIST 数据集及 FY-4A 卫星云图数据集上进行实验, 结果表明, 在 FY-4A 卫星云图数据集和 MovingMNIST 数据集上, SmartCrevNet 的均方误差(MSE)分别比 CrevNet 降低了 7.3% 和 6.1%, 结构相似性(SSIM)分别提升了 7.9% 和 1.2%, 预测效果优于 CrevNet 和传统的视频预测算法。

关键词 遥感; 云图预测; FY-4A 卫星; 门控循环单元; 轻量型注意力模块

中图分类号 TP751.1

文献标志码 A

DOI: 10.3788/LOP231059

Feature-Enhanced Cloud Image Prediction Algorithm Based on Spatio-Temporal Attention Gated Recurrent Unit

Zhang Xiuzai^{1,2}, Li Jingxuan², Yang Changjun^{3,4*}, Feng Xuan⁵

¹Jiangsu Province Atmospheric Environment and Equipment Technology Collaborative Innovation Center, Nanjing University of Information Science & Technology, Nanjing 210044, Jiangsu, China;

²School of Electronic and Information Engineering, Nanjing University of Information Science & Technology, Nanjing 210044, Jiangsu, China;

³Key Laboratory of Radiometric Calibration and Validation for Environmental Satellites, National Satellite Meteorological Center (National Center for Space Weather), China Meteorological Administration, Beijing 100081, China;

⁴Innovation Center for FengYun Meteorological Satellite (FYSIC), Beijing 100081, China;

⁵Technology and Engineering Center for Space Utilization, Chinese Academy of Sciences, Beijing 100094, China

Abstract Changes in various weather phenomena are accompanied by the movement of clouds. Continuous satellite cloud images obtained by meteorological satellites contain considerable spatio-temporal sequence information; that is, continuous satellite cloud images have significant time sequence characteristics, which can be used as basic information for cloud image prediction. Cloud image prediction is essentially a video prediction problem in which the spatio-temporal sequence characteristics of cloud images are analyzed and processed. To accurately predict the change in the position of

收稿日期: 2023-04-10; 修回日期: 2023-04-25; 录用日期: 2023-05-15; 网络首发日期: 2023-05-25

基金项目: 第二次青藏高原综合科学考察研究项目(2019QZKK0105)、中美人工智能战略比较研究(22BZZ080)、国家自然科学基金青年基金项目(41905033)

通信作者: *yangcj@cma.gov.cn

clouds, by focusing on the unstable and nonlinear motion characteristics of clouds, a video prediction algorithm called SmartCrevNet for cloud image prediction based on the CrevNet video prediction model is proposed. In this algorithm, a spatiotemporal attention gated recurrent unit (STA-GRU), along with the lightweight attention module spatial group-wise enhance (SGE), is introduced into the original two-way autoencoder module of CrevNet to enhance the ability of feature extraction without increasing the amount of calculation. The algorithm was tested on the public dataset MovingMNIST and the FY-4A satellite cloud image dataset. The results show that, on the FY-4A satellite cloud image dataset and MovingMNIST dataset, the mean square error (MSE) of SmartCrevNet is, respectively, 7.3% and 6.1% lower than that of CrevNet, and the structural similarity (SSIM) is increased by 7.9% and 1.2%, respectively. The prediction effect is better than that of CrevNet and traditional video prediction algorithms.

Key words remote sensing; cloud image prediction; FY-4A; gated recurrent unit; lightweight attention module

1 引言

卫星云图预测利用气象卫星获取的连续云图数据来预测云图位置变化,其本质是一种视频预测。视频预测根据输入的连续图像帧预测输出未来的连续图像帧,这种序列到序列的预测主要利用神经网络中的循环神经网络(RNN)实现。目前,常用于视频预测算法的长短期记忆网络(LSTM)^[1]及门控循环单元(GRU)^[2]都是RNN的变种。

基于RNN的视频预测方法起步较晚,2014年Ranzato等^[3]将RNN应用到视频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模型。Srivastava等^[4]将LSTM引入视频预测领域,使用LSTM分别作为预测模型的编码器与解码器实现视频预测,并点明卷积神经网络(CNN)与LSTM结合提升预测准确度的可行性。Shi等^[5]借鉴Srivastava的思想将CNN与LSTM结合提出适用于图像预测的ConvLSTM,根据网络中信息流的传输设计encoding-forecasting(EF)结构,并将该预测模型用于降水预测且取得了不错的效果。之后,Shi等^[6]进一步优化了EF结构,提出较ConvLSTM更加轻量化的ConvGRU,并引入可学习卷积提出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年—2019年,Wang等^[7-9]在ConvLSTM的基础上融入3D卷积、时空记忆单元、梯度高速公路等,先后提出PredRNN^[7]、PredRNN++^[8]以及E3D-LSTM^[9],为视频预测技术的发展作出了重大贡献。近几年,随着深度学习领域的生成式对抗网络(GAN)^[10]与自注意力机制(self-attention)^[11]的火热,Xu等^[12]与Lin等^[13]分别将GAN与自注意力机制融入LSTM中进行视频预测,取得了较好的效果。以上视频预测算法为了尽可能保留图像的信息量需要占用极大的内存空间,大大限制了算法的应用范围。2020年,Yu等^[14]提出一种条件可逆网络CrevNet,该模型在特征提取过程中几乎无信息丢失、内存消耗较低且计算效率和预测精准度较高,有效解决了传统视频预测算法运行时显存不足、内存不足、计算速度较慢的问题。

由于云移动的不稳定、非线性等特点,基于深度学

习方法的云移动预测难度较大。2017年,Hong等^[15]将自编码器与ConvLSTM相结合以提取时序云图中的时空特征,在编码器与解码器之间加入跳跃连接以提升预测结果的精确度。2021年,蔡朋艳^[16]将3D卷积与U-Net相结合作为云图预测模块,并使用GAN来训练预测网络以提升预测模型对时空特征的提取能力。云的运动受到大气环流、地形、气压等多因素的影响,在运动过程中会出现扩散、消融等情况,针对云的移动特点,云移动预测模型的设计应当从以下两个方面入手:

1)针对无规则的云图运动,应当更加重视预测模块对云图短期运动特征与长期运动特征的综合利用,兼顾提取具有全局依赖和局部依赖的时空特征,才能使预测网络更精确地学习到云的运动趋势;

2)针对云大规模无规律分散分布,应当对输入的时序云图进行更加全面且充分的特征提取,尽可能地减少薄云碎云及边界云在特征提取时的信息流失。

CrevNet因其在保证模型轻量化的同时能提升预测的精准度而广受好评,本文从以上两个方面对CrevNet进行针对性的改进,基于CrevNet提出SmartCrevNet模型。具体工作如下:

1)设计了一种改进的时空注意力门控循环预测模块(STA-GRU)。该模块在ConvGRU的基础上增加空间隐藏状态,引入3D卷积网络使模型能很好地捕捉图像的时空序列特征^[17]。另外为解决云移动的不稳定、非线性等问题,模型的预测模块需要兼顾提取序列云图中的具有全局依赖和局部依赖的时空特征。因此,该模块又引入注意力思想,将历史信息模块(HIM)融入预测模块中,得到STA-GRU。

2)针对云运动中的扩散、消融等问题,改进CrevNet中的特征提取模块,将轻量级注意力模块(SGE)融入双向自编码器,在不增加计算量的情况下,可增强云图特征的语义信息和减少特征提取时的信息流失^[18]。

3)从中国国家卫星气象中心获取风云四号卫星(FY-4A)的原始遥感数据,经处理得到长三角地区的2021年6月—10月的卫星云图,制作FY-4A云图数据集。将SmartCrevNet及对比模型在该数据集上进行实验,并对比分析实验效果。

2 相关理论基础

2.1 卫星云图预测基本原理

在卫星云图预测中,一般利用神经网络根据前几帧时序云图预测后几帧时序云图。若观测的每帧时序云图都被记录在 $H \times M$ 大小的网格中,则每帧时序云图可以用一个张量 $\mathbf{x} \in \mathbf{R}^{H \times W}$ 表示。将观测的每帧时序云图表示为 $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \dots$, 云图预测问题就是利用连续观测的 m 帧时序云图预测后 n 帧时序云图:

$$\tilde{\mathbf{x}}_{t+1}, \dots, \tilde{\mathbf{x}}_{t+n} = \underset{\mathbf{x}_{t+1}, \dots, \mathbf{x}_{t+n}}{\operatorname{argmax}} \rho(\mathbf{x}_{t+1}, \dots, \mathbf{x}_{t+n} | \mathbf{x}_{t-m+1}, \dots, \mathbf{x}_t). \quad (1)$$

2.2 门控循环单元

LSTM 已被证实在视频预测中的重要作用,而 GRU 与 LSTM 一样也是为了解决传统 RNN 存在的梯度消失等问题而被提出的。GRU 有着与 LSTM 几乎相同的实验效果,且更易于计算,本研究旨在探索 GRU 在云图预测中的可靠性。

GRU 基本框架如图 1 所示, $\mathbf{H}_{t-1}, \mathbf{X}_t$ 分别表示前一时刻的隐藏状态输入、当前时刻的数据输入。与 LSTM 的 3 个控制门不同,GRU 只有 2 个控制门,图 1 中 \mathbf{R}'_t 与 \mathbf{Z}'_t 分别代表重置门(reset gate)和更新门(update gate)。GRU 的关键方程为

$$\begin{cases} \mathbf{R}'_t = \sigma(\mathbf{X}_t \mathbf{W}_{xr} + \mathbf{H}_{t-1} \mathbf{W}_{hr} + \mathbf{b}_r) \\ \mathbf{Z}'_t = \sigma(\mathbf{X}_t \mathbf{W}_{xz} + \mathbf{H}_{t-1} \mathbf{W}_{hz} + \mathbf{b}_z) \\ \tilde{\mathbf{H}}_t = \tanh[\mathbf{X}_t \mathbf{W}_{xh} + (\mathbf{R}'_t \odot \mathbf{H}_{t-1}) \mathbf{W}_{hh} + \mathbf{b}_h] \\ \mathbf{H}_t = \mathbf{Z}'_t \odot \mathbf{H}_{t-1} + (1 - \mathbf{Z}'_t) \odot \tilde{\mathbf{H}}_t \end{cases} \quad (2)$$

式中: σ 与 \tanh 为两种激活函数; \odot 为哈达玛乘积; \mathbf{W}

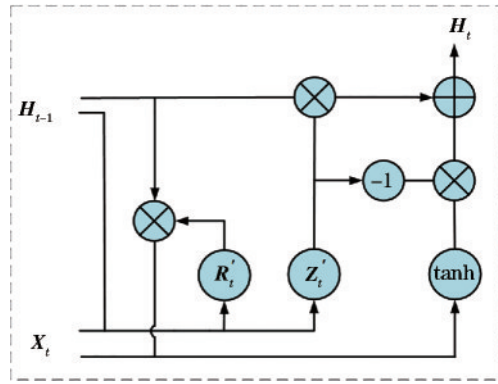


图 1 GRU 基本框架

Fig. 1 Basic framework of GRU

与 \mathbf{b} 分别表示输入门中对应数据的权重与偏差,如 \mathbf{W}_{xr} 表示数据 \mathbf{X} 输入重置门的权重, \mathbf{W}_{hr} 是隐藏状态 \mathbf{H} 输入重置门的权重, \mathbf{b}_r 表示输入重置门总数据的偏差值。

重置门 \mathbf{R}'_t 可以忽略输入的前一时刻隐藏状态 \mathbf{H}_{t-1} 中不相关信息,再加上输入数据 \mathbf{X} , 可得到候选隐藏状态 $\tilde{\mathbf{H}}$ 。

\mathbf{Z}'_t 为更新门,可以选择性遗忘前一时刻的隐藏状态和记忆当前时刻的候选隐藏状态,输出为当前时刻的隐藏状态 \mathbf{H}_t 。 \mathbf{H}_t 在视频预测中也代表输出的预测数据。

3 模型设计

3.1 基于时空注意力门控循环单元的特征增强预测算法 SmartCrevNet

CrevNet 在减少特征提取中信息流失的同时,保证了预测结果的精准度,减少了计算量,非常适合改进成针对云图预测的模型。CrevNet 框架如图 2 所示。

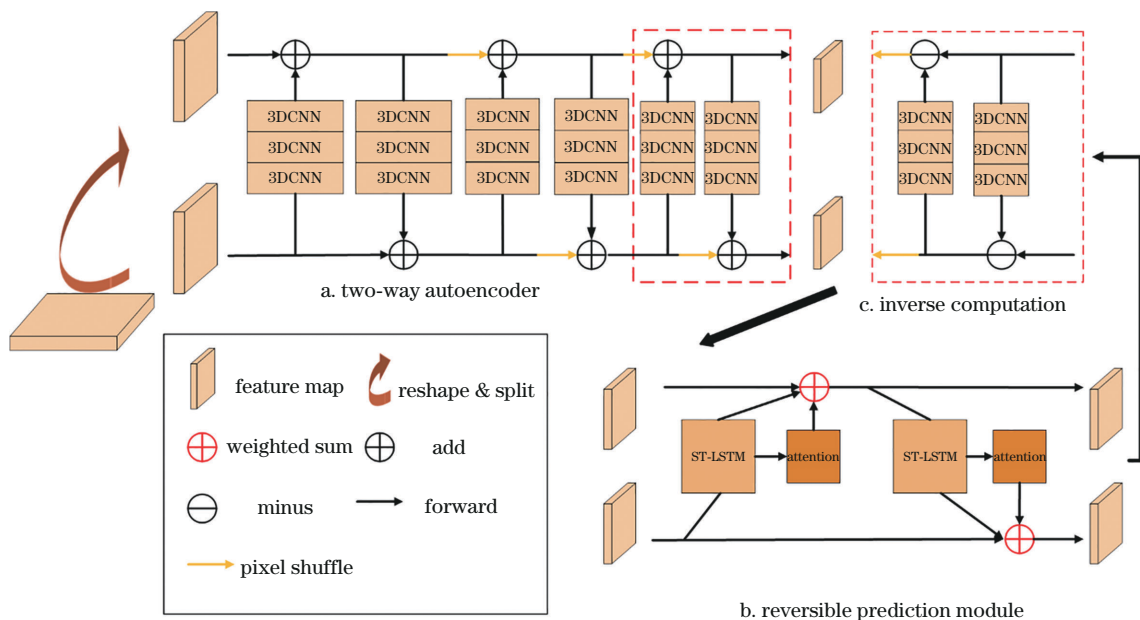


图 2 CrevNet 整体框架

Fig. 2 Overall framework of CrevNet

从云图序列中取出图像帧,将图像帧重构为两组图像数据输入双向自编码器进行特征提取,进而将提取的特征传递给双向循环预测模型,再将包含预测信息的特征数据传递给逆向计算模块进行解码运算,最终得到预测数据。设 $\mathbf{x}_t \in \mathbf{R}^{D \times W \times H \times C}$ 表示图像序列中的第 t 帧图像,其中, W, H, C, D 分别为图像帧的宽、高、通道数和深度。给定图像序列 $\mathbf{x}_{0:t-1}$ 输入 CrevNet 中得到的预测结果为 $\tilde{\mathbf{x}}_t$:

$$\tilde{\mathbf{x}}_t = E^{-1} \left\{ F \left[E(\mathbf{x}_{t-1}) | \mathbf{x}_{0:t-2} \right] \right\}, \quad (3)$$

式中: E 为自编码器; E^{-1} 为解码器(编码器的逆运算); F 为循环预测模块,用来桥接编码器和解码器。

SmartCrevNet 基于 CrevNet 改进而来,其框架如图 3 所示,该模型改进了 CrevNet 中的特征提取模块,可以强化双向自编码器的特征提取性能,增强了云图的语义信息。还设计了一种更适用于时序云图的预测模块 STA-GRU 来增强模型的预测性能。

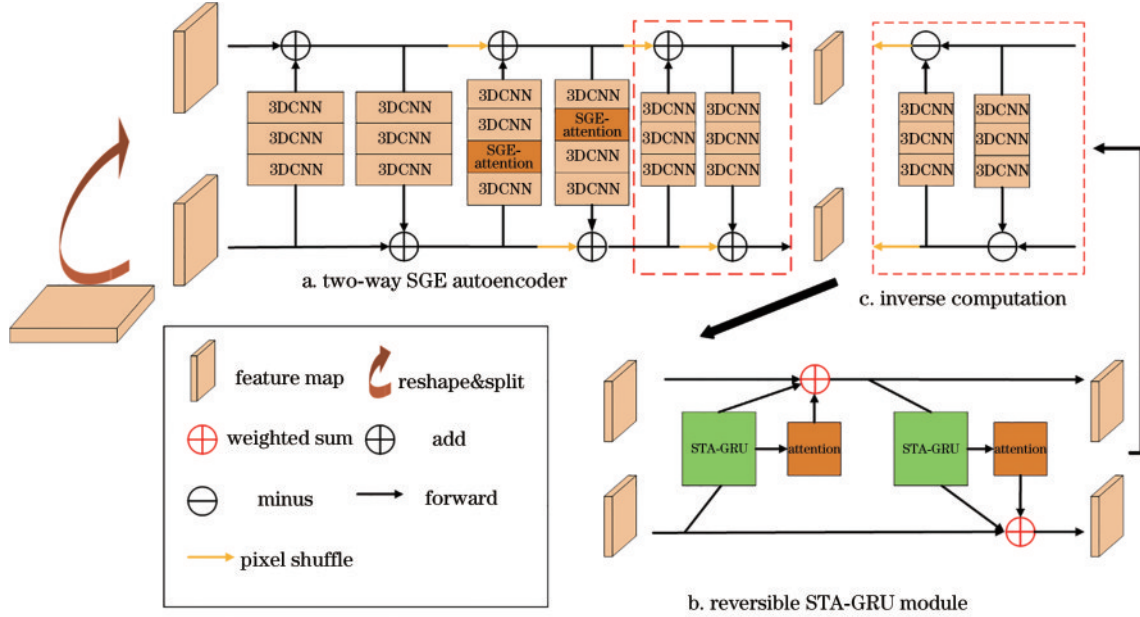


图 3 SmartCrevNet 整体框架
Fig. 3 Overall framework of SmartCrevNet

3.2 双向 SGE 自编码器

时序云图富含复杂而分散的云图特征信息,为充分获取特征,减少零散云图特征信息流失,在 CrevNet 的自编码器模型的基础上引入轻量型注意力模块 SGE,以增强特征提取时对云图语义特征的学习,高效提取云图特征信息。

双向 SGE 自编码器如图 4 所示,输入 \mathbf{x} 首先被重构为 \mathbf{x}_1 与 \mathbf{x}_2 两组。将 \mathbf{x}_1 与 \mathbf{x}_2 输入双向 SGE 自编码器

的两个端口, \mathbf{x}_1 经过多层 3D 卷积及激活等操作进行特征提取后与 \mathbf{x}_2 相加作为新的 $\tilde{\mathbf{x}}_2$ 。编码运算可表示为

$$\begin{cases} \tilde{\mathbf{x}}_2 = \mathbf{x}_2 + \delta_1(\mathbf{x}_1) \\ \tilde{\mathbf{x}}_1 = \mathbf{x}_1 + \delta_2(\mathbf{x}_2) \end{cases} \quad (4)$$

式中: $\tilde{\mathbf{x}}_2$ 和 $\tilde{\mathbf{x}}_1$ 为 \mathbf{x}_2 和 \mathbf{x}_1 更新后的值; δ 为多层 3D 卷积与激活的复合操作。式(4)为编码运算,解码运算是编码的逆运算,可表示为

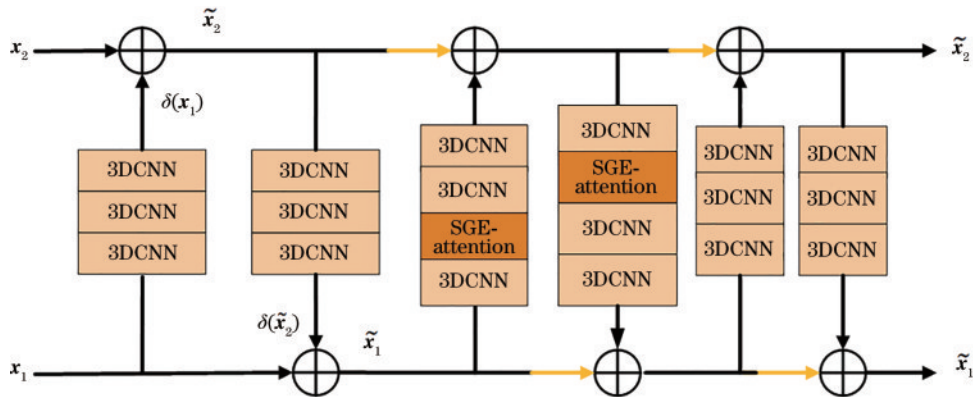


图 4 双向 SGE 自编码器
Fig. 4 Two-way SGE autoencoder

$$\begin{cases} \mathbf{x}_2 = \bar{\mathbf{x}}_2 - \delta_1(\mathbf{x}_1) \\ \mathbf{x}_1 = \bar{\mathbf{x}}_1 - \delta_2(\mathbf{x}_2) \end{cases} \quad (5)$$

将单个序列数据分成两组,交替提取特征再依次相加,可以极大减少特征提取过程中的信息丢失。为保证数据的可逆性和减少数据冗余,通过像素洗牌(pixel shuffle)^[19]将特征图像从 (w, h, c) 转变成 $(w/n, h/n, c \times n^2)$,既增加了通道数,又减小了特征图片大小。

增加特征图片通道数可以使网络从云图中学到更丰富的特征,但是云图运动的不稳定、非线性、易扩散等特点,使得 3D 卷积在提取云图特征时,会丢失部分新生成的或扩散出去的云图信息。因此,需要增强云图的语义特征。本文在像素洗牌后的 3D 卷积中引入轻量型注意力 SGE 模块,该模块在几乎不增加计算量和参数量的情况下,用局部信息和全局信息的相似性指导语义特征的增强,抑制云图中存在的噪声。

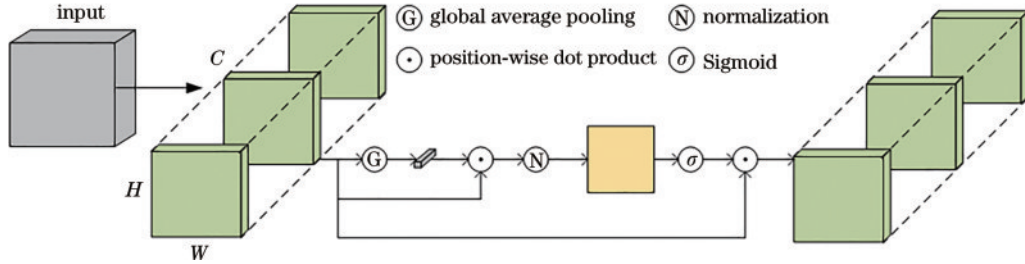


图 5 轻量型注意力模型 SGE
Fig. 5 Lightweight attention model SGE

该模块将输入的特征图按照通道数分成 n 个特征组,每组特征与其全局池化后的特征进行点积操作以此获取局部特征与全局特征的相似性,得到初始的注意力掩码,再经过归一化与 Sigmoid 激活函数后得到最终的注意力掩码,将该注意力掩码与原始特征组中每个位置的特征进行点积,获得语义增强后的特征数据。

3.3 时空注意力门控循环预测单元

所设计的预测模块 STA-GRU 在获取到双向 SGE 自编码模块输出的特征信息后,利用两组时空隐藏状态单元和历史信息模块充分获取前几帧云图的时空信息,可得到更好的预测结果。

STA-GRU 模型框架如图 6 所示。该模型将 ConvGRU 中的 2D 卷积换成 3D 卷积,使模型对目标的短时状态及局部运动的感知更加敏感。为充分利用云图序列中的时间信息与空间信息,STA-GRU 模型基于传统预测模型 ConvGRU,引入空间隐藏状态单元 M_t^i ,如图 6 中右半部分所示。空间隐藏状态单元 M_t^i 与原有的隐藏状态 H_t^i 分别负责结构中时间信息与空间信息的流通,使得两隐藏状态各司其职,缓解 H_t^i 存取时空信息的压力。另外 ConvGRU 仅用前一帧图预测后一帧图,会忽略历史时空信息,导致预测结果不理

3.2.1 轻量型注意力模块 SGE

在进行视频预测或者图像分类时都需要利用卷积神经网络提取图像特征信息,但是图像中不同局部特征的重要性不同,要使神经网络在特征提取时只关注特定的图像特征而非全部特征,就需要引入注意力机制^[20]。一个完整的特征图是由许多个子特征组成的,而子特征被划分到一个个特征组中,在特征提取时这些互有差异的特征组经过相同方式处理后,会受到背景噪声影响,从而导致提取到错误的图像特征。SGE 轻量级注意力模块中的每个特征组都会生成注意力掩码,这些注意力掩码代表当前子特征组内部各个子特征的重要程度,随着神经网络的不断训练,可以改善每个特征组内部语义特征的空间分布,针对性地学习语义信息和抑制噪声。另外 SGE 中的注意力掩码仅由每个子特征组内的全局和局部特征的相似性来决定,所以几乎不增加训练负担。轻量型注意力模型 SGE 结构如图 5 所示。

想。因此,在模型中引入历史信息模块(如图 6 中左上部所示),该模块可以充分利用历史隐藏状态 $H_{t-\tau,t-1}^i$ 提升网络对目标整体运动感知的敏感度。STA-GRU 模型可表示为

$$\begin{cases} R_t = \sigma(X_t * W_{xr} + H_{t-1}^i * W_{hr} + b_r) \\ Z_t = \sigma(X_t * W_{xz} + H_{t-1}^i * W_{hz} + b_z) \\ \tilde{H}_t^i = \tanh[X_t * W_{xh} + (R_t \odot H_{t-1}^i) W_{hh} + b_h] \\ S_{\text{attention}} = \text{Softmax}[R_t \cdot (H_{t-\tau,t-1}^i)^T] \cdot H_{t-\tau,t-1}^i \\ H_t^i = Z_t \odot \text{LayerNorm}(H_{t-1}^i + S_{\text{attention}}) + (1 - Z_t) \odot \tilde{H}_t^i \end{cases} \quad (6)$$

式中: σ 与 \tanh 分别是 Sigmoid 激活函数和 \tanh 激活函数; Softmax 为归一化指数函数, $*$ 代表 3D 卷积操作; \odot 与 \cdot 分别表示哈达玛卷积和矩阵乘积。 R_t 、 Z_t 与 \tilde{H}_t^i 分别代表隐藏状态 H_t^i 的重置门、更新门和候选隐藏状态。式(6)第 4 行为历史信息模块的公式化,在历史信息模块中,将包含有前 τ 帧云图信息的历史隐藏状态 $H_{t-\tau,t-1}^i$ 与 R_t 相乘,再经过归一化指数函数后获得 τ 帧历史云图的信息权重,这些权重信息与历史隐藏状态对应相乘获得整体的注意力分数($S_{\text{attention}}$),再将注意力

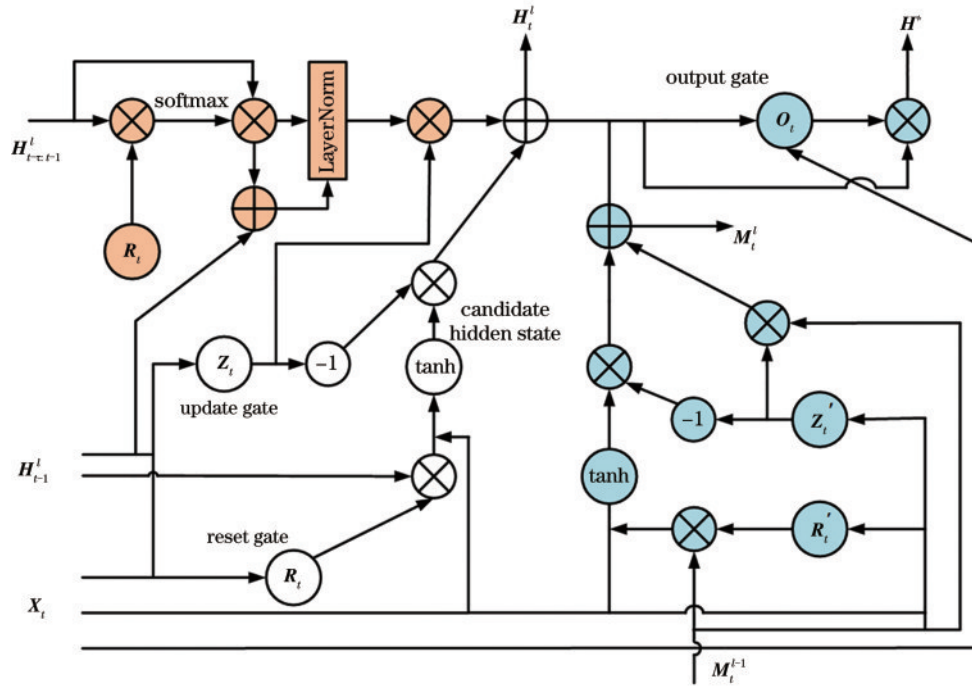


图 6 STA-GRU 结构

Fig. 6 STA-GRU structure

分数与前一时刻的隐藏状态 H_{t-1}^i 由 LayerNorm (层归一化) 整合, 完成对过去 τ 帧云图隐藏状态的信息选择。

$$\left\{ \begin{array}{l} R_t^i = \sigma(X_t * W_{xr}^i + M_{t-1}^{i-1} * W_{mr} + b_r^i) \\ Z_t^i = \sigma(X_t * W_{xz}^i + M_{t-1}^{i-1} * W_{mz} + b_z^i) \\ \tilde{M}_t^i = \tanh[X_t * W_{xm} + (R_t^i \odot M_{t-1}^{i-1}) W_{mm} + b_m] \\ M_t^i = Z_t^i \odot M_{t-1}^{i-1} + (1 - Z_t^i) \odot \tilde{M}_t^i \\ O_t = \sigma(X_t * W_{xo} + H_t^i * W_{ho} + M_t^i * W_{mo} + b_o) \\ H^i = O_t \odot \tanh(W_{1 \times 1} * [H_t^i, M_t^i]) \end{array} \right. \quad (7)$$

模型在原始 ConvGRU 的基础上增加一组新的重

置门 R^i 和更新门 Z^i 以指导构建空间隐藏状态单元 M^i 。 M^i 可在网络中分担隐藏状态 H^i 的压力, 获取更多时空运动信息。最后将 H^i 和 M^i 连接起来并利用 1×1 卷积降维, 再经过输出门 O , 获得最终的融合时空隐藏状态 H^* 。

3.3.1 可逆预测结构

预测结构的级联方式与自编码器类似, 也是使用循环输入交替预测再相加的方式, 以此来减少预测数据流通过程中的信息损失, 同时也确保输入和输出之间的像素对齐, 以兼容双向 SGE 自编码器的可逆性, 可逆预测结构如图 7 所示。

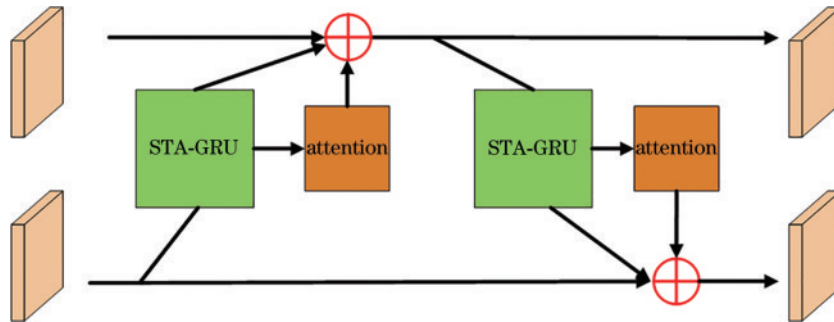


图 7 可逆 STA-GRU 预测结构

Fig. 7 Reversible STA-GRU module

可逆预测结构可表示为

$$\left\{ \begin{array}{l} H^* = \text{STA-GRU}(x_t^1, H_{t-1}^1, M_{t-1}^1) \\ G_t = \sigma[W_2 * \text{ReLU}(W_1 * H^* + b_1) + b_2] \\ \hat{x}_t^2 = (1 - G_t) \odot x_t^2 + G_t \odot H^* \end{array} \right. \quad (8)$$

式中: x_t^1, x_t^2 为单条图像帧的两组特征; H^* 为 STA-GRU

输出的融合隐藏状态数据; G_t 为一种软注意力模型。

4 实验与分析

4.1 实验环境及模型参数

使用 FY-4A 卫星云图数据集以及公共数据集 MovingMNIST 进行实验。在 SmartCrevNet 中, 隐藏

层通道数设为 32,双向自编码器循环层数设置为 36,可逆预测模块循环层数设置为 8。所有的模型均使用 L1+L2 损失函数,并利用 Adam^[21] 优化器进行训练,初始学习率为 5×10^{-4} ,设置了自适应学习率模块 StepLR。实验在 PyTorch 上实现,并使用 NVIDIA GeForce RTX 3060 12G GPU 训练。

4.2 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移动速度更快,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代表性、挑战性。FY-4A 是中国第 2 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 6 月—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 3488 张云图,按照序列长度为 8 滑动取值,共获得 3481 个云图时间序列,随机选择 2780 个时间序列作为训练集,701 个时间序列作为验证集。云图样本的经纬度范围为 $115^{\circ}2'E \sim 123^{\circ}5'E$ 、 $26^{\circ}6'N \sim 35^{\circ}2'N$,像素大小为 256×256 ,空间分辨率为 4 km,STA-GRU 中的 τ 设置为 4。

经反复实验评估,在已有的实验算力情况下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,在气象部门业务应用上,4 h 跨度的云图预测可以应用到气象灾害突发预警、飞机航线规划、农作物耕作规划等领域。综上所述,将云图预测的时间跨度设置为 4 h,即利用前 4 h 的云图帧来预测后 4 h 的云图帧。

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) 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 $-1 \sim 1$ 之间,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。

4.2.1 云图数据集模型对照实验及结果分析

将 SmartCrevNet 模型分别与经典视频预测模型 ConvLSTM、ConvGRU、PredRNN、PredRNN++ 以及新型预测模型 CrevNet 进行对照实验。各模型定量评价如表 1 所示,所有指标均是预测的 4 个帧的平均值,更低的 MSE、MAE 与更高的 SSIM、PSNR 代表图像预测精度更高。

表 1 不同预测算法(4 帧→4 帧)的定量评价表

Table 1 Comparative analysis of various prediction algorithms (4 frames→4 frames)

Method(4→4)	MSE / $10^{-3} \downarrow$	MAE / $10^{-3} \downarrow$	SSIM \uparrow	PSNR \uparrow
ConvLSTM	27.41	104.19	0.448	15.62
ConvGRU	26.81	99.07	0.444	15.72
PredRNN	25.30	97.11	0.457	15.97
PredRNN++	25.11	95.06	0.451	16.00
CrevNet	19.64	72.41	0.482	17.07
SmartCrevNet	18.21	67.34	0.520	17.40

由表 1 可知,SmartCrevNet 在 MSE、MAE、SSIM、PSNR 等 4 种指标上均取得了最好的效果。预测云图序列示例如图 8 所示,其中, input 为输入的前 4 帧云图, G.T. 为实际的后 4 帧云图。ConvLSTM 与 ConvGRU 作为最传统的预测模型,各项评价指标均为最低水平,预测效果并不理想,特别是 ConvLSTM 的预测图像含有较多的噪声信息。PredRNN 将时间与空间信息区分开,利用两个 LSTM 分别处理时间信息与空间信息,相比于 ConvLSTM 更有效捕捉到了云图时空信息。PredRNN++ 与 PredRNN 类似,通过级联 LSTM 构建了新的预测模型,同时加入梯度高速公路单元进一步缓解了梯度消失问题,对云运动趋势的预测能力进一步增强,其预测精度相比 PredRNN 也有所提高。CrevNet 使用的预测模块与 PredRNN 相同,双向自编码器与可逆预测模块极大地减少了特征提取和序列预测过程中的信息损失,减少内存占用,使得 CrevNet 的预测结果强于 PredRNN。SmartCrevNet 与 CrevNet 因双向编码器的存在都能很清晰地预测前两帧云图,但 SmartCrevNet 包含了改进后的双向 SGE 自编码器以及预测模型 STA-GRU,使其具有更好的长期预测能力,因此 SmartCrevNet 后两帧云图的预测效果优于 CrevNet。

各模型有关 MSE 与 SSIM 的逐帧分析如图 9 所示,可以看出,在 SSIM 评价标准下,随着预测时间的深入,SmartCrevNet 模型预测精度的衰减是最缓慢的。

4.2.2 云图数据集消融实验及结果分析

为进一步分析 SmartCrevNet 中新模块的作用,对双向自编码器及可逆预测模块进行消融实验,实验研究包括移除其中一个模块或替换为另一个模块。实验结果如表 2 所示。

实验使用了相同的数据集并设置了相同的实验参数,表 2 中的 w/o SGE 代表当前模型没有添加 SGE 模块, SmartCrevNet+SGE+HIM 为所提完整 SmartCrevNet 模型。SGE 可以指导特征提取过程图像中的语义增强,为检验 SGE 在 SmartCrevNet 中的作用,将 SGE 从模型中去除,与完整的 SmartCrevNet 进行对比。从表 2 可知,去除 SGE 的 SmartCrevNet 模型的各项评价指标与完整的 SmartCrevNet 相比均有退化,表明 SGE 对预测起正向作用。

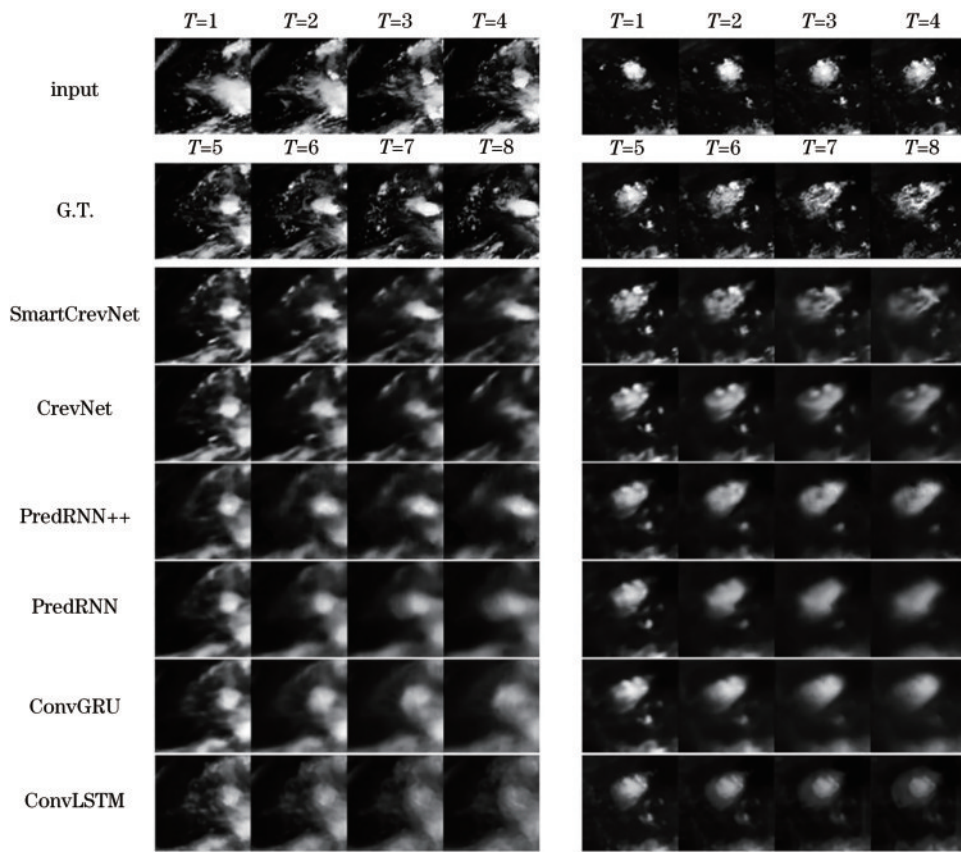


图 8 卫星云图的两组预测示例

Fig. 8 Two prediction examples from satellite cloud images

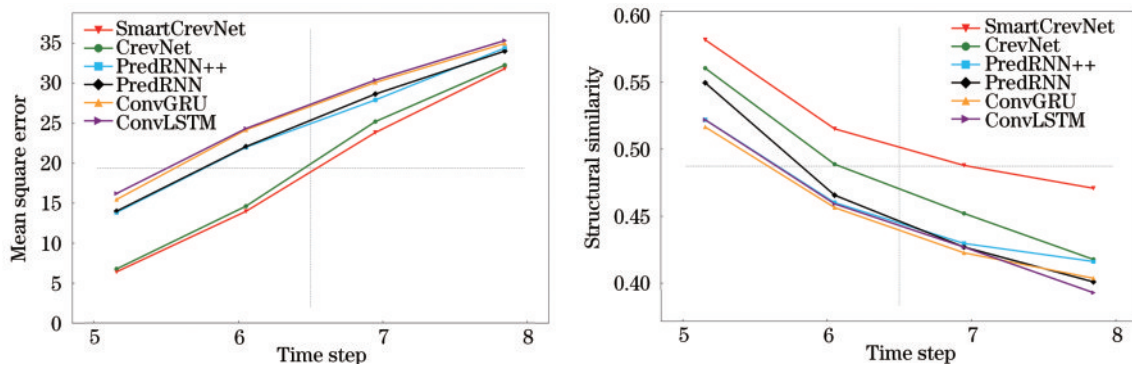


图 9 卫星云图数据集上不同模型 MSE 与 SSIM 的预测逐帧比较

Fig. 9 Frame-wise MSE and SSIM comparison of the different models on the satellite cloud map dataset

表 2 双向自编码器及预测模块的消融实验

Table 2 Two-way autoencoder and prediction module ablation experiments

Model(4→4)	MSE / 10 ⁻³ ↓	MAE / 10 ⁻³ ↓	SSIM ↑	PSNR ↑
CrevNet	19.64	72.41	0.482	17.07
SmartCrevNet w/o SGE	19.47	69.36	0.518	17.11
SmartCrevNet w/o HIM	19.64	69.16	0.518	17.06
SmartCrevNet+SE+HIM	19.13	68.08	0.511	17.18
SmartCrevNet+CBAM+HIM	19.34	69.68	0.504	17.14
SmartCrevNet+SGE+HIM	18.21	67.34	0.520	17.40

实验还尝试了两个与 SGE 具有类似作用的注意力模块 convolutional block attention module (CBAM)^[22] 以及 squeeze-and-excitation (SE)^[23]。从整体的评估指标来看,SGE 取得了比较好的效果。将 SmartCrevNet 中的 SGE 去除后,此时该模型与 CrevNet 仅存在预测模块上的差异,SmartCrevNet 的预测模块为 STA-GRU,而 CrevNet 的预测模块为 ST-LSTM,通过两组实验数据对比可知,SmartCrevNet 的预测模块要优于 CrevNet 的预测模块。为继续验证预测模块中 HIM 的作用,将 HIM 取消掉,此时 SmartCrevNet 的预测模块为 ST-GRU,预测时仅考虑前一帧云图信息,忽视历史云图信息,表 2 中消融实验将无 HIM 的模型与完整

的 SmartCrevNet 进行对比,体现了 HIM 对 SmartCrevNet 的重要作用。

从表 2 实验结果可以看出,完整 SmartCrevNet 模型整体优于消融实验的其他模型,说明了所提模型用于云图预测的可行性与有效性。

4.3 MovingMNIST 数据集实验

MovingMNIST 移动手写数字数据集是视频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 64×64 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,每个序列都是独一无二的。数据集设置为 10000 个序列,其中,训练集由 8000 个序列组成,测试集由 2000 个序列组成,长度为 10, batch size 设置为 16。

4.3.1 MovingMNIST 模型对照实验及结果分析

将 SmartCrevNet 分别与 ConvLSTM、PredRNN、PredRNN++ 和 CrevNet 在 MovingMNIST 数据集上进行实验对比。MovingMNIST 数据集对比实验及预测示例如图 10 所示。

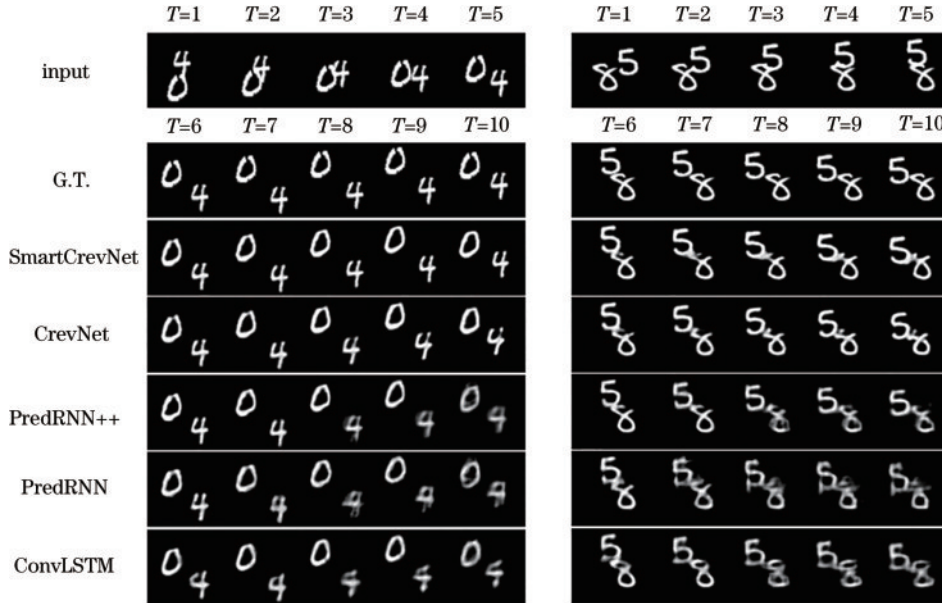


图 10 两组预测示例

Fig. 10 Two sets of prediction examples

表 3 不同方法在 MovingMNIST 数据集上的定量评估
Table 3 Quantitative evaluation of different methods on Moving MNIST

Model	MSE / 10^{-3} ↓	SSIM ↑
ConvLSTM	11.562	0.882
PreRNN	9.496	0.905
PredRNN++	8.344	0.917
CrevNet	5.603	0.925
SmartCrevNet	5.261	0.936

在 MovingMNIST 数据集上, SmartCrevNet 的 MSE 相较于 CrevNet 降低了 6.1%, SSIM 提升了 1.2%, 预测结果均优于 CrevNet 及传统的视频预测算法。MovingMNIST 数据集内部图像特征及移动方式都较为单一,各模型预测结果明显好于云图预测数据集。

图 10 实验对比了每个模型在 MovingMNIST 实际预测中的表现。两组数字分别在输入阶段与输出阶段有融合的趋势。“04”的预测较为简单,在此预测中每个模型都较好地预测到了数字移动的趋势,但是 SmartCrevNet 预测的 5 张图的清晰度与数字的完整性均高于其他模型。“58”的预测中,输入序列图像内的

“5”与“8”一直处于紧贴状态,预测时两个数字的信息会互相干扰,增加了预测的难度, ConvLSTM、PredRNN、PredRNN++ 的预测图像清晰度较低且输出的预测数字有不同程度的崩坏, CrevNet 的预测保证了预测图像的清晰度,但“5”在预测过程中逐渐变形,“8”没有和“5”很好地区分开来。相比而言, SmartCrevNet 的预测结果更为精确,保证了数字运动趋势的准确性、预测结果的清晰度。

5 结 论

在 CrevNet 预测模型的基础上设计了一种基于时空注意力门控循环单元的特征增强卫星云图预测模型。该模型改进了 CrevNet 中的特征提取模块,在不增加计算量的情况下增强特征提取中云图的语义信息,还设计了一种更适合时序云图的预测模块 STA-GRU,使得模型预测时能充分利用历史云图信息,以此得到更好的预测结果。实验结果表明, SmartCrevNet 在卫星云图数据集和 MovingMNIST 移动手写数字数据集上的预测结果优于原模型 CrevNet 及 PredRNN 等传统模型。

此外,云图预测的重要难点在于大尺度卫星云图预测时资源占用过大,且预测清晰度较低。本研究在实验中仅仅使用了 256×256 的云图影像进行预测,可能难以适用于更大的云图影像预测。因此,减少大尺度云图特征提取时的资源占用及信息损失,并输出更加清晰的预测云图,将是今后研究工作的努力方向。

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[EB/OL]. (2014-06-03) [2023-02-06]. <https://arxiv.org/abs/1406.1078>.
- [3] Ranzato M, Szuam A, Bruna J, et al. Video (language) modeling: a baseline for generative models of natural videos[EB/OL]. (2014-12-20)[2023-02-06]. <https://arxiv.org/abs/1412.6604>.
- [4] Srivastava N, Mansimov E, Salakhutdinov R. Unsupervised learning of video representations using LSTMs[C]//Proceedings of the 32nd International Conference on International Conference on Machine Learning-Volume 37, July 6-11, 2015, Lille, France. New York: ACM Press, 2015: 843-852.
- [5] Shi X J, Chen Z R, Wang H, et al. Convolutional LSTM Network: a machine learning approach for precipitation nowcasting[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems -Volume 1, December 7-12, 2015, Montreal, Canada. New York: ACM Press, 2015: 802-810.
- [6] Shi X J, Gao Z H, Lausen L, et al. Deep learning for precipitation nowcasting: a benchmark and a new model [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems, December 4-9, 2017, Long Beach, CA, USA. New York: ACM Press, 2017: 5622-5632.
- [7] Wang Y B, Long M S, Wang J M, et al. PredRNN: recurrent neural networks for predictive learning using spatiotemporal LSTMs[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems, December 4-9, 2017, Long Beach, California, USA. New York: ACM Press, 2017: 879-888.
- [8] Wang Y B, Gao Z F, Long M S, et al. PredRNN++: towards a resolution of the deep-in-time dilemma in spatiotemporal predictive learning[EB/OL]. (2018-04-17) [2023-02-03]. <https://arxiv.org/abs/1804.06300>.
- [9] Wang Y, Jiang L, Yang M H, et al. Eidetic 3D LSTM: a model for video prediction and beyond[C]//7th International Conference on Learning Representations, ICLR 2019, May 6-9, 2019, New Orleans, LA, USA. La Jolla: ICLR, 2019.
- [10] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. *Communications of the ACM*, 2020, 63(11): 139-144.
- [11] Vaswani A, Shazeer N, Parmar N, et al. Attention is all You need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems, December 4-9, 2017, Long Beach, California, USA. New York: ACM Press, 2017: 6000-6010.
- [12] Xu Z, Du J, Wang J J, et al. Satellite image prediction relying on GAN and LSTM neural networks[C]//ICC 2019-2019 IEEE International Conference on Communications (ICC), May 20-24, 2019, Shanghai, China. New York: IEEE Press, 2019.
- [13] Lin Z H, Li M M, Zheng Z B, et al. Self-attention ConvLSTM for spatiotemporal prediction[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(7): 11531-11538.
- [14] Yu W, Lu Y, Easterbrook S, et al. Efficient and information-preserving future frame prediction and beyond [C]//8th International Conference on Learning Representations, ICLR 2020, April 26-30, 2020, Addis Ababa, Ethiopia. La Jolla: ICLR, 2020.
- [15] Hong S, Kim S, Joh M, et al. PSIQUE: next sequence prediction of satellite images using a convolutional sequence-to-sequence network[EB/OL]. (2017-11-29) [2023-02-05]. <https://arxiv.org/abs/1711.10644>.
- [16] 蔡朋艳. 基于 FY-4A 卫星的云检测与云图预测方法研究[D]. 南京: 南京信息工程大学, 2021.
Cai P Y. Research on cloud detection and cloud image prediction method based on FY-4A satellite[D]. Nanjing: Nanjing University of Information Science & Technology, 2021.
- [17] Liu K, Liu W, Gan C A, et al. T-C3D: temporal convolutional 3D network for real-time action recognition [J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018, 32(1): 12333.
- [18] Li X, Hu X L, Yang J. Spatial group-wise enhance: improving semantic feature learning in convolutional networks[EB/OL]. (2019-05-23) [2023-02-03]. <https://arxiv.org/abs/1905.09646>.
- [19] Shi W Z, Caballero J, Huszár F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 1874-1883.
- [20] Mnih V, Heess N, Graves A, et al. Recurrent models of visual attention[EB/OL]. (2014-06-24) [2023-02-03]. <https://arxiv.org/abs/1406.6247>.
- [21] Kingma D P, Ba J. Adam: a method for stochastic optimization[EB/OL]. (2014-12-22) [2023-02-03]. <https://arxiv.org/abs/1412.6980>.
- [22] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[M]//Ferrari V, Hebert M, Sminchisescu C, et al. *Computer vision-ECCV 2018. Lecture notes in computer science*. Cham: Springer, 2018: 3-19.
- [23] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7132-7141.