

基于多尺度运动记忆模型的遥感云图预测方法

张永宏^{1,2}, 孙书林¹, 龚蒙¹, 王俊飞³, 马光义³

(1. 南京信息工程大学自动化学院, 江苏 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, 江苏 南京 210044;

3. 南京信息工程大学电子与信息学院, 江苏 南京 210044)

摘要: 针对现有深度学习模型难以捕获云团运动模式导致云图长期预测结果模糊、准确度低的问题, 提出一种基于多尺度运动记忆模型(MSMM_Net)的遥感云图预测方法。该模型采用空间多尺度记忆流和运动差分记忆流相融合的双分支记忆流架构, 分别提取输入图片序列隐含的高低频空间特征和序列运动特征, 从而同时获得图片的全局信息、细节信息和运动信息, 在预测阶段融合双分支记忆, 缓解特征丢失问题并增强模型对云团运动轨迹的预测能力。在此基础上, 使用像素损失和边缘损失相结合的融合损失函数指导模型的训练, 强化模型对图片边缘细节的关注度, 促使模型生成清晰的预测图片。实验结果表明, 与基准模型 PredRNN 相比, MSMM_Net 在 Moving MNIST 数据集上的均方误差(MSE)降低了 31.71%, 在可学习感知图像块相似性指标(LPIPS)上降低了 64.7%, 在遥感卫星云图数据集上, 峰值信噪比(PSNR)和结构相似性指数(SSIM)指标分别提升了 5.51% 和 5.38%, 表明该模型生成的预测图片序列与真实图片序列更加相似, 能够有效提升长期预测准确率。

关键词: 云图预测; 时空序列预测; 深度学习; 循环卷积网络; 遥感卫星

源代码链接: https://github.com/BlueBerrrry/MSMM_Net

中图分类号: TP391

文献标志码: A

DOI: 10.19678/j.issn.1000-3428.0069950

Remote Sensing Cloud Image Prediction Method Based on Multi-scale Motion Memory Model

ZHANG Yonghong^{1,2}, SUN Shulin¹, GONG Meng¹, WANG Junfei³, MA Guangyi³

(1. School of Automation, Nanjing University of Information Science and Technology, Nanjing 210044, Jiangsu, China;

2. Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology,

Nanjing University of Information Science and Technology, Nanjing 210044, Jiangsu, China;

3. School of Electronic and Information Engineering,

Nanjing University of Information Science and Technology, Nanjing 210044, Jiangsu, China)

【Abstract】 Existing deep learning models find it difficult to capture cloud motion patterns, resulting in long-term cloud prediction results that are fuzzy and low in accuracy. To address this problem, this study proposes a remote sensing cloud image prediction method based on a Multi-Scale Motion Memory Network (MSMM_Net). This model adopts a dual-branch memory-flow architecture that combines spatial multi-scale and motion-differential memory flows. It extracts high- and low-frequency spatial features and sequence motion features hidden in the input image sequence, thereby simultaneously obtaining global, detail, and motion information of the image. In the prediction stage, dual-branch memory is fused to alleviate the problem of feature loss and enhance the ability of the model to predict the trajectory of cloud clusters. On this basis, a fusion loss function combining pixel and edge losses is used to guide model training, enhance the model's attention to image edge details, and promote the generation of clear predicted images. Experimental results show that, compared with the benchmark model PredRNN, MSMM_Net reduces the Mean Square Error (MSE) by 31.71% on the Moving MNIST dataset and the Learned Perceptual Image Patch Similarity (LPIPS) by 64.7%. On the remote sensing satellite cloud image dataset, the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) indicators improve by 5.51% and 5.38%, respectively, indicating that the predicted image sequence generated by the model is more similar to the real image sequence and can effectively improve long-term prediction accuracy.

【Key words】 cloud image prediction; spatio-temporal sequence prediction; deep learning; recurrent convolutional network; remote sensing satellite

基金项目: 国家重点研发计划(2021YFE0116900); 国家自然科学基金(42175157); 风云应用开创性项目(FY-APP-2022.0604)。

作者简介: 张永宏, 男, 教授、博士, 主研方向为遥感大数据分析、深度学习; 孙书林、龚蒙, 本科生; 王俊飞, 硕士研究生; 马光义, 博士研究生。

收稿日期: 2024-06-03

修回日期: 2024-08-20

E-mail: zyh@nuist.edu.cn

0 引言

卫星云图是由地球同步卫星上搭载的遥感仪器拍摄得到的云层图像,蕴含着丰富的气象信息^[1]。在气象预报领域,覆盖范围广、时间分辨率高的卫星云图^[2]已经成为一种非常重要的数据获取方式,在海雾监测^[3]、降水估计^[4-5]、台风特征分析^[6-7]与等级预测^[8]、太阳能可用性分析^[9-10]等多个领域都发挥着关键作用。其中,云光学厚度、云顶温度等云结构特征参数直接影响着降水概率和降水强度^[11],因此,有效提取卫星云图序列的动态变化特征并实现卫星云图的准确预测,对降水临近预报具有重要意义。

块匹配法和光流法是云图预测领域常用的传统方法,但是块匹配法建立的线性模型难以准确模拟云团运动,光流法的计算也极易出现误差。随着深度学习的飞速发展,使用神经网络解决云图预测问题也成为热门研究方向,并取得了一定的成果^[12-15],证明了深度学习方法在云图预测上的巨大潜力。然而,现有大多数研究使用的卫星数据时间分辨率不高,且模型设计无法有效提取云图的多尺度特征,也没有充分利用图片序列包含的运动特性,导致准确预报时效较短,长期预测的图片严重模糊,不能观测到有效的云结构特征。

针对上述问题,本文构建了一种新的多尺度运动记忆网络(MSMM_Net)模型。将运动差分注意力模块(MDAM)嵌入运动差分时空记忆单元(MDSTM)中,以增强模型对图片序列运动特征的提取能力。同时,堆叠不同隐藏状态数的 MDAM 单元,分别提取图片的高低频特征,由空间多尺度记忆流和运动差分记忆流在网络中传递多尺度空间特征,最后在预测阶段通过跳跃连接融合不同尺度的特征,根据图像序列的多尺度时空信息生成该时间步的预测图像,并将此时的记忆信息传入下一时间步,以降低长期预测产生的叠加误差。在训练过程中,使用像素损失和边缘损失相融合的损失函数,从像素和边缘 2 个维度衡量预测图片序列和真实图片序列之间的差异,促使模型在学习图片内容的同时保留高频特征,生成清晰的预测图片。在 Moving MNIST 数据集和高时间分辨率的 GOES-16 卫星云图数据集上进行对比实验,以验证 MSMM_Net 在长期预测中的准确率提升效果。

1 相关研究

传统的云图预测方法主要包括块匹配法和光流

法。块匹配法使用从天空成像设备[如全天空成像仪(WISIs)^[16]]中获得的云运动矢量(CMV_s),根据当前帧与下一帧的重叠情况来预测云团的运动。文献[17]结合块匹配算法、最佳候选块搜索法和向量中值正则化开发了云运动估计模型,成功使用卫星红外图像估计了云的运动场。文献[18]将互相关法(CCM)和互谱分析(CSA)作为匹配准则来估计云的运动,并额外增加了质量控制措施,提高了预测可靠性。然而,使用块匹配法建立的模型是线性的,无法精确模拟云团的非线性运动。光流是指空间中的运动物体在被观测平面上像素点的瞬时运动速度,可以描述观测物体在时间序列图像之间运动时发生的图像亮度模式变化^[19]。根据光流原理计算云团的运动速度,能够建立符合云团运动特征的非线性模型。文献[20]使用光流法跟踪云特征,根据特征轨迹得出云的方向和速度等运动信息。文献[21]使用互相关法和光流法跟踪地面云团的运动并进行预测。文献[22]提出了一种变分光流(VOF)的方法来确定云运动的亚像素精度。文献[23]则利用密集光流模型结合 DIS(Dense Inverse Search)光流算法和后向平流方案,来外推红外卫星云图。光流法的核心是光流场的计算,但是云图序列中云团的灰度会随着时间的推移而发生变化,从而影响计算结果,导致预测出现误差。

卫星云图预测任务是基于输入的历史卫星云图序列,预测未来时刻的卫星云图序列,其本质上属于时空序列预测问题。近年来,随着机器学习技术的蓬勃发展,许多研究者开始使用深度学习方法来解决时空序列预测问题。文献[24]使用卷积运算取代全连接长短期记忆网络(FC-LSTM)中的哈德曼卷积运算,构建可以处理图像序列的卷积长短期记忆网络(ConvLSTM),使模型能够同时学习到图像序列的空间和时间信息。然而,标准卷积只能不断提取固定区域的空间信息,将其用于提取位置频繁变化的自然运动时效率不高。为此,文献[25]对编码器-解码器结构进行改进,并使用新的可学习卷积构建轨迹门循环单元(TrajGRU)模型,增强了模型的抗失真性。但是,这样的结构使得逐层提取的空间信息仅在各个层之间随时间流传递,下一时间步的底层接收不到高层空间信息,文献[26]认为时空预测学习应该在统一的记忆池中记忆空间特征和时间变化,因此,他们将信息流改为“之”字形,并且将 LSTM 单元替换为可同时提取记忆状态和时间表示的时空长短期记忆单元(ST-LSTM),提出了 PredRNN 模型。这个改进虽然提高了模型的记忆

能力,但同时面临着梯度消失的问题。为此,文献[27]又使用新的 Casual LSTM 单元和梯度高速单元(GHU)构建了 PredRNN++模型,缓解了梯度消失问题,使模型能够自适应地学习长短期信息。除以上基于循环神经网络(RNN)结构的模型之外,文献[28]使用 U-Net 作为基础结构,结合注意力机制并使用更轻量的深度可分离卷积,提出了 SmaAt-UNet 网络,较好地预测了云层和降水图像的二值图像。

2 研究方法

图片序列预测问题需要模型完成对输入图片序

列的特征提取与特征重建 2 个步骤,这不仅要求网络充分获取每一帧图片的空间信息,而且同时要捕获图片序列之间的运动趋势,并随着时间步的递增保留这些重要特征的记忆,从而在特征重建时进行正确预测。因此,本文提出一个端到端的多尺度运动记忆网络,其整体框架如图 1 所示。编码器接收当前时间步的输入图片 X_t 及其对应的运动差分图 Diff_X ,与上一时间步的存储单元 C 、水平传递的时间状态 T 、隐藏状态 H 一起进入编码阶段进行特征提取,使用双分支记忆流(即空间多尺度记忆流 S 和运动差分记忆流 M)分别传递图片序列的空间特

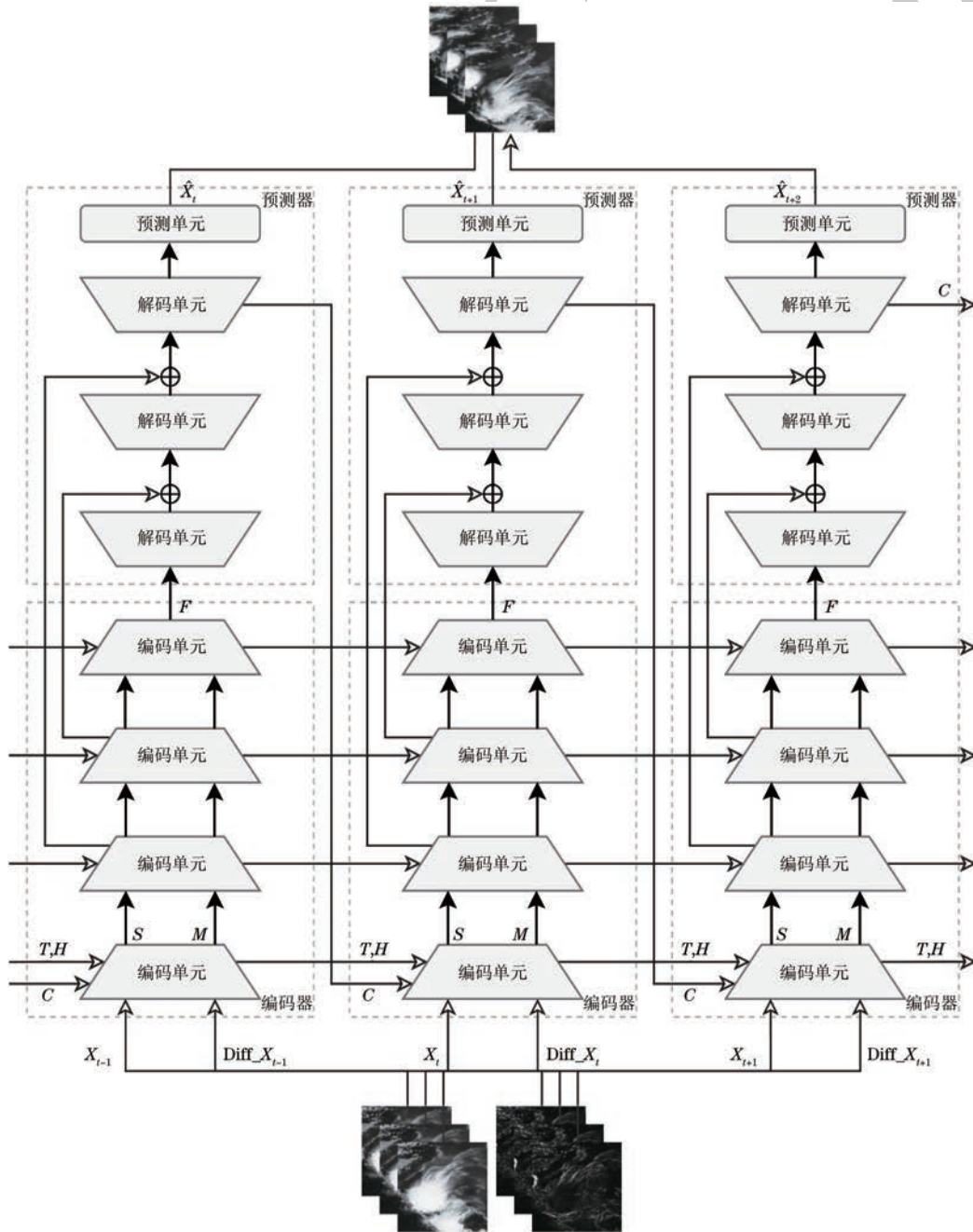


图 1 多尺度运动记忆网络整体结构

Fig.1 Overall structure of multi-scale motion memory network

征和时间特征,最后一个编码单元负责融合双分支记忆,得到的时空融合记忆 F 流入预测器,用于在预测阶段重建时空特征,生成预测的下一时间步图片 \hat{X}_{t+1} 并将记忆传递给下一时间步,如此递推,生成每一个图片帧的预测结果,最终得到预测图片序列。

2.1 空间多尺度记忆流

PredRNN^[26] 的网络结构如图 2 所示,它使用 ST-LSTM 单元同时保存时间和空间信息,并采用“之”字形记忆流将前序时间步的高层空间信息传输给后序时间步底层,有效提高了时空序列预测的准确度。但是, PredRNN 仅使用相同隐藏状态数的 ST-LSTM 单元进行堆叠,只能不断重复提取相同尺度的特征,对于具有丰富信息的卫星云图而言,在不改变特征图大小的情况下提取出的特征信息都是相同层次的,网络很难同时捕获到卫星云图包含的底层运动变化信息和高层云团细节信息,导致长期准确预测难以维持。因此,本文将 ST-LSTM 单元扩展为 MDSTM 单元,其具体结构如图 3 所示,并与卷积块、最大池化块一起封装为编码单元,每个编码单元中 MDSTM 单元的隐藏状态数都是不同的,以提取图片不同尺度的空间特征,这些特征在网络的层与层之间流动更新,构成空间多尺度记忆流,帮助网络同时获取图片的高低频信息。

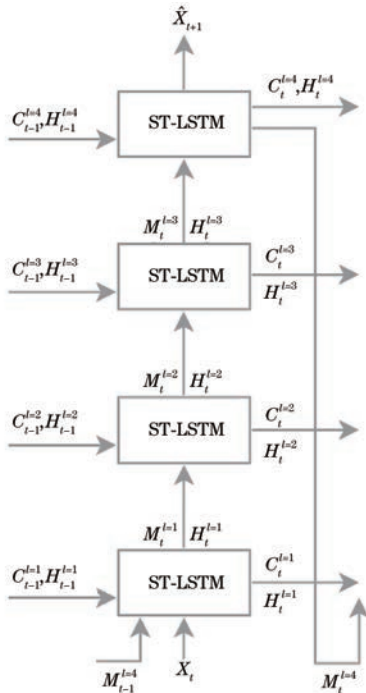


图 2 PredRNN 网络结构

Fig. 2 PredRNN network structure

MDSTM 单元在每一个时间步的第一层接收对应的输入图片 X_t 、差分运动图 Diff_X 和隐藏状

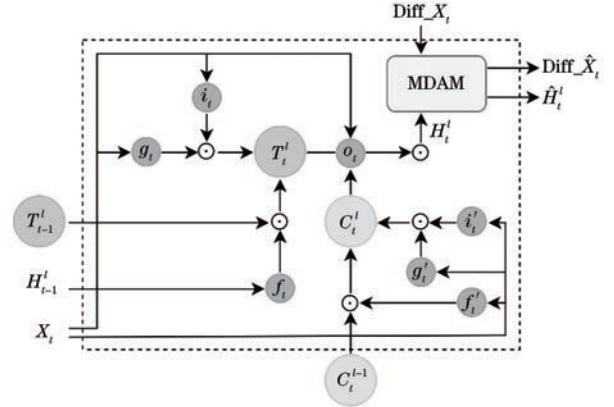


图 3 MDSTM 结构

Fig. 3 MDSTM structure

态 H , 使用门控机制提取特征信息,其具体计算公式如下:

$$\begin{cases}
 g_t = \tanh(W_{xg} * X_t + W_{hg} * H_{t-1}^i + b_g) \\
 i_t = \sigma(W_{xi} * X_t + W_{hi} * H_{t-1}^i + b_i) \\
 f_t = \sigma(W_{xf} * X_t + W_{hf} * H_{t-1}^i + b_f) \\
 T_t^i = f_t \odot T_{t-1}^i + i_t \odot g_t \\
 g'_t = \tanh(W'_{xg} * X_t + W'_{cg} * C_{t-1}^{i-1} + b'_g) \\
 i'_t = \sigma(W'_{xi} * X_t + W'_{ci} * C_{t-1}^{i-1} + b'_i) \\
 f'_t = \sigma(W'_{xf} * X_t + W'_{cf} * C_{t-1}^{i-1} + b'_f) \\
 C_t^i = f'_t \odot C_{t-1}^{i-1} + i'_t \odot g'_t \\
 o_t = \sigma(W_{xo} * X_t + W_{ho} * H_{t-1}^i + W_{to} * T_t^i + W_{co} * C_t^i + b_o) \\
 H_t^i = o_t \odot \tanh(W_{1 \times 1} * [T_t^i, C_t^i]) \\
 (\text{Diff}_X, \hat{H}_t^i) = \text{MDAM}(\text{Diff}_X, H_t^i)
 \end{cases}
 \quad (1)$$

式中: T_t^i 是时间记忆单元,保存前一个时间步的时间记忆信息; C_t^i 是存储单元,在当前时间步的不同层之间保存空间记忆,底层的存储单元由前一时间步的顶层传递而来; σ 和 \tanh 分别代表 Sigmoid 激活函数和 tanh 激活函数; $*$ 和 \odot 分别表示卷积运算和哈德曼积运算; g_t, i_t 和 f_t 是用来更新 T_t^i 的输入调制门、输入门和遗忘门; g'_t, i'_t 和 f'_t 是用来更新 C_t^i 的另一套输入调制门、输入门和遗忘门; o_t 表示输出门; W 和 b 表示对应的权重和偏移量。 T_t^i 和 C_t^i 拼接后通过 1×1 的卷积层降维得到融合的隐藏状态 H_t^i ,最后与 Diff_X 一起通过运动差分注意力模块 MDAM 提取出运动差分记忆 Diff_X 和空间多尺度记忆 \hat{H}_t^i 。

通过 MDSTM 单元的空间多尺度记忆流和运动差分记忆流进入最大池化层提取语义信息,存储单元 C 则通过卷积层提取,经过编码器的 4 层编码单元后共得到 3 种不同尺度的空间记忆,其中,最小

尺度的空间记忆与同一层的运动差分记忆融合后进入预测器中,根据提取的时空特征逐层恢复为原图像大小,每经过一层使用反卷积的解码单元后加入经过跳跃连接得到的对应尺度的空间记忆,如此可以让网络同时获得图像的底层和高层信息,从而恢复图像的低频特征。最后,通过一个卷积核大小为 1×1 的预测单元生成对下一帧图片的准确预测。为了在网络中存储长期记忆信息,存储单元 C 的记忆流纵向经过编码器和预测器,存储每一帧的多尺度时空特征,最后带着这一时间步的高层信息流向下一时间步的底层 MDSTM 单元,以便进行下一时间步的预测。

2.2 运动差分记忆流

如前文所述,大部分时空序列预测模型都没有注重提取序列隐含的时序特征,只是在努力训练模型生成与每一帧标签图片相似的图片,实质上并没有获取并理解图像序列之间的运动趋势,导致在预报时效变长时准确度急剧恶化。因此,本文设计了

一个运动差分注意力模块,将其嵌入运动差分时空记忆单元中,不断更新运动差分记忆流,以帮助编码器获得图像序列的时序特征。

输入到神经网络的数据中往往包含非常丰富的信息,但是信息的重要程度各不相同,为了使神经网络具备自主判断信息重要性的能力,研究人员引入能够模拟人类视觉和认知系统的注意力机制,使神经网络能够选择性地注意其中更重要的信息。基于注意力机制构建的卷积块注意力模块可以自适应地学习图片数据中不同像素的注意力权重,具有计算速度快、占用空间小等优点,非常适用于处理数据量较大的图片序列数据^[29]。因此,本文基于卷积块注意力模块构建运动差分注意力模块,具体结构如图 4 所示。其中,通道注意力模块和空间注意力模块使用平均池化和最大池化的方法,配合全连接层和卷积层得到通道权重和空间位置权重,与原图像相乘后得到增强通道和空间注意力的特征编码图。

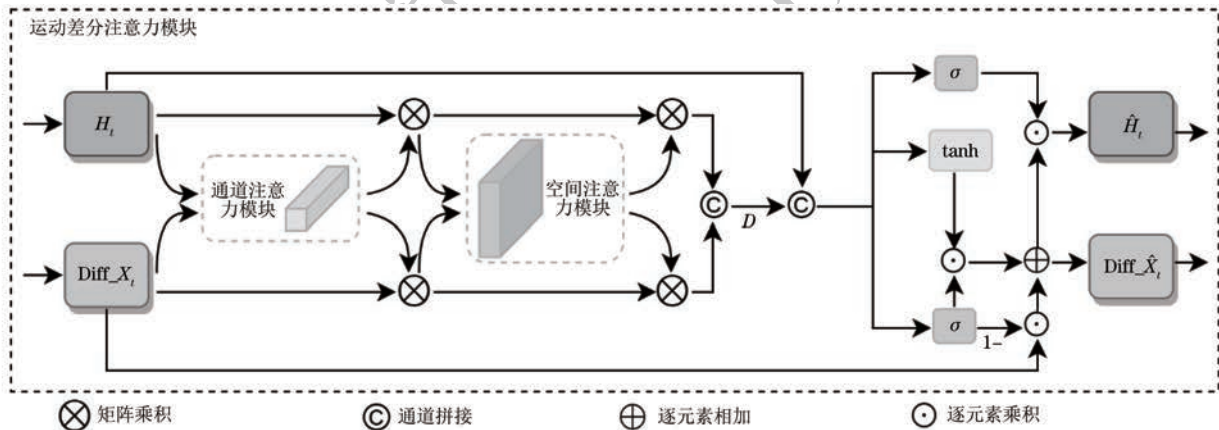


图 4 运动差分注意力模块

Fig. 4 Motion difference attention module

首先,运动差分注意力模块接收使用帧差法得到的差分运动图 $\text{Diff_}X_t$ 和当前时间步的隐藏状态 H_t ,分别通过通道注意力模块和空间注意力模块计算出运动特征编码图和隐藏状态编码图,在通道维度上进行拼接后得到聚合后的差分特征 D 。为了将差分特征保留在内存记忆中并沿着时间步传播,使用门控机制更新差分运动图和隐藏状态,具体更新公式如下:

$$\begin{cases} i_t = \sigma(W_{di} * D + W_{hi} * H_t + b_i) \\ f_t = \tanh(W_{df} * D + W_{hf} * H_t + b_f) \\ \text{Diff_}\hat{X}_t = (1 - i_t) \circ \text{Diff_}X_t + i_t \circ f_t \\ o_t = \sigma(W_{do} * D + W_{ho} * H_t + b_o) \\ \hat{H}_t = o_t \circ \text{Diff_}\hat{X}_t \end{cases} \quad (2)$$

式中: i_t 表示输入门; f_t 表示融合门; o_t 表示输出

门; \circ 表示点积运算; W 和 b 表示各个不同的权重和偏置; $\text{Diff_}\hat{X}_t$ 和 \hat{H}_t 分别表示更新后的差分运动图和隐藏状态。

通过以上的差分特征提取和记忆更新机制,运动差分记忆流与空间多尺度记忆流一起在层叠的编码单元中更新并传递,使网络能够在保留全局时空信息的同时强化对当前时间步差分特征的注意力,更好地提取到图片序列的运动信息。

2.3 损失函数

图像回归任务中最常用的损失函数是以 L1 和 L2 为代表的像素损失,它直接计算 2 个图像在像素级别的差异,能够确保图片的颜色、光照或整体对比度没有重大变化,提高生成图片的整体准确度^[30]。但是,像素损失不会考虑图片中的高频特征,这部分

特征又恰恰是人类感知图像质量的重要评判标准,只采用像素损失进行训练会导致模型无法捕捉图像中的上下文关系和全局结构信息,尤其是在图像序列预测任务中,误差的累积使得模型更加倾向于生成模糊和平滑的图片,在长期预测中得到不具备参考性的结果。因此,本文在像素损失的基础上融入一部分边缘损失,迫使模型重视边缘细节信息的生成,整体损失函数如下所示:

$$L_{\text{total}} = L_{\text{pixel}} + L_{\text{edge}} \quad (3)$$

式中: L_{pixel} 表示像素损失; L_{edge} 表示边缘损失。

在像素损失中,L1 损失倾向于产生稀疏的误差分布,有助于减少图像中的模糊并且保留锐利的边缘,但对噪声和异常值不敏感;L2 损失对大误差的惩罚力度大,对小误差较能容忍,会导致模型在预测时避免出现大的像素强度变化,从而生成比较模糊的图像。使用 L1+L2 的损失函数能够同时利用这 2 种损失的优势,兼顾局部和全局信息,提升生成图像的质量,因此,本文像素损失的计算公式如下:

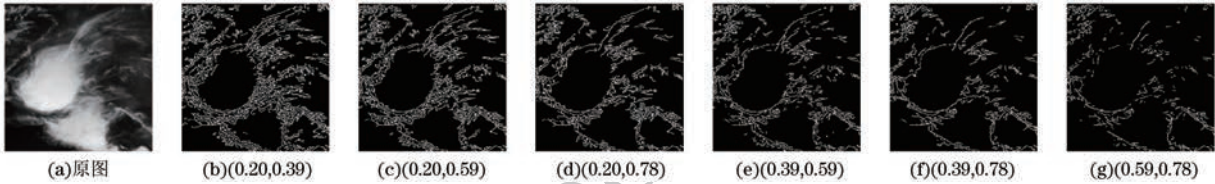


图 5 Canny 边缘检测算法高低阈值取值对比示例

Fig.5 Comparison example of high and low threshold values for Canny edge detection algorithm

首先,对标签图片序列和预测图片序列分别使用 Canny 边缘检测算法得到对应的边缘图,然后计算标签边缘图与预测边缘图之间的差异得到边缘损失,具体计算公式如下:

$$L_{\text{edge}} = \frac{\sum_{i=1}^N |E_i - \hat{E}_i|}{N} + \frac{\sum_{i=1}^N (E_i - \hat{E}_i)^2}{N} \quad (5)$$

式中: E 为标签边缘图片序列; \hat{E} 为预测边缘图片序列。

3 实验与分析

3.1 数据集

为了客观评估本文模型的预测性能和在不同类

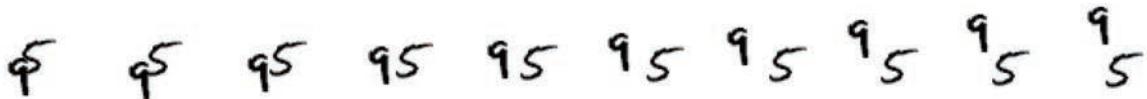


图 6 Moving MNIST 数据集示例

Fig.6 Example of Moving MNIST dataset

3.1.2 GOES-16 卫星云图数据集

GOES-16 是美国国家航空航天局(NASA)和美国国家海洋和大气管理局(NOAA)运营的 GOES-R 系列地球同步运行环境卫星中的第一颗,

$$L_{\text{pixel}} = \frac{\sum_{i=1}^N |Y_i - \hat{Y}_i|}{N} + \frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{N} \quad (4)$$

式中: Y 为标签图像序列; \hat{Y} 为预测图像序列; i 为像素标号; N 为总像素数。

Canny 边缘检测算法使用双阈值法计算得到图像的边缘图,具有较高的准确性和抗噪性。其中,高阈值和低阈值分别用于识别强边缘和弱边缘,较高的高阈值在过滤掉更多噪声的同时更加容易丢失细节,较低的低阈值则会将更多的弱边缘与强边缘相连,得到更连续的边缘图,但也会导致更多伪边缘的产生,一般低阈值会设为高阈值的 $1/3 \sim 1/2$ 。在本文中,所有用于训练的图片均进行了归一化处理,像素值在 $0 \sim 1$ 之间。由不同阈值计算出的对应边缘图对比示例如图 5 所示,将高、低阈值分别设置为 0.39 和 0.78 时,既能保留清晰的云团边缘,又能避免背景信息的过度干扰,最小化干扰信息带来的误差。

型数据集上的泛化能力,在 2 个图片像素大小不同的数据集上进行实验,分别为时空序列预测领域常用的 Moving MNIST 数据集^[31]和自制的 GOES-16 卫星云图数据集。

3.1.1 Moving MNIST 数据集

如图 6 所示, Moving MNIST 数据集包括 10 000 条长度为 20 帧、大小为 64×64 像素的灰度图序列,其中,每张灰度图包含 2 个位置和方向不停变化且能从边界回弹的手写数字。本文将每条图片序列的前 10 帧作为输入数据,后 10 帧作为标签数据,整个数据集按照 3:1:1 的比例划分为训练集、验证集和测试集。

它使用先进基线成像仪获取地球高时空分辨率图,包含 16 个光谱波段,时间分辨率为 15 min,能够有效监测云位置的变化情况。其中,第 13 波段与其他红外波段相比对水蒸气吸收的敏感性较低,有助于

识别及分类云和其他大气特征,也有利于云顶亮度温度、云粒径等云结构参数的估计,广泛应用于降水强度预测等领域。因此,本文选择 2019 年 6 月—

9 月 GOES-16 卫星遥感云图第 13 波段的数据进行数据提取、几何校正、辐射定标等预处理操作,制作卫星云图数据集,如图 7 所示。

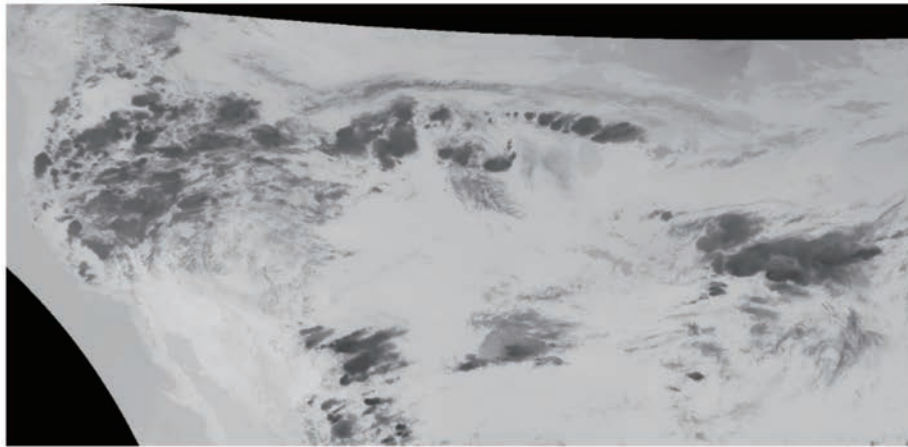


图 7 GOES-16 卫星第 13 波段示例

Fig.7 Example of the 13th band of GOES-16 satellite

卫星云图的原始分辨率为 $1\,483 \times 722$ 像素,直接使用时对计算机存储空间和性能要求过高,也会带来更高的时间成本,因此,本文使用 Python 程序将其划分为大小为 384×384 像素的区域卫星云图。过滤云团面积过小甚至无云的区域后按时间顺序排列,使用滑动窗口法取长度为 24 帧的连续序列,两帧的时间间隔为 15 min,训练样本为前 12 帧,标签数据为后 12 帧,即用前 3 h 的历史云图序列预测后 3 h 的未来云图序列。为了消除数据取值范围差异产生的影响,对原始范围为 150~350 像素的云图进行最小最大规范化处理,将像素值映射到 $[0, 1]$ 区间,公式如下所示:

$$x^* = \frac{x - V_{\min}}{V_{\max} - V_{\min}} \quad (6)$$

式中: x 为样本数据的像素值; V_{\min} 为样本数据的最小值; V_{\max} 为样本数据的最大值。

经过以上预处理,得到序列长度为 24 帧、分辨率为 384×384 像素、像素值范围在 $[0, 1]$ 之间的区域卫星云图序列数据集。数据集总序列数为 11 543 个,其中 6 927 个作为训练集,2 308 个作为验证集,2 308 个作为测试集。

3.2 实验环境与设置

为了验证多尺度运动记忆模型的预测性能,分别在 2 个实验环境下使用 Moving MNIST 数据集和 GOES-16 卫星云图数据集进行对比实验,选择经典预测模型 ConvLSTM^[24]、ConvGRU^[25]、PredRNN^[26]、PredRNN++^[27] 以及新型预测模型 E3D-LSTM^[32]、CrevNet^[33] 作为对比模型。其中, Moving MNIST 实验在 Windows 11 操作系统下进行,计算机 CPU 为

Intel Core i9 2.50 GHz, GPU 为 NVIDIA GTX 3060, 显存为 6 GB; GOES-16 卫星云图实验在 Windows 10 操作系统下进行,计算机 CPU 为 Intel Core i7 3.40 GHz, GPU 为 NVIDIA GTX 3060, 显存为 24 GB。2 个实验的软件环境均为 Python 3.9, 使用 PyTorch 2.0.1^[34] 作为深度学习框架进行训练。所有模型均使用 Adam 优化器训练 60 个 Epoch, 初始学习率为 0.000 1, Batch Size 为 4。为了获得最好的泛化性能同时防止过拟合,使用早停法进行实验。同时,使用 OpenCV 和 GDAL 等开源库进行数据集预处理,使用 OpenCV 提供的 Canny 函数得到边缘图,高、低阈值分别设为 0.78 和 0.39。

3.3 评价指标

本文使用 4 种评价指标来对模型预测的图片序列质量进行定量分析。首先,是回归任务常用的评价指标均方误差(MSE),它从像素级别上对预测图片和标签图片的差别进行计算,其计算公式如下所示:

$$\text{MSE}(Y, \hat{Y}) = \frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{N} \quad (7)$$

式中: Y 为真实的区域未来卫星云图; \hat{Y} 为模型预测的区域未来卫星云图。

可学习感知图像块相似性指标(LPIPS)通过学习一个神经网络模型来模拟人类对图像的视觉感知,使用图像块之间的感知距离评估预测图像的质量^[35],一般使用 CNN 对图像的局部块进行特征提取,并计算块之间的相似性得分。相似性得分表示两幅图像在感知上的差异,范围在 $0 \sim 1$ 之间,数值越小表示图像之间的感知差异越小,生成的图像质

量越高。本文选择 AlexNet 作为预训练网络模型提取预测图像和标签图像的特征表示。

峰值信噪比(PSNR)表示峰值信号的最大功率与噪声的平均功率之比,常用单位为 dB,是一种基于误差灵敏度估计预测图像质量的方法。PSNR 值越大,两幅图像越相似。PSNR 计算公式如下所示:

$$PSNR(Y, \hat{Y}) = 10 \lg \left(\frac{M_{MAX,I}^2}{MSE(Y, \hat{Y})} \right) \quad (8)$$

式中: $M_{MAX,I}$ 表示可能的最大像素值。

结构相似性指数(SSIM)假设人眼会提取图像中的结构化信息,通过图像的亮度、对比度和结构 3 个方面衡量两幅图像之间的结构相似性^[36],其值在 0~1 之间,当 SSIM 值为 1 时,认为两幅图像完全相同。SSIM 计算公式如下所示:

$$SSIM(Y, \hat{Y}) = \frac{(2\mu_Y\mu_{\hat{Y}} + c_1)(2\sigma_{Y\hat{Y}} + c_2)}{(\mu_Y^2 + \mu_{\hat{Y}}^2 + c_1)(\sigma_Y^2 + \sigma_{\hat{Y}}^2 + c_2)} \quad (9)$$

式中: $\mu_Y, \mu_{\hat{Y}}$ 分别表示 Y, \hat{Y} 的平均值,是对图像亮度的估计; $\sigma_Y^2, \sigma_{\hat{Y}}^2$ 分别表示 Y, \hat{Y} 的方差,是对图像对比度的估计; $\sigma_{Y\hat{Y}}$ 表示 Y 和 \hat{Y} 的协方差,是对图像结构相似度的度量; c_1 和 c_2 是用于防止分母为 0 的正常数。

3.4 实验结果分析

3.4.1 Moving MNIST 上的对比实验

为了客观评价模型的预测能力,分别计算模型的定量评价指标在所有预测帧上的平均值,结果如表 1 所示,最优结果加粗标注,并画出各指标的逐帧变化趋势图,如图 8 所示。由表 1 可知,本文提出的 MSMM_Net 在 MSE、LPIPS、PSNR 和 SSIM 指标上均取得了最好的成绩。与基准模型 PredRNN 相比,多尺度运动记忆模型的 MSE 从 1 908.91 降至 1 303.50, LPIPS 从 0.227 降至 0.080,同时 PSNR 和 SSIM 分别提升了 11.7% 和 7.95%,预测能力显著提升。

表 1 Moving MNIST 数据集上的定量对比结果

Table 1 Quantitative comparison results on the Moving MNIST dataset

模型	MSE ↓	LPIPS ↓	PSNR ↑ /dB	SSIM ↑
ConvLSTM	3 196.30	0.262	14.09	0.736
ConvGRU	3 109.52	0.253	14.46	0.735
PredRNN	1 908.91	0.227	15.73	0.805
PredRNN++	1 782.60	0.214	16.03	0.805
E3D-LSTM	1 695.07	0.188	16.24	0.817
CrevNet	1 634.04	0.189	16.52	0.821
MSMM_Net	1 303.50	0.080	17.57	0.869

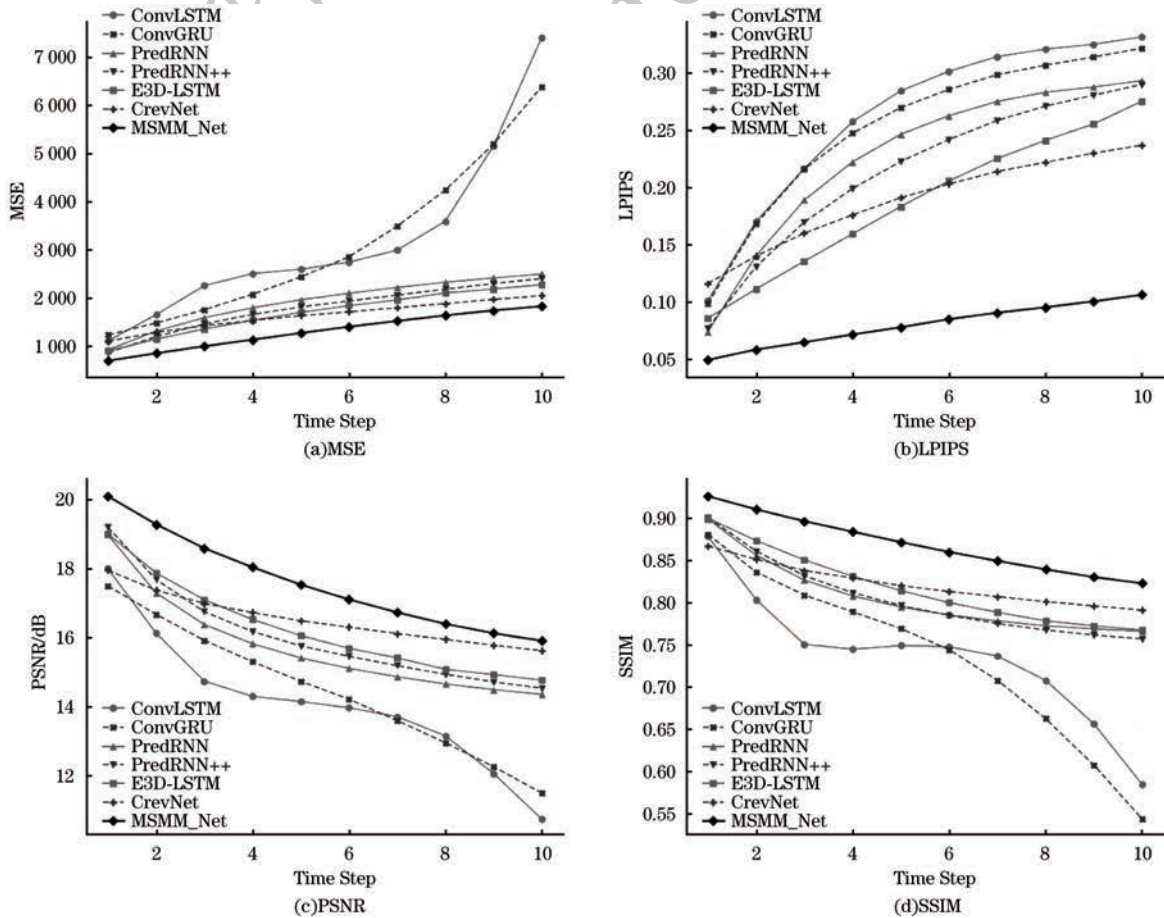


图 8 Moving MNIST 数据集上的评价指标趋势

Fig.8 Trends in evaluation metrics on the Moving MNIST dataset

各模型预测出的手写数字序列示例如图 9 所示。可以看出:ConvLSTM 仅能准确预测出第一帧的数字;ConvGRU 的预测质量略有提升,但是生成的图片背景颜色较深,说明它并没有捕捉到运动的主体,仅靠生成平均值来获得更低的损失;PredRNN 和 PredRNN++ 在预测时间过半前后会出现严重模糊,无法辨认数字;E3D-LSTM

LSTM 与 CrevNet 生成的数字清晰度明显提升,但是随着时间的推移,无法保持正确的数字形态;本文提出的 MSMM_Net 在所有时间帧上都与标签图片保持高度一致,且数字边缘清晰,说明该模型同时捕获了数字的运动轨迹和形状特征,且在较长的预测时间内保留了重要记忆,具有更高的预测性能。



图 9 Moving MNIST 数据集上的预测效果对比

Fig.9 Comparison of prediction effect on the Moving MNIST dataset

3.4.2 GOES-16 卫星云图上的对比实验

各模型在 GOES-16 卫星云图数据集上的定量评价结果如表 2 所示,每个指标均为所有预测帧的平均值,同时画出各指标的逐帧变化趋势图,如图 10 所示。

表 2 GOES-16 数据集上的定量对比结果
Table 2 Quantitative comparison results on the GOES-16 dataset

模型	MSE ↓	LPIPS ↓	PSNR ↑ /dB	SSIM ↑
ConvLSTM	845.76	0.222	19.99	0.623
ConvGRU	846.28	0.217	19.76	0.630
PredRNN	778.12	0.219	20.50	0.632
PredRNN++	768.59	0.220	20.92	0.645
E3D-LSTM	702.61	0.214	20.85	0.648
CrevNet	651.69	0.217	20.94	0.652
MSMM_Net	581.70	0.209	21.63	0.666

与 Moving MNIST 数据集不同,卫星云图相比手写数字具有更加复杂的背景,导致模型受到的干扰信息量更大,同时运动主体的形状也会随着时间推移而变化,这大大提高了模型预测的难度。如表 2 所示,各个评价指标的变化范围与 Moving MNIST 实验结果相比变小了许多,但 MSMM_Net 依然保持最佳的性能表现,在指标最好的同时精度

衰减也较为缓慢,更适合长期预测任务。

选取 2 种不同运动形态的云团对比 MSMM_Net 与其他模型的预测效果,结果如图 11 和图 12 所示,其中,Input frames 为输入的前 12 帧云图,Ground Truth 为实际的后 12 帧云图。图 11 中左上角的云团在不断向右扩大,图 12 则为气旋云图,虽然气旋本身并没有太大的位置变动,但其周围的云在不停环绕着风眼中心按逆时针方向转动,并吸收边缘的细小云团进入气旋。由于 ConvLSTM 和 ConvGRU 的单元状态只进行水平方向的更新,下一时间步的底层记忆单元无法接收到上一时间步的高层记忆单元提取的特征,因此在它们的预测中云团的形状和位置几乎没有变化,且边缘最早开始模糊。增加“之”字形数据流和时空记忆单元的 PredRNN 与 PredRNN++ 对云团的纹理细节保留时间更长,也学习到了云团扩大的趋势,但预测出的云团形状与真实云团不符。E3D-LSTM 通过 3D 卷积与 RNN 模型的融合,增强了模型对运动的感知能力,CrevNet 的条件可逆设计能够减少特征提取中的信息丢失,因此在它们的预测中云团均具备了正确的运动趋势,但是随着时间的推移,容易受到噪声的影响,云团运动方向偏离现实。相比之下,MSMM_Net 使用运动差分注意力模块充分提取云

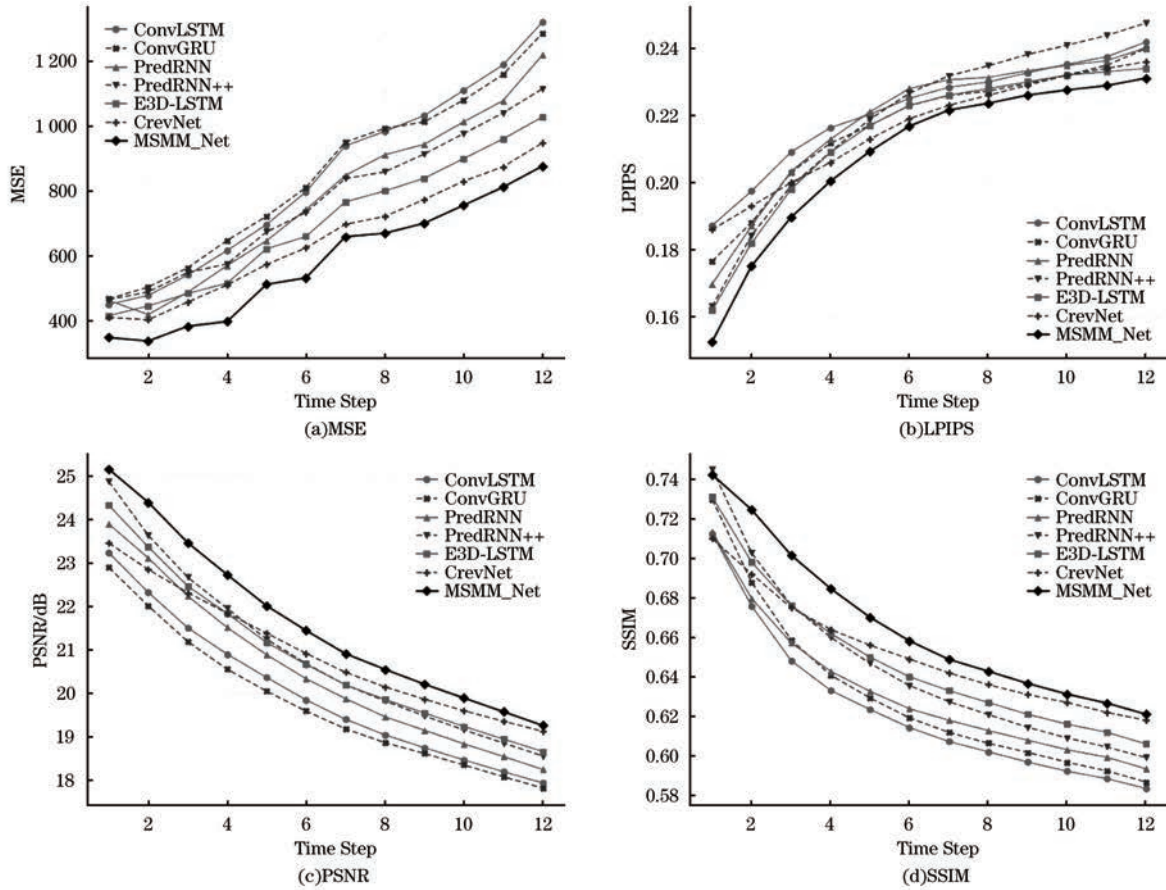


图 10 GOES-16 数据集上的评价指标趋势

Fig. 10 Trend of evaluation indicators on the GOES-16 dataset

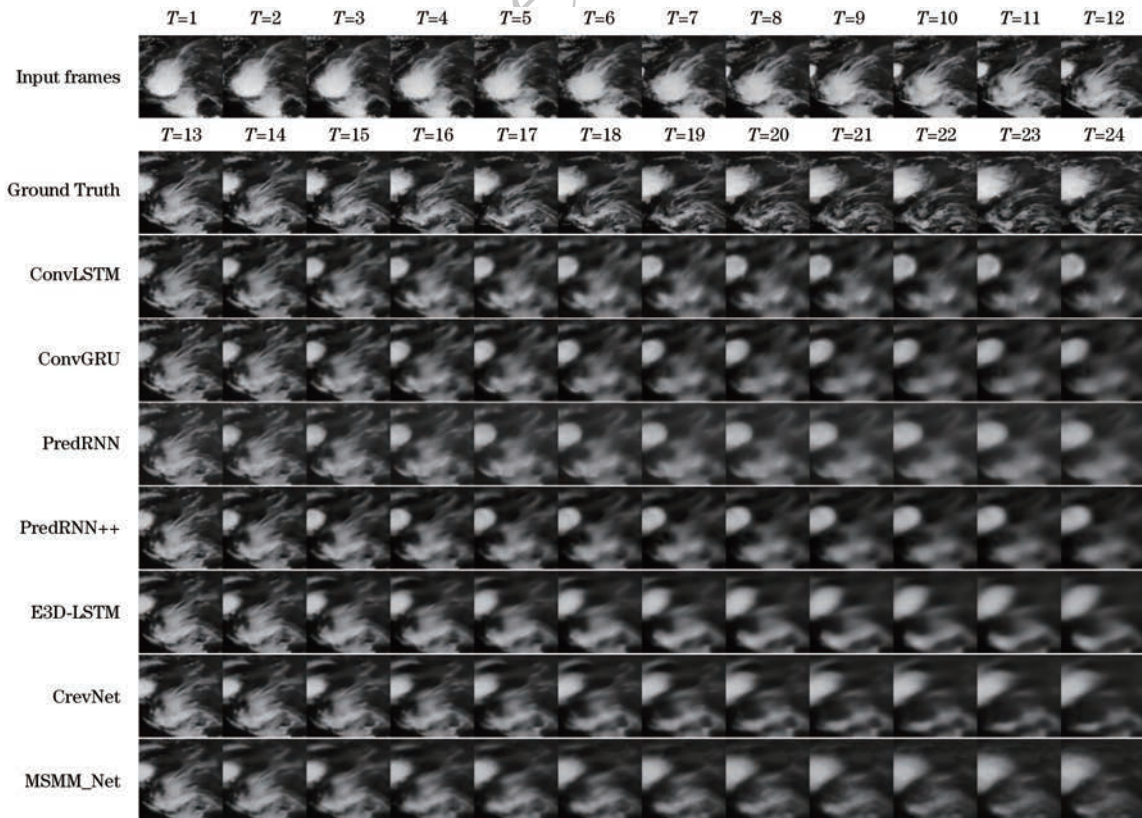


图 11 GOES-16 数据集的预测效果对比 1

Fig. 11 Comparison 1 of prediction effect of GOES-16 dataset

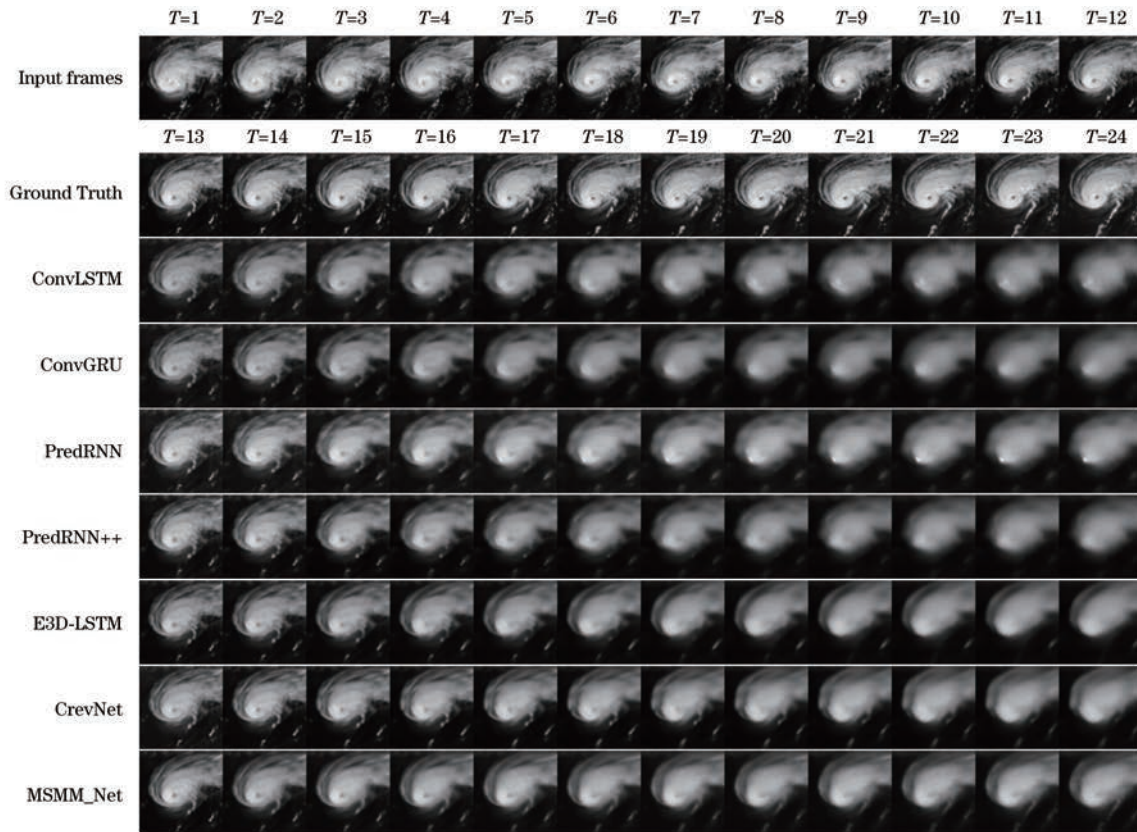


图 12 GOES-16 数据集的预测效果对比 2

Fig. 12 Comparison 2 of prediction effect of GOES-16 dataset

团的运动特征,正确预测了云团的运动趋势,另外,受益于能够融合云图高频次信息和低频次信息的空间多尺度记忆流,该模型预测的图片细节保留程度也最高,在气旋云图的最后一帧时依然能观测到风眼。因此,本文提出的 MSMM_Net 比其他模型拥有更高的时空特征提取能力,能够正确预测云团的

运动趋势,生成细节丰富的预测图片,并维持长时间的准确预测,提升了长时间时序预测能力。

3.4.3 GOES-16 卫星云图上的消融实验

为了进一步分析 MSMM_Net 中每个改进的作用,在 GOES-16 数据集上进行一系列消融实验,结果如表 3 所示。

表 3 GOES-16 数据集上的消融实验结果

Table 3 Results of ablation experiments on the GOES-16 dataset

模型	MDAM	多尺度架构	损失函数	MSE ↓	LPIPS ↓	PSNR ↑ /dB	SSIM ↑
基准模型			像素损失	778.12	0.219	20.50	0.632
变体模型 1			边缘损失	792.84	0.221	20.32	0.620
变体模型 2			融合损失	752.02	0.217	20.70	0.638
变体模型 3	✓		像素损失	652.65	0.210	21.02	0.655
变体模型 4		✓	像素损失	600.45	0.210	21.50	0.659
本文模型	✓	✓	融合损失	581.70	0.209	21.63	0.666

基准模型为 PredRNN,同时使用 L1+L2 的像素损失进行训练;变体模型 1 同样为 PredRNN,但单独使用边缘损失进行训练;变体模型 2 依然不改变网络结构,使用同时包含像素损失和边缘损失的融合损失进行训练,以此探究融合损失对预测结果的影响。由表 3 可知,单独使用边缘损失的变体模型 1 效果没有超过基准模型,说明像素损失在图像生成任务中的作用是不可或缺的,但变体模型 2 的

各指标值均优于基准模型,体现了边缘损失作为辅助数据对模型预测的正向作用。正如前文所述,像素损失在图像生成任务中具有高度的优势,但同时也有无法捕捉上下文关系的缺点,因此,在使用像素损失的基础上根据任务目标融入其他损失函数,是提高图像生成质量的有效途径。变体模型 3 是在基准模型的基础上加入 MDAM 运动差分注意力模块得到的,能够探究加入运动差分记忆流后对模型预

测性能的影响。由表 3 可知,变体模型 3 的 MSE 比基准模型降低了约 16.12%,LPIPS 降低了约 4.11%,PSNR 从 20.50 dB 增加到 21.02 dB,SSIM 从 0.632 增加到 0.655,证明在模型中加入运动差分记忆流能够有效增强模型的预测能力。变体模型 4 是使用多个不同尺度且没有嵌入 MDAM 的运动差分时空记忆单元构成的,目的是分析空间多尺度记忆流对云图预测任务的贡献。由表 3 可知,变体模型 4 的 MSE 与基准模型相比降低了约 22.83%,LPIPS 与变体模型 3 相同,PSNR 和 SSIM 与基准模型相比分别增长了 1 dB 和 0.027,说明空间多尺度记忆流也能够明显提升模型的时序特征提取能力,促进模型生成正确的预测结果。

本文提出的 MSMM_Net 包含了消融实验测试的所有模块,能够大幅提升模型的预测精度,证明了其用于长期云图预测的可行性和有效性。

4 结束语

本文提出了一种用于遥感云图预测的多尺度运动记忆模型 MSMM_Net。使用双分支记忆流结构增强模型对图片序列的时空特征提取能力;针对传统方法对图片序列运动特征提取不足的问题,提出运动差分记忆流来捕捉运动特征,使模型充分理解主体的运动趋势;针对特征提取过程中底层信息容易被遗忘的问题,使用空间多尺度记忆流在网络中传递图片的高低频空间特征;在训练时使用像素损失和边缘损失融合的损失函数,强化模型对边缘细节的关注度,提高模型的预测准确度。实验结果表明,MSMM_Net 的预测图片序列与标签图片序列的相似度较高,能够维持长达 3 h 的准确预测,验证了模型的有效性。尽管在实验中 MSMM_Net 相比其他模型在预测图像序列的清晰度上有所提升,但仍然存在一定的模糊问题。这可能是由于 RNN 在预测过程中将生成帧作为真实帧来参与后续帧预测,导致误差随着时间的推移而逐渐累积,从而增大了预测不确定性,生成了更加模糊的结果。因此,在保留 RNN 捕捉时序特征优势的同时,减少不确定性误差的积累,将是未来的研究方向。

参考文献

- [1] 张弛,刘钧,李旭光,等.基于可见光-红外图像信息融合的云状识别方法[J].气象与环境学报,2018,34(1):82-90. ZHANG C, LIU J, LI X G, et al. Cloud recognition method based on visible-infrared image information fusion [J]. Journal of Meteorology and Environment, 2018, 34(1): 82-90. (in Chinese)
- [2] XIAN D, ZHANG P, GAO L, et al. Fengyun meteorological satellite products for earth system science applications[J]. Advances in Atmospheric Sciences, 2021, 38(8): 1267-1284.
- [3] 方旭源,金炜,符冉迪,等.注意力机制下多尺度特征融合生成对抗网络的白天海雾监测[J].遥感学报,2023,27(12):2736-2747. FANG X Y, JIN W, FU R D, et al. Daytime sea fog detection based on multi-scale feature fusion of generated adversarial network under attention mechanism[J]. National Remote Sensing Bulletin, 2023, 27(12): 2736-2747. (in Chinese)
- [4] ROJAS Y, MINDER J R, CAMPBELL L S, et al. Assessment of GPM IMERG satellite precipitation estimation and its dependence on microphysical rain regimes over the mountains of south-central Chile[J]. Atmospheric Research, 2021, 253: 105454.
- [5] 黄杰,张永宏,马光义,等.改进 U-Net 的风云四号卫星降水估计算法研究[J].计算机工程与应用,2023,59(11):285-293. HUANG J, ZHANG Y H, MA G Y, et al. Research on precipitation estimation algorithm from Fengyun-4 satellite based on improved U-Net [J]. Computer Engineering and Applications, 2023, 59(11): 285-293. (in Chinese)
- [6] 郑倩,毛程燕,于丽华,等.台风利奇马(1909)与台风摩羯(1814)云特征对比[J].应用气象学报,2022,33(1):43-55. ZHENG Q, MAO C Y, DING L H, et al. Comparison of cloud characteristics between typhoon Lekima (1909) and typhoon Yagi(1814) [J]. Journal of Applied Meteorological Science, 2022, 33(1): 43-55. (in Chinese)
- [7] 宁子杰,包文雯,陈峥蓉,等.台风“浪卡”(2016)暴雨特点及卫星云图特征分析[J].气象研究与应用,2021,42(3):88-93. NING Z J, BAO W W, CHEN Z R, et al. Typhoon Nangka (2016) rainstorm characteristics and satellite cloud image characteristics analysis [J]. Journal of Meteorological Research and Application, 2021, 42(3): 88-93. (in Chinese)
- [8] 郑宗生,刘敏,胡晨雨,等.基于 Seq2Seq 和 Attention 的时序卫星云图台风等级预测[J].遥感信息,2020,35(4):16-22. ZHENG Z S, LIU M, HU C Y, et al. Prediction of typhoon grade with time series typhoon satellite images based on Seq2Seq and Attention [J]. Remote Sensing Information, 2020, 35(4): 16-22. (in Chinese)
- [9] BOSCH J L, KLEISSL J. Cloud motion vectors from a network of ground sensors in a solar power plant[J]. Solar Energy, 2013, 95: 13-20.
- [10] HUANG H, XU J, PENG Z Z, et al. Cloud motion estimation for short term solar irradiation prediction [C]// Proceedings of the IEEE International Conference on Smart Grid Communications (SmartGridComm). Washington D. C., USA: IEEE Press, 2013: 696-701.
- [11] 王磊,周毓荃,蔡森,等.华北云特征参数与降水相关性的研究[J].气象与环境科学,2019,42(3):9-16. WANG L, ZHOU Y Q, CAI M, et al. Study on correlation between cloud characteristic parameters and precipitation in North China[J]. Meteorological and Environmental Sciences, 2019, 42(3): 9-16. (in Chinese)
- [12] 吴现,吐松江.卡日,王海龙,等.基于时空长短时记忆神经网络的地基云图预测算法[J].计算机工程,2024,50(3):298-305. WU X, TUSONGJIANG K, WANG H L, et al. Ground-based cloud map prediction algorithm based on spatio-temporal long short-term memory neural network [J]. Computer Engineering, 2024, 50(3): 298-305. (in Chinese)
- [13] MA B, YANG N, CUI C G, et al. Cloud position forecasting based on ConvLSTM network [C]// Proceedings of the 5th International Conference on Power and Renewable Energy (ICPRE). Washington D. C., USA: IEEE Press,

- 2020: 562-565.
- [14] XU Z, DU J, WANG J J, et al. Satellite image prediction relying on GAN and LSTM neural networks[C]//Proceedings of the 2019 IEEE International Conference on Communications (ICC). Washington D. C., USA: IEEE Press, 2019: 1-6.
- [15] 严义杰. 基于深度学习的云移动预测方法研究[D]. 哈尔滨: 哈尔滨工业大学, 2020.
YAN Y J. Research on cloud movement forecasting method based on deep learning [D]. Harbin: Harbin Institute of Technology, 2020. (in Chinese)
- [16] SHIELDS J, KARR M, TOOMAN T, et al. The whole sky imager-a year of progress [EB/OL]. [2024-05-05]. <https://www.semanticscholar.org/paper/The-Whole-Sky-Imager-A-Year-of-Progress-Shields-Karr/239291dffb867b02e5e7ab97c6d215ff60e98ddf>.
- [17] BRAD R, LETIA I A. Cloud motion detection from infrared satellite images [EB/OL]. [2024-05-05]. https://www.researchgate.net/publication/228557524_Cloud_motion_detection_from_infrared_satellite_images.
- [18] JAMALY M, KLEISSL J. Robust cloud motion estimation by spatio-temporal correlation analysis of irradiance data[J]. Solar Energy, 2018, 159: 306-317.
- [19] GIBSON J J. The perception of the visual world [M]. Boston, USA: Houghton Mifflin, 1950.
- [20] WOOD-BRADLEY P, ZAPATA J, PYE J. Cloud tracking with optical flow for short-term solar forecasting [EB/OL]. [2024-05-05]. <https://www.semanticscholar.org/paper/Cloud-tracking-with-optical-flow-for-short-term-Wood-Bradley-Zapata/fe698691c892cae10b8a81a5e2b90f24809b2f74/figure/0>.
- [21] DISSAWA D M L H, EKANAYAKE M P B, GODALIYADDA G M R I, et al. Cloud motion tracking for short-term on-site cloud coverage prediction[C]//Proceedings of the 17th International Conference on Advances in ICT for Emerging Regions (ICTer). Washington D. C., USA: IEEE Press, 2018: 1-6.
- [22] CHOW C W, BELONGIE S, KLEISSL J. Cloud motion and stability estimation for intra-hour solar forecasting[J]. Solar Energy, 2015, 115: 645-655.
- [23] 石玉立, 施声伟. 气流算法在 FY-4A 红外图像外推中的精度评价研究[J]. 兵器装备工程学报, 2021, 42(9): 150-158, 224.
SHI Y L, SHI S W. Research on accuracy evaluation of optical flow algorithm in FY-4A infrared image extrapolation[J]. Journal of Ordnance Equipment Engineering, 2021, 42(9): 150-158, 224. (in Chinese)
- [24] SHI X J, CHEN Z R, WANG H, et al. Convolutional LSTM network: a machine learning approach for precipitation nowcasting [EB/OL]. [2024-05-05]. <https://arxiv.org/abs/1506.04214>.
- [25] SHI X J, GAO Z H, LAUSEN L, et al. Deep learning for precipitation nowcasting: a benchmark and a new model [EB/OL]. [2024-05-05]. <https://arxiv.org/abs/1706.03458>.
- [26] WANG Y B, LONG M S, WANG J M, et al. PredRNN: recurrent neural networks for predictive learning using spatiotemporal LSTMs [EB/OL]. [2024-05-05]. <https://arxiv.org/abs/2103.09504>.
- [27] 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]. [2024-05-05]. <https://arxiv.org/abs/1804.06300>.
- [28] TREBING K, STAŃCZYK T, MEHRKANOON S. SmaAt-UNet: precipitation nowcasting using a small attention-UNet architecture [J]. Pattern Recognition Letters, 2021, 145: 178-186.
- [29] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [EB/OL]. [2024-05-05]. <https://arxiv.org/abs/1807.06521>.
- [30] SEIF G, ANDROUTSOS D. Edge-based loss function for single image super-resolution [C]//Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Washington D. C., USA: IEEE Press, 2018: 1468-1472.
- [31] SRIVASTAVA N, MANSIMOV E, SALAKHUTDINOV R. Unsupervised learning of video representations using LSTMs [C]// Proceedings of the 32nd International Conference on Machine Learning. Washington D. C., USA: IEEE Press, 2015: 843-852.
- [32] WANG Y, JIANG L, YANG M H, et al. Eidetic 3D LSTM: a model for video prediction and beyond [EB/OL]. [2024-05-05]. <https://openreview.net/pdf?id=BLIKS2AqtX>.
- [33] YU W, LU Y, EASTERBROOK S, et al. Efficient and information-preserving future frame prediction and beyond [EB/OL]. [2024-05-05]. https://openreview.net/pdf?id=BlEY_pVYvB.
- [34] PASZKE A, GROSS S, MASSA F, et al. PyTorch: an imperative style, high-performance deep learning library [EB/OL]. [2024-05-05]. <https://arxiv.org/abs/1912.01703>.
- [35] ZHANG R, ISOLA P, EFROS A A, et al. The unreasonable effectiveness of deep features as a perceptual metric [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 586-595.
- [36] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity [J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.

文字编辑 吴云芳
栏目编辑 宋圆