E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035
中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS
©中国图象图形学报版权所有

中图法分类号:P412.27; TP391.41　　**文献标识码**: A　　**文章编号**: 1006-8961(2026)09-3412-15

论文引用格式:He Q, Zang Z Y and Hao Z Z. 2026. CloudPredUNet: a satellite cloud image prediction network incorporating frequency domain self-attention. Journal of Image and Graphics, 31(9):3412-3426(贺琪, 臧正源, 郝增周. 2026. 融合频域自注意力的卫星云图预测网络CloudPredUNet. 中国图象图形学报, 31(9):3412-3426)[DOI:10.11834/jig.250559]

融合频域自注意力的卫星云图预测网络CloudPredUNet

贺琪[1],臧正源[1,2],郝增周[2*]

1. 上海海洋大学信息学院,上海　201306; 2. 自然资源部第二海洋研究所卫星海洋环境监测预警全国重点实验室,杭州　310012

摘　要:**目的**　现有卫星云图预测方法在应对云团非线性运动时存在局限:传统方法难以建模复杂动态,深度学习模型则常受计算复杂、感受野有限或时间依赖建模不足的制约。为此,提出融合频域自注意力的卫星云图预测网络CloudPredUNet。**方法**　CloudPredUNet基于UNet架构,编码器引入频域注意力模块,通过对图像块进行傅里叶变换与频域互相关,以较低计算成本实现全局特征聚合;解码器结合多尺度空洞卷积与空间注意力,增强细节重建;时间映射器融合多分支异感受野卷积与通道重标定,以加强长程时空依赖建模。**结果**　在FY-2D数据集上的实验表明,CloudPredUNet在模型效率与预测精度上均领先。CloudPredUNet在参数量(0.101 M)和计算量(5 GFLOPs)远低于主流对比模型的同时,在4项关键评价指标上均取得最优值:MAE(mean absolute error)为8.393、MSE(mean squared error)为188.922、SSIM(structural similarity index measure)为0.825、PSNR(peak signal to noise ratio)为25.367 dB。消融实验验证了各模块的有效性与互补性,完整模型相比基线在MAE上提升约11.63%。多帧预测结果显示,该模型在连续5帧预测中每一帧的MAE均为最低,且在第5帧预测误差(10.589)较次优的MIM模型(10.852)进一步降低约2.4%,显示出更优的抗误差累积能力。可视化对比进一步证实,所提方法在云团生长、合并及台风眼结构保持等复杂场景中,能更完整地维持细节与运动一致性。**结论**　本文工作为气象业务中实时、精准的云图短时预测提供了轻量高效的解决方案。未来将拓展数据集覆盖范围、探索多模态数据融合,并进一步优化频域注意力机制以提升重建质量。

关键词:卫星云图预测;时空预测;FY-2D卫星;频域自注意力;空间注意力

CloudPredUNet: a satellite cloud image prediction network incorporating frequency domain self-attention

He Qi[1], Zang Zhengyuan[1,2], Hao Zengzhou[2*]

1. College of Information Science, Shanghai Ocean University, Shanghai　201306, China; 2. State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou　310012, China

Abstract: **Objective** Accurate short-term prediction of satellite cloud imagery is crucial for modern meteorological operations, including severe weather nowcasting, aviation safety, and disaster preparedness. However, this task remains profoundly challenging due to the inherently nonlinear and chaotic dynamics of atmospheric motion. Cloud systems undergo

收稿日期:2025-11-17;**修回日期**:2026-01-21;**预印本日期**:2026-01-28

*通信作者:郝增周　hzyx80@sio.org.cn

基金项目:浙江省自然科学基金项目(LZJMZ24D060001);国家自然科学基金项目(42376194)

Supported by: Natural Science Foundation of Zhejiang Province, China (LZJMZ24D060001); National Natural Science Foundation of China (42376194)

complex spatiotemporal transformations, such as genesis, dissipation, deformation, rotation, merging, and splitting, which are difficult to capture using conventional forecasting techniques. Traditional approaches, including optical flow and block-matching methods, primarily estimate displacement vectors based on local brightness constancy or regional correlation. While computationally manageable, these methods often fail to model the nonstationary and nonrigid transformations characteristic of real cloud evolution, leading to considerable error accumulation over extended prediction horizons. **Method** To overcome these challenges, this study proposes CloudPredUNet, a novel and lightweight neural network architecture for satellite cloud image sequence prediction. The model is built upon a U-Net skeleton, renowned for its effective encoder-decoder structure with skip connections, which is well-suited for dense prediction tasks. CloudPredUNet introduces three core innovative modules designed to enhance global feature aggregation, detail reconstruction, and long-range spatiotemporal dependency modeling, respectively. First, in the encoder pathway, we design a frequency domain self-attention (FDSA) module. Instead of applying global Fourier transforms, which can be sensitive to noise and computationally heavy for high-resolution images, our module operates on localized image patches. The input features are projected and split into query (Q), key (K), and value (V) components. The Q and K tensors are partitioned into nonoverlapping patches. A 2D fast Fourier transform (FFT) is applied to each patch, transitioning the representation into the frequency domain. A Hadamard (element-wise) product between the spectra of Q and K patches is then performed, effectively computing cross-correlation in the frequency domain, a process equivalent to spatial convolution but with considerably lower computational complexity (O(HW log (HW))). The result is transformed back to the spatial domain via inverse FFT, and then gated with the V component through element-wise multiplication. This design enables efficient aggregation of global contextual information across the entire image while preserving local texture details, addressing the limited receptive field of standard convolutions. Second, the decoder pathway incorporates a multiscale spatial attention (MSA) module to recover high-resolution details and suppress irrelevant background regions during upsampling. This module employs a series of cascaded depthwise separable convolutions with progressively increasing dilation rates (e. g. , d = 1, 3, 5, 7). This multibranch structure captures features at multiple scales, mitigating the gridding artifacts common in single-dilation dilated convolutions. The outputs from all branches are concatenated and processed through a spatial attention mechanism. This mechanism generates an attention map by applying max-pooling and average-pooling across the channel dimension, followed by a small multilayer perceptron (MLP) and a sigmoid activation. This map highlights structurally important regions (e. g. , cloud edges, cores) and attenuates uniform areas, thereby refining the reconstruction quality. Third, a dedicated temporal mapper is introduced between the encoder and decoder to model the evolution of features over time explicitly. Its core component is a spatiotemporal feature extraction (SFE) module. This module first merges the time and channel dimensions of the encoded feature sequence. The merged features are then split into four subsets. One subset is preserved, while the other three are processed by distinct convolutional branches: a standard 3 × 3 depthwise convolution, a 1 × 11 horizontal strip convolution, and an 11 × 1 vertical strip convolution. This heterogeneous receptive field design allows the model to capture isotropic cloud expansion as well as anisotropic motion patterns (e. g. , shear, directional advection). The outputs are fused and further refined by a channel recalibration block, which employs channel-wise max and average pooling followed by an MLP to weight the importance of different temporal channels adaptively, enhancing the modeling of long-range dependencies. The complete CloudPredUNet model is trained end-to-end using the mean squared error (MSE) loss function, optimizing it to minimize pixel-wise prediction error directly. **Result** Extensive experiments were conducted on a dataset constructed from the FY-2D geostationary meteorological satellite, covering the region 26° E—146° E, 60° S—60° N. The dataset was temporally and spatially partitioned to ensure no overlap between training, validation, and independent test sets. CloudPredUNet was evaluated against a comprehensive suite of state-of-the-art models, including traditional optical flow, autoregressive models (ConvLSTM, PredRNN, MIM), and nonautoregressive models (MMVP, SimVP, TAU). Results demonstrate the superior performance and remarkable efficiency of CloudPredUNet. In terms of model efficiency, CloudPredUNet is exceptionally lightweight, containing only 0. 101 million parameters and requiring approximately five GFLOPs per forward pass. This represents a drastic reduction compared with competitors; for instance, its parameter count is less than one-fifth of the next most efficient model (MMVP), and its computational cost is nearly two orders of magnitude lower than ConvLSTM. Regarding prediction accuracy, CloudPredUNet achieved the best

scores across all four primary evaluation metrics on the test set: MAE of 8.393, MSE of 188.922, SSIM of 0.825, and PSNR of 25.367. This consistent outperformance indicates its exceptional capability in minimizing prediction error while maintaining high structural fidelity and image quality relative to the ground truth. A critical test for any sequence prediction model is its performance in multistep forecasting. CloudPredUNet demonstrated outstanding robustness against error accumulation. In a five-frame-ahead prediction task, it secured the lowest MAE for every single predicted frame. Notably, by the fifth frame, its MAE (10.589) was approximately 2.4% lower than that of the second-best model, MIM (10.852), underscoring its superior stability for longer-term predictions. Ablation studies systematically validated the contribution of each proposed module. Removing any of the three core modules (FDSA, MSA, SFE) led to a measurable drop in performance. The complete model provided an 11.63% improvement in MAE over a baseline U-Net without these modules, confirming their effectiveness and complementary nature. Qualitative visual analysis further reinforced these quantitative findings. In complex meteorological scenarios such as cloud growth, cloud merging, and the preservation of a typhoon eye's characteristic structure, CloudPredUNet generated predictions that were visually more coherent, detailed, and consistent with the actual evolution compared with other models. It successfully maintained sharp boundaries and realistic textures where other models produced blurring, fragmentation, or loss of critical features. **Conclusion** This study addresses key limitations in satellite cloud image prediction, namely, the modeling of nonlinear cloud dynamics, the efficient capture of global spatial dependencies, and the mitigation of multistep error accumulation, by introducing CloudPredUNet. The core innovations lie in its integrative design: The frequency domain self-attention module enables efficient global context modeling; the multiscale spatial attention module enhances detail reconstruction; and the temporal mapper with its spatiotemporal feature extraction module strengthens long-range evolutionary pattern learning. The experimental outcomes affirm that CloudPredUNet successfully bridges the gap between high predictive accuracy and operational efficiency. It establishes a new state-of-the-art on the FY-2D benchmark while being orders of magnitude more parameter- and compute-efficient than prevailing methods. This combination makes it a highly promising candidate for real-time, operational short-term cloud forecasting systems where computational resources may be constrained. Future research will focus on several avenues to advance this work toward greater practical utility and robustness, including the following: 1) expanding the spatiotemporal diversity of training and testing datasets to evaluate and improve the model's generalizability across different seasons, climatic zones, and synoptic conditions; 2) investigating multimodal fusion frameworks that incorporate ancillary meteorological data fields (e.g., wind vectors, humidity, atmospheric pressure) to inject more physical constraints into the learning process, potentially improving the prediction of physically coherent evolutions; and 3) further refining the frequency-domain processing mechanism, potentially by integrating adaptive filters or multiresolution analysis, to better handle high-frequency noise and further enhance the reconstruction of subtle cloud microstructures. Through these efforts, we aim to evolve CloudPredUNet from a high-performing research prototype into a reliable and versatile tool for operational meteorology.

Key words: satellite cloud image prediction; spatiotemporal prediction; FY-2D satellite; frequency domain self-attention; spatial attention

0 引　言

随着气象卫星技术的飞速发展,高频次、高分辨率的卫星云图已成为现代气象监测与预报的核心数据源。云图清晰地展示了云系的宏观分布、纹理结构及亮度温度等关键信息,为分析云团的生消、移动以及强对流天气的演变提供了不可或缺的视觉依据。准确的短时云图预测对于暴雨、台风等极端天气的预警、航空安全、农业生产及公众日常生活保障具有至关重要的作用。然而,传统的气象业务系统在云图预测方面仍面临严峻挑战:一方面,卫星数据的传输与接收过程可能存在延迟或丢失,影响了预报的实时性;另一方面,大气运动本质上是混沌且非线性的,云团在移动过程中会经历复杂的形变、旋转、合并与分裂等过程,这使得依靠人工经验或简单的线性外推方法难以捕捉其真实的动态演变规律。因此,探索能够自动、精准地预测未来云图序列的智能方法,已成为提升气象预报技术水平的前沿课题。

如图1所示,时空预测方法总体上可分为自回

归与非自回归两大框架。为系统梳理相关技术的发展脉络，下文将首先阐述自回归框架的核心机制与代表性模型，随后分析非自回归框架为提升效率所做的改进与面临的挑战，最后综述应用于卫星云图预测的具体方法，从而为本文提出的新模型奠定理论基础。

自回归框架是时空预测中的经典范式，其核心机制是循环递归，即模型以上一时刻的输出作为下一时刻的输入，逐步、顺序地生成未来帧序列。这类模型通常基于循环神经网络构建，尤其擅长建模时间依赖性。ConvLSTM（Shi等，2015）开创性地将传统LSTM（long short-term memory）中的全连接操作替换为卷积运算，使其能够同时捕捉空间特征和时间动态，为深度学习应用于时空序列预测奠定了基石。此后，一系列改进模型被提出以增强性能。例如，PredRNN（predictive recurrent neural network）（Wang等，2017）引入了“时空记忆流”机制，通过在不同时间层之间传递记忆状态，增强了短期时空动态的建模能力。为进一步缓解梯度消失问题，PredRNN++（Wang等，2018），引入梯度高速公路单元和Causal-LSTM模块。为提升对局部运动的感知能力，E3D-LSTM（Wang等，2019）将3D卷积融入循环单元，增强了模型对短期时空特征的提取能力。尽管自回归模型在捕获复杂时间动态方面表现出色，但其序列生成的本质导致了推理速度较慢，且存在误差累积问题，即前期预测的偏差会直接影响后续帧的生成质量。

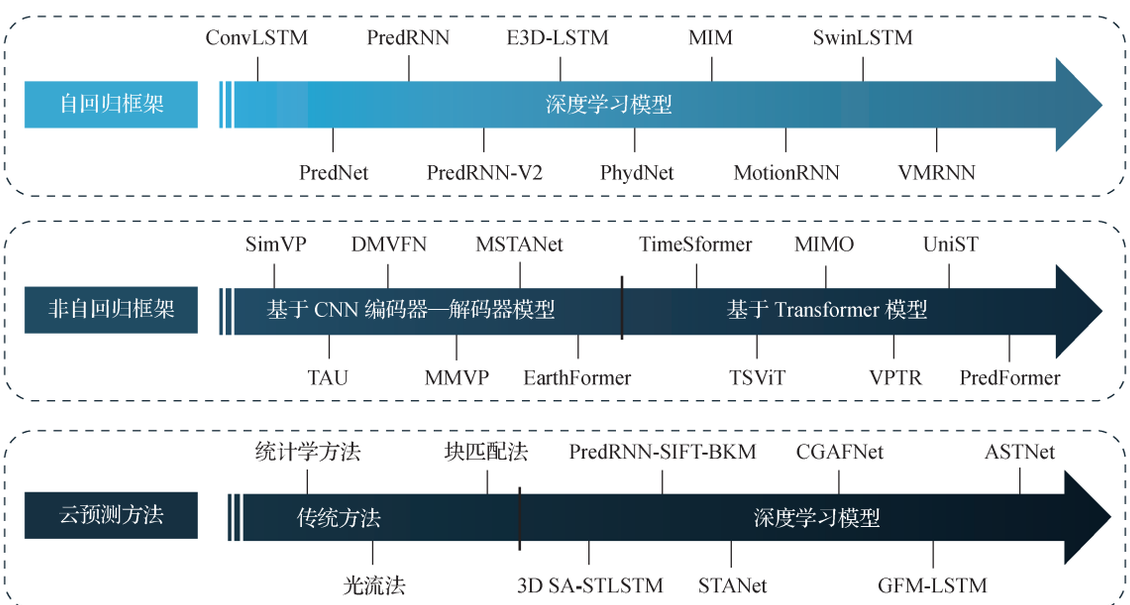


图1　时空预测框架与云预测方法

Fig. 1　Spatiotemporal prediction framework and cloud-based prediction method

为了克服自回归模型在推理效率上的瓶颈，非自回归框架应运而生。该类模型摒弃了逐步递归的方式，采用编码器—解码器结构，将整个历史帧序列一次性编码为潜在表示，然后并行地解码出所有未来预测帧，从而极大地提升了计算速度。早期方法如深度像素流（Liu等，2017）和PredCNN（Xu等，2018）依赖卷积神经网络（convolutional neural network，CNN）进行特征提取与重建。随后，SimVP（simpler yet better video prediction）（Gao等，2022）通过简单堆叠CNN块学习时空演化，显示出CNN在高效建模中的潜力。基于这一发现，时间注意力单元（Tan等，2023）提出了将时间注意力分解为帧内注意力和帧间注意力的方法，并得到了很好的性能。在特定应用如北极海冰预测中，WRANet（wavelet-based multi-scale residual aggregation network）（弓政等，2025）进一步扩展了非自回归架构，通过引入小波多尺度特征提取模块与渐进残差聚合结构，在保留高频细节与多层次特征的同时实现高效并行预测，显著提升了海冰密集度预测的精度与效率。另一方面，受视觉Transformer成功的启发，基于Transformer的模型也被引入时空预测。例如，PredFormer（Tang等，2024）采用纯Transformer架构，通过自注意力机制有效捕获长程时空依赖，解决了CNN感受野有限和循环神经网络（recurrent neural network，

RNN)计算成本高的问题;Ye和Bilodeau(2023)基于高效局部时空分离注意力的视频预测VPTR(video prediction Transformer),在保持性能的同时降低计算复杂度。非自回归模型虽然在推理速度上优势明显,但其一次性预测所有未来的特性使其在捕捉精细的长期时间依赖性方面面临挑战,有时可能导致预测结果过于平滑或丢失细节。

然而,现有的卫星云图时空序列预测方法仍面临诸多挑战:一方面,基于卷积神经网络的模型大多依赖层叠卷积操作,虽然能有效提取局部空间特征,但难以捕获云系在大范围空间中的全局依赖关系,导致对云团形变、旋转等非线性运动的建模能力有限;另一方面,非自回归框架在时间维度上面临长期依赖性捕捉不足或细节丢失的问题,尤其在多步预测中,误差累积与特征平滑现象响了预测结果的清晰度与结构完整性。

卫星云图序列预测的传统方法主要围绕如何从连续的图像中量化云系的运动信息展开,可大致归纳为统计学方法、光流法和块匹配法三大类别(郑行钰等,2024)。1)统计学方法的核心在于分析连续帧之间云图特征的统计相关性以估算位移矢量。该方法直观且计算量相对较小,例如Endlich等人(1971)开发的SATS(sequences of applications technology satellite)系统通过特征量提取和模板匹配追踪云团(Endlich等,1971),而Smith和Phillips(1972)构建的WINDCO(WIND computation)系统则利用二维互相关分析在数字图像序列上追踪云的运动。2)光流法是一种基于像素级别的运动估计技术,通过计算图像序列中像素强度随时间变化的梯度来获取每个像素的瞬时运动速度场。Dissawa等人(2017)将光流法与交叉相关法结合,实现了短时(如3分钟)的云位置预测;顾轶等人(2023)则利用光流法获取云运动矢量并进行外推,通过引入拉普拉斯算子刻画云层扩散,提升了预测精度。3)块匹配法将云图分割成多个小块(或区域),通过在后续帧中搜索最相似的匹配块来计算云体的运动矢量。龚克等人(2000)将GMS(geostationary meteorological satellite)云图分割为小块并匹配参考图以获取运动矢量,实现了量化预测;Brad和Letia(2002)则提出了结合最优候选块搜索和矢量值正则化的改进算法,以提升块匹配的准确性和鲁棒性。这3类传统方法为卫星云图运动预测奠定了重要的技术基础。然而,它们共同面临的挑战在于对云系的非线性变化(如生消、旋转)捕捉能力有限,预测误差会随时间步长增加而累积,且难以有效处理大规模复杂天气系统。这些局限性催生了后续深度学习等新方法的探索与应用。

卫星云图序列预测在深度学习方法取得了进展,这些模型通过捕捉云团的复杂时空动态,有效提升了预测精度和实用性。总体而言,当前研究主要围绕增强时空特征提取和优化计算效率展开:例如,方巍等人(2023)提出的基于3D卷积和自注意力的模型融合了短期趋势和长期依赖关系;Lu等人(2023)开发的STANet(spatio-temporal-aware network)利用注意力机制统一建模瞬态变化和累积趋势;康奇秀等人(2024)设计的CGAFNet(ConvGRU attention fusion network)通过门控循环注意力和主副损失函数改善了中小尺度云团预测;Lian等人(2024)提出的Rep-SSCIPN(re-parameterized sequence-to-sequence satellite cloud imagery prediction network)采用重新参数化序列到序列结构以降低计算成本;而Ren等人(2024)提出的SAM-Net(self-attention memory network)则引入自注意力记忆模块处理长时空依赖性。此外,多任务学习的融合策略通过联合优化相关任务,旨在提升预测结果的实用性和清晰度。这类方法直接应对预测图像模糊、细节丢失的挑战。例如,姜苏城(2024)提出的GFM-LSTM(gated fusion and motion aware LSTM)模型通过门控融合与运动感知模块改善序列预测中的空间信息保留,并进一步设计了AMST(attention multi-scale Transformer)超分辨率重建网络以增强图像纹理细节。同样,李佳欣(2024)在完成ASTNet(attention spatio-temporal network)预测研究后,也提出了SOUL(a new satellite cloud super-resolution method)超分辨率模型,利用多尺度Transformer架构提升重建能力。这种"预测—重建"的协同范式,能够输出更清晰、细节更丰富的最终结果,体现了多任务学习的优势。

为应对上述挑战,本文提出了一种融合频域自注意力的云图预测网络CloudPredUNet。该模型在编码阶段引入了频域注意力模块,通过对输入特征进行局部图像块划分、傅里叶变换及频域互相关计算,以较低复杂度实现全局特征聚合与细粒度纹理增强;解码路径采用多尺度空洞卷积与空间注意力机制,有效恢复细节并抑制冗余背景;时间映射器则

通过多分支异感受野卷积与通道重标定策略，强化长程时空特征的提取与融合。CloudPredUNet 在保持高效推理的同时，提升了云图序列预测的精度与视觉质量。本文的主要贡献如下：

1）提出了一种新颖的 CloudPredUNet 网络，通过频域自注意力设计，提升了卫星云图序列的预测精度。与现有频域注意力方法（如傅里叶神经网络）通常在全图尺度进行频谱变换不同，本文设计了局部补丁划分策略。该方法通过局部—全局平衡的频域注意力设计，在增强云图结构化表征能力的同时，降低了计算复杂度，实现了高效且精准的长程依赖建模。

2）解码器空间注意力模块通过串联不同膨胀率的深度可分离卷积构建多尺度感受野，并结合空间双池化机制（最大池化与平均池化）生成空间注意力权重图。该权重图能自适应增强云图中结构区域（如云核边缘），抑制均匀背景，从而在克服单尺度空洞卷积栅格化伪影的同时，提升对云图多尺度结构的细节重建能力。

3）开发了时间映射器中的时空特征提取模块，设计了一种多分支异感受野卷积结构，融合了方形卷积核与条形卷积核以分别捕获各向同性与各向异性的时空演化模式，并引入基于双池化的通道重标定机制对时间维度特征进行自适应加权。相较于依赖全局自注意力的时空 Transformer 方法，该模块在保持对长程时空依赖有效建模的同时，降低了模型的参数量与计算复杂度，更适用于长序列卫星云图的高效预测。

CloudPredUNet 在 FY-2D 数据集上多项评价指标均优于现有主流方法，证明了其在卫星云图时空预测任务中的有效性与先进性。

1 方　法

1.1 问题的定义

卫星云图的预测不同于传统的时间序列任务，在关注时间信息变化的同时，还需要关注空间的特征变化，因此，卫星云图序列预测可看做是一个时空序列预测任务。其本质就是给定一个连续时间序列长度为 T_{in} 的多通道时空观测序列：$\boldsymbol{X}=\left(\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_{T_{\text{in}}}\right),\boldsymbol{X}_t\in\mathbf{R}^{C\times H\times W}$ 作为模型的输入，预测未来 T_{out} 个时间步：$\hat{\boldsymbol{Y}}=\left(\hat{\boldsymbol{Y}}_1,\hat{\boldsymbol{Y}}_2,\cdots,\hat{\boldsymbol{Y}}_{T_{\text{out}}}\right),\ \hat{\boldsymbol{Y}}_k\in\mathbf{R}^{C\times H\times W}$，其中，$C$ 为通道数，H 为高度，W 为宽度，实现从历史到未来的映射函数：$\hat{\boldsymbol{Y}}=f_\theta(\boldsymbol{X}),f_\theta:\mathbf{R}^{T_{\text{in}}\times C\times H\times W}\rightarrow\mathbf{R}^{T_{\text{out}}\times C\times H\times W}$，其中 θ 为可学习参数。

1.2 模型概述

为实现高精度的卫星云图序列预测任务，本文提出了一种融合频域与多尺度时空注意力的云图预测网络 CloudPredUNet，整体结构如图 2（a）所示。首先，将前 5 个时刻的云图序列输入模型，经逆瓶颈与多级编码器逐层提取特征；编码阶段引入频域注意力模块，通过局部分块的傅里叶变换与逆变换进行相关建模，提升全局与细粒度纹理表征。其次，在时间映射器中使用时空特征提取模块，提取多尺度的时空特征，随后使用时间尺度采样，实现 T_{in} 到 T_{out} 的转换。在解码路径中采用多尺度空洞深度卷积与通道融合的空间注意力模块增强细节恢复，并辅以大核卷积加上时间注意力的时空增强块强化不同感受野下的结构感知；上采样阶段利用 Pixel Shuffle 方法逐步重建空间分辨率。最终，输出未来 5 个时刻的云图序列，实现对卫星云图动态演变的高质量预测。

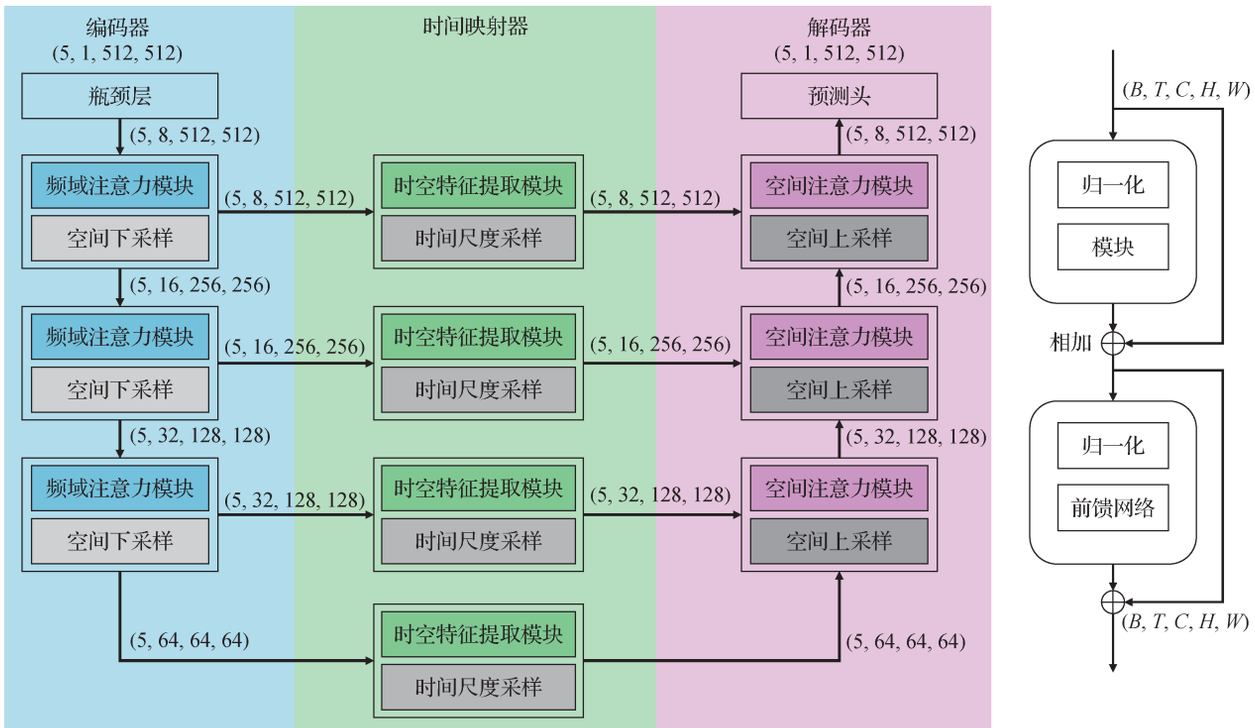
块结构如图 2（b）所示，编码器频域注意力模块、解码器空间注意力模块和时间映射器时空特征提取模块都被块结构所包裹（即下面的模块），块结构由残差 + 前馈网络（feed-forward network，FFN）+ 归一化（batch normalization，BN）+ BA（block attention）模块组成，使用可学习缩放因子 α_1 与 α_2，以提高深层网络训练的稳定性。块结构计算式为

$$\boldsymbol{X}^{(1)}=\boldsymbol{X}+\alpha_1 BA\left(BN\left(\boldsymbol{X}\right)\right) \tag{1}$$

$$\boldsymbol{X}^{(2)}=\boldsymbol{X}^{(1)}+\alpha_2 FFN\left(BN\left(\boldsymbol{X}^{(1)}\right)\right) \tag{2}$$

1.3 编码器频域注意力模块

传统卷积仅在固定局部窗口内建模，难以在不增加深度的情况下获得全局依赖，而基于全局自注意力的相关性计算，其复杂度为 $\mathrm{O}\left((HW)^2\right)$，在高分辨率输入下会带来难以承受的计算和内存负担。频域变换可以将空间卷积转化为频谱逐元素相乘，从而用较低的 $\mathrm{O}\left(HW\log(HW)\right)$ 复杂度实现跨区域相关聚合。本文引入“局部补丁划分加上频域互相关（Hadamard 积）”策略：利用补丁尺度在全局与局部之间取得平衡；通过局部傅里叶变换抑制高频噪声，同时保留频谱结构；结合深度卷积与通道混合增强低频与高频之间的信息流动。



(a) 模型结构　　(b) 块结构

图2　CloudPredUNet模型概述图

Fig. 2　CloudPredUNet model overview diagram ((a) model structure map; (b) block structure map)

编码器频域注意力模块如图3所示。给定输入特征 $\boldsymbol{X} \in \mathbf{R}^{B\times T\times C\times H\times W}$，首先特征投影与划分：经过通道卷积得到张量 $\boldsymbol{H} \in \mathbf{R}^{B\times T\times 6C\times H\times W}$，沿通道维度均分为 $\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V} \in \mathbf{R}^{B\times T\times 2C\times H\times W}$，随后将 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 按设定的补丁尺寸(patch_size = 16)划分为局部块，得到 $\boldsymbol{Q}_p$ 与 $\boldsymbol{K}_p$；接着频域互相关计算：对划分后的 $\boldsymbol{Q}_p$ 和 $\boldsymbol{K}_p$ 执行二维快速傅里叶变换，得到频谱表示 $\mathcal{F}_Q = FFT\left(\boldsymbol{Q}_p\right)$ 和 $\mathcal{F}_K = FFT\left(\boldsymbol{K}_p\right)$，在频域进行逐元素乘法(Hadamard 积)，实现互相关 $\mathcal{F}_O = \mathcal{F}_Q \odot \mathcal{F}_K$，该操作在频域等价于空间域的相关或卷积运算，但计算效率更高；然后逆变换、门控融合与输出投影：对 $\mathcal{F}_O$ 执行逆快速傅里叶变换(inverse fast Fourier transform，IFFT)，恢复至空间域并折叠复原到整图，与值 $\boldsymbol{V}$ 通过逐元素相乘进行门控融合；最后进行输出投影 $\boldsymbol{Y} = \boldsymbol{W}_1\left(\boldsymbol{V}\odot IFFT\left(\mathcal{F}_O\right)\right)$。编码器频域注意力模块表示为

$$FA(\boldsymbol{X}) = \boldsymbol{W}_1\left(\boldsymbol{V}\odot\left(IFFT\left(FFT\left(\boldsymbol{Q}_p\right)\odot FFT\left(\boldsymbol{K}_p\right)\right)\right)\right) \tag{3}$$

式中，FFT 表示快速傅里叶变换，$IFFT$ 表示快速傅里叶变换的逆变换，该方法使用多头(头数为8)频域注意力，$\boldsymbol{W}_i(i = 1, \cdots, 5)$ 为权重矩阵。

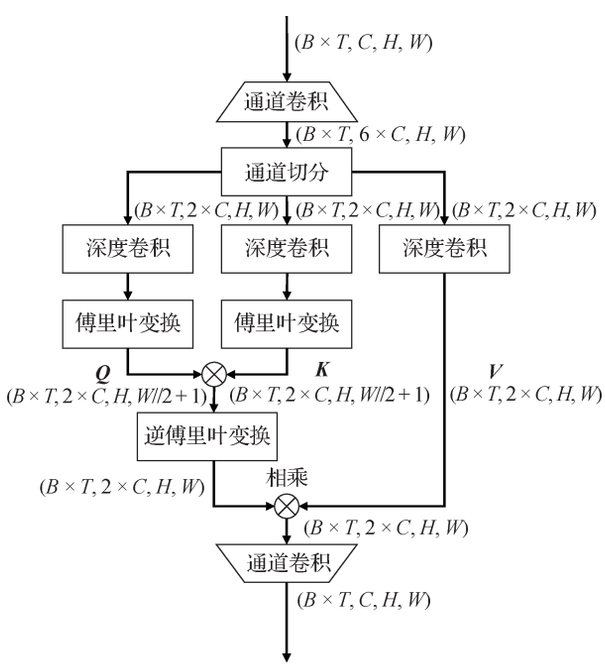


图3　编码器频域注意力模块

Fig. 3　Encoder frequency domain attention module

1.4　解码器空间注意力模块

单尺度卷积对复杂尺度变化(边缘、细纹理与大范围区域形态)具有偏置，而单一尺度的空洞卷积易产生栅格化伪影。为解决多尺度上下文聚合与结构

性抑制冗余背景之间的矛盾，本文串联多膨胀率深度卷积分支（形成逐级感受野扩张的层次），并通过级联特征的通道重新组合及双通道池化生成空间注意图，以在增强结构对比的同时抑制重复响应。这一设计利用轻量的深度可分离卷积降低参数量，并与解码器多尺度融合无缝耦合，提升重建阶段的细节与尺度自适应鲁棒性。

解码器空间注意力模块如图4所示。首先输入 $\boldsymbol{X} \in \mathbf{R}^{B\times T\times C\times H\times W}$，依次通过串联的多空洞深度可分卷积分支得到 $\boldsymbol{Y}_0 = f_{d=1}(\boldsymbol{X})$，$\boldsymbol{Y}_1 = f_{d=r_1}(\boldsymbol{Y}_0 + \boldsymbol{X})$，$\boldsymbol{Y}_2 = f_{d=r_2}(\boldsymbol{Y}_1 + \boldsymbol{X})$，$\boldsymbol{Y}_3 = f_{d=r_3}(\boldsymbol{Y}_2 + \boldsymbol{X})$，然后级联压缩并激活 $\boldsymbol{F} = \phi\left(\left[\boldsymbol{Y}_0, \boldsymbol{Y}_1, \boldsymbol{Y}_2, \boldsymbol{Y}_3\right]\boldsymbol{W}_2\right)$；接着对 $\boldsymbol{F}$ 沿着空间维度做最大和平均池化得到 $Max_c(\boldsymbol{F})$ 和 $Avg_c(\boldsymbol{F})$，沿着通道维度拼接后经多层感知机（multilager perceptron，MLP）与激活函数得到空间注意力权重 $\boldsymbol{A} = \sigma\left(MLP\left[Max_{hw}(\boldsymbol{F}), Avg_{hw}(\boldsymbol{F})\right]\right)$，进而加权并投影输出 $\boldsymbol{Y} = \boldsymbol{W}_3 * (\boldsymbol{F}\odot\boldsymbol{A})$。解码器空间注意力模块表示为

$$\mathrm{SA}(\boldsymbol{X}) = \boldsymbol{W}_3 * \left(\boldsymbol{F}\odot\sigma\left(MLP\left[Max_{hw}(\boldsymbol{F}), Avg_{hw}(\boldsymbol{F})\right]\right)\right) \tag{4}$$

$$\boldsymbol{F} = \phi\left(\left[f_1(\boldsymbol{X}), f_{r_1}(\cdot), f_{r_2}(\cdot), f_{r_3}(\cdot)\right]\boldsymbol{W}_2\right) \tag{5}$$

式中，σ 表示 sigmoid 激活函数，ϕ 表示 GELU（Gaussian error linear unit）激活函数。

1.5 时间映射器时空特征提取模块

时空特征提取模块通过多分支异感受野卷积结构协同提取长程空间依赖与时间维度语义特征，在扩大感受野的同时保持计算效率。其中，由方形深度卷积与横向、纵向条形卷积组成的空间分支，能够有效捕捉各向异性的形态演化模式；而通过最大池化与平均池化双路汇聚的通道注意力机制，则对通道堆叠数据的时间信息进行动态权重标定，强化关键时间步的语义表达。该设计实现了空间结构建模与时间通道筛选的解耦与互补：空间卷积分支专注形态演变的连续性与方向性特征；通道重标定机制则保障时间维度语义不被空间卷积淹没，二者通过加权融合形成耦合增强。

时间映射器时空特征提取模块如图5所示。输入 $\boldsymbol{X} \in \mathbf{R}^{B\times T\times C\times H\times W}$，将时间维度 T 和通道维度 C 合并。首先将张量沿着通道维度，按照5/8、1/8、1/8和1/8的比例切分为4个子集 $\left[\boldsymbol{U}_0, \boldsymbol{U}_{hw}, \boldsymbol{U}_w, \boldsymbol{U}_h\right]$，$\boldsymbol{U}_0$ 保持不变，其余通道分别施加 3 × 3 深度卷积与 1 × 11、11 × 1 条形卷积得到 $\boldsymbol{V}_{hw}, \boldsymbol{V}_w, \boldsymbol{V}_h$，再沿着通道维度拼接并投影 $\boldsymbol{Z} = \boldsymbol{W}_4 * \left[\boldsymbol{U}_0, \boldsymbol{V}_{hw}, \boldsymbol{V}_w, \boldsymbol{V}_h\right]$；然后对 $\boldsymbol{Z}$ 沿着通道维度做最大与平均池化生成 $Max_c(\boldsymbol{Z})$ 和 $Avg_c(\boldsymbol{Z})$，两者相加经 MLP 与激活函数得到通道注意力权重 $\boldsymbol{A} = \sigma\left(MLP\left[Max_c(\boldsymbol{F}), Avg_c(\boldsymbol{F})\right]\right)$，最后加权输出 $\boldsymbol{Y} = \boldsymbol{W}_5 * (\boldsymbol{Z}\odot\boldsymbol{A})$。时间映射器时空特征提取模块表示为

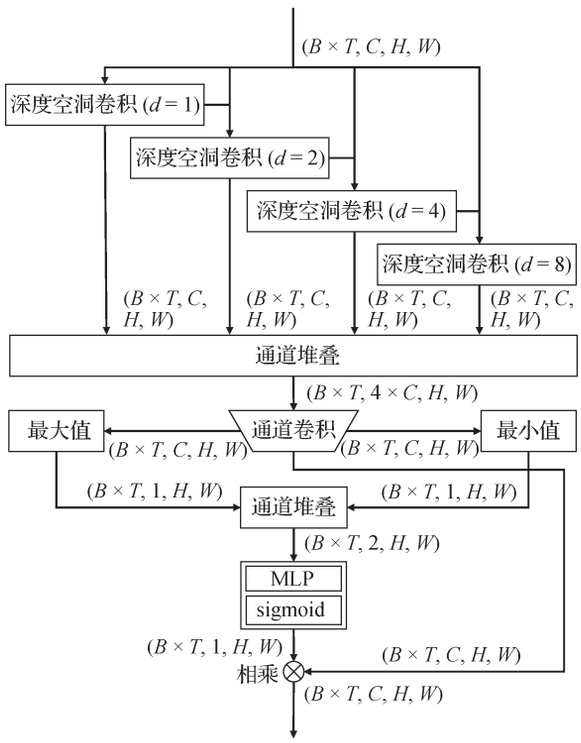


图4 解码器空间注意力模块

Fig. 4 Decoder spatial attention module

$$STA(\boldsymbol{X}) = \boldsymbol{W}_5 * \left(\boldsymbol{Z}\odot\sigma\left(\mathrm{MLP}\left[\mathrm{Max}_c(\boldsymbol{Z}), \mathrm{Avg}_c(\boldsymbol{Z})\right]\right)\right) \tag{6}$$

$$\boldsymbol{Z} = \boldsymbol{W}_4 * \left[\boldsymbol{U}_0, f_3(\boldsymbol{U}_{hw}), f_{1\times 11}(\boldsymbol{U}_w), f_{11\times 1}(\boldsymbol{U}_h)\right] \tag{7}$$

2 网络训练

2.1 数据集

2.1.1 数据源

本研究采用静止气象卫星——中国的风云二号D星（FY-2D）的观测数据。FY-2D卫星于2006年12月8日发射成功，定点于86.5°E赤道上空。FY-2D空间覆盖范围为：26°E—146°E，60°S—60°N，如图6所示。

FY-2D红外通道共有4个波段，空间分辨率为5 km，如表1所示。为完成本文研究，收集了卫星数

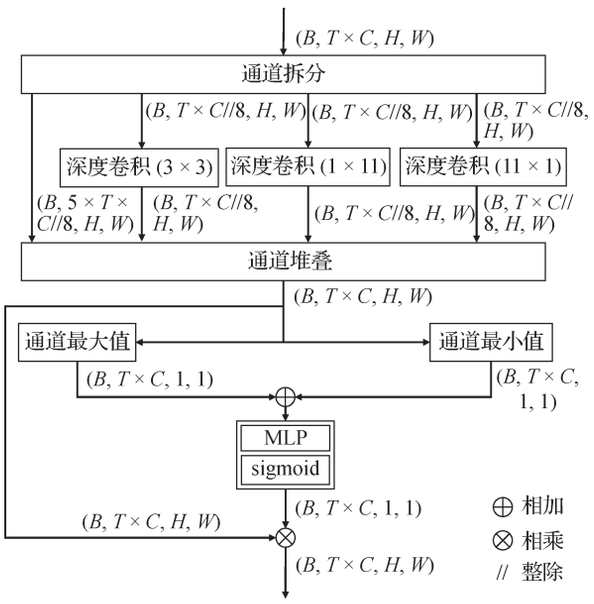


图5　时间映射器时空特征提取模块

Fig. 5　Time mapper spatiotemporal feature extraction module

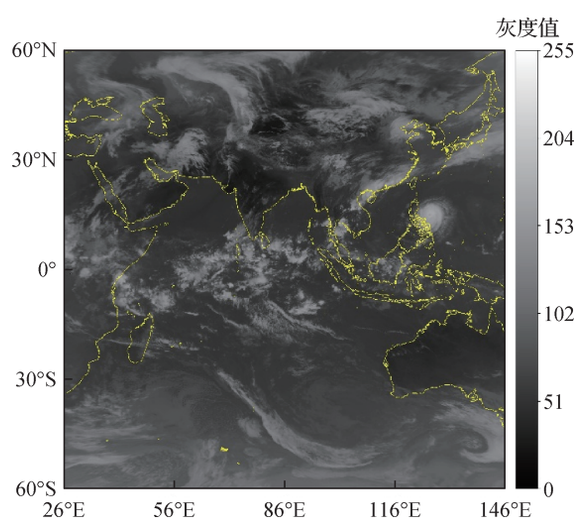


图6　FY-2D卫星图像

Fig. 6　FY-2D satellite imagery

据，时间跨度自2015年5月9日5时至2015年5月9日21时，每隔一个整点进行一次样本选取，总计获得17个原始文件。

表1　FY-2D波段信息

Table 1　FY-2D band information

波段	光谱范围/μm	分辨率/km
IR1(红外)	10.3～11.3	5
IR2(红外)	11.5～12.5	5
IR3(红外)	6.3～7.6	5
IR4(红外)	3.5～4.0	5

2.1.2　数据划分

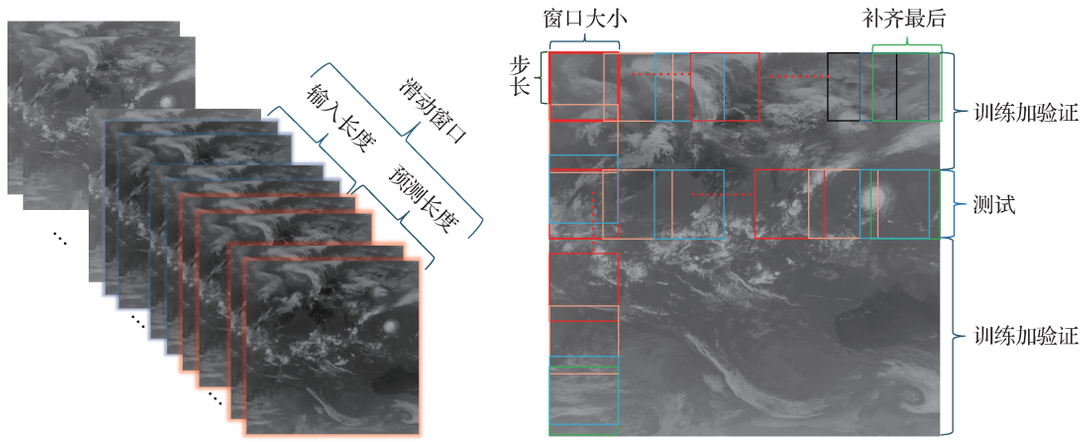
FY-2D数据集的划分方式如图7所示，在总时长为17帧的时间尺度上，以前N帧预测后M帧，采用滑动窗口方式(步长为1)生成时间样本，共得到$(12-N-M+1)$组数据。图像原始尺寸为3 000×3 000像素，在空间维度上采用窗口大小为Q、步长为K的滑动裁剪策略，从而生成$\left(\frac{3\,000-Q}{K}+1\right)$组空间区域。其中，测试集在空间上独立，仅选取一行数据，与训练集和验证集无任何空间区域重叠。具体实现中，时间上采用输入前5帧预测后5帧的滑动窗口设置，共生成8个长度为10帧的时间窗口(步长为1帧)；空间上使用512×512像素的窗口进行滑动裁剪(步长100像素)，得到390个训练/验证子区域和26个测试子区域。最终，训练集和验证集共包含3 120个样本(390个空间区域×8个时间窗口)，按8∶2比例划分为训练集2 496个样本和验证集624个样本；测试集共计208个样本(26个空间区域×8个时间窗口)。

2.2　实验环境和参数设置

本实验的硬件环境为搭载NVIDIA GeForce RTX 3080显卡的windows计算机平台，软件环境基于Python 3.9.21与PyTorch 2.1.0深度学习框架，并启用CUDA 12.1进行GPU加速计算。模型输入数据均为五维张量，维度依次表示批量大小、时间步长、通道数、图像高度与宽度，具体尺寸为(16, 5, 1, 512, 512)，即每批次输入16个样本，每个样本包含5帧单通道512×512像素分辨率的序列图像。训练过程中使用AdamW优化器，初始学习率设为1×10^{-5}，批量大小为16，最大训练轮次为100，并采用基于验证集损失的学习率衰减策略：若验证损失连续5轮未下降，则将学习率减半，以促进模型收敛至更优状态。

2.3　评价指标

选用均方误差(mean square error，MSE)、平均绝对误差(mean absolute error，MAE)、结构相似性(structural similarity index measure，SSIM)和峰值信噪比(peak signal to noise ratio，PSNR)评估模型的预测性能(Setiadi，2021)。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



(a) 时间窗口划分图　　(b) 空间窗口划分图

图 7　FY-2D 数据集的划分方式

Fig. 7　Division method of the FY-2D dataset((a)temporal window division diagram;(b)spatial window division diagram)

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、SSIM 和 PSNR 越大，说明模型的准确性越好，预测结果也越好。

2.4　损失函数

为了有效地引导模型学习气象云图在时空维度上的演化规律，并确保预测结果在像素级精度上与真实数据保持一致，本工作选择均方误差作为模型的训练损失函数。均方误差是回归预测任务中最为经典和广泛应用的损失函数之一，其定义为预测值与真实值之间差值的平方的期望。对于多个预测帧 $\hat{\boldsymbol{Y}} \in \mathbf{R}^{T\times C\times H\times W}$ 及其对应的真实帧 $\boldsymbol{Y} \in \mathbf{R}^{T\times C\times H\times W}$，其损失函数的计算式为

$$L_{\mathrm{MSE}} = \frac{1}{T\times C\times H\times W}\sum_{t=1}^{T}\sum_{i=1}^{C}\sum_{j=1}^{H}\sum_{k=1}^{W}(Y_{t,i,j,k} - \hat{Y}_{t,i,j,k})^2 \tag{8}$$

式中，C、H 和 W 分别表示特征图的通道数、高度和宽度。在加入时间维度后，用 T 表示时间步长或帧数。此时，$Y_{t,i,j,k}$ 与 $\hat{Y}_{t,i,j,k}$ 则分别代表在时间步(t)、通道(i)、高度(j)、宽度(k)位置上的真实值与模型预测值。

3　实验结果及分析

3.1　模型对比实验分析

在 FY-2D 数据集上的综合评估结果表明，本文所提出的 CloudPredUNet 模型在各项性能指标上均表现出色。实验选取了包括传统光流法在内的 7 种代表性先进模型作为基准，覆盖了经典的循环结构方法(如 ConvLSTM、PredRNN、MIM)以及近年来兴起的无循环架构(如 MMVP、SimVP、TAU)，从而保证对比的广泛性和代表性。

如表 2 所示，CloudPredUNet 模型与其他模型相比，在效率与预测精度方面均展现出优势。在计算效率上，CloudPredUNet 的表现尤为突出，该模型参数量仅 0.101 M，不足 MMVP 模型(参数量次低)的 1/5；在计算复杂度方面，其每秒浮点运算次数(floating point operations per second，FLOPs)(5 G)远低于其他对比模型，如 ConvLSTM 的计算开销是其 87.6 倍。这充分体现了 CloudPredUNet 极为轻量的特性，具备在计算资源受限环境下部署的巨大潜力。在预测性能方面，CloudPredUNet 在 4 项关键评价指标上均领先。它取得了最低的 MAE(8.393)和 MSE(188.922)，表明其预测结果与真实值之间的误差最小、预测精度最高。同时，其 SSIM(0.825)和 PSNR(25.367 dB)均为最高，证明了其在保持云图结构相似性和图像质量方面的卓越能力。综合来看，CloudPredUNet 成功地在“轻量化”与“高性能”之间取得了最佳平衡。它不仅以极低的模型复杂度和计算成本超越了所有基于循环的基线模型，同时也优于现有的先进无循环方法。

根据表 3 所示的多帧预测结果，可以清晰地观察到所有模型的预测误差(MAE)均随着预测时长的增加而逐步上升，这一趋势符合序列预测任务的普遍规律。然而，在此过程中，本文提出的 CloudPredUNet 模型展现出了卓越且稳定的预测性能。具体而言，在从第 1 帧到第 5 帧的整个预测序列中，

表2　不同模型实验评价指标结果

Table 2　Results of experimental evaluation metrics for different models

模型	参数量/M	FLOPs/G	MAE ↓	MSE ↓	SSIM ↑	PSNR/dB ↑
光流法	–	–	10.842	340.338	0.748	18.448
ConvLstm	3.823	438	9.061	226.995	0.805	24.570
PredRNN	24.920	919	9.642	226.743	0.821	24.575
MIM	47.972	1 377	8.799	213.199	0.811	24.842
MMVP	0.475	52	10.561	266.048	0.817	23.881
SimVP	4.280	79	9.416	233.351	0.784	24.450
TAU	4.617	111	8.952	223.446	0.795	24.639
CloudPredUNet(本文)	0.101	5	**8.393**	**188.922**	**0.825**	**25.367**

注:加粗字体表示各列最优结果。↓表示值越低越好,↑表示值越高越好。模型的参数量和浮点运算数(FLOPs)在相同输入配置下计算,光流法是传统算法,不计算这两项指标。“–”表示无参数,无法计算。

CloudPredUNet在每一时间步上均取得了最低的MAE值。这不仅说明了该模型在短期预测中具备最高的起始精度,也证明了其在多帧预测任务中始终保持最佳的预测稳定性。值得注意的是,随着预测帧数的推进,CloudPredUNet相对于其他模型的性能优势呈现出持续性的保持甚至扩大趋势。例如,在预测第5帧时,其MAE较次优的MIM模型(10.852)进一步降低了0.263,展现出更强的长期鲁棒性。

表3　不同模型多帧预测MAE指标结果

Table 3　Results of multi-frame mean absolute error for different models

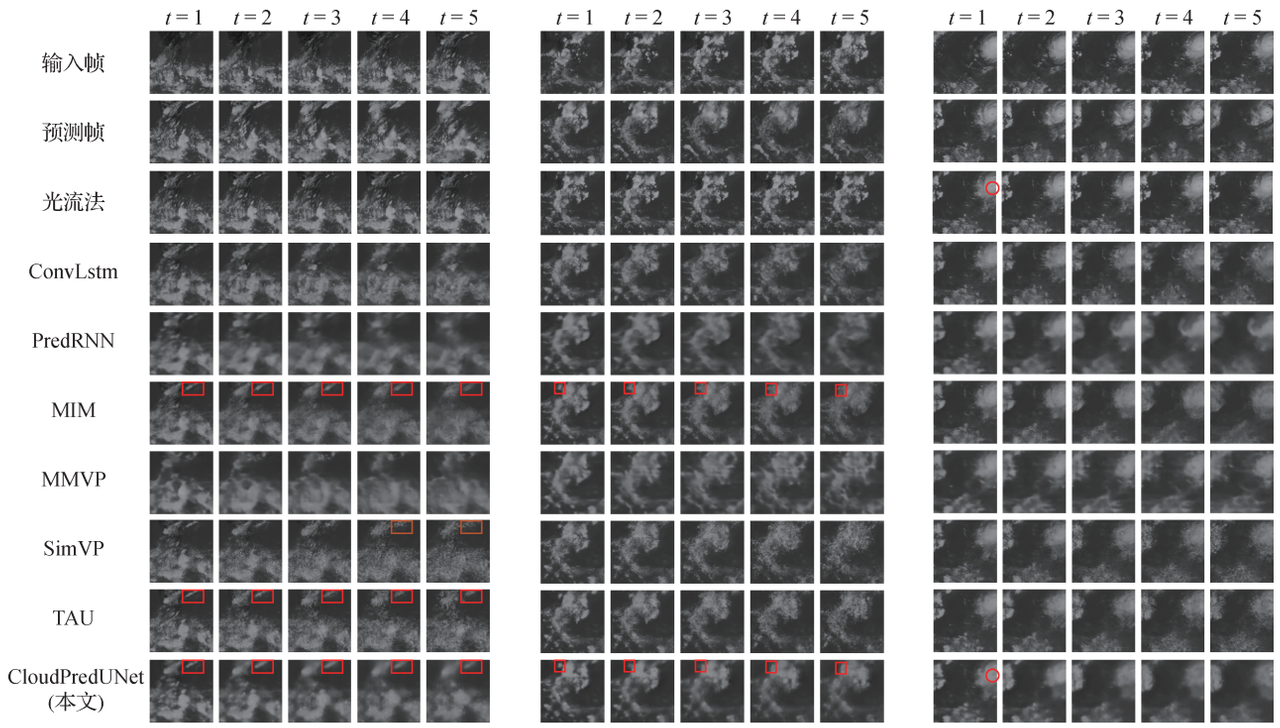
模型	预测第1帧	预测第2帧	预测第3帧	预测第4帧	预测第5帧
光流法	7.563	9.281	11.408	12.510	13.450
ConvLstm	5.813	8.120	9.572	10.548	11.254
PredRNN	6.053	8.215	9.697	10.755	11.488
MIM	5.848	7.891	9.216	10.189	10.852
MMVP	7.190	9.517	11.104	12.140	12.854
SimvP	6.164	8.388	9.851	10.887	11.790
TAU	5.569	7.864	9.376	10.514	11.437
CloudPredUNet(本文)	**5.473**	**7.263**	**8.778**	**9.865**	**10.589**

注:加粗字体表示各列最优结果。

根据图8在FY-2D数据集上的可视化预测结果,可以直观比较各模型的性能差异。图中第1行为输入帧序列,第2行为真实帧,后续各行展示了不同模型的预测输出。从3类典型气象过程(分别对应图8的3列)的建模效果来看:在云团生长场景中,与真实序列相比,MIM、TAU和CloudPredUNet较完整地模拟了云团由小变大的动态演变过程(图8(a),红色框选区域),而SimVP的预测结果则出现了局部离散化现象(图8(a),橙色框选区域);ConvLSTM、PredRNN和MMVP在该场景下未能有效刻画出明显的云团生长过程。在云团合并场景中,多数模型对多云团融合行为的模拟存在不足,仅MIM与CloudPredUNet较准确地再现了这一过程(图8(b),红色框选区域)。在台风眼结构特征提取方面,CloudPredUNet与光流法均能捕捉到台风眼的关键形态(图8(c),红色圆圈区域),但光流法的预测结果在时序上变化幅度较小。综合以上可视化对比可知,CloudPredUNet在建模复杂气象系统的时空演变方面表现相对较好,尤其在云团生长、合并及台风眼结构保持等任务上,呈现出较高的模拟能力和结构一致性。

3.2　云团生长阶段场景分析

针对图8(a)所呈现的云团生长阶段典型场景,本研究综合运用光流法与密度分割方法,从运动特征与内部结构两个维度分析。该时段图像记录了云团在生长初期的动态形态与空间结构,为理解其发展机制提供了关键时序节点。基于光流法,本文一方面通过全网格特征点追踪刻画云系的整体运动趋势,另一方面聚焦云团边缘局部光流场,揭示边界处的扩展动力学;同时,借助密度分割技术对云图灰度



(a) 云生长图　(b) 云合并图　(c) 台风眼特征图

图 8　不同模型的预测结果图

Fig. 8　Prediction results of different models ((a) cloud growth maps; (b) cloud merging maps; (c) typhoon eye feature maps)

层级进行色彩映射，直观呈现云及其周边区域的热力结构演变。以下将结合上述方法，系统阐述该阶段云团在运动行为与内部组织方面的典型特征，并深入探讨其动力与微物理含义。

边界光流分析显示(图 9(c))，云团轮廓上的运动矢量呈现出系统性的外向辐射模式，超过 85% 的边界点的光流方向指向云团外部，这种边界动力学特征充分表明云团正处于快速生长扩展期。在宏观尺度上，整体光流矢量图(图 9(b))不仅清晰地展示了云团随主导气流平移的轨迹，还揭示了云系内部不同区域的运动差异。特别值得注意的是，在发展中期阶段(图 9(b)，蓝色框选区域)，云团前端出现光流矢量的发散特征，而后缘保持汇聚状态，这种不对称结构反映了云团在发展过程中与周边环境场的复杂相互作用。

密度分割序列(图 9(d))细致地展示了个体云团的完整生长过程，其核心区的演变轨迹具有重要的指示意义。在发展初期，云核呈现出高度凝聚的状态；随着云团的生长，核心区范围逐渐扩大，在密度分割图上表现为高密度区域的颜色变浅。这一现象揭示了云核内部的动力和微物理过程——上升气流可能减弱或扩散，导致云顶高度或云水含量的空间分布趋于平均化。而在发展的后续阶段，可以观察到云团范围经历了一个动态的汇聚和重组过程，其核心区随之变得更加集中和紧凑，在图像上表现为颜色再度变深，这可能预示着新的强上升运动中心的形成或云内动力过程的再次增强。

3.3　消融实验

为系统验证 CloudPredNet 中各核心模块对预测性能的贡献，本研究设计了完整的消融实验，通过在基准模型上逐步引入编码器频域注意力模块、解码器空间注意力模块和时空特征提取模块，并在控制变量的基础上，综合评价各组件对模型性能的影响。

表 4 展示了不同模块组合在 FY-2D 数据集上的 MAE 与 SSIM 结果，反映了各模块对预测精度的影响。实验从仅包含基础结构的模型(模型 1)开始(MAE 为 9.497，SSIM 为 0.800)，逐步加入各模块进行性能对比。结果表明，每个模块均能独立带来明显性能提升：加入编码器频域注意力模块后(模型 2)，MAE 降至 8.751，SSIM 提升至 0.812；加入解码器空间注意力模块后(模型 3)，MAE 进一步降低至 8.595，SSIM 提高至 0.825；而时空特征提取模块的加入(模型 4)使得 MAE 下降至 8.640，SSIM 增至

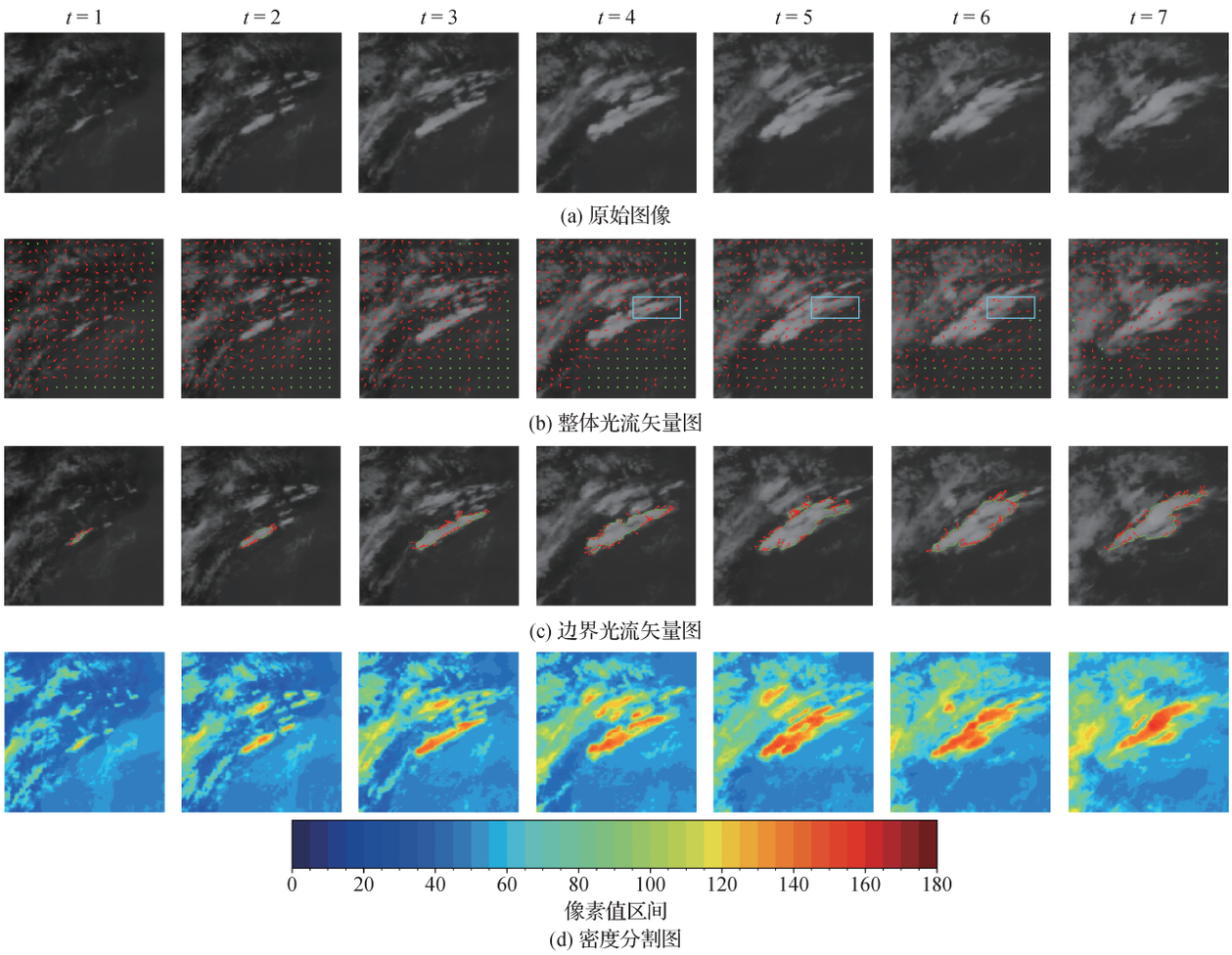


图9　云团生长阶段场景分析

Fig. 9　Scenario analysis of cloud cluster growth stages ((a) original images; (b) overall optical flow vector maps; (c) boundary optical flow vector maps; (d) density segmentation maps)

0.818。当3个模块共同集成于完整模型(模型5)中时,模型取得最优性能,MAE降低至8.393(相较于基准模型提升了约11.62%),SSIM提升至0.835(提升了约4.38%),明显优于任何单一模块的配置。这一结果充分说明,各模块在捕捉全局频域特征、重建空间细节以及建模长程时间依赖方面具有有效性和互补性,从而验证了CloudPredUNet整体架构设计的合理性与必要性。

表4　消融实验评价指标结果

Table 4　Results of ablation experiment evaluation indicators

模型	编码器	解码器	时间映射器	MAE↓	SSIM↑
模型1	–	–	–	9.497	0.800
模型2	√	–	–	8.751	0.812
模型3	–	√	–	8.595	0.825
模型4	–	–	√	8.640	0.818
模型5	√	√	√	**8.393**	**0.835**

注:"√"代表使用该模块,"–"代表不使用。加粗字体表示各列最优结果,↓表示值越低越好,↑表示值越高越好。

4　结　论

本文围绕卫星云图序列预测中非线性运动建模困难、全局依赖捕捉不足及多步预测误差累积等关键问题展开研究。传统方法在刻画云团生消、旋转等复杂动态方面存在局限,而现有深度学习模型在长时序建模与结构保持方面仍有提升空间。为此,本文提出融合频域注意力的云图预测网络CloudPredUNet,通过模块化设计与协同优化,提升了预测精度与结构完整性。

本研究的主要工作与成果可概括如下:提出基

于U-Net架构的CloudPredUNet模型，在编码阶段引入频域注意力模块，通过对局部图像块进行傅里叶变换与频域互相关，以较低计算复杂度实现全局特征聚合；解码路径采用多尺度空洞卷积与空间注意力机制，增强细节重建能力；时间映射器结合多分支异感受野卷积与通道重标定策略，强化长程时空依赖建模。整体设计兼顾了高效计算与高精度预测，适用于云图序列的时空演化建模。

在FY-2D数据集上的实验表明，CloudPredUNet在参数量（0.101 M）和计算量（5 G FLOPs）均低于对比模型的同时，在MAE、MSE、SSIM和PSNR 4项关键指标上均取得最优结果。消融实验验证了各模块的有效性与互补性，完整模型相较于基线在MAE上提升约11.62%。可视化结果进一步显示，该模型能够更好地保持云团生长、合并及台风眼等复杂结构的细节完整性，尤其在多步预测中表现出较强的稳定性与抗误差累积能力。

尽管CloudPredUNet在卫星云图预测任务中取得了较好的效果，本研究仍存在一些局限性：首先，模型在极端天气系统（如强对流云团的快速生消）下的泛化能力有待进一步验证；其次，本研究所使用的FY-2D数据集时间跨度较短（仅2015年5月9日5时至21时，共16 h），空间范围也限于26°E—146°E、60°S—60°N区域，未能覆盖不同季节、不同气候背景下的云图变化，可能影响模型的鲁棒性与泛化能力；此外，频域变换虽然提升了全局建模效率，但对高频噪声较为敏感，在某些情况下可能影响局部细节的还原。因此，模型在当前数据上表现出的优越性主要反映其在该时段与该区域的有效性，而在不同季节、不同气候系统（如高纬、极地或强对流频发区域）以及更长时序预测中的性能仍需进一步验证。

针对上述局限性，未来的研究工作将重点围绕以下方向展开：首先，构建覆盖多季节、多气候区域的长时间序列云图数据集，以系统评估并提升模型在不同气象条件下的适应性与泛化能力；其次，探索多模态数据融合机制，引入如风场、湿度场等辅助气象要素，以增强对极端天气过程中云团生消、旋转等非线性演变的物理建模能力；此外，将进一步优化频域注意力机制，抑制高频噪声干扰，并结合空间局部增强方法，以提升对云图细微结构的重建质量。通过这些改进，旨在推动该预测方法向更稳健、更实用的业务化方向演进。

参考文献（References）

Brad R and Letia I A. 2002. Cloud motion detection from infrared satellite images//Proceedings of 2002 SPIE 4875, Second International Conference on Image and Graphics. Hefei, China: SPIE: 408-412 [DOI: 10.1117/12.477174]

Dissawa D M L H, Ekanayake M P B, Godaliyadda G M R I, Ekanayake J B and Agalgaonkar A P. 2017. Cloud motion tracking for short-term on-site cloud coverage prediction//Proceedings of the 17th International Conference on Advances in ICT for Emerging Regions (ICTer). Colombo, Sri Lanka: IEEE: 1-6 [DOI: 10.1109/ICTER.2017.8257803]

Endlich R M, Wolf D E, Hall D J and Brain A E. 1971. Use of a pattern recognition technique for determining cloud motions from sequences of satellite photographs. Journal of Applied Meteorology, 10 (1): 105-117 [DOI: 10.1175/1520-0450 (1971) 010<0105: UOAPRT>2.0.CO;2]

Fang W, Li J X and Lu W H. 2023. Research on satellite cloud image prediction based on 3D convolution and self-attention. Journal of Nanjing University (Natural Science), 59(1): 155-164 (方巍, 李佳欣, 陆文赫. 2023. 基于3D卷积和自注意力机制的卫星云图预测研究. 南京大学学报(自然科学), 59(1): 155-164) [DOI: 10.13232/j.cnki.jnju.2023.01.015]

Gao Z Y, Tan C, Wu L R and Li S Z. 2022. SimVP: simpler yet better video prediction//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE: 3160-3170 [DOI: 10.1109/CVPR52688.2022.00317]

Gong K, Ye D L and Ge C H. 2000. A method for geostationary meteorological satellite cloud image prediction based on motion vector. Journal of Image and Graphics, 5(4): 349-352 (龚克, 叶大鲁, 葛成辉. 2000. 卫星云图预测的运动矢量方法. 中国图象图形学报, 5(4): 349-352) [DOI: 10.3969/j.issn.1006-8961.2000.04.016]

Gong Z, Zhang J L, Gao F, Gan Y H and Dong J Y. 2025b. Wavelet-based multiscale residual aggregation network (WRANet) for arctic sea ice forecasting. Journal of Image and Graphics, 31(5): 1557-1568 (弓政, 张家亮, 高峰, 甘言海, 董军宇. 2026. 小波多尺度残差聚合北极海冰预测网络WRANet. 中国图象图形学报, 31(5): 1557-1568 [DOI:10.11834/jig.250318]

Gu Y, Han C, Liu J X, Liu S G and Xing W. 2023. Research on large area cloud forecasting method based on satellite cloud images. Chinese Space Science and Technology, 43(2): 165-173 (顾轶, 韩潮, 刘建勋, 刘升刚, 邢炜. 2023. 基于卫星云图的大区域云层预测方法. 中国空间科学技术, 43(2): 165-173) [DOI: 10.16708/j.cnki.1000-758X.2023.0031]

Jiang S C. 2024. Research on Prediction and Reconstruction of Satellite Cloud Images Based on Deep Learning. Nanjing: Nanjing University of Information Science and Technology (姜苏城. 2024. 基于深

度学习的卫星云图预测与重建研究. 南京: 南京信息工程大学)[DOI: 10.27248/d.cnki.gnjqc.2024.001331]

Kang Q X, Du D S, Chen L F, Ou X F and Ye C Z. 2024. Satellite cloud image nowcasting based on CGAFNet. Journal of Tropical Meteorology, 40(6): 1074-1084 (康奇秀, 杜东升, 陈立福, 欧小锋, 叶成志. 2024. 基于CGAFNet的卫星云图临近预报研究. 热带气象学报, 40(6): 1074-1084)[DOI: 10.16032/j.issn.1004-4965.2024.094]

Li J X. 2024. Research on Satellite Cloud Image Prediction and Reconstruction Methods Based on Deep Learning. Nanjing: Nanjing University of Information Science and Technology (李佳欣. 2024. 基于深度学习的卫星云图预测与重建方法研究. 南京: 南京信息工程大学)[DOI: 10.27248/d.cnki.gnjqc.2024.000631]

Lian J, Wu S X, Huang S R and Zhao Q. 2024. A novel sequence-to-sequence based deep learning model for satellite cloud image time series prediction. Atmospheric Research, 306: #107457 [DOI: 10.1016/j.atmosres.2024.107457]

Liu Z W, Yeh R A, Tang X O, Liu Y M and Agarwala A. 2017. Video frame synthesis using deep voxel flow//Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE: 4473-4481 [DOI: 10.1109/ICCV.2017.478]

Lu Z Y, Zhou Z Y, Li X and Zhang J F. 2023. STANet: a novel predictive neural network for ground-based remote sensing cloud image sequence extrapolation. IEEE Transactions on Geoscience and Remote Sensing, 61: #4701811 [DOI: 10.1109/TGRS.2023.3268503]

Ren Y Z, Ye J Y, Wang X C, Xiao F J and Liu R J. 2024. SAM-Net: spatio-temporal sequence typhoon cloud image prediction net with self-attention memory. Remote Sensing, 16(22): #4213 [DOI: 10.3390/rs16224213]

Setiadi D R I M. 2021. PSNR vs SSIM: imperceptibility quality assessment for image steganography. Multimedia Tools and Applications, 80(6): 8423-8444 [DOI: 10.1007/s11042-020-10035-z]

Shi X J, Chen Z R, Wang H, Yeung D Y, Wong W K and Woo W C. 2015. Convolutional LSTM network: a machine learning approach for precipitation nowcasting//Proceedings of the 29th International Conference on Neural Information Processing Systems (NIPS). Montreal, Canada: MIT Press: 802-810

Smith E A and Phillips D R. 1972. Automated cloud tracking using precisely aligned digital ATS pictures. IEEE Transactions on Computers, C-21(7): 715-729 [DOI: 10.1109/t-c.1972.223574]

Tan C, Gao Z Y, Wu L R, Xu Y J, Xia J, Li S Y, et al. 2023. Temporal attention unit: towards efficient spatiotemporal predictive learning//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE: 18770-18782 [DOI: 10.1109/CVPR52729.2023.01800]

Tang Y J, Qi L, Xie F, Li X T, Ma C and Yang M H. 2024. Video prediction transformers without recurrence or convolution [EB/OL]. [2024-10-10]. https://arxiv.org/pdf/2410.04733v3.pdf

Wang Y B, Gao Z F, Long M S, Wang J M and Yu P S. 2018. 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: 5110-5119

Wang Y B, Jiang L, Yang M H, Li L J, Long M S and Li F F. 2019. Eidetic 3D LSTM: a model for video prediction and beyond//Proceedings of the 7th International Conference on Learning Representations. New Orleans, USA: OpenReview.net

Wang Y B, Long M S, Wang J M, Gao Z F and Yu P S. 2017. PredRNN: recurrent neural networks for predictive learning using spatiotemporal LSTMs//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc.: 879-888

Xu Z R, Wang Y B, Long M S and Wang J M. 2018. PredCNN: predictive learning with cascade convolutions//Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm, Sweden: ijcai. org: 2940-2947 [DOI: 10.24963/IJCAI.2018/408]

Ye X and Bilodeau G A. 2023. Video prediction by efficient transformers. Image and Vision Computing, 130: #104612 [DOI: 10.1016/j.imavis.2022.104612]

Zheng X Y, Fang W and Tao E Y. 2024. Review of deep learning in satellite cloud image movement prediction. Remote Sensing Information, 39(6): 1-11 (郑行钰, 方巍, 陶恩屹. 2024. 深度学习在卫星云图移动预测中的研究综述. 遥感信息, 39(6): 1-11)[DOI: 10.20091/j.cnki.1000-3177.2024.06.001]

作者简介

贺琪,女,教授,主要研究方向为海洋大数据存储和云计算。
E-mail:qihe@shou.edu.cn
郝增周,通信作者,男,研究员,主要研究方向为海洋遥感。
E-mail:hzyx80@sio.org.cn
臧正源,男,硕士研究生,主要研究方向为遥感图像处理。
E-mail: zzy19991219@163.com