

基于上下文注意力的多帧臭氧总量数据重建方法

赵伯霖¹, 黄飞虎¹, 胡金蓉¹, 张艳²

(1. 成都信息工程大学 计算机学院, 四川 成都 610225; 2. 国家卫星气象中心研究所, 北京 100081)

摘要: 大气臭氧层在地外环境系统中扮演着至关重要的角色, 全面的臭氧数据对于科研人员分析臭氧层状态、预测其对全球气候变化的影响十分重要。由于卫星自身传感器故障或轨道扫描宽度限制, 数据采集过程普遍存在信息缺失现象。为此, 设计一种基于上下文注意力的多帧臭氧总量数据重建方法。首先, 构建基于多头功能的三维可变形残差编码器, 以对远距离依赖关系进行编码; 其次, 使用多层次上下文注意力模块连接编码器与解码器, 用于提高对复杂数据结构的建模能力; 最后, 引入卷积选择性聚合网络作为解码器, 为模型提供局部性信息补充。实验结果表明, 该模型在风云三号紫外臭氧总量探测仪臭氧总量产品数据集上的 PSNR 和 SSIM 两项客观评价指标均优于对照方法。所提模型为臭氧数据重建提供了一种有效的解决方案, 对于大气科学研究与臭氧层保护策略制定具有重要科学价值。

关键词: 臭氧质控; 深度学习; 上下文注意力; 可变形卷积

DOI: 10.11907/rjdk.241353

开放科学(资源服务)标识码(OSID):



中图分类号: TP391

文献标识码: A

文章编号: 1672-7800(2025)006-0032-09

Multi-frame Total Ozone Data Reconstruction Method Based on Contextual Attention

ZHAO Bolin¹, HUANG Feihu¹, HU Jinrong¹, ZHANG Yan²

(1. School of Computer Science, Chengdu University of Information Technology, Chengdu 610225, China;

2. National Satellite Meteorological Center, Beijing 100081, China)

Abstract: The atmospheric ozone layer plays a crucial role in the extraterrestrial environmental system, and comprehensive ozone data is essential for researchers to analyze the state of the ozone layer and predict its impact on global climate change. Due to sensor failures or limited orbit scanning width of the satellite itself, there is a common phenomenon of information loss in the data acquisition process. To this end, design a multi frame ozone total data reconstruction method based on contextual attention. Firstly, construct a 3D deformable residual encoder based on multi head functionality to encode long-range dependencies; Secondly, a multi-level contextual attention module is used to connect the encoder and decoder, in order to improve the modeling ability of complex data structures; Finally, a convolutional selective aggregation network is introduced as a decoder to provide local information supplementation for the model. The experimental results show that the model outperforms the control method in both PSNR and SSIM objective evaluation indicators on the ozone total product dataset of the Fengyun-3 ultraviolet ozone total detector. The proposed model provides an effective solution for ozone data reconstruction and has important scientific value for atmospheric science research and the development of ozone layer protection strategies.

Key Words: ozone quality control; deep learning; contextual attention; deformable convolution

收稿日期: 2024-04-10

扫描二维码阅读全文:



基金项目: 四川省科技计划项目(2023YFQ0072)

作者简介: 赵伯霖(1997-), 男, 成都信息工程大学计算机学院硕士研究生, 研究方向为计算机视觉; 黄飞虎(1986-), 男, 博士, 成都信息工程大学计算机学院讲师, 研究方向为时空大数据分析; 胡金蓉(1979-), 女, 博士, 成都信息工程大学计算机学院副教授、硕士生导师, 研究方向为人工智能、数字图像处理、气象多源多模态数据融合分析技术理论与应用; 张艳(1977-), 女, 博士, 国家卫星气象中心研究所高级工程师、硕士生导师, 研究方向为卫星质控。本文通讯作者: 黄飞虎。

0 引言

臭氧作为地球大气层中的一种微量气体,能够有效阻挡太阳光中强烈的紫外线辐射,在维护空气质量和生态系统的健康稳定方面发挥着至关重要的作用^[1-2]。深入探究臭氧在全球范围内的分布状况及变化规律能够加深对全球气候变化现象的理解。因此,构建一套长期且完善的臭氧气候数据集尤为重要。由于传感器故障和轨道扫描宽度限制等因素,卫星采集的臭氧总量数据普遍存在缺失现象,设计能有效重建卫星臭氧总量数据中缺失部分的算法成为一个亟待解决的问题。

1 相关工作

1.1 基于统计学的方法

以往,大多数研究采用基于统计学的方法来重建缺失臭氧数据,如均值法和2D线性插值技术等^[3-4],也有少数研究采用反距离加权法和克里金插值法等^[5-6]。然而,均值法忽略了时间单位尺度内的动态变化特征,而2D线性插值技术在数据缺失较多的情况下精度不足。总体来说,基于统计学的数据重建方法通常需要大量先验知识作为支撑,且受限于有限的特征表示和固定的算法结构,重建结果在准确性和空间一致性上存在不足。

1.2 深度学习方法

随着深度学习技术的发展,基于深度神经网络的方法凭借其强大的表征能力对臭氧数据进行像素特征提取,逐渐取代了传统的基于统计学的方法。根据重建信息来源不同,基于深度神经网络的臭氧数据重建方法分为基于空间分布特征信息重建方法和联合时空特征信息重建方法两类^[7-8]。

基于空间分布特征信息重建方法利用观测数据中的空间关联性推断未知位置的空间特征分布,以恢复原始对象或场景的空间分布情况。例如,Pathak等^[9]率先提出基于卷积神经网络的上下文编码器(Context Encoders),从而学习到更丰富的图像特征,提升模型对图像的语义理解能力。然而,该模型对不规则掩膜的泛化能力有限,限制了其在处理复杂数据缺失问题时的应用。为解决这一问题,Liu等^[10]采用部分卷积处理不规则掩膜,搭配掩码更新机制自动生成更合理的掩膜预测。然而,部分卷积掩膜共享导致修复网络的学习能力受限,在较为复杂的图像修复任务处理中仍存在内容错乱等问题。针对网络学习能力较弱的问题,Zeng等^[11]将生成式对抗网络与Transformer^[12]模块结合,提出AOT(Aggregated Contextual Transformations)模型,通过聚合上下文信息的变换来填补图像中的缺失部分,显著提升了模型的学习能力。虽然AOT具有

较强的局部预测复杂全局特征像素的能力,但在处理大范围缺失区域时重建效果较差。Zhao等^[13]提出Co-Modulated生成对抗网络,通过将条件输入与风格表征进行协同调制,结合条件和无条件图像生成实现对大范围缺失图像的补全,但该网络对于细节特征的修复效果不够理想。为兼顾全局结构和纹理细节,Zheng等^[14]提出一种将全局码调制与空间码调制级联的CM-GAN(Cascaded Modulation Generative Adversarial Networks),该网络在处理细节特征的同时能够更好地将全局上下文注入空间域内,以达到对细节特征的精细化修复。基于空间分布特征信息重建方法在数据大面积缺失或具有复杂空间结构的情况下,由于没有加入可靠的相似时空特征辅助数据,只能依赖于自身空间域上的有效信息,重建结果无法满足高精度需求。

联合时空特征信息重建方法是一种综合利用时序和空间数据的技术,旨在将时序与空间信息结合起来,以提高数据信息量。例如,Zeng等^[15]首次将Transformer应用于视频修复领域中,提出一种联合时空变换网络(Spatial-Temporal Transformer Network, STTN)。该网络将相邻帧和远程帧作为输入,使用多头注意力机制计算不同尺度下空间特征块的相似性,从而填充所有输入帧数据中的缺失区域。然而,由于只计算空间块中的注意力,网络难以对快速运动的短时连续性进行建模。为提高时间一致性,Liu等^[16]提出解耦时空变换模型,将时间解耦模块与空间解耦模块解耦合,两个模块以交替的方式堆叠,有利于时空信息的交互并提升修复效果。针对边缘细节纹理捕捉缺陷问题,Liu等^[17]提出FuseFormer模型,通过新颖的软分割和软合成操作细粒度特征融合,在图像或视频处理中能灵活处理非局部的依赖关系。考虑到光流引导具有较好的时间一致性以及Transformer具有全局建模能力的优点,Zhang等^[18]提出一种光流引导的Transformer框架FGT(Flow-Guided Transformer),利用光流传播外观的运动变化特征来指导Transformer中注意力的提取。随后,Zhang等^[19]进一步改进模型,提出FGT++,使用光流来引导自注意力机制,从而缓解特征退化。然而,FGT++在处理具有大范围缺失或时空分布特征异质性显著的区域时,由于光流估计的精度容易受缺失区域大小和运动范围的影响,会使重建精度难以得到保证。

为改善上述方法存在的问题,本文提出一种基于上下文注意力的多帧臭氧总量数据重建方法,主要贡献点为:①利用多头机制实现远距离像素级依赖关系的特征聚合,同时通过运动自适应补偿机制确保重建数据的时空一致性;②设计具有多层次结构的上下文注意力模块,通过补充全局信息并重新加权像素获取时空上下文信息,从而提高模型的建模能力;③引入卷积选择性聚合(Convolution Selective Aggregation, CSA)模块,通过自适应地聚合多尺度特征提高模型的表征能力和泛化能力。

2 模型构建

臭氧总量数据重建模型采用编码器—解码器结构,包括编码器、多层次上下文注意力(Multi-Hierarchical Contextual Attention, MHCA)和解码器3个模块。整体网络结构如图1所示。

模型的输入为4帧连续的臭氧总量数据。首先,编码器模块通过6个自适应时空聚合(Adaptive Spatiotemporal

Aggregation, ASA)模块完成自适应运动补偿与特征提取。该模块在特征提取过程中学习额外的偏移量,能够更灵活地提取臭氧数据中的有效信息;其次,解码器模块对6个ASA模块进行特征解码,融合编码器的多层特征,有助于保留更多臭氧结构和细节信息。此外,为了扩大感受野并提高特征利用率,为每一层的特征解码过程设计一种MHCA模块。该模块对编码器每层的特征进行融合后将多次进入解码器模块,从而加速臭氧数据的重建过程。以下分别对3个模块的设计与实现进行详细阐述。

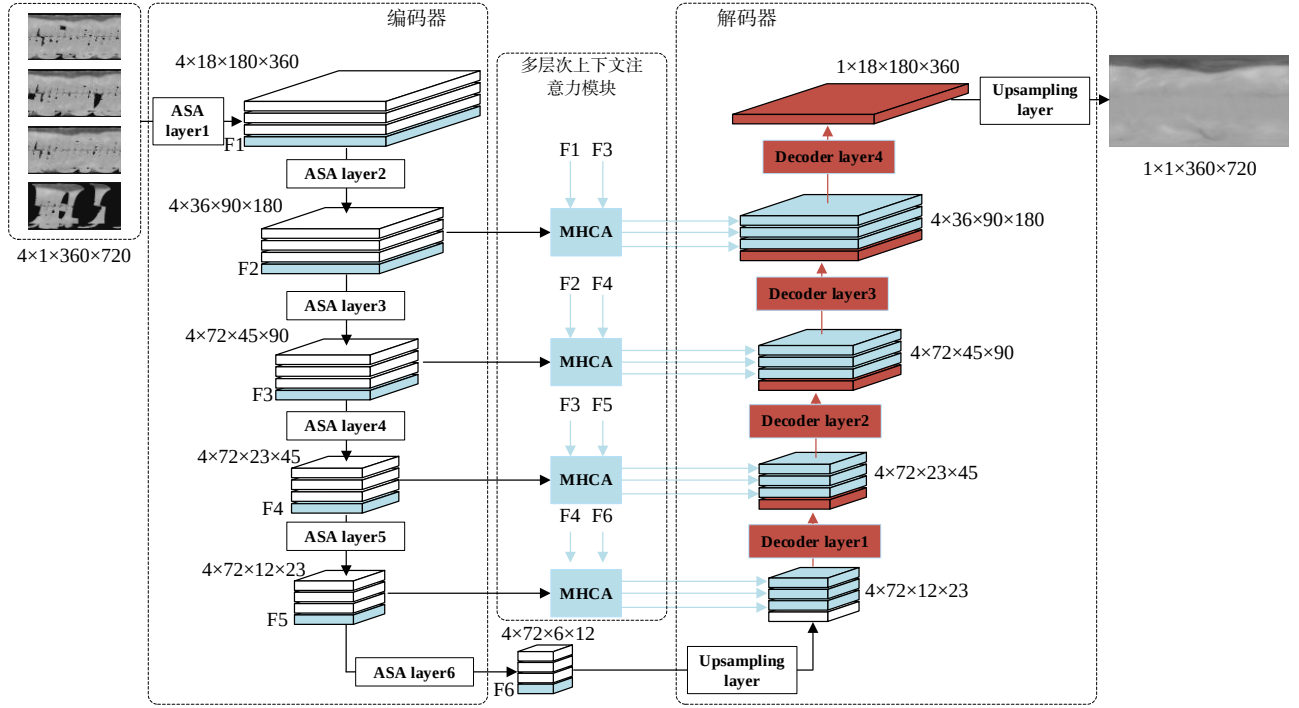


Fig. 1 Overall network structure

图1 整体网络结构

2.1 编码器模块

编码器由6个ASA模块线性组成。ASA模块由以多组可变形三维(Multi-group Deformable 3D, MD3D)卷积为核心的残差模块和深度可分离前馈模块(Depthwise Feed-forward Module, DWF)两部分组成。具体结构如图2所示。

ASA模块通过调控MD3D卷积的参数来控制输出特征图的大小,其余模块的参数均保持不变。具体参数设置如表1所示。

2.1.1 MD3D卷积

MD3D卷积是在可变形三维卷积(Deformable 3D Convolution, D3D)的基础上创新性地引入权重系数和分组的概念而来^[20]。结构如图3所示。为解决模块在聚合特征时会引入一些无关区域的问题,本文设计权重系数 m ,其取值范围为 $[0, 1]$ 。权重系数能够通过与区域特征向量的相乘操作有效区分出相关区域与不相关区域,进而确保特征提取的准确性和效率。为增加模块的自适应空间聚合能力,将时空聚合过程划分为 G 组,每个组都有单独的

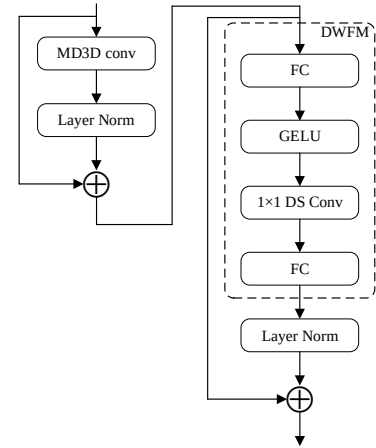


Fig. 2 ASA module structure

图2 ASA模块结构

采样偏移和权重系数,因此单个卷积层上的不同组可以具有不同的空间聚合模式,从而有效地从不同位置的不同表示子空间学习更丰富的信息。

定义第 k 层模块的特征输入为 $Feat_{in} \in \mathbb{R}^{C_k \times T_k \times W_k \times H_k}$,

Table 1 Specific parameter configuration of the encoder

表 1 编码器具体参数配置

模块名称	输出特征图大小	MD3D 卷积配置
ASA layer1	[4, 18, 180, 360]	kernel size=7×7, stride=2, padding=3
ASA layer2	[4, 36, 90, 180]	kernel size=5×5, stride=2, padding=2
ASA layer 3	[4, 72, 45, 90]	kernel size=5×5, stride=2, padding=2
ASA layer 4	[4, 72, 23, 45]	kernel size=3×3, stride=2, padding=1
ASA layer 5	[4, 72, 12, 23]	kernel size=3×3, stride=2, padding=1
ASA layer 6	[4, 72, 6, 12]	kernel size=3×3, stride=2, padding=1

其中 C_k 、 T_k 、 W_k 、 H_k 分别为第 k 层特征的通道数、时间步数、宽和高。首先,使用线性层,按照通道维度将第 k 层模块的特征输入 $Feat_{in}$ 划分成 G 组,得到每组的输入特征图 $x_g \in \mathbb{R}^{(C_k/G) \times T_k \times W_k \times H_k}$;其次,通过效率更高的深度可分离卷积得到每个组的采样偏移量和权重系数,其中采样偏移量指卷积核的方向向量,可使卷积核的形态更贴近特征物;再次,将每组的采样偏移量、权重系数和分割后的特征输入作为 D3D 卷积的输入;最后,将每组的输出按位相加,得到模块输出 $Feat_{out} \in \mathbb{R}^{C_{k+1} \times T_{k+1} \times (W_k/2) \times (H_k/2)}$,其中 C_{k+1} 、 T_{k+1} 分别表示第 $k+1$ 层特征的通道数和时间步数。经过上述改进,得到 MD3D 卷积的表达式为:

$$y(p_0) = \sum_{g=1}^G \sum_{n=1}^N \omega_g m_{gn} x_g (P_0 + P_n + \Delta P_{gn}) \quad (1)$$

式中: G 表示聚合组的总数。对于第 g 个组, ω_g 表示每组内的共享投影权重; N 表示采样点总数,为卷积核尺寸的平方; m_{gn} 表示第 g 组中第 n 个采样点的权重系数; x_g 表示第 g 组的输入特征图; ΔP_{gn} 表示第 g 组中网格采样位置 P_n 相对应的偏移量。在模块的最后加入层归一化 (Layer Normalization, LN) 操作,可以减少内部协变量偏移并加速模型训练。

2.1.2 DWFM

为进一步增强特征提取模块的表征能力,在 MD3D 后

引入 DWFM。首先,利用全连接层对输入特征进行精细的提取与映射,将原始向量空间高效映射到更高维度的隐藏表示空间,进而提升模型的表征深度和泛化能力;其次,引入 GELU 激活函数对特征进行非线性变换,以提高训练稳定性;再次,采用深度可分离卷积聚合局部特征,通过轻量级卷积深入挖掘输入序列中复杂的特征表示和相互关系^[21];最后,再次利用全连接层将隐藏表示空间映射到输出向量空间,并借助 LN 层加快模型的训练速度。模块中涉及的计算公式为:

$$Feat_{out} = \mathcal{F}_{FC}(\mathcal{F}_{DWConv}(\mathcal{F}_{GELU}(\mathcal{F}_{FC}(Feat_{in})))) + Feat_{in} \quad (2)$$

式中: $Feat_{in} \in \mathbb{R}^{C \times W \times H}$ 表示输入特征图; $Feat_{out} \in \mathbb{R}^{C \times W \times H}$ 表示输出特征图; $\mathcal{F}_{FC}(\cdot)$ 表示全连接层操作; $\mathcal{F}_{GELU}(\cdot)$ 表示 GELU 激活函数; $\mathcal{F}_{DWConv}(\cdot)$ 表示深度可分离卷积。

2.2 MHCA 模块

为弥补 ASA 模块在轴向时空聚合中可能引入的信息损失,本文引入 MHCA 模块,通过一种多输入机制获取自相邻编码级的上下文信息,提高了模型对上下文信息的敏感度,从而有针对性地支持臭氧数据的重建任务。该模块结构如图 4 所示。

MHCA 模块包含 3 个输入和 3 个输出。输入分别定义为 $X \in \mathbb{R}^{C_X \times T_X \times H_X \times W_X}$ 、 $L \in \mathbb{R}^{C_L \times H_L \times W_L}$ 以及 $H \in \mathbb{R}^{C_H \times H_H \times W_H}$,其中 $C_L = 2C_X = 4C_H$ 、 $H_L = 2H_X = 4H_H$ 、 $W_L = 2W_X = 4W_H$ 。首先,通过 $1 \times 1 \times 1$ 卷积对当前层次的特征图 X 进行降维,以获得大小为 $H_X \times W_X \times C_X$ 的输出特征图 $\tilde{X}$;其次,通过 3×3 卷积层对低维特征图 L 进行下采样,获得大小为 $H_X \times W_X \times C_X$ 的输出特征图 $\tilde{L}$;再次,通过 3×3 反卷积对高维特征图 H 进行上采样操作,获得大小为 $H_X \times W_X \times C_X$ 的输出特征图 $\tilde{H}$;接着,将特征图 $\tilde{X}$ 依次通过 ReLU 激活函数、 $1 \times 1 \times 1$ 的逐通道卷积和 sigmoid 激活函数,得到大小为 $H_X \times$

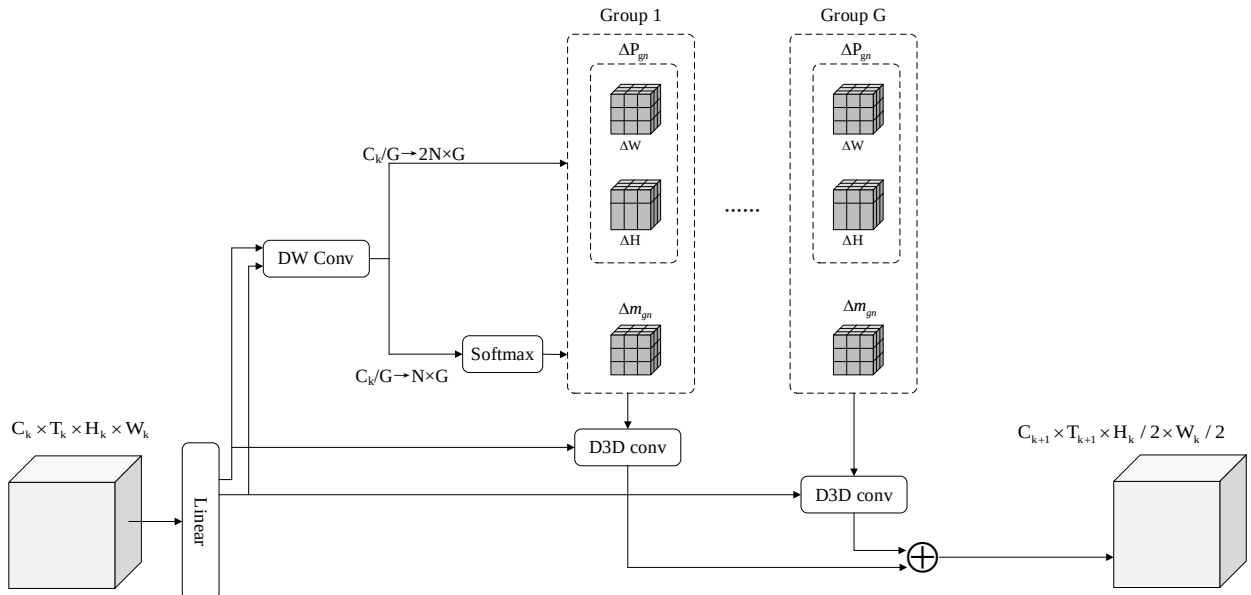


Fig. 3 MD3D convolution structure

图 3 MD3D 卷积结构

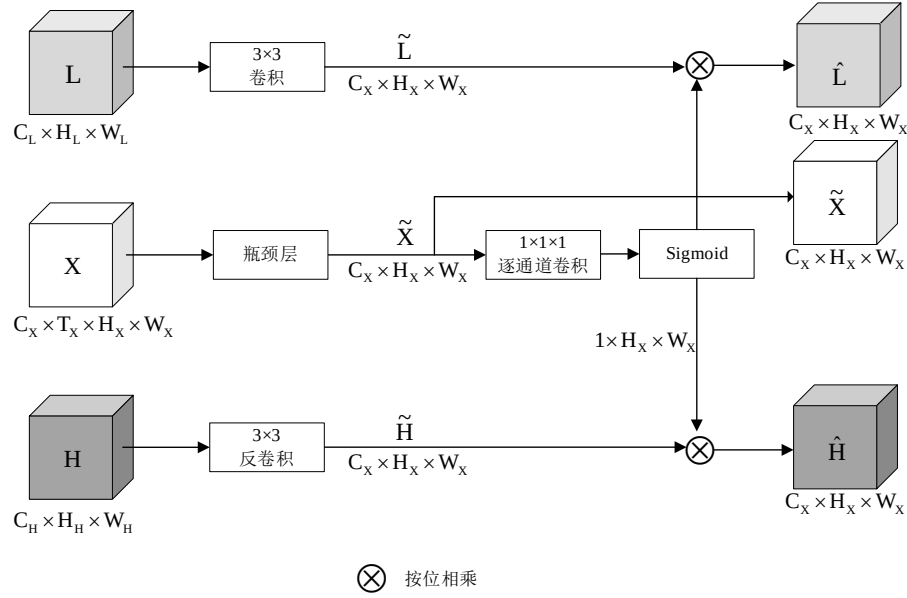


Fig. 4 MHCA module structure

图4 MHCA模块结构

$W_X \times 1$ 的注意力加权图 β ; 最后, 将注意力加权图依次逐通道乘以低维特征图 $\tilde{L}$ 和高维特征图 $\tilde{H}$, 使得 $\tilde{X}$ 的每个信道从相邻特征层中学习注意力信息, 得到低维特征图 $\hat{L}$ 和高维特征图 $\hat{H}$ 。该步骤的计算公式为:

$$\hat{L} = \tilde{L} * \beta \quad (3)$$

$$\hat{H} = \tilde{H} * \beta \quad (4)$$

式中: $*$ 表示广播相乘。将低维特征图 $\hat{L}$ 、高维特征图 $\hat{H}$ 和降维后的当前层次特征图 $\tilde{X}$ 作为模型输出。

2.3 解码器模块

解码器由 4 个 CSA 模块和 2 个基于亚像素卷积的上采样模块组成^[22]。具体参数配置如表 2 所示。

Table 2 Specific parameter configuration of the decoder

表2 解码器具体参数配置

模块名称	输出特征图大小	配置
Upsampling layer	[1, 72, 12, 23]	1×1 Conv2d, sub-pixel conv
Decoder layer1	[1, 72, 23, 45]	1×1 Conv2d, sub-pixel conv, CSAM
Decoder layer2	[1, 72, 45, 90]	1×1 Conv2d, sub-pixel conv, CSAM
Decoder layer3	[1, 36, 90, 180]	1×1 Conv2d, sub-pixel conv, CSAM
Decoder layer4	[1, 18, 180, 360]	1×1 Conv2d, sub-pixel conv, CSAM
Upsampling layer	[1, 1, 360, 720]	1×1 Conv2d, sub-pixel conv

2.3.1 CSA 模块

CSA 模块融合多尺度特征, 使网络能够更有效地保留图像的轮廓、边缘等细节特征, 并采用残差结构提高网络稳定性。具体结构如图 5 所示。

首先, 模型输入为 4 个相同大小的特征图 $Feat_{in} \in \mathbb{R}^{H \times W \times C}$, 对所有输入特征矩阵按元素求和, 融合特征信息, 以帮助模型自适应地调节感受野。计算公式为:

$$U = \tilde{X} + \hat{L} + \hat{H} + D_k \quad (5)$$

其次, 为提高模型效率并降低复杂性, 模块通过 4 组

卷积分类器将融合后得到的特征 U 压缩成 $z_a \in \mathbb{R}^{C_k \times 1 \times 1}$ 、 $z_b \in \mathbb{R}^{C_k \times 1 \times 1}$ 、 $z_c \in \mathbb{R}^{C_k \times 1 \times 1}$ 、 $z_d \in \mathbb{R}^{C_k \times 1 \times 1}$ 4 个特征序列, 其中卷积分类器在进行特征提取与分类时能够保留输入特征图的空间信息。与全连接层相比, 卷积分类器更能捕捉特征图中的空间关系和上下文信息, 从而更准确地选择合适的通道。

再次, 模型利用训练后的 4 个特征序列 z_a, z_b, z_c, z_d , 沿通道维度进行 Softmax 激活函数操作, 从而建立每个输入的通道注意力图 a, b, c, d 。表示为:

$$a = \frac{e^{z_a}}{e^{z_a} + e^{z_b} + e^{z_c} + e^{z_d}} \quad (6)$$

$$b = \frac{e^{z_b}}{e^{z_a} + e^{z_b} + e^{z_c} + e^{z_d}} \quad (7)$$

$$c = \frac{e^{z_c}}{e^{z_a} + e^{z_b} + e^{z_c} + e^{z_d}} \quad (8)$$

$$d = \frac{e^{z_d}}{e^{z_a} + e^{z_b} + e^{z_c} + e^{z_d}} \quad (9)$$

最后, 每个输入与其相应的通道注意力序列相乘, 进而自适应地选择图像的不同尺度信息。使用 1×1 卷积将各层次特征图降至 $1/4$ 通道数, 在通道维度上拼接特征图, 进行特征聚合。

2.3.2 上采样模块

上采样模块接受 CSA 模块输出的低维特征表示作为输入, 通过上采样操作将低维特征映射回原始数据的空间, 旨在重建出与原始输入尽可能相似的数据, 从而实现臭氧数据的有效重构。为提高臭氧数据的精度和细节保留能力, 采用残差亚像素卷积层进行上采样操作, 其结构如图 6 所示。

首先, 利用 1×1 卷积对输入特征进行跨通道交互, 从而提高网络的表达能力; 其次, 使用亚像素卷积对特征进行上采样, 以保留更多细节信息, 同时引入 Relu6 激活函数

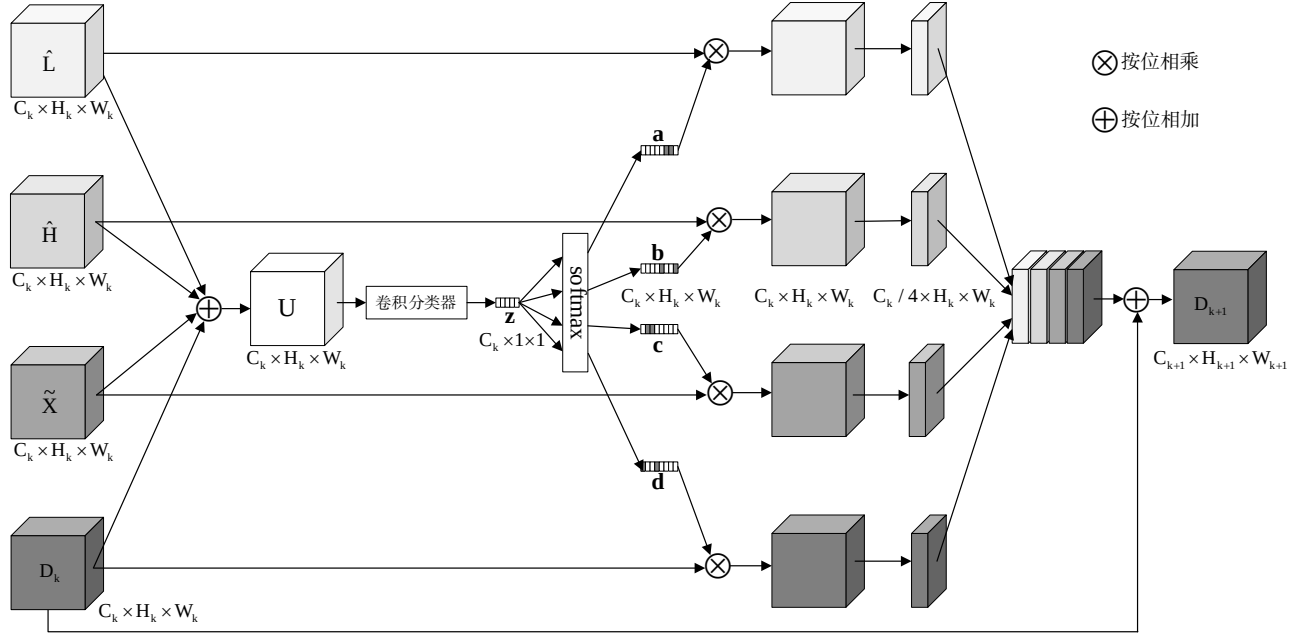


Fig. 5 CSA module structure

图5 CSA 模块结构

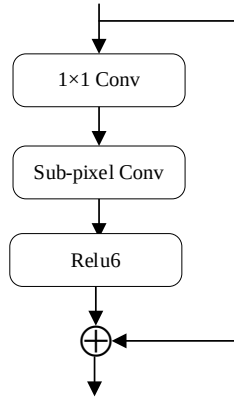


Fig. 6 Upsampling module structure

图6 上采样模块结构

对特征进行非线性变换,以提高训练稳定性;最后,将输出特征与输入特征相加,以防止梯度爆炸。

3 实验方法与结果分析

3.1 实验环境与参数设置

使用 PyTorch 框架实现模型并在 NVIDIA GeForce 2080Ti 上进行训练。输入数据为连续的 4 帧 TOU 臭氧总量数据,输入数据尺寸为 $4 \times 1 \times 360 \times 720$, batch size 设置为 24, 迭代 2 000 个 epoch, 并使用学习率为 10^{-4} 的 ADAM 优化器对训练过程中的参数进行优化。其他参数为 $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 10^{-8}$ 。

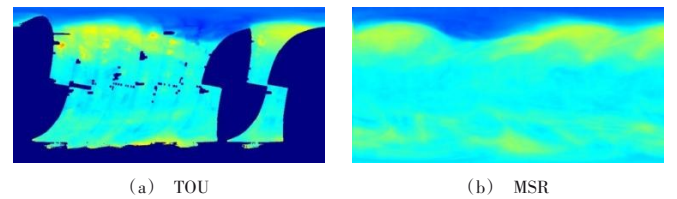
3.2 数据集

风云三号紫外臭氧总量探测仪 (Total Ozone Unit, TOU) 臭氧总量产品数据集是基于 TOU 观测的反演结果得到的数据集,数据分辨率为 360×720 , 共计 3 600 帧数

据^[23]。本文将 TOU 臭氧总量产品数据集作为模型的输入。

多传感器再分析 (Multi Sensor Reanalysis, MSR) 臭氧总量数据集是利用多个卫星传感器观测数据进行臭氧总量的再分析,进而生成的连续、一致且高质量的臭氧总量数据集,分辨率为 360×720 ^[24]。选取与 TOU 数据集中相同天数的数据,共计 3 600 帧,作为模型训练的目标。二者比较见图 7。

为增加数据集的多样性、降低过拟合程度,需要对数据进行预处理。本文采用 Z-score 标准化方法对 TOU 数据集和 MSR 数据集进行标准化处理。将 TOU 数据集与 MSR 数据集按照天数对应,组成 3 600 组连续臭氧总量数据,并按照 8:2 的比例分为训练集和测试集。



(a) TOU

(b) MSR

Fig. 7 Dataset comparison

图7 数据集比较

3.3 损失函数

在构造损失函数时需考虑到图像数据重建结果的全局一致性和细节合理性。本文将感知损失、总变分损失、缺失损失与背景损失组合起来对网络模型进行训练。

感知损失 $L_{perceptual}$ 通过比较生成图像与真实图像在感知空间中的相似性来量化生成图像数据的质量^[25]。计算公式为:

$$L_{\text{perceptual}} = \sum_i \frac{1}{C_i H_i W_i} \left\| \phi_i(I_{\text{out}}) - \phi_i(I_{\text{gt}}) \right\|_1 \quad (10)$$

式中: ϕ_i 表示加载了预训练参数的 VGG 特征提取网络的第 i 层^[26]; I_{out} 表示生成图像数据; I_{gt} 表示真实图像数据; C_i 、 H_i 、 W_i 分别表示经过 i 层特征提取操作后所获得特征图的通道数、高和宽。

总变分损失 L_{tv} 用于增强生成图像数据的空间平滑度, 消除边界伪影问题^[27]。计算公式为:

$$L_{\text{tv}} = \sum_{H, W} \left(\left\| d_{h, w} - d_{h+1, w} \right\|_1 + \left\| d_{h, w} - d_{h, w+1} \right\|_1 \right) \quad (11)$$

式中: $d_{h, w}$ 表示位置 (h, w) 处的数据值。

对于缺失损失 L_{hole} , 计算 I_{gt} 与 I_{out} 中缺失区域的 L1 距离; 对于背景损失 L_{valid} , 计算 I_{gt} 与 I_{out} 中背景区域的 L1 距离。计算公式分别为:

$$L_{\text{hole}} = \left\| (1 - M) \odot (I_{\text{out}} - I_{\text{gt}}) \right\|_1 \quad (12)$$

$$L_{\text{valid}} = \left\| M \odot (I_{\text{out}} - I_{\text{gt}}) \right\|_1 \quad (13)$$

式中: M 为初始的 mask, 其值为 1 表示未损坏区域, 值为 0 表示损坏区域。

综上所述, 本文算法的总损失函数定义为:

$$L_{\text{total}} = \lambda_{\text{perceptual}} L_{\text{perceptual}} + \lambda_{\text{tv}} L_{\text{tv}} + \lambda_{\text{hole}} L_{\text{hole}} + \lambda_{\text{valid}} L_{\text{valid}} \quad (14)$$

式中: λ_{tv} 、 λ_{hole} 、 λ_{valid} 、 $\lambda_{\text{perceptual}}$ 表示每个损失函数对应的系数。模型所涉及的损失函数乘以对应的系数后再相加, 得到最终损失函数。

3.4 评价指标

采用 L1 Loss、结构相似性 (Structural Similarity Index Measure, SSIM) 和峰值信噪比 (Peak Signal-to-Noise Ratio, PSNR) 作为评价输出数据的指标^[28-29]。

SSIM 是一种用于评估两幅图像数据相似度的指标, 包含亮度、对比度和结构信息。通常来说, SSIM 值越接近 1, 表示图像数据越相似。其计算公式为:

$$SIM(I_{\text{out}}, I_{\text{gt}}) = \frac{(2\mu_{I_{\text{out}}}\mu_{I_{\text{gt}}} + C_1)(2\sigma_{I_{\text{out}}I_{\text{gt}}} + C_2)}{(\mu_{I_{\text{out}}}^2 + \mu_{I_{\text{gt}}}^2 + C_1)(\sigma_{I_{\text{out}}}^2 + \sigma_{I_{\text{gt}}}^2 + C_2)} \quad (15)$$

式中: $\sigma_{I_{\text{out}}}$ 、 $\mu_{I_{\text{out}}}$ 分别为输出图像数据 I_{out} 的方差和均值; $\sigma_{I_{\text{gt}}}$ 、 $\mu_{I_{\text{gt}}}$ 分别为真实图像数据 I_{gt} 的数据值方差和均值; $\sigma_{I_{\text{out}}I_{\text{gt}}}$ 为 I_{out} 与 I_{gt} 的协方差; C_1 和 C_2 为用于维持稳定的常数, 避免模型出现除数为 0 的情况。

PSNR 通过测量原始信号与经过处理后信号之间的差异来衡量信号质量的失真程度。其值越高, 表示信号质量越好, 噪声越小。计算公式为:

$$MSE = \frac{1}{HW} \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} [I_{\text{out}}(i, j) - I_{\text{gt}}(i, j)]^2 \quad (16)$$

$$PSNR = 10 \times \lg \left(\frac{MAX^2}{MSE} \right) \quad (17)$$

式中: $I_{\text{out}}(i, j)$ 、 $I_{\text{gt}}(i, j)$ 分别表示生成图像数据与真实图像数据所对应位置的数据值; MAX 表示单个图像点的最大

大值。

3.5 比较实验

在 TOU 臭氧总量产品数据集上采用 PSNR、SSIM、L1 Loss 作为评判臭氧数据重建水平的指标, 进行了定量分析。参与比较的模型包括单帧补全模型 AOT^[7]、Co-Mod^[8]、CM-GAN^[10] 以及多帧补全模型 FuseFormer^[13]、FGT^[14]、FGT++^[15]。实验结果如表 3 所示, 最优结果使用粗体标注。可以看出, 本文方法的各项指标相较其他比较模型均有一定优势, PSNR 和 SSIM 分别达到 32.55 dB 和 0.965。其中, 本文模型的 PSNR 和 SSIM 相较主流图像补全模型 CM-GAN 分别高出 0.82 dB 和 0.018, L1 Loss 降低了 0.94。尽管 CM-GAN 采用全局调制与空间调制级联的结构, 并利用傅里叶卷积来整合长程特征依赖, 从而扩大了有效感受野, 使重建结果更加逼真和准确, 但是傅里叶变换在生成复杂干净的纹理方面表现不佳, 且该模型对时域信息的利用不足, 导致其表现不如本文模型。本文模型引入 CSA 模块以及多尺度融合机制, 增加了对细节信息的敏感度, 有助于生成细节纹理。此外, 本文模型的 PSNR 和 SSIM 相较主流视频补全模型 FGT++ 分别提高了 0.28 dB 和 0.05, L1 Loss 降低了 0.19。虽然 FGT++ 引入了光流引导特征集成和特征传播模块以提高重建质量, 但对于变化量较大的视频以及大面积缺失的情况仍有不足, 导致模型效果不理想。而本文模型设计了 MD3D 卷积, 增加了时空聚合能力和远距离建模能力, 具有较强的鲁棒性。

Table 3 Comparison of indicators of various models on the TOU total ozone product dataset

模型	PSNR ↑	SSIM ↑	L1 Loss ↓
AOT	28.98	0.927	5.37
Co-Mod	31.69	0.948	4.35
CM-GAN	31.73	0.947	4.32
FuseFormer	31.09	0.947	4.49
FGT	32.03	0.953	3.94
FGT++	32.27	0.960	3.57
本文模型	32.55	0.965	3.38

为进一步验证本文模型在重建臭氧数据方面的性能, 进行定性实验比较。图 8 为本文模型与比较模型的重建性能比较。可以看出, AOT、Co-Mod 和 CM-GAN 等单帧模型由于未能处理时域信息, 导致信息量不足, 模型重建的臭氧数据细节精度不高。FuseFormer 模型虽然能够处理时域信息, 但其捕捉局部纹理不准确的缺点无法精确地重建细节信息。相比之下, FGT 和 FGT++ 模型通过引入光流引导传播模块, 保持了良好的时间一致性, 在重建数据细节方面优于 FuseFormer 模型。然而, 光流引导传播也可能降低语义结构的连贯性, 导致过度平滑和模糊轮廓, 因此在重建精度上仍不及本文模型。本文模型不仅能有效补全缺失数据, 而且在细节丰富性方面优于其他比较模型, 重建效果与 MSR 数据集最为接近。

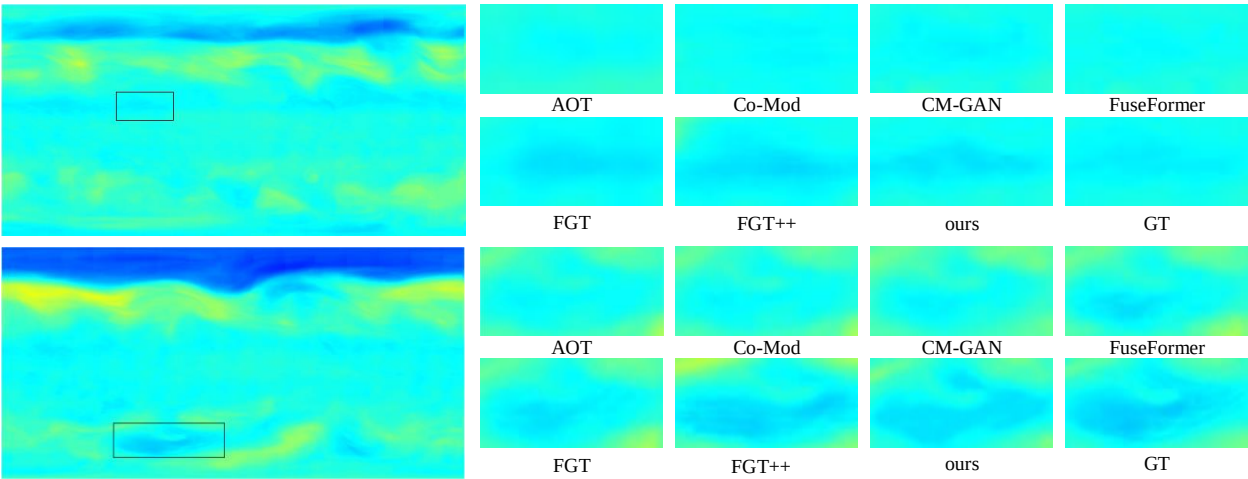


Fig. 8 Comparison of reconstruction performance of various models on TOU ozone total product dataset
图8 各模型在TOU臭氧总量产品数据集上的重建性能比较

3.6 消融实验

在TOU臭氧总量产品数据集上进行消融实验,有针对性地去除模型中的关键模块,并比较模块去除前后模型的PSNR值和L1 Loss值,以验证各模块对模型性能的影响。首先,为检验ASA模块对模型性能的影响,将该模块中的D3D v2结构替换为D3D结构,并将其命名为-ASA模型;其次,为验证MHCA模型对模型性能的影响,去除MHCA模块,改为直接连接,将其命名为-MHCA模型;最后,为验证CSA模块的作用,剔除该模块,采用特征直接拼接的方式,将其命名为-CSANet模型。消融实验结果如表4所示。

Table 4 Results of ablation experiments
表4 消融实验结果

模型	PSNR ↑	L1 Loss ↓
-ASA	32.10	3.53
-MHCA	32.23	3.47
-CSANet	31.99	3.62
本文模型	32.55	3.38

可以看出,-ASA模型的PSNR值达到32.10,L1损失为3.53,相较本文模型的PSNR值下降了0.45 dB,表明ASA模块对网络性能有提升作用,多头分组机制和建立远距离像素之间的相关性有助于提升模型的补全能力,可以生成更加真实的图片数据。-MHCA模型的PSNR相较本文模型降低了0.32 dB,这是由于MHCA模块能够帮助模型充分结合上下层次的上下文信息,进而提高模型补全质量。-CSANet模型的PSNR相较于本文模型降低了0.56 dB,同时增加了0.24的L1损失,证实了CSAM模块可提高网络对不同尺度信息的感知能力,以更有效地捕获图像中的语义信息,从而帮助模型生成更为精细的重建数据。

4 结语

为提高臭氧数据质量,本文提出基于上下文注意力的多帧臭氧总量数据重建方法。首先,采用MD3D卷积增强

远距离特征提取能力;其次,采用CSA模块增强模型的鲁棒性和适应能力,并提升重建精度;最后,采用MHCA模块获取更多上下文信息,进一步提高重建数据的准确性。与其他主流模型比较实验验证了本文模型的优越性,同时消融实验证明了模型各模块的有效性。然而,当重建大范围的臭氧总量数据时,本文方法仍然存在一定的局限性。计划在后续工作中引入更为准确的地面监测站数据作为先验信息,并将其融入到多模态重建算法中,进一步提升重建数据的准确性。

参考文献:

[1] LI K, JACOB D J, SHEN L, et al. Increases in surface ozone pollution in China from 2013 to 2019: anthropogenic and meteorological influences [J]. Atmospheric Chemistry and Physics, 2020, 20(19): 11423–11433.

[2] ARCHIBALD A T, NEU J L, ELSHORBANY Y F, et al. Tropospheric ozone assessment report: a critical review of changes in the tropospheric ozone burden and budget from 1850 to 2100[J]. Elementa: Science of the Anthropocene, 2020, 8(1): 34.

[3] MALINOVIĆ M S, RADOVANOVIĆ M M, MIJATOVIĆ Z, et al. Reconstruction and variability of high daily erythral ultraviolet doses and relationship with total ozone, cloud cover, and albedo in Novi Sad (Serbia) [J]. International Journal of Climatology, 2022, 42(16): 9088–9100.

[4] HUANG G. Missing data filling method based on linear interpolation and lightgbm [J]. Journal of Physics: Conference Series, 2021, 1754(1): 012187.

[5] LIU Z N, YU X Y, JIA L F, et al. The influence of distance weight on the inverse distance weighted method for ore-grade estimation [J]. Scientific Reports, 2021, 11(1): 2689.

[6] BELKHIRI L, TIRI A, MOUNI L. Spatial distribution of the groundwater quality using kriging and Co-kriging interpolations [J]. Groundwater for Sustainable Development, 2020, 11: 100473.

[7] ZHANG X, ZHAI D, LI T, et al. Image inpainting based on deep learning: a review[J]. Information Fusion, 2023, 90: 74–94.

[8] QUAN W, CHEN J, LIU Y, et al. Deep learning-based image and video inpainting: a survey[J]. International Journal of Computer Vision, 2024, 132(7):2367–2400.

- [9] PATHAK D, KRAHENBUHL P, DONAHUE J, et al. Context encoders: feature learning by inpainting [C]//Las Vegas: 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016.
- [10] LIU G, REDA F A, SHIH K J, et al. Image inpainting for irregular holes using partial convolutions [C]//Proceedings of the European Conference on Computer Vision, 2018: 85–100.
- [11] ZENG Y, FU J, CHAO H, et al. Aggregated contextual transformations for high-resolution image inpainting [J]. IEEE Transactions on Visualization and Computer Graphics, 2022, 29(7): 3266–3280.
- [12] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [DB/OL]. <https://arxiv.org/abs/1706.03762>.
- [13] ZHAO S, CUI J, SHENG Y, et al. Large scale image completion via co-modulated generative adversarial networks [DB/OL]. <https://arxiv.org/abs/2103.10428>.
- [14] ZHENG H, LIN Z, LU J, et al. Image inpainting with cascaded modulation gan and object-aware training [C]//European Conference on Computer Vision, 2022: 277–296.
- [15] ZENG Y, FU J, CHAO H. Learning joint spatial-temporal transformations for video inpainting [C]//16th European Conference on Computer Vision, 2020: 528–543.
- [16] LIU R, DENG H, HUANG Y, et al. Decoupled spatial-temporal transformer for video inpainting [DB/OL]. <https://arxiv.org/abs/2104.06637>.
- [17] LIU R, DENG H, HUANG Y, et al. Fuseformer: fusing fine-grained information in transformers for video inpainting [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 14040–14049.
- [18] ZHANG K, FU J, LIU D. Flow-guided transformer for video inpainting [C]//European Conference on Computer Vision, 2022: 74–90.
- [19] ZHANG K, PENG J, FU J, et al. Exploiting optical flow guidance for transformer-based video inpainting [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(7): 4977–4992.
- [20] YING X, WANG L, WANG Y, et al. Deformable 3d convolution for video super-resolution [J]. IEEE Signal Processing Letters, 2020, 27: 1500–1504.
- [21] CAO J, LI Y, SUN M, et al. Do-conv: depthwise over-parameterized convolutional layer [J]. IEEE Transactions on Image Processing, 2022, 31: 3726–3736.
- [22] SUN Y, CHEN J, LIU Q, et al. Learning image compressed sensing with sub-pixel convolutional generative adversarial network [J]. Pattern Recognition, 2020, 98: 107051.
- [23] XIE M, GU M, HU Y, et al. A study on the retrieval of ozone profiles using FY-3D/HIRAS infrared hyperspectral data [J]. Remote Sensing, 2023, 15(4): 1009.
- [24] NAOE H, MATSUMOTO T, UENO K, et al. Bias correction of multi-sensor total column ozone satellite data for 1978–2017 [J]. Journal of the Meteorological Society of Japan. Ser. II, 2020, 98(2): 353–377.
- [25] LIU Y, CHEN H, CHEN Y, et al. Generic perceptual loss for modeling structured output dependencies [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 5424–5432.
- [26] SENGUPTA A, YE Y, WANG R, et al. Going deeper in spiking neural networks: VGG and residual architectures [J]. Frontiers in Neuroscience, 2019, 13: 95.
- [27] NGUYEN-TRUONG H, NGUYEN K N A, CAO S. Srgan with total variation loss in face super-resolution [C]//2020 7th NAFOSTED Conference on Information and Computer Science, 2020: 292–297.
- [28] BAKUROV I, BUZZELLI M, SCHETTINI R, et al. Structural similarity index (SSIM) revisited: a data-driven approach [J]. Expert Systems with Applications, 2022, 189: 116087.
- [29] SETIADI D R I M. PSNR vs SSIM: imperceptibility quality assessment for image steganography [J]. Multimedia Tools and Applications, 2021, 80(6): 8423–8444.

(责任编辑:尹晨茹)