

自注意力映射超时相遥感图像超分辨率重建

唐晓天^{1,2}, 杨雪^{1,3}, 李峰¹, 马骏², 梁亮⁴

1. 中国空间技术研究院钱学森空间技术实验室, 北京 100094;

2. 河南大学 软件学院, 开封 475004;

3. 南京大学 国际地球系统科学研究所, 南京 210023;

4. 清华大学 电子工程系, 北京 100084

摘要:针对视频卫星超时相数据高时间分辨率的特点, 单帧超分辨率重建算法无法充分利用其时空信息, 提出一种基于自注意力的超时相遥感图像超分辨率重建模型。对超时相序列帧先依据帧速率划分为多个时间组, 利用3D卷积对空间及时间同时建模的特点提取不同时间组下的时空信息, 之后通过改进的广泛自注意力残差块增加注意力计算范围并完成高动态映射, 在完成超时相序列帧建模的同时提取其丰富的空间细节信息, 最后将多个时间组的特征融合, 通过亚像素卷积提升分辨率并完成重建。所提算法的优点在于, 对超时相数据的多时间组处理充分利用了其丰富的时空信息, 改进的自注意力块可在无配准情况下完成序列帧的建模并提高空间细节信息的提取能力。在高分四号数据上的实验表明, 本文算法的主观视觉效果及客观评价指标均优于对比算法, 在高分四号数据两倍上采样重建时, 较双三次插值算法在PSNR值上提升了2.49 dB以上, 且在4倍重建时依然有很强的重建性能, 较双三次插值算法有巨大提升。实验结果表明, 该方法拥有较好的超分辨率重建效果, 有利于超时相数据在各领域的应用。

关键词: 遥感, 超时相数据, 超分辨率重建, 深度学习, 多时间组融合特征, 广泛自注意力, 高分四号

中图分类号: TP701/P2

引用格式: 唐晓天, 杨雪, 李峰, 马骏, 梁亮. 2024. 自注意力映射超时相遥感图像超分辨率重建. 遥感学报, 28(6): 1551-1559

Tang X T, Yang X, Li F, Ma J and Liang L. 2024. Super-resolution reconstruction of hypertemporal remote sensing images based on self-attention. National Remote Sensing Bulletin, 28(6): 1551-1559 [DOI: 10.11834/jrs.20221825]

1 引言

随着航空航天技术的迅速发展, 卫星, 无人机等设备被越来越多的应用到日常信息获取中。卫星的观测方式从传统的定时单一拍摄模式发展到现在多谱段, 多时相, 多模式相结合的观测特点, 获取的遥感数据范围广泛, 信息量大。近年随着中国遥感相机中面阵CMOS探测器的投入使用, 可完成超时相成像, 在重访周期1天以内的可称为超时相数据。其连拍的工作模式可以在短时间内(分钟级, 甚至秒级)获取多幅高重叠率的低分辨率图像(李峰等, 2021)。在实时天气预测、突发自然灾害、应急救援等多个领域有着十分重要的应用价值(聂娟等, 2018)。

超分辨率重建技术SR(Super-Resolution)可以从低分辨率图像中恢复出高分辨率图像。高分辨率遥感图像可以提供丰富的地物纹理细节信息, 在进行运动目标检测与识别方面具有很大的优势。传统的图像超分辨率重建技术包括基于插值的方法, 基于重构的方法和基于学习的方法。基于插值的方法主要有最邻近插值法, 双线性插值法等, 插值重建方法简单易实现, 但会导致重建结果边缘模糊, 细节信息丢失等。重构方法中如基于最大后验概率MAP和凸集投影法(许丽娜和何鲁晓, 2017)等, 这些算法迭代次数过多, 计算成本较大。基于学习的方法利用先验知识, 经过训练样本集获得LR与HR图像间的映射关系进行图像的重建(贺智和贺丹, 2020; 杨雪等, 2022)。

收稿日期: 2022-01-05; 预印本: 2022-07-19

基金项目: 国家重点研发计划(编号: 2020YFA0714100); 高分辨率对地观测系统重大专项(编号: 30-Y30F06-9003-20/22)

第一作者简介: 唐晓天, 研究方向为多时相超分辨率重建。E-mail: 631719950@qq.com

通信作者简介: 杨雪, 研究方向为超时相数据处理。E-mail: yangxue@qxslab.cn

随着深度学习方法在遥感图像超分辨率重建中的广泛应用。Haut等(2019)将视觉注意机制集成到残差网络中,学习遥感图像的高频信息;Dong等(2021)提出一种密集采样超分辨率网络来探索遥感图像的大尺度SR重建;还有研究者提出基于生成对抗网络的边缘增强网络(EEGAN),用于鲁棒卫星图像SR重建以及对噪声不敏感的对抗学习(Jiang等,2019)。多帧超分辨率重建算法中,有使用门控再流单元模块作为融合模块进行重建(Arefin等,2020);通过设计端到端网络,从多张未配准的低分辨率图像中恢复出高分辨率图像(Molin等,2020)。一些深度学习视频超分辨率重建算法对于多帧遥感图像重建也具有很高的参考价值,如利用估算光流实现序列帧对齐与超分辨率重建(Caballero等,2017;Wang等,2020);结合时间可变形卷积进行特征对齐进行超分重建(Tian等,2020);将3D卷积与可变形卷积结合对序列帧影像进行超分重建(Ying等,2020)。

单帧视频卫星超分辨率重建算法仅利用视频帧自身的信息进行超分辨率重建,重建效果有限。本文结合超时相数据特点。提出基于自注意力的

超分辨率重建模型,该模型通过端到端的网络可直接挖掘低分辨率中的高频信息,从多帧低分辨率图像中恢复出高分辨率图像,在高分四号数据集中进行实验验证,取得了较好的重建结果。

2 本文算法

2.1 整体网络结构

本文提出的多时间组自注意力网络模型如图1(a)所示。该模型以7帧序列帧作为输入,每帧图像首先经过一层 3×3 的卷积提取空间特征,然后根据序列帧速率划分为3个时间组,各个时间组通过时空特征融合模块(图1(b))提取该时组下的时空特征,之后送入广泛自注意力模块(如图1(c))中计算广泛注意力,即计算特征权重并再次分配,得到对重建最有利的注意力特征。最后将不同时间组的注意力特征拼接并融合,得到多时间组融合特征,交由亚像素卷积层提升分辨率。将生成的HR残差图像和上采样的LR帧相加,得到最终的超分辨率重建结果,训练时注重低分辨率图像与高分辨率图像间的高频残差部分,节约低频部分的运算耗费时间。

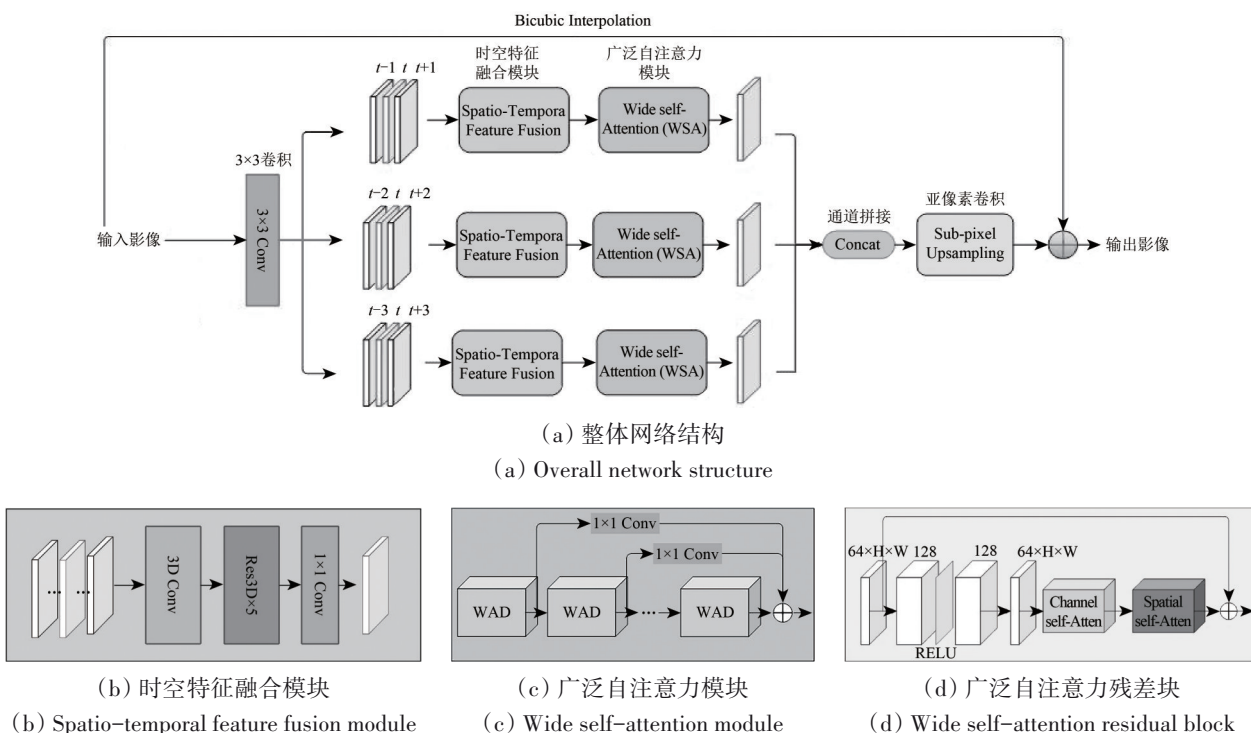


图1 多时间组自注意力网络模型结构图

Fig. 1 Structure diagram of multi-temporal self-attention network model

$$SR = MTSA(f_{i-3}, \dots, f_i, \dots, f_{i+3}) \oplus Up(f_i) \quad (1)$$

式中, SR 表示网络超分辨率重建结果, $MTSA(\cdot)$ 表示自注意力网络模型, f_i 表示输入的序列帧, $Up(\cdot)$ 表示上采样操作。

2.2 多时间组特征融合

超时空数据具有时间的有序性及空间的连续性。序列帧间成像时间间隔不同, 每帧之间都含有一定的高频成分, 为充分利用这种信息, 提出一种多时间组特征融合方法, 根据帧速率将序列帧划分为多个时间组, 通过观测多个时间状态下序列帧的特征信息, 挖掘超时空数据间的时空相关性。各个时间组先通过时空特征融合模块提取各组内的序列帧时空特征。时空特征融合模块(图1(b))由3D卷积、3D残差块以及 1×1 的2D卷积组成。3D卷积能够对空间与时间同时建模(Tran等, 2015), 可提取序列帧中的时空特征。2D卷积提取空间特征, 3D残差块由5组3D卷积残差结构组成, 提取序列帧的深度特征; 最后使用一层 1×1 的卷积来整合时间组内的特征。各个时间组的输出 F_g 可以表示为

$$F_g = H_{STF}(I_g) \quad (2)$$

式中, I_g 表示不同的时间组, $H_{STF}(\cdot)$ 表示时空特征融合模块。

各时间组提取完序列帧的时空特征后, 将其送入广泛自注意力模块(图1(c)), 计算特征权重, 并再次分配得到广泛注意力特征, 具体的广泛注意力计算在2.3节中介绍。将各个时间组计算出的注意力特征拼接融合, 得到多时间组融合特征:

$$F_{MT} = f_{conv}(cat(F_1, F_2, F_3)) \quad (3)$$

式中, f_{conv} 表示卷积操作, $cat(\cdot)$ 表示通道拼接操作, F_i 为各时间组特征。多时间组融合特征中整合了多个时间组下的注意力特征, 对序列帧中的时空信息利用更加充分, 重建的效果更好。

2.3 广泛自注意力

注意力机制(Vaswani等, 2017)模仿了生物观察行为的内部过程, 能够帮助关注重点信息, 以及控制全局信息, 擅长序列任务的处理。卷积操作则善于提取细节特征, 两者的结合渐渐成为趋势(Srinivas等, 2021; Dai等, 2021)。注意力计算中通常需要池化等数据压缩操作, 对于分类

或分割任务可以去除一些冗余信息, 但应用于重建过程会导致像素信息丢失, 影响重建精度。因此, 本文在超分重建模型中引入了广泛自注意力模块, 利用多个非线性函数的组合来增加注意力范围, 完成高动态映射。广泛自注意力模块输入的是序列帧时空特征融合模块得到的时空特征, 模块包括通道注意力与空间注意力。注意力计算首先对向量间进行相似度计算得到特征的权重值, 之后将权重与对应的特征加权求和得到注意特征。通过对时空特征的权重计算, 无需配准也可以找到序列帧中对重建更加有利的特征。

广泛自注意力整体结构采用递归残差的形式(图1(c)), 可以更好的学习权值像素间的长期依赖关系。内部具体的广泛自注意力残差块如图1(d)所示, 输入特征首先经过宽激活卷积, 通过扩展RELU前的特征, 使更多的信息通过。即将输入特征的通道数量由64个增加到128个, 以此来获取更多的像素信息。之后依次通过广泛通道自注意力(图2(a))与广泛空间自注意力(图2(b))计算注意力特征。通道注意力中将查询向量 W_q 完全折叠, 值向量 W_v 进行低压缩(保留一半通道维度和所有空间维度), 然后在查询向量上使用 softmax 函数增加注意力范围, 最后通过 sigmoid 函数进行动态映射。通道注意力 F_{CA} 的计算可表示为

$$F_{CA} = F \odot f_{sig}\left(\text{LayerNorm}\left(f_{soft}(W_q) \otimes W_v\right)\right) \quad (4)$$

式中, F 表示输入特征, $f_{sig}(\cdot)$ 表示 sigmoid 函数, $f_{soft}(\cdot)$ 表示 softmax 函数, W_q 和 W_v 分别表示查询向量和值向量, LayerNorm 表示通道方向上的归一化操作。

输入特征经过通道注意力计算后, 将通道注意力特征送入后续的空间注意力块进行空间注意力计算, 处理方式同通道自注意力相似, 可用公式表示为

$$F_{SA} = F_{CA} \odot f_{sig}\left(W_v \otimes f_{soft}\left(\text{GlobalPooling}(W_q)\right)\right) \quad (5)$$

经过通道注意与空间注意计算后可得到最终的广泛注意力特征。广泛自注意力中的特征扩展和低程度压缩可以有效防止像素信息的丢失, Softmax 和 Sigmoid 函数的结合增加了注意力计算范围并完成高动态范围映射。从而为序列帧建模的同时保证了像素细节信息的提取。

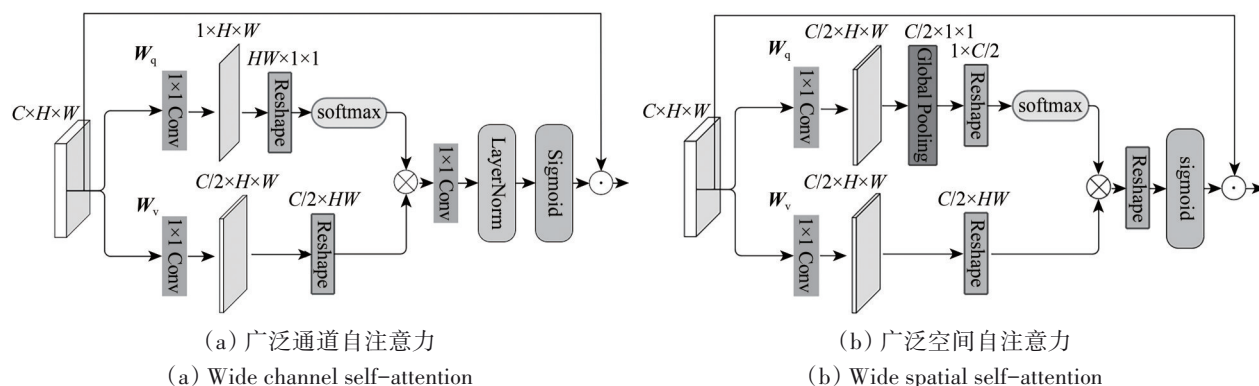


图2 广泛自注意力结构图

Fig. 2 Wide self-attention structure diagram

3 数据结果处理与分析

3.1 实验数据

本文使用了高分四号超时相遥感图像数据集对算法的适用性进行验证, 数据采集自地球同步遥感卫星, 图像拍摄时间为2016年8月20日, 图像尺寸大小为10240×10240, 部分海南岛地区图像如图3所示。

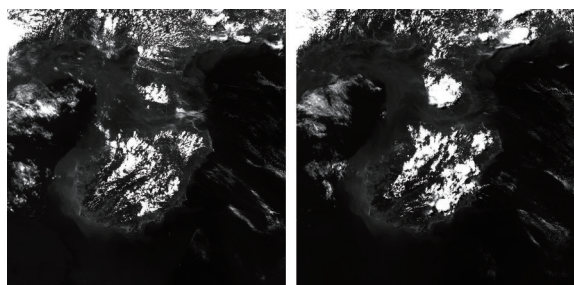


图3 高分四号海南岛地区

Fig. 3 Hainan Island region of the GF-4 data

3.2 实验设置

在网络训练中, 从高分四号数据集中裁切出1024×1024的图像作为HR真实图像, 为了得到HR-LR对应的训练数据对, 将HR真实图像下采样后的图像作为LR图像。进行训练时会从LR图像中随机裁切32×32大小的块作为网络输入, 并在HR图像中裁切出相对应的块, 并利用旋转、翻转等操作进行数据扩充, 共取得40000个训练样本, 以此来提高网络的泛化能力。网络的训练批次大小batch-size设置为32, 梯度优化算法采用自适应矩估计Adam优化器 (Kingma和Ba, 2015), 初始

学习率设置为 4×10^{-4} 。损失函数采用Charbonnier损失函数 (Lai等, 2019), 经实验测试, 相较于均方误差损失函数 (MSE), Charbonnier函数的收敛速度更快, 最终结果的表现也较好。

$$L(SR, HR) = \sqrt{\|SR - HR\|^2 + \varepsilon^2} \quad (6)$$

式中, SR 表示网络超分辨率重建结果, HR 表示真实图像, ε 设置为 1×10^{-3} 。

实验中的数据评测指标采用超分辨率中常用的峰值信噪比 (PSNR) 与结构相似性 (SSIM), PSNR用来计算原始图像与重建图像间的像素误差, 数值越大则表明重建图像失真越小; SSIM用来计算原始图像与重建图像的相似度程度, 数值越大则表示重建图像与原始图像差距越小。

本文的实验环境为NVIDIA Quadro P6000 GPU, 深度学习框架为Pytorch。

3.3 高分四号超时相数据对比实验

本节将本文算法MTSA与其他超分辨率算法进行比较, 包括双三次插值算法Bicubic, 深度学习算法FSRCNN (Dong等, 2016), 深度学习算法ESPCN ((Shi等, 2016), 深度学习算法EDSR (Lim等, 2017), 残差通道注意力网络RCAN (Zhang等, 2018), 级联注意力网络FCCAN (Fan等, 2020)。通过不同上采样因子 (2倍、3倍、4倍) 在高分四号数据集上的重建结果进行比较。

各算法在高分四号数据上的重建结果客观评价指标如表1、表2所示, 两倍重建的主观视觉效果如图4所示, 4倍重建的主观视觉效果如图5所示。测试数据为高分四号海南岛东部与西部裁切大小为1024×1024的图像块。

表 1 不同算法在高分四号数据集上的重建结果定量指标 (PSNR)

Table 1 Quantitative index of reconstruction results of different algorithms on GF-4 data (PSNR)

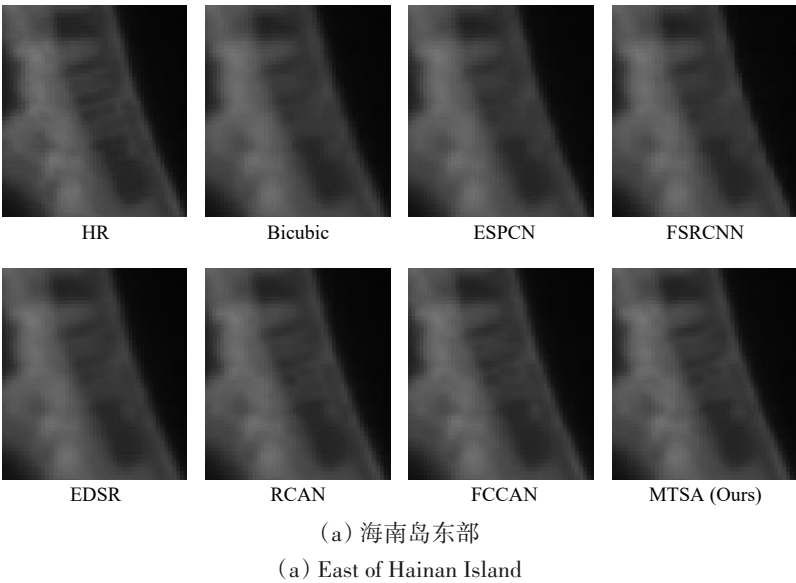
数据集	不同算法/dB							
	采样因子	Bicubic	ESPCN	FSRCNN	EDSR	RCAN	FCCAN	MTSA(本文)
海南岛东部	2 倍	45.97	45.49	45.43	46.99	47.21	47.90	48.63
	3 倍	41.39	40.23	39.44	42.78	43.55	44.32	46.14
	4 倍	38.07	36.95	36.50	39.22	39.73	40.66	43.46
海南岛西部	2 倍	42.97	42.91	43.24	43.92	44.35	44.93	45.46
	3 倍	38.82	37.48	37.16	39.54	40.15	40.86	42.62
	4 倍	36.07	34.91	34.75	36.56	37.05	37.72	39.96

表 2 不同算法在高分四号数据集上的重建结果定量指标 (SSIM)

Table 2 Quantitative index of reconstruction results of different algorithms on GF-4 data (SSIM)

数据集	不同算法/dB							
	采样因子	Bicubic	ESPCN	FSRCNN	EDSR	RCAN	FCCAN	MTSA(本文)
海南岛东部	2 倍	0.9873	0.9862	0.9848	0.9883	0.9891	0.9895	0.9903
	3 倍	0.9768	0.9712	0.8937	0.9788	0.9805	0.9821	0.9847
	4 倍	0.9628	0.9575	0.9297	0.9664	0.9646	0.9674	0.9774
海南岛西部	2 倍	0.9766	0.9768	0.9775	0.9796	0.9815	0.9822	0.9833
	3 倍	0.9513	0.9383	0.9337	0.9549	0.9582	0.9627	0.9701
	4 倍	0.9208	0.9058	0.9028	0.9250	0.9270	0.9305	0.9529

从表 1 及表 2 中可以看到，在高分四号海南岛数据上，2 倍、3 倍、4 倍超分辨率重建时本文算法的客观评价指标均优于对比算法，拥有先进的重建性能。海南岛东部两倍重建时较 Bicubic 算法在 PSNR 上提升 2.66 dB，海南岛西部两倍重建时较 Bicubic 算法在 PSNR 上提升 2.49 dB。进行较大因子超分辨率重建时算法依然保持着很高的重建性能，如在 4 倍超分辨率重建时，Bicubic 算法在海南岛东部的 PSNR 值仅有 38.07 dB，本文算法为 43.46 dB，获得了巨大提升，主观视觉效果如图 5 所示，传统插值及 ESPCN、FSRCNN、EDSR 等算法在 4 倍超分重建时较为模糊，仅能识别大致轮廓。基于注意力的算法 RCAN、FCCAN 以及本文算法重建结果更为清晰，能够恢复出更多的细节纹理信息，更接近于真实图像。



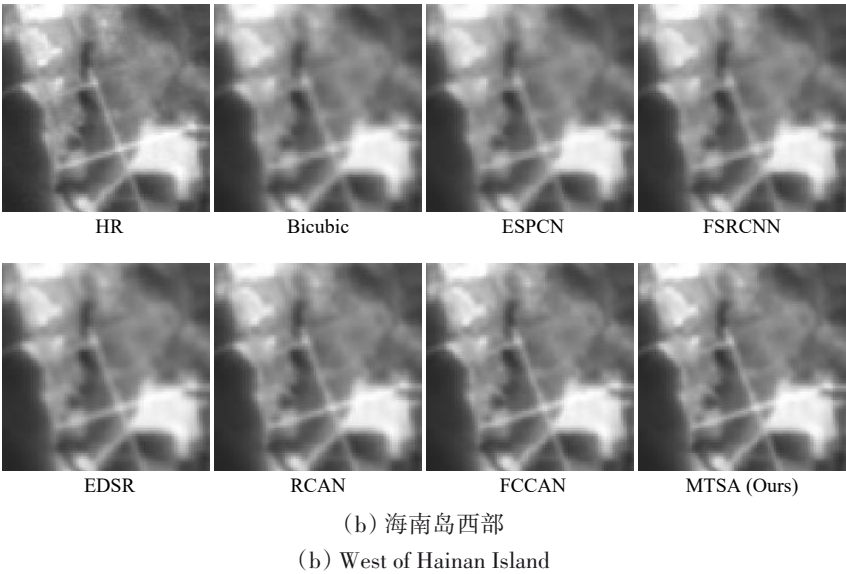


图4 不同算法在高分四号数据集上2倍重建视觉效果
Fig. 4 Double reconstruction visual effect of different algorithms on GF-4 data

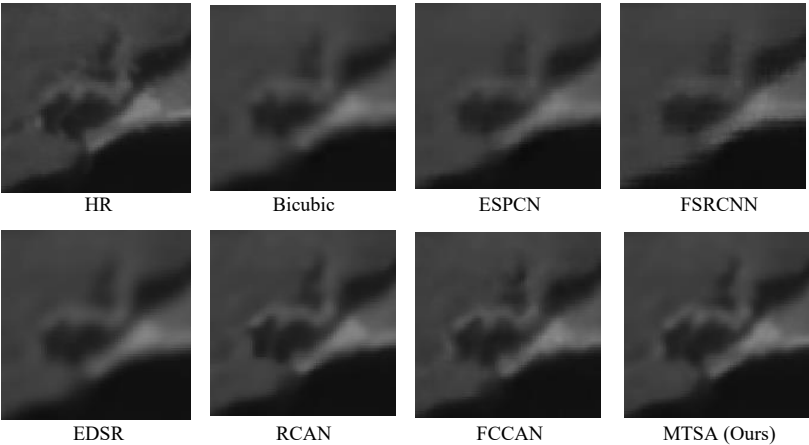


图5 不同算法在高分四号数据集上4倍重建视觉效果
Fig. 5 Quad puple reconstruction visual effect of different algorithms on GF-4 data

表3展示了不同算法在高分四号数据集上2倍重建所需的计算时间。可以看出，各算法在该数据的计算时间上，传统插值算法 Bicubic 所需时间最短，深度网络算法 EDSR 及 RCAN 所需时间较长，本文算法的计算时间适中。

3.4 消融实验

本节将设计多组消融实验对算法各模块的有效性进行验证。第1组消融实验移除多时间组特征融合模块对其验证；第2组消融实验移除广泛自注意力模块（通道注意力+空间注意力）进行验证；第3组消融实验移除通道注意力模块进行验证；第4组消融实验移除空间注意力进行验证。此外还将讨论算法中特征通道参数的选择。

表3 不同算法在高分四号数据集上的计算时间
Table 3 Calculation time of different algorithms on GF-4 data

算法	计算时间/s
Bicubic	0.23
ESPCN	0.83
FSRCNN	1.15
EDSR	58.78
RCAN	46.23
FCCAN	8.21
MTSA(本文)	11.83

表4对各种网络结构两倍超分重建的结果进行了分析，通过对比移除多时间组特征融合模块前后的重建结果，可以证明该模块的有效性，同理

可证得广泛自注意力模块及其中的通道自注意力与空间自注意力的有效性。

表 4 不同网络结构的重建结果分析

网络结构	PSNR/dB	SSIM
MTSA 移除多时间组特征融合	47.91	0.9893
MTSA 移除广泛自注意力模块(通道+空间)	47.37	0.9891
MTSA 移除通道自注意力	48.17	0.9897
MTSA 移除空间自注意力	48.32	0.9898
MTSA(本文)	48.63	0.9903

表5对宽激活特征通道参数对重建结果的影响进行了分析，可以发现，随着特征通道数的增加，重建结果的PSNR值也在上升，但在通道数为128之后继续增加数量并不能有效提升重建质量。

表 5 宽激活特征通道参数对重建结果的影响分析

特征通道数	PSNR/dB	SSIM
32	46.71	0.9875
64	48.06	0.9896
128	48.63	0.9903
256	48.30	0.9898
512	48.21	0.9898

3.5 真实图像实验

本节针对高分四号真实图像进行实验，对海南岛地区的图像块不进行下采样，将原始图像作为输入进行重建。图6左侧图像为海南岛部分地区原始图像，右侧为本文算法重建结果。可以看出重建后图像的道路河流等细节信息更加清晰，验证了本文算法的有效性。

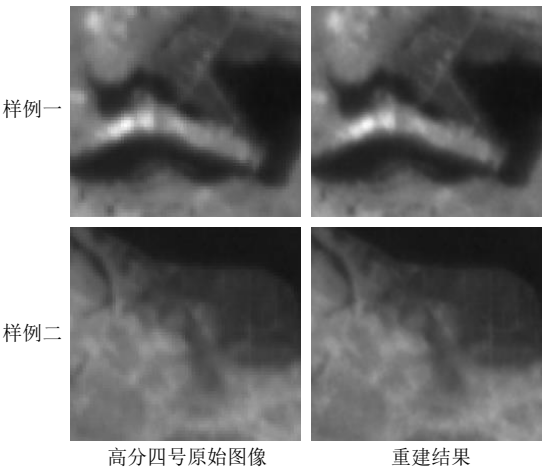


图 6 高分四号原始图像重建
Fig. 6 GF-4 original image reconstruction

4 结 论

本文针对超时相数据特点，设计了一个端到端的学习网络能够从序列帧低分辨率图像中恢复出高分辨率图像；提出一种多时间组特征融合方法，提取多个时间维度下的序列帧时空信息，充分挖掘超时相数据中丰富的时空域相关信息；广泛自注意力残差块帮助扩展特征量并减小压缩，利用多个非线性函数的组合来增加注意力范围，完成高动态映射，保留图像重建后的高频细节信息。在高分四号数据上的对比实验表明，客观评测指标和主观视觉效果均优于对比算法，验证了本文算法对超时相数据超分辨率重建的有效性与先进性，拥有较好的重建性能。本文算法在计算时间上有待优化，与其他算法相比虽提升了精度，但计算时间也增加了不少，后续研究中将考虑算法结构的优化，如改变网络的残差结构等，在保证精度的同时减少计算时间。

参考文献(References)

Arefin M R, Michalski V, St-Charles P L, Kalaitzis A, Kim S, Kahou S E and Bengio Y. 2020. Multi-image super-resolution for remote sensing using deep recurrent networks//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Seattle: IEEE: 816-825 [DOI: 10.1109/CVPRW50498.2020.00111]

Caballero J, Ledig C, Aitken A, Acosta A, Totz J, Wang Z, and Shi W. 2017. Real -time video super -resolution with spatio -temporal networks and motion compensation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 4778-4787 [10.1109/CVPR.2017.304]

Dai Z H, Liu H X, Le Q V and Tan M X. 2021. CoAtNet: marrying convolution and attention for all data sizes//Proceedings of the Advances in Neural Information Processing Systems 34. [s.l.]: NeurIPS: 3965-3977

Dong C, Loy C C, He K M and Tang X O. 2016. Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(2): 295-307 [DOI: 10.1109/TPAMI.2015.2439281]

Dong X Y, Sun X, Jia X P, Xi Z H, Gao L R and Zhang B. 2021. Remote sensing image super-resolution using novel dense-sampling networks. IEEE Transactions on Geoscience and Remote Sensing, 59(2): 1618-1633 [DOI: 10.1109/TGRS.2020.2994253]

Fan Z H, Dan T T, Yu H H, Liu B Y and Cai H M. 2020. Single fundus image super-resolution via cascaded channel-wise attention network//42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Montreal: IEEE: 1984-1987 [DOI: 10.1109/EMBC44109.2020.9176428]

Haut J M, Fernandez-Beltran R, Paoletti M E, Plaza J and Plaza A.

2019. Remote sensing image superresolution using deep residual channel attention. *IEEE Transactions on Geoscience and Remote Sensing*, 57(11): 9277-9289 [DOI: 10.1109/TGRS.2019.2924818]
- He Z and He D. 2020. Deep learning-based super-resolution for GF-4 satellite imagery. *Journal of Remote Sensing (in Chinese)*, 24(12): 1500-1510 (贺智, 贺丹. 2020. 基于深度学习的高分四号卫星图像超分辨率重建. *遥感学报*, 24(12): 1500-1510) [DOI: 10.11834/jrs.20208397]
- Jiang K, Wang Z Y, Yi P, Wang G C, Lu T and Jiang J J. 2019. Edge-enhanced GAN for remote sensing image superresolution. *IEEE Transactions on Geoscience and Remote Sensing*, 57(8): 5799-5812 [DOI: 10.1109/TGRS.2019.2902431]
- Kingma D P and Ba J. 2015. Adam: A method for stochastic optimization//3rd International Conference on Learning Representations. San Diego: ICLR
- Lai W S, Huang J B, Ahuja N and Yang M H. 2019. Fast and accurate image super-resolution with deep Laplacian pyramid networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(11): 2599-2613 [DOI: 10.1109/TPAMI.2018.2865304]
- Li F, Yang X, Lu X T, Xin L, Lu M and Zhang N. 2021. A new hyper-temporal imaging mode for spaceborne CMOS cameras. *National Remote Sensing Bulletin*, 25(1): 514-525 (李峰, 杨雪, 鲁啸天, 辛蕾, 鹿明, 张南. 2021. 面向星载CMOS相机的超时相工作方式研究. *遥感学报*, 25(1): 514-525) [DOI: 10.11834/jrs.20210607]
- Lim B, Son S, Kim H, Nah S and Lee K M. 2017. Enhanced deep residual networks for single image super-resolution//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu: IEEE: 1132-1140 [DOI: 10.1109/CVPRW.2017.151]
- Molini A B, Valsesia D, Fracastoro G and Magli E. 2020. DeepSUM: deep neural network for super-resolution of unregistered multi-temporal images. *IEEE Transactions on Geoscience and Remote Sensing*, 58(5): 3644-3656 [DOI: 10.1109/TGRS.2019.2959248]
- Nie J, Deng L, Hao X L, Liu M and He Y. 2018. Application of GF-4 satellite in drought remote sensing monitoring: a case study of Southeastern Inner Mongolia. *Journal of Remote Sensing (in Chinese)*, 22(3): 400-407 (聂娟, 邓磊, 郝向磊, 刘明, 贺英. 2018. 高分四号卫星在干旱遥感监测中的应用. *遥感学报*, 22(3): 400-407) [DOI: 10.11834/jrs.20187067]
- Shi W Z, Caballero J, Huszar F, Totz J, Aitken A P, Bishop R, Rueckert D and Wang Z H. 2016. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE: 1874-1883 [DOI: 10.1109/CVPR.2016.207]
- Srinivas A, Lin T Y, Parmar N, Shlens J, Abbeel P and Vaswani A. 2021. Bottleneck transformers for visual recognition//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE: 16514-16524 [DOI: 10.1109/CVPR46437.2021.01625]
- Tian Y P, Zhang Y L, Fu Y and Xu C L. 2020. TDAN: temporally-deformable alignment network for video super-resolution//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE: 3357-3366 [DOI: 10.1109/CVPR42600.2020.00342]
- Tran D, Bourdev L, Fergus R, Torresani L and Paluri M. 2015. Learning spatiotemporal features with 3D convolutional networks//2015 IEEE International Conference on Computer Vision. Santiago: IEEE: 4489-4497 [DOI: 10.1109/ICCV.2015.510]
- Vaswani A, Shazzer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser L and Polosukhin I. 2017. Attention is all you need//Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc.: 5998-6008
- Wang L G, Guo Y L, Liu L, Lin Z P, Deng X P and An W. 2020. Deep video super-resolution using HR optical flow estimation. *IEEE Transactions on Image Processing*, 29: 4323-4336 [DOI: 10.1109/TIP.2020.2967596]
- Xu L N and He L X. 2017. GF-4 images super resolution reconstruction based on POCS. *Acta Geodaetica et Cartographica Sinica*, 46(8): 1026-1033 (许丽娜, 何鲁晓. 2017. 基于凸集投影的高分四号卫星影像超分辨率重建. *测绘学报*, 46(8): 1026-1033) [DOI: 10.11947/j.AGCS.2017.20170070]
- Yang X, Li F, Lu M, Xin L, Lu X T and Zhang N. 2022. New super-resolution reconstruction method based on Mixed Sparse Representations. *National Remote Sensing Bulletin*, 26(8): 1685-1697 (杨雪, 李峰, 鹿明, 辛蕾, 鲁啸天, 张南. 2022. 混合稀疏表示模型的超分辨率重建. *遥感学报*, 26(8): 1685-1697) [DOI: 10.11834/jrs.20219409]
- Yang Y and Newsam S. 2010. Bag-of-visual-words and spatial extensions for land-use classification//Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems. San Jose: ACM: 270-279 [DOI: 10.1145/1869790.1869829]
- Ying X Y, Wang L G, Wang Y Q, Sheng W D, An W and Guo Y L. 2020. Deformable 3D convolution for video super-resolution. *IEEE Signal Processing Letters*, 27: 1500-1504 [DOI: 10.1109/LSP.2020.3013518]
- Zhang Y L, Li K P, Li K, Wang L C, Zhong B N and Fu Y. 2018. Image super-resolution using very deep residual channel attention networks//2018 15th European Conference on Computer Vision. Munich: Springer: 294-310 [DOI: 10.1007/978-3-030-01234-2_18]

Super-resolution reconstruction of hypertemporal remote sensing images based on self-attention

TANG Xiaotian^{1,2}, YANG Xue^{1,3}, LI Feng¹, MA Jun², LIANG Liang⁴

1.China Qian Xuesen Space Technology Laboratory, China Academy of Space Technology, Beijing 100094, China;

2.College of Software, Henan University, Kaifeng 475004, China;

3.International Institute for Earth System Science, Nanjing University, Nanjing 210023, China;

4.Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

Abstract: Video satellite hypertemporal data have the characteristics of high temporal resolution, while the single-frame super-resolution reconstruction algorithm can only use the information of the image frame itself, and the reconstruction effect is limited. Therefore, how to utilize fully and effectively the rich spatiotemporal information in hypertemporal data in the super-resolution reconstruction of video satellite images is an issue of interest.

Aiming at the characteristics of hypertemporal data, this study proposes a self-attention-based super-resolution reconstruction model of hypertemporal remote sensing images. The model can mine high-frequency information from low-resolution images through an end-to-end network. High-resolution images are recovered from multiple frames of low-resolution images. First, the hypertemporal sequence frames are divided into multiple time groups according to the frame rate, and the spatial and temporal information under different time groups are extracted by using the characteristics of 3D convolution to model space and time simultaneously. It pays attention to the calculation range and completes high dynamic mapping, extracts rich spatial detail information while realizing hypertemporal sequence frame modeling, and finally fuses the features of multiple time groups and completes the reconstruction through subpixel convolution to improve the resolution. The advantage of the proposed algorithm is that the multitime group feature fusion method can extract the spatiotemporal information of sequence frames in multiple time dimensions and fully mine the rich spatiotemporal-related information in the hypertemporal data; the improved self-attention block can be used without registration. It can complete the modeling of sequence frames and improve the extraction ability of detailed spatial information.

Experiments on the GF-4 dataset show that the subjective visual effect and objective evaluation index of the algorithm are better than those of the comparison algorithm. When the GF-4 dataset is reconstructed twice, the PSNR value is improved by more than 2.49 dB compared with the bicubic interpolation algorithm, and it still has a strong reconstruction performance when the reconstruction is four times, which is a huge improvement compared with the bicubic interpolation algorithm. Experimental results show that the method has good super-resolution reconstruction effect, which is beneficial to the application of hypertemporal data in various fields.

The self-attention-based super-resolution model of hypertemporal remote sensing images proposed in this study fully extracts the spatiotemporal information in hypertemporal data by dividing multiple time groups and calculating the attention features in each time group. The combination of multitemporal group feature fusion and self-attention enables the modeling of overphase sequence frames while ensuring the ability to extract detailed information. Comparative experiments on the GF-4 dataset show that the algorithm in this study is superior to the compared algorithm in terms of objective evaluation indicators and subjective visual effects, verifying the effectiveness and advancement of the algorithm in super-resolution reconstruction of hypertemporal data and the algorithm's improved reconstruction performance. However, the proposed algorithm must be optimized in terms of calculation time. In a follow-up research, the algorithm structure will be optimized (e.g., changing the residual structure of the network) to reduce the calculation time while ensuring accuracy.

Key words: remote sensing, hyper-temporal data, super-resolution reconstruction, deep learning, fusion feature of multiple time groups, wide self-attention, GF-4

Supported by National Key Research and Development Program of China (No.2020YFA0714100); High Resolution Earth Observation System Major Project (No.30-Y30F06-9003-20/22)