基于随机置换的红外图像超分辨率Transformer

熊承义[ab],何缘[ab*],高志荣[c],曹雨轩[ab],李明月[ab],王薇[ab]
(中南民族大学 a.电子信息工程学院; b.智能无线通信湖北省重点实验室; c.计算机学院,湖北 武汉 430074)

摘 要 红外图像超分辨率具有广泛应用.当前,视觉Transformer在提升图像超分辨率性能方面展现了良好潜能,但如何有效平衡系统复杂度引起了许多研究者的兴趣.提出了一种基于随机置换的Transformer红外图像超分辨率方法,实现了在较好控制系统复杂度的同时进一步改进图像重构性能.特别地,设计了一种结合随机置换的窗自注意力模块代替原始Swin Transformer的移位窗自注意力模块,通过对特征图在空间维度的随机置换运算有效改进了窗自注意力的全局依赖关系的学习能力.基于多个公开的红外图像数据集的实验比较结果验证了本文方法的有效性.

关键词 红外图像超分辨率;视觉Transformer;全局建模;随机置换
中图分类号 TP391.4 **文献标志码** A **文章编号** 1672-4321(2026)04-0505-09
doi:10.20056/j.cnki.ZNMDZK.20250847

Transformer of infrared image super-resolution based on random permutation

Xiong Chengyi[ab], He Yuan[ab*], Gao Zhirong[c], Cao Yuxuan[ab], Li Mingyue[ab], Wang Wei[ab]
(South-Central Minzu University, a. College of Electronic and Information Engineering; b. Hubei Key Lab of Intelligent Wireless Communication; c. College of Computer Science, Wuhan 430074, Hubei China)

Abstract Infrared image super-resolution has wide applications. Currently, the Visual Transformer has shown great potential in improving the performance of image super-resolution, but how to effectively balance system complexity has attracted the interest of many researchers. A random permutation based Transformer for infrared image super-resolution is proposed, which achieves further improvement in image reconstruction performance while controlling system complexity well. Specifically, a window self-attention module combining random permutation is designed to replace the original Swin Transformer's shifted window self-attention module, effectively improving the learning ability of global dependency relationships of window self-attention through random permutation operations in the spatial dimension of the feature map. Experimental comparison results based on multiple public infrared image datasets validate the effectiveness of this method.

Keywords infrared image super-resolution; Visual Transformer; global modeling ability; random permutation

红外图像(Infrared Image)是通过捕获物体反射的辐射能量所形成的图像,该类图像可以抵抗恶劣天气条件干扰,在军事侦察、航空航天及居家看护等军事和民生领域具有深远的影响[1].然而,在红外图像成像和传输的过程中,受到红外图像成像设备和红外波段的波长等影响,红外图像的分辨率通常低于可见光图像且图像细节信息不丰富[2].因此,为了获得更加适合在实际应用场景中使用的高质量红外图像,提高红外图像的分辨率十分必要.单图超分辨率(Single Image Super-Resolution,SISR)是计算机视觉领域中的一项重要技术,旨在从退化的单幅低分辨率(Low Resolution,LR)图像中恢复出对应

收稿日期 2024-05-18
作者简介 熊承义(1969-),男,教授,博士,研究方向:图像处理与模式识别,E-mail:xiongcy@mail.scuec.edu.cn
基金项目 多谱信息处理技术国家重点实验室基金资助项目(6142113210303);中央高校基本科研业务专项资金资助项目(CZY21013)

的高分辨率(High Resolution,HR)图像[3].在日常生产和生活中,超分辨率技术凭借其操作简便且成本低廉的优点,在监控[4]、医疗成像[5]等多个领域被广泛运用.

近年来,深度学习的蓬勃发展为红外图像超分辨率提供了广阔的发展平台.He等[6]提出了一种具有多个感受野的深度神经网络级联结构,通过不同大小的感受野网络更有针对性地重建出高分辨率红外图像.Han等[7]从低分辨率的近红外图像及对应的高分辨率可见光图像中分别提取高频分量,将两分量concat后经卷积神经网络(Convolutional Neural Network,CNN)得到合成的高分辨率近红外图像.Du等[8]提出的网络交替输入彩色图像和红外图像,将红外图像与具有较高分辨率的彩色图像相结合来提高红外图像的空间分辨率.虽然这些基于CNN的超分网络都达到了可观的图像重构效果,但是卷积操作依赖于设定的卷积核大小,且每个卷积核只能感知到输入图像的局部信息.有研究发现:在高频区域中,红外图像通常缺少可见光图像中普遍存在的纹理信息,而具有语义和结构信息的全局信息占比更大.在降采样的过程中这些高频结构会极大地遭到破坏[9].因此,在进行超分辨率重建红外图像时,捕获其全局信息十分重要.

伴随着Transformer框架的问世[10],基于Transformer的超分辨率重构网络进入大众视野,并因其出色的建模长距离关系的能力深受青睐.但该类方法对计算资源的消耗巨大,应用受限,如何有效平衡重构性能和计算复杂度是一个亟须解决的问题.Swin Transformer[11]网络提供了一种新的解决思路,该网络在窗口内进行自注意力计算并提出移位窗口策略增进窗口之间的信息交互.Liang等[12]则将Swin Transformer的思想运用到图像复原领域,提出的SwinIR网络有效地平衡了重构性能和计算复杂度.Zhang等[13]交替使用稀疏注意力块和密集注意力块从而构成ART网络,在增大感受野的同时兼顾局部信息,取得了优异的重构效果.EMT网络[14]则提出混合Transformer块,通过在不同方向上的移位操作增强了基于Transformer的网络对局部信息的整合能力.基于窗的自注意力方法有效提高了Transformer的运用效率,具有优良的性能.然而,这些基于自然图像的超分辨率方法未考虑红外图像对比度低、成像细节相较于自然图像更加模糊等特性,当将这些网络直接运用到红外图像领域时,往往不能取得理想的效果[2].

Xiao等[15]于2023年提出的用于图像去噪、去雨等领域的ShuffleFormer网络具有良好的性能.此网络通过随机洗牌操作,在局部窗口内对非局部信息建模,在复杂度不变的基础上辅助网络更好地进行图像复原任务.受到这一启发,本文将该思想运用到图像超分辨率领域,并结合红外图像缺少细节纹理且存在更多全局语义信息的特点,提出了基于随机置换的Transformer红外图像超分辨率方法.该方法在空间维度上对特征图进行随机置换运算后再在窗口内计算自注意力,打破了窗口自注意力的局部性壁垒,同时增强了网络的全局建模能力.本文提出的方法可以更有效建模图像的全局信息,在保持线性复杂度的情况下改进图像恢复质量,较好平衡了网络的全局建模能力与计算复杂度.在多个红外图像数据集上的实现结果表明,本文方法相较于同类方法有着较好的竞争优势.

1 提出的方法

本节详细介绍提出的基于随机置换的Transformer红外图像超分辨率网络.通过在空间维度上对特征图进行随机置换运算,增强了窗口自注意力对全局依赖关系的学习能力,改进了基于窗口自注意力的Transformer网络的局部性限制.不同于基于滑动局部窗口的Transformer网络,本文方法在进行自注意力计算之前经随机置换运算打乱特征图的原始像素排列,使得特征图中任意位置的像素都具有相同的概率进入到同一个局部窗口内.虽然后续的自注意力计算仍然在局部窗口中进行,但参与计算的像素来自图像的各个部分,弥补了局部窗口对全局信息建模能力的不足.在自注意力计算结束后,进行相对应的逆置换运算恢复图像.这样的方法既可以不增加计算复杂度,又提高了全局建模能力.

1.1 网络结构

本文提出的随机置换的Transformer红外超分辨率重构网络的整体结构如图1所示,该网络主要由三部分组成,分别是:浅层特征提取模块(Shallow Feature Extraction Module,SFEM)、深层特征提取模块(Deep Feature Extraction Module,DFEM)和图像重建模块(Reconstruction Module,RM).其中,深层特征提取模块是整个网络的重点,该部分将在1.2节中展开进行详细介绍.

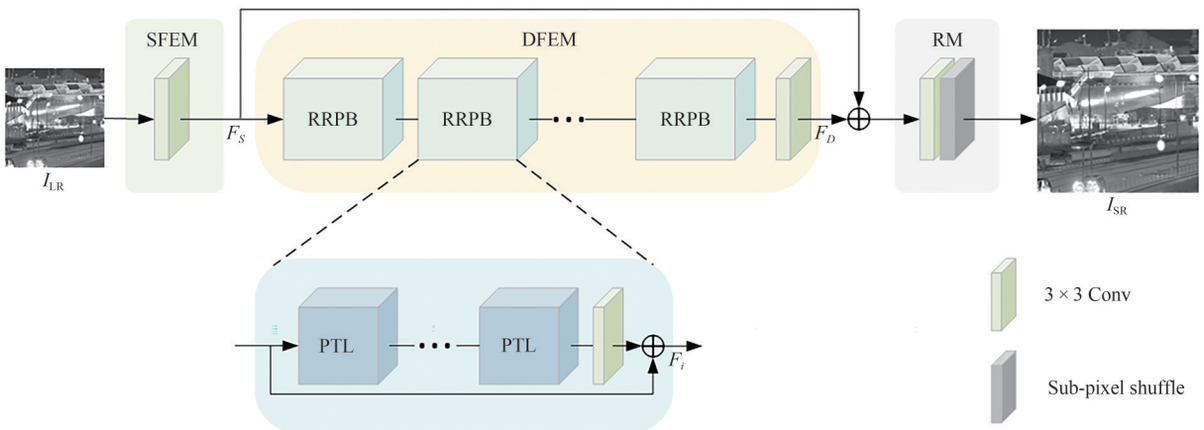


图 1　网络的整体结构框图

Fig. 1　Block diagram of the whole network structure

（1）浅层特征提取模块．该模块使用一个卷积核大小为 3×3 的卷积层对输入的 LR 图像提取浅层特征 F_s，该过程可表示为：

$$F_s = H_{\mathrm{SFEM}}(I_{\mathrm{LR}}), \tag{1}$$

其中：$H_{\mathrm{SFEM}}(\cdot)$ 表示卷积操作，I_{LR} 代表输入的低分辨率图像．之后，提取到的浅层特征 F_s 会传入深层特征提取模块中进一步提取特征．除此以外，F_s 还会直接经残差连接送入到图像重建模块，达到在重建图像中保留输入图像的低频特征信息的目的．

（2）深层特征提取模块．该模块通过堆叠多个残差随机置换块（Residual Random Permutation Block，RRPB）对上一阶段所得的浅层特征 F_s 进一步提取特征信息．其中，每个随机置换块由多个置换 Transformer 层（Permutated Transformer Layer，PTL）和一个 3×3 的卷积层组成．假设共有 k 个 RRPB，则第 i 个 RRPB 的输出 $F_i(1 \leqslant i \leqslant k)$ 可表示为：

$$F_i = H_{\mathrm{RRPB}}(F_{i-1}),\ i = 1, 2, \cdots, k, \tag{2}$$

其中，$H_{\mathrm{RRPB}}(\cdot)$ 表示残差随机置换块用来提取图像深层特征的操作．在残差随机置换块堆叠结束后，使用一个卷积核大小为 3×3 的卷积层对提取到的特征进行融合，得到最终的深层特征 F_{D}，该过程可表示为：

$$F_D = \mathrm{Conv}(F_1, F_2, \cdots, F_k). \tag{3}$$

（3）图像重建模块．该模块以浅层特征 F_s 和深层特征 F_D 作为输入，经由一个 3×3 的卷积层和一个亚像素卷积层[16]，恢复出对应大小的超分辨率图像 I_{SR}，该过程可表示为：

$$I_{\mathrm{SR}} = H_{\mathrm{REC}}(F_S + F_D), \tag{4}$$

其中，$H_{\mathrm{REC}}(\cdot)$ 表示图像重建模块中卷积层和亚像素卷积层的操作．

1.2　置换 Transformer 层

全局信息主要负责提供图像的上下文信息和结构信息，可以帮助网络更好地理解图像内容，从而恢复出更加真实的图像细节．受 ShuffleFormer 网络[15]的启发，本文堆叠多个连续的 RRPB 模块在提取图像深层特征的同时增强网络的全局建模能力，其中每个 RRPB 由多个置换 Transformer 层和一个 3×3 的卷积层构成．

置换 Transformer 层是 RRPB 的核心，该模块基于窗口自注意力，通过对特征图进行随机置换运算提高模型于局部窗口内对全局依赖关系的建模能力，其结构如图 2 所示．在图像超分辨率重建任务中，局部信息和全局信息都发挥着十分重要的作用．因此，本文在置换 Transformer 层中交替使用窗口多头自注意力（Window Multi-head Self-attention，WMSA）和新提出的置换窗口多头自注意力（Permutated Window Multi-head Self-attention，PWMSA）以兼顾局部和全局信息．此外，为了避免网络受到随机置换运算所带来的影响，本文在置换 Transformer 层中加入了卷积核大小为 3×3 的深度卷积层对位置信息进行建模．在 PTL 中，输入特征首先经过层归一化（LN）结构和深度卷积层得到具有位置信息的特征．随后该特征被送入窗口多头自注意力（WMSA）模块，将所得结果与输入特征相加后再经 LN 和多层感知机（MLP）结构对特征进行进一步增强．将第 l 个 PTL 的输入特征表示为 z^{l-1}，则该过程可表示为：

$$\hat{z}^l = \mathrm{WMSA}\left(\mathrm{Dconv}\left(\mathrm{LN}\left(z^{l-1}\right)\right)\right) + z^{l-1}, \tag{5}$$

$$z^l = \mathrm{MLP}\left(\mathrm{LN}\left(\hat{z}^l\right)\right) + \hat{z}^l, \tag{6}$$

其中：WMSA（·）代表窗口多头自注意力操作，Dconv（·）代表 3×3 的深度级卷积层，LN（·）代表层归一化

操作,MLP(·)代表多层感知机操作.

随后,将上一部分所得的特征z^l作为下一阶段的输入,在计算注意力时采用新提出的PWMSA模块,其余过程与上一阶段相同.该过程可表示为:

$$\hat{z}^{l+1} = \mathrm{PWMSA}\left(\mathrm{Dconv}\left(LN\left(z^l\right)\right)\right) + z^l, \quad (7)$$

$$z^{l+1} = \mathrm{MLP}\left(\mathrm{LN}\left(\hat{z}^{l+1}\right)\right) + \hat{z}^{l+1}, \quad (8)$$

其中,PWMSA(·)代表置换窗口多头自注意力操作.

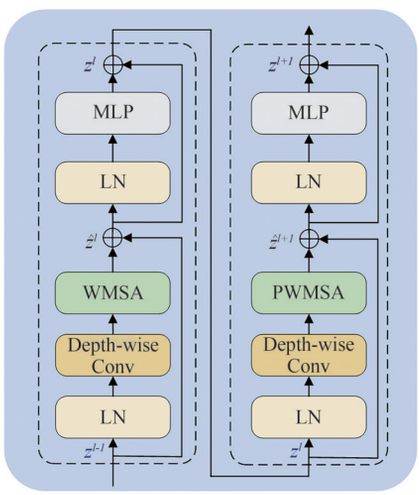


图2 置换Transformer层的内部结构图

Fig. 2 Internal structural of the permutated Transformer layer

新提出的PWMSA模块在窗口自注意力中加入了随机置换运算和逆置换运算.随机置换运算包含在空间维度分别基于行向和列向的随机置换,具体过程可表示为:

$$X_{\mathrm{PH}} = \mathrm{RPH}(X), \quad (9)$$

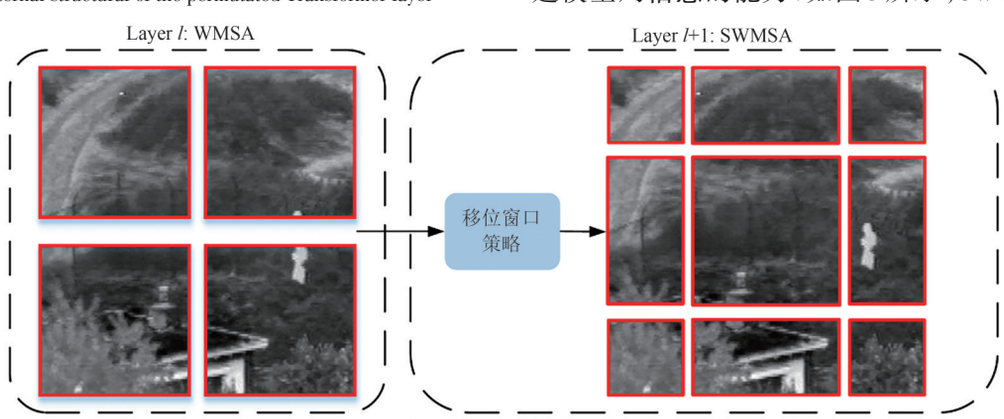
$$X_P = \mathrm{RPW}\left(X_{\mathrm{PH}}\right), \quad (10)$$

其中:RPH(·)、RPW(·)分别代表行向和列向的随机置换运算,X_{PH}代表行向随机置换运算后的结果,X_{P}代表列向随机置换运算后的结果.结合了随机置换运算的窗口自注意力模块PWMSA的计算复杂度为:

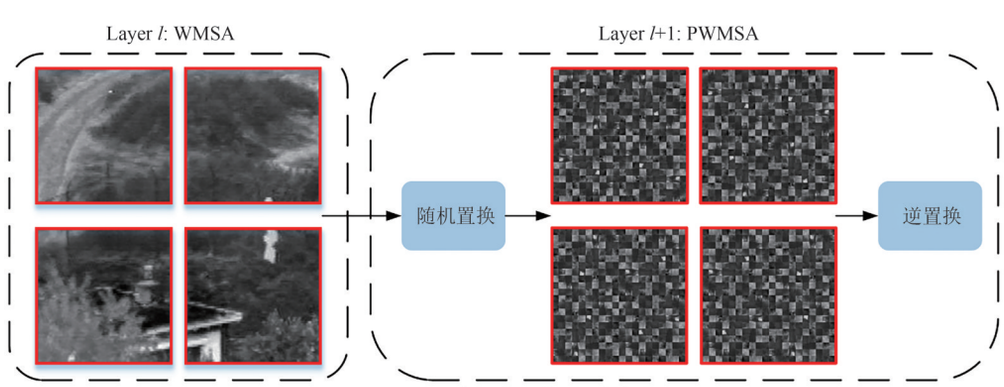
$$\Omega(\mathrm{PWMSA}) = 4hwM^2 + 2M^2hwC, \quad (11)$$

其中:h和w表示输入图像的长和宽,C表示图像的通道数,M表示窗口大小.因此,PWMSA的计算复杂度与图像的大小呈线性相关,与WMSA相等.

与Swin Transformer[11]中的移位窗口多头自注意力(Shifted Window Multi-head Self-attention,SWMSA)相比,本文提出的PWMSA通过随机置换运算打乱图像的像素排列,在对窗口不做任何调整的情况下使得窗口内包含图像各个位置的信息,提高了网络建模全局信息的能力.如图3所示,SWMSA利用移



(a) 基于移位窗口的窗自注意力



(b) 基于随机置换的窗自注意力

图3 移位窗口自注意力和置换窗口自注意力对比图

Fig. 3 Comparison of SWMSA and PWMSA

位窗口策略对不同窗口的信息进行交互，但是移位操作的移动范围受到限制，导致窗口间的信息交互被局限在一定的范围内；而本文所使用的 PWMSA 在空间维度上对特征图进行随机置换运算，使得参与自注意力计算的 token 来源于图像的任意位置，相较于 SWMSA 的移位窗口策略更加灵活．自注意力的置换等变性[10]在理论上给予了 PWMSA 进行随机置换运算的支持，其思想是：在局部窗口内进行自注意力计算之前，先在空间维度上对输入图像进行随机置换运算．虽然之后仍然是在相同大小的局部窗口内进行自注意力计算，但图像经过随机置换运算后其任意位置的像素都有相同的概率进入到同一窗口内参与后续的计算，改进了窗口自注意力学习全局依赖关系的能力．在自注意力计算结束后，通过与随机置换相对应的逆置换运算恢复图像原有的像素排列．随机置换运算和逆置换运算是一个可逆变换对，不会引起输入图像信息的丢失．

2　实验

2.1　实验设置和实验数据

在模型训练阶段，考虑到红外数据集通常难以获取且数量较少，不足以满足基于深度学习的算法需要大量数据进行训练的需求．因此，在训练中使用迁移学习的方法．具体而言，首先将在超分辨率领域中广泛使用的 DIV2K[17]自然图像数据集作为训练集进行训练，得到预训练模型，再将预训练模型在 IR100[18]红外数据集上进行训练微调．其中，IR100 数据集包含了 100 张使用车载红外摄像头拍摄的街景图像．训练开始前，对 DIV2K 数据集中共 800 幅图像和 IR100 数据集中共 100 幅图像分别以放大因子（×2、×3、×4）进行双三次下采样，得到对应的低分辨率图像作为模型的输入图像．另外，通过水平翻转和随机翻转等方法增强训练图像的数据多样性．在训练过程中，采用 Adam 优化器进行优化，并将损失函数设置为 L1 函数．训练的初始学习率设为 1×10^{-4}，当训练迭代次数达到 250000、400000、450000 和 475000 时，将学习率调整为上一阶段的一半，总迭代次数为 500000 次．实验中 RRPB 的个数为 6，每个 RRPB 包含 6 个 PTL．训练时设置 batch-size 为 16，并将输入图像随机裁剪成大小为 32×32 像素的图像块输入网络进行训练．

在模型测试阶段，使用 result-A[18]、result-C[18]、Thermal 50[19]以及 CDNMRF[6]四个红外图像数据集作为测试集来验证所提方法的有效性．其中，Thermal 50 是 Rivadeneira 等[19]所提出的数据集中的测试集部分．这四个红外数据集包含了野外、室内、人物、街景和建筑物等不同场景，共 114 张图像．网络性能通过峰值信噪比（PSNR）和结构相似性指数（SSIM）两个通用的评价指标在 Y 通道上的计算结果体现．所有的实验均基于 Pytorch 框架，使用两张 NVIDIA GTX 1080Ti GPU 显卡完成．

2.2　实验结果及分析

为验证本文方法的有效性，选取了 5 个具有代表性的超分辨率方法：SRCNN[20]、EDSR[21]、SwinIR[12]、ART[13]和 DASR[9]与本文方法分别在×2、×3 和×4 三种放大因子下进行比较．其中，SRCNN 和 EDSR 是图像超分辨率领域中经典的基于 CNN 的算法，SwinIR 和 ART 算法是在 Transformer 框架下构建的自然图像超分辨率算法，而 DASR 算法是基于 Transformer 的红外图像超分辨率算法．上述 5 个算法的网络结构设置均与原论文相同，与本文方法在同等条件下训练后进行比较．

2.2.1　定量指标分析

本文方法与其他算法在四个红外测试集上的对比结果如表 1 所示，表 2 则以 Thermal 50 测试集为例，展现了本文方法与其他算法在 4 倍上采样时的网络参数以及平均 PSNR．

由表 1 可得，在不同的放大因子下，本文方法在所有数据集上均优于其他方法的结果．具体来说，在×2 放大因子下，本文方法在 result-C 数据集上的优势最为明显，与次好的 ART 模型相比，就 PSNR 而言高了 0.11 dB，就 SSIM 指标而言高了 0.0015．在×3 放大因子下，本文方法在各测试集上均处于领先位置．在 result-A、result-C 和 CDNMRF 数据集上，本文方法比第二好的模型在 PSNR 指标上分别高了 0.05、0.03 和 0.04 dB．在 Thermal 50 数据集上则比第二优的模型的 PSNR 高了 0.09 dB．在×4 放大因子下，本文方法在所有数据集上同样均取得最高的 PSNR 和 SSIM．以 PSNR 指标为例，在 result-C 和 CDNMRF 数据集上本文方法比次优结果分别高出 0.03 dB 和 0.02 dB，在 result-A 和 Thermal 50 数据集上则比次优结果都高出 0.04 dB．此外，从表 2 可以看出，在 Thermal 50 数据集上，本文方法的参数量虽比取得次优结果的 SwinIR 模型高，但是取得了优势较大的重构性能改进．

综上所述，本文方法在 3 种放大因子下均获得

表1 不同算法在×2、×3、×4放大因子下的平均PSNR和SSIM

Tab. 1 Average PSNR/SSIM of different algorithms at ×2, ×3, ×4 amplification factors

模型	scale	result-A		result-C		Thermal 50		CDNMRF
		PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
SRCNN	×2	36.29	0.9146	34.24	0.9283	38.61	0.9717	40.39	0.9552
EDSR		37.28	0.9243	38.15	0.9370	41.92	0.9808	43.00	0.9577
SwinIR		38.41	0.9318	39.32	0.9438	43.17	0.9825	44.12	0.9656
ART		38.60	0.9348	39.53	0.9463	43.32	0.9833	44.28	0.9667
DASR		38.61	0.9351	39.52	0.9461	43.34	0.9836	44.32	0.9670
本文		**38.66**	**0.9365**	**39.64**	**0.9478**	**43.44**	**0.9842**	**44.38**	**0.9671**
SRCNN	×3	33.63	0.8535	34.30	0.8712	35.23	0.9270	38.70	0.9384
EDSR		34.39	0.8685	35.07	0.8852	37.05	0.9481	40.85	0.9459
SwinIR		35.54	0.8802	36.25	0.8957	38.31	0.9523	42.12	0.9558
ART		35.59	0.8814	36.31	0.8972	38.33	0.9531	42.27	0.9575
DASR		35.60	0.8827	36.27	0.8969	38.32	0.9530	42.30	0.9578
本文		**35.65**	**0.8829**	**36.34**	**0.8983**	**38.42**	**0.9541**	**42.34**	**0.9581**
SRCNN	×4	32.49	0.8255	31.92	0.8046	31.73	0.8198	35.65	0.9233
EDSR		32.24	0.8168	32.85	0.8366	33.80	0.9046	38.80	0.9366
SwinIR		33.55	0.8374	34.18	0.8555	35.19	0.9145	40.16	0.9475
ART		33.57	0.8382	34.19	0.8560	35.18	0.9148	40.20	0.9488
DASR		33.48	0.8372	34.06	0.8550	35.13	0.9146	40.24	0.9492
本文		**33.61**	**0.8388**	**34.22**	**0.8569**	**35.23**	**0.9151**	**40.26**	**0.9495**

注:加粗字体表示最优结果;下划线字体表示次优结果.

表2 在×4放大因子下不同算法的参数量及在Thermal 50数据集上的平均PSNR

Tab. 2 Average PSNR and parameters of different algorithms on Thermal 50 dataset at ×4 amplification factor

模型	SRCNN	EDSR	SwinIR	ART	DASR	Ours
参数量/M	**0.05**	43.09	11.90	16.55	1.28	16.66
PSNR/dB	31.73	33.80	35.19	35.18	35.13	**35.23**

注:加粗字体表示最优结果;下划线字体表示次优结果.

了良好重构性能提升.究其原因,是因为本文方法通过随机置换运算,使网络在窗口内更有效地利用了图像的全局信息,提高了网络的建模能力.

2.2.2 定性指标分析

为更全面地体现本文方法的有效性,在放大因子为4时,将本文方法与其他算法的重构结果在视觉效果上进行了对比,对比结果如图4-7所示.

图4展现了在4倍放大因子下对测试图像result-C_result2的重构结果,并局部放大了店铺招牌部分.可以看出:本文方法重构出的字母"NERO"更为清晰,尤其是在"N""O"两个字母上,本文重构出了更清晰的字母轮廓.图5展现了在4倍放大因子下对测试图像Thermal50_0042的重构结果,并局部放大了栏杆部分.从图5可看出,本文方法和DASR重构出的图像更接近原始图像,但本文方法恢复出的栏杆粗细更均匀,且线条扭曲程度更小.而其他对比方法重构出的栏杆都出现了不同程度的扭曲,且恢复出的栏杆较细,与原始图像不符.图6展现了在4倍放大因子下对测试图像Thermal50_0046的重构结果,并局部放大了屋顶部分.其中,只有本文方法和SwinIR恢复出了正确的房顶瓦片纹理,但本文重建出的瓦片线条更加清晰,重构出的瓦片细节较SwinIR而言更加真实.图7则展现了在4倍放大因子下对测试图像CDNMRF_07的重构结果,并局部放大了键盘部分.同样地,其他比较方法都未能恢复出正确的键盘格子方向,将本应该是水平的键盘格重建成了斜纹格,出现了不同程度的失真,只有本文方法重建出正确的键盘方向.

以上结果表明:本文方法不仅在定量性能比较上表现优势明显,而且在定性的视觉效果比较上也优于其他算法,能得到更清晰的高分辨率图像.

2.3 消融实验及分析

2.3.1 置换Transformer层的有效性分析

为验证本文方法使用置换Transformer层的有效性,比较了3种不同的自注意力组合形式,包括:全部使用置换窗自注意力的PWMSA+PWMSA方法,置换窗自注意力在前而原始窗自注意力在后的PWMSA+WMSA,以及原始窗自注意力在前而置换窗自注意力在后的WMSA+PWMSA.在放大因子为4的比较结果如表3所示.

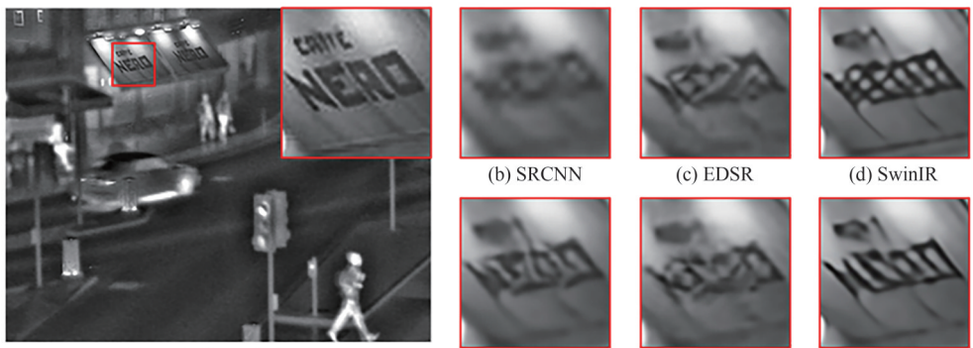


图 4　不同算法对 result-C 中的 result2 在× 4 放大因子下的重构结果

Fig. 4　Reconstruction results of result2 in result-C by different algorithms at × 4 amplification factor

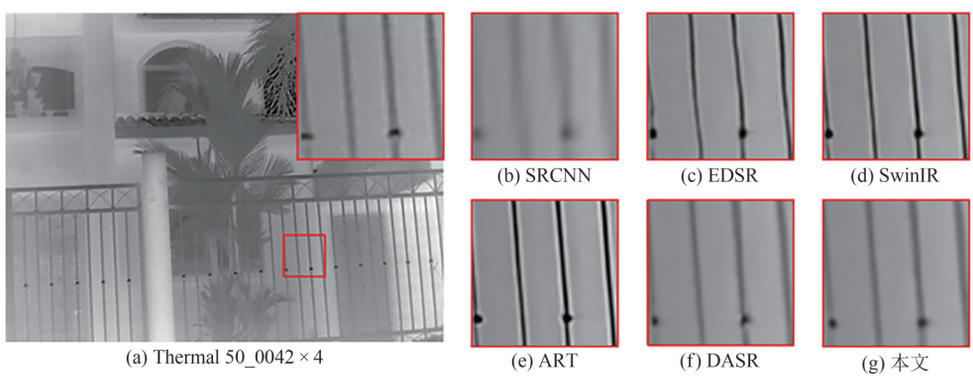


图 5　不同算法对 Thermal 50 中的 0042 在× 4 放大因子下的重构结果

Fig. 5　Reconstruction results of 0042 in Thermal 50 of different algorithms at × 4 amplification factor

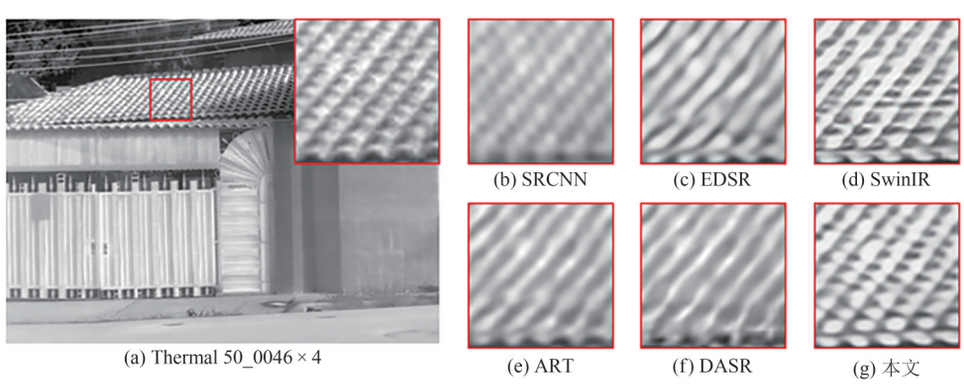


图 6　不同算法对 Thermal 50 中的 0046 在× 4 放大因子下的重构结果

Fig. 6　Reconstruction results of 0046 in Thermal 50 of different algorithms at × 4 amplification factor

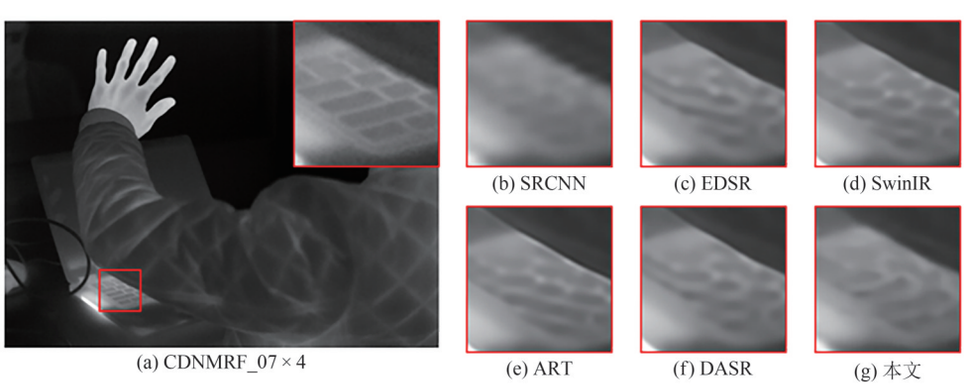


图 7　不同算法对 CDNMRF 中的 07 在× 4 放大因子下的重构结果

Fig. 7　Reconstruction results of 07 in CDNMRF of different algorithms at × 4 amplification factor

从表3结果可见:交替使用WMSA和PWMSA的方法在性能上优于只使用PWMSA的方法,且参数量更少;WMSA在前PWMSA在后的方法略优于PWMSA在前而WMSA在后的方法.因此,本文在置换Transformer层中交替使用WMSA和PWMSA,且先进行WMSA后进行PWMSA.

表3 自注意力组合形式对重建性能的影响

Tab. 3 Effect of combination of self-attention on the reconstruction performance

自注意力组合形式	参数量/M	Thermal 50		CDNMRF	
		PSNR/dB	SSIM	PSNR/dB	SSIM
PWMSA+PWMSA	16.74	33.40	0.9077	37.91	0.9435
PWMSA+WMSA	16.66	33.81	0.9108	38.33	0.9447
WMSA+PWMSA(本文)	16.66	**33.86**	**0.9121**	**38.35**	**0.9456**

注:加粗字体表示最优结果.

2.3.2 置换Transformer层中深度级卷积的有效性分析

在置换Transformer层中,引入了深度级卷积以提高模型对位置信息建模的能力.为了验证深度级卷积进行位置建模的有效性,给出了放大因子为2的情况下的重构性能比较,结果如表4所示.

根据表4结果可见:引入深度级卷积后明显提升了模型的重构质量,在定量评价指标上得到更优.因此,本文方法在置换Transformer层中加入深度级卷积以重构出更高质量的图像.

表4 深度级卷积对重建性能的影响

Tab. 4 Effect of depth-wise convolution on the reconstruction performance

方法	Depth-wise Conv	参数量/M	result-A		Thermal 50	
			PSNR/dB	SSIM	PSNR/dB	SSIM
Base	×	16.51	37.56	0.9244	41.01	0.9763
本文	√	16.58	**37.58**	**0.9251**	**41.10**	**0.9766**

注:加粗字体表示最优结果.

3 结语

本文针对红外图像的退化特性,研究了一种基于随机置换的红外图像超分辨率Transformer网络.通过在窗口自注意力中结合随机置换运算,有效地平衡了网络的重构性能和计算复杂度,增强了窗自注意力对全局依赖关系的学习能力.本文在置换Transformer层中交替使用WMSA和PWMSA,兼顾了全局信息和局部信息的提取,在控制模型计算复杂度的情况下提高了全局建模能力.本文方法与同类方法相比,可以恢复出更真实的红外图像纹理细节,有效提升了重构性能,具有一定的应用价值.

参 考 文 献

[1] Wang C, Li X. Infrared image feature analysis and application in the field of image encryption [C]//2021 IEEE 4th International Conference on Automation, Electronics and Electrical Engineering (AUTEEE). Shenyang: IEEE, 2021: 578-583.

[2] Danaci K I, Akagunduz E. A survey on infrared image & video sets [J]. Multimedia Tools and Applications, 2023, 83(6): 16485-16523.

[3] Wang Z, Chen J, Hoi S C H. Deep learning for image super-resolution: A survey [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(10): 3365-3387.

[4] Zhang L, Zhang H, Shen H, et al. A super-resolution reconstruction algorithm for surveillance images [J]. Signal Processing, 2010, 90(3): 848-859.

[5] Teh I, Mcclymout D, Carruth E, et al. Improved compressed sensing and super-resolution of cardiac diffusion MRI with structure-guided total variation [J]. Magnetic Resonance in Medicine, 2020, 84(4): 1868-1880.

[6] He Z, Tang S, Yang J, et al. Cascaded deep networks with multiple receptive fields for infrared image super-resolution [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 29(8): 2310-2322.

[7] Han T Y, Kim Y J, Song B C. Convolutional neural network-based infrared image super resolution under low light environment [C]//2017 25th European Signal Processing Conference (EUSIPCO). Kos: IEEE, 2017: 803-807.

[8] Du J, Zhou H, Qian K, et al. RGB-IR cross input and sub-pixel upsampling network for infrared image super-resolution [J]. Sensors, 2020, 20(1): 281.

[9] Liang S, Song K, Zhao W, et al. DASR: Dual-Attention Transformer for infrared image super-resolution [J]. Infrared Physics & Technology, 2023, 133: 104837.

[10] Vaswani A, Shazeer N, Parmar N, et al. Attention is All you Need [J]. Advances in Neural Information Processing Systems, 2017:30.

[11] Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2021: 9992-10002.

[12] Liang J, Cao J, Sun G, et al. SwinIR: Image restoration using swin transformer [C]//2021 IEEE/CVF International Conference on Computer Vision Workshops

(ICCVW). Montreal: IEEE, 2021: 1833-1844.

[13] Zhang J, Zhang Y, Gu J, et al. Accurate image restoration with attention retractable transformer [C]// International Conference on Learning Representations. Kigali: ICLR, 2023: 1-13.

[14] Zheng L, Zhu J, Shi J, et al. Efficient mixed transformer for single image super-resolution [J]. Engineering Applications of Artificial Intelligence, 2024, 133: 108035.

[15] Xiao J, Fu X, Zhou M, et al. Random shuffle transformer for image restoration[C]//Proceedings of the 40th International Conference on Machine Learning. Honolulu: ACM, 2023: 38039-38058.

[16] Shi W, Caballero J, Huszar F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 1874-1883.

[17] Agustsson E, Timofte R. NTIRE 2017 challenge on single image super-resolution: Dataset and Study[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Honolulu: IEEE, 2017: 1122-1131.

[18] Huang Y, Jiang Z, Lan R, et al. Infrared image super-resolution via transfer learning and PSRGAN[J]. IEEE Signal Processing Letters, 2021, 28: 982-986.

[19] Rivadeneira R E, Suárez P L, Sappa A D, et al. Thermal image SuperResolution through deep convolutional neural network[C]//International Conference on Image Analysis and Recognition. Waterloo: ICIAR, 2019: 417-426.

[20] Dong C, Loy C C, He K, et al. Image super-resolution using deep convolutional networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(2): 295-307.

[21] Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Honolulu: IEEE, 2017: 1132-1140.

（责编&校对　刘钊）