



## 矿井图像的多维特征与残差注意力网络超分辨率重建方法

程健 米立飞 李昊 李和平 王广福 马永壮

### 引用本文:

程健, 米立飞, 李昊, 等. 矿井图像的多维特征与残差注意力网络超分辨率重建方法[J]. 煤炭科学技术, 2024, 52(11): 117–128.

CHENG Jian, MI Lifei, LI Hao. Coalmine image super-resolution reconstruction via fusing multi-dimensional feature and residual attention network[J]. Coal Science and Technology, 2024, 52(11): 117–128.

在线阅读 View online: <https://doi.org/10.12438/cst.2024-1055>

### 您可能感兴趣的其他文章

#### Articles you may be interested in

#### 基于深度神经网络的煤矿井下低光照图像增强算法

Deep neural network-based image enhancement algorithm for low-illumination images underground coal mines  
煤炭科学技术. 2023, 51(9): 231–241 <https://doi.org/10.12438/cst.2022-1626>

#### 融合金字塔结构与注意力机制的煤矿井下巡检机器人PT目标检测算法

Research on mine underground inspection robot target detection algorithm based on pyramid structure and attention mechanism coupling  
煤炭科学技术. 2024, 52(6): 206–215 <https://doi.org/10.12438/cst.2023-1071>

#### 基于阈值和权重Census变换的煤矿井下图像立体匹配算法

A stereo matching algorithm for coal mine underground images based on threshold and weight under Census transform  
煤炭科学技术. 2024, 52(6): 216–225 <https://doi.org/10.12438/cst.2023-1169>

#### 基于同态滤波方法的煤矿井下图像增强技术研究

Research on coal mine underground image recognition technology based on homomorphic filtering method  
煤炭科学技术. 2023, 51(3): 241–250 <https://doi.org/10.13199/j.cnki.cst.2021-0774>

#### 基于CFS-YOLO算法的复杂工况环境下煤矸图像识别方法

Image recognition method of coal gangue in complex working conditions based on CES-YOLO algorithm  
煤炭科学技术. 2024, 52(6): 226–237 <https://doi.org/10.12438/cst.2023-1967>

#### 基于深度神经网络的煤岩显微图像划痕检测与修复

Scratch detection and restoration of coal photomicrograph via deep neural network  
煤炭科学技术. 2023, 51(S2): 275–284 <https://doi.org/10.13199/j.cnki.cst.2023-0058>



关注微信公众号, 获得更多资讯信息



移动扫码阅读

程 健, 米立飞, 李 昊, 等. 矿井图像的多维特征与残差注意力网络超分辨率重建方法[J]. 煤炭科学技术, 2024, 52(11): 117–128.  
CHENG Jian, MI Lifei, LI Hao, *et al.* Coalmine image super-resolution reconstruction via fusing multi-dimensional feature and residual attention network[J]. Coal Science and Technology, 2024, 52(11): 117–128.

# 矿井图像的多维特征与残差注意力网络超分辨率重建方法

程 健<sup>1,2,3,4</sup>, 米立飞<sup>1,2,3</sup>, 李 昊<sup>2,4</sup>, 李和平<sup>1,2,3,4</sup>, 王广福<sup>2,4</sup>, 马永壮<sup>2,4</sup>

(1. 煤炭科学研究总院, 北京 100013; 2. 煤炭科学研究总院有限公司 矿山人工智能研究院, 北京 100013; 3. 煤炭智能开采与岩层控制全国重点实验室, 北京 100013; 4. 天地科技股份有限公司, 北京 100013)

**摘 要:** 煤矿井下环境复杂, 受光照、煤尘、水雾的影响, 采集的图像往往存在细节模糊、纹理缺失等问题, 低分辨率的矿井图像对煤矿安全监控的智能化发展带来诸多制约。图像超分辨率重建作为一种重要的图像处理技术, 旨在从矿井低分辨率图像中恢复出清晰的高分辨率图像, 从而显著提升煤矿智能监测与安全管理可靠性。针对矿井图像边缘纹理信息缺失、细节模糊不清等质量退化问题, 笔者提出一种矿井图像的多维特征与残差注意力网络超分辨率重建方法。首先, 采用多分支网络将动态卷积与通道注意力机制进行并行融合, 以“水平-通道”“垂直-通道”交互方式来捕获不同的空间统计特性。其次, 设计了一种递归稀疏自注意力机制, 在线性复杂度下聚合代表性特征图, 自适应选择权重分配, 减少计算过程中的信息冗余。最后, 基于标准多头自注意力机制和残差连接方式构建深层特征提取的基本单元, 将获得的特征信息与浅层特征通过跳跃连接共同输入重建模块, 完成超分辨率矿井图像重建。实验结果表明, 笔者所提方法在客观评价指标和主观视觉分析上较现有主流算法均有明显提升。在矿井数据集的测试中, 2 倍和 4 倍缩放因子下的图像相似性(LPIPS)平均降低 10.97%、9.91%, 峰值信噪比(PSNR)平均提升 4.10%、2.30%, 证明了该方法在恢复矿井图像结构和纹理细节上的有效性。

**关键词:** 煤矿井下图像; 超分辨率重建; 注意力机制; 残差网络; 递归算法

**中图分类号:** TD76; TP18; TP391.41 **文献标志码:** A **文章编号:** 0253-2336(2024)11-0117-12

## Coalmine image super-resolution reconstruction via fusing multi-dimensional feature and residual attention network

CHENG Jian<sup>1,2,3,4</sup>, MI Lifei<sup>1,2,3</sup>, LI Hao<sup>2,4</sup>, LI Heping<sup>1,2,3,4</sup>, WANG Guangfu<sup>2,4</sup>, MA Yongzhuang<sup>2,4</sup>

(1. China Coal Research Institute, Beijing 100013, China; 2. Research Institute of Mine Artificial Intelligence, Chinese Institute of Coal Science, Beijing 100013, China; 3. State Key Laboratory for Intelligent Coal Mining and Strata Control, Beijing 100013, China; 4. Tiandi Science and Technology Co., Ltd., Beijing 100013, China)

**Abstract:** The complex underground environment of coal coalmines, influenced by lighting, coal dust, and water mist, often results in collected images with blurred details and missing textures, leading to decreased image resolution and posing significant limitations to the intelligent development of coal coalmine safety monitoring. Image super-resolution reconstruction, an essential image processing technology, aims to recover clear high-resolution images from low-resolution coalmine images, thereby significantly enhancing the reliability of intelligent monitoring and safety management in coal coalmines. To address issues such as the loss of edge texture information and blurring of details in coalmine images, a coalmine image super-resolution reconstruction method integrating multi-dimensional features and residual attention networks is proposed. First, a multi-branch network is employed to parallelly integrate dynamic convolution and channel attention mechanisms, capturing different spatial statistical characteristics through “horizontal-channel” and “vertical-channel” interactions.

收稿日期: 2024-07-19 策划编辑: 常 琛 责任编辑: 李雅楠 DOI: 10.12438/cst.2024-1055

基金项目: 国家重点研发计划资助项目(2023YFC2907600); 天地科技股份有限公司科技创新创业资金专项重点资助项目(2021-TD-ZD002, 2022-2-TD-ZD001)

作者简介: 程 健(1974—), 男, 四川平昌人, 研究员, 博士生导师, 博士。E-mail: jiancheng@tsinghua.org.cn

通讯作者: 李和平(1978—), 男, 湖北荆门人, 教授级高级工程师, 博士生导师, 博士。E-mail: lih@ccetg-bigdata.com

tions. Secondly, a recursive sparse self-attention mechanism is designed to aggregate representative feature maps under linear complexity, adaptively selecting weight distribution and reducing information redundancy during computation. Finally, the basic unit of deep feature extraction is constructed based on the standard multi-head self-attention mechanism and residual connection, with the obtained feature information and shallow features jointly input into the reconstruction module via skip connections to complete super-resolution reconstruction of coalmine images. Experimental results indicate that the proposed method significantly outperforms existing mainstream algorithms in both objective evaluation metrics and subjective visual analysis. In tests on the coalmine dataset, LPIPS (Learned Perceptual Image Patch Similarity) decreases by an average of 10.97% and 9.91%, while PSNR (Peak Signal-to-Noise Ratio) increases by an average of 4.10% and 2.30% for 2x and 4x scaling factors, respectively, demonstrating the method's effectiveness in restoring the structure and texture details of coalmine images.

**Key words:** coalmine underground images; super-resolution reconstruction; attention mechanism; residual network; recursive algorithm

## 0 引言

矿山数字化、智能化建设是新一轮科技革命下的战略选择<sup>[1]</sup>。高质量的井下图像有助于增强场景的可视化效果和分析能力,提高矿井边缘设备和智能平台的计算性能。由于井下环境存在低光、煤尘干扰等复杂条件,同时受成像设备和成像技术的影响,图像往往存在细节缺失、边缘模糊等问题。因此,研究井下图像超分辨率重建技术以提升成像质量,对实现井下智能化生产和管理意义重大<sup>[2]</sup>。

图像超分辨率重建作为一种重要的图像处理技术,旨在从低分辨率图像中恢复出清晰的高分辨率图像。在井下目标检测与跟踪、安全监控智能分析等视觉计算任务中具有广泛应用<sup>[3]</sup>。通过图像超分辨率重建,可以提高矿井图像的视觉质量和细节程度,使其更适合于人眼观察和矿井智能平台的实际任务分析,增强图像的可用性和应用范围。图像超分辨率重建技术通常通过利用图像的局部和全局特征,结合统计模型或深度学习方法来实现。

传统超分辨率重建技术主要包括插值法、重构法、统计法和基于学习的方法<sup>[4]</sup>。插值法包括双线性插值、双三次插值和样条插值。其核心思想是利用现有像素点信息来估算新增像素点的值。基于重构的方法将超分辨率重建视为图像退化的逆问题。这类方法通过总变差正则化、稀疏表示等策略来恢复图像细节,但是不仅计算复杂度高,处理速度也较慢。基于统计的方法通过构建图像概率模型来推断高分辨率图像,如利用最大后验估计和贝叶斯估计来完成结果预测。基于学习的方法借助低分辨率和高分辨率图像之间的映射关系来生成高分辨率的图像。IRANI<sup>[5]</sup>提出的迭代反投影法是一种典型的传统超分辨率重建方法。该方法通过对图像插值和滤波处理来逐步降低投影空间中的像素误差,能够保持图像局部特征平滑,但该方法非常依赖初始图像

质量,同时在迭代过程中容易放大噪声。

基于深度学习的方法在性能上显著优于传统方法<sup>[6]</sup>。其中,卷积神经网络的出现为图像超分辨率领域带来突破性进展<sup>[7]</sup>。SRCNN<sup>[8]</sup>(Super-Resolution Convolutional Neural Network)首次将卷积神经网络用于超分辨率重建任务,通过将低分辨率图像放大到目标尺寸,然后使用三层卷积网络拟合非线性映射,输出高分辨率图像结果。作为一种端到端学习算法, SRCNN在保持轻量结构的同时也具有较优的重建质量。KIM等<sup>[9]</sup>引入递归神经网络结构,提出了DRCN(Deeply-Recursive Convolutional Network)。DRCN最大的特点在于多次应用相同的卷积层,重复执行递归计算而不会增加参数数量。

随着网络层次的加深,卷积神经网络极易出现梯度消失或爆炸的问题,同时带来计算资源的消耗,训练成本愈发变高。为了解决这一问题,研究者们将残差网络引入超分辨率重建工作中。残差单元以跳层连接的方式来帮助梯度更顺畅的回流,使网络可以向更深层次的方向发展。为了在提高收敛速度的同时保持较高鲁棒性, VDSR<sup>[10]</sup>(Very Deep Super-Resolution)首次将残差网络引入图像超分辨率重建任务中,并使用残差学习和自适应梯度裁剪来加速网络迭代。程德强等<sup>[11]</sup>提出一种多通道递归残差网络,通过复用残差块来优化提升训练效率,同时引入多通道网络和交叉学习机制来融合更多特征信息,对于弱边缘纹理特征的矿井图像展现出较好的重建性能。陈亮亮<sup>[12]</sup>提出了一种双向一致性级联超分辨率重建方法,通过交替计算和递归计算增强传统残差单元的特征表达能力,再利用矿井图像的超分辨率重建和退化的双向一致性学习机制联合监督图像重建,提高了矿井重建图像的感知特征。LIM等<sup>[13]</sup>提出的EDSR(Enhanced Deep Super-Resolution)通过优化残差网络结构,引入多尺度系统,能够在单一模型中重建不同尺度的高分辨率图像。受对抗网络启



发,杨宏业等<sup>[14]</sup>将纹理转移融合到生成对抗网络来重建高分辨率煤矿图像。该方法通过构造生成器和判别器实现数据分布的学习和生成,在重建任务中无需成对数据,对于图像复杂结构和纹理的恢复具有一定的准确性。WANG 等<sup>[15]</sup>在对抗神经网络中引入深度卷积来应对矿井图像的低分辨率恢复问题,在矿井图像保真度上取得良好效果。但容易导致对图像特定特征的过拟合,在应对不同井下场景时缺乏鲁棒性。田子健等<sup>[16]</sup>结合生成对抗思想搭建生成对抗式主体模型框架,使用目标图像域而非单一参考图像驱动判别器监督生成器的训练,避免了矿井图像局部细节丢失。

上述基于神经网络的图像超分辨率重建方法主要关注更宽或更深的架构设计,而忽略了中间层特征之间的相关性。DAI 等<sup>[17]</sup>提出的 SAN(Second-order Attention Network)通过使用二阶特征统计来学习特征之间的相关性,并设计局部残差注意力组用于提高网络表征能力。LIANG 等<sup>[18]</sup>结合了 CNN 和 Transformer 的优势提出 SwinIR(Image Restoration using Swin transformer),利用局部注意力机制处理大尺寸图像,并通过移位窗口方案模拟长距离依赖,在图像特征和注意力权重上具有较好的交互效果,但捕捉更大范围全局上下文的能力需要进一步提升,尤其对于具有明显全局结构的矿井图像需要关注全局特征信息来保持图像的视觉一致性。

在矿井图像的超分辨率重建中,由于矿井环境中的光照不均匀,导致图像往往昏暗且存在强烈阴影和高亮区域<sup>[19]</sup>,其中暗部区域较低的信噪比水平导致噪声增加。矿井生产过程中产生的粉尘使得图像细节缺失,这进一步加剧了矿井图像处理的复杂性。另外井下环境主要由煤层、岩石和设备构成,图像大部分呈现黑色、灰色等单调的色调,在低对比度环境下细节容易与背景混淆,增加了特征提取和细

节增强的难度。现有方法在井下特定成像条件下泛化能力较差,容易造成高频信息丢失,甚至可能重建出假细节。针对上述挑战,笔者提出一种矿井图像的多维特征与残差注意力网络超分辨率重建方法。该方法能够有效提升图像在各维度上的表征能力,优化深层特征之间的交互作用。同时结合递归泛化方法,在全局尺度上完成建模,对低纹理矿井图像的超分辨率重建具有一定的优势。笔者主要贡献包括:

- 1)提出了多维交互注意力机制,采用并行融合策略增强不同通道之间的信息整合,优化静态卷积,更加关注矿井图像高频信息。
- 2)设计了递归稀疏自注意力机制,捕获全局空间信息的同时通过选择算子实现自适应特征聚合,增强矿井图像边缘和轮廓等关键特征。
- 3)设计了残差协同模块,通过融合注意力块和残差结构实现特征映射,完成复杂纹理和结构的有效提取和呈现。

## 1 整体架构和残差混合注意力

### 1.1 整体架构

如图 1 所示,笔者所提矿井图像超分辨率重建网络包括浅层特征提取、残差协同模块组和特征重建模块 3 个部分。图像首先经过卷积层提取浅层特征,浅层特征也直接输出到重建模块以保留低频信息。残差协同模块组由残差协同模块(Residual Collaborative Modules, RCM)构成基本单元,通过多次引入残差连接完成特征融合。在特征重建模块中,使用亚像素卷积对融合后的特征进行上采样,完成图像高质量超分辨率重建。

图中,  $I_{LR}$  为输入的低分辨率图像,  $I_{SR}$  为输出的超分辨率重建图像。浅层特征提取部分主要利用  $3 \times 3$  卷积提取图像低频信息。该层输出可表示为

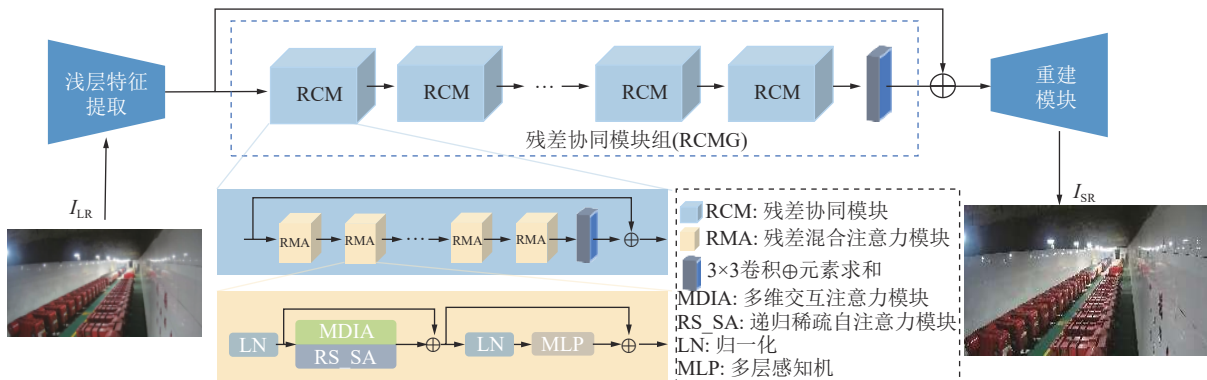


图 1 矿井图像的多维特征与残差注意力网络超分辨率重建网络结构

Fig.1 Coalmine image super-resolution reconstruction via fusing multi-dimensional feature and residual attention network

$$F_S = H_f(I_{LR}) \quad (1)$$

其中,  $F_S \in R^{H \times W \times C}$  为提取的浅层特征;  $H_f(\cdot)$  为  $3 \times 3$  卷积算子构成的特征提取函数。然后, 将  $F_S$  输入到残差协同模块组 (Residual Collaborative Modules Group, RCMG) 进行深层特征映射。在图像多维度上感知和提取高频信息, 在全局和局部区域完成建模。该层输出可表示为

$$F_D = H_d(F_S) \quad (2)$$

其中,  $F_D \in R^{H \times W \times C}$  为提取的深层特征;  $H_d(\cdot)$  为由一系列残差协同模块和一个  $3 \times 3$  卷积层构成的深层特征提取函数, 具体表示为

$$F_i = H_{RCM_i}(F_{i-1}), i = 1, 2, \dots, k \quad (3)$$

$$F_D = H_f(F_k) \quad (4)$$

其中,  $F_k$  为中间层特征;  $H_{RCM_i}(\cdot)$  为第  $i$  个 RCM 块;  $H_f(\cdot)$  为卷积层。最后, 将  $F_S$  和  $F_D$  输入到重建模块, 通过亚像素卷积对融合特征进行上采样, 该层输出表示为

$$I_{SR} = H_R(F_S + F_D) \quad (5)$$

其中,  $H_R(\cdot)$  为重建模块函数。笔者利用 L1 损失来优化网络参数:

$$\mathcal{L}_{L1} = \frac{1}{N} \sum_{i=1}^N \|I_{HR} - I_{SR}\|_1 \quad (6)$$

式中:  $N$  为训练集中样本对个数,  $I_{HR}$  为真实高分辨率图像。

## 1.2 残差混合注意力

鉴于标准 Transformer 块在长距离依赖建模中的优势, 笔者在多头自注意力块的基础上设计残差混合注意力模块。

如图 2 所示, LayerNorm 为归一化层, MLP 为多层感知机。该模块具体结构为: 首先将输入特征进行归一化处理, 多维交互注意力块 (MDIA) 和递归稀

疏自注意力机制 (RS\_SA) 对上一层输出加权融合, 并再次进行归一化处理, 最后通过多层感知机完成非线性激活。其中各层间通过残差连接充分利用局部和全局信息。对于给定的输入特征  $X$ , 残差协同模块的整个计算过程可表示为

$$X_1 = LN(X) \quad (7)$$

$$X_2 = DIA(X_1) + RS_{SA}(X_1) + X \quad (8)$$

$$Y = MLP(LN(X_2)) + X_2 \quad (9)$$

式中:  $X_1$  和  $X_2$  表示中间特征,  $Y$  表示模块最终输出,  $LN$  表示归一化操作,  $MLP$  表示多层感知机。

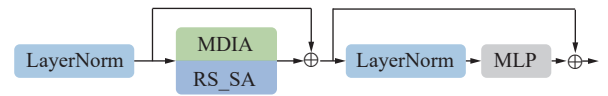


图 2 残差混合注意力 (RMA)

Fig.2 Residual Mixed Attention(RMA)

## 2 深层特征提取模块

### 2.1 多维交互注意力模块

为了更好地捕捉特征之间的交互关系, 增强不同通道间的信息整合能力, 笔者提出了多维交互注意力 (Multi-Dimensional Interactive Attention, MDIA) 网络模块。如图 3 所示, Transpose 为维度旋转操作, FC 为全连接, ReLU 为激活函数。首先将特征  $F$  输入 MDIA 模块的每个分支中。在维度旋转分支中, 通过转置操作分别从“水平-通道”、“垂直-通道”进行特征学习。为了避免复杂的降维和升维过程, 保留原始通道特征的信息完整性, 在该分支中引入高效通道注意力机制<sup>[20]</sup> (Efficient Channel Attention, ECA)。

在 MDIA 模块网络顶部分支和底部分支中, 分别定义  $F^W$  和  $F^H$  为沿着水平和垂直维度转置后的特征图。其中以顶部分支为例,  $F^W$  输入 ECA 模块, 其

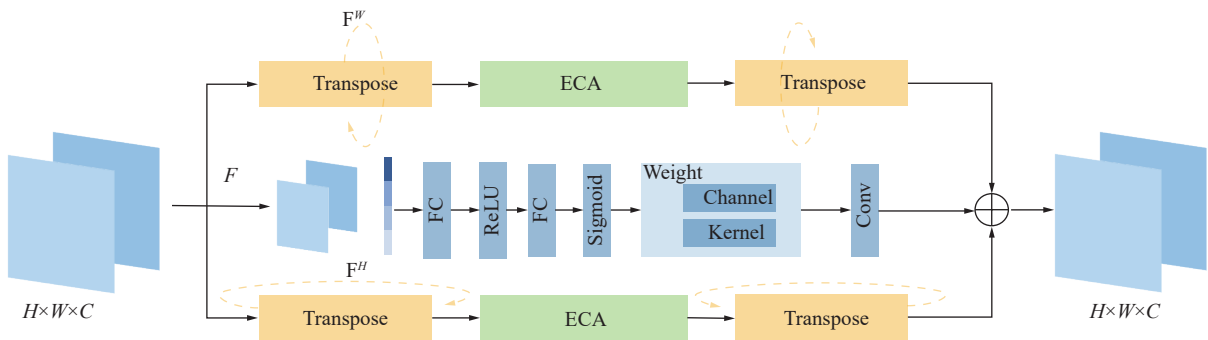


图 3 多维交互注意力模块 (MDIA)

Fig.3 Multi-Dimensional Interactive Attention module (MDIA)

中卷积核大小 $k$ 可表示为

$$k = \left\lceil \frac{\log_2 C}{\gamma} + \frac{b}{\gamma} \right\rceil \quad (10)$$

其中,  $C$  是输入通道数,  $\gamma$  和  $b$  是超参数。然后通过一维卷积来捕捉通道间依赖关系:

$$F_{out}^W = \text{Conv1D}_k(F^W) \quad (11)$$

式中:  $\text{Conv1D}_k(\cdot)$  为一维卷积操作,  $k$  为卷积核大小。

输出特征图  $F_{out}^W$  沿水平方向进行一次逆旋转, 获得与原始输入形状相同的特征图。

在网络中间的动态卷积分支中, 首先将  $F$  输入到自适应平均池化层进行池化操作, 将输入特征图的高度和宽度分别压缩为 1, 得到  $B \times C \times 1 \times 1$  大小的特征图, 接下来用  $1 \times 1$  大小的卷积核进行降维处理, 使用 ReLU 作为激活函数。然后分别对通道数、卷积核进行 Sigmoid 函数处理, 得到各自维度上的注意力权重, 通过激活程度来调整每个维度的重要性。最后, 对 3 个分支的所有输出进行平均聚合, 通过不同维度生成的关注权重重新校准, 输出最终的精细化特征图。

该模块在垂直和水平上对特征图进行维度旋转, 旋转后的特征图分别输入到不同的分支中, 并通过注意力机制调整它们之间的交互权重, 优化特征图各维度之间在空间结构上的协同作用, 在全局尺度上完成建模。同时, 考虑到卷积在局部感受野上特征提取的优势, 采用并行策略的方式引入动态卷积, 自适应地调整卷积核的大小形状, 更好地感知图像中的边缘细节和纹理信息, 有效提高超分辨率重建质量。

## 2.2 递归稀疏自注意力模块

由于 Transformer 在自然语言处理中的出色表现, 其核心机制自注意力也被广泛应用于图像处理领域。该机制通过捕获输入数据之间的交互, 对长距离依赖关系进行建模。ViT<sup>[21]</sup> (Vision Transformer, ViT) 的计算复杂度与特征块数量成平方关系, 限制了其在图像超分辨率重建中的应用。CAT<sup>[22]</sup> (Cross Attention in vision Transformer, CAT) 利用不同头部

的水平和垂直矩形窗口注意力来并行扩展注意力区域并聚合不同窗口的特征, 将 CNN 的平移不变性和局部性的归纳偏差合并到 transformer 中来捕获全局信息, 但忽略了空间依赖性的建模。Swin<sup>[23]</sup> (Shifted Windows, Swin) 将图像划分成互不重合的窗口, 然后在不同的窗口中进行自注意力计算, 其计算复杂度和图像尺寸呈线性关系, 但窗口之间的信息交互受到窗口大小和移动策略的限制, 模型对长距离依赖关系的捕捉仍有一定的局限。针对以上网络存在的问题, 笔者提出一种递归稀疏自注意力机制 (Recursive Sparse Self-Attention, RS\_SA) 网络模块 (图 4), 通过递归聚合以线性复杂度利用全局上下文, 同时引入 Top-k 选择算子来保留关键的自注意力值来增强特征聚合。首先, 通过递归循环单个深度卷积捕获全局信息, 其中循环次数  $T$  的计算方式如下:

$$T = \log_{s_r} \frac{H}{h} \quad (12)$$

其中,  $s_r$  为卷积的步幅大小,  $h$  为常数,  $H$  为图像高度。通过动态选择递归次数  $T$ , 来灵活处理不同大小输入。然后, 对获得的初始聚合映射进行细化, 使用  $3 \times 3$  深度卷积、 $1 \times 1$  多个卷积层计算获得具有代表性的映射。整体计算过程可表示为:

$$\hat{X} = f_r(f_r(\dots(f_r(X)))) \quad (13)$$

$$X_r = f_{1 \times 1} f_{3 \times 3}(\hat{X}) \quad (14)$$

其中,  $f_r$  为  $s_r$  步幅的深度卷积,  $f_{3 \times 3}$  为  $3 \times 3$  深度卷积,  $f_{1 \times 1}$  为  $1 \times 1$  逐点卷积。在标准的 Transformer 中, 自注意力输出可由式 (12) 描述:

$$\text{Att}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\lambda}\right)V \quad (15)$$

其中,  $Q$  为查询,  $K$  为键,  $V$  为值。由于标准 Transformer 基于全连接的注意范式容易导致特征交互过程中的信息冗余, 因此对上述  $X_r$  进行张量重塑 (如图 4 “reshape” 操作), 之后计算所有  $Q$  和  $K$  之间的像素对相似度, 利用转置注意矩阵  $M$  屏蔽低注意权重分配的值, 在  $M$  上自适应选择 Top-k 分数的权重值。其中,  $k$  为动态控制稀疏度大小的可调参数。只有在

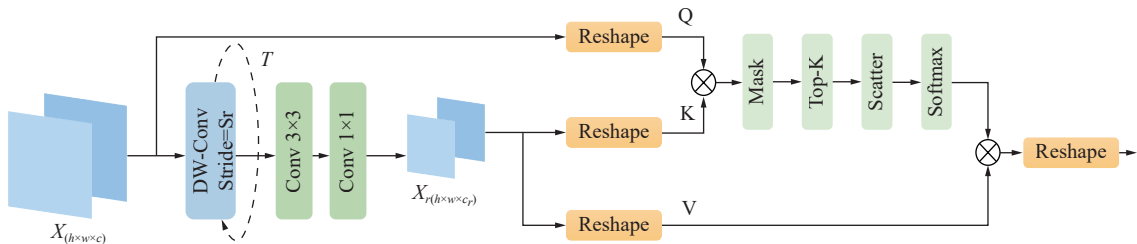


图 4 递归稀疏自注意力 (RS\_SA)

Fig.4 Recursive Sparse Self-Attention (RS\_SA)



给定阈值内的 Top-k 分数才会进行归一化计算,对于小于阈值的分数,通过离散函数 Scatter 在给定的索引上将其概率置 0。

$$SAtt(Q, K, V) = softmax\left(T_k\left(\frac{QK^T}{\lambda}\right)\right)V \quad (16)$$

其中,  $T_k$  为选择算子。最后,将每个注意力头的输出拼接在一起,利用线性投影得到矩阵计算结果。

### 3 实验条件和评价指标

#### 3.1 实验环境和数据集

笔者的实验环境如下:CPU 为 i9-12900K/64G, GPU 为 RTX4090/24G, 操作系统为 Ubuntu22.04, 编程语言为 Python3.8.18, CUDA 版本为 12.0, 深度学习框架为 Pytorch1.9.1。



图 5 矿井图像数据集部分示例  
Fig.5 Examples of coalmine image dataset

#### 3.2 评价指标

笔者实验性能评价采用峰值信噪比<sup>[28]</sup>(Peak Signal-to-Noise Ratio, PSNR)和结构相似性指数 SSIM<sup>[29]</sup>(Structure Similarity Index Measure, SSIM)来初步评估图像重建算法的效果。PSNR 以分贝(dB)为单位,其值越高表示图像质量越好。为了进一步评估 2 个图像之间的像素差异,使重建效果更加符合矿井真实场景的视觉感知。笔者在矿井数据集的实验中引入图像相似性度量标准 LPIPS<sup>[30]</sup>(Learned Perceptual Image Patch Similarity, LPIPS)和视觉信息保真度 VIF<sup>[31]</sup>(Visual Information Fidelity, VIF)。LPIPS 利用了预训练的 CNN 模型来学习感知距离,这使得它能够更好地与人类感知系统的感知一致性。取值范围为 [0,1], LPIPS 越接近 1 表明图像质量越好。VIF 是 Sheikh 等结合自然图像统计模型、图像失真模型和人眼视觉系统模型提出的图像质量评价指标。其取值范围为 [0,1], VIF 越大表明图像与主观视觉有更高的一致性。

### 4 实验结果分析

#### 4.1 客观指标分析

将文本算法在 DIV2K 上训练结果与其他超分

笔者分别使用公开数据集 DIV2K<sup>[24]</sup>和煤矿井下图像数据集进行分组实验。DIV2K 是 1 个权威的单图像超分辨率数据集,它包含 1 000 个不同场景的图像对。笔者所用矿井图像是由 KBA18-矿用本安型摄像仪在多个煤矿场景拍摄获取的,该摄像仪分辨率为 1 920×1 080,环境照度为 10 Lux,采集的图像包括井下巷道、工作面和设备间等场景。笔者使用双三次下采样来构成训练所需的图像对。矿井图像数据集部分示例如图 5 所示。

为了提升训练速度,同时减少内存占用,将训练图片裁剪为具有重叠的 480×480 的子图像块。然后提取每个图块像的尺寸和通道数存储为元信息。笔者测试集使用矿井图像测试集和基准数据集 Set5<sup>[25]</sup>、Set14<sup>[26]</sup>、Urban100<sup>[27]</sup>。

分辨率重建算法在基准数据集上进行对比实验。设置两组实验(表 1),选择与经典算法 SRCNN 及 CSR-CNN<sup>[32]</sup>;基于残差网络的 VDSR 及 EDSR 算法;基于注意力机制的 SPAN<sup>[33]</sup>、ELAN<sup>[34]</sup>、DAT<sup>[35]</sup>和 SwinIR 进行 PSNR 和 SSIM 性能比较。表 1 展示了上述方法在基准数据集 Set5、Set14、Urban100 上缩放因子为×2、×4 的客观指标对比。由表 1 数据所示,笔者算法与目前主流算法相比,在各数据集上总体表现出较优性能。在 2 倍缩放因子下,笔者算法的 PSNR 指标较次优算法 DAT 分别提升 0.09、0.51、0.27 dB,在 Set14 基准数据集上,笔者算法的 SSIM 指数略低于 DAT 算法 0.001 5,在 Set5 和 Urban100 上分别提升了 0.001 8、0.032 3。在 4 倍缩放因子下,笔者算法的 PSNR 指数在 3 个基准数据集上分别提升 0.17、0.89、0.58 dB, SSIM 指数分别提升 0.032 9、0.097 7、0.040 0。

为了验证笔者算法在矿井图像上的有效性,同样选取上述 8 个算法在井下图像数据上训练完成后进行比对。分别设置 2 倍缩放因子和 4 倍缩放因子两组实验,同时引入 LPIPS 和 VIF 两个指标来进一步探究图像之间的感知差异。本组实验分别选取综放工作面、井下水仓、变电所场景各 5 张作为测试数

据集,实验结果如表 2 所示,在 2 倍和 4 倍缩放因子下,本算法在 PSNR 和 SSIM 指标上均取得了最优结果。在 2 倍缩放因子下,笔者算法相较于次优算法

DAT 在 LPIPS 指标上降低 0.026 3(10.97%),在 PSNR 指标上提升 1.312 3 dB(4.10%)。在 4 倍缩放因子下,本算法相较于次优算法在 LPIPS 指标上降低 0.028 8

表 1  不同超分算法在 3 种基准数据集上的性能对比

Table 1  Performance comparison of different super-resolution algorithms on three benchmark datasets

| 尺度因子 | 模型     | Set5          | Set14         | Urban100      |
|------|--------|---------------|---------------|---------------|
|      |        | PSNR↑/SSIM↑   | PSNR↑/SSIM↑   | PSNR↑/SSIM↑   |
| ×2   | SRCNN  | 36.66/0.954 2 | 32.45/0.906 7 | 29.50/0.894 6 |
|      | CSRCNN | 37.45/0.957 0 | 34.34/0.924 0 | 29.88/0.902 0 |
|      | VDSR   | 37.53/0.958 7 | 33.03/0.912 4 | 30.76/0.914 0 |
|      | EDSR   | 38.11/0.960 2 | 33.92/0.919 5 | 32.93/0.935 1 |
|      | SPAN   | 38.27/0.961 4 | 34.34/0.924 0 | 33.34/0.938 4 |
|      | ELAN   | 38.11/0.960 9 | 33.82/0.919 6 | 32.62/0.932 8 |
|      | SwinIR | 38.35/0.962 0 | 34.14/0.922 7 | 33.40/0.939 3 |
|      | DAT    | 38.34/0.961 9 | 34.43/0.924 7 | 33.54/0.940 2 |
|      | Our    | 38.43/0.963 7 | 34.94/0.923 2 | 33.81/0.972 5 |
| ×4   | SRCNN  | 30.49/0.862 8 | 27.50/0.731 5 | 24.47/0.722 9 |
|      | CSRCNN | 31.01/0.870 2 | 28.47/0.772 0 | 24.62/0.728 0 |
|      | VDSR   | 31.35/0.883 8 | 28.01/0.767 4 | 25.18/0.752 4 |
|      | EDSR   | 32.46/0.896 8 | 28.80/0.787 6 | 26.64/0.803 3 |
|      | SPAN   | 32.63/0.900 2 | 28.87/0.788 9 | 26.82/0.808 7 |
|      | ELAN   | 32.47/0.898 3 | 28.81/0.786 8 | 26.60/0.801 5 |
|      | SwinIR | 32.72/0.902 1 | 28.94/0.791 4 | 27.07/0.816 4 |
|      | DAT    | 32.74/0.901 3 | 29.02/0.791 4 | 27.14/0.814 9 |
|      | Our    | 32.91/0.934 2 | 29.91/0.889 1 | 27.72/0.854 9 |

表 2  不同算法在矿井数据集上的性能对比

Table 2  Performance comparison of different algorithms on coalmine dataset

| 尺度因子 | 模型     | PSNR↑    | SSIM↑   | LPIPS↓  | VIF↑    |
|------|--------|----------|---------|---------|---------|
| ×2   | SRCNN  | 28.461 9 | 0.950 3 | 0.297 1 | 0.601 2 |
|      | CSRCNN | 30.623 8 | 0.959 6 | 0.293 4 | 0.610 3 |
|      | VDSR   | 30.640 4 | 0.966 4 | 0.283 1 | 0.626 5 |
|      | EDSR   | 30.714 8 | 0.967 5 | 0.279 9 | 0.715 4 |
|      | SPAN   | 30.836 6 | 0.970 0 | 0.277 6 | 0.751 8 |
|      | ELAN   | 31.071 9 | 0.971 3 | 0.276 8 | 0.778 9 |
|      | SwinIR | 31.808 4 | 0.975 5 | 0.265 4 | 0.832 4 |
|      | DAT    | 32.014 6 | 0.971 6 | 0.238 1 | 0.869 3 |
|      | Our    | 33.326 9 | 0.979 6 | 0.211 8 | 0.882 0 |
| ×4   | SRCNN  | 23.124 5 | 0.802 1 | 0.398 0 | 0.587 2 |
|      | CSRCNN | 23.615 3 | 0.815 3 | 0.312 8 | 0.598 1 |
|      | VDSR   | 23.982 7 | 0.827 6 | 0.313 5 | 0.631 8 |
|      | EDSR   | 24.781 4 | 0.843 2 | 0.310 6 | 0.638 7 |
|      | SPAN   | 25.475 6 | 0.857 5 | 0.308 2 | 0.680 5 |
|      | ELAN   | 26.109 2 | 0.876 2 | 0.305 4 | 0.726 4 |
|      | SwinIR | 26.724 8 | 0.874 3 | 0.293 1 | 0.770 2 |
|      | DAT    | 27.349 3 | 0.883 7 | 0.290 5 | 0.790 3 |
|      | Our    | 27.979 2 | 0.888 5 | 0.261 7 | 0.814 7 |



(9.91%), 在 VIF 指标上提升 0.022 4(3.09%)。

## 4.2 主观性能分析

为了进一步验证笔者方法的有效性, 选取 5 个

代表性算法在×4 因子下矿井图像上进行超分辨率重建结果的视觉图对比。如图 6 所示, 笔者分别在井下水仓、综放工作面、变电所 3 种场景进行主观视觉

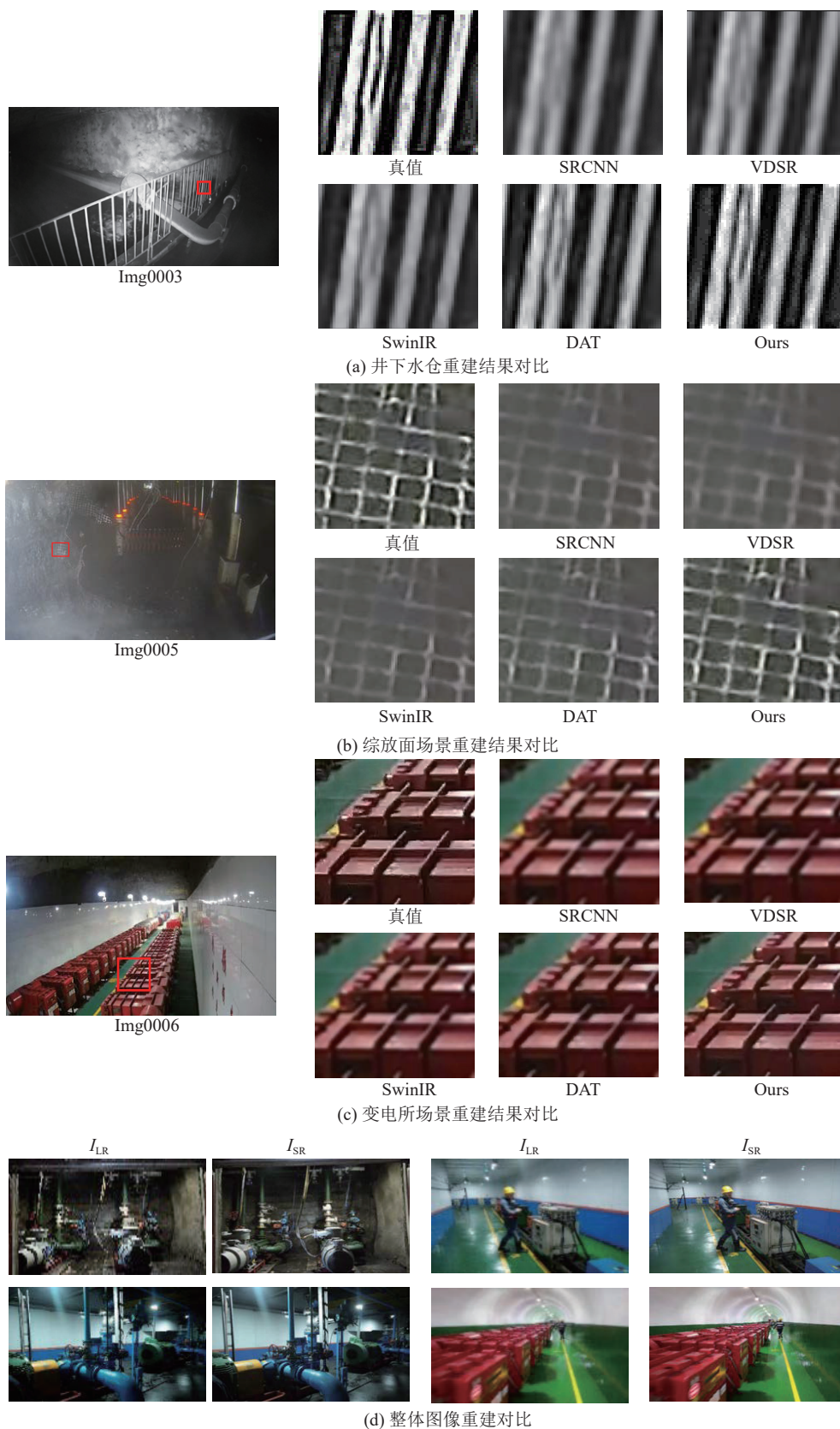


图 6 ×4 缩放因子下矿井图像的重建结果对比

Fig.6 Comparison of reconstruction results for ×4 images in coalmine dataset

对比。同时,在图 6d 中展示了本算法对低分辨率矿井图像整体重建结果的视觉对比,其中,  $I_{LR}$  为低分辨率图像,  $I_{SR}$  为超分辨率重建图像。图 6 中,截取水仓中具有明显纹理结构的栏杆、综放工作面中铁丝网、变电所中电力设备进行对比,模型 SRCNN 基本完成了图像中纹理方向的恢复,但边缘部分存在较明显的模糊, VDSR 算法的重建结果较 SRCNN 在图像左侧栏杆处的纹理层次有了进一步提升,可见引入残差结构对特征信息具有较好的提取效果。SwinIR 和 DAT 相较前者在局部结构上更加接近真值,在栏杆边缘部分整体轮廓较清晰。表明注意力机制在保持特征连续性、恢复边缘轮廓关键特征上有明显优势。笔者算法在几何结构上具有更加清晰的视觉效果,在图 6a 左侧细节密集部分较真值图像基本取得了较好的重建结果。对于图 6b 中铁丝网的图像恢复中,模型 SRCNN、VDSR、SwinIR 均未能对重建后的煤尘遮挡部分进行有效辨识。而笔者算法重建后的网面轮廓具有较高的视觉质量,对于几何结构和边缘具有显著优势。

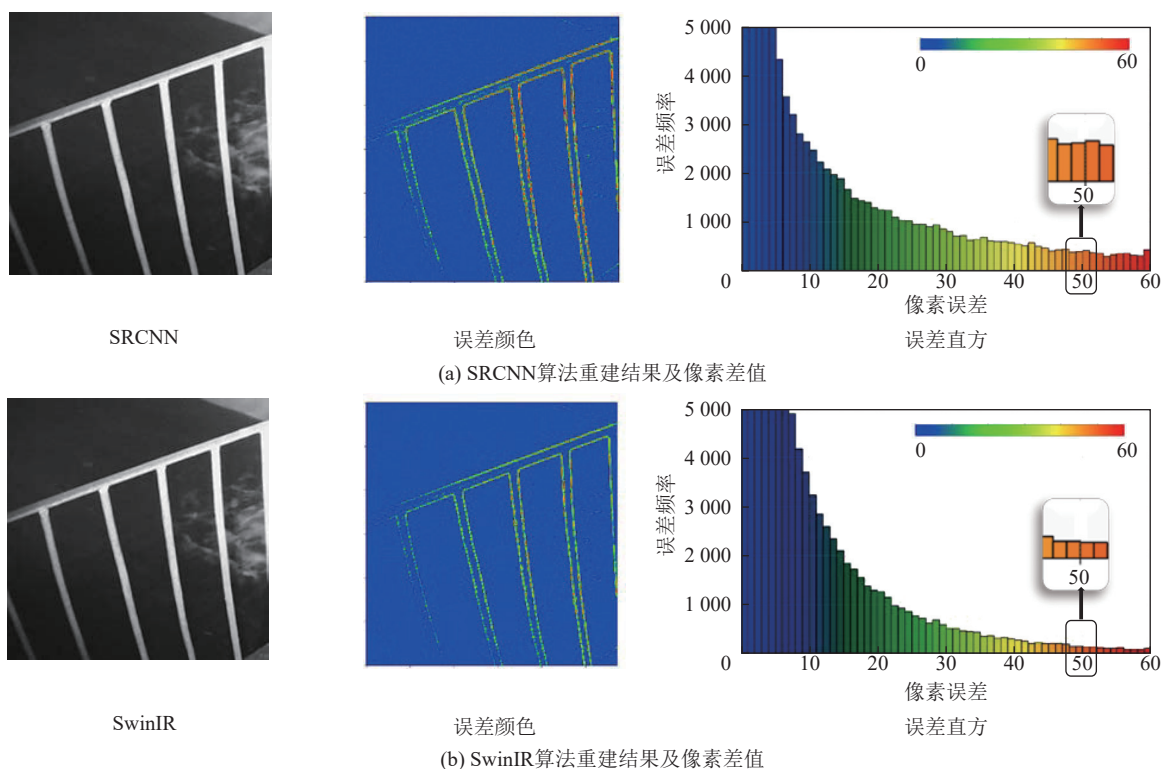
为了提供一种更加直观且易于理解的方式来展现各算法间的结果对比,基于像素差值来生成各模型输出图像与原始高分辨率图像的误差颜色图和误差直方图。如图 7 所示,使用蓝色到红色的映射来表示像素差值的大小,设定红色表示最大差值 60,颜色越趋于蓝色表示差值越小,即超分辨率重建效果

越好。在误差直方图中,横坐标表示像素误差值,同样设定相同的颜色映射。纵坐标表示误差频率,即每个差值像素数量的分布情况。选取井下水仓图像进行实验,设置经典的 SRCNN 算法作为比较基准,同时采用目前主流的 SwinIR、DAT 来验证笔者算法的优越性。由图 7 可知, SRCNN 在整体栏杆上表现出较大差值,尤其背光较暗位置的红色区域较深。SwinIR 和 DAT 整体呈现代表低差值的绿色区域,但栏杆结构连接处及轮廓部分存在明显像素差异。笔者算法输出的结果图基本为低差值颜色,在边缘位置没有较大像素差变化,对栏杆暗部区域也较好地完成了图像恢复。

为了验证超分辨率重建后的图像是否符合人类视觉系统对图像质量的感知,笔者引入平均意见分数 MOS (Mean Opinion Score, MOS)。与上述评估指标不同的是, MOS 指数不使用像素差异作为计算依赖,采用人类主观评价规律对图像清晰度、锐度、模糊和噪声等做出得分预测,从而更加准确地衡量图像质量。笔者基于 Resnet 的一个基线,在 ModelScope 平台上对 2 倍和 4 倍缩放因子下的矿井图像试验结果进行无参考评估(实验结果见表 3)其中指标分数范围为 [0,1],值越大代表图像质量越好。实验结果表明,本算法在两组实验下均取得最优结果,在 2 倍和 4 倍缩放因子下分别提升 0.023 2、0.034 0。

### 4.3 消融实验

为了进一步验证笔者算法中各模块的有效性,



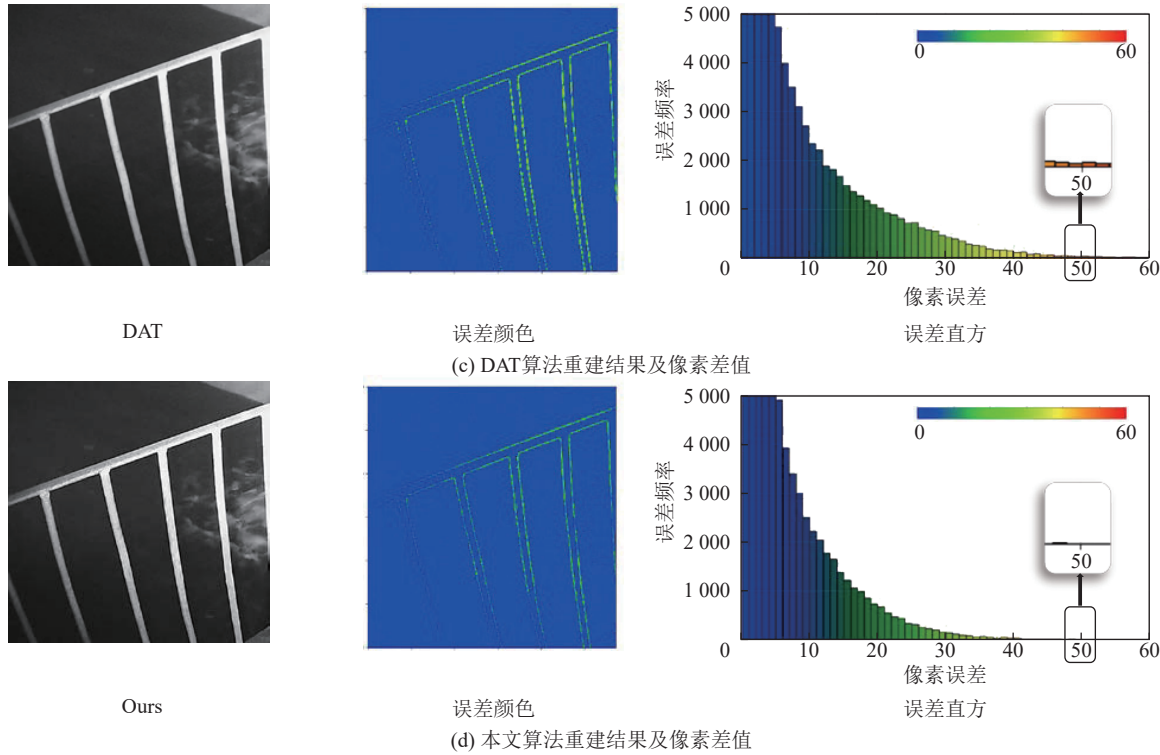


图 7 不同算法重建结果的像素差值颜色图对比

Fig.7 Comparison of pixel difference color maps of reconstruction results from different algorithms

表 3 不同算法在 MOS 评价指标上的对比

Table 3 Comparison of different algorithms on mos evaluation metrics

| 尺度因子   | SRCNN   | CSRCNN  | SPAN    | VDSR    | SwinIR  | DAT     | Our            |
|--------|---------|---------|---------|---------|---------|---------|----------------|
| ×2缩放因子 | 0.527 0 | 0.549 2 | 0.599 2 | 0.610 8 | 0.632 6 | 0.719 4 | <b>0.742 6</b> |
| ×4缩放因子 | 0.434 2 | 0.478 2 | 0.488 9 | 0.541 2 | 0.562 5 | 0.570 1 | <b>0.604 1</b> |

设置 4 组试验, 分别在矿井图像和基准数据集 Set5、Set14、Urban100 上进行×2 缩放因子下的 PSNR 和 SSIM 指标对比。其中 MSA 是多头自注意力, MDIA 是多维交互注意力, RS\_SA 为递归稀疏自注意力。表 4 中, 模型 a 为单独使用多头自注意力所训练出的模型; 模型 b 为单独使用递归稀疏自注意力训练出的模型; 模型 c 为单独使用多维交互注意力模块所训练出的模型; 模型 d 为使用多维交互注意力模块和递归稀疏自注意力模块所训练出的模型。从表 4 可以看出, 模型 b 相较模型 a 的 PSNR 值要高

0.53 dB, 说明利用递归聚合方式和算子动态选择机制在处理全局和局部特征时有更好的精度表现; 模型 d 相较模型 b 在各项数据集上均有所提升, 尤其在矿井图像上 PSNR 提升了 0.65 dB, 可见引入多维交互注意力增强了空间结构上的协同作用, 捕获了更多特征图信息; 模型 d 相较模型 c 在矿井图像上 PSNR 提升了 0.54 dB, 由此可见, 在 MDIA 和 RS\_SA 二者融合策略下, 模型获得更优的信息交互和特征聚合效果。

表 4 2 倍缩放因子下不同模块对模型性能的影响对比

Table 4 Comparison of the impact of different modules on model performance under 2x scaling factor

| 模型 | 模块  |      |       | 矿井图像                 | Set5                 | Set14                | Urban100             |
|----|-----|------|-------|----------------------|----------------------|----------------------|----------------------|
|    | MSA | MDIA | RS_SA | PSNR↑/SSIM↑          | PSNR↑/SSIM↑          | PSNR↑/SSIM↑          | PSNR↑/SSIM↑          |
| a  | √   | ×    | ×     | 32.15/0.961 0        | 37.74/0.952 5        | 33.28/0.915 6        | 32.45/0.927 8        |
| b  | ×   | ×    | √     | 32.68/0.971 2        | 37.89/0.962 1        | 34.61/0.919 8        | 33.18/0.944 7        |
| c  | ×   | √    | ×     | 32.79/0.964 3        | 38.01/0.959 7        | 34.68/0.916 9        | 33.03/0.954 1        |
| d  | ×   | √    | √     | <b>33.33/0.973 8</b> | <b>38.43/0.963 7</b> | <b>34.94/0.923 2</b> | <b>33.81/0.972 5</b> |



## 5 结 论

1)提出了一种矿井图像的多维特征与残差注意力网络超分辨率重建方法。该方法利用多维交互注意力模块来提取矿井图像中的高频信息,通过动态卷积增强矿井图像通道及卷积核各维度之间的信息整合能力。设计递归稀疏自注意力机制,实现更有效的特征聚合,以减少边缘区域的通道冗余。最后,基于标准多头自注意力模块,设计深层特征提取的基本单元,同时融合浅层特征,进行矿井图像超分辨率重建。

2)通过在煤矿井下图像及基准数据集与其他模型的大量实验对比分析,证明了笔者方法在各项评估指标上均表现出较优结果,在主观视觉分析上也具有显著优势。笔者提出的重建方法对基于矿井图像的智能识别、安全监测等视觉计算任务也具有广泛应用价值。

3)存在一些需要继续研究和探索的地方:在矿井图像数据集的构建中,笔者所采用的基于双三次下采样的矿井图像退化模型尚未考虑运动模糊及附加噪声的影响。因此,笔者后续研究工作仍需进一步对真实矿井图像退化方式进行建模,构建更加贴合实际矿井应用的超分辨率方法。

## 参考文献(References):

- [1] 王国法, 庞义辉, 任怀伟, 等. 智慧矿山系统工程及关键技术研究与实践[J]. 煤炭学报, 2024, 49(1): 181–202.  
WANG Guofa, PANG Yihui, REN Huaiwei, et al. System engineering and key technologies research and practice of smart mine[J]. Journal of China Coal Society, 2024, 49(1): 181–202.
- [2] 程健, 李昊, 马昆, 等. 矿井视觉计算体系架构与关键技术[J]. 煤炭科学技术, 2023, 51(9): 202–218.  
CHENG Jian, LI Hao, MA Kun, et al. Architecture and key technologies of coalmine underground vision computing[J]. Coal Science and Technology, 2023, 51(9): 202–218.
- [3] 程健, 陈亮, 王凯, 等. 一种多特征融合的复杂场景动态目标跟踪算法[J]. 中国矿业大学学报, 2021, 50(5): 1002–1010.  
CHENG Jian, CHEN Liang, WANG Kai, et al. Multi-feature fusion dynamic target tracking algorithm for complex scenes[J]. Journal of China University of Mining & Technology, 2021, 50(5): 1002–1010.
- [4] 张艳青, 马建红, 韩颖, 等. 真实场景下图像超分辨率重建研究综述[J]. 计算机工程与应用, 2023, 59(8): 28–40.  
ZHANG Yanqing, MA Jianhong, HAN Ying, et al. Review of research on real-world single image super-resolution reconstruction [J]. Computer Engineering and Applications, 2023, 59(8): 28–40.
- [5] IRANI M, PELEG S. Improving resolution by image registration[J]. CVGIP: Graphical Models and Image Processing, 1991, 53(3): 231–239.
- [6] 李佳星, 赵勇先, 王京华. 基于深度学习的单幅图像超分辨率重建算法综述[J]. 自动化学报, 2021, 47(10): 2341–2363.  
LI Jiaxing, ZHAO Yongxian, WANG Jinghua. A review of single image super-resolution reconstruction algorithms based on deep learning[J]. Acta Automatica Sinica, 2021, 47(10): 2341–2363.
- [7] LEPCHA D C, GOYAL B, DOGRA A, et al. Image super-resolution: a comprehensive review, recent trends, challenges and applications[J]. Information Fusion, 2023, 91: 230–260.
- [8] DONG C, LOY C C, TANG X O. Accelerating the super-resolution convolutional neural network[M]//Lecture notes in computer science. Cham: Springer International Publishing, 2016: 391–407.
- [9] KIM J, LEE J K, LEE K M. Deeply-recursive convolutional network for image super-resolution[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA. IEEE, 2016: 1637–1645.
- [10] KIM J, LEE J K, LEE K M. Accurate image super-resolution using very deep convolutional networks[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA. IEEE, 2016: 1646–1654.
- [11] 程德强, 郭昕, 陈亮亮, 等. 多通道递归残差网络的图像超分辨率重建[J]. 中国图象图形学报, 2021, 26(3): 605–618.  
CHENG Deqiang, GUO Xin, CHEN Liangliang, et al. Image super-resolution reconstruction from multi-channel recursive residual network[J]. Journal of Image and Graphics, 2021, 26(3): 605–618.
- [12] 陈亮亮. 光照不均匀场景单幅图像超分辨率重建方法研究[D]. 徐州: 中国矿业大学, 2022.  
CHEN Liangliang. Research on super-resolution reconstruction method of single image in uneven illumination scene[D]. Xuzhou: China University of Mining and Technology, 2022.
- [13] 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, HI, USA. IEEE, 2017: 1132–1140.
- [14] 杨宏业, 赵银娣, 董霁红. 基于纹理转移的露天矿区遥感图像超分辨率重建[J]. 煤炭学报, 2019, 44(12): 3781–3789.  
YANG Hongye, ZHAO Yindi, DONG Jihong. Remote sensing image super-resolution of open-pit mining area based on texture transfer[J]. Journal of China Coal Society, 2019, 44(12): 3781–3789.
- [15] WANG X T, XIE L B, DONG C, et al. Real-ESRGAN: training real-world blind super-resolution with pure synthetic data[C]//2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Montreal, BC, Canada. IEEE, 2021: 1905–1914.
- [16] 田子建, 吴佳奇, 张文琪, 等. 基于 Transformer 和自适应特征融合的矿井低照度图像亮度提升和细节增强方法[J]. 煤炭科学技术, 2024, 52(1): 297–310.  
TIAN Zijian, WU Jiaqi, ZHANG Wenqi, et al. An illuminance improvement and details enhancement method on coal mine low-light images based on Transformer and adaptive feature fusion[J]. Coal Science and Technology, 2024, 52(1): 297–310.



- [17] DAI T, CAI J R, ZHANG Y B, et al. Second-order attention network for single image super-resolution[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA. IEEE, 2019: 11057–11066.
- [18] LIANG J Y, CAO J Z, SUN G L, et al. SwinIR: image restoration using swin transformer[C]//2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Montreal, BC, Canada. IEEE, 2021: 1833–1844.
- [19] 王满利, 张航, 李佳悦, 等. 基于深度神经网络的煤矿井下低光照图像增强算法[J]. *煤炭科学技术*, 2023, 51(9): 231–241.  
WANG Manli, ZHANG Hang, LI Jiayue, et al. Deep neural network-based image enhancement algorithm for low-illumination images underground coal mines[J]. *Coal Science and Technology*, 2023, 51(9): 231–241.
- [20] WANG Q L, WU B G, ZHU P F, et al. ECA-net: efficient channel attention for deep convolutional neural networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA. IEEE, 2020: 11531–11539.
- [21] HAN K, WANG Y H, CHEN H T, et al. A survey on vision transformer[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(1): 87–110.
- [22] LIN H Z, CHENG X, WU X Y, et al. CAT: cross attention in vision transformer[C]//2022 IEEE International Conference on Multimedia and Expo (ICME). Taipei, China. IEEE, 2022: 1–6.
- [23] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada. IEEE, 2021: 9992–10002.
- [24] LI Y W, ZHANG Y L, TIMOFTE R, et al. NTIRE 2023 challenge on efficient super-resolution: methods and results[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Vancouver, BC, Canada. IEEE, 2023: 1922–1960.
- [25] BEVILACQUA M, ROUMY A, GUILLEMOT C, et al. Low-complexity single-image super-resolution based on nonnegative neighbor embedding[C]//Proceedings of the British Machine Vision Conference 2012. Surrey. British Machine Vision Association, 2012.
- [26] ZEYDE R, ELAD M, PROTTER M. On single image scale-up using sparse-representations[M]//BOISSONNAT J D, CHENIN P, COHEN A, et al, eds. Lecture notes in computer science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012: 711–730.
- [27] HUANG J B, SINGH A, AHUJA N. Single image super-resolution from transformed self-exemplars[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA. IEEE, 2015: 5197–5206.
- [28] SETIADI D R I M. PSNR vs SSIM: imperceptibility quality assessment for image steganography[J]. *Multimedia Tools and Applications*, 2021, 80(6): 8423–8444.
- [29] SARA U, AKTER M, UDDIN M S. Image quality assessment through FSIM, SSIM, MSE and PSNR—a comparative study[J]. *Journal of Computer and Communications*, 2019, 7(3): 8–18.
- [30] ZHANG R, ISOLA P, EFROS A A, et al. The unreasonable effectiveness of deep features as a perceptual metric[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. IEEE, 2018: 586–595.
- [31] KUO T Y, SU P C, TSAI C M. Improved visual information fidelity based on sensitivity characteristics of digital images[J]. *Journal of Visual Communication and Image Representation*, 2016, 40: 76–84.
- [32] ZHANG J W, WANG Z X, ZHENG Y H, et al. Cascaded convolutional neural network for image super-resolution[M]//Communications in computer and information science. Cham: Springer International Publishing, 2021: 361–373.
- [33] WAN C, YU H Y, LI Z Q, et al. Swift parameter-free attention network for efficient super-resolution[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, WA, USA. IEEE, 2024: 6246–6256.
- [34] ZHANG X D, ZENG H, GUO S, et al. Efficient long-range attention network for Image super-resolution[M]//Lecture notes in computer science. Cham: Springer Nature Switzerland, 2022: 649–667.
- [35] CHEN Z, ZHANG Y L, GU J J, et al. Dual aggregation transformer for image super-resolution[C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France. IEEE, 2023: 12278–12287.