

融合密集连接与多注意力机制的星载红外遥感图像超分辨率网络

徐新昊¹, 王 俊^{1,2}, 王 峰^{1*}, 孙胜利²

(1. 复旦大学信息科学与工程学院 电磁波信息科学教育部重点实验室, 上海 200433;

2. 中国科学院上海技术物理研究所, 上海 200083)

摘要: 星载红外遥感图像在环境监测和军事侦察等领域具有重要应用价值。然而, 由于器件工艺限制、大气扰动和传感器噪声等因素, 这类图像面临分辨率不足和细节纹理模糊的问题, 严重限制了后续分析和处理的准确性。为了克服这些困难, 提出了一种新的超分辨率生成对抗网络模型 (Generative adversarial network, GAN), 该模型将密集连接与 Swin Transformer 架构相融合, 实现跨层特征传递和上下文信息的有效利用, 同时增强了模型的全局特征提取能力。此外, 将传统的跳跃连接改进为基于多尺度通道注意力机制 (Multi-scale channel attention, MSCA) 的特征融合, 使网络能够更加灵活地整合多尺度特征, 提升了特征融合的质量和效率。随后, 通过构建联合损失函数, 以全面优化生成器的性能, 进一步提升超分辨率图像质量。所提算法在不同数据集上进行对比测试均取得了明显提升, 且超分后的图像在目标检测等下游任务中, 也展现出更高的性能, 验证了该算法在星载红外遥感图像超分辨率中的有效性和应用潜力。

关键词: 星载红外遥感; 超分辨率重建; 注意力机制; 生成对抗网络; 联合损失函数

中图分类号: O43

文献标识码: A

Infrared remote sensing image super-resolution network by integration of dense connection and multi-attention mechanism

XU Xin-hao¹, WANG Jun^{1,2}, WANG Feng^{1*}, SUN Sheng-li²

(1. Key Laboratory for Information Science of Electromagnetic Waves (MoE), School of Information Science and Technology, Fudan University, Shanghai, 200433, China;

2. Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai, 200083, China)

Abstract: Space-borne infrared remote sensing images have crucial application value in the fields of environmental monitoring and military reconnaissance. Nonetheless, due to limitations in technologies, atmospheric disturbances, and sensor noise, these images suffer from insufficient resolution and blurred texture details, severely restricting the accuracy of subsequent analysis and processing. To address these issues, a new super-resolution generative adversarial network model is proposed. This model integrates dense connections with the Swin Transformer architecture to achieve effective cross-layer feature transmission and contextual information utilization while enhancing the model's global feature extraction capabilities. Furthermore, we improve the traditional residual connections with multi-scale channel attention-based feature fusion, allowing the network to more flexibly integrate multi-scale features, thereby enhancing the quality and efficiency of feature fusion. A combined loss function is constructed to comprehensively optimize the performance of the generator. Comparative tests on different datasets demonstrate significant improvements with the proposed algorithm. Additionally, the super-resolved images exhibit higher performance in downstream tasks such as object detection, confirming the effectiveness and application potential of the algorithm in space-borne infrared remote sensing image super-resolution.

Key words: space-borne infrared remote sensing, super-resolution, attention mechanism, generative adversarial network, joint loss

基金项目: 国家自然科学基金 (61991421)

作者简介 (Biography): 徐新昊 (2000—), 男, 河南郑州人, 硕士研究生, 主要研究领域为星载红外图像超分辨率、目标检测跟踪, E-mail: xuxh22@m.fudan.edu.cn

* 通讯作者 (Corresponding author): E-mail: fengwang@fudan.edu.cn

PACS:42. 55. Px, 78. 55. Cr, 78. 67. De, 68. 65. Fg

引言

红外遥感技术利用目标自身的热辐射特性进行探测,因而不易受到外部电磁干扰,即使在恶劣的电磁环境下也能保持稳定的工作性能^[1]。作为一种被动遥感技术,红外遥感的隐蔽性强,其探测行为不易被察觉,为军事侦察和安全监控提供了一种有效的手段。红外辐射的强度与物体的温度密切相关,赋予了红外遥感全天候工作的能力和一定的穿透能力。这些独特的优势使得红外遥感技术在环境监测^[2]、灾害预警^[3]、军事侦察^[4]领域得到了广泛的应用。但是由于红外衍射限、传感器工艺与硬件系统性能等因素,使得星载红外遥感图像与可见光遥感图像相比存在空间分辨率低,图像纹理细节模糊等问题,限制了目标检测与跟踪等任务的精度。为此需要进行图像超分辨率重建(super-resolution, SR),以提高星载红外遥感图像的分辨率,恢复出更多的细节纹理信息。

图像超分辨率算法是一种通过软件算法将低分辨率(Low Resolution, LR)图像重建为高分辨率(High Resolution, HR)图像的技术^[5],目前图像超分辨率算法主要有三种^[6]:基于插值的超分辨率算法、基于重建的超分辨率算法和基于学习的超分辨率算法,插值方法简单快速,不涉及复杂计算,但其结果可能过于锐利容易出现锯齿^[7],缺乏真实图像的平滑性^[8];基于重建的超分辨率方法通过从输入图像中提取所有有用的信息,将整个超分辨率过程视为信息提取和信息融合的过程。通过利用输入图像的先验知识和全局重构约束作为正则项,构建代价函数并求解逆问题^[9]。这种方法能够保持图像边缘的清晰度,但在恢复纹理细节方面效果有限。基于学习的方法利用机器学习技术训练先验模型,以填补低分辨率图像中缺失的高频信息,从而生成丰富的高频纹理细节^[10]。然而,这种方法可能会生成虚假信息造成图像失真。Dong 等人^[11]首次将卷积神经网络(Convolutional neural network, CNN)引入图像超分辨率重建领域,提出 SRCNN 算法,使得深度学习在该领域的应用越来越广泛。此后,研究人员设计出超分辨率 GAN 网络(SRGAN)并证实基于 GAN 网络的超分方法具有更高的感知质量^[12]。

红外图像超分辨率技术沿用了可见光图像领域的处理方法,但由于成像系统的不同,退化噪声

的特性也有所差异,且红外图像通常呈现单一颜色,缺乏显著的梯度和高频先验信息^[13],这导致常规可见光图像超分辨率算法的效果难以满足红外图像的实际工程应用需求。因此,需要针对红外图像的特定场景进行算法设计,以提升超分辨率效果。Choi 等人^[14]假设红外图像与对应的可见光图像有相似的高频信息,直接估计与可见光图像边缘信息相对应的红外图像边缘的高频部分并构建包含重建约束和正则化约束的目标函数,用于改善高频图像中的弱纹理区域并减少异常值的影响。但该方法依赖先验信息且需要高质量的图像数据对,未能充分利用红外图像特征。文献[15]提出了一种基于区域显著度的红外图像超分辨率重建方法,改进了稀疏编码的特征提取过程,并构建字典进行特征筛选。通过计算稀疏特征来确定显著性区域并进行字典重建。然而,显著区域的选择依赖于显著性检测,其准确性会直接影响超分辨率的效果。Liu 等^[16]将红外梯度先验信息加入,并利用相同分辨率的同场景可见光图像进行超分辨率重建,但是红外图像和可见光图像存在域间差异。为适应红外图像的特点,文献[17]简化了 SRGAN 的模型结构,进行对抗训练实现单帧红外图像超分辨率。Huang 等人^[18]提出了一种基于迁移学习的超分 GAN 网络,能够弥合不同特征空间之间的差距,提高小样本红外图像的超分辨率重建质量,该方法依赖于小样本数据集进行训练,导致在面对多样化且复杂的红外场景时泛化性受限,仅通过峰值信噪比(Peak signal-to-noise ratio, PSNR)与结构相似度(Structural similarity index, SSIM)难以反应图像实际应用效果。Zou 等人^[19]提出了一种基于跳跃连接的卷积神经网络,用于红外图像的超分辨率重建。该网络通过卷积层和反卷积层来提取图像特征并重建红外图像。由于网络深度较浅,学习到的特征较为单一,难以提取红外图像的深层次信息。为了有效提取红外特征,[20]提出了一种多级跳跃连接的深度残差卷积网络,并利用拼接的方法进行局部特征融合增强效果,然而,简单地拼接特征可能难以突出关键特征,从而限制了超分辨率结果的优化。尽管 CNN 在图像超分辨率领域取得了显著进展,但它们通常面临局部感受野的限制。这意味着网络可能无法充分捕捉图像的全局上下文信息。在重

建过程中,虽然在 PSNR 等像素级评价指标上取得了较好的成绩^[12],但实际超分辨率图像却显得相对平滑,难以满足实际工程应用的需求。相比之下,Transformer^[21]具备全局上下文感知能力,能够提取全局特征,并根据长距离依赖关系更好地恢复图像细节^[22],因此得到了广泛的应用,但是 Transformer 进行特征提取时自注意力机制将每个 Token 视作同等重要的信息,需处理大量冗余特征,影响模型性能^[23]。

为解决上述问题,本文提出了一种新型超分辨率 GAN 网络,用于星载红外遥感图像的超分辨率重建。生成器基于密集残差连接^[28]与 Swin Transformer 架构^[24]设计,充分利用其大感受野和自注意力机制捕捉图像的全局依赖关系,并实现特征的跨层传递,更好地理解图像结构。针对特定的图像超分任务,设计多尺度通道注意力特征融合模块,用其取代传统的跳跃连接,嵌入进注意力特征提取模块,增强有用特征的影响,提升模型表征能力。为避免单一像素损失函数导致超分图像过于平滑和失真,构建了新的联合损失函数,并根据 GAN 网络的特定将不同类型的损失函数整合到统一框架中。该联合损失函数考虑了模型训练稳定性、感知质量以及对异常值的鲁棒性,有效地指导模型学习并恢复图像的重要纹理细节。为验证图像超分辨率结果的有效性,除了客观的图像评估指标外,还结合了目标检测等下游任务,评估图像超分辨率在实际应用中的性能表现。从不同角度证明了该网络对星载红外遥感图像的特征提取和细节恢复的有效性和

准确性。

1 本文方法

由于星载红外遥感图像包含的纹理细节少且分辨率低,为了确保超分辨率图像的质量,网络需要学习到图像中足够的信息,即使少量信息的丢失也可能导致重建质量低下。因此在生成器结构中,本文融合了密集残差连接与 Swin Transformer 设计出深度注意力特征提取模块,不仅能够有效提取不同尺度的图像特征,增强梯度传递,而且能够捕捉图像特征的长距离依赖关系避免关键特征的丢失,在特征融合方面,由于浅层特征与深层特征的语义和尺度信息的不一致,单一的线性组合难以有效利用关键特征,通过注意力机制沿特征通道维聚合多尺度上下文信息,并进行动态加权以突出重要特征,抑制不必要的特征。最后,设计 VGGNet^[33]结构的鉴别器,共同组成 GAN 网络用于星载红外遥感图像的超分辨率。

1.1 融合密集连接与多注意力机制的 GAN 网络架构

该网络生成器利用 Swin Transformer 模块构建密集连接(Dense block),设计深度注意力特征提取模块(Attention Feature Extraction Block, AFEB),使其具备更强的捕捉全局依赖关系和理解图像结构的能力。

如图 1 所示,网络结构主要由生成器和鉴别器组成,生成器包含浅层特征提取、基于 AFEB 的深层特征提取以及图像上采样三个部分。具体来说,生成器首先对输入的 LR 图像 $I_{LR} \in R^{H \times W \times C_{in}}$ 利用 3×3

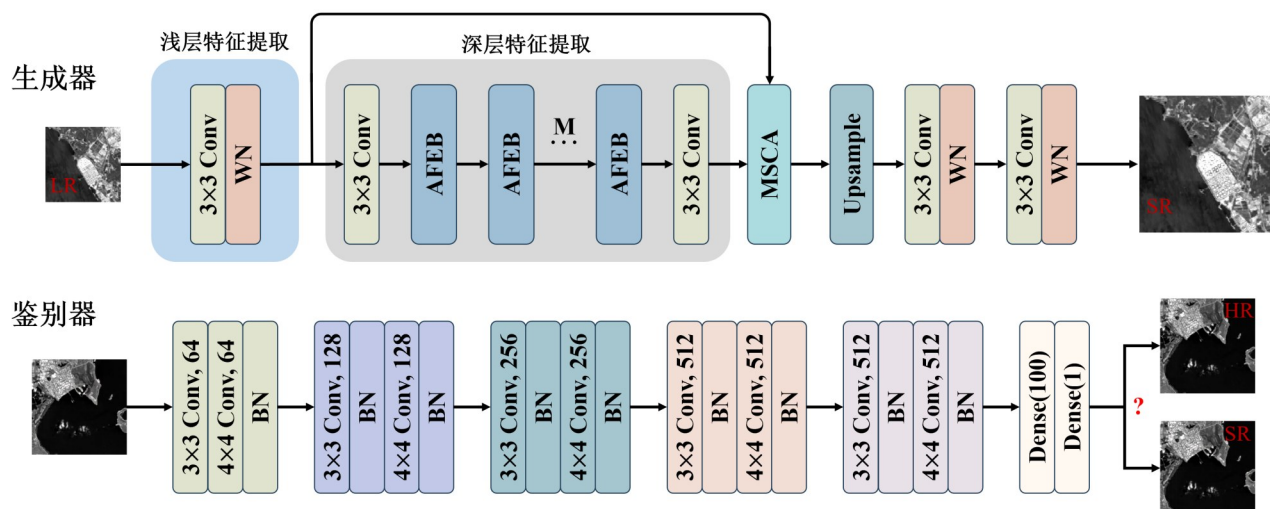


图1 网络结构

Fig. 1 Original image block tensor construction

卷积 $H_{SF}(\cdot)$ 进行浅层特征 $F_0 \in R^{H \times W \times C}$ 的提取,

$$F_0 = H_{SF}(I_{LR}), \quad (1)$$

其中, H, W, C^{in} 表示输入图像的尺寸, C 表示输出浅层特征图的通道数, 通过浅层特征提取将输入图像从灰度空间映射到高维特征空间, 提高视觉表示的抽象性和可学习性, 并在卷积层后引入权值归一化层 (Weight Normalization, WN)^[25] 以稳定整个生成器的训练。随后通过 $H_{DF}(\cdot)$ 进行深层特征 $F_{DF} \in R^{H \times W \times C}$ 提取,

$$F_{DF} = H_{DF}(F_0), \quad (2)$$

$H_{DF}(\cdot)$ 表示深层特征提取, 其中包含了 M 个注意力特征提取模块 (Attention feature extraction block, AFEB), 每个 AFEB 模块包括 N 个 Swin Transformer 模块和三个 3×3 的卷积层, 其中 M 和 N 为可调控的网络超参数。经过 AFEB 特征提取后, 通过一个 3×3 的卷积层 H_{conv} 进行深度特征信息的聚合。因此, 每一个中间特征可以分别并表示为 F_1, F_2, \dots, F_M , 经过深度特征提取器后的 F_{DF} 可以表示为,

$$\begin{aligned} F_i &= H_{AFEB}(F_{i-1}), i = 1, 2, \dots, M \\ F_{DF} &= H_{conv}(F_M) \end{aligned} \quad (3)$$

经过特征提取后, 深层特征与浅层特征一并输入到多尺度通道注意力特征融合模块 (Multi-scale channel attention, MSCA)。该模块沿通道维度对多尺度特征信息进行聚合, 并通过自适应动态加权突出重要特征。最后, 通过图像上采样模块, 实现高分辨率星载红外遥感图像的重构。

$$I_{SR} = H_{SR}(H_{MSCA}(F_0, F_{DF})), \quad (4)$$

其中, $H_{MSCA}(\cdot)$ 表示利用注意力机制融合模块, $H_{SR}(\cdot)$ 包含图像上采样模块和特征映射模块, 具体地, 将融合后的特征图输入包含最邻近插值上采样

和一个卷积层的上采样模块, 输出的上采样特征图经过两个卷积层映射生成 SR 图像。

本文基于 VGGNet^[33] 结构设计网络鉴别器, 旨在区分真实的高分辨率红外图像与生成的超分辨率红外图像。该鉴别器将 3×3 步长为 1 的卷积核与 4×4 步长为 2 的卷积核相结合, 采用 Leaky ReLU 作为激活函数, 对样本进行特征提取与下采样。输入图像经过四次组合卷积处理后, 利用两个全连接层输出图像的判别结果。在 GAN 网络的整个训练过程中交替训练生成器和鉴别器, 生成器不断优化超分辨率图像的生成性能, 以产生更接近真实高分辨图像的结果, 并进一步促使鉴别器提高判别能力, 以更准确地区分真实与生成图像。最终, 在对抗过程中促使 GAN 网络实现星载红外遥感图像的超分辨率。

1.2 深度注意力特征提取模块

在星载红外遥感图像的超分辨率研究中, CNN 主要关注局部特征, 虽然在某些情况下能够有效处理图像细节, 但对于全局特征和长距离依赖关系的捕捉能力较弱^[26], 在图像的重建中, 全局信息能够提供图像中不同区域之间的相互关系, 确保超分辨率图像的整体一致性和结构完整性, 避免仅依赖局部信息可能导致伪影和不一致性, 从而进一步提升重建质量^[27]。相比于 CNN, Swin Transformer 结合了 Transformer 和 CNN 的优势。由于局部注意力机制, 它在处理大图像时受益于 CNN。同时, 通过利用滑动窗口架构, 使其可以长距离依赖建模和全局特征提取方面发挥 Transformer 的优势。

因此本文利用密集连接方法^[28]与 Swin Transformer Block (STB) 相结合, 构建注意力特征提取模块 AFEB, 其结构如图 2 所示。

以 6 个 STB 模块为例进行密集连接, 利用 STB

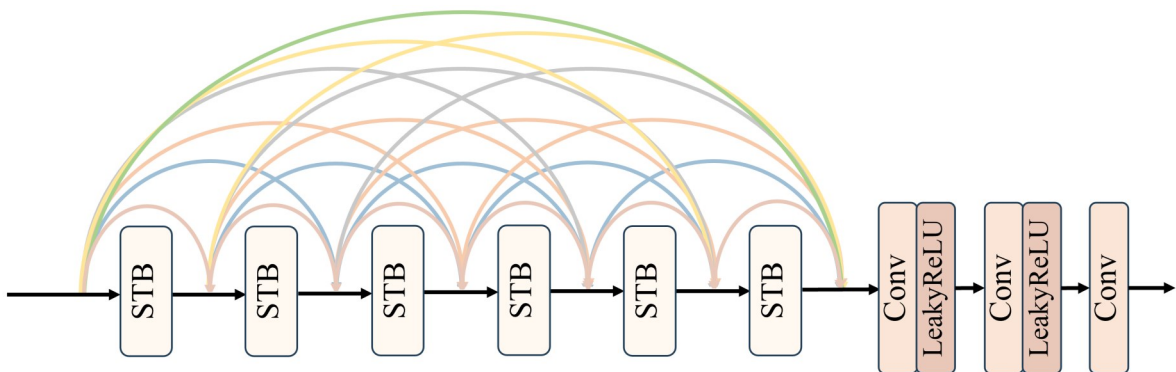


图2 注意力特征提取模块 AFEB

Fig. 2 Attention feature extract block (AFEB)

来处理不同尺度的图像信息,凭借密集连接实现多层次特征的有效传递,允许每个STB模块的输出与之前所有层的输出相结合,有助于梯度传播,减少梯度消失或爆炸的问题,并在网络中实现更丰富的特征表示,减少信息的丢失,使网络不仅能够关注局部特征,还能够在图像全局范围内捕捉和融合关键信息,从而提升图像质量。

图3给出了STB结构,将特征图分割成多个不重叠的窗口,通过窗口化多头注意力机制(Windows Multi-head self-attention, W-MHSA)来建立窗口内元素的依赖关系。具体地,每个子特征图 $feature^{h \times w \times d}$ 前两维被展平得到特征矩阵,计算查询矩阵 Q 、键矩阵 K 和值矩阵 V 。

$$\begin{cases} Q = XW^Q \\ K = XW^K, \\ V = XW^V \end{cases} \quad (5)$$

其中, $X \in R^{(h \times w) \times d}$ 表示每个子特征图的嵌入矩阵, h 和 w 表示原始子特征图的高和宽, $(h \times w)$ 表示二者的乘积, d 作为特征通道数, $W^Q \in R^{d \times d_k}$ 、 $W^K \in R^{d \times d_k}$ 和 $W^V \in R^{d \times d_v}$ 为模型的可学习权重矩阵,这些矩阵在模型的训练过程中逐渐优化,以便更好的捕捉图像信息。通过 Q 和 K 点乘对不同像素间的相互依赖关系进行衡量。利用 softmax 函数将相互依赖关系转换为注意力权重后与值矩阵 V 点乘,得到最终注意力头的全局信息,其表达式如下

$$\text{head}_i(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (6)$$

其中 d_k 表示可学习权重矩阵的列数, $\sqrt{d_k}$ 作为尺度因子防止点乘结果过大而影响梯度传递, head_i 表示第 i 个注意力头的全局信息。

最后,所有注意力头的输出被按列拼接在一起,并通过一个线性层投影来进行全局信息整合。

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W_o, \quad (7)$$

其中, h 表示注意力头的数量, $\text{Concat}(\cdot)$ 表示对矩

阵进行拼接操作, $W_o \in R^{(d_k \times h) \times (h \times w)}$ 是可训练权重矩阵。

这种窗口化的自注意力机制能够显著减少计算复杂度和内存消耗^[24]。同时,引入了跨窗口连接机制,让不同窗口之间加以传递信息,使其能够处理较大的图像尺寸并捕捉到更大范围的依赖关系。每个STB模块包含一个多层感知机(Multi-layer perceptron, MLP)^[29], 本文将该结构设计为包含两个全连接层,并使用 ReLU 作为激活函数。在整个前向传播过程中,输入特征的维度为网络训练时设置的STB模块的深层特征嵌入维度(Embedded Dimension)。以图3中将深层特征嵌入维度设置为180为例,输入首先经过第一个全连接层,特征维度映射为360,经过激活函数后特征维度保持不变。随后,通过第二个全连接层将特征维度投影回180,使得输出张量的形状与输入的特征维度保持一致,MLP不改变特征图的尺度大小。

1.3 多尺度通道注意力特征融合

在图像超分辨率领域中,浅层特征中包含了图像的边缘和纹理等信息,有助于细节还原。深层特征可以反应图像的全局信息和上下文关系,对于理解图像整体结构具有重要意义。因此,将深层特征与浅层特征融合可以提升模型的表达能力和性能。而简单的拼接和相加进行特征融合可能会导致特征尺度不匹配,难以有效区分不同特征的重要性,导致重要特征信息作用降低,使得融合效果不理想,影响重建效果。

文献[30]证明注意力机制的引入,自适应地对多尺度特征和通道特征进行加权融合,能够有效地突出重要特征、抑制冗余信息,提升特征表达能力。本文针对超分辨率重建任务,移除了[30]中原有的批归一化(Batch Normalization, BN)层,设计了新的通道注意力机制特征融合模型 MSCA,提高模型训练的稳定性,避免伪影的产生^[31],同时增强了网络泛化能力。

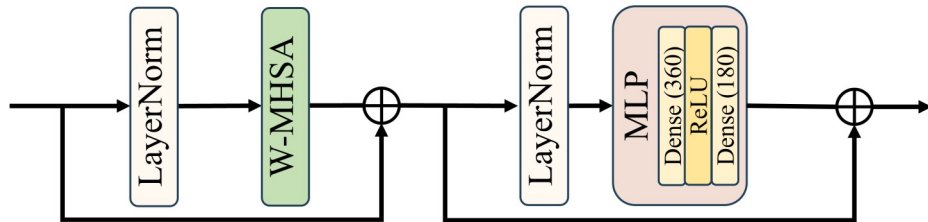


图3 STB结构

Fig. 3 Structure of Swin Transformer block (STB)

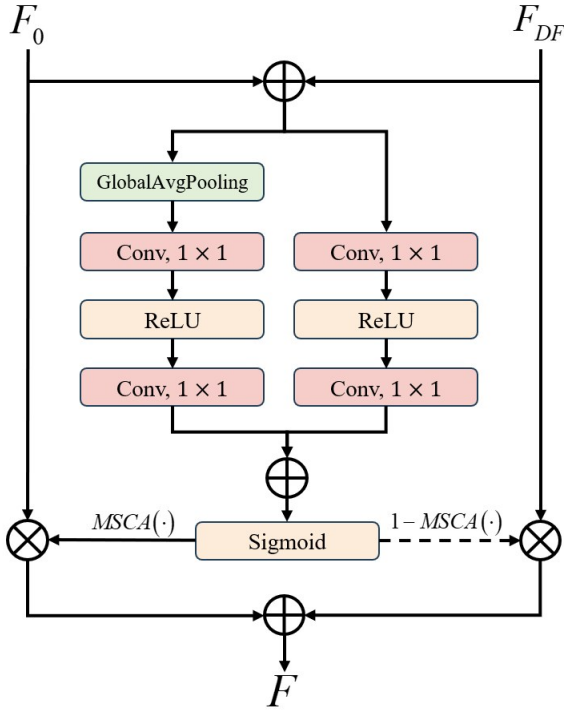


图4 多尺度通道注意力机制 MSCA

Fig. 4 Structure of Multi-scale channel attention (MSCA)

MSCA 模块在通道维度上聚合多尺度上下文信息,通过逐点卷积聚焦输入特征的细节,完成局部特征表达,并通过自适应平均池化提取全局上下文信息,捕捉全局特征。在前向传播中,这两个注意力特征相加后,通过 sigmoid 激活生成的权重对输入特征进行加权融合。

针对超分辨率重建任务设计的多尺度通道注意力特征融合模块的结构如图 4 所示,其中 $F_0, F_{DF} \in R^{H \times W \times C}$, $F \in R^{H \times W \times C}$ 表示融合后的特征。图 4 可以表达为:

$$H_{MSCA}(F_0, F_{DF}) = MSCA(F_0, F_{DF}) \otimes F_0 + (1 - MSCA(F_0, F_{DF})) \otimes F_{DF}, \quad (8)$$

$$F = H_{MSCA}(F_0, F_{DF}), \quad (9)$$

其中, \oplus 表示按张量元素逐项相加,经过 sigmoid 函数之后,融合权重 $MSCA(F_0, F_{DF})$ 作为图 4 中指向左侧 \otimes 的黑色实线,图 4 中指向右侧 \otimes 的黑色虚线代表 $1 - MSCA(F_0, F_{DF})$,权重均是元素介于 0 和 1 之间的张量,这使得网络能够在浅层特征 F_0 和深层特征 F_{DF} 之间进行加权平均。

1.4 损失函数

在 GAN 网络中,损失函数常包括,像素损失、感知损失和对抗损失,模型的训练过程也是对损失函数的优化过程,如今,很多图像超分辨率算法都倾向于

使用优化平均绝对误差 (Mean absolute error, MAE) 的 l_1 损失进行训练^{[22][23][29]},这样训练通常会取得较为可观的峰值信噪比 (Peak Signal-to-Noise Ratio, PSNR)^[17],因为 PSNR 与 MAE 是密切相关的。然而,这样的像素损失函数只关注图像像素级别的相似度而忽略了图像的感知质量和结构信息。

为此,本文利用 Charbonnier 损失和多尺度结构相似性 (Multi-scale Structural Similarity, MS-SSIM) 损失构建联合损失函数 (Joint loss)。相比于传统的 l_1 损失,Charbonnier 损失能够更好地处理异常值并保持图像亮度和颜色信息,在接近零值时表现为类似于 l_2 损失,梯度变化更为平滑,有助于训练过程的稳定性和收敛性。MS-SSIM 损失有助于恢复图像的高频信息,提升图像质量。因此,联合损失函数不仅能够保留图像细节,还能提升整体视觉效果。

$$L^{Joint} = \alpha L^{MS-SSIM} + (1 - \alpha) L^{char}, \quad (10)$$

$$L^{MS-SSIM} = 1 - \sum_{i=1}^S \omega_i \cdot SSIM_i(I^{HR}, I^{SR}), \quad (11)$$

$$L^{char} = \frac{1}{H \times W} \sum_i \sum_j \sqrt{(I_{ij}^{HR} - I_{ij}^{SR})^2 + \varepsilon^2}, \quad (12)$$

其中, α 为联合损失函数的权重, $L^{MS-SSIM}$ 和 L^{char} 分别对应多尺度结构相似度损失和 Charbonnier 损失, I^{HR} 表示高分辨率图像, I^{SR} 表示超分辨率图像, S 为不同尺度的个数, ω_i 则是对应尺度的权重, $SSIM_i$ 是对应尺度的结构相似度, ε 作为常量。[32] 中研究了不同损失函数项对于图像质量指标的监督效应,并提供了最佳参数设置方案,因此本文采用其中对于 $L^{MS-SSIM}$ 的权重设置,即 $\alpha = 0.84$, $\omega = [0.5, 1, 2, 4, 8]$ 。

感知损失主要关注 SR 图像的感知质量,使用预训练网络 VGG-19^[33] 的卷积层提取图像的纹理和结构信息,本文选择输出第四个卷积层后对应的特征图加以计算,

$$L^{Perceptual} = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W (F(I^{HR}) - F(I^{SR}))^2, \quad (13)$$

式中, $L^{Perceptual}$ 表示感知损失, $F(\cdot)$ 表示使用预训练网络 VGG-19 进行特征提取并输出的特征图。

此外,利用对抗损失函数分别训练生成器和鉴别器以判别 SR 图像的真实性,其表达式如下,

$$L^G = -E_{I^{SR}} [\log D(G(I^{LR}))], \quad (14)$$

$$L^D = -E_{I^{HR}} [\log D(I^{HR})] - E_{I^{SR}} [\log (1 - D(G(I^{LR})))], \quad (15)$$

其中, L^G 表示生成器损失, L^D 表示鉴别器损失。利

用判别器和生成器的对抗性训练使得生成器不断改进其生成能力,以欺骗判别器,从而实现更好的超分辨率效果。

综上所述,本文用于进行超分辨率的生成器的损失函数表达式为

$$L^{Total} = L^{Joint} + L^{Perceptual} + 10^{-3} L^G, \quad (16)$$

结合对抗损失、感知损失以及联合损失,对网络进行训练,可以在多个层次上提升超分辨率图像的质量,实现更逼真和细节丰富的生成结果。

2 实验分析

2.1 数据集介绍

本文采用公开的星载红外遥感图像数据集 ISDD^[35] 和 NUDT-SIRST-Sea^[36] 来制作超分辨率图像数据集,其中 ISDD 数据集是目前首个公开的星载红外遥感舰船检测数据集,数据采集自 Landsat8 共包含 1284 张红外遥感图像,涵盖近岸和离岸场景的不同气候条件的图像,共标注了 3061 个舰船实例;NUDT-SIRST-Sea 共包含 48 张 10000 × 10000 的星载红外图像,其中包含了 17598 个精确标注的红外弱小船只。本文在 NUDT-SIRST-Sea 数据集中选取包含近岸和离岸船只的场景,划分出 1600 张 512 × 512 的图像构建数据集。

结合高斯模糊核与 Bicubic 退化模型分别对 ISDD 和 NUDT-SIRST-Sea 进行四倍下采样得到 128 × 128 的低分辨率图像,进而构建高-低分辨率星载红外图像数据对,并按 6:1:3 划分训练集、验证集和测试集。

2.2 训练参数设置

本文所有实验在同一台服务器上进行,服务器配备 Ubuntu 16.04 操作系统,40 个 Intel(R) Xeon(R) Silver 4210 CPU @ 2.20GHz CPU 和 4 个 NVIDIA RTX3090 GPU。实验中的 batch size 设置为 16,生成器和鉴别器均选用 Adam 优化器进行训练优化,为避免陷入 GAN 网络陷入局部最优,先对生成器进行迭代次数为 50000 的训练,随后对整个 GAN 网络进行训练,总迭代次数设置为 400000,初始学习率为 10^{-4} ,每隔 100000 次学习率衰减为原来的一半。

2.3 评估指标

超分辨率图像质量是评估超分辨率算法性能的重要指标。图像质量的评价方法可以分为主观评价和客观评价两类。主观评价方法依赖于人的主观感受来评估图像质量;而客观评价方法则通过量化指标来衡量图像质量。本文选择了三种常用的

图像质量评价指标 PSNR、SSIM 和 FID (Fréchet Inception Distance) 来计算红外超分辨率图像的质量,并在对比实验中通过目标检测的下游任务来评估不同超分辨率算法的性能。

PSNR 是一种基于对应像素点间误差的客观评价指标,均方误差 (MSE) 越小,则 PSNR 越大,SR 图像越接近于对比图像。式 (12) 给出了 PSNR 的计算公式,其中, MAX_I 表示图像像素的最大值。

$$MSE = \frac{1}{H \times W} \sum_i \sum_j (I_{ij}^{HR} - I_{ij}^{SR})^2, \quad (17)$$

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right), \quad (18)$$

SSIM 是一种用于衡量两幅图像之间相似度的指标,从亮度、对比度和结构三个方面评估图像的相似性,能够更好地反映人眼感知的图像质量。其计算公式为

$$SSIM(x, y) = \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (19)$$

其中, μ_x 和 μ_y 分别为 HR 图像和 SR 图像的均值, σ_x^2 和 σ_y^2 分别为 HR 图像和 SR 图像的方差, σ_{xy} 是 HR 图像和 SR 图像的协方差, C_1 和 C_2 是为了使输出结构稳定而引入的常数。SSIM 取值范围在 0 到 1 之间,数值越接近于 1 则图像质量越好。

FID 是通过比较 SR 图像和 HR 图像在特征空间中的分布来评估图像质量^[37],其表达式为:

$$FID = \|\mu_{HR} - \mu_{SR}\|^2 + \text{Tr} \left(\Sigma_{HR} + \Sigma_{SR} - 2(\Sigma_{HR} \Sigma_{SR})^{\frac{1}{2}} \right), \quad (20)$$

其中 μ_{HR} 和 μ_{SR} 分别表示 HR 图像和 SR 图像在 Inception 网络特征空间中的均值, Σ_{HR} 和 Σ_{SR} 表示协方差矩阵。FID 值越小,生成的图像质量就越高。

在进行对比实验中,除了利用上述客观图像评价指标外,本文还验证了超分辨率图像在实际应用场景中的有效性。使用 YOLOv9 网络^[38] 对超分辨率图像进行目标检测训练和测试,并利用 mAP50 和 mAP50-90 评估目标检测性能,从而进一步评估超分辨率图像的质量。

2.4 实验验证

2.4.1 消融实验

通过实验验证本文所提出的网络结构中各个子模块的有效性。消融实验制定了 5 个不同的测试模型,将前向传递未进行密集连接且特征融合选用残差连接和使用 l_1 损失训练的模型作为 Base 模型

Model1; Base模型的前向传递中引入密集连接的模型作为Model2,其中AFEB的数量 M 设置为4,STB数量 N 设置为6;使用MSCA替换Model2中传统的残差模块模型作为Model3模型。Model4则在Model3的基础上,将 l_1 损失替换为由Charbonnier损失和MS-SSIM损失构建的联合损失函数。Model5作为加入鉴别器后的生成对抗网络模型。所有模型使用相同的训练集和验证集以及网络配置进行训练,尺度因子为4,通过测试集平均PSNR、SSIM和FID对各个模型进行性能评估。

由表1和表2可知,对于没有密集连接、MSCA和联合损失函数所训练的Model1模型测试结果的PSNR和SSIM最小且FID最大。由于密集连接处将不同层次的特征进行相加,因此不增加额外的可训练参数。随着密集连接、MSCA、联合损失函数和鉴别器的逐步添加,模型的PSNR和SSIM增加且FID减小,相较于前三种模型结果,Model5在两类数据集上的测试结果PSNR和SSIM最大且FID最小。五种模型加以对比,分别证明了各个子模块的有效性。

为评估各个子模块对模型计算复杂度的影响,研究统计了五个子模型的可训练参数量、每秒浮点运算次数(FLOPs)以及模型推理单帧图像的平均时间,结果如表3所示。可以看出,相加形式的密集连

接和联合损失函数的引入均未增加可训练参数量,对FLOPs的影响也微乎其微。MSCA模块和鉴别器的加入提高了模型训练的计算复杂度。但由于模型推理过程仅通过生成器进行图像超分辨率,因此Model4与Model5生成器结构相同,二者推理时间相同。

2.4.2 对比实验

为进一步验证本文提出的融合密集连接与多注意力的星载红外遥感图像超分辨率算法的有效性,本研究将其与传统的Bicubic插值方法以及ESRGAN^[39]、CRAFT-SR^[40]、SAFMN^[41]、TransENet^[42]、DSSR^[43]、等基于深度学习的超分辨率算法分别在ISDD和NUDT-SIRST-Sea数据集上进行对比。除了利用PSNR、SSIM和FID等客观的图像评价指标外,本研究还使用YOLOv9目标检测模型对超分辨图像进行训练和推理,以获得mAP50和mAP50-95指标,比较该算法在下游任务中的可延展性。

实验过程中,将AFEB的数量 M 设置为4,STB数量 N 设置为6,与不同超分辨率算法在ISDD数据集和NUDT-SIRST-Sea数据集上的对比实验结果分别列在表4和表5中,其中加粗数字表示对比结果中相应指标的最优值,加下划线数字表示对比结果中相应指标的次优值。从这些结果可以看出,本文设计的超分辨率方法在两种星载红外遥感图像数

表1 ISDD数据集消融实验结果

Table 1 Ablation results of ISDD

	Model1	Model2	Model3	Model4	Model5
Dense block		√	√	√	√
MSCA			√	√	√
Joint loss				√	√
Discriminator					√
PSNR/dB	46.5374	46.8726	46.9220	46.9741	47.0499
SSIM	0.9764	0.9785	0.9789	0.9795	0.9801
FID	12.4895	11.1970	10.9424	9.9607	9.7600

表2 NUDT-SIRST-Sea数据集消融实验结果

Table 2 Ablation results of NUDT-SIRST-Sea

	Model1	Model2	Model3	Model4	Model5
Dense block		√	√	√	√
MSCA			√	√	√
Joint loss				√	√
Discriminator					√
PSNR/dB	41.7496	41.7587	41.7732	41.7789	41.7804
SSIM	0.9415	0.9419	0.9427	0.9447	0.9450
FID	16.4868	15.8964	15.7469	15.1687	14.5104

表3 各个子模块对模型计算复杂度的影响

Table 3 impact of each submodule on the computational complexity of the model

	Model1	Model2	Model3	Model4	Model5
可训练参数量/ 10^6	3.2576	3.2576	3.2598	3.2598	7.0915
FLOPs/ 10^9	58.3334	58.3334	58.3512	58.3512	59.1132
平均推理时间/s	0.0344	0.0346	0.0358	0.0362	0.0362

数据集上的 PSNR 和 SSIM 指标均取得了最大值,分别达到了 47.0499dB、0.9801 和 41.7804dB、0.9450,并且 FID 指标最小分别为 9.76 和 14.5104,在客观评价指标明显优于其他算法。此外,与其他几种算法相比,本文设计的超分辨率算法重建图像在利用 YOLOv9 算法进行目标检测时置信度阈值和 IOU (Intersection over union) 阈值分别设置为 0.2 和 0.45 的情况下,也取得了最高的 mAP50 和 mAP50-95,其中红外舰船检测数据集 ISDD 上达到了 0.947 和 0.530,在空间红外微小舰船检测数据集 NUDT-SIRST-Sea 上达到 0.578 和 0.340。这表明该算法相比于其他超分辨率算法不仅在图像重建质量上占据优势,而且在下游任务中也展现出良好的延展性。

表4 不同超分辨率算法在 ISDD 测试集上的对比结果

Table 4 Comparison results of different super resolution algorithms on the ISDD test set

评估指标	HR	bicubic	ESRGAN	CRAFT-SR	SAFMN	DSSR	TransENet	Ours
PSNR	\	46.2894	45.3329	<u>46.9941</u>	46.9751	46.3847	44.8646	47.0499
SSIM	\	0.9751	0.9692	<u>0.9796</u>	0.9784	0.9771	0.9735	0.9801
FID	\	18.9760	16.3093	<u>9.9029</u>	9.9545	11.3072	19.5806	9.7600
mAP50	0.950	0.940	0.925	0.938	0.944	0.934	0.943	<u>0.947</u>
mAP50-95	0.536	0.525	0.507	0.523	0.527	0.515	0.525	<u>0.530</u>

表5 不同超分辨率算法在 NUDT-SIRST-Sea 测试集上的对比结果

Table 5 Comparison results of different super resolution algorithms on the NUDT-SIRST-Sea test set

评估指标	HR	bicubic	ESRGAN	CRAFT-SR	SAFMN	DSSR	TransENet	Ours
PSNR	\	41.1219	39.1022	<u>41.7775</u>	41.6659	41.6141	40.7012	41.7804
SSIM	\	0.9301	0.9243	<u>0.9435</u>	0.9430	0.9413	0.9373	0.9450
FID	\	33.0765	41.1734	14.7453	<u>14.6553</u>	15.0255	19.3459	14.5104
mAP50	0.667	0.528	0.552	0.555	0.551	0.552	0.554	<u>0.578</u>
mAP50-95	0.425	0.298	0.320	0.332	0.330	0.324	0.323	<u>0.340</u>

表6 不同超分辨率算法的计算复杂度对比

Table 6 Comparison of computational complexity of different super-resolution algorithms

	ESRGAN	CRAFT-SR	SAFMN	DSSR	TransENet	Ours
可训练参数量/ 10^6	31.1974	0.7534	0.2395	8.9023	37.4589	7.0915
FLOPs/ 10^9	293.7135	12.5374	3.8521	1578.6186	12.0685	59.1132
平均推理时间/s	0.0516	0.0539	0.0107	0.0742	0.3744	0.0362

本文从 ISDD 和 NUDT-SIRST-Sea 的测试集中选取边缘信息丰富的图像进行放大显示,具体如图 5 和图 6 所示。

由图 5 和图 6 可以明显看出通过传统的 bicubic 所生成的图像细节模糊,特别是在高频区域,纹理信息丢失严重,ESRGAN 虽然在某些图像细节恢复上有所改进,但依旧存在一些模糊和严重的伪影。虽然 DSSR 较好地重建了图像的细节,提升了边缘清晰度,但是在整体图像的部分细节还原中存在偏差,造成局部失真,高频细节区域仍有一定的模糊和伪影。TransENet 较为清晰地还原了图像的纹理细节,但在部分区域存在明显的断层失真现象。CRAFT-SR、SAFMN 和本文方法在细节恢复、边缘处理和整体质量上的视觉效果相近,生成的图像细节丰富、边缘清晰且自然,几乎没有伪影的产生。但是,在客观的图像评价指标上,和下游任务实际应用评估中,本文方法取得了更为明显的提升。综合上述实验结果,本文建立的网络结构能够有效提升星载红外遥感图像超分辨率的效果。

表 6 总结了不同超分辨率算法的计算复杂度,比较了模型可训练参数量、FLOPs 和单帧图像的平均推理时间。

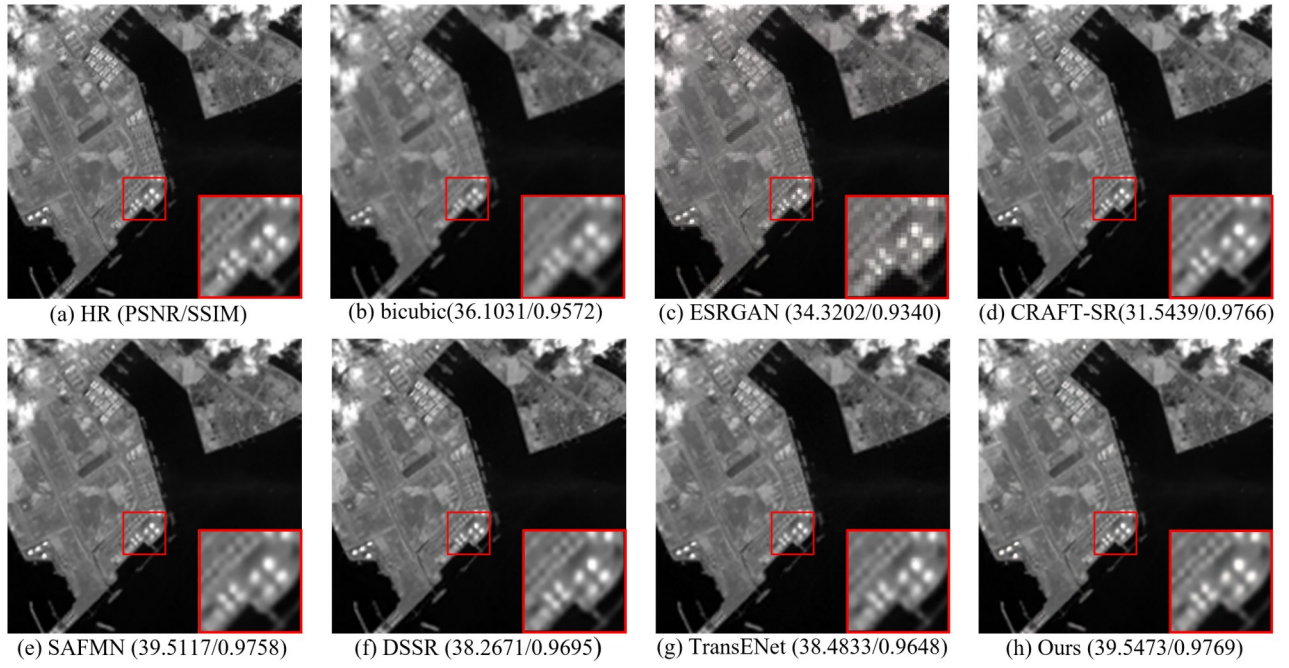


图5 不同超分辨率算法在ISDD测试集上的重建图像对比

Fig 5 Comparison of reconstructed images by different super resolution algorithms on the ISDD test set

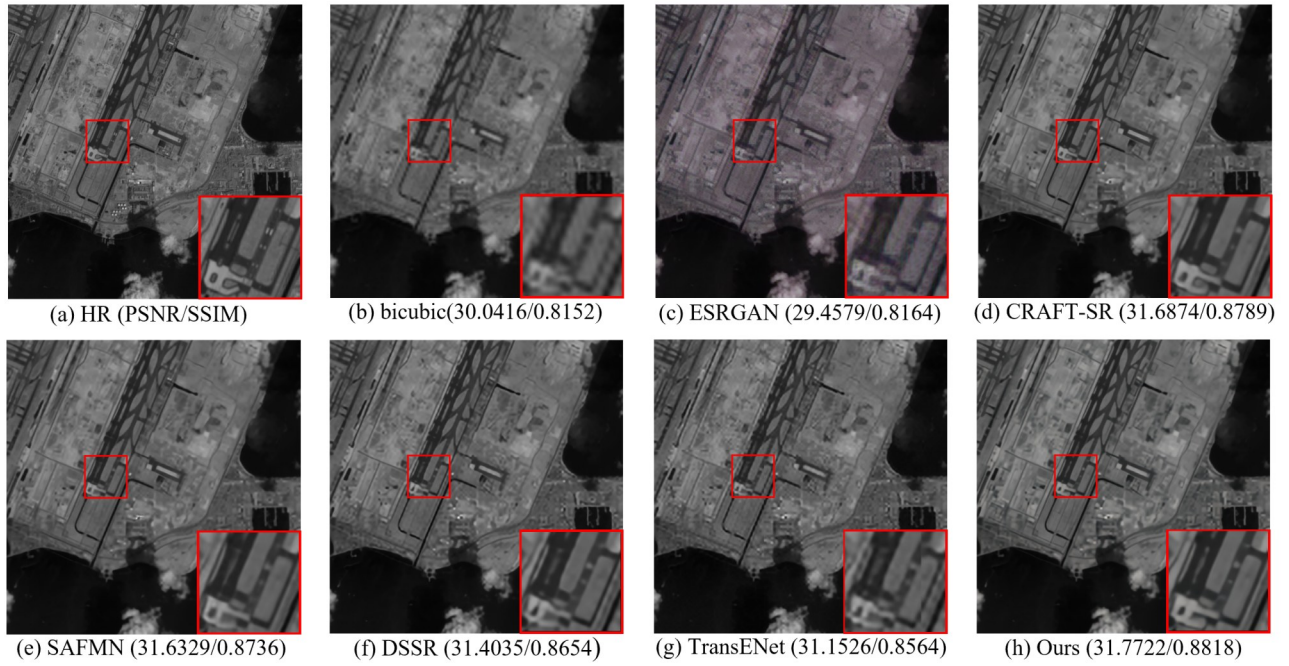


图6 不同超分辨率算法在NUDT-SIRST-Sea测试集上的重建图像对比

Fig 6 Comparison of reconstructed images by different super resolution algorithms on the NUDT-SIRST-Sea test set

均推理时间。ESRGAN 和本文模型的参数量和 FLOPs 为各自生成器与鉴别器的总和。相较于其他算法,本文算法具有较低的计算复杂度和合理的推理时间,展现出了良好的性能和效率平衡。

3 结论

为进一步解决星载红外遥感图像面临的分辨率不足和细节纹理模糊问题,本文提出了一种新的注意力增强型生成对抗网络模型。该模型结合了密集连接与 Swin Transformer 架构,通过设计深度注意力特征提取模块,有效提取不同尺度的图像特

征,并实现多层次特征的有效传递。同时,模型能够捕捉图像特征的长距离依赖关系,避免关键特征的丢失。我们用多尺度通道注意力特征融合模块替代了传统残差连接,突出有效特征,确保深层特征和浅层特征的有效融合。此外,设计的联合损失函数进一步优化了模型的性能。不同数据集上的实验结果表明,该方法能够有效恢复星载红外图像的纹理细节,提高分辨率并避免伪影产生。生成的超分辨率图像在下游任务中也表现出良好的性能,进一步证明了算法的实用性和有效性。

References

- [1] Gade R, Moeslund T B. Thermal cameras and applications: a survey[J]. *Machine vision and applications*, 2014, **25**: 245–262.
- [2] Jiang K, Wang Z, Yi P, et al. ATMFN: Adaptive-threshold-based multi-model fusion network for compressed face hallucination[J]. *IEEE Transactions on Multimedia*, 2019, **22**(10): 2734–2747.
- [3] Arrue B C, Ollero A, De Dios J R M. An intelligent system for false alarm reduction in infrared forest-fire detection[J]. *IEEE Intelligent Systems and their Applications*, 2000, **15**(3): 64–73.
- [4] Zhang J, Zhou X, Li L, et al. A combined stripe noise removal and deblurring recovering method for thermal infrared remote sensing images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, **60**: 1–14.
- [5] Liu A, Liu Y, Gu J, et al. Blind image super-resolution: A survey and beyond[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2022, **45**(5): 5461–5480.
- [6] Wang X, Sun L, Chehri A, et al. A review of GAN-based super-resolution reconstruction for optical remote sensing images[J]. *Remote Sensing*, 2023, **15**(20): 5062.
- [7] Rukundo O, Cao H. Nearest neighbor value interpolation[J]. *arXiv preprint arXiv:1211.1768*, 2012.
- [8] Keys R. Cubic convolution interpolation for digital image processing[J]. *IEEE transactions on acoustics, speech, and signal processing*, 1981, **29**(6): 1153–1160.
- [9] Chen H, Xue J, Zhang S, et al. Image super-resolution based on adaptive cospase regularisation[J]. *Electronics Letters*, 2014, **50**(24): 1834–1836.
- [10] Gauvain J L, Lee C H. Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains[J]. *IEEE transactions on speech and audio processing*, 1994, **2**(2): 291–298.
- [11] 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.
- [12] Ledig C, Theis L, Huszar F, et al. Photo-realistic single image super-resolution using a generative adversarial network[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4681–4690.
- [13] Gao Z, Chen J. Maritime infrared image super-resolution using cascaded residual network and novel evaluation metric[J]. *IEEE access*, 2022, **10**: 17760–17767.
- [14] Choi K, Kim C, Kang M H, et al. Resolution improvement of infrared images using visible image information[J]. *IEEE signal processing letters*, 2011, **18**(10): 611–614.
- [15] HUANG Shuo, HU Yong, GONG Cai-Lan, ZHENG Fu-Qiang. Saliency region super-resolution reconstruction algorithm for infrared images based on sparse coding[J]. *J. Infrared Millim. Waves*, 2020, **39**(3): 388–395. (黄硕, 胡勇, 巩彩兰, 等. 基于稀疏编码的红外显著区域超分辨率重建算法[J]. *红外与毫米波学报*), 2020, **39**(03): 388–395.
- [16] Liu S, Yang Y, Li Q, et al. Infrared image super resolution using gan with infrared image prior[C]//2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP). IEEE, 2019: 1004–1009.
- [17] SHAO Bao-Tai, TANG Xin-Yi, JIN Lu, LI Zheng. Single frame infrared image super-resolution algorithm based on generative adversarial nets[J]. *J. Infrared Millim. Waves*, 2018, **37**(4): 427–432. (邵保泰, 汤心溢, 金璐, 等. 基于生成对抗网络的单帧红外图像超分辨率算法[J]. *红外与毫米波学报*), 2018, **37**(04): 427–432.
- [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] Zou Y, Zhang L, Liu C, et al. Super-resolution reconstruction of infrared images based on a convolutional neural network with skip connections[J]. *Optics and Lasers in Engineering*, 2021, **146**: 106717.
- [20] DENG Wei, CHEN Jianfei, ZHANG Sheng. Super-Resolution Reconstruction of Thermal Infrared Image in Deep Residual Network With Skip Connections[J]. *Electronics Optics & Control*, 2023, **30**(3): 27–32. (邓伟, 陈建飞, 张胜. 深层跳线残差网络热红外图像超分辨率重建[J]. *光电与控制*), 2023, **30**(03): 27–32.
- [21] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. *Advances in neural information processing systems*, 2017, 30.
- [22] Yang F, Yang H, Fu J, et al. Learning texture transformer network for image super-resolution[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 5791–5800.
- [23] Xiao Y, Yuan Q, Jiang K, et al. TTST: A top-k token selective transformer for remote sensing image super-resolution[J]. *IEEE Transactions on Image Processing*, 2024.
- [24] Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 10012–10022.
- [25] Salimans T, Kingma D P. Weight normalization: A simple reparameterization to accelerate training of deep neural networks[J]. *Advances in neural information processing systems*, 2016, 29.
- [26] Long Y, Wang X, Xu M, et al. Dual self-attention Swin transformer for hyperspectral image super-resolution[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, **61**: 1–12.
- [27] Yang F, Yang H, Fu J, et al. Learning texture transformer network for image super-resolution[C]//Proceedings of

- the IEEE/CVF conference on computer vision and pattern recognition. 2020: 5791–5800.
- [28] Iandola F, Moskewicz M, Karayev S, *et al.* Densenet: Implementing efficient convnet descriptor pyramids[J]. *arXiv preprint arXiv:1404.1869*, 2014.
- [29] Liang S B, Song K, Zhao W, *et al.* DASR: Dual-Attention Transformer for infrared image super-resolution[J]. *Infrared Physics & Technology*, 2023, **133**: 104837.
- [30] Dai Y, Gieseke F, Oehmcke S, *et al.* Attentional feature fusion[C]//Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2021: 3560–3569.
- [31] Lim B, Son S, Kim H, *et al.* Enhanced deep residual networks for single image super-resolution[C]//Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2017: 136–144.
- [32] Zhao, Hang, *et al.* "Loss functions for image restoration with neural networks." *IEEE Transactions on computational imaging* 3.1 (2016): 47–57.
- [33] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. *arXiv preprint arXiv:1409.1556*, 2014.
- [34] Jolicoeur-Martineau A. The relativistic discriminator: a key element missing from standard GAN[J]. *arXiv preprint arXiv:1807.00734*, 2018.
- [35] Han Y, Liao J, Lu T, *et al.* KCPNet: Knowledge-driven context perception networks for ship detection in infrared imagery[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, **61**: 1–19.
- [36] Wu T, Li B, Luo Y, *et al.* MTU-Net: Multilevel TransU-Net for space-based infrared tiny ship detection[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, **61**: 1–15.
- [37] Heusel M, Ramsauer H, Unterthiner T, *et al.* Gans trained by a two time-scale update rule converge to a local nash equilibrium[J]. *Advances in neural information processing systems*, 2017, 30.
- [38] Wang C Y, Yeh I H, Liao H Y M. Yolov9: Learning what you want to learn using programmable gradient information[J]. *arXiv preprint arXiv:2402.13616*, 2024.
- [39] Wang, Xintao, *et al.* "Esrgan: Enhanced super-resolution generative adversarial networks." *Proceedings of the European conference on computer vision (ECCV) workshops*. 2018.
- [40] Li, Ao, *et al.* "Feature modulation transformer: Cross-refinement of global representation via high-frequency prior for image super-resolution." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023.
- [41] Sun, Long, *et al.* "Spatially-adaptive feature modulation for efficient image super-resolution." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023.
- [42] Lei S, Shi Z, Mo W. Transformer-based multistage enhancement for remote sensing image super-resolution[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, **60**: 1–11.
- [43] Li F, Wu Y, Bai H, *et al.* Learning detail-structure alternative optimization for blind super-resolution[J]. *IEEE Transactions on Multimedia*, 2022, **25**: 2825–2838.