

基于层次交互动态注意力与序列学习的图像超分辨率重建*

莫恒辉 魏霖静

(甘肃农业大学信息科学技术学院 兰州 730070)

摘要:针对图像超分辨率中网络关注度不足、模块间协同性弱和深层特征表征消失等问题,提出了一种结合层次交互动态注意力与序列学习单元的多级残差聚合超分辨率重建模型。模型采用多层次特征融合与跳跃连接的网络结构,结合不同层次的特征,捕获从低到高的不同级别的信息,生成更丰富和准确的表征。模块里通过残差连接避免梯度消失,实现深度网络的平滑损失空间和灵活增加。提出动态层次融合注意力模块动态计算各个特征的重要性权重,进行有选择性的特征融合,并通过序列学习单元捕获更长范围的上下文信息。提出多尺度特征融合模块将不同感受野的特征信息提取融合,以挖掘更深层的特征表示。模块尾部引入轻量化无参注意力机制自适应加权特征图,恢复图像高频细节。实验结果表明,相较于主流算法,提出的算法在各种公开测试集(Set5、Set14、BSD100、Urban100、Manga109)上进行3倍超分辨率重建评估,平均峰值信噪比提升约0.47 dB,平均结构相似性提升约0.006 8。且在遥感图像超分辨率重建方面展示出实际应用潜力。证明其在图像超分辨率重建方面的优越性。

关键词:图像超分辨率;注意力机制;序列学习;多级残差聚合网络;峰值信噪比

中图分类号: TP391.4 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Image super-resolution reconstruction based on hierarchical interactive dynamic attention and sequence learning

Mo Henghui Wei Linjing

(College of Information Science and Technology, Gansu Agricultural University, Lanzhou 730070, China)

Abstract: To address issues like insufficient network focus, weak synergy between modules, and loss of deep feature representations in image super-resolution, a multi-level residual aggregation super-resolution reconstruction model is presented. This model integrates hierarchical interactive dynamic attention with sequence learning units, featuring a network structure with multi-level feature fusion and skip connections for capturing diverse information levels more richly and accurately. Residual connections prevent gradient vanishing, ensuring smooth, flexible enhancements in deep networks. The dynamic hierarchical fusion attention module dynamically assigns importance weights to each feature for selective fusion, complemented by sequence learning units that broaden the contextual scope. A multi-scale feature fusion module combines features from different receptive fields to explore deeper representations. At the end, a lightweight, parameter-free attention mechanism adaptively weights feature maps, restoring high-frequency details. Experimental results demonstrate that this model surpasses mainstream algorithms in 3x super-resolution reconstruction across multiple public datasets (Set5, Set14, BSD100, Urban100, Manga109), with average improvements of about 0.47 dB in PSNR and 0.006 8 in SSIM, showcasing its potential for practical remote sensing applications and its superiority in the domain.

Keywords: super-resolution; attention mechanism; sequence learning; multi-level residual aggregation network; PSNR

收稿日期:2024-01-08

* 基金项目:科技部国家外专项目(G2022042005L)、甘肃省高等学校产业支撑项目(2023CYZC-54)、甘肃省重点研发计划(23YFWA0013)、兰州市人才创新创业项目(2021-RC-47)、2020年甘肃农业大学研究生教育研究项目(2020-19)、2021年甘肃农业大学校级“三全育人”试点推广教学研究项目(2022-9)、2022年甘肃农业大学校级专业综合改革项目(2021-4)资助

0 引言

超分辨率重建(super-resolution, SR)是计算机视觉的关键研究方向,旨在从低分辨率图像(low-resolution, LR)中恢复出高清晰度的高分辨率(high-resolution, HR)图像。深度学习技术的兴起,特别是卷积神经网络(convolutional neural networks, CNN)飞速发展,超分辨率重建技术在医学成像、矿井分析和文本增强^[1-3]等多个领域展现广泛应用,比如提升医疗诊断的精确度和优化矿井信息的精度。因此,超分辨率重建吸引了众多研究者的关注。

早期研究人员通过插值方法,如最近邻、双线性 and 双三次插值^[4]等,难以达到令人满意的高质量结果。基于深度学习的超分辨率重建研究最早由 Dong 等^[5]提出的超分辨率卷积神经网络(super-resolution convolutional neural network, SRCNN)模型。为了加强特征提取并提高运算速度, Kim 等^[6]提出极深超分辨率网络(very deep super-resolution, VDSR)。VDSR 使用 20 层的残差卷积网络端到端的使模型学习到更加复杂和高层次的图像特征。然而,其深层网络结构导致计算成本较高,且在训练过程中容易出现过拟合的风险,限制了模型在资源受限环境下的应用。Lim 等^[7]提出增强的深度残差超分辨率网络(enhanced deep residual network for super-resolution, ED-SR),通过引入残差缩放,通过引入更多的残差块和更大的卷积核,去除了批量归一化层,从而有效地学习到更复杂多样的图像特征,并预防了梯度爆炸问题。但是,更大的模型规模可能导致效率下降,使其在实际应用中需要更多的计算资源。Ahn 等^[8]提出一种轻量级基于级联残差网络(cascading residual network, CARN),CARN 通过级联残差网络架构实现模块内参数的共享和重用。Zheng 等^[9]提出轻量级的信息多重蒸馏网络(information multi-distillation network, IMDN),通过信息多重蒸馏网络以及自适应裁剪策略与通道注意力方法,有效地减少了模型的大小和计算量。这些轻量化模型虽然在轻量化方面取得了进展,但可能牺牲了特征提取的深度和复杂性,限制了在处理高度复杂图像时的性能。Li 等^[10]针对盲目增加网络深度并不能有效改善网络,提出了一种新颖的多尺度残差网络(multi-scale residual network, MSRN),通过引入不同大小的卷积核并使特征交互,充分挖掘了图像的特征,且在多数先进的方法上取得优异的表现。然而,多尺度设计增加了模型的复杂度,可能在一定程度上影响了训练和推理的效率。Liang 等^[11]提出基于 Swin Transformer 架构的 SwinIR 模型,引入有效的长距离依赖捕获能力,通过优化的注意力机制和灵活的窗口策略在处理复杂图像时表现出色。但是,这种高级的架构需要显著的计算资源,尤其是在训练阶段,限制了其在计算受限的环境下的适用。

尽管目前超分辨率重建技术已经相当成熟,但依旧面临一些问题和挑战。其中,一个主要问题是网络关注

度不足^[12],网络关注度的缺乏意味着在处理多模态信息时,网络无法区分特征之间的重要性层级,导致模型在训练和推理过程中可能对所有输入特征分配相等的关注度和权重。这种无差别的特征处理方式导致模型无法精确捕获和优化那些对最终任务性能更为关键的信息和细节。再者,多数模型中各个模块之间缺乏足够的协同,降低了特征学习和表征的准确性,从而制约了整体性能的提升。此外,由于网络结构过深,一些网络结构过于复杂,导致重要的特征信息在传递过程中消失,影响最终的重建效果^[13-14]。

针对上述问题,本文提出基于层次交互动态注意力与序列学习的图像超分辨率重建算法。通过整合特殊的多层次特征融合、跳跃连接网络结构、残差连接、动态层次融合注意力机制、多尺度特征融合模块和序列学习单元以及轻量化无参注意力机制等结构和方法,以提升特征表征的全面性和准确性,加快网络收敛,精确捕获关键特征,提高细节重建效果与模型性能。为图像重建技术提供坚实的理论和基础。

1 本文方法

1.1 超分辨率网络总体结构

为了应对现有超分辨率方法的限制,本文提出了创新的动态层次超分辨率网络(dynamic hierarchical super-resolution, DHSR),如图 1 所示。该模型融合了层次交互残差网络和注意力机制,以实现更加准确且细致的图像重建。整个模型可被划分为 3 大部分,浅层特征提取、深层特征提取与融合、高分辨率图像重建。

首先,LR 图像经过一个由 1×1 卷积构成的浅层特征提取模块。然后,通过由 3 个主动态残差模块(main residual dynamic module, MRDM)组成的残差网络提取网络的深层特征与融合。最后,通过像素混洗完成最终重建。整个 SR 过程可以表示为:

$$F_0 = C_{1 \times 1}(I^{LR}) \quad (1)$$

式中: F_0 表示经过浅层特征提取模块后的特征图; $C_{1 \times 1}$ 为 1×1 卷积; I^{LR} 为模型输入低分辨率图片。

$$F_{MRDM} = C_{1 \times 1}(f_{MRDM}^n(f_{MRDM}^{n-1}(\dots f_{MRDM}^0(F_0)\dots))) \oplus F_0 \quad (2)$$

式中: F_{MRDM} 表示经过深层特征提取与融合模块之后的特征图; f_{MRDM}^i 表示经过第 i 个 MRDM 模块输出特征图; \oplus 表示特征图相加。

$$I^{SR} = C_{1 \times 1}(PS(C_{1 \times 1}(F_{MRDM}))) \quad (3)$$

式中: I^{SR} 表示最后重建的超分辨率图像; PS 表示像素混洗模块。相比传统的双三次插值,能学习更复杂的模式。通过上采样和卷积的分解,减少了特征图混淆,更有效地保留了图像细节。

1.2 MRDM 模块设计

如图 1 所示,本文提出的动态层次超分辨率网络采用

了一种创新的层次结构以实现更精确的图像超分辨率。网络的主体部分由连续的级联残差融合模块组成,包括动态层级注意力机制(dynamic hierarchical attention mod-

ule, DHAM)和其进一步扩展的动态组(dynamic block, DB)。这些模块针对图像的不同细节级别和内容特性进行操作,充分利用了深度学习的层次特性。

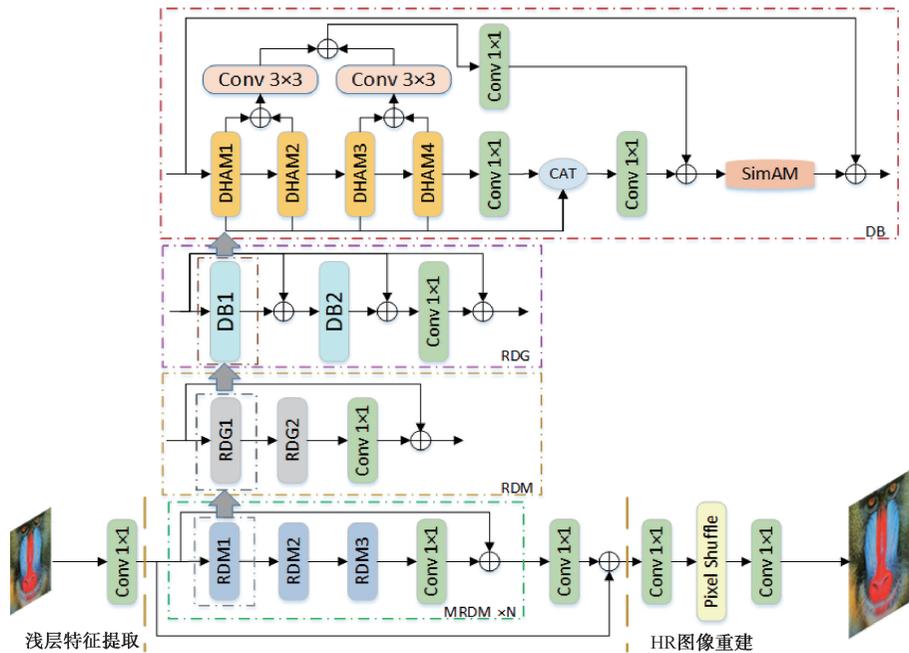


图1 本文算法网络结构

Fig.1 Algorithmic network structure of this paper

首先,在 DHSR 网络的顶部,对于 MRDM 块数量的选择,出于对网络做轻量级的扩展与计算机显存的考虑,本文选择一个 MRDM 作为网络深层提取模块。其中,每个 MRDM 利用 3 个动态残差模块(residual dynamic module, RDM)来做增强输入特征,并为后续模块提供基础。

接着,对于 RDM 采取瓶颈结构进行对动态残差组(residual dynamic group, RDG)数量选择缩小至两个。对于 RDM 选择简单的残差连接结构,进一步保证梯度流动的稳定性。过程可表示为:

$$F_{RDM} = C_{1 \times 1}(F_{RDG}^n(\dots F_{RDG}^0(I_{RDM})\dots)) \oplus I_{RDM} \quad (4)$$

式中: F_{RDM} 表示经过 RDM 模块之后的特征图; F_{RDG}^i 表示经过第 i 个 RDG 模块输出特征图; I_{RDM} 表示 RDM 模块输入特征图。

RDG 模块是由两个 DB 模块组成,采用密集连接结构输入特征与每个 RDG 块输出特征融合作为下一个 RDG 块的输入,有助于梯度流、特征复用和信息传递,同时提高参数效率并增强模型的正则化能力,从而提升网络的性能。其过程表示为:

$$F_{RDG} = C_{1 \times 1}(F_{DB}^n(\dots F_{DB}^0(I_{RDG}) \oplus I_{RDG} \dots) \oplus I_{RDG}) \oplus I_{RDG} \quad (5)$$

式中: F_{RDG} 表示经过 RDG 模块之后的特征图; F_{DB}^i 表示经过第 i 个 DB 模块输出特征图; I_{RDG} 表示 RDG 模块输入特征图。

每个 DB 模块展现了一个多层次的结构设计。首先

核心由 4 个 DHAM 模块组成。模块逐个处理输入特征,各自产生其特有的特征响应。DB 模块的设计中采用了双路特征融合策略。第 1 路,通过通道维度的拼接,该策略直接对所有 DHAM 模块的输出进行累积,强调了特征的全局融合和信息完整性。第 2 路,设计对称卷积组模块,特定的 3×3 卷积被用于每对模块的输出并进一步做多层次特征融合提取,强调局部特征的优化和空间信息的细致处理,这些处理后的特征被汇总并通过另一个 1×1 卷积进行综合。最终将两路特征通过矩阵相加的方式进行融合。两路特征的设计增强网络对于不同尺度特征的处理能力,同时为后续的特征融合提供坚实基础。

1.3 SimAM

DB 模块中巧妙地融入了 SimAM。这是一种先进的技术,SimAM 能够在网络层中直接推断出三维的注意力权重,充分考虑空间和通道维度的相关性的同时,避免增加额外的参数量^[15]。

在超分辨率重建任务中,SimAM 为携带关键信息的神经元分配更高的权重。通过对特定特征的相邻神经元施加空间抑制,SimAM 能够减少复杂背景对特征识别的干扰,从而更突出地展现该特征的核心属性^[16]。其作用原理基于能量函数发挥其作用,灵感来源于神经科学中的空域抑制现象。在这一现象中,活跃的神经元会抑制其周围的非活跃神经元。因此,神经元的重要性可以通过这一能量函数来衡量。能量函数形式表示为:

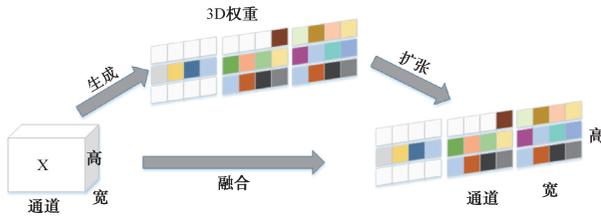


图2 SimAM原理

Fig. 2 SimAM principle

$$e_t(w_t, b_t, y, x_i) = (y_t - \hat{t})^2 + \frac{1}{M-1} \sum_{i=1}^{M-1} (y_0 - \hat{x}_i)^2 \quad (6)$$

$$\hat{t} = (w_t t + b_t) \quad (7)$$

$$\hat{x}_i = (w_t x_i + b_t) \quad (8)$$

式中: t 表示目标神经元; x_i 表示其他神经元; b_t 表示偏置; M 表示神经元数量。上述公式的最小化过程可以视为增强神经元 t 与其他神经元间的线性区分能力。为简化操作,将标量 y_t 和 y_0 用二进制标签分别表示为 1 和 -1,最终通过添加正则项 λ 对能量函数整理表达为:

$$e_t(w_t, b_t, y, x_i) = \frac{1}{M-1} \sum_{i=1}^{M-1} (-1 - \hat{x}_i)^2 + (1 - \hat{t})^2 + \lambda w_t^2 \quad (9)$$

$$w_t = -\frac{2(t - \mu_t)}{(t - \mu_t)^2 + 2\sigma_t^2 + 2\lambda} \quad (10)$$

$$b_t = -\frac{1}{2}(t + \mu_t)w_t \quad (11)$$

式中: μ_t 和 σ_t 分别表示图像的均值和方差。因此最终简化的最小能量函数如下:

$$e_t^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda} \quad (12)$$

$$\tilde{X} = \text{sigmoid}\left(\frac{1}{e_t^*}\right) \odot X \quad (13)$$

由式(12)可知,当神经元 t 与周围神经元区别越大时, e_t^* 表现为越小,作为式(13)分母则表现越重要,通过 Sigmoid 抑制过大的值,从而满足注意力机制的定义,得到最终的特征加权注意力特征图。

二路融合的特征通过 SimAM 并通过 1×1 卷积混洗特征与 DB 模块的输入进行残差连接作为最后的输出。使 DB 模块关注对最终任务更为关键的特征区域,同时忽略不太重要的信息。

1.4 动态层次融合注意力机制

动态卷积的引入为深度学习领域带来了创新的突破,与传统的卷积操作不同,为每个输入样本动态生成了卷积核。这种动态性为模型提供了额外的适应性,允许它对不同的输入具有更大的响应能力^[17]。LR 空间中包含冗余的低频成分以及少量的高频成分,这些少量的高频特征往往与图像中的细节和纹理信息有关,而利用注意力机制可以引导模型更加关注这些高频细节,但并非所有注意力模

块均有益^[18]。传统的方法中,多种注意力机制的特征融合往往采用简单的方法,如串联或并联,这意味着所有的特征都被等同对待。这种方法并未充分挖掘不同注意力机制下特征的差异性和重要性,同时可能导致在重复或相似的特征上消耗过多的计算资源。

为此,本文提出动态层级注意力机制,进一步拓展了动态卷积的概念,通过动态地调整不同注意力机制的权重,使模型能够在不同上下文中识别和强调关键的特征。特征提取部分使用特征提取模块(feature extraction, FE)和多尺度特征融合模块(multi-scale feature fusion module, MFFM)做模块深层特征的挖掘。并将不同的注意力机制模块与层级交互结构融合,以增强模型的表征能力。动态地为不同注意力机制模块赋予相应权重,以确定在给定的上下文特征中具有更为关键作用的注意力机制模块。且最终输出特征经过序列学习单元进行整合和优化以确保模型输出的稳健性。网络结构如图3所示。

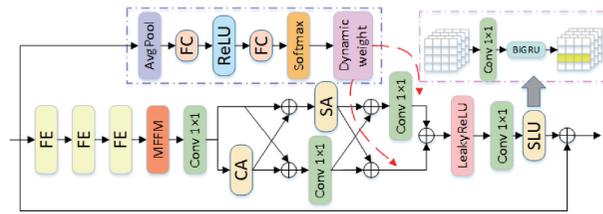


图3 动态层级注意力模块结构

Fig. 3 DHAM structure

DHAM 网络分别由层次化特征提取和融合路径和动态注意力动态调制路径两个主要路径组成。层次化特征提取和融合路径首先依赖于多个 FE 模块做特征提取,网络结构如图4所示。FE 通过一系列 3×3 卷积操作将输入特征进行变换和处理且 ReLU 函数对重要神经元进行激活,旨在学习特征的高级表示。

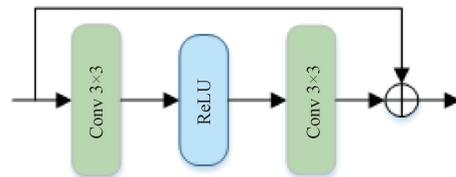


图4 特征提取模块结构

Fig. 4 FE structure

经过多个 FE 模块后特征通过 MFFM 模块进一步做多尺度的深层特征提取与融合。整合后的特征通过 1×1 卷积做特征混洗操作,进一步捕获输入特征之间的复杂关系。随后,特征分别进入层级交互结构且与通道注意力机制(channel attention, CA)与空间注意力机制(spatial attention, SA)进行融合表示。

通道注意力机制通过突出重要的通道特征,有助于模型更好地理解图像中的关键信息,如纹理、边缘和结

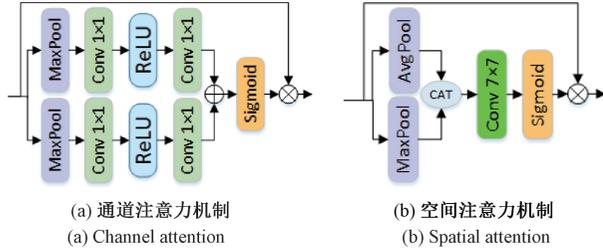


图5 两种注意力机制结构

Fig. 5 Two attention mechanism structures

构^[19],从而生成更精确和清晰的超分辨率图像。还有助于减少不必要的信息冗余,提高了计算效率。

与此同时,空间注意力机制专注于不同位置的特征交互,使模型能够更好地捕捉到图像中的空间细节和结构^[20]。通过动态调整每个像素的权重,空间注意力机制能够更好地处理不同区域的信息,特别是在存在缩放、旋转或畸变的情况下,有助于提高超分辨率重建的稳健性。

为使通道和空间注意力机制能更高效的特征复用弥补忽视掉的特征,将其与层次交互结构融合。具体过程可表示为:

$$F_{ca1} = CA(I_0) \quad (14)$$

$$F_{sa1} = SA(I_0 \oplus F_{ca1}) \quad (15)$$

$$F_{ca2} = C_{1 \times 1}(F_{ca1} \oplus I_0) \quad (16)$$

$$F_{ca_over} = F_{ca2} \oplus F_{sa1} \quad (17)$$

$$F_{sa_over} = C_{1 \times 1}(F_{ca2} \oplus F_{sa1}) \quad (18)$$

式中: I_0 表示进入层次交互结构之前的特征图; CA 表示通过通道注意力机制模块; SA 表示通过空间注意力机制模块。经过交融后的特征信息提供了更全面的图像上下文理解,从而使模型能够更好地进行超分辨率重建,生成更高质量的图像结果。

另一条注意力动态调制路径计算权重如式(19)所示。 $f_d(\cdot)$ 表示注意力动态调制模块,对输入 x_n 进行全局平均池化,然后通过两个具有 ReLU 激活函数的全连接层生成两个分支的注意力权重。通过这种方式,动态注意力模块能够从整个图像中捕获特征信息,增加了模型的感受野。 γ_n 会根据输入特征的变化自适应的调整。

$$\gamma_n = f_d(x_n) \quad (19)$$

为促进动态注意力模块的学习引入了和为一的约束条件,即 $\gamma_{ca} + \gamma_{sa} = 1$, 约束条件旨在减小卷积核空间的复杂度,从而简化权重 γ 的学习过程。因此,采用 Softmax 函数来生成两个分支的归一化注意力权重。最终与层次交互结构融合后的各自注意力特征结果相乘得到动态调制后的结果。

$$F_{cs} = SU(C_{1 \times 1}(lRelu(\gamma_{ca} \cdot F_{ca_over} + \gamma_{sa} \cdot F_{sa_over}))) \quad (20)$$

式中: $lRelu$ 为 LeakyReLU 激活函数; γ_{ca} 、 γ_{sa} 为动态调制后的权重; SU 为序列学习单元。输出特征经过序列学

习单元得到最终高度优化的特征表示。

1.5 MFFM 模块

多尺度特征融合模块是本文的一个核心组成部分,如图6所示。通过不同尺度和扩展率的膨胀卷积,强化了网络在捕获全局上下文信息方面的能力,从而实现更精确的特征提取和融合。

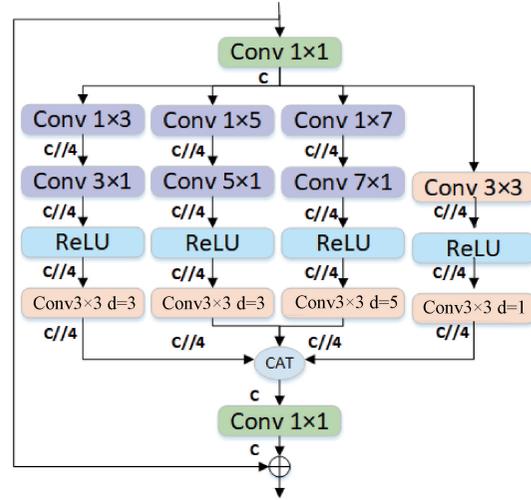


图6 多尺度特征融合模块结构

Fig. 6 MFFM structures

首先,模型通过 1×1 的卷积层对输入特征图进行线性变换。

$$F_{1 \times 1} = C_{1 \times 1}(I_{MFFM}) \quad (21)$$

式中: I_{MFFM} 为 MFFM 模块的输入特征图。

接着,网络分支化为4个路径以提取特征,为获得更多的空间上下文信息,且不增加过多的参数和计算复杂度,分支1~3均采用分步卷积的方式首先沿特征图宽度方向做 $1 \times K$ 卷积并做通道上的降维至 $C//4$ 操作,捕获该方向上的空间关系。随后,沿另一个高度方向做 $K \times 1$ 卷积,整合之前捕获的信息并获取双方向的空间关系。最终得到与 $K \times K$ 相当的感受野卷积后的特征图。随后使用 3×3 的膨胀卷积根据不同扩张率对卷积的感受野进行扩张。关注特征图不同方向上的纹理特征的同时,进一步整合局部与全局的特征差异。

$$F_{MFFM_i} = C_{3 \times 3}^d(F_{relu}(C_{k \times 1}(C_{1 \times k}(F_{1 \times 1})))) \quad (22)$$

式中: $C_{1 \times k}$ 、 $C_{k \times 1}$ 表示不同 K 值的分步卷积,本文选择 $K = 3, 5, 7$; $C_{3 \times 3}^d$ 表示不同扩张率 d 的膨胀卷积, $d = 3, 5$; F_{MFFM_i} 表示不同分支的处理结果, $i = 1, 2, 3$ 。

第4分支与前3个分支有所不同,直接采用两个 3×3 的连续卷积操作,确保网络不丢失原始特征的重要信息,并为其他3个分支提供一个基准或参照。

$$F_{MFFM_4} = C_{3 \times 3}(F_{relu}(C_{3 \times 3}(F_{1 \times 1}))) \quad (23)$$

最后,将4个分支的输出合并为通道数为 C 的特征图。并经过最终的 1×1 卷积混洗操作并与 MFFM 模块的输入特征做残差连接输出最终的结果,形成一个综合多

尺度特征的特征图,确保网络在不同的上下文情境中都能够提供出色的性能。最终结果表示为:

$$F_{MFFM} = C_{1 \times 1}(Cat(F_{MFFM_1}, F_{MFFM_2}, F_{MFFM_3}, F_{MFFM_4}) \oplus I_{MFFM}) \quad (24)$$

式中: Cat 表示特征图通道上 Concat 操作; F_{MFFM} 为 MFFM 模块的最终输出。

1.6 序列学习单元(SLU)模块

在图像超分辨率重建领域,考虑图像中的序列特征对提高重建质量至关重要。且对于具有特定纹理或模式的图像,如动物皮毛或建筑物表面,正确识别和重建这些序列特征显著提升图像的细节和质量。

为了有效捕捉这些序列关系,本文引入双向门控循环单元(BiGRU)来捕获这些特征之间的空间序列关系。如图4所示,SLU模块结合了卷积层和BiGRU。首先,卷积层负责从低分辨率图像中提取基本的空间特征。随后,将这些特征按行或列重构成一维的序列,并将它们输入到BiGRU中。BiGRU通过学习这些特征间的序列关系,更新其内部状态,更准确地推测每个位置的细节。这种处理方式可以通过式(25)和(26)来表示。最后,输出的序列特征再次整合为二维的特征图,为后续的超分辨率重建做准备。

$$F_h = GRU(x_{t_h}, h_{t_h-1}) \quad t_h = 0, 1, \dots, H \quad (25)$$

$$F_w = GRU(x_{t_w}, h_{t_w-1}) \quad t_w = 0, 1, \dots, W \quad (26)$$

式中: H 和 W 分别是特征图的高度和宽度; h_t 表示隐层状态; GRU 表示对输入特征的非线性变换,把不同位置特征编码成对应的 GRU 隐层状态; x_t 表示输入特征; t_h 和 t_w 表示输入特征不同沿高度和宽度方向构建特征向量。

序列学习单元的设计不仅恢复了基本的图像结构,还捕获了图像中的复杂序列信息,从而得到更高质量的超分辨率图像。

2 模型训练及结果分析

2.1 实验环境配置

实验在具有高性能硬件配置的平台进行的,包括 NVIDIA GeForce RTX 3090 Ti 显卡、Intel Core i9-13900HQ@5.40GHz CPU,运行在 Ubuntu 18.04 LTS 操作系统上。深度学习框架采用 Pytorch 1.8.1。软件环境方面,采用 Pycharm 2.022.3、OpenCV 3.4.6 和 Python 3.9.1,确保实验的高效执行。

2.2 数据集构建

DIV2K 数据集是高质量的图像数据集,专为图像恢复任务而设计。DIV2K 数据集包括 800 张训练图像,100 张验证图像和 100 张测试图像。由于测试数据集的基准真值尚未公开,本文在 5 个公开标准基准数据集(Set5、Set14、BSD100、Urban100、Manga109)进行了测试比较。

2.3 模型训练参数设置

训练集的数据随机裁剪为 96×96 大小的低分辨率输

入补丁和对应的高分辨率补丁。通过随机水平翻转和 90° 旋转来增强训练数据。且通过减去 DIV2K 数据集的各自的平均 RGB 值来预处理图像。模型采用 ADAM^[21] 优化器和 L1 损失函数进行训练,设置一阶动量参数 $\beta_1 = 0.9$ 、二阶动量参数 $\beta_2 = 0.999$ 和稳定常数 $\epsilon = 10^{-8}$ 。批次大小设置为 16。学习率初始化为 10^{-4} ,并在每 2×10^5 个批量更新时减半。经过多轮参数调整,最终将 DIV2K 模型的训练周期设置为 500。

2.4 实验结果与分析

1) 评价指标

为了精确衡量研究中算法的效果,本文采用了两个广泛应用的指标来评估网络重建的性能,分别是峰值信噪比(peak signal to noise ratio, PSNR)和结构相似性(structural similarity, SSIM)。两个指标的数值越高,表示图像重建效果越好。

PSNR 表示峰值信号与噪声的平均能量之比,是评估图像质量的一个指标。其计算公式为:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2 \quad (27)$$

$$PSNR = 10 \cdot \lg\left(\frac{MAX_I^2}{MSE}\right) \quad (28)$$

式中: MSE 表示生成器模型输出的图像与实际高清图像之间的均方误差; MAX_I 表示图像颜色的最大值; m 和 n 表示图像的高宽; $I(i, j)$ 表示坐标 (i, j) 的标签图像像素强度值; $K(i, j)$ 表示坐标 (i, j) 的重建图像像素强度值。

SSIM 是基于人类视觉系统评估图像质量的指标,主要评价图像退化对结构信息的影响。综合考虑了亮度、对比度等视觉感知因素。其计算公式为:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (29)$$

式中: μ_x 和 μ_y 表示在 x 、 y 方向上的平均值; σ_x 、 σ_y 和 σ_{xy} 表示在 x 和 y 方向上的标准差和协方差; c_1 和 c_2 是用来保持数值稳定性的常数。

2) 客观评价

本文 DHSR 模型与 Bicubic、VDSR、EDSR、CARN、IMDN、MSRN、SwinIR 模型在 5 个公开数据集上进行 2、3、4 倍超分辨率重建的 PSNR 和 SSIM 测试结果如表 1 所示。其中以加粗标记的是本文算法取得的最佳结果,而下划线标记的是次优的结果。由表 1 可以观察到,DHSR 在多个放大倍数下均表现优异。特别是在 Set14 和 Urban100 数据集上,DHSR 在 2、3 和 4 倍放大时展现了显著的性能提升。在 Set14 上 2 倍放大时,DHSR 的 PSNR 为 33.95 dB,微弱领先于 SwinIR;在 3 倍放大时,PSNR 达到 30.58 dB,略高于 SwinIR。在 Urban100 的 4 倍放大中,DHSR 以 26.64 dB 的 PSNR 和 0.803 的 SSIM 维持领先。此外,Manga109 上的结果显示,DHSR 在所有放大倍数下都有出色表现,尤其是在 3 倍放大时,其 PSNR 和 SSIM 分别达到 34.03 dB 和 0.947 6。

表 1 不同算法的定量分析
Table 1 Quantitative analysis of different algorithms

方法	放大倍数	参数量 /K	PSNR(dB) ↑ /SSIM ↑				
			Set5	Set14	BSD100	Urban100	Manga109
Bicubic		—	33.48/0.925 0	30.17/0.868 2	26.80/0.833 1	29.32/0.837 1	30.72/0.921 2
VDSR		665	37.53/0.959 0	33.03/0.910 0	31.90/0.896 0	30.76/0.914 0	37.22/0.975 0
EDSR		1 370	37.99/0.960 4	33.57/0.917 5	32.16/0.899 4	31.98/0.927 2	38.45/0.977 0
CARN	×2	1 592	37.76/0.959 0	33.52/0.916 6	32.09/0.897 8	31.92/0.925 6	38.36/0.976 5
IMDN		694	38.00/0.960 5	33.63/0.917 7	32.19/0.899 6	32.17/0.928 3	38.88/0.977 4
MSRN		630	38.08/0.960 5	33.74/0.917 0	32.23/0.901 3	32.22/0.932 6	38.82/0.986 8
SwinIR		878	38.14/0.961 1	33.86/0.920 6	32.31/0.901 2	32.76/0.934 0	38.77/0.977 1
本文算法		1 532	38.09/0.960 9	33.95/0.921 7	32.37/0.901 4	32.72/0.934 2	38.77/0.977 3
Bicubic		—	30.91/0.877 0	27.55/0.774 2	27.21/0.738 5	24.40/0.734 5	26.95/0.855 6
VDSR		665	33.67/0.921 0	29.78/0.832 0	28.83/0.799 0	27.14/0.829 0	32.01/0.934 0
EDSR		1 500	34.37/0.927 0	30.28/0.841 7	29.09/0.805 2	28.15/0.852 7	33.49/0.943 8
CARN	×3	1 592	34.29/0.925 5	30.29/0.840 7	29.06/0.803 4	28.06/0.849 3	33.50/0.944 0
IMDN		703	34.36/0.927 0	30.32/0.841 7	29.09/0.804 6	28.17/0.851 9	33.61/0.944 5
MSRN		630	34.38/0.926 2	30.34/0.839 5	29.08/0.804 1	28.08/0.855 4	33.44/0.942 7
SwinIR		886	34.62/0.928 7	30.54/0.846 0	29.20/0.807 7	28.66/0.862 4	33.95/0.946 9
本文算法		1 555	34.65/0.929 0	30.58/0.846 3	29.24/0.808 3	28.82/0.865 7	34.03/0.947 6
Bicubic		—	28.37/0.799 7	25.95/0.700 7	26.90/0.664 0	23.14/0.657 7	24.80/0.786 6
VDSR		665	31.13/0.883 0	27.95/0.768 0	27.29/0.726 0	25.18/0.754 0	28.83/0.870 7
EDSR		1 500	32.46/0.896 8	28.80/0.787 6	27.71/0.742 0	26.64/0.803 3	31.02/0.914 8
CARN	×4	1 592	32.13/0.893 7	28.60/0.780 6	27.58/0.734 9	26.07/0.783 7	30.47/0.908 7
IMDN		715	32.21/0.894 8	28.58/0.781 1	27.56/0.735 3	26.04/0.783 8	30.45/0.907 5
MSRN		630	32.07/0.890 3	28.60/0.775 1	27.52/0.727 3	26.04/0.789 6	30.17/0.903 4
SwinIR		897	32.44/0.897 6	28.77/0.785 8	27.69/0.740 6	26.47/0.798 0	30.97/0.914 2
本文算法		1 581	32.49/0.898 1	28.80/0.788 1	27.73/0.742 0	26.64/0.803 4	31.08/0.915 4

这些数据表明,与其他先进算法相比,本文提出的算法不仅在单一的指标或测试集上展现出优越性,且在各种场景和条件下都保持卓越和稳定的性能。结合定性实验,本文算法重建图像与高分辨率图像之间差距最小,在实际应用中具有很高的价值。

3) 主观评价

本文在 5 个标准数据集上对自然图片进行了 2~4 倍超分辨率测试。对比结果如图 7~9 所示,可以观察到,在 Set14 中的 baboon 图像中,DHSR 在毛发和面部纹理恢复上,较 Bicubic 和 VDSR 在细节和颜色还原上展现了出色的平衡,且呈现更规则纹理。对于 Urban100 的 img062 图像,特别关注建筑物的窗户和红色雕塑的细节。虽然 DHSR 在红色雕塑的恢复上与其他模型恢复效果大致相同,但其在建筑物窗户的网格信息与纹理走向上与 SwinIR 和 MSRN 相比能有较好的细节特征表示。在 Set5 中的 woman 图像,DHSR 在头饰恢复方面尤为突出,亮度、对比度、色彩及针织纹理效果均趋近原图,且相较其他模型有更少模糊和伪影。BSD100 中的飞机图像中,DHSR 对于飞机身上的“A”标志的恢复与其他方法持平但飞机机翼和尾翼的恢复上略胜于其他方法,且明显重建效果优于

IMDN、MSRN、SwinIR,减少了几何结构失真的现象。此外,在 Manga109 漫画数据集中,AisazuNihaIrena 图像中人物脸部、耳朵和头发上的颜色与线条细节相对重要。Bicubic 方法有明显的模糊现象,而 EDSR、CARN、IMDN 和 SwinIR 等相对恢复了更多的细节。但 EDSR 和 DHSR 在线条的清晰度和连续性上表现得更好。尤其是 DHSR,其重建的头发细节与 HR 图像接近,线条既清晰又连续,且整体颜色恢复上更有动漫的蜡像感。对于 ReveryEarth 图像,角色的头发与配饰上相对重要。Bicubic 的处理结果同样显得模糊。其他模型虽然增强了某些部分的清晰度,但在头发线条和配饰的轮廓上还存在一些不足。DHSR 在配饰的轮廓和头发线条上,其清晰度和连续性都与原图相当,具有较好的重建效果。

2.5 遥感数据实验设计

为了探究模型在遥感数据方面的通用性和性能,本文选取了 AID 数据集^[22]作为实验的基础。AID 数据集来源于 Google Earth,其中包含了 30 个不同类别的遥感场景图像,总共约 10 000 张,每个类别大约有 220~420 张图像。每张图像的像素尺寸为 600×600。为了进行测试,

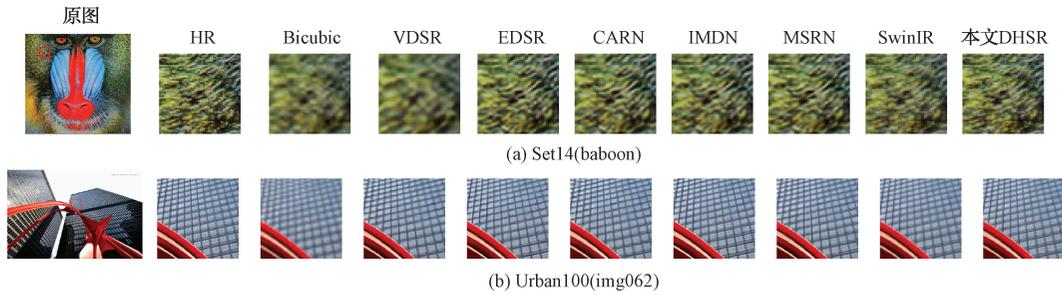


图7 不同模型2倍重建结果对比

Fig.7 Comparison chart of $2\times$ reconstruction results of different models

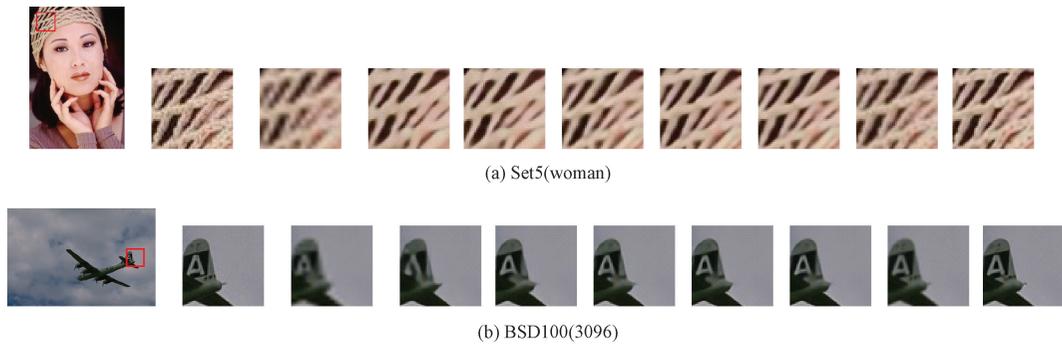


图8 不同模型3倍重建结果对比

Fig.8 Comparison chart of $3\times$ reconstruction results of different models

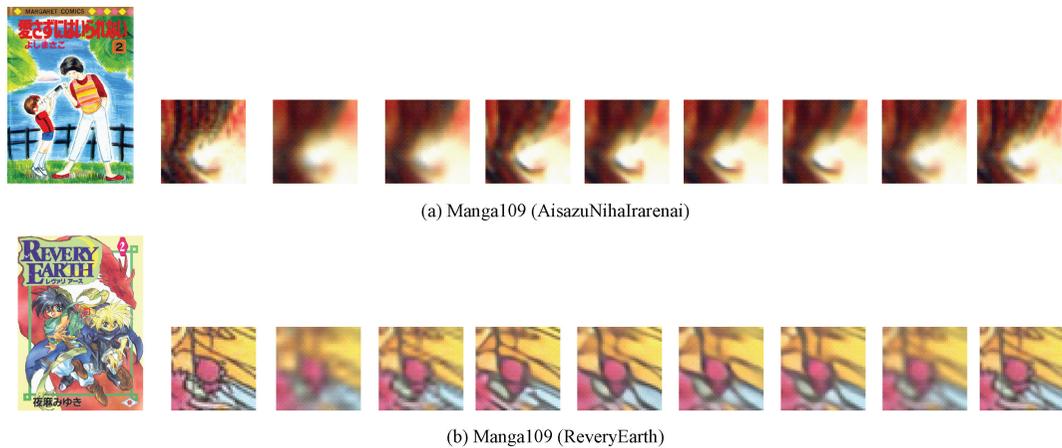


图9 不同模型4倍重建结果对比

Fig.9 Comparison chart of $4\times$ reconstruction results of different models

本文从每个类别中随机抽取5张图像,共计150张,组成了专门的测试集。通过这种方式,可以确保测试结果的多样性和代表性,同时也能全面评估DHSR模型在处理遥感数据方面的效果。

在AID遥感数据集上对不同超分辨率算法进行定量分析的结果如表2所示。在2、3和4倍放大比例下,Bicubic插值法以及各超分辨率方法VDSR、EDSR、CARN、IMDN、MSRN、SwinIR和本文提出的DHSR算

法的PSNR和SSIM指标。从表2可以看出,对于2和4倍的放大比例,本文算法在PSNR和SSIM两个指标上均优于其他方法,其中在2倍放大时,DHSR的PSNR和SSIM指标分别达到37.07 dB和0.9427,而在4倍放大时达到30.81 dB和0.8063。尤其在3倍放大情况下,DHSR算法的PSNR和SSIM结果为33.02 dB和0.8712,相比于次优的SwinIR模型有0.070 dB和0.0005的提升,进一步证实了本文算法在处理遥感图像

数据上同样具有通用性和优越性。尤其在更高的放大比例下,凸显了 DHSR 算法在维持遥感图像细节和纹理方

面的能力。证明了本文算法在提升遥感图像质量方面的潜力。

表 2 遥感数据集定量分析

Table 2 Quantitative analysis of remote sensing datasets

放大倍数	PSNR(dB) ↑ /SSIM ↑							
	Bicubic	VDSR	EDSR	CARN	IMDN	MSRN	SwinIR	本文算法
×2	35.159/0.900 3	36.478/0.934 4	36.721/0.937 8	36.982/0.940 2	37.995/0.941 7	37.002/0.941 9	37.015/0.942 4	37.071/0.942 7
×3	31.261/0.803 6	32.045/0.833 9	32.489/0.846 2	32.874/0.856 8	32.901/0.860 1	32.925/0.869 5	32.954/0.869 7	33.024/0.871 2
×4	29.331/0.719 5	29.852/0.742 9	30.801/0.805 7	30.398/0.762 7	30.564/0.772 7	30.639/0.779 2	30.669/0.803 7	30.811/0.806 3

AID 遥感数据集上 3 倍超分辨率重建的结果对比如图 10 所示。由图 10(a)可以看出,DHSR 算法明显改善了对飞机表面的纹理和细节,与 Bicubic 和 VDSR 相比,提供了更加锐利和连贯的视觉效果。相对于 SwinIR 算法,DHSR 在细节的锐度和连续性方面也表现出了明显的优势。由图 10(b)可以看出,DHSR 算法在建

筑物屋顶纹理和小车细节的恢复上尤为出色,展现了对复杂场景中细节的精确捕捉能力,维持建筑物和车辆的形状与纹理信息。这些对比结果验证了 DHSR 算法在维持遥感图像的细节和纹理方面同样具有显著优势,适合于对图像细节有高度要求的遥感图像分析应用。

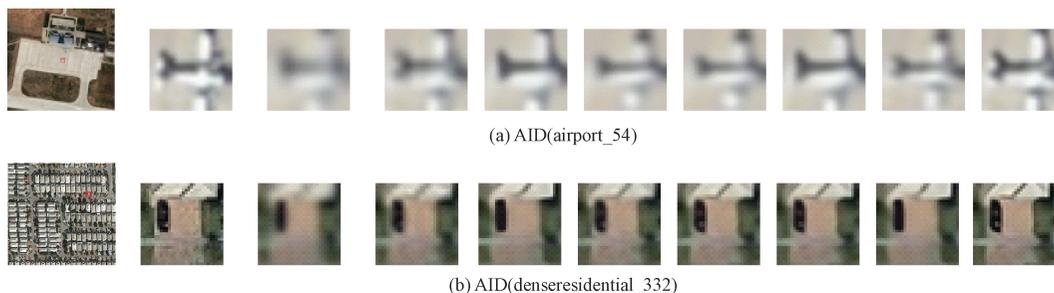


图 10 3 倍遥感图像重建结果对比

Fig. 10 Comparison of 3× remote sensing image reconstruction results

2.6 消融实验结果

为更好的验证提出的算法对原有网络优化的有效性,设计了如下消融实验进行模块分析,且为保证实验的准确性,所有实验的训练周期数均设置为 300 次,其中基准的 SR 网络为本文 DHSR 网络结构,在此基础上做模块增添、替换、修改操作。

- 1) 仅用基准的 SR 网络模型进行学习,将 DHAM 模块用普通 3×3 卷积块进行代替。
- 2) 在上述基准策略上,引入 SimAM 注意力机制模块。
- 3) 保持基准策略不变,加入 DHAM 模块,但移除其中的 MFFM 和 SLU 模块。
- 4) 结合第 3 种策略,加入 MFFM 模块,去掉 SLU 模块。
- 5) 基于第 3 种策略,同时加入 MFFM、SLU 模块。
- 6) 全部策略结合到基准 SR 网络模型。

上述 6 种消融实验都以 Set14×2 图像集的重建结果作为评估标准。

消融实验如表 3 所示,可以看出各个模块对模型性能的积极影响。基准 SR 网络模型的 PSNR 为 33.31 dB 和

SSIM 为 0.914 9。在引入 SimAM 模块后,PSNR 和 SSIM 分别提高了 0.007 dB 和 0.004,突显了特征重加权在提高图像细节准确性上的作用。当加入 DHAM 模块,重建结果提高了 0.45 dB 的 PSNR 和 0.28 的 SSIM,模型性能得到显著提升,证明了其在增强特征提取上的有效性。在 DHAM 基础上 MFFM 和 SLU 模块的加入使重建结果继续提升,还确保了更均衡和细致的图像恢复。最终,基于基准 SR 网络模型加入完整策略达到最高的 33.95 dB 的 PSNR 和 0.921 7 的 SSIM,这些数值结果均证实了所提方法在图像超分辨率重建任务上的显著优势。

2.7 注意力机制性能分析

本文通过在 Set14×3 数据集上的消融实验,评估了 DHAM 模块在超分辨率重建的性能。实验对比了不同数量的 DHAM 模块和对称卷积组模块的配置对 PSNR 和 SSIM 指标的影响。实验结果如表 4 所示,其中数字代表使用的 DHAM 模块的数量。

实验结果表明,使用 1 个 Conv3×3 卷积时,每增加一个 DHAM 模块,PSNR 指标平均提升约 0.8,同时 SSIM 指标提高约 0.001 5。当使用 2 个 Conv3×3 卷积和 4 个 DHAM 模块时,性能显著提升,PSNR 达到 30.58 dB,

表 3 消融实验结果

Table 3 Ablation experiment results

Baseline	SimAM	DHAM	MFFM	SLU	PSNR/dB ↑	SSIM ↑
✓					33.31	0.914 9
	✓				33.38	0.915 3
		✓			33.76	0.917 7
		✓	✓		33.82	0.919 4
		✓		✓	33.79	0.918 0
		✓	✓	✓	33.89	0.920 6
✓	✓	✓	✓	✓	33.95	0.921 7

表 4 DHAM 和对称卷积组模块数量探究

Table 4 Research on the number of DHAM and symmetric convolution group modules

Set × 3	(PSNR(dB) ↑ / SSIM ↑)		
	1Conv 3 × 3	2Conv 3 × 3	3Conv 3 × 3
2DHAM	30.30/0.839 2	—	—
4DHAM	30.41/0.842 0	30.58/0.846 3	—
6DHAM	30.45/0.843 2	—	30.50/0.843 9

SSIM 达到 0.846 3,所有配置中的最佳表现。然而,当使用 3 个 Conv3 × 3 卷积和 DHAM 模块数量增至 6 个时,性能反而略有下降,PSNR 降至 30.50 dB,SSIM 降至 0.843 9。这表明虽然更多的 DHAM 模块可能提供更复杂的特征提取能力,但同时可能引入冗余或过拟合问题,导致性能降低。因此,选择 2 个 Conv3 × 3 卷积和 4 个 DHAM 模块为本文 SR 网络中的最佳配置。

3 结 论

本文提出的 DHSR 模型通过层次交互动态注意力机制与序列学习解决了图像超分辨率重建的关键问题。模型通过多层次特征融合与跳跃连接加速收敛,并通过动态注意力机制与多尺度特征融合精准捕获关键信息,实现细节恢复。在公开数据集(Set5、Set14、BSD100、Urban100、Manga109)上的实验可以看出,DHSR 性能卓越,特别是在峰值信噪比和结构相似性指标上优于现有算法。此外,DHSR 在遥感图像超分辨率任务中表现出良好的通用性和稳定性。尽管如此,DHSR 在处理极低分辨率或高噪声图像时面临挑战,未来工作将致力于提高模型的鲁棒性,并探索其在特定领域如医学影像分析的应用潜力。DHSR 模型的成功为图像超分辨率重建领域提供了新的研究方向和技术框架。

参 考 文 献

[1] 谢巧雪,马宗庆,祝连庆,等. 基于结构相似约束生成对抗网络的视网膜 OCT 图像去噪算法[J]. 电子测量与仪器学报, 2023, 37(3): 11-20.

XIE Q X, MA Z Q, ZHU L Q, et al. Retinal OCT image denoising algorithm based on structural similarity constrained adversarial network[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(3): 11-20.

[2] 程德强,陈杰,寇旗旗,等. 融合层次特征和注意力机制的轻量化矿井图像超分辨率重建方法[J]. 仪器仪表学报, 2022, 43(8): 73-84.
CHEN D Q, CHEN J, KOU Q Q, et al. A super-resolution reconstruction method for lightweight mine images by fusing hierarchical features and attention mechanism [J]. Chinese Journal of Scientific Instrument, 2022, 43(8): 73-84.

[3] 杨梦薇,肖秦琨,朱毅琳. 基于深度残差网络的图像超分辨率重建[J]. 国外电子测量技术, 2022, 41(4): 170-175.
YANG M W, XIAO Q K, ZHU Y L. Image super-resolution reconstruction based on deep residual network [J]. Overseas Electronic Measurement Technology, 2022, 41(4): 170-175.

[4] 李晨霖,王仕成,张金生,等. 基于改进的 Kriging 插值方法构建地磁基准图[J]. 计算机仿真, 2018, 35(12): 278-282.
Li CH L, WANG SH CH, ZHANG J SH, et al. Construction of geomagnetic reference map based on improved Kriging interpolation method[J]. Computer Simulation, 2018, 35(12): 278-282.

[5] DONG C, LOY C C, TANG X. Accelerating the super-resolution convolutional neural network [C]. Computer Vision-ECCV 2016: 14th European Conference, 2016: 391-407.

[6] KIM J, LEE J K, LEE K M. Accurate image super-resolution using very deep convolutional networks[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 1646-1654.

[7] LIM B, SON S, KIM H, et al. Enhanced deep residual networks for single image super-resolution[C]. IEEE Conference on Computer Vision and Pattern

- Recognition Workshops, 2017: 136-144.
- [8] AHN N, KANG B, SOHN K A. Fast, accurate, and lightweight super-resolution with cascading residual network [C]. European conference on computer vision (ECCV), 2018: 252-268.
- [9] ZHENG H, GAO X, YANG Y, et al. Lightweight image super-resolution with information multi-distillation network [C]. 27th ACM International Conference on Multimedia, 2019: 2024-2032.
- [10] LI J, FANG F, MEI K, et al. Multi-scale residual network for image super-resolution [C]. European Conference on Computer Vision (ECCV), 2018: 517-532.
- [11] LIANG J, CAO J, SUN G, et al. SWINIR: Image restoration using swin transformer [C]. IEEE/CVF International Conference on Computer Vision, 2021: 1833-1844.
- [12] LI Z, LI G, LI T, et al. Information-growth attention network for image super-resolution [C]. 29th ACM International Conference on Multimedia, 2021: 544-552.
- [13] HUANG G, LIU Z, VAN D, et al. Densely connected convolutional networks [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4700-4708.
- [14] 辛元雪, 朱凤婷, 史朋飞, 等. 基于改进增强型超分辨率生成对抗网络的图像超分辨率重建算法[J]. 电子测量技术, 2022, 45(4): 126-133.
- XIN Y X, ZHU F T, SHI P F, et al. Image super-resolution reconstruction algorithm based on improved enhanced super-resolution generative adversarial network[J]. Electronic Measurement Technology, 2022, 45(4): 126-133.
- [15] QIN X, LI N, WENG C, et al. Simple attention module based speaker verification with iterative noisy label detection [C]. ICASSP 2022 — 2022 IEEE International Conference on Acoustics, 2022: 6722-6726.
- [16] YANG L, ZHANG R Y, LI L, et al. Simam: A simple, parameter-free attention module for convolutional neural networks [C]. International Conference on Machine Learning, 2021: 11863-11874.
- [17] CHEN Y, DAI X, LIU M, et al. Dynamic convolution: Attention over convolution kernels [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11030-11039.
- [18] WANG L, WANG Y, LIANG Z, et al. Learning parallax attention for stereo image super-resolution [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 12250-12259.
- [19] ZHU X, CHENG D, ZHANG Z, et al. An empirical study of spatial attention mechanisms in deep networks [C]. IEEE/CVF International Conference on Computer Vision, 2019: 6688-6697.
- [20] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 1-10.
- [21] KINGMA D P, BA J. Adam: A method for stochastic optimization [J]. ArXiv preprint ArXiv: 1412.6980, 2014.
- [22] XIA G S, HU J, HU F, et al. AID: A benchmark data set for performance evaluation of aerial scene classification [J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(7): 3965-3981.

作者简介

莫恒辉, 硕士研究生, 主要研究方向为计算机视觉、机器学习。

E-mail: 973733507@qq.com

魏霖静, 博士, 教授, 主要研究方向为农业大数据、智能计算。

E-mail: wlj@gsau.edu.cn