

全局-局部注意力引导的红外图像恢复算法

刘晓朋^{1,2}, 张涛^{1,2}

(1. 江南大学 人工智能与计算机学院, 江苏 无锡 214122; 2. 中国船舶科学研究中心, 江苏 无锡 214122)

摘要: 针对真实世界的红外图像恢复算法中存在的图像模糊、纹理失真、参数过大等问题, 提出了一种用于真实红外图像的全局-局部注意力引导的超分辨率重建算法。首先, 设计了一种跨尺度的全局-局部特征融合模块, 利用多尺度卷积和 Transformer 并行融合不同尺度的信息, 并通过可学习因子引导全局和局部信息的有效融合。其次, 提出了一种新颖的退化算法, 即域随机化退化算法, 以适应真实红外场景图像的退化域。最后, 设计了一种新的混合损失函数, 利用权重学习和正则化惩罚来增强网络的恢复能力, 同时加快收敛速度。在经典退化图像和真实场景红外图像上的测试结果表明, 与现有方法相比, 该算法恢复的图像纹理更逼真, 边界伪影更少, 同时参数总数最多可减少 20%。

关键词: 域随机化退化算法; 跨尺度融合; 红外图像超分辨率; 生成对抗网络

中图分类号: TP394.1 文献标识码: A 文章编号: 1001-8891(2024)07-0791-11

Global-Local Attention-Guided Reconstruction Network for Infrared Image

LIU Xiaopeng^{1,2}, ZHANG Tao^{1,2}

(1. School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China;

2. China Ship Scientific Research Center, Wuxi 214122, China)

Abstract: To solve the problems of image blur smoothing, texture distortion, and excessively large parameters in real-world infrared-image recovery algorithms, a global-local attention-guided super-resolution reconstruction algorithm for infrared images is proposed. First, a cross-scale global-local feature fusion module utilizes multi-scale convolution and a transformer to fuse information at different scales in parallel and to guide the effective fusion of global and local information by learnable factors. Second, a novel domain randomization degradation model accommodates the degradation domain of real-world infrared images. Finally, a new hybrid loss based on weight learning and regularization penalty enhances the recovery capability of the network while speeding up convergence. Test results on classical degraded images and real-world infrared images show that, compared with existing methods, the images recovered by the proposed algorithm have more realistic textures and fewer boundary artifacts. Moreover, the total number of parameters can be reduced by up to 20%.

Key words: domain randomization degradation algorithm, cross-scale fusion, infrared image super-resolution, generative adversarial network

0 引言

单图像超分辨率 (Super-Resolution, SR) 是近年来一个活跃的研究课题, 旨在从低分辨率 (Low-Resolution, LR) 图像中重建出高分辨率 (High-Resolution, HR) 图像^[1-2], 因其具有较高的实用价值而受到广泛关注。事实上, HR 图像包含了更清晰的纹理特征和更多的细节信息, 在监控设备、红外检测和医学影像领域^[3-5]都有广泛的应用。随着卷积神经网络

(Convolutional Neural Networks, CNN) 的发展。2016 年, Dong 等人首次提出了 (Super-Resolution Convolutional Neural Network, SRCNN)^[6] 用于图像恢复, 此后, 基于 CNN 的方法在 SR 领域上迎来了蓬勃的发展^[7-10]。

早期基于 CNN 的 SR 模型主要专注于精细的架构设计, 例如残差学习^[11], 亚像素卷积^[12], 多尺度融合^[13]和密集连接^[14]等, 而忽略了图像退化的复杂性。尽管与传统模型方法相比, 这些方法在峰值信噪比

收稿日期: 2023-02-26; 修订日期: 2023-03-31。

作者简介: 刘晓朋 (1998-), 男, 陕西汉中人, 硕士研究生, 主要从事深度学习, 图像处理。E-mail: 6201910027@stu.jiangnan.edu.cn。

基金项目: 船舶总体性能创新研究开放基金项目 (14422102)。

(Peak Signal-To-Noise Ratio, PSNR) 方面取得了优越的性能, 但它们仍然无法在真实世界的红外图像上取得良好的效果。主要原因是生成的图像过于平滑, 看起来不太真实。首先, 这可能是因为传统的基于 CNN 的方法对全局依赖建模效果不佳, 无法适应现实世界红外图像中急剧变化的全局-局部信息。其次, 现有的 SR 方法都使用固定的退化模型, 退化过程相对简单, 退化类型单一。但真实场景的退化类型和数据场景要复杂得多, 退化过程是未知的。因此现有的模型无法模拟现实世界中复杂变换的红外图像。此外, 作为 CNN 的替代方案, 基于 Transformer 结构^[15-16]的图像恢复模型也在 SR 领域蓬勃发展。这些方法通常将输入图像划分为固定大小的块, 并独立处理每个块。然而, 它们也存在新的问题。首先, 恢复的图像会在每个小块周围引入边界伪影。其次, 相邻的图像块之间不能生成局部依赖。这使得生成的图像存在局部伪影的问题。

在这项工作中, 我们的目标是学习一种更强大的图像超分辨率模型, 该模型可以广泛应用于现实世界的图像。首先, 我们提出了一种新的图像退化算法, 它结合了随机洗牌策略^[17]和二次退化方法^[18]来构建随机模型, 并对图像的不同区域进行随机退化。它可以自动适应图像的退化域, 以更好地模拟真实世界的红外图像。随后, 为了适应退化空间的随机性, 我们构建了一个稳健的 SR 模型, 命名为全局-局部注意力引导超分辨率网络 (Global-Local Attention-Guided Super-Resolution, GLAGSR), 它由一个生成器 (Generator) 和一个鉴别器 (Discriminator) 组成。具体来说, 在生成器中引入了 Swin Transformer 块和多尺度卷积块, 通过聚合全局和局部的特征来增强细节并生成逼真的纹理。鉴别器采用更强大的 Transformer 结构, 通过捕捉图像不同区域之间的关系, 准确反映低级纹理特征。最后, 考虑到现有损失函数的局限性, 我们探索了一种新的混合损失函数, 它结合了梯度损失、像素损失、感知损失和 GAN (Generative Adversarial Network) 损失, 使用解码器-编码器映射来学习权重, 并利用正则化惩罚来平衡模型的损失和收益, 稳定训练。基于这些思想, 我们训练的 GLAGSR 模型, 在复杂的随机退化图像中实现局部细节增强和随机退化之间的平衡。

本文的贡献如下: ①为真实世界的红外图像提供了一种全局-局部注意力引导的超分辨率网络。特别是, 构建了一个结合多尺度卷积模块和 Swin Transformer 模块的跨尺度结构, 旨在捕获全局和局部信息。②提出了一种新的图像退化算法, 即域随机化

退化算法, 它针对图像的不同区域使用随机退化模型来适应现实世界红外图像的退化域。③探索了一种新的混合损失函数, 它结合了梯度损失、像素损失、感知损失和 GAN 损失项, 以提高捕获全局-局部信息的能力。特别是, 还引入了权重学习和正则化惩罚项, 以平衡模型的损失和收益, 同时稳定训练。

1 相关工作

1.1 图像超分辨率

近年来, 随着深度学习的蓬勃发展, 基于学习的方法 (基于 CNN 的方法) 在 SR 领域逐渐取代了传统的基于模型的方法。2016 年, Dong 等人首次提出了 SRCNN^[6]。此后 SRResNet^[19] (Super-Resolution Residual Network), EDSR^[9] (Enhanced Deep Residual Networks for Single Image Super-Resolution) 和 RDN^[14] (Residual Dense Network) 等网络进一步促进了 SR 的发展。然而这些模型只是简单地与原始像素的比较, 生成的图像趋于模糊化^[20]。为了解决这个问题, 研究人员专注于更精细的网络结构, 并尝试设计更好的特征提取块, 例如残差块^[11]、密集块^[14]和多尺度块^[13, 21]等。为了提高模型的表示能力, 其中一些方法在 CNN 网络中引入了注意力机制, 例如通道注意力^[22], 空间注意力^[23]和非局部注意力^[2, 16]。同时, 为了生成逼真的 SR 图像, Wang 等人在 SR 任务中引入了感知损失^[24], 并使用生成对抗网络来训练 SR 模型。尽管他们进一步提高了图像质量, 但仍然存在局限性。例如, 基于注意力的 CNN 结构无法对全局信息进行建模, 这使得生成的 SR 图像在某些区域出现了扭曲的线条。

作为 CNN 的替代方案, Transformer 也被引入图像恢复^[15-16, 25]领域。Essers 等人使用基于 Transformer 结构的 GAN^[15]来了解图像的上下文丰富的视觉区域, 并设法捕获全局细节, 成功合成了高分辨率图像。但是, 整个架构中存在一些特定的设计模块, 限制了模型的通用性。Jiang 等人将 Transformer 的结构作为 GAN 网络^[16]的主要结构, 完全摒弃了 CNN 结构, 通过数据增强和多任务训练, 模型的最大性能几乎与 GAN 模型相当。然而, 生成的 HR 图像会丢失局部纹理信息, 无法反馈真实世界的高分辨率图像。在最近的工作中, Liu 等人提出了 Swin Transformer^[25]并设计了一种滑动窗口机制来捕捉全局信息中的局部细节信息。这种技术结合了 CNN 和 Transformer 的优点。但是, 在平衡全局依赖和局部增强方面仍然存在问题, 不能很好地融合全局-局部特征。为了解决这些问题, 本文提出了一个跨尺度特征融合块, 它聚合全局和局部特征并自动平衡全局-局部信息依赖关系以生

成更高质量的 SR 图像。

1.2 退化模型

在 SR 领域, 早期退化模型大多采用传统的双三次下采样^[19, 24, 26]。尽管它们在经典的退化数据集上取得了很好的效果, 但它们无法适应广泛的真实红外数据集。事实上, 真实世界的退化类型和数据场景要复杂得多。对于特定退化的超分模型采用任意的 LR 输入, 就会出现较大的域差异, 进而产生较差的重建结果^[27]。此外, 同一幅真实图像也存在不同的退化方式, 没有统一的退化范式。为了进一步缩小 HR 图像与真实世界图像之间的差距, 近年来, 许多研究人员将注意力集中在未知退化模型的学习上。

现有的未知退化模型的方案可以大致分为显式建模和隐式建模。显式建模方法通常由模糊、下采样、噪声和 JPEG (Joint Photographic Experts Group) 压缩叠加组成。在最新的研究中, Zhang 等提出了一种随机洗牌的策略^[17]来合成更实用的退化图像。然而, 退化过程的数量是固定的, 基于退化模型生成的 SR 图像往往会在纹理区域产生“气泡”伪影。在最新的研究中, Wang 等人提出了一个高阶退化建模策略^[18]来合成退化图像。然而, 一些恢复的图像有扭曲的线条, 特别是在建筑和室内场景中。隐式建模方法则使用训练数据来学习 LR-HR 映射, 试图获取图像的退化过程, Cai 等人建立了具有配对 LR/HR 图像的 SISR (Single-Image Super-Resolution) 数据集^[12], 但该方法仅适用于训练图像定义的退化, 不能推广。Lugmayr 等人^[27]提出通过使用一致性损失循环来学习深度退化映射, 然后生成用于监督训练的 LR/HR 对。然而, 收集数据的过程很复杂, 学到的模型仅限于这些 LR 图像域, 不能很好地泛化。换言之, 现有的退化模型在处理真实图像退化的复杂性方面仍然不足。在这项工作中, 我们提出了一个灵活的域随机化退化模型。该模型合成的 LR 图像用于覆盖更广泛的红外图像退化空间, 为后续训练 SR 网络提供广泛的样本。

2 图像重建算法框架

我们的主要目标是学习一种更适合真实世界红外图像的新型 SR 网络, 同时尽可能提高 SR 图像的整体感知质量。在这一部分, 首先描述了提出的新型退化模型。然后, 描述了提出的全局-局部注意力引导的超分辨率网络, 重点讨论了创新模块中特征融合的细节。接下来, 给出每个组件的详细信息。

2.1 退化模型

目前, 显示建模中通常采用广义的经典退化模型。它确定了一个固定的退化序列{模糊、下采样、噪

声、JPEG 压缩}。该模型表示为:

$$L_x = D(H_y : \theta) = \{(H_y \otimes k) \downarrow_{\theta} + n\}_{\text{JPEG}} \quad (1)$$

式中: L_x 代表 LR 图像; D 代表退化函数; H_y 代表对应的 HR 图像; θ 代表退化过程的参数。

但是这种固定的退化过程限制了退化模型的表达能力, 图像的退化范围被限制在固定的区间内。此外, 现实世界的图像可能存在区域退化, 即图像的不同区域以不同方式和不同程度退化。然而, 目前传统的退化模型对同一幅图像的所有区域都使用相同的退化模型, 这意味着图像退化的方向是单一的, 无法模拟真实世界 LR 图像中的复杂变换。

为了更好地模拟红外图像退化过程, 我们将构建一个强大的退化模型, 它可适应红外图像空间的退化。具体地说: 首先, 我们使用随机混洗策略改进固定退化序列{模糊、下采样、噪声、JPEG 压缩}, 它利用随机算法动态更新参数(例如随机选择模糊核的类型、下采样方法、噪声类型和 JPEG 压缩质量因子的大小等)来改变退化过程。其次, 考虑到图像可能会经历多次退化, 我们引入二阶退化建模方案, 对图像进行两次随机退化, 生成随机二阶退化模型。最后, 考虑到图像存在区域退化, 即不同区域有不同程度的退化, 我们随机划分图像的区域, 对这些区域使用随机二阶退化模型。基于这些操作, 我们设计了一种区域随机退化算法。

值得注意的是, 由于随机性相同的退化模式可能出现在图像的不同区域, 同一区域的两个退化过程也可能相同。这意味着其他退化模型, 例如传统的退化模型和 BSRGAN^[17] (Blind Image Super-Resolution Generative Adversarial Network) 以及 Real-ESRGAN^[18] (Training Real-World Blind Super-Resolution with Pure Synthetic Data) 中提出的退化模型, 都是我们的特例。

我们构建的退化模型可以表示为:

$$D''(X : \theta) = \text{Random} \{(X \otimes k) \downarrow_{\theta} + n\}_{\text{JPEG}} \quad (2)$$

$$L_x = \begin{cases} (D^2 \circ D^1)(X_1 : \theta) & X_1 \in [t_1 : t_2, k_1 : k_2] \\ (D^2 \circ D^1)(X_2 : \theta) & X_2 \in [t_1 : t_2, k_1 : k_2] \end{cases} \quad (3)$$

式中: D^2 代表二阶退化; D^1 代表一阶退化; Random 操作表示使用随机因素调节传统退化建模过程的随机化策略; X 代表输入的高分辨率图像; X_1 和 X_2 分别代表图像 X 不同区域的, t_1 、 t_2 、 k_1 和 k_2 分别表示图像不同区域的隶属边界。

图 1 为提出的退化模型。对于 HR 图像, 随机选

择一个区域（区域的大小和位置是随机设置的）并通过随机调整退化序列{模糊、下采样、噪声、JPEG压缩}并设置不同的参数来对其进行退化。重新设置其他剩余区域的参数以生成新的退化模型。使用上述退化算法，将能够在 HR 图像的大型数据集进行执行，以生成无限的、完美对齐的训练图像 HR/LR 对，为后续的 SR 网络学习提供大量样本，从而更好地恢复真实世界的红外图像。

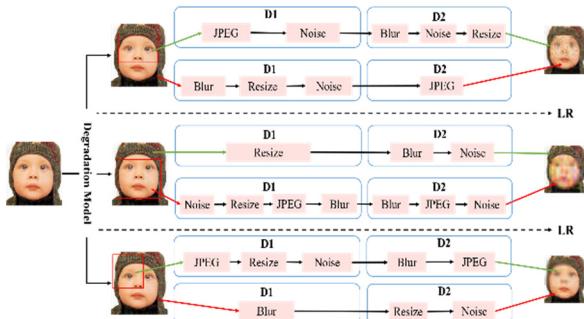


图 1 比例因子为 2 的退化模型的示意图

Fig.1 Schematic illustration of the proposed degradation model for a scale factor of 2

2.2 网络架构

如 2.1 节所述，提出的退化算法合成了具有多样性的 LR 图像，更好地模拟了现实世界中复杂变换的退化图像。这解决了以前的 SR 方法中退化模式和退化空间受限的问题。然而，由于 LR 图像中内容信息的存在巨大变换，现有的网络无法很好地捕获全局-局部信息。为了解决这个问题，提出了一种全局-局部注意引导的超分辨率网络 (GLAGSR)。特别地，设计了一种更精细的结构来自适应地融合不同尺度上的信息。如图 2(左)所示，网络包括一个多尺度生成器和一个 Transformer 鉴别器。

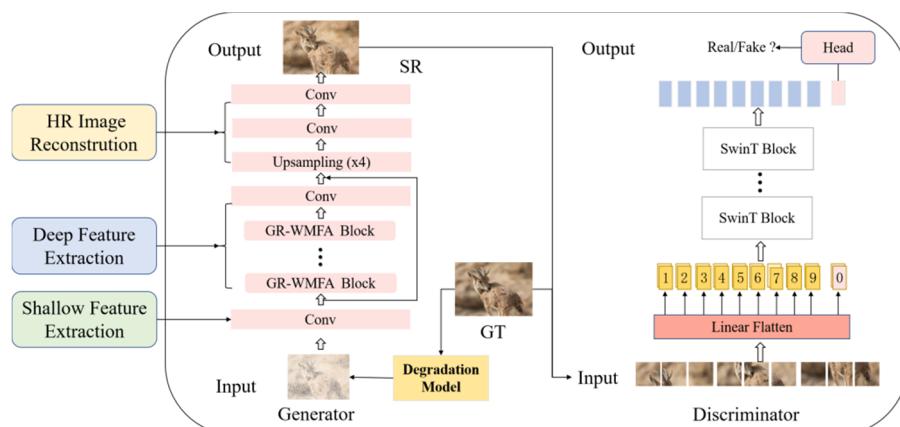


图 2 用于图像恢复的 GLAGSR 的架构

Fig.2 The architecture of the proposed GLAGSR for image restoration

2.2.1 多尺度生成器

本文提出了一种基于全局-局部特征融合的多尺度生成器。该生成器由 3 个模块组成：浅层特征提取模块 (Shallow Feature Extraction)、深层特征提取模块 (Deep Feature Extraction) 和 HR 图像重建模块 (HR Image Reconstruction)。如图 2 所示，通过使用这 3 个模块从 LR 图像中逐步提取信息，生成最终的 SR 图像，该方法的更多细节如下。

浅层特征提取模块：卷积层擅长早期视觉处理，优化更稳定，效果更好^[28]。因此，针对 2.1 节构造的退化模型所合成的 LR 图像 $I_{LR} \in R^{C \times H \times W}$ (其中 C , H 和 W 分别代表图像的通道数、高度和宽度)。我们使用 3×3 的卷积层 $C_{\text{first}}(\cdot)$ 来提取浅层特征 $F_{\text{shallow}} \in R^{C \times H \times W}$ ：

$$F_{\text{shallow}} = C_{\text{first}}(I_{LR}) \quad (4)$$

深层特征提取模块：基于获得的浅层特征 F_{shallow} ，深层特征 $F_{\text{deep}} \in R^{C \times H \times W}$ 可以通过以下方式计算：

$$F_{\text{deep}} = \text{MF}_{\text{feature}}(F_{\text{first}}) \quad (5)$$

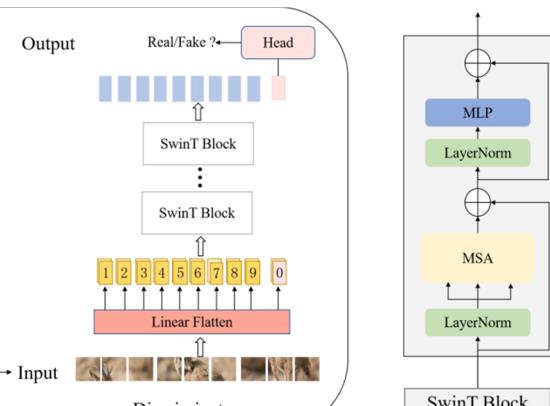
式中： $\text{MF}_{\text{feature}}(\cdot)$ 是一个多尺度特征提取模块。它包含 k 分组的残差全局局部特征融合块 (Grouped Residual Global-local feature fusion, GR-GFF) 和一个 3×3 的卷积层。更具体地说，中间特征 $F_{d1}, F_{d2}, \dots, F_{dk}$ 和输出深度特征 F_{deep} 被逐块提取为：

$$F_{di} = G_{\text{grfi}}(F_{di-1}), \quad i=1,2,3, \dots, k \quad (6)$$

$$F_{\text{deep}} = C_{\text{last}}(F_{dk}) \quad (7)$$

式中： $G_{\text{grfi}}(\cdot)$ 表示第 i 个 GR-GFF 块； C_{last} 代表最后一个卷积层。

特别的是，我们在多尺度特征模块中提出了一个全局-局部特征融合 (Global-local Feature Fusion, GFF) 块。如图 3 所示，采用双分支结构，融合了 CNN 关注局部信息和 Transformer 关注全局信息的优势，将



全局-局部信息进行耦合。首先, 我们引入了一个多尺度卷积块 (Multi-Scale Convolution Block, MC Block) 和一个 Swin Transformer 块 (SwinT Block)。其中 MC Block 采用了不同的卷积核与空洞卷积在整个图像上获得更全的感受野, 用于提取局部特征, 如图 4(a)所示。而 SwinT Block 将图像划分为不同的补丁块, 进行全局特征提取, 如图 2(右)所示。然后在不同的模块中引入注意机制 (attention), 通过聚焦图像通道和 3 种空间信息, 在抑制低频噪声的同时有选择地增强高频细节。其中, “MC Block+attention” 模块更注重局部信息。将其并入一个局部信息选择 (LIS) 模块, 通过该模块可以提取具有丰富局部信息的特征映射 L_f 。相反, “SwinT Block+attention” 模块更注重全局信息。将其并入全局信息引导 (GIG) 模块, 通过该模块可以提取具有丰富全局信息的特征映射。接下来, 利用可学习的权重因子对不同的特征进行权重分配, 并利用参数的可学习性自适应地平衡不同特征的贡献。

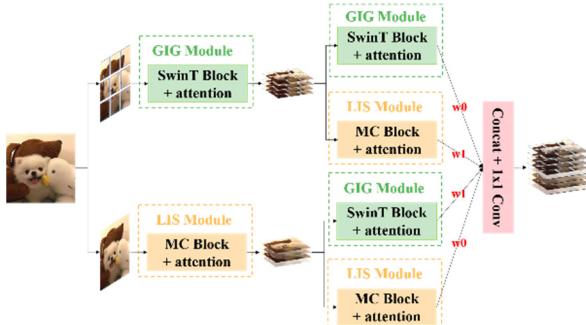


图 3 全局-局部特征融合块

Fig.3 Global-local feature fusion block

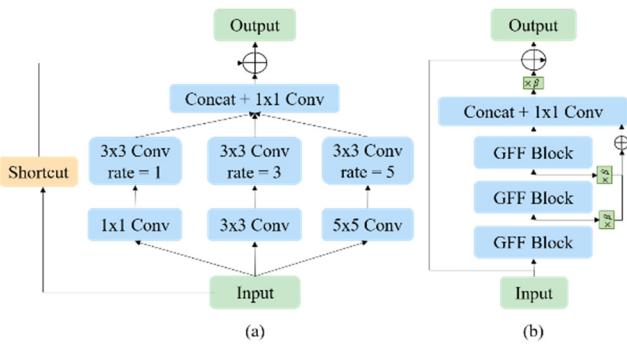


图 4 多尺度卷积块(a)和分组残差 GFF 块(b), β 是残差比例参数

Fig.4 The multi-scale convolution block (a)and the grouped residual GFF block (GR-GFF Block) (b), and β is the residual scaling parameter

因此, 最终整合全局-局部信息形成的特征被描述为:

$$GL_f = w_0(G_{f,1} + L_{f,1}) + w_1(G_{f,2} + L_{f,2}) \quad (8)$$

式中: w_0 和 w_1 平衡不同特征的系数, 为后续全局-局

部特征有更好的聚合。

然后, 为了充分捕捉 LR 图像中包含的丰富低频信息, 受残差块^[11]和密集块^[14]等结构的启发, 我们进一步设计了分组残差 GFF 块 (Grouped Residual GFF, GR-GFF)。如图 4(b)所示, 该结构由 3 个全局-局部融合特征块 GL_f 组成:

$$G_{grfi} = F_{cat}\{GL_{f1} \oplus GL_{f2} \oplus GL_{f3}\} + x_i \quad (9)$$

式中: F_{cat} 是 Channel 拼接层, x_i 是输入。这种设计有两个好处。首先, Transformer 可以看成是空间变化卷积的具体实例化^[28], 而多分支卷积可以在局部空间获得极大的感受野, 二者结合可以完美提取不同的空间特征。其次残差连接提供从不同块到重构模块的基于身份的连接^[25], 允许不同级别的特征的聚合。

HR 图像重建模块: 通过合并浅层和深层特征, 重建高分辨率图像 I_{HR} :

$$I_{HR} = F_{sr}(F_{deep} + F_{shallow}) \quad (10)$$

式中: F_{sr} 是重构模块的函数。这里的重建模块, 使用一种新的上采样方法, 补丁扩展^[25]。它首先将特征划分为不同的块, 然后对这些特征加一个线性层, 将特征维数增加到原始维数的整数倍, 然后, 利用重排操作将输入特征的分辨率扩大到输入分辨率的整数倍, 从而实现上采样。

2.2.2 Transformer 鉴别器

由于退化模型扩展了退化图像的空间, 模型的训练也变得具有挑战性, 因此鉴别器需要更强大的能力来区分真实性和复杂的训练输出。此外, 局部纹理将对鉴别器产生重要影响。为了精确捕捉局部信息, 我们借鉴了 Transgan^[16]的鉴别器结构, 在 Swin Transformer 结构的基础上构建了一个强大的多尺度鉴别器, 如图 2(左)所示。鉴别器的基本块是一个 SwinT 块编码器。我们在每个编码器块递归地构造变换块, 以便在每个阶段都可以提取不同的纹理细节。

首先, 将图像分成 8×8 的小块作为输入, 并且给予首位加上标记 “[cls]”, 然后通过将这些小块编号添加可学习的位置编码。

$$DF_f = F_{flatten}(I_{HR}^* \text{ or } I_{GT}^*) \quad (11)$$

式中: $F_{flatten}$ 是一个展平层 (其用于将输入 “展平”, 即使多维输入变为一维); I_{HR}^* 表示由生成器生成的图像; I_{GT}^* 代表真实图像。然后, 我们将 DF_f 放入 SwinT 块编码器中提取多阶段特征图。更具体地说, 中间特征 $F_{e,0}, F_{e,1}, \dots, F_{e,t}$ 和输出深度特征 DF_l 被逐块提取, 如下:

$$F_{e,0} = DF_f, \quad F_{e,i} = F_{STB,i}(F_{e,i-1}), \quad i = 1, 2, \dots, t \quad (12)$$

$$DF_l = F_{e,t} + DF_f \quad (13)$$

式中: $F_{e,i}$ 是 SwinT 块编码器; DF_l 是最后一个 SwinT

块编码器。特别是, $F_{\text{STB},i}$ 是一个 Swin Transformer 层, 如图 2(右)所示, 它首先将输入特征图划分为 $S \times S$ 不重叠局部窗口以重塑特征输入。然后, 它分别计算每个窗口的标准自注意力, 即多头注意力层 (Multi-head Self-Attention, MSA)。具体来说, 对于局部窗口特征 $F_f \in R^{S^2 \times C}$ (C 是图像通道数), 查询矩阵 \mathbf{Q} 、键矩阵 \mathbf{K} 和值矩阵 \mathbf{V} 计算如下:

$$\mathbf{Q} = F_f \mathbf{H}_Q, \mathbf{V} = F_f \mathbf{H}_V, \mathbf{K} = F_f \mathbf{H}_K \quad (14)$$

式中: \mathbf{H}_Q 、 \mathbf{H}_V 和 \mathbf{H}_K 是在不同窗口之间共享的投影矩阵。随后, 通过局部窗口中的自我注意机制计算注意矩阵:

$$A(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}(\mathbf{Q}\mathbf{K}^\top \sqrt{d} + b)\mathbf{V} \quad (15)$$

式中: A 代表注意力; b 是一种可学习的相对位置编码。然后, 使用多层感知器 (Multi-layer Perceptron, MLP) 进行进一步的特征变换, 该多层感知器具有两个完全连接的层, 其间具有 GELU 非线性。在 MSA 和 MLP 之前增加了层归一化 (LayerNorm, LN) 层, 两个模块都采用了剩余连接。整个过程表示为:

$$F_{i-1} = \text{MSA}(\text{LN}(F_{e,i-1})) + F_{e,i-1} \quad (16)$$

$$F_{\text{STB},i} = \text{MLP}(\text{LN}(F_{i-1})) + F_{i-1} \quad (17)$$

最后, 将提取的特征映射 DF_{last} 通过 SwinT 块编码器后, 由分类头获得标签 [CLS], 并最终输出真假预测:

$$D_f = \text{DF}_{\text{last}}[\text{cls}] \quad (18)$$

2.3 损失函数

在 SR 领域, 用于指导网络优化的损失函数有很多, 如像素损失 L_1 ^[22]、GAN 损失 L_g ^[24]、感知损失 L_p ^[17, 24] 等。虽然这些方法能够获得较高的 PSNR 值, 但恢复的图像过于平滑, 并失去了细节的纹理信息。此外, 由于难以平衡多个损失的贡献, 现有的 SR 网络都不能充分利用多个损失形成一个组合优化的损失函数。为了在获得较高的 PSNR 值的前提下, 生成高感知的图像, 我们提出了一种新的混合损失函数 L_m , 它结合了梯度损失 L_d 、 L_1 损失、感知损失 L_g 和 GAN 损失 L_p , 并利用权重学习和正则化惩罚在模型的稳定性损失和收益之间取得平衡。其表示为:

$$L_m = L_1 + \alpha L_g + \beta L_p + \gamma L_d + \tau |W|_1 \quad (19)$$

式中: α 、 β 和 γ 是平衡不同损失项的系数; τ 是正则化系数; W 是网络的权重参数。详细的损失说明如下:

$$L_1 = \|I_{\text{HR}} - I_{\text{GT}}\| \quad (20)$$

$$L_g = \min(-\log(D(G(I_{\text{HR}})))) \quad (21)$$

$$L_p = \|\phi(I_{\text{HR}}) - \phi(I_{\text{GT}})\|_2^2 \quad (22)$$

$$L_d = \|\nabla(I_{\text{HR}}) - \nabla(I_{\text{GT}})\|_1 \quad (23)$$

式中: I_{GT} 表示地面真实图像; $D(\cdot)$ 和 $G(\cdot)$ 分别表示鉴别器和生成器的输出; $\phi(\cdot)$ 表示特征图的输出; L_1 表示像素损失, 这里计算为恢复图像 I_{HR} 与真实图像 I_{GT} 对应的像素位置的值的误差。 L_p 表示感知损失, 它使用预先训练的 (Visual Geometry Group 19, VGG19) 网络提取的特征来测量图像之间的感知差异。 L_g 表示 GAN 损失, 即表示生成器和鉴别器之间的对抗的损失。 L_d 表示正则化的梯度损失, $\nabla(I_{\text{HR}})$ 和 $\nabla(I_{\text{GT}})$ 分别表示恢复图像 I_{HR} 的梯度和真实图像 I_{GT} 的对应梯度。上述梯度损失可以帮助防止 SR 网络产生模糊的边界。 $\nabla(I_{\text{HR}})$ 和 $\nabla(I_{\text{GT}})$ 的数学公式如下:

$$\nabla(I_{\text{HR}}) = \frac{\partial^2 I_{\text{HR}}(x, y)}{\partial x^2} + \frac{\partial^2 I_{\text{HR}}(x, y)}{\partial y^2} \quad (24)$$

$$\nabla(I_{\text{GT}}) = \frac{\partial^2 I_{\text{GT}}(x, y)}{\partial x^2} + \frac{\partial^2 I_{\text{GT}}(x, y)}{\partial y^2} \quad (25)$$

对于经典的双三次退化图像, 仅使用与之前工作相同 L_1 损失来优化它的 PSNR 值, 以保证公平性和有效性, 对于真实世界红外图像, 使用 L_1 损失、感知损失和 GAN 损失的组合方式来提高视觉质量。

3 实验

在本章中, 为了评估我们提出的算法的性能, 我们分别对传统的双三次退化图像和合成退化图像进行训练。在经典退化图像数据集 Set5^[10]、Set14^[8] 和其他数据集^[7] 以及真实红外图像数据集 (这里采用了艾睿光电红外开源数据库^[21]) 上进行了测试。

3.1 数据集与实验设置

训练细节: 对于经典的退化图像, 退化模型采用传统的双三次下采样。我们使用 DIV2K^[29] 数据集进行训练。LR 图像块大小设置为 48, 并且 GR-GFF 块数量、通道数、窗口大小和 GFF 块数量通常分别被设置为 6、180、8 和 3。对于真实红外图像, 退化模型是构建的综合退化模型 (见 2.1 节)。我们使用 DIV2K^[29]、Flickr2K^[9]、Wed^[21] 和 OST^[21] 数据集进行训练, LR 图像块大小设置为 64, GR-GFF 块个数、通道数、窗口大小和 GFF 块个数分别设置为 6、220、8 和 3。

用 DGXA 100 GPU 训练模型, 使用 Adam 优化器^[17], 批量大小为 32。首先以 2×10^{-4} 的学习率来训练生成器, 以 L_1 损失作为损失函数, 尽可能地提高图像的 PSNR 值, 然后在经典的退化图像上进行测试。以 1×10^{-4} 的学习率迭代训练生成器和鉴别器。使用

L1 损失、感知损失（使用预训练的 VGG19 网络中提取特征图作为感知损失）、GAN 损失和梯度损失的组合方式，权重分别为{1, 1, 0.1, 0.2}，然后在真实红外图像上进行测试。

退化细节：对于经典的退化图像，使用了传统的双三次下采样退化模型。将 HR 图像划分为大小为 128×128 的块。对于这个 HR 块，在整个图像区域上使用单一的双三次退化模型来获得 LR 图像。对于合成退化图像，使用由随机二阶退化模型和传统退化模型组成的综合退化模型（见第 2.1 节）。为了平衡退化模型的复杂性和有效性，将 HR 图像划分为大小为 256×256 的块。对于这个 HR 块随机选择一个大小为 50~200 的区域进行随机二阶降级。然后，对于剩余的区域，使用随机二阶退化。最后，得到了合成的 LR

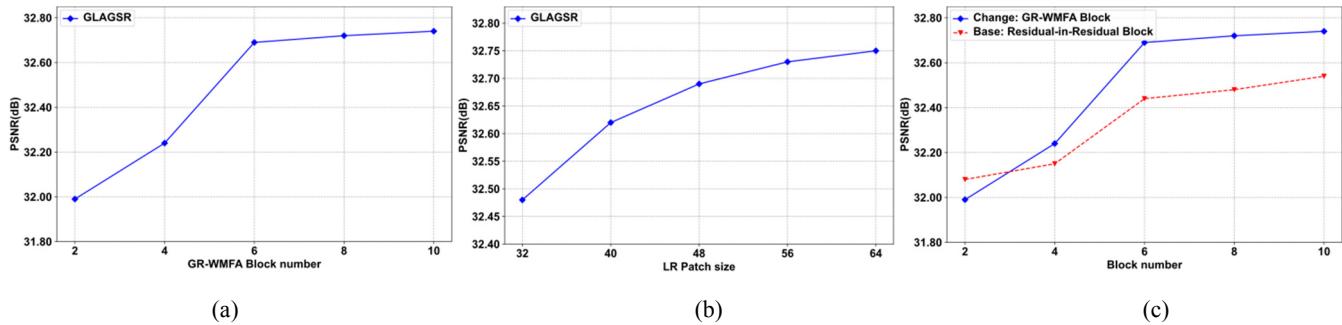


图 5 GLAGSR 不同设置下的消融研究. (a)不同的 GR-GFF 数量块; (b)不同的 LR 图像块; (c)不同的数量块

Fig.5 Ablation study on different settings of GLAGSR . (a)Different GR-GFF block numbers; (b) Different Patch sizes; (c) Different block numbers

GFF 块数目和权重因子的消融研究：表 1 显示了 GFF Block 的 3 种不同的数量对 PSNR 和 (Structural Similarity, SSIM) 的影响。从表中我们可以得出结论，当 GFF 块数为 3 时，PSNR 的性能最佳。而其他两种带来的改善很小，这可能是因为块数过小时提取的特征不充分，块数过大时提取的特征噪声过大。

表 1 GFF 模块数量设计的消融研究表

Table 1 Ablation study on GFF block number design

GFF block number	2	3	4
PSNR	32.60	32.68	32.64
SSIM	0.8999	0.9010	0.9011

特别是，当 GFF 块的数量为 3 时，SSIM 的值是次佳的。这可能特别是，当使用较小的块数时，一些视觉特征（如亮度和对比度等）没有被充分捕获。因此，当 GFF 块的数量为 4 时，就 SSIM 而言，可以获得最佳的性能，从而允许网络捕获更详细的符合人类视觉特征的特征，从而产生更好的性能。

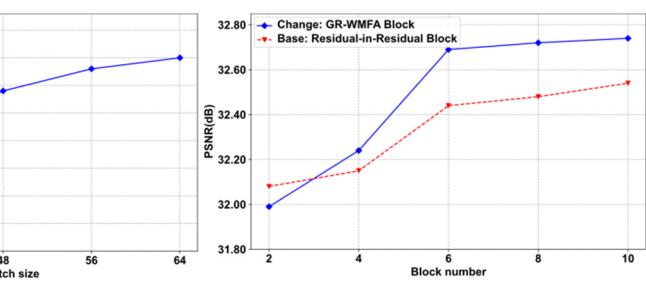
表 2 显示了 GFF 模块的权重学习因子 $w_0:w_1$ 对性能的影响。从表 2 可以看出， w_0 和 w_1 的比值为 2 时

图像。

3.2 消融实验

对于消融实验，在经典的双三次退化数据集 DIV2K^[29]上训练 GLAGSR，并在 Set5^[10]上进行测试，这里放大因子为 4。

GR-GFF 块数目和 LR 图像块大小的消融研究：分别在图 5(a)和(b)中展示 GR-GFF 块数和 LR 图像尺寸对模型性能的影响。结果表明，GLAGSR 模型的性能与 LR 块大小和 GR-GFF 块数呈正相关。然而，随着 GR-GFF 块数目和 LR 图像尺寸的增加，性能增益逐渐饱和。为了平衡模型性能和参数量，将 GR-GFF 块数量固定为 6，将 LR 图像块大小固定为 48，以获得相对较小的模型，这也加速了模型的训练时间。



可以获得最佳的 PSNR 和 SSIM 值。

表 2 权重因子的消融研究表

Table 2 Ablation study on weight factor

Weight Factor	$w_0:w_1=1$	$w_0:w_1=2$	$w_0:w_1=0.5$
PSNR	32.15	32.68	32.04
SSIM	0.8887	0.9012	0.8786

GR-GFF 结构设计的消融研究：在图 5(c)中给出了 GR-GFF 模块结构设计对模型性能的影响。由于 GR-GFF 模块的结构设计借鉴了 (Enhanced Super-Resolution Generative Adversarial Networks, ESRGAN) 的工作成果，为了验证模型的有效性，在两种不同的设计（基于 ESRGAN^[24]的残余致密块的设计和基于 ESRGAN 的改进的 GR-GFF 模块的设计）上进行了实验。从图中可以看出，两个结构块的数量与 PSNR 值呈正相关。

尽管基于 ESRGAN 的残差密集块的设计方法在块数为 2 时可以获得较高的 PSNR 值。但随着块的数量的提升，基于 ESRGAN 改进的 GR-GFF 模块设计可以获得更好的性能。其原因在于提出的分组残差结

构可以灵活地优化从每个 GFF 块中提取的多尺度特征, 进一步提高网络表示能力。

鉴别器消融研究: 首先, 通过实验证明了两种不同形式的鉴别器的有效性, 即原始的基本 TransGan 鉴别器结构和改进的基于 TransGan 的 Swin 变压器鉴别器。从图 6 中可以观察到, 原始的基本 TransGan 鉴别器结构不能恢复详细的纹理 (砖和灌木), 甚至在灌木枝上引入令人不快的伪影。与之相反, 使用改进的基于 TransGan 的 SwinT 块鉴别器可以产生更多的局部细节。同时, 这也增加了训练的不稳定性。因此, 设计了一个正则化惩罚项, 可以在稳定训练动态的同时提高纹理恢复能力。

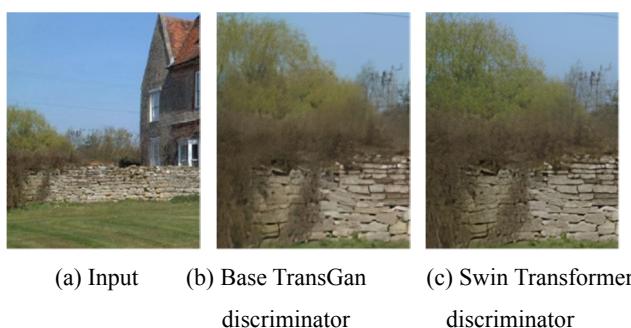


图 6 鉴别器设计的消融研究

Fig.6 Ablation of two discriminator designs

损失函数的消融研究: 为了证明提出的混合损耗的有效性, 在 GLAGSR 结构的基础上进行了一系列使用不同损耗的实验。如表 3 所示, 首先单独使用像素损失 L_1 , 感知损失 L_p , GAN 损失 L_g 和梯度损失 L_d 进行训练, 最后使用混合损失进行训练。结果表明, 综合损失可以达到最佳的效果, 这验证缓和损失函数的有效性。

3.3 对比实验

双三次退化图像的超分辨率重建: 表 4 显示了 GLAGSR 与最先进的方法之间的定量比较, 包括

表 4 基准数据集上双三次退化图像的超分辨率性能 (PSNR/SSIM) 与最新方法的定量比较

Table 4 Quantitative comparison of super-resolution performance (average PSNR/SSIM) with the state-of-the-art methods for bicubic degradation images on benchmark datasets

Method	Scale	Training dataset	Set5 ^[10]		Set14 ^[8]		BSD100 ^[7]		Urban100 ^[7]	
			PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
SRCNN ^[6]	× 2	DIV2K	36.66	0.9542	32.45	0.9067	31.36	0.8879	29.50	0.8946
EDSR ^[9]	× 2	DIV2K	38.11	0.9602	33.92	0.9195	32.32	0.9013	32.93	0.9773
RDN ^[14]	× 2	DIV2K	38.24	0.9614	34.01	0.9212	32.34	0.9017	33.39	0.9353
RCAN ^[22]	× 2	DIV2K	38.27	0.9614	34.12	0.9216	32.41	0.9027	33.34	0.9384
SAN ^[28]	× 2	DIV2K	38.31	0.9620	34.07	0.9213	32.42	0.9028	33.10	0.9370
HAN ^[23]	× 2	DIV2K	38.27	0.9614	34.16	0.9217	32.42	0.9027	33.35	0.9385
NLSA ^[2]	× 2	DIV2K	38.34	0.9618	34.08	0.9231	32.43	0.9027	33.42	0.9394

SRCNN^[6], EDSR^[9], RCAN^[22] (Residual Channel Attention Networks), SAN^[28] (Second-order Attention Network), HAN^[23] (Holistic Attention Network), RDN^[14] 和 NLSA^[2] (Non-Local Sparse Attention)。如表 4 所示, 当在 DIV2K 上训练时, 我们提出的 GLAGSR 在 4 个基准数据集上的所有尺度因子上都取得了较佳的性能。特别是 HAN^[23] 和 RCAN^[22] 引入了空间注意机制和信道注意机制, NLSA^[2] 引入了非局部注意机制。虽然这些方法通过其设计的注意机制增加了接受领域, 但没有一种方法比 GLAGSR 模型表现得更好。这是因为高频信息比低频信息更难恢复, 因此如果模型能够针对具有不同特征的兴趣点进行优化, 那么 SR 结果会更好。然而, 以往的方法往往对提取的特征一视同仁, 忽视了网络资源分布对各种信息的影响, 从而限制了其表示能力。相比之下, 本文设计了一个基于参数自学习的加权多尺度特征聚合模块, 利用多尺度卷积块和 SwinT 块提取不同尺度的特征, 通过注意机制和加权学习因子引导全局-局部信息的融合, 自适应地匹配图像的特征域, 从而提高模型的表示能力。

表 3 损失函数的消融研究表

Table 3 Ablation study of the proposed hybrid loss

Index	Loss1	Loss2	Loss3	Loss4	Loss5
L_1	✓	✗	✗	✗	✓
L_p	✗	✓	✗	✗	✓
L_g	✗	✗	✓	✗	✓
L_d	✓	✗	✗	✓	✓
PSNR	32.62	32.58	32.53	32.51	32.68
SSIM	0.9000	0.8994	0.8985	0.8987	0.9011

Continued Table 4

GLAGSR (Ours)	$\times 2$	DIV2K	38.37	0.9616	34.17	0.9221	32.48	0.9029	33.49	0.9395
SRCNN ^[6]	$\times 3$	DIV2K	36.66	0.9542	32.45	0.9067	31.36	0.8879	29.50	0.8946
EDSR ^[9]	$\times 3$	DIV2K	34.76	0.9290	30.66	0.8481	29.32	0.8104	29.02	0.8685
RDN ^[14]	$\times 3$	DIV2K	34.58	0.9280	30.53	0.8447	29.23	0.8079	28.46	0.8582
GLAGSR (Ours)	$\times 3$	DIV2K	34.90	0.9314	30.80	0.8498	29.40	0.8130	29.55	0.8751
SRCNN ^[6]	$\times 4$	DIV2K	30.84	0.8628	27.50	0.7513	26.90	0.7101	24.52	0.7221
EDSR ^[9]	$\times 4$	DIV2K	32.46	0.8968	28.80	0.7876	27.71	0.7420	26.64	0.8033
RDN ^[14]	$\times 4$	DIV2K	32.47	0.8990	28.81	0.7871	27.72	0.7419	26.61	0.8028
RCAN ^[22]	$\times 4$	DIV2K	32.63	0.9002	28.87	0.7889	27.77	0.7436	26.82	0.8087
SAN ^[28]	$\times 4$	DIV2K	32.64	0.9003	28.92	0.7888	27.78	0.7436	26.79	0.8068
HAN ^[23]	$\times 4$	DIV2K	32.64	0.9002	28.90	0.7890	27.80	0.7442	26.85	0.8094
NLSA ^[2]	$\times 4$	DIV2K	32.59	0.9000	28.87	0.7891	27.78	0.7444	26.96	0.8109
GLAGSR (Ours)	$\times 4$	DIV2K	32.80	0.9029	29.03	0.7928	27.89	0.7461	27.02	0.8135

此外, GLAGSR 模型实现了高性能的同时, 大大地减少了模型参数量。图 7 表明, 与现有的方法相比, GLAGSR 方法获得了更好的 PSNR, 同时参数量减少了 20% 以上。

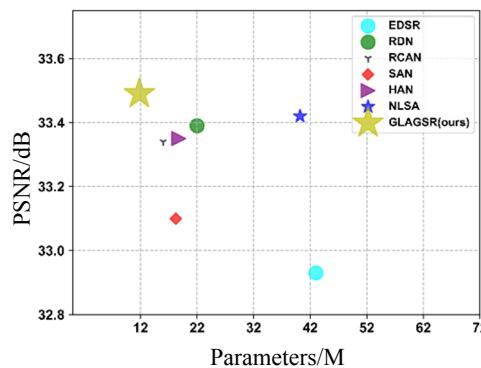


图 7 对于 Urban100 上的图像 SR($\times 2$), PSNR 结果与不同方法的参数总数相比较

Fig.7 PSNR results vs the total number of parameters for different methods for image SR ($\times 2$) on Urban100

真实场景红外图像的超分辨率重建: 为了评估 GLAGSR 在红外图像上的性能, 在收集的压缩图像数据集上进行了测试。由于没有地面真实的高质量图像, 我们只提供了与代表性模型 ESRGAN^[24], BSRGAN^[17], SwinIR^[30] (Image Restoration Using Swin Transformer), RealSR^[12] (Real-World Super-Resolution,) 和 Real-ESRGAN^[18] 的视觉比较。此外, 现有的感知质量指标不能很好地反映人类在细粒度尺度上的实际感知偏好^[18]。因此, 我们对测试图像执行非参考图像质量评估 (Image Quality Assessment, IQA) 指标 (评估指标包括 Natural Image Quality Evaluator, NIQE^[31], Non-Reference Quality Metric,

NRQM^[32])。几个有代表性的视觉样本如图 8 所示。

从图 8 可以看出, GLAGSR 在去除伪影和恢复纹理细节方面优于现有的 (State Of The Arts, SOTA) 方法, IQA 评估结果更好。其他方法可能导致各种各样的视觉问题。例如, ESRGAN 引入了图像平滑和图像细节变得过于模糊; RealSR 不能添加非自然的纹理; Real-ESRGAN 可能产生扭曲的线条等, 即使是最具代表性的 SwinIR 也会丢失一些细节, 纹理平滑。

4 结语

红外图像恢复是一个非常活跃的领域, 有着广泛的应用前景。然而, 现有的模型仍有一定的局限性, 如不能适应图像退化空间的变化, 恢复后的图像存在伪影和线条失真等。在这项工作中, 我们提出了一个强大的图像恢复模型 GLAGSR, 它包括一个多尺度生成器和一个基于 Transformer 结构的鉴别器。特别是, 开发了一个新的自适应学习特征融合框架, 融合全局和局部特征, 以获得更全面的融合特征表示, 可以自适应匹配图像特征域。为了更好地模拟真实图像的退化, 设计了一种新的退化算法来扩展图像的退化边界。最后, 为了解决图像边界的平滑恢复问题, 还提出了一种混合损失算法, 它结合了正则梯度损失、L1 损失、感知损失等来优化 GLAGSR, 并最终生成更高质量的图像。大量实验表明, 该算法在双三次退化图像和真实红外图像上都获得了较好的效果, 证明了 GLAGSR 模型的有效性。在未来, 我们将进一步将我们的模型扩展到其他图像恢复任务, 如卫星图像增强、水下图像修复等。

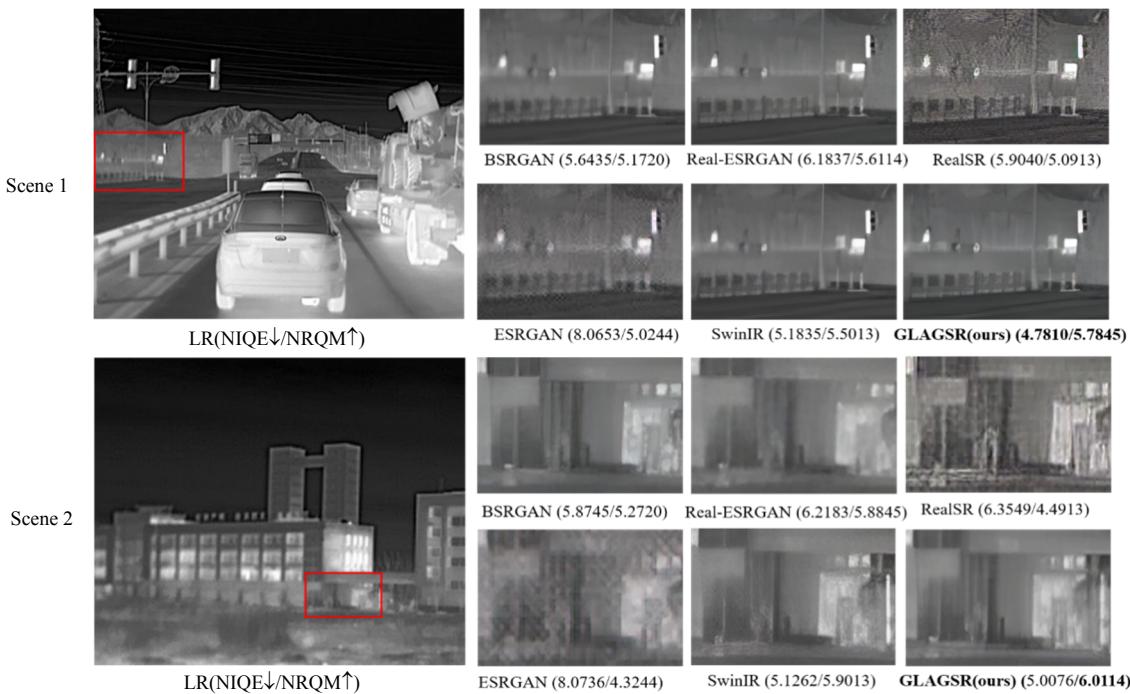


图8 超分辨率($\times 4$)方法在红外图像上的视觉比较

Fig.8 Visual comparison of super-resolution ($\times 4$) methods on real-world infrared images

参考文献：

- [1] 胡德敏, 闵天悦. 改进型轻量级GAN的红外图像超分辨率方法[J]. 小型微型计算机系统, 2022, 43(8): 1711-1717.
HU Demin, MIN Tianyue. Infrared image super-resolution via improved lightweight GAN[J]. *Journal of Chinese Computer Systems*, 2022, 43(8): 1711-1717.
- [2] MEI Y, FAN Y, ZHOU Y. Image super-resolution with non-local sparse attention[C]//Proc. CVPR, 2021: 3517-3526. DOI: 10.1109/CVPR46437.2021.00352.
- [3] 马敏慧, 王红茹, 王佳. 基于改进的MSRCR-CLAHE融合的水下图像增强算法[J]. 红外技术, 2023, 45(1): 23-32.
MA Minhui, WANG Hongru, WANG Jia. An underwater image enhancement algorithm based on improved MSRCR-CLAHE fusion[J]. *Infrared Technology*, 2023, 45(1): 23-32.
- [4] ZHANG D, SHAO J, LI X, et al. Remote sensing image super-resolution via mixed high-order attention network[J]. *IEEE Trans. Geosci. Remote Sens.*, 2021, 59(6): 5183-5196.
- [5] 刘云峰, 赵洪山, 杨晋彪, 等. 基于GNR先验的电力设备热成像超分辨率方法[J]. 红外技术, 2023, 45(1): 40-48.
LIU Yunfeng, ZHAO Hongshan, YANG Jinbiao, et al. Super resolution method for power equipment infrared imaging based on gradient norm-ratio prior[J]. *Infrared Technology*, 2023, 45(1): 40-48.
- [6] DONG C, Loy C, HE K, et al. Learning a deep convolutional network for image super-resolution[C]//Proc. Computer Vision, ECCV, 2014: 184-199. DOI: 10.1007/978-3-319-10593-2_13.
- [7] 聂丰英, 侯利霞, 万里勇. 自适应双边滤波与方向梯度的红外图像增强[J]. 红外技术, 2022, 44(12): 1309-1315.
- NIE Fengying, HOU Lixia, WAN Liyong. Infrared image enhancement based on adaptive bilateral filtering and directional gradient[J]. *Infrared Technology*, 2022, 44(12): 1309-1315.
- [8] Timofte R, Rothe R, Van Gool L. Seven ways to improve example-based single image super resolution[C]//Proc. CVPR, 2016: 1865-1873. DOI: 10.1109/CVPR.2016.206
- [9] Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution[C]//Proc. CVPRW, 2017: 136-144. DOI: 10.1109/CVPRW.2017.151
- [10] Bevilacqua M, Roumy A, Guillemot C, et al. Low-complexity single-image super-resolution based on nonnegative neighbor embedding[C]//Proc. BMVC, 2012: 135-141. DOI: 10.5244/C.26.135.
- [11] TANG Y, GONG W, CHEN X, et al. Deep inception-residual Laplacian pyramid networks for accurate single-image super-resolution[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2019, 31(5): 1514-1528.
- [12] CAI J, ZENG H, YONG H, et al. Toward real-world single image super-resolution: a new benchmark and a new model[C]//Proc. ICCV, 2019: 3086-3095. DOI: 10.1109/ICCV.2019.00318.
- [13] SHI W, Caballero J, Huszár F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]//Proc. CVPR, 2016: 1874-1883. DOI: 10.1109/CVPR.2016.207.
- [14] ZHANG Y, TIAN Y, KONG Y, et al. Residual dense network for image super-resolution[C]//Proc. CVPR, 2018: 2472-2481. DOI: 10.1109/CVPR.2018.00262.
- [15] Esser P, Rombach R, Ommer B. Taming transformers for high-resolution

- image synthesis[C]//Proc. CVPR, 2021: 12873-12883. DOI: 10.1109/CVPR46437.2021.01268.
- [16] JIANG Y, CHANG S, WANG Z. Transgan: two transformers can make one strong GAN[J/OL]. *Computer Vision and Pattern Recognition*, 2021, 34: <https://arxiv.org/abs/2102.07074>.
- [17] ZHANG K, LIANG J, Van Gool L, et al. Designing a practical degradation model for deep blind image super-resolution[C]//Proc. ICCV, 2021: 4771-4780. DOI: 10.1109/ICCV48922.2021.00475.
- [18] WANG X, XIE L, DONG C, et al. Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data[C]//Proc. ICCVW, 2021: 1905-1914. DOI: 10.1109/ICCVW54120.2021.00217
- [19] Ledig C, Theis L, Huszar F, et al. Photo-realistic single image super-resolution using a generative adversarial network[C]//Proc. CVPR, 2017: 4681-4690. DOI: 10.1109/CVPR.2017.19.
- [20] ZHOU Y, WU G, FU Y, et al. Cross-MPI: cross-scale stereo for image super-resolution using multiplane images[C]//Proc. CVPR, 2021: 14842-14851. DOI: 10.1109/CVPR46437.2021.01460.
- [21] 烟台艾睿光电科技有限公司. 艾睿光电红外开源数据库[EB/OL]. [2023-02-26]. http://openai.raytrontek.com/apply/Super_resolution.html/.
- IRay Technology Co., Ltd. IRay Optoelectronic Infrared Open Source Database[EB/OL]. [2023-02-26]. http://openai.raytrontek.com/apply/Super_resolution.html/.
- [22] ZHANG Y, LI K, LI K, et al. Image super-resolution using very deep residual channel attention networks[C]//Proc. ECCV, 2018: 286-301. DOI: 10.1007/978-3-030-01234-2_18.
- [23] NIU B, WEN W, REN W, et al. Single image super-resolution via a holistic attention network[C]//Proc. ECCV, 2020, 12357: 191-207.
- [24] WANG X, YU K, WU S, et al. ESRGAN: Enhanced super-resolution generative adversarial networks[C]// Proc. ECCV, 2019: 63-79. DOI: 10.1007/978-3-030-11021-5_5.
- [25] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proc. ICCV, 2021: 9992-10002. DOI: 10.1109/ICCV48922.2021.00986.
- [26] WANG Y, WANG L, WANG H, et al. Resolution-aware network for image super-resolution[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2018, 29(5): 1259-1269.
- [27] Andreas Lugmayr, Martin Danelljan, Radu Timofte. Unsupervised learning for real-world super-resolution[J]. *ICCV Workshop*, 2019, 2(3): 3408-3416.
- [28] DAI T, CAI J, ZHANG Y, et al. Second-order attention network for single image super-resolution[C]//Proc. CVPR, 2019: 11057-11066. DOI: 10.1109/CVPR.2019.01132.
- [29] Agustsson E, Timofte R. Ntire 2017 challenge on single image super-resolution: Dataset and study[C]// Proc. CVPRW, 2017: 126-135. DOI: 10.1109/CVPRW.2017.150.
- [30] LIANG J, CAO J, SUN G, et al. SwinIR: Image restoration using swin transformer[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 1833-1844. DOI: 10.1109/ICCVW54120.2021.00210.
- [31] Blau Y, Mechrez R, Timofte R, et al. The 2018 PIRM challenge on perceptual image super-resolution[C]//Proceedings of the European Conference on Computer Vision (ECCV), 2018, DOI: 10.1007/978-3-030-11021-5_21
- [32] MA C, YANG C Y, YANG X, et al. Learning a no-reference quality metric for single-image super-resolution[J]. *Computer Vision and Image Understanding*, 2017, 158: 1-16.

(上接第 781 页)

- [15] 王新彦, 易政洋. 基于改进 YOLOv5 的割草机器人工作环境障碍物检测方法研究[J]. 中国农机化学报, 2023, 44(3): 171-176. DOI: 10.13733/j.jcam.issn.2095-5553.2023.03.024.
- WANG Xinyan, YI Zhengyang. Research on obstacle detection method of mowing robot working environment based on improved YOLOv5[J]. *Journal of Chinese Agricultural Mechanization*, 2023, 44(3): 171-176. DOI: 10.13733/j.jcam.issn.2095-5553.2023.03.024.
- [16] 徐正军, 张强, 许亮. 一种基于改进 YOLOv5s-Ghost 网络的交通标志识别方法[J]. 光电子·激光, 2023, 34(1): 52-61. DOI: 10.16136/j.joel.2023.01.0216.
- XU Zhengjun, ZHANG Qiang, XU Liang. Research on obstacle detection method for working environment of lawn mower robot based on improved YOLOv5[J]. *Journal of Optoelectronics Laser*, 2023, 34(1): 52-61. DOI: 10.16136/j.joel.2023.01.0216.
- [17] 党宏社, 党晨, 张选德. 基于改进 YOLOv5s 的交通标志识别算法[J]. 实验技术与管理, 2022, 39(9): 97-102. DOI: 10.16791/j.cnki.sjg.2022.09.016.
- DANG Hongshe, DANG Chen, ZHANG Xuande. Traffic sign recognition algorithm based on improved YOLOv5s[J]. *Experimental Technology and Management*, 2022, 39(9): 97-102. DOI: 10.16791/j.cnki.sjg.2022.09.016.