引文格式
刘茜娜，顾津锦，董超. 图像背景在图像超分辨率中的作用研究[J].集成技术,2023,?(?):??

Citing format
Liu XN, Gu JJ, Dong C. Investigating the Function of Image Background in Image Super Resolution[J]. Journal of Integration Technology,2023,?(?):??

图像背景在图像超分辨率中的作用研究

刘茜娜[1,2]，顾津锦[3]，董超[1*]

[1]（中国科学院深圳先进技术研究院 深圳 518055）

[2]（中国科学院大学，北京 101408）

[3]（悉尼大学）

摘要：图像超分辨率是底层视觉领域的一项代表性任务，相关研究发现图像某个像素位置的重建质量与其周围的背景存在关联性，基于该现象，本文探索了通过分割输入图像来解释网络的新视角，提出了一种简单组合数据集，该数据集具有丰富的信息量，但单张图中只包含单一的纹理信息，我们证明了与目标区域纹理相近的背景更有利于模型在该区域的超分辨率重建，通过对比分析注意力机制与传统卷积神经网络，结论显示注意力结构更能帮助网络关注长程有效信息。

关键词：超分辨率；网络可解释性；图像背景；数据集
doi: 10.12146/j.issn.2095-3135.20230215001

Investigating the Function of Image Background in Image Super Resolution

LIU Xina[1,2], Gu Jinjin[3], DONG Chao[1*]

[1] (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China)
[2](University of Chinese Academy of Sciences, Beijing 101408, China)
[3](The University of Sydney)

Abstract: As a representative low-level vision problem, image super resolution (SR) aims to reconstruct the high-resolution image from its low-resolution counterpart. For a long time, the analysis of SR tasks is based on the whole image, while little works observe the input partition. In this paper, we find that the restoration quality of a certain position is inseparable from its surrounding image background. This phenomenon provides us a new perspective to explain the networks by splitting the input image. We construct a new hybrid dataset, of which the foreground and background contain only one kind of texture information. And then, we prove that the similar background could benefit the network restoration. By analyzing similarity and difference between the attention mechanism and the traditional CNN network, we show that the attention structure could help the network focus on long-range effective information. Moreover, a data enhancement method to improve the network final performance and potential future works are also proposed.

来稿日期：2023-02-15 修回日期：yyyy-mm-dd
作者简介：刘茜娜，硕士研究生，研究方向为图像处理；顾津锦，博士研究生，研究方向为图像处理；董超(通讯作者)，研究员，研究方向为计算机视觉，E-mail：chao.dong@siat.ac.cn；

Key words: Super-Resolution; Network Interpretability; Image Background; Dataset

1 引 言

图像超分辨率（Image Super-Resolution, SR，以下简称“超分”）是一项经典的底层视觉任务，旨在从低分辨率输入中恢复高分辨率图像。继单帧超分辨率卷积神经网络（Super-Resolution Convolutional Neural Networks, SRCNN）[1]成功将卷积神经网络引入超分任务后，许多工作设计了新的网络结构[2, 3, 4, 5, 6]，大大提升了网络的拟合能力。近年来，伴随着注意力机制[7, 8, 9]和变换神经网络（Transformer Neural Networks, Transformer）结构[10, 11, 12]的加入和改进，超分模型取得了新一轮的性能飞跃。尽管这些工作取得了新的进展，它们成功的原因仍然神秘，因为研究人员只能看到测试结果，却无法解释模型的行为。这种仅仅依靠性能驱动的网络分析方式限制了更好的结构的诞生。

在图 1 中，针对红色方框内的区域，仅仅改变方框周围的补充像素的多少，增强型深度残差网络（Enhanced Deep Residual Networks, EDSR）[5]就对该区域给出了大相径庭的恢复结果。这一现象让作者对方框周围的像素所起到的作用产生了疑问：邻域的加入是不是必要的？是不是有些相邻的像素反而对图像重建有害？什么样的邻域是对恢复结果有益的？以上问题启发本项研究从一个全新的视角来分析超分网络的重建过程：通过分割输入图像来观察和解释网络。在本文中，选择输入图像的中心区域作为分析对象“前景”，将周围的其他像素视为分析“背景”。实验验证了前面的猜想，有些邻域的出现是不必要的，甚至是对结果有害的。此外，为了支持这种分析方法，作者在自然图像之外制作了一个户外场景简单组合数据集(Outdoor Scene Simple Combined Dataset, OSSCD)。这些数据的特殊之处在于，作为前景的中心区域与背景来自不同的图像。本项工作通过改变前景的邻域以探索超分任务中背景的影响。借助新的分析视角和新的数据集，本项研究观察到，当网络重建一个区域时，附近的像素与该区域越相似，该区域的性能越好。经过验证，该结论对领域内几个代表性网络结构都成立。代表性网络指代的是：EDSR[5]，深度残差通道卷积网络(Very Deep Residual Channel Attention Networks, RCAN) [9]，和变换神经网络(Image Restoration Using Swin Transformer, SwinIR) [10]。值得一提的是，文章中将这一发现应用于网络行为的解释，成功发现了注意力机制和 Transformer 结构与传统卷积神经网络（Convolutional Neural Networks, CNN）网络之间的异同，注意力结构的确能帮助网络关注长程有效信息。

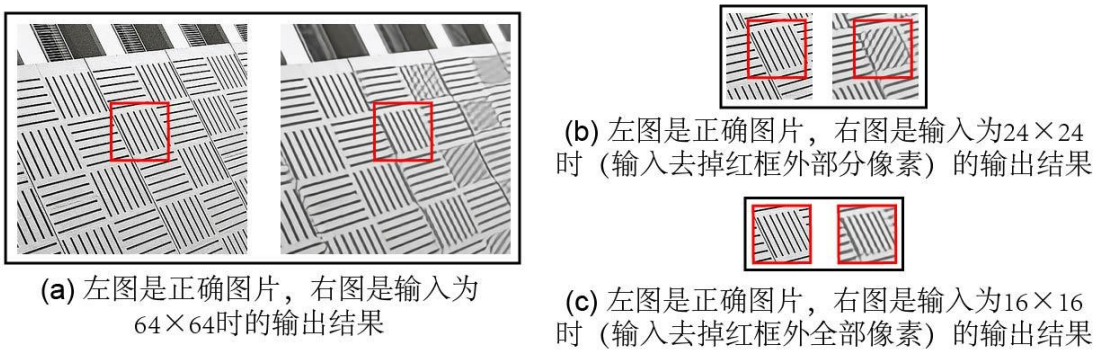

图 **1** EDSR 的超分结果

Fig. 1 SR results of EDSR

本文的主要贡献有 3 点：（1）提出了新的分析视角，即将输入图像拆分分析；（2）制作了更适合解释网络的 OSSCD 数据集；（3）拓展解释网络结构和网络行为，提出并简单验证了本项工作未来可能的发展方向。

2 国内外的研究现状

2.1 超分辨率

作为底层视觉中的一项代表性任务，超分（SR）的目的是通过学习高分辨率（High Resolution, HR）和低分辨率（Low Resolution, LR）图像对的映射，从 LR 输入中重建 HR 图像。SRCNN[1]是首个基于 CNN 提出的超分网络。在此之后，研究人员开发了大量超分深度学习模型[2, 3, 4, 5, 6]，包括残差体系结构[13]、循环体系结构[14]等。在注意力机制得到推广后，后来的许多网络增加了对注意力模块的设计和使用[7, 8, 9]。研究人员认为，注意力结构可以帮助网络获得关于恢复目标的更详细的信息，并抑制网络利用其他的无用信息。事实上，这些新的基于注意力的网络确实比传统的 CNN 网络性能更好。近日，Transformer 在超分领域也取得了突破性进展[10, 11, 12]，得到更清晰的恢复结果已成为现实。在这些网络中，SwinIR[10]利用移位窗口对长程依赖进行建模，因其优异的性能而受到了广泛关注。然而，迄今为止，在超分领域仍然缺乏对这些结果的解释。

2.2 可解释性

自深度学习被应用于计算机视觉以来，许多工作关注了神经网络的可解释性。在高级视觉任务中，对深度神经网络的解释方法已经得到了较为广泛的关注和应用[15, 16]。比如在卷积神经网络的基础上，引入反卷积神经网络[17]，即利用卷积层中卷积核的转置来进行反卷积，进而可视化出每个网络层学习到的特征，该种方法可以对隐层特征进行有效的定性分析。也有人在深度网络中引入注意力机制[18, 19, 20]，即在不影响模型效果的前提下，引入注意力向量，对特征及网络中的隐层特征赋予不同的权重，并在训练过程中对该权重进行学习，这样就可以得到各个特征对于模型学习的重要性程度，从而达到解释模型的效果。解释网络的一个好方法是归因（显著性图），它明确地可视化了模型预测的结果是由哪部分输入负责的。基于显著性图，许多方法[21, 22, 23, 24, 25]获得了人类可理解的归因表示。超分网络同样继承了深度学习和深度神经网络难以解释的性质。最近，有工作提出了一种专门针对超分网络的解释方法，称为局部归因图（Local Attribution Map, LAM）[26]，以定位影响网络输出的输入特征。此外，在底层视觉领域，深度退化表示[27]发现，在模型测试期间，特征图会根据退化类型聚集在一起。现有的解释工作有助于研究人员从不同的角度解释网络的工作机制，为更好的设计铺平了道路。遗憾的是，类似的研究还很少，加深对超分网络可解释性的探究是当前超分领域亟待解决的问题。

2.3 数据集

作为超分领域常用的训练集，DIV2K[28]的数据质量已经得到了广泛认可，也便利了研究人员训练更好出的模型。此外，Set5[29]、Set14[30]、BSD100[31]、Manga109[32]和 Urban100[33]等数据集也被广泛用于训练和测试。除了这些语义复杂的数据集，空间特征变换生成式对抗网络[34]提出了户外场景数据集（Outdoor Scene Dataset, OST），这是一个具有丰富纹理的数据集，但单张图中只包含单一的纹理，如动物毛发、建筑物的砖块、或是水的波纹等等。实验表明，OST 数据集也能够使网络学习到有效的信息。

3 图像背景的作用

3.1 超分网络中的特殊现象

超分领域内有许多用于评估图像质量的指标，其中一个较为重要的是峰值信噪比（Peak Signal to Noise Ratio, PSNR）。如图 2 所示，现阶段有两种常见的测试图像质量的方法。第一种是将整张测试图像输入训练好的超分网络，获得恢复结果并计算与原始高分辨率图像的差异，PSNR 可以量化这种差异。另一种方法是将整图切割成小块，分别输入

网络，先将获得的结果进行组合，再使用组合结果作为最终恢复结果进行评估计算。当使用的硬件设备计算能力不足时，研究人员们通常会采用第二种评估方法。但在这两种测量方法下，PSNR 值的计算结果有可能会存在不小的差距（单图差距超过 1dB），切片的方法总是可以获得更高的量化数值，而关于这一差距的解释至今并不明确。在上述两种方法中，整图输入和切片输入作对比，受影响最大的部分是小切片图的边缘，这些像素的边缘（图 2（b）中的红线附近）变化很大，周围像素在剪切测试时会完全消失。因此，很容易联想到，是否是这种“背景”的缺失导致了两种测试方法之间的差异？但迄今为止，没有类似的办法帮助理解当网络重建某部分图像区域时，相邻像素甚至远处像素带来的影响。这启发作者设计一种分析方法，将图像的目标区域与其他区域分开研究。这显然是一种新的分析视角。从这个角度来看，本项研究希望找出未被选择的像素在超分过程中扮演的角色。

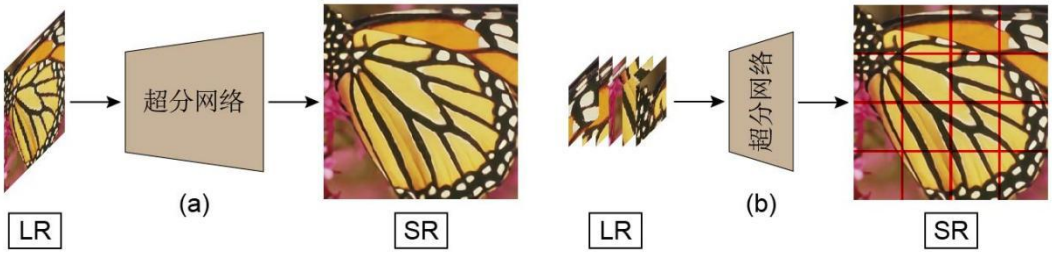

图 **2** 两种常见的测试方法

Fig. 2 Two common test methods

3.2 分析方法

目前，对超分的研究主要集中在整张退化图像的恢复上，很少有方法关注所选择的区域如何受到其附近像素的影响。如果这种效应能够得到解释，研究人员就可以了解更多的网络行为，这对于完整图像的恢复显然是有意义的。启发本文找到新的分析视角的是解释工具 LAM，这是一项从超分网络输入中定位重要像素的工作。该文对图像中心的 16 × 16 局部块进行了归因分析，输入图的其余部分被视为网络恢复中心域的补充输入。在选定图像的部分区域之后，LAM 可以确定网络为重建该部分对附近像素的利用率。受此启发，本文建议分离图像，操作步骤如下：首先，选择图像的中心区域作为前景；然后在前景固定的前提下不断地变换背景，也就是说，为这个小区域改变它的邻域，以便观察网络对前景恢复的变化。改变结果如图 3（a）所示。对于相同的训练模型，在网络重建过程中，仅改变区域附近的像素是否会对该部分的恢复结果产生很大影响？如果有，这一现象主要出现在哪些情况下？找到、衡量和利用这种影响，更多地了解网络行为特性，是文章设计这一分析方法的主要目的。

3.3 简单组合数据集

关于邻域像素的作用，根据 3.2 节中提到的方法，容易联想到：在改变背景的过程中，背景和前景之间的关系发生了变化，中心域重建结果的不同之处就是这样产生的。这是衡量图像背景所产生作用的关键。对于这种关系，显然，背景和前景之间存在相似性和差异性，二者的相似性和差异性与中心区域恢复程度之间的关系是现阶段需要关注的问题。但新的问题是，类似图 3（a）中的自然图像信息含量往往很高，这给分析图片之间的关系带来了很大的困难。因此，为了更容易地将所选区域与其背景分离并降低分析的干扰和复杂程度，需要重新建立一个可控的数据集。便于衡量图像背景对图像中心影响的新数据集必须满足两个特征：首先是单一前景到替换背景的多样性。在确定中心区域的内容之后，

需要为该前景替换多个背景。对于某个前景，良好的性能（高 PSNR）意味着当前背景对模型恢复这部分纹理是相对有利的，低 PSNR 则含义相反。其次是前景和背景纹理的单调性。使用自然图像进行分析的困难在于，如果背景纹理复杂多样，即使发现中心区域由于替换了某个背景而显示出良好的结果，也很难确定背景多种纹理的哪个部分起到了关键作用。此时建立图像之间相关性的方法会受到干扰，结果可能无法控制。综上，现阶段常见的超分数据集并不满足此时的分析需求。

空间特征变换生成式对抗网络[34]提出了使用先验类别信息来解决超分纹理不真实的问题，并设计制作了户外场景数据集 OST。该数据集中包含各类纹理，且每张图像中只具有单调的信息。基于 OST，本项工作制作了 OSSCD 数据集，图 3（b）给出了部分示例。简单是指组合图片的前景和背景仅包含一种纹理信息（背景中的纹理不一定与前景中的纹理相同），如草地、天空、建筑物等。在这类组合数据中，背景和前景纹理丰富但并不复杂，一方面保证了信息含量，能够使网络学习有效的信息；另一方面，新数据比自然图像更容易分析背景和前景之间的相关性。以这批数据作为分析对象，在中心区域的恢复过程中，模型除了利用该区域自己的纹理之外，只能使用背景中的另一种纹理。

图 **3** 组合数据集

Fig. 3 Combined dataset

3.4 方法概括

综上所述，基于超分任务中的一些特殊现象，本文提出了一种新的视角来分析超分辨率网络的重建过程，即通过分割输入图像来观察和解释网络。目标是观察图像背景在图像超分辨率中的作用，确定是否有些背景对像素的重建有害。为了方便分析，本章提出了一个简单组合数据集，该数据集将和自然组合数据集一起支撑下文中的实验和分析。

4 实验结果

4.1 数据集收集

为了保证自然组合图片和 OSSCD 之间的结论一致，下述实验将遵守先在自然组合图片中找到规律，再扩展到 OSSCD 数据集的原则。首先从 DIV2K 验证集[24]和 Urban100[29]中采样出 300 张大小为 128×128 的子图像，这批数据中的 150 张用作前景源，另外 150 张用作背景源。从前景数据集中，选择中心 32×32 的区域逐一与背

景源中的图像组合，这一操作为本阶段带来了 22500 张组合图像。其次，遵循解释复杂案例的原则，手动删除具有重复和不可识别内容的图像，使其中心内容有意义。最终作者筛选出了 100 组数据，每组 100 张，共 10000 张。单个组中的前景内容是相同的，只有中心之外的背景被不断替换（其中一个是该中心区域的原始背景）。此外，本阶段以相同的方式准备了 10000 个 OSSCD 数据，这批数据的前景和背景均来自于 OST 数据集。

4.2 背景的不同作用

前文中提到，将整图先切割成小块再输入网络测试，获得的重建结果可能更好。本章节将通过实验验证这一结果。LAM 中收集了 150 幅对超分网络具有挑战性的图像作为分析的测试集。这些图像同样来自 DIV2K 验证集和 Urban100，大小为 256×256，作者选择在不同超分网络之间具有低平均 PSNR 性能和高方差的子图像。本章节的实验使用了 LAM 提出的 150 张测试集，尝试恢复完整图像，或舍弃部分背景，将中间的 64×64 区域直接输入训练好的网络。此时选择 56×56 大小的中心区域作为目标对象，测量其 PSNR 并进行比较。缩小测试区域是为了消除边缘损坏带来的干扰。在表 1 中，很容易发现许多区域从其背景中单独拿出后恢复出了更好的性能。在舍弃背景后，近一半的中心区域（62 / 150）被 EDSR 恢复的更好。对 RCAN 和 SwinIR，超过 20%的数据中也有同样的现象。这也验证了之前的想法，不是所有的背景都能给前景的恢复带来收益。对于带有注意力机制的 RCAN[9] 和经典的 Transformer 结构 SwinIR 来说，网络似乎只关注了应该关注的部分，补充像素的加入对中心区域的恢复损害不大。但对于传统的 CNN 网络 EDSR 来说，盲目扩大背景范围，很有可能损伤网络的表征能力。这不禁使作者好奇，什么样的背景对前景恢复有害，又是什么样的补充输入对中心区域的恢复有益呢？这显然是本项工作下一步需要解决和回答的问题。

表 **1**　模型对中心区域的恢复
（输入为全图 / 部分时）

Table 1　Models' recovery of central area
(When the input is full / partial)

模型类别	恢复结果	
	全图更好	部分更好
EDSR	88	62
RCAN	117	33
SwinIR	111	39

4.3 相似性和恢复效果

在发现了背景发挥着不同的作用后，本章节将通过实验探索它的各类作用场景。该阶段从 DIV2K 数据集中采样高分辨率图像进行训练，最后选择经过训练的 EDSR、RCAN 和 SwinIR 模型进行测试。实验证实了上文中的猜想，中心块在改变周边背景后恢复结果有了很大变化。除此之外，本阶段的实验还带来了一些新的发现。例如，尽管数据集中为单个中心区域提供了 99 种不同的背景，它仍然更倾向于原始的那张，这块区域在其自身背景下恢复得更准确，PSNR 更高，正如图 4 中展示的那样。然而，值得注意的是，一些其他背景看上去也有助于前景的恢复。经过主观分析，很明显的是，“友好”的背景看起来与前景本身

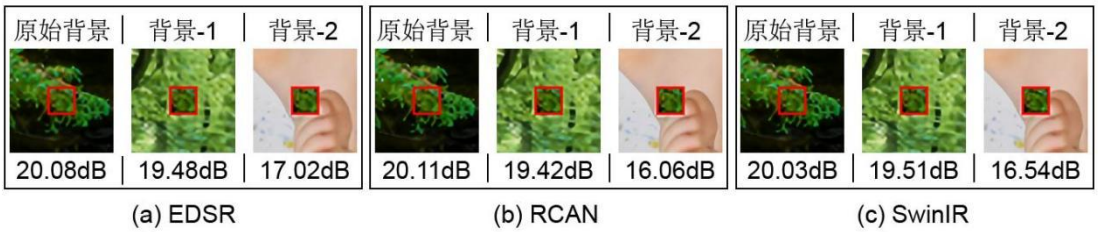

图 **4** 更换背景后的恢复结果（图中为模型输出）

Fig. 4 Recovery after changing the background

非常相似。在 100 组数据中，每个组中重建良好的那些中心块显然具有与自身高度相似的新邻居。这一结论适用于本章使用的所有三个测试模型。如果组合图像的前景和背景看起来相似，则 EDSR、RCAN 和 SwinIR 都可以很好地恢复中间部分。图 5 也证实了这一结论。这说明网络在恢复某一部分区域时，周边补充像素的输入并不是越多越好，相似的信息才是有用的，它们能帮助网络学习和提升对类似纹理的恢复。当邻域都是无用和有害信息时，网络对这些像素不可避免的利用会使中心区域面临恢复较差的困境。

5 探索发现

上文的发现是很有趣的，但更重要的是这些发现的进一步用途。图 6 是几个经典的超分模型对同一张输入的重建结果，模型分别是 EDSR、RCAN 和 SwinIR。显然，EDSR 和 RCAN 错误地还原了中央红框中的纹理。为了解释网络错误的原因，本章节分别分析了这几个模型的超分结果。仅就 EDSR 而言，全图的恢复效果都不够准确，而除了周围的像素之外，似乎没有任何位置能给中央区域条纹方向的判断带来错误的指导。但对于同一张输入，前景与背景都相同的情况下，SwinIR 对全图都做出了较为准确的判断。那么，是什么干扰因素使 RCAN 和 EDSR 频繁发挥失常？Transformer 网络和 CNN 网络的表现在此类情况下有什么区别和联系？基于提出的新视角，将一幅图像拆开并把前景和背景单独看待，本章节想解释不同类型网络的工作机制。

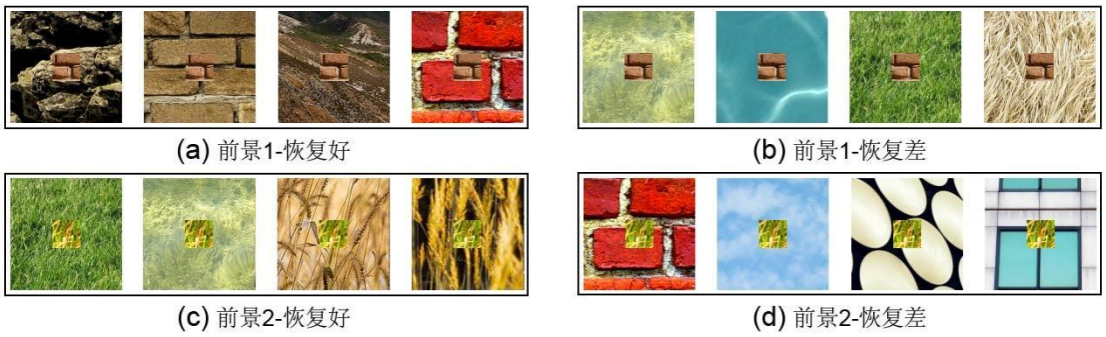

图 **5** 前景重建好 / 不好的情况

Fig. 5 Foreground with high / low performance

显然，这些网络本身有很大的结构差异。EDSR 是一个基本的卷积神经网络，RCAN 是带有注意力机制的 CNN，SwinIR 是一个经典的 Transformer 结构。SwinIR 的作者提到，网络中的非局部稀疏注意力结构（Non-Local-Sparse-Attention, NLSA）是基于非局部注意机制设计的。研究人员认为，具有注意力结构的网络能够忽略无关信息，更加关注提升性

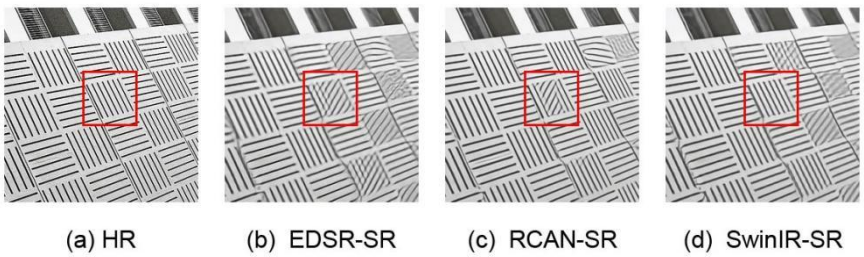

图 **6** 不同模型的超分结果

Fig. 6 Results of different models

能需要关注的关键信息[35, 36]。这些未经验证的网络特性能反映在上文提出的新方法中吗？为便于分析，本阶段使用 OST 数据集制作了一批新数据，图 7 给出了六种方法的示例来组合相同的前景和背景。完整图像大小为 128×128，其中心区域为 16×16，周围背景的大小分别为 0（即没有外来背景的原始图像）、32、48、64、80 和 96。依照第 4.3 节中的规律，EDSR、RCAN 和 SwinIR 都在前景和背景相似的情况下具备更好的恢复效果。因此，本阶段数据集的前景和背景并不相似，随着背景越来越大，对中心恢复有用的信息离中心越来越远。这里使用此种方法来观察网络捕获远程信息的能力。依托于这批新数据，本阶段获得的研究结果总结如下：

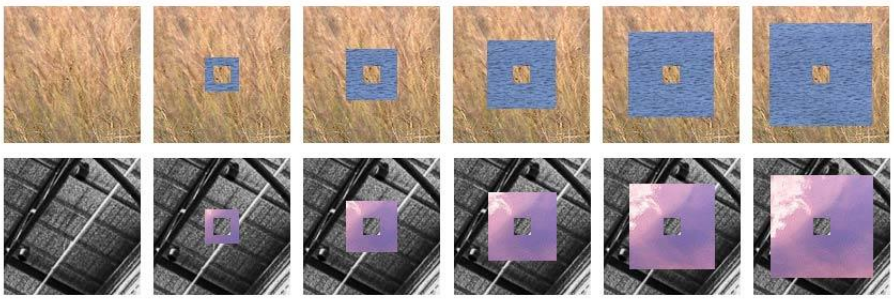

图 **7** 背景大小不同的新数据集

Fig. 7 Dataset with different size of background

（1）PSNR 跃升。对于中心区域的恢复，EDSR 显然不满意目标附近无关的背景。五种尺寸的外来背景都导致了它对中心区域的恢复质量下降（21.34dB 到 19.76dB 及以下）。反观 RCAN，在背景大小为 32 时就展现出了差异性。虽然此时中心域 PSNR 仍然没有使用原始背景时（26.82dB）那么高，但很明显，当背景从 48 缩小到 32（相似的像素更接近中心）时，PSNR 会发生跃升（18.88dB 到 21.05dB）。这里，注意力结构似乎帮助 RCAN 跨过了无用信息、关注到了相似信息因而获得了更好的重建结果。对于 SwinIR，这个数值上的“跃升”在背景大小为 32 和 48 的情况下都有出现（背景 48 时 20.61dB，背景 32 时 21.45 dB，背景 0 时 26.33 dB）。具体数值可见于图 8、图 9 和图 10。但是，数值结果带来了新的疑惑，为什么在无关背景加入后，EDSR 反而是效果最好的模型呢？

（2）对有用信息的关注。在第 4 节中，文章介绍了超分领域最近提出的解释工具 LAM。因此，当发现中心区域的恢复随着有用信息的接近而改变，同时又有（1）中提出的问题时，本文尝试使用 LAM 查看组合图片的归因图，这可能有助于解释这些现象。通过分析图 8 中的归因结果可以发现，对于 EDSR，当背景大小为 32 时，网络试图争取一些有用的信息，但不可避免地利用到了很多无关的信息。图中标红的像素点表示网络利用了该位置的信息来重建红色方框内的区域，颜色越深代表利用率越高。归因图显示 EDSR 同时用到了背景中的像素和背景外的像素。这也就是说，尽管在恢复图像中心的过程中，EDSR

使用了一些相似的有效信息，但它缺乏规避无效信息的能力，这就是为什么此时中心区域的 PSNR 不高。当背景从 32 进一步扩展到 48 或更大时，网络开始对有效信息漠不关心。显然，它的远距离信息捕捉能力还不够强。对于 RCAN，与 EDSR 一样，它只能在背景大小为 32 × 32 时找到有用的信息。但在图 9 中可以观察到，与 EDSR 不同的是，注意力结构有助于 RCAN 专心关注关键信息、减少对无用像素的使用，反映在归因图中就是对背景的利用减少。因此，PSNR 在此处跃升。而对 SwinIR 来说，它的视野显然更为广阔。在图 10 中，可以发现当背景大小为 48 时，网络依然可以关注到障碍之外的有用信息。SwinIR 的 PSNR 指数还显示，当背景大小为 32 和 48 时，中心区域的恢复结果都能变得更好。不过，RCAN 和 SwinIR 对无关信息的忽略能力还是不够强大，它们对长程信息的争取伴随着对无关信息的利用，最终导致了模型性能的损伤。这既解释了（1）中的问题，也说明了现阶段的注意力机制还有一定的进步空间。

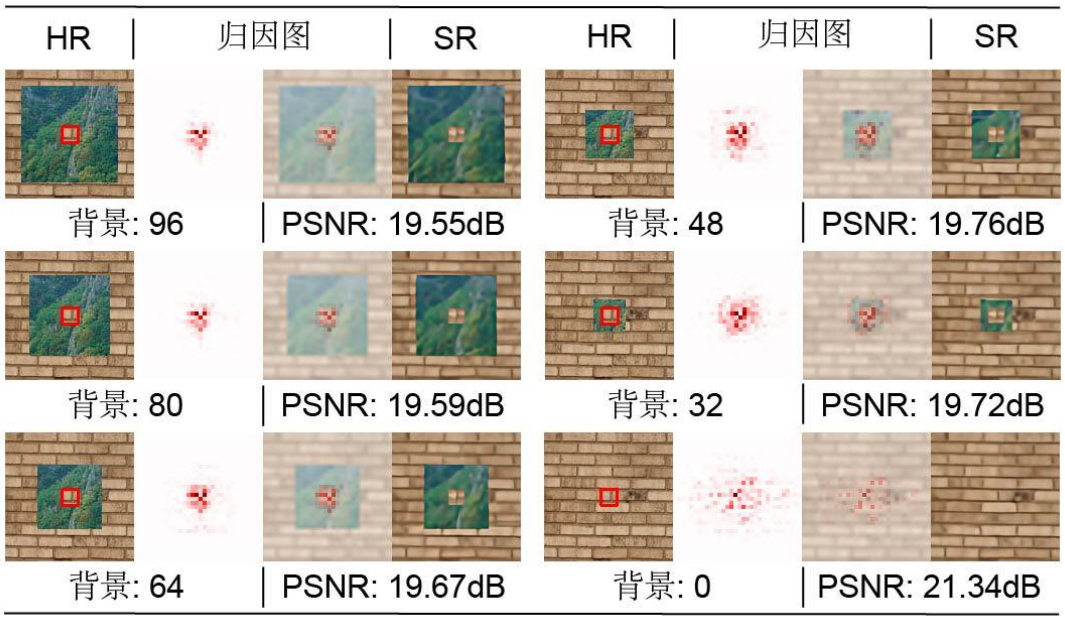

图 **8** EDSR 的具体数值和归因图

Fig. 8 Specific values and attribution map of EDSR

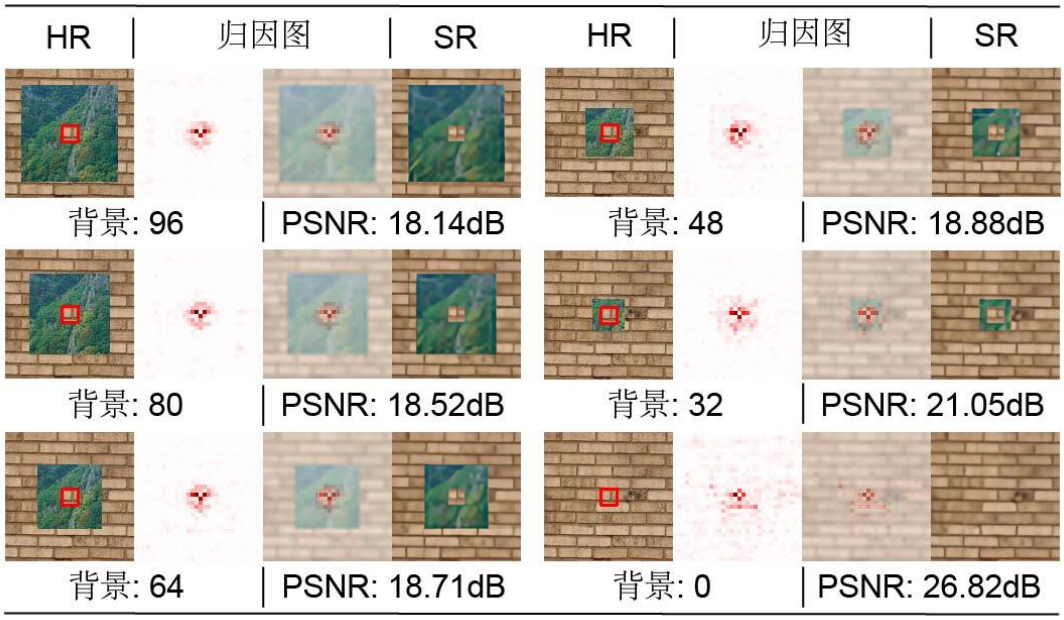

图 **9** RCAN 的具体数值和归因图

Fig. 9 Specific values and attribution map of RCAN

这样，结合解释工具 LAM，从背景和前景单独分析的角度，本章节解释了模型结构设计带来的网络行为的异同。注意力结构能帮助模型更专心地关注关键信息，规避对无用像素的使用，但这种规避能力是有限的。RCAN 和 SwinIR 在无关信息较多的情况下甚至达不到 EDSR 的恢复效果。

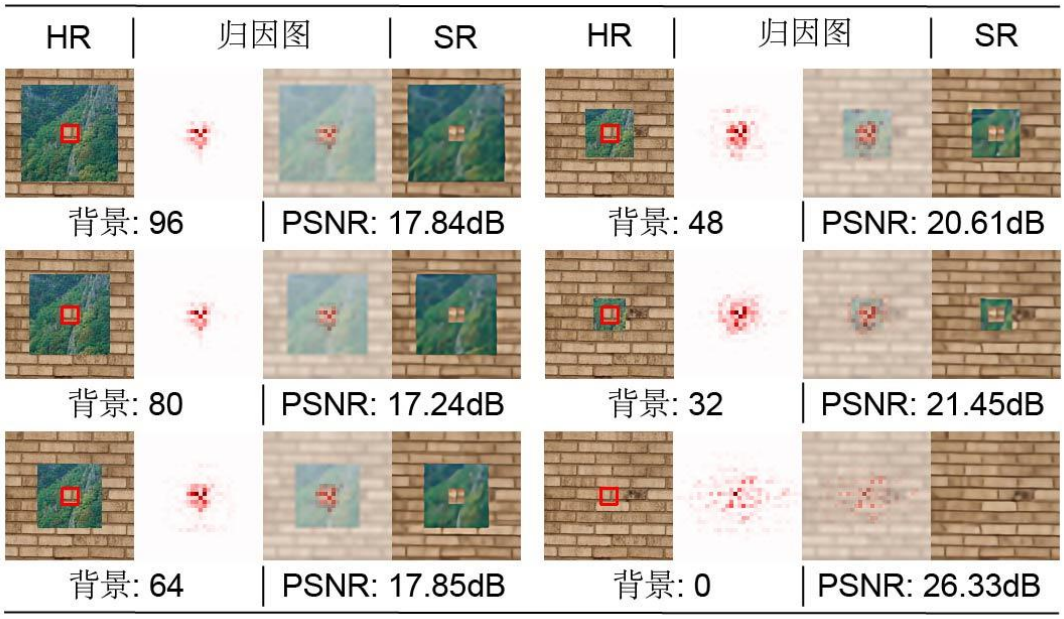

图 **10** SwinIR 的具体数值和归因图

Fig. 10 Specific values and attribution map of SwinIR

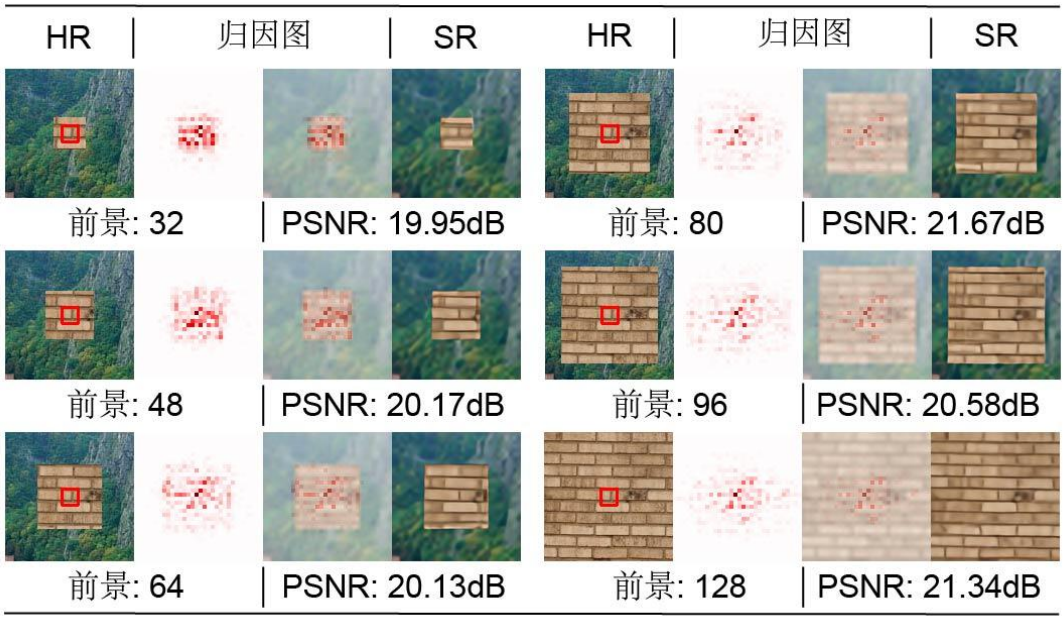

图 **11** EDSR 的具体数值和归因图(方法二)

Fig. 11 Specific values and attribution map of EDSR (Method 2)

此外，因为 RCAN 和 SwinIR 是具备大感受野的模型，本章节不仅从局部中心区域扩散非相似背景，也从非中心区域扩散（整体图像边缘向内扩散）了无关区域，如图 11、图 12 和图 13，以更完备地研究不同背景对不同模型的影响情况。观察可知，当无关背景从整体图像边缘向内扩散时，EDSR 发生了一定程度上的性能下降；RCAN 作为带有注意力结构的模型，更专心地关注关键信息，规避了对无用像素的使用，性能下降不大。但对于具备最大感受野的 SwinIR 来说，强大的长程信息利用能力使它的表征能力下降了近 2dB。显然，对有害背景的注意和利用导致了这一结果。

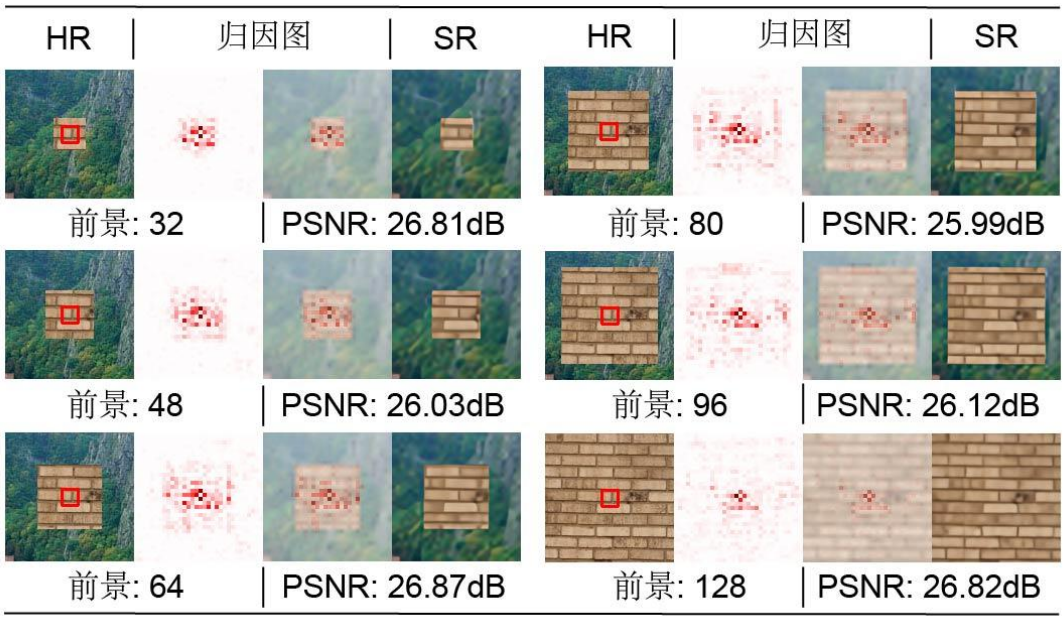

图 **12** RCAN 的具体数值和归因图(方法二)

Fig. 12 Specific values and attribution map of RCAN (Method 2)

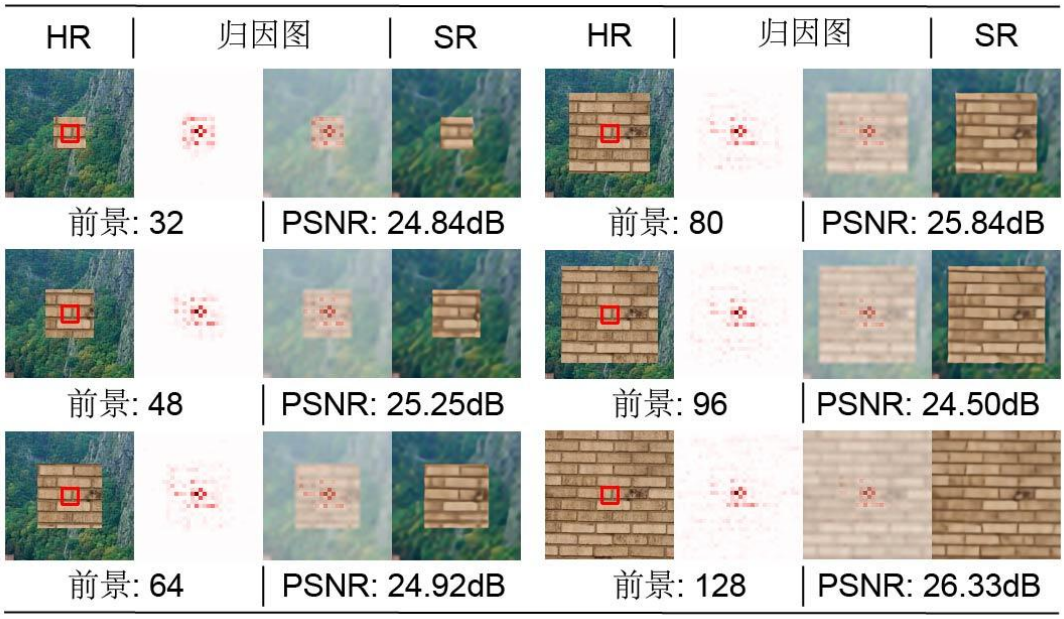

图 **13** SwinIR 的具体数值和归因图(方法二)

Fig. 13 Specific values and attribution map of SwinIR (Method 2)

6 未来方向

本项工作的关键在于发现和揭示现象，关于未来这类方法能否有更多的数值结论和更具体的应用，下文做了以下尝试和展望。

6.1 量化相似性

许多研究人员通过先提取图像本身的特征，再计算这些特征的差异性来衡量图像之间的关系。已经有许多从图像中提取特征的方法被广泛应用。Haralick[37]提出了灰度共生矩阵来描述纹理特征，还有一些工作关注图像中的颜色信息[38]。在这之中，格拉姆矩阵常被用在风格迁移的深度网络中[39, 40, 41, 42, 43]。其具体操作是使用神经网络提取图像的浅层特征图，然后计算特征图的格拉姆矩阵值。这个计算出的矩阵结合了图像的纹理、边缘和颜色特征。因此，完成风格迁移，通常目标是优化（降低）目标图像和风格图像之间的格拉姆差异。格拉姆矩阵具有公认的分析图像之间的相似性和差异性的功能。

前文中提到，研究过程中确实发现主观上相似的背景有助于前景的重建。因此，可以先使用格拉姆矩阵来量化前景与背景之间的相似性，下一步的目标是找到二者相似性与前景恢复性能之间的数值关联。具体做法是，先将前景和背景图像分别输入到 VGG16[44]网络中，用第一卷积层之后获得的特征图作为计算图像格拉姆矩阵的基础。关于这一操作，许多作品[45, 46] 中都有类似的讨论，这是风格迁移任务中的常见操作。然而，格拉姆矩阵只被用于掌握整个图像的一般风格，其自身的局限性在于，当图片之间的风格差距较小时，它不能非常准确地反映相似性。这也是该项研究下一阶段需要解决的主要问题，在量化格拉姆差和中心域 PSNR 的相关性时，结果并不明显。

6.2 数据增强方法

第 4.3 节提到当中心区域周围有与它类似的像素时，该区域的性能会更好。依靠这一发现，能否通过重新排布图像来提高图像中某区域的恢复性能呢？图 14 和图 15 中展示了可能的重排结果。即，选定某个区域作为重建目标，然后为该目标创建高度相似的背景。

背景可能由其自身组成，也可能由它翻转、旋转和缩放的结果组成。翻转可以是水平翻转或竖直翻转，旋转角度有顺时针 90°、180°、270° 三种，缩放倍数为 2 或 4，缩放方法为最近邻插值。最终生成的图像面积为原始区域面积的 25 倍，即长和宽都是它的 5 倍。将这样的图像输入到训练好的网络中进行测试时，可以发现目标区域确实恢复得更好（PSNR 更高）。请注意，此处的 PSNR 不是针对整个画面，它仅指所选区域的恢复程度。

从该方法很容易联想到测试时数据增强（Test-Time Augmentation, TTA），这是一种以计算资源为代价消除干扰的做法。TTA 的具体操作为，测试时将原始数据做不同形式的增强后再输入网络，如水平翻转、放大缩小和中心旋转等，然后逆回结果，取它们的平均值作为最终输出。不过，与 TTA 不同，本阶段的工作是为选定区域添加了有用的邻居，这显然是一种不同的数据增强方式。相比较于结果随所选的增强方法而变化的 TTA，本处的测试只需进行一次，结果不会因测试次数而改变。为了公平地显示这两种方法之间的差异，在消耗相同算力的前提下，本章节将 TTA 的结果与所提出方法的恢复结果进行了比较。使用 TTA 的做法为，选择一个要重建的区域进行 25 次数据增强测试。

结果表明，TTA 最多可带来 0.1dB 的改善。在某些情况下，由于增强方法的不当选取，结果可能比不使用 TTA 更糟糕。相比之下，除了稳定地提高 PSNR，使用本文中的方法没有损害性能的风险。图 14 中给出了两个显著改进的示例。不过，这种方法也有其局限性。当测试图像如图 15 所示随机生成时，大多数选定的测试区域（白色方框中）不会有很大的性能提高（但仍然没有降低性能的风险）。本文从 DIV2K 数据集中随机选取了 200 张图，执行上述随机增强操作，注意，这些数据大部分和图 15 中的类似，并非来自于人工挑选。表 2 给出了统计结果，可以发现，这种增强方法的增益稳定，没有损害性能的风险。当所选区域在原始图像中特殊时，即原始图像中该区域相似的部分占比较少时，此测试增强方法对性能是友好的，网络需要“看到”更多类似的纹理来锻炼对此特殊区域的恢复能力。

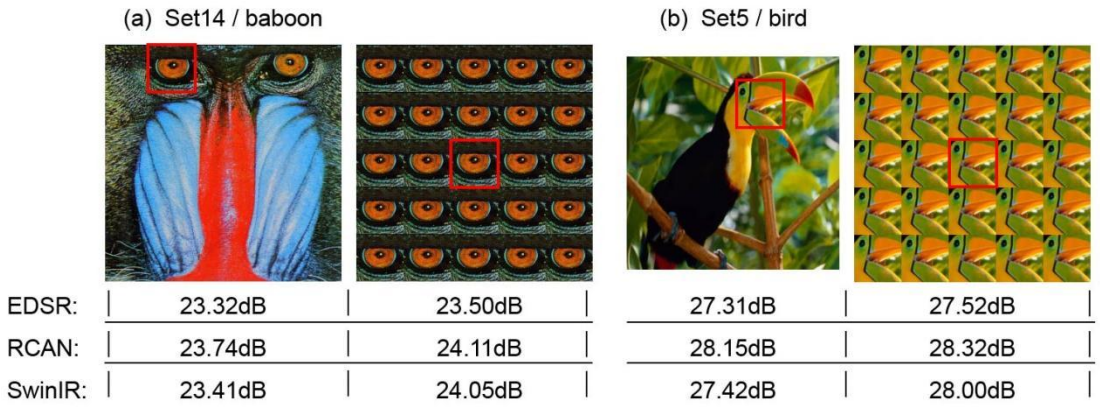

图 **14** 改进较大的示例

Fig. 14 Examples of great improvement

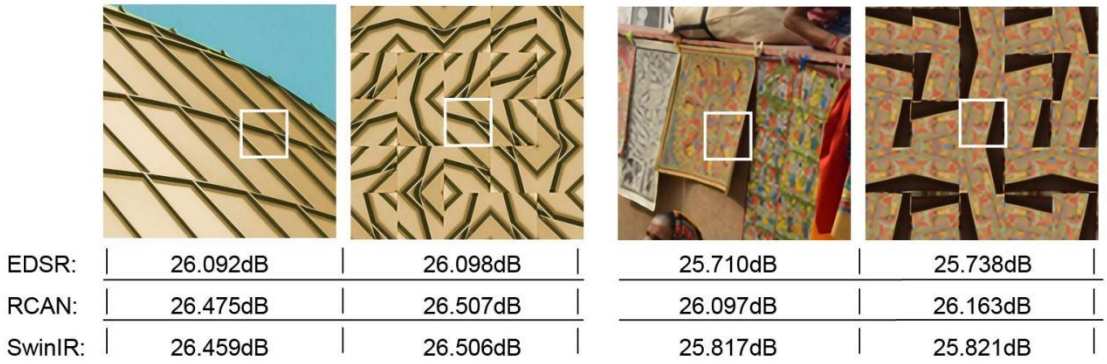

图 **15** 改进较小的示例

Fig. 15 Examples of less improvement

表 **2** 模型对中心区域的恢复
（输入为原图 / 增强图时）

Table 2 Models' recovery of central area (When the input is original / enhanced)

模型类别	恢复结果	
	原图(dB)	增强图(dB)
EDSR	23.02	23.28
RCAN	23.11	23.39
SwinIR	23.24	23.33

7 结语

本项工作发现，超分任务中某个位置的重建质量与其周围的“背景”密不可分，当需要恢复的区域与其附近的像素相似时，性能会更好。为证实这一结论，本研究设计了分析网络的新方法，即目标区域与其他区域的分割分析，并制作了针对该问题的数据集。不仅如此，借助这一发现，文章进一步探索了网络的工作机制，发掘了量化相似性的可能性，并提出了一种重做图像背景的数据增强方法。希望这项工作能为超分任务的分析带来新的视角，进而帮助研究人员们更好地理解网络行为，指导设计更好的网络和评估算法。

参考文献

[1] 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.

[2] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 1646-1654.

[3] Ledig C, Theis L, Huszár 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.

[4] Kong X, Zhao H, Qiao Y, et al. Classsr: A general framework to accelerate super-resolution networks by data characteristic[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 12016-12025.

[5] 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.

[6] Zhang Y, Tian Y, Kong Y, et al. Residual dense network for image super-resolution[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 2472-2481.

[7] Chen H, Gu J, Zhang Z. Attention in attention network for image super-resolution[J]. arXiv preprint arXiv:2104.09497, 2021. https://arxiv.org/abs/2104.09497

[8] Dai T, Cai J, Zhang Y, et al. Second-order attention network for single image super-resolution[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 11065-11074.

[9] Zhang Y, Li K, Li K, et al. Image super-resolution using very deep residual channel attention networks[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 286-301.

[10] 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.

[11] Lu Z, Liu H, Li J, et al. Efficient transformer for single image super-resolution[J]. arXiv preprint arXiv:2108.11084, 2021. https://arxiv.org/abs/2108.11084

[12] 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.

[13] Dong C, Loy C C, Tang X. Accelerating the super-resolution convolutional neural network[C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14. Springer International Publishing, 2016: 391-407.

[14] Kim J, Lee J K, Lee K M. Deeply-recursive convolutional network for image super-resolution[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 1637-1645.

[15] Mahendran A, Vedaldi A. Visualizing deep convolutional neural networks using natural pre-images[J]. International Journal of Computer Vision, 2016, 120: 233-255.

[16] Zhou B, Bau D, Oliva A, et al. Interpreting deep visual representations via network dissection[J]. IEEE transactions on pattern analysis and machine intelligence, 2018, 41(9): 2131-2145.

[17] Yosinski J, Clune J, Nguyen A, et al. Understanding neural networks through deep visualization[J]. arXiv preprint arXiv:1506.06579, 2015. https://arxiv.org/abs/1506.06579

[18] Zhao B, Wu X, Feng J, et al. Diversified visual attention networks for fine-grained object classification[J]. IEEE Transactions on Multimedia, 2017, 19(6): 1245-1256.

[19] Xiao T, Xu Y, Yang K, et al. The application of two-level attention models in deep convolutional neural network for fine-grained image classification[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 842-850.

[20] Jaderberg M, Simonyan K, Zisserman A. Spatial transformer networks[J]. Advances in neural information processing systems, 2015, 28.

[21] Lundberg S M, Lee S I. A unified approach to interpreting model predictions[J]. Advances in neural information processing systems, 2017, 30.

[22 Shrikumar A, Greenside P, Kundaje A. Learning important features through propagating activation differences[C]//International conference on machine learning. PMLR, 2017: 3145-3153.

[23] Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: Visualising image classification models and saliency maps[J]. arXiv preprint arXiv:1312.6034, 2013. https://arxiv.org/abs/1312.6034

[24] Springenberg J T, Dosovitskiy A, Brox T, et al. Striving for simplicity: The all convolutional net[J]. arXiv preprint arXiv:1412.6806, 2014. https://arxiv.org/abs/1412.6806

[25] Sundararajan M, Taly A, Yan Q. Axiomatic attribution for deep networks[C]//International conference on machine learning. PMLR, 2017: 3319-3328.

[26] Gu J, Dong C. Interpreting super-resolution networks with local attribution maps[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 9199-9208.

[27] Liu Y, Liu A, Gu J, et al. Discovering Distinctive" Semantics" in Super-Resolution Networks[J]. arXiv preprint arXiv:2108.00406, 2021. https://arxiv.org/abs/2108.00406

[28] Agustsson E, Timofte R. Ntire 2017 challenge on single image super-resolution: Dataset and study[C]//Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2017: 126-135.

[29] Bevilacqua M, Roumy A, Guillemot C, et al. Low-complexity single-image super-resolution

based on nonnegative neighbor embedding[J]. 2012.

[30] Yang J, Wright J, Huang T S, et al. Image super-resolution via sparse representation[J]. IEEE transactions on image processing, 2010, 19(11): 2861-2873.

[31] Martin D, Fowlkes C, Tal D, et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics[C]//Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001. IEEE, 2001, 2: 416-423.

[32] Matsui Y, Ito K, Aramaki Y, et al. Sketch-based manga retrieval using manga109 dataset[J]. Multimedia Tools and Applications, 2017, 76: 21811-21838.

[33] Huang J B, Singh A, Ahuja N. Single image super-resolution from transformed self-exemplars[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 5197-5206.

[34] Wang X, Yu K, Dong C, et al. Recovering realistic texture in image super-resolution by deep spatial feature transform[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 606-615.

[35] Cohen M R, Maunsell J H R. Attention improves performance primarily by reducing interneuronal correlations[J]. Nature neuroscience, 2009, 12(12): 1594-1600.

[36] Muqeet A, Iqbal M T B, Bae S H. HRAN: Hybrid residual attention network for single image super-resolution[J]. IEEE Access, 2019, 7: 137020-137029.

[37] Haralick R M, Shanmugam K, Dinstein I H. Textural features for image classification[J]. IEEE Transactions on systems, man, and cybernetics, 1973 (6): 610-621.

[38] Han J, Ma K K. Fuzzy color histogram and its use in color image retrieval[J]. IEEE Transactions on image Processing, 2002, 11(8): 944-952.

[39] Gatys L A, Ecker A S, Bethge M, et al. Controlling perceptual factors in neural style transfer[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 3985-3993.

[40] Gupta A, Johnson J, Alahi A, et al. Characterizing and improving stability in neural style transfer[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 4067-4076.

[41] Jing Y, Yang Y, Feng Z, et al. Neural style transfer: A review[J]. IEEE transactions on visualization and computer graphics, 2019, 26(11): 3365-3385.

[42] Li Y, Wang N, Liu J, et al. Demystifying neural style transfer[J]. arXiv preprint arXiv:1701.01036, 2017. https://arxiv.org/abs/1701.01036

[43] Li Y, Fang C, Yang J, et al. Universal style transfer via feature transforms[J]. Advances in neural information processing systems, 2017, 30.

[44] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014. https://arxiv.org/abs/1409.1556

[45] Shen F, Yan S, Zeng G. Neural style transfer via meta networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 8061-8069.

[46] Yanai K. Unseen style transfer based on a conditional fast style transfer network[J]. 2017.