第 11 卷 第 5 期
2022 年 9 月
集 成 技 术
JOURNAL OF INTEGRATION TECHNOLOGY
Vol. 11 No. 5
Sep. 2022

引文格式：
李哲远, 陈翔宇, 乔宇, 等. 注意力机制在单图像超分辨率中的分析研究 [J]. 集成技术, 2022, 11(5): 58-79.
Li ZY, Chen XY, Qiao Y. Research of single image super resolution based on attention mechanism [J]. Journal of Integration Technology, 2022, 11(5): 58-79.

注意力机制在单图像超分辨率中的分析研究

李哲远[1,2] 陈翔宇[1,3] 乔 宇[1] 董 超[1*]
井 焜[4] 刘辰飞[4] 许野平[4] 陈英鹏[4]

[1](中国科学院深圳先进技术研究院 深圳 518055)
[2](西北工业大学 西安 710072)
[3](澳门大学 澳门 999078)
[4](神思电子技术股份有限公司 济南 250098)

摘 要 基于卷积神经网络的单图像超分网络性能已经远超传统算法，为进一步提升网络表征能力及网络性能，许多研究在网络架构中使用了注意力机制。该文首先回顾注意力机制在单图像超分中的研究，并将其划分为基于一阶注意力机制和基于高阶注意力机制两类方法；然后，对比基于注意力机制的超分网络在网络规模、内存占用、计算量、网络损失类型和注意力机制架构差异，验证了不同注意力机制模块的性能差异，并使用最新的超分可视化分析工具为实验提供侧面证明；最后，分析和讨论基于注意力机制的算法研究在处理真实退化图像方面存在的挑战，指出超分技术发展的关键瓶颈及未来发展方向。

关键词 深度学习； 超分辨率； 注意力机制；计算机视觉；神经网络
中图分类号 TP 39；TP 751.1 **文献标志码** A doi: 10.12146/j.issn.2095-3135.20211209001

Research of Single Image Super Resolution Based on Attention Mechanism

LI Zheyuan[1,2] CHEN Xiangyu[1,3] QIAO Yu[1] DONG Chao[1*]
JING Kun[4] LIU Chenfei[4] XU Yeping[4] CHEN Yingpeng[4]

[1](*Shenzhen Institute of Advanced Technology*, *Chinese Academy of Sciences*, *Shenzhen* 518055, *China*)
[2](*Northwestern Polytechnical University*, *Xi'an* 710072, *China*)
[3](*University of Macau*, *Macao* 999078, *China*)
[4](*Synthesis Electronics Technology Co*., *Ltd*., *Jinan* 250098, *China*)
[*]*Corresponding Author*: *chao.dong@siat.ac.cn*

收稿日期：2021-12-09 **修回日期**：2022-01-03
作者简介：李哲远，本科生，研究方向为图像处理；陈翔宇，博士研究生，研究方向为图像处理；乔宇，研究员，研究方向为计算机视觉；董超(通讯作者)，副研究员，研究方向为计算机视觉，E-mail：chao.dong@siat.ac.cn；井焜，应用研究员，研究方向为人工智能；刘辰飞，高级工程师，研究方向为人工智能与图像处理；许野平，应用研究员，研究方向为人工智能与图像处理；陈英鹏，硕士研究生，研究方向为人工智能与图像处理。

Abstract　CNN-based methods have achieved notable performance in the research of single image super resolution domain. To further improve the representation ability and performance of networks, most research works have adopted the attention mechanism. In this survey, we introduce a taxonomy for the attention based super-resolution networks and classify existing methods into two categories: first-order and second-order attention. We also provide comparisons between the models in terms of network scale, memory footprint, type of network losses and important architectural differences for attention implementation. An analysis tool from recent network interpretation works is applied to verify the improvements of the evolving attention mechanism. Finally, we analyze and discuss challenges in processing real degraded images, and point out the problems and potential topics in future research work.

Keywords　deep learning; super resolution; attention mechanism; computer vision; neural network

1　引　言

单图像超分辨率(Single Image Super Resolution，SISR，以下简称“超分”)，旨在解决从低分辨率(Low-Resolution，LR)图像重建相应高分辨率(High-Resolution，HR)图像的问题，改善图像的细节和纹理，提升视觉质量。目前，超分技术已广泛应用于各领域，包括遥感、视频监控、医疗图像，以及图像分割、物体识别等高层视觉任务的预处理过程。超分技术在工业界和学术界都备受关注。

尽管大量研究已经提出了许多高效的方法来推动超分网络性能不断快速发展，但超分问题始终是一个长期存在的基础问题，在很多方面有待推进。由于超分是一个高度病态的问题，存在多个高分辨率图像与相应的低分辨率图像相对应，因此超分任务极具挑战性。此外，随着超分放大倍数的增加，问题的病态程度加剧，需要更多的先验信息来重建丢失的像素。

近年来，神经网络和深度学习——计算机视觉和模式识别研究中应用最为广泛的方法，利用大规模数据的强大学习能力克服了传统算法严重依赖手工特征的缺点，在计算机视觉领域取得了瞩目的成功。随着深度学习相关理论和技术的发展，研究人员已注意到卷积神经网络(Convolutional Neural Network，CNN)的潜力。Dong 等[1]最先在超分领域提出先驱性的工作——超分卷积神经网络(Super-Resolution Convolutional Neural Network，SRCNN)，探索了设计有效的超分网络的可能性。随后研究人员将最初应用于高层视觉任务和自然语言处理以增强深度网络表达能力的注意力机制应用在单图像超分网络上，使网络拟合能力大大增强，同时达到了最优的性能，这些先进的网络包括二阶注意力网络(Second-Order Attention Network，SAN)[2]、综合注意力网络(Holistic Attention Network，HAN)[3]、残差通道注意力网络(Residual Channel Attention Network，RCAN)[4]和 Swin 图像恢复网络(Image Restoration Using Swin Transformer，SwinIR)[5]等。

为了分析注意力机制在超分问题中的作用，以及不同注意力机制的有效性和效率，本文对注意力机制进行全面的分类和研究，总结了注意力机制的原理和发展过程。本文根据注意力机制的统计原理将相关网络分为两大类：基于一阶注意力的超分网络和基于高阶注意力的超分网络。为了进一步对比不同注意力机制的有效性和效率，本文设计了不同注意力机制模块的性能对比实

验，验证了部分注意力机制模块的相对性能。

本文的主要贡献有 4 点：(1)对比了不同特性的注意力机制的网络架构和统计原理；(2)根据注意力机制的统计原理提出了一种新的分类方式；(3)设计了不同注意力机制模块的性能对比实验；(4)总结现有研究的局限性，归纳展望了多个未来发展方向。

2 国内外的研究现状

针对图像超分的问题，国内外研究人员提出的各种算法和模型大致可以分为两类：一类是基于神经网络的深度学习算法[1-7]，另一类是模型传统算法[8-9]。由于篇幅所限，本文只介绍基于深度神经网络的超分算法，使网络专注于具有更多信息的通道。

自 SRCNN[1]成功地将深度学习网络应用于超分任务以来，各种有效和更深层次的超分方法架构被陆续提出。Tai 等[10]提出的超深度持久记忆网络(Very Deep Persistent Memory Network，MemNet)利用长期记忆网络进行多任务图像复原。Wang 等[11]将稀疏编码领域的知识与深度 CNN 结合，并训练级联网络逐步恢复图像。为了缓解梯度爆炸现象，降低模型复杂度，Kim 等[12]提出深度递归卷积网络(Deeply-Recursive Convolutional Network，DRCN)。Lai 等[13]提出的拉普拉斯金字塔超分辨率网络(Laplacian Pyramid Super-Resolution Network，LapSR)采用金字塔框架，通过 3 个子网络逐步生成×8 图像。Lim 等[7]通过去除批量归一化层修改了残差网络(Residual Network，ResNet)[14]，这极大提高了超分效果。

目前，注意力机制已成功应用于基于深度卷积神经网络的图像增强方法，帮助网络忽略无关信息而专注于重要信息。Zhang 等[4]提出的残差通道注意力网络(Residual Channel Attention Network，RCAN)允许网络专注于更多信息的通道。Choi 等[15]利用空间注意力机制，构建了 SelNet 超分网络。Dai 等[2]提出使用二阶统计量的注意力模块，使用二阶特征统计自适应地细化特征。在单层信息被充分利用的情况下，Niu 等[3]提出一个融合层注意力机制和通道空间注意力机制的整体注意力网络，以研究不同层、通道和位置的相互作用。Liang 等[5]结合卷积神经网络和基于自注意力机制的 Transformer，提出更具表达能力的 SwinIR，利用移位窗口对长程依赖进行建模，进一步提升了超分效果。

除了上述基于均方误差(Mean Square Error，MSE)最小化的方法外，研究人员还提出了感知约束以实现更佳视觉质量[16]的方法。SRGAN[17]使用生成对抗网络(Generative Adversarial Networks，GAN)，通过引入多任务损失，包括均方误差损失、感知损失[18]和对抗性损失[19]，来预测高分辨率输出。Zhang 等[20]根据纹理相似度从参考图像进一步转移纹理以增强纹理。

3 超分方法中的注意力机制

注意力机制是一种将可用计算资源偏向分配于信息量最大的信号的方法，应用于多个领域的研究，包括序列学习[21]、图像中的定位[22]和理解以及图像字幕[23]等。在这些应用中，注意力机制可以作为一个运算模块合并到高级抽象层，以便在模态之间进行适配。注意力机制最先由 Bahdanau 等[24]提出并应用于机器翻译。在 2017 ILSVRC 竞赛中，Hu 等提出冠军模型——SENet[25]，率先开发通道注意力机制，根据不同通道的重要程度，挖掘模型不同渠道相互依存关系。简而言之，注意力机制帮助网络忽略无关信息而专注于重要信息[26-27]。目前，注意力机制已成功应用于基于深度卷积神经网络的图像增强方面。

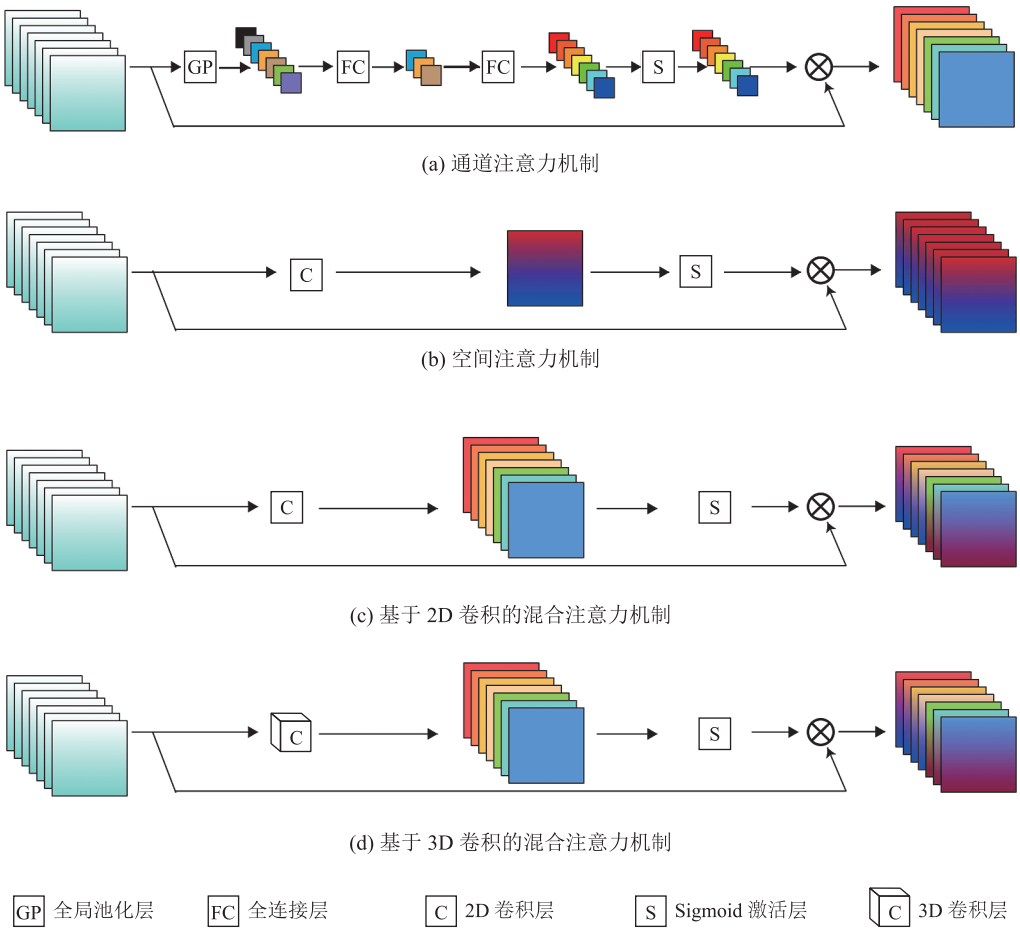
本文根据注意力机制的统计原理将相关网络分为两大类——基于一阶注意力的超分网络和基于高阶注意力的超分网络。一阶注意力超分网络的核心是使用特征的一阶统计量(如平均强度)生成注意力权重，而高阶注意力超分网络是利用高阶统计量(如协方差)或自相似性方法生成注意力权重。为进一步说明对应的注意力机制，本文对每种注意力机制列举一个具体的方法。

3.1　一阶注意力机制

3.1.1　通道注意力机制

随着深度学习的发展，Lim 等[7]充分发掘卷积神经网络在深度和广度两方面的潜力，提出了增强深度超分网络(Enhanced Deep Super-Resolution Network，EDSR)和多尺度深度超分网络(Multi-Scale Deep Super-Resolution Network，MDSR)。简单地提高网络的深度或广度已经很难获得较大的性能提升，如何进一步提高超分网络性能和构建可训练的网络成为当前亟待解决的问题。为了解决该问题，通道注意力机制被引入超分方法，其原理是通过对特征图的处理，生成 $1\times1\times C$ 特征权重，并捕捉每个通道之间的关系和重要程度，最后将该特征权重与每个对应的通道相乘，得到最终的加权特征图。

Zhang 等[4]提出的通道注意力机制的重要代表网络——残差通道注意力网络(Residual Channel Attention Network，RCAN)，在残差通道注意力模块(Residual Channel Attention Block，RCAB)中使用了图 1(a)所示的通道注意力机



(a) 通道注意力机制

(b) 空间注意力机制

(c) 基于 2D 卷积的混合注意力机制

(d) 基于 3D 卷积的混合注意力机制

图 1　注意力机制示意图

Fig. 1　Schematic diagram of the attention mechanisms

制。RCAB 包含 2 层卷积，1 个 ReLU[28]激活层和 1 个通道注意力模块，以及 1 个链接模块首尾的短连接。在通道注意力模块中，特征图通过非局部平均池化层，计算每个特征图的平均值，将特征图由 $H\times W\times C$ 压缩成 $1\times 1\times C$。为了从聚合信息中完全捕获通道维度的依赖关系，RCAN 借鉴 Hu 等[25]的方法，首先使用 1×1 的卷积层来收缩特征图以达到 $1\times 1\times C/r$，其中 r 是收缩比(Reduction Ratio)，设置 $r=16$。然后经过 1 个 ReLU 激活层，通过 1×1 的卷积层重新恢复到初始大小为 $1\times 1\times C$ 特征权重。最后经过 Sigmoid 函数，输出特征权重。新生成的特征权重已捕捉到每个通道之间的关系和重要程度，因此将该特征权重与每个对应的通道相乘，可得到最终的加权特征图。

3.1.2 空间注意力机制

空间注意力机制和通道注意力机制的作用方式类似，但它生成和原特征图大小相同的二维特征权重，旨在加强空间域上的“注意力”，使网络更加关注空间某些特定像素的信息，忽略冗余信息，凝聚模型的处理能力。尽管特征权重维度从一维扩展到了二维，但是超分问题和高层视觉问题有所不同，超分问题旨在恢复图像的边缘细节和高频纹理。在残差网络的作用下，低频信息通过长连接保留，而高频信息在网络主干中恢复，这与空间注意力机制的作用重复，故纯粹的空间注意力机制网络表现平庸。

受通道注意力机制的启发，Choi 等[15]提出选择单元，利用空间注意力机制，构建 SelNet 超分网络。如图 1(b)所示，选择单元由特征映射和选择模块组成。选择模块依次由 ReLU 激活层、卷积核大小为 1×1 的卷积层和 Sigmoid 函数层组成。选择模块计算空间域中的权重，生成二维的特征权重。

3.1.3 混合注意力机制

生成一维特征权重的方法存在忽略空间位置信息的局限性，生成二维特征权重的方法没有充分利用通道间的相互依赖关系。因此研究人员开始尝试结合两者的特点，发掘通道和位置之间的依赖关系，生成三维特征权重，进一步增强网络的表征能力。

3.1.3.1 基于 2D 卷积的混合注意力机制

Zhao 等[29]在通道注意力机制和空间注意力机制的启发下，提出像素注意力机制，并构建超分网络——像素注意力网络(Pixel Attention Network，PAN)。通道注意力机制通过空间非局部池化层生成一维特征权重，空间注意力机制通过通道池化层生成二维特征权重，但这些注意力机制在超分任务中效果不明显。如图 1(c)所示的像素注意力机制进一步使用像素级的三维特征权重，同时移除池化层，可以显著提高性能。

Zhao 等[29]提出的高效超分网络 PAN 主要由像素注意力自校准模块(Self-Calibrated block with Pixel Attention，SC-PA)和像素注意力上采样模块(Upsampling block with Pixel Attention，U-PA)组成。SC-PA 分为上下两部分，上层负责更高层的特征提取，下层负责维护原始信息。SC-PA 结构简单，没有复杂的连接和尺度变换操作，使用多个 2D 卷积核生成三维特征权重，更利于硬件加速。

U-PA 负责图像的重建步骤。目前，很少有超分网络重点研究重建步骤的结构，通常使用反卷积或像素混洗层和规则卷积。然而，这种结构多余且低效。为了进一步提升模型的效率，PAN 在 U-PA 卷积层中加入像素注意力模块(Pixel Attention，PA)，同时使用邻近上采样算法，进一步减少参数量。

3.1.3.2 基于 3D 卷积的混合注意力机制

空间注意力机制关注特征的平面维度，没有充分利用通道维度信息，而通道注意力机制又忽略平面信息。基于此，使用 3D 卷积捕获全部维度的信息，使通道空间注意力模块利用强大的表

达能力来描述连续通道的通道间和通道内信息。

通道注意力机制已被证明可以有效保留每一层信息丰富的特征，然而通道注意力机制将每个特征通道视为一个单独的过程，忽略了不同层之间的相关性。为了解决该问题，Niu 等[3]提出的混合注意力网络 HAN 在 RCAN 基础上增加了层注意力模块(Layer Attention Module，LAM)和基于 3×3×3 的 3D 卷积的通道空间注意力模块(Channel-Spatial Attention Module，CSAM)，为层、通道和位置之间的整体相互依赖性建模。

如图 1(d)所示，对于特征图 $\boldsymbol{F}_N \in R^{H\times W\times C}$，网络将 $\boldsymbol{F}_N$ 通过 3D 卷积层捕获通道和空间特征来生成注意力权重。HAN 通过将 3×3×3 的 3D 卷积核与 $\boldsymbol{F}_N$ 中多个相邻通道构建的立方体进行卷积操作，产生三维注意力特征，然后通过 Sigmoid 函数生成三维注意力权重。这样，HAN 的 CSAM 可以提取有效的特征来描述连续通道的通道间和通道内信息。

3.2　高阶注意力机制

3.2.1 基于高阶统计量的注意力机制

Dai 等[2]认为大多数基于 CNN 的超分模型都没有考虑特征的相互依赖性，尽管 SENet 通过重新调整通道特征的方法利用特征通道的相互依赖性，但使用全局平均池化层会导致网络忽略高于一阶的统计量，从而限制网络性能。Lin 等[30]和 Li 等[31]的相关工作也表明在深度卷积神经网络中，二阶统计量比一阶统计量的表达能力更强。

Dai 等[2]提出的 SAN 在通道注意力机制的基础上，增加了非局部增强残差组和二阶通道注意力模块。其中，二阶通道注意力模块，使用全局协方差池化层代替传统通道注意力机制中使用的一阶池化层，如全局平均池化层。全局协方差池化层可以通过以下一系列公式表示：

$$\boldsymbol{\Sigma}=\boldsymbol{X}\bar{\boldsymbol{I}}\boldsymbol{X}^{\mathrm{T}} \tag{1}$$

其中，$H\times W\times C$ 的特征图重塑为 $WH\times C$ 的特征矩阵 $\boldsymbol{X}$；$s\times s$ 的矩阵 $\bar{\boldsymbol{I}}$ 表示为 $\bar{\boldsymbol{I}}=\frac{1}{s}\left(\boldsymbol{I}-\frac{1}{s}\boldsymbol{1}\right)$，$\boldsymbol{I}$ 为 $s\times s$ 的特征矩阵，$\boldsymbol{1}$ 表示全部元素为 1 的矩阵。得到对称半正定协方差矩阵 $\boldsymbol{\Sigma}$ 后，通过以下特征值分解：

$$\boldsymbol{\Sigma}=\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{T}} \tag{2}$$

其中，$\boldsymbol{U}$ 为 1 个正交矩阵；$\boldsymbol{\Lambda}=diag\left(\lambda_1,\cdots,\lambda_C\right)$ 为具有非递增顺序特征值的对角矩阵。最后经过如下的归一化过程：

$$\hat{\boldsymbol{Y}}=\boldsymbol{\Sigma}^{\alpha}=\boldsymbol{U}\boldsymbol{\Lambda}^{\alpha}\boldsymbol{U}^{\mathrm{T}} \tag{3}$$

Li 等[31]研究表明，α=1/2 更具表达力，故 SAN 也在网络架构中沿用该参数。最后，归一化的协方差矩阵表示为 $\hat{\boldsymbol{Y}}=\left[\boldsymbol{y}_1,\cdots,\boldsymbol{y}_c\right]$，$C$ 维的通道统计量 z 由以下公式计算可得：

$$z_c=H_{GCP}\left(\boldsymbol{y}_c\right)=\frac{1}{C}\sum_{i}^{C}\boldsymbol{y}_c\left(i\right) \tag{4}$$

其中，$H_{GCP}\left(\cdot\right)$ 表示全局协方差池化操作。

3.2.2 基于卷积神经网络的非局部注意力机制

自注意力机制最早应用于自然语言处理领域，通过计算单词间的互相影响解决长距离依赖问题。在纯卷积神经网络中，卷积运算 1 次只处理 1 个局部邻域，难以充分利用非局部的信息。为了充分利用自然图像中的自相似性特征，加强网络对于重复纹理的注意力，研究人员在超分领域引入了图 2 所示的自注意力模块，生成自注意力特征权重来增强网络的特征提取和恢复能力。Liu 等[6]将自注意力机制应用于超分，和循环网络结合，提出了非局部循环网络(Non-Local Recurrent Network，NLRN)；Zhang 等[32]结合自注意力机制和残差网络 ResNet[14]提出了残差非局部注意力网络(Residual Non-Local Attention Network，RNAN)；Dai 等[2]在二阶通道注意力的基础上，融合区域自注意力机制，提出了 SAN。

对于给定的特征 $\boldsymbol{X}$，自注意力机制生成特征权重 $\boldsymbol{Z}_{i,j}$ 可以用以下公式来表示：

$$\boldsymbol{Z}_{i,j}=\sum_{g,h}\frac{\exp\left\{\phi\left(\boldsymbol{X}_{i,j},\boldsymbol{X}_{g,h}\right)\right\}}{\sum_{u,v}\exp\left(\phi\left(\boldsymbol{X}_{i,j},\boldsymbol{X}_{g,h}\right)\right)}\psi\left(\boldsymbol{X}_{g,h}\right) \tag{5}$$

其中，$(i,j),(g,h),(u,v)$是 $\boldsymbol{X}$ 的坐标；ψ 是特征变换函数；ϕ 是衡量相似性的函数。

$$\phi\left(\boldsymbol{X}_{i,j},\boldsymbol{X}_{g,h}\right)=\theta\left(\boldsymbol{X}_{i,j}\right)^{\mathrm{T}}\delta\left(\boldsymbol{X}_{g,h}\right) \tag{6}$$

其中，θ 和 δ 是特征变换函数。

与一般的利用空间自相似性原理的非局部注意力机制不同的是，HAN[3]的层注意力机制通过聚合多个网络中间特征图，并重塑聚合特征图，生成层间相似性特征权重。

HAN 的密集连接和跳过连接允许绕过浅层信息，但并没有利用不同层之间的相互依赖性。相比之下，HAN 的层注意力模块将每一层的特征图视为对特定类的响应，且不同层的响应相互关联。对层间特征使用非局部注意力机制以获取不同深度特征之间的依赖关系，使网络可以为不同深度的特征分配不同的注意力权重，并自动提高特征的提取能力。

3.2.3 基于图神经网络的非局部注意力机制

Zhou 等[33]提出的跨尺度内部图神经网络(Internal Graph Neural Network，IGNN)的灵感来自传统的基于自我实例[30]的超分方法。IGNN 的原理来自于经过统计验证的跨尺度切片重复属性[34]，因为在自然图像中，局部切片往往跨尺度重复多次。IGNN 将跨尺度相似块之间的这种内部相关性建模为一个图，其中每个切片是一个顶点，边是来自两个不同尺度切片之间的相似系数。基于该图结构，IGNN 能够处理不规则的图数据，并有效探索跨尺度递归属性。如图 3，相较于传统方法将跨尺度切片作为约束，IGNN 利用图模块来聚合高分辨率图像，包括图构建和切片聚合两个步骤。与其他超分网络仅从外部数据学习从低分辨率向高分辨率的映射不同，IGNN 充分利用了从低分辨率图像发现的 k 个最可能的高分辨率切片来恢复更详细的纹理。

在图构建环节，IGNN 首先通过 19 层的视觉几何群网络(Visual Geometry Group Network，VGG)[35]的前三层为低分辨率图像 $\boldsymbol{I}_L$ 和其下采样

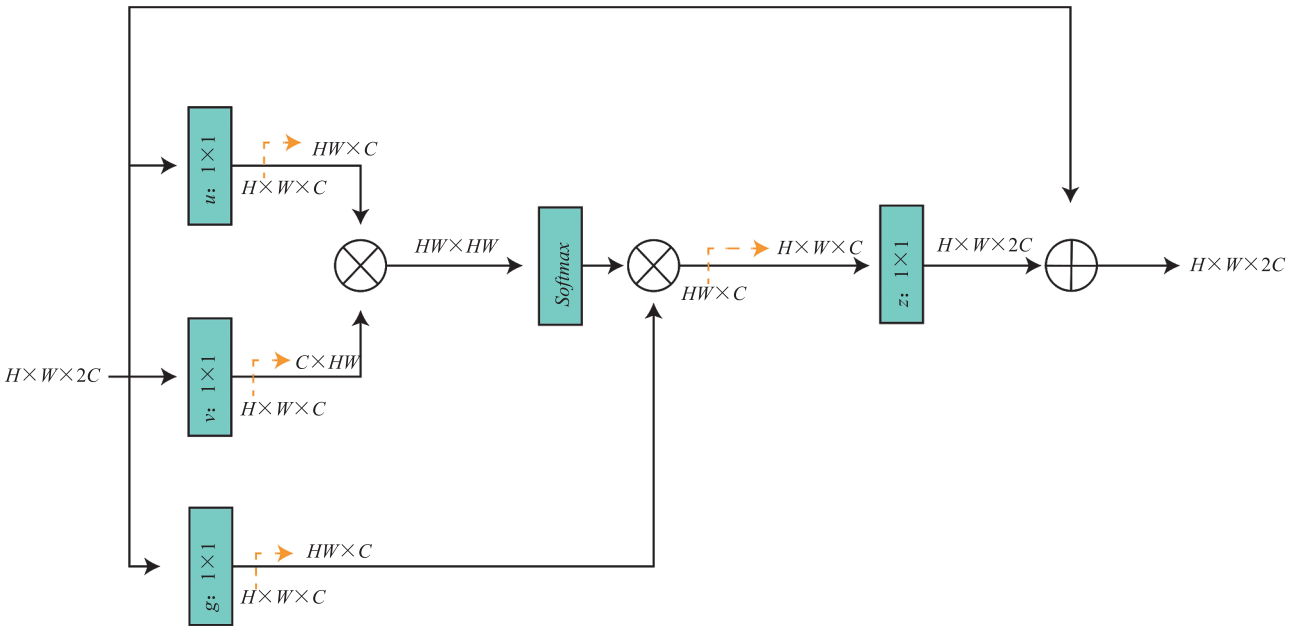


图 2 非局部注意力机制示意图

Fig. 2 Pipeline of non-local attention mechanism

图 $\boldsymbol{I}_{L\downarrow S}$ 生成特征图 $\boldsymbol{E}_L$ 和 $\boldsymbol{E}_{L\downarrow S}$。然后对每个待查询切片在 $\boldsymbol{E}_{L\downarrow S}$ 中搜索 k 个最近的邻接切片来动态构建跨尺度图 $\mathcal{G}_k$。将切片从 $\boldsymbol{E}_{L\downarrow S}$ 尺度的 k 个邻接切片映射到 $\boldsymbol{E}_L$ 尺度后，构建的跨尺度图 $\mathcal{G}_k$ 可以为每个待查询切片提供切片对。最终该跨尺度图可以表示为 $\mathcal{G}_k(\upsilon,\epsilon)$，顶点 υ 是低分辨率特征图 $\boldsymbol{E}_L$ 中的切片和高分辨率图像中的对应 k 个邻接切片，边 ϵ 包含了高分辨率切片和对应的 k 个邻接切片的相关系数。

在切片聚合部分，IGNN 借鉴边缘卷积(Edge-Conditioned Convolution)[36]的思想聚合了 k 个以边缘标签为条件的高分辨率切片，其过程可以表示为公式(7)：

$$\boldsymbol{F}_{L\uparrow S}^{q,l,s}=\frac{1}{\delta_q(\boldsymbol{F}_L)}\sum_{n_r\in S_q}\exp\left(\mathcal{F}_\theta\left(\boldsymbol{D}^{n_r\to q}\right)\right)\boldsymbol{F}_L^{n_r,ls},\forall q \quad (7)$$

其中，$\boldsymbol{F}_L^{n_r,ls}$ 表示输入 $\boldsymbol{F}_L$ 中第 r 个邻接 $ls\times ls$ 高分辨率切片；$\boldsymbol{F}_{L\uparrow S}^{q,l,s}$ 表示查询位置的高分辨率输出切片。通过“切片-图像”转换操作[37]，IGNN 将输入切片聚合成输出特征图 $\boldsymbol{I}_{L\uparrow S}$。然后使用一个边缘子网络(Edge-Conditioned Sub-Network，ECN)从嵌入特征 $\boldsymbol{E}_L$ 通过切片差异 $\boldsymbol{D}^{n_r\to q}$ 来计算每个邻接切片的聚合权重，即公式中的 $\mathcal{F}_\theta\left(\boldsymbol{D}^{n_r\to q}\right)$。公式中的 $\exp(\cdot)$表示底数为自然系数的指数函数，$\delta_q(\boldsymbol{F}_L)=\sum_{n_r\in S_q}\exp\left(\mathcal{F}_\theta\left(\boldsymbol{D}^{n_r\to q}\right)\right)$表示归一化因子。

为了进一步利用 $\boldsymbol{F}_{L\uparrow S}$，IGNN 使用一个下采样嵌入子网络(Downsampled-Embedding Sub-Network，DEN)对 $\boldsymbol{F}_{L\uparrow S}$ 进行下采样，然后拼接 $\boldsymbol{F}_L$ 生成 $\boldsymbol{F}_L'$，用于网络的后续层。

3.2.4 自注意力机制

大多数基于卷积神经网络的方法专注于精细的架构设计，如残差学习和密集连接等。虽然与传统基于模型的方法[38-39]相比，性能有了显著提升，但它们存在两个源于卷积层的基本问题。第一，图像和卷积核之间的交互与内容无关，使用相同的卷积核来恢复不同的图像区域不是最合理的选择；第二，在局部处理原理下，卷积对长程依赖建模效果不佳。Transformer[21]使用了一种自注意力机制来捕获上下文之间的全局交互，并在多个视觉问题[40-41]中表现出良好的性能，但会产生两个缺点：一是边界像素不能利用切片外的相邻像素进行图像恢复；二是恢复的图像可能会在每个切片周围引入边界伪影，虽然这个问题可以通过切片重叠来缓解，但它不可避免会引入额外的计算负担。

2021 年，Liu 等[22]提出的 Swin Transformer 融合了 CNN 和 Transformer 的优点，潜力巨大。如图 4 所示，其关键模块 Swin Transformer Layer 由多头自注意力模块(Multi-Head Self-Attention，

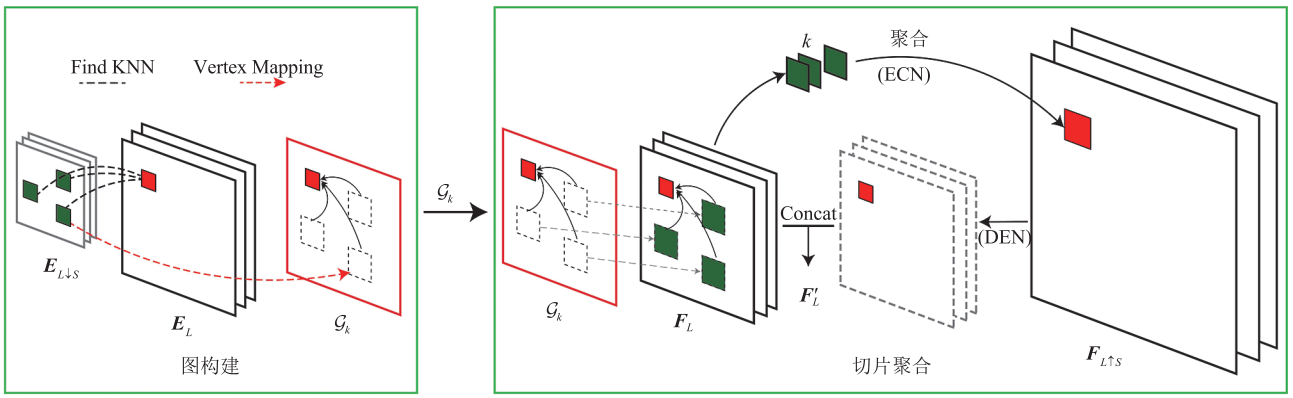


图 3　跨尺度内部图神经网络原理示意图[33]

Fig. 3　Schematic diagram of IGNN[33]

MSA)和多层感知器(Multi-Layer Perceptron，MLP)组成。

Liang 等[5]在 Swin Transformer 的基础上，提出一种图像恢复模型——SwinIR。该模型一方面利用局部注意力机制具有处理大尺寸图像的优势，另一方面具有 Transformer 的优势，可以利用图 5 所示的移窗机制对长程依赖进行建模。如图 5 所示，给定大小为 $H\times W\times C$ 的输入特征图，首先将输入划分为不重叠的大小为 $M\times M$ 的局部窗口，使其重塑为 $HW/M^2\times M^2\times C$ 大小的特征图，其中 HW/M^2 为窗口总数。然后，分别计算每个窗口的标准自注意力权重。对于局部窗口特征图 $\boldsymbol{X}\in\mathbf{R}^{M^2\times C}$，查询、键和值矩阵 $\boldsymbol{Q}$、$\boldsymbol{K}$ 和 $\boldsymbol{V}$ 的计算方式如下：

$$\boldsymbol{Q}=\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{Q}},\boldsymbol{K}=\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{K}},\boldsymbol{V}=\boldsymbol{X}\boldsymbol{P}_{\boldsymbol{V}} \tag{8}$$

其中，$\boldsymbol{P}_{\boldsymbol{Q}}$、$\boldsymbol{P}_{\boldsymbol{K}}$ 和 $\boldsymbol{P}_{\boldsymbol{V}}$ 是跨不同窗口共享的投影矩阵。故局部窗口中的注意力权重计算如下：

$$Attention(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=SoftMax\left(\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}/\sqrt{d}+\boldsymbol{B}\right)\boldsymbol{V} \tag{9}$$

其中，$\boldsymbol{B}$ 是可学习的相对位置编码，d 是查询、键的维度。而后并行执行 h 次注意力函数并将结

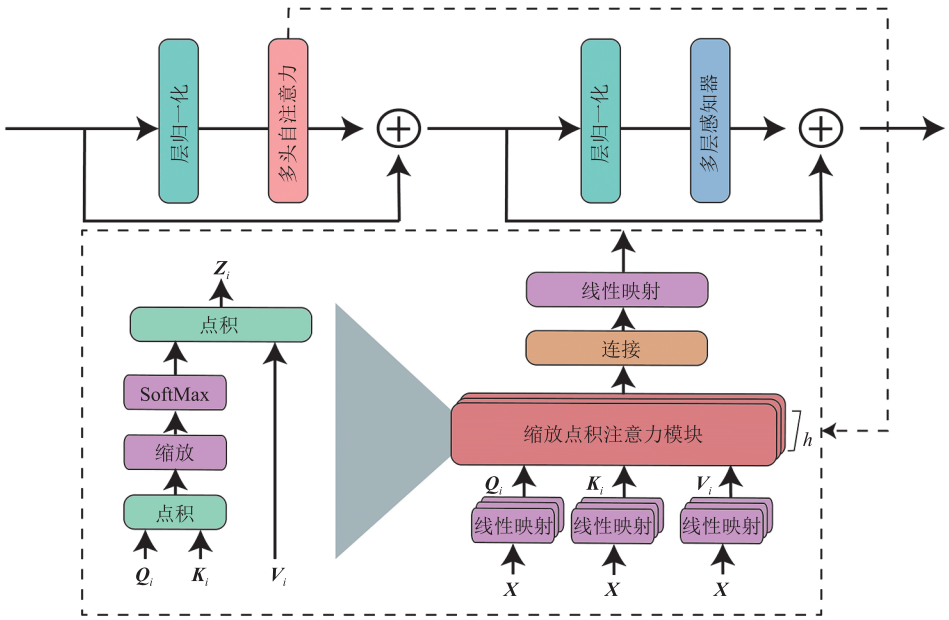


图 4 Swin Transformer 层示意图

Fig. 4 Pipeline of Swin Transformer Layer

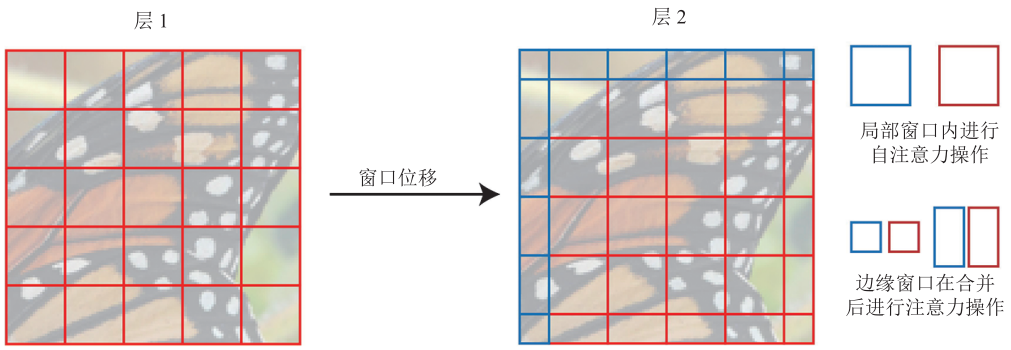


图 5 移窗机制示意图

Fig. 5 An illustration of shifted windows

果连接起来用于 MSA 模块。

4　实验评估

4.1　数据集和实验设置

本文使用 DIV2K 数据集[42]作为训练数据集，其中包含 800 张训练图像。训练使用的低分辨率图像由高分辨率图像经过 Matlab 双三次下采样获得，同时在训练过程中对 800 张训练图像随机旋转 90°、180°、270° 并水平翻转进行数据增强。

所有网络都采用了阶梯式学习率下降的方法，初始学习率为 10^{-4}，每迭代 2×10^5 次学习率下降一半，最小学习率设为 $\epsilon=10^{-8}$，共训练 8×10^5 次迭代。统一使用了 Adam 优化器[43]，其中 Adam 优化器的超参 $\beta_1=0.9$，$\beta_2=0.999$，批量大小(Batch Size)设置为 32，采用低分辨率图像为 48×48 大小的图像切片作为训练输入。本文所有网络均使用 Pytorch 框架[44]，在 Nvidia GeForce RTX 3090 上训练。

本文使用 5 个标准基准数据集：Set5[45]、Set14[46]、B100[47]、Urban100[48]、Manga109[49]进行评估。高分辨率结果通过 YCbCr 空间的 Y 通道上的峰值信噪比(Peak Signal to Noise Ratio，PSNR)和结构相似性指数[50](Structural Similarity Index，SSIM)进行评估。

目前，研究人员已分析比较多个损失函数，包括 $\mathcal{L}_1$ 损失函数[2,7,51]、$\mathcal{L}_2$ 损失函数[1,17,52]和生成对抗损失函数[17]。本文主要聚焦注意力机制，故不讨论使用生成对抗损失函数的算法。早期一些算法 SRCNN[1]、FSRCNN[53]等都使用 $\mathcal{L}_2$ 损失函数，而在基于注意力机制的超分算法中，所有算法均使用了 $\mathcal{L}_1$ 损失函数。

4.2　定量分析

表 1 展示了超分算法在峰值信噪比和结构相似性指数度量的结果。SAN[2]、HAN[3]、RCAN[4]和 SwinIR[5]均表现出了优秀的性能，其中 HAN 的性能在卷积神经网络汇总中表现最佳，SwinIR 在所有测试集中的表现与其他方法都拉开了显著的差距。

4.3　参数量分析

表 2 展示了不同超分网络的对比分析。其中 AIM 2020[54]比赛中的高效超分网络 PAN 模型的超分效果较差，但激活次数(M)、卷积数量、浮点运算量(G)、参数量(M)、显存占用(M)和平均计算时间(s)最优。RNAN 与 PAN 相比，尽管超分效果领先较大，但模型规模急剧膨胀，且由于网络中的残差非局部注意力模块(Residual Non-Local Attention Block，RNAB)的特性，在目标图像尺寸较大的情况下，显存占用会急剧提升，以致在测试过程中，Urban100 和 Manga109 数据集不能直接在 Nvidia GeForce RTX 3090 上运行。如表 1 所示，在 Set5 数据集的 4 倍超分结果中，IGNN 相比于 RNAN 虽然在 PSNR 指标上提升了 0.08 dB，但由于该网络使用 VGG19 和基于多切片聚合图神经网络的原因，网络的参数量、运算量、显存占用和计算时间等与其他网络相比都产生了非常大的差距。本文所调研的模型中，计算成本和运行时间成本最大的网络是 RNAN，但其效果并非最优。RCAN、SAN 和 HAN 的超分性能优秀，且激活次数、运算量和参数量接近。其中，SAN 的卷积层虽然较少，但由于网络的区域级非局部模块需要生成协方差矩阵和非局部注意力权重矩阵，在较大的图像上，会产生大量额外的计算开销和显存占用，以致在测试过程中 Urban100 和 Manga109 数据集不能直接在 Nvidia GeForce RTX 3090 上运行。

4.4　消融实验

不同的注意力机制基于不同的原理，通过“注意”层、通道、位置之间的信息，增强网络表达能力和恢复能力。为了进一步比较不同注意力模块的性能，表 3 展示了在相同骨干网络下，

表 1 不同超分网络的定量结果

Table 1 Quantitative results of different SR networks

方法	超分倍数	Set5		Set14		BSD100		Urban100		Manga109	
		PSNR (dB)	SSIM	PSNR (dB)	SSIM	PSNR (dB)	SSIM	PSNR (dB)	SSIM	PSNR (dB)	SSIM
EDSR	×2	38.11	0.960 2	33.92	0.919 5	32.32	0.901 3	32.93	0.935 1	39.10	0.977 3
RCAN	×2	38.27	0.961 4	34.12	0.921 6	32.41	0.902 7	33.34	0.938 4	39.44	0.978 6
PAN	×2	38.00	0.960 5	33.59	0.918 1	32.18	0.899 7	32.01	0.927 3	38.70	0.977 3
IGNN	×2	38.24	0.961 3	34.07	0.921 7	32.41	0.902 5	33.23	0.938 3	39.35	0.978 6
HAN	×2	38.27	0.961 4	34.16	0.921 7	32.41	0.902 7	33.35	0.938 5	39.46	0.978 5
RNAN	×2	38.17	0.961 1	33.87	0.920 7	32.32	0.901 4	32.73	0.934 0	39.23	0.978 5
SAN	×2	38.31	0.962 0	34.07	0.921 3	32.42	0.902 8	33.10	0.937 0	39.32	0.979 2
SwinIR	×2	38.35	0.962 0	34.14	0.922 7	32.44	0.903 0	33.51	0.940 1	39.70	0.979 4
EDSR	×3	34.65	0.928 0	30.52	0.846 2	29.25	0.809 3	28.80	0.865 3	34.17	0.947 6
RCAN	×3	34.74	0.929 9	30.65	0.848 2	29.32	0.811 2	29.09	0.870 2	34.44	0.949 9
PAN	×3	34.40	0.927 1	30.36	0.842 3	29.11	0.805 0	28.11	0.851 1	33.61	0.944 8
IGNN	×3	34.72	0.929 8	30.66	0.848 4	29.31	0.810 5	29.03	0.869 6	34.39	0.949 6
HAN	×3	34.75	0.929 9	30.67	0.848 3	29.32	0.811 0	29.10	0.870 5	34.48	0.950 0
RNAN	×3	34.66	0.929 4	30.52	0.846 0	29.26	0.805 9	28.76	0.864 8	34.22	0.948 3
SAN	×3	34.75	0.930 0	30.59	0.847 6	29.33	0.811 2	28.93	0.867 1	32.30	0.949 4
SwinIR	×3	34.89	0.931 2	30.77	0.850 3	29.37	0.812 4	29.29	0.874 4	34.74	0.951 8
EDSR	×4	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
RCAN	×4	32.63	0.900 2	28.87	0.788 9	27.77	0.743 6	26.82	0.808 7	31.22	0.917 3
PAN	×4	32.13	0.894 8	28.61	0.782 2	27.59	0.736 3	26.11	0.785 4	30.51	0.909 5
IGNN	×4	32.57	0.899 8	28.85	0.789 1	27.77	0.743 4	26.84	0.809 0	31.28	0.918 2
HAN	×4	32.64	0.900 2	28.90	0.789 0	27.80	0.744 2	26.85	0.809 4	31.42	0.917 7
RNAN	×4	32.49	0.898 2	28.83	0.787 8	27.72	0.742 1	26.61	0.802 3	31.09	0.914 9
SAN	×4	32.64	0.900 3	28.92	0.788 8	27.78	0.743 6	26.79	0.806 8	31.18	0.916 9
SwinIR	×4	32.72	0.902 1	28.94	0.791 4	27.83	0.745 9	27.07	0.816 4	31.67	0.922 6

表 2 不同超分网络的模型对比

Table 2 Model summary of different SR networks

方法	激活次数 (M)	卷积核	每秒浮点运算次数 (G)	参数量 (M)	显存占用 (M)	平均处理时间-Set14 (s)	损失函数
RCAN	1 815.098 7	815	1 044.025 3	15.592 4	604.476	0.086 647	$\mathcal{L}_1$
PAN	270.532 6	121	32.190 8	0.272 4	299.845	0.027 809	$\mathcal{L}_1$
HAN	1 826.829 6	819	1 075.442 3	16.071 8	846.308	0.087 234	$\mathcal{L}_1$
IGNN	2 195.701 8	136	4 820.970 0	49.512 6	1 3801.464	0.230 422	$\mathcal{L}_1$
RNAN	885.194 8	235	545.965 8	9.255 0	5 212.662	0.133 644	$\mathcal{L}_1$
SAN	1 867.711 8	498	1 066.046 9	15.860 5	7 775.214	0.396 803	$\mathcal{L}_1$
SwinIR	185.598 0	12	500.559 5	11.900 2	723.638	0.120 190	$\mathcal{L}_1$

注：运算量(G)、激活次数(M)和显存占用(M)，均在低分辨率图像尺寸为256×256的图像上进行测试

不同注意力模块的性能表现，所有网络模型均取最后一次训练结果。

HAN 中的层注意力模块和通道空间注意力模块对网络性能都产生显著的提升效果。在

RCAN 的基础上添加通道注意力模块后，网络在 Urban100 和 Manga109 数据集的超分效果分别提升了 0.118 8 dB 和 0.154 4 dB。使用层注意力模块后，网络 SSIM 指标有轻微提升。

PAN 性能最差，表明简单堆叠像素注意力模块并不能获得更好的性能。对比表 4，像素注意力机制在 RCAN 骨干网络上性能较差不是因为像素注意力模块的参数量较小，而是因为网络结构的限制。在加大网络深度后，PAN 的性能仍然没有明显提升。

SAN 使用协方差特征矩阵传递特征矩阵性能明显优于直接传递特征矩阵的 RCAN，在 Urban100 测试集上超分效果提升了 0.110 6 dB，表明二阶统计量对于网络的性能有显著的提升。

表 4 展示了以 MDSR 的参数量为基准，在近似参数量的情况下各个注意力模块的性能。HAN 相比于 RCAN 主要改进的是 LAM 和 CSAM。在整体参数量相对 HAN 原网络较小的情况下，HAN-L 中的 LAM 没有如原模型对网络起到增强作用，甚至对模型起到一定的负作用。而在不使用 LAM 的模型 HAN-L-woLA 中，CSAM 对 HAN 的性能明显增强，相比于 RCAN 在大

表 3　RCAN 骨干网络下不同注意力模块的性能对比

Table 3　Comparison of different attention modules based on the same backbone

方法	超分倍数	Set5		Set14		BSD100		Urban100		Manga109	
		PSNR (dB)	SSIM	PSNR (dB)	SSIM	PSNR (dB)	SSIM	PSNR (dB)	SSIM	PSNR (dB)	SSIM
HAN-wLA	×4	32.575 5	0.900 1	28.769 7	0.787 5	27.714 7	0.742 2	26.612 9	0.803 3	31.004 3	0.916 4
HAN-woLA	×4	32.527 2	0.899 6	28.776 9	0.787 4	27.709 9	0.741 9	26.612 5	0.803 0	31.028 6	0.916 2
SAN-wNL	×4	32.461 6	0.898 6	28.789 8	0.787 0	27.677 5	0.740 4	26.523 0	0.799 8	30.928 3	0.914 5
SAN-woNL	×4	32.462 6	0.898 6	28.772 4	0.786 9	27.713 7	0.741 7	26.604 3	0.802 5	31.034 5	0.915 7
PAN	×4	32.362 0	0.897 9	28.779 8	0.786 1	27.694 9	0.740 7	26.508 0	0.799 5	30.840 1	0.913 6
RCAN	×4	32.482 5	0.890 0	28.722 0	0.786 1	27.686 2	0.741 4	26.493 7	0.800 5	30.874 2	0.914 7

注：w(o)LA 表示(不)使用层注意力模块；w(o)NL 表示(不)使用非局部注意力机制(网络训练过程中均未加载预训练模型)

表 4　参数量相近的不同注意力网络的性能对比

Table 4　Comparison of different attention networks with comparable parameters

方法	超分倍数	Set5		Set14		BSD100		Urban100		Manga109		Parameters (M)
		PSNR (dB)	SSIM	PSNR (dB)	SSIM	PSNR (dB)	SSIM	PSNR (dB)	SSIM	PSNR (dB)	SSIM	
MDSR	×4	32.350 0	0.897 5	28.750 1	0.786 0	27.685 4	0.740 3	26.487 1	0.798 9	30.855 4	0.913 7	7.954 0
HAN-L-wLA	×4	32.461 1	0.898 4	28.787 6	0.786 5	27.692 5	0.741 4	26.537 1	0.800 1	30.928 4	0.915 1	8.8665 1
HAN-S-wLA	×4	32.500 6	0.898 5	28.746 6	0.785 6	27.670 3	0.740 1	26.528 0	0.799 1	30.934 1	0.914 8	7.920 7
HAN-L-woLA	×4	32.513 9	0.899 0	28.772 7	0.786 8	27.699 5	0.740 9	26.557 7	0.801 0	30.916 2	0.915 0	8.593 1
SAN-wNL	×4	32.381 3	0.897 6	28.772 4	0.786 4	27.688 5	0.740 0	26.513 5	0.798 8	30.885 7	0.914 0	8.099 8
SAN-woNL	×4	32.446 5	0.898 3	28.764 5	0.786 2	27.696 9	0.740 7	26.526 3	0.800 1	30.977 0	0.914 5	8.096 6
RNAN-wNL	×4	32.192 1	0.894 8	28.653 8	0.783 1	27.589 0	0.737 6	26.194 8	0.789 9	30.404 5	0.907 5	7.474 5
RNAN-woNL	×4	32.244 2	0.895 8	28.731 0	0.785 4	27.657 4	0.739 0	26.360 5	0.794 9	30.731 6	0.911 7	7.457 8
PAN	×4	32.424 2	0.898 0	28.760 1	0.786 4	27.694 6	0.740 4	26.506 7	0.799 1	30.886 5	0.914 1	7.818 5
RCAN	×4	32.451 3	0.898 5	28.733 2	0.786 4	27.694 2	0.741 1	26.548 1	0.800 2	30.849 4	0.914 0	8.148 8

注：w(o)LA 表示(不)使用层注意力模块；w(o)NL 表示(不)使用非局部机制；HAN-L 表示使用较大参数量的 HAN 网络；HAN-S 表示使用较小网络参数量的 HAN 网络(网络训练过程中均未加载预训练模型)

部分指标上均有明显提升。HAN-S-wLA 相比于 RCAN 在 Set5、Set14 数据集测试具有明显的性能优势，其在 Manga109 数据集上以 0.08 dB 大幅强于 RCAN，但是在 BSD100 和 Urban100 数据集上 PSNR 指标落后约 0.02 dB，SSIM 指标落后约 0.001。在参数量较小且没有使用 RL-NL 模块的情况下，SAN 与采用一阶统计量的 RCAN 性能接近，没有表现出明显的优势。高效超分网络 PAN 在增加网络宽度和深度后，性能明显提升，与 RCAN 的性能相当。RNAN 在减少网络参数量后，性能急剧下降。

4.5 训练技巧

在训练过程中，PAN 放大倍数为 4 的网络可以直接使用以上设置获得理想的效果，而 IGNN、SAN、HAN、RNAN 和 RCAN 直接使用 4.1 节中的训练设置会与原文的效果产生一定差距，尤其是 HAN 容易出现梯度爆炸导致训练崩溃，需要加载 RCAN 或 HAN 预训练模型才能使训练正常进行。HAN、SAN、RNAN 和 RCAN 需要通过加载×2 预训练模型，设置学习率为 10^{-5}，训练 2×10^5 次迭代，才可以得到原文的结果。

4.6 可视化

图 6 展示了各网络在“YumeiroCooking”和“img_004”图像×4 超分的可视化对比。在“YumeiroCooking”可视化对比中，通过视觉感受和 PSNR/SSIM 指标量化分析可得，HAN、RCAN 的视觉效果明显优于其他超分网络。其中，HAN 的效果最佳，是唯一 PSNR 高于 30 dB 的网络，线条边界清晰，方向准确；RCAN 效果次佳，方向准确，但右侧条纹密集区域轻微模糊。其他网络均存在明显的边缘模糊和条纹方向错误问题。观察“img_004”，其中 IGNN、RCAN 和 SAN 的视觉效果和量化指标较好。尽管在 IGNN 恢复的白色网格中存在一些黑色伪影，且将椭圆形网格恢复成了接近方形的情况，但网络视觉效果和指标均为最佳。而 RCAN 和 SAN 只生成了部分网格，SAN 在黑色网格部分效果明显弱于 RCAN。

量化性能最好的 SwinIR 在图 6 没有表现出非常强的复原效果。为了进一步验证 SwinIR 的性能，如图 7 所示，本文挑选 Urban100 数据集中场景更加复杂的“img_073”和“img_076”，对比图像边缘和复杂结构叠加人脸纹理的超分效果。对比“img_073”的边缘大楼效果，PAN、RNAN、IGNN 和 SAN 都产生了方向错误的线条，整个大楼的透视角度完全错误。HAN 的输出结果尽管透视角度准确，但与 RCAN 和 SwinIR 相比，边缘明显更为模糊。相似的效果在“img_076”也有所体现，仅有 HAN、RCAN、SwinIR 输出了清晰的纹理，但均受到人脸皮肤纹理的干扰，将矩形结构错误恢复成六边形结构。

总体来说，SwinIR、HAN 和 RCAN 的视觉效果最佳。HAN 和 RCAN 对单一纹理表现出非常强的恢复效果，而 SwinIR 能够较好地处理复杂场景和图像边缘。HAN 和 RCAN 的优秀视觉效果说明，基于通道注意力机制的主干网络明显提升单一纹理的复原效果，而对于复杂场景和叠加纹理，基于 Transformer 的自注意力机制表现出非常强的表征能力。

4.7 局部归因图分析

局部归因图是由 Gu 等[55]提出的一个超分网络解释性方法，继承了积分梯度的方法，可以直观展示网络的实际感受野。如图 8 所示，通过分析不同算法的局部归因图，观察到不同算法的“视野”，分析不同算法的行为模式。图中的红色高亮部分表示对超分图片影响最大的部分，理论上，对于相同的局部切片，越大的局部归因图代表网络利用了更多像素中的信息。

从图 8 分析可得，除了高效超分网络 PAN 以外，RCAN、RNAN、SAN、HAN 均通过非局

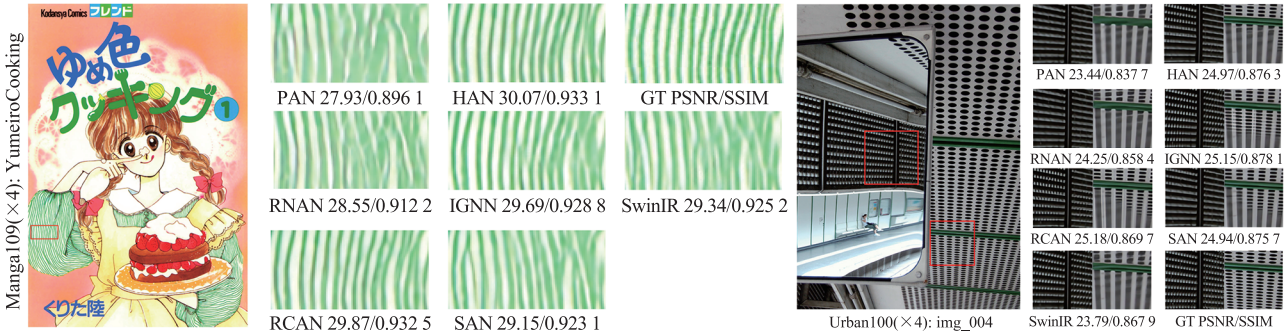


图 6　Urban100 测试集中 img_004 和 Manga109 测试集中 YumeiroCooking 在×4 超分的可视化对比

Fig. 6　Visual comparison for ×4 SR on img_73 in Urban100 and YumeiroCooking in Manga109

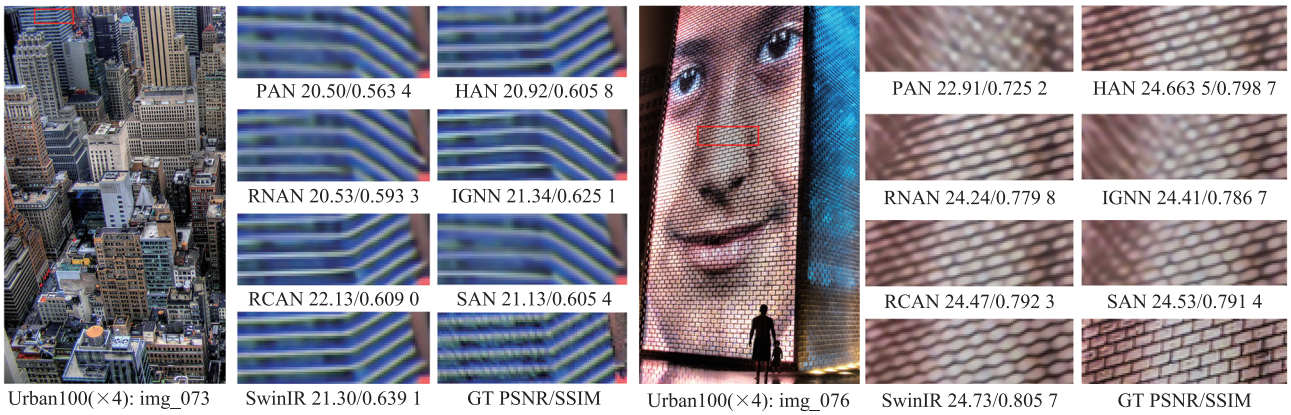


图 7　Urban100 测试集中 img_073 和 img_076 在×4 超分的可视化对比

Fig. 7　Visual comparison for ×4 SR on img_073 and img_076 in Urban100 datasets

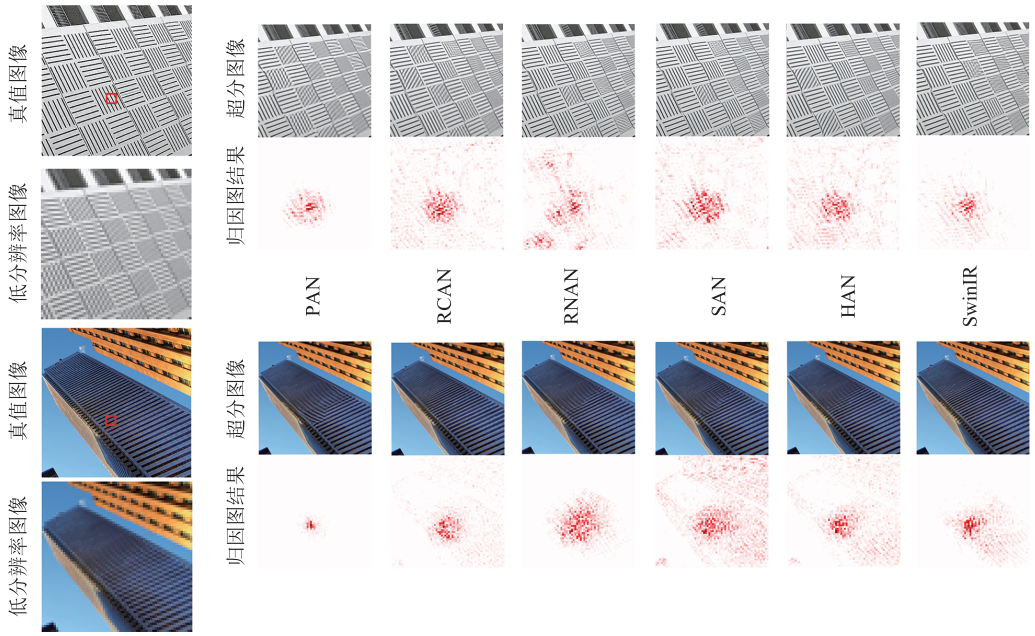


图 8　超分结果和不同网络局部归因图

Fig. 8　Comparison of the SR results and LAM attribution results of different SR networks

部的算法利用全局信息，PAN 将“注意力”仅放在局部切片附近，未能成功复原竖直纹理，且边缘模糊。而 RCAN、SAN 则“注意”到切片以外的相似结构，尝试使用这些额外信息恢复该局部切片，尽管受到干扰，出现一些异常斜纹，但边缘细节相对 PAN 的恢复效果更加清晰。对该切片成功复原的网络有 HAN 和 SwinIR。在局部归因图中观察到，相对 RCAN 和 SAN，HAN 和 SwinIR 网络利用了该切片左上方和左下方的竖直结构，恢复效果边缘锐利，纹理正确。但对于该切片左上方和左下方的相似竖直纹理，HAN 复原失败，产生了斜向纹理；SAN 成功复原，但有严重斜向伪影；SwinIR 复原效果最佳，边缘清晰，几乎无斜向伪影，但目标切片右上方的纹理复原没有 SAN 和 HAN 的效果清晰。

5 注意力机制和超分领域发展和应用的关键瓶颈及未来方向

5.1 注意力机制在超分任务上的结构设计

基于 CNN 的超分网络结构已被充分研究和探索，主要骨干网络均使用类似 EDSR 的单分支残差网络结构，并结合长连接和短连接。该类结构一方面增强了深度网络的可训练性，一定程度上缓解了深度超分网络一直存在的梯度消失和梯度爆炸问题。另一方面，低分辨率输入和特征中有许多冗余的低频信息可以通过这些连接传递，使超分网络能更专注于恢复损失的高频信息。

在 CNN 网络基础上，RCAN 通过引入全局池化层，扩大网络的感受野，提升网络的信息获取范围。PAN 使用基于 2D 卷积的混合注意力机制，在参数量极小的情况下获得可观的超分效果。HAN 在 RCAN 基础上，在网络末尾使用基于 3D 卷积的混合注意力机制，配合基于二阶自注意力机制的层注意力模块，进一步提升网络的表征能力。SAN 在 RCAN 基础上，使用全局协方差池化层代替全局平均池化层，以及二阶统计量方法增强网络的表征能力。

但基于 CNN 结构的网络仍存在两个源于卷积层的基本问题导致性能出现瓶颈：第一，图像和卷积核之间的交互与内容无关，使用相同的卷积核来恢复不同的图像区域不是最合理的选择；第二，在局部处理原理下，卷积对长程依赖建模效果不佳。

为了解决上述问题，研究人员开始尝试将自然语言处理中表现突出的自注意力机制引入超分方法。早期方法如 NLRN 和 RNAN 仅简单地对中间特征图使用自注意力操作，网络的性能提升有限，同时牺牲了网络的计算量和运算速度。IGNN 则尝试通过图神经网络，挖掘图像跨尺度特征融合的方法，获得较为可观的性能，但存在图像失真、网络并行化程度低和运算速度慢的问题。

尽管基于 Transformer 的方法依靠极富竞争力的建模能力在多个高层视觉任务中取得不俗的表现，但在超分领域基于 Transformer 的方法仍然不多。直到 SwinIR 在超分任务上的优秀表现刷新了研究人员对于 Transformer 在超分任务上的认知。通过局部切片和移窗机制，SwinIR 解决了自注意力机制网络计算量爆炸和切片边缘模糊的问题，充分利用窗口内部的局部信息和移窗时的长程依赖，同时保证了网络的并行性，使超分网络性能达到新高度。

但 SwinIR 显然不是超分网络的最终形态，基于高阶注意力机制的网络结构仍有待进一步发掘，尤其基于 Transformer 的网络展现出非常大的潜力。目前，基于 Transformer 的网络结构设计由于计算量庞大，自注意力机制的应用仍然停留在固定大小的特征切片上。在 4.6 节可视化部分也可以观察到，对于部分处于临界恢复效果的图像，SwinIR 相比于传统 CNN 方法没有表现出明显的优势，也容易受到重叠纹理的干扰。

基于 Swin Transformer 的结构本质是对全局自注意力机制的一种稀疏化表示，不可避免在某种程度上牺牲模型对全局信息的表征能力。然而，仅通过扩大窗口的尺寸来扩大感受野的方式虽然能在一定程度上改善该问题并使得模型能力得到一定提升，但由于该结构的计算量与窗口大小的平方成正比，会大量增加计算成本并与基于窗口的稀疏化自注意力机制的出发点相悖。对此，可能的改进方式包括：一是在基于窗口的自注意力结构下增强模型利用全局信息的能力；二是改进自注意力机制的计算方式，将复杂度 $O(n^2)$的注意力权重计算降低至 $O(n\log n)$甚至 $O(n)$。

5.2　真实图像退化方式不可控

目前，几乎所有基于注意力机制的超分网络在生成数据时，都默认使用理想的双三次下采样算法得到低分辨率图像，和实际应用存在较大差异。实际应用的退化方式复杂，存在成像设备不同、图像处理算法不同、压缩方式不同等引起的不同退化问题。在不同退化处理过程损失的信息也各有差异，退化方式不匹配的问题使基于注意力机制的超分网络在实际应用中效果较差，产生严重的伪影问题。如果将特定退化对应的超分模型应用于任意低分辨率输入，超分输出与目标高分辨率图像之间将存在极大的域间隙，从而导致质量较差的结果。

为了对未知退化类型的低分辨率图像进行超分增强，学界提出了另一种盲超分(Blind Super-Resolution)方法，包括：具有迭代内核校正的盲超分辨率方法(Iterative Kernel Correction，IKC)[56]、深度交变网络(Deep Alternating Network，AN)[57]、变体盲超分辨率(Variant Blind Super-Resolution，VBSR)[58]、核建模超分辨率网络(Kernel Modeling Super-Resolution Network，KMSR)[59]、真实增强生成对抗超分网络(Real-World Enhanced Generative Adversarial Network for Image Super-Resolution，Real-ESRGAN)[60]等。盲超分通过基于方程扩展的显式建模和基于外部数据集内固有分布的隐式建模方法，尝试缩小自然图像域和输出图像域之间的差距。显式建模方法将模糊、下采样、噪声和 JPEG 压缩经典退化模型组合生成真实退化模型。

然而，现实世界的退化太复杂，无法通过多个退化模型的简单组合进行建模。因此，以上方法在现实世界的样本中容易失败。隐式建模试图绕过显式建模的步骤，利用数据分布学习和生成对抗网络(GAN)获得退化模型。隐式建模通过数据分布隐式定义退化过程，且现有隐式建模的方法均需要外部数据集进行训练。然而，它们仅限于训练数据集的退化，不能很好地推广到分布外图像。

现有方法通常声称专注于现实世界的设置，实际上假设了某个场景，如某些数码相机拍摄的图像。事实上，真实世界的图像在其潜在的退化类型上大有不同，为特定退化类型设计的超分模型容易在另一种退化类型上表现较差，而造成不同退化的主要因素有 3 个，包括获取图像的设备、图像处理算法和存储导致的图像退化。

上述讨论的现实世界的图像都有自己的退化和挑战。以往的工作通常专注于单一类型的真实图像，如智能手机拍摄的图像，这极大限制了它们在不同场景的表现。未来，期望看到对不同类型的真实世界图像的更多探索以及更加综合可靠的真实退化数据集。研究出针对每种不同类型均有效的解决方案，应该是超分研究的最终目标。

本文涵盖的大部分方法，尤其是具有显式退化建模和外部数据集的方法，需要“低分辨率-高分辨率”图像对来优化和评估超分模型。然而，由于难以获得真实的配对数据，到目前为止只有少数真实世界的数据集，大多数方法仍然从高分辨率图像合成低分辨率输入。使用精心设计的技术和先进的数字设备构建的真实图像数

据集屈指可数，包括 City100[61]、DRealSR[62]和 RealSR[63]。其中，DRealSR 数据量最大，每个超分倍数有 800 个图像对，并通过调整成像设备的焦距来捕获高分辨率图像及其相应的低分辨率观测值，然后将图像对精确对齐并校正颜色。与合成数据相比，这些真实世界的数据集是在真实环境中研究盲超分的重要基准。然而，构建真实世界的数据集既耗时又昂贵，并且由于不同成像系统之间的复杂差异，也无法涵盖所有场景。希望未来会出现规模更大、场景更加复杂、退化方式更接近真实场景的数据集，为超分发展提供坚实的数据支持。

5.3 大规模预训练模型

回顾注意力机制从一阶向高阶的发展历程，不难看出，注意力机制仍处于快速发展阶段，不断有新网络在现有基础上改进，在测试集上显示出更优秀的超分效果。尤其是 SwinIR 的出现，使自注意力机制摆脱了不适用于超分领域的质疑，不仅展现了最先进的超分性能，还减少了模型的计算量，显存利用更加高效。

尽管近年来超分模型发展迅速，但当前先进的超分模型过于专注单个任务，如双三次下采样，结果为多个任务或环境单独开发了数千个模型，消耗了大量的计算资源，却无法处理真实场景下的复杂退化类型。行业中成千上万的长尾任务，即各种退化方式，是人工智能研究和应用面临的一个重大障碍。通用人工智能方法将“通用智能”作为一个不同的属性，理应关注人工智能模型的通用性、适应性和灵活性。

视觉和语言是通用人工智能不可或缺的两种模式。在语言方面，通用语言模型(General Vision Model，GLM)取得了令人瞩目的进展。BERT[64]和 GPT-3[65]等大规模预训练语言模型已显示出开发 GLM 的潜力，这些 GLM 通过使用情景学习和即时学习，不需要进行反向传播，在控制大模型训练成本的同时，有益于广泛的语言相关下游任务。此外，随着与任务无关的训练目标的出现，可以通过扩展网络爬行数据和模型容量以及计算预算来稳步提高预训练的性能增益。

GLM 的成功激发了大规模超分预训练模型学习的新方向。从事大规模监督、自监督和跨模态预训练的先驱在有限范围的下游视觉任务上表现出一定的普遍性。然而，设计可靠的大规模超分预训练方法仍具挑战性。大多数已有的工作主要利用一个监督信号源，在单独的监督下进行单调预训练生成在特定场景中表现良好的模型，但如果目标是获得可推广到大量下游任务(甚至是目前不可知的任务)的“真实”大规模超分预训练模型，单一监督则无法提供足够的表征能力。如何实现不同退化类型超分视觉任务的通用网络，集成各种监督信号的、高度可扩展的上游预训练模型，以及针对多样化任务设计灵活的下游网络，将会是超分大规模预训练模型的关键突破点。

5.4 图像超分的评估机制

超分方法通常通过图像质量评估指标(Image Quality Assessment，IQA)，如 PSNR、SSIM，测量重建图像和真实图像之间的相似性来评估图像恢复效果。随着真实场景的退化类型引入超分领域，如何评估复杂退化图像的超分效果成为研究人员面临的新问题。一些非参考图像质量评估方法，如 Ma[66]和感知指数(Perceptual Index，PI)[67]，被引入评估感知驱动超分方法。在某种程度上，这些图像质量评估方法是超分领域取得长足进步的主要原因之一。然而，虽然新算法在不断刷新指标数值，但定量结果和感知质量之间不一致的情况却越发明显，甚至出现指标失灵的情况。Blau 等[68]认为 PI 与人类感知更相关，但具有高 PI 数值的算法，如 RankSRGAN[69]仍会在恢复图像中生成明显不真实的伪影。

在超分领域中，网络设计远远超越了图像质量评估指标的发展，现有的图像评估机制存在的问题随之暴露，这迫使研究者必须重新思考超分

任务的有效评估方法。

首先，现有的量化指标和人类的感知效果还有显著的差距。为了更加精准地评估网络性能，为网络发展提供有效的指导，自然图像质量评价(Natural Image Quality Evaluator，NIQE)[70]、成对偏好的感知图像错误评估(Perceptual Image-Error Assessment through Pairwise Preference，PieAPP)[71]、学习感知图像切片相似度(Learned Perceptual Image Patch Similarity，LPIPS)[72]和用于全参考图像质量评估的加权平均深度图像质量度量(Weighted Average Deep Image QuAlity Measure for Full-Reference Image Quality Assessment，WaDIQaM)[73]等指标被相继提出。Gu 等[74]在现有的图像质量评估指标的基础上，进一步提出了一个大规模数据集——感知图像处理算法数据集(Perceptual Image Processing Algorithms，PIPAL)，为图像质量评估指标的改进提供新的基准。

其次，随着生成对抗网络在超分任务中的广泛应用，尽管基于生成对抗网络的方法输出的图像由于网络生成虚假的纹理导致 PSNR 和 SSIM 指标较低，但输出图像的视觉效果却远优于传统 CNN 方法。研究人员对基于生成对抗网络的方法普遍采用 NIQE 和 LPIPS 来评估模型输出，但基于深度学习方法的指标存在依赖人工超参和内容敏感导致的指标不稳定问题，还有待进一步研究。

再者，现有的图像评估指标还没有针对模型处理过程以及模型本身进行评估的机制。随着超分方法处理的低分辨率图像由简单的双三次下采样向复杂下采样方式转变，针对不同图像退化方式、网络的泛化能力和模型处理过程的评估机制，是一个非常有指导意义的方向。

6　讨论与分析

现有的研究主要依据网络结构[75]和特征图利用的维度[76]对超分网络进行分类，且停留在网络性能对比，没有对网络关键模块和重要构成机制进行系统的对比分析。而本文对现有的基于注意力机制的深度学习超分辨率方法，依据注意力机制的数学统计原理，对网络进行了系统的分类；对关键模块性能进行了全面的定量定性分析。通过广泛的定量和定性比较，注意到现有方法的以下趋势：

(1)网络模型利用的数学统计量由一阶向高阶转变。

(2)表现最好的方法开始发掘卷积神经网络以外的网络结构。

(3)真实退化的低分辨率图像正逐渐代替简单的双三次下采样。

总的来说，近年来超分辨率性能得到了极大的提升，但仍存在一些亟待解决的关键问题。本文总结归纳了这些问题，并提出一些潜在的研究方向。值得注意的是，现实世界场景对先进的超分方法的限制正在逐步解除，最先进的方法在复杂退化的图像超分上表现出越来越强的性能。尽管本文分析对比的模型数量较小和量化指标相对单一，但希望能够为研究人员进一步发展超分提供帮助。

参 考 文 献

[1] Dong C, Loy CC, He KM, et al. Learning a deep convolutional network for image super-resolution [C] // Proceedings of the European Conference on Computer Vision, 2014: 184-199.

[2] Dai T, Cai JY, Zhang YB, 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.

[3] Niu B, Wen WL, Ren WQ, et al. Single image super-resolution via a holistic attention network [C] // Proceedings of the European Conference on

Computer Vision, 2020: 191-207.

[4] Zhang YL, Li KP, Li K, et al. Image super-resolution using very deep residual channel attention networks [C] // Proceedings of the European Conference on Computer Vision, 2018: 286-301.

[5] Liang JY, Cao JZ, Sun GL, et al. SwinIR: image restoration using Swin Transformer [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 1833-1844.

[6] Liu D, Wen BH, Fan YC, et al. Non-local recurrent network for image restoration [Z/OL]. arXiv Preprint, arXiv:1806.02919, 2018. https://arxiv.org/abs/1806.02919.

[7] 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.

[8] Huang SY, Sun J, Yang Y, et al. Robust single-image super-resolution based on adaptive edge-preserving smoothing regularization [J]. IEEE Transactions on Image Processing, 2018, 27(6): 2650-2663.

[9] Yang JC, Wright J, Huang T, et al. Image super-resolution as sparse representation of raw image patches [C] // Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, 2008: 1-8.

[10] Tai Y, Yang J, Liu XM, et al. MemNet: a persistent memory network for image restoration [C] // Proceedings of the IEEE International Conference on Computer Vision, 2017: 4539-4547.

[11] Wang ZW, Liu D, Yang JC, et al. Deep networks for image super-resolution with sparse prior [C] // Proceedings of the IEEE International Conference on Computer Vision, 2015: 370-378.

[12] Kim J, Lee JK, Lee KM. Deeply-recursive convolutional network for image super-resolution [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 1637-1645.

[13] Lai WS, Huang JB, Ahuja N, et al. Fast and accurate image super-resolution with deep Laplacian pyramid networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 41(11): 2599-2613.

[14] He KM, Zhang XY, Ren SQ, et al. Deep residual learning for image recognition [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.

[15] Choi JS, Kim M. A deep convolutional neural network with selection units for super-resolution [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017: 154-160.

[16] Sajjadi MSM, Scholkopf B, Hirsch M. EnhanceNet: single image super-resolution through automated texture synthesis [C] // Proceedings of the IEEE International Conference on Computer Vision, 2017: 4491-4500.

[17] 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.

[18] Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets [J]. Advances in Neural Information Processing Systems, 2014, 27: 2672-2680.

[19] Johnson J, Alahi A, Li FF. Perceptual losses for real-time style transfer and super-resolution [C] // Proceedings of the European Conference on Computer Vision, 2016: 694-711.

[20] Zhang YL, Tian YP, 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.

[21] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [C] // Proceedings of the Advances in Neural Information Processing Systems, 2017: 5998-6008.

[22] 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.

[23] Lu J, Xiong CM, Parikh D, et al. Knowing when to look: adaptive attention via a visual sentinel for image captioning [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 375-383.

[24] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate [Z/OL]. arXiv Preprint, arXiv:1409.0473, 2014. https://arxiv.org/abs/1409.0473.

[25] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.

[26] Zhang ZF, Wang ZW, Lin Z, et al. Image super-resolution by neural texture transfer [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 7982-7991.

[27] Hu YT, Li J, Huang YF, et al. Channel-wise and spatial feature modulation network for single image super-resolution [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 30(11): 3911-3927.

[28] Nair V, Hinton GE. Rectified linear units improve restricted Boltzmann machines [C] // Proceedings of the 27th International Conference on Machine Learning, 2010: 807-814.

[29] Zhao HY, Kong XT, He JW, et al. Efficient image super-resolution using pixel attention [C] // Proceedings of the European Conference on Computer Vision, 2020: 56-72.

[30] Lin TY, RoyChowdhury A, Maji S. Bilinear CNN models for fine-grained visual recognition [C] // Proceedings of the IEEE International Conference on Computer Vision, 2015: 1449-1457.

[31] Li PH, Xie JT, Wang QL, et al. Is second-order information helpful for large-scale visual recognition [C] // Proceedings of the IEEE International Conference on Computer Vision, 2017: 2070-2078.

[32] Zhang YL, Li KP, Li K, et al. Residual non-local attention networks for image restoration [Z/OL]. arXiv Preprint, arXiv:1903.10082, 2019. https://arxiv.org/abs/1903.10082.

[33] Zhou S, Zhang J, Zuo W, et al. Cross-scale internal graph neural network for image super-resolution [J]. Advances in Neural Information Processing Systems, 2020, 33: 3499-3509.

[34] Plötz T, Roth S. Neural nearest neighbors networks [C] // Proceedings of the 32nd International Conference on Neural Information Processing Systems, 2018: 1095-1106.

[35] Zontak M, Irani M. Internal statistics of a single natural image [C] // Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, 2011: 977-984.

[36] Simonovsky M, Komodakis N. Dynamic edge-conditioned filters in convolutional neural networks on graphs [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 3693-3702.

[37] Lefkimmiatis S. Universal denoising networks: a novel CNN architecture for image denoising [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 3204-3213.

[38] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [Z/OL]. arXiv Preprint, arXiv:1409.1556, 2014. https://arxiv.org/abs/1409.1556.

[39] Dabov K, Foi A, Katkovnik V, et al. Image denoising by sparse 3-D transform-domain collaborative filtering [J]. IEEE Transactions on Image Processing, 2007, 16(8): 2080-2095.

[40] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with Transformers [C] // Proceedings of the European Conference on Computer Vision, 2020: 213-229.

[41] Chen HT, Wang YH, Guo TY, et al. Pre-trained image processing transformer [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 12299-12310.

[42] 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.

[43] Kingma DP, Ba J. Adam: a method for stochastic optimization [Z/OL]. arXiv Preprint, arXiv:1412.6980, 2014. https://arxiv.org/abs/1412.6980.

[44] Paszke A, Gross S, Massa F, et al. Pytorch: an imperative style, high-performance deep learning library [J]. Advances in Neural Information Processing Systems, 2019, 32: 8026-8037.

[45] Bevilacqua M, Roumy A, Guillemot C, et al. Low-complexity single-image super-resolution based on nonnegative neighbor embedding [C] // Proceedings of the British Machine Vision Conference, 2012: 135.1-135.10.

[46] Zeyde R, Elad M, Protter M. On single image scale-up using sparse-representations [C] // Proceedings of the 7th International Conference on Curves and Surfaces, 2010: 711-730.

[47] 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, 2001: 416-423.

[48] Huang JB, 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.

[49] Matsui Y, Ito K, Aramaki Y, et al. Sketch-based manga retrieval using Manga109 dataset [J]. Multimedia Tools and Applications, 2017, 76(20): 21811-21838.

[50] Wang Z, Bovik AC, Sheikh HR, et al. Image quality assessment: from error visibility to structural similarity [J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.

[51] Haris M, Shakhnarovich G, Ukita N. Deep back-projection networks for super-resolution [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 1664-1673.

[52] Shi WZ, Caballero J, Huszár F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 1874-1883.

[53] Dong C, Loy CC, Tang XO. Accelerating the super-resolution convolutional neural network [C] // Proceedings of the European Conference on Computer Vision, 2016: 391-407.

[54] Zhang K, Danelljan M, Li YW, et al. AIM 2020 challenge on efficient super-resolution: methods and results [C] // Proceedings of the European Conference on Computer Vision, 2020: 5-40.

[55] Gu JJ, 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.

[56] Gu JJ, Lu HN, Zuo WM, et al. Blind super-resolution with iterative kernel correction [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 1604-1613.

[57] Huang Y, Li S, Wang L, et al. Unfolding the alternating optimization for blind super resolution [J]. Advances in Neural Information Processing Systems, 2020, 33: 5632-5643.

[58] Cornillere V, Djelouah A, Wang YF, et al. Blind image super-resolution with spatially variant degradations [J]. ACM Transactions on Graphics (TOG), 2019, 38(6): 1-13.

[59] Zhou RF, Susstrunk S. Kernel modeling super-resolution on real low-resolution images [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 2433-2443.

[60] Wang XT, Xie LB, Dong C, et al. Real-ESRGAN: training real-world blind super-resolution with pure synthetic data [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 1905-1914.

[61] Chen C, Xiong ZW, Tian XM, et al. Camera lens super-resolution [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 1652-1660.

[62] Wei PX, Xie ZW, Lu HN, et al. Component divide-and-conquer for real-world image super-resolution

[C] // Proceedings of the European Conference on Computer Vision, 2020: 101-117.

[63] Cai JR, Zeng H, Yong HW, et al. Toward real-world single image super-resolution: a new benchmark and a new model [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 3086-3095.

[64] Devlin J, Chang MW, Lee K, et al. Bert: pretraining of deep bidirectional Transformers for language understanding [Z/OL]. arXiv Preprint, arXiv:1810.04805, 2018. https://arxiv.org/abs/1810.04805.

[65] Brown T, Mann B, Ryder N, et al. Language models are few-shot learners [J]. Advances in Neural Information Processing Systems, 2020, 33: 1877-1901.

[66] Ma C, Yang CY, 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.

[67] Blau Y, Michaeli T. The perception-distortion tradeoff [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 6228-6237.

[68] 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) Workshops, 2018: 1-22.

[69] Zhang WL, Liu YH, Dong C, et al. RankSRGAN: generative adversarial networks with ranker for image super-resolution [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 3096-3105.

[70] Mittal A, Soundararajan R, Bovik AC. Making a “completely blind” image quality analyzer [J]. IEEE Signal Processing Letters, 2012, 20(3): 209-212.

[71] Prashnani E, Cai H, Mostofi Y, et al. PieAPP: perceptual image-error assessment through pairwise preference [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 1808-1817.

[72] Zhang R, Isola P, Efros AA, et al. The unreasonable effectiveness of deep features as a perceptual metric [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 586-595.

[73] Bosse S, Maniry D, Müller KR, et al. Deep neural networks for no-reference and full-reference image quality assessment [J]. IEEE Transactions on Image Processing, 2017, 27(1): 206-219.

[74] Gu JJ, Cai HM, Chen HY, et al. Image quality assessment for perceptual image restoration: a new dataset, benchmark and metric [Z/OL]. arXiv Preprint, arXiv:2011.15002, 2020. https://arxiv.org/abs/2011.15002.

[75] Anwar S, Khan S, Barnes N. A deep journey into super-resolution: a survey [J]. ACM Computing Surveys (CSUR), 2020, 53(3): 1-34.

[76] Zhu HY, Xie C, Fei YQ, et al. Attention mechanisms in CNN-based single image super-resolution: a brief review and a new perspective [J]. Electronics, 2021, 10(10): 1187.