

聚合残差注意力网络的单图像超分辨率重建

彭晏飞, 张曼婷*, 张平甲, 李健, 顾丽睿

辽宁工程技术大学电子与信息工程学院, 辽宁 葫芦岛 125105

摘要 针对现有生成式对抗网络对单图像进行超分辨率重建时存在特征信息挖掘不足、算法复杂度高及训练不稳定的问题,提出一种聚合残差注意力网络的单图像超分辨率重建方法。首先,以聚合残差模块作为基本残差块构造生成器,降低计算复杂度,在每个残差块中引入具有三维权重的注意力模块作为网络主通道,在不引入其他参数情况下捕获更多的高频信息。其次,在鉴别器中采用谱归一化处理,对鉴别器网络参数进行限制,从而稳定训练过程。最后,采用拟合性更好的 Swish 激活函数,提高网络的特征提取能力。将鲁棒性更好的 Charbonnier 损失函数作为像素损失,同时加入正则化损失抑制图像噪点,提升图像的空间平滑性。实验结果表明,所提方法得到的四倍放大的超分辨率重建图像在 Set5、Set14、BSD100 三个公开数据集上的峰值信噪比平均值提高了 1.54 dB,结构相似性平均值提高了 0.0457,重建图像拥有更好的清晰度和更为丰富的高频细节。

关键词 超分辨率; 生成对抗网络; 残差网络; 注意力机制; 谱归一化

中图分类号 TP391 文献标志码 A

DOI: 10.3788/LOP220752

Single-Image Super-Resolution Reconstruction Aggregating Residual Attention Network

Peng Yanfei, Zhang Manting*, Zhang Pingjia, Li Jian, Gu Lirui

School of Electronic and Information Engineering, Liaoning Technical University,
Huludao 125105, Liaoning, China

Abstract A single-image super-resolution reconstruction method based on aggregated residual attention network is proposed to solve the problems for insufficient feature information mining, high algorithm complexity, and unstable training in the super-resolution reconstruction for a single image in existing generative countermeasure networks. First, the aggregated residual module is used as the basic residual block to construct a generator, to reduce computational complexity. In each residual block, an attention module with a three-dimensional weight is introduced as the main channel to capture additional high-frequency information without other parameters. Second, the discriminator network parameters are limited via spectral normalization to stabilize the training process. Finally, the Swish activation function with improved fitting is used to improve the feature extraction ability of the network. The Charbonnier loss function with enhanced robustness is used as the pixel loss, and the regularization loss is added to suppress image noise to improve spatial smoothness. The experimental results show that the average value of the peak signal-to-noise ratio and structural similarity of images reconstructed using the proposed method on Set5, Set14, and BSD100 public datasets increase by 1.54 dB and 0.0457, respectively. Therefore, the reconstructed images have a better resolution and richer high-frequency detail than the original image.

Key words super resolution; generative adversarial network; residual network; attention mechanism; spectral normalization

1 引言

图像超分辨率重建旨在从低分辨率(LR)图像中

重建对应的高分辨率(HR)图像。目前为止,图像超分辨率(SR)重建技术已经取得了非凡的成绩,在效果上越来越清晰和真实,并广泛应用于许多现实领域,如监

收稿日期: 2022-02-16; 修回日期: 2022-03-15; 录用日期: 2022-04-06; 网络首发日期: 2022-04-16

基金项目: 国家自然科学基金(61772249)、辽宁省高等学校基本科研项目(LJKZ0358)、辽宁工程技术大学双一流学科创新团队资助项目(LNTU20TD-27)

通信作者: *1016422506@qq.com

控设备^[1]、遥感图像^[2]、医学影像^[3]、雷达成像^[4]、显微成像^[5]等。超分辨率重建方法主要分为基于插值的方法^[6]、基于重建的方法^[7]、基于学习的方法^[8]。得益于深度学习的发展,基于学习的方法从训练样本集中学习 LR 图像与 HR 图像之间的关系,通过学习端到端的映射推测对应 SR 图像,重建效果超越前两种方法和传统浅层学习方法。

Dong 等^[9]提出的 super-resolution convolutional neural network(SRCNN)是将深度卷积神经网络引入到单图像超分辨率(SISR)重建领域的开山之作,使用双三次插值将输入图像放大,通过训练三层卷积网络实现图像超分辨率重建,性能显著提升,但将插值 LR 作为模型训练的输入也增加了计算开销。为解决此问题,Shi 等^[10]提出 efficient sub-pixel convolutional neural network(ESPCN),将对 HR 图像进行双三次下采样得到的 LR 图像作为输入,把子像素卷积层放在最后来实现重建操作,提高重建效率的同时降低了内存。Kim 等^[11]借鉴深度学习中的残差学习思想^[12],将网络结构加深至 20 层,获得超分辨率网络(VDSR),利用全局残差思想提高网络收敛速度。在 VDSR 的基础上, Kim 等^[13]受递归学习的启发,设计 deeply-recursive convolutional network(DRCN),在多层网络上共享参数,减少参数量,重建效果优于 VDSR。Zhang 等^[14]将残差模块和密集模块连接,构成 residual dense network(RDN)。紧接着 Zhang 等^[15]提出 residual in residual(RIR)结构来构建更深层次的模型,并在每个残差块中引入通道注意力机制,形成残差通道注意力网络(RCAN)。这些算法均以增加更深层次的卷积层来提高峰值信噪比(PSNR)和结构相似性(SSIM),但模型仍存在参数量大、内存占用大、测试和训练速度慢等问题。最近, Ayazoglu^[16]提出一种超轻量的超分辨率网络(XLSR),该网络以 aggregated residual transformations networks(ResNeXt)结构^[17]为基本模块,去除跳跃连接,提高运算速度。在网络输出处添加 Clipped ReLU 模块,实现重构质量和运行时间的平衡,该模型网络参数量仅为 VDSR 的 1/30。

根据 Goodfellow 等^[18]提出的生成对抗网络(GAN)在图像生成工作中的优势, Ledig 等^[19]将 GAN 应用到 SR 领域,提出超分辨率生成对抗网络(SRGAN),用感知损失使生成图像与真实图像更接近,但对纹理细节的重构中存在模糊和伪影问题,同时 GAN 的引入具有训练困难、模型难于收敛等问题。Wang 等^[20]提出增强型超分辨率生成对抗网络(ESRGAN),将残差密集块作为基本网络构建单元并去掉批量归一化(BN)层,该网络使得重建图像更加锐化,质量更好。辛元雪等^[21]对 ESRGAN 进行改进,使用多尺度密集连接模块并且在其后添加通道注意力机制,以调整不同通道的特征响应值,从而使得模型中的深层信息得以更好提取。

针对现有方法存在特征信息挖掘不足、算法复杂度高及训练不稳定的问题,本文改进了 SRGAN,提出一种聚合残差注意力网络的单图像超分辨率方法。以聚合残差网络为残差块构成生成器,采用与其相同的分组拓扑结构对图像进行特征提取,在每个分组拓扑结构中引入具有三维权重的无参注意力模块(SimAM)^[22]提高网络性能,在不需要额外超参数调整的前提下,加强了高频信息在网络中的作用,增强网络的表征能力。在鉴别器中引入谱归一化处理,稳定了训练过程,提高了模型的泛化能力。像素损失函数使用抗噪性更好的 Charbonnier 损失函数,同时加入正则化损失来稳定 GAN 的训练。所提模型中的激活函数均使用 Swish 激活函数。通过以上方法,可均衡提升图像感知效果与客观评价值,并且减少了大量训练时间。

2 所提方法内容

2.1 残差块设计

受 ResNeXt 启发,将分组卷积思想加入到 SRGAN 中,简化训练的同时提高网络性能。ResNeXt 是一种简单的高度模块化的残差结构, Xie 等通过大量实验得出在网络准确度相似的情况下, ResNeXt 比 ResNet 深度更浅显、计算复杂度更低的结论。ResNeXt 将相同的并行分支的路径数提取为分组数(cardinality),减少超参数的使用,防止模型过拟合的风险,提高模型泛化能力。

本文将 SRGAN 中的 ResNet 更改为 ResNeXt,如图 1 所示。图 1(a)为原始 SRGAN 的 ResNet 结构,网络参数量约为 74×10^3 ,本文采用一种平行堆叠相同拓扑结构的三层卷积层代替原始典型残差块的卷积层。通过实验表明,当分组数 cardinality 为 8 时,模型性能达到最高,本文残差块模型的参数量约为 35×10^3 ,比 SRGAN 的参数量减少 53%,降低了计算复杂度。残差块有两种等价的设计:图 1(b)采用 8 个相同的三层卷积分支,在残差块的末端聚合相加输出;图 1(c)在残差块中间采用组卷积操作,利用小卷积核结合多个感受野提取特征。所提结构如图 1(c)所示,相比而言,结构更为清晰和简洁。所提模型将 ResNeXt 的分组卷积思想应用于超分辨率领域,相对于经典残差网络, ResNeXt 参数少、深度大的特点可以降低网络模型复杂度,提高模型精度,节省模型训练时间,增强网络特征表达能力。

2.2 注意力机制

注意力机制起源于人眼视觉系统对于观察到的物体会选择视觉区域中的特定部分,然后集中关注部分有用信息而忽略其他无用或不重要的信息这一生物特征。现有的注意力机制大多分别将空间注意力与通道注意力并行或串行组合。在 squeeze-and-excitation network(SENNet)^[23]中通道注意力 SE 模块通过全局平

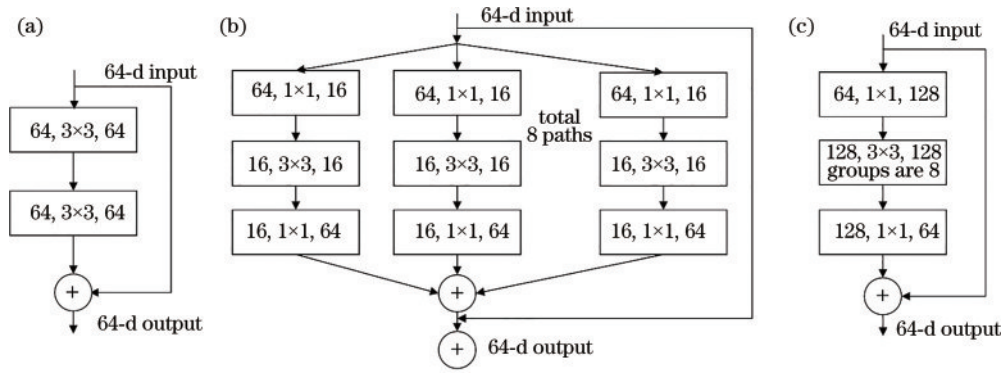


图1 ResNet和ResNeXt的对比。(a)ResNet结构;(b)ResNeXt结构;(c)所提残差结构

Fig. 1 Comparison of ResNet and ResNeXt. (a) ResNet structure; (b) ResNeXt structure; (c) proposed residual structure

均池化和全连接层学习特征映射通道之间的关系,自适应地分配每个特征通道的权重,提高网络表征能力。然而,人脑的两种注意力形式往往是协同工作促进信息选择的,利用空间维度的二维权重不能提取足够的图像信息。最近 SimAM 在不增加额外参数的情况下访问特征图的三维注意力权重,在改善卷积神经网络的表征能力上具有灵活性和有效性,更适合于超分辨率重建任务。

在图像超分辨率重建过程中,低频信息与高频信息

区域重建难度不同,而现有的深度网络通常对两种区域分配相同的权重,往往会弱化高频信息的重要性^[24],从而导致重建质量较低,视觉感官效果较差,捕捉图像中的高频信息显得尤为重要。为进一步增强 ResNeXt 网络结构的特征提取能力,将注意力机制添加至每个残差块中的第二个卷积层后,如图 2 所示,通过优化能量函数为特征图 X 推断三维注意力权重,以挖掘特征图中每个神经元的重要性,找到含有丰富高频信息的神经元并赋予其更高权重,以有效提高网络表达能力。

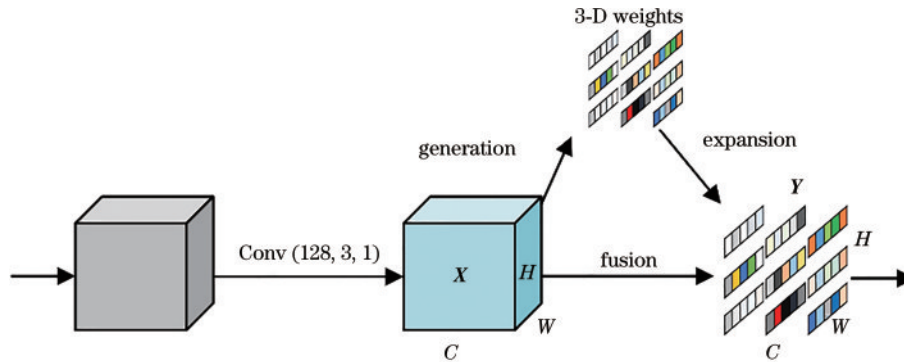


图2 注意力机制

Fig. 2 Attention mechanism

在神经科学中,信息丰富的神经元通常表现出与周围神经元不同的放电模式,并且激活神经元通常会抑制周围神经元,即空域抑制^[25]。具有空域抑制效应的神经元应当被赋予更高的重要性,从而提高网络性能。目标神经元 t 的最小能量 e_t 为

$$e_t^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda}, \quad (1)$$

式中: $\hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i$, $\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{\mu})^2$, $\lambda = 10^{-4}$; t 和 x_i 分别是输入特征 X 的单个通道中的目标神经元和其他神经元; $X \in \mathbb{R}^{C \times T \times V}$, 其中 C 表示通道数, T 表示时间维度, V 表示空间维度; $N = T \times V$ 指在某个通道上所有神经元的个数。基于空域抑制现象,神经元的重要性可以通过 $1/e_t^*$ 得到,能量越低,神经元 t 与周围神经元的区别越大,重要性越高。

最后增强重要特征,以便提升网络性能,采用 Sigmoid 函数控制向量的输出范围,抑制神经元中的过大异常值。对特征进行增强处理,公式为

$$Y = \text{Sigmoid}\left(\frac{1}{E}\right) \otimes X, \quad (2)$$

式中: E 表示 e_t^* 在所有空间和通道维度上的汇总; \otimes 表示各元素相乘。

2.3 谱归一化处理

超分辨率中,生成对抗网络的应用提高了重建图像的质量,然而也存在网络不收敛、梯度消失和训练不稳定等问题。Miyato 等^[26]提出谱归一化生成对抗网络(SNGAN),将谱归一化(SN)应用于鉴别器网络。本文借鉴 SNGAN,在鉴别器网络的所有卷积层中使用谱归一化处理,并去掉传统批量归一化层,通过约束 GAN 的鉴别器的每一层网络的权重矩阵的谱范数来

约束鉴别器的 Lipschitz 常数,可以防止训练过程中出现模式坍塌现象,确保训练后的模型对测试数据的扰动表现出轻微的敏感性,从而增强 GAN 在训练过程中的稳定性,为生成器提供更多梯度信息,得到令人满意的超分辨率重建图像。

2.4 损失函数构建

SRGAN 的损失函数包含内容损失和对抗损失两部分,使得生成网络可以生成高质量 SR 图像,鉴别网络可以更准确区分真样本和假样本。损失函数的表达式为

$$l^{\text{SR}} = l_{\text{content-loss}}^{\text{SR}} + 10^{-3} l_{\text{Gen}}^{\text{SR}} \quad (3)$$

本文将均方误差(MSE)损失函数替换为鲁棒性更好的 Charbonnier 损失函数来构建像素损失函数,并加入 total variation(TV)正则化损失来抑制生成图像的噪声。

2.4.1 内容损失

SRGAN 的内容损失函数也分为两部分:MSE 损失和 VGG 损失。由于 MSE 损失函数直接优化高分辨率与低分辨率各个像素之间的平方差,对异常点较为敏感,虽然收敛效果较好,但使用原始度量的方法,导致图像有振铃或伪影情况,人眼感官效果较差。所以本文采用 Charbonnier 损失来代替 MSE 损失保证重建图像低频部分的正确性。Charbonnier 损失函数对网络异常值具有较好的监督能力,可以有效抑制图像出现的振铃和伪影现象,表达式为

$$l_{\text{Charbonnier}}(\mathbf{y}, \hat{\mathbf{y}}) = E_{\mathbf{x}, \mathbf{y} \sim P_{\text{data}}(\mathbf{x}, \mathbf{y})} [\sigma(\mathbf{y} - \hat{\mathbf{y}})], \quad (4)$$

式中: \mathbf{y} 表示原始图像; $\hat{\mathbf{y}}$ 表示 SR 图像; $\sigma(x) = \sqrt{x^2 + \epsilon^2}$ 为 Charbonnier 损失的惩罚函数, ϵ 取值为 $\epsilon = 10^{-7}$ 。

由于像素损失可能会导致图像中高频信息丢失、图像细节不足和轮廓不清晰,所以采用基于 VGG 网络提取特征的损失,使用预训练好的 VGG16 网络提取特征的损失值。VGG 损失是重建图像的特征表示与原始图像之间的欧氏距离,表达式为

$$l_{\text{VGG}} = \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} \left\{ \phi_{i,j}(\mathbf{I}^{\text{HR}})_{x,y} - \phi_{i,j}[G_{\theta_c}(\mathbf{I}^{\text{LR}})]_{x,y} \right\}^2, \quad (5)$$

式中: $W_{i,j}$ 和 $H_{i,j}$ 表示在 VGG 网络中的特征图的维度; $\phi_{i,j}$ 表示 VGG 网络中第 i 层最大池化层后的第 j 个卷积层的特征图; $G_{\theta_c}(\mathbf{I}^{\text{LR}})$ 为重建图像。

2.4.2 对抗损失

SRGAN 将 GAN 的生成部分加到感知损失中。对于对抗损失函数,将其与 SRGAN 保持一致。

$$l_{\text{adv}} = \sum_{i=1}^N -\log_{10} D_{\theta_d}[G_{\theta_g}(\mathbf{I}^{\text{LR}})], \quad (6)$$

式中: $D_{\theta_d}[G_{\theta_g}(\mathbf{I}^{\text{LR}})]$ 表示鉴别器将生成器生成的图像重建为自然图像的概率。

2.4.3 TV 损失

在图像超分辨率重建过程中,随着放大因子的增大,图像中的噪声也会放大,产生非常不好的视觉体验。引入 TV 损失,保持图像光滑完整性的同时去除生成图像的噪声,防止图像过于像素化。图片中相邻像素之间的差值可以使用 TV 损失来解决,TV 损失的公式为

$$l_{\text{TV}} = \|\hat{\mathbf{y}}\|_{\text{TV}}. \quad (7)$$

2.4.4 总损失函数

综上所述,总的损失函数为

$$l_{\text{total}} = \alpha l_{\text{Charbonnier}} + \beta l_{\text{VGG}} + \chi l_{\text{adv}} + \theta l_{\text{TV}}, \quad (8)$$

式中: $\alpha=1, \beta=6 \times 10^{-3}, \chi=10^{-3}, \theta=2 \times 10^{-8}$ 。

2.5 激活函数

在超分辨率重建中,当大的梯度流经神经元时,某些神经元无法学习从而导致神经元“死亡”。所有激活函数使用 Swish,其具备稀疏、平滑和非单调的特性,可有效消除神经元“死亡”现象,加快训练速度,更好地学习 LR 到 HR 图像之间的映射关系指导图像重建,增强特征提取能力,促进模型收敛。激活函数的公式为

$$G(\mathbf{x}) = \mathbf{x} * \text{Sigmoid}(\mathbf{x}). \quad (9)$$

2.6 网络模型

在 SRGAN 模型基础上,生成器使用聚合残差块,构成残差组,减少网络参数,增强网络表征能力;并在残差组内添加注意力机制,充分提取特征图特征,使得特征图中的重要特征得以加强,提高重建质量;然后对原有生成对抗网络中的鉴别器进行谱归一化处理,稳定网络训练;最后在损失函数上进行改进,用鲁棒性更好、抗噪性更强的 Charbonnier 损失函数代替 L2 损失函数,加入 TV 正则项改善图像噪音和平滑性。改进的超分辨率模型的作用过程如图 3 所示。改进模型不仅拥有高的客观评价,视觉体验也更满意。

2.6.1 生成器

图 4 为生成器网络结构。在网络设计上,采用 5 个带有注意力机制的聚合残差块,每个残差块包含 3 个卷积层,卷积核大小分别为 $1 \times 1, 3 \times 3, 1 \times 1$,通道数分别为 128、128、64;在残差块中间采用组卷积,分组数 cardinality 为 8,并在组卷积层后引入 SimAM 注意力机制,以确保网络提取更精细的特征信息;在卷积层后加入 batch normalization(BN)层,为了防止网络出现梯度消失问题,在每个 BN 层后应用 Swish 激活函数,以确保神经元不会丢失。使用两个子像素卷积层对图像进行上采样,并在生成器网络的最后一层应用 Tanh 激活函数。

2.6.2 鉴别器

鉴别器网络结构如图 5 所示,其中 k 代表卷积核个数, n 代表通道数, s 代表步长。为了避免图像过度下采样,鉴别器具有类似 VGG 的结构,它采用 8 组卷积层,每一层有包含相似数量的信息,通过逐渐减小特征

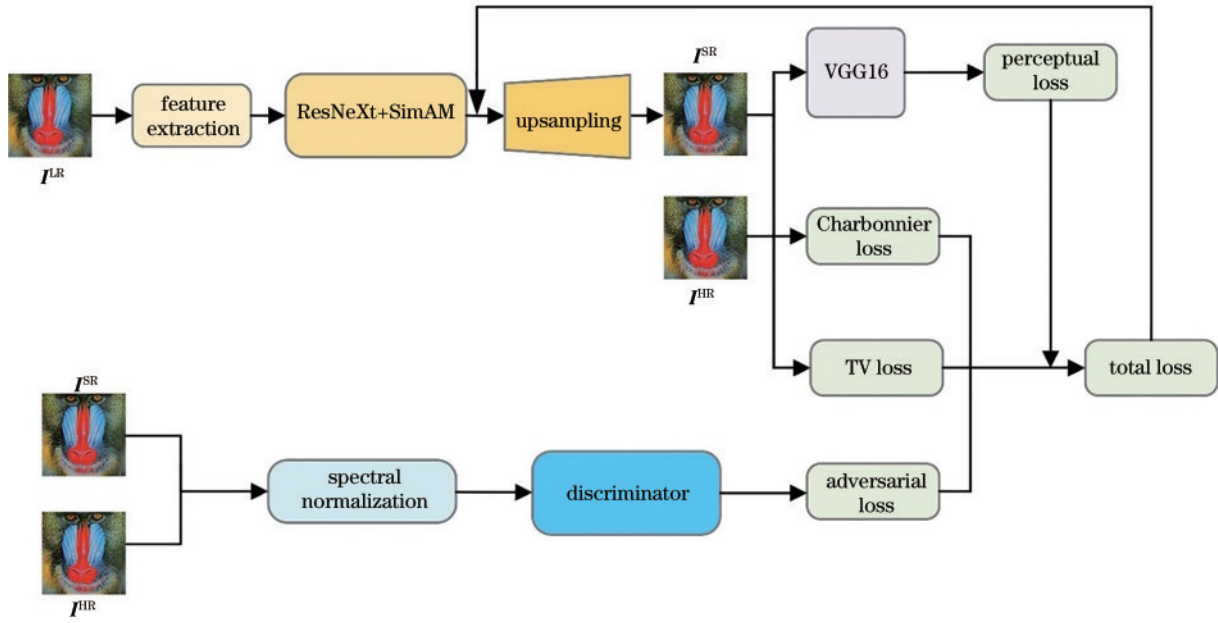


图 3 改进网络的流程

Fig. 3 Flowchart of improved network

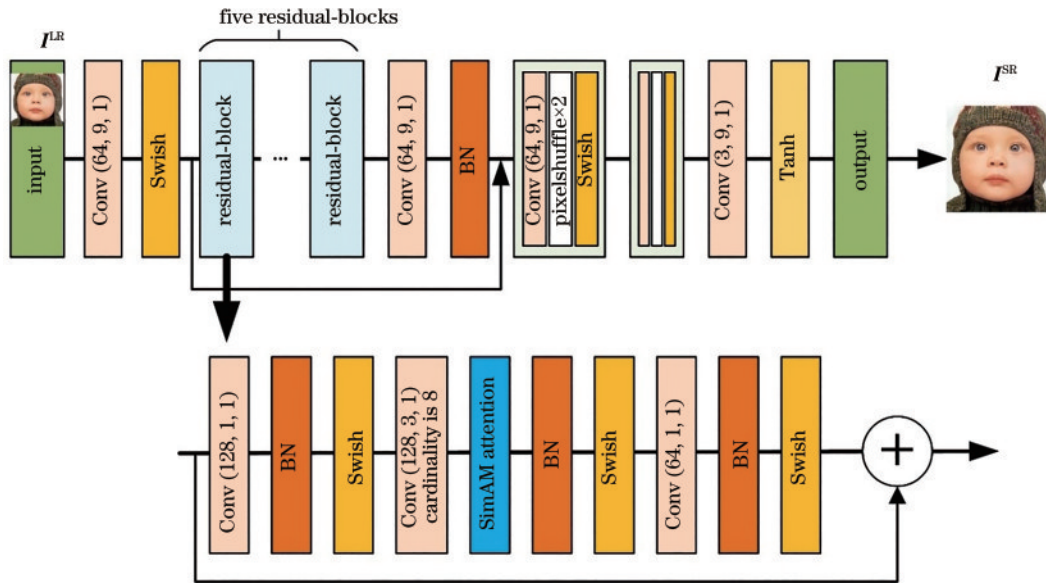


图 4 生成器网络结构

Fig. 4 Generator network structure

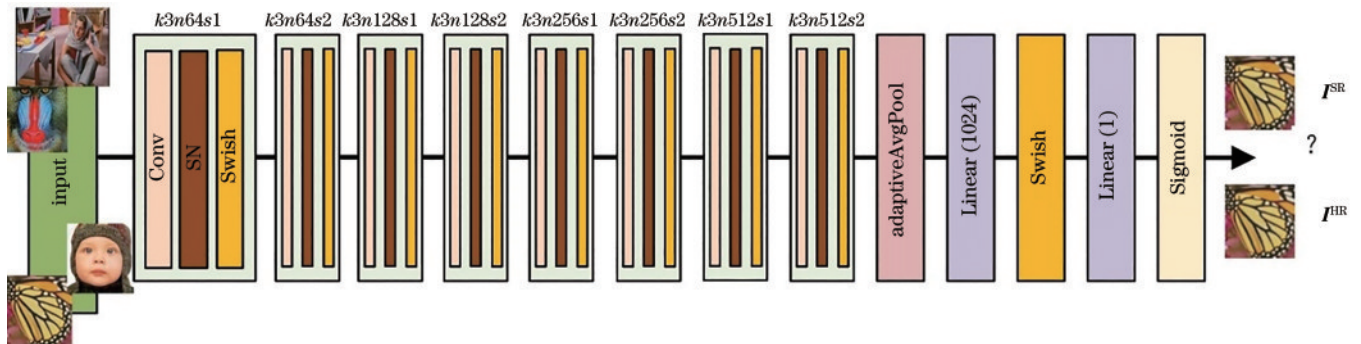


图 5 鉴别器网络结构

Fig. 5 Discriminator network structure

图的大小并扩展通道深度提高鉴别能力。与 SRGAN 模型中的鉴别器不同,该鉴别器在每个卷积层后加入谱归一化处理,并使用 Swish 激活函数,防止在训练过程中出现梯度消失和模式崩溃。最后采用 Sigmoid 函数使鉴别器网络的最终输出是一个单一的值,表示输入图像是生成器生成的或是真实的。

3 实验结果与分析

3.1 实验环境

本实验的硬件设备为 AMD Ryzen 7 4800H with Radeon Graphics 2.90 GHz 处理器,内存 16G,显卡为 NVIDIA GeForce RTX2060 6G;软件环境为 Win10 操作系统,深度学习框架为 PyTorch,编程语言为 Python。

3.2 实验设置

使用的训练数据集是 DIV2K 训练图像数据集^[27],DIV2K 数据集由 800 张训练图像和 100 张验证图像组成。对 Set5^[28]、Set14^[29]、BSD100^[30] 三个通用数据集进行测试。BSD100 数据集由 100 张图像组成;Set5 数据集由 5 张图像组成;Set14 数据集由 14 张图像组成。

在训练阶段,将 HR 图像裁剪到 88×88 大小图像

进行训练,使用放大因子为 4 的双三次插值方法对 HR 图像进行下采样,生成相应的 LR 图像。batchsize 设置为 8,放大因子设置为 4,本模型使用 Adam 算法对收敛速度进行优化,学习率设置为 10^{-4} ,共训练 200 个 epoch。

比较了 Bicubic 算法、SRCNN 算法、ESPCN 算法、SRGAN 算法、ESRGAN 算法、XLSR 算法。为了评估模型性能,在 YCbCr 通道的 Y 通道(即亮度)上计算峰值信噪比 (PSNR) 和结构相似性 (SSIM)。较高的 PSNR 值对应 SR 图像失真较小效果更佳,接近 1 的 SSIM 值表示两幅图像的结构更相似。

3.3 结果分析

3.3.1 客观评估结果与分析

1) Cardinality 的取值对模型性能的影响

为了探究 Cardinality 取值对模型性能的影响,分别选取数值为 4, 8, 16, 32, 64。图 6 表示不同 cardinality 取值下的在 Set5 数据集上的 PSNR 值和 SSIM 值。通过图 6 可知,随着 cardinality 值的增加,PSNR 和 SSIM 值呈现先增加后下降趋势,继续提高 cardinality 值后 PSNR 不会提高。当 cardinality 为 8 时 PSNR 和 SSIM 值最高,故将 8 作为聚合残差块的分组数。

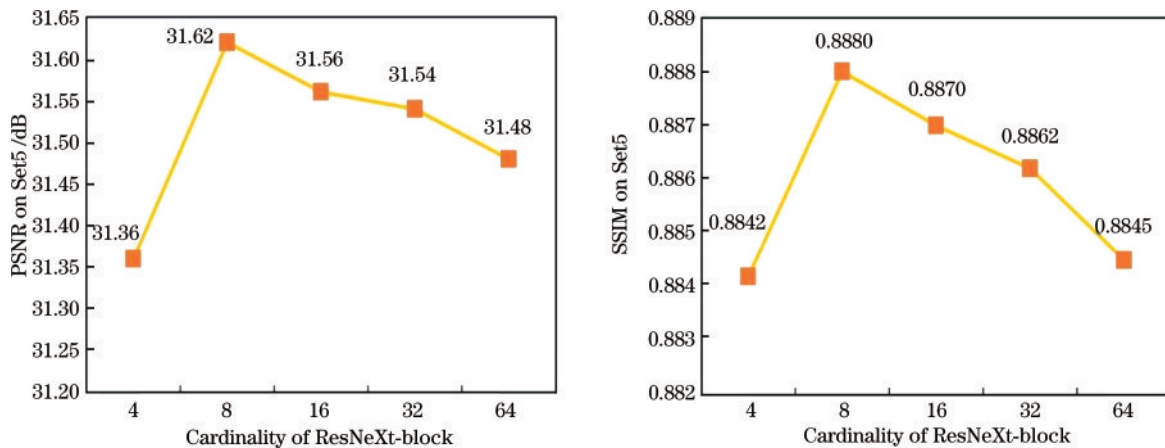


图 6 Cardinality 取值对模块性能的影响

Fig. 6 Effect of cardinality number on module performance

2) 模型中改进模块的性能评估

为了验证所提方法在原始 SRGAN 模型上的有效性,使用以下结构进行消融实验。为了保证实验的公平性,在 SRGAN 模型基础上,在同等参数条件下,分别训练 200 个 epoch,获得重建尺度为 4 时的 PSNR 值和 SSIM 值。“Baseline”表示原始 SRGAN 模型;“Baseline+ResNeXt”表示将经典 ResNet 残差块替换为 ResNeXt 聚合残差块的模型;“Baseline+SimAM”表示在残差块内引入具有三维权重的无参注意力的模型;“Baseline+SN”表示在鉴别器中采用谱归一化的模型;“Baseline+Charbonnier”表示将 Charbonnier 损失函数作为像素损失函数的模型;“Ours”表示所提模型。实验结果如表 1 所示,引入四个改进模块后,

SRGAN 模型的性能均有所提高。实验结果表明:在 Set14 数据集上,当模型引入聚合残差块时,PSNR 值提高 0.6 dB,SSIM 值提高 0.0250;当模型引入具有三维权重的无参注意力时,PSNR 值提高 0.19 dB,SSIM 值提高 0.0134;当模型引入谱归一化处理时,PSNR 值提高 0.24 dB,SSIM 值提高 0.0246;当模型引入 Charbonnier 损失函数时,PSNR 值提高 0.11 dB,SSIM 值提高 0.0155。综上所述,相对于经典残差块,聚合残差块对模型性能的提高效果显著,网络表征能力更好。通过消融实验,证实了四个模块的有效性。

3) 注意力机制性能评估

为验证在聚合残差块内引入的具有三维权重的注

表 1 不同模块组合在 Set14 数据集上 PSNR 值和 SSIM 值
Table 1 PSNR and SSIM of different module combinations on Set14 dataset

Method	PSNR /dB	SSIM
Baseline	27.33	0.7517
Baseline + ResNeXt	27.83	0.7767
Baseline + SimAM	27.52	0.7651
Baseline + SN	27.57	0.7763
Baseline + Charbonnier	27.44	0.7672
Ours	28.50	0.7848

注意力机制的作用,分别进行三组放大因子为 4 的实验,使用 BSD100 数据集计算 PSNR 值,如表 2 所示。在 SRGAN 模型引入聚合残差块的情况下,分别探究具有三维权重的注意力机制和通道间的注意力机制对模型性能的影响。从表 2 得出,使用具有三维权重的注

意力机制的模型比使用通道间的注意力机制的模型的 PSNR 值提高了 0.043 dB,比不使用注意力机制的模型提高 0.121 dB。图 7 进一步展示了表 2 中三种模型在 BSD100 数据集上的视觉效果,使用具有三维权重注意力机制的模型能较好捕捉高频信息,在色彩与纹理细节的重建方面均优于其他模型,验证了具有三维权重的注意力机制的有效性。

表 2 不同模块组合在 BSD100 数据集上的 PSNR 值
Table 2 PSNR value of different module combinations on BSD100 dataset

Method	PSNR /dB
ResNeXt	27.047
ResNeXt+SE	27.125
ResNeXt+SimAM	27.168

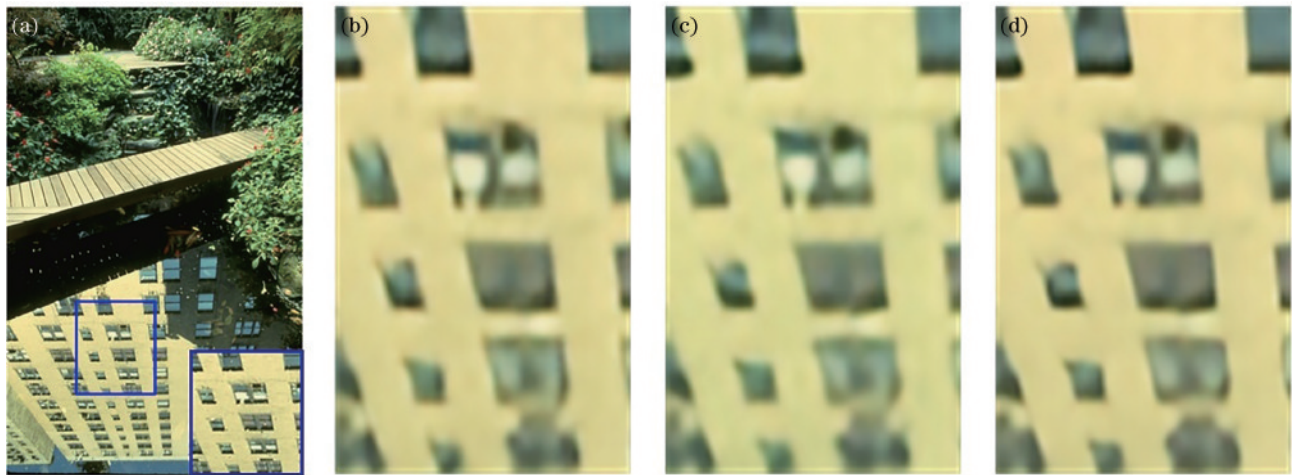


图 7 BSD100 数据集上图像实验对比结果。(a)原图;(b)ResNeXt;(c)ResNeXt+SE;(d)ResNeXt+SimAM
Fig. 7 Experimental comparison result for image on BSD100 dataset. (a) Original image; (b) ResNeXt; (c) ResNeXt+SE; (d) ResNeXt+SimAM

4) 模型整体性能评估

为探究模型整体性能,对所提模型得到的超分辨率重建结果与 Bicubic、SRCNN、ESPCN、SRGAN、ESRGAN、XLSR 这几种算法进行比较,为了实验公平性,所有实验的训练集和验证集均为 DIV2K。表 3 和表 4 为不同 SR 重建算法的 PSNR 和 SSIM 值的比较,黑体数据对应最佳结果。当图像放大因子为 4 时,所提模型实现了最佳的实验结果。如表 3 所示,在 Set5、Set14、BSD100 数据集上,所提模型的 PSNR 值比 Bicubic 模型平均提高 2.34 dB,比 SRCNN 模型平

均提高 1.89 dB,比 ESPCN 模型平均提高 1.50 dB,比 SRGAN 模型平均提高 1.22 dB,比 ESRGAN 模型平均提高 1.69 dB,比 XLSR 模型平均提高 0.61 dB;所提模型的 SSIM 值比 Bicubic 模型平均提高 0.0660,比 SRCNN 模型平均提高 0.0477,比 ESPCN 模型平均提高 0.0538,比 SRGAN 模型平均提高 0.0335,比 ESRGAN 模型平均提高 0.0614,比 XLSR 模型平均提高 0.0115。综上所述,所提模型的 PSNR 和 SSIM 值比其他模型更佳。结果表明,所提模型中的聚合残差注意力网络比其他模型能获取到更多的图像特征,获

表 3 在三种测试集上不同 SR 算法在 4 倍放大因子下的平均 PSNR 值
Table 3 Average PSNR of different SR algorithms on three test sets at 4× magnification factor unit: dB

Dataset	Scale	Bicubic	SRCNN	ESPCN	SRGAN	ESRGAN	XLSR	Ours
Set5	4	28.41	29.15	29.66	29.82	30.47	30.87	31.62
Set14	4	26.09	26.32	26.86	27.33	26.61	27.69	28.50
BSD100	4	25.95	26.32	26.45	26.64	25.32	27.07	27.34

表 4 三种测试集上不同 SR 算法在 4 倍放大因子下的平均 SSIM 值

Table 4 Average SSIM of different SR algorithms on three test sets at 4× magnification factor

Dataset	Scale	Bicubic	SRCNN	ESPCN	SRGAN	ESRGAN	XLSR	Ours
Set5	4	0.8128	0.8291	0.8336	0.8471	0.8518	0.8738	0.8880
Set14	4	0.7184	0.7361	0.7301	0.7517	0.7139	0.7729	0.7848
BSD100	4	0.6712	0.6921	0.6754	0.7011	0.6505	0.7192	0.7277

取更多高频信息,提高网络特征表达能力和泛化能力,促进网络重建过程,能得到较高客观评价价值。

3.3.2 主观结果展示与分析

在不同超分辨率重建方法中裁剪相同的区域并放大预测区域进行比较,结果如图 8~10 所示。从图 8“baby”左眼局部放大图可以观察到:通过 Bicubic 重建方法处理后,SR 图像非常模糊,图像出现失真现象并缺少细节信息;SRCNN 和 ESPCN 二者均以 MSE 作为目标函数来指导模型进行训练,相比于 Bicubic 方法,得到的图像更清晰,但均出现了模糊、

振铃效应并且锯齿状严重;SRGAN 算法的重建图像振铃效应基本消失,但缺乏视觉上的真实感,过于平滑;ESRGAN 的重建效果最为清晰,边缘明显,但客观评价价值比所提模型低;XLSR 算法的重建图像与所提模型的重建图像在视觉上较为相近,但重建图像边缘区域出现模糊现象;所提模型能够恢复较多的高频细节,与原始图像颜色匹配更好,并且拥有最佳的 PSNR 值和 SSIM 值,这些结果证明了所提模型的有效性。所提模型可以恢复更多的高频细节并生成清晰的 SR 图像。

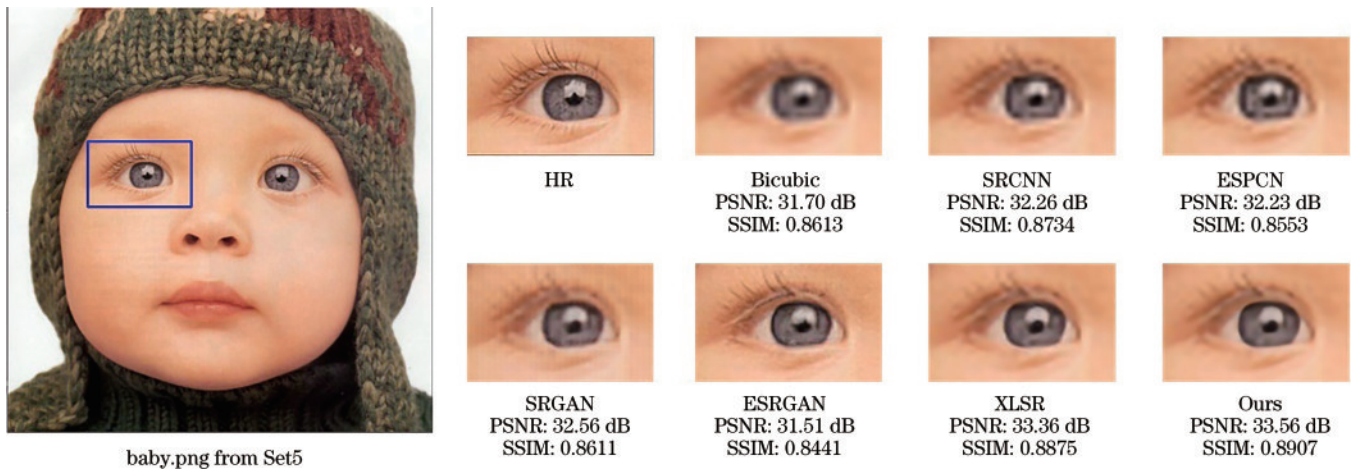


图 8 在 Set5 中“baby”重建效果对比

Fig. 8 Comparison of reconstruction effect of "baby" on Set5 dataset

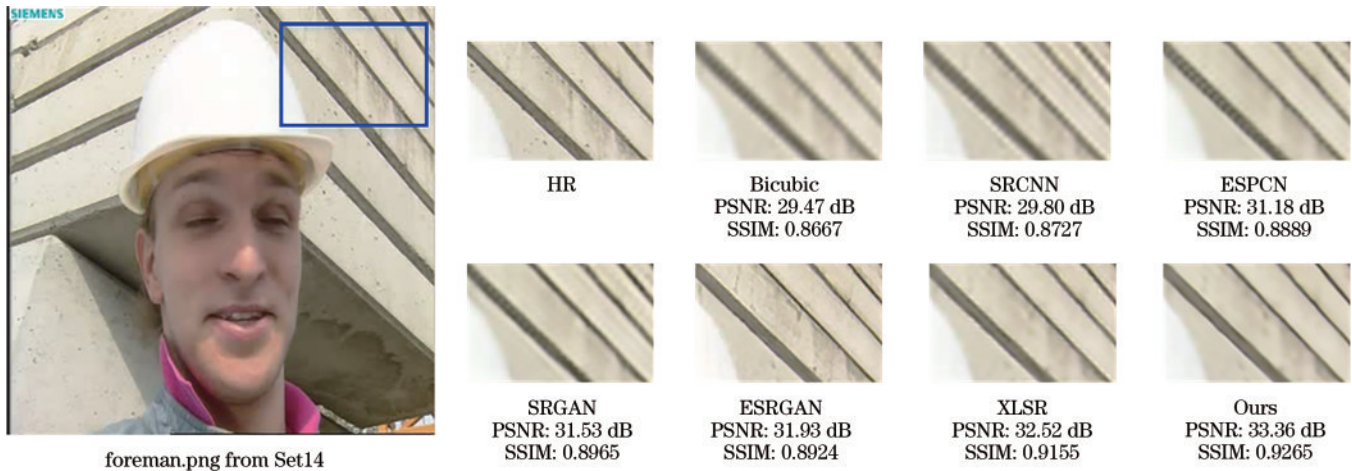


图 9 在 Set14 中“foreman”重建效果对比

Fig. 9 Comparison of reconstruction effect of "foreman" on Set14 dataset

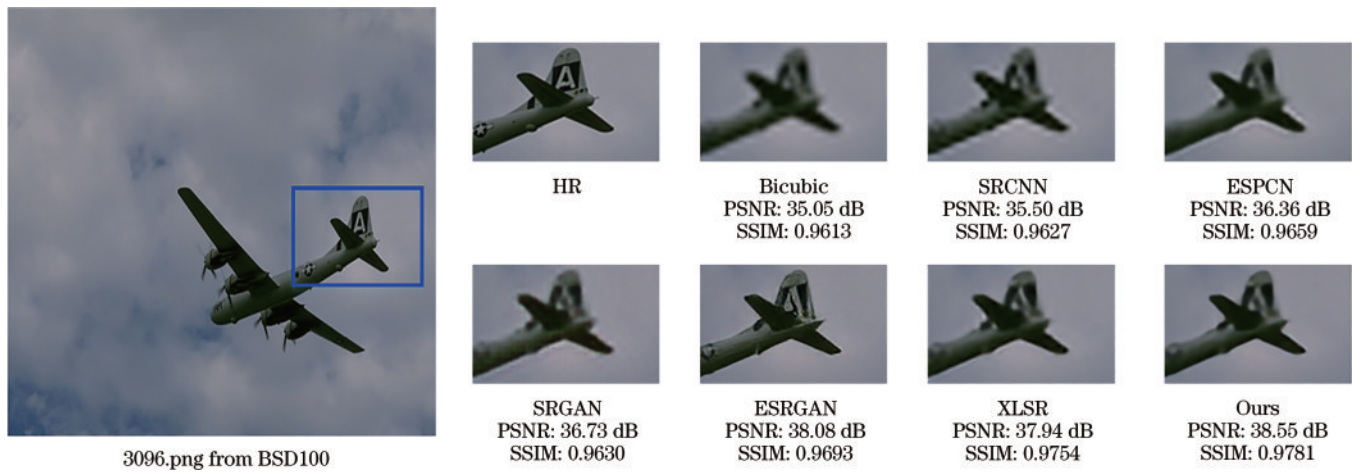


图 10 在 BSD100 中“3096”重建效果对比

Fig. 10 Comparison of reconstruction effect of “3096” on BSD100 dataset

4 结 论

为解决重建 SR 图像特征提取不足、算法计算复杂度过高、训练不稳定等问题,提出了一种基于聚合残差注意力网络的单图像超分辨率重建方法。在生成器网络中,采用带有注意力机制的聚合残差网络提取图像特征信息,并使用拟合性更好的 Swish 激活函数提高网络表达能力,优化图像生成质量。采用 Charbonnier 损失函数和 TV 正则项重新设计损失函数,指导网络生成具有更多细节和清晰轮廓的超分辨率图像。在鉴别器网络中,使用谱归一化处理提高网络的稳定性和鉴别能力。实验结果表明:所提模型在客观评价上均优于其他经典方法;在模型参数量上,比原始 SRGAN 残差块模型的参数量减少 53%;在视觉效果上,能恢复更多高频细节,但略低于 ESRGAN,部分图像真实感有待提升。在未来研究中,将进一步优化网络框架模型和损失函数,提高视觉感知质量。

参 考 文 献

- [1] Yang X M, Wu W, Liu K, et al. Long-distance object recognition with image super resolution: a comparative study[J]. IEEE Access, 2018, 6: 13429-13438.
- [2] 黄硕, 胡勇, 顾明剑, 等. 基于深度学习的红外遥感目标超分辨率检测算法[J]. 激光与光电子学进展, 2021, 58(16): 1610015.
Huang S, Hu Y, Gu M J, et al. Super-resolution infrared remote-sensing target-detection algorithm based on deep learning[J]. Laser & Optoelectronics Progress, 2021, 58(16): 1610015.
- [3] Wang Y K, Teng Q Z, He X H, et al. CT-image of rock samples super resolution using 3D convolutional neural network[J]. Computers & Geosciences, 2019, 133: 104314.
- [4] Heckel R, Morgenshtern V I, Soltanolkotabi M. Super-resolution radar[J]. Information and Inference: A Journal of the IMA, 2016, 5(1): 22-75.
- [5] 杨飞凡, 李晖, 彭晶, 等. 基于图正则化的高分辨率光场显微成像研究[J]. 光学学报, 2021, 41(9): 0918001.
Yang F F, Li H, Peng J, et al. Research on microscopic imaging of high resolution light field based on graph regularization[J]. Acta Optica Sinica, 2021, 41(9): 0918001.
- [6] Zhu S Y, Zeng B, Zeng L Y, et al. Image interpolation based on non-local geometric similarities and directional gradients[J]. IEEE Transactions on Multimedia, 2016, 18(9): 1707-1719.
- [7] Beck A, Teboulle M. Convergence rate analysis and error bounds for projection algorithms in convex feasibility problems[J]. Optimization Methods and Software, 2003, 18(4): 377-394.
- [8] 李勇, 金秋雨, 赵怀慈, 等. 基于改进残差密集网络的高光谱重建[J]. 光学学报, 2021, 41(7): 0730001.
Li Y, Jin Q Y, Zhao H C, et al. Hyperspectral image reconstruction based on improved residual dense network [J]. Acta Optica Sinica, 2021, 41(7): 0730001.
- [9] Dong C, Loy C C, He K M, et al. Image super-resolution using deep convolutional networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(2): 295-307.
- [10] Shi W Z, Caballero J, Huszár F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 1874-1883.
- [11] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 1646-1654.
- [12] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 26-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [13] Kim J, Lee J K, Lee K M. Deeply-recursive

- convolutional network for image super-resolution[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 26-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 1637-1645.
- [14] Zhang Y L, Tian Y P, Kong Y, et al. Residual dense network for image super-resolution[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 2472-2481.
- [15] Zhang Y L, Li K P, Li K, et al. Image super-resolution using very deep residual channel attention networks[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 294-310.
- [16] Ayazoglu M. Extremely lightweight quantization robust real-time single-image super resolution for mobile devices [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 19-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 2472-2479.
- [17] Xie S N, Girshick R, Dollár P, et al. Aggregated residual transformations for deep neural networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 5987-5995.
- [18] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[EB/OL]. (2014-06-10) [2021-02-05]. <https://arxiv.org/abs/1406.2661>.
- [19] Ledig C, Theis L, Huszár F, et al. Photo-realistic single image super-resolution using a generative adversarial network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 105-114.
- [20] Wang X T, Yu K, Wu S X, et al. ESRGAN: enhanced super-resolution generative adversarial networks[M]//Leal-Taixé L, Roth S. Computer vision-ECCV 2018 workshops. Lecture notes in computer science. Cham: Springer, 2019, 11133: 63-79.
- [21] 辛元雪, 朱凤婷, 史朋飞, 等. 基于改进增强型超分辨率生成对抗网络的图像超分辨率重建算法[J]. 激光与光电子学进展, 2022, 59(4): 0420002.
Xin Y X, Zhu F T, Shi P F, et al. Super-resolution reconstruction algorithm of images based on improved enhanced super-resolution generative adversarial network [J]. Laser & Optoelectronics Progress, 2022, 59(4): 0420002.
- [22] Yang L, Zhang R Y, Li L, et al. Simam: a simple, parameter-free attention module for convolutional neural networks[C]//Proceedings of the 38th International Conference on Machine Learning, July 18-24, 2021, Virtual Event. Cambridge: PMLR, 2021: 11863-11874.
- [23] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 42(8): 2011-2023.
- [24] 雷鹏程, 刘丛, 唐坚刚, 等. 分层特征融合注意力网络图像超分辨率重建[J]. 中国图象图形学报, 2020, 25(9): 1773-1786.
Lei P C, Liu C, Tang J G, et al. Hierarchical feature fusion attention network for image super-resolution reconstruction[J]. Journal of Image and Graphics, 2020, 25(9): 1773-1786.
- [25] Webb B S, Dhruv N T, Solomon S G, et al. Early and late mechanisms of surround suppression in striate cortex of macaque[J]. The Journal of Neuroscience, 2005, 25(50): 11666-11675.
- [26] Miyato T, Kataoka T, Koyama M, et al. Spectral normalization for generative adversarial networks[EB/OL]. (2018-02-16) [2022-02-05]. <https://arxiv.org/abs/1802.05957>.
- [27] Timofte R, Agustsson E, van Gool L, et al. Ntire 2017 challenge on single image super resolution: methods and results[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 114-125.
- [28] 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, September 3-7, 2012, Surrey, UK. London: British Machine Vision Association, 2012: 1-10.
- [29] Zeyde R, Elad M, Protter M. On single image scale-up using sparse-representations[M]//Boissonnat J D, Chenin P, Cohen A, et al. Curves and surfaces. Lecture notes in computer science. Heidelberg: Springer, 2012, 6920: 711-730.
- [30] 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, July 7-14, 2001, Vancouver, BC, Canada. New York: IEEE Press, 2001: 416-423.