

文章编号:1007-2780(2022)03-0367-09

结合双注意力和结构相似度量的 图像超分辨率重建网络

黄友文, 唐 欣*, 周 斌

(江西理工大学 信息工程学院, 江西 赣州 341000)

摘要:针对低分辨率图像到高分辨率图像的映射函数解空间极大,导致超分辨率重建模型难以产生细致纹理的问题,本文提出一种结合双注意力和结构相似度量的图像超分辨率重建网络。以改进的 U-Net 网络模型作为基本结构,引入针对低级别视觉任务的数据增强方法,增加样本多样性。编码器部分由卷积层和自适应参数线性整流激活函数(Dynamic ReLU)组成。同时引入了一种残差双注意力模块(Residual Dual Attention Module, RDAM)与像素重组(PixelShuffle)模块共同构成解码器,通过上采样操作,逐级放大图像。为了使生成图像更加符合人眼视觉特性,采用了一种结合结构相似度量准则的损失函数,增强网络约束。实验结果表明:重建图像的质量对比 SRCNN,在 Set5、Set14、BSD100 和 Urban100 标准测试集上的平均 PSNR 提升约 1.64 dB, SSIM 提升约 0.047。本文方法能够使重建的图像纹理细节更丰富,有效地减少了映射函数可能的解空间。

关键词:超分辨率; U-Net 网络; 数据增强; 双注意力; 结构相似

中图分类号: TP391.4 **文献标识码:** A **doi:** 10.37188/CJLCD.2021-0178

Image super-resolution reconstruction network with dual attention and structural similarity measure

HUANG You-wen, TANG Xin*, ZHOU Bin

(School of Information Engineering, Jiangxi University of Science and Technology,
Ganzhou 341000, China)

Abstract: Aiming at the problem that the solution space of mapping function from low resolution image to high resolution image is extremely large, which makes it difficult for super-resolution reconstruction models to generate detailed textures, this paper proposes a image super resolution that combines dual attention and structural similarity measure. With the improved U-Net network model as the basic structure, the data augmentation methods for low-level vision tasks are introduced to increase sample diversity. The encoder is composed of a convolution layer and an adaptive parameter linear rectifier function (Dynamic ReLU). At the same time, a residual dual attention module(RDAM) is introduced, which forms a decoder together with the PixelShuffle module. The image is enlarged gradually

收稿日期:2021-07-06; 修订日期:2021-09-27.

基金项目:江西省教育厅科技项目(No.GJJ180443)

Supported by Science and Technology Project of the Education Department of Jiangxi Province (No. GJJ180443)

* 通信联系人, E-mail: tangxin.2108@163.com

through the up-sampling operation. In order to make the generated image more in line with the human visual characteristics, a loss function combined with structural similarity measurement criteria is proposed to enhance the network constraints. The experimental results show that the average PSNR of the quality of the reconstructed image on the Set5, Set14, BSD100 and Urban100 standard test sets is improved by about 1.64 dB, and the SSIM is improved by about 0.047 compared with SRCNN. The proposed method can make the reconstructed image texture more detailed and reduce the possible solution space of the mapping function effectively.

Key words: super-resolution; U-net network; data augmentation; dual attention; structural similarity

1 引言

图像超分辨率(Super-resolution, SR)重建技术是从单帧或多帧低分辨率(Low resolution, LR)图像中重建出高分辨率(High resolution, HR)图像的过程,是近年来计算机视觉领域的一个重要研究方向,经历了基于插值的、重建和学习 3 个发展阶段^[1]。随着深度学习的发展,基于学习的图像超分辨率重建算法成为研究的热点。Dong 等人^[2]提出卷积神经网络的超分辨率(Super-Resolution Convolutional Neural Networks, SRCNN),通过 3 层卷积结构实现图像重建,但其网络层次太浅,不能充分提取图像特征,且它对超参数的变化非常敏感。Kim 等人^[3]提出极深卷积网络的超分辨率,将网络加深到 20 层,同时引入残差思想减轻了网络训练的负担,但增加网络深度会加大模型的参数量和计算量,产生梯度消失的问题。Kim 等人^[4]又提出深层递归神经网络(Deeply-Recursive Convolutional Network, DRCN),使用深层递归的方法,在卷积层间共享参数,缓解了梯度消失问题,但 DRCN 网络层间信息流动性较差,没有充分挖掘图像不同层次间的特征。Lai 等人^[5]提出基于 Laplacian 金字塔的超分辨率,该网络通过残差学习和逐级上采样恢复 HR 图像,移除插值放大的预处理,减少了图像重建时间。Lim 等人^[6]提出增强的深度残差网络的超分辨率,通过去除残差块中的批归一化层,扩大模型的尺寸来提升重建质量。Zhang 等人^[7]提出残差密集网络(Residual Dense Network, RDN),该网络通过密集连接的卷积层提取并融合局部特征,然后进行全局融合。Zhang 等人^[8]提出深层残差注意力的超分辨率

(Residual Channel Attention Networks, RCAN),该网络在残差块中加入通道注意力,使模型能够学习更多通道特征,提高网络的判别能力。Ledig 等人^[9]提出超分辨率生成对抗网络(Super-Resolution Generative Adversarial Network, SRGAN),该算法通过两个网络交替迭代训练,但 SRGAN 生成的高分辨率图像相对原图较模糊。Wang 等人^[10]又提出增强的超分辨率生成对抗网络,该网络使用密集残差块作为生成网络的主体,获得了较好的重建效果,但生成的图像存在过多的伪细节。Xiao 等人^[11]提出可逆缩放网络(Invertible Rescaling Networks, IRN),将 HR 图像经小波变换分解为低频分量和高频分量作为网络的输入,同时网络在缩放过程中使用指定分布后的潜在变量,来捕获丢失的信息分布,然后以此逆变换重建 HR 图像,该方法能有效提升重建模型的性能。

为进一步解决 LR 图像到 HR 图像的映射函数解空间极大,导致 SR 模型性能有限,难以产生纹理细致、边缘清晰图像的问题,本文在 U-Net 模型^[12]的基础上提出一种结合双注意力和结构相似度量的图像超分辨率重建网络。本文主要创新包括:(1)该网络引入针对低级别视觉任务的数据增强模块对训练集进行预处理,增加训练样本的多样性,提高模型的泛化能力。(2)在改进的 U-Net 网络模型中引入残差双注意力模块,将特征提取集中在关键位置和通道中,残差块能有效缓解由于网络加深造成的梯度消失问题。(3)结合对偶回归损失和结构相似度量的损失约束,增强网络约束,有效提高生成图像的质量。本文方法在客观评价指标上优于其他对比的方法,有效减少了映射函数可能的解空间。

2 结合双注意力和结构相似度量的超分辨率重建网络

2.1 数据增强

训练样本有限时,数据增强能够增加样本的多样性。目前常用的数据增强方法大多是为高级别视觉任务开发的,而超分辨率重建是一种典型的低级别视觉任务,需要根据全局关系决定局部像素点的处理方式。本文引入针对低级别视觉任务的数据增强方法,主要包括 RGB、Blend、CutMix^[13]、Mixup、CutMixup 和 CutBlur^[14]。RGB 是打乱 RGB 像素的排列;Blend 是添加固定像素值;CutMix 是随机切除图像中的一块区域,并使用另一个图像替换随机切除的区域,使图像的上下文边界产生急剧变换;Mixup 是随机将两个图像按照比例混合;CutMixup 方法结合了 CutMix 与 Mixup,可以使边界效应和混合图像上下文信息的比率最小化;CutBlur 是切割一个低分辨率图像块,并粘贴到相应的高分辨率图像区域。根据文献^[14]中的结论,采用混合增强方法(MOA)显著提高了各种场景的性能,尤其当模型规模很大并且数据集是来自于现实环境中。

2.2 网络结构

本文提出一种结合双注意力和结构相似度量的超分辨率重建网络,网络结构如图 1 所示。

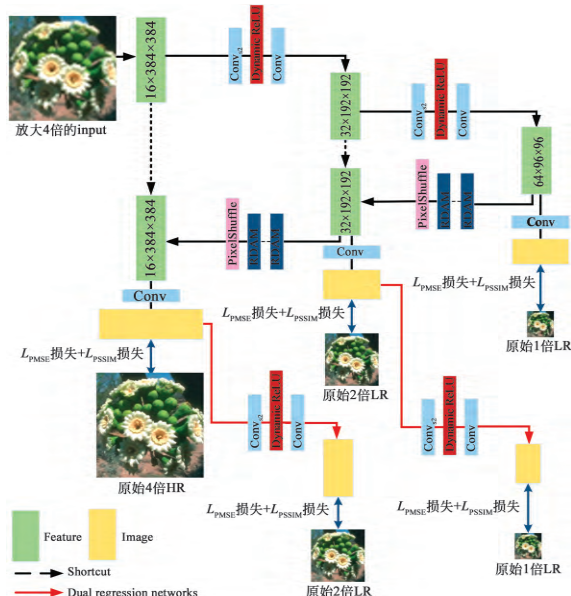


图 1 网络结构

Fig.1 Network structure

首先采用 Bicubic 方式对输入图像放大 4 倍,之后通过卷积提取该 4 倍图像的特征,维度是 $16 \times 384 \times 384$ 。下采样模块由 $\text{Conv}_{s2} + \text{Dynamic ReLU} + \text{Conv}$ 构成,其中 Conv_{s2} 是步长为 2 的卷积。通过下采样模块将输入特征逐级缩小,得到 2 倍和 1 倍的 LR 图像,下采样操作完成。生成的特征经过卷积层得到 1 倍的 LR 图像,与原始 1 倍 LR 图像比较,得到重建损失。上采样模块由 30 个 RDAM 模块和 PixelShuffle 构成,RDAM 由残差和双注意力模块组成,将残差映射依次输入到双注意力模块中产生注意力权重,再经过残差块进行差值计算,通过原始输入和残差结果进行求和输出。PixelShuffle 像素重组的主要功能是将一个输入为 $H \times W$ 的特征图通过卷积和多通道间的重组得到大小为 $rH \times rW$ 的特征图,其中 r 为上采样因子, $r=2$ 。通过上采样模块将输入特征逐级放大,生成 2 倍和 4 倍的图像,上采样操作完成。虚线箭头部分利用跳跃连接将下采样输出和上采样输入的特征图连接起来,相当于将低层特征拼接到相应的高层特征上,创造了信息的传播路径。图 1 中红色箭头部分表示对偶回归网络(Dual regression networks)^[15],通过反向下采样操作得到 2 倍和 1 倍的 LR 图像,与原始的 LR 图像比较,得到对偶回归网络损失,增加网络的约束。蓝色双箭头部分表示本文损失函数的实现路径。

2.3 残差双注意力模块

何恺明等人提出残差网络(Residual Neural Network)^[16],通过残差学习能有效简化模型训练,缓解网络加深造成的梯度消失问题。本文将残差块引入超分辨率重建网络中,与双注意力模块^[17]构成残差双注意力模块(Residual Dual Attention Module, RDAM),结构如图 2 所示。

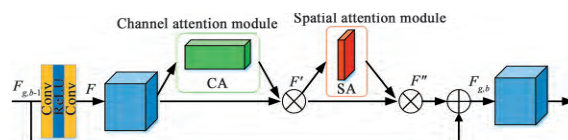


图 2 残差双注意力模块

Fig.2 Residual dual attention module

首先对输入特征进行卷积操作得到残差映射,完成特征维度匹配。由于 Batch Norm 对图像来说类似于对比度的拉伸,图像经过 Batch

Norm 层后,其色彩分布会被归一化,破坏了图像原本的对比度信息,影响图像输出的质量。本文在残差映射块中去掉 Batch Norm 层,由两个标准卷积层和 ReLU 激活函数构成。其计算过程如公式(1)所示:

$$F = W_1 \delta(W_0 F_{g,b-1}) \quad (1)$$

其中 F 表示从输入 $F_{g,b-1}$ 中学习到的残差映射, δ 表示 ReLU 激活函数, W_0 和 W_1 分别表示残差块中卷积层的权重。

注意力机制是人类视觉所特有的大脑信号处理机制,其主要思想是强化特征,以获取更多关注目标的细节信息。图 2 中的注意力模块是由通道注意力(ChannelAttention, CA)和空间注意力(SpatialAttention, SA)依次级联的双注意力模块^[17],从通道和空间两个维度提取注意力特征。

CA 模块首先对残差映射 F 进行全局平均池化和最大池化操作,分别表示为 F_{avg}^c 和 F_{max}^c ,随后经过多层感知器网络(Multi-Layer Perceptron, MLP)为每个通道产生权重。将结果进行元素求和,通过 sigmoid 激活函数后输出通道注意力的权重信息 $M_c \in R^{C \times 1 \times 1}$, $W_2 \in R^{C \times 16 \times C}$ 和 $W_3 \in R^{C \times C \times 16}$ 是 MLP 的两层参数, σ 表示 sigmoid 激活函数,如公式(2)所示:

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F))) + \text{MLP}(\text{MaxPool}(F)) = \sigma(W_3(W_2(F_{\text{avg}}^c) + W_3(W_2(F_{\text{max}}^c)))) \quad (2)$$

$M_c(F)$ 与输入特征信息 F 进行元素点乘运算得到 F' , 如公式(3)所示:

$$F' = M_c(F) \otimes F \quad (3)$$

接着 SA 模块沿通道方向对输入特征信息 F' 进行平均池化和最大池化操作,分别表示为 $F_{\text{avg}}^s \in R^{1 \times H \times W}$ 和 $F_{\text{max}}^s \in R^{1 \times H \times W}$ 并将其按照 cat 方式拼接,再经过一个 7×7 的卷积层对此特征映射进行卷积操作,经过激活函数后输出空间注意力映射 $M_s \in R^{1 \times H \times W}$ 。如公式(4)所示:

$$M_s(F) = \sigma(f^{7 \times 7}([\text{AvgPool}(F); \text{MaxPool}(F)])) = \sigma(f^{7 \times 7}([F_{\text{avg}}^s; F_{\text{max}}^s])) \quad (4)$$

$M_s(F')$ 与 F' 进行点乘运算得到最终残差注意力特征信息 F'' , 如公式(5)所示:

$$F'' = M_s(F') \otimes F' \quad (5)$$

最后将此残差注意力特征信息与原始输入进行元素求和,如公式(6)所示,其中 $F_{g,b-1}$ 和 $F_{g,b}$

分别表示残差双注意力模块的输入和输出。

$$F_{g,b} = F_{g,b-1} + F'' \quad (6)$$

2.4 激活函数

在神经网络中常用的激活函数有 ReLU、LeakyReLU 和 PReLU 等,这些激活函数的参数往往是固定的。文献[18]中提出 Dynamic ReLU 激活函数,该函数能够根据输入特征动态地调整对应的分段函数,有效提升了网络的表达能力。

Dynamic ReLU 是分段函数 $f_{\theta(x)}(x)$, 其参数由超函数 $\theta(x)$ 根据 x 计算得到。如公式(7)所示,超函数 $\theta(x)$ 用于获得激活函数的参数, $f_{\theta(x)}(x)$ 用于获得激活函数的输出值,系数 (a_c^k, b_c^k) 由超函数 $\theta(x)$ 计算得到, K 为函数个数, C 为通道数目。

$$\theta(x) = [a_1^1, \dots, a_C^1, \dots, a_1^K, \dots, a_C^K, b_1^1, \dots, b_C^1, \dots, b_1^K, \dots, b_C^K]^T \quad (7)$$

$$y_c = f_{\theta(x)}(x_c) = \max_{1 \leq k \leq K} \{a_c^k(x) x_c + b_c^k(x)\} \quad (8)$$

2.5 损失函数

本文中改进的 U-net 网络的损失函数由对偶回归网络损失和重建损失两部分叠加组成,通过均方误差和结构相似度进行度量。

2.5.1 均方误差

均方误差表示生成图像与原始图像间存在的均方误差值(Mean Square Error, MSE), 公式如式(9)所示,其中 M 与 N 分别为图像的高度和宽度, $X(i, j)$ 和 $Y(i, j)$ 分别为原始图像和生成图像。

$$\text{MSE} = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N (X(i, j) - Y(i, j))^2 \quad (9)$$

文中 MSE 损失公式如(10)、(11)所示:

$$L_{\text{PMSE}} = \sum_{i=1}^W L_{\text{MSE}}(P(x_i), y_i) \quad (10)$$

$$L_{\text{DMSE}} = \sum_{i=1}^Q L_{\text{MSE}}(D(P(x_i)), x_i) \quad (11)$$

其中: $P(x_i)$ 为改进的 U-Net 网络, $D(P(x_i))$ 为对偶回归网络, $W=3$ 是上采样逐级放大 1 倍、2 倍和 4 倍; $Q=2$ 是反向下采样逐步缩小 2 倍和 1 倍。

2.5.2 结构相似度

图像之间的像素相似度往往不能充分反映重建图像的好坏,利用结构相似度(Structural Simi-

larity, SSIM) 能够衡量图像之间的结构相似程度, SSIM 是对亮度、对比度和结构等不同因素间的整体估算, SSIM 与人类视觉感知系统 (Human vision system, HVS) 相似。如式(12)所示:

$$\text{SSIM}(x, y) = \frac{(2\mu_x \mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \quad (12)$$

其中: μ_x 和 μ_y 为像素均值, 是对亮度的估计; σ_x^2 和 σ_y^2 为方差, 是对比度的估计; σ_{xy} 为图像的协方差; $c_1 = (k_1 L)^2$; $c_2 = (k_2 L)^2$; L 是像素的取值区间, 一般的 $k_1 = 0.01$, $k_2 = 0.03$, $L = 255$ 。SSIM 损失计算公式如(13)所示, N 代表图片数量:

$$L_{\text{SSIM}}(x, y) = \frac{1}{N} \sum_{i=1}^N 1 - \text{SSIM}(x, y). \quad (13)$$

本文中 SSIM 损失计算公式如(14)、(15)所示:

$$L_{\text{PSSIM}} = \sum_{i=1}^W L_{\text{SSIM}}(P(x_i), y_i), \quad (14)$$

$$L_{\text{DSSIM}} = \sum_{i=1}^Q L_{\text{SSIM}}(D(P(x_i)), x_i). \quad (15)$$

2.5.3 对偶回归网络损失

受对偶回归网络^[15]的启发, 引入对偶回归网络损失, 用于评估重建图像经下采样得到的 LR 图像与参考图像的差异。通过生成 4 倍的 HR 图像进行下采样得到 2 倍和 1 倍的 LR 图像, 与真实的 LR 图像比较, 得到对偶回归网络损失 L_D , 用于对偶回归网络的训练, 如公式(16)所示, 其中 λ 和 β 为权重参数, 本文采用 $\lambda = 0.1$, $\beta = 10$ 。

$$L_D = \lambda L_{\text{DMSE}} + \beta L_{\text{DSSIM}} = \sum_{i=1}^Q \lambda L_{\text{MSE}} \times (D(P(x_i)), x_i) + \sum_{i=1}^Q \beta L_{\text{SSIM}}(D(P(x_i)), x_i). \quad (16)$$

2.5.4 重建损失

重建损失是用于评估改进的 U-net 网络重建得到的图像与参考图像的差异, 网络通过下采样得到 1 倍的 LR 图像和上采样生成 2 倍和 4 倍的图像, 与真实的 HR 图像比较, 得到重建损失 L_P , 如公式(17)所示, α 为权重参数, 本文采用 $\alpha = 20$ 。

$$L_P = L_{\text{PMSE}} + \alpha L_{\text{PSSIM}} = \sum_{i=1}^W L_{\text{MSE}}(P(x_i), y_i) + \sum_{i=1}^W \alpha L_{\text{SSIM}}(P(x_i), y_i). \quad (17)$$

对改进的 U-net 网络进行训练时采用公式(18)所示的总损失 L_{Total} 进行约束, 其中 L_P 为重建损失, L_D 为对偶回归网络损失。

$$L_{\text{Total}} = L_P + L_D. \quad (18)$$

3 实验及结果分析

实验平台硬件配置为 Intel Xeon 3104 处理器, 16 GB 内存、NVIDIA 2080Ti 显卡, 并使用 64 位操作系统 Ubuntu 18.04 和 PyTorch 深度学习框架进行训练和测试。

3.1 数据集及设置

实验采用 DIV2K 数据集对网络模型进行训练, 该数据集包含 800 张 2K 高质量图像。本文选取 Set5、Set14、BSD100 和 Urban100 为测试集, 分别包括 5, 14, 100, 100 张高分辨率图像。DIV2K 数据集中每张高清图片都通过 Bicubic 操作得到了对应的缩小 2 倍和缩小 4 倍的低分辨率图像。对于每次训练迭代, 本文以相同的概率从 RGB、Blend、CutMix、Mixup、CutMixup 和 CutBlur 这 6 种增强方法中随机选择一种增强方法对输入的训练数据进行预处理, 增加了训练样本的多样性, 进一步提升模型的泛化能力。在训练过程中, 采用 Adam 优化函数来优化网络, 其中 $\beta_1 = 0.9$, $\beta_2 = 0.99$, 网络的初始学习率设置为 10^{-4} , 采用余弦退火算法对学习率进行自动调整。

3.2 客观实验结果分析

为验证算法的效果, 本文采用峰值信噪比 (Peak Signal-to-Noise Ratio, PSNR) 和结构相似度 (Structural Similarity, SSIM) 来评估改进的网络。本文在 Set5、Set14、BSD100 和 Urban100 标准测试集上验证 4 倍超分辨率重建的性能, 并与一些当前代表性的超分辨率重建方法进行比较, 结果参见表 1。其中对比方法的 PSNR 和 SSIM 数值均来自文献^[19]中的结果。从表 1 可以看出, 本文方法在 4 个测试集上的 PSNR 和 SSIM 数值均获得了较好的结果。对于 Set5、Set14、BSD100 和 Urban100 测试集, 本文方法与 SRCNN 相比, PSNR 分别提升 2.06, 1.36, 0.87, 2.24 dB, SSIM 值分别提升 0.036 5, 0.036 6, 0.032 9, 0.083 7。与 SRFBN 相比, PSNR 分别提升 0.07, 0.05, 0.05, 0.16 dB, SSIM 值分别提升 0.001, 0.001 1, 0.002 1, 0.004 3。对于 Urban 100 测试集, 本文方法与 DBPN 相比, PSNR 提升

表 1 各超分辨率方法的客观评价对比

Tab.1 Objective evaluation comparison of various super-resolution methods

Scale	Method	Parameters/K	Set5	Set14	BSD100	Urban100
			PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
×4	SRCNN ^[2]	57	30.48/0.862 8	27.50/0.751 3	26.90/0.710 1	24.52/0.722 1
	VDSR ^[3]	665	31.35/0.883 8	28.02/0.767 8	27.29/0.725 2	25.18/0.752 5
	DRCN ^[4]	1 775	31.53/0.885 4	28.03/0.767 3	27.24/0.723 3	25.14/0.751 1
	LapSRN ^[5]	813	31.54/0.886 6	28.09/0.769 4	27.32/0.726 4	25.21/0.755 3
	SRGAN ^[9]	1 500	32.05/0.891 0	28.53/0.780 4	27.57/0.735 4	26.07/0.783 9
	CARN ^[20]	1 592	32.13/0.893 7	28.60/0.780 6	27.58/0.734 9	26.07/0.783 7
	EDSR ^[6]	43 000	32.46/0.896 8	28.80/0.787 6	27.71/0.742 0	26.64/0.803 3
	DBPN ^[21]	10 400	32.47/0.898 0	28.82/0.786 0	27.72/0.740 0	26.38/0.794 6
	RDN ^[7]	21 900	32.47/0.899 0	28.81/0.787 1	27.72/0.741 9	26.61/0.802 8
	SRFBN ^[22]	3 631	32.47/0.898 3	28.81/0.786 8	27.72/0.740 9	26.60/0.801 5
	Ours	4 810	32.54/0.899 3	28.86/0.787 9	27.77/0.743 0	26.76/0.805 8

0.38 dB,SSIM 值提升 0.011 2。

同时本文也比较了在 Set5 标准测试集下,不

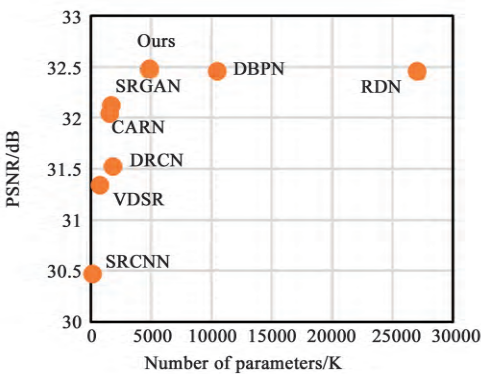


图 3 Set5 测试集下超分辨率重建方法的参数数量和 PSNR 的关系

Fig.3 Relationship between the number of parameters of the super resolution reconstruction method and PSNR under Set5 test set

同超分辨率重建方法的参数数量和 PSNR 均值的关系,对比结果参见图 3。其中 SRCNN 方法的网络结构只包含 3 层,故参数量较小,但其 PSNR 值最低。DBPN 和 RDN 方法参数量较大,本文方法的参数量只有 RDN 的大约 25%,本文使用较少的参数量获得了较好的重建效果。

3.3 RDAM 模块对网络的影响

为验证 RDAM 模块对网络性能的影响,进行了 5 组 4 倍超分辨率重建实验,分别比较了网络中采用残差双注意力模块 RDAM、仅采用残差通道注意力模块 RCAB 和 3 组空白对照实验:(1)None,不使用残差块和双注意力机制;(2)仅使用残差块,不使用双注意力机制;(3)仅使用双注意力机制,不使用残差块时,在 4 个标准测试集下 PSNR 和 SSIM 均值的变化。实验结果如表 2 所示,通过 3 组空白对照实验可以看出,网络在采用 RDAM 模块时,PSNR 和 SSIM 数值均最高,

表 2 不同残差注意力模块对重建效果的影响

Tab.2 Influence of different residual attention modules on the reconstruction effect

Method	Set5	Set14	BSD100	Urban100
	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
None	32.18/0.894 1	28.45/0.778 0	27.37/0.728 3	26.41/0.795 1
仅使用残差块	32.29/0.895 0	28.57/0.780 6	27.46/0.730 8	26.43/0.795 4
仅使用双注意力机制	32.36/0.895 8	28.64/0.781 3	27.52/0.733 6	26.54/0.799 8
RCAB ^[8]	32.47/0.898 3	28.72/0.785 7	27.63/0.735 9	26.65/0.803 4
RDAM	32.54/0.899 3	28.86/0.787 9	27.77/0.743 0	26.76/0.805 8

可以说明当残差块和双注意力机制结合时,能有效地提升网络的性能。

3.4 数据增强对网络的影响

为验证数据增强对网络的影响,进行了3组4倍超分辨率重建实验,分别比较了网络中不采用数据增强方法(None)、采用一种 CutBlur

增强方法和采用混合增强方法(MOA)时PSNR和SSIM均值的变化。从表3可以看出,在4个标准测试集中,采用混合数据增强方法(MOA)与不采用数据增强方法(None)相比,平均PSNR提升约0.15 dB,平均SSIM提升约0.003 7。

表3 不同数据增强方法对重建效果的影响

Tab.3 Influence of different data enhancement methods on the reconstruction effect

Method	Set5	Set14	BSD100	Urban100
	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
None	32.44/0.896 4	28.75/0.785 7	27.67/0.741 3	26.49/0.797 8
Ours(CutBlur)	32.48/0.898 0	28.77/0.786 2	27.69/0.741 8	26.55/0.799 9
Ours(MOA)	32.54/0.899 3	28.86/0.787 9	27.77/0.743 0	26.76/0.805 8

3.5 损失函数对网络的影响

为验证损失函数对网络性能的影响,进行了5组4倍超分辨率重建实验,分别比较了网络中采用本文的损失函数 $L_P + L_D$ 和4组对照实验:(1)仅对网络采用重建损失约束 L_P ;(2)仅对网络采用对偶回归网络损失约束 L_D ;(3)仅对网络采用MSE损失约束 $L_{PMSE} + \lambda L_{DMSE}$;(4)仅对网络采

用SSIM损失约束 $\alpha L_{PSSIM} + \beta L_{DSSIM}$ 时,在4个标准测试集下PSNR和SSIM均值的变化。实验结果参见表4,可以看出,网络采用本文的损失函数 $L_P + L_D$ 时,PSNR和SSIM数值均最高,说明对改进的U-net网络采用重建损失 L_P 和对偶回归网络损失 L_D 共同约束时,能最大化提升网络的性能。

表4 不同损失函数对重建效果的影响

Tab.4 Influence of different loss functions on the reconstruction effect

Method	Set5	Set14	BSD100	Urban100
	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
L_P	32.34/0.895 6	28.52/0.779 5	27.45/0.730 5	26.42/0.793 6
L_D	32.32/0.895 1	28.55/0.780 0	27.43/0.728 1	26.44/0.794 5
$L_{PMSE} + \lambda L_{DMSE}$	32.41/0.896 1	28.68/0.781 0	27.59/0.735 4	26.61/0.802 8
$\alpha L_{PSSIM} + \beta L_{DSSIM}$	32.43/0.897 0	28.72/0.785 6	27.63/0.740 0	26.64/0.803 3
$L_P + L_D$	32.54/0.899 3	28.86/0.787 9	27.77/0.743 0	26.76/0.805 8

3.6 主观实验结果分析

PSNR和SSIM客观评价图像质量的结果与人眼主观视觉感知不是完全一致的。为了验证本文方法的有效性,本文与其他超分辨率重建方法对比,在Set14、BSD100和Urban100测试集上选取部分图像的4倍超分辨率重建结果,参见图4。便于观察,本文将对对比图像进行了局部裁剪和放大。其中,红色方框中图像是放大区域,各重建方法标注在图像的下方。

从主观视觉效果可以看出,本文方法重建后的超分辨率图像效果优于其他的方法,更好地恢复了图像的边缘和纹理,使细节清晰可见。在图4(a)中,对于银饰和难以恢复的衣领细节,SRCNN和VDSR方法遭遇边缘失真的现象,DRN方法重建的图像较为清晰,但本文方法对于弯曲的银饰细节上恢复效果更好,观感接近原图。在图4(b)中,对于斑马纹理细节,SRCNN方法学习到的高频特征少,恢复的图像边缘较为模糊,DBPN方法重建出

的蝴蝶纹理出现了一些噪点,本文方法重建出的斑马画面更加纯净,接近原 HR 图像。在图 4(c)中,对于石块纹理的细节,SRCNN 和 VDSR 遭遇了不同程度的边缘失真,本文方法相对于 DBPN 和

RDN,能够重建更锋利的石块边缘细节,效果接近原图。在图 4(d)中,本文方法相对于其他的方法,在玻璃窗和难以恢复的树枝细节上,表现得更加清晰,边缘细节更丰富。



图 4 不同方法下超分辨率重建后的效果图比较

Fig.4 Comparison of renderings after four times super resolution reconstruction under different methods

4 结 论

由于图像超分辨率重建映射函数的解空间极大,往往导致重建模型性能有限,不能产生纹理细致、边缘清晰的图像。本文提出了一种结合双注意力和结构相似度量的图像超分辨率重建网络。该网络采用残差双注意力模块,结合数据增强方

法、自适应参数线性整流激活函数,增加均方误差和结构相似损失约束,有效提高了生成图像的质量,使生成的图像更逼近真实图像。实验结果表明,本文方法对比 SRCNN 方法,在 Set5、Set14、BSD100 和 Urban100 测试集上的平均 PSNR 提升约 1.64 dB,SSIM 提升约 0.047。本文方法能够更好地重建图像纹理和边缘细节,在客观评价和主观视觉上获得了较好的效果。

参 考 文 献:

- [1] 陈宗航,胡海龙,姚剑敏,等.基于改进生成对抗网络的单帧图像超分辨率重建[J].液晶与显示,2021,36(5): 705-712.
CHEN Z H, HU H L, YAO J M, *et al.* Single frame image super-resolution reconstruction based on improved generative adversarial network [J]. *Chinese Journal of Liquid Crystals and Displays*, 2021, 36(5): 705-712. (in Chinese)
- [2] 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.
- [3] 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*. Las Vegas; IEEE, 2016: 1646-1654.
- [4] 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 (CVPR)*. Las Vegas; IEEE, 2016: 1637-1645.
- [5] LAI W S, HUANG J B, AHUJA N, *et al.* Deep Laplacian pyramid networks for fast and accurate super-resolution [C]//2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu; IEEE, 2017:

5835-5843.

- [6] LIM B, SON S, KIM H, *et al.* Enhanced deep residual networks for single image super-resolution [C]//2017 *IEEE Conference on Computer Vision and Pattern Recognition Workshops*. Honolulu: IEEE, 2017: 1132-1140.
- [7] 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*. Salt Lake City: IEEE, 2018: 2472-2481.
- [8] ZHANG Y L, LI K P, LI K, *et al.* Image super-resolution using very deep residual channel attention networks [C]//*Proceedings of 15th European Conference on Computer Vision*. Munich: Springer, 2018: 294-310.
- [9] 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 (CVPR)*. Honolulu: IEEE, 2017: 105-114.
- [10] WANG X T, YU K, WU S X, *et al.* ESRGAN: enhanced super-resolution generative adversarial networks [C]//*Proceedings of European Conference on Computer Vision*. Munich: Springer, 2019: 63-79.
- [11] XIAO M Q, ZHENG S X, LIU C, *et al.* Invertible image rescaling [C]//16th *European Conference on Computer Vision*. Glasgow: Springer, 2020: 126-144.
- [12] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation [C]//18th *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Munich: Springer, 2015: 234-241.
- [13] YUN S, HAN D, CHUN S, *et al.* CutMix: regularization strategy to train strong classifiers with localizable features [C]//2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul: IEEE, 2019: 6022-6031.
- [14] YOO J, AHN N, SOHN K A. Rethinking data augmentation for image super-resolution: a comprehensive analysis and a new strategy [C]//2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2020: 8372-8381.
- [15] GUO Y, CHEN J, WANG J D, *et al.* Closed-loop matters: dual regression networks for single image super-resolution [C]//2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2020: 5406-5415.
- [16] 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*. Las Vegas: IEEE, 2016: 770-778.
- [17] WOO S, PARK J, LEE J Y, *et al.* CBAM: convolutional block attention module [C]//*Proceedings of 15th European Conference on Computer Vision*. Munich: Springer, 2018: 3-19.
- [18] CHEN Y P, DAI X Y, LIU M C, *et al.* Dynamic ReLU [EB/OL]. [2020-05-12]. <https://arxiv.org/abs/2003.10027>.
- [19] ANWAR S, KHAN S, BARNES N. A deep journey into super-resolution: a survey [J]. *ACM Computing Surveys*, 2020, 53(3): 60.
- [20] AHN N, KANG B, SOHN K A. Fast, accurate, and lightweight super-resolution with cascading residual network [C]//*Proceedings of 15th European Conference on Computer Vision*. Munich: Springer, 2018: 256-272.
- [21] HARIS M, SHAKHNAROVICH G, UKITA N. Deep back-projection networks for super-resolution [C]//2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018: 1664-1673.
- [22] LI Z, YANG J L, LIU Z, *et al.* Feedback network for image super-resolution [C]//2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019: 3862-3871.

作者简介:



黄友文(1982—),江西赣州人,博士,副教授,2009年于同济大学获得博士学位,研究方向为智能信息处理、嵌入式应用和图像处理。E-mail: ywhuang@jxust.edu.cn



唐欣(1996—)女,湖北黄石人,硕士研究生,2018年于武汉工程大学邮电与信息工程学院获得学士学位,研究方向为图像处理。E-mail: tangxin.2108@163.com