

文章编号: 2095-2163(2023)01-0012-07

中图分类号: TP311

文献标志码: A

GCA: 一种结合注意力机制的图像超分辨率重建模型

彭学桂

(上海理工大学 光电信息与计算机工程学院, 上海 200093)

摘要: 深度学习在图像超分辨率任务上的重建精度和计算性能均表现优异。通常的方法是将高分辨率图片通过简单的插值法进行下采样得到低分辨率图片, 然后将其输入到超分网络中生成高清图片。然而, 真实世界的低质量图片往往存在噪声, 且其模糊状态按一定分布呈现, 直接将通过插值得到的低分辨率图片来训练的模型在真实世界图片上的效果并不好。针对这一问题, 本文在数据集的构造上采用了一种“降级”结构, 通过退化算法估计各种模糊核以及真实的噪声分布, 使得作为网络输入的低分辨率图片与真实图像共享一个公共域; 提出一种结合注意力机制的 GCA 模型, 将生成模块得到的超分辨率图片与标签图像共同输入判别模块的胶囊网络中进行二分类, 最终达到判别模块无法区分超分辨率图片和对应的高分辨率图片。在 City100 数据集上的实验表明, 本文提出的 GCA 模型在真实世界图像上取得了更好的效果。

关键词: 图像超分辨率; 胶囊网络; 注意力机制; 模糊核

GCA: An image super-resolution reconstruction model that combines attention mechanisms

PENG Xuegui

(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

[Abstract] Deep learning has performed well in reconstruction accuracy and computational performance on image hyper-resolution tasks. The usual approach is to take a low-resolution picture by sampling it by simple interpolation and then inputting it into a super-divided network to generate a high-definition picture. However, real-world low-quality pictures tend to be noisy, and their fuzzy state is presented in a certain distribution, and models trained directly by interpolation of low-resolution images do not work well in real-world images. In response to this problem, this paper adopts a "degraded" structure in the construction of the data set. The degradation algorithm is used to estimate various fuzzy cores and real noise distribution, so that low-resolution pictures as network inputs share a common domain with real images. At the same time, a GCA model combined with an attention mechanism is proposed, which combines the super-resolution image obtained by the generating module with the label image into the Capsule network of the distinguishing module for two classifications, and finally achieves that the distinguishing module cannot distinguish between the super-resolution picture and the corresponding high-resolution picture. Experiments on the City100 data set show that the GCA model proposed in this paper has achieved better results in real-world images.

[Key words] image super-resolution; capsule network; attention mechanism; fuzzy core

0 引言

单幅图像的超分辨重建 (Single Image Super Resolution, SISR) 是计算机视觉中的子任务, 目标是从一幅低分辨率 (Low Resolution, LR) 图像中恢复出高清图像 (High Resolution, HR)。随着图像处理技术的日趋成熟, 超分辨率重建作为一种提高图像分辨率的方法, 其应用领域也越来越广, 已被用于人脸识别、卫星遥感图像、银行安全监控系统等领域。

深度卷积神经网络 (CNN) 在图像处理方面效果显著。Dong^[1] 等人在 2014 年提出 SRCNN, 第一次将卷积网络应用于超分辨率 (Super Resolution, SR) 重建; Lan^[2] 等人为了提高网络的计算效率, 设计了一种密集型的轻量级网络, 可以用于更强的多尺度特征表达和特征相关学习; Shi^[3] 等人提出的高效率亚像素卷积模型 (Efficient Sub-Pixel Convolution Networks, ESPCN), 模型引入亚像素卷积, 是一种高效、快速、无参的像素重排列上采样方

作者简介: 彭学桂 (1997-), 男, 硕士研究生; 主要研究方向: 计算机视觉与图像。

收稿日期: 2022-03-30

式;Hui^[4]等人提出了一种新的感知图像超分辨率方法,通过构建一个阶段式网络,逐步产生视觉上高质量的结果;Tong^[5]等人构建了 SRDenseNet,在稠密块(Dense block)中将每一层的特征都输入给之后的所有层,使所有层的特征都串联起来,而不是像 ResNet 那样直接相加。这样的结构给整个网络带来了减轻梯度消失、加强特征传播、支持特征复用、减少参数数量的优点;Zhang^[6]等人设计了一种无监督的多循环模型结构,以生成对抗网络(Generative Adversarial Networks, GAN)作为基础单元,将有噪声和模糊的低分辨率向量映射到一个无噪声的低分辨率向量空间;Wang^[7]等人在超分辨率重建生成对抗模型(Super Resolution Generative Adversarial Networks, SRGAN)的基础上,去除网络中所有的正则化层,并将残差单元由串联变为密集连接,得到增强的超分辨率重建生成对抗模型。

近年来诞生了各种各样的超分辨率网络,但超分辨率重建本身依旧存在很多问题。首先,超分辨率重建是一个病态问题,对于同一个低分辨率图像,其对应的高分辨率图像的解并不唯一,即存在不稳定性,为了约束解空间,通常需要可靠的先验信息;其次,随着重建图片的尺寸扩大,问题的复杂性也会增加。在更高的因子下,恢复丢失的细节信息会更加复杂,经常导致错误信息的再现。此外,网络输出图片质量的评估并不简单,量化指标与人的感知只有松散的关联。最关键的,这些网络大多数采用简单的双三次下采样,从高质量图像构建低分辨率和高分辨率对用于训练,可能会丢失与频率相关的细节,在现实世界的图像超分辨率中的效果并不好。

针对上述问题,本文提出一种 GCA 网络。首先通过一个图像退化框架估计各种模糊核以及真实的噪声分布,可以获得与真实世界图像共享一个公共域的图像,使网络输入的低分辨率图片与高分辨率图片处在一个域中,提高了模型在现实世界图片上测试的效果;其次,生成模块使用残差注意力稠密块(Residual in Residual Dense, RRDB)和残差块(Residual Network, ResNet)对图片高低频信息进行融合,为了使网络在各层次特征信息融合的过程中更专注于图片的高层次信息,采用注意力机制,将 RRDB 模块输出后的特征向量作为高层次特征信息的参数,再使用 sub-pixel 网络上采样得到超分辨率图片;在判别模块中使用胶囊网络(Capsule Net)对生成模块输出的超分辨率图片及真实的高分辨率图片进行判别。

1 相关工作

早期的超分辨率重建技术可以分为3类:基于插值,基于重建,基于学习的方法。基于插值的方法在传统方法中相对来说实现简单,且应用广泛,但是这些线性的模型不利于恢复高频细节信息;基于学习的稀疏表示技术通过使用先验知识增强了线性模型的能力,假设任意的自然图像可以被字典的元素稀疏表示,需要大量计算资源;基于重建的方法通常用于多帧图像的超分辨率重建,该技术需要结合先验知识。

单幅图片超分辨率重建是一个逆问题,对于一个低分辨率图像,可能存在许多不同的高分辨率图像与之对应,因此通常在求解高分辨率图像时会加一个先验信息进行规范化约束。在传统的方法中,这个先验信息可以通过若干成对出现的低-高分辨率图像的实例中学到。而基于深度学习的超分辨率重建(Super Resolution, SR)通过神经网络直接学习分辨率图像到高分辨率图像的端到端的映射函数。

基于卷积神经网络(Convolutional Neural Networks, CNN)解决单图像超分辨率重建的初步方法是通过3层的卷积网络对输入图片做卷积处理。Singh^[8]等人提出将图像超分辨率问题分为多个子问题,然后利用神经网络来处理的方法。为了解决信息流随网络深度而削弱的问题, Xu^[9]等人根据循环神经网络(Recurrent Neural Network, RNN)的原理设计了一个轻量级的基于反馈的递归神经网络(Feedback Recurrent Neural Network FRNN); Li^[10]等人根据该思路,提出了基于反馈连接和类循环神经网络结构的 SRFBN 反馈网络; Zhang^[11]等人采用了一种基于深度卷积神经网络的多层降级超分辨率模型(Super Resolution Multiple Degradations, SRMD); Fritsche^[12]等人提出了通过无监督方式训练网络来学习退化率的模型(Distant Supervision GAN, DSGAN)。

目前,注意力机制已经逐渐运用在图像及其它一些任务上,取得了显著的效果。Yang^[13]等人提出将残差网络和稠密跳跃连接结合起来; Huang^[14]等设计了一个双路径的注意力网络,将残差连接和稠密连接组合,以此来进行不同层次特征信息的互补; Fu^[15]等人设计了一个尺度注意力模块,通过引入尺度因子作为先验知识来学习低分辨率图像的判别特征,利用坐标信息和比例因子的二次多项式预测像素级重构核,实现任意比例因子的超分辨率。

2 问题定义

图像超分辨率的重建目的是从 LR 图像中恢复出相应的 HR 图像。为了得到相对应的成对数据集,通常的做法是将高分辨率图片降级,以此来获得低分辨率图片。一般来说, LR 图像 I_{LR} 建模为退化的输出,式(1):

$$I_{LR} = D(I_{HR}; \delta) \quad (1)$$

其中, D 为退化映射函数; I_{HR} 为对应的 HR 图像; δ 为退化函数中的超参数。

往往降级退化过程是未知的,只是得到 LR 图片。需要从 LR 图片中恢复和真实标签样本 I_{HR} 一致的超分辨率图片 \tilde{I}_{SR} , 式(2):

$$\tilde{I}_{SR} = F(I_{LR}; \theta) \quad (2)$$

其中, F 是超分辨率模型, θ 代表整个模型 F 的超参数。

虽然退化过程是未知的,并且可能受到各种因素的影响,如:压缩伪像、各向异性退化、传感器噪声

和散斑噪声,但是随着不断地尝试对退化映射建模,现在大多数工作将降级直接建模为单个下采样操作,式(3):

$$D(I_{HR}; \delta) = (I_{HR}) \downarrow_s, \{s\} \subset \delta \quad (3)$$

其中, \downarrow_s 为下采样操作, s 为比例因子。

事实上,大多数的超分辨率数据集构造都是通过这种方式,下采样一般采用双三次插值。

本文采用一种新的退化映射函数,式(4):

$$D(I_{HR}; \delta) = (I_{HR} * k) \downarrow_s + n_\varepsilon, \{k, s, \varepsilon\} \subset \delta \quad (4)$$

其中, $I_{HR} * k$ 表示模糊核 k 和 HR 图像 I_{HR} 之间的卷积, n_ε 是以 ε 为标准差的噪声。

实验表明,该降级函数得到的 LR 图片比简单的插值下采样更接近真实世界的图片。

3 模型

GCA 模型的整体结构如图 1 所示。模型分为 3 部分,分别为预处理模块(Preprocessed Module)、生成器模块(Generator Module)以及判别器模块(Discriminator Module)。

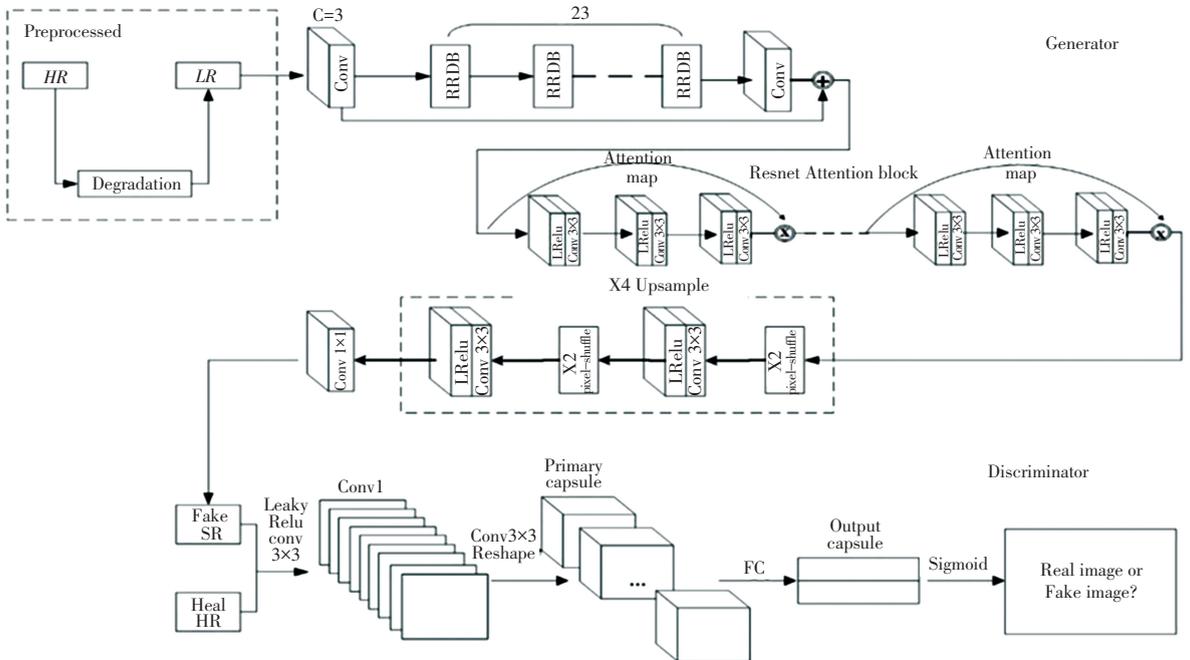


图 1 GCA 的整体模型

Fig. 1 The GCA overall model

3.1 数据集预处理

本文使用一种基于核估计和噪声注入的真实图像退化方法。假设 LR 图像通过式(5)退化方法获得。

$$I_{LR} = (I_{HR} * k) \downarrow_s + n \quad (5)$$

其中, k 和 n 分别表示模糊核和噪声。

为了更准确地模拟退化方法,本文从真实世界图像中明确地估计核和噪声,在获得估计的核和噪声块后,建立一个退化池,用于将高分辨率图像退化为符合真实世界模糊和噪声分布的低分辨率图片,从而生成用于训练 GCA 网络的图像对,该降级过程

简单描述见表 1。

表 1 数据集预处理算法

Tab. 1 Data set preprocessing algorithm

输入: 真实图像集 X , 超分辨率图像集 Y , 下采样因子 s

输出: 真实世界的高低分辨率图像对: $\{I_{LR}, I_{HR}\}$

- 1: 随机初始化模糊核池 $K = \emptyset$
- 2: 随机初始化噪声池 $N = \emptyset$
- 3: 对于所有的 I_{src} 按照 $I_{src} \in X$ 处理
- 4: 估计 k 从 I_{src} 通过公式 4
- 5: 加入 k 到 K
- 6: 裁剪 n 从 I_{src}
- 7: 如果 n 满足公式 7 则
- 8: 加入 n 到 N
- 9: 结束
- 10: 结束
- 11: 对于所有的 I_{HR} 按照 $I_{HR} \in Y$ 处理
- 12: 随机选择 $k_i \in K, n_j \in N$
- 13: 生成 I_{LR} 通过 k_i 和 n_j
- 14: 结束
- 15: 返回 $\{I_{LR}, I_{HR}\}$

本文采用核估计算法从真实图片中明确地估计核, 估计的模糊核应该满足式(6)约束:

$$\arg \min_k \left\| (I_{src} * k) \downarrow_s - I_{src} \downarrow_s \right\|_1 + \left| 1 - \sum k_{i,j} \right| + \left| \sum k_{i,j} m_{i,j} \right| + \left| 1 - D((I_{src} * k) \downarrow_s) \right| \quad (6)$$

其中, $(I_{src} * k) \downarrow_s$ 是核函数为 k 的下采样 LR 图片, $I_{src} \downarrow_s$ 是理想核函数下采样后的图片。

因此最小化二者之差就是期望下采样的低分辨率图片保留原更多的低频信息, 式(6)中的第二项是为了约束 k 的总和为 1, 第三项是 k 的惩罚边界, 后面的判别模块是为了保证生成的图片和原图片尽可能相似, 也就是处于同一个域, 更符合真实世界的图像。

通过以上约束获得模糊核后, 将所有的模糊核放入退化池中, 然后从退化池中随机选择一个模糊

核来对 HR 图片进行降级退化处理, 该过程可以表述如式(7):

$$I_D = (I_{HR} * k_i) \downarrow_s, i \in \{1, 2, \dots, m\} \quad (7)$$

其中, I_D 代表下采样后的图片, k_i 代表从 $\{k_1, k_2, \dots, k_m\}$ 中随机选定的模糊核。

由于真实世界图片并不是纯净的, 往往受到环境的影响而存在一定噪声。为了更好地拟合真实图片, 需要在经过模糊核下采样后的图片中加入噪声。本文直接从原数据集 X 的图片中采集噪声块, 一般情况下, 内容越丰富, 其像素值的方差是越大的, 因此, 本文设计了一个筛选规则来收集方差在一定范围内的块, 简化如式(8):

$$\sigma(n_i) < v \quad (8)$$

其中, $\sigma(\cdot)$ 表示计算方差的函数, v 是方差的最大值。

获得了一系列的噪声块后, 将所有的噪声块放入退化池, 从中随机裁剪噪声块加入到下采样后的图片中, 以此来完成噪声注入, 表示为式(9):

$$I_{LR} = I_D + n_i, i \in \{1, 2, \dots, l\} \quad (9)$$

其中, n_i 是从退化池随机裁剪的噪声块。

经过模糊核和噪声注入退化后的 I_{LR} 更加符合真实世界的图像。

3.2 生成模块

生成模块本质上就是一个生成函数 G , 对于给定的 LR 图片, 可以输出对应的 HR 图片。本文采用 RRDB 块和 Resnet 块作为生成网络的基础块, 注意力机制的原理便是将前面部分的特征向量作为当前层高层次信息的参数。假设数据集 Preprocessed 中下采样的比例因子为 r , 用一个大小为 $W \times H \times C$ 的张量来表示 I_{LR} 图片, $rW \times rH \times C$ 的张量则表示 I_{LR} 图片和 I_{HR} 。生成模块作为一个前馈网络 G_{θ_C}, θ_C 代表网络中的超参数, $\theta_C = \{W_{1:L}; b_{1:L}\}$, 通过超分辨率网络中特定的损失函数 L^{SR} 得到。训练时, 数据集集中的高分辨率图片表示为 $I_{HR}^n (n = 1, \dots, N)$, 低分辨率图片表示为 $I_{LR}^n (n = 1, \dots, N)$ 。

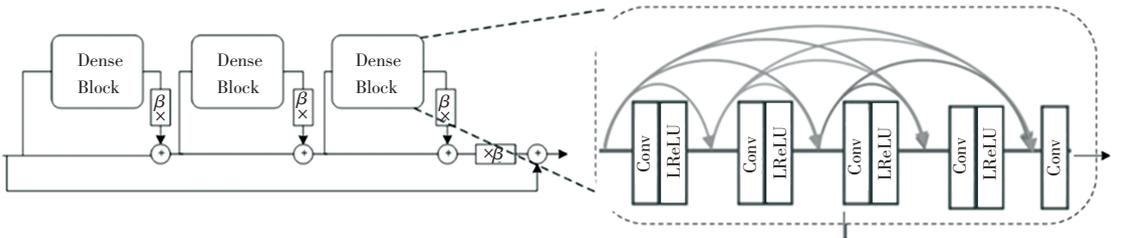


图 2 RRDB 块结构

Fig. 2 RRDB block structure

作为生成模块的基本块 RRDB,其结构如图 2 所示。首先将 Preprocessed 过程后得到的 LR 图片经过 9×9 卷积对输入图片进行一个全局信息提取,再输入到 RRDB 块中对信息进一步整合,具体来说,稠密块(Dense Block)的每一层使用 64 个步长为 1 的 3×3 卷积核,每层的输出采用 LRelu 激活函数,其与 Relu 函数的不同之处在于将小于 0 的数值进行保留,这对于超分辨率这样对细节信息要求较高的任务非常适合。RRDB 的每一个 Dense Block 之后会有一个残差缩放因子 β ,即在将残差添加到主路径之前,通过乘以 0 和 1 之间的常数来缩小残差,因为随着网络层数的加深,会存在一定的冗余信息,越靠近输入的层其所含信息越丰富,残差因子相对更大,这也是一种自注意力机制。

RRDB 块之后,使用 Resnet 块来整合网络中各层的信息,同时加速网络,保证了梯度信息能够有效的传递。损失函数从判别模块开始反向传播至生成模块时,经过了很多层,越深的网络隐藏参数越多,在反向传播的过程中也越容易梯度弥散,而加入 Resnet 块可避免这种情况。Resnet 块共有 3 个,每个包含 3 层卷积神经网络,每层网络由 64 个 3×3 的卷积核构成,在卷积核前面加了一个 LRelu 预激活函数。

为了使生成模块输出的图片和 Preprocessed 中的 HR 图片能输入到判别模块进行分类,需要将 Resnet 块后的图片进行上采样,使之与 HR 图片同等尺寸。此处本文使用像素混洗(Pixel-Shuffle)对图片进行放大,其过程可以表示为式(10):

$$I_{SR} = f^L(I_{hid}) = PS(W_L * f^{L-1}(I_{hid}) + b_L) \quad (10)$$

其中, I_{hid} 表示处于低维空间的图片, PS 是像素混洗算子,其将 $W \times H \times C \times r^2$ 的张量重新排列成大小为 $rW \times rH \times C$ 的张量。

从数学角度来讲,该算子的像素重排列可以表示为式(11):

$$PS(T)_{x,y,c} = T_{\lfloor \frac{x}{r} \rfloor, \lfloor \frac{y}{r} \rfloor, C \cdot r \cdot \text{mod}(y,r) + C \cdot \text{mod}(x,r) + c} \quad (11)$$

其中, x, y, c 分别为尺寸大小以及通道数; mod 为模运算符; C 为比例因子; T 则为组合排列函数。

假设输入是一个三维的张量 $x \times y \times c$, 通过 PS 周期为 T 的运算,其可以变成其它尺寸的张量。

在 Pixel-Shuffle 层前设置 3×3 的卷积层,经过卷积得到通道数为 r^2 与输入图像大小一样的特征图像;再将特征图像每个像素的 r^2 个通道重新排列成一个 $r \times r$ 的区域,对应高分辨率图像中一个 $r \times r$ 大小的子块,从而大小为 $W \times H \times r^2$ 的特征图像被

重新排列成 $rW \times rH \times 1$ 的高分辨率图片。另外, Pixel-Shuffle 只在最后一层对图像大小做变换,前面的卷积运算由于在低分辨率图像上进行,因此效率会较高。

3.3 判别模块

判别模块是由胶囊网络组成的一个二分类器,将生成模块输出的 I_{SR} 作为虚假图片,数据集预处理过程的 I_{HR} 作为真实图片,判别模块的目的是尽可能地区分出 I_{SR} 和 I_{HR} 。Discriminator 只训练判别模块的参数;Generator 的训练是把生成模块和判别模块两个网络连在一起,冻结判别模块的学习率,以此来对两个模块进行交替训练,最终使得 Discriminator 中的分类器无法区分 I_{SR} 和 I_{HR} 。生成模块输出的 fake 图片表示为 $I_{SR}^n (n = 1, \dots, N)$, 生成模块和判别模块的对抗训练定义如式(12):

$$\min_{\theta_G} \max_{\theta_D} E_{I_{HR} \sim P_{\text{train}}(I_{HR})} [\log D_{\theta_D}(I_{HR})] + E_{I_{LR} \sim P_G(I_{LR})} [\log(1 - D_{\theta_D}(G_G(I_{LR})))] \quad (12)$$

其中, G_G 代表生成器; D_{θ_D} 代表判别器; E 则代表训练过程中的比例因子。

用判别模块训练来实现 \max_{θ_D} , 其由胶囊网络组成,分别是 Conv1 层、PrimaryCaps 层、DigitCaps 层。为了实现 I_{SR} 和 I_{HR} 的 0-1 二分类,本文将 DigitCaps 层改为两个胶囊,最后通过计算胶囊向量的模块作为分类的概率值。由于胶囊网络中不存在池化层,对细节信息的捕获非常敏感,且在 Mnist 数据集上对数字图片分类的结果较好,非常适合图片分类任务。

和 CNN 中的标量神经元不一样,胶囊网络中的神经元是向量,胶囊网络通过动态路由算法将低级胶囊的预测结果传输到高级胶囊。Conv1 层有 256 个 9×9 卷积核, primary 层是一个卷积胶囊层,通过 6×6 的卷积核和 reshape 操作得到,最后通过全连接的矩阵转换得到两个胶囊,分别对应 0-1 分类。

3.4 损失函数

将均方误差作为损失函数,因为图像超分辨率重建的常用评价指标 PSNR 与之存在一定的数值运算关系,可以得到更高的 PSNR 值。学习端到端的重建映射函数 f^L 需要计算网络参数,用 θ 代表,即式(13):

$$L(\theta) = \frac{1}{n} \sum_{i=1}^n \|f(y_i; \theta) - x_i\|^2 \quad (13)$$

其中, y_i 和 x_i 为高低分辨率图像, n 为训练样本的数量。

4 实验

为了验证模型的有效性,做了一定的对比实验和消融实验,还使用 Set5、Set14、BSD100 公共基准数据集进行测试实验。

4.1 数据集介绍

Set5 是一个经典数据集,只包含婴儿、鸟、蝴蝶、头和人的 5 幅测试图像。

Set14 公开数据集包含更多类别,但图像数量依然很少,只有 14 张图像。

BSD100 是一个具有 100 张测试图像的经典数据集,从自然图像到特定对象,如植物、人、食物等。

City100 包括 City100 NikonD5500 和 City100 iPhoneX,分别表示单反相机和智能手机相机下的分辨率和视场退化,以 100 张不同城市场景的明信片作为成像对象。HR 和 LR 图像分别在 55 mm 和 18 mm 的焦距下拍摄。

4.2 实验细节

通过数据集预处理,得到相对应的高低分辨率图像对,通过随机截取低分辨率图像的 48×48 ,以及高分辨率图像对应位置的 192×192 ,并分别将其旋转与翻转,进行数据增强处理,以此得到更多的训练数据集。采用 Adam 优化器来优化模型,动量设置为 0.9,权重衰减系数为 0.000 1,共训练代数 200,此外采用多步学习策略,初始学习率为 0.01,在适当的迭代次数后,以一定比例降低学习率。

所有实验均在 windows10 系统搭载的 12 GB NVIDIA GTX 1080 Ti GPU 和 3.6 GHz Intel® i7CPU 上完成的,在 Pytorch-GPU 1.13.1 环境下使用 python3.7 实现了模型。

4.3 评价指标

PSNR:峰值信噪比是一个表示信号最大可能功

率和影响其表示精度的破坏性噪声功率的比值,经常作为图像重建等领域中重建质量好坏的评价标准,其值越大,表明图像质量越好;

SSIM:越大表示输出图像和无失真图像的差距越小,即图像质量越好。

4.4 对比模型

为了进行直观的对比,将本文所提模型的 SR 结果与 SRCNN、ESRGAN、DPAN、WDN 等现有模型进行对比,这些模型在超分任务中性能较好,具有代表性,且与本文提出的算法具有一定的相关性。SRCNN 模型首次将卷积神经网络应用于图像超分辨率任务;ESRGAN 在 SRGAN 的基础上去除所有的 BN 层。DPAN^[14]则构建了一个双路径的注意力模型;WDN^[8]是最近提出的一种将超分辨率任务分而治之的算法,通过多路径的模型将超分辨率重建任务分为多个子任务,并行计算从而大大提高了模型的效率。

表 2 对比实验结果

Tab. 2 Compare the results of the experiment

Method	SRCNN	ESRGAN	DPAN	WDN	Ours
PSNR/db	19.74	24.86	27.84	27.37	28.53
SSIM	0.42	0.66	0.725	0.728	0.869

基线模型的 PSNR 和 SSIM 指标在公共数据集上的实验结果见表 2,本文提出的模型分别为 28.53 和 0.869,比 WDN 要更好。和 SRCNN、ESRGAN、DPAN、WDN 比较,在两个评价指标上,本文的方法表现出了一定的优势,这说明本文提出的 GCA 模型在图像超分辨率重建上效果良好,直观效果如图 3 所示。为了证明本文所提算法中各个模块的有效性,将进行消融实验,结果见表 3。



图 3 模型效果可视化

Fig. 3 Model effect visualization

表3 消融实验结果

Tab. 3 Ablation experiment

Method	Generator (only)	Generator+ Preprocessed	Generator+ Preprocessed+ Attention map	GCA
PSNR/db	25.7	26.34	27.1	28.46
SSIM	0.62	0.691	0.75	0.83

表4 Set5、Set14、BSD100 数据集上的对比实验

Tab. 4 Comparison experiments on Set5, Set14, BSD100 datasets

	Set5		Set14		BSD100	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
SRCNN	28.43	0.840 1	26.10	0.702 5	25.96	0.667 5
ESRGAN	31.56	0.898 0	28.87	0.786 0	27.64	0.741 3
DPAN	32.53	0.899 2	29.34	0.789 0	27.77	0.743 6
WDN	32.77	0.904 2	29.56	0.790 3	27.83	0.746 4
Ours	33.20	0.909 1	29.97	0.798 1	27.92	0.751 9

将 GCA 算法中的 Generator 作为基线模型, City100 数据集中包含高低分辨率图像对,进行消融实验时,各个模块之间可以独立存在,互不影响,这也保证了消融实验的可靠性。另外,从实验结果上来看也证明了 Preprocessed 模块在构造低分辨率图片上有效果。表3中的实验数据表明 GCA 算法中的各个模块在超分辨率任务中是有效的。在 Set5、Set14 和 BSD100 公共数据集上的实验结果见表4,通过与基线模型的对比可以看出本文提出的 GCA 算法在 PSNR 和 SSIM 指标上都是要更好的。

5 结束语

本文提出 GCA 算法来提高真实世界图片的超分辨率重建效果,数据集的预处理上采用核估计和噪声注入,目的是为了获取处于同一域的高分辨率和低分辨率图片对,生成模块利用 RRDB 块和 ResNet 块促进了模型中不同层之间的信息流动和特征重用,判别模块则利用胶囊网络对超分辨率图片和高清图进行判别分类,通过生成模块和判别模块二者之间的交替博弈来进行训练。在 City100 数据集上的实验表明,本文采用的模型在真实世界图片上的效果优异,与其它方法相比,在 PSNR 和 SSIM 评价指标上,均取得较好的效果。目前的模型主要针对的是较低倍数的超分辨率重建图片,在未来的工作中,将致力于输出更高倍数图片的研究。

参考文献

- [1] DONG C, LOY C C, HE K, et al. Learning a deep convolutional network for image super-resolution[C]//European conference on computer vision. Springer, Cham, 2014: 184-199.
- [2] LAN R, SUN L, LIU Z, et al. MADNet: a fast and lightweight network for single-image super resolution[J]. IEEE transactions on cybernetics, 2020, 51(3): 1443-1453.
- [3] SHI W, CABALLERO J, HUSZÁR F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 1874-1883.
- [4] HUI, ZHENG, et al. Progressive perception-oriented network for single image super-resolution[J]. Information Sciences, 2021, 546: 769-786.
- [5] TONG T, LI G, LIU X, et al. Image super-resolution using dense skip connections [C]//Proceedings of the IEEE international conference on computer vision. 2017: 4799-4807.
- [6] ZHANG Y, LIU S, DONG C, et al. Multiple cycle-in-cycle generative adversarial networks for unsupervised image super-resolution[J]. IEEE transactions on Image Processing, 2019, 29: 1101-1112.
- [7] WANG X, YU K, WU S, et al. ESRGAN: Enhanced super-resolution generative adversarial networks [C]// Proceedings of the European conference on computer vision (ECCV) workshops. 2018: 63-79.
- [8] SINGH V, MITTAL A. WDN: A Wide and Deep Network to Divide-and-Conquer Image Super-Resolution[J]. IEEE Journal of Selected Topics in Signal Processing, 2020, 15(2): 264-278.
- [9] XU W, SONG H, ZHANG K, et al. Learning lightweight Multi-Scale Feedback Residual network for single image super-resolution [J]. Computer Vision and Image Understanding, 2020, 197-198: 103005.
- [10] LI Z, YANG J, LIU Z, et al. Feedback network for image super-resolution [C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 3867-3876.
- [11] ZHANG K, ZUO W, ZHANG L. Learning a single convolutional super-resolution network for multiple degradations [C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 3262-3271.
- [12] FRITSCHÉ M, GU S, TIMOFTE R. Frequency separation for real-world super-resolution [C]//2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). IEEE, 2019: 3599-3608.
- [13] YANG Y, LIU Q, ZHANG M, et al. Bi-path network coupling for single image super-resolution [J]. Multimedia Tools and Applications, 2019, 78(15): 21981-21998.
- [14] HUANG Z, LI W, LI J, et al. Dual-path attention network for single image super-resolution [J]. Expert Systems with Applications, 2020, 169(1): 114450.
- [15] FU Y, CHEN J, ZHANG T, et al. Residual scale attention network for arbitrary scale image super-resolution [J]. Neurocomputing, 2021, 427: 201-211.