

使用 VGG 能量损失的单图像超分辨率重建*

丁玲^{1,2}, 丁世飞¹, 张健¹, 张子晨¹

¹(中国矿业大学 计算机科学与技术学院, 江苏 徐州 221116)

²(中国矿业大学 徐海学院, 江苏 徐州 221008)

通讯作者: 丁世飞, E-mail: dingsf@cumt.edu.cn; 张健, E-mail: 597409675@qq.com



摘要: 单幅图像的超分辨率重建(single image super-resolution, 简称 SR)是一项重要的图像合成任务。目前,在基于神经网络的 SR 任务中,常用的损失函数包括基于内容的重构损失和基于生成对抗网络(generative adversarial network, 简称 GAN)的对抗损失。但是,基于传统的 GAN 的超分辨率重建模型(SRGAN)在判别器接收高分辨率图像作为输入时,输出判别信号不稳定。为了缓解这个问题,在 SRGAN 以及常用的 VGG 重构损失框架上,设计了一个稳定的基于能量的辅助对抗损失,称为 VGG 能量损失。该能量损失使用重构损失中的 VGG 编码部分,针对 VGG 编码设计相应的解码器,构建一个 U-Net 自编码结构 VGG-UAE,利用 VGG-UAE 的重构损失表示能量,并使用该能量函数为生成器提供梯度;基于追踪能量函数的思想,VGG-UAE 使生成器生成的高分辨率样本追踪真实数据的能量流。实验部分验证了使用 VGG 能量损失将比使用传统的 GAN 损失可以生成更有效的高分辨率图像。

关键词: 单图像超分辨率重建;生成对抗网络;能量函数;自动编码器

中图法分类号: TP391

中文引用格式:丁玲,丁世飞,张健,张子晨.使用 VGG 能量损失的单图像超分辨率重建.软件学报,2021,32(11):3659-3668.
http://www.jos.org.cn/1000-9825/6053.htm

英文引用格式: Ding L, Ding SF, Zhang J, Zhang ZC. Single image super-resolution reconstruction based on VGG energy loss. Ruan Jian Xue Bao/Journal of Software, 2021,32(11):3659-3668 (in Chinese). http://www.jos.org.cn/1000-9825/6053.htm

Single Image Super-Resolution Reconstruction Based on VGG Energy Loss

DING Ling^{1,2}, DING Shi-Fei¹, ZHANG Jian¹, ZHANG Zi-Chen¹

¹(School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China)

²(Xuhai College, China University of Mining and Technology, Xuzhou 221008, China)

Abstract: Single image super-resolution (SR) is an important task in image synthesis. Based on neural nets, the loss function in the SR task commonly contains a content-based reconstruction loss and a generative adversarial network (GAN) based regularization loss. However, due to the instability of GAN training, the generated discriminative signal of a high-resolution image from the GAN loss is not stable in the SRGAN model. In order to alleviate this problem, based on the commonly used VGG reconstruction loss, this study designs a stable energy-based regularization loss, which is called VGG energy loss. The proposed VGG energy loss in this study uses the VGG encoder in the reconstruction loss as an encoder, and designs the corresponding decoder to build a VGG-U-Net auto encoder: VGG-UAE; by using the VGG-UAE as the energy function, which can provide gradients for the generator, the generated high-resolution samples track the energy flow of real data. Experiments verify that a generative model using the proposed VGG energy loss can generate more effective high-resolution images.

Key words: single image super-resolution reconstruction; generative adversarial net; energy function; auto encoder

随着深度神经网络的研究,许多计算机视觉任务取得了突破性的进展^[1-5]。其中,单幅图像的超分辨率重建

* 基金项目: 国家自然科学基金(61976216, 61672522)

Foundation item: National Natural Science Foundation of China (61976216, 61672522)

收稿时间: 2020-01-15; 修改时间: 2020-03-09; 采用时间: 2020-04-07

研究在神经网络的支持下日趋完善^[6-8].SR 以单幅低分辨率图像为输入,输出与之对应的清晰的高分辨率图像.SR 任务有着广泛的应用,例如辅助数码变焦、视频中单帧图像的高分辨率复现、行人身份识别等.如何以低分辨率图像为基础生成清晰稳定的高分辨率图像,一直以来都是研究的热点问题^[9-13].

早期,大多关于 SR 任务的模型是基于插值方法实现的,其中最具有代表性的是基于稀疏表示的模型^[14-16].这类模型假设任意的自然图像可以由图像字典中的元素稀疏表示,那么模型可以根据图像字典重建高分辨率图像.但是这类方法计算复杂,需要大量计算资源,且此类方法在恢复图像的细节方面表现不佳.随着深度学习的发展,神经网络被引入到 SR 任务中.基于神经网络的 SR 任务都是以监督学习的方式实现的,从神经网络的角度出发,需要建立由低分辨率图像到高分辨率图像的像素级别映射^[7].从统计学的角度看,这个过程可以认为是建立条件概率 $p(y|x)$,其中, x 是输入的低分辨率图像, y 是与之对应的高分辨率图像.通过训练,使神经网络学习到低分辨率图像的统计特征,并据此恢复高分辨率图像,即由训练数据集泛化到测试数据集^[8-10].基于深度神经网络的图像超分辨率重建大致可以分为两个研究方向.

- 一部分研究工作着眼于如何构造基于内容的损失函数、并优化神经网络的结构,其中,损失函数往往选为均方误差损失或者 $L1$ 范数等变体^[8],研究的重点集中于网络模型的设计以及损失函数的优化,例如使用 VGG 网络提取的特征构建基于内容的损失函数,引入 ResNet、DenseNet 以及 RNN 网络优化模型的生成器等^[9,10,13].为了有效地构建基于内容的损失函数,学者们在 $L1$ 和 $L2$ 范数的基础上,使用训练好的深度神经网络结构提取深度特征,利用该特征判断生成的高分辨率图像与原高分辨率图像之间的差异,并把该差异作为损失函数,训练生成模型^[13],常用于 SR 任务中的神经网络为 VGG-16.该网络在 ImageNet 数据集上进行训练后,被用于高分辨率图像的特征提取过程,通常把基于 VGG 特征构建的内容损失称为 VGG 重构损失.这部分研究工作的目标是使构建的超分辨率图像在整体上具有更优的评价指标,但是这些超分辨率图像往往在高频细节上的表现不理想^[11].
- 随着 GAN 的发展^[17-19],学者们除了着眼于优化神经网络结构从而构建更有效的生成网络和内容损失之外,另一方面的研究集中于构建损失函数中的正则化项,从而生成高频细节更加丰富的超分辨率图像(尽管这类超分辨率图像在峰值信噪比等指标上表现并不理想,但可以还原丰富的图像细节).在这部分工作中,学者们将 GAN 引入到 SR 任务,使用对抗损失作为辅助损失优化损失函数,从而获得更加清晰锐利的生成图像^[11,12].然而,由于 GAN 训练的不稳定性,尽管生成的图像总体上更加锐利,但是会导致输入为高分辨率图像时判别器产生的判别信号不稳定,从而导致神经网络生成的部分图像块模糊或出现锐利的噪声块^[11].本文认为,导致判别信号不稳定的原因有两个.
 - 第一,判别器需要对高分辨率的图像进行处理,并尽可能地通过输出值区分伪造图像和真实图像,当图像的分辨率过高的时候,图像块的细微差距很难反映在一般判别器的输出值上.
 - 第二,目前在 SR 任务中使用的 GAN 损失本身是不稳定的,这种不稳定性是由 GAN 的判别器造成的,无论是使用 f 散度的 GAN 还是使用 Wasserstein 距离的 Wasserstein GAN,其判别器都是以概率的形式度量真实数据与生成样本之间的距离.一方面, f 散度难以识别高维空间中低维流形之间的距离,而 Wasserstein 距离中的 Lipschitz 条件也是近似的,它们本身作为判别距离时就存在不稳定性;另一方面,训练判别器的目的是使其对数据和样本之间的微小距离敏感,尽管生成器和判别器之间的 Nash 平衡理论上存在,但对于 SR 任务中使用的生成器而言,其生成的样本受网络参数初始化、正则化、dropout 等技术手段的影响,本身存在一定的波动,尤其是本文所关注的 $4\times$ SR 任务,由于生成器的层数较多,其波动性也相对较大,而基于概率距离测度的判别器对这种波动性是敏感的.此外,SR 任务中所使用的判别器也没有很好地利用高分辨率图像的有效特征.

为了解决上述问题,生成更加锐利的图像,本文在常用的 VGG 重构损失基础上,设计了一个稳定、有效的基于能量的对抗辅助损失.使用能量函数作为判别器取代传统判别器的优势在于,把数据编码为能量的过程考虑了神经网络本身的波动性,且当数据的能量流构建之后,可以使用生成器追踪该能量流.追踪数据能量流的另一个优势在于当能量趋近于 0 时,判别器不再产生额外的梯度,因此基于能量的对抗生成网络是相对稳定的.为了

构建相对稳定的辅助能量损失,本文借鉴统计力学中玻尔兹曼分布的概念以及基于能量的 GAN 模型^[18],玻尔兹曼分布建立了能量和概率之间的关系,根据玻尔兹曼分布,能量越低对应样本的概率就越大,当能量函数建模了真实的高分辨率图像时(真实的高分辨率图像都对应着低能量值,理想情况下,这个能量趋近于 0),其对应的概率分布可以认为是真实数据的分布 P_{data} ,因此假设服从数据分布的样本具有低能量,那么当传递给判别器的样本能量足够低时,就可以认为该样本服从数据分布,由此构建的生成对抗网络可以看作是模型分布追踪数据的能量流.基于这个思路,本文使用自编码器(auto encoder,简称 AE)建模能量函数,AE 的重构误差天然适合表示隐式形式的能量函数.

- 首先,使用真实的高分辨率图像训练 AE,AE 的重构误差表示数据的能量,当样本接近数据的能量流时,AE 的重构误差小,对应于低能量,根据玻尔兹曼分布,该样本以高概率属于数据分布,而生成器的训练目标一方面最小化重构损失、一方面追踪该能量函数.为了使真实数据与生成的高分辨率图像之间的微小差异反映在基于能量的对抗损失中,本文需要使用图像的有效特征构建能量,且该特征应当使能量函数对图像微小的变化敏感.因此,表示能量的 AE 使用构建重构损失时使用的 VGG 编码部分,并设计相应的可训练的解码器组成 U-Net 自编码结构,该自编码结构称为 VGG-UAE,利用 VGG-UAE 的重构误差表示能量,并把该能量用作判别器,训练判别器使数据流具有低能量.
- 然后,利用该判别器训练生成器,使生成的高分辨率图像追踪数据的能量流.

该对抗损失称为 VGG 能量损失.因此,本文的目标函数由 VGG 重构损失和 VGG 能量损失两部分组成,训练过程同时最小化这两个损失、完成 SR 任务.在本文的实验部分,我们通过实验验证了使用 VGG 能量作为对抗损失,不仅可以提高神经网络在 SR 任务中的表现,还可以使生成的高分辨率图像相对稳定.

本文的主要贡献可以概括如下.本文在 SR 任务中构建了基于能量的对抗辅助损失,称为 VGG 能量损失.为了更加稳定的完成 SR 任务,我们使用能量函数,使模型追踪数据的能量流;为了使高分辨率图像的微小差异反映在基于能量的对抗损失中,能量损失使用训练好的 VGG 编码部分和可训练的解码部分组成 VGG-UAE 结构,从而使用更加有效的重构损失作为能量损失.

本文首先介绍 SR 任务的常用方法以及存在的问题.第 2 节为本文提出的基于 VGG 能量的 SR 模型:VGG-SR-GAN.第 3 节通过实验验证本文提出的方法在 SR 任务中的有效性.第 4 节为总结与展望.

1 相关工作

与本文相关的研究是使用神经网络完成 SR 任务.目前,在 SR 任务中最常使用的神经网络是深度卷积神经网络,模型以一个低分辨率图像($Input_L$)作为输入,输出与之对应的高分辨率图像($Fake_H$);通过最小化 $Fake_H$ 与真实的高分辨率图像($Real_H$)之间的距离,训练深度卷积神经网络,从而完成由低分辨率图像到高分辨率图像之间的映射.但是,当使用传统的均方误差($Real_H - Fake_H$)² 作为重构损失函数时(基于均方误差的度量指标在图像处理领域被称为峰值信噪比),生成的 $Fake_H$ 虽然在峰值信噪比指标上表现良好,但是图像的细节不够清晰锐利.一方面,这是由于使用 $L2$ 重构损失作为损失函数会促使图像中新生成的像素点趋近于其周边的像素的平均值;另一方面,最小化高分辨率图像之间的 $L2$ 距离会产生一个平均结果,不能很好地反映像素级别的微小差异.为了生成更加清晰锐利的高分辨率图像,在深度卷积神经网络的基础上,学者们使用高分辨率图像深度特征之间的距离作为重构损失,并调整了目标函数的形式,将目标函数由传统的重构损失拓展到重构损失+对抗辅助损失的形式;使用基于对抗的辅助损失来优化高分辨率图像的细节.具体而言,对于训练集中的高分辨率图像,首先使用高斯模糊将其变成原尺寸 1/16 的低分辨率图像($height$ 和 $width$ 各变为原来的 1/4),然后把该低分辨率图像传递到基于深度卷积神经网络的生成器中作为输入,输出该图像对应的高分辨率图像 $Fake_H$,然后将 $Fake_H$ 和 $Real_H$ 输入给训练好的 VGG 网络中,提取深度特征,将深度特征之间的距离作为重构损失,该重构损失称为 VGG 重构损失.模型同时最小化 $Fake_H$ 与 $Real_H$ 的对抗损失,SRGAN 论文中把这种损失取名为 Perceptual Loss(本文称其为对抗损失).这个损失本质上是基于对抗正则化的重构误差,其中,使用 VGG 重构损失是为了防止图像模糊,图像模糊是因为使用峰值信噪比作为损失会导致生成的图像的所有像素都趋向于总

误差的平均值,因此虽然图像的峰值信噪比指标表现得非常理想,但是生成的超分辨率图像仍然模糊.而 VGG 重构损失度量高分辨率图像特征之间的距离,因此能够更有效地反映图像之间的微小差异.对抗损失的加入,更多的作用是提供正则化,生成细节更锐利的图像.SRGAN 的网络结构以及部分生成图像如图 1 所示.

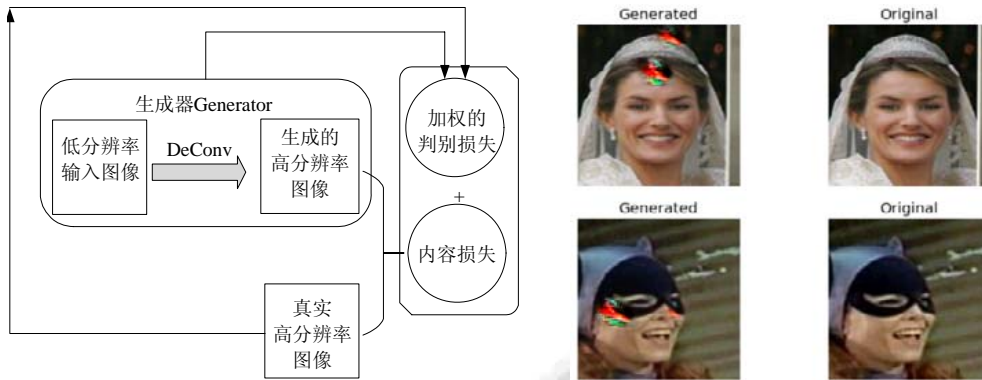


Fig.1 Diagram of a SRGAN model and some generated unreal images

图 1 SRGAN 的模型示意图以及生成的失真的图像

对于 SRGAN,其主体还是使用两个图像之间的相似度度量作为损失,对抗损失只起到辅助作用.但是实际应用中,使用 SRGAN 生成的超分辨率图像会存在不稳定的问题.从图 1 右图可以看出,虽然 SRGAN 生成的图像具有很高的还原度,但是也会存在不稳定的问题.我们已知:使用传统的 end-to-end 方法直接训练神经网络来完成 SR 任务尽管不能得到清晰的 4×超分辨率重建,但是不会出现图 1 中不稳定的情形.结合先前的分析可以推断,SRGAN 的不稳定是由对抗损失引起的.为了克服上述问题、生成更加锐利的图像,本文在常用的 VGG 重构损失基础上设计了一个稳定、有效的基于对抗的辅助损失.为了构建相对稳定的对抗辅助损失,本文将高分辨率图像编码为能量函数的形式,并使用该能量函数替代 GAN 中的判别器.假设服从数据分布的样本具有低能量,那么当传递给判别器的样本对应的能量足够低时,就可以认为该样本服从数据分布,由此构建的生成对抗网络可以看作是模型分布追踪数据的能量流.追踪数据能量流的优势在于:当能量趋近于 0 时,判别器不再产生额外的梯度,因此基于能量的对抗生成网络是相对稳定的.从损失函数的角度看,内容损失部分使用了 VGG 网络编码的特征,那么这个 VGG 网络也可以用作能量函数中的特征编码器,只需针对 VGG 编码设计一个解码器,就可以设计一个能量函数形式的判别器,并建立内容损失和对抗损失之间的联系,从而稳定训练过程.

2 基于 VGG 能量的 SR 模型

本文所做的工作是,在 VGG 重构损失的基础上构建了基于能量的对抗辅助损失,称为 VGG 能量损失.为了更加稳定地完成 SR 任务,我们使用能量函数的概念,使模型追踪数据的能量流;为了使高分辨率图像像素的微小差异反映在基于能量的对抗损失中,能量损失使用 VGG 编码部分和可训练的解码部分组成自编码结构——VGG-UAE.该模型称为 VGG-SRGAN,其模型结构图如图 2 所示.

如图 2 所示,左侧部分表示由低分辨率图像到高分辨率图像的生成器 Generator;右侧部分表示基于 VGG 能量的判别器,其中,VGG 编码网络部分使用训练好的 VGG 网络模型,在 VGG-SRGAN 中不训练 VGG 编码网络的参数.此时,VGG 编码器被用作一个给定的特征提取器.因此,可以使用 VGG 编码网络完成特征编码、构建内容损失(图 2 中的 content loss),同时利用 VGG 编码网络编码能量函数,只需根据编码完成解码,构造基于 VGG 编码的自编码器,使用 VGG-AE 的损失作为能量即可.该自编码器使用 U-Net 网络结构,称为 VGG-UAE.整个过程中,VGG 编码部分不需要训练.对于 VGG-SRGAN,其生成器(generator)为基于残差的卷积神经网络,输入为低分辨率图像,通过逐步的卷积和上采样操作得到 4×超分辨率图像.判别器为 VGG-UAE,网络的目标是使数据的能量尽可能低,并以数据的能量为指标,使生成的样本追踪数据的能量流.追踪数据能量流的优势在于:当能量

趋近于 0 时,判别器不再产生额外的梯度,因此基于能量的对抗生成网络是相对稳定的.为了使真实数据与生成的高分辨率图像之间的微小差异反映在基于能量的对抗损失中,本文使用图像的有效特征构建能量.该特征可以使能量函数对图像微小的变化敏感.基于这个思路,本文使用自编码器编码数据,以 AE 的重构损失作为能量.

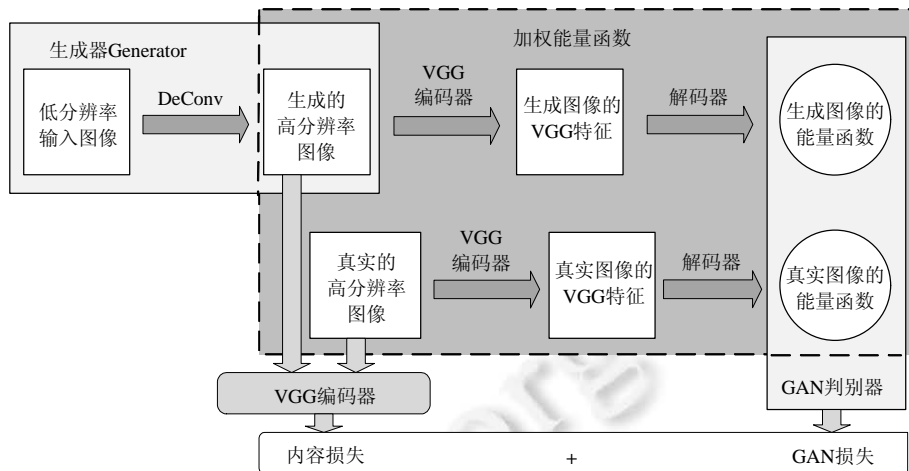


Fig.2 Diagram of a VGG-SRGAN model
图 2 VGG-SRGAN 的模型示意图

不同于先前基于能量的 GAN 模型,在本文中,为了提取更加有效的编码特征,所使用的 AE 基于 ImageNet 数据集训练好的 VGG 编码网络提取有效的图像特征;为了得到更有效的重构损失作为能量,本文针对 VGG 编码器设计相应的可训练的解码器组成基于 VGG 编码的 U-Net 结构,训练判别器使数据流具有低的能量;然后用该判别器训练生成器,使生成的高分辨率图像样本追踪数据的能量流.该对抗损失称为 VGG 能量损失.

基于 VGG 编码的 U-Net:VGG-UAE 示意图表示如下(如图 3 所示),本文使用该网络对数据进行编码,得到的重构损失记为 $U(x)$.如前文所述,VGG-SRGAN 中的目标函数包括两个部分:基于 VGG 编码的重构损失 L_{con} 和基于 VGG 能量的对抗损失 L_{GAN} .VGG 编码器的特征提取过程记为 $VGG_{feature}(\cdot)$.

根据我们定义的生成器 Generator,我们把两个损失定义如下:

$$L_{con}=[VGG_{feature}(Real_H)-VGG_{feature}(Fake_H)]^2 \tag{1}$$

其中,Real_H 表示真实的高分辨率图像(真实数据),Fake_H 表示由生成器得到的与真实图像对应的伪造的高分辨率图像:

$$Fake_H=Generator(Input_L) \text{ where } resolution(Fake_H)=4 \times resolution(Input_L) \tag{2}$$

根据公式(2),可以使用 Generator,由输入的低分辨率图像 Input_L 生成对应的高分辨率图像 Fake_H.公式(1)利用 VGG 编码器提供了基于 VGG 编码的重构损失,接下来我们定义基于 VGG 能量的对抗损失:

$$L'_{GAN} = E_{x \sim Fake_H}[U(x)] - E_{x \sim Real_H}[U(x)] \text{ where } Fake_H = Generator(Input_L) \tag{3}$$

其中,U(x)为能量函数,表现为 AE 的形式.该 AE 基于使用 VGG 编码器的 U-Net 网络输出重构损失,并将该重构损失作为能量函数:

$$U(x)=AE_Loss(VGG_{feature}(Image_H)) \tag{4}$$

其中,对于真实的高分辨率图像 Real_H,得到与之对应的能量 $U_{real}(x)$,模型的目标之一是使伪造的高分辨率图像能够追踪能量 $U_{real}(x)$.能量 U 可以看作势能面,数据应当位于势能的最低点,因此在目标函数中,我们加入能量函数的梯度惩罚项,使数据点的能量梯度为 0.由此,公式(3)可以表达为如下形式:

$$L_{GAN} = L'_{GAN} + \lambda E_{x \sim Real_H}[\|\nabla_{\theta} U(x)\|^2] \tag{5}$$

因此,VGG-SRGAN 的目标函数可以表示为

$$L=L_{con}+\epsilon L_{GAN} \quad (6)$$

其中, ϵ 是损失的比例系数.

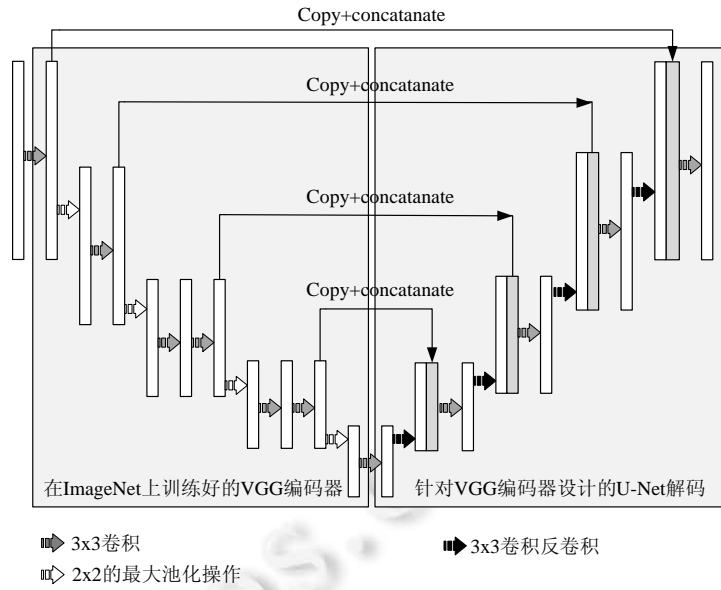


Fig.3 Diagram of a VGG coding based U-Net: VGG-UAE

图3 基于 VGG 编码的 U-Net: VGG-UAE 示意图

网络的训练以 GAN 的方式进行,首先使用生成器 Generator 从低分辨率图像生成高分率图像,然后使用 ImageNet 数据集上训练好的 VGG 网络提取特征,利用 VGG 特征构建损失函数 L_{con} ;针对 VGG 特征,构建相应解码器组成基于 U-Net 架构的 VGG-UAE 模型,把 VGG-UAE 的重构损失用作损失函数 L_{GAN} ;根据损失函数,利用梯度下降方法更新 VGG-UAE 中的参数和 Generator 中的参数,完成模型的训练.VGG-SRGAN 的训练过程如算法 1 所示.

算法 1. VGG-SRGAN 训练算法.其中, $VGG_{feature}$ 为训练好的固定参数的 VGG 网络, ϵ 和 λ 是超参数.

for $t=1$ to T (迭代次数) **do**

从数据集中取高分率图像 $Real_H$.

针对真实的高分辨率图像 $Real_H$,对其进行 $4\times$ 降采样,得到低分辨率图像 $Input_L$.

#由低分辨率图像生成高分率图像 $Fake_H$ 和 VGG 特征:

$$Fake_H=Generator(Input_L); VGG_{feature}(Fake_H); VGG_{feature}(Real_H).$$

#构建基于内容的损失 L_{con} :

$$L_{con}=[VGG_{feature}(Real_H)-VGG_{feature}(Fake_H)]^2.$$

#构建 VGG-UAE 的能量函数:

$$U(x)=AE_Loss(VGG_{feature}(Image_H)).$$

#构建对抗正则化损失 L_{GAN} :

$$L_{GAN}=E_{x\sim Input_L}[U(Generator(x))]-E_{x\sim Real_H}[U(x)]+\lambda E_{x\sim Real_H}[\|\nabla_{\theta}U(x)\|^2].$$

#根据损失函数计算 Generator 参数的梯度:

$$\nabla_{\theta_{Generator}} \{[VGG_{feature}(Generator(Input_L))]^2 - E_{x\sim Input_L}[U(Generator(x))]\}.$$

#根据损失函数计算 VGG-UAE 中能量 U 的参数的梯度:

$$\nabla_{\theta_U} \{E_{x\sim Real_H}[U(x)] - E_{x\sim Input_L}[U(Generator(x))] + \lambda E_{x\sim Real_H}[\|\nabla_{\theta}U(x)\|^2]\}.$$

end for

梯度的更新可以使用任何标准的基于梯度的学习规则,我们在实验中使用动量方法.

3 实验

我们的实验数据集为 Set5^[6],Set14^[20],BSD100^[21]和 CelebA^[22],在 CelebA 上,使用 128×128 的图像进行训练.所有实验在真实图像数据集上采用 4×分辨率缩,输入图像 *Input_L* 是真实图像分辨率的 1/16.为了进行定量的比较,本文使用的指标为峰值信噪比(peak signal-to-noise ratio,简称 PSNR)和结构相似性(structural SIMilarity,简称 SSIM).本文的对比实验所采用的方法为 bicubic interpolation、SRCNN^[9]、SelfE×SR^[23]、SRResNet 以及基于 VGG-22 的 SRGAN^[11](因为基于 VGG-22 的 SRGAN 在量化指标上表现更好,在针对生成的高分辨率图像定性分析中,采样基于 VGG-54 的 SRGAN 结果,该结果从定性分析上生成的图像相对 VGG-22 更加有效),相关数据获取于 <https://github.com/jbhuang0604/SelfE×SR>.本文使用的 VGG 编码器和 VGG-UAE 中使用的 VGG 网络架构为 VGG-19,该网络为 ImageNet 上预训练的模型.为了方便计算和比较,我们将高分辨率图像的像素规则化到区间[-1,1].在梯度下降算法中,我们选择 Adam 方法优化梯度,并使用参数 $\beta_1=0.9$,学习率为 10^{-4} , L_{con} 和 L_{GAN} 的比例系数固定为 10^{-3} ,迭代次数为 10^5 .本文使用的 Generator 中含有 16 个残差块.本文的实验结果见表 1.

Table 1 PSNR and SSIM on different datasets based on different generative models

表 1 不同的生成模型在不同数据集上的 PSNR 和 SSIM

Set5	Bicubic	SRCNN	SelfE×SR	SRResNet	SRGAN	VGG-SRGAN	HR
PSNR	28.44	30.10	30.17	32.09	29.68	30.02	∞
SSIM	0.821 4	0.859 1	0.880	0.901 9	0.847 2	0.851 1	1
Set14	Bicubic	SRCNN	SelfE×SR	SRResNet	SRGAN	VGG-SRGAN	HR
PSNR	25.27	27.06	27.46	28.49	26.12	27.07	∞
SSIM	0.741 6	0.780 7	0.796 4	0.818 4	0.750 1	0.761 1	1
BSD100	Bicubic	SRCNN	SelfE×SR	SRResNet	SRGAN	VGG-SRGAN	HR
PSNR	25.87	26.69	26.79	27.39	25.02	26.07	∞
SSIM	0.690 2	0.722 1	0.732 0	0.759 5	0.658 9	0.662 1	1

表 1 展示了本文提出的 VGG-SRGAN 以及对算法、真实图像的 PSNR 和 SSIM 指标.从定量的比较来看,本文提出的方法虽然优于 SRGAN,但是不及传统的 SRResNet.究其原因,基于 ResNet 及其改进的网络模型在损失函数上使用基于内容的损失,其形式与 PSNR 和 SSIM 指标非常类似,因此在测试中,相对于 SRGAN 和本文提出的方法更有优势.为了直观地展示实验中的生成图像,本文将生成的部分图像展示如图 4 所示.

如图 4 所示,细节部分是我们将图像放大同样倍数后截取细节,从图 4 可以看出,我们提出的 VGG-SRGAN 模型能够有效地生成与真实图像对应的高分辨率图像,并对图像的细节部分(纹路、断点、凸起等)有一定的还原能力.为了进行直观的对比,我们将定量分析中效果最好的 SRResNet 和 SRGAN 在本数据上的生成图像进行对比.

图 5 展示了 SRResNet、SRGAN、VGG-SRGAN 以及真实图像的细节对比,其中,左数第 1 行第 1 幅图像为 SRResNet 生成的图像,第 2 行第 1 幅图像为上面图像中对应的放大细节;第 1 行的第 2 幅图像是基于 SRGAN 生成的,其中,VGG 特征编码网络层数为 52 层;第 1 行的第 3 幅图像是 VGG-SRGAN 生成的图像,其中,VGG 编码器和 VGG-UAE 中 VGG 的层数都是 19.如图 5 所示,无论是 SRGAN 还是我们提出的 VGG-SRGAN 都能生成比 SRResNet 更加清晰锐利的高分辨率图像,具体体现在细节、纹路、凸起和断点上.SRGAN 和 VGG-SRGAN 相比,生成的图像虽然更加锐利,但是其细节点处的噪声也更多.VGG-SRGAN 在保持一定细节信息的同时,还使生成的图像更加干净,同时在信噪比等指标上也优于 SRGAN.其他生成高分辨率图像与真实图像的对比实例如图 6 所示.

如图 6 所示,第 1 行图像和第 2 行图像来自数据集 Set14、Set5 和 BSD100,最后一行数据来自于 CelebA 数据集.对于 Set14、Set5 和 BSD100 数据集,由于样本数较少,在算法中我们保证足够多的迭代次数,从而尽可能降低真实图像和生成图像之间的损失;但是对于 CelebA 数据集,由于样本数较多,在总迭代次数一定的前提下,

单一样本训练的迭代次数相对较少,因此生成的高分辨率图像不够清晰,可以通过增加迭代次数改善实验结果.

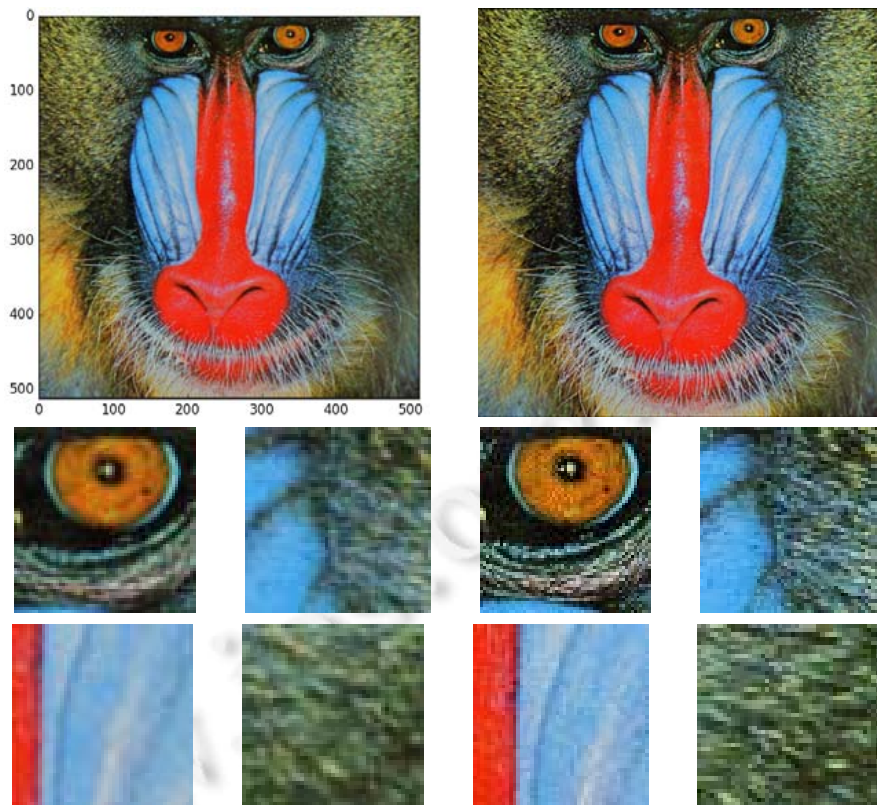


Fig.4 Overall and detail comparison of a generated image of a VGG-SRGAN (left) and the corresponding real image (right)

图4 基于 VGG-SRGAN 的生成图像(左)与真实图像(右)的整体和细节对比

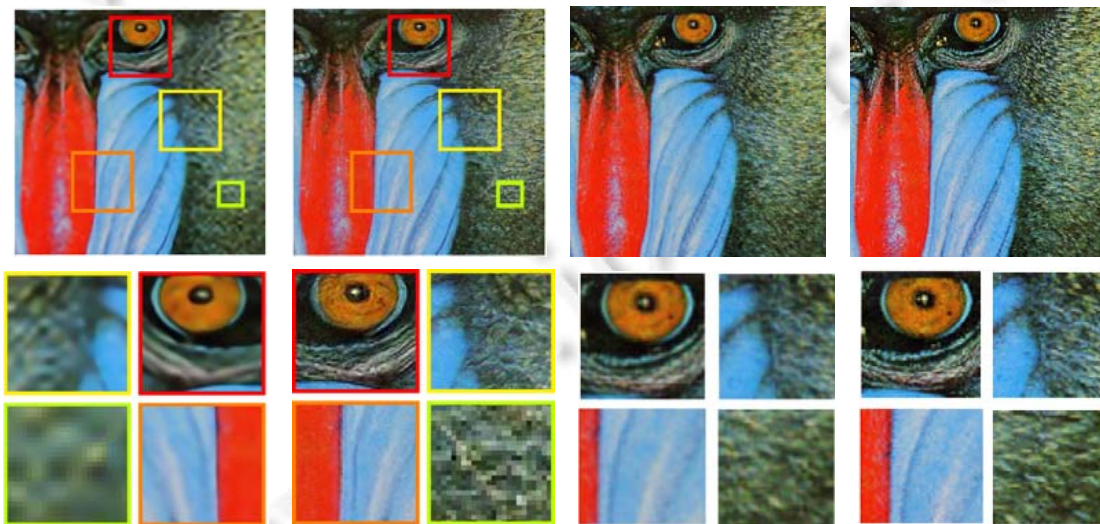


Fig.5 Comparison of a generated image of a VGG-SRGAN, that of comparative models, and the real image

图5 基于 VGG-SRGAN 的生成图像与其他模型和真实图像的对比

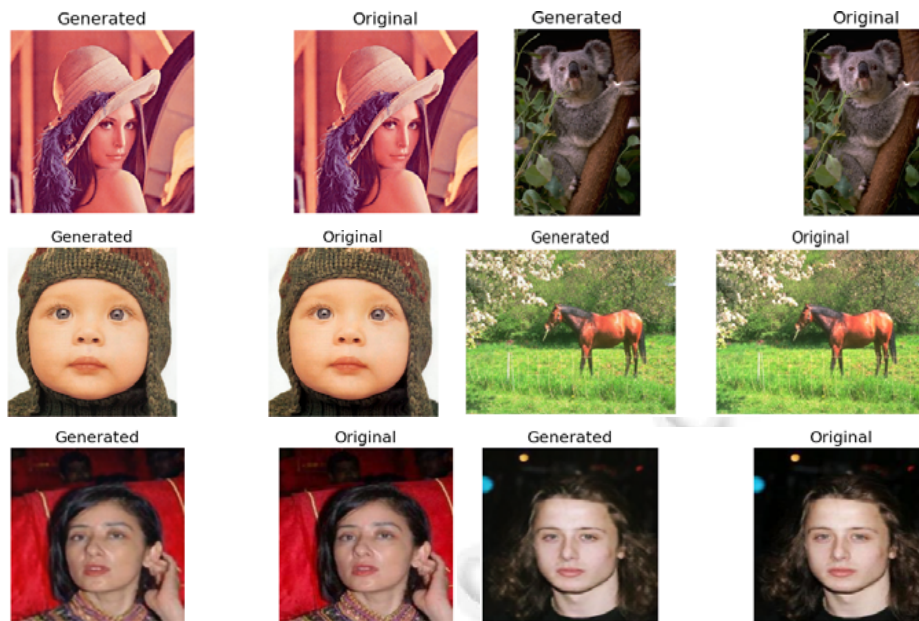


Fig.6 Overall comparison of generated images of a VGG-SRGAN (left) and the corresponding real images (right)
图 6 基于 VGG-SRGAN 的生成的多幅图像(左)与真实图像(右)的整体对比

在实验部分,本文验证了提出的算法所生成的高分辨率图像能够拟合真实图像的统计特性,对微小的图像细节能够实现较高精度的还原,且生成的高分辨率图像更加稳定.

4 总结与展望

为了从低分辨率图像生成更加锐利清晰的 $4\times$ 高分辨率图像,本文提出了基于 VGG 能量损失的生成模型 VGG-SRGAN,该模型以低分辨率的图像作为输入,使用在 ImageNet 上训练的固定权值的 VGG 网络作为特征提取器,提取高分辨率图像的特征,然后根据特征构建基于内容的损失函数 L_{con} 和基于 VGG 能量的对抗辅助损失 L_{GAN} ,其中,该 VGG 能量损失使用 VGG-UNet 网络的形式构建为 AE 结构 VGG-UAE,使用该 AE 的重构损失作为能量.使用能量函数形式的对抗损失作为辅助项能使神经网络的训练更加稳定,且可以生成清晰锐利的高分辨率图像.本文的实验部分验证了所提算法的有效性,本文提出的 VGG-SRGAN 不仅能够恢复原始图像的纹理细节,且有更少的噪声点、更加平滑.然而,本文所使用是算法在定量指标上弱于传统的基于内容损失的网络模型,如 SRResNet,究其原因,在于目前常用的定量分析指标大多与内容损失相关,因此不能很好地反映生成图像的细节特征.我们接下来的工作是在优化模型的同时,寻找更加适合 SR 任务的评价指标.

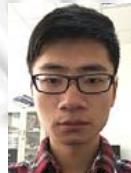
References:

- [1] He KM, Zhang X, Ren S, *et al.* Deep residual learning for image recognition. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2016. 770–778.
- [2] Ding SF, Zhang N, Zhang J, *et al.* Unsupervised extreme learning machine with representational features. Int'l Journal of Machine Learning and Cybernetics, 2017,8(2):587–595.
- [3] He Y, Zhang J, Shan H, *et al.* Multi-task GANs for view-specific feature learning in gait recognition. IEEE Trans. on Information Forensics & Security, 2019,14(1):102–113.
- [4] Zhang N, Ding SF, Zhang J, *et al.* An overview on restricted Boltzmann machines. Neurocomputing, 2017,275(31):1186–1199.
- [5] Seeliger K, Guclu U, Ambrogioni L, *et al.* Generative adversarial networks for reconstructing natural images from brain activity. NeuroImage, 2018,181(1):775–785.

- [6] Bevilacqua M, Roumy A, Guillemot C, *et al.* Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In: Proc. of the British Machine Vision Conf. 2012.
- [7] Borman S, Stevenson RL. Super-resolution from image sequences—A review. In: Proc. of the Midwest Symp. on Circuits and Systems. 1998. 374–378.
- [8] Bruna J, Sprechmann P, LeCun Y. Super-resolution with deep convolutional sufficient statistics. In: Proc. of the Int'l Conf. on Learning Representations. 2016.
- [9] Dong C, Loy CC, He K, *et al.* Image super-resolution using deep convolutional networks. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2016,38(2):295–307.
- [10] Dosovitskiy A, Brox T. Generating images with perceptual similarity metrics based on deep networks. In: Proc. of the Advances in Neural Information Processing Systems. 2016. 658–666.
- [11] Ledig C, Theis L, Huszar F, *et al.* Photo-realistic single image super-resolution using a generative adversarial network. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2017. 105–114.
- [12] Zhu JY, Park T, Isola P, *et al.* Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proc. of the IEEE Int'l Conf. on Computer Vision. 2017. 2223–2232.
- [13] Zeng K, Ding SF. Single image super-resolution using a polymorphic parallel CNN. Applied Intelligence, 2019,49(1):292–300.
- [14] Yang J, Wright J, Huang TS, *et al.* Image super-resolution as sparse representation of raw image patches. In: Proc. of the 2008 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. 2008. 1–8.
- [15] Wright J, Ma Y, Mairal J, *et al.* Sparse representation for computer vision and pattern recognition. Proc. of the IEEE, 2010,98(6): 1031–1044.
- [16] Dong W, Zhang L, Shi G, *et al.* Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. IEEE Trans. on Image Processing, 2010,20(7):1838–1857.
- [17] Goodfellow I, Pouget-Abadie J, Mirza M, *et al.* Generative adversarial nets. In: Proc. of the Int'l Conf. on Neural Information Processing Systems. 2014. 2672–2680.
- [18] Zhao J, Mathieu M, LeCun Y. Energy-based generative adversarial network. arXiv preprint arXiv:1609.03126, 2016.
- [19] Gulrajani I, Ahmed F, Arjovsky M, *et al.* Improved training of Wasserstein GANs. In: Proc. of the Advances in Neural Information Processing Systems. 2017. 5769–5779.
- [20] Zeyde R, Elad M, Protter M. On single image scale-up using sparse-representations. In: Proc. of the Curves and Surfaces. Springer-Verlag, 2012. 711–730.
- [21] 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. In: Proc. of the IEEE Int'l Conf. on Computer Vision, Vol.2. 2001. 416–423.
- [22] Liu ZW, Luo P, Wang XG, *et al.* Deep learning face attributes in the wild. In: Proc. of the Int'l Conf. on Computer Vision. 2015. 3730–3738.
- [23] Huang JB, Singh A, Ahuja N. Single image super-resolution from transformed self-exemplars. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2015. 5197–5206.



丁玲(1994—),女,讲师,主要研究领域为机器学习,数据挖掘.



张健(1990—),男,博士,讲师,CCF 专业会员,主要研究领域为机器学习,模式识别.



丁世飞(1963—),男,博士,教授,博士生导师,CCF 杰出会员,主要研究领域为智能信息处理,人工智能与模式识别,机器学习,数据挖掘,粗糙集,软计算,大数据分析,云计算.



张子晨(1992—),男,博士生,主要研究领域为机器学习,模式识别.