文章编号:1006-1037(2019)01-0106-06
doi:10.3969/j.issn.1006-1037.2019.02.20

基于改进的深度卷积神经网络的单图像超分辨率重建

刘世豪,李 军
(青岛大学计算机科学技术学院,青岛 266071)

摘要:为解决现有的超分辨率模型不能很好的恢复图像的纹理细节和模型训练困难等问题,结合现有的残差网络和GoogleNet中的Inception模块对其进行改进。通过将5×5的卷积核替换为两个级联的3×3的卷积核、使用LeakyReLU作为激活函数和删除池化层等方法对原始的Inception模块进行改进,然后在模型中多次级联改进后的Inception模块。实验结果表明,与双三次插值算法、SRCNN和VDSR算法相比,改进后的模型能获得更高的峰值信噪比(PSNR)和结构相似性指数(SSIM),并且在视觉效果上也有明显的改善。

关键词:深度卷积神经网络;图像处理;超分辨率;Inception
中图分类号:TP391.4 **文献标志码**:A

单图像超分辨率致力于从单张低分辨率图像中获取视觉效果更好的高分辨率图像。因为获得的高分辨率图像保留了一些重要或关键的信息用于之后的图像处理,所以单图像超分辨率得到了越来越广泛的研究,并被广泛地应用于各个领域,比如视频监控、医学图像、人脸识别、卫星图像等。图像超分辨率方法主要分为三大类:基于插值、基于重建和基于学习的方法。基于插值的方法,像双三次插值[1]、最近邻插值[2]等,典型上是采用一个固定函数的内核去估计高分辨率图像未知的像素。虽然基于插值的方法能够以简单并有效的方式重建出高分辨率图像,但是它们产生了更平滑的边和模糊的细节,这种结果对于实际应用来说是不太满意的。基于重建的方法通常引入了某些图像先验知识或限制,如迭代反投影法[3]、凸集投影法[4]、最大后验概率法[5]等。虽然这些方法有效地保留了几何结构,但是未能增加足够的高频细节。基于学习的方法试图从大量LR-HR样本中学习一种映射,然后用学习到的映射去重建目标高分辨率图像。最近几年,基于学习的方法[6]已经变成了研究的焦点,大量的算法被提出,比如邻域嵌入[7]、稀疏编码[8]等。最近,基于卷积神经网络的方法由于它简单的结构和更好的重建质量得到了相当高的关注。Dong等[9]第一次证明了卷积神经网络能够被成功应用在图像超分辨率问题上,提出了SRCNN,并且训练了一个三层的卷积神经网络用来学习低分辨率图像和高分辨率图像之间的映射。SRCNN是一个浅层网络,仅仅由三个卷积层组成,包括补丁提取层、非线性映射层和重建层。他们试图去增加网络的深度来提高模型性能,但是实验结果表明更深的模型表现的比浅层模型更糟,最终得出结论,更深的网络并不总能得到更好的结果。然而随着网络层数的加深,网络可以在低分辨率图像和高分辨率图像之间学习到更复杂的映射关系。随后,肖进胜[10]等通过调整卷积核大小、加入池化层等方法在SRCNN的基础上进行改进取得了更好的结果。郭晓[11]等通过级联HD-CN模型解决了放大倍数无法按需选择的问题,并且引入了深度边缘滤波器来减少级联误差。为了在一个大的图像区域内利用更多的上下文信息,Kim等[12]受到VGG-net的启发,通过使用非常深的卷积神经网络,提出了一种使重建质量更好的超分辨率方法VDSR。VDSR通过多次使用3×3的卷积核加深网络的深度,实验结果表明,增加网络的深度确实能够提高重建性能。虽然VDSR已经在重建质量实现了优异的性能,但是仍有一些不足,它试图去级联多个小的卷积核来增加网络的深度以提高重建效果,但是却存在对区域特征利用不足的问题,易导致重建后的图像纹理细节丢失;并且由于网络的层数过深,容易使训练过程不

收稿日期:2018-11-01
通讯作者:李军,男,博士,研究员,主要研究方向为数据分析计算机工程仿真等。

稳定,从而导致梯度消失或梯度爆炸的问题。针对以上问题,本文通过对原始的 Inception 模块进行改进缓解了传统模型重建出的图像纹理细节丢失的问题,并且训练过程中使用学习率衰减稳定了模型的训练过程。实验结果表明,改进后的模型可以有效地提高单图像超分辨率重建的效果。

1　模型改进

1.1　Inception 模块的改进

为了更加充分地利用图像中隐含的信息,Szegedy 等[13]提出了最初用于图像分类问题的 Inception 模块,如图 1(a)所示,该模块包含池化层和多个卷积核大小不同的卷积层,输出时将不同的卷积层进行合并,这种通道上的合并可以使网络提取更复杂的特征。受 GoogleNet 的启发,改进的模型针对超分辨率问题对原始的 Inception 模块进行了改进,并使用改进后的 Inception 模块来提升单图像超分辨率的性能。由于原始的 Inception 模块中使用了 5×5 的卷积核,导致网络的参数过多,影响了网络的收敛速度。考虑到一个 5×5 的卷积核和两个 3×3 的卷积核感受野的大小是相同的,但是一个 5×5 的卷积核需要 25 个参数,而两个 3×3 的卷积核只需要 18 个参数,所以为了减少参数,本文主要使用了 3×3 的卷积核。

原始的 Inception 模块中使用 ReLU[14]作为激活函数,如果学习率设置的太大,可能会导致网络中一些神经元"死亡",即神经元的梯度不会更新。即使学习率较小,这种情况也有可能发生。所以本文的激活函数选择 LeakyReLU[15]

$$\mathrm{LeakyReLU}(x)=\begin{cases}x, x\geqslant 0\\ ax, x<0\end{cases} \tag{1}$$

其中,a 为固定值,通常是一个很小的常数。该函数在一定程度上缓解了神经元"死亡"的问题,当输入为负值时,LeakyReLU 会有非 0 输出值,从而获得一个小的梯度。

在图像分类等领域中,池化层通过下采样操作可以降低输出结果的维度,加快模型的收敛速度并且也可以有效地减少过拟合现象。Kim 等[16]提出,在超分辨率和降噪等图像恢复问题中,图像细节是非常重要的,由于池化层的下采样操作会造成信息丢失,所以深度学习模型面对这些问题时最好不要使用池化层。

针对上述问题,本文对原始的 Inception 模块进行了相应的改进,使用 LeakyReLU 代替 ReLU 作为激活函数,删除了其中的池化层,并且将 5×5 的卷积核替换为两个级联的 3×3 的卷积核。改进的 Inception 模块结构如图 1(b)所示。

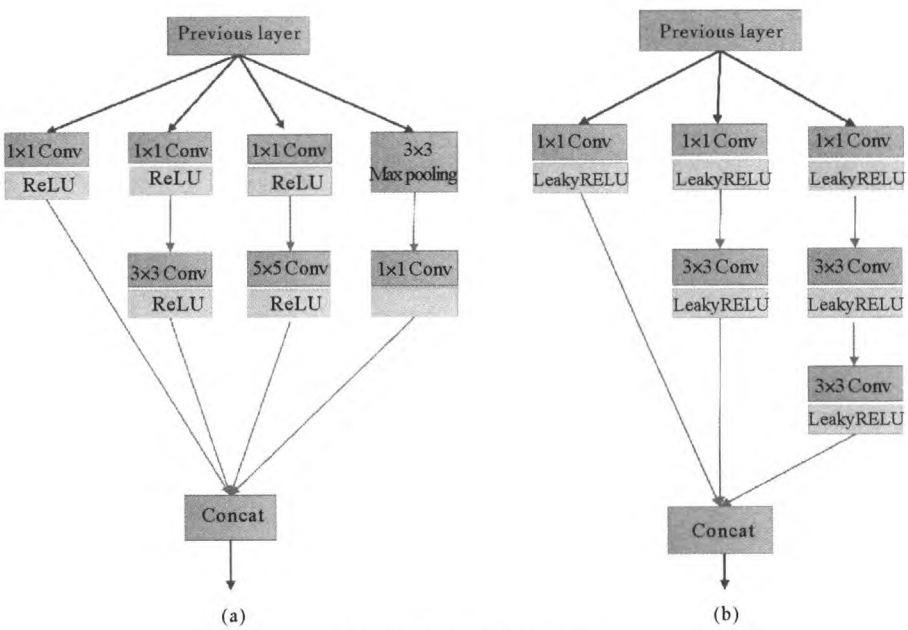

图 1　Inception 模块的改进

(a)原始的 Inception 模块;(b)改进后的 Inception 模块。

1.2 深度CNN的模型结构的改进

本文提出的模型与VDSR相比:使用改进后的Inception模块来代替普通的卷积层;使用LeakyReLU激活函数代替ReLU激活函数;使用学习率衰减来稳定训练过程,降低模型损失。提出的网络结构如图2所示。模型的输入为低分辨率图像,输出为双三次插值后的图像与训练得到的残差相加后得到的图像。

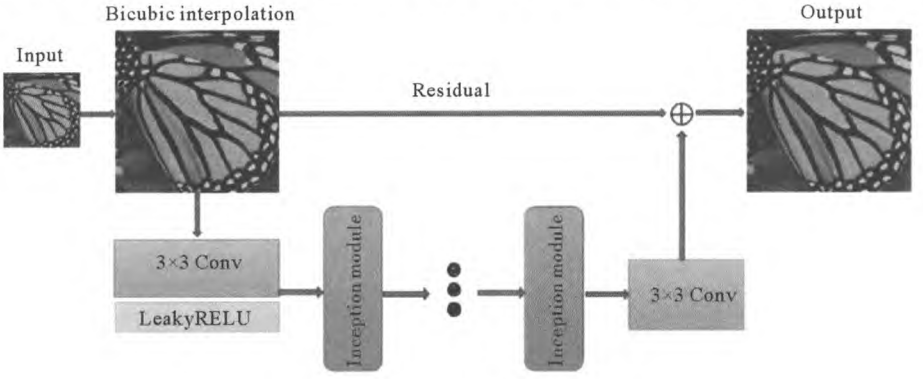

图2 改进的深度CNN的模型结构

在SRCNN中,模型直接学习低分辨率图像LR与高分辨率图像HR之间的映射关系

$$F(X) = W_3 * A(W_2 * A(W_1 * X + B_1) + B_2) + B_3 \tag{2}$$

其中,X为原始的LR图像,W_x和B_x分别表示第x层卷积核的权值和偏差,A表示激活函数,SRCNN中使用修正线性单元ReLU。F即模型学习到的映射。

实际上,LR图像和HR图像之间拥有很多相似的低频信息,而如何利用这些相似的信息在SRCNN中并没有体现出来。为了充分地利用这些相似的低频信息,本文引入了残差学习[17]的方法。加入残差后,模型学习的映射从$Y = F(X)$变为$Y = G(X) + X = F(X)$,其中,$G(X)$为LR图像和HR图像间的残差映射,也可以表示为图像中的高频信息。所以,模型只需要学习$G(X)$即可,即图像中高频信息的映射。

在对LR图像进行超分辨率重建时,可以先将RGB图像转换成YCbCr图像,只需要对YCbCr图像中的Y通道(亮度通道)进行处理,剩余两个通道则直接使用双三次插值方法放大到目标尺寸,三个通道融合后的图像即为最终重建后图像。本文首先通过双三次插值法将低分辨率图像上采样到目标大小,然后插值后的图像经过3×3的卷积核和一个Inception模块将通道数增加到64个,经过LeakyReLU激活函数后输出,这一部分相当于SRCNN中的特征提取层。随后级联的m个Inception模块作为非线性映射部分,将输出后的结果输入到一个Inception模块和只含有1个3×3的卷积核的卷积层中输出残差图像,这一部分相当于SRCNN中的重建层。最后将预测出的图像高频细节与双三次插值后的图像相加得到最终的结果。

1.3 评价标准和损失函数

常用的评价图像质量的指标有峰值信噪比(Peak Signal to Noise Ratio,PSNR)和结构相似性指数(Structural Similarity Index,SSIM)[18]等,本文使用PSNR和SSIM对重建后的图像进行评价。PSNR值越大表示图像失真越少,SSIM值越接近1表示图像越相似,1表示完全相似。

本文采用均方误差(MSE)作为神经网络的损失函数,目标是训练一个端到端的映射$F: y = F(x)$,其中,x为LR图像双三次插值后的结果,y为重建后的图像。给定训练集$\{(x^i, y^i)\}_{i=1}^N$,优化目标为

$$L(\theta) = \min_{\theta} \frac{1}{2N} \sum_{i=1}^{N} \| F(x_i;\theta) - y_i \|_F^2$$

其中,L为损失函数,θ为需要训练的参数,N为训练集中图像的数目,y_i和$F(x_i;\theta)$分别表示原始的高分辨率图像和重建后的高分辨率图像。

2 实验与分析

实验环境:硬件配置是Intel Core(R) i5-3570k CPU,16GB内存,Nvidia GeForce GTX1080显卡,软件

使用的是 Tensorflow 1.4.0,Matlab 2016b,Cuda 8.0,Cudnn5.1。

目前基于卷积神经网络的超分辨率算法最常用的训练数据集为 Yang[19] 的 91 张图像和 BSD500[20] 中的 200 张图像。为了客观地比较本文和其他超分辨率算法的结果,实验也使用这 291 张图像作为数据集。由于数据集规模较小,为了使模型更充分地利用这些图像来学习映射关系,所以采用数据增强的方法来扩大数据集的规模,主要采用以下三种方式:(1) 缩放:每幅图像都缩放为原来的 0.9、0.8、0.7、0.6 倍;(2) 旋转:每幅图像都被旋转 90°、180°和 270°;(3) 翻转:每幅图像都进行水平和垂直翻转。最终可以获得 $5\times4\times3=60$ 倍数目的图像来形成 HR 图像集合 $\{Y\}$。模型将不同放大倍数的图像一起训练使得模型可以处理不同倍数的超分辨率重建问题[12]。为了产生训练集,首先通过双三次插值法并使用不同的采样因子下采样原始的 HR 图像 $\{Y\}$ 以形成相应的 LR 图像 $\{X\}$,然后以步长为 41 和 $n\times41(n=2,3,4)$ 分别裁剪 LR 和 HR 图像得到共 287088 个 LR/HR 图像补丁对。最终,使用这些 LR/HR 图像补丁对作为神经网络的训练数据。

实验中使用 Adam 算法来更新网络的权值,epoch = 80,batch-size = 64。由于权值初始化方法会影响模型的收敛速度和最终结果,本文使用了 Xavier 初始化[21]方法。学习率决定了神经网络的参数达到最优的速度,当模型开始训练时,以一个较大的学习率开始直到损失不会大幅改变,当训练接近结束时,使用较小的学习率能加快网络找到损失函数最优值的速度。本文设置初始学习率为 5×10^{-4},并且随着迭代次数的增加,学习率按照二次多项式衰减,最终衰减到 10^{-5}。其余参数均使用 TensorFlow 默认设置。

为了验证提出的算法的有效性和正确性,本文进行了多次对比实验。影响实验结果的主要因素有卷积核的数目和 Inception 模块数 m 的选择。图 3(a)为卷积核数为 64 时测试集上平均 PSNR 的变化情况,(b)为卷积核数为 128 的情况,可见,卷积核数目为 128 时的 PSNR 值要比卷积核数目为 64 时有略微的提高。但当选择更多的卷积核数目比如 256 时,通过实验显示 PSNR 值并没有更好的提升,反而会增加网络的参数,减缓网络的训练速度。高层次的视觉任务(图像分类、目标检测等)和低层次的视觉任务(图像去噪、图像恢复等)都证明了网络模型结构越深,模型的性能越好。模型层数越深,越能获得更大的感受野,因此模型就能够获得更多的上下文信息,这些信息能够给高频细节的预测提供更多的帮助。但是在实验中发现当 m 超过 10 时,由于网络层数太深导致梯度没有更新,所以最终模型选用的 Inception 模块数为 10。

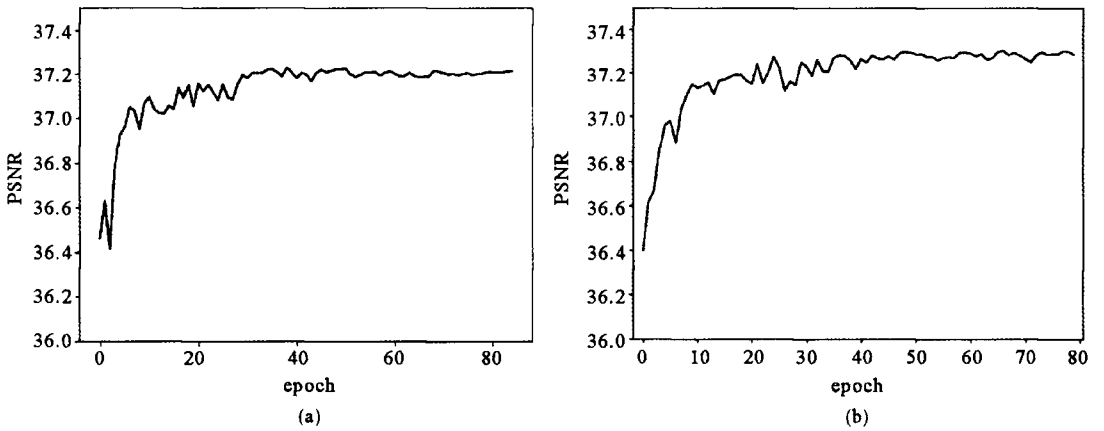

图 3　不同卷积核数目下的 PSNR 对比

(a)卷积核数目为 64 时 PSNR 变化情况;(b)卷积核数目为 128 时 PSNR 变化情况。

为了验证模型的重建性能,实验采用图像超分辨率领域通用的 Set5 和 Set14[22] 数据集进行测试,并将实验结果与 Bicubic、SRCNN 和 VDSR 算法进行对比。

表 1 为在 Set5 和 Set14 测试集下,不同算法不同放大倍数的平均 PSNR 和 SSIM 值。从表中可以看出,用改进后的模型进行超分辨率重建的结果在 PSNR 和 SSIM 值上均有一定程度的提高。

为了使本文方法和其他算法做更直观地对比,图 4 左侧为从 Set14 数据集中选取的一张图像 comic,右侧展示了在放大 4 倍情况下,不同算法对其重建后的高分辨率图像的部分区域。从图中可以看出,虽然 SRCNN 和 VDSR 方法重建后的结果比 Bicubic 的结果略好,但是在手指边缘处仍比较模糊,而本文方法的

重建后的结果在边缘处更为清晰。

表1 不同放大倍数下的平均PSNR和SSIM

Dataset	Scale	Bicubic	SRCNN	VDSR	Proposed
		PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Set5	×2	33.66/0.9299	36.34/0.9542	37.24/0.9587	37.43/0.9604
	×3	30.39/0.8682	32.39/0.9090	33.40/0.9189	33.52/0.9201
	×4	28.42/0.8104	30.09/0.8628	31.07/0.8773	31.21/0.8810
Set14	×2	30.23/0.8688	32.18/0.9063	32.79/0.9115	32.99/0.9198
	×3	27.54/0.7742	29.00/0.8209	29.65/0.8303	29.82/0.8330
	×4	26.00/0.7027	27.20/0.7503	27.85/0.7634	27.99/0.7672

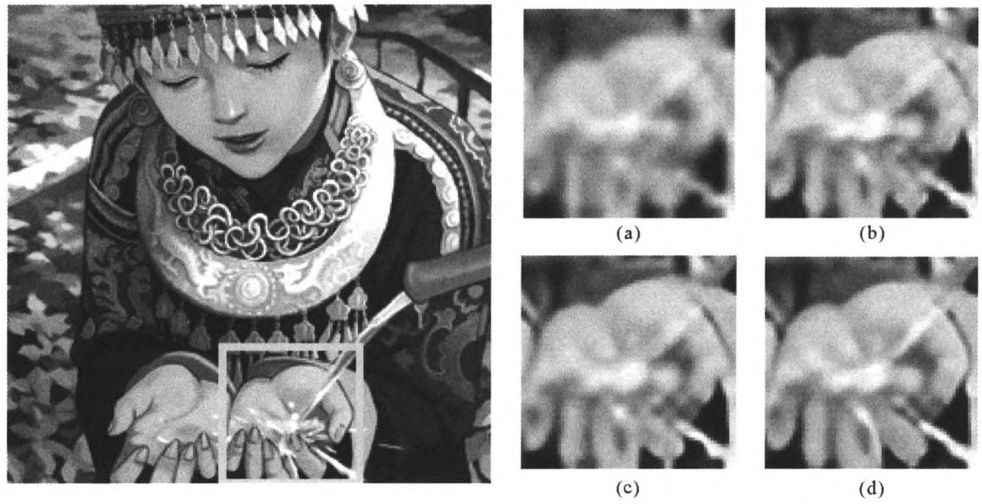

图4 不同算法放大图像4倍的部分区域对比

(a)Bicubic;(b)SRCNN;(c)VDSR;(d)Proposed

3 结论

本文结合Inception模块和残差网络对现有的超分辨率模型进行了改进,并与双三次插值、SRCNN和VDSR算法在同一测试集上进行了对比。实验显示,改进模型的超分辨率结果在PSNR、SSIM和视觉效果上都有较好的改善,证明改进后的Inception模块可以很好地用于超分辨率问题。由于模型在训练之前仍需要进行双三次插值操作,导致对图像进行处理的时间有所增加,在未来的工作中,将尝试结合不同的神经网络层和模型结构进一步提高重建图像的质量,同时提高图像处理速度,进而实现对视频的超分辨率重建。

参考文献

[1] Keys R. Cubic convolution interpolation for digital image processing[J]. IEEE transactions on acoustics, speech, and signal processing, 1981, 29(6): 1153-1160.

[2] Fattal R. Image upsampling via imposed edge statistics[C]//ACM transactions on graphics (TOG). ACM, 2007, 26(3): 95—103.

[3] Irani M, Peleg S. Improving resolution by image registration[J]. CVGIP: Graphical models and image processing, 1991, 53(3): 231-239.

[4] Stark H, Oskoui P. High-resolution image recovery from image-plane arrays, using convex projections[J]. JOSA A, 1989, 6(11): 1715-1726.

[5] Schultz R R, Stevenson R L. Improved definition video frame enhancement[C]//Acoustics, Speech, and Signal Processing, 1995. ICASSP-95., 1995 International Conference on. IEEE, 1995, 4: 2169-2172.

[6] Freeman W T, Pasztor E C, Carmichael O T. Learning low-level vision[J]. International journal of computer vision, 2000, 40(1): 25-47.

[7] Gao X, Zhang K, Tao D, et al. Image super-resolution with sparse neighbor embedding[J]. IEEE Transactions on Image Processing, 2012, 21(7): 3194-3205.

[8] Yang J, Wright J, Huang T S, et al. Image super-resolution via sparse representation[J]. IEEE transactions on image processing, 2010, 19(11): 2861-2873.

[9] Dong C, Loy C C, He K, et al. Image super-resolution using deep convolutional networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 38(2): 295-307.

[10] 肖进胜, 刘恩雨, 朱力, 等. 改进的基于卷积神经网络的图像超分辨率算法[J]. 光学学报, 2017, 37(3): 96-104.

[11] 郭晓, 谭文安. 基于级联深度卷积神经网络的高性能图像超分辨率重构[J]. 计算机应用, 2017, 37(11): 3124-3127.

[12] Kim J, Kwon Lee J, Mu Lee K. Accurate image super-resolution using very deep convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 1646-1654.

[13] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.

[14] Nair V, Hinton G E. Rectified linear units improve restricted boltzmann machines[C]//Proceedings of the 27th international conference on machine learning (ICML-10). 2010: 807-814.

[15] Maas A L, Hannun A Y, Ng A Y. Rectifier nonlinearities improve neural network acoustic models[C]//Proc. icml. 2013, 30(1): 3.

[16] Kim J, Kwon Lee J, Mu Lee K. Deeply-recursive convolutional network for image super-resolution[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 1637-1645.

[17] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[18] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE transactions on image processing, 2004, 13(4): 600-612.

[19] Yang J, Wright J, Huang T S, et al. Image super-resolution via sparse representation[J]. IEEE transactions on image processing, 2010, 19(11): 2861-2873.

[20] Arbelaez P, Maire M, Fowlkes C, et al. Contour detection and hierarchical image segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2011, 33(5): 898-916.

[21] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks[C]//Proceedings of the thirteenth international conference on artificial intelligence and statistics. 2010: 249-256.

[22] Zeyde R, Elad M, Protter M. On single image scale-up using sparse-representations[C]//International conference on curves and surfaces. Springer, Berlin, Heidelberg, 2010: 711-730.

Single Image Super-resolution Reconstruction Based on Improved Deep Convolutional Neural Network

LIU Shi-hao, LI Jun

(College of Computer Science and Technology, Qingdao University, Qingdao 266071, China)

Abstract: In order to solve the problem that the existing super-resolution model can not restore the texture details of the image and the difficulty of model training, the existing model is improved by combining the existing residual network and the Inception module in GoogleNet. The original Inception module is improved by replacing the 5×5 convolution kernel with two cascaded 3×3 convolution kernels, using LeakyReLU as the activation function, and deleting the pooling layer, and then improved Inception module is cascaded multiple times in the model. The experimental results show that compared with the bicubic interpolation, SRCNN and VDSR algorithm, the improved model can obtain higher peak signal-to-noise ratio (PSNR) and structural similarity (SSIM), and also has obvious improvement in visual effects.

Keywords: deep convolutional neural network; image processing; super-resolution; Inception