

基于去卷积的快速图像超分辨率方法

孙超¹, 吕俊伟¹, 李健伟², 仇荣超¹

¹海军航空大学控制工程系, 山东 烟台 264001

²海军航空大学电子信息工程系, 山东 烟台 264001

摘要 针对目前基于样例学习的图像超分辨率方法难以同时满足快速运算和生成高质量图像的问题, 提出一种基于去卷积的快速图像超分辨率方法。设计新型网络模型, 以低分辨率图像作为输入图像, 利用卷积层进行特征提取与表示; 利用去卷积层对图像特征放大膨胀, 再以池化层浓缩特征图, 提炼出对结果更敏感的特征; 以亚像素卷积层实现特征映射与图像融合, 获得高分辨率图像。在图像集上进行测试, 相比其他方法, 本文方法的测试结果具有较高的峰值信噪比, 且平均每秒能处理 24 幅以上大小为 320 pixel×240 pixel 的图像, 表明该方法不仅可以生成更高质量的图像, 且具有较高的处理速度, 能满足视频实时处理要求。

关键词 图像处理; 超分辨率; 深度学习; 卷积神经网络

中图分类号 TN219 **文献标识码** A

doi: 10.3788/AOS201737.1210004

Method of Rapid Image Super-Resolution Based on Deconvolution

Sun Chao¹, Lü Junwei¹, Li Jianwei², Qiu Rongchao¹

¹Department of Control Engineering, Naval Aeronautical University, Yantai, Shandong 264001, China

²Department of Electronic and Information Engineering, Naval Aeronautical University,
Yantai, Shandong 264001, China

Abstract In view of existing problems of image super-resolution method based on sample-learning, which is difficult to operate rapidly and generate high quality image at the same time, a rapid image super-resolution method based on deconvolution is proposed. A new type of network model is designed and low resolution images are taken as input images directly, and then convolution layer is used to extract and represent features. Deconvolution layer is used to enlarge image feature maps, and the following pooling layer is used to concentrate the feature maps and extract features which are more sensitive to the results. Moreover, sub-pixel convolution layer is applied to features mapping and images fusion simultaneously and the super-resolution image could be obtained. The proposed method is tested on images of test datasets, and compared with other methods. The test results of the proposed method have higher peak signal to noise ratio (PSNR) and can process more than 24 images in size of 320 pixel×240 pixel per second, which shows that the proposed rapid image super-resolution method based on deconvolution can not only generate images with higher quality, but also satisfy the requirement of real-time video processing.

Key words image processing; super resolution; deep learning; convolution neural network

OCIS codes 100.6640; 100.2000; 110.3010

1 引 言

图像包含的信息量大小相当程度上取决于分辨率。提高图像的分辨率, 能够改善视觉效果, 使图像更符合人类或计算机的处理要求, 在监控设备、卫星图像和医学影像等领域都有举足轻重的应用价值。由于受到工艺水平等因素的限制, 难以单纯依靠硬件革新来提高图像的分辨率。因此利用软件技术, 通过图像处理的方式来提高图像分辨率显得尤为重要。

收稿日期: 2017-06-29; **收到修改稿日期:** 2017-08-01

基金项目: 国家自然科学基金(61471382, 61401495)、山东省自然科学基金(ZR2016FQ17)

作者简介: 孙超(1990—), 男, 博士研究生, 主要从事图像处理、目标识别方面的研究。E-mail: lemony1314@163.com

导师简介: 吕俊伟(1960—), 男, 博士, 教授, 主要从事模式识别、目标检测方面的研究。E-mail: ljwei369@163.com

由一帧或多帧低分辨率(LR)的观测图像来恢复高分辨率(HR)图像或图像序列的技术,称为图像的超分辨率(SR)技术^[1]。现有的超分辨率方法主要分为插值法、基于重建和基于学习 3 类。其中插值法最简单直观,但难以有效地重现图像细节信息,生成的图像比较模糊。基于重建的方法是利用先验知识,以图像退化模型为基础进行图像超分辨率重建,包括凸集投影法、最大后验概率法、迭代反投影法等。近年来,基于样例学习的图像超分辨率方法备受欢迎,如邻域嵌入^[2]、固定领域回归(ANR)^[3]、调整的固定领域回归(A+)^[4]方法等。Yang 等^[5-6]提出基于稀疏表示的超分辨率方法,利用稀疏编码(SC)将高低分辨率图像块分别表示为高低分辨率字典,通过机器学习建立字典之间的映射关系。随着深度学习的兴起,基于深度学习的超分辨率方法也得到了长足的发展,且与传统方法相比,具有更好的超分辨率效果。Dong 等^[7]率先提出了基于卷积神经网络的图像超分辨率(SRCNN)算法,采用深度学习直接对高低分辨率图像训练学习,实现了一个端到端的学习架构,省去了特征提取的预处理和后续的高分辨率图像块聚合过程。文献[8]引入更深的网络进行超分辨率计算,借鉴残差网络(ResNet)的思想减轻网络的学习负担,提升学习效率,并构造了一个适用放大不同尺度图片的网络。文献[9]提出使用更多的卷积层来增加网络的感受野,同时为了避免过多的网络参数,引入递归神经网络,效果相比 SRCNN 有很大的提升。郑向涛等^[10]提出一种自外而内的图像超分辨率复原算法,同时考虑外在样例和内在自相似性,利用图像的先验知识进行正则化约束。

但上述方法参数较多、计算量较大,导致运算速度较慢,无法对视频进行实时超分辨率重建。为此,Dong 等^[11]在 SRCNN 方法的基础上进行改进,以低分辨率图像作为网络输入,减小特征图维度和卷积核尺寸,提出了一种快速超分辨率(FSRCNN)方法,减少了网络训练参数,并相应地提出一种能够实时处理视频的简约版超分辨率(FSRCNN-s)方法。Shi 等^[12]提出一种利用亚像素卷积层的超分辨率(ESPCN)方法,将得到的特征图重新排列,得到高分辨率图像,大幅提高了运算速度,满足实时视频处理要求。然而这两种方法在缩减网络参数、减少计算量的同时,超分辨率性能相比采用更深网络的方法也有所下降。故需要寻求一种方法,既能满足算法的实时性要求,同时也具有良好的图像超分辨率性能。

本文提出一种基于去卷积的快速图像超分辨率(RSRD)方法。与同类实时的超分辨率方法相比,该方法具有更优异的超分辨率性能;与其他基于卷积神经网络的超分辨率方法相比,RSRD 方法不仅能够满足对视频进行实时超分辨率重建的要求,而且也具有较高的超分辨率性能,能够生成较高质量的高分辨率图像,仅次于利用更深的网络模型、以计算量与时间为代价的 VDSR^[8]、DRCN^[9]。

2 基于深度学习的图像超分辨率方法对比

基于深度学习的图像超分辨率方法主要分为 4 个步骤:1)构建训练所需的外部图像数据库;2)搭建合适的网络框架;3)利用样本数据库对网络进行训练,优化网络参数,获得图像的特征表示及先验知识;4)将低分辨率图像作为训练优化后网络的输入,得到最终的高分辨率图像。其中网络的搭建和训练是至关重要的环节,网络框架大多以卷积神经网络为模型。

与传统方法相比,该方法在取得更为理想效果的同时,也存在一些问题。

1)网络输入的低分辨率图像经过双三次插值(Bicubic)后,按比例放大到要求的图像尺寸,如图 1(a)~(c)中的 SRCNN^[7]、VDSR^[8]、DRCN^[9]方法。这些方法在网络的初始均存在一个预定义的上采样操作,这一操作相比以低分辨率图像直接作为输入时,其网络的运算复杂度呈比例增长,例如当放大比例为 f 时,以上采样后图像作为网络输入相比低分辨率图像作为输入,网络的参数增长了 $f^2 - 1$ 倍,训练及测试所花费的时间也随之增多。

2)现有超分辨率方法在处理图像时,图像在网络过程中的尺寸一般分为两种情况。一种是保持不变:以上采样后的图像作为网络输入,经过若干卷积层,最后得到与上采样后图像尺寸相同的高分辨率图像,如图 1(a)~(c)所示。另一种是成比例放大(网络末端):将低分辨率图像直接作为网络输入,经过若干卷积层,在最后一层通过某种方法进行比例放大,得到要求尺寸的高分辨率图像,如图 1(d)、(e)中的 ESPCN、FSRCNN 方法所示。前者会导致网络训练参数过多、计算量增大、运算速度较慢,无法对视频进行实时超分辨率重建;后者用时减少,能够对视频进行实时处理,但网络结构单一,且超分辨率效果相比前者有所下降。

3)为了保证输出图像的尺寸符合 2)中的要求,用于超分辨率方法的网络结构中通常仅含有卷积层、激

活层和损失函数层,而没有去卷积层、池化层等,如 SRCNN、VDSR。这是由于去卷积层相当于卷积层的逆操作使上一层网络得到的特征图尺寸变大;而池化层会使特征图的尺寸缩小,这两种层均会大幅改变网络最终输出图像的尺寸。

针对上述问题,提出 RSRD 方法,其网络结构如图 1(f)所示。以低分辨率图像作为网络输入,通过卷积层对图像块进行特征提取与表示;利用去卷积层和池化层的组合,共同完成网络的非线性映射阶段;再利用亚像素卷积层,以微小的计算量对特征图重新排列,组合之后得到最终的高分辨率图像。

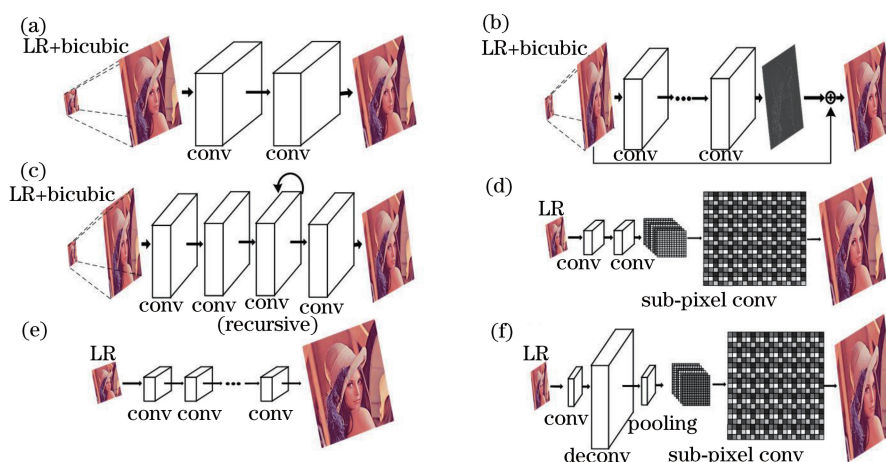


图 1 基于深度学习的图像超分辨率方法网络结构图。

(a) SRCNN; (b) VDSR; (c) DRCN; (d) ESPCN; (e) FSRCNN; (f) RSRD

Fig. 1 Network architectures of image super-resolution methods based on deep-learning.

(a) SRCNN; (b) VDSR; (c) DRCN; (d) ESPCN; (e) FSRCNN; (f) RSRD

相比现有的基于卷积神经网络的超分辨率方法,RSRD 主要区别于以下 3 个方面。

1) 网络结构。在基于卷积神经网络的超分辨率方法中加入去卷积层和池化层的组合,既能对网络训练过程中的特征图进行放大与收缩,提炼出对结果更为敏感的特征图,同时也能更好地完成网络的非线性映射。

2) 速度。所用的网络首端以低分辨率图像为输入图像,减少了网络的运算负担,提高了算法运行速度;中间采取去卷积层与池化层的组合,未引入更多的网络参数;末端利用亚像素卷积层对网络得到的特征图按特定顺序排列,以微小的工作量为代价,组合得到高分辨率图像。因此具有较高的运算速度,以大小为 $320 \text{ pixel} \times 240 \text{ pixel}$ 的图像 3 倍超分辨率重建为例,平均每秒能够处理 28.7 frame 图像,达到 24 frame/s 的视频实时处理要求。

3) 性能。采用去卷积层和池化层的组合在图像的超分辨率重建上有良好应用,性能不仅优于 FSRCNN、ESPCN 这 2 种快速图像超分辨率方法,也优于常规基于卷积神经网络的 SRCNN、SRCNN-Ex^[13],仅次于采用深度网络、以大量运算时间为代价的 VDSR、DRCN。图 2 展示了 RSRD 方法及其他经典方法在 3 倍放大尺度下,Set14 测试集的峰值信噪比(PSNR)与各自运算所花费的时间。

3 RSRD 方法

RSRD 方法不仅具有较高的超分辨率性能,同时也具有较快的运算速度,能够对视频进行实时超分辨率重建。网络结构如图 1(f)所示,分为特征提取、特征放缩、图像融合 3 个部分,与现有基于卷积神经网络的超分辨率方法的网络结构不同,利用去卷积层和池化层的组合搭配,对网络中的特征图进行放缩,不仅能够很好地完成网络的非线性映射,也能够提取出对结果更为敏感的特征图,使网络具有更优异的超分辨率性能。

3.1 特征提取

以未经过上采样操作的低分辨率图像作为网络的输入图像,先经过卷积层对图像进行特征提取,因为卷积运算的重要特点就是使原信号的特征增强,并且能够降低噪声影响。

用一组预先训练的基,例如离散余弦变换(DCT),哈尔变换(Haar)等,提取图像块的特征,这是图像重

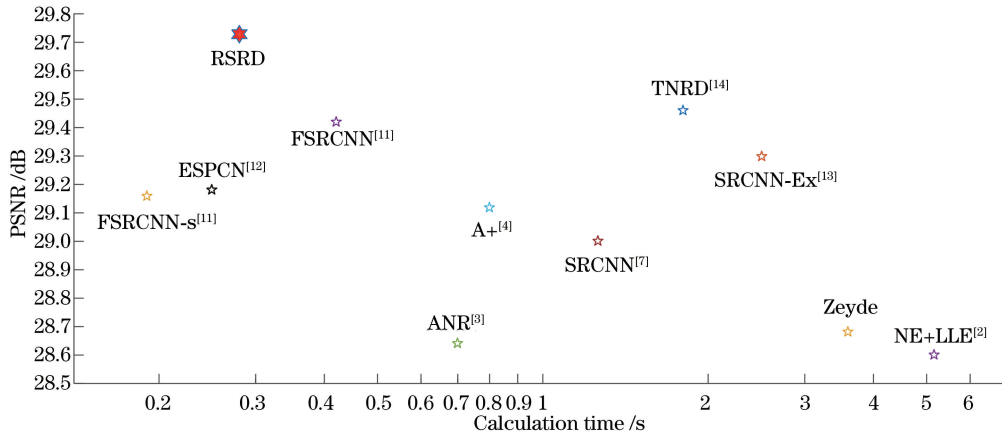


图2 当图像放大3倍时各超分辨率方法在Set14测试集上的PSNR与运算时间关系

Fig. 2 PSNR versus calculation time for different methods performing super-resolution over Set14 with magnification factor of 3

建领域中常用方式,即等效于用一组卷积核在图像上滑动(即卷积操作),得到一组特征图,该操作可表示为

$$F_1(Y) = g(W_1 * Y + B_1), \quad (1)$$

式中 W_1 表示一组数量为 n_1 、大小为 $c \times k_1 \times k_1$ 的卷积核(c 为输入图像的通道数量, k_1 为卷积核的尺寸), B_1 为 n_1 维的偏差,*为卷积操作, $g(x)$ 为激活函数, Y 为输入的低分辨率图像, $F_1(Y)$ 为经过卷积层后得到的 n_1 幅特征图。

由于人眼对图像的亮度信息更为敏感,因此在对低分辨率图像进行超分辨率重建时,与现有方法的处理方式相同。先将RGB图像转换成YCbCr图像,再对YCbCr图像中的Y通道(亮度通道)图像进行超分辨率处理,剩余2个通道则直接采取Bicubic方法进行插值放大到目标尺寸,3个通道的图像融合后即最终的高分辨率图像。因此,超分辨率训练网络的输入图像通道数 $c=1$ 。

激活函数 $g(x_i)$ 采用带参数的修正线性单元(PReLU),能够在一定程度上起到正则效果。相比修正线性单元(ReLU)而言,PReLU只增加了极少量的参数,但网络的收敛速度会有所提升,其公式表示为

$$g(x_i) = \begin{cases} x_i, & \text{if } x_i > 0 \\ \alpha_i x_i, & \text{if } x_i < 0 \end{cases}, \quad (2)$$

也可表示为

$$g(x_i) = \max(0, x_i) + \alpha_i \min(0, x_i), \quad (3)$$

式中 α_i 为可学习参数,初始化 α 为0.25,当网络反向更新参数时,更新公式可表示为

$$\Delta\alpha_{i+1} = \mu\Delta\alpha_i + \epsilon \frac{\partial E}{\partial \alpha_i}, \quad (4)$$

其中 μ 和 ϵ 分别为动量和学习率, E 为损失函数。

借鉴牛津大学VGG小组提出的基于卷积神经网络的视觉识别算法[15]采用的VGG-ILSVRC-16-layers模型,网络中卷积层选取 3×3 的卷积核,即设置 $k_1=3$,相比SRCNN方法在第1层选取的 9×9 的卷积核,减少了参数的数量,也能够提取足够的图像特征。

对图像进行卷积时,若不对边界进行零填充(zero padding, p),每层得到的特征图尺寸相比之前会减小,使得最后通过网络训练的输出图像尺寸小于输入图像。如在SRCNN中, $p=0$ 导致训练网络中输出图像尺寸小于输入图像,输出、输入图像尺寸分别为 $21 \text{ pixel} \times 21 \text{ pixel}$ 、 $33 \text{ pixel} \times 33 \text{ pixel}$ 。这意味着SRCNN只能有效利用图像中间部分的信息,可能忽略了图像边缘的信息。因此在卷积操作前,对边界进行零填充,即 $p=(k_1-1)/2=1$,假定网络输入图像的尺寸为 $H \times W$,则输入图像经过卷积层后得到的特征图与输入图像尺寸保持不变,均为 $H \times W$ 。

3.2 特征放缩

在图像的超分辨率重建过程中,需要对图像提取出的特征进行非线性映射,从低层特征逐步抽象到高层

特征,这也是图像重建过程中最重要、对性能影响最大的一个环节。现有基于卷积神经网络的图像超分辨率方法中,均是利用一系列的卷积层和激活层的组合来实现非线性映射过程。

利用去卷积层与池化层的组合对图像卷积得到的特征图进行放缩、提炼,提取出对结果更为敏感的特征,更好地完成超分辨率网络的非线性映射阶段。

1)利用去卷积层,将卷积层得到的特征图按设计的比例放大。输入图像经过卷积层卷积后,特征图尺寸与输入图像保持不变,均为 $H \times W$,然后经过去卷积层,特征图尺寸呈比例增大,表示为

$$F_2(Y) = g [W_2 \otimes F_1(Y) + B_2] \quad (5)$$

式中 W_2 为一组数量为 n_2 ,大小为 $n_1 \times k_2 \times k_2$ 的去卷积核(k_2 为去卷积核的尺寸); B_2 为 n_2 维的偏差; \otimes 为去卷积操作; $F_2(Y)$ 为经过去卷积层处理之后得到的 n_2 幅特征图。

设置去卷积核的步长为 s_2 ,边缘零填充为 p_2 ,当图像放大因子为 f 时,去卷积操作后的特征图尺寸则为 $[s_2(H-1)-2p_2+qk_2] \times [s_2(W-1)-2p_2+k_2] = fH \times fW$ 。实验中,当 $f=2$ 时,分别设置 $s_2=2, p_2=2, k_2=6$;当 $f=3$ 时, $s_3=3, p_3=2, k_3=7$;当 $f=4$ 时, $s_4=2, p_4=2, k_4=6$ 。

2)利用池化层可以减少输出的特征向量、降低特征图的维度,同时提高网络的训练速度,而且能够改善网络的训练结果,不容易出现过拟合的情况^[16],该操作公式表示为

$$F_3(Y) = W_3 \odot F_2(Y) \quad (6)$$

式中 W_3 表示大小为 $n_2 \times k_3 \times k_3$ 的滤波器(k_3 为池化层滤波器的尺寸); \odot 为池化操作; $F_3(Y)$ 为经过去卷积层之后得到的 n_2 幅特征图。

设置池化层的滤波器大小为 $k_3=f$,步长为 f ,池化方式采用最大池化。经过池化层之后,特征图的尺寸由 $fH \times fW$ 重新恢复为 $H \times W$,即通过去卷积层和池化层的组合之后,尺寸恢复初始状态,与网络的输入图像大小仍保持一致。

去卷积层和池化层组合,相当于对提取的特征图进行了膨胀再浓缩操作。去卷积层是对特征的放大、膨胀操作,池化层是对特征的缩小、浓缩操作,两者结合能够提取出对结果更为敏感、有效的特征,因此能够更高效地完成网络的非线性映射阶段。

3.3 图像融合

在网络的图像融合阶段,将上一阶段得到的特征图经过亚像素卷积层,得到最终的高分辨率图像:

$$Y_{SR} = F_4(Y) = R [W_4 * F_3(Y) + B_4] \quad (7)$$

式中 $F_4(Y)$ 、 Y_{SR} 为网络最终得到的高分辨率图像, W_4 为一组数量为 n_4 、大小为 $n_2 \times k_4 \times k_4$ 的卷积核, B_4 为偏差, $R(x)$ 为排列操作。设置 $n_4=f^2, k_4=3$,步长为 1,边缘零填充为 1。

亚像素卷积层实际包含 2 个步骤。

1)卷积。与网络初始经过的卷积层相同,经过卷积后得到的特征图尺寸大小与输入图像相同,但特征图的通道数变为 f^2 个。

2)排列。将 f^2 个通道中所有相应位置的像素按预先设定的顺序进行排列,组合成一个大小为 $f \times f$ 的图像区域,该图像区域即对应于所求高分辨率图像中的一个 $f \times f$ 大小的图像子块。通过这种方式,将最后得到的大小为 $f^2 \times H \times W$ 的特征图重新排列,最终组合成一个 $1 \times fH \times fW$ 的高分辨率图像。这一操作仅仅是对图像的重新排列,没有进行卷积运算,因而花费的时间很少。

3.4 损失函数

网络的输入图像在经过特征提取与表示、特征放缩、图像融合 3 个阶段之后,最后得到高分辨率图像。通常情况下,用一组真实图像 $\{X_i\}$ 和网络预测图像 $\{Y_{SR}\}$ 的均方误差(MSE)来表示损失函数,利于后期以 PSNR 作为高分辨率图像的质量评价指标,PSNR 值越高,说明图像失真越小,即图像的质量越高。该损失函数公式表示为

$$L = \frac{1}{n} \sum_{i=1}^n \|Y_{SR} - X_i\|^2 \quad (8)$$

式中 n 为训练集的数目。

4 实验设置

实验所用计算机搭载的是 Intel Core i7-6700K 的 CPU, 含 16G 内存, 并用英伟达 GTX1080 的显卡进行加速, 软件配置的是 64 位 Ubuntu 16.04 LTS 的操作系统和 Caffe、Mtalab 2014a、CUDA 8.0、Cudnn 5.1、OpenCV2.7。

4.1 数据集

在将 RSRD 方法与现有方法进行比较、分析时, 实验采用文献[4]中提供的公开数据集 Timofte dataset, 其中包括 91 幅训练图像以及 2 个测试数据集 Set5 和 Set14, 该数据集被应用于众多的单幅图像超分辨率方法测试中。Berkeley Segmentation Datasets 中的 BSD300 及 BSD500 也分别提供了 100、200 幅图像用作测试^[17]。

训练集的大小在一定程度上能够影响超分辨率方法的实际性能, 训练集越大, 方法训练后的超分辨率效果越好。因此以 Timofte dataset 中的 91 幅图像作为网络训练优化阶段的训练集, 在优化之后的模型中, 采用文献[8]中使用的 291 幅图像作为训练图像, 以此来提升超分辨率方法的效果。这些图像均具有良好的画质, 适用于超分辨率的网络训练。

4.2 训练样本及策略

参照文献[11]中的方法, 首先将训练样本集中的图像按比例进行降采样, 从而得到对应的低分辨率图像; 然后设置子图像大小为 $f_{\text{sub}}=7$, 按步长为 3 进行裁剪, 将训练集裁剪成一系列的子图像作为网络训练阶段的输入。与文献[11]中策略相同, 通过 2 种方式对训练图像的数量进行增强: 1) 缩放, 对每幅图像进行降采样, 缩放到原先尺寸的 60%、70%、80%、90%; 2) 旋转, 将每幅图像分别旋转 90°、180°、270°, 如此便能够将训练图像的数量增大 19 倍 ($5 \times 4 - 1 = 19$)。

网络的学习策略采用随机梯度下降法 (SGD), 利用负梯度 $\nabla L(W)$ 、上一次的更新值 V_i 和当前权重 W_i 来计算当前的更新值 V_{i+1} 和权重 W_{i+1} , 即

$$V_{i+1} = \mu V_i - \epsilon \nabla L(W_i), W_{i+1} = W_i + V_{i+1}, \quad (9)$$

式中动量 μ 为 0.9, 学习率 ϵ 为 10^{-3} , 当训练过程中损失达到稳定时, 学习率 ϵ 缩小为原先的 1/10, 多次重复该过程。

网络整体的训练过程如下: 1) 生成训练样本数据集; 2) 调用网络的前向过程, 训练样本依次经过卷积层、激活层、去卷积层、激活层、池化层、亚像素卷积层、损失层, 最终计算网络的输出和损失; 3) 调用网络的后向过程, 计算梯度; 4) 根据随机梯度下降法, 利用梯度更新参数, 对网络进行迭代优化; 5) 重复 2)~4) 的过程, 直到网络的损失达到最小, 则表明网络性能已达到最优。

4.3 网络框架的细节选择

为了测试网络的不同设置对图像超分辨率性能的影响, 从 3 个方面进行了不同的设置和相应的实验测试, 以期找出最优设置。对比实验以 $f=3$ 为例, 分别以 Set5 和 Set14 作为测试集, 求出测试集中各图像峰值信噪比的均值, 并作出该值随迭代次数增长的曲线。

1) 层数选择

通常图像超分辨率方法中的非线性映射阶段由一系列的卷积层组成, 适当增加卷积层的个数, 能够提高网络的超分辨率性能。RSRD 方法的非线性映射阶段由去卷积层与池化层组成, 而网络中去卷积层与池化层组合的个数会影响网络的超分辨率性能。为此进行两组观察实验, 分别测试一个去卷积层与池化层组合和两个去卷积层与池化层组合的网络性能, 测试结果如图 3 所示。由图 3 可以看出, 在网络训练的初始阶段, 2 个去卷积层与池化层组合的性能更佳, 但随着迭代次数的增多, 两者差距越来越小, 最后只保持了可以忽略的微弱性能优势。因此综合考量下, 仅选取一个去卷积层与池化层组合实现网络功能。

2) 去卷积核尺寸选择

卷积核、去卷积核尺寸的不同能够直接改变网络的感受野, 从而影响到网络的超分辨率性能。仅对网络中去卷积尺寸的选取进行研究, 分别以 $k_2=5$ 与 $k_2=7$ 进行实验, 实验结果如图 4 所示。由图 4 可以看出, 当去卷积核尺寸 $k_2=7$ 时, 网络的超分辨率性能更佳, 在训练初始时就保持领先; 随着迭代次数增加, 优势逐渐增大, 当训练趋于结束时, PSNR 值在 Set5 与 Set14 测试集上分别领先 $k_2=5$ 时 0.16 dB、0.06 dB。因此,

最终设定网络的去卷积核尺寸 $k_2=7$ 。

3) 激活函数选择

深度神经网络具有丰富表达能力的影响因素,除了“深”之外,还有一个重要因素即非线性处理单元,也就是激活函数。不同的激活函数对网络的超分辨率性能的影响不同。分别选取 tanh、ReLU、PReLU 3 种激活函数进行测试,测试结果如图 5 所示。由图 5 可以看出,激活函数 PReLU 表现最优,ReLU 次之,因此最终选取 PReLU 函数作为网络的激活函数。

综合以上考量,RSRD 方法最后采用的训练网络如图 6 所示,该图为 3 倍放大比例下的训练网络。

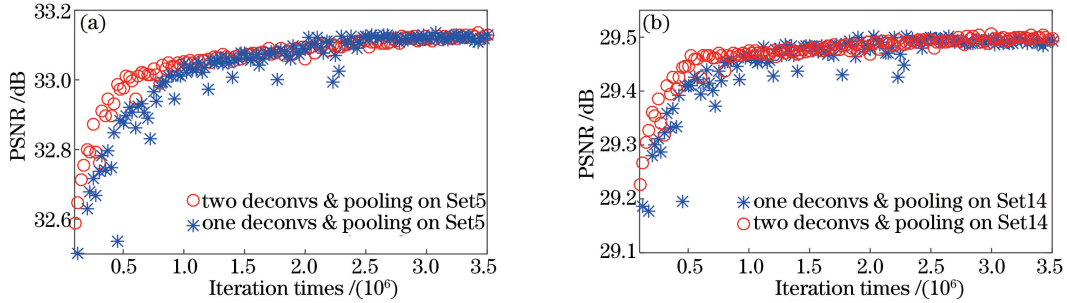


图 3 不同层数下测试集的 PSNR 均值随迭代次数增长的趋势图。(a) Set5; (b) Set14

Fig. 3 Trend of mean PSNR of dataset with iteration rising under different layers. (a) Set5; (b) Set14

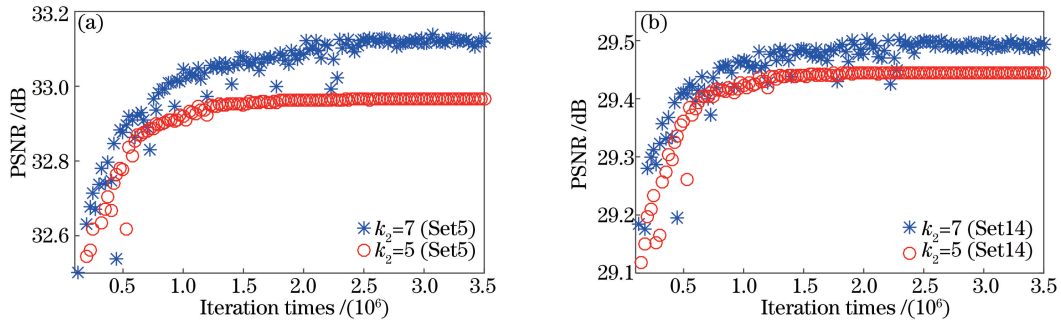


图 4 不同去卷积核尺寸下测试集的 PSNR 均值随迭代次数增长的趋势图。(a) Set5; (b) Set14

Fig. 4 Trend of the mean PSNR of dataset with iteration rising under different deconvolution kernel sizes. (a) Set5; (b) Set14

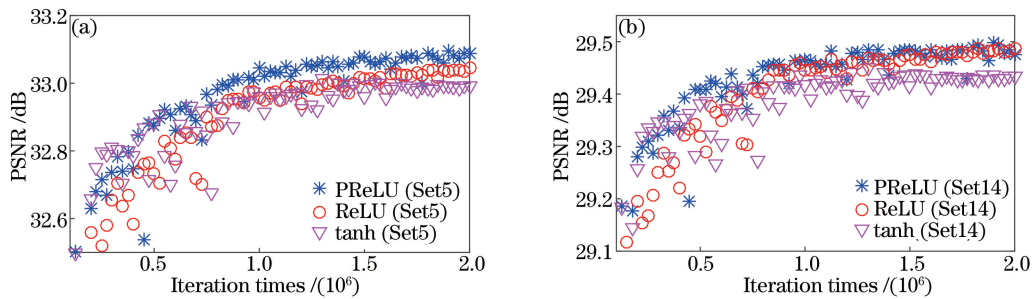


图 5 不同激活函数下测试集的 PSNR 均值随迭代次数增长的趋势图。(a) Set5; (b) Set14

Fig. 5 Trend of the mean PSNR of dataset with iteration rising under different active functions. (a) Set5; (b) Set14

5 结果及分析

为了验证 RSRD 方法的图像超分辨率性能,实验采用超分辨率重建领域通用的测试集 Set5、Set14、BSD100、BSD200 进行测试,并将测试结果与 Bicubic、RFL^[17]、SRCNN^[7]、SRCNN-Ex^[13]、SCN^[18]、FSRCNN-s^[10]、FSRCNN^[10]、ESPCN^[11]、Self-Ex^[19]、VDSR^[8]、DRCN^[9] 方法对比。

表 1、表 2 分别展示了当训练集为 91 幅^[4]、291 幅^[8] 图像时,各方法分别在 2 倍、3 倍、4 倍图像放大比例下,在各图像测试集上所有图像经超分辨率重建后 PSNR 的平均值。从表 1 和表 2 可以看出,RSRD 方法相比传统基于样例学习,但无法对视频进行实时重建的超分辨率方法(如 Self-Ex^[19]、RFL^[17]、SCN^[18]、

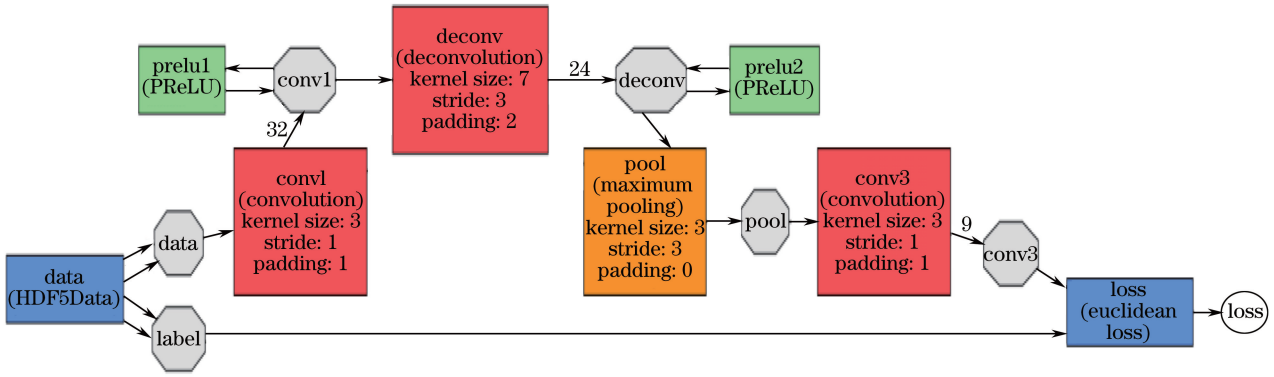


图 6 适用于 3 倍图像放大比例下的训练网络

Fig. 6 Proposed training network applied to magnification factor of 3

SRCNN、SRCNN-Ex 等), 具有更高的 PSNR 值, 即能够重建出与原图更为相似的结果图, 具有更好的图像超分辨率性能; 相比快速超分辨率 FSRCNN-s、ESPCN 方法, RSRD 方法在满足对视频实时处理性的前提下, PSNR 值也得到了较大的提升, 图像超分辨率重建能力也得到了增强。从表 2 可以看出, 相比采用更深的训练网络、以计算量和时间为代价的 VDSR、DRCN 方法, RSRD 方法在各图像集上的测试结果相对较低, 图像重建能力略逊一筹, 但为了具有较快的图像处理速度、能够达到实时处理视频的要求, 采取了较为简便的网络结构。

表 1 以 91 幅图像为训练集, 不同放大比例下各测试集经不同方法处理后峰值信噪比的平均值

Table 1 Mean PSNR of test datasets processed by all methods with different magnification factors trained by 91 images

Dataset	Scale	Bicubic	RFL	SRCNN	SRCNN-Ex	SCN	FSRCNN-s	FSRCNN	ESPCN	RSRD
Set5	2	33.66	36.54	36.33	36.67	36.76	36.53	36.94	36.74	37.22
Set14	2	30.23	32.26	32.15	32.35	32.48	32.22	32.54	32.47	32.83
BSD100	2	29.56	31.36	31.32	31.50	31.49	31.30	31.42	31.27	31.58
BSD200	2	29.70	31.38	31.34	31.53	31.63	31.44	31.73	31.53	31.88
Set5	3	30.39	32.43	32.45	32.83	33.04	32.55	33.06	32.55	33.15
Set14	3	27.54	29.05	29.01	29.26	29.37	29.08	29.37	29.09	29.51
BSD100	3	27.21	28.22	28.21	28.4	28.49	28.27	28.4	28.26	28.49
BSD200	3	27.26	28.25	28.27	28.47	28.54	28.32	28.55	28.34	28.58
Set5	4	28.42	30.14	30.15	30.45	30.64	30.04	30.55	30.27	30.74
Set14	4	26.00	27.24	27.21	27.44	27.62	27.12	27.50	27.39	27.69
BSD100	4	25.96	26.75	26.69	26.83	26.95	26.71	26.90	26.81	26.98
BSD200	4	25.97	26.76	26.72	26.88	26.96	26.73	26.92	26.82	26.99

表 2 以 291 幅图像为训练集, 不同放大比例下各测试集经不同方法处理后峰值信噪比的平均值

Table 2 Mean PSNR of test datasets processed by all methods with different magnification factors trained by 291 images

Dataset	Scale	RFL	Self-Ex	SRCNN	SCN	FSRCNN-s	FSRCNN	ESPCN	VDSR	DRCN	RSRD
Set5	2	36.54	36.67	36.49	36.65	36.71	37.02	36.82	37.49	37.58	37.30
Set14	2	32.28	32.35	32.22	32.29	32.36	32.67	32.58	33.00	33.01	32.96
BSD100	2	31.21	31.49	31.18	31.36	31.50	31.68	31.59	31.85	31.80	31.75
BSD200	2	37.26	31.51	31.20	31.39	31.52	31.82	31.65	31.90	31.86	32.08
Set5	3	32.58	32.58	32.58	32.75	32.81	33.19	32.87	33.62	33.77	33.42
Set14	3	29.13	29.13	29.16	29.28	29.36	29.46	29.40	29.73	29.70	29.74
BSD100	3	28.29	28.30	28.29	28.41	28.41	28.52	28.49	28.77	28.73	28.69
BSD200	3	28.31	28.32	28.34	28.48	28.49	28.64	28.56	28.80	28.76	28.80
Set5	4	30.28	30.3	30.31	30.48	30.53	30.75	30.71	31.30	31.49	31.12
Set14	4	27.32	27.33	27.40	27.49	27.50	27.62	27.51	27.95	27.97	28.00
BSD100	4	26.82	26.86	26.84	26.90	26.94	27.02	26.98	27.20	27.18	27.18
BSD200	4	26.82	26.89	26.88	26.91	26.94	27.04	26.99	27.21	27.20	27.20

表 3 展示了与现有一些方法在网络输入、层数、不同层的使用情况、效果、实时性上的对比情况。从图 3

可以看出,利用去卷积层和池化层组合来实现网络的非线性映射,不仅能够实时对图像进行超分辨率重建,且超分辨率性能也领先于传统的基于卷积神经网络的超分辨率方法,仅次于 VDSR 和 DRCN 方法。

表 3 RSRD 方法与其他方法的对比

Table 3 Comparisons between the proposed RSRD and other methods

Method	Network input	Layer	Deconvolution	Pooling	Real-time	Accuracy	Speed
SRCNN	LR + bicubic	3	No	No	No	No. 7	No. 5
VDSR	LR + bicubic	20	No	No	No	No. 2	No. 6
DRCN	LR + bicubic	5(recursive)	No	No	No	No. 1	No. 7
ESPCN	LR	3	No	No	Yes	No. 5	No. 2
FSRCNN-s	LR	5	Yes	No	Yes	No. 6	No. 1
FSRCNN	LR	8	Yes	No	No	No. 4	No. 4
RSRD(ours)	LR	4	Yes	Yes	Yes	No. 3	No. 3

为了更直观地显示 RSRD 方法与其他方法的效果对比,图 7~9 分别展示了 2 幅测试集中图像与 1 幅实际采集图像经 Bicubic、ANR、SRCNN、A+、ESPCN、FSRCNN-s 与 RSRD 方法进行超分辨率重建后的效果图,并取局部进行放大对比。

图 7 为 3 倍放大比例下,7 种方法对 BSD200 中 118031.bmp 图像的重建结果,以及对屋顶转轮与天空交接部分的局部放大图。由图 7 可以看出,除了图 7(e)中 A+与图 7(h)中的 RSRD 方法外,其他方法的结果均无法辨别出转轮向外伸展的支撑木棍。与 A+方法相比,RSRD 方法能够产生更清晰、锐利的木棍和屋顶边缘,有更好的辨识度,与原图最相似,具有最好的视觉效果。

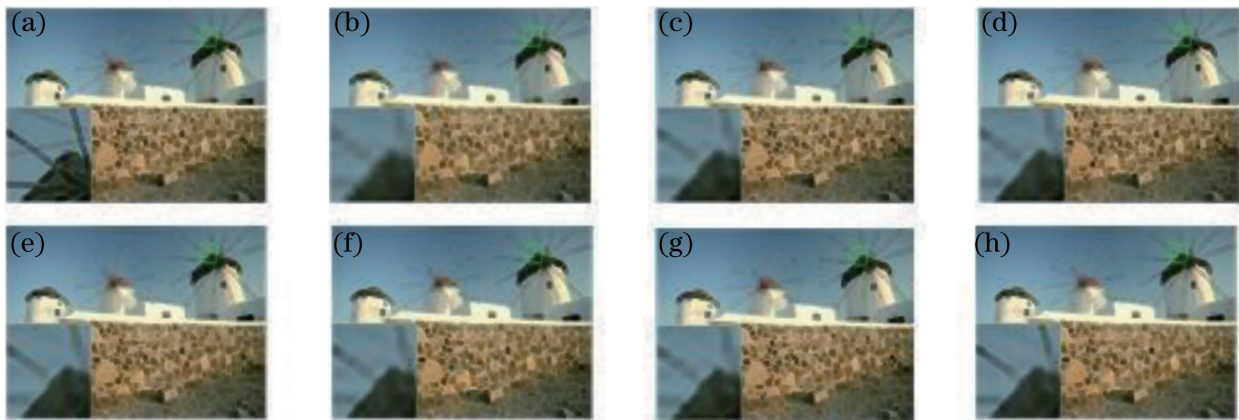


图 7 3 倍放大比例下,BSD200 测试集中 118031.bmp 经各方法处理后的整体及局部效果对比图。

(a) 真实图像; (b) Bicubic(24.82 dB); (c) ANR(25.40 dB); (d) SRCNN(25.70 dB);

(e) A+(25.70 dB); (f) ESPCN(25.81 dB); (g) FSRCNN-s(25.74 dB); (h) RSRD(26.18 dB)

Fig. 7 Whole and local comparisons of 118031.bmp in BSD200 processed by different methods with magnification factor of 3.

(a) Real image; (b) Bicubic(24.82 dB); (c) ANR(25.40 dB); (d) SRCNN(25.70 dB);

(e) A+(25.70 dB); (f) ESPCN(25.81 dB); (g) FSRCNN-s(25.74 dB); (h) RSRD(26.18 dB)

图 8 为 4 倍放大比例下,7 种方法对 Set14 中 foreman.bmp 图像的重建结果,以及对人物帽子与背后建筑相接部分的局部放大图。人物背后建筑有黑色条纹区域,在原图中能够清晰辨别,Bicubic 结果最为模糊,难以分辨物体;ANR、FSRCNN-s 方法的结果略好,但仍无法辨识出目标;SRCNN、ESPCN 方法的结果边缘较为模糊;A+与 RSRD 方法的结果有相对清晰的边缘,但 RSRD 方法结果的对比度更强,更为锐利。

图 9 为 2 倍放大比例下,7 种方法对实际采集图像(拍摄目标为烟台海润一品住宅楼)的重建结果,以及取屋顶与蓝天交界处进行了 15 倍的局部放大图。从图 9 可以看出,Bicubic 处理得到的图像视觉效果最差,PSNR 值仅为 29.88 dB,ANR、SRCNN、A+方法的结果略好,PSNR 值相比 Bicubi 方法也有所提高,但屋顶边缘仍比较模糊;FSRCNN-s、ESPCN 方法的结果边缘较优,相对清晰,图像的 PSNR 值也达到 31.85 dB;而 RSRD 方法的结果相比其他方法,有最为清晰的物体边缘,对比度也最为明显,有良好的图像辨识度,且 PSNR 值为 32.47 dB,相比 FSRCNN-s、ESPCN 方法提高了 0.62 dB。

收集 100 幅大小为 320 pixel×240 pixel 的图像,在配置单一 CPU、时钟频率为 2.6 GHz 的计算机上进



图 8 4 倍放大比例下,Set14 测试集中 foreman.bmp 经各方法处理后的整体及局部效果对比图。

(a) 真实图像; (b) Bicubic(29.57 dB); (c) ANR(30.80 dB); (d) SRCNN(31.50 dB);

(e) A+(32.20 dB); (f) ESPCN(32.02 dB); (g) FSRCNN-s(31.52 dB); (h) RSRD(32.89 dB)

Fig. 8 Whole and local comparisons of foreman.bmp in Set14 processed by different methods with magnification factor of 4.

(a) Real image; (b) Bicubic(29.57dB); (c) ANR(30.80 dB); (d) SRCNN(31.50 dB);

(e) A+(32.20 dB); (f) ESPCN(32.02 dB); (g) FSRCNN-s(31.52 dB); (h) RSRD(32.89 dB)

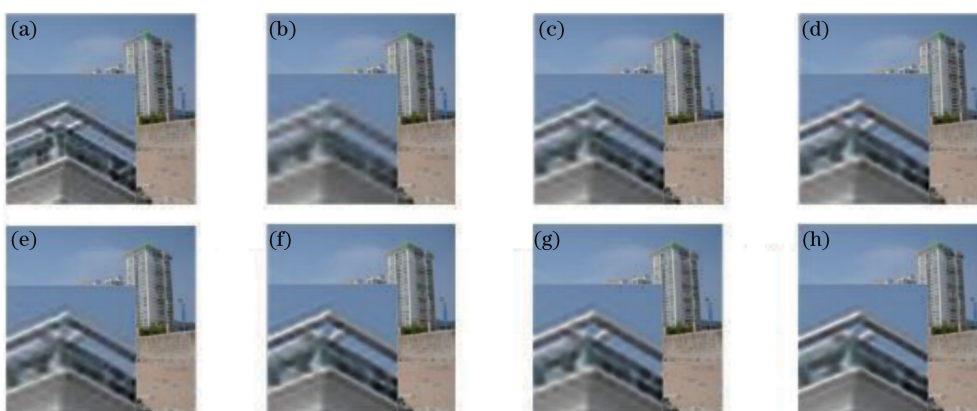


图 9 2 倍放大比例下,真实采集图像经各方法处理后的整体及局部效果对比图。

(a) 真实图像; (b) Bicubic(29.88 dB); (c) ANR(30.90dB); (d) SRCNN(31.65dB);

(e) A+(31.30dB); (f) ESPCN(31.85 dB); (g) FSRCNN-s(31.85 dB); (h) RSRD(32.47 dB)

Fig. 9 Whole and local comparisons of real image processed by different methods with magnification factor of 4.

(a) Real image; (b) Bicubic(29.88 dB); (c) ANR(30.90 dB); (d) SRCNN(31.65 dB);

(e) A+(31.30 dB); (f) ESPCN(31.85 dB); (g) FSRCNN-s(31.85 dB); (h) RSRD(32.47 dB)

行超分辨率重建。2 倍、3 倍、4 倍放大比例下,所花费的时间分别为 3.32,3.48,3.73 s,即平均每秒分别处理 30.1,28.7,26.8 frame 图像,满足视频 24 frame/s 图像的处理要求。

6 结论

针对现有基于卷积神经网络的超分辨率方法网络结构形式单一、难以兼顾速度与精度的问题,提出了一种基于去卷积的快速图像超分辨率,即 RSRD 方法,并设计了一种新型网络模型。对大小为 $320 \text{ pixel} \times 240 \text{ pixel}$ 的图像进行测试,能够每秒对 24 frame 以上图像进行超分辨率重建;并与其他方法在通用测试集上进行了对比实验,RSRD 方法生成的图像具有更锐利的边缘和清晰的轮廓,有更好的图像辨识度,且具有更高的峰值信噪比。实验结果表明 RSRD 方法具有较快的处理速度、更优异的超分辨率性能,但与采用更深网络的超分辨率方法相比仍略逊一筹。因此下一步研究方向是在保证良好处理速度的前提下,进一步优化网络模型,提高网络的超分辨率性能。

参 考 文 献

- [1] Su Heng, Zhou Jie, Zhang Zhihao. Survey of super-resolution image reconstruction methods[J]. Acta Automatica Sinica, 2013, 39(8): 1202-1213.
苏衡, 周杰, 张志浩. 超分辨率图像重建方法综述[J]. 自动化学报, 2013, 39(8): 1202-1213.
- [2] Chang H, Yeung D Y, Xiong Y. Super-resolution through neighbor embedding[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2004, 1: 275-282.
- [3] Timofte R, de Smet V, van Gool L. Anchored neighborhood regression for fast example-based super-resolution[C]. IEEE International Conference on Computer Vision, 2013: 1920-1927.
- [4] Timofte R, de Smet V, van Gool L. A+: Adjusted anchored neighborhood regression for fast super-resolution[C]. Asian Conference on Computer Vision, 2014, 9006: 111-126.
- [5] Yang J, Wright J, Huang T, *et al.* Image super resolution via sparse representation[J]. IEEE Transactions on Image Processing, 2010, 19(11): 2861-2873.
- [6] Yang J, Wright J, Ma Y, *et al.* Image super-resolution as sparse representation of raw image patches[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2008: 1-8.
- [7] Dong C, Loy C C, He K, *et al.* Learning a deep convolutional network for image super-resolution[C]. European Conference on Computer Vision, 2014: 184-199.
- [8] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 1646-1654.
- [9] Kim J, Lee J K, Lee K M. Deeply-recursive convolutional network for image super-resolution[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 1637-1645.
- [10] Zheng Xiangtao, Yuan Yuan, Lu Xiaoqiang. Single image super-resolution restoration algorithm from external example to internal self-similarity[J]. Acta Optica Sinica, 2017, 37(3): 0318006.
郑向涛, 袁媛, 卢孝强. 自外而内的单幅图像超分辨率复原算法[J]. 光学学报, 2017, 37(3): 0318006.
- [11] Dong C, Loy C C, Tang X. Accelerating the super-resolution convolutional neural network[C]. European Conference on Computer Vision, 2016: 391-407.
- [12] 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]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 1874-1883.
- [13] Dong C, Chen C L, He K, *et al.* Image super-resolution using deep convolutional networks[J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2016, 38(2): 295-307.
- [14] Chen Y, Pock T. Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(6): 1256-1272.
- [15] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J/OL]. Computer Science, 2014. [2017-06-10]. <https://arxiv.org/pdf/1409.1556.pdf>.
- [16] He K, Zhang X, Ren S, *et al.* Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1906.
- [17] 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]. IEEE International Conference on Computer Vision, 2001: 416-423.
- [18] Schulter S, Leistner C, Bischof H. Fast and accurate image upscaling with super-resolution forests[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3791-3799.
- [19] Wang Z, Liu D, Yang J, *et al.* Deeply improved sparse coding for image super-resolution[C]. IEEE International Conference on Computer Vision, 2015: 370-378.
- [20] Huang J B, Singh A, Ahuja N. Single image super-resolution from transformed self-exemplars[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2015: 5197-5206.