

# 基于 CNN 的图像超分辨率重建方法

王 容, 张永辉, 张 健, 张帅岩

(海南大学 信息科学技术学院, 海南 海口 570228)

**摘 要:** 为同时满足较好的超分辨率重建效果和实时处理要求, 提出一种基于 CNN 的图像超分辨率重建方法。以低分辨率图像作为输入, 采用  $1 \times 1$  的小卷积核进行降维和扩维处理, 减少网络的参数, 利用反卷积与池化的组合提取出对结果更为敏锐的特征, 通过反卷积进行上采样重建, 易于实现图像不同比例放大。实验结果表明, 相比 FSRCNN-s、ESPCN 等方法, 该方法实现了更好的重建效果, 平均每秒能处理 24 张以上尺寸为  $320 \times 240$  的图像, 满足对视频超分辨率重建的实时性要求。

**关键词:** 图像处理; 卷积神经网络; 反卷积; 池化; 超分辨率重建

**中图法分类号:** TP391.41 **文献标识号:** A **文章编号:** 1000-7024 (2019) 06-1654-06

**doi:** 10.16208/j.issn1000-7024.2019.06.027

## Image super-resolution reconstruction based on CNN

WANG Rong, ZHANG Yong-hui, ZHANG Jian, ZHANG Shuai-yan

(College of Information Science and Technology, Hainan University, Haikou 570228, China)

**Abstract:** To satisfy better super-resolution reconstruction effects and real-time processing requirements at the same time, an image super-resolution reconstruction method based on convolutional neural network was proposed. The low-resolution image was used as input, and the  $1 \times 1$  small convolution kernel was used for dimensionality reduction and expansion to reduce the network parameters, the combination of deconvolution and pooling was used to extract features that were more sensitive to the results, deconvolution was used for upsampling and reconstruction, and it was easy to achieve different scales of image magnification. Experimental results show that compared with FSRCNN-s, ESPCN and other methods, the proposed method not only achieves better reconstruction effects, but processes more than 24 images with a size of  $320 \times 240$  per second, which meet the real-time requirements for video super-resolution reconstruction.

**Key words:** image processing; convolutional neural network; deconvolution; pooling; super-resolution reconstruction

## 0 引 言

基于插值的方法、基于重建的方法以及基于学习的方法是当前对图像超分辨率研究的 3 个主要方向。基于插值的方法最为简单且运算速度快, 如双三次插值法 (bicubic<sup>[1]</sup>), 但得到的高分辨率图像边缘模糊; 基于重建的方法包括最大后验概率估计法、迭代反向投影法<sup>[2]</sup>等; 基于学习的方法近年来备受追捧, 其中基于邻居嵌入的方法<sup>[3-5]</sup>和基于稀疏表示的方法<sup>[6]</sup>都能得到较好的超分辨率性能。目前, 基于深度学习的超分辨率方法的研究也取得了显著的

成果。Dong 等<sup>[7]</sup>把卷积神经网络应用到图像超分辨率重建中, 提出了 SRCNN, 但重建效果不佳, 运算速度较慢。随后, Dong 等<sup>[8]</sup>在 SRCNN 的基础上进行改进, 提出了一种能够实时处理视频的超分辨率方法 FSRCNN-s, FSRCNN-s 的网络运算速度可达到 SRCNN 的 17.36 倍, 但超分辨率重建效果与 SRCNN 相近, 无法同时满足高质量和实时性要求。

本文提出一种基于 CNN 的图像超分辨率重建方法, 直接将低分辨率图像输入至网络, 通过卷积操作提取图像特征, 利用反卷积与池化的组合完成非线性映射, 最后通过

收稿日期: 2018-04-23; 修订日期: 2019-04-24

基金项目: 海南省自然科学基金项目 (618MS027); 海南大学大学生创新创业基金项目 (Hdcxcyxm201704)

作者简介: 王容 (1994-), 女, 湖南衡阳人, 硕士研究生, 研究方向为深度学习、嵌入式系统; 张永辉 (1974-), 男, 山东德州人, 博士, 教授, 博士生导师, 研究方向为嵌入式系统、智能检测技术、数字信号处理和过程控制等; 张健 (1974-), 男, 湖南长沙人, 博士研究生, 讲师, 研究方向为机器视觉; 张帅岩 (1993-), 男, 河南许昌人, 硕士研究生, 研究方向为嵌入式系统。

E-mail: 1759387159@qq.com

反卷积进行上采样重建。实验结果表明, 本文所提方法能够达到较好的重建效果并满足对视频超分辨率重建的实时性要求。

## 1 相关理论

### 1.1 FSRCNN-s

FSRCNN-s 是在 SRCNN 基础上的改进, 网络结构层次相比 SRCNN 层次加深, 并且使用反卷积进行上采样, 代替了传统的双三次插值法。FSRCNN-s 的网络结构如图 1

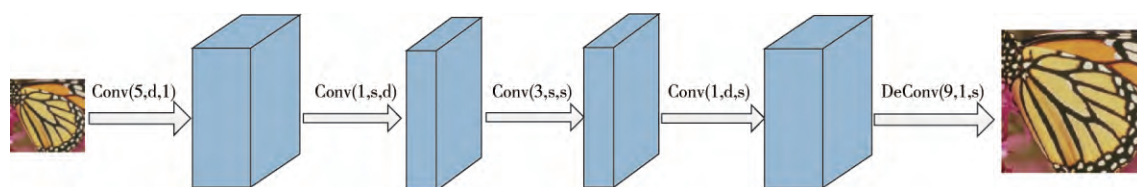


图 1 FSRCNN-s 的网络结构

FSRCNN-s 的网络运行速度可达到 SRCNN 的 17.36 倍, 但超分辨率重建效果与 SRCNN 相近, 不能同时实现良好重建性能和实时性。

### 1.2 ESPCN

SRCNN 首先利用双三次插值 (Bicubic) 方法, 把输入图像放大到与输出相同的尺寸, 再输入到网络中, 使得卷积操作都是在高分辨率图像上进行, 相对在低分辨率图像上做卷积计算而言, 增加了计算复杂度, 效率降低。ESPCN 是 Shi 等<sup>[9]</sup>提出的一种使用亚像素卷积神经网络的超分辨率算法, 该算法直接在低分辨率图像空间进行特征提取, 最后通过亚像素卷积层实现超分辨率重建。图 2 所示为 ESPCN 的网络结构。

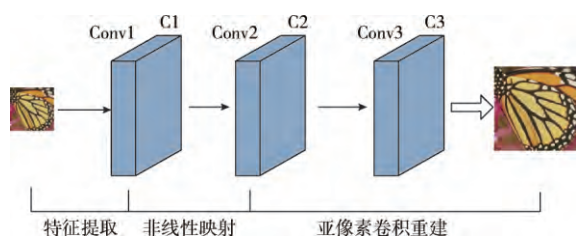


图 2 ESPCN 的网络结构

ESPCN 由两个卷积层和一个亚像素卷积层构成, 其中亚像素卷积层包括卷积和排列两个步骤。将未经任何处理操作的低分辨率图像作为输入, 通过 3 个卷积层的卷积操作, 得到和输入图像相同尺寸大小的特征图, 特征图的通道数为  $r^2$ , 然后将特征图  $r^2$  个通道中的所有像素点按照已经设定的顺序进行重新排列组合, 从而得到尺寸为  $r \times r$  的图像块。采用该排列方式可以将尺寸为  $r^2 \times H \times W$  的特征图重新排列成为一个  $1 \times rH \times rW$  的高分辨率图像。

所示。

FSRCNN-s 由 4 个卷积层和 1 个反卷积层组成, 可分为 5 个部分: 第 1 层为特征提取部分, 采用 32 个  $5 \times 5$  的卷积核提取输入图像的特征; 第 2 层采用 5 个  $1 \times 1$  的卷积核对第 1 层输出的特征图的数量进行收缩; 第 3 层为非线性映射部分, 利用 5 个  $3 \times 3$  的卷积核来完成; 第 4 层采用 32 个  $1 \times 1$  的卷积核扩张特征图; 第 5 层采用  $9 \times 9$  反卷积对低分辨率图像上采样到高分辨率图像, 并对高分辨率图像块进行重组, 实现最终的重建。

ESPCN 缩减了网络参数, 降低计算复杂度, 大幅度地提升了运算速度, 达到了实时视频处理的要求, 但是其超分辨率重建性能相比其它方法有所下降。

## 2 本文方法

在 FSRCNN-s 和 ESPCN 的基础上, 本文提出一种超分辨率重建方法, 利用反卷积与池化的组合代替卷积层完成非线性映射, 并使用反卷积取代亚像素卷积层, 同时添加了收缩层和扩张层, 不仅提升了超分辨率重建性能, 也具有较快的网络运算速度, 能够满足实时处理的要求。该算法的网络结构由特征提取、收缩、非线性映射、扩张、重建 5 个部分组成, 图 3 展示了该网络的网络结构, 其中 Conv1、Conv2、Conv3 表示卷积层, pool 表示池化层, deconv1、deconv2 表示反卷积层。

### 2.1 特征提取

本文将未进行任何预处理的原始低分辨率图像作为输入数据, 通过 Conv1 卷积层的卷积操作对输入的低分辨率图像进行提取特征。

提取特征的过程即卷积操作, 采用一组卷积核在图像上进行从左到右、从上到下的滑动, 然后得到一组特征图, 卷积操作公式如式 (1) 所示

$$F_1(Y) = \sigma(W_1 * Y + b_1) \quad (1)$$

式中:  $W_1$  ——  $n_1$  个大小为  $m \times k_1 \times k_1$  的卷积核 ( $m$  为输入图像的通道数,  $k_1$  为卷积核的尺寸),  $Y$  —— 输入的原始图像,  $*$  —— 卷积操作,  $b_1$  ——  $n_1$  维的偏差,  $\sigma(x)$  —— 激活函数,  $F_1(Y)$  —— 经过卷积层后得到的一组数量为  $n_1$  的特征图。

本文在对图像进行卷积时, 对边界进行零填充 (zero padding), 即  $padding = (k_1 - 1) / 2$ , 使得经过卷积后得到

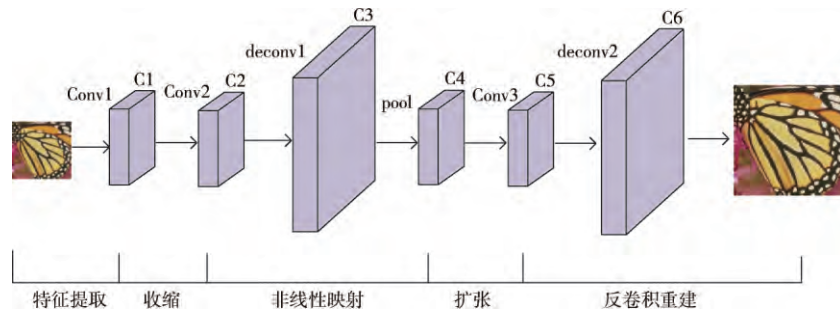


图3 本文方法的网络结构

的特征图与输入图像的尺寸保持不变。因此, Conv1 卷积层采用 56 个尺寸大小  $5 \times 5$  的卷积核, 即设置  $k_1 = 5$ ,  $padding = 2$ 。激活函数  $\sigma(x_i)$  采用带参数的修正线性单元 PReLU 函数, 加快网络收敛速度。

## 2.2 非线性映射

本文利用反卷积与池化的组合对特征提取阶段生成的一组特征图进行放大和缩小, 提取出对结果更为敏锐的特征, 从而更好地完成非线性映射阶段。

经过反卷积操作后, 得到的特征图尺寸成比例放大, 反卷积操作过程如式 (2) 所示

$$F_2(Y) = \sigma(W_2 \otimes F_1(Y) + b_2) \quad (2)$$

式中:  $W_2$  ——  $n_2$  个大小为  $n_1 \times k_2 \times k_2$  的反卷积核 ( $k_2$  为反卷积核的尺寸),  $F_1(Y)$  —— 特征提取生成的特征图,  $\otimes$  —— 反卷积操作,  $b_2$  ——  $n_2$  维的偏差,  $\sigma(x)$  —— 激活函数,  $F_2(Y)$  —— 经过反卷积层后得到的一组数量为  $n_2$  的特征图。

池化即为下采样, 与卷积操作相似, 但不经过反向传播的修改, 能够减少特征图的维数, 有助于降低网络训练的时间。该操作如式 (3) 所示

$$F_3(Y) = W_3 \otimes F_2(Y) \quad (3)$$

式中:  $W_3$  ——  $n_2 \times k_3 \times k_3$  的滤波器 ( $k_3$  为滤波器的尺寸),  $F_2(Y)$  —— 经过反卷积层后得到的  $n_2$  张特征图,  $\otimes$  —— 池化操作,  $F_3(Y)$  —— 经池化处理后得到的  $n_2$  张特征图。

设输入图像尺寸为  $H \times W$ , 放大因子为  $f$ , 反卷积核的尺寸为  $k_2$ , 步长为  $s$ , 零填充为  $p$ , 经反卷积后的特征图尺寸为  $[s(H-1)-2p+k_2] \times [s(W-1)-2p+k_2] = fH \times fW$ ; 池化层滤波器的尺寸为  $k_3 = f$ , 步长为  $stride = f$ , 池化方式为最大池化, 经过池化后恢复为  $H \times W$ 。本文反卷积与池化的组合设置如下: 当  $f = 2$  时,  $k_2 = 6$ ,  $s = 2$ ,  $p = 2$ ,  $k_3 = 2$ ;  $stride = 2$ ; 当  $f = 3$  时,  $k_2 = 7$ ,  $s = 3$ ,  $p = 2$ ,  $k_3 = 3$ ;  $stride = 3$ ; 当  $f = 4$  时,  $k_2 = 8$ ,  $s = 4$ ,  $p = 2$ ,  $k_3 = 4$ ;  $stride = 4$ 。

## 2.3 反卷积重建

通过反卷积层对低分辨率图像上采样到高分辨率图像, 并对高分辨率图像块进行重组, 实现最终的重建。本文设

计的网络模型中将 deconv2 层的反卷积核尺寸设置为  $9 \times 9$ ,  $padding = 4$ , 当放大因子  $f = 2$  时, 设置步长为 2, 当放大因子  $f = 3$  时, 设置步长为 3, 当放大因子  $f = 4$  时, 设置步长为 4。

## 2.4 收缩与扩张

收缩: 采用 12 个尺寸为  $1 \times 1$  的小卷积核对 Conv1 层生成的特征图进行降维处理, 以此减少网络的参数, 降低计算复杂度, 帮助提高网络的训练速度。

扩张: 采用 56 个尺寸为  $1 \times 1$  的小卷积核对经 pool 层处理得到的特征图进行扩维, 改善重建效果。扩张相当于收缩的逆过程。

## 2.5 损失函数

本文使用均方误差 (mean squared error, MSE) 作为损失函数, 如式 (4) 所示

$$L = \frac{1}{N} \sum_{m=1}^N \|Y - X_m\|^2 \quad (4)$$

其中,  $Y$  为网络预测图像,  $\{X_m\}$  为一组真实图像。

## 3 实验与分析

为了对所提出的算法进行实验分析, 首先对于软件环境配置: 在服务器上安装了 Ubuntu 16.04 LTS 64 位操作系统, 然后在系统中搭建了 Caffe<sup>[10]</sup> 开源框架, 同时配置了 cuDNN 实现加速, 在对算法测试时采用了 Matlab R2014a; 对于系统硬件配置: 服务器配置了 Intel Xeon E5 型号的处理器和 GTX 1080 Ti 型号的显卡, 以及 32 GB 的内存。

### 3.1 数据集与训练样本

本文实验采用的训练数据集为 91 images<sup>[4]</sup> 和 General-100<sup>[8]</sup>, 采用的测试数据集为 Set5<sup>[11]</sup>、Set14<sup>[4]</sup>、BSD200<sup>[12]</sup>。

为了提升训练图像数据的有效性, 本文参照文献 [8] 中的策略, 对数据集 91 images 和 General-100 进行数据扩大: ① 旋转: 将每张训练图像分别旋转  $90^\circ$ 、 $180^\circ$ 、 $270^\circ$ ; ② 将每张训练图像缩放至其原始尺寸的 0.6、0.7、0.8、0.9 倍, 最后得到 3820 ( $191 \times 4 \times 5$ ) 张图像作为训练集。实验使用数据集 Set5 的图像作为训练过程中的验证集。

### 3.2 结果及分析

本文实验在放大比例为 2、3、4 时均进行 1500 万次迭代训练, 学习速率设置为 0.001, 动量设置为 0.9, 将测试结果和 Bicubic<sup>[1]</sup>、SRCNN<sup>[7]</sup>、A+<sup>[4]</sup>、ESPCN<sup>[9]</sup>等方法进行对比。本实验使用峰值信噪比 PSNR 和结构相似性指数

SSIM 作为生成的高分辨率图像的质量评估标准。

表 1 展示了当训练集为 91 images<sup>[4]</sup>时, 各超分辨率重建方法分别在 2 倍、3 倍、4 倍图像放大比例下, 各测试集中所有图像重建后 PSNR 的平均值。经 Bicubic、SRCNN、FSRCNN-s、ESPCN 处理所得数据请参见文献 [13] 表 1。

表 1 91 images<sup>[4]</sup>时, 不同放大比例下各测试集经不同方法处理后 PSNR 的平均值

数据集	BSD200	Set5	Set14	BSD200	Set5	Set14	BSD200	Set5	Set14
Scale	2	2	2	3	3	3	4	4	4
PSNR	31.63	36.81	32.29	28.41	32.71	29.10	26.83	30.27	27.25

表 2 展示了当训练集为 91 images<sup>[4]</sup> + General-100<sup>[8]</sup>时, 本文方法分别在 2 倍、3 倍、4 倍图像放大比例下, 各测试集中所有图像重建后 PSNR 的平均值。经 Bicubic、A+、SRCNN、FSRCNN-s 处理所得数据请参见文献 [8]

表 4。

表 3 展示了当训练集为 91 images<sup>[4]</sup> + General-100<sup>[8]</sup>时, 各超分辨率重建方法分别在 2 倍、3 倍、4 倍图像放大比例下, 各测试集中所有图像重建后 SSIM 的平均值。

表 2 91 images<sup>[4]</sup> + General-100<sup>[8]</sup>时, 不同放大比例下各测试集经不同方法处理后 PSNR 的平均值

数据集	BSD200	Set5	Set14	BSD200	Set5	Set14	BSD200	Set5	Set14
Scale	2	2	2	3	3	3	4	4	4
PSNR	31.62	36.78	32.32	28.93	32.87	29.19	26.92	30.51	27.39

表 3 91 images<sup>[4]</sup> + General-100<sup>[8]</sup>时, 不同放大比例下各测试集经不同方法处理后 SSIM 的平均值

Dataset	Scale	Bicubic <sup>[1]</sup>	ANR <sup>[4]</sup>	A+ <sup>[4]</sup>	SRCNN <sup>[7]</sup>	FSRCNN-s <sup>[8]</sup>	本文算法
Set5	2	0.9299	0.9499	0.9544	0.9521	0.9532	<b>0.9546</b>
Set14	2	0.8687	0.9004	0.9056	0.9039	0.9052	<b>0.9063</b>
BSD200	2	0.8625	0.8968	0.9026	0.9024	0.9027	<b>0.9030</b>
Set5	3	0.8675	0.8968	0.9088	0.9033	0.9055	<b>0.9097</b>
Set14	3	0.7736	0.8093	0.8188	0.8145	0.8167	<b>0.8215</b>
BSD200	3	0.7638	0.7981	0.8078	0.8038	0.8058	<b>0.8093</b>
Set5	4	0.8104	0.8419	0.8603	0.8530	0.8499	<b>0.8623</b>
Set14	4	0.7019	0.7353	0.7471	0.7413	0.7423	<b>0.7519</b>
BSD200	4	0.6949	0.7253	0.7359	0.7291	0.7312	<b>0.7394</b>

根据表 1 中的数据, 采用本文方法重建的图像的 PSNR 的值相比快速超分辨率重建方法 FSRCNN-s、ESPCN 有所提升, 即本文方法实现了更好的超分辨率重建性能; 从表 2 和表 3 也可以看出经过本文方法重建后的图像质量均高于其它方法。结果表明, 本文算法在重建效果上超过了实时图像超分辨率算法 FSRCNN-s 以及 ESPCN。

为了测试本文算法的实时性, 实验收集了 100 张尺寸大小为 320×240 的图像并对其进行超分辨率重建, 在 2 倍、3 倍以及 4 倍放大比例下, 平均每张图片的处理时间分

别为 0.041 s、0.041 s、0.040 s, 即平均每秒分别处理 24.4、24.4、25 张图像, 能够满足视频每秒处理 24 帧图像的实时性要求。

图 4 展示了 Bicubic、ANR<sup>[4]</sup>、SRCNN、FSRCNN-s 等各种方法在 91 images<sup>[4]</sup>时 2 倍放大比例下, 测试集 Set14 中“comic”图像的重建效果对比; 图 5 展示了各方法在 91 images<sup>[4]</sup>时 3 倍放大比例下, 测试集 Set5 中“butterfly \_\_ GT”图像的重建效果对比; 图 6 展示了各方法在 91 images<sup>[4]</sup>时 4 倍放大比例下, 测试集 Set14 中“monarch”图像

的重建效果对比。

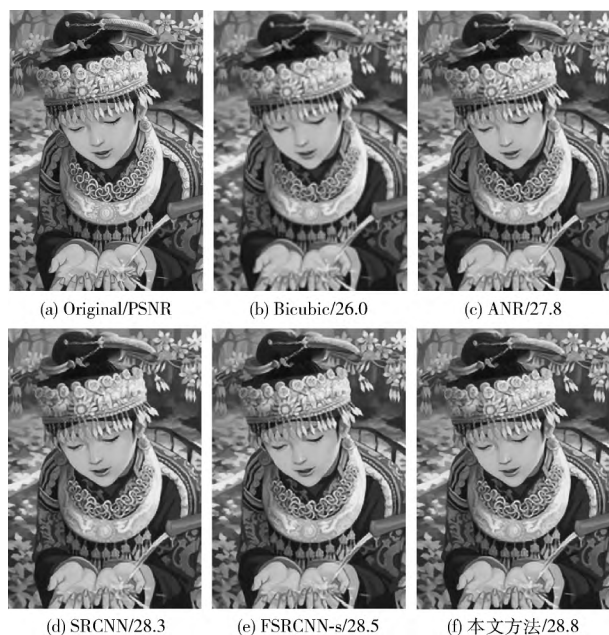


图 4 2 倍放大比例下“comic”的重建效果对比

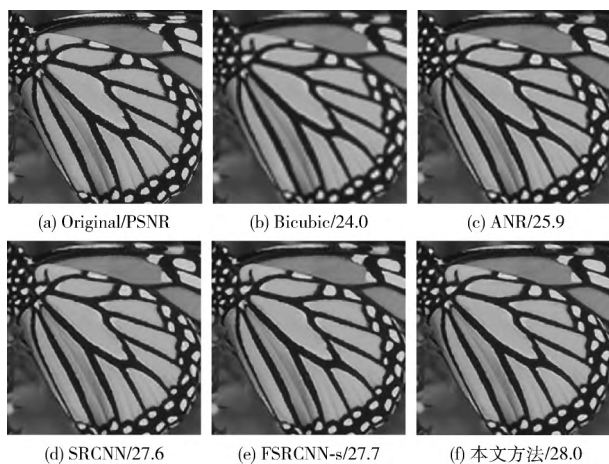


图 5 3 倍放大比例下“butterfly\_GT”的重建效果对比

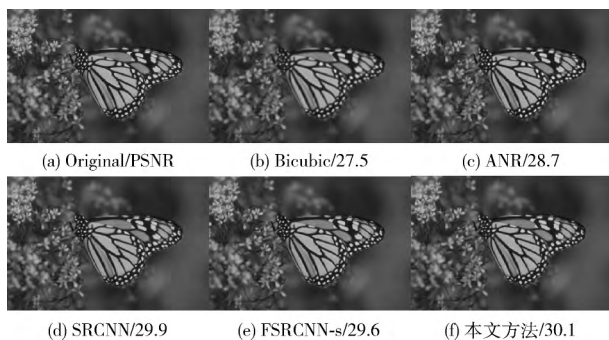


图 6 4 倍放大比例下“monarch”的重建效果对比

#### 4 结束语

针对现有的实时图像超分辨率算法 ESPCN 和 FSRCNN-s 难以同时满足较好的超分辨率重建效果和实时处理要求的问题, 本文提出了一种基于 CNN 的图像超分辨率重建方法, 直接以未做任何预处理的低分辨率图像为输入数据, 利用卷积层提取原始图像表征, 采用尺寸为  $1 \times 1$  的卷积核对卷积得到的特征图进行降维, 减少网络的参数, 降低计算复杂度, 利用反卷积与池化的组合对特征图进行放大和缩小, 提取出对结果更为敏锐的特征, 再采用  $1 \times 1$  的小卷积核对特征图进行括维, 改善重建效果, 最后通过反卷积进行上采样重建。实验结果表明, 相比 Bicubic、SRCNN、FSRCNN-s、ESPCN 等方法, 本文所提方法不仅实现了更好的图像重建效果, 同时能够平均每秒能处理 24 张以上尺寸为  $320 \times 240$  的图像, 满足对视频进行超分辨率重建的实时性要求, 但是该网络的训练时间较长, 图像重建效果也有待提升。因此, 未来将继续研究更优的网络, 进一步提高超分辨率重建性能, 加快网络收敛速度, 并且能够保持较快的处理速度。

#### 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 (in Chinese). [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] Yang X, Zhang Y, Zhou D, et al. An improved iterative back projection algorithm based on ringing artifacts suppression [J]. Neurocomputing, 2015, 162 (C): 171-179.
- [3] Timofte R, De V, Gool L V. Anchored neighborhood regression for fast example-based super-resolution [C] //IEEE International Conference on Computer Vision. IEEE Computer Society, 2013: 1920-1927.
- [4] Timofte R, Smet V D, Gool L V. A+: Adjusted anchored neighborhood regression for fast super-resolution [C] //Asian Conference on Computer Vision. Springer, 2014: 111-126.
- [5] Yang C Y, Yang M H. Fast direct super-resolution by simple functions [C] //IEEE International Conference on Computer Vision. IEEE, 2014: 561-568.
- [6] Zhou L Y, Su C X, Cao Y F. Image super-resolution via sparse representation [J]. Computer Engineering & Design, 2016, 37 (12): 3290-3294.
- [7] Dong C, Chen C L, He K, et al. Image super-resolution using deep convolutional networks [J]. IEEE Transactions on

- Pattern Analysis & Machine Intelligence, 2016, 38 (2): 295-307.
- [8] Dong C, Chen C L, Tang X. Accelerating the super-resolution convolutional neural network [C] //European Conference on Computer Vision. Springer, 2016: 391-407.
- [9] Shi W, Caballero J, Huszar 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. IEEE Computer Society, 2016: 1874-1883.
- [10] Jaderberg M, Simonyan K, Vedaldi A, et al. Reading text in the wild with convolutional neural networks [J]. International Journal of Computer Vision, 2014, 116 (1): 1-20.
- [11] Bevilacqua M, Roumy A, Guillemot C, et al. Single-image super-resolution via linear mapping of interpolated self-examples [J]. IEEE Transactions on Image Processing, 2014, 23 (12): 5334-5347.
- [12] Martin D R, Fowlkes C C, Malik J. Learning to detect natural image boundaries using local brightness, color, and texture cues [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2004, 26 (5): 530-549.
- [13] Schuler S, Leistner C, Bischof H. Fast and accurate image upscaling with super-resolution forests [C] //Computer Vision and Pattern Recognition. IEEE, 2015: 3791-3799.

(上接第 1610 页)

- [10] YANG Ronghua, YOU Yangsheng, WU Hao. A method for calculating curvature of scattered point clouds on mixed surfaces [J]. Science of Surveying and Mapping, 2014, 39 (6): 129-131 (in Chinese). [杨荣华, 游扬声, 吴浩. 一种混合曲面散乱点云曲率计算方法 [J]. 测绘科学, 2014, 39 (6): 129-131.]
- [11] YANG Ronghua, HUA Xianghong, YOU Yangsheng. Local n-TH order surface fitting and curvature estimation based on scattered-point cloud data [J]. Journal of Geodesy and Geodynamics, 2013, 33 (3): 141-143 (in Chinese). [杨荣华, 花向红, 游扬声. 基于散乱点的局部 n 次曲面拟合及其曲率计算 [J]. 大地测量与地球动力学, 2013, 33 (3): 141-143.]
- [12] Marani R, Renò V, Nitti M, et al. A modified iterative closest point algorithm for 3D point cloud registration [J]. Computer-Aided Civil and Infrastructure Engineering, 2016, 31 (7): 515-534.