

中图分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2017)12-1690-11

论文引用格式: Li Y F, Fu R D, Jin W, Ji N. Image super-resolution using multi-channel convolution[J]. Journal of Image and Graphics, 2017, 22(12): 1690-1700. [李云飞, 符冉迪, 金炜, 纪念. 多通道卷积的图像超分辨率方法[J]. 中国图象图形学报, 2017, 22(12): 1690-1700.] [DOI: 10.11834/jig.170325]

# 多通道卷积的图像超分辨率方法

李云飞 符冉迪 金炜 纪念

宁波大学信息科学与工程学院, 宁波 315211

**摘要:** 目的 超分辨率技术在实际生活中具有较为广泛的应用。经典的基于卷积神经网络的超分辨率(SRCNN)方法存在重建图像纹理结构模糊以及网络模型训练收敛过慢等问题。针对这两个问题,在SRCNN的基础上,提出一种多通道卷积的图像超分辨率(MCSR)算法。方法 通过增加残差链接,选择MSRA初始化方法对网络权值进行初始化,加快模型收敛;引入多通道映射提取更加丰富的特征,使用多层 $3 \times 3$ 等小卷积核代替单层 $9 \times 9$ 等大卷积核,更加有效地利用特征,增强模型的超分辨率重构效果。结果 MCSR迭代 $4 \times 10^6$ 次即可收敛,在Set5与Set14数据集上边长放大3倍后的平均峰值信噪比分别是32.84 dB和29.28 dB,与SRCNN相比提升显著。结论 MCSR收敛速度更快,并且可以生成轮廓清晰的高分辨率图像,超分辨率效果更加优秀。

**关键词:** 图像超分辨率;深度学习;卷积神经网络;多通道卷积;残差学习

## Image super-resolution using multi-channel convolution

Li Yunfei, Fu Randi, Jin Wei, Ji Nian

Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo 315211, China

**Abstract:** **Objective** Super-resolution (SR) technology is the method for satisfying the demand for high-quality images. The method was first proposed in the 1960s, and its goal is to obtain one or a series of high-resolution (HR) image (s) using one or a sequence of low-resolution (LR) image (s). SR technology not only can improve the visual performance of images but also can help improve the analysis and processing of images, including object recognition, image retrieval, and object detection. SR technology is widely used in real life, such as in video surveillance system, medical image processing, and remote sensing image processing. Traditional methods, such as interpolation-, reconstruction-, and learning-based algorithms, cannot achieve desirable SR results and short SR times. In recent years, a modern convolutional neural network (CNN)-based method called super-resolution CNN (SRCNN) has been proposed. The SRCNN method is a deep learning method for single-image SR and directly learns an end-to-end mapping between LR and HR images. This method achieves better performance in SR results and SR times than do the traditional ones but still presents several limitations. SRCNN uses stacked CNN structure and Gauss initialization method, thereby resulting in slow convergence and time-consuming model

收稿日期: 2017-06-29; 修回日期: 2017-09-14; 预印本日期: 2017-09-19

基金项目: 国家自然科学基金资助项目(61471212); 浙江省自然科学基金资助项目(LY16F010001); 宁波市自然科学基金资助项目(2016A610091)

第一作者简介: 李云飞(1994—),男,2016年于宁波大学攻读电子与通信工程硕士学位,主要研究方向为图像超分辨率、图像检索。  
E-mail: lyf940622@163.com

通信作者: 符冉迪,副教授, E-mail: furandi\_nbu@163.com

**Supported by:** National Natural Science Foundation of China (61471212); Natural Science Foundation of Zhejiang Province of China (LY16F010001); Ningbo Municipal Natural Science Foundation(2016A610091)

training. Furthermore, SRCNN exhibits poor nonlinear mapping capability and simple feature extraction because it comprises only three layers of convolution kernels. The method generates unclear HR images of blurry texture. An image SR method based on multi-channel CNN (MCSR) is proposed to resolve the aforementioned issues. **Method** MCSR adopts two different strategies, namely, residual CNN model and MSRA initialization method, to accelerate the convergence of model training. Given that residual CNN possesses an identity mapping from input to output, the model training aims to explicitly model the residual image, which is the difference between HR and LR images. This change is advantageous considering that LR and HR images share the same information to a large extent. The MSRA initialization method can maintain activation and back-propagated gradient variances when moving up or down the network. Both schemes result in substantially fast convergence. At the same time, the two schemes are suggested to improve the performance of image SR. The deeper the CNN structure, the better the performance of CNN. MCSR replaces the large convolution kernel, such as  $9 \times 9$ , as chosen by SRCNN with several layers of small convolution kernel, such as  $3 \times 3$ . As a result, MCSR obtains seven layers of convolution kernel and exhibits enhanced capability of nonlinear mapping. In addition to deepening, MCSR is widened to multi-channel on the nonlinear mapping part. Precisely, the basic MCSR possesses four channels of one layer of  $3 \times 3$  convolution kernels, two layers of stacked  $3 \times 3$  convolution kernels, one layer of  $1 \times 5$  convolution kernels, and one layer of  $5 \times 1$  convolution kernels. Experimental results show that different channels produce dissimilar feature maps. In particular, the  $3 \times 3$  channel produces local feature maps, the  $2 \times 3 \times 3$  channel produces relative global feature maps, the  $1 \times 5$  channel extracts transversal textural features, and the  $5 \times 1$  channel extracts vertical textural feature. Furthermore, MCSR possesses an extra one layer of  $1 \times 1$  convolution kernel for compressing the dimension of the feature map, thereby providing the method with powerful nonlinear capability. Powerful nonlinear mapping capability and diverse feature maps can result in good SR performance. **Result** The proposed MCSR is trained on Image91 dataset, the same as SRCNN, and tested on Set5, Set14, and BSD200 datasets. Experimental results demonstrate that MCSR converges within  $4 \times 10^6$  backprops whereas SRCNN converges at least  $1.5 \times 10^7$  backprops. The average peak signal-to-noise ratios (PSNRs) with an upscaling factor 3 on Set5, Set14, and BSD200 are 32.84 dB, 29.28 dB, and 29.03 dB and increase by 0.45 dB, 0.27 dB, and 0.38 dB, respectively, compared with those for SRCNN. Structural similarity image measurement also achieves considerable improvement. With regard to subjective effect, MCSR can produce high-quality HR images of clear texture. The produced images barely show shadow and ripple effects. These findings indicate that MCSR achieves good SR performance. Notably, we propose an extra method called MCSR-Ex, which extends the MCSR method to five channels. The additional channel consists of three layers of  $3 \times 3$  convolution kernels and improves the PSNR by approximately 0.1 dB on Set5 dataset on average. **Conclusion** In this study, a new SR method called MCSR is proposed. On the one hand, the combination of residual model and MSRA initialization method can significantly accelerate the convergence of model training. On the other hand, the suggested two schemes, which include widening the CNN model to multi-channel and deepening the CNN model to seven layers, can considerably improve the performance of image SR. In other words, the good SR performance is attributed to extracting various feature maps and using feature maps.

**Key words:** image super-resolution; deep learning; convolution neural network; multi-channel convolution; residual learning

## 0 引言

超分辨率(SR)技术是计算机视觉领域的一门重要学科,目的是利用已有的低分辨率(LR)图像生成对应的高分辨率(HR)图像。HR图像具有较高的像素密度,可以提供更加丰富的细节信息,有利于图像、视频的进一步处理。图像超分辨率技术在医学图像处理、遥感卫星图像处理、视频监控等领域有比

较广泛的应用。

自1964年Harris提出<sup>[1]</sup>以来,SR领域已经形成基于插值的方法<sup>[2-3]</sup>、基于重建的方法<sup>[4-5]</sup>、基于学习的方法<sup>[6-7]</sup>3类主流方法,其中基于学习的SR方法是近年来图像SR领域的热点方向,它的基本思路是通过大量样本训练得到LR图像与HR图像之间的映射关系,并藉此作为先验知识进行重建。Yang等人<sup>[6]</sup>提出基于稀疏编码的图像超分辨(Sc-SR)算法,对高、低分辨率图像块进行联合训练得到

相应的一个完备字典,通过字典进行高低分辨率图像块的稀疏关联,取得非常不错的效果。Zeyde 等人<sup>[7]</sup>在 Yang 的基础上应用 K-SVD 算法<sup>[8]</sup>对字典的训练过程进行改进,提高了训练字典速度,使得重建的图像在主观与客观评价指标上均有所改善,但是重建过程的计算复杂度高,HR 图像生成时间过长。为了实现图像的实时处理,Timofte 等人<sup>[9]</sup>提出锚点邻域回归(ANR)算法,把稀疏编码的思想用于最近邻域嵌入算法 NE<sup>[10]</sup>之上,大大降低重建过程的计算复杂度,但是重建图像的质量还有待提高。

Krizhevsk 等人<sup>[11]</sup>凭借卷积神经网络(CNN)这一利器以绝对优势赢得了 2012 年的 ImageNet 图像分类大赛,同时也引起了新一轮的研究 CNN 的热潮。之后几年,不断涌现出像 ZFNet<sup>[12]</sup>、GoogLeNet<sup>[13]</sup>、ResNet 等人<sup>[14]</sup>优秀的 CNN 模型。实践证明,用于分类任务的 CNN 模型迁移到其他任务也可以取得非常优秀的效果<sup>[15-17]</sup>。因此,Dong 等人<sup>[18]</sup>在基于样本学习的 SR 算法基础上提出了超分辨率卷积神经网络(SRCNN)。SRCNN 首次将 CNN 引入 SR 任务中,将传统 SR 方法的分步处理整合到一个深度学习模型中,大大简化了 SR 工作流程。SRCNN 的提出证明将深度学习,特别是 CNN 应用到 SR 任务中是非常合适的,发展前景广阔。

同传统的 SR 算法相比,SRCNN 在 SR 性能上有不错的提升。但 SRCNN 在以下两个方面依旧具有局限性:1) SRCNN 的模型在训练过程中收敛速度过于缓慢;2) SRCNN 在重建 HR 图像时,利用到的区域特征单一并且利用效率不高,最终导致重建的 HR 图像纹理模糊,SR 性能有待提高。

针对上述两个不足,作出以下改进:1) 采取 MSRA 权值初始化方法与残差学习 CNN 模型的组合,加快模型训练的收敛速度;2) 提出使用多通道卷积来提取丰富的区域特征,同时加深模型增强模型的非线性映射能力,更有效地利用提取到的特征进行 HR 图像重建。

最终,提出了一种多通道超分辨率卷积神经网络(MCSR)模型。实验证明,同 SRCNN 相比,MCSR 收敛速度更快:SRCNN 至少需要迭代  $1.5 \times 10^7$  次才可以收敛,而 MCSR 只需要  $4 \times 10^6$  次;超分辨效果更加优秀:在 Set5 与 Set14 数据集放大 3 倍边长后每幅图片的平均峰值信噪比(PSNR)为 32.84 dB 和

29.28 dB,同 SRCNN 的 32.39 dB 和 29.01 dB 分别提升 0.45 dB 和 0.27 dB。当放大倍数扩展到 4 倍以及 8 倍时,MCSR 的表现依旧优于 SRCNN 等方法。

## 1 相关工作

### 1.1 SRCNN

SRCNN 的超分辨率流程如下:对于一幅原始图像,首先采用双三次线性插值到目标尺寸,然后送入 SRCNN 模型进行多层卷积映射,最终得到想要的 HR 图像。其中,将原始图像插值到目标尺寸是唯一的预处理工作。为方便表示,下文提到的 LR 图像均表示将原始图像进行双三次(Bicubic)插值后得到的图像。

SRCNN 具有由 3 层卷积核组成,尺寸分别为:  $9 \times 9$ 、 $1 \times 1$ 、 $5 \times 5$ ,相应的每层卷积核数分别是:64、32、1。SRCNN 对 3 层卷积核的解释分别是:

输入特征提取、非线性映射以及高分辨率图重构,结构如图 1 所示。第 1 层用于将输入 LR 图像表示为 64 幅不同的特征图(feature map);第 2 层将其映射为 32 幅特征图,不同于上一层,这一层的特征图可以通过组合进而表示成 HR 图像;第 3 层就是对第 2 层的特征图进行组合表示,得到 HR 图像。

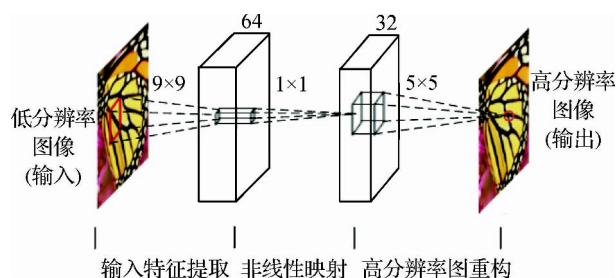


图 1 SRCNN 结构图

Fig. 1 Structure of SRCNN

在训练过程中,首先建立训练集时,将来自 image91 的 91 幅 HR 图像通过双三次插值进行  $1/N$  倍下采样,然后再进行  $N$  倍的上采样得到对应的 91 幅 LR 图像。高、低分辨率图各切成  $24 \times 24$  幅  $13 \times 13$  的图像块,分别作为训练集的标签以及数据。训练时,输入一幅 LR 图像块,经过 3 层的卷积和非线性映射计算,重构一幅目标图像,并且与 HR 标签图像块进行像素级别的均方误差(MSE)计算,然后将误差进行反向传递更新权值。权值更新策略采用随

机梯度下降法(SGD),batchsize 设定为 128,即计算 128 幅图像的误差后权值更新一次,作为迭代一次。经过大量训练,最终学习到 LR 图像到 HR 图像端的映射。测试时输入一幅 LR 图像即可输出一幅对应的 HR 图像,而不需要其他操作。

同时 Dong 等人<sup>[19]</sup>将 SRCNN 的卷积核尺寸扩展到  $9 \times 9, 5 \times 5, 5 \times 5$ ,并将其命名为 SRCNN-Ex。第 2 层的卷积核尺寸由  $1 \times 1$  扩展到  $5 \times 5$ ,这导致:参数数目由 8 032 增加到 57 184;感受野由  $13 \times 13$  增加到  $17 \times 17$ ,最终使得超分辨率效果提升了 0.19 dB( Set5 为例)。我们认为这样扩展网络换取性能的方式是低效的,即依旧没有改变特征提取单一以及特征利用不足的状况。

值得一提的是,Dong 等人<sup>[18]</sup>中通过增加一层  $1 \times 1$  的卷积核将 SRCNN 扩展到 4 层,训练之后发现这种做法反而会损失 SR 性能。于是,他们认为深层的网络并不会提升模型的 SR 性能。然而,本文提出的 MSCR 通过将模型扩展到 7 层大大的提升了模型的 SR 性能。

## 1.2 PreLU 与 MSRA 初始化

SRCNN 选择使用修正线性单元(ReLU)函数作为激活函数,即

$$f(x_i) = \begin{cases} x_i & x_i > 0 \\ 0 & x_i < 0 \end{cases} \quad (1)$$

和 Sigmoid 函数相比,ReLU 可以加快模型训练速度,缩短模型收敛时间,同时一定程度上抑制梯度消失现象。但是 ReLU 强制将负数值置零的特性常导致许多特征的丢失。

因此,He 等人<sup>[20]</sup>提出一种新的激活函数:含参修正线性单元(PreLU),表达式为

$$f(x_i) = \begin{cases} x_i & x_i > 0 \\ a_i x_i & x_i < 0 \end{cases} \quad (2)$$

对比式(1)(2)可以发现,PreLU 引入一个新的参数: $a_i$ , $a_i$  的值通常被设定为非常小,如 0.01 等。这样,既可以保留 ReLU 可以修正数据分布以及加速收敛的优点,同时又可以不完全丢失特征。因此,采用 PreLU 函数作为 MSCR 的激活函数。

另一方面,模型训练过程中对网络权值进行合理的初始化是非常重要的。初始化方法的不同选择可以直接决定一个模型是否可以收敛、收敛速度快慢以及最终的收敛状态如何。Glorot 等人<sup>[21]</sup>发现当

输入权值与输出权值在前向传播与反向传播过程中保持相同的分布时,模型的性能最优秀:模型可以快速收敛并且收敛的结果更优秀。为了达到这个目的,提出一种简便的初始化方法:Xavier 初始化。它

规定第  $i$  层的权值  $w_i$  服从均值为 0,方差为  $\frac{2}{n_i + n_{i+1}}$  的分布,并且建议采用均匀分布形式  $w_i \sim U(-\frac{\sqrt{6}}{\sqrt{n_i + n_{i+1}}}, \frac{\sqrt{6}}{\sqrt{n_i + n_{i+1}}})$ ,使得输出保持和输

入相同的均值为 0,方差为  $\sigma_x^2$  的分布,其中  $n_i$  表示当前层的参数数目, $n_{i+1}$  表示下一层的参数数目。

Xavier 初始化大大提高了模型性能,但是它的应用前提是激活函数在零点附近是线性的,因此只适用于 sigmoid 函数,不适用于 ReLU 以及 PReLU 函数。

之后,He 等人<sup>[20]</sup>在 Xavier 的基础上提出一种适用于 ReLU 以及 PReLU 函数的权值初始化的方法:

MSRA(microsoft research)初始化。它规定第  $i$  层的权值  $w_i$  服从均值为 0,方差为  $\frac{2}{n_i}$  的分布,通常采用

高斯分布形式, $w_i \sim N(0, \sqrt{\frac{2}{n_i}})$ ,同样输入和输出保持相同的分布,加快收敛速度同时提高模型性能。SRCNN 中采用均值为 0,方差为 0.01 的高斯分布对模型的权值进行初始化,MCSR 采用 MSRA 初始化方法对模型的权值进行初始化。

## 2 本文方法

本文提出的 MSCR 主要包括以下几个步骤:

- 1) 构建模型训练所需数据集;
- 2) 构建 CNN 模型,并确定相关的超参数;
- 3) 训练模型,优化网络参数,得到图像超分辨率的先验知识;

- 4) 输入低分辨率图像,测试模型的 SR 效果。

其中步骤 1) 首先对原始数据集进行增强(详见 3.1 节),然后采取 SRCNN 的方法建立数据集。下面对步骤 2) 3) 进行详细的阐述。

### 2.1 网络结构

本文提出的 MSCR 和 SRCNN 相比:增加残差链接、选择 MSRA 对网络的权值进行初始化加快模型收敛;使用多层的小卷积核( $3 \times 3$  等)代替单层的大卷积核( $9 \times 9$  等)加深网络模型,同时将非线性映射



部分改为局部多通路映射,增强模型 SR 性能。本文提出的 CNN 模型由 5 个部分组成,分别是:残差连接、输入特征提取部分、维度压缩部分、非线性映射部分以及高分辨率图像重建部分,网络结构如图 2 所示。

在 SRCNN 模型中,模型直接通过训练学习由 LR 图像  $x$  到 HR 图像  $y$  的映射关系。在实际情况中,高、低分辨率图像拥有大量的相似信息:低频信息,而 SR 任务的主要内容就是在保存原有的低频信息的同时增强 LR 图像的高频信息。而这种思想在 SRCNN 的映射关系中并没有得到体现。为了更

好的利用图像中的低频信息,我们引入一种更加有效的映射学习方法:残差学习<sup>[14]</sup>。MCSR 增加了一条残差通道,将输入的低分辨率图像直接送到输出端。训练 SRCNN 模型的目的是直接学习一幅 LR 图像到 HR 图像的映射关系,函数关系式为  $y = H(x)$ 。加入残差通道后映射关系式变为  $y = F(x) + x = H(x)$ ,其中  $F(x)$  为 LR 与 HR 图像之间的残差映射。而残差映射常表现为图像中的高频信息,故模型训练的目的转变为学习 LR 图像到图像纹理细节等高频信息的映射。实验证明,残差学习的引入可以大大加速模型的收敛速度。

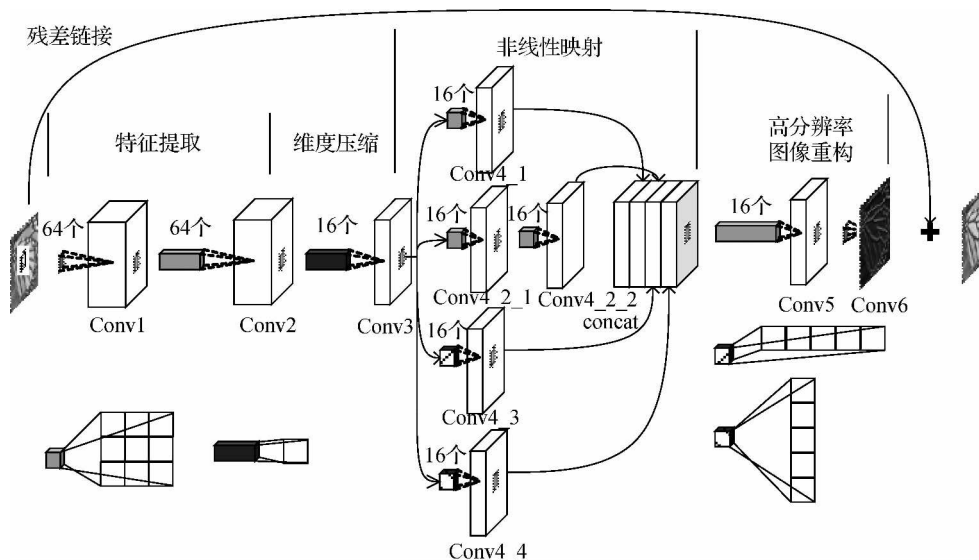


图 2 MCSR 结构图

Fig. 2 Structure of MCSR

Hinton 等人<sup>[22]</sup>将神经网络推向了深度学习的阶段,并证明深层的网络结构具有更加强大的非线性映射能力。深层网络结合各种模型压缩的方法<sup>[11-13]</sup>,可以在保证相同感受野情况下减少参数的数目。例如:一层尺寸为  $5 \times 5$  的卷积核可以被两层尺寸为  $3 \times 3$  的卷积核代替(假设步长均为 1,输入输出均为  $C$  通道)。这样做带来两个好处:1) 非线性映射由一层变成了两层,增加了局部的非线性映射能力;2) 参数数目由  $5 \times 5 \times C^2$  减少到  $2 \times (3 \times 3 \times C^2)$ ,降低了参数数目。因此,MCSR 用多个  $3 \times 3$  卷积核代替 SRCNN 中的  $9 \times 9$  卷积核与  $5 \times 5$  卷积核,将网络结构扩展到 7 层,包括 6 层的  $3 \times 3$  卷积核以及一层的  $1 \times 1$  卷积核。具体来说,我们采用两层的  $3 \times 3$  卷积核作为模型的特征提取部分,同样采用两层的  $3 \times 3$  卷积核作为模型的高分辨率重构

部分。

Szegedy 等人<sup>[13]</sup>在图像分类任务中使用  $1 \times 1$  卷积核来增加模型的非线性映射能力,并取得了非常优秀的效果。本模型在特征提取之后加入  $1 \times 1$  卷积核,除了增强一定的非线性映射能力之外,更重要的是对特征图进行维度压缩,从而减少参数。本文将厚度为 64 的 conv2 通过  $1 \times 1$  卷积压缩到 16 层,然后进行多通道的卷积,压缩的比率达到了 25%。

受 GoogleNet 的启发,MCSR 模型对非线性映射部分进行局部多通道扩展。每个通道具有不同层次、不同结构的卷积核。具体的,MCSR 包含四条不同的通道,从上到下依次为:conv4\_1,由一层尺寸为  $3 \times 3$  的卷积核组成;conv4\_2,由两层尺寸均为  $3 \times 3$  的卷积核组成;conv4\_3,由一层尺寸为  $5 \times 1$  的卷积核组成;conv4\_4,由一层尺寸为  $1 \times 5$  的卷积核组

成。之后对四个通道的卷积结果进行粘和 (concat) 作为整个非线性映射部分的输出, 同时作为高分辨率图像重构部分的输入。

4 条不同的通道, 可以提取不同的特征。conv4\_1 和 conv4\_2 通道相比: conv4\_1 具有一层, 每个像素点是关于 conv3 中  $3 \times 3$  个像素点的映射, 偏向提取局部特征; conv4\_2 具有两层, 每个像素点是关于

conv3 中  $5 \times 5$  个像素点的映射, 提取的特征相对更全面。conv4\_1、conv4\_3 和 conv4\_4 通道相比: 3 条通道都是一层, 但是 conv4\_1 是  $3 \times 3$ , conv4\_3 是  $5 \times 1$ , conv4\_4 是  $1 \times 5$ , 不同的卷积核尺寸提取了不同的特征图。conv4\_1 提取局部细节, conv4\_3 主要提取了纵向的纹理细节, conv4\_4 主要提取了横向的纹理细节, 如图 3 所示。

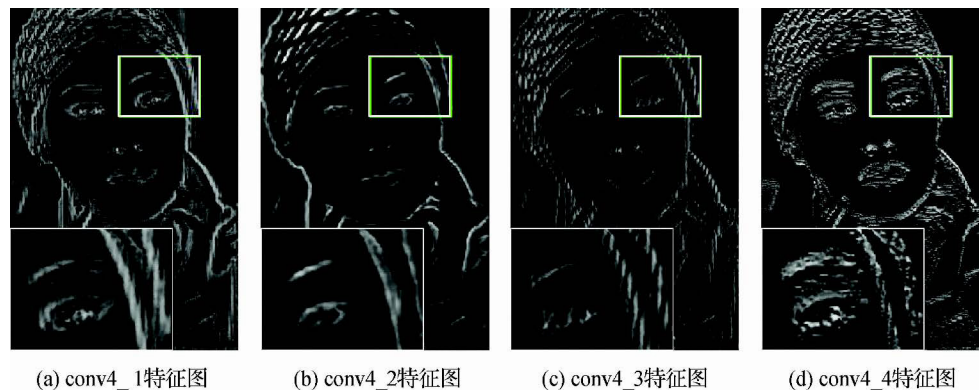


图3 非线性映射部分不同通道输出

Fig.3 Outputs of different channels in Non - Liner Mapping; ((a) feature map from conv4\_1; (b) feature map from conv4\_2; (c) feature map from conv4\_3; (d) feature map from conv4\_4)

选择 Set5 数据集集中的 Woman 图作为输入, 在每条通道的 16 幅输出特征图中各选择一幅激活度较高的边缘特征图, 进行对比度增强后作为图 3 的子图。观察发现, 同样是对 woman 图的边缘特征进行提取: 1) 与双层的、具有更大感受野的 conv4\_2 相比单层的 conv4\_1、conv4\_3、conv4\_4 提取的特征会比较的粗糙。2) 图 3(c) 中脸颊以及手指等位置的特征比较清晰, 纵向特征更丰富, 即 conv4\_3 提取了大量纵向边缘特征; 图 3(d) 中嘴唇以及下巴等位置的轮廓比较清晰, 横向特征更加丰富, 即 conv4\_4 提取了大量的横向边缘特征。

本文最终提出的 MCSR 模型拥有 57 296 个参数, 相较于 SRCNN 的 8 032 个有较大提升, 同 SRCNN - Ex 的 57 184 个相差不大, 可以认为处于同一参数数量级。同时本文对提出 CNN 模型进行扩展: 非线性映射部分增加一条含有 3 层  $3 \times 3$  卷积核的通道, 命名为 MCSR - Ex, 参数数目增加了  $(3 \times 3 \times 16 \times 16) \times 3 = 6 912$ , 总数目达到了 64 208 个。多种模型进行 SR 效果对比, 以此证明是本文提出的多通道卷积等策略增强了 SR 性能。

## 2.2 其他设置

选择 PReLU 函数作为本模型的激活函数, 相应

地使用 MSRA 进行模型初始化。本模型采取与 SRCNN 相同的学习率, 即基础学习率为  $10^{-4}$ , 最后的高分辨率重构部分学习率设定为  $10^{-5}$ , 较小的学习率可以保证模型更稳定地拟合映射关系。为了缩短迭代一次的时间, batchsize 设定为 64。遗忘率设定为 0.9, 没有采用权值衰减。

## 3 实验及分析

针对两个目的: 1) 加快模型收敛速度; 2) 增强模型 SR 性能, 所做的改进是合理的。设计两组实验: 第 1 组, 对比观察不同策略对模型收敛的影响; 第 2 组, 对比观察 MCSR 对 SR 结果的提升。

### 3.1 实验说明

实验在 ubuntu14.04 系统下进行, 硬件配置为 Intel i7-4720HQ @ 2.6 GHz、GTX-970M、6 GB 显存、16 GB 内存。基于深度学习框架 Caffe 训练并生成 CNN 模型, HR 图像生成工作在 Matlab 2014a 平台展开。

一般情况下, 训练集越大训练得到的模型会更加优秀。例如: SRCNN - Ex 在 Image91 库上进行训练, Set5 的平均 PSNR 可以达到 32.58 dB, 而在 Ima-

geNet 库上进行训练时, Set5 的平均 PSNR 可以达到 35.75 dB。本文选择 Image91 数据集作为唯一的原始训练库, 不同于 SRCNN 的是: 本文对训练集进行了扩展。首先是旋转, Image91 数据集的每幅图都依次旋转  $90^\circ$ 、 $180^\circ$  以及  $270^\circ$ , 于是图像总数扩展到之前的 4 倍。其次是裁剪, 上一步产生了每幅图都被缩放为原尺寸的 0.6、0.7、0.8、0.9 倍。经过两步处理, 图像总数变成未处理之前的  $5 \times 4 = 20$  倍。之后采取和 SRCNN 相同的方法生成许多对的 HR 图像块以及 LR 图像块, 分别作为标签和输入数据训练模型。与此同时, 选择 Set5、Set14 以及 BSD200 共 3 个数据集进行 SR 性能测试, 并对部分图像的重构效果进行展示。

选择包括: Bicubic、ScSR 算法、ANR 算法、SRCNN、SRCNN-Ex 以及 ESRCN<sup>[23]</sup> 在内的多种方法作为对比。其中 SRCNN、SRCNN-Ex 以及 ESRCN 使用和本文相同的 image91 图像库进行训练。实验选择 PSNR 以及结构相似度 (SSIM) 作为客观评价指标。PSNR 比较了 HR 图像与原始图之间像素级的误差, SSIM 从亮度、对比度、结构 3 个方面衡量图像的相似度, 通常认为两个参数越高, SR 效果越好。本章节最后展示 3 幅图像基于不同方法的高分辨率结果, 作为主观评价。

### 3.2 实验结果及分析

针对上文提出的两种提升模型收敛速度的策略: 增加残差链接以及使用 MSRA 初始化, 设计两个实验, 对比观察两种策略对模型收敛状况的影响, 实验结果如图 4 所示。蓝色曲线代表本文提出的 MCSR 模型, 作为基准; 红色曲线表示 MCSR 移除残差链接后的模型收敛情况; 黑色曲线表示权值初始化为均值为 0, 方差为 0.01 的高斯分布 (SRCNN 的策略) 后 MCSR 模型收敛情况。测试集为 Set5 数据集, 放大倍数为 3 倍。

对比图 4 的蓝色曲线与黑色曲线发现: 蓝色曲线一直处于黑色曲线上方, 蓝色曲线迭代  $2.5 \times 10^6$  次开始收敛而黑色曲线直到迭代  $3 \times 10^6$  次也没有收敛的迹象, 这证明增加残差链接可以加快模型收敛速度。对比蓝色曲线与红色曲线发现: 蓝色曲线一直处于红色曲线上方, 红色曲线一直没有收敛并且有 PSNR 降低的倾向, 说明 MSRA 初始化可以加快模型收敛速度, 同时也可以提高模型的性能。综上所述, 我们可以认为 MCSR 采用的 MSRA 初始化与残

差学习的组合可以有效地加快模型收敛速度。

图 5 展示了不同模型在 Set5 数据集上的收敛曲线, 由于不同 CNN 模型收敛时刻相差过大, 仅展示到 SRCNN-Ex 收敛时刻, SRCNN 并未达到收敛状态。观察发现 MCSR 收敛速度远超 SRCNN 以及 SRCNN-Ex 模型。MCSR 在  $4 \times 10^6$  次就达到收敛状态, SRCNN-Ex 达到收敛状态需要迭代  $8 \times 10^6$  次, 而 SRCNN 则需要更久 (图 5 中未展示)。可知, 同时采用残差链接以及 MSRA 初始化参数两个策略大大提升了模型的收敛速度。

观察图 5 可以发现本文模型结果优于 SRCNN-Ex 以及 SRCNN, 平均每幅图 PSNR 值超 SRCNN-Ex 大约 0.3 dB, 超过 SRCNN 大约 0.5 dB。同时发现, 在 MCSR 的基础上提出的 MCSR-Ex 模型具有更好的性能, 同 MCSR 相比提升了约 0.1 dB。有理由相信随着卷积通道的进一步丰富, 模型的 SR 性能会进一步提升。

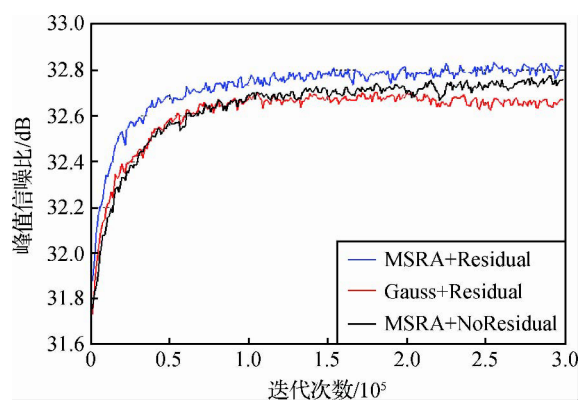


图 4 不同方法在 Set5 数据集的收敛曲线

Fig. 4 Convergence curves for different strategies on the Set5 dataset

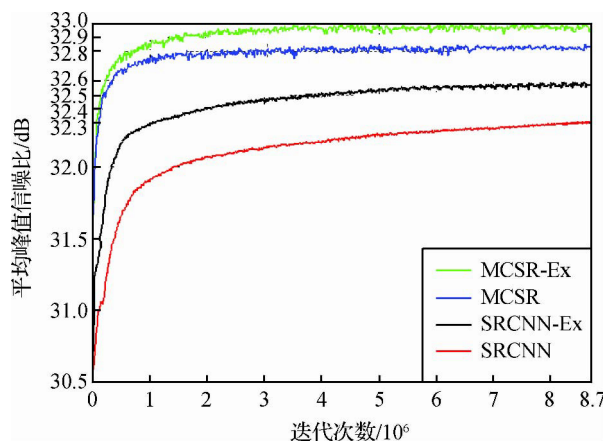


图 5 不同方法在 Set5 数据集的收敛曲线

Fig. 5 Convergence curves for different strategies on the Set5 dataset

表1列出了不同SR算法在Set5数据集的PSNR以及SSIM数值。表2列出了不同SR算法在Set14以及BSD200数据集的平均PSNR值。

表1 不同超分辨率算法在Set5数据集的PSNR以及SSIM对比

Table 1 The comparison of different SR strategies for PSNR( dB) and SSIM on Set5 dataset

	图像	倍数	Bicubic	ScSR	ANR	SRCNN	SRCNN-Ex	ESRCN	MCSR	MCSR-Ex
PSNR/dB	Baby	3	33.92	34.31	35.06	35.01	35.03	34.88	35.01	35.10
	Bird	3	32.59	34.09	34.49	34.91	35.10	35.16	35.41	35.57
	Butterfly	3	24.04	25.59	25.85	27.58	27.96	28.15	28.80	29.09
	Head	3	32.90	33.21	33.63	33.55	33.64	33.29	33.59	33.62
	Woman	3	28.57	29.90	30.33	30.92	31.20	31.28	31.43	31.57
	Baby	4	31.77	31.10	33.01	32.98	33.01	—	32.98	33.14
	Bird	4	30.18	29.87	31.78	31.97	32.25	—	32.41	32.49
	Butterfly	4	22.09	22.41	23.50	25.07	25.23	—	25.89	26.11
	Head	4	31.59	31.09	32.28	32.18	32.36	—	32.34	32.49
	Woman	4	26.46	26.33	27.83	28.21	28.41	—	28.81	28.92
SSIM	Baby	3	0.904 2	0.921 0	0.922 5	0.920 9	0.920 5	0.920 6	0.919 1	0.920 5
	Bird	3	0.924 9	0.943 3	0.949 4	0.949 4	0.949 8	0.951 5	0.951 8	0.925 3
	Butterfly	3	0.819 7	0.875 1	0.871 8	0.901 1	0.908 1	0.909 8	0.921 3	0.924 4
	Head	3	0.800 1	0.822 3	0.824 1	0.821 5	0.823 2	0.823 4	0.823 0	0.824 1
	Woman	3	0.889 1	0.913 2	0.917 0	0.923 6	0.926 1	0.926 8	0.928 0	0.929 0
	Baby	4	0.857 7	0.838 4	0.881 1	0.877 8	0.896 0	—	0.876 1	0.878 5
	Bird	4	0.871 8	0.864 2	0.901 7	0.901 8	0.900 1	—	0.905 1	0.910 3
	Butterfly	4	0.733 8	0.765 4	0.788 6	0.841 6	0.850 8	—	0.864 1	0.863 3
	Head	4	0.755 1	0.735 5	0.776 2	0.772 5	0.776 7	—	0.777 4	0.782 1
	Woman	4	0.831 2	0.824 7	0.861 8	0.870 9	0.888 6	—	0.891 3	0.898 9
PSNR/dB	平均值	3	30.40	31.42	31.87	32.39	32.58	32.55	32.84	32.99
		4	28.42	28.15	29.69	30.08	30.31	—	30.48	32.63
SSIM	平均值	3	0.867 6	0.894 9	0.896 9	0.903 3	0.905 5	0.905 4	0.908 6	0.910 4
		4	0.810 0	0.805 6	0.841 8	0.852 9	0.862 4	—	0.862 8	0.866 6

表2 不同超分辨率算法在Set14以及BSD200数据集的平均PSNR对比

Table 2 The comparison of different SR strategies for average PSNR ( dB) on Set5&amp;BSD200 datasets

									/dB
数据集	倍数	Bicubic	ScSR	ANR	SRCNN	SRCNN-Ex	ESRCN	MCSR	MCSR-Ex
Set14	3	27.53	28.31	28.64	29.01	29.10	29.08	29.28	29.39
	4	26.00	25.94	26.85	27.20	27.44	—	27.55	27.62
BSD200	3	27.70	28.20	28.39	28.65	28.78	28.65	29.03	29.10
	4	26.39	26.26	26.50	26.71	26.84	—	27.41	27.55



对比发现 MCSR 同传统 SR 算法相比在 PSNR 指标以及 SSIM 指标方面均有不同程度的提升。

训练速度方面,每迭代  $1 \times 10^4$  次: SRCNN 用时 61 s; SRCNN-Ex 用时 107 s; MCSR 用时 80 s。为了衡量不同方法的计算效率,以 Set5 数据集以及 Set14 数据集作为测试集,统计不同方法的所需的平均每幅图的 SR 重构时间,结果如表 3 所示。

从表 3 可以看出,ESRCN 所需时间最短,本文方法所需时间略高于 SRCNN-Ex,但是差距不大。考虑到 MCSR 在 SR 性能上的提升,其计算时间的牺牲是可以接受的。图 6 以及图 7 展示的两幅图像分别来自 Set14 数据集以及 Set5 数据集,放大 3 倍。

表 3 不同超分辨率算法平均重建时间对比

Table 3 The comparison of different SR strategies for average reconstruction time

数据集	SRCNN	SRCNN-Ex	ESRCN	MCSR	MCSR-Ex
Set5	0.35	1.09	0.11	1.12	1.28
Set14	0.68	2.15	0.26	2.21	2.37

我们分别放大显示 ppt3 图的标题边界区域以及 Butterfly 图的斑点区域,对比观察不同方法 SR 性能的差异。观察图 6 发现,MCSR 基本消除了黑色标题周围的灰色阴影,和其他 SR 算法相比提升明显。观察图 7 发现,MCSR 基本消除了斑点周围的振铃现象,而这一现象在其他算法的结果中均不同程度地存在。为了进一步验证本文算法的有效性,我们将放大倍数扩大到 8 倍并对比观察不同算法的表现,结果如图 8 所示。可以非常明显的观察到,MC-SR 算法可以产生更加清晰地边界线。综上所述可知,同传统方法相比本文提出的 MCSR 可以生成更加细腻清晰的边缘,和原图更加相似,更加符合人眼主观感受。

4 结 论

本文主要通过对 SRCNN 模型进行加深以及加宽处理达到提高 SR 性能的目的。其中加深操作通过将大尺寸的卷积核替换为多层  $3 \times 3$  卷积核达到;加宽操作通过将非线性映射部分改为多通道卷积实



图 6 放大 3 倍时 ppt3 图

Fig. 6 The “ppt3” image with an upscaling factor 3 ( (a) original image; (b) Bicubic result; (c) ScSR result; (d) ANR result; (e) SRCNN result; (f) SRCNN-Ex result; (g) MCSR result; (h) MCSR-Ex result)

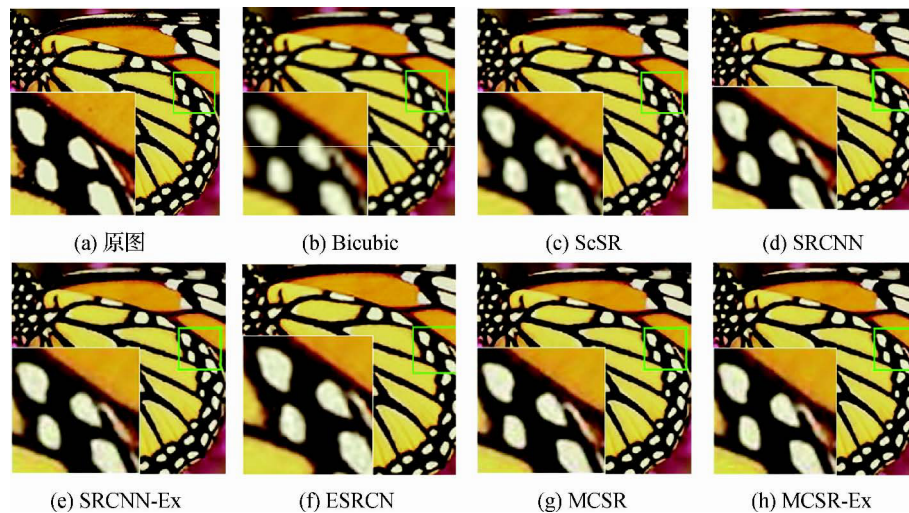


图 7 放大 3 倍时 butterfly 图

Fig. 7 The “butterfly” image with an upscaling factor 3 ( (a) original image; (b) Bicubic result; (c) ScSR result; (d) SRCNN result; (e) SRCNN – Ex result; (f) ESRCN result; (g) MCSR result; (h) MCSR – Ex result)

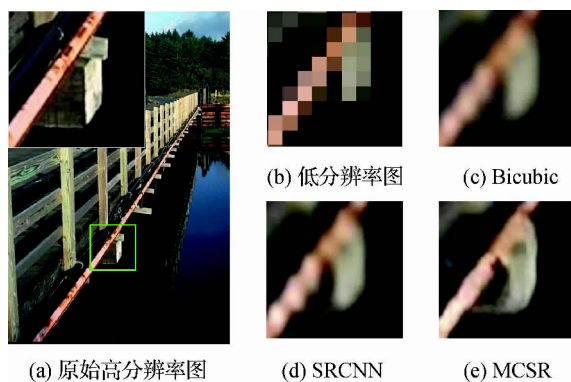


图 8 放大 8 倍时 22013 图

Fig. 8 The “22013” image with an upscaling factor 8 ( (a) original HR image; (b) LR image; (c) Bicubic result; (d) SRCNN result; (e) MCSR result)

现 不同通道采用不同形式的卷积操作 提取不同的图像特征。另外,我们增加从模型的输入直达输出的残差链接并采用 MSRA 初始化方法,进而加快模型收敛速度。实验结果显示 MCSR 模型在主观与客观两种评价方式上均优于经典方法。这表明:若要提升模型的 SR 性能,必然要求模型可以提取更加丰富的 LR 图像特征并且更有效地利用特征生成 HR 图像,将来可以沿这条路线做更多的工作。

## 参考文献(References)

- [1] Harris J L. Diffraction and resolving power [J]. Journal of the Optical Society of America (1917-1983), 1964, 54: 931-933. [DOI: 10.1364/JOSA.54.000931]
- [2] Ur H, Gross D. Improved resolution from subpixel shifted pictures [J]. CVGIP: Graphical Models and Image Processing, 1992, 54(2): 181-186. [DOI: 10.1016/1049-9652(92)90065-6]
- [3] Parker J A, Kenyon R V, Troxel D E. Comparison of interpolating methods for image resampling [J]. IEEE Transactions on medical imaging, 1983, 2(1): 31-39. [DOI: 10.1109/TMI.1983.4307610]
- [4] Irani M, Peleg S. Improving resolution by image registration [J]. CVGIP: Graphical models and image processing, 1991, 53(3): 231-239. [DOI: 10.1016/1049-9652(91)90045-L]
- [5] Schultz R R, Stevenson R L. A Bayesian approach to image expansion for improved definition [J]. IEEE Transactions on Image Processing, 1994, 3(3): 233-242. [DOI: 10.1109/83.287017]
- [6] Yang J, Wright J, Huang T, et al. Image super-resolution as sparse representation of raw image patches [C]//Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008: 1-8. [DOI: 10.1109/CVPR.2008.4587647]
- [7] Zeyde R, Elad M, Protter M. On single image scale-up using sparse representations [C]// International Conference on Curves and Surfaces. Springer-Verlag, 2010: 711-730. [DOI: 10.1007/978-3-642-27413-8\_47]
- [8] Aharon M, Elad M, Bruckstein A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation [J]. IEEE Transactions on Signal Processing, 2006, 54(11): 4311-4322. [DOI: 10.1109/TSP.2006.881199]

- [9] Timofte R , De V , Gool L V. Anchored neighborhood regression for fast example-based super-resolution [C]// IEEE International Conference on Computer Vision. IEEE , 2013: 1920-1927. [DOI: 10.1109/ICCV.2013.241]
- [10] Chang H , Yeung D Y , Xiong Y. Super-resolution through neighbor embedding [C]//IEEE Computer Society Conference on Computer Vision & Pattern Recognition. IEEE Computer Society , 2004: 275-282. [DOI: 10.1109/CVPR.2004.1315043]
- [11] Krizhevsky A , Sutskever I , Hinton G E. ImageNet classification with deep convolutional neural networks [C]// International Conference on Neural Information Processing Systems. Curran Associates Inc. 2012: 1097-1105. [DOI: 10.1145/3065386]
- [12] Zeiler M D , Fergus R. Visualizing and understanding convolutional networks [C] //European conference on computer vision. Springer , Cham , 2014: 818-833. [DOI: 10.1007/978-3-319-10590-4\_53]
- [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. [DOI: 10.1109/CVPR.2015.7298594]
- [14] 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. [DOI: 10.1109/CVPR.2016.90]
- [15] Girshick R , Donahue J , Darrell T , et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587. [DOI: 10.1109/CVPR.2014.81]
- [16] He K , Zhang X , Ren S , et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [C]//European Conference on Computer Vision. Springer , Cham , 2014: 346-361. [DOI: 10.1007/978-3-319-10578-9\_23]
- [17] Babenko A , Slesarev A , Chigorin A , et al. Neural codes for image retrieval [C]//European conference on computer vision. Springer , Cham , 2014: 584-599. [DOI: 10.1007/978-3-319-10590-4\_38]
- [18] Dong C , Loy CC , 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. [DOI: 10.1109/TPAMI.2015.2439281]
- [19] Dong C , Loy CC , He K , et al. Learning a deep convolutional network for image super-resolution [C]//European Conference on Computer Vision. Springer , Cham , 2014: 184-199. [DOI: 10.1007/978-3-319-10593-2\_13]
- [20] He K , Zhang X , Ren S , et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification [C]//Proceedings of the IEEE international conference on computer vision. 2015: 1026-1034. [DOI: 10.1109/ICCV.2015.123]
- [21] Glorot X , Bengio Y. Understanding the difficulty of training deep feedforward neural networks [J]. Journal of Machine Learning Research , 2010 , 9: 249-256.
- [22] Hinton G E , Osindero S , Teh Y W. A fast learning algorithm for deep belief nets [J]. Neural computation , 2006 , 18 ( 7 ) : 1527-1554. [DOI: 10.1162/neco.2006.18.7.1527]
- [23] 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. [DOI: 10.1109/CVPR.2016.207]