

# 基于小波域的深度残差网络图像超分辨率算法\*



段立娟<sup>1,2,3</sup>, 武春丽<sup>1,2,3</sup>, 恩擎<sup>1,2,3</sup>, 乔元华<sup>4</sup>, 张韵东<sup>5</sup>, 陈军成<sup>1</sup>

<sup>1</sup>(北京工业大学 信息学部, 北京 100124)

<sup>2</sup>(可信计算北京市重点实验室, 北京 100124)

<sup>3</sup>(信息安全等级保护关键技术国家工程实验室, 北京 100124)

<sup>4</sup>(北京工业大学 应用数理学院, 北京 100124)

<sup>5</sup>(数字多媒体芯片技术国家重点实验室(北京中星微电子有限公司), 北京 100191)

通讯作者: 陈军成, E-mail: juncheng@bjut.edu.cn

**摘要:** 单幅图像超分辨率(SISR)是指从一张低分辨率图像重建高分辨率图像.传统的神经网络方法通常在图像的空间域进行超分辨率重构,但这些方法常在重构过程中忽略重要的细节.鉴于小波变换能够将图像内容的“粗略”和“细节”特征进行分离,提出一种基于小波域的深度残差网络(DRWSR).不同于其他传统的卷积神经网络直接推导高分辨率图像(HR),该方法采用多阶段学习策略,首先推理出高分辨率图像对应的小波系数,然后重建超分辨率图像(SR).为了获取更多的信息,该方法采用一种残差嵌套残差的灵活可扩展的深度神经网络.此外,提出的神经网络模型采用结合图像空域与小波域的损失函数进行优化求解.所提出的方法在 Set5、Set14、BSD100、Urban100 等数据集上进行实验,实验结果表明,该方法的视觉效果和峰值信噪比(PSNR)均优于相关的图像超分辨率方法.

**关键词:** 单幅图像超分辨率;小波变换;卷积神经网络;残差块

**中图法分类号:** TP391

中文引用格式: 段立娟,武春丽,恩擎,乔元华,张韵东,陈军成.基于小波域的深度残差网络图像超分辨率算法.软件学报,2019,30(4):941-953. <http://www.jos.org.cn/1000-9825/5663.htm>

英文引用格式: Duan LJ, Wu CL, En Q, Qiao YH, Zhang YD, Chen JC. Deep residual network in wavelet domain for image super-resolution. Ruan Jian Xue Bao/Journal of Software, 2019,30(4):941-953 (in Chinese). <http://www.jos.org.cn/1000-9825/5663.htm>

## Deep Residual Network in Wavelet Domain for Image Super-resolution

DUAN Li-Juan<sup>1,2,3</sup>, WU Chun-Li<sup>1,2,3</sup>, EN Qing<sup>1,2,3</sup>, QIAO Yuan-Hua<sup>4</sup>, ZHANG Yun-Dong<sup>5</sup>, CHEN Jun-Cheng<sup>1</sup>

<sup>1</sup>(Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China)

<sup>2</sup>(Beijing Key Laboratory of Trusted Computing, Beijing 100124, China)

<sup>3</sup>(National Engineering Laboratory for Key Technologies of Information Security Level Protection, Beijing 100124, China)

<sup>4</sup>(College of Applied Science, Beijing University of Technology, Beijing 100124, China)

<sup>5</sup>(State Key Laboratory of Digital Multimedia Chip Technology (Vimicro Corp.), Beijing 100191, China)

**Abstract:** Single Image Super Resolution (SISR) refers to the reconstruction of high resolution images from a low resolution image. Traditional neural network methods typically perform super-resolution reconstruction in the spatial domain of an image, but these methods

\* 基金项目: 国家重点研发计划(2017YFC0803705); 国家自然科学基金(61572004, 61771026); 北京市自然科学基金-市教委联合资助项目(KZ201910005008); 青海省创新平台建设专项(2016-ZJ-Y04)

Foundation item: National Key R&D Program of China (2017YFC0803705); National Natural Science Foundation of China (61572004, 61771026); Key Project of Beijing Municipal Education Commission (KZ201910005008); Innovation Platform Construction of Qinghai Province of China (2016-ZJ-Y04)

本文由“多媒体数据的知识关联与理解专题”特约编辑蒋树强研究员、刘青山教授、孙立峰教授、李波教授推荐.

收稿时间: 2018-04-15; 修改时间: 2018-06-13; 采用时间: 2018-09-30

often ignore important details in the reconstruction process. In view of the fact that wavelet transform can separate the “rough” and “detail” features of image content, this study proposes a wavelet-based deep residual network (DRWSR). Different from other traditional convolutional neural networks, the high-resolution image (HR) is directly derived. This method uses a multi-stage learning strategy to first infer the wavelet coefficients corresponding to the high-resolution image and then reconstruct the super-resolution image (SR). In order to obtain more information, the method uses a flexible and scalable deep neural network with residual nested residuals. In addition, the proposed neural network model is optimized by combining the loss function of image space and wavelet domain. The proposed method is carried out on Set5, Set14, BSD100, Urban100, and other datasets. The experimental results show that the proposed visual effect and peak signal-to-noise ratio (PNSR) are better than the related image super-resolution method.

**Key words:** single image super-resolution; wavelet transform; convolutional neural network; residual block

## 1 引言

在图像处理领域,为了获得更高分辨率的图像,通常采用超分辨率方法来重建低分辨率图像的细节信息.一般来讲,超分辨率技术(super-resolution,简称 SR)是指从一副或者多幅低分辨图像重建出高分辨图像的数字图像处理技术.单幅图像超分辨率(single image super-resolution,简称 SISR)的重点在于如何从一张低分辨率图像重建出高分辨率图像的局部细节.该问题已经发展成为图像处理领域一个重要的研究方向<sup>[1]</sup>.因其可以恢复一些高频细节,该技术被广泛应用于需要大量细节信息的图像处理领域,如医学成像<sup>[2]</sup>、卫星成像<sup>[3]</sup>、人脸认证<sup>[4-8]</sup>以及公共安全监控领域等等.

基于实例的 SR 方法通过使用大型图像数据集学习从低分辨率(low resolution,简称 LR)图像到高分辨率(high resolution,简称 HR)图像的映射,已经证明能够达到较好的效果.许多机器学习算法,包括字典学习<sup>[9]</sup>、局部线性回归<sup>[10]</sup>和随机森林<sup>[11]</sup>等都被应用到这个领域中.近些年来,基于卷积神经网络(convolutional neural network,简称 CNN)的方法,凭借其强大的学习能力被广泛应用于计算机视觉任务之中,在目标识别、分割、光流和超分辨率领域,均取得了卓越的进步.尽管基于卷积神经网络的图像超分辨率方法较传统方法取得了较大的突破,但仍存在许多问题.大部分现有的 SISR 方法<sup>[12-19]</sup>的训练过程是在图像空间中依靠逐像素均方误差的方式使网络输出尽可能地接近 HR 图像,但这种方式趋向于产生模糊和过平滑的输出,缺乏细节信息.而且,目前的工作只适合一些小且特定的缩放系数(如 x2 或者 x4).因此,SISR 问题仍然有待进一步的探究和发展.

小波变换已被证明是一种高效的特征提取算法,常被用来重现和存储多分辨率图像<sup>[20]</sup>.如图 1 所示,它可以表示一张图像在不同水平上的上下文和纹理信息.空间小波系数本身是稀疏的,因此,它能够使网络学习更加容易.

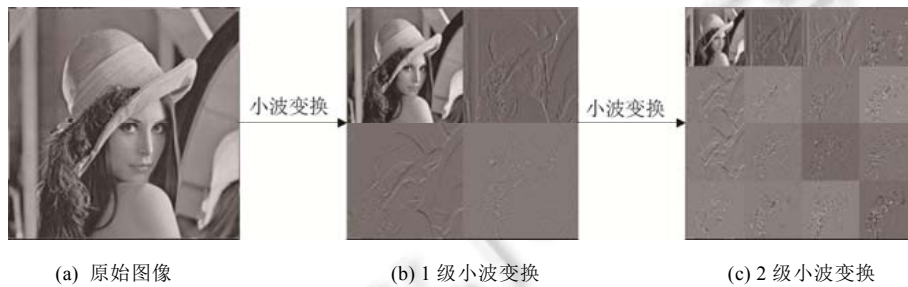


Fig.1 The example of Haar wavelet transform

图 1 Haar 小波变换示例

DWSR<sup>[21]</sup>利用小波变换的方式将重建 HR 图像转变为推断 HR 图像一系列相关的小波系数.该方法在与当前最优算法超深度超分辨率网络(VDSR)效果相当的情况下,实现了更简单的运算量和更快的速度.本文提出的方法在此方法的基础上加以改进.在输入方面,为减少计算量从而实现 LR 到 HR 的实时处理,本文直接将 LR 图像作为网络的输入.在网络架构方面,本文设计的框架分为 3 部分.特征提取网络、推理网络和重建网络.为了更

好地捕捉上下文之间的信息,以及更好地推断出缺失的细节,特征提取网络采用残差块相连的方式<sup>[22]</sup>,同时将每个残差块的输出进行跨连,作为推理网络的输入.此外,网络最终输出的小波系数个数(即网络最终输出的通道数)不再是固定值,而会随尺度而发生改变.在损失函数方面,本文不仅只考虑小波系数之间的残差损失,同时考虑最终经过小波逆变换后的 SR 图像与 HR 图像之间的残差损失,进行双重约束,实现更优的效果.

本文第 1 节对解决单幅图像的超分辨率问题的方法进行概述和总结,并简单概述本文方法的基本思想.第 2 节对已有的基于卷积神经网络的超分辨率方法进行总结.第 3 节具体描述本文方法的框架和损失函数的设计.第 4 节展示本文方法与其他相关方法在不同数据集上的结果对比,总结本文方法的优势和不足.第 5 节对本文进行总结,并对未来值得关注的研究方向进行初步探讨.

## 2 相关工作

从观察到的低分辨率图像推测合成的高分辨率图像是典型的病态逆问题.现有的算法可以根据技术手段分为两类:基于重建的方法和基于学习的方法<sup>[23]</sup>.传统学习模型的特征提取和表达能力有限,很大程度上限制了图像超分辨率重建效果的大幅度提升.而近年来基于深度学习的方法因其具有从大规模数据中学习知识的强大能力被引入到解决 SR 的问题中.

随着基于深度卷积神经网络在高视觉水平任务上的迅速发展,大量基于 CNN 的方法也被应用到计算机低视觉水平的任务中,如图像超分辨率、图像去噪等.SRCNN<sup>[12]</sup>是首个将卷积神经网络引入到图像超分辨率重建领域的神经网络模型.其网络结构由 3 个卷积层组成,分别代表图像块的特征提取、表示和重构.通过引入卷积神经网络,该方法显著提高了传统方法的重构进度.虽然 SRCNN 运作良好,但仍存在缺乏语境信息、单一尺度放大以及收敛速度慢等问题.为了解决这一问题,有研究者提出具有深层次网络的 VDSR<sup>[14]</sup>(超深度超分辨率网络).该方法基于用于图像分类的 VGG<sup>[24]</sup>网络结构设计,通过训练 HR 和 LR 图像之间的残差,并使用更高的学习速率来加速收敛.同时,通过权重共享,实现网络参数较少且重构性能较好的图像多尺度放大.此外,DRCN<sup>[15]</sup>在 VDSR 的基础上增加递归连接,实现图像层之间的信息反馈以及上下文信息关联,进一步提升效果.同时,将模型压缩成 5 层,降低训练难度.

上述方法均利用卷积神经网络学习 MR 图像(middle resolution image)到 HR 图像的映射,其中,MR 图像是通过 LR 图像利用双三次插值的方法进行上采样获取的.为了实现从 LR 到 HR 的直接映射,FSRCNN<sup>[14]</sup>利用反卷积层替代 SRCNN 模型中双三次插值的操作.在移除上采样的操作之后,该模型可以学习从低分辨率图像到高分辨率图像的直接映射,同时实现超过 40 倍速度的提升.FSRCNN 模型仅包括卷积层和反卷积层,其中卷积层对于不同的放大倍数共享卷积层的权重,因此,FSRCNN 能够利用一个单独的模型处理不同的尺度.ESPCN<sup>[17]</sup>是在 FSRCNN 的基础上设计 sub-pixel 卷积层实现上采样的操作,该方法有效地减少了总计算复杂度,能够合成一个没有棋盘伪像的干净图像.LapSRN<sup>[19]</sup>是最近提出的单幅图像超分辨率方法之一,该模型包括一个基于拉普拉斯金字塔的特征提取和图像重建部分的级联框架,并且使用 Charbonnier 损失函数代替 L2 范数损失函数,取得了较好的超分辨率重构效果.

上述基于卷积神经网络的超分辨率方法针对图像的空间域进行处理,目的是将像素值直接作为网络输出进行重构,这通常会使得结果趋向于模糊和过平滑.为了解决这一问题,PLSR<sup>[25]</sup>将感知损失代替逐像素差损失,进而优化超分辨率网络,以保证获取更多的语义信息.该策略虽然未能在 PSNR 指标上取得较好的效果,但在视觉上更加逼真,能够带来更多好的细节和边缘.SRGAN<sup>[26]</sup>在 PLSR 的基础上,将感知损失和 GAN 模型相结合,生成更真实和锐利的图像.该方法利用 GAN 中的 Generator 网络生成高分辨图像,再由 Discriminator 网络进行判断,最终得到满足条件的高分辨率图像.但是该方法提出的基于 VGG 分类网络的感知损失不能够精确地捕捉超分辨率任务中必要的细节信息.于是,SRPGAN<sup>[27]</sup>提出一种基于 GAN 模型中 Discrimination 网络的感知损失,并且使用 Charbonnier 损失函数.该方法重建的超分辨率图像更加锋利和逼真.

综上所述,目前解决该问题主要有 3 种思路:第一,设计不同的网络架构去重建高分辨率图像;第二,利用感知损失和 GAN 生成模型生成更真实和锐利的图像;第三,考虑在变换域的情况下处理该问题.小波变换将提供

的图像内容“过程”和“细节”分离的特性,与输入 LR 图像恢复图像细节的超分辨率问题具有一定相似性.因此,本方法考虑变换到小波域进行超分辨率重构.目前已有一些基于小波的方法解决超分辨率问题,但其中大部分集中在多幅图像的超分辨问题,这些方法<sup>[28-31]</sup>利用多帧低分辨率图像推断出高分辨率图像缺失的细节.针对单幅图像超分辨问题,虽然也有一些小波域的插值方法被研究,但其训练有限和预测程序简单,不足以处理普通输入图像,其 SR 结果也远差于基于深度学习的超分辨率方法.

DWSR<sup>[21]</sup>是首个在小波域下利用深度卷积神经网络结合低频和高频信息子带之间的互补信息的方法.该方法把重建高分辨率图像的问题转换为推理一系列小波变换系数,这样产生的图像边缘具有更少的伪像.该方法采用的网络是由一系列相同大小的卷积层直接连接而成,缺乏层与层之间的信息融合,而且没有考虑尺度与小波分解级数之间的对应关系,所有尺度经过最终网络训练后的结果都是尽可能地接近 HR 图像与 MR 图像一级小波分解系数之间的残差.为了更好地捕捉上下文之间的信息,并推断出更多缺失的细节,本方法在 DWSR<sup>[21]</sup>的基础上作了进一步的改进,直接将 LR 图像作为网络的输入,并且设计更为合理的网络结构.此外,本文提出的网络结构输出的小波系数个数不再是固定值,而是随着尺度的变化而发生改变.

### 3 基于小波域的深度残差网络模型

本文提出的基于小波域的深度残差网络,将小波变换的思想与深度残差网络相结合,并结合图像空间损失和小波系数损失,对网络训练加强约束.该模型应用于图像超分辨率重构问题,能够将图像各个层次的特征分隔开来,捕捉更多缺失的细节信息.

#### 3.1 小波变换

图 1 显示二维小波变换的结果,可以直观地看出,如果把变换后第 1 个小波图像表示为 LR 图像,那么其他的小波图像就是想要得到的缺失的细节.因此,本文将小波变换引入到 SISR 问题当中.本文方法的重点在于强调小波变换对于超分辨率重建问题的有效性,因此,选择最简单的哈尔小波即可满足实现描述不同频带信息的要求,当然也可以使用其他的小波.将 HR 图像利用哈尔小波变换为一系列尺寸相同的小波图像作为神经网络的训练目标.如图 2 所示,更高水平的变换就是循环地进行低通滤波和高通滤波以及下采样的操作.

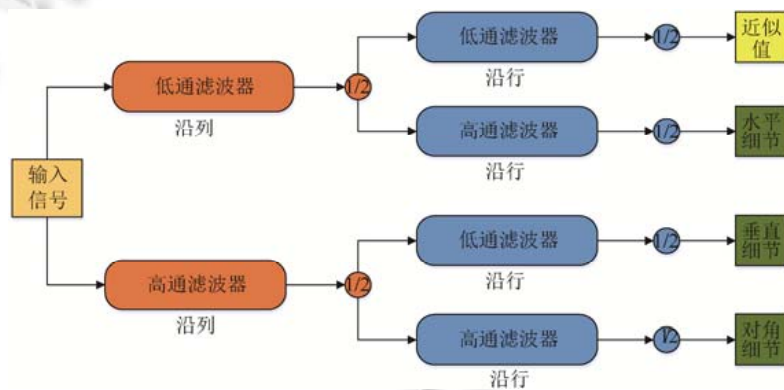


Fig.2 The procedure of 1-level 2D DWT

图 2 1 级二维小波变换过程

#### 3.2 网络架构

本文提出的方法基于空域到小波域的变换来实现超分辨率图像的重建.为了更加有效地提取特征并降低运算的复杂度,本文直接将 LR 图像作为网络的输入,最终映射到 HR 图像的小波域下.从 LR 到最终重建 HR 的整个过程其实是多个网络,但是我们把这多个网络当作一个网络进行统一、有效的学习.整个网络的学习属于多阶段,利用推理网络输出的一系列小波图像属于第 1 阶段,第 2 阶段是根据第 1 阶段的结果进行小波逆变

换,得到最终的超分辨率图像(SR).最后将小波系数的损失和图像空间像素的损失一起以梯度的形式反传回整个网络进行更新.如图3所示,本文提出的神经网络结构包括特征提取网络、推理网络和重建网络3部分.其中,特征提取网络从低分辨率图像中提取特征,推理网络则将提取的特征表示为一系列HR图像与对应MR图像小波系数的差异,重建网络利用小波逆变换重建SR图像.为了捕捉更多缺失的细节,本文利用小波系数和图像空间像素的两种损失构建一个鲁棒的损失函数来计算SR与HR的相似度,从而优化本文提出的神经网络.在第3节中将会详细阐述损失函数.

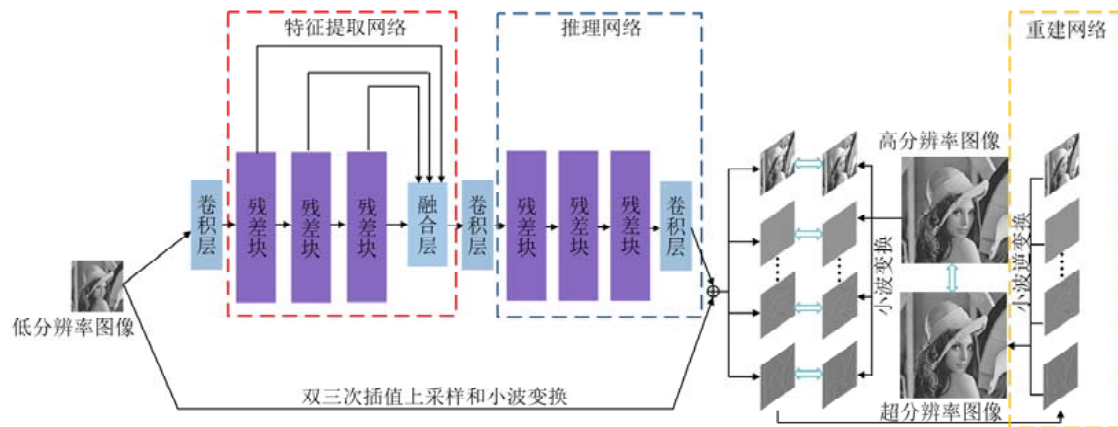


Fig. 3 Overall framework. The two parts corresponding to the two-way arrow are used to calculate the loss during network training

图3 总体框架.其中双向箭头对应的两部分在网络训练时进行损失计算

### 3.2.1 特征提取网络

特征提取网络将LR图像作为输入,通过神经网络前向传播表示为一系列特征图.特征提取网络由多个残差块级联构成.其中,残差块由两个具有相同核大小以及过滤器数目的卷积构成,其输出是由输入和经过连续两个卷积之后的结果相加构成.每个残差块作为一个单元,每个单元的输出都被传递到下一个单元,并同时进行跨连作为推理网络的输入.所有的卷积层共享相同的核大小:3×3.而为了保持特征图的大小与输入一致,将步长(stride)和边缘填充(pad)都设置为1.同时,为了获取更多、更丰富的信息,卷积层过滤器的数目随着网络的深入不断增长.

### 3.2.2 推理网络

推理网络将特征提取网络的输出作为输入,但因其维度比较大,所以使用1×1的卷积实现减少特征维度的目的.同时,为了保证推理出的小波图像与LR图像具有相同尺寸,对推理网络中所有卷积层参数的设置与特征提取网络保持一致,即卷积核大小设置为3×3,步长和边缘填充设置为1.因为小波分解系数的高度独立性,所以无需考虑通道之间的关系.不同尺度网络最终输出通道的个数是缩放倍数的平方,每个通道表示一一对应的小波图像.如图3所示,将网络最终的输出与MR图像的小波分解系数相加,然后通过小波逆变换转换到原始图像空间.正如典型的残差学习网络一样,本文提出的模型也致力于学习残差输出.因为不同尺度的小波变换图像尺寸是不同的,其对应网络最终输出的通道数也是不同的,所以需要学习多个网络.但是,不同尺度图像的训练只有最后一层卷积的参数设置有所不同,其他卷积层对于不同尺度共享权重,因此只对一个尺度的网络从头开始训练,其他尺度在已训练好的模型上进行微调即可.

### 3.2.3 重建网络

重建网络是将推理网络的输出与LR经上采样和小波变换后的结果相加作为输入,利用小波逆变换,将一系列小波图像生成对应的一张超分辨率图像,得到最终的结果.依据中间结果(一系列的小波系数)和最终SR图像,本文提出一种灵活且约束能力更强的损失函数去优化网络,它由小波系数损失和图像空间像素损失两部分



组成,总的损失函数被定义为

$$l_{total} = \lambda l_{wave} + (1 - \lambda) l_{img} \quad (1)$$

其中, $\lambda$ 和  $1-\lambda$ 分别代表小波系数损失和图像空间损失的权重,具体细节会在第 4 节加以详细说明.

**小波系数损失.**本文设定输入 LR 图像为  $x$ ,标签高分辨图像为  $y$ ,输入 LR 经双三次插值上采样后的 MR 图像为  $y_b$ , $\theta$ 是待优化的一系列网络参数.通用单幅图像超分辨率网络旨在学习给定低分辨率输入  $x$  与高分辨率图像  $y$  之间的映射关系.而本文方法提出的网络结构是学习低分辨率输入  $x$  与高分辨率图像  $y$  小波变换后系数之间的关系,网络的输出尽可能地与对应高分辨率图像小波分解后的系数接近.本文定义缩放系数  $\{r, r \geq 2\}$ 、小波变换的级数为  $m$  以及变换后小波系数的个数为  $W_N$ ,它们之间的映射关系为  $m = \log_2^r W_N$ ,  $W_N = r^2$ .

同时,将图像  $y$  和  $y_b$  进行小波分解后的系数分别表示为

$$W_{Cy} = (c_1, c_2, c_3, \dots, c_{W_N-1}, c_{W_N}),$$

$$W_{Cy_b} = (c_1, c_2, c_3, \dots, c_{W_N-1}, c_{W_N}).$$

两者分解后系数的差异(即残差)可计算为

$$\Delta W_C = \Delta W_{Cy} - \Delta W_{Cy_b} = (c_1, c_2, c_3, \dots, c_{W_N-1}, c_{W_N}) \quad (2)$$

式(2)的结果为神经网络致力于学习的目标,即  $f_{\theta}(x)$ .在图像空间里最常用的损失函数是对 HR 图像和 SR 图像逐像素进行均方误差操作.本文同样采用这种方式,不同之处是对图像对应的小波系数进行这一操作.即:

$$l_{wave} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{W_N} \|f_{\theta,j}(x) - \Delta W_{C,j}\|_2^2 \quad (3)$$

其中, $n$  为 batchsize 的大小, $i$  代表批处理图像中的第  $i$  张图像, $j$  为小波系数序列中的第  $j$  个系数.

**图像空间损失.**本文提出的神经网络结构学习到的是 MR 图像与 HR 图像小波变换后系数之间的差异,将输出的结果经过小波逆变换即可得到学习到的残差图像  $I_{\Delta SR}$ ,即:

$$I_{\Delta SR} = idwt(f_{\theta}(x)) \quad (4)$$

在得到残差图像后,与原始输入图像经过双三次插值上采样后的 MR 图像  $y_b$  相加即可得到最终的超分辨率图像  $I_{SR}$ :

$$I_{SR} = I_{\Delta SR} + y_b \quad (5)$$

考虑到之前图像空间的损失计算,为获取纹理与平滑之间的一个平衡,在小波系数损失的基础上,添加原始图像空间的损失,计算图像空间的残差损失.

$$l_{img} = \frac{1}{n} \sum_{i=1}^n \|I_{\Delta SR} - (y - y_b)\|_2^2 \quad (6)$$

其中, $n$  和  $i$  的含义参考公式(3).

## 4 实验结果及分析

本文使用 D2VIK 数据集中的 800 张图像作为训练集,其所有图像至少有一个轴上存在 2K 像素(垂直或水平).本文采用 64 的步长将数据集切割成大小为  $128 \times 128$  的图像块,获得大约 50 万左右的切块用于网络训练.在批训练中,每次随机选择 256 个高分辨图像块作为标签,利用双三次内核插值法进行下采样获取低分辨率图像块,作为网络的输入.

对于测试集,本文对 5 个公共基准数据集进行了实验:SET5<sup>[32]</sup>、SET14<sup>[33]</sup>、BSD100<sup>[34]</sup>、URBAN100<sup>[35]</sup>和 MANGA109<sup>[36]</sup>.SET5、SET14 和 BSD100 数据集中的图像由自然场景组成,URBAN100 数据集包含具有挑战性的城市景观图像以及不同频段的细节,MANGA109 是日本漫画的数据集.

### 4.1 训练细节

在训练时,本文将 HR 图像对应的小波系数作为训练标准,其中小波分解级数通过与缩放系数的映射关系来获取.然后对每一次迭代后的结果利用公式(1)进行损失计算.同时采用 ADAM 的优化器(参数设置: $\beta_1=0.9, \beta_2=$

0.999)更新权重和偏置.学习率  $lr$  初始化为  $2e-4$ ,并将学习率衰减因子设置为 0.005.网络的输入可以是单通道的灰度图或 3 通道的彩色图像,本文网络的训练选择后者.如表 1 所示,特征提取网络由 3 个残差块组成,其中过滤器数目从 64~256 依次呈二倍增多.特征提取网络融合的特征经过  $1 \times 1$  卷积实现降维,作为推理网络的输入.推理网络同样由 3 个残差块组成,与特征提取部分相反的是,过滤器数目从 256~64 依次呈 2 倍减少,最后再经过一个卷积获取指定的通道数  $3 \times r^2$  ( $r$  是缩放倍数).

同时,开始时先利用单个的损失函数进行训练,即( $\lambda=0$ )和( $\lambda=1$ ),通过训练好的结果来设定权重,获取最优的参数选择.最终,设定参数为  $\lambda=0.99$ .

**Table 1** The specific settings of the network parameters proposed by this method

**表 1** 本文方法提出的网络参数的具体设置

网络	层操作	DRWSR_x2	DRWSR_x4	DRWSR_x8
特征提取网络	卷积	3×3 conv, 64, stride 1, padding 1		
	残差块(1)	[3×3 conv, 64, stride 1, padding 1]		
	卷积	1×1 conv,128,stride 1, padding 0		
	残差块(2)	[3×3 conv, 128, stride 1, padding 1]		
	卷积	1×1 conv,256,stride 1, padding 0		
	残差块(3)	[3×3 conv, 256, stride 1, padding 1]		
推理网络	卷积	1×1 conv,256,stride 1, padding 0		
	残差块(4)	[3×3 conv, 256, stride 1, padding 1]		
	卷积	1×1 conv,128,stride 1, padding 0		
	残差块(5)	[3×3 conv, 128, stride 1, padding 1]		
	卷积	1×1 conv,64,stride 1, padding 0		
	残差块(6)	[3×3 conv, 64, stride 1, padding 1]		
重建网络	卷积	3×3 conv, 12	3×3 conv, 48	3×3 conv, 192
	小波逆变换	Haar 小波变换		

### 4.2 损失函数分析

为证明所使用损失函数的优越性,本文针对不同的损失函数做了一系列对比实验.在表 2 中展示了分别利用小波系数损失函数、图像空间损失函数以及结合两者的损失函数训练后的结果(缩放系数为 x4,数据集 Set 5 和 Set 14).实验结果表明,无论是峰值信噪比还是结构相似性,采用结合方式产生的效果都是最优的,进一步证明两者结合这一策略是可行且有效的.

**Table 2** Quality evaluation of different loss functions

**表 2** 不同损失函数的质量评价

损失函数	尺度	Set 5		Set 14	
		PSNR	SSIM	PSNR	SSIM
$l_{wav}$	x4	31.79	0.888	28.24	0.774
$l_{img}$		31.70	0.886	28.14	0.773
$l_{total}$		<b>31.81</b>	<b>0.888</b>	<b>28.28</b>	<b>0.774</b>

### 4.3 质量评价

本文采用广泛被用于评价图像质量的两个指标来判断 SR 结果的好坏:峰值信噪比(PSNR)和结构相似性(SSIM).将本文提出的方法与相关优异的方法进行比较分析.对于缩放尺度,采用 2x、4x 和 8x 测试模型.

表 3 中总结了本文方法与其他方法在不同数据集(Set 5,Set 14,BSD100,Urbn100,Manga109)以及不同尺度上(x2,x4 和 x8)结果的(峰值信噪比和结构相似性)比较.最好的 3 个结果分别以加粗、下划线和阴影表示.可以直观地看出,本文提出的方法在结果上并没有超过 EDSR<sup>[37]</sup>.究其原因,在网络结构上,两者都采用残差块作为网

络的基本组成单元.所不同的是,本文提出的方法仅使用了6个残差块,EDSR<sup>[37]</sup>使用了32个.而之前的相关工作已经可以证明,结构很深的网络使得能够根据更多的像素即更大的区域来预测目标像素信息,超分辨率重建的效果也更好.但在其他方面,本文提出的方法也具有一定的优势,如训练速度和运行速度更快、硬件要求低等.而与其他方法相比,本文方法对于缩放尺度 x4 和 x8,实验结果最优.

Table 3 Quality evaluation of different algorithms

表3 不同方法的质量评价

算法	尺度	SET 5		SET 14		BSD100		URBAN100		MANGA109	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	x2	33.69	0.931	30.25	0.87	29.57	0.844	26.89	0.841	30.86	0.936
A+ <sup>[10]</sup>		36.69	0.955	32.32	0.906	31.24	0.887	29.25	0.895	35.37	0.968
SRCNN <sup>[12]</sup>		36.72	0.955	32.51	0.908	31.38	0.889	29.53	0.896	35.76	0.968
SCN <sup>[13]</sup>		36.58	0.954	32.35	0.905	31.26	0.885	29.52	0.897	35.51	0.967
VDSR <sup>[14]</sup>		37.53	0.959	33.05	0.913	31.90	0.896	30.77	0.914	37.22	0.975
DWSR <sup>[21]</sup>		37.43	0.957	33.07	0.911	31.80	0.894	30.46	0.916	24.08	0.846
DRCN <sup>[15]</sup>		37.63	0.959	33.06	0.912	31.85	0.895	30.76	0.914	37.63	0.974
LapSRN <sup>[19]</sup>		37.52	0.959	33.08	0.913	31.80	0.895	30.41	0.910	37.27	0.974
DRRN <sup>[16]</sup>		37.74	0.959	33.23	0.914	32.05	0.897	31.23	0.919	37.92	0.976
EDSR <sup>[37]</sup>		<b>38.11</b>	<b>0.960</b>	<b>33.92</b>	<b>0.919</b>	<b>32.32</b>	<b>0.901</b>	<b>32.93</b>	<b>0.935</b>	<b>39.10</b>	<b>0.977</b>
DRWSR(ours)	37.47	0.957	32.35	0.909	31.91	0.894	31.45	0.920	38.02	0.977	
Bicubic	x4	28.43	0.811	26.01	0.704	25.97	0.67	23.15	0.66	24.93	0.79
A+ <sup>[10]</sup>		30.32	0.86	27.34	0.751	26.83	0.711	24.34	0.721	27.03	0.851
SRCNN <sup>[12]</sup>		30.5	0.863	27.52	0.753	26.91	0.712	24.53	0.725	27.66	0.859
SCN <sup>[13]</sup>		30.41	0.863	27.39	0.751	26.88	0.711	24.52	0.726	27.39	0.857
VDSR <sup>[14]</sup>		31.35	0.883	28.02	0.768	27.29	0.726	25.18	0.754	28.83	0.887
DWSR <sup>[21]</sup>		31.39	0.883	28.04	0.770	27.25	0.724	25.26	0.755	24.23	0.816
DRCN <sup>[15]</sup>		31.54	0.884	28.03	0.768	27.24	0.725	25.14	0.752	28.98	0.887
LapSRN <sup>[19]</sup>		31.54	0.885	28.19	0.772	27.32	0.727	25.21	0.756	29.09	0.890
DRRN <sup>[16]</sup>		31.68	0.888	28.21	0.772	27.38	0.728	25.44	0.764	29.46	0.896
EDSR <sup>[37]</sup>		<b>32.46</b>	<b>0.897</b>	<b>28.80</b>	<b>0.788</b>	<b>27.71</b>	<b>0.742</b>	<b>26.64</b>	<b>0.803</b>	<b>31.02</b>	<b>0.915</b>
DRWSR(ours)	31.81	0.888	28.28	0.774	27.43	0.731	25.66	0.773	30.18	0.903	
Bicubic	x8	24.4	0.658	23.1	0.566	23.67	0.548	20.74	0.516	21.47	0.65
A+ <sup>[10]</sup>		25.53	0.693	23.89	0.595	24.21	0.569	21.37	0.546	22.39	0.681
SRCNN <sup>[12]</sup>		25.33	0.690	23.76	0.591	24.13	0.566	21.29	0.544	22.46	0.695
SCN <sup>[13]</sup>		25.59	0.706	24.02	0.603	24.3	0.573	21.52	0.56	22.68	0.701
VDSR <sup>[14]</sup>		25.93	0.724	24.26	0.614	24.49	0.583	21.7	0.571	23.16	0.725
DWSR <sup>[21]</sup>		-	-	-	-	-	-	-	-	-	-
DRCN <sup>[15]</sup>		25.93	0.723	24.25	0.614	24.49	0.582	21.71	0.571	23.2	0.724
LapSRN <sup>[19]</sup>		26.15	0.738	24.35	0.620	24.54	0.586	21.81	0.581	23.39	0.735
DRRN <sup>[16]</sup>		26.18	0.738	24.42	0.622	24.59	0.587	21.88	0.583	23.60	0.742
EDSR <sup>[37]</sup>		<b>26.97</b>	<b>0.775</b>	<b>24.94</b>	<b>0.640</b>	<b>24.80</b>	<b>0.596</b>	<b>22.47</b>	<b>0.620</b>	<b>24.58</b>	<b>0.778</b>
DRWSR(ours)	26.74	0.758	24.74	0.632	24.74	0.592	22.23	0.601	24.25	0.757	

图4~图9展示了从上述几种数据集挑选的例子图像在不同方法重建后视觉效果比较的情况.如图4和图7所示,本文方法生成的图像线条不仅没有扭曲,而且更加清晰和锐利,而其他结果中线条边缘比较模糊,有的甚至还有重影的现象.图5所示本文方法产生的结果比较好地保留了字母的边缘信息,视觉上看起来更加完整和规范.从图6中可以看出,本文的结果较好地保留了两个桥洞之间的石缝,而其他方法显示的结果并不能很好地体现这一细节信息.如图8所示,本文的结果将斑马身上的纹路更加清晰地恢复出来,而其他的结果则较为模糊.在图9中,能够明显地看出来,本文方法产生的结果将人行道上白色线条笔直且清晰地恢复出来,边缘的部分也更加锐利.而其他结果则比较模糊,线条边缘呈锯齿状.

总的来说,由示例图可以很明显地看出,无论是缩放系数为 x4 还是 x8,相较除 EDSR<sup>[37]</sup>以外的其他方法,本文采用的方法能够更好地保留细节信息,同时使边缘部分更加锐利和清晰,拥有更好的重建效果.



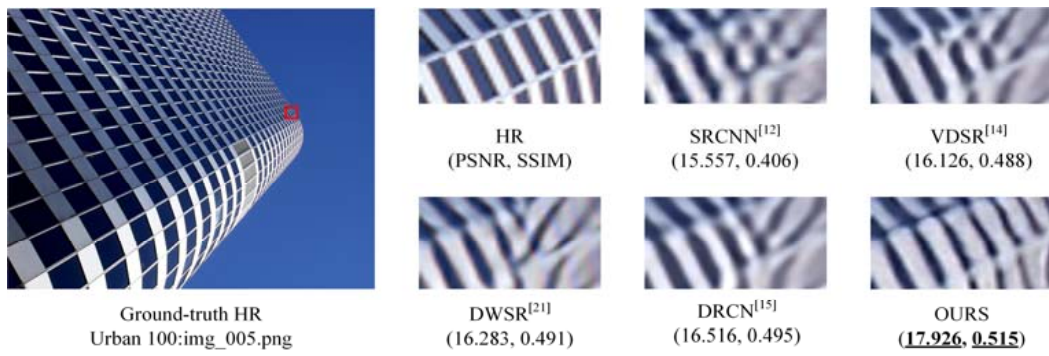


Fig.4 Super-resolution results of “img\_005.png (Urban 100)” with scale factor x4

图 4 缩放因子为 x4 的图像 img\_005.png(Urban100)的超分辨率结果

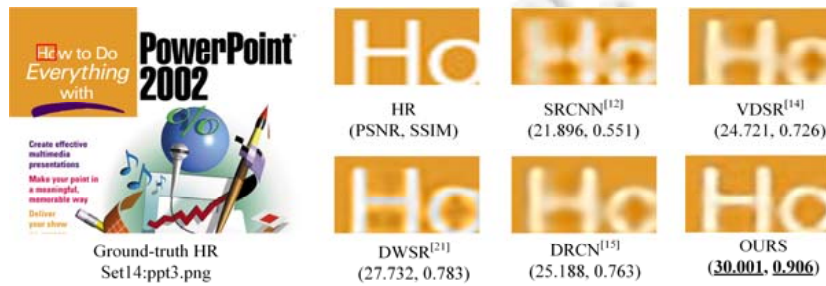


Fig.5 Super-resolution results of “ppt3.png (Set14)” with scale factor x4

图 5 缩放因子为 x4 的图像 ppt3.png(Set14)的超分辨率结果

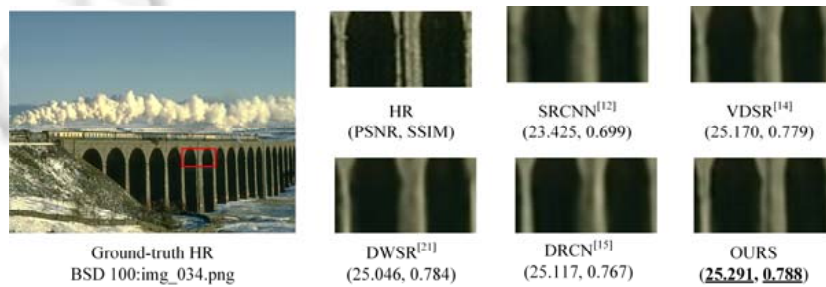


Fig.6 Super-resolution results of “img\_034.png (BSD100)” with scale factor x4

图 6 缩放因子为 x4 的图像 img\_034.png(BSD100)的超分辨率结果

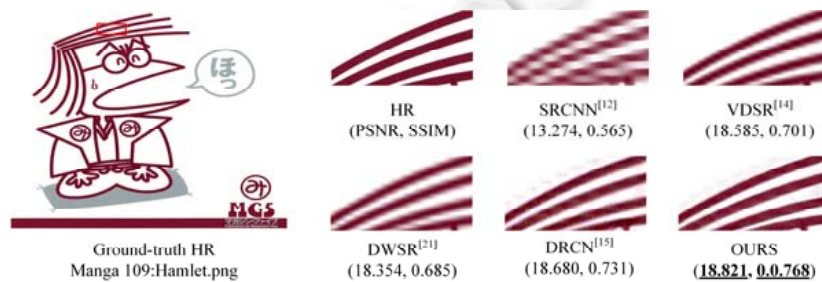


Fig.7 Super-resolution results of “Hamlet.png (Manga 109)” with scale factor x8

图 7 缩放因子为 x8 的图像 Hamlet.png(Manga 109)的超分辨率结果

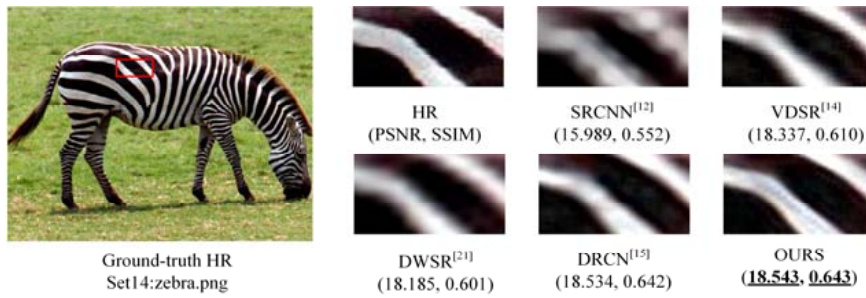


Fig.8 Super-resolution results of “zebra.png (Set14)” with scale factor x8  
 图 8 缩放因子为 x8 的图像 zebra.png(Set14)的超分辨率结果

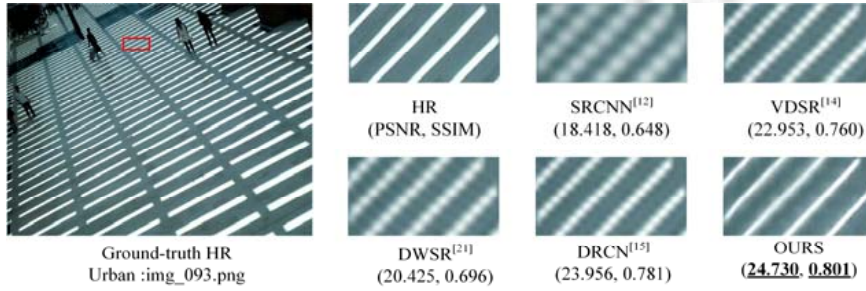


Fig.9 Super-resolution results of “img\_93.png (Urban 100)” with scale factor x8  
 图 9 缩放因子为 x8 的图像 img\_93.png(Urban 100)的超分辨率结果

4.3.1 与 SRGAN 对比

对于利用对抗网络生成的 SR 图像,单从 PSNR 和 SSIM 指标上来看,效果并不好.但从图 10 所示的视觉效果上来看,重建后的 SR 图像更加逼真和锐利.但是如图 11 所示,对细节部分进行放大操作后可以看见,SRGAN 生成的图像多了一些伪像,合成了很多奇怪的纹路.这一结果不利于后续高水平计算机视觉任务的进行.



Fig.10 Super-resolution results of “00016.png (CelebA)” with scale factor x2  
 图 10 缩放因子为 x2 的图像 00016.png(CelebA)的超分辨率结果

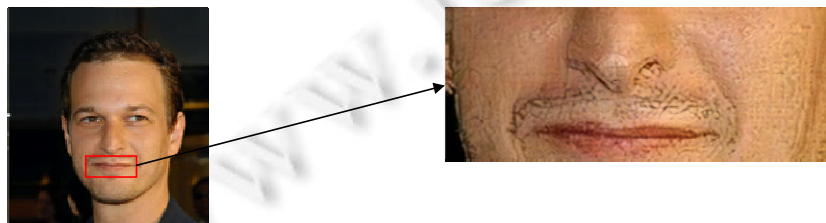


Fig.11 Result of zooming in the details of image generated by SRGAN  
 图 11 对 SRGAN 生成图像细节放大的结果

#### 4.4 局限性

从结果可以看出,本文提出的网络在缩放系数  $x_4$  和  $x_8$  的情况下,效果会有大幅度的提升.对于小的缩放倍数  $x_2$ ,效果也不错.但是并没有超越其他最优的方法.通过分析小波变换的特性可以发现,随着缩放因子的增大,小波变换的水平也在提高,图像的细节信息进一步被细化,可以通过网络更好地去推理图像缺失的细节,从而实现更好的效果.实验结果也进一步说明,利用小波分解的方式更适合于较大的缩放系数.同时,考虑到缩放倍数与最终输出的小波系数个数的映射关系,本文的方法只能处理特定的  $2^n$  的缩放倍数,无法处理一般方法中采用的  $x_3$  的情况.

#### 5 结论

本文提出一种基于小波域的深度残差网络的图像超分辨率算法.在网络设计方面,不仅结构简单,并且与其他神经网络结构相比,能够更加有效地捕捉图像细节信息.同时,使用跨连和残差学习的方式来减少训练模型的难度.在损失函数方面,本文同时使用原始图像空间域和小波域下产生的损失,加强网络训练的约束.实验结果表明,本文提出的算法在视觉效果和峰值信噪比(PNSR)方面都取得了更好的效果.

#### References:

- [1] Glasner D, Bagon S, Irani M. Super-resolution from a single image. In: Proc. of the IEEE Int'l Conf. on Computer Vision. 2009. 349–356. [doi: 10.1109/ICCV.2009.5459271]
- [2] Shi W, Caballero J, Ledig C, Zhuang X, Bai W, Bhatia K. Cardiac image super-resolution with global correspondence using multi-atlas PatchMatch. In: Proc. of the 16th Int'l Conf. on Medical Image Computing and Computer-Assisted Intervention—MICCAI 2013. 2013. 9–16.
- [3] Thornton. MW, Atkinson PM, Holland D. Sub-pixel mapping of rural land cover objects from fine spatial resolution satellite sensor imagery using super-resolution pixel-swapping. Int'l Journal of Remote Sensing, 2006,27(3):473–491. [doi: 10.1080/01431160500207088]
- [4] Bilgazyev E, Efraty B, Shah SK, Kakadiaris IA. Improved face recognition using super-resolution. In: Proc. of the Int'l Joint Conf. on Biometrics (IJCB). 2011. 1–7.
- [5] Jian M, Lam KM. Simultaneous hallucination and recognition of low-resolution faces based on singular value decomposition. IEEE Trans. on Circuits & Systems for Video Technology, 2015,25(11):1761–1772. [doi: 10.1109/TCSVT.2015.2400722]
- [6] Jian M, Lam KM, Dong J. A novel face-hallucination scheme based on singular value decomposition. Pattern Recognition, 2013, 46(11):3091–3102.
- [7] Jian M, Lam KM, Dong J. Facial-feature detection and localization based on a hierarchical scheme. Information Sciences, 2014, 262(3):1–14.
- [8] Jian M, Lam KM. Face-image retrieval based on singular values and potential-field representation. Signal Processing, 2014,100(7): 9–15.
- [9] Yang J, Wright J, Huang TS, *et al.* Image super-resolution via sparse representation. IEEE Trans. on Image Processing, 2010, 19(11):2861–2873.
- [10] Timofte R, Smet VD, Gool LV. A+: Adjusted anchored neighborhood regression for fast super-resolution. In: Proc. of the Asian Conf. on Computer Vision. Springer Int'l Publishing, 2014. 111–126.
- [11] Schulter S, Leistner C, Bischof H. Fast and accurate image upscaling with super-resolution forests. In: Proc. of the Computer Vision and Pattern Recognition. IEEE, 2015. 3791–3799.
- [12] Dong C, Chen CL, He K, Tang X. Image super-resolution using deep convolutional networks. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2016,38(2):295–307. [doi: 10.1109/TPAMI.2015.2439281]
- [13] Wang Z, Liu D, Yang J, Han W, Huang T. Deep networks for image super-resolution with sparse prior. In: Proc. of the IEEE Int'l Conf. on Computer Vision. IEEE, 2016. 370–378.
- [14] Kim J, Lee JK, Lee KM. Accurate image super-resolution using very deep convolutional networks. In: Proc. of the Computer Vision and Pattern Recognition. IEEE, 2015. 1646–1654.

- [15] Kim J, Lee JK, Lee KM. Deeply-recursive convolutional network for image super-resolution. In: Proc. of the Computer Vision and Pattern Recognition. IEEE, 2015. 1637–1645.
- [16] Tai Y, Yang J, Liu X. Image super-resolution via deep recursive residual network. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. IEEE Computer Society, 2017. 2790–2798.
- [17] Shi W, Caballero J, Huszar F, Totz J, Aitken AP, Bishop R, Rueckert D, Wang ZH. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proc. of the Computer Vision and Pattern Recognition. IEEE, 2016. 1874–1883.
- [18] Dong C, Chen CL, Tang X. Accelerating the super-resolution convolutional neural network. In: Proc. of the European Conf. on Computer Vision (ECCV). Springer-Verlag, 2016. 391–407. [doi: 10.1007/978-3-319-46475-6\_25]
- [19] Lai WS, Huang JB, Ahuja N, Yang MH. Deep Laplacian pyramid networks for fast and accurate super-resolution. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2017.
- [20] Mallat S. Wavelets for a vision. Proc. of the IEEE, 1996,84(4):604–614. [doi: 10.1109/5.488702]
- [21] Guo T, Mousavi HS, Vu TH, Monga V. Deep wavelet prediction for image super-resolution. In: Proc. of the Computer Vision and Pattern Recognition Workshops. IEEE, 2017. 1100–1109.
- [22] He K, Zhang XY, Ren SH, Sun J. Deep residual learning for image recognition. In: Proc. of the Computer Vision and Pattern Recognition. 2015. 770–778.
- [23] Sun X, Li XG, Li JF, Zhuo L. Review on deep learning based image super-resolution restoration algorithms. Acta Automatica Sinica, 2017,43(5):697–709 (in Chinese with English abstract).
- [24] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: Proc. of the Int'l Conf. on Learning Representations. 2015.
- [25] Johnson J, Alahi A, Li FF. Perceptual losses for real-time style transfer and super-resolution. In: Proc. of the European Conf. on Computer Vision. 2016. 694–711. [doi: 10.1007/978-3-319-46475-6\_43]
- [26] Ledig C, Theis L, Huszar F, Caballero J, Cunningham A, Acosta A, Aitken A, Tejani A, Totz J, Wang ZH, Shi WZ. Photo-realistic single image super-resolution using a generative adversarial network. In: Proc. of the Computer Vision and Pattern Recognition. 2016. 105–114.
- [27] Wu B, Duan H, Liu Z, Sun GY. SRPGAN: Perceptual generative adversarial network for single image super resolution. In: Proc. of the Computer Vision and Pattern Recognition. 2017.
- [28] El M, Wahed S. Image enhancement using second generation wavelet super resolution. Int'l Journal of Physicalences, 2007,2(6): 149–158.
- [29] Ji H. Robust wavelet-based super-resolution reconstruction: Theory and algorithm. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2009,31(4):649–660. [doi: 10.1109/TPAMI.2008.103]
- [30] Demirel H, Izadpanahi S, Anbarjafari G. Improved motion-based localized super resolution technique using discrete wavelet transform for low resolution video enhancement. In: Proc. of the IEEE European Conf. on Signal Processing. 2009. 1097–1101.
- [31] Robinson MD, Toth CA, Lo JY, Farsiu S. Efficient Fourier-wavelet super-resolution. IEEE Trans. on Image Processing, 2010, 19(10):2669–2681. [doi: 10.1109/TIP.2010.2050107]
- [32] Bevilacqua M, Roumy A, Guillemot C, Alberi-Morel ML. Low-complexity single image super-resolution based on nonnegative neighbor embedding. In: Proc. of the British Machine Vision Conf. 2012.
- [33] Roman Z, Michael E, Matan P. On single image scale-up using sparse-representations. In: Proc. of the Int'l Conf. on Curves and Surfaces. Springer-Verlag, 2012. 711–730.
- [34] Arbelaez P, Maire M, Fowlkes C, Malik J. Contour detection and hierarchical image segmentation. IEEE Trans. on Pattern Analysis & Machine Intelligence, 2011,33(5):898–916. [doi: 10.1109/TPAMI.2010.161]
- [35] Huang JB, Singh A, Ahuja N. Single image super-resolution from transformed self-exemplars. In: Proc. of the Computer Vision and Pattern Recognition. IEEE, 2015. 5197–5206. [doi: 10.1109/CVPR.2015.7299156]
- [36] Matsui Y, Ito K, Aramaki Y, *et al.* Sketch-based manga retrieval using Manga109 dataset. Multimedia Tools & Applications, 2017, 76(20):21811–21838. [doi: 10.1007/s11042-016-4020-z]

- [37] Lim B, Son SH, Kim HW, Nah SJ, Lee KM. Enhanced deep residual networks for single image super-resolution. In: Proc. of the Computer Vision and Pattern Recognition. 2017.

附中文参考文献:

- [23] 孙旭,李晓光,李嘉锋,卓力.基于深度学习的图像超分辨率复原研究进展.自动化学报,2017,43(5):697-709.



段立娟(1973—),女,山西晋中人,博士,教授,博士生导师,CCF 高级会员,主要研究领域为图像处理,机器学习.



乔元华(1969—),女,博士,教授,博士生导师,主要研究领域为机器学习,动力系统.



武春丽(1994—),女,硕士,主要研究领域为图像处理,深度学习.



张韵东(1968—),男,教授级高工,博士生导师,主要研究领域为人工智能,芯片设计,视频处理.



恩擎(1992—),男,硕士,主要研究领域为深度学习,图像处理,弱监督场景解析.



陈军成(1980—),男,博士,讲师,CCF 专业会员,主要研究领域为软件测试,大数据,机器学习.