

多尺度残差聚合特征网络图像超分辨重建

何立风^{1,2}, 苏亮亮^{1*}, 周广彬¹, 袁朴¹, 陆泊帆¹, 于佳佳¹¹陕西科技大学电子信息与人工智能学院, 陕西 西安 710021;²日本爱知县立大学信息科学学院, 爱知 长久手 480-1198

摘要 针对现有基于深度卷积神经网络模型的图像超分辨重建技术存在图像特征提取尺度单一和中间层次特征利用不充分等问题,提出了一种多尺度残差聚合特征网络模型。首先,该模型利用不同扩展系数的扩展卷积和残差连接设计了一种混合扩展卷积残差块(HERB),有效地提取到图像多个尺度的特征信息;其次,引入了一种特征聚合机制(AM),解决了网络中间层次特征利用不充分的问题。在常用的 5 种数据集上进行的实验结果表明,所提网络模型在主观视觉效果和客观评价指标上都比其他模型具有更好的性能。

关键词 图像处理; 超分辨重建; 多尺度特征信息; 扩展卷积; 残差连接; 聚合机制

中图分类号 TP391

文献标志码 A

doi: 10.3788/LOP202158.2410011

Image Super-Resolution Reconstruction Based on Multi-Scale Residual Aggregation Feature Network

He Lifeng^{1,2}, Su Liangliang^{1*}, Zhou Guangbin¹, Yuan Pu¹, Lu Bofan¹, Yu Jiajia¹¹ School of Electronic Information and Artificial Intelligence, Shaanxi University of Science & Technology, Xi'an, Shaanxi 710021, China;² School of Information Science and Technology, Aichi Prefectural University, Nagakute, Aichi 480-1198, Japan

Abstract Aiming at the problems of single image feature extraction scale and insufficient utilization of middle level features in the existing image super-resolution reconstruction technology based on depth convolution neural network model, a multi-scale residual aggregation feature network model for image super-resolution reconstruction is proposed. First, the proposed network model uses expanded convolutions with different expanded coefficients and residual connection to construct a hybrid expanded convolution residual block (HERB), which can effectively extract multi-scale feature information of an image. Second, a feature aggregation mechanism (AM) is used to solve the problem of insufficient utilization of features among middle levels of the network. Experiments results on five commonly used data sets show that the proposed network model has better performance than other models in subjective visual effect and objective evaluation index.

Key words image processing; super-resolution reconstruction; multi-scale feature information; extended convolution; residual connection; aggregation mechanism

OCIS codes 100.6640; 200.4260; 100.3010

1 引言

近年来,图像超分辨重建问题得到了国内外研究学者的广泛关注,目前研究最多的是单幅图像超

分辨重建(SISR)。SISR 的任务是从退化的低分辨率(LR)图像重建出对应的高分辨率(HR)图像,但这是一个病态的逆问题,因为一个 LR 图像与多个 HR 图像对应,不同的方法重建的 HR 图像也不同。

收稿日期: 2021-04-07; 修回日期: 2021-04-29; 录用日期: 2021-05-18

基金项目: 国家自然科学基金(61971272)

通信作者: *1211516382@qq.com

2410011-1

SISR 主要应用于卫星遥感图像^[1]和医学图像^[2]等领域。

随着深度学习的发展,越来越多基于深度卷积神经网络的图像超分辨方法被提出。2014 年,Dong 等^[3]将深度卷积网络引入到 SISR 领域,提出了一种包含三层卷积的深度神经网络,即 SRCNN。相比于传统的稀疏编码方法^[4],SRCNN 具有更为先进的性能,从而引发了深度学习在超分辨领域的热潮。在此基础上,Dong 等^[5]又提出了 FSRCNN 方法,Shi 等^[6]提出了 ESPCN 方法,两种方法分别使用反卷积层和亚像素卷积层,直接将不经过插值放大的 LR 图像输入网络中,加速网络的训练。已有研究表明,更深更宽的网络通常可以带来更好的性能,但是更深更宽的网络往往会带来网络训练困难的问题。为了解决这一问题,He 等^[7]提出了一种深度残差网络(ResNet)。ResNet 的提出不仅使得较深的网络得以训练,而且网络性能也得到了进一步的提升。因此,Kim 等^[8]在之前研究学者的基础上提出了 VDSR 方法,进一步将网络的深度提升到 20 层,同时引入全局残差学习来缓解训练的难度,使网络性能又得到了很大的提升。Kim 等^[9]又提出了 DRCN 方法,该方法利用递归思想精简了网络,并且达到了与 VDSR 相近的效果。Tai 等^[10]受到 VDSR 和 DRCN 的启发,在它们的基础上提出了 DRRN 方法,该方法增加了网络深度,采用局部残差和全局残差结合的方式使性能得到了进一步的提升。

尽管上述的 SISR 方法有了很大的改进,但仍然存在一定的局限性,它们的每一层网络都采用单一尺寸的卷积核,因此不能充分地提取图像的特征信息。针对该问题,文献^[11-12]利用不同尺寸大小的卷积核获得不同大小的感受野,以获取多尺度特征信息,并取得了一定的成效,但是这些方法无形中增加了网络中的参数量。此外,大部分方法都仅将

残差学习作为一种策略来缓解训练难度,并没有对中间层特征输出图进行充分利用,而这些特征输出图往往可以帮助重建出更多的高频细节信息。

针对以上问题,本文提出了一种多尺度残差聚合特征网络图像超分辨重建方法。首先,所提方法采用不同扩展系数的扩展卷积,在不增加网络参数的情况下,获得不同大小的感受野,从而提取图像的多尺度特征信息;其次,在引入短残差连接和长残差连接的同时,借鉴密集连接思想^[13-15],设计了一种聚合机制(AM)。该机制充分利用中间层的特征,减少了中间层特征在网络后期的丢失。实验结果表明,由所提模型重建的超分辨图像在主观和客观的性能上都优于现有方法。

2 所提多尺度残差聚合特征网络图像超分辨重建方法

2.1 网络结构概述

图 1 为所提多尺度残差聚合特征网络模型,该模型主要可以分为 3 个模块,分别为浅层特征提取模块、深层特征提取模块及重建模块。浅层特征提取模块由一个卷积层组成,直接提取不经插值放大的低分辨图像的浅层特征,并将其输入到深层特征提取模块当中,用以提取更深层次的特征信息。深层特征提取模块包含多个多尺度特征提取模块,如图 2 所示,多尺度特征提取模块由多个混合扩展卷积残差块(HERB)及末端的聚合机制组成,混合的扩展卷积残差块用以提取图像多个尺度的特征信息,聚合机制用以聚合多个混合扩展卷积残差块输出的特征。此外,短残差连接可以将更多的浅层信息传递到更深的网络。重建模块先通过上采样层对网络特征提取层提取到的特征信息进行通道像素信息重组,提升图像的分辨率,再通过一个 3×3 的卷积层重建出视觉效果良好的 RGB 三通道高分辨图像。

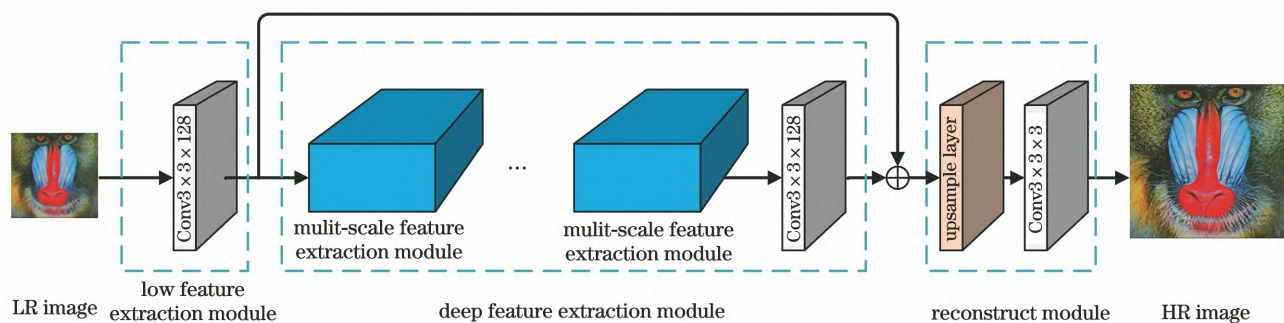


图 1 多尺度残差聚合特征网络结构

Fig. 1 Structure of multi-scale residual aggregation feature network

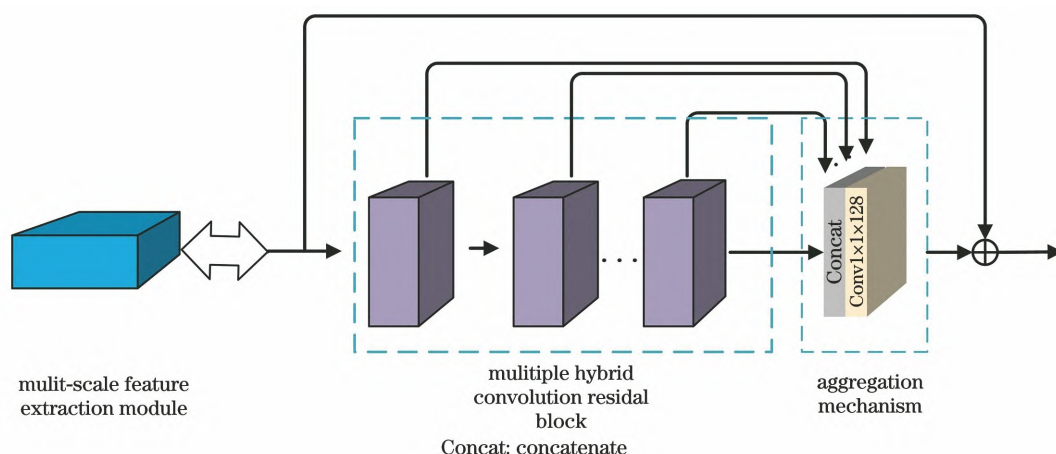


图 2 多尺度特征提取模块

Fig. 2 Multi-scale feature extraction module

2.2 浅层特征提取模块

主流的浅层特征提取方法都直接将不经过插值放大的低分辨率图像输入到浅层特征提取模块当中。直接将小尺寸的图像输入网络当中可以降低网络的计算复杂度,加速网络的训练。受此启发,所提方法也采用这种方式提取浅层的图像特征,所提方法的浅层特征提取模块使用一个 3×3 的卷积层提取图像浅层特征。记给定的 RGB 三通道的低分辨率图像为 I_{LR} ,提取到地浅层特征为 F_0 ,有

$$F_0 = H_{\text{Conv}}(I_{LR}), \quad (1)$$

式中: $H_{\text{Conv}}(\cdot)$ 表示卷积操作。提取到浅层特征 F_0 后,将其输入到深层特征提取模块当中,以提取更深层次的特征。同时,通过一个长残差连接,将浅层特征 F_0 直接输入到网络的重建部分。因为浅层提取层提取的信息包含很多图像的原始信息,这些原始信息可以辅助重建出被忽略的高频细节部分。

2.3 深层特征提取模块

图 3 为扩展系数分别为 1,2,3 的 3 种扩展卷积核。扩展系数为 1 的扩展卷积核相当于普通的卷积核,扩展系数为 2 和 3 的扩展卷积核实际上是在扩展系数为 1 的卷积核中间用相应比例系数的空洞填充

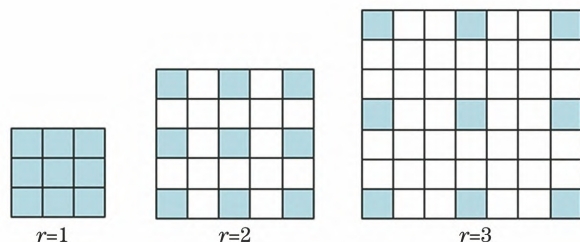


图 3 不同扩展系数的扩展卷积核

Fig. 3 Extended convolution kernel with different expansion coefficients

充得到的,空洞在实际的网络计算中填充为 0。从图中可以看出,不同扩展系数的扩展卷积核的参数数目没有发生改变,但是感受野大小随着扩展系数的增大呈指数增涨。所提网络采用这 3 种卷积核的组合,利用不同大小的感受野获得多尺度的特征信息,这是所提网络模型的核心部分。

所提深层特征提取模块包含多个多尺度特征提取模块,而多尺度特征提取模块由多个混合扩展卷积残差块和末端聚合机制构成。混合扩展卷积残差块是受文献[16]的启发而设计的,与图 4(a)所示的普通残差块(RB)仅从单尺度上获取图像的特征信息不同的是,混合扩展卷积残差块利用不同扩展系数的扩展卷积从 3 个尺度捕获图像的特征信息,使得提取的图像信息更为丰富,可为图像重建提供更多有用信息,如图 4(b)所示。

浅层特征提取模块得到的特征图经过 1×1 的卷积核降维后,由 128 维降为 64 维,目的是减少网络中的参数,降低网络的规模,使得网络计算的复杂度不会太高。为了多尺度地提取图像的特征信息,文献[17]采用了 3×3 和 5×5 的 2 种不同尺寸大小的卷积核;所提方法采用了上述扩展系数为 1,2,3,参数数目都为 3×3 ,数量为 64 的 3 种卷积核。在相同数量卷积核的情况下,所提方法参数减少了 20.6%,且具有更大的感受野,能够捕获更多尺度的特征信息。此外,所提方法也考虑到直接将经过 3 个尺度分支得到的特征输出图进行合并时,尽管网络可以获得一个很大的有效感受野,但同时也引入了不必要的网格伪影。图 5 为单个像素与 3×3 扩展卷积核(扩展系数 $r=2$)卷积操作之后的网格伪影。所提方法在得到 3 个尺度的特征输出之后,没有直接对其进行融合,而是对其进行逐像素相加

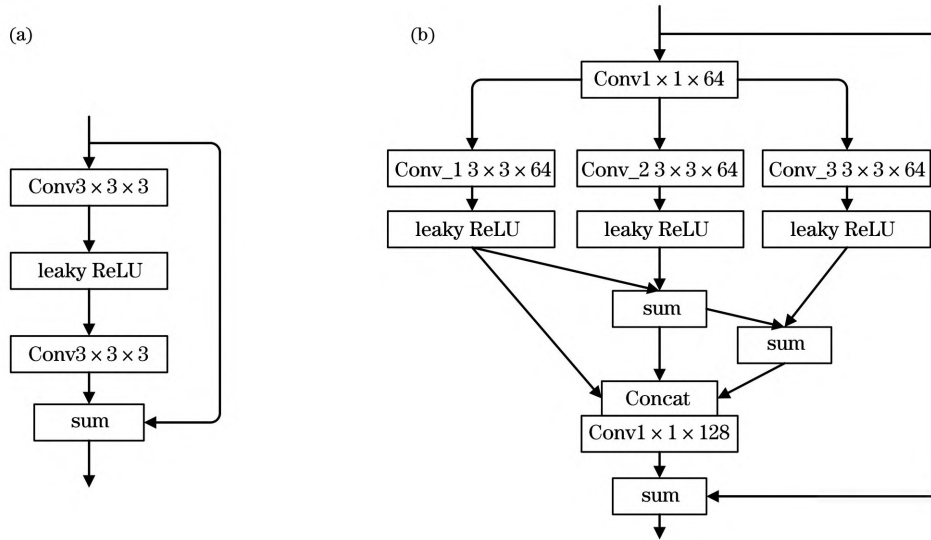


图 4 不同模块结构比较。(a)普通残差块;(b)混合扩展卷积残差块

Fig. 4 Comparison of different module structures. (a) Ordinary residual block; (b) hybrid extended convolution residual block

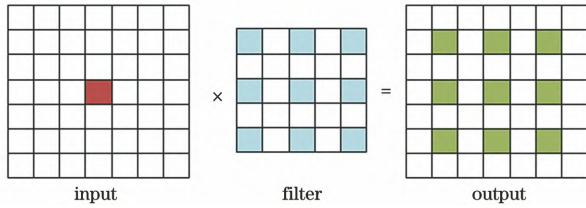


图 5 单像素与 3×3 扩展卷积核(扩展系数 $r=2$) 卷积的网格伪影

Fig. 5 Gridding artifact with a single pixel convolved with a 3×3 extended convolutional kernel (expansion coefficient $r=2$)

操作,再将像素相加之后的特征图重新融合成新的

128 维特征图。为验证该操作的有效性,对中间层的某一个特征输出图进行了可视化操作。可视化的特征输出图如图 6 所示,图 6(a)为 RGB 图,图 6(b)为未采用分支输出特征通道的逐像素相加操作而直接进行融合得到的特征图,图 6(c)为在分支输出特征通道融合之前进行逐像素相加操作得到的特征图。从图中可以看出,逐像素相加操作能够减少网格伪影的产生,能够很好地避免扩展卷积的空洞所带来的不良视觉效果。

最后,为使网络训练更容易、性能更好,所提方法对每个混合扩展卷积块引入的局部残差进行连接。

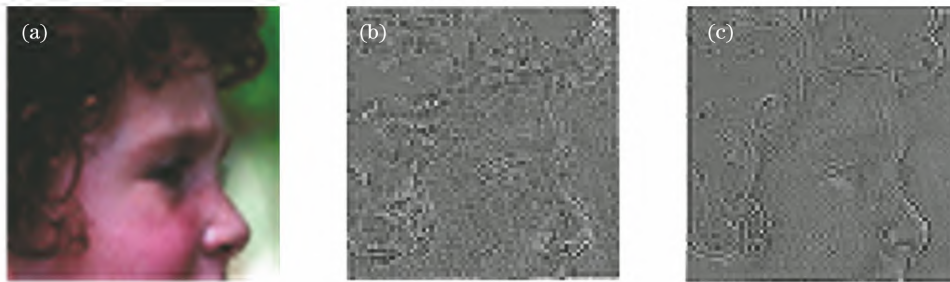


图 6 可视化的特征输出图。(a)RGB 图;(b)无逐像素相加操作;(c)具有逐像素相加操作

Fig. 6 Diagram of visual feature output. (a) RGB image; (b) without per-pixel addition operation; (c) with per-pixel addition operation

$$\mathbf{B}_n^i = \mathbf{B}_n^{i-1} + H_{\text{Concat}}(\mathbf{F}_n^{i,1}, \mathbf{F}_n^{i,1} + \mathbf{F}_n^{i,2}, \mathbf{F}_n^{i,1} + \mathbf{F}_n^{i,2} + \mathbf{F}_n^{i,3}), \quad (2)$$

式中: \mathbf{B}_n^i 表示第 n 个多尺度特征提取模块第 i 个混合扩展卷积残差块的特征输出; $\mathbf{F}_n^{i,1}$ 、 $\mathbf{F}_n^{i,2}$ 和 $\mathbf{F}_n^{i,3}$ 表示 3 种不同尺度分支的特征输出; $H_{\text{Concat}}(\cdot)$ 表示

3 个尺度特征的融合操作。其中

$$\mathbf{F}_n^{i-1} = H_{\text{Conv}}(\mathbf{B}_n^{i-1}), \quad (3)$$

$$\mathbf{F}_n^{i,1} = \delta[H_{\text{Conv}_1}(\mathbf{F}_n^{i-1})], \quad (4)$$

$$\mathbf{F}_n^{i,2} = \delta [H_{\text{Conv}_2}(\mathbf{F}_n^{i-1})], \quad (5)$$

$$\mathbf{F}_n^{i,3} = \delta [H_{\text{Conv}_3}(\mathbf{F}_n^{i-1})], \quad (6)$$

式中: δ 表示 leaky ReLU 非线性激活函数; \mathbf{F}_n^{i-1} 表示经过 1×1 卷积核降维得出的输出; $H_{\text{Conv}_1}(\cdot)$ 、 $H_{\text{Conv}_2}(\cdot)$ 和 $H_{\text{Conv}_3}(\cdot)$ 分别表示 3 种不同尺度下的卷积操作。

多尺度特征提取模块的末端含有聚合机制,目的是为了充分利用中间层次的特征。从图 2 中可以看出,聚合机制添加在每一个多尺度特征提取模块的末尾,通过引入了一种弱密集连接^[13],将多尺度特征提取模块中的多个混合扩展卷积残差块的输出特征聚合在一起,得到总数较大的聚合特征图。最后使用 1×1 的卷积核对这些特征图进行融合,获得特征信息非常丰富的输出,使得重建出的高分辨率图像具有更多的高频细节信息。

$$\mathbf{F}_n = H_{\text{Concat}}(\mathbf{B}_n^1, \mathbf{B}_n^2, \dots, \mathbf{B}_n^i), \quad (7)$$

式中: \mathbf{F}_n 表示第 n 个多尺度特征提取模块的特征输出,它融合了多个混合扩展卷积残差块的输出特征。

2.4 重建模块

重建模块由一个上采样层和普通卷积层组成,其中上采样层通过亚像素卷积层^[6]来提升图像的分辨率,普通卷积层用于重构 RGB 三通道高分辨率图像。记重建模块获得的深层特征提取模块和浅层特征提取模块叠加的特征输出图的大小为 $H \times W \times C$, 其中 H 、 W 和 C 分别表示特征输出图的长、宽和通道数。首先,将重建模块获得的特征输出图输入到大小为 $3 \times 3 \times r'^2 C$ 的亚像素卷积层,其中 r' 为放大因子,并进行通道像素信息的周期重组得到大小为 $r' H \times r' W \times C$ 的特征输出图,实现特征图分辨率的提升;然后,再经过大小为 $3 \times 3 \times 3$ 的卷积层实现 RGB 三通道高分辨率图像的重构。

$$\mathbf{F}_{\text{out}} = H_{\text{Conv}}\{\text{PS}[H_{\text{sub_Conv}}(\mathbf{F}_0 + \mathbf{F}_n)]\}, \quad (8)$$

式中: $H_{\text{sub_Conv}}(\cdot)$ 表示亚像素卷积操作; $\text{PS}(\cdot)$ 表示通道信息周期重组操作,对特征张量进行重排列; \mathbf{F}_{out} 表示网络最终的输出,即重构出的高分辨率图像。

2.5 损失函数

L1 损失函数相较于 L2 损失函数而言更接近人的视觉感观,所提方法和很多其他方法^[18-19]一样采用 L1 损失函数来获得重建出的超分辨率图像 \mathbf{Y}_{SR} 与 ground-truth 图像 \mathbf{Y}_{HR} 之间的差异,通过最小化损失函数来优化网络,使得网络重建出的超分辨率图像更接近 ground-truth 图像。假设给定训练集 $\{\mathbf{X}_i, \mathbf{Y}_i\}^N$, 其中 N 是训练集的一个 mini-batch 的

数量, \mathbf{X}_i 是低分辨率图像, \mathbf{Y}_i 是 ground-truth 图像,损失函数 L_{loss} 的表达式为

$$L_{\text{loss}} = \frac{1}{N} \|\mathbf{F}(\mathbf{X}_i) - \mathbf{Y}_i\|_1, \quad (9)$$

式中: $\mathbf{F}(\mathbf{X}_i)$ 表示网络的输出,即重建的超分辨率图像。

3 实验与分析

实验的设备条件为: Intel(R) Core(TM) i7-8700K CPU, 主频为 3.70 GHz, 内存为 32 GB 的 PC 及 NVIDIA GeForce RTX2070 的 GPU, 显存为 8 GB。实验使用 PyCharm 编译平台, 配置 tensorflow 深度学习环境, 用 python 语言实现模型的构建。

所提方法采用图像处理领域公认具有权威性的指标: 峰值信噪比 (PSNR) 和结构相似性 (SSIM) 作为客观评价指标, 实验中的这两个指标的值都是在图像 Y 通道上计算得到的。

3.1 数据集与训练细节

所提方法使用 DIV2K 数据集^[20] 的 800 张高分辨率训练图像作为训练集。在实验中, 由于 DIV2K 数据集的图像过大, 通常的办法是将一张原始大图按照一定的步幅裁剪成多张小图像, 并将裁剪出的小图像作为新的高分辨率图像。按照这样的方法, 所提方法共裁剪出 32592 张可作为高分辨率图像的小图像。同时, 为了避免网络出现过拟合现象, 通过使用随机水平翻转和 90 度旋转的方式总共获得了 97776 张高分辨率图像, 并利用双三次插值方法对高分辨率图像进行降采样, 得到对应的低分辨率图像, 总共获得 97776 对低分辨率图像与高分辨率图像用于网络的训练。实验测试集为 Set5 (5 张)^[21]、Set14 (14 张)^[22]、BSD100 (100 张)^[23]、Manga109 (109 张)^[24] 和 Urban100 (100 张)^[25] 这 5 种常用的测试集。

训练时, 每次随机抽取 16 张尺寸为 48×48 的低分辨率图像的子图及所对应的高分辨率图像的子图进行训练。所提方法采用 ADAM 优化器, 优化器参数为默认参数, 即控制权重分配的指数衰减率 $\beta_1 = 0.9$, 控制梯度平方的指数衰减率 $\beta_2 = 0.999$, 初始学习率为 1×10^{-4} , 网络每反向传播 20 万次, 学习率减小一半。所提方法在 $r' = 2, r' = 3, r' = 4$ 三个放大因子上评估模型, 所有实验的结果均是在网络反向传播 30 万次时得到的。

3.2 网络模型分析

通过实验分析了不同模块和设计在所提模型中的作用和贡献, 主要包括多尺度特征提取模块的数

量、混合扩展卷积残差块的数量、聚合机制的引入及与普通残差块的对比。

为了检验多尺度特征提取模块的数量和混合扩展卷积残差块的数量对网络深度的影响,从这两种模块开始分别做实验。假设多尺度特征提取模块的数量为 N_1 ,混合扩展卷积残差块的数量为 N_2 ,一开始固定多尺度特征提取模块的数量 $N_1=2$,混合扩展卷积残差块的数量 $N_2=2,3,\dots,11$ 。因为 N_2 为 1 的时候,网络太浅,所提取的特征信息不够,所以直接从 2 开始实验,并与 Set14($r'=4$)上的测试结果进行比较。表 1 中的结果为混合扩展卷积残差块的数量对网络训练过程中训练一个 mini-batch

表 1 混合扩展卷积残差块的数量、平均时间和 PSNR 之间的关系

Table 1 Relationship between number of hybrid extended convolution residual blocks, average time, and PSNR

N_2	2	3	4	5	6	7	8	9	10	11
PSNR /dB	28.335	28.398	28.428	28.444	28.463	28.490	28.522	28.532	28.507	28.506
Average time /s	0.107	0.138	0.176	0.195	0.216	0.235	0.254	0.277	0.299	0.320

在确定了混合扩展卷积残差块的数量为 8 之后,开始检验多尺度特征提取模块的数量对网络训练过程中训练一个 mini-batch 的平均时间和在 Set14 测试集上测得的平均 PSNR 值的影响,同样与 Set14($r'=4$)的测试结果进行比较。多尺度特征提取模块的数量 N_1 分别为 1,2,3,4。表 2 中的结果为多尺度特征提取模块的数量对平均时间和 PSNR 值的影响。从表中可以看出:随着模块数量的增加,测试集数据的平均 PSNR 的值基本上是有

表 2 多尺度特征提取模块的数量、平均时间和 PSNR 之间的关系

Table 2 Relationship between number of multi-scale feature extraction modules, average time, and PSNR

N_1	1	2	3	4
PSNR /dB	28.431	28.522	28.527	28.533
Average time /s	0.171	0.254	0.341	0.438

表 3 三种模型在 5 种数据集上的平均 PSNR/SSIM

Table 3 Average PSNR/SSIM of three models on 5 data sets

PSNR unit: dB

Data set	M_HERB	M_RB+AM	M_HERB+AM
Set5	32.11/0.8938	32.01/0.8930	32.03/0.8933
Set14	28.49/0.7797	28.41/0.7756	28.52/0.7805
BSD100	27.53/0.7351	27.45/0.7312	27.57/0.7361
Manga109	30.17/0.9055	30.10/0.9011	30.30/0.9072
UrBan100	25.85/0.7792	25.77/0.7789	25.99/0.7846

的平均时间和在 Set14 测试集上测得的平均 PSNR 值的影响。从表中可以看出:随着混合扩展卷积残差块数量的增加,PSNR 的值都是有提升的;但是从 $N_2=8$ 之后,PSNR 值增加的幅度开始变小;在 $N_2=10$ 时,PSNR 的值不但没有增加,反而减小。原因是随着混合扩展卷积残差块数量的增加,网络虽然变得更深,但有用的信息还是会丢失,造成网络训练困难,从而得出质量不佳的结果。虽然在混合扩展卷积残差块的数量为 9 的时,PSNR 值也是有增加的,但是增加的幅度很小。为了权衡时间成本与 PSNR,所提方法选择混合扩展卷积残差块的数量为 8。

提升的;但是从数量为 3 开始,提升的幅度特别小,只提升了不到 0.01 dB。然而随着模块数量的增加,网络的参数量在成倍增加,导致几乎需要花费成倍的时间训练网络。同样为了综合考虑 PSNR 的值和时间成本,所提方法选择多尺度特征提取模块的数量为 2。

为了验证聚合机制及混合扩展卷积残差块的有效性,使用普通残差块替代混合扩展卷积残差块,在其他条件均一致的情况下进行了实验,同时也对未去除聚合机制和去除聚合机制的网络进行了消融实验。总的来说,对 3 种网络模型进行了对比实验,3 种模型分别为所提网络模型去除聚合机制之后的模型(M_HERB)、包含聚合机制的普通残差块模型(M_RB+AM)、包含聚合机制的所提网络模型(M_HERB+AM)。表 3 和图 7 分别为 3 种网络模型在 5 种数据集($r'=4$)上的客观指标结果和主观视觉结果,表中最佳的结果用粗体标记。从表 3 可以看

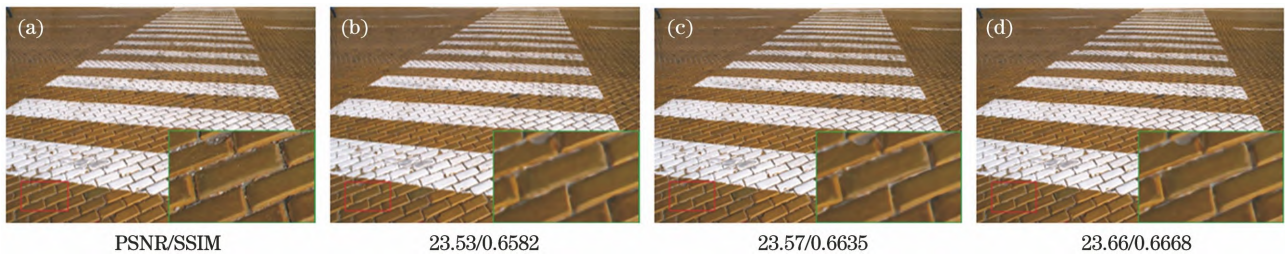


图 7 三种模型在 Urban100 的图像“img091”上的重建结果。(a)原图;(b)M_HERB;(c)M_RB+AM;(d)M_HERB+AM

Fig. 7 Reconstruction results of the three models in the Urban100 image “img091”. (a) Original drawing; (b) M_HERB; (c) M_RB+AM; (d) M_HERB+AM

出:M_HERB+AM 在大多数数据集上都是表现最好的,尤其是在 Manga109 和 UrBan100 数据集上发挥了它的优势,PSNR 值相比 M_HERB 分别增加了 0.13 dB 和 0.14 dB,相比 M_RB+AM 分别增加了 0.20 dB 和 0.22 dB,说明了所提多尺度模型相比单尺度模型能够提取到更丰富的特征信息,引入聚合机制对于图像重建的质量也有一定的提升。此外,从图 7 中可以看出,M_HERB+AM 和 M_RB+AM 重建出的地砖轮廓更锐利、更清晰。实验结果表明,在网络中使用聚合机制确实可以保留更多的中间层次特征,减少中间层特征在网络传播过程中的丢失,从而可以有效地利用这些丰富的特征信息进行图像重建。

3.3 实验结果与对比

为了说明所提方法的有效性,对所提方法与一

些现有方法在 5 种常用测试集上进行了比较,包括 SRCNN^[4]、VDSR^[8]、DRRN^[10]、SRMDNF^[26] 和 IMRSR^[12]。客观的评价指标为 PSNR 和 SSIM 这两种具有权威性的指标,并且分别在 $r'=2$, $r'=3$, $r'=4$ 上进行比较。因为 SRCNN, VDSR, DRRN, SRMDNF 这些方法没有在 Manga109 数据集上的数据,所以本实验组利用这些方法已提供的公开训练模型在 Manga109 数据集上进行了评估,得到了 $r'=2$, $r'=3$, $r'=4$ 上的平均 PSNR 和 SSIM 值。对于 IMRSR 方法,它没有提供公开的训练模型,所以本实验组只是利用它原文中的指标来进行比较,没有将它的网络模型在 Manga109 数据集上进行评估。这些方法在 5 种测试集上的客观评价结果如表 4 所示,表中最佳的结果用粗体标记,其次的结果用下划线标记。

表 4 不同方法在不同测试集的平均 PSNR/SSIM

Table 4 Average PSNR/SSIM of different methods on different test sets

PSNR unit: dB

Method	r'	Set5	Set14	BSD100	Manga109	UrBan100
Bicubic	2	33.66/0.9299	30.24/0.8688	29.56/0.8431	30.80/0.9339	26.88/0.8403
SRCNN ^[4]	2	36.66/0.9542	32.45/0.9067	31.36/0.8879	35.60/0.9663	29.50/0.8946
VDSR ^[8]	2	37.53/0.9590	33.05/0.9130	31.90/0.8960	37.22/0.9750	30.77/0.9140
DRRN ^[10]	2	37.74/0.9591	33.23/0.9136	32.05/0.8973	37.60/0.9736	31.23/0.9188
SRMDNF ^[26]	2	37.79/0.9601	33.32/0.9159	32.05/0.8985	38.07/0.9761	31.33/0.9204
IMRSR ^[12]	2	37.78/0.9643	33.26/0.8488	32.00/0.9073		31.00/0.9235
Proposed method	2	37.89/0.9603	33.41/0.9159	32.07/0.8986	38.24/0.9763	31.62/0.9237
Bicubic	3	30.39/0.8682	27.55/0.7742	27.21/0.7385	26.95/0.8556	24.46/0.7349
SRCNN ^[4]	3	32.75/0.9090	29.30/0.8215	28.41/0.7863	30.48/0.9117	26.24/0.7989
VDSR ^[8]	3	33.67/0.9210	29.78/0.8320	28.83/0.7990	32.01/0.9340	27.14/0.8290
DRRN ^[10]	3	34.03/0.9244	29.96/0.8349	28.95/0.8004	32.42/0.9359	27.53/0.8378
SRMDNF ^[26]	3	34.12/0.9254	30.04/0.8382	28.97/0.8025	33.00/0.9403	27.57/0.8398
IMRSR ^[12]	3	33.91/0.9312	29.88/0.8488	28.80/0.8166		27.00/0.8403
Proposed method	3	34.18/0.9255	30.16/0.8389	28.99/0.8033	33.01/0.9413	27.77/0.8450

(续表 4)

Method	r'	Set5	Set14	BSD100	Manga109	UrBan100
Bicubic	4	28.42/0.8104	26.00/0.7027	25.96/0.6675	24.89/0.7866	23.14/0.6577
SRCNN ^[4]	4	30.48/0.8628	27.50/0.7513	26.90/0.7101	27.58/0.8555	24.52/0.7221
VDSR ^[8]	4	31.35/0.8830	28.02/0.7680	27.29/0.7726	28.83/0.8870	25.18/0.7540
DRRN ^[10]	4	31.68/0.8888	28.21/0.7721	27.38/0.7284	29.18/0.8914	25.44/0.7638
SRMDNF ^[26]	4	31.96/0.8925	28.35/0.7787	27.49/0.7337	30.09/0.9024	25.68/0.7731
IMRSR ^[12]	4	31.59/0.8957	28.19/0.7892	27.30/0.7469		25.15/0.7714
Proposed method	4	32.03/0.8933	28.52/0.7805	27.57/0.7361	30.30/0.9072	25.99/0.7846

从表 4 中可以看出:所提方法无论对平均 PSNR 值还是平均 SSIM 值,在这 5 种测试集上相比其他 5 种现有方法都表现出最佳或者其次的结果。对于平均 PSNR 值,所提方法的数值都是最高的,比 SRMDNF 方法基本上高出 0.1~0.3 dB,比同样是基于多尺度的 IMRSR 方法,最高高出近 0.8 dB,说明了所提混合扩展卷积块及聚合机制的有效性。对于平均 SSIM 值,总的来说,所提方法基本上和 SRMDNF 和 IMRSR 方法不相上下,虽然在有些测试集上有差距,但差距甚微。综合

来看,所提方法性能相较现有的方法有明显的提升。

为了说明所提方法在主观视觉上的效果,在 $r'=4$ 的情况下和上述现有方法进行视觉感观比较。之所以采用 $r'=4$,是因为更大的放大因子意味着网络将要重建出更多的图像信息,这就更能显示出一个方法的强大性能。鉴于篇幅有限,在 Set14、BSD100、Manga109 和 UrBan100 中各拿出一幅具有代表性的图像进行局部放大并进行比较。4 幅重建出的图像如图 8 所示,其中第一列为

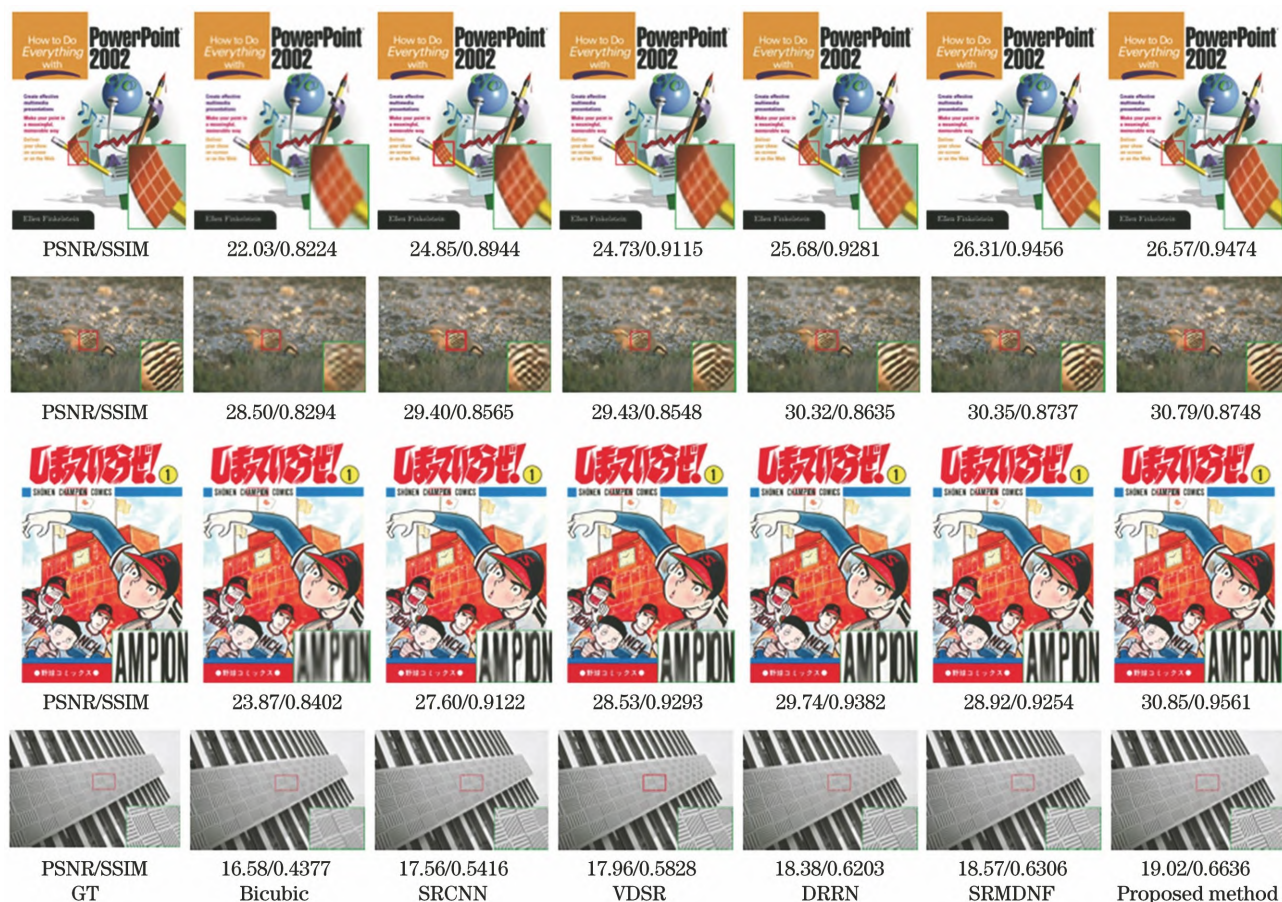


图 8 各类方法下图像重建效果的比较

Fig. 8 Comparison of image reconstruction effects under various methods

ground-truth, 其余 6 列从左至右依次是 Bicubic、SRCNN、VDSR、DRRN、SRMDNF 及所提方法。

从 4 幅重建后的图像局部放大部位可以看出: Bicubic 和 SRCNN 重建出的图像都很模糊, 没有将动物身上及建筑物上的纹理轮廓重建出来; VDSR 方法能够重建出部分的纹理轮廓; DRRN 和 SRMDNF 方法的结果相较前面的方法清晰了不少, 基本上可以重建出部分图像的纹理轮廓; 而相比 DRRN 和 SRMDNF 方法, 所提方法能够重建出更多的纹理轮廓信息, 重建的图像也更加清晰、更加锐利, 有更好的视觉感观。如图 8 的第 2 幅图像, 所提方法将动物身上的纹理部分基本重建出来了, 并且有纯净的视觉感受。再比如最后 1 幅建筑物图像的轮廓细节, 对于有些部分, 所提方法未将其完全重建出来, 但 DRRN 和 SRMDNF 这些先进的方法完全不能重建出来。这样一比较, 就更显示出了所提方法的优越性。

所提方法在主观视觉上的优势, 说明多尺度残差聚合特征网络具有强大的重建性能, 相比现有方法, 能够提取到更多的有用特征信息, 帮助重建出视觉效果更佳的图像。

4 结 论

提出了一种多尺度残差聚合特征网络模型的图像超分辨重建方法, 使用一种混合扩展卷积残差块提取了图像的多尺度特征信息, 与现有的基于多尺度模型的方法相比, 能够使用更少的参数, 获得更大的感受野, 捕获更多的特征信息。同时, 与现有大多数方法不同的地方在于所提方法不仅引入残差策略, 而且还引入一种特征聚合机制, 既减少了浅层特征信息在向更深层次传播过程中的丢失, 又充分利用了各中间层次的特征, 为重建高分辨率图像提供丰富的层次特征信息。实验结果表明: 所提方法在客观指标上基本上都优于现有方法; 在主观视觉上, 也表现出了非常好的重建效果, 重建出的图像大多数的纹理轮廓都非常清晰, 并且更加真实。不过, 所提方法重建出的图像中也还存在部分的纹理轮廓重建不足的情况, 下一步将针对这一情况着重研究, 尽可能重建出与高分辨率图像更接近的图像。

参 考 文 献

- [1] Ma C, Rao Y M, Cheng Y A, et al. Structure-preserving super resolution with gradient guidance [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 7766-7775.
- [2] Hu F, Lin Y, Hou M D, et al. Super-resolution reconstruction of cytoskeleton image based on deep learning[J]. Acta Optica Sinica, 2020, 40(24): 2410001.
胡芬, 林洋, 侯梦迪, 等. 基于深度学习的细胞骨架图像超分辨重建[J]. 光学学报, 2020, 40(24): 2410001.
- [3] Dong C, Loy C C, He K M, et al. Learning a deep convolutional network for image super-resolution [C]//Fleet D, Pajdla T, Schiele B, et al. Computer vision-ECCV 2014. Lecture notes in computer science. Cham: Springer, 2014, 8692: 184-199.
- [4] Yang J C, Wright J, Huang T S, et al. Image super-resolution via sparse representation[J]. IEEE Transactions on Image Processing, 2010, 19(11): 2861-2873.
- [5] Dong C, Loy C C, Tang X O. Accelerating the super-resolution convolutional neural network [M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9906: 391-407.
- [6] Shi W Z, Caballero J, Huszar F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 1874-1883.
- [7] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [8] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 1646-1654.
- [9] Kim J, Lee J K, Lee K M. Deeply-recursive convolutional network for image super-resolution [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 1637-1645.
- [10] Tai Y, Yang J, Liu X M. Image super-resolution via deep recursive residual network[C]//2017 IEEE Conference on Computer Vision and Pattern

- Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 2790-2798.
- [11] Li J C, Fang F M, Mei K F, et al. Multi-scale residual network for image super-resolution [M] // Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11212: 527-542.
- [12] Chen X Y, Zhang W J, Sun W Z, et al. Super-resolution reconstruction of images based on multi-scale and multi-residual network[J]. Laser & Optoelectronics Progress, 2020, 57(18): 181009.
陈星宇, 张伟劲, 孙伟智, 等. 基于多尺度与多重残差网络的图像超分辨率重建[J]. 激光与光电子学进展, 2020, 57(18): 181009.
- [13] Tong T, Li G, Liu X J, et al. Image super-resolution using dense skip connections[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 4809-4817.
- [14] Xi Z H, Yuan K P. Super-resolution image reconstruction based on residual channel attention and multilevel feature fusion[J]. Laser & Optoelectronics Progress, 2020, 57(4): 041504.
席志红, 袁昆鹏. 基于残差通道注意力和多级特征融合的图像超分辨率重建[J]. 激光与光电子学进展, 2020, 57(4): 041504.
- [15] Li B, Ma L. Super-resolution reconstruction of densely connected generative adversarial network images[J]. Laser & Optoelectronics Progress, 2020, 57(22): 221011.
李斌, 马璐. 密集连接的生成对抗网络图像超分辨率重建[J]. 激光与光电子学进展, 2020, 57(22): 221011.
- [16] Mehta S, Rastegari M, Caspi A, et al. ESPNet: efficient spatial pyramid of dilated convolutions for semantic segmentation[M] // Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11214: 561-580.
- [17] Cao F L, Liu H. Single image super-resolution via multi-scale residual channel attention network [J]. Neurocomputing, 2019, 358: 424-436.
- [18] Hui Z, Wang X M, Gao X B. Fast and accurate single image super-resolution via information distillation network[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 723-731.
- [19] Li Z, Yang J L, Liu Z, et al. Feedback network for image super-resolution[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 3862-3871.
- [20] Timofte R, Agustsson E, van Gool L, et al. Ntire 2017 challenge on single image super-resolution: Methods and results [C] // Proceedings of the IEEE conference on computer vision and pattern recognition workshops, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 114-125.
- [21] Bevilacqua M, Roumy A, Guillemot C, et al. Low-complexity single-image super-resolution based on nonnegative neighbor embedding [C] // Proceedings of the British Machine Vision Conference 2012, September 3-7, 2012, Surrey, UK. Wales: British Machine Vision Association, 2012: 135.1-135.10.
- [22] Zeyde R, Elad M, Protter M. On single image scale-up using sparse-representations[M] // Boissonnat J D, Chenin P, Cohen A, et al. Curves and surfaces. Lecture notes in computer science. Heidelberg: Springer, 2012, 6920: 711-730.
- [23] Arbeláez P, Maire M, Fowlkes C, et al. Contour detection and hierarchical image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(5): 898-916.
- [24] Matsui Y, Ito K, Aramaki Y, et al. Sketch-based manga retrieval using manga109 dataset[J]. Multimedia Tools and Applications, 2017, 76(20): 21811-21838.
- [25] Huang J B, Singh A, Ahuja N. Single image super-resolution from transformed self-exemplars[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 5197-5206.
- [26] Zhang K, Zuo W M, Zhang L. Learning a single convolutional super-resolution network for multiple degradations [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 3262-3271.