

信号处理  
*Journal of Signal Processing*  
ISSN 1003-0530, CN 11-2406/TN

## 《信号处理》网络首发论文

题目: 基于渐进式特征增强网络的超分辨率重建算法  
作者: 杨勇, 吴峥, 张东阳, 刘家祥  
网络首发日期: 2020-09-16  
引用格式: 杨勇, 吴峥, 张东阳, 刘家祥. 基于渐进式特征增强网络的超分辨率重建算法. 信号处理. <https://kns.cnki.net/kcms/detail/11.2406.tn.20200915.1131.003.html>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式 (包括网络呈现版式) 排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊 (光盘版)》电子杂志社有限公司签约, 在《中国学术期刊 (网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊 (网络版)》是国家新闻出版广电总局批准的网络连续型出版物 (ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

# 基于渐进式特征增强网络的超分辨率重建算法

杨勇 吴峥 张东阳 刘家祥

(江西财经大学信息管理学院, 江西南昌 330032)

**摘要:**为了在图像重建质量和网络参数之间取得较好的平衡, 本文提出一种基于渐进式特征增强网络的超分辨率 (Super-Resolution, SR) 重建算法。该方法主要包含两个模块: 浅层信息增强模块和深层信息增强模块。在浅层信息增强模块中, 首先利用单层卷积层提取低分辨率 (Low-Resolution, LR) 图像的浅层信息, 再通过我们设计的多尺度注意力块来实现特征的提取和增强。深层信息增强模块先利用残差学习块学习图像的深层信息, 然后将得到的深层信息通过设计的多尺度注意力块来获得增强后的深层多尺度信息。最后我们利用跳转连接的方式将首层得到的浅层信息和深层多尺度信息进行像素级相加得到融合特征图, 再对其进行上采样操作, 得到最终的高分辨率 (High-Resolution, HR) 图像。实验结果表明, 相比于一些主流的深度学习超分辨率方法, 本文方法重建得到的图像无论是主观效果还是客观指标, 都取得了更好的效果。

**关键词:** 超分辨率重建; 浅层信息; 深层信息; 多尺度信息; 跳转连接

中图分类号: TP391.41

文献标识码: A

## Progressive Feature Enhancement Network for Image Super-Resolution Reconstruction

Yang Yong Wu Zheng Zhang Dongyang Liu Jiaxiang

(School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, Jiangxi 330032, China)

**Abstract:** In order to achieve a better balance between image restoration quality and network parameters, this paper proposes a super-resolution (SR) method based on progressive feature enhancement network. The method mainly consists of two modules: shallow information enhancement module and deep information enhancement module. In the shallow information enhancement module, firstly, the shallow information of low-resolution (LR) image is extracted by single convolution layer, and then we design the multi-scale attention block to achieve feature extraction and enhancement. The deep information enhancement module first uses the residual learning block to learn the deep information of the image and then employs the designed multi-scale attention block to obtain the enhanced deep multi-scale information. Finally, we use a skip-connection to fuse the shallow information obtained at the first layer and the deep multi-scale information at the pixel level and then up-sample the fusion feature maps to obtain the final high-resolution (SR) image. The experimental results show that, compared with some mainstream deep learning SR methods, the image reconstructed by the proposed method has achieved better results in both subjective and objective indicators.

**Key words:** super-resolution reconstruction; shallow information; deep information; multi-scale information; skip-connection

# 1 引言

图像超分辨率重建是计算机视觉领域的重要任务，目的在于将低分辨率（Low-Resolution, LR）图像重建得到高分辨率（High-Resolution, HR）图像。由于一幅 LR 图像往往有多种形态的 HR 图像映射，因此，该重建过程是一种不适定性问题。为了获得更加符合视觉任务的 HR 图像，许多超分辨重建方法已被提出来。这些方法主要分为三类，分别为：基于插值的方法<sup>[1]</sup>、基于模型的方法<sup>[2]</sup>和基于学习的方法<sup>[3]</sup>。

以双线性插值和双三次插值为代表的早期的插值方法简单方便，但是其重建得到的图像质量已经无法满足日益发展的应用需求。为了适应更加灵活、更高质量的重建需求，基于模型的方法应运而生。具有代表性的方法包括全变分模型<sup>[4]</sup>、非局部先验模型<sup>[5]</sup>和稀疏表示模型<sup>[6]</sup>。相较于插值的方法，基于模型的方法更加灵活且能够获得更加高质量的图像。但这些方法依然存在较大的缺陷：（1）这些算法在优化和重建过程中需要耗费大量的时间，不利于其在应用领域的推广；（2）过度依赖于先验信息，在没有先验条件的情况下，图像的重建质量急剧降低。

卷积神经网络强大的非线性表达能力在计算机领域取得了前所未有的成功，并且其端到端的训练模式与图像超分辨率重建思想完美契合。Dong 等人<sup>[7]</sup>第一次将深度学习引入到超分辨领域中，他们设计了一个简单的包含三个卷积层的超分辨率重建方法，命名为 SRCNN，该方法在亮度通道上对图像进行修复，相比于之前的基于插值和基于模型的方法，SRCNN 无论在重建质量还是在性能上都获得了巨大的提升。为了进一步提升网络的性能，Kim 等人将方法 VDSR<sup>[8]</sup>和 DRCN<sup>[9]</sup>的网络层数增加到了 20 层，在重建质量上，相比于 SRCNN 取得了显著的提升。值得一提的是，He 等人<sup>[10]</sup>的残差网络（ResNet）提出以后，在一定程度上改善了网络加深后的梯度爆炸或消失现象，使得深度网络更加易于加深和训练。随后，残差网络几乎被运用于所有的深度学习网络中，在超分辨领域中也同样得到广泛的运用。Lim 等人<sup>[11]</sup>通过移除传统残差网络中批归一化层构建了一种增强的深度超分辨率网络（EDSR），该网络主要通过扩大模型的尺寸来提高网络的性能，其重建出来的效果又

进一步得到提升。由于这种模式在重建图像质量上取得了较大的进步，使得更宽更深网络变得越来越流行。随后，Zhang 等人<sup>[12]</sup>利用通道注意力机制设计了 RCAN 网络，该网络的层数达到了 400+层。而 Dai 等人<sup>[13]</sup>的 SAN 网络在 RCAN 的基础上提出二阶注意力机制，但网络的参数依然巨大。在如此深的网络情况下，毫无疑问他们的网络能够取得更好的指标。但是，随着网络宽度和深度的增加，不可避免地带来了庞大的网络参数，比如 EDSR 的参数到达了惊人的 44M，RCAN 也高达 15M。庞大的网络参数占用了过多的计算资源，这给其在应用领域的推广带来了阻碍，并且过于巨大的网络也容易造成网络的过拟合等问题。

为了减少网络参数，Hui 等人<sup>[14]</sup>设计出了蒸馏神经网络来充分利用网络的浅层信息。Li 等人<sup>[15]</sup>构建了一种循环神经网络，利用循环结构将深层信息通过循环网络回传到初始层并且将其编码映射到浅层信息中。Yang 等人<sup>[16]</sup>利用多尺度、多级别的 U 型网络结构来进行优化，以减少网络参数。尽管这些网络在控制网络参数和保证图像重建质量上都取得了较好的效果，但在控制参数和提升图像质量上还有进一步的提升空间。

基于上述分析，本文在网络参数的减少和重建图像细节的增强方面，提出了一种渐进式特征增强网络的超分辨率重建方法。主流的基于深度学习的图像超分辨方法大多都是利用 3×3 的卷积核提取特征，受到感受野大小的限制，提取到的局部特征不明显。因此，往往会通过改变特征图的大小或者改变卷积核的大小来增加网络感受野进行图像特征的提取。由于改变特征图的大小不可避免地会丢失掉图像的部分特征信息，而采用不同的卷积核提取特征则会增加网络的参数且不易于训练。针对这一问题，我们在不改变卷积核及特征图大小的前提下，利用扩张卷积的方法增加感受野来提取多个尺度的特征。基于人眼视觉观察机制中不同特征引起人眼的关注度不同的这一原理，本文在多个尺度特征提取之后引入注意力机制，实现对多尺度的特征进行分配权重参数，起到特征信息增强的作用。基于上述两个思想，本文构建了一个多尺度注意力模块来实现浅特征层信息和深层特征信息的增强。本文所提出的超分辨率重建网络的主

要贡献如下：

(1)我们设计了一个端到端的渐进式特征增强网络的图像超分辨率重建算法。在公共数据集上的实验结果表明，相比于其它的方法，我们的网络取得了更好的效果。

(2)我们设计了一个多尺度注意力机制模块，模块中利用扩张卷积的方式来提取不同尺度的特征信息，并引入注意力机制对不同尺度的特征图分配权重，来实现图像特征信息的增强。其中，扩张卷积的方式在不增加网络参数及不丢失特征信息的同时扩大了卷积核的感受野，且使网络更容易训练。

(3)在整个网络结构中，我们设计了两种信息增强模块，即浅层特征信息增强模块和深层特征信息增强模块。这两个模块中都采用了多尺度注意力机制模块，并针对提取的浅层信息和深层信息来实现不同尺度的特征提取和增强。

## 2 相关工作

在本节中，我们将介绍与本文工作相关的一些工作，包含基于深度学习的单幅图像超分辨率重建、图像重建的多尺度结构以及注意力机制等。

### 2.1 基于深度学习的单幅图像超分辨重建

深度学习的基本思想是以数据为驱动，通过一系列的隐藏层来实现输入到输出的映射，而单幅图像超分辨率的目的在于通过输入 LR 图像来得到符合人眼视觉的 HR 图像。因此比较容易将深度学习引入到单幅图像超分辨领域中。基于深度学习的单幅图像超分辨率重建的一般过程如下：首先，通过对数据进行处理得到包含 LR 和 HR 图像对的数据集，通常数据集样本数量要求比较大且包含的信息越丰富越好；然后设计算法的网络框架，一般网络结构越深，提取到的细节信息越丰富，得到的结果越好；最后通过训练得到最终的从 LR 图像到超分辨率图像的映射模型。在 Dong 等人<sup>[7]</sup>的推动下，深度学习被广泛应用于单幅图像超分辨率重建过程中。相比于之前的方法，无论在主观视觉感知上还是在客观指标上都获得了质的提升。但是目前的主流深度学习超分辨率方法热衷于提升客观指标，不断地将网络加大加深，忽视了网络模型的大小对于其在应用领域推广带来的影

响。为解决这一问题，本文提出了渐进式特征增强的图像超分辨率算法，在控制网络参数的同时取得了较好的重建效果。

### 2.2 图像重建的多尺度结构

多尺度的本质是对信号进行不同粒度的采样，即在不同的尺度下能够观测到不同的特征。源于多尺度这一特性，该结构已成为计算机视觉研究的热点之一。一些学者也在探索多尺度在图像复原任务中的作用，Han 等人<sup>[17]</sup>提出了双态递归网络 (DSRN)，通过联合低分辨率和高分辨率尺度上的信息来实现图像超分辨率。具体来说，DSRN 中的递归信号通过延迟反馈的方式来进行两个尺度间的信息交换。多尺度残差网络 (MSRN)<sup>[18]</sup>通过使用不同尺度的卷积核来提取图像在不同尺度下的特征。Yang 等人<sup>[16]</sup>提出多级多尺度图像超分辨率网络 (M2SR)，利用残差 U 型网络和注意力 U 型网络提取图像的多尺度特征，增强网络的表达能力。本文在上述思想的基础上，设计出多尺度注意力块来进一步增强网络提取多尺度特征的能力。

### 2.3 注意力机制

注意力机制来源于人的感知理论，指的是人类的视觉系统会在感兴趣的区域上投入更多的关注<sup>[19]</sup>。近些年来，许多视觉任务利用注意力机制来提升卷积神经网络的性能。比如在图像和视频分类任务中<sup>[20, 21]</sup>。Wang 等人<sup>[20]</sup>通过引入非局部空间注意力构建了非局部神经网络。与之相对应的，Hu 等人<sup>[21]</sup>提出了通道注意力机制模块 (SENet) 来处理通道域的关系并且取得了显著的效果。

最近，注意力机制也被引用到图像超分辨率领域并且使得图像重建质量获得显著提高。然而。这些网络不论是引用通道注意力机制还是空间注意力机制都是在单一尺度下，忽略了多尺度在图像重建中的作用。因此，为了强化多尺度结构提取特征效率，我们将通道注意力与多尺度相结合，进一步增强多尺度结构提取图像特征。

## 3 本文方法

为了更好地重建出 HR 图像的细节信息，本文提出一种基于渐进式特征增强网络的超分辨率重建方法。该网络整体结构如图 1 所示，主要包含两个模块：浅层信



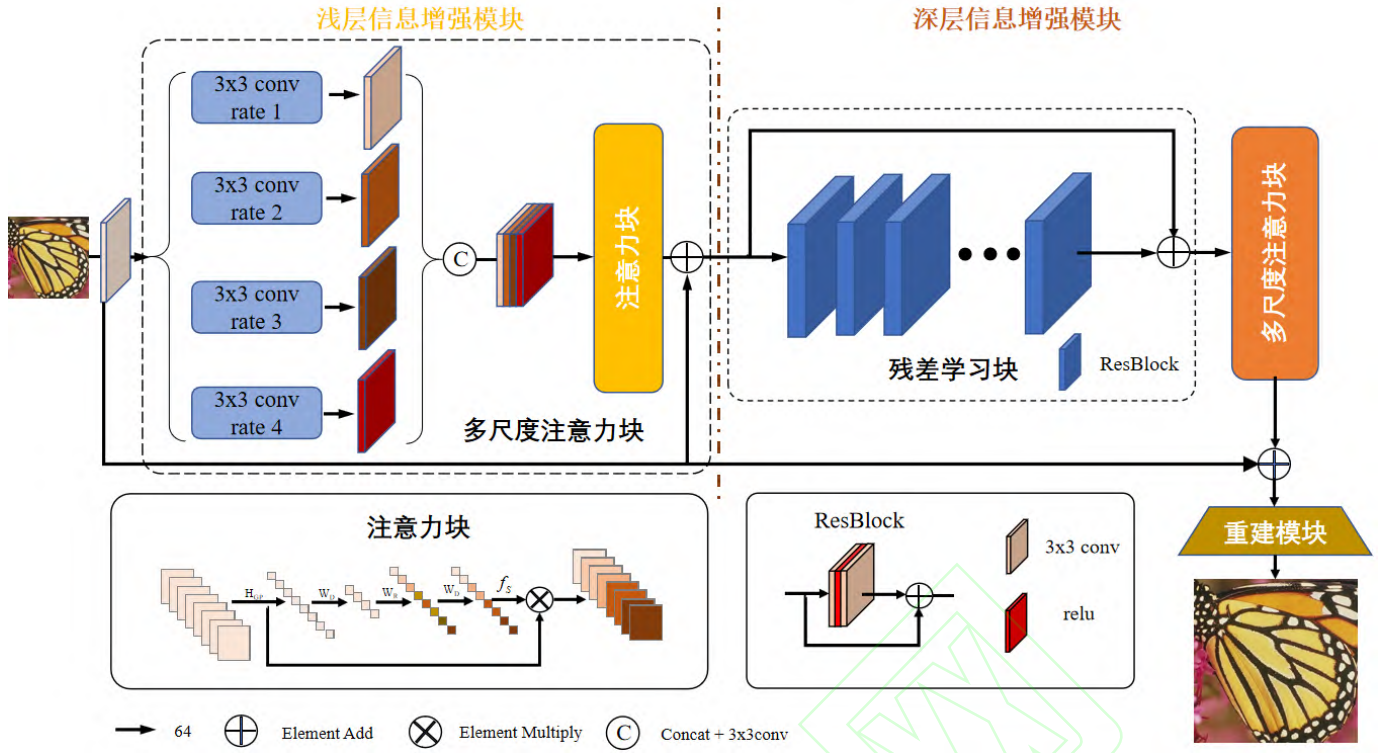


图 1 网络框架流程图

Fig.1 Flow chart of the proposed network

息增强模块和深层信息增强模块。该方法重建 HR 图像的过程描述如下。首先，我们对数据集进行预处理来增强网络的表达能力；然后，LR 图像先通过一个卷积层获得图像的浅层信息，再利用一个多尺度注意力模块提取图像的浅层多尺度特征并对其进行增强；接着，增强后的特征信息再通过一个残差学习模块获得图像的深层特征，同样经过一个多尺度注意力模块达到特征信息增强的目的；最后，利用一个跳跃连接将图像初始的浅层特征与深层多尺度特征进行像素级相加并通过一个上采样操作得到最终的 HR 图像。

### 3.1 数据预处理

为了增强网络的泛化能力，防止过拟合，我们对数据集进行了预处理。首先我们将训练集图像大小比例分别调整为 1, 0.9, 0.8, 0.7, 0.6，并且对得到的五组图像进行 180°翻转，由此得到样本量为原数据十倍的训练数据集。这些操作可以增加数据集包含的信息量，丰富数据特征，可增强网络的鲁棒性和泛化能力，在一定程度上也强化了网络的学习能力。

### 3.2 网络结构描述

超分辨率重建过程主要体现在对图像纹理特征的重

建，因此本文通过构建一种渐进式的增强网络来实现图像纹理特征的重构和增强。该网络如图 1 所示，主要由两大模块组成，其中浅层特征增强模块是由一个卷积层和一个多尺度注意力块构成，深层信息增强模块是由一个残差学习块和一个多尺度注意力块构成。下面主要对网络中的浅层信息增强模块、深层信息增强模块、及最后图像重建过程进行详细描述。

#### 3.2.1 浅层信息增强模块

首先将接收到的红、绿、蓝三个通道的 LR 图像，通过一个卷积层获得初始的浅层特征图。该卷积层利用大小为 3 的卷积核获得 64 个特征图，其操作如下：

$$F_0 = \text{Con}_1(I_{LR}) \quad (1)$$

其中  $I_{LR}$  代表输入的 LR 图像， $\text{Con}_1(\bullet)$  表示第一层卷积层的卷积操作， $F_0$  是提取到的初始浅层特征。

然后，将获得的初始浅层特征通过一个多尺度注意力块实现浅层信息的增强，该模块主要由多尺度特征提取部分和注意力机制部分组成。其中多尺度特征的提取结构不同于传统的多尺度结构，我们的结构既不改变特征图的大小也无需改变卷积核的大小。我们采用大小相同的 3×3 的卷积核利用扩张卷积的方式来提取图像四个尺度的特征，即卷积核大小为 3×3，膨胀率分别为 1、2、

表 1 在数据集 Set5、Set14、BSD100、Urban100、Manga109 上放大倍数分别为 2、3、4 的平均 PSNR(dB)和 SSIM 的结果比较

Table 1 The average results of PSNR/SSIM with scale factor  $\times 2$ ,  $\times 3$  and  $\times 4$  on datasets Set5, Set14, BSD100, Urban100 and Manga109

数据集	尺度	Bicubic	SRCNN	VDSR	DRCN	SRMDF	M2SR	Our
Set5	2	33.66/0.9299	36.66/0.9542	37.53/0.9587	37.63/0.9588	37.79/0.9601	38.01/0.9607	<b>38.17/0.9611</b>
	3	30.39/0.8682	32.75/0.9090	33.66/0.9213	33.82/0.9226	34.12/0.9254	34.43/0.9275	<b>34.61/0.9286</b>
	4	28.42/0.8104	30.48/0.8628	31.35/0.8838	31.53/0.8854	31.96/0.8925	32.23/0.8952	<b>32.35/0.8969</b>
Set14	2	30.24/0.8688	32.45/0.9067	33.03/0.9124	33.04/0.9118	33.32/0.9159	33.72/0.9202	<b>33.93/0.9205</b>
	3	27.55/0.7742	29.30/0.8215	29.77/0.8314	29.76/0.8311	30.04/0.8382	30.39/0.8440	<b>30.49/0.8456</b>
	4	26.00/0.7027	27.50/0.7513	28.01/0.7674	28.02/0.7670	28.35/0.7787	28.67/0.7837	<b>28.79/0.7864</b>
BSD100	2	29.56/0.8431	31.36/0.7863	31.90/0.8960	31.85/0.8942	32.05/0.8985	32.17/0.8997	<b>32.31/0.9011</b>
	3	27.21/0.7385	28.41/0.7863	28.82/0.7976	28.80/0.7963	28.97/0.8025	29.11/0.8056	<b>29.18/0.8074</b>
	4	25.96/0.6675	26.90/0.7101	27.29/0.7251	27.23/0.7233	27.49/0.7337	27.60/0.7373	<b>27.66/0.7393</b>
Urban100	2	26.88/0.8403	29.50/0.8946	30.76/0.9140	30.75/0.9133	31.33/0.9204	32.20/0.9295	<b>32.74/0.9344</b>
	3	24.46/0.7349	26.24/0.7989	27.14/0.8279	27.15/0.8276	27.57/0.8398	28.29/0.8551	<b>28.57/0.8615</b>
	4	23.14/0.6577	24.52/0.7221	25.18/0.7524	25.14/0.7510	25.68/0.7731	26.19/0.7889	<b>26.43/0.7966</b>
Manga109	2	31.05/0.935	35.72/0.968	37.16/0.974	37.57/0.973	38.07/0.976	38.71/0.9772	<b>39.08/0.9779</b>
	3	26.95/0.856	30.48/0.912	32.01/0.934	32.31/0.936	33.00/0.940	33.59/0.9447	<b>33.97/0.9469</b>
	4	25.15/0.789	27.66/0.858	28.82/0.886	28.97/0.886	30.09/0.902	30.51/0.9093	<b>30.94/0.9135</b>

3、4。其中，每个尺度的卷积块组成方式为 conv+ReLU+conv，即在两个卷积层之间加一层激活函数。该结构能够更加充分地提取多尺度特征、参数更少且易于训练。其结构表示公式如下：

$$F_i = Con_i(relu(Con_i(F_0))) \tag{2}$$

$$F_{MF} = H_c(F_1, F_2, F_3, F_4) \tag{3}$$

式中， $Con_i(\bullet)$  表示膨胀率为  $i$  的  $3\times 3$  卷积核， $relu(\bullet)$  代表 ReLU 激活函数， $F_i$  是提取到的不同尺度特征。 $H_c(\bullet)$  表示对多尺度特征的融合，包含 Concat 操作和一层卷积操作。 $F_{MF}$  表示融合得到的图像多尺度特征。

对于提取到的图像多尺度局部与非局部特征，我们希望能够更加有效地利用这些不同的特征，而不是对于每个尺度特征以同等方式对待。因此，本文将通道注意力加入到多尺度结构之后来有选择性地增强网络提取的特征，通道注意力结构块如图 1 中左下方的方框所示。具体操作如下：首先，利用全局平均池化函数将输入的特征图压缩成与通道数相对应的一维向量；然后对向量进行一个全连接的编解码，即将向量维度降低到原输入维度的 1/16，再经过一个 ReLU 激活函数之后再扩张到输入的维度。这一过程的主要目的在于构建通道之间的

相关性。相比于直接使用一个全连接层，其优点在于：（1）拉大通道间的距离，具有更多的非线性，能够更好地拟合通道之间的复杂关系；（2）相比于全连接操作，在计算量和参数量上有巨大的降低。最后，将得到的权重参数运用 Sigmoid 函数归一化映射到 0~1 之间再加权到最初输入的每个通道特征图中，这样就完成了对多尺度特征图权重参数的分配。

对于经过多尺度结构和注意力机制结构获得的多尺度特征，我们利用跳转连接将初始浅层特征与其进行像素级相加，这种局部残差的方式能够加快训练的收敛速度，同时在一定程度上改善网络的性能，获得浅层信息的增强。浅层信息增强模块的公式表示如下：

$$F_{MSAB} = H_{AB}(F_{MB}) + F_0 \tag{4}$$

式中的  $F_{MB}$  表示公式 3 得到的浅层多尺度特征， $H_{AB}(\bullet)$  表示为通道注意力结构的操作， $F_{MSAB}$  是经过浅层信息增强模块得到的图像特征。

### 3.2.2 深度信息增强模块网络

网络的浅层和深度学习获得的特征所体现的图像信息是不同的，因此，我们针对浅层信息增强模块得到的图像特征信息，进一步使用残差学习模块来学习不同尺



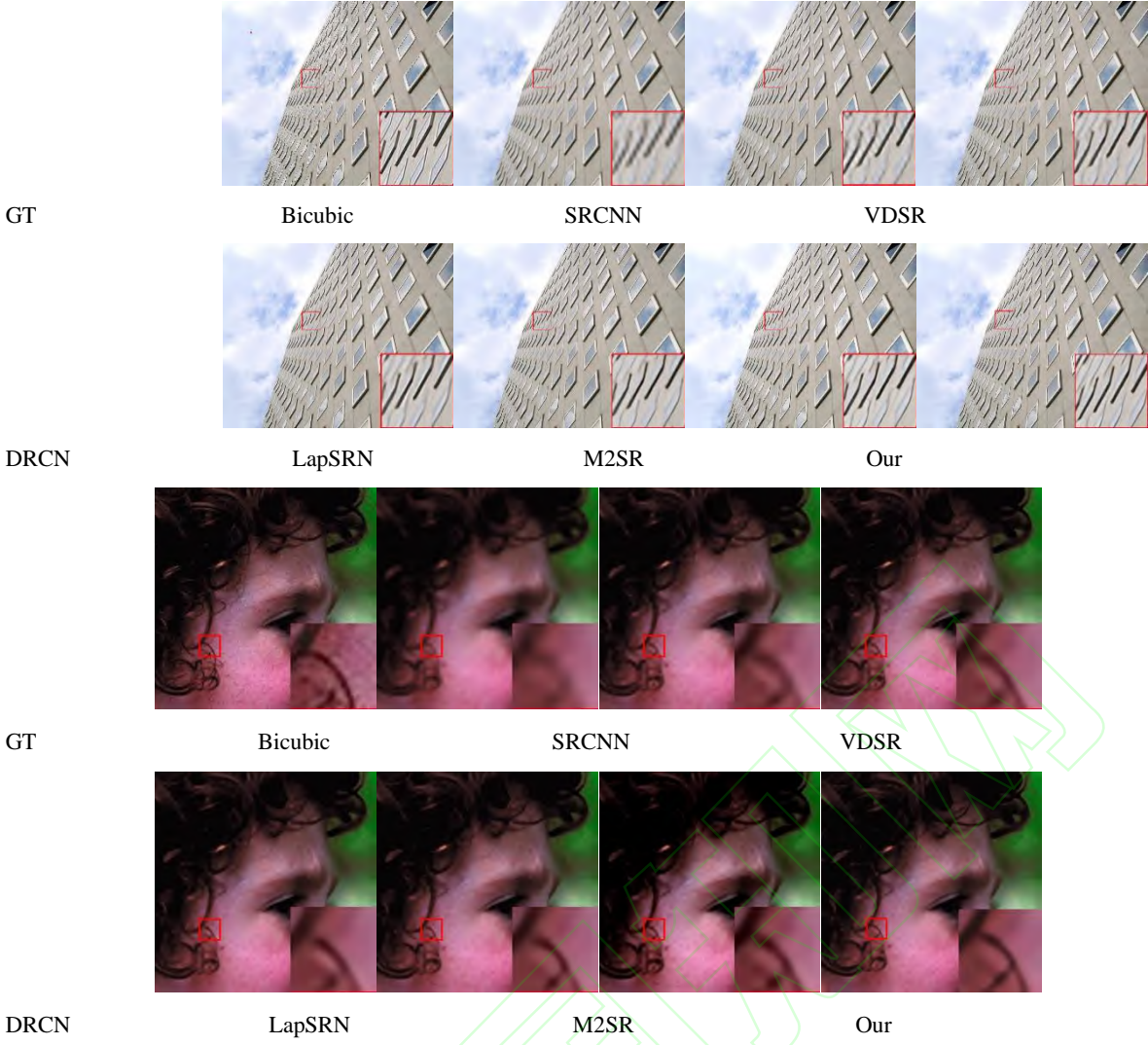


图 2 图像 Urban\_010 和图像 Set5\_head 在 4 倍下的效果对比图  
Fig.2 The SR results of Urban\_010 and Set5\_head with scale factor  $\times 4$

度特征图之间的关联性以及深层的特征信息，并利用多尺度注意力模块来实现深层信息的增强。其中，残差学习模块如图 1 中所示，该残差结构包含 32 个残差单元组成，每个残差单元有两个  $3 \times 3$  卷积层，在两个卷积层中间接入一层激活函数 ReLU。整个残差单元的输入和输出通道数均为 64。

接着，将残差学习模块得到的特征信息输入到一个多尺度注意力块，通过四层卷积来获得不同尺度的特征信息，并利用通道注意力块获得不同特征图的权重系数，来自适应地增强或抑制不同尺度的结构特征，获得增强后的深度特征信息。

### 3.2.3 图像重建

在上述两个模块学习到的特征信息基础上，我们利用跳转连接将初始的浅层特征与深层增强模块学习到的特征进行像素级相加，构建全局残差网络，输出重建后

的特征图。然后，将重建后的特征图经过一个上采样模块来获取最终的 HR 图像。上采样模块包含一个  $3 \times 3$  卷积层、一个激活函数 ReLU 和 PixelShuffle 函数，卷积层的输入通道数为 64，输出通道数为 3。整个网络结构可以表示为：

$$I_{SR} = (H_{DFEB}(H_{SFEB}(I_{LR})) + F_0) \uparrow \quad (5)$$

其中  $I_{SR}$  表示为重建得到的 HR 图像， $I_{LR}$  代表 LR 图像， $H_{SFEB}(\bullet)$  和  $H_{DFEB}(\bullet)$  分别代表了浅层信息增强模块和深层信息增强模块的操作。 $\uparrow$  是对图像进行上采样操作。

## 4 实验设置及结果分析

### 4.1 实验数据集

在实验过程中，我们选用了 DIV2K<sup>[22]</sup> 作为我们网络的训练集。该数据集由 800 张训练集图片和 100 张验证集图片组成。为了充分利用数据集，增强网络的表达能

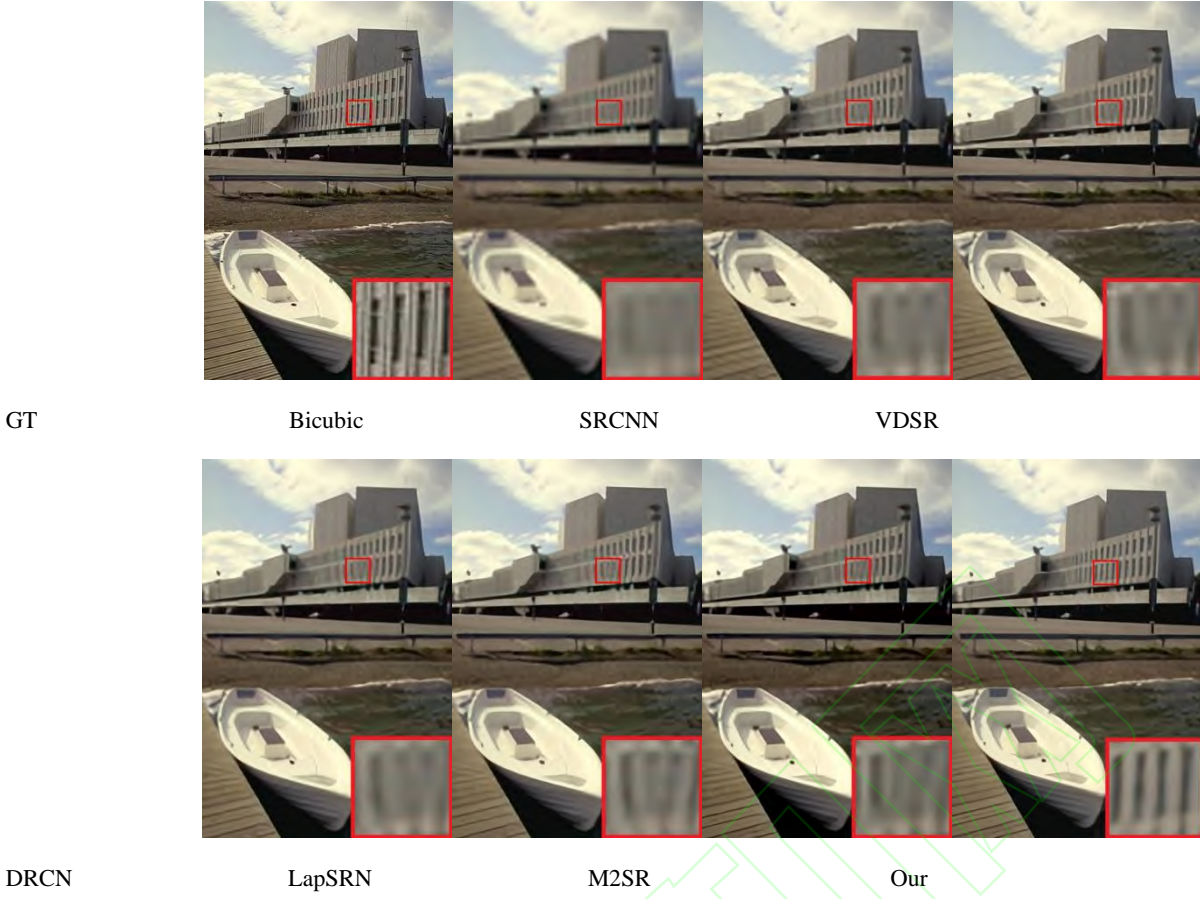


图 3 图像 78004 (BSD100) 在放大 4 倍下的效果图

Fig.3 The SR results of 78004(BSD100) with scale factor  $\times 4$

力，我们对训练数据集进行了预处理。关于测试集，我们选用 5 个标准数据集，分别为：Set5<sup>[23]</sup>，Set14<sup>[24]</sup>，BSD100<sup>[25]</sup>，Urban100<sup>[26]</sup>，Manga109<sup>[27]</sup>。其中测试集 BSD100 包含有多种风格类型的图片，Urban100 为各种类型的建筑物图片，Manga109 为各种类型的卡通图片。这 5 个测试集具有丰富多样的信息，能够很好地验证超分辨率方法的有效性。

## 4.2 客观评价指标

为了客观评价方法的性能，实验中我们采用 PSNR 和 SSIM<sup>[28]</sup>作为客观评价指标。PSNR 与均方差（Mean Square Error, MSE）有一定的相关性。PSNR 值越大时，网络的输出与真实图像之间的均方差越小，这就代表了重建出的图像质量越高。PSNR 定义如下所示：

$$MSE(y, \hat{y}) = \frac{1}{2} \left\| y - \hat{y} \right\|^2 \quad (6)$$

$$PSNR(y, \hat{y}) = 20 \log(s) - 10 \log MSE(y, \hat{y}) \quad (7)$$

其中， $y$  表示原始的真实图像， $\hat{y}$  表示网络重建得到的 HR 图像， $s$  表示了最大的可能像素值（ $s$  取值为 255）。

人眼视觉感知结构相似指数（Structural Similarity Index, SSIM）评价指标通常用来衡量两幅图像之间的相似度，被用来弥补在相同的 PSNR 情况下呈现的不同质量感知问题。SSIM 的取值范围在 0~1 之间，数值越接近于 1 表示图像失真越小，算法性能越好。公式为：

$$SSIM(y, \hat{y}) = \frac{(2u_y u_{\hat{y}} + c_1)(2\sigma_{y\hat{y}} + c_2)}{(u_y^2 + u_{\hat{y}}^2 + c_1)(\sigma_y^2 + \sigma_{\hat{y}}^2 + c_2)} \quad (8)$$

其中  $u_y$  和  $u_{\hat{y}}$ 、 $\sigma_y$  和  $\sigma_{\hat{y}}$  分别表示原图像  $y$  和超分图像  $\hat{y}$  的均值和标准差，而  $\sigma_{y\hat{y}}$  则表示交叉协方差， $c_1$  和  $c_2$  为常数，通常  $c_1 = (0.01s)^2$ ， $c_2 = (0.03s)^2$ ， $s$  的涵义同公式(7)。

## 4.3 损失函数

我们采用  $L_1$  损失函数来优化提出的网络。对于给出的训练集  $\{I_{LR}^i, I_{HR}^i\}_{i=1}^N$  包含了  $N$  个 LR 和 HR 图像对。我们的网络目标是训练图像对并利用  $L_1$  损失函数来进行优化，公式如下所示：

$$L(\theta) = \frac{1}{N} \sum_{i=1}^N \|H_{PFEN}(I_{LR}^i) - I_{HR}^i\|_1 \quad (9)$$

$H_{PFEN}(\bullet)$  表示网络重建结果。 $\|\bullet\|_1$  定义为  $L_1$  范数，



而  $\theta$  表示了网络中的参数。

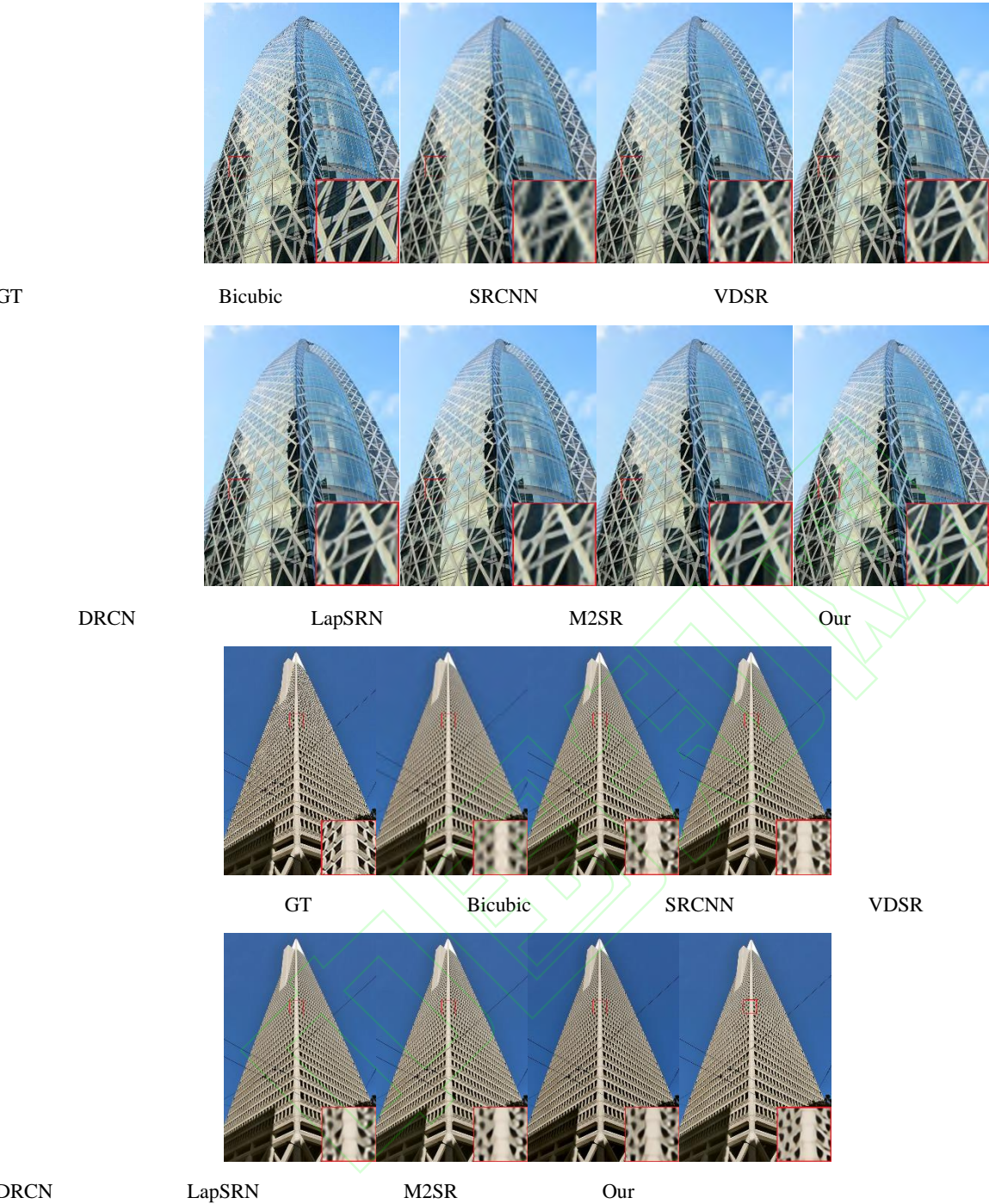


图 4 图像 image\_039(Urban100)和 image\_048(Urban100)在放大 4 倍情况下的效果对比图  
Fig.4 The SR results of image\_039 (Urban100) and image\_048(Urban100) with scale factor  $\times 4$

4.4 训练配置

本文的模型采用 ADAM 优化函数来优化网络，网络的初始学习率为设置为  $3 \times 10^{-4}$ ，并且每迭代  $2 \times 10^5$  次学习率减半。本文所提出的方法实现测试的硬件环境搭配 Intel(R)Xeon(R)CPU(2.6 GHz)、内存 32 GB、配置 NVIDIA GeForce GTX 1080 Ti GPU 的计算机。软件环境为 64 位 Win10 操作系统,PyTorch 框架和 Matlab R2014a。

表 2 多尺度注意力模块和残差学习模块在数据集 Set5 放大 4 倍下结果比较

learning block with scale factor $\times 4$ on Set5			
多尺度注意力块	✓	✗	✓
残差学习块	✗	✓	✓
PSNR	31.52	32.30	<b>32.35</b>
SSIM	0.8851	0.8965	<b>0.8969</b>

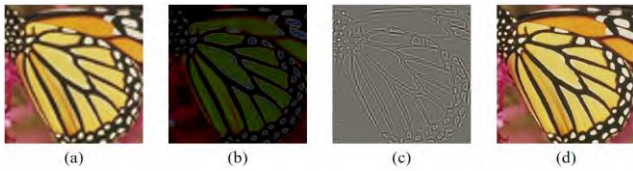


图 5 网络的不同深度输出结果。(a) 第一层卷积输出结果；(b) 浅层特征增强模块输出结果；(c) 深层残差特征图；(d) 模型的最终输出结果。

Fig.5 Results of the network at different depths. (a) The result of first layer convolution; (b) The result of shallow feature enhancement module; (c) The feature map of deep residual; (d) The final result of our model.

#### 4.5 实验结果与分析

实验中，我们将本文方法与现阶段一些具有代表性的方法作对比，其中包含 Bicubic、SRCNN<sup>[7]</sup>、VDSR<sup>[8]</sup>、DRCN<sup>[9]</sup>、MS-LapSRN<sup>[29]</sup>和 M2SR<sup>[16]</sup>。为了公平比较，我们将所有的方法在公共测试集 Set5、Set14、BSD100、Urban100 和 Manga109 上进行实验测试，然后对于不同测试集上得到的 PSNR 和 SSIM 指标值分别取平均值。获得的结果列于表 1 中，表中红色加粗字体表示在指标上取得的最好结果。从表中可以看出我们的方法获得的 PSNR 和 SSIM 值都高于其它的对标方法获得的结果值，比如在数据集 Set5 上放大 2 倍的情况下我们的模型重建图像的 PSNR 和 SSIM 值相比于排序第二的 M2SR 方法分别提高了 0.16 (dB) 和 0.004，相比于其它方法我们的指标提升得更加明显。表 1 中的客观指标的实验对比结果证明了我们方法的有效性。

在主观表现方面，将我们的方法也与上述主流方法进行了对比，其中 GT (Ground Truth) 代表原始 HR 图像。图 2、图 3、图 4 展现了各种方法的在不同测试图像

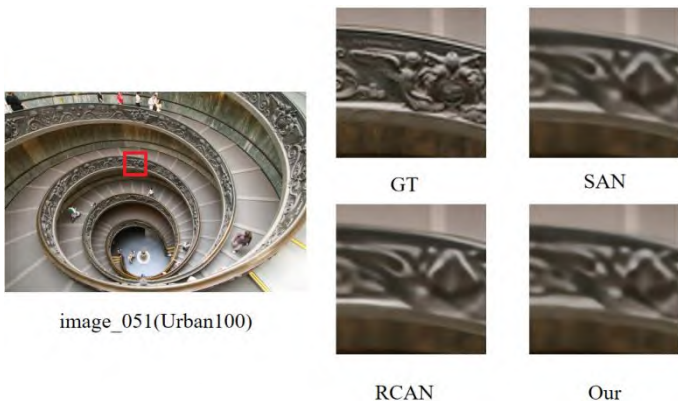


图 6 图像 image\_051(Urban100)在放大 4 倍情况下的效果对比图

Fig.6 The SR results of image\_051 (Urban100) with scale factor  $\times 4$

表 3 模型大小在数据集 Set5 放大 2 倍情况下的效果比较

Table 3. Model size and performance comparison on Set5( $\times 2$ )

模型	EDSR	DBPN	RCAN	SAN	Our
参数	43M	10M	16M	15.7M	<b>3M</b>
PSNR	38.11	38.09	38.27	<b>38.31</b>	38.17
SSIM	0.9602	0.9600	0.9614	<b>0.9620</b>	0.9611

中对比效果。为了突出对比效果，我们选取了图像的局部区域使用双三次插值的方法放大了 4 倍。通过观察对比图可以看出，我们的方法重建的图像拥有更丰富的细节，更清晰的边缘且更加符合人眼视觉，结果看起来也更清晰。尤其在图 3 的局部放大部分可以清晰地观察到，其它的方法重建出的图像细节很模糊，甚至无法重建出图像的边缘信息，而我们的方法重建出的细节更加清晰，具有较好的识别度。这些结果也表明了，我们的方法在主观表现上也取得了更优的效果。

#### 4.6 消融实验

对于残差学习模块和多尺度注意力模块的有效性，我们在测试集 Set5 中对图像放大 4 倍的情况下进行了剥离实验来验证本文模型的优越性。表 2 给出了网络包含多尺度注意力模块和残差学习模块中一种或者两种的情况下的实验结果。从表中可以看出，当我们的网络同时包含多尺度注意力模块和残差学习模块时 PSNR 值相比于只包含多尺度注意力模块和只包含残差学习模块的情况下分别提高了 0.83 dB 和 0.05 dB，而在 SSIM 上也获得了最大值。

为了更好的展示网络中两个增强模块的功能，我们对两个模块学习到的特征图进行了可视化，其结果如图 5 所示。其中图(a)表示网络在第一层卷积输出的结果，图(b)和图(c)分别代表浅层特征增强模块的输出结果和深层残差的特征图，图(d)为整个模型的最终输出结果。从图(b)和(c)可以看出，浅层特征增强模块学习到了图像的结构信息，而深层特征增强模块学习到了更多的图像纹理细节。实验结果表明了本文网络中的两个增强模块起到了很好的纹理细节补充的作用。

#### 4.7 参数分析

为了进一步验证本文提出模型的有效性，我们在参数的数量方面和图像重建质量方面将本文方法与当前公



认取得效果最好的一些深度学习的超分辨率方法进行了比较, 这些方法包括 EDSR<sup>[11]</sup>, DBPN<sup>[30]</sup>, RDN<sup>[31]</sup>, RCAN<sup>[12]</sup>和 SAN<sup>[13]</sup>, 参数结果如表 3 所示。从表中可以看出我们的方法在取得了较好的客观指标的同时, 大幅缩减了网络的参数量。由此可以证明, 我们的网络在图像重建质量和模型压缩上取得了更好的平衡, 即我们的模型在参数较少时也能获得较好的 PSNR 和 SSIM 结果。虽然本文的模型与 RCAN 均是利用 SEblock<sup>[21]</sup>来对通道域进行增强, 但 RCAN 模型是在单一尺度下对通道域赋予权值, 而我们的网络考虑了多尺度信息在通道域上的差异性, 利用注意力机制来增强多尺度信息的表达。在主观视觉效果上, 如图 6 所示, 我们的方法与目前客观指标上最优的 RCAN<sup>[12]</sup>和 SAN<sup>[13]</sup>方法进行比较, 取得了相近的重建质量, 但我们的模型参数仅约是上述两种模型参数的 1/5。

## 5 结论

本文提出了一种基于渐进式特征增强网络的超分辨率重建方法。该方法着重于对浅层信息和深层信息的特征增强, 其增强的过程是基于设计的一种多尺度注意块, 利用扩张卷积的方式来高效地获取图像多尺度下的结构信息并利用通道注意力机制来增强提取的特征。不同于以往的专注于加深和加宽网络的深度学习超分辨率方法, 该网络的构建既考虑到了图像细节的重建也考虑了模型参数量问题。实验结果表明, 本文提出的方法无论在客观指标还是在主观表现下都取得了较好的效果。本文方法的不足之处在于, 当放大倍数较大时在纹理丰富的区域会出现平滑现象。未来, 我们的工作目标是在较少参数的前提下, 设计出更加高效合理的多尺度结构提取更丰富的特征来消除平滑现象, 从而获得更好的客观指标和主观效果。

## 参考文献

- [1] Zhang L, Wu X. An Edge-guided Image Interpolation Algorithm via Directional Filtering and Data Fusion[J]. IEEE Transactions on Image Processing, 2006, 15(8):2226-2238.
- [2] 汪雪林, 文伟, 彭思龙. 基于小波域局部高斯模型的图像超分辨率[J]. 中国图象图形学报, 2004, 9(8): 941-946.  
Wang Xue-lin, Wen Wei, Peng Si-long. Image Super Resolution Based on Wavelet-domain Local Gaussian Model [J]. Journal of Image and Graphics, 2004, 9(8): 941-946.(in Chinese)
- [3] Wang S, Zhang L, Liang Y and Pan Q. Semi-coupled Dictionary Learning with Applications to Image Super-Resolution and Photo-sketch Synthesis[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, 2012: 2216-2223.
- [4] 班晓征, 李志华, 李贝贝, 徐敏达. 基于改进广义全变分的稀疏图像重建算法[J]. 激光与光电子学进展, 2018, 55(11): 245-255.  
Ban Xiaozheng, Li Zhihua, Li Beibei, Xu Minda. Sparse Image Reconstruction Based on Improved Total Generalized Variation [J]. Laser & Optoelectronics Progress, 2018, 55(11): 245-255. (in Chinese)
- [5] Zhang K, Gao X, Tao D, et al. Single Image Super-Resolution with Non-Local Means and Steering Kernel Regression[J]. IEEE Transactions on Image Processing, 2012, 21(11): 4544-4556.
- [6] 林格平, 马晓川, 鄢社锋, 郝程鹏, 林津丞. 无字典的超分辨率多径稀疏信道估计方法[J]. 信号处理, 2017, 33(9): 1239-1247.  
LIN Ge-ping, MA Xiao-chuan, YAN She-feng, HAO Cheng-peng, LIN Jin-cheng. Super-Resolution Multipath Sparse Channel Estimation With No Dictionary. Journal of Signal Processing, 2017, 33(9): 1239-1247. (in Chinese)
- [7] Dong C, Loy C C, Tang X. Accelerating the Super-Resolution Convolutional Neural Network[C]//2016 European Conference on Computer Vision. ECCV, 2016: 391-407.
- [8] Kim J, Lee J K, and Lee K M. Accurate Image Super-Resolution Using Very Deep Convolutional Networks[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, 2016: 1646-1654
- [9] Kim J, Lee J K, and Lee K M. Deeply-Recursive Convolutional Network for Image Super-Resolution[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, 2016: 1637-1645.
- [10] He K, Zhang X, Ren S. and Sun J. Deep Residual Learning for Image Recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, 2016: 770-778.
- [11] Lim B, Son S, Kim H, et al. Enhanced Deep Residual Networks for Single Image Super-Resolution[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. CVPRW, 2017: 1132-1140.
- [12] Zhang Y, Li K, Li K, et al. Image Super-Resolution Using Very Deep Residual Channel Attention Networks[C]//2018 European Conference on Computer Vision. ECCV, 2018: 294-310.
- [13] Dai T, Cai J, Zhang Y, et al. Second-Order Attention Network for Single Image Super-Resolution[C]//2019 IEEE Computer Vision and Pattern Recognition. CVPR, 2019: 11065-11074.
- [14] Hui Z, Gao X, Yang Y, et al. Lightweight Image Super-Resolution with Information Multi-distillation Network[C]//2019 ACM Multimedia. ACMM, 2019: 2024-2032.
- [15] Li Z, Yang J, Liu Z, Yang X, G. Jeon and Wu W. Feedback Network for Image Super-Resolution[C]//2019 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, 2019: 3862-3871.
- [16] Yang Y, Zhang D, Huang S, et al. Multilevel and Multiscale Network for Single-Image Super-Resolution [J]. IEEE Signal Processing Letters, 2019, 26(12): 1877-1881.



- [17] Han W, Chang S, Liu D, Yu M, M. Witbrock and Huang T. Image Super-Resolution via Dual-State Recurrent Networks[C] // 2018 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, 2018: 1654-1663.
- [18] Li J, Fang F, Mei K, et al. Multi-scale Residual Network for Image Super-Resolution[C] // 2018 European Conference on Computer Vision. ECCV, 2018: 527-542.
- [19] Itti L, Koch C, Niebur E, et al. A Model of Saliency-based Visual Attention for Rapid Scene Analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11): 1254-1259.
- [20] Wang X, Girshick R, Gupta A, et al. Non-local Neural Networks[C]//2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2018: 7794-7803.
- [21] Hu J, Shen L and Sun G. Squeeze-and-Excitation Networks[C]//2018 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, 2018: 7132-7141.
- [22] Timofte R, Agustsson E, Van Gool L, et al. Ntire 2017 Challenge on Single Image Super-Resolution: Methods and Results[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. CVPRW, 2017: 114-125
- [23] Bevilacqua M, Roumy A, Guillemot C, et al. Low-complexity Single Image Super-Resolution Based on Nonnegative Neighbor Embedding[J]. 2012.
- [24] Zeyde R, Elad M, Protter M. On Single Image Scale-Up Using Sparse-Representations[C]//2010 IEEE International Conference on Curves and Surfaces. ICCS, 2010: 711-730.
- [25] Martin D, Fowlkes C, Tal D, et al. A Database of Human Segmented Natural Images and Its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics[C]//2001 IEEE International Conference on Computer Vision and Pattern Recognition. CVPR, 2001, 2: 416-423.
- [26] 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, 2015: 5197-5206.
- [27] 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.
- [28] Wang Z, Bovik A C, Sheikh H R, et al. Image Quality Assessment: From Error Visibility to Structural Similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.
- [29] Lai W, Huang J, Ahuja N, et al. Fast and Accurate Image Super-Resolution with Deep Laplacian Pyramid Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 41(11): 2599-2613
- [30] Haris M, Shakhnarovich G, Ukita N, et al. Deep Back-Projection Networks for Super-Resolution[C]//2018 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, 2018: 1664-1673.
- [31] Zhang Y, Tian Y, Kong Y, Zhong B and Fu Y. Residual Dense Network for Image Super-Resolution[C]//2018 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, 2018: 2472-2481.

## 作者简介



**杨勇男**，1976 年生，湖北鄂州人。江西财经大学信息管理学院教授，主要研究方向为图像处理、机器学习、模式识别。  
E-mail: [greatyangy@126.com](mailto:greatyangy@126.com)



**吴峥男**，1993年生，江西宜春人。江西财经大学信息管理学院硕士生研究生，主要研究方向为图像处理、深度学习。  
E-mail: [724549669@qq.com](mailto:724549669@qq.com)



**张东阳男**，1992年生，河南驻马店人。江西财经大学信息管理学院硕士生研究生，主要研究方向为图像处理、深度学习。  
E-mail: [zdysuyi@163.com](mailto:zdysuyi@163.com)



**刘家祥男**，1996年生，江西吉安人。江西财经大学信息管理学院硕士研究生，主要研究方向为深度学习、图像处理。  
E-mail: [forworkliu@gmail.com](mailto:forworkliu@gmail.com)