

# 递归式多阶特征融合图像超分辨率算法

佟骏超, 费加罗, 陈靖森, 李恒, 丁丹丹

杭州师范大学 国际服务工程学院, 杭州市, 311121

**摘要:** **目的** 近年来, 卷积神经网络在解决图像超分辨率的问题上取得了巨大成功, 不同结构的网络模型相继被提出。通过学习, 这些网络模型对输入图像的特征进行抽象、组合, 进而建立了从低分辨率的输入图像到高分辨率的目标图像的有效非线性映射。在该过程中, 无论是图像的低阶像素级特征, 还是高阶各层抽象特征, 都对像素间相关性的挖掘起了重要作用, 影响着目标高分辨图像的性能。而目前典型的超分辨率网络模型, 如 SRCNN、VDSR、LapSRN 等, 都未充分利用这些多层次的特征。**方法** 本文提出一种充分融合网络多阶特征的图像超分辨率算法: 该模型基于递归神经网络, 由相同的单元串联构成, 单元间参数共享; 在每个单元内部, 从低阶到高阶的逐级特征被级联、融合, 以获得更丰富的信息来强化网络的学习能力; 在训练中, 采用基于残差的策略, 单元内使用局部残差学习, 整体网络使用全局残差学习, 以加快训练速度。**结果** 本文所提出的网络模型在通用 4 个测试集上, 针对分辨率放大 2 倍、3 倍、4 倍的情况, 与深层超分辨率网络 VDSR 相比, 平均分别能够获得 0.24dB、0.23dB、0.19dB 的增益。**结论** 实验结果表明, 所提出的递归式多阶特征融合图像超分辨率算法, 有效提升了性能, 特别是在细节非常丰富的 Urban100 数据集上, 该算法对细节的处理效果尤为明显, 图像客观质量与主观质量都得到显著改善。

**关键词:** 图像超分辨率; 卷积神经网络; 特征融合; 递归神经网络; 残差学习

## Multi-level feature fusion image super-resolution algorithm with recursive neural network

Tong Junchao, Fei Jialuo, Chen Jingsen, Li Heng, Ding Dandan

*Institute of Service Engineering, Hangzhou Normal University, Hangzhou 311121, China*

**Abstract:** **Objective** The recovery of a high resolution (HR) image or video from its low resolution (LR) counterpart, which is referred to as super resolution (SR), has attracted lots of studies in computer vision community. The SR problem is inherently ill-posed because the HR image or video actually does not exist. To address this issue, lots of methods have been proposed. There are some typical methods like bilinear or bicubic interpolation, Lanczos resampling, and internal patch recurrence. Recently, learning based methods, like sparse coding, random forest and convolutional neural networks (CNN), are exploited to create a mapping between LR and HR images. Particularly, the CNN based scheme has achieved great performance improvement. Different network models have been proposed, like SRCNN, VDSR, LapSRN, and DRRN. These models abstract and combine the features of LR image to establish an

**基金项目:** 本课题得到国家自然科学基金(No. 61502133、No. 61602145)与浙江省自然科学基金(No. LQ15F010001)资助。丁丹丹(通信作者), 女, 1983 年生, 博士, 讲师, 主要研究领域为图像处理、视频压缩与编码。E-mail: DandanDing@hznu.edu.cn.

**收稿日期:** 2018-06-20; **修回日期:** 2018-07-12;

Supported by: Project supported by the National Natural Science Foundation of China (Grant No. 61502133 and No. 61602145)

effective nonlinear mapping from LR input images to HR target images. In this process, both low-level and high-level features play a significant role in exploring the correlation between pixels as well as improving the performance of the restored HR images. However, in the typical SR network models mentioned above, features of previous layer are fed into the next layer directly, where the multi-level features are not fully utilized. Inspired by the recent DenseNet, we propose to concatenate and fuse multi-level features from multi-layers. Although multi-level features are exploited in this way, the amount of parameters is huge, which costs long training time and large storage. Therefore, we further propose to employ recursive network architecture for parameter sharing. The overall model aims to developing an efficient CNN model, which can utilize the multi-level features of CNN to improve the SR performance whilst control the number of model parameters within an acceptable range. **Method** We propose an image super-resolution model that fully makes use of multi-level features. The proposed multi-feature fusion recursive network (MFRN) is based on recursive neural network, including the same units in series. The information of features is passed along the basic unit of MFRN, named as multi-feature fusion unit (MFU). Parameters are shared among these basic units, thus the required amount of parameters is reduced effectively. Within each MFU, the input status is obtained from the previous unit with continuous memory mechanism. Then, the features from low-level to high-level are concatenated and fused, thus to obtain abundant features to describe the image. Finally, the useful features are extracted and enhanced, which can describe the mapping relationship between LR and HR accurately. With regard to the training process, the residual learning strategy, including local residual learning inside each units and global residual learning through the whole network, is adopted to accelerate the training speed. Specifically, the global residual learning strategy is employed in the training of overall MFRN and the local residual learning is for MFU. By combining the strategies above, the training difficulty is reduced efficiently and the typical phenomenon like network degradation and gradient vanish can be avoided. In terms of the cost function, the averaged mean square error over the training set is minimized. With the proposed cost function and training methods, we train single model for multiple scales. **Result** We use 291 pictures from public databases as the training set. In addition, data augmentation (rotation or flip) is used. Images with different scales ( $\times 2$ ,  $\times 3$  and  $\times 4$ ) are all included in the training set. Therefore, only single model is trained for all different scales. During training process, we adopt the adaptive learning rate and adjustable gradient clipping to boost the convergence rate while suppressing exploding gradients. We test 4 network models with different number of MFUs, which is corresponding to 29, 37, 53 and 81 layers, respectively. By comparing the convergence rate and performance, it is found that the network with 9 MFUs achieves the best performance. Hence we adopt 9 MFUs in the final CNN model. Although the proposed network is as deep as 37 layers, it elegantly converges at 230 epochs and obtains significant gains. The dominant evaluation criterions of image quality including PSNR, SSIM and IFC are employed for performance assessment of restored images. The experimental results show that the proposed model achieves averaged PSNR gains of 0.24dB, 0.23dB, and 0.19dB, respectively, compared to the very deep convolutional networks VDSR with the general 4 test sets for  $\times 2$ ,  $\times 3$ , and  $\times 4$  resolution. Especially in the data set Urban100 that contains rich details, the proposed MFRN significantly improves the quality of restored images. In addition, the subjective quality of restored images is also illustrated. It can be seen that MFRN can produce relatively sharper edges than the other methods. **Conclusion** A multi-level feature fusion image super resolution algorithm based on recursive neural network, referred to as MFRN, is proposed in this paper. The MFRN consists of multiple multi-level feature fusion units. Several recursive units are stacked to learn the residual image between the HR and LR images. With the recursive learning scheme, parameters are shared among units, thus effectively reduces the number of network parameters. Within each unit, features of different levels are concatenated and fused to provide intensive description of the images. In this way, the proposed MFRN can extract and enhance useful feature adaptively, which leads to accurate mapping between LR and HR images. During the training procedure, we adopt a local residual learning inside each units and a global residual learning through the whole network. As a result, single model is trained for different scales.

The experimental results show that the proposed MFRN greatly improves the performance. Especially in the Urban100 data set, MFRN outperforms up to 0.4dB PSNR gains compared to the classical VDSR model. Compared with the basic recursive network DRRN, up to 0.14dB PSNR improvement is obtained. With regard to the subjective quality, MFRN is specialized at handling the details of images. The visual perception of images is significantly improved.

**Key words :** image super-resolution; convolutional neural network; feature fusion; recursive neural network; residual learning

## 0 引言

图像超分辨率是根据低分辨率 (Low-Resolution, LR) 图像获得对应的高分辨率 (High-Resolution, HR) 图像的过程。目前, 图像超分辨率已经被广泛用于多种领域, 如医学成像<sup>[1]</sup>、卫星成像<sup>[2]</sup>和电子监控<sup>[3]</sup>等。

对于一张低分辨率图像, 在图像采集的模数转换过程中, 往往并未采样得到对应的高分辨率图像, 即高分辨图像并不真实存在, 因此, 图像超分辨率问题是一个不适定问题。而实际在解决该问题时, 为使算法有明确的学习目标并能够客观评价算法性能, 通常假定高分辨率为真实图像, 通过下采样获得低分辨率图像, 该低分辨率图像再经过算法映射得到恢复的高分辨率图像。基于上述假设, 涌现出许多图像超分辨率算法, 如基于自相似性的超分辨率算法 SelfEx<sup>[4]</sup>、基于随机森林的超分辨率算法 RFL<sup>[5]</sup>方法, 等等。

近年来, 借助于强大的学习能力, 深度神经网络, 尤其是卷积神经网络, 为解决图像超分辨问题提供了新思路。Dong 等人最早进行了尝试, 所提出的超分辨率卷积神经网络 (Super-Resolution Convolutional Neural Network, SRCNN)<sup>[6]</sup>包括 3 层, 分别使用  $9 \times 9$ 、 $1 \times 1$ 、 $5 \times 5$  尺寸的卷积核提取图像特征。SRCNN 是一种典型的端到端的网络模型, 其作者也尝试了更深的 4 层或 5 层的结构, 性能并未有显著提升但训练时间十分长。可能正是因为这一点, SRCNN 提出, 在超分辨率问题中, 网络深度不一定“越深越好”。

在 2015 年的 ILSVRC 分类比赛中, 残差网络<sup>[7]</sup>取得了令人瞩目的成功。通过训练残差, 网络训练时间被大大缩减, 这也使得更深的网络结构成为可能。Kim 等人将残差网络用于图像超分辨率问题, 所提出的深层超分辨率网络 (Very Deep Convolutional Networks, VDSR)<sup>[8]</sup>将网络层数加深到 20 层。在 BSD100<sup>[9]</sup>基准测试集上, 针对分辨率放大 4 倍, VDSR 性能比 SRCNN 平均提高了 0.39dB。此后, 图像超分辨率网络几乎都采用残差训练方法, 网络深度也越来越深, 也似乎越来越表明在该问题上网络“越深越好”。

不可避免地, 随着网络深度的增加, 参数规模也越来越大。因此, 出现了从不同角度加深网络深

度并控制参数规模的方法, 意在以较低代价获得性能增益。Shi 等认为, 在网络学习之前通过双三次插值算法将低分辨率图像插值到目标分辨率, 增加了计算成本。他们提出了一种亚像素卷积神经网络 (Efficient Sub-Pixel Convolutional Neural Network, ESPCN)<sup>[10]</sup>, 将上采样操作设计在网络的末端, 有效减少了参数数量, 也降低了计算复杂度。Lai 等设计了“金字塔”式的 LapSRN<sup>[11]</sup>网络结构, 该结构有三级, 每级完成一次 2 倍的上采样操作。在每级内部, 首先通过若干卷积层来提取特征, 然后通过一层反卷积层将提取的特征尺寸上采样 2 倍。通过连续的三级结构, 最终可得到  $\times 8$  的目标分辨率图像。与 VDSR 与 ESPCN 不同, LapSRN 在网络的中部进行反卷积操作, 由于尺寸是逐步放大, 一定程度上减少了网络参数的数目。结果表明, LapSRN 网络收敛速度较快, 在 BSD100 基准测试集上, 针对分辨率放大 4 倍的情况, LapSRN 与 VDSR 相比, PSNR 性能平均提升了 0.03dB。

上述方法都是通过改变上采样的位置来减少参数数量。Kim 等人另辟蹊径, 通过递归的方式来共享参数, 达到同样的目标。他们所提出深度递归卷积网络 (Deeply-Recursive Convolutional Network, DRCN)<sup>[12]</sup>, 采用链式结构构造了一个 16 层的递归网络。为方便训练, DRCN 使用了递归监督减少参数规模, 并使用跳跃连接加快梯度在网络中的流动, 提升了性能。实验结果表明, 相对于 VDSR, DRCN 在四个通用测试集 Set5<sup>[18]</sup>、Set14<sup>[13]</sup>、BSD100<sup>[9]</sup>和 Urban100<sup>[4]</sup>, 针对分辨率放大 4 倍, 性能提升了  $-0.06\text{dB} \sim 0.18\text{dB}$ , 值得注意的是, 在 BSD100 数据集上, DRCN 性能降低了 0.06dB。进一步地, Tai 等提出深度递归残差网络 (Deep Recursive Residual Network, DRRN)<sup>[14]</sup>, 同样借助于递归学习策略实现权值共享, 并在网络的训练过程中多次利用残差学习加快训练过程。得益于这些策略的使用, DRRN 在提高性能的同时有效的控制了参数数目。DRRN 给出了 20 层和 52 层的实现, 实验表明, 在 Urban100<sup>[4]</sup>基准测试集上, 针对分辨率放大 4 倍, 与 VDSR 相比, 平均分别获得 0.17dB 和 0.24dB 的性能增益。

上述网络从不同角度加深了网络深度, 显著提升了性能。然而, 这些网络普遍都采用卷积层串联的方式, 即每个卷积层仅接受上一层的输出, 没有充分利用各卷积层所提取的分层特征信息。为此,

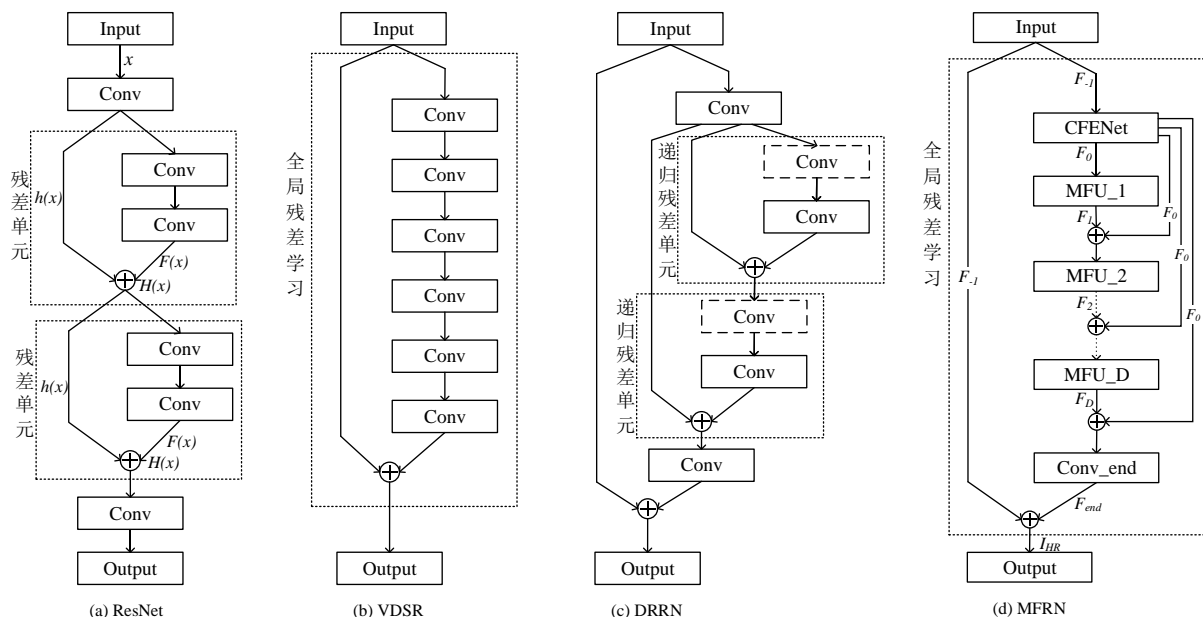


图 1 相关工作及所提出的多阶特征融合的递归网络结构，为了表达简洁，图中并未给出批量数据标准化与激活函数过程。(a) ResNet 网络简化结构，其中虚线框内为一个残差单元；(b) VDSR，基于全局残差进行训练；(c) DRRN，虚线框内为递归残差单元，每个递归残差单元包括两个卷积层，递归单元相同位置的卷积层共享参数。(d) MFRN，基于递归网络架构，由多阶特征融合单元 MFU 组成，MFU 之间共享参数

Fig. 1 The network structure of proposed multi-feature fusion recursive network and state-of-art methods, where the batch normalize(BN) process and activate functions are omitted for simplicity. (a) ResNet: components in the dotted box denote one single residual unit; (b) VDSR: the global residual training is employed; (c) DRRN: components in the dotted box denote a recursive residual unit, which includes 2 convolution layers; convolutional parameters at the same position are shared among untis; (d) MFRN: it is constituted of several MFUs with recursive network structure, and parameters are shared among MFUs

我们提出了递归式多阶特征融合图像超分辨率算法 (Multi-Feature Fusion Recursive Network, MFRN)。MFRN 基于递归网络架构，其基本组成单元为“多阶特征融合单元 (Multi-Feature Fusion Unit, MFU)”。MFU 之间仍采用特征逐级连续传递的方式，并可以共享参数。在 MFU 内部，卷积层提取的多阶特征被进一步级联、融合，提高了网络的拟合精度。

本文的主要贡献包括：

- 1) 基于递归神经网络架构设计算法模型，特征信息在递归的基本单元——多阶特征融合单元间递归传递，形成权值共享，从而有效减少了参数数量。尽管所提出的网络深度多达 37 层，但能够在 230 个 epoch 后收敛，并获得明显性能增益。
- 2) 每个多阶特征融合单元通过连续内存机制<sup>[15]</sup>从前一单元获取状态，然后对本单元内的从低阶到高阶的各级特征进行级联、

融合，以达到对有效信息的精确筛选，从而能够更加深度刻画低分辨率图像与高分辨率图像之间的映射关系。

- 3) 在网络训练过程中，充分利用残差网络的优势。具体地，在整体递归网络中使用全局残差学习策略，在多阶特征融合单元内使用局部残差学习策略。通过结合两种策略，有效降低了深层网络的训练难度，避免了网络退化、梯度弥散等问题。

## 1 超分辨率相关工作

受递归网络、深层网络与残差网络的启发，所提出的模型结合了递归网络架构的优势，进一步加深了网络层数，并借助于残差网络的方式进行训练。下文将对残差网络 ResNet、深层超分辨率网络 VDSR、递归网络 DRRN 进行介绍，三种网络的结构如图 1 所示。

## 1.1 ResNet

对于普通的神经网络，当达到一定深度时，会出现梯度弥散问题，即梯度在反向传播中会逐渐消失，从而导致无法对网络前几层的权重进行调整，使深层网络无法收敛。所以，直接增加网络的深度，往往不能提高网络的性能，还可能引起网络退化问题。为此，He 等人设计了易训练的、非常深的残差网络 ResNet<sup>[7]</sup>。在此之前，神经网络一般拟合的是输入与输出之间的潜在映射，而残差网络通过在残差单元中增加捷径连接，使单元中的卷积层拟合残差映射，如图 1(a)所示。设  $x$  表示输入， $H(x)$  是残差单元的潜在映射，残差映射  $F(x)$  可表示为：

$$\hat{x} = U(x) = \sigma(F(x, W) + h(x)) \quad (1)$$

式中， $\hat{x}$  是残差单元的输出， $h(x)$  是恒等映射： $h(x) = x$ ， $W$  表示权重（为表达简洁，略去了偏置项）， $\sigma$  表示 ReLU 激活函数。实验结果表明，残差网络更方便训练，从而很容易通过增加深度来提高网络性能。例如，He 等<sup>[7]</sup>所设计的网络结构深达 152 层，该网络在每个残差单元都使用了局部残差学习策略，在 ImageNet 分类中，直接将错误率降低到了 3.57%。

## 1.2 VDSR

与残差网络在每个残差单元中都使用局部残差学习不同，VDSR<sup>[8]</sup>引入了全局残差学习策略，即在输入的低分辨率图像和输出的高分辨率图像之间进行全局残差学习。如图 1(b)所示，右侧的残差分支包括 20 层卷积层，并且所有的卷积核尺寸都为  $3 \times 3$ ，网络的感受范围达到  $41 \times 41$ 。可见，深度的增加有效地扩大了网络感受范围，从而能够更好的挖掘图像像素间的关联度，提高网络性能。然而，随着深度的增加，网络的训练愈发困难，为此，VDSR 采用了全局残差学习和自适应梯度裁剪策略来加快训练速度。实验结果表明，在 NVIDIA GeForce GTX TITAN Z 上，VDSR 仅需要训练 4 小时左右就可以收敛。

## 1.3 DRRN

除了给训练带来困难，深度的增加还将导致参数数目增加，进而引起两个问题：一是网络容易过拟合，二是模型大，存储和重现有困难。为解决上述问题，Tai 等人提出了基于递归神经网络共享参数的 DRRN<sup>[14]</sup>结构。如图 1(c)所示，DRRN 中的递

归残差单元都拥有一个共同的输入，每个单元包含 2 个卷积层，单元间对应位置的卷积层共享参数。通过参数共享，DRRN 的参数数量仅为 VDSR 的一半。

综上，上述基于深度神经网络的图像超分辨率算法从训练方法、参数规模等角度提出解决思路，在提升性能的同时，也降低了训练复杂度。在这些算法中，信息的流动都采用卷积层依次串联的方式。然而，不同的卷积层提取了图像不同维度的特征，这些特征对图像超分辨率的恢复都有贡献。本文提出了卷积后各层特征再融合的思想，设计了多阶特征融合单元，以自适应融合所提取的多阶特征。同时，充分利用上述网络结构的局部残差学习、全局残差学习以及递归学习等策略，来加快网络收敛的速度和减少参数数量，最终形成融合多阶特征的递归网络模型。

## 2 多阶特征融合的递归网络模型

### 2.1 网络结构

如图 1(d)所示，所提出的网络包括 3 部分：粗特征提取网络(Coarse Feature Extraction Net, CFENet)、多阶特征融合单元 MFU 以及全局残差学习。定义  $I_{LR}$  和  $I_{HR}$  分别为低分辨率与高分辨率图像。先使用双三次插值算法，将低分辨率图像插值到目标分辨率，作为网络的输入，如公式(2)所示。

$$F_{-1} = H_{bic}(I_{LR}) \quad (2)$$

$F_{-1} = H_{bic}(\cdot)$  为双三次插值函数， $F_{-1}$  为所提取的图像粗特征。因此，可以得到：

$$F_0 = H_{CFEN}(F_{-1}) \quad (3)$$

$H_{CFEN}$  为图像粗特征提取网络，其功能是通过一层卷积操作提取图像的浅层特征来作为多阶特征融合单元的输入。设该网络包括  $D$  个多阶特征融合单元， $F_d$  表示第  $d$  个多阶特征融合单元的输出，如公式(4)所示。

$$F_d = H_{MFU,d}(F_{d-1} + F_0) = H_{MFU,d}(H_{MFU,d-1}(\cdots H_{MFU,2}(F_{-1} + F_0)\cdots)) \quad (4)$$

$H_{MFU,d}$  表示第  $d$  个多阶特征融合单元， $H_{MFU,d}$  是一个复合函数，包括卷积操作和 ReLU 激活函数。

在每个  $H_{MFU,d}$  内, 不同卷积层提取的图像特征被先级联、后融合, 同样, 我们使用递归学习策略, 在 MFU 之间实现权值共享。

经过  $D$  个 MFU, 网络最终获得特征信息  $F_D$ , 利用  $F_{-1}$  与  $F_{end}$  构成全局残差学习架构, 更多细节将在 2.3 节展开。

给定一个训练集  $\{x^{(i)}, \mathbb{X}^{(i)}\}_{i=1}^N$ ,  $N$  表示训练集中训练数据的个数,  $\mathbb{X}^{(i)}$  是真实高分辨率图像,  $x^{(i)}$  是低分辨率图像, 定义损失函数为:

$$L(\theta) = \frac{1}{2N} \sum_{i=1}^N \|\mathbb{X}^{(i)} - I_{HR}(x^{(i)})\|^2 \quad (5)$$

$\theta$  表示权重参数的集合,  $I_{HR}(\cdot)$  为所获得的高分辨率图像, 2.3 节给出了更详细的介绍。通过小批量梯度下降算法和反向传播来优化目标函数方程。

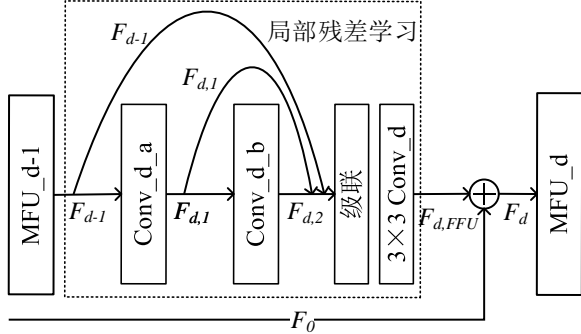


图 2 所提出的多阶特征融合单元 MFU 的内部网络结构  
Fig. 2 Network structure of the proposed multi-feature fusion unit (MFU)

## 2.2 多阶特征融合单元

如图 2 所示, 所提出的多阶特征融合单元 MFU 基于残差学习结构, 将各个卷积层提取的分层特征进行级联、融合, MFU 与 DenseNet 有类似思想, 但不相同, 3.4 节详细对比了 MFU 与稠密网络 DenseNet<sup>[16]</sup>及相关超分辨率网络模型。

定义  $F_{d-1}$  和  $F_d$  分别为多阶特征融合单元的输入和输出, 两者都含有  $Q_0$  个特征映射图。例如, 第  $d$  个多阶特征融合单元的第二个卷积层可表示为:

$$F_{d,2} = \sigma(W_{d,2}[F_{d-1}, F_{d,1}, F_{d,2}]) \quad (6)$$

$\sigma$  为 ReLU 激活函数,  $W_{d,2}$  为第二个卷积层的权重, 为了表达的简洁, 略去了偏置项。设  $F_{d,1}$  和  $F_{d,2}$  分别包含  $q_1$  和  $q_2$  个特征映射图,  $[F_{d-1}, F_{d,1}, F_{d,2}]$  表示特

征映射图的级联操作。因此, 第  $d$  个多阶特征融合单元有  $Q_0 + (q_1 + q_2)$  个特征映射图。

由图 2 可见, 第  $d-1$  个多阶特征融合单元的特征映射图将直接输入至第  $d$  个多阶特征融合单元中进行级联操作。然后, 借助于  $3 \times 3$  的卷积层来自适应筛选特征并控制输出的信息, 如公式(7)所示。

$$F_{d,FFU} = H_{FFU}^d([F_{d-1} + F_{d,1} + F_{d,2}]) \quad (7)$$

式中,  $H_{FFU}^d$  表示在第  $d$  个多阶特征融合单元内使用  $3 \times 3$  的卷积核来实现特征融合。

## 2.3 局部与全局残差学习

在每个多阶特征融合单元内部, 我们使用局部残差学习策略来加快梯度在网络中的流动, 同时也可以进一步改进网络的表达能力和减轻训练一个非常深的网络所带来的梯度消失问题, 从而获得更好的性能, 如公式(8)所示。

$$F_d = F_0 + F_{d,FFU} \quad (8)$$

此外, 整个网络的训练还采用了全局残差策略, 如公式(9)所示。

$$I_{HR} = F_{-1} + F_{end} \quad (9)$$

$F_{-1}$  表示由双三次插值得到的粗特征, 在采用全局残差学习之前, 前向卷积层所提取的所有特征都在 MFU 中被充分结合、筛选与提取, 经过  $D$  个 MFU, 这些多阶特征被自适应地融合并形成  $F_D$ 。  $F_{end}$  由  $F_D$  与粗特征求和再经过卷积得到。经过全局残差学习, 得到了恢复的高分辨率图像  $I_{HR}$ 。

# 3 实验结果及分析

## 3.1 实验条件

本文使用的训练集中共包括 291 张图片, 其中 91 张图片来自 Yang 等人<sup>[17]</sup>提供的数据集, 其余 200 张图像来自伯克利分割数据集<sup>[9]</sup>。测试集选用的是 Set5<sup>[18]</sup>、Set14<sup>[13]</sup>、BSD100<sup>[9]</sup>和 Urban100<sup>[4]</sup>, 这些测试集分别有 5 张、14 张、100 张和 100 张图像。

在训练中, 使用旋转和翻转两种方式来增加训练数据: (1) 旋转, 将每张原始图像旋转  $90^\circ$ 、 $180^\circ$ 、 $270^\circ$ ; (2) 翻转, 将旋转后的图像再进行镜像翻转。经过上述操作, 每一张原始图像新增了 7

个版本，训练集扩充到 2328 张。另外，不同的分辨率放大倍数共享同一个网络模型，即只训练一个模型。网络模型设置小批量数据为 20，动量参数为 0.9，每个卷积层有 128 个卷积核，卷积核尺寸为  $3 \times 3$ ，权值衰减参数为  $10^{-4}$ 。

我们采用与 He 等<sup>[19]</sup>相同的办法进行权重初始化，即将每层的权重从高斯分布中采样进行初始化，高斯分布的均值为 0，实验表明利用这种权重初始化方法配合使用 ReLU 激活函数，训练效果更佳。采用自适应学习率调整策略，初始学习率设为 0.1，设置每 10 个 epoch，学习率减半。由于所设定的初始学习率比较大，在训练中使用自适应梯度衰减策略来加速收敛的速度，同时防止梯度爆炸。也就是说，初始学习率较大会造成梯度较大，将梯度限定在  $[-\alpha/\eta, \alpha/\eta]$  之间，式中  $\eta$  是当前的学习率， $\alpha=0.01$  为梯度衰减参数。该网络基于 Caffe 平台实现，使用 NVIDIA GeForce GTX 1080TI(11G 显存)，对图像亮度分量进行训练。

### 3.2 实验结果比较与分析

为进一步确定多阶特征融合单元 MFU 的个数  $D$ ，我们分别对  $D=7, 9, 13, 20$  进行了实验，它们分别对应 29 层、37 层、53 层、81 层的网络。不

同深度网络的收敛速度与性能对比如图 3 所示。显然，当  $D=9$  时，网络获得最优性能。最终，我们采用 9 个特征融合单元即 37 层网络结构，在 230 个 epoch 后收敛。

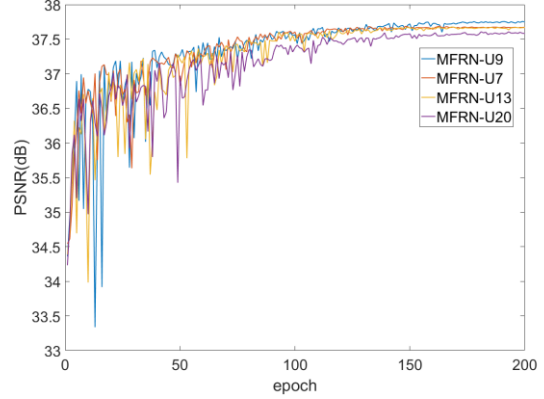


图 3 基于 Set5 数据集，针对分辨率放大 2 倍， $D=7,9,13,20$  所对应的网络在 200 次 epoch 内的收敛速度与性能对比

Fig. 3 Comparison on performance and coverage speed of different networks: scale factor  $\times 2$ ,  $D=7,9,13,20$ , Set5 dataset

表 1 基准测试集下的 PSNR/SSIM 指标对比（加粗为最优性能）

Table 1 Average PSNR/SSIM on benchmark dataset (bold font indicates the best performance)								
Dataset	Scale	Bicubic	SRCNN	VDSR	DRCN	LapSRN	DRRN_B1U9	MFRN
Set5	$\times 2$	33.66/0.9299	36.66/0.9542	37.53/0.9587	37.63/0.9588	37.52/0.959	37.66/0.9589	<b>37.75/0.9595</b>
	$\times 3$	30.39/0.8682	32.75/0.9090	33.66/0.9213	33.82/0.9226	33.82/0.922	33.93/0.9234	<b>33.95/0.9242</b>
	$\times 4$	28.42/0.8104	30.48/0.8628	31.35/0.8838	31.53/0.8854	31.54/0.885	31.58/0.8864	<b>31.62/0.8882</b>
Set14	$\times 2$	30.24/0.8688	32.45/0.9067	33.03/0.9124	33.04/0.9118	33.08/0.913	33.19/0.9133	<b>33.24/0.9139</b>
	$\times 3$	27.55/0.7742	29.30/0.8215	29.77/0.8314	29.76/0.8311	29.87/0.832	29.94/0.8339	<b>29.97/0.8347</b>
	$\times 4$	26.00/0.7027	27.50/0.7513	28.01/0.7674	28.02/0.7670	<b>28.19/0.772</b>	28.18/0.7701	28.17/0.7716
BSD100	$\times 2$	29.56/0.8431	31.36/0.8879	31.90/0.8960	31.85/0.8942	31.80/0.895	32.01/0.8969	<b>32.04/0.8975</b>
	$\times 3$	27.21/0.7385	28.41/0.7863	28.82/0.7976	28.80/0.7963	28.82/0.798	28.91/0.7992	<b>28.93/0.7999</b>
	$\times 4$	25.96/0.6675	26.90/0.7101	27.29/0.7251	27.23/0.7233	27.32/0.727	27.35/0.7262	<b>27.38/0.7280</b>
Urban100	$\times 2$	26.88/0.8403	29.50/0.8946	30.76/0.9140	30.75/0.9133	30.41/0.910	31.02/0.9164	<b>31.16/0.9181</b>
	$\times 3$	24.46/0.7349	26.24/0.7989	27.14/0.8279	27.15/0.8276	27.07/0.828	27.38/0.8331	<b>27.45/0.8356</b>
	$\times 4$	23.14/0.6577	24.52/0.7221	25.18/0.7524	25.14/0.7510	25.21/0.756	25.35/0.7576	<b>25.42/0.7619</b>

表 2 基准测试集下的 IFC 指标对比（加粗为最优性能）

Table 2 Average IFC on benchmark dataset (bold font indicates the best performance)



Dataset	Scale	Bicubic	SRCNN	VDSR	DRCN	LapSRN	DRRN_B1U9	MFRN
Set5	×2	6.083	8.036	8.569	8.326	<b>9.010</b>	8.583	8.753
	×3	3.580	4.658	5.221	5.202	5.194	5.241	<b>5.362</b>
	×4	2.329	2.991	3.547	3.502	3.559	3.581	<b>3.687</b>
Set14	×2	6.105	7.784	8.178	8.025	<b>8.501</b>	8.181	8.357
	×3	3.473	4.338	4.730	4.686	4.662	4.732	<b>4.842</b>
	×4	2.237	2.751	3.133	3.066	3.147	3.147	<b>3.238</b>
BSD100	×2	5.695	7.242	7.169	7.220	<b>7.715</b>	7.43	7.620
	×3	3.168	3.879	4.043	4.070	4.057	4.129	<b>4.222</b>
	×4	1.993	2.412	2.627	2.587	2.677	2.669	<b>2.743</b>
Urban100	×2	6.245	7.989	8.645	8.527	8.907	8.653	<b>8.947</b>
	×3	3.620	4.584	5.194	5.187	5.168	5.259	<b>5.421</b>
	×4	2.361	2.963	3.496	3.412	3.530	3.536	<b>3.673</b>

在评价所重建的高分辨率图像的质量时，与VDSR、DRCN、DRRN等类似，对图像边缘部分进行了裁剪，针对分辨率放大2倍、3倍、4倍的情况，每个边裁分别剪掉2、3、4行像素。

表1对所提出的MFRN网络与双三次插值Bicubic、SRCNN、VDSR、DRCN、LapSRN、DRRN等方法的峰值信噪比PSNR与结构相似性SSIM性能进行了对比。显然地，在所有数据集的不同缩放因子下，MFRN模型的PSNR与SSIM两个指标几乎都优于其它网络模型。尤其是在细节非常丰富的Urban100基准测试集上，MFRN极大的提升了重建图像的质量，针对分辨率放大2倍、3倍、4倍的情况，相较于基础递归网络结构DRRN，分别获得0.08dB, 0.04dB, 0.03dB的增益。值得注意的是，表1给出的LapSRN网络针对不同放大倍数分别训练了不同的模型，在此前提下，LapSRN在Set14上的平均性能比所提出的网络模型高0.02dB。此外，本文还使用了信息保真度准则IFC<sup>[20]</sup>指标对图像质量进行了评价，IFC被誉为与主观评价打分结果最相似的评价标准<sup>[21]</sup>。如表2所示，我们计算了所提出的MFRN的IFC指标，并引用了PSyCo<sup>[22]</sup>和LapSRN作者所给出的关于其它网络的实验结果。通过对比发现，在大部分放大倍数与数据集上，MFRN的IFC性能有明显提

升。同样，在针对不同放大倍数单独训练了模型的前提下，仅在前三个数据集放大2倍的这种情况，LapSRN的性能稍高于MFRN。

各方法所获得的高分辨率图像的主观质量比较如图4所示。可以看出，MFRN能够更好地保留图像的细节。以图4所给出的局部图像为例，MFRN能够重建出相对清晰的边缘，而其他方法得到的结果都过于平滑，线条偏模糊，甚至得到了错误的结果。

本文所提出的MFRN与稠密网络(Dense Convolutional Networks, DenseNet)<sup>[16]</sup>都采用了多维特征图级联的思想。受DenseNet启发，在单张图像超分辨率问题上，先后有基于稠密网络的跳跃连接超分辨率网络(Dense Skip Connections Network, SRDenseNet)<sup>[23]</sup>与残差稠密网络(Residual Dense Network, RDN)<sup>[15]</sup>被提出。SRDenseNet采用DenseNet的稠密块作为网络的基本组成单元，单元之间采用稠密的跳跃方式连接。RDN采用全局与局部稠密连接结合的策略，在每个稠密块中，每一层所提取的特征信息以一种前馈的方式连接到之后的每一层。SRDenseNet与RDN都获得了可观的增益，表明了充分多层次特征图对图像性能提升具有积极作用。



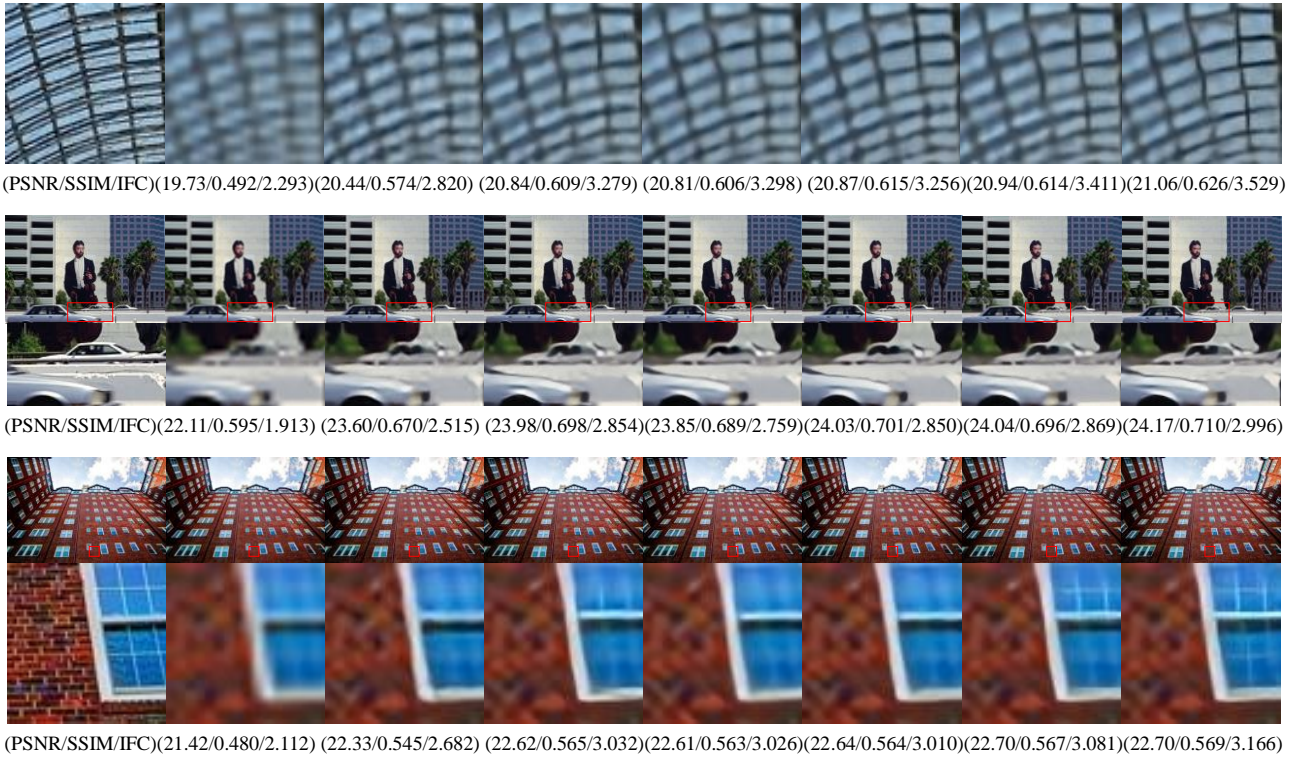


图 4 各方法所获得的高分辨图像的主观质量对比

Fig. 4 The subjective quality comparison of super resolution results from different methods

然而，上述不断迭代的级联将导致通道数与网络参数数目急剧增加，造成瞬时显存的占用率剧增，极易引起显存溢出，从而给硬件设备带来极大挑战。虽然本文所提出的 MFRN 同样也利用了网络不同卷积层所提取的图像多层次特征，但更注重对参数规模的控制，是一种“全局稀疏，局部稠密”的网络。其特点体现在：(1) 在 MFU 内部采用稠密增量学习机制，借助于局部残差，在已有的特征图基础上继续追加学习新特征；(2) MFU 之间不存在稠密连接，而是一种稀疏连接，后续 MFU 只需要在输入特征图的基础上，继续学习高维度特征。(3) MFU 之间共享参数，网络深度的增加并不影响参数规模，从而有效降低了算法的硬件需求，同等环境下能够加快图像处理速度。

图 5 给出了各种方法的性能与参数对比，尽管 RDN 达到了所列举方法中的最优性能，但其参数数量高达 1500 万，是 MFRN 的 8.8 倍。由于 SRDenseNet 并未给出基于 Urban100 数据集，针对分辨率放大 2 倍的性能，因此在图中未绘制。但经过统计，SRDenseNet 的参数数目达 430 万，是 MFRN 的 2.4 倍。借助于递归网络的优势，MFRN 实现了

参数共享，参数数目为 170 万，在同等规模参数的网络中，获得了最优性能。

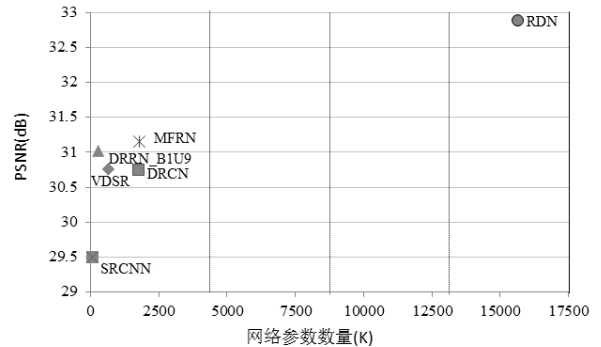


图 5 基于 Urban100 数据集，针对分辨率放大 2 倍，各方法的性能与参数数量对比

Fig. 5 Comparison on performance and number of parameters of different methods: scale factor  $\times 2$ , Urban100 dataset

## 4 结论

本文提出了递归式多阶特征融合图像超分辨率算法 MFRN。MFRN 采用“全局稀疏，局部稠密”的结构，由多阶特征融合单元 MFU 组成，并通过递

归学习，形成 MFU 间权值共享，从而有效减少网络的参数数量。在每个 MFU 内，采用稠密连接，从低阶到高阶的各级特征被级联、融合，并对特征信息自适应的筛选，从而能够更加精准的刻画低分辨率图像和高分辨率图像之间的映射。在 MFU 之间，采用稀疏连接，后续 MFU 只需要在输入特征图的基础上，继续提取高维度特征。在训练过程中，整体递归网络使用全局残差学习策略，多阶特征融合单元使用局部残差学习策略，两种策略的结合有效降低了训练深层网络的难度。实验结果表明，与同等参数规模的几种典型网络相比，MFRN 能够达到最优性能。在主观质量上，MFRN 对细节的处理效果尤为明显，图像质量得到显著改善。

### 参考文献

- [1] Shi W, Caballero J, Ledig C, et al. Cardiac Image Super-resolution With Global Correspondence Using Multi-atlas Patchmatch[M]// Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013. Springer Berlin Heidelberg, 2013:9-16.
- [2] M. W. Thornton, P. M. Atkinson, D. A. Holland. Sub - pixel mapping of rural land cover objects from fine spatial resolution satellite sensor imagery using super - resolution pixel - swapping[J]. International Journal of Remote Sensing, 2006, 27(3):473-491.
- [3] Wilman, W. W. Z, and P. C. Yuen. "Very low resolution face recognition problem." Fourth IEEE International Conference on Biometrics: Theory Applications and Systems IEEE, 2010:1-6., Yuen P C. Very low resolution face recognition problem[C]// Fourth IEEE International Conference on Biometrics: Theory Applications and Systems. IEEE, 2010:1-6.
- [4] Huang J B, Singh A, Ahuja N. Single image super-resolution from transformed self-exemplars[C]// Computer Vision and Pattern Recognition. IEEE, 2015:5197-5206.
- [5] Schuler S, Leistner C, Bischof H. Fast and accurate image upscaling with super-resolution forests[C]// Computer Vision and Pattern Recognition. IEEE, 2015:3791-3799.
- [6] Dong C, Chen C L, Tang X. Accelerating the super-resolution convolutional neural network[C]// European Conference on Computer Vision. Springer, Cham, 2016:391-407.
- [7] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[J]. // Conference on Computer Vision and Pattern Recognition (CVPR) 2015:770-778.
- [8] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2016:1646-1654.
- [9] 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]// Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on. IEEE, 2002:416-423 vol.2.
- [10] Shi W, Caballero J, Huszar F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[J]. //2016:1874-1883.
- [11] Lai W S, Huang J B, Ahuja N, et al. Deep laplacian pyramid networks for fast and accurate super-resolution[J]. 2017.
- [12] Kim J, Lee J K, Lee K M. Deeply-recursive convolutional network for image super-resolution[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2016:1637-1645.
- [13] Roman Zeyde, Michael Elad, Matan Protter. On single image scale-up using sparse-representations[C]// International Conference on Curves and Surfaces. Springer-Verlag, 2012:711-730.
- [14] Tai Y, Yang J, Liu X. Image super-resolution via deep recursive residual network[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2017:2790-2798.
- [15] Zhang Y, Tian Y, Kong, et al. Residual dense network for image super-resolution[C]// IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018:2472-2481
- [16] Gao Huang, Zhuang Liu, Laurens van der Maaten, et al. Densely connected convolutional networks[J]. 2016.
- [17] Yang J, Wright J, Huang T S, et al. Image super-resolution via sparse representation.[J]. IEEE Transactions on Image Processing, 2010, 19(11):2861-2873.
- [18] Bevilacqua M, Roumy A, Guillemot C, et al. Low-complexity single image super-resolution based on nonnegative neighbor embedding[J]. Bmvc, 2012.
- [19] He K, Zhang X, Ren S, et al. Delving deep into rectifiers: surpassing human-level performance on imageNet classification[J]. 2015:1026-1034.
- [20] Sheikh H R, Bovik A C, Veciana G D. An information fidelity criterion for image quality assessment using natural scene statistics[J]. IEEE Transactions on Image Processing, 2005, 14(12):2117-2128.
- [21] Yang C Y, Ma C, Yang M H. Single-image super-resolution: a benchmark[C]// European Conference on Computer Vision. Springer, Cham, 2014:372-386.

- [22] Pérez-Pellitero E, Salvador J, Ruiz-Hidalgo J, et al. Psycho: manifold span reduction for super resolution[C]// Computer Vision and Pattern Recognition. IEEE, 2016:1837-1845.
- [23] Tong T, Li G, Liu X, et al. Image super-resolution using dense skip connections[C]// IEEE International Conference on Computer Vision. IEEE Computer Society, 2017:4809-4817.

#### 第一作者简介:



佟骏超, 1994 年生,男,硕士研究生,主要研究方向为图像处理、视频压缩。

E-mail:771034446@qq.com

#### 通信作者:



丁丹丹, 女, 讲师, 主要研究方向为图像处理、视频压缩与编码,

E-mail:DandanDing@hznu.edu.cn

#### 其他作者简介:

费加罗, 男, 硕士研究生, 主要研究方向为图像处理, E-mail:398080890@qq.com

陈靖森, 男, 本科生, 主要研究方向为图像处理, E-mail:empcjs@gmail.com

李恒, 男, 本科生, 主要研究方向为图像处理, E-mail:poptreehengli@gmail.com