

# 基于递归残差网络的图像超分辨率重建

周登文<sup>1</sup> 赵丽娟<sup>1</sup> 段然<sup>1</sup> 柴晓亮<sup>1</sup>

**摘要** 深度卷积神经网络在单图像超分辨率重建方面取得了卓越成就,但其良好表现通常以巨大的参数数量为代价.本文提出一种简洁紧凑型递归残差网络结构,该网络通过局部残差学习减轻训练深层网络的困难,引入递归结构保证增加深度的同时控制模型参数数量,采用可调梯度裁剪方法防止产生梯度消失/梯度爆炸,使用反卷积层在网络末端直接上采样图像到超分辨率输出图像.基准测试表明,本文在重建出同等质量超分辨率图像的前提下,参数数量及计算复杂度分别仅为 VDSR 方法的 1/10 和  $1/(2n^2)$ .

**关键词** 递归结构, 残差学习, 卷积神经网络, 深度学习, 超分辨率

**引用格式** 周登文, 赵丽娟, 段然, 柴晓亮. 基于递归残差网络的图像超分辨率重建. 自动化学报, 2019, 45(6): 1157–1165

**DOI** 10.16383/j.aas.c180334

## Image Super-resolution Based on Recursive Residual Networks

ZHOU Deng-Wen<sup>1</sup> ZHAO Li-Juan<sup>1</sup> DUAN Ran<sup>1</sup> CHAI Xiao-Liang<sup>1</sup>

**Abstract** Despite the great success in single image super-resolution reconstruction achieved by deep convolutional neural network, the number of the computational parameters is often very large. This paper proposes a concise and compact recursive residual network. The local residual learning method is adopted to mitigate the difficulty of training very deep network, the recursive structure is configured to control the number of model parameters while increasing the model depth, the adjustable gradient clipping strategy is applied to prevent the gradient disappearance/gradient explosion, and a deconvolutional layer is set to directly up sample the image to a super-resolution image at the end of the residual network. According to benchmark tests, in the premise that the same quality super-resolution image is reconstructed, the number of parameters and the computational complexity of the proposed method are reduced to about 1/10 and  $1/(2n^2)$  of VDSR, respectively.

**Key words** Recursive structure, residual learning, convolutional neural network, deep learning, super-resolution

**Citation** Zhou Deng-Wen, Zhao Li-Juan, Duan Ran, Chai Xiao-Liang. Image super-resolution based on recursive residual networks. *Acta Automatica Sinica*, 2019, 45(6): 1157–1165

单图像超分辨率 (Single image super-resolution, SISR) 重建是一种经典的计算机视觉问题,旨在从一个给定的低分辨率 (Low-resolution, LR) 图像中恢复高分辨率 (High-resolution, HR) 图像. 由于 SISR 恢复了高频信息,因此被广泛应用于需要更多图像细节的领域,如医疗成像<sup>[1]</sup>、卫星成像<sup>[2]</sup>、安全监控<sup>[3]</sup> 等.

现有超分辨率 (Super-resolution, SR) 图像重建技术主要分为三大类: 基于插值的 SR<sup>[4]</sup> 技术, 基于重建的 SR<sup>[5]</sup> 技术和基于学习的 SR<sup>[6–16]</sup> 技术. 目前的 SR 算法大多基于学习 (或基于图像

块), 通过学习 LR 图像和 HR 图像之间的映射来重建 SR 图像. 作为 SISR 的先驱卷积神经网络 (Convolutional neural network, CNN)<sup>[17]</sup> 模型, 超分辨率卷积神经网络 (SRCNN)<sup>[6]</sup> 可以用端到端的方式学习 LR/HR 间的非线性映射, 其性能显著优于传统的非深度学习 (Deep learning, DL) 方法. 之后, Dong 等在文献 [7] 中又提出加速超分辨率卷积神经网络 (FSRCNN), 旨在加速 SRCNN. 该模型直接从原始 LR 图像 (无插值) 学习映射到 HR 图像, 大大减小了计算复杂度. 然而, 以上两种模型均使用浅层网络直接学习原始映射函数, 图像恢复质量不佳. 2016 年, Kim 等提出了基于 CNN 的 SISR 算法—VDSR<sup>[8]</sup>. 其网络有 20 个卷积层, 引入了全局残差学习, 并以较高的学习率训练模型, SISR 性能得到较大幅度提升. 但其需要在 CNN 学习之前通过双三次插值对 LR 图像进行上采样, 计算成本较高, 内存占用较大.

近两年来, SISR 方法大多通过加深网络提升性能<sup>[9–11]</sup>. 在网络层数加深的同时, 参数数量、内存空

收稿日期 2018-05-22 录用日期 2018-11-05  
Manuscript received May 22, 2018; accepted November 5, 2018  
北京市自然科学基金 (4162056) 和中央高校基本科研业务费专项资金 (2018ZD06) 资助  
Supported by Beijing Natural Science Foundation (4162056) and the Fundamental Research Funds for the Central Universities (2018ZD06)  
本文责任编辑 王亮  
Recommended by Associate Editor WANG Liang  
1. 华北电力大学控制与计算机工程学院 北京 102206  
1. School of Control and Computer Engineering, North China Electric Power University, Beijing 102206

间和计算复杂度也随之增加. 受计算能力、内存空间和功耗等方面的限制, 这些方法在移动或嵌入式等资源受限的设备中实用价值不大. 本文基于 VDSR 方法, 构建出一种新型递归残差网络结构 (RRSR), 旨在不损失图像恢复质量的前提下, 构建结构更紧凑、模型参数更少、计算复杂度更低的网络模型, 提升 SISR 方法在资源受限的移动或嵌入式等设备上的实用价值. 具体来说, 本文方法主要有以下几点贡献:

1) RRSR 引入局部残差学习传递更多图像信息. 在 VDSR 中, 残差图像是从网络的输入和输出估计的, 称为全局残差学习. 此外, 如视觉识别<sup>[18]</sup> 和图像恢复<sup>[19]</sup> 中观察到的那样, 非常深的网络可能会遇到性能退化问题, 其原因在于图像经过多层传输后, 会丢失大量的细节信息. 为解决该问题, RRSR 引入局部残差学习的增强型残差块结构, 其恒等分支不仅能够将深层图像细节传送到后层, 还有助于梯度流动.

2) RRSR 采用残差块的递归结构减少参数数量. RRSR 还将递归结构引入到残差块中, 构成一个递归残差网络, 并且权重集合在这些残差块之间共享, 参数数量大大减少, 模型结构更加紧凑.

3) RRSR 通过反卷积操作降低计算复杂度. 计算复杂度与输入图像的大小成正比, 本文模型使用原始未插值的 LR 图像作为输入, 而 VDSR 则是将插值后的 LR 图像作为输入, 其大小约为 RRSR 的  $n^2$  倍. 同时, 我们使用的网络层数更少, 经计算, VDSR 的计算复杂度约为 RRSR 的  $2n^2$  倍.

## 1 相关工作

### 1.1 SRCNN

SRCNN 共由三部分组成: 特征提取、特征的非线性映射、SR 图像重建. SRCNN 将插值后的 LR 图像作为输入, 直接输出 SR 图像, 从而证明了 CNN 可以用端到端的方式学习图像间的非线性映射. 虽然图像性能比传统的非 DL 方法有了很大的提升, 但其层数较浅, 学习能力有限, 不能重建出高质量的 SISR 图像.

### 1.2 FSRCNN

FSRCNN 的提出旨在加速 SRCNN, 该方法重新设计了 SRCNN 网络, 提出一个紧凑的沙漏形 CNN 结构. 该模型在网络末端引入反卷积层, 并重新制定了映射层, 在映射之前先缩小输入特征维度, 映射之后再扩大特征维度, 使模型提速 40 多倍, 甚至图像有更好的恢复质量. 受 FSRCNN 启发, 所提方法同样使用反卷积层对图像进行上采样, 但性能比 FSRCNN 有显著提升.

### 1.3 ResNet

残差网络 (ResNet) 的主要思想是根据输入学习残差函数而非原始函数, 这使得深层网络的训练更加简单, 而且能由更深的网络获得更好的性能.

将输入表示为  $x$ , 底层映射表示为  $H(x)$ , 残差映射定义为:

$$F(x) := H(x) - x \quad (1)$$

则残差块结构如下:

$$\hat{x} = U(x) = \sigma(F(x, W) + h(x)) \quad (2)$$

其中,  $\hat{x}$  是残差块的输出,  $h(x)$  是一个恒等映射<sup>[18]</sup> 且  $h(x) = x$ ,  $W$  是一组权重 (省略偏差以简化符号),  $F(x, W)$  是待学习的残差映射, 函数  $\sigma$  表示 Relu<sup>[20]</sup> 激活函数,  $U$  表示残差块函数. 对于堆叠两个卷积层的基本残差块,  $F(x, W) = W_2\sigma(W_1x)$ .

残差映射比原始映射更容易优化. 在极端情况下, 如果某个恒等映射是最优的, 那么将残差变为 0 比用非线性层的堆叠来拟合恒等映射更为简单. 公式  $F(x) + x$  可以通过前馈神经网络的短连接来实现, 恒等的短连接并不增加额外的参数和计算复杂度. 残差块结构如图 1 所示.

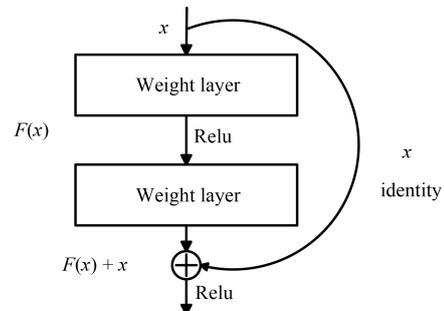


图 1 残差块结构<sup>[18]</sup>

Fig. 1 Residual block structure<sup>[18]</sup>

残差网络的提出, 打破了加深网络层数不能提升性能的说法. 然而, 原始 ResNet 被提出来是用于解决更高层次的计算机视觉问题, 如图像分类和目标检测. 将 ResNet 架构直接应用于图像超分辨率重建的低级视觉问题可能不是最佳模型. 本文方法借鉴残差学习的思想, 使网络能够进行局部残差学习, 但所提方法采用了一种新的残差块结构, 与 ResNet 中的链模式不同.

### 1.4 VDSR

VDSR 使用非常深的 CNN (20 层卷积层), 使得精度显著提升. 然而, 训练非常深的网络, 收敛速度是关键. 若仅通过提高学习率来加快收敛速度, 可能会导致梯度消失/梯度爆炸. 为解决该问题, Kim

等没有采用 ResNet 中的归一初始化和中间归一化, 而是提出一种可调整的梯度裁剪方法.

梯度裁剪技术通常用于训练循环神经网络<sup>[21]</sup>, 其在 CNN 中的用法十分有限, 常用的策略之一就是 将单个梯度裁剪到预定义的范围  $[-\theta, \theta]$ . CNN 通常使用随机梯度下降方法训练网络, 用梯度乘上学习率以调整步长来更新梯度. 如果使用高学习率, 则需要将  $\theta$  调整到一个很小的值以避免产生梯度爆炸. 但是随着学习率变小, 有效梯度 (梯度乘以学习率) 接近于零, 如果学习率以几何速率降低, 训练则会以指数级迭代次数收敛. 对于最大收敛速度, VDSR 将梯度裁剪为  $[-\theta/\gamma, \theta/\gamma]$ , 其中  $\gamma$  表示当前学习速率<sup>[8]</sup>. 可调梯度裁剪能够使网络快速收敛, 本文方法同样采用可调梯度裁剪来防止梯度消失/梯度爆炸.

VDSR 借鉴 ResNet 思想, 在网络中引入了全局残差学习, 即插值后的 LR 输入图像与 HR 输出图像之间的残差学习. 与 VDSR 不同的是, RRSR 采用局部残差学习, 即在网络内部堆叠多个残差块, 使网络能够学到更多有用的信息.

VDSR 在 CNN 学习之前要通过双三次插值增加 LR 图像的分辨率, 但同时也增加了计算成本. 尽管 VDSR 生成图像的质量有显著提升, 但是深层网络结构需要庞大的参数数量, 相较于紧凑模型, 大模型需要更多的存储空间.

### 2 本文方法

本文方法旨在不损失 VDSR 图像恢复质量的前提下, 改进 VDSR 模型, 使其网络结构更加简洁紧凑. VDSR 共 20 层卷积层, 通过在深层网络中多次级联小型滤波器, 使得精度显著提升. 然而, 深层网络结构会产生大量参数, 占用过多存储空间. 并且, 该模型是将插值后的 LR 图像作为输入, 其大小为原始 LR 图像的  $n^2$  倍 ( $n$  为放大因子), 而计算复杂度与输入图像的大小成正比, 该操作使计算复杂度呈指数级增长.

针对上述问题, 本文提出一种递归残差网络. 该模型使用局部残差学习而非 VDSR 所用全局残差学习来训练深层网络, 更有助于信息传输和梯度流动; 在残差块中引入递归结构, 使得参数减少, 模型更加紧凑; 将未插值的 LR 图像作为输入, 最后在网络末端使用反卷积层直接上采样到 SR 输出图像, 使得计算复杂度大大降低. 如图 2 所示, 网络整体由三部分组成: 特征提取、残差函数的非线性映射、SR 图像重建.

图 3 示出本文模型的具体网络结构. LR 输入图像经过一层卷积层和一层 Relu 层提取特征, 然后将提取到的特征输入到若干残差块中, 递归学习残差映射函数. 最后, 在网络末端使用反卷积层直接对

学到的残差图像进行上采样, 重建出 SR 输出图像.

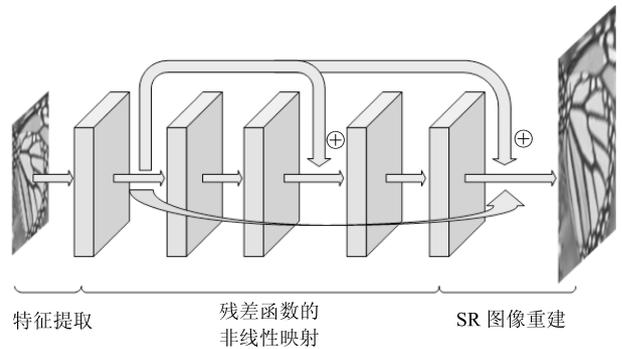


图 2 网络结构示意图

Fig. 2 Network structure diagram

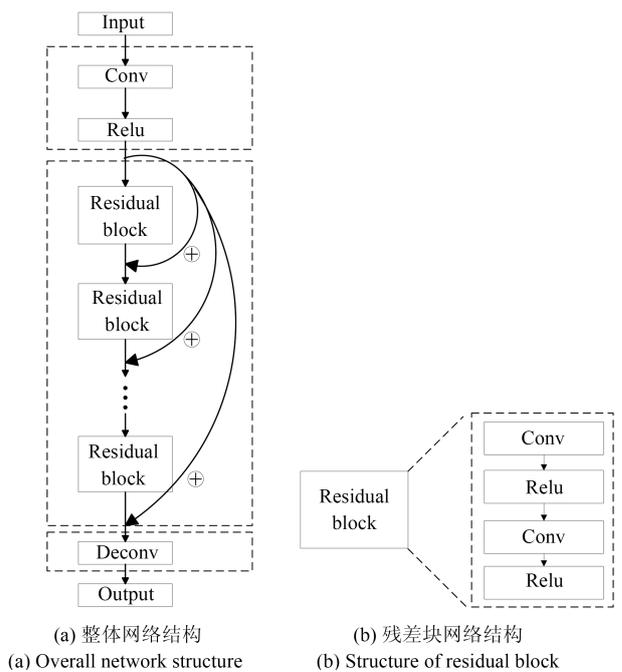


图 3 RRSR 具体网络结构图

Fig. 3 The specific network structure of RRSR

第 2.1 ~ 2.3 节分别详细介绍了局部残差学习、递归结构和反卷积层; 第 2.4 节探讨控制网络结构的两大参数对网络性能的影响.

#### 2.1 局部残差学习

最近, 残差网络 ResNet 在低级到高级计算机视觉任务中均表现出优秀的性能. 所提方法采用 ResNet 中的残差学习思想以及跳跃连接, 构成增强型残差块结构. 与 ResNet 的残差块不同, RRSR 先通过卷积层提取 LR 输入图像的高频特征, 之后每经过两层卷积层, 都加上第一层卷积层提取的特征图像, 即残差块中所有恒等分支的输入均保持相同. 该方法不仅能够向网络深层传递更多图像信息, 其恒等分支还有助于训练期间梯度的反向传播, 避免

出现过拟合现象<sup>[22]</sup>. 本文将残差块函数定义为:

$$H^u = R(H^{u-1}) = F(H^{u-1}, W) + H^0 \quad (3)$$

其中,  $H^u$  是第  $u$  个残差块的输出结果,  $R$  表示残差块函数,  $F(H^{u-1}, W)$  是待学习的残差映射,  $W$  为权重集合,  $H^0$  是经过第一层卷积层输出的特征图像.

### 2.2 递归结构

近年来, 各种 SISR 方法大都通过加深网络层数提高性能. 但是, 随着层数的增加, 参数数量也在增大, 导致计算量和存储空间需求的增加. 为了控制模型参数, 本文将递归结构引入到残差块中, 使得权重在残差块之间共享, 并且模型更加紧凑<sup>[23]</sup>. 图 4 给出递归块的基本结构: 在递归块的开头引入卷积层和 Relu 层, 然后叠加多个残差块,  $H^0$  为每个残差块的恒等映射,  $H^u$  为经过第  $u$  个残差块的输出结果, 由此形成递归结构.  $B$  表示递归结构中包含的残差块数.

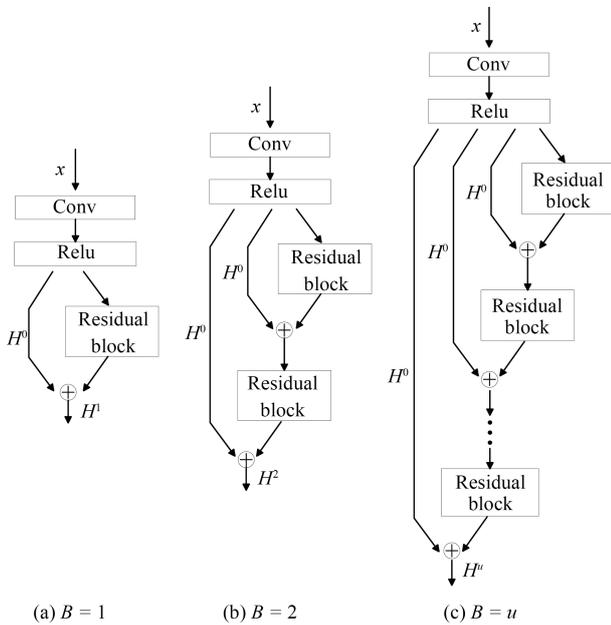


图 4 递归块结构

Fig. 4 Recursive block structure

根据式 (3), 可得第  $u$  个残差块的结果为:

$$H^u = R^{(u)}(H^0) = R(R(\dots(R(H^0))\dots)) \quad (4)$$

通过式 (4) 可以看出, 第  $u$  个残差块的结果可由残差块函数  $R$  递归得到.

### 2.3 反卷积层

RRSR 的网络末端为反卷积层, 该层使用一组反卷积滤波器对输出的残差图像进行上采样. 不同

于传统的插值方法, 反卷积是可以学习到图像特征的上采样内核, 可被视为卷积的逆运算. 对于卷积操作, 当滤波器移动步幅为  $k$  时, 输出为输入的  $1/k$ ; 反之, 反卷积操作的输出则为输入的  $k$  倍. 当  $k$  等于放大因子  $n$  时, 通过反卷积层可直接输出所需尺寸的 SR 图像, 并降低网络计算复杂度.

RRSR 及 VDSR 的计算复杂度可分别用式 (5) 和式 (6) 进行计算:

$$O\{(f_1^2 m_1 + m_1 f_2^2 m_2 + \dots + m_{L-2} f_{L-1}^2 m_{L-1} + m_{L-1} f_L^2 m_L \cdot n^2) \cdot S_{\text{input}}\} \quad (5)$$

$$O\{(f_1^2 m_1 + m_1 f_2^2 m_2 + m_2 f_3^2 m_3 + \dots + m_{L-1} f_L^2 m_L) \cdot S_{\text{input}}\} \quad (6)$$

其中,  $f_i$  ( $i = 1, 2, \dots, L$ ) 表示第  $i$  层滤波器的大小,  $m_i$  ( $i = 1, 2, \dots, L$ ) 表示第  $i$  层滤波器的个数,  $S_{\text{input}}$  是输入图像的大小,  $n$  为放大因子.

本文方法卷积层的参数设置与 VDSR 相同, 均使用 64 个大小为  $3 \times 3$  的滤波器; 最后一层 RRSR 使用 1 个大小为  $5 \times 5$  的滤波器, VDSR 使用 1 个大小为  $3 \times 3$  的滤波器. 由式 (5) 及式 (6) 可知, 网络计算复杂度与输入图像的大小成正比, VDSR 输入图像的大小约为本文方法输入图像的  $n^2$  倍. 此外, VDSR 网络共有 20 层, 而本文方法只有 12 层, 故 VDSR 的计算复杂度约为本文方法的  $2n^2$  倍.

### 2.4 U 和 C 的设置

本节将探讨各种  $U$ 、 $C$  组合所构成不同结构的递归残差网络对模型性能的影响, 其中  $U$  代表残差块数,  $C$  代表每个残差块中的卷积层数. 如图 5 所示, 本文在 6 到 20 层的深度范围内对几个不同网络结构进行采样, 测试结果均为各个网络的  $\times 3$  模型在 Set5 测试集上的平均峰值信噪比 (Peak signal to noise ratio, PSNR) 值.

为清楚地展示单个参数对 RRSR 网络性能的影响, 本文采用控制变量法进行分析. 在图 5 中, 曲线代表具有相同网络层数的不同网络结构, 曲线上的数字表示层数, 点的颜色对应右侧相应的 PSNR 值. 当  $C$  为 2 时, 对比  $U3C2$ 、 $U4C2$  和  $U5C2$  的网络性能, 其 PSNR 值随着  $U$  的增大而增加; 当  $U$  为 3 时, 对比  $U3C2$ 、 $U3C3$  和  $U3C4$  的网络性能, 其 PSNR 值随着  $C$  的增大而增加. 以上结果表明网络性能的提升可能与  $U$  和  $C$  成正相关, 但这两种情况随着  $U$ 、 $C$  的增加, 网络层数也在增加. 为进一步验证网络层数对模型性能的影响, 将网络层数固定为某一值, 以 14 层网络为例, 对比  $U6C2$ 、 $U4C3$  和  $U3C4$  网络的 PSNR, 发现 PSNR 值随着残差块数量的增加而增加. 并且由于递归网络结构的特点, 当  $C = 2$  时使用参数数量最少. 因此将  $C$  固定为 2,

通过改变  $U$  来选择最佳参数. 据图 5, 当  $U$  增加到 5 时, 网络性能达到饱和, 故本文将 12 层的  $U5C2$  网络作为最终模型.

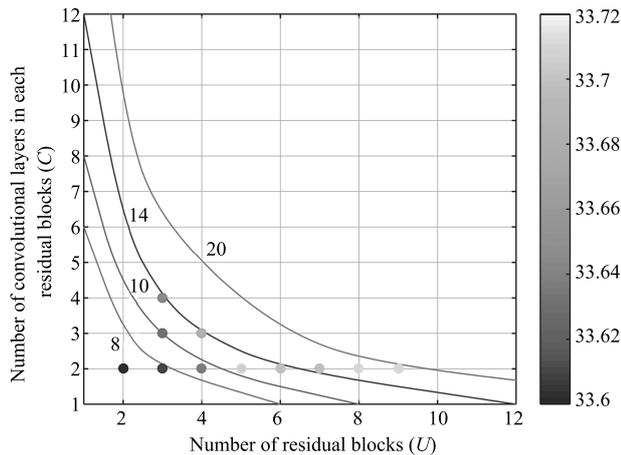


图 5 各种  $U$ 、 $C$  组合所构成网络的性能对比图

Fig. 5 The performance of various networks at  $U$  and  $C$  combinations

### 3 实验细节

本节将对实验部分进行详细说明, 并展示所提方法与多种 SISR 方法的比较结果.

#### 3.1 数据集

##### 3.1.1 训练集

不同的基于学习的 SISR 方法使用不同的训练集. SRCNN 使用 ImageNet 训练集. FSRCNN 使用 General-100 训练集, 其中为 100 幅无压缩 bmp 格式图像, 并联合使用 Yang 等<sup>[24]</sup> 的 91 幅图像作为训练图像. VDSR 使用文献 [25] 中的 291 幅图像作为训练集, 该训练集在 Yang 等<sup>[24]</sup> 的 91 幅图像集中又添加了 Berkeley Segmentation Dataset<sup>[26]</sup> 中的 200 幅图像. 为充分利用训练集, VDSR 通过缩放和旋转的方式对训练集进行数据增强. 为将本文方法与 VDSR 进行客观公正的比较, 本文采用与 VDSR<sup>[8]</sup> 相同的训练集, 并进行相同的数据增强.

##### 3.1.2 测试集

测试阶段采用三个广泛使用的基准测试集 Set5<sup>[27]</sup>、Set14<sup>[28]</sup> 和 BSD100<sup>[26]</sup> 进行测试, 它们分别包含 5 幅、14 幅和 100 幅图像.

#### 3.2 训练细节

##### 3.2.1 训练样本

对于输入图像, 首先采用放大因子  $n$  对原始训练图像进行下采样, 成为 LR 图像. 然后将 LR 图像裁剪成一组具有步幅  $k$ , 大小为  $f_{\text{sub}} \times f_{\text{sub}}$  像素的子图像, 并从对应的真实图像中裁剪出相应的大小为

$(nf_{\text{sub}})^2$  像素的 HR 子图像, 这些 LR/HR 子图像对即为训练样本. 为保证映射过程中图像大小不变, 卷积层均采用“0”填充. 影响子图像大小的另一因素是反卷积层. 在用 Caffe 包<sup>[29]</sup> 训练 RRSR 时, 反卷积滤波器将生成大小为  $(nf_{\text{sub}} - n + 1)^2$  而不是  $(nf_{\text{sub}})^2$  的输出图像. 因此, 还需在 HR 子图像上裁剪  $(n-1)$  像素的边界. 最后, 对于  $\times 2$ ,  $\times 3$  和  $\times 4$  模型, 分别将 LR/HR 子图像对的大小设为  $25^2/49^2$ ,  $17^2/49^2$  和  $13^2/49^2$ .

##### 3.2.2 参数设置

本文所提网络深度为 12 层, 对于每层卷积层, 都使用 64 个大小为  $3 \times 3$  的滤波器; 对于反卷积层, 滤波器的大小为  $5 \times 5$ , 步幅等于放大因子  $n$ . 训练时, 图像批处理的大小为 128, 动量为 0.9, 权重衰减参数为 0.0001. 初始学习率设定为 0.1, 然后每隔 15 代学习率减半, 120 代后学习停止, 损失函数为均方误差 (Mean square error, MSE) 函数. 本文实验环境: Caffe 框架, MATLAB (R2015b) 平台和一个 NVIDIA GTX Titan X GPU (12 GB 内存).

##### 3.2.3 训练策略

关于权重初始化问题, 所提方法卷积层采用 He 等<sup>[30]</sup> 提出的“msra”方法进行初始化; 反卷积层则使用均值为 0, 标准差为 0.001 的高斯分布随机初始化. 将 Relu 函数作为激活函数. 训练阶段, 首先训练  $\times 3$  模型直至饱和. 然后, 使用  $\times 3$  模型训练好的参数初始化  $\times 2$  和  $\times 4$  模型的权重, 并进行微调, 收敛速度远比重新训练要快.

#### 3.3 模型分析

为了说明局部残差学习和递归结构的有效性, 本文以 12 层 RRSR $\times 3$  模型为模型 I, 然后在其基础上通过删除或改变跳跃连接方式, 构造网络结构不同但深度相同的模型 II 和模型 III, 并用 Set5 测试集验证其性能, 测试结果见表 1. 其中, 模型 II 是将模型 I 的残差函数学习改为原始映射函数学习, 此模型类似于 FSRCNN<sup>[7]</sup>, 没有跳跃连接, 因此也就不存在递归结构; 模型 III 是在模型 II 的基础上根据本文所提跳跃连接方式增加跳跃连接, 构成分别包含卷积层数为 2-4-4 的残差块, 此模型仍然没有递归结构. 由表 1 可知, 局部残差学习能够向网络深处传递更多有效图像信息, 学到更多图像特征, 使得图像恢复质量有较大幅度提升. 递归结构的加入, 不仅使参数数量大幅降低, 同时还提升了图像的恢复质量, 网络结构更加紧凑, 性能更优.

表 1 不同 RRSR 组件构成的  $\times 3$  模型在 Set5 测试集上的平均 PSNR 值及参数数量

Table 1 Average PSNR and number of parameters when different RRSR components are turned on or off, for scale factor  $\times 3$  on testset Set5

局部残差	递归结构	PSNR (dB)	参数数量 ( $\times 10^3$ )
×	×	33.27	371
✓	×	33.58	371
✓	✓	33.70	39

### 3.4 与先进方法的比较

近年来非 DL 方法<sup>[23, 26, 31]</sup>和 DL 方法<sup>[6, 7, 13, 32, 33]</sup>通常用于基准测试, 本节与这些 SISR 方法进行了定量定性的比较。

表 2 提供了本文方法与 Bicubic、SelfEx<sup>[31]</sup>、SRCNN<sup>[12]</sup>、FSRCNN<sup>[7]</sup>和 VDSR<sup>[8]</sup>方法的定量评估总结。其中, 文献 [8, 12, 31] 的结果从文献 [8] 中引用, FSRCNN 的结果来自我们的重新测试, 代码为作者公开发布的源代码。由表 2 可知, 本文方法生成图像的平均 PSNR 值和结构相似性 (Structure similarity index measurement, SSIM) 与 VDSR 不相上下, 但比其他 SISR 方法重建图像的质量有大幅提升。同时, 图 6 示出, 在此种情况下, 本文方法使用参数更少。

图 7 和图 8 展示了上述各种方法的定性评估结果。通过对图像细节的放大, 能够从视觉效果上直观地评估各种 SISR 方法重建图像的质量。在图 7 和图 8 中, 第一行图像为不同 SISR 方法对同一幅图像的重建结果; 第二行为对应方法重建图像中矩形框标注区域的放大图像; 第三行为对应方法重建图像的 PSNR 值与 SSIM 值。图 7 示出, 对于桥上木板的重建, 除了 VDSR, 其他方法严重模糊, 甚至条

纹产生畸变, 而本文方法的恢复图像中木板缝隙清晰、条纹平行, 相较之前方法图像的恢复质量, 本文方法有大幅提升。类似地, 在图 8 中, 本文方法完美地重建了蝴蝶翅膀上的花纹, 而其他方法生成的图像有明显的振铃现象, 且边缘比较模糊。

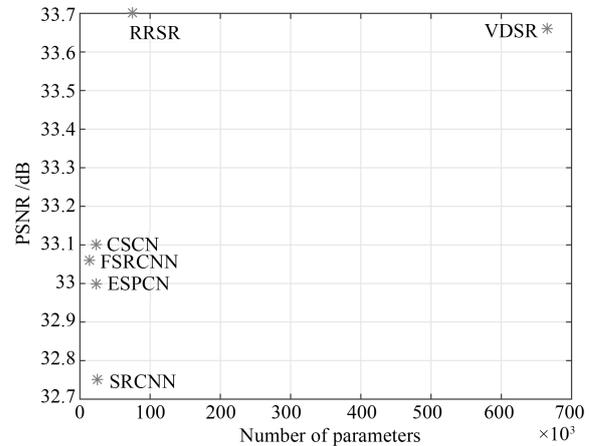


图 6 各种 SISR 方法的  $\times 3$  模型在 Set5 测试集上的平均 PSNR 值及参数数量

Fig. 6 Average PSNR and number of parameters on the testset Set5 for scale factor  $\times 3$  of various SISR methods

图 9 示出各种 SISR 方法  $\times 4$  模型在测试集 Set14 上的平均运行时间及平均 PSNR 值。由于 SRCNN 和 FSRCNN 的测试代码基于 CPU 实现, 为保证比较的公平性, Lai 等<sup>[10]</sup>在 MatConvNet 工具<sup>[34]</sup>中使用相同的网络权重对其进行重建, 以测量 GPU 上的运行时间。本文使用与图 9 中其他方法相同的测试环境, 即配有 NVIDIA Titan X GPU (12 GB 内存) 的机器重新测试 VDSR 及 RRSR 方法的运行时间。由图 9 可知, 所提 RRSR 方法运行速度比其他 SISR 方法都快。

表 2 各种 SISR 方法的  $\times 2$ ,  $\times 3$  和  $\times 4$  模型在测试集 Set5、Set14 和 BSD100 上的平均 PSNR 值与 SSIM 值

Table 2 Average PSNR/SSIMs of various SISR methods for scale factor  $\times 2$ ,  $\times 3$  and  $\times 4$  on Set5, Set14 and BSD100

测试集	放大倍数	Bicubic	SelfEx	SRCNN	FSRCNN	VDSR	RRSR
		PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Set5	$\times 2$	33.66/0.9299	36.49/0.9537	36.66/0.9542	37.00/0.9558	37.53/0.9587	37.55/0.9588
Set5	$\times 3$	30.39/0.8682	32.58/0.9093	32.75/0.9090	33.16/0.9140	33.66/0.9213	33.70/0.9208
Set5	$\times 4$	28.42/0.8104	30.31/0.8619	30.48/0.8628	30.71/0.8657	31.35/0.8838	31.32/0.8836
Set14	$\times 2$	30.24/0.8688	32.22/0.9034	32.42/0.9063	32.64/0.9088	33.03/0.9124	33.04/0.9125
Set14	$\times 3$	27.55/0.7742	29.16/0.8196	29.28/0.8209	29.43/0.8242	29.77/0.8314	29.75/0.8307
Set14	$\times 4$	26.00/0.7027	27.40/0.7518	27.49/0.7503	27.60/0.7535	28.01/0.7674	28.00/0.7675
BSD100	$\times 2$	29.56/0.8431	31.18/0.8855	31.36/0.8879	31.51/0.8906	31.90/0.8960	31.91/0.8961
BSD100	$\times 3$	27.21/0.7385	28.29/0.7840	28.41/0.7863	28.52/0.7897	28.82/0.7976	28.78/0.7969
BSD100	$\times 4$	25.96/0.6675	26.84/0.7106	26.90/0.7101	26.97/0.7128	27.29/0.7251	27.25/0.7249

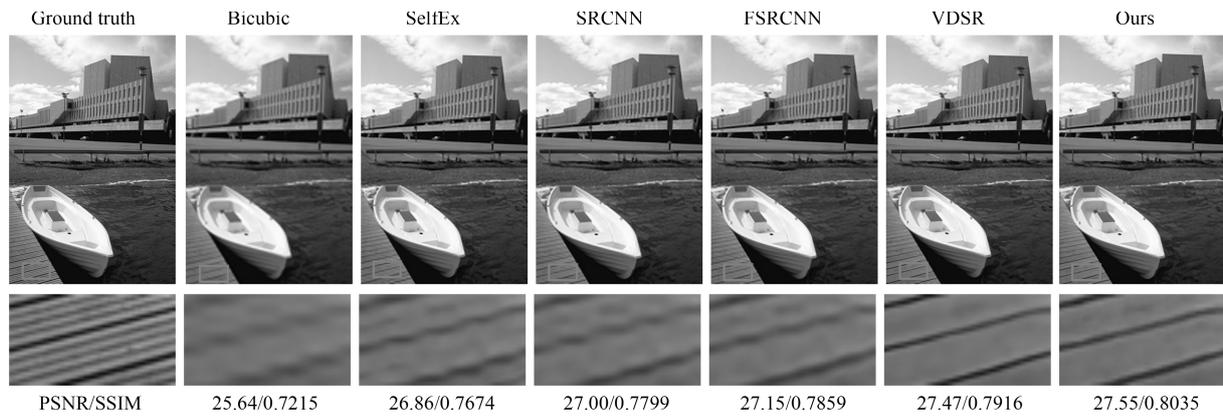


图 7 测试集 BSD100 中的“img\_092”重建结果对比图

Fig. 7 A comparison of the reconstruction results of “img\_092” in the testset BSD100

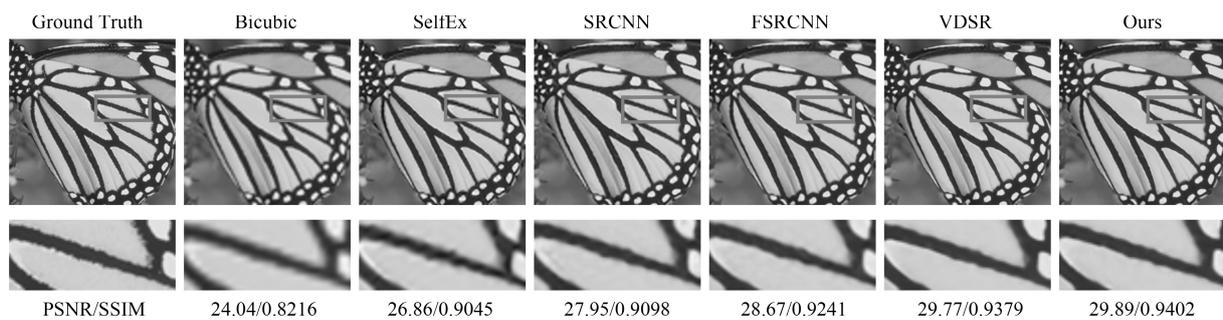
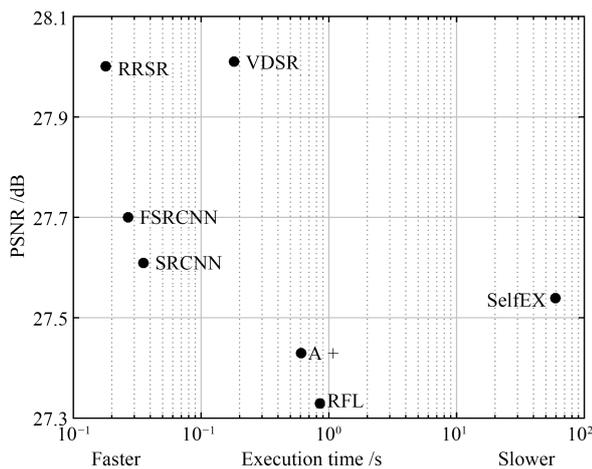


图 8 测试集 Set5 中的“butterfly”重建对比图

Fig. 8 A comparison of the reconstruction results of “butterfly” in the testset Set5

图 9 各种 SISR 方法的  $\times 4$  模型在测试集 Set14 上的平均运行时间及平均 PSNR 值<sup>[10]</sup>Fig. 9 Speed and average PSNR of various SISR methods on the Set14 with scale factor  $\times 4$ <sup>[10]</sup>

## 4 结论

本文提出一种在残差块中引入递归结构的新型残差网络, 所采用跳跃连接和递归结构可有效减轻网络携带特征信息的负担, 实现高质量超分辨率图像的重建. 与 VDSR 方法相比, 本文在重建出同等

质量超分辨率图像的前提下, 网络结构更紧凑、模型参数更少、计算复杂度更低, 在存储资源和计算能力有限的移动设备上具有潜在实用价值. 后续将进一步研究如何使用同一网络生成多尺度超分辨率图像, 以及设计更紧凑的网络结构, 重建更高质量的超分辨率图像.

## References

- 1 Oktay O, Bai W, Lee M, Guerrero R, Kamnitsas K, Caballero J, et al. Multi-input cardiac image super-resolution using convolutional neural networks. In: Proceedings of the 2016 International Conference on Medical Image Computing and Computer-assisted Intervention. Athens, Greece: Springer, Cham, 2016. 246–254
- 2 Luo Y, Zhou L, Wang S, Wang Z Y. Video satellite imagery super resolution via convolutional neural networks. *IEEE Geoscience and Remote Sensing Letters*, 2017, 14(12): 2398–2402
- 3 Rasti P, Uiboupin T, Escalera S, Anbarjafari G. Convolutional neural network super resolution for face recognition in surveillance monitoring. In: Proceedings of the 2016 International Conference on Articulated Motion and Deformable Objects. Palma de Mallorca, Spain: Springer, Cham, 2016. 175–184
- 4 Lu Zhi-Fang, Zhong Bao-Jiang. Image interpolation with predicted gradients. *Acta Automatica Sinica*, 2018, 44(6):

- 1072–1085  
(陆志芳, 钟宝江. 基于预测梯度的图像插值算法. 自动化学报, 2018, **44**(6): 1072–1085)
- 5 Xiong Jiao-Jiao, Lu Hong-Yang, Zhang Ming-Hui, Liu Qie-Gen. Convolutional sparse coding in gradient domain for MRI reconstruction. *Acta Automatica Sinica*, 2017, **43**(10): 1841–1849  
(熊娇娇, 卢红阳, 张明辉, 刘且根. 基于梯度域的卷积稀疏编码磁共振成像重建. 自动化学报, 2017, **43**(10): 1841–1849)
- 6 Dong C, Chen C L, He K M, Tang X O. Learning a deep convolutional network for image super-resolution. In: Proceedings of the 13th European Conference on Computer Vision, Zurich Switzerland: Springer, Cham, 2014. 184–199
- 7 Dong C, Chen C L, Tang X O. Accelerating the super-resolution convolutional neural network. In: Proceedings of the 14th European Conference on Computer Vision. Amsterdam, The Netherlands: Springer, Cham, 2016. 391–407
- 8 Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016. 1646–1654
- 9 Lim B, Son S, Kim H, Nah S, Lee K M. Enhanced deep residual networks for single image super-resolution. In: Proceedings of the 2017 IEEE Computer Vision and Pattern Recognition Workshops. Honolulu, USA: IEEE, 2017. 1132–1140
- 10 Lai W S, Huang J B, Ahuja N, Yang M H. Deep laplacian pyramid networks for fast and accurate super-resolution. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017. 5835–5843
- 11 Ledig C, Theis L, Huszar F, Caballero J, Cunningham A, Acosta A, et al. Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017. 105–114
- 12 Dong C, Chen C L, He K M, Tang X O. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, **38**(2): 295–307
- 13 Timofte R, De Smet V, Van Gool L. A+: Adjusted anchored neighborhood regression for fast super-resolution. In: Proceedings of the 2015 Asian Conference on Computer Vision. Singapore, Singapore: Springer, Cham. 2015. 111–126
- 14 Hu Chang-Sheng, Zhan Shu, Wu Cong-Zhong. Image super-resolution based on deep learning features. *Acta Automatica Sinica*, 2017, **43**(5): 814–821  
(胡长胜, 詹曙, 吴丛中. 基于深度特征学习的图像超分辨率重建. 自动化学报, 2017, **43**(5): 814–821)
- 15 Kim J, Lee J K, Lee K M. Deeply-recursive convolutional network for image super-resolution. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, Nevada, USA: IEEE, 2016. 1637–1645
- 16 Yang W H, Feng J S, Xie G S, Liu J Y, Guo Z M, Yan S C. Video super-resolution based on spatial-temporal recurrent residual networks. *Computer Vision and Image Understanding*, 2018, **168**: 79–92
- 17 LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, **86**(11): 2278–2324
- 18 He K M, Zhang X Y, Ren S Q, Sun J. Deep residual learning for image recognition. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, Nevada, USA: IEEE, 2016. 770–778
- 19 Mao X J, Shen C H, Yang Y B. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain: NIPS, 2016. 2810–2818
- 20 Nair V, Hinton G E. Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning. Haifa, Israel: ICML, 2010. 807–814
- 21 Pascanu R, Mikolov T, Bengio Y. On the difficulty of training recurrent neural networks. In: Proceedings of the 30th International Conference on Machine Learning. Atlanta, USA: ICML, 2013, **28**(3): 1310–1318
- 22 Perez-Pellitero E, Salvador J, Ruiz-Hidalgo J, Rosenhahn B. PSyCo: Manifold span reduction for super resolution. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016. 1837–1845
- 23 Huang Y, Wang W, Wang L. Video super-resolution via bidirectional recurrent convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, **40**(4): 1015–1028
- 24 Yang J C, Wright J, Huang T S, Ma Y. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 2010, **19**(11): 2861–2873
- 25 Schuler S, Leistner C, Bischof H. Fast and accurate image upscaling with super-resolution forests. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE, 2015. 3791–3799
- 26 Martin D, Fowlkes C, Tal D, Malik J. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proceedings of the 2002 International Conference on Computer Vision. Vancouver, BC, Canada: IEEE, 2002. 416–423
- 27 Bevilacqua M, Roumy A, Guillemot C, Alberi Morel M L. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In: Proceedings of the 23rd British Machine Vision Conference. Surrey UK: BMVA Press, 2012. 1–10
- 28 Zeyde R, Elad M, Protter M. On single image scaleup using sparse-representations. In: Proceedings of the 2012 International Conference on Curves and Surfaces. Avignon, France: Springer, Berlin, Heidelberg, 2012. 711–730

- 29 Jia Y Q, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, et al. Caffe: convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM International Conference on Multimedia. New York, USA: ACM, 2014. 675–678
- 30 He K M, Zhang X Y, Ren S Q, Sun J. Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: Proceedings of the 2015 International Conference on Computer Vision. Santiago, Chile: IEEE, 2015. 1026–1034
- 31 Huang J B, Singh A, Ahuja N. Single image super-resolution from transformed self-exemplars. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE, 2015. 5197–5206
- 32 Shi W Z, Caballero J, Huszar F, Totz J, P Aitken A, Bishop R, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016. 1874–1883
- 33 Wang Z W, Liu D, Yang J C, Han W, Huang T. Deep networks for image super-resolution with sparse prior. In: Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago, Chile: OALib Journal, 2015. 370–378
- 34 Vedaldi A, Lenc K. MatConvNet: convolutional neural networks for MATLAB. In: Proceedings of the 23rd ACM International Conference on Multimedia. New York, USA: ACM, 2015. 689–692



**周登文** 华北电力大学控制与计算机工程学院教授. 主要研究方向为图像去噪, 图像去马赛克, 图像插值和图像超分辨率. 本文通信作者.

E-mail: zdw@ncepu.edu.cn

(**ZHOU Deng-Wen** Professor at the School of Control and Computer Engineering, North China Electric Power University. His research interest covers image denoising, image demosaicing, image interpolation and image super-resolution. Corresponding author of this paper.)

His research interest covers image denoising, image demosaicing, image interpolation and image super-resolution. Corresponding author of this paper.)



**赵丽娟** 华北电力大学控制与计算机工程学院硕士研究生. 2015 年获得邢台学院数学与信息技术学院学士学位. 主要研究方向为计算机视觉和深度学习.

E-mail: zhaolj97@163.com

(**ZHAO Li-Juan** Master student at the School of Control and Computer Engineering, North China Electric Power University. She received her bachelor degree from the School of Mathematics and Information Technology, Xingtai University in 2015. Her research interest covers computer vision and deep learning.)



**段然** 华北电力大学控制与计算机工程学院硕士研究生. 2016 年获得北京理工大学计算机学院学士学位. 主要研究方向为计算机视觉和深度学习.

E-mail: 1162227075@ncepu.edu.cn

(**DUAN Ran** Master student at the School of Control and Computer Engineering, North China Electric Power University. She received her bachelor degree from the School of Computing, Beijing Institute of Technology in 2016. Her research interest covers computer vision and deep learning.)



**柴晓亮** 华北电力大学控制与计算机工程学院硕士研究生. 2016 年获得平顶山学院信息工程学院学士学位. 主要研究方向为图像处理和图像超分辨率.

E-mail: 13051603700@163.com

(**CHAI Xiao-Liang** Master student at the School of Control and Computer Engineering, North China Electric Power University. He received his bachelor degree from the School of Information Engineering, Pingdingshan University in 2016. His research interest covers image processing and image super-resolution.)