

# 图像超分辨率重建研究综述

唐艳秋<sup>1</sup>, 潘 泓<sup>1</sup>, 朱亚平<sup>2</sup>, 李新德<sup>1</sup>

(1. 东南大学自动化学院, 江苏南京 210096; 2. 中国传媒大学信息与通信工程学院, 北京 100024)

**摘 要:** 图像超分辨率重建(Super-resolution Reconstruction, SR)是由一张或多张低分辨率图像得到高分辨率图像的过程. 近年来, SR 技术不断发展, 在许多领域被广泛应用. 本文在回顾 SR 技术发展历史的基础上, 全面综述了 SR 技术各个时期的代表性方法, 重点介绍了基于深度学习的图像超分辨率工作. 我们从模型类型、网络结构、信息传递方式等方面对各种算法进行了详细评述, 并对比了其优缺点. 最后探讨了图像超分辨率技术未来的发展方向.

**关键词:** 图像超分辨率; 深度学习; 图像处理; 方法综述

**中图分类号:** TP391 **文献标识码:** A **文章编号:** 0372-2112 (2020)07-1407-14

**电子学报 URL:** <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2020.07.022

## A Survey of Image Super-Resolution Reconstruction

TANG Yan-qiu<sup>1</sup>, PAN Hong<sup>1</sup>, ZHU Ya-ping<sup>2</sup>, LI Xin-de<sup>1</sup>

(1. School of Automation, Southeast University, Nanjing, Jiangsu 210096, China;

2. School of Information and Communication Engineering, Communication University of China, Beijing 100024, China)

**Abstract:** Image super-resolution reconstruction (SR) aims to obtain high-resolution images from one or more low-resolution images. Recently, SR has been developing and widely applied in different fields. This survey retrospects the history of SR technique and provides a comprehensive overview of representative SR methods, with an emphasis on recent deep learning-based approaches. We elaborate the details of various deep learning-based SR methods, including their strengths and weakness, in terms of the deep learning model, architecture, and message pass. Finally, we discuss the possible research directions on SR technique.

**Key words:** image super-resolution reconstruction; deep learning; image processing; methods review

## 1 引言

图像超分辨率重建(Super-resolution Reconstruction, SR)是指采用图像处理和机器学习技术,从同一场景中已有的一张或多张低分辨率(Low-Resolution, LR)图像构建高分辨率(High-Resolution, HR)图像的技术<sup>[1]</sup>. 图像SR在医学影像分析<sup>[2]</sup>、安全监控<sup>[3]</sup>、遥感成像<sup>[4]</sup>等领域有广泛应用. 图像SR最早在1964年和1968年分别由Harris<sup>[5]</sup>和Goodman<sup>[6]</sup>提出,他们的方法被称为Harris-Goodman频谱外推法. 1984年, Tsai和Huang<sup>[7]</sup>等人提出在频域中由LR图像序列复原HR图像的方法后, SR技术逐渐引起人们注意并被广泛研究. 从Tsai和Huang开创性工作以来,人们陆续提出了各种SR技术,包括频域到空域的方法,以及从信号处理到机器学习

的方法. 2014年左右,随着深度学习的兴起,人们将深度学习技术用于SR问题,提出了一系列基于深度学习的SR方法,逐步提升了SR算法的性能. 图1展示了图像SR研究工作的发展路线.

传统图像SR算法包括基于重构和基于样例学习的方法,这些方法普遍难以恢复高频细节信息,导致重建图像模糊且计算复杂、实时性不高等问题,不适用于大放大因子(如 $\times 4$ 、 $\times 8$ )的要求. 为解决这些问题,近年来人们将深度学习用于图像SR中. 如今,基于深度学习的图像SR已逐渐成为主流方法.

随着对SR问题研究的深入,大量方法不断涌现. 一些学者相继对该领域工作进行了综述. 早期综述文章<sup>[8-10]</sup>主要总结了传统SR算法,综述了常用SR算法及当时的研究成果,由于深度学习尚未用于SR领域,

收稿日期:2019-05-13;修回日期:2019-11-20;责任编辑:孙瑶

基金项目:国家自然科学基金(No. 61671151, No. 61573097, No. 91748106);江苏省自然科学基金(No. BK20181265);流程工业综合自动化国家重点实验室基金(No. PAL-N201704);中国传媒大学优秀博导组项目(No. CUC2019A009);中国传媒大学重大攻关培育项目-媒介事件中的AI新闻生产系统与关键技术(No. CUC19ZD003)

因此这些综述没有涉及基于深度学习的 SR 算法. 近期工作<sup>[11~14]</sup>对基于深度学习的 SR 技术进行了综述, 但涉及算法较少, 分类不够全面. 本文在上述工作基础上, 对图像 SR 研究的发展进行了系统的总结(整体框架见图

2). 在传统方法基础上, 重点补充了最新基于深度学习图像 SR 技术的研究成果, 希望为相关领域的后续研究提供有益参考.

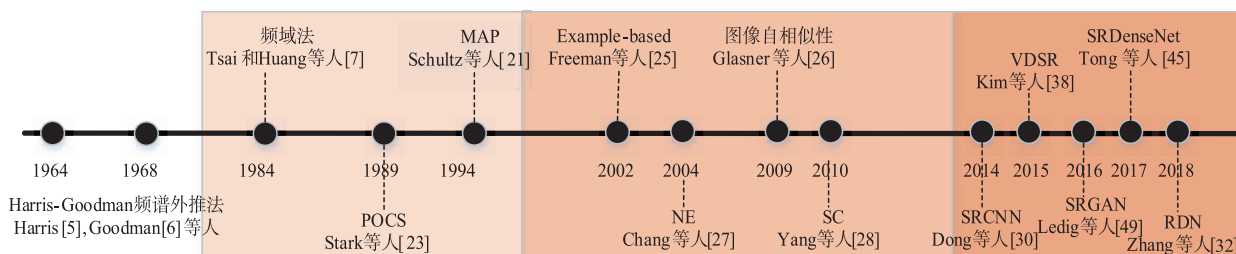


图1 图像SR方法的发展历程

## 2 图像 SR 的主要方法

如图 2 所示, 按算法原理, 我们将基于图像的 SR 方法分为基于重构的方法、基于样例学习的方法和基于深度学习的方法三大类. 目前, 基于深度学习的 SR 算法是该领域的研究热点.

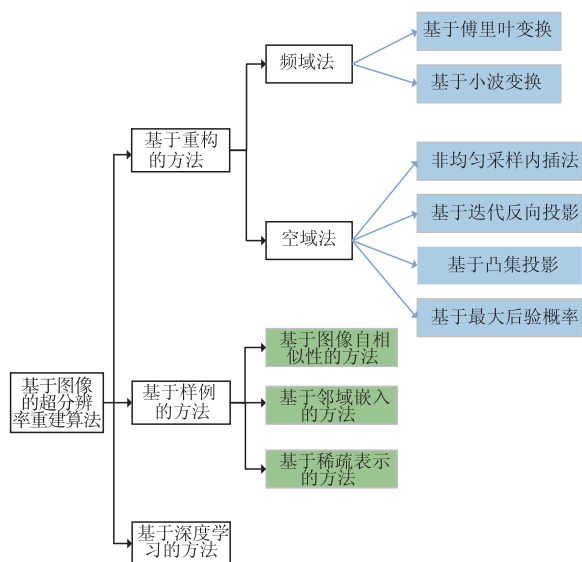


图2 基于图像的超分辨率重建方法分类

### 2.1 基于重构的方法

基于重构的图像 SR 通常要对 LR 图像序列进行亚像素精度对齐, 得到高低分辨率图像间的运动偏移量, 构建观测模型中的空间运动参数, 并应用不同的约束条件求解 HR 图像. 基于重构的图像 SR 方法主要分为频域法和空域法两大类.

#### 2.1.1 频域法

1984 年, Tsai 和 Huang<sup>[7]</sup>首次提出了在频域内提高图像分辨率的思想, 在假定 LR 图像生成模型后, 分别对 LR 图像和原始 HR 图像进行傅里叶变换, 在频域中建立起二者之间的线性关系, 重建出 HR 图像. 该方法

假设图像中不存在运动模糊和观测噪声, 同时忽略了光学系统的点扩散函数, 因此只适合理想图像退化模型. 后续改进算法分别采用递归最小二乘法<sup>[15]</sup>、离散 DCT 变换<sup>[16]</sup>和小波变换<sup>[17,18]</sup>消除图像中的观测噪声、空间模糊和相对物体运动, 有效提高了重建图像质量, 并加快了算法速度.

频域重建方法简单, 运行速度快, 但该方法处理复杂退化模型的能力有限且难以加入先验知识.

#### 2.1.2 空域法

空域法对影响图像成像效果的空域因素(如光学模糊、运动模糊等)建模. 因此, 更接近实际应用情况. 常用的空域 SR 方法主要包括非均匀采样内插法(NUI)<sup>[19]</sup>、基于迭代反投影(IBM)的方法<sup>[20]</sup>、基于最大后验概率(MAP)的方法<sup>[21,22]</sup>、基于凸集投影(POCS)的方法<sup>[23]</sup>和基于 MAP-POCS 的混合算法<sup>[24]</sup>等.

### 2.2 基于样例学习的方法

基于样例学习的方法利用先验知识, 通过给定的样例训练得到 LR 与 HR 图像之间的映射关系, 进行 HR 图像的重建. 该方法包括训练集构建、特征学习和高频细节重建等三个步骤. 训练集构建通过对 HR 图像进行降采样等降质操作得到对应的 LR 图像; 特征学习则学习 LR→HR 图像的映射关系; 高频细节重建根据学习到的 LR→HR 图像映射关系, 从 LR 图像中恢复出 HR 图像. 根据训练样本库来源及匹配特征域的不同, 可分为基于图像自相似性、基于邻域嵌入和基于稀疏表示的方法. 在基于图像自相似性的方法中, 训练样本集来自输入图像本身. 在基于邻域嵌入和基于稀疏表示的方法中, 训练样本集来自外部数据库, 与输入图像无关.

#### 2.2.1 基于图像自相似性的方法

该方法假设图像局部内容及结构在不同尺度间或同一尺度内反复出现. 训练集样本通常由对输入图像做多尺度变换得到的不同分辨率下的图像组成.

Freeman 等人<sup>[25]</sup>认为两个相似的 LR 图像块对应

的 HR 图像块也应相似,两个相邻 HR 图像块重叠部分的像素值应该一致. 他们使用马尔科夫网络分别对 HR、LR 图像对之间,相邻 HR 图像块之间进行建模,对输入的 LR 图像搜索拼接得到对应的 HR 图像. Glasner 等人<sup>[26]</sup>将同一尺度下图像自相似性和跨尺度的图像相似性相结合,通过搜索一粘贴的方式逐级放大图像,填充高频细节,实现高分辨率重建.

### 2.2.2 基于邻域嵌入的方法

该方法<sup>[27]</sup>假设 HR 图像与其对应的 LR 图像块在特征空间中具有相似的局部流形,并用样例图像块对输入图像进行拟合,通过样例 LR 图像块训练得到拟合系数,并将该拟合系数用于 HR 图像重建. 利用邻域嵌入结合其它约束条件,如最小二乘算法,局部线性嵌入和非负最小二乘法得到一组权值,使得 LR 图像与其在样本中对应的  $K$  个最近邻图像块加权得到的 LR 图像误差最小,对 HR 图像块应用这组系数进行加权即可得到重建的 HR 图像.

邻域嵌入法对样本集依赖性小,缺点在于最近邻图像块数目  $K$  需人为选定.

### 2.2.3 基于稀疏表示的方法

基于稀疏表示的方法<sup>[28,29]</sup>通过稀疏分解将图像变换到稀疏域,并假设样本库中 LR-HR 图像对共享同一稀疏系数. 对输入的 LR 图像利用 LR 字典计算其稀疏系数,将同样的系数应用于 HR 字典上生成 HR 特征,得到最终的重建 HR 图像. 稀疏表示法克服了邻域嵌入法需要确定最近邻图像块数目的缺点,可扩展性强,对噪声不敏感,但重建图像质量和计算复杂度均与字典大小相关. 当字典不够完备时,图像边缘细节较模糊.

## 3 基于深度学习的图像 SR

2014 年 Dong 等人<sup>[30]</sup>提出一个含三个卷积层的网络 - SRCNN (Super-Resolution CNN),首次将深度学习用于图像 SR. 此后,基于深度学习的 SR 网络不断出现并改进. 如图 3 所示,我们根据网络类型将基于深度学习的图像 SR 方法分为基于卷积神经网络 (CNN) 的 SR 方法和基于其它深度学习模型的 SR 方法. 其中使用 CNN 结构的 SR 方法,可进一步分为基于 SRCNN 及其改进算法的网络、基于先验知识的网络、基于残差结构的网络和基于密集连接的网络. 有些网络如 MemNet<sup>[31]</sup>、RDN<sup>[32]</sup> 则同时借鉴了残差结构和密集连接,而基于残差结构的 SR 方法中又包含了基于生成对抗网络 (Generative Adversarial Net, GAN) 的 SR 方法.

### 3.1 基于 SRCNN 及其改进算法

SRCNN<sup>[30]</sup>借鉴了基于稀疏编码的图像 SR 思想,通过卷积操作完成图像块提取、特征非线性映射和重建过程三个步骤,其网络结构如图 4 所示. 首先通过双三

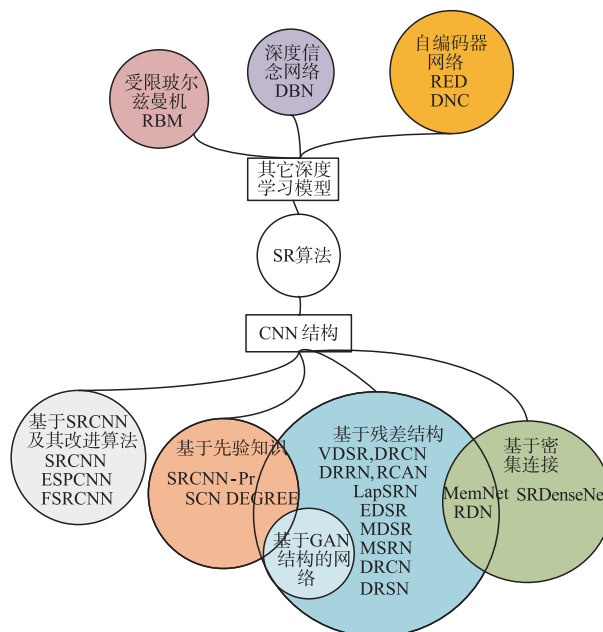


图3 基于深度学习的图像超分辨率重建方法分类

次线性插值将 LR 图像插值到 HR 图像尺寸作为输入. 第一层卷积对图像密集采样,得到一系列图像块,每个图像块表示成一个高维向量,这些高维向量组成  $n_1$  个特征图;第二层卷积将表示 LR 图像的  $n_1$  个特征图非线性映射到表示 HR 图像的  $n_2$  个特征图;第三层卷积采用全连接方式将预测的 HR 图像块整合起来生成一张 HR 图像.

SRCNN 将真实 HR 图像和重建 HR 图像间的均方误差 (MSE) 作为损失函数 (式 (1) 所示) 训练网络

$$L(\theta) = \frac{1}{n} \sum_{i=1}^n \|F(Y_i; \theta) - X_i\|^2 \quad (1)$$

式中,  $L(\theta)$  为损失函数,  $\theta$  为网络优化参数,  $n$  为训练样本数量,  $Y_i$  为训练集中第  $i$  幅 LR 图像,  $F(Y_i; \theta)$  为由  $Y_i$  映射得到的 HR 图像,  $X_i$  为第  $i$  幅真实 HR 图像.

虽然 SRCNN 在图像重建质量和速度上均优于传统方法,但存在以下问题:①没有利用任何先验知识;②只适用于单放大因子的图像 SR,针对不同的放大因子,网络模型需要重新训练;③由于输入需要先插值至目标尺寸大小,整个图像重建过程均在 HR 空间中进行,导致计算量大,训练速度慢;④网络比较浅,结构简单,卷积核感受野较小 ( $13 \times 13$ ) 不能很好地利用图像上下文信息.

针对 SRCNN 需要将 LR 图像先放大至 HR 空间学习 LR→HR 特征映射,造成计算量大、重建时间长的缺点,ESPCNN (Efficient Sub-Pixel CNN)<sup>[33]</sup> 和 FSRCNN (Fast Super-Resolution by CNN)<sup>[34]</sup> 算法直接将 LR 图像送入网络提取特征,并学习 HR 特征图. 不同之处在于,ESPCNN 利用增加卷积核数量和亚像素卷积得到 HR

特征图,并通过通道重组得到 HR 图像.而 FSRCNN 使用更小的卷积核和更多的映射层改变特征维数,实现多通道非线性映射和特征融合,最后通过反卷积层得到 HR 图像.实验结果表明,ESPCNN 能取得与 SRCNN 相当的重建质量,FSRCNN 的重建质量较 SRCNN 略好.但两种算法的重建时间仅为 SRCNN 的 1/10 (ESPCNN) 和 1/38 (FSRCNN).

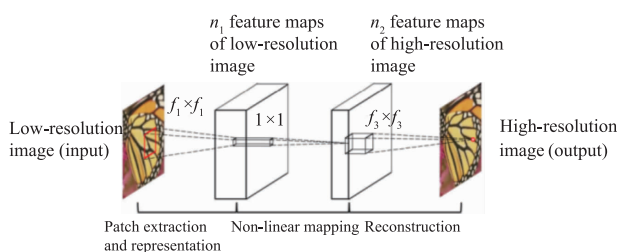


图4 SRCNN网络结构图<sup>[30]</sup>

### 3.2 基于先验知识的图像 SR 网络

针对 SRCNN 算法未能充分利用领域及先验知识的缺点,人们陆续提出改进方法,用图像梯度、边缘、稀疏性和自相似性等先验知识指导网络设计,使重建图像具有丰富的高频细节和更好的视觉效果.

#### (1) SRCNN-Pr

SRCNN-Pr<sup>[35]</sup> 利用图像边缘先验信息和不同尺度间的图像自相似特性进行图像多尺度放大.输入的 LR 图像同时经过多个放大因子对应的网络实现多任务学习,图像在不同尺度间的自相似性使得网络可以进行权值共享,缓解参数过多的问题.网络末端卷积层用于提取重建图像的边缘特征.在 SRCNN 损失函数的基础上,通过添加重建图像的梯度信息和目标 HR 图像梯度信息一致的约束条件来引入图像边缘梯度先验.整个网络损失函数由重建损失  $Loss_s$  和梯度信息损失  $Loss_{pr}$  两部分组成,如式(2)所示,其中  $\alpha$  是权重系数.

$$Loss = Loss_s + \alpha Loss_{pr} \quad (2)$$

将先验知识和特征提取相结合是 SRCNN-Pr 方法的创新之处.但领域先验知识对重建图像质量提升有限,如何实现高质量的高放大倍数图像重建是 SRCNN-Pr 需要解决的问题.

#### (2) SCN

SCN (Sparse Coding Based Network)<sup>[36]</sup> 通过神经网络实现基于稀疏编码的图像 SR.它利用 LISTA (Learned Iterative Shrinkage and Thresholding Algorithm) 将基于稀疏表示的 SR 方法中的稀疏表示、映射和重建三个模块纳入同一个稀疏编码网络 SCN 中,并通过级联 SCN 实现任意尺度的图像放大.

SCN 的思想与传统基于稀疏编码的 SR 方法完全一致,但使用深度网络实现并进行训练,避免了传统稀

疏编码方法需要大量优化计算的缺点.相比 SRCNN 这类基于数据驱动的网络,稀疏编码网络的每一层都具有清楚的物理意义,可解释性强,通过 LISTA 子网络结构加入图像的先验知识可以使得重建图像有更好的视觉效果.

### (3) DEGREE

DEGREE (Deep Edge Guided Recurrent Residual Network) 是 Yang 等人<sup>[37]</sup> 提出的一种加入边缘先验知识的深度残差网络.该网络包含四个部分:①LR 边缘提取层,提取图像边缘作为先验信息,将 LR 图像及其边缘特征图输入网络;②递归残差网络,将图像分解为边缘、纹理等成分,作为不同的子频带;③HR 边缘预测层,实现各成分的重建;④残差子带融合层,将各子频带的重建信息融合生成 HR 图像.

网络的损失函数由边缘损失和重建损失组成,均采用均方误差形式. DEGREE 将自然图像先验信息融入到图像 SR 中,可以更好地恢复细节,获得更好的视觉效果,但网络复杂,运行速度慢.在大放大因子下,重建效果不好.

### 3.3 基于残差结构的图像 SR 方法

一般而言,网络越深越宽,其感受野越大,能利用的上下文信息越多,学习能力及层次特征表征能力也越强.近年来,越来越多学者尝试在加深网络的同时使用跳连接来增强网络的特征学习能力,应用残差学习,梯度裁剪,参数共享等策略解决深度网络参数量大,不易收敛以及梯度爆炸/梯度消失等问题.我们根据网络中残差学习方式的不同、层间信息传递的不同以及实现多尺度放大方法的不同对各种基于残差结构的图像 SR 方法进行梳理.

#### 3.3.1 基于不同残差学习的方法

##### (1) VDSR

VDSR (Very Deep CNN for SR)<sup>[38]</sup> 是首个将残差结构用于图像 SR 问题的深度网络.该方法认为 LR 图像携带的低频信息与 HR 图像的低频信息相近,因此只需学习 HR 图像和 LR 图像之间的高频残差信息即可. VDSR 网络结构如图 5 所示,它将插值后得到具有目标尺寸大小的 LR 图像输入网络,再将这个图像与网络学到的残差相加得到最终的 HR 重建图像.

VDSR 的创新之处在于:

(I) 网络深度增加至 20 层.感受野由 SRCNN 中的  $13 \times 13$  变为  $41 \times 41$ ,网络能利用更多的上下文信息来预测图像细节.

(II) 将残差学习融入 SR 问题.残差图像比较稀疏,极大减轻了网络计算负担,加快了收敛速度. VDSR 还应用自适应梯度裁剪,将梯度限制在某一范围内,避免了梯度爆炸问题.



(III) VDSR 将不同放大倍数的图像混合在一起训练,使得模型可以解决不同放大倍数的图像 SR 问题。

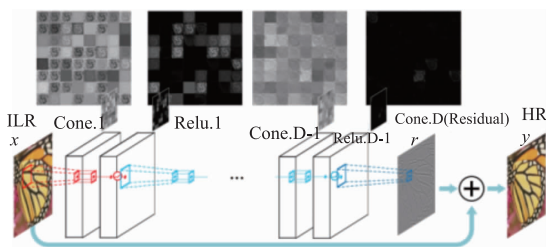


图5 VDSR网络结构图<sup>[38]</sup>

## (2) DRRN

VDSR 的基本思想来源于全局残差学习, DRCN (Deep Recursive Residual Network for SR)<sup>[39]</sup> 将全局残差学习与单权重的递归学习以及多目标优化结合起来, Tai 等人<sup>[40]</sup> 受 VDSR 和 DRCN 的启发, 提出了一种结合多路径模式的局部残差学习和全局残差学习以及多权重递归学习的 SR 方法——DRRN. 该网络一共 52 层, 相较于 VDSR 的 20 层网络结构, 网络规模更大, 计算更复杂, 重建图像质量提升不显著。

## (3) RCAN

输入图像含有丰富的低频信息, 现有基于 CNN 的 SR 网络无区别地对待这些信息的每个通道, 缺乏跨特征通道的区分学习能力, 限制了网络的表达能力. 因此, Zhang 等人<sup>[41]</sup> 提出了一个加入通道注意力机制的 400 层残差网络 RCAN (Residual Channel Attention Networks), RCAN 是首个将注意力机制应用于 SR 问题的网络. 为了缓解深度网络的训练难度, RCAN 采用 RIR (Residual In Residual) 结构, RIR 由 10 个残差块组通过一个长跳连接连接, 缓解了信息传递负担, 并能学习粗粒度残差信息, 稳定训练过程. 每个残差组由 20 个残差模块通过短跳连接相连, 使网络的主要部分学习 LR 的高频特征, 残差模块中加入通道注意力机制使 RCAN 可以自适应地学习更多有用的通道特征. 结果表明, 加入通道注意力机制可以将 PSNR 均值提升约 0.4 dB.

### 3.3.2 基于不同信息传递的方法

这类方法主要利用层间、尺度间信息流动加强不同层级和尺度的特征学习, 以提高重建图像的质量。

#### (1) DRCN

DRCN (Deeply Recursive Convolutional Network)<sup>[39]</sup> 将递归神经网络和残差学习应用于图像 SR 中, 整个模型分为三个模块: ①嵌入网络——用于特征提取; ②推理网络——使用递归结构加强层间信息传递和上下文信息联系, 学习 HR 图像与 LR 图像的残差, 恢复图像的高频部分; ③重建网络——将每一层的卷积结果和输入图像相加可得一个重建结果, 对所有卷积做相同

操作可得到多个重建结果, 再把它们加权平均可以得到最终的输出. 在网络结构上, DRCN 中递归层之间参数共享, 可有效减少网络参数, 同时用递归监督方法缓解梯度消失/爆炸问题, 在训练过程中, 将每个递归层的输出误差和总输出误差之和作为网络目标函数。

#### (2) DSRN

Liao 和 Poggio 等人<sup>[42]</sup> 证明了一个权重共享的残差网络等同于一个浅层循环神经网络 (RNN). 基于此, Han 等人<sup>[43]</sup> 发现包含 DRCN 和 DRRN 可以统一等效成一个只将 LR 空间信息传递到 HR 空间的单状态 RNN 网络. 为了能同时利用 LR 和 HR 空间的特征, 文献<sup>[43]</sup> 提出一种双状态循环网络——DSRN (Dual-State Recurrent Network). DSRN 中信息流动方向包括 LR→HR、HR→LR、HR→HR 和 LR→LR 四个方向, 这种设计加强了不同分辨率之间的信息共享. 实验结果表明, 使用参数共享后的网络重建图像 PSNR 提高 0.2 dB, 双状态循环网络比单状态循环网络提高 0.3 dB 左右, 且重建图像有更好的视觉效果, 这也表明加强不同分辨率图像间的信息传递, 充分利用不同分辨率下的分层特征有助于更好地恢复图像的高频细节, 这一点和文献<sup>[32, 44, 45]</sup> 中算法核心思想相吻合, 只是实现方式不同。

### 3.3.3 基于多尺度重建的方法

上述两类方法都是单放大因子模型, 无法得到不同放大因子的重建图像. 基于残差网络的多尺度图像 SR 方法主要包括 3 种。

#### (1) LapSRN

SRCNN 中需要将 LR 图像插值放大到理想分辨率尺寸输入网络, 增加了计算开销, 也会导致重建伪影. ESPCNN<sup>[33]</sup> 中使用亚像素卷积层替换插值, 但简单的浅层网络并不能很好地学到 LR→HR 图像间的复杂映射. 文献<sup>[46]</sup> 提出了一个基于拉普拉斯金字塔模型的网络结构 LapSRN (Laplacian Pyramid Networks for SR). 如图 6 所示, LapSRN 通过残差学习和逐级放大恢复 HR 图像. 每个放大模块完成两倍上采样操作, 这些放大模块结构相同, 模块间参数共享, 模块级联完成逐级放大, 做高倍上采样的同时, 也能得到中间低倍上采样结果的输出。

为了避免 L2 型损失函数带来的边缘平滑和模糊问题, LapSRN 中使用式 (3) 所示的 charbonnier 损失函数。

$$\begin{aligned} L(\hat{y}, y, \theta) &= \frac{1}{N} \sum_{i=1}^N \sum_{s=1}^L \rho(\hat{y}_s^{(i)} - y_s^{(i)}) \\ &= \frac{1}{N} \sum_{i=1}^N \sum_{s=1}^L \rho((\hat{y}_s^{(i)} - x_s^{(i)}) - r_s^{(i)}) \end{aligned} \quad (3)$$

其中:  $N$  是图像数目;  $L$  是总的级数;  $\rho$  是 Charbonnier 惩

罚函数;  $s$  是级联层数;  $\hat{y}_s^{(i)}$  是第  $s$  级第  $i$  张重建图像;  $y_s^{(i)}$  是第  $s$  级的真实 HR 图像;  $x_s^{(i)}$  是第  $s$  级第  $i$  张 LR 图像;  $r_s^{(i)}$  是学习到的第  $s$  级第  $i$  张 HR 图像和 LR 图像之间的残差。

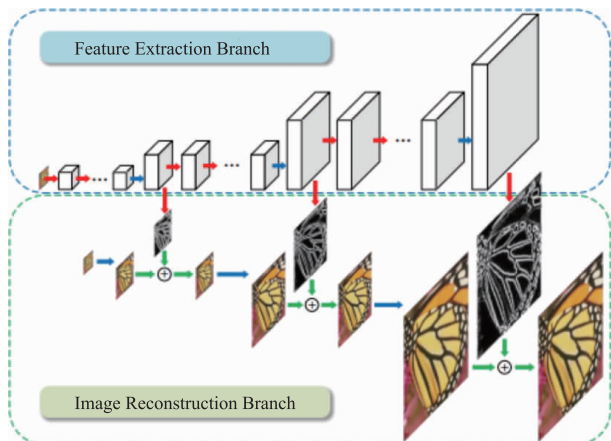


图6 LapSRN网络结构<sup>[46]</sup>

## (2) EDSR/MDSR

文献[47]去除残差块中的批归一化(BN)层,设计了一个用于单尺度SR的模型——EDSR(Enhanced Deep Residual Networks for SR)和多尺度SR的模型——MDSR。

训练EDSR时,先训练低倍数的上采样模型,接着用低倍数下上采样模型得到的预训练模型初始化高倍数的上采样模型,从而减少高倍数上采样模型的训练时间,同时训练结果也更好。MDSR是一个能同时进行不同上采样倍数放大的网络,MDSR中间部分和EDSR一样,只是在网络前面添加了不同放大倍数的预训练模型来匹配不同放大倍数的情况。在网络最后,将不同倍数的上采样结构平行排列获得不同倍数的输出结果。

## (3) MSRN

MSRN(Multi-Scale Residual Network)<sup>[44]</sup>是一种可以利用多尺度分层特征,进行任意尺度放大的图像SR网络,如图7所示。网络分为两部分:

(I)特征提取部分。由多尺度残差模块MSRB和分层特征融合HFFS模块组成。MSRB完成多尺度特征融合,每个MSRB单元使用局部残差学习减少计算复杂度,利用 $1 \times 1$ 卷积核组成瓶颈层提取各层特征中的有用信息,实现层次特征融合。

(II)重建部分。完成图像的放大,pixel shuffle层使该网络能完成图像任意尺度的放大。当使用8个MSRB模块时,实验效果和EDSR方法相当,但网络参数只有后者的1/7。

表1从网络特点和算法优点两方面比较了基于残差结构的网络。

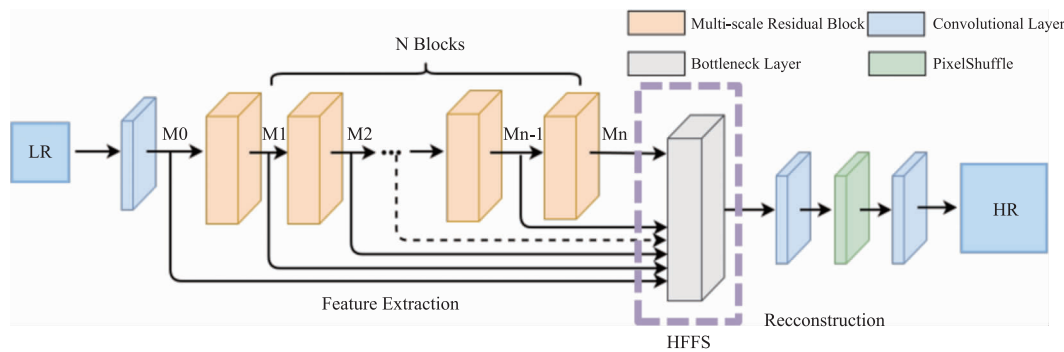


图7 MSRN网络结构图<sup>[44]</sup>

表1 基于残差结构的网络对比

算法	网络特点	算法优点
VDSR <sup>[38]</sup>	全局残差学习	加深网络,增加图像感受野;残差缓解训练负担
DRRN <sup>[40]</sup>	局部残差学习 + 全局残差学习 + 多权重递归学习	加深网络,提升重建图像质量
RCAN <sup>[41]</sup>	残差学习 + 通道注意力机制	自适应地学习更多有用的通道特征
LapSRN <sup>[46]</sup>	基于拉普拉斯金字塔模型;L2型损失函数	多放大因子逐级放大;解决边缘平滑和模糊问题
EDSR/MDSR <sup>[47]</sup>	用单尺度预训练模型加快多尺度模型训练	去除残差块中BN层,提升了训练速度;多尺度放大
MSRN <sup>[44]</sup>	多尺度特征融合	利用多尺度下图像分层特征,实现图像任意放大
DRCN <sup>[39]</sup>	递归网络 + 全局残差学习,加强层间信息传递	递归层间参数共享,有效减少网络参数,加强信息传递
DSRN <sup>[43]</sup>	采用循环卷积网络传递LR和HR之间信息	加强不同分辨率图像之间的信息传递

### 3.4 基于密集连接的图像 SR 方法

2016 年, Huang 等人<sup>[48]</sup> 提出密集连接网络——DenseNet, DenseNet 的基本思路与 ResNet 一致, 但它的组成单元是密集块. 密集块通过将所有卷积层的特征以串接方式连接起来, 有效减轻了网络梯度消失的问题, 并加强了层间特征的传播. 此外, DenseNet 通过特征在通道上的连接实现特征重用, 这些特点让 DenseNet 在参数和计算成本更少的情形下实现比 ResNet 更优的性能. DenseNet 的出现为 SR 问题的研究注入了新活力.

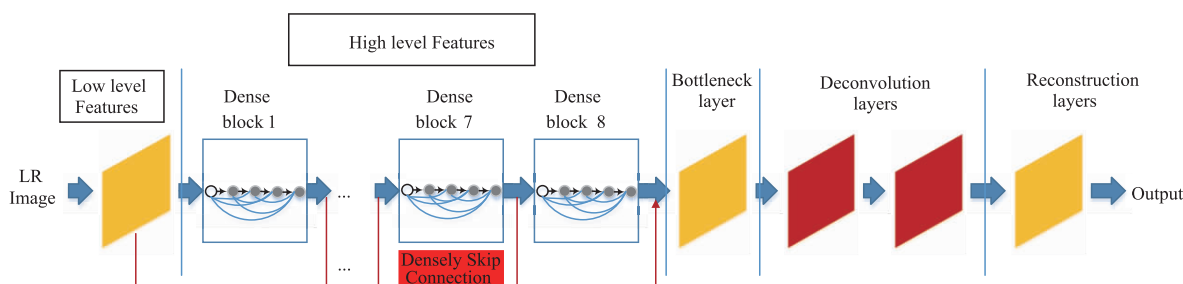


图8 SRDenseNet网络结构<sup>[45]</sup>

#### (2) MemNet

MemNet (Deep Persistent Memory Network) 模型<sup>[31]</sup> 是一个由递归单元和门控单元组成的内存块作为基本单元的深度学习网络. 它采用全局密集连接, 相较于 SRDenseNet 的局部密集连接, MemNet 考虑了前面状态/层对后续状态/层的影响, 解决了深度网络中的长期依赖问题. MemNet 的缺点在于前面层到后面层的直接连接, 限制了长期连接的能力, 使得局部信息不能完全被利用.

#### (3) RDN

考虑到之前的模型 (VDSR, DRCN, MemNet 等) 没有充分利用 LR 图像的分层特征, 文献[32]提出了一种结合局部残差学习和全局残差学习, 能充分利用所有层信息的网络——RDN (Residual Dense Network for SR). 该网络使用一系列串联的残差密集块提取并融合局部特征, 建立连续记忆机制, 最大程度地保留前面一层的信息, 传递层间信息, 对获得的局部特征进行全局融合, 自适应地学习得到全局密集分层特征.

由于 BN 层会消耗大量内存单元, 增加计算复杂度, 池化层会丢弃像素级信息. 因此, RDN 中取消了 BN 层和池化层. 该网络能对不同退化模型得到的 LR 图像进行高分辨率重建, 得到不错的重建效果.

### 3.5 基于残差结构与基于密集连接结构的网络比较

在基于残差结构的 VDSR 和基于密集连接的 SR-

#### (1) SRDenseNet

SRDenseNet (DenseNet for SR)<sup>[45]</sup> 首次将密集块结构应用到 SR 问题上. SRDenseNet 可以分成四个部分, 其网络结构如图 8 所示. 首先由一个卷积层提取 LR 图像的低层特征, 然后多个密集块通过密集跳连接, 将前面密集块的特征传递到后面密集块, 逐级学习高层特征, 并由反卷积层对图像进行上采样, 最后通过一个卷积层生成高分辨率输出. 该网络优点在于, 通过不同层之间的密集跳连接, 实现不同深度层次互补特征的共享, 从而进一步提升重建图像质量.

DenseNet 模型基础上, 各种新的网络陆续出现. 总体而言, 相关工作集中在三个方面: 前期工作通过增加网络深度提升性能. 研究如何在增加网络深度的同时, 减少网络参数, 加快网络训练速度. 这类网络 (如 VDSR, DRCN) 的层间信息流动性差, 没有充分挖掘图像不同层次的特征, 但重建速度和质量比 SRCNN 好; 第二类方法在增加网络深度的同时, 从网络内部模块间的连接、模块各层的连接方式 (跳连接、密集连接和密集跳连接) 和不同层次特征的融合方式 (局部和全局残差学习) 着手, 加强网络内部信息的流动性, 充分利用不同尺寸和层次的图像信息, 获取高频特征重建 HR 图像. 典型例子如 RDN 的基本模块内部密集连接, 模块间局部信息和全局信息融合. 这类网络结构较复杂, 参数较多, 计算复杂度高; 第三类方法将新的学习机制引入现有网络框架中, 如 RCAN 将注意力机制用于残差网络, 采用 RIR 结构, 结合长跳连接和短跳连接提高重建图像质量. 表 2 从网络输入、基本模块、模块间连接关系、是否包含残差学习、层间信息流动性和网络复杂度等方面, 比较了基于残差学习和基于密集连接的 SR 网络. 可以看出, 从 VDSR, DRCN, DRRN, SRDenseNet, MemNet, RDN 到 RCAN, 网络结构越来越复杂, 网络内部层间的信息流动性逐步提高, 不同层次间的信息被充分利用, 同时重建图像质量也逐步提升.



表 2 基于残差学习与基于密集连接类型的网络比较

算法	网络输入	基本模块	模块间连接方式	残差学习	层间信息流动性	网络复杂度
VDSR <sup>[38]</sup>	LR + 双三次插值	卷积层	串联	是	差	一般(20 层)
DRCN <sup>[39]</sup>	LR + 双三次插值	卷积层	串联	是	差	一般(20 层)
DRRN <sup>[40]</sup>	LR + 双三次插值	残差单元	串联	是	差	较复杂(52 层)
SRDenseNet <sup>[45]</sup>	LR + 双三次插值	密集块	密集跳连接	否	较好	较复杂(68 层)
MemNet <sup>[31]</sup>	LR + 双三次插值	内存块	密集连接	是	好	复杂(80 层)
RDN <sup>[32]</sup>	LR	残差密集块	级联	是	很好	很复杂(149 层)
RCAN <sup>[41]</sup>	LR	加入通道注意力机制的残差单元	串联,短跳连接	是	好	很复杂(400 层)

### 3.6 基于 GAN 的 SR 方法

近年来 GAN 在图像生成领域中取得了巨大的成功,越来越多学者尝试将 GAN 用于图像 SR 任务,希望生成高质量的 HR 重建图像. 基于 GAN 的 SR 方法中,网络内部都使用了残差模块,因此我们将基于 GAN 的 SR 方法统一并入基于残差结构的 SR 方法中. 根据算法内容,我们将其分为四大类(如图 9 所示):①大放大因子下图像 SR 的 GAN 网络;②加入先验知识的 GAN 网络;③学习复杂退化模型的 GAN 网络;④实现传统 SR 方法的 GAN 网络.

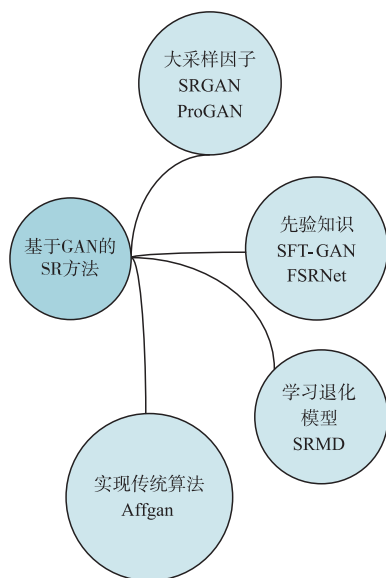


图9 基于GAN的SR方法分类

#### 3.6.1 基于 GAN 的 SR 网络模型

##### (1) 大放大因子下图像 SR 的 GAN 模型

大部分传统重建方法和基于深度学习的重建方法通常在网络前端或末端进行放大,这种方式适合于较小的放大因子(如放大因子为 2~3),当放大因子提高到 4~8 倍时,重建效果不理想. 人们提出了一些实现大放大因子下图像 SR 的 GAN 网络模型.

##### (a) SRGAN

Ledig<sup>[49]</sup>等人提出 SRGAN(GAN for SR)来恢复 4 倍上采样因子图像的高频细节. SRGAN 由 G(Genera-

tor)网和 D(Discriminator)网组成. G 网是一个残差网络,通过 LR 图像生成 HR 图像,由 D 网判断输入图像是由 G 网生成的,还是数据库中的真实图像. 当 G 网和 D 网达到博弈平衡时,该网络即可用于 SR.

SRGAN 使用式(4)所示的感知损失函数  $l^{\text{SR}}$ ,感知损失由内容损失  $l_{\text{X}}^{\text{SR}}$  和对抗损失  $l_{\text{Gen}}^{\text{SR}}$  两部分组成. 内容损失部分与文献[50]中的特征重建损失完全一致. 对抗损失关注纹理细节,使得生成网络的输出更接近自然图像.

$$l^{\text{SR}} = l_{\text{X}}^{\text{SR}} + 10^{-3} l_{\text{Gen}}^{\text{SR}} \quad (4)$$

实验结果表明 SRGAN 在 4 倍上采样因子时,其重建图像的平均 PSNR 值较 SRCNN 降低 1dB 左右,较 DRCN 降低 1.3~2dB 左右,但图像的纹理细节更丰富逼真,视觉效果更好. 该方法缺点在于网络结构复杂,需要训练两个网络,训练过程较长.

##### (b) ProGAN

图像 SR 过程是图像恢复高频细节并放大的过程,逐级放大重建是图像 SR 常用的策略,如 LapSRN<sup>[46]</sup> 每次实现 2 倍放大,直至实现 4 倍、8 倍高放大倍数的图像重建. 文献[51]将逐级重建思想和 GAN 结合起来,提出了一种基于渐进式 GAN 的图像多尺度逐级 SR 网络——ProGAN(Progressive GAN),有效实现了高倍数的图像 SR,同时 ProGAN 重建图像的 PSNR 值比同类逐级重建网络 LapSRN 平均提高 0.05~0.13dB.

##### (2) 基于先验知识的 GAN 模型

这类方法将包括物体结构或分割信息的先验知识融入网络或损失函数构造中,各网络的损失函数也在 SRGAN 感知损失函数的基础上,综合考虑基于像素、高层次特征和先验知识的各类损失,从而更好地提升重建图像的纹理或细节逼真度. 基于先验知识的 GAN 模型 SR 方法主要包括两种.

##### (a) SFT-GAN

SRGAN 通过感知损失获得包含丰富细节的 HR 重建图像,因为没有加入先验知识,这些细节可能并不真实. 文献[52]提出了一种考虑类别先验的 SR 重建网络——SFT-GAN(GAN with Spatial Feature Transform for SR). 该网络也是一种基于残差学习的 GAN,生成器由条件网络和 SR 重建网络组成. 条件网络将表达语义先



验类别的语义分割概率图作为输入,实时生成空间特征变换层,共享语义先验知识,利用类别先验知识恢复出与所属语义类别特征一致的纹理.判别器不但要判断生成图像是否真实,还需要预测输入的类别信息.通过实验,作者验证了加入类别先验信息可以恢复出更真实的图像纹理细节.

### (b) FSRNet

文献[53]将人脸先验知识加入网络,设计了人脸 SR 网络——FSRNet (Face SR with Facial Priors),整个网络由生成器和判别器两部分组成,生成器网络首先恢复出粗精度的 HR 图像,然后分别进入细粒度图像 SR 编码器提取图像特征,进入先验估计网络估计人脸特征点和分割信息先验知识,最后汇总至细粒度 SR 解码网络中,重构出目标 SR 图像. FSRNet 判别器的损失函数  $L_{\text{FSRNet}}$  如式(5)所示,在特征损失函数  $l_p$  和 GAN 损失  $l_{\text{GAN}}$  线性组合(权重系数分别为  $\gamma_p$  和  $\gamma_{\text{GAN}}$ )的基础上加入了人脸先验损失  $l_F$ . 人脸特征的先验知识能明显提升人脸图像 SR 质量.相比人脸特征点先验,分割先验知识包含更多信息,更能显著影响网络性能,而且全局分割图往往比局部分割图更有用.

$$L_{\text{FSRNet}} = l_F + \gamma_{\text{GAN}} l_{\text{GAN}} + \gamma_p l_p \quad (5)$$

### (3) 学习复杂退化模型的 GAN

一般用于图像 SR 的 LR 训练集,是通过对 HR 图像进行下采样得到的,该过程可视为理想的退化模型.而自然图像的退化过程往往更复杂且未知,为适应复杂退化模型的 SR 问题,文献[54]提出了一种维度拉伸策略,将退化过程的两个关键因素(即模糊核和噪声水平)作为网络输入,使得单个卷积 SR 网络可以处理多个甚至是不均匀退化类型的 SR 模型.文献[55]使用一个基于 GAN 结构的 SR 方法解决该问题.该网络由一个学习 HR 图像退化为 LR 图像的 GAN (High-to-Low GAN) 和一个学习从 LR 图像重建 HR 图像的 GAN (Low-to-High GAN) 组成.首先用不配对的 LR、HR 图像训练 High-to-Low GAN,学习图像的退化过程.训练完成后,将 LR 图像和与之对应的 HR 图像作为 Low-to-High GAN 网络的输入完成图像 SR.两个网络的损失函数均由 GAN 损失和基于像素的损失经过线性加权构成.

### (4) 与传统 SR 方法相结合的 GAN 模型

传统 MAP 算法必须利用图像先验知识对图像分布建模,但图像分布往往不可预知,导致建模困难且模型不准确导致重建图像质量不佳.文献[56]提出了一种将 GAN 模型与传统 MAP 算法相结合的 SR 方法——AffGAN (Affine Projected GAN),将极大似然估计问题转化为求解重构图像分布  $q_\theta$  与真实 HR 图像分布  $p_Y$  的 KL 散度  $-\text{KL}[q_\theta \| p_Y]$  的极小值问题,并最小化 GAN 损失函数.训练时通过对样本加入实例噪声增加 GAN 模型训练的稳定性.

## 3.7 基于其它深度学习的 SR 方法

CNN 被广泛应用于图像 SR 问题中,与此同时,也有学者提出将受限玻尔兹曼机(RBM),深度信念网络(DBN),自编码器(AE)等深度学习模型用于图像 SR.

文献[57]提出一种使用 RBM 训练稀疏字典,进行图像 SR 的方法.其中 RBM 的可见层由 LR、HR 图像块构成,隐藏层代表 LR、HR 图像字典对共享的稀疏系数,通过极大似然估计得到字典元素.该方法的本质仍是稀疏表示<sup>[28]</sup>,只是将字典构造和训练过程通过 RBM 完成,是传统方法和深度学习方法的结合.

将多层 RBM 堆叠起来可构成 DBN. Zhou 等人<sup>[58]</sup>利用 DBN 的层次结构学习 LR、HR 图像之间的非线性映射关系,通过训练和重建两个步骤完成图像 SR.在训练过程中,将 HR 图像分解成多个图像块并变换到 DCT 域,得到的二维 DCT 系数用于训练 DBN.重建时,将 LR 图像插值到目标尺寸后分解为图像块,变换到 DCT 域,利用训练好的 DBN 恢复丢失的高频信息.最后通过 DCT 逆变换重建 HR 图像.

文献[59]提出了一种基于堆协同局部自编码实现图像逐级放大的网络——DNC (Deep Network Cascade).该方法在多尺度图像块中进行非局部自相似搜索,增强输入图像块的高频纹理细节,抑制噪声,整合重叠图像块的兼容性.文献[60]提出了由对称的卷积层-反卷积层构成的极深层残差自编码器网络 RED (Residual Encoder-Decoder Networks).RED 中每个卷积层和反卷积层间使用跳线连接,解决梯度消失问题.卷积层获取图像的抽象内容,反卷积层放大特征尺寸并恢复图像细节.

## 4 讨论与分析

从 SRCNN 诞生以来,各种基于深度学习的 SR 方法不断涌现,比较总结上述各种代表性的 SR 方法,我们可以发现:

(1)在网络选择上,基于 CNN 的 SR 方法仍然是目前深度学习在 SR 应用中的主流方法,也有部分学者尝试使用 RBM、DBN 和 AE 等其它深度学习模型进行图像 SR.

(2)从网络设计和训练策略来看,基于 CNN 的 SR 方法从 SRCNN 开始,到基于残差结构的 VDSR 和基于密集连接的 SRDenseNet,再到基于 GAN 的 SR 方法,网络结构朝着更深、更宽、连接更复杂的趋势发展.深度网络能够充分利用图像上下文信息,扩大卷积感受野,提取图像高层语义特征.复杂的连接形式便于加强层间信息传递和特征传播,充分利用图像的层次特征,但会带来训练慢,难收敛等问题.针对上述问题,通常可使用跳连接、密集连接、残差学习、梯度裁剪和参数共享等策略减轻网络负担,加快学习速度,缓解梯度消失和爆炸.

(3)在网络输入层,一般方法将 LR 图像或经过上

采样的 LR 图像输入网络. 将后者输入网络会导致整个网络的运算在 HR 空间中进行, 不仅增加了网络运算复杂度, 也会使重建图像边缘模糊, 丢失细节特征.

(4) 针对多放大倍数下的 HR 图像重建问题, 现有模型主要通过以下四种方法解决.

(a) 训练不同放大因子对应的子网络, 得到不同放大因子下的重建图像. 该方法的缺点是训练好的网络只适用于固定放大因子的 SR 任务, 当放大因子变化时, 整个网络要重新训练.

(b) 在网络末端加入并调整反卷积层和 pixel shuffle 层 (如 FSRNet 和 MSRN) 或加入亚像素卷积层 (如 ESPCNN, MDSR, RDN 和 RCAN), 对重建图像进行不同倍数放大.

(c) 采用级联结构, 通过逐级放大形式得到不同放大倍数的重建图像 (如 SCN, LapSRN, ProGAN 和 DNC).

(d) 上述方法都只能实现整数倍数的图像放大, 为实现任意倍数的图像 SR, Mao 等人<sup>[61]</sup>引入元学习构造 Meta-Upscale Module 代替传统放大模块, 并嵌入到 MDSR, RDN 和 RCAN 等网络结构中, 实现了由单一模型进行任意尺度的图像 SR.

(5) 领域先验知识可以指导设计基于深度学习的 SR. 如 SRCNN-Pr、LapSRN、DNC 等方法利用图像不同

尺度间的自相似性来恢复图像高频细节, 实现图像逐级 SR. 而 SCN、SFT-GAN 和 FSRNet 等方法在子网络构造中分别加入稀疏先验, 类别先验及人脸特征先验知识来指导 SR. SRCNN-Pr 则利用图像梯度先验知识, 设计边缘损失和纹理损失函数, 指导网络训练.

(6) 融合领域先验知识和深度学习框架的 SR 方法比单纯使用基于数据驱动的深度学习方法能获得更符合人眼视觉特性, 更合理重建 HR 图像.

(7) 大部分网络的损失函数是基于像素差异的 L1 型、L2 型损失函数. L1 型损失函数收敛性比 L2 型好, 针对 L2 型损失函数带来的振铃现象, LapSRN 提出使用 charbonnier 损失函数代替 L2 型损失函数, 获得了更好的图像细节. 基于像素差异的损失函数并不能反映重建图像和真实图像间的主观感知差异, 基于 GAN 的 SR 方法中普遍结合感知损失函数, 生成高质量的重建图像并获得视觉效果的提升.

表 3 比较了不同放大因子下, 一些典型图像 SR 方法在 Set5、Set14、BSD100、Urban100 测试集上的实验结果, 实验结果均取自相关文献. 其中, SC 和 NE + LLE 分别是基于稀疏编码和基于流形学习的传统图像 SR 算法, RED 是基于自编码器网络的深度学习 SR 算法, 其余算法都是基于 CNN 的深度学习 SR 算法.

表 3 各种典型图像超分辨率重建方法实验结果 (PSNR/SSIM) 的比较

	数据集	性能指标	Set5			Set14			BSD100			Urban100		
			×2	×3	×4	×2	×3	×4	×2	×3	×4	×2	×3	×4
传统方法	Bicubic <sup>[30]</sup>	PSNR	33.66	30.39	28.42	30.24	27.55	26.00	29.56	27.21	25.96	26.88	24.46	23.14
		SSIM	0.9299	0.8682	0.8104	0.8688	0.7742	0.7027	0.8431	0.7385	0.6675	0.8403	0.7349	0.6577
	SC <sup>[28]</sup>	PSNR	—	31.42	—	—	28.31	—	—	—	—	—	—	—
		SSIM	—	0.8821	—	—	0.7954	—	—	—	—	—	—	—
	NE + LLE <sup>[27]</sup>	PSNR	35.77	31.84	29.61	31.76	28.60	26.81	—	—	—	—	—	—
		SSIM	0.9490	0.8956	0.8402	0.8993	0.8076	0.7331	—	—	—	—	—	—
深度学习方法	SRCNN <sup>[30]</sup>	PSNR	36.66	32.75	30.48	32.42	29.28	27.49	31.36	28.41	26.90	29.50	26.24	24.52
		SSIM	0.9542	0.9090	0.8628	0.9063	0.8209	0.7503	0.8879	0.7863	0.7101	0.8946	0.7989	0.7221
	VDSR <sup>[38]</sup>	PSNR	37.53	33.82	31.53	33.04	29.76	28.02	30.76	27.14	25.18	30.76	27.14	25.18
		SSIM	0.9587	0.9226	0.8854	0.9118	0.8311	0.7670	0.9140	0.8279	0.7524	0.9140	0.8279	0.7524
	DRCN <sup>[39]</sup>	PSNR	37.63	33.82	31.53	33.04	29.76	28.02	31.85	28.80	27.23	30.76	27.14	25.18
		SSIM	0.9588	0.9226	0.8854	0.9118	0.8311	0.7670	0.8942	0.7963	0.7233	0.9140	0.8279	0.7524
	LapSRN <sup>[46]</sup>	PSNR	37.52	33.82	31.54	33.08	29.79	28.19	31.80	28.82	27.32	30.41	27.07	25.21
		SSIM	0.9591	0.9227	0.8855	0.9130	0.8320	0.7720	0.8950	0.7973	0.7280	0.9101	0.8272	0.7553
	SRGAN <sup>[49]</sup>	PSNR	—	—	32.05	—	—	28.53	—	—	27.57	—	—	26.07
		SSIM	—	—	0.8910	—	—	0.7804	—	—	0.7354	—	—	0.7839
	RED <sup>[59]</sup>	PSNR	37.669	33.82	31.51	32.94	29.61	27.86	31.99	28.93	27.40	—	—	—
		SSIM	0.9599	0.9230	0.8869	0.9144	0.8341	0.7718	0.8974	0.7994	0.7290	—	—	—
	MDSR <sup>[47]</sup>	PSNR	38.11	<b>38.11</b>	<b>32.50</b>	33.85	30.44	28.72	32.29	29.25	27.72	32.84	28.79	<b>26.67</b>
		SSIM	0.9602	<b>0.9602</b>	0.8973	0.9198	0.8452	0.7857	0.9007	0.8091	0.7418	0.9347	<b>0.8655</b>	<b>0.8041</b>
	RDN <sup>[32]</sup>	PSNR	<b>38.24</b>	34.71	32.47	<b>34.01</b>	<b>30.57</b>	<b>28.81</b>	<b>32.34</b>	<b>29.26</b>	<b>27.72</b>	<b>32.89</b>	<b>28.80</b>	26.61
		SSIM	<b>0.9614</b>	0.9296	<b>0.8990</b>	<b>0.9212</b>	<b>0.8468</b>	<b>0.7871</b>	<b>0.9017</b>	<b>0.8093</b>	<b>0.7419</b>	<b>0.9353</b>	0.8653	0.8028

由结果可见,基于深度学习的方法较基于流形学习和稀疏编码等传统方法在图像重建质量上有较大提升.就基于深度学习的 SR 方法而言,图像自身的领域先验知识对提高重建图像质量具有十分重要的指导作用. RDN 将密集连接、跳连接、局部特征融合和全局特征融合整合起来,形成了一个复杂的网络,它充分利用图像各层次特征,在多数情况下取得了最好的重建效果,这也验证了网络越深,内部信息流动性越好,越有利于提升图像 SR 质量.

## 5 总结与展望

图像 SR 问题经历了基于重构、基于样例学习和基于深度学习的三个发展阶段.目前,基于深度学习(特别是以 CNN 为代表)的 SR 方法逐渐成为主流方法.基于深度学习的 SR 方法在模型选择、网络结构设计、领域先验知识引入和损失函数优化等方面不断创新.尽管图像 SR 研究取得了较大进展,但仍存在以下问题需要解决:

(1)网络选择和设计.目前大部分基于深度学习的 SR 方法采用 CNN 模型,如何将其它深度学习模型用于图像 SR 值得探索.此外,一般来说网络结构越深,重建效果越好,但也会导致训练困难,不易收敛等问题.如何权衡网络的复杂性和重建图像质量的矛盾是需要进一步研究的问题.

(2)如何更有效地将特定应用场合中的领域知识、先验知识和深度学习框架结合起来,提升现有基于深度学习的图像 SR 重建性能是值得探索的方向.领域和先验知识是人类在相关领域经过长期研究得到的正确知识和经验,具有重要的指导意义.而深度学习能充分利用大数据训练模型,自动提取和选择 SR 问题中有用的特征,学习 LR 到 HR 的非线性映射.若能充分利用领域和先验知识指导深度学习模型,不仅可以提取富含上下文信息的高层语义特征,进而学习更复杂的 LR 到 HR 映射关系,也能保证这种映射关系的合理性,从而进一步提高基于深度学习的 SR 模型的重建性能和可解释性.

(3)如何将新型学习机制引入图像 SR.注意力机制在自然语言处理、目标检测和分类等领域取得了成功.最近,Zhang 等人<sup>[41]</sup>将通道注意力机制引入 SR 问题,验证了通道注意力机制有利于图像恢复.如何将一种或多种注意力机制和现有网络结构相结合,并应用于 SR 问题是有潜力的研究方向之一.

(4)目前多数 SR 方法通过对 HR 图像下采样得到 LR 图像来模拟自然情况下图像退化过程,而实际应用中的 LR 图像通常存在运动变形、光学模糊、降采样和噪声污染等多种复杂降质因素.如何学习得到可以处

理复杂图像退化模型的网络有待探讨.

(5)无论是传统样例学习方法,还是基于深度学习的图像 SR 都是属于有监督的学习方法.在很多实际应用中,很难事先找到对应的 HR 图像.因此,探索基于弱监督学习和无监督学习的图像 SR 方法是未来图像 SR 技术发展的方向之一.

## 参考文献

- [1] Tsai R Y. Multiframe image restoration and registration [J]. *Advance Computer Visual and Image Processing*, 1984, 1: 317 - 339.
- [2] 吴秀秀,肖珊,张煜. 基于配准的肺 4D-CT 图像超分辨率重建[J]. *电子学报*, 2015, 43(2): 383 - 386.  
WU Xiu-xiu, XIAO Shan, ZHANG Yu. Registration based super-resolution reconstruction for lung 4D-CT image [J]. *Acta Electronica Sinica*, 2015, 43(2): 383 - 386. (in Chinese)
- [3] Jia Z, Wang H, Xiong Z, et al. Fast face hallucination with sparse representation for video surveillance [A]. *First Asian Conference on Pattern Recognition (ACPR)* [C]. USA: IEEE, 2011. 179 - 183.
- [4] Hu M G, Wang J F, Ge Y. Super-resolution reconstruction of remote sensing images using multifractal analysis [J]. *Sensors*, 2009, 9(11): 8669 - 8683.
- [5] Harris J L. Difraction and resolving power [J]. *Journal of the Optical Society of America*, 1964, 54(7): 931 - 936.
- [6] Goodman J W, Cox M E. *Introduction to Fourier Optics* [M]. McGraw-Hill, 1968.
- [7] Tsai R Y, Huang T S. Multiframe image restoration and registration [J]. *Advances in Computer Vision and Image Processing*, 1984, 1: 317 - 339.
- [8] Borman S, Stevenson R. *Spatial Resolution Enhancement of Low-resolution Image Sequences: A Comprehensive Review with Directions for Future Research*, Technical Report [R]. Laboratory Image and Signal Analysis, University of Notre Dame, 1998.
- [9] Park S C, Min K P, Kang M G. Super-resolution image reconstruction: a technical overview [J]. *Signal Processing Magazine IEEE*, 2003, 20(3): 21 - 36.
- [10] 苏衡,周杰,张志浩. 超分辨率图像重建方法综述 [J]. *自动化学报*, 2013, 39(8): 202 - 213.
- [11] Huang D, Liu H. A short survey of image super resolution algorithms [J]. *Journal of Computer Science Technology Updates*, 2015, 2(2): 19 - 29.
- [12] Hayat K. Super-resolution via deep learning [J]. *arXiv Preprint*, 2017, arXiv:1706.09077.
- [13] Sun X, Xiao-Guang L I, Jia-Feng L I, et al. Review on deep learning based image super-resolution restoration al-



- gorithms[J]. *Acta Automatica Sinica*, 2017, 43(5): 697 – 709.
- [14] Yang W, Zhang X, Tian Y, et al. Deep learning for single image super-resolution: A brief review [J]. *arXiv Preprint*, 2018, arXiv:1808.03344.
- [15] Kim S P, Bose N K, Valenzuela H M. Recursive reconstruction of high resolution image from noisy under sampled multiframe [J]. *IEEE Transactions on Acoustics Speech & Signal Processing*, 1990, 38(6): 1013 – 1027.
- [16] Rhee S, Kang M G. Discrete cosine transform based regularized high-resolution image reconstruction algorithm [J]. *Optical Engineering*, 1999, 38(8): 1348 – 1356.
- [17] 王相海, 张晓阳, 毕晓昀, 陶兢喆. 小波域多角度轮廓模板变分模型的单幅图像超分辨率重建[J]. *电子学报*, 2018, 46(9): 2256 – 2262.
- WANG Xiang-hai, ZHAO Xiao-yang, BI Xiao-yun, TAO Jing-zhe. Single image super-resolution reconstruction approach based on multi-angle contour templates variational calculus model in wavelet domain [J]. *Acta Electronica Sinica*, 2018, 46(9): 2256 – 2262. (in Chinese)
- [18] Nguyen N, Milanfar P. An efficient wavelet-based algorithm for image super-resolution [A]. *Proceedings of the 2000 International Conference on Image Processing [C]*. USA: IEEE, 2000. 351 – 354.
- [19] Panagiotopoulou A, Anastassopoulos V. Super-resolution image reconstruction employing Kriging interpolation technique [A]. *The 14th International Workshop on Systems, Signals and Image Processing and 6th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services [C]*. USA: IEEE, 2007. 144 – 147.
- [20] Irani M, Peleg S. Super resolution from image sequences [A]. *Proceedings of the 10th International Conference on Pattern Recognition [C]*. USA: IEEE, 1990. 115 – 120.
- [21] Schultz R R, Stevenson R L. An Bayesian approach to image expansion for improved definition [J]. *IEEE Transactions on Image Processing*, 1994, 3(3): 233 – 242.
- [22] Schultz R R, Stevenson R L. Improved definition video frame enhancement [A]. *International Conference on Acoustics, Speech, and Signal Processing (ICASSP-95) [C]*. USA: IEEE, 1995. 2169 – 2172.
- [23] Stark H, Oskoui P. High-resolution image recovery from image-plane arrays, using convex projections [J]. *Journal of the Optical Society of America A Optics & Image Science*, 1989, 6(11): 1715.
- [24] Elad M, Feuer A. Restoration of a single super-resolution image from several blurred, noisy, and under sampled measured images [J]. *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, 1997, 6(12): 1646 – 58.
- [25] Freeman W T, Jones T R, Pasztor E C. Example-based super-resolution [J]. *IEEE Computer Graphics & Applications*, 2002, 22(2): 56 – 65.
- [26] Glasner D, Bagon S, Irani M. Super-resolution from a single image [A]. *International Conference on Computer Vision [C]*. USA: IEEE, 2009. 349 – 356.
- [27] Chang H, Yeung D Y, Xiong Y. Super-resolution through neighbor embedding [A]. *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2004) [C]*. USA: IEEE, 2004. I-I.
- [28] Yang J, Wright J, Huang T S, et al. Image super-resolution via sparse representation [J]. *IEEE Trans on Image Process*, 2010, 19(11): 2861 – 2873.
- [29] 詹曙, 方琪, 杨福猛, 常乐乐, 闫婷. 基于耦合特征空间下改进字典学习的图像超分辨率重建 [J]. *电子学报*, 2016, 44(5): 1189 – 1195.
- ZHAN Shu, FANG Qi, YANG Fu-meng, CHANG Le-le, YAN Ting. Image super-resolution reconstruction via improved dictionary learning based on coupled feature space [J]. *Acta Electronica Sinica*, 2016, 44(5): 1189 – 1195. (in Chinese)
- [30] Dong C, Loy C C, He K, et al. Learning a deep convolutional network for image super-resolution [A]. *European Conference on Computer Vision [C]*. Springer, Cham, 2014. 184 – 199.
- [31] Tai Y, Yang J, Liu X, et al. MemNet: A persistent memory network for image restoration [A]. *IEEE International Conference on Computer Vision [C]*. USA: IEEE, 2017. 4549 – 4557.
- [32] Zhang Y, Tian Y, Kong Y, et al. Residual dense network for image super-resolution [A]. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [C]*. USA: IEEE, 2018. DOI: 10.1109/CVPR.2018.00262.
- [33] Shi W, Caballero J, Huszár, Ferenc, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network [A]. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [C]*. USA: IEEE, 2016. DOI: 10.1109/CVPR.2016.207.
- [34] Dong C, Loy C C, Tang X. Accelerating the super-resolution convolutional neural network [A]. *European Conference on Computer Vision [C]*. Springer, Cham, 2016. 391 – 407.
- [35] Liang Y, Wang J, Zhou S, et al. Incorporating image priors with deep convolutional neural networks for image super-resolution [J]. *Neurocomputing*, 2016, 194: 340 – 347.
- [36] Wang Z, Liu D, Yang J, et al. Deep networks for image

- super-resolution with sparse prior [A]. Proceedings of the IEEE International Conference on Computer Vision [C]. USA:IEEE,2015. 370 – 378.
- [37] Yang W, Feng J, Yang J, et al. Deep edge guided recurrent residual learning for image super-resolution [J]. IEEE Transactions on Image Processing, 2017, 26 ( 12 ): 5895 – 5907.
- [38] Kim J, Kwon Lee J, Mu Lee K. Accurate image super-resolution using very deep convolutional networks [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. USA:IEEE,2016. 1646 – 1654.
- [39] Kim J, Lee J K, Lee K M. Deeply-recursive convolutional network for image super-resolution [A]. Computer Vision and Pattern Recognition [C]. USA:IEEE, 2016. 1637 – 1645.
- [40] Tai Y, Yang J, Liu X. Image super-resolution via deep recursive residual network [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. USA:IEEE Computer Society, 2017. 2790 – 2798.
- [41] Zhang Y, Li K, Li K, et al. Image super-resolution using very deep residual channel attention networks [A]. Proceedings of the European Conference on Computer Vision (ECCV) [C]. USA:IEEE,2018. 286 – 301.
- [42] Liao Q, Poggio T. Bridging the gaps between residual learning, recurrent neural networks and visual cortex [J]. arXiv Preprint, 2016, arXiv:1604.03640.
- [43] Han W, Chang S, Liu D, et al. Image super-resolution via dual-state recurrent networks [J]. arXiv Preprint, 2018, arXiv:1805.02704.
- [44] Li J, Fang F, Mei K, et al. Multi-scale residual network for image super-resolution [A]. Proceedings of the European Conference on Computer Vision (ECCV) [C]. USA:IEEE,2018. 517 – 532.
- [45] Tong T, Li G, Liu X, et al. Image super-resolution using dense skip connections [A]. IEEE International Conference on Computer Vision [C]. USA:IEEE Computer Society, 2017. 4809 – 4817.
- [46] Lai W S, Huang J B, Ahuja N, et al. Deep laplacian pyramid networks for fast and accurate super-resolution [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. USA:IEEE Computer Society, 2017. 5835 – 5843.
- [47] Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution [A]. Computer Vision and Pattern Recognition Workshops [C]. USA:IEEE,2017. 1132 – 1140.
- [48] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. USA:IEEE,2017. 4700 – 4708.
- [49] Ledig C, Theis L, Huszar F, et al. Photo-realistic single image super-resolution using a generative adversarial network [J]. arXiv Preprint, 2016, arXiv:1609.04802.
- [50] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution [A]. European Conference on Computer Vision [C]. Springer, Cham, 2016. 694 – 711.
- [51] Wang Y, Perazzi F, McWilliams B, et al. A fully progressive approach to single-image super-resolution [J]. arXiv Preprint, 2018, arXiv:1804.02900.
- [52] Wang X, Yu K, Dong C, et al. Recovering realistic texture in image super-resolution by deep spatial feature transform [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. USA:IEEE, 2018. 606 – 615.
- [53] Chen Y, Tai Y, Liu X, et al. FSRNet: End-to-end learning face super-resolution with facial priors [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. USA:IEEE,2018. 2492 – 2501.
- [54] Zhang K, Zuo W, Zhang L. Learning a single convolutional super-resolution network for multiple degradations [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. USA:IEEE,2018. DOI:10.1109/CVPR.2018.00344.
- [55] Bulat A, Yang J, Tzimiropoulos G. To learn image super-resolution, use a GAN to learn how to do image degradation first [A]. Proceedings of the European Conference on Computer Vision (ECCV) [C]. USA:IEEE,2018. 185 – 200.
- [56] Sønderby C K, Caballero J, Theis L, et al. Amortised map inference for image super-resolution [J]. arXiv Preprint, 2016, arXiv:1610.04490.
- [57] Gao J, Guo Y, Yin M. Restricted Boltzmann machine approach to couple dictionary training for image super-resolution [A]. IEEE International Conference on Image Processing [C]. USA:IEEE,2014. 499 – 503.
- [58] Zhou Y, Qu Y, Xie Y, et al. Image super-resolution using deep belief networks [A]. International Conference on Internet Multimedia Computing and Service [C]. USA:ACM,2014. 28.
- [59] Cui Z, Chang H, Shan S, et al. Deep network cascade for image super-resolution [A]. Computer Vision-ECCV2014 [M]. Springer International Publishing, 2014. 49 – 64.
- [60] Mao X J, Shen C, Yang Y B. Image restoration using convolutional auto-encoders with symmetric skip connections [J]. arXiv Preprint, 2016, arXiv:1606.08921.
- [61] Hu X, Mu H, Zhang X, et al. Meta-SR: A magnification-arbitrary network for super-resolution [J]. arXiv Preprint, 2019, arXiv:1903.00875.

## 作者简介



**唐艳秋** 女,1995 年生,四川达州人. 东南大学自动化学院控制科学与工程专业硕士研究生,研究方向为机器学习、数字图像处理等.  
E-mail:220171510@seu.edu.cn



**潘 泓(通信作者)** 男,1974 年生. 博士,副研究员,主要研究方向为机器学习、深度学习、计算机视觉、医疗图像处理、机器视觉、多媒体信号处理.  
E-mail:enhpan@seu.edu.cn

**朱亚平** 女,1977 年生. 博士,教授,主要研究方向为机器学习、深度学习、图像处理、模式识别.

E-mail:zhuyaping@cuc.edu.cn

**李新德** 男,1975 年生. 博士,教授,主要研究方向为机器感知与学习、数据挖掘、机器视觉、智能信息融合和不确定推理、智能机器人.

E-mail:xindeli@seu.edu.cn