**图像超分辨率重建**

# 概述

高分辨率图像具有更密集的像素，包含更多精细有效的信息，这些信息在医学影像处理[[[1]](#endnote-1)-,[[2]](#endnote-2),[[3]](#endnote-3)]、安全监控[[[4]](#endnote-4)-[[5]](#endnote-5)]、遥感成像[[[6]](#endnote-6)-][[[7]](#endnote-7)][[[8]](#endnote-8)]、视频转换[[[9]](#endnote-9)-][[[10]](#endnote-10)]等领域至关重要，因此获取高分辨率图像具有十分重要的意义。然而，图像获取过程中会受到硬件条件限制，成像环境变化和图像传输信道干扰等因素影响；欠采样效应造成图像的频谱交叠，使获取的图像产生变形效应；大气扰动，拍摄图像与传感器之间的相对运动造成的变形、光学系统以及相对运动造成的模糊、下采样和系统噪声等，都会导致图像质量出现不同程度的下降。采用增加传感器阵列采样密度的办法来提高图像分辨率和消除变形效应的方法往往面临昂贵的代价和技术困难，因此从软件方面着手来提高图像分辨率有着极大的现实意义和应用价值[[[11]](#endnote-11)]。

图像超分辨率重建(Super resolution image reconstruction, SRIR 或 SR)是指用信号处理和图像处理的方法, 通过软件算法的方式将已有的低分辨率 (Low-resolution, LR) 图像转换成高分辨率(High-resolution, HR) 图像的技术[[[12]](#endnote-12)]。图像超分辨率重建的概念最早在1964年和1968年分别由Harris[[[13]](#endnote-13)]和Goodman[[[14]](#endnote-14)]提出，他们的方法被称为Harris-Goodman频谱外推法。1894年，Tsai和Huang[[[15]](#endnote-15)]等人提出频域中由低分辨率图像序列复原高分辨率图像的方法后，超分辨率重建技术逐渐被广泛研究。

传统的图像超分辨率重建算法包括基于插值，基于重构和基于学习的方法，这些方法普遍难以恢复丰富的高频信息，导致重建图像模糊；计算复杂，实时性不高；不适用于大上采样因子（如4,8）的要求。为了解决这些问题，有学者将深度学习应用于图像的高分辨率重建中。通过加深加宽深度网络，增加网络内部连接，加入图像边缘，梯度，类别等先验知识，优化感知损失函数，网络可以更好的学习到低分辨率图像和高分辨率图像的映射关系，从而重建出高质量的图像细节；通过残差学习，参数共享，梯度裁剪等策略减少网络参数，避免梯度爆炸，加快网络收敛速度，实现图像的实时高分辨率重建；通过逐级放大的方式使得网络能实现大放大倍数的高分辨率重建。如今，基于深度学习的超分辨率逐渐成为主流方法。

随着对图像超分辨率重建问题研究的深入，大量的相关成果不断涌现。为此, 众多学者逐渐展开了对图像高分辨率重建的综述工作，对该领域进行了归纳与总结。早期的综述文章[[[16]](#endnote-16)-][[[17]](#endnote-17)][[[18]](#endnote-18)]主要介绍传统的超分辨率算法的研究情况，不断更新的理论和方法有待整理。文献[18]按照输入输出情况将超分辨率问题分为基于重建的超分辨率、视频超分辨率、单帧图像超分辨率三大类，全面综述了该问题的常用算法及当前的最新研究成果，由于深度学习尚未应用于图像超分辨率重建领域，因此该文献不包含基于深度学习的算法综述。文献[19][[[19]](#endnote-19)]比较全面的介绍了现有图像高分辨率重建方法，但基于深度学习的方法仅介绍了SRCNN。文献[20-22][[[20]](#endnote-20)-][[[21]](#endnote-21)][[[22]](#endnote-22)]仅对基于深度学习的图像超分辨率重建技术进行了综述，文献[20]介绍了图像和视频的高分辨率重建现有算法和常用数据库，但仅限于基于卷积神经网络的方法，且没有将提到的算法进行归纳对比和总结；文献[21]从网络类型、网络结构、训练方法等方面详细地分析了现有技术的优势和不足，但涉及的算法较少，不全面；文献[22]从网络结构和优化目标出发较为全面的阐述，分析整理了基于深度学习的图像超分辨率重建技术，但分类不够明确，且不涉及传统的超分辨率重建技术。这些综述工作极大地推动并促进了该领域的研究与发展。然而, 随着时代的发展, 图像高分辨率重建技术也在突飞猛进, 近几年来该领域的研究成果可谓日新月异，各种优秀算法也是层出不穷，对这些工作及成果进行梳理与总结, 将会有效地推动和促进计算机视觉领域的发展。本文试图对图像高分辨率重建问题及其解决方法进行全面的梳理和总结，整体框架如图2所示，在传统的基于插值，基于重建和基于学习的高分辨率重建技术上重点补充并更新了基于深度学习的高分辨率重建方法。，即本文的主要贡献。

# 图像超分辨率重建算法

理想高分辨率图像一般经过几何运动变形、光学模糊、降采样以及噪声污染等一系列过程得到低分辨率图像，这一过程称为图像退化过程，如图1所示。

图1 图像退化过程

由图像的退化过程可以得到观测模型为：

（1）

gk表示某一幅低分辨率图像，z表示所求的高分辨率图像。分别表示几何运动矩阵，模糊矩阵和下采样矩阵，是加性噪声。

图像高分辨率重建即是通过图像处理技术，机器学习算法等手段从一张低分辨率图像或一个低分辨率图像序列中估计出高分辨率图像。图像高分辨率重建是一个病态问题。

图像超分辨率重建算法按照其原理通常可以分为基于插值的方法，基于重构的方法和基于学习的方法，如图2所示，其中基于深度学习的高分辨率重建算法是该领域的研究热点。



图2 图像超分辨率重建技术框架

## 2.1 基于插值的方法

基于插值的方法[[[23]](#endnote-23)]利用一些特定核函数确定的权值对某个待插值点周围的像素点进行加权平均来得到该像素点的值，对不同的核函数，由于权重可事先计算得到，重建时只需通过查表即可得到。传统的插值方法有以下三种，表1比较了其优缺点。

表1 基于插值的方法比较

|  |  |  |  |
| --- | --- | --- | --- |
| 方法 | 原理 | 优点 | 缺点 |
| 最近邻插值[[[24]](#endnote-24)] | 待插值的点由它最近点的灰度值决定 | 计算简单，插值速度快 | 重建图片产生块效应，图像边缘产生不同程度的锯齿效应 |
| 双线性插值[[[25]](#endnote-25)] | 分别在水平和垂直方向上进行线性插值 | 克服最近邻插值引起的图像边缘锯齿效应和块效应 | 重建图像边缘平滑 |
| 双立方插值[[[26]](#endnote-26)] | 利用待插值点周围上下左右四个领域内的16个像素点进行三次插值 | 消除图像边缘锯齿效应和块效应，重建图像的视觉效果好于前两种算法 | 计算量大，运算时间长，实时性差 |

在传统插值方法基础上，不断有学者提出了新的插值算法。文献[27-29][[[27]](#endnote-27)-],[[[28]](#endnote-28)]，[[[29]](#endnote-29)]结合传统插值方法和小波变换使重建图像具有更高的信噪比和更好的视觉效果，Zhang等人[[[30]](#endnote-30)]提出了一种基于定向滤波和数据融合的边缘引导非线性插值方法，可以保持边缘清晰度并减少振铃伪影。 Shi等人[[[31]](#endnote-31)]提出了一种基于边缘的扩展方法。该方法通过对由双线性或双三次插值得到的扩展图像，使用Canny边缘检测器来减少锯齿形边缘噪声，并修改边缘邻域中的像素值以减少重构模糊。

插值重建算法原理简单，运算速度快，实时性好，但该方法难以修复图像的高频细节，尤其当放大因子增加时，该方法的性能会急剧地下降，出现高频细节丢失、边缘模糊等问题。

## 2.2 基于重构的方法

基于重建的超分辨率技术即求解观测模型（式1）的逆问题，通常需要对多张低分辨率图片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，应用不同的约束条件（如平滑性、能量有限性和非负性等）求解高分辨图像。在基于重构的超分辨率图片重建方法中，主要分为频域法和空域法。

### 2.2.1 频域法

1984 年， Tsai 和 Huang [12]首次提出了在频域内提高图像分辨率的思想，利用一个低分辨率图像序列重建一幅高分辨率图片。它的主要思想是假定低分辨率图像生成模型后，分别对低分辨率图像和原始高分辨率图像进行离散傅里叶变换和连续傅里叶变换, 在频域中建立起二者之间的线性关系，从而重建出高分辨率图片。该方法基于图像中不存在运动模糊和观测噪声的假设，同时忽略光学系统的点扩散函数，因此只适合简单理想的图像退化模型。 Kim 等人[[[32]](#endnote-32)]在超分辨率重建过程中考虑观测噪声和空间模糊等因素，将递归最小二乘法应用于高分辨率图像重建的求解。文献[[[33]](#endnote-33)]使用离散余弦变换减少了运算量，并使用多通道自适应规整化参数来减轻高分辨率重建问题的病态性。

传统的基于频域的方法值只能处理输入低分辨率图像之间存在全局整体运动的情况，为了能更好的处理图像场景中存在相对运动物体的情况，有文献[[[34]](#endnote-34)-]，[[[35]](#endnote-35)]提出了基于小波变换的图像超分辨率方法，提高图像的局部质量。

频域重建方法算法简单，运行速度快，但该方法建立在整体平移运动的图像退化模型基础上，处理复杂的退化模型能力有限，且难以加入先验知识。因此，该方法没有成为超分辨率重建算法研究的主流。

### 2.2.2 空域法

空域方法是对影响低分辨率图像成像效果的空间域因素( 包括光学模糊、运动模糊等) 进行建模。因此，基于空间域的超分辨率重建方法更接近于实际应用情况。目前，常用的空域超分辨率图像重建方法有非均匀采样内插法、迭代反投影法、基于概率的方法、凸集投影法和 MAP /POCS 混合算法等。

1. 非均匀采样内插法（Non-uniform interpolation）

非均匀采样内插法的本质仍是对未知像素点的拟合。首先对低分辨率图像序列进行配准，在目标图像特定位置的内容约束下，进行插值得到高分辨率图像。非均匀采样内插法简单直观，计算效率高，但是这种算法需要假定所有低分辨率图像的噪声和模糊特征都是相同的，在图像复原阶段忽略了插值阶段的误差，难以加入先验约束知识，适应性不好，导致重建效果不佳，应用较少。

1. 迭代反投影法（Iterative back projection, IBP）

迭代反投影法[[[36]](#endnote-36)]首先通过对原始低分辨率图像进行插值得到预估的高分辨率图像，利用先验知识建立观测模型，将输入到观测模型中得到预估的低分辨率图像。

（2）

计算与原始低分辨率图像之间的差值，将此误差反向投影来修正得到新的高分辨图片：

（3）

其中表示后向投影算子。不断重复该迭代过程，直到误差满足要求。

该方法原理简单，收敛快，但是难以加入高分辨率图像的先验信息，反投影算子较难选择、通常解不唯一，并且每次迭代的误差均匀地累加到重建图像上，所以图像边缘存在一定程度的锯齿。

（3）凸集投影法（Projection onto convex sets, POCS）

凸集投影法基本过程是利用图像的先验知识获得一定的约束条件（平滑性，非负性和能量有限性，目标图像峰值像素的约束等），把约束条件定义为一系列约束凸集合，然后通过求解这些集合的交集获得高分辨率图像解空间。对于超分辨率图像空间中的任意一点，经过多次迭代运算， 可以得到一个满足所有凸集约束条件的收敛解。

该算法思想比较简单，易加入先验知识，较好地恢复了图像的边缘和细节。但初始点的随机选择会导致解不唯一，不稳定，且迭代过程收敛慢、计算量大。

（4）最大后验概率MAP

最大后验概率 (Maximum a posterior, MAP)方法是由Schultz等人[[[37]](#endnote-37)-]，[[[38]](#endnote-38)]提出的一种基于概率的算法，其基本思想是高分辨率图像出现的后验概率在低分辨率图像序列已知的前提下达到最大。

若将高分辨率图像H和低分辨率图像L 视为两个随机过程，则有：

（4）

表示对高分辨率图片H的估计，表示当高分辨率图像为H时，对应的低分辨率图像为L的概率。为H的先验概率，体现图像的先验知识，称为正则化项。

基于最大后验概率的高分辨率重建方法研究主要集中在先验知识（如边缘先验[[[39]](#endnote-39)]）的加入，正则化项的形式（Tikhonov 正则项[[[40]](#endnote-40)]，全变差正则化[[[41]](#endnote-41)]，双边全变差正则化[[[42]](#endnote-42)]等），正则化项的系数选取方法（利用L 曲线[[[43]](#endnote-43)]，U 曲线[[[44]](#endnote-44)]求解），MAP与其他算法的结合（如MAP-POCS 算法[[[45]](#endnote-45)]）。

该方法在解中加入先验约束，可以确保解存在并且唯一，收敛稳定性高，但是收敛慢，计算量大，边缘保持能力不如 POCS方法， 重构得到的高分辨率图像的细节易被平滑掉。MAP-POCS 算法结合了这两种算法的优点，被认为是基于重构的超分辨率算法最好的方法。

表2比较了基于频域和基于空域重构方法的特点及复杂度。可见，MAP-POCS 算法与其他算法相比，具有更好的综合性能。

表2 基于频域和空域的SR方法比较

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 频域方法 | 非均匀图像插值 | IBP | POCS | MAP | MAP-POCS 算法 |
| 图像先验知识 | 难加入 | 难加入 | 难加入 | 易加入 | 易加入 | 易加入 |
| 解是否唯一 | 是 | 是 | 否 | 否 | 是 | 是 |
| 复杂退化模型处理能力 | 差 | 中等 | 中等 | 高 | 高 | 高 |
| 计算复杂度 | 小 | 中等 | 中等 | 大 | 大 | 大 |
| 边缘保持能力 | 差 | 差 | 差 | 较好 | 差 | 好 |
| 适应性 | 较差 | 差 | 中等 | 高 | 高 | 高 |

## 2.3 基于学习的方法

基于学习的方法也称为基于样例（Exampled-based）的图像超分辨率重建，其主要思想是学习低分辨率图像与高分辨率图像之间的映射关系，获得先验知识，用于高分辨率图像的重建。

基于学习的高分辨率重建方法一般包含训练集的构建、特征学习和高频细节重建三个步骤。按照训练集的样本组成可以将基于学习的方法分为基于内部库的方法和基于外部库的方法。

### 2.3.1 基于内部库的方法

基于内部库的方法假设图像具有自相似性，即局部的图像内容及结构在图像不同尺度间或同一图像尺度内反复出现。训练集样本通常由不同分辨率图像金字塔组成。

Freeman等人[[[46]](#endnote-46)]提出相似的两个低分辨率图像块对应的高分辨率图像块应相似，相邻的两个高分辨率图像块重叠部分像素值应该一致。他们使用马尔科夫网络分别对高低分辨率图片间，邻接的高分辨率图像块之间进行建模，对输入的低分辨率图像搜索拼接得到对应的高分辨率图像。Glasner等人[[[47]](#endnote-47)]将同一尺度下图像自相似性和跨尺度的图像相似性结合，通过搜索—粘贴的方式逐级放大图片，填充高频细节实现高分辨率重建。在此基础上Freedman等人[[[48]](#endnote-48)]在图像块搜索阶段采取全局搜索和局部搜索相结合，加快搜索速度。为快速稳定地进行高分辨率重建，这两种方法均采用小的放大因子不断迭代至需要的放大倍数。

### 基于外部库的方法

基于外部库的方法主要是学习高低分辨率图像块之间的映射关系。训练集的样本由同一尺寸的高分辨率图像(HR)—低分辨率(LR)图像对构成。根据学习算法的结构层数，可进一步分为基于浅层结构的重建算法额基于深度学习的重建算法。基于领域嵌入算法和基于稀疏表示的方法是基于浅层结构的代表算法。

（1）基于邻域嵌入的方法

基于领域嵌入的基本思想是假设高分辨率图像与其对应的低分辨率图像块在特征空间中具有相似的局部流形。利用NE+LS(Neighbor Embedding + Least Square), NE+LLE[[[49]](#endnote-49)](Neighbor Embedding + Locally Linear Embedding), NE+NNLS[[[50]](#endnote-50)](Neighbor Embedding + Non-Negative Least Square)得到一组权值，使得低分辨率图像与其在样本中对应的K个最近邻图像块加权得到的低分辨率图像误差最小，对高分辨率图像块应用这组系数进行加权即可得到重建的高分辨率图像。

邻域嵌入方法简单直接，对于样本集的依赖性小，但是最近邻图像块数目需人为选定，易导致过拟合或者欠拟合的现象，从而影响重建图像质量。

（2）基于稀疏表示（Sparse Coding, SC）的方法

基于稀疏表示的方法[[[51]](#endnote-51)]在LR-HR样本库中建立一对共享系数并能产生最小重建误差的过完备字典，对待重建的低分辨率图像利用LR字典计算其稀疏表示系数，将同样的系数应用于HR字典上生成HR特征，得到最终的重建高分辨率图像。

使用稀疏表示法是对基于领域嵌入方法的拓展，克服了需要固定最近邻图像块数目的缺点，可扩展性增强，且对噪声不敏感，但图像的重建质量和计算复杂度均与字典大小成正相关，当建立的字典完备性不强则会导致图像边缘细节模糊。

基于外部库的方法还包括基于深度学习的高分辨率重建方法，这类方法将放在第三章重点讨论。

表3 常用的基于学习方法的算法

|  |  |  |
| --- | --- | --- |
| 算法名称 | 基本原理 | 特点 |
| Self-Ex[[[52]](#endnote-52)] | 基于图像自相似性 | 利用图像内部信息，不需要建立外部库，但搜索时运算量大 |
| NE+[49] | 基于高低分辨率图像具有相似的局部流形 | 需要建立LR-HR样本库学习低分辨率图像到高分辨率图像的映射关系，能获取更多高频信息，生成的高分辨率图像细节更丰富 |
| SC[51] | 基于图像的稀疏表示 |
| K-SVD[[[53]](#endnote-53)] | 利用K-SVD方法进行高效的过完备字典建立，是对SC方法的改进 |
| ANR(Anchored Neighborhood Regression)[[[54]](#endnote-54)] | 稀疏表示和邻域嵌入方法的结合，利用全局协同编码加快运行速度 |

# 基于深度学习的图像超分辨率重建

深度学习方法在计算机视觉领域表现出巨大的潜力, 它可以通过建立强有力的模型和设计高效的学习策略来克服过度拟合, 并且神经网络可以灵活地通过增加新的非线性激活函数或特定功能的层来更好地拟合训练数据。2014年Dong等人[[[55]](#endnote-55)]提出一个仅含三个卷积层的网络---SRCNN（Super-resolution CNN），首次将深度学习应用于图像超分辨率重建领域，从此，基于深度学习的高分辨率重建网络不断出现改进。根据网络类型可将基于深度学习的图像超分辨率重建分为基于卷积神经网络的SR方法和基于其他网络结构（如受限玻尔兹曼机，深度信念网络，自编码器网络）的SR方法，其中使用卷积神经网络结构的SR方法根据网络的层数进一步分为基于浅层网络和基于深层网络的SR方法两大类，基于深层网络的方法可以进一步分为基于残差结构和基于密集连接两大类，有网络如MemNet[72], RDN[73]同时借鉴了这两种结构，其中基于残差结构的SR方法中有很特殊的一类方法，即基于GAN的SR方法。



图3 基于深度学习的SR方法分类

## 3.1 基于浅层网络的SR方法

### 3.1.1 SRCNN系列网络

（1）SRCNN网络

SRCNN[55]借鉴了基于稀疏编码的图像超分辨率重建的思想，通过卷积操作完成图像块的提取，特征非线性映射和重建过程三个步骤，网络结构如图4所示。首先将低分辨率图像（LR）通过双三次线性插值到目标高分辨率图像大小作为输入，利用第一层卷积对图像密集采样为一个个图像块，每个图像块表示成一个高维向量，这些高维向量组成个特征图，第二层卷积主要将表示低分辨率图像的个特征图非线性映射成代表高分辨率图像的个特征图，第三层卷积采用全连接方式将预测得到的高分辨率图像块整合生成一张高分辨率图像。

训练阶段将高分辨率图像和对应的重建高分辨率图像间的均方误差（Mean Squared Error, MSE）作为损失函数，如式5所示，采用随机梯度下降法（Stochastic Gradient Descent, SGD）最小化损失函数，进一步提升训练速度。当放大因子为2,3,4时，该方法在最常用的测试集Set5, Set14上的重建图像的平均峰值信噪比（Peak Signal to Noise Ratio，PSNR）比传统方法（三次插值[26]，SC[51]，ANR[52]等）高至少0.3dB, 平均重建速度较最快的传统方法提升两倍左右。该方法在图像重建速度和重建质量上均优于传统方法，是深度学习应用于图像高分辨率重建问题的典范和基础，但该网络主要存在以下问题：1）只适用于单放大因子的图像高分辨率重建，针对不同的方法因子，该网络需要重新训练；2）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高分辨率空间进行，导致计算量大，训练速度慢；3）网络比较浅，感受野小（13）不能很好地能利用的图像上下文信息。

(5)

其中表示损失函数，表示网络优化参数，n表示训练样本集大小，表示低分辨率图像，表示重建高分辨率图像，表示原高分辨率图像。

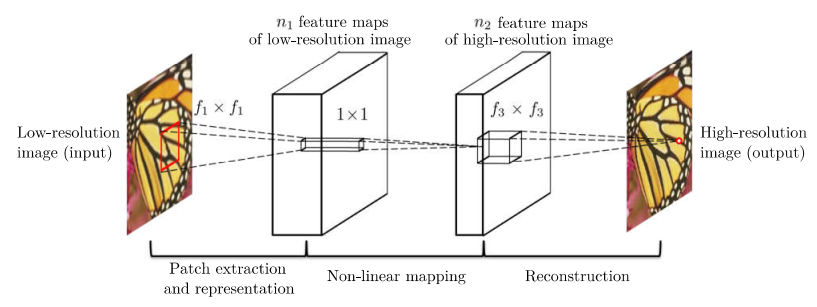


图4 SRCNN网络结构图

（2）FSRCNN（Fast Super-Resolution by CNN）

FSRCNN[[[56]](#endnote-56)]是对之前SRCNN的改进，主要在三个方面：1）直接将原始的低分辨率图像输入到网络中，不需要先通过双三次插值方法放大尺寸，可以减少计算量，加快训练速度；2）使用更小的卷积核和使用更多的映射层改变特征维数，实现多通道的非线性映射和特征融合；3）通过微调最后一层反卷积层来训练不同上采样因子的模型。

该网络可以分为五个部分: 1）特征提取：直接是对原始的低分辨率图像进行操作；2）收缩：通过应用1×1的卷积核进行降维，减少网络的参数，降低计算复杂度；3）非线性映射：采用两个串联的3×3的卷积核可以替代SRCNN中一个5×5的卷积核，减少运算参数；4）扩张：低维度特征带来的重建效果不佳，因此用1×1的卷积核进行扩维，即是收缩的逆过程；5）反卷积层：实现图片放大。

FSRCNN重建图像质量较之前的方法略好，其优势在于运行速度十分快，当放大因子为2,3,4时，该方法在最常用的测试集Set5[98], Set14[99]，BSD200[100]上重建一副高分辨率图像平均所需时间是SRCNN的十分之一，甚至更快，达到毫秒级，可以用于实时的图像高分辨率重建。

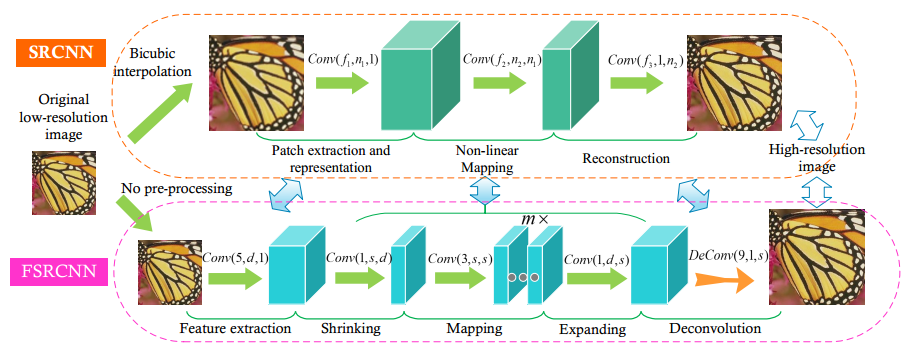


图5 FSRCNN网络结构图

（3）ESPCN（Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network）

文献[57][[[57]](#endnote-57)]提出了一种直接在低分辨率图像尺寸上提取特征，计算得到高分辨率图像的高效方法--ESPCN网络。该网络主要有两部分：1）隐藏层：原始低分辨率图像通过两个卷积层组成的隐藏层后，得到的r2×H×W (r是图像的目标放大倍数，H，W分别是目标图像的高度和长度)的特征图像。2）亚像素卷积层：将每个像素的r2个通道重新排列成一个r×r的区域，对应于高分辨率图像中的一个r×r大小的子块，从而特征图像被重新排列成1×rH×rW大小的高分辨率图像。

只在最后一层对图像大小做变换，前面的卷积运算由于在低分辨率图像上进行，因此效率会较高。在Set14上进行图像放大三倍的高分辨率重建测试，每幅图像的平均重建时间是4.7ms，ESPCN网络可以用于图像和视频的实时重建，图像重建质量和SRCNN差不多。

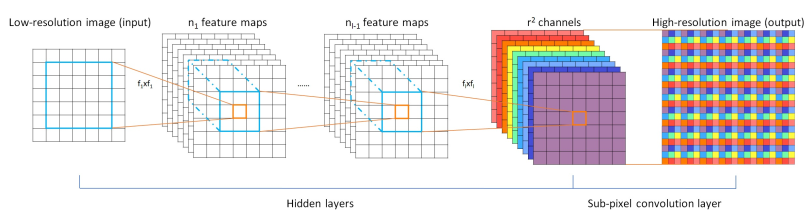


图6 ESPCN网络结构图

### 3.1.2 加入领域先验知识的浅层网络

图像超分辨率重建问题中可以利用图像的自相似性，稀疏性等领域知识（domain expertise）和图像的边缘梯度信息，类别信息等先验知识(prior knowledge)指导图像超分辨率重建，使得重建图像具有丰富的高频细节和更好的视觉效果。在网络中加入领域先验知识并无统一的框架和方法。

（1）SRCNNN-Pr

Liang等人[[[58]](#endnote-58)]提出了利用图像边缘先验信息和不同放大尺度间的图像自相似性实现图像多尺度放大的网络SRCNNN-Pr，其网络结构如图7所示。输入的低分辨率图同时经过多放大因子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的第二层卷积间可以进行权值共享，缓解参数过多的问题。网络末端的卷积层用于提取重建图像边缘特征，在SRCNN的损失函数中添加重建图像的梯度信息和目标HR的梯度信息一致的约束条件来加入图像边缘梯度先验。网络的损失函数由重建损失和梯度信息损失两部分组成，如式6所示，其中是权重系数。

（6）

实验结果表明，当放大因子为3时，在Set5测试集上，该方法训练速度比SRCNN快，重建图像PNSR值较SRCNN高0.01~0.02db，将先验知识和深度学习相结合是SRCNN-Pr方法的创新所在，但领域先验知识指导图像超分辨率提升对重建图像的质量提升有限，如何实现高质量的大放大倍数图像重建是SRCNN-Pr待解决的问题。

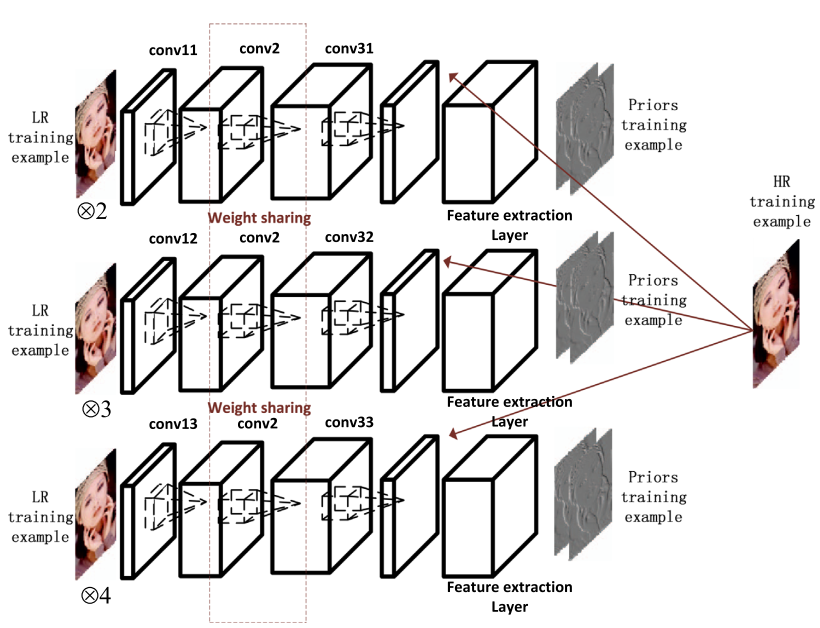


图7 SRCNN-Pr网络结构图

（2）SCN(Sparse Coding Based Network)

Wang等人[[[59]](#endnote-59)]提出一种通过神经网络实现基于稀疏编码的高分辨图像重建的网络—SCN。该方法主要思想是使用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，重建三个模块纳入同一个稀疏编码网络SCN中， 通过级联SCN可实现任意尺度的图像放大。 SCN网络结构如下图7所示，将低分辨率图像通过卷积层H分割为图像块并提取其特征表示为y，y经过LISTA网络（图中红色框部分）可以得到稀疏编码，稀疏编码通过代表高分辨率图像的字典的卷积层得到高分辨率图像块x，卷积层G将得到的高分辨率图像块x连接重建出一幅高分辨率图像。

该网络的思想与传统的基于稀疏编码的高分辨率重建方法完全一致，使用深度学习的方法进行网络的训练，可以避免传统方法需要大量优化计算的缺点，同时相比SRCNN这类完全基于数据驱动的网络而言，稀疏编码网络的每一层都具有很清楚的物理意义，可解释性强，通过LISTA子网络结构加入图像的先验知识使得重建图像有更好的视觉效果，实验结果表明，当放大因子为2,3,4时，SCN的PNSR值较传统的基于稀疏表示的方法高0.3∼1.6dB，较SRCNN方法提高0.2∼0.5dB。

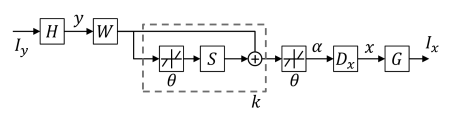


图8 SCN网络结构图

（3）DEGREE（Deep Edge Guided Recurrent Residual network）

Yang等人[[[60]](#endnote-60)]提出基于深度边缘指导的反馈残差网络DEGREE。该网络包含四个部分，1）LR边缘提取层：提取图像边缘作为先验信息，然后低分辨率图像及其边缘特征图作为网络输入；2）递归残差网络：将图像信号分解为边缘、纹理等成分，作为不同的子频带；3）HR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目标高分辨率图像。

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。实验结果表明放大因子为2,3,4时，重建图像的PNSR值较SRCNN放大提升了0.4~1.1db，将自然图像先验信息嵌入到图像高分辨率重建中，可以更好的恢复细节，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，效果不好。

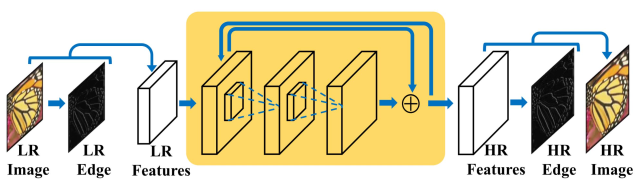


图9 DEGREE网络结构图

（4）ZSSR（Zero-Shot Super-Resolution）

Shocher等人[[[61]](#endnote-61)]提出了一种不需要预训练，在测试时利用图像自相似性训练的同时 进行高分辨率图像重建的自监督卷积神经网络方法ZSSR。实验结果表明在理想退化情况下，该方法得到的重建图像质量和当时最好的SR方法EDSR相当，复杂退化情况下，重建图像PNSR值比EDSR高1~2db，且由于该网络不需要预训练过程，网络规模小，运行速度更快。

以上对浅层的图像超分辨率重建网络进行了梳理，在SRCNN网络基础上，有研究从加快运算速度，解决大放大因子重建问题，加入图像先验等方面着手解决图像超分辨率问题，表3给出对以上算法的总结对比。

表4 浅层图像超分辨率图像重建网络算法对比

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 算法 | 网络层数 | 算法目标 | 领域先验知识及其加入形式 | 损失函数 |
| SRCNN | 3 | SR与CNN结合 | -- | L2范数 |
| FSRCNN | 8 | 加快运算速度 | -- | L2范数 |
| ESPCN | 3 | 逐级放大重建 | -- | L2范数 |
| SRCNNN-Pr | 3 | 加入梯度先验 | 梯度先验：边缘提取层+梯度信息损失函数 | L2范数 |
| SCN | 7 | SC方法网络化 | 稀疏先验：LISTA子网络 | L2范数 |
| DEGREE | 10 | 加入边缘先验 | 边缘先验：边缘提取层+边缘信息损失函数 | L2范数 |
| ZSSR | 8 | 适应复杂退化模型 | 图像自相似性：全连接层 | L1范数 |

## 3.2 基于深层网络的SR方法

一般而言，网络越深越宽，其视野域越大，能利用的上下文信息越多，学习能力及层次特征表征能力越强，因此，随着深度学习在图像超分辨率重建领域的应用，越来越多学者尝试在加深网络的同时使用跳连接（skip connection）和密集连接（dense connection）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。

### 3.2.1 基于残差结构的SR方法

（1）VDSR （Very Deep Convolutional Networks for SR）

VDSR[[[62]](#endnote-62)]是第一个将残差结构应用于图像超分辨率重建问题的深度网络，其网络结构如图10所示。该方法认为低分辨率图像携带的低频信息与高分辨率图像的低频信息相近，因此只需要学习高分辨率图像和低分辨率图像之间的高频部分残差即可。VDSR将插值后得到的变成目标尺寸的低分辨率图像作为网络的输入，再将这个图像与网络学到的残差相加得到最终的网络的输出。

VDSR有四个创新之处：1）加深了网络结构(20层)，使得越深的网络层拥有更大的感受野。感受野由SRCNN的13\*13变为41\*41。图像的感受野越大意味着网络能利用更多的上下文信息来预测图像细节。2）采用残差学习，残差图像比较稀疏，大部分值都为0或者比较小，减轻了网络负担，加快收敛速度。VDSR还应用了自适应梯度裁剪(Adjustable Gradient Clipping)，将梯度限制在某一范围，也能够加快收敛过程，避免梯度爆炸问题。3）VDSR在每次卷积前都对图像进行补0操作，保证了所有特征图和最终的输出图像在尺寸上保持一致，使得重建图像边缘更清晰。4）VDSR将不同倍数的图像混合在一起训练，这样训练出来的一个模型就可以解决不同放大倍数的图片超分辨率问题。

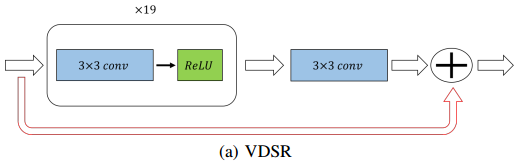


图10 VDSR网络结构图

（2） DRCN

Kim等人进一步提出了一个与VDSR类似的网络DRCN [[[63]](#endnote-63)]，其网络结构如图11所示。DRCN将递归神经网络(Recursive Neural Network)结构和残差学习思想应用于图像的超分辨率重建中。DRCN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的信息传递和上下文信息联系，学习高分辨率图像与低分辨率图像的差，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得到一个重建结果，从而共得到D个重建结果，再把它们加权平均得到最终的输出。在训练过程中，递归层之间进行参数共享，避免参数过多的情况，采用递归监督方法缓解梯度消失/爆炸问题，将每一个递归层输出的误差和总输出的误差之和作为待优化的目标函数。

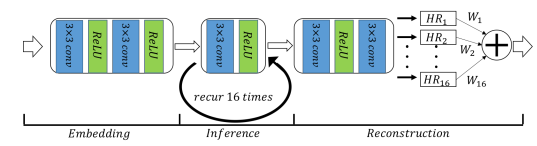


图11 DRCN网络结构图

（3） DRRN（Deep Recursive Residual Network for SR）

VDSR基于全局残差学习，DRCN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[[[64]](#endnote-64)]受到VDSR和DRCN的启发，提出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的递归学习的SR方法--DRRN。DRRN一共52层，相较于VDSR的20层，网络规模大，计算复杂，重建图像质量仅有小幅度的提升。

（4） EDSR

文献[65][[[65]](#endnote-65)]去除残差块中的批归一化（Batch Normalization, BN）层建立起一个用于单尺度高分辨率重建的网—EDSR(图12左)和多尺度高分辨率重建的网—MDSR（图12右）。

EDSR在训练时先训练低倍数的上采样模型，接着用训练低倍数上采样模型得到的参数来初始化高倍数的上采样模型，这样能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的网络，MDSR的中间部分和EDSR一样，只是在网络前面添加了不同放大倍数的预训练好的模型来减少不同倍数的输入图片的差异。在网络最后，不同倍数上采样的结构平行排列来获得不同倍数的输出结果。

文献提出来的两个网络最大的特点是1）去掉BN层，减少内存使用，提升训练速度；2）充分利用图像不同尺度间的内部关系，用低倍数的上采样模型来训练高倍数的模型，减少了模型参数，加快训练，且能提升图片的重建质量。

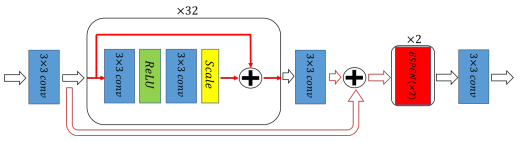
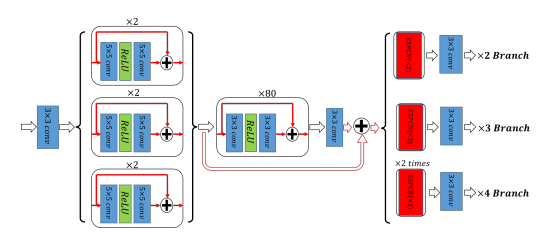
 

图12（左）EDSR网络结构图；（右）MDSR网络结构图

（5）LapSRN（Laplacian Pyramid Networks for SR）

SRCNN在输入图像进网络前，需要使用预先定义好的上采样操作来获得目标的空间尺寸，这样的操作增加了额外的计算开销，同时也会导致可见的重建伪影。而在ESPCN[57]使用了亚像素卷积层来替换上采样操作，但浅层简单的网络并不能很好地学到低分辨率图像到高分辨率图像复杂的映射。文献[66][[[66]](#endnote-66)]提出了一个基于拉普拉斯金字塔模型的网络结构LapSRN，通过残差学习和逐级放大操作恢复高分辨率图片。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，在做高倍上采样时，也能得到中间低倍上采样结果的输出。

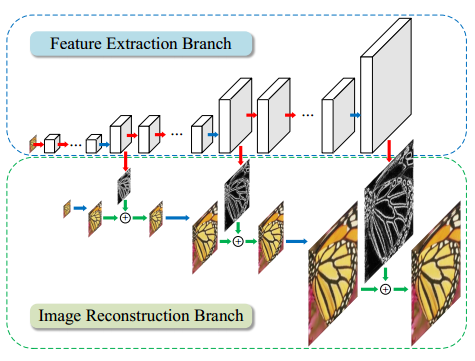


图13 LapSRN网络结构

为了避免L2型损失函数带来的边缘平滑和模糊问题，文中采用使用charbonnier损失函数，如式7所示。

（7）

其中是训练集中的图像数目，表示Charbonnier惩罚函数，s是级联的级数，取1,2,3；表示第s级第i张重建图像，表示第s级的目标高分辨率图像，表示第s级的第i张低分辨率图像，表示学习到的第s级的第i张高分辨率图像和低分辨率图像之间的残差。

（6） DSRN（the Dual-State Recurrent Network）

Liao and Poggio等人[[[67]](#endnote-67)]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（Recurrent Neural Network, RNN），如图14（a）（b）（c）所示。基于此，Han等人[[[68]](#endnote-68)]发现包含递归计算和权重共享的残差网络如DRCN和DRRN可以统一等效成一个展开的单状态RNN网络，如图11（d）（e）所示。

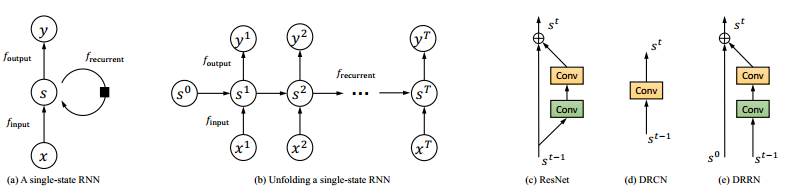


图14 （a）单状态RNN；（b）展开的单状态RNN；（c）残差网络；（d）DRCN网络；（e）DRRN网络

单状态循环网络只将LR空间信息传递到HR空间，为了能同时利用LR和HR空间的特征，文献[]提出一种双状态循环网络--DSRN，其网络结构如图14（a）所示，图14（b）是图14（a）的展开结构，等效于T个参数共享的残差块级联，图中箭头表示两个状态（：高分辨率状态和：低分辨率状态）间的传递函数，表5给出了该网络的传递函数作用及其实现方式，可以看出，该网络信息流动方向包括LR-HR(黄色), HR-LR(绿色), HR-HR(橙色), LR-LR(蓝色)，这种设计加强了不同分辨率之间信息共享。该网络仍采用重建图像和原图像的均方误差作为损失函数，T个残差块的输出分别和输入低分辨率图像相加得到T重建结果，再把它们平均得到最终的输出，这一点与DRCN类似。实验结果表明使用参数共享的网络比不使用的重建图像PNSR值高0.2db，双状态循环网络比单状态循环网络高出0.3db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之间信息传递，充分利用不同分辨率图像之间的分层特征能够帮助更好的恢复图像的高频细节，这一点也是在文献[69,71,73]所提出的算法核心思想所在，只是实现的方式不同。

表5 DSRN网络的传递函数及其作用和实现方式

|  |  |  |
| --- | --- | --- |
| 传递函数 | 作用 | 实现方式 |
|  | 将输入低分辨率图像x变换到特征空间 | 含跳连接的两层卷积层 |
|  | 从LR特征空间到HR特征空间传递信息 | 反卷积层 |
|  | 从HR特征空间到LR特征空间传递信息 | 卷积层 |
|  | 从HR特征空间到HR特征空间传递信息 | 标准残差模块 |
|  | 从LR特征空间到LR特征空间传递信息 | 标准残差模块 |
|  | 将预测的HR特征转换为图像HR图像 | 一层卷积层 |

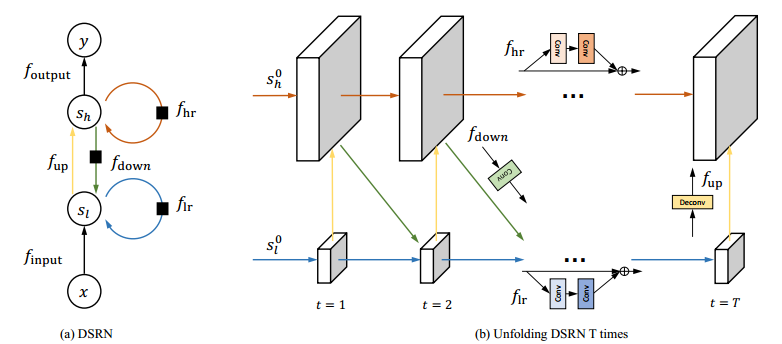


图15 （a）DSRN网络结构图；（b）展开的DSRN网络结构图

（7）MSRN（multi-scale residual network）

文献[69][[[69]](#endnote-69)]提出一种可以利用多尺度下图像分层特征实现任意尺度放大的图像超分辨率重建网络MSRN，如图所示，网络结构与MemNet[72]类似，只是基本构成模块和模块内部连接不同。网络主要分为两部分：1）特征提取部分：特征提取模块由多尺度残差模块MSRB（multi-scale residual block） 和分层特征融合HFFS（hierarchical feature fusion structure）两部分组成，MSRB主要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\*1卷积核组成瓶颈层提取各层次特征中的有用信息，实现层次特征融合；重建部分主要完成图像的放大，pixelshuffle层使得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，但模型的参数量只有它的1/7，极大的减少了网络的训练负担。

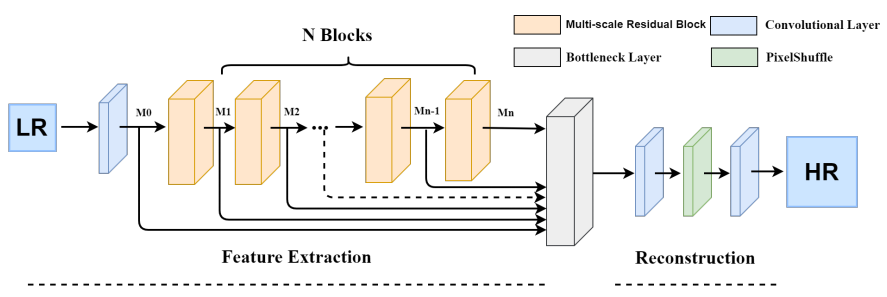


图16 MSRN网络结构图

### 3.2.2 基于密集连接的网络

2016年，有学者提出密集连接网络DenseNet(Densely Connected Convolutional Networks)[[[70]](#endnote-70)]，DenseNet 的基本思路与残差网络一致，但它组成单元是密集块(dense block)，结构如图所示，将密集块中每一层的特征都输入给之后的所有层，使所有层的特征都串联(concatenate)起来，而不是像ResNet那样直接相加。这样的结构给整个网络带来了减轻梯度消失问题、加强特征传播、支持特征复用、减少参数数量的优点。DenseNet的出现为SR问题的研究注入了新的活力。

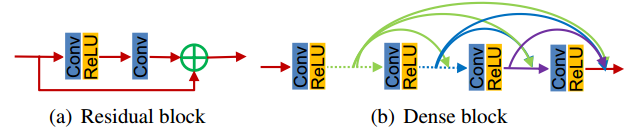


图17 （a）残差块；（b）密集块

（1）SRDenseNet（Image Super-Resolution Network Using Dense Skip Connections）

SRDenseNet[[[71]](#endnote-71)]首先将密集块结构应用到了超分辨率问题上。SRDenseNet可以分成四个部分，其网络结构如图18所示：一个卷积层学习低层的特征；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，逐级学习高层的特征；通过反卷积层对图像上采样；最后通过一个卷积层生成高分辨率输出。实验结果表明，借助密集块可以将低层特征和高层特征相结合来图像超分辨率重建性能，使用不同层之间的密集跳连接可以利用不同深度层的特征之间的信息互补性，进一步提升重建图像质量。

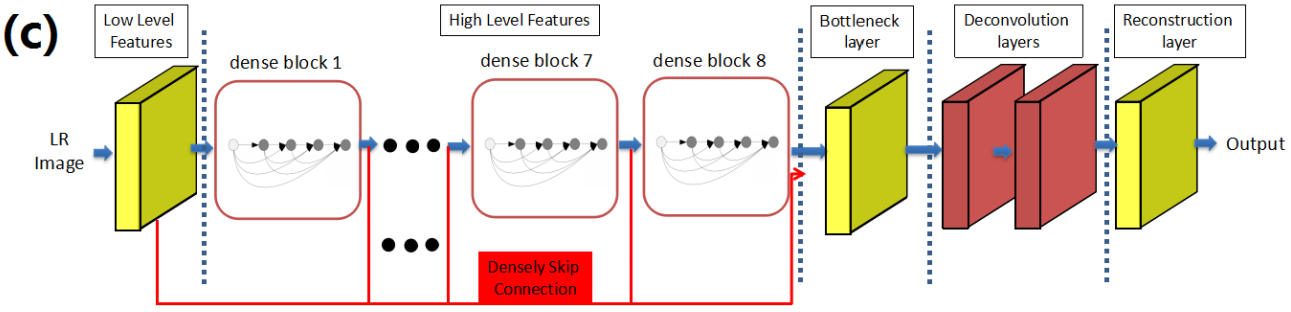


图18 SRDenseNet网络结构

（2） MemNet(deep persistent memory network)

MemNet 网络[[[72]](#endnote-72)]是一个由递归单元（recursive unit）和门控单元（gate unit）组成的内存块作为基本单元的深度持续记忆网络，与DRCN网络结构极相似，它主要解决深度网络中的长期依赖问题（long-term dependency problem），考虑了前面的状态／层对后续的状态／层的影响，可以用于图像去噪，图像超分辨率重建和PEG解锁问题。

（3）RDN（Residual Dense Network for Image Super-Resolution）

MemNet前面层到后面层的直接连接，限制了长期连接的能力，局部信息不能完全被利用。此外，MemNet网络在高分辨率空间中提取特征，增加了计算复杂度。考虑到之前的网络（VDSR[62],DRCN[63],MemNet等）没有充分利用低分辨率图像的分层特征，文献[73][[[73]](#endnote-73)]提出一种结合局部残差学习和全局残差学习，能充分利用所有层携带的信息的网络--RDN。 该网络主要有四个部分，网络结构如图所示：1）浅层特征提取部分：对于输入的低分辨率图片，使用两个卷积层提取浅层特征，2） 局部特征提取部分：然后使用串联的一系列RDB（结构如图19所示）提取丰富的局部特征，允许前一个RDB（Residual dense block）单元的输出直接连接到后面所有的RDB单元（称为密集连接，dense connections）建立连续记忆机制，最大程度的保留前面一层的信息，传递层间信息。3）局部特征融合模块：利用RDB的局部特征融合去除局部特征中冗余的信息，学习更多有效的特征，增强网络的表达能力。4）全局特征融合部分：对获得的密集局部特征进行全局融合，从而自适应地学习全局密集分层特征。前面所有的操作均在低分辨空间中实现，使得计算复杂度小，加快运算速度 5）图片的重建和放大。

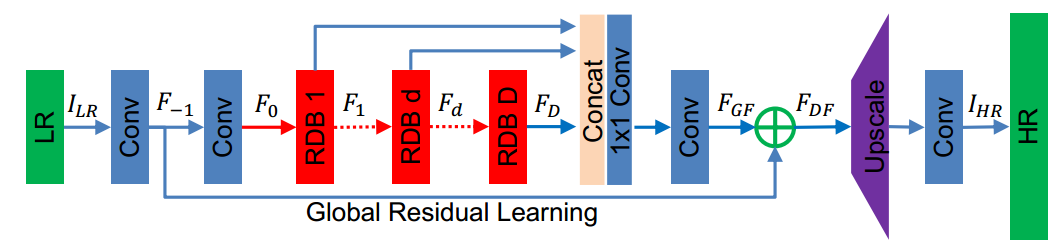


图19 RDN网络结构图

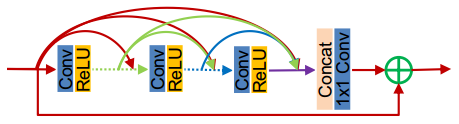


图20 RDB单元结构图

由于BN层会导致GPU内存大量消耗，增加计算复杂度，池化层会丢弃像素级信息，因此，RDN网络设计时取消了BN层和池化层。该网络能对不同的退化模型如模糊下采样，降采样混合高斯噪声和双三次降采样得到的低分辨率图像进行高分辨率重建，得到不错的效果。

在基于残差结构的网络DRCN[63]和基于密集连接的网络SRDenseNet[71]基础上，各种新的网络结构不断被提出来，主要朝两个方向发展：1）前期主要是通过加深网络提升性能，这种方法主要需要研究的问题是如何在加深网络提升性能的同时减轻网络参数量，加快网络训练速度。这种方法（VDSR[62],DRCN[63]）网络层间信息流动性差，不能挖掘图像不同层次的特征，但图像重建速度质量比SRCNN好；2）过渡到主要从内部模块间的连接和模块内部各层的连接方式（跳连接，密集连接和密集跳连接）及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，充分挖掘利用不同尺寸不同层次的图像信息，以获取高频特征用于高分辨率图像重建，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。这类网络往往连接复杂，网络规模大，参数量多，计算复杂度高。表4比较了基于残差学习典型网络DRCN和基于密集连接的系列网络，可以看出，从DRCN，SRDenseNet，MemNet到RDN，网络结构内部连接越来越复杂，逐步提高网络内部层间的信息流动性，尽可能考虑前一层的输出对后面层的影响，充分利用不同层次信息，以提升重建图像质量。

表6 存在层间信息流动的网络比较

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 算法 | 输入 | 基本模块 | 基本模块间连接方式 | 残差学习 | 层间信息流动性 | 损失函数 | 网络复杂度 |
| VDSR[62] | LR+双三次插值 | 卷积层 | 无密集连接的链式结构 | 是 | 差 | L2 | 一般 |
| SRDenseNet[71] | LR+双三次插值 | 密集块 | 过渡层 | 否 | 较好 | L1 | 一般 |
| MemNet[69] | LR+双三次插值 | 内存块 | 密集连接 | 是 | 好 | L2 | 复杂 |
| RDN[73] | LR | RDB | 全局信息融合方式 | 是 | 很好 | L2 | 复杂 |

### 3.2.3 基于GAN的SR方法

基于GAN的SR方法中，网络内部都使用了残差模块，因此将基于GAN的SR方法统一归为基于残差结构的SR方法这一大类。从Ledig [[[74]](#endnote-74)]等人提出将生成对抗网络应用于图像高分辨率重建领域以来，基于GAN的图像超分辨率重建方法越来越多，根据其算法研究内容，可以分为四大类，如图所示：1）解决大放大因子图像超分辨率重建问题的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方法的网络。



图21 基于GAN的SR方法分类

基于GAN网络的SR方法较其他CNN网络结构的SR方法最大的不同点在于损失函数。因此这个部分先介绍GAN网络涉及的损失函数，然后介绍各类基于GAN的SR网络。

（a）GAN网络的生成对抗损失

GAN网络的生成对抗损失由生成器的生成损失和判别器的对抗损失两部分组成，其表达式如式所示。生成损失和对抗损失分别如式8，式9所示。

（8）

（9）

Arjovsky 等人[[[75]](#endnote-75)]对GAN网络的损失函数进行改进，得到WGAN网络，其损失函数如式所示，WGAN网络解决了训练不稳定的问题，其生成损失和对抗损失分别如式10，式11所示。在基于GAN网络的SR问题中，有文献[81,87]使用WGAN损失代替GAN损失。

（10）

（11）

（b）基于像素的损失

（12）

其中，表示重建图像，表示原高分辨率图像。

（c）GAN网络的感知损失

（13）

基于像素的损失并不能捕捉到输出图像和真实图像间的感知差异，使得重建图像缺乏高频细节，图像过于平滑。而在图像特征可视化[[[76]](#endnote-76)]，纹理合成[[[77]](#endnote-77)]和图像风格转换[[[78]](#endnote-78)]等领域的研究表明可以使用感知损失函数捕捉更高层次的图像特征差异来生成高质量的图像。Johnson等人[[[79]](#endnote-79)]首次在高分辨率重建问题中提出使用感知损失代替基于像素的均方误差函数，使生成图片和目标图片在语义和风格上更相似。感知损失即是图像（可以是低分辨率图像或重建图像）通过VGG16或VGG19网络提取到的图像高层次的特征与网络提取到的目标高分辨率图像特征之间的均方误差。实验结果表明该方法重建出的图像相较于基于像素损失的方法在PNSR,SSIM值上略微有所下降，但图像包含更丰富真实的高频纹理细节，带来视觉感官上的提升。从此，感知损失函数在基于GAN的SR方法被广泛采用。

这里列出的损失函数是基于所阅读文献的共性总结出来的，具体而言，各种GAN的SR方法对损失函数的选择和组合不尽相同，且每项损失函数的表达式也略微有差别。比如SRGAN[84]方法中将对经典GAN网络的对抗损失函数进行简化，再结合式15表示的感知损失函数作为其感知损失函数， Enhancenet[[[80]](#endnote-80)]中对感知损失进行数学运算得到纹理匹配损失，结合感知损失和GAN对抗损失作为网络优化目标，Super-FAN[87] 和FSRNet[88]的损失函数中分别加入表示先验知识的热图损失项和先验损失项，FSR-GAN[[[81]](#endnote-81)]网络中的感知损失函数项是低分辨率图像与重建高分辨率图像之间的感知损失。SRFeat[[[82]](#endnote-82)]网络中先利用基于像素的损失来预训练生成器，然后利用重建图像与目标图像间的感知损失、特征辨别器的对抗损以及GAN对抗损失的线性组合训练GAN网络。

表7 基于GAN的SR方法损失函数对比

|  |  |  |
| --- | --- | --- |
| 算法 | 损失函数 | 算法目标 |
| SRGAN[84] | 感知损失和对抗损失加权组合 | 大上采样因子的SR问题 |
| Enhancement[80] | 感知损失+GAN对抗损失+纹理匹配损失 | 加入纹理损失恢复真实的图像细节，减少棋盘格效应 |
| ProGAN[85] | 感知损失+GAN损失 | 4倍，8倍高分辨率重建放大 |
| SFT-GAN[86] | 感知损失和对抗损失 | 加入类别先验知识，恢复真实的高频信息 |
| Super-FAN[87] | 像素损失，感知损失，热图损失，WGAN损失加权组合 | 同时实现人脸图像的高分辨率重建和人脸特征点定位 |
| FSRNet[88] | 先验损失，GAN损失，感知损失加权组合 | 人脸特征先验能明显提升人脸图像高分辨率重建质量 |
| FSR-GAN[81] | WGAN生成损失+低分辨率图像与高分辨率图像间的感知损失函数 | 改进损失函数提升图像超分辨率网络性能 |
| SRFeat[82] | 第一阶段：基于像素的损失  第二阶段：重建图像与目标图像间的感知损失、GAN对抗损失和特征对抗损失的线性组合 | 加入特征判别器提升重建图像质量 |
| SRMD[91] | GAN损失和基于像素的损失线性加权 | 解决复杂退化模型 |
| Affgan[92] | GAN损失 | 实现传统算法 |

1. 解决大放大因子的图像超分辨率重建问题的GAN网络

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，这种重建方式适合于放大因子为2,3的情况，当放大因子为4,8时，效果不尽如人意。为实现大放大因子的图像超分辨率重建，文献[]提出使用基于CNN的图像逐级放大重建方法LapSRN，文献[][[[83]](#endnote-83)]提出一种不断迭代地进行上采样和下采样的网络--DBPN，为每个阶段的映射误差提供误差反馈机制，本质是IBP算法的思想，网络采用深度串联方式直接利用不同分辨率特征，在上下采样阶段使用密集连接方式增加特征复用，实验结果表明即使只是用浅层网络也能实现8倍放大的图像超分辨率重建。解决大放大因子的图像超分辨率重建问题的GAN网络包括SRGAN和ProGAN。

1. SRGAN(Generative Adversarial Network for SR)

Ledig [[[84]](#endnote-84)]等人提出了生成对抗网络SRGAN来恢复4倍上采样因子图像的高频细节。SRGAN 由 G (generative) 网和D (Discriminator)网组成，网络结构如图所示。G网是一个残差网络，通过低分辨率的图像生成高分辨率图像，由D网判断输入图像是由G网生成的，还是数据库中的原图像。当G网和D网达到博弈平衡时候，该网络即可用于超分辨率重建。

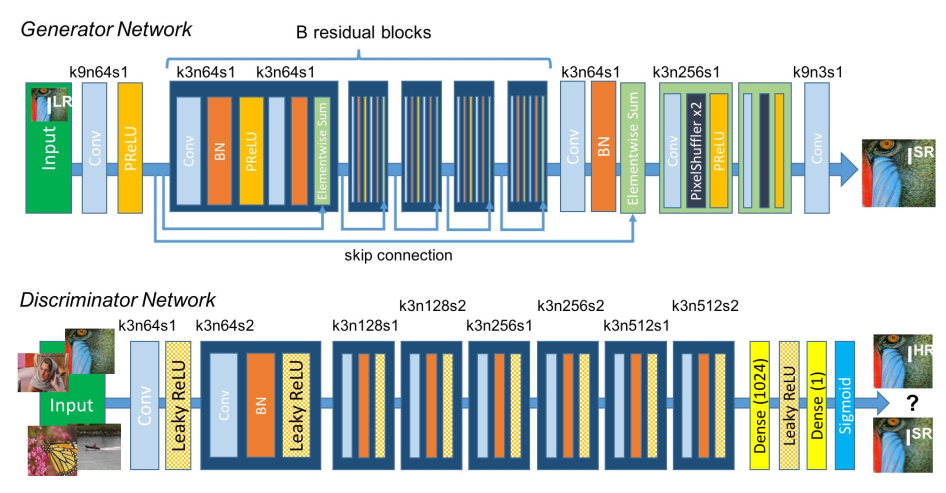


图22 SRGAN网络结构

SRGAN使用感知损失函数，感知损失由内容损失和对抗损失两部分组成。内容损失部分是与文献[]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。

（14）

将原始高分辨率图像和G网络生成的高分辨率图像输入到一个预训练好的VGG19网络中，计算这两幅图像经过VGG19网络每一层输出特征的欧式距离之和即是内容损失。

（15）

其中，表示VGG网络中特征图谱的维度。

优化对抗损失 ，使得判别器将训生成器生成图像判定为自然图像的概率最小。

(16)

高分辨率重建问题一般采用图像的PNSR值作为重建的图像质量的评价指标，图像的信噪比一定程度上可以表征图像质量的好坏，但PNSR值的图像不一定具有好的视觉效果。因此该文作者在客观评价指标PNSR, SSIM基础上增加了平均意见得分（MOS: mean-opinion-score）作为主观评价指标。实验结果表明SRGAN方法可以解决4倍上采样因子下的图像超分辨率重建问题，在Set5,Set14和BSD100测试集上，重建图像的平均PNSR值较SRCNN方法低1db左右，较DRCN方法低1.3~2db左右，但图像MOS值最高，图像的纹理细节更丰富逼真，视觉效果更好，但该网络复杂，需要训练两个网络，计算量大，耗时长。

1. ProGAN（A Fully Progressive Approach to Single-Image Super-Resolution）

图像超分辨率重建过程是图像恢复高频细节放大的过程，图像的放大主要有两类方法，一是直接放大法，包含直接对低分辨率图像上采样[55]或在网络的最后一层使用反卷积[56]或者亚像素卷积进行放大[57]，前者的图像重建在高分辨率空间进行，计算量大，后者是对前者的改进，这两种方法通仅适用于2倍，3倍的图像放大；二是逐级放大法，如LapSRN每次实现2倍采样放大，直至实现4倍，8倍高放大倍数的图像重建。文献[85][[[85]](#endnote-85)]将这种逐级重建的思想和GAN网络联合起来提出一种由渐进生成网络（progressive generator）和渐进判别网络(progressive discriminator)组成的图像多尺度逐级高分辨率重建网络—ProGAN，实验结果表明，该网络能有效实现图像4倍，8倍高分辨率重建放大，而现有SRGAN最多只能实现4倍放大，同时ProGAN网络的重建图像PNSR值比同类逐级重建网络LapSRN平均高出0.05~0.13db。

（2）加入先验知识的GAN网络

1. SFT-GAN（GAN with Spatial Feature Transform for SR）

SRGAN通过感知损失获得了包含丰富细节的高分辨率重建图像，但因为没有加入先验知识，这些细节并不自然，不真实。文献[86][[[86]](#endnote-86)]提出一种考虑类别先验的高分辨率重建网络，其网络结构如图23所示。该网络也是一种基于残差学习的生成对抗网络，生成器由条件网络和高分辨率重建网络组成，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，实时地产生空间特征调制层（Spatial Feature Transform Layer, SFT layer）共享的语义先验知识，重建网络由16个残差块和空间特征调制层组成，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片是否真实，还需要预测输入的类别信息。SFT-GAN网络也使用感知损失函数作为优化目标，实验结果表明加入图像类别先验可以促使图像恢复更真实的纹理细节。

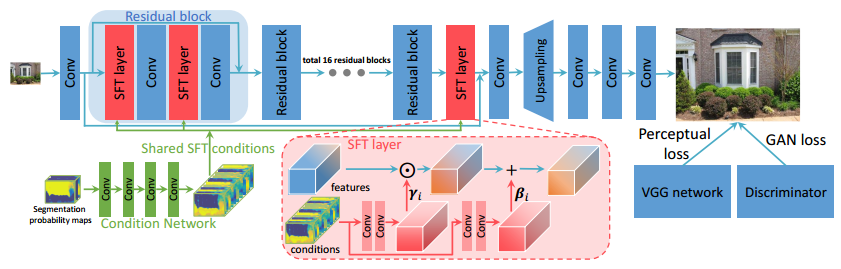


图23 SFT-GAN结构图

1. Super-FAN及FSRNet（Face Super-Resolution Network with Facial Priors）

使用图像高分辨率重建技术提升人脸图像分辨率是SR具体应用领域之一。应用GAN这种网络结构，结合人脸先验信息和图像超分辨率技术是当前人脸超分辨率重建研究的热点之一。

文献[87][[[87]](#endnote-87)]提出首个集脸超分辨率和特征点定位的端到端网络-- Super-FAN。该算法的基本思想是通过热图回归（heatmap regression）将一个人脸对齐（face alignment）的子网络FAN集成到基于GAN的图像超分辨率重建网络中并优化新的热图损失（heatmap loss），同时实现人脸图像的高分辨率重建和人脸特征点定位。Super-FAN的网络结构如图24所示，由生成器，判别器和人脸对齐子网络FAN组成，网络的损失函数由基于像素的损失，感知损失，生成对抗损失和热图损失四部分线性组合而成。加入热图损失可以加强低分辨率人脸图像和高分辨率人脸图像的结构相似性，同时实现人脸特征定位，实验结果表明该网络能对除了某些极端情况下的任意姿态人脸图像进行对齐和高分辨率重建，且重建图像的质量优于SRGAN方法。

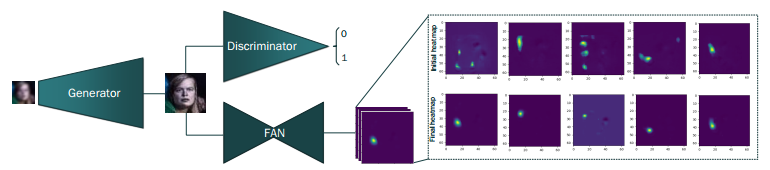


图24 Super-FAN网络结构图

文献[88][[[88]](#endnote-88)]将人脸先验知识加入网络实现人脸图像超分辨率识别的网络—FSRNet，该网络由生成器和判别器两部分组成，其中生成器网络如图所示，粗粒度网络恢复一个粗精度的高分辨率图片，然后将该图像分别送入细粒度网络的两个分支，第一个分支是一个用于提取图像特征的细粒度图像超分辨率编码器，第二个分支是一个从粗粒度重建图像中估计人脸特征点和分割信息先验知识的先验估计网络，最后两个分支结果汇总至一个细粒度超分辨率解码网络中，重构出目标高分辨率图像。FSRNet判别器网络结构和功能与SRGAN的判别器一致。该网络的损失函数在感知损失和GAN损失的基础上加入了先验损失，实验结果表明先验估计网络是整个网络的核心所在，人脸特征先验能明显提升人脸图像高分辨率重建质量，且分割先验相比人脸特征点先验包含更多信息，更能显著影响网络性能；全局分割图比局部分割图更有用；包含人脸特征点越多，重建图像质量越好。

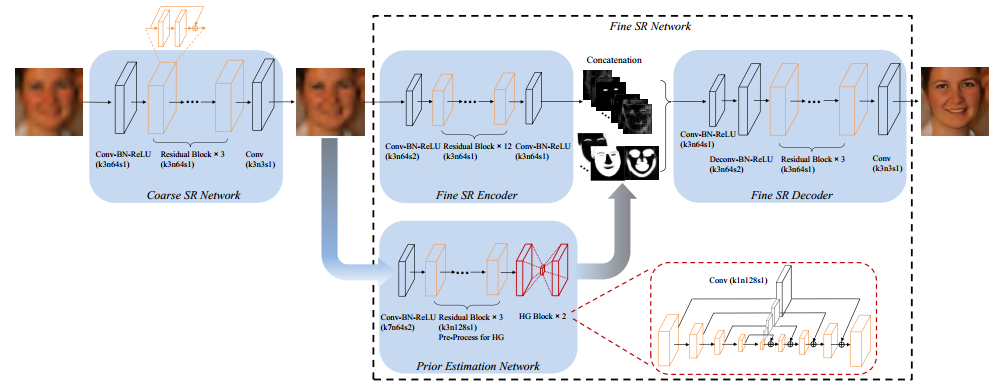


图25 FSRNet网络结构图

以上提到的两种人脸图像超分辨率重建网络是目前效果最好的两种方法，但即使是这样，重建图像的PNSR和SSIM值也比其他自然图像低，这主要是由于在低分辨率情况下难以估计出人脸特征点，导致人脸对齐效果不够准确。相对于其他自然图像，人脸本身这个类别具有丰富的先验信息，如何更有效地挖掘利用这些先验知识以重建出高质量的人脸特征图是未来研究方向之一。

表8给出了对加入先验知识的GAN算法的比较，可以看出图像先验知识通过网络，损失函数的形式加入到图像超分辨率重建网络中，各网络的损失函数也在SRGAN的感知损失函数基础上综合考虑基于像素，基于高层次特征，基于先验的各类损失，从而更好地提升重建图像质量。

表8 加入先验知识的GAN网络比较

|  |  |  |
| --- | --- | --- |
| 算法 | 先验知识及其加入方式 | 损失函数 |
| SFT-GAN[86] | 类别先验：  条件网络 | 感知损失函数 |
| Super-FAN[87] | 人脸特征点先验：  FAN网络及热图损失函数 | 基于像素的损失，感知损失，生成对抗损失和热图损失四部分线性组合 |
| FSRNet[88] | 人脸特征及分割先验：  先验估计网络及先验损失函数 | 感知损失，生成对抗损失及先验损失三部分线性组合 |

（3）学习复杂退化模型的GAN网络

现今图像超分辨率重建方法使用的低分辨率训练集一般是对高分辨率图片进行下采样得到的，忽略了传感器噪声，图像压缩，运动模糊等因素，这个过程称为理想的退化模型，而自然图像的退化过程往往很复杂且未知，为了能适应复杂退化模型的超分辨率问题求解，Shocher等人[[[89]](#endnote-89)]提出了图像自相似性进行复杂退化情况下的高分辨率图像重建的自监督卷积神经网络方法ZSSR。文献[90][[[90]](#endnote-90)]提出了一种维度拉伸策略使得单个卷积超分辨率网络能够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，训练得到可以处理多个甚至是退化空间不均匀的退化类型的超分辨网络模型。文献[91][[[91]](#endnote-91)]提出使用一个基于GAN的SR方法结构解决该问题。该网络由一个学习高分辨率图像退化为低分辨率图像的GAN网络（High-to-Low GAN）和一个学习从低分辨率图像重建高分辨率图像的GAN网络（Low-to-High GAN）组成，网络结构如图所示，首先利用不配对的高低分辨率图像训练High-to-Low GAN，学习图像的退化过程，训练完成后将低分辨率图像和与之对应的高分辨率图像作为Low-to-High GAN网络输入完成图像高分辨率重建，两个网络的损失函数均由GAN损失和基于像素的损失线性加权而成，实验结果表明虽然该文献仅对人脸这一类图像进行实验，但该方法对经过自然复杂退化过程得到的图片有很好的重建效果，提升了SR技术应用于实际问题的可能性。

（4）与传统SR方法的结合的GAN网络

传统的MAP算法必须利用图像先验对图像的分布建模，但图像分布往往不可预知，导致建模过程困难，建立的模型不准确使得重建图像质量不佳。文献[92][[[92]](#endnote-92)]提出一种将GAN网络的思想与传统基于空域的SR方法--MAP算法结合的SR方法-- AffGAN（affine projected GAN），将原本的极大似然估计问题的目标函数转化为求解重构图像的分布与目标HR图像的KL散度--的极小值问题，最小化等同于最小化GAN损失函数，通过对样本加入实例噪声（instance noise）的方法增加GAN网络训练的稳定性，获取具有良好视觉效果的重建高分辨率图像。

## 3.3 基于其他深度学习模型的SR方法

卷积神经网络被广泛引用于SR问题的研究中，与此同时，也有学者提出将受限玻尔兹曼机（Restricted Boltzmann Machine, RBM），深度信念网络（Deep Belief Network, DBN），自编码器（AutoEncoder, AE）等网络结构用于解决图像的高分辨率重建问题。

文献[93][[[93]](#endnote-93)]提出一种使用RBM训练稀疏字典对实现图像高分辨率重建的方法。其中RBM的可见层由高低分辨率图像块构成，隐藏层表示高低分辨率图像字典对共享的稀疏系数，通过极大似然估计得到隐藏层与可见层间的连接参数，即字典元素。采用对比散度算法加快训练过程。可以通过增加RBM的隐藏层层数构成深度玻尔兹曼机（Deep Boltzmann Machine, DBM）。该方法的本质仍是稀疏表示[51]，只是将字典的训练过程用网络完成，是传统方法和深度学习方法的结合。

深度信念网络DBN由多个RBM层组成。Zhou等人[[[94]](#endnote-94)]利用深度信念网络的层次结构来学习低分辨率图片到高分辨率图像之间的非线性映射关系。文献[94]通过训练和重建两个步骤完成图像高分辨率重建。在训练过程中，将HR图像分为图像块，变换到DCT域，得到的二维的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，利用训练得到的DBN来恢复丢失的高频信息。最后通过DCT逆变换获得重建图像。

文献[95][[[95]](#endnote-95)]提出一种基于堆协同局部自编码实现图片逐级放大的网络DNC(Deep Network Cascade)。该网络中由基本模块SCLA级联而成，每个基本模块都由CLA （collaborative local auto-encoder）和NLSS(non-local self-similarity)单元组成，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[96][[[96]](#endnote-96)]提出了由对称的卷积层-反卷积层构成的自编码器网络RED(very deep Residual Encoder-Decoder Networks)。网络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。卷积层用来获取图像的抽象内容，反卷积层用来放大特征尺寸并且恢复图像细节。

## 3.3 网络共性分析与差异比较

表总结对比了对基于深度学习的代表性SR方法。

表9 基于深度学习的SR方法

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 算法 | 输入 | 多尺度 | 残差学习 | 训练策略 | 先验知识 | 重建方式 | 损失函数 |
| SRCNN[55] | LR+bicubic | N | N | SGD | -- | 直接 | L2 |
| FSRCNN[56] | LR | Y | N | SGD | -- | 直接 | L2 |
| ESPCN[57] | LR | Y | N | -- | -- | 直接 | L2 |
| SRCNN-Pr[58] | LR | Y | N | SGD | 梯度先验  自相似性 | 直接 | L2 |
| SCN[59] | LR+bicubic | Y | N | SGD | 稀疏先验 | 逐级 | L2 |
| DEGREE[60] | LR | N | Y | SGD  递归监督 | 边缘先验 | 直接 | L2 |
| ZSSR[61] | LR | N | N |  | 自相似性 | 直接 | L1 |
| VDSR[62] | LR+bicubic | Y | Y | SGD  梯度裁剪 | -- | 直接 | L2 |
| DRCN[63] | LR+bicubic | N | N | SGD  递归监督 | -- | 直接 | L2 |
| DRRN[64] | LR+bicubic | N | Y | SGD  梯度裁剪 | -- | 直接 | L2 |
| EDSR[65] | LR | N | Y | Adam | -- | 直接 | L1 |
| LapSRN[66] | LR | Y | Y | Momentum | -- | 逐级 | charbonnier |
| DSRN[68] | LR+bicubic | Y | Y | SGD | -- | 直接 | L2 |
| MSRN[69] | LR+bicubic | Y | Y | ADAM |  | 直接 | L1 |
| SRDenseNet[71] | LR+bicubic | N | N | 否 | -- | 直接 | L1 |
| MemNet[72] | LR+bicubic | N | Y | SGD | -- | 直接 | L2 |
| RDN[73] | LR | N | Y | Adam | -- | 直接 | L2 |
| SRGAN[84] | LR | N | Y | Adam | -- | 直接 | L2 |
| ProGAN[85] | LR | Y | Y | Curriculum Learning | -- | 逐级 | L2 |
| Enhancenet[80] | LR | N | Y | Adam | -- | 直接 | L2 |
| SFT-GAN[86] | LR | N | Y | Adam | 类别先验 | 直接 | L2 |
| Super-FAN[87] | LR | N | Y | rmsprop | 人脸特征点先验 | 直接 | L2 |
| FSRNet[88] | LR+bicubic | N | Y | RMSprop | 人脸特征及分割先验 | 直接 | L2 |
| SRFeat[82] | LR | N | Y | Adam | -- | 直接 | L2 |
| SRMD[91] | LR | N | Y | Adam | -- | 直接 | L2 |
| Affgan[92] | LR | N | Y | GD | -- | 直接 | GAN损失 |
| RBM[93] | LR+bicubic | N | N | 对比散度 | 稀疏先验 | 直接 | -- |
| DBN[94] | LR+bicubic | N | N | SGD | -- | 直接 | -- |
| RED[95] | LR | N | Y | Adam | -- | 直接 | L2 |
| DNC[96] | LR | Y | N | BFGS | -- | 逐级 | L2 |

对表中符号说明如下：

1. ‘BI(bicubic)’表示双三次插值，’N’ 表示是，‘Y’表示否；
2. Momentum, SGD, Adam, BFGS均是优化算法；
3. L2表示L2范数损失函数，L1表示L1范数损失函数，charbonnier损失函数形式见式7，对于基于GAN的SR方法的损失函数已经在表7给出。

从表中可以得出如下结论：

1. 在网络选择上，基于卷积神经网络的SR方法是主流，也有部分学者尝试使用其他结构网络，如受限玻尔兹曼机，深度置信网络，自编码器网络等进行图像超分辨率重建，重建图像质量取得了不同程度的提高。
2. 在网络设计和训练策略上，基于CNN的SR方法从SRCNN[55]开始，到第一个使用残差结构的VDSR网络，到基于密集连接的网络SRDenseNet[71]，再到基于GAN的SR方法，图像超分辨率重建网络朝着向更深，更宽，连接更复杂的发展趋势。深度网络能够充分利用图像上下文信息，扩大图像视野域，提取图像高层特征，深入挖掘图像信息，复杂的连接形式能够加强层间的信息传递和特征传播，充分利用图像的层次特征。但这会带来训练慢，难收敛的问题。通常使用跳连接，密集连接，残差学习，梯度裁剪，参数共享等策略减轻网络负担，加快学习速率，缓解梯度消失/爆炸。
3. 一般对高分辨率图像下采样得到的低分辨率图像作为训练集和测试集，这个过程中，图像退化模型简单，为此，ZSSR[61]，SFT-GAN[86]， SRMD[91]等网络用于重建存在光学模糊，运动模糊，传感器噪声等复杂退化因素的自然图像。
4. 在SR网络输入上，一般将低分辨率图像或对其上采样得到的目标尺寸大小的低分辨率图像作为网络输入。将后者作为输入会导致整个网络的运算在高分辨率空间进行，提高了计算复杂度，网络负担加大，也会使得重建图像边缘模糊，丢失细节特征。
5. 一些网络能够灵活的实现图像不同倍数的高分辨率重建，满足不同的放大需求。
6. 领域先验知识可以用于指导基于深度学习的SR方法。如SRCNN-Pr[58], LapSRN[66], SFT-GAN[86], DNC[96]等网络利用图像不同尺度间的自相似性来恢复丢失的高频细节，实现图像的逐级高分辨率重建。而SCN[59], SFT-GAN[86], Super-FAN[87]和FSRNet[88]通过增加子网络分别加入稀疏先验，类别先验及人脸特征先验知识，用来指导超分辨率重建。SRCNN-Pr[58]网络将图像梯度先验知识反映在边缘损失和纹理损失函数中，指导网络训练。
7. 一般的深度学习方法适用于图像2倍，3倍放大，大上采样因子（如4倍，8倍）的图像高分辨率重建主要通过LapSRN[66]和ProGAN[85]网络中的逐级放大策略。
8. 损失函数的设计。大部分网络的损失函数是基于像素的L1, L2型损失函数。L1型损失函数收敛性比L2好，针对L2型损失函数带来的振铃现象，LapSRN[66]提出使用charbonnier损失函数代替获得更好的图像细节。基于像素差异的损失函数并不能体现到输出图像和真实图像间的感知差异，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，获得视觉上的提升。

# 相关技术讨论

## 4.1 常用数据库

图像超分辨率重建问题中最常用的数据有91-images，DIV2K dataset[[[97]](#endnote-97)], Set5[[[98]](#endnote-98)], Set14[[[99]](#endnote-99)], BSD100[[[100]](#endnote-100)]和Urban100[[[101]](#endnote-101)]。91-images由91张自然图片，通常需要分为图像块；DIV2K dataset[[[102]](#endnote-102)]由800张训练图像，100张验证图像和100张测试图像组成，Set5和Set14分别由5张和14张图片组成，BSD 100和Urban100均由100张图片组成。

## 4.2 常用评价指标

常用来客观评价重建图像质量的指标包括峰值信噪比（Peak Signal to Noise Ratio, PNSR）和结构相似性（Structural Similarity, SSIM）SSIM，主观评价一般是根据观察者从视觉感官效果出发，对重建图像的优劣打分，无统一的标准。

1. PNSR

峰值信噪比代表了信号的最大功率和噪声功率的比值，用来度量重建高分辨率图像与真实高分辨率图像之间的像素差值，单位为 dB。

（17）

其中Peak 是图像像素值可取值的最大值，MSE表示重建图像与目标图像间像素的均方误差，见式17。

1. SSIM

图像间的像素相似度不能准确反映重建效果的质量，因此提出使用SSIM从亮度、对比度、结构三方面度量图像间的相似性，取值范围为[0,1]，SSIM定义如式18所示，其值越大就表示两幅图像结构相似程度越高，最大的值可取 1。

（18）

其中和分别代表 x 和 y 的平均值，表示对亮度的估计，和分别代表 x和 y 的标准差，表示对比度的估计，分母是协方差，表示对结构的估计，参数和是常数。

## 基于深度学习的典型SR方法实验结果与分析

从一个基于深度学习的SR网络—SRCNN于2014诞生以来，基于深度学习的SR方法层出不穷，表10列出了放大因子为2,3,4下的情况下，基于深度学习的典型SR方法在Set5, Set14, BSD100, Urban100测试集中实验结果均值，并对最好的结果进行了加粗。

表10 基于深度学习的典型SR方法实验结果比较

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 数据集 | 放大倍数 | SRCNN[55] | VDSR[62] | DRCN[63] | LapSRN[66] | SRGAN[84] | RED[95] | MDSR[65] | RDN[73] |
| Set5 | ×2  ×3  ×4 | 36.66/0.9542  32.75/0.9090  30.48/0.8628 | 37.53/0.9587  33.66/0.9213  31.35/0.8838 | 37.63/0.9588  33.82/0.9226  31.53/0.8854 | 37.52/0.9591  33.82/0.9227  31.54/0.8855 | --  --  32.05/0.8910 | 37.66/0.9599  33.82/0.9230  31.51/0.8869 | 38.11/0.9602  **38.11/0.9602**  **32.50**/0.8973 | **38.24/0.9614**  34.71/0.9296  32.47/**0.8990** |
| Set14 | ×2  ×3  ×4 | 32.42/0.9063  29.28/0.8209  27.49/0.7503 | 33.03/0.9124  29.77/0.8314  28.01/0.7674 | 33.04/0.9118  29.76/0.8311  28.02/0.7670 | 33.08/0.9130  29.79/0.8320  28.19/0.7720 | --  --  28.53/0.7804 | 32.94/0.9144  29.61/0.8341  27.86/0.7718 | 33.85/0.9198  30.44/0.8452  28.72/0.7857 | **34.01/0.9212**  **30.57/0.8468**  **28.81/0.7871** |
| BSD100 | ×2  ×3  ×4 | 31.36/0.8879  28.41/0.7863  26.90/0.7101 | 30.76/0.9140  27.14/0.8279  25.18/0.7524 | 31.85/0.8942  28.80/0.7963  27.23/0.7233 | 31.80/0.8950  28.82/0.7973  27.32/0.7280 | --  --  27.57 / 0.7354 | 31.99/0.8974  28.93/0.7994  27.40/0.7290 | 32.29/0.9007  29.25/0.8091  27.72/0.7418 | **32.34/0.9017**  **29.26/0.8093**  **27.72/0.7419** |
| Urban100 | ×2  ×3  ×4 | 29.50/0.8946  26.24/0.7989  24.52/0.7221 | 30.76/0.9140  27.14/0.8279  25.18/0.7524 | 30.75/0.9133  27.15/0.8276  25.14/0.7510 | 30.41/0.9101  27.07/0.8272  25.21/0.7553 | --  --  26.07 / 0.7839 | --  --  -- | 32.84/0.9347  28.79/**0.8655**  **26.67/0.8041** | **32.89/0.9353**  **28.80**/0.8653  26.61/0.8028 |

SRCNN是第一个将深度学习和SR方法结合起来的方法；VDSR首次将残差学习思想应用于SR问题，通过加深网络，扩大图像视野域，相较SRCNN，重建图像质量得到了提升，DRCN从训练策略上减少深度网络的参数量，提升网络训练速度；LapSRN网络提出了逐级放大重建高分辨率图像的思想，在4倍上采样情况下表现较SRCNN和VDSR方法都好，为实现大上采样因子的图像高分辨率重建奠定了基础；SRGAN是第一个利用GAN结构结合感知损失函数解决SR问题的网络，为后续基于GAN结构的SR方法提供了思路，RED是一个结构简单性能优良的自编码器网络，这为使用除了CNN以外的其他网络结构解决SR问题提供了可能性；MDSR利用图像自相似的领域先验知识实现图像多尺度的放大重建，从实验结果来看，图像自身的领域先验知识对提升重建图像质量具有十分重要的作用；RDN是一个通过密集连接、跳连接、局部特征融合和全局特征融合充分利用图像各层次特征的复杂网络，表中RDN是最好的SR方法，这说明网络越深，网络内部信息流动性越好，越有利于图像超分辨率重建。

# 总结与展望

图像的超分辨率重建问题经历了基于插值，基于重构和基于学习的三个发展阶段。以基于深度学习为代表的基于学习的方法是目前主流的SR方法。基于深度学习的SR方法在网络模型选择，网络结构的变化，先验知识的加入，训练函数的优化等方面不断创新，成为了研究热点。但仍存在以下问题需要解决。

1）网络的选择和设计。目前大部分基于深度学习的SR方法基于卷积神经网络模型，如何利用其它类别的深度学习模型的应用于图像超分辨率重建值得探索。此外，一般来说，网络越深，效果越好，但会导致训练困难，如何权衡网络的复杂性和重建图像质量的矛盾也是徐亚进一步研究的问题。

2）针对特定应用场合的超分辨率算法。例如人脸、 文字、 指纹、 车牌这些专用场合中具有较多的先验知识，如何利用这些先验知识指导网络结构设计，代价函数构造以及训练方式的改进，使得重建图像质量有较大提高。

3）针对复杂的图像退化模型。实验中通常对高分辨率图像下采样得到低分辨率图像，而现实生活中的低分辨率图像通常存在运动变形、光学模糊、降采样以及噪声污染等多种复杂降质因素。如何学习得到一个可以处理复杂图像退化模型的网络是未来可研究的方向之一。

4）大放大倍数的图像高分辨率重建。一般针对2倍，3倍放大因子的图像高分辨率重建取得不错的效果，当放大因子为4，8时效果急剧下降，放大因子更大时，现有方法有待改进。设计针对大放大倍数的图像超分辨率重建算法是亟待研究的方向。

**参考文献**

1. Yang S, Sun Y, Chen Y, et al. Structural similarity regularized and sparse coding based super-resolution for medical images[J]. Biomedical Signal Processing & Control, 2012, 7(6):579-590. [↑](#endnote-ref-1)
2. Zhang M, Desrosiers C, Qu Q, et al. Medical image super-resolution with non-local embedding sparse representation and improved IBP[C]// IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2016:888-892. [↑](#endnote-ref-2)
3. Wang Y H, Li J B, Fu P. Medical Image Super-resolution Analysis with Sparse Representation[C]// Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing. IEEE Computer Society, 2012:106-109. [↑](#endnote-ref-3)
4. Liu H C, Li S T, Yin H T. Infrared surveillance image super resolution via group sparse representation[J]. Optics Communications, 2013, 289(4):45-52. [↑](#endnote-ref-4)
5. Uiboupin T, Rasti P, Anbarjafari G, et al. Facial image super resolution using sparse representation for improving face recognition in surveillance monitoring[C]// Signal Processing and Communication Application Conference. IEEE, 2016:437-440. [↑](#endnote-ref-5)
6. 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. [↑](#endnote-ref-6)
7. 杨学峰. 遥感图象频域和空域超分辨重建技术研究[D]. 哈尔滨工业大学, 2011. [↑](#endnote-ref-7)
8. Hong Z, Song W, Hai T, et al. Remote Sensing Images Super Resolution Reconstruction Based on Multi-scale Detail Enhancement[J]. Acta Geodaetica Et Cartographica Sinica, 2016. [↑](#endnote-ref-8)
9. Liao R, Tao X, Li R, et al. Video Super-Resolution via Deep Draft-Ensemble Learning[C]// IEEE International Conference on Computer Vision. IEEE, 2015:531-539. [↑](#endnote-ref-9)
10. Cheng M H, Lin N W, Hwang K S, et al. Fast video super-resolution using artificial neural networks[C]// International Symposium on Communication Systems, Networks & Digital Signal Processing. IEEE, 2012:1-4. [↑](#endnote-ref-10)
11. 卓力, 王素玉, 李晓光. 图像/视频的超分辨率复原[M]. 人民邮电出版社, 2011. [↑](#endnote-ref-11)
12. Tsai R Y, Huang T S. Multiframe image restoration and registration[C]// Advances in computer vision and image processing. 1984. [↑](#endnote-ref-12)
13. Harris J L. Difiraction and resolving power. Journal of the Optical Society of America, 1964, 54(7): 931-936 [↑](#endnote-ref-13)
14. Goodman J W, Cox M E. Introduction to Fourier Optics[M]. McGraw-Hill, 1968. [↑](#endnote-ref-14)
15. Tsai R Y, Huang T S. Multiframe image restoration and registration. Advances in Computer Vision and Image Processing, 1984, 1: 317-339 [↑](#endnote-ref-15)
16. Borman S, Stevenson R. Spatial Resolution Enhancement of Low-resolution Image Sequences: A Comprehensive Review with Directions for Future Research, Technical Report, Laboratory Image and Signal Analysis, University of NotreDame, 1998 [↑](#endnote-ref-16)
17. 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. [↑](#endnote-ref-17)
18. 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):1202-1213. [↑](#endnote-ref-18)
19. Dongdong Huang，Heng Liu，A Short Survey of Image Super Resolution Algorithms，Journal of Computer Science Technology Updates，Vol. 2(2), pp. 19-29, 2015. [↑](#endnote-ref-19)
20. Hayat K. Super-Resolution via Deep Learning[J]. 2017. [↑](#endnote-ref-20)
21. Sun X, Xiao-Guang L I, Jia-Feng L I, et al. Review on Deep Learning Based Image Super-resolution Restoration Algorithms[J]. Acta Automatica Sinica, 2017, 43(5):697-709. [↑](#endnote-ref-21)
22. Yang W, Zhang X, Tian Y, et al. Deep Learning for Single Image Super-Resolution: A Brief Review[J]. 2018. [↑](#endnote-ref-22)
23. Sigitani T, Iiguni Y, Maeda H. Image interpolation for progressive transmission by using radial basis function networks[J]. IEEE Transactions on Neural Networks, 2002, 10(2):381-390. [↑](#endnote-ref-23)
24. Schultz R R, Stevenson R L. A Bayesian approach to image expansion for improved definition.[J]. Image Processing IEEE Transactions on, 1994, 3(3):233-242. [↑](#endnote-ref-24)
25. Hou H, Andrews H. Cubic splines for image interpolation and digital filtering[J]. IEEE Transactions on Acoustics Speech & Signal Processing, 1978, 26(6):508-517. [↑](#endnote-ref-25)
26. Li X, Orchard M T. New edge-directed interpolation[J]. IEEE Trans Image Process, 2001, 10(10):1521-1527. [↑](#endnote-ref-26)
27. Lin H. Research on the Algorithms of Super resolution Image Reconstruction with Wavelet and Interpolation. Xidian University (Chinese) 2005; p. 7-14 [↑](#endnote-ref-27)
28. Adamczyk K, Walczak A. Digital images interpolation with wavelet edge extractors[C]// Human System Interactions. IEEE, 2010:399-405. [↑](#endnote-ref-28)
29. 蔡念, 张海员, 张楠,等. 基于小波的双线性插值误差补偿算法的图像放大[J]. 激光与红外, 2010, 40(5):558-562. [↑](#endnote-ref-29)
30. Zhang L, Wu X. An edge-guided image interpolation algorithm via directional filtering and data fusion[J]. IEEE Trans Image Process, 2006, 15(8):2226-2238. [↑](#endnote-ref-30)
31. Shi H, Ward R. Canny edge based image expansion[C]// IEEE International Symposium on Circuits and Systems. IEEE, 2002:I-785-I-788 vol.1. [↑](#endnote-ref-31)
32. Kim S P, Bose N K, Valenzuela H M. Recursive reconstruction of high resolution image from noisy undersampled multiframes[J]. Acoustics Speech & Signal Processing IEEE Transactions on, 1990, 38(6):1013-1027. [↑](#endnote-ref-32)
33. Rhee S, Kang M G. Discrete cosine transform based regularized high-resolution image reconstruction algorithm[J]. Optical Engineering, 1999, 38(8):1348-1356. [↑](#endnote-ref-33)
34. Nguyen N, Milanfar P. An efficient wavelet-based algorithm for image superresolution[J]. 2000, 2:351-354 vol.2. [↑](#endnote-ref-34)
35. Ji H. Robust Wavelet-Based Super-Resolution Reconstruction: Theory and Algorithm[M]. IEEE Computer Society, 2009. [↑](#endnote-ref-35)
36. Irani M, Peleg S. Super resolution from image sequences[C]// [1990] Proceedings. 10th International Conference on Pattern Recognition. IEEE, 2002:115-120 vol.2. [↑](#endnote-ref-36)
37. Schultz R R, Stevenson R L. A Bayesian approach to image expansion for improved definition[J]. Image Processing IEEE Transactions on, 1994, 3(3):233-242. [↑](#endnote-ref-37)
38. Schultz R R, Stevenson R L. Improved definition video frame enhancement[C]// International Conference on Acoustics, Speech, and Signal Processing. IEEE, 1995:2169-2172 vol.4. [↑](#endnote-ref-38)
39. Chantas G K, Galatsanos N P, Woods N A. Super-resolution based on fast registration and maximum a posteriori reconstruction[J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2007, 16(7):1821. [↑](#endnote-ref-39)
40. Hennings-Yeomans P H, Baker S, Kumar B V K V. Simultaneous super-resolution and feature extraction for recognition of low-resolution faces[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2008:1 - 8. [↑](#endnote-ref-40)
41. Babacan S D, Molina R, Katsaggelos A K. Total variation super resolution using a variational approach[C]// IEEE International Conference on Image Processing. IEEE, 2008:641-644. [↑](#endnote-ref-41)
42. Farsiu S, Robinson M D, Elad M, et al. Fast and robust multiframe super resolution[J]. Image Processing IEEE Transactions on, 2004, 13(10):1327-1344. [↑](#endnote-ref-42)
43. Bose N K, Lertrattanapanich S, Koo J. Advances in superresolution using L-curve[C]// IEEE International Symposium on Circuits and Systems. IEEE, 2001:433-436 vol. 2. [↑](#endnote-ref-43)
44. Yuan Q, Zhang L, Shen H, et al. Adaptive Multiple-Frame Image Super-Resolution Based on U-Curve[J]. IEEE Transactions on Image Processing, 2010, 19(12):3157-3170. [↑](#endnote-ref-44)
45. Elad M, Feuer A. Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images[J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 1997, 6(12):1646-58. [↑](#endnote-ref-45)
46. Freeman W T, Jones T R, Pasztor E C. Example-based super-resolution[J]. Computer Graphics & Applications IEEE, 2002, 22(2):56-65. [↑](#endnote-ref-46)
47. Glasner D, Bagon S, Irani M. Super-resolution from a single image[C]// IEEE, International Conference on Computer Vision. 2009:349-356. [↑](#endnote-ref-47)
48. Freedman G, Fattal R. Image and video upscaling from local self-examples[J]. Acm Transactions on Graphics, 2011, 30(2):1-11. [↑](#endnote-ref-48)
49. Chang H, Yeung D Y, Xiong Y. Super-Resolution through Neighbor Embedding[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington. 2004:I-275- I-282 Vol.1. [↑](#endnote-ref-49)
50. M. Bevilacqua, A. Roumy, C. Guillemot, and M. Alberi-Morel, “Low-complexity single-image super-resolution based on nonnegative neighbor embedding,” in British Machine Vision Conference, BMVC 2012, Surrey, UK, September 3-7, 2012, 2012, pp. 1–10. [↑](#endnote-ref-50)
51. Yang J, Wright J, Huang T S, et al. Image super-resolution via sparse representation[J]. IEEE Trans Image Process, 2010, 19(11):2861-2873. [↑](#endnote-ref-51)
52. J. B. Huang, A. Singh, and N. Ahuja, “Single image super-resolution from transformed self-exemplars,” in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2015, pp. 5197–5206. [↑](#endnote-ref-52)
53. 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. [↑](#endnote-ref-53)
54. Timofte R, De V, Gool L V. Anchored Neighborhood Regression for Fast Example-Based Super-Resolution[C]// IEEE International Conference on Computer Vision. IEEE Computer Society, 2013:1920-1927. [↑](#endnote-ref-54)
55. Dong C, Chen C L, He K, et al. Learning a Deep Convolutional Network for Image Super-Resolution[J]. 2014, 8692:184-199. [↑](#endnote-ref-55)
56. Dong C, Chen C L, Tang X. Accelerating the Super-Resolution Convolutional Neural Network[J]. 2016:391-407. [↑](#endnote-ref-56)
57. Shi W, Caballero J, Huszár F, et al. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2016:1874-1883. [↑](#endnote-ref-57)
58. Liang Y, Wang J, Zhou S, et al. Incorporating image priors with deep convolutional neural networks for image super-resolution[J]. Neurocomputing, 2016, 194(C):340-347. [↑](#endnote-ref-58)
59. Wang Z, Liu D, Yang J, et al. Deep Networks for Image Super-Resolution with Sparse Prior[J]. 2015:370-378. [↑](#endnote-ref-59)
60. Yang W, Feng J, Yang J, et al. Deep Edge Guided Recurrent Residual Learning for Image Super-Resolution[J]. IEEE Transactions on Image Processing, 2016, PP(99):1-1. [↑](#endnote-ref-60)
61. Shocher, Assaf, Cohen, Nadav, Irani, Michal. "Zero-Shot" Super-Resolution using Deep Internal Learning[J]. 2017. [↑](#endnote-ref-61)
62. Kim J, Lee J K, Lee K M. Accurate Image Super-Resolution Using Very Deep Convolutional Networks[J]. 2015:1646-1654. [↑](#endnote-ref-62)
63. Kim J, Lee J K, Lee K M. Deeply-Recursive Convolutional Network for Image Super-Resolution[C]// Computer Vision and Pattern Recognition. IEEE, 2016:1637-1645. [↑](#endnote-ref-63)
64. 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. [↑](#endnote-ref-64)
65. Lim B, Son S, Kim H, et al. Enhanced Deep Residual Networks for Single Image Super-Resolution[C]// Computer Vision and Pattern Recognition Workshops. IEEE, 2017:1132-1140. [↑](#endnote-ref-65)
66. Lai W S, Huang J B, Ahuja N, et al. Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2017:5835-5843. [↑](#endnote-ref-66)
67. Liao Q, Poggio T. Bridging the Gaps Between Residual Learning, Recurrent Neural Networks and Visual Cortex[J]. 2016. [↑](#endnote-ref-67)
68. Han W, Chang S, Liu D, et al. Image Super-Resolution via Dual-State Recurrent Networks[J]. 2018. [↑](#endnote-ref-68)
69. Juncheng Li, Faming Fang, Kangfu Mei, Guixu Zhang; The European Conference on Computer Vision (ECCV), 2018, pp. 517-532 [↑](#endnote-ref-69)
70. Huang G, Liu Z, Laurens V D M, et al. Densely Connected Convolutional Networks[J]. 2016:2261-2269. [↑](#endnote-ref-70)
71. 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. [↑](#endnote-ref-71)
72. Tai Y, Yang J, Liu X, et al. MemNet: A Persistent Memory Network for Image Restoration[C]// IEEE International Conference on Computer Vision. IEEE, 2017:4549-4557. [↑](#endnote-ref-72)
73. Zhang Y, Tian Y, Kong Y, et al. Residual Dense Network for Image Super-Resolution[J]. 2018. [↑](#endnote-ref-73)
74. Ledig C, Theis L, Huszar F, et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network[J]. 2016:105-114. [↑](#endnote-ref-74)
75. Arjovsky M, Chintala S, Bottou L. Wasserstein GAN[J]. 2017. [↑](#endnote-ref-75)
76. Simonyan, K., Vedaldi, A., Zisserman, A.: Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint arXiv:1312.6034 (2013) [↑](#endnote-ref-76)
77. Gatys, L.A., Ecker, A.S., Bethge, M.: Texture synthesis using convolutional neural networks. In: Advances in Neural Information Processing Systems 28. (May 2015) [↑](#endnote-ref-77)
78. Gatys, L.A., Ecker, A.S., Bethge, M.: A neural algorithm of artistic style. arXiv preprint arXiv:1508.06576 (2015) [↑](#endnote-ref-78)
79. Johnson J, Alahi A, Li F F. Perceptual Losses for Real-Time Style Transfer and Super-Resolution[J]. 2016:694-711. [↑](#endnote-ref-79)
80. Sajjadi, M. S. M., Schölkopf, B., & Hirsch, M. (2016). Enhancenet: single image super-resolution through automated texture synthesis. [↑](#endnote-ref-80)
81. Weimin Tan, Bo Yan, and Bahetiyaer Bare. Feature Super-Resolution: Make Machine See More Clearly. CVPR 2018 [↑](#endnote-ref-81)
82. Seong-Jin Park, Hyeongseok Son, Sunghyun Cho, et al. SRFeat: Single Image Super-Resolution with Feature Discrimination.The European Conference on Computer Vision (ECCV), 2018, pp. 439-455 [↑](#endnote-ref-82)
83. Haris M , Shakhnarovich G , Ukita N . Deep Back-Projection Networks For Super-Resolution[J]. 2018. [↑](#endnote-ref-83)
84. Ledig C, Theis L, Huszar F, et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network[J]. 2016:105-114. [↑](#endnote-ref-84)
85. Wang Y , Perazzi F , Mcwilliams B , et al. A Fully Progressive Approach to Single-Image Super-Resolution[J]. 2018. [↑](#endnote-ref-85)
86. Wang X, Yu K, Dong C, et al. Recovering Realistic Texture in Image Super-resolution by Deep Spatial Feature Transform[J]. 2018. [↑](#endnote-ref-86)
87. Bulat A , Tzimiropoulos G . Super-FAN: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with GANs[J]. 2017. [↑](#endnote-ref-87)
88. Chen Y , Tai Y , Liu X , et al. FSRNet: End-to-End Learning Face Super-Resolution with Facial Priors[J]. 2017. [↑](#endnote-ref-88)
89. Shocher, Assaf, Cohen, Nadav, Irani, Michal. "Zero-Shot" Super-Resolution using Deep Internal Learning[J]. 2017. [↑](#endnote-ref-89)
90. Zhang K, Zuo W, Zhang L. Learning a Single Convolutional Super-Resolution Network for Multiple Degradations[J]. 2017. [↑](#endnote-ref-90)
91. Bulat A , Yang J , Tzimiropoulos G . To learn image super-resolution, use a GAN to learn how to do image degradation first[J]. 2018. [↑](#endnote-ref-91)
92. Sønderby, Casper Kaae, Caballero J , Theis L , et al. Amortised MAP Inference for Image Super-resolution[J]. 2016. [↑](#endnote-ref-92)
93. Gao J, Guo Y, Yin M. Restricted Boltzmann machine approach to couple dictionary training for image super-resolution[C]// IEEE International Conference on Image Processing. IEEE, 2014:499-503. [↑](#endnote-ref-93)
94. Zhou Y, Qu Y, Xie Y, et al. Image Super-Resolution Using Deep Belief Networks[C]// International Conference on Internet Multimedia Computing and Service. ACM, 2014:28. [↑](#endnote-ref-94)
95. Cui Z, Chang H, Shan S, et al. Deep Network Cascade for Image Super-resolution[M]// Computer Vision – ECCV 2014. Springer International Publishing, 2014:49-64. [↑](#endnote-ref-95)
96. Mao X J, Shen C, Yang Y B. Image Restoration Using Convolutional Auto-encoders with Symmetric Skip Connections[J]. 2016. [↑](#endnote-ref-96)
97. R. Timofte, E. Agustsson, L. Van Gool, M.-H. Yang, L. Zhang, et al. Ntire 2017 challenge on single image superresolution: Methods and results. In CVPR 2017 Workshops. [↑](#endnote-ref-97)
98. Bevilacqua, M., Roumy, A., Guillemot, C., & Alberi-Morel, M. (2012). Low-Complexity Single-Image Super-Resolution based on Nonnegative Neighbor Embedding. BMVC. [↑](#endnote-ref-98)
99. Zeyde, R., Elad, M., & Protter, M. (2010). On Single Image Scale-Up Using Sparse-Representations. Curves and Surfaces. [↑](#endnote-ref-99)
100. Martin, D.R., Fowlkes, C.C., Tal, D., & Malik, J. (2001). A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics. ICCV. [↑](#endnote-ref-100)
101. Huang, J. B. , Singh, A. , & Ahuja, N. . (2015). Single image super-resolution from transformed self-exemplars. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society. [↑](#endnote-ref-101)
102. R. Timofte, E. Agustsson, L. Van Gool, M.-H. Yang, L. Zhang, et al. Ntire 2017 challenge on single image superresolution: Methods and results. In CVPR 2017 Workshops. [↑](#endnote-ref-102)