

# 基于深度学习的超分辨率算法综述

刘雪晨

21821276

浙江大学

3140101216@zju.edu.cn

## 摘要 (Abstract)

图像超分辨率技术是计算机视觉领域的一个经典问题,传统的超分辨率算法常由特征提取、特征变换、特征重构系列操作构成,在不同阶段使用不同方法是传统超分辨率算法的主流研究思路。近年来,随着卷积神经网络被应用于图像分类、检测等任务中,其端到端的结构整合了上述工作流程,且有着不俗的特征提取与转化能力,近期许多超分辨率研究就基于卷积神经网络展开,本文针对基于深度学习的超分辨率算法的发展历程进行了调研,分析其在网络结构、损失函数等方面的演变,并通过实验比较上述网络的性能演化。实验表明,与大多数计算机视觉任务相似,更深的网络层数带来效果的提升;而针对超分辨率问题专门设计的网络结构也为这一任务提出了解决思路。

## 1. 研究背景

单幅图像超分辨率 (Single image super-resolution, SR) 是计算机视觉领域一个经典的问题。传统的超分辨率算法中,一个热门的研究领域是基于实例

(example-based) 的超分辨率方法,一般是利用相同图像的局部相似性[1,2],或是从外部低分辨率和高分辨率样本对之间学习映射函数[3,4],基于外部实例的方法需要丰富的样本进行学习,且在建模方面有很大难度。

基于外部实例的图像超分辨率算法中,一个代表性的算法是基于稀疏编码 (sparse-coding-based) 的方法[5]。该方法由几个步骤组成流水线操作,首先,从图像提取密集且有重叠的图像块并进行预处理(如减去均值)。然后这些图像块经由低分辨率字典编码,编码后的数值被映射到高分辨率字典中,反向映射成为高分辨率的图像块,最终将密集、重叠的高分辨率图像聚合,输出最终的完整高分辨率图像。在此类方法中,相关研究主要集中在映射词典的优化[5]和建模方法的设计上[3]。

上述算法由一个特征提取、特征变换(映射)、特征重构流水线组成,其工作流程与卷积神经网络(Convolutional Neural Network, CNN)类似[8],CNN

通常由许多大小不同的卷积核和稠密链接的全连接层组成,应用到图像处理领域,网络底层的卷积核首先从输入图像中提取特征,然后经由全连接层完成一系列非线性变换,最后根据不同任务完成图像输出图像分类结果或是重构结果。CNN将图像处理流程中的特征提取、特征变换、特征重构过程隐含在网络内部,无需人工筛选特征,特征变换也在网络内部完成,是一个端到端(End-to-End)的图像处理系统。

自从Dong等人[9]首先引入浅层卷积神经网络来处理超分辨率问题,近年来基于深度学习的方法由于其优越的重建性能而引起了广泛的关注,与传统超分辨率算法相比,基于深度学习的超分辨率算法使用端到端的神经网络模型完成特征提取、特征映射和特征重构流程,模型结构简洁,且有着不俗的表现,本文将对近年来使用深度学习的超分辨率算法做一简述,并简单介绍其发展历程。

## 2. 相关工作

传统基于实例学习的算法学习低分辨率和高分辨率图像块之间的映射,相关研究在如何学习字典映射表示和如何对映射进行建模之间存在着差异。Freeman等人[6]将字典直接表示为低/高分辨率的图像对,并在低分辨率空间中找到输入图像块的最近邻(Nearest Neighbor, NN),其相应的高分辨率图像块用于重建SR图像。Chang等人[3]使用流形嵌入技术作为最近邻策略的替代方案。Yang等人[7]更是使用更复杂的稀疏编码方法来完成字典之间的映射。这种基于系数编码的方法是传统超分辨率算法中效果最好的一类方法。

卷积神经网络可以追溯到几十年前[8],并在几年前爆发了空前的研究热潮,部分原因在于其在图像分类方面的成功[10]。在将卷积神经网络用于图像超分辨率领域之前,已经有一些使用深度学习技进行图像恢复的研究。多层感知器(multi-layer perceptron, MLP)由许多层一一相连的神经元神经元组成,被使用于自然图像去噪[11]和后去模糊去噪[12],另外,更深一步的应用使用卷积神经网络对自然图像去噪[13]。

Dong等人[9]证明了基于稀疏编码的超分辨率方法[5]可以被看作是一个卷积神经网络。如图1所示,在基

于稀疏编码的方法中,考虑从输入图像中提取 $f_1 * f_1$ 的低分辨率图像块,将该图像块减去均值,然后投影到低分辨率字典上,如果字典大小为 $n_1$ ,这相当于使用 $n_1$ 个线性滤波器( $f_1 * f_1$ )对输入图像滤波,如图1中左侧所示。基于稀疏编码器的方法接下来将大小为 $n_1$ 的字典映射到大小为 $n_2$ 的字典中,这一过程可理解为非线性映射,对应卷积神经网络中卷积核大小 $1 * 1$ 的卷积操作,如图1中间部分所示。接下来将上述 $n_2$ 个系数投影到另一个字典上以产生高分辨率的图像块,最后对重叠的高分辨率图像块进行平均,这相当于 $n_2$ 特征图上的线

性卷积。如果重建后的高分辨率图像块大小为 $f_3 * f_3$ ,则对应的线性滤波器尺寸即为 $f_3 * f_3$ ,如图1右侧所示。

上述讨论表明基于稀疏编码的超分辨率算法可以被视为一种卷积神经网络,更进一步地,在基于稀疏编码的方法中,并非上述所有步骤的优化都经过了研究,而在卷积神经网络中,低分辨率字典、高分辨率字典、非线性映射以及图像正则化和重构算法都涉及到要优化的滤波器,而滤波器的优化随着网络的反向传播(backpropagation)自动完成了优化,因此卷积神经网络的算法优化了包含所有操作的端到端的映射,因此也取得了令人瞩目的效果。

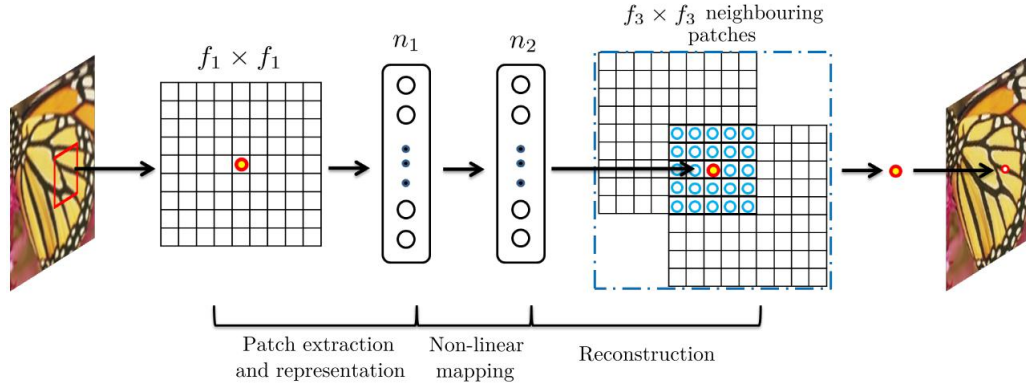


图1 从卷积神经网络角度解释基于稀疏编码的方法

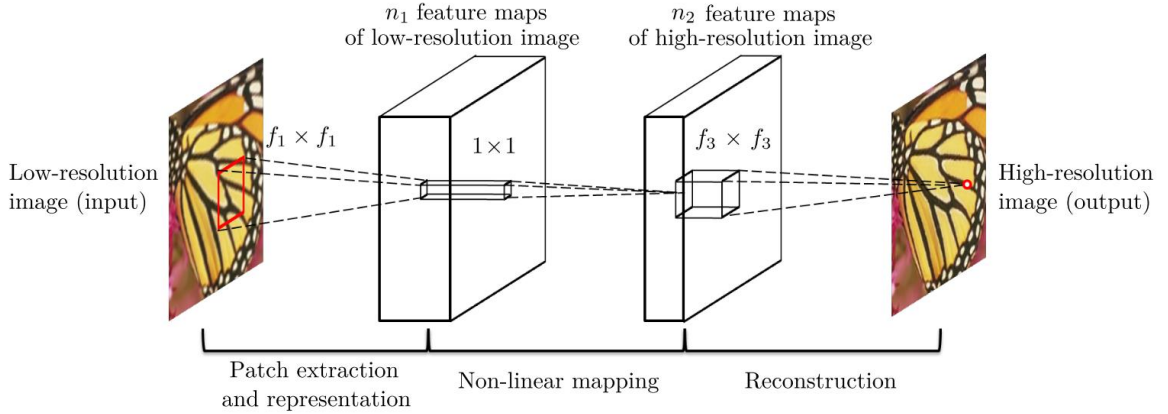


图2 SRCNN网络结构图

### 3. 基于机器学习的超分辨率算法

#### 3.1. SRCNN

通过与基于稀疏编码的方法进行对比,Dong等人[9]首次将卷积神经网络应用于超分辨率问题,提出了SRCNN(Super-Resolution Convolutional Neural Network),其结构如图2所示,其网络结构由图1类

比而来,将把低分辨率图像块映射到字典这一过程用 $f_1 * f_1$ 的卷积操作代替,生成 $n_1$ 个特征图,然后使用 $1 * 1$ 卷积操作完成基于稀疏编码中两个字典之间的映射变换,之后使用 $f_3 * f_3$ 卷积核将 $n_2$ 个特征图映射为高分辨率图像,并完成重建。

在训练时,使用均方误差(Mean Squared Error, MSE)作为损失函数,MSE逐像素点计算输入图像与输出图像之间的差距,并进行平均,公式如下:

$$MSE = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W (X(i, j) - Y(i, j))^2 \quad \#(1)$$

其中，W和H为输出图像的宽和高，端到端的映射将输入图像映射为目标大小的输出图像X，并将输出图像与真实（Ground Truth，GT）图像Y逐像素点计算偏差值，计算得到的均方误差使用标注你反向传播的随机梯度下降算法最小化损失函数[14]，作者指出，使用MSE作为损失函数有利于提高峰值信噪比（Peak Signal to Noise Ratio，PSNR），PSNR使量化评估图像恢复质量广泛使用的度量方法，其公式如下：

$$PSNR = 10 \log_{10} \left( \frac{(2^n - 1)^2}{MSE} \right) \quad \#(2)$$

其中n为每像素的比特数（bit），一般取8，即像素灰度等级为256，PSNR的单位是Db，数值越大表示失真越小。由于在优化过程中，使用MSE作为损失函数，直接优化了PSNR中的分母部分，其余部分均为常数，可以认为SRCNN将PSNR指标设定为学习目标，相较于传统算法，优化目标更加明确，也取得了更好的效果。

### 3.2. VDSR

自从SRCNN将卷积神经网络引入超分辨率任务中后，许多研究着手于通过提升网络性能来改进效果，基于深度学习方法的优点主要来自其两个关键因素，即网

络深度和跳跃链接（skip connection）[15]，多层的网络结构提供了强大的功能来拟合复杂的低分辨率和高分辨率之间的映射，同时用更大的感受野保留更多的上下文信息；跳跃链接可以有效缓解因网络层数增加带来的梯度消失/爆炸问题（gradient vanishing/exploring）。

Kim等人首先将上述两种思想引入了超分辨率问题[16]，他们指出，输入的低分辨率图像和输出的高分辨率图像在低频处十分相似，差别主要体现在高频细节部分，低频的相似部分再训练时带来了大量的时间消耗，因而在输入低分辨率图像和输出高分辨率图像之间引入跳跃链接，直接学习两者之间高频部分的残差。提出VDSR（Very Deep Convolutional Networks for Super Resolution），其结构如图3所示。

与SRCNN相比，VDSR首先网络层数加深到20层，使深层的网络拥有更大的感受野。文章选取3\*3的卷积核，则深度为D的网络感受野大小为(2D+1)×(2D+1)的感受野，在进行细节恢复时，深层网络提供了更加丰富的上下文信息。引入的跳跃链接学习低分辨率图像之间高频部分的残差，其数值较小，收敛速度跟快，且解决了梯度爆炸问题，作者指出低分辨率图像中有部分信息（轮廓）与高分辨率图像相似，是一种冗余的学习，因此直接通过跳跃链接将低频图像与网络输出相加，网络部分只负责学习差距较大的高频部分（细节）信息，提升了学习效率。

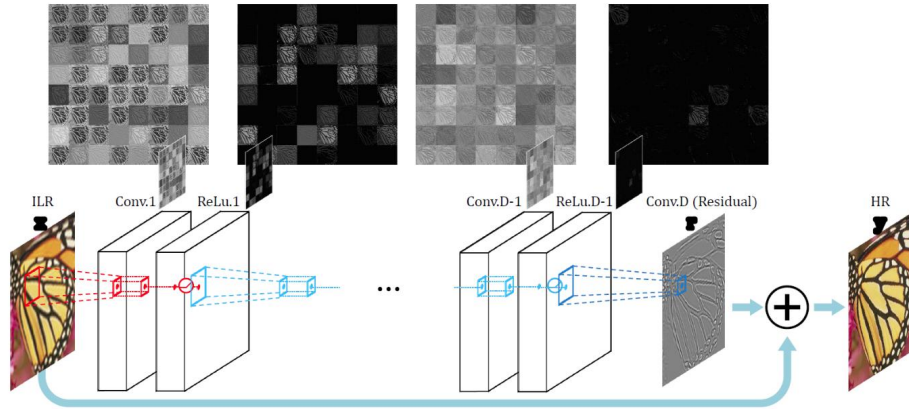


图3 VDSR网络结构图

### 3.3. DRCN

随着网络深度的增加，参数的数量也会增加，庞大的网络结构将使用大量计算资源并遇到过拟合（Overfitting）问题。为了减少网络参数，通常采用递归结构的递归神经网络（Recursive neural networks, RNN）[17]。

Kim等人首次将递归神经网络结构应用在超分辨率任务中[18]，同时使用跳跃链接，加深网络结构以增

加网络感受野，提出了DRCN（Deeply-Recursive Convolutional Network for Image Super-Resolution）模型，其结构如图4，分为三个模块：Embedding network用于提取低分辨率图像特征；Inference network用于将低分辨率特征映射到高分辨率；Reconstruction network从高分辨率图像中重建高分辨率图像。

其中，Inference network使用递归结构，由多层参数共享的卷积网络层组成，其展开后结构如图5所示， $H_1$ 到 $H_D$ 共享参数，输入通过每个 $H_i$ 后，结果传入Reconstruction network，与输入图像相加，得到D个重

建结果 $y_i$ ，按式(3)加权平均后得到总输出，训练过程同时对总输出 $y$ 和每个分支的输出 $y_i$ 进行优化。

$$y = \sum_{i=1}^D w_i \cdot y_i \quad (3)$$

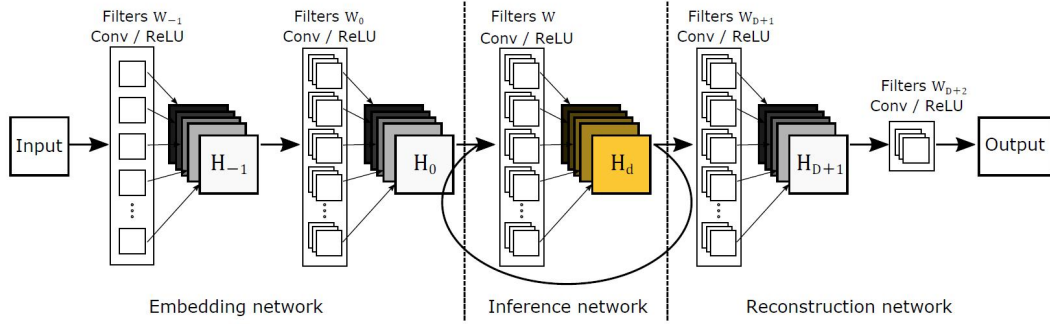


图 4 DRCN网络结构图

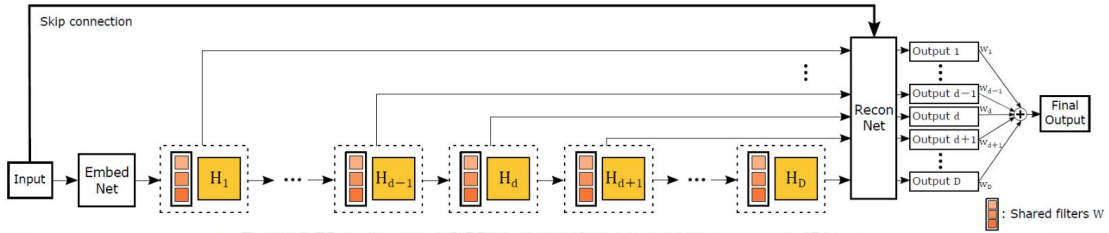


图 5 DRCN网络结构展开图

### 3.4. DRRN

受到DRCN的启发，Tai等人提出了递归神经网络应用于超分辨率问题的另一种方法[19]，使用更深的网络结构，提出DRRN (Deep Recursive Residual Network)，作者在论文中给出了对比图，如图 6 所示。DRRN中的每个递归网络层都有一个共同的输入，输入图像经过第一个卷积层后的输出结果。每个递归模块包含 2 个卷积层，在不同递归模块中，对应位置的卷积层（图中颜色相同）共享参数。相比VDSR和DRCN的全局残差学习，DRRN还包含了不同路径下局部残差的学习。

### 3.5. EDSR

Lim等人在Ledig等人提出的SRResNet[20]上进行修改，去掉了SRResNet中多余的模块，提出了EDSR (Enhanced Deep Residual Networks) [21]，文中指出，SRResNet直接将Resnet的残差结构[15]应用于超分辨率领域，引入了与任务不匹配的结构，阻碍了训练效果，如图 7 所示。

作者认为批规范化处理 (Batch normalization, BN) [22]适用与复杂的计算机视觉问题，如分类和检测，不适用于超分辨率这样的底层问题，影响了训练效果，且

占用许多计算资源。去除BN操作后，在相同的计算资源下，EDSR可以加深网络层数或增加特征提取数量，从而获得更好的性能表现，且网络结构如图 8 所示。

### 3.6. SRFBN

在生物学认知理论中，连接皮层视觉区域的反馈链接可以将响应信号从告诚区域传输到低阶区域[23]，受此现象的启发，有算法将反馈机制应用于神经网络[24]，采用自下而上的反馈机制，将高级信息传递回先前的层并改进低级编码信息，构建循环神经网络

(Recurrent Neural Network, RNN)，与递归神经网络不同，循环神经网络的每一次循环都有新的输入，且受到之前循环输出的影响，从而实现反馈机制。

受此影响，Li等人将反馈机制与循环神经网络应用于超分辨率问题[25]，提出SRFBN (Super-Resolution Feedback Network)，通过反馈链接使用高级别信息改进低级别信息，其结构如图 9 所示，该模型使用反馈块 (deefback block, FB) 作为循环神经网络中的隐藏状态 (hidden state)，每次迭代的隐藏状态流入下一次迭代以调制输入，为了确保隐藏状态包含高分辨率图像信息，对每次迭代的输出结果计算损失函数，从而粗粒度的高分辨率图像的信息就可以促进低分辨率图像重建更好的高分辨率图像。



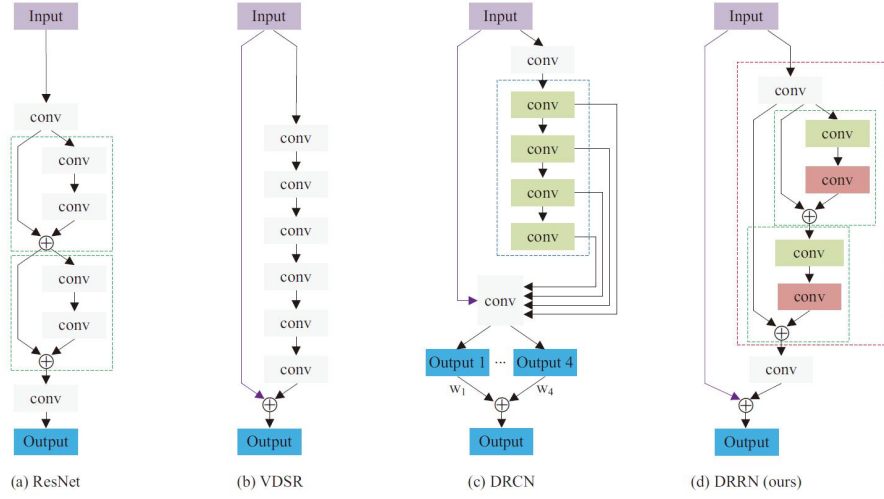


图 6 VDSR、DRCN、DRRN结构对比图

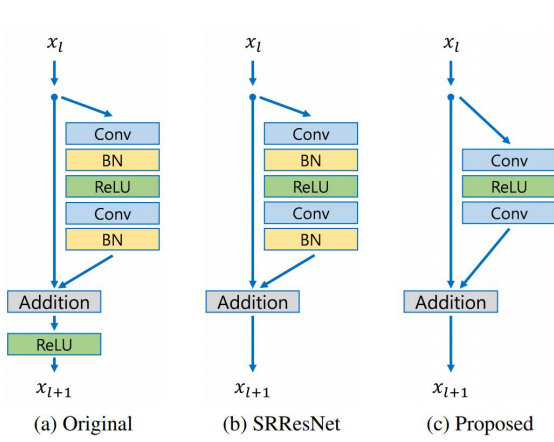


图 7 ResNet、SRResNet、EDSR残差模块对比图

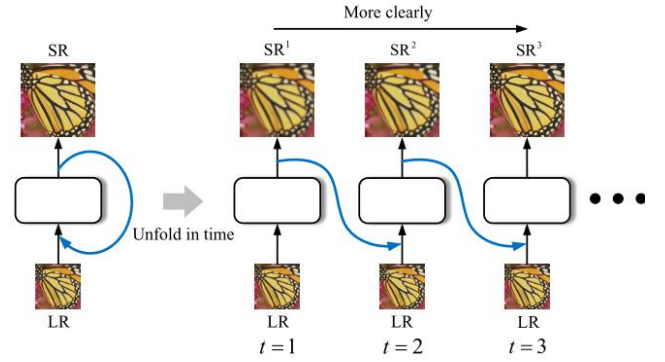


图 9 SRFBN网络结构图

## 4. 实验对比

### 4.1. 评估指标

为了评估上述算法的高分辨率恢复性能,使用峰值信噪比PSNR与结构相似性 (structural similarity, SSIM) 作为评估指标。

PSNR是用于定量评估图像恢复质量的广泛使用的度量,在一定程度上与人眼感知质量相关,其公式如式(2)所示。

SSIM从亮度、对比度、结构三方面度量图像相似性,是一种更加全面的图像质量评估指标,其公式如下图所示:

$$\mu_X = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X(i, j), \sigma_X^2 = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W (X(i, j) - \mu_X)^2$$

$$\sigma_{XY} = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W ((X(i, j) - \mu_X)(Y(i, j) - \mu_Y))$$

$$l(X, Y) = \frac{2\mu_X\mu_Y + C_1}{\mu_X^2 + \mu_Y^2 + C_1}, c(X, Y) = \frac{2\sigma_X\sigma_Y + C_2}{\sigma_X^2 + \sigma_Y^2 + C_2}, s(X, Y) = \frac{\sigma_{XY} + C_3}{\sigma_X\sigma_Y + C_3}$$

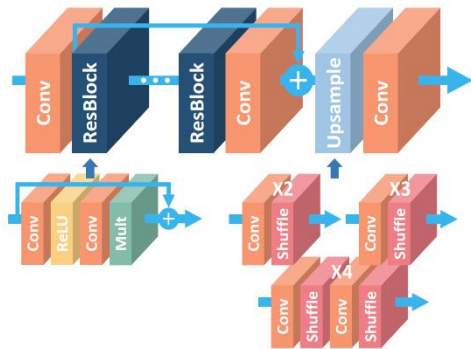


图 8 EDSR网络结构图

$$SSIM(X, Y) = l(X, Y) \cdot c(X, Y) \cdot s(X, Y)$$

#### 4.2. 数据集

使用DIVK[27]和Flickr2K作为训练数据，训练时使用数据增强，并在五个标准数据集上评估PSNR和SSIM指标下的结果，分别是Set5[28]，Set14[29]，B100[30]，Urban[31]和Manga109[32]

#### 4.3. 对比结果

使用相同数据集训练对上述提到的SRCNN、VDSR、DRRN、EDSR网络进行训练，并在基准数据集上进行测试，得到结果如表1所示。

从表中结果可以看出，相较于传统的双三次插值算法，基于深度学习的超分辨率算法效果提升十分明显，且网络层数的加深、残差模块的引入、迭代结构与反馈机制的引入，模型的效果在稳步提升，底层网络表达能力的提升带来了上层任务效果的提升，恢复的高分辨率图像更加清晰，细节也更加明显（图10）。

表1 基于深度学习的超分辨率算法性能对比

<i>Dataset</i>	<i>Scale</i>	<i>Bicubic</i>	<i>SRCNN</i>	<i>VDSR</i>	<i>DRRN</i>	<i>EDSR</i>	<i>SRFBN</i>
<b>Set5</b>	x2	33.66/0.9299	36.66/0.9542	37.53/0.9590	37.74/0.9591	38.11/0.9600	<b>38.18/0.9611</b>
<b>Set14</b>	x2	30.24/0.8688	32.45/0.9067	33.05/0.9130	33.23/0.9136	<b>33.92/0.9195</b>	<b>33.90/0.9203</b>
<b>B100</b>	x2	29.56/0.8431	31.36/0.8879	31.90/0.8960	32.05/0.8973	32.32/0.9013	<b>32.34/0.9015</b>
<b>Urban100</b>	x2	26.88/0.8403	29.50/0.8946	30.77/0.9140	31.23/0.9188	<b>32.93/0.9351</b>	32.80/0.9341
<b>Manga109</b>	x2	30.30/0.9339	35.60/0.9663	37.22/0.9750	37.60/0.9736	39.10/0.9773	<b>39.28/0.9784</b>

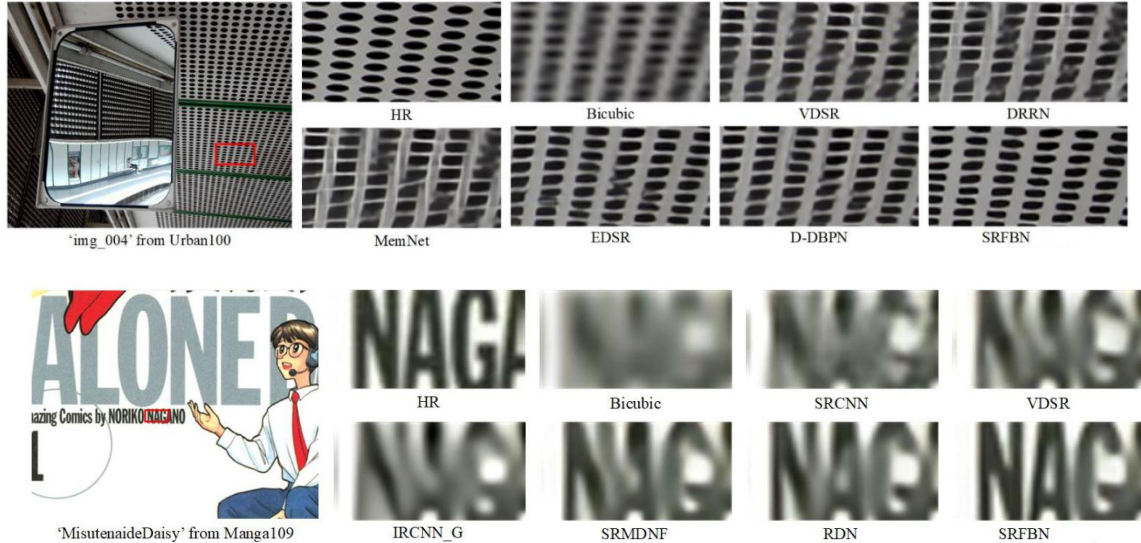


图10 超分辨率输出结果

#### 5. 结论

本文针对基于深度学习的超分辨率算法的发展历程进行了调研，分析其在网络结构、损失函数等方面的演变，并通过实验比较了上述网络的性能演化。

值得注意的是，深度学习在图像领域的应用推动了基于深度学习的超分辨率算法的发展，许多模型结构与思想都来源于图像处理领域的其他问题，如常用的特征提取网络来自于图像分类问题，循环结构与递归结构来自于时序性的视频图像处理任务，不同任务之间方法相互借鉴，共同推动了计算机视觉领域的发展。

## 参考文献

- [1] Freedman, G., Fattal, R.: Image and video upscaling from local self-examples. *TOG* 30(2), 12 (2011)
- [2] Yang, J., Lin, Z., Cohen, S.: Fast image super-resolution based on in-place example regression. In: *CVPR*. pp. 1059–1066 (2013)
- [3] Chang, H., Yeung, D.Y., Xiong, Y.: Super-resolution through neighbor embedding. In: *CVPR* (2004)
- [4] Chang, H., Yeung, D.Y., Xiong, Y.: Super-resolution through neighbor embedding. In: *CVPR* (2004)
- [5] Yang, J., Wright, J., Huang, T., Ma, Y.: Image super-resolution as sparse representation of raw image patches. In: *CVPR*. pp. 1–8 (2008)
- [6] Freeman, W.T., Jones, T.R., Pasztor, E.C.: Example-based super-resolution. *Computer Graphics and Applications* 22(2), 56–65 (2002)
- [7] Yang, J., Wright, J., Huang, T., Ma, Y.: Image super-resolution as sparse representation of raw image patches. In: *CVPR*. pp. 1–8 (2008)
- [8] Zeyde, R., Elad, M., Protter, M.: On single image scale-up using sparse representations. In: *Curves and Surfaces*, pp. 711–730. Springer (2012)
- [9] C. Dong, C. C. Loy, K. He, and X. Tang. Learning a deep convolutional network for image super-resolution. In *ECCV*, 2014.
- [10] Glasner, D., Bagon, S., Irani, M.: Super-resolution from a single image. In: *ICCV*. pp. 349–356 (2009)
- [11] Burger, H.C., Schuler, C.J., Harmeling, S.: Image denoising: Can plain neural networks compete with BM3D? In: *CVPR*. pp. 2392–2399 (2012)
- [12] Schuler, C.J., Burger, H.C., Harmeling, S., Scholkopf, B.: A machine learning approach for non-blind image deconvolution. In: *CVPR*. pp. 1067–1074 (2013)
- [13] Jain, V., Seung, S.: Natural image denoising with convolutional networks. In: *NIPS*. pp. 769–776 (2008)
- [14] LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86(11), 2278–2324 (1998)
- [15] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, 2016.
- [16] J. Kim, J. Kwon Lee, and K. Mu Lee. Accurate image superresolution using very deep convolutional networks. In *CVPR*, 2016.
- [17] Goller, C.; Küchler, A. (1996). Learning task-dependent distributed representations by backpropagation through structure. *IEEE Neural Networks*. 1. pp. 347–352.
- [18] J. Kim, J. Kwon Lee, and K. Mu Lee. Deeply-recursive convolutional network for image super-resolution. In *CVPR*, 2016.
- [19] Y. Tai, J. Yang, and X. Liu. Image super-resolution via deep recursive residual network. In *CVPR*, 2017.
- [20] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *CVPR*, 2017.
- [21] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee. Enhanced deep residual networks for single image super-resolution. In *CVPRW*, 2017
- [22] Ioffe, Sergey; Szegedy, Christian (2015). "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift".
- [23] J. M. Hup, A. C. James, B. R. Payne, S. G. Lomber, P. Girard, and J. Bullier. Cortical feedback improves discrimination between figure and background by v1, v2 and v3 neurons. *Nature*, 1998. 2
- [24] M. F. Stollenga, J. Masci, F. Gomez, and J. Schmidhuber. Deep networks with internal selective attention through feedback connections. In *NIPS*. 2014.
- [25] Li, Zhen et al. "Feedback Network for Image Super-Resolution." *CoRR* abs/1903.09814 (2019): n. pag.
- [26] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *TIP*, 2004.
- [27] E. Agustsson and R. Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *CVPRW*, 2017.
- [28] M. Bevilacqua, A. Roumy, C. Guillemot, and M. AlberiMorel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In *BMVC*, 2012.
- [29] R. Zeyde, M. Elad, and M. Protter. On single image scale-up using sparse-representations. In *Curves and Surfaces*, 2010.
- [30] D. R. Martin, C. C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *ICCV*, 2001.
- [31] J. Huang, A. Singh, and N. Ahuja. Single image superresolution from transformed self-exemplars. In *CVPR*, 2015.
- [32] Y. Matsui, K. Ito, Y. Aramaki, A. Fujimoto, T. Ogawa, T. Yamasaki, and K. Aizawa. Sketch-based manga retrieval using manga109 dataset. *Multimedia Tools and Applications*, 2017.