

基于 DCT 变换和零次学习的刑侦图像超分辨率^①



徐 健^{1,2}, 李新婷¹, 牛丽娇¹

¹(西安邮电大学 通信与信息工程学院, 西安 710121)

²(电子信息现场勘验应用技术公安部重点实验室, 西安 710121)

通信作者: 徐 健, E-mail: xujian@xupt.edu.cn

摘 要: 图像超分辨率在视频侦查领域有重要作用. 基于卷积神经网络的超分辨率算法通常在训练时输入人工合成的低分辨率图像, 学习高、低分辨率图像的映射, 很难应用于视频侦查领域. 真实低分辨率图像退化过程复杂未知, 且大都经过压缩算法的处理, 存在人工压缩痕迹, 导致超分辨率图像出现假纹理. 针对真实场景下的低分辨率图像提出一种基于离散余弦变换 (DCT) 和零样本学习的超分辨率算法. 该算法利用图像内部的重复相似性特点, 采用输入图像自身的子图像进行训练. 不同于以往超分辨率网络的输入, 所提算法采用子图像的离散余弦变换系数作为超分辨率网络的输入, 避免网络对输入图像的压缩痕迹进行放大, 减少假纹理. 在标准数据集和真实刑侦图像上的实验结果表明所提算法能减少图像中由压缩痕迹导致的假纹理.

关键词: 图像超分辨率重建; 压缩图像; 零样本学习; 离散余弦变换; 刑侦图像; 深度学习; 生成对抗网络 (GAN)

引用格式: 徐健, 李新婷, 牛丽娇. 基于 DCT 变换和零次学习的刑侦图像超分辨率. 计算机系统应用, 2022, 31(5): 291-297. <http://www.c-s-a.org.cn/1003-3254/8472.html>

Criminal Investigation Image Super-resolution Based on DCT and Zero-shot Learning

XU Jian^{1,2}, LI Xin-Ting¹, NIU Li-Jiao¹

¹(School of Communication and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710121, China)

²(Key Laboratory of Electronic Information Application Technology for Scene Investigation (Ministry of Public Security), Xi'an 710121, China)

Abstract: Image super-resolution (SR) plays an important role in video-based criminal investigation. SR algorithms based on convolutional neural networks are usually trained with the input of artificially synthesized low-resolution images to learn the mapping between high-resolution and low-resolution images, and thus they are difficult to be applied in the field of video-based criminal investigation. The degradation process of real low-resolution images is complex and unknown, and most of them are processed by compression algorithms, leading to false textures in SR images. Therefore, a new SR algorithm based on discrete cosine transform (DCT) and zero-shot learning is proposed for the video-based criminal investigation images in real scenarios. The algorithm takes advantage of the repetitive similarity within the images and utilizes sub-images from the input image for training. Different from the input of the previous SR network, the proposed algorithm takes the DCT coefficients of the sub-images as the input of the SR network to avoid magnifying compression artifacts of the input image and reduce false textures. The experimental results on the standard datasets and real criminal investigation images show that the proposed algorithm can reduce the false texture caused by compression artifacts.

Key words: image super-resolution reconstruction; compressed image; zero-shot learning; discrete cosine transform (DCT); criminal investigation images; deep learning; generative adversarial networks (GAN)

① 基金项目: 陕西省社会科学基金 (2021ZX15); 国家自然科学基金 (61601362, 62071380, 41874173); 西安邮电大学研究生创新基金 (CXJLZ202001)

收稿时间: 2021-07-31; 修改时间: 2021-08-20; 采用时间: 2021-09-03; csa 在线出版时间: 2022-04-11

随着智能化产业的不断发展,视频监控以其直观、准确、及时和信息内容丰富而广泛应用于许多场合,在安防系统中的重要性日益突出,成为技术安全防范最有力的手段.图像超分辨率(super-resolution, SR)可以有效地提升视频图像质量.对低分辨率图像本身进行高质量的重建,对于发现线索、拓展侦查手段和范围、突破嫌疑人口供和证明犯罪具有重要作用.随着深度学习在图像重建方面的发展,基于卷积神经网络(convolutional neural network, CNN)^[1-3]的SR方法取得了显著的成果.由于真实LR-HR图像对难以获取,许多基于CNN的SR方法通常在人工合成的图像对上进行训练,例如对HR图像进行双三次下采样^[4]得到的LR图像,作为LR-HR训练样本对.然而,真实场景下的LR图像退化过程复杂未知,与双三次下采样得到的LR图像的分布存在差异,由人工合成的训练样本训练的网络往往不适用于真实LR图像的超分辨率.

尽管有越来越多的复杂退化模型模拟真实图像的退化情况,但与真实LR图像之间仍存在差异,例如, Sun等人^[5]通过训练退化生成网络来减小合成LR和真实LR图像之间的域差, Guo等人^[6]提出对偶回归网络对真实LR图像附加约束,通过对偶学习估计下采样核.为了使网络适用于真实场景下的低分辨率图像, Shocher等人^[7]提出一种无监督零样本超分辨率算法(zero-shot super-resolution, ZSSR),该算法对输入图像进行下采样,训练网络学习下采样图像到输入图像之间的映射,待网络收敛后,对输入图像进行超分辨率,仅利用图像内部的重复相似性对网络进行训练,能处理任意模糊核下的真实LR图像,普适性强,但应用于视频侦查效果欠佳.

在真实场景下,为了减小存储与传输的数据量,采集的低分辨率图像都经过压缩,因此都存在压缩人工痕迹.在JPEG(joint photographic experts group)^[8]压缩中,由离散余弦变换(discrete cosine transform, DCT)将图像从空域变换到频域,通过对DCT系数进行量化编码,实现图像压缩,为了压缩数据量, DCT系数仅保留低频处的系数.由于缺乏高频处的系数,图像会出现振铃似的人工痕迹.在低分辨率图像中,由于JPEG压缩是分小块进行的,人眼的分辨率有限,振铃似的人工痕迹不是太明显.但是,当低分辨率图像经过超分辨率之后,小块变成了大块,人工痕迹也经过了超分辨率,导致人工痕迹非常明显.因此,本文提出一种基于离散余

弦变换和零样本学习的图像超分辨率算法,从图像压缩原理出发,对低分辨率图像进行DCT变换,取前几个DCT系数.在训练神经网络时,使网络学习低分辨率图像的前几个DCT系数与高分辨率图像之间的映射,避免超分辨率图像中出现假纹理.

1 相关工作

1.1 静止图像压缩原理

图像压缩是指去除图像数据中的冗余信息,在不影响视觉感官的基础上节约计算机存储. JPEG^[8]是静止图像压缩的最常见的一种方式,其基本原理如图1所示,对图像进行 8×8 大小的分块DCT变换, DCT系数反映出图像块中的能量分布情况,对DCT系数进行量化、编码,再进行逆DCT变换得到压缩后的图像.其中,低频分量对应于空域里图像的低频信息,包含了图像的大部分内容,且人眼对图像中的低频信息较为敏感;高频分量对应于空域图像中的细节信息,去除高频分量对于图像的视觉影响不大,尽管去除高频分量后在 8×8 的小方块上会出现一些人工痕迹,但是由于方块较小,这些人工痕迹并不明显.但是,当图像超分辨率后, 8×8 的小方块会被放大,这些压缩产生的人工痕迹也会一起被放大,超分辨率图像出现假纹理,影响图像视觉效果.

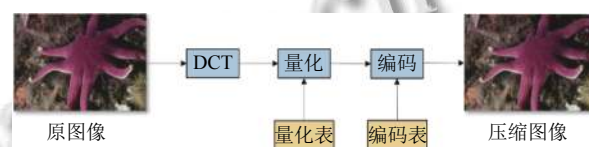


图1 JPEG图像压缩算法的示意图

1.2 零样本超分辨率

ZSSR^[7]是一种自监督或无监督算法,用于学习图像特有的内部信息.考虑到图像内部的重复相似性,使用核估计的方法估计输入图像的退化核,在输入图像上进行不同缩放因子的下采样构造成对训练样本,且通过在水平和垂直方向上旋转、翻转的方式增加训练样本.然后在增广后的训练样本上训练用于超分辨率的小型卷积神经网络,学习下采样图像与对应高分辨率图像间的映射,经过数千次的梯度更新后,再对输入图像进行超分辨率测试,得到超分辨率结果.整个超分辨率过程不依赖于任何外部训练样本或预训练过程,非理想条件下获得的超分辨率图像更接近真实场景的

模糊、噪声、人工痕迹等,适用于任意模糊核生成的低分辨率图像。

2 本文方法

本文提出一种基于离散余弦变换和零样本学习的图像超分辨率算法,具体结构图如图2所示,结合图像

压缩原理,训练网络学习输入图像的 DCT 系数与高分辨率图像之间的映射,主要包含下采样、频域转换、超分辨率 3 个步骤。本文算法采用 ZSSR 为基础框架,对下采样图像进行 DCT 变换,训练特定的小型 CNN 网络学习下采样图像 DCT 系数与对应高分辨率图像之间的映射关系,实现压缩图像的超分辨率。

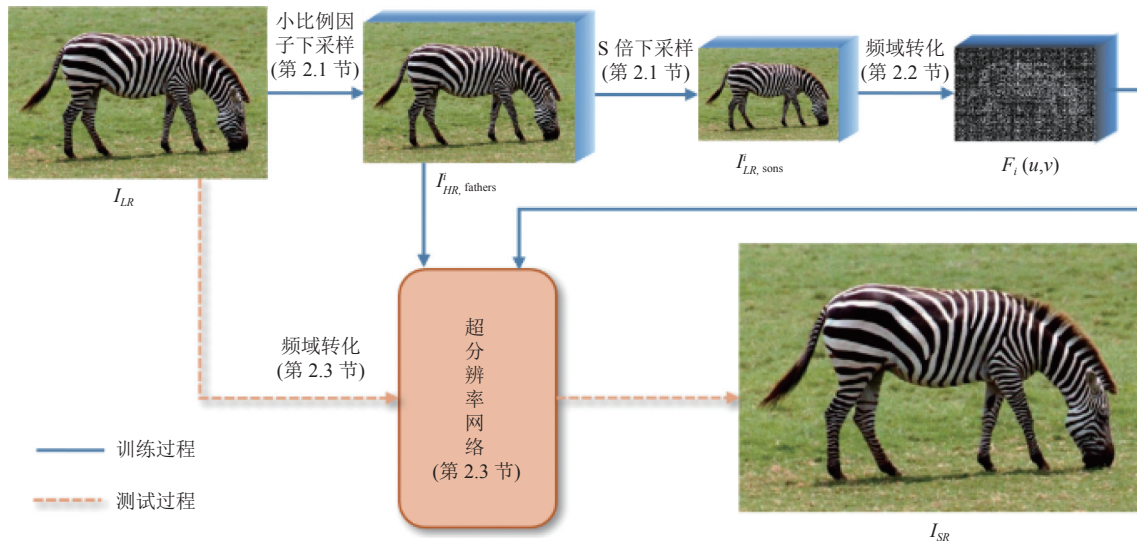


图2 所提算法流程图

2.1 下采样阶段

零样本学习仅依赖于单张图像内部的统计特性,所有的训练样本都来自于图像本身。下采样阶段的主要目的是获取训练 CNN 所需要的高低分辨率样本对。如图2所示,对输入图像 I_{LR} 先进行不同小比例因子的下采样得到下采样图 $I_{HR,fathers}^i$:

$$I_{HR,fathers}^i = D_{\downarrow s_i}(I_{LR}), 0 \leq i \leq n \quad (1)$$

其中, $D_{\downarrow s_i}(\cdot)$ 表示不同倍数 s_i 的下采样操作,再对 $I_{HR,fathers}^i$ 进行期望倍数为 s 的下采样:

$$I_{LR,sons}^i = D_{\downarrow s}(I_{HR,fathers}^i), 0 \leq i \leq n \quad (2)$$

将 $I_{LR,sons}^i$ 看作低分辨率子样本, $I_{HR,fathers}^i$ 看作与之对应的高分辨率子样本,构成训练 CNN 需要的 LR-HR 样本对 $(I_{LR,sons}^i, I_{HR,fathers}^i)$ 。相比其他标准数据集里的成对训练样本,仅通过不同下采样因子得到的训练样本对数量远远不够。因此,对 LR-HR 样本对分别在水平和垂直方向上进行不同角度的旋转和翻转,进一步增加 LR-HR 训练样本对。

2.2 频域转换

由于真实低分辨率图像大都经过压缩算法处理,存在人工压缩痕迹,使得超分辨率图像出现假纹理。这是因为许多超分辨率算法在超分辨率时将人工痕迹也进行了放大。结合图像压缩原理,使网络学习 DCT 系数与高分辨率图像之间的映射,能有效缓解人工痕迹被放大而出现假纹理的问题。具体来说,对低分辨率图像 $I_{LR,sons}^i$ 进行二维 DCT 变换:

$$\begin{cases} F_i(u,v) = \frac{2}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f_i(x,y)C(u)C(v)W(u,v,x,y) \\ x, u = 0, 1, 2, \dots, M-1; y, v = 0, 1, 2, \dots, N-1 \end{cases} \quad (3)$$

$$W(u,v,x,y) = \cos \frac{(2x+1)u\pi}{2M} \cos \frac{(2y+1)v\pi}{2N} \quad (4)$$

其中, $f_i(x,y)$ 为低分辨率图像 $I_{LR,sons}^i$ 的像素值矩阵, x 和 y 分别表示图像各像素点的位置, M 和 N 为低分辨率图像的大小, $F_i(u,v)$ 为 DCT 系数矩阵。将图像分为 8×8 大小的图像块进行分块 DCT 变换,对于每一个图像块

的 DCT 系数矩阵, 左上角的系数值较大, 为低频分量, 包含原始图像中的大部分内容. 对 DCT 系数矩阵进行量化, 保留低频系数, 将高频系数置零, 可以达到压缩图像的目的. 本文选取包含少量低频系数的 DCT 系数矩阵, 作为 CNN 网络的输入.

2.3 超分辨率网络

将低分辨率图像的 DCT 系数输入 CNN 网络, 训

练网络学习低分辨率图像 $I_{LR,sons}^i$ 的 DCT 系数与高分辨率图像 $I_{HR,fathers}^i$ 之间的映射, 采用重建损失训练网络. 网络结构如图 3 所示, 采用全卷积网络, 8 个隐层, 每层有 64 个通道, 并在每层后面跟 ReLU 激活函数. 训练网络经过数千次梯度更新后, 网络收敛, 将输入图像 I_{LR} 进行 DCT 变换后, 由训练好的 CNN 网络进行超分辨率, 获取的超分辨率结果不包含假纹理.

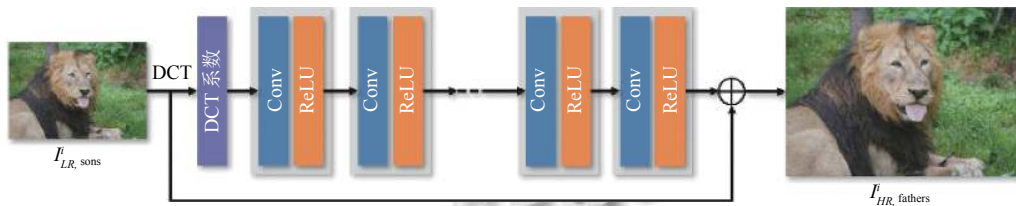


图 3 网络结构图

3 实验分析

3.1 实验环境与设置

本文算法的训练数据与测试数据都来自输入图像本身, 训练时采用 Adam 优化器^[9], 初始化学习率为 0.001, 定期对重建误差进行线性拟合, 如果标准偏差比线性拟合的斜率大, 就把学习率除以 10, 直到学习率降为 10^{-6} 为止, 超分辨率倍数为 2. 实验环境为 Ubuntu 20.04, 深度学习框架为 TensorFlow 2.0, 32 GB 内存主频为 DDR4 3 200 MHz, GPU 为 NVIDIA GTX1080Ti.

3.2 结果与分析

在标准数据集 Set5^[10]、Set14^[11]、BSD100^[12]、Urban100^[13] 上进行了实验测试分析, 使用零样本超分辨率 (zero-shot super-resolution, ZSSR)^[7]、纹理迁移网络超分辨率 (learning texture Transformer network for image super-resolution, TTSR)^[14]、深度交替网络 (deep alternating network, DAN)^[15]、多尺度超分辨率 (multi-scale image super-resolution, MSWSR)^[16]、残差通道注意力网络 (residual channel attention networks, RCAN)^[17]

等 5 种算法与本文算法进行 2 倍放大的结果如表 1 所示, 其中, Bicubic 为双三次插值算法, ZSSR、TTSR 为无监督算法, DAN、RCAN、MSWSR 为监督算法. PSNR (peak signal-to-noise ratio) 为峰值信噪比, SSIM (structural similarity) 为结构相似度, 表中的数据为 PSNR/SSIM 的平均值. 考虑到标准数据集中的人工合成低分辨率图像为无损压缩格式, 本文算法在训练时输入为未量化的 DCT 系数, 由表 1 可以看出, 本文算法的 PSNR 值要高于 ZSSR 算法平均 0.065 dB, 高于 TTSR 算法平均 0.885 dB. 有监督的算法 (例如 DAN、RCAN 和 MSWSR 等) 均比无监督的算法效果好, 但是无监督的算法仅牺牲了一部分性能指标, 就可以避免长时间的学习过程, 大大降低了算法复杂度. 有监督的算法通常需要针对不同的放大倍数分别进行参数训练 (例如 2 倍和 3 倍放大需要分别训练神经网络), 每一次训练都要消耗大量时间, 但是无监督的算法能够用于任何放大倍数, 适应性较强, 更适合实际应用场合.

表 1 不同算法在标准数据集上的 PSNR/SSIM 结果

| 算法 | Set5 | Set14 | BSD100 | Urban100 |
|---------|----------------------|----------------------|----------------------|----------------------|
| Bicubic | 28.82/0.857 7 | 26.02/0.763 4 | 25.92/0.731 0 | 23.14/0.725 8 |
| DAN | 36.33/0.935 4 | 32.07/0.906 8 | 31.16/0.921 3 | 30.60/0.902 0 |
| RCAN | 37.25/0.956 7 | 33.08/0.901 3 | 31.64/0.892 8 | 30.41/0.909 2 |
| MSWSR | 37.48/0.958 0 | 33.23/0.912 3 | 31.86/0.892 7 | 31.22/0.926 0 |
| ZSSR | 31.08/0.878 6 | 28.35/0.793 3 | 27.92/0.763 2 | 25.25/0.761 8 |
| TTSR | 30.25/0.862 4 | 27.89/0.782 5 | 26.56/0.744 3 | 24.62/0.747 2 |
| 本文算法 | 31.12/0.878 9 | 28.41/0.806 8 | 27.98/0.774 2 | 25.35/0.782 8 |

图4展示了对比算法与本文算法在标准数据集上的超分辨率结果图,图4(a)为低分辨率图像,图4(b)为原始高分辨率图像,图4(c)为Bicubic超分辨率结果,图4(d)为ZSSR超分辨率结果,图4(e)为TTSR超分辨率结果,图4(f)为DAN超分辨率结果,图4(g)为RCAN超分辨率结果,图4(h)为MSWSR超分辨率结果,图4(i)为本文算法结果.从图中可见,图4(i)的视觉效果与图4(d)和图4(e)的视觉效果相接近,这是因为标准数据上的图像为无损压缩图像,不包含人工压缩痕迹,而本文算法虽然是针对压缩图像的超分辨率算法,但在人工合成的低分辨率图像上仍然具有良好的超分辨率效果.为进一步验证本文算法的有效性,采用真实低分辨率图像进行测试,由于真实低分辨率图像经过了压缩算法

的处理,本文算法分别选取包含3个DCT系数和1个DCT系数的系数矩阵对真实图像进行超分辨率,结果如图5、图6所示,图5(a)和图6(a)为真实低分辨率图像,图5(b)和图6(b)为ZSSR超分辨率结果,图5(c)和图6(c)为TTSR超分辨率结果,图5(d)和图6(d)为RCAN超分辨率结果,图5(e)和图6(e)为MSWSR超分辨率结果,图5(f)和图6(f)为本文算法保留一个系数的超分辨率结果,图5(g)和图6(g)为本文算法保留3个系数的超分辨率结果.图5为人物图,可以看出与图5(a)~图5(f)相比,图5(g)中人物的额头、下巴之类的部位的假纹理明显减少.图6为车牌图像,可以看出与图6(a)~图6(f)相比,图6(g)的后保险杠和车牌上的假纹理明显减少,车牌的轮廓和字符的边缘也更清晰.

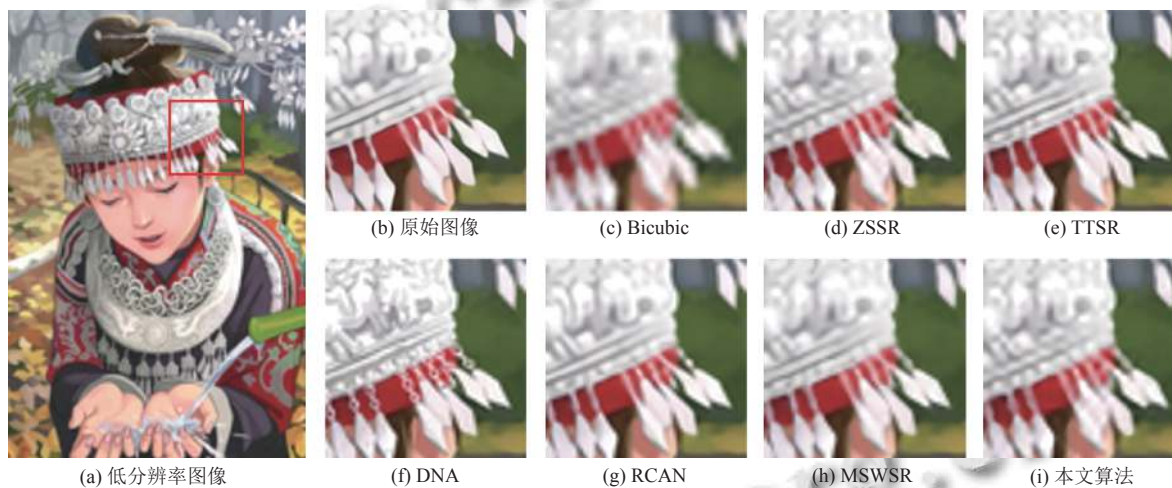


图4 Set14数据集超分辨率结果



图5 真实LR图像(一)超分辨率结果



图6 真实LR图像(二)超分辨率结果

4 结论

基于CNN的超分辨率网络通常学习低分辨率图像到高分辨率图像的映射,对于真实低分辨率图像而言,退化过程复杂未知,且缺少对应的高分辨率图像。另外,由于真实场景下的低分辨率图像大都经过压缩算法的处理,存在人工压缩痕迹,许多算法对真实低分辨率图像进行超分辨率时会将压缩痕迹放大,导致超分辨率图像出现假纹理。在自监督算法ZSSR的基础上,本文提出一种基于离散余弦变换和零样本学习的超分辨率算法,对低分辨率图像进行DCT变换,训练超分辨率网络学习DCT系数与高分辨率图像之间的映射,避免压缩痕迹被放大。观察实验结果可以看出,本文算法在无损压缩的图像上具有与最新无监督算法相接近的性能,在真实低分辨率图像上,能够有效地减少假纹理,获得良好的视觉效果。

参考文献

- 1 Tian CW, Xu Y, Zuo WM, *et al.* Coarse-to-fine CNN for image super-resolution. *IEEE Transactions on Multimedia*, 2021, 23: 1489–1502. [doi: [10.1109/TMM.2020.2999182](https://doi.org/10.1109/TMM.2020.2999182)]
- 2 Niu B, Wen WL, Ren WQ, *et al.* Single image super-resolution via a holistic attention network. *Proceedings of the 16th European Conference on Computer Vision*. Glasgow: Springer, 2020. 191–207.
- 3 Zhang KB, Wang Z, Li J, *et al.* Learning recurrent residual regressors for single image super-resolution. *Signal Processing*, 2019, 154: 324–337. [doi: [10.1016/j.sigpro.2018.09.002](https://doi.org/10.1016/j.sigpro.2018.09.002)]

- 4 Ruangsang W, Aramvith S. Efficient super-resolution algorithm using overlapping bicubic interpolation. *Proceedings of 2017 IEEE Global Conference on Consumer Electronics*. Nagoya: IEEE, 2017. 1–2. [doi: [10.1109/GCCE.2017.8229459](https://doi.org/10.1109/GCCE.2017.8229459)]
- 5 Sun W, Gong D, Shi QF, *et al.* Learning to zoom-in via learning to zoom-out: Real-world super-resolution by generating and adapting degradation. *IEEE Transactions on Image Processing*, 2021, 30: 2947–2962. [doi: [10.1109/TIP.2021.3049951](https://doi.org/10.1109/TIP.2021.3049951)]
- 6 Guo Y, Chen J, Wang JD, *et al.* Closed-loop matters: Dual regression networks for single image super-resolution. *Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2020. 5406–5415. [doi: [10.1109/CVPR42600.2020.00545](https://doi.org/10.1109/CVPR42600.2020.00545)]
- 7 Shocher A, Cohen N, Irani M. Zero-shot super-resolution using deep internal learning. *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 3118–3126. [doi: [10.1109/CVPR.2018.00329](https://doi.org/10.1109/CVPR.2018.00329)]
- 8 程芳. 基于DCT变换的JPEG图像压缩方式的应用研究. *电脑知识与技术*, 2021, 17(10): 228–229, 232. [doi: [10.14004/j.cnki.ckt.2021.1098](https://doi.org/10.14004/j.cnki.ckt.2021.1098)]
- 9 Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv: 1412.6980, 2014.
- 10 Bevilacqua M, Roumy A, Guillemot C, *et al.* Low-complexity single-image super-resolution based on nonnegative neighbor embedding. *Proceedings of British Machine Vision Conference*. Guildford: BMVA Press, 2012: 135.1–135.10. [doi: [10.5244/C.26.135](https://doi.org/10.5244/C.26.135)]

- 11 Zeyde R, Elad M, Protter M. On single image scale-up using sparse-representations. Proceedings of the 7th International Conference on Curves and Surfaces. Avignon: Springer, 2010. 711–730. [doi: [10.1007/978-3-642-27413-8_47](https://doi.org/10.1007/978-3-642-27413-8_47)]
- 12 Martin D, Fowlkes C, Tal D, *et al.* A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. Proceedings of the 8th IEEE International Conference on Computer Vision. Vancouver: IEEE, 2001. 416–423. [doi: [10.1109/ICCV.2001.937655](https://doi.org/10.1109/ICCV.2001.937655)]
- 13 Huang JB, Singh A, Ahuja N. Single image super-resolution from transformed self-exemplars. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015. 5197–5206. [doi: [10.1109/CVPR.2015.7299156](https://doi.org/10.1109/CVPR.2015.7299156)]
- 14 Yang FZ, Yang H, Fu JL, *et al.* Learning texture transformer network for image super-resolution. Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 5791–5799. [doi: [10.1109/CVPR42600.2020.00583](https://doi.org/10.1109/CVPR42600.2020.00583)]
- 15 Huang Y, Li S, Wang L, *et al.* Unfolding the alternating optimization for blind super resolution. Advances in Neural Information Processing Systems, 2020, 33: 5632–5643.
- 16 Zhang HR, Xiao J, Jin Z. Multi-scale image super-resolution via a single extendable deep network. IEEE Journal of Selected Topics in Signal Processing, 2021, 15(2): 253–263. [doi: [10.1109/JSTSP.2020.3045282](https://doi.org/10.1109/JSTSP.2020.3045282)]
- 17 Zhang YL, Li KP, Li K, *et al.* Image super-resolution using very deep residual channel attention networks. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 294–310. [doi: [10.1007/978-3-030-01234-2_18](https://doi.org/10.1007/978-3-030-01234-2_18)]