

Research on Super-Resolution of Marine Ship Image Based on Target Detection

Kun Zhang, Tianwei Li

Department of Navigation, Dalian Naval Academy, Dalian Liaoning
Email: 597402096@qq.com

Received: Jun. 11st, 2019; accepted: Jun. 28th, 2019; published: Jul. 4th, 2019

Abstract

Aiming at the problem that the effective pixels in the image of marine ships account for a small proportion in the total pixels, a super-resolution method based on target detection network is proposed. The method consists of two stages, combining with the bicubic transform, to restore the sharpness of the image from coarse to fine step by step. Firstly, in the first stage, super-resolution regions in the original image are detected through the target detection network. Then, in the second stage, the corresponding regions are adjusted to the specified resolution by bicubic transformation, and then the image details are enhanced by generating the countermeasure network. Finally, the experimental results on the self-built dataset show that compared with the traditional method and the existing super-resolution reconstruction algorithm based on deep neural network, this algorithm not only has the best visual effect, but also improves the peak signal-to-noise ratio (PSNR) of the dataset by an average of 0.79 dB and the structural similarity (SSIM) by an average of 0.04, which proves the effectiveness of the algorithm.

Keywords

Target Detection, Generation of Countermeasure Network, Super-Resolution, U-Net, PatchGAN

基于目标检测的海上舰船图像超分辨率研究

张 坤, 李天伟

海军大连舰艇学院航海系, 辽宁 大连
Email: 597402096@qq.com

收稿日期: 2019年6月11日; 录用日期: 2019年6月28日; 发布日期: 2019年7月4日

摘 要

针对海上舰船图像有效像素在整体像素中占比小的问题, 提出一种基于目标检测网络的超分辨率方法。该

方法包含两个阶段, 结合bicubic变换, 逐步地将图像的清晰度从粗到细地进行恢复。首先, 第一阶段通过目标检测网络, 检测出原图像中需要超分辨率的区域, 然后, 第二阶段将对应区域通过bicubic变换调整至指定分辨率, 而后通过生成对抗网络增强图像细节。最终在自建数据集上的实验结果表明, 与传统方法和现有基于深度神经网络的超分辨率重建算法相比, 该算法不仅图像视觉效果最好, 而且在数据集上的峰值信噪比(PSNR)平均提高了0.79 dB, 结构相似性(SSIM)平均提高了0.04, 证明了该算法的有效性。

关键词

目标检测, 生成对抗网络, 超分辨率, U-Net, PatchGAN

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近年来, 图像超分辨率技术得到了长足的发展, 结合比较有效的深度学习技术, 该方法的精度已经得到了较大的提高, 并已经广泛应用于各种领域, 如视频监控、医学成像、高清晰度电视、遥感、手机与数码相机等[1] [2]。本文算法的应用背景为提高海上远距离舰船目标图像的分辨率, 主要应用于单幅海上舰船图像的超分辨率, 该问题是一个不适定的逆问题, 旨在从低分辨率(Low-Resolution, LR)图像中恢复出一个高分辨率(High-Resolution, HR)图像。如图1。



Figure 1. Marine ship image

图1. 海上舰船图像

目前传统超分辨率算法在运行时存在以下问题: 1) 由于远距离舰船图像的像素在整体图像所占比例较少, 大部分像素为用户并不关心的海浪天空等区域, 因此在运行传统算法的时, 大部分运算时间应用到了无关的区域, 不符合使用者的使用意图, 而且运行效率低下。2) 在应用深度学习技术的超分辨率算法中, 如果对整张图像进行学习, 则算法使用的卷积神经网络不能有效学习舰船目标的特征, 而很可能把海浪天空的特征误认为舰船的特征, 从而无法有效对舰船图像进行重建。基于以上两点原因, 本文提出一种基于目标检测网络的超分辨率算法。

2. 研究背景及现状

2.1. 目标检测算法

自从 AlexNet 在比赛中使用卷积神经网络进而大幅度提高了图像分类的准确率, 便有学者尝试将深度学习应用到目标检测中。在这方面, 主要有两种主流的算法: 一类是结合 region proposal、CNN 网络

的, 基于分类的 R-CNN 系列目标检测框架(two stage); 另一类则是将目标检测转换为回归问题的算法(single stage) [3]。虽然 FasterRCNN 算法是目前主流的目标检测算法之一, 但是速度上并不能满足实时的要求。随后出现像 YOLO, SDD 这一类的算法逐渐凸显出其在速度上的优势。YOLO [4]算法的网络设计策略延续了 GoogleNet [5]的核心思想, 真正意义上实现了端到端的目标检测, 且发挥了速度快的优势。YOLO 采用以 cell 为中心的多尺度区域取代 region proposal, 舍弃了一些精确度以换取检测速度的大幅提升, 检测速度可以达到 45 f/s, 足以满足实时要求。

2.2. 超分辨率算法

传统的 SISR 的方法包括基于插值的算法, 基于凸集投影法的算法等, 但由于深度学习在计算机视觉领域的突破性进展, 人们尝试在超分辨率问题中引入神经网络, 通过构建深层次的网络进行端到端的训练来解决图像超分辨率重建问题[6]。SRCNN (super-resolution convolutional neural network) [7]是最早运用深度学习方法在 LR 与 HR 之间建立端到端映射的 SISR 模型, 其输入图像采用了插值预处理的方法。Ledig 等人[8]基于 GAN 提出了一种用于图像超分辨率的生成对抗网络 SRGAN, 通过生成式和判别器的交替执行, 充分提取高频信息。由于海上舰船图片成对采集, 因此适用于使用建立像素到像素映射关系的超分辨率算法[9], 论文借鉴 pix2pix [10]的算法设计。

3. 基于目标检测网络的超分辨率重建模型

本文所设计的模型是一种基于目标检测算法的模型, 目的在于检测出整幅中用户感兴趣的区域, 而后再对指定区域进行超分辨率放大, 从而减少算法运行的时间, 并更好的重建目标区域图像的边缘和纹理, 本文的方法由两阶段组成, 如图 2 所示。

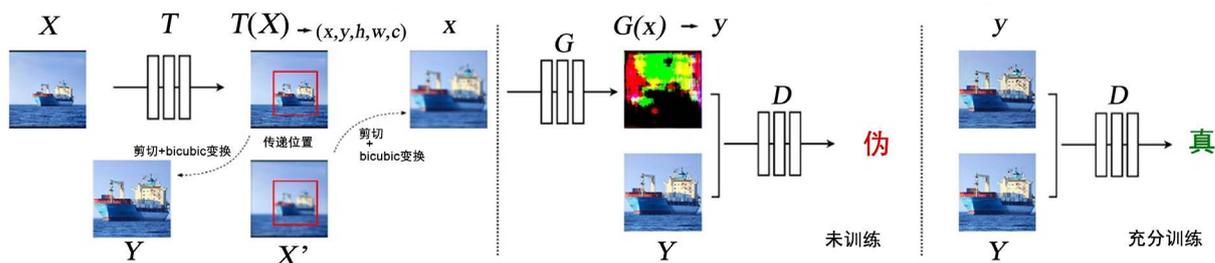


Figure 2. Model structure
图 2. 模型结构

X 为原图, X' 为 X 的退化图像, Y 为 X 中用户感兴趣的区域, x 为 X' 中用户感兴趣的区域, y 为生成网络生成的图像。T 为目标检测网络, 用于获取 X 中的 $(x, y, h, w, confidence)$ 信息, G 为图像生成网络, D 为鉴别网络。

3.1. 目标检测卷积神经网络结构设计

目标检测网络 T 的结构设计借鉴了 yoloV3 中 Darknet-53 的神经网络设计, 其候选框基于 anchor 候选框机制, 其原理图如图 3:

网络实际的预测值为, t_x, t_y, t_w, b_h 根据上图中的四个公式计算得到预测框的中心点坐标和宽高 b_x, b_y, b_w, b_h 。其中 c_x, c_y 为当前 grid 相对于左上角 grid 偏移的 grid 数量。

图 3 所示 $\sigma(t)$ 函数为 logistic 函数, 将坐标归一化到 0~1 之间。最终得到的 b_x, b_y 为归一化后的相对于 gridcell 的值。 p_w, p_h 与 groundtruth 重合度最大的 anchor 框的宽和高。实际在使用中, 将 b_w, b_h 也归

一化到 0~1, 实际程序中的 p_w , p_h 为 anchor 的宽, 高和 feature map 的宽, 高的比值。最终得到的 b_w , b_h 为归一化后相对于 anchor 的值。

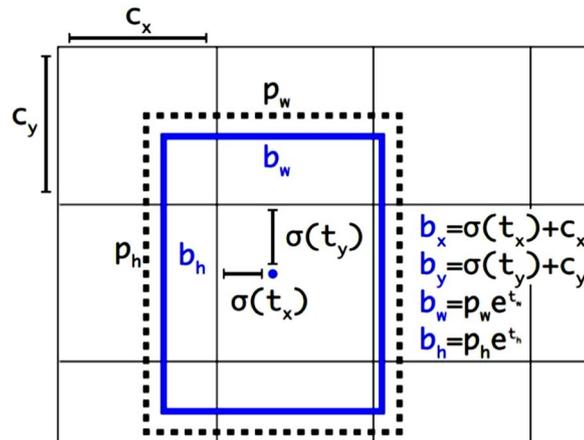


Figure 3. Anchor Candidate Box

图 3. Anchor 候选框

卷积神经网络结构如图 4:

Type	Filters	Size	Output
Convolutional	32	3 × 3	256 × 256
Convolutional	64	3 × 3 / 2	128 × 128
1x	Convolutional	32	1 × 1
	Convolutional	64	3 × 3
	Residual		128 × 128
Convolutional	128	3 × 3 / 2	64 × 64
2x	Convolutional	64	1 × 1
	Convolutional	128	3 × 3
	Residual		64 × 64
8x	Convolutional	256	3 × 3 / 2
	Convolutional	128	1 × 1
	Convolutional	256	3 × 3
	Residual		32 × 32
8x	Convolutional	512	3 × 3 / 2
	Convolutional	256	1 × 1
	Convolutional	512	3 × 3
	Residual		16 × 16
4x	Convolutional	1024	3 × 3 / 2
	Convolutional	512	1 × 1
	Convolutional	1024	3 × 3
	Residual		8 × 8
Avgpool		Global	
Connected		1000	
Softmax			

Figure 4. Convolutional neural network architecture

图 4. 卷积神经网络结构

该网络的特点在于使用了连续的 3 × 3 和 1 × 1 的卷积基层, 简化了 resnet 神经网络, 减少了检测时间。

在 yoloV3 损失函数中, 需要关注 4 个信息, 分别为位置信息: (x, y), 选中框的长宽: (w, h), 识别的置信度: confidence, 识别出物体的类别: class。由于在舰船的超分辨率任务中, 识别的目标只有舰船一类, 所以为了简化算法, 提高运算速度, 本文算法只是用前三项作为算式函数内容, 因此的到如下公式:

$$\begin{aligned}
loss = & \lambda_{coord} \sum_{i=0}^{l.w*l.h} \sum_{j=0}^{l.n} 1_{ij}^{obj} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \\
& + \lambda_{coord} \sum_{i=0}^{l.w*l.h} \sum_{j=0}^{l.n} 1_{ij}^{obj} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right] \\
& + \lambda_{noobj} \sum_{i=0}^{l.h*l.w} \sum_{j=0}^{l.n} 1_{ij}^{noobj} (C_i - \hat{C}_i)^2 + \lambda_{obj} \sum_{i=0}^{l.h*l.w} \sum_{j=0}^{l.n} 1_{ij}^{obj} (C_i - \hat{C}_i)^2
\end{aligned}$$

其中 λ_{coord} 为位置错误的权重, λ_{noobj} 为没有 *object* 的候选框的置信度权重, λ_{obj} 为有 *object* 的候选框的置信度权重, 1_{ij}^{obj} 判断第 i 个栅格中的第 j 个候选框是否包含物体, 1_{ij}^{noobj} 判断是否有物体的中心落在 i 中, x_i, y_i 为实际的坐标值, \hat{x}_i, \hat{y}_i 为预测的坐标值, w_i, h_i 为实际候选框的宽和高, \hat{w}_i, \hat{h}_i 为预测候选框的宽和高, C_i 为实际的类别, \hat{C}_i 为预测的类别。

3.2. 图像生成卷积神经网络结构设计

图像生成网络 G 的设计为 U-Net 结构, U-Net [11] 是德国 Freiburg 大学模式识别和图像处理组提出的一种全卷积结构。和常见的先降采样到低维度, 再升采样到原始分辨率的编解码(Encoder-Decoder)结构的网络相比, U-Net 的区别是加入 skip-connection, 对应的 featuremaps 和 decode 之后的同样大小的 featuremaps 按通道拼(concatenate)一起, 用来保留不同分辨率下像素级的细节信息, U-Net 对提升细节的效果非常明显。

3.3. 图像判别卷积神经网络结构设计

图像判别网络 G 为 PatchGAN [12], 该网络不是以整个图像的作为输入, 而是以小的 patch 来进行的。把一副图像划分为 $N \times N$ 个 patch 后, 对于每一块进行上述的那个操作。可以发现当 $N = 1$ 的时候, 相当于逐像素进行了, 当 $N = 256$ (图像大小为 256 的话), 就是一幅图像的操作。最后将一张图片所有 patch 的结果取平均作为最终的判别器输出。实验发现当 $N = 70$ 的时候, 效果最好[10]。

3.4. 超分辨率网络的目标函数

构建好生成网络和判别网络结构后, 设定目标函数。生成器 G 不断的尝试最小化下面的目标函数, 而 D 则通过不断的迭代去最大化这个目标函数。目标函数分为两个部分, 第一部分为生成对抗网络的损失函数, 公式为:

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y} [\log D(x, y)] + \mathbb{E}_{x,z} [\log(1 - D(x, G(x, z)))]$$

第二部分为 L1 正则化补偿函数, 公式为:

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z} [\|y - G(x, z)\|_1]$$

所以最终的目标函数是:

$$G = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G)$$

4. 实验结果与分析

由于目前并没有针对舰船的图片数据集, 本文实验使用项目自行采集的数据集作为训练集, 该数据集由实际采集的舰船照片和网络爬取的舰船照片组成。其中实际采集的舰船照片为同一目标不同分辨率的图像, 经测算高分辨率图像经双线性差值与高斯模糊函数相叠加的方式作为退化函数能有效的模拟低分辨率图像, 而后将此退化函数应用到采集的舰船照片上, 得到本文使用的数据集。舰船图像的原始像

素大小的 256×256 , 经退化函数处理得到不同分辨率的低分辨率图像。在 4 倍缩小尺度下, 低分辨率图像像素为 128×128 , 在 9 倍缩小尺度下, 低分辨率图像像素为 85×85 , 在 16 倍缩小尺度下, 低分辨率图像像素为 64×64 。一张高分辨舰船图像和一张对应的低分辨率图像为一组图像。数据集分为训练图像、验证图像和测试图像三个子集, 其中训练图像包含 3000 组图像, 验证图像包含 50 组图像, 测试图像包含 50 组图像。

本文实验通过 Adam 优化方法对网络进行训练。利用“步长”(step)策略调整学习率, 初始学习率权重为 $2e-5$, 指数衰减率为 0.999, 调整系数 gamma 为 0.5, 最大迭代次数为 10000。训练网络用机器配置为: i7-7700, 16G 内存, GTX1070Ti8G。

本文所提出的算法与现有的 5 种超分辨率的方法相比较, 分别为 bilinear, bicubic [13], discogan [14], pix2pix [10], pocs [15]。由于本文使用的算法只对选定的图像位置进行超分比率, 因此进行图像质量评价时为保证评价的公平性, 也仅对选定区域进行比较评价。利用本文算法得到的位置信息, 对各类算法得的图像进行截取, 而后利用峰值信噪比(PSNR)和结构相似度(Structual Similarity, SSIM)对截取的原图像和生成图像进行比较, 得到重建图像的评价值, 比较 $\times 4$, $\times 9$, $\times 16$ 不同尺度下的效果, 结果如表 1 所示。

Table 1. Test results on datasets using different super-resolution methods

表 1. 使用不同超分辨率方法在数据集上的测试结果

算法	重建倍数	Psnr	Ssim	Time (ms)
Yolo2pix		24.549	0.791	77.8
discogan		23.401	0.724	13.2
pix2pix	$\times 4$	23.891	0.731	12.9
pocs		18.948	0.631	2034
bilinear		23.266	0.752	8.8
bicubic		23.864	0.777	13.6
Yolo2pix		22.791	0.715	78.2
discogan		22.031	0.647	13.3
pix2pix	$\times 9$	22.154	0.652	13.1
pocs		17.549	0.581	2154
bilinear		21.815	0.682	8.8
bicubic		22.148	0.699	14.8
Yolo2pix		21.746	0.675	78.3
discogan		21.050	0.627	13.3
pix2pix	$\times 16$	21.047	0.636	13.2
pocs		16.758	0.532	2253
bilinear		20.958	0.629	9.6
bicubic		21.241	0.645	15.2

从表 1 中可以看出本文算法在测试集上的 PSNR, SSIM 超过了其他超分辨率方法, 但是相比于传统算法有选定区域的运算, 因此相比于其他超分辨率算法, 运行时间稍长。但这种劣势会随着舰船在所述图像中所占像素的减小而抵消。

为了直观的比较成像效果, 将本算法的生成图像和其他算法生成的图像排列对比, 实验结果如图 5、图 6、图 7 所示。

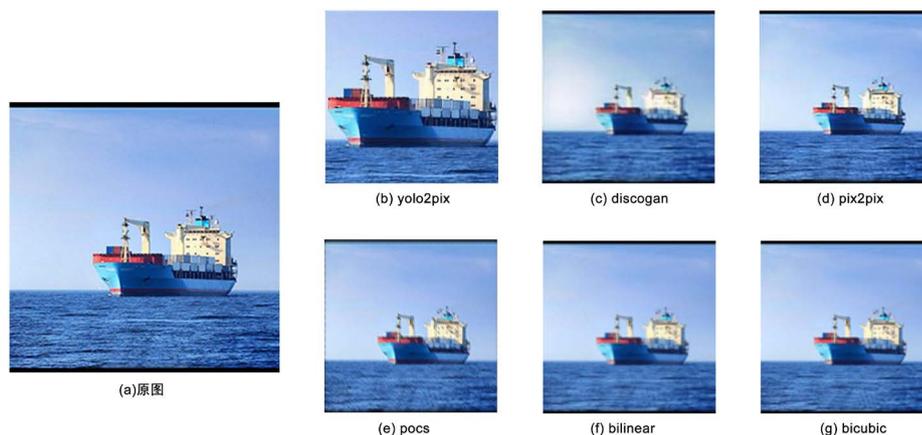


Figure 5. Comparison of results of different super-resolution methods at 4-fold scaling ratio
图 5. 4 倍缩放比率下不同超分辨率方法的结果比对

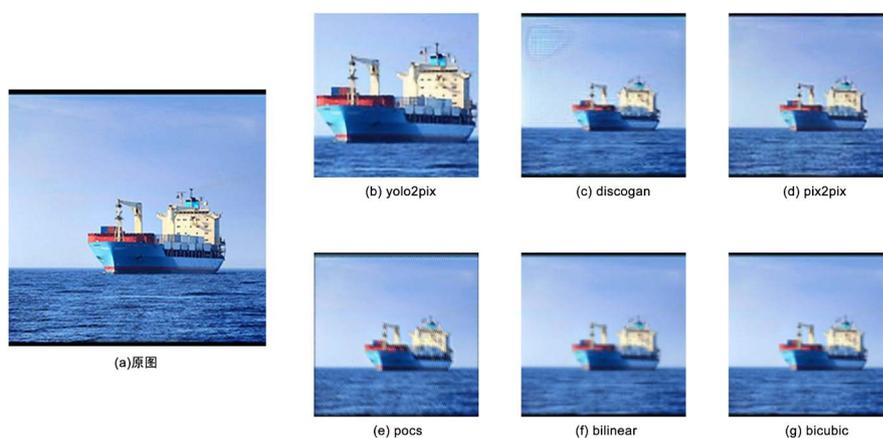


Figure 6. Comparison of results of different super-resolution methods at 9-fold scaling ratio
图 6. 9 倍缩放比率下不同超分辨率方法的结果比对



Figure 7. Comparison of results of different super-resolution methods at 16-fold scaling ratio
图 7. 16 倍缩放比率下不同超分辨率方法的结果比对

从数据上看,排除效果较差的凸集投影(pocs)算法,本文算法相比于其他算法,在重建倍数 4 下,PSNR 最少提高 0.658 dB,平均提高 0.94 dB,SSIM 最少提高 0.014,平均提高 0.045。在重建倍数 9 下,PSNR 最少提高 0.637 dB,平均提高 0.75 dB,SSIM 最少提高 0.016,平均提高 0.045。在重建倍数 16 下,PSNR 最少提高 0.505 dB,平均提高 0.67 dB,SSIM 最少提高 0.03,平均提高 0.04。总体来说,在数据集上的峰值信噪比(PSNR)平均提高了 0.79 dB,结构相似性(SSIM)平均提高了 0.04。人工观察效果来看,本文方法的整体视觉效果更好,在重建倍数 4 下,本文算法恢复了更多的船艇结构细节,其船体部分棱角更加分明,在重建倍数 16 下,船体上部分窗户细节恢复的比较真实。但是船体的侧面细节纹理没有有效重建。

5. 结语

本文提出基于目标检测网络的图像超分辨率重建方法,该方法首先框选出需要超分辨率的区域,而后对选定区域使用基于 GAN 的超分辨率算法,从而达到提升图像清晰度的目的。第一阶段利用 Darknet-53 网络快速检测目标,获得目标的位置尺寸和置信度信息。第二阶段利用生成对抗模型,利用 U-Net 结构和 PatchGAN 构建了端对端的超分辨率模型。结合自建舰船数据库,让网络更有针对性的学习舰船结构特征,从而从低分辨率图像中恢复出拥有更多纹理信息的高分辨率图像。本文方法与其他超分辨率算法相比,不论是在主观重建效果还是客观评价标准上都有所提高的,且重建出的图像具有更高的质量并显示更精细的细节。在实际应用中,一幅画面中可能出现不止一艘船艇,本文使用的 Darknet-53 网络对图像不同部分一次性给出全部置信度参数,可以一次检测多艘船艇,因此本文算法可以满足实际使用的需求,但在检测小目标时算法效果有待考证,此为下一步研究方向。

参考文献

- [1] 李欣, 崔子冠, 朱秀昌. 超分辨率重建算法综述[J]. 电视技术, 2016, 40(9): 1-9.
- [2] Glasner, D., Bagon, S. and Irani, M. (2009) Super-Resolution from a Single Image. In: *Proceedings of the 2009 IEEE 12th International Conference on Computer Vision*, IEEE, Piscataway, 349-356. <https://doi.org/10.1109/ICCV.2009.5459271>
- [3] 郑伟成. 基于深度学习的目标检测算法综述[C]//中国计算机用户协会网络应用分会. 中国计算机用户协会网络应用分会 2018 年第二十二届网络新技术与应用年会论文集. 北京: 中国计算机用户协会网络应用分会, 北京联合大学北京市信息服务工程重点实验室, 2018: 5.
- [4] Redmon, J., Divvala, S., Girshick, R., et al. (2015) You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of CVPR 2015*, Boston, 7-12 June 2015, 779-788. <https://doi.org/10.1109/CVPR.2016.91>
- [5] Zhong, Z., Jin, L. and Xie, Z. (2015) High Performance off Line Handwritten Chinese Character Recognition Using Google Net and Directional Feature Maps. *International Conference on Document Analysis and Recognition*, Nancy, 23-26 August 2015, 846-850. <https://doi.org/10.1109/ICDAR.2015.7333881>
- [6] 南方哲, 钱育蓉, 行艳妮, 赵京霞. 基于深度学习的单图像超分辨率重建研究综述[J/OL]. 计算机应用研究, 1-7. <http://kns.cnki.net/kcms/detail/51.1196.TP.20181219.1953.003.html>, 2019-04-03.
- [7] Dong, C., Loy, C.C., He, K., et al. (2014) Learning a Deep Convolutional Network for Image Super-Resolution. In: *Computer Vision ECCV 2014*, Springer International Publishing, Berlin, 184-199. https://doi.org/10.1007/978-3-319-10593-2_13
- [8] Ledig, C., Theis, L., Huszar, F., et al. (2016) Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. *Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 105-114. <https://doi.org/10.1109/CVPR.2017.19>
- [9] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y. (2014) Generative Adversarial Nets. In: *Advances in Neural Information Processing Systems*, The MIT Press, Cambridge, 2672-2680.
- [10] Isola, P., Zhu, J.Y., Zhou, T., et al. (2016) Image-to-Image Translation with Conditional Adversarial Networks. *Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 5967-5976. <https://doi.org/10.1109/CVPR.2017.632>
- [11] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation.

-
- [12] Li, C. and Wand, M. (2016) Pre-Computed Real-Time Texture Synthesis with Markovian Generative Adversarial Networks. *European Conference on Computer Vision*, Amsterdam, 8-16 October 2016, 702-716. https://doi.org/10.1007/978-3-319-46487-9_43
- [13] Keys, R. (2003) Cubic Convolution Interpolation for Digital Image Processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **29**, 1153-1160. <https://doi.org/10.1109/TASSP.1981.1163711>
- [14] Kim, T. (2017) Learning to Discover Cross-Domain Relations with Generative Adversarial Networks.
- [15] 杨庆怡, 黄灿, 张琼. 基于 POCS 算法的超分辨率图像重建技术[J]. 计算机光盘软件与应用, 2012(16): 61-62.

知网检索的两种方式:

1. 打开知网首页: <http://cnki.net/>, 点击页面中“外文资源总库 CNKI SCHOLAR”, 跳转至: <http://scholar.cnki.net/new>, 搜索框内直接输入文章标题, 即可查询;
或点击“高级检索”, 下拉列表框选择: [ISSN], 输入期刊 ISSN: 2325-6753, 即可查询。
2. 通过知网首页 <http://cnki.net/>顶部“旧版入口”进入知网旧版: <http://www.cnki.net/old/>, 左侧选择“国际文献总库”进入, 搜索框直接输入文章标题, 即可查询。

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: jisp@hanspub.org