doi:10.3772/j.issn.1002-0470.2019.05.003

基于深度强化学习和循环卷积神经网络的图像恢复算法①

杨海清② 徐勇军③ 王明雪

(浙江工业大学信息与通信工程学院 杭州 310014)

摘 要 图像恢复是指通过算法来实现恢复出图像残缺部分的信息用来补全残缺的图像。在最近几年由于深度学习的快速发展,深度学习在图像恢复领域也取得了很好的效果。然而现有的方法都需要海量的数据来训练实现图像恢复的研究。本文结合强化学习中的 actor-critic 算法和生成式对抗网络的网络结构提出了一种新的图像决策算法和循环卷积神经网络结构来实现图像恢复。在数据集 CelebA,BSDS500,Pascal Voc2012 上的实验表明,在较少数据量的情况下,该方法有效地恢复出了残缺图像的信息,与流行的图像恢复算法相比取得了较好的恢复效果。

关键词 强化学习,图像恢复,深度学习,生成式对抗网络,循环卷积

0 引 言

图像恢复是一种从破损图像中将图像恢复还原的技术,是计算机视觉领域中的热门研究领域之一。最近几年的一些先进算法[1-3]基本都采用了生成式对抗神经网络和编解码神经网络的方法来实现图像的复原。文献[1,2,4,5]采用的图像恢复算法需要有大量的数据来训练它们的神经网络,然而在很多情况下并没有海量的训练数据来训练目标网络。

深度强化学习在图像标注[6,7]、句子生成[8]、电脑游戏[9]和控制理论[10]等方面都有很广泛的研究应用。本文结合了 actor-critic[11]和生成式对抗神经网络的模型结构来设计整个网络。整个网络由图像生成网络和图像判决网络两部分组成。图像生成网络主要用于对缺损图像的恢复生成,在训练过程中,判决网络需要对生成的图片和标准图像的判决值进行对比判决直到完成训练。

在深度强化学习算法应用于图像恢复的研究上,本研究主要做了以下工作:

(1) 提出了新的循环卷积神经网络,使得在图像恢复效果依然较好的前提下大量减少了训练数据。

(2) 将强化学习算法 actor-critic 和生成式对抗神经网络结合构造新的网络框架,相比于原有的 actor-critic 网络,网络数据量更少,训练速度更快。

(3) 从网络的不同层提取特征进行多尺度融合,增强了网络的鲁棒性

1 相关工作

在图像恢复领域卷积神经网络的应用十分普遍,文献[1,2,4,5,12]的算法都是结合了卷积神经网络,取得了很好的实验效果。卷积神经网络在非线性问题的拟合上有很好的效果,经常被用来作为特征提取的工具。文献[1]用卷积神经网络来作为提取图像纹理特征和上下文联系的工具。文献[5]用神经网络来构造深度生成模型。受文献[13]的启发,本文使用扩展卷积来代替自编解码网络中的全连接部分,有效地提高了卷积网络的感受野,对特

① 浙江省科技计划(2017C37054)资助项目。
② 男,1971 年生,博士,副教授;研究方向:计算机视觉及应用;E-mail: yanghq@zjut.edu.cn
③ 通信作者,E-mail: xuyongjun@zjut.edu.cn
(收稿日期:2018-09-25)

征的提取更为全面。本文使用文献[14]中的 leaky-relu 代替了 relu 来做为对应的激活函数。在实验中发现 leaky-relu 更适合于网络。

强化学习是一种激励学习,智能体通过和环境交互来获得相应的激励,通过激励来判决当前策略是否有效。在与环境的不断交互中来优化自己的行为策略。在图像恢复方面,可以把缺损图像当做原始环境,图像生成网络通过给图像不同的像素填充来实现对环境的互动,生成的图像为更新后的环境,判决网络给出当前生成的图像是否正确恢复了对应点的像素来给出激励,指导生成网络的策略更新。随着强化学习的被重视和深度学习的迅猛发展,深度强化学习也越来越得到大家的认可。文献[10]用深度 Q 学习使计算机的游戏能力达到了人类的游戏水平,文献[9]通过结合蒙特卡洛树搜索和深度神经网络结合设计了 Go 算法,文献[15]通过结合 actor-critic 模型实现了视觉导航系统的设计。但是强化学习的方法目前在图像恢复方面的研究还是很少,文献[16]用强化学习的方法作为整个图像恢复网络中的一个工具链来实现图像恢复。本文借鉴了文献[11]中提出的 actor-critic 强化学习算法和生成对抗网络(generative adversarial nets,GANs)的实现理念来设计生成网络和判决网络。

2 图像恢复算法

本节定义了本文在图像恢复中的算法和网络结构,并介绍了本研究的训练过程。

2.1 问题描述

把图像恢复作为一个判决过程。在强化学习中判决过程是智能体与环境交互的过程,在这一过程中智能体会尝试多种行为使得最后的回报达到最大化。在图像恢复问题上给定缺损图片 $I_{\text{corrupted}}$,目标是得到完整的图片 I_{complete},行为就是填充的像素 $pixel_{c\times n\times m}$,$c$ 是输入图片的通道数,$n\times m$ 是图片的维度。生成网络根据策略 p_π 来选择行为,生成图片。判决网络通过状态值函数 v_θ 来判断生成图片质量。初始环境为输入的缺损图片 $I_{\text{corrupted}}$,当 $pixel^t_{c\times n\times m}$ 作用于 $I_{\text{corrupted}}$ 后,环境更新,直到最后生成标准图像 I_{label}。

2.2 状态空间和动作空间

在强化学习的决策过程中会有一系列的动作产生以及相应的状态。在问题描述中,在 t 时刻的状态 s_t 是由缺损图像 $I_{\text{corrupted}}$ 和动作 $pixel^t_{c\times n\times m}$ 组成。动作空间为每个像素点的像素变化,每个像素的变化范围为[0, 255],图像的像素空间为图像通道数乘以像素个数。

2.3 图像生成网络

图 1 所示为图像生成网络的网络结构以及相关参数设置。网络在接收到输入图片后经过网络的 n 次循环运算,然后输出图片。循环运算可以得到图像更为丰富的特征信息,取代了深层网络的设计方法,有效地减少了网络的参数,使得对图像纹理和空间结构的学习更加全面,学习效率更高。用 p_π 来表示根据策略 π 选择动作的概率,当前状态 s_t,网络根据策略生成行为 $a_t = pixel^{t+1}_{c\times n\times m}$ 和更新后的环境状态 s_{t+1}。将更新后的状态 s_{t+1} 作为新的环境输入网络进行循环运算。生成网络可表示为

$$s_t = \{I_{\text{corrupted}}, pixel^t_{c\times n\times m}\} \tag{1}$$

$$s_{t+1} = p_\pi(a_t, s_t) \tag{2}$$

2.4 判决网络

图 2 为判决网络的网络结构。判决网络也称为值网络,该网络用来计算生成图像和标准图像的状态值大小。判决网络表示如下:

$$v = CNN_v(p_\pi(a_t, s_t) \mid \lambda) \tag{3}$$

$$r = -\sum_i^c \sum_j^n \sum_k^m \mid pixel^t_{i\times j\times k} - pixel^{label}_{i\times j\times k} \mid \tag{4}$$

$$q = r + \gamma \times v \tag{5}$$

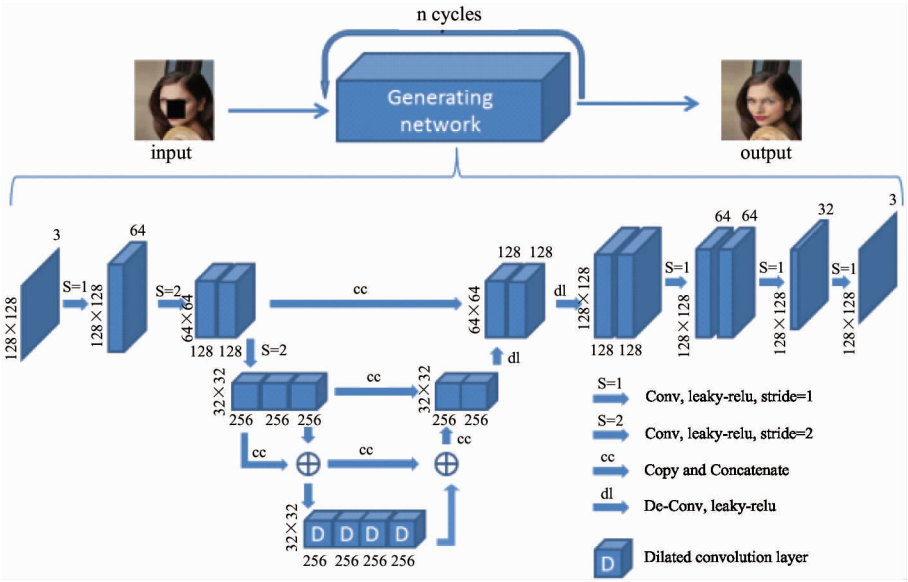
式中,λ 为判决网络的参数,r 是当前环境对当前动作的奖励值,CNN_v 表示判决网络,γ 为衰减系数,q 为最后得到的状态值。

2.5 整体网络结构

完整的图像恢复网络由图像生成和图像判决两部分组成。受文献[17]的启发,本文采用了 3 段跳跃式连接用于特征提取以减小梯度消失或爆炸对网络的影响。整体网络的表达式如下所示:

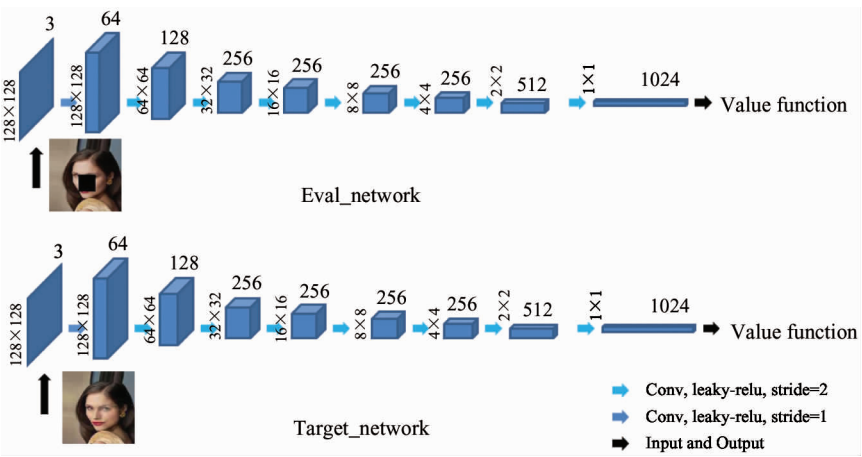
$$s_t = \{I_{\text{corrupted}}, pixel^t_{c\times n\times m} \mid \theta_t\} \tag{6}$$

$$v^{\text{eval}} = CNN^{\text{eval}}_v(p_\pi(a_t, s_t) \mid \lambda^{\text{eval}}_t) \tag{7}$$



图像生成网络作为策略生成网络对当前状态来选择对应的动作。我们结合了编解码式的网络结构和残差算法，用扩展卷积代替了全连接层，用端到端的方法训练网络。

图 1　图像生成网络结构图



判决网络由孪生网络构成，一个是估计网络输入为生成图像，另一个为目标网络输入为标准图像，图像输入大小均为 128 ×128 像素。最后生成相应状态值来判断图像优劣。

图 2　判决网络结构图

$$v^{\text{target}} = CNN_v^{\text{target}}(I_{\text{label}} \mid \lambda_t^{\text{target}}) \tag{8}$$

$$\lambda_t^{\text{target}} = \lambda_{t+T}^{\text{eval}} \tag{9}$$

$$TD_{\text{error}} = r_t + v_\lambda^{\text{target}} - v_\lambda^{\text{eval}} \tag{10}$$

$$\text{loss}_p = \mid p_\pi(a_t, s_t) - I_{\text{label}} \mid \tag{11}$$

其中，θ_t 是生成网络的参数权重，T 是时间间隔，每过时间 T 判决网络中的目标网络就会用估计网络的参数来更新自己的网络参数。$\lambda_t^{\text{target}}$ 和 λ_t^{eval} 分别是目标网络和估计网络的参数权重。梯度更新表示如下：

$$\nabla G_v = \frac{\partial TD_{\text{error}}}{\partial \lambda_t^{\text{eval}}} \tag{12}$$

$$\nabla G_p = \frac{\partial loss_p}{\partial \theta_t} \tag{13}$$

受文献[1,18]的启发，本文用预训练的 VGG-19 生成的特征图来定义内容损失，表示如下：

$$loss_{\text{content}} = \frac{\mid \phi_i(p_\pi(a_t, s_t)) - \phi_i(I_{\text{label}}) \mid}{C_i \times H_i \times W_i} \tag{14}$$

其中，ϕ_i 是从 VGG-19 的 i_th 卷积层提取的特征，C_i、H_i 和 W_i 分别为对应特征的通道数、长和宽。根

据文献[4]中的实验经验,本文采用了 VGG-19 的 *relu*3 _1 和 *relu*4 _1 两层的特征。

为了得到更好的颜色恢复,加入了颜色损失来优化颜色恢复,根据文献[18]中所提到的高斯模糊算法,颜色损失表达式如下:

$$loss_{\text{color}} = \| C_b - L_b \|_2^2 \tag{15}$$

$$C_b(i,j) = \sum_{k,w} C(i+k, j+w) \cdot G(k,w) \tag{16}$$

$$G(k,w) = A\exp\left(-\frac{(k-\mu_c)^2}{2\sigma_c} - \frac{(w-\mu_l)^2}{2\sigma_l}\right) \tag{17}$$

其中,C_b 和 L_b 为 $I_{\text{corrupted}}$ 和 I_{label} 的模糊表示。$G(k,w)$ 为高斯模糊,本文中常数项大小为 $A = 0.053$,$\sigma_{cl} = 3$,$\mu_{cl} = 0$,常数的取值根据实验中的参数调试来确定。总的损失函数表示为

$$loss_{\text{total}} = loss_p + loss_{\text{color}} + 10^{-4} \cdot loss_{\text{content}} \tag{18}$$

3 实验与结果

本节对比了不同损失函数下的图像恢复效果,本文提出的结合深度强化学习的图像恢复算法实现是基于 TensorFlow 实现的。硬件设备上使用了一个 NVIDIA 1060 GPU 用作图像处理加速。在神经网络的训练中,每个数据集分别随机抽取 100 张图片作为训练集,抽取 50 张作为验证集,50 张作为测试集。在生成网络训练部分,训练回合次数为 100,批次大小为 10。在每一回合的训练中,设置 30 次循环训练,本实验发现在 30 次循环网络对图像的优化表现得不那么明显。在循环初始把输入图片作为状态 S,输出图片作为下一个状态 $S\_$,进入第二次循环后,将 $S\_$ 作为初始状态输入网络,一直到达到目标值或者达到 30 次循环结束本回合。图 3 展示了在相同训练样本和训练次数的情况下,本算法与 PM[19](PatchMatch),EPLL[20],CE[21] 和 SI[3](semantic image inpainting with deep generative models),FF[22](free-form image inpainting with gated convolution)之间的恢复能力的对比。本实验在 CelebA[23],BSDBS500[24],Pascal Voc2012[25]这些公开数据集上都做了对比实验。表 1 中对比了不同损失函数对图像恢复的影响。图 4 展示了不同损失函数情况下,网络对图像恢复的效果。

图 3 不同算法间的修复结果对比

表 1 不同损失函数的结果对比
(PSNR 值越大表示恢复的越好)

Method	Mean L2 Loss(%)	PSNR
$loss_p$	1.24e-3	22.76
$loss_{total}$	7.34e-4	25.13
本文($loss_{\text{total}}$ + residual learning)	4.68e-4	27.28

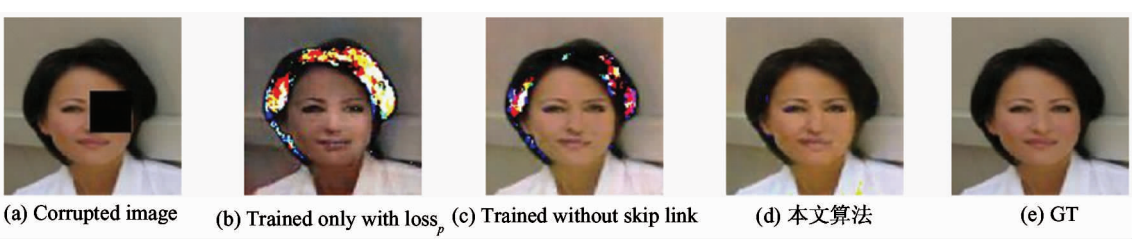


图 4 不同损失函数的生成图像对比

图 5 展示了该算法在 CelebA,BSDS500,Pascal Voc2012 数据集上的实验结果。图片从上到下排列分别为原图,遮掩图,修复图。其中人脸部分为 CelebA 数据集实验结果。自然图像部分为 BS-

DS500 数据集上的实验结果。车和猫部分是在 Pascal Voc2012 数据集上的实验结果。图 6 展示了不同尺度以及在非固定位置的缺损情况下，本文算法对图像修复的能力。图 3 展示了 PM[24]，EPLL[23]，CE[25]，SI[3]，FF[22]算法和本文算法在图像修复上的能力对比。

图 4　CelebA，BSDS500，Pascal Voc2012 数据集上的实验结果

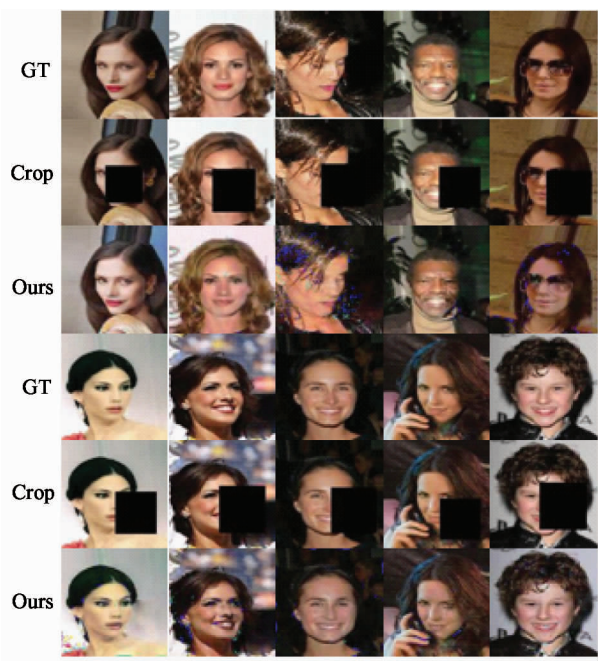


图 6　不同位置和遮盖大小情况下的修复结果

4　结　论

与原有的深度强化学习的神经网络结构相比，本研究去除了生成网络的孪生结构并添加了循环机制，在判决网络部分保留了孪生结构。循环机制的设置使得生成网络可以依靠极少的数据量来达到一个较好的学习状态，判决网络的孪生结构保证了整个网络的鲁棒性，两者结合有效地提高了网络的运算性能。

在将来的工作中将会对三维立体的图像缺损进行恢复还原的尝试。

参考文献

[1] Yang C，Lu X，Lin Z，et al. High-resolution image inpainting using multi-scale neural patch synthesis[C]. In：2017 IEEE Conference on Computer Vision and Pattern Recognition，Honolulu，USA，2017. 4076-4084

[2] Iizuka S，Simo-Serra E，Ishikawa H. Globally and locally consistent image completion[J]. *ACM Transactions on Graphics*，2017，36(4)：1-14

[3] Yeh R A，Chen C，Lim T Y，et al. Semantic image inpainting with deep generative models[C]. In：2017 IEEE Conference on Computer Vision and Pattern Recognition，Honolulu，USA，2017. 5485-5493

[4] Yang C，Lu X，Lin Z，et al. High-resolution image inpainting using multi-scale neural patch synthesis[C]. In：2017 IEEE Conference on Computer Vision and Pattern Recognition，Honolulu，USA，2017. 4076-4084

[5] Yeh R A，Chen C，Lim T Y，et al. Semantic image inpainting with deep generative models[C]. In：2017 IEEE Conference on Computer Vision and Pattern Recognition，Honolulu，USA，2017 . 5485-5493

[6] Zhang L，Sung F，Liu F，et al. Actor-critic sequence training for image captioning[J]. *arXiv*：1706. 09601v2，2017

[7] Ren Z，Wang X，Zhang N，et al. Deep reinforcement learning-based image captioning with embedding reward[C]. In：2017 IEEE Conference on Computer Vision and Pattern Recognition，Honolulu，USA，2017. 1151-1159

[8] Bahdanau D，Brakel P，Xu K，et al. An actor-critic algorithm for sequence prediction[J]. *arXiv*：1607. 07086，2016

[9] Silver D，Huang A，Maddison C J，et al. Mastering the game of Go with deep neural networks and tree search[J]. *Nature*，2016，529(7587)：484-489

[10] Volodymyr M，Koray K，David S，et al. Human-level control through deep reinforcement learning[J]. *Nature*，2015，518(7540)：529-533

[11] Konda V. Actor-critic algorithms[J]. *SIAM Journal on Control & Optimization*，2003，42(4)：1143-1166

[12] Zhang K，Zuo W，Gu S，et al. Learning deep CNN denoiser prior for image restoration[C]. In：2017 IEEE Conference on Computer Vision and Pattern Recognition，

Honolulu, USA , 2017. 2808-20817

[13] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions[J]. *arXiv*:1511.07122v3,2015

[14] He K, Zhang X, Ren S, et al. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification[C]. In: 2015 IEEE International Conference on Computer Vision, Santiago, Chile, 2015. 1026-1034

[15] Zhu Y, Mottaghi R, Kolve E, et al. Target-driven visual navigation in indoor scenes using deep reinforcement learning[J]. *arXiv*:1609.05143,2017

[16] Yu K, Dong C, Lin L, et al. Crafting a toolchain for image restoration by deep reinforcement learning[J]. arXiv:1804.03312v1, 2018

[17] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition , Las Vegas, USA, 2016 . 770-778

[18] Ignatov A, Kobyshev N, Timofte R, et al. DSLR-quality photos on mobile devices with deep convolutional networks[C]. In: 2017 IEEE International Conference on Computer Vision. Venice, Italy, 2017. 3297-3305

[19] Barnes C, Shechtman E, Finkelstein A, et al. PatchMatch: a randomized correspondence algorithm for structural image editing[J]. *ACM Transactions on Graphics*, 2009, 28(3): 24:1-24:11

[20] Zoran D, Weiss Y. From learning models of natural image patches to whole image restoration[C]. In: 2011 International Conference on Computer Vision, Washington, USA , 2011. 479-486

[21] Pathak D, Krahenbuhl P, Donahue J, et al. Context encoders: feature learning by inpainting[C]. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition , Las Vegas, USA, 2016. 2536-2544

[22] Yu J H, Lin Z, Yang J M, et al. Free-form image inpainting with gated convolution [J]. *arXiv*: 1806.03589v1,2018

[23] Liu Z, Luo P, Wang X, et al. Deep learning face attributes in the wild[C]. In: 2015 IEEE International Conference on Computer Vision, Santiago, Chile, 2015. 3730-3738

[24] Arbelaez P, Maire M, Fowlkes C, et al. Contour detection and hierarchical image segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(5):898-916

[25] Gidaris S, Komodakis N. Object detection via a multi-region and semantic segmentation-aware CNN model[C]. In: 2015 IEEE International Conference on Computer Vision, Santiago, Chile, 2015. 1134-1142

Image restoration algorithm based on deep reinforcement learning and circular convolutional neural network

Yang Haiqing, Xu Yongjun, Wang Mingxue

(College of Information Engineering, Zhejiang University of Technology, Hangzhou 310014)

Abstract

Image restoration refers to the use of an algorithm to recover the image of the defective part of the image to complement the defective image. In recent years, due to the rapid development of deep learning, deep learning has achieved good results in the field of image restoration. However, existing methods require massive amounts of data to train the study of image restoration. A new image decision algorithm and a new circular convolutional neural network structure are proposed to realize image restoration by combining the actor-critic algorithm in reinforcement learning and the network structure of the generated confrontation network. Experiment results of the datasets CelebA, BSDS500, Pascal Voc2012 show that, in the case of less data, the proposed method can effectively recover the information of the defective image, which is better than the popular image restoration algorithm.

Key words: reinforcement learning, image restoration, deep learning, generative adversarial networks, circular convolution