·图形图像处理· 文章编号：1000-3428(2019)04-0223-05 文献标志码：A 中图分类号：TP391

基于条件生成对抗网络的咬翼片图像分割

蒋 芸，谭 宁，张 海，彭婷婷

（西北师范大学 计算机科学与工程学院，兰州 730000）

摘 要： 现有基于 U 型网络（U-Net）的咬翼片图像分割方法将咬翼片 X 射线图像分割成龋齿、牙釉质、牙本质、牙髓、牙冠、修复体和牙根管 7 个部分，但分割准确率偏低。为此，提出一种改进的咬翼片图像分割方法，将条件生成对抗网络与 U-Net 相结合对咬翼片进行分割，使判别器与生成器相互优化，获得具有更多上下文信息的分割特征图。实验结果表明，改进方法的 Dice 系数相比 U-Net 方法提升了 0.133，分割准确率更高。

关键词： 生成对抗网络；图像分割；深度学习；U 型网络；数据增强

中文引用格式： 蒋芸，谭宁，张海，等. 基于条件生成对抗网络的咬翼片图像分割[J]. 计算机工程，2019，45(4)：223-227.

英文引用格式： JIANG Yun，TAN Ning，ZHANG Hai，et al. Bitewing radiography image segmentation based on conditional generative adversarial network[J]. Computer Engineering，2019，45(4)：223-227.

Bitewing Radiography Image Segmentation Based on Conditional Generative Adversarial Network

JIANG Yun，TAN Ning，ZHANG Hai，PENG Tingting

（School of Computer Science and Engineering，Northwest Normal University，Lanzhou 730000，China）

【Abstract】 The existing bitewing radiography image segmentation method based on U-Net divides the X-ray image of the bitewing radiography into caries，enamel，dentin，pulp，crown，prosthesis and root canal，but the segmentation accuracy is low. So，this paper proposes an improved method to segmentation bitewing radiograpy images. The conditional Generative Adversarial Network（cGAN） combined with U-Net to segmentation the bitewing radiograpy images. It optimizes the discriminator and the generator to obtain a segmentation feature map with more context information. Experimental results show that the Dice coefficient of the improved method is improved by 0.133 compared to the U-Net method，and the segmentation accuracy is higher.

【Key words】 Generative Adversarial Network（GAN）；image segmentation；deep leaning；U-Net；data enhancement

DOI： 10.19678/j.issn.1000-3428.0050445

0 概述

图像分割是指根据灰度、颜色、纹理和形状等特征将图像划分成若干互不交迭的区域，并使这些特征在同一区域内呈现出相似性，而在不同区域间呈现出明显的差异性。常用的图像分割方法有阈值分割[1]、区域分割[2]、边缘检测分割[3]等。医学图像分割一直是生物医学工程领域的研究热点，对于人体的不同组织和器官的分割问题有不同方法。近年来，无监督学习成为研究热点，变分自编码器[4]、生成对抗网络（Generative Adversarial Network，GAN）[5]等无监督模型受到越来越多的关注。在人工智能高速发展的时代，GAN 的提出不仅满足了相关领域的研究和应用需求，也带来了新的发展动力。特别是在图像和视觉领域中对 GAN 的研究和应用较为广泛，已经可以通过随机数字生成人脸、从低分辨率图像生成高分辨率图像[6]等。此外，GAN 已经开始逐渐应用到医学图像处理中，例如通过术前磁共振图像直接生成病人特定的超声探头，从而诱导前列腺运动的模型[7]，在超声探头的空间位置上有条件地对解剖学图像进行采样[8]，检测恶性前列腺癌[9]等问题的研究中。

由于龋齿会破坏牙齿结构传播细菌性疾病，因此特别是对于儿童，龋齿会影响儿童的发育，对颌面部造成影响，龋洞会破坏牙胚。牙科医生主要根据咬翼片 X 射线图像来诊断和治疗龋齿。自动化龋齿病变检测技术为牙科医生提供潜在的诊断数据，并有助于识别各种疾病的迹象[10]。目前，牙科鉴定过程为手工进行，非常耗时且主观性差。为解

基金项目： 国家自然科学基金（61163036）；甘肃省自然科学基金（1606RJZA047）；甘肃省高校研究生导师科研项目（1201-16）；西北师范大学第三期“知识与创新工程”科研骨干项目（nwnu-kjcxgc-03-67）。

作者简介： 蒋 芸（1970—），女，教授、博士，主研方向为图像分割、数据挖掘、粗糙集理论及应用；谭 宁、张 海、彭婷婷，硕士研究生。

收稿日期： 2018-02-07 **修回日期：** 2018-04-11 **E-mail：** tanning2315@126.com

决该问题,文献[11]提出基于小波变换的图像多分辨率梯度信息的牙齿图像分割方法。文献[12]在牙缝约束下利用模糊连接法对单颗牙齿进行分割。文献[13]建立一种自动且高效的牙齿隔离方法,能够实现上下颌分离、单齿隔离、过分割验证和下分割检测。文献[14]研究了一种基于细菌优化的模糊聚类算法,用于牙齿图像分割。文献[15]使用 U-Net 对咬翼片进行分割,将咬翼片 X 射线图像分割成龋齿、牙釉质、牙本质、牙髓、牙冠、修复体和牙根管 7 个部分,但是分割准确率偏低。本文在上述研究的基础上,引入条件生成对抗网络对咬翼片 X 射线进行分割,由于判别器有无限的建模能力,因此不断对生成器分割结果与标准分割图像的差异进行判断,生成器和判别器互相对各自的模型进行优化,使得生成器分割结果无限接近标准分割图像。

1 相关理论

近年来,生成对抗网络受到了越来越多研究人员的广泛关注,成为无监督[15]复杂概率分布学习的主流方法。生成器(G)和判别器(D)采用当前热门的深度神经网络[16]。

1.1 生成对抗网络

生成对抗网络由 2 个神经网络组成,包含生成器(G)和判别器(D)。生成器(G)的目的是生成近似真实数据分布 p_{data}的样本来欺骗判别器(D),使判别器(D)无法区分输入的数据来自真实数据 p_{data}还是生成器(G)。判别器(D)的目的是正确区分数据样本来自真实的数据分布 p_{data}还是生成器(G),当输入的数据来自真实的数据分布 p_{data}时,判别器(D)的目标是使输出的概率 $D(x)\approx 1$,当输入的数据来自生成器(G)时,判别器(D)的目标是使 $D(G(z))\approx 0$,同时生成器(G)的目标是使 $D(G(z))\approx 1$。2 个网络在一个极小-极大的游戏中相互迭代、相互竞争优化各自的网络参数,使目标函数达到那什均衡[17],当收敛时,期望 $p_{\text{data}}=p_g$,其中 p_g 是生成器(G)生成的数据样本分布。本文使用以下目标函数优化生成器的参数(θ_G)和判别器的参数(θ_D):

$$\max_D \min_G(\theta_G,\theta_D)=\mathbb{E}_{x\sim p_{\text{data}}}[\lg D(x)]+\mathbb{E}_{z\sim p_z(z)}[\lg(1-D(G(z)))] \tag{1}$$

其中,x 取自于真实的数据分布 p_{data},z 取自于先验分布 $p_z(z)$(例如正态分布),$\mathbb{E}(\cdot)$表示计算期望值。

1.2 条件生成对抗网络

条件生成对抗网络(condition GAN,cGAN)[18]是生成器(G)和判别器(D)都输入一个额外的信息 y 作为条件,y 可以是任何类型的辅助信息(例如手写体,y 可以表示为数字类别)。额外辅助信息 y 与样本数据 x 进行拼接后输入到生成器(G)和判别器(D),使生成对抗网络扩展为条件模型,目标函数转换为:

$$\min_G \max_D \mathcal{L}(D,G)=\mathbb{E}_{x\sim p_{\text{data}}(x)}[\lg D(x|y)]+\mathbb{E}_{z\sim p_z(z)}[\lg(1-D(G(x|y)))] \tag{2}$$

2 基于 cGAN 的咬翼片图像分割方法

2.1 损失函数

生成对抗网络的生成模型需要学习一个从随机噪声 z 到输出图像 y 之间的映射关系,$G:z\rightarrow y$。条件生成对抗网络[18]需要学习图片 x 和随机噪声 z 到输出图像 y 之间的映射,$G:\{x,z\}\rightarrow y$。

实验中所用的条件生成对抗网络的目标函数描述如下:

$$\mathcal{L}_{\text{cGAN}}(G,D)=\mathbb{E}_{x\sim p_{\text{data}}(x),y}[\lg D(x,y)]+\mathbb{E}_{x\sim p_{\text{data}}(x),z\sim p_z(z)}[\lg(1-D(x,G(x,z)))] \tag{3}$$

其中,x 为输入的图片样本数据,y 是通过人工标注的标准分割图像,生成器(G)尝试最小化目标函数,判别器(D)尝试最大化目标函数:

$$G^*=\min_G \max_D \mathcal{L}_{\text{cGAN}}(G,D)$$

通过实验发现将 cGAN 的目标函数与传统损失函数(像 L_1 距离)相结合的分割效果会更好,L_1 的距离函数如下:

$$\mathcal{L}_{L_1}(G)=\mathbb{E}_{x\sim p_{\text{data}}(x),y,z\sim p_z(z)}[\|y-G(x,z)\|] \tag{4}$$

其中,y 是分割的目标图像,$G(x,z)$是训练样本数据输入生成器中生成的分割图像。最终目标函数变换为:

$$G^*=\arg\min_G \max_D \mathcal{L}_{\text{cGAN}}(G,D)+\lambda\mathcal{L}_{L_1}(G) \tag{5}$$

由于 L_1 距离函数会产生模糊化效果,因此通过加入超参 λ 进行控制。当生成器能够对输入的咬翼片 X 射线图像进行准确分割时,对参数的改变较敏感,超参 λ 的选择非常重要,因此设置超参 $\lambda=100$ 较合理。

2.2 模型结构

生成器(G)的训练过程(见图 1)分为 2 个部分:1)输入一张咬翼片 X 射线图像到生成器(G),生成器(G)输出一张分割后的图像,比较生成器(G)输出的分割图像和人工标注的标准分割图像(Ground truth)之间的误差,通过误差调整生成器(G)的权重。2)输入一对假图像(咬翼片 X 射线图像和生成器通过该咬翼片生成的分割图像)到判别器(D)中,由于生成器(G)是生成一张无限接近于目标的分割图像,使判别器(D)误认为生成器(G)输出的分割图像是人工标注的标注图像,期待判别器输出的结

果为“1”。比较判别器(*D*)输出的结果与标准的正确结果“1”之间的误差,从而优化生成器(*G*)的权重,使其生成的分割图像更加接近目标分割图像。

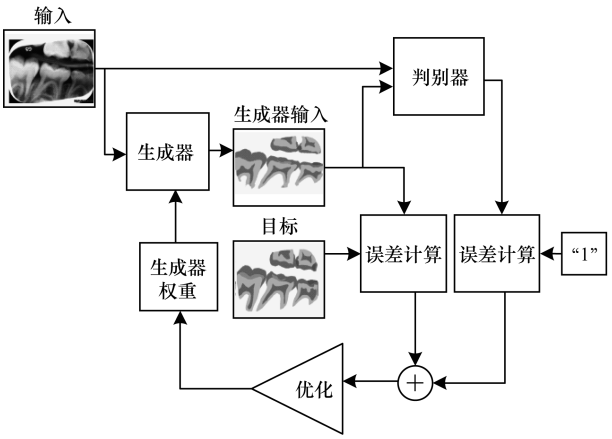


图 1　生成器(*G*)的训练过程

判别器(*D*)的训练过程(见图 2)同样分为 2 个部分:1)输入一对真实的图像对(咬翼片 X 射线图像和人工标注的标准目标分割图像)到判别器(*D*),判别器(*D*)输出判别后的结果。由于已知输入的是真实的数据样本,因此判别器(*D*)输出的结果应该为“1”,通过比较判别器(*D*)输出的结果与标准答案“1”之间的差值,然后优化判别器(*D*)的权重。2)输入一对假图像对(咬翼片 X 射线图像和生成器输出的分割图像)到判别器(*D*)中,已知输入的数据样本为假,所以判别器(*D*)输出的结果应该为“0”,比较判别器(*D*)输出的结果与标准答案“0”之间的差值,然后调整判别器(*D*)的权重,使判别器(*D*)能够正确区分图像是人工标注的标准分割图像还是由生成器(*G*)生成的分割图像。

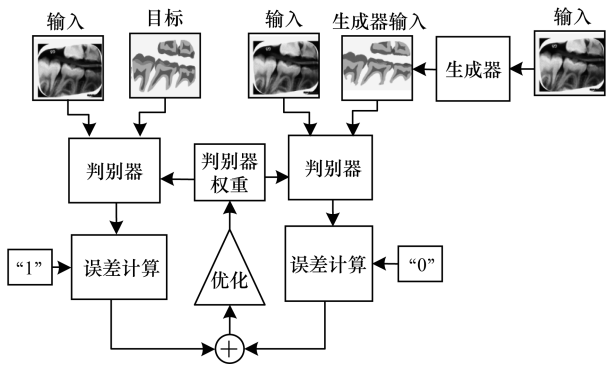


图 2　判别器(*D*)的训练过程

2.3　数据增强

由于只能提供少量咬翼片 X 射线样本图像对网络结构进行训练,但深度神经网络训练大量参数需要大量的训练样本,会造成数据饥饿,因此本文通过数据增强的方式对训练样本进行扩充,从而解决该问题。数据增强对于提高网络分类准确率、鲁棒性以及防止过拟合至关重要。通过对图像旋转、水平翻转、垂直翻转、平移变换、图像灰度值的变化等方法对训练数据进行增强。

2.4　网络结构

本文基于深度卷积对抗神经网络(DCGAN)[19]和条件生成对抗网络进行实现,使用 DCGAN 中推荐的训练参数进行训练,训练时使用 Adam 优化算法[20]($\beta_1=0.5$、$\beta_2=0.999$、$\varepsilon=10^{-8}$),学习率 $lr=0.0002$,*mini-batch* =2,每层输出结果进行批量归一化,从而减少每层之间的依赖性,提高各网络层之间的独立性[21-22]。训练 20 个周期,输出的分割图像大小为 256 像素×256 像素,生成器(*G*)和判别器(*D*)具体的网络结构如表 1、表 2 所示。

表 1　生成器(*G*)的网络结构

操作	核大小	步幅	卷积核数	归一化	激活函数
e_1:卷积	5×5	2×2	64	是	Leaky ReLU
e_2:卷积	5×5	2×2	128	是	Leaky ReLU
e_3:卷积	5×5	2×2	256	是	Leaky ReLU
e_4:卷积	5×5	2×2	512	是	Leaky ReLU
e_5:卷积	5×5	2×2	512	是	Leaky ReLU
e_6:卷积	5×5	2×2	512	是	Leaky ReLU
e_7:卷积	5×5	2×2	512	是	Leaky ReLU
e_8:卷积	5×5	2×2	512	是	Leaky ReLU
d_1:反卷积	5×5	2×2	512+512	是	ReLU
d_2:反卷积	5×5	2×2	512+512	是	ReLU
d_3:反卷积	5×5	2×2	512+512	是	ReLU
d_4:反卷积	5×5	2×2	512+512	是	ReLU
d_5:反卷积	5×5	2×2	256+256	是	ReLU
d_6:反卷积	5×5	2×2	128+128	是	ReLU
d_7:反卷积	5×5	2×2	64+64	是	ReLU
d_8:反卷积	5×5	2×2	3	—	ReLU
全连接	—	—	3	—	tanh

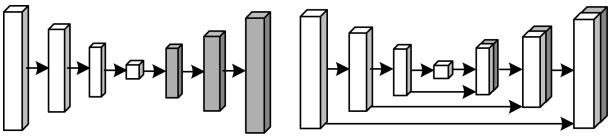
表 2　判别器(*D*)的网络结构

操作	核大小	步幅	卷积核数	归一化	激活函数
h_0:卷积	5×5	2×2	64	是	Leaky ReLU
h_1:卷积	5×5	2×2	128	是	Leaky ReLU
h_2:卷积	5×5	2×2	256	是	Leaky ReLU
h_3:卷积	5×5	2×2	512	是	Leaky ReLU
h_4:全连接	—	—	16×16×512	否	—
全连接	—	—	1	否	Sigmoid

生成器中使用 U-Net 网络结构[23],在 d_1、d_2、d_3 3 层使用 *dropout*,每层随机删除 50% 的节点以防止过拟合。

现有分割方法的网络结构多数使用编码器-解码器[24],该网络结构通过向下采样,逐渐降低采样层,直到达到一个瓶颈层,将提取的信息变为一个一维向量,此时过程被逆转,逐渐向上采样,最后还原成图像。该网络结构要求所有的信息流通过所有的网络层,包括瓶颈层。在许多图像翻译应用中,输入和输出之间共享大量低级信息,但通常期望这些信息能直接通过网络。

为使生成器(G)能够避免出现信息瓶颈,本文使用跳远连接,遵循 U-Net 网络结构,具体操作是将第 i 层和第 $n-i$ 层进行跳远连接,每个跳远连接只是简单地将第 i 层网络输出的所有通道和第 $n-i$ 层的所有输出进行连接(n 为网络结构的总层数),作为第 $n-i+1$ 层节点的输入。生成器网络结构如图 3 所示。

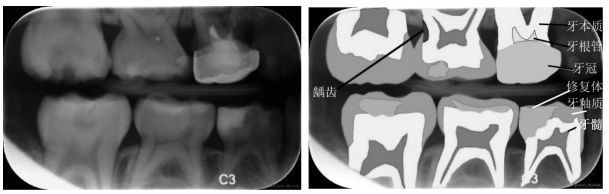


(a)编码器-解码器结构 (b)U-Net网络结构

图 3 生成器网络结构

3 实验结果与分析

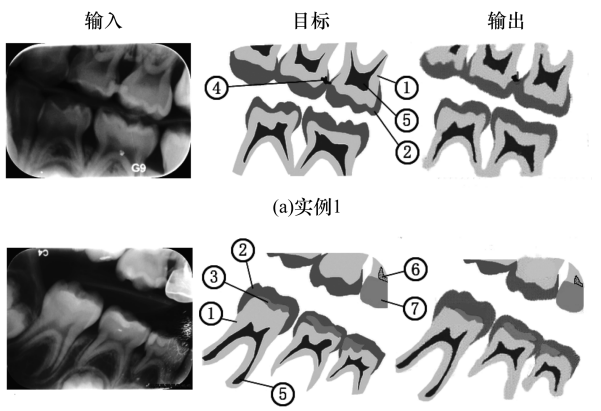
实验目标是研究一种用于龋齿检测的自动图像分割方法,识别 7 个部位,包括龋齿、牙釉质、牙本质、牙髓、牙冠、牙根管和修复体,并进行标记。图 4(a)为咬翼片 X 射线图像,图 4(b)为该图像由医生进行手工标记的 7 个部位。



(a)原始咬翼片X射线图像 (b)标记7个部位的咬翼片图像

图 4 训练样本与分割标签图像

训练的数据样本来自 80 个患者的咬翼片 X 射线图像和医生手工标记的标准分割图像对,由于图像的分辨率不统一,因此将图像重新进行缩放至 256 像素×256 像素,并且归一化灰度值到[-1,1]的范围内。通过数据增强的方法将训练数据集扩充到 28 800 对,使用这些数据集训练本文模型,训练出的模型对咬翼片 X 射线图像的分割效果如图 5 所示。可以看出,本文方法对于常见类型牙釉质、牙本质、牙髓能够进行较好的分割。



(a)实例1

(b)实例2

① 牙本质 ② 牙釉质 ③修复体 ④ 龋齿 ⑤ 牙髓 ⑥ 牙根管 ⑦ 牙冠

图 5 基于 cGAN 与 U-Net 的咬翼片 X 射线图像分割结果

实验通过精确率、准确率、真阳率、假阳率、Dice 系数这些主要指标来评估本文方法的性能。训练出的模型对测试集(40 个患者的咬翼片 X 射线图像)进行分割后,通过模型分割出的 7 个部位与医生手工标记的标准结果进行对比(见表 3),可以看出对常见部位(牙釉质、牙本质、牙髓)的分割相似度超过 75%,对龋齿、牙冠、牙根管等其他部位的分割效果还需进一步提高。

表 3 测试集上不同部位分割的性能比较结果

部位	精确率	真阳率	假阳率	Dice 系数
龋齿	0.418	0.768	0.973	0.584
牙釉质	0.588	0.886	0.923	0.759
牙本质	0.641	0.797	0.848	0.781
牙髓	0.542	0.878	0.941	0.751
牙冠	0.494	0.790	0.901	0.567
修复体	0.513	0.792	0.946	0.663
牙根管	0.433	0.721	0.958	0.597

本文方法通过结合 U-Net 和 cGAN 后,图像分割的 Dice 系数达到 0.697,而文献[25]采用 U-Net 分割的 Dice 系数最高为 0.564,相对来说提高了0.133。表 4 为 2 个分割模型的评估性能比较。

表 4 测试集上 2 个分割模型的性能比较结果

模型	精确率	真阳率	假阳率	Dice 系数
U-Net	0.453	0.613	0.983	0.564
cGAN + U-Net	0.546	0.784	0.956	0.697

实验结果表明,条件生成对抗网络对于图像分割具有较好的性能,对牙齿的常见部位(牙本质、牙釉质、牙髓)分割较准确,由于数据增强的原因,因此有大量的训练样本可以对这些部位进行训练。对于龋齿、牙根管等这些部位的分割准确率较低,如果有更多的数据样本对模型进行训练,那么分割的 Dice 系数将会更高。标准目标分割图像中有时并未标记牙齿,加大了模型训练难度,若假设提供的训练数据标注更加严格,则模型整体性能将更好。

4 结束语

本文通过使用条件生成对抗网络结合 U 型网络对咬翼片 X 射线图像进行分割,将其分割成龋齿、牙冠、牙釉质、牙本质、牙髓、牙根管和修复体 7 个部位。实验结果表明,与基于 U-Net 的分割方法相比,本文方法的咬翼片分割准确率更高,特别是对于常见部位(牙釉质、牙本质、牙髓)的分割准确率超过 75%。下一步将对包含龋齿、牙根管等部位的训练数据样本增多情况下的图像分割准确率和性能做进一步研究与测试。

参考文献

[1] AL-AMRI S S,KALYANKAR N V. Image segmentation by using threshold techniques[EB/OL]. [2018-01-16]. http://www. arXiv preprint arXiv:1005.

[2] KANG J K. Moving region segmentation using parse motion cue from a moving camera[C]//Proceedings of the 12th International Conference Intelligent Autonomous Systems. Berlin,Germany:Springer,2012:257-264.

[3] GAUR P. Recognition of 2D barcode images using edge detection and morphological operation [EB/OL]. [2018-01-16]. https://ijcsmc. com/docs/papers/April2014/V3I4201499b59. pdf.

[4] KINGMA D P, MAX W. Auto-encoding variational Bayes[EB/OL]. [2018-01-16]. https://arxiv. org/abs/1312. 6114.

[5] GOODFELLOW I J,POUGET-ABADIE J,MIRZA M, et al. Generative adversarial nets[EB/OL]. [2018-01-16]. https://arxiv. org/abs/1406. 2661.

[6] LEDIG C,THEIS L, HUSZAR F, et al. Photo-realistic single image super-resolution using a generative adversarial network[EB/OL]. [2018-01-16]. https://arxiv. org/abs/1609. 04802.

[7] HU Y P, GIBSON E, VERCAUTEREN T, et al. Intraoperative organ motion models with an ensemble of conditional generative adversarial networks [C]// Proceedings of International Conference on Medical Image Computing and Computer-assisted Intervention. Berlin,Germany:Springer,2017:368-376.

[8] HU Y P, GIBSON E, LEE L L. Freehand ultrasound image simulation with spatially conditioned generative adversarial networks[C]//Proceedings of International Workshop on Reconstruction and Analysis of Moving Body Organs. Berlin,Germany:Springer,2017:105-115.

[9] KOHL S,BONEKAMP D,SCHLEMMER H. Adversarial networks for the detection of aggressive prostate cancer [EB/OL]. [2018-01-16]. https://arxiv.org/abs/1702. 08014.

[10] TAN P L, EVANS R W, MORGAN M V. Caries, bitewings,and treatment decisions[J]. Australian Dental Journal,2002,47(2):138-141.

[11] 臧晶,宋凯. 基于多分辨率分析理论的口腔 CT 图像的增强[J]. 沈阳电力高等专科学校学报,2004,6(3):48-50.

[12] 杨玲,王中科,王云鹏. 牙缝约束下模糊连接法对单颗牙齿的分割[J]. 计算机工程与设计,2009,30(21):5031-5036.

[13] LIN P L,HUANG P W,CHO Y S,et al. An automatic and effective tooth isolation method for dental radiographs[J]. Opto-electronics Review,2013,21(1):126-136.

[14] 莫宏兵,赵刚. 基于细菌模糊聚类算法的牙齿图像分割技术研究[J]. 黑龙江医药科学,2014,34(1):108-109.

[15] RONNEBERGER O,FISCHER P,BROX T. Dental X-ray image segmentation using a U-shaped deep convolutional network[EB/OL]. [2018-01-16]. http://www-o. ntust. edu. tw/ ~ cweiwang/ISBI2015/challenge2/isbi2015_Ronneberger. pdf.

[16] SUTSKEVER I,JOZEFOWICZ R,GREGOR K, et al. Towards principled unsupervised learning [EB/OL]. [2018-01-16]. https://arxiv. org/abs/1511. 06440

[17] GOODFELLOW I, BENGIO Y, COURVILLE A. Deep learning[M]. Cambridge,UK:MIT Press,2016.

[18] RATLIff L J,BURDEN S A,SASTRY S S. Characterization and computation of local Nash equilibria in continuous games[C]//Proceedings of the 51st Annual Allerton Conference on Communication, Control, and Computing. Monticello,USA:IEEE Press,2013:917-924.

[19] MIRZA M, OSINDERO S. Conditional generative adversarial nets [EB/OL]. [2018-01-16]. https://arxiv. org/abs/1411. 1784.

[20] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks [EB/OL]. [2018-01-16]. https://arxiv. org/abs/1511. 06434.

[21] KINGMA D, BA J. Adam: a method for stochastic optimization [EB/OL]. [2018-01-16]. https://arxiv. org/abs/1412. 6980v8.

[22] SALIMANS T,GOODFELLOW I,ZAREMBA W,et al. Improved techniques for training gans [EB/OL]. [2018-01-16]. https://arxiv. org/abs/1606. 03498.

[23] GULRAJANI I, AHMED F, ARJOVSKY M, et al. Improved training of wasserstein GANs [EB/OL]. [2018-01-16]. https://arxiv. org/abs/1704. 00028v3.

[24] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation [C]//Proceedings of MICCAI'15. Berlin, Germany:Springer,2015:234-241.

[25] PATHAK D, KRAHENBUHL P, DONAHUE J, et al. Context encoders:feature learning by inpainting[EB/OL]. [2018-01-16]. https://arxiv. org/abs/1604. 07379.

编辑　陆燕菲