

## 基于多尺度编码-解码网络的皮肤病变图像分割

杨国亮,洪志阳,许楠

江西理工大学电气工程与自动化学院,江西 赣州 341000

**【摘要】**针对皮肤病变图像分割在医疗诊断中的作用,提出一种基于多尺度编码-解码网络的皮肤病变图像分割算法。该算法继承了SegNet网络结构的训练速度快、训练模型存储小等特点,采用多尺度输入的方式增强了网络对皮肤病变图像的充分学习。此外,在编码网络中的pool2层输出一个二进制双线性插值的中间预测特征图到解码层的最后一层卷积块进行级联输入提高最终的分割精度。实验结果表明,采用多尺度编码-解码网络对皮肤病变图像分割具有极好的效果,在其他医学图像分割方面也能进行广泛应用。

**【关键词】**皮肤病变;多尺度编码-解码网络;SegNet;二进制双线性插值

**【中图分类号】**R318;TP391.41

**【文献标志码】**A

**【文章编号】**1005-202X(2019)02-0199-06

### Segmentation of skin lesion image based on multi-scale encoder-decoder network

YANG Guoliang, HONG Zhiyang, XU Nan

School of Electrical Engineering and Automation, Jiangxi University of Science and Technology, Ganzhou 341000, China

**Abstract:** An image segmentation algorithm based on multi-scale encoder-decoder network is proposed for the segmentation of skin lesion image in medical diagnosis. The proposed algorithm inherits the characteristics of SegNet network structure, such as fast training speed and small training model storage. And the multi-scale input method enhances the ability of network to comprehensively learn the skin lesion image. In addition, the output of a layer of binary bilinear interpolated intermediate prediction features to the final layer of convolution blocks of the decoder layer in the pool2 layer of the encoder network is cascaded to increase the final segmentation accuracy. The experimental results show that using multi-scale encoder-decoder network can achieve an excellent segmentation of skin lesion image, and that the proposed network can also be widely used in other medical image segmentations.

**Keywords:** skin lesion; multi-scale encoder-decoder network; SegNet; binary bilinear interpolation

### 前言

皮肤病变往往容易导致皮肤癌,而在众多的皮肤癌病症中,恶性黑色素瘤是致命的,占皮肤癌死亡率的75%<sup>[1]</sup>。据世界卫生组织统计,全球每年至少发生约200万到300万例非黑色素瘤皮肤癌患者和132 000例恶性黑色素瘤皮肤癌患者。恶性黑色素瘤在全球的发病率持续上升,而对于早期的恶性黑色素瘤进行切除,其治愈率高达95%<sup>[2]</sup>。但是,黑色素瘤早期阶段的检测是很

困难的,因为它们类似于常见的痣或者由于皮肤表面的毛发、皮肤的颜色、毛细血管、皮肤病变区域与未病变区域之间的对比度较低等因素而影响检测。皮肤病变的临床评估是通过眼睛检查或临床照相来确定可疑病变,然后通过皮肤镜进一步检查。因此,皮肤病变图像的计算分析在皮肤癌预防中非常重要<sup>[3]</sup>。研究人员通过开发算法和计算机辅助诊断系统(Computer Aided Diagnosis, CAD)来帮助皮肤科医生分割出皮肤病变区域用以诊断恶性黑色素瘤。

在过去的几十年中,关于皮肤病变图像分割的研究一直在持续进行。传统的图像分割算法中,皮肤病变图像分割的方法主要包括基于阈值<sup>[4-5]</sup>的区域增长<sup>[6-7]</sup>分割方法以及基于活动轮廓的方法<sup>[8-9]</sup>等。而近年来,随着深度学习的不断发展,基于深度卷积神经网络(Convolution Neural Network, CNN)<sup>[10]</sup>的分割方法首先被应用在图像分割领域。但其存在着存储

**【收稿日期】**2018-08-22

**【基金项目】**国家自然科学基金(51365017)

**【作者简介】**杨国亮,博士,教授,研究方向:模式识别与图像处理, E-mail: ygliang30@126.com;洪志阳,硕士研究生,研究方向:模式识别与图像处理, E-mail: 1293340610@qq.com;许楠,硕士研究生,研究方向:模式识别与图像处理, E-mail: 563622733@qq.com

方面开销过大、计算效率低下,丢失关键信息等问题。因此,在2015年Shelhamer等<sup>[11]</sup>提出了全卷积神经网络(Fully Convolutional Networks, FCN)用于图像分割领域,取得了较好的分割效果。此后,研究者借助FCN网络结构的思想提出了不同类型的分割网络算法并取得了令人满意的分割结果。其中,在这众多的分割网络算法中主要有Badrinarayanan等<sup>[12]</sup>提出的SegNet网络,Chen等<sup>[13]</sup>提出的DeepLab网络,Ronneberger等<sup>[14]</sup>提出的U-Net网络。

由于上述网络的提出,皮肤病变图像分割也迎来了新的发展并取得一系列的成果。本文也正是基于以上网络的设计思路,在SegNet网络结构的设计思想基础上提出一种基于多尺度编码-解码网络(Multi-Scale Encoder-Decoder Network, MSEDN)的皮肤病变分割算法。该算法是通过在SegNet网络编码-解码的结构上增加了多尺度的输入的形式,同时在编码层中的池化层2(pool2)输出一个二进制双线性插值的中间预测特征图到解码层的最后一层卷积块进行级联输入提高最终的分割精度。MSEDN网络不仅继承了SegNet网络结构训练速度快、训练模型存储小等特点,同时采用多尺度输入的方式增强网络对皮肤病变图像信息的充分学习,输出二进制双线性插值的中间预测特征图<sup>[15]</sup>到解码层的最后一层卷积块进行级联输入使得编码层提取的特征图信息能够进行充分解码输出。

## 1 SegNet网络

SegNet网络是Badrinarayanan等<sup>[12]</sup>提出来的,用以解决无人驾驶或智能机器人的图像语义分割问题的深度网络。其主要由编码网络和解码网络两部分组成,最后将解码网络的输出特征图输入到一个Softmax分类器中得到最终的分割结果。

此外,SegNet网络是一个对称网络,其编码网络结构跟VGG16网络<sup>[16]</sup>结构一样,包括卷积层(Convolution)、激活层(ReLU)、池化层(Pool)以及批量归一化层(Batch Normalization, BN)5个卷积块共13层组成。其解码结构也由5个上采样块共13层组成,包括上采样层(Upsample)、Convolution层、ReLU层和BN层。编码层与解码层之间是一一对应的关系,将图像输入到编码网络结构的第1个卷积块通过卷积激活池化操作提取图像特征输出到下一个卷积块以及相应的解码网络结构的最后一部分的上采样块进行图像的复原。同理,编码网络剩余的4个卷积块以及解码网络的剩余的4个上采样块也是一一对应的。这样通过编码网络不断提取图像的特征向下

传输并传输到与之对应的解码网络,对图像进行解码还原使得整个网络构成一个对称网络,最后通过SoftMax分类器实现图像的分割。

SegNet网络是在FCN网络结构设计思想上进一步改进得到的,FCN网络相较于VGG16网络其最大的特点是将两层全连接层改为卷积层,并在其后进行上采样层,采用跳跃融合的结构。而SegNet网络则直接将VGG16网络的两层全连接层去掉,进行上采样以及编码网络与解码网络之间信息的直接编解码。在SegNet网络中,其解码网络是整个网络最重要的部分。解码网络中的上采样(Upsample)的指数(Indices)与编码网络中的指数一样,这样可以保留分割图像中的高频信息,并大大减少训练参数,提高图像边界的描述以及加快训练速度,降低内存等。

然而,SegNet网络的编码网络结构在提取输入图像特征时,由于输入尺度的原因可能导致图像的部分关键特征信息未能提取。同时,SegNet网络的编码结构和解码结构均有5个部分,网络相对复杂易造成过拟合,在皮肤病变图像分割结果上仍存在些许不足。因此,在此基础上本文提出了一种基于MSEDN的皮肤病变图像分割算法。

## 2 多尺度编码-解码网络

MSEDN是在SegNet网络结构设计思路的基础上进行修改而得到的网络,采用SegNet网络的对称网络结构的形式。其不仅继承了SegNet网络结构的训练速度快、训练模型存储小等特点,同时考虑到SegNet网络结构中编码层结构根据VGG16网络结构进行编写的,每一层提取的输入图像的特征图是有限的,不能够较为完整的将整个输入图像的特征都提取到编码层中进行编码,存在信息提取缺失。因此,对编码层采用多尺度输入的方式,增加编码层对输入图像特征图信息的提取,并将不同尺度输入的编码层进行融合,通过最大池化层传递到相应的解码层以及下一编码层中,使得网络对图像的学习进一步加强。在编码网络的第2层卷积块pool2输出二进制双线性插值的中间预测特征图到解码层的最后一层卷积块进行级联输入使得编码层提取的特征图信息能够充分解码输出,利于最后对皮肤病变图像的分割,提高分割精度。

MSEDN网络主体上也是一种编码-解码网络,其中编码网络结构主体部分跟VGG16网络结构相似,主要是由4个卷积块组成<sup>[17]</sup>,包括Convolution层、ReLU层、BN层和Pool层。同时,借助inception-v2网络<sup>[18]</sup>结构形式使得编码网络为多尺度输入结构。而

解码层结构相应地由4个上采样块组成,包括Convolution层、ReLU层、Upsample层以及BN层,其与SegNet网络的解码网络结构一样。最后,在解

码网络后面加一个SoftMax分类器,实现整体网络对皮肤病变图像的分割,达到最佳的分割效果。MSEDN网络结构示意图如图1所示。

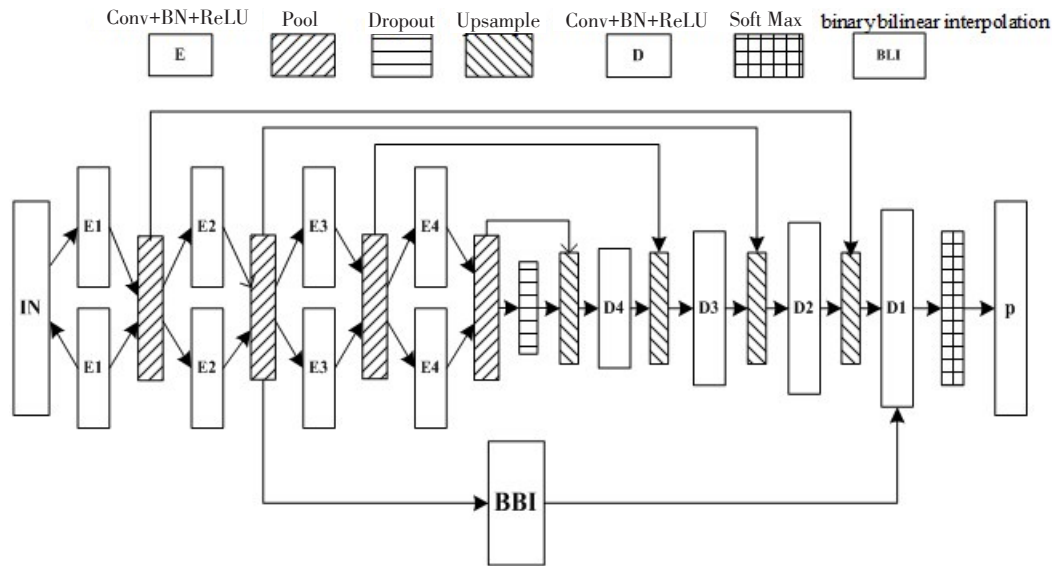


图1 MSEDN网络结构示意图

Fig.1 Schematic diagram of multi-scale encoder-decoder network (MSEDN)

如图1所示,MSEDN网络的整体架构以“退出(dropout)”层为中心将整个网络分为两部分,即编码网络和解码网络。在编码网络中每个卷积块采用多尺度输入的方式,皮肤病变图像通过不同尺度的E1层进行卷积、批量归一化、激活操作,然后输入到pool1层中。采用最大池化的方式提取关键信息传输到下一个卷积块,同时将提取的关键信息相应地输入到解码网络中的上采样块中进行解码。当信息传输到pool4层后,通过dropout层防止信息过拟合,然后传递到解码网络的Upsample4层进行上采样并继续逐层解码传输。最后,通过SoftMax分类器得到最后的皮肤病变图像的分割结果。此外,在pool2层输出一个二进制双线性插值(Binary Bilinear Interpolation, BBI)的中间预测层到D1层,实现编码网络中间高频信息进一步传输到解码网络进行图像还原,提高图像边界的描述及分割精度。

对于BBI来说,其是将两个不同变量的插值函数进行线性插值扩展,主要意思即在两个方向分别进行一次线性插值。假设图像上相邻的4个点分别为 $Q_{11}=(x_1, y_1)$ 、 $Q_{12}=(x_1, y_2)$ 、 $Q_{21}=(x_2, y_1)$ 、 $Q_{22}=(x_2, y_2)$ ,求未知函数 $f$ 在点 $p(x, y)$ 的值。因此,先在 $x$ 方向进行线性插值,得:

$$f(R_1) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{11}) + \frac{x - x_1}{x_2 - x_1} f(Q_{21}) \quad (1)$$

$$f(R_2) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{12}) + \frac{x - x_1}{x_2 - x_1} f(Q_{22}) \quad (2)$$

其中, $R_1=(x, y_1)$ 、 $R_2=(x, y_2)$ ,然后,在 $y$ 方向进行线性插值,得:

$$f(p) \approx \frac{y_2 - y}{y_2 - y_1} f(R_1) + \frac{y - y_1}{y_2 - y_1} f(R_2) \quad (3)$$

因此,综上求得 $f(x, y)$ :

$$\begin{aligned} f(x, y) \approx & \frac{f(Q_{11})}{(x_2 - x_1)(y_2 - y_1)}(x_2 - x)(y_2 - y) \\ & + \frac{f(Q_{21})}{(x_2 - x_1)(y_2 - y_1)}(x - x_1)(y_2 - y) \\ & + \frac{f(Q_{12})}{(x_2 - x_1)(y_2 - y_1)}(x_2 - x)(y - y_1) \\ & + \frac{f(Q_{22})}{(x_2 - x_1)(y_2 - y_1)}(x - x_1)(y - y_1) \end{aligned} \quad (4)$$

即:

$$\begin{aligned} p \approx & \frac{(x_2 - x)(y_2 - y)}{(x_2 - x_1)(y_2 - y_1)} Q_{11} + \frac{(x - x_1)(y_2 - y)}{(x_2 - x_1)(y_2 - y_1)} Q_{21} \\ & + \frac{(x_2 - x)(y - y_1)}{(x_2 - x_1)(y_2 - y_1)} Q_{12} + \frac{(x - x_1)(y - y_1)}{(x_2 - x_1)(y_2 - y_1)} Q_{22} \end{aligned} \quad (5)$$

### 3 实验内容

#### 3.1 数据集与实验参数设置

本文实验采用的数据集是2016年国际生物医学图像之皮肤病变分析黑色素瘤检测挑战赛(ISBI 2016 Skin Lesion Analysis Towards Melanoma Detection Challenge)的公共数据集<sup>[19]</sup>,该数据集包括900张训练图像,379张测试图像,以及相应的人工分

割的真实标签图像。由于数据集中的图像大小,分辨率等不同导致其对MSEDN网络的训练结果会产生一定的影响,因此将数据集图像进行预处理。通过对图像进行简单的尺寸变换、裁剪、翻转等操作,使得数据集符合MSEDN网络的训练要求。

本实验将数据集图像分辨率统一设置为 $512 \times 512$ ,同时将人工分割的真实标签图像进行处理制作索引图,便于整个网络的训练及测试。同时,实验是基于深度学习框架-Caffe框架<sup>[20]</sup>实现的,所有的运算都在一块内存为8 G的Nvidia GeForce GT750M的显卡上来加速网络训练。在MSEDN网络训练过程中,采用随机下降梯度法(Stochastic Gradient Descent, SGD)来更新网络权值,网络的初始学习率设置为0.001,动量(momentum)设置为0.9,权重衰减为0.000 5,最大迭代次数为4万次。在网络的测试阶段,将测试集中的图像逐张输入到训练好的MESEDN网络模型中,通过网络的前向传播与反向求导的方式计算测试集图像的分割结果,并输出每张测试图像的分割效果图。

### 3.2 评估指标

本文对皮肤病变图像的分割性能的评估指标由ISBI 2016 Lesion Analysis Towards Melanoma Detection Challenge提供,作为皮肤病变图像最终分割性能的评判标准。对于皮肤病变图像分割性能的评估指标主要包括:逐像素分割精度(Accuracy, Ac)、Dice相似度(Dic)、Jaccard指数(Jac)。其中,对于Accuracy、Dic以及定义如下:

$$Ac = \frac{N_p + N_m}{N_p + N_{fp} + N_{fn} + N_m} \quad (6)$$

$$Jac = \frac{N_p}{N_p + N_{fn} + N_{fp}} \quad (7)$$

$$Dic = \frac{2 N_p}{2N_p + N_{fn} + N_{fp}} \quad (8)$$

其中, $N_p$ 、 $N_m$ 、 $N_{fp}$ 和 $N_{fn}$ 分别表示真阳性、真阴性、假阳性和假阴性数,它们都是像素级上定义的。若病变像素的预测是病变,则认为病变像素是真阳性;否则被认为是假阴性。如果非病变像素的预测是非病变的,则认为真阴性;否则判定为假阳性。

### 3.3 实验结果与分析

通过对ISBI 2016皮肤病变图像数据集中的进行训练和测试,来验证本文所提出的MSEDN分割网络的有效性以及准确性。

**3.3.1 MSEDN网络的选择** MSEDN网络在对皮肤病变图像数据集进行训练时,其训练损失(loss)值变化图如图2所示。

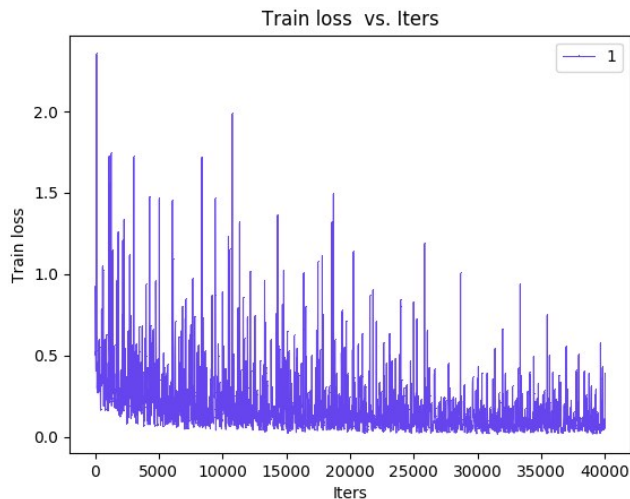


图2 MSEDN网络训练loss变化图  
Fig.2 Changes of MSEDN training loss

观察图2可以发现MSEDN网络在训练过程中,随着迭代次数不断增加,loss值逐渐趋于稳定。当网络训练迭代趋于最大迭代次数4万次,loss值基本稳定,这时所得训练模型具有一定准确性。此外,MSEDN网络是在SegNet网络结构的设计思路的基础上进行修改而得到的网络,而SegNet网络又是在FCN网络结构基础上提出的。因此,通过对比MSEDN网络与SegNet网络、FCN网络分别在皮肤病变图像的分割结果,从而验证MSEDN网络的有效性和准确性。本文采用逐像素分割精度Ac、Dic、Jac 3项评估指标来对最后的分割结果进行评估。实验是在Caffe中进行的,结果如表1所示。

表1 MSEDN网络、SegNet网络以及FCN网络分割结果  
Tab.1 Segmentation results of MSEDN network, SegNet network, and FCN network

Network	Accuracy	Dic	Jac
FCN	0.941 3	0.886 4	0.813 7
SegNet	0.936 9	0.912 5	0.841 4
MSEDN	0.973 7	0.958 4	0.920 1

FCN: Fully convolutional networks; MSEDN: Multi-scale encoder-decoder network

通过观察表可知,MSEDN网络在皮肤病变图像的各项评估指标上均明显高于FCN网络与SegNet网络,表明MSEDN网络对于皮肤病变图像分割具有显著效果,验证了MSEDN网络的有效性及准确性。

最后,根据MSEDN网络在皮肤病变图像分割的各项评估指标上取得的显著结果,得到其最终的分割效果图,如图3所示。

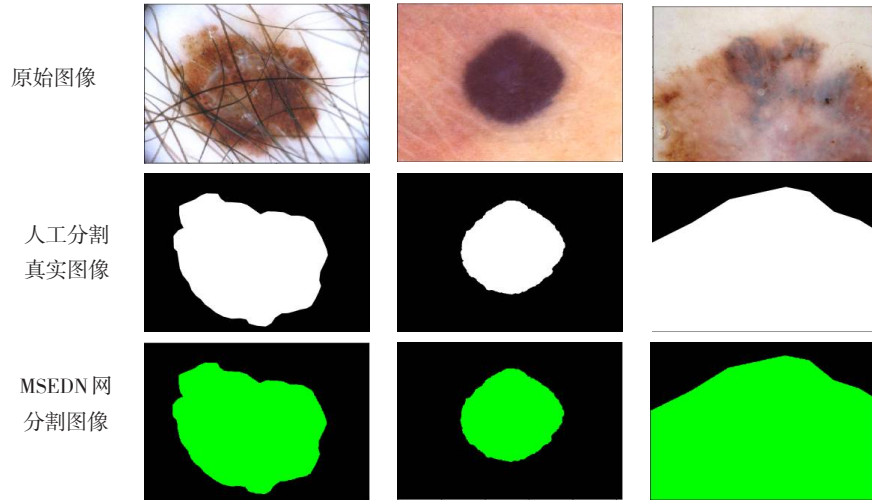


图3 皮肤病变图像分割效果图

Fig.3 Segmentation of skin lesion images

**3.3.2 对比与评估** 本文实验数据集采用是 ISBI 2016 皮肤病变图像的公共数据集,在其官方网站提供部分参赛人员在皮肤病变图像分割各项评估指标结果。选取官网中前5名皮肤病变图像分割各项评估结果与MSEDN网络的分割结果进行比较,结果如表2所示。

表2 ISBI 2016 前5名皮肤病变图像的分割结果与本文实验结果对比  
Tab.2 Comparison of segmentation results of skin lesion images in top 5 ISBI 2016 patients

Method	Accuracy	Dic	Jac
EXB	0.953	0.910	0.843
CUMED	0.949	0.897	0.829
Mahmudur	0.952	0.895	0.822
SFU-mail	0.944	0.885	0.811
TMUteam	0.946	0.888	0.810
MSEDN	0.974	0.958	0.920

MSEDN 网络各项评估指标均显著超过 ISBI 2016 前5名,无论是在逐像素分割精度、Dice 相似度,还是 Jaccard 指数均有很大提升,表明该网络在皮肤病变图像分割上具有有效性和准确性,所提算法优势显著。

## 4 结论

针对皮肤病变图像分割在医疗诊断中所起的作用,本文所提出的基于MSEDN的皮肤病变图像分割算法有效地帮助医生进行诊断。通过采用多尺度的输入方式对输入图像进行编码传输到相应的解码网

络以及下一层网络。同时,在 pool2 层输出一个二进制双线性插值的中间预测层到解码网络的 D1 层,实现编码网络中间高频信息进一步传输到解码网络进行图像还原,提高图像边界的描述及分割精度。MSEDN 网络在皮肤病变图像分割上所取得的成果,表明该分割算法对比其他的分割算法具有一定的优势,验证了该网络的有效性及其准确性。此外,算法的所有计算均是在同一块 GPU 上实现的,虽然对最终的分割结果产生些许影响,但总体达到实验的最佳要求。最后,MSEDN 网络在皮肤病变图像的成功,表明了其在其他医学图像分割存在着广泛的应用前景。

## 【参考文献】

- [1] JERANT A F, JOHNSON J T, SHERIDAN C D, et al. Early detection and treatment of skin cancer[J]. Am Fam Physician, 2000, 62(2): 357-368.
- [2] BALCH C M, GERSHENWALD J E, SOONG S J, et al. Final version of 2009 AJCC melanoma staging and classification[J]. J Clin Oncol, 2009, 27(36): 6199-6206.
- [3] OLIVEIRA R B, MARRANGHELLO N, PEREIRA A S, et al. A computational approach for detecting pigmented skin lesions in macroscopic images[J]. Expert Syst Appl, 2016, 61(C): 53-63.
- [4] GARNAVI R, ALDEEN M, CELEBI M E, et al. Border detection in dermoscopy images using hybrid thresholding on optimized color channels[J]. Comput Med Imaging Graph, 2011, 35(2): 105-115.
- [5] CAVALCANTI P G, YARI Y, SCHARCANSKI J. Pigmented skin lesion segmentation on macroscopic images[C]//Image and Vision Computing New Zealand. IEEE, 2010: 1-7.
- [6] CELEBI M E, KINGRAVI H A, IYATOMI H, et al. Border detection in dermoscopy images using statistical region merging[J]. Skin Res Technol, 2008, 14(3): 347-353.
- [7] LI C, KAO C Y, GORE J C, et al. Minimization of region-scalable fitting energy for image segmentation[J]. IEEE Trans Image Process, 2008, 17(10): 1940-1949.

- [8] ERKOL B, MOSS R H, STANLEY R J, et al. Automatic lesion boundary detection in dermoscopy images using gradient vector flow snakes[J]. *Skin Res Technol*, 2005, 11(1): 17-26.
- [9] MA Z, TAVARES J M. A novel approach to segment skin lesions in dermoscopic images based on a deformable model[J]. *IEEE J Biomed Health Inform*, 2017, 20(2): 615-623.
- [10] 廖晓磊, 赵涓涓. 基于改进自生成神经网络的肺部CT序列图像分割[J]. *计算机科学*, 2017, 44(8): 296-300.
- LIAO X L, ZHAO J J. Segmentation of lung CT image sequences based on improved self-generating neural networks[J]. *Computer Science*, 2017, 44(8): 296-300.
- [11] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation[J]. *IEEE Trans Pattern Anal Mach Intell*, 2017, 39(4): 640-651.
- [12] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: a deep convolutional encoder-decoder architecture for scene segmentation[J]. *IEEE Trans Pattern Anal Mach Intell*, 2017, 39(12): 2481-2495.
- [13] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. *IEEE Trans Pattern Anal Mach Intell*, 2018, 40(4): 834-848.
- [14] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation [C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2015: 234-241.
- [15] IQBAL S, GHANI M U, SABA T, et al. Brain tumor segmentation in multi-spectral MRI using convolutional neural networks (CNN) [J]. *Microsc Res Tech*, 2018, 81(4): 419-427.
- [16] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [Z]. arXiv preprint arXiv, 2014: 1409-1556.
- [17] ISOBE S, ARAI S. Deep convolutional encoder-decoder network with model uncertainty for semantic segmentation[C]//IEEE International Conference on Innovations in Intelligent Systems and Applications. IEEE, 2017.
- [18] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[J]. *Comput Sci*, 2015: 2818-2826.
- [19] Skin lesion analysis toward melanoma detection: a challenge at the International Symposium on Biomedical Imaging (ISBI) 2016, hosted by the International Skin Imaging Collaboration (ISIC)[EB/OL]. [https://challenge.kitware.com/#challenge/n/ISIC\\_2016%3A\\_Skin\\_Lesion\\_Analysis\\_Towards\\_Melanoma\\_Detection](https://challenge.kitware.com/#challenge/n/ISIC_2016%3A_Skin_Lesion_Analysis_Towards_Melanoma_Detection).
- [20] JIA Y Q, SHELHAMER E, DONAHUE J, et al. Caffe: convolutional architecture for fast feature embedding[C]//Proceedings of the 22nd ACM international conference on Multimedia. New York: ACM, 2014: 675-678.

(编辑:薛泽玲)