JOURNAL OF IMAGE AND GRAPHICS

主办：中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2020
09
VOL.25

ISSN1006-8961
CN11-3758/TB

遥感边缘智能技术 *P1719*

中国图象图形学报

刊名题字：宋健 | 月刊（1996年创刊）

第25卷第9期（总第293期）
2020年9月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

主管单位 中国科学院
主办单位 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

主　　编 顾行发
编辑出版 《中国图象图形学报》编辑出版委员会
通信地址 北京市海淀区北四环西路19号
邮　　编 100190
电子信箱 jig@radi.ac.cn
电　　话 010-58887035
网　　址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号
总 发 行 北京报刊发行局
订　　购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
（邮政信箱：北京399信箱　邮编：100048）
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian | Monthly, Started in 1996

Superintended by Chinese Academy of Sciences
Sponsored by Aerospace Information Research Institute, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Gu Xingfa
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@radi.ac.cn
Telephone 010-58887035
Website www.cjig.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048,P.R.China))
Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

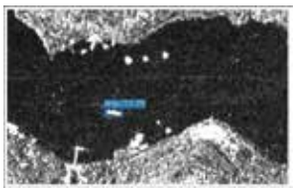
国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元

2020年9月 第25卷 第9期（总第293期）
中国图象图形学报
Zhongguo Tuxiang Tuxing Xuebao

目次

第一人称视角下的社会力优化多行人跟踪(第1869页)

分形曲线生成的频域方法(第1904页)



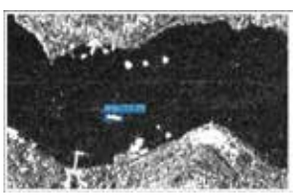
双向特征融合的数据自适应SAR图像舰船目标检测模型(第1943页)

Volume 25, Number 9
Published September 16, 2020

CONTENTS
JOURNAL OF IMAGE AND GRAPHICS

Multi-pedestrian tracking optimized by social force model under first-person perspective(P1869)

Fractal curve generation method based on the frequency domain (P1904)



Data-adaptive single-shot ship detector with a bidirectional feature fusion module for SAR images (P1943)

中图法分类号:TP391　　**文献标识码**: A　　**文章编号**: 1006-8961(2020)09-1915-15

论文引用格式: Hou X D, Zhao Y H, Liu H P, Guo H Y, Yu X X and Ding M Y. 2020. Optic disk segmentation by combining UNet and residual attention mechanism. Journal of Image and Graphics,25(09):1915-1929(侯向丹,赵一浩,刘洪普,郭鸿湧,于习欣,丁梦园. 2020. 融合残差注意力机制的 UNet 视盘分割. 中国图象图形学报,25(09):1915-1929)[DOI:10.11834/jig.190527]

融合残差注意力机制的 UNet 视盘分割

侯向丹[1,2],赵一浩[1,2],刘洪普[1,2],郭鸿湧[2],于习欣[2],丁梦园[2]

1. 河北工业大学人工智能与数据科学学院, 天津　300401; 2. 河北省大数据计算重点实验室, 天津　300401

摘　要: **目的**　青光眼和病理性近视等会对人的视力造成不可逆的损害,早期的眼科疾病诊断能够大大降低发病率。由于眼底图像的复杂性,视盘分割很容易受到血管和病变等区域的影响,导致传统方法不能精确地分割出视盘。针对这一问题,提出了一种基于深度学习的视盘分割方法 RA-UNet(residual attention UNet),提高了视盘分割精度,实现了自动、端到端的分割。**方法**　在原始 UNet 基础上进行了改进。使用融合注意力机制的 ResNet34 作为下采样层来增强图像特征提取能力,加载预训练权重,有助于解决训练样本少导致的过拟合问题。注意力机制可以引入全局上下文信息,增强有用特征并抑制无用特征响应。修改 UNet 的上采样层,降低模型参数量,帮助模型训练。对网络输出的分割图进行后处理,消除错误样本。同时,使用 DiceLoss 损失函数替代普通的交叉熵损失函数来优化网络参数。**结果**　在 4 个数据集上分别与其他方法进行比较,在 RIM-ONE(retinal image database for optic nerve evaluation)-R1 数据集中,F 分数和重叠率分别为 0.957 4 和 0.918 2,比 UNet 分别提高了 2.89% 和 5.17%;在 RIM-ONE-R3 数据集中,F 分数和重叠率分别为 0.969 和 0.939 8,比 UNet 分别提高了 1.5% 和 2.78%;在 Drishti-GS1 数据集中,F 分数和重叠率分别为 0.966 2 和 0.934 5,比 UNet 分别提高了 1.65% 和 3.04%;在 iChallenge-PM 病理性近视挑战赛数据集中,F 分数和重叠率分别为 0.942 4 和 0.891 1,分别比 UNet 提高了 3.59% 和 6.22%。同时还在 RIM-ONE-R1 和 Drishti-GS1 中进行了消融实验,验证了改进算法中各个模块均有助于提升视盘分割效果。**结论**　提出的 RA-UNet,提升了视盘分割精度,对有病变区域的图像也有良好的视盘分割性能,同时具有良好的泛化性能。

关键词:青光眼;UNet;深度学习;视盘分割;预训练;注意力机制;DiceLoss

Optic disk segmentation by combining UNet and residual attention mechanism

Hou Xiangdan[1,2], Zhao Yihao[1,2], Liu Hongpu[1,2], Guo Hongyong[2], Yu Xixin[2], Ding Mengyuan[2]

1. School of Artificial Intelligence, Hebei University of Technology, Tianjin　300401, China;

2. Hebei Provincial Key Laboratory of Big Data Computing, Tianjin　300401, China

Abstract: **Objective** Glaucoma and pathologic myopia are two important causes of irreversible damage to vision. The early detection of these diseases is crucial for subsequent treatment. The optic disk, which is the starting point of blood vessel convergence, is approximately elliptical in normal fundus images. An accurate and automatic segmentation of the optic disk from fundus images is a basic task. Doctors often diagnose eye diseases on the basis of the colored fundus images of patients. Browsing the images repeatedly to make appropriate diagnoses is a tedious and arduous task for doctors. Doctors are likely

收稿日期:2019-10-30;**修回日期**:2020-03-14;**预印本日期**:2020-03-21

基金项目:国家自然科学基金项目(U1813222);国家重点研发计划项目(2018YFB1306900)

Supported by:National Natural Science Fundation of China(U1813222);National Key Research and Development Program of China (2018YFB1306900)

to miss some subtle changes in the image when they are tired, resulting in missed diagnoses. Therefore, using computers to segment optic disks automatically can help doctors in the diagnosis of these diseases. Glaucoma, pathologic myopia, and other eye diseases can be reflected by the shape of the optic disk; thus, an accurate segmentation of the optic disk can assist doctors in diagnosis. However, achieving an accurate segmentation of optic disks is challenging due to the complexity of fundus images. Many existing methods based on deep learning are susceptible to pathologic regions. UNet has been widely used in medical image segmentation tasks; however, it performs poorly in optic disk segmentation. Convolution is the core of convolutional neural networks. The importance of information contained in different spatial locations and channels varies. Attention mechanisms have received increasing attention over the past few years. In this study, we present a new automatic optic disk segmentation network based on UNet to improve segmentation accuracy. **Method** According to the design idea of UNet, the proposed model consists of an encoder and a decoder, which can achieve end-to-end training. The ability of the encoder to extract discriminative representations directly affects the segmentation performance. Achieving pixel-wise label data is expensive, especially in the field of medical image analysis; thus, transfer learning is adopted to train the model. Given that ResNet has a strong feature extraction capability, the encoder adopts a modified and pretrained ResNet34 as the backbone to achieve hierarchical features and then integrates a squeeze-and-excitation (SE) block into appropriate positions to enhance the performance further. The final average pooling layer and the fully connected layer of ResNet34 are removed, but the rest are kept. The SE block can boost feature discriminability, which includes SE operations. The SE block can model the relationship between different feature map channels to recalibrate channel-wise feature responses adaptively. In the encoder, all modules, except for four SE blocks, use the pretrained weights on ImageNet (ImageNet Large-Scale Visual Recognition Challenge) as initialization, thereby speeding up convergence and preventing overfitting. The input images are downsampled for a total of five times to extract abstract semantic features. In the decoder, 2×2 deconvolution with stride 2 is used for upsampling. Five upsampling operations are conducted. In contrast to the original UNet decoder, each deconvolution, except for the last one, outputs a feature map of 128 channels, thus reducing model parameters. The shallow feature map preserves more detailed spatial information, whereas the deep feature map has more high-level semantic information. A set of downsampling layers enlarges the receptive field of the network but causes a loss of detailed location information. The skip connection between the encoder and decoder can combine high-level semantic information with low-level detailed information for fine-grained segmentation. The feature map in the encoder first goes through a 1×1 convolution layer, and then the output of 1×1 convolution is concatenated with the corresponding feature map in the decoder. Using skip connection is crucial in restoring image details in the decoder layers. Lastly, the network outputs a two-channel probability map for the background and the optic disk; this map has the same size as the input image. The network utilizes the last deconvolution with two output channels, followed by SoftMax activation, to generate the final probability map of the background and the optic disk simultaneously. The segmentation map predicted by the network is rough; thus, postprocessing is used to reduce false positives. In addition, DiceLoss is used to replace the traditional cross entropy loss function. Considering that the training images are limited, we first perform data augmentation, including random horizontal, vertical, and diagonal flips, to prevent overfitting. An NVidia GeForce GTX 1080Ti device is used to accelerate network training. We adopt Adam optimization with an initial learning rate of 0.001. **Result** To verify the effectiveness of our method, we conduct experiments on four public datasets, namely, RIM-ONE (retinal image database for optic nerve evaluation) -R1, ONE-R1, RIM-ONE-R3, Drishti-GS1, and iChallenge-PM. Two evaluation metrics, namely, F score and overlap rate, are computed. We also provide some segmentation results to compare different methods visually. The extensive experiments demonstrate that our method outperforms several other deep learning-based methods, such as UNet, DRIU, DeepDisc, and CE-Net, on four public datasets. In addition, the visual segmentation results produced by our method are more similar to the ground truth label. Compared with the UNet results in RIM-ONE-R1, RIM-ONE-R3, Drishti-GS1, and iChallenge-PM, the F score (higher is better) increases by 2.89%, 1.5%, 1.65%, and 3.59%, and the overlap rate (higher is better) increases by 5.17%, 2.78%, 3.04%, and 6.22%, respectively. Compared with the DRIU results in RIM-ONE-R1, RIM-ONE-R3, Drishti-GS1, and iChallenge-PM, the F score (higher is better) increases by 1.89%, 1.85%, 1.14%, and 2.01%, and the overlap rate (higher is better) increases by 3.41%, 3.42%, 2.1%, and 3.53%, respec tively. Compared with the DeepDisc results in RIM-ONE-R1, RIM-ONE-R3, Drishti-GS1, and iChallenge-PM, the

F score (higher is better) increases by 0.24%, 0.01%, 0.18%, and 1.44%, and the overlap rate (higher is better) increases by 0.42%, 0.01%, 0.33%, and 2.55%, respectively. Compared with the CE-Net results in RIM-ONE-R1, RIM-ONE-R3, Drishti-GS1, and iChallenge-PM, the F score (higher is better) increases by 0.42%, 0.2%, 0.43%, and 1.07%, and the overlap rate (higher is better) increases by 0.77%, 0.36%, 0.79%, and 1.89% respectively. We also conduct ablation experiments on RIM-ONE-R1 and Drishti-GS1. Results demonstrate the effectiveness of each part of our algorithm. **Conclusion** In this study, we propose a new end-to-end convolutional network model based on UNet and apply it to the optic disk segmentation problem in practical medical image analysis. The extensive experiments prove that our method outperforms other state-of-the-art deep learning-based optic disk segmentation approaches and has excellent generalization performance. In our future work, we intend to introduce some recent loss functions, focusing on the segmentation of the optic disk boundary.
Key words: Glaucoma; UNet; deep learning; optic disc segmentation; pre-trained; attention mechanism; DiceLoss

0 引　言

青光眼会对患者造成不可逆的视力损害(Mary 等,2016)。青光眼的发生率逐渐增高,但人们却未对其给予足够的重视。视盘,全称是视神经盘,也叫视神经乳头,在正常眼底图像中,视盘通常近似椭圆形,是血管汇聚的起点,血管从起点向四周扩散。眼科医生常常根据病人的彩色眼底图像进行眼科疾病诊断,视盘分割往往是后续疾病诊断中很重要的一步。视杯在视盘内部区域,视杯盘比是诊断青光眼的重要依据,视杯盘比越大,患青光眼的概率也越大。因此对视杯的准确分割往往需要先将视盘区域提取出来。

随着医疗影像技术的发展,医院每天都会产生大量的医学影像,重复浏览图像并做出相应判断对于医生来说是一项烦琐并且艰巨的任务。更重要的是,对医学影像做出诊断是一个相对主观的判断过程,会受到医生的经验和疲劳程度的影响。医生在疲惫的时候很有可能遗漏图像中某些细微变化之处,从而导致漏诊、误诊等情况发生。因此,利用计算机自动、高效地进行视盘分割,有助于眼科医生进行眼底疾病的诊断。

用于视盘分割的方法主要有传统方法和基于深度学习的方法。传统方法主要基于边缘检测、模板匹配和形变模型等。例如 Aquino 等人(2010)提出采用形态学边缘检测方法分割视盘边界,吴鑫鑫和肖志勇(2018)采用圆形或椭圆形霍夫变换的方法,Lowell 等人(2004)采用局部可形变模型,Cheng 等人(2013)提出基于超像素点分类的方法等。眼底图像比较复杂,视盘区域分割往往会受到血管等其他区域的影响。传统方法需要人工提取图像特征,当对比度不强或者有病变区域影响时,分割效果不好。因此,提取具有判别性的图像特征和实现自动、方便的模型优化方法是至关重要的。

随着深度学习技术的发展,以数据驱动、自动提取相关特征的卷积神经网络(convolutional neural network,CNN)在自然图像分割任务上取得了比传统方法更好的分割效果,例如 FCN(fully convolutional networks)(Long 等,2015)等。因此把深度学习相关技术引入到医学影像处理中,利用卷积神经网络来进行视盘分割的研究越来越多,并取得了优于传统分割方法的分割结果。M-Net(multi-label deep network)(Fu 等,2018)用来进行视盘和视杯的联合分割,在 UNet 基础上,增加了多尺度输入,引入了深度监督思想,在中间层添加额外的损失函数。此外,还引入了极坐标展开操作。CE-Net(context encoder network)(Gu 等,2019)提出了一个上下文编码模块,由一个多尺度的密集空洞卷积模块和一个残差多路径池化模块构成,可以多角度捕获具有高水平语义信息的特征。Sevastopolsky(2017)提出了一个修改后的 UNet 分割框架,在每个卷积层后使用了 Dropout 操作。和 FCN 相似,DRIU(deep retinal image understanding)(Maninis 等,2016)舍弃了 VGG(visual geometry group)网络(Simonyan 和 Zisserman,2015)中的全连接层,设计了两个特殊分支,一个分支进行视盘分割,另一个分支进行血管分割。Al-Bander 等人(2018)将全卷积 DenseNet 应用于视盘分割任务,密集连接(Huang 等,2017)提高了特征复用,减少了网络参数,有利于网络优化。Shankaranarayana 等人(2017)将残差连接思想(He 等,2016)和 UNet(Ronneberger 等,2015)结合,提出了 Res-UNet,并且结合生成对抗网络思想完成对视盘和视杯的联合分割。

随着迁移学习的发展,把预训练好的模型在小

样本数据集上进行微调,解决了样本少导致的网络过拟合问题。DeepDisc(Gu 等,2018)中提出基于空洞卷积(Yu 和 Koltun,2016)和空间金字塔池化模块(Zhao 等,2017)的视盘分割网络,空洞卷积可以增大感受野,空间金字塔池化模块可以多角度地聚合上下文信息,使用预训练的 ResNet34 作为特征提取网络。Yu 等人(2019)提出一个用预训练的 ResNet-34 作为编码层的 U 型结构,用于分割视盘和视杯。

虽然基于卷积神经网络的视盘分割方法在一定程度上取得了优于传统方法的效果,但是仍然存在一些问题,例如:由于眼底图像复杂,视盘区域很容易受其他病变区域影响,导致编码器提取特征能力不够,无法提取出有效的特征;用于训练神经网络的样本少,很容易导致过拟合问题;视盘分割方法往往对某一数据集分割效果好,但是对于新的数据集分割结果差,泛化性能差。为了解决上述问题,本文主要有以下几点贡献:

1)基于 UNet 设计并实现了一个可以实现端到端训练的编码器—解码器网络结构 RA-UNet(residual attention UNet),使用融合注意力机制的 ResNet34 作为特征提取网络,具有更强的特征提取能力。

2)精简 UNet 的上采样层,减少网络优化参数;浅层特征图先经过 1×1 卷积,再与深层特征图拼接,可以促进信息融合。

3)使用 DiceLoss 损失函数代替交叉熵损失函数,有助于提升分割精度。

4)在 RIM-ONE(retinal image database for optic nerve evaluation)-R1(Fumero 等,2011),RIM-ONE-R3(Fumero 等,2011),Drishti-GS1(Sivaswamy 等,2015)和 iChallenge-PM(iChallenge-pathological myopia)(Fu 等,2019)这 4 个数据集上验证了 RA-UNet 的有效性和泛化性。

1 相关技术

1.1 残差网络

卷积神经网络在图像分类、分割和目标检测等计算机视觉领域取得了巨大成功。在图像分类任务中,CNN 一般由卷积层、池化层、激活函数和全连接层组成。通过堆叠卷积层来增加网络深度,往往可以提高网络提取图像特征的能力。但是,随着卷积神经网络深度的逐渐加深,网络会变得越来越难以训练,进而出现网络性能退化问题。ResNet(He 等,2016)通过引入跳跃连接结构进一步加深了网络深度,解决了梯度消失问题,提升了网络性能。

图 1 是构成 ResNet18 和 ResNet34 的基本残差单元,由恒等连接路径和残差路径组成。残差路径由两个 3×3 卷积层、批标准化(batch normalization)和 ReLU(rectified linear units)激活函数构成,然后把两条路径的结果相加即为输出。同时,跳跃连接没有引入额外的参数量和计算复杂度。

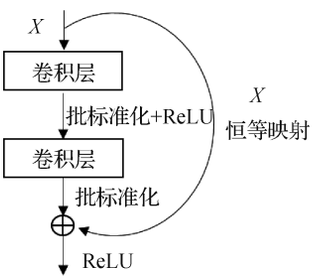


图 1 残差连接单元

Fig. 1 Residual block

1.2 UNet

Ronneberger 等人(2015)提出 UNet 结构,并将其应用到医学图像细胞分割中,在小样本数据集上取得了优异的性能。如图 2 所示,该网络分为编码器(encoder)和解码器(decoder)两部分,二者是对称结构。编码器对应的是图像下采样过程,解码器对应的是特征图上采样过程,并且相应的编码器和解码器之间存在着跳跃连接,这些跳跃连接可以帮助上采样层恢复图像的细节信息。编码器通过典型的卷积神经网络结构来提取图像的特征信息,由 3×3 卷积层、ReLU 函数和 2×2 最大池化层构成,共进行了 4 次下采样,每次进行池化操作后,特征图尺寸下降,同时通道数会翻倍。解码器通过 2×2 反卷积层(或转置卷积)来进行上采样,逐渐恢复图像信息。和编码器部分相对应,解码器部分共进行了 4 次上采样,每次上采样都会扩大特征图尺寸,同时通道数减半。将编码器和解码器相对应的特征图进行拼接(concatenation),能够用浅层网络保存较好的细节位置信息来辅助分割。UNet 网络一共包含 23 个卷积层。

1.3 迁移学习

深度学习往往需要大规模数据来进行网络模型

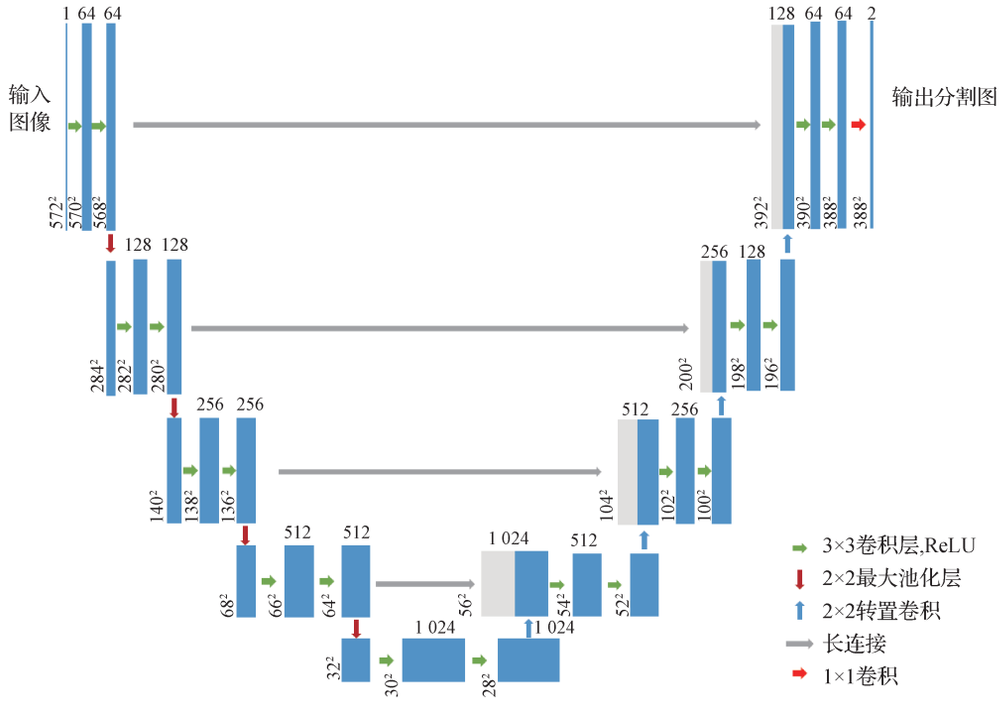


图 2　UNet 网络结构

Fig. 2　The architecture of original UNet

的训练和优化，但是在医学图像分割领域，有逐像素标注的图像数量往往很少。在小规模数据集上训练网络模型往往会出现过拟合问题。同时，让医生进行像素标注是耗时和乏味的，且容易受医生的主观影响。当医生累的时候，也常常会有错误的判断。迁移学习可以解决数据量少导致的深度卷积神经网络模型不好训练的问题。首先在大规模的图像数据集上（例如 ImageNet）进行网络训练，然后在小规模数据集上微调该模型。可以大大降低训练模型所需的时间，并且取得更好的结果。

2　改进的网络模型

分割彩色眼底图像视盘区域过程如图 3 所示，输入图像为 RGB 3 通道的彩色图像，先将其输入网络模型中，输出分割好的视盘区域图像。其中，视盘区域为白色，背景区域为黑色，实现了自动、端到端的图像分割。然后经过后处理操作，就得到了更精细化的视盘分割图。

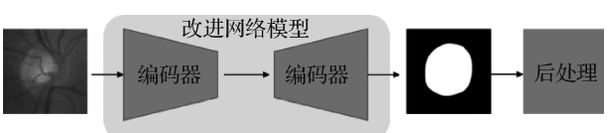


图 3　视盘分割流程图

Fig. 3　The pipeline of proposed method

2.1　注意力机制

注意力机制在图像分类、语义分割和自然语言处理等领域受到广泛关注。卷积层是卷积神经网络的核心，卷积操作在局部感受野上无差别地融合了特征图的空间和通道信息，但是不同空间位置、不同通道所包含信息的重要程度不一样。SE（squeeze and excitation）模块（Hu 等，2018）能够显示建模通道之间的关系，增强重要特征，抑制无用的特征。如图 4 所示，SE 模块分为 Squeeze 和 Excitation 两个操作。输入为 $\boldsymbol{X}$，其维度为 $\mathbf{R}^{H\times W\times C}$，$H$、$W$ 和 C 分别表示特征图的高、宽和通道数。首先经过 Squeeze 操作，沿着空间维度（$\mathbf{R}^{H\times W}$）通过全局平均池化方式来聚合全局信息，生成一个 $\mathbf{R}^{1\times 1\times C}$ 维度的通道描述符。为了充分利用 Squeeze 聚合的全局信息，再经过 Excitation 操作，来捕获特征图各通道之间的相互依赖关系。该操作先经过一个全连接层，把通道数由 c 缩小为 c/r，然后经过 ReLU 函数，参数 r 可以控制 SE 模块的计算量，r 的大小设置不同，对网络性能影响效果也会不同（r 的大小选择将在实验部分具体讨论）。然后再经过一个全连接层，通道数由 c/r 扩大为 c，紧接着是一个 Sigmoid 函数。至此，

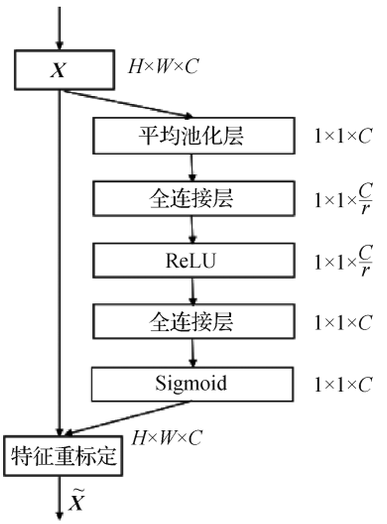


图 4　SE 模块

Fig. 4　The SE block

生成了一个 $\mathbf{R}^{1\times1\times C}$ 维度的向量，然后将其和输入 $\boldsymbol{X}$ 进行逐通道相乘，就完成了对输入特征图的通道重标定。如图 5 所示，经过特征重标定后的特征图不同通道重要性不同，重要的信息被放大，不重要的信息被减弱。

SE 模块引入的参数量主要取决于两个全连接层。先只考虑加入一个 SE 模块的情况，若输入特征图具有 C 个通道，假设 r 为 2，则引入的新的参数量为 C^2；假设 r 为 8，则引入的新的参数量为 $C^2/4$。SE 模块简单高效、轻量级，引入的参数量可以忽略不计。SE 模块的有效性主要有两点：1）通过全局平均池化操作，可以在提取图像特征时引入全局上下文信息；2）通过两个全连接层，建立了特征图跨通道之间的联系。

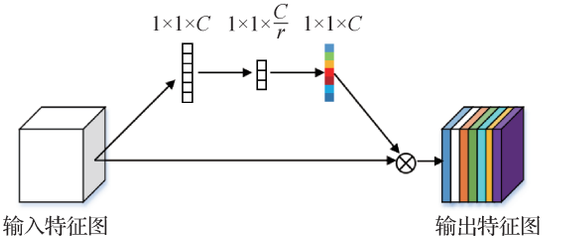


图 5　特征图通道重标定

Fig. 5　Recalibrate the feature map

2.2　RA-UNet 网络模型

按照 UNet 的设计思想，RA-UNet 主要分为编码器和解码器两部分（如图 6 所示），能够实现端到端的训练。

编码器提取有代表性的图像特征的能力，对分割的最终性能影响很大。传统的 UNet 编码器，输入图像依次经过两个相连的带 ReLU 函数的 3×3 卷积层和 2×2 最大池化层，一共进行了 4 次下采样。

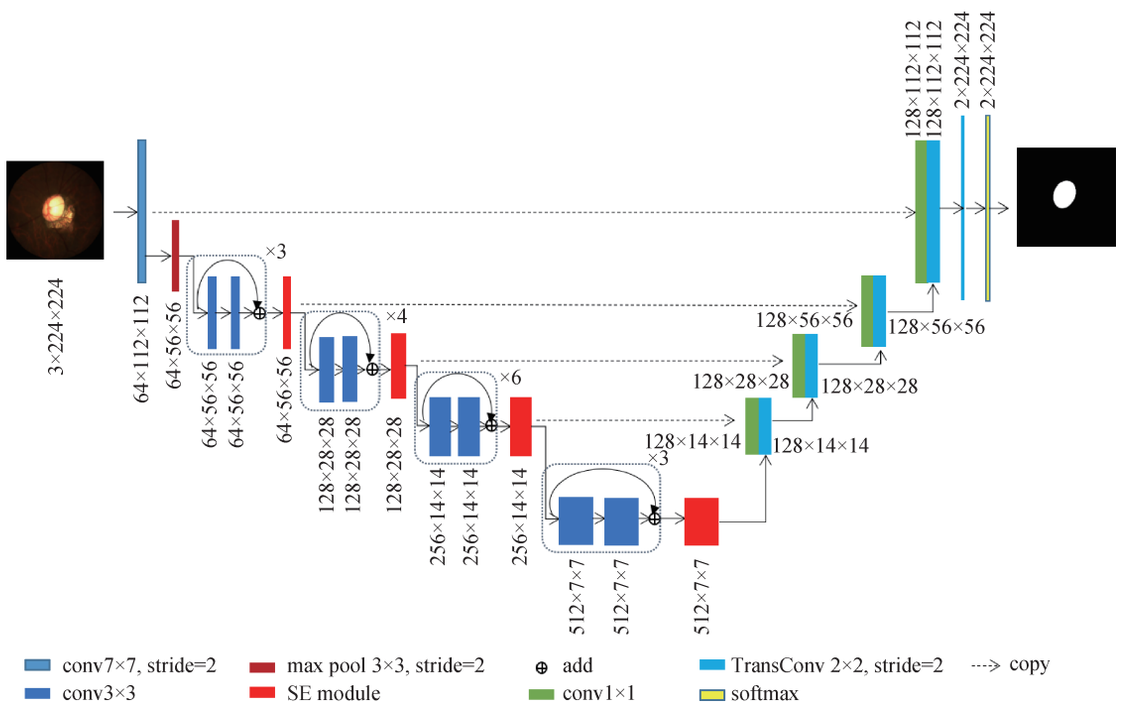


图 6　RA-UNet

Fig. 6　RA-UNet

因为 ResNet 提取特征能力更强，所以采取修改的 ResNet34 作为编码器，并且融合了 SE 模块，来提取图像的分层特征，增强特征提取能力。

图 7(a)移除了 ResNet34 最后的平均池化层和全连接层，保留了第 1 个 7×7 卷积层和最大池化层，以及个数分别为 3，4，6，3 的残差模块。如图 1 所示，残差模块由两个 3×3 卷积层、批标准化、ReLU 激活函数和恒等连接构成。其中，批标准化操作可以将卷积输出变为以 0 为均值、1 为方差的标准正态分布，有助于提升网络训练速度和解决梯度消失问题。图 7(b)在除了第 1 个 7×7 卷积层以外的其余几个残差模块后添加了 SE 模块。为了解决数据少导致的过拟合问题，编码器中除了 4 个 SE 模块，其他部分都是使用在 ImageNet 数据集上训练好的权重，加快了网络收敛速度，防止了过拟合。SE 模块采用均值为 0、方差为 1 的高斯分布权重初始化方式。SE 模块中的参数 r 设置为 8。共进行了 5 次下采样，输入图像尺寸大小为 224×224，最后被下采样为 7×7 像素大小。原始的 UNet 网络中 3×3 卷积层没有使用填充 0(padding)策略，这使得每次卷积输出尺寸都会减小。所以，提出的网络中每个卷积层都采用了 padding 策略，其中 7×7 卷积填充为 3，3×3 卷积填充为 1，这使得卷积前后特征图尺寸大小一致。

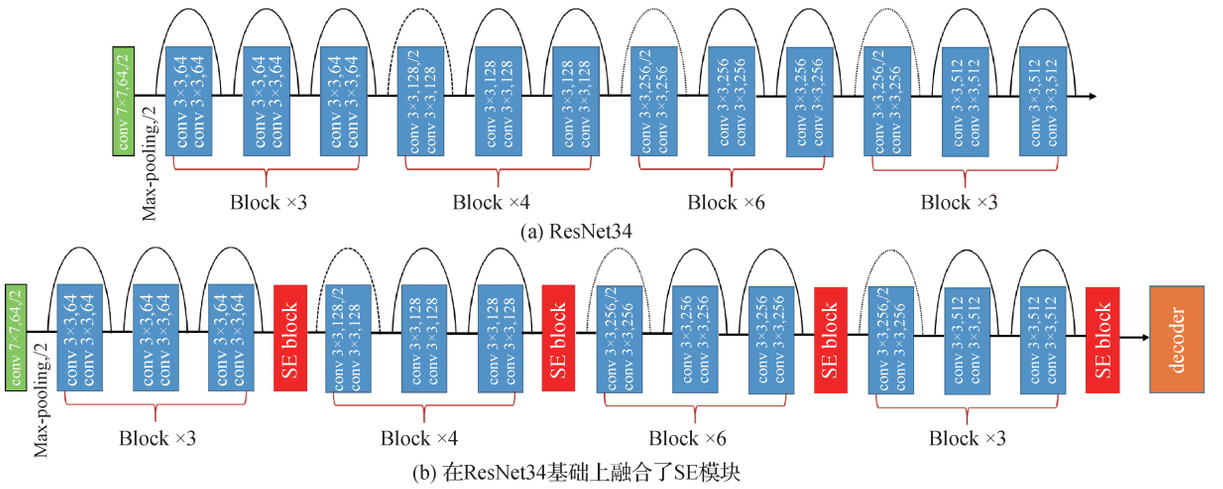


图 7　编码器结构

Fig. 7　Architecture of the encoder((a) ResNet34; (b) the combination of ResNet34 and SE block)

编码器提取了具有高水平语义信息的特征，解码器需要恢复原图信息。和 UNet 中的上采样方式一样，采用的是步长为 2 的 2×2 转置卷积，每次上采样，特征图的长和宽都会扩大为原来的 2 倍，共进行了 5 次上采样操作，最后恢复到了原图尺寸 224×224 像素大小。和 UNet 解码器不同之处是，每次上采样后输出的特征图通道数都是 128(不包括最后一个反卷积层)，并且去掉了原始 UNet 解码器中的 3×3 卷积层，减少了计算量和参数量，简化了模型。浅层网络保存较好的位置细节信息，而深层网络具有较高水平的语义特征信息。编码器中的下采样层造成了图像细节位置信息的丢失，这使得解码器在上采样时很难恢复。所以，在相应的编码器和解码器之间，仍旧存在着跳跃连接，可以用浅层网络的细节位置信息帮助解码器更好地恢复图像信息。如图 8 所示，在跳跃连接过程中，编码器中的特征图先经过 1×1 卷积层(没有带 ReLU 函数和批标准化层)，输出 128 通道的特征图。1×1 卷积层的作用主要有两点：1)降低编码器中特征图的通道数，方便与解码器中具有相同通道数的特征图进行拼接(concatenation)；2)编、解码器特征图包含的信息差异太大，先经过 1×1 卷积层，缩小二者之间的差异，更能有效地融合二者信息，然后再与解码器中相应的特征图进行拼接。特征图拼接之后依次经过批标准化层(Ioffe 和 Szegedy，2015)和 ReLU 函数，再进行 2×2 转置卷积。

最后一个 2×2 转置卷积输出一个 2 通道的特征图，然后经过 softmax 激活后，输出 2 通道的和原图尺寸大小相同的概率图。语义分割本质上属于逐像素点的分类问题，有视盘区域和背景区域两个种

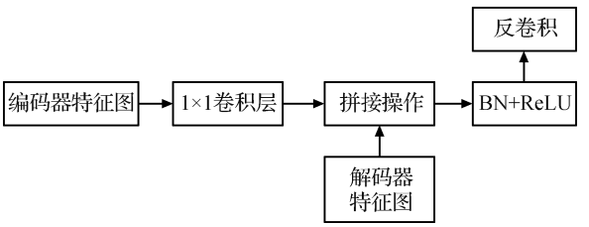


图 8　浅层和深层信息融合

Fig. 8　Combination of shallow and deep information

类,所以输出 2 通道的概率图。概率图中每个像素位置都有对应于每个种类(视盘或背景)的概率值。通过取出每个像素位置处最大概率值的索引,就可以输出模型的分割效果图。

2.3　损失函数

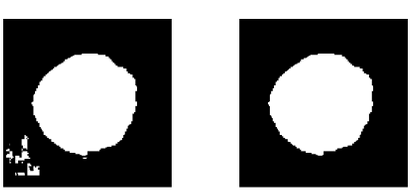
语义分割问题仍旧是像素点的分类问题,最常用的损失函数就是交叉熵损失函数。医学图像往往存在正负样本不平衡问题,使用交叉熵损失函数不能很好解决这一问题。本文使用 DiceLoss 函数(Milletari 等,2016)来替代传统的交叉熵损失函数,即

$$L_{\text{Dice}} = 1 - \sum_{k=1}^{K} \frac{2w_k \sum_{i=1}^{N} p_{(k,i)} g_{(k,i)}}{\sum_{i=1}^{N} p_{(k,i)} + \sum_{i=1}^{N} g_{(k,i)}} \tag{1}$$

式中,N 表示像素点个数。K 表示种类个数,设置为 2,包括视盘区域和背景两个种类。$p_{(k,i)} \in [0,1]$,表示像素点预测为种类的概率,也就是网络最后的 softmax 层输出的概率值。$g_{(k,i)} \in \{0,1\}$,表示像素点 i 属于种类 k 的标签值。本文设置 $w_k = \frac{1}{K}$。

2.4　后处理操作

直接从网络输出的分割图可能会有一些边缘噪声点或一些视盘区域不连续现象,故需要对其进行一些后处理操作,来精细化分割结果,如图 9(a)所示。通过 8 连通域来找出分割图中的所有连通域;计算各连通域的面积大小;最后只保留面积最大的连通域,删除其余的连通域。如果得到的最大连通域有孔洞,就进行孔洞填充。经过后处理后的视盘分割图如图 9(b)所示。



(a) 原始输出　(b) 后处理

图 9　后处理操作

Fig. 9　Post processing

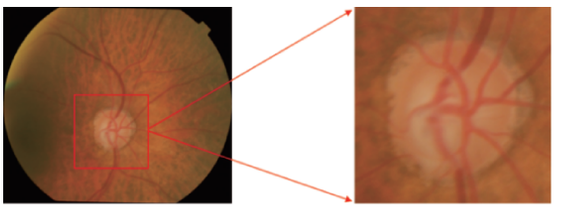
((a) raw output; (b) the output of post processing)

3　实验结果与分析

3.1　数据集

在 4 个公开的彩色眼底视盘分割数据集上分别进行了对比实验,来验证网络模型的性能以及泛化能力,4 个数据集分别为 Drishti-GS1, RIM-ONE-R1, RIM-ONE-R3 和 iChallenge-PM。

Drishti-GS1 数据集(Sivaswamy 等,2015)包含 101 幅图像(31 幅正常眼底图像和 70 幅青光眼图像),官方已经将其分为训练集 50 幅和测试集 51 幅,故无需再人工划分训练集和测试集。每幅图像都由 4 位专家进行像素标注,把 4 位专家的平均标注结果作为金标准。该数据集提供的是整幅眼底图像,因为视盘区域只占一小部分,为了防止血管、黄斑等无关区域对视盘分割结果的影响,采用 Wang 等人(2019)方法中使用的视盘区域提取方法先找出视盘中心,然后以其为中心从整图中裁剪出大小为 512×512 像素的视盘区域(如图 10 所示)。



(a) 原始图像　(b) 裁剪的视盘区域

图 10　裁剪感兴趣区域

Fig. 10　Crop the region of optic disc

((a) raw image; (b) the region of interest)

RIM-ONE 数据集(Fumero 等,2011)有 3 个发行版本,分别包含 169、455 和 159 幅彩色眼底图像。在第 1 个发行版本 RIM-ONE-R1 和第 3 个发行版本 RIM-ONE-R3 上均进行了对比实验。RIM-ONE-R1 数据集同时提供了 5 个专家的逐像素标注结果,把第 1 个专家的标注结果作为金标准,用来进行训练和测试。该数据集提供的是裁剪后的视盘区域图像,感兴趣区域已经从整幅眼底图像中裁剪出来。把 169 幅图像分为训练集 105 幅和测试集 64 幅,分别用于训练模型和测试模型。

RIM-ONE-R3 数据集包含 159 幅彩色眼底图像，同时提供了两个专家的标注结果，把两个专家的平均标注结果作为金标准。按 Yu 等人(2019)方法中的划分训练集、测试集方法，按照 8∶2 的比例，随机选 127 幅图像作为训练集，剩余的 32 幅图像作为测试集。由于提供的不是裁减感兴趣区域后的图像，故先用 Wang 等人(2019)方法中使用的视盘区域提取方法，从原图中裁剪出 512×512 像素大小的视盘区域。

iChallenge-PM 挑战赛数据集(Fu 等，2019)有 1 200 幅彩色眼底图像，分为训练集 400 幅、验证集 400 幅和测试集 400 幅。先由中山大学中山眼科中心的 7 位眼科医生分别进行逐像素标注，然后再由另外一位高级专家合并为单一标注，最后存储为 BMP(Bitmap)格式图像。由于只能获得训练集的图像标注信息，所以只能在提供的 400 幅训练集图像中进行对比试验。训练集中有 19 幅图像不包含视盘区域，所以将这 19 幅图像删除，还剩 381 幅带标注信息的图像。从这 381 幅彩色眼底图像中划分出 305 幅图像用于训练，76 幅图像用于测试。同样，由于提供的是整图，先用 Wang 等人(2019)方法中使用的视盘区域提取方法，从原图中裁剪出 800×800 像素大小的视盘区域。

3.2 实验细节

实验使用的深度学习框架是 PyTorch 0.4.0，计算机操作系统为 Ubuntu 16.04，同时使用了 GPU(graphic processing unit)来加速网络模型的训练和测试，显卡型号为 GeForce GTX 1080Ti。采用 Adam 优化器，因为 Adam 算法可以在训练时自适应地调节学习率，且有更快的收敛速度，初始学习率设置为 0.001。训练阶段 batch size 设置为 6，测试阶段 batchsize 设置为 1，共训练了 150 轮。

由于视盘分割数据集数量较少，为了防止过拟合，在训练阶段对每幅图像分别进行 3 种数据增强操作，包括随机水平翻转、随机垂直翻转和随机对角翻转。通过数据增强，每幅训练图像可以变为 8 幅。此外，考虑到计算资源有限，故先把图像统一缩放为 224×224 像素，再送入网络模型中进行训练和测试。

3.3 评价指标

和曹新容等人(2018)提出的视盘分割评价指标一样，使用 F 分数(用 F_{score} 表示)和重叠率(用 S 表示)两个评价指标，计算为

$$P = \frac{TP}{TP + FP} \tag{2}$$

$$R = \frac{TP}{TP + FN} \tag{3}$$

$$F_{score} = \frac{2 \times P \times R}{P + R} \tag{4}$$

式中，P 表示精确率，R 表示召回率，F_{score} 是二者的调和均值，介于 0 与 1 之间，越接近于 1，表示结果越好。TP 表示网络输出的视盘区域，实际上也是视盘区域；TN 表示网络输出的背景区域，实际上也是背景区域；FP 表示网络输出的视盘区域，但实际上是背景区域；FN 表示网络输出的背景区域，但实际上是视盘区域。

$$S = \frac{Area(A \cap B)}{Area(A \cup B)} \tag{5}$$

式中，A 表示专家标注的视盘区域，B 表示网络模型输出的视盘区域，$Area$ 函数表示面积。重叠率 S 介于 0 到 1 之间，越接近于 1，表示视盘重叠面积越大，分割结果越好。

3.4 实验结果与分析

3.4.1 网络模型各模块消融实验

实验主要分为训练和测试两个阶段，首先在训练集上进行训练，然后在测试集中进行模型测试，得到分割结果。由于实验中存在一些超参数和可变项，故先进行对比实验，证明实验中的各项设置均为最优。主要包括以下 4 点：1)SE 模块中的超参数 r 大小对性能影响的结果比较；2)SE 模块在分割网络中的添加位置对性能影响的结果比较；3)不同损失函数对性能影响的结果比较；4)是否加载预训练权重对性能影响的结果比较。该部分对比实验在 RIM-ONE-R1 和 Drishti-GS1 两个数据集上进行。

表 1 是 SE 模块中超参数 r 大小对性能影响的结果比较。如表 1 所示，分别设置 r 大小为 2、8、12 和 16，当 r 为 8 时，在两个数据集上均取得最优结果。故在后续其他实验中均将 r 设置为 8。

如图 11 所示，SE 模块可以加入到编解码网络结构的不同位置，分别为：

1)P0：不添加 SE 模块；

2)P1：在编码器后(不包括第 1 个 7×7 卷积层后)；

3) P2：在编码器后(包括第 1 个 7×7 卷积层后)；

表 1　SE 模块中参数 r 大小对性能影响

Table 1　Comparison results of different values of r in SE block

参数 r	RIM-ONE-R1		Drishti-GS1	
	F 分数	重叠率	F 分数	重叠率
2	0.954 0	0.912 0	0.962 8	0.928 2
8	**0.957 4**	**0.918 2**	**0.966 2**	**0.934 5**
12	0.949 8	0.904 4	0.963 2	0.929 1
16	0.948 0	0.901 2	0.962 1	0.927 0

注：加粗字体为最优值。

4）P3：在解码器后；

5）P4：在编码器和解码器后（不包括第 1 个 7×7 卷积层）；

6）P5：在编码器最后一层后。

表 2 是 SE 模块的不同位置对性能影响的比较结果。如表 2 所示，当将 SE 模块放在 P1 位置时，对模型性能提升最大。尤其是在 RIM-ONE-R1 数据集中，当在 P1 位置加入 SE 模块后，重叠率 S 比不加 SE 模块时提高了 0.93%。同时也观察到，若在第 1 个 7×7 卷积层后加入SE模块，性能会下降，这

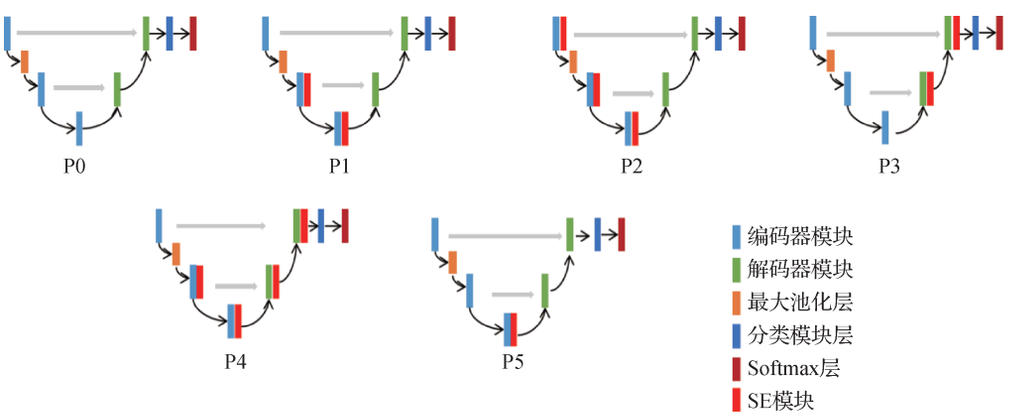


图 11　SE 模块位置

Fig. 11　Position of the SE block

表 2　SE 模块不同位置对性能影响

Table 2　Impact of different positions of SE block on performance

SE 模块位置	RIM-ONE-R1		Drishti-GS1	
	F 分数	重叠率	F 分数	重叠率
P0	0.952 3	0.908 9	0.965 1	0.932 6
P1	**0.957 4**	**0.918 2**	**0.966 2**	**0.934 5**
P2	0.951 2	0.906 9	0.961 2	0.925 3
P3	0.953 7	0.911 4	0.963 0	0.928 7
P4	0.952 8	0.909 9	0.965 6	0.933 6
P5	0.954 4	0.912 7	0.964 0	0.930 6

注：加粗字体为最优值。

可能是因为浅层特征图距离输入层太近，感受野太小，包含的信息语义水平太低。在其余实验中均将 SE 模块放在 P1 位置。

表 3 是不同的损失函数对结果影响的比较。通过对比可知，使用 DiceLoss 的结果要优于传统的交叉熵损失函数。其中，在 RIM-ONE-R1 数据集中，使

表 3　不同损失函数对性能影响

Table 3　Impact of different loss functions on performance

损失函数	RIM-ONE-R1		Drishti-GS1	
	F 分数	重叠率	F 分数	重叠率
交叉熵损失函数	0.949 4	0.903 7	0.964 9	0.932 2
DiceLoss	**0.957 4**	**0.918 2**	**0.966 2**	**0.934 5**

注：加粗字体为最优值。

用 DiceLoss 的 F 分数和平均重叠率 S 比使用交叉熵损失函数分别提高了 0.8% 和 1.45%。

表 4 为是否加载预训练权重对结果影响的比较。通过对比可知，加载预训练权重提升了分割效果。在 Drishti-GS1 数据集中，加载预训练权重的 F 分数和平均重叠率 S 比不加载预训练权重时分别提高了 1.42% 和 2.6%。

3.4.2　与其他视盘分割方法对比试验

为了验证 RA-UNet 在视盘分割任务上的有效性和泛化性，在4个数据集上分别和4个基准模型进行

表 4　加载预训练权重对性能影响

Table 4　Comparison results of pre-trained or not

加载预训练权重	RIM-ONE-R1		Drishti-GS1	
	F 分数	重叠率	F 分数	重叠率
否	0.941 6	0.889 7	0.952	0.908 5
是	**0.957 4**	**0.918 2**	**0.966 2**	**0.934 5**

注：加粗字体为最优值。

对比实验，为了公平对比，4 个深度学习的基准模型，分别为 UNet（Ronneberger 等，2015）、DRIU（deep retinal image understanding）（Maninis 等，2016）、DeepDisc（Gu 等，2018）和 CE-Net（Gu 等，2019）。同时，还列出了其他方法在相应文献中的实验结果。表 5 是在 Drishti-GS1 的测试集中与其他方法的对比结果。由表 5 可知，RA-UNet 的平均 F 分数和平均重叠率 S 为 0.966 2 和 0.934 5，分别比 UNet 提高了 1.65% 和 3.04%，RA-UNet 的重叠率 S 是最优的。

表 5　不同方法 Drishti-GS1 数据集的对比结果

Table 5　Comparison results of different methods on Drishti-GS1 dataset

方法	F 分数	重叠率
UNet	0.949 7	0.904 1
DRIU	0.954 8	0.913 5
DeepDisc	0.964 4	0.931 2
CE-Net	0.961 9	0.926 6
曹新容等人（2018）	0.934 0	0.885 0
Al-Bander 等人（2018）	0.949 0	0.904 2
Son 等人（2018）	**0.967 4**	—
Edupuganti 等人（2018）	0.967 0	—
RA-UNet（本文）	0.966 2	**0.934 5**

注：加粗字体为最优值，“—”表示文献未给出结果。

表 6 是在 RIM-ONE-R3 的测试集中与其他方法的对比结果。如表 6 所示，RA-UNet 的平均 F 分数和平均重叠率 S 为 0.969 和 0.939 8，分别比 UNet 提高了 1.5% 和 2.78%。同时，在与其他文献列出的结果比较中，两个指标都达到了最优。

表 7 是在 RIM-ONE-R1 和 iChallenge-PM 的测试集中与其他方法的对比结果。在 RIM-ONE-R1 数据集和 iChallenge-PM 挑战赛数据集中，由于相应文献较少，故只与 4 个基准模型作对比。如表 7 所示，

表 6　不同方法 RIM-ONE-R3 数据集的对比结果

Table 6　Comparison results of different methods on RIM-ONE-R3 dataset

方法	F 分数	重叠率
UNet	0.954 0	0.912 0
DRIU	0.950 5	0.905 6
DeepDisc	0.968 9	0.939 7
CE-Net	0.967 0	0.936 2
Son 等人（2018）	0.954 6	—
Zilly 等人（2017）	0.942 0	0.890 0
Al-Bander 等人（2018）	0.903 6	0.828 9
Sevastopolsky（2017）	0.950 0	0.890 0
Yu 等人（2019）	0.961 0	0.925 6
Wang 等人（2019）	0.968 0	—
RA-UNet（本文）	**0.969 0**	**0.939 8**

注：加粗字体为最优值，“—”表示文献未给出结果。

表 7　不同方法 RIM-ONE-R1 和 iChallenge-PM 数据集的对比结果

Table 7　Comparison results of different methods on RIM-ONE-R1 and iChallenge-PM dataset

方法	RIM-ONE-R1		iChallenge-PM	
	F 分数	重叠率	F 分数	重叠率
UNet	0.928 5	0.866 5	0.906 5	0.828 9
DRIU	0.938 5	0.884 1	0.922 3	0.855 8
DeepDisc	0.955 0	0.914 0	0.928 0	0.865 6
CE-Net	0.953 2	0.910 5	0.931 7	0.872 2
RA-UNet（本文）	**0.957 4**	**0.918 2**	**0.942 4**	**0.891 1**

注：加粗字体为最优值。

在 RIM-ONE-R1 数据集中，RA-UNet 的平均 F 分数和平均重叠率 S 为 0.957 4 和 0.918 2；在 iChallenge-PM 数据集中，RA-UNet 的平均 F 分数和平均重叠率 S 为 0.942 4 和 0.891 1。因为 iChallenge-PM 数据集的图像有病变区域影响，故会对模型有所影响。但是，RA-UNet 仍能取得最优分割结果。RA-UNet 在 4 个数据集上的优异性能表现，表明了该方法有效且泛化性能良好。

表 8 是 5 种网络的参数量和计算量对比结果。如表 8 所示，RA-UNet 的参数量和 GFLOPs（Giga floating-point operations per second）分别为 22.09 M 和 6.31。RA-UNet 在性能提升的同时，参数量和计

表 8　不同方法的参数量和计算量对比结果

Table 8　Comparison of different methods in terms of parameters and FLOPs

网络模型	参数量/M	GFLOPs
UNet	31.04	45.87
DRIU	15.07	24.81
DeepDisc	9.96	3.53
CE-Net	29.00	6.50
RA-UNet(本文)	22.09	6.31

算量也没有大大增加,主要原因有:1)残差连接的引入,在缓解梯度消失问题的同时,也大大降低了网络模型的计算量;2)上采样阶段的每个反卷积(除了最后一个)输出特征图通道数为 128,相对应的编解码结构中的特征图拼接之后,无须像原始 UNet 解码器中那样再经过 3×3 卷积层,这精简了网络结构,减少了参数量。

如图 12 和图 13 所示,从 4 个数据集中各选出了 4 幅图像,可视化 5 种模型的分割结果。UNet 的分割结果容易受眼底图像中血管的影响;DRIU方法分割出的视盘边界比较粗糙;DeepDisc 有错分现象。由于病理性近视彩色眼底图像在视盘区域周围往往有大范围的病变区域,这会影响视盘区域的分割。UNet、DRIU、CE-Net 和 DeepDisc 方法容易受

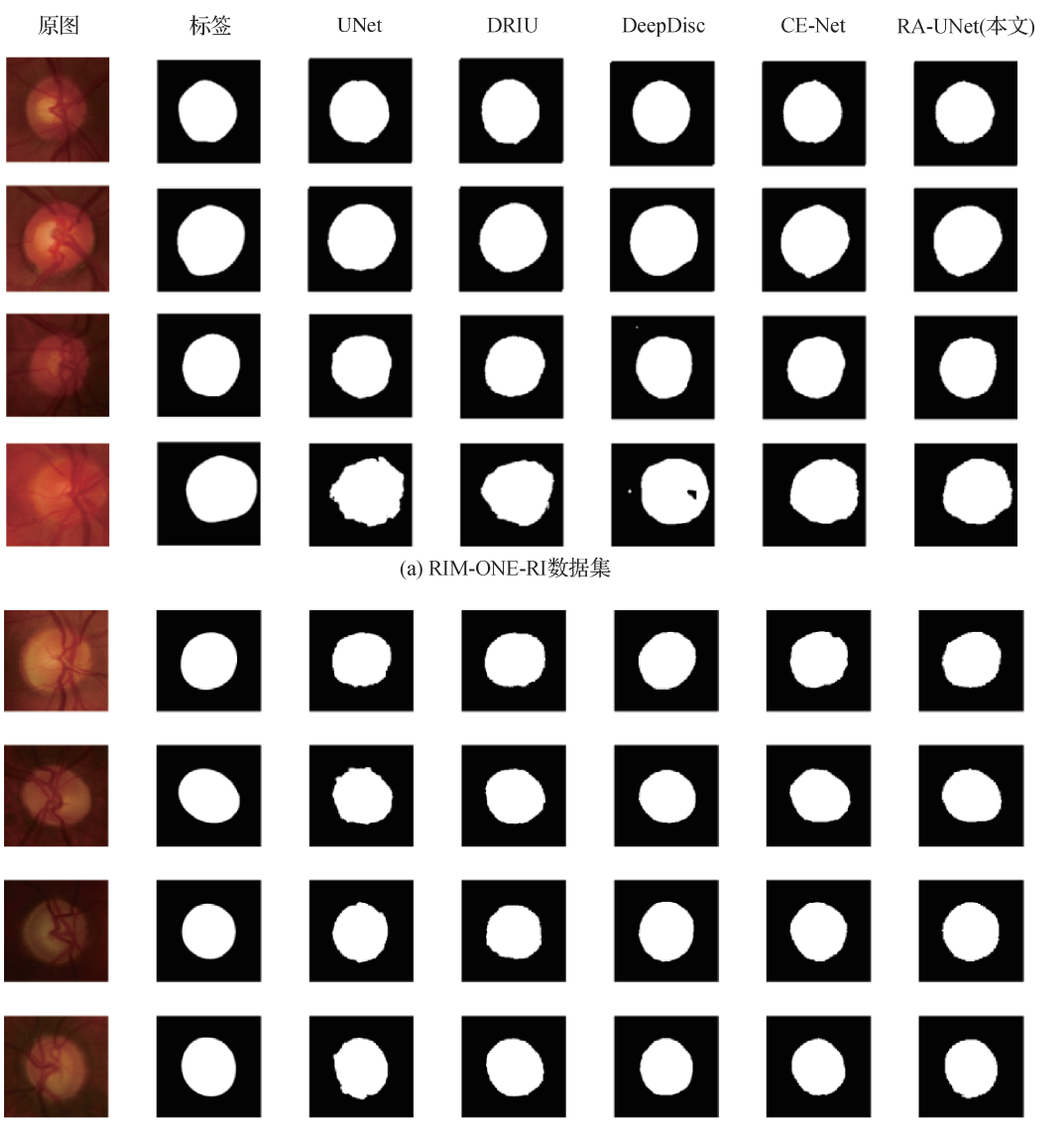


图 12　RIM-ONE-R1 和 RIM-ONE-R3 数据集分割图

Fig. 12　Visual segmentation examples of different models on RIM-ONE-R1 and RIM-ONE-R3 datasets((a) RIM-ONE-R1; (b) RIM-ONE-R3)

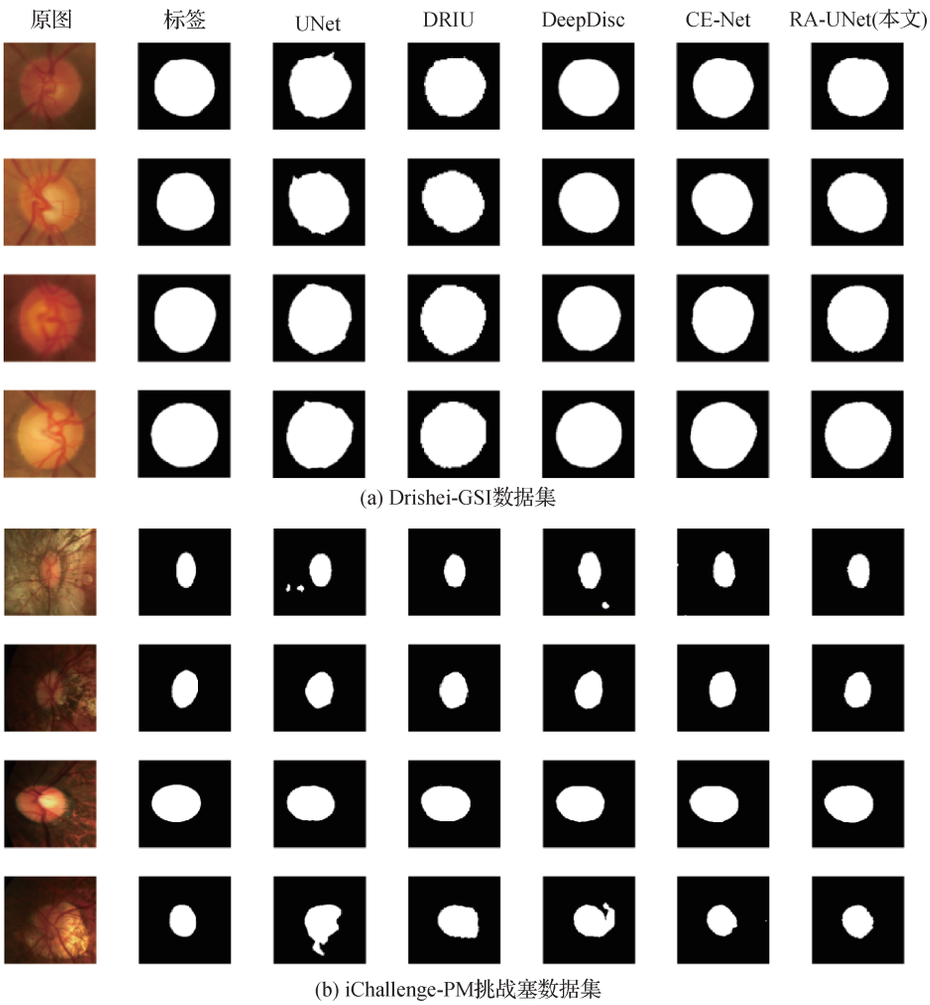


图 13　Drishti-GS1 和 iChallenge-PM 数据集分割图

Fig. 13　Visual segmentation examples of different models in Drishti-GS1 and iChallenge-PM datasets ((a) Drishti-GS1; (b) iChallenge-PM)

病变区域影响，导致不能很好地分割出视盘边界。相对来说，RA-UNet 能够更好地从病变区域中识别出视盘区域，分割结果也更接近于专家实际标注结果。

4　结　论

彩色眼底图像中视盘的形状可以反映青光眼和病理性近视等眼科疾病，自动精确地分割出视盘区域可以辅助医生进行疾病诊断。但是，血管和病变区域等会影响视盘区域的分割。本文提出一种基于 UNet 的眼底图像视盘分割网络 RA-UNet，实现端到端的自动分割。RA-UNet 采用融合注意力机制的预训练 ResNet34 作为编码器，能够提取出更加有效的图像特征；精简的解码器在减少参数的同时，通过跳跃连接进一步细化了分割效果；使用 DiceLoss 损失函数提升了分割精度。在 4 个数据集上做了对比实验，均取得了不错的结果，验证了本文方法的有效性和泛化性。对于有病变区域影响的图像，RA-UNet 也能有效地分割出视盘。但是，也注意到分割出的视盘边界有的不是很平滑。所以，在接下来的工作中，可以引入一些最新的损失函数，重点关注视盘边界的分割。考虑到视盘的形状近似椭圆，故也可以加入这一先验知识来约束分割。同时，也可以使用生成对抗网络来进行视盘分割。

参考文献(References)

Al-Bander B, Williams B M, Al-Nuaimy W, Al-Taee M A, Pratt H and Zheng Y L. 2018. Dense fully convolutional segmentation of the

optic disc and cup in colour fundus for glaucoma diagnosis. Symmetry, 10(4): #87 [DOI: 10.3390/sym10040087]

Aquino A, Emilio M, Gegúndez-Arias M E and Marin D. 2010. Detecting the optic disc boundary in digital fundus images using morphological, edge detection, and feature extraction techniques. IEEE Transactions on Medical Imaging, 29(11): 1860-1869 [DOI: 10.1109/TMI.2010.2053042]

Cao X R, Xue L Y, Lin J W and Yu L. 2018. A novel method of optic disk segmentation based on visual saliency and rotary scanning. Journal of Biomedical Engineering, 35(2): 229-236 (曹新容, 薛岚燕, 林嘉雯, 余轮. 2018. 基于视觉显著性和旋转扫描的视盘分割新方法. 生物医学工程学杂志, 35(2): 229-236) [DOI: 10.7507/1001-5515.201706013]

Cheng J, Liu J, Xu Y W, Yin F S, Wong D W K, Tan N M, Tao D C, Cheng C Y, Aung T and Wong T Y. 2013. Superpixel classification based optic disc and optic cup segmentation for glaucoma screening. IEEE Transactions on Medical Imaging, 32(6): 1019-1032 [DOI: 10.1109/TMI.2013.2247770]

Edupuganti V G, Chawla A and Kale A. 2018. Automatic optic disk and cup segmentation of fundus images using deep learning//Proceedings of the 25th IEEE International Conference on Image Processing. Athens: IEEE: 2227-2231 [DOI: 10.1109/ICIP.2018.8451753]

Fu H Z, Cheng J, Xu Y W, Wong D W K, Liu J and Cao X C. 2018. Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. IEEE Transactions on Medical Imaging, 37(7): 1597-1605 [DOI: 10.1109/TMI.2018.2791488]

Fu H Z, Li F, Orlando J L, Bogunović H, Sun X, Liao J G, Xu Y W, Zhang S C and Zhang X L. 2019. PALM: pathologic myopia challenge[EB/OL]. [2019-10-09]. http://dx.doi.org/10.21227/55pk-8z03

Fumero F, Alayon S, Sanchez J L, Sigut J and Gonzalez-Hernandez M. 2011. Rim-One: an open retinal image database for optic nerve evaluation//Proceedings of the 24th International Symposium on Computer-Based Medical Systems. Bristol: IEEE: 1-6 [DOI: 10.1109/CBMS.2011.5999143]

Gu Z W, Cheng J, Fu H Z, Zhou K, Hao H Y, Zhao Y T, Zhang T Y, Gao S H and Liu J. 2019. CE-Net: context encoder network for 2D medical image segmentation. IEEE Transactions on Medical Imaging, 38(10): 2281-2292 [DOI: 10.1109/TMI.2019.2903562]

Gu Z W, Liu P, Zhou K, Jiang Y M, Mao H Y, Cheng J and Liu J. 2018. DeepDisc: optic disc segmentation based on atrous convolution and spatial pyramid pooling//Proceedings of the 1st International Workshop on Computational Pathology and Ophthalmic Medical Image Analysis. Granada: Springer: 253-260 [DOI: 10.1007/978-3-030-00949-6_30]

He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: 770-778 [DOI: 10.1109/CVPR.2016.90]

Hu J, Shen L and Sun G. 2018. Squeeze-and-excitation networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE: 7132-7141 [DOI: 10.1109/CVPR.2018.00745]

Huang G, Liu Z, van der Maaten L and Weinberger K Q. 2017. Densely connected convolutional networks//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE: 2261-2269 [DOI: 10.1109/CVPR.2017.243]

Ioffe S and Szegedy C. 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift[EB/OL]. 2015-03-02[2019-10-09]. https://arxiv.org/pdf/1502.03167.pdf

Long J, Shelhamer E and Darrell T. 2015. Fully convolutional networks for semantic segmentation//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE: 3431-3440 [DOI: 10.1109/CVPR.2015.7298965]

Lowell J, Hunter A, Steel D, Basu A, Ryder R, Fletcher E and Kennedy L. 2004. Optic nerve head segmentation. IEEE Transactions on Medical Imaging, 23(2): 256-264 [DOI: 10.1109/TMI.2003.823261]

Maninis K K, Pont-Tuset J, Arbeláez P and Van Gool L. 2016. Deep retinal image understanding//Proceedings of the 19th International Conference on Medical Image Computing and Computer-Assisted Intervention. Athens: Springer: 140-148 [DOI: 10.1007/978-3-319-46723-8_17]

Mary M C V S, Rajsingh E B and Naik G R. 2016. Retinal fundus image analysis for diagnosis of glaucoma: a comprehensive survey. IEEE Access, 4: 4327-4354 [DOI: 10.1109/ACCESS.2016.2596761]

Milletari F, Navab N and Ahmadi S A. 2016. V-Net: fully convolutional neural networks for volumetric medical image segmentation//Proceedings of the 4th International Conference on 3D Vision (3DV). Stanford: IEEE: 565-571 [DOI: 10.1109/3DV.2016.79]

Ronneberger O, Fischer P and Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation//Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]

Sevastopolsky A. 2017. Optic disc and cup segmentation methods for glaucoma detection with modification of U-Net convolutional neural network. Pattern Recognition and Image Analysis, 27(3): 618-624 [DOI: 10.1134/S1054661817030269]

Shankaranarayana S M, Ram K, Mitra K and Sivaprakasam M. 2017. Joint optic disc and cup segmentation using fully convolutional and adversarial networks//Proceedings of the International Workshop, FIFI 2017, and the 4th International Workshop, OMIA 2017, Held in Conjunction with MICCAI 2017: Fetal, Infant and Ophthalmic Medical Image Analysis. Québec City: Springer: 168-176 [DOI: 10.1007/978-3-319-67561-9_19]

Simonyan K and Zisserman A. 2015. Very deep convolutional networks for large-scale image recognition[EB/OL]. 2015-04-10[2019-10-09]. https://arxiv.org/pdf/1409.1556.pdf

Sivaswamy J, Krishnadas S R, Chakravarty A, Joshi G D, Ujjwal and Syed T A. 2015. A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis. JSM Biomedical Imaging Data Papers, 2(1): #1004

Son J, Park S J and Jung K H. 2018. Towards accurate segmentation of retinal vessels and the optic disc in fundoscopic images with generative adversarial networks. Journal of digital imaging, 32(3): 499-512 [DOI: 10.1007/s10278-018-0126-3]

Wang S J, Yu, L Q, Yang X, Fu C W and Heng P A. 2019. Patch-based output space adversarial learning for joint optic disc and cup segmentation. IEEE Transactions on Medical Imaging, 38(11): 2485-2495 [DOI: 10.1109/TMI.2019.2899910]

Wu X X and Xiao Z Y. 2018. Automatic algorithm for fast parting optical fundus disc based on multi-circle. Optical Technique, 44(5): 586-591 (吴鑫鑫, 肖志勇. 2018. 基于多圆快速分割眼底视盘的自动算法. 光学技术, 44(5): 586-591) [DOI: 10.13741/j.cnki.11-1879/o4.2018.05.012]

Yu F and Koltun V. 2016. Multi-Scale context aggregation by dilated convolutions[EB/OL]. 2016-04-30[2019-10-09]. https://arxiv.org/pdf/1511.07122.pdf

Yu S, Xiao D, Frost S and Kanagasingam Y. 2019. Robust optic disc and cup segmentation with deep learning for glaucoma detection. Computerized Medical Imaging and Graphics, 74: 61-71 [DOI: 10.1016/j.compmedimag.2019.02.005]

Zhao H S, Shi J P, Qi X J, Wang X G and Jia J Y. 2017. Pyramid scene parsing network//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE: 6230-6239 [DOI: 10.1109/CVPR.2017.660]

Zilly J, Buhmann J M and Mahapatra D. 2017. Glaucoma detection using entropy sampling and ensemble learning for automatic optic cup and disc segmentation. Computerized Medical Imaging and Graphics, 55: 28-41 [DOI: 10.1016/j.compmedimag.2016.07.012]

作者简介

侯向丹，1976 年生，女，副教授，主要研究方向为数字图像处理，智能算法。
E-mail: 2837158@qq.com

刘洪普，通信作者，男，讲师，主要研究方向为数字图像处理，智能算法，机器学习。
E-mail: liuii@scse.hebut.edu.cn

赵一浩，男，硕士研究生，主要研究方向为基于深度学习的医学图像处理。E-mail: 1152048717@qq.com

郭鸿湧，男，正高级工程师，主要研究方向为智能信息处理。
E-mail: guohongyong@163.com

于习欣，男，硕士研究生，主要研究方向为深度学习，3D 点云。E-mail: 1367456185@qq.com

丁梦园，女，硕士研究生，主要研究方向为深度学习。
E-mail: 1061540619@qq.com