计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2022,31(2):342−349 [doi: 10.15888/j.cnki.csa.008286]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于可变形卷积和语义嵌入式注意力机制的眼球超声图像分割方法①

盛克峰[1,2], 李文星[1]

[1](贵州大学 计算机科学与技术学院, 贵阳 550025)
[2](贵州大学 密码学与数据安全研究所, 贵阳 550025)
通信作者: 盛克峰, E-mail: shengkeifeng@163.com

摘　要: 眼球区域分割是医学超声图像处理和分析的关键步骤, 由于临床设备采集的眼球超声图像具有噪声干扰、区域模糊、边缘灰度相似等缺点, 从而导致现有的方法不能准确地分割出眼球区域, 因此本文基于可变形卷积提出了一种语义嵌入的注意力机制的分割方法. 首先使用可变形卷积替代传统的卷积, 提高本文网络对眼球区域的表征能力; 其次构建语义嵌入的注意力机制, 融合不同层之间的语义信息, 增强目标区域中的显著特征, 减少对背景区域的错误分割, 从而提升网络的分割准确度; 最后, 为了验证本文模型的分割性能, 分别与现有的 3 种深度学习分割模型进行对比, 在超声眼球图像分割数据集上, 本文方法获得了最高的准确度; 充分验证了本文的模型有较好的分割能力和鲁棒性.

关键词: 深度学习; 语义分割; 特征融合; 注意力机制

引用格式: 盛克峰,李文星.基于可变形卷积和语义嵌入式注意力机制的眼球超声图像分割方法.计算机系统应用,2022,31(2):342–349. http://www.c-s-a.org.cn/1003-3254/8286.html

Eyeball Ultrasound Image Segmentation Based on Deformable Convolution and Semantic Embedded Attention Mechanism

SHENG Ke-Feng[1,2], LI Wen-Xing[1]

[1](College of Computer Science and Technology, Guizhou University, Guiyang 550025, China)
[2](Institute of Cryptography and Data Secuity, Guizhou University, Guiyang 550025, China)

Abstract: The segmentation of eyeball areas is a key step in medical ultrasound image processing and analysis. Since the eyeball ultrasound images collected by clinical equipment have disadvantages including noise interference, blurred areas, and similar edge gray levels, the existing methods cannot accurately segment eyeball areas. Therefore, this study proposes a semantic embedded attention mechanism for eyeball segmentation based on deformable convolutions. Firstly, deformable convolutions, instead of traditional convolutions, are used to improve the representational ability of the network in eyeball areas. Secondly, a semantic embedded attention mechanism is constructed to fuse semantic information among different layers, enhance the salient features in the target area, and reduce the wrong segmentation of the background area, thereby improving the segmentation accuracy of the network. Finally, in order to check the segmentation performance, the proposed model in this study is compared with three existing deep learning segmentation models, and it obtains the highest accuracy on the segmentation data set of ultrasound eyeball images, fully verifying that this model has better segmentation ability and robustness.

Key words: deep learning; semantic segmentation; feature fusion; attention mechanism

① 基金项目: 贵州省科技重大专项 (20183001)
收稿时间: 2021-04-06; 修改时间: 2021-04-29; 采用时间: 2021-05-11; csa 在线出版时间: 2022-01-17

眼球超声图像可以为临床提供丰富的眼球信息辅助医生诊断, 分割眼球区域是分析医学图像非常重要的手段, 其分割的效果会影响后续分析. 一般情况下, 超声图像中眼球区域分割需要临床医生进行手动分割和标注, 消耗大量的人力和资源. 除此之外, 临床超声设备容易受噪声影响, 采集到的图像容易不可避免的具有区域模糊、边缘灰度相似等缺点, 传统基于阈值的分割方法和基于轮廓的分割方法并不能准确地将眼球区域分割出来.

越来越多的专家引入深度学习的方法处理医学图像. 将卷积神经网络应用到图像语义分割当中, 可以高效地从大量的样本中学习丰富的图像特征信息, 显著提高分割的速度和精度. 利用卷积神经网络和深度学习的分割方法在医学图像分割中取得不俗的表现, 其分割精度接近于临床专家手动标注. 因此, 在 2014 年 Long 等[1] 提出基于全卷积神经网络 (fully convolutional networks for semantic segmentation, FCN) 的图像语义分割, 首次将深度学习应用到语义分割当中; 该模型是一个编码-解码架构的神经网络, 允许任意尺寸大小的图像输入, 降低了图像处理的难度. FCN 使用卷积层代替传统卷积神经网络中的全连接层, 又提出了通过跳跃连接将包含语义信息的高层特征和包含位置信息的低层特征进行融合以达到较精确的分割效果. 2015 年 Ronneberger 等[2] 基于 FCN 提出了图像分割的 U 形神经网络 (U-net architecture, Unet) 通过跳跃连接, 融合高低层的特征信息以增强解码器恢复局部细节的能力, 尤其在生物医学图像数据集上的分割中取得了不俗的效果. 但在多次卷积和下采样时, 会造成空间位置信息和边缘轮廓像素的丢失, 而原始 Unet 全卷积网络使用跳跃连接的方式不能充分利用低层语义信息, 只能粗略地对图像进行语义分割. Zhang 等[3] 为了将更多的语义信息引入低级特征, 提出了语义监督和语义嵌入分支, 同时为了将更多的空间信息嵌入到高层特征中提出了通道分辨率嵌入和密集相邻预测. Lin 等[4] 提出了语义嵌入分支的 Unet, 用于分割边缘模糊图像. Gu 等[5] 基于 Unet 提出了适用于 2D 医学图像分割的上下文编码器网络, 其目的就是进一步提取高层信息, 尽可能保留更多的空间信息. 虽然以上基于深度学习的分割方法在语义分割中取得不俗的表现, 但在针对眼球超声图像的分割存在分割精度不高的问题.

为进一步有效地提取关键信息特征, 2014 年 Mnih 等[6,7] 在图像分类中引入了注意力机制, 用于关注输入图像的最相关区域, 提高网络的学习能力. Jaderberg 等[8] 提出了空间变换网络, 使卷积神经网络具有空间变换的能力, 让网络能够学习到图像的平移、尺度变换、旋转的不变性. Hu 等[9] 提出了一个新的架构单元 (sequeze and excitation block, SE block) 通过对图像特征通道间的相互依赖关系进行操作, 通过学习的方式来自动获取到每个特征通道的重要程度, 去提升关键特征并抑制无用的特征. Vaswani 等[10] 提出了利用注意力机制将编码器-解码器连接起来, 摒弃了递归的网络结构, 节省大量的训练时间. Wang 等[11]提出了残差注意力网络, 在不同层次的特征上进行学习, 极大地减少计算量又达到了较高的准确度. Oktay等[12] 在 2018 年提出了基于 Unet 的注意力机制 (attention Unet), 利用注意力机制在输入图像中抑制不相关区域的同时突出目标的显著特征. Alom 等[13] 在 Unet 的基础上将残差网络和 RCNN (regions with convolutional neural networks) 结合在一起, 使用残差模块可以进行深层网络的训练, 同时在不过多增加参数量的情况下提高分割能力, 在循环残差层将特征相加有利于特征提取. 2017 年 Dai 等[14] 提出了可变形卷积层替换传统的卷积层, 可变形卷积对形状的几何信息进行建模, 能够有效地学习不同形状的目标. 2019 年 Zhu 等[15] 在网络中引入更多的可变形卷积层, 增强网络的学习能力, 通过可变形卷积模块的调制机制, 减小无关的图像区域对特征的影响. 2017 年, Zhang 等[16]提出了用于细胞分割与分类的可变形卷积的 Unet 分割网络, 利用可变形卷积解决了尺寸、形状存在巨大差异的细胞之间难以分割的问题. 2019 年 Deng 等[17]提出了一种约束的可变形卷积语义分割算法, 该算法在输入特征图上利用可变形卷积有效地对目标的几何形状进行学习. 2019 年 Sun 等[18] 提出了一种用于胃癌区域的分割方法, 利用可变形卷积和 Atrous 空间金字塔池化模块进行多尺度的语义分割.

2019 年 Takikawa 等[19] 提出了用于语义分割的门控形状的卷积神经网络 Gated-SCNN, 将形状信息作为单独分支即形状流, 用门控连接双流 CNN 架构, 高效地去除噪声且专注地处理边界相关信息. 2020 年 Niu 等[20] 提出了混合多重注意力网络 HMANet, 从通道和空间的注意力本身出发, 自适应地捕获全局信息, 通过嵌入通道注意力来计算并更新权重, 引入注意力机制后的卷积神经网络模型可以提高网络的学习能力和分割精度. 为了提高网络的分割精度, 许多学者使用可变形卷

积去提升网络对目标区域的感知[14–17]. 除此之外, 随着注意力机制在计算机视觉中不断发展[6,7], 更多学者将注意力机制引入到语义分割中[19,20], 在特征图的空间域和通道域上增强目标区域特征, 抑制不相关的背景区域.

上述特征融合和注意力机制的思想为本文方法提供了更多的思路, 针对具有更多噪声干扰和轮廓模糊的眼球超声图像, 本文提出了一种语义嵌入分支的注意力机制的图像分割模型, 提高超声图像眼球区域分割的准确度和模型的鲁棒性. 本文的主要思想是通过对图像不同层次之间的语义特征进行融合从而提高模型对超声图像中眼球区域的分割能力; 引入注意力机制可以在突出关键区域特征的同时抑制不相关区域的特征响应, 减少错误的分割; 而不同层次语义特征的融合可以在保留图像细节纹理的同时减少全局语义信息的丢失. 本文主要的研究工作如下.

1) 在编码器-解码器 Unet 的基础上, 本文提出在第 3 次和第 4 次下采样过程中使用可变形卷积替代传统的卷积, 可变形卷积可以自适应地调整网络的感受野, 更好地适应目标区域的形状, 提高卷积神经网络的特征表达能力和分割精度.

2) 为了充分利用超声图像中眼球区域多尺度特征, 在上采样过程中通过构建语义嵌入的注意力机制, 生成具有权重系数的特征图, 通过对权重系数的更新从而达到突出超声图像中重要空间位置的关键信息和抑制不相关的背景区域的目的.

3) 将本文方法对比 3 种不同的深度学习分割方法, 基于可变形卷积的语义嵌入注意力机制, 更好地感知超声图像中眼球信息, 对超声图像中的背景噪声和眼球区域的边缘模糊, 实现更加精准且鲁棒的分割效果. 实验结果表明, 相较于其他的分割方法, 本文模型在超声眼球图像数据集上可以取得最高的分割精度, 像素准确率达到 98.15%.

1 基础知识

1.1 语义嵌入分支

不同层次的特征融合是语义分割中的一种重要的方法, 在卷积神经网络中低层特征图的分辨率较高, 包含丰富的空间位置信息, 但是其语义层次较低; 而高层特征具有更强的语义信息, 但是分辨率很低, 特征图的细节信息较少. 如图 1 所示, 本文中的语义嵌入分支在尽量不增加模型的参数和复杂度同时, 将高层的特征图经过一次核大小为 3×3 的卷积层和上采样操作, 再乘上来自较低层的特征图, 实现不同层之间的特征融合, 弥补高低层特征之间的差异, 减少了图像的噪声和灰度相似带来不良影响, 提高了模型的学习能力. 同时, 有助于后续上采样操作, 还原出更多的图像细节信息. 语义嵌入分支输入第l阶段编码器得到的特征图x_l, 同时来自较高层的特征图x_{l+1}通过核大小为 3×3 的卷积层, 经过上采样操作, 将图像分辨率增加一倍, 使通道数减少到原来的一半. 使得特征图的大小和通道的数量和来自低层的特征图保持一致. 最后, 将高层特征上采样后的特征图和低层的特征图相乘得到特征图y_l. 其中计算过程:

$$\bar{x}_{l+1} = Upsample(Conv(x_{l+1})) \tag{1}$$

$$y_l = \bar{x}_{l+1} \cdot x_l \tag{2}$$

其中, x_{l+1}经过 3×3 卷积再进行上采样, 得到特征图$\bar{x}_{l+1}$, 此时特征图$\bar{x}_{l+1}$的大小和低层特征图x_l一致, 最后将特征图$\bar{x}_{l+1}$乘上特征图x_l得到输出的特征图y_l.

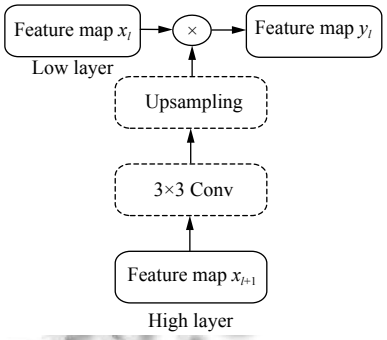


图 1 语义嵌入分支

1.2 注意力机制

语义分割中的注意力机制[21–23] 主要包括通道域和空间域, 通道域主要是对特征图的通道进行处理; 由于超声图像的特征, 本文中的注意力机制主要是在特征图的空间域上进行操作, 如图 2 所示; 注意力机制是由核大小为 1×1 的卷积层、非线性 ReLU 层, Sigmoid 层等组成. 核大小为 1×1 的卷积层可以减少网络参数, 调整特征图的通道数量; ReLU 激活函数增强模型的学习能力, 解决训练时梯度消失等问题; Sigmoid 函数用于特征图取值的归一化, 得到取值在[0 , 1]之间的概率图可以加速网络的收敛. 用输入特征图乘上 Sigmoid 函数后的特征图α_i, 可以得到每个像素都具有权重的特征图, 在图像中显著性特征取得较大的值, 在不相关的背景区域中像素取得较小的值, 从而增强显著特征和抑制不相关的区域, 为上采样操作提供更加精细的特

征, 从而有利于对本文中超声图像中眼球区域的分割.

本文的注意力机制, 重新调整了编码器的输出特征图, 更新特征图的权重, 可以实现对显著性区域的关注. 最后将具有注意力的特征图进行跳跃连接.

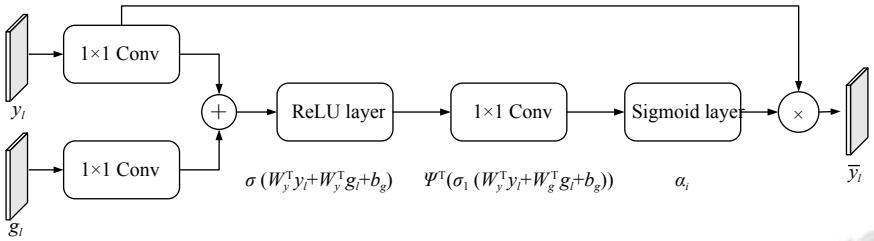


图 2 注意力机制

1.3 可变形卷积

由于卷积神经网络中卷积核的大小是固定的, 其感受野具有一定的局限性, 不能很好地感知超声图像中眼球区域的几何形状, 为了高效地提取眼球区域中关键语义信息, 本文在卷积层中引入可变形卷积, 可变形卷积模块可以在训练过程中学习偏移量来改变空间中的采样位置, 可变形卷积结构如图 3.

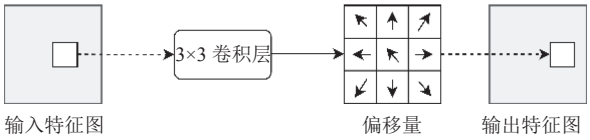


图 3 可变形卷积

图 3 中, 输入特征图x经过 3×3 的卷积, 其目的是在训练网络的过程中学习偏移量的大小, 生成具有偏移量的参数的特征图y. 可变形卷积在标准卷积中的每一个采样点位置上都加了一个可学习的偏移Δa_n, 可变形卷积使用偏移$\{\Delta a_n|n=1,\cdots,N\}$将区域 R 的每个点进行位移, 如以下公式所示:

$$y(a_0)=\sum_{a_n\in R} w(a_n)\cdot x(a_0+a_n) \tag{3}$$

$$y(a_0)=\sum_{a_n\in R} w(a_n)\cdot x(a_0+a_n+\Delta a_n) \tag{4}$$

其中, w为权重, R 为采样区域, R 为$\{(-1,-1),(-1,0),\cdots,(0,1),(1,1)\}$, a_0为输出特征图y中的点, a_n为采样区域R的所有采样点, 由于Δa_n为小数, 所以采用双线性插值计算$x(a_0+a_n+\Delta a_n)$的值.

2 模型的网络结构

本文模型新提出的卷积神经网络以 Unet 为基础架构, 如图 4 所示, 其结构主要是由编码器、语义嵌入的注意力机制、解码器组成. 编码器包括图像的输入和 4 次下采样过程; 输入分辨率为 96×96 的图像通过卷积核大小为 3×3 卷积层, 其中包括 3×3 卷积、批标准化 (batch normalization, BN)、ReLU 激活函数, 最后一个步长为 2 的最大池化层; 其中最大池化层用来实现下采样操作, 每次执行下采样操作都将特征通道数增加一倍, 图像的大小缩小一倍, 提取高层的语义信息. 如图 4, 第 3 次和第 4 次下采样过程中使用可变形卷积层, 可变形卷积相较于传统的卷积, 不再局限于固定的感受野, 可变形卷积网络可以适应目标区域的变化. 在卷积神经网络中增加更多的可变形卷积层, 提高卷积神经网络对图像相关区域的表征能力. 引入可变形卷积代替传统的卷积层, 对超声图像中眼球区域更好地感知, 能够为分割提供更加有效的特征.

下采样过程由池化操作来实现, 得到高层图像的语义信息, 经过语义嵌入分支实现不同层之间的特征融合, 使本文的网络充分学习超声眼球图像的丰富信息. 其次, 语义嵌入分支输出的特征图经过注意力机制, 生成具有注意力的特征图. 本文在中间过程中使用语义嵌入分支和注意力机制, 是由于最后一次下采样得到的特征使用跳跃连接就可以实现特征融合.

在本文网络结构的中间阶段, 其主要的思想是融合不同层之间的特征, 使用语义嵌入分支取代 Unet 中特征信息的直接通过跳跃连接与相应层的特征融合操作, 将低层特征信息和高层特征融合解决了上采样操作带来的低层特征信息的丢失问题和高层特征细节信息不足的问题.

本文模型语义嵌入分支主要是融合 3 个不同层的特征图, 下采样后的特征图y_l, 来自低层特征图x_l, x_{l+1}进行特征融合.

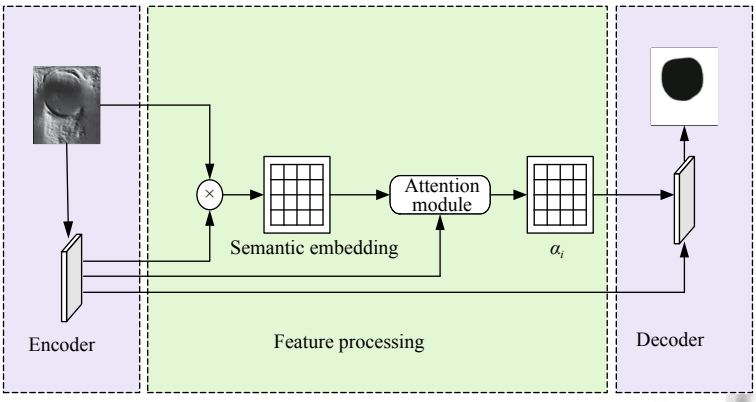


图 4 本文模型的结构

本文模型使用的注意力机制如图 2 所示, 输入编码器第l层的特征图y_l, 同样经过 1×1 卷积运算操作得到特征图$W_y^{\mathrm{T}}y_l$. 通过l+1 层的特征图上采样后的特征图g_l, 经过 1×1 卷积运算操作得到特征图$W_y^{\mathrm{T}}g_l$. 将上两步得到的特征图$W_y^{\mathrm{T}}g_l$和$W_y^{\mathrm{T}}y_l$进行相加后再进行非线性操作 ReLU 得到$\sigma\left(W_y^{\mathrm{T}}y_l+W_g^{\mathrm{T}}g_l+b_g\right)$, 随后再使用 1×1 卷积运算得到特征图$q_{att}$, 最后经过 Sigmoid 激活函数得到最终的注意力系数 attention coefficient (α_i). 用公式表示为:

$$q_{att}=\psi^{\mathrm{T}}\left(\sigma_1\left(W_y^{\mathrm{T}}y_l+W_g^{\mathrm{T}}g_l+b_g\right)\right)+b_\psi \tag{5}$$

$$\alpha_l=\sigma_2\left(q_{att}\right) \tag{6}$$

$$\bar{y}=\alpha_l\cdot y_l \tag{7}$$

其中, 偏置项 $b_\psi\in R$, $b_g\in R$, $\alpha_i\in[0,1]$. 利用注意力系数乘上输入的特征图y_l得到具有权重的特征图$\bar{y}_l$, 从而突出图像中显著区域, 有助于实现精准的分割. 由于图像分辨率在多次下采样的处理后较小, 本文网络仅在第 2 阶段和第 3 阶段使用了两次语义嵌入的注意力机制, 为后续的处理提供更加丰富特征.

解码器包括 4 次上采样过程, 首先是特征图经过一个上采样操作再和语义嵌入的注意力机制处理后的特征图进行连接, 克服了上采样操作造成的特征信息不足问题, 经过卷积核大小为 3×3 的卷积层, 每个卷积层后跟一个 ReLU 层, 同时上采样使特征通道数减半, 图像恢复原来分辨率大小的两倍, 多次上采样操作后恢复图像的大小. 最后经过一个核大小为 1×1 的卷积输出特征图.

在原始的 Unet 中由于高层特征对细节的感知能力差, 本文模型利用可变形卷积具有更好的感受野这一优势, 极大地提高网络对眼球区域的学习能力, 为语义嵌入注意力机制提供更多细节信息的特征, 其次本文模型从突出目标中显著性特征的角度出发, 通过构建语义嵌入注意力机制提高网络对眼球分割区域的注意力, 从而实现减少背景区域的错误分割.

本文以编码器-解码器的 Unet 为基础架构但又不局限于其本身, 针对眼球超声图像的特性, 本文模型基于可变形卷积, 构建语义嵌入注意力机制, 在不过多增加网络参数量的情况下, 本文模型具有良好的分割能力.

3 实验数据与参数

3.1 数据集

本文的数据集来自于临床采集到的图像, 该数据集包括 668 张超声眼球图像, 每张图像都有专家手动标注出眼球区域的掩膜, 本文将数据集中 500 张图像用于训练和 168 张图像用于测试. 为了解决样本过少容易造成过拟合的问题, 在训练之前本文对数据进行了预处理, 通过几何平移、随机缩放、旋转等图像增强方法对数据集进行数据增强, 最终得到 4 290 张图像作为训练集, 利用增强后的数据进行模型训练可以增加模型的鲁棒性和泛化能力.

3.2 实验的参数

实验环境基于 Linux 操作系统、Intel Xeon(R)5218 CPU、内存 32 GB、GeForce RTX 2080Ti GPU, 使用 CUDA 加速网络训练, 网络的训练采用了 PyTorch 1.0 深度学习框架, 实验参数主要是动量为 0.9 的 Adam 优化器, 初始学习率为 0.001, 200 个 epoch, batchsize 的大小为 128 张.

本文模型采用深度学习常用的交叉熵损失函数 (CrossEntropyError LossFunction), 交叉熵损失函数将每个像素点的类别预测概率与相应的图像掩膜数据进

行计算, 最后将结果求平均值, 图像分割后的结果和平均值相关. 交叉熵损失函数公式如下:

$$L=-\frac{1}{N}\sum_{i=1}[y_i\cdot\log_2(p_i)+(1-y_i)\cdot\log_2(1-p_i)] \quad (8)$$

其中, N为总的样本数, y_i表示第i个样本的标签, p_i表示第i个样本预测为正的概率.

4 实验及分析

4.1 模型的评估标准

为了验证本文模型的算法的有效性, 本文的网络模型在眼球超声图像数据集中进行了训练和测试. 本文使用了图像分割中常用的评估标准, 其中包括像素的准确率PA, 交并比IoU, 平均交并比$mIoU$. 像素准确率是指所有分类正确的像素数占全部像素的比例. 其中在$n+1$个类中, p_{ij}是本属于第i类却被分到第j类的像素数量, p_{ii}代表的是分类正确的正例像素的数量, PA的计算公式如下:

$$PA=\frac{1}{n+1}\sum_{i=0}^{n}\frac{p_{ii}}{\sum_{j=0}^{n}p_{ij}} \quad (9)$$

交并比是将图像真实分割的所有像素点p_{ii}和预测图像的分割所有像素点两个像素集合的交集和并集的比值, 而平均交并比则是将所有类的IoU取平均值. 其中IoU、$mIoU$的计算公式如下:

$$IoU=\frac{p_{ii}}{\sum_{j=0}^{n}p_{ij}+\sum_{j=0}^{n}p_{ji}-p_{ii}} \quad (10)$$

$$mIoU=\frac{1}{n+1}\sum_{i=0}^{n}\frac{p_{ii}}{\sum_{j=0}^{n}p_{ij}+\sum_{j=0}^{n}p_{ji}-p_{ii}} \quad (11)$$

4.2 对比实验

在训练过程中, 如图 5 所示, 对比目前较流行的 U 形分割网络[2]、语义嵌入的分割网络 (semantic-embedding Unet)[4]、注意力机制的分割网络 (attention Unet)[16] 和本文提出的语义嵌入的注意力机制分割网络的损失函数发现, 在 100 个 epoch 的时候, 本文模型的损失函数收敛最快, 表明在训练过程中可以更快学习到目标特征. 在 200 个 epoch 时候, 损失函数不再下降, 趋于稳定.

为了验证本文模型的分割能力和泛化能力, 在眼球超声图像测试集上得到结果如表 1 所示, 本文模型的准确率达到了 98.15%. 由于原始 U 形网络局限于使用跳跃连接进行特征融合, 没有注意到不同层特征之间的差异; 语义嵌入的分割网络根据超声图像的特点融合了高低层特征, 但基于传统的卷积, 不能对目标区域的几何信息进行更好地学习; 注意力机制有助于突出目标的显著特征, 因此本文基于可变形卷积 (deformable convolution), 分别对比 3 种分割网络.

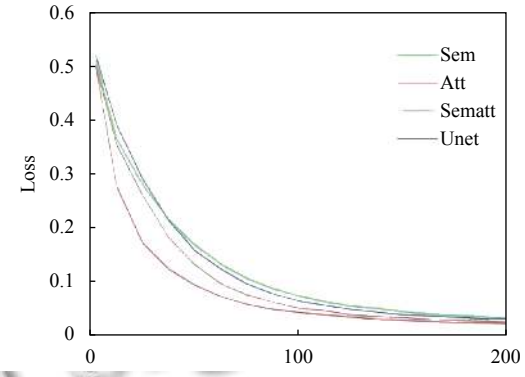


图 5 损失函数的变化曲线

表 1 对比实验的结果

网络	*IoU*	*mIoU*	*PA* (%)
Unet	0.945 5	0.897 7	95.04
Unet+Deform	0.969 1	0.906 3	96.35
SSnet	0.974 4	0.914 4	96.77
SSnet+Deform	0.966 9	0.897 8	97.14
Attention	0.954 1	0.862 3	96.14
Attention+Deform	0.956 6	0.877 5	96.01
SSnet+Attention	**0.976 5**	**0.924 4**	**97.97**
本文模型	**0.978 5**	**0.930 8**	**98.15**

通过表 1 中的数据定量分析发现, 使用可变形卷积的 Unet 比原始 Unet 的交并比提高了 2.36%, 有效地证明了可变形卷积比传统卷积有更好的感受野. 本文模型相比于原始 Unet[2]、语义嵌入分割网络[4]、注意力机制分割网络[16] 的像素准确度分别提高了 2.75%、1.38%、2.01%, 说明用可变形卷积对眼球区域进行高效的特征表示和用语义嵌入注意力机制增强显著性特征的有效性.

4.3 消融实验

如表 2 所示, 在下采样过程中不同阶段使用可变形卷积, 在低阶段使用可变形卷积并没有提升效果. 在第 3 个和第 4 个下采样过程中, 可变形卷积提升本文网络的表征能力. 尤其是第 3 阶段提到的特征图, 针对超声图像的区域模糊、边缘灰度相近等缺点, 可变形卷积提升网络对眼球区域的敏感度, 有助于提取更加丰富的语义信息, 间接地提高网络的分割能力, 所以本文模型选择在第 3 次下采样之后使用可变形卷积.

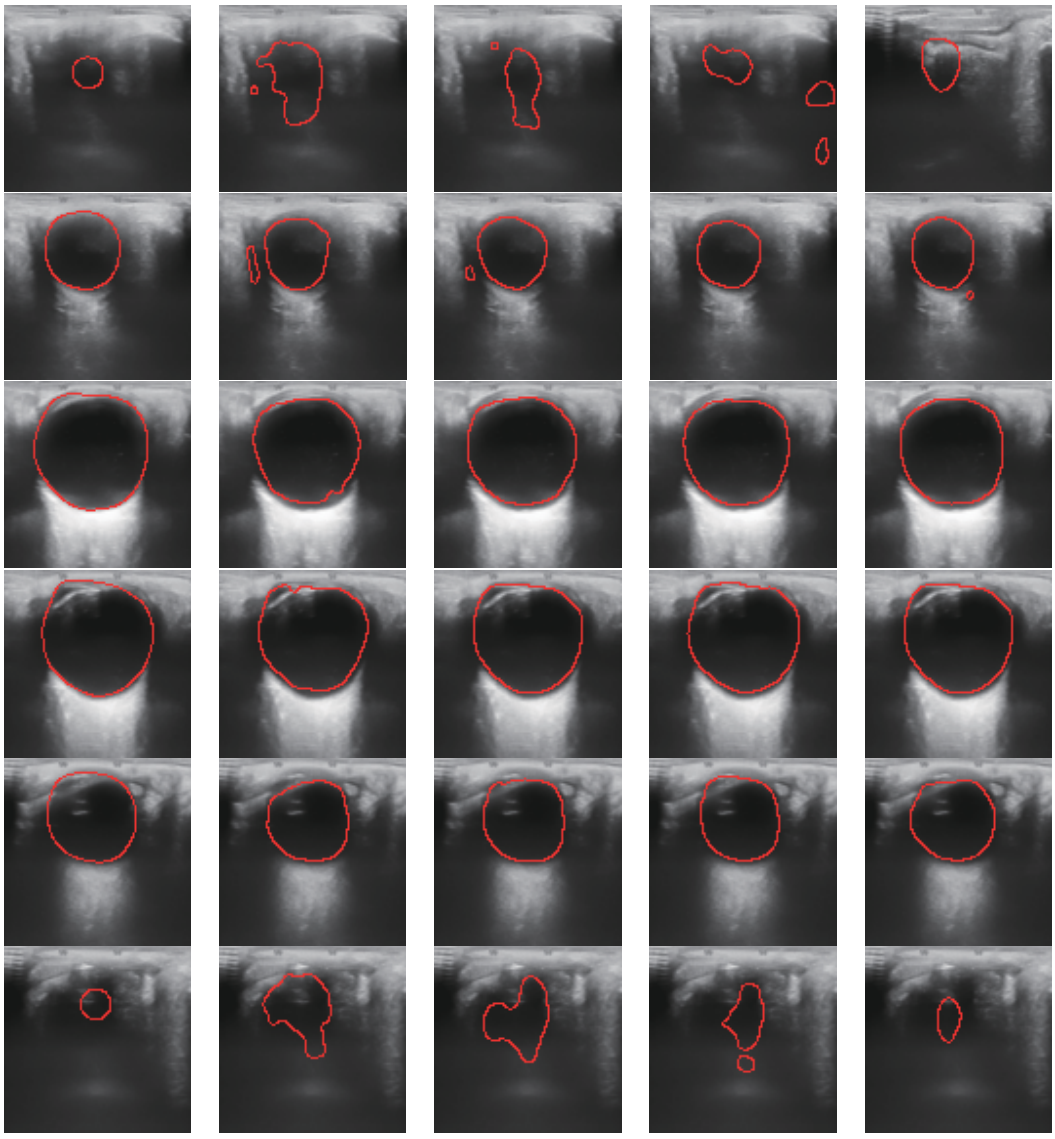
4.4 分割结果的可视化

如图 6 所示, 更加直观地对本文的卷积神经网络分割结果进行了可视化, 在眼球超声图像测试集上的分割

结果是二值图像, 因此本文提取边缘轮廓叠加到原图上, 红色部分就是提取到轮廓. 如第 1 行和第 6 行所示, 由于原始 Unet 受跳跃连接的限制, 提取后的特征并不能很好的还原更多的细节信息, 容易产生错误分割. 如图中第 3 列和第 4 列, 由于超声图像的灰度相近, 语义嵌入分割网络和注意力机制分割网络的结果容易造成过度分割. 最后一列是本文模型分割结果, 不仅能够实现对超声图像中眼球的边缘轮廓的平滑分割, 还减少了对背景区域的错误分割, 分割出来的区域更加合理.

表 2 不同层使用可变形卷积的结果

网络	*IoU*	*mIoU*	*PA* (%)
Conv1	0.958 6	0.865 6	96.31
Conv2	0.951 6	0.872 5	96.56
Conv3	0.969 0	0.915 1	97.34
Conv4	0.970 9	0.910 3	97.50



(a) 原图 (b) Unet (c) Attention-Unet (d) SS-Unet (e) 本文网络

图 6 测试集分割结果

本文模型的语义嵌入分支能够融合高低层的特征以此达到准确地分割, 注意力机制通过对超声图像中眼球区域的显著特征进行增强, 解决了超声图像灰度相似不易分割的困难.

5 结束语

本文针对具有轮廓模糊、灰度相似的眼球超声图像数据集, 从多尺度特征融合和注意力机制的角度出发, 提出了一种基于语义嵌入分支的注意力卷积神经

网络的语义分割方法, 用于超声图像中眼球区域的分割. 本文模型改进 Unet 的编码器和解码器网络结构, 利用可变形卷积提高模型对目标区域的感知, 构建语义嵌入分支实现了不同层之间的语义信息特征融合. 进一步在语义嵌入分支的基础上引入注意力机制, 突出了图像显著性特征, 抑制不相关的区域, 提高模型的学习目标特征的能力, 提高图像边缘分割的准确度; 为了验证本文模型分割的准确性和泛化能力, 将训练后的模型在测试集进行预测; 实验结果表明, 本文模型在 3 个评估标准上取得更高的分割精度, 证明了本文的模型能够克服眼球超声图像的缺点, 实现了较为精准的分割.

参考文献

1 Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Proceeding of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston: IEEE, 2015. 3431–3440.

2 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich: Springer, 2015. 234–241.

3 Zhang ZL, Zhang XY, Peng C, *et al*. ExFuse: Enhancing feature fusion for semantic segmentation. Proceedings of the 15th European Conference. Munich: Springer, 2018. 273–288.

4 Lin FC, Liu CB, Xie HT. Semantic-embedding and shape-aware U-Net for ultrasound eyeball segmentation. 2019 IEEE International Conference on Multimedia and Expo (ICME). Shanghai: IEEE, 2019. 892–897.

5 Gu ZW, Cheng J, Fu HZ, *et al*. CE-Net: Context encoder network for 2D medical image segmentation. IEEE Transactions on Medical Imaging, 2019, 38(10): 2281–2292. [doi: 10.1109/TMI.2019.2903562]

6 Mnih V, Heess N, Graves A, *et al*. Recurrent models of visual attention. Proceedings of the 27th International Conference on Neural Information Processing Systems, Volume 2. Montreal: ACM, 2014. 2204–2212.

7 Ba J, Mnih V, Kavukcuoglu K. Multiple object recognition with visual attention. arXiv: 1412.7755, 2014.

8 Jaderberg M, Simonyan K, Zisserman A, *et al*. Spatial transformer networks. Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal: ACM, 2015. 2017–2025.

9 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 7132–7141.

10 Vaswani A, Shazeer N, Parmar N, *et al*. Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: ACM, 2017. 6000–6010.

11 Wang F, Jiang MQ, Qian C, *et al*. Residual attention network for image classification. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017. 6450–6458.

12 Oktay O, Schlemper J, Le Folgoc L, *et al*. Attention U-Net: Learning where to look for the pancreas. arXiv: 1804.03999v1, 2018.

13 Alom Z, Hasan M, Yakopcic C, *et al*. Recurrent residual convolutional neural network based on U-Net (R2U-net) for medical image segmentation. arXiv: 1802.06955, 2018.

14 Dai JF, Qi HZ, Xiong YW. Deformable convolutional networks. 2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE, 2017. 764–773.

15 Zhu XZ, Hu H, Lin S. Deformable ConvNets V2: More deformable, better results. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019. 9300–9308.

16 Zhang M, Li X, Xu MJ. Image segmentation and classification for sickle cell disease using deformable U-Net. arXiv: 1710.08149, 2017.

17 Deng MY, Yang H, Li T, *et al*. Restricted deformable convolution-based road scene semantic segmentation using surround view cameras. IEEE Transactions on Intelligent Transportation Systems, 2020, 21(10): 4350–4362. [doi: 10.1109/TITS.2019.2939832]

18 Sun MY, Zhang GH, Dang H, *et al*. Accurate gastric cancer segmentation in digital pathology images using deformable convolution and multi-scale embedding networks. IEEE Access, 2019, 7: 75530–75541. [doi: 10.1109/ACCESS.2019.2918800]

19 Takikawa T, Acuna D, Jampani V, *et al*. Gated-SCNN: Gated shape CNNs for semantic segmentation. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2019. 5228–5237.

20 Niu RG, Sun X, Tian Y, *et al*. Hybrid multiple attention network for semantic segmentation in aerial images. IEEE Transactions on Geoscience and Remote Sensing, 2021. [doi: 10.1109/TGRS.2021.3065112]

21 Woo S, Park J, Lee J Y, *et al*. CBAM: Convolutional block attention module. Proceedings of the 15th European Conference. Munich: Springer, 2018. 3–19.

22 Li X, Wang WH, Hu XL, *et al*. Selective kernel networks. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019. 510–519.

23 Cao Y, Xu JR, Lin S, *et al*. GCNet: Non-local networks meet squeeze-excitation networks and beyond. 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). Seoul: IEEE, 2019. 1971–1980.