

引用格式: 刘梦瑶, 刘茹涵, 姚一静, 等. 应用 DeeplabV3+网络实现小儿髋关节超声图像识别[J]. 声学技术, 2022, 41(2): 235-239. [LIU Mengyao, LIU Ruhan, YAO Yijing, et al. Application of DeeplabV3+ network in ultrasonic image recognition of pediatric hip joint[J]. Technical Acoustics, 2022, 41(2): 235-239.] DOI: 10.16300/j.cnki.1000-3630.2022.02.013

应用 DeeplabV3+网络实现小儿髋关节 超声图像识别

刘梦瑶¹, 刘茹涵², 姚一静¹, 余倩³, 高乙惠³,
王芮¹, 盛斌², 姜立新¹

(1. 上海交通大学医学院附属仁济医院超声医学科, 上海 200127; 2. 上海交通大学电子信息与电气工程学院计算机系, 上海 200240;
3. 上海交通大学附属第六人民医院, 上海 200233)

摘要: 利用 Graf 法进行发育性髋关节发育不良(Developmental Dysplasia of the Hip, DDH)诊断时主要依靠骨-软骨交界面、股骨头、滑膜皱襞、关节囊及软骨膜、孟唇、软骨顶、骨性顶这 7 个解剖结构进行解剖验证, 而初级医生对上述结构识别困难, 因此文章提出了一种基于 DeeplabV3+的网络模型用于 7 个结构的分割识别。首先对纳入的 106 例图像进行手动标记和预处理, 之后将其分别输入 DeeplabV3+和 U-Net 两种网络模型中, 最终对其预测图表现和分割性能进行比较。与目前 DDH 图像分割中常用且表现优越的 U-Net 网络相比, DeeplabV3+网络的预测图包含的结构较多, 边界分割也较清晰, 其图像分割评价指标如相似性系数、豪斯多夫距离和平均豪斯多夫距离平均值的表现也优于 U-Net 网络。文章利用 DeeplabV3+网络实现了 DDH 超声图像的 7 个结构分割, 对临床医生进行后续图像的角度测量和分型诊断具有重要意义。

关键词: 发育性髋关节发育不良; 超声; 图像分割; 网络模型; DeeplabV3+

中图分类号: R445.1

文献标志码: A

文章编号: 1000-3630(2022)-02-0235-05

Application of DeeplabV3+ network in ultrasonic image recognition of pediatric hip joint

LIU Mengyao¹, LIU Ruhan², YAO Yijing¹, YU Qian³, GAO Yihui³,
WANG Rui¹, SHENG Bin², JIANG Lixin¹

(1. Department of Ultrasound, Ren Ji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200127, China;
2. Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China;
3. Department of Ultrasound, Shanghai Jiaotong University Affiliated No. 6 Hospital, Shanghai 200233, China)

Abstract: The diagnosis of developmental dysplasia of the hip (DDH) by Graf method mainly depends on the seven key structures, involving chondro-osseous border, femoral head, cartilaginous roof, synovial fold, labrum, bony roof and joint capsule. However, it is difficult for junior doctors to identify these structures. Therefore, a network model based on Deeplabv3+ for the segmentation of these seven structures is proposed in this paper. Firstly, 106 images were manually labeled and preprocessed, and then they were input into Deeplabv3+ and U-net network models respectively. Finally, the segmentation performances were compared. Compared with U-Net network which is commonly used and well-behaved in DDH segmentation, the image prediction of Deeplabv3+ network contain more structures and clearer boundary, and the main evaluation indices of segmentation, such as the average values of dice similarity coefficient, Hausdorff distance, average Hausdorff distance, also showed a better performance. The Deeplabv3+ network is first used to achieve segmentation of seven structures in DDH ultrasound images, which is of great significance for angle measurement and classification diagnosis.

Key words: developmental dysplasia of the hip (DDH); ultrasound; segmentation; network model; DeeplabV3+

收稿日期: 2021-10-08; 修回日期: 2021-11-03

基金项目: 国家自然科学基金(81771850)资助项目。

作者简介: 刘梦瑶(1996—), 女, 山西人, 硕士研究生, 研究方向为小儿髋关节超声图像的自动化分型诊断。

通信作者: 姜立新, E-mail: jinger_28@sina.com

0 引言

发育性髋关节发育不良(Developmental Dysplasia of The Hip, DDH)是新生儿最常见的先天性

疾病之一,目前超声检查是针对小儿早期 DDH 最普遍且最有用的影像诊断方法,规模化的新生儿 DDH 超声筛查正在国内兴起^[1-3]。超声检查因其无创、安全、易行、无辐射、费用较低、并可动态观察,特别适用于 DDH 高危人群的筛查及治疗后的连续随访。1980 年奥地利学者 Graf 开创的 DDH 超声检查的静态方法——简称 Graf 法^[4-5],主要利用 5~7.5 MHz 线阵探头取得包括 7 个主要解剖学标志的髋关节冠状切面,通过对骨-软骨交界面、股骨头、滑膜皱襞、关节囊及软骨膜、孟唇、软骨顶、骨性顶这 7 个主要解剖结构进行解剖验证,并在此基础上进行后续的可用性检测和测量分型。

然而,在超声图像上上述重要解剖结构的准确识别对从医时间较短、资历较浅(简称低年资)的医师和初学者而言较为困难,如孟唇由于其解剖位置和形态的个体差异较大,髋白的骨性顶和软骨顶由于软组织的覆盖不易显示全貌,辨认起来并不容易^[6]。

另外由于在临床实践中 DDH 超声诊断主要依靠人工取图及测量,其随操作者水平变化显著,因此最近科研人员们提出了一些自动分割的方法。Quader 等^[7]提出用置信加权结构相位对称(Confidence-weighted Structured Phase Symmetry, CSPS)特征来分割三维超声图像上髋关节不同平面骨结构,减少了软组织假阳性,提高骨分割效率。Pandey 等^[8]使用阴影峰(Shadow Peak, SP)进一步简化了骨阴影特征提取方法,在精确性和速度上都比 CSPS 有一定的提高。El-Hariri 等^[9]为了提高骨定位的准确性,首先利用基于相位的特征提取,应用空间解剖先验来消除软组织假阳性,准确分割髋骨和髋白的轮廓;其次使用深度学习中较流行的单通道和多通道输入的 U-Net 观察到深度学习特征的性能优于增强工程特征,如 SP 和 CSPS。

虽然上述方法在髋关节超声图像的分割上取得了一定的突破,但其高度依赖人工提取的特征,在新数据的鲁棒性和通用性上仍面临很大的挑战。此外,目前的研究只关注了孟唇和骨性顶的分割,还未有研究对髋关节图像的 7 个主要解剖结构及其相对位置进行分割。而对上述 7 个关键结构进行人工图像标注是一项费时费力的工作。因此,本研究将利用在图像分割领域多个公开数据集中表现优越的网络模型——DeeplabV3+网络的基础上,探讨其对 DDH 二维冠状位超声图像上 7 个解剖结构的分割价值,并将其与目前 DDH 图像分割中常用的表现优越的 U-Net 网络进行比较,探索最适合用于本研究中 DDH 超声图像分割的网络模型。

1 材料与方 法

1.1 试验材料

本研究以疑似髋关节发育不良或存在 DDH 高危因素、并进行超声筛查的 0~6 月患儿作为研究对象,其纳入及排除标准如下所示。

1.1.1 疑似标准

(1) 臀部或大腿内侧褶皱不对称;(2) 双下肢长度不等及活动程度不一致;(3) 单侧髋关节外展活动受限;(4) 早产儿且体重低于平均值。

1.1.2 高危因素

(1) 女性;(2) 臀先露;(3) 巨大儿;(4) DDH 家族史;(5) 胎儿过度成熟;(6) 婴儿襁褓;(7) 羊水过少;(8) 其他引起体位性变形的宫内因素。

1.1.3 排除标准

(1) 排除不合格的超声图像(如存在 7 个主要解剖结构显示不全/不清楚,不满足中间平面等问题);(2) 排除脑瘫、关节挛缩、化脓性髋关节关节炎等疾病所致的 DDH 图像;(3) 排除其他髋关节疾病及肢体畸形患者的图像;(4) Graf 法 III 型及以上 DDH 患者的图像。

1.1.4 受试者分组

本文纳入了 106 名患儿的 106 张二维超声冠状面图像,超声图像由 5~7.5 MHz 线性换能器采集,图像深度为 40~55 mm,将其按 4:1 比例随机分为训练集和测试集,即训练集包括 86 张图片,测试集包括 20 张图片,其基本信息如表 1 所示。

表 1 DDH 图像分割所用病例基本信息表
Table 1 Basic characteristics of the study participants

数据集	数量	中位年龄(四分位距*)/d	人数		分型	
			男	女	I 型	II 型
训练集	86	53(13-113)	32	54	59	27
测试集	20	49(8-98)	6	14	14	6

注: *四分位距表示第一四分位数与第三四分位数的差距。

1.2 方 法

1.2.1 手动标记

106 张二维超声冠状面图像中包括骨-软骨交界面、股骨头、滑膜折痕、关节囊及软骨膜、孟唇、软骨顶、骨性顶在内的 7 个关键结构,被一位经验丰富的超声医师手动标记出其边界,手动标记示意图如图 1 所示。将经过预处理后的训练集和测试集分别输入 DeeplabV3+网络和 U-Net 网络这两种网络模型中进行训练。

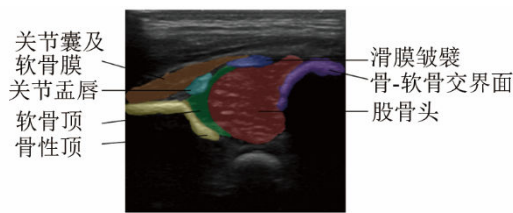


图 1 小儿髋关节二维冠状位超声图像手动标记图
Fig.1 Manual labeling of ultrasonic image of hip joint

1.2.2 预处理操作

本研究使用图像处理技术裁剪了超声图片中包含病患敏感信息的部分。为了使小儿髋关节结构的形状不失真，将裁剪得到的超声图片用 0 填充至标准大小，图像尺寸为 572×748 像素，并且所有的图片都采用 3 通道的彩色图像模式。为了应对数据量较少的问题，本文对训练集中的原图及其对应的标注图进行数据增强操作，包括伽马变换、旋转、归一化等，将图片扩增至 1 204 幅。

1.2.3 网络模型

本研究所采用的 DeeplabV3+是图像语义分割领域较为主流的网络模型，在实际生活中应用较广^[10]。DeeplabV3+网络具有多尺度的卷积层，同时具备编码-解码双重模块，能够实现对图像所包含结构的精细分割，能够平衡分割精度和分割效率，其分割性能优异。其中 Deeplabv3 网络作为编码模块，能够实现目标特征分层嵌套提取和多尺度上下文信息提取。而解码模块则将 Deeplabv3 骨干网络中产生的低级特征和高层抽象特征相融合，最后对特征采用 4 倍非线性上采样输出预测结果，其结构如图 2 所示。

U-Net 网络是语义分割领域较早使用的全卷积网络算法之一，因其使用包含压缩路径和扩展路径的左右对称的 U 形结构而得名，并在医学图像分割领域有着较广泛的应用^[11]。而且 U-Net 网络使用跳跃链接结构将上、下采样层连接，可将下层提取到的特征直接传递给上层，这也是 U-Net 网络分割精度较高的主要原因。

1.2.4 评价指标

为了评估本网络模型的分割效果，实验使用相似性系数(Dice Similarity Coefficient, DSC)、豪斯多夫距离(Hausdorff Distance, HD)、平均豪斯多夫距离(Average Hausdorff Distance, AHD)作为评价指标。DSC、HD、AHD 是医学图像分割领域的主要评价指标，其中 DSC 用来度量医生标注区域与预测区域之间的重叠部分，DSC 取值范围为 0~1，DSC 越接近 1 则表示两个区域越相似；HD 和 AHD 则是用于衡量边界分割的指标，用来描述两组轮廓之间的相似程度，值越小代表两者之间轮廓越相似。

1.2.5 网络训练

将同一数据组分别输入 DeeplabV3+网络和 U-Net 网络，分别对 86 例 DDH 二维超声图像进行训练，训练后保存模型并对测试集中的 20 例图像进行预测。本试验采用 Python 语言并在 TensorFlow 和 Keras 框架下实现，使用交叉熵损失函数和 Adma 训练算法，迭代轮次为 100，学习率为 1×10^{-5} ，最后计算 7 个解剖结构的 DSC、HD、AHD 等评价指标，并对两种网络的预测结果进行比较，评估两种网络的效果。

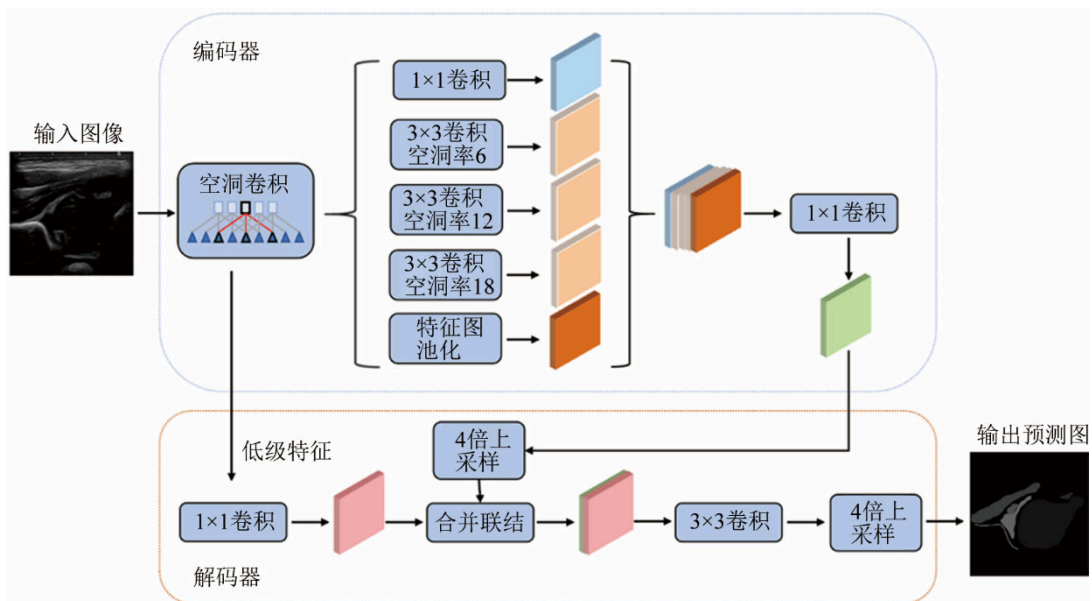
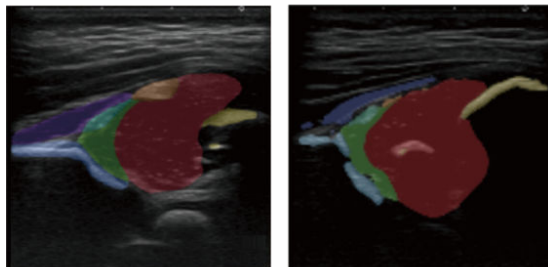


图 2 Deeplabv3+网络结构模式图
Fig.2 The pattern diagram of Deeplabv3+ network

2 结果

2.1 两种网络的预测结果图比较

将 DDH 二维超声图像分别输入 DeeplabV3+网络和 U-Net 网络这两个网络中进行分割预测,从预测的结果来看,DeeplabV3+网络有较好的分割效果(如图 3(a)所示)。利用 U-Net 网络预测得到的分割效果较差,不仅无法完全分割出 7 个解剖结构,比如对于关节孟唇和滑膜皱襞这种微小精细结构,同时对每个结构的边界分割也较为粗糙,尤其是在各结构毗邻处,如骨性顶和软骨顶交界处、软骨顶与孟唇交界处等(如图 3(b)所示)。



(a) DeeplabV3+网络 (b) U-Net 网络

图 3 DeeplabV3+网络和 U-Net 网络的预测结果图

Fig.3 The predicted results of Deeplabv3 + network and U-Net network

2.2 两种网络的分割性能定量比较

本研究采用 DeeplabV3+网络和 U-Net 网络这两种网络模型来分割 DDH 中的 7 个关键结构,最终利用 DeeplabV3+网络得到 7 个结构的 DSC、HD、AHD 平均值分别为 86.50%、10.22、0.39,相比 U-Net 网络的 84.45%、13.06、0.50,各项指标均表现较好,结果如表 2 所示。这两种网络模型在股骨头、骨性顶、关节孟唇及骨-软骨交界面这几个骨

表 2 DeeplabV3+网络和 U-Net 网络的分割性能比较

Table 2 Comparison of segmentation performance between Deeplabv3 + Network and U-Net Network

序号	DeeplabV3+			U-Net		
	DSC/%	HD	AHD	DSC/%	HD	AHD
1	88.58	8.77	0.30	87.07	12.42	0.40
2	91.53	13.63	0.33	90.45	14.14	0.38
3	81.74	19.30	0.75	79.31	22.61	0.76
4	81.59	7.29	0.47	77.30	11.27	0.83
5	82.35	6.14	0.36	79.76	7.98	0.49
6	88.20	6.08	0.26	86.48	9.09	0.34
7	91.53	10.36	0.22	90.76	13.94	0.27
平均	86.50	10.22	0.39	84.45	13.06	0.50

注:序号 1~7 分别代表关节孟唇、股骨头、滑膜皱襞、关节囊与软骨膜、软骨顶、骨-软骨交界面、骨性顶。

性结构上的分割性能较好, DSC 值高达 86.48%~91.53%;而在滑膜皱襞、关节囊及软骨膜、软骨顶上的分割性能则相对较差, DSC 值为 77.30%~82.35%。

3 讨论

由于在临床实践中初级医师对重要解剖结构的理解程度不一,且 DDH 超声诊断随操作者水平变化显著,因此研究者相继提出了一些自动分割方法。El-Hariri 等^[9]使用 U-Net 网络模型对髋关节三维超声图像中的髋臼和髌骨进行分割,其 DSC 值为 86%,与 Quader 等^[7]提出的置信加权结构相位对称特征提取方法(CSPS, DSC=81%)以及 Pandey 等^[8]使用的阴影峰(SP, DSC=75%)相比, U-Net 网络的分割性能较为优异。然而,本研究发现,与目前 DDH 图像分割中常用且表现优越的 U-Net 网络相比, DeeplabV3+网络作为另外一种图像分割领域最常用的网络模型,对 DDH 二维超声图像中 7 个结构的分割效果更好。从预测出来的超声图像中可以看出, DeeplabV3+网络的 DSC 平均值较高,为 $86.50\% \pm 5.20\%$,表示模型预测的区域与高年资医生标注的区域重合范围较大,二者重合率较高;而从预测图中也可以看出, DeeplabV3+网络^[10]在可分割出的结构数目和单一结构边界分割中相对表现较好,其能够较好识别出 DDH 判别时必需的 7 个解剖结构。另外,代表边界分割准确性的两个参数:HD 和 AHD 的平均值分别为 10.22 ± 7.32 和 0.39 ± 0.44 ,明显低于 U-Net 网络模型的值,表明 DeeplabV3+网络对这 7 个结构边界的检测能力较高,在预测图上 7 个结构的边界分割也相对清晰可辨。这可能是由于 DeeplabV3+网络加入了编码-解码的思想,对编码器中的 Xception 进行调整,以及加入了类似于 U-Net 网络的解码器结构,解码时能够与低级信息相连,解决了编码时经过步长卷积或者池化层后的细节信息丢失,有助于恢复图像边缘信息。因此, DeeplabV3+网络利用其具有的多尺度卷积层,能够实现对髋关节二维超声图像中 7 个主要结构的精细分割,具有良好的分割性能。对于超声医生后续利用孟唇中点、骨性顶中的平直髌骨面和髌骨下缘最低点进行角度测量将会有很大的帮助。

除此之外,这两种网络对上述 7 个重要解剖结构的分割精确度具有相同的趋势,都对骨性顶、股骨头这两个结构的分割准确度较高,而对软骨顶、滑膜皱襞、关节囊及软骨膜的识别精度较低,原因

在于髋关节中骨性结构在超声图像上显示为强回声, 边界清楚, 轮廓锐利, 其包含的强回声像素点与周围的低回声软组织灰阶水平对比较为明显, 因此很容易被肉眼和网络模型识别出来。而识别精度较低的 3 个结构则是软组织结构, 其回声较低甚至呈无回声, 边界欠清, 因此较难识别。

综上所述, 对于这种语义分割任务, 不仅需要注重图像分割的内容和边缘等细节, 往往还需要结合更多详细的空间信息, 比如图像中各解剖结构间的相对位置关系等, 从整体上来判断分析各个结构的边界。另外, 由于本研究所纳入的手动标记样本量较少, 使得网络模型缺乏充分的训练, 预测效果也会受到很大影响。因此在今后的研究中, 我们将进一步扩大手动标记的图像样本量, 在此基础上引入结构位置信息处理网络、注意力机制等^[12-14], 进一步提高本网络模型对于 DDH 超声图像中重要解剖结构的分割精确度和可解释性。

4 结论

本文应用 DeeplabV3+网络模型, 利用其强大的编码、解码能力, 将其所提取的低级特征用于对 DDH 二维冠状位超声图像的像素级精准分割, 初步实现了对 DDH 超声图像上 7 个解剖结构的分割, 相比于目前 DDH 图像分割中常用且表现优越的 U-Net 网络而言, 内容及边界识别都更加精细, 具有较高的分割性能, 对后续 DDH 超声图像的标准性判定、角度测量和分型诊断具有重要意义。

参 考 文 献

- [1] JACKSON J C, RUNGE M M, NYE N S. Common questions about developmental dysplasia of the hip[J]. *American Family Physician*, 2014, **90**(12): 843-850.
- [2] LODER R T, SKOPELJA E N. The epidemiology and demographics of hip dysplasia[J]. *ISRN Orthopedics*, 2011, 2011: 238607.
- [3] 边臻, 陈涛. 发育性髋关节发育不良临床诊疗指南(0~2 岁)[J]. *中华骨科杂志*, 2017, **37**(11): 641-650.
- [4] 于静淼, 陈涛. 超声在发育性髋关节发育不良诊疗中的临床应用[J]. *中华医学超声杂志(电子版)*, 2015, **12**(1): 6-10.
- [5] GRAF R. Hip sonography: background; technique and common mistakes; results; debate and politics; challenges[J]. *Hip International*, 2017, **27**(3): 215-219.
- [6] MOSTOFI E, CHAHAL B, ZONOBI D, et al. Reliability of 2D and 3D ultrasound for infant hip dysplasia in the hands of novice users[J]. *European Radiology*, 2019, **29**(3): 1489-1495.
- [7] QUADER N, HODGSON A J, MULPURI K, et al. Automatic evaluation of scan adequacy and dysplasia metrics in 2-D ultrasound images of the neonatal hip[J]. *Ultrasound in Medicine & Biology*, 2017, **43**(6): 1252-1262.
- [8] PANDEY P U, QUADER N, GUY P, et al. Ultrasound bone segmentation: a scoping review of techniques and validation practices[J]. *Ultrasound in Medicine & Biology*, 2020, **46**(4): 921-935.
- [9] EL-HARIRI H, MULPURI K, HODGSON A, et al. Comparative evaluation of hand-engineered and deep-learned features for neonatal hip bone segmentation in ultrasound[C]//*Medical Image Computing and Computer Assisted Intervention-MICCAI 2019*, 2019.
- [10] CHEN L C, ZHU Y K, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//*Computer Vision-ECCV 2018*, 2018.
- [11] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation[C]//*Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015*, 2015.
- [12] ISLAM M A, JIA S, BRUCE N D B. How much position information do convolutional neural networks encode? [EB/OL]. 2020: arXiv: 2001.08248[cs.CV]. <https://arxiv.org/abs/2001.08248>.
- [13] SUN H, LI C, LIU B Q, et al. AUNet: attention-guided dense-upsampling networks for breast mass segmentation in whole mammograms[J]. *Physics in Medicine and Biology*, 2020, **65**(5): 055005.
- [14] FU J, LIU J, JIANG J, et al. Scene segmentation with dual relation-aware attention network[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, **32**(6): 2547-2560.

[1] JACKSON J C, RUNGE M M, NYE N S. Common questions about developmental dysplasia of the hip[J]. *American Family*