

• 听力重建 •

三维卷积神经网络在中耳手术相关颞骨影像
自动分割中的应用探索*柯嘉¹ 吕弈² 杜雅丽¹ 王君臣² 王江¹ 孙世龙¹ 马芙蓉¹

[摘要] 目的:研究基于三维卷积神经网络 3D U-net 神经网络的深度学习方法,在临床常规颞骨 CT 影像中,对中耳手术相关的迷路、面神经和听骨进行全自动分割的可行性。方法:随机调取门诊常规颞骨 CT 检查患者的薄层 CT 扫描数据 30 例为正常结构组,既往耳蜗、面神经和听骨形态或走行变异者各 1 例为异常结构组。所有数据由两位临床医生在 Mimics 20.0 软件中,对面神经、迷路和听骨 3 个结构进行手工初分割和精细分割,同时利用 3D U-Net 对上述数据进行深度学习。分别对正常结构组中测试集 5 例和异常结构组中的迷路、听骨和面神经进行手工分割与自动分割的 Dice 相似指数(DSC)比较。结果:利用 3D U-net 网络结构对常规颞骨 CT 中迷路、听骨和面神经进行自动分割,其 DSC 分别为 0.79 ± 0.03 、 0.64 ± 0.05 和 0.49 ± 0.09 ;对异常的迷路、听骨和面神经的识别,其 DSC 也可达到 0.71、0.54 和 0.40。结论:根据颞骨解剖特点,采用 3D U-net 神经网络结构,可以实现对迷路、听骨和面神经的全自动化分割,并获得接近手工分割的精度,该方法可行、快捷、准确度高。

[关键词] 深度学习;卷积神经网络;中耳手术;面神经;自动分割

doi:10.13201/j.issn.2096-7993.2020.10.002

[中图分类号] R764.9 **[文献标志码]** A

Application of 3D U-net in automatic segmentation of middle ear
surgery structures in temporal bone CTKE Jia¹ LV Yi² DU Yali¹ WANG Junchen² WANG Jiang¹
SUN Shilong¹ MA Furong¹

(¹Department of Otorhinolaryngology Head and Neck Surgery, Third Hospital, Peking University, Beijing, 100191, China; ²School of Mechanical Engineering and Automation, Beihang University)

Corresponding author: MA Furong, E-mail: furongma@126.com

Abstract Objective: To study the feasibility of fully automatic segmentation of labyrinth, facial nerve and ossicles in clinical routine temporal bone CT images based on 3D U-net neural network. **Method:** Clinical data were divided into two groups: ① Normal group: data were randomly assigned from 30 patients for routine temporal bone CT examination; ② Abnormal group: cochlear, ossicles and facial nerve morphology variation of 1 case each. The structures of facial nerve, labyrinth and ossicles were manually initial segmented and fine segmented by 2 clinicians with Mimics 20.0. Three-dimensional convolutional neural network(3D U-Net) was selected to conduct deep learning on the same data. The dice similarity coefficient(DSC) was used as the evaluation index. **Result:** The 3D U-net neural network was used to automatically segment the labyrinth, ossicles and facial nerve in the routine temporal bone CT. In the normal group, the DSC of labyrinth, ossicles and facial nerve were 0.79 ± 0.03 , 0.64 ± 0.05 and 0.49 ± 0.09 , respectively. In the abnormal group, the DSC of these structures were 0.71, 0.54 and 0.40. **Conclusion:** According to the anatomical characteristics of the temporal bone, the labyrinth, ossicles and the facial nerve can be totally automatic segmented by 3D U-net neural network, and the accuracy was closed to that of manual segmentation. This method is feasible, fast and accurate.

Key words deep learning; convolutional neural network; middle ear surgery; facial nerve; automatic segmentation

*基金项目:国家自然科学基金(No:61701014,61911540075);首都卫生发展科研专项项目资助(首发-2016-2-4094);北京大学第三医院临床重点项目(No:BYSSY2017025)

¹北京大学第三医院耳鼻咽喉头颈外科(北京,100191)

²北京航空航天大学机械工程及自动化学院

通信作者:马芙蓉,E-mail: furongma@126.com

颞骨 CT 检查是进行鼓室成形、中耳重建手术之前必须的一项检查,有助于准确地了解病变的范围、听骨结构的连续性以及重要解剖结构如面神经、半规管、耳蜗等的相互位置关系及病变破坏情况,帮助我们提前对整体手术进行合理的规划,选择合适的听骨假体和重建方式。但是一些解剖结构的异常在以平面形式呈现的 CT 图像上有时很难直观地发现,在图像分析上也费时费力。目前人工智能已经广泛应用于医学影像和病理诊断中^[1],在肺、乳腺、心脏、颅脑、肝脏、前列腺、骨骼等部位均得到了广泛应用。根据不同病变的影像学特性,采用不同的深度学习的算法来进行处理,可提升影像诊断的准确性和时效性,减轻医生的工作强度,也能够让临床专家丰富的阅片经验在广大医生中得到普及,然而在耳鼻喉科领域这一应用尚处于探索阶段。本研究采用自行设计的 3D U-net-plus 网络结构,尝试对颞骨内的面神经、迷路及听骨进行自动识别及分割,并在一些变异的解剖结构中进行验证,以期对未来的术前评估提供参考。

1 资料与方法

1.1 临床资料

正常结构组:随机调取门诊常规行颞骨 CT 检查患者的薄层 CT 扫描数据 30 例,均为 18 岁以上成人颞骨 CT,颞骨结构已完成发育。扫描设备为德国 SIEMENS / SOMATOM Definition Flash 64 排螺旋 CT,层厚 0.60 mm,层间距 0.30 mm,像素 0.412 mm。除外中耳炎症性病变、先天性中耳内耳畸形,选择正常颞骨影像中薄层轴位骨窗 CT 数据。其中 25 例作为训练集,用于神经网络结构的训练;另外 5 例作为测试集,用于对训练的结果进行验证。

异常结构组:选取既往诊断过 Mondini 畸形患者的薄层 CT 1 例、先天性听骨链发育畸形患者的薄层 CT 1 例,及术中发现有面神经走行变异的患者 CT 扫描数据 1 例,扫描参数同上。

观察指标:重点观察上述两组耳蜗、听骨链、面神经 3 个解剖结构的空间位置,耳蜗转数,听骨是否融合、分离或缺失及面神经的走行情况等。

1.2 研究方法

1.2.1 解剖结构手工分割 手工分割由两位在颞骨 CT 阅片方面有丰富经验的临床医生完成。将上述数据导入 Mimics 20.0 软件中,进行三维重建。CT 的阈值在 -1000~2000 范围内。分别对面神经(自茎乳孔至内听道底),迷路(包括耳蜗、前庭、半规管管腔部分)和听骨进行手工分割。其中迷路的分割所选择的蒙罩(mask)阈值在 -1000~1000,面神经和听骨均直接编辑。在轴位、冠状位及矢状位上分别对目标结构进行编辑,获得初分割的结果,经区域生长,三维重建获得三维重建后的

效果图。对照重建后的解剖结构,再对上述初分割数据进行精细分割,主要在上述结构与邻近解剖结构交界区的部位,逐一像素区进行比对,最终生成精细分割的三维重建数据,生成 STL 文件。

1.2.2 深度学习方法的建模及实现 选择三维卷积神经网络(convolutional Neural Networks)3D U-Net 作为深度学习中的神经网络结构部分。采用翻转、拉伸、变形等手段对训练集的数据进行扩增,获得 1250 组训练数据。选择自适应矩估计(Adam)优化器作为迭代过程中的优化器。权重交叉熵(weighted cross entropy)的方法,作为损失函数进行计算。训练过程通过可视化手段实时监控损失函数和验证数据集上的分割精度,对收敛速度和趋势进行判断并及时调整网络参数。以手工分割的结果作为金标准,采用 Dice 相似指数(Dice similarity coefficient, DSC)作为分割精度评价标准,用于反映自动分割与手工分割的相似指数,分别对正常结构组和异常结构组中的面神经、迷路、听骨链的分割精度进行测试。其中 $DSC = 2 \times \text{预测正确的结果} / (\text{真实结果} + \text{预测结果})$,其越接近于 1,说明分割的准确度越高。

2 结果

2.1 手工分割与自动分割结果

利用 Mimics 软件获得面神经、迷路及听骨的初分割、精细分割图,利用 3D U-net 网络结构对颞骨 CT 数据进行自动分割训练,其分割显示见图 1。

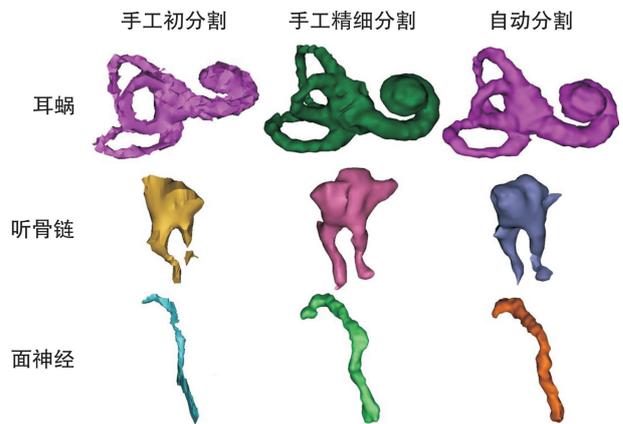


图 1 采用手工分割与自动分割对正常解剖结构进行分割

2.2 利用神经网络对正常的迷路、听骨链和面神经进行自动分割的精度

完成 3D U-net 网络结构的训练后,分别对正常结构组中训练集的 1250 例样本和测试集中的 5 例样本进行面神经、迷路及听骨的自动分割与手工分割图像的精度比较,见表 1。

2.3 利用神经网络对异常的迷路及听骨链和面神经进行自动分割

同理,利用 3D U-net 网络结构对异常的迷路、

听骨链和面神经进行自动分割,获得的自动分割与手工分割图像的对比见图 2。对于 Mondini 畸形患者的影像资料,3D U-net 网络结构可以自动识别出 1 周半的耳蜗;听骨链畸形患者术中证实为砧骨长脚末端变细,砧镫连接中断,自动分割中也可以看到砧骨长脚变短;面神经走行异常患者术中证实为面神经向下方遮盖前庭窗,通过比较自动分割与手动分割,可以看到两者识别的面神经其中心线是重合的。

表 1 自动分割与手工分割的 DSC 比较

	例数	迷路	听骨	面神经
正常结构组				
训练集	1 250	0.81±0.02	0.63±0.05	0.54±0.06
测试集	5	0.79±0.03	0.64±0.05	0.49±0.09
异常结构组	1	0.71	0.54	0.40

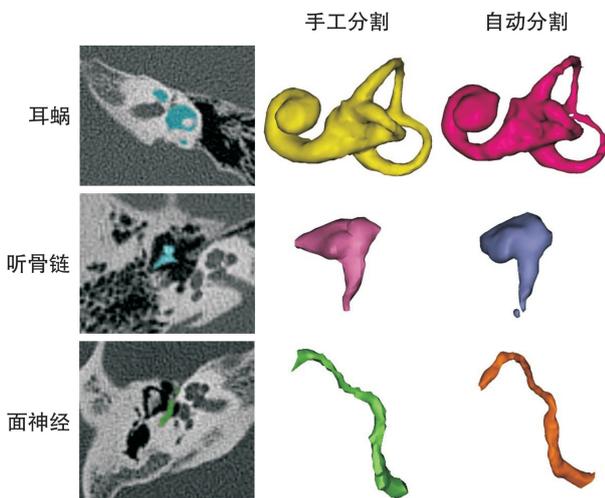


图 2 采用手工分割与自动分割对异常解剖结构进行分割

3 讨论

深度神经网络是新一代人工智能技术的代表,是一种模仿人类大脑构造设计的数学模型。它通过深层非线性网络结构无限逼近复杂函数,以获取数据的原始特征,进而取代传统的由人工提取数据的方法,实现让机器自主学习并提取数据特征^[2]。目前深度学习已广泛应用于医学领域,特别是涉及到大量图像数据的医学影像领域,在图像分割、图像配准、影像复原、影像生成与增强等方面,开展了大量的研究;广泛应用于影像中肺结节的诊断、儿童自闭症的早期诊断、乳腺癌的筛查、皮肤癌的检测、前列腺及脑组织内肿瘤的发现等很多专科领域^[3]。它的出现,极大地解放了临床医技人工分割图像的劳动,且其诊断的准确率接近人工水平。

中耳手术最基本也是最困难的技术,就是要掌握颞骨解剖。颞骨 CT 为我们了解解剖提供了影像学参考。由于颞骨由不同气化程度的骨质构成,

其中的解剖标志深处于骨质之内或骨质内部的空腔内,这些结构既有骨性的,如骨性耳蜗、半规管、听骨等,又有软组织性质的,如面神经、血管、肌腱等,多以管状结构为主,解剖结构体积小、结构细长,使得这些标志的分割和识别较为困难,因此深度学习在颞骨影像的图像分割中并没有得到广泛应用。本研究中,我们选取在颞骨解剖中最重要的 3 个解剖结构进行了尝试。

3.1 听骨

自 1956 年 Wullstein^[4] 提出了围绕听骨链重建(ossicular chain reconstruction)为中心的鼓室成形术的分类方法,中耳手术的目的不再仅仅是彻底清除病变,重建完整、连续、活动的听骨链,改善术后听力逐渐受到重视。因此,术前准确获取听骨信息,重建听骨形态,对于中耳植入假体的选择起到了一定的参考作用。在影像上,听骨结构的优点是,听骨位于含气的中耳腔内,其骨性结构与周边含气腔的密度差异较大,使得听骨特别是锤骨和砧骨的骨性轮廓较易辨别;缺点是听骨结构较小,特别是镫骨,在临床常用的颞骨 CT 中,往往在 CT 上仅有 1 层,且骨质密度较低。受中耳炎病变的影响,听骨被破坏从而在影像学上部分或全部缺失的现象很常见。本研究中,我们使用 3D U-net 网络,对听骨的识别与人工识别的相似度为 0.64,主要原因是听骨的整体体积较小,由于 CT 阈值的变化,听骨边缘的识别差异会在听骨整体中占有较高的比例。Fauser 等^[5] 使用 0.2 mm×0.2 mm×0.4 mm 的超高清影像学数据,利用 2D U-net 神经网络结合主动形态模型的方法,对面神经、耳蜗、听骨等结构实现了高精度分割。Elfarnawany 等^[6] 使用同步辐射相位对比成像(synchrotron radiation phase-contrast imaging)及显微 CT(micro-CT)对颞骨成像,可以更清晰地跟踪和勾画镫骨。随着放射性技术的发展,听骨链可以被更为精细地分割,并作为深度学习的训练集,从而大大提高自动分割的精度。

3.2 迷路

包括耳蜗、前庭、半规管,虽然属于内耳的结构,但是却跟中耳手术密切相关,如外半规管是中耳手术中定位面神经的重要解剖标志。迷路分为外侧的骨迷路和内侧的膜迷路两部分,其中膜迷路位于骨迷路内侧含有淋巴液的管腔中。在 CT 中,骨迷路与其内侧充满液体的管腔之间有较为清晰的边界,而骨迷路外侧与颞骨的骨质之间往往没有明确的界限。因此在迷路的分割上,我们实际上分割的是迷路中央的管腔部分。Powell 等^[7] 采用联合骨密度模型和感兴趣区描记方法进行的半自动分割方法,获得了 0.8 的 Dice 相似度。由于这部分从影像上容易识别,因此 3D U-net 网络学习后,

自动分割的 Dice 相似度可达到 0.79,也是这 3 个结构中自动分割效果最好的一个结构。本研究挑选了 10 例前庭导水管扩大合并 Mondini 畸形患者的 CT,经 3D U-net 网络学习后,也获得了 0.71 的自动分割效果。提示该网络模型能够很好地识别出异常的耳蜗结构。

3.3 面神经

面神经是中耳手术中最为重要的一个解剖结构,由于面神经走行较长,通过单一层面的观察不易发现面神经的走行或形态变异。如果利用大数据分析及自动分割等人工智能手段,可以在术前对面神经的解剖变异提出预警,提前发现诸如面神经遮窗等异常,将有助于在术前进行合理的手术预期,术中加强防范。本研究中,利用 3D U-net 网络学习后,面神经自动分割的 Dice 相似度仅有 0.49。面神经在影像上识别的难度最大,自动分割的效果不佳,分析原因主要有:①面神经呈微小管状结构,直径仅有 0.8~1.7 mm,呈软组织密度,与周围组织分辨力不高,易与邻近的气房相混淆;②面神经结构细长,同时存在两个弯曲^[8],使得常规颞骨 CT 扫描下轴位、冠状位或矢状位均难以完整显示面神经的全程;③面神经走行范围内有其他软组织结构混淆,如水平段上方与鼓膜张肌伴行,锥体段与镫骨肌毗邻,也给面神经的识别带来了困难;④面神经常常有走行上的变异,或有分支,或因面神经骨管的缺失有出现局部疝出等可能,因此,即使是有经验的临床医生,有时也要在不同的重建位上反复移动层面,才可能准确地识别出面神经。Noble 等^[9]提出了一种基于主动形状模型的识别方法来识别在常规 CT 影像中部分可见的结构,在面神经的识别上可达到 0.3 mm 的精度;Reda 等^[10]将基于“图谱”的方法与统计模型算法相结合,对儿童和成人患者的面神经进行分割,自动生成的结构与手工生成的结构平均误差分别为 0.23 mm 和 0.155 mm;Powell 等^[7]采用密度模型与标记后的感兴趣区进行比对,在面神经的分割相似度上也高达 0.7。本研究对比分割的效果图发现,采用 3D U-net plus 神经网络获得的面神经走行的中央线与手工分割的重复性较好。主要在神经的直径上与手工分割有较大差异,这与分割时选择的阈值有一定关系。随着学习网络的优化,这一精度也将得到进一步的提高。

近年来,人工智能在影像学的应用逐渐受到关注,本研究使用的 3D U-net 模型是在卷积神经网络

的基础上提出的,其特点是可对整个扫描的所有影像直接训练,从而更有效地利用全局信息。与既往人工智能基于正常的解剖结构信息与 CT 上感兴趣区域相匹配的方法不同,该方法在异常的解剖结构中也能够获得较为满意的分割相似度。在本研究中,对 Mondini 畸形、听骨链畸形等均获得了较好的识别。这是一种基于全自动的分割方法,虽然尚不能达到要求的精度,但随着训练样本量的增加以及算法的不断优化,这一精度还会得到不断提升。该方法的使用,将大大解放临床医生的工作负担,通过自动分割、重建三维结构,可以有效地发现异常结构的分布规律,在术前进行有效的规划,在手术方案设计、听骨假体的选择以及手术模拟培训上发挥更大的作用。

参考文献

- [1] 潘亚玲,王晗琦,陆勇. 人工智能在医学影像 CAD 中的应用[J]. 国际医学放射学杂志,2019,42(1):3-7.
- [2] 夏黎明,沈坚,张荣国,等. 深度学习技术在医学影像领域的应用[J]. 协和医学杂志,2018,9(1):10-14.
- [3] 宫进昌,赵尚义,王远军. 基于深度学习的医学图像分割研究进展[J]. 中国医学物理学杂志,2019,36(4):420-424.
- [4] Wullstein H. Theory and practice of tympanoplasty [J]. Laryngoscope,1956,66(8):1076-1093.
- [5] Fauser J,Stenin I,Bauer M, et al. Toward an automatic preoperative pipeline for image-guided temporal bone surgery[J]. Int J Comput Assist Radiol Surg, 2019,14(6):967-976.
- [6] Elfarnawany M,Rohani SA,Ghomashchi S, et al. Improved middle-ear soft-tissue visualization using synchrotron radiation phase-contrast imaging [J]. Hear Res,2017,354:1-8.
- [7] Powell KA,Liang T,Hittle B, et al. Atlas-based segmentation of temporal bone anatomy[J]. Int J Comput Assist Radiol Surg,2017,12(11):1937-1944.
- [8] 马晓波,赵守琴,李洁,等. 正常耳颞骨内面神经形态分析[J]. 中国耳鼻咽喉头颈外科,2015,22(6):287-289.
- [9] Noble JH,Labadie RF,Majdani O, et al. Automatic segmentation of intracochlear anatomy in conventional CT[J]. IEEE Trans Biomed Eng,2011,58(9):2625-2632.
- [10] Reda FA,Noble JH,Rivas A, et al. Automatic segmentation of the facial nerve and chorda tympani in pediatric CT scans [J]. Med Phys, 2011, 38 (10): 5590-5600.

(收稿日期:2020-06-04)