

多模医学影像阅片交互系统

21172710 韩霖

摘要

医学影像作为一种重要的辅助诊疗手段如今被大规模运用在临床实践中，通过CT，MRI等非侵入式诊断方法，医生能够准确快速的对患者体内的病变位置，程度和性质有准确的了解，从而有针对性地制定诊疗方案。但是每天大量产生的医学影像数据也给影像科医生带来了巨大的挑战。一方面大型医院的影像科医生平均每天通常需要看50~100组影像并编写报告，工作负担极大；另一方面一些医疗资源匮乏的地区面临着空有设备没有能读片的医生或者医生水平有限的情况。本文提出多模医学影像阅片交互系统旨在通过语音，手势，触摸板，键鼠等多模交互提升影像科医生编写诊断报告的效率，并通过计算机视觉方法对待诊断的影像进行理解，更好的领会诊断医生的意图，在报告编写过程中进行结构化自动补全，甚至通过主动交互给医生提供诊断提示，提升生成报告的质量。

领域现状

目前患者在医院进行影像检查的大概流程如下。当主治医生认为病人有必要进行影像学检查时会让病人统一到影像科进行检查。影像科医生需要结合主治医生的需求判断病人是否需要提前服用显影剂，影像采集过程中的使用的剂量等，并指导病人进行检查。待病人的影像数据采集完毕，医生需要为患者出具诊断报告，供主治医生参考。

每天影像科一位医生发出报告的数量一般在50份以上，由此不难看出医生每天的工作负担都很重。按看一组扫描8分钟计算，发出一天的所有报告就需要超过6个小时，这并没有包括指导病人进行检查的时间。按每份报告100字计，一天需要用键盘输入超过5000字[9]。目前患者在进行影像检查后不能立即拿到报告，其中急诊报告大约需要等待30分钟，普通门诊报告大约1小时，住院病人报告一般保证当天发出。综上可以看出当前的纯键鼠交互模式效率很低，医生撰写报告的过程很累，患者等待的时间也很长。随着医疗条件的改善和影像科技技术的发展，患者进行的影像学诊断会越来越多，但影像科作为一个辅助科室很难吸引大量的医生加入。综上，影像科医生的工作负担仍然会加重，因此研究如何提升医学影像报告编写过程中的交互效率十分有意义。

通过查阅国内外的文献，当前针对医学影像方面人机交互的研究大多集中在一些具体的场景[1][3][4][7]，比如如何利用手势在影像数据中进行3D导航[2]，利用VR设备对影像数据进行展示[8]等。针对提升医生阅片和编写报告效率和质量的完整系统的研究基本没有。一方面这体现了对这一领域研究的缺乏，另一方面笔者认为更多的研究成果直接以产品的形式被应用在了临床实践，这也提示了提出的人机交互方式需要与实际临床需求紧密结合。

创新点和系统细节

系统的主要创新点在于两方面：通过语音，体感，触摸板和键鼠的多模交互方式提升交互效率；基于计算机视觉和自然语言处理技术对影像内容和报告文本进行分析，提供数据自动补全，内容纠错，在限定场景下主动交互，提供诊断提示。

近年来随着深度学习技术，尤其是循环神经网络和注意力机制的发展，语音识别技术的准确率和鲁棒性已经得到了极大的改善[15]。相比目前常用的键盘输入方法，语音识别输入更加

高效。正常人的语速通常在键盘输入速度两倍以上。和通用的语音识别输入系统相比，影像报告撰写场景使用的词汇比较有限，类似脏器名称，疾病名称这种词语会频繁出现，针对这些高频词对识别系统进行训练能有效保证识别准确率。此外语音识别不会打断医生阅片的思路，阅片和打字很难同时进行但是在阅片过程中口述观察到的症状非常自然。

语音交互主要包含两类内容，一类是通过语音识别直接输入报告的内容，另一类是语音指令。用户通过说出唤醒词，鼠标点击屏幕按钮或键盘快捷键让系统开始收音和识别。系统对用户说的内容进行分析，查找其中满足一定格式的语音指令，比如输入脏器体积，输入病灶位置等，并通过计算机视觉方法对影像进行处理以满足用户需求。除语音指令外的所有内容都被原文输入报告编辑区，并利用自然语言处理技术进行自动纠错。

目前的阅片系统大都使用鼠标在3D的影像中进行导航，鼠标只包含点按，拖动和滚轮滚动操作比较有限，很多操作，比如放大和缩小，都需要在菜单中进行选择或者结合键盘快捷键才能做到。触摸板支持的操作远超过鼠标，比如可以区分轻按重按，单指双指三指操作以及双指拉伸动作等，在阅片过程中常用的片内移动，切换片层，放大缩小，改变窗宽窗位等操作都可以用触摸板完成，降低交互的复杂度并提升效率。

自然语言处理技术主要用于报告编写过程中的自动补全与聚焦和报告完成后的校验。在影像报告中有很多固定的书写结构，比如在某位置可见某病灶或某脏器内有多少病变。在传统的交互方式中，医生需要将片子手动聚焦到发生病变的部分，在片子中用标注工具添加箭头指向病变具体位置，并在报告中输入相应内容。通过自然语言处理技术可以在用户语音输入过程中识别到用户的意图，对下文内容进行自动补全并自动聚焦影像和添加标注。在报告完成后，自然语言处理技术可以分析报告内容，对其中的语法，专有名词使用进行分析纠错。此外可以结合计算机视觉对报告的内容进行校验。

用户每天长时间固定姿势久坐可能会引发肩颈，腰等部位的疾病，在系统中引入一些体感交互能让用户在工作过程中对一些关键部位进行活动，避免工作中疲劳和长期积累形成疾病。精确的体感交互需要引入专业设备，比如景深摄像头，这用户实际使用中会增加额外的成本，而且对效率提升不是很显著。因此系统只基于RGB摄像头做一些简单的交互。比如在CT读片过程中针对特定器官有固定的窗宽窗位，切换不同的窗宽窗位操作可以用面向屏幕向左或向右偏头实现。这个操作活动了颈椎而且非常明显，识别率高，交互可靠。

最后针对一些常见的病灶或形态学特征，系统可以集成计算机视觉识别模块。当医生完成报告编写，系统对类似肺结界，肝脏肿瘤等可以识别的很准确的病症进行扫描，结合报告的自然语言处理结果，主动提示医生可能存在漏诊的现象。利用计算机视觉和自然语言处理结合的方法提升报告内容的质量。

验证实验

系统的主要目标是提升交互效率，此外在限定场景下给出诊断提示，针对这两点设计实验。针对交互效率，实验需要选取同等年资，职称的医生，随机分成实验组和对照组。实验组教学如何使用新的多模交互系统，并实际在工作中使用一周，对照组仍使用传统的基于鼠标和键盘的交互方法。实验期间，尽量保证两组医生阅片数量和内容一致，方法是让同一个科室的两名医生分别使用传统方法和新系统撰写报告，所有的患者两人交替分配。统计的变量包括：读片时间，编写报告时间（类似与统计输入速度，从有一个输入开始计时，中间中断超过1分钟则停止计时），修订审阅报告时间，从开始读片到发出报告的总时间，点按交互次数，报告字数，规范程度，诊断质量以及一些衍生指标类似看片时间占总时间比例，输入时间占总时间比例，报告每字时间等。

针对主动交互提升报告质量设计实验，实验需要一些参与工作时间较短的医生，收集一些他们近期撰写的诊断报告，对报告质量进行评价，包括字数，规范程度，是否有误诊漏诊等。之后教学新的交互方式，使用一周比较熟练之后再收集一些用新系统撰写的报告，统计同样的指标。需要注意的时间跨度不可以很长，最开始收集的报告和熟练使用新系统的报告时间跨度不应该超过一个月，而且中间医生不应该进行了进修等会明显提升诊断水平的学习活动。

预期结论

第一个实验预期结果：使用新交互方式的医生撰写报告用的总时间更短，在撰写报告过程中，读片时间占比最大，输入时间和修订时间占比较低。相比对照组，读片时间占比提升，输入时间和修订时间占比下降，点按交互次数明显下降，报告字数更多，平均每字时间少，报告规范程度高，诊断质量基本持平。第二个实验预期结果：医生使用新的交互方式之后报告的字数变多，报告整体质量提升，漏诊情况明显减少。通过这些指标能够证明系统确实能够提升医生出报告的效率，并且系统能给年资低的医生有效的辅助。

未来展望

多模交互系统的主要应用场景在医院的影像科室中。目前因为医院中影像科的待遇和辅助科室地位的问题，也有一些影像人选择成立私营的影像服务机构，提供阅片和3D重建等业务，在这些机构中也可以用到本系统。系统中包含自然语言处理，语音识别，计算机视觉等多种机器学习技术，积累更多的数据可以有效提升算法的表现。比如针对不同口音的地区收集语音数据并对语音识别模块进行改善，通过自然语言处理技术生成影像数据对应的标注，并提供给计算机视觉系统进行训练等。通过更多数据和实际临床应用的淬炼，希望系统未来能有效的提升医生撰写报告的效率，从目前的完全手动输入编辑到未来医生只作为一个决策者，利用语音，触摸板等方式下达简单的指令或进行简单的点选就能快速生成报告。希望系统能够帮助医生有效降低报告编辑时间，从现在的50%以上的时间都在编辑报告到未来超过90%的时间可以用来阅片，只用小于10%的时间实际将阅片发现形成报告。

参考文献

- [1]Preim, Bernhard. "HCI in medical visualization." *Dagstuhl Follow-Ups*. Vol. 2. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2011.
- [2]Soutschek, Stefan, et al. "3-d gesture-based scene navigation in medical imaging applications using time-of-flight cameras." *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 2008.
- [3]王世英, et al. 医学影像诊断多媒体教学系统的研究与开发. Diss. 2000.
- [4]王蕊, 王屹峰, 俞任飞, 郑琪. 影像科医生的烦恼：我们就像影子[J]. 钱江晚报. 2017.12.5
- [5]王天明,熊曾,周晖,罗维军,唐海雄,刘进康. COVID-19规范化影像学诊断报告的设计、验证与临床实践[J]. 中南大学学报(医学版),2020,45(03):229-235+221.
- [6]曲滨鹏,缪佳. 基于LabVIEW的DICOM医学影像处理技术分析[J]. 电子世界,2020(06):184-185.
- [7]万兵,汪利琴,刘沛武. B/S模式下三维可视化医学影像系统的设计[J]. 中国医疗设备,2019,34(08):105-107+119.
- [8]杨晨彬,高瞻. 基于WebVR的医学体数据可视化[J]. 计算机应用与软件,2019,36(05):101-107.
- [9]付昕. 医学影像科信息系统设计与实现[D]. 河北科技大学,2019.
- [10]黄永刚. 区域医疗医学影像交互信息共享方法仿真[J]. 计算机仿真,2019,36(04):103-106.
- [11]李亚鹏. 医学影像云存储系统的设计与实现[D]. 郑州大学,2019.
- [12]刘婧,李坤,艾杨. 人工智能在医疗中的应用[J]. 信息与电脑(理论版),2019(05):150-151.
- [13]冯浩哲,张鹏,徐欣楠,郝鹏翼,吴福理,吴健,陈为. 面向3D CT影像处理的无监督推荐标注算法[J]. 计算机辅助设计与图形学学报,2019,31(02):183-189.
- [14]刘兆明,郭景,柯永振. 面向Web的远程图像可视化系统的研制[J]. 计算机应用与软件,2018,35(06):190-196+262.
- [15]刘诚然. 基于深度学习的远场语音识别技术研究[D]. 战略支援部队信息工程大学,2019.

