

医学图像关键点检测深度学习方法与挑战

李居朋¹, 王颖慧², 李 刚²

(1. 北京交通大学电子信息工程学院, 北京 100044; 2. 北京大学口腔医学院, 北京 100081)

摘要: 作为众多医学图像处理的前提和关键, 医学图像关键点检测具有重要的理论研究和应用价值. 由于个体间差异性和个体内歧义性的影响, 以及更高的临床应用定位精度的要求, 医学解剖关键点检测面临着巨大的挑战. 鉴于深度学习技术在医学图像关键点检测乃至整个医学图像处理领域都表现出了强大的实力, 本文全面检索发表于顶级医学期刊和会议论文集的医学图像关键点研究成果并进行了详细的梳理和综述. 从计算机视觉任务角度简述医学图像关键点检测及其存在的难点; 总结了深度学习技术在医学图像关键点检测中的基本框架, 详细论述了医学图像关键点检测的分类问题和回归分析两种不同类型的解决思路; 最后探讨了医学图像关键点检测深度学习方法面临的挑战、主要应对策略和开放的研究方向.

关键词: 医学图像处理; 关键点检测; 深度学习; 卷积神经网络

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 0372-2112(2022)01-0226-12

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20200725

Research and Challenges of Medical Image Landmark Detection Based on Deep Learning

LI Ju-peng¹, WANG Ying-hui², LI Gang²

(1. School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China;

2. Peking University School of Stomatology, Beijing 100081, China)

Abstract: As an entrance and challenge for many medical images processing, it is clinically essential to study on the medical image landmark detection and localization. Due to inter-individual variability and intra-individual ambiguity, as well as higher accuracy requirements of clinical application, the detection of medical anatomical landmarks was facing enormous challenges. In view of strength of deep learning of medical image landmark detection and the entire medical image processing field, we comprehensively retrieves relevant papers published in the top medical journals and conference proceedings to conduct a detailed review of these research findings. First of all, we briefly introduce difficulties in medical image landmark detection from the view of computer vision tasks. Secondly, we describe basic framework in medical image landmark detection, and discuss two different categories: classification and regression landmark detection solutions. Finally, we discuss the challenges and practicable strategies in deep learning for medical image landmark detection, as well as open research.

Key words: medical images processing; landmark detection; deep learning; convolutional neural networks

1 引言

医学解剖学意义上的“关键点”定义为人体普遍具有特定特征的、位置和拓扑上存在对应关系的一些点或者曲线^[1]. 医学图像解剖学关键点的自动检测是医学图像处理研究领域一个重要而活跃的课题, 作为众多医学图像分析应用的前提基础, 已被广泛用于医学图像配准^[2-4]、组织分割^[5-8]、参数测量^[6,7,9,10]、病理诊

断^[1,11-16]以及治疗规划^[17-20]、手术引导^[21-24]或其他医学图像处理的初始化^[25]等. 形式上, 对应给定的医学图像 X , 关键点检测算法预测关键点集合 L 的位置 $x = \{(x_1, y_1, z_1), (x_2, y_2, z_2), \dots, (x_L, y_L, z_L)\}$, 其中 x 、 y 和 z 表示关键点坐标.

然而由于人体解剖结构的多样性, 尤其是潜在局部相似关键点的情况下, 精确而鲁棒的解剖学关键点

定位变得充满挑战^[26]. 在过去的几十年里,医学图像关键点检测取得了较多的研究进展,当前主要的解决方案可以分为五大类,即基于知识、模式匹配、统计学习、混合技术和深度学习的方法. 第一类是利用人类对关键点结构知识模拟手动检测过程^[27,28],但由于模式过于复杂,无法随图像复杂度的增加而制定规则. 随后,一些研究人员采用了模式匹配搜索的策略^[29,30],但这类方法对个体间的差异异常敏感. 考虑到全局空间约束和关键点位置局部信息的同等重要性,基于统计学习的关键点检测方法脱颖而出,如“主动形状模型”^[31]和“主动外观模型”^[32],也出现了一些基于上述混合技术的方法^[33]. 在 IEEE ISBI 2014 和 Grand Challenge 2015 挑战赛中^[19,20],结合随机森林回归投票和统计形状分析技术的两个框架性能表现良好^[17,18]. 后续多个研究都以 Grand Challenge 数据集 (<https://grand-challenge.org/>) 为基础开始了相关技术研究^[34-36]. 以上这些传统的检测方法超出了本文的讨论范围,如有兴趣可参考相关的研究论文.

最新的深度学习技术在计算机视觉领域取得了巨大的成功,激发了国内外学者们将其应用于医疗图像分析的研究热情,已在医学图像分类、检测、分割、配准和检索等方面表现出传统技术无法比拟的性能^[37,38],正如哈佛大学医学院 Wells 教授给出的判断,应用深度学习解决医学图像分析任务是本领域的发展趋势^[39,40]. 自 2016 年开始,已有多位专家充分利用深度学习技术的多级语义自动学习特征,克服先前方法在特征定义和提取中的局限性,提出了多种用于医学关键点检测的有效解决方案. 在此基础上,本文依托于课题组在国家自然科学基金项目中的相关研究工作,聚焦于深度学习技术在医学图像关键点检测这一特定应用领域的研究现状和挑战,使用谷歌学术搜索引擎 (<https://scholar.google.com/>), 设定检索关键词包括 medical images AND (landmark detection OR landmark localization OR landmark digitization), 范围涵盖了医学图像处理领域顶级的期刊和知名的国际会议论文集 (包括 Medical Image Analysis 等在内的多个刊源), 以及最新发表在 arXiv 网站上的论文,并逐一筛选出以深度学习技术重点解决医学图像关键点检测问题的文献. 据知,这是第一份关于医学图像关键点检测的深度学习论文综述,相信这份清单对于相关领域的学习者或研究者而言将是一个很好的助力.

2 相关领域研究

图像关键点本质上是一种特征,是对图像中一个固定区域或者空间物理关系的抽象描述,描述的是一定邻域范围内的上下文关系^[41]. 医学图像关键点检测

与人脸关键点检测 (Facial Landmark Detection)^[42]、人体关键点检测 (Human Pose Estimation)^[43]、手势关键点检测 (Hand Pose Estimation)^[44]、服装关键点检测 (Fashion Landmark Detection)^[45] 等研究内容 (图 1) 是计算机视觉研究领域中的核心任务和热点问题,也是相关的更深层次应用的基础. 同属于关键点检测范畴的这些任务间的研究目标和方法存在着一定的相通之处. 医学图像关键点检测中很多思路来自于具有更多研究基础的人脸、人体等检测研究成果. 为了更好地展开问题的讨论,下文将对相关研究进行简要论述.

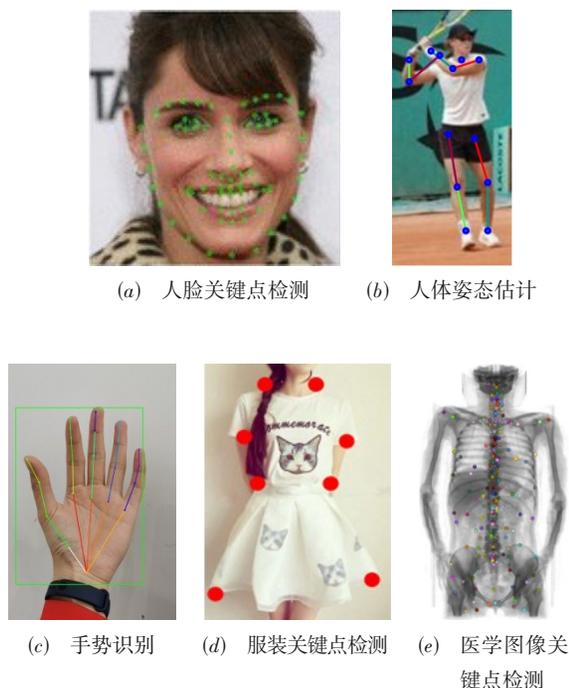


图 1 多个不同的关键点检测任务实例

注:(a)~(b)分别来自参考文献[43]和文献[42],(c)由腾讯音视频实验室提供,(d)选自 FashionAI 全球挑战赛,(e)来自参考文献[1]

传统的图像关键点检测主要包括基于模型匹配、约束模型、形态回归等方法,但由于容易受到姿态变化、物体遮挡等因素的影响,关键点检测性能提升有限从而大大限制了技术的实际应用. 2012 年 Hinton 课题组为了证明深度学习的潜力,首次参加 ImageNet 图像识别比赛,其通过构建的 CNN 网络 AlexNet 一举夺得冠军,也正是由于该比赛 CNN 吸引了众多研究者的注意,深度学习开始迎来超级发展时期,借助深度学习技术,图像关键点检测也完成了从传统方法到深度学习的转变^[44]. 此后,研究者们提出了人脸检测的 Face++ 版 DCNN、TCNN、DAN 框架,以及人体检测的 Convolutional Pose Machines、Stacked Hourglass Network 等诸多优秀的关键点检测网络模型,关键点检测与定位性能得到不断提升,相关领域的更多研究可参考文献[43,45~47]

等综述类文献。

用于人体、人脸等关键点检测的研究思路和方法可以作为医学图像关键点检测研究的借鉴,但医学图像关键点与上述问题在图像数据类型、检测精度要求等方面也存在着区别。依据 Bookstein 等人^[41]给出的医学关键点定义将其分为三类:①相邻组织间的位置;②具有最大曲率或局部形态突变的位置;③几何形态上极值点的位置。其中第①类关键点多被冠以特定医学名称或标签,具有明确的解剖学意义和更可靠的点一点对应关系,被广泛用于医学研究与临床应用。

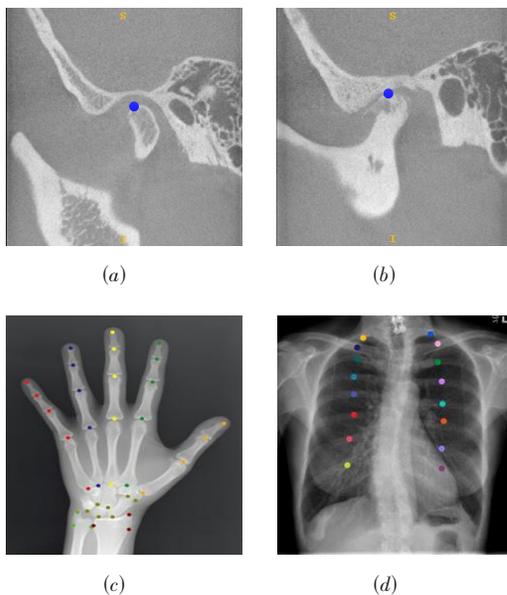


图2 医学图像关键点形态表现差异

注:(a)和(b)为颞下颌关节髁突顶点的个体间差异(颞下颌关节图像由北京大学口腔医学院放射科提供),(c)和(d)展示了个体内关键点的歧义性,(c)和(d)分别来自参考文献[29]和文献[23]

医学图像关键点存在以下特点:①患者个体间关键点形状存在差异性,而且这种差异可能表现得非常巨大,例如颞下颌关节髁突顶(蓝点)受多方面的影响,因而其外观在个体间存在很大的差异[图2(a)和(b)];②患者个体内关键点形状存在歧义性,人体可能存在与待检测关键点具有相似局部外观的多个点,最极端示例是在重复的人体骨骼结构上定义的关键点类别,包括人体中的手部骨骼和胸部肋骨[图2(c)和(d)]等。

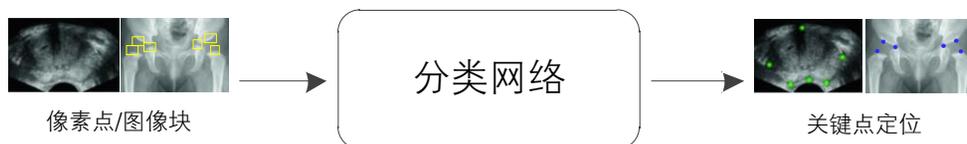


图4 基于分类的关键点检测框架

注:主要包括像素点和图像块两种不同的分类方法

正是由于这些特点的存在,使得医学图像关键点检测问题存在巨大的挑战。在以上检测难点共存的情况下,如何提升医学图像关键点检测的定位精度是众多研究者的关注重点。本文首先对医学图像关键点检测方法的国内外研究现状进行分类与整理;然后探讨并归纳医学图像分析深度学习方法的挑战及主要应对策略;最后给出对该领域相关技术发展趋势的思考与讨论。

3 医学图像关键点检测深度学习方法的研究现状

医学图像关键点检测深度学习的方法广泛采用监督学习的方式,即利用一组关键点标注数据样本训练与调整深度学习网络的参数,使其达到所要求分类或者回归性能的过程。根据学习问题的类型可以将现有的医学图像关键点检测深度学习方法划分为两大类:一类利用像素点分类方式解决,另一类则是关键点坐标回归的方法。图3统计了这两类方法的研究论文数量,其中采用分类框架的方法远低于基于回归分析的研究,针对每一类方法的类型细分将在后续研究现状分析中给出更为详细的梳理与讨论。

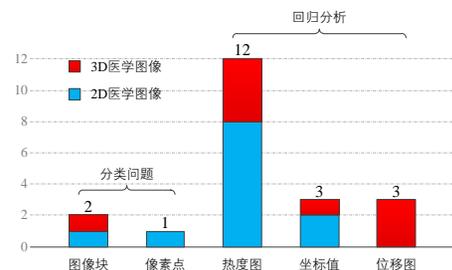


图3 医学图像关键点检测深度学习方法分类统计

注:其中将其作为分类问题的研究论文仅有3篇,远少于回归分析的解决方案,基于热度图的回归方法更是占据研究的重要位置

3.1 基于分类的关键点检测方法

分类问题(Classification Problem)是有监督学习中的核心问题,用于解决要预测样本属于哪个或者哪些预定义类别,此时输出变量通常取有限的离散值。如何将关键点定位问题转化为分类问题,现有的基于深度学习的医学图像关键点检测方法提出了两种问题解决思路(图4)。

3.1.1 基于像素点的分类检测方法

Tuysuzoglu 等人^[48]在研究直肠超声 2D 图像的前列腺关键点定位问题时,提出了一种对抗性的多任务深度学习方法,对超声图像中的每个像素分配 7 种类别的概率分布(Probability Distribution)特征向量,将每个关键点和背景分配单独的类别标签.分类网络学习像素到概率分布的映射,在映射后的特征图中找到每一类关键点标签的极大值作为定位位置.经 32 位直肠超声检查患者的 4799 幅图像实验测试,6 个关键点的平均定位误差为 3.56 mm.

3.1.2 基于图像块的关键点检测方法

准确检测和识别骨盆解剖学关键点是诊断髋关节发育不良(Developmental Dysplasia of Hip, DDH)的关键步骤. Liu 等人^[9]提出一种用于盆骨 X-Ray 2D 图像的 FR-DDH 关键点检测网络,将关键点检测任务转换为关键点局部邻域图像块的分类问题,以最匹配的检测区域中心作为关键点坐标.通过含有 9813 例盆骨 X-Ray 图像的数据集验证了 FR-DDH 关键点检测网络的关键点定位(平均误差为 1.24 mm)精度.与其相似的一份研究工作, Zheng 等人^[10]针对头颈 CT 扫描中的颈动脉分

叉关键点检测问题,首先由浅层网络完成所有体素为中心的图像块筛选以获取少量候选区域,然后使用深层网络结合 Haar 小波等特征进行更准确的后续图像块分类,在 455 例患者的头颈部 CT 数据集上进行了颈动脉分叉检测的定量评估,平均误差降低到 2.64 mm 的定位精度.

3.2 基于回归的关键点检测方法

回归分析(Regression Analysis)是确定两种或两种以上变量之间相互依赖的定量关系的统计分析方法.回归分析侧重从定量关系的分析直接输出实数数值,而分类处理的输出通常为若干指定的类别标签.基于回归分析的关键点检测方法,依据学习网络回归输出数据类型的不同,可分为坐标值回归(Coordinate Regression)、热度图回归(Heat-Map Regression)和位移图回归(Displacement Regression)三种不同类别,图 5 给出了基于回归分析的关键点检测框架.回归网络将整幅图像或者图像块作为数据输入,由不同的 Ground Truth 设置不同的网络输出,对热度图和位移图一般再经过后处理获得关键点的坐标.

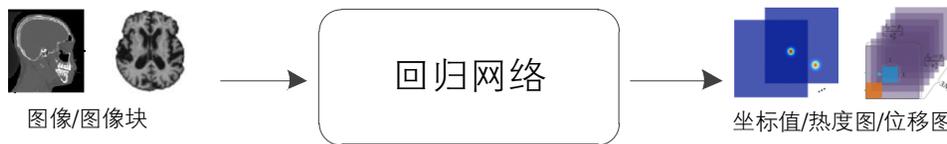


图5 基于回归分析的关键点检测框架

注:一般将整幅图像或者图像块作为网络输入,通过网络回归输出关键点的坐标值、热度图或位移图

3.2.1 基于坐标值回归的检测方法

深度学习网络通过输入的医学图像回归出关键点坐标是一种最直接解决思路,对于给定的含有 L 个关键点的 3D 图像 X_n ,网络输出层一般设计为 1 个或多个全连接层,通过端到端的(End-to-End)训练方式直接回归出 $3L$ 长度的向量,即为 L 个关键点的坐标(或归一化的坐标).

Andermatt 等人^[49]在研究 3D MRI 图像中髓脑沟关键点定位问题时,构建了由三个下采样 MD-GRU 层、全连接层和 Tanh 激活函数层组成的定位网络,网络输出经全连接层和 LReLU 层回归为关键点的坐标数据.经 1218 例图像的训练和测试,该方法的平均定位误差为 1.70 mm,与神经病学专家标注精度相当. Tiulpin 等人^[16]借助堆叠沙漏网络(Stacked Hourglass Network)完成输入图像到特征表示,由 2D Soft-Max 层回归每一个关键点的坐标.对膝盖骨 X-Ray 图像中 16 个关键点定位测试结果,关键点正确估计比例(Percentage of Correct Key-Points, PCK)参数在定位偏差为 2.50 mm 时达到 90.91%. Zhang 等人^[14]提出了两阶段的、面向任务的

深度学习网络(Two-stage Task-Oriented Deep Learning, T²DL)实现颅脑 3D T1W-MRI 图像关键点的自动检测.第一阶段采用基于 CNN 的回归模型使用数百万个图像块作为网络输入,旨在学习局部图像斑块和目标解剖学关键点之间的空间距离关系.第二阶段进一步建模图像块之间的相关性,与第一阶段 CNN 共享相同的网络结构和权重直接回归 $3L$ 长度的向量即 L 个关键点的空间坐标,同时增加额外卷积层实现大规模关键点检测.

3.2.2 基于热度图回归的检测方法

关键点检测的最终任务是输出预测关键点位置的坐标,然而直接通过学习网络输出坐标进行优化学习是一个极其非线性的过程,而且损失函数对权重的约束会比较弱,因此, Tompson 等人^[42]提出了构造一个中间态热度图作为网络回归输出,再经过非极大值抑制(Non-Maximum Suppression, NMS)等算法寻找并确定关键点的坐标.

一般地对于给定 L 个关键点,将坐标为 $\mathbf{x}_i \in \mathbb{R}^d$ 目标关键点 $L_i (i=1, 2, \dots, L)$ 的 d 维热度图 $g_i(\mathbf{x}): \mathbb{R}^d \rightarrow \mathbb{R}$ 定义

为高斯函数,计算式为

$$g_i(\mathbf{x}; \sigma_i) = \frac{\gamma}{(2\pi)^{d/2} \sigma_i^d} \exp\left(-\frac{\|\mathbf{x} - \hat{\mathbf{x}}_i\|_2^2}{2\sigma_i^2}\right) \quad (1)$$

因此,目标关键点坐标附近的热度图像素具有较高的值,并在远离关键点的位置平滑而迅速减小.引入比例因子 γ 可以有效避免回归网络训练期间由于高斯函数中微小值而引起的不稳定.对于每个维度 d ,标准偏差 σ_i 定义了关键点 L_i 的热度图中高斯函数的峰宽.在网络推断时,获取热度图中最高值的坐标并将其作为每个关键点 L_i 的预测坐标 $\hat{\mathbf{x}}_i \in \mathbb{R}^d$,计算式为

$$\hat{\mathbf{x}}_i = \arg \max_{\mathbf{x}} h_i(\mathbf{x}; w, b) \quad (2)$$

其中,预测热度图 $h_i(\mathbf{x}; w, b)$ 中参数 w 和 b 表示网络权重和偏置参数.

Payer等人^[50]提出一种全卷积空间配置网络(Spatial Configuration-Net, SCN)架构,局部外观模块(Local Appearance Model)映射局部外观特征回归生成候选关键点热度图,空间配置模块(Spatial Configuration Model)着重于减少形状歧义以提高对关键点错误识别的鲁棒性,在数量有限的手部2D/3D医学图像上获得了良好的关键点定位性能.相关工作经作者进一步完善后发表在医学图像处理顶级期刊Medical Image Analysis上^[51].更多相似的方法可参考表1列出的文献.

表1 基于热度图回归分析的关键点检测研究方法

文献编号	网络模型	数据类型	应用对象	方法特点描述
[25]	3D U-Net	CT	耳蜗	由FCN回归左右内耳关键点热度图,结合假阳性抑制以正确检测左、右内耳关键点
[4]	U-Net	X-Ray	肋骨	利用FCN回归16个关键点热度图,经条件随机场评估关键点间的空间信息完成局部细化
[52]	U-Net	X-Ray	颅骨	由粗略到精细两个阶段回归19个关键点的热度图,利用阶段间的Attention-Guide机制指导推理的扩展探索策略
[27]	VGG-19	X-Ray	颅骨	结合VGG-19和特征金字塔融合模块回归19个关键点热度图,经通过逐像素回归投票得出坐标
[53]	Flat-Net	MRI矢状位断层图像	声道	设计独特的Flat-Net网络结构用于回归MRI矢状位断层2D图像中21个关键点的热度图
[22]	DI2IN	CT	脊柱	结合传统Encoder-Decoder网络和监督网络设计DI2IN网络回归12个关键点热度图,经稀疏表示优化定位精度
[23]	SHN	高清手术图像	手术器具	采用堆叠沙漏网络回归高清图像中手术器具的3个关键点热度图,经关键点定位用于机器人手术控制
[24]	U-Net	X-Ray, 手术图像	血管、手术钳	采用U-Net回归关键点热度图,结合新颖的自适应损失函数设计,根据训练统计信息更新目标精度
[54]	FCN	超声	腹部	采用Skip-Connection全卷积网络回归14个关键点热度图,与视图分类组成多任务学习模型提高检测精度

3.2.3 基于位移图回归的检测方法

关键点定位任务中的位移图一般定义为与关键点间位移参数的特定数据形式,具有 V 体素的3D图像 \mathbf{X}_n 表示从该体素到特定轴空间中某个关键点的位移.也就是说, \mathbf{X}_n 中的第 l 个关键点有3个位移图(即 D_n^{lx} 、 D_n^{ly} 和 D_n^{lz}),分别对应于 x 、 y 和 z 轴三个方向上的偏移.从而给定 L 个关键点,每个输入图像 \mathbf{X}_n 则都有 $3L$ 个位移图.

在解决颌面部CBCT图像多个关键点检测问题中,Zhang等人^[6]提出了上下文指导的全卷积网络(Fully Convolutional Network, FCN)用于联合颅骨骨骼分割和关键点数字化,首先训练FCN-1学习图像空间体素相对于关键点的位移图以捕获CBCT图像空间上下文信息,然后结合原始图像经多任务的FCN-2网络以共同执行骨骼分割和关键点检测.在此工作基础上,Zhang

等人^[7]增加了更多的实验测试工作并将相关内容发表在Medical Image Analysis期刊,15个关键点的平均定位误差为 1.10 ± 0.71 mm.

在研究胎儿3D超声图像关键点自动检测问题时,Li等人^[55]提出了一种新颖的基于图像块的迭代网络(Patch-based Iterative Network, PIN),网络学习图像块与关键点之间的空间位移关系,推理时使用迭代的、稀疏的采样方法将图像块引导至目标关键点位置.实验采用72例人工标注图像进行网络的训练和性能测试,获得了 5.47 ± 4.23 mm的定位误差.

3.3 定位评价指标与常用数据集

为了验证关键点检测的准确性,常用的点对点误差(Point-to-point Error for Landmark, PEL)衡量参数定义为

$$PEL(L_i) = \left(\sum_{i=1}^N \|\hat{\mathbf{x}}_i - \mathbf{x}_i\| \right) / N \quad (3)$$

其中, N 表示测试图像数量, $\mathbf{x}_i \in \mathbb{R}^d$ 表示标记的关键点坐标, $\hat{\mathbf{x}}_i \in \mathbb{R}^d$ 表示网络推断识别结果. 将 PEL 的平均值定义为平均点对点误差 (Average Point-to-Point Errors, APE), 其计算式为

$$\begin{aligned} APE(L) &= \frac{\sum_{i=1}^L PEL(L_i)}{L} \\ &= \frac{\sum_{i=1}^L \sum_{j=1}^N \|\hat{\mathbf{x}}_j - \mathbf{x}_j\|}{NL} \end{aligned} \quad (4)$$

其中, L 表示每幅图像的关键点总数. 为了衡量点对点误差的分散程度, 一般在上述两衡量参数后面加入对应的标准差, 变为 $PEL(L_i) \pm \text{Std}(\text{mm})$ 和 $APE(L_i) \pm \text{Std}(\text{mm})$ 形式.

关键点成功检测率 (Successful Detection Rate, SDR) 是另外一个常用的定位精度评价指标, 定义为关键点 L_i 位于一系列定位精度范围 $\Delta =$

$\{1.0 \text{ mm}, 2.0 \text{ mm}, 3.0 \text{ mm}, \dots\}$ 内的百分比, 计算式为

$$SDR(L_i) = \# \frac{\{i: \|\hat{\mathbf{x}}_i - \mathbf{x}_i\| \leq \Delta\}}{N} \times 100\% \quad (5)$$

定位精度范围 Δ 中的精度值可以依据实际适当调整.

利用深度学习进行医学图像关键点检测时, 获取大规模的学习训练样本数据集非常困难, 且需要临床专家标注, 因此目前公开可用的医学图像关键点检测数据集偏少, 为了解决数据集的限制, 可以从两个方面尝试解决大规模的医学数据标注样本数据的问题.

一方面, 医学图像领域的挑战赛及大型公开数据集的出现对医学图像分析的发展有着极大的帮助, 自 2007 年以来, MICCAI、ISBI 和 SPIE 等医学成像研讨会组织挑战赛数据集已经成为惯例, 开放了大量用于基准研究的医学数据集^[56], 收录在网站 <http://www.grand-challenge.org/>. 另外通过追踪相关研究文献中研究者们给出的数据集, 表 2 给出本文整理的可用于医学图像关键点检测的图像库及其链接.

表 2 医学图像关键点检测可用数据集及参数

数据集	类型	规模	链接
Digital Hand Atlas Dataset	X-Ray	895	https://ipilab.usc.edu/research/baaweb/
Cephalometric Dataset	X-Ray	400	http://www-o.ntust.edu.tw/~cweiwang/
Pathological Spine CT Dataset	CT	242	https://biomedica.doc.ic.ac.uk/data/spine/
DDH Dataset	X-Ray	9813	https://github.com/liuboss1992/FR-DDH
Indiana Chest X-ray Dataset	X-Ray	7470	https://openi.nlm.nih.gov/
Osteoarthritis Initiative Dataset	X-Ray	748	https://oai.epi-ucsf.org/datarelease/
Instrument Tracking Dataset	可见光	860	https://endovissub-instrument.grand-challenge.org/

另一方面, 临床医学数据的收集为深度学习提供了另一个解决方案. 2016 年我国科技部已经启动国家重点研发计划“精准医学研究”专项, 可望在解决医学图像隐私问题的同时, 为深度学习在医学图像处理领域的研究提供可用的大规模医疗数据集. 目前, 本文项目组承担的国家自然科学基金项目“三维多模态融合图像诊治颞下颌关节紊乱病的关键技术研究”进展顺利, 通过与北京大学口腔医学院和解放军 306 医院合作, 以临床医生标注的颞下颌关节 CBCT 图像以及 MRI 图像两种模态对应的 5 组关键点信息为金标准训练学习的样本集, 结合热度图回归搭建了带有注意力机制的端到端的 3D FCN 关键点深度学习检测框架, 以 103 对 CBCT 图像进行测试, 初步实验结果 $APE = 2.13 \pm 1.84 \text{ mm}$. 在进一步数据整理的基础上, 计划将其发布并作为一份 3D 医学关键点检测的公开数据集.

4 挑战与对策

深度学习技术以自动学习和应用多级语义特征, 很大程度上克服了传统方法在特征定义和提取中的局

限性, 但基于深度学习的医学图像分析 (特别的针对研究相对较少的医学图像关键点检测) 上依然存在巨大挑战^[55]. 在上述研究现状分析的基础上, 本文尝试将基于深度学习的医学图像关键点检测面临的挑战总结为以下三点: ①通常只有数量有限的且带有医生标注的医学图像数据集可用, 难以处理学习模型训练过程中的过拟合问题; ②医学关键点高精度的检测需要综合应用医学图像多种信息, 设计能够挖掘和综合应用多种信息的深度学习网络具有较大的难度; ③医学图像 (特别是 3D 医学图像) 数据量过大, 规模巨大的网络模型参数对 GPU 内存提出了更高的要求, 同时这也对关键点检测的实时性带来了更大的挑战. 那么为了提高特征表示能力和关键点定位的准确性以满足临床的实际应用, 当标注数据集样本量不足时该怎么处理? 如何利用关键点的空间上下文信息 (Spatial Context Information) 以及与医学图像处理的其他任务间的互补关系? 如何降低医学图像数据量大对计算空间和时间的压力? 目前, 主要的应对策略如下文所述.

4.1 医学图像标注数据不足的对策

数据是深度学习算法研究所需的核心资源,深度学习方法在多个自然图像处理任务中的成功很大程度上归功于高达百万级别图像库的支撑,但医学图像由于疾病病例稀缺等,因此可获取的图像数量普遍偏少(一般在几十至几百数量级),而更高质量的医学图像标注将耗费大量人力和时间,因此在医学影像领域获取大量且具有高可靠性的标注数据是基于深度学习的医学图像处理研究的首要挑战。

针对这一挑战,Urschler等人^[26]将关键点定位任务分为两个更简单的子问题,以减少对大型训练数据集的总体需求,实验结果证实了即使在可用训练图像数量有限的情况下,也可获得2D和3D医学图像关键点的良好定位性能.Zhang等人^[6,7]采用两阶段的、面向任务的深度学习网络实现脑部图像关键点检测,同时使用数百万个图像块作为网络输入回归局部图像块和关键点之间的空间距离关系以进一步降低有限训练数据的影响。类似地,Li等人^[55]在网络训练时利用卷积神经

络学习图像块与解剖学关键点之间的空间关系,推理时PIN网络使用迭代地、稀疏地采样方法将图像块引导至目标关键点位置,实验采用72例人工标注图像进行网络的训练和性能测试,获得了 5.47 ± 4.23 mm的定位误差。

从以上研究思路来看,并没有出现类似于处理自然图像数据量不足问题中常用的迁移学习、数据增广或者生成式对抗网络(Generative Adversarial Networks, GAN)样本生成等技术^[57],本文认为这正是由医学图像关键点检测中数据集特有的需求所决定,特征点标记是位于图像空间中一个精确的坐标向量,而非类似与目标分割、分类等问题中区域的标记。文献^[50,51]倾向于将复杂的关键点定位问题分解为多个子问题,以降低有限数据训练大型网络的难度。文献^[36,58]结合了特殊应用中的医师注视点信息用于替代标注,这种方法与精确的数据标注还是存在较大的偏差,在一些对定位精度要求相对偏低的场合是一种选择。而文献^[55]提出的PIN模型的迭代优化思路(图6)最为新颖,值得研究者参考和学习。

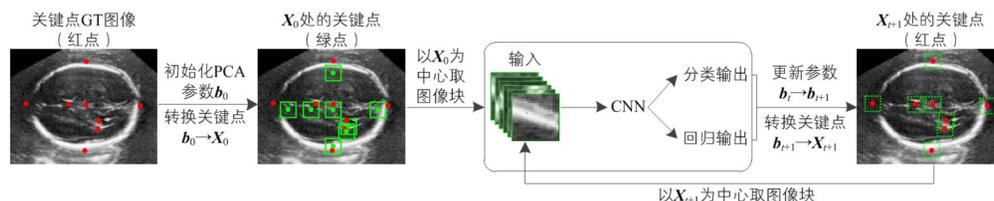


图6 文献[55]提出的多个关键点检测的迭代更新网络结构

4.2 深度学习网络设计的对策

深度学习算法的设计归根结底是适用于特定应用场景需求的CNN网络架构的设计。如何定义网络的“适用性”? 本文认为,网络能够有效挖掘数据中可用于表示待解决特定问题信息的能力,是衡量网络适用性的唯一标准。针对这一问题,建议从三个角度思考:①人体组织存在较大的近似性,这就决定了关键点的空间信息具有相当固定的关系,形成了特有的空间上下文信息,在设计提取关键点局部信息的同时应更加重视空间关系的全局信息;②采用更加符合人类认知过程的多阶段关键点检测网络,将关键点检测任务划分为从粗略到精细的推理过程,以提高检测的正确性和精准度;③合理结合医学图像处理的其他任务,设计多任务处理的学习网络。

4.2.1 空间上下文信息应用

关键点位置的图像亮度信息是最直接可用的信息,除此之外,医学图像关键点之间的空间位置分布(也称为空间上下文信息)一般具有相对稳定且相对统一的特性。这对医学图像关键点的检测起到显著的帮助作用。同时也有相关研究表明,即使对专家注释者/

临床医师而言,空间上下文信息在手工标注具有挑战性的医学关键点过程中也是必不可少的,特别是在图像信号或线索很少的区域,将关键点的拓扑/空间先验信息整合到检测任务中是一个活跃的研究领域且有着广泛的应用。Zhang等人^[25]使用低维形状模型捕获内耳对之间的空间关系,并使用此先验信息进一步评估了后处理步骤中检测到的内耳对的合理性。Liu等人^[9]挖掘盆骨关键点空间局部相关性巧妙地将检测任务转换为局部邻域图像块的分类问题。Tuysuzoglu等人^[48]利用待检测的6个解剖关键点均位于前列腺边界上为已知的拓扑/空间先验信息,网络显式地学习关键点标志性位置特征,实现了通过使用空间上下文告知地标位置来改善边界不明确的区域的检测性能。Mader等人^[4]利用条件随机场(Conditional Random Field, CRF)规范化建模肋骨关键点间的空间关系,最后在局部子图上优化关键点的定位推断。

4.2.2 多阶段学习网络架构设计

图像处理中的很多算法都符合人类由粗到细的认知过程的视觉机制,更加适合图像的变换信息处理,如多分辨率处理等。对于医学图像关键点检测问题,多个

基于深度学习方法的研究同样表明,由粗到细的检测技术显示了更好的解剖学关键点检测与定位的精度. Zhong 等人^[52]提出了两阶段注意导向的深度回归模型(Attention-Guided Deep Regression Model, AGDRM)关键点检测框架, Andermatt 等人^[49]提出了两阶段多维门控循环单元(Multi-Dimensional Gated Recurrent Units, MD-GRUs)网络. Zheng 等人^[10]采用了浅层网络和深层网络相结合的检测网络. Chen 等人^[27]提出了结合特征提取模块、注意力特征金字塔融合(Attentive Feature Pyramid Fusion, AFPF)模块和预测模块实现关键点端到端检测的深度框架. 这些研究成果表明,多阶段的关键点检测框架可以更加有效地提高关键点的定位精度.

4.2.3 多任务学习网络框架设计

多任务学习(Multi-Task Learning)是一种基于共享表示(Shared Representation)技术将多个相关的任务综合在一起学习的机器学习方法,充分利用任务之间所富含的关联信息,提升单任务学习网络的泛化(Network Generalization)性能. 研究者们将医学图像关键点检测任务与其相关联的诸如分割等医学任务相结合,展示了优异的检测性能.

Zhang 等人^[6,7]提出了一个上下文指导的全卷积网络(FCN)用于联合颅骨骨骼分割和关键点数字化处理两个任务. Duan 等人^[15]将心脏磁共振(Cardiac Magnetic Resonance, CMR)图像双心室分割与关键点检测相结合构建多任务学习网络——同步分段和地标本地化网络(Simultaneous Segmentation and Landmark Localization Network, SSLN),以此网络输出配合地图集传播实现具有解剖学意义的双心室分割. Tuysuzoglu 等人^[48]基于解剖学关键点均位于光滑封闭的前列腺边界这一先验知识,提出了一种多任务学习网络,在学习标志性位置的同时,还建立了学习前列腺轮廓的机制,通过预测每

个关键点位置之外的完整边界轮廓,以增强整体网络的上下文感知能力并提高关键点的检测性能.

在医学图像关键点检测问题的研究中关于深度学习网络模型相关的研究成果最多,设计多阶段、多任务的网络结构,同时将关键点空间上下文信息引入网络学习内容之中,用来提升医学关键点检测的精准度,这正体现了网络更深层、更全面地学习和挖掘医学图像中与关键点有关的信息是提升性能的唯一途径的思路,同时也存在网络参数调优的需求,以进一步改善检测任务的性能^[24].

4.3 医学图像计算量大的对策

利用小型数据集中训练诸如 CNN 等网络是一项艰巨的任务,另外医学图像数据量过大易于造成网络学习的过拟合问题,再者也会需要更大容量的 GPU 存储方可实现网络的训练和推理. 因此,难以以端到端的方式利用有限的医学成像数据实现准确的医学关键点检测模型的训练. 如何有效解决医学图像对深度学习算法带来的训练与计算困难也是研究者们无法回避的问题.

一方面,通过多种图像降采样方式降低图像数据量的大小. Li 等人^[55]在特定点 x 、 y 和 z 三个方向(或者说医学图像的横断面、矢状位和冠状位三个断层方向)上各提取一幅二维图像,将其叠加在一起组成三通道的 2D 图像作为网络输入,实验验证了该方法能提供与全 3D 图像相近的识别性能. Duan 等人^[15]从心脏 CMR 三维体数据中选择位于基底部(basal)、中部(mid-cavity)和心尖部(apical)轴位上的三个切片作为多通道矢量图像(2.5D)作为网络数据输入. Yang 等人^[22]从 x 、 y 、 z 轴方向将 3D 图像转换为三组 2D 图像,分别对每个轴采用 CNN 分类方式实现 2D 图像包含有特定关键点的概率分布. 三种方法如图 7 所示.

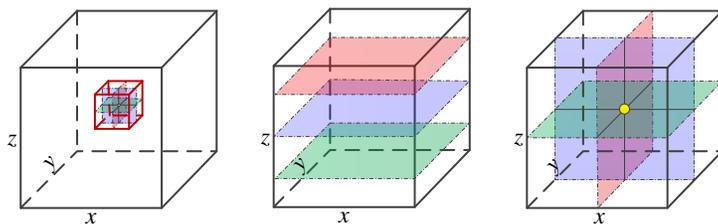


图7 三种不同的3D医学图像抽取方式可以有效降低医学图像数据量大对网络计算带来的压力

另一方面,可选择更为轻量级的深度学习网络作为特征提取的主干网络,以解决使用类似 V-Net 的体系结构对大量关键点检测时遇到的内存问题. Mader 等人^[4]从深度学习网络结构简化入手,采用参数更少的卷积姿态机(Convolutional Pose Machines, CPM)神经网络体系结构; Probst 等人^[23]为解决手术

导航立体视觉中 full-HD 图像(分辨率为 1920 像素 × 1080 像素)中的工具钳尖端 3 个关键点的快速检测,提出了利用堆叠沙漏网络将关键点检测问题转化为热度图回归的任务. Tiulpin 等人^[16]采用相同的网络结构解决膝部 X 线图像中解剖关键点自动定位的问题.

医学图像(特别是3D医学图像)数据量巨大是造成网络模型参数占用动辄几百兆字节甚至上G字节的存储空间,而网络训练阶段的误差反向传播更是带来所占GPU内存的成倍增大. 现有的方法分别从降低网络输入数据和降低网络规模两个方面入手:文献[15, 22, 55]充分利用了医学图像的特点,通过多种数据抽取降维方式实现数据量的大幅降低,从而得以利用通用的GPU计算卡完成原本不可完成的任务,并通过实验验证了数据抽取处理并未明显降低关键点检测精度;文献[4, 16, 23]则选用参数量较U-Net网络更少的CPM、SHN网络解决问题.

5 开放的研究方向思考与探讨

深度学习有效促进了不同尺度和不同任务信息之间的融合,使得信息的结合方式由平面开始向立体方法发展,对于医学图像关键点检测模型的发展具有突出的实际意义. 正因为如此,本文对医学图像关键点检测的深度学习进行了梳理和综述. 从本文筛选出的在此研究领域具有特别贡献的论文研究成果可知,研究者们结合各自的医学应用需求开展的特定医学图像上的关键点深度学习检测算法中,大多有意识地去解决医学标注数据少、医学图像数据量大对CNN网络带来的训练与预测中的问题,并取得了良好的检测与定位性能(需要注意的是,由于各自任务目标、数据等不具有统一对比性,这里并未给出关键点定位精度的统计).

尽管已经取得了良好的检测效果,但进一步提高医学解剖关键点的检测精度并将实现研究成果的临床应用还有较长的路要走. 本文作者结合自身在这个领域的研究经验对其存在的关键难题和开放的研究方向给出一些思考和讨论:①医学图像关键点个体间差异性和个体内歧义性的固有特征无法回避,在这种情况下要实现高精度医学关键点检测,就需要建立一个充分挖掘医学图像关键点信息并能够综合应用信息的最优深度学习网络框架,同时考虑可用的小规模医学图像数据集对网络规模的限制;②深度学习只是解决问题的一种方法,同样存在各种各样的局限性,在很好地把握深度学习方法优劣势的基础上,能够结合几十年累积起来的经典的图像处理方法和思路不失为明智的选择,同时还要关注能对医学图像关键点检测带来启发的其他计算机视觉、机器学习领域的新成果.

6 结束语

医学图像关键点检测是一个十分重要的研究领域,具有重要的研究价值和广泛的应用前景. 本文对近

年来医学图像关键点检测的深度学习研究方法进行了分类梳理和详细综述,在所面临的挑战问题和研究趋势方面,本文亦抛砖引玉,希望为相关科研人员进一步深入了解医学图像关键点检测问题并开展相关研究尽微薄之力.

参考文献

- [1] HANAOKA S, SHIMIZU A, NEMOTO M, et al. Automatic detection of over 100 anatomical landmarks in medical CT images: A framework with independent detectors and combinatorial optimization[J]. *Medical Image Analysis*, 2017, 35: 192-214.
- [2] GE J W, SAEIDI H, OPFERMANN J D, et al. Landmark-guided deformable image registration for supervised autonomous robotic tumor resection[C]//*Medical Image Computing and Computer Assisted Intervention-MICCAI 2019*. Cham, GER: Springer, 2019: 320-328.
- [3] PETERLÍK I, COURTECUISSÉ H, ROHLING R, et al. Fast elastic registration of soft tissues under large deformations[J]. *Medical Image Analysis*, 2018, 45: 24-40.
- [4] MADER A O, BERG J, FABRITZ A, et al. Localization and labeling of posterior ribs in chest radiographs using a CRF-regularized FCN with local refinement[C]//*Medical Image Computing and Computer Assisted Intervention-MICCAI 2018*. Cham, GER: Springer, 2018: 562-570.
- [5] TOROSDAGLI N, LIBERTON D K, VERMA P, et al. Deep geodesic learning for segmentation and anatomical landmarking[J]. *IEEE Transactions on Medical Imaging*, 2019, 38(4): 919-931.
- [6] ZHANG J, LIU M, WANG L, et al. Joint craniomaxillofacial bone segmentation and landmark digitization by context-guided fully convolutional networks[J]. *Medical Image Computing and Computer-Assisted Intervention*, 2017, 10434: 720-728.
- [7] ZHANG J, LIU M X, WANG L, et al. Context-guided fully convolutional networks for joint craniomaxillofacial bone segmentation and landmark digitization[J]. *Medical Image Analysis*, 2020, 60: 101621.
- [8] ZHANG J, SAHA A, ZHU Z, et al. Hierarchical convolutional neural networks for segmentation of breast tumors in MRI with application to radiogenomics[J]. *IEEE Transactions on Medical Imaging*, 2019, 38(2): 435-447.
- [9] LIU C B, XIE H T, ZHANG S C, et al. Misshapen pelvis landmark detection by spatial local correlation mining for diagnosing developmental dysplasia of the hip[C]//*Medical Image Computing and Computer Assisted Intervention-*

- MICCAI 2019. Cham, GER: Springer, 2019: 441-449.
- [10] ZHENG Y F, LIU D, GEORGESCU B, et al. 3D deep learning for efficient and robust landmark detection in volumetric data[C]//Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Cham, GER: Springer, 2015: 565-572.
- [11] AUBERT B, VAZQUEZ C, CRESSON T, et al. Toward automated 3D spine reconstruction from biplanar radiographs using CNN for statistical spine model fitting[J]. IEEE Transactions on Medical Imaging, 2019, 38(12): 2796-2806.
- [12] NOOTHOUT J M H, VOS B D D, WOLTERINK J M, et al. CNN-based landmark detection in cardiac CTA scans[EB/OL]. [2018]. <https://arxiv.org/abs/1804.04963>.
- [13] GREWAL M, DEIST T M, WIERSMA J, et al. An end-to-end deep learning approach for landmark detection and matching in medical images[EB/OL]. [2020]. <https://arxiv.org/abs/2001.07434>.
- [14] ZHANG J, LIU M X, SHEN D G. Detecting anatomical landmarks from limited medical imaging data using two-stage task-oriented deep neural networks[J]. IEEE Transactions on Image Processing, 2017, 26(10): 4753-4764.
- [15] DUAN J M, BELLO G, SCHLEMPER J, et al. Automatic 3D Bi-ventricular segmentation of cardiac images by a shape-refined multi-task deep learning approach[J]. IEEE Transactions on Medical Imaging, 2019, 38(9): 2151-2164.
- [16] TIULPIN A, MELEKHOW I, SAARAKKALA S. Knee anatomical landmark localization using hourglass networks[EB/OL]. [2019]. <https://arxiv.org/abs/1907.12237>.
- [17] IBRAGIMOV B, LIKAR B, PEMUS F, et al. Automatic cephalometric x-ray landmark detection by applying game theory and random forests[C]//Proceedings of ISBI International Symposium on Biomedical Imaging. Cham, GER: Springer, 2014: 1-8.
- [18] LINDNER C, COOTES T F. Fully automatic cephalometric evaluation using random forest regression-voting[C]//Proceedings of ISBI International Symposium on Biomedical Imaging. Cham, GER: Springer, 2015: 1-8.
- [19] WANG C W, HHANG C T, HSIEH M C, et al. Evaluation and comparison of anatomical landmark detection methods for cephalometric X-ray images: A grand challenge[J]. IEEE Transactions on Medical Imaging, 2015, 34(9): 1890-1900.
- [20] WANG C W, HUANG C T, LEE J H, et al. A benchmark for comparison of dental radiography analysis algorithms [J]. Medical Image Analysis, 2016, 31: 63-76.
- [21] BIER B, GOLDMANN F, ZAECH J N, et al. Learning to detect anatomical landmarks of the pelvis in X-rays from arbitrary views[J]. International Journal of Computer Assisted Radiology and Surgery, 2019, 14(9): 1463-1473.
- [22] YANG D, XIONG T, XU D G, et al. Automatic vertebra labeling in large-scale 3D CT using deep image-to-image network with message passing and sparsity regularization [C]//Information Processing in Medical Imaging. Cham, GER: Springer, 2017: 633-644.
- [23] PROBST T, MANINIS K K, CHHATKULI A, et al. Automatic tool landmark detection for stereo vision in robot-assisted retinal surgery[J]. IEEE Robotics and Automation Letters, 2018, 3(1): 612-619.
- [24] TEIXEIRA B, TAMERSOY B, SINGH V, et al. Adaloss: Adaptive loss function for landmark localization[EB/OL]. [2019]. <http://arxiv-export-lb.library.cornell.edu/pdf/1908.01070>.
- [25] ZHANG D Q, WANG J N, NOBLE J H, et al. HeadLocNet: Deep convolutional neural networks for accurate classification and multi-landmark localization of head CTs[J]. Medical Image Analysis, 2020, 61: 101659.
- [26] URSCHLER M, EBNER T, ŠTERN D. Integrating geometric configuration and appearance information into a unified framework for anatomical landmark localization [J]. Medical Image Analysis, 2018, 43: 23-36.
- [27] CHEN R N, MA Y X, CHEN N L, et al. Cephalometric landmark detection by attentive feature pyramid fusion and regression-voting[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2019. Cham, GER: Springer, 2019: 873-881.
- [28] LÉVY-MANDEL A D, VENETSANOPOULOS A N, TSOTSOS J K. Knowledge-based landmarking of cephalograms[J]. Computers and Biomedical Research, 1986, 19(3): 282-309.
- [29] EL-FEGHI I, SID-AHMED M A, AHMADI M. Automatic localization of craniofacial landmarks for assisted cephalometry[J]. Pattern Recognition, 2004, 37(3): 609-621.
- [30] CARDILLO J, SID-AHMED M A. An image processing system for locating craniofacial landmarks[J]. IEEE Transactions on Medical Imaging, 1994, 13(2): 275-289.
- [31] COOTES T F, TAYLOR C J, COOPER D H, et al. Active shape models-their training and application[J]. Computer Vision and Image Understanding, 1995, 61(1): 38-59.
- [32] COOTES T F, EDWARDS G J, TAYLOR C J. Active appearance models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(6): 681-685.

- [33] YUE W, YIN D, LI C, et al. Automated 2-D cephalometric analysis on X-ray images by a model-based approach [J]. *IEEE Transactions on Biomedical Engineering*, 2006, 53(8): 1615-1623.
- [34] LINDNER C, WANG C W, HUANG C T, et al. Fully automatic system for accurate localisation and analysis of cephalometric landmarks in lateral cephalograms[J]. *Scientific Reports*, 2016, 6: 33581.
- [35] ARIK S Ö, IBRAGIMOV B, XING L. Fully automated quantitative cephalometry using convolutional neural networks[J]. *Journal of Medical Imaging (Bellingham, Wash.)*, 2017, 4(1): 014501.
- [36] DROSTE R, CAI Y F, SHARMA H, et al. Ultrasound image representation learning by modeling sonographer visual attention[C]//*Information Processing in Medical Imaging*. Cham, GER: Springer, 2019: 592-604.
- [37] LITJENS G, KOOI T, BEJNORDI B E, et al. A survey on deep learning in medical image analysis[J]. *Medical Image Analysis*, 2017, 42: 60-88.
- [38] 田娟秀, 刘国才, 谷珊珊, 等. 医学图像分析深度学习方法研究与挑战[J]. *自动化学报*, 2018, 44(3): 401-424.
TIAN J X, LIU G C, GU S S, et al. Deep learning in medical image analysis and its challenges[J]. *Acta Automatica Sinica*, 2018, 44(3): 401-424. (in Chinese)
- [39] WELLS W M III. Medical Image Analysis - past, present, and future[J]. *Medical Image Analysis*, 2016, 33: 4-6.
- [40] SHEN D, WU G, SUK H I. Deep learning in medical image analysis[J]. *Annual Review of Biomedical Engineering*, 2017, 19: 221-248.
- [41] BOOKSTEIN F L. *Morphometric Tools for Landmark Data*[M]. Cambridge, UK: Cambridge University Press, 1992.
- [42] TOMPSON J, JAIN A, LECUN Y, et al. Joint training of a convolutional network and a graphical model for human pose estimation[C]//*Proceedings of the 27th International Conference on Neural Information Processing Systems*. New York, USA: ACM, 2014: 1799-1807.
- [43] WU Y, JI Q. Facial landmark detection: A literature survey[J]. *International Journal of Computer Vision*, 2019, 127(2): 115-142.
- [44] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [45] DANG Q, YIN J Q, WANG B, et al. Deep learning based 2D human pose estimation: A survey[J]. *Tsinghua Science and Technology*, 2019, 24(6): 663-676.
- [46] LI R, LIU Z Y, TAN J R. A survey on 3D hand pose estimation: Cameras, methods, and datasets[J]. *Pattern Recognition*, 2019, 93: 251-272.
- [47] CHENG W H, SONG S J, CHEN C Y, et al. Fashion meets computer vision[J]. *ACM Computing Surveys*, 2021, 54(4): 1-41.
- [48] TUYSUZOGLU A, TAN J, EISSA K, et al. Deep adversarial context-aware landmark detection for ultrasound imaging[C]//*Medical Image Computing and Computer Assisted Intervention-MICCAI 2018*. Cham, GER: Springer, 2018: 151-158.
- [49] ANDERMATT S, PEZOLD S, AMANN M, et al. Multi-dimensional gated recurrent units for automated anatomical landmark localization[EB/OL].[2017].<https://arxiv.org/abs/1708.02766>.
- [50] PAYER C, ŠTERN D, BISCHOF H, et al. Regressing heatmaps for multiple landmark localization using CNNs [C]//*Medical Image Computing and Computer-Assisted Intervention-MICCAI 2016*. Cham, GER: Springer, 2016: 230-238.
- [51] PAYER C, ŠTERN D, BISCHOF H, et al. Integrating spatial configuration into heatmap regression based CNNs for landmark localization[J]. *Medical Image Analysis*, 2019, 54: 207-219.
- [52] ZHONG Z S, LI J, ZHANG Z X, et al. An attention-guided deep regression model for landmark detection in cephalograms[C]//*Medical Image Computing and Computer Assisted Intervention-MICCAI 2019*. Cham, GER: Springer, 2019: 540-548.
- [53] ESLAMIM, NEUSCHAEFER-RUBEC, Serrurier A. Automatic vocal tract landmark localization from midsagittal MRI data[EB/OL].[2020].<https://arxiv.org/abs/1907.07951>.
- [54] XU Z B, HUO Y K, PARK J, et al. Less is more: Simultaneous view classification and landmark detection for abdominal ultrasound images[C]//*Medical Image Computing and Computer Assisted Intervention-MICCAI 2018*. Cham, GER: Springer, 2018: 711-719.
- [55] LI Y W, ALANSARY A, CERROLAZA J J, et al. Fast multiple landmark localisation using a patch-based iterative network[C]//*Medical Image Computing and Computer Assisted Intervention-MICCAI 2018*. Cham, GER: Springer, 2018: 563-571.
- [56] GREENSPAN H, VAN GINNEKEN B, SUMMERS R M. Guest editorial deep learning in medical imaging: Over-

view and future promise of an exciting new technique[J]. IEEE Transactions on Medical Imaging, 2016, 35(5): 1153-1159.

- [57] SHORTEN C, KHOSHGOFTAAR T M. A survey on image data augmentation for deep learning[J]. Journal of Big Data, 2019, 6(1): 1-48.
- [58] DROSTE R, CHATELAIN P, DRUKKER L, et al. Discovering salient anatomical landmarks by predicting human gaze[C]//2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). Iowa City, IA, USA: IEEE, 2020: 1711-1714.

作者简介



李居朋 男,1981年8月出生于江苏省沛县.2009年获得北京交通大学博士学位.2019年赴北卡罗来纳大学教堂山分校研修,现为北京交通大学副教授、研究生导师,主要研究方向为医学图像分割及多模态配准等.
E-mail:lijupeng@bjtu.edu.cn



王颖慧 女,1992年1月出生于河北省保定市.2018年毕业于首都医科大学长学制(本硕连读)口腔医学专业.现为北京大学口腔医学院口腔颌面医学影像学在读博士研究生,主要研究方向为数字化口腔颌面医学影像.



李 刚 男,1970年2月出生于黑龙江省哈尔滨市.2004年于瑞典卡罗林斯卡医学院获博士学位.现为北京大学教授、博士生导师,中国电子学会会员,中华口腔医学会口腔颌面放射专委会副主任委员,口腔医学院医学影像科副主任,主要研究方向为数字化口腔医学影像的基础和临床研究等.
E-mail:kqgang@bjmu.edu.cn