



(12) 发明专利申请

(10) 申请公布号 CN 111914925 A

(43) 申请公布日 2020. 11. 10

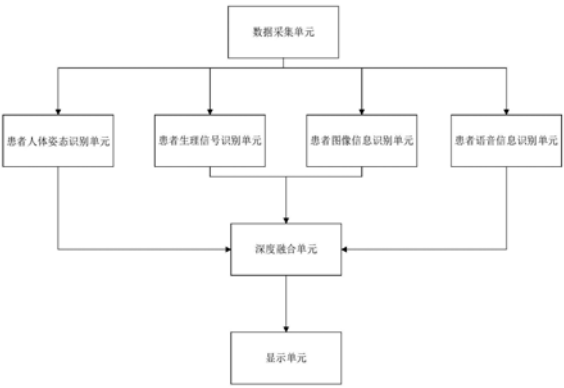
(21) 申请号 202010740442.5  
(22) 申请日 2020.07.28  
(71) 申请人 复旦大学  
地址 200433 上海市杨浦区邯郸路220号  
(72) 发明人 张立华 杨鼎康 翟鹏 董志岩  
(74) 专利代理机构 上海科盛知识产权代理有限公司 31225  
代理人 赵继明  
(51) Int. Cl.  
G06K 9/62 (2006.01)  
G06N 3/04 (2006.01)  
G06N 3/08 (2006.01)  
G06K 9/00 (2006.01)  
G06K 9/46 (2006.01)  
A61B 5/11 (2006.01)

A61B 5/00 (2006.01)  
A61B 5/0488 (2006.01)  
A61B 5/0402 (2006.01)  
A61B 5/0496 (2006.01)  
A61B 5/0205 (2006.01)  
A61B 5/055 (2006.01)  
A61B 5/053 (2006.01)  
A61B 6/00 (2006.01)  
A61B 6/03 (2006.01)  
A61B 8/00 (2006.01)

权利要求书2页 说明书9页 附图6页

(54) 发明名称  
一种基于深度学习的患者行为多模态感知与分析系统  
(57) 摘要

本发明涉及一种基于深度学习的患者行为多模态感知与分析系统,包括数据采集单元、患者人体姿态识别单元、患者生理信号识别单元、患者图像信息识别单元、患者语音信息识别单元以及深度融合单元;通过对采集到的患者姿态、生理、图像和语音等多模态数据实现预处理和感兴趣区提取和诊断,深度融合单元采用多模态二维特征和三维特征融合网络结构,由2D深度学习网络获取初步分割结果,并通过3D深度学习网络在初步分割结果的基础上,获取患者行为检测结果。与现有技术相比,本发明实现了更加精准的评估患者行为,准确定位病灶,显著提高了患者病理走向预测准确度,为患者行为科学干预与医疗流程智能优化手段的实施提高有力的基础保障。



1. 一种基于深度学习的患者行为多模态感知与分析系统,其特征在于,包括数据采集单元、患者人体姿态识别单元、患者生理信号识别单元、患者图像信息识别单元、患者语音信息识别单元以及深度融合单元,所述数据采集单元用于获取多模态患者数据,所述数据采集单元分别连接所述患者人体姿态识别单元、患者生理信号识别单元、患者图像信息识别单元和患者语音信息识别单元;所述深度融合单元分别连接所述患者人体姿态识别单元、患者生理信号识别单元、患者图像信息识别单元和患者语音信息识别单元;

所述深度融合单元采用2D深度学习网络获取初步分割结果,并通过3D深度学习网络在所述初步分割结果的基础上,获取患者行为检测结果。

2. 根据权利要求1所述的一种基于深度学习的患者行为多模态感知与分析系统,其特征在于,所述2D深度学习网络和3D深度学习网络均采用带残差结构的ResUNet构建。

3. 根据权利要求1所述的一种基于深度学习的患者行为多模态感知与分析系统,其特征在于,所述深度融合单元的数据处理过程包括以下步骤:

2D深度学习网络处理步骤:2D深度学习网络对多维特征数据进行处理,并将处理结果由2D特征转换为3D特征,获取第一3D特征图;

3D深度学习网络处理步骤,将所述第一3D特征图与多维特征数据合并,一同输入到3D深度学习网络中,得到第二3D特征图;

融合处理步骤:将所述第一3D特征图和第二3D特征图进行求和计算后,依次输入2D+3D融合层、卷积计算层和分类预测层,获取3D的患者行为检测结果。

4. 根据权利要求3所述的一种基于深度学习的患者行为多模态感知与分析系统,其特征在于,所述2D深度学习网络的输入端通过四个特征通道分别对所述患者人体姿态识别单元、患者生理信号识别单元、患者图像信息识别单元和患者语音信息识别单元传输的多维特征数据进行处理。

5. 根据权利要求3所述的一种基于深度学习的患者行为多模态感知与分析系统,其特征在于,所述第一3D特征图包括模态信息分割概率图和背景分割概率图,所述3D深度学习网络的输入端通过六个特征通道分别对所述模态信息分割概率图、背景分割概率图以及患者人体姿态识别单元、患者生理信号识别单元、患者图像信息识别单元和患者语音信息识别单元传输的多维特征数据。

6. 根据权利要求1所述的一种基于深度学习的患者行为多模态感知与分析系统,其特征在于,所述患者人体姿态识别单元的执行过程包括以下步骤:

角度空间特征表示步骤,用于根据人体单元节点的骨架数据,提取患者姿态的角度空间特征;

节点动量特征表示步骤,用于通过将目标节点抽象为物理质点的方式,获取患者姿态的节点动量特征;

第一模态输出步骤,用于根据所述患者姿态的角度空间特征和节点动量特征,通过混淆矩阵,提取患者骨骼数据的空间运动信息,实现模态输出。

7. 根据权利要求1所述的一种基于深度学习的患者行为多模态感知与分析系统,其特征在于,所述患者生理信号识别单元的执行过程包括以下步骤:

外周生理信号收集步骤,用于通过医学影像学技术设备采集外周生理信号;

噪声消除预处理步骤,用于通过滤波器滤波和小波变换,对所述外周生理信号进行预

处理；

外周生理特征提取步骤，用于对预处理后的所述外周生理信号进行特征提取；

第二模态输出步骤，用于根据所述特征提取的结果，实现模态输出。

8. 根据权利要求1所述的一种基于深度学习的患者行为多模态感知与分析系统，其特征在于，所述患者图像信息识别单元包括依次连接的图像信息采集模块、图像预处理模块和配准模块，所述图像预处理模块首先对彩色图像通过分量法和加权平均法实现灰度化；然后进行图像的几何变换，并按照变换关系进行灰度差值，最后结果图像增强技术，完成图像的预处理。

9. 根据权利要求1所述的一种基于深度学习的患者行为多模态感知与分析系统，其特征在于，所述患者语音信息识别单元包括依次连接的语音信息采集模块、语音预处理模块和信息特征提取模块，所述语音预处理模块用于通过加窗和端点检测的方式对语音信息进行预处理。

10. 根据权利要求1所述的一种基于深度学习的患者行为多模态感知与分析系统，其特征在于，所述患者行为多模态感知与分析系统还包括显示模块，该显示模块连接所述深度融合单元。

## 一种基于深度学习的患者行为多模态感知与分析系统

### 技术领域

[0001] 本发明涉及患者行为分析领域,尤其是涉及一种基于深度学习的患者行为多模态感知与分析系统。

### 背景技术

[0002] 随着深度学习技术的不断发展,在处理许多单模态的感知型机器学习任务中,深度神经网络相对传统的信息处理方法,取得了很大的优势和信息处理效果。例如循环神经网络和递归神经网络(RNN)的提出,对于患者病历文本信息和语音信息的序列问题的处理,实现了极为成功的工程化推广和医疗诊断的应用;AlexNet、ResNet等模型的提出,在患者行为视频信息领域的任务处理上甚至超越了人类的表现。

[0003] 将深度学习技术应用到模式识别领域进行各个模态的特征提取和选择,最常用于特征提取的深度网络是卷积神经网络(Convolutional neural network,CNN),它由多个单层卷积神经网络进行多次堆叠而成。单层卷积神经网络一般包括卷积、非线性变换和下采样三个阶段,CNN经常被用于提取视觉模态的特征,对于其他模态的特征也可使用CNN进行特征提取。在神经语言模型的基础上,大量的深度神经网络结构被改良并进一步应用于自然语言处理任务,如RNN、LSTM被广泛地应用于文本分类、实体识别等任务。由于RNN能够出色地学习序列样本中不同时刻的信息及其相互关系,RNN结构在机器翻译、对话生成等序列分析及序列生成任务中的优势极为突出。

[0004] 面向急诊、ICU、护理、隔离病房或代谢舱等复杂医疗场景下,基于深度学习的患者行为的多维度感知算法传统的缺点是无法有效感知患者的精细行为和细粒度,对基于患者行为的医疗行为实施的合规性无法做出准确判断,同时目前对于患者行为的分析和研究,大多数医院和医疗数据中心依然停留在人工样本采集分析和自动化的单模态分析阶段,尽管部分机构已经展开了对多模态数据的感知分析,但是缺乏对多模态数据的兼容性处理和考量,很大程度上制约了患者行为和下一步的医疗结局的数据分析和研究。

[0005] 现有的深度学习方法大多数应用于患者的图像信息处理方面,对多模态数据的处理依然存在网络设计复杂,训练速度慢以及梯度问题无法解决等缺陷,无法很好的发挥多模态信息融合可以使单模态异构特征以及多模态异构特征在时间和空间线索维度上互为补充的优势。例如虽然基于图像信息的特征提取中,RGB的表现纹理信息丰富,但却容易受到光强干扰,较难以表达时序线索;虽然基于三维人体动作捕捉系统得到的人体位姿惯性数据时序变化线索较容易提取和表达,但是却难以表现纹理信息等。

[0006] 上述问题导致患者行为采集样本的位置不准确,延迟了对患者行为的评估和治疗,或者出现过度诊断以及过度治疗。

### 发明内容

[0007] 本发明的目的就是为了克服上述现有技术存在的缺陷而提供一种更加准确可靠的基于深度学习的患者行为多模态感知与分析系统。

[0008] 本发明的目的可以通过以下技术方案来实现：

[0009] 一种基于深度学习的患者行为多模态感知与分析系统，包括数据采集单元、患者人体姿态识别单元、患者生理信号识别单元、患者图像信息识别单元、患者语音信息识别单元以及深度融合单元，所述数据采集单元用于获取多模态患者数据，所述数据采集单元分别连接所述患者人体姿态识别单元、患者生理信号识别单元、患者图像信息识别单元和患者语音信息识别单元；所述深度融合单元分别连接所述患者人体姿态识别单元、患者生理信号识别单元、患者图像信息识别单元和患者语音信息识别单元；

[0010] 所述深度融合单元采用2D深度学习网络获取初步分割结果，并通过3D深度学习网络在所述初步分割结果的基础上，获取患者行为检测结果。

[0011] 进一步地，所述2D深度学习网络和3D深度学习网络均采用带残差结构的ResUNet构建。

[0012] 进一步地，所述深度融合单元的数据处理过程包括以下步骤：

[0013] 2D深度学习网络处理步骤：2D深度学习网络对多维特征数据进行处理，并将处理结果由2D特征转换为3D特征，获取第一3D特征图；

[0014] 3D深度学习网络处理步骤，将所述第一3D特征图与多维特征数据合并，一同输入到3D深度学习网络中，得到第二3D特征图；

[0015] 融合处理步骤：将所述第一3D特征图和第二3D特征图进行求和计算后，依次输入2D+3D融合层、卷积计算层和分类预测层，获取3D的患者行为检测结果。

[0016] 进一步地，所述2D深度学习网络的输入端通过四个特征通道分别对所述患者人体姿态识别单元、患者生理信号识别单元、患者图像信息识别单元和患者语音信息识别单元传输的多维特征数据进行处理。

[0017] 进一步地，所述第一3D特征图包括模态信息分割概率图和背景分割概率图，所述3D深度学习网络的输入端通过六个特征通道分别对所述模态信息分割概率图、背景分割概率图以及患者人体姿态识别单元、患者生理信号识别单元、患者图像信息识别单元和患者语音信息识别单元传输的多维特征数据。

[0018] 进一步地，所述患者人体姿态识别单元的执行过程包括以下步骤：

[0019] 角度空间特征表示步骤，用于根据人体单元节点的骨架数据，提取患者姿态的角度空间特征；

[0020] 节点动量特征表示步骤，用于通过将目标节点抽象为物理质点的方式，获取患者姿态的节点动量特征；

[0021] 第一模态输出步骤，用于根据所述患者姿态的角度空间特征和节点动量特征，通过混淆矩阵，提取患者骨骼数据的空间运动信息，实现模态输出。

[0022] 进一步地，所述患者生理信号识别单元的执行过程包括以下步骤：

[0023] 外周生理信号收集步骤，用于通过医学影像学技术设备采集外周生理信号；

[0024] 噪声消除预处理步骤，用于通过滤波器滤波和小波变换，对所述外周生理信号进行预处理；

[0025] 外周生理特征提取步骤，用于对预处理后的所述外周生理信号进行特征提取；

[0026] 第二模态输出步骤，用于根据所述特征提取的结果，实现模态输出。

[0027] 进一步地，所述患者图像信息识别单元包括依次连接的图像信息采集模块、图像

预处理模块和配准模块,所述图像预处理模块首先对彩色图像通过分量法和加权平均法实现灰度化;然后进行图像的几何变换,并按照变换关系进行灰度差值,最后结果图像增强技术,完成图像的预处理。

[0028] 进一步地,所述患者语音信息识别单元包括依次连接的语音信息采集模块、语音预处理模块和信息特征提取模块,所述语音预处理模块用于通过加窗和端点检测的方式对语音信息进行预处理。

[0029] 进一步地,所述患者行为多模态感知与分析系统还包括显示模块,该显示模块连接所述深度融合单元。

[0030] 与现有技术相比,本发明具有以下优点:

[0031] (1) 本发明通过结合采集单元、患者人体姿态识别单元、患者生理信号识别单元、患者图像信息识别单元、患者语音信息识别单元、深度融合单元以及显示单元,实现患者不同模态异构数据的充分采集,同时依据不同模态数据的特点在识别单元实现数据的预处理和筛查配准,融合单元所述的深度融合卷积神经网络模型不仅仅学到患者行为本身的时空特征,同时包含其空间纹理特征,实现更加精准的评估患者行为,准确定位病灶,显著提高患者病理走向预测准确度,为患者行为科学干预与医疗流程智能优化手段的实施提高有力的基础保障。

[0032] (2) 本发明深度融合单元通过利用多模态数据的二维和三维特征实现综合的多模态融合,大大提高了模型的泛化能力和患者行为的识别预测准确度,解决了跨模态数据时空异构性与患者行为多维度感知准确性需求的矛盾,为下一步实现低延时,高精度,兼顾不同医疗场景的智慧患者行为感知和医疗流程优化提供客观依据,提高了患者诊治效率和水平。

[0033] (3) 本发明数据采集单元充分应用医疗场景下的医学影像等技术和传感器技术结合,采集了患者行为的多模态数据信息,同时区别传统的分析方法,分别应用不同识别单元的相关原理和技术对模态数据实现了进一步的处理和特征提取。

[0034] (4) 患者语音信息识别单元的语音预处理模块在通过加窗、端点检测等方式对语音信号进行预处理,一方面为了消除患者自身不可避免的伪影干扰,另一方面尽可能的减少医疗音频采集设备产生的环境噪声,提高患者语音质量,为后续的特征提取提供保障。

## 附图说明

[0035] 图1为本发明基于深度学习的患者行为多模态感知与分析系统的原理框图;

[0036] 图2为本发明提供的患者人体姿态识别单元的原理框图;

[0037] 图3为本发明提供的患者生理信号识别单元的原理框图;

[0038] 图4为本发明提供的患者图像信息识别单元的原理框图;

[0039] 图5为本发明提供的患者语音信息识别单元的原理框图;

[0040] 图6为本发明深度融合单元采用的基于深度融合卷积神经网络的多模态特征融合流程示意图。

## 具体实施方式

[0041] 下面结合附图和具体实施例对本发明进行详细说明。本实施例以本发明技术方案

为前提进行实施,给出了详细的实施方式和具体的操作过程,但本发明的保护范围不限于下述的实施例。

#### [0042] 实施例1

[0043] 如图1所示,本实施例提供一种基于深度学习的患者行为多模态感知与分析系统,包括数据采集单元、患者人体姿态识别单元、患者生理信号识别单元、患者图像信息识别单元、患者语音信息识别单元、深度融合单元以及显示模块,数据采集单元用于获取多模态患者数据,数据采集单元分别连接患者人体姿态识别单元、患者生理信号识别单元、患者图像信息识别单元和患者语音信息识别单元;深度融合单元分别连接患者人体姿态识别单元、患者生理信号识别单元、患者图像信息识别单元、患者语音信息识别单元和显示模块。

[0044] 数据采集单元通过医疗场景下的医学影像技术提供的多传感器设备和自建的数据库获取多模态患者数据。

[0045] 深度融合单元主要包括2D-ResUNet结构、3D-ResUNet结构以及2D+3D融合层。所述的多模态2D-ResUNet结构主要包括四种模态的信息编码器、多模态卷积结构、解码器以及跳跃连接结构;所述的3D-ResUNet结构用于接收2D结构的分割概率信息,实现多维的多特征融合。

[0046] 显示模块用于输出通过深度融合模块后的测试集的患者行为检测结果和行为评估。

[0047] 下面对各部分进行详细描述。

#### [0048] 1、患者人体姿态识别单元

[0049] 如图2所示,患者人体姿态识别单元的执行过程包括以下步骤:

[0050] 角度空间特征表示步骤,用于根据人体单元节点的骨架数据,提取患者姿态的角度空间特征;

[0051] 节点动量特征表示步骤,用于通过将目标节点抽象为物理质点的方式,获取患者姿态的节点动量特征;

[0052] 第一模态输出步骤,用于根据患者姿态的角度空间特征和节点动量特征,通过混淆矩阵,提取患者骨骼数据的空间运动信息,实现模态输出。

[0053] 具体描述如下:

[0054] 1.1) 角度空间特征表示步骤将提取的25个人体单元节点的骨架数据转换为20个节点表示简化的人体骨架表示,计算整个数据集的平均值和标准偏差。通过标准化操作,令处理后的数据符合 $\mu=0, \sigma=1$ 的标准正态分布。对处理后的20个关节进行排列组合提取平面,结果将会产生 $C_{20}^3$ 种可能的平面。这些可能值通过如下公式表示:

[0055]  $C_p = \{P_i, P_j, P_k\}, p \in N^+[1, 1140], i, j, k \in N^+[1, 20]$

[0056] 式中, $C_p$ 为总可能平面集合, $P_i$ 为人体关节特征点*i*的取值概率, $P_j$ 为人体关节特征点*j*的取值概率, $P_k$ 为人体关节特征点*k*的取值概率, $N^+[1, 1140]$ 为特征因子的分布水平范围。

[0057] 此模块从这些点中选取三个点。可以得到所需要的向量,如下公式表示:

$$[0058] \quad \begin{cases} V_p(1) = (P_i^x - P_j^x, P_i^y - P_j^y, P_i^z - P_j^z) \\ V_p(2) = (P_i^x - P_k^x, P_i^y - P_k^y, P_i^z - P_k^z) \\ V_p(3) = (P_k^x - P_j^x, P_k^y - P_j^y, P_k^z - P_j^z) \end{cases}$$

[0059] 式中,  $V_p(1)$  为关节数据向量, 在笛卡尔坐标系参照下,  $P_i^x$  为x平面关节特征点i的取值概率,  $P_j^x$  为x平面关节特征点j的取值概率,  $P_k^x$  为x平面关节特征点z的取值概率。

[0060] 最后将提取后的数据特征描述子堆叠为  $F \times H \times W$  的张量表示形式, 其中F表示从序列中抽取的帧数, H, W分别为长度和宽度。

[0061] 1.2) 节点动量特征表示步骤将目标节点抽象为一个物理质点, 其质量为1, 则目标质点的动量可以表示为  $p=mv$ , 速度可以对质点单位时间运动的距离进行微分获得, 假设患者姿态质点的运动轨迹处处可微, 则表示如下:

$$[0062] \quad f'(t) = \lim_{\Delta t \rightarrow 0} \frac{y_{t+1} - y_{t-1}}{2\Delta t}$$

[0063] 之后将提取后的Mop堆叠为  $F \times H \times W$  的几何流表示, 其中F表示从序列中抽取的帧数, H, W分别为长度和宽度。

[0064] 1.3) 模态输出: 通过角度空间特征和节点动量特征表示患者姿态信息后, 在网络层构造批量归一化和激活函数, 建立混淆矩阵, 挖掘骨骼数据的空间运动信息, 实现模态输出。

[0065] 2、患者生理信号识别单元

[0066] 如图3所示, 患者生理信号识别单元通过医疗场景下的CT、B超等医学影像学技术设备采集到的呼吸、皮肤电等外周生理信号时, 不免会受到患者穿戴各种传感器设备产生的磁干扰和功率干扰, 同时在对患者的采集过程中, 患者会不自主的产生不可避免的各种肢体动作, 产生噪声和基线漂移等干扰。

[0067] 患者生理信号识别单元的执行过程包括以下步骤:

[0068] 外周生理信号收集步骤, 用于通过医学影像学技术设备采集外周生理信号;

[0069] 噪声消除预处理步骤, 用于通过滤波器滤波和小波变换, 对外周生理信号进行预处理;

[0070] 外周生理特征提取步骤, 用于对预处理后的外周生理信号进行特征提取;

[0071] 第二模态输出步骤, 用于根据特征提取的结果, 实现模态输出。

[0072] 具体描述如下:

[0073] 2.1) 利用滤波器和小波变换, 令外周信号更加纯净, 确保输出模态的识别准确性。此单元考虑到患者外周生理信号强度微弱、不同信号频率范围内的强度不同的特点, 利用滤波器对各种信号进行滤波, 截取强度大的频率范围的生理信号便于后续模态识别的进行。

[0074] 2.2) 本单元得到的患者有限生理信号的频率范围为, 脉搏信号有效频率范围  $0.5\text{Hz} \sim 20\text{Hz}$  之间, 肌肉电信号有效频率范围  $50\text{Hz} \sim 150\text{Hz}$  之间, 心电信号有效频率范围  $0\text{Hz} \sim 40\text{Hz}$  之间, 皮肤电信号有效频率范围  $0\text{Hz} \sim 2\text{Hz}$  之间。结合脉搏信号、心电信号和皮肤点信号的频率范围较小, 频率波动幅度较大的特点, 对于基线漂移的低频部分, 采用小波变换进行预处理, 将低频部分利用小波变换中的小波分解提出, 将提出以后剩余信号利用分组重



构实现组合得到预处理数据。

[0075] 2.3) 结合该单元提取外周生理信号主要有水平眼电图、垂直眼电图、颌骨肌电图、斜方肌肌电图、电流皮肤相应和呼吸振幅以及皮肤温度等。利用频域特征提取法在信号强度大的频率范围提取有效的特征,用于光谱功率、功率谱密度和功能谱能量等。结合过滤信号的频率范围,分别提取信号的平均数、中位数、标准偏差等特征。

[0076] 3、患者图像信息识别单元

[0077] 如图4所示,患者图像信息识别单元包括图像信息采集模块、图像预处理模块和配准模块三个方面。具体如下:

[0078] 3.1) 此单元通过患者疾病临床过程中提供的多种医学影像,例如,如计算机X线断层扫描)、磁共振成像、单光子发射计算机断层成像、超声图像、电阻抗图像等不同的医学图像实现对不同基于患者行为的模态信息的收集、多模态的采集患者的图像信息可以充分显示形态成像方法的高分辨力和精准的定位能力。力图尽可能的挖掘影像信息,用于患者抗法治疗的收束定位和放疗计划设计。

[0079] 3.2) 患者图像信息的质量好坏直接影响了识别算法的效果精度,因此在配准之前需要先对图像进行预处理。图像预处理模块首先对彩色图像通过分量法和加权平均法实现灰度化,以减少所需处理的数据量,提高模块整体的处理速度。之后进行图像的几何变换,通过平移、转置、镜像、旋转和缩放对采集的图像进行处理,用于改正图像采集模块的系统误差和仪器误差。此外,还利用灰度插值按照变换关系,避免输出图像的像素映射到输入图像非整数坐标的可能。最后结合图像增强技术,针对特点医疗场景的应用场合,有目的的强调图像的整体或者局部特性,扩大图像中不同患者特征之间的差别,抑制不感兴趣的特征,丰富信息量和图像的判读识别效果,满足后续分析的需要。

[0080] 3.3) 配准模块实现对图像的分割来提取特征。此模块直接依照图像感兴趣区域(ROI)中的生理特征进行分析,将这些特征与图像的边、轮廓、表面和跳跃性特征。例如角落、线的交叉点和高曲率点等。之后按照先验知识选择一定的分割阈值对图像进行自动分割,从而得到图像的几何特征。然后进行变换操作。将一幅图像中的坐标点变换到另一幅图像的坐标系中,经过图像分割后,求解两幅图像中对应点的变换。

[0081] 4、患者语音信息识别单元

[0082] 如图5所示,患者语音信息识别单元包含了语音信息采集、语音预处理和信息特征提取三个模块,具体步骤如下:

[0083] 4.1) 语音信息采集模块通过医疗诊断记录中所收集到的专项语音数据集合和丰富的通用问诊数据集实现患者行为语音数据的补充和完善,这些数据集包含了医疗场景下丰富的患者语音信息,涉及患者的问诊记录、患者治疗日常言语信息以及康复阶段的情绪言语化表现。

[0084] 4.2) 语音预处理模块中,由于患者的语音信号具有短时平稳性,此模块对信号进行分帧处理。紧接着进行加窗处理,加窗的目的是可以认为对抽样n附近的语音波形加以强调而对波形的其余部分加以减弱。对语音信号的各个短段进行处理,即对各个短段进行某种变换或施以某种运算。这里使用矩形窗,其定义为:

[0085] 
$$W(n) = \begin{cases} 1, & 0 < n \leq N - 1 \\ 0, & else \end{cases}$$

[0086] 式中,  $W(n)$  为矩形窗函数表达式,  $n$  为矩形窗序列长度,  $N$  为大于  $n+1$  的某一自然数, 具体用来规定  $n$  的上边界范围。

[0087] 这里的矩形窗主瓣较窄, 具有较高的频率分辨率。

[0088] 之后采用双门限比较法来进行端点检测。双门限比较法以短时能量  $E$  和短时平均过零率  $Z$  作为特征, 结合  $Z$  和  $E$  的优点, 使检测更为准确, 有效降低系统的处理时间, 能排除无声段的噪声干扰, 从而提高了语音信号的处理性能。

[0089] 4.3) 信息特征提取模块中通过对短时能量、短时平均幅度和短时平均过零率三个特征提取, 实现语音模态的输出。取一帧信号, 计算某帧的语音信号的短时能量, 之后实现语音信号的短时平均幅度, 定义如下:

$$[0090] \quad M(i) = \sum_{n=0}^{L-1} |y_i(n)|$$

[0091] 短时平均幅度是一帧语音信号能量大小的表征, 它与短时能量的区别在于计算时不论采样值的大小, 不会因为取二次方而造成较大的差异, 在患者行为语音信号处理会带来很大的好处。短时能量和短时平均幅度的有作用主要是: 作为区分清浊音, 区分声母韵母, 区分有话段和无话段的指标。短时平均过零率表示一帧语音中语音信号波形穿过横轴(零电平)的次数。其计算公式为:

$$[0092] \quad Z(i) = \frac{1}{2} \sum_{n=0}^{L-1} |\text{sgn}[y_i(n)] - \text{sgn}[y_i(n-1)]|$$

[0093] 式中,  $\text{sgn}[*]$  为阶跃函数,  $L$  为语音信号序列取值长度。

[0094] 由于浊音具有较低的过零率而清音具有较高的过零率, 过零率可以用来区分清音和浊音, 在端点检查中有一定的运用。

[0095] 5、深度融合单元

[0096] 深度融合单元采用2D深度学习网络获取初步分割结果, 并通过3D深度学习网络在初步分割结果的基础上, 获取患者行为检测结果。

[0097] 2D深度学习网络和3D深度学习网络均采用带残差结构的ResUNet构建。

[0098] 深度融合单元的数据处理过程包括以下步骤:

[0099] 2D深度学习网络处理步骤: 2D深度学习网络对多维特征数据进行处理, 并将处理结果由2D特征转换为3D特征, 获取第一3D特征图;

[0100] 3D深度学习网络处理步骤, 将第一3D特征图与多维特征数据合并, 一同输入到3D深度学习网络中, 得到第二3D特征图;

[0101] 融合处理步骤: 将第一3D特征图和第二3D特征图进行求和计算后, 依次输入2D+3D融合层、卷积计算层和分类预测层, 获取3D的患者行为检测结果。

[0102] 2D深度学习网络的输入端通过四个特征通道分别对所述患者人体姿态识别单元、患者生理信号识别单元、患者图像信息识别单元和患者语音信息识别单元传输的多维特征数据进行处理。

[0103] 第一3D特征图包括模态信息分割概率图和背景分割概率图, 3D深度学习网络的输入端通过六个特征通道分别对所述模态信息分割概率图、背景分割概率图以及患者人体姿态识别单元、患者生理信号识别单元、患者图像信息识别单元和患者语音信息识别单元传输的多维特征数据。

[0104] 如图6所示,具体实施时,本实施例深度融合单元采用基于深度融合卷积神经网络的多模态特征融合流程,包括2D-ResUNet结构(2D深度学习网络)、3D-ResUNet结构(3D深度学习网络)以及2D+3D融合层,整个模型的输入为多维特征,其中2D-ResUNet结构的特征通道数 $c=4$ 表示四种模态 $T_1, T_2, T_3, T_4$ 。此结构中假设函数 $T$ 表示3D特征转化为2D的过程, $T^{-1}$ 表示该转换的逆过程,则 $I_{2d}=T(I)$ ,得到四种模态2D网络表示为 $f_{2d}$ ,3D网络表示为 $f_{3d}$ ,则多模态2D经多模态2D-ResUNet后的特征和概率图分别如下

$$[0105] \quad F_{2d} = f_{2d}(I_{2d-T_1}, I_{2d-T_2}, I_{2d-T_3}, I_{2d-T_4}; \theta_{2d})$$

$$[0106] \quad F_{2d} \in R^{b \times 384 \times 384 \times 16}$$

$$[0107] \quad y_{2d} = f_{2dcls}(F_{2d}; \theta_{2dcls})$$

[0108] 其中的 $\theta_{2d}, \theta_{2dcls}$ 分别为卷积网络和预测网络的参数,卷积网络和预测网络是2D-ResUNet结构的组成部分,卷积网络层进行多模态数据的卷积运算;预测网络层将解码后的多模态特征进行提取,输出分类特征进行预测。为了将2D网络所得到的结果与3D网络相结合, $F_{2d}$ 与 $y_{2d}$ 需要经过以下变换,得到对应的3D特征图:

$$[0109] \quad \hat{F}_{2d} = T^{-1}(F_{2d}), \hat{F}_{2d} \in R^{1 \times 384 \times 384 \times b \times 16}$$

$$[0110] \quad \hat{y}_{2d} = T^{-1}(y_{2d}), \hat{y}_{2d} \in R^{1 \times 384 \times 384 \times b \times 2}$$

[0111] 为此将 $\hat{y}_{2d}$ 与 $I$ 合并,一同输入3D-ResUNet中,得到3D网络的特征图:

$$[0112] \quad F_{3d} = f_{3d}(I, \hat{y}_{2d}; \theta_{3d}), F_{3d} \in R^{1 \times 384 \times 384 \times b \times 16}$$

[0113] 通过求和变化后得到 $Z$ ,再输入2D+3D融合层 $f_{HF}$ ,进行卷积计算得到 $H$ ,在经过预测分类层 $f_{HFcls}$ ,得到3D的多模态行为预测结果 $y_H$ 。表示如下:

$$[0114] \quad Z = \hat{F}_{2d} + F_{3d}$$

$$[0115] \quad H = f_{HF}(Z; \theta_{HF})$$

$$[0116] \quad y_H = f_{HFcls}(H; \theta_{HFcls})$$

[0117] 2D-ResUNet和3D-ResUNet的网络层次结构采用带残差结构的ResUNet构建模型很好的解决了梯度消失问题,有效实现了模型的训练和测试。

[0118] 2D-ResUNet网络结构主要包括四种模态的编码器、多模态卷积结构、解码器以及编码器-解码器间的跳跃连接结构。编码器由带残差结构的卷积块与最大池化层组成,解码器由带残差结构的卷积块和反卷积层组成,将四种模态经编码器中的每一池化层进行池化,再将经池化后的四种特征图进行卷积运算,从而实现多模态2D特征的融合。具体实现过程为四种模态图像经过相同层次的卷积和池化后的特征图具有相同的尺寸,通过在深度的方向上对四种特征图进行合并,可生成深度为4的3D特征图,再通过大小为 $(1, 1, 4)$ 的卷积核以 $(1, 1, 4)$ 为步长对该特征图进行卷积,从而将特征图深度转化为1,实现四种模态特征的深度融合。经多模态卷积后的特征图,通过跳跃连接结构与解码器相同层次的特征图相加,从而减少反卷积过程中的信息丢失。

[0119] 所述的3D-ResUNet网络结构的输入为多模态2D-ResUNet的分割概率图 $y_{2d}$ 经变换所得到的概率图 $\hat{y}_{2d}$ 与模型的原始输入合并后的特征图;其中3D-ResUNet网络结构的输入为6通道,分别为四种模态以及多模态2D-ResUNet网络所得患者行为模态信息与背景两种区域分割概率图。将其输入到带有跳跃连接以及残差结构的3D-ResUNet网络中,即可实现

3D分割。因此,该网络将四种模态作为输入的四通道,将多模态2D-ResUNet分割概率图作为输入的二通道,从而将2D网络快速分割结果用于指导3D模型分割,实现了基于多模态的患者行为的准确感知和评估。

[0120] 以上详细描述了本发明的较佳具体实施例。应当理解,本领域的普通技术人员无需创造性劳动就可以根据本发明的构思做出诸多修改和变化。因此,凡本技术领域技术人员依本发明的构思在现有技术的基础上通过逻辑分析、推理或者有限的实验可以得到的技术方案,皆应在由权利要求书所确定的保护范围内。

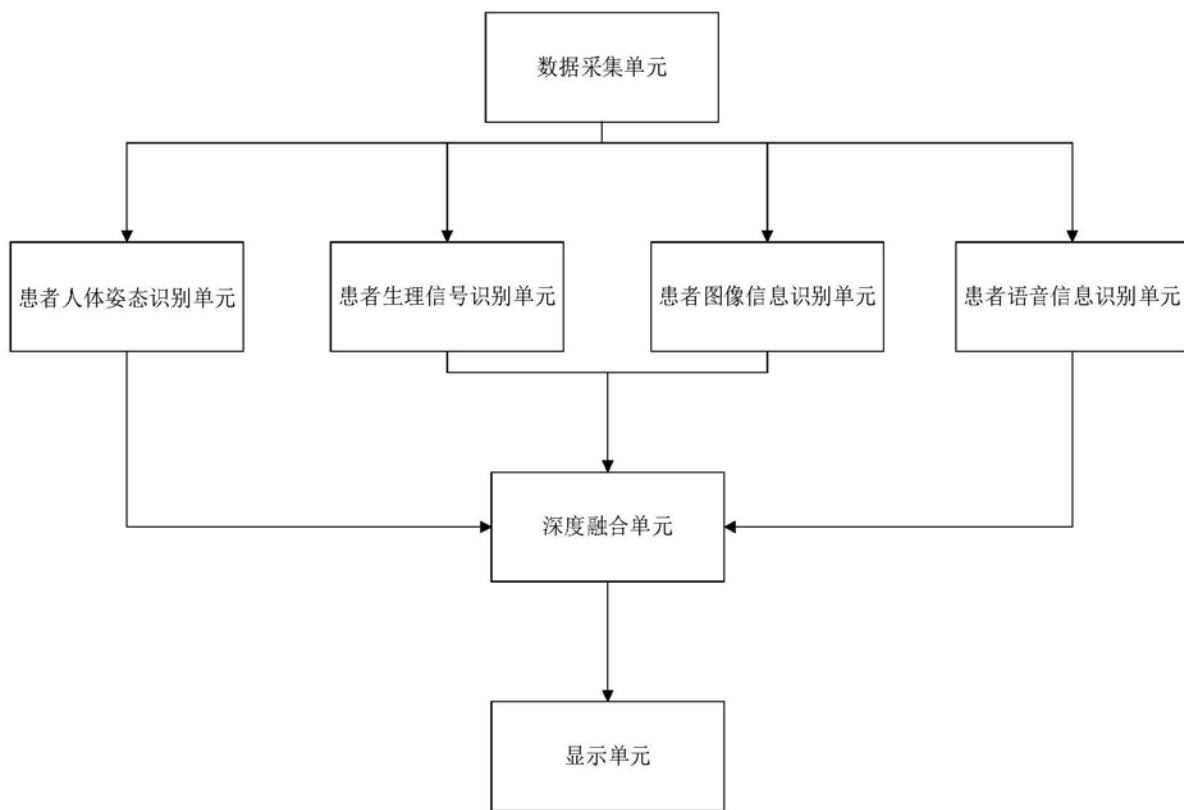


图1

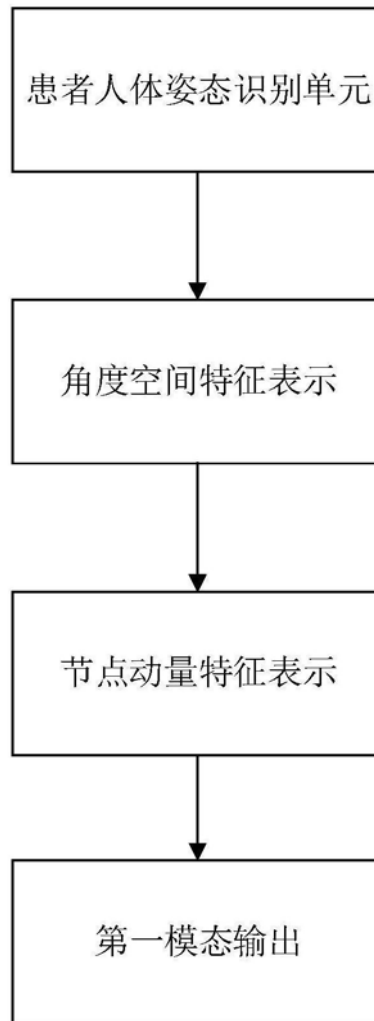


图2

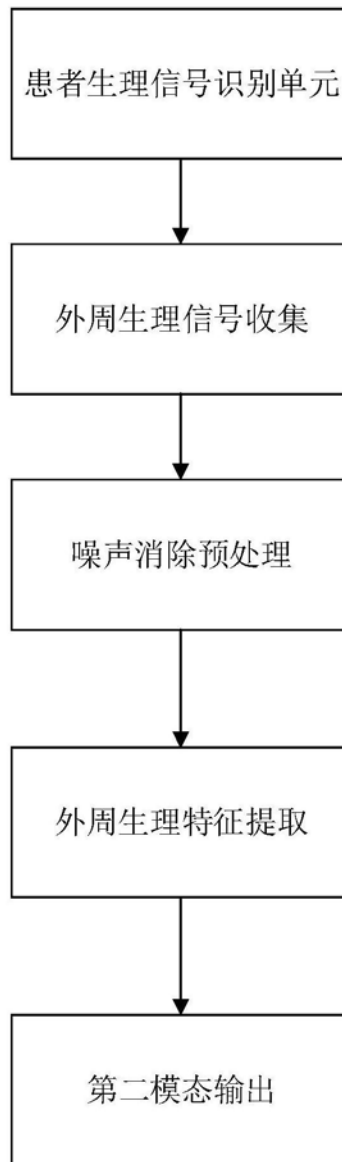


图3

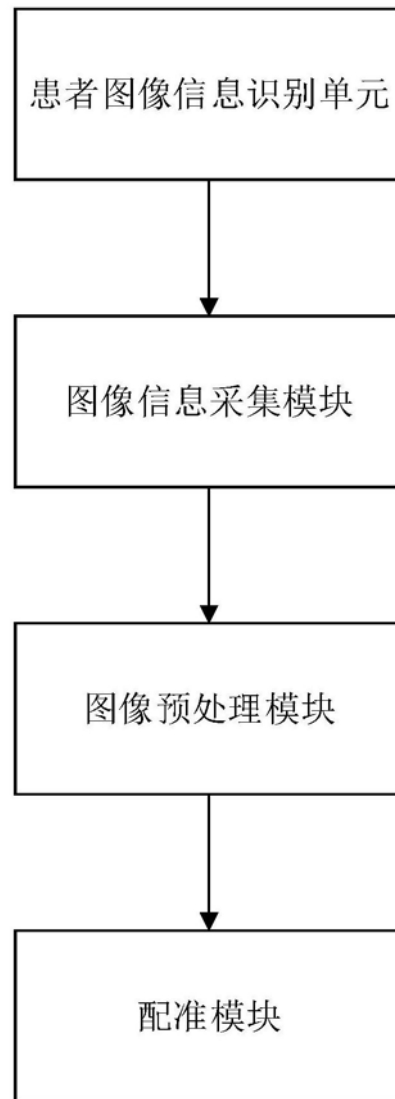


图4



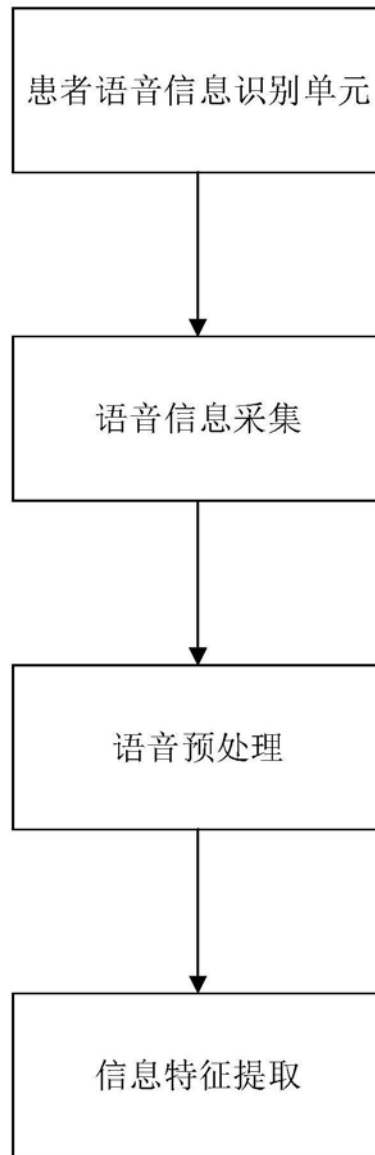


图5

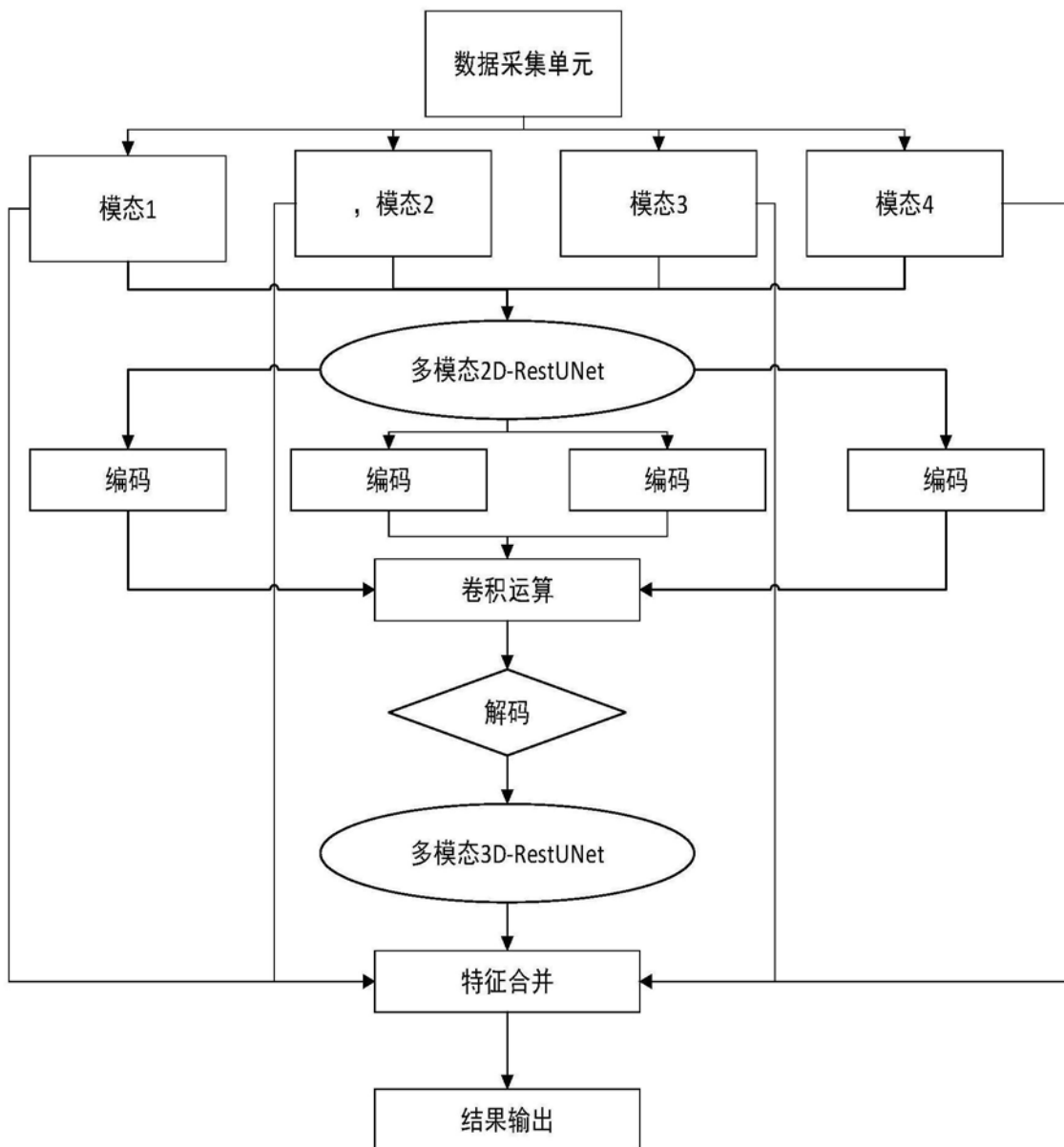


图6