



(12) 发明专利申请

(10) 申请公布号 CN 116359846 A

(43) 申请公布日 2023. 06. 30

(21) 申请号 202310236507.6

G06N 3/08 (2023.01)

(22) 申请日 2023.03.13

(71) 申请人 东南大学

地址 210096 江苏省南京市玄武区四牌楼2号

(72) 发明人 王帅 梅洛瑜 曹东江 史瑞签

(74) 专利代理机构 南京众联专利代理有限公司 32206

专利代理师 薛雨妍

(51) Int. Cl.

G01S 7/02 (2006.01)

G01S 13/88 (2006.01)

G06F 18/23 (2023.01)

G06F 18/25 (2023.01)

G06N 3/0499 (2023.01)

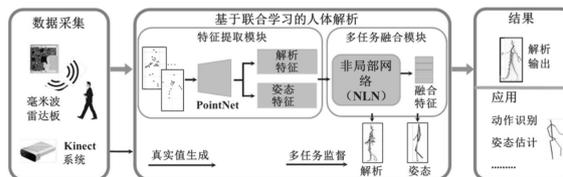
权利要求书4页 说明书9页 附图2页

(54) 发明名称

一种基于联合学习的动态毫米波雷达点云人体解析方法

(57) 摘要

本发明其所属技术领域为物联网感知,设计了一种基于联合学习的动态毫米波雷达点云人体解析方法,尤其是基于毫米波感知的人体解析。当前以人为中心的毫米波感知大多集中于动作识别、姿势估计等场景,但它们通常无法获取毫米波点云的语义信息,即无法分辨每个雷达点对应的身体部位信息,因此需要一种动态毫米波雷达点云人体解析方案。本方法的内容包括:对于毫米波点云数据,首先进行聚类,之后使用多任务学习模型联合执行人体解析和姿势估计任务进行特征提取,再通过非局部网络进行多任务特征融合,最终输出结果为带注释语义标签的点云。



1. 一种基于联合学习的动态毫米波雷达点云人体解析方法,其特征在於:包括以下步骤:

步骤1:通过多任务特征提取模块解决毫米波点云的稀疏性;采用多任务学习模型联合执行人体解析的主要任务和姿态估计的辅助任务;辅助任务可以有效引导人类解析网络提取代表受试者姿势的高级结构特征;由於人体姿势和人体解析之间具有强相关性,人体姿势相关特征有助于提高解析网络在预测语义标签时的准确性和鲁棒性;对于人类解析和姿态估计任务,多任务学习模型并行提取相应的特征;

步骤2:通过多任务特征融合模块解决镜面反射造成点云数据中身体部位缺失问题;从非局部网络中获得灵感,设计了一种多任务特征融合方法,结合了任务内注意和任务间注意机制,实现了从全局角度聚合主体的时空特征;

步骤3:多任务监督,在系统离线训练阶段,本发明采用Kinect系统,用来得到人体解析标签和姿态估计标签的真实值。

2. 根据一种基于联合学习的动态毫米波雷达点云人体解析方法,其特征在於:所述步骤1具体来说,多任务学习模型可分为包括点模块、帧模块和特征聚合模块,模型的输入为长度为 s 的帧序列,每一帧包含 n 个点,每个点包含 d 个特征维度;

步骤1.1:人体解析特征提取;

在点模块,对于 t 时刻对应帧的点集合 C_t 中的任一雷达点 $p_{i,t}$,采用多呈感知机得到点的高维特征表示,即点特征 $e_{i,t}^H$;其公式为:

$$e_{i,t}^H = MLP(p_{i,t}; \theta_e)$$

其中 θ_e 表示MLP的可学习参数, H 表示人体解析任务;

在帧模块,对于每个雷达点 $p_{i,t}$ 的点特征 $e_{i,t}^H$,首先编码为更高维度的特征表示形式 $h_{i,t}^H$;其公式为:

$$h_{i,t}^H = MLP(e_{i,t}^H; \theta_h)$$

其中 θ_h 表示MLP的可学习参数;

步骤1.2:人体姿态特征提取;

对于人体姿态特征的提取,使用与人体解析特征提取略微不同的网络架构;具体来说,对于任一帧的帧特征 g_t^P ,使用长短期记忆网络进行处理;其公式为:

$$r_t^P = LSTM(g_t, r_{t-1}; \theta_r)$$

其中 θ_r 表示LSTM的参数;

最后,将 r_t^P 与人体姿态任务的点特征 $e_{i,t}^P$ 连接,得到姿态估计任务下该帧中每个点的特征向量 $z_{i,t}^P$;其公式为:

$$z_{i,t}^P = [e_{i,t}^P; r_t^P]。$$

3. 根据一种基于联合学习的动态毫米波雷达点云人体解析方法,其特征在於:所述步骤1.1中进一步,将帧中所有点的点特征 $h_{i,t}^H$ 聚合为帧特征 g_t^H ,以提取帧的全局信息;其公

式为：

$$g_t^H = \sum_{i=1}^N A(h_{i,t}^H; \theta_a) \times h_{i,t}^H$$

其中N为t时刻对应帧包含点的个数，A()代表注意力函数，为 θ_a 表示注意力函数的可学习参数；

最后，将帧特征 g_t^H 连接到点特征 $e_{i,t}^H$ ，获得人体解析任务下该帧中每个点的特征向量 $z_{i,t}^H$ ；其公式为：

$$z_{i,t}^H = [e_{i,t}^H; g_t^H]。$$

4. 根据权利要求1所述的一种基于联合学习的动态毫米波雷达点云人体解析方法，其特征在于：所述步骤2分别使用两个并行的NLN执行人体解析和姿态估计任务；具体如下：

步骤2.1：任务内注意力机制，对于人体解析任务，解析NLN将一系列解析特征作为输入，并执行任务自我关注，以聚合不同帧的特征；这将生成一个全局上下文，用于对每个帧中的身体部位进行分类，以解决由于局部帧中的镜面反射而丢失身体部位的问题；

步骤2.2：任务间注意力机制；为了融合人体解析和姿态估计任务中的特征，需找出解析特征和姿态特征之间的相关性，采用任务间注意力机制，以计算解析特征与姿态特征的时空相关性；对于解析任务，本方法将姿态估计特征输入解析NLN，首先对解析特征矩阵 Z^H 和姿态特征矩阵 Z^P 进行线性变换，之后对结果进行点积和归一化，最终得到人体解析任务的任務间注意力矩阵 a^{H-P} ；

步骤2.3特征聚合：利用任务内和任务间注意力矩阵，本发明融合所有帧中的人体解析特征和姿势估计特征，以预测特定帧中点的身体部分。

步骤2.4：模型输出；人体解析特征 Y^H 和姿态估计特征 Y^P 分别由多层感知机和全连接神经网络处理，最终的输出为人体部位分类信息和人体骨骼关键点位置信息。

5. 根据权利要求4所述的一种基于联合学习的动态毫米波雷达点云人体解析方法，其特征在于：

所述步骤2.1：对于解析任务，首先将所有时间序列内所有点的点特征 $z_{i,t}^H$ 堆叠为特征矩阵 Z^H ，之后，将 Z^H 经线性变换转换得到嵌入向量 $W_\theta^H Z^H$ 和 $W_\phi^H Z^H$ ；进一步，为了估计每组帧中点之间的时空相关性，通过非线性函数对嵌入向量进行点积和归一化处理，得到人体解析任务下的任务内注意力矩阵 a^H 。公式如下：

$$a^H = \sigma[W_\theta^H Z^H (W_\phi^H Z^H)^T]$$

其中 σ 表示非线性函数， W_θ^H 和 W_ϕ^H 代表线性变换的参数；

类似地，在姿态估计任务中，结果同样的处理过程，得到相应任务内注意力矩阵 a^P ；公式如下：

$$a^P = \sigma[W_\theta^P Z^P (W_\phi^P Z^P)^T]$$

其中 σ 表示非线性函数, W_{θ}^P 和 W_{ϕ}^P 代表线性变换的参数, Z^P 表示姿态估计任务下所有时间序列内所有点的点特征 $z_{i,t}^P$ 堆叠得到的特征矩阵。

6. 根据权利要求4所述的一种基于联合学习的动态毫米波雷达点云人体解析方法, 其特征在于: 所述步骤2.2的具体的公式如下:

$$a^{H \rightarrow P} = \sigma[W_{\theta}^{H \rightarrow P} Z^H (W_{\phi}^{H \rightarrow P} Z^P)^T]$$

其中 $W_{\theta}^{H \rightarrow P}$ 和 $W_{\phi}^{H \rightarrow P}$ 表示线性变换的权重参数;

类似地, 将人体解析特征输入姿态估计NLN中, 得到姿态估计任务的任务间注意力矩阵 $a^{P \rightarrow H}$; 公式如下:

$$a^{P \rightarrow H} = \sigma[W_{\theta}^{P \rightarrow H} Z^P (W_{\phi}^{P \rightarrow H} Z^H)^T]$$

其中 $W_{\theta}^{P \rightarrow H}$ 和 $W_{\phi}^{P \rightarrow H}$ 表示线性变换的权重参数。

7. 根据权利要求4所述的一种基于联合学习的动态毫米波雷达点云人体解析方法, 其特征在于: 所述步骤2.3具体来说, 对于人体解析任务, 首先将特征矩阵 Z^H 分别线性变换为 $W_g^H Z^H$ 和 $W_g^{H \rightarrow P} Z^H$, 之后分别与任务内注意力矩阵 a^H 和任务间注意力矩阵 $a^{H \rightarrow P}$ 相乘, 分别得到任务内特征和任务间特征, 并依据所有特征与当前帧的相关性计算所有特征的加权和; 最后, 将任务内特征和任务间特征连接, 结果与原始特征 Z^H 逐元素相加并生成最终的聚合后的人体解析特征 Y^H ; 公式如下:

$$Y^H = W_y^H [a^H W_g^H Z^H; a^{H \rightarrow P} W_g^{H \rightarrow P} Z^H] + Z^H$$

其中 W_g^H 和 $W_g^{H \rightarrow P}$ 分别为线性变换参数;

对于姿态估计任务, 采用与上述方法同样的方式, 最终得到聚合后的姿态估计特征 Y^P ; 其公式为:

$$Y^P = W_y^P [a^P W_g^P Z^P; a^{P \rightarrow H} W_g^{P \rightarrow H} Z^P] + Z^P$$

其中 W_g^P 和 $W_g^{P \rightarrow H}$ 分别为线性变换参数。

8. 根据权利要求1所述的一种基于联合学习的动态毫米波雷达点云人体解析方法, 其特征在于: 所述步骤3需要说明的是: Kinect系统仅用于离线训练阶段, 但在推理阶段不需要; 对于人体解析任务, 采用交叉熵损失来最小化每个点的预测身体部位分类和真实分类之间的误差; 公式如下:

$$L_H = -\frac{1}{N} \sum_{n=1}^N \sum_{k=1}^K y_k^n \log p(\hat{y}_k^n)$$

其中 N 表示点的个数, K 为语义标签的类数, y_k^n 是输出为0或1的函数, 当 y_k^n 为1时表明样本 n 属于类别 k , $p(\hat{y}_k^n)$ 是样本 b 属于类别 k 的预测概率;

对于姿态估计任务, 采用均方误差来最小化骨架关节的预测位置和实际位置之间的误

差;公式如下:

$$L_P = \frac{1}{M} \sum_{m=1}^M \| \hat{p}_m - p_m \|$$

其中 $\| \cdot \|$ 表示L2范式, \hat{p}_m 和 p_m 分别表示骨骼关节的预测值和实际值,M表示选取的骨骼关节个数;网络架构是端到端训练的,系统整体的监督函数为:

$$L = \gamma L_H + \beta L_p$$

其中, γ 和 β 是超参数。

一种基于联合学习的动态毫米波雷达点云人体解析方法

技术领域

[0001] 本发明涉及于物联网感知技术领域,尤其涉及一种基于联合学习的动态毫米波雷达点云人体解析方法。

背景技术

[0002] 感知和理解人类活动在以人为中心的智能应用中扮演着越来越重要的角色。传统方法采用摄像头或身体接触式传感器,易受恶劣环境影响,存在隐私问题。在人类感知方面,近年来出现了使用毫米波雷达的研究热潮,并证明了其在手势和活动识别、姿势估计、身份识别等方面的有效性。但这些任务通常无法明确获取毫米波点云的语义信息,即难以分辨每个雷达点对应的身体部位信息,实现人体解析。

[0003] 在人体感知应用中,不断需要细粒度的身体部位信息,而这种语义信息的缺乏极大地限制了毫米波雷达在日常生活中成为人本计算的使能技术。同时,将语义信息作为额外的信息输入通道可以使人类感知任务更具鲁棒性。各种计算机视觉任务已经证实,包含语义信息输入可以显著提高姿势估计、活动识别和人物识别的准确性,这种优势对于毫米波雷达来说更为突出,因为毫米波点云本质上比视觉传感器的图像质量更低。因此需要一种技术方案,实现动态毫米波雷达点云人体解析任务,得到带有身体部位语义信息标签的点云。

[0004] 毫米波点云的稀疏特性使得特征提取面临挑战:受限于单芯片和天线尺寸,毫米波点云极其稀疏,这种稀疏性使得即使用人类肉眼也很难从点云中感知到人体的详细结构。采用现有深度神经网络模型提取包含人体结构信息(如姿势)的特征具有挑战性,这将直接影响人体解析任务。

[0005] 镜面反射造成毫米波点云数据中身体部位缺失:受限于低成本毫米波雷达较小的天线孔径,大部分人体反射信号没有返回到传感器,出现镜面反射,导致点云中身体部位缺失,最终导致错误的解析结果。

发明内容

[0006] 为解决上述问题,本发明公开了一种基于联合学习的动态毫米波雷达点云人体解析方法;本发明对于毫米波点云数据,首先进行聚类,之后通过多任务学习模型联合执行人体解析和姿态估计任务进行特征提取,再通过非局部网络进行多任务特征融合,最终输出结果为带注释语义标签的点云。

[0007] 具体的技术方案如下:

[0008] 步骤1:通过多任务特征提取模块解决毫米波点云的稀疏性;采用多任务学习模型联合执行人体解析的主要任务和姿态估计的辅助任务;辅助任务可以有效引导人类解析网络提取代表受试者姿势的高级结构特征;由于人体姿势和人体解析之间具有强相关性,人体姿势相关特征有助于提高解析网络在预测语义标签时的准确性和鲁棒性;对于人类解析和姿态估计任务,多任务学习模型并行提取相应的特征;步骤2:通多任务特征融合模块解

决镜面反射造成点云数据中身体部位缺失问题;从非局部网络(NLN)中获得灵感,设计了一种多任务特征融合方法,结合了任务内注意和任务间注意机制,实现了从全局角度聚合主体的时空特征;

[0009] 步骤3:多任务监督,在系统离线训练阶段,本发明采用Kinect系统,用来得到人体解析标签和姿态估计标签的真实值。

[0010] 进一步的,所述步骤1具体来说,多任务学习模型可分为包括点模块、帧模块和特征聚合模块,模型的输入为长度为s的帧序列,每一帧包含n个点,每个点包含d个特征维度;

[0011] 步骤1.1:人体解析特征提取;

[0012] 在点模块,对于t时刻对应帧的点集合 C_t 中的任一雷达点 $p_{i,t}$,采用多呈感知机(MLP)得到点的高维特征表示,即点特征 $e_{i,t}^H$;其公式为:

$$[0013] \quad e_{i,t}^H = MLP(p_{i,t}; \theta_e)$$

[0014] 其中 θ_e 表示MLP的可学习参数,H表示人体解析任务。

[0015] 在帧模块,对于每个雷达点 $p_{i,t}$ 的点特征 $e_{i,t}^H$,首先编码为更高维度的特征表示形式 $h_{i,t}^H$;其公式为:

$$[0016] \quad h_{i,t}^H = MLP(e_{i,t}^H; \theta_h)$$

[0017] 其中 θ_h 表示MLP的可学习参数;

[0018] 步骤1.2:人体姿态特征提取;

[0019] 对于人体姿态特征的提取,使用与人体解析特征提取略微不同的网络架构;具体来说,对于任一帧的帧特征 g_t^P ,使用长短期记忆网络(LSTM)进行处理;其公式为:

$$[0020] \quad r_t^P = LSTM(g_t, r_{t-1}; \theta_r)$$

[0021] 其中 θ_r 表示LSTM的参数。

[0022] 最后,将 r_t^P 与人体姿态任务的点特征 $e_{i,t}^P$ 连接,得到姿态估计任务下该帧中每个点的特征向量 $z_{i,t}^P$;其公式为:

$$[0023] \quad z_{i,t}^P = [e_{i,t}^P; r_t^P]。$$

[0024] 进一步的,所述步骤1.1中,将帧中所有点的点特征 $h_{i,t}^H$ 聚合为帧特征 g_t^H ,以提取帧的全局信息;其公式为:

$$[0025] \quad g_t^H = \sum_{i=1}^N A(h_{i,t}^H; \theta_a) \times h_{i,t}^H$$

[0026] 其中N为t时刻对应帧包含点的个数,A()代表注意力函数,为 θ_a 表示注意力函数的可学习参数;

[0027] 最后,将帧特征 g_t^H 连接到点特征 $e_{i,t}^H$,获得人体解析任务下该帧中每个点的特征

向量 $\mathbf{z}_{i,t}^H$;其公式为:

$$[0028] \quad \mathbf{z}_{i,t}^H = [\mathbf{e}_{i,t}^H; \mathbf{g}_t^H]。$$

[0029] 进一步的,所述步骤2分别使用两个并行的NLN执行人体解析和姿态估计任务;具体如下:

[0030] 步骤2.1:任务内注意力机制,对于人体解析任务,解析NLN将一系列解析特征作为输入,并执行任务自我关注,以聚合不同帧的特征;这将生成一个全局上下文,用于对每个帧中的身体部位进行分类,以解决由于局部帧中的镜面反射而丢失身体部位的问题;

[0031] 步骤2.2:任务间注意力机制;为了融合人体解析和姿态估计任务中的特征,需找出解析特征和姿态特征之间的相关性,采用任务间注意力机制,以计算解析特征与姿态特征的时空相关性;对于解析任务,本方法将姿态估计特征输入解析NLN,首先对解析特征矩阵 Z^H 和姿态特征矩阵 Z^P 进行线性变换,之后对结果进行点积和归一化,最终得到人体解析任务的跨任务注意力矩阵 $\mathbf{a}^{H \rightarrow P}$;

[0032] 步骤2.3特征聚合:利用任务内和跨任务注意力矩阵,本发明融合所有帧中的人体解析特征和姿势估计特征,以预测特定帧中点的身体部分。

[0033] 步骤2.4:模型输出;人体解析特征 Y^H 和姿态估计特征 Y^P 分别由多层感知机和全连接神经网络处理,最终的输出为人体部位分类信息和人体骨骼关键点位置信息。

[0034] 进一步的,所述步骤2.1:对于解析任务,首先将所有时间序列内所有点的点特征 $\mathbf{z}_{i,t}^H$ 堆叠为特征矩阵 Z^H ,之后,将 Z^H 经线性变换转换得到嵌入向量 $W_\theta^H Z^H$ 和 $W_\phi^H Z^H$ 。进一步,为了估计每组帧中点之间的时空相关性,通过非线性函数对嵌入向量进行点积和归一化处理,得到人体解析任务下的任务内注意力矩阵 \mathbf{a}^H 。公式如下:

$$[0035] \quad \mathbf{a}^H = \sigma[W_\theta^H Z^H (W_\phi^H Z^H)^T]$$

[0036] 其中 σ 表示非线性函数, W_θ^H 和 W_ϕ^H 代表线性变换的参数。

[0037] 类似地,在姿态估计任务中,结果同样的处理过程,得到相应任务内注意力矩阵 \mathbf{a}^P ;公式如下:

$$[0038] \quad \mathbf{a}^P = \sigma[W_\theta^P Z^P (W_\phi^P Z^P)^T]$$

[0039] 其中 σ 表示非线性函数, W_θ^P 和 W_ϕ^P 代表线性变换的参数, Z^P 表示姿态估计任务下所有时间序列内所有点的点特征 $\mathbf{z}_{i,t}^P$ 堆叠得到的特征矩阵。

[0040] 进一步的,所述步骤2.2的具体的公式如下:

$$[0041] \quad \mathbf{a}^{H \rightarrow P} = \sigma[W_\theta^{H \rightarrow P} Z^H (W_\phi^{H \rightarrow P} Z^P)^T]$$

[0042] 其中 $W_\theta^{H \rightarrow P}$ 和 $W_\phi^{H \rightarrow P}$ 表示线性变换的权重参数。

[0043] 类似地,将人体解析特征输入姿态估计NLN中,得到姿态估计任务的跨任务注意力矩阵 $\mathbf{a}^{P \rightarrow H}$;公式如下:

$$[0044] \quad \mathbf{a}^{P \rightarrow H} = \sigma[W_\theta^{P \rightarrow H} Z^P (W_\phi^{P \rightarrow H} Z^H)^T]$$

[0045] 其中 $W_{\theta}^{P \rightarrow H}$ 和 $W_{\phi}^{P \rightarrow H}$ 表示线性变换的权重参数。

[0046] 进一步的,所述步骤2.3具体来说,对于人体解析任务,首先将特征矩阵 Z^H 分别线性变换为 $W_g^H Z^H$ 和 $W_g^{H \rightarrow P} Z^H$,之后分别与任务内注意力矩阵 a^H 和任务间注意力矩阵 $a^{H \rightarrow P}$ 相乘,分别得到任务内特征和任务间特征,并依据所有特征与当前帧的相关性计算所有特征的加权和;最后,将任务内特征和任务间特征连接,结果与原始特征 Z^H 逐元素相加并生成最终的聚合后的人体解析特征 Y^H ;公式如下:

$$[0047] \quad Y^H = W_y^H [a^H W_g^H Z^H; a^{H \rightarrow P} W_g^{H \rightarrow P} Z^H] + Z^H$$

[0048] 其中 W_g^H 和 $W_g^{H \rightarrow P}$ 分别为线性变换参数;

[0049] 对于姿态估计任务,采用与上述方法同样的方式,最终得到聚合后的姿态估计特征 Y^P ;其公式为:

$$[0050] \quad Y^P = W_y^P [a^P W_g^P Z^P; a^{P \rightarrow H} W_g^{P \rightarrow H} Z^P] + Z^P$$

[0051] 其中 W_g^P 和 $W_g^{P \rightarrow H}$ 分别为线性变换参数。

[0052] 进一步的,所述步骤3需要说明的是:Kinect系统仅用于离线训练阶段,但在推理阶段不需要;对于人体解析任务,采用交叉熵损失来最小化每个点的预测身体部位分类和真实分类之间的误差;公式如下:

$$[0053] \quad L_H = -\frac{1}{N} \sum_{n=1}^N \sum_{k=1}^K y_k^n \log p(\hat{y}_k^n)$$

[0054] 其中N表示点的个数,K为语义标签的类数, y_k^n 是输出为0或1的函数,当 y_k^n 为1时表明样本n属于类别k, $p(\hat{y}_k^n)$ 是样本b属于类别k的预测概率;

[0055] 对于姿态估计任务,采用均方误差来最小化骨架关节的预测位置和实际位置之间的误差;公式如下:

$$[0056] \quad L_P = \frac{1}{M} \sum_{m=1}^M \| \hat{p}_m - p_m \|^2$$

[0057] 其中 $\| \cdot \|$ 表示L2范式, \hat{p}_m 和 p_m 分别表示骨骼关节的预测值和实际值,M表示选取的骨骼关节节点个数;网络架构是端到端训练的,系统整体的监督函数为:

$$[0058] \quad L = \gamma L_H + \beta L_P$$

[0059] 其中, γ 和 β 是超参数。

[0060] 本发明的有益效果::

[0061] 本发明设计了一种基于联合学习的动态毫米波雷达点云人体解析方法,可以生成带注释语义标签的点云,解决了当前以人为中心的毫米波感知无法获取毫米波点云的语义信息的问题,达到了约92%的准确率和84%的IoU准确率,预测的语义标签可分别将两个下游任务(姿势估计和动作识别)的性能提高约18%和6%。

附图说明

- [0062] 图1:系统总体结构示意图。
 [0063] 图2:多任务特征提取模块结构示意图。
 [0064] 图3:多任务特征融合模块结构示意图。
 [0065] 图4:不同场景下的系统准确率。

具体实施方式

[0066] 下面结合附图和具体实施方式,进一步阐明本发明,应理解下述具体实施方式仅用于说明本发明而不适用于限制本发明的范围。需要说明的是,下面描述中使用的词语“前”、“后”、“左”、“右”、“上”和“下”指的是附图中的方向,词语“内”和“外”分别指的是朝向或远离特定部件几何中心的方向。

[0067] 如图1所示,本实施例的一种基于联合学习的动态毫米波雷达点云人体解析方法,具体方法步骤如下:

[0068] 如图2所示,步骤1:通过“多任务特征提取模块”解决毫米波点云的稀疏性

[0069] 针对毫米波点云稀疏性,现有深度神经网络模型难以提取包含人体结构信息的特征问题,本发明采用多任务学习模型联合执行人体解析(主要任务)和姿态估计(辅助任务)。辅助任务可以有效引导人类解析网络提取代表受试者姿势的高级结构特征。由于人体姿势和人体解析之间具有强相关性,人体姿势相关特征有助于提高解析网络在预测语义标签时的准确性和鲁棒性。对于人类解析和姿态估计任务,多任务学习模型并行提取相应的特征,具体来说,多任务学习模型可分为包括点模块、帧模块和特征聚合模块,模型的输入为长度为s的帧序列,每一帧包含n个点,每个点包含d个特征维度。

[0070] 步骤1.1:人体解析特征提取;

[0071] 在点模块,对于t时刻对应帧的点集合 C_t 中的任一雷达点 $p_{i,t}$,采用多呈感知机(MLP)得到点的高维特征表示,即点特征 $e_{i,t}^H$ 。其公式为:

$$[0072] \quad e_{i,t}^H = \text{MLP}(p_{i,t}; \theta_e)$$

[0073] 其中 θ_e 表示MLP的可学习参数,H表示人体解析任务。

[0074] 在帧模块,对于每个雷达点 $p_{i,t}$ 的点特征 $e_{i,t}^H$,首先编码为更高维度的特征表示形式 $h_{i,t}^H$ 。其公式为:

$$[0075] \quad h_{i,t}^H = \text{MLP}(e_{i,t}^H; \theta_h)$$

[0076] 其中 θ_h 表示MLP的可学习参数。

[0077] 进一步,将帧中所有点的点特征 $h_{i,t}^H$ 聚合为帧特征 g_t^H ,以提取帧的全局信息。其公式为:

$$[0078] \quad g_t^H = \sum_{i=1}^N A(h_{i,t}^H; \theta_a) \times h_{i,t}^H$$

[0079] 其中N为t时刻对应帧包含点的个数,A()代表注意力函数,为 θ_a 表示注意力函数的

可学习参数。

[0080] 最后,将帧特征 g_t^H 连接到点特征 $e_{i,t}^H$,获得人体解析任务下该帧中每个点的特征向量 $z_{i,t}^H$ 。其公式为:

$$[0081] \quad z_{i,t}^H = [e_{i,t}^H; g_t^H]$$

[0082] 步骤1.2:人体姿态特征提取;

[0083] 对于人体姿态特征的提取,使用与人体解析特征提取略微不同的网络架构。具体来说,对于任一帧的帧特征 g_t^P ,使用长短期记忆网络(LSTM)进行处理。其公式为:

$$[0084] \quad r_t^P = \text{LSTM}(g_t, r_{t-1}; \theta_r)$$

[0085] 其中 θ_r 表示LSTM的参数。

[0086] 最后,将 r_t^P 与人体姿态任务的点特征 $e_{i,t}^P$ 连接,得到姿态估计任务下该帧中每个点的特征向量 $z_{i,t}^P$ 。其公式为:

$$[0087] \quad z_{i,t}^P = [e_{i,t}^P; r_t^P]$$

[0088] 如图3所示,步骤2:通过“多任务特征融合模块”解决镜面反射造成点云数据中身体部位缺失问题

[0089] 为了解决镜面反射造成点云数据中身体部位缺失问题,本发明从非局部网络(NLN)中获得灵感,设计了一种多任务特征融合方法,该方法结合了任务内注意和任务间注意机制,实现了从全局角度聚合主体的时空特征。具体来说,本方法分别使用两个并行的NLN执行人体解析和姿态估计任务。

[0090] 步骤2.1:任务内注意力机制;

[0091] 对于人体解析任务,解析NLN将一系列解析特征作为输入,并执行任务自我关注,以聚合不同帧的特征。这将生成一个全局上下文,用于对每个帧中的身体部位进行分类,以解决由于局部帧中的镜面反射而丢失身体部位的问题。更具体地,对于解析任务,首先将所有时间序列内所有点的点特征 $z_{i,t}^H$ 堆叠为特征矩阵 Z^H ,之后,将 Z^H 经线性变换转换得到嵌入向量 $W_\theta^H Z^H$ 和 $W_\phi^H Z^H$ 。进一步,为了估计每组帧中点之间的时空相关性,通过非线性函数(如softmax函数)对嵌入向量进行点积和归一化处理,得到人体解析任务下的任务内注意力矩阵 a^H 。公式如下:

$$[0092] \quad a^H = \sigma[W_\theta^H Z^H (W_\phi^H Z^H)^T]$$

[0093] 其中 σ 表示非线性函数, W_θ^H 和 W_ϕ^H 代表线性变换的参数。

[0094] 类似地,在姿态估计任务中,结果同样的处理过程,得到相应任务内注意力矩阵 a^P 。公式如下:

$$[0095] \quad a^P = \sigma[W_\theta^P Z^P (W_\phi^P Z^P)^T]$$

[0096] 其中 σ 表示非线性函数, W_θ^P 和 W_ϕ^P 代表线性变换的参数, Z^P 表示姿态估计任务下

所有时间序列内所有点的点特征 $z_{i,t}^P$ 堆叠得到的特征矩阵。

[0097] 步骤2.2:任务间注意力机制;

[0098] 为了融合人体解析和姿态估计任务中的特征,需找出解析特征和姿态特征之间的相关性,本方采用任务间注意力机制,以计算解析特征与姿态特征的时空相关性。实现上,对于解析任务,本方法将姿态估计特征输入解析NLN,首先对解析特征矩阵 Z^H 和姿态特征矩阵 Z^P 进行线性变换,之后对结果进行点积和归一化,最终得到人体解析任务的任务间注意力矩阵 $a^{H \rightarrow P}$ 。公式如下:

$$[0099] \quad a^{H \rightarrow P} = \sigma[W_{\theta}^{H \rightarrow P} Z^H (W_{\phi}^{H \rightarrow P} Z^P)^T]$$

[0100] 其中 $W_{\theta}^{H \rightarrow P}$ 和 $W_{\phi}^{H \rightarrow P}$ 表示线性变换的权重参数。

[0101] 类似地,将人体解析特征输入姿态估计NLN中,得到姿态估计任务的任务间注意力矩阵 $a^{P \rightarrow H}$ 。公式如下:

$$[0102] \quad a^{P \rightarrow H} = \sigma[W_{\theta}^{P \rightarrow H} Z^P (W_{\phi}^{P \rightarrow H} Z^H)^T]$$

[0103] 其中 $W_{\theta}^{P \rightarrow H}$ 和 $W_{\phi}^{P \rightarrow H}$ 表示线性变换的权重参数。

[0104] 步骤2.3:特征聚合;

[0105] 利用任务内和任务间注意力矩阵,本发明融合所有帧中的人体解析特征和姿势估计特征,以预测特定帧中点的身体部分。

[0106] 具体来说,对于人体解析任务,首先将特征矩阵 Z^H 分别线性变换为 $W_g^H Z^H$ 和 $W_g^{H \rightarrow P} Z^H$,之后分别与任务内注意力矩阵 a^H 和任务间注意力矩阵 $a^{H \rightarrow P}$ 相乘,分别得到任务内特征和任务间特征,并依据所有特征与当前帧的相关性计算所有特征的加权和。最后,将任务内特征和任务间特征连接,结果与原始特征 Z^H 逐元素相加并生成最终的聚合后的人体解析特征 Y^H 。公式如下:

$$[0107] \quad Y^H = W_y^H [a^H W_g^H Z^H; a^{H \rightarrow P} W_g^{H \rightarrow P} Z^H] + Z^H$$

[0108] 其中 W_g^H 和 $W_g^{H \rightarrow P}$ 分别为线性变换参数。

[0109] 对于姿态估计任务,采用与上述方法同样的方式,最终得到聚合后的姿态估计特征 Y^P 。其公式为:

$$[0110] \quad Y^P = W_y^P [a^P W_g^P Z^P; a^{P \rightarrow H} W_g^{P \rightarrow H} Z^P] + Z^P$$

[0111] 其中 W_g^P 和 $W_g^{P \rightarrow H}$ 分别为线性变换参数。

[0112] 步骤2.4:模型输出;

[0113] 人体解析特征 Y^H 和姿态估计特征 Y^P 分别由多层感知机(MLP)和全连接神经网络(FC)处理,最终的输出为人体部位分类信息和人体骨骼关键点位置信息。

[0114] 步骤3:多任务监督

[0115] 在系统离线训练阶段,本发明采用Kinect系统,用来得到人体解析标签和姿态估计标签的真实值。需要说明的是,Kinect系统仅用于离线训练阶段,但在推理阶段不需要。

对于人体解析任务,本发明采用交叉熵损失来最小化每个点的预测身体部位分类和真实分类之间的误差。公式如下:

$$[0116] \quad L_H = -\frac{1}{N} \sum_{n=1}^N \sum_{k=1}^K y_k^n \log p(\hat{y}_k^n)$$

[0117] 其中N表示点的个数,K为语义标签的类数, y_k^n 是输出为0或1的函数,当 y_k^n 为1时表示样本n属于类别k, $p(\hat{y}_k^n)$ 是样本n属于类别k的预测概率。

[0118] 对于姿态估计任务,本发明采用均方误差(MSE)来最小化骨架关节的预测位置和实际位置之间的误差。公式如下:

$$[0119] \quad L_P = \frac{1}{M} \sum_{m=1}^M \| \hat{p}_m - p_m \|^2$$

[0120] 其中 $\| \cdot \|^2$ 表示L2范式, \hat{p}_m 和 p_m 分别表示骨骼关节的预测值和实际值,M表示选取的骨骼关节个数。

[0121] 本发明设计的网络架构是端到端训练的,系统整体的监督函数为:

$$[0122] \quad L = \gamma L_H + \beta L_P$$

[0123] 其中, γ 和 β 是超参数。

[0124] 如图4所示,为不同场景下的系统准确率。

[0125] 实施例1:应急救援

[0126] 在应急救援中,救援队员往往需要紧密配合,将工具精准无误的交到队友手中。但在诸如火灾救援场景下,存在着大量的烟雾,传统基于摄像头的成像设备难以正常工作,基于毫米波的成像设备在恶劣环境中依然具有鲁棒性。本发明通过增加带有身体部位的额外语义标签作为额外输入,可以提高雷达识别人员手部的准确率,从而在烟雾环境中协助救援人员精准完成工具交接。

[0127] 实施例2:动作识别

[0128] 在某些场景下,需要精准感知人类的动作。如在养老院场景中,用以健康监测的毫米波设备需要准确无误的识别出老人的摔倒动作。本发明通过增加点云的身体部位语义信息,可以提升毫米波设备的动作识别性能。

[0129] 实施例3:身份识别

[0130] 由于摄像头可能造成隐私侵犯问题,近年来,越来越多的毫米波设备被部署于诸如仓库、办公室等私密场景中,用以代替摄像头进行监控。不同于相机成像含有丰富的语义信息,毫米波成像具有固有的稀疏性,在识别人员身份上处于劣势。本发明通过增加点云的身体部位语义信息,可以提升毫米波设备的人员身份识别性能。

[0131] 实施例4:自动驾驶

[0132] 在自动驾驶领域,帮助车辆理解和识别行人的动作具有重大意义,本发明可以应用于车载毫米波雷达设备,以提升自动驾驶车辆对行人的动作识别能力,遇到突发情况时可以及时做出预判和响应。

[0133] 本发明方案所公开的技术手段不仅限于上述实施方式所公开的技术手段,还包括由以上技术特征任意组合所组成的技术方案。

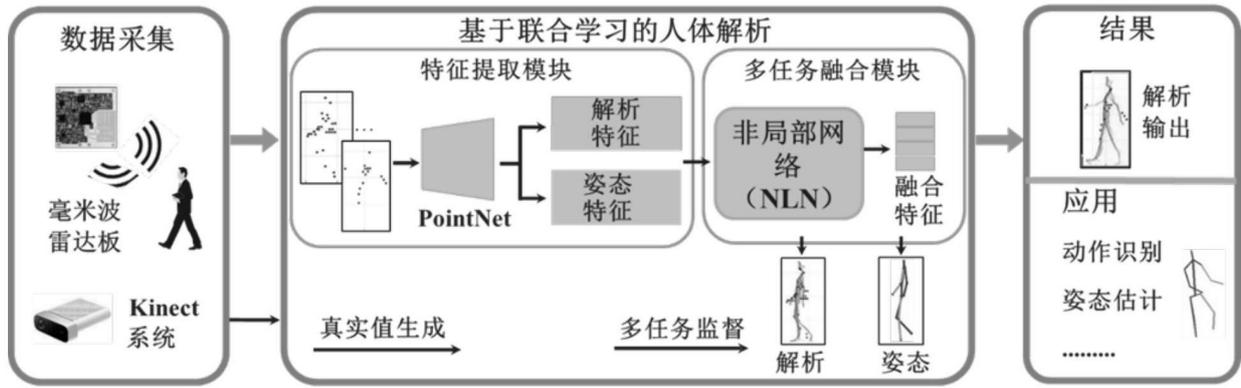


图1

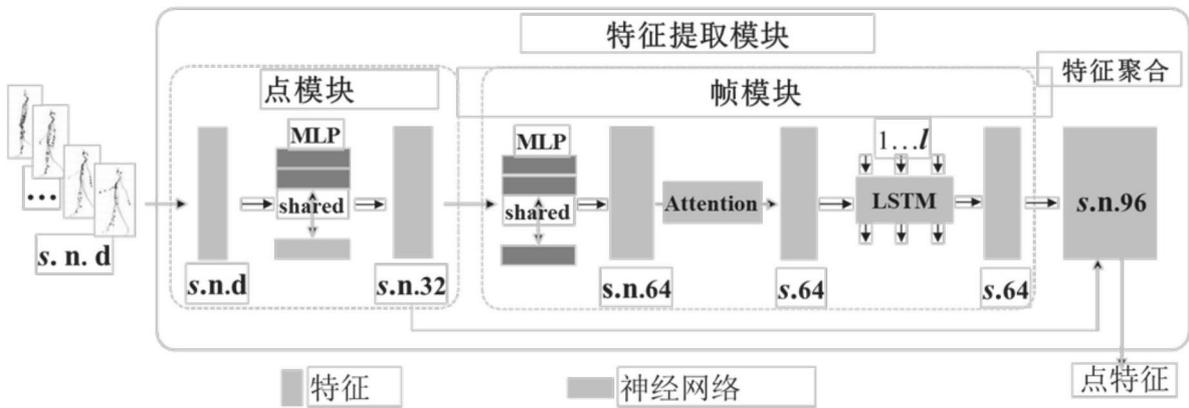


图2

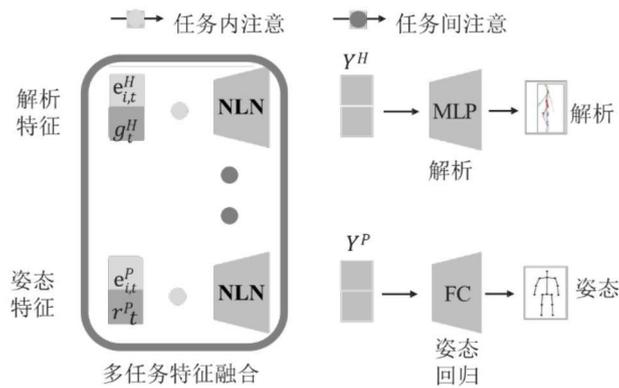


图3

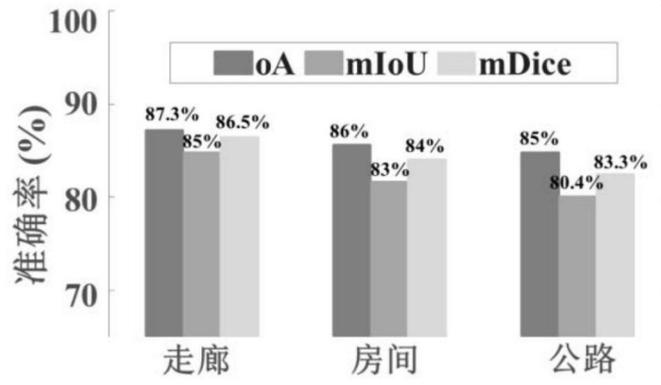


图4