

基于毫米波雷达稀疏点云的人体行为识别方法

李育臣, 张之江, 曾丹, 李佳
(上海大学 通信与信息工程学院, 上海 200444)

摘要: 目前利用毫米波雷达进行人体行为识别的方法在复杂场景下无法很好地区分相似动作, 与此同时模型的鲁棒性和抗干扰能力也相对较差; 针对以上两个问题, 提出了一种通用的基于毫米波雷达稀疏点云的人体行为识别方法, 该方法首先利用 K-means++ 聚类算法对点云进行采样, 然后使用基于注意力特征融合的点云活动分类网络进行人体行为特征的提取和识别, 该网络可以兼顾点云的空间特征以及时序特征, 对稀疏点云的运动有灵敏的感知能力; 为了验证所提出方法的有效性和鲁棒性, 分别在 MMAActivity 数据集和 MMGesture 数据集上进行了实验, 并在两个数据集上取得 97.50% 和 94.10% 的准确率, 均优于其它方法; 此外, 进一步验证了 K-means++ 点云采样方法的有效性, 相较于随机采样, 准确率提升了 0.4 个百分点, 实验结果表明所提出方法能够有效地提升人体行为识别的准确率, 且模型具有较好的泛化能力。

关键词: 毫米波雷达; 人体感知; 行为识别; 稀疏点云; 特征融合

Human Activity Recognition Method Based on Millimeter Wave Radar Sparse Point Clouds

LI Yuchen, ZHANG Zhijiang, ZENG Dan, LI Jia

(School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China)

Abstract: At present, human behavior recognition methods based on millimeter wave radar cannot distinguish similar actions in complicated scenes. In addition, these methods have the characteristics of low robustness and interference resistance. To address the above two issues, a universal human behavior recognition method based on millimeter wave radar sparse point clouds is proposed. Firstly, the method samples the point cloud using the K-means++ clustering algorithm, and then adopts a point cloud activity classification network based on attentional feature fusion for the extraction and recognition of human behavior features, which can consider both the spatial and temporal features of point clouds and has the sensitive perception of sparse point cloud motion. In order to verify the effectiveness and robustness of the proposed method, the experiments are conducted on the MMAActivity dataset and MMGesture dataset, respectively, with the accuracy of 97.50% and 94.10% on both datasets, outperforming other methods. Furthermore, the effectiveness of the K-means++ point cloud sampling method is further verified, and compared to random sampling, the accuracy is improved by 0.4%. The experimental results show that the proposed method can effectively promote the accuracy of human behavior recognition, and the model has a strong generalization ability.

Keywords: millimeter-wave radar; human perception; behavior recognition; sparse point clouds; feature fusion

0 引言

随着人口老龄化程度的加深以及护理服务需求的增多, 老人跌倒监测以及其他人体行为感知逐渐成为辅助生活的关键技术^[1-3]。目前, 摄像头、WIFI、雷达等设备均可以用于非接触式行为感知。利用摄像头监控人体行为已经成为非常成熟的技术^[4-6], 但是也存在一些问题阻碍了其发展。在家庭环境下, 人们越来越关注的隐私问题限制了摄像监控的应用; 同时, 摄像头对环境的干扰比较敏感, 易受到光线条件和周围物体遮挡的影响。此外, 利用 WIFI^[7]也可用于探测人体的行为, 但它们精度较低。相比于其他的无线设备, 毫米波雷达具有精度高和探测距离较远的优点,

并且不会涉及隐私问题, 也不易受到光照影响, 在黑暗环境下也可正常工作, 具有一定的穿透性, 因此毫米波雷达在行为感知方面具有很大的优势^[8-10]。

目前较为主流的手段是利用多普勒雷达收发毫米波, 然后使用时频分析法将原始信号映射到时频图上, 并观察不同行为的微多普勒效应^[2, 9, 11]。但是这种基于微多普勒图的行为识别有一定的局限性, 即需要对雷达回波信号进行运算量比较大的时频分析, 同时特征的定义和分类器的设计也比较复杂, 并且只通过速度信息来判断跌倒在某些场合的准确度并不高。Singh 等人^[12]在 2019 年证明了稀疏点云的时间窗口体素表示可以用于人类行为识别。通过将点

收稿日期: 2023-03-22; 修回日期: 2023-04-25。

作者简介: 李育臣(1998-), 男, 硕士研究生。

张之江(1968-), 男, 博士, 教授。

曾丹(1982-), 女, 博士, 教授。

引用格式: 李育臣, 张之江, 曾丹, 等. 基于毫米波雷达稀疏点云的人体行为识别方法[J]. 计算机测量与控制, 2024, 32(2): 198-205.

云体素化, 利用基于深度学习的分类器进行了测试, 可以达到甚至超过基于微多普勒图特征分类的效果。但该方法计算量较大, 实时处理能力较弱, 并且复杂环境中的噪声对点云体素化的效果有较大影响。Gong 等人^[13]尝试用图的方式来表示点云, 虽然取得了不错的分类效果, 但是由于毫米波雷达的特殊性, 对于人体行为识别系统的抗干扰能力、实时性以及鲁棒性有着很大的要求, 在满足实时性的同时也要具备很强的抗干扰能力和模型泛化能力, 该方法并不能满足要求。因此, 如何在复杂的环境中准确地、快速地识别人类行为仍然存在着巨大挑战。

现有方法无法在复杂环境下有效地提取稀疏点云的空间特征和时序特征, 同时模型的抗干扰能力和泛化能力较差。综上, 提出了一种通用的基于毫米波雷达稀疏点云的人体行为识别方法, 该方法根据毫米波雷达所产生点云的特性, 既保留了点云的原始特征, 提高了网络计算的效率, 同时使用基于注意力特征融合的点云活动分类网络 (MM-PCANet, MMwave radar point cloud activity network) 进行稀疏点云的特征提取和识别。在 MMAActivity 数据集^[12]和 MMGesture 数据集^[14]上评估了所提的方法, 分别取得 97.50% 和 94.10% 的准确率, 均优于所有其它基线。从而证明了所提方法具有很好的鲁棒性, 也适用于解决毫米波雷达点云相关的其它问题。

1 相关工作

1.1 毫米波雷达点云生成

在过去几年中, 单片毫米波雷达有着快速的发展和运用, 美国德克萨斯仪器公司的毫米波雷达就是其中一个具有代表性的设备, 可以利用这些毫米波雷达来捕获人体活动相关的点云。点云指的是一系列包含空间坐标、速度以及其它信息的离散点集合, 以此来表示人体在空间上的形状以及姿态等信息。毫米波雷达点云生成的整体流程如图 1 所示。

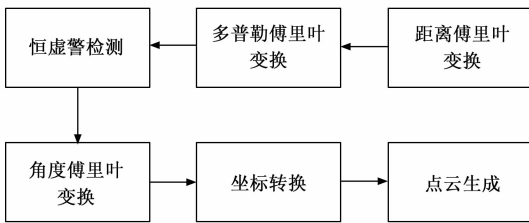


图 1 毫米波雷达点云生成过程

这些毫米波雷达采用调频连续波 (FMCW), 其在一个调频周期内发射信号^[15]可以表示为:

$$S_T(t) = A_T \cos \left\{ 2\pi \left[f_c t + \pi \frac{B}{T_c} t^2 + \varphi(t) \right] \right\} \quad (1)$$

式中, A_T 为发射信号的幅值, f_c 为载波中心频率, B 为带宽, T_c 为信号调频周期, $\varphi(t)$ 为相位噪声。经过目标和环境反射后, 接收天线得到回波信号, 回波信号可以表示为:

$$S_R(t) = A_R \cos \left\{ 2\pi \left[f_c (t - t_d) + \pi \frac{B}{T_c} (t - t_d)^2 + \varphi(t - t_d) \right] \right\} \quad (2)$$

式中, A_R 为发射信号的幅值, t_d 表示从雷达发射至接收到人体反射信号的时间, $t_d = R/c$, 其中 R 表示目标距离, c 为光速。

发射信号与回波信号混频处理并低通滤波得到中频信号, 对中频信号执行距离快速傅里叶变换 (FFT, fast fourier transform) 运算以分离不同的频率分量, 从而获得每个目标与雷达之间的距离。对中频信号执行多普勒 FFT 操作以测量中频信号的相位变化, 来获取目标的速度信息。

除此之外为了得到目标的空间坐标 x 、 y 和 z , 还需进行角度估计, 利用天线之间的相位差计算出目标的方位角和俯仰角。在经过角度估计之后, 获取了目标的距离 R 、方位角 θ 以及俯仰角 φ 。假设笛卡尔坐标系中的点坐标为 (x, y, z) , 其中 x 表示横轴, y 表示纵轴, z 表示高度, 则雷达坐标系转换为笛卡尔坐标系的计算如式 (3) 所示:

$$\begin{aligned} x &= R * \cos(\varphi) * \sin(\theta) \\ y &= R * \cos(\varphi) * \cos(\theta) \\ z &= R * \sin(\varphi) \end{aligned} \quad (3)$$

以上是毫米波雷达生成点云的大致流程, 至此就可以获取到目标的三维坐标、距离以及速度等信息, 供后续数据集构建、数据处理以及数据可视化使用。连续多帧的点云信息就是后续方法的输入, 每一帧包含 n 个离散的点, 每个点都包含空间位置 x 、 y 和 z 以及速度等信息。

1.2 点云分类网络 PointNet++

在上述所提到的 MM-PCANet 当中, 采用 PointNet++^[16]作为主干网络, 从 3D 点云中提取空间分层特征。PointNet^[17]是使用深度学习的方法提取点云特征的先驱工作, 之后出现了类似于 PointNet++ 等许多新颖的基于深度学习的点云网络, 但大多数都是以 PointNet 为基础。在 PointNet 当中使用多层感知机 (MLP, multilayer perceptron) 来进行点云的高级特征表示, 并选择用最大池化来进行特征的聚合, PointNet 提取特征的方式是对所有点云数据提取一个全局特征, 这会带来很大的局限性, 即无法提取到点云的空间局部特征, 使得它很难对复杂场景进行分析。PointNet++ 采用了分层抽取特征的思想, 把每一次特征提取操作称为 Set Abstraction。Set Abstraction 又分为三部分, 分别是采样层、分组层、特征提取层, 它能够在不同尺度提取局部特征, 通过多层网络结构得到更深层次的特征, 同时兼顾全局特征以及空间局部特征。因此, 在点云稀疏或其它复杂的场景下, PointNet++ 有着更好的特征提取能力。

2 方法

2.1 点云采样策略

在实际的人体行为数据采集过程中, 不同帧的点云数量是不相同的, 在一些复杂的动作当中, 生成的点云数量可能会比简单动作的要多, 为了使每帧点云数量保持一致, 作为后续分类网络的输入, 同时方便比较每帧固定不同点云数量时分类方法的整体性能差异, 采用基于 K-means++ 聚类算法进行点云的采样, 如果当前帧的点云个数小于

设定的阈值，则通过重复一些原有点和聚类中心点，完成点云的扩充。如果大于设定的阈值，则只保留聚类中心点，剔除中心点以外的其它点。整体流程如图 2 所示。

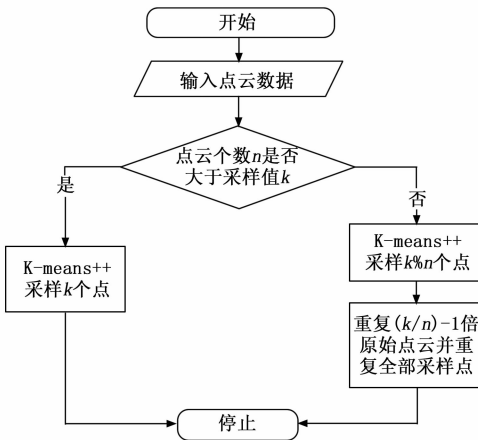


图 2 点云采样整体流程

2.1.1 方法介绍

K-means 聚类是最常用的聚类算法。它的基本思想是通过迭代寻找 K 个簇的一种划分方案，使得聚类结果对的损失函数最小。它的优点是计算复杂度较低，收敛速度较快，但是容易受初始值的影响，为了减少初始值带来的影响，选择利用 K-means++ 中选择初始聚类中心的方法初始化聚类中心。通过给出具体的 K 值，将点云数据聚类成 K 个簇，通过保留或重复聚类中心点，完成点云的采样。具体步骤如下：

- 1) 先从数据当中随机选择一个点作为聚类中心。
- 2) 对于其它每个点，都计算其与最近的一个聚类中心的欧式距离。
- 3) 用加权的方式选取聚类中心点，离已确定的中心点距离越远，越有可能被选择为下一个中心点。
- 4) 重复步骤 2) 和 3)，直到 K 个聚类中心选择完成。
- 5) 针对每个点计算其到所有聚类中心的欧式距离，并将其归属到距离最小的聚类中心所对应的类别。
- 6) 针对每个簇类，重新计算该簇的聚类中心并对聚类中心进行更新。
- 7) 重新迭代 5)、6) 两步，直到簇的聚类中心不发生改变或达到迭代上限。
- 8) 迭代完成后，保留每个簇的聚类中心作为最终的结果。

2.1.2 参数分析

随机选取一帧点云个数为 109 的数据，将聚类中心分别设置为 16、32、64、128 和 256，代表着将点云个数采样至对应的数值，迭代次数设置为 300 次，选择用欧式距离作为距离的衡量标准，采样的效果如图 3 所示，分别对应不同采样值的采样效果，可以看出利用所提出的 K-means++ 聚类算法可以很好地保留关键的特征点以及原始的点云空间特征，在进行采样的同时，不会造成点云空间结构

的变化。

2.2 分类网络结构

基于毫米波雷达的特殊性，人体行为识别需要面临如下两个挑战：第一是所采集到人体行为点云较为稀疏；第二是人体的行为并不是单帧的三维点云，而是多帧三维点云的聚合，因此如何从这些连续且无序的三维点云中高效地提取特征是主要问题。如图 4 所示，是所提出的 MM-PCANet 分类网络的整体结构图，其中主要包含基础模块、全局帧模块、全局点模块以及注意力特征融合模块等。

2.2.1 基础模块

该模块的输入维度是 $T \times N \times D$ ，其中 T 代表着时间序列帧的个数， N 代表每帧当中点云的数量， D 代表每个点的特征个数，在所提出的方法中，每个点有 4 个特征，分别是空间位置 x 、 y 、 z 以及速度。该模块的作用是利用 MLP 提取出每一帧点云的高级表示，同时为了提取到帧间的局部特征以及减少计算量，采用了帧间分组与采样模块^[18]，其包含两个部分，即帧间分组模块和采样模块。帧间分组模块可以通过在短周期内找到 k 个相邻点来构造局部点云集，从而能够在更密集的点云中学习出更精细的行为特征，当人体快速移动时，也可以学习更多的全局运动信息。采样模块旨在通过在每个时刻采样点云特征来收集局部时空信息，以降低计算成本并识别不同尺度的人体行为。

2.2.2 全局帧模块

因为人体的行为是连续帧所叠加而成的，所以帧与帧之间的时序信息是很重要的，全局帧模块旨在利用长短时记忆网络^[19] (LSTM, long short-term memory) 来提取连续帧之间的时序信息。LSTM 是循环神经网络的一种特殊自回归模型，通过加入长期记忆单元和改变门结构，缓解了循环神经网络中的梯度弥散和梯度爆炸问题，已广泛应用在序列数据处理中。该模块具体的步骤是首先经过了帧间分组与采样模块以及 MLP 多层感知机，进一步将每一帧的点云特征进行更高级的表示，然后送入 LSTM 层提取连续帧之间的时序信息，最后送入注意力特征融合模块和后续全局点模块所输出的特征相融合。具体计算方法如公式 (4) 所示：

$$\begin{aligned}
 f_t &= \sigma(\mathbf{W}_f[\mathbf{h}_{t-1}, \mathbf{y}_t] + \mathbf{b}_f) \\
 i_t &= \sigma(\mathbf{W}_i \cdot [\mathbf{h}_{t-1}, \mathbf{y}_t] + \mathbf{b}_i) \\
 \tilde{C}_t &= \tanh(\mathbf{W}_c \cdot [\mathbf{h}_{t-1}, \mathbf{y}_t] + \mathbf{b}_c) \\
 C_t &= f_t * C_{t-1} + i_t * \tilde{C}_t \\
 o_t &= \sigma(\mathbf{W}_o[\mathbf{h}_{t-1}, \mathbf{y}_t] + \mathbf{b}_o) \\
 \mathbf{h}_t &= o_t * \tanh(C_t)
 \end{aligned} \tag{4}$$

式中， \mathbf{y}_t 为时刻 t 神经元的输入， \mathbf{h}_t 为时刻 t 神经元的输出， \mathbf{W} 代表权重矩阵， \mathbf{b} 代表阈值向量， σ 代表神经元激活函数， f_t 为遗忘门的输出值， \tilde{C}_t 为输入门结果， i_t 决定 \tilde{C}_t 是否加入 t 时刻的状态， o_t 代表输出门的状态， C_{t-1} 代表上一时刻的细胞状态， C_t 代表当前时刻的细胞状态。

2.2.3 全局点模块

上述的全局帧模块是以每一帧的点云信息作为整体，

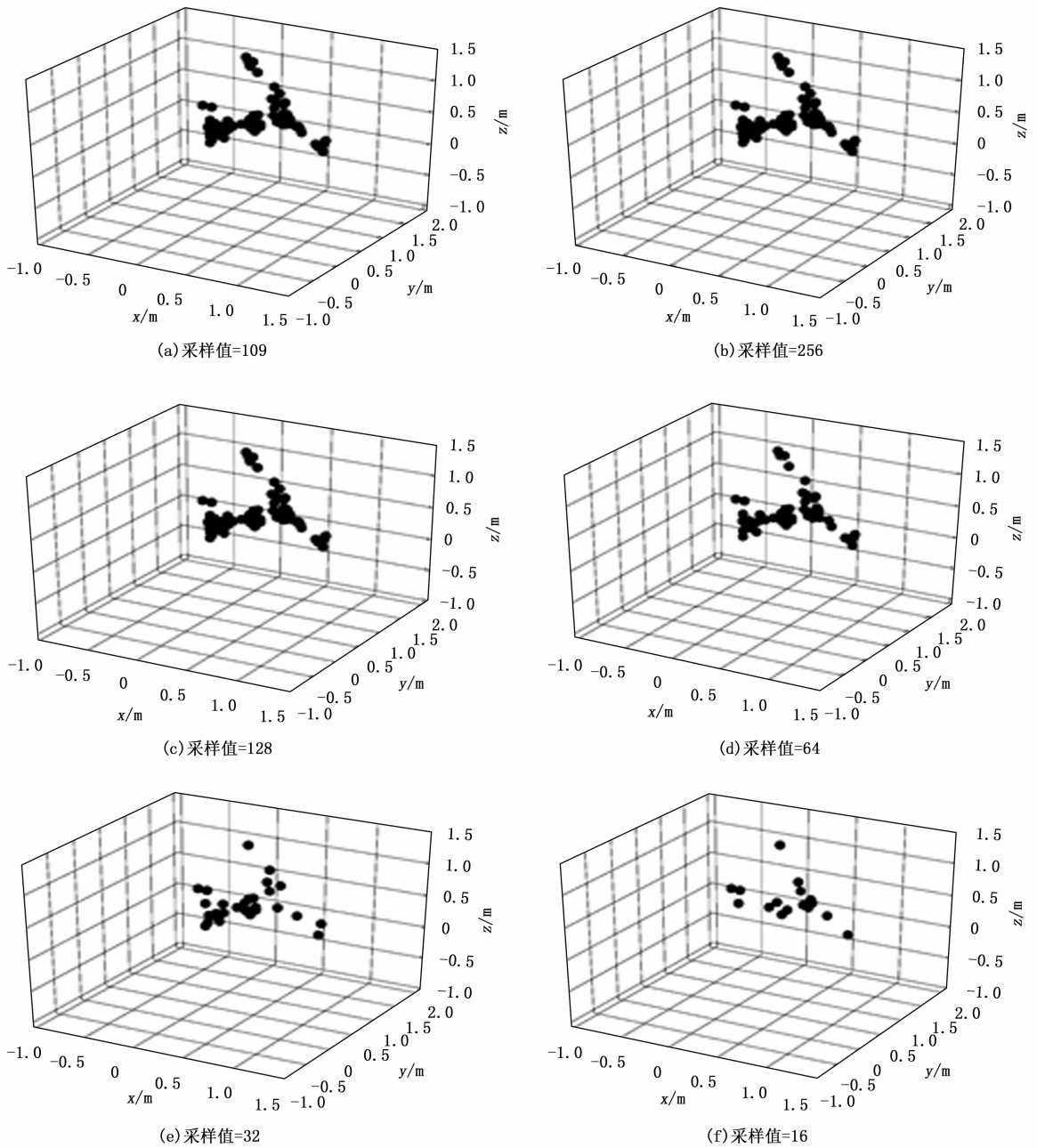


图 3 点云采样效果图

提取帧与帧之间的时序信息, 但是由于点云的特殊性, 不仅要关注帧与帧之间的时序信息, 还要关注每一帧点云中每个点在整个时间序列当中的时序信息, 这样有助于模型获取每个点在整个行为过程中的一个特征变化, 从而进一步的提取到更有用的点特征。受 PointLSTM^[20] 的启发, 通过共享的 LSTM 层对每个点云的时序特征进行提取。

假设包含着行为信息的一个点云序列有 T 帧, 每一帧可以表示为 $(P^{(1)}, P^{(2)}, \dots, P^{(T)})$, 并且每帧包含着相同的点云个数 $P^{(t)} = \{p_i^{(t)} \mid i=1, 2, \dots, n_t\}$ 。每一个 $p_i^{(t)}$ 都由两部分组成, 一部分是空间坐标向量 $x_i^{(t)}$, 一部分是 m 维

的特征向量 $f_i^{(t)}$, $N_{\Delta t}(x_i^{(t)})$ 表示 $P^{(t+\Delta t)}$ 帧中 $p_i^{(t)}$ 的相邻点, 结合式 (4) 可以将 LSTM 的计算公式表示为:

$$h^{(t)}, c^{(t)} = LSTM(y^{(t)}, h^{(t-1)}, c^{(t-1)}) \quad (5)$$

其中: 在不同帧中每个点都有着独立的神经元输出 $h_i^{(t)}$ 和细胞状态 $c_i^{(t)}$, 对于每一个组当前点和前一帧的相邻点 $(p_i^{(t)}, p_j^{(t-1)})$, 其中 $p_j^{(t-1)} \in N_{-1}(x_i^{(t)})$, 可以将 LSTM 的更新表示为式 (6):

$$\begin{aligned} y_{i,j}^{(t)} &= [x_i^{(t)} - x_j^{(t-1)}; f_i^{(t)}] \\ \tilde{h}_{i,j}^{(t)}, \tilde{c}_{i,j}^{(t)} &= LSTM(y_{i,j}^{(t)}, h_j^{(t-1)}, c_j^{(t-1)}) \end{aligned} \quad (6)$$

式中, $[\cdot]$ 表示的是向量拼接, 在上式中代表着将当前点

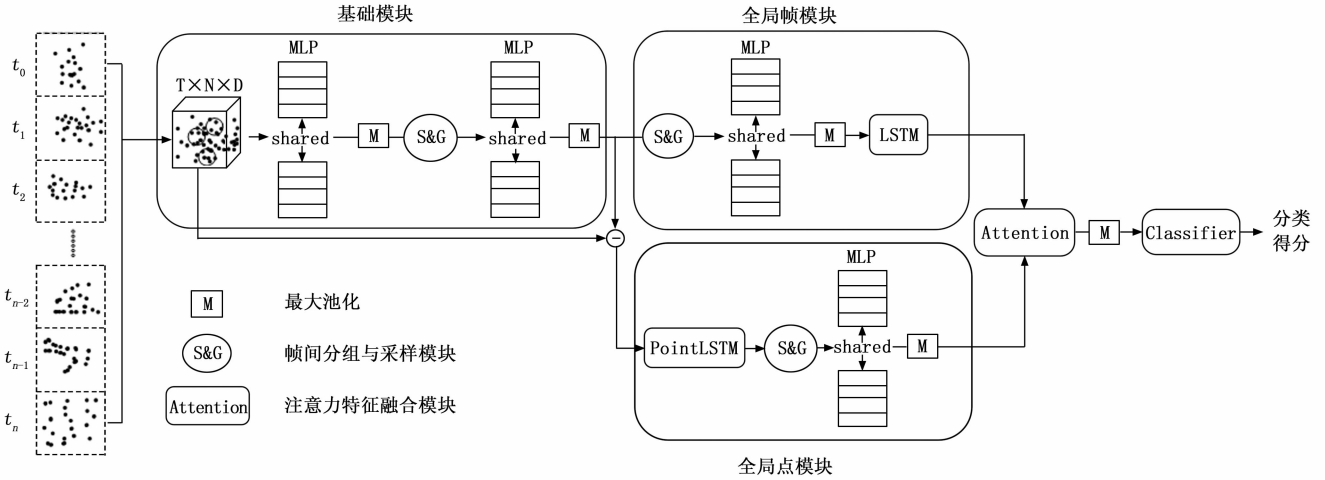


图 4 MM-PCANet 分类网络

的坐标向量和相邻点坐标向量相减之后和当前点特征向量完成拼接。 $\tilde{h}_{i,j}^{(t)}$ 和 $\tilde{c}_{i,j}^{(t)}$ 代表着当前点与不同相邻点的神经元输出和细胞状态。对于当前点最终的状态输出是将其与不同相邻点的输出进行最大池化操作得到：

$$\begin{aligned} h_i^{(t)} &= g(\tilde{h}_{i,1}^{(t)}, \tilde{h}_{i,2}^{(t)}, \dots, \tilde{h}_{i,n,t-1}^{(t)}) \\ c_i^{(t)} &= g(\tilde{c}_{i,1}^{(t)}, \tilde{c}_{i,2}^{(t)}, \dots, \tilde{c}_{i,n,t-1}^{(t)}) \end{aligned} \quad (7)$$

式中， $h_i^{(t)}$ 和 $c_i^{(t)}$ 是当前点神经元输出和细胞状态， g 表示最大池化操作。

2.2.4 注意力特征融合模块

上述通过全局帧模块和全局点模块之后，分别提取到了帧与帧之间的时序信息以及每一帧点云中点的空间和时序特征，最后利用基于注意力的特征融合模块^[21]将这两部分的特征融合。如图 5 所示，其中 MS-CAM 模块是当中提到的多尺度通道注意力模块^[21]，该模块是为了更好地融合语义和尺度不一致的特征，解决了融合不同尺度给出的特征时出现的问题。同时针对初始特征集成对于注意力特征融合影响比较大的问题，使用额外一层注意力特征融合模块生成更好的初始特征，故称之为迭代注意力特征融合模块。具体的计算方法如下：

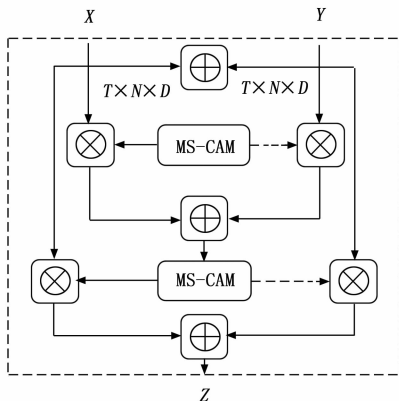


图 5 迭代注意力特征融合模块

$$Z = M(X \cup Y) \otimes X + (1 - M(X \cup Y)) \otimes Y$$

$$X \cup Y = M(X + Y) \otimes X + (1 - M(X + Y)) \otimes Y \quad (8)$$

式中， M 是 MS-CAM 多尺度通道注意力模块^[21]， X 和 Y 是两个维度 $T \times N \times D$ 的输入特征， Z 是维度 $T \times N \times D$ 的融合特征。

2.2.5 损失函数

多分类交叉熵 (Multiclass Cross-entropy) 是一种常用的损失函数，用于衡量多分类问题中模型输出的概率分布与真实标签的差距，具体计算方法如下：

$$L = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^K y_{ic} \log(p_{ic}) \quad (9)$$

式中， N 代表样本数， K 代表类别数， y_{ic} 代表真实标签中第 i 个样本属于类别 c 的概率， p_{ic} 代表模型预测的第 i 个样本属于类别 c 的概率。

3 实验

3.1 数据集

在 MMActivity 数据集^[12]和 MMGesture 数据集^[14]两个毫米波雷达人体数据集上进行了验证，从而来证明所提方法的鲁棒性和通用性。

1) MMActivity 数据集：该数据集是第一个使用毫米波雷达收集人体行为点云信息的数据集，使用 IWR 1 443 毫米波雷达采集来自两个不同实验者的 5 种人体行为数据，5 种行为分别是跳跃、拳击、行走、蹲起和开合跳跃，每个行为在两秒内完成，采样频率是每秒 30 帧，一共收集了 93 min 的数据。

2) MMGesture 数据集：该数据集是第一个使用毫米波雷达收集手势点云信息的数据集，使用的是 IWR 1 443 毫米波雷达，将在其远程方案的数据当中进行实验，一共有 79 个不同性别的参与者执行 4 种手势，包含左滑动，右滑动，敲击和旋转，其中不仅直接感应了手势动作，而且在有某些堵塞物的情况下进行了感测，收集的点云包含空间坐标 x 、 y 和 z 、速度以及反射强度 5 个特征信息。

3.2 训练和推理

该工作在单块的 3090TI 显卡上完成模型的训练和推理, 其余工作是在具有 Intel i7-10750H 的处理器以及 16GB 运行内存的笔记本电脑上完成, 使用了 PyTorch 深度学习框架实现所有模型。设置初始学习率为 0.0001, 并采用衰减率为 0.005 的阶梯式学习率衰减的 Adam 优化器^[22]来训练模型, 训练集和测试集的比例为 8:2, 采用的是多分类交叉熵损失函数。其中上述基础模块里两个 MLP 层输出的特征维度分别为 64 和 128, 全局帧模块和全局点模块里 MLP 层输出的特征维度分别为 256 和 1024, LSTM 的层数为 2 且单向, 输入输出的特征维度分别为 256 和 1024, PointLSTM 输入输出的特征维度分别为 132 和 256。在 3 个帧间分组与采样模块中, 分别选择相邻的 4、8、4 个点进行分组, 并且点云采样率为 0.5。

3.3 基线

本小节将所提出的 MM-PCANet 划分为 3 个主要部分作为基线, 以此来验证网络每一部分的特征提取能力以及所做出的改进。

1) 基础模块+全局帧模块 (Baseline A): 在该基线中选择保留了基础模块和全局帧模块, 并且将注意力特征融合模块删除, 其余保持不变。该基线可以作为参照以此来证明全局帧模块的作用。

2) 基础模块+全局点模块 (Baseline B): 在该基线中选择保留了基础模块和全局点模块, 并将注意力特征融合层去掉, 其余保持不变。该基线可以作为参照以此来证明全局点的作用。

3) 基础模块+全局帧模块+全局点模块+特征拼接 (Baseline C): 在该基线中选择保留了基础模块、全局帧模块以及全局点模块, 并选择直接用特征拼接的方法代替注意力特征融合层, 其余保持不变, 以此来证明全局帧模块和全局点模块的互补作用以及注意力特征融合层的作用。

3.4 指标

通过以下 4 个指标来评判所提出模型的可行性。

准确率: 准确率是最常见的评价指标, 指的是被对分的样本数除以所有的样本数, 通常来说, 正确率越高, 分类器越好。

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (10)$$

精确率: 又称为查准率, 指的是在模型预测结果为正样本中, 真正样本所占的百分比。

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

混淆矩阵: 主要用于比较分类结果和实际测得值, 可以把分类结果的精度显示在一个混淆矩阵里面, 并直观地表明所提出模型在哪一类样本里面表现得不是很好。

推理速度: 神经网络执行一次前向传播所需要的时间是推理时间。通常用一秒钟内模型能够执行的推理次数来表示模型的推理速度, 单位用 fps 表示。

4 结果

4.1 消融实验

为了验证 MM-PCANet 网络每个模块的有效性以及所作出的改进, 利用 K-means++ 方法将点云采样至 64 个点, 分别在 MMAActivity 和 MMGesture 两个数据集上进行了实验。如图 6 和图 7 所示, 分别按照 3.3 所提出的 3 个基线进行了测试, 其中 Baseline A 在两个数据集的准确率分别达到了 93.6% 和 77.6%, 说明所提出的基础模块和全局帧模块是具备一定的序列点云特征提取能力。Baseline B 在两个数据集的准确率分别达到了 95.3% 和 85.3%, 证明了全局点模块的特征提取效果要略微强于全局帧模块, 能够更准确地完成人体行为的识别。Baseline C 在两个数据集的准确率分别达到了 96.5% 和 89.5%, 该基线将全局帧模块和全局点模块同时利用起来, 并通过直接拼接的方式进行特征融合, 效果要强于 Baseline A 和 Baseline B, 证明全局帧模块和全局点模块具有互补作用, 不仅提取了帧与帧之间的序列信息同时也很好地提取到了帧中点云的时序的信息。用注意力特征融合模块代替了特征的直接拼接, 形成了所提出的 MM-PCANet 模型, 在两个数据集的准确率分别达到了 97.50% 和 94.10%, 直接说明了注意力特征融合模块的有效性, 能够更好地融合全局帧模块和全局点模块所输出的特征信息。

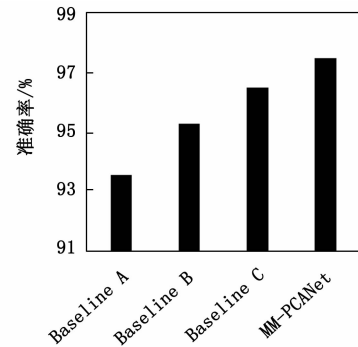


图 6 在 MMAActivity 数据集上的表现

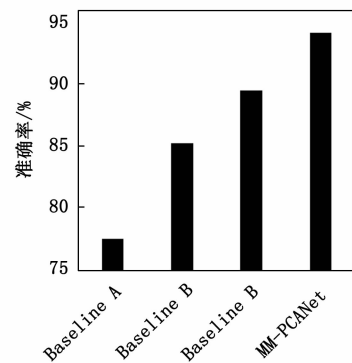


图 7 在 MMGesture 数据集上的表现

4.2 点云采样实验

为了验证 K-means++ 采样方法的有效性, 在两个数

数据集上分别利用 K-means++ 采样和随机采样在不同采样值的情况下进行了实验,如图 8 所示,上方的两条线分别代表 MM-PCANet 在 MMAActivity 数据集上两种采样方法的表现,下方的两条线分别代表在 MMGesture 数据集上两种采样方法的表现,可以得出,在不同采样值下 K-means++ 采样方法比随机采样方法准确率平均提升了 0.4 个百分点,证实了所提出的 K-means++ 采样方法的有效性。

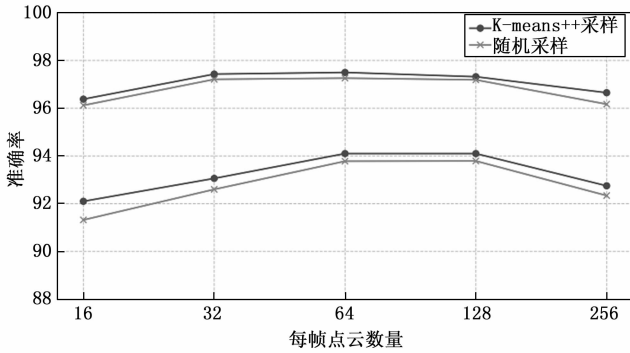


图 8 不同点云采样值在两个数据集上的表现

4.3 方法对比实验

从表 1 中可以看出 8 种不同方法在 MMAActivity 数据集上的表现,其中不同的模型点云的表示方式不同,将点云体素化表示的 RadHAR^[12] 准确率表现不是很好,随着体素分辨率的增加,该方法的空间成本和计算复杂性呈立方增长,同时点云的密度不均匀也会造成体素分布的不均匀,导致空间特征的丢失,在稀疏点云的数据集上这种问题会更加严重。除此之外是传统的直接对点云处理网络 PointNet^[17]、PointLSTM^[20] 以及 Pantomime^[23] 和用图表示点云的 MMPointGNN^[13] 分类网络,分类准确率均不如所提出的 MM-PCANet,该准确率达到了 97.50%,是表 1 所有方法当中表现最好的,造成这种现象更多的是因为上述 4 个基线网络无法兼顾帧序列的时序特征和具体点云的时序特征,而 MM-PCANet 可以很好地解决这个问题。如图 9 所示 MM-PCANet 在 MMAActivity 数据集的混淆矩阵可以看出整体的分类效果是非常好的,尽管由于跳跃和行走之间的相似性,两者有点混淆,但并不影响 MM-PCANet 的整体优良表现。

表 1 不同方法在 MMAActivity 数据集上的表现

模型	点云表示	准确率/%	精确率/%
RadHAR	体素	90.47	90.98
PointNet+LSTM	点云	93.37	93.62
PointLSTM	点云	94.97	95.27
Pantomime	点云	94.19	94.95
MMPointGNN	图	96.97	97.20
MM-PCANet	点云	97.50	97.78

从图 10 中可以看出,由于独特的运动轨迹,旋转手势具有最高的准确性,不容易被误判。其他 3 个手势的准确度不够高,因为这 3 种手势的运动模型相似。表 2 是不同方



图 9 MM-PCANet 在 MMAActivity 数据集的混淆矩阵

法在 MMGesture 数据集上的表现,其中 PointNet+LSTM 的准确率仅为 70.41%,这是因为受限于 PointNet 的空间特征提取能力,在较为稀疏的手势数据集上表现较差。PointLSTM 和 MMPointGNN 的准确率分别为 90.35% 和 92.67%,而 MM-PCANet 有着更高的准确率 94.10%,表明所提出的 MM-PCANet 在基于毫米波雷达稀疏点云的手势识别场景有着更好的特征提取能力。

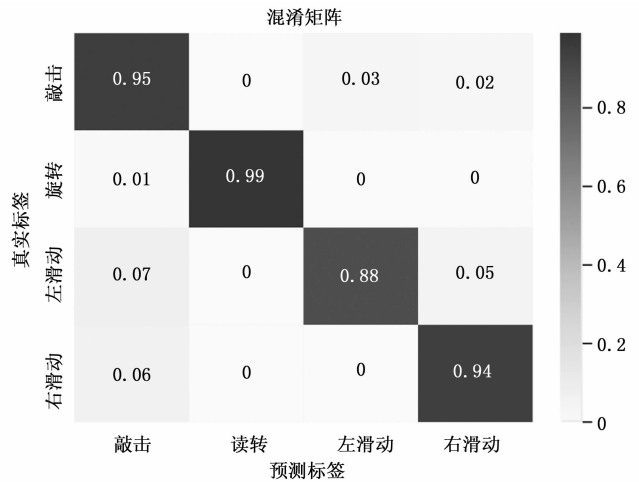


图 10 MM-PCANet 在 MMGesture 数据集的混淆矩阵

表 2 不同方法在 MMGesture 数据集上的表现

模型	点云表示	准确率/%	精确率/%
PointNet+LSTM	点云	70.41	70.47
PointLSTM	点云	90.35	90.58
MMPointGNN	图	92.67	92.69
MM-PCANet	点云	94.10	94.26

4.4 推理速度

如表 3 所示,分别在两个数据集上测试了 MM-PCANet 的平均推理速度,可以看出 Batch Size 越大,网络的平均推理速度也越快,在 Batch Size 为 1 时,在这两个数据集上的平均推理速度分别达到了每秒 19.40 次和每秒 50.02 次,也完全满足使用毫米波雷达进行人体感知的实时性。

表 3 不同 Batch Size 下 MM-PCANet 的平均推理速度 fps

数据集	Batchsize-1	Batchsize-8	Batchsize-16
MMActivity	19.40	128.85	212.83
MMGesture	50.02	273.76	489.70

5 结束语

本文提出了一种通用的基于毫米波雷达稀疏点云的人体行为识别方法, 该方法直接对人体原始点云进行处理, 无需将点云进行体素化或者其它方式的表示, 利用 K-means+ 算法进行点云采样, 在不破坏点云空间结构的同时有效地提高了模型的计算效率。此外, 使用 MM-PCANet 进行人体行为特征的提取和识别, 兼顾点云的空间特征以及时序特征, 对稀疏点云的运动有着灵敏的感知能力, 同时也具有较强的抗干扰能力。实验结果表明该方法在两个毫米波雷达人体点云数据集 (MMActivity 和 MMGesture) 上的表现均优于所有其它基线, 证明所提出方法的鲁棒性和通用性。但由于要融合多方面的特征, 导致网络结构略显复杂, 因此未来会尝试将网络变得更轻量化。

参考文献:

[1] WANG Y, WU K, NI L M. WiFall: device-free fall detection by wireless networks [J]. *IEEE Transactions on Mobile Computing*, 2017, 16 (2): 581-94.

[2] LI W, ZHANG D, LI Y, et al. Real-time fall detection using mmwave radar [C] // *ICASSP 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022: 16-20.

[3] JIN F, SENGUPTA A, CAO S. MmFall: fall detection using 4-D mmwave radar and a hybrid variational RNN autoencoder [J]. *IEEE Transactions on Automation Science and Engineering*, 2022, 19 (2): 1245-57.

[4] CIPPITELLI E, FIORANELLI F, GAMBI E, et al. Radar and RGB-depth sensors for fall detection: a review [J]. *IEEE Sensors Journal*, 2017, 17 (12): 3585-604.

[5] WITHANAGE K I, LEE I, BRINKWORTH R, et al. Fall recovery subactivity recognition with RGB-D cameras [J]. *IEEE Transactions on Industrial Informatics*, 2016, 12 (6): 2312-20.

[6] 岳彬, 余大兵, 常心悦, 等. 基于智能视觉物联网的油库人员行为识别与监测系统 [J]. *电子测量技术*, 2020, 43 (3): 128-131.

[7] THARIQ AHMED H F, AHMAD H C V A. Device free human gesture recognition using Wi-Fi CSI: a survey [J]. *Engineering Applications of Artificial Intelligence*, 2020, 87 (1): 1-19.

[8] YU C, XU Z, YAN K, et al. Noninvasive human activity recognition using millimeter-wave radar [J]. *IEEE Systems Journal*, 2022, 16 (2): 3036-47.

[9] TIWARI G, GUPTA S. An mmwave radar based real-time contactless fitness tracker using deep CNNs [J]. *IEEE Sensors Journal*, 2021, 21 (15): 17262-17270.

[10] KIM Y, ALNUJAIM I, OH D. Human activity classification based on point clouds measured by millimeter wave MIMO radar with deep recurrent neural networks [J]. *IEEE Sensors Journal*, 2021, 21 (12): 13522-13529.

[11] ZHANG R, CAO S. Real-time human motion behavior detection via CNN using mmwave radar [J]. *IEEE Sensors Letters*, 2019, 3 (2): 1-4.

[12] SINGH A D, SANDHA S S, GARCIA L, et al. RadHAR: human activity recognition from point clouds generated through a millimeter-wave radar [C] // *Proceedings of the 3rd ACM Workshop on Millimeter-wave Networks and Sensing Systems - mmNets'19*, 2019: 51-56.

[13] GONG P, WANG C, ZHANG L. MMPoint-GNN: graph neural network with dynamic edges for human activity recognition through a millimeter-wave radar [C] // *2021 International Joint Conference on Neural Networks (IJCNN)*, 2021: 1-7.

[14] LIU H, ZHOU A, DONG Z, et al. M-Gesture: person-independent real-time in-air gesture recognition using commodity millimeter wave radar [J]. *IEEE Internet of Things Journal*, 2022, 9 (5): 3397-415.

[15] AHMAD A, ROH J C, WANG D, et al. Vital signs monitoring of multiple people using a FMCW millimeter-wave sensor [C] // *Proceedings of the 2018 IEEE Radar Conference (RadarConf18)*, 2018: 1450-1455.

[16] QI C R, YI L, SU H, et al. PointNet++: deep hierarchical feature learning on point sets in a metric space [C] // *Proceedings of the 31st Annual Conference on Neural Information Processing Systems (NIPS)*, 2017: 5105-5114.

[17] QI C R, SU H, MO K, et al. PointNet: deep learning on point sets for 3D classification and segmentation [C] // *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017: 77-85.

[18] MIN Y, CHAI X, ZHAO L, et al. FlickerNet: adaptive 3D gesture recognition from sparse point clouds [C] // *Proceedings of the British Machine Vision Conference*, 2019: 105.

[19] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. *Neural Comput.*, 1997, 9 (8): 1735-80.

[20] MIN Y, ZHANG Y, CHAI X, et al. An efficient point LSTM for point clouds based gesture recognition [C] // *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020: 5760-5769.

[21] DAI Y M, GIESEKE F, OEHMCKE S, et al. Attentional feature fusion [C] // *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2021: 3559-3568.

[22] KINGMA D P, BA J. Adam: a method for stochastic optimization [J]. *Computer Science*, 2014 (12): 69-80.

[23] PALIPANA S, SALAMI D, LEIVA L A, et al. Pantomime: mid-air gesture recognition with sparse millimeter-wave radar point clouds [J]. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2021, 5 (1): 1-27.