DOI:10.3969/j.issn.1672-2337.2024.05.012

基于毫米波雷达的三维特征自适应融合 人体动作识别算法

李程辉¹,蒋俊正^{2,3},周 芳⁴

(1.桂林电子科技大学信息与通信学院,广西桂林 541004; 2.西安电子科技大学杭州研究院,浙江杭州 311231;3.西安电子科技大学电子工程学院,陕西西安 710071; 4.中国计量大学信息工程学院,浙江杭州 310018)

摘 要:目前基于毫米波雷达的人体动作识别方法存在人体动作特征利用不足导致识别精度低,与多维特征融合方式简单导致模型参数量增加的问题。针对以上问题,提出了一种基于毫米波雷达的三维特征自适应融合人体动作识别方法。首先,对雷达回波信号进行时频分析以获取人体动作距离、多普勒与角度特征,并在时域上拼接构建三维特征数据集。然后,设计了一种带有特征自适应融合器的三分支卷积神经网络,实现对三维特征数据集的高维抽象特征提取与多维特征的自适应融合。最后,通过活动分类器得到人体动作检测结果。实验结果表明,所提方法的人体动作识别平均准确率可达到96.41%,受试者工作特征曲线下面积(Area Under the Curve, AUC)值可达到0.984.优于单维特征识别方法与多种三维特征融合方法。

关键词:毫米波雷达;动作识别;时频分析;特征融合;卷积神经网络

中图分类号:TN957.51 文献标志码:A 文章编号:1672-2337(2024)05-0569-09

引用格式:李程辉,蒋俊正,周芳.基于毫米波雷达的三维特征自适应融合人体动作识别算法[J].雷达科学 与技术,2024,22(5):569-577.

LI Chenghui, JIANG Junzheng, ZHOU Fang. Human Motion Recognition Algorithm with Three-Dimensional Feature Adaptive Fusion Based on Millimeter-Wave Radar [J]. Radar Science and Technology, 2024, 22(5):569-577.

Human Motion Recognition Algorithm with Three-Dimensional Feature Adaptive Fusion Based on Millimeter-Wave Radar

LI Chenghui¹, JIANG Junzheng^{2,3}, ZHOU Fang⁴

(1. School of Information and Communication, Guilin University of Electronic Technology, Guilin 541004, China;

2. Hangzhou Institute of Technology, Xidian University, Hangzhou 311231, China;

3. School of Electronic Engineering, Xidian University, Xi'an 710071, China;

4. College of Information Engineering, China Jiliang University, Hangzhou 310018, China)

Abstract: At present, the human motion recognition methods using millimeter-wave radar exist the problems of low recognition accuracy caused by underutilization of human motion features and the increased number of model parameters due to the simple fusion method with multi-dimensional features. In view of the above problems, a three-dimensional feature adaptive fusion method for human motion recognition based on millimeter-wave radar is proposed. First, the time-frequency analysis of radar echo signals is carried out to obtain the features of human action distance, Doppler and angle, and the 3D feature dataset is constructed by accumulating them in the time domain. Second, a three-branch convolutional neural network with a feature adaptive fusion device is designed to realize the extraction of high-dimensional abstract features from 3D feature dataset and the effective fusion of multi-dimensional features. Finally, the human motion detection results are obtained by the activity classifier. The experimental results show that the average accuracy of human action recognition of the proposed method can reach 96.41%, and the area under the curve (AUC) value can reach 0.984, which is better than the single-dimensional feature recognition method and a variety of three-dimensional feature fusion methods.

收稿日期: 2024-03-25; 修回日期: 2024-04-28

基金项目:国家自然科学基金资助项目(No.62261014, 62171146);广西创新驱动发展专项(No.桂科 AA21077008);广西自然科学杰 出青年基金项目(No.2021GXNSFFA220004);广西科技基地和人才专项(No.桂科 AD21220112);广西研究生教育创新计 划项目(No.YCBZ2023137)

Key words: millimeter-wave radar; human motion recognition; time-frequency analysis; feature fusion; convolutional neural network

0 引 言

随着智能设备的普及与人工智能技术的快速 发展,人机交互技术变得愈加重要。人体动作识 别技术作为推动人机交互发展的关键技术之一, 在智能家居、健康监测与护理、安防与监控等诸多 领域具有重要应用价值,成为近年来国内外的研 究热点^[1]。

目前,用于人体动作识别的设备可分为接触 式与非接触式两种。常见的接触式设备包括智能 手机、手表或由加速度计、振动陀螺仪等传感器组 合而成的人体动作检测装置[24]。然而,接触式设 备需要直接接触或佩戴在用户身上,这往往会对 用户的自由度与舒适度造成一定的限制。因此, 许多研究人员将目光转向了非接触式的人体动作 识别方式。通常可以使用 RGB 视频、WiFi 或毫米 波雷达实现非接触式人体动作识别。基于视频的 人体动作识别技术发展成熟,其一般采用不同的 深度学习技术,对RGB视频或人体骨架数据进行 处理以实现人体动作的识别[5-6]。但是基于视觉的 人体动作识别方法对光线条件的依赖度高,并且 存在泄露用户隐私的风险。基于WiFi的方法具有 成本低、易于部署、保护隐私的优点[7-8]。然而, WiFi容易受到其他无线信号的影响,并且它的分 辨率较低,稳定性也较差。相比之下,毫米波雷达 有着分辨率高、设备尺寸小以及不依赖光源的特 点,为非接触式人体动作识别提供了一种可行的 解决方案。同时,深度学习技术的快速发展推动 了基于雷达的人体动作识别技术的进步。文献 [9]系统介绍了调频连续波(Frequency Modulated Continuous Wave, FMCW) 雷达的基本工作原理与 基于离散傅里叶变换的信号处理技术。文献[10] 使用毫米波雷达采集人体的点云信息,并选取其 中一点作为质心点云,然后根据所获取的点云计 算人体运动的异常水平,当异常水平的尖峰与质 心点云高度的下降同时发生时,则检测为跌倒。 文献[11]与文献[12]根据人体动作微多普勒特 征,分别使用光学字符识别技术以及深度卷积神 经网络(Deep Convolutional Neural Network, DCNN) 实现了对多种人体动作的识别。文献[13]将人体 动作距离特征与微多普勒特征进行融合得到双流 融合特征,并将其输入至 Bi-LSTM 网络以实现对不 同人体动作的识别。文献[14]提出了一种多分支 卷积神经网络结构,用于提取手势目标的距离、多 普勒与角度参数的高维特征,并采用特征拼接融 合的方式对手势进行识别。以上工作说明了深度 学习技术应用于雷达人体动作识别的可行性。然 而,现有基于毫米波雷达的人体动作识别方法存 在以下问题:第一,因为使用毫米波雷达生成的点 云数据具有稀疏性与随机性,所以基于点云的方 法难以对较为轻微的动作进行识别;第二,对人体 动作特征信息的利用不够充分,例如文献[11]与 文献[12] 仅利用了人体动作的微多普勒特征,文 献[13]仅利用了人体动作的距离特征与多普勒特 征,两者都缺乏对人体动作角度特征的利用;第 三,对多维特征融合的方式较为简单,文献[14]采 用特征拼接的方式对多维特征进行融合,而这种 方式会导致模型参数量以及计算复杂度的增加, 同时容易引入过拟合与网络冗余信息增加的问题。

针对上述问题,本文提出了基于毫米波雷达 的三维特征自适应融合人体动作识别算法。首 先,采用二维快速傅里叶变换(2-Dimensional Fast Fourier Transform, 2D-FFT) 与多信号分类(Multiple Signal Classification, MUSIC)算法^[15]对雷达回波信 号处理得到人体动作距离、多普勒与角度特征,并 在时间上拼接构建三维特征数据集。其次,为了 能够同步提取三维特征数据集中的高维抽象特 征,设计了一种三分支卷积神经网络结构,解决了 特征输入维度不足的问题。然后,提出了一种特 征自适应融合方案,解决了由特征拼接融合方法 带来的模型参数量与计算复杂度增加的问题,减 少了模型过拟合的风险。最后通过活动分类器获 得识别结果。为了证明本文方法的有效性,将其 与单维特征识别方法、多维特征拼接融合与平均 融合方法进行对比。实验结果表明,采用多维特 征自适应融合方法的网络具有更高的人体动作识

别精度与AUC值。

算法描述 1

本文提出了基于毫米波雷达的三维特征自适 应融合人体动作识别系统。如图1所示,该系统主 要包括FMCW雷达信号处理、多维特征数据集构 建与三维特征自适应融合卷积神经网络(Three-Dimensional Feature Adaptive Fusion Network, 3D -FAFNet)设计部分。首先,采用距离维-FFT与多普 勒维-FFT对雷达回波信号进行处理以获取人体动 作距离、多普勒特征,采用MUSIC算法获取人体动 作角度特征,并在时域上拼接构建三维特征数据 集。其次,设计了 3D-FAFNet 网络,实现对不同人 体动作的识别。



图1 人体动作识别系统框图

1.1 FMCW 雷达信号处理

本文采用FMCW毫米波雷达进行信号采集, 其系统框图如图2所示。雷达系统从发射天线发 射LFM 信号x_T(t),表示为^[9]

$$x_{\mathrm{T}}(t) = A_{\mathrm{T}} \cos\left(2\pi f_{c}t + \pi St^{2}\right) \tag{1}$$

$$S = B/T_{a}$$
(2)

式中, A_{T} 为发射信号的幅度, f_{s} 为起始频率,S为调 频斜率,B为带宽,T。为扫频周期。





当雷达前方存在目标时,部分信号能量被反 射至接收天线,得到接收信号x_R(t),表示为

 $x_{\rm R}(t) = A_{\rm R} \cos(2\pi f_{\rm c}(t-\tau) + \pi S(t-\tau)^2) \quad (3)$ 式中 $A_{\rm B}$ 为接收信号的幅度, τ 为时延。然后,将发 射信号与接收信号进行混频得到混频信号 $x_{\mu}(t)$ = $x_{T}(t) \cdot x_{R}(t)$,同时使用低通滤波器(Low-Pass Filter, LPF)滤除混频信号的高频分量,得到中频信号为

$$x_{\rm IF} = \frac{1}{2} A_{\rm T} A_{\rm R} \cos(2\pi f_{\rm e} \tau + 2\pi S \tau t - \pi S \tau^2)$$
(4)

可知中频信号的频率与初始相位分别为

$$f_{\rm IF} = S\tau \tag{5}$$

$$\phi_{\rm IF} = 2\pi f_{\rm e} \tau - \pi S \tau^2 \approx 2\pi f_{\rm e} \tau \tag{6}$$

假设人体动作目标与雷达之间的距离为R,结 合公式(5)可得

$$R = \frac{c\tau}{2} = \frac{c}{2S} f_{\rm IF} \tag{7}$$

式中,c=3×10⁸ m/s为光速。

为了估计目标的速度,雷达会发射多个时间 间隔为T。的线性调频脉冲信号。若目标的速度为 v,则在时间间隔 T_c 内目标的位移 $\Delta R = vT_c$ 。结合 公式(6)、(7)知相邻线性调频信号之间的相位差 $\Delta \phi = 4\pi \Delta R / \lambda = 4\pi v T_c / \lambda$,则目标速度为

$$v = \frac{\lambda \Delta \phi}{4\pi T_{\rm c}} \tag{8}$$

式中, $\lambda = clf_c$ 为信号波长。

如图3所示,可利用多个天线之间的雷达回波 信号的相位差测量目标的角度信息。由于天线之 间存在间隔,到达两个相邻接收天线的信号所经 历的相对延迟可以被确定为 $\Delta \tau = d \sin \theta / c$,其中,d 为天线间距,θ为接收信号的到达角。结合公式 (6)可得两个相邻天线处所接收信号的相位差为

 $\Delta \varphi = 2\pi f_c \Delta \tau = 2\pi d \sin \theta / \lambda$ (9)则目标角度为

$$\theta = \sin^{-1}\left(\frac{\lambda\Delta\varphi}{2\pi d}\right) \tag{10}$$

1.2 数据集构建

为了获取人体动作目标相对于雷达距离、多 普勒与角度随时间变化的信息,本文对雷达回波



图3 角度估计原理图

信号进行时频分析以提取人体动作目标在每一帧 上的距离、多普勒与角度特征,在时间上拼接分别 得到距离-时间图(Range-Time-Map, RTM)、多普勒-时间图(Doppler-Time-Map, DTM)以及角度-时间图 (Angle-Time-Map, ATM)。

本文以48帧数据为观测时长,每一帧的持续时间为40ms。人体动作目标在一帧时间内相对 于雷达的位置近乎不变。在一帧数据内,对每个 扫频信号执行 Range-FFT 求出其对应的距离估计 谱,然后在帧内求平均得到一帧数据的人体动作 目标距离特征,对多帧数据的距离特征进行拼接 可得到 RTM。在得到一帧数据内所有扫频信号的 距离估计谱后,在慢时间维上执行 Doppler-FFT 得 到距离-多普勒联合估计结果,通过谱峰搜索提取 一帧数据的人体动作目标多普勒特征,经过拼接 多帧数据的多普勒特征后得到 DTM。本文采用的 雷达具有 2个发射天线以及4个接收天线,通过时 分复用技术可构建8路信号向量矩阵,从而可以利 用 MUSIC 算法提取人体动作目标的角度特征,将 多帧数据的角度特征拼接后得到ATM。

本文选取的人体动作包括站立(Stand)、靠近 雷达行走(Walk Close to the radar, WC)、远离雷达 行走(Walk Away from the radar, WA)、平行雷达行 走-从左至右(Walk Parallel to the radar-Left to Right, WP-LR)、平行雷达行走-从右至左(Walk Parallel to the radar-Right to Left, WP-RL)、下蹲(Squat)共6 类。为了使数据集更加多样化,实验邀请了年龄、 身高和体重各不相同的11名志愿者进行数据采 集。每名志愿者重复执行每种动作50次,每种人 体动作采集550次,共计3300组数据。在图4中, 本文给出了完整的人体动作图以及不同人体动作 对应的RTM、DTM、ATM结果,易知,不同人体动作 对应的 RTM、DTM、ATM 表现出了不同的物理特 征。例如,在靠近雷达行走动作中,人体动作目标 不断接近雷达,故其对应的RTM呈现为一条逐步 上升的曲线,DTM的能量主要集中在上半区域。 而远离雷达行走动作对应的 RTM、DTM 结果则呈 现了相反的变化趋势。因为靠近雷达行走动作与 远离雷达行走动作的行走路线接近为一条直线, 故这两个动作的ATM结果波动较小,类似于一条 直线。

1.3 3D-FAFNet 网络设计

卷积神经网络能够通过局部感知性捕捉图像



边缘、纹理等局部特征,并采用堆叠多个卷积层与 池化层的方式逐层提取图像数据的高维抽象特 征,在图像数据处理方面具有巨大的优势。为了 有效提取 RTM、DTM 与 ATM 数据的高维抽象特 征,并实现对多维特征的自适应融合,本文设计了 一种三维特征自适应融合卷积神经网络。如图 5 所示,3D-FAFNet由特征提取器、特征融合器和活 动分类器组成。其中,3D-FAFNet的特征融合器采 用的是特征自适应融合方法。



图 5 3D-FAFNet 结构图

1.3.1 特征提取器

3D-FAFNet的特征提取器由3个相同的特征 提取模块(Feature Extraction Module, FEM)构成, 是一种三分支卷积神经网络结构。其中,3个特征 提取模块的输入依次为 RTM、DTM 以及 ATM 数 据,每个特征提取模块都是由 MobileNetV3-Small¹⁶网络改进而来,将其中的SE通道注意力机 制^[17]更改为ECA通道注意力机制^[18],目的在于避 免SE通道注意力机制中的降维处理对通道注意力 预测带来的副作用,更加有效地捕获局部跨通道 交互信息。如图6所示,FEM由4种Block组成。 其中,Block1与Block2中的深度卷积核(Depthwise convolution, Dwise)的步长为2, Block3与Block4中 深度卷积核的步长为1。当步长为1时,输出图像 的长和宽与输入图像相同;当步长为2时,输出图 像的长和宽缩短为原来的一半。Block2与Block4 包含ECA通道注意力机制,而Block1与Block3则 不包含。每个 Block 中的 Activation 表示激活函数 的类型,M×M则表示深度卷积核的尺寸大小。 1.3.2 特征融合器

特征融合器能够融合所提取的RTM、DTM、 ATM高维抽象特征,可以更充分地利用不同的人 体动作特征,获得更丰富的特征表示,如图7所示, 不同的特征融合方法拥有不同的特点,每种特征 融合器的输入为所提取的RTM、DTM、ATM高维抽



象特征**F**_{RTM}、**F**_{DTM}与**F**_{ATM}。

如图 7(a)为特征自适应融合器结构图。首先 对每个高维抽象特征使用全局平均池化得到压缩 后的特征表示z:

$$\boldsymbol{z}_i = \frac{1}{n} \sum_{j=1}^n \boldsymbol{F}_i(j), \boldsymbol{z} \in \mathbb{R}^{3 \times 1}$$
(11)

式中,n为每一类高维抽象特征的长度,i表示高维 抽象特征的类别,i = 1,2,3, $F_i(j)$ 表示第i类高维抽 象特征的第j个元素。然后采用两个带有不同激 活函数的全连接层(Fully-Connected layer, FC)获取 特征权重因子s:

 $s = \sigma(g(z,w)) = \sigma(w_2\delta(w_1z)), s \in \mathbb{R}^{3\times 1}$ (12) 式中, $w_1, w_2 \in \mathbb{R}^{3\times 3}, \delta$ 为H-Swish激活函数^[19], σ 为Sigmoid激活函数。最后,将每个高维抽象特征与其相对应的特征权重进行相乘,并以相加的方式完成特征的自适应融合,得到融合特征 F_{fusion} 为

$$\boldsymbol{F}_{\text{fusion}} = \sum_{i=1}^{3} \boldsymbol{s}_{i} \boldsymbol{F}_{i}$$
(13)

如图 7(b)所示,作为特征自适应融合方法的 消融方法,本文人工地为每个高维抽象特征赋予 1/3 的特征权重并以加权求和的方式完成特征融合, 设计了特征平均融合器,此时的融合特征为

$$\boldsymbol{F}_{\text{fusion}} = \frac{1}{3} \sum_{i=1}^{3} \boldsymbol{F}_i \tag{14}$$

易知,相比于特征自适应融合方法,特征平均融合 方法不能自适应地为不同的人体动作特征赋予不 同的权重,缺乏一定的灵活性,从而限制了网络的 特征融合能力。

如图7(c)所示,文献[14]采用特征拼接的方 法对多维特征融合,得到的融合特征为

$$F_{\text{fusion}} = [F_{\text{RTM}}, F_{\text{DTM}}, F_{\text{ATM}}]$$
(15)
式中, [•,•]为向量首尾拼接符。

卷积神经网络模型的参数量通常与模型的复杂度相关联。卷积神经网络全连接层的参数量为 *T*_{in} · *T*_{out} + *T*_{out},其中*T*_{in}为输入向量的长度,即为融 合特征的长度,*T*_{out}为输出向量的长度。可知对于 同一种网络模型,特征拼接融合方法在全连接层 产生的参数量约为特征自适应融合与特征平均融 合方法的3倍,而过多的参数量会导致网络计算复 杂度的增加,同时容易引入网络冗余信息增加与 过拟合的问题^[20]。

1.3.3 活动分类器

活动分类器由两层全连接层组成,其输入为融合特征 F_{fusion}。在第一层全连接层后引入 H-Swish激活函数,以增强模型的非线性表达能力,并在第二层全连接层后使用 Softmax 激活函数,获得每一类人体动作的检测概率:

 $P = \rho(w_4 \delta(w_3 F_{\text{fusion}})), P \in \mathbb{R}^{6 \times 1}$ 式中, $w_3 \in \mathbb{R}^{n \times m}, w_4 \in \mathbb{R}^{m \times 6}, m$ 为第一个全连接层 的输出通道数, ρ 为 Softmax 激活函数。为了与 MobileNetV3-Small 网络保持一致,本文选取 n = 576, $m = 1024_{\circ}$

本文的损失函数为交叉熵损失函数,交叉熵 能够衡量两个分布的差异程度,其公式定义如下:

$$H(P,Q) = -\sum_{k=1}^{6} P(x_k) \log(Q(x_k))$$
(17)

式中,**P**(x_k)表示真实概率分布,**Q**(x_k)表示预测概 率分布,x_k表示人体动作的类别。通过计算交叉熵 损失函数,能够得知两个分布的差异程度,之后在 反向传播过程中不断更新模型的权重参数,并计 算特征权重因子s,达到三维特征自适应融合的 目的。

2 实验分析与讨论

2.1 实验平台与环境

如图 8 所示,本文使用德州仪器(Texas Instruments, TI)公司的 IWR1642Boost 毫米波雷达开发 板以及 DCA1000 高速数据采集卡完成对数据的采 集。雷达扫频周期为 38 μs,帧周期为 40 ms,带宽 约为4 GHz。实验环境为一个空旷的会议室,雷达 距离前方墙体约 2.7 m,距离地面约 1.3 m,实验人 员在雷达与墙体之间的范围内执行不同的动作。 3D-FAFNet 网络由 Pytorch 深度学习框架搭建而 来,采用 NVIDIA 公司的 GTX 1070 显卡实现对模 型的加速训练与测试。



 (a) 实验环境
 (b) 设备

 图 8 实验环境与设备图

2.2 单维特征网络验证

为了探究不同特征对人体动作识别贡献程度 的大小,并证明三维特征数据集相较于单维特征 数据集的优势,本文将 3D-FAFNet 网络修改为单 分支卷积神经网络结构,即只保留一个特征提取 模块作为网络的特征提取器,同时去除特征融合 器。在实验中,本文将每种数据集划分为训练集、 验证集与测试集。训练集用于模型的训练与网络 参数的更新,验证集用于选择训练过程中生成的 最佳模型,测试集用于评估所选择模型的泛化能 力。其中,训练集包含2160组数据,验证集包含 540组数据,测试集包含600组数据,且训练集、验 证集与测试集中的数据都是互不重合的。为了让 模型具有更好的适应性,对于在不同时刻、不同位 置执行人体动作的实验人员仍具有较高的人体动 作识别准确率,同时为了减少模型的计算复杂度, 将图像尺寸均随机缩放为96×96,并进行灰度归一 化处理。采用带有动量与权重衰减的随机梯度下 降优化器,初始学习率设置为0.02,动量为0.9,并 基于余弦衰减的方法^[21]设置学习率调度器。

卷积神经网络在测试集上的准确率变化曲线 能够有效反映模型的泛化能力。如图9所示为经 过训练与验证后所得到的RTM、DTM、ATM最佳模 型在各自测试集上进行30轮测试的准确率变化曲 线。可见,不同数据集最佳模型在各自测试集上 的平均准确率从高到低依次为ATM、DTM、RTM。 说明ATM数据集能够提供更多的人体动作特征差 异信息,具有更好的人体动作表征能力。同时,由 ATM数据集训练与验证后保存的最佳模型具有更 强的泛化能力。



2.3 三维特征网络验证

为了证明本文提出的特征自适应融合器对网络性能的提升,本文将 3D-FAFNet 中的特征自适应融合器替换为特征平均融合器,构建三维特征平均融合网络(Three-Dimensional Feature Mean Fusion Network, 3D-FMFNet),与三维特征自适应融合网络进行比较。如图10为采用不同特征融合方法的网络经过训练与验证后得到的最佳模型在三维特征测试集上的准确率变化曲线。由图10可知,采用特征自适应融合方法的人体动作识别准确率明显高于特征平均融合方法,这表明相比于特征平均融合方法,特征自适应融合方法能够自动地为不同的人体动作特征赋予不同的权重,让网络能够更好地关注重要的特征信息并抑制不重

要的特征信息,从而有效提升网络的人体动作识 别性能。



STO 5D-IMINCI-5D-IMINCIE-研干文化画文

为了验证3D-FAFNet人体动作识别方法相比 于同类算法的优势,本文选用 MobileNetV3-Small^[16]与RDA-TNet^[14]作为对比算法。其中, MobileNetV3-Small网络为单维特征网络。而RDA-TNet 为三维特征网络, 它将简化后的 VGG16 网络 作为特征提取模块,并以特征拼接的方式实现三 维特征的融合。基于 3D-FAFNet 的构建方法,本 文将 MobileNetV3-Small 网络修改为三分支卷积神 经网络结构,并分别加入特征平均融合方法与特 征自适应融合方法以构建 MobileNet-V3平均融合 网络与MobileNetV3自适应融合网络。同时,本文 将RDA-TNet中的特征拼接融合方法分别替换为 特征平均融合方法与特征自适应融合方法,构建 RDA-T平均融合网络以及 RDA-T 自适应融合网 络,并将它们作为本文的对比算法。通常可以采 用受试者工作特征(Receiver Operating Characteristic, ROC)曲线^[22]与AUC值评估不同模型的综合性 能。如图11与图12所示,为了更加充分地评估不 同模型的泛化能力与综合性能,本文分别绘制了 RDA-TNet与其余特征自适应融合网络在三维特征 测试集上的准确率变化曲线与ROC曲线。比较可 知,3D-FAFNet的平均准确率与AUC值均高于其 他网络,说明3D-FAFNet具有更强的泛化能力与 综合性能。同时,RDA-T自适应融合网络的平均 准确率与AUC值略高于RDA-TNet,这表明相比于 特征拼接融合方法,特征自适应融合方法拥有更 好的特征融合能力。





复杂度与在单维特征测试集、三维特征测试集上 的精度比较。如表2所示为3D-FAFNet在三维特 征测试集上进行随机一次测试得到的人体动作分 类混淆矩阵。对比可知,基于三维特征人体动作 识别方法的平均准确率与AUC值均高于基于单维 特征人体动作识别方法,说明三维特征数据集能 够提供更加丰富的人体动作差异信息,体现了构 建三维特征人体动作识别网络的必要性。同时, 相比于采用三维特征平均融合方法与拼接融合方 法的不同网络结构,采用特征自适应融合方法的 网络结构的平均准确率提高了1.6%~4%, AUC值 提高了0.01~0.04,证明了本文所提特征自适应融 合方法的有效性与通用性,可以适用于不同的网 络结构。最后,本文所提出的3D-FAFNet的平均 准确率达到了96.41%, AUC值达到了0.984, 相比 于其他特征融合方法的网络,将准确率提高了 4%~6%,将AUC值提高了0.03~0.05,对比单维特 征识别方法则具有更加显著的提升,体现了3D-FAFNet的有效性与可靠性。此外,相较于RDA-TNet多分支网络结构,3D-FAFNet网络的模型参数 量与时间复杂度大幅减少,单个样本的处理时间 仅为35.2 ms,为后续的实时人体动作识别提供了 有力支撑。

网络结构	数据集	参数量/ 10 ⁶ Params	时间复杂度/ 10 ⁹ FLOPs	单个样本处 理时间/ms	平均准确率/ %	AUC值
RDA-T ^[14] 单维特征网络	RTM				84.22	0.852
	DTM	29.96	0.740	67.89	84.99	0.862
	ATM				86.69	0.898
MobileNetV3 ^[16] 单维特征网络	RTM				84.03	0.823
	DTM	1.52	0.011	35.90	85.72	0.866
	ATM				87.53	0.870
本文 单维特征网络	RTM DTM 1.06			84.97	0.844	
		1.06	0.010	31.26	86.05	0.872
	ATM				88.33	0.924
RDA-TNet ^[14]		89.67	2.214	126.62	91.22	0.958
RDA-T ^[14] 平均融合网络	三维特征 数据集	56.71	2.180	81.69	89.94	0.942
MobileNetV3 ^[16] 平均融合网络		3.37	0.032	40.94	91.07	0.934
3D-FMFNet		1.99	0.030	38.32	92.41	0.944
RDA-T ^[14] 自适应融合网络		56.71	2.180	75.44	92.81	0.968
MobileNetV3 ^[16] 自适应融合网络		3.37	0.032	37.75	93.02	0.948
3D-FAFNet		1.99	0.030	35.21	96.41	0.984

表1 不同人体动作识别方法的算法复杂度与精度对比

真实	预测类别								
类别	Stand	WA	WC	WP-LR	WP-RL	Squat			
Stand	99	0	0	1	0	0			
WA	5	92	0	0	1	2			
WC	0	0	94	6	0	0			
WP-LR	1	0	0	99	0	0			
WP-RL	0	0	0	3	93	4			
Squat	2	0	0	0	0	98			

表2 人体动作混淆矩阵

3 结束语

本文提出了一种基于毫米波雷达的三维特征 自适应融合人体动作识别算法。首先,采用二维 FFT变换与MUSIC算法对雷达回波信号进行处 理,得到人体动作的距离特征、多普勒特征与角度 特征,并在时域上拼接构建 RTM、DTM 以及 ATM 数据集。然后,提出了一种三分支卷积神经网络 结构与多维特征自适应融合方案,拓宽了网络的 人体动作特征输入维度,实现了对多维特征的有 效融合。最后,使用活动分类器实现对不同人体 动作的识别。相比于单维特征识别方法与多种多 维特征识别方法,所提方法能够有效提升人体动 作识别准确率与AUC值,证明了 3D-FAFNet 的有 效性与可靠性。后续的研究工作将从以下2个角 度出发:1)采集更加丰富的人体动作类型,如跌 倒、抽搐等威胁人体健康的异常动作,增强模型的 实用性;2)设计一种鲁棒高效的生成对抗网络,解 决制备大量数据集困难的问题。

参考文献:

- [1] RAY A, KOLEKAR M H. Transfer Learning and Its Extensive Appositeness in Human Activity Recognition: A Survey [J]. Expert Systems with Applications, 2023, 65 (6):122538-122570.
- [2] BODHE R, SIVAKUMAR S, SAKARKAR G, et al. Outdoor Activity Classification Using Smartphone Based Inertial Sensor Measurements [J]. Multimedia Tools and Applications, 2024, 15(1):1-27.
- [3] ALEXAN A I, ALEXAN A R, ONIGA S. Real-Time Machine Learning for Human Activities Recognition Based on Wrist-Worn Wearable Devices [J]. Applied Sciences, 2024, 14(1):329-349.

- [4] ZHANG Haoran, XU Linhai. Multi STMT: Multi Level Network for Human Activity Recognition Based on Wearable Sensors[J]. IEEE Trans on Instrumentation and Measurement, 2024, 73(3):1-12.
- [5] CAMARENA F, GONZALEZ-MENDOZA M, CHANG L, et al. An Overview of the Vision - Based Human Action Recognition Field [J]. Mathematical and Computational Applications, 2023, 28(2):61-80.
- [6] XIN Wentian, LIU Ruyi, LIU Yi, et al. Transformer for Skeleton-Based Action Recognition: A Review of Recent Advances[J]. Neurocomputing, 2023 (3):164-186.
- [7] MOGHADDAM M G, SHIREHJINI A A N, SHIRMOHAM-MADI S. A WiFi - Based Method for Recognizing Fine -Grained Multiple-Subject Human Activities[J]. IEEE Trans on Instrumentation and Measurement, 2023, 72(1):1-13.
- [8] GONG Danei, LIU Kezhong, PEI Dashuai, et al. Wi-Watch: WiFi-based Vigilant-Activity Recognition for Ship Bridge Watchkeeping Officers [J]. IEEE Trans on Instrumentation and Measurement, 2023, 72(2):645-662.
- [9] LI Xinrong, WANG Xiaodong, YANG Qing, et al. Signal Processing for TDM MIMO FMCW Millimeter-Wave Radar Sensors[J]. IEEE Access, 2021, 9(3):167959-167971.
- [10] JIN Feng, SENGUPTA A, CAO Siyang. mmFall: Fall Detection Using 4-D mmWave Radar and a Hybrid Variational RNN AutoEncoder[J]. IEEE Trans on Automation Science and Engineering, 2022, 19(2):1245-1257.
- [11] 蒋留兵,吴岷洋,车俐.FMCW 雷达基于光学字符识别的连续动作识别研究[J].雷达科学与技术,2023,21 (1):74-81.
- [12] CHAKRABORTY M, KUMAWAT H C, DHAVALE S V, et al. DIAT-µ RadHAR (Micro-Doppler Signature Dataset) & µ RadNet (A Lightweight DCNN)—for Human Suspicious Activity Recognition [J]. IEEE Sensors Journal, 2022, 22(7):6851-6858.
- [13] 屈乐乐,张丁元,杨天虹,等.基于双流特征融合的 FMCW 雷达人体连续动作识别[J].雷达科学与技术, 2022,20(5):565-572.
- [14] 王勇,吴金君,田增山,等.基于FMCW 雷达的多维参数手势识别算法[J].电子与信息学报,2019,41(4): 822-829.
- [15] ELBIR A M. DeepMUSIC: Multiple Signal Classification via Deep Learning[J]. IEEE Sensors Letters, 2020, 4(4):1-4.
- [16] HOWARD A, SANDLER M, CHU G, et al. Searching for MobileNetV3 [C]// 2019 IEEE/CVF International Conference on Computer Vision, Seoul, (下转第586页)