文章编号:1671-4598(2024)10-0069-08 DOI:10.16526/j.cnki.11-4762/tp.2024.10.010 中图分类号:TN957.52 文献标识码:A

基于双模态门控特征融合的跌倒检测方法

郭夏迪，曹炳尧
(上海大学 特种光纤与光接入网重点实验室，上海 200444)

摘要：使用单一传感器进行人体跌倒检测的方法不能充分捕捉动作特征，摄像头在光线较差时无法获得高质量图像，毫米波雷达的点云稀疏性降低了远距离目标信息的有效性；针对上述问题，提出了一种基于双模态门控特征融合的跌倒检测方法；使用雷达和摄像头同步检测，雷达分支根据时间-距离图和微多普勒图获得融合特征，视觉分支提取目标的光学特征；将两种特征送入门控融合模块，根据权重整合特征信息，在输出层实现分类；设计了雷达分支和整体网络的相关实验，雷达分支融合方法的平均准确率是91.7%，优于单一特征方法；整体网络的门控融合方法的准确率是94.1%，相比特征相加融合和首尾拼接融合方法分别高出3.0%和1.8%；充分表明该方法能够提升人体跌倒检测的性能。

关键词：毫米波雷达；视觉；双模态；特征融合；门控融合；跌倒检测

Falling Detection Method Based on Bi-modal Gated Feature Fusion

GUO Xiadi，CAO Bingyao
(Key Laboratory of Specialty Fiber and Optics Access Networks，Shanghai University，Shanghai 200444，China)

Abstract：A single sensor is used to detect the method for human falling，which cannot adequately capture motion features，cameras can not obtain high-quality images under poor lighting conditions，and the point cloud sparsity of millimeter-wave radar reduces the effectiveness of remote target information. To solve these problems，a falling detection method based on bi-modal gated feature fusion is proposed. The method uses a radar and camera to synchronously detect the images. The radar branch obtains the fusion feature based on the time-distance map and micro-Doppler map，and the visual branch extracts the optical feature of the target. The two features are sent to the gated fusion module，and the feature information is integrated according to the weight to realize the classification at the output layer. Experiments of the radar branch and the overall network are designed，the average accuracy of the radar branch fusion method is 91.7%，which is better than that of the single feature method. The accuracy of the gated fusion method in the overall network is 94.1%，which is 3.0% and 1.8% higher than that of the feature addition fusion and fore-tail fusion，respectively. It fully demonstrates that the method can effectively improve the performance of human falling detection.

Keywords：millimeter-wave radar；visual；bi-modal；feature fusion；gated fusion；falling detection

0 引言

随着人口老龄化趋势的加剧，老年群体的安全与健康成为社会的重点议题。根据统计表明，跌倒成为对独居老人构成重大威胁的意外之一[1]，平均每5次跌倒事件中会有一次造成骨折或头部损伤等严重伤害[2]。老年人跌倒后通常无法及时自救，会造成更严重的后果。因此，人体跌倒检测技术对于实施有效救援、降低伤害后果有至关重要的作用。

在早期阶段，人体跌倒检测技术主要依赖于体积小巧、成本较低的可穿戴设备，使用加速度计、陀螺仪等传感器来确定佩戴者是否跌倒。但此类设备的用户使用舒适度欠佳，佩戴位置的改变也会影响检测结果[3]。随着计算机视觉的快速发展，相关学者使用机器学习或深度学习从视觉信息提取特征进行训练。摄像头采集的视觉数据能够提供详细的人体纹理和运动信息，显著提升了跌倒检测的可靠性。但受限于视角范围，无法完整捕捉近距离目标的肢体运动，可能会丢失重要的运动信息。并且视觉信息容易受到光照条件的影响，难以满足全天候、稳定性的应用需求[4]。近年来，基于毫米波雷达的跌倒检测技术由于其自身的优越性受到越来越多的关注。毫米波雷达具有广阔的探测视角，对环境和光照变化的敏感性较低，可以采集目标的三维空间信息，为检测提供稳定且立体的数据支持。但毫米波雷达的点云数据具有稀疏性，单人目标的点云大多只有数个点，且点云数量会随着目标距离的增加而减少，限制了有效信息的获取[5]，在复杂场景中可能导致误检率增加。

为了克服单一模态的局限性，多模态融合方法已经成为研究热点和发展趋势。多模态融合通过整合来自不同传感器的数据，充分利用各自的优势，提升检测任务的准确

收稿日期:2024-04-10； **修回日期**:2024-05-07。
基金项目:上海市科委项目(22511100902)。
作者简介:郭夏迪(1999-)，女，硕士研究生。
通讯作者:曹炳尧(1985-)，男，博士，实验师。
引用格式:郭夏迪，曹炳尧.基于双模态门控特征融合的跌倒检测方法[J].计算机测量与控制，2024，32(10)：69-76.

性和鲁棒性。在人体跌倒检测的研究中，决策级融合和特征级融合是目前最流行的多模态融合策略。决策级融合将不同模态数据送入对应的网络训练，采用决策融合模型将独立的决策信息结合，产生综合性的判定结果。文献［6］提出了一种基于 4D 成像雷达和摄像头的双任务感知框架，使用雷达提取人体目标的运动特征，并通过摄像头捕捉光学图像，根据实验选择最优网络参数进行两个分支的加权平均融合，实现对人体动作的分类。决策级融合虽然在整合多模态信息方面提供了一种直接的途径，但其处理方式通常局限于对各独立模态输出结果的简单数学运算，例如求和或平均，未能深入挖掘和利用不同模态间潜在的相互作用与协同。特征级融合致力于将不同模态在特征层面进行结合[7]，使得检测模型能够更有效地学习和捕捉跨模态数据之间的相关性。文献［8］获取雷达的微多普勒谱图和摄像头的步态能量图，通过身体空间注意力模块提取步态能量图中的身体部位特征，同时通过长短时序建模模块提取微多普勒谱图中的运动特征，在多尺度特征空间中融合两种模态的特征，为人体动作识别任务提供具有代表性的步态特征。文献［9］聚焦于室外场景中的人体步态分类问题，探讨了雷达和视频结合的不同融合策略，包括决策级融合和特征级融合。其中，决策级融合将两种模态信息分别送入卷积神经网络和长短期记忆网络级联构成的网络架构中，对两种模态的预测结果取平均值，得到最终分类决策。特征级融合采用卷积神经网络分别提取每种模态的特征，将两种特征拼接后送入长短期记忆网络，得到最终结果。

在多模态融合领域，尽管决策级融合和特征级融合的研究有所进展，但现有方法在提取模态特征与挖掘不同模态间深层次关联方面仍存在不足。在人体动作识别与跌倒检测应用领域，基于多模态融合的研究处于初期阶段，大部分研究仍依赖于单一模态的数据。因此，本文提出了一种基于毫米波雷达与视觉模态互补性的门控特征融合网络，旨在实现信息的有效整合。该网络结构的设计综合考虑了不同模态对整体融合过程的贡献度，通过门控机制筛选和优化特征表示，过滤掉冗余信息并保留关键特征。这种方法不仅提升了网络对人体动作的语义理解，而且增强了对复杂环境变化的适应性，有效缓解了由于光线变化、目标不完整或图像模糊等因素对检测性能的影响，增强了任务的泛化能力和实用性。

1 系统设计

本文所提出的跌倒检测的算法方案如图 1 所示，具体流程为：同时启动两种传感器的原始数据采集工作，其一采集毫米波雷达传感器的数据，通过一系列数据预处理操作得到雷达的时间-距离特征图和微多普勒特征图；其二采集摄像头的数据，利用 FFmpeg 工具和抽帧操作处理原始视频流，得到包含人体运动信息的动作视频帧序列。把处理后的数据分别送到设计的分支网络中进行多层特征提取，得到两种模态的深层次信息。将两种模态信息通过门控模块进行特征的优化组合，增强特征的多样性和稳定性，最终送到分类器进行跌倒检测的决策。

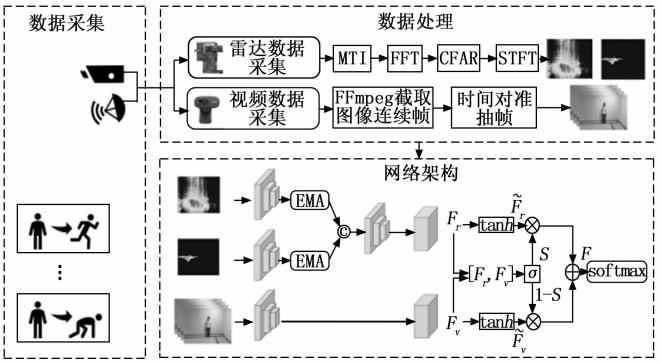

图 1 算法方案图

2 毫米波雷达特征提取分支

2.1 数据预处理

2.1.1 构建调频连续波雷达模型

调频连续波（FMCW，frequency modulated continuous wave）雷达在扫频周期发射频率连续变化的信号，不仅可以高精度测量，而且功耗较低，可以长时间、大规模地收集数据，从而深入理解和分析人类行为的复杂性和多样性。

假设 FMCW 雷达发射锯齿形信号[10]，信号为：

$$S_T(t) = exp\left[j2\pi\left(f_c t + \frac{1}{2}\mu t^2\right)\right] \tag{1}$$

式中，f_c 信号载频，μ 为调频斜率，表达式为：

$$\mu = \frac{B}{T} \tag{2}$$

式中，B 为调频带宽，T 为调频脉冲的周期。接收天线收到的目标反射的回波信号 S_R（t）表示为：

$$S_R(t) = \exp\left\{j2\pi\left[f_c(t-\tau) + \frac{1}{2}\mu(t-\tau)^2\right]\right\} \tag{3}$$

式中，τ 为接收信号与发射信号的时延，假设目标的初始距离为 R，v 以速度远离雷达，则：

$$\tau = \frac{2(R+vt)}{c} \tag{4}$$

式中，c 为光速。将发射信号和接收信号在混频器中进行混频操作，通过低通滤波后生成中频信号（IF，intermediate frequency），表示为：

$$S_{IF}(t) = \exp\left[j2\pi\left(f_c\tau - \frac{1}{2}\mu\tau^2 + \mu t\tau\right)\right] \tag{5}$$

将式（4）代入式（5）可得：

$$S_{IF}(t) = \exp\left\{j2\pi\left[\frac{2\mu(R+vt)t}{c} + \frac{2f_c(R+vt)}{c} - \frac{2\mu(R+vt)^2}{c^2}\right]\right\} \tag{6}$$

式中，第三项的分母是 c^2，与前两项的量级相差过大，可忽略不计。在室内近距离检测中信号的处理时间在 ms 或 μs 量级，t^2 可以忽略。简化可得：

$$S_{IF}(t) = \exp\left\{j2\pi\left[\left(\frac{2f_c v}{c} + \frac{2\mu R}{c}\right)t + \frac{2f_c R}{c}\right]\right\} \tag{7}$$

假设雷达发射 M 个线性调频信号，每个信号个采样点，得到数字中频信号为：

$$Y[n,m]=\exp\left\{j2\pi\left[\left(\frac{2f_c v}{c}+\frac{2\mu R}{c}\right)nT_f+\frac{2R(nT_f+mT_s)}{\lambda}\right]\right\} \quad (8)$$

式中，n，m 分别代表快时间采样轴和慢时间采样轴的标号，$n=0$，1，2，…，$N-1$；$m=0$，1，2，…，$M-1$。T_f，T_s 分别是快、慢时间轴的采样间隔。

2.1.2　提取时间-距离和微多普勒特征

FMCW 雷达的数字中频信号可以按照快时间维、慢时间维和接收天线维组成一个立方体[11]，通过信号处理提取时间、距离、速度、多普勒和角度等关键特征，雷达信号的处理流程如图 2 所示。

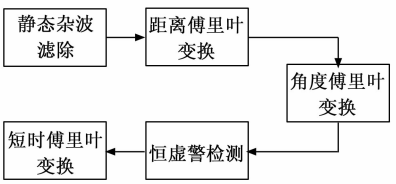

图 2　雷达信号处理流程图

在密闭的室内场景中，信号中含有由地面、墙壁、家居和其他设备等静止物体产生的杂波信号，干扰后续的信号分析结果。所以使用运动目标指示（MTI，moving target indication）滤波技术，通过将相邻采样周期的信号相减，有效地滤除信号中的静态杂波。在快时间维执行距离快速傅里叶变换（FFT，fast Fourier transform）[12]获取目标的距离信息，进一步将距离信息在时间维度有序堆叠，生成时间-距离特征图。时间-距离特征图直观地表现了人体目标在一定时间内相对于雷达的动态距离变化信息。

在接收天线维执行角度 FFT 解析相位变化，获得目标的距离-角度矩阵，利用恒虚警（CFAR，constant false alarm rate detector）[13]技术对矩阵进行目标检测。在慢时间维对定位的目标 Bin 实施短时傅里叶变换（STFT，short timefourier transform）[14]，监测信号频率的变化，获得当前时刻的速度信息。以一定步长滑动窗口后，按时间排列就可以得到整个信号的微多普勒特征图。对输入信号，STFT 的计算过程如下：

$$\mathrm{STFT}(t,\omega)=\int_{-\infty}^{\infty}\omega(\tau-T)s(\tau)e^{-j\omega\tau}\mathrm{d}\tau \quad (9)$$

式中，$\omega(t)$ 是窗函数，选用汉宁窗，窗的长度设置为 255。STFT (t,ω) 代表微多普勒特征，反映了人体运动时躯干和四肢的动作变化特征。本文针对室内的单人场景进行分析，图 3 是 4 类人体日常动作经过上述信号处理后生成的特征图，分别是行走、坐下、站起身和摔倒。

2.2　毫米波雷达特征融合网络

2.2.1　2D CNN

卷积神经网络（CNN，convolutional neural networks）

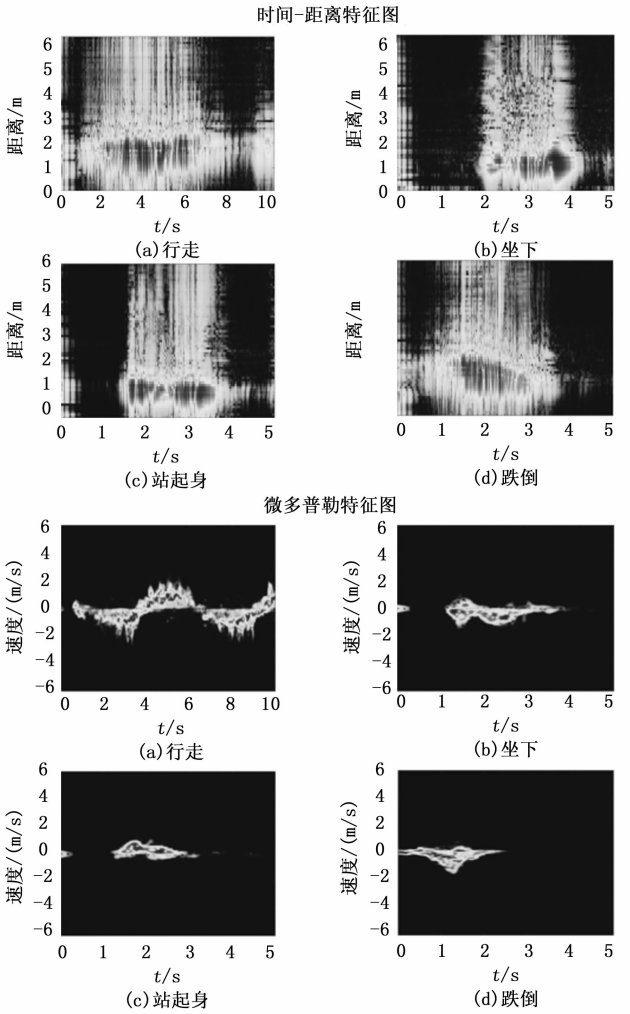

图 3　毫米波雷达特征图

利用局部操作来提取输入数据的局部特征，具有强大的表征学习能力。本文使用并行的 2D CNN 网络提取抽象的特征信息，计算公式[15]为：

$$v_{ij}^{xy}=\mathrm{ReLU}\left(\sum_m\sum_{p=0}^{H_i-1}\sum_{q=0}^{W_i-1}w_{ijm}^{pq}v_{(i-1)m}^{(x+p)(y+q)}+\boldsymbol{b}_{ij}\right) \quad (10)$$

式中，ReLU（·）表示单元的激活函数，v_{ij}^{xy} 表示在第 i 层的第 j 个特征图在（x，y）像素的输出值，m 表示与当前特征图相连的第 $i-1$ 层的特征图的索引量，H_i、W_i 分别表示第 i 层二维卷积核的高度和宽度，w_{ijm}^{pq} 表示连接第 m 个特征图的卷积核的权重，$\boldsymbol{b}_{ij}$ 表示第 i 层的第 j 个特征图的偏置矩阵。ReLU 激活函数计算方法为：

$$f(x)=\begin{cases}x,x>0\\0,x<0\end{cases}=max(0,x) \quad (11)$$

不同的卷积核的设计能够实现对图像特征的不同层次和尺度的提取。当卷积核的尺寸较大时，其感受野相应增大，能够捕捉到图像的全局特征，但计算量的显著增加，导致网络训练速度减缓。核尺寸较小时，感受野减小，可以集中于图像的局部特征，有助于网络的快速训练和部署。

2.2.2 EMA 注意力机制

为了关注 2D CNN 提取特征中的关键信息，引入高效的多尺度注意力模块（EMA，efficient multi-scale attention module），不仅可以对通道间的信息进行编码以调整不同通道的重要程度，还可以精确地将空间结构信息保留到通道中[16]。EMA 模块的结构如图 4 所示。

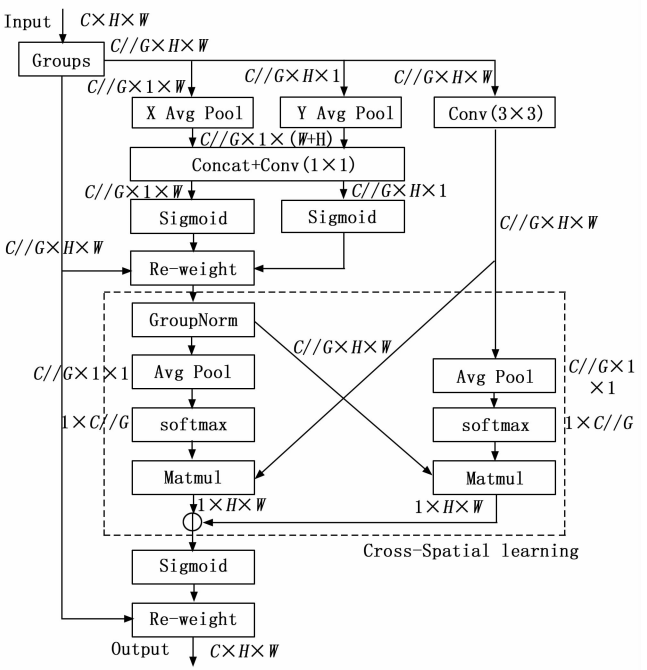

图 4 EMA 模块

首先将输入的特征图沿通道维度分成 G 个子特征，表示为：

$$X=[X_0,X_i,\cdots,X_{G-1}],X_i\in R^{C//G\times H\times W} \tag{12}$$

然后采用并行网络分支在中间特征中收集更多的上下文信息，提取划分后的特征图的注意力权重。3 个平行路径分别是两个 1×1 卷积分支和一个 3×3 卷积分支，对两个 1×1 卷积分支，将子特征图分别沿 H 和 W 两个方向进行 1D 全局平均池化，获得相应的特征层 $C//G\times1\times W$ 和 $C//G\times H\times1$。合并两个并行分支，将 H 和 W 转置到相同维度进行堆叠，并使用 1×1 卷积获得特征层 $C//G\times1\times(W+H)$。再分离两个并行分支，经过非线性 Sigmoid 函数拟合，获得 H 和 W 维度的注意力情况，与原有的特征相乘得到 1×1 卷积分支的输出。在 3×3 卷积分支使用 3×3 的卷积捕捉多尺度特征表示，扩大特征空间，得到 3×3 卷积分支的输出。

为了捕捉像素级别的关系，构建了跨维度交互模块实现特征聚合。模块的输入是 1×1 卷积分支的输出和 3×3 卷积分支的输出，引入 2D 全局平均池化，计算公式为：

$$z_c=\frac{1}{H\times W}\sum_j^H\sum_i^W x_c(i,j) \tag{13}$$

分别对两个卷积分支的输出进行 2D 全局平均池化，经过 softmax 激活后与另一个卷积分支的输出进行点积运算，将生成的两组空间注意力权重值经过 Sigmoid 激活，与原特征相乘获得特征信息，最终的输出与输入的尺寸相同。

2.2.3 雷达分支网络架构

针对毫米波雷达模态，特征提取分支通过对时间-距离图和微多普勒图的分析，捕捉人体动作的细微特征。将两个特征图并行地输入到 CNN 网络中，并引入 EMA 模块，使网络不仅能够理解全局特征，而且更加精准地关注和表征关键局部特征。最后对双分支进行特征融合，为跌倒检测任务提供丰富的雷达模态信息[17]。雷达分支的网络实现流程如下：

1）为了适应网络的输入要求，需要对时间-距离图和微多普勒图进行处理。首先将图片尺寸调整为 224×224×3，然后将像素值转化到 0～1 范围。时间-距离图和微多普勒图作为两个分支的输入，旨在从两个特征图获取更多语义信息。

2）利用 3×3 的卷积（Kernelsize＝3，stride＝2）提取浅层特征，输出通道数为 32。然后添加批量归一化（BN，batch normalization）层将特征图标准化，从而加快模型训练过程和收敛速度。并采用 ReLU 激活函数，改解决梯度消失的问题，提高模型训练效率和性能。

3）添加 EMA 注意力机制，提高了重要特征的权重，有助于模型更全面地捕捉人体动作在不同尺度上的细微变化。

4）经过池化核尺寸为 3×3 的最大池化层（Kernelsize＝3，stride＝3）后，双分支输出的特征图拼接融合，获得的高阶特征图包含时间－距离和微多普勒的综合信息。

5）融合特征图通过卷积核尺寸为 3×3 的卷积层（Kernel _ size＝3，stride＝3）、BN 层、ReLU 激活函数和 2×2 最大池化层（Kernel _ size＝2，stride＝2）。最终通过 Flatten 层将多维特征图展平为一维特征向量，为后续任务提供了强有力的特征支持。

3 视觉特征提取分支

针对视觉模态，2D CNN 只能对一帧图像进行卷积操作，没有考虑连续帧之间含有的运动信息。相对于 2D CNN，3D CNN 可以在时间维度进行运算。Tran[18]等人提出的 C3D（Convolutional 3D）模型利用 3D CNN 提取空间和时间特征，将连续帧图像输入到网络训练，对视频数据进行端到端的处理。参考 C3D 网络架构的设计思路，构建视觉特征提取网络，符合该分支的功能需求。

对于摄像头捕捉的原始视频数据，使用 FFmpeg 工具转换为 RGB 连续帧序列。为了保证与毫米波雷达数据的时间对准，根据时间戳抽出图像序列的多余帧，并对图像尺寸大小进行调整，将预处理后得到的 3×16×112×112 的帧序列作为分支输入。

网络模型主要包含 3 个 3D 卷积层和 3 个 3D 池化层。利用 3×3×3 的卷积（Kernelsize＝3，stride＝2）在时空维度提取视觉特征信息，输出通道数为 64。在卷积层后嵌入最大池化层（Kernelsize＝1×2×2，stride＝1×2×2），控

制模型的复杂度、减少计算量，有效防止了模型的过拟合。再通过两层通道数分别为 128、256 的卷积和池化核大小均为 2×2×2 的池化操作，逐层提炼和强化特征表达，得到视觉运动特征。最终展平为一维特征向量作为视频分支的特征表示。该分支的设计充分考虑了视觉数据的复杂性和动态性，能够有效地捕捉人体动作的视觉特征。

4　BMGFFN 模型

本文设计了一种双模态门控特征融合网络（BMGFFN，bi-modal gated feature fusion network），网络包括雷达特征提取分支、视觉特征提取分支和门控融合模块。BMGFFN 的核心目标是从多模态数据中提取强大的中间表示，从而全面反映不同模态的特征并指导后续的检测过程。

门控机制如图 5 所示，双曲正切函数作为门控的激活函数分别对两个模态的内部进行编码。对于输入模态 F_r、F_v，门控神经元（图 5 中由 σ 节点表示）负责调节每个模态特征对于整体输出的贡献度。这种机制确保了网络在接收新样本时，能够通过考虑所有模态的联合特征向量，动态地决定每种模态对于样本内部表示的贡献程度。

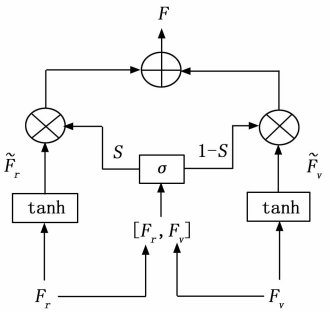

图 5　门控机制

门控机制的数学计算表示为[19]：

$$\widetilde{F}_r = \tanh(W_r, F_r) \tag{14}$$

$$\widetilde{F}_v = \tanh(W_v, F_v) \tag{15}$$

$$S = \sigma(W_s \cdot [F_r, F_v]) \tag{16}$$

$$F = S\widetilde{F}_r + (1-S)\widetilde{F}_v \tag{17}$$

式中，W_r、W_v、W_s 是可学习的权重参数，S 是门控模块的权值系数，权值系数越大，对最终融合特征的贡献度就越高。tanh 是双曲正切函数，σ 是 Sigmoid 函数。

BMGFFN 模型采用门控机制优化雷达和视觉数据模态的特征表示，具体实现流程如下。将每个动作对应的毫米波雷达特征图和视频连续帧序列作为输入，送入 BMGFFN 模型，通过雷达分支和视觉分支提取出毫米波雷达模态的特征向量 $\boldsymbol{F}_r$ 和视觉模态的特征向量 $\boldsymbol{F}_v$。鉴于两个特征提取分支的网络架构的差异性，为了便于后续门控融合模块的处理，将两种模态得到的 $\boldsymbol{F}_r$ 和 $\boldsymbol{F}_v$ 使用全连接（FC，fully connected）层统一映射到 1024 维，并通过双曲正切函数进行激活。同时，经过 FC 层的 $\boldsymbol{F}_r$ 和 $\boldsymbol{F}_v$ 在通道维度拼接，生成初级融合特征，根据式（16）计算得到权值系数 S，该系数可以根据模态间的互补性动态地赋予两个模态不同的权重，根据式（17）将系数与各模态的特征向量进行逐元素相乘，再相加得到最终的特征表示 $\boldsymbol{F}$，实现特征的自适应融合。

模型的输出层采用交叉熵损失函数，量化了模型输出的概率分布与真实标签之间的差距，数学表达式为：

$$L_{ce} = -y\log\hat{y} - (1-y)\log(1-\hat{y})$$
$$\begin{cases} -\log\hat{y} & y=1 \\ -\log(1-\hat{y}) & y=0 \end{cases} \tag{18}$$

式中，y 表示真实概率，$\hat{y}$ 表示预测概率。交叉熵损失对误判的分类赋予较大的惩罚，引导模型集中于边界模糊或预测困难的样本，有利于提高分类的准确性。

5　实验与分析

5.1　模型环境

模型在 Python 3.7 环境下执行，实验环境采用 AMD R7 5800H 和 NVIDIAGeForce RTX 3050Ti GPU，依托于 PyTorch 深度学习框架在 GPU 中进行训练。

5.2　实验数据集

5.2.1　自建数据集

本实验搭建了定制的传感器平台，包含一个毫米波雷达传感器和一台摄像头。毫米波雷达设备采用德州仪器公司的 IWR6843AOPEVM、MMWAVEICBOOST 承载卡平台以及 DCA1000EVM 组合。IWR6843AOPEVM 有 3 个发射通道和 4 个接收通道，支持水平和垂直方向各 120°的广阔视角，参数配置如表 1 所示。摄像头采用 LogiC270 高清摄像头，帧率为 30 FPS，可以通过串口直接与计算机相连。

表 1　毫米波雷达参数配置

参数	数值
起始频率/GHz	60
扫频带宽/GHz	1.78
调频斜率/（MHz/μs）	56.25
Tx 起点/μs	1
ADC 采样点数	96
Chirp 数	128

进行数据采集时，毫米波雷达传感器使用三脚架放置在距离地面 170 cm 的位置，摄像头通过支架固定在毫米波雷达传感器正上方 5 cm 的位置，保证摄像头的光轴垂直于毫米波雷达芯片。

考虑到实验场地需要符合实际的人体日常活动场景，设定实验区域的大小为 4 m×5 m。实验人员共有 6 名受试者，年龄、身高与体重等基础信息不同。数据集采用毫米波雷达和摄像头联合采集的方法，每组样本的特征包含雷达数据预处理后获得的时间-距离特征图、微多普勒特征图以及摄像头数据预处理后获得的动作帧序列。实验人员在实验区域内做规定动作来获取样本数据，本数据集采集了 4 种典型的人体行为作为测试动作，共计 420 组样本，包括 150 组跌倒行为和 270 组正常行为，其中正常行为有行走、

坐下、站起身。

5.2.2 Glasgow 公开数据集[20]

该数据集是 Glasgow 大学采集的毫米波雷达数据集，采用了 Anocortek 公司的 FMCW 雷达，工作在 5.8 GHz 频段，带宽 400 MHz。引入该数据集作为雷达数据的补充，参与毫米波雷达特征融合网络的性能验证。选取共计 588 组动作样本，包括跌倒、行走、坐下、站起身。

5.3 评价指标

通过下面介绍的 4 个评价指标评判模型的性能。TP 是跌倒行为被判定为跌倒的个数，TN 是正常行为被判定为非跌倒的个数，FP 是正常行为被判定为跌倒的个数，FN 是跌倒行为被判定为非跌倒的个数。

准确率：最常见的指标，代表正确检测的样本占所有样本的比重。

$$Acc = \frac{TP + TN}{TP + FN + FP + TN} \tag{19}$$

召回率：代表正确检测是跌倒行为的样本占所有跌倒样本的比重。

$$Re = \frac{TP}{TP + FN} \tag{20}$$

精确率：代表正确检测是跌倒行为的样本占所有检测为跌倒行为的样本的比重。

$$Pr = \frac{TP}{TP + FP} \tag{21}$$

F_1 值：召回率和精确率的调和平均，综合评价方法的性能。

$$F_1 = \frac{2 \times Re \times Pr}{Re + Pr} \tag{22}$$

5.4 评估结果分析

5.4.1 毫米波雷达特征融合实验

本节主要通过实验验证基于时间-距离图和微多普勒图的毫米波雷达分支网络，分别在两个数据集评估了跌倒检测任务的性能。设置了 3 组实验：1）仅时间-距离特征进行检测；2）仅微多普勒特征进行检测；3）融合两种特征进行检测。实验中使用交叉熵函数作为损失项，通过 Adam 优化器对参数进行更新，最小化损失并且提高模型性能。初始学习率为 0.001，通过 StepLR 更新学习率，batch _ size 设置为 16，共计进行 100 轮训练。

1）在 Glasgow 公开数据集中的对比实验：首先在 Glasgow 公开数据集进行实验，3 组实验的损失值和准确率随训练次数变化的曲线如图 6 和图 7 所示。

在 60 轮训练后，实验的损失值下降变缓，准确率变化趋于平稳。3 种特征的检测性能比较结果如表 2 所示。

表 2 Glasgow 公开数据集雷达模态性能检测结果 %

方法	准确率	召回率	精确率	F_1 值
时间-距离特征	80.2	79.5	68.9	73.8
微多普勒特征	92.8	92.3	87.8	90.0
融合特征	94.6	94.9	90.2	92.5

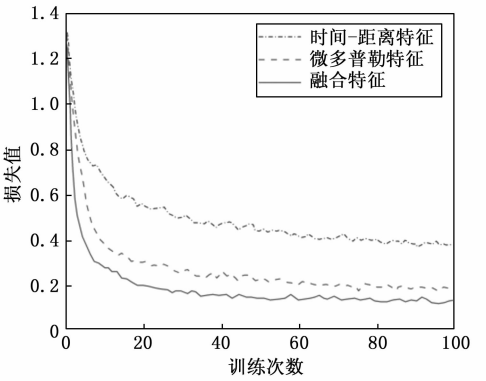

图 6 三组实验的损失值曲线

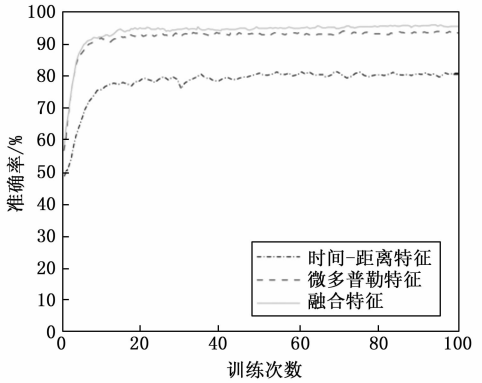

图 7 3 组实验的准确率曲线

由 3 组实验的性能检测结果可以得出，融合两种特征的方法展现了更强的检测性能。融合特征的准确率分别比单一特征高出 1.8%和 14.4%，召回率分别提升了 2.6%和 15.4%，精确率也分别提高了 2.4%和 21.3%，精确率 F_1 值分别比单一特征高出 2.5%和 18.7%。

2）在自建数据集中的对比实验：取自建数据集中的雷达数据集进行实验，3 组实验的性能检测结果如表 3 所示。

表 3 自建数据集雷达模态性能检测结果 %

方法	准确率	召回率	精确率	F_1 值
时间-距离特征	74.4	71.7	62.3	66.7
微多普勒特征	85.1	85.0	76.1	80.3
融合特征	88.7	88.3	81.5	84.8

可以发现，在自建数据集中，融合特征的方法在所有性能中的表现也优于单一特征方法。进一步验证了融合特征方法在跌倒任务中的有效性，增强了结论的客观性和可信度。在两个数据集中，准确率对比如图 8 所示。

通过结果发现，自建数据集的准确率低于公开数据集准确率。为了更贴近老年人日常活动的实际状态，实验人员放缓了行动速度，模拟老年人可能的行为特征，增加了数据集的复杂性。由于跌倒行为与坐下行为的运动特征在速度减缓情况下表现相似，因此会出现特征重叠，对分类算法的准确性造成影响。该模块的性能提升为整体网络架

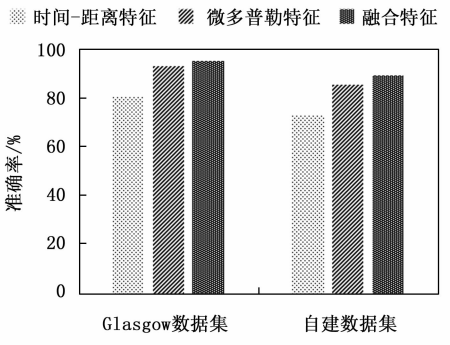

图 8　不同数据集的准确率比较

构的高效工作提供了技术支撑。

5.4.2　BMGFFN 实验

模型训练使用交叉熵作为损失函数，使用 Adam 优化器更新模型参数，初始学习率为 0.000 5，采用 StepLR 更新学习率，batch _ size 设置为 32，共计进行 200 轮训练。

1）网络结构组合对比实验：为了探索 BMGFFN 网络的最优网络结构，在自建数据集对不同深度的 2D CNN 和 3D CNN 进行组合实验，分别代表雷达分支和视觉分支的卷积层数，最终确定取得最高准确率的结构。根据初步的实验分析，卷积核大小并非影响准确率的关键参数，所以 2D CNN 和 3D CNN 设置固定的卷积核大小为 33 和 333，实验结果如表 4 所示。

表 4　不同网络结构准确率对比结果

2D CNN 深度	3D CNN 深度	准确率/%
1	1	86.3
2	1	89.9
3	1	90.5
4	1	87.5
1	2	89.3
2	2	91.7
3	2	92.9
4	2	91.1
1	3	92.3
2	3	94.1
3	3	93.5
4	3	91.1
1	4	85.7
2	4	88.7
3	4	82.1
4	4	76.2

根据表 4 可以确定准确率最佳的网络结构是雷达分支和视觉分支分别选取 2 层 2D CNN 和 3 层 3D CNN 的组合。该组合可以达到最高的检测准确率，不仅可以有效提取目标的关键特征，避免了因网络过浅而导致的特征提取能力不足，同时也防止了因网络过深可能引起的过拟合现象。

2）融合方法对比实验：为了验证本文提出 BMGFFN 方法的有效性，在自建数据集进行测试，设计了对比实验：（1）仅使用毫米波雷达模态特征进行检测；（2）仅使用视觉模态特征进行检测；（3）双模态特征相加融合进行检测，直接将两个模态分支提取的向量的对应元素相加，得到的向量维度与单模态提取的特征维度一致；（4）双模态首尾拼接融合进行检测，在指定维度上将两个特征向量合并，保留了更多的原始特征信息；（5）双模态门控特征融合进行检测，将两个模态分支提取的向量送入门控模块进行动态融合[21]。

对单模态特征与多模态特征融合的检测结果进行评估，性能检测结果如表 5 所示。

表 5　不同融合方法性能检测结果　　%

方法	准确率	召回率	精确率	F_1 值
毫米波雷达	88.7	88.3	81.5	84.8
视觉	89.3	86.7	83.9	85.3
特征相加融合	91.1	90.0	85.7	87.8
首尾拼接融合	92.3	91.7	87.3	89.4
门控特征融合	94.1	93.3	90.3	91.8

由表 5 可知，多模态特征融合的性能表现均优于单模态，本文提出的门控特征融合在跌倒检测任务中的整体性能优异，准确率达到了 94.1%，比特征相加融合和首尾拼接融合的准确率分别提高了 3.0%、1.8%，召回率比两种融合方式分别高出 3.3%、1.6%，精确率分别提升了 4.6%、3.0%，综合评价指标 F1 值也分别提高了 4.0%、2.4%。门控特征融合方法不仅可以弥补单一模态信息的局限性，而且引入 Sigmoid 函数动态地调控特征权值。Sigmoid 函数具有平滑性和输出值在［0，1］内的特点，在处理多种因素时实现更精确的权重分配。实验表明，门控特征融合方法在跌倒检测中有更佳的特征提取与整合能力。

6　结束语

本文针对室内场景中人体跌倒行为，提出了一种基于毫米波雷达和视觉双模态的门控特征融合方法。本方法同步采集了毫米波雷达的时间-距离特征图和微多普勒特征图，以及摄像头的视频帧序列。使用毫米波雷达特征融合模块获得目标在三维空间的运动距离和速度分布，并使用视觉模块捕捉目标的光学运动特征。最后引入门控机制，根据不同模态的信息动态调整特征权重，生成更丰富、鲁棒的动作特征表示，用于跌倒检测任务。经过实验表明，相较于其他对比方法，BMGFFN 方法的表现更加优异，准确率达到了 94.1%，F_1 值达到了 91.8%，证明了该方法可以更精确地检测老年人的跌倒行为。

参考文献：

[1] 林伟权，刘　慧，利耀辉，等. 2014－2018 年广州市 60 岁及以上老年人跌倒/坠落伤害流行病学特征［J］. 中华疾病控制杂志，2020，24（3）：269－273.

[2] STERLING D A，O'CONNOR J A，BONADIES J. Geriatric

falls: injury severity is high and disproportionate to mechanism [J]. Journal of Trauma and Acute Care Surgery, 2001, 50 (1): 116 - 119.
[3] 陈达理, 刘雪红. 基于自适应样本熵的穿戴式传感器人体活动识别 [J]. 武汉理工大学学报, 2022, 44 (8): 91 - 96.
[4] 毕春艳, 刘 越. 基于深度学习的视频人体动作识别综述 [J]. 图学学报, 2023, 44 (4): 625 - 639.
[5] 林凤泰, 严蕾蕾, 张 慧, 等. 基于最近迭代点的毫米波雷达点云数据处理方法 [J]. 信号处理, 2023, 39 (2): 288 - 297.
[6] SONG Y, DAI Y, JIN T, et al. Dual-task human activity sensing for pose reconstruction and action recognition using 4D imaging radar [J]. IEEE Sensors Journal, 2023, 23 (19): 23927 - 23940.
[7] 李晓欢, 霍科辛, 颜晓凤, 等. 基于特征加权视觉增强的雷视融合车辆检测方法 [J]. 公路交通科技, 2023, 40 (2): 182 - 189.
[8] SHI Y, DU L, CHEN X, et al. Robust gait recognition based on deep CNNs with camera and radar sensor fusion [J]. IEEE Internet of Things Journal, 2023, 10 (12): 10817 - 10832.
[9] DE JONG R J, DE WIT J J M, UYSAL F. Classification of human activity using radar and video multimodal learning [J]. IET Radar, Sonar & Navigation, 2021, 15 (8): 902 - 914.
[10] LI X, WANG X, YANG Q, et al. Signal processing for TDM MIMO FMCW millimeter-wave radar sensors [J]. IEEE Access, 2021, 9: 167959 - 167971.
[11] EUGIN H, LEE J. Hardware architecture design and implementation for FMCW radar signal processing algorithm [C] //Proceedings of the 2014 Conference on Design and Architectures for Signal and Image Processing. IEEE, 2014: 1 - 6.
[12] SONG M, LIM J, SHIN D J. The velocity and range detection using the 2D-FFT scheme for automotive radars [C] //2014 4th IEEE International Conference on Network Infrastructure and Digital Content. IEEE, 2014: 507 - 510.
[13] XU C, WANG F, ZHANG Y, et al. Two-level CFAR algorithm for target detection in mmWave radar [C] //2021 International Conference on Computer Engineering and Application (ICCEA). IEEE, 2021: 240 - 243.
[14] JIN F, ZHANG R, SENGUPTA A, et al. Multiple patients behavior detection in real-time using mmWave radar and deep CNNs [C] //2019 IEEE Radar Conference (Radar-Conf). IEEE, 2019: 1 - 6.
[15] HE J, REN Z, ZHANG W, et al. Fall detection based on parallel 2DCNN-CBAM with radar multidomain representations [J]. IEEE Sensors Journal, 2023, 23 (6): 6085 - 6098.
[16] OUYANG D, HE S, ZHANG G, et al. Efficient multi-scale attention module with cross-spatial learning [C] //ICASSP 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023: 1 - 5.
[17] WEI H, LI Z, GALVAN A D, et al. IndexPen: two-finger text input with millimeter-wave radar [J]. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2022, 6 (2): 1 - 39.
[18] TRAN D, BOURDEV L, FERGUS R, et al. Learning spatiotemporal features with 3D convolutional networks [C] //Proceedings of the IEEE International Conference on Computer Vision, 2015: 4489 - 4497.
[19] AREVALO J, SOLORIO T, MONTES-Y-GOMEZ M, et al. Gated multimodal networks [J]. Neural Computing and Applications, 2020, 32 (14): 10209 - 10228.
[20] FIORANELLI F, SHAH S A, LI H, et al. Radar sensing for healthcare [J]. Electronics Letters, 2019, 55 (19): 1022 - 1024.
[21] CHAIB S, LIU H, GU Y, et al. Deep feature fusion for VHR remote sensing scene classification [J]. IEEE Trans-actions on Geoscience and Remote Sensing, 2017, 55 (8): 4775 - 4784.

(上接第 68 页)

[9] BREWER K R W. Some consequences of temporal aggregation and systematic sampling for ARMA and ARMAX models [J]. J Econom, 1973, 1 (2): 133 - 154.
[10] 王丽娜, 肖冬荣. 基于 ARMA 模型的经济非平稳时间序列的预测分析 [J]. 武汉理工大学学报: 交通科学与工程版, 2004, 28 (1): 133 - 136.
[11] CHIB S, GREENBERG E. Bayes inference in regression models with ARMA (p, q) errors [J]. Journal of Econometrics, 1994, 64 (1/2): 183 - 206.
[12] 张晋昕, 王亚拉, 何大卫. 含缺失值时间序列的 ARMA 模型拟合 [J]. 中国卫生统计, 2000, 17 (4): 197 - 199.
[13] 王 锋, 杨 荣, 黄 攀, 等. 基于 ARMA 模型的海上风电功率预测研究 [J]. 电力系统装备, 2023 (7): 118 - 120.
[14] 张海勇, 马孝江, 盖 强. 非平稳信号的一种 ARMA 模型分析方法 [J]. 电子与信息学报, 2002, 24 (7): 992 - 996.
[15] 李瑞莹, 康 锐. 基于 ARMA 模型的故障率预测方法研究 [J]. 系统工程与电子技术, 2008, 30 (8): 1588 - 1591.
[16] 王 宏, 亚 萍, 龚正虎. PCAR: 基于主成分分析的网络关键路径发现算法 [J]. 计算机工程与科学, 2008 (6): 1 - 4.
[17] 江 丽. 基于粒子群与模拟退火算法的 BP 网络学习方法研究 [D]. 合肥: 安徽大学, 2013.
[18] 曹 杰. 基于 SVM 的网络流量特征降维与分类方法研究 [D]. 吉林: 吉林大学, 2017.
[19] 郭秀英. 预测决策的理论与方法 [M]. 北京: 化学工业出版社, 2010.
[20] WAN S, LAN Y, XU J, et al. Match-SRNN: Modeling the Recursive Matching Structure with Spatial RNN [J]. Computers & Graphics, 2016, 28 (5): 731 - 745.
[21] 杜 奕. 时间序列挖掘相关算法研究及应用 [D]. 合肥: 中国科学技术大学, 2007.