计算机科学
COMPUTER SCIENCE

基于特征融合的毫米波雷达行为识别算法

韩崇, 樊卫北, 郭澳

引用本文

韩崇, 樊卫北, 郭澳. 基于特征融合的毫米波雷达行为识别算法[J]. 计算机科学, 2024, 51(12): 181-189.

HAN Chong, FAN Weibei, GUO Ao. Millimeter Wave Radar Human Activity Recognition Algorithm Based on Feature Fusion [J]. Computer Science, 2024, 51(12): 181-189.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.231200170

基于特征融合的毫米波雷达行为识别算法

韩 崇 樊卫北 郭 澳
南京邮电大学计算机学院 南京 210023

摘 要 基于毫米波雷达的人体行为识别方法以远程非接触的方式捕获人类活动的电磁波信号并进行识别,不受烟雾和光线等的干扰,具有一定的隐私保护性,是当前的一个研究热点。针对现有的算法存在特征输入单一、模型结构复杂、泛化能力验证性不够等问题,提出了基于双分支特征融合卷积神经网络(Two Steam Features Fusion Convolutional Neural Network,2S-FCNN),使用搭载注意力机制的残差神经网络作为骨干网络,并行输入时间距离图和时间速度图,采用特征加权分数融合的方式融合特征后进行分类识别,实现了较高的识别准确率。在公开数据集和自建数据集上与现有的其他算法进行了深入的对比实验,实验结果表明所提算法在识别率和泛化能力方面都具有良好的性能。

关键词:毫米波雷达;行为识别;特征融合;注意力机制
中图分类号 TP391.41

Millimeter Wave Radar Human Activity Recognition Algorithm Based on Feature Fusion

HAN Chong,FAN Weibei and GUO Ao
College of Computer,Nanjing University of Posts and Telecommunications,Nanjing 210023,China

Abstract The human activity recognition method based on millimeter-wave radar captures the electromagnetic wave signals of human activities in non-contact way for recognition. It is not easily interfered by smoke and light,which has a certain degree of privacy protection,and has become a research hotspot at present. However,the existing algorithms have some problems,such as single feature input,complex model structure,and insufficient generalization ability verification. A human activity recognition algorithm with two stream feature fusion convolutional neural network is proposed,named 2S-FCNN,which uses the residual neural network equipped with attention mechanism as the backbone network,inputs the time-distance image and the time-velocity image in parallel,and uses the feature weighted score fusion method to fuse the features for classification and recognition,so as to achieve a high recognition accuracy. A set of in-depth comparative experiments are conducted with other existing algorithms on both public and self built datasets,and the experimental results show that the proposed algorithm has good performance in recognition rate and generalization ability.

Keywords Millimeter wave radar,Human activity recognition,Feature fusion,Attention mechanism

1 引言

近年来,人类行为识别(Human Activity Recognition,HAR)是一个热门研究领域。在人机互动的游戏领域[1],通过HAR可以更加智能地响应玩家的动作;在医疗监护领域[2],HAR可以提供实时的监测,必要时向医生或病人家属发出警报;在医疗康复领域,HAR可以监测患者的运动能力[3],根据患者的反馈进行调整,从而帮助患者康复;在安全保障领域[4-5],HAR可以监测和识别异常行为,从而提高安全性。除此之外,HAR还可以被应用在智能家居、交通安全等领域。

基于人体行为识别的发展现状,现有方法大体上可以分为3类:基于光学摄像头的人体行为识别、基于可穿戴式设备的人体行为识别,以及基于无线电信号的人体行为识别。其中无线电信号的种类又可以分为Wi-Fi信号、RFID信号和雷达射频信号。

基于光学的传统人体行为识别一直都是计算机视觉领域的研究热点,在性能和识别准确率上都取得了一定的成果。从深度学习网络类型上可以将基于光学摄像头的行为识别划分为双流模型和时空模型。Simonyan等[6]最早提出经典的时间流和空间流融合方式,实现对人体行为的判断。Tran等[7]提出了3D卷积神经网络(Convolutional 3D Network,C3D),其使用3D卷积和3D池化,能够从连续视频帧中同时获取到时空特征。虽然基于光学的人体行为识别在现阶段已经能够达到较高的识别精度,但是仍然存在两个严重的问题:1)光学设备采集数据依赖于场景光线,如果光线昏暗,或者在黑夜环境下,基于光学的人体行为识别系统会受到严重影响;2)光学设备在一些特殊场景中采集到的数据存在隐私问题。

到稿日期:2023-12-25 返修日期:2024-05-06
基金项目:国家自然科学基金(62272242)
This work was supported by the National Natural Science Foundation of China(62272242).
通信作者:韩崇(hc@njupt.edu.cn)

基于可穿戴式监测设备的人体行为识别算法，一般通过内置的陀螺仪和速度传感器获取人体在活动时瞬间造成的加速度变化数据，从而进行活动分类。Yin 等[8]提出一种基于人体穿戴式无线传感器的异常行为检测方法，通过核非线性回归(Kernel NonLinear Regression，KNLR)从一般正态模型导出异常活动模型，并使用支持向量机对结果进行分类。虽然可穿戴式监测设备在检测时能够达到较高的灵敏度，但佩戴设备会影响人体舒适度，故难以得到大规模的推广和应用。

近年来，基于 Wi-Fi 和 RFID 无线电信号的人体行为识别算法逐渐受到研究者的广泛关注。基于 Wi-Fi 的人体行为识别算法中，Ma 等[9]设计了深度神经网络，从信道状态信息(Channel State Information，CSI)数据的不同角度学习与人的位置和行为相关的特征，实现了对个人的定位与坐、站、行走等行为的识别。基于 RFID 的人体行为识别算法不同于基于 Wi-Fi 的识别算法，其设备需要人来携带电子标签。Yao 等[10]设计了一种基于 RFID 的被动式行为识别系统，该系统基于接收信号强度指示器(Receive Signal Strength Indicator，RSSI)所采集的数据，再结合支持向量机，对 23 种人体行为的识别准确率达到了 90%以上。虽然基于 Wi-Fi 和 RFID 的人体行为识别系统均能够在某些场景下对人体的日常行为完成分类，但是 Wi-Fi 和 RFID 信号存在着容易受其他信号干扰的缺陷。

在无线信号设备中，毫米波雷达利用介于微波和远红外之间的毫米波进行探测，具有带宽较高的优点，可实现高精度的多普勒、距离和角度探测。相比视觉传感器，毫米波雷达具备全天候、远距离的探测性能，不受环境光照强度影响，具备很强的穿透性能，因此使用毫米波雷达进行人体行为识别逐渐成为一个研究热点。基于毫米波雷达的人体行为识别系统的工作原理是通过接收雷达的回波信号，并对该信号进行一系列的信号处理操作以生成 2D 的微多普勒特征图像。通过将射频信号转化为 2D 图像，就能够使用深度学习的方法对图像进行分类和识别。基于雷达的无接触式人体行为识别算法，不仅能够有效地保护隐私问题，还能够在夜间光线昏暗的情况下工作，因此具有很高的研究和应用价值。

然而在现有的基于雷达特征行为的识别工作中，神经网络的输入通常为单一特征，主要为时间-距离图、时间-多普勒图、距离-多普勒图中的一种，随着行为更为复杂，分类数量更多，以单一特征作为输入的模型的分类性能逐渐下降，并且区分易混淆的行为的能力也会变弱。速度信息能够很好地反映人体运动过程中的一些状态的变化，对人体的行为信息进行表述。但距离域同样包含丰富的信息，当人体靠近或远离雷达时，肢体运动也可以在距离信息中表现出来，故本文在特征上考虑对速度和距离特征进行融合，提出了以时间距离图和时间速度图作为输入的双分支特征融合卷积神经网络 2S-FCNN(Two Steam Features Fusion Convolutional Neural Network)，以此实现人体行为识别任务。

本文的主要贡献如下：

1)针对当前基于雷达人体行为识别数据集较少的问题，本文使用商用德州仪器(Texas Instruments，TI)IWR6843 毫米波雷达采集人体行为数据，创建一个分类数更多、动作更复杂的人体行为识别数据集。

2)随着行为更为复杂，分类数量更多，以单一特征作为输入的模型的分类性能逐渐下降。针对采集到的人体行为雷达回波数据，进行滤波、加窗等预处理后，以双流的方式将速度时间和距离时间特征分别输入 ResNet18 主干网络，配合注意力机制模块细化特性信息，使用特征加权分数进行特征融合，提出了多特征融合的人体行为识别算法。

3)将本文提出的算法模型和现有算法模型在自建数据集和公开数据集上进行了对比实验，并从准确率、精准率、召回率、F1-Score、模型的计算量、参数量和推理时间 7 个维度对其进行了综合评估。

2 相关工作

当前在基于雷达射频信号的人体行为识别工作中，广泛使用了基于微多普勒频谱图实现人体行为识别的算法。Taylor 等[11]提出了一种基于雷达微多普勒频谱图、图像分类和机器学习算法的行为识别系统，使用 Glasgow 大学提供的公开数据集，在采用卷积神经网络(CNN)和主成分分析(Principal Component Analysis，PCA)算法作为降维方法进行数据增强后，实现了 95.30%的最佳分类准确度。该论文中使用单通道的索引图像进行分类，但这种方法可能会丢失 RGB 图像中的一些重要信息，从而对分类性能产生不利影响。因此，本文在特征图的输入上，考虑采用完整的 RGB 通道以保留更多的信息，以期获得更好的识别性能。Saeed 等[12]使用残差网络模型 ResNet[13]完成对从 FMCW 雷达提取的微多普勒特征的行为识别。结果表明，该方法的总体准确率达到了 96%。然而，该团队并未公布具体使用的 ResNet 网络参数的相关信息。Han 等[14]采用非接触式脉冲无线电超宽带(Impulse Radio Ultra-WideBand，IR-UWB)雷达传感器和 CNN 模型构建了行为识别系统，使用 IR-UWB 传感器采集被试者的运动信号，并对其进行滤波处理，以消除环境噪声的影响。接着，将滤波后的信号转换为 2D 图像，并利用 CNN 模型对不同类别的图像进行分类，从而实现对跌倒事件的识别。该方法能够有效地区分“跌倒”和其他“非跌倒”日常生活活动，并且在实验中获得了高达 96%的准确率。但该团队的自建数据集只设计了 5 个活动类型和 3 个测试者的数据进行实验，数据样本数量较少，不能很好地反映实际应用场景，对算法的泛化能力存在一定的影响。Victoria 等[15]提出了一种基于塔式卷积神经网络(TowerCNN)的人类活动识别方法，他们利用 Glasgow 大学提供的公开数据集，提取出微多普勒特征图。接着将这些图像分别转化为 RGB，HSV 和 LAB 3 种颜色空间下的特征图像，并使用这些特征图像作为 TowerCNN 的输入，对不同类别的特征图像进行训练和优化。该方法通过融合不同颜色空间下的特征来提升分类性能，并利用 TowerCNN 的三分支结构融合不同通道的信息，有效地实现了对人类活动的准确识别。实验结果表明，该方法取得了 97.58%的准确率。虽然该算法在识别精度上取得了较好的结果，但其重点是对图像的色彩频谱进行分析，因此在雷达特征图色彩信息并不丰富的实验环境下，其模型能力会受到限制。Abdu 等[16]使用 VGG19[17]和 AlexNet[18]，并采用迁移学习的方式，以时间速度图作为特征输入，成功地将分类准确率分别提升至 96.86%和 94.57%。随后，

该团队进一步使用双分支网络，其中 VGGNet 和 AlexNet 分别作为特征提取网络，以时间速度图作为输入，通过并行提取高阶特征，并结合通道注意力机制进行特征融合，最终使用 CCA(Canonical Correlation Analysis)方法进行特征融合后输出至 SVM 进行分类预测。该方法在 Glasgow 公开数据集上取得了 99.70%的准确度。该算法存在的问题是，模型使用的是参数量大且推理能力较差的 VGGNet 和 AlexNet 网络，网络训练和推理所需时间较长，并且 Abdu 等仅在 Glasgow 公开数据集上验证了模型的性能，因此对该模型的泛化能力的评估有所影响。

总的来说，现有的基于雷达的行为识别分类算法在分类数量较少、动作较为简单的 Glasgow 公开数据集或自建数据集上能够取得较高的准确率，但是仍然存在一定的问题，并且在利用雷达射频信号所处理特征信息来提高识别精度问题上，仅考虑了速度信息，且侧重于对微多普勒特征图视觉信息的分析。在行为识别的过程中，距离域同样包含非常重要的信息，人体在远离靠近雷达的过程中，不同部位的肢体都会有不同的表现。所以本工作不仅使用了时间速度图，同时还从时间距离图中提取特征，以帮助判断行为结果。

针对现有雷达人体行为算法存在的问题，本文提出双分支特征融合卷积神经网络 2S-FCNN，并在公开数据集和自建数据集上验证其性能。

3　调频连续波雷达测距原理

雷达测距是利用电磁波的传播速度和回波时间来计算目标物体与雷达之间的距离的。雷达发射出一束电磁波，当它遇到目标时会被反射回来，雷达接收器接收到这个回波信号并测量出回波时间，然后根据电磁波在空气中的传播速度和回波的时间差计算出目标物体与雷达之间的距离。

图 1 给出了同一线性调频脉冲信号的频率时间函数。FMCW 信号被称为啁啾(Chirp)，也叫做调频信号，调频信号的频率随时间线性增加。设 Chirps 的起始频率为 f_s，截止频率为 f_e，持续时间为 T，带宽为 B，频率的变化率同时也是该线性调频脉冲的斜率，用 S 表示。调频信号可以用正弦函数表示如下：

$$x_T(t)=\cos(2\pi f_s t+\pi S t^2) \tag{1}$$

其瞬时频率可以通过对式(1)求导得到，如式(2)所示：

$$f(t)=\frac{1}{2\pi}\frac{d}{dt}(2\pi f_s t+\pi S t^2)=f_s+St \tag{2}$$

从图 1 可得到斜率 S 的计算式为：

$$S=\frac{B}{T} \tag{3}$$

图 1 中黄色实线表示发射信号 $XT(t)$，蓝色实线表示回波信号 $XR(t)$，它们之间的时延 t_s 表示接收天线接收到发射天线发射调频信号与返回接收天线的时间差。假设目标物体距离雷达的距离为 d，光速为 c，则时延为：

$$t_s=\frac{2d}{c} \tag{4}$$

随着 FMCW 信号 $XT(t)$ 的频率随时间线性增加，$XT(t)$ 和 $XR(t)$ 的瞬时频率，即 $fT(t)$ 和 $fR(t)$ 在混合时是不同的。可以通过低通滤波器滤除频率为 $fT(t)+fR(t)$ 的信号分量，同时允许频率为 $fT(t)-fR(t)$ 的信号分量通过。最后可以得到中频信号，表示如下：

$$x_{IF}(t)=LPF\{x_T(t)x_R(t)\}=A\cos(2\pi f_{IF}t+\Phi_{IF}) \tag{5}$$

$$f_{IF}=f_R(t)-f_T(t)=St_s=S\frac{B}{T} \tag{6}$$

其中，f_{IF} 表示中频信号的频率，雷达的混频器会将发射信号和回波信号混合得到中频信号 $X_{IF}(t)$，也称为拍频、中频、差或者差拍信号；Φ_{IF} 表示中频信号的相位。中频信号的表达式为：

$$x_{IF}(t)=Ae^{j(2\pi f_{IF}t+\Phi_{IF})} \tag{7}$$

$$\Phi_{IF}=2\pi f_s t_s+\pi S t_s^2\approx 2\pi f_s t_s \tag{8}$$

对中频信号进行 ADC 采样，随后进行 FFT 以提取信号的频率信息。假设 FFT 得到频谱的谱峰值对应的频率为 f_m，则有以下公式推导，得出距离信息。

$$f_m=St_s=\frac{2Sd}{c}=\frac{2Bd}{cT} \tag{9}$$

$$d=\frac{cTf_m}{2B} \tag{10}$$

根据相位差的变化，可以计算出目标物体的速度。具体来说，假设目标物体的速度为 v，则回波信号的频率变化量(即多普勒频移)为：

$$\Delta f=\frac{2v}{\lambda} \tag{11}$$

其中，λ 为发射信号的波长。根据工作频率 f_0 和调制频率 f_{mod} 能够计算出波长 $\lambda=\frac{c}{f_0-f_{mod}}$。又因为相位差是与时间和频率的乘积成正比的，因此相位差可以表示为：

$$\Delta\Phi=2\pi\Delta f t_s=\frac{4\pi v}{\lambda}t_s \tag{12}$$

故，推导出速度的计算式 v：

$$v=\frac{\lambda}{4\pi t_s}\Delta\Phi \tag{13}$$

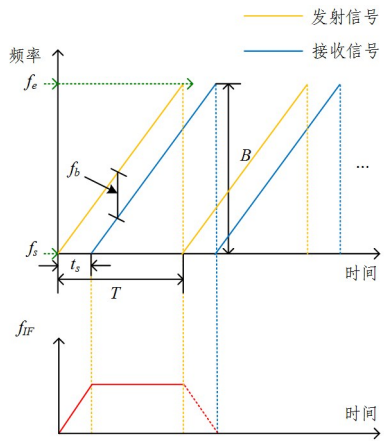

图 1　调频连续波雷达频率时间关系图(电子版为彩图)

Fig. 1　FMCW radar frequency-time chart

4　本文方法 2S-FCNN

本文提出的基于毫米波雷达的人体行为识别方法分为 3 个阶段。第一阶段利用 IWR6843 毫米波雷达平台在场景中收集不同用户行为对应的雷达回波数据。第二阶段在对回波数据进行数据重排、滤波、加窗等预处理操作后，再对数据进行 FFT 操作以提取距离信息，并对射频帧进行叠加得到时间距离图；然后在时间距离图的距离维度上进行 STFT，得到

距离-多普勒频移矩阵，在多普勒维度上进行 FFT，得到速度-距离矩阵；最后在距离维度上进行堆叠，得到速度时间特征。第三阶段将获取到的距离时间图和速度时间图并行输入 2S-FCNN，并使用特征加权分数融合策略对特征进行融合从而实现行为识别。在网络的基础框架上，使用 ResNet18 作为网络架构，并使用 CBAM 注意力机制细化特征信息。本文工作的整体算法框架如图 2 所示。

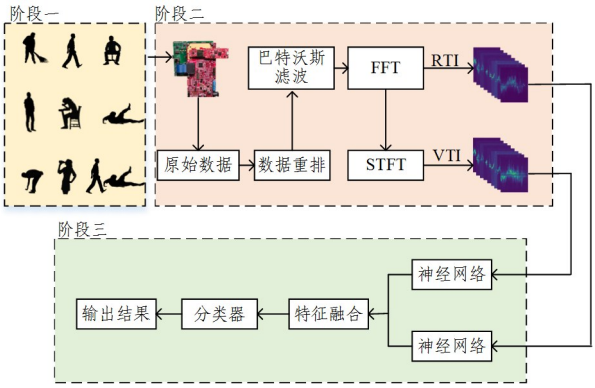

图 2 算法框架

Fig. 2 Framework of the proposed algorithm

4.1 速度时间和距离时间特征获取

在提取速度和距离信息之前，需要先对采集到的数据进行直流去除、杂波滤除以及频谱泄露抑制操作。直流指雷达接收机输出的恒定电压信号，通常表示为 DC(Direct Current)。直流信号(即零频分量)通常被认为是杂波或背景噪声的一部分，因为它代表了被测物体反射回来的雷达波的平均功率。由于直流信号的功率较大，如果不进行处理就会对后续的雷达信号处理和解调产生干扰。传统的去除直流的方法是使用回波信号减去回波信号的平均功率，或者减去几十帧内的平均功率(动态减除)，效果不佳。本文使用巴特沃斯滤波器对噪声和直流进行滤除操作，并使用汉宁窗改善频谱泄露。

巴特沃斯滤波器(Butterworth)是一种常用的信号处理滤波器，其作用是增强输入信号中的某些频率分量，而抑制或削弱其他频率分量[19]。因此在信号处理中，巴特沃斯滤波器通常被用来实现信号的预处理、降噪、滤波等功能。对于巴特沃斯滤波器而言，n 阶巴特沃斯低通滤波器的振幅和频率关系可用式(14)表示：

$$|H(j\Omega)|=\frac{1}{\sqrt{1+\left(\frac{\Omega}{\Omega_c}\right)^{2n}}} \tag{14}$$

其中，Ω 表示工作频率，Ω_c 表示截止频率(中心频率)。本文中使用 4 阶的巴特沃斯低通滤波器来去除直流。

时间距离特征的提取方法为在每一帧数据中选取通道进行采样，然后对采样点进行一维 FFT 就能得到距离信息，在累计了所有的射频帧后就得到了时间距离特征图。

在无线信号回波中，相对于波源的运动目标会引起波的频率或波长的变化，反映在回波信号的相位变化上。在计算距离 FFT 后，通过对信号做多普勒 FFT，相当于对短时间内的距离变化求导得到其速度信息，随后在距离维度上进行堆叠，就可得到速度时间特征图。具体过程为：通过在时间-距离矩阵上做距离 FFT，得到距离-多普勒频移矩阵；接下来在多普勒维度上进行二次 FFT，就可得到速度-距离矩阵；最后在距离维度上进行堆叠，得到速度时间特征图。

完成以上步骤获得时间速度图和时间距离图以后，将特征图保存为 2D 图像，可应用于神经网络训练与分类。图 3 给出了处理后的结果。

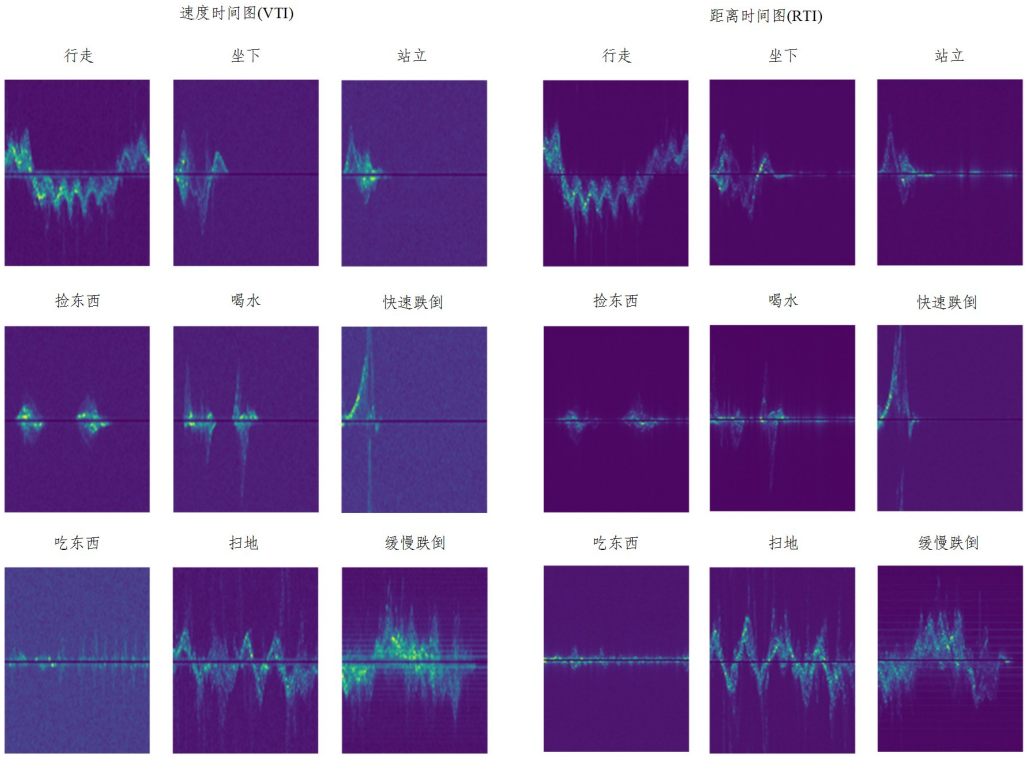

图 3 雷达回波信号处理结果图

Fig. 3 Processing results of radar echo signals

分析频谱图像可以发现，每个活动在时间速度和时间距离信息上都有独特的表现，可以使用这两个信息对各个活动进行区分。CNN 模型可以学习这些属性，有效地实现人体活动分类。

4.2　注意力机制

注意力在人类感知系统中起着至关重要的作用。人类视觉系统的一个显著特征是它不必直接处理整个场景。相反，人类会考虑场景中的一系列镜头，并仔细选择重要位置进行聚焦。注意力机制（Attention Mechanism）是一种模仿人类视觉注意机制的技术，该技术允许模型动态地聚焦于输入序列中具有不同权重的部分，从而更好地捕获上下文信息。

在注意力机制的使用上，本文选择了轻量级的注意力模块[20] CBAM(Convolutional Block Attention Module)，它可以在通道和空间维度上进行计算。CBAM 注意力机制能够让网络在学习的过程中更关注于图像中的目标，它包含 2 个独立的子模块：通道注意力模块（CAM）和空间注意力模块（SAM），在使用的过程中串行计算。本工作将 CBAM 模块添加在网络的头部和尾部。在对特征图输入网络进行第一次卷积后输入 CBAM 模块，能够帮助模型细化低级特征图，使得神经网络在后期的特征提取中学习到更加聚焦于目标的特征。由于本文使用了特征融合的方法，而 CBAM 模块不仅能够对图片的空间信息进行分析，还能够对通道维度进行细化的特征提取，因此其非常契合本工作的需求。同时，添加在网络尾部的 CBAM 模块能够让融合后的特征在空间和维度信息上进一步细化，使得网络在学习的过程中更关注图像中的物体。

具体来说，CBAM 的输入为 $F\in R^{C*H*W}$，然后是通道注意力模块一维卷积 $M_c\in R^{C*1*1}$，将卷积结果与原图相乘，将 CAM 输出结果作为输入，进行空间注意力模块的二维卷积 $M_s\in R^{C*1*1}$，再将输出结果与原图相乘。计算过程如式(15)所示：

$$F'=M_c(F)\otimes F \tag{15}$$

$$F''=M_s(F')\otimes F' \tag{16}$$

通道注意力模块将输入的 Feature maps 经过两个并行的 MaxPool 层和 AvgPool 层，将特征图从 $C\times H\times W$ 变为 $C\times 1\times 1$ 的大小，然后经过 Share MLP 模块。在该模块中，先将通道数压缩为原来的 $1/r$（Reduction，减少率），再扩张到原通道数，经过 ReLU 激活函数得到两个激活后的结果。将这两个输出结果进行逐元素相加，再通过一个 Sigmoid 激活函数得到 Channel Attention 的输出结果，最后将这个输出结果乘原图，变回 $C\times H\times W$ 的大小。计算过程如式(17)所示：

$$M_c(F)=\partial(MLP(AvgPool(F))+MLP(MaxPool(F))) \tag{17}$$

空间注意力模块将 Channel Attention 的输出结果通过最大池化和平均池化得到两个 $1\times H\times W$ 的特征图；然后经过 Concat 操作对两个特征图进行拼接，通过 7×7 卷积变为 1 通道的特征图；再经过一个 Sigmoid 得到 Spatial Attention 的特征图；最后将输出结果乘原图变回 $C\times H\times W$ 大小。计算过程如下所示：

$$M_s(F)=\partial(f^{7*7}([AvgPool(F);MaxPool(F)])) \tag{18}$$

4.3　特征融合

深度学习领域的多特征融合的策略首先经过神经网络提取出高阶的特征信息，然后对高阶的信息特征进行融合处理。常用的特征融合方式有通道拼接、特征相加、特征相乘以及加权融合等。特征通道拼接 contact 方式将多个输入产生的高阶特征信息按照通道维度进行拼接，获得一个和特征图尺寸大小一样、通道数变为多个特征通道总和的高阶特征信息；对于大小和通道完全相同的高阶特征信息，也可以直接对高阶的特征信息相加或相乘进行融合，这样融合后的特征维度和某一个单一特征所提取出的高阶特征信息维度完全相同。

本文采用的特征融合方式与前三种方式不同，采用多分支输入网络，在网络的最后阶段使用 SoftMax 分类器计算网络输出 x，得到来自各分支网络对应的 SoftMax 分数，然后将它们的分数按照权重相加，计算过程如式(19)和式(20)所示：

$$\mathrm{softmax}(x)_i=\frac{\exp(x)_i}{\sum_{j=1}^{n}\exp(x_j)} \tag{19}$$

$$FusionScore=P_1*\mathrm{softmax}(Net_1(x_k)_{k\in(1\sim c)})+P_2*\mathrm{softmax}(Net_2(x_k)_{k\in(1\sim c)}) \tag{20}$$

其中，P_1 和 P_2 表示权重系数，总和为 1；$Net(x)$ 表示神经网络的输出；c 表示分类数。该方法不需要在计算的过程中控制特征的维度信息。

4.4　骨干网络

本文使用的神经网络架构以 ResNet18 作为主体，配合 CBAM 注意力机制，网络框架如图 4 所示。ResNet18 是一种深度卷积神经网络，ResNet18 的大体构成分为 3 个部分，分别是 Stem，Body 以及 Head。Stem 由 1 个卷积和 1 个最大池化层构成；Body 包含 4 个 layer，每个 layer 由 2 个 Block 组成；Head 则包含 1 个平均池化和 1 个全连接层，总计由 18 个卷积层和 1 个全连接层组成。

具体来说，雷达热图以 $3\times 128\times 128$ 三维张量的形式输入至 ResNet 的第一层。第一层是 7×7 的卷积层（*kernel size*=7，*stride*=2），64 个输出通道，*padding*=3，不改变特征图大小。随后执行正则化（Batch Normalization）和 ReLU 激活函数操作，并输入 CBAM 注意力机制中，然后是一个 3×3 的最大池化层（*kernel size*=3，*stride*=2）。接下来的二、三、四、五层（layer），每一层均包含一个残差块，每个残差块都包含两层卷积操作，最后特征图会被进行一次全局平均池化，得到 1024 维度的特征向量，随后输入全连接层的 1000 个概率分布上，最终输出分类结果，并经过 SoftMax 分类器计算得到最终输出。在另一个以时间距离图作为输入的分支网络，进行相同的计算。

为了使网络能够提取更为细化的特征，并且不破坏 ResNet18 本身的结构，使得模型能够继续加载 ResNet18 的预训练模型，加快网络收敛，本文在 ResNet18 的基础上在 Head 位置额外添加了一层全连接层，在 Stem 和 Head 位置添加了 CBAM 注意力机制。在特征融合方面，本文工作采用了加权分数融合策略，以提高最终识别效率。

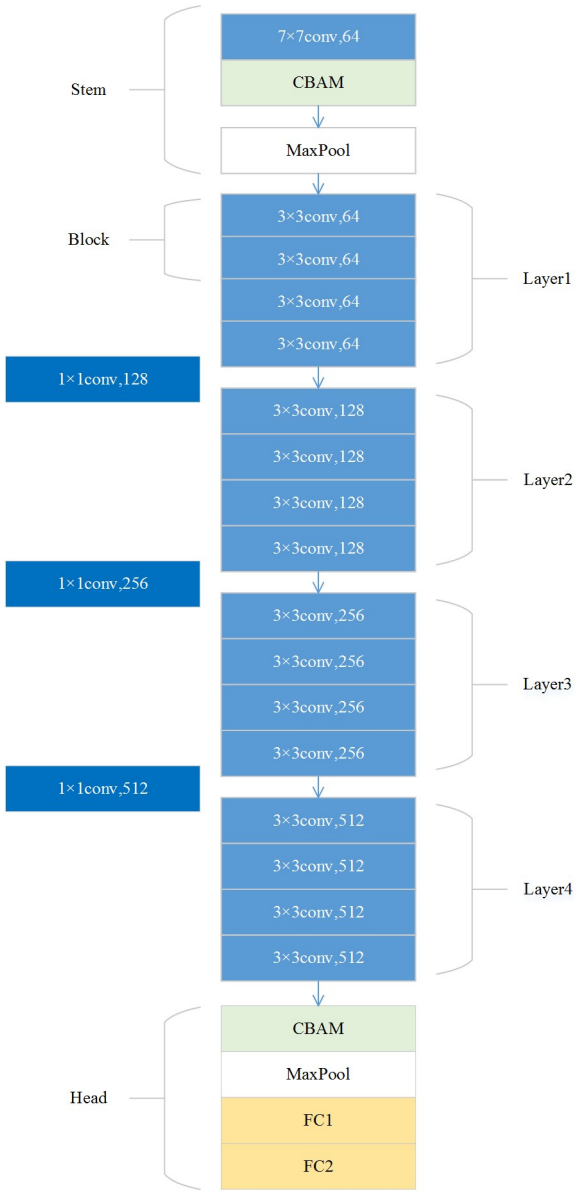

图 4 模型网络结构

Fig. 4 Model network structure

5 实验结果分析

5.1 数据集介绍

1)Glasgow 公开数据集

Glasgow 公开数据集是一个可供研究人员进行雷达人体活动识别应用的基准数据集，包含了 56 个年龄在 21～98 岁的被试人员在 9 个不同环境中(包括室内和室外)进行 6 种不同的活动。每个测试人员都要重复进行行走、起立、坐在椅子上、弯腰捡起物品、从玻璃或瓶子中喝水等活动。由于安全原因，只有部分被试人员模拟了前倾摔倒。Glasgow 数据集录制过程中所使用的实验设备是由 Ancortek 公司制造的 FM-CW 雷达，工作频段为 5.8 GHz，带宽为 400 MHz。

2)自建数据集

本文同时进行了自建数据集的采集，共安排了 10 名实验对象，在场景中分别进行如图 5 所示的动作。图 6 则展示了在录制坐下行为时的实际场景。实验设备是 TI IWR-6843 商用毫米波雷达，其工作频率为 60 GHz。数据集行为动作涉及行走、坐、站、拾起物体、喝水、摔倒、吃饭、扫地和走路时缓慢跌倒 9 种行为。

图 5 人体行为示意图

Fig. 5 Diagram of human behaviors

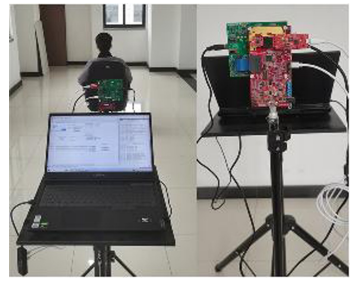

图 6 场景和设备展示图

Fig. 6 Diagram of experimental scenario and equipment display

将数据集进行随机划分，80%用于训练，20%用于测试。使用 0.001 的学习率和随机梯度下降作为优化器对所有模型进行训练。此外，在整个实验中训练轮数为 30，批次为 32。模型运行在搭载有 GTX2080 显卡的服务器上。

5.2 实验对比与分析

1)基准模型实验结果

为了对 2S-FCNN 模型的性能进行比较和评估，首先将基准骨干网络 ResNet18 在 Glasgow 公开数据集和自建数据集上进行实验。为了确保公平性，在公开数据集和自建数据集进行相同分类对象和分类数量的 6 分类实验。在实验中，采用迁移学习(Transfer Learning，TL)方法，加载 ResNet18 预训练模型，冻结网络卷积层的参数，并微调分类器。在数据集上进行训练，分别基于速度时间图特征(VTI)和距离时间图特征(RTI)取得了如图 7 所示的结果。

通过实验结果可以发现，骨干网络 ResNet18 在公开数据集和自建数据集上，使用时间速度图作为网络输入时的性能优于使用时间距离图。在对 ResNet18 进行迁移学习后模型的性能有所提升，再加上通道注意力机制细化模型的特征图以后，ResNet18 模型能够取得较好的结果。

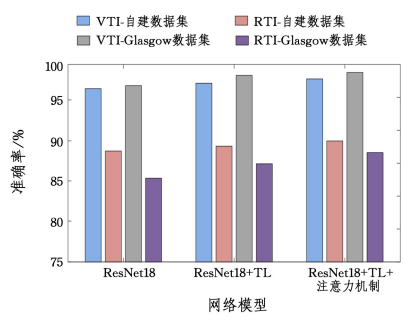

图 7 基准模型实验结果

Fig. 7 Experimental results of baseline model

2）融合策略分析

接下来将介绍基于时间速度图和时间距离图双输入的融合模型在公开数据集和自建数据集上的实验结果。为此，将6分类Glasgow公开数据集以及9分类自建数据集样本按8∶2的比例随机划为训练集和测试集进行实验。

融合模式1　Resnet18＋TL＋Attention＋Concat，指按照通道拼接的融合策略，网络分别对时间速度和时间距离图进行高阶特征提取，按照通道拼接。

融合模式2　Resnet18＋TL＋Attention＋FeatureAdd，指直接相加的特征图融合策略中，采用了与按通道拼接类似的操作，将两个分支的特征矩阵直接相加，得到相加融合特征图。

融合模式3　Resnet18＋TL＋Attention＋FeatureMulti，指直接相乘的特征图融合策略，与融合模式2相同，将两个分支的特征矩阵直接相乘得到相乘融合特征图。

2S-FCNN　本文采用的融合策略是基于权重的特征融合方式。

表1列出了不同融合策略下的实验结果。

表1　融合策略实验结果

Table 1　Experimental results of different fusion strategies

模型	准确率/%	
	公开数据集	自建数据集
融合模式1	98.75	97.42
融合模式2	97.66	96.54
融合模式3	95.32	93.41
2S-FCNN	99.70	99.49

在融合模式1中，由于两个输入的特征提取网络结构相同，因此它们提取出的特征具有相同的维度，可以直接按照通道拼接。拼接后的特征图通道数增加到原先的两倍，再进一步通过卷积操作获取更深层次的特征信息。实验结果表明，与仅使用时间速度图作为单一输入的网络相比，采用该融合策略能够获得更好的性能。

在直接相加和相乘的特征图融合策略2和策略3中，采用了与按通道拼接类似的操作，将两个分支的特征矩阵直接相加或相乘以得到一个新的特征图。随后，对相加或者相乘后的特征图进行进一步卷积处理，但是相加和相乘两种策略下的识别精度低于仅使用时间速度图作为输入的结果。

本文采用的特征融合策略是分别对时间速度图和时间距离图进行完整的网络计算，并使用SoftMax分类器分别得到两个分支网络的预测分数。然后，按照权重对它们的结果进行融合，并根据聚合后的分数进行预测。对于权重的划分方案上，进行了表2所列的4种实验。

表2　权重分数融合/权重划分实验结果

Table 2　Experimental results of weight score fusion/weight division

融合权重	准确率/%	
	公开数据集	自建数据集
5∶5	95.74	94.34
7∶3	98.43	97.16
8∶2	99.70	99.49
9∶1	98.57	98.12

从表2可以看出，采用5∶5比例对双分支网络的结果进行预测，最终得到的结果劣于仅使用单一输入的结果。这是因为在行为分类任务中，单独使用时间速度图的效果优于单独使用时间距离图，直接相加会导致整体结果变差。同样地，如果时间速度图的权重占比过大，则会降低时间距离图在整体行为预测中的作用。通过多次实验，确定时间速度图和时间距离图的最优权重比为8∶2。

3）基于Glasgow公开数据集的实验结果及对比

本文最终采用加权融合的形式，将双输入特征融合模型与当前的模型在公开数据集上进行了10次实验，并获得了总体准确率、召回率、精确率和F1-Score的对比结果，具体如表3所列。

表3　对比实验结果（公开数据集）

Table 3　Comparative experimental results(public dataset)

模型	准确率/%	召回率	精准率	F1-Score
ResNet18[13]	98.40	0.9221	0.9180	0.9200
Taylor 等[11]	95.30	0.9917	0.9927	0.9919
VGG19[17]	96.86	0.9583	0.9603	0.9591
Victoria 等[15]	97.58	0.9500	0.9500	0.9500
Abdu 等[16]	99.77	0.9689	0.9704	0.9696
2S-FCNN	99.70	0.9976	0.9975	0.9970

在公开数据集上的实验结果表明，仅使用时间速度图作为输入的ResNet18网络，在应用了通道注意力机制和迁移学习后，其准确率可达到98.40%。而本文提出的特征融合模型在该数据集上的准确率可达99.70%，接近Abdu 等[16]提出的VGG＋Alex＋CCA＋SVM模型架构的最优准确率（99.77%）。因此，2S-FCNN在公开数据集上的准确率能够逼近当前最优算法。

对模型的性能的评价，不能仅局限于准确率的高低，还应该对模型的大小、推理时间和计算量进行评估。因此本文还对模型的计算量、参数数量和推理时间进行了比较。模型推理时间的计算是在设置BatchSize为32的情况下在GTX-2080显卡上计算所得。实验结果的具体数据如表4所列。从表中可以看出，本文提出的双输入特征融合网络2S-FCNN在公开数据集上的整体准确率低于Abdu 等[16]提出的VGG＋Alex＋CCA＋SVM模型，但是在模型计算量、参数量和推理时间上性能更加优越。

表4　模型计算量、参数量和推理时间对比

Table 4　Comparison of model computation, parameter quantity, and inference time

模型	Flops	Params	Time Inference/ms
ResNet18[13]	1.01Gmac	11.27×10^6	214
VGG19[17]	4.92Gmac	33.68×10^6	623
Taylor 等[11]	674.35Mmac	24.32×10^6	301
Victoria 等[15]	979.28Mmac	27.87×10^6	412
Abdu 等[16]	4.96Gmac	30.26×10^6	368
2S-FCNN	2.04Gmac	22.83×10^6	284

由于公开数据集所包含的分类数量过少，仅包含6个行为动作，并且行为动作之间的区别明显，并不包含复杂动作，为了更进一步地验证2S-FCNN模型的性能，接下来将在分类数更多、动作更为复杂的自建数据集上进行对比。

4）基于自建数据集的实验结果及对比

本文将9分类自建数据集按8∶2的比例随机划分为训练集和验证集，2S-FCNN进行了10折实验并取平均结果后，得到了基于SoftMax权重相加融合策略的双输入模型的总体准确率和平均准确率，具体实验数据如表5所列。

表 5 2S-FCNN10 折实验结果(自建数据集)

Table 5 10-fold experimental results of 2S-FCNN(self-built dataset)

(%)

序号	准确率
1	99.43
2	99.24
3	99.43
4	99.43
5	99.62
6	99.81
7	99.25
8	99.62
9	99.43
10	99.62
AVG	99.49

在自建数据集上,2S-FCNN 进行 10 次实验,平均准确率能够达到 99.488%。表 6 列出了 2S-FCNN 和当前的主流模型在自建数据集上总体准确率、精准率、召回率以及 F1-Score 的对比结果。

通过对比结果可以发现,在自建数据集上,VGG19 表现最差,特别是在扫地和走路时缓慢跌倒等行为分类上不能很好地进行区分,而且随着分类行为的增加和行为复杂性的提高,其网络预测能力也有所下降。这可能是因为扫地的动作、走路时缓慢跌倒的动作以及行走动作都包含了移动过程,导致网络在识别的过程中不能准确区分。与此不同的是,本文提出的双输入模型使用时间速度图和时间距离图,能够结合更多特征和行为活动信息,因此行为分类能力优于其他模型。

表 6 对比实验结果(自建数据集)

Table 6 Comparative experimental results(self-built dataset)

模型	准确率/%	召回率/%	精准率/%	F1-Score
2S-FCNN	99.49	99.76	99.75	0.9970
ResNet18[13]	97.02	95.83	96.03	0.9591
VGG19[17]	83.78	96.89	97.04	0.9696
Victoria 等[15]	92.71	92.21	91.80	0.9200
Abdu 等[16]	97.97	99.17	99.27	0.9919
Taylor 等[11]	94.23	95.00	95.00	0.9500

在自建数据集中,对当前先进模型进行了 10 次实验,每个行为分类的平均准确率如表 7 所列。

表 7 模型每个行为预测结果对比

Table 7 Comparison of predicted results for each behavior in different models

模型	行走	坐下	站立	捡东西	喝水	快速跌倒	吃东西	扫地	缓慢跌倒
2S-FCNN	99.5	99.9	99.9	99.8	100	100	99.1	99.8	100
ResNet18[13]	97.4	96.2	97.7	98.3	97.6	99.4	96.23	95.5	96.6
VGG19[17]	57.6	93.3	90.4	96.7	89.4	84.0	98.4	65.2	75.0
Taylor 等[11]	96.2	94.1	94.2	93.2	94.5	96.3	95.4	87.5	90.2
Abdu 等[16]	98.0	97.1	98.5	97.1	98.9	99.7	97.4	89.3	90.4
Victoria 等[15]	95.9	96.7	97.9	95.8	93.0	94.6	94.7	80.0	73.9

结束语 本文对基于毫米波雷达的行为识别算法进行了研究,提出了一种基于雷达和深度学习的特征融合网络 2S-FCNN,阐述了数据采集、数据处理、特征提取和网络架构。最后在公开数据集和自建数据集进行了对比实验,对 2S-FCNN 与现有其他毫米波雷达行为识别算法进行了多方面的实验对比,从准确率、精准率、召回率、F1-Score、模型计算量、模型参数量和模型推理时间这 7 个维度对模型进行了评价。实验对比结果验证了本文提出的算法在识别性能、泛化能力、模型的复杂度等方面都有较好的性能。本文的研究主要是针对分割好的数据对特定行为进行识别,采用少样本或者无样本学习方法对未知行为进行分类识别将是接下来的研究方向。

参 考 文 献

[1] DING C Y, LIU K, LI G, et al. Spatio-Temporal Weighted Posture Motion Features for Human Skeleton Action Recognition Research[J]. Chinese Journal of Computers, 2020, 43(1): 29-40.

[2] ZHANG X P, JI J H, WANG L, et al. Overview of video based human abnormal behavior recognition and detection methods [J]. Control and Decision, 2022, 37(1): 14-27.

[3] DENG M L, GAO Z D, LI L, et al. Overview of Human Behavior Recognition based on Deep Learning[J]. Computer Engineering and Applications, 2022, 58(13): 14-26.

[4] MAQSOOD R, BAJWA U I, SALEEM G, et al. Anomaly recognition from surveillance videos using 3D convolution neural network[J]. Multimedia Tools and Applications, 2021, 80(12): 18693-18716.

[5] LIANG J, ZHU H, ZHANG E, et al. Stargazer: A transformer-based driver action detection system for intelligent transportation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 3160-3167.

[6] SIMONYAN K, ZISSERMAN A. Two-stream convolutional networks for action recognition in videos[J]. arXiv: 1406.2199, 2014.

[7] 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.

[8] YIN J, YANG Q, PAN J J. Sensor-based abnormal human-activity detection[J]. IEEE Transactions on Knowledge and Data Engineering, 2008, 20(8): 1082-1090.

[9] MA Y, ARSHAD S, MUNIRAJU S, et al. Location-and person-independent activity recognition with WiFi, deep neural networks, and reinforcement learning[J]. ACM Transactions on Internet of Things, 2021, 2(1): 1-25.

[10] YAO L, SHENG Q Z, LI X, et al. Compressive representation for device-free activity recognition with passive RFID signal strength[J]. IEEE Transactions on Mobile Computing, 2017, 17(2): 293-306.

[11] TAYLOR W, DASHTIPOUR K, SHAH S A, et al. Radar sensing for activity classification in elderly people exploiting micro-doppler signatures using machine learning[J]. Sensors, 2021, 21(11): 3881.

[12] SAEED U，SHAH S Y，SHAH S A，et al. Discrete human activity recognition and fall detection by combining FMCW RADAR data of heterogeneous environments for independent assistive living[J]. Electronics，2021，10(18)：2237.

[13] HE K，ZHANG X，REN S，et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016：770-778.

[14] HAN T，KANG W，CHOI G. IR-UWB sensor based fall detection method using CNN algorithm[J]. Sensors，2020，20(20)：5948.

[15] VICTORIA A H，MARAGATHAM G. Activity recognition of FMCW radar human signatures using tower convolutional neural networks[J/OL]. Wireless Networks，2021：1-17. https://doi. org/10. 1007/s11276-021-02670-7.

[16] ABDU F J，ZHANG Y，DENG Z. Activity classification based on feature fusion of FMCW radar human motion micro-Doppler signatures[J]. IEEE Sensors Journal，2022，22(9)：8648-8662.

[17] SIMONYAN K，ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv：1409. 1556，2014.

[18] KRIZHEVSKY A，SUTSKEVER I，HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM，2017，60(6)：84-90.

[19] ALI A S，RADWAN A G，SOLIMAN A M. Fractional Order Butterworth Filter：Active and Passive Realizations[J]. IEEE Journal on Emerging and Selected Topics in Circuits and Systems，2013，3(3)：346-354.

[20] WOO S，PARK J，LEE J Y，et al. Cbam：Convolutional block attention module[C]// Proceedings of the European Conference on Computer Vision(ECCV). 2018：3-19.

HAN Chong，born in 1985，Ph. D，associate professor，master supervisor，is a member of CCF (No. C3132M). His main research interests include wireless sensing and RF computing.

（责任编辑：杨雪敏）

产学研专家秀湖论剑，共话生成式可视媒体未来之路

2024 年 11 月 22 日至 24 日，第二十三期 CCF 秀湖会议在苏州 CCF 业务总部 & 学术交流中心举办，会议为期三天，就“生成式可视媒体”进行深入交流和研讨。浙江大学鲍虎军教授、上海交通大学马利庄教授、北京大学汪国平教授、北京大学彭宇新教授、复旦大学张新鹏教授、上海科技大学虞晶怡教授、中科院软件所田丰研究员等 20 余位学术界专家以及来自华为、VIVO、网易伏羲、联想、腾讯等多位工业界专家出席会议（名单后附）。秀湖会议 AC 主席、CCF 副理事长、中国科学院院士、清华大学胡事民教授在开幕式上致辞并做特邀报告。本次会议由执行主席浙江大学周昆教授和大连理工大学杨鑫教授主持。

CCF 会士、副理事长、中国科学院院士、清华大学胡事民教授以“可视媒体生成：网络模型与编程框架”为题作特邀报告。胡教授回顾了可视媒体生成的发展历程，从计算框架的角度对可视媒体生成的挑战进行了阐述，他希望广大学者关注可视媒体生成的基础性难题，共同打造国产基础生态。

CCF 会士、浙江大学鲍虎军教授以“AI 原生的交互图形技术与系统”为题作特邀报告。鲍教授首先介绍对 AI 时代计算机图形学发展的思考，然后以团队在 AI 原生的神经图形绘制技术及其流水线软硬件架构方面的研究进展作为具体案例，验证并展示上述构想的可能性。

北京大学汪国平教授带以“智能制造中的 CAD/CAE 一体化产品设计优化系统”为题作特邀报告。汪教授首先探讨了 CAD/CAE 的需求背景与存在的问题，然后介绍了团队在基于体细分的通用 CAD/CAE 一体化系统、基于云计算平台的 CAD-CAE 统一设计优化系统等系列研究成果。

北京大学彭宇新教授以“基于细粒度多模态分析的 AIGC”为题作特邀报告。彭教授针对当前 AIGC 普遍存在的逻辑错误问题，以及多模态大模型缺乏细粒度视觉感知与运动分析能力的局限性，提出了“多模态大模型－细粒度视觉感知－细粒度运动分析－AIGC”的研究布局，分享了团队在上述 4 个研究方向的近期研究进展，并对 AIGC 与多模态大模型面临的主要挑战与未来发展方向进行了讨论与展望。

上海交通大学马利庄教授以“面向影视媒体的 NeRF 与 3DGS 技术”为题作特邀报告。马教授点明了 NeRF 与 3DGS 各自的优势与现有的一些问题，并针对部分问题分享了团队在 NeRF 及 3DGS 方法优化上的一系列研究成果。报告提出 NeRF 在渲染效率优化、训练效率优化、重建质量优化等方面还有很大提升空间；而 3DGS 方法未来应该向质量优化、大场景建模，多模态信号建模等方向发展。

会议最后半天，与会嘉宾围绕“生成式可视媒体”进行了专题研讨，一方面在此前三天会议讨论的基础上进行总结并形成共识，另一方面就如何共同推动生成式可视媒体的发展发出倡议。经过热烈的讨论，嘉宾们围绕生成式可视媒体的科学技术问题、生成式可视媒体的共性平台、生成式可视媒体的刚需应用场景等方面形成了初步的共识。

据 CCF 微信公众号