

手机端多模态低头族危险感知与预警^{*}

金磊 吉翔 邓丽云 徐少杰 王晗

(南通大学交通与土木工程学院 南通 226019)

摘 要: 随着智能手机产品的发展与畅销,不分场合随时玩手机的低头族大量涌现;针对低头族依赖手机导致道路交通事故频发问题,提出一种基于手机端多模态低头族危险感知与预警系统。首先,利用手机端的重力加速度基于模糊控制规则实时监测行为,包括:走路看手机、上下楼梯看手机、静止看手机、手持手机走路、手机揣兜走路;然后,使用手机后视摄像机图像基于分组快速空间金字塔池化的轻量化 YOLO 网络实时描述用户周围环境,包括:楼梯、斑马线、低照明环境、积水坑、正常路面。最后,面向安卓系统构建状态-环境多模态低头族危险判定模型;并根据判定结果利用声音、画面、震动信号给予低头族听觉、视觉、触觉立体式预警信号;减少低头族跌伤、碰撞等潜在危险。在线实验表明,本文提出的手机端多模态低头族危险感知模型准确性高、鲁棒性强、实时性好,能够针对低头族常见的危险状态实现有效的主动预警。

关键词: 低头族危险感知与预警;移动手机终端;多模态判定模型;重力加速度传感器;后视摄像头

中图分类号: TP391 **文献标识码:** A **国家标准学科分类代码:** 520.20

Multimodal hazard sensing and warning for bowed-head tribe based on mobile terminals

Jin Lei Ji Xiang Deng Liyun Xu Shaojie Wang Han

(School of Transportation and Civil Engineering, Nantong University, Nantong 226019, China)

Abstract: With the development and popularity of smartphone products, a large number of bowed-head tribes have emerged who play mobile phones at any time regardless of the occasion; for the frequent occurrence of traffic accidents caused by bowed-head tribes' dependence on mobile phones, a multimodal bowed-head tribes' hazard perception and warning system based on mobile phones is proposed. First, gravity acceleration on the mobile phone side is used to monitor behaviors in real time based on fuzzy control rules, including Walking and looking at the mobile phone, Walking up and down stairs, Looking at the mobile phone at rest, Walking with the mobile phone in hand, Walking with the mobile phone in pocket; and then the user's environment is described in real time using the mobile phone's rear view camera images based on the grouping of fast spatial pyramids pooled in the lightweight YOLO network, including: stairs, crosswalks, low-light environments, puddles, and normal road surfaces. Finally, a state-environment-multimodal hazard detection model is constructed for the Android system; and based on the detection results, audible, visual, and tactile three-dimensional warning signals are given to the bowed tribe by using sound, image, and vibration signals to reduce the potential hazards of the bowed tribe such as fall injury and collision. Online experiments show that the proposed multimodal threat perception model for mobile phones is highly accurate, robust, and real-time, and is able to achieve effective proactive warning for the common threat states of bowed heads.

Keywords: hazard sensing and early warning for the bowed-head tribe; mobile phone terminals; multimodal decision models; gravity acceleration sensors; rear view cameras

0 引 言

随着我国经济水平的快速提高,智能制造产业与高科

技电子产品的高速发展,智能手机正在蚕食人们生活的碎片时间;越来越多的人沉迷于手机娱乐活动^[1-3]。21 世纪智能手机的问世衍生出不分时间、场合随时玩手机的低头

族群体。随着低头族队伍的日渐壮大,楼梯上、水池旁、甚至在斑马和低照明区域均出现他们的身影。长时间保持低头看手机的状态,容易引起视觉疲劳、手腕麻木,进而引发颈椎病和精神涣散等身心上的损伤;同时,无视周边环境潜在的危险、持续低头玩手机,可能会发生上下楼梯时踏空、过马路时不避让车辆、路过水坑时滑倒摔伤等事故,甚至会影响到道路交通安全、威胁生命安全^[4-5]。

现有低头族危险检测方法可以根据系统和传感器的布置位置分为:监控设备端检测系统、可穿戴设备端检测系统和移动手机端检测系统。面向监控设备端,王文雅^[6]用改进的深度监督下的小型目标检测(tiny-deeply supervised object detection, Tiny-DSOD)轻量化网络检测办案区域中的行人低头族目标。陈浩^[7]用改进的 MobileNetv2 网络对车载摄像头下驾车低头族进行实时检测与语音警示,取得较好预警效果。施冬梅等^[8]在空间流通道采用卷积神经网络提取空间特征值,时间流通道采用单次多框检测算法(single shot multibox detector, SSD)检测人眼,对车载摄像头画面中驾车低头族的危险行为进行检测,提升了准确率。管文华等^[9]通过腕关节坐标来模拟低头族姿态,利用手臂与头部的相对位置信息通过 Yolov3+Vgg16 融合网络实现对低头族危险行为检测与预警,算法通用性较强。现有监控设备端的检测方法检测范围有限,其只能实现对固定、单一场景的低头族检测与预警,面向移动的行人无法覆盖其全天的活动范围。面向可穿戴设备端,Bi等^[10]设计了一种基于现成无线耳机内置加速度计的系统,通过元学习感知(meta-learning sensing, MetaSensing)头部手势识别模型利用少量样本提高识别能力,以此来判断是否存在低头玩手机的行为,并进行相应的语音预警,该系统实用性较强。Schwebel等^[11]设计了一种基于蓝牙信标技术的低头族预警系统,当分心的行人接近一个有潜在危险的街角时,系统直接通过分散分心的装置发出警告,通过视觉听觉方式对不同状态下的低头族进行预警,该系统通过不断地系统预警提醒来使低头族养成良好习惯。贾森等^[12]提出了一种面向低头族的颈部疲劳基于表面肌电信号(surface electromyography, SEMG)的幅频联合分析模型(joint analysis of EMG spectrum and amplitude, JASA),通过采集用户颈部肌肉的表面肌电信号,利用疲劳联合分析模型对地铁低头族进行检测并发起疲劳预警,为开发智能穿戴系统提供了借鉴意义。郭毅博等^[13]提出一种基于可穿戴设备的深度学习数据特征提取方法,将从加速度计、磁力计、陀螺仪等可穿戴设备中的传感器收集到的时序数据在灵活设置卷积的基础上引入分支结构来多尺度提取原始数据的深度特征,通过将各分支的数据特征融合后作为下一个卷积的输入,以此来准确检测包括低头族行为在内的多种人体行为。现有可穿戴设备仍然需要用户额外购买该设备,且部分可穿戴设备体积较大、不方便,影响用户的正常生活^[14]。面向移动手机端,丁阁文等^[15]通过手机端提取到

的加速度计和陀螺仪数据特征,采用二次多项式核函数的支持向量机(support vector machine with quadratic polynomial kernel, Q-SVM)模型结合“多票判决策略”实时对低头族行为进行检测,提高了分类准确率和鲁棒性。Zhuang等^[16]利用手机GPS、加速度计等组合信号,通过模糊数学(fuzzy mathematics, FM)与邓普斯特-沙费尔证据理论(dempster-shafer evidence theory, DST)的多信息融合算法,对十字路口的低头族进行有效识别并预警。虽然提高了准确率,但缺少了周边环境的危险检测。Zhuo等^[17]提出了一种基于极端随机树算法(extremely randomized trees, ET)的低头族实时活动分类方法,通过智能手机的惯性测量单元(inertial measurement unit, IMU)采集低头族走路和浏览手机、打字、看视频等活动数据来对低头族玩手机行为进行分类。综上所述,现有手机端检测方法存在如下缺陷:1)多采用手机端单一传感器的观测数据,基于单模态观测数据的低头族行为识别模型准确性较低;2)多聚焦于低头族玩手机行为的识别,忽略了其周边的环境信息,无法实现低头族当前潜在危险状态的判别。本文主要的工作内容如下:

1)为了解决模型的尺寸过大问题,本文采用分组快速空间金字塔池化(group fast spatial pyramid pooling, GFSP)的轻量化YOLO网络模型,本文选择GFSP模块和卷积块的注意力机制模块(convolutional block attention module, CBAM)替换快速空间金字塔池化(spatial pyramid pooling fast, SPPF)结构,用轻量化幽灵卷积(ghost convolution, GhostConv)替换颈部网络中的普通卷积。

2)为了解决单模态检测误检问题,本文同时采集手机端的重力加速度序列和后视摄像头画面的两种不同模态的观测数据,构建基于用户行为和周边环境的多模态低头族危险状态判别模型。

3)为了解决单一预警的遗漏问题,本文根据识别结果通过手机画面、语音播报和震动信号给予低头族“视-听-触”多感官立体式全方面预警,提高了预警的实时性和有效性。

1 手机端多模态低头族危险感知与预警系统

本文提出的手机端多模态低头族危险感知与预警系统框架分为5个模块依次为手机端用户传感器数据采集模块、用户运动状态识别模块(行为识别)、用户周围环境识别模块(场景识别)、用户低头族危险状态识别模块(危险感知)、手机端用户预警显示模块组成,通过行为识别和场景识别相结合实时监测危险状态并给予低头族用户多感官全面预警,如图1所示。

1.1 基于GFSP-YOLO模型的手机用户周边环境识别

在YOLO系列模型中,YOLOv5^[18]因其参数体量大、运行速度快等特点被广泛应用于工业实际项目。然而,面

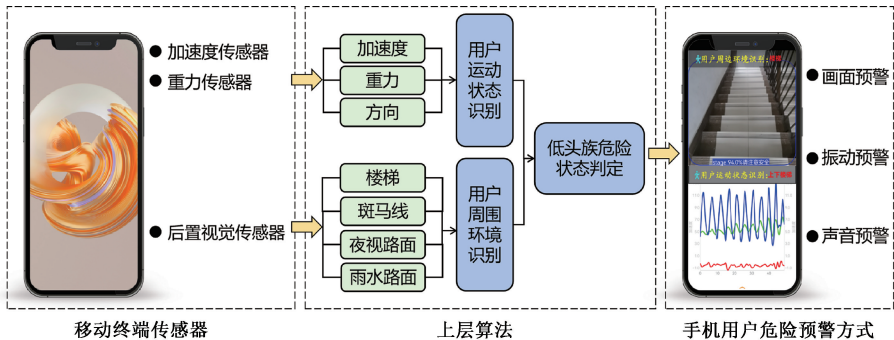


图 1 手机端多模态低头族危险感知与预警系统框架

向手机端的安卓系统嵌入式环境中,面向后视摄像头实时图像检测处理,YOLOv5 仍因为相对计算能力需求较大、内存消耗较多等^[19]问题无法满足本系统实时场景识别的需求。因此本文提出了一种基于分组快速空间金字塔池化的轻量化 GFSP-YOLO 模型,该网络分别由输入端、主干网络、颈部网络和头部网络 4 个模块组成,如图 2 所示。

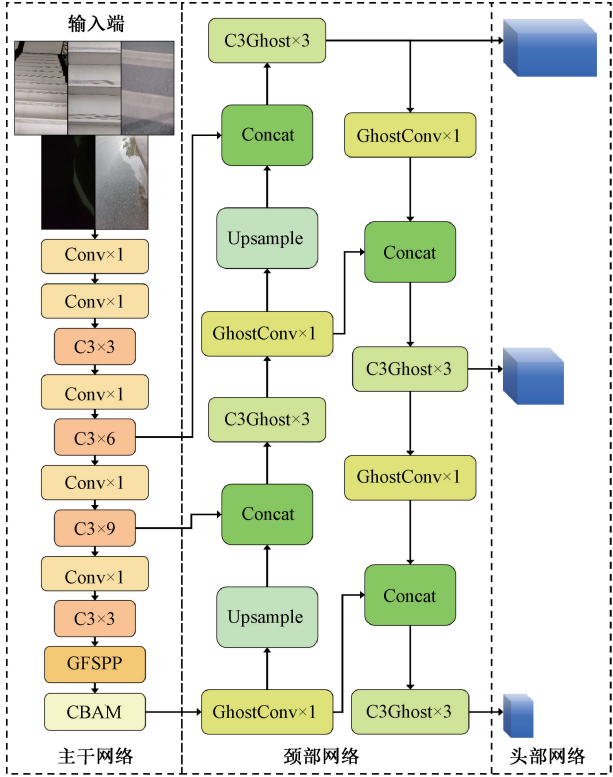


图 2 轻量化 GFSP-YOLO 网络结构

在主干网络中,将原始 YOLOv5 中的 SPPF^[20]替换为 GFSP 和 CBAM 注意力模块^[21]的网络结构。新的主干网络结构主要由 5 个 Conv 模块、4 个 C3 模块、1 个 GFSP 结构和 CBAM 模块组成,该模块共输出大、中、小 3 个尺度的图像特征;该结构在空间金字塔池化结构的基础上结合 CSPDarknet53 的思想^[22]以及分组卷积的思想^[23]。首先,将同一尺度的特征图经过 3 个尺度为 5×5 的最大池化层,

且经过每一个池化层都将输出 1 个分支,然后将池化后的结果在通道维度上与原始特征图进行拼接,在保持感受野不变的情况下获得了速度的提升,有利于网络提取目标细节的特征信息,并且引入分组卷积的思想将 GFSP 结构中的所有普通卷积块都替换为分组卷积 (group convolution, GConv),如图 3 所示。

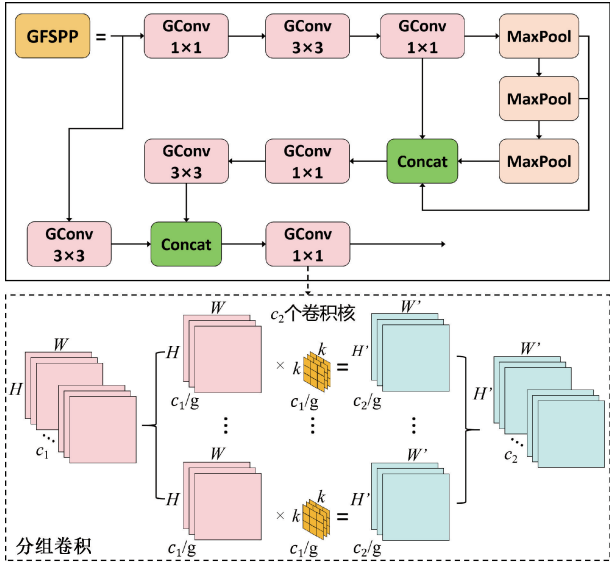


图 3 GFSP 结构

在输入矩阵经过普通卷积和分组卷积后,输入矩阵参数量 P 计算公式分别如式(1)和(2)所示。

$$P_{\text{conv}} = H \times W \times c_1 \times c_2 \quad (1)$$

$$P_{\text{groupconv}} = H \times W \times c_1 \times c_2 \times \frac{1}{g} \quad (2)$$

式中: H, W 和 c_1 分别代表输入矩阵的长、宽和通道数, c_2 代表卷积核的数量, g 代表分组数量, P_{conv} 表示经过普通卷积参数量, $P_{\text{groupconv}}$ 表示经过分组卷积参数量。

由式(2)可知,经过指定组数为 g 的分组卷积后,与普通卷积相比,原来普通卷积后的参数量是分组卷积的 g 倍,分组卷积使得计算复杂度以及输入矩阵的参数量进一步减小。同时,将注意力模块 CBAM 串联在 GFSP 结构后作为主干网络的最后一层,CBAM 由通道注意力模块、

空间注意力模块两部分组成,将通道注意力和空间注意力处理的特征图相加用于提高网络检测性能的精度。

在颈部网络中,采用 Slim Neck 的范式,结合 GhostConv 轻量化卷积模块,将其嵌入到颈部网络中,其网络结构主要由 4 个 GhostConv 模块、4 个 C3Ghost 模块、4 个通道拼接模块和两个上采样模块组成。如图 4 所示, GhostConv 模块由 1 个 CBS 模块和 1 个 DWConv 模块组成,将原特征图先经过一个 CBS 模块后分为两个分支,其一支进行 DWConv 操作,另一支不进行任何操作,然后将两个分支进行通道拼接作为 GhostConv 模块的输出, GhostConv 卷积模块可以最大程度地提取并融合信息,减少冗余的特征,确保颈部网络轻量化的同时保证信息的完整性^[24]; C3Ghost 模块主要由 3 个 CBS 模块和 1 个 GhostBottleneck 模块^[25]组成,其结构与 C3 模块相同,只是替换其中的 GhostBottleneck 模块,而 GhostBottleneck 模块是由两个串联的 GhostConv 模块和一个并行的 DWConv 模块进行元素相加操作并输出。

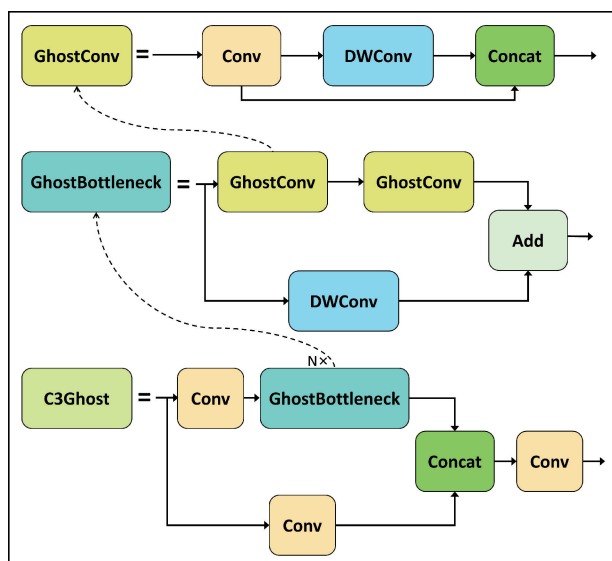


图4 GhostConv、GhostBottleneck、C3Ghost 结构

1.2 基于重力加速度模糊控制规则的手机用户行为识别

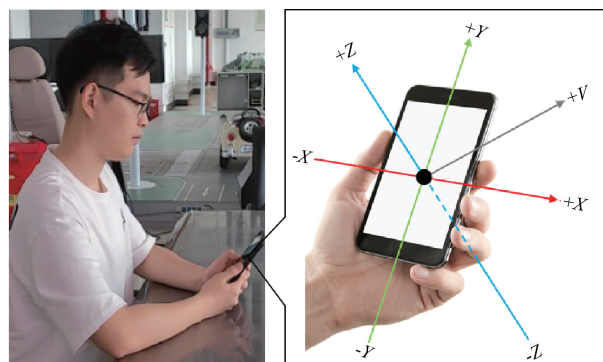
微电子机械系统(micro electro mechanical systems, MEMS)重力加速度计利用微纳技术制造微小的机械结构和电子器件,将传统的机械加速度计集成到微小的芯片中来感应芯片加速度量。其中, MEMS 重力加速度计包括微小的质量块和微型弹簧系统,当质量块受到加速度的作用而发生微小的位移时,这个位移通过微电子传感器转换成电信号,最终得到手机端感应的重力加速度数据结果^[26]。手机 MEMS 重力加速度计能够测量手机在三维空间中的加速度,其三维空间坐标系如图 5(a)所示。其中, X 轴表示手机的水平方向坐标轴,其方向与手机屏幕的宽度方向保持一致; Y 轴表示手机的垂直方向坐标轴,其方向与屏幕

的高度方向保持一致; Z 轴表示手机的垂直方向坐标轴,其方向与 X-Y 平面垂直,与屏幕表面保持垂直。当手机平放在水平表面上时, X 轴正方向指向手机的右侧, Y 轴正方向指向手机的顶部, Z 轴正方向指向手机的屏幕外部。为了更有效地描述手机用户的行为和运动状态,本文在原有 X、Y、Z 三维坐标系的加速度分量基础上,引入了三轴“矢量和”加速度(V),并将其作为第 4 个坐标轴 V 轴的方向。其单位为 m/s^2 , 如式(3)所示。

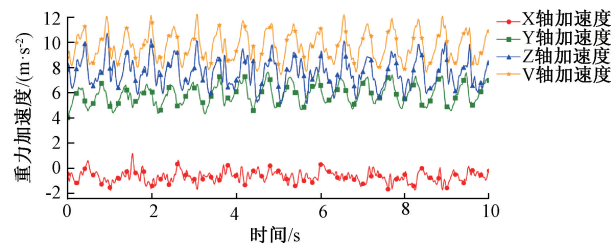
$$a_v = \sqrt{a_x^2 + a_y^2 + a_z^2} \quad (3)$$

式中: V 轴中的加速度 a_v 大于 0, 其他 3 轴中的加速度 a_x , a_y , a_z 均有负值出现。

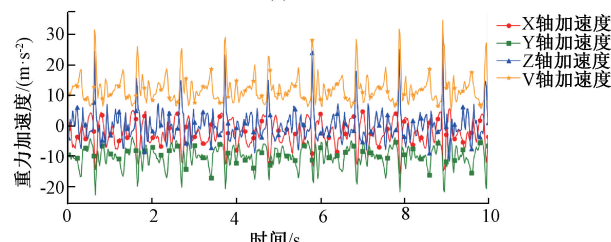
图 5 给出了本文利用手机 3 轴重力加速度计采集的 10 s 低头族 5 种常见的运动状态(行为)时序数据曲线, 包括: 走路看手机, 如图 5(b)所示; 手机揣兜走路, 如图 5(c)所示; 手持手机走路, 如图 5(d)所示; 静止看手机, 如图 5(e)所示; 上下楼梯看手机, 如图 5(f)所示。通过分析图 5 所示的 5 种低头族运动状态的重力加速度曲线可知, 不同运动状态的加速度数据在幅值、频率、相关性上具有较为明显的差异, 反映了低头族在不同活动状态下的运动特征和行为模式的变化。



(a) 重力加速度的四轴图



(b) 走路看手机



(c) 手机揣兜走路

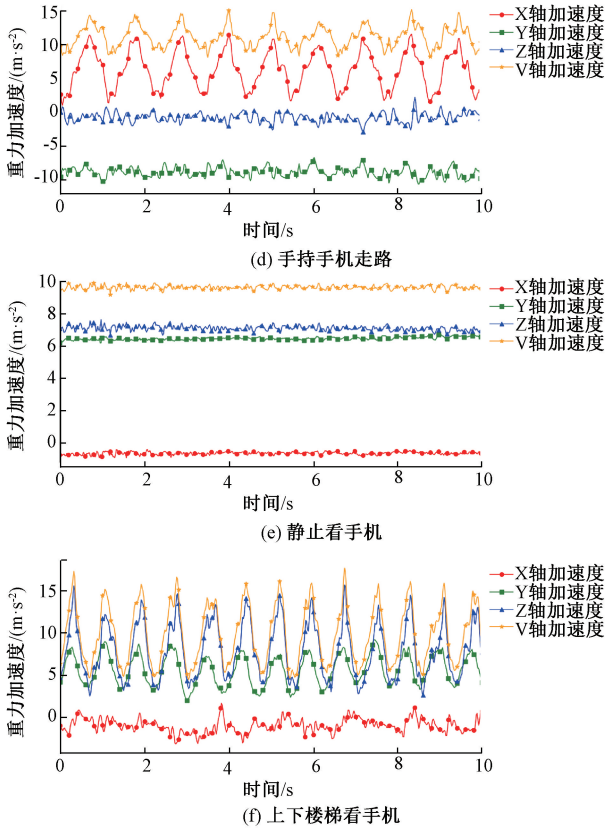


图 5 坐标轴设定与不同行为的时序数据实例

模糊控制是一种基于模糊经验知识对被控对象进行控制的方法,它不依赖于精确的数学模型;其能够迅速适应系统动态变化和外部环境变化^[27]。因此,本文将手机用户行为作为被控对象,利用模糊控制原理对用户行为识别展开模糊控制器设计,提出了基于多统计学特征的低头族行为模糊判定算法表,如表 1 所示。具体步骤如下:

首先,分别对 10 s 内采集到的 500 个重力加速度 3 轴分量值以及合加速度值进行总体均值 μ 、总体方差 σ^2 、总体变异系数 c_v 以及总体相关系数 $\rho_{A,B}$ 计算,如式(4)~(7)所示。

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i \quad (4)$$

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 \quad (5)$$

$$c_v = \frac{\sigma}{\mu} \quad (6)$$

$$\rho_{A,B} = \frac{E[(A - \mu_A)(B - \mu_B)]}{\sigma_A \sigma_B} \quad (7)$$

式中: N 取 500, x_i 表示重力加速度在第 i 点采集到的加速度值, σ 表示 10 s 内 500 个重力加速度值的标准差, c_v 用来表现数据离散程度的绝对值, $E[\cdot]$ 表示期望, $\rho_{A,B}$ 表示重力加速度 3 轴分量和合加速度轴分量中的任意两轴 A 轴加速度和 B 轴加速度之间的相关系数。

表 1 多统计学特征的低头族行为模糊判定算法表

算法 1: 基于多统计学特征的低头族行为模糊判定模型

输入: 将 X, Y, Z, V 四轴加速度时间序列值 $a_X(t), a_Y(t), a_Z(t), a_V(t)$ 模糊化处理; N 个连续四轴加速度时间序列值的均值 $\mu_X, \mu_Y, \mu_Z, \mu_V$; 初始化: $N \leftarrow 500$ 。

输出: 将低头族行为向量 $\hat{r} = (r_1, r_2, r_3, r_4, r_5)$ 解模糊化处理, \hat{r} 下标从 1~5 依次为静止看手机、手机揣兜走路、持手机走路、走路看手机、上下楼梯看手机。

步骤 1) 计算 V 轴加速度总体方差阈值, 以模糊阈值 $\sigma_V^2 = 0.3$ 为界输出函数值, 其式为:

$$f_1(\sigma_V^2) = \begin{cases} r_1, & \sigma_V^2 \leq 0.3 \\ \text{步骤 2),} & \sigma_V^2 > 0.3 \end{cases}$$

步骤 2) 在步骤 1) $\sigma_V^2 > 0.3$ 的条件下计算 Y 轴加速度均值阈值, 以模糊阈值 $\mu_Y = 0$ 为界输出函数值, 其式为:

$$f_2(\mu_Y) = \begin{cases} \text{步骤 3),} & \mu_Y < 0 \\ \text{步骤 4),} & \mu_Y \geq 0 \end{cases}$$

步骤 3) 在步骤 2) $\mu_Y < 0$ 的条件下计算 X 轴和 V 轴加速度的相关系数阈值, 以模糊阈值 $\rho_{X,V} = 0.73$ 为界输出函数值, 其式为:

$$f_3(\rho_{X,V}) = \begin{cases} r_2, & \rho_{X,V} < 0.73 \\ r_3, & \rho_{X,V} \geq 0.73 \end{cases}$$

步骤 4) 在步骤 2) $\mu_Y \geq 0$ 的条件下, 计算 V 轴加速度总体变异系数阈值, 以模糊阈值 $c_v(V) = 0.75$ 为界输出函数值, 其式为:

$$f_4[c_v(V)] = \begin{cases} r_4, & c_v(V) < 0.75 \\ r_5, & c_v(V) \geq 0.75 \end{cases}$$

Return

然后, 上述统计学特征值进行模糊化处理, 并根据多次实验结果确定相应特征值的模糊阈值; 建立如图 6 所示

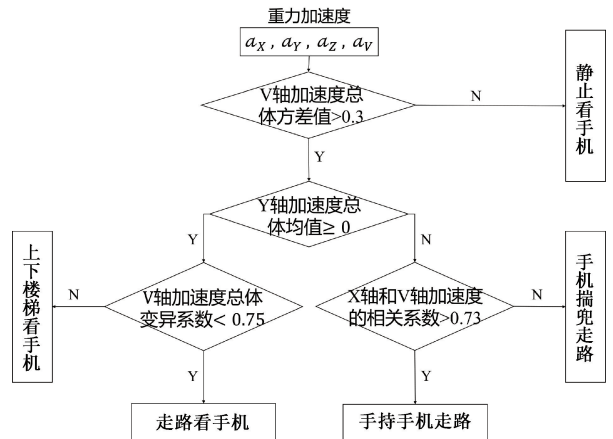


图 6 面向“低头族”行为判定的模糊控制规则

的模糊控制规则,并依据模糊控制规则建立基于多统计学特征的低头族行为模糊判定模型依次实现低头族的行为判定;最后,通过解模糊化输出低头族行为中5种低头族常见的行为,包括:静止看手机、手机揣兜走路、持手机走路、走路看手机、上下楼梯看手机。

1.3 基于多模态决策树的低头族危险状态识别

决策树(decision tree, DT)是一种用于分类、聚类和预测的建模方法,通过“分而治之”的方式将问题的搜索空间分成子集,根据节点引出的弧表示可能的问题答案,叶节点则代表问题解决方案的预测输出^[28]。关键在于选择恰当的分类属性规则,以优化决策树的结构并提高准确性。随着训练样本和属性空间的扩大,优化算法以有效处理大规模数据成为研究焦点,尤其在面对多构化不规则数据和跨学科研究的挑战^[29]。

结合1.1节用户周边环境(场景)识别模块和1.2节用户运动状态(行为)识别模块的识别结果,本文设计了面向手机端安卓系统的“行为-环境”多模态低头族危险状态识别决策树模型,如图7所示。首先,将基于GFSP-YOLO

模型的用户周边环境检测结果作为周边环境特征向量;将基于多统计学特征阈值的用户行为模糊判定模型的用户行为检测结果作为行为特征向量;基于两种不同属性的特征向量,构建多模态低头族危险状态判定决策树模型。具体方法如下:首先,设定“行为-环境”多模态低头族危险状态识别决策树模型的输入,将5种用户周边环境特征向量(低照明环境、斑马线路面、楼梯区域、水坑路面、正常路面)和5种用户行为向量(上下楼梯看手机、走路看手机、静止看手机、手机揣兜走路、手持手机走路)作为决策树模型的输入信号,输入端输入特征维度为10维;然后,根据特征属性作为判别条件,设计3层决策节点集合(共5个节点);最后,输出端输出识别结果类别为11类,分别包括4种安全状态和7种低头族危险状态。其中,4种安全状态包括手机揣兜走路、手持手机走路、正常路面下走路看手机、正常路面下静止看手机;7种危险状态分别包括:上下楼梯看手机、斑马线走路看手机、湿滑路面走路看手机、低照明环境走路看手机、斑马线区域静止看手机、湿滑路面区域静止看手机、低照明环境静止看手机。

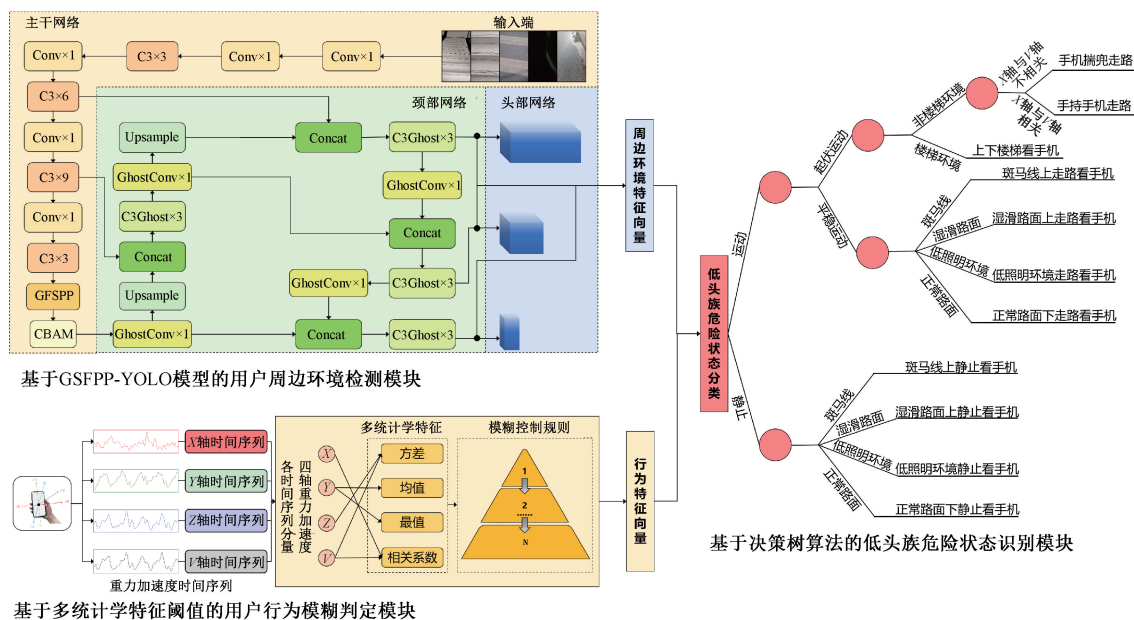


图7 行为-环境多模态低头族危险判定决策树模型

1.4 基于视-听-触觉立体预警与用户界面

为简洁直观地向用户提供有效的预警信息,本系统按照周边环境(行为)与运动状态(行为)两大检测方向,将用户界面分为上、下两个部分。上端显示后视摄像头捕捉到的实时周边路面场景情况,同时显示周边环境的识别结果、画面和语音震动预警信息。下端显示用户的实时运动状态检测变化情况,如“上下楼梯看手机”、“走路看手机”、“静止看手机”、“手机揣兜走路”、“持手机走路”5种不同运动状态(行为)。图8给出了一组在线识别实例:当系统利用手机重力加速度计检测的时间序列数据识别出用户处于危险行为时(包括上下楼梯看手机、走

路看手机、静止看手机),自动开启手机后视摄像头并利用后视摄像头图像实时识别用户当前脚下的环境;如识别结果为楼梯区域、斑马线区域、湿滑路面区域、低照明区域时,则根据利用多模态决策树算法判定手机用户处于低头状态,同时周边环境为危险环境。此时,当用户手机声音开启时,手机利用语音播报提醒用户处于低头族危险行为如“上下楼梯,请不要看手机,注意安全”,当用户手机声音关闭时,手机利用振动传感器对用户进行预警提醒。同时,手机画面切换为后视摄像头实时弹窗画面并锁定屏幕,利用听觉-视觉-触觉立体预警对手机用户给予警告。

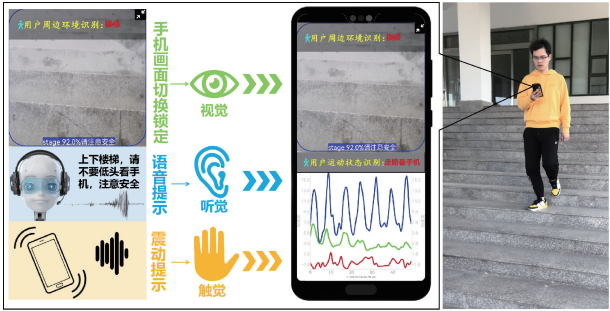


图 8 视-听-触觉立体预警与用户界面

2 实验结果与分析

2.1 实验数据集及实验环境

本文自制环境识别数据集,通过 26 名志愿者利用手机的后置摄像头采集脚下路面环境图像数据集以 7 : 2 : 1 的比例划分为训练集 6 174 张、验证集 1 764 张和测试集 882 张。环境识别数据集包括 5 种场景类别,如图 9(a)所示,包括“楼梯区域”1 924 张、“斑马线区域”1 913 张、“路面积水区域”1 895 张、“低照明环境”1 541 张和“正常路面”1 547 张,样本实例如图 9(b)所示。

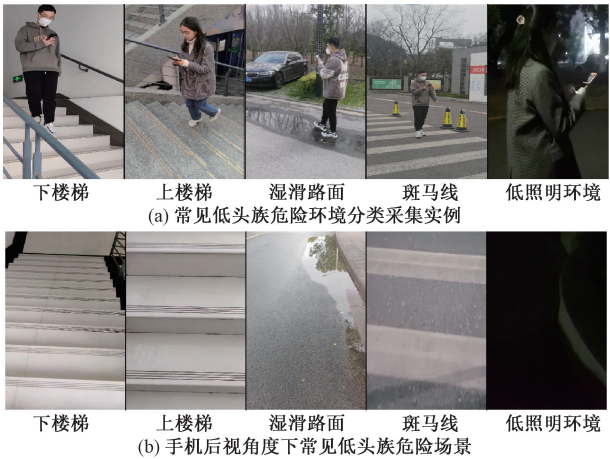


图 9 环境识别数据示例

基于重力加速度计的用户行为数据,在采集环境识别数据的同时,利用手机端的重力加速度计手机 X 轴、Y 轴、Z 轴 3 个方向的原始数据时间序列。自制手机用户重力加速度数据集(行为数据集)来源于 26 名志愿者的 3 862 份不同品牌的智能手机端的加速度传感器数值的时间序列数据,6 种行为的重力加速度时间序列数据采集实例分别为:上、下楼梯看手机、手机揣兜走路、手持手机走路、走路看手机、静止看手机,如图 10 所示。26 人共采集 3 862 份数据集,每一份数据集均包含 500 个数据点,重力加速度频率为 50 Hz,每份时长均为 10 s。其中,持手机走路 785 份、手机揣兜里 704 份、上下楼梯看手机 871 份、静止看手机 628 份、走路看手机 874 份,不同个体采集实验数据量的分布如图 11 所示。

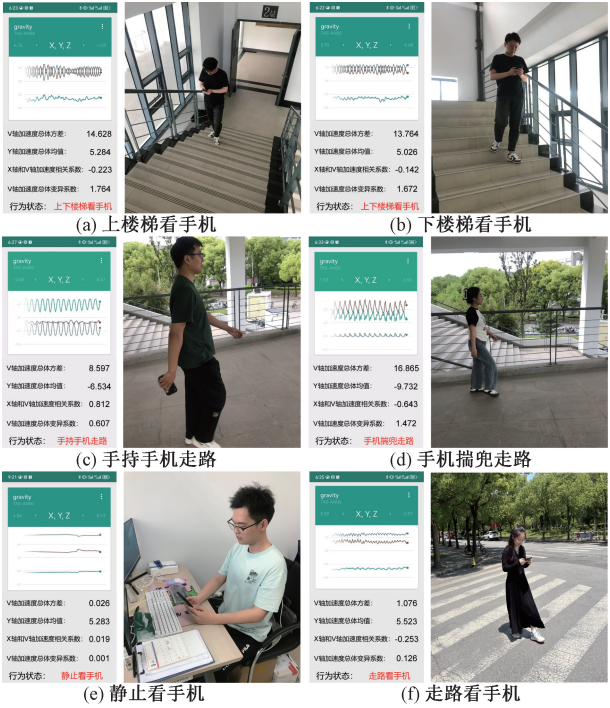


图 10 手机用户运动状态(行为)数据样本示例

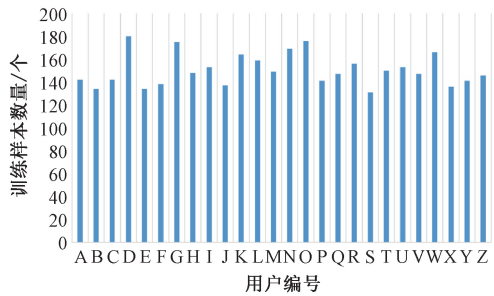


图 11 不同志愿者运动状态样本数量分布图

本文训练、测试模型的实验环境包括离线模型训练环境和在线模型测试环境两个部分。其中,离线模型训练采用 Windows10 服务器进行实验,训练生成 PT 模型。然后,将 PT 模型转成 TF 模型通过 Android Studio 部署到对应移动手机终端上,实现手机端的实时用户周边环境识别检测,离线训练服务器与在线测试手机端环境的具体实验配置如表 2 所示。

表 2 实验环境

服务器端参数	配置	手机端参数	配置
GPU	GTX 1050	手机型号	HUAWEI
CUDA 版本	11.3	处理器	Kirin 990 5 G
操作系统	Windows10	内存	8 GB
编程语言	Python3.8	操作系统	Android

2.2 模型评价指标

采用平均精度均值(mean average precision, mAP)和

准确率分别作为用户周边环境识别模型和用户运动状态(行为)识别模型的评价指标。假设检验分为 TP 、 TN 、 FP 、 FN 4 种情况, TP 表示设定场景样本被检测为设定场景样本的次数, TN 表示非设定场景样本没有被检测为设定场景样本的次数, FP 表示非设定场景样本被检测为设定场景样本的次数, FN 表示设定场景样本没有被检测为设定场景样本的次数。在 4 种假设检验的基础上获取精确率(precision, P)、召回率(recall, R)、准确率(accuracy, A)、 F_1 值的指标, 其相关定义如式(8)~(11)所示。

$$P = \frac{TP}{TP + FP} \tag{8}$$

$$R = \frac{TP}{TP + FN} \tag{9}$$

$$A = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

$$F_1 = \frac{2 \times P \times R}{P + R} \tag{11}$$

式中: R 表示检测为正例的设定场景样本数量在实际为正例的设定场景样本数量的比例; A 表示检测正确的样本数量在所有样本数量中的比例; F_1 值表示精确率和召回率的调和平均数。

采用混淆矩阵作为行为识别的另一个评价指标, 其矩阵的每一行代表一个真实的类别, 每一列代表的是预测的类别。对应 AP 和 mAP 的评价指标定义如式(12)~(13)所示。

$$AP = \frac{\sum P_i}{N} \tag{12}$$

$$mAP = \frac{\sum AP_i}{K} \tag{13}$$

式中: AP 表示某个类别的平均精度, $\sum P_i$ 表示验证集中该类别所有精确率之和, N 表示含有该类别的图像数量, mAP 表示所有类别的平均精度, K 表示分类数量。

2.3 手机用户周边环境识别结果

为了验证本文提出的轻量化环境识别模型 GFSP- YOLO 的有效性, 统一设定 epochs 为 200, batch-size 为 8, imgsz 为 640, optimizer 为 SGD, workers 为 0, lr 为 0.01。图 12 给出了召回率、精确率、 mAP 值、 F_1 值曲线的可视化指标评价结果, 如图 12(a)~(d) 所示, 随着置信度数值的上升, 精确率逐渐升高, 召回率逐渐降低, F_1 值随着置信度的变化先升高后降低, 呈抛物线状, 当置信度为 0.8 时, F_1 值均在 0.9 以上。如图 12(c) 所示, 所有类别 mAP 值曲线下面积(area under the curve, AUC)接近于 1, AUC 接近于 1 表明 GFSP- YOLO 模型辨别能力较强。表 3 给出训练模型 GFSP- YOLO 在 5 个类别场景下每组 882 张测试图像数据的测试结果, 所有类别平均精度均值 mAP 为 0.994, F_1 值为 0.992, 测试结果表明目标检测具有较高的准确性和平衡性。

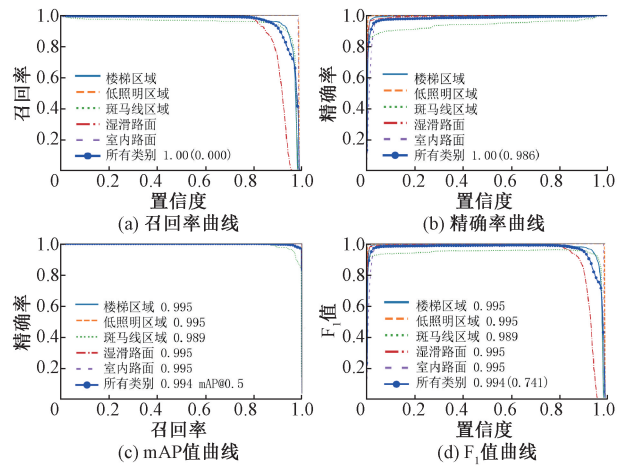


图 12 GFSP- YOLO 模型行为识别性能评价结果

表 3 用户周边环境测试集识别指标

类别	图像数	F_1	P	R	mAP
所有类别	882	0.992	0.99	0.994	0.994
楼梯	284	0.995	0.995	0.995	0.995
湿滑路面	153	0.997	1	0.995	0.995
斑马线	156	0.974	0.96	0.989	0.989
低照明环境	149	0.996	0.997	0.995	0.995
室内环境	140	0.997	1	0.995	0.995

图 13 给出了不同模型的环境识别比较结果实例。通过比较可知, 在相同 5 种类别场景下, YOLOv5s 存在场景错检的情况: 将室内环境错检为湿滑路面, 将楼梯区域错检为斑马线区域, 如图 13(a) 所示; 且相较于本文 GFSP- YOLO 模型, YOLOv5s 的检测置信度显著降低了 30%~50%。这表明 YOLOv5s 模型的泛化能力不足、鲁棒性较差; 而 GFSP- YOLO 模型通过 GFSP 模块在原来的 SPPF 模块基础上将普通卷积替换成了 GConv, 减少了大量参数的运算量, 选择最相关的特征, 减少了模型的冗余,

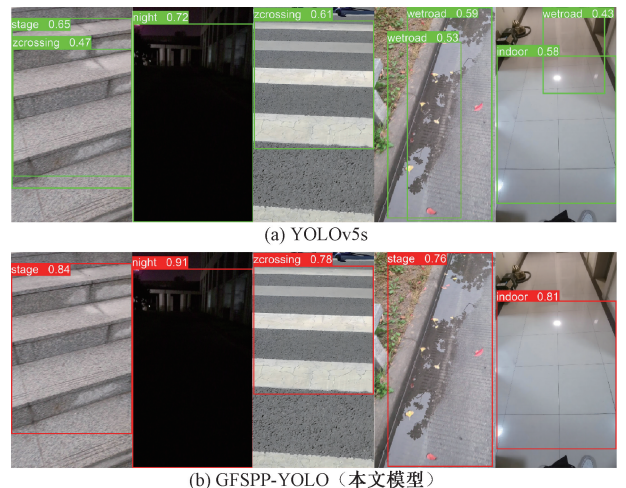


图 13 模型比较结果

提高了其效率和准确性;同时在 GFSP模块后添加了 CBAM 模块关注了最重要的特征区域,进一步提升了模型的检测性能。最后,在颈部网络上通过 GhostConv 模块,在保持模型性能的基础上降低了相关计算量和内存的需求,大幅度提升了模型的检测速率。

表 4 给出了环境识别比较实验,与原始 YOLOv5s 模型相比,GFSP-YOLO 计算量降低了 10%,参数量减少了

3.5%,mAP@0.5 和 mAP@0.5 : 0.95 分别提高了 2.6% 和 6.0%,模型内存减少了 6%并且 FPS 提高了 11.5%。与轻量化模型 YOLOv4n 和 YOLOv7-tiny 相比,GFSP-YOLO 在提高检测精度的同时,也保持了轻量化模型的检测速度,移动端 GFSP-YOLO 部署到移动端上占用的模型内存较低仅为 10.8 Mb,符合移动终端轻量化神经网络的部署,移动端 FPS 为 21 fps。

表 4 用户环境识别比较实验结果

模型	GFLOPS	参数量/ Mb	mAP@0.5	mAP@ 0.5 : 0.95	F1 值	服务器端 模型内存/Mb	移动端 模型内存/Mb	FPS (服务器)	FPS (手机)
YOLOv4-tiny ^[30]	16.2	5.64	0.941	0.829	0.952	22.4	18.9	63	21
YOLOv7-tiny ^[31]	13.0	6.37	0.959	0.871	0.967	13.1	11.2	60	19
YOLOv5s ^[32]	16.0	7.06	0.968	0.867	0.973	13.7	11.6	54	12
本文方法	14.4	6.81	0.993	0.919	0.990	12.8	10.8	65	21

2.4 用户行为识别结果

采用 3 862 份时间序列数据进行基于多统计学特征的低头族行为模糊判定模型测试,结果如表 5 所示,算法平均准确率为 91.27%,处理时间为 0.005 s,能够满足手机端的实时监测性能需求。

表 5 低头族行为模糊判定模型算法性能

准确率	刷新速率	处理时间
91.27%	200 ms	0.005 s

图 14 给出了统计学阈值模糊规则判定模型的准确性混淆矩阵,可以看出静止看手机、持手机走路、手机揣兜行走的运动状态识别准确性较高;而走路看手机和上下楼梯看手机由于 X、Y、Z 3 个方向的时间序列曲线较为相似,存在一定的混淆情况。

2.5 多模态低头族危险状态识别结果

为验证本文提出的多模态低头族危险状态识别方法的有效性,选择了最新 8 种低头族危险检测算法作为比较对象。其中,4 种利用图像信号识别的方法,包括 Tiny-DSOD^[6]、

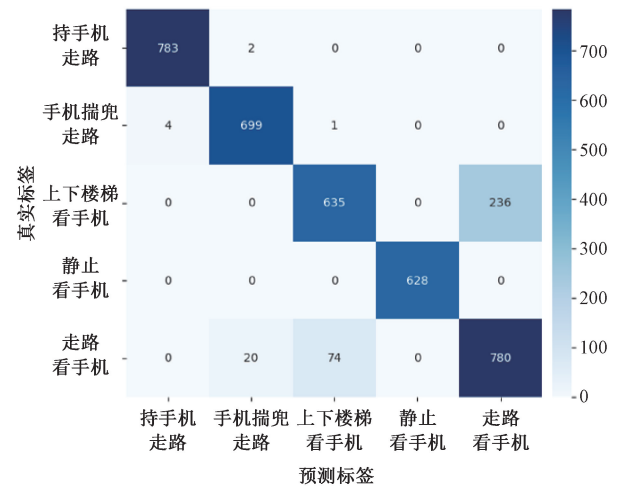


图 14 低头族行为模糊判定模型混淆矩阵

MobileNetV2^[7]、AM-LSTM^[8]和 Yolov3+Vgg16^[9];4 种利用重力加速度信号进行识别的方法:Q-SVM^[15]、DST^[16]、ET^[17]、LSTM^[33]。采用检测准确性、处理时间和模型体量作为评价指标,其性能指比较结果如表 6 所示。

表 6 低头族危险状态识别对比实验

类型	方法	检测模型	准确率/%	处理时间/s	模型大小/Mb
单模态	图像	Tiny-DSOD ^[6]	86.24	0.09	35.2
	图像	MobileNetV2 ^[7]	88.43	0.18	28.5
	图像	AM-LSTM ^[8]	83.28	0.36	96.4
	图像	Yolov3+Vgg16 ^[9]	83.92	0.07	29.3
	加速度	Q-SVM ^[15]	86.90	0.48	48.6
	加速度	DST ^[16]	88.21	0.05	27.1
	加速度	ET ^[17]	78.60	0.04	25.4
	加速度	LSTM ^[33]	88.72	0.64	89.8
多模态	图像+加速度	决策树(本文)	90.63	0.01	17.31

通过表 6 结果可知,在模型体量上,本文采用的决策树模型的树状数据结构避免了大体量的参数。相比于 Q-SVM、LSTM 等大体量参数模型,其模型大小仅为 17.31 Mb,这使得移动端模型占用的内存更小。在运算速度上,决策树模型的计算复杂度较低、处理时间仅为 0.01 s,远低于前文的深度神经网络模型,有效保障了手机端环境运行的实时性。在准确率上,多模态决策树模型能迅速捕捉输入特征之间的非线性关系,在一定程度上降低了小样本数据过拟合的风险,使得模型泛化能力更强,其准确率为 90.63%,在 9 种比较方法中最高。此外,较于传统单模态模型,本文提出的多模态低头族危险识别模型,可分别通过重力加速度计和手机后视摄像头的观测数据完成对手机用户的行为和周边环境的感知与描述;并通过决策树模型的树状结构实现不同属性信息之间互补与协同,有效提高复杂环境下低头族危险状态判别的准确性和鲁棒性。

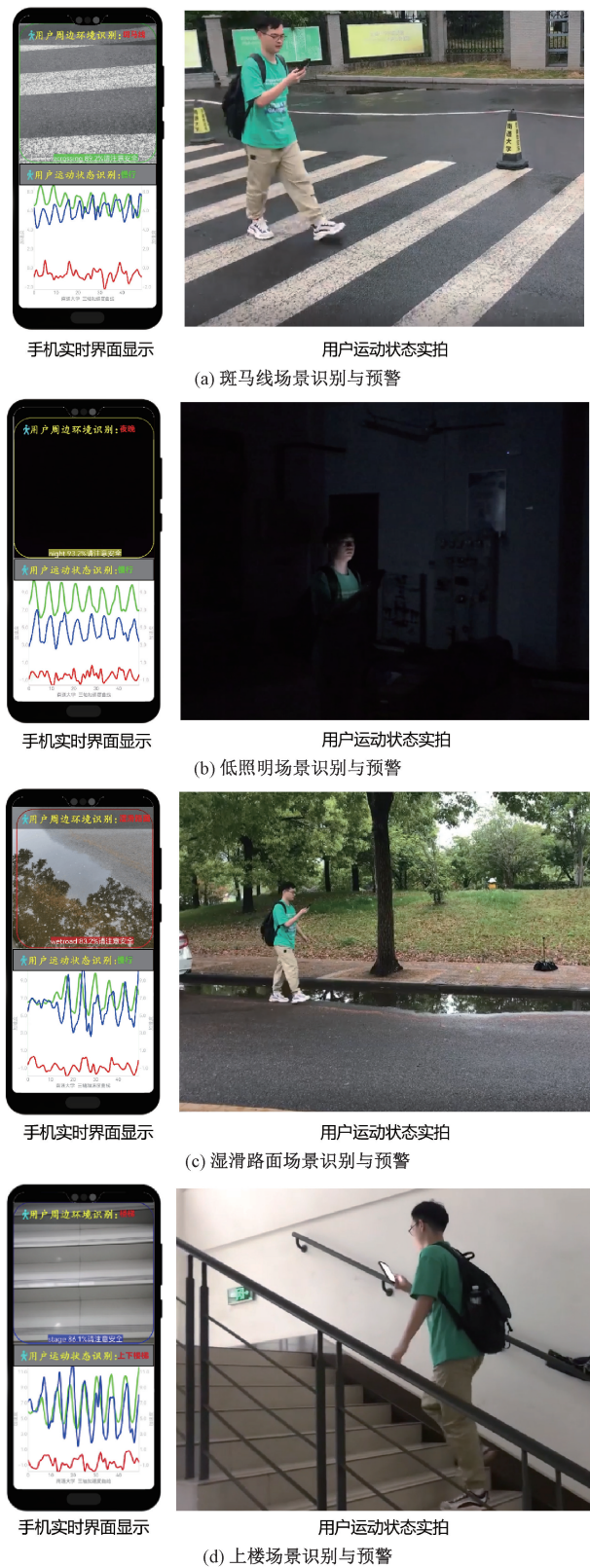
为了全面评估本系统在不同手机环境下的性能,本文使用了 4 种 Android 操作系统的不同品牌和型号的智能手机进行了测试。根据表 7 所示结果,观察到在不同品牌和型号下,低头族危险状态识别系统的模型大小并未呈现显著差异,均稳定维持在约为 17 Mb。处理时间方面,所有手机均表现出极短的响应时间,皆在 20 ms 以内。而系统的整体准确率则保持在 90%左右的水平。尽管不同品牌和型号的手机在该系统性能表现上略有差异,但综合来看,其性能表现依然保持着较为稳定和可靠的特征。这些结果表明,在不同手机环境下,低头族危险状态识别系统呈现出了一致性的模型大小、响应时间和准确率,这意味着本系统具有较好的普适性和稳定性,适用于多种 Android 设备。

表 7 不同类型手机低头族危险状态识别对比实验

手机品牌	手机型号	模型大小/Mb	处理时间/ms	准确率/%
华为	Mate 30	17.31	12	90.63
小米	Mi 10	17.47	15	90.42
三星	Galaxy S20	17.49	15	90.28
OPPO	Reno 5 Pro	17.53	18	90.26

图 15 给出了本文多模态低头族危险感知与预警系统的在线测试结果实例。如图 15 所示,测试场景依次包括过斑马线走路看手机、低照明环境走路看手机、湿滑路面走路看手机、上楼看手机、下楼看手机、持手机行走、手机揣兜行走、端坐静止看手机等 8 个测试场景。其中,前 5 个实例属于为危险状态,因此其对应的手机预警软件界面出现图像、声音、震动、红色文字等多维度预警信号;如图 15(a)所示,斑马线区域走路看手机时,手机在震动和画面预警的同时,出现“斑马线区域请不要低头看手机,注意

四周安全!”的语音提示。而后 3 种测试场景属于相对安全的状态,如图 15(f)~(h)所示,软件无报警信息。



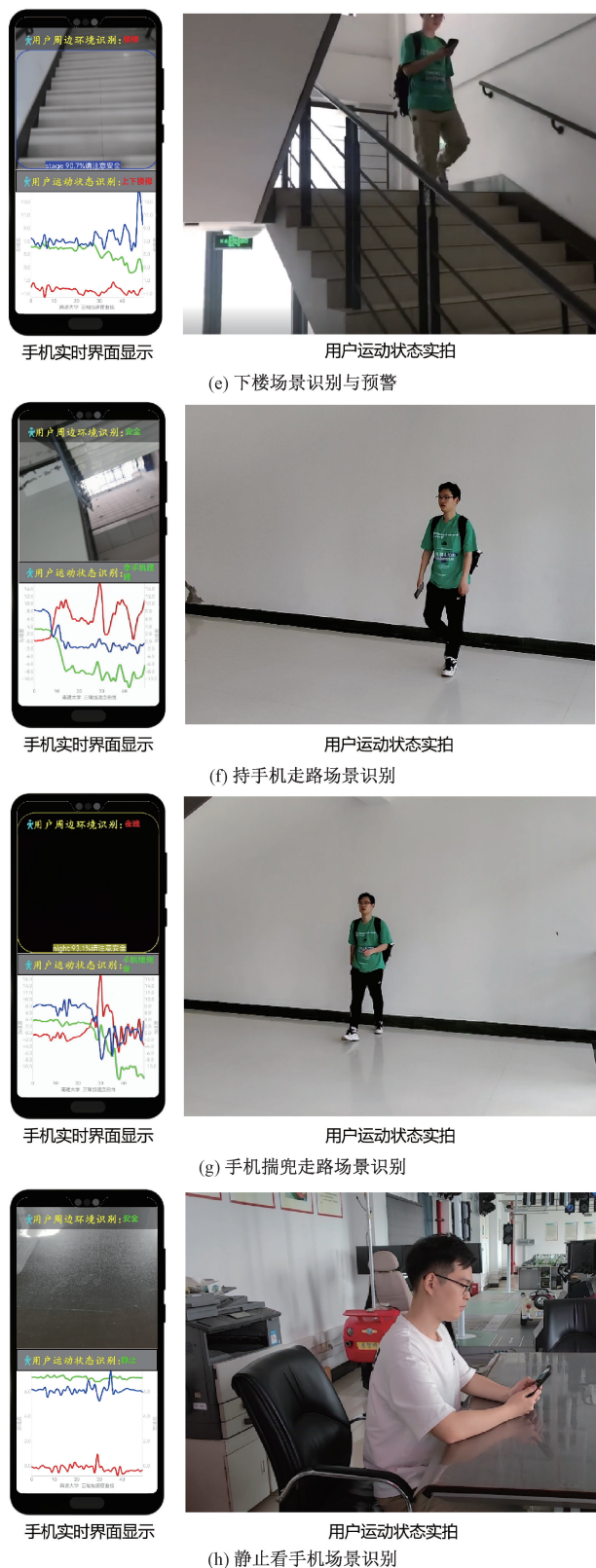


图 15 系统的在线测试结果实例

3 结 论

随着低头族队伍的日渐壮大,因低头看手机导致的跌落、碰撞等道路交通安全事故频发,已经成为影响大众出行安全的重要因素。现有基于手机端的低头族危险检测方法,均采用单一传感器的观测数据,准确性低鲁棒性差。针对上述问题,本文将手机用户的运动状态(行为)与周边环境感知数据相结合,提出基于手机端的多模态低头族危险感知与预警系统;利用手机重力加速度和后视摄像机的观测数据,多角度描述低头族可能出现的危险状态;在手机端构建多模态决策树危险判定模型,实时融合场景识别与行为分类的结果;采用文字、图像、声音、震动多种信号实现“听觉-视觉-触觉”立体式预警;经实验表明,相较于传统方法本文方法准确性高、鲁棒性强、实时性好;可以有效加强大众的交通安全意识、提升低头族的自我约束力;逐渐引导相关群体改正不良生活习惯,树立健康生活理念。

参考文献

- [1] SCHNEIDER F M, HITZFELD S. I ought to put down that phone but I phub nevertheless: Examining the predictors of phubbing behavior[J]. Social Science Computer Review, 2021, 39(6): 1075-1088.
- [2] FRACKOWIAK M, HILPERT P, RUSSELL P S. Partner's perception of phubbing is more relevant than the behavior itself: A daily diary study[J]. Computers in Human Behavior, 2022, 134: 107323.
- [3] BEUKEBOOM C J, POLLMANN M. Partner phubbing: Why using your phone during interactions with your partner can be detrimental for your relationship [J]. Computers in Human Behavior, 2021, 124: 106932.
- [4] 乔婧,陶瑞,孙立山,等. 基于元胞自动机的低头族行人效率模型[J]. 北京工业大学学报, 2023, 49(1): 44-52.
- [5] KIM H S, KIM G H, CHO Y Z. Prevention of smombie accidents using deep learning-based object detection[J]. ICT Express, 2022, 8: 618-625.
- [6] 王文雅. 办案区场景玩手机行为识别的深度学习算法[D]. 西安:西安电子科技大学, 2021.
- [7] 陈浩. 车载危险驾驶行为报警系统的研究与实现[D]. 成都:电子科技大学, 2022.
- [8] 施冬梅,肖锋. 基于改进长短时记忆网络的驾驶行为检测方法研究[J]. 汽车工程, 2021, 43(8): 1203-1209, 1262.
- [9] 管文华,林春雨,杨尚蓉,等. 基于人体关节点的低头异常行人检测[J]. 计算机科学, 2021, 48(5): 163-169.
- [10] BI H L, LIU J J. CSEar: Metalearning for head gesture recognition using earphones in Internet of healthcare things[J]. IEEE Internet of Things Journal, 2022,(9): 23176-23187.

- [11] SCHWEBEL D C, HASAN R, GRIFFIN R, et al. Reducing distracted pedestrian behavior using Bluetooth beacon technology: A crossover trial[J]. *Accident Analysis & Prevention*, 2021, 159: 106253.
- [12] 贾森, 杨钟亮, 陈育苗. 面向地铁低头族的颈部疲劳 sEMG-JASA 评价模型[J]. *智能系统学报*, 2020, 15(4): 705-713.
- [13] 郭毅博, 孟文化, 范一鸣, 等. 基于可穿戴传感器数据的人体行为识别数据特征提取方法[J]. *计算机辅助设计与图形学学报*, 2021, 33(8): 1246-1253.
- [14] 蔡木生. 基于智能手机传感器的人体行为识别技术研究与实践[J]. *单片机与嵌入式系统应用*, 2022, 22(3): 56-60, 65.
- [15] 丁阁文, 丁绪星, 许蓉, 等. 基于智能手机传感器的人类行为识别研究[J]. *无线电通信技术*, 2023, 49(3): 566-576.
- [16] ZHUANG Y, FANG Z X. Smartphone zombie context awareness at crossroads: A multi-source information fusion approach[J]. *IEEE Access*, 2020, 8: 101963-101977.
- [17] ZHUO S J, SHERLOCK L, DOBBIE G, et al. Real-time smartphone activity classification using inertial sensors—recognition of scrolling, typing, and watching videos while sitting or walking[J]. *Sensors*, 2020, 20(3): 655.
- [18] 朱美强, 汤力凡, 王鸿璞, 等. 机械臂倒液任务中透明容器的液位高度视觉检测与控制[J]. *仪器仪表学报*, 2023, 44(10): 283-293.
- [19] WANG C Y, BOCHKOVSKIY A, LIAO H Y. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR)*, 2023: 7464-7475.
- [20] HUANG Z, WANG J, FU X, et al. DC-SPP-YOLO: Dense connection and spatial pyramid pooling based YOLO for object detection[J]. *Information Sciences*, 2020, 522: 241-258.
- [21] 刘会家, 肖懂, 滕杰, 等. 基于 CBAM-ResNet 和多域特征融合的配电网故障选线方法[J]. *国外电子测量技术*, 2023, 42(8): 10-18.
- [22] HSU W Y, LIU W Y. Ratio-and-scale-aware YOLO for pedestrian detection[J]. *IEEE Transactions on Image Processing*, 2020, 30: 934-947.
- [23] PARK S, SHIN Y G. Generative convolution layer for image generation[J]. *Neural Networks*, 2022, 152: 370-379.
- [24] ZHANG N, NEX F, VOSELMAN G, et al. Lite-Mono: A lightweight CNN and transformer architecture for self-supervised monocular depth estimation [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR)*, 2023: 18537-18546.
- [25] WANG B W, LI L Z, NAKASHIMA Y, et al. Learning bottleneck concepts in image classification [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR)*, 2023: 10962-10971.
- [26] 杨金显, 蔡纪鹏, 尹凤帅, 等. 随钻测量钻具重力加速度提取方法[J]. *电子测量与仪器学报*, 2023, 37(10): 145-152.
- [27] PAN Y, WU Y, LAM H K. Security-based fuzzy control for nonlinear networked control systems with DoS attacks via a resilient event-triggered scheme[J]. *IEEE Transactions on Fuzzy Systems*, 2022, 30(10): 4359-4368.
- [28] CARREIRA-PERPINAN M Á, GABIDOLLA M, ZHARMAGAMBETOV A. Towards better decision forests: forest alternating optimization [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR)*, 2023: 7589-7598.
- [29] SAGI O, ROKACH L. Approximating XGBoost with an interpretable decision tree [J]. *Information Sciences*, 2021, 572: 522-542.
- [30] TANG Y, ZHOU H, WANG H, et al. Fruit detection and positioning technology for a Camellia oleifera C. Abel orchard based on improved YOLOv4-tiny model and binocular stereo vision[J]. *Expert Systems with Applications*, 2023, 211: 118573.
- [31] ZHU J, QIN C, CHOI D. YOLO-SDLWD: YOLOv7-based small target detection network for infrared images in complex backgrounds[J]. *Digital Communications and Networks*, 2023. DOI: 10.1016/j.dcan.2023.11.001.
- [32] DONG X, YAN S, DUAN C. A lightweight vehicles detection network model based on YOLOv5 [J]. *Engineering Applications of Artificial Intelligence*, 2022, 113: 104914.
- [33] 陈开华. 基于图像分析的手机传感器行为识别研究[D]. 呼和浩特: 内蒙古大学, 2021.

作者简介

金磊, 硕士研究生, 主要研究方向为机器视觉和图像处理技术。

E-mail: 583824968@qq.com

吉翔, 硕士研究生, 主要研究方向为深度学习和图像处理技术。

E-mail: 1016275126@qq.com

邓丽云, 本科, 主要研究方向为目标识别和数据挖掘技术。

E-mail: 2033177359@qq.com

徐少杰, 硕士研究生, 主要研究方向为智能传感器和模式识别。

E-mail: 1371091034@qq.com

王晗(通信作者), 博士, 副教授, 硕士生导师, 主要研究方向为计算机视觉与智能传感器。

E-mail: hanwang@ntu.edu.cn