(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 119669847 A
(43)申请公布日 2025.03.21

(21)申请号 202411719406.5

(22)申请日 2024.11.28

(71)申请人 上海交通大学
地址 200240 上海市闵行区东川路800号

(72)发明人 薛广涛 陈柱西 赵鼎华 周峻韬 丁典 陈奕超

(74)专利代理机构 上海科盛知识产权代理有限公司 31225
专利代理师 胡紫薇

(51)Int.Cl.
G06F 18/241(2023.01)
G10L 21/0208(2013.01)
G06F 18/2131(2023.01)
G06V 10/44(2022.01)
G06V 10/764(2022.01)
G06V 10/82(2022.01)
G06N 3/045(2023.01)
G06N 3/0464(2023.01)
G06F 3/01(2006.01)
G06F 123/02(2023.01)

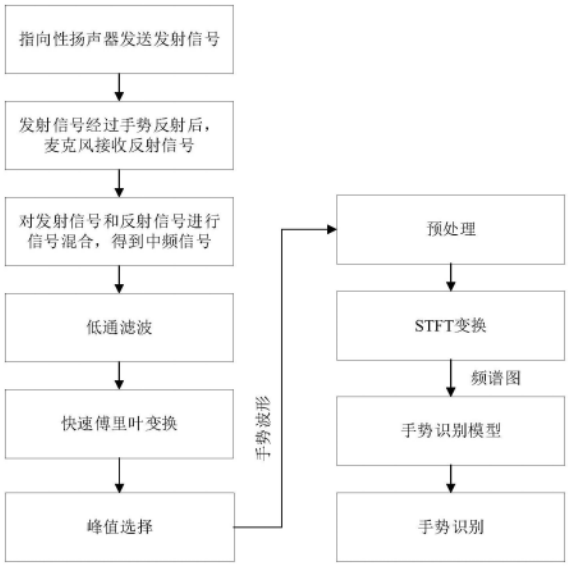
权利要求书2页 说明书5页 附图4页

(54)发明名称

一种基于指向性扬声器的车内手势识别方法及系统

(57)摘要

本发明涉及一种基于指向性扬声器的车内手势识别方法及系统,其中方法包括以下步骤:利用指向性扬声器采集手势数据,并提取包含手势的音频片段,得到手势波形;对手势波形进行预处理;对预处理后的手势波形进行STFT变换,转换成频谱图,并对频谱图进行均匀裁剪后作为手势识别模型的输入;利用基于ResNet网络构建的手势识别模型进行手势识别。与现有技术相比,本发明具有识别准确率高且识别速度快等优点。



CN 119669847 A

1.一种基于指向性扬声器的车内手势识别方法,其特征在于,包括以下步骤:

利用指向性扬声器采集手势数据,并提取包含手势的音频片段,得到手势波形;

对手势波形进行预处理;

对预处理后的手势波形进行STFT变换,转换成频谱图,并对频谱图进行均匀裁剪后作为手势识别模型的输入;

利用基于ResNet网络构建的手势识别模型进行手势识别。

2.根据权利要求1所述的一种基于指向性扬声器的车内手势识别方法,其特征在于,所述提取包含手势的音频片段的方法为:

将指向性扬声器发出的发射信号和所述发射信号经过手势反射后由麦克风接收的反射信号进行信号混频;

对混频后的信号依次进行低通滤波和快速傅里叶变换处理,得到频谱数据;

对频谱数据进行峰值检测,当检测到第三高峰且所述第三高峰的峰值超过预设阈值时,记录一个时间窗口,所述时间窗口以检测到第三高峰的时刻为时间起点,以所述第三高峰下降到预设阈值以下的时刻为时间终点,将所述时间窗口所对应的混频信号作为手势波形。

3.根据权利要求1所述的一种基于指向性扬声器的车内手势识别方法,其特征在于,所述预处理具体为:考虑到电流中存在的噪声,对手势音频片段使用谱减法进行降噪处理。

4.根据权利要求1所述的一种基于指向性扬声器的车内手势识别方法,其特征在于,将均匀裁剪后的频谱图输入到手势识别模型中,频谱图被预处理为一个3×224×224的输入,然后通过卷积层、通道注意力层与空间注意力层和最大池化层进行处理,得到维数为64×64×56的第一特征图;

将第一特征图输入到ResNet18的残差模块中,经过包括卷积、正则化、恒等映射在内的残差运算,得到一个512×7×7的第二特征图;

所述第二特征图通过通道注意、空间注意和平均池化层进一步处理,得到512张1×1的第三特征图;

将第三特征图作为全连接层的输入,经过三层全连接层处理得到手势分类结果并输出。

5.根据权利要求4所述的一种基于指向性扬声器的车内手势识别方法,其特征在于,所述卷积层和池化层的核大小为3×3,填充为1,步幅为2。

6.一种基于指向性扬声器的车内手势识别系统,其特征在于,包括:

指向性扬声器,用于向车内需进行手势识别的位置发送发射信号,所述发射信号为频率范围在70KHz-80KHz之间、啁啾为0.1s的调频连续波信号;

麦克风,用于接收发射信号经过手势反射后的反射信号;

信号处理模块,用于基于发射信号和反射信号提取包含手势的音频片段,得到手势波形,对手势波形进行预处理,并对预处理后的手势波形进行STFT变换,转换成频谱图;

手势识别模块,所述手势识别模块以均匀裁剪后的频谱图作为输入,利用基于ResNet网络构建的手势识别模型进行手势识别。

7.根据权利要求6所述的一种基于指向性扬声器的车内手势识别系统,其特征在于,所述提取包含手势的音频片段的方法为:

将指向性扬声器发出的发射信号和所述发射信号经过手势反射后由麦克风接收的反射信号进行信号混频;

对混频后的信号依次进行低通滤波和快速傅里叶变换处理,得到频谱数据;

对频谱数据进行峰值检测,当检测到第三高峰且所述第三高峰的峰值超过预设阈值时,记录一个时间窗口,所述时间窗口以检测到第三高峰的时刻为时间起点,以所述第三高峰下降到预设阈值以下的时刻为时间终点,将所述时间窗口所对应的混频信号作为手势波形。

8.根据权利要求6所述的一种基于指向性扬声器的车内手势识别系统,其特征在于,所述预处理具体为:考虑到电流中存在的噪声,对手势音频片段使用谱减法进行降噪处理。

9.根据权利要求6所述的一种基于指向性扬声器的车内手势识别系统,其特征在于,将均匀裁剪后的频谱图输入到手势识别模型中,频谱图被预处理为一个3×224×224的输入,然后通过卷积层、通道注意力层与空间注意力层和最大池化层进行处理,得到维数为64×64×56的第一特征图;

将第一特征图输入到ResNet18的残差模块中,经过包括卷积、正则化、恒等映射在内的残差运算,得到一个512×7×7的第二特征图;

所述第二特征图通过通道注意、空间注意和平均池化层进一步处理,得到512张1×1的第三特征图;

将第三特征图作为全连接层的输入,经过三层全连接层处理得到手势分类结果并输出。

10.根据权利要求9所述的一种基于指向性扬声器的车内手势识别系统,其特征在于,所述卷积层和池化层的核大小为3×3,填充为1,步幅为2。

一种基于指向性扬声器的车内手势识别方法及系统

技术领域

[0001] 本发明涉及汽车控制技术领域,尤其是涉及一种基于指向性扬声器的车内手势识别方法及系统。

背景技术

[0002] 随着汽车产业尤其是新能源汽车的蓬勃发展,车辆系统越来越智能化,用户与车辆的交互也越来越多,越来越复杂。但现有的交互方式中,使用车机屏幕或是用户的智能手机都会对用户的驾驶行为产生较大影响;使用物理按键虽然熟练后对驾驶影响较小却比较死板,难以更改预设操作。并且随着车机屏幕的越来越大,留给物理按键的空间也越来越有限。同时,声学传感器越发普遍地运用在了驾驶环境中。一些研究也表明,手势交互是消费者所想要的,它可以有效避免交互时分心所造成的事故。

[0003] 目前基于声学的手势识别系统,由于车内环境复杂的干扰,在车载环境使用时具有挑战性。虽然已经有研究使用波束成形来感应手势操作,但现有的通过扬声器阵列的大尺寸波束成形使得其难以在车辆中部署。

[0004] 目前的定向扬声器可以提供一个定向声场,这在车载环境中是可用的。尽管车辆中不同的人通常有不同的收听需求,但定向声波可以根据不同需求进行调整以满足不同用户。然而,当前的产品没有对定向扬声器的传感潜力进行深入挖掘。在利用定向扬声器进行手势识别的过程中所面临的挑战如下:1.驾驶时车辆环境是一个具有不同物体带来的多径反射的复杂环境,难以准确识别接受信号的具体来源。2.收发频率的冲突,为了能够在不影响向用户发送声音的前提下接收声音,必须谨慎设计两种声波的频段,避免串扰。3.用户的手势变化通常快速而微妙,这使得识别精度难以保证。

发明内容

[0005] 本发明的目的是为了提供一种基于指向性扬声器的车内手势识别方法及系统,利用定向扬声器产生的具有高度指向性的声波,通过频率分析,从接收到的频率中过滤出用户手势所产生的频率,然后使用基于ResNet的深度学习对不同频谱图进行分类,从而在车内实现高精度的手势检测。

[0006] 本发明的目的可以通过以下技术方案来实现:

[0007] 一种基于指向性扬声器的车内手势识别方法,包括以下步骤:

[0008] 利用指向性扬声器采集手势数据,并提取包含手势的音频片段,得到手势波形;

[0009] 对手势波形进行预处理;

[0010] 对预处理后的手势波形进行STFT变换,转换成频谱图,并对频谱图进行均匀裁剪后作为手势识别模型的输入;

[0011] 利用基于ResNet网络构建的手势识别模型进行手势识别。

[0012] 所述提取包含手势的音频片段的方法为:

[0013] 将指向性扬声器发出的发射信号和所述发射信号经过手势反射后由麦克风接收

的反射信号进行信号混频;

[0014] 对混频后的信号依次进行低通滤波和快速傅里叶变换处理,得到频谱数据;

[0015] 对频谱数据进行峰值检测,当检测到第三高峰且所述第三高峰的峰值超过预设阈值时,记录一个时间窗口,所述时间窗口以检测到第三高峰的时刻为时间起点,以所述第三高峰下降到预设阈值以下的时刻为时间终点,将所述时间窗口所对应的混频信号作为手势波形。

[0016] 所述预处理具体为:考虑到电流中存在的噪声,对手势音频片段使用谱减法进行降噪处理。

[0017] 将均匀裁剪后的频谱图输入到手势识别模型中,频谱图被预处理为一个3×224×224的输入,然后通过卷积层、通道注意力层与空间注意力层和最大池化层进行处理,得到维数为64×64×56的第一特征图;

[0018] 将第一特征图输入到ResNet18的残差模块中,经过包括卷积、正则化、恒等映射在内的残差运算,得到一个512×7×7的第二特征图;

[0019] 所述第二特征图通过通道注意、空间注意和平均池化层进一步处理,得到512张1×1的第三特征图;

[0020] 将第三特征图作为全连接层的输入,经过三层全连接层处理得到手势分类结果并输出。

[0021] 所述卷积层和池化层的核大小为3×3,填充为1,步幅为2。

[0022] 一种基于指向性扬声器的车内手势识别系统,包括:

[0023] 指向性扬声器,用于向车内需进行手势识别的位置发送发射信号,所述发射信号为频率范围在70KHz-80KHz之间、啁啾为0.1s的调频连续波信号;

[0024] 麦克风,用于接收发射信号经过手势反射后的反射信号;

[0025] 信号处理模块,用于基于发射信号和反射信号提取包含手势的音频片段,得到手势波形,对手势波形进行预处理,并对预处理后的手势波形进行STFT变换,转换成频谱图;

[0026] 手势识别模块,所述手势识别模块以均匀裁剪后的频谱图作为输入,利用基于ResNet网络构建的手势识别模型进行手势识别。

[0027] 所述提取包含手势的音频片段的方法为:

[0028] 将指向性扬声器发出的发射信号和所述发射信号经过手势反射后由麦克风接收的反射信号进行信号混频;

[0029] 对混频后的信号依次进行低通滤波和快速傅里叶变换处理,得到频谱数据;

[0030] 对频谱数据进行峰值检测,当检测到第三高峰且所述第三高峰的峰值超过预设阈值时,记录一个时间窗口,所述时间窗口以检测到第三高峰的时刻为时间起点,以所述第三高峰下降到预设阈值以下的时刻为时间终点,将所述时间窗口所对应的混频信号作为手势波形。

[0031] 所述预处理具体为:考虑到电流中存在的噪声,对手势音频片段使用谱减法进行降噪处理。

[0032] 将均匀裁剪后的频谱图输入到手势识别模型中,频谱图被预处理为一个3×224×224的输入,然后通过卷积层、通道注意力层与空间注意力层和最大池化层进行处理,得到维数为64×64×56的第一特征图;

[0033] 将第一特征图输入到ResNet18的残差模块中,经过包括卷积、正则化、恒等映射在内的残差运算,得到一个512×7×7的第二特征图;

[0034] 所述第二特征图通过通道注意、空间注意和平均池化层进一步处理,得到512张1×1的第三特征图;

[0035] 将第三特征图作为全连接层的输入,经过三层全连接层处理得到手势分类结果并输出。

[0036] 所述卷积层和池化层的核大小为3×3,填充为1,步幅为2。

[0037] 与现有技术相比,本发明具有以下有益效果:

[0038] (1)本发明通过使用指向性扬声器,实现了对驾驶员手势的指向性检测,可以有效的提高驾驶员的驾驶体验和驾驶安全。

[0039] (2)本发明通过将注意力层引入ResNet神经网络,将测量到的音频信号转化为频谱图,通过深度学习的方式实现了手势识别,在保证识别准确率的同时提高了识别速度。

[0040] (3)在不同服装以及不同性别、不同年龄的人群条件下验证了本发明仍然能够保持良好的任务准确性。

附图说明

[0041] 图1为本发明的方法流程图;

[0042] 图2为一种实施例中的手势及其对应的频谱图;

[0043] 图3为一种实施例中手势识别模型的网络结构示意图;

[0044] 图4为一种实施例中手势识别的结果准确性示意图。

具体实施方式

[0045] 下面结合附图和具体实施例对本发明进行详细说明。本实施例以本发明技术方案为前提进行实施,给出了详细的实施方式和具体的操作过程,但本发明的保护范围不限于下述的实施例。

[0046] 本实施例提供一种基于指向性扬声器的车内手势识别方法,如图1所示,包括以下步骤:

[0047] S1,利用指向性扬声器采集手势数据,并提取包含手势的音频片段,得到手势波形。

[0048] 具体的,包括以下步骤:

[0049] S11,指向性扬声器发出的频率范围在70KHz-80KHz之间、啁啾为0.1s的FMCW(Frequency-Modulated Continuous Wave,调频连续波)信号,所述发射信号经过手势反射后由麦克风接收,得到反射信号,将发射信号和反射信号进行信号混频。

[0050] S12,对混频后的信号依次进行低通滤波和快速傅里叶变换处理,得到频谱数据。

[0051] S13,对频谱数据进行峰值检测,频谱中的最高峰和次高峰分别为发射信号与人体产生的反射信号,因此,检测到的第三高峰为手势信号。当检测到第三高峰且所述第三高峰的峰值超过预设阈值S时,记录一个时间窗口t,所述时间窗口以检测到第三高峰的时刻为时间起点,以所述第三高峰下降到预设阈值S以下的时刻为时间终点,将所述时间窗口t所对应的混频信号作为手势波形。

[0052] S2,对手势波形进行预处理:考虑到电流中存在的噪声,对手势音频片段使用谱减法进行降噪处理,优化信号的信噪比。

[0053] S3,对预处理后的手势波形进行STFT变换,转换成频谱图,如图2所示,并对频谱图进行均匀裁剪后作为手势识别模型的输入。

[0054] S4,利用基于ResNet网络构建的手势识别模型进行手势识别。

[0055] 如图3所示,将均匀裁剪后的频谱图输入到手势识别模型中,频谱图被预处理为一个3×224×224的输入,然后通过卷积层、通道注意力层与空间注意力层和最大池化层进行处理,得到维数为64×64×56的第一特征图;

[0056] 将第一特征图输入到ResNet18的残差模块中,经过包括卷积、正则化、恒等映射在内的残差运算,得到一个512×7×7的第二特征图;

[0057] 所述第二特征图通过通道注意、空间注意和平均池化层进一步处理,得到512张1×1的第三特征图;

[0058] 将第三特征图作为全连接层的输入,经过三层全连接层处理得到手势分类结果并输出。

[0059] 本实施例中,卷积层和池化层的核大小为3×3,填充为1,步幅为2。

[0060] 本实施例收集了6000组手势数据,将其分为70%训练集,20%验证集与10%的测试集,对神经网络进行训练。在测试集上,整体手势识别F1得分值为92.3%,平均精度为91.7%,表明本发明在车辆环境下实现手势交互的可行性和适用性,并且测试了不同手套对结果的影响,测试结果显示在最差情况下也有82.5%的识别准确率。

[0061] 除了ResNet网络之外,本实施例以VGGNet,LeNet5等神经网络进行相同的实验验证,但其表现都不如ResNet神经网络。在引入了注意力模块之后,本发明兼具良好的识别速度与准确性。图4展示了本发明方法的结果准确性,其中,颜色越深表示精度越高。

[0062] 本实施例还提供一种基于指向性扬声器的车内手势识别系统,包括:

[0063] 指向性扬声器,用于向车内需进行手势识别的位置发送发射信号,所述发射信号为频率范围在70KHz-80KHz之间、啁啾为0.1s的调频连续波信号;

[0064] 麦克风,用于接收发射信号经过手势反射后的反射信号;

[0065] 信号处理模块,用于基于发射信号和反射信号提取包含手势的音频片段,得到手势波形,对手势波形进行预处理,并对预处理后的手势波形进行STFT变换,转换成频谱图;

[0066] 手势识别模块,所述手势识别模块以均匀裁剪后的频谱图作为输入,利用基于ResNet网络构建的手势识别模型进行手势识别。

[0067] 所属领域的技术人员可以清楚地了解到,为描述的方便和简洁,所述描述的模块的具体工作过程,可以参考前述方法实施例中的对应过程,在此不再赘述。

[0068] 综上,大量实验验证了本发明在车载环境下通过声波提取用户手势的可行性。本发明提出了一种使用高指向性transducer组成的阵列的声学感知系统,与传统声学系统相比,本发明能够在不与可听频段相互干扰的前提下实现指向性的手势检测。此外,本发明还提出了一种手势识别方法,能够从接受到的波形中过滤掉声音主径与人体反射,进而通过基于ResNet的深度学习算法提取出具体的手势信息。由于手势识别需要更快速的图像处理速度,本发明向ResNet中引入了注意力机制。通过大量实验测试得出了该系统的检测准确度为91.7%,证明了本发明在车辆环境下实现手势交互的可行性和适用性。

[0069] 以上详细描述了本发明的较佳具体实施例。应当理解,本领域的普通技术人员无需创造性劳动就可以根据本发明的构思做出诸多修改和变化。因此,凡本技术领域中技术人员依据本发明的构思在现有技术的基础上通过逻辑分析、推理、或者有限的实验可以得到的技术方案,皆应在权利要求书所确定的保护范围内。

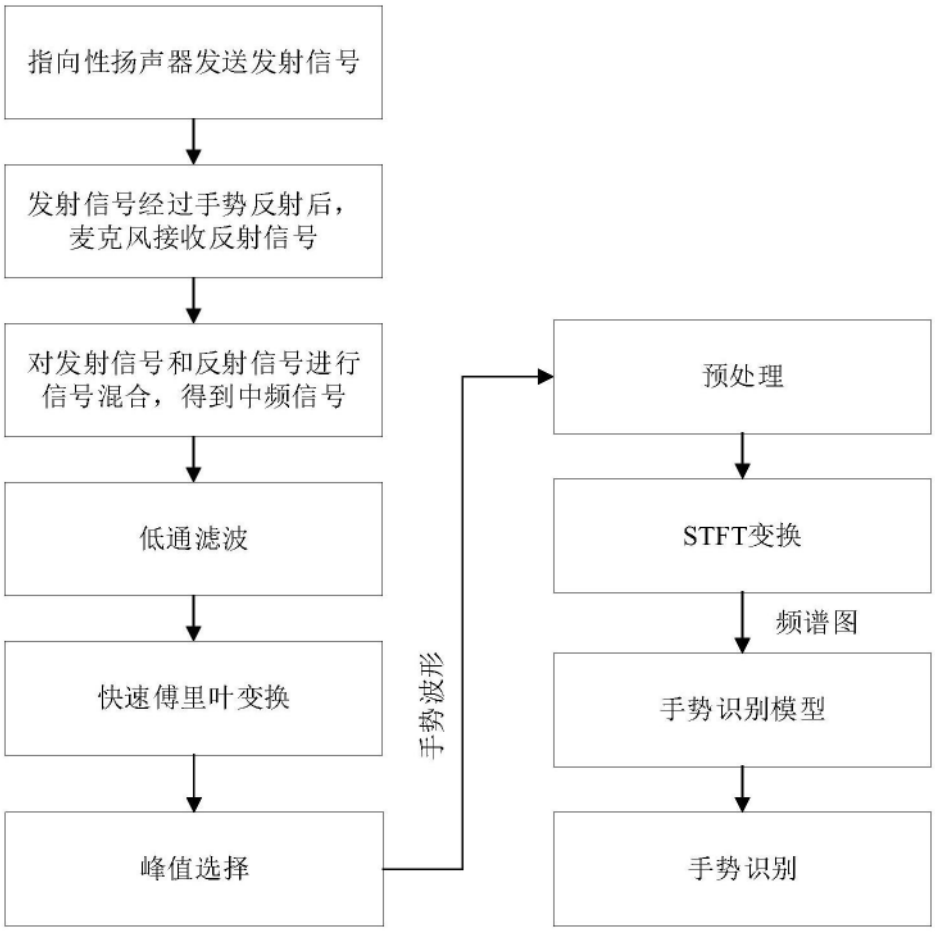

指向性扬声器发送发射信号
发射信号经过手势反射后，麦克风接收反射信号
对发射信号和反射信号进行信号混合，得到中频信号
低通滤波
快速傅里叶变换
峰值选择
手势波形
预处理
STFT变换
频谱图
手势识别模型
手势识别

图1

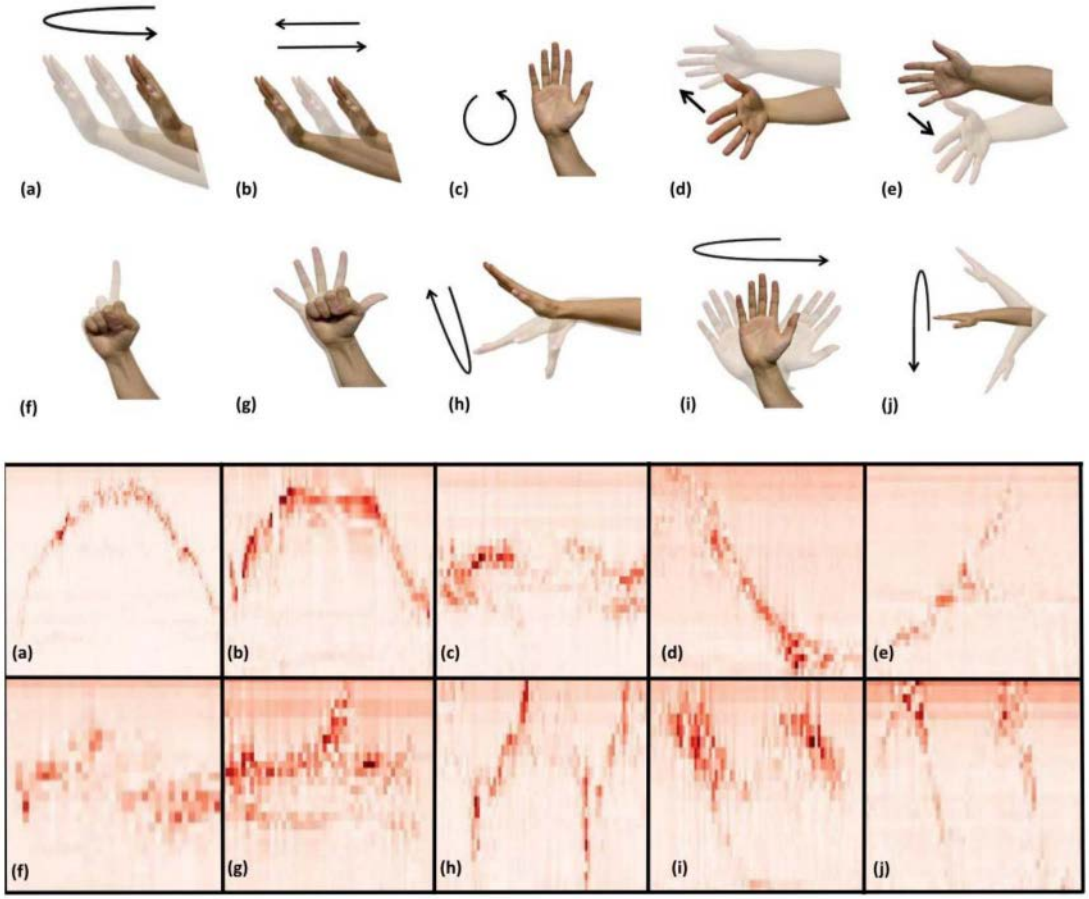


图2

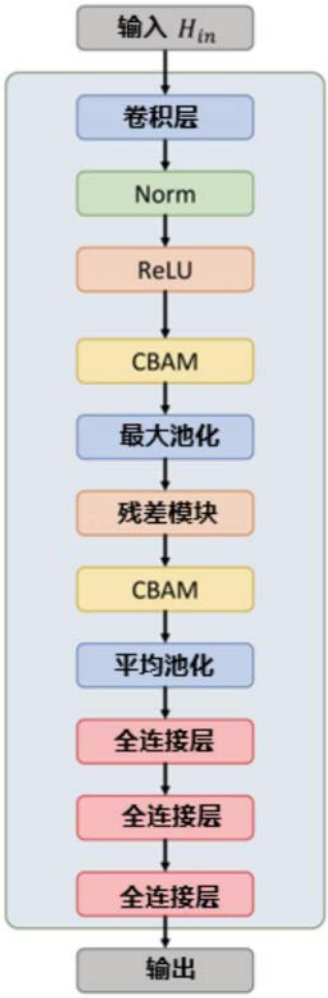

输入 H_{in}
卷积层
Norm
ReLU
CBAM
最大池化
残差模块
CBAM
平均池化
全连接层
全连接层
全连接层
输出

图3

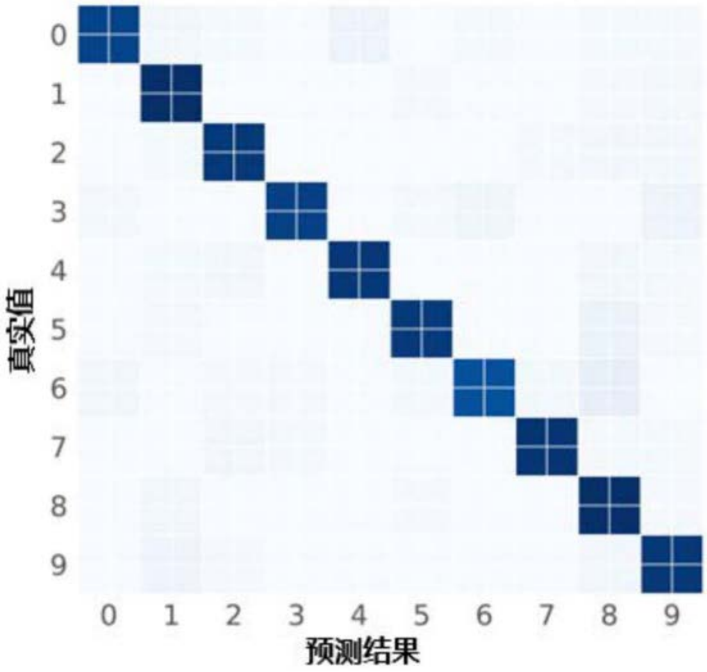


图4