



上海科技大学
ShanghaiTech University

本科毕业论文（设计）

题 目： 多传感器手势识别和人机交互图形控制

学生姓名： 刘承奇

学 号： 2018533249

入学年份： 2018

所在学院： 信息科学与技术学院

攻读专业： 计算机科学与技术

指导教师： 张玉瑶

上海科技大学

2022 年 6 月



上海科技大学
ShanghaiTech University

THESIS

Subject: Multi-sensor Gesture Recognition and Human-computer
Interaction Graphic Control

Student Name: Chengqi Liu

Student ID: 2018533249

Year of Entrance: 2018

School: School of Information Science and Technology

Major: Computer Science and Technology

Advisor: Yuyao Zhang

ShanghaiTech University

Date: 6 / 2022

上海科技大学

毕业论文(设计)学术诚信声明

本人郑重声明：所呈交的毕业论文（设计），是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

作者签名：刘承奇

日期：2022 年 6 月 1 日

上海科技大学

毕业论文（设计）版权使用授权书

本毕业论文（设计）作者同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权上海科技大学可以将本毕业论文（设计）的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本毕业论文（设计）。

保 密 ☐，在____年解密后适用本授权书。

本论文属于

不保密 ☒。

（请在以上方框内打“√”）

作者签名：

刘承奇

指导教师签名：

张玉瑶

日期：2022 年 6 月 1 日

日期：2022 年 6 月 1 日



多传感器手势识别和人机交互图形控制

摘要

人机交互（HCI）是一种研究用户和计算机之间通过各种接口进行交互的计算机技术，其目的是创建更加自然和直接的用户界面，以增加计算机产品的易用性。利用多种传感器进行手势识别的人机交互控制技术是一种具有广泛前景的人机交互技术，它用途广泛且不会对用户造成创伤。模式识别技术是通过传感器检测小范围手势动作，并从中提取大量信息的一种有效方式。与传统的单一视角深度学习算法相比，多视角卷积神经网络能够从数据中提取判别性强的特征，从而提升模型的分类精度。我们在现有研究的基础上，利用OYMotion Technologies公司生产的gForcePro+肌电臂环携带的多种传感器收集被试者上臂的表面肌电信号（sEMG），构建了5个基于不同手势方案的数据集。我们进而设计了一种多视角卷积神经网络模型用于手势识别，从中选取了一种最合适的手势方案，最终实现了一种对用户友好的人机交互图形控制软件。该软件将便于手术医生在操纵医疗手术机器人过程中同时调整医学影像，也可广泛用于其它应用场景。

关键词： 人机交互，模式识别，表面肌电信号，多视角卷积神经网络，手势识别



MULTI-SENSOR GESTURE RECOGNITION AND HUMAN-COMPUTER INTERACTION GRAPHIC CONTROL

ABSTRACT

HCI (human-computer interaction) is a computer technology that studies the interaction between users and computers through various interfaces. Its purpose is to create a more natural and direct user interface to increase the ease of use of computer products. The HCI using multi-sensor gesture recognition is a promising human-computer interaction technology. It is widely used and will not cause trauma to users. Pattern recognition is an effective way to detect small-scale gestures through sensors and extract a large amount of information from them. Compared with the traditional single-view deep learning algorithm, the multi-view CNN (convolutional neural network) can extract highly discriminative features from the data, so as to improve the classification accuracy. On the basis of the existing research, we collected the sEMG (surface electromyographic signals) of the upper arm of the subjects by using a variety of sensors carried by the gForcePro+ armband produced by OYMotion technologies, and constructed five data sets based on different gesture schemes. We then design a multi-view CNN model for gesture recognition, select the most appropriate gesture scheme, and finally implement a user-friendly HCI graphics control software. The software will facilitate surgeons to adjust medical images while manipulating surgery robot, and can also be widely used in other application scenarios.

Key words: HCI, pattern recognition, sEMG, multi-view CNN, gesture recognition



目 录

第一章 绪论	1
1.1 人机交互	1
1.2 模式识别	2
1.2.1 姿态估计	2
1.2.2 手势识别	2
1.3 实验简述	3
1.3.1 场景及需求	3
1.3.2 实验流程	4
1.3.3 实验难点	5
1.3.4 最终效果	5
第二章 材料和方法	7
2.1 肌电信号数据集构建	7
2.2 模型训练	11
2.3 软件搭建	12
2.3.1 实时数据采集部分	12
2.3.2 手势识别部分	12
2.3.3 图形控制部分	13
2.3.4 进程间通信	14
第三章 实验结果	15
第四章 讨论	17
4.1 对实验的探讨	17
4.1.1 实验过程	17
4.1.2 手势方案	17
4.1.3 采集过程	17
4.1.4 滑动窗口大小	18
4.2 后续改进方向	18
4.2.1 添加加速度计数据	18
4.2.2 提升采样率	18
4.2.3 修改进程间通信方式	18
4.2.4 与姿态估计相结合	19
第五章 结论	20
参考文献	21
致谢	23



第一章 绪论

1.1 人机交互

人机交互（HCI）是一种研究用户和计算机之间通过各种接口进行交互的计算机技术，其目的是创建更加自然和直接的用户界面，以增加计算机产品的易用性^[1]（图1-1）。尽管基于WIMP（窗口、图标、菜单和指针）的传统人机交互范式已被广泛采用，但随着计算机设备的小型化和使用功能的多样化，各种新型人机交互方式也不断被提出^[1]。

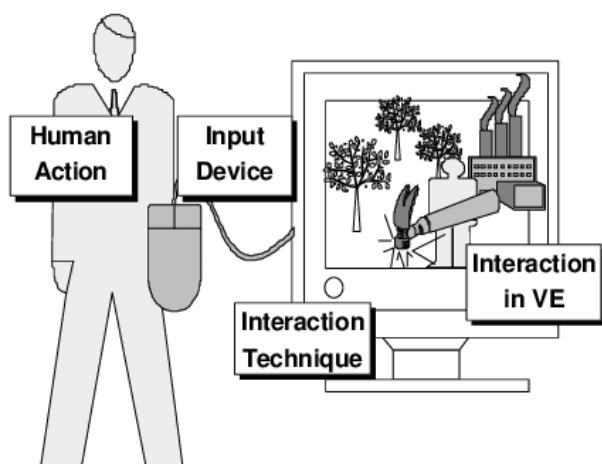


图 1-1 人机交互原理

利用多种传感器进行手势识别的人机交互控制技术是一种具有广泛前景的人机交互技术。该技术可利用多种传感器（如相机、皮肤上的电极或加速度计）收集信号并识别出用户当前的手势，以控制计算机执行对应的功能。这种交互方式易于使用且不会对用户造成创伤，它可以设计成臂环等可以随时穿戴装饰品，也可以设计在手表、服装等日常用品上。它既可以用于难以使用传统人机交互方式（如鼠标、键盘）的截肢者，以帮助他们控制假肢或康复设备（图1-2）；也可以用于未截肢的普通人，以增强用户体验或为传统计算机设备提供一些扩展功能^[2]。

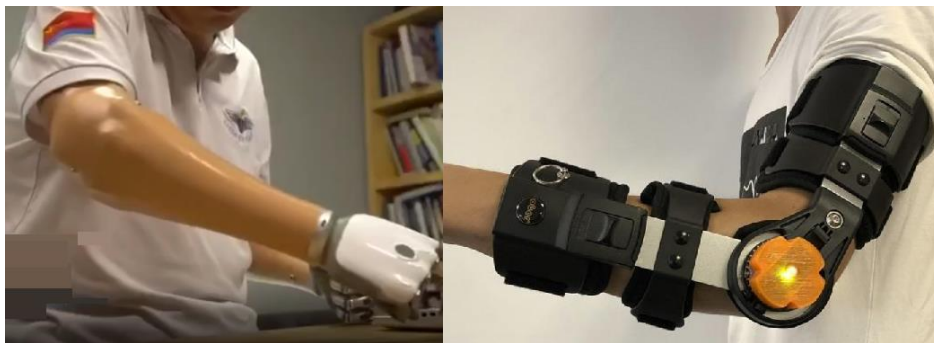


图 1-2 利用多传感器手势识别的假肢和康复设备



1.2 模式识别

模式识别技术被认为是通过传感器检测小范围手势动作，并从中提取大量信息的一种有效方式^[3]。在实现通过手部动作进行人机交互的方面，基于相机、皮肤上的电极和加速度计等不同传感器的选择，目前有手势识别和姿态估计两种主要的工作方向。

1.2.1 姿态估计

姿态估计一般利用相机拍摄用户的手势动作，再利用计算机视觉技术进行模式识别（图 1-3）。该方法被认为具有一定难度，主要的障碍在于肢体在拍摄时相互遮挡^[4]，用户的手部位置也可能因光照、背景等因素的影响而识别失败^[5]，从而影响人机交互效果。目前，姿态估计在基于外观、基于运动和基于骨架等多种识别方法上均有广泛的进展，许多方法的识别精度都可达到 90% 以上^[5]。

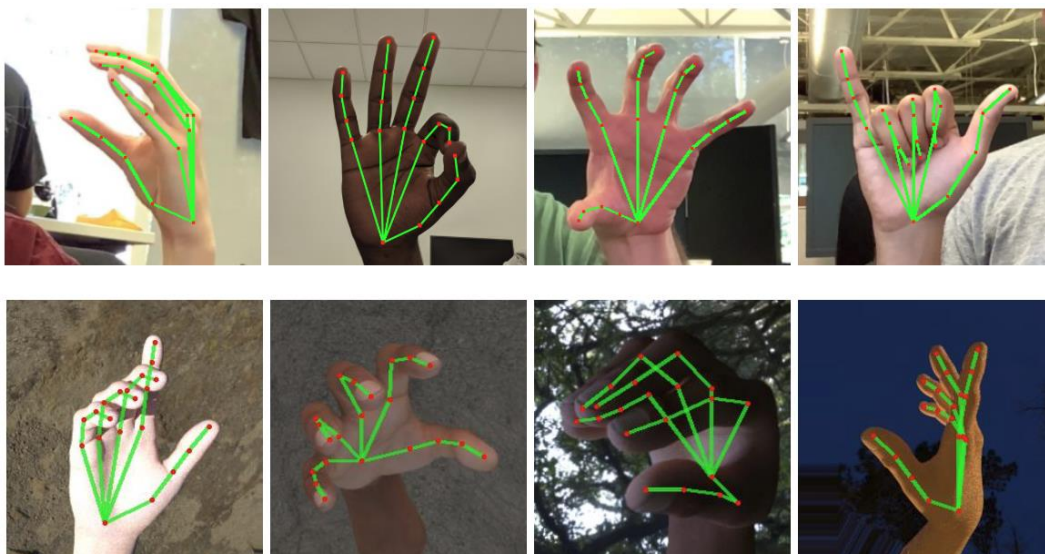


图 1-3 利用计算机视觉技术的手部姿态估计技术

1.2.2 手势识别

手势识别一般采用紧贴皮肤的多组电极收集用户的表面肌电信号（sEMG），再利用机器学习方法进行模式识别（图 1-4）。一些研究也会同时安装加速度计采集信号，将加速度计的数据与表面肌电信号数据同时使用以增强识别精度^[4]。该方法的主要障碍在于不同被试者的肌肉和脂肪分布可能有很大的差异，因此通常需要为每个被试者单独训练模型；此外用户的多块肌肉的表面肌电信号存在互相干扰，可能因此影响识别效果^[4]。目前，基于支持向量机（SVM）、随机森林（RF）、线性判别分析（LDA）和卷积神经网络（CNN）的方法都取得了广泛进展并能达到 90% 以上的识别精度^{[4][6]}。

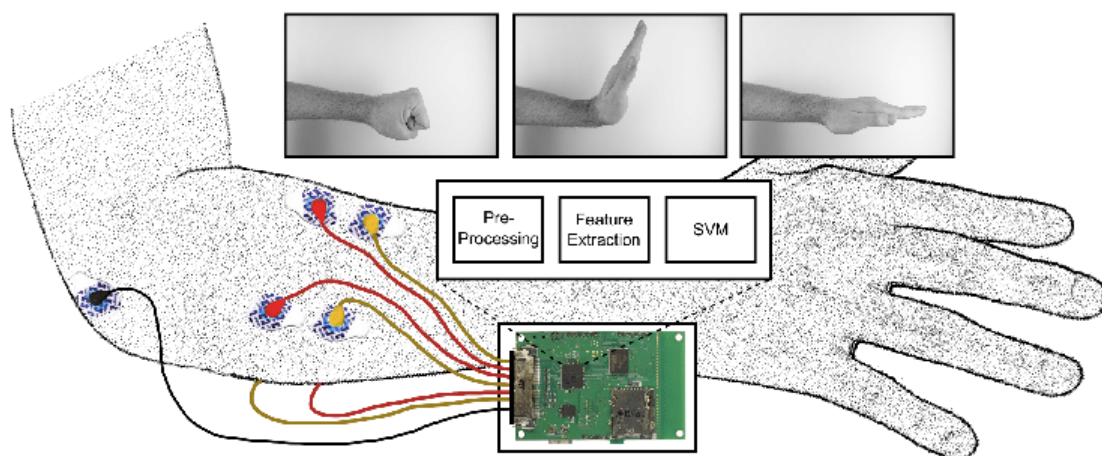


图 1-4 利用表面肌电信号的手势识别技术

1.3 实验简述

1.3.1 场景及需求

医疗手术机器人是一种智能外科手术工具，它可以为患者减小手术时的感染风险、缩短住院时间并减小手术造成的不适，也可以为外科医生增强手术时的精度、灵活性和可视化程度^[7]（图1-5）。医疗手术机器人在国内的发展已处于高速增长期，市场规模也在逐年增大^[8]。医生在执行手术的过程中需随时调整医疗影像所显示的视野范围，以便更精准地定位到执行手术的部位。然而，在医生操纵过程中，由于手部的活动范围受限而无法方便地对医学影像同时进行控制。针对此问题，人机交互图形控制技术可识别医生在操纵过程中额外做出的简单动作，从而辅助医生操纵医疗手术机器人的同时对医疗影像进行调整，因而可能提供一种良好的解决方案。



图 1-5 达芬奇外科手术机器人

我们采访了海军军医大学第一附属医院（上海长海医院）的一位具有多年使用达芬奇外科手术机器人经验的医生，并对该手术机器人进行了实地测量。医生希望能通过与操纵手术机器人类似的简单手部动作直接对医学影像进行控制，然而该医学影像设备的控制位置和手术机器人相距较远，导致手术时存在不便。医生还希望尽量减小或避免对手术机器人进行改装。而目前市场上尚不存在一种现有的产品能够完全满足该需求。我们实地调查发现，该手术机器人的操纵空间较为狭窄，且在狭小的空间中已包含两个操纵杆、多个按



钮和踏板（图1-6）。测量表明，医生的操纵位置和机器人主体支架间仅有约20厘米的水平距离（图1-7）。



图 1-6 达芬奇外科手术机器人的操纵位置



图 1-7 测量结果

在该距离下若采用姿态估计的方式安装相机进行手部动作识别，则最多只能安装两台相机（前部和右部各一台），且加装相机的过程可能需要对手术机器人进行轻微的改装。但若采用手势识别的方法，则可借助无线信号传输的肌电臂环进行人机交互，因此手势识别是我们的首选方案。

1.3.2 实验流程

我们在现有研究的基础上，设计了5种可行的人机交互手势方案，并利用gForcePro+肌电臂环携带的多种传感器收集被试者上臂的表面肌电信号，构建了基于5个不同手势方案的



数据集。我们进而设计了一种多视角卷积神经网络模型。在综合模型识别精度和影响用户体验的多种因素后，我们在5种方案中选出了一最合适的手势方案，并利用该手势方案最终实现了一种人机交互图形控制软件（图1-8）。该软件将便于手术医生在操纵医疗手术机器人过程中同时调整医学影像，也可广泛用于其它应用场景。

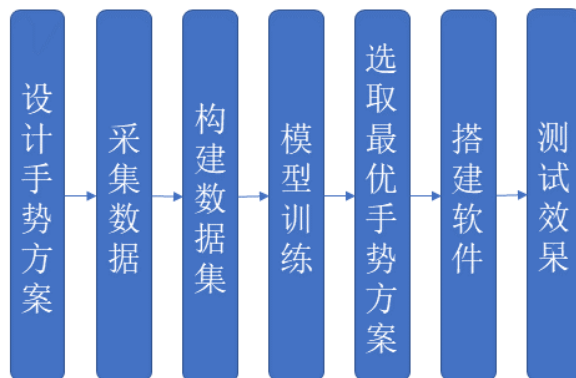


图 1-8 实验流程图

1.3.3 实验难点

实验的难点一方面在于设计并选择出一种实用性强、精度足够高、对用户友好的手势方案，对此我们在设计手势方案时进行了充分的讨论，随后在现有研究的基础上对设计的手势方案进行了训练和筛选。另一方面的难点在于将不同编程语言和框架下实现的功能相互整合，对此我们调研了并尝试了多种方案，最终结合用户需求实现了最合适的一种。此外，实验中涉及大量复杂的参数需要调整，若对每个参数都逐一调整和测试将耗费大量的时间，对此我们首先调研已有的研究成果，提前确定一些参数最优的范围，再在此基础上固定一些参数不变以控制变量，仅对最可能影响实验效果的部分参数进行调整。

1.3.4 最终效果

在我们完成的工作中，在提前收集好用户数据并完成模型训练的情况下，用户可以通过佩戴在上臂的无线肌电臂环，通过做出12个不同的手势动作实时控制远程计算机设备上的3D重建医学影像平移、旋转、缩放和复位。该软件可实时绘制接收到的用户上臂表面肌电信号图像（图1-9）。该肌电臂环易于穿戴且能够发射蓝牙信号，用户只需在计算机设备的USB接口处插入蓝牙接收设备并运行该软件即可操控。经测试，用户可在约12米的距离内进行远程无线操纵，软件识别精度约为90%，延迟约为300毫秒。



图 1-9 软件使用效果

第二章 材料和方法

2.1 肌电信号数据集构建

我们使用 OYMotion Technologies 公司生产的 gForcePro+肌电臂环收集被试者做出手势动作时产生的信号。该设备包含呈轴对称分布的 8 个高灵敏度且带有差分干电极的表面肌电信号 (sEMG) 传感器、1 个 9 轴惯性测量单元 (IMU) 运动传感器, 并可通过蓝牙 BLE 4.2 进行通信 (图 2-1) (各相关参数详见^[9])。该设备的 8 个差分干电极由弹力松紧带链接, 从而可以紧贴被试者的上臂肌肉表面以收集表面肌电信号。这些电极在使用时不必涂抹导电凝胶, 因而便于数据采集和使用。



图 2-1 gForcePro+肌电臂环内部结构

一些研究表明, 采用机器学习的模式识别方法对于电极放置的具体位置不敏感, 只需要保证被试者采集数据和使用臂环时电极处于相同的位置^[10]。因而, 我们未对被试者佩戴肌电臂环的位置做出严格规定。我们只让被试者在采集数据时将肌电臂环佩戴于上臂接近肘关节处肌肉比较明显的位置 (图 2-2), 并在采集过程中保持佩戴位置不动。随后我们使用油性笔在被试者的手臂上做出记号并对佩戴位置进行拍摄, 以保证在被试者后续使用臂环时仍能佩戴在相同的位置。

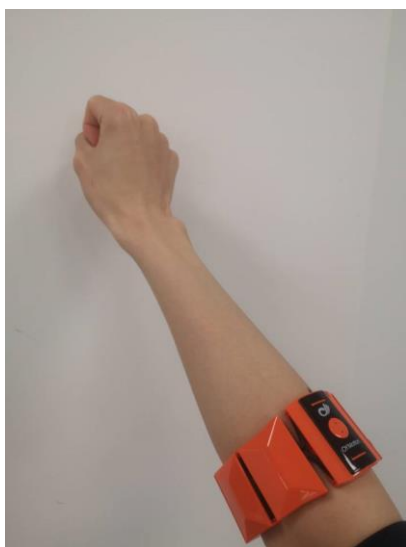


图 2-2 肌电臂环佩戴位置

在数据采集过程中，被试者佩戴好臂环坐在笔记本电脑前，电脑播放一个包含提示信息及手势相关图像的视频，该视频被设定好按照一定的时间节点发出提示。被试者的手臂自然前伸并放在一个高度合适的平台上（图 2-3）。我们在旁边架设一台小型相机拍摄采集数据的全部过程，并将该视频信息用于切分采集到的肌电信号以构建数据集。被试者在参加实验前需同意自愿参加该实验。

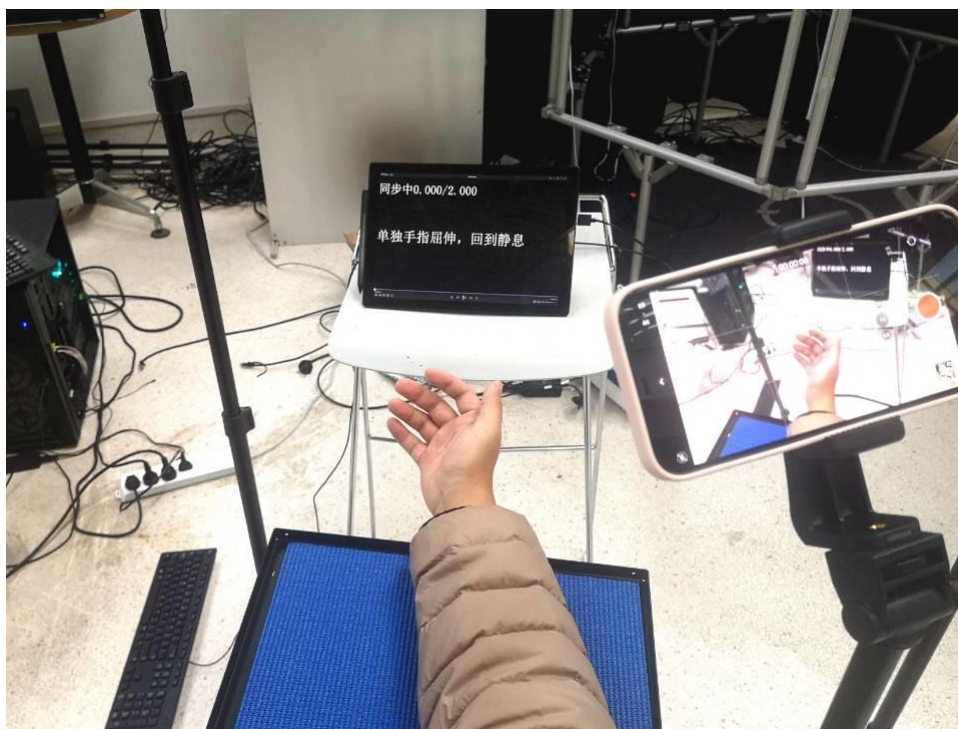


图 2-3 数据采集场景

我们基于一些手势识别中常用的手势（图 2-4）^[11]和该软件所需的 11 种功能设计了 5 种不同的手势方案，如表 2-1 所示。其中手势方案 1 要求被试者做出动态的手势，手势方案 2 到方案 5 要求被试者做出静止的手势并维持肌肉用力。随后我们根据每一种手势方案采集相同被试者在相同位置佩戴臂环并做出手势时的前臂表面肌电信号数据，最终构建出 5 个数据集。采集每种手势方案的数据时所使用的参数如表 2-2 所示。为防止被试者在重复



做手势动作时产生疲劳因而影响数据质量，我们让被试者在两次重复的手势动作之间休息 2 秒，在每组不同的手势动作之间休息 15 秒。该提示信息将在采集数据时通过预先编辑好的视频自动提示给被试者。

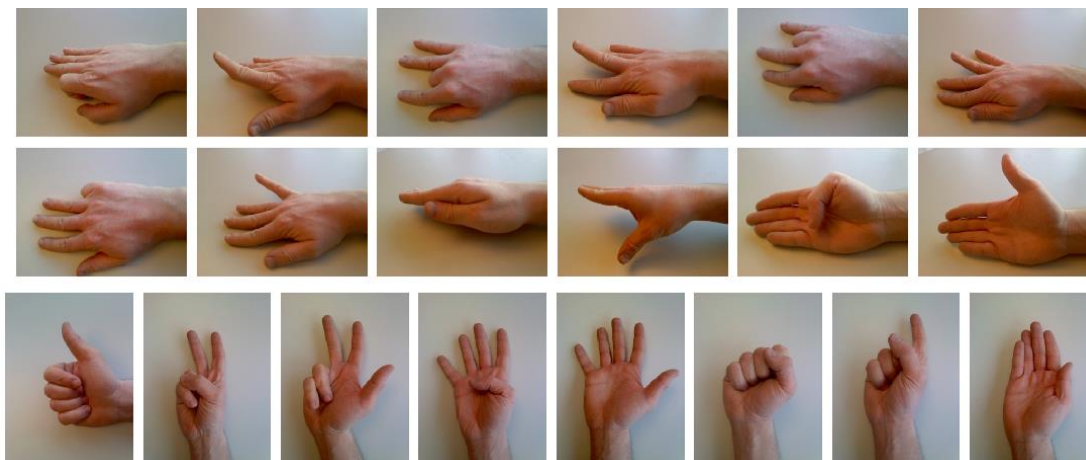


图 2-4 基本手势动作



表2-1 5种手势方案及对应功能

序号	对应功能	手势方案1	手势方案2	手势方案3	手势方案4	手势方案5
1	放大					
2	缩小					
3	向左平移					
4	向右平移					
5	向上平移					
6	向下平移					
7	向左旋转					
8	向右旋转					
9	向上旋转					
10	向下旋转					
11	复位					
12	静止状态					



表 2-2 数据采集相关参数

名称	数值
被试者人数	1
手势数量	15
重复次数	9
数据总量	135
手势持续时间	3s
采样率	500 Hz
数据精度	12-bit
电极数量	8

我们对数据进行了预处理（图 2-5）。表面肌电信号数据首先经过 gForcePro+肌电臂环自带的 20-500 Hz 硬件带通滤波器滤波以消除高频随机噪声干扰^[9]。肌电臂环将采集到的原始数据存放在二进制格式的文件中，我们将其导出并以整数形式表示每个数据点。随后我们将表面肌电信号数据绘制成折线图并生成动态视频，再利用 Adobe Premiere 软件将肌电信号视频和相机录制的视频进行对齐。我们随后根据对齐信息对表面肌电信号数据进行切分，将被试者休息时的数据丢弃，将其余各组手势数据打好标签。随后我们以长度为 200 个数据点（0.4 秒）的滑动窗口对数据进行窗口化，再进行标准化使每个数据点都落在 [-1,1] 的区间内。

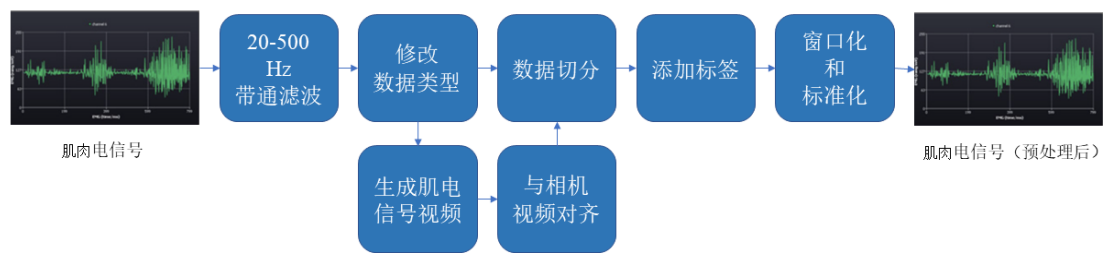


图 2-5 数据预处理流程图

2.2 模型训练

相关研究表明，利用多种传统的机器学习方法在手势识别任务中能得到 90% 以上的识别精度，利用深度学习方法也能得到相似甚至更好的效果^[4]。Wentao 等人提出了一种基于多视角特征提取的深度学习方法，并在多个手势识别任务中取得了比传统端到端神经网络更好的效果^[12]。基于相似的原理，我们设计了结构如图 2-6、图 2-7 所示的神经网络。该神经网络模型利用 4 种不同尺寸的卷积核进行卷积，以从表面肌电信号数据中提取 4 种不同尺度的特征，再将这些特征传入一个由全连接层和线性整流函数（ReLU）构成的简单网络中进行分类。整个过程可以被看作是基于 4 个不同的视角对数据提取特征再进行分类。多视角卷积神经网络能够从数据中提取判别性强的特征，从而提升模型的分类精度，而这些特征是容易被传统的单一视角深度学习方法所忽略的^[12]。

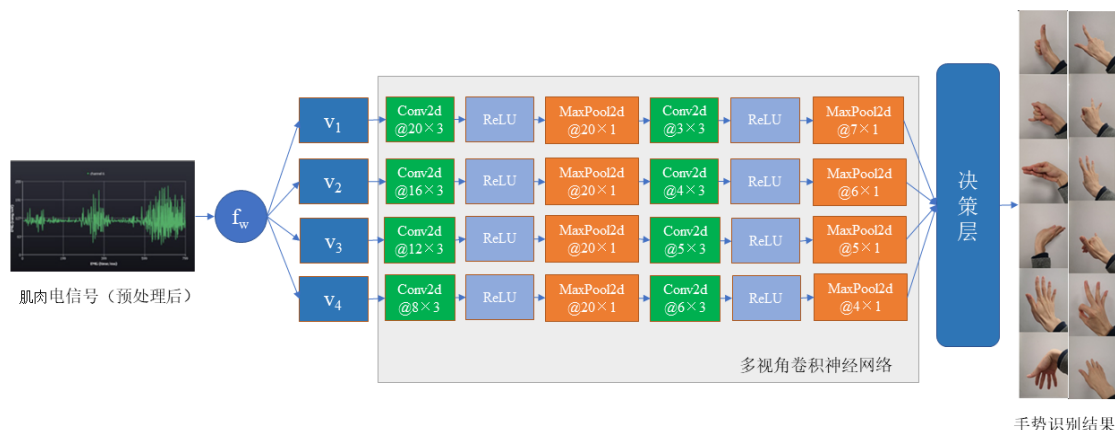


图 2-6 神经网络总结构



图 2-7 决策层结构

Wentao 等人在对比使用不同视角数量情况下的分类精度时，发现分类精度大致随视角数量的增加而提升，但在视角数量超过 3 之后，精度进一步提升的程度非常小^[12]。我们在权衡识别精度和模型的复杂度后，将视角的数量设定为 4。

我们对每个数据集都进行 200 个 epoch 的训练，并使用 9 折交叉验证的方式验证模型的分类准确性。我们将训练过程中神经网络在验证集上取得最优识别精度时各节点的权重进行存储，用于后续搭建软件中的手势识别部分。

2.3 软件搭建

我们将该利用多传感器手势识别技术的人机交互图形控制软件的架构分为实时数据采集、实时分类和图形控制三部分，将三部分分别封装为进程，再实现进程间的相互通信。

2.3.1 实时数据采集部分

OYMotion Technologies 公司生产的 gForcePro+肌电臂环提供了开源的开发软件 oym8CHWave，该软件基于 Qt5.15.2 和 MSVC2019-64bit 开发，实现了在肌电臂环设备和计算机之间通过蓝牙 BLE 4.2 进行的实时数据传输功能^[9]。肌电臂环实时采集的表面肌电信号数据通过蓝牙传输至计算机，并以二进制格式的文件存储，传输延迟估计在百毫秒级别。

我们对开源软件 oym8CHWave 进行了适当修改，使原有的批处理式文件存储方式改为实时的分段存储方式。软件每采集 256 个数据点进行一次数据存储，从而方便实时对数据进行处理。

2.3.2 手势识别部分

我们基于 Python 3.9、Pytorch 1.10 和 CUDA 11.3 搭建了一个基于训练获得的神经网络模型进行手势识别的软件。该部分软件实时读取二进制数据文件，并采用与构建数据集时完全一致的数据预处理方式，再将预处理后的数据传入神经网络中预测手势结果。每 256 个数据点经过长度为 200 的窗口处理为 56 组表面肌电信号数据，每组数据将产生一个预测手势结果，最终获得 56 个预测结果。该部分软件随后选取 56 个结果的众数，并将该手势信号转化为键盘按键信号，传入下一个图形控制部分的进程。我们采用选取众数的判断方式是为了减小预测误差以获得平滑的操纵效果，从而增强人机交互的体验。



2.3.3 图形控制部分

我们基于 Python 3.9，利用 OpenGL 和 PyQTGraph 的算法库开发了图形控制部分。该部分可控制 3D 重建的医学影像，并对影像的大小和位置进行实时操作。我们基于控制影像所需的方式，抽象出 12 种具体的图形控制功能，并将各功能与键盘上适当位置的按键相绑定（表 2-3），使其接收按下按键发出的信号从而对医学影像进行对应的操作。



表 2-3 12 种图形控制功能及对应按键

序号	对应功能	对应按键
1	放大	Q
2	缩小	E
3	向左平移	J
4	向右平移	L
5	向上平移	I
6	向下平移	K
7	向左旋转	A
8	向右旋转	D
9	向上旋转	W
10	向下旋转	S
11	复位	O
12	静止状态	Z

2.3.4 进程间通信

由于本软件的三个部分分别涉及不同的编程语言和程序框架，且各部分中均有多个复杂的数据结构需初始化，将其直接整合为一个软件是复杂且低效的。因此，我们将三部分分别封装为进程，再通过进程间相互通信的方式满足最终需求。常见的进程间通信方案有管道通信、本地 Socket 通信、文件共享、消息队列等多种。我们在实时数据采集部分和手势识别部分的进程之间采用文件共享的形式通信，在手势识别部分和图形控制部分的进程之间采用模拟和接收键盘按键信号的形式进行通信，方式如图 2-8 所示。

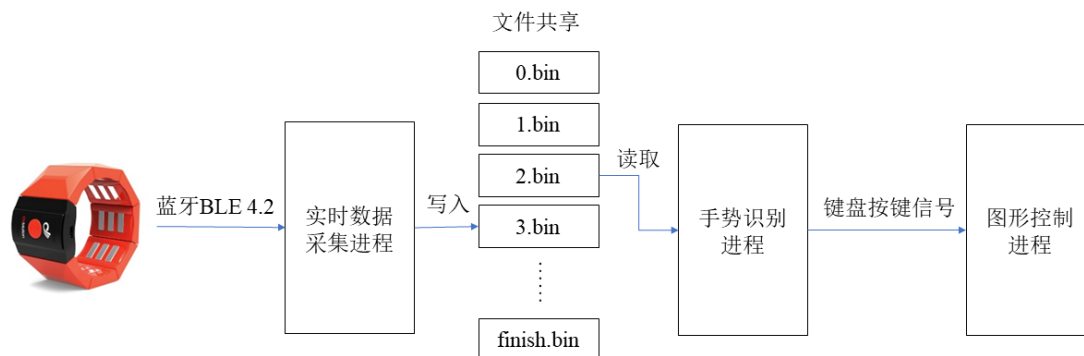


图 2-8 进程间通信方式

由于用于数据采集的开源软件 oym8CHWave 将采集到的数据以二进制文件的形式存储，我们认为直接利用该机制在前两个进程之间采用文件共享的通信方式是最直接的，因此在数据采集和手势识别的进程之间采用文件共享的方式通信。我们设定，仅当数据采集进程开始下一个数据文件的写入后（即下一个数据文件存在后），图形控制进程才能进行读取。否则图形控制进程将等待，从而避免数据冲突。

由于我们希望该人机交互图形控制软件能够广泛应用于多种计算机应用，且目前多种传统计算机应用软件均支持接收键盘发出的信号，因此我们让手势识别部分模拟发出按键信号，图形控制部分模拟接收按键信号。若需将该手势识别软件应用于其它传统计算机软件的人机交互，只需将图形控制部分替换为目标软件，然后修改手势识别部分模拟的键盘按键信号即可。



第三章 实验结果

我们对 5 个数据集都进行了 200 个 epoch 的训练,使用 9 折交叉验证的方式验证模型的分类准确性。该方法将每个数据集的 1/9 用于构建验证集,8/9 用于构建训练集。在训练过程中,神经网络模型在 5 个数据集的训练集和验证集上的识别精度变化如图 3-1 所示,可见 200 个 epoch 的训练数量足以使模型收敛。

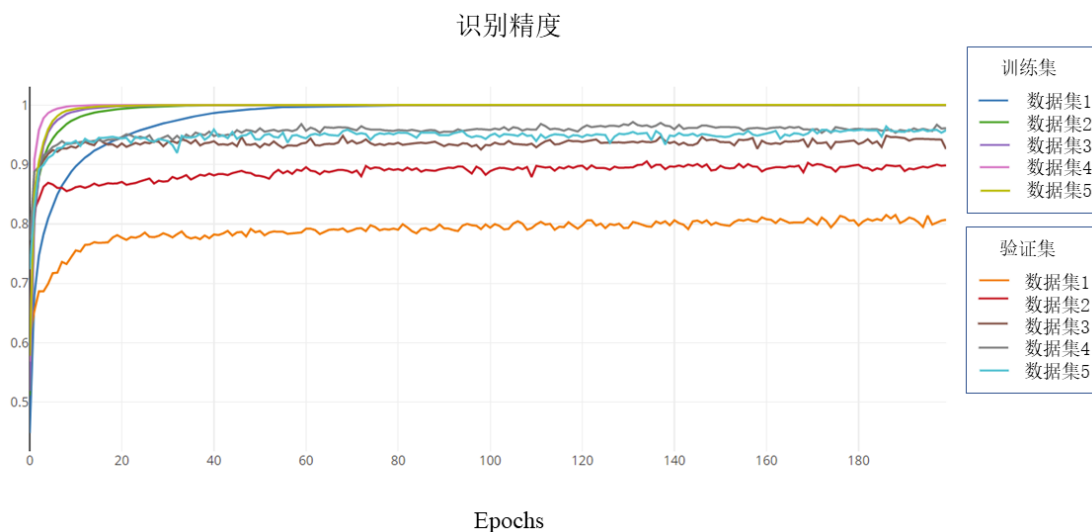


图 3-1 训练过程中识别精度的变化

在训练过程中,该神经网络模型在 5 个数据集的验证集上分别取得的最优识别精度如表 3-1 所示。该模型在由动态手势构成的数据集 1 上的识别精度明显较低,且收敛速度也明显更慢,因此我们首先排除该手势方案。在由静态手势构成的数据集 2-5 的 4 个数据集中,神经网络模型的识别精度均在 90% 以上,其中对数据集 3、4、5 的识别精度最好。

表 3-1 在验证集上的识别精度

数据集	精度
1	0.8150
2	0.9051
3	0.9483
4	0.9713
5	0.9638

我们最终选择了数据集 3 作为最终人机交互的手势方案,原因是该手势方案精度较高且对用户更加友好。在软件的控制效果方面,我们希望略有差别的相似功能(如控制图像向四个方向平移)能由略有差别的相似手势来控制,并且我们希望用户做出手势时的动作方向也能够和控制图形移动的方向尽量一致。

我们让用户佩戴好肌电臂环,在蓝牙连接的情况下逐渐远离具有医学影像的计算机设备,直至用户的操纵出现困难,并测量出此时计算机设备和用户之间的距离。经测试,用户可在约 12 米的距离内进行正常的远程操纵。我们将用户多次重复正常操纵的过程拍摄视频并统计软件的识别精度,以及测量用户做出手势到软件做出反应之间的延迟。经测试,



软件识别精度约为90%，延迟约为300毫秒。

我们最终实现的软件能够基本实现图形控制所需的相关功能。尽管在使用时仍有数百毫秒级别的延迟，但该延迟并不会明显影响用户体验，基本处于用户可接受的范围内。尽管该软件已采用多种方式使操纵效果更平滑以减小误差，但有时仍会对用户做出的手势发生误判，从而对 3D 重建的医学影像做出一个错误的操作。在发生错误操纵的情况下，用户通常能够迅速做出反馈而调整手势，从而使图像操纵回归正常状态。我们认为这种偶然性的判断失误是可以接受的。



第四章 讨论

4.1 对实验的探讨

4.1.1 实验过程

本研究的第一个目的是设计一个识别精度良好且尽量对用户友好的手势方案，以用于人机交互图形控制。第二个目的是最终实现该人机交互图形控制软件，并使得该软件的架构也可广泛用于其它常见的应用场景。经过具体的实验和分析，我们建议首先应当尽可能设计多种手势方案，再对每一种方案构建数据集并逐一训练模型，最终权衡优劣选取最优的一种手势方案构建软件。我们建议利用连续判断取众数的方式使操纵平滑，并调整窗口大小、每次接收信号后图像移动的幅度等相关参数，以提升用户体验。

4.1.2 手势方案

我们在实验中发现，由于用采集表面肌电信号的方式识别手势而非使用相机拍摄的方式识别手势，因此一些外观有所差异但肌肉用力方式相似的手势可能无法被很好地区分。我们建议在设计手势方案时尽量避免引入一些过于相似的手势，否则很可能降低模型的识别效果。此外，我们在采集数据的过程中发现，部分被试者做出一些难度较大的手势时非常窘迫，以至于难以做出标准动作（图 4-1），因此人机交互的手势方案中应当尽量避免引入这类高难度手势。



图 4-1 部分难度较大的手势

4.1.3 采集过程

由于构建的表面肌电信号数据集直接影响训练出的神经网络各节点权重，进而影响到模



型的识别精度,从而影响最终的软件效果,因此在最初构建数据集时,数据的质量非常重要。我们建议在数据采集时适当增加用户在做出两个手势的休息间隔,并且避免长时间连续采集同一位被试者的手势,以防止其手部肌肉疲劳,从而影响采集到的数据质量。我们建议在采集前首先让被试者熟悉视频和相关手势动作,以避免被试者出现重大失误。若被试者在数据采集过程中连续失误,则该组数据应当立即重采。Angkoon 等人建议在数据采集时,尽量让被试者使用不同力度做出同一个手势时的数据都包含在数据集内,以提升最终的识别精度^[6]。我们也建议在采集数据时避免让被试者以同样的力度刻板地重复相同动作。

4.1.4 滑动窗口大小

对数据窗口化时滑动窗口的大小也是影响识别精度和软件效果的重要因素。增加窗口大小可能提升识别精度,但也会增加数据处理时造成的延迟,最终设定的窗口大小需权衡这两个因素确定。研究表明,滑动窗口的大小应当保证在 200 个数据点以上且应尽可能大一些^[13],但数据处理的延迟又不宜大于 300 毫秒^[14]。Arjan 等人为平衡识别精度和延迟的两个因素,提出了移动误差率(Movement Error Rate)的新测量标准,认为一部分识别误差是由延迟产生的^[15]。但是在本实验中,我们未将由于延迟造成的误差计入识别精度的错误中,因此未采用移动误差率的测量标准。由于蓝牙传输造成的延迟已在百毫秒级别,我们选择将窗口大小设置为最低限度的 200 个数据点,以尽量减小延迟。由于不同实验所使用的数据采集设备不同,所产生的数据采集环境不同,且在构建软件时也可采用多种不同的方案,因此我们建议窗口大小应当基于实际情况具体调整,使用户获得最优的交互体验。

4.2 后续改进方向

4.2.1 添加加速度计数据

一些研究表明,与单纯使用表面肌电信号进行预测相比,添加加速度计的数据能够明显提升模型识别精度,降低预测误差^{[4][15]}。Robin 等人对增加加速度计前后的信号进行对比分析后得出结论,加速度计的信号的加入有助于模型减少假阳性的预测误差,并认为这种提升效果能够广泛应用于截肢者和未截肢者^[4]。本实验使用的 gForcePro+肌电臂环提供了 9 轴惯性测量单元(IMU)运动传感器。但由于添加更多信号可能使数据处理过程和神经网络模型更加复杂,我们只收集了用户的表面肌电信号数据进行人机交互图形控制。在 Robin 等人的实验中,添加加速度计的数据使得模型识别精度在不同的数据集中上升了约 0.5-1.5%^[4];在 Arjan 等人的实验中,添加加速度计的数据使得模型在相同延迟的情况下移动误差率(Movement Error Rate)明显降低^[15]。因此,我们认为添加加速度计数据可能是一个在避免增加延迟的情况下提升模型识别效果的合理改进方向。

4.2.2 提升采样率

Angkoon 等人对采集表面肌电信号时不同的采样率对模型识别精度的影响进行了深入研究,发现对于模型分类异常重要的信息几乎全部集中在高频信号的波段中^[6]。若将采样率从 1000Hz 降低至 200Hz,则模型的识别精度将显著下降^[6]。Wilson 等人甚至发现,若采用高通滤波器将 20Hz-120Hz 的低频表面肌电信号全部过滤,甚至能够提升模型识别精度^[16]。我们认为,提升采样率尽管能够提升识别精度,但也可能因增加单位时间内数据点的总数从而增加数据处理延迟。此外, gForcePro+肌电臂环在 12-bit 数据模式下,采样率最高仅能设置为 500Hz,在 8-bit 数据模式下最高采样率才能设置为 1000Hz。虽然 8-bit 数据模式下的每个数据点精度将不如 12-bit 模式下高,但我们依然认为采用 1000Hz、8-bit 数据模式有可能对人机交互效果带来提升,因此有进行进一步实验的必要。

4.2.3 修改进程间通信方式

操作系统进行文件读写会产生毫秒级别的延迟,具体延迟的大小根据文件大小和系统架构的不同而有所差别。由于我们使用共享文件的方式在数据采集进程和手势识别进程之间进



行通信，因此连续的文件读写可能造成了较为明显的延迟。若采用操作系统提供的管道通信或其他进程间通信方式有可能减小该延迟，从而提升人机交互效果。此外，若采用同一种编程语言将软件的各个部分重写，则易于将其统一封装并进行交互，但这样做的成本较高。我们认为进一步测试其它的进程间通信方式有可能在减小延迟方面取得进展。

4.2.4 与姿态估计相结合

许多利用相机和计算机视觉技术进行姿态估计的方法可达到 90% 以上的识别精度^[5]。尽管使用单台相机拍摄时可能因肢体遮挡带来识别困难，但可通过架设不同角度的多台相机来避免遮挡。在本实验的场景中可安装最多两台相机。我们推测，与利用上臂的表面肌电信号识别相比，利用计算机视觉技术进行手势姿态估计可能更好地识别手的不同姿态，一些姿态不同但控制肌肉方式相似的手势也可能被更好地识别。因此我们认为有必要尝试姿态估计的方案，并将效果与本实验结果相比较。甚至在后续实验中可尝试将两种模式识别方式相结合。



第 5 章 结论

我们设计了5种不同的手势方案，并利用gForcePro+肌电臂环携带的多种传感器收集被试者上臂的表面肌电信号，构建了基于5个不同手势方案的数据集。我们进而设计了一种多视角卷积神经网络模型，在综合模型识别精度和影响用户体验的多种因素后，我们从中选出了一种最合适的手势方案，并利用该手势方案最终实现了一种人机交互图形控制软件。用户可在约12米的距离内通过手势进行远程无线操纵，软件识别精度约为90%，延迟约为300毫秒。该软件将便于手术医生在操纵医疗手术机器人过程中同时调整医学影像，也可广泛用于其它应用场景。



参考文献

- [1] TURK M, ROBERTSON G G. Perceptual user interfaces (introduction) [J]. ACM, 2000, 43(3):32–34.
- [2] PHINYOMARK A, QUAIN F, LAURILLAU Y. The Relationship between Anthropometric Variables and Features of Electromyography Signal for Human-Computer Interface [M]// GANESH R. NAIK. Applications, Challenges, and Advancements in Electromyography Signal Processing. Pennsylvania: IGI Global, 2014:321–353.
- [3] OSKOEI M A, HU H. Myoelectric control systems—A survey [J]. Biomedical Signal Processing & Control, 2007, 2(4):275–294.
- [4] AYLING R, JOHNSON C G, LI L, et al. Pose Estimation from Electromyographical Data using Convolutional Neural Networks [J]. IEEE, 2000, 653–656.
- [5] OUDAH M, AL-NAJI A, CHAHL J. Hand Gesture Recognition Based on Computer Vision: A Review of Techniques [J]. Journal of Imaging, 2020, 6(8):73.
- [6] ANGKOON P, RAMI N K, ERIK S. Feature Extraction and Selection for Myoelectric Control Based on Wearable EMG Sensors [J]. Sensors, 2018, 18(5):1615.
- [7] UC Health. Benefits of Robotic Surgery [EB/OL]. (2022-05-29)[2022-05-29]. <https://www.uchealth.com/services/robotic-surgery/patient-information/benefits/>
- [8] 前瞻产业研究院. 预见 2022:《2022 年中国手术机器人行业全景图谱》(附市场规模、竞争格局和发展前景等) [EB/OL]. (2022-04-12)[2022-05-29]. <https://finance.sina.com.cn/roll/2022-04-12/doc-imcwipii3798216.shtml>
- [9] OYMotion Technologies Co., Ltd. gForce EMG Armband User Guide [EB/OL]. (2021-08-21)[2022-05-29]. <https://oymotion.github.io/assets/downloads/gForce-EMG-ARMBAND-User-Guide-202108.pdf>
- [10] HARGROVE L, ENGLEHART K, HUDGINS B. A training strategy to reduce classification degradation due to electrode displacements in pattern recognition based myoelectric control [J]. Biomedical Signal Processing & Control, 2008, 3(2):175–180.
- [11] ATZORI M, GIJSBERTS A, HEYNEN S, et al. Building the Ninapro database: A resource for the biorobotics community [C] // 2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob). IEEE, 2012: 1258–1265.
- [12] WEI W, DAI Q, WONG Y, et al. Surface-Electromyography-Based Gesture Recognition by Multi-View Deep Learning [J]. IEEE Transactions on Biomedical Engineering, 2019, 66(10): 2964–2973.
- [13] YENTES J M, HUNT N, SCHMID K K, et al. The appropriate use of approximate entropy and sample entropy with short data sets [J]. Annals of biomedical engineering, 2013, 41(2):349–365.
- [14] ASGHARIOSKOEI M, HU H. Myoelectric control systems—A survey [J]. Biomedical Signal Processing and Control, 2007, 2(4):275–294.
- [15] GIJSBERTS A, ATZORI M, CASTELLINI C, et al. Movement Error Rate for Evaluation of Machine Learning Methods for sEMG-Based Hand Movement Classification [J]. IEEE Transactions on Neural Systems & Rehabilitation Engineering, 2014, 22(4):735–44.



-
- [16] WILSON A W, LOSIER Y G, PARKER P A, et al. A bus-based smart myoelectric electrode amplifier-system requirements [J]. IEEE Transactions on Instrumentation & Measurement, 2011, 60(10):3290 - 3299



致谢

感谢上海科技大学信息科学与技术学院张玉瑶教授对本科研项目的细心、耐心、努力指导。感谢上科大信息科学与技术学院蒋颖文琪、姚奕忱、杨佳文同学对完成该科研项目所做出的努力指导和在解决问题思路上的帮助。感谢 OYMotion Technologies 公司为本科研项目提供 gForcePro+肌电臂环设备及相关开源软件 oym8CHWave。感谢 OYMotion Technologies 公司的 CEO 倪华良先生在百忙之中对本科研项目涉及到的肌电臂环技术做出多次线上答疑,为本科研项目提供大力支持。感谢海军军医大学第一附属医院(上海长海医院) 任善成教授提供珍贵的参观达芬奇医疗手术机器人的机会,以及观摩任善成教授现场操纵该手术机器人完成肿瘤切除手术的机会。

此科研项目完成期间正值上海市爆发新冠疫情。因此,本科研项目能够最终顺利完成还需特别感谢参与上海科技大学新冠疫情期间后勤保障工作的全体教师、职工、学生志愿者、援沪医疗人员和其它工作人员。我在完成实验和论文撰写的多数时间都被隔离封控在宿舍楼内,若缺少了食堂工作人员制作饭菜、志愿者配送物资、保洁人员保持宿舍楼内的环境卫生、同学在最艰难的时刻互相帮助,我和其它项目参与人员将难以维持正常的生活,此科研项目也可能无法完成。在此向大家给予诚挚的感谢。

感谢上海科技大学提供完善的基础设施和丰富的科研资料,这对在特殊时期完成本科研项目起到了至关重要的作用。

最后,感谢我的同学与我 4 年来的和睦相处,感谢我的父母对我完成学业的支持和帮助。