

# 本科生研究计划（PRP）研究论文

## （第 46 期）

论文题目: 基于视觉的运动动作识别与评估算法研究

项目负责人: 许可 学院(系): 电子信息与电气工程学院

指导教师: 许可 学院(系): 电子信息与电气工程学院

参与学生: 马张诚、谢子修、杨昊澄

项目执行时间: 2024 年 9 月 至 2025 年 3 月

# 基于视觉的运动动作识别与评估算法研究

机械与动力工程学院：马张诚 523010910130

指导老师：许可 电子信息与电气工程学院

## 摘要

本文研究了基于视觉的运动动作识别与评估算法，旨在开发一种智能体测系统，以提高体育测试的准确性和效率和公平性。项目探讨了智慧体测算法研发的意义及必要性，介绍了人体姿态估计的传统方法和深度学习方法，重点研究了 YOLOv8 深度学习与图结构模型结合的混合方法，以及基于 KNN 分类器的动作状态识别方法，用于人体姿态识别。通过构建图形用户界面(GUI)，实现了对深蹲、引体向上等动作的识别和计数。实验结果表明，该系统能够有效进行人体姿态估计和动作计数，为体育测试的智能化发展提供了新的思路和方法。最后，文章总结了研究成果，并对未来的研究方向提出了建议。

**关键词：**人体姿态识别，YOLOv8，KNN 算法，智能体测

## ABSTRACT

This paper investigates a vision-based motion recognition and evaluation algorithm, aiming to develop an intelligent physical testing system to enhance the accuracy, efficiency, and fairness of physical testing. The study explores the significance and necessity of developing intelligent physical testing algorithms and introduces traditional methods and deep learning methods for human pose estimation. The focus is on the hybrid method combining YOLOv8 deep learning with graph-structured models, as well as a method based on MediaPipe pose estimation and a KNN classifier. By constructing a graphical user interface (GUI), the system is able to recognize and count movements such as squats and pull-ups. Experimental results demonstrate that the system can effectively perform human pose estimation and movement counting, providing new ideas and methods for the intelligent development of physical testing. Finally, the research achievements are summarized, and suggestions for future research directions are proposed.

**Keywords:** Human Pose Recognition, YOLOv8, KNN Algorithm, Intelligent Physical Testing

## 1. 绪论

### 1.1 智慧体测算法研发意义及必要性

随着国家对全民健身战略的深入推进，体育事业的重要性日益凸显。2020 年 10 月，中共中央办公厅、国务院办公厅发布的《关于全面加强和改进新时代学校体育工作的意见》明确指出，学校体育是实现立德树人根本任务、提升学生综合素质的基础性工程。近年来，从中央财政到地方各级财政都加强了健身设施建设力度，多举措支持全民健身服务设施的供给，一些先进的人工智能技术也开始在体育运动中得到应用。<sup>i</sup>

然而，尽管国家在体育事业上投入了大量资源，但在实际的中高考体育测试中，绝大多数地区仍采用人工计数或传感器辅助型计数。传统人工计数存在准确性问题，容易受到个人因素的影响；而传感器辅助类仪器则因天气条件、受测人员体型差异等问题，可能导致计时计数不准确。

为了解决传统测试方法中的诸多弊端，提升体育测试的准确性和效率、公平性，开发一种基于计算机视觉的智能体测系统显得尤为重要。该系统通过摄像头采集视频图像，利用人体姿态估计模型检测人体不同区域骨骼关键点的位置坐标，结合不同运动项目的行为特征，分析学生具体进行的体质测试项目，并判断其动作的规范程度，最终输出运动数目和距离等成绩信息。这种智能体测算法不仅能够有效避免人工和传感器辅助测试的误差，还能节约人力成本，提升测试效率，为体育测试的智能化发展提供新的技术支撑。

本研究旨在探索并验证人体姿态估计在多项体质测试项目中应用的可行性和可靠性，选用适合算法，编写配套组件，完善具体的项目测试功能，实现体测任务的智能化。不仅解决了传统测试方法中的诸多问题，还为体育测试的智能化发展提供了新的思路和方法，具有重要的理论和实践意义。

1.2 人体姿态估计简述

传统人体姿态估计方法主要依赖手工设计的特征提取器和基于模型的算法，包括模板匹配、边缘检测和几何约束等，通过迭代优化或回归模型实现对人体轮廓和关键点的初步估计。图结构模型通过将人体表示为关键点及其连接关系的图结构，利用概率模型描述关键点的形状和运动约束，实现了对人体姿态的表示。随着机器学习技术的发展，传统方法开始结合支持向量机、随机森林等算法，提高了姿态估计的准确性和鲁棒性，但仍存在泛化能力差、计算效率低等问题。

深度学习的兴起推动了姿态估计技术的革新，二维单人姿态估计成为研究热点。基于坐标回归的方法直接通过神经网络预测关键点的 2D 坐标，具有模型简单、效率高的特点，但在复杂场景下准确率较低，且难以捕捉关键点间的关联。例如，通过卷积神经网络提取图像特征并逐步细化关键点坐标的回归，能够显著提高定位精度。基于热图检测的方法则通过生成关键点的概率分布热图实现定位，虽然计算复杂，但在鲁棒性和关键点关联表达上表现更优。例如，通过多阶段预测和修正逐步提升关键点定位精度，能够有效应对复杂场景中的挑战。这些方法通过深度网络的端到端学习，显著提升了姿态估计的性能。

表 1-1 单人姿态估计方法对比

类别	算法	优点	缺点
基于坐标回归	DeepPose、DEKR 等	模型简单；时间效率高	通用性较低；准确率较低
基于热图检测	CPM、HRNet 等	准确率高；关节连接关系逻辑清晰	网络结构复杂；时间效率低

多人姿态估计则需要检测图像中多个人体的关键点并将其组合成人体骨架，分为自顶向下和自底向上两种策略。自顶向下方法先检测人体目标再估计姿态，定位精度高但计算成本较大。例如，通过多阶段细化关键点定位，能够提高复杂场景下的准确性。自底向上方法则先检测图像中的所有关键点，再根据空间关系和人体结构将其分组到不同个体，具有实时性较高的特点，但在复杂背景和遮挡情况下易出现误判。

表 1-2 多人姿态估计方法对比

类别	算法	优点	缺点
自顶向下	AlphaPose、HRNet 等	定位精度较高；思路清晰	计算成本高；实时性较差
自底向上	DeepCut、OpenPose 等	受人数影响小；实时性较高	复杂背景、遮挡情况影响较大；易出现误判、匹配错误等问题

总的来说，传统姿态估计方法为后续研究奠定了基础，但深度学习方法在鲁棒性、准确性和效率上均实现了显著突破。单人姿态估计通过坐标回归和热图检测两种策略，解决了关键点定位和关联表达的问题；多人姿态估计则通过自顶向下和自底向上两种方法，适应了不同场景下的需求。这些技术的不断发展为人体姿态估计在智能体测等领域的应用提供了有力支持。<sup>ii</sup>

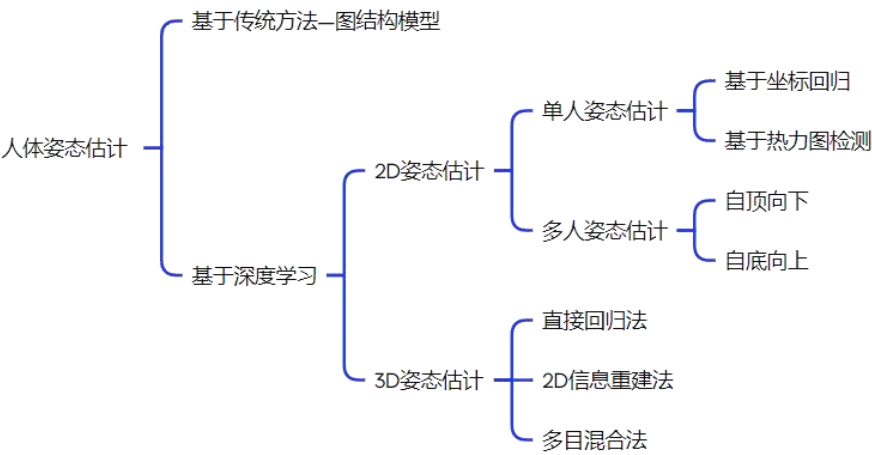


图 1-1 姿态估计分类图

## 2. YOLOv8 与图结构模型混合方法实现人体姿态识别

### 2.1 YOLOv8 人体姿态识别原理

YOLOv8 是一种先进的深度学习模型，专门设计用于实时目标检测任务。在人体姿态识别领域，YOLOv8 能够识别图像中的人体关键点，如头部、手肘、脚腕等，也能实现关键点的位置预测。检测速度快、准确率高，适合用于需要实时反馈的应用场景如视频监控、运动分析等。

YOLOv8 模型的架构包括 Backbone 网络（用于特征提取）、Neck 网络（用于特征融合）和 Head 网络（用于预测）。Backbone 网络从输入图像中提取特征，Neck 网络处理这些特征，增强模型对不同尺度目标的检测能力。Head 网络是模型的输出部分，对关键点的位置和类别进行预测。YOLOv8 通过端到端的训练方式，直接从图像像素到关键点坐标进行学习，大大简化了特征工程的过程，并提高了模型的泛化能力。人体姿态识别任务中，YOLOv8 首先识别出图像中的人体关键点，并利用其位置信息来进行后续的姿态分析，为进一步的动作识别和分析提供了基础。<sup>iii</sup>

### 2.2 YOLOv8 与图结构模型的结合

在本项目中，首先利用 YOLOv8 模型，从输入的图像或视频中来识别人体关键点，用作后续数学处理的节点。

图结构模型将人体表示为一个图，节点代表人体的关键点，边代表关键点间的连接关系。通过分析连接关系，从而推断出人体的姿态。例如，在深蹲动作中，我们可以通过分析膝关节两端的角度变化来判断动作的开始和结束。

以深蹲部分代码为例：

```
def count_squats(self, results, frame):
    keypoints = results[0].keypoints.xy[0].cpu().numpy()
    # 提取关键骨骼点
    left_hip = keypoints[11]
    left_knee = keypoints[13]
    left_ankle = keypoints[15]
    right_hip = keypoints[12]
    right_knee = keypoints[14]
    right_ankle = keypoints[16]
    # 计算髋关节、膝关节和踝关节之间的夹角
    left_angle = self.calculate_angle(left_hip, left_knee, left_ankle)
    right_angle = self.calculate_angle(right_hip, right_knee, right_ankle)
    # 使用左右腿的平均角度
    avg_angle = (left_angle + right_angle) / 2
    # 判断动作
    if avg_angle < 95 and not self.is_start_position:
        self.squat_count += 1
        self.is_start_position = True # 标记动作完成
    elif avg_angle > 120:
        self.is_start_position = False # 重置状态
    # 更新计数显示
    self.count_label.config(text=f"动作计数: {self.squat_count}")
    # 在帧上绘制计数信息
    cv2.putText(frame, f"Squats: {self.squat_count}", (10, 30), cv2.FONT_HERSHEY_SIMPLEX, 1,
    (0, 255, 0), 2)
    return frame
```

具体数学描述会在 4.2.4 中详细介绍

### 2.3 GUI 设计

为了使人体姿态识别系统更加用户友好和易于操作，本研究设计了一个图形用户界面（GUI）。该 GUI 基于 Tkinter 库（简单易用、集成于 Python）构建，用户可以通过它来选择动作门类、开始和停止检测、上传视频等。本实验中，通过 LabelFrame、Label、Combobox、Button 等组件的使用布局，实现动作选择、控制按钮和数据展示面板的设计。其中，动作选择下拉菜单允许用户从预定义的动作列表中选择一个动作，控制按钮根据用户的操作来控制检测过程的开始和停止，数据展示

面板则实时显示检测结果，使用户能够直观地了解当前的动作计数。为用户提供了友好的友好的操作界面，还通过实时反馈增强了用户的交互体验，相较于其他视觉识别模块有显著优势。进一步研究深化后，使得本项目既适用于专业的运动分析场景，也适合普通用户在家中或健身房中使用，进行自我运动监测和评估，能够很好的实现智慧体测系统的既定目标。

## 2.4 实际效果

开始运行代码后，会显示如图 2-1 的 GUI 界面，此时未选中所需要检测的运动门类，此时开始检测（摄像头实时检测）和上传视频都无法选中。

在下拉框里选择相应的运动门类后，可以点击开始检测或者上传视频。

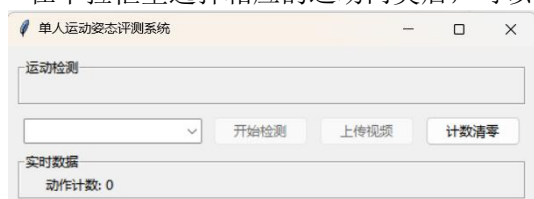


图 2-1 GUI 界面

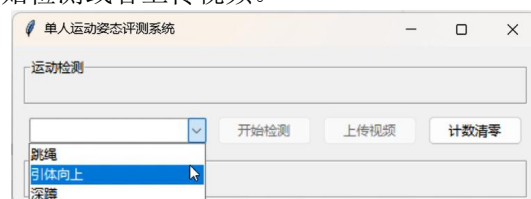


图 2-2 GUI 界面

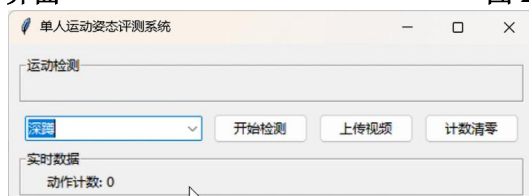


图 2-3 GUI 界面

以深蹲为例，“开始检测”后会调用本地摄像头开始检测，当检测到髋关节膝关节连线与膝关节踝关节连线（即图 2-5 中橙色两条线）夹角超过 120 度至小于 95 度一次为一次计数。动作门类和实时数量会显示在左上角，如 Squats: 6, Squats: 14。



图 2-4 摄像头检测实况

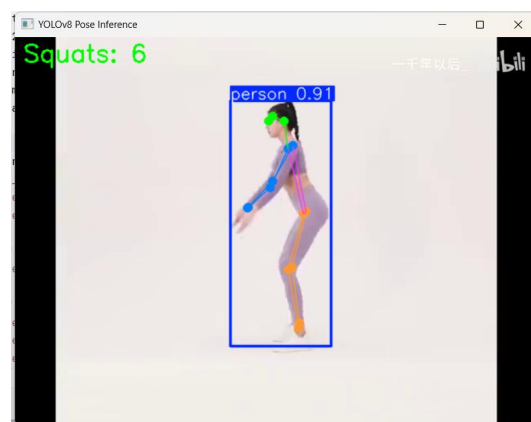


图 2-5 上传视频后分析



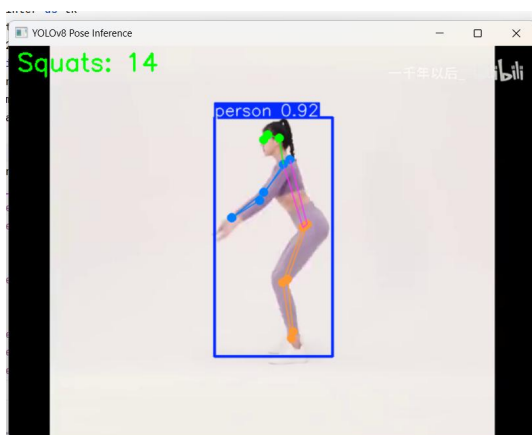


图 2-6 上传视频后分析

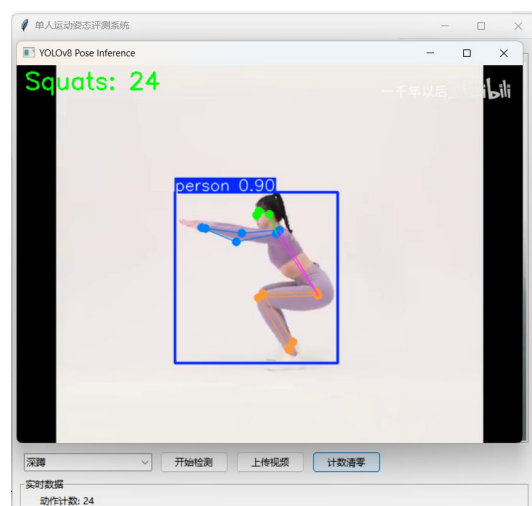


图 2-7 识别结束显示动作计数



图 2-8 计数清零按钮



图 2-9 计数清零按钮效果

计数全部完毕后，共完成多少次深蹲会计入到 GUI 页面中的实时数据-动作计数部分，如图 2-7 中的“动作计数：24”。

此时，按下“计数清零”按钮，动作计数会变成 0，便于下一次检测的开始。

引体向上板块同理，上传相关引体向上视频后，会进行计数，每次头部高于手腕至头部低于手肘为一次计数，计数完成后会显示在动作计数中，如“动作计数：25”。



图 2-10 引体向上视频分析

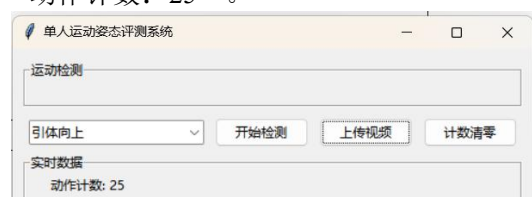


图 2-11 引体向上视频计数

## 2.5 总结

该方法利用 YOLOv8 的机器学习成果，能够直接识别出人体各个部位，并通过对人体姿态的数学描述，实现了人体姿态识别和动作计数，并通过 GUI 界面与用户交互，能够实现一些常用的基本功能。

本方法的缺点是对于运动视频的录制/拍摄视角要求高，视角较偏时计数可能不准；同时对于运动开始前或结束后的一些其他动作，有误识别的可能，最终结果可能偏大。

### 3. 基于 KNN 算法的机器学习人体姿态识别

在该方法中，我也做出了一个 GUI 界面，对深蹲、引体向上等运动进行识别；不同的是，我更多的使用了 KNN 机器学习算法，而非基础的数学关系，实现人体姿态识别。

#### 3.1 数据准备与关键点提取

首先，需要收集大量的训练样本图像，这些图像应涵盖不同的动作状态，如深蹲的“下”和“上”位置，引体向上的“下”和“上”位置。本研究未自行采集视频，而是复用公开开源的运动视频数据。对于每帧图像，调用 MediaPipe Pose 模型自动检测人体 33 个关键点的二维坐标，并将结果保存为 CSV 文件。

#### 3.2 归一化处理和特征向量表示

接着，通过定义类，将检测到的关键点转换为特征向量。该过程中，先将所有姿势进行归一化，使得不同的数据来源得到的躯干尺寸和垂直躯干方向相同。然后提取特征向量，计算特定部位之间的距离，如手腕和肩膀、脚踝和臀部以及两个手腕之间的距离，并将其进一步储存为特征向量。

#### 3.3 利用 KNN 进行回归及利用 EMA 减少噪声

KNN (K-最近邻算法) 是机器学习中的一种分类与回归方法，根据测试样本最近邻域的 K 个训练样本的类别，通过距离来预测测试样本的类别。本研究使用 KNN，利用从图像中提取的特征向量来识别人体姿态。

实际代码中，KNN 通过计算输入姿态特征向量与训练样本特征向量之间的距离来进行分类。本题中，使用了两种不同的距离度量进行两次 KNN 搜索：一次是最小坐标距离，用于过滤掉与目标样本几乎相同但在特征向量中有区别的样本，防止将相近动作错误归类；然后是平均坐标距离，用于在第一次搜索中找到最近的姿势簇。

为了减少数据集的特殊点（即噪声），防止其对实验结果产生影响，此处使用 EMA（指数移动平均）方法对分类结果进行平滑处理。EMA 是一种递归方法，通过给予最近的观测更大的权重，实现平滑时间序列的效果。EMA 平滑的计算公式如下：

$$v_t = \beta \cdot v_{t-1} + (1 - \beta) \cdot \theta_t$$

$v_t$  表示前  $t$  条的平均值， $\beta$  是加权重值（一般设为 0.9-0.999）。

对于每个新的姿态分类结果，更新数据窗口，并重新计算平滑值。这种方法减少了由于随机噪声导致的分类波动，提高了计数的稳定性。<sup>iv</sup>

#### 3.4 计数

本计数器对动作归属的概率进行判定，结合 KNN 算法学习到的进入阈值（enter\_threshold）和退出阈值（exit\_threshold）来确定何时进入和退出特定的姿态。当某动作门类的概率首次超过进入阈值时，计数器标记进入该姿态类别；一旦概率下降到低于退出阈值，计数器标记退出该姿态类别，并对计数器进行加一。

#### 3.5 实际演示

本章研究过程中同样设计了一个 GUI 界面，为用户交互增加便利性。在 GUI 界面上，可以对检测模式和动作类别进行选择，同样可以选择本地导入视频检测或者调用摄像头检测。由于本章算法不是实时反映结果，而是视频处理好之后另存到相应文件夹中，故没有实时计数功能，而是换成了退出按钮。

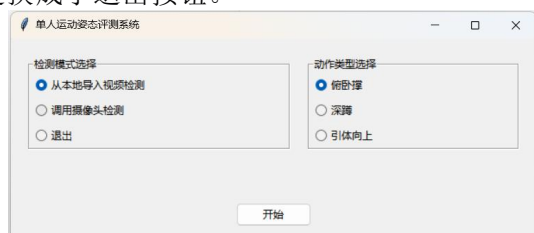


图 3-1 GUI 界面

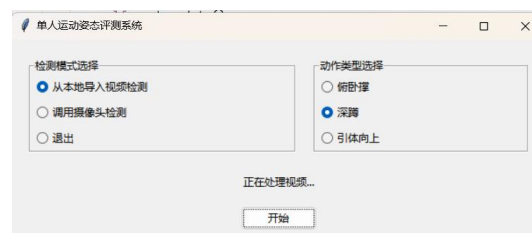


图 3-2 视频处理中展示



图 3-3 视频处理完成展示



图 3-5 深蹲动作计数



图 3-7 引体向上动作计数

视频处理结束，输出保存在

图 3-4 视频处理结束展示



图 3-6 深蹲动作计数完成



图 3-8 引体向上动作计数完成

以深蹲为例，上传视频后，会显示“正在处理视频...”，视频处理完成后，会输出“视频处理结束，输出保存在 xxxxxxxxxxxx（指地址）”。

结果同样是对人体一些姿态点进行标记，并在右上角有计数功能显示。具体结果如图 3-5 至 3-8 所示。

### 3.6 总结

相较于第 2 章的纯几何规则方法，本方法通过引入数据驱动的 KNN 分类器，具有显著优势，能更好地区分有效动作与无关肢体晃动，减少误计；对非标准但接近规范的动作具有一定容错能力。

但同时，本方法需要预先收集和标注训练数据；计算流程较长，无法满足高实时性需求；性能受限于训练数据的覆盖范围。

总体而言，该方法更适合离线视频分析场景，而第 2 章方法更适合实时交互式体测。

## 4. 总结

本项目研究了国家对于体育事业的重视投入、及研发智慧体育系统的必要性和重要意义。从“YOLOv8 结合图模型混合方法”和“KNN 机器学习算法”两个方面展开研究，分别编写、调试代码，处理数据，并设计 GUI 界面便于用户交互，最后都能有不错的人体姿态估计实际效果，项目初见成效。



囿于时间、能力有限，本项目未能更深一步研究发展。在今后可以考虑进一步优化“YOLOv8 结合图模型混合方法”，提高其算法准确度，或优化“YOLOv8 结合图模型混合方法”，增强其实时反馈性，提高交互效率。

同时，由于互相沟通的失误，未能与队友产生很好的合作效果，文中所有研究均为自行调研设计完成。在今后研究中要注意与队友的合作，以团队的力量提高效率和成果，达到“众人拾柴火焰高”的效果。

## 5. 参考文献

- [1] 李晟,宋可儿,欧阳柏强,等.基于人体姿态识别的立定跳远动作智能评估系统[J].人工智能,2022,(2):75-87.
- [2] 李豪杰,林守勋,张勇东.基于视频的人体运动捕捉综述[J].计算机辅助设计与图形学学报,2006,(11):1645-1651.
- [3] 李宾皓,李颖,孙宇飞,等.目标检测、人体姿态估计算法叠加的监控视频分析方法[J].电子技术与软件工程,2020,(7):143-147.
- [4] 曹晓瑜.基于 Mediapipe 单人运动姿态评测系统的研究与开发[D].湖北师范大学,2023.

## 6. 谢辞

非常感谢许可老师给我这次机会，让我对计算机视觉、YOLOv8、KNN 算法、Mediapipe 有了初步的了解与学习，在体验科研的过程中得到了很多长进和快乐，许可老师的鞭策也对我的持续学习起到了重要作用。感谢谢子修、杨昊澄两位同伴，在交流之中我学到了很多，没有你们的支持，本项目难以完成。感谢丁屹旻学长的陪伴与守候。最后感谢我自己，有决心、有毅力将这个项目做到自己能力范围内的极致。