

基于多模态大语言模型的数字厨师与智能烹饪系统

李鑫源¹, 李柏^{1,2}, 孙跃硕¹, 张坦探¹, 田永林³, 殷烛炎¹, 王飞跃^{4,5,6}

(1. 湖南大学机械与运载工程学院, 湖南 长沙 410082;

2. 湖南大学整车先进设计制造技术全国重点实验室, 湖南 长沙 410082;

3. 中国科学院自动化研究所多模态人工智能系统全国重点实验室, 北京 100190;

4. 中国科学院自动化研究所复杂系统管理与控制国家重点实验室, 北京 100190;

5. 中国科学院大学人工智能学院, 北京 100049;

6. 澳门科技大学创新工程学院工程科学系, 澳门 999078)

摘要: 面向高质量和精准烹饪的需求, 提出一种基于多模态大语言模型的数字厨师与智能烹饪方法。离线阶段利用视觉、声音、温度等多源传感器记录专业厨师的连续操作, 将图像与多轮问答文本融合, 建立烹饪专家知识库, 并采用低秩适配方法对预训练多模态大语言模型进行微调, 以构建能够理解烹饪意图的多模态大语言模型。在线阶段将实时感知的数据转换为图文输入微调后的大语言模型, 经模型分析后生成烹饪指令, 引导用户完成相应的烹饪动作。以煎牛排任务为例, 搭建了智能烹饪软硬件系统并进行实验验证。实验结果表明, 经过微调后的智能烹饪系统能有效控制牛排的熟度与品质, 相较于微调前的模型, 显著提升了烹饪指令的合理性和针对性。

关键词: 多模态大语言模型; 数字厨师; 智能烹饪; 烹饪机器人; 专家系统; 人工智能

中图分类号: TP391.4

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202448

Digital chefs and intelligent cooking systems based on multimodal large language model

LI Xinyuan¹, LI Bai^{1,2}, SUN Yueshuo¹, ZHANG Tantan¹, TIAN Yonglin³, YIN Zhuyan¹, WANG Fei-Yue^{4,5,6}

1. College of Mechanical and Vehicle Engineering, Hunan University, Changsha 410082, China

2. State Key Laboratory of Advanced Design and Manufacturing Technology for Vehicle, Hunan University, Changsha 410082, China

3. State Key Laboratory for Multi-modal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

4. State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

5. School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China

6. Department of Engineering Science, Faculty of Innovation Engineering, Macau University of Science and Technology, Macao 999078, China

Abstract: A digital chef and an intelligent cooking method were proposed to achieve high-quality, precise cooking results. In the offline phase, visual, auditory and thermal sensors record professional chefs' continuous cooking operations. The collected frame-by-frame images and multi-round Q&A texts form a culinary expert knowledge base. A low-rank adaptation method was applied to fine-tune a pretrained multimodal large language model, enabling it to understand cooking intentions. In the online phase, real-time sensory data were converted into image-text inputs for the fine-tuned model, which then generated cooking instructions to guide users through the cooking steps. A hardware-software cooking system was implemented and tested with a pan-frying steak task. Experimental results show that the fine-tuned system effectively con-

收稿日期: 2024-09-27; 修回日期: 2024-11-08

通信作者: 李柏, libai@zju.edu.cn; 王飞跃, feiyue.wang@ia.ac.cn

基金项目: 国家自然科学基金项目 (No.62103139); 湖南省芙蓉计划湖湘青年英才项目 (No.2023RC3115)

Foundation Items: The National Natural Science Foundation of China (No.62103139), Lotus Youth Talent Program of Hunan Province (No.2023RC3115)

trols the steak's doneness and quality, and significantly improves the accuracy and rationality of cooking instructions compared to the model before fine-tuning.

Key words: multimodal large language model, digital chef, intelligent cooking, cooking robot, expert system, artificial intelligence

0 引言

随着生活水平的提升和健康管理意识的增强,人们对饮食品质的需求日益增长,传统烹饪方法在满足多样化饮食需求时逐渐显现出诸多不足^[1]。如何在保证食物感官体验的同时兼顾营养价值,已成为消费者和烹饪从业者面临的重要挑战。这一挑战不仅涉及食材选择和加工工艺,还包括对烹饪过程的精准有效控制,而对于非专业人士来说这些因素往往难以兼顾^[2]。

机器人烹饪技术通过先进的智能控制和多传感器实时反馈机制,能够有效减少厨师在烹饪过程中可能出现的人为操作失误。这不仅确保了烹饪过程的精准性和一致性,还避免了人为因素导致的菜肴品质波动^[3]。特别是对于一些步骤复杂、需要细致控制的菜肴,机器人烹饪技术展现出巨大的应用潜力^[4]。与传统人工操作相比,机器人烹饪技术不仅能够保持食材质量和口感的一致性,还能缩短烹饪时间,提高整体效率。因此,机器人烹饪技术不仅适用于日常家庭烹饪,也在餐饮行业中具有广泛的应用前景。

近年来,随着大语言模型和多模态交互技术的迅速发展,烹饪机器人的适应性和智能决策能力有望进一步提升^[5]。大语言模型使机器人能够更好地理解和处理人类通过自然语言表达的个性化饮食需求,如用户对菜肴口味、营养成分和烹饪风格的具体要求。同时,多模态交互技术通过整合视觉、声音、温度等多种传感器数据,使机器人能够实时感知烹饪环境和食材状态。这些技术的结合使烹饪机器人能够精确调节烹饪过程中的关键参数,如温度、湿度和时间,从而提高菜肴的烹饪精准度和整体效率。借助这些先进的人工智能技术,烹饪机器人不仅能够不同厨房环境中灵活操作,还能满足用户多样化的烹饪需求,显著增强烹饪的灵活性和高效性。因此,本文旨在研究如何利用多模态大语言模型开发一套智能烹饪方法,并基于此方法构建软硬件集成的数字厨师平台,以进一步推动智能烹饪技术的发展和應用。

1 智能烹饪与大语言模型

随着经济的发展,人们对饮食品质的要求日益提升。从主观角度来说,人们更注重食物的色泽、香气、风味以及用餐环境的品质^[6];从客观角度来看,则更加关注饮食的营养价值^[7]。饮食的色泽、香气、风味以及用餐环境在很大程度上影响人们的饮食选择和行为,若这些因素无法被满足,容易导致人们的食欲下降^[8]。同时,合理的营养摄入作为维持人体正常生理功能和健康状态的关键因素,不仅在预防疾病和促进康复方面发挥重要作用,还能提升生活质量和工作效率^[9]。因此,日常饮食需综合考虑人们的主观和客观需求,以实现饮食的全面平衡。然而,对于大多数人来说,在日常烹饪中兼顾饮食的主观与客观需求是一项极具挑战性的任务。从主观需求出发,厨师需选择合适的烹饪方法,不仅要提升食物的色、香、味,还要创造愉悦的用餐体验;从客观需求来看,人体所需的各种营养元素分布在不同的食材中,因此合理搭配食材至关重要。这要求厨师既具备精湛的烹饪技巧,又拥有丰富的营养学知识和美学素养。而对于非专业人士而言,兼顾这些需求更是难上加难。

多模态大语言模型技术与智能烹饪的结合,有望有效解决上述难题。多模态大语言模型通过整合多种数据类型,展现出强大的分析与推理能力,能够生成并理解与人类语言相关的复杂内容^[10]。这种分析与推理能力为在日常烹饪中同时兼顾食物的主观与客观需求提供了便捷的解决途径。在多样化的烹饪数据上对多模态大语言模型进行微调,模型可以掌握全面的烹饪知识,从而形成一个融合多模态数据(包括视觉、温度、声音、气味等)的烹饪专家系统。该系统能够深入理解烹饪过程中的细微差别,从多个维度为用户提供个性化的建议和解决方案,这是传统烹饪方案难以实现的。结合多模态大语言模型的智能烹饪系统是未来烹饪领域的重要发展方向之一,为解决日常烹饪中的难题提供了新契机。

在探索多模态大语言模型与烹饪结合的过程

中,选择典型且易于分析的烹饪任务至关重要,而牛排正符合这一要求。作为广受欢迎的国际美食,牛排因其独特风味、丰富的营养价值和优雅的用餐体验而深受消费者喜爱^[11]。其广泛普及促使大量牛排烹饪教程的出现,为多模态大语言模型的训练提供了丰富的素材。牛排的烹饪过程相对简单,主要包括翻面和温度调整两个关键步骤。尽管操作简单,实际过程却蕴含丰富的技巧与经验。不同的烹饪操作会显著影响牛排的最终质量,尤其是在口感、色泽和嫩度等方面表现出明显差异^[12]。这些特点使牛排的烹饪过程与多模态大语言模型的结合具有合理性,并能够从多个维度分析与评估烹饪效果。

本文提出一种基于多模态大语言模型的智能烹饪方法。首先,作为大语言模型微调的前置步骤,离线构建了多模态烹饪专家知识库,利用多种传感器记录专业厨师的连续烹饪过程,将多模态数据转化为逐帧图像与相应的多轮问答文本。随后,基于上述知识库数据,采用低秩适配(low-rank adaptation, LoRA)方法对LLaVA(large language and vision assistant)大模型进行微调,从而完成智能烹饪系统的离线学习过程。在智能烹饪系统的在线使用阶段,通过传感器实时捕获烹饪过程中的视觉、听觉、温度等多模态数据,将其转化为图文信息并按一定频率输入微调后的大语言模型,进而生成自然语言格式的烹饪指令,指导整个烹饪过程直至烹饪结束。以煎牛排这一具体烹饪任务为例,在实践中检验本文所开发智能烹饪软硬件系统的有效性。本文的主要贡献如下。

(1) 提出一套基于多模态大语言模型的智能烹饪方法体系。

(2) 提出一套适用于烹饪任务的多模态专家知识库构建方法以及在实践中基于GPT-4的简便处理策略。

(3) 搭建功能完善的智能烹饪软硬件平台,并在实践中检验本文所提智能烹饪方法的有效性。

2 相关研究

随着烹饪技术的不断发展,研究重点经历了从探索影响饮食行为的因素,到提升厨师专业素养,再到智能烹饪系统的逐步完善。这一演变的核心目标是更好地平衡饮食中的主观需求与客观需求,不仅推动了烹饪技术的进步,也为个性化饮食方案的

制定提供了新机遇,促进了烹饪领域向科学化和精细化方向发展。

在探讨不同因素对人们饮食行为影响的研究中,研究者从多个角度进行分析。季鸿崑^[13]指出,饮食美学在烹饪过程中不仅提升了食物的艺术性,还通过文化表达丰富了烹饪内涵,使其超越单纯的味觉体验,成为一种深层次的文化活动。陈梦晓^[14]将食品造型与雕塑艺术结合,食品造型不仅满足饮食需求,还是文化与艺术的载体。研究强调,将雕塑艺术元素融入食品外观设计,不仅能够提升食品的美学价值和文化意义,还能增强感官吸引力和消费体验。这种融合不仅增强了食品的视觉冲击力,还能传递特定的文化信息,进一步丰富食品在日常生活中的功能。Morquecho-Campos等^[15]则探讨了食物气味对食欲、食物偏好和摄入量的影响,研究结果表明,特定食物气味能够显著增加食欲,尤其在蛋白质类食物中表现更为明显。

为提升厨师的专业素养,研究者从食物的色泽、香气、风味、营养以及饮食美学等方面入手,致力于帮助厨师提高综合能力。侯兵^[16]在对我国烹饪高等教育发展的研究中指出,高质量的烹饪教育对于培养具有综合素养的专业人才至关重要。研究结果表明,系统化的教育体系和专业培训可以有效提升烹饪从业者的理论水平和实践能力,不仅满足菜品在视觉和味觉上的高要求,还能最大限度保留和提升食物的营养价值。李楠等^[17]的研究进一步表明,烹饪温度和时间是影响食品营养成分的关键因素之一。研究发现,合理控制温度和时间有助于减少营养物质的损失,而适度的温度与时间调控可以在保障食物感官质量的同时,最大限度保留其中的营养成分,为提高膳食整体营养水平提供了科学依据。Suhairom等^[18]提出星级厨师素质模型,该模型涵盖影响厨师职业可持续发展的六大关键素质,包括技术素质、非技术素质、个人素质、自我概念、身体状态和动机,为提升厨师职业素养提供了理论指导。

借助智能化技术手段,学者们开发了一系列智能烹饪系统,并探讨了未来烹饪系统的发展方向。吴晓蒙等^[19]提出智慧餐饮4.0概念,强调通过融合人工智能与新型食品科学技术,推动餐饮行业的智能化发展。研究指出,数字孪生技术可模拟烹饪大师的技艺,使机器人精准再现传统烹饪技法,为餐饮行业的技术进步和专业化发展提供了系统性的解

决方案。高莹婷等^[20]开发了一种基于电子鼻的烹饪方式检测方法,该方法能够高效分类蒸煮、煎炒和油炸3种烹饪方式,平均准确率达95%。该方法不仅帮助用户记录烹饪过程,还能优化烹饪方式,减少营养流失,为智能厨房场景下的饮食管理提供了技术支持。Liu等^[21]提出一种机器人双臂协作框架,结合图神经网络和Transformer模型,实现类人动态双臂翻炒操作。研究证明,机器人能够在不明确动态模型条件下,通过视觉反馈调整动作,达到目标翻炒效果。这种方法不仅解决了翻炒过程中内容物变形导致难以建模的问题,还为复杂双臂非抓取操作任务提供了重要的技术支持。

综上所述,烹饪领域的研究从多个角度探讨了饮食行为的影响因素,并通过提升厨师素养和引入智能化系统,显著推动了烹饪技术的发展。这些研究不仅深化了对饮食文化的理解,还为制定兼顾主观需求与客观需求的饮食方案提供了科学依据。然而,尽管已有研究在智能烹饪系统和厨师专业培训方面取得了显著进展,但在多模态数据整合与大语言模型应用于烹饪决策支持方面仍存在不足,特别是在实现个性化和精准化的烹饪指导、优化多传感器数据的实时处理以及提升烹饪系统的自适应能力方面,尚有进一步探索的空间^[22-23]。基于此,本文利用多模态大语言模型开发了一个智能烹饪系统,并构建了软硬件集成的数字厨师平台,以填补现有研究的空白,进一步提升智能烹饪系统的智能决策和操作能力,从而为未来烹饪技术的科学化和精细化发展提供新的解决方案。

3 基于多模态大语言模型的智能烹饪方法

本文提出一种基于多模态大语言模型的在线智能烹饪方法,该方法利用视觉、温度、声音等多源

感知数据,并通过自然语言指令辅助用户逐步完成烹饪过程。

3.1 智能烹饪系统总体架构

智能烹饪系统总体架构如图1所示,智能烹饪系统的整体流程分为离线学习与在线应用两个阶段。在离线阶段,通过多模态传感器连续采集专业厨师的示范烹饪过程,并对获取的视觉信息进行标注,从而构建图文对应的多模态烹饪专家知识库。基于该知识库提供的训练样本,采用LoRA方法对多模态大语言模型进行微调,使其具备生成适宜烹饪建议的能力。在进入在线应用阶段后,经过微调的智能烹饪模型可实时处理传感器上传的多模态数据,并在一定时间间隔内输出自然语言烹饪指令,从而引导用户高效、准确地完成菜肴的烹饪。

3.2 多模态烹饪专家知识库的构建

本节将详细介绍多模态烹饪专家知识库的离线构建过程。多模态烹饪专家知识库离线构建流程如图2所示,主要包括专家烹饪数据的连续采集与记录、对采集数据进行多轮问答形式的标注两个步骤,经过这两个步骤即可形成供模型微调使用的多模态烹饪专家知识库。

在专家烹饪过程记录环节,需要利用传感器对专业厨师的实际烹饪过程进行全面记录。传感器的配置应根据具体烹饪任务的需求进行调整,但视觉传感器是不可或缺的基础配置,用于持续记录食材颜色、纹理、形态的变化过程,厨师的烹饪动作以及厨师对烹饪过程的语音评价等。根据菜肴和烹饪手法的特点,还可以增设红外传感器监测食材表面的热力分布,或利用声音传感器捕捉烹饪过程中油脂受热、食材翻动等产生的声学特征。在必要情况下,还可引入气味传感器(如电子鼻)对风味释放进行定性与定量分析,从而为数据融合与智能决策

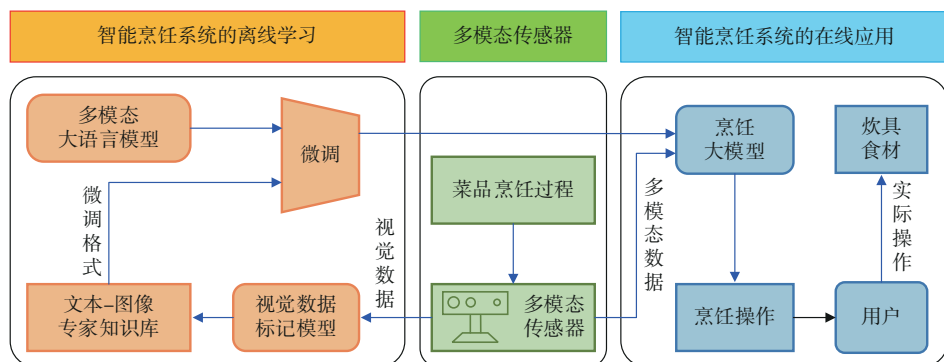


图1 智能烹饪系统总体架构

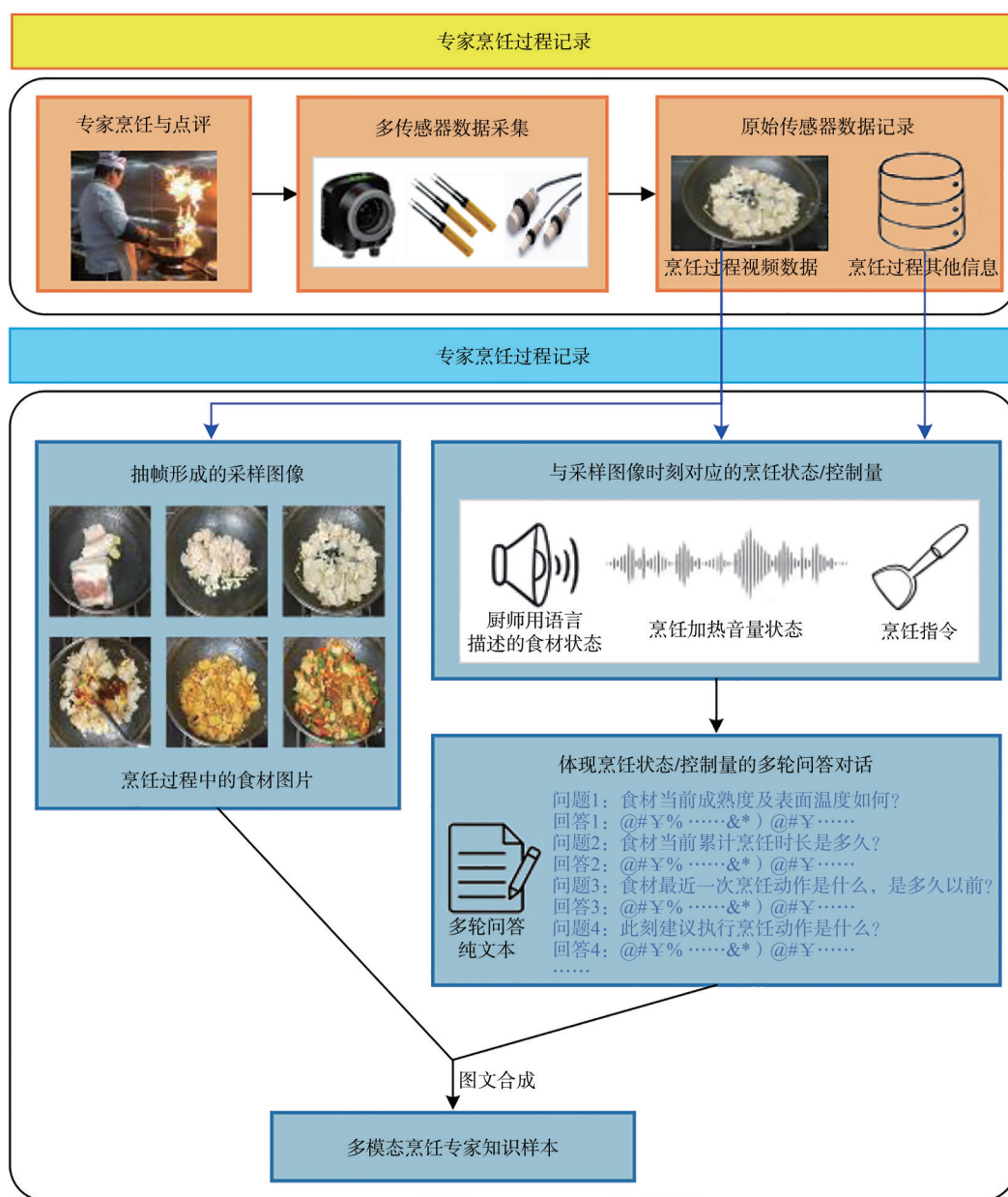


图2 多模态烹饪专家知识库离线构建流程

提供更全面的支持。

在训练样本标注环节,为了使多模态大语言模型能够充分利用图像及其关联数据,需要将每帧采样图像的状态和控制信息转化为多轮问答格式的文本。具体而言,需要为每张采样图像生成对应的问答文本信息,包括烹饪历史动作、当前时刻、当前食材熟度、当前食材质地以及下一步操作建议等。通过将问答文本与相应的采样图像绑定,可以构建完整的多模态烹饪专家知识样本。对所有采样图像执行此类转换和标注操作,最终生成由大量多模态烹饪专家知识样本组成的多模态烹饪专家知识库,

为后续模型微调提供高质量的训练数据。

需要强调的是,理想的训练数据采集应尽可能精准地还原专家的每项烹饪动作及相应的食材状态(如温度、湿度、熟度等),以实现图像、状态量和控制量(如火候调整、翻面操作等)的高度关联。然而,由于本研究无法获得国际顶尖厨师的原始数据,采用了一种基于OpenAI开发的GPT-4大语言模型的替代方案。具体来说,本文将原始视频中的所有帧图像按时间顺序排列,以 $2N$ 为间隔抽取关键帧,将这些关键帧输入GPT-4,并要求GPT-4根据当前食材状态画面生成当前时刻的食材状态量以

及应施加的烹饪控制量。GPT-4通过分析每帧图像内容，为每个关键帧生成对应的状态量（如食材热度）和控制量（如火候调整、是否关火），本文将GPT-4推测的控制量视为专家动作，并将其与该关键帧及其前后 N 帧图像信息同时进行关联。需要再次强调的是，这个基于GPT-4的标注策略仅为临时解决方案。如果条件允许，应尽量获取真实的专家烹饪数据，以确保训练样本的准确性和可信度。

3.3 多模态大语言模型的微调

本节介绍了多模态大语言模型微调的方法。微调过程包括两个主要步骤：（1）选择合适的预训练基础大模型；（2）在该基础上进行针对性微调训练，使预训练基础大模型能够适应特定任务。

在选择合适的预训练基础大模型时，本研究选择了LLaVA作为预训练基础大模型。这一选择的原因在于LLaVA已通过大规模图文指令数据完成预训练，具备较强的图像理解和自然语言处理能力^[24]。此外，LLaVA在多模态推理和对话任务中表现优异，能够满足烹饪场景中复杂多变的多模态信息处理需求^[25-26]。

在进行针对性微调训练时，本研究采用低秩适配LoRA方法对LLaVA进行微调。LoRA的核心思路是在冻结预训练模型权重的同时，在Transformer结构的每层中注入低秩分解矩阵，从而显著减少微调过程中需要更新的参数数量^[27]。该方法不仅能够降低计算和存储成本，还能为模型快速适应和任务切换提供高效的解决方案。在本研究中，LoRA方法针对烹饪场景中的视觉信息进行微调，以增强模型对烹饪图像中关键信息的识别和处理能力。通过这一过程，LLaVA能够更准确地理解与烹饪相关的视觉输入，并生成精准的烹饪控制指令。具体而言，LoRA方法通过在LLaVA模型的权重矩阵中引入一个低秩的偏置项 ΔW 实现微调，微调后模型的参数矩阵 W_{update} 更新为：

$$W_{\text{update}} = W_0 + \Delta W \quad (1)$$

其中， $W_0 \in \mathbb{R}^{d \times k}$ 为模型微调前的参数矩阵，即LLaVA模型的原始参数矩阵， d 为原始参数矩阵行数， k 为原始参数矩阵列数。原始参数矩阵 W_0 与偏置项 ΔW 相加得到微调后模型的参数矩阵。 ΔW 与 W_0 的行列数应相同， ΔW 表示为：

$$\Delta W = B \cdot A \quad (2)$$

其中， $B \in \mathbb{R}^{d \times r}$ ， $A \in \mathbb{R}^{r \times k}$ ， $r \ll \min(d, k)$ 为 ΔW 的

内在秩。

在训练过程中，将偏置矩阵 ΔW 分解为两个低秩矩阵 A 和 B ，从而降低计算开销。基于LoRA的大语言模型微调方法基本原理如图3所示，在训练过程中，LLaVA基础模型的参数矩阵 W_0 保持不变，只调整低秩矩阵 A 和 B 的参数，从而改变偏置矩阵 ΔW ，以达到微调的目的。

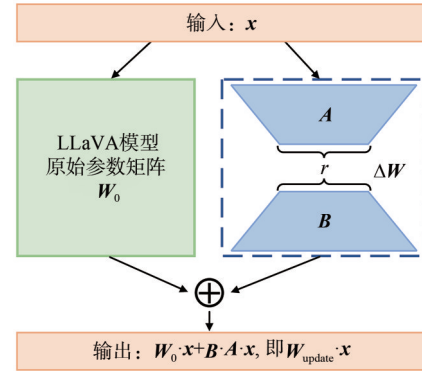


图3 基于LoRA的大语言模型微调方法基本原理

本文利用LoRA方法，结合多模态烹饪专家离线知识库对LLaVA多模态大语言模型进行微调，得到参数矩阵的偏置矩阵 ΔW 。将偏置矩阵 ΔW 与LLaVA基础模型参数矩阵 W_0 进行代数相加，得到微调参数矩阵 W_{update} 。至此，大语言模型的离线微调过程完成。

4 实验与讨论

本节围绕智能烹饪系统在煎牛排任务中的实际表现展开实验与讨论。

4.1 智能烹饪软硬件系统的搭建

以普通家庭厨房环境中的煎牛排任务为例，本节介绍智能烹饪软硬件系统的搭建与部署方案，总体效果如图4所示。

智能烹饪系统的硬件部分主要由传感器、计算机和炊具3部分组成，每类设备均为单台（套）配置。传感器硬件设备包括：（1）用于获取牛排烹饪过程中视觉图像数据的S-YUE SY011HD视觉传感器（图5（a））；（2）用于捕捉牛排表面热力分布图像的MELEXIS MLX90640红外传感器（图5（b））；（3）用于采集烹饪过程中音量时序信号的SOONSALL XM8165B声音传感器（图5（c））。上述3类传感器协同工作，为后续的数据处理与分析提供了多维度、细粒度的感知数据。



图4 牛排智能烹饪软硬件系统



图5 牛排智能烹饪系统中的传感器硬件设备

在计算机硬件设备方面，智能烹饪系统配置了一台本地便携式计算机和一台高性能云端服务器。本地便携式计算机采用惠普（HP），内置英伟达（NVIDIA）GeForce RTX 4060显卡，拥有8 GB显存及1 830 MHz GPU频率，主要负责对传感器原始数据进行初步处理，并将生成的多模态图像数据与文本信息上传至云服务器，同时将云服务器返回的烹饪指令展示给用户。云服务器则采用NVIDIA Tesla V100-SXM2显卡，拥有32 GB显存与1 290 MHz GPU频率，负责对上传的多模态信息进行大语言模型推理计算，进而生成高质量的烹饪动作指令。通过本地与云端协同处理，智能烹饪系统实现了数据的有效流转与处理，从而保证烹饪过程的高效性。

在执行机构方面，智能烹饪系统选用由燃气灶、煎锅和食品夹具组成的基础炊具组合，用于执行多模态大语言模型生成的自然语言格式的烹饪指令。燃气灶作为热源控制单元，煎锅用于承载食材，食品夹具则辅助食材的翻转和位置调整。

牛排智能烹饪系统的软件层面以Ubuntu 20.04为基础操作平台，在本地便携式计算机与云端高性

能服务器上均采用统一的系统环境，以确保应用程序在不同计算节点之间的兼容性和稳定性。软件开发语言采用Python 3.10.15，该版本在稳定性和依赖兼容性方面具有良好的表现，为数据处理、模型推理和网络通信等功能模块提供了可靠的底层支持。借助Python丰富的生态社区与第三方库资源，该系统能够高效解析和融合多模态数据，并快速集成深度学习框架，为多模态大语言模型在云端服务器上的部署与运行奠定了坚实的技术基础。

图6给出了智能烹饪系统的硬件设备布局与数据传输链路。在实际应用过程中，视觉与红外传感器固定于煎锅上方，对准正在加热的牛排，以便连续获取高分辨率的视觉图像与热力分布数据。声音传感器则置于煎锅附近，用于捕捉烹饪过程中由加热、油脂翻腾及微观化学变化产生的音量时序信息。上述3类传感器协同工作，实时将原始数据流（如图7所示）传输至本地便携式计算机进行初步处理与多模态融合。经过数据处理与融合后，本地便携式计算机生成可供分析的多模态图像数据及其对应的文本提示信息，并将其上传至云服务器，在

云端执行多模态大语言模型的推理计算与烹饪指令生成任务。

云服务器上部署的多模态大语言模型对收到的数据进行分析与推理,生成相应的烹饪动作指令,并通过网络回传至本地便携式计算机,在显示器上以直观的形式呈现给厨师。厨师根据指令调整炉火大小、翻转牛排或改变烹饪时间,从而使烹饪过程在多模态感知与智能决策的引导下有序进行。这一数据传输链路确保了智能烹饪系统从多模态感知、

数据本地预处理、云端推理到本地执行的高效协作,使智能烹饪过程更加高效、准确和可控。

4.2 面向煎牛排任务的多模态大语言模型离线学习实验

针对煎牛排这一具体任务,本节详细介绍智能烹饪系统中多模态大语言模型的离线学习实现方案。离线学习过程主要包括两个环节:(1)构建多模态烹饪专家知识库;(2)利用该知识库对多模态大语言模型进行微调。

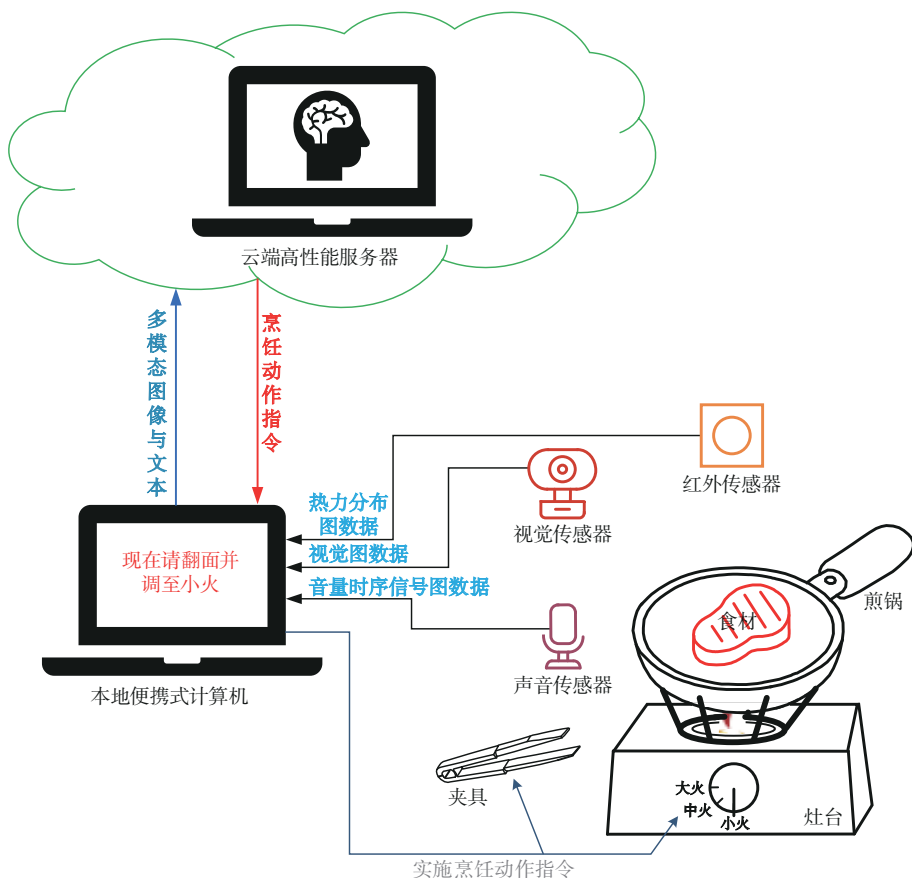


图6 牛排智能烹饪系统在线应用过程中的硬件设备及信息链路

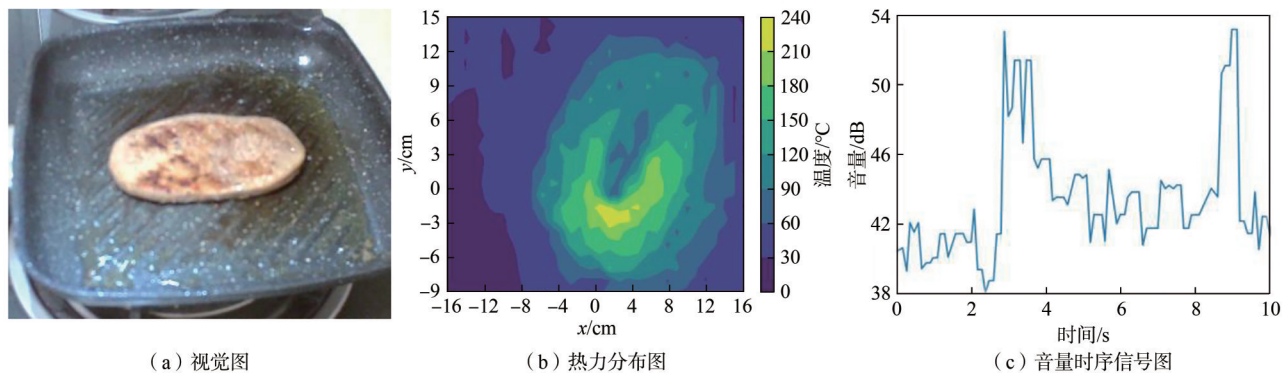


图7 传感器硬件设备采集到的多模态原始图像数据

在构建多模态烹饪专家知识库的过程中，本文首先将牛排烹饪的连续图像数据输入GPT-4进行自动化标注，如图8所示。在标注过程中，为每个图像帧附加时间戳信息，并提示GPT-4从专业厨师的视角分析牛排的当前熟度，从而生成相应的烹饪动作指令建议。该过程将原始图像数据转化为富含语义信息的文本标注数据，为后续模型微调奠定基础。需要特别说明的是，考虑LLaVA在英文问答任务中的优异表现，知识库中所使用的文本信息均采用英文描述。

由于从GPT-4获得的文本标注数据仅包含一组简单的时序信息和指令描述，为便于LLaVA模型的微调与训练，本文将这些文本信息组织为多轮对话格式，并存储到多模态烹饪专家知识库中。具体而言，每组对话数据均包含一张图像的文件路径（如图9中的b37-1.png）及其对应的两轮问答。其中，

每组数据的唯一编码id（如b37-1）用于标识该对话与图像的关联关系；对话中human字段表示用户侧的提问，gpt字段则为模型给出的回答，两者围绕牛排熟度判断与烹饪控制指令展开。通过上述流程，本文共获得1 133组数据，每组数据包含一张图像及两轮问答信息。这些经过自动标注与对话格式转化后形成的数据集合，构成了多模态烹饪专家知识库的核心内容，并为后续利用LoRA方法对LLaVA模型进行微调提供了充足且高质量的训练样本。

针对多模态大语言模型的微调，本文采用上述步骤构建的多模态烹饪专家知识库，对预训练完成的LLaVA-v1.5-13b多模态大语言模型进行微调。微调过程在高性能云服务器上进行，设置学习率为 2×10^{-4} 、训练轮次为10、批次大小为8。经过这一微调步骤，本文成功获得了专用于牛排智能烹饪任务的多模态大语言模型。



The steak is currently being cooked at the 262nd second.
1. Please determine the doneness of the steak shown in the image and explain the basis for your judgment.
2. You are a chef skilled in cooking steaks. You are currently preparing a steak, and its current state is as shown in the image. What should be done at this moment to cook the steak? (Only address actions needed at this specific time point, focusing on adjusting the heat and flipping the steak, without considering future steps).
Please answer these two questions in order.



1. Doneness of the steak and basis for judgment

The steak appears to be in the range of medium-rare to medium. This assessment is based on the steak's surface color, which shows a light brownish hue with some retained pink, and the presence of visible juices on the surface. The juices seem to be slightly cloudy, indicating that the internal temperature has not yet reached higher levels associated with more done stages like medium-well or well-done. The texture of the steak also appears tender, consistent with this level of doneness.

2. Actions to take at this moment

At this specific moment, maintain medium heat to avoid overcooking the outer layer while allowing the internal temperature to rise evenly. Flip the steak to cook the other side, ensuring balanced heat distribution and achieving a more even sear. Avoid pressing down on the steak, as this could release juices and reduce its tenderness.

图8 使用GPT-4对视图图数据进行自动化标注

```

{
  "id": "b37-1",
  "image": "b37-1.png",
  "conversations": [
    {
      "from": "human",
      "value": "<image>\nWhat is the doneness level of the steak in the picture?"
    },
    {
      "from": "gpt",
      "value": "The steak appears to be in the range of medium-rare to medium. This assessment is based on the steak's surface color, which shows a light brownish hue with some retained pink, and the presence of visible juices on the surface. The juices seem to be slightly cloudy, indicating that the internal temperature has not yet reached higher levels associated with more done stages like medium-well or well-done. The texture of the steak also appears tender, consistent with this level of doneness."
    },
    {
      "from": "human",
      "value": "What steps should be taken to cook a steak at this moment?"
    },
    {
      "from": "gpt",
      "value": "At this specific moment, maintain medium heat to avoid overcooking the outer layer while allowing the internal temperature to rise evenly. Flip the steak to cook the other side, ensuring balanced heat distribution and achieving a more even sear. Avoid pressing down on the steak, as this could release juices and reduce its tenderness."
    }
  ]
}

```

图9 多模态烹饪专家知识库数据存储格式

4.3 智能烹饪系统的实际烹饪表现与理解能力实验验证

本节对该系统在实际烹饪过程中的表现和理解能力进行实验验证。

4.3.1 实验操作

在实际烹饪表现的对照实验中，人类用户根据智能烹饪系统在微调前后生成的烹饪动作指令，在家庭厨房环境中完成牛排的烹饪任务。本文实验选用厚度、大小和质量基本相同的牛排，以确保实验的一致性和可比性。多模态图像在云服务器上的处理流程如下。

(1) 热力分布图处理

将热力分布图输入多模态大语言模型（以下简称“模型”），提示模型当前正在煎牛排，并提供牛排烹饪热力分布图。模型分析并提取关键信息，以自然语言形式输出热力分布信息。

(2) 音量时序信号图处理

将音量时序信号图输入模型，提示模型当前正在煎牛排，并提供过去 10 s 的音量时序信号图。模型分析并提取关键信息，以自然语言形式输出音量时序信号信息。

(3) 提示信息综合处理

将生成的热力分布信息和音量时序信号信息记录为提示信息 INFO_1；将当前牛排烹饪的时间戳和此次开始烹饪后的历史动作记录为提示信息 INFO_2；将牛排烹饪经验与常识（如最大/最小翻

面频率、烹饪油温及烹饪时长等）记录为提示信息 INFO_3；将预期输出的烹饪动作格式及指令编码规则记录为提示信息 INFO_4，具体编码规则如表 1 所示；将上述提示信息 INFO_1、INFO_2、INFO_3 和 INFO_4 整合为完整提示信息 INFO_ALL。

表1 牛排烹饪动作指令与数字编码的对应关系

烹饪动作指令	数字编码
不翻面用小火	1
翻面用小火	2
不翻面用中火	3
翻面用中火	4
不翻面用大火	5
翻面用大火	6
关火并终止烹饪	0

(4) 向模型提问并获取答案

将视觉图与提示信息 INFO_ALL 作为模型输入，并询问当前烹饪状态下建议采取的烹饪操作。获取模型输出的答案，用于指导用户执行烹饪任务。

在上述实验方法的基础上，智能烹饪系统被设定为每隔 30 s 生成一次烹饪动作指令。然而，生成的自然语言信息长度存在差异，可能导致烹饪动作指令的生成时间间隔超过 30 s。针对这一问题，在实际实验过程中，若超过 30 s 仍未生成新的烹饪动作指令，系统会自动将烹饪温度调整至最低状态，以避免高温持续加热导致的牛排过度烹饪或烧焦，

同时等待指令的生成。

4.3.2 烹饪方法的有效性

微调前后智能烹饪系统在牛排烹饪过程中生成的操作指令如表2所示,在实验过程中,厨师严格执行烹饪系统生成的指令。实验结果表明,微调前系统在牛排已明显焦糊的情况下,仍未生成烹饪终止指令,导致厨师不得不在第13次操作后提前停止进一步烹饪动作;微调后系统能够在适当时刻生成烹饪终止指令,确保牛排达到理想熟度。

大语言模型微调前后牛排的烹饪结果对比如图10所示,微调前系统烹饪的牛排存在明显焦糊现象,而微调后系统生成的指令更精准,指导完成的牛排熟度更理想。此外,图11展示了微调后系统在煎牛排过程中每隔15 s采集的图像序列。从初期的生牛排到中期的轻微焦化,再到后期的均匀熟度,图像序列直观展示了系统在不同时间点生成的烹饪指令对食材外观的精准控制和优化效果。

4.3.3 烹饪模型的理解能力

为了验证模型的理解能力,本文设计了一种实验方法,通过实际烹饪过程中采集的多模态图像,以自然语言生成烹饪动作指令,对多模态大语言模

表2 微调前后智能烹饪系统在牛排烹饪过程中生成的操作指令

操作次序	微调前	微调后
第1次操作	3	3
第2次操作	3	3
第3次操作	4	2
第4次操作	4	2
第5次操作	3	2
第6次操作	2	2
第7次操作	4	2
第8次操作	2	2
第9次操作	2	2
第10次操作	3	0
第11次操作	3	—
第12次操作	3	—
第13次操作	3	—

型的理解能力进行可视化验证。实验具体场景如表3所示,选取牛排烹饪过程中的前期、中期和后期3个关键时间节点,分别使用对应的热力分布图、音量时序信号图和视觉图对模型的理解能力进行测试与评估。

在牛排智能烹饪理解能力实验中,热力分布图



(a) 微调前烹饪结果



(b) 微调后烹饪结果

图10 大语言模型微调前后的牛排烹饪结果对比

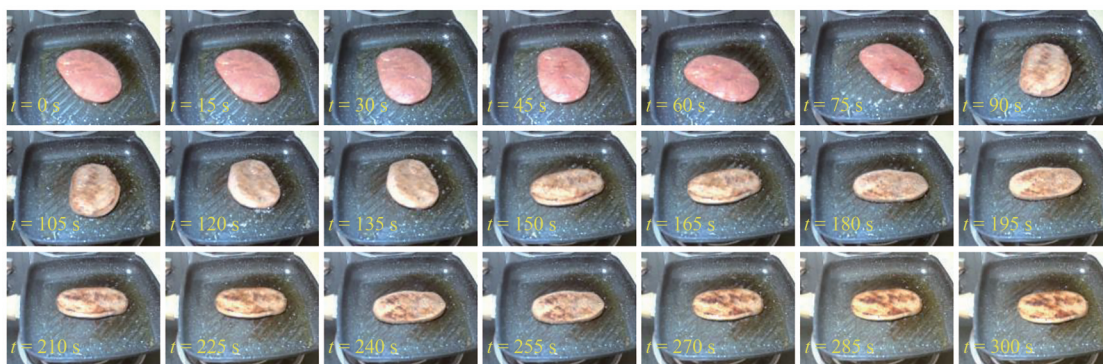


图11 烹饪实验中牛排外观的变化过程

表3 牛排智能烹饪理解力实验场景设置

场景序号	热力分布图	音量时序信号图	视图图
实验场景1			
实验场景2			
实验场景3			

和音量时序信号图的提取和处理方式与实际烹饪实验保持一致,旨在生成自然语言形式的热力分布信息和音量时序信号信息。在处理完成后,将视觉图像输入多模态大语言模型,并将生成的热力分布信息和音量时序信号信息作为提示信息输入模型中。同时,提示信息还包括当前牛排烹饪的时间戳、历史烹饪动作记录、牛排烹饪常识以及数字编码形式的烹饪动作指令集等。由于多模态大语言模型 LLaVA-v1.5-13b 对英文的理解能力较强,本文实验中的所有文本信息均采用英文描述。针对实验设置的3个场景,微调前后模型生成的自然语言形式的烹饪动作指令如表4所示。

4.4 实验结果分析

本节对4.3节的实验结果进行讨论,进一步分析本文所提智能烹饪方法的性能。从表2的实验结果可以看出,微调后的智能烹饪系统在实际烹饪过程中表现出更符合厨师操作习惯的特征。首先,微调后系统在前两次操作中采用中火加热,随后自第3次操作起持续使用小火,整个烹饪过程中仅发生一次温度切换,显著减少了频繁调温的情况。相比之下,微调前系统在第5~8次操作期间(约90 s内)频繁调整烹饪温度,平均每30 s调整一次烹饪温度,这种模式明显不符合厨师的操作规律。然

后,微调后系统在翻面频率上更加合理,约每30 s翻面一次,确保牛排两面受热均匀、成熟度一致;而微调前系统在相应时间段内未执行翻面操作,导致单侧持续受热,最终该侧出现焦糊现象。最后,在食材熟度判断方面,微调后系统能够在牛排达到理想熟度时及时生成烹饪终止指令,避免了过度烹饪,使成品更接近理想的口感和品质。这些表现表明,微调后的智能烹饪系统在温度管理、翻面间隔及烹饪时机选择上更符合厨师的经验与操作习惯,进一步验证了多模态大语言模型在烹饪过程指导中的有效性和实用性。

在实际烹饪过程中,微调后的智能烹饪系统的菜品品质更高。如图10(a)所示,微调前系统烹饪的牛排表面呈深褐色,出现过度上色甚至焦糊现象,视觉效果不理想;而微调后系统烹饪的牛排如图10(b)所示,表面颜色均匀,呈金黄色,更符合食客的期望。图11进一步展示了微调后系统在烹饪过程中牛排外观的均匀变化过程,实验结果表明,两面受热一致,颜色由红色逐渐变为浅褐色,最终达到理想的金黄色,并及时结束烹饪过程,从而获得优质牛排。从图10和图11可以看出,微调后的智能烹饪系统能够更好地完成烹饪任务,产出符合食客对色泽、口感及熟度需求的高质量牛排。此

表 4 牛排智能烹饪理解能力的实验结果

场景序号	模型	烹饪动作指令
实验场景 1	微调前模型	Based on the visual information in the image, the steak is being cooked on a grill or stovetop at a temperature of 120 degrees Fahrenheit. The sound variation information during the past 10 seconds shows a gradual increase in the sound wave, indicating that the cooking process is progressing. As a professional chef, the following steps should be taken to cook the steak at this moment: 1. Check the steak's color and texture: Since the steak is being cooked at a high temperature, it should have a seared exterior and a juicy, pink interior. If the steak is not cooked to the desired level, adjust the heat accordingly. 2. Flip the steak: Since the steak is being cooked on a grill or stovetop, it is essential to flip it halfway through the cooking process to ensure even cooking and prevent the formation of a tough, overcooked exterior. 3. Monitor the cooking process: Keep an eye on the steak's progress, adjusting the heat as needed to maintain the desired temperature and cooking time. 4. Remove the steak from the grill or stovetop: Once the steak is cooked to the desired level, carefully remove it from the grill or stovetop using a pair of tongs or a spatula. 5. Rest the steak: Allow the steak to rest for a few minutes before slicing and serving, as this will help the juices redistribute throughout the meat, resulting in a more flavorful and tender dish. By following these steps, the steak will be cooked to perfection, resulting in a delicious and satisfying meal.
	微调后模型	As a professional chef, the current action should be to flip the steak so that the other side can also be seared. Flipping the steak after it has been seared to a certain degree is a standard practice to ensure even cooking on both sides. After flipping, continue to sear the steak at the current temperature and observe the sound of the cooking process. If the sound of the steak cooking changes significantly, it may be necessary to adjust the heat. If the sound is too loud, indicating that the steak is searing too quickly, it's advisable to lower the heat slightly to prevent the outer layer from burning while ensuring that the interior is cooked to the desired level. Conversely, if the sound is too soft, indicating that the steak is searing too slowly, you may need to increase the heat slightly to ensure that the interior of the steak is thoroughly cooked.
实验场景 2	微调前模型	Based on the visual information in the image, the steak is being cooked on a grill or a stovetop, and the sound variation information suggests that the sound has not significantly changed during the past 10 seconds. The temperature information indicates that the steak is being cooked at 120 degrees Celsius. As a professional chef, the following steps should be taken to cook the steak at this moment: 1. Check the steak's color and texture: Since the steak is being cooked at 120 degrees Celsius, it should be cooked to a medium-rare level. The steak should have a slightly pink center and a slightly charred exterior. 2. Flip the steak: If the steak is not yet cooked to the desired level, carefully flip it over to ensure even cooking. 3. Monitor the cooking time: Keep an eye on the cooking time to ensure that the steak is cooked to the desired level without overcooking it. 4. Remove the steak from the grill or stovetop: Once the steak is cooked to the desired level, carefully remove it from the grill or stovetop using a pair of tongs or a spatula. 5. Rest the steak: Allow the steak to rest for a few minutes before slicing and serving, as this will help the juices redistribute and enhance the flavor of the steak. By following these steps, the steak will be cooked to the desired level, ensuring a delicious and well-prepared dish.
	微调后模型	Based on the visual information in the image, the steak has just started cooking, and the surface is still the original color. This indicates that the steak has not undergone any significant searing on this side yet. As a professional chef, at this point, you should continue cooking the steak and wait for the surface to develop a golden-brown sear before flipping it. The heat should be kept at medium to medium-high to ensure that the steak sears on the outside while the interior gradually heats up. Once one side of the steak has developed a beautiful sear, you can flip it to cook the other side to your desired level of doneness.
实验场景 3	微调前模型	Based on the visual information in the image, the steak is being cooked on a grill or frying pan at a high temperature of 180 degrees Celsius. As a professional chef, the following steps should be taken to cook the steak at this moment: 1. Monitor the cooking process closely, as the high temperature can cause the steak to cook quickly and potentially overcook or burn if not attended to. 2. Flip the steak occasionally to ensure even cooking and prevent the formation of a crust on one side while the other side remains undercooked. 3. Use a meat thermometer to check the internal temperature of the steak, ensuring it reaches the desired level of doneness (rare, medium, or well-done). 4. Once the steak is cooked to the desired level, remove it from the grill or frying pan and let it rest for a few minutes before slicing and serving. The sound variation information during the past 10 seconds suggests that the cooking process is progressing, which means that the steak is likely approaching the desired level of doneness. The temperature information at the current moment indicates that the steak is being cooked at a high temperature, which is appropriate for achieving the desired level of doneness.
	微调后模型	Based on the visual information in the image, it can be seen that one side of the steak has been seared to a certain degree, and the surface color is relatively uniform, indicating that the cooking process has been even. At this point, it's time to flip the steak to sear the other side. Regarding the heat, since the steak has been seared to some extent, it's advisable to maintain the current heat level of 125 degrees Celsius to ensure that the other side of the steak is also seared evenly.

外，从表 4 的结果可以看出，微调后模型生成的自然语言烹饪指令更具针对性与可操作性。微调后模型明确指出了具体的烹饪动作（如何时翻面、何时调整火力），使厨师或机器人能够直接依据指令执行相应操作；而微调前模型仅提供笼统的建议和注

意事项，缺乏明确的操作指令，难以满足实际应用需求。

5 结束语

本文提出一种基于多模态大语言模型的智能烹

饪方法,并在此基础上搭建了软硬件一体化的数字厨师。首先,在大语言模型微调的前置阶段,本文构建了多模态烹饪专家知识库,通过多种传感器记录专业厨师在烹饪过程中的连续操作,借助传感器收集视觉、听觉和温度等多模态数据,并将其转化为逐帧图像及对应的多轮问答文本,为微调模型提供了丰富的知识基础。随后,本文采用LoRA方法对LLaVA多模态大语言模型进行微调,使其能够更好地适应烹饪任务中复杂的多模态信息处理需求,从而完成智能烹饪系统的离线学习过程。当在线应用智能烹饪系统时,通过实时传感器采集烹饪过程数据(包括视觉、听觉和温度信息),将这些数据转化为图文信息,并以固定频率输入微调后的LLaVA模型中。模型分析输入数据后生成自然语言形式的烹饪指令,实时指导用户完成烹饪操作,直至菜肴制作完成。本文以牛排烹饪为例,验证了本文开发的智能烹饪软硬件系统在实际烹饪过程中的有效性,并展示了其如何在烹饪任务中智能地辅助用户,提升烹饪体验。

本文提出的智能烹饪方法具有以下潜在的发展方向。首先,应进一步提高智能烹饪系统的自动化集成程度。本文方法生成的烹饪指令以文字形式呈现,且对应的烹饪操作是标准化动作。这一特点为智能烹饪方法与自动控制硬件(如柔性机械臂)结合奠定了基础。未来可将标准化控制指令与柔性机械臂集成,最终实现数字厨师从指令生成到实际操作的全流程无人化烹饪。

然后,应进一步提升多模态数据采集的维度与精度。本文搭建的智能烹饪系统通过多模态传感器采集温度、听觉和视觉等数据,为烹饪过程提供了基础感知能力。然而,现有感知渠道未能全面记录人类烹饪专家的微观动作(如手部细微动作)和烹饪食材的细微状态变化(如食材的纹理变化、气味释放等),导致人类资深烹饪专家的经验无法被完整继承到多模态烹饪专家知识库中,进而可能造成烹饪品质的偏差。

最后,应重视智能烹饪系统执行机构的仿生设计,以确保烹饪系统的工作空间能够充分覆盖人类厨师在烹饪过程中的动作空间。目前,市场上的智能烹饪机器人主要集中在简单烹饪操作上,如牛排翻煎。然而,人类厨师具备更高的灵活性和精准度,尤其在复杂烹饪任务中表现突出。因此,未来智能烹饪系统应重点关注微观动作的精准捕捉,并

在硬件设计中实现这些动作的高度复现。这一研究方向将有助于更真实地模仿人类的关节运动和手部细节,从而实现高水平的烹饪动作再现,尤其是在食材处理的精准控制和灵活翻动等微小动作的复制上。

通过上述3个方面的技术升级,未来的智能烹饪系统有望实现更高水平的自动化与智能化,具备更强的适应性与灵活性,最终实现从文字指令到烹饪操作的全程无人化控制,进一步提升烹饪过程的效率和品质。

参考文献:

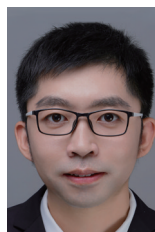
- [1] 陈涵,姜慧.我国社区餐饮发展现状、问题与对策[J].商业经济研究,2023(11):30-33.
CHEN H,JIANG H. Present situation, problems and countermeasures of community catering development in China[J]. Journal of Commercial Economics, 2023(11): 30-33.
- [2] 杨铭铎,陈健.对本科烹饪专业建设几个问题的探讨[J].美食研究,2022,39(3):45-50.
YANG M D, CHEN J. Discussion on several practical problems in the construction of undergraduate culinary major[J]. Journal of Researches on Dietetic Science and Culture, 2022, 39(3): 45-50.
- [3] 左武举.烹饪机器人与智能烹饪市场[J].现代食品,2022,28(9):83-85.
ZUO W J. Cooking robots and smart cooking market[J]. Modern Food, 2022, 28(9): 83-85.
- [4] 雷鸣,管声启,王杰,等.新型商用炒菜机器人搅拌系统结构与轨迹分析[J].西安工程大学学报,2019,33(6):660-665.
LEI M, GUAN S Q, WANG J, et al. Structural design and trajectory analysis of a new commercial cooking robot mixing system[J]. Journal of Xi'an Polytechnic University, 2019, 33(6): 660-665.
- [5] 邓鹏,唐文涛,罗静.机器人大模型发展与挑战[J].电子测量与仪器学报,已录用.
DENG P, TANG W T, LUO J. Robotic large model development and challenges[J]. Journal of Electronic Measurement and Instrumentation, accepted.
- [6] 王婉娇.中国传统饮食文化的发展及当代价值[J].食品与机械,2024,40(6):241.
WANG W J. The development and contemporary value of traditional diet culture in China[J]. Food and Machinery, 2024, 40(6): 241.
- [7] 青平,王玉泽,李剑,等.大食物观与国民营养健康[J].农业经济问题,2023,44(5):61-73.
QING P, WANG Y Z, LI J, et al. The greater food approach and national nutritional health[J]. Issues in Agricultural Economy, 2023, 44(5): 61-73.
- [8] 李程,常静一,朱婧.食欲的评价方法[J].卫生研究,2023,52(5):832-836.
LI C, CHANG J Y, ZHU J. Evaluation method of appetite[J]. Journal of Hygiene Research, 2023, 52(5): 832-836.
- [9] 赵海,王路,滕立新,等.北京市大中中学生抑郁流行现状及与饮食相关行为的关系[J].中国学校卫生,2022,43(5):715-717.
ZHAO H, WANG L, TENG L X, et al. Prevalence of depression and the association between depressive symptoms and dietary behaviors among students in Beijing[J]. Chinese Journal of School Health, 2022,

- 43(5): 715-717.
- [10] 薛羽, 张逸轩. 深层神经网络架构搜索综述[J]. 信息安全, 2023, 23(9): 58-74.
XUE Y, ZHANG Y X. Survey on Deep Neural Architecture Search[J]. Netinfo Security, 2023, 23(9): 58-74.
- [11] 张淼, 贾洪峰, 张振宇, 等. 我国牛排品质识别技术研究现状[J]. 食品工业, 2023, 44(2): 175-180.
ZHANG M, JIA H F, ZHANG Z Y, et al. Research situation on recognition technology of steak in China[J]. The Food Industry, 2023, 44(2): 175-180.
- [12] 姚力为, 邓静, 易宇文, 等. 烹饪熟度对牛排挥发性风味物质的影响研究[J]. 中国调味品, 2024, 49(9): 47-53.
YAO L W, DENG J, YI Y W, et al. Study on effect of degree of doneness on volatile flavor substances of steak[J]. China Condiment, 2024, 49(9): 47-53.
- [13] 季鸿崑. 饮食美感和饮食风味[J]. 扬州大学烹饪学报, 2005, 22(1): 11-16.
JI H K. Dietetic aesthetic perception and dietetic flavor[J]. Cuisine Journal of Yangzhou University, 2005, 22(1): 11-16.
- [14] 陈梦晓. 雕塑艺术在食品造型中的应用研究[J]. 食品研究与开发, 2020, 41(6): 2.
CHEN M X. Research on the application of sculpture in food modeling[J]. Food Research and Development, 2020, 41(6): 2.
- [15] MORQUECHO-CAMPOS P, GRAAF K D, BOESVELDT S. Smelling our appetite? The influence of food odors on congruent appetite, food preferences and intake[J]. Food Quality and Preference, 2020, 85: 103959.
- [16] 侯兵. 新时代我国烹饪高等教育发展审视: 成就、挑战与展望: 基于扬州大学烹饪与营养教育专业办学40周年的思考[J]. 美食研究, 2023, 40(1): 1-10.
HOU B. Research on the development of Chinese higher culinary education in the new era: achievements, challenges and prospects: for the 40th anniversary of the founding of the school of tourism and cuisine of Yangzhou University[J]. Journal of Researches on Dietetic Science and Culture, 2023, 40(1): 1-10.
- [17] 李楠, 王昆仑, 王起赫, 等. 餐饮食品营养成分影响因素分析[J]. 中国食品卫生杂志, 2023, 35(8): 1192-1198.
LI N, WANG K L, WANG Q H, et al. Factors influencing the nutritional composition of catering food[J]. Chinese Journal of Food Hygiene, 2023, 35(8): 1192-1198.
- [18] SUHAIROM N, MUSTA'AMAL A H, MOHD A N F, et al. Quality culinary workforce competencies for sustainable career development among culinary professionals[J]. International Journal of Hospitality Management, 2019, 81: 205-220.
- [19] 吴晓蒙, 李想, 刘楠, 等. 智慧餐饮4.0: 融合人工智能的未来餐饮发展新方向[J]. 食品科学技术学报, 2024, 42(1): 10-19.
WU X M, LI X, LIU N, et al. Smart catering 4.0: new development direction of future catering integrated with artificial intelligence technology[J]. Journal of Food Science and Technology, 2024, 42(1): 10-19.
- [20] 高莹婷, 胡懿昕, 路奇. 基于电子鼻的智能厨房烹饪方式检测[J]. 计算机辅助设计与图形学学报, 2023, 35(2): 185-194.
GAO Y T, HU Y X, LU Q. Cooking method detection based on electronic nose in smart kitchen[J]. Journal of Computer-Aided Design & Computer Graphics, 2023, 35(2): 185-194.
- [21] LIU J J, CHEN Y T, DONG Z P, et al. Robot cooking with stir-fry: bi-manual non-prehensile manipulation of semi-fluid objects[J]. IEEE Robotics and Automation Letters, 2022, 7(2): 5159-5166.
- [22] 张腾超, 田永林, 林飞, 等. 平行旅游: 基础智能驱动的智慧出游服务[J]. 智能科学与技术学报, 2024, 6(2): 164-178.
ZHANG T C, TIAN Y L, LIN F, et al. Parallel tourism: foundation intelligence driven smart trip services[J]. Chinese Journal of Intelligent Science and Technology, 2024, 6(2): 164-178.
- [23] 田永林, 王兴霞, 王雨桐, 等. RAG-PHI: 检索增强生成驱动的平行人与平行智能[J]. 智能科学与技术学报, 2024, 6(1): 41-51.
TIAN Y L, WANG X X, WANG Y T, et al. RAG-PHI: searching for parallel people and parallel intelligence driven by enhanced generation[J]. Chinese Journal of Intelligent Science and Technology, 2024, 6(1): 41-51.
- [24] LIU H T, LI C Y, WU Q Y, et al. Visual instruction tuning[J]. Advances in Neural Information Processing Systems, 2024, 36: 1-25.
- [25] 李耕, 王梓烁, 何相腾, 等. 从ChatGPT到多模态大模型: 现状与未来[J]. 中国科学基金, 2023, 37(5): 724-734.
LI G, WANG Z S, HE X T, et al. From ChatGPT to large multimodal model: present and future[J]. Bulletin of National Natural Science Foundation of China, 2023, 37(5): 724-734.
- [26] JIANG Y, YAN X Y, JI G P, et al. Effectiveness assessment of recent large vision-language models[J]. Visual Intelligence, 2024, 2(1): 1-17.
- [27] HU E J, SHEN Y, WALLIS P, et al. LoRA: Low-rank adaptation of large language models[EB]. arXiv preprint, 2021, arXiv: 2106.09685.

[作者简介]



李鑫源 (1999-), 男, 湖南大学机械与运载工程学院硕士生, 主要研究方向为大语言模型、具身智能、柔性机械臂建模与控制。



李柏 (1989-), 男, 博士, 湖南大学机械与运载工程学院副教授, 主要研究方向为自主无人系统运动规划、基于智能车竞赛的新工科教育、计算最优控制理论、柔性机械臂运动控制以及智能烹饪系统。



孙跃硕 (2002-), 男, 湖南大学机械与运载工程学院硕士生, 主要研究方向为自主无人系统决策规划、数值优化以及具身智能。



张坦探 (1990-), 男, 博士, 湖南大学机械与运载工程学院副教授, 主要研究方向为智能控制、自动驾驶以及人工智能。



田永林 (1994-), 男, 博士, 中国科学院自动化研究所多模态人工智能系统国家重点实验室助理研究员, 主要研究方向为平行系统、自动驾驶以及场景工程。



王飞跃 (1961-), 男, 博士, 中国科学院自动化研究所复杂系统管理与控制国家重点实验室主任, 澳门科技大学特聘教授, 主要研究方向为平行系统的方法与应用、社会计算、平行智能、知识自动化。



殷焯炎 (1995-), 女, 湖南大学机械与运载工程学院科研助理, 主要研究方向为运筹优化、智能控制以及多模态大语言模型。