分类号 _____ 密级 _____

重庆邮电大学研究生学位论文
中期考核报告

中文题目	面向车载嵌入式设备的 智能语音对话方法研究
英文题目	**Research on Intelligent Voice Dialogue Method for Vehicle Embedded Devices**
学　　号	**S200303078**
姓　　名	黄子恒
学位层次	硕士研究生
学位类别	工学硕士
学科专业	控制科学与工程
研究方向	神经网络理论及应用
指导教师	李鹏华　教授
完成日期	**2023 年 3 月 4 日**

<table>
<tr><td colspan="8">一、研究生简况</td></tr>
<tr><td>姓名</td><td>黄子恒</td><td>性别</td><td>男</td><td>年龄</td><td>24</td><td>专业类型</td><td>学术型</td></tr>
<tr><td colspan="2">导师
/副导师或联系人</td><td colspan="2">李鹏华</td><td colspan="2">所在团队名称
团队负责人</td><td colspan="2">汽车电子与嵌入式系统理论、技术及应用重庆高校创新团队
团队负责人：冯辉宗</td></tr>
<tr><td colspan="2">外出实习单位</td><td colspan="2">无</td><td colspan="2">外出实习时间</td><td colspan="2">无</td></tr>
<tr><td colspan="2">开题时间</td><td colspan="2">2022.1.8</td><td colspan="2">开题状态</td><td colspan="2">正常</td></tr>
<tr><td colspan="2">学籍异动情况</td><td colspan="6">无</td></tr>
<tr><td colspan="8">二、学位论文工作进展自查</td></tr>
<tr><td>分项</td><td colspan="3">开题报告内容</td><td colspan="4">中期完成情况</td></tr>
<tr><td>论文题目</td><td colspan="3">面向车载嵌入式设备的智能语音对话方法研究</td><td colspan="4">面向车载嵌入式设备的智能语音对话方法研究</td></tr>
<tr><td>研究目标</td><td colspan="3">（1）针对基于深度编—解码器的自动语音识别模型参数量庞大的问题，探索基于残差分组线性变换的解码器结构。给出残差分组线性变换的具体形式，建立与注意力机制的拓扑关系，降低模型的参数量和计算复杂度。

（2）针对基于显式联合建模的自然语言理解模型泛化能力不足的问题，探索基于层级交互的全局—局部交互网络。给出全局—局部交互网络的具体结构，设计高效的层级信息融合算法，提高模型的泛化能力和预测精度。

（3）针对“云—端”方式运行的车载智能语音对话存在数据安全隐患的问题，探索面向车载嵌入式设备的本地语音</td><td colspan="4">（1）提出了一种基于残差分组线性变换的解码器结构，并将其嵌入原始的编—解码器网络中，进而大幅度降低模型的参数量和计算复杂度。通过上述算法构建的网络模型在公开中文数据集 AISHELL-1 上参数量和计算量分别为 20.4M 和 5.3B，并达到 6.67%的字错误率，与对比方法中准确率最高的 HA-Transformer 相比（5.96%），字错误率略高，但参数量和计算量分别相对下降了 48.35%和 77.16%。在引入语言模型后，字错误率可进一步降低至 6.38%。

（2）提出了一种基于标签感知的图交互模型，主要由标签映射模块和全局图交互模块构成，前者可以获取原始话语与标签语义之间的相关性以提供丰富的先验知识，后者可以对语句级别的意图—槽位交互过程进行建模以提供全局优化，进而提高模型的交互能力和预测精度。通过上述算法构建的网络模型在公开数据集 MixATIS 和 MixSnips 上分别实现了 49.9%和 77.3%的整体准确率。</td></tr>
</table>

	对话系统集成与验证。针对嵌入式设备进行模型训练、移植和集成，提出面向实际应用场景的适配性优化方案，搭建离线运行的车载智能语音对话平台。	（3）搭建了面向车载嵌入式设备的本地智能语音对话系统。具体地，首先选取 Nvidia Jetson TX2 作为车载嵌入式设备并进行刷机、配置环境等操作，然后根据实际应用场景收集、创建驾驶数据集，接着将研究内容（1）和（2）的模型在驾驶数据集上进行训练，最后集成、移植网络模型至 TX2 并围绕搭建全套硬件平台，进而实现数据安全、自然实时的离线智能语音对话。通过上述方法搭建的系统在利龙项目的验收环节达到了 97%的通过率，平均响应时间为 0.87±0.12s。	
研究内容	**（1）基于残差分组线性变换解码器的自动语音识别** 研究基于残差分组线性变换的解码器结构，首先引入残差分组线性变换方法，设计含有扩张和收缩操作的缩放单元，然后建立逐块缩放机制并将缩放单元放置于解码器自注意力模块前，最后耦合词嵌入层（Embedding Layer）和前馈网络（Feedforward Network, FFN），实现自动语音识别模型参数量和计算复杂度的降低。 **（2）基于全局—局部交互网络的自然语言理解** 研究基于层级交互的全局—局部交互网络，首先构造局部槽位交互层建立槽位自交互通道，其次构造全局意图—槽位交互层建立意图和槽位的双向交互通道，最后设计层级信息融合算法，高效聚合所有交互信息，实现自然语言	已完成研究内容（1）的算法设计和相关实验。 提出了一种基于残差分组线性变换的解码器结构，该结构关键模块为“钻石”型缩放单元，其内部采用稀疏连接，同一组神经元共享相同的权重矩阵，实现模型参数量和计算复杂度的降低。在 AISHELL-1 上实现了 6.67%的字错误率以及 20.4M 的参数量和 5.3B 的计算量，三项评价指标均优于对比方法中的 Speech-Transformer 和 STBD，与对比方法中准确率最高的 HA-Transformer 相比（5.96%），字错误率略高，但参数量和计算量分别相对下降了 48.35%和 77.16%。在引入语言模型后，字错误率可进一步降低至 6.38%。 已完成研究内容（2）的算法设计和相关实验。 提出了一种基于标签感知的图交互模型，主要由标签映	完成百分比 1.100% 2.100% 3.95%

	理解模型泛化能力和预测精度的提升。 **（3）面向车载嵌入式设备的智能语音对话系统** 研究面向车载嵌入式设备的智能语音对话系统，首先根据车身结构选择嵌入式设备并配置相应的软件环境，其次面向企业和用户需求构建特定场景下的语音、文本数据集，接着训练适合特定场景的自动语音识别、自然语言理解等模型，依据车载嵌入式设备的特性进行模型移植与集成，最后根据实际使用情形提出适配性优化方案并搭建硬件平台，实现数据安全、自然实时的离线智能语音对话。	射模块和全局图交互模块构成。具体地，基于最佳线性逼近构建标签映射模块，将意图标签特征自适应地融合到话语特征中，以增强模型的表征能力；然后基于图注意力网络搭建全局图交互模块，对语句级别的意图—槽位交互过程进行建模以提供全局优化，实现模型交互能力和预测精度的提升。在公开数据集 MixATIS 和 MixSnips 上分别达到了 49.9%和 77.3%的整体准确率，相较于对比方法中性能最好的 SDJN，分别提高了 5.3%和 1.6%。 已完成研究内容（3）的系统搭建和相关实验。部分数据结果待整理。 搭建了面向车载嵌入式设备的本地智能语音对话系统。具体地，首先选取 Nvidia Jetson TX2 作为嵌入式设备并进行刷机等操作；然后根据实际应用场景，开发硬件、软件平台用于收集驾驶数据集，经过清洗、标注后得到近 52 小时、共包含 57938 条音频的驾驶数据集；接着将研究内容（1）和（2）的网络模型在驾驶数据集上进行训练，分别达到了 7.26%的字错误率和 95.76%的整体准确率；最后将训练好的网络模型集成、移植至 TX2 并围绕搭建全套硬件平台，实现数	

		据安全、自然实时的离线智能语音对话。在利龙项目的验收环节达到了 97%的通过率，平均响应时间为 0.87±0.12s。	
创新	（1）针对基于深度编—解码器的自动语音识别模型参数量庞大的问题，在解码器中使用基于残差分组线性变换的缩放单元耦合各层，实现自动语音识别模型参数量和计算复杂度的降低； （2）针对基于显式联合建模的自然语言理解模型泛化能力不足的问题，搭建基于层级交互的全局—局部交互网络，设计层级信息融合算法，高效融合所有交互信息，实现自然语言理解模型泛化能力和预测精度的提升； （3）针对“云—端”方式运行的车载智能语音对话存在数据安全隐患的问题，搭建基于车载嵌入式设备的本地运行平台，提出面向实际应用场景的适配性优化方案，实现数据安全、自然实时的离线智能语音对话。	（1）针对基于深度编—解码器的自动语音识别模型参数量庞大的问题，提出了一种基于残差分组线性变换的解码器结构，该结构关键模块为“钻石”型缩放单元，其内部采用稀疏连接，同一组神经元共享相同的权重矩阵，进而实现了模型参数量和计算复杂度的降低。 （2）针对基于显式联合建模的自然语言理解模型交互能力不足的问题，提出了一种基于标签感知的图交互模型，其中标签感知模块可以获取原始话语与标签语义之间的相关性以提供丰富的先验知识，全局图交互模块可以对语句级别的意图—槽位交互过程进行建模以提供全局优化，进而提高了模型的交互能力和预测精度； （3）针对“云—端”方式运行的车载智能语音对话存在数据安全隐患的问题，搭建了面向车载嵌入式设备的本地智能语音对话系统。首先选取 Nvidia Jetson TX2 作为车载嵌入式设备并进行刷机、配置环境等操作，然后根据实际应用场景收集、创建驾驶数据集，接着将研究内容（1）和（2）的模型在驾驶数据集上进行训练，最后集成、移植网络模型至 TX2 并围绕搭建全套硬件平台，进而实现了数据安全、自然实时的离线智能语音对话。	
学术指标	（1）自动语音识别模型在 AISHELL-1 上字错误率不高于 8%，参数量不高于 25M，计算量不高于 10B；在 TED-LIUM2 上词错误率不高于 13%，参数量不高于 25M，计算量不高于 8B；	（1）自动语音识别模型在 AISHELL-1 上达到了 6.67%的字错误率及 20.4M 的参数量和 5.3B 的计算量；在 TED-LIUM2 上达到了 11.86%的词错误率及 20.3M 的参数量和 4.6B 的计算量；	

	（2）自然语言理解模型整体准确率在 MixATIS 和 MixSnips 上整体准确率分别不低于 45%和 75%； （3）面向车载嵌入式设备的智能语音对话系统通过率不低于 95%，平均响应时间不高于 1s。	（2）自然语言理解模型在 MixATIS 和 MixSnips 上分别达到了 49.9%和 77.3%的整体准确率； （3）面向车载嵌入式设备的本地智能语音对话系统通过率为 97%，平均响应时间为 0.87s。
成果指标	发表论文 1 篇，公开专利 1 件。	录用 EI 会议论文 1 篇，受理专利 1 件。
其他情况	实际采用的研究方法： 课题主要包含自动语音识别模型的构建、自然语言理解模型的构建和面向车载嵌入式设备的智能语音对话系统的搭建三部分。在自动语音识别模型方面，首先通过频谱增强技术对语音信号频谱特征进行增强，接着用堆叠的编码器实现高层次特征的提取，然后让标注文本的嵌入表示与编码器提取的特征通过解码器进行充分交互以得到输出序列，并与标注文本进行比较产生损失，通过反向传播更新网络权值，不断迭代上述过程，最终得到训练好的模型；在自然语言理解模型方面，包含意图识别和槽位填充两个子任务，首先利用共享编码器得到话语表示，然后在标签感知模块和意图解码器的作用下得到意图序列，接着将意图序列送入全局图交互模块得到槽位序列，定义联合训练损失以更新网络权值，不断迭代上述过程，最终得到训练好的模型；在面向车载嵌入式设备的智能语音对话系统方面，首先选取 Nvidia Jetson TX2 作为车载嵌入式设备并进行刷机、配置环境等操作，然后根据实际应用场景收集、创建驾驶数据集，接着将上述两个模型在驾驶数据集上进行训练，最后将训练好的模型集成、移植至 TX2 并围绕搭建硬件平台，实现数据安全、自然实时的离线智能语音对话。	
	实际执行的进度安排： 2021.12~2022.01 阅读相关文献，整理相关理论依据并与导师保持交流。完成文献整理工作，撰写综述。 2022.03~2022.04 收集、处理所需数据集。完成数据集整理工作。 2022.05~2022.08 设计自动语音识别模型并在数据集上训练，进行模型评估与优化。完成自动语音识别模型训练与优化。 2022.09~2022.12 设计自然语言理解模型并在数据集上训练，进行模型评估与优化。完成自然语言理解模型训练与优化。	

<table>
<tr><td></td><td colspan="7">2023.01~2023.03
搭建车载智能语音对话平台，撰写毕业论文。完成平台搭建和论文撰写。
2023.04~2023.05
论文定稿，准备答辩。</td></tr>
<tr><td></td><td colspan="7">中期考核时提交的材料清单（含系统演示文档、发表成果等）：
系统演示文档一份
论文一篇
专利一篇</td></tr>
<tr><td rowspan="4">工作小结</td><td colspan="2">是否按开题报告执行</td><td colspan="5">是</td></tr>
<tr><td colspan="2">对变更开题内容的说明</td><td colspan="5">无</td></tr>
<tr><td colspan="2">工作进展情况说明</td><td colspan="5">按计划进行</td></tr>
<tr><td colspan="2">目前存在的问题和举措</td><td colspan="5">暂无</td></tr>
<tr><td colspan="8">三、毕业条件自查</td></tr>
<tr><td rowspan="2">课程学分情况</td><td colspan="2">学位课</td><td>18</td><td colspan="5" rowspan="2">是否完成规定学分？如无，说明原因及补救措施：
是</td></tr>
<tr><td colspan="2">非学位课</td><td>16</td></tr>
<tr><td rowspan="4">成果情况</td><td>排名</td><td>类型(论文、专利、获奖等)</td><td colspan="2">内容
(按文献格式)</td><td>级别</td><td>状态
(发表/录用/已投/计划中)</td><td>是否论文工作相关</td><td>可否获得毕业资格</td></tr>
<tr><td>2</td><td>论文</td><td colspan="2">LAGIM: A Label-Aware Graph Interaction Model for Joint Multiple Intent Detection and Slot Filling, CCDC 2023</td><td>EI</td><td>录用</td><td>是</td><td>是</td></tr>
<tr><td>2</td><td>专利</td><td colspan="2">李鹏华, 黄子恒, 张奕辉, 谢潇, 刘学超, 唐培渊. 一种基于全局—局部对比学习的跨语言自然语言理解方法</td><td></td><td>已受理</td><td>是</td><td>是</td></tr>
<tr><td>其他情况</td><td colspan="7">如有其他和毕业相关的情况，在此说明：
无</td></tr>
</table>

四、学位论文写作计划

列出学位论文的目录计划，一般应给出全部章节的一级和二级标题（此部分对梳理学位论文工作并按期完成撰写很重要，务必在导师指导下撰写并提交）。

注：以上内容根据情况添页。

论文中期考核管理

版本说明	提交版本：首次提交。
学生承诺	本人提交的研究内容和材料属实。 学生签字：黄子恒　2023年3月3日
导师意见	同意中期答辩 导师签字：[illegible]　2023年3月3日

考核组成员	姓名	职称	所在单位
组长	[illegible]	教授	重庆邮电大学
成员	[illegible]	高级工程师	重庆邮电大学
成员	[illegible]	副教授	重庆邮电大学
成员	[illegible]	讲师	重庆邮电大学
成员	[illegible]	讲师	重庆邮电大学

考核组意见	同意通过中期考核。 组长签字：[illegible]　2023年3月5日
导师团队意见	同意通过 负责人：[illegible]　2023年3月6日
学院意见	同意　魏旻 院长：[illegible]　2023年3月27日 重庆邮电大学 自动化学院

备注：本页单独打印在一页。