附件7：

**东北师范大学本科生毕业论文开题报告**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **学生姓名** | 潘仲秋 | **学号** | 2017012838 | | **专业** | 软件工程 |
| **指导教师** | 杨喜权 | **职称** | 副教授 | | | |
| **审查小组** | **组长：杨喜权 成员：**肖永鹏，封子军，朱铁英 | | | | | |
| **论文题目** | 基于深度学习的视障人士视听转换app | | | | | |
| **论文类型** | 研究性论文 | | | **选题来源** | 应用实践 | |
| **一、选题依据、研究目标与内容、研究意义，研究方法和手段**  1、选题依据：  **学术价值：**   1. 随着人工智能技术的兴起，图像特征提取技术和自然语言处理技术都得到了长足的进步，本系统使用的图像语义识别技术将两者结合进行结合，利用模型去理解图片的内容然后用自然语言去表达它们之间的关系，并最终将自然语言转化为语音。 2. 图像语义识别技术目前主要运用于为视频生成摘要、为未标注图像加标题和图片审核等途径，本系统创新的利用图片描述技术为视障人士服务，用户输入的图片或视频可转化为语音描述传给用户，让盲人“看到”世界。   **应用价值：**   1. 图像语义识别——帮盲人“看到“”世界 2. 出行辅助——包括红绿灯识别，公交车牌号识别，以及盲道筛选等服务 3. 学习辅助——识别黑板板书和非盲文书本内容 4. 生活辅助——识别物体和志愿者热线   **国内外研究现状分析：**   1. 国内：国内相关AP已有的视障人士辅助APP多为辅助阅读类，如盲人听书等，少数的识物类APP，如天使眼，以及志愿者服务APP启明瞳。但仍无法满足视障者出行的需求，即时性图像信息获取的要求，也不能帮助视障者更好的了解世界。   国外：AR结合VR技术的Sightplus：可以在手机头盔佩戴使用，并使用操作手柄进行类似上面  视频调控放大倍率焦距。盲用高科技智能导航眼镜：该眼镜可将视觉信号转化为听觉信号，为视  障人士出行和感知世界提供便利。虽然国外的视障人士生活辅助技术更为成熟，但仍需要硬件支  持，而且费用较昂贵，不利于普及。  2、研究目标与内容：  **目标**：通过本课题的研究，开发出一个基于深度学习技术，可将用户输入的图片或视频进行理解，  并转化为自然语言描述，再进一步转化为语音输出给用户，让盲人“看到”世界。  **内容**：  APP：  1.通过JavaMVC模式链接MySQL数据库，实现后台数据和逻辑处理。  2.通过vue.js实现跨平台APP开发。  算法：  采用深度学习的图像语义识别方法，实现图像转自然语言，并最终转化为语音。   1. 研究意义：   市场已有的视障人士辅助APP多为辅助阅读类，如盲人听书等，少数的识物类APP，如天使眼，  以及志愿者服务APP启明瞳。相比较其他项目，本项目创新的将图像语义识别技术运用于视障人士辅  助，通过识别用户输入的图片，将其内容进行识别，并将其语义转化为语音进行输出。可用于图像  描述——帮盲人“看到“世界，出行辅助——包括红绿灯识别，公交车牌号识别，以及盲道筛选等  服务，学习辅助——识别黑板板书和非盲文书本内容，生活辅助——识别物体和志愿者热线。  本项目融合了市场现有视障人士辅助APP的功能，且主打的图像语义识别功能更可以帮助视障人士更  好的了解世界，“看到”世界。  4、研究方法和手段：  具体步骤：  1.App构建思路：   1. 前端：采用vue全家桶系列（vue-router，vuex，vue-resource），再加上构建工具vue-cli以及组件库ant-design-vue，构建完整的app前端样式。采用 axios ，json等技术实现与后端的数据数据交互，实现前后端分离的开发模式。采用radius服务器，保证网络的安全。采用混合模式应用（Hybrid App）,基于HTML5低成本跨平台开发优势又兼具Native App特质的Hybrid App更加具有竞争力。 2. 后端：采用java实现，主要是java+tomcat+mysql，可以采用ssm（Spring+Springmvc+Mybatis）框架。处理的业务，主要包括处理前端的请求，对数据库进行操作，并将结果返回给前台，需要搭建服务器、有扎实java基础，具有网站搭建的后端(javaEE)体系的知识。 3. 功能实现：   前端调用手机相机API，让用户拍摄照片，返回到后台，经过算法处理转换成文字在转换为语音，返回到前端输出给用户。  2.算法实现：  核心技术：DCNN（Deep Convolutional Neural Network，深度卷积网络），LSTM（Long Short Term Memory，长短时记忆网络） i.图片转语音：  简单地将图像信号直接作为输入无法达到很好的效果，原因是原始的图像信号含  有太多的噪声。所以我们需要引入DL（Deep Learning，深度学习）在机器视觉中  最核心的部件：CNN。在DCNN的高层神经元输出可以表示图像的紧致的高层语  义信息，所以此图像文字描述方法的基本思想就是利用了DCNN生成图像的高层  抽象语义向量，将其作为语言生成模型LSTM的输入进行序列到序列  （sequence to sequence）的转换，本质上就是从一段不定长的序列转换为另一段  不定长的序列，实现sequence to sequence最有效的方法即为LSTM，一种带门的  RNN（Recurrent Neural Network，递归神经网络)，它可以将源语言编码为一个固  定长度含丰富语义的向量，然后作为解码网络的隐藏状态去生成目标语言，得到\  文本后交给APP系统进行转语音操作。 ii.数据集方面：  我们的核心功能需要解决的是两部分的问题，第一部分是图像转换，第二部分是  图像识别。对于图像转换部分，我们将使用mscoco、Flicker30k等数据集进行模  型训练，对于图像识别部分，我们将使用**ImageNet、CIFAR**等数据集进行模型训  练。  iii.训练步骤：  固定CNN参数，训练LSTM语言模型500K  CNN参数：在ImageNet数据集预训练(pre-trained)好的参数  训练拆分：一句话n个词 -> n-1组训练序列  细调CNN参数，CNN&LSTM一起训练100K | | | | | | |
| **二、论文提纲、进度安排**  1、论文提纲:  摘要 |  Abstract ||  目录  第一章 绪论  1.1研究背景和意义  1.2国内外研究现状  1.3研究内容与方法  第二章 图像语义识别技术  2.1图像语义识别定义  2.2图像语义识别的基本过程  2.3 图像语义识别的衡量指标  第三章 基于深度学习的图像语义识别算法  3.1 DCNN算法原理  3.2 LSTM算法原理  3.3数据集  3.3.1 mscoco数据集  3.3.2 ImageNet数据集  第四章 算法和APP实现  4.1视障人士视听转化系统  4.2图片语义识别  4.2.1利用DCNN实现图片转化为语义信息  4.2.2利用mscoco进行模型训练  4.3语义转化为语音输出  4.3.1利用LSTM实现语义转化为语音输出  4.3.2利用ImageNet进行模型训练  4.4训练步骤  4.4.1 固定CNN参数训练LSTM  4.4.2 细调CNN参数  4.5实验总结  第五章 工作总结与展望  5,1工作总结  5.2工作展望  参考文献  致谢  2、进度安排：  第一阶段：12月，确定选题，拟定开题报告，进行文献收集和分析、平台的安装学习。  第二阶段：1月，研究LSTM，DCNN算法原理，复现代码。  第三阶段：2月，按照技术路线训练模型，完成系统制作。  第四阶段：3月，结合文献资料与实验积累的结果撰写论文初稿。  第五阶段：4月，对论文初稿进行修正并与指导老师交流想法，不断改进论文。  第六阶段：5月，按学校关于毕业论文设计的具体要求对论文进行全面完善，完成论文。 | | | | | | |
| **三、参考文献**  [1]兜哥.AI安全之对抗样本入门[M]，机械工业出版社，2019.  [2]凌祥,纪守领,任奎.面向深度学习系统的对抗样本攻击与防御[J],CCF,2018,14(6):11-17.  [2]雷盛川. 针对图像识别的对抗样本防御系统设计与实现[D].北京邮电大学,2019.  [3]潘文雯,王新宇,宋明黎,陈纯.对抗样本生成技术综述[J/OL].软件学报:1-17[2019-12-24].  [4]周飞燕,金林鹏,董军.卷积神经网络研究综述[J].计算机学报,2017,40(06):1229-1251.  [5]陈超,齐峰.卷积神经网络的发展及其在计算机视觉领域中的应用综述[J].计算机科学，2019,  46(03):63-73.  [6]耿颖.使用Python语言的GUI可视化编程设计[J].单片机与嵌入式系统应用, 2019,19(02):  20-22+44. | | | | | | |
| **指导教师意见：**  **签字： 年 月 日** | | | **审查小组意见：**  **组长签字： 年 月 日** | | | |

注：1.选题类型：研究性论文、文献综述、调查报告、实验报告、毕业设计、毕业创作

2.选题来源：科研项目、社会实践、应用实践、专业实习、其它