華南師範大學
SOUTH CHINA NORMAL UNIVERSITY

2019—2020年度学生课外科研一般课题

结题报告

课题名称：基于深度学习的图像描述

第一作者：刘杰聪

课题组成员：李海宏、谢淦煜、赖乐贤

指导教师：梁军

所在院系：软件学院

结项日期：2020 年 4 月

共青团华南师范大学委员会制

2020.4

华南师范大学2019—2020年度学生课外科研一般课题

结题申请表

<table>
<tr><td>课题名称</td><td colspan="5">基于深度学习的图像描述</td></tr>
<tr><td>课题形式</td><td colspan="5">☐个人课题　　■集体课题</td></tr>
<tr><td>学科类别</td><td colspan="5">☐哲学　☐社会　☐法律　☐经济　☐管理　☐教育
☐政治　☐历史　☐文学　☐艺术　☐体育　☐数理
■信息技术　☐机械与控制　☐生命科学　☐能源化工</td></tr>
<tr><td>课题类别</td><td colspan="5">☐哲学社会科学类调查报告和学术论文
☐自然科学类学术论文
■科技发明制作类
☐创作成果</td></tr>
<tr><td rowspan="2">指导教师</td><td>姓名</td><td>职称</td><td colspan="2">学 院</td><td>联系电话</td></tr>
<tr><td>梁军</td><td>工程师</td><td colspan="2">软件学院</td><td>13632368330</td></tr>
<tr><td rowspan="9">作者简介</td><td rowspan="5">第一作者</td><td>姓名</td><td>刘杰聪</td><td>性别</td><td colspan="2">男</td></tr>
<tr><td>出生年月</td><td>1999年1月</td><td>学历</td><td colspan="2">本科</td></tr>
<tr><td>学院年级</td><td colspan="4">软件学院2018级</td></tr>
<tr><td>联系地址</td><td colspan="4">茂名市电白区观珠镇</td></tr>
<tr><td>联系电话</td><td colspan="4">13580070189</td></tr>
<tr><td rowspan="4">其他作</td><td>姓名</td><td>性别</td><td colspan="2">学历</td><td>所在学院</td></tr>
<tr><td>李海宏</td><td>男</td><td colspan="2">本科、2018级</td><td>软件学院</td></tr>
<tr><td>谢淦煜</td><td>男</td><td colspan="2">本科、2018级</td><td>软件学院</td></tr>
</table>

	者	赖乐贤	男	本科、2018 级	软件学院

结题申请

本小组已经完成院级课题《基于深度学习的图像描述》，申请结题

课题负责人签字：刘杰聪

2020 年 4 月 25 日

经费支出概况

支出内容	金 额
《深度学习》*4	288
《深度学习入门之 pytorch》*4	56
《 Python 从入门到精通》*4	170
（备注：以上均为小组自费）	0
总计	514

课题指导教师评定意见

(1)评定内容说明：课题研究是否按计划开展，研究内容有无明显抄袭和 不实之处，研究成果质量如何，研究者是否主动征求导师指导意见并就研究相关问题与导师交流等。
(2)评定情况及分数说明：85 分以上为优秀，70—85 分为良好，60—70 为及格，60 以下为不及格。

综合评定：

该项目主要识别图像，并用一个语句来描述该图像的含义。模型是基于 CNN-LSTM 的编码器-解码器框架，并利用注意力机制提取重要区域，并优化了界面，实验表明，该项目有较好的识别效果。该项目达

	到结题要求，同意结题。 综合得分：88 指导教师签字：梁军 2020 年　4 月　29 日
学院专家评审委员会意见	主审专家签字： 年　月　日
学院课外科技创新领导小组意见	领导签字： 部门盖章： 年　月　日
学校课外	

科技创新领导小组意见意见	领导签字： 部门盖章： 年　月　日
备注	1、本报告中指导老师评定意见、学院专家评审委员会意见和学院课外科技创新领导小组意见**均须手写**，不得用打印，否则该申请表作废； 2、除 1 中所述三项内容外其余内容均须用计算机打印，文字力求精练、准确，不得使用非规范用语； 3、课题形式、学科类别、课题类别的方框均采用涂黑（■）的形式进行选择 4、提交本申请表时需同时提交本申请表的电子文档； 5、结题成果请按《学生课外科研课题成果（文本）格式要求》另附材料说明。

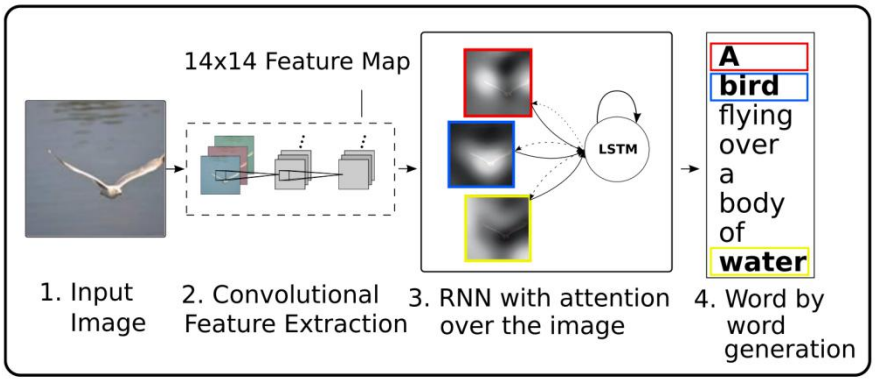
课题内容简介及课题成果

课题内容简介：

图像描述，顾名思义，就是给计算机提供图片，让计算机识别出图片，并以自然语言表述出来。基于深度学习的方法的优势是可以直接从大量的训练数据中学习图像到描述语句的映射，实现端到端的训练，并能产生更精确的图像描述，在性能上远远优于传统方法。

本小组采用的深度学习框架是 pytorch，其设计思路是线性、直观且易于使用的；采用的模型是基于 CNN-RNN（Encoder-Decoder） 的编码器-解码器框架：对整个图像描述生成过程进行建模，分成两个部分：基于卷积神经网络的图像编码器和基于循环神经网络的句子解码器。编码，即将输入序列转化成一个固定长度的向量；解码，即将之前生成的固定向量再转化成输出序列。模型首先利用 CNN 对图像进行有效的特征提取与编码，然后利用 RNN 及其变体 LSTM 生成描述句子。另外本课题引入了注意力机制：解码器的每个时刻在生成一个单词的时候，注意力机制模块根据之前已经预测出的单词信息来提示当前时刻应该关注图像的哪些重要区域，而不是漫无目的地关注整幅图像，然后利用与本时刻相关性高的图像区域特征来生成单词。通俗地来讲，注意力机制就是让网络在解码的时候能够“集中注意力”在编码输出的某些部分上。注意力机制的引入大大提高了我们采用基于编码器-解码器框架的图像描述模型的性能。

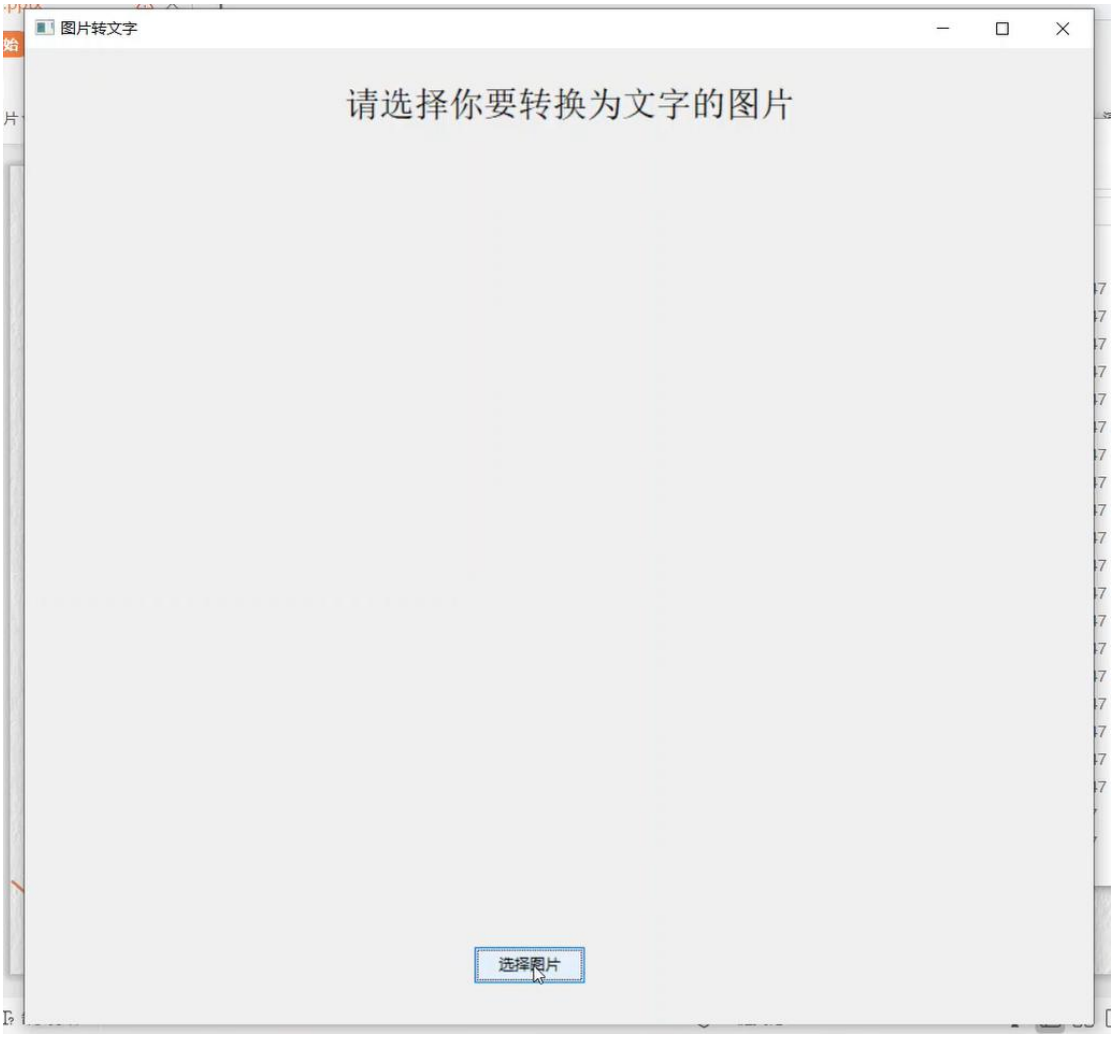
Encoder-Decoder 框架示意图：



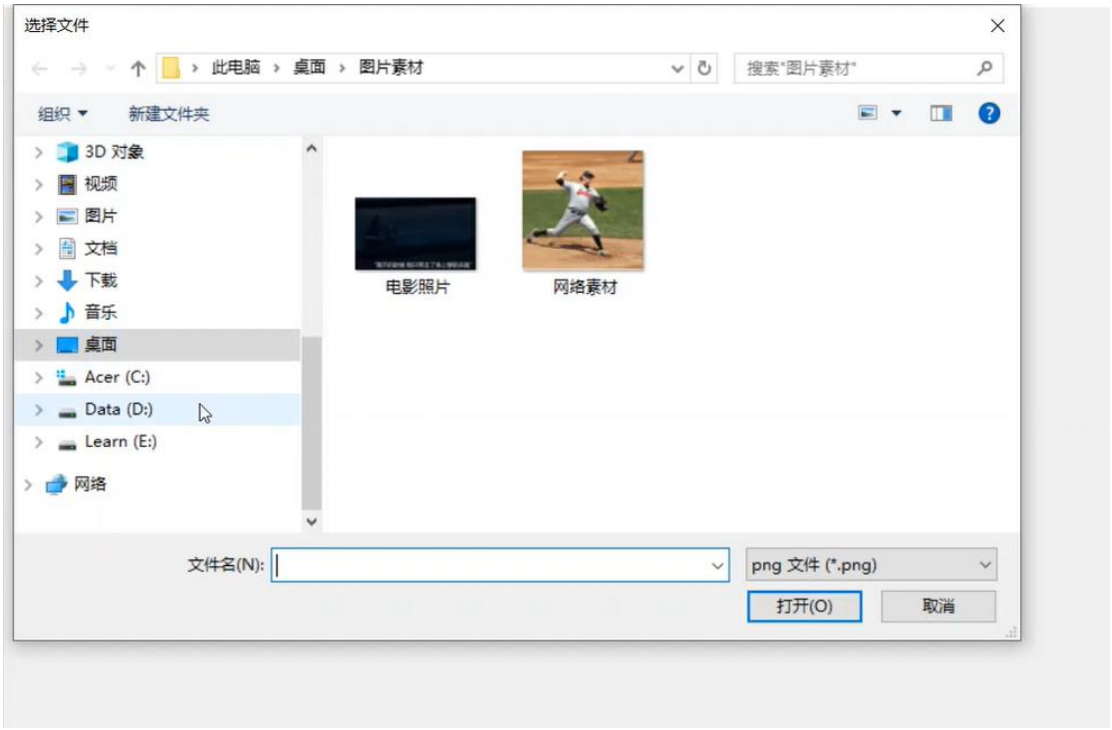
课题成果：

本小组课题成果是一个打包好的软件，实现了申报时的预期目标，由于代码量较大（接近 4G），所以就不展示代码了，下面是软件运行截图，详细的软件演示可见演示视频。

这是软件的打开界面，比较简洁：

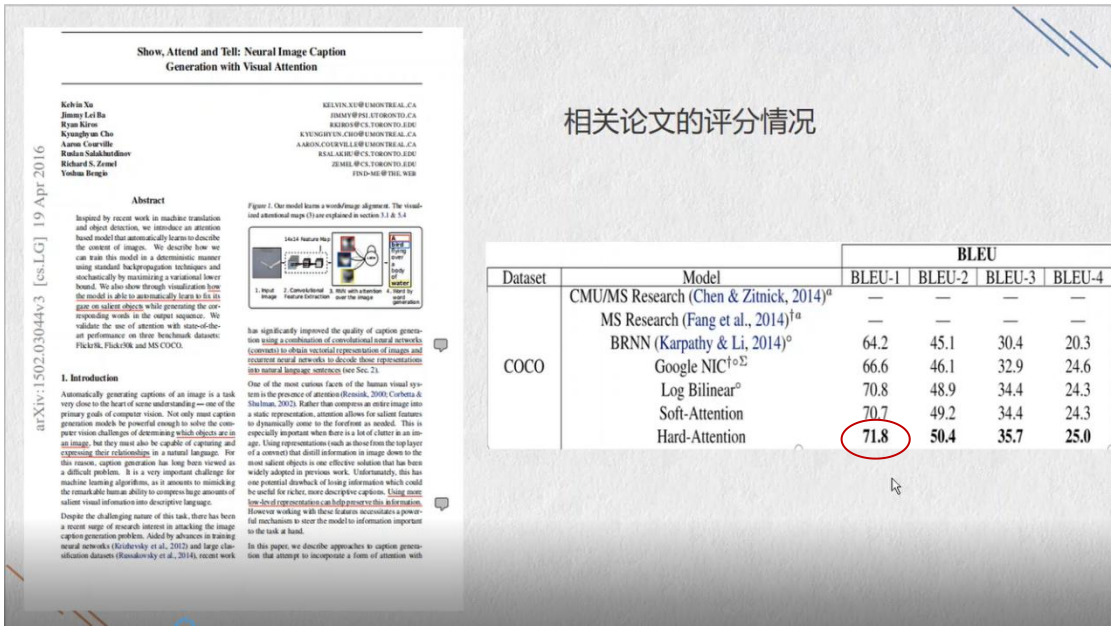


打开后，点击选择图片，会弹出文件选择框供选择：



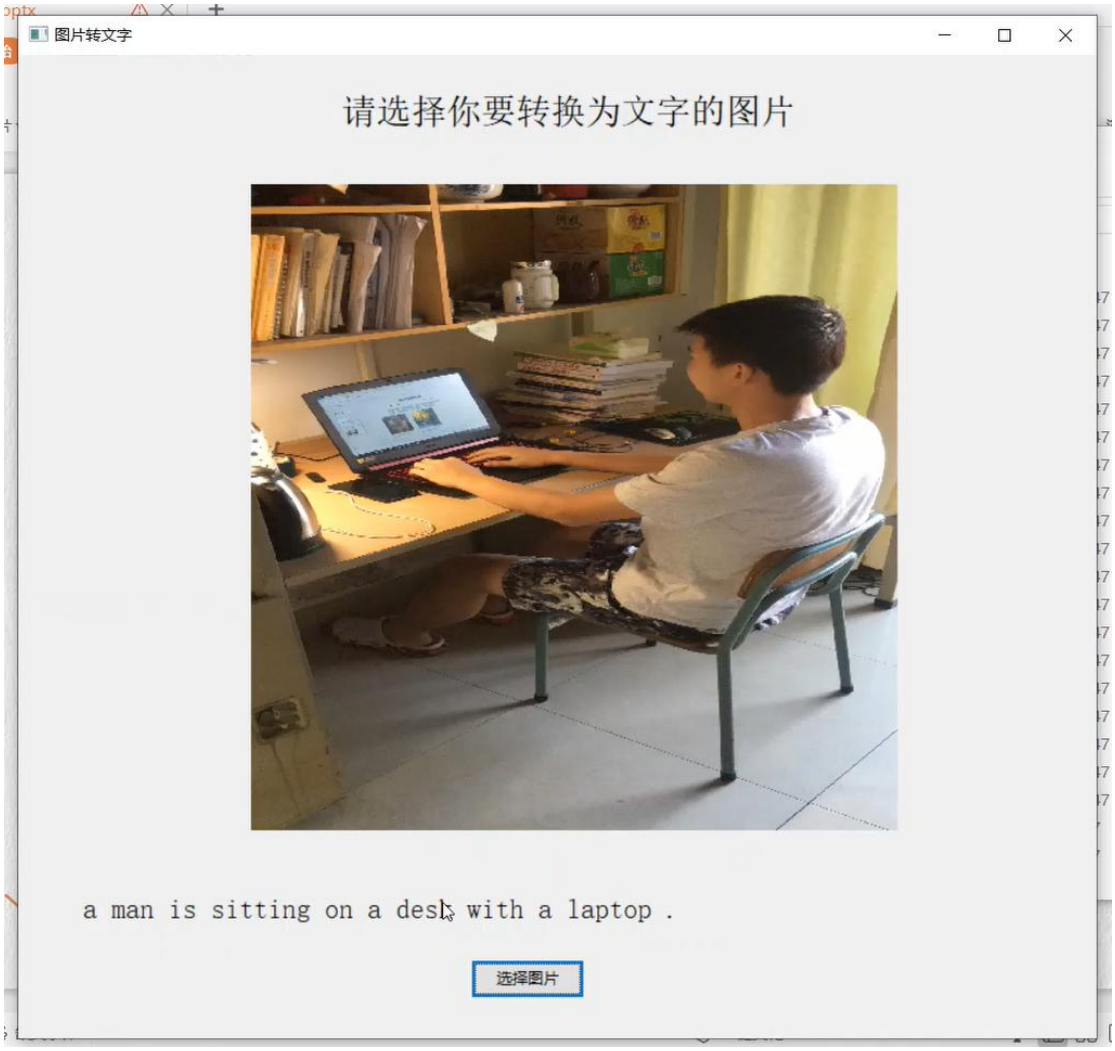
选择完成后，第一次加载会比较久，因为软件会调用一些模型文件，然后进行识别描述，将描述句子显示在图片下方：

图片的描述评估分数 BLEU 分数已达到 69，接近了参考的相关论文分数 71：



```
100%|██████████| 633/633 [04:59<00:00,  2.48it/s]BLEU SCORE:[0.6840128708252998, 0.4471250062524604 3, 0.28219346445189475, 0.18541787110010358]
100%|██████████| 633/633 [04:59<00:00,  2.11it/s]
```

下面是一些图片描述的结果：




图片转文字
请选择你要转换为文字的图片
a baseball player is swinging a bat at a baseball game .
选择图片

请选择你要转换为文字的图片

a woman is holding a cell phone in her hand .

选择图片

请选择你要转换为文字的图片

a cat sitting on a bench in a room .

选择图片

请选择你要转换为文字的图片

a man and a woman in a suit and tie .

选择图片

总体而言，软件界面简单，功能专一（即图像描述）。图片描述得不算特别精确，但大体描述都还过得去，部分描述比较差劲，当然也有部分描述得相当好，这受限于训练模型的数据集以及训练次数等，毕竟采用的数据集 MSCOCO 里面的事物以及对应词汇量有限；另外，由于小组电脑性能不太好，训练时间太长，所以训练次数相应少很多。但小组算是完成了预期目标，并且有点成果。在这个过程中小组成员付出了许多，也收获了许多，比如接触了深度学习相关知识以及体会了团委项目的整个过程，小组从开始的对项目不知所措到现在总算有点头绪。十分感谢学校及学院提供的平台以及资源，感谢指导老师的珍贵建议与教导，小组成员定会铭记在心，继续前行！