

本科毕业设计（论文）任务书

一、题目

基于深度学习的跨模态图像文本匹配研究

二、研究主要内容

本课题来源于国家自然科学基金青年项目，关注多模态计算前沿领域中的跨模态图像-文本匹配任务。本课题的研究主要基于深度学习方法，拟采用卷积神经网络完成图像的特征提取，拟采用循环神经网络完成自然语言的建模，而后通过不同模态特征之间的对齐学习，最终实现图像和自然语言语句的语义准确对齐，进而准确完成利用图像搜索对应的语言描述，以及利用语言描述搜索其对应图像的跨模态双向检索。作为跨模态语义对齐学习中的典型研究任务，相关研究成果在多模态计算领域具备重要的理论和应用价值。

三、主要技术指标

实验基于大规模数据集 **Flickr30K** 进行，作为跨模态检索领域中的代表性研究任务之一，主要采用召回率（**Recall**）指标进行检索准确率评估，一般计算 **Recall@1**，**Recall@5** 及 **Recall@10** 作为技术指标。

四、进度和要求

2023 年 2 月 1 日- 2023 年 2 月 15 日，文献调研
2023 年 2 月 16 日- 2023 年 2 月 28 日，数据获取及预处理
2023 年 3 月 1 日- 2023 年 3 月 15 日，开题准备
2023 年 3 月 16 日- 2023 年 3 月 31 日，深度学习模型搭建及调试
2023 年 4 月 1 日- 2023 年 4 月 15 日，深度学习模型搭建及调试
2023 年 4 月 16 日- 2023 年 4 月 30 日，检索性能评估，模型改进
2023 年 5 月 1 日- 2023 年 5 月 15 日，参数优化
2023 年 5 月 16 日- 2023 年 5 月 31 日，结果整理，论文写作
2023 年 6 月 1 日- 2023 年 6 月 15 日，论文写作，答辩准备

五、主要参考书及参考资料

- [1]. Faghri F, Fleet D J, Kiros J R, et al. Vse++. Improving visual-semantic embeddings with hard negatives[C]//The British Machine Vision Conference (BMVC). 2018.
- Ren P, Xiao Y, Chang X, et al. A survey of deep active learning[J]. ACM computing surveys (CSUR), 2021, 54(9): 1-40.
- [2]. Lee K H, Chen X, Hua G, et al. Stacked cross attention for image-text

matching[C]//European Conference on Computer Vision (ECCV). 2018: 201-216.Ian Goodfellow , Yoshua Bengio 编著.《深度学习》.人民邮电出版社出版. 2017 年 7 月.

[3].《深度学习》，Ian Goodfellow 著，人民邮电出版社

学生学号 2019302973 学生姓名 _____

指导教师 _____ 专业负责人 _____