**实训营第3组启动会**

**一、项目介绍**

**项目1名称：**（Image captioning）图片描述生成

**项目2名称：**（Video captioning）视频描述生成

**任务1、2简介：**Image/Video captioning是近几年非常热门的跨模态视觉任务研究领域之一，输入图片或者视频生成细致的文本描述，该研究可帮助解决视频、图片、多模态信息理解等多方面的任务需求，具有重要的研究价值与未来应用前景。



图 1 图片描述生成示例

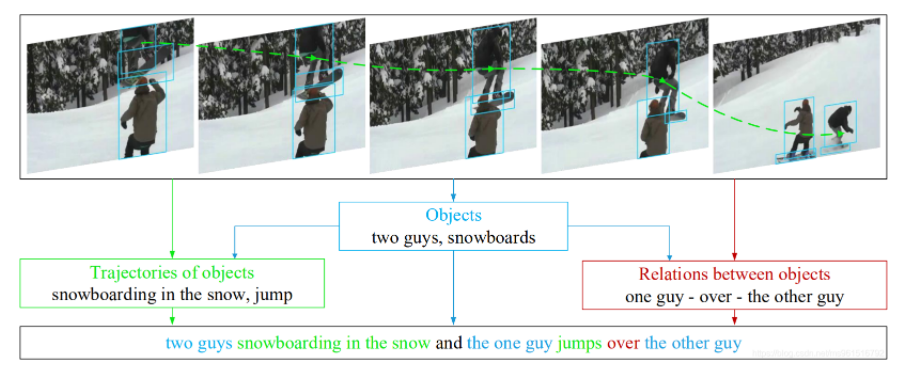
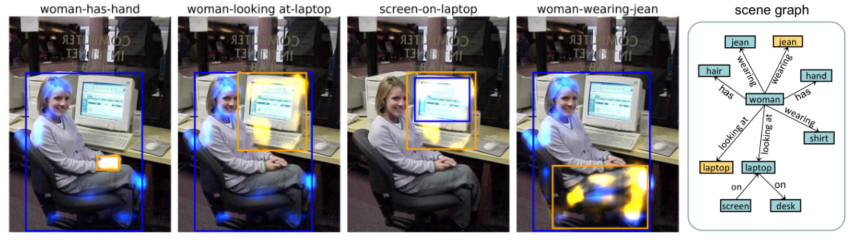


图 2 视频描述生成示例

**项目3名称：**（Image scene graph generation）图片场景图生成

**任务3简介：**Image scene graph generation的研究可以追溯到2017年，李飞飞团队研究对于图像的“细粒度分析”中，将目标检测与关系检测相结合，生成“目标-关系”之间的图状结构，即由目标和关联关系共同构成的场景图。这项研究，可帮助推进图片/视频的细粒度分析以及图片/视频内容理解研究朝前迈出历史性的一步。



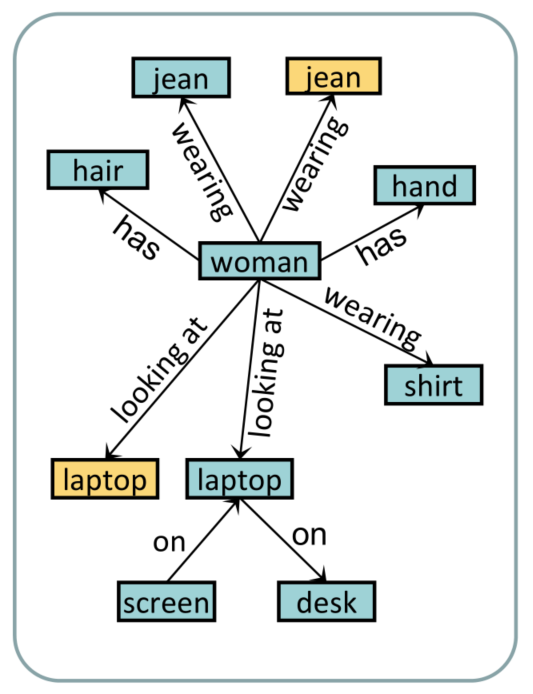


图 3 场景图生成示例

1. **已有成果介绍**

中科闻歌多模态内容理解平台

1. **学习材料及文献**

项目1：Group-based Distinctive Image Captioning with Memory Attention

项目2：Hierarchical Modular Network for Video Captioning（CVPR 2022）

项目3：Stacked Hybrid-Attention and Group Collaborative Learning for Unbiased Scene Graph Generation（CVPR 2022）

1. **第一阶段任务分配（2022/04/28-2022/05/13）**

**任务1：进行研究课题领域的技术调研，撰写调研报告（5000字以上，图文并茂）（2周时间）；**

**任务2：调研和整理近2年该领域的Github相关开源代码，整理到调研报告中；**

**任务3：选择1-2个开源代码，配置环境，运行算法，研究代码；**

**任务4：每周五，简单整理PPT汇报本周工作进展。**