

研究背景

研究背景、研究问题、研究现状与研究内容

研究问题

对多粒度视觉语言信息的利用不充分

语言偏见导致的泛化能力不足

对视觉语言输入变化的鲁棒性不足

低资源场景下学习能力不足

研究内容

在不同尺度上同等程度地表征视频的视觉和语言内容，并在表征整合时兼顾表的多样性

将分布外泛化表述为组合泛化，提出图生成建模策略，用域内数据来生成已有视觉概念的新组合

提出相关信息瓶颈理论，通过平衡表征中信息的冗余与压缩，促进其收敛到最小充分统计量

通过参数有效微调减少可训练参数量以避免过拟合，采用专家混合的方式增加模型容量以提升其性能

研究总结

研究内容的总结与展望