

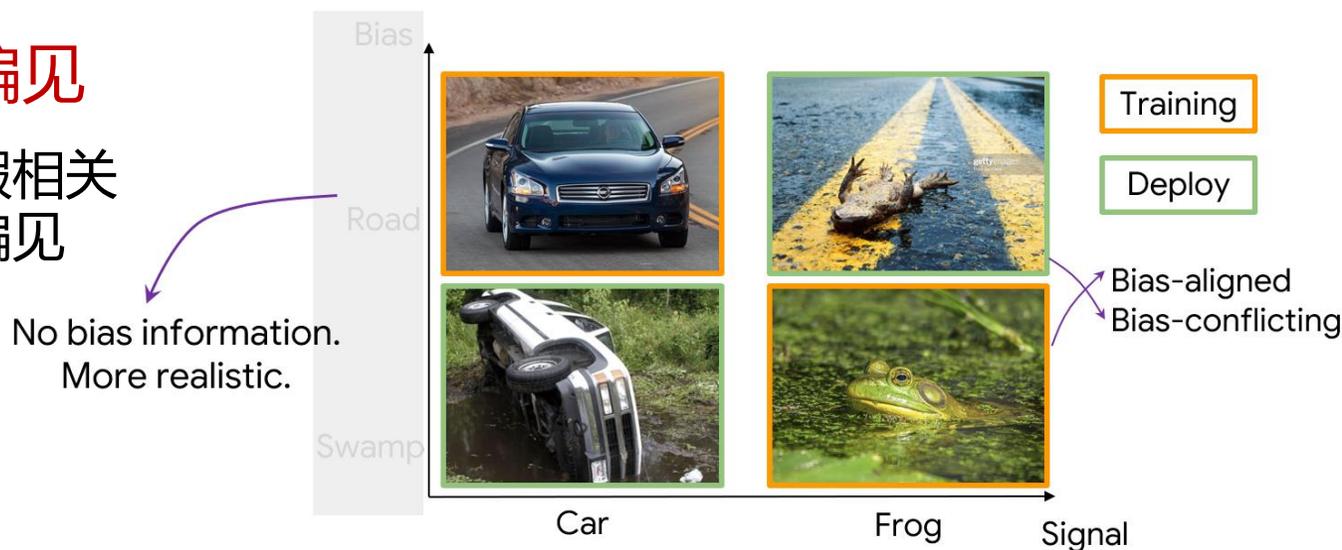
2022年终总结

Rui Hu

2023.1.10

研究方向

- 所在小组研究方向：**模型测试调试**，这个方向的目标是使深度学习模型能够和传统软件一样进行“测试”和“调试”
 - 测试：发现模型问题
 - 调试：修复模型问题
- 个人目前研究方向：**视觉模型去偏见**
 - 解决因数据偏差导致的模型学习虚假相关性的问题，主要关注的是无监督去偏见



工作总结

- 研究论文

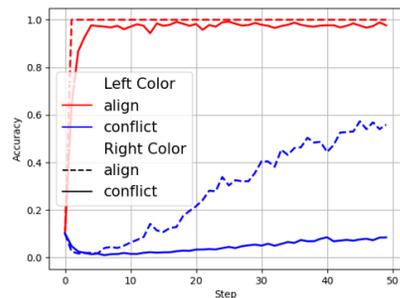
- 面向多偏见的模型去偏，准备中
- 基于反例对比学习的模型虚假相关性缓解，已发表MM22，辅助金强

- 技术项目

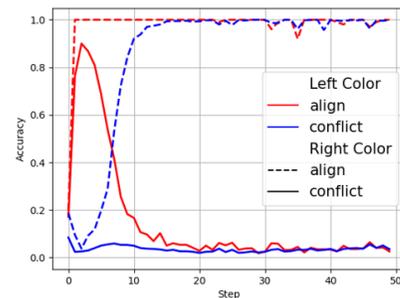
- 兵棋决策归因项目，第一版已上线

- 其他

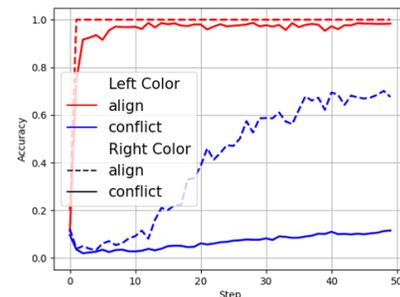
- 22秋机器学习助教



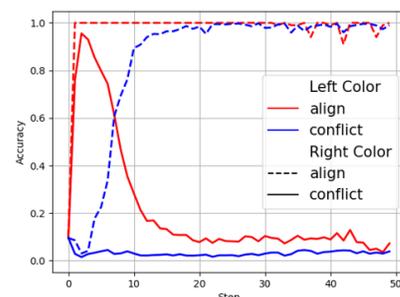
(a)Vanilla



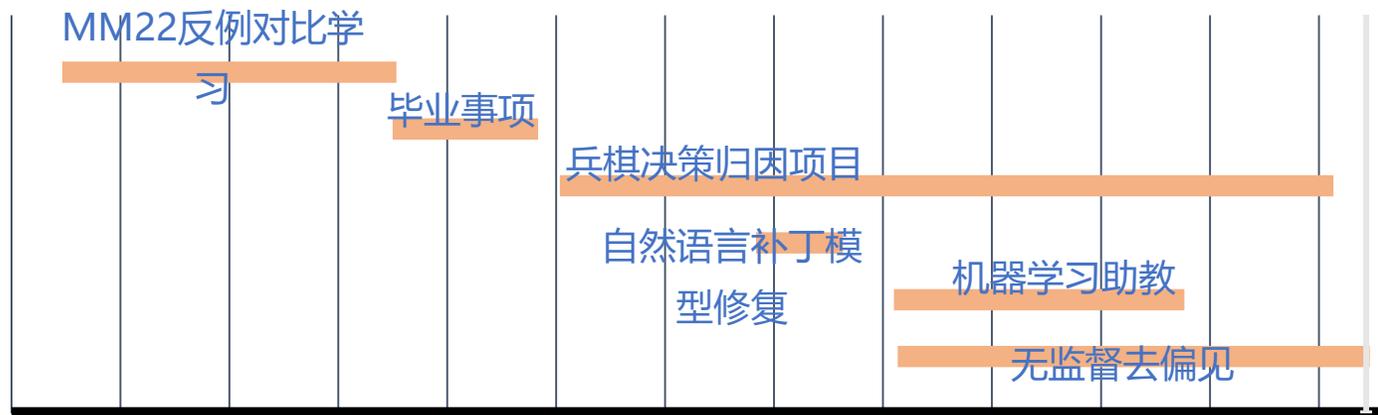
(b)lff



(c)Debian



(d)DisEnt



2022/1/1
2022/1/31
2022/3/2
2022/4/1
2022/5/1
2022/5/31
2022/6/30
2022/7/30
2022/8/29
2022/9/28
2022/10/28
2022/11/27
2022/12/27

入学至今工作量总结

工作	认定积分
素问前期项目	1
辅助MM22论文	0.5
兵棋决策归因项目	0.5
总计	2

2023工作计划

- 面向多bias的模型去偏
- 视觉模型的特征学习优先级研究
- 基于自然语言补丁(多模态)的模型修复

争取形成自己的研究框架

