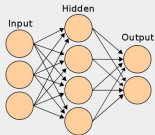
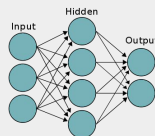


视觉模型  
预训练



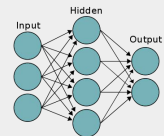
语言模型  
预训练



迁移  
权重



视觉模型



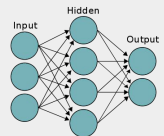
语言模型



迁移  
权重

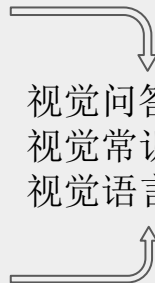


视觉模型



语言模型

视觉问答\  
视觉常识推理  
视觉语言导航



预训练阶段

模态对齐

联合微调