写在前面

很长时间没有写文章了，主要是这段时间在忙 `multi-shot` 数据集采集和复习 `408` 的工作。唉，南哪的夏令营考核太繁杂了，简直比考研还苦难，本校何苦为难本校呢？

Anyway，现在开始要做一个关于计算机视觉和心理学交叉的项目，有很多论文要看。所以开几个 markdown 做一下论文阅读笔记。

学长在发给我们论文之前已经写好了一个大纲，所以这一篇也从大纲的基础上向外延伸。

AI视觉心理：相关文献与资料

1. 心理理论与应用

实验平台的理论基础，箱庭测验与视觉投射理论

关注：1）箱庭疗法，沙盘的基础理论书籍；2）同理心AI心理治疗，一系列论文的组织逻辑是怎样的？如何将AI算法和心理理论结合起来？3）组内心理相关的工作，初步了解沙盘测评与疗愈

- 沙盘与箱庭理论
  - 箱庭疗法 张日昇等
    - 这是一本五百多页的书，里面是很详细的，从箱庭疗法的理论依据、其中元素的象征意义到具体的案例，在此就不细读，可以以后再来参考，当作工具书使用。
- 实例应用：同理心AI心理治疗
  - **A Computational Approach to Understanding Empathy Expressed in Text-Based Mental Health Support**
    - 这是一篇用机器学习方法识别 “共情” 的文章。提出了一个叫做 ***EPITOME*** *(**E**m**P**athy **I**n **T**ext-based, asynchr**O**nous **ME**ntal health conversations)* 的 “共情” 识别模型。
    - Abstract: 共情对成功的心理健康支持至关重要。共情测量主要发生在同步、面对面的环境中，可能无法转化为异步、基于文本的情境。由于数百万人使用基于文本的平台进行心理健康支持，了解这些情境中的共情至关重要。在这项工作中，我们提出了一种计算方法，以了解共情在在线心理健康平台中的表达方式。我们开发了一个新颖的理论基础的统一框架，用于描述基于文本的对话中的共情沟通。我们收集并分享了一组包含10k个（帖子、回复）配对的语料库，使用这个共情框架进行了注释，并提供了支持注释的证据（理由）。我们开发了一个基于多任务RoBERTa的双编码器模型，用于识别对话中的共情并提取其预测的理由。实验证明，我们的方法能够有效地识别共情对话。我们进一步应用这个模型来分析235k个心理健康互动，并显示用户不会随着时间自学共情，揭示了共情培训和反馈的机会。
    - 在技术上看起来还是有 Transformer 的影子在的，首先将 seeker 的数据 encode，然后将 rationale 的数据 encode，两者合成一个 attention，再将这个 attention 与 encoded rationale 作为输出分别输入 Empathy Identifier 和 Rationale Extractor。这被叫做 “Bi-Encoder Model with Attention”
  - **Towards Facilitating Empathic Conversations in Online Mental Health Support: A Reinforcement Learning Approach**
    - 这个工作把上一篇论文的 “识别工作” 抬高到了应用层面，把自然语言改写成更加具有同理心的形式。
    - Abstract: 在线点对点支持平台使数百万寻求和提供心理健康支持的人进行对话成为可能。如果成功，基于网络的心理健康对话可以提高治疗的获取性，并减少全球疾病负

担。心理学家反复证明，共情，即理解和感受他人情感和经历的能力，是在支持性对话中取得积极结果的关键组成部分。然而，最近的研究表明，在线心理健康平台上高度共情的对话是罕见的。在本文中，我们致力于改善在线心理健康支持对话中的共情。我们引入了一项新任务，即共情改写，旨在将低共情的对话帖子转化为更高的共情水平。学习这样的转换是具有挑战性的，并需要对共情有深入的理解，同时通过文本流畅性和对话上下文的特定性来保持对话质量。在这里，我们提出了 **PARTNER**，一个深度强化学习（RL）代理程序，它学会对帖子进行句子级别的编辑，以增加表达的共情水平，同时保持对话质量。我们的RL代理利用了一个基于转换器语言模型的策略网络，该模型是从GPT-2进行了调整，它执行生成候选共情句子并将这些句子添加到适当位置的双重任务。在训练过程中，我们奖励能够增加帖子中共情水平的转换，同时保持文本流畅性、上下文特定性和多样性。通过自动和人工评估的组合，我们证明了Partner成功地生成了更具共情、具体和多样化的回复，并且优于与样式转换和共情对话生成等相关任务的NLP方法。这项工作对于在基于网络的平台上促进共情对话具有直接的意义。
        - 在技术方面，主要的思路是 “在原来的基础上插入有同理心的句子”。模型使用 Seeker Post 和 Response Post 作为输入，经过多个 Transformer 模块，分别用 Position Classifier 决定插入位置、用 Sentence Generator 生成句子，对 Response Post 进行 rewrite。评价标准由四个评价函数组成：Change in empathy reward (r_e)，Text fluency reward (r_f)，Sentence coherence reward (r_c)，Mutual information reward (r_m)。假若改写后的句子不符合标准，会再输入一次来 rewrite，直到结果令人满意。
    - **Human–AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support**
        - 这篇论文和上一篇差别不大，仍然是增加机器生成语句的共情性。不过我似乎没在这篇文章找到任何技术相关的内容，而是人机交互方式？它们并没有对模型进行 evaluate（采用的是上一篇里的 **PARTNER**）而是对不同的采用 ai 的方式产生的效果进行 evaluate，最后发现 “基于AI的建议进行重写（而并不是照抄）”效果最好。
        - Abstract: 人工智能（AI）的进步使得系统能够增强和与人类合作执行简单的机械任务，如安排会议和语法检查文本。然而，这种人工智能与人类的合作对于更复杂的任务，如进行共情对话，存在挑战，因为AI系统在处理复杂的人类情感和这些任务的开放性质方面面临困难。在这里，我们专注于点对点心理健康支持，这是一个共情对成功至关重要的场景，并研究AI如何与人类合作，在文本的在线支持性对话期间促进点对点的共情。我们开发了**HAILEY**，一个AI-循环代理，提供及时反馈，帮助提供支持的参与者（同龄支持者）更具共情地回应那些寻求帮助的人（支持寻求者）。我们在 TalkLife（N = 300），一个大型在线点对点支持平台上，进行了一项非临床随机对照试验，与真实世界的同龄支持者合作。我们展示了我们的人机合作方法导致了同龄人之间对话共情的整体增加19.6%。此外，我们发现，在将自己标识为在提供支持方面遇到困难的同龄支持者的子样本中，共情增加了38.9%。我们系统地分析了人机合作的模式，并发现同龄支持者能够直接和间接地使用AI反馈，而不会过度依赖AI，同时报告了反馈后的自我效能的改善。我们的发现表明，基于反馈驱动的AI-循环写作系统有潜力在开放性、社交性和高风险任务，如共情对话中赋予人类力量。
- 心理沙盘相关工作
    - 基于证据中心设计理论的智能心理测评：构建与应用
        - **HIST** 模型，中科院的工作。
        - Abstract: 有效测量和评估抑郁、焦虑、强迫等心理健康维度是心理测量学的重要问题，本研究构建了基于证据中心设计理论的智能心理测评方法，旨在结合经典测量理论和人工智能最新技术开展智能心理测评研究。在测评环境方面，智能心理测评方法引入电子沙盘游戏作为测评环境，相较量表测评和视觉投射测验，该环境为被试提供了更自由、真实、开放的表达空间；在分析技术方面，本方法通过人工智能技术实现了对沙盘作品象征性意义的自动识别和心理分层评估模型的构建，从而对被测者的心理健康状况进行综合评估。经实验验证，基于证据中心设计理论的智能心理测评方法在信度和效度方面表现良好，对心理健康障碍的预警作用显著，具备对强迫、抑郁、焦虑三个心理健康维度的测评能力。基于证据中心设计理论的智能心理测评方法为心理健康问题的测评

提供了可靠的工具，为人工智能与心理学的跨学科结合提供了有效的研究范式，能有效促进生成式人工智能技术在心理学领域的应用。
    - 模型结构："心理维度 - 诊断要素 - 沙盘主题"，识别沙盘主题（通过摆放的元素和位置提取特征 "主题"） → 获得诊断要素（情绪低落、思维迟缓、精力缺乏、消极预期、易敏感……） → 获取心理维度（抑郁、焦虑、强迫）。
    - 这被称为 "证据中心设计"，从情境中获得指标。
- AI心世界测评原理与数据效果
    - 是一个产品介绍。
- 基于游戏的心理测评
    - 是一篇综述论文。
- 我的世界论文
    - 摘要：本文介绍了在《我的世界》中实施的第一个基于游戏的智力评估，这是一款非常受欢迎的电子游戏，销量超过2亿份。在游戏的三维沉浸式环境中，实现了基于矩阵的模式补全任务（PC）、心理旋转任务（MR）和空间构建任务（SC）。PC旨在衡量归纳推理能力，而MR和SC则是空间能力的衡量指标。我们测试了129名年龄在10至12岁之间的儿童，分别进行了基于《我的世界》的测试和等价的纸笔测试。所有三个量表符合拉斯模型（Rasch model），并且具有中等可靠性。在区分PC和SC之间的区分方面，因子效度良好，但在MR方面没有找到明确的因子。在潜在水平上，使用《我的世界》和传统测试测量的能力之间的相关性很高（r = 0.72），收敛效度良好。子测试级别的相关性在中等范围内。此外，我们发现从游戏环境中收集的行为日志数据对《我的世界》测试的表现具有很高的预测能力，并在较小程度上也预测了传统测试的得分。我们确定了与空间推理能力相关的许多行为特征，证明了分析粒度化行为数据的实用性，除了传统的答题格式之外。总的来说，我们的研究结果表明，《我的世界》是一个适合于基于游戏的智力评估的平台，并鼓励未来的工作探索在纸上或传统计算机测试中不可行的基于游戏的问题解决任务。
    - 这篇论文并不像我预期的那样是类似沙盒测试的物品摆放分析，而是利用mc进行一系列测试，让我觉得有点捞。主要评估方式是与传统心理测试方法做对比。
- **A Hierarchical Theme Recognition Model for Sandplay Therapy**
    - **HI**erarchical **S**andplay **T**heme recognition
    - Abstract: 沙盘游戏疗法作为心理投射的关键工具，测试者构建一个场景来反映他们内心世界，而精神分析师则审视测试者的心理状态。在这个过程中，识别沙盘游戏图像的主题（即识别内容和情感色调）是促进更高水平分析的关键步骤。与传统的视觉识别不同，后者仅关注基本信息（例如类别、位置、形状等），沙盘游戏主题识别需要考虑图像的整体内容，然后依赖于分层知识结构来完成推理过程。然而，沙盘游戏主题识别的研究受到以下挑战的阻碍：（1）收集高质量且足够的沙盘游戏图像，并配对专家分析以形成科学数据集具有挑战性，因为此任务依赖于专业的沙盘游戏环境。（2）主题是综合且高水平的信息，使得难以直接采用现有的工作来完成此任务。总之，我们从以下几个方面应对了上述挑战：（1）基于对挑战的仔细分析（例如，小规模数据集和复杂信息），我们提出了 **HIST**（分层沙盘游戏主题识别）模型，该模型融合了外部知识来模拟精神分析师的推理过程。（2）以分裂主题（代表性且均匀分布的主题）为例，我们提出了一个名为 SP^2（沙盘游戏分裂）的高质量数据集，用于评估我们提出的方法。实验结果表明，与其他基线方法相比，我们的算法表现出更优异的性能，并且消融实验证实了整合外部知识的重要性。我们期待这项工作将有助于沙盘游戏主题识别的研究。相关数据集和代码将持续发布。
    - 结构：输入图片 → 语义识别 → 根据外部信息提取特征 (External Knowledge Incorporation Module) → 沙盘主题识别 (Theme Classification Module)
    - SP^2 数据集，Considering the challenge of gathering high-quality sandplay images paired with theme annotation from psychoanalysts, we take the representative split theme as the example, and construct the SP2 dataset.

2. 视觉理解与分析

早年的一些工作，使用Deep Learning算法解决经典视觉任务

未使用预训练模型，End-to-End

关注：1）在什么数据集上验证？数据集的难度与挑战如何？2）方法的基本框架是怎样的，创新点在何处？3）算法的效果如何，相比其他方法优势在哪里？

- Image Caption
  - **Auto-encoding_and_Distilling_Scene_Graphs_for_Image_Captioning**
    - end-to-end encoder-decoder模型存在一个问题：当将一张包括未见过的场景输入到网络中时，返回的结果仅仅就是一些显著的object，比如“there is a dog on the floor”，这样的结果与object detection几乎没有区别

      认知上的证据表明，基于视觉的语言并非是end-to-end的，而是与高层抽象的符号相关。

      例如，对于一张图片，scene abstraction是“helmet-on-human”和"road dirty"，我们则可以生成"a man with a helmet in contryside"通过使用一个常识：country road is dirty，这种推断就是inductive bias。

      本文将inductive bias融合到encoder-decoder中来进行image captioning，利用符号推理和端到端多模型特征映射互补，通过scene graph($\mathcal{G}$)来bridge它们，一个scene graph($\mathcal{G}$)是一个统一的表示，它连接了以下几个部分

      - objects(or entities)
      - their attributes
      - their relationships in an image($\mathcal{I}$) or a sentence($\mathcal{S}$)，通过有向边表示

      作者提出了Scene Graph Auto-Encoder(SGAE)，作为一个句子重建网络，其过程是 $\mathcal{S} \rightarrow \mathcal{G} \rightarrow \mathcal{D} \rightarrow \mathcal{S}$
      其中 $\mathcal{D}$ 是一个 可训练的字典，用来记录结点特征，$\mathcal{S} \rightarrow \mathcal{G}$ 使用现成的scene graph language parser，$\mathcal{D} \rightarrow \mathcal{S}$ 是一个可训练的RNN decoder，注意 $\mathcal{D}$ 是"juice"——即language inductive bias，在训练SGAE中得到，通过将 $\mathcal{D}$ 共享给encoder-decoder的pipline：$\mathcal{I} \rightarrow \mathcal{G} \rightarrow \mathcal{D} \rightarrow \mathcal{S}$，即可利用语言先验来指导端到端模型，具体的 $\mathcal{I} \rightarrow \mathcal{G}$ 是一个visual scene graph detector，引入multi-modal GCN来进行 $\mathcal{G} \rightarrow \mathcal{D}$ 的过程，来补足detection的不足之处，有趣的是，$\mathcal{D}$ 可以被视作为一个working memory，用来从 $\mathcal{I} \rightarrow \mathcal{S}$ re-key encoded nodes，以更小的domain gap来得到一个更一般的表达。

      - **Contrubution**
      - 一个先进的SGAE模型，可以学习language inductive bias的特征表达
      - 一个multi-model 图卷积网络，用来调节scene graph到视觉表达
      - 一个基于SGAE的 encoder-decoder image captioner with a shared dictionary guiding the language decoding

      搬运自 Auto-Encoding Scene Graphs for Image Captioning 论文阅读笔记-CSDN博客
  - **Dense_Captioning_with_Joint_Inference_and_Visual_Context**
    - Abstract: Dense Captioning 是一种新兴的计算机视觉主题，用于理解具有密集语言描述的图像。其目标是从图像中密集地检测视觉概念（例如对象、对象部分以及它们之间的交互），并为每个概念标注一个简短的描述性短语。我们确定了密集字幕面临的两个关键挑战，在解决该问题时需要妥善处理这两个挑战。首先，每个图像中密集的视觉概念注释与高度重叠的目标区域相关联，使得每个视觉概念的准确定位具有挑战性。其次，大量的视觉概念使得仅通过外观来识别它们变得困难。我们提出了一个基于两个新思想（联合推理和上下文融合）的新模型管道，以缓解这两个挑战。我们以系统的方式设计了模型架构，并对架构的变化进行了彻底评估。我们的最终模型，紧凑高效，实现

了在Visual Genome [23]上密集字幕的最先进准确度，相对于先前最佳算法取得了73%的相对增益。定性实验还展示了我们模型在密集字幕中的语义能力。
        - 与传统标注任务的区别是，Dense Captioning 要求尽可能详细地描述目标的特征。
        - 结构：通过卷积分别获得 region proposal、region feature、context feature，region feature 的分割结果 (基于 Faster R-CNN)(获取 RoI) 可以获得一个 detection scrore，由 region feature 和 context feature 获得 caption (Faster R-CNN & LSTM 缝合) 和 bounding box；
        - 验证数据集：Visual Genome dataset (主要), Flicker30k, MS COCO；
    - **Dense_Relational_Captioning_Triple-Stream_Networks_for_Relationship-Based_Captioning**
        - 提出了 **DenseCap** 模型，**relational captioning** 方法。
        - 验证数据集：Visual Genome，MS COCO，UCF101 (human interactions with other objects or surroundings)；
    - DenseCap_Fully_Convolutional_Localization_Networks_for_Dense_Captioning
    - Dense-Captioning_Events_in_Videos
- Visual Question Answering
    - FVQA_Fact-Based_Visual_Question_Answering
    - Out of the Box：Reasoning with Graph Convolution Nets for Factual Visual Question Answering
    - Visual_Genome
    - zs-f-vqa

3. 语言模型

国内外主流的Large Language Model及相关应用

关注：1）不同语言模型的性能、综合表现如何？参数量与推理开销？2）语言模型接口设计、API调用、微调

- 大语言模型
    - ChatGPT：https://chat.openai.com/
    - 通义千问：通义 (aliyun.com)
    - 书生：https://github.com/internLM/int
    - 智谱清源：https://chatglm.cn/
- 实例应用
    - EmoLLM：SmartFlowAI/EmoLLM: 心理健康大模型、LLM、The Big Model of Mental Health、Finetune、InternLM2、Qwen、ChatGLM、Baichuan、DeepSeek、Mixtral、LLama3 (github.com)
    - MindChat：•X-D-Lab/MindChat: 👍MindChat（漫谈）——心理大模型：漫谈人生路, 笑对风霜途 (github.com)

4. 沙盒模拟与智能体

利用沙盒环境，构建仿真场景，从而模拟并研究人类/智能体在社会环境中相互作用与进化的一类工作

关注：1）如何构建环境；2）使用了什么方法、模拟了人类的哪些特性；3）采用什么算法，取得怎样的效果

- Agent Sims: An Open-Source Sandbox for Large Language Model Evaluation
- Emergent Social Learning via Multi-agent Reinforcement Learning
- Humanoid Agents: Platform for Simulating Human-like Generative Agents
- Social diversity and social preferences in mixed-motive reinforcement learning