

Multimodal-Roadmap-for-freshman

本项目用于Multimodal领域新手的学习路线，包括该领域的经典论文，项目及课程。旨在希望学习者在一定的时间内达到对这个领域有较为深刻的认知，能够自己进行的独立研究。

前2个月精读20篇+实现，然后把深度学习基础过一遍，接下来2个月把领域新的东西泛读+精读过一遍然后做一些别人的idea（帮忙跑实验or洗数据or对方向进行调研），最后两个月开始想一些独立的idea去面向会议进行投稿研究，到投出第一篇为止就算度过freshman阶段。

良好的科研习惯

- 每天早上刷刷arxiv：第一时间看到各个领域最新的文章，获取最新的科研进展。
- 论文分类整理：按照会议或方向，分类整理论文，方便高效查找论文。
- 保证每日科研效率：确保大块的连续的科研时间，如早上或晚上，无人打扰。

领域可以follow人的工作

吴琦老师，何晓冬老师，张含望老师，马林老师，段楠老师，Christopher D. Manning, Jiasen Lu, Paul Pu Liang, 可以关注VALSE公众号，参加各种人工智能大会(智源大会，院士峰会等)

公众号

- 机器之心
- Paperweekly
- 极市平台
- AI科技评论
- 新智元
- 专知
- Arxiv每日学术速递
- 量子位
- CVer

主要关注的顶会

CVPR, ICCV, ECCV, NIPS, ICLR, ACL, EMNLP

重要程度：oral论文（带code）> poster论文（带code）> oral论文（无code）> poster（无code）> workshop

适当阅读science, nature, pami和ijcv的文章，扩大视野

常用网站&必读资料

- [multimodal论文](#)
- [image captioning论文](#)
- [visual grounding论文](#)
- [一文纵览 Vision-and-Language 领域最新研究与进展](#)
- [从 Vision 到 Language 再到 Action, 万字漫谈三年跨域信息融合研究](#)

- [Arxiv](#)
- [Paper with code](#)
- [Aldeadline](#)
- [Overleaf](#)
- Latex
- [张士峰学长分享](#)
- [CVPR Rebuttal](#)
- [画图：PPT，画神经网络图](#)
- [PPT制作（用于组会汇报）](#)
- [Grammarly](#)
- [Linggle](#)
- [提高科研论文写作效率的小工具](#)

代码训练

- [调参](#)
- Numpy
- [新手教程](#)
- [Pytorch常用代码段](#)

单步调试（每篇论文都有code）+精读(About 2 months)

5-20篇论文代表基本理解了相关主题，或许对于进一步理解技术实现足够了。

50-100篇论文，你将会对这一领域有一个非常好的理解。

关注点

- 1、Describe what the authors of the paper aim to accomplish, or perhaps did achieve.这篇论文作者的目标是什么，或者也许已经实现了什么。
- 2、If a new approach/technique/method was introduced in a paper, what are the key elements of the newly proposed approach?如果文中引入了一种新方法/技术，那么这一新提出的方法/技术的关键要素是什么？
- 3、What content within the paper is useful to you?论文中，有哪些内容对你有用。
- 4、What other references do you want to follow?你还想关注哪些参考资料/文献？

深度学习基础

- [《动手学深度学习》](#)
- [《机器学习&深度学习》](#)
- 西瓜书《机器学习》
- 花书《深度学习》
- 《统计机器学习》
- 《The book of why》探讨因果
- 斯坦福CS231n 视觉部分
- 斯坦福CS224n transform/generation部分
- [Tutorial on Multimodal Machine Learning](#)

Week 1

《Attention is all you need》

《[Multimodal intelligence: Representation learning, information fusion, and applications](#)》 综述

Week 2

《Self-critical Sequence Training for Image Captioning》

《Bottom-Up and Top-Down Attention for Image Captioning and Visual Question Answering》

Week 3

《SCA-CNN-Spatial and Channel-wise Attention in Convolutional Networks》

《Show, Attend and Tell--Neural Image Caption Generation with Visual Attention》

[BLEU](#)、[Meteor](#)、[ROUGE](#)、[CIDEr](#) 和 [SPICE](#)

Week 4

《Show, Control and Tell--A Framework for Generating Controllable and Grounded Captions》

《[Spice: Semantic propositional image caption evaluation](#)》

《Knowing When to Look: Adaptive Attention via A Visual Sentinel for Image Captioning》

Week 5

《[Unsupervised Image Captioning](#)》

《[Conceptual Captions: A Cleaned, Hypernymed, Image Alt-text Dataset For Automatic Image Captioning](#)》

《Show and Tell: A Neural Image Caption Generator》

Week 6

《BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding》

《Corpus-Guided Sentence Generation of Natural Images》

《Hierarchical Question-Image Co-Attention for Visual Question Answering》

Week 7

《Baby Talk: Understanding and Generating Simple Image Descriptions》

《Long-term Recurrent Convolutional Networks for Visual Recognition and Description》

《[Auto-Encoding Scene Graphs for Image Captioning](#)》

Week 8

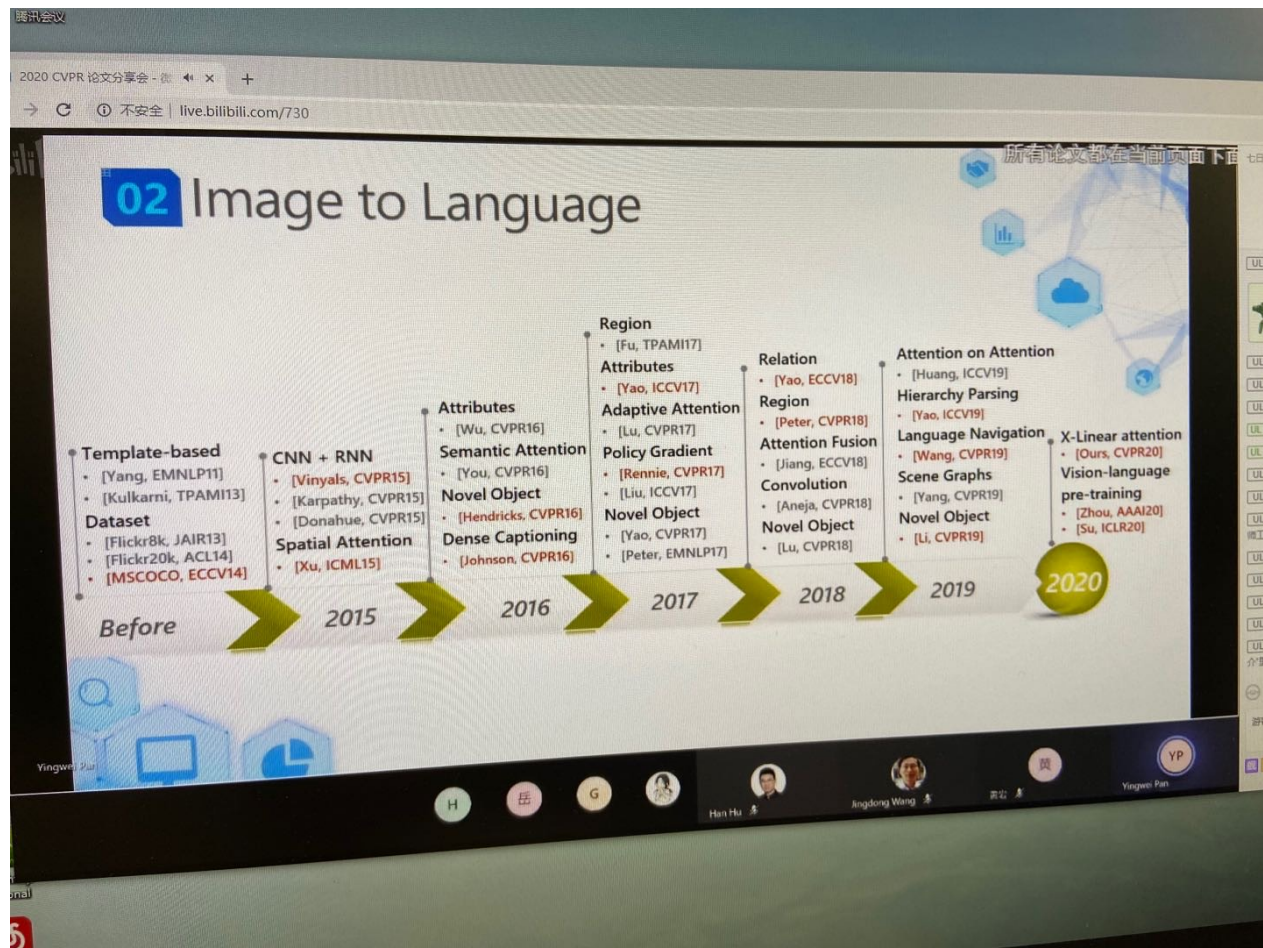
《Aligning Visual Regions and Textual Concepts for Semantic-Grounded Image Representations》

《[Attention on Attention for Image Captioning](#)》

《VL-BERT: Pre-training of Generic Visual-Linguistic Representations》

Week 9

泛读当年CVPR所有论文(尤其是oral)，可以思考其他领域和multimodal的关联，对整个AI发展有个趋势的认知。然后做一份阅读报告作为第一阶段的总结。



实验坑&工具&trick (About 2 months)

Linux学习

服务器学习

Pytorch

Tensorflow

实验trick

实验遇到的坑

未来的方向(About 2 months)

2020年: self-supervised learning+multimodal

Unsupervised/semi-supervised+multimodal

可控制的image caption/visual grounding

visual reasoning

symbolicism+multimodal

visual grounding+传统CV/NLP任务

(仅供参考)