

大会主旨报告

[大会主旨报告 - 特征编码与数字视网膜](#)

[大会主旨报告 - 下一代深度学习的思考与若干问题](#)

[大会主旨报告 - 计算机视觉--从孤立到系统性方法](#)

年度进展评述

[神经网络模型轻量化设计](#)

[基于NeRF的三维视觉](#)

[类脑智能](#)

[医学影像检测与分割](#)

[扩散概率模型的前沿进展](#)

[CV大模型VIMER](#)

[基于隐式神经表示的三维重建、渲染与定位](#)

[开放视觉感知](#)

[开放域目标检测/识别](#)

[视频动作分析与识别](#)

[视觉自监督学习](#)

[遥感目标检测](#)

[人脸感知计算](#)

[多模态融合感知](#)

[图表示学习](#)

[文档智能](#)

Workshops

围绕手机的计算影像学

[当A\(academia\)碰到I\(industry\): 图像复原和增强任务中AI的困境和机遇](#)

[卷帘快门相机: 模型、优化与学习](#)

[影像的时空分辨率增强](#)

[如何设计一个真正意义上的智能相机](#)

[Zero-Shot Image Restoration Using Denoising Diffusion Null-Space Model](#)

[Panel](#)

大模型对CV/PR的挑战与机会

[国产E级智算平台上视觉大模型训练及应用](#)

[在大模型背景下，重新审视计算机视觉的发展方向](#)

[多模态预训练的进展回顾与展望](#)

[视觉数据多样性带来的机遇与挑战](#)

[视觉基础模型的安全性与可信](#)

[面向多模态多任务的通用视觉感知大模型](#)

[基于解耦学习的可逆纵列神经网络](#)

[Panel](#)

人工智能与科学计算

[盘古气象大模型：3D全球高分辨率气象预报方法](#)

[从计算机视觉到AI4Science-挑战与机遇](#)

[量子图机器学习及其化学应用初探](#)

[融合HPC和AI的药物分子设计](#)

[非均匀物理数据分布的偏微分方程（PDEs）的通用求解算子](#)

[Panel](#)

开放域的学习方法与视觉任务

[开放域物体识别算法](#)

[从视觉内容生成到视觉任务自动化](#)

[面向开放域的虚实迭代强化学习](#)

[多模态开放域检测大模型及应用](#)

[混合大模型](#)

[动态开放环境的具身持续学习方法](#)

[大规模基础模型在开放世界任务中的研究与应用](#)

[Panel](#)

特征编码与数字视网膜 (Digital Retina System)

高文 北京大学 工程院院士

背景

- 计算机：采集-》预处理-》检测识别-》场景理解
- 人类：视网膜采集信号-》视神经神经元进行编码-》眼动，凝视等注意机制-》形成理解

目的

设计类人的视觉编码系统，主要针对类视神经的编码模式，如何设计注意力，如何协调与传统编码的关系

1. 视觉编码决定了下游算法的性能瓶颈（识别，检测）

回顾视觉编码发展：

1. Canny, Sift等filter算子(LBP,SIFT,HOGMSURF,ORB)
2. 深度学习

认知心理：

感官数据：10e9 byte/s 视皮层：10e3 bytes/s

=》大脑只留下了简化后的表征，而且压缩倍率很高

2. 根据香农理论，任何失真度，只要码长足够，总可以找到一种编码
3. 大脑的编码模式近似贝叶斯公式
4. 联结主义模型
5. 人类视觉编码特征
视觉编码约1.5秒：低码率，粗粒度，关注语义
6. 应用
城市大脑，主要克服端-边-云环境中的传输，复用瓶颈

- [Digital Retina: A Way to Make the City Brain More Efficient by Visual Coding](#)
- [Towards Coding for Human and Machine Vision: A Scalable Image Coding Approach](#)
- [Video Coding for Machines: A Paradigm of Collaborative Compression and Intelligent Analytics](#)

领域进展

- AVS3 8K超高清视觉编码标准
 - 2019 AVS3 8k解码芯片发布并在2021用于冬奥会
- CDVA
 - [Compact Descriptors for Video Analysis: the Emerging MPEG Standard](#)
- 模型压缩与增量更新
- 数字视网膜体系12部分将从端-边-云部署到城市
 - 采用无监督, 元学习等技术减少人工标注, 模型自适应
 - 鹏城-大圣 鹏城-云脑 (深圳)
 - **大模型已经可以向30+下游任务做迁移, 并取得性能提升**

下一代深度学习的思考与若干问题

焦李成 西安电子科技大学

深度学习的milestone技术都不在中国, **中国学者下一步该做什么样的研究?**

(c: 类脑智能, 突破data-driven的限制是焦老师认为的未来的方向, 焦老师对于模型可解释性, 理论性比较看重, 非常讨厌造大黑箱)

神经网络先驱

- P.warl BP
- 福岛彦邦 1980 卷积神经网络

深度学习: motivation都是类脑智能 (c: 似乎焦老师比较喜欢深度学习往类脑智能发展, 喷了下大模型可解释性和高功耗, 而且好像做类脑就会比较喜欢玩表征学习, 高文老师也是这样。)

方向:

- 优化问题, 设计高于BP的学习算法
- 表征学习不能靠网络结构堆砌, 而是要从优化的角度思考
- 复数域网络
- 人类先验建模赋能模型设计
- 多尺度几何建模 (c: 和我的课题是否相关? SIFT等算子是否能帮助劣化模型的建模)
- 从物理, 生物, 推导深度学习的新发展

CV系统性方法

陈熙霖 中科院计算所所长 IEEE fellow

背景

- 人类因为学习占领了地球，学习需要视觉
- 人的学习是持续性，和探索性的
- 理解才是模型的发展方向

领域发展：

- Pos: 开封开源，门槛低
- Neg: 性能驱动（打榜），问题过于细分

系统化方法：

- SAM: 分久必合，小领域，小问题开始合并，背后是巨大的数据支持
- Question as all: 问答学习
- 多模态
- **具身智能：让智能在仿真环境中学习**

大模型是希望还是终结？

未来：

- 催生AI 体系结构，大模型成为组件

模型 = 模型复杂度 x 数据

模型成熟度 = 算力 / 模型复杂

(cc:和焦老师高老师不一样，显然陈老师是支持大模型发展的，乐，学术观点不一样，真有意思)

要发展更简单的通用算法（模型），在专用算法上做复杂化

构建智能体，终身学习，观察学习，探索学习

(c:这种level的大佬都是关注大问题啊，感觉智能学习的进展肯定是颠覆领域的，但也是很难的，)

智能系统的评价

多维度的评价

神经网络模型轻量化设计

南京大学 吴建鑫

- Minimum variance unbiased N:M Sparsity for the neural gradient
- DepGraph: 结构化剪枝-不同层的参数之间的依赖关系被解的很好
- Deep Partial Updating: 边缘设备模型参数小部分更新

ViT轻量化：

- EdgeViT: 无需蒸馏，三星上时延与CNN差不多但是FLOPS却多很多 (c: 不能只看flops) token选择性的attention并加入interpolation操作

- RIFormer
没有token mixer, 替换multi-head self attention layer (MHSA) 为仿射变换, (软硬件结合)
- EfficientFormer
部分结构没有MHSA, 软硬件结合
- 知识蒸馏:
- Ensembling weak checkpoints
做teacher-student, 实际上用中间部分没有完全收敛的模型做teacher效果会更好
- DIST: Distillation from a stronger teacher
- Decoupled Knowledge Distillation
不用目标类, 非目标类做k-1分类, 目标类做二分类

小样本, 零样本压缩:

- Feature mimicking
- Compressing Transformer:
PCA做矩阵分析, 压缩线性层

基于NeRF的三维视觉

清华大学 刘焯斌

(c:orz PPT过的贼快)

Nerf 使用神经网络还原可见光成像过程 (cc: 能用来做劣化模型建模吗)

1. 效率优化:

- 体素化
instantNGP 秒级收敛
- 体素分解
- 利用稀疏几何

2. 动态建模

- 动态前景感知 (c: 前景!)
- 人体重建
AvatatRex, Posevocab
- 场景交互

3. 可泛化的nerf重建

- 基于扩散模型

4. 编辑

光影编辑, 人像编辑

5. 场景建模

无边界的nerfi建模, 比如城市级

相机姿态估计

类脑智能

中科院自动化所 张兆翔

实现更智能更低功耗强学习能力的类脑智能

1. 神经元模型

(c: SNN spiking neural network好像是个蛮重要的工作)

- 膜电位发放: 脉冲啥的 ($\phi(\otimes \square \otimes; \varphi)$)
- 神经元特异化: 如SNN与transformer结合
- 神经元建模: 脉冲网络版resnet

2. 神经网络结构

- 生物视觉
- 分析SNN, ANN, 发现SNN更align小鼠的视觉机制???
- 丘脑结构启发
- 彩票网络: 稀疏网络媲美大模型

3. 训练方式

- 事件驱动的SNN学习
- 多模态学习

总结

SNN是主体, 生物脑是发展方向, 正在吸收大模型的技术-

医学影像检测与分割

夏勇 西北工业大学

背景

医学影像成片类别多, 应用主要是在疾病检测, 愈后分析等

1. 小数据到大数据

器官, 目标疾病种类多样化

- MOTS (Multi-Organ and Tumor Segmentation Benchmark)

2. 从小模型到大模型

STU-Net (Params: 1.4 billion)

应用

1. 诊断报告

通用诊断报告生成 (S4M: Generating Radio..)

2. 多模态

2D-2D

Image-Text: BiomedGPT

3. 可交互

可交互分割

基于SAM的交互分割

挑战

1. 需要更多的数据: 标注质量, 隐私
2. 多模态
3. efficiency? 更快的诊断, 对于交互式模型也很重要

CV大模型VIMER

王井东 百度视觉首席科学家

企业比较关注: 自监督, 弱监督, 半监督, 多任务

自监督, 弱监督, 半监督帮助模型预训练

自监督发展

- 数据增强
- contrastive learning
- MAE

百度现有工作: Context Autoencoder

应用:

工业视觉大模型, 人体建模大模型, OCR大模型

大模型 Dense-recognition (语义分割, 目标检测):

- Accelerating DETR

基于隐式神经表示的三维重建、渲染与定位

浙江大学 章国锋

使用隐式方法对三维场景进行表示

Nerf (隐式神经表达+神经渲染)

真实场景下的问题

自由编辑, 任意场景定位, 支持下游应用

1. 物体不可解耦

希望做到场景物体分离, 单独进行建模, 编辑

Solution: two-branch 场景分支只建模背景, 物体分支则相反

2. 编辑方式单一

只有刚性编辑, 颜色编辑, 材质替换

Solutions: 实现非刚性编辑, 局部纹理编辑填充

基于Mesh框架实现 (Mesh由NeuS重建)

3. 三维场景编辑方案少, 且不方便

Solutions:

1. 单张图像实现建模及编辑, 几何编辑场与纹理编辑场+双体积渲染+神经编码的几何先验 (形状先验)

2. 基于文本对纹理做编辑

4. 基于体素3维建模

未来:

隐式SLAM, 大场景建模, 复杂应用场景建模, 多模态

开放视觉感知

刘偲 北京航空航天大学

问题: 对于陌生物体, 模型容易失效, 但人类可以处理陌生物体, 做出判断

开放视觉感知:

1. 从数据中获取知识

- 人在回路: 通过增量学习实现未知类别感知
- 多任务
- 图像文本匹配: clip
- 知识迁移: 把大模型的知识迁移出来

(OADP) Object-Aware Distillation Pyramid: 伪标签与多尺度特征蒸馏

2. 从交互引入概念

- 视觉交互

- 1. 视觉样例，提供输入样例，无需微调模型（c: 感觉类似in context learning）

- 2. 视觉提示，提供输入输出对样例，无需微调（c: 真-in context learning）

- 语言交互

- 1. 渐进式理解？

未来：

具身系统：在真实环境下做交互，在虚拟世界做智能体的交互学习

开放域目标检测/识别

侯淇彬 南开大学

借助clip等文本编码器，帮助视觉在为训练的类别上完成预测。（open-vocabulary）

视频动作分析与识别

南京大学 王利民

领域发展：

- 1. 时序差分网络（TDN）

- 2. Vision Transformer

- 3. ViViT

- 4. Uniformer

- 5. Video MAE(V1/V2)

视觉自监督学习

微软亚研院 胡瀚

技术进展趋势

- 1. 掩码图像的改进

- BEiT -> MAE -> MASKFeat->MVP(已CLIP特征作为重建目标)->BEiT-V2

- 2. 大模型友好

- 模型容量的扩展持续改进这NLP任务性能

- 3. 小模型

4. 挖掘性质

MIM Dark secrets

5. 模态拓展

遥感目标检测

程gong 西北工业大学

挑战:

- 有向目标检测
- 篇幅大, 数量多, 目标类型复杂
- 高轨卫星目标特别小, 深度学习, 传统算法性能都不够好

进展:

- 有向目标检测:
Oriented R-CNN for Object Detection
- 细粒度目标识别: transformer 目标旋转敏感

人脸感知计算

邓伟洪 北京邮电

无监督方法:

- 编辑 EigenGAN (ICCV 2021,和Stylegan? ?)
- Debiased 方法
- Privacy Protection
- 识别结果的解释
- Facial expression recognition
 - 姿态解耦
 - 复合表情
- Gaze estimation
- Lip Reading

多模态融合感知

西安电子科技大学 邓成

1. 多查询+响应

单查询向量不合适，无法表达足够的信息

2. 多模态预训练模型

DALLE2 (CLIP+DDPM)

GPT4

3. 多模态感知

基于任务分解

基于层级语义关联 (把文字表述的视频提取出来)

基于联合自注意力

4. 多模态内容生成

未来：大模型从单个到多个，从黑盒到白盒，从认知到推理

图表示学习

中国人民大学 魏哲巍

1. 空域GNN

- GIN
- k-hop
- GD-WL ICLR2023 Outstanding 解决双连通判别问题

2. 谱域GNN

- 同配图，异配图 与 滤波器

3. GNN 训练策略

- 自动化神经网络搜索
PaSca
- 分布式训练
SANCUS
- 联邦学习
FS-GNN
- 训练过平滑问题
DropMessage

4. 几何图学习

AI for science

每个节点具备坐标，旋转如何保持不变性

5. Benchmark

Graphworld LongRangeGraph

总结：方兴未艾，理论基础，应用，benchmark偏少

当A (academia) 碰到I (industry): 图像复原和增强任务中AI的困境和机遇

张磊 港理工

数据量越来越大，算力需求越来越大

challenge:

退化无关的修复:

多退化建模 网络结构改进 多任务通用模型

academia: 刷榜

set5, 14和现在拍摄的图像其实差别很大

industry: 刷配置

academia: 模拟仿真数据

industry: 真实传感器

academia: PSNR/SSIM/LPIPS指标与人类质量感知不一致

ind: 用户满意 (质感, 油画感etc)

ac: flops param

ind: 时延功耗, 算子算量, 内存

现有方法 手机, 相机设备间噪声不一致, sr方法无法解决真实场景

现有评价指标完全不够:

小文字修复

实用高效小模型, 硬件友好

, 新型结构

通用大模型? deblur, SR是有冲突的

看好stable diffusion

生成式复原，复原式生成（复原与生成的结合如何平衡）

A 从实际问题角度思考

I: i提供数据设备支持

卷帘快门相机：模型、优化与学习

戴玉超 西北工业大学

CMOS相机成为主流

卷帘快门，逐行曝光。

针对全局曝光的模型可能失效，尤其是高速摄像

1. RS 相机几何模型

- 基于GS的模型需要改进来适配RS模型
- 加入时间维度
- CVPR 2016

2. 复原

RS correction

帮助3D重建

Joint Appearance and Motion Learning Efficient Rolling shuffle corrector

3. 插帧

从两张连续的卷帘照片构造一段GS视频

光流法

影像的时空分辨率增强

大连理工大学 贾旭

时：插帧

空：超分

模型设计

- 模型搜索
 - 遗传算法搜索卷积数量， channel等等来找到合适的超分模块
- Transformer
 - transformer 对纹理建模
- Recurrent Structure
 - BasicVSR, BasicVSR++
- Explicit Temporal difference

- Look back and forth (cvpr 2022)
- Real world Degradation
 - Unsupervised degradation representation learning
 - degradation aware adaptation
- VSR on compressed Videos
编码导致的噪声与信息丢失
- Multi-sensor
event camera (非匀速捕帧)
统一模型 (一个模型完成去模糊, 插帧等一系列任务)
 - neural Image Re-Exposure (实现重曝光)
 - neural film 矫正

未来:

- diffusion model
- large pre-trained model
- multi-sensor fusion

如何设计一个真正意义上的智能相机

薛天帆 香港中文大学

什么是智能相机?

应用人工智能算法 ×

解决实际问题 ✓

相机pipeline:

sensor 采集 -> 前处理 -> 后处理

采集端: 如何获取多源信息

- 长曝光?
- 多曝光
 - HDG
 - 多帧去反射 (反射移动的快)
- 不同摄像头的融合, 比如rgb+红外
- 同一张图 dual pixel

处理: 数据与模型

- 难标注，相较于检测分类任务，图像增强 GT难定义
- 相机设备发展太快了
- 如何合成真实噪声
 - raw数据噪声好建模
synthesize raw from srgb, 加噪声再做isp
 - 眩光合成
基于scattering reflection建模来合成眩光
- 网络设计
融合传统算法
 - 4k数据的图像处理
Patch 的图像变换可以是现性的, joint bilateral
 - noise

未来:

1. simulation 深入到sensor级别的
2. lens+sensor+ML 为硬件设计算法
3. VR, 相机采集3D信息, 相机能否和人眼一样好, 8k60fps如何做图像处理

计算手机panel

1. 手机端应用大模型

- 手机端的算力发展会跟上模型发展
- 小模型蒸馏
- 端云结合
- 可能会有全新的问题出现
- 视频处理的瓶颈比图像处理大得多 (unet都不一定能支持)

2. 手机影像还存在的问题

- 全pipeline的设计, 而不是纯算法设计 (软硬协同)
- 手机处理三维影像
- 手机处理场景 (新设备)
-

3. 审稿

审稿人不应该太拘泥于性能上marginal的提升, 而要更多关注方法的insight, 可行性

在大模型背景下，重新审视计算机视觉的发展方向

谢凌曦 华为

AGI Artificial General Intelligence

definition: maximizing reward in an environment

GPT目前最接近通用人工智能: the spark!

可以理解笑话, 写代码, debug

观点:

- CV is a superset of NLP
- Language alone is insufficient

challenges:

- CV不同任务间模型区别很大

统一的思路:

- 开放域识别
 - SAM
 - Gato
 - 大模型guided视觉理解

大一统的本质问题

chat这个交互式任务是最重要的

交互任务的设计是最重要的: *proxy task(分割, 检测, etc) is dying!*

未来:

具身智能 PALM-E ENTN

从NLP大模型到CV大模型可能需要新的方法论

视觉数据多样性带来的机遇与挑战

黄高 清华

从数据角度聊模型进展

data-centric ai

数据多样性: 长尾问题。。

1. semantic data augmentation (提高泛化)

deep feature interpolation for image content changes!!

(cc:类似用stylegan做数据增强)

(nips 2019) implicit semantic data augmentatino

2. generalized curriculum training

curriculum trainig(先学习简单样本, 加速训练)

定义简单, 复杂样本, 难, 耗时高

easy data -》 easy pattern

- 从频域角度分析 (先学习低频)
- 从数据增广角度 (先学习没有做强数据增广的样本)

3. Dynamic inference

根据样本来决定处理方式, 使用多少模型参数, 注意力机制 (加速运算)

视觉基础模型的安全性可信性

董胤蓬 清华

基础模型 (foundation model) : 预训练的大模型, 可适配下游任务

all system might inherit the same problem bias from the foundation model

backdoor attacks: 在训练阶段加入 poisoned sample

distribution shifts: viewpoint change就可以让模型出问题

SAM GPT的攻击手段都是存在的

1. 能否利用预训练模型的脆弱性, 攻击fintune模型?

Pretrained adversarial perturbations (nips 2022)

不同深度层捕捉不同语义级别的特征: low-level层在finetune时变化不大

uniform Gaussian sampling

2. 大模型用来解决攻击问题

鲁棒性的提高伴随着精确度的下降

利用backdoor做模型版权保护

面向多模态多任务的通用视觉感知大模型

商汤 朱锡洲

感知模型发展

从感知的角度区分不同领域，现在在思考通用感知

开发模型：VisionLLM 可基于文本调整模型做的任务

未来：开放世界智能体

challenge：场景不确定，任务时长长

基于强化学习：需要目标定义（局限于特定任务），reward非常稀疏，训练开销大

基于LLM：分解复杂任务为简单任务，以文本形式学习，动作接口通用（简单任务）

GIMT

(GHOST IN MINECRAFT)

分解：将目标定义为（object, count, material, tool, info）

LLM planner：子任务通过文本记录成功的action 序列（由LLM总结）

使用语义封装结构化接口

基于解耦学习的可逆纵列神经网络

张祥雨 旷视

表征学习：

IB 准则 information bottleneck：表征学习跟task是耦合的

occlusion invariant feature 对底层视觉就不合适

大模型问题：通用意味着不知道丢掉什么信息是合适的

solution：将不变性特征解耦到不同维度

不压缩(丢失)特征信息

解耦学习十年前就有，但是通用大模型才是合适的应用

GLOM:

- 依据时间逐步解耦（类似resnet是前向就完成解耦，而人类则是多次前向（cc:DDPM???））
- 特征是层级化的
- 问题：contrastive loss来让信息不丢失（GLOM 人做梦来获得负样本 :））

reversible network来保证信息不丢失：既可以从输入算到输出，从输出算到输入
但过去的架构不合理：

revcol

架构适合大模型

- 维护一个column的feature
- 只有一个超参 column 数量

未来：

- 多模态除了对齐问题，还有上下文逻辑的问题
- 自主学习
- 生成式模型与判别式模型的同意
- 表征不变性的挖掘
- 视频的处理

盘古气象大模型：3D全球高分辨率气象预报方法

华为 谢凌曦

人类经验是否有缺陷？

有，那我们才需要ai

气象预防

ai相比较超算解数值模型，功耗视觉差10000+倍

fourcastnet, graphcast: 全球超高分辨率

era5 2000T

era6 2006

- 3D神经网络

- 绝对坐标！（全球）

-

盘古-weather: 首个超过传统方法的ai方法

ensemble: 每次输入加入一点小噪声

预测存在的问题，极端天气倾向于平滑，比如17级台风变成13级

问题是核心，不是解法！

合理的metric很重要

从计算机视觉到AI4Science-挑战与机遇

欧阳万里 上海ai lab

22年开始 ai4science

science: 高通量实验与计算 但是不懂人工智能，机器学习

与cv类似的挑战: **小样本 少标注 数据形式多样 可解释性**

独特的: 需要更多科学背景知识

已有方向:

人工智能加速研发材料, 加速电池研究

问题: 样本少到几乎可以说是无中生有

人工智能分子动力学

人工智能转录组学

人工智能+大气海洋模拟

风乌模型

时序问题解决数据问题: 数据即标注 (插帧, 下一帧blabla)

多模态 (温度, 风速, 对位差 etc是不同的模态):

不同模态不同编码器 + 单个transformer 模型

量子图机器学习及其化学应用初探

上海交通大学 严骏驰

通过量子力学特性来获得计算加速

技术方向:

- 超导电路
- 光学
- 离子
- 拓扑
- 半导体

经典计算的问题:

- 冯诺依曼体系算力满足不了现在的需求

量子系统 并行优势, 算法又是

quantum 网络 (cc:感觉有点像类脑计算的spiking NN, 模块替换)

量子力学用波函数描述粒子状态

利用变分原理解波函数($\langle \varphi | \hat{H} | \varphi \rangle$)

量子结构搜索 🤖

量子三维图表征

融合HPC和AI的药物分子设计

中山大学

十年十亿筛选一万分子10个临床才能出一个成药

人类基因组产生巨量数据

研究方向：蛋白质结构预测-靶标发现

如何对蛋白质进行合适表征，对化合物进行合适表征，信息融合？

蛋白质：先预测一维二维结构，再解决三维

BRNN-LSTM -> AlphaGo (三维)

一维二维是否足够描述3D

transformer 应用到化合物逆合成

Drug-VQA

更准确的化合物分子表示：

图卷积 orz

更准确的蛋白质表示：

图卷积orz

跨细胞图像尺度的药物设计

药物知识图谱

未来：

高性能计算+AI

非均匀物理数据分布的偏微分方程 (PDEs) 的通用求解算子

没听懂 orz

人工智能与科学技术panel

1. AI4Science领域各做各的，不成体系，未来是这样吗？

- 会是这样， 但不成体系也没有关系， 一个是AI4s是问题导向的， 二是科研鼓励多样性
- 但应当鼓励开源开放

2. AI4Science就业问题

- 10年前做计算机最好的工作可能是去中国移动， ai4science可能就是未来的就业大需求， 目前是蓝海， 而且目前是增长的行业
- 做图像的职位非常紧缺， 如果做AI4Science， 职位门槛低一些

- o ai4science 从science上能学会如何抽象，凝练问题
 - o 基础科学的爆发才是人类真正的跃迁
3. AI4Science是否需要自己的科学问题 (c:通解?)
- o 应当建立科学范式 (类似力学定理)
 - o 会有结合的更紧密地方法结合，需要专家的意见去改进更专有化的发展
 - o 要推动开源开放，并不一定要技术上有范式
4. AI4Science对研究者的要求
- o AI领域研究风格不好 (刷点，而不研究问题本质)
 - o 科研的Sense很重要 (找准问题，持续深耕，e.g., 盘古weather是一个人做的)
 - o 要有信仰，做研究要有深度有广度，要真正解决问题
 - o 关于刷点问题，不要因为点数拒稿，关注创新，点数只是证明，让新方法成长，发表，不要让一个方向dominate (PAMI副编，欧阳万里)
 - o 沟通能力很重要，同时要渴望能解决一个科学问题
5. AI4Science和生物机理的结合不够紧密
- o 建模问题的能力很重要
 - o 目前的合作确实比较浅层，希望学生的步调能慢下来，学生能比如博士阶段只做一两个有影响力的问题吗?
 - o AI的顿悟可能会出现来解决一些目前的故有问题
6. AI4Science如何促进数据共享
- o 要去思考为什么大家不愿意共享? 背后的忧虑是什么?

开放域物体识别算法

侯淇彬 南开大学

借助clip等文本编码器，帮助视觉在为训练的类别上完成预测。(open-vocabulary)

未来

从视觉内容生成到视觉任务自动化

面向开放域的虚实迭代强化学习

>张伟楠 上海交通大学

强化学习传统问题：仿真环境和真实环境存在gap

sim2real framework: simulate-env trained + real-world finetune

- adv: data efficiency + safety

具身智能!!

问题: real-world 也是local sample

终极目标: 单机器人适用所有场景所有任务

问:

gpt能用不, 交互性能很好啊?

- 语言指令的交互与控制模块的交互是不一样的, 两者应当有机结合

多模态开放域检测大模型及应用

梁小丹 中山大学

具身智能

visual-language pretraining -》 visual-language detection pretraining

detclip

concept enrichment

沿着concept enrichment继续想

unifying dense captioning and open-world detection pretraining

rm-prt: robot manipulation 数据集: 模拟了星巴克的咖啡厅

realistic robotic manipulation

future: robot-chatgpt

关键是simulator与platform

混合大模型

港大 罗平

主要讲 RAPHAEL这个工作

open-vocabulary

BG: DDPM

challenges: 人对视觉artifact 容忍度低,

思路: 不同的东西用不同的token画在一个区域里

动态开放环境的具身持续学习方法

张幸幸 启元实验室

continual learning “活到老, 学到老”

数据增量出现, 数据分布变化

灾难遗忘 (遗忘了关键知识)

与online learning 不一样

continual learning收到的数据任务可以是不一样的, 分布可以有本质区别

损失函数平坦性, 任务差异性

future: 具身持续学习

开放域panel

1. 开放域学习的挑战和难点

数据分布很难覆盖所有场景, 开放域的评价标准也很难