

VALSE 2023 无锡

视觉与学习青年学者研讨会

会议时间：2023年6月10-12日

会议地点：无锡太湖国际博览中心

主办单位：中国人工智能学会

中国图象图形学学会

承办单位：江南大学

无锡国家高新技术产业开发区管理委员会

协办单位：江苏省人工智能学会

无锡市计算机学会

中国图象图形学学会青年工作委员会

VALSE 2023 6-10

VALSE-大会特邀报告

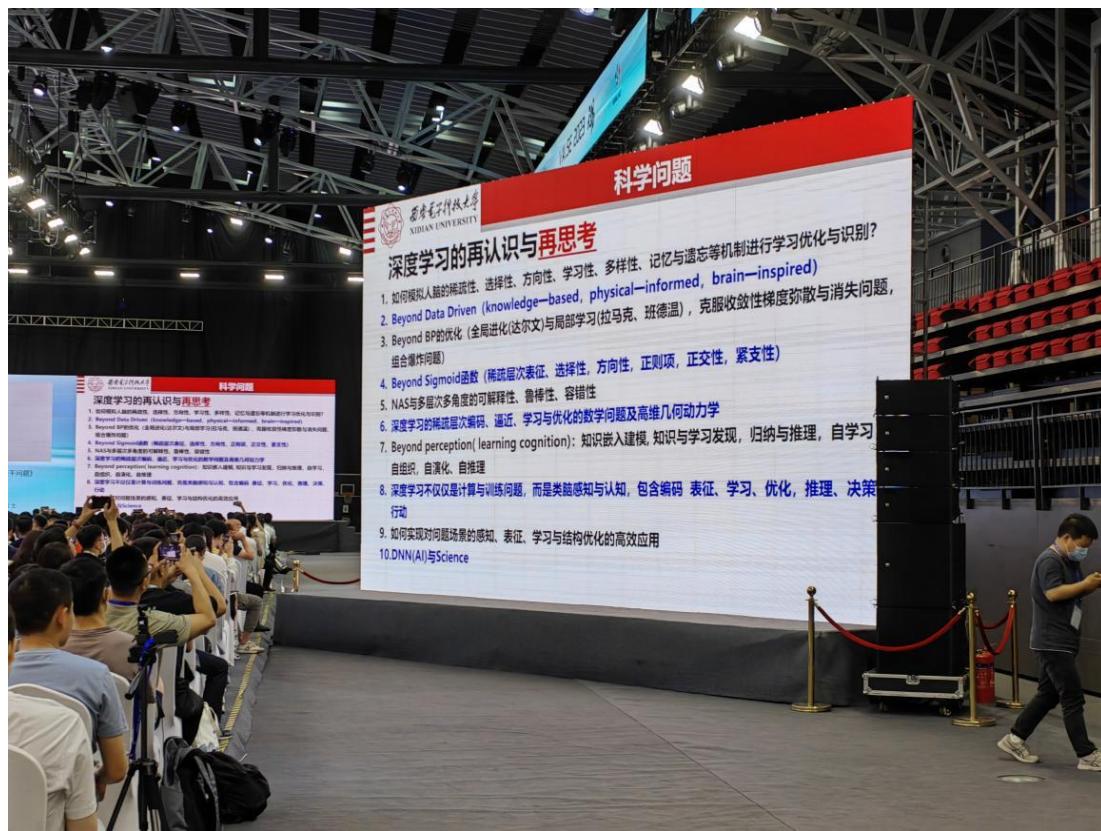
汇报人：焦李成

单位：西安电子科技大学

研究方向：智能感知与图像理解、深度学习与类脑计算、进化优化与遥感解译

人物简介：焦李成，欧洲科学院外籍院士，俄罗斯自然科学院外籍院士，IEEE Fellow。现任西安电子科技大学华山杰出教授、计算机科学与技术学部主任、人工智能研究院院长、智能感知与图像理解教育部重点实验室主任、教育部科技委学部委员、教育部人工智能科技创新专家组专家、国家级领军人才首批入选者、教育部长江学者计划创新团队负责人、“一带一路”人工智能创新联盟理事长，陕西省人工智能产业技术创新战略联盟理事长，中国人工智能学会第六-七届副理事长，IEEE/IET/CAAI/CAA/CIE/CCF Fellow，连续八年入选爱思唯尔高被引学者榜单。主要研究方向为智能感知与图像理解、深度学习与类脑计算、进化优化与遥感解译。曾获国家自然科学奖二等奖、吴文俊人工智能杰出贡献奖、霍英东青年教师奖、全国模范教师称号、中国青年科技奖、及省部级以上科技奖励十余项。

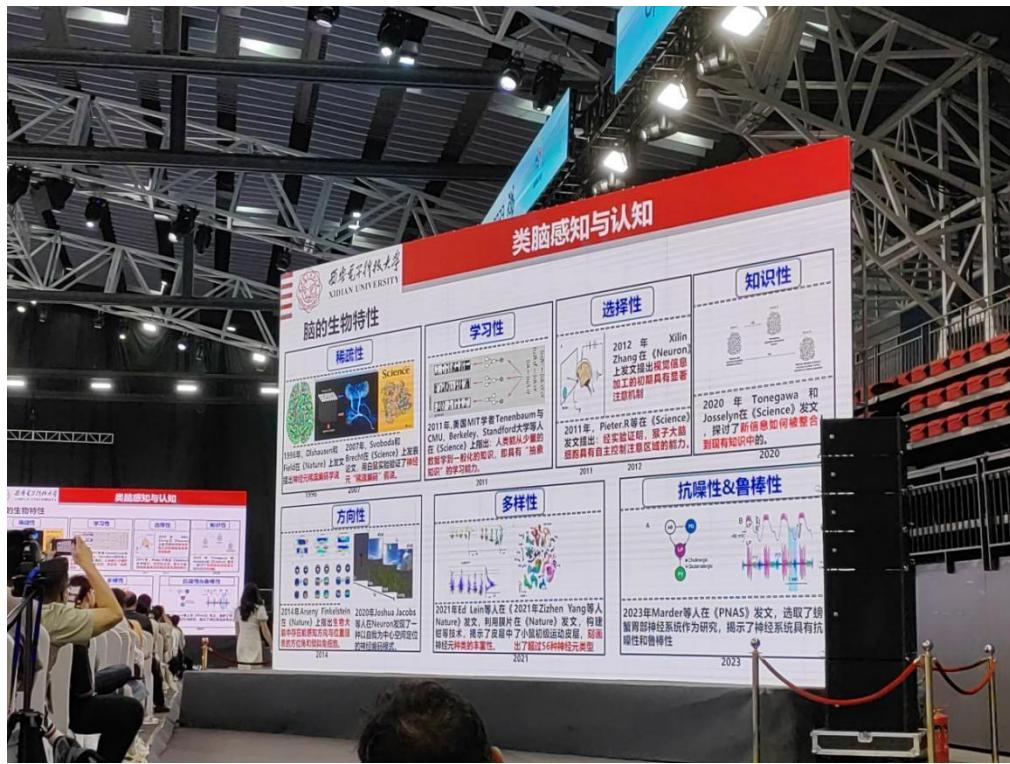
报告主题：下一代深度学习的思考与若干问题

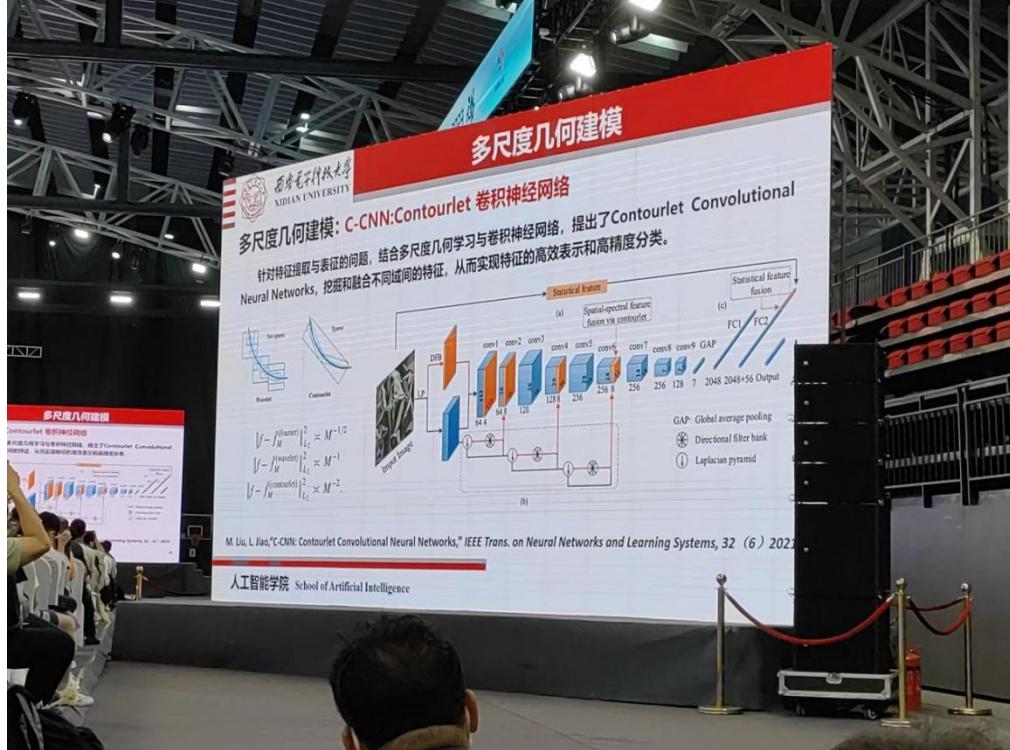
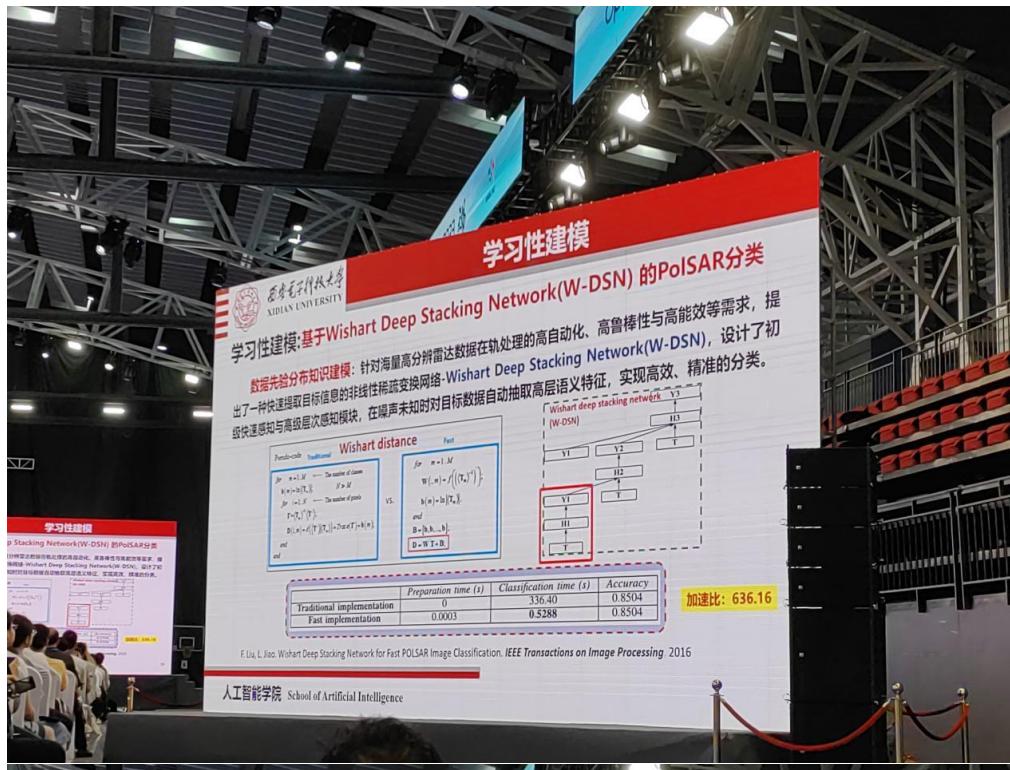


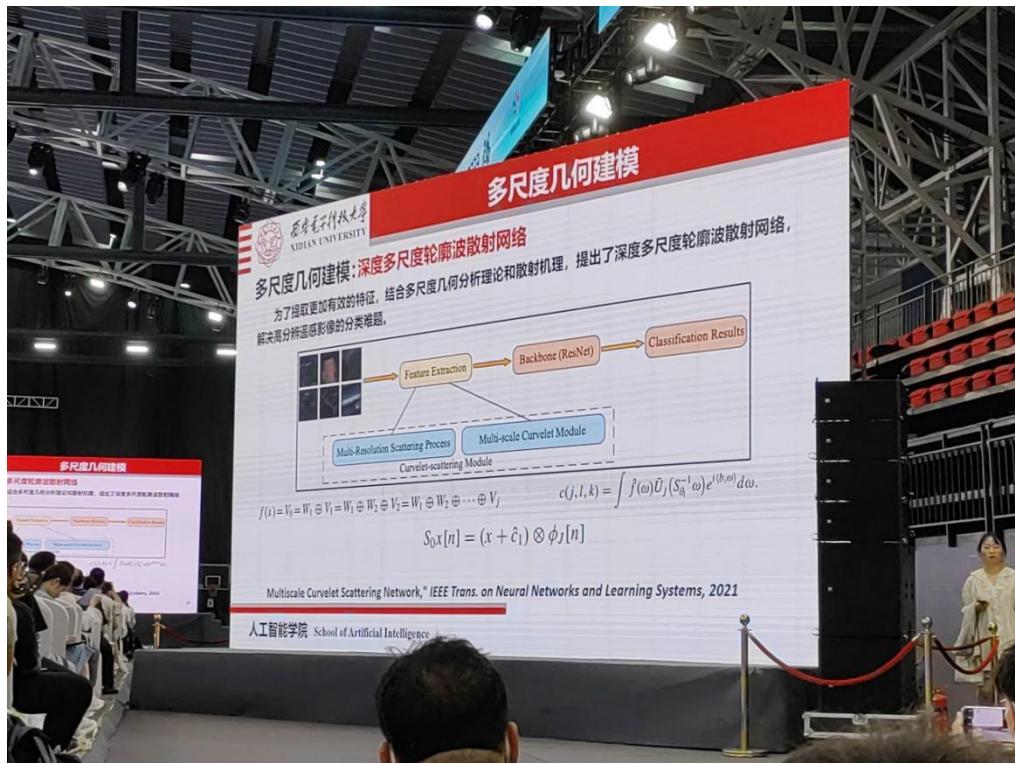
报告总结:尽管深度学习技术得到了长足的发展，并在诸多领域取得了显著成果。然而其在发展过程中，存在诸多理论问题，需要研究人员进一步研究和关注。因此，本报告着重探讨了深度学习基础理论相关的研究。首先，回顾了深度学习的思想起源与发展历程。随后，报告人讨论了对深度学习再认识与再思考，从而引出应突破的基础理论。



报告人从类脑启发、物理启发和进化启发等三个方面讨论了深度学习的表征、学习与优化理论。最后，给出了对下一代深度学习的一些思考：







汇报人：陈熙霖

单位：中国科学院计算技术研究所

研究方向：图像理解与计算机视觉、模式识别、图像处理、多模式人机接口

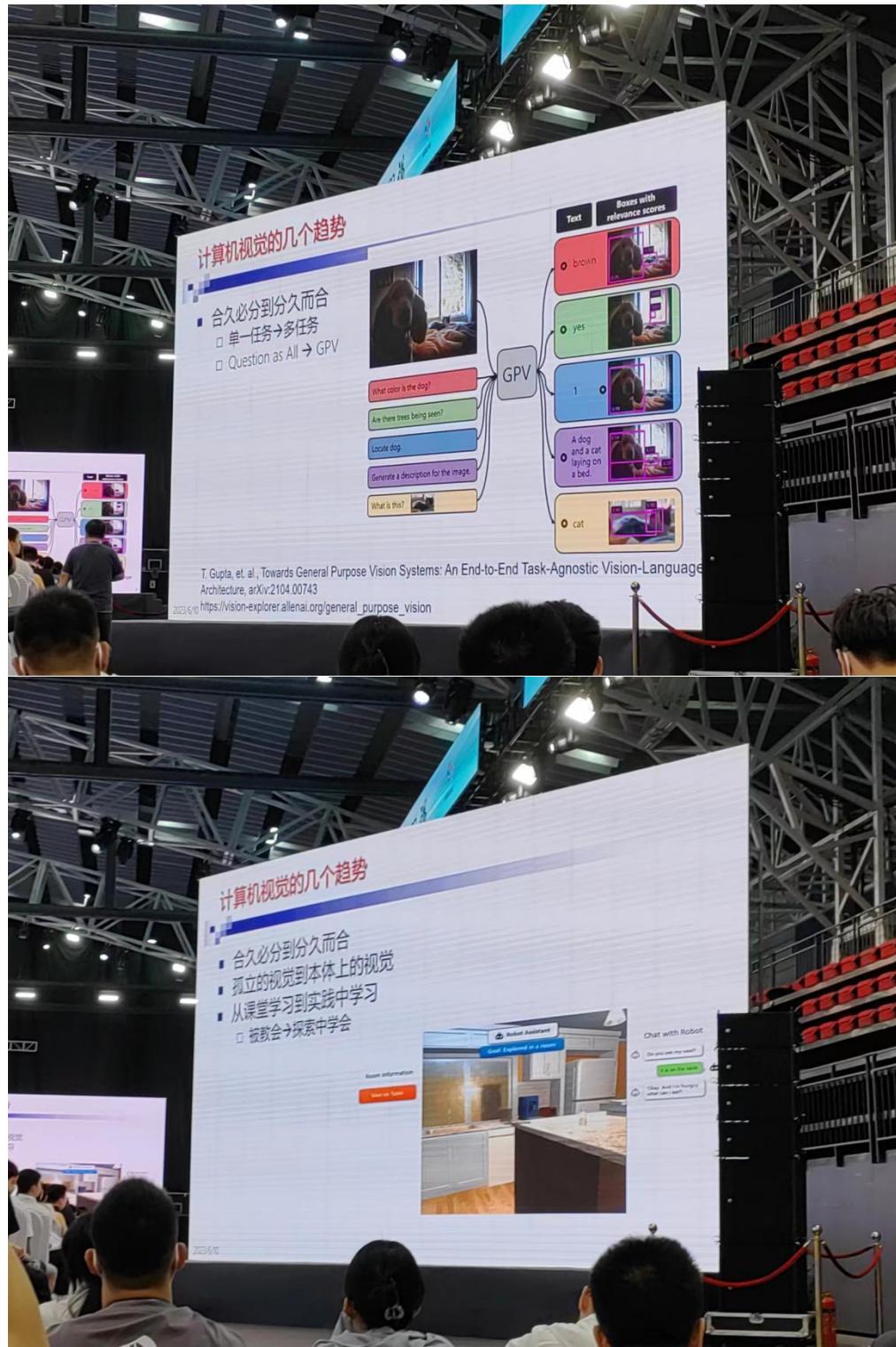
报告主题：计算机视觉--从孤立到系统性方法

个人简介：陈熙霖，中国科学院计算技术研究所研究员，ACM / IEEE / IAPR Fellow, 中国计算机学会会士。主要研究领域为计算机视觉、模式识别、多媒体技术以及多模式人机接口。先后主持多项自然科学基金重大、重点项目、国家杰出青年基金、973 计划课题等的研究。现任/曾任 IEEE Trans. on Image Processing 和 IEEE Trans. on Multimedia 的 Associate Editor、Journal of Visual Communication and Image Representation 的 Senior Associate Editor, 以及计算机学报副主编、人工智能与模式识别副主编等。担任过 IEEE FG 2013 / IEEE FG 2018 / IEEE VCIP 2022 等大会主席，并十多次担任 CVPR、ICCV、ECCV、NeurIPS 等会议的领域主席。



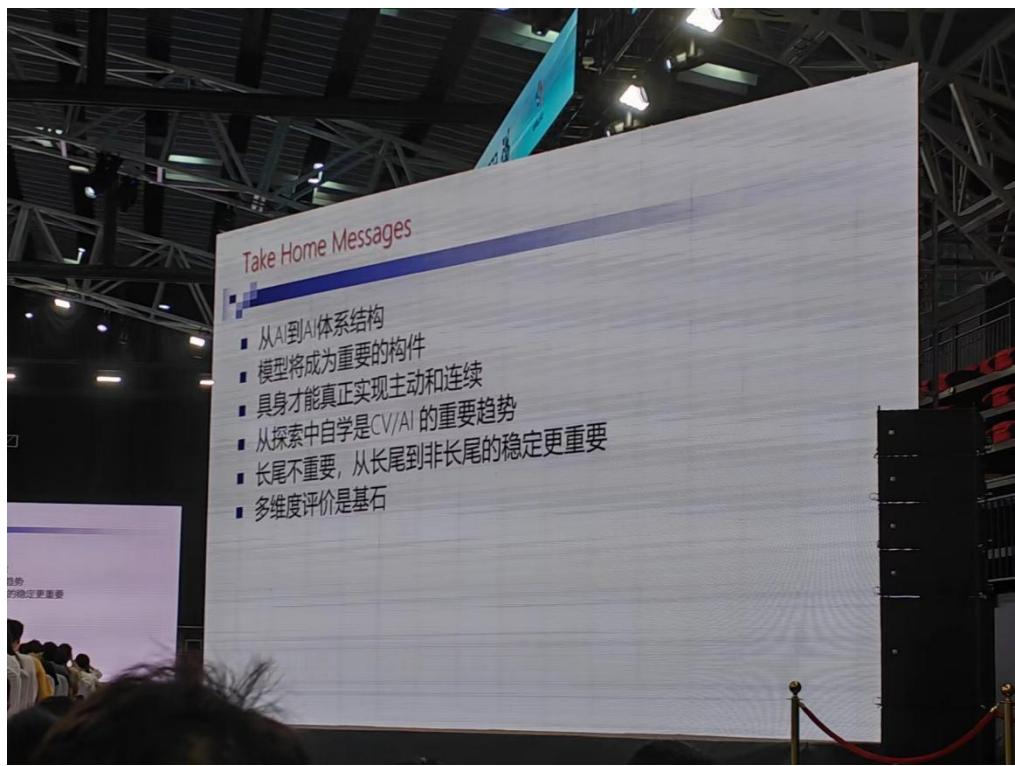
报告总结：在开放世界中通过观察积累、类比模仿到推理理解、交互尝试直至干预反馈是高等动物从外界获得经验与知识的基本手段与过程。在 AI 领域，很长时间以来的研究范式是以孤立算法为核心的单点研究，同时，现实世界中广泛存在着样本分布不均、任务多样性等问题。对以往的孤立研究范式而言，这些问题显然是难以克服的困难，因此需要从系统化的角度探索融合多模态信息，构建从感到探、从被动到主动的系统性学习体系。报告提出了计算机视觉研究的几个趋

势：“合久必分到分久而合”、“从孤立视觉到本体上的视觉”、“从课堂学习到实践中学习”。





最后报告人对于 CV 的发展趋势作了相关总结：



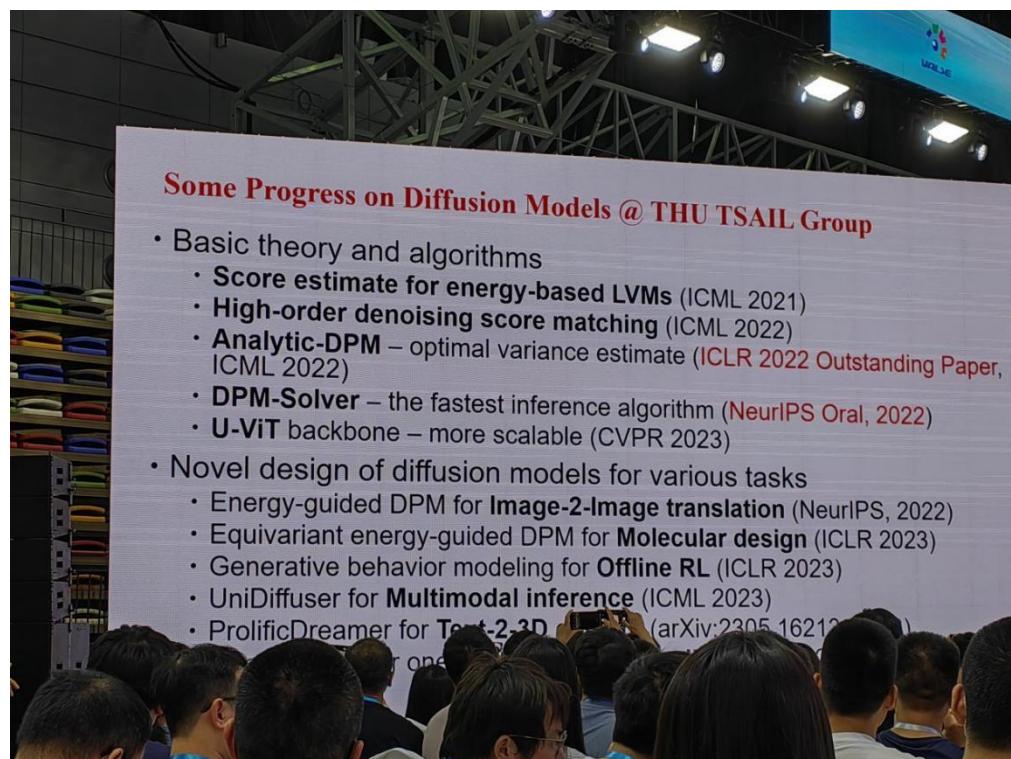
汇报人：朱军

单位：清华大学

研究方向：机器学习、贝叶斯方法、深度学习

个人简介：朱军，清华大学计算机系 Bosch AI 教授、IEEE Fellow，清华大学人工智能研究院副院长，曾任卡内基梅隆大学兼职教授。2001-2009 年获清华大学学士和博士学位，主要从事机器学习研究，担任国际著名期刊 IEEE TPAMI 的副主编，担任 ICML、NeurIPS、ICLR 等（资深）领域主席 20 余次。获求是杰出青年奖、科学探索奖、中国计算机学会自然科学一等奖、吴文俊人工智能自然科学一等奖、ICLR 国际会议杰出论文奖等，入选万人计划领军人才、中国计算机学会青年科学家、MIT TR35 中国先锋者等。

报告主题：扩散概率模型的前沿进展



报告总结：AIGC 发展迅速，扩散概率模型是 AIGC 的关键技术之一，在文图生成、3D 生成等方面取得显著进展。该报告将介绍扩散概率模型的若干进展，包括扩散概率模型的基础理论和高效算法、大规模多模态扩散模型以及 3D 生成等内容。报告只是针对扩散模型在 CV 领域中各个方向的应用做了大致描述与介绍，并没有详细梳理扩散模型发展的脉络。更为详细的脉络可见 tutorial：李崇轩博士对扩散模型的介绍。

VALSE 2023 6-11

Workshops-围绕手机的计算影像学

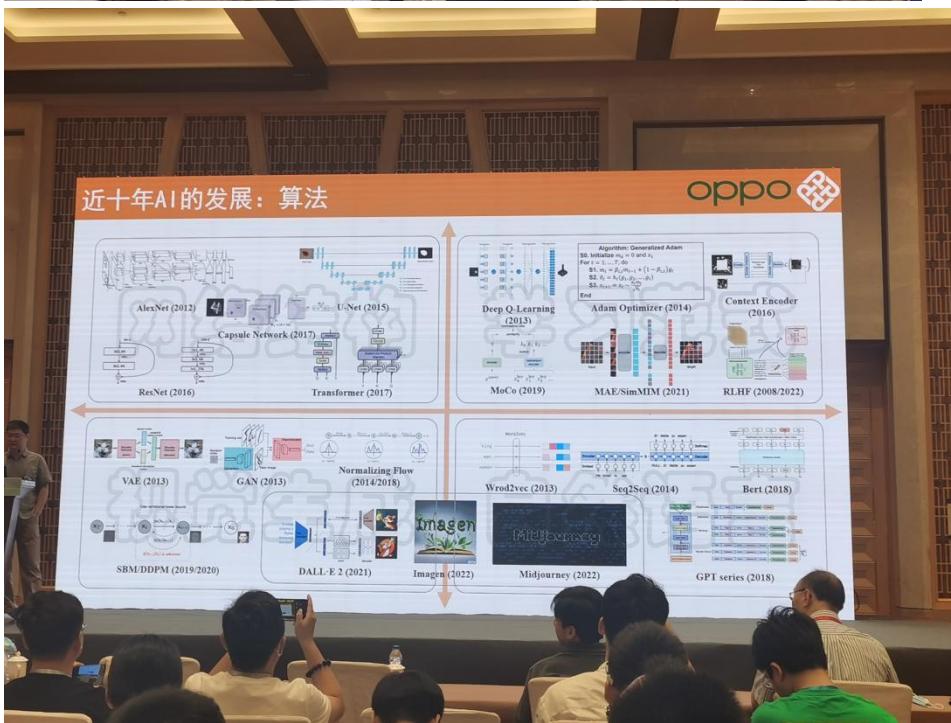
汇报人：张磊

单位：OPPO

个人简介：张磊教授（IEEE Fellow）于 2006 年加入香港理工大学电子计算学系，2017 年起任职讲座教授。张磊教授长期致力于计算机视觉、图像处理、模式识别等方向的研究，是底层视觉方面的国际权威学者。张教授是 IEEE Trans. on Image Processing (TIP) 的高级编委, IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI) 、SIAM Journal of Imaging Sciences 等多个国际期刊的编委。从 2015 年至 2022 年，张教授连续被评为 ClarivateAnalytics Highly Cited Researcher。张磊教授目前也于 OPPO 研究院任 AI 科学家，从事 AI 影像前沿技术的研发。

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

报告总结：对近年来的底层视觉任务的相关处理算法进行了一个总结。底层视觉任务包括但不限于图像恢复、去噪、去模糊、超分辨、去雨、去雾、去反射等等。将现有基于学习的算法分为四大类：无监督/自监督 (zero-shot) ,多退化过程建模，网络结构改进和多任务通用模型。系统的分析了学术界和工业界对于底层视觉问题处理上的差异，最后指出未来的研究方向在于通用的图像恢复大模型，但难点在于退化过程复杂多样。



Tips: AI 领域, 算法 (包括网络结构、训练方式、损失函数) 都有很快的发展和变化, 但也开始出现一些瓶颈。



Tips: 近年来，大模型的快速发展逐渐打破瓶颈，有着一统 AI 界的趋势



Tips: 回顾一下早期基于 AI 的在各种任务（超分、去噪、去模糊等等）上的图像复原、增强方法（其实有很多都是他们组的工作）



Tips: 逐渐发展至今，从实现的角度来说，主要现有的 AI 方法主要分为四大类：无监督、退化建模、结构改进、多任务通用。（我们前面的工作属于从无监督出发，在往多任务通用模型上面靠。）



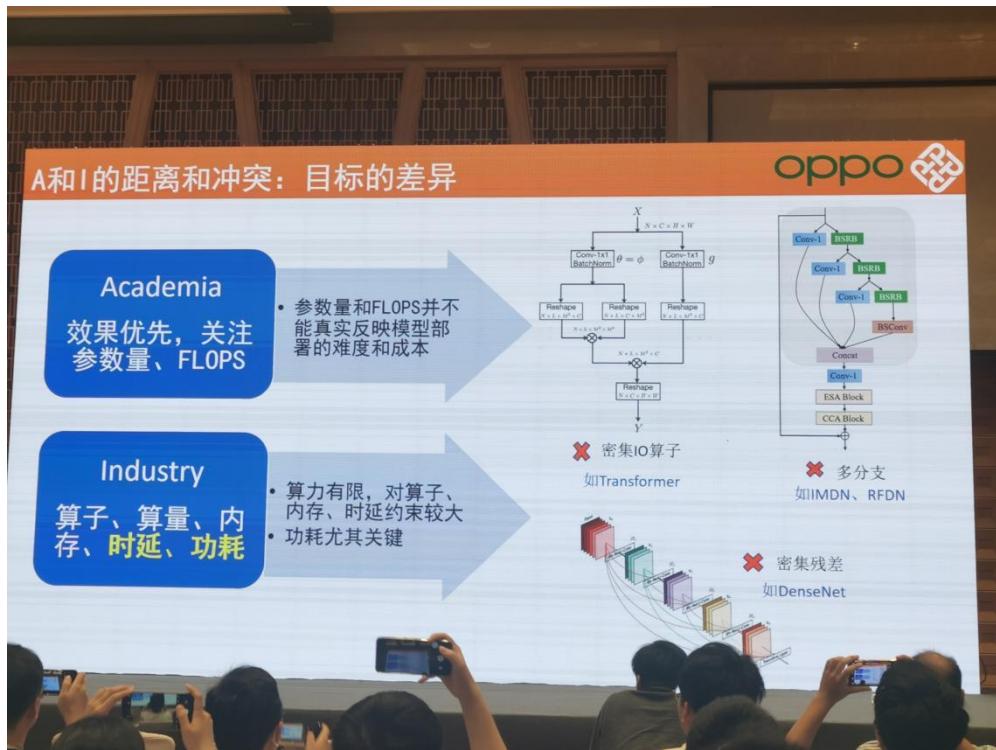
Tips: 总结学术界和工业界都在关心什么问题，都在刷指标，虽然指标上性能好，但是实际有没有用需要打个问号。但学术和工业是分不开的，做学术的应该多考虑工业中的实际问题，才能做出更有用的算法出来。



Tips：学术界哪怕是大家都在用的公开数据集，图像退化与实际问题中的 Real-image 仍有较大差距。



Tips：哪怕是大家都在用的客观评价指标 PSNR、SSIM，在有真值的情况下也不能完全反应算法性能的好坏。



Tips: 学术界考虑的计算复杂度，与实际相差较大。



Tips: 数据的问题。



Tips: 评价指标的问题



Tips: 现有方法聚焦于各自任务场景，不具备通用性，极需通用的大模型解决所有的图像复原问题。



Tips: 联系方式，对我们的启发在于如何建模更接近 real image 的退化，以及在设计算法的时候考虑对不同退化场景的统一求解，但难点在于这种大模型对于计算资源要求比较高，理论性也不一定强，一般的组可能都做不了。

汇报人：戴玉超

单位：西北工业大学

个人简介：戴玉超，男，西北工业大学电子信息学院教授、博士生导师，国家级青年人才。主要研究工作集中在机器视觉、智能感知、图像处理、人工智能等领域，聚焦复杂动态场景的三维重建与感知、深度学习和几何模型融合的稠密匹配、新型仿生视觉传感器和计算成像等问题。主持国家自然科学基金、科技部科技创新 2030 “新一代人工智能”重大研究计划子课题、JKW 领域基金重点项目等科研项目。近年来在 TPAMI、IJCV、ICCV、CVPR、NeurIPS 等国际顶级期刊和会议上发表论文 70 余篇，谷歌学术引用超过 7700 次，H 因子 43。先后获得 CVPR 2012 最佳论文奖（大陆高校 30 年来首次获得该奖项）、陕西省自然科学奖一等奖、中国图象图形学学会青年科学家奖、火箭军“智箭火眼”人工智能挑战赛全国第一名、IEEE CVPR 2020 最佳论文奖提名、ECCV 2020 鲁棒计算机视觉挑战赛双目深度估计赛道冠军和光流估计赛道亚军、CVPR 2017 非刚性结构与运动恢复挑战赛最佳算法奖、APSIPA 2017 年度峰会最佳深度学习/机器学习论文奖等奖项。担任 APSIPA 杰出讲者和 CVPR、ICCV、NeurIPS 等国际顶级会议领域主席，ACCV 2022 宣传主席，中国图象图形学报青年编委。

报告主题：卷帘快门相机：模型、优化与学习



Tips：因为考虑的问题不是很相关，只能从方法的思路上进行理解，从建模、优化和学习三个方面进行科研是比较好的方式和习惯，值得我们学习。问题建模从

卷帘快门相机的成像机理出发，逐行或逐列的退化相同，短时间则可简单建模成匀速线性运动模糊核。优化和学习则分别采用帧与帧之间的先验，逆过程先验、光流估计等辅助求解，并从 2D 推广到 3D，考虑的问题场景由易到难，全面系统的阐述了卷帘快门相机的成像提高应该考虑的因素，可以说是卷帘快门相机成像的专家了。

汇报人：贾旭

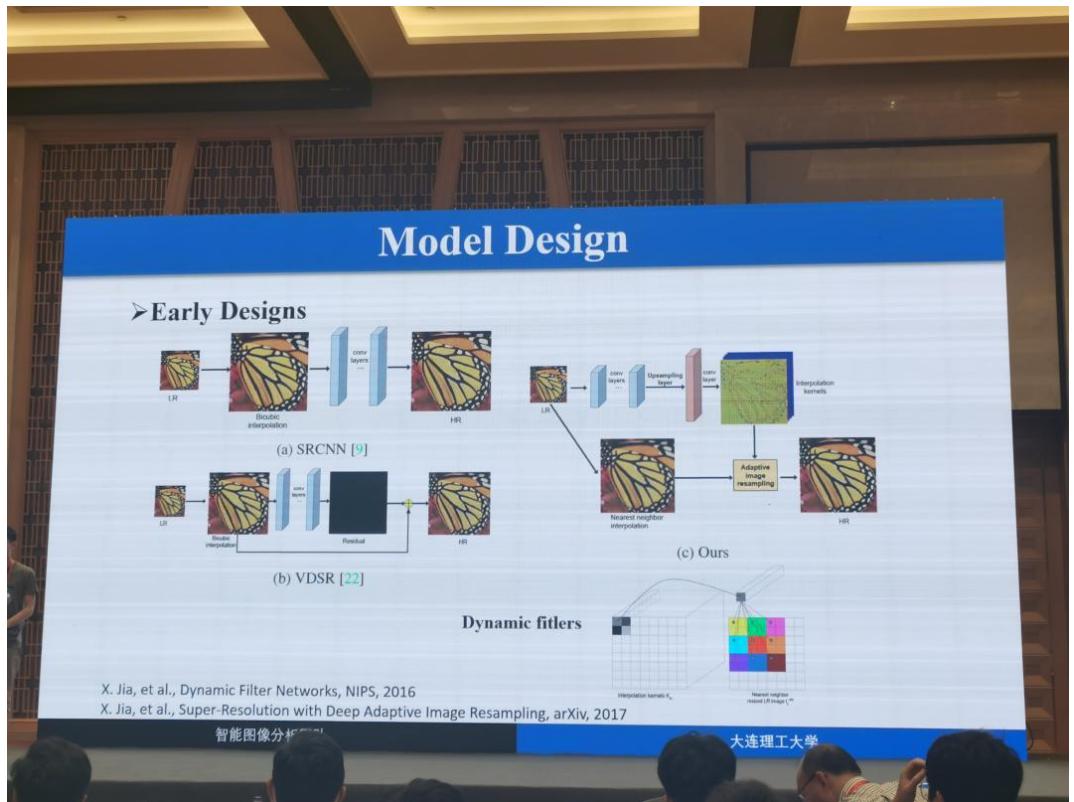
单位：大连理工大学

个人简介：贾旭，大连理工大学人工智能学院院长聘副教授，辽宁省智能感知与理解人工智能重点实验室骨干成员，博士毕业于比利时鲁汶大学，是从 Tinne Tuytelaars 教授和 Luc Van Gool 教授。主要研究方向包括图像和视频的增强与生成、视觉目标检测跟踪以及类脑视觉。迄今在计算机视觉和机器学习领域顶级会议及期刊发表论文 30 余篇, Google Scholar 引用 6000 余次, 申请国内外专利 10 余项。曾在 Google Research, 商汤科技, 华为诺亚方舟实验室等从事研究工作。主持并参与国家自然科学基金、科技部科技创新 2030 重大项目、科技委项目以及华为等企业合作项目若干项。

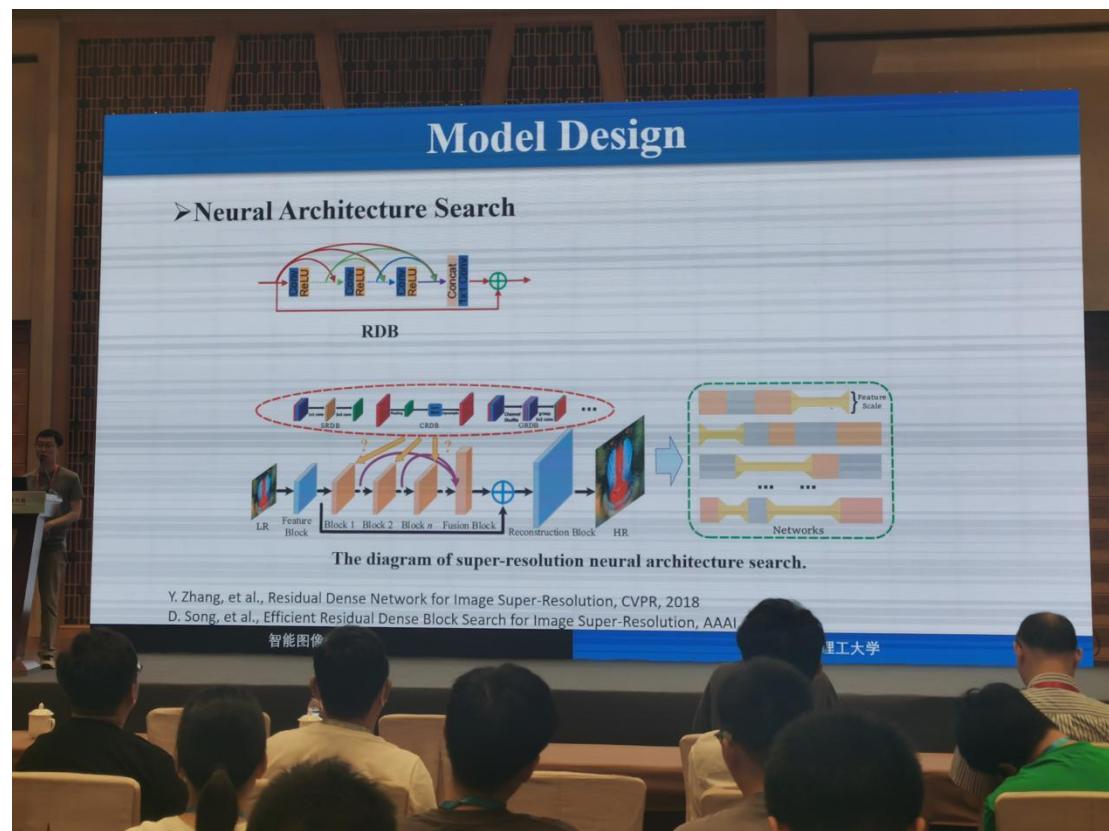
报告主题：影像的时空分辨率增强

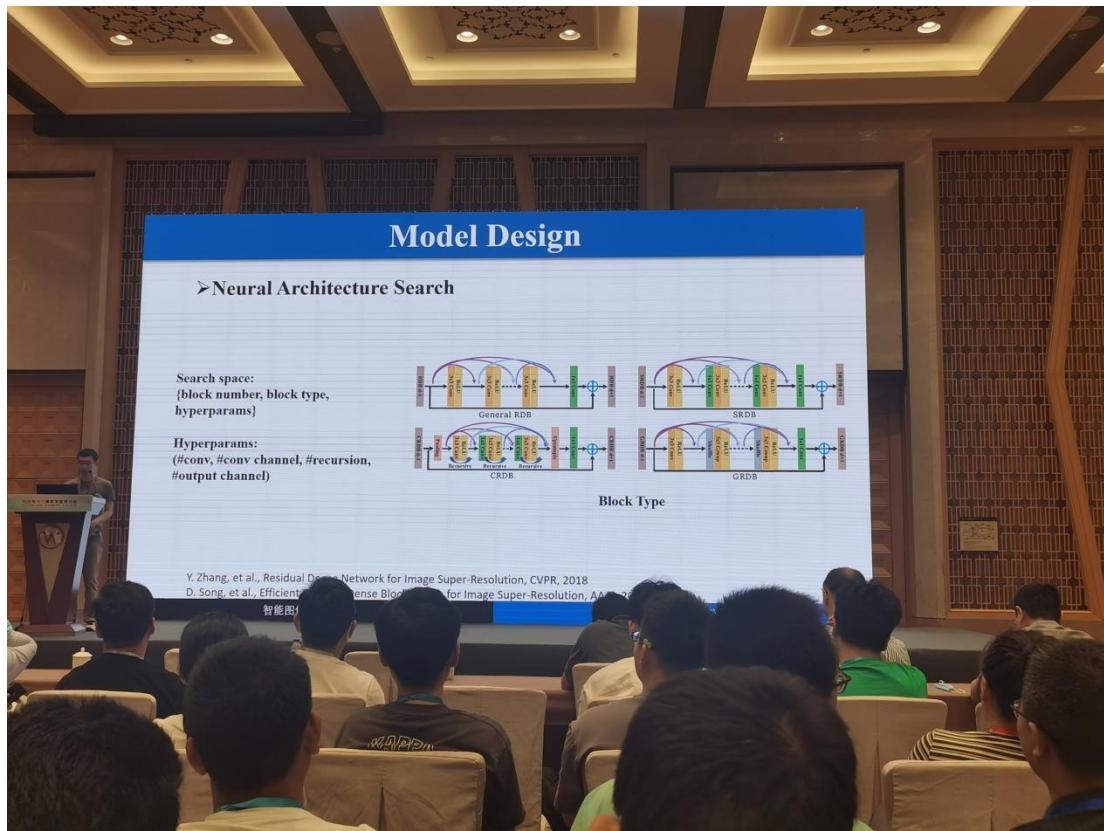
报告总结：对于底层计算机视觉任务，对不同位置通过编码的方式预测不同的模糊核，以及网络结构搜索展开了一系列研究。发现 transformer 具有较好的细节纹理特征提取学习能力，将 CNN 提取的特征与 transformer 相结合。最后针对 DASR 的工作进行了进一步的改进，将无监督学习到的退化编码进行了进一步的排序，提高了 BSR 的算法性能和泛化性能。



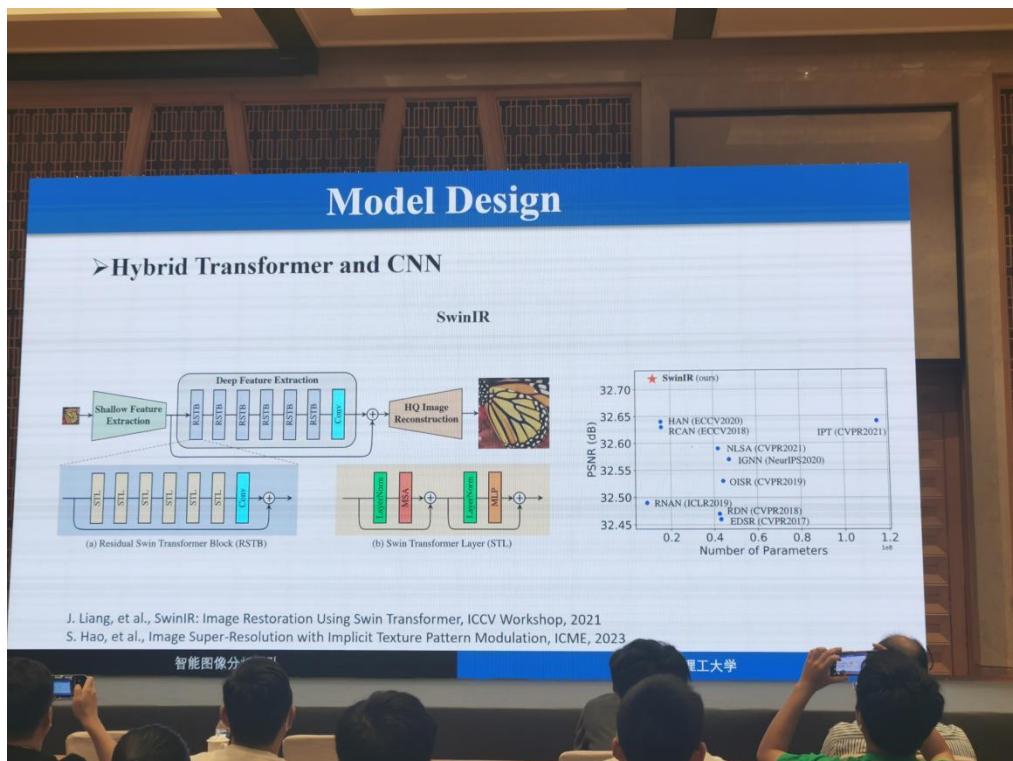


Tips: 早期的模型设计比较简单粗暴，图像每个位置的噪音、模糊、下采样都是一样的，但是实际中可能会更加复杂。

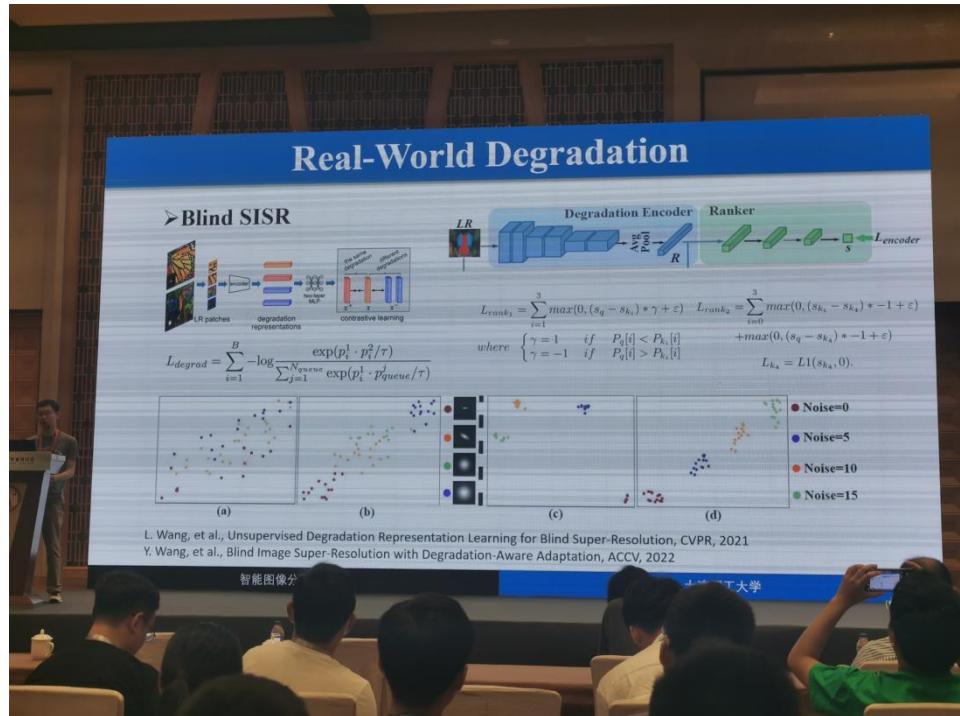




Tips: 对于网络结构，大部分方法采用固定的网络结构，但其实也有网络结构搜索（NAS）的方法，可供我们参考，大概思路是在几种 block 和几种链接方式上进行搜索。他们也做了一些工作和创新。

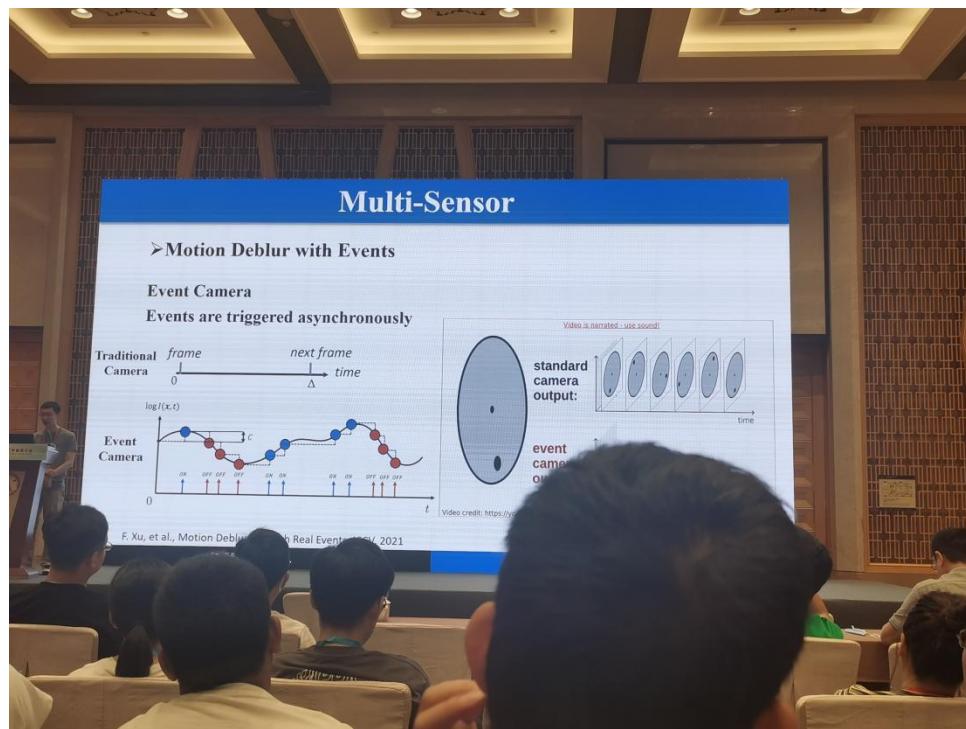


Tips: 他们发现 transformer 提取到的图像特征与 CNN 不同。(有点 heuristic)
他们把这里俩特征结合了一下，SR 效果更好。



Tips: 他们把 DASR 的退化因变量进行了进一步的排序，提高了算法性能。这里的排序还需要进一步看论文[1]研究一下。

[1] Blind image super-resolution with degradation-aware adaptation



Tips: 多传感器，对我们参考意义不大。



Tips: 指出了三个方向，
1: diffusion 可以用来作为基础模型；
2: 多变、复杂的退化需要大量的 pre-training；
3: 多传感器融合也是一个方向。

汇报人：薛天帆

单位：香港中文大学

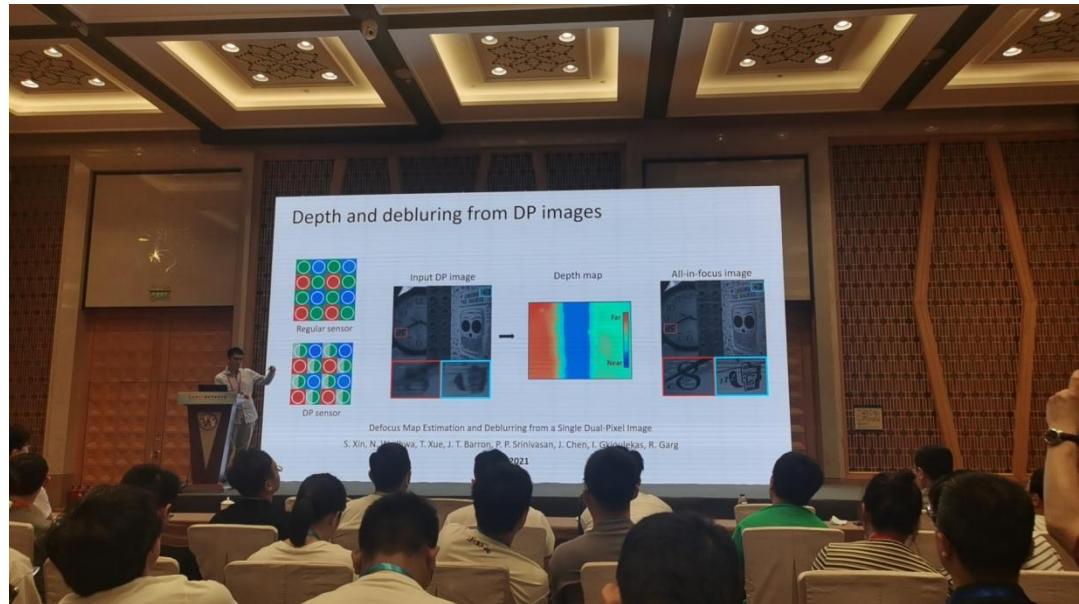
个人简介：薛天帆教授是香港中文大学讯息工程系的 vice chancellor assistant professor。在此之前，他在谷歌研究院的计算摄影团队担任主任工程师，工作了五年以上，有着丰富的产学研结合经验。他毕业于麻省理工学院计算机科学与人工智能实验室，导师为 William T. Freeman 教授。2011 年在香港中文大学讯息工程系获得哲学硕士(M.Phil.)学位，2009 年在清华大学计算机科学与技术系获得本科学位。他的研究重点是计算摄影、计算机视觉和图形学以及机器学习。他的多项研究技术在学界和工业届都有很大反响：他研究的去反射技术被 Google Photoscan 使用，拥有超过 1000 万用户；他研究的快速双边学习技术被整合进了 Google Tensor 芯片；他研究的夜景算法活动 DPRReview 年度最佳创新奖，并被应用与谷歌 Pixel 手机中，也启发了很多其他夜景算法的开发。他也同时担任多个顶级会议和期刊的审稿人工作，并担任了 CVPR 2020 的网络主席，WACV 2023 和 CVPR 2023 领域主席。他的研究领域涵盖了计算摄影、机器学习、计算机视觉和计算机图形学。特别地，他致力于构建智能摄像头系统，并与视觉世界进行交互。这些技术可以惠及多个领域，包括人工智能视觉系统、手机摄像头、自动驾驶和医学成像等。

报告主题：如何设计一个真正意义上的智能相机

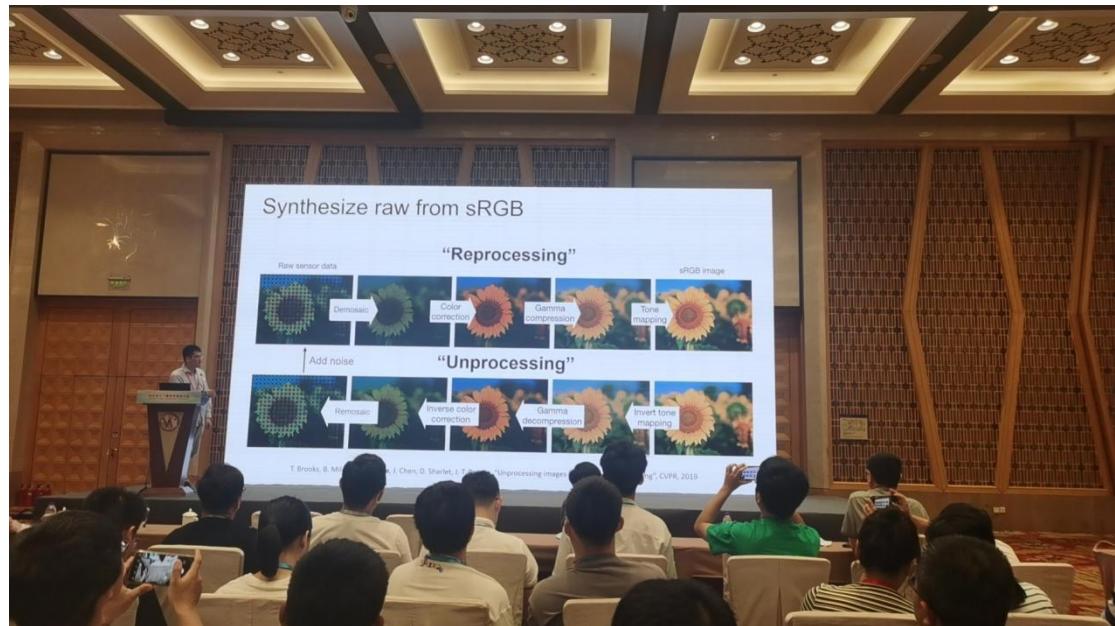
报告总结：针对单张图像本身信息量有限的问题，考虑在成像端使用多个摄像头，单张图像考虑深度图，针对 noise 和 blur 的配对数据（符合真实场景的）难以获得，以合成真实噪声和炫光两个场景为例进行了介绍。最后在介绍了一种将传统方法与深度学习方法相结合的方法，在图像风格迁移场景下进行了验证。



Tips: 从流程上看，图像包括传感器获取回波，然后成像，然后增强处理三个步骤，现有方法大多聚焦于第三个，但是前两个也应该考虑。



Tips: 传感器端，多传感器信息融合。



Tips: 成像阶段，考虑基于成像过程的退化



Tips: 最后算法端

汇报人：张健

单位：北京大学

个人简介：张健，北京大学深圳研究生院信息工程学院助理教授/研究员、博导生导师，深圳市海外高层次人才。分别于 2007 年、2009 年、2014 年获得哈尔滨工业大学数学与应用数学学士学位、计算机科学与技术硕士学位及计算机应用博士学位，并先后在北京大学、香港科技大学和沙特国王科技大学做博士后访问研究。主要从事底层视觉、计算成像以及 AIGC 方面研究，共计发表包括 TPAMI、IJCV、TIP、CVPR、ECCV、ICCV、ICLR 等高水平权威国际期刊/会议论文 100 余篇。目前，负责北京大学视觉信息智能学习实验室(VILLA)。近两年，带领 VILLA 以第一或通讯作者在 CCF A 类会议/期刊上发表论文 20 余篇，Google Scholar 引用超过 4500 次，h-index 值为 35，i10-index 值为 63。多次获得国际/国内期刊会议最佳论文奖，连续三年入选人工智能与图像处理领域全球前 2% 顶尖科学家榜单。**个人主页：**<https://jianzhang.tech/>。

报告主题：Zero-Shot Image Restoration Using Denoising Diffusion Null-Space Model

Most existing Image Restoration (IR) models are task-specific, which can not be generalized to different degradation operators. In this work, we propose the Denoising Diffusion Null-Space Model (DDNM), a novel zero-shot framework for arbitrary linear IR problems, including but not limited to image super-resolution, colorization, inpainting, compressed sensing, and deblurring. DDNM only needs a pre-trained off-the-shelf diffusion model as the generative prior, without any extra training or network modifications. By refining only the null-space contents during the reverse diffusion process, we can yield diverse results satisfying both data consistency and realness. We further propose an enhanced and robust version, dubbed DDNM+, to support noisy restoration and improve restoration quality for hard tasks. Our experiments on several IR tasks reveal that DDNM outperforms other state-of-the-art zero-shot IR methods. We also demonstrate that DDNM+ can solve complex real-world applications, e.g., old photo restoration. Code: <https://github.com/wyhuai/DDNM>.

报告总结：提出了一种基于 diffusion 的单张图像恢复方法，他们任务传统迭代算法本质上也是在做去噪，diffusion 也是在做去噪，从优化的角度可以很好的将二者结合起来，做一个不需要预训练的在线迭代更新生成的 stable diffusion model，并在图像超分、去模糊、去噪等多个应用场景都有着非常广泛的应用，并且文章中有推导和证明，跟我们的工作相关性较大，可以好好深度研究并 follow。



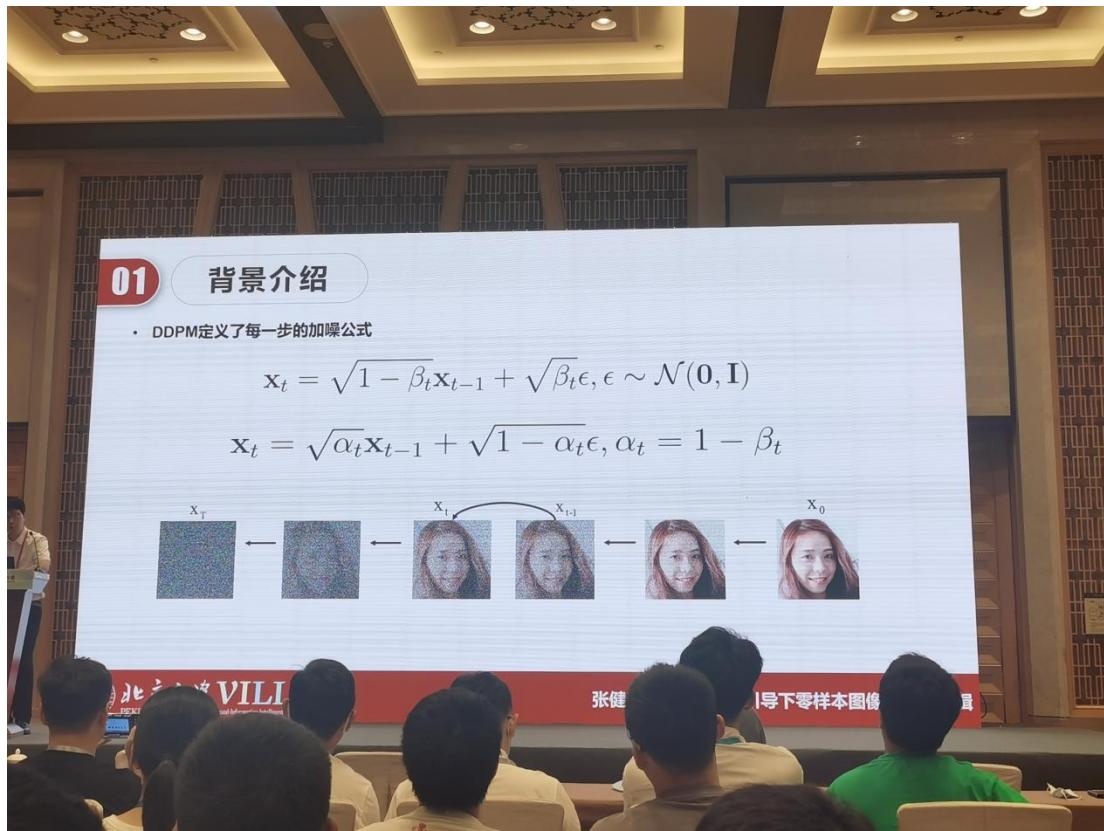
Tips: 首先将 diffusion 的基础知识和背景，然后推广到 zero-shot 任务场景。



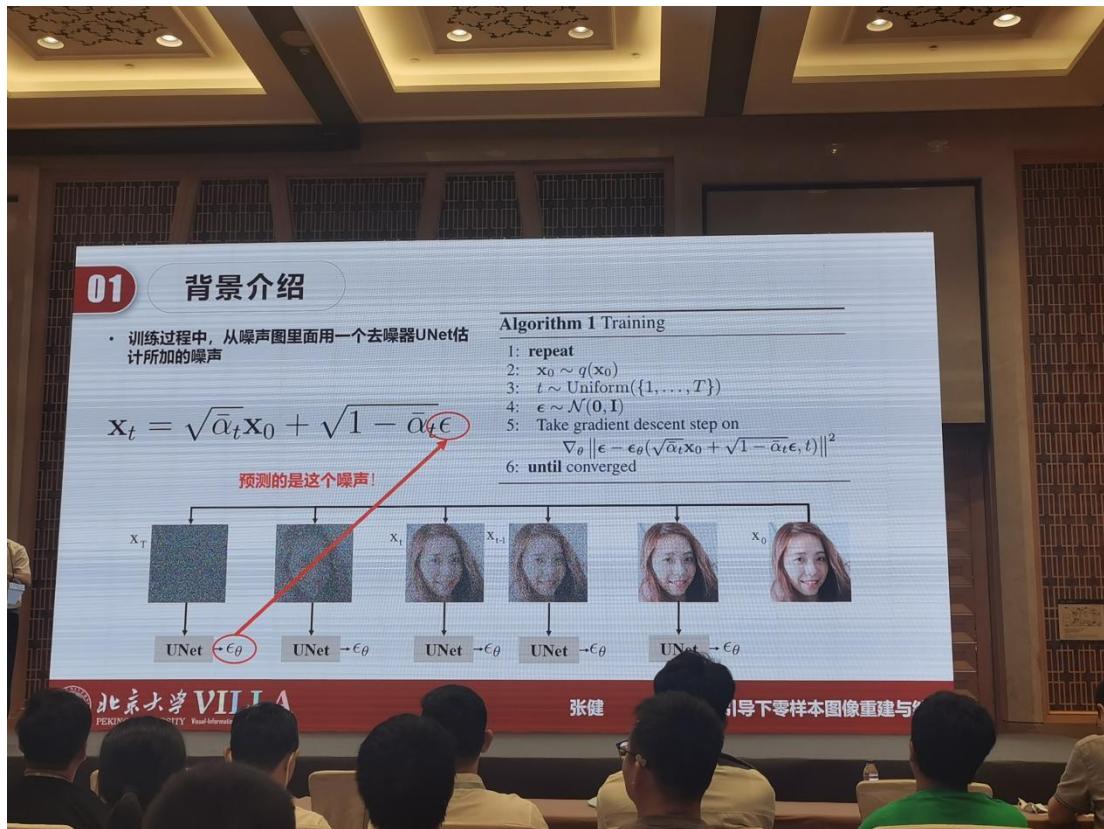
Tips: Diffusion 性能非常好非常强大，在各个领域的生成任务上都能起到很好的效果，但是需要大模型、大量数据和大量的 GPU 计算资源。



Tips: 分为扩散阶段和采样阶段两个过程



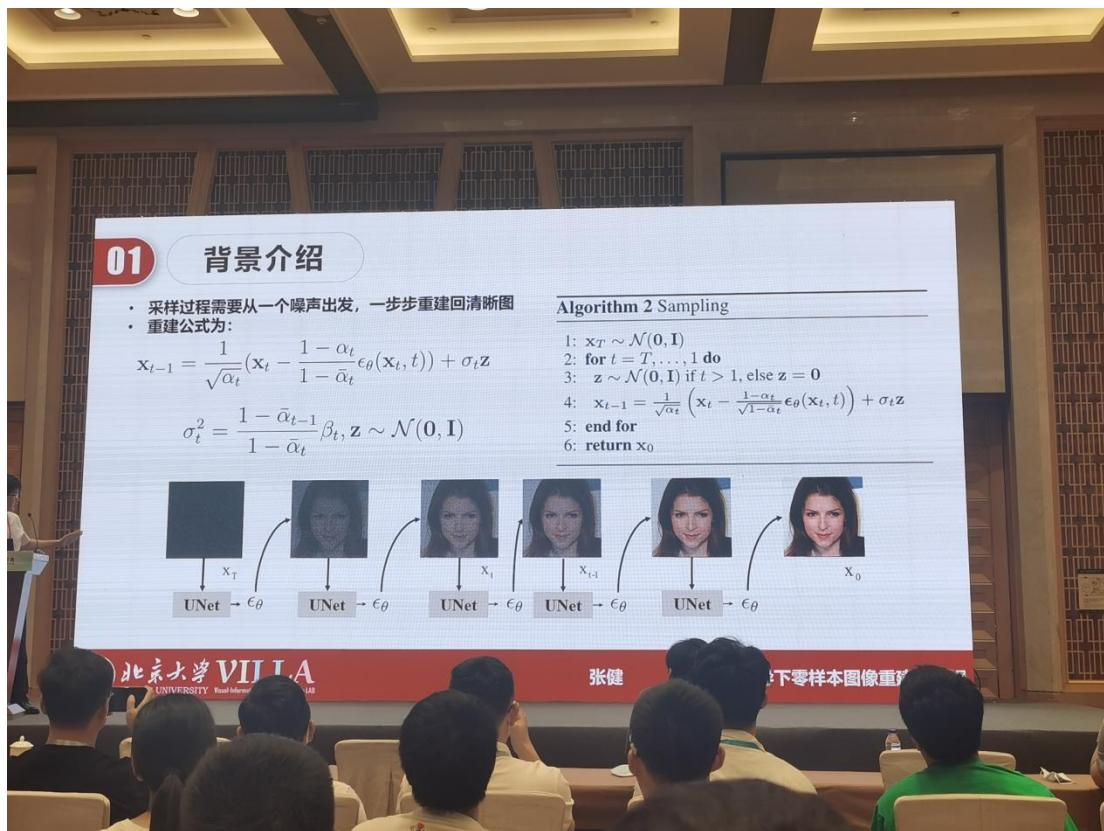
Tips: 加噪公式，原文中有推导证明



Tips: 网络预测的是噪声



Tips: 时间异步的网络参数共享



Tips: 采样恢复过程中，逐步去掉预测的噪声



总结：DDPM 两个阶段都有核心更新过程的公式，还有证明推导，DDPM 是最基本最原始的 Diffusion model，有很多理论证明过程在里面。



Tips：经典的图像恢复任务，压缩感知、核磁共振、医学成像、雷达成像、图像增强等领域都有着非常强大的应用。



Tips: 主流方法分为传统迭代优化求解算法和基于黑盒神经网络的方法



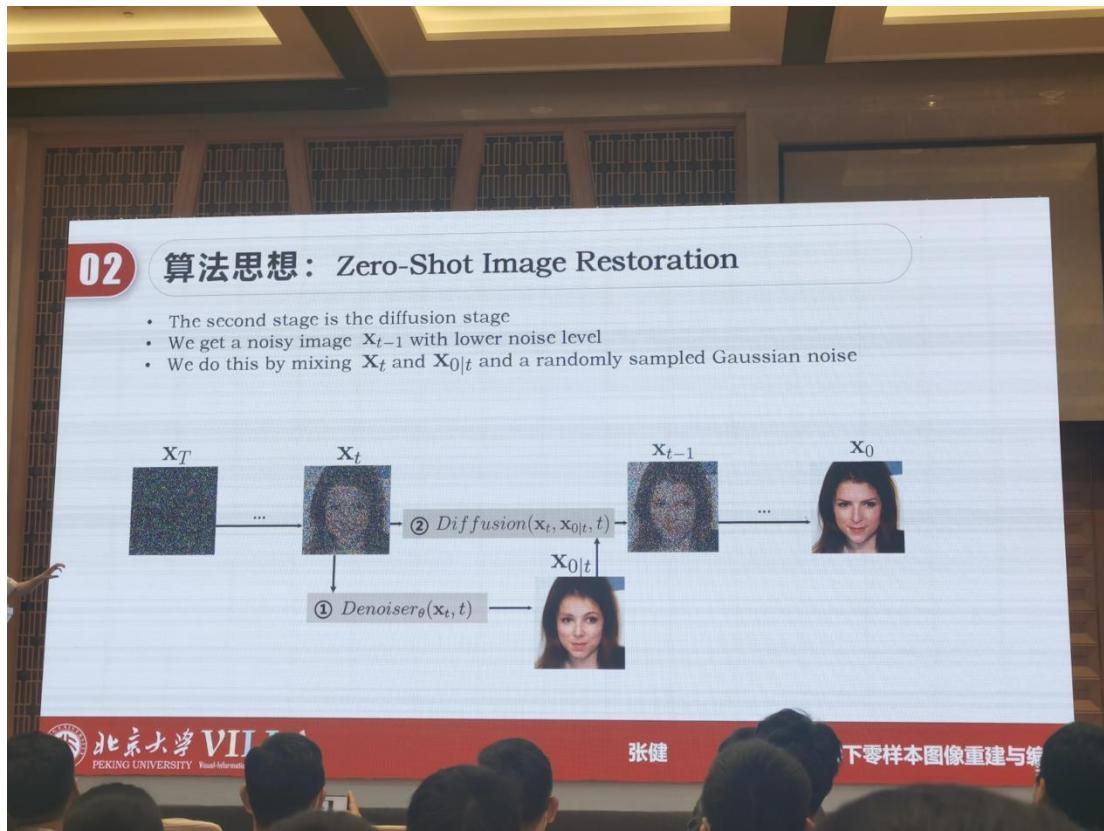
Tips: ISTA-Net, 其实就是 Unfolding



Tips: 主要分为图像复原和图像编辑任务两方面来解释



Tips: 两方面应用介绍所提算法，有开源代码



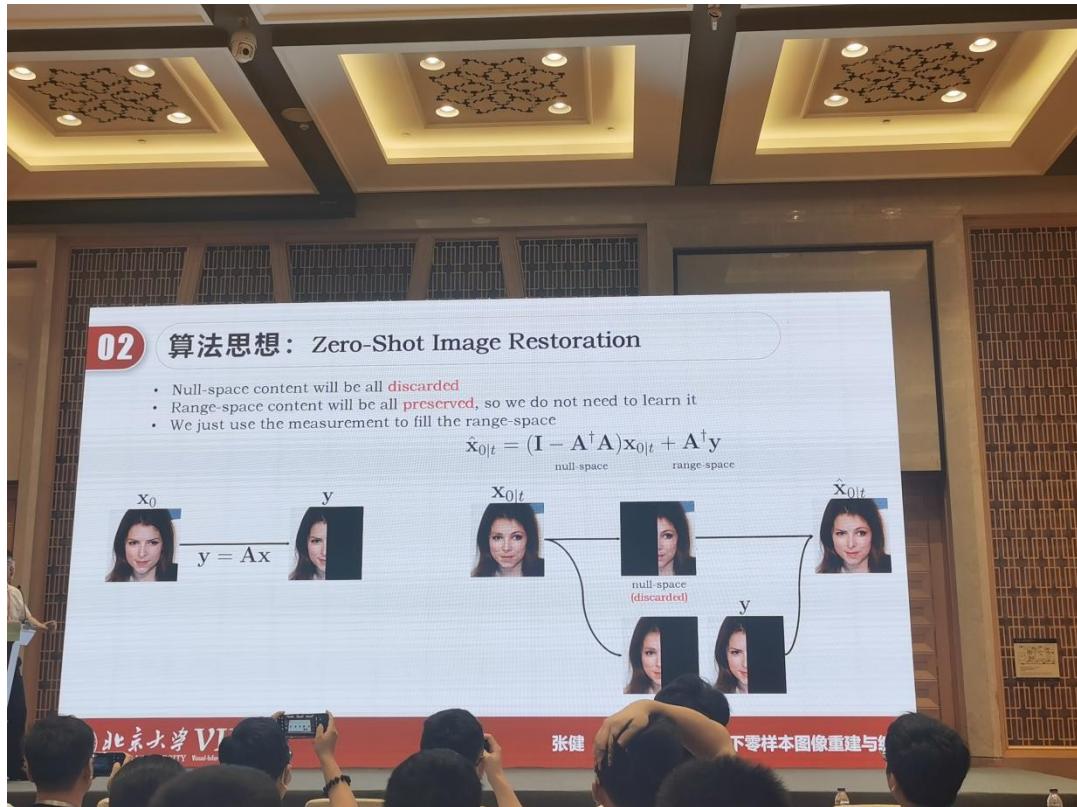
Tips: 算法思想



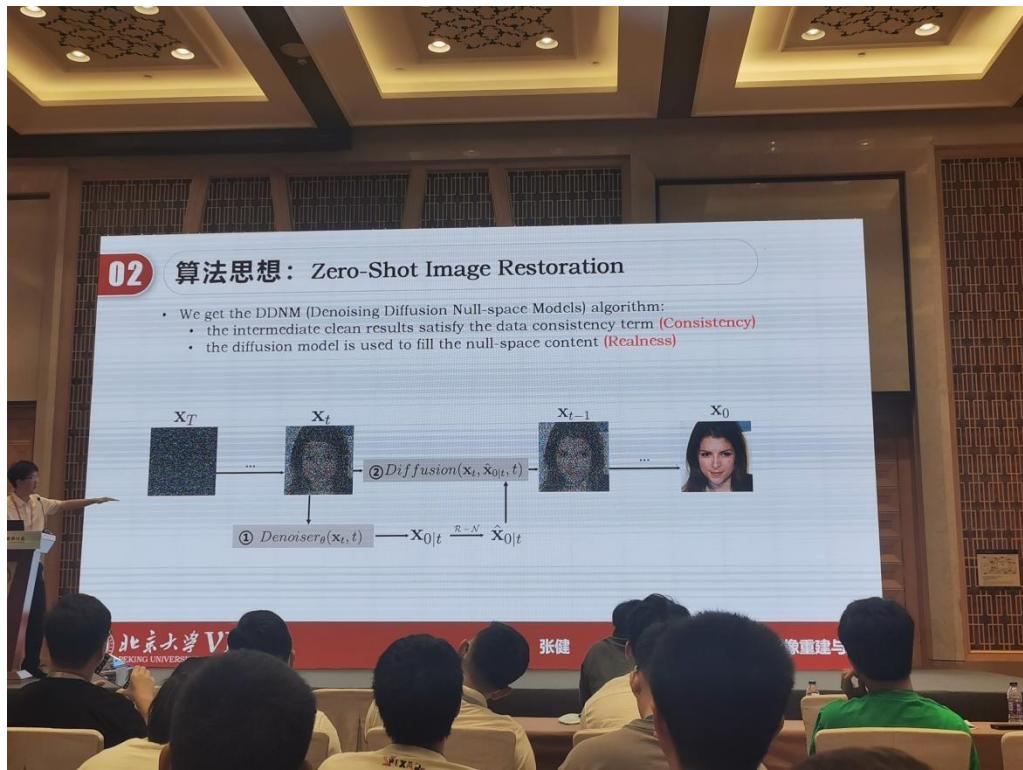
Tips: 考虑一个线性退化模型，我们一定可以得到它的一个伪逆，而且伪逆应该是不唯一的。



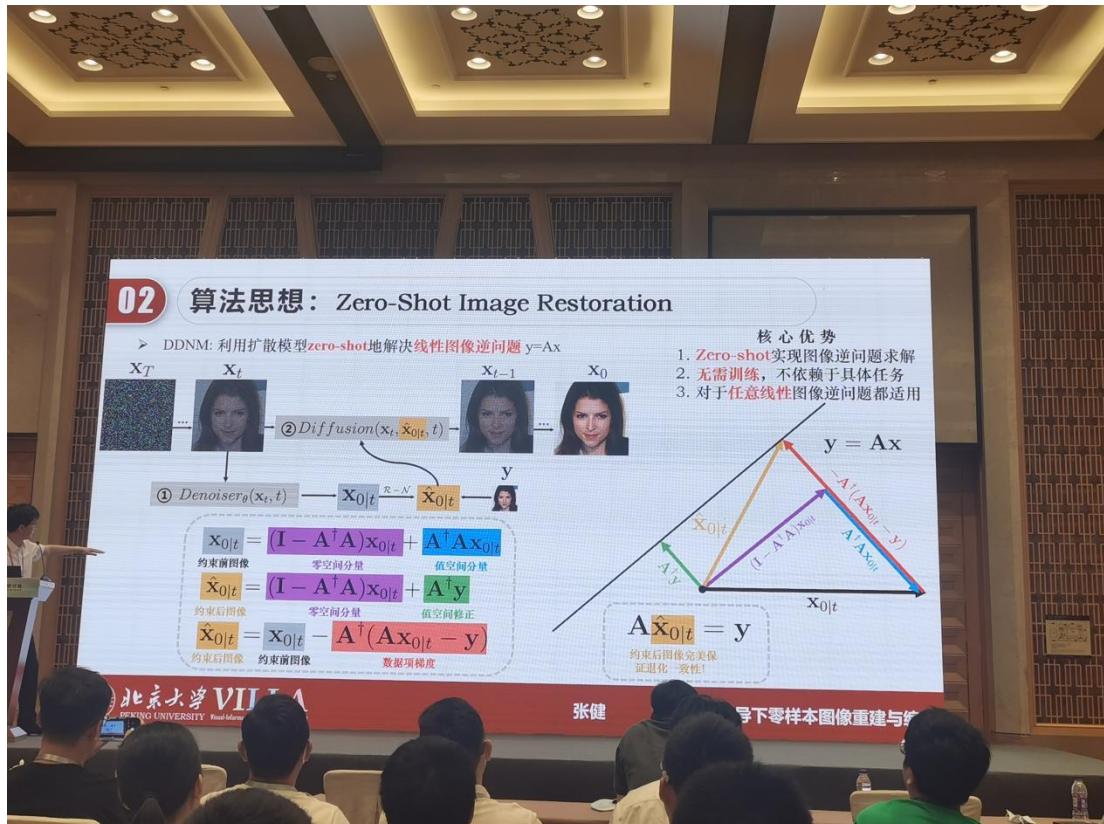
Tips: 分解操作是显然恒等的



Tips: 引入 Y, 等式仍然成立



Tips: 最终算法分为两项，一项满足图像回复任务的一致性，另一项通过 diffusion 实现图像的真实性。



Tips: 总结优点，zero-shot, training-free and generalization-ability



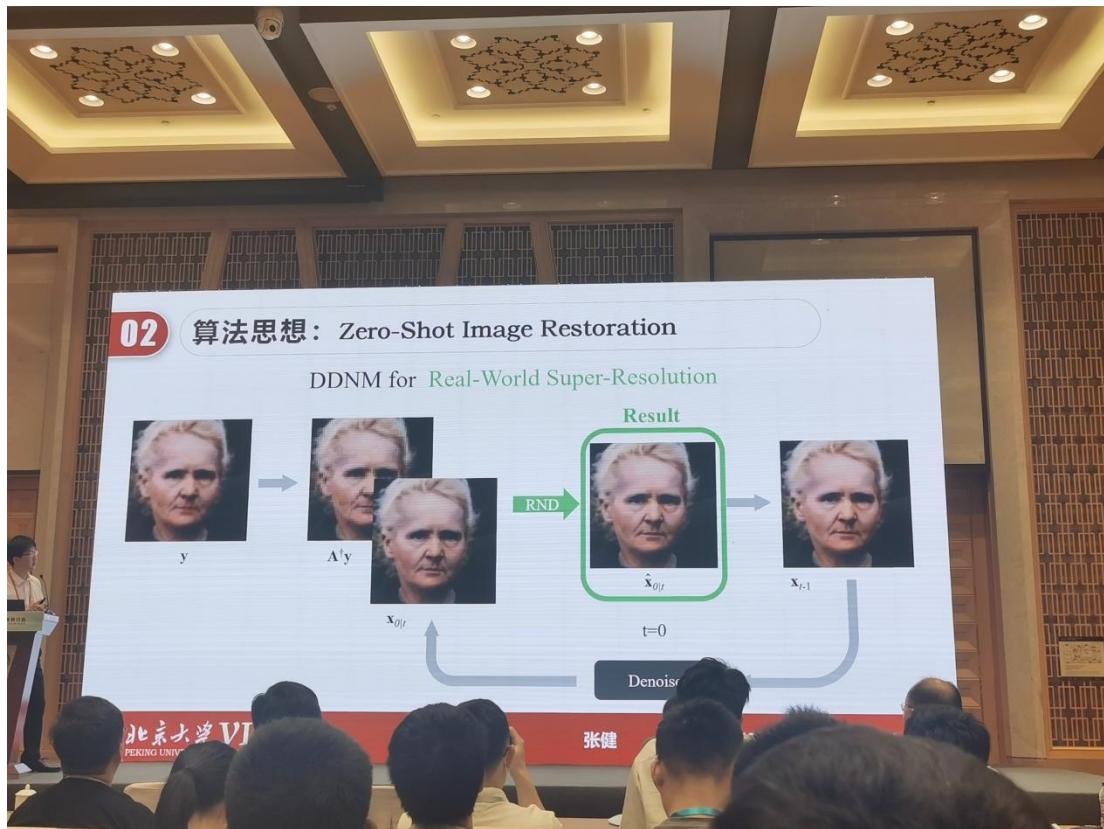
Tips:其实算法实现在公式上非常简单，只是加了一步。



Tips:详细推导证明过程，建议直接看论文原文。



Tips:由于退化过程是线性的，可以考虑退化级联。



Tips:真实图像复原，这里我觉得可能是有问题的，因为在真实图像中矩阵 A 是未知的，按照前面的描述，推导，该算法成功的关键在于要已知且线性的矩阵 A，

真实图像的退化可能也不是线性的。



Tips:一些实验结果





Tips:我觉得这里算 trick, 启发式的滑动拼接的方式实现对长图的 IR。



Tips:这里是另外一个工作, 类似的思路解决图像编辑问题。



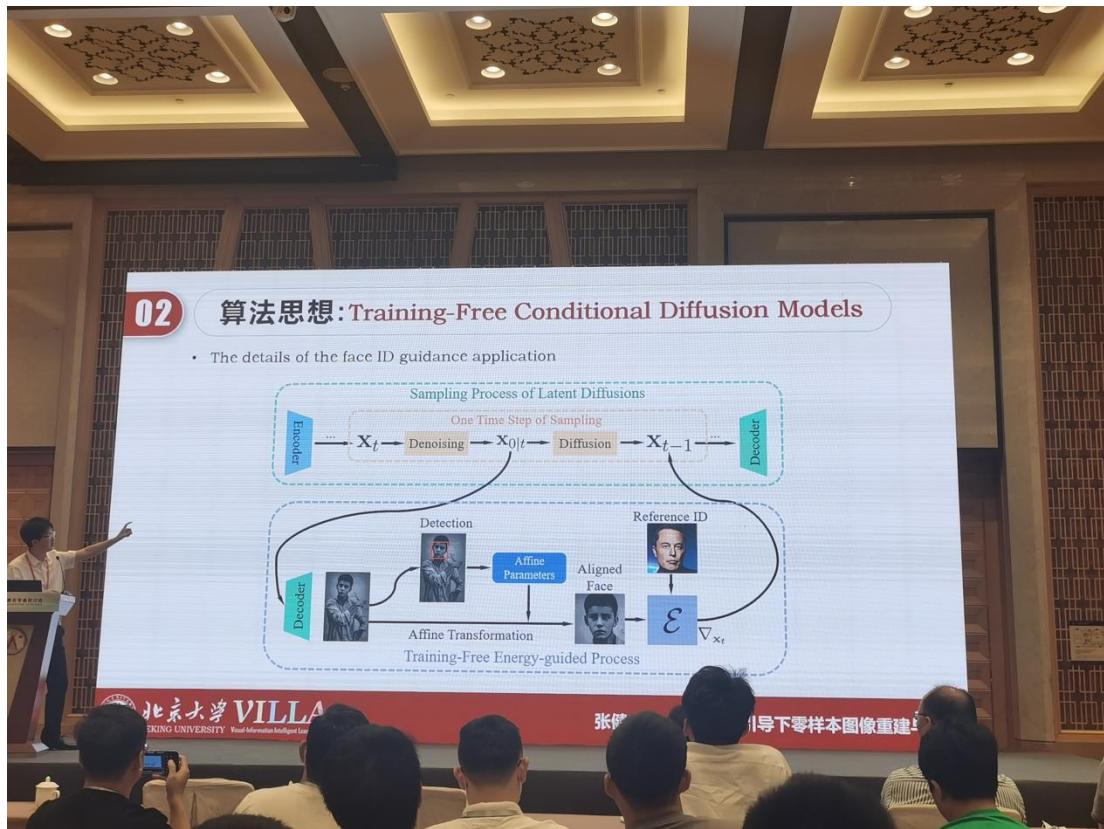
Tips: 在上述工作基础上可以加入条件，我觉得这里相当于是可以引入先验。



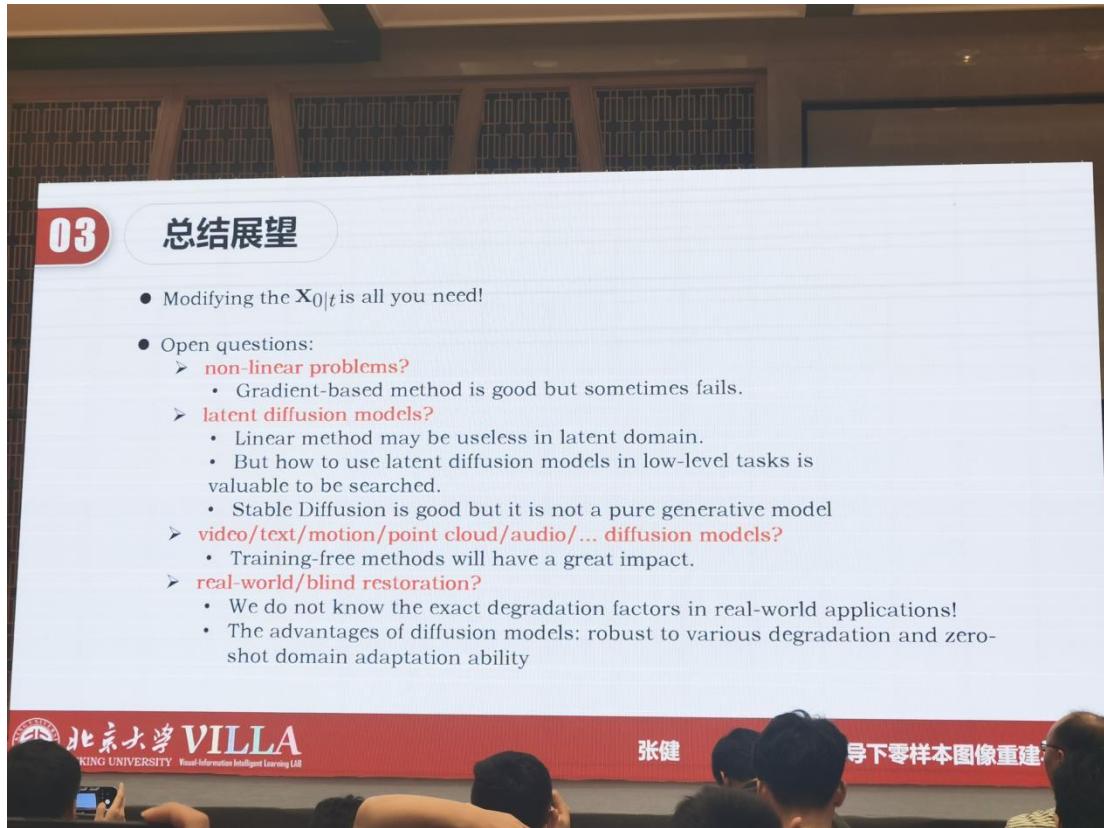


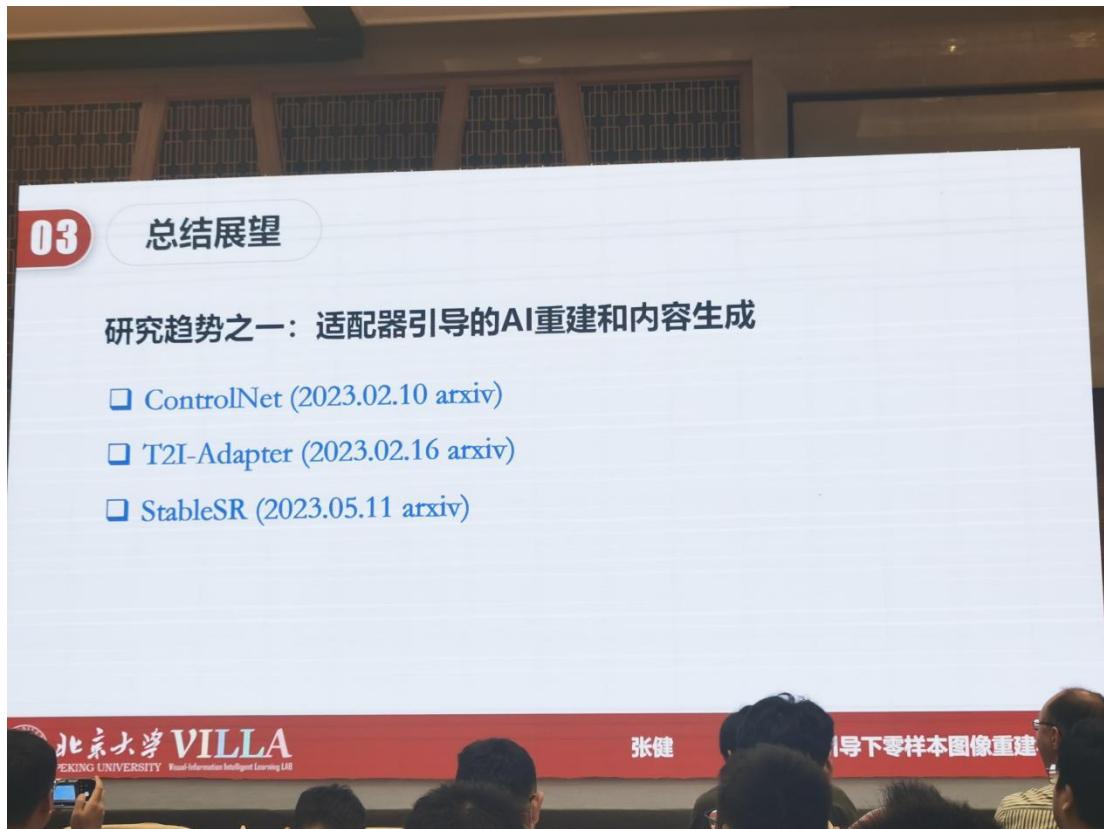
Tips:各种任务举例。



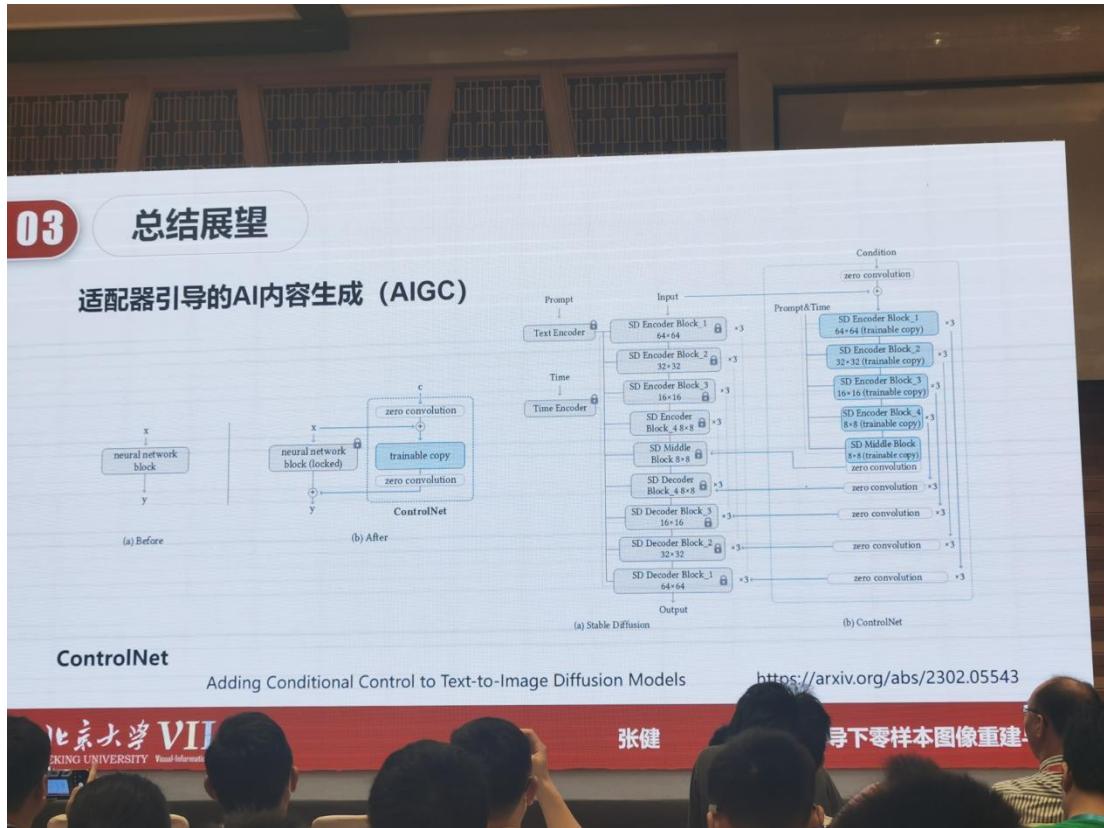


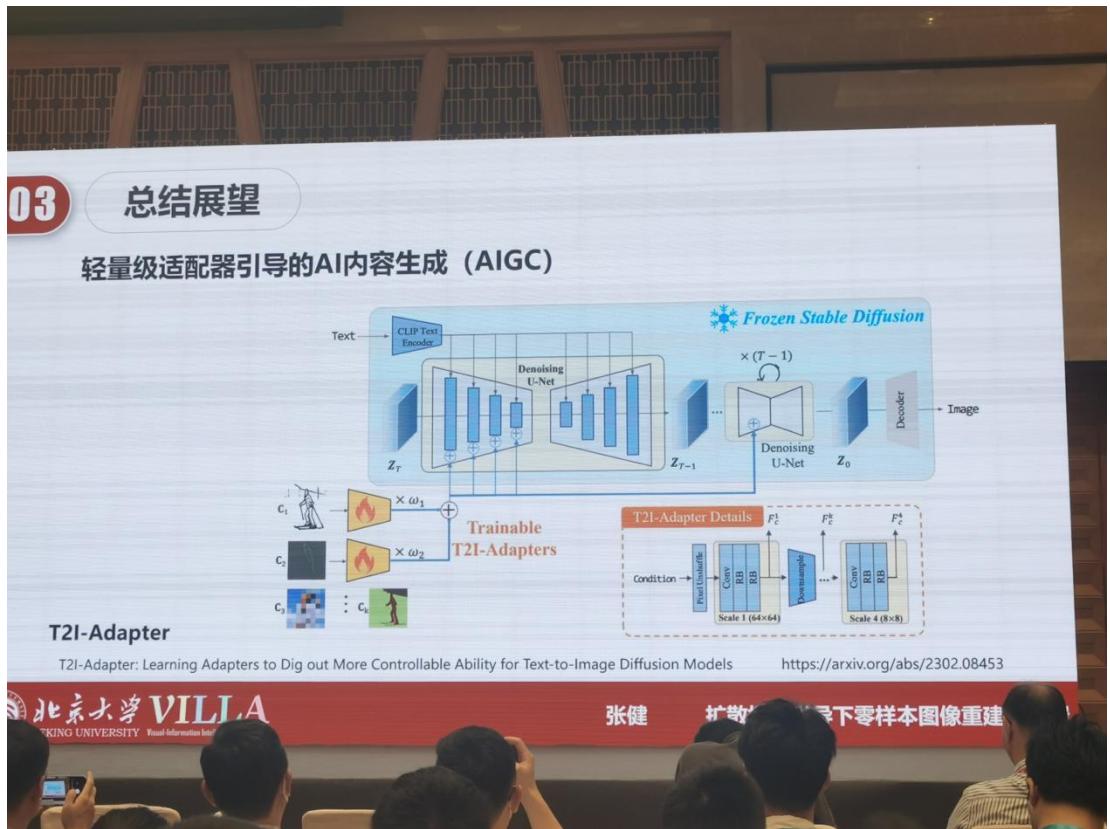
Tips: 实现流程，这里我觉得不涉及太多理论，主要是将各个领域的一些模型方法引入上述方法的框架中，进行组合，理论性不强，但是应用意义较大。





Tips:总结了Diffusion的优势也是未来的发展方向，同时training-free方法的影响力也将进一步扩大。然后给出了AIGC的前沿研究。





03

总结展望

➤ 强泛化能力



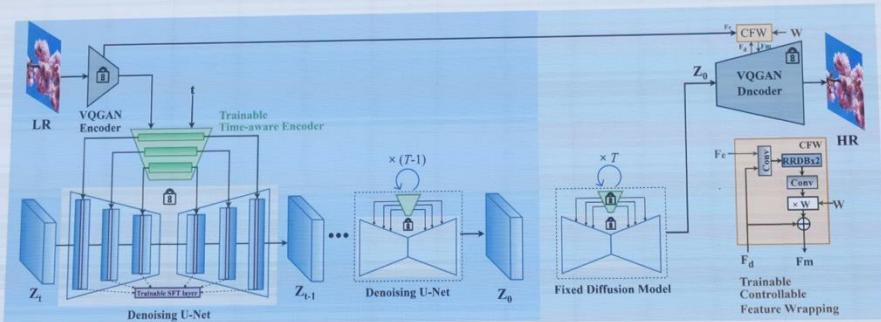
北京大学 VILLA
KING UNIVERSITY Visual Information Laboratory

张健

下零样本图像重建

03

总结展望



StableSR

Exploiting Diffusion Prior for Real-World Image Super-Resolution

<https://arxiv.org/abs/2305.07015>

北京大学 VILLA
KING UNIVERSITY Visual Information Laboratory

张健
下零样本图像重建



Tips:具体的应用介绍，给出了个人主页、实验室主页。

Workshops-视觉内容生成

汇报人：卢志武

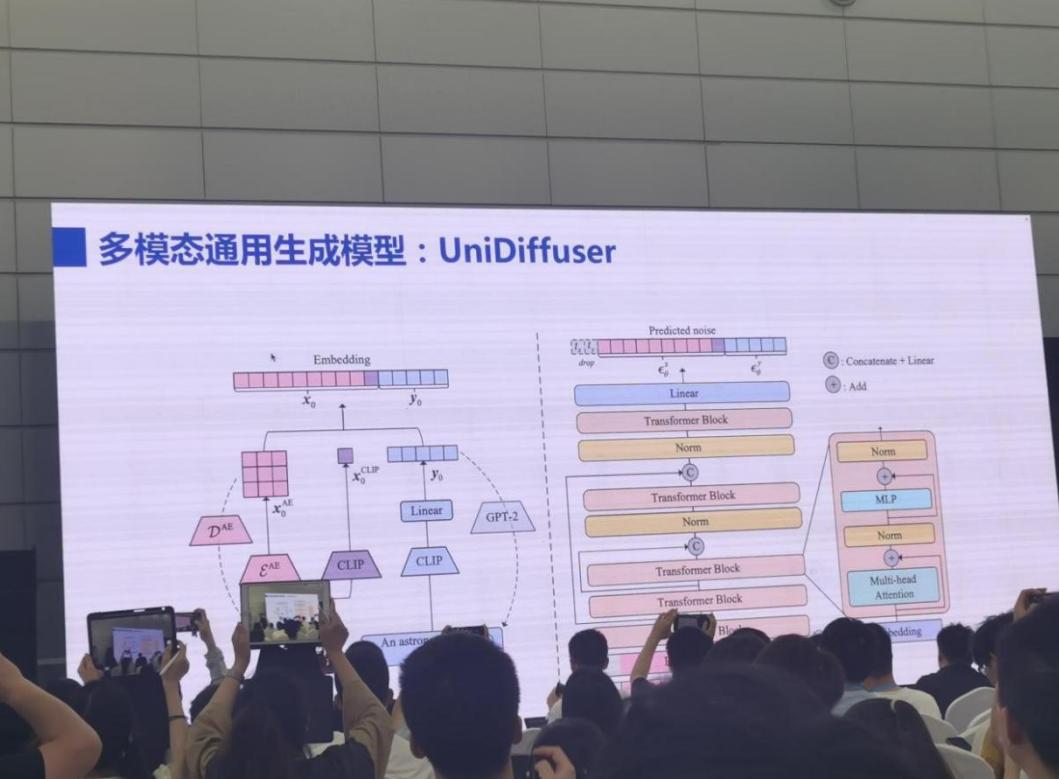
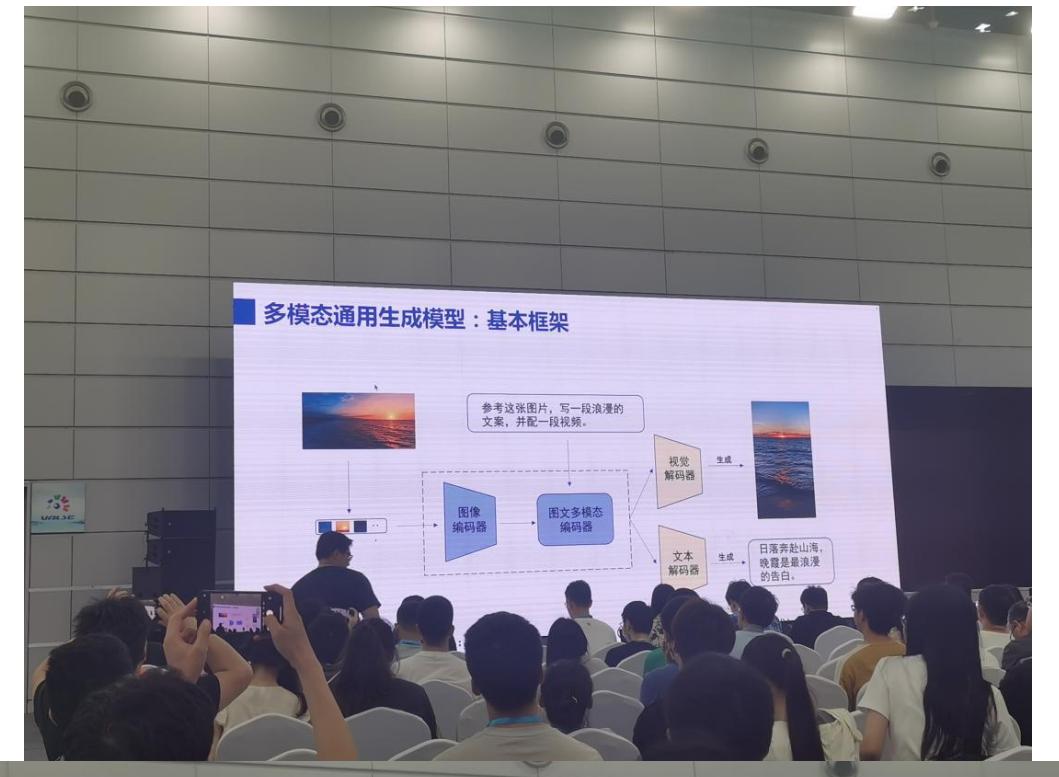
单位：中国人民大学

个人简介：卢志武博士，中国人民大学高瓴人工智能学院教授，博士生导师。2005年毕业于北京大学数学科学学院，获理学硕士学位；2011年毕业于香港城市大学计算机系，获PhD学位。研究方向为机器学习、计算机视觉等。设计首个公开的中文通用图文预训练模型文澜 BriVL，并发表于Nature Communications。

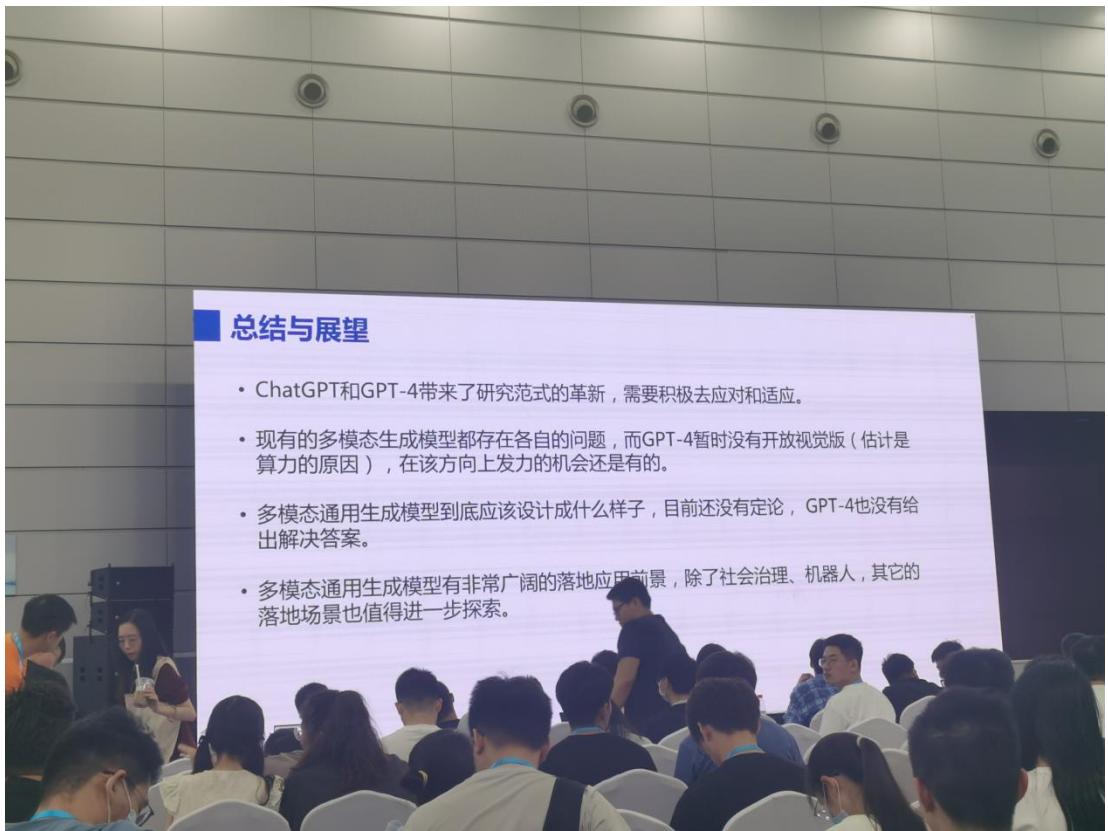
报告主题：多模态通用生成模型的基本框架与最新进展



Tips: 报告回顾了一篇发表在 nature communications 的 paper 《**towards artificial general intelligence via a multimodal foundation model**》。



Tips: 然后对未来的多媒体的研究与发展给出他的观点：通用大模型配上不同领域的专家模型的框架（大+小），可以实现对语言、文字、图像、视频等多种模态信息的生成。在网络结构上以 UNet、diffusion、transformer 为主，并给出了大量的实际应用场景。



总结与展望

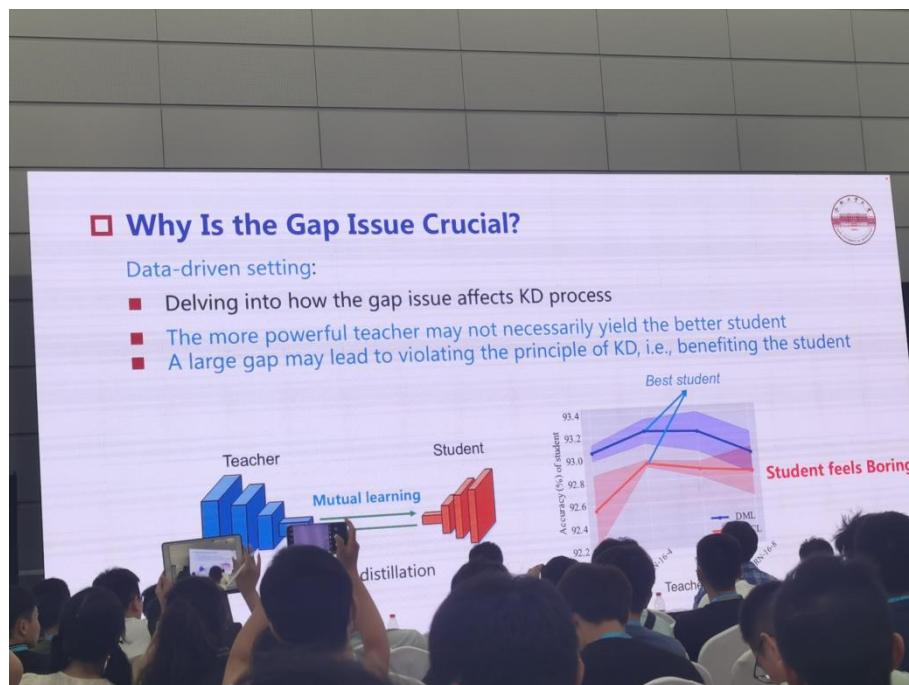
- ChatGPT和GPT-4带来了研究范式的革新，需要积极去应对和适应。
- 现有的多模态生成模型都存在各自的问题，而GPT-4暂时没有开放视觉版（估计是算力的原因），在该方向上发力的机会还是有的。
- 多模态通用生成模型到底应该设计成什么样子，目前还没有定论，GPT-4也没有给出解决答案。
- 多模态通用生成模型有非常广阔的应用前景，除了社会治理、机器人，其它的落地场景也值得进一步探索。

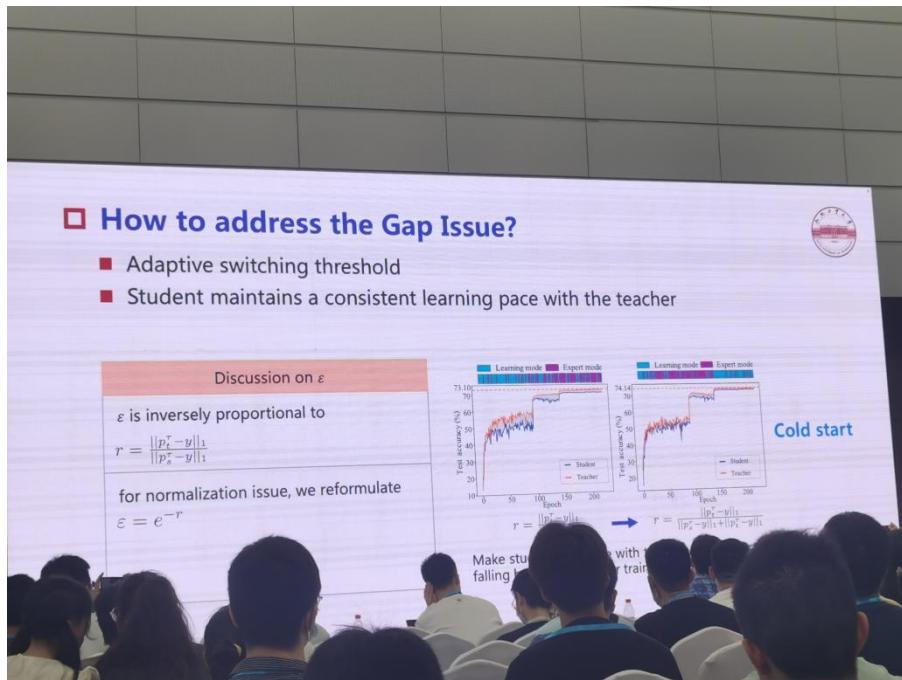
汇报人：王杨

单位：合肥工业大学

个人简介：合肥工业大学多媒体实验室 教授，博士生导师。入选安徽省高层次人才计划，担任信息搜索领域顶级杂志 ACM Transactions on Information Systems (CCF Rank A) 副编(Associate Editor, 2019 至今)。入选斯坦福大学 2022 年 9 月统计发布的人工智能与图像处理领域 前 2% 全球顶级科学家。至今在模式识别, 多媒体计算相关领域的顶级杂志与会议上发表文章 80 篇, 其中 ESI 高被引文章 7 篇, 并且全部进入 top 1% 列表, 发表源包括 Artificial Intelligence (Elsevier), International Journal of Computer Vision (IJCV), IEEE TIP, ACM TOIS, Machine Learning (Springer), IEEE TKDE, VLDB Journal, CVPR, ECCV, ACM SIGIR, ACM KDD, AAAI, IJCAI, ACM Multimedia, SCIENCE CHINA Information Sciences (中国科学:信息科学), 其中两篇论文入选 paperdigest2021/03 版本的 IJCAI 最有影响力文章之一, 主持国家自然科学基金联合基金重点项目, 面上项目等, 同时担任国家自然科学基金 优秀青年基金(海外)项目, 面上项目评审专家。谷歌学术引用 5000+, H-因子 34。

报告主题：基于知识蒸馏场景下的样本生成

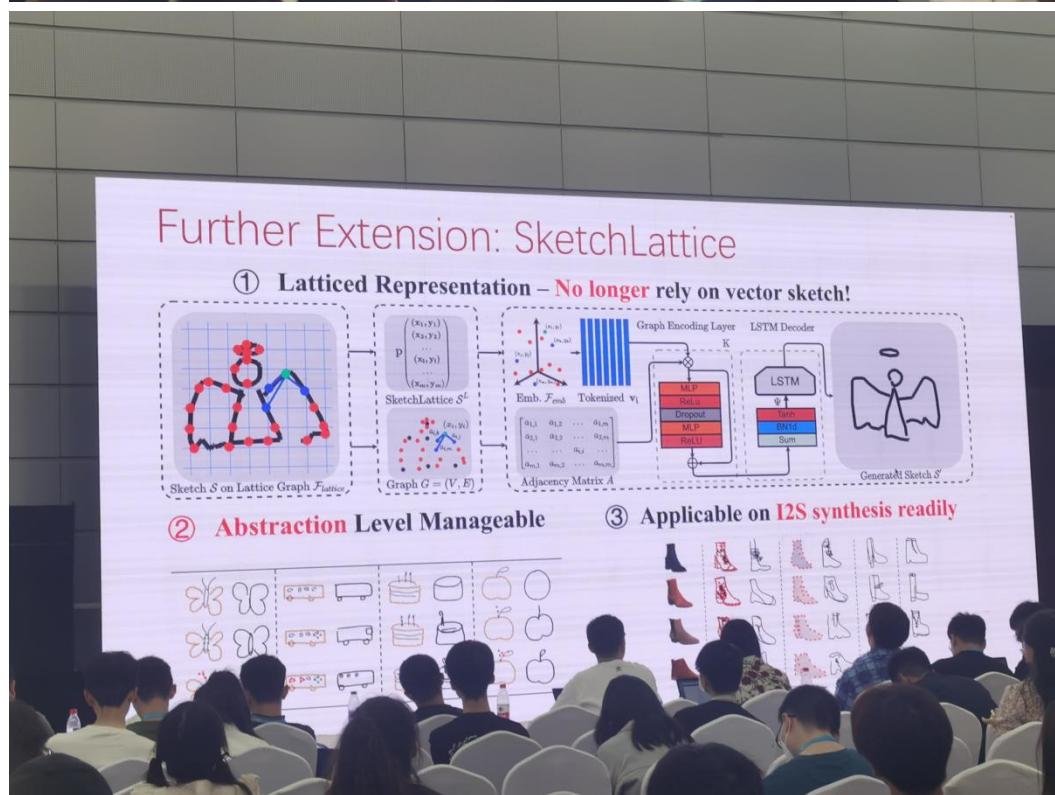
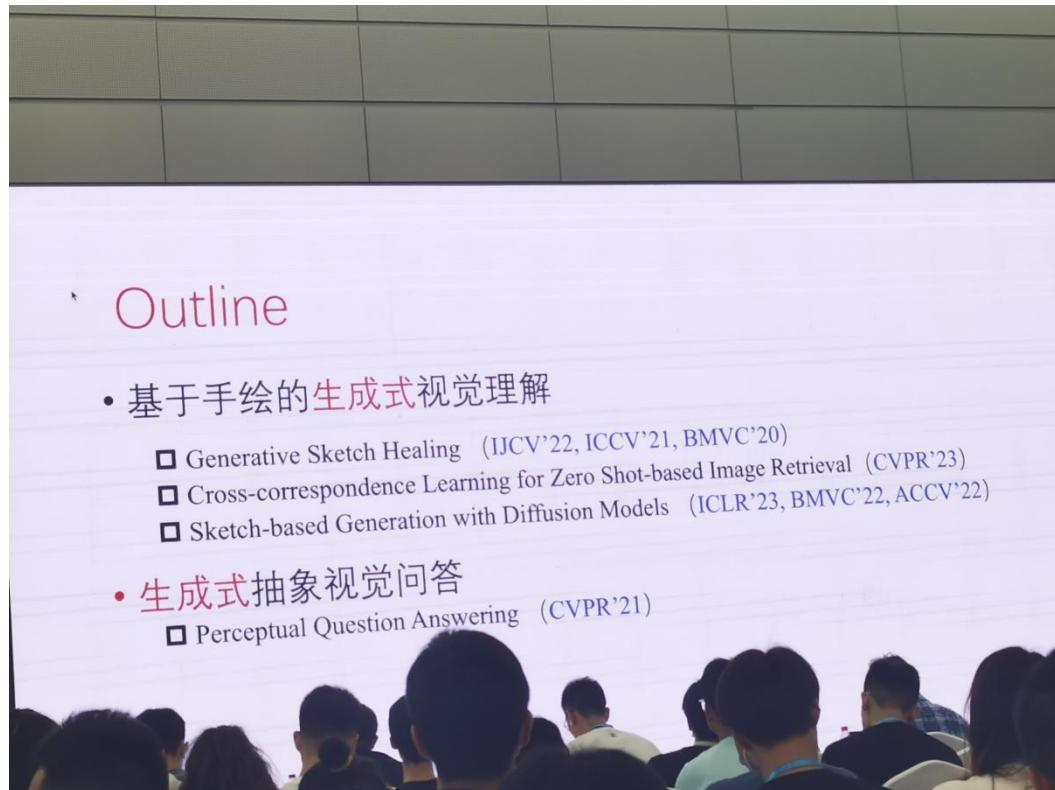


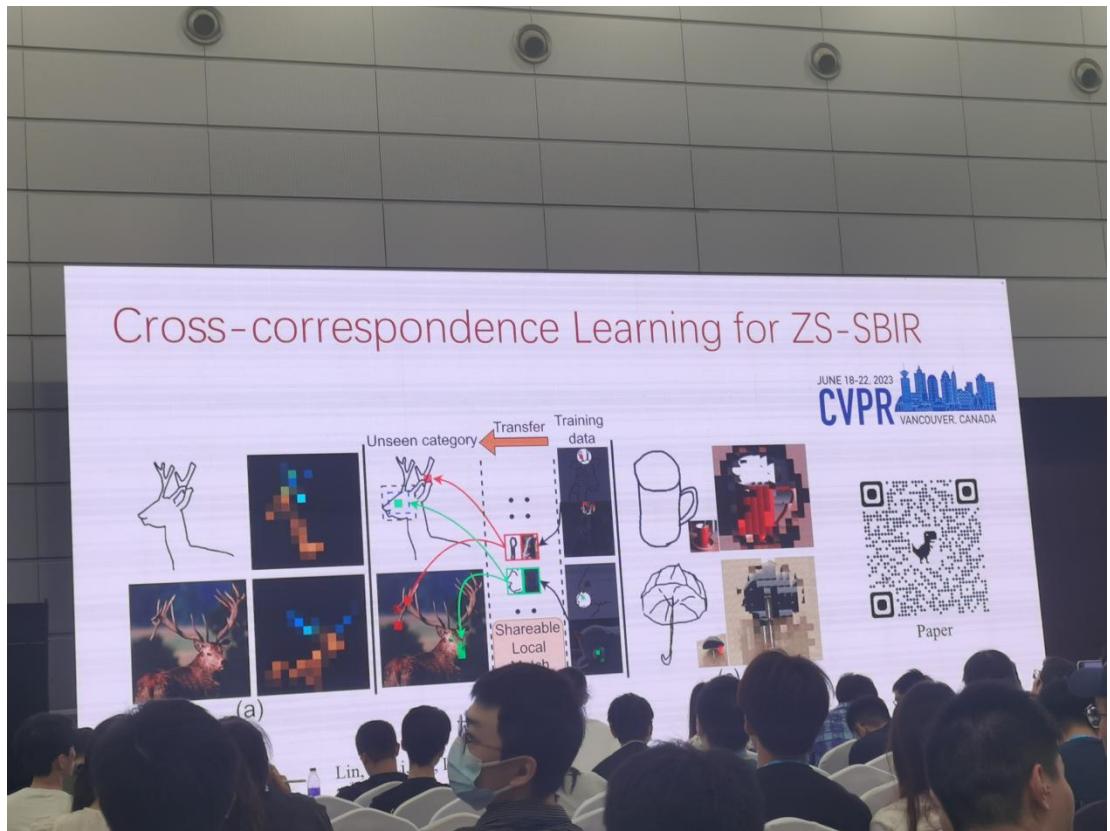


Tips: 由于不太了解知识蒸馏，从思想的角度理解王老师的观点在于：由于老师网络和学生网络在很多方面都有较大差异，穷尽地让老师全力指导，或者照搬自己的学习方式去训练学生网络不一定是最有效的，应该适当给学生网络一定的松弛，自由度，可能会学的更好，这个也是契合我们组的思想：适当的松弛和扰动可能会实现更好的优化性能。

汇报人：齐勇刚

单位：北京邮电大学





Tips: 生成式视觉理解相当于图像编辑的任务，23 年的 CVPR 做了一个 zero-shot 的工作，可以好好读一读。

汇报人：古纾旸

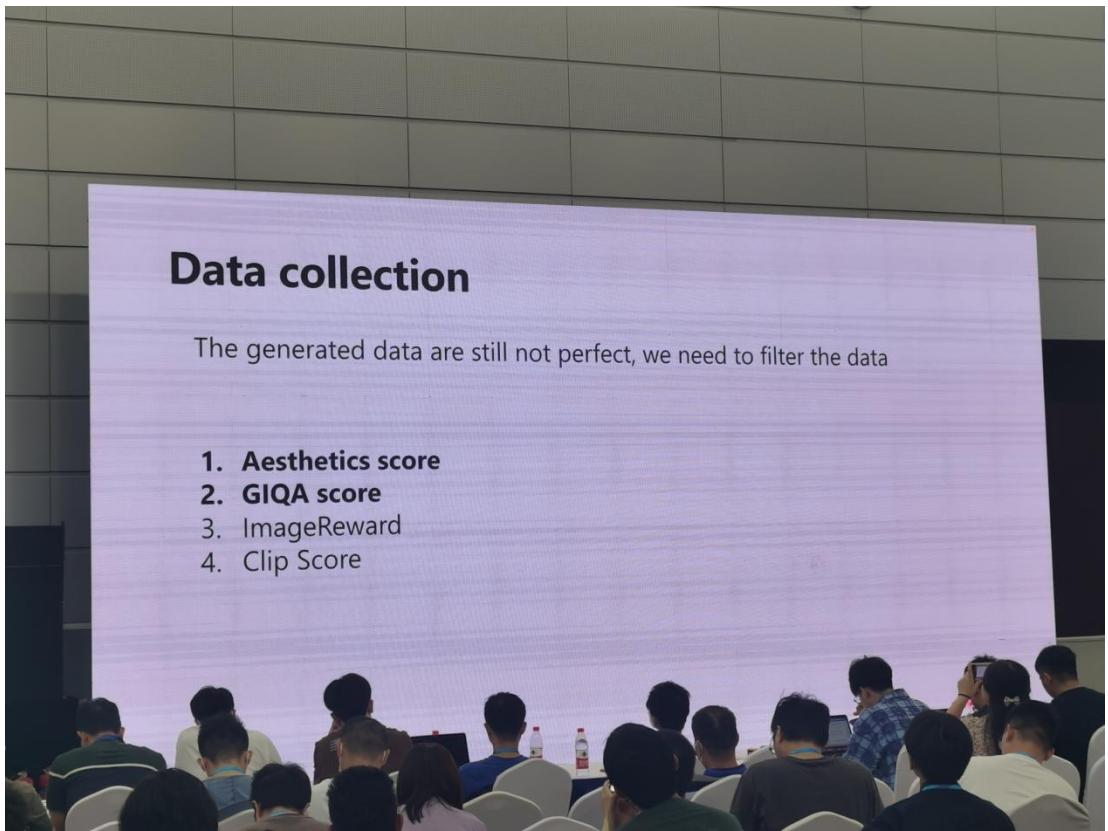
单位：微软亚洲研究院

个人简介：古纾旸，博士，在中国科学技术大学自动化系于 2017 年和 2022 年分别获得学士和博士学位，现为微软亚洲研究院研究员，主要研究方向为计算机视觉中的生成模型，特别是生成对抗网络和扩散模型的理论及其在 2D 和 3D 数据中的应用，以及对生成结果的质量评估等。目前已在 CVPR, ICCV, ECCV 等会议上发表多篇论文并担任多个会议与期刊的审稿人。

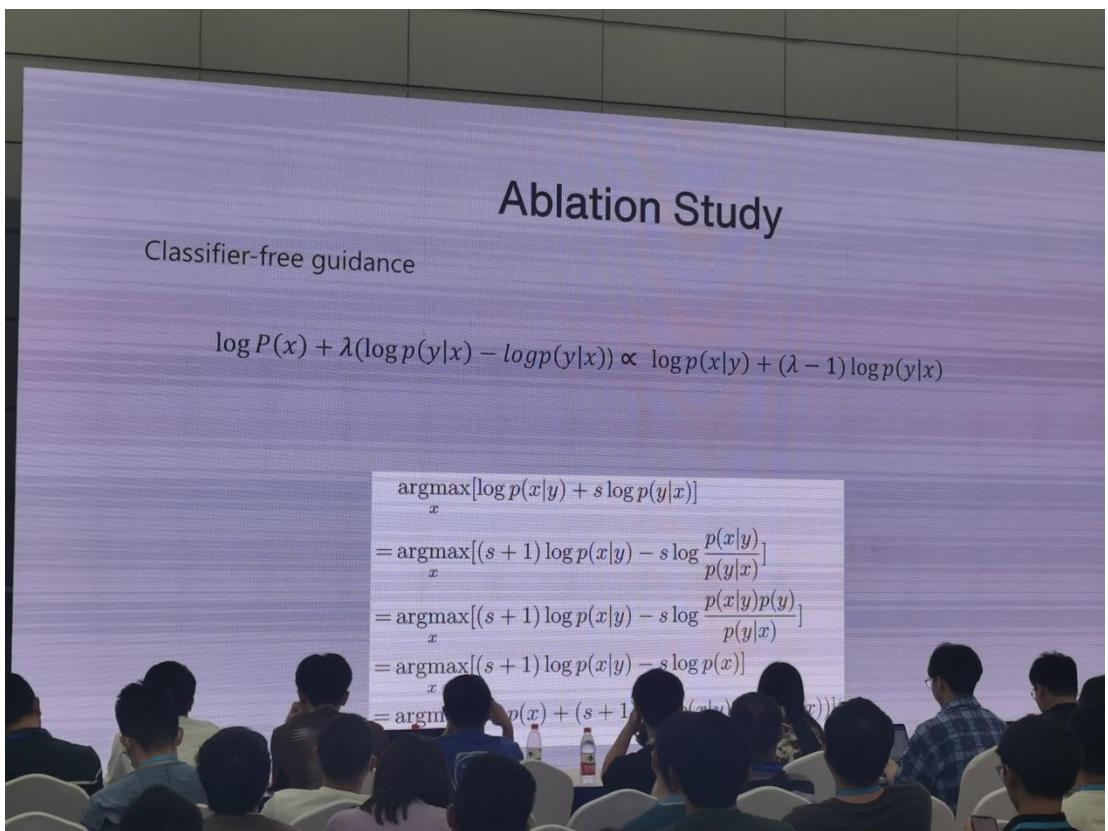
报告主题： From Paint by Example to Generalized Image Editing

Recently, language-guided image editing techniques have become increasingly successful. However, these techniques often lack local control. In our study, for the first time, we investigate exemplar-guided image editing as a way to achieve more precise control. We achieve this by disentangling and re-organizing the source image and the exemplar using self-supervised training. To avoid the copy-paste trivial solution, we propose an information bottleneck and strong augmentations. Additionally, to ensure controllability, we design an arbitrary shape mask for the exemplar image. Our framework involves a single forward of the diffusion model without iterative optimization. This new technology allows for controllable editing of in-the-wild images with high fidelity. By combining it with language-guided editing methods, we propose a generalized image editing framework that can handle various image editing operations based on language instructions.

报告总结：报告紧紧围绕在数据生成领域中，配对数据严重不足的问题，做了一系列的工作，包括训练数据增广，图像的局部切片，多种图像特征的提取，网络结构与模块多样性，多种下游任务辅助等等。



Tips: 对于 Data collection, 总共主要有以下四种途径, 但是他们通过实验发现, 主要是前两种比较有效。





Tips: 所提出的 Classifier-Free 方法经过了一系列的推导、实验在图像编辑生成方面实现了 SOTA 的性能，并验证了算法的每个部分的有效性。

[1] Paint by example: Exemplar-based image editing with diffusion models

汇报人：易冉

单位：上海交通大学



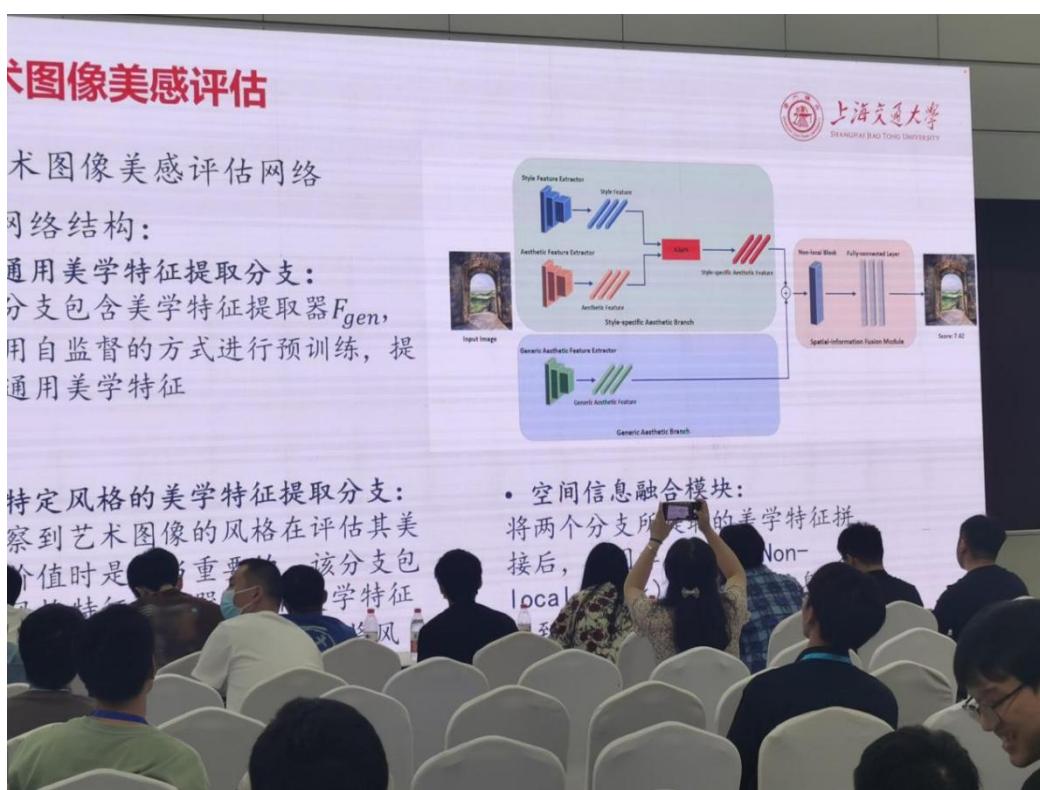
Tips: 报告主要内容从单模态的艺术肖像画生成任务，到 2D 多模态任务，然后到 3D 多模态任务生成，最后对生成图像评估指标进行了探索。



Tips: 改进了网络结构、损失函数以及新的数据集，应用创新意义较大，理论性不强。



Tips: 改进了损失函数、编码、评估指标，应用创新意义较大，理论性不强。



Tips: 提出了信息融合模块。



Tips: 提供了联系方式。

Workshops-目标检测与风格分割

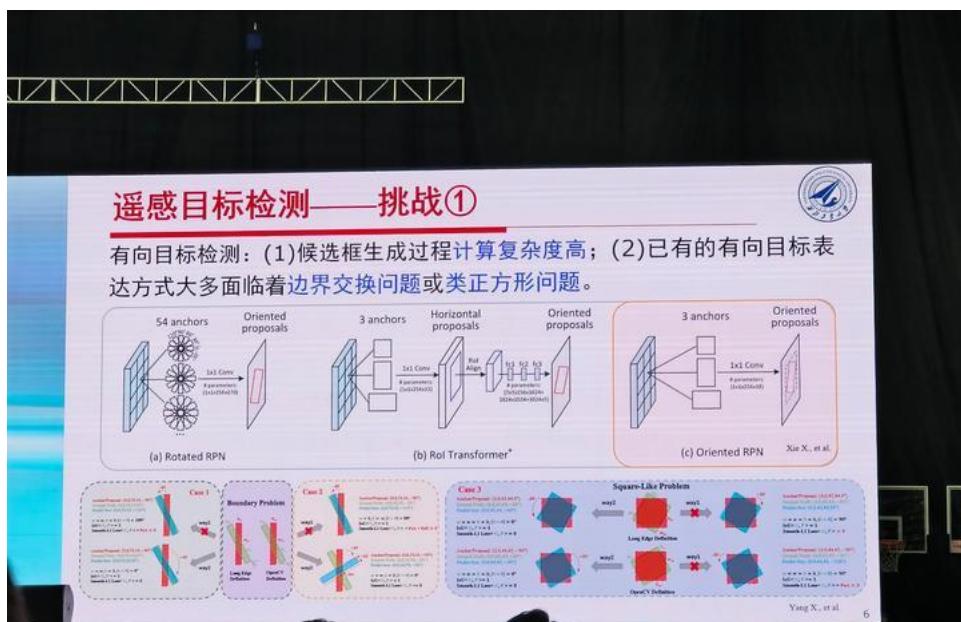
汇报人：程堪

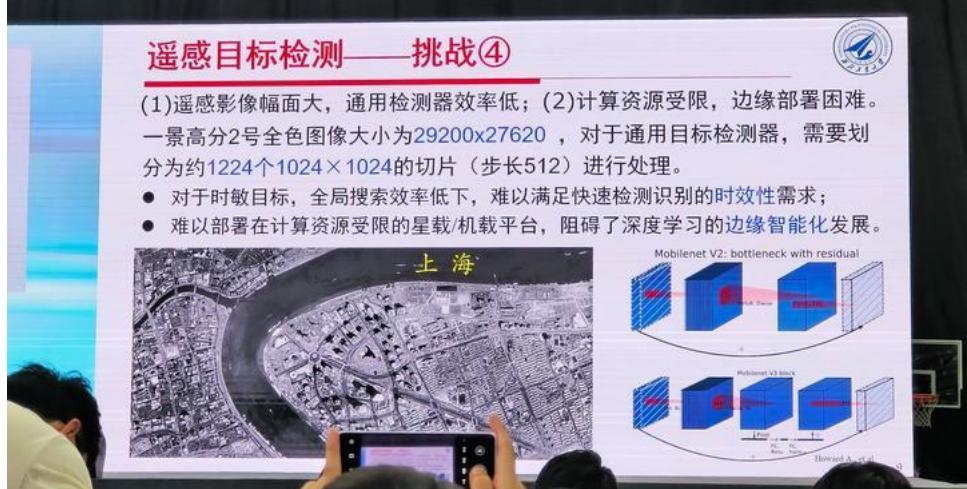
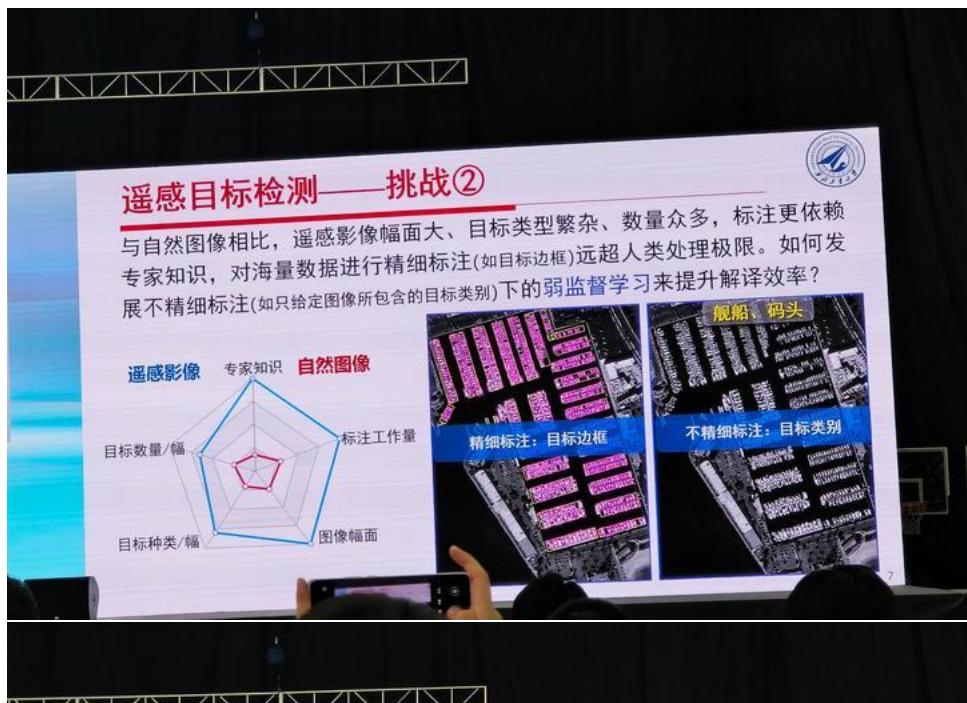
单位：西北工业大学

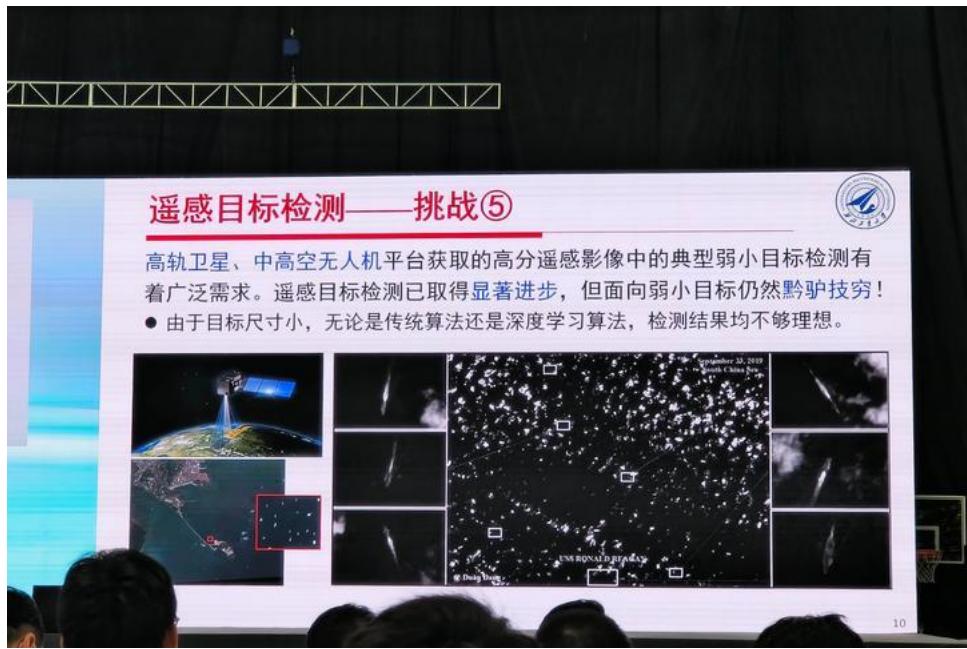
个人简介：程堪，西北工业大学长聘教授，博士生导师，信息融合技术教育部重点实验室副主任，入选国家“万人计划”青年拔尖人才，连续3年入选科睿唯安“全球高被引科学家”和爱思唯尔“中国高被引学者”。主要研究方向为光学遥感图像理解、计算机视觉等。以第一作者/通讯作者发表论文60余篇，包括PIEEE、TPAMI、CVPR、ICCV等，谷歌学术总引用1.3万余次，4篇第一作者论文单篇引用大于1000次，3篇论文入选年度中国百篇最具影响国际学术论文，获得2021年度IEEE TCSVT最佳论文奖、2021年度和2023年度IEEE地球科学与遥感学会最高影响力论文奖（IEEE GRSS Highest Impact Paper Award）等学术奖励，获得吴文俊人工智能技术发明一等奖、陕西省科学技术一等奖等4项省部级科技奖励，担任IEEE GRSM、ISPRS JPRS、JRS等多个国际期刊编委。

报告主题：遥感目标检测

报告总结：遥感目标检测是空天地海一体化观测系统的一项关键技术。与自然图像相比，遥感图像具有目标方向多变、目标类型及数量繁杂、特定领域样本稀缺、成像视角单一等特点，此外，不同平台、不同光照、天气条件、大气参数等都会对遥感图像获取产生影响。这些综合因素使得遥感目标检测面临着更大的挑战和更多的难点问题。具体来说，报告人总结了遥感目标检测的五大挑战：





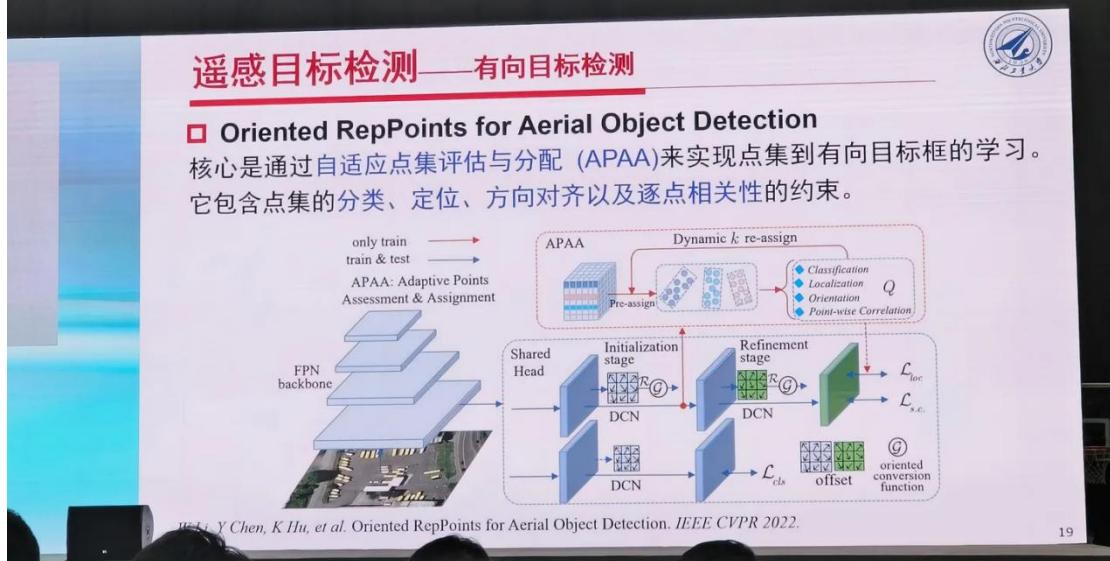
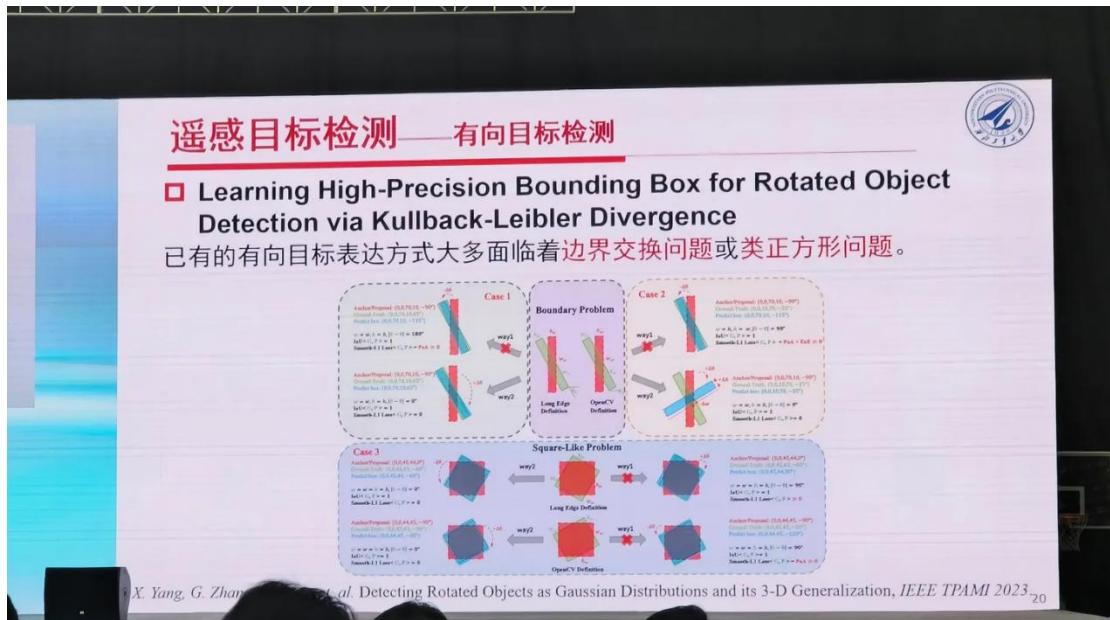


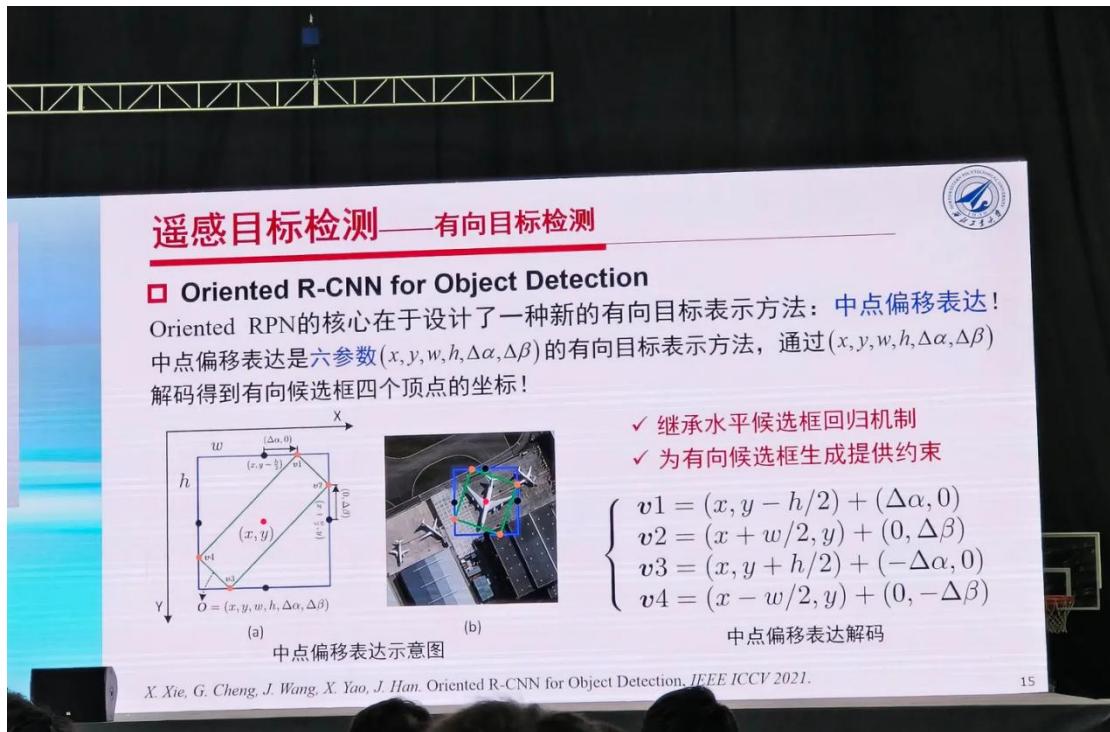
后续针对有向目标检测、弱监督目标检测、细粒度型号识别、高效目标检测、弱小目标检测展开相关介绍。具体工作可参考相关论文：

有向目标检测：

- [1] Learning High-Precision Bounding Box for Rotated Object Detection via Kullback-Leibler Divergence, NeurIPS 2021;
- [2] Phase-shifting coder: Predicting accurate orientation in oriented object detection, CVPR 2023;







弱监督目标检测：

[3] SOOD: Towards Semi-Supervised Oriented Object Detection , CVPR 2023;

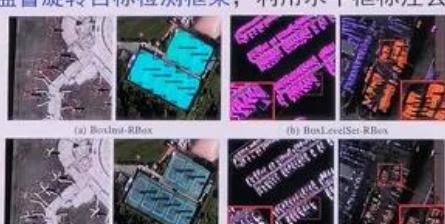
[4] H2RBox: Horizontal Box Annotation is All You Need for Oriented Object Detection;

遥感目标检测——弱监督目标检测

□ **H2RBox: Horizontal Box Annotation is All You Need for Oriented Object Detection**

已有的有向目标检测算法主要基于旋转框标注，利用水平框标注训练旋转框能够节省标注成本，并且优化大量的现有数据集。

- 提出了一种弱监督旋转目标检测框架，利用水平框标注去训练旋转目标检测器！



X. Yang, G. Zeng, et al. H2RBox: Horizontal Box Annotation is All You Need for Oriented Object Detection. ICML 2023. 28

遥感目标检测——弱监督目标检测

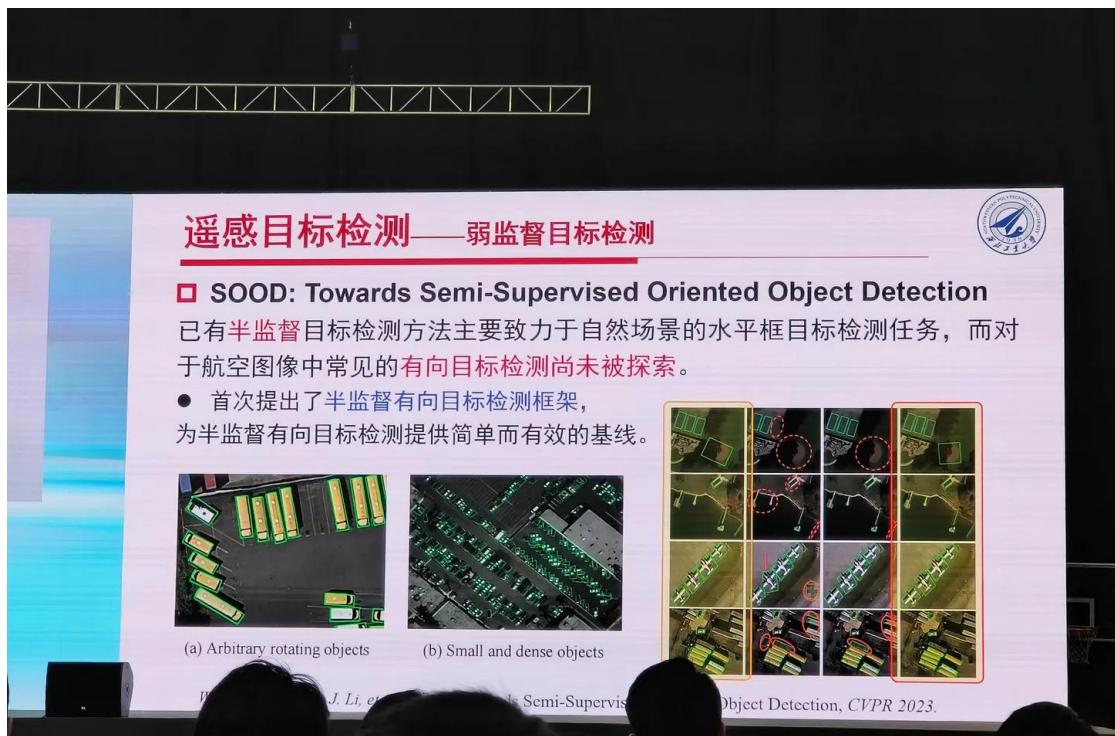
□ **SOOD: Towards Semi-Supervised Oriented Object Detection**

旋转感知自适应加权损失：考虑目标的方向，通过方向差动态地对每个伪标签/预测对进行加权。全局一致性损失：考虑目标的分布，从全局的角度来衡量伪标签和预测的相似性。

- 核心在于设计两个损失，加强学生和教师网络模型预测的一致性。

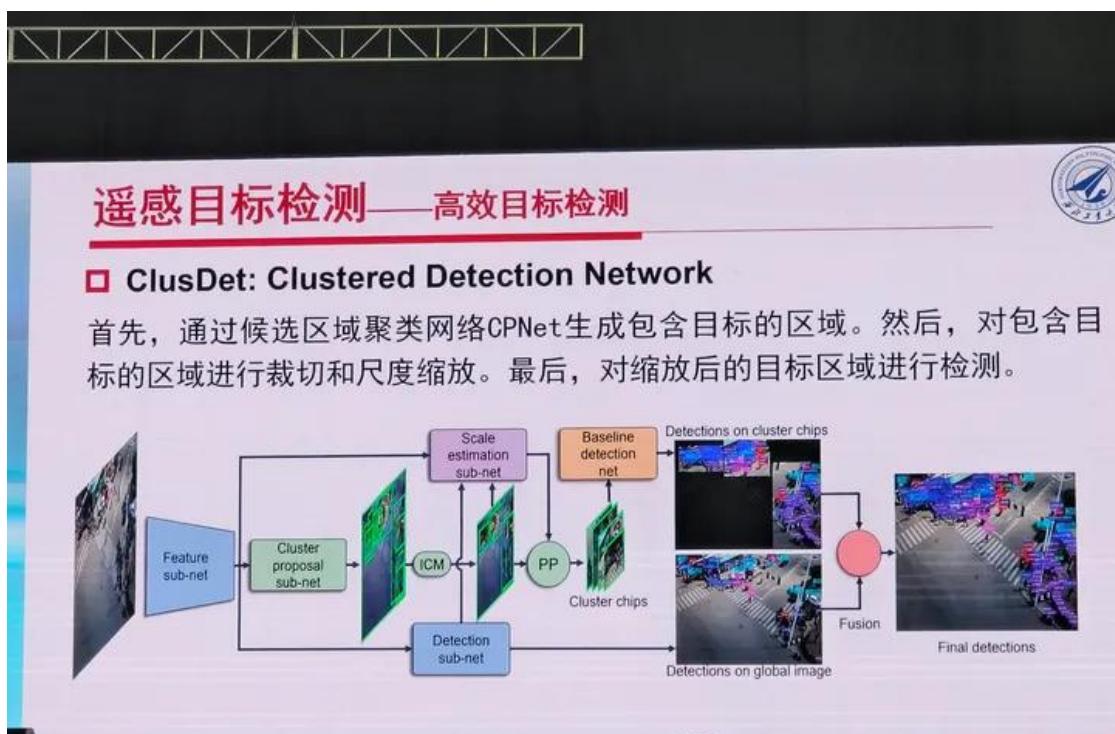


W. Li, J. Li, et al. SOOD: Towards Semi-Supervised Oriented Object Detection. CVPR 2023. 28



细粒度型号识别：

[5] FAIR1M: A benchmark dataset for fine-grained object recognition in high-resolution remote sensing imagery;



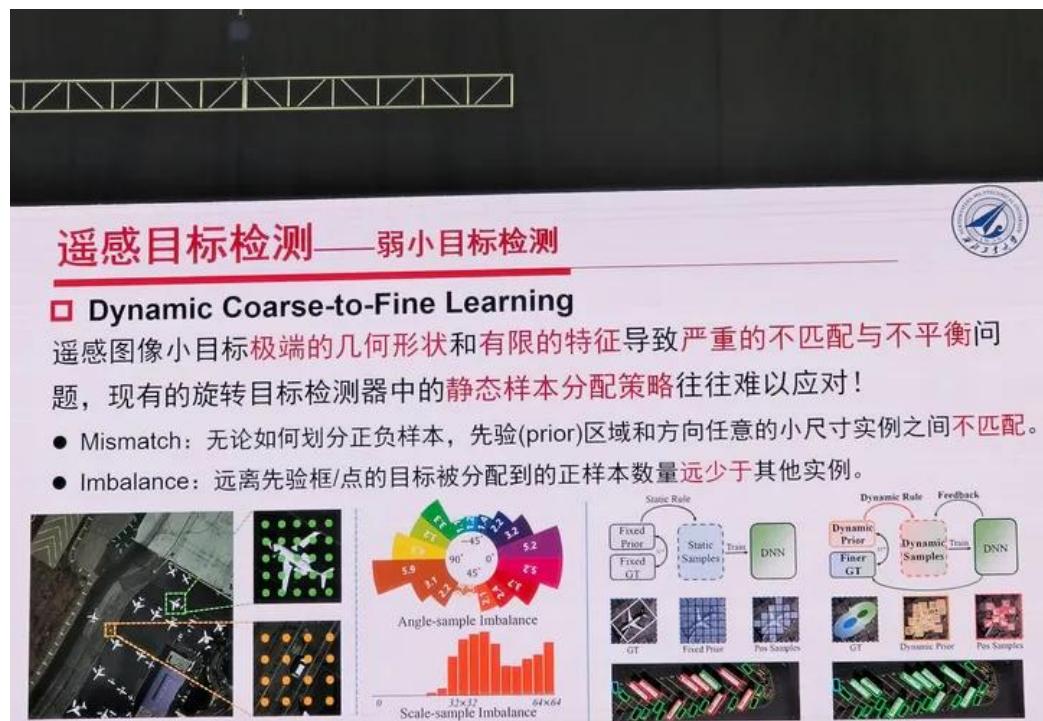
高效目标检测：

[6] Aerial Image Object Detection Method Based on Adaptive ClusDet Network



弱小目标检测：

[7] Dynamic Coarse-to-Fine Learning for Oriented Tiny Object Detection, CVPR 2023



汇报人：张兆祥

单位：中科院自动化所

个人简介：张兆翔，中国科学院自动化研究所研究员、博士生导师，入选教育部长江学者，国家万人计划青年拔尖人才等。主要研究方向包括脑启发的神经网络建模、视觉认知学习、面向开放环境的场景感知与理解，在本领域 TPAMI、IJCV、JMLR 等顶刊和 CVPR、ICCV、ICLR、NeurIPS 等顶会发表论文 100 余篇，授权专利 20 余项，承担了国家自然科学基金重点项目、国家自然科学基金企业联合重点项目、国家重点研发项目等一系列国家级科研项目，是 IEEE 高级会员，中国计算机学会 CCF 杰出会员、中国人工智能学会 CAAI 杰出会员、中国计算机学会 CCF 杰出演讲者，担任或曾担任 IEEE T-CSVT、IEEE T-BIOM、Pattern Recognition 等知名期刊编委，是 CVPR、ICCV、NeurIPS、AAAI、IJCAI、ACM MM 等知名国际会议的领域主席（Area Chair）。

报告主题：基于多传感器融合的视觉物体检测与分割



报告总结：视觉物体检测与分割是计算机视觉领域的核心问题，具有重要理论意义和应用价值。当前，在以自动驾驶、无人机、服务机器人等为代表的具身智能系统中，视觉场景感知涉及的传感器往往类型多样、数目多样。在多种传感器条件下，如何针对特定的传感器类型设计高效的视觉场景感知算法，如何对多种传感器的信息加以有机融合，具有重要的创新价值，也是推动实际应用性能的关键。

多传感器融合动机：



首先是“看得多”：对于多帧传感器的融合，主要针对二维图像与三维点云



针对二维图像的时序融合，核心是2D、3D物体结构几何约束、物体位姿约束等重投影约束。相关论文：

[1] **Densely Constrained Depth Estimator for Monocular 3D Object Detection, ECCV 2022;**

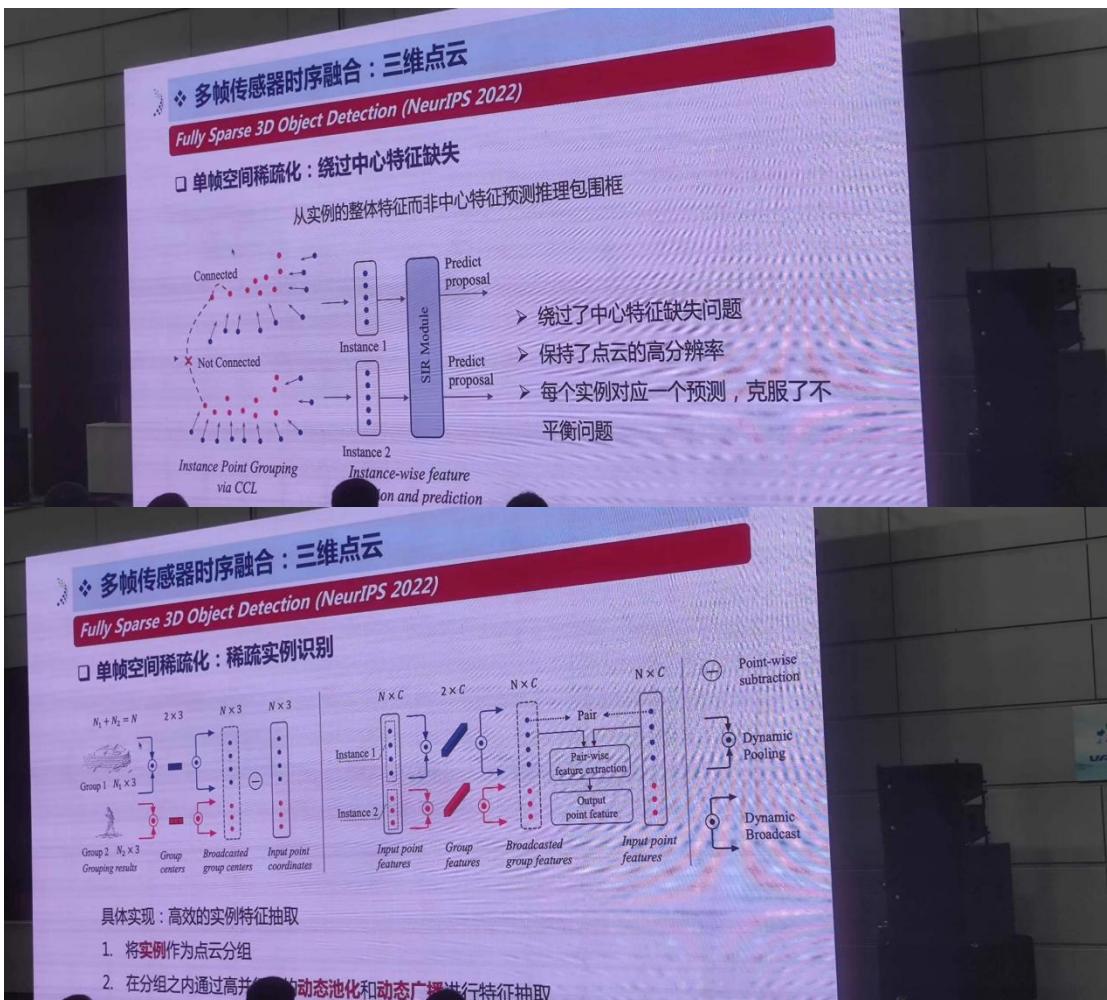
[2] **3D Video Object Detection with Learnable Object-Centric Global Optimization, CVPR 2023.**



针对 3D 点云的时序融合，相关论文：

[3] Fully Sparse 3D Object Detection (NeurIPS 2022);

[4] Super Sparse 3D Object Detection (TPAMI 2023)



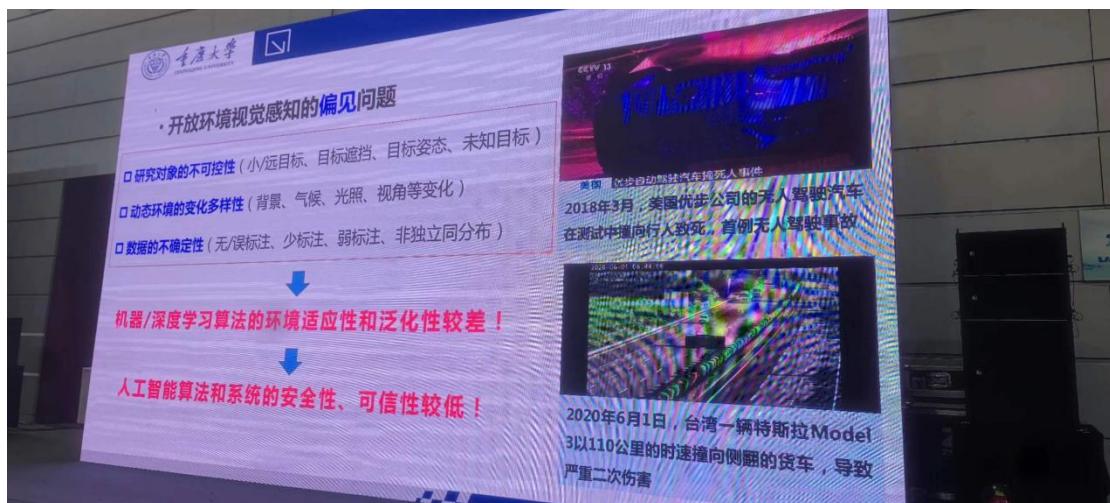
汇报人：张磊

单位：重庆大学

个人简介：张磊，重庆大学微电子与通信工程学院教授，博导。主要研究兴趣为开放环境视觉感知中的域迁移泛化和模型鲁棒性问题。发表 IEEE Transactions(TPAMI、IJCV、TIP 等)以及 CVPR/ICCV/ECCV/ICML/AAAI 等论文 100 余篇，专著 1 部，被引用 5 千余次。担任 IEEE Trans. Instrumentation and Measurement 和 Neural Networks 等期刊编委以及 ACM MM/CVPR/ICCV/ICLR/ICML/NeurIPS/AAAI/IJCAI 等会议领域主席或程序委员等，曾获吴文俊人工智能优秀青年奖、ACM SIGAI 中国新星奖、重庆市十佳科技青年奖。以第 1 完成人获得吴文俊人工智能自然科学奖和重庆市自然科学奖 2 项。

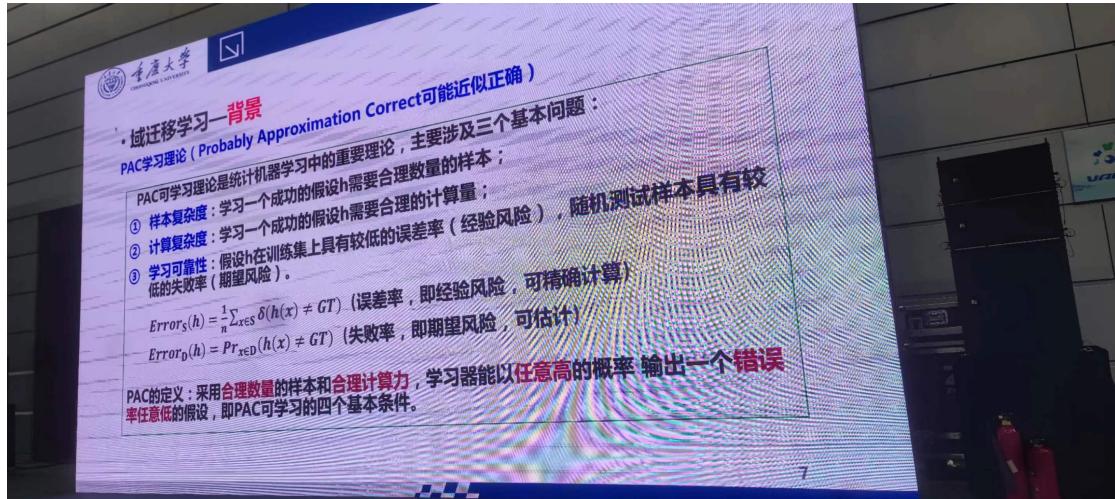
报告主题：开放环境视觉感知之开放域目标检测

报告总结：目标检测作为计算机视觉中最基础的任务之一，已在深度学习引擎的推动下得到快速发展，检测精度及泛化能力显著提升。尽管如此，在实际应用中，由于数据不确定性、环境不可控性以及算法特异性，现有目标检测算法在开放域下的适应能力依然较差，难以满足开放、动态、复杂物理环境下的视觉感知与应用。开放环境视觉感知的挑战：





针对以上问题与挑战,报告人的研究方向主要集中在跨域不变特征和域共有特征的表征学习,部分工作集中在module化的多尺度对齐工作。域迁移学习的背景、概念以及理论:



重庆大学
CHONGMING UNIVERSITY

域迁移学习一理论 (Domain Adaptation或适配)

目标域的期望误差上界：由源域期望误差，域间分布差异和联合误差共同界定。

Ben-David定理(NIPS'06)：令 H 为VC维是 d 的假设空间，一个维度为 m 的标记样本 x ，通过 R 函数将 x 映射到 S 后，则对于假设空间 H 中的每个假设 h ，则以至少 $1-\delta$ 的概率，下式成立：

$$e_T(h) \leq \hat{e}_S(h) + \sqrt{\frac{4}{m} \left(d \log \frac{2em}{d} + \log \frac{4}{\delta} \right)} + d_H(\mathcal{D}_S, \mathcal{D}_T) + \epsilon$$

即，假设 h 在目标域 T 上的期望风险（误差）可以被精确的界定（三角不等式证明）。

1) 设计大量迁移自适应学习模型与方法，实现跨域迁移学习；
 2) 从本质上讲，domain adaptation是一个minimax问题。

视角1： $MMD[\mathcal{F}, p, q] := \sup_{f \in \mathcal{F}} |\mathbb{E}_{x \sim p}[f(x)] - \mathbb{E}_{z \sim q}[f(z)]|$

视角2： $\min_f d_H(S, T) \Leftrightarrow \max_{h \in \mathcal{H}} \{err_S(h(x)) + err_T(h(x))\}$

DG假设待泛化的目标域数据不可知，且目标域数据无限大 (out-of-distribution)。

Tricky problem: 如何评估所有可能的目标数据的期望风险？

目标期望风险的精确计算：假设所有分布的目标域数据均服从某种潜在的超分布 P ，为使模型 $h(\cdot)$ 泛化到不可见的任务，计算其期望风险：

$$\mathcal{E}(h) := \mathbb{E}_{P_{XY} \sim P} \mathbb{E}_{(x,y) \sim P_{XY}} [\ell(h(P_X, x), y)]$$

显然，精确计算未来目标任务的期望风险是不可实现的。（理想真的要化为泡影吗？）

目标期望风险的有限估计：定义有限域或分布的监督数据 $\{U|x, y\}$ （source domains），同样服从该潜在的超分布 P ，则经验估计为：

$$\hat{\mathcal{E}}(h) := \frac{1}{M} \sum_{i=1}^M \frac{1}{n^i} \sum_{j=1}^{n^i} \ell(h(U^i, x_j^i), y_j^i)$$

有限估计虽解燃眉之急，但总要面对世事无常。

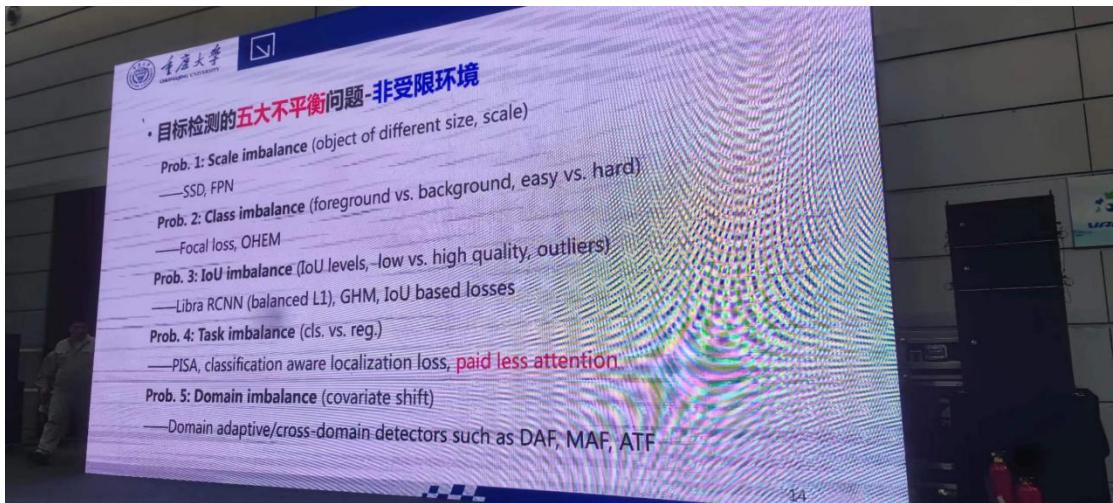
即，泛化到某个特定目标域 T 的风险上界到底如何？

DG理论上界：对于目标域 T ，其分布 P_X^t 坐落于源域分布的凸包内，即 $\sum_{i=1}^M \pi_i P_X^i = P_X^t$ 。仅考虑协变量偏移，那么其泛化风险上界为：

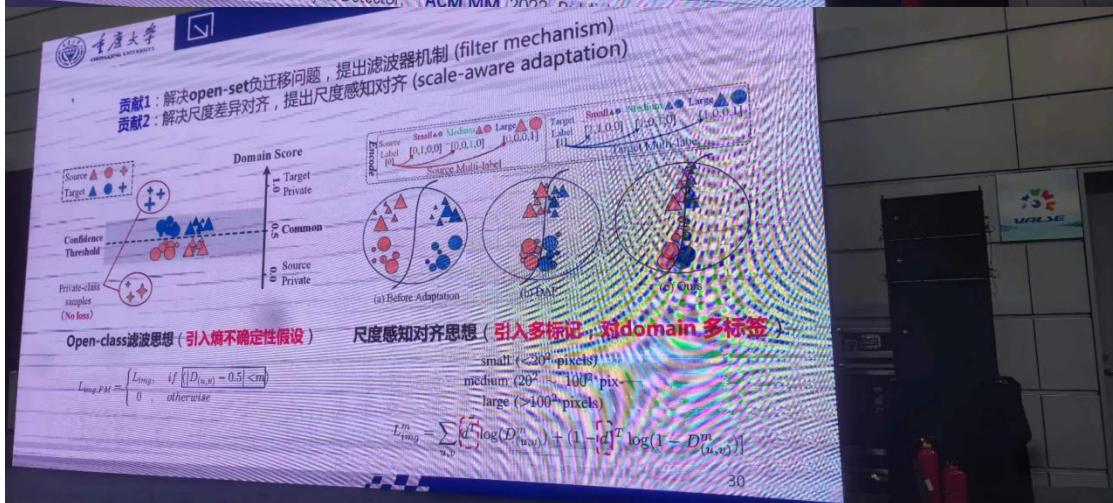
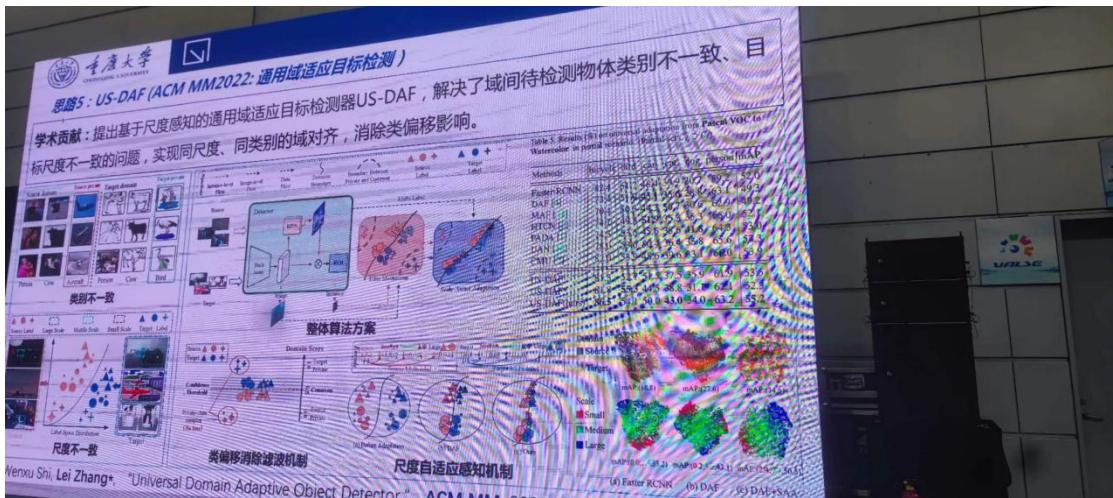
$$\epsilon^t(h) \leq \sum_{i=1}^M \pi_i^* \epsilon^i(h) + \frac{\gamma + \rho}{2} + \lambda_{H_t}(P_X^t, P_X^t)$$

其中， $\gamma := \min_{\pi \in \Delta_M} d_H(P_X^t, \sum_{i=1}^M \pi_i P_X^i)$ （域偏移：不能太大）
 $\rho := \sup_{P_X^t, P_X^i \in \Delta_X} d_H(P_X^t, P_X^i)$ （凸包的直径：不能太大）

目标检测的五大问题：



报告人提出的相关工作：主要贡献是：解决了 open-set 负迁移问题，提出了滤波器机制；解决了尺度差异问题，提出了尺度感知对齐机制。



Workshops-视觉知识和多重知识表达

汇报人：付彦伟

单位：复旦大学

个人简介：付彦伟，博士，复旦大学大数据学院青年研究员，博士生导师，东方学者、国家青年千人计划学者。2014 年获得伦敦大学玛丽皇后学院博士学位，2015.01-2016.07，在美国匹兹堡迪士尼研究院任博士后研究员。付博士发表高水平论文 100 多篇:在 IEEE TPAMI 发表通讯作者/第一作者论文 11 篇，论文曾获得 IEEE ICME 2019 最佳论文，获得美国发明专利 6 项、中国专利 10 多项。研究方向侧重于基于迁移学习的多个任务，如 3D /4D 物体的建模；神经网络稀疏化学习、机械臂抓取；图像编辑及修复等。

报告主题：先验信息引导的图片内容生成与编辑



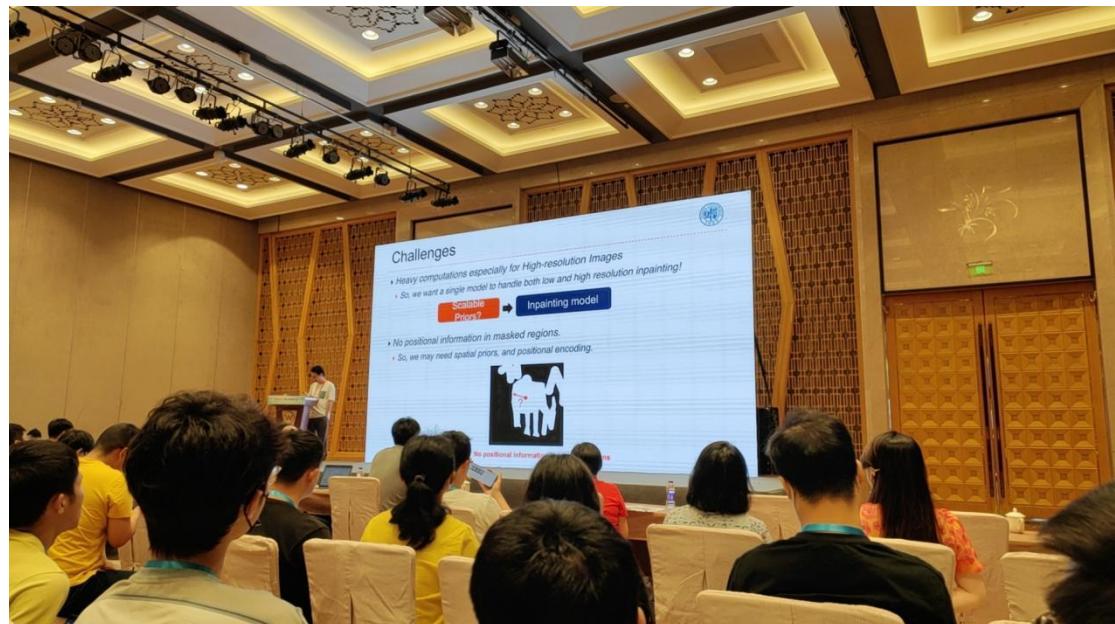
Paper-List

- › Image Inpainting
 - › Learning a Sketch Tensor Space for Image Inpainting of Man-made Scenes. Chenjie Cao, Yanwei Fu. ICCV 2021
 - › Incremental Transformer Structure Enhanced Image Inpainting with Masking Positional Encoding. Qiaole Dong, Chenjie Cao, Yanwei Fu. CVPR 2022
 - › Learning Prior Feature and Attention Enhanced Image Inpainting. Chenjie Cao, Qiaole Dong, Yanwei Fu. ECCV 2022
 - › ZITS++: Image Inpainting by Improving the Incremental Transformer on Structural Priors. Chenjie Cao, Qiaole Dong, Yanwei Fu. IEEE TPAMI, to appear
- › Image Manipulation
 - › ManiTrans: Entity-Level Text-Guided Image Manipulation via Token-wise Semantic Alignment and Generation. Jianan Wang, Guansong Lu, Hang Xu, Zhenguo Li, Chunjing Xu, Yanwei Fu. CVPR 2022
 - › Wang et al. Entity-Level Text-Guided Image Manipulation, arxiv 2023


Chenjie Cao Qiaole Dong Jianan Wang

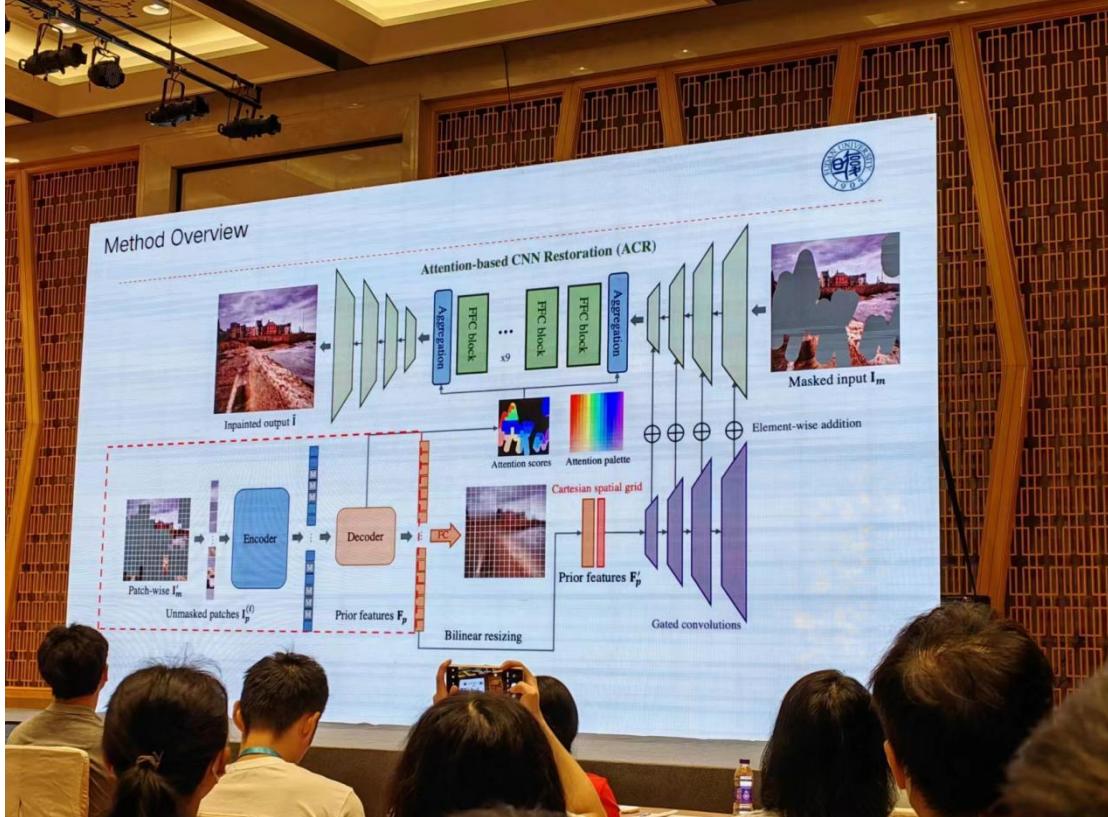
报告总结:多模态学习成为 AI 的热门技术趋势之一,尤其是在新内容生成方面。而得益于深度神经网络的发展,以图像修复、编辑及内容生产为代表的新一代图像内容智能技术,近年来取得了巨大的进步。以 Transformers、生成对抗网络、Diffusion model 等为代表的技术,成就了众多优秀的图像生成模型。先验信息指导,对图片内容生成和编辑尤为重要。相关工作在该领域取得了成功,促进了更多的模型/算法性能改进和有效的图像合成。本次报告主要介绍了报告人的相关工作。

首先是图像编辑任务所面临的困难与挑战: **1、希望得到一个模型能够同时对低分辨、高分辨图像进行图像编辑; 2、掩码的区域没有相应的位置信息。**

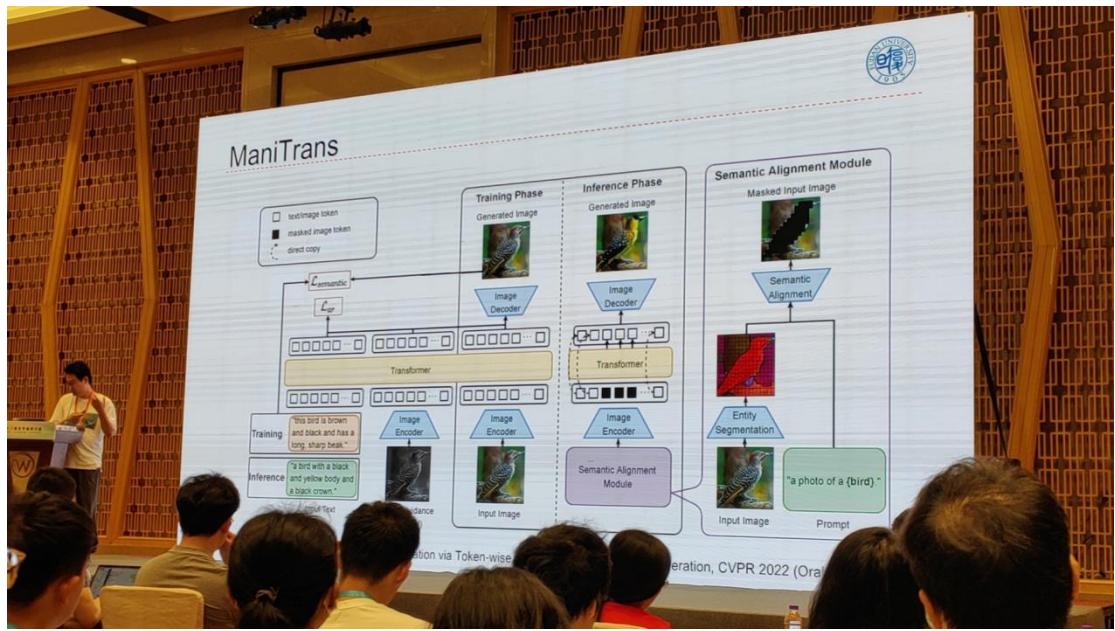


报告人提出了 MST 模型, **主要创新点在于提取图像中的线条信息作为先验引入图像编辑任务中。**

后续进一步结合 MAE (mask autoencoder) 模型对图像中的先验特征进行提取。但是,该先验特征并未有显性的物理含义,只是在实际实验中对掩码图像恢复效果有提升,对于 ISAR 图像、红外图像等具有显著物理意义的图像需要进一步考虑 MAE 模型的适应性改进。



值得关注的是，报告人最近的工作将传统的图像编辑任务与多模态任务进行了结合，提出了文本引导的实体级图像操作 ManiTrans：



模型在训练阶段利用结合了语义对齐损失 $L_{semantic}$ 指导 transformer 捕捉文本与图像的对应关系；测试推断阶段则是利用目标文本、prompt 与原始图像，将原始图像中的目标进行掩码，再由训练好的 transformer 进行修改。

汇报人：谢伟迪

单位：上海交通大学

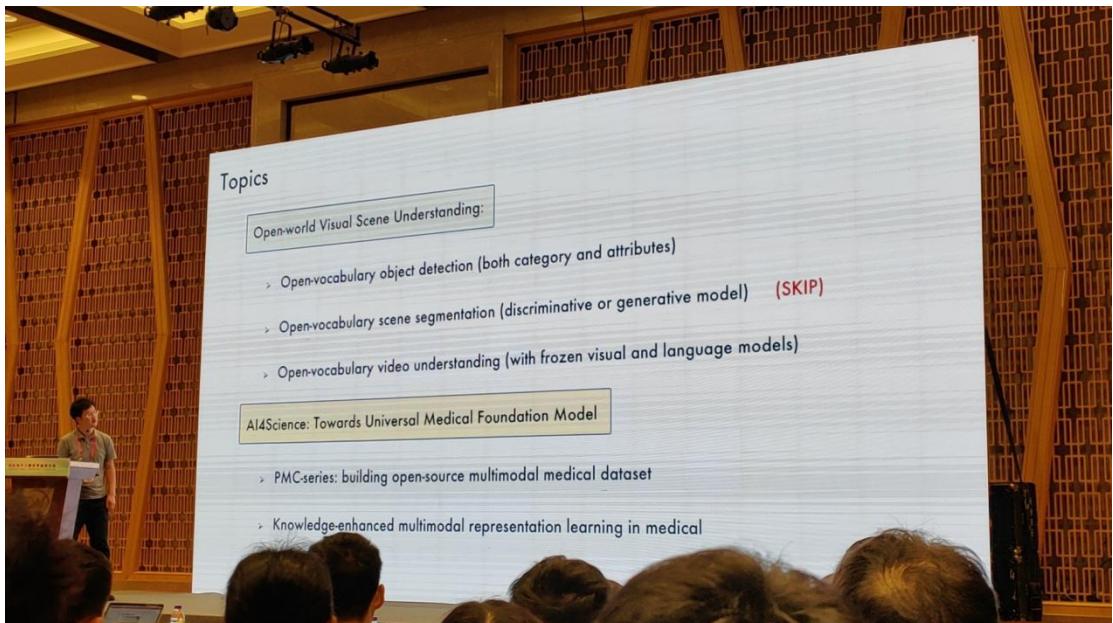
个人简介：谢伟迪，上海交通大学副教授、博士生导师。2018 年于英国牛津大学视觉几何组 (VGG) 获得博士学位，先后担任牛津大学博士后研究员，高级研究学者，获 Oxford-Google DeepMind Scholarship, Magdalen Award (China-Oxford Scholarship Funds), Oxford Excellence Award, 上海市领军人才，科技部“新一代人工智能”重大项目青年项目负责人。发表论文 40 余篇，Google Scholar 引用超 6500 次，开源多个标准领域数据集合，包括 VGGFace2, Voxceleb VGGSound, MoCA，下载量超 25 万次。担任 CVPR2023, NeurIPS2023 领域主席。主要研究领域为大规模多模态表征学习。

个人主页：<https://weidixie.github.io>.

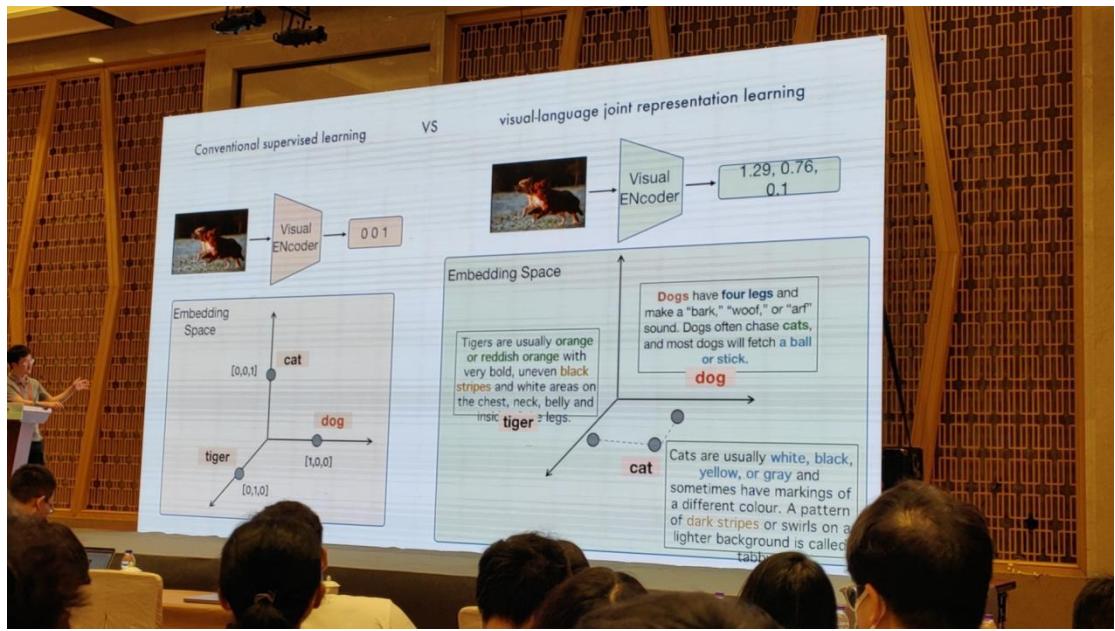
报告主题：基于知识驱动的多模态表征学习



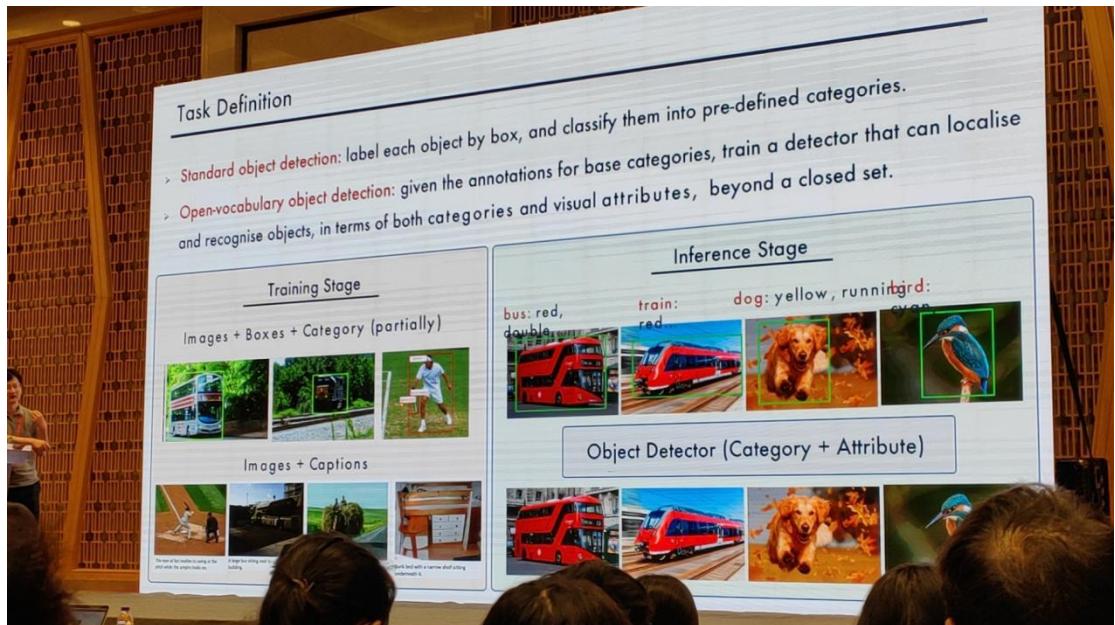
报告总结：该报告主要分为两部分，第一部分介绍了在知识引导下的目标类别、属性检测与视频理解，第二部分则与医学图像相关。



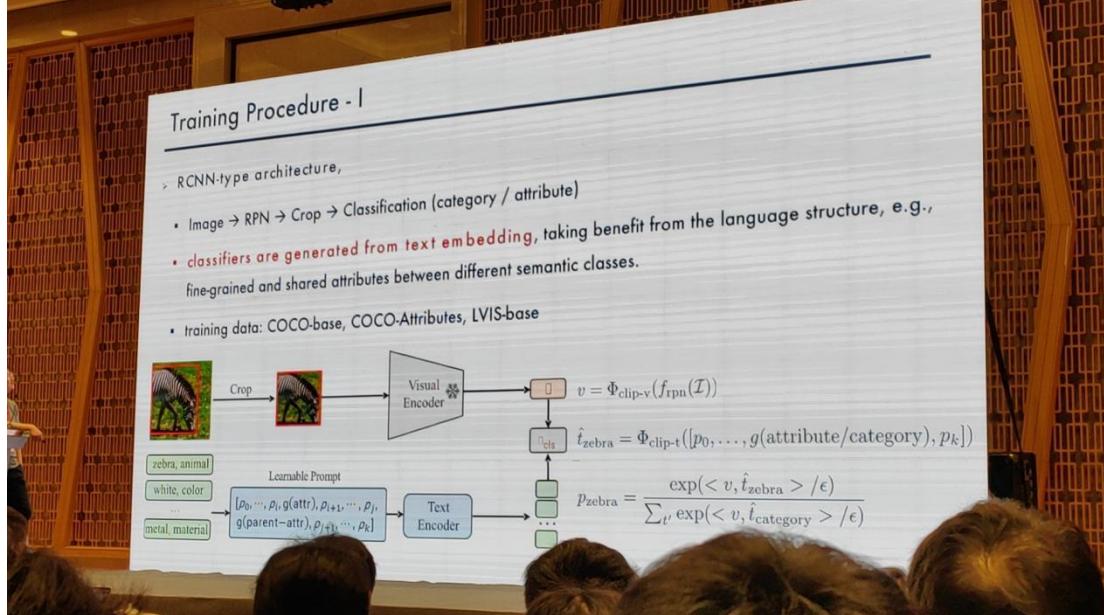
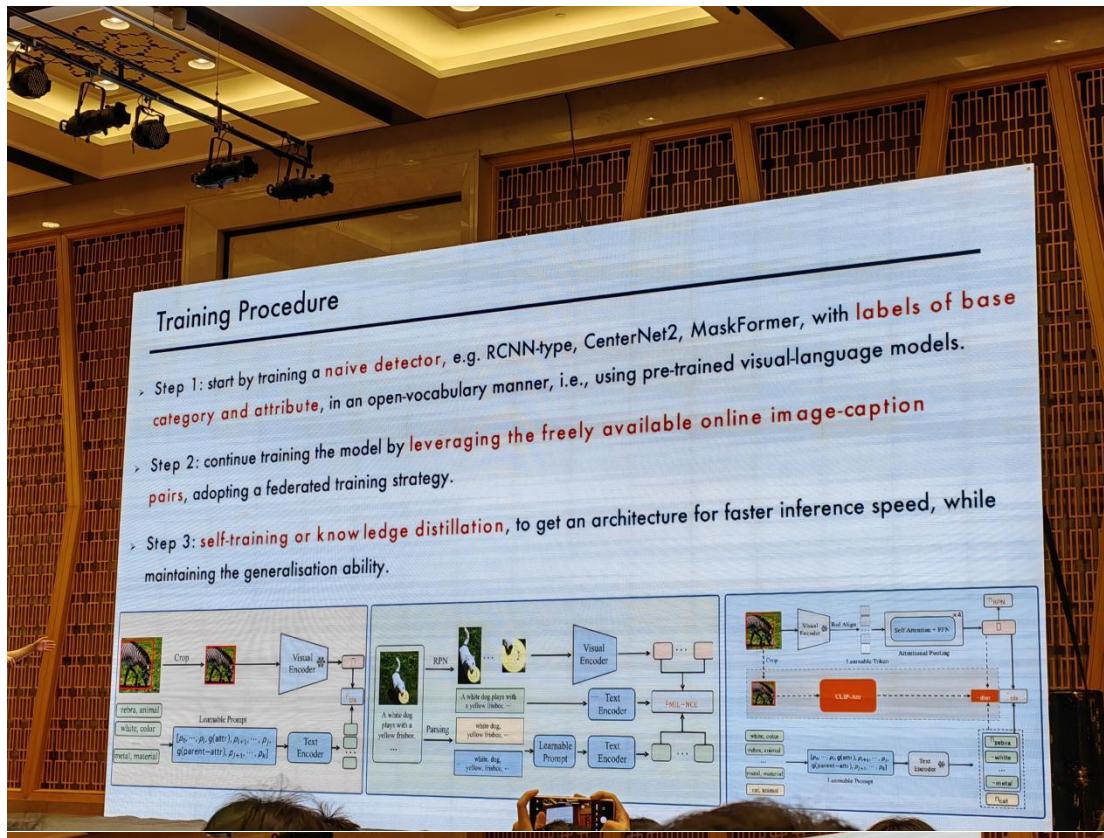
第一部分的第一小节主要基于 CLIP 的架构对图像进行目标类别、属性的检测。现有图像、语句 embedding 过程的缺点：不能体现出与其他词语、token 之间的关联。



报告人对其提出的开放词目标检测给出了定义，即目标检测识别不仅能划定目标区域的检测框，还能更预测目标物体的类别与属性。

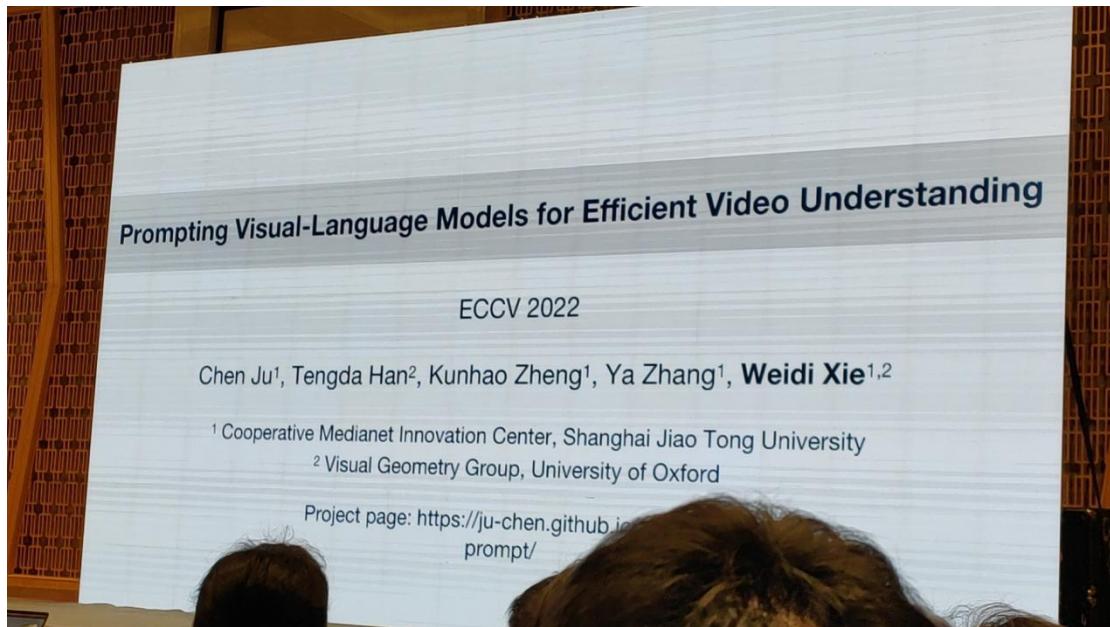


报告人采用三个阶段对模型进行训练，模型、模块基于现有工作。

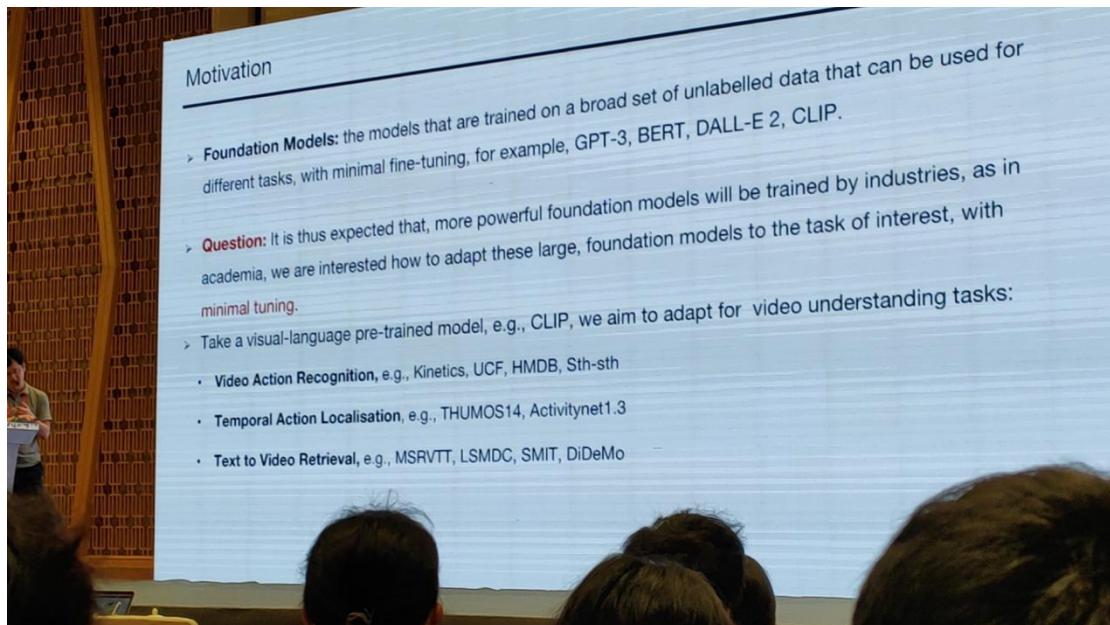


在后续工作中，报告人阐述知识驱动表征学习在面向真实场景中的应用，例如，开放集物体检测、开放集动作识别、开放集物体分割、电影自动语音字幕等多个领域的应用效果，并探讨如何进一步提高其性能和可解释性。

近期工作：视频理解



主要包含三个任务：动作识别、动作定为、文本-to-视频



VALSE 2023 6-12

VALSE2023 汇报整理 6-12

Workshops-AI for science

汇报人：杨跃东

单位：中山大学

报告题目：融合 HPC 和 AI 的药物分子设计



Tips: AI 在制药领域的应用，给出了总体平台和框架。杨老师认为 AI for Science 的重点应该是 Science，AI 对我们来说只是工具，如何将运用好 Science 领域的专家知识结合 AI 求解问题是未来的趋势。

汇报人：谢凌曦

单位：华为

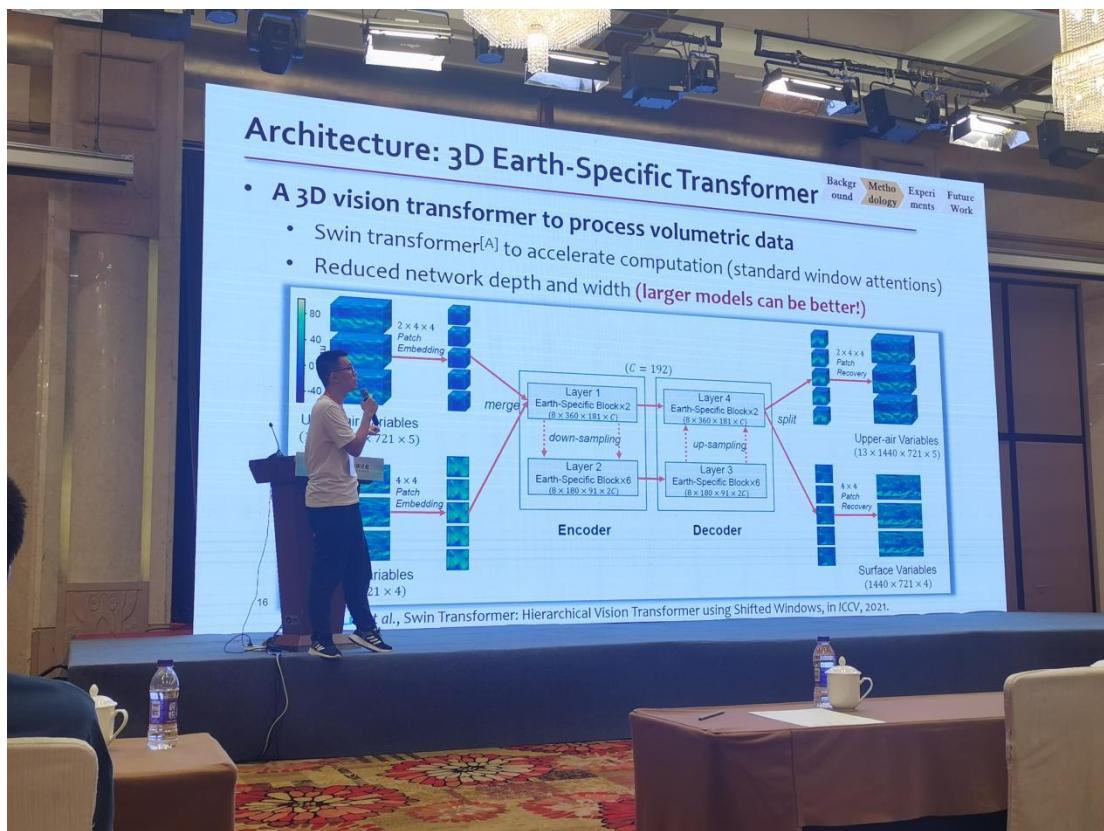
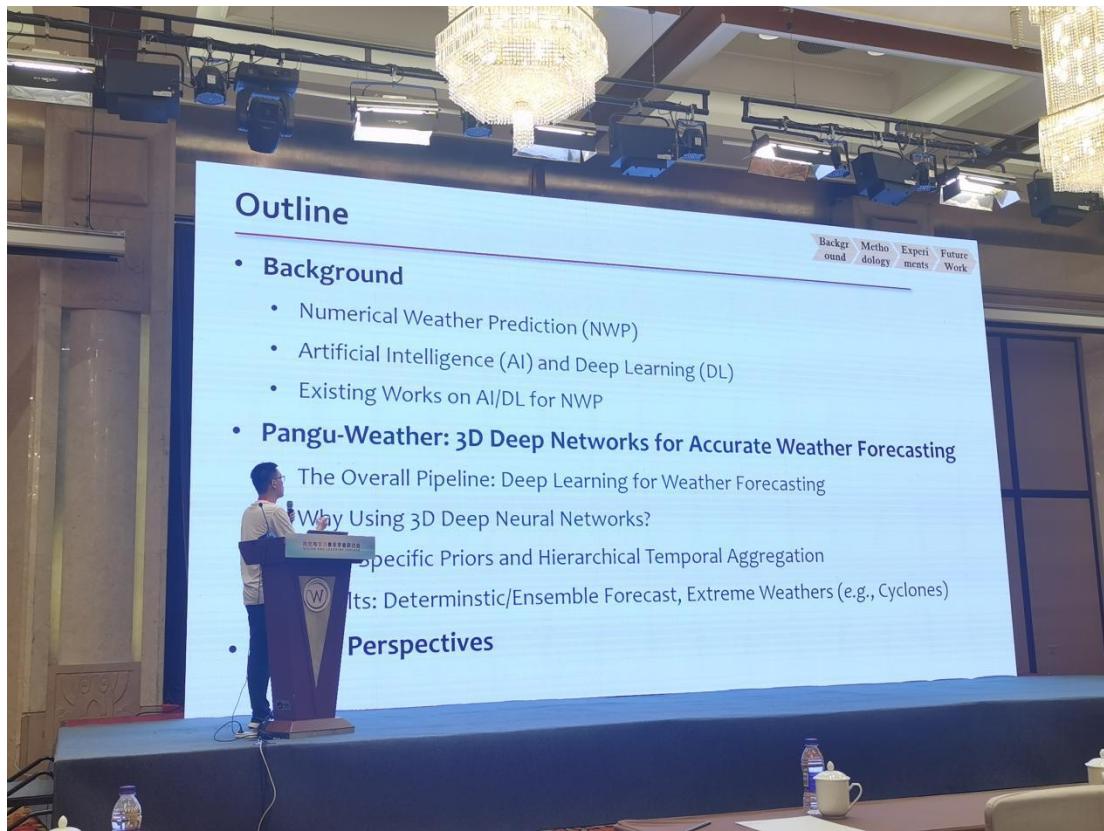
盘古气象大模型：3D 全球高分辨率气象预报方法

报告总结：数值天气预报在每日天气预报、极端灾害预警、气候变化预测等领域发挥着重要作用，但是随着算力增长的趋缓和物理模型的逐渐复杂化，传统数值预报的瓶颈日益突出。研究者们开始挖掘新的气象预报范式，如使用深度学习方法预测未来天气。在数值方法应用最广泛的中长期预报任务中，AI 预报方法精度仍然显著低于数值预报方法。本次报告中，我将介绍盘古气象大模型，一种新的高分辨率全球 AI 气象预报系统。盘古气象大模型是首个超过传统数值预报精度的 AI 方法，1 小时-7 天预测精度均高于传统数值方法（欧洲气象中心的 operational IFS），同时预测速度提升 10000 倍，能够在秒级时间内提供全球气象预报。古气象模型的水平空间分辨率达到 0.25 度，时间分辨率为 1 小时，覆盖 13 层垂直高度，可以精准地预测位势、湿度、风速、温度、海平面气压等气象特征。作为基础模型，盘古气象大模型能够直接应用于下游气象预报场景，例如在热带风暴轨迹预测中，盘古气象大模型的预测精度显著超过欧洲气象中心的细网格预报结果。日前，相关论文已经被 nature 正刊接收 三篇代表性工作：

[1] Duan, K., Bai, S., Xie, L., Qi, H., Huang, Q., Tian, Q. Centernet: Keypoint triplets for object detection. ICCV, 2019, (pp. 6569-6578), Google Citation 1869;

[2] Pangu-Weather: A 3D High-Resolution Model for Fast and Accurate Global Weather Forecast, arXiv, 2022, Google Citation 10;

[3] Xu Y, Xie L, Dai W, et al. Pc-darts: Partial channel connections for memory-efficient architecture search. ICLR, 2020, Google Citation 588;





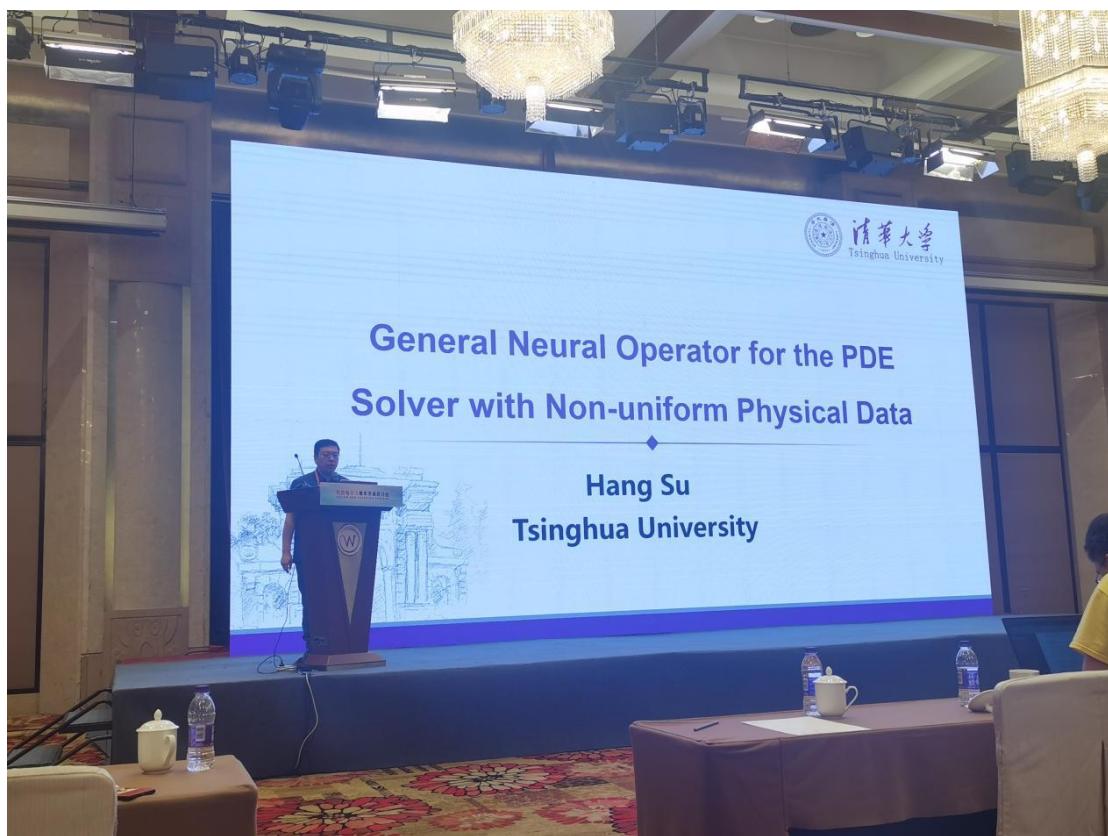
Tips: 谢老师全面系统的介绍了如何运用现在的 AI 大模型来解决气象预报的问题, 端到端的 AI 算法在有大量的数据的前提下可以实现更好的性能和计算时间, 但是数据、计算资源仍然是问题, 但是在精心调参、设计网络以及优化训练策略的情况下, AI 算法的网络结构以及迭代次数是可以大大缩减的, 只是轻微的降低算法性能。

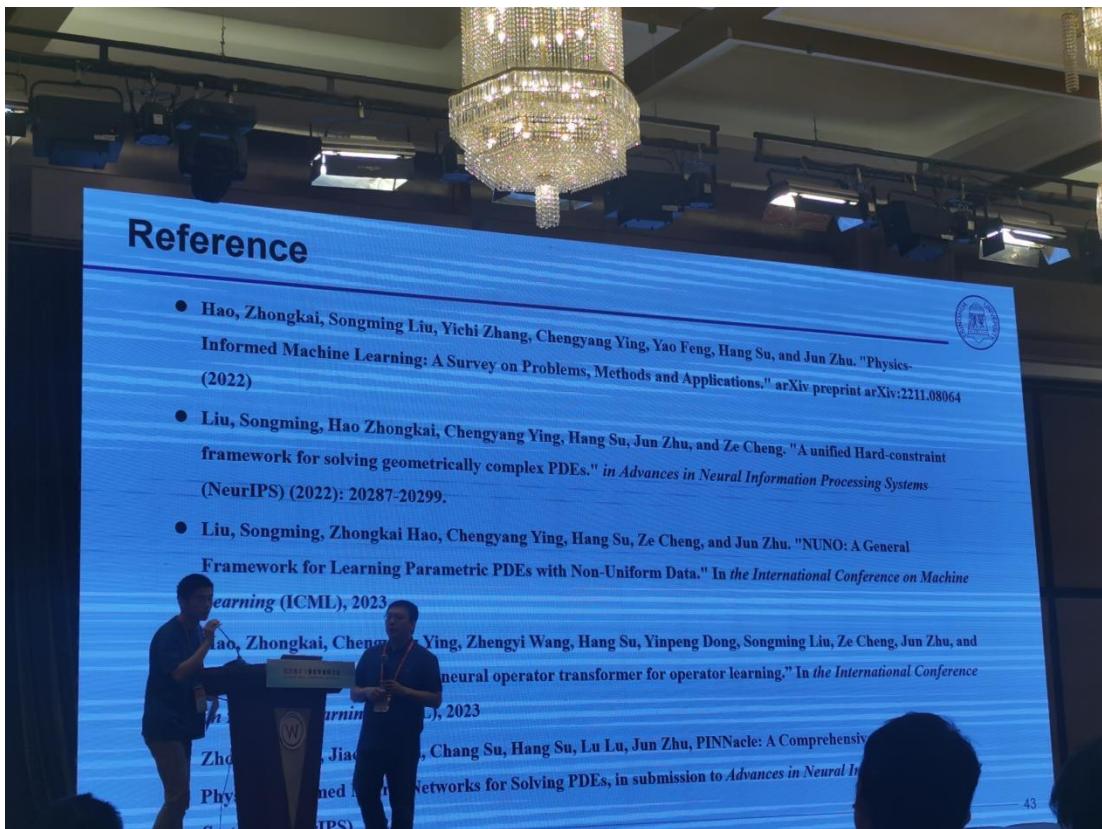
汇报人：苏航

单位：清华大学

报告题目：非均匀物理数据分布的偏微分方程（PDEs）的通用求解算子

报告总结：物理信息驱动的机器学习（PIML）作为一种新型机器学习范式，有效地将物理的先验知识和实验数据融合，赋能机器学习模型在面临高维度和不确定性问题时，得到更贴合物理规律的解决方案。本报告专注于非均匀物理数据分布的偏微分方程（PDEs）求解问题，探讨基于机器学习的神经算子求解方法。首先，对于几何边界条件复杂的 PDE 求解问题，我们引入了混合有限元方法的 Extra Field，构建了统一而高效的求解框架。随后，我们提出了通用神经运算符变换器（GNOT），这是一个基于变换器的学习运算符框架，能够处理多个输入函数和不规则网格。最后，我们引入了非均匀神经运算符（NUNO），这是一个针对非均匀数据设计的高效运算符学习的通用框架。同时，为了建立了全面、一致的基准测试，评估不同物理信息神经网络（PINNs）方法在解决偏微分方程中的效果，我们建立了一个 PINN 的评估测试基准 PINNacle，以期能为科学机器学习方法的发展提供有力工具。





[1] NUNO: A General Framework for Learning Parametric PDEs with Non-Uniform Data;

[2] A unified Hard-constraint framework for solving geometrically complex PDEs;

[3] Bi-level physics-informed neural networks for pde constrained optimization using broyden's hypergradients.

Tips: 苏老师从优化理论的角度出发，提出了一个新的物理信息驱动的机器学习框架，并给出了一些证明，在各种实验条件下，都有着非常好的性能和泛化性能。

汇报人：欧阳万里

单位：上海人工智能实验室

报告题目：从计算机视觉到 AI4Science-挑战与机遇

报告总结：以深度学习为代表的人工智能算法取得了飞速的发展，并大规模地应用到人类的生产生活实践中。Valser 们大量学者关注计算机视觉问题，讲者的科研经历也是如此。另一方面，将人工智能技术应用到科学研究，利用人工智能算法解决当前科学的未解问题已经成为产学研关注的重点。本次报告将介绍这两个不同科研课题的共性、区别以及介绍 AI4Science 这一对于 Valser 们比较新的课题中的挑战与机遇。同时将介绍上海人工智能实验室在 AI4Science 研究（包括材料、生物、气象、天文）的既有工作和未来探索。作为其中的一个工作，将介绍实验室最近的中期天气预报大模型“风鸟”。



[1] Chen, K., Han, T., Gong, J., Bai, L., Ling, F., Luo, J.J., Chen, X., Ma, L., Zhang, T., Su, R. and Ci, Y., 2023. FengWu: Pushing the Skillful Global Medium-range Weather Forecast beyond 10 Days Lead. arXiv preprint arXiv:2304.02948.

[2] Pang, J., Chen, K., Shi, J., Feng, H., Ouyang, W., & Lin, D. (2019). Libra r-cnn: Towards balanced learning for object detection. In

Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 821-830), Google Citation 1100

[3] Chen, K., Pang, J., Wang, J., Xiong, Y., Li, X., Sun, S., ... & Lin, D. (2019). Hybrid task cascade for instance segmentation. In **Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4974-4983)**, Google Citation 973

Tips: 欧阳老师是最早一批做 AI 计算机视觉的，也有很多的开创性工作，但是现在开始往 AI for science 方向发展，他认为传统 AI 研究逐渐陷入瓶颈，突破的关键是结合 science 领域积累的数据、经验、知识。

Workshops-多模态大模型与提示学习

报告人：左旺孟

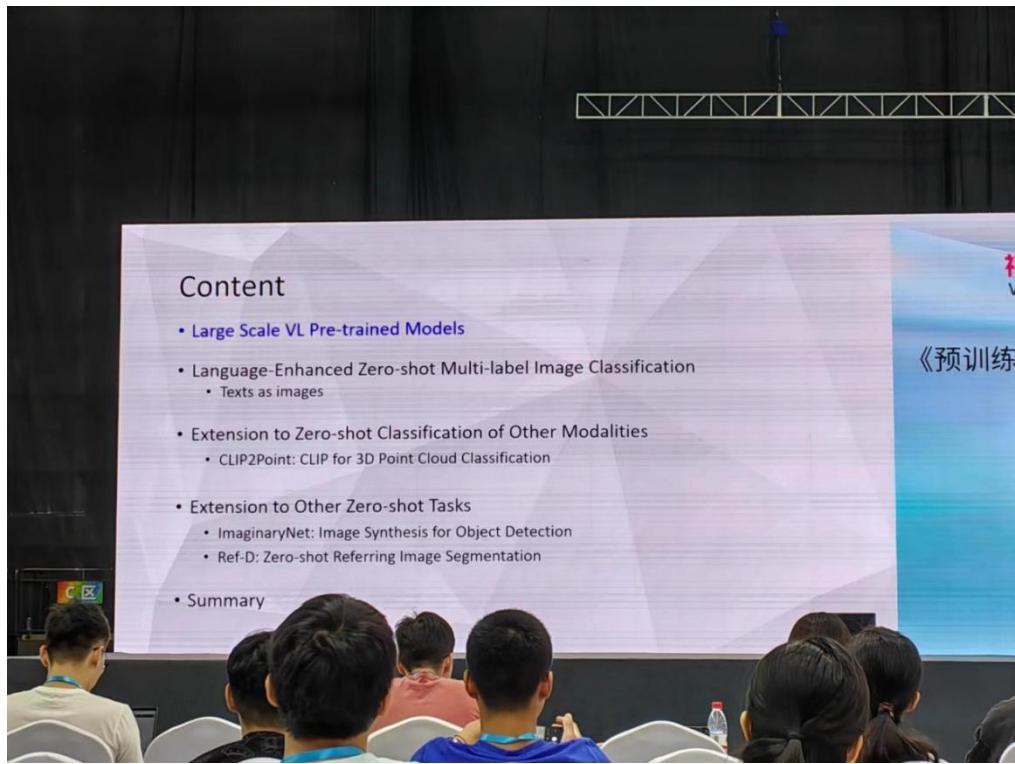
单位：哈尔滨工业大学

个人简介：左旺孟，哈尔滨工业大学计算机学院教授、博士生导师。主要从事图像增强与复原、图像编辑与生成、物体检测与目标跟踪、图像与视频分类等方面的研究。在 CVPR/ICCV/ECCV 等顶级会议和 T-PAMI、IJCV 及 IEEE Trans. 等期刊上发表论文 100 余篇。曾任 ICCV2019、CVPR2020/2021、ECCV 2022 等顶级会议领域主席，现任 IEEE T-PAMI 和 T-IP 等期刊编委。

报告主题：预训练模型和语言增强的零样本视觉学习

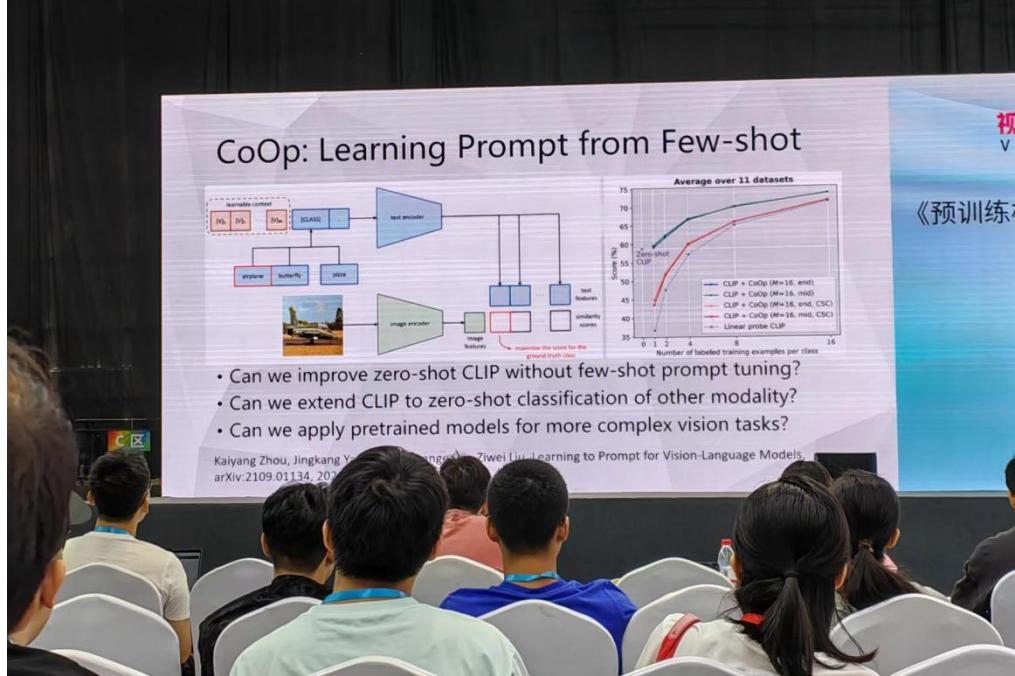
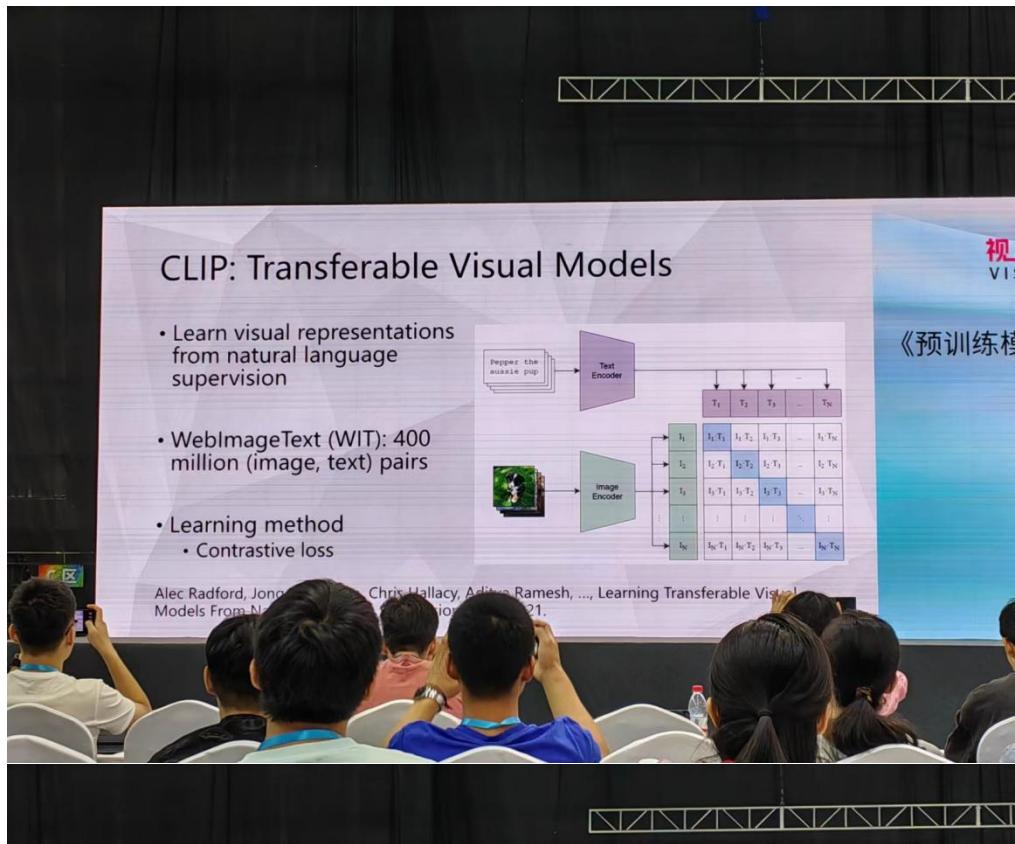
报告总结：



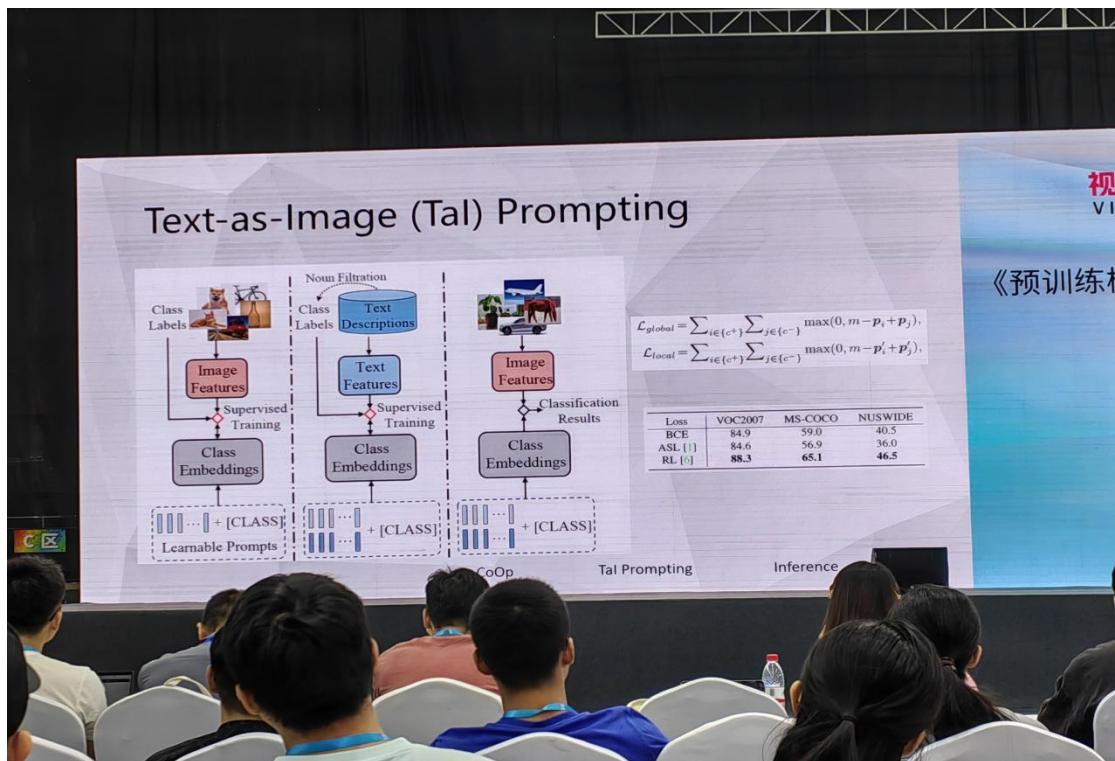
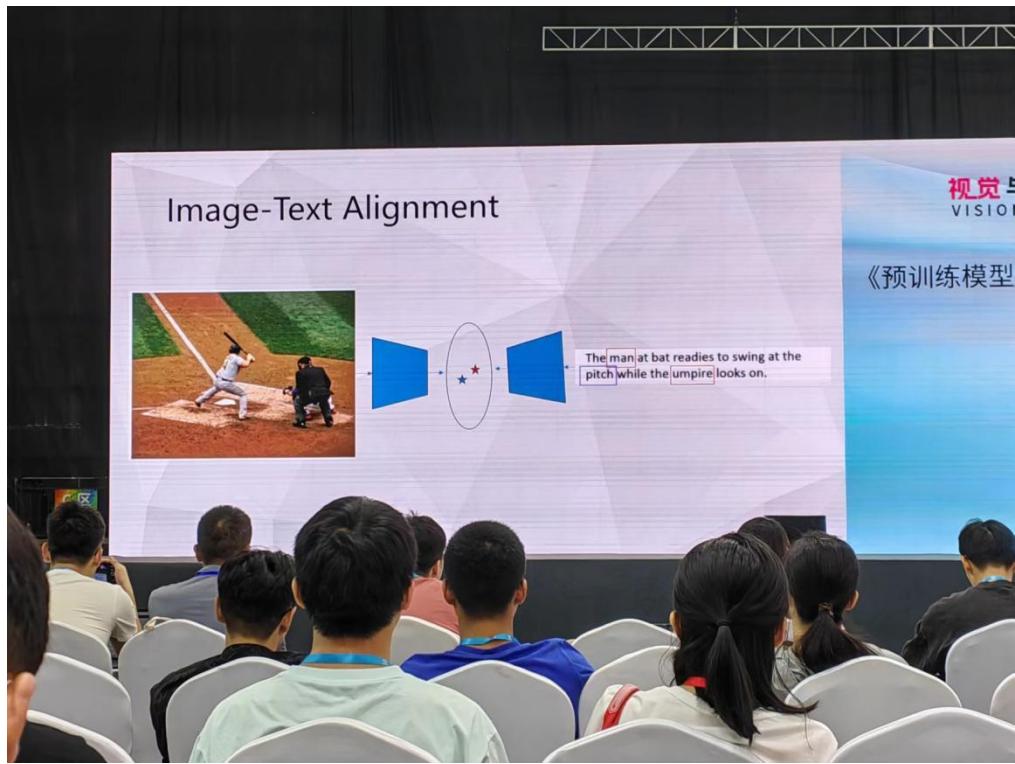


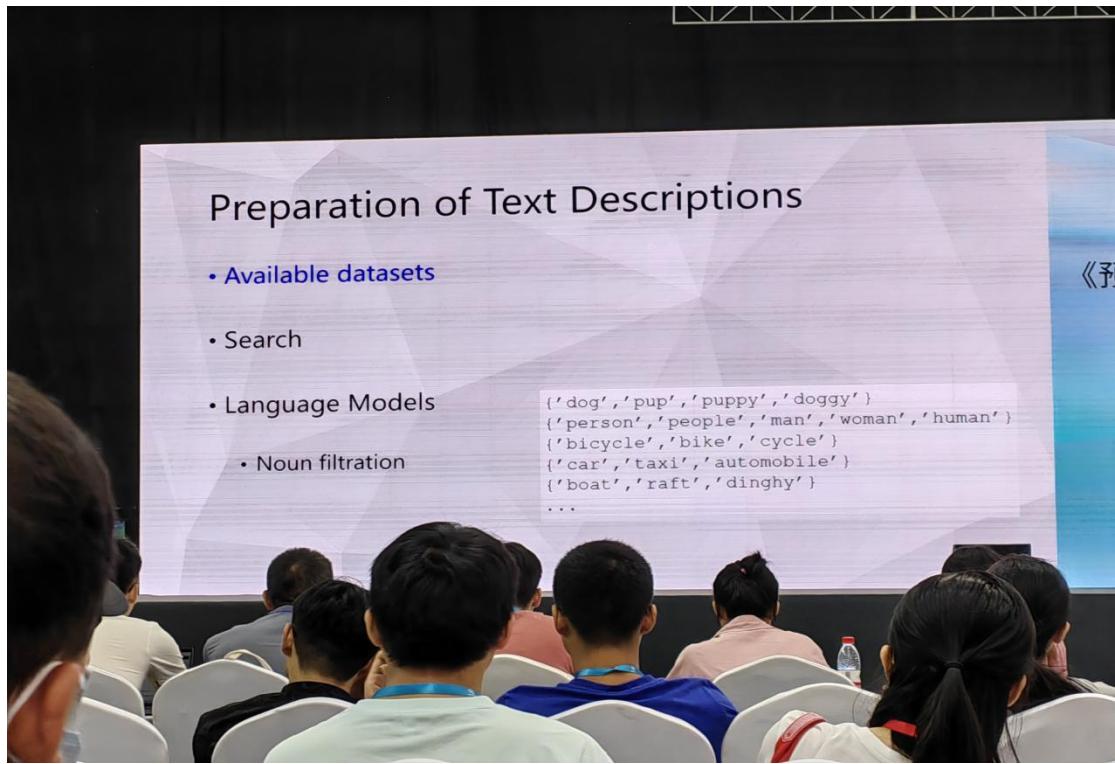
报告总结主要分为四个部分：首先是传统的大尺度下预训练的视觉模型；其次是语言增强的零样本多标签图像分类模型；零样本分类在其他模态下的扩展以及在其他零样本任务上的扩展，最后是总结。

零样本传统预训练视觉模型：CLIP、CoOp

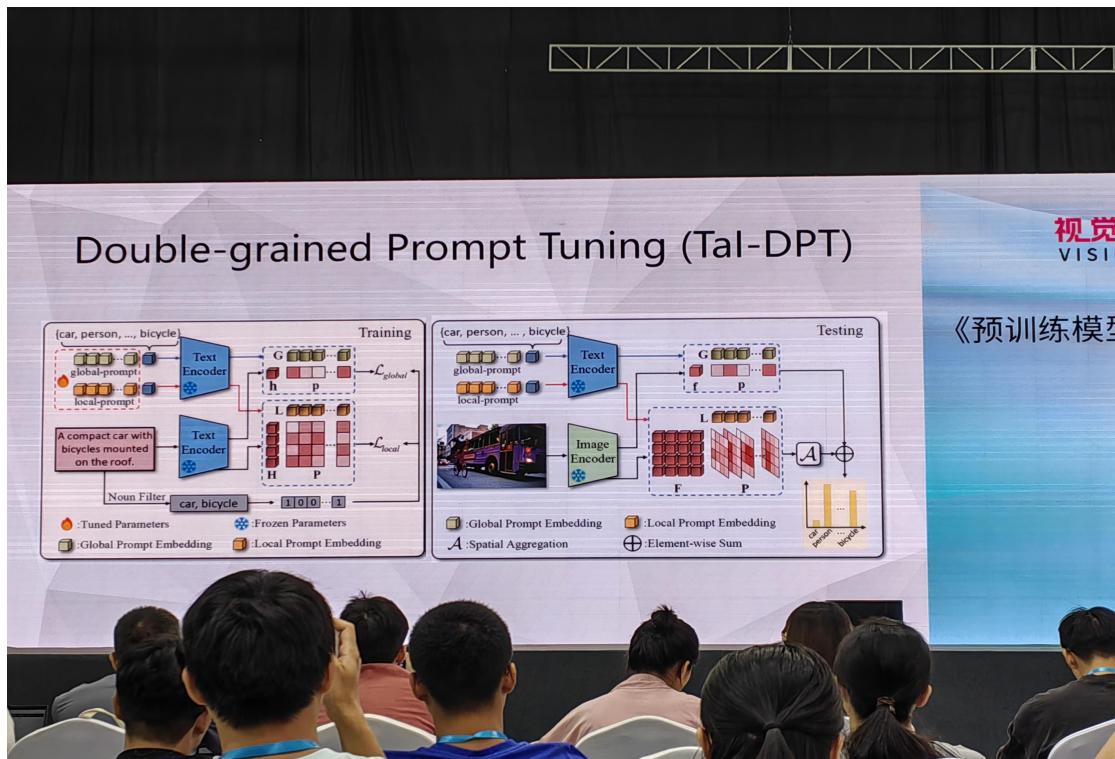


接下来考虑多标签的图像 zero-shot 多标签分类问题, 报告人想利用语句中丰富的信息来引导模型, 首要问题就是利用文本特征替换图像特征, 该部分需要利用 prompt 部分。



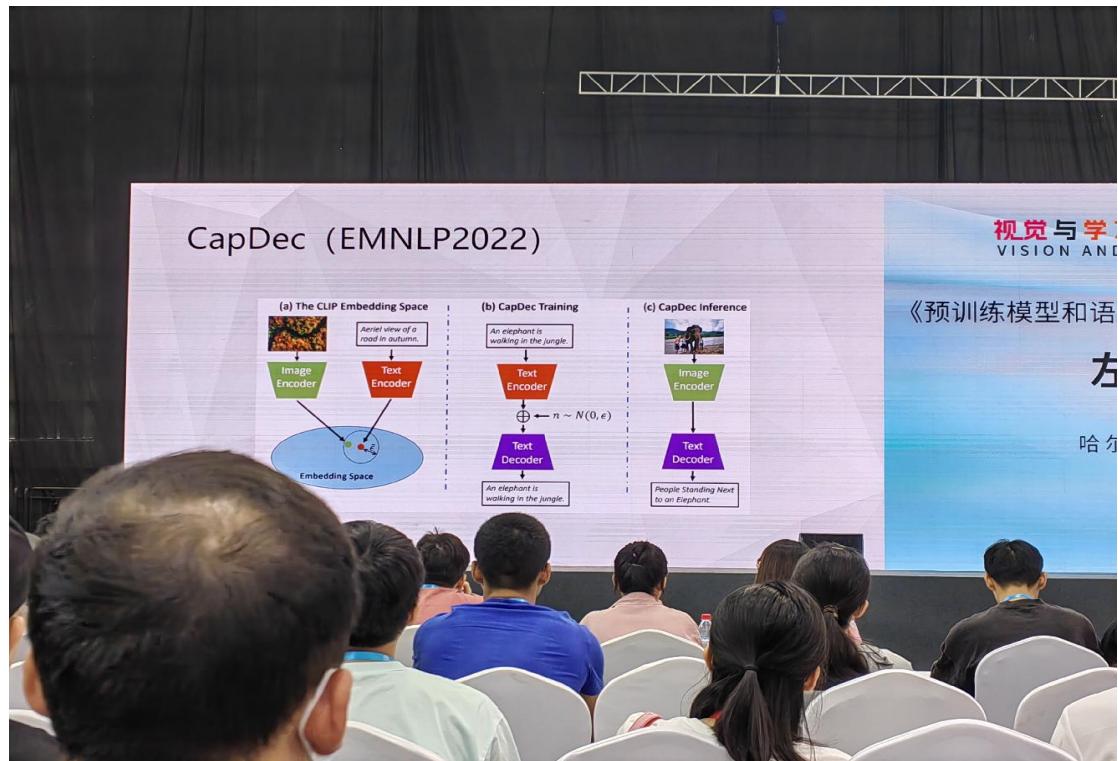


在数据预处理部分，对于文本信息需要构建一个名词过滤器提取出语句中的名词以及语义相近的名词（作为分类类别使用）。报告者提出的相关模型：



该模型训练阶段输入皆为文本语句，整体结构除了 prompt 部分以外，皆为预训练好的模型。其中，global loss 关注输入的语句里面是否有与 prompt 对应的标签类别信息，local loss 则关注输入语句里面是否有与 prompt 相近的类别信

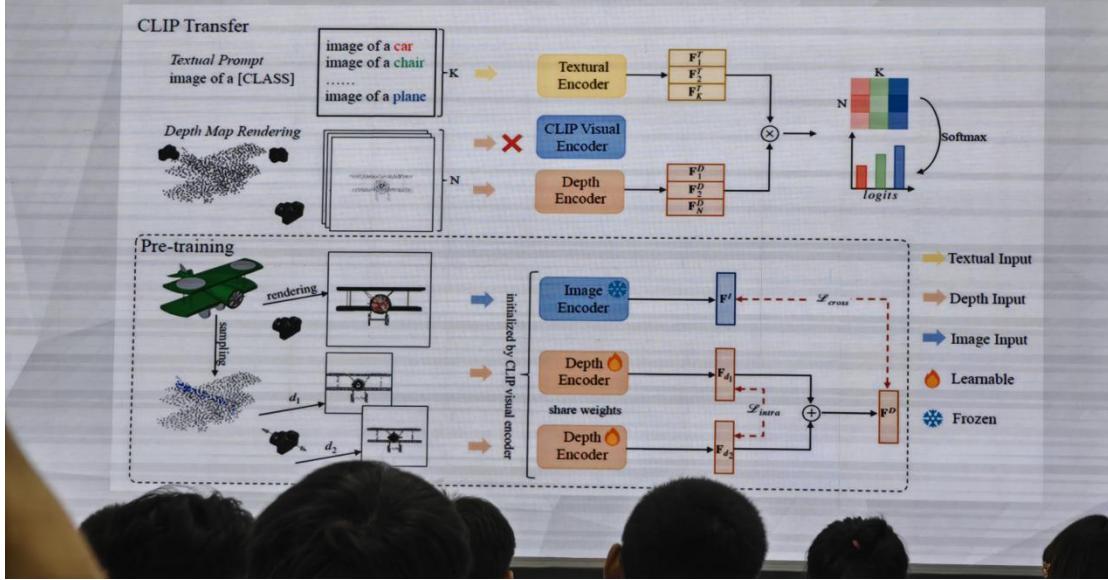
息。测试阶段将文本语句信息替换为图像，实现对图像的 zero-shot 多标签分类：此时输入语句改为图像，global 损失不变，local 变为与图像中的 token 相比。后续跟进的一些相关工作。该工作主要的创新点在于，文本特征与图像特征对齐后还存在 gap，需要在文本信息提取的特征加入高斯噪声，以期减少两模态特征之间的距离，提高模型鲁棒性。



工作拓展：

该方法在更多模态下的应用，例如 3D 点云。

CLIP2Point



主要思想很简单：对齐 3D 点云与图像，对齐图像与文本，实现 3D 点云与文本的对齐。该思想的集大成者是最近提出的 ImageBind，以图像为中间媒介实现不同模态之间的对齐。

Tutorial-扩散模型

时光倒转万物生：扩散模型与AIGC

VALSE 2023 扩散模型讲习班

汇报人：李崇轩

中国人民大学 高瓴人工智能学院



汇报人：李崇轩

单位：中国人民大学

个人简介：博士，中国人民大学准聘助理教授，博士生导师。2019年博士毕业于清华大学，2021年加入中国人民大学高瓴人工智能学院。研究方向为深度概率学习，相关工作发表于机器学习领域重要国际会议、期刊40余篇。代表性工作有：一致性理论下最优的半监督GAN方法 Triple-GAN；扩散概率模型在最大似然意义下的最优反向方差估计 Analytic-DPM；文图通用的多模态扩散概率大模型 Unidiffuser。李崇轩曾荣获机器学习领域重要国际会议 ICLR 杰出论文奖，吴文俊人工智能自然科学奖一等奖，吴文俊人工智能优秀青年奖，中国计算机学会优秀博士学位论文奖，北京市科技新星，中国博士后创新人才支持计划，主持国家自然科学基金面上项目，教育部产学研结合协同育人项目等。

报告总结：扩散概率模型是一类新涌现的深度生成模型，这类方法逐步地对先验噪声分布去噪得以逼近数据分布。目前，扩散概率模型在数据合成质量、采样的多样性和数据密度估计等指标下取得了超越 VAE、GAN、FLOW 等经典深度生成模型的结果，也在诸多领先的视觉生成大模型中得到应用，有望成为未来通用生成式智能的重要组成部分。本次课程会介绍扩散概率模型的基本原理与代表性工作，介绍其在可控生成、多模态建模和三维场景建模等方面的前沿进展，并对扩散概率模型的下一步发展做简单的展望。

报告提纲

- 概览
- 扩散模型学习算法
- 扩散模型采样算法
- 大规模扩散模型
- 扩散模型与 AIGC
- 展望

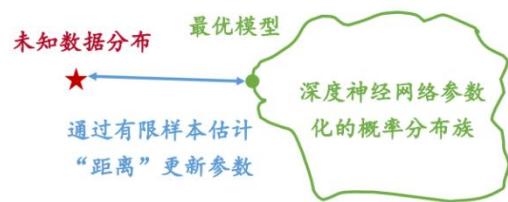
Tips: 报告主要分为两个半场，上半场讲前三个，偏理论基础和原理，下半场讲后面三个，偏应用。

深度生成模型的基本原理

- 核心问题：高维、复杂的联合概率分布的表示、**学习与推断**



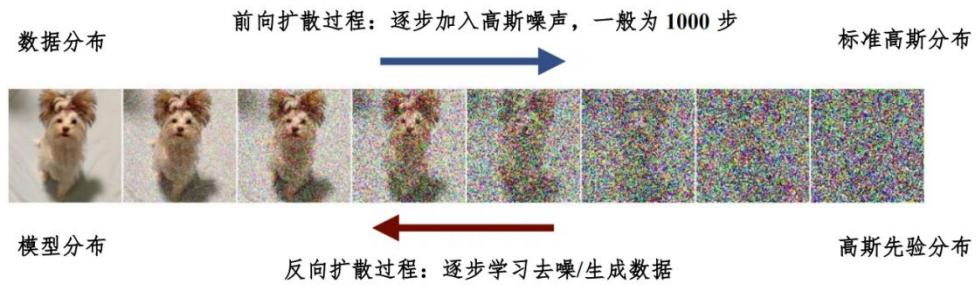
经典图像数据：超过十万维的多峰分布



Tips: 生成模型解决的核心问题是复杂高维的联合概率分布的学习，一张图像可以视为概率分布上的一个点。参数化的 NN 学的是概率分布的表示，学习和优化的过程是使得 NN 离未知分布更近，表示的更好。最终的推断是在学到的分布中进行采样。

扩散模型

- 理论：一定条件下，前向扩散过程对任意输入分布均可逆，且逆过程形式不变
- 直觉：将生成数据这一复杂问题转化为不同噪声级别下的去噪问题（相对简单）



Tips: 扩散模型概览

如何表示高维空间中的联合概率分布？



两种等价理解

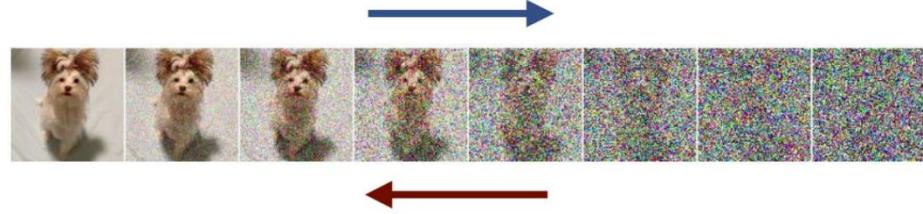
1. 层次化隐变量模型（变分自编码器）
2. 多层次去噪评分匹配（基于评分函数的模型）

自回归 规整流 对抗网络 扩散模型

概率表示

Tips: 生成网络本质上都是在表示高维空间中的联合概率分布，有两个不同的等价理解。

扩散模型



核心思想：学习一个反向过程去噪

- 从任意数据分布出发，均可得到同样的高斯分布（表达能力强）
- 存在唯一的对应反向过程，同样是高斯核马尔科夫链（可学习）

Tips: 扩散模型的表达能力强且可学习

扩散模型

可学习的反向高斯核马尔科夫链



模型分布

$$p(x^{(0)}) \approx q(x^{(0)})$$

高斯先验分布

$$p(x^{(T)}) = \mathcal{N}(x^{(T)}; 0, I)$$

$$p(x^{(t-1)}|x^{(t)}) = \mathcal{N}(x^{(t-1)}; f_{\mu}(x^{(t)}, t), f_{\Sigma}(x^{(t)}, t))$$

理解一：层次化隐变量模型，定义了概率密度 $p(x_0) = \int p(x_0|x_1)p(x_1|x_2)\dots p(x_{T-1}|x_T)p(x_T)dx_{1\dots T}$

模型训练

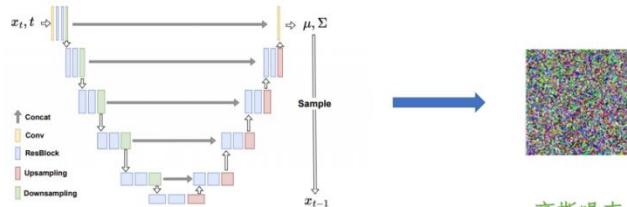
Jonathon et al., NeurIPS 2021

理解二：手工设置方差，做最大似然估计，等价于同时处理 1000 个不同层级的去噪任务

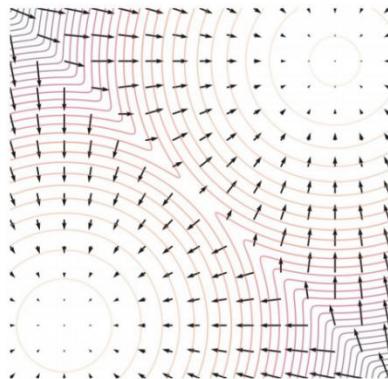
$$\mathbb{E}_{p_D(x_0), \epsilon} \mathbb{E}_{t \sim U[1, 2, 3, \dots, T]} \|\epsilon_{\theta}(x_t, t) - \epsilon\|^2$$

随机噪声大小 噪声预测网络 高斯噪声

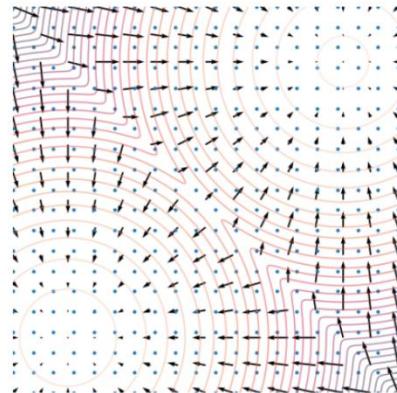
带噪图片，噪声大小与采样的 t 有关
接收带噪图片和时间，预测噪声



评分函数

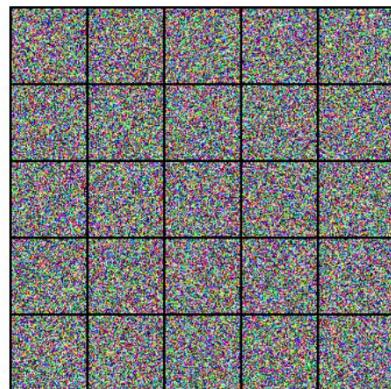
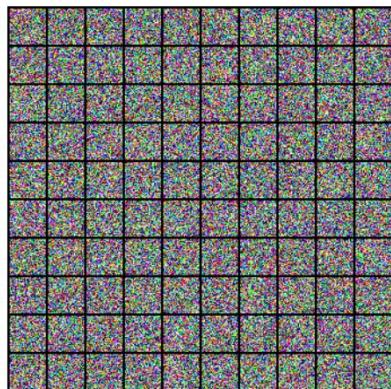


评分函数 $\nabla_x \log p(x)$ (箭头)，指向局部增大概率密度的方向



LD MCMC：按照评分函数方向更新
并注入合适的噪声，可以采样

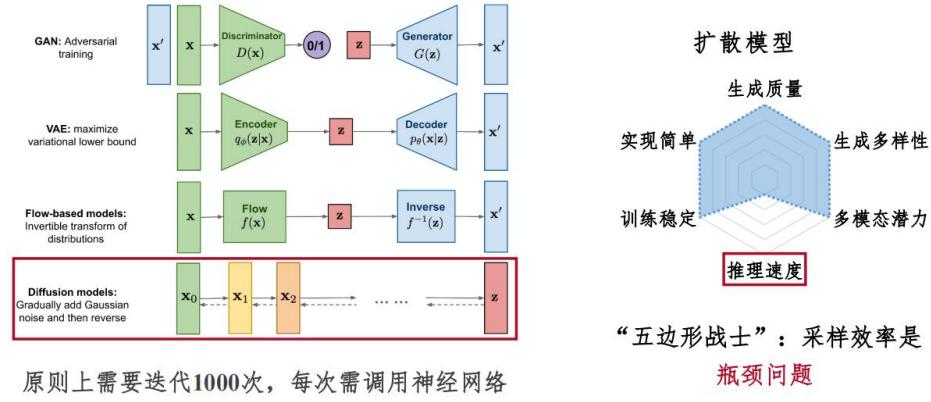
从扩散模型中采样



隐变量模型的祖先采样 或 退火郎之万动力学

Tips: 更新过程可以等效为利用 LD MCMC 对评分函数的更新方向进行采样。

扩散模型的瓶颈问题：采样效率低



Tips: 其实就是实际的 test 成本较高。

高效采样

Review of diffusion models: Yang et al, arxiv 2022

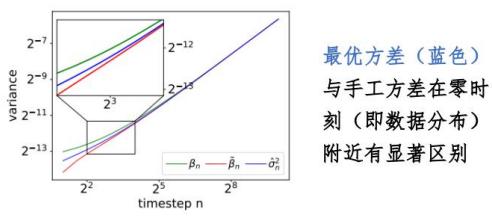


扩散模型的最优采样方差理论 (隐变量模型视角出发)

- 证明最大似然意义下最优采样方差闭式解，改变了手工设计方差的范式
- 发表于机器学习领域旗舰国际会议 ICLR 2022，获杰出论文奖（接收率 0.15%）

定理：扩散概率模型在最大似然意义下关于评分函数/去噪函数的最优采样方差闭式解如下：

$$\sigma_t^{*2} = \frac{\beta_t}{1-\beta_t} \left(1 - \beta_t E_{q_t(x_t)} \frac{\|\nabla \log q_t(x_t)\|^2}{d} \right).$$

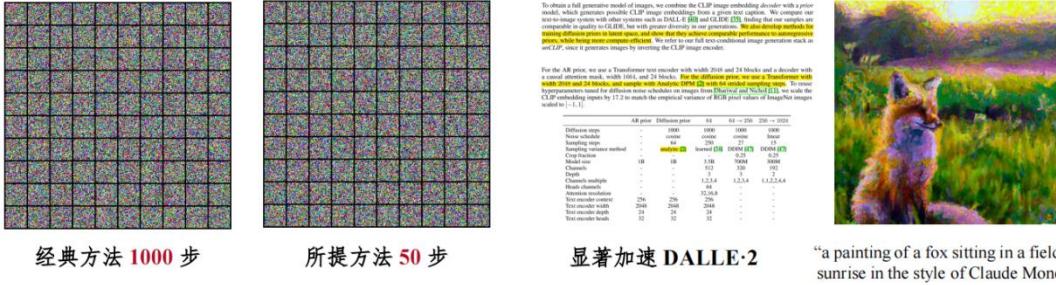


Bao F, Li C, Zhu J, et al. Analytic-DPM: an analytic estimate of the optimal reverse variance in diffusion probabilistic models. ICLR 2022.

Tips: 从采样的角度对扩散模型进行改进

Analytic-DPM

- 无需额外训练，保证合成样本质量不变，加速 **20-80 倍**
- 作为核心技术部署于 **OpenAI** 公司发布的领先文到图生成大模型 **DALLE·2**



扩散概率模型的常微分方程离散化

- 针对扩散概率模型半线性等特点，设计等价常微分方程的离散化解析形式
- 发表于机器学习领域旗舰国际会议 **NeurIPS 2022**, 口头报告 (接收率 1.7%)

经典龙格库塔法 $\mathbf{x}_t = \mathbf{x}_s + \int_s^t \left(f(\tau) \mathbf{x}_\tau + \frac{g^2(\tau)}{2\sigma_\tau} \epsilon_\theta(\mathbf{x}_\tau, \tau) \right) d\tau$ 整体黑盒泰勒展开并做差分近似

所提 DPM-Solver $\mathbf{x}_{t_{i-1} \rightarrow t_i} = \frac{\alpha_{t_i}}{\alpha_{t_{i-1}}} \tilde{\mathbf{x}}_{t_{i-1}} - \alpha_{t_i} \sum_{n=0}^{k-1} \hat{\epsilon}_\theta^{(n)}(\hat{\mathbf{x}}_{\lambda_{t_{i-1}}}, \lambda_{t_{i-1}}) \int_{\lambda_{t_{i-1}}}^{\lambda_{t_i}} e^{-\lambda} \frac{(\lambda - \lambda_{t_{i-1}})^n}{n!} d\lambda + \mathcal{O}(h_i^{k+1})$

解析形式 神经网络部分差分近似 解析形式 高阶小量

Lu C, et al. DPM-Solver: A Fast ODE Solver for Diffusion Probabilistic Model Sampling in Around 10 Steps. NeurIPS 2022.

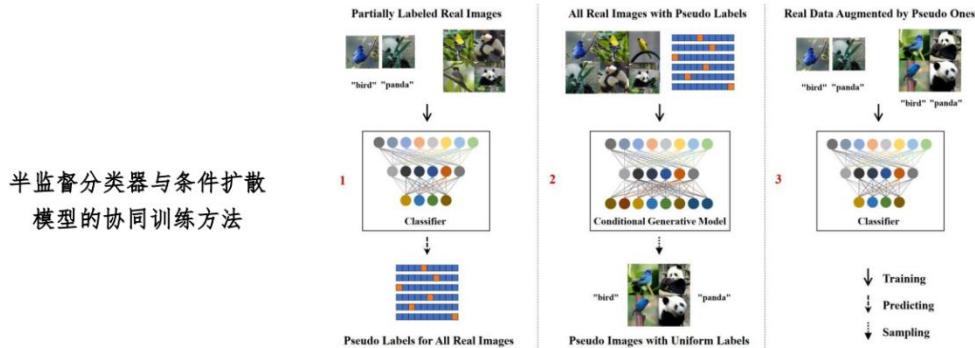
Tips: 显著减小采样次数，加速采样

总结：扩散模型的表示、学习与推断

- 表示
 - 两种等价概率建模方式：隐变量模型 vs. 评分函数估计
 - 网络结构：卷积 vs. Transformer
- 学习
 - 噪声预测：离散化训练 vs. 连续化训练（也有其他等价预测目标）
- 推断
 - 迭代去噪：随机微分方程离散化 vs. 常微分方程离散化
 - 加速推断：针对扩散模型对应微分方程的结构得到解析解

DPT: 半监督扩散概率模型

- 在少量标注下，如何训练条件扩散概率模型并控制生成样本的语义？



You Z, et al. Diffusion Models and Semi-Supervised Learners Benefit Mutually with Few Labels. Arxiv preprint 2023.

Classifier guidance

A Nichol et al., ICML 2021

原始出发点：如何把无条件扩散模型转为条件模型

$$\text{贝叶斯公式} \quad p(x | c) = \frac{p(c | x) p(x)}{p(c)}$$

似然函数	先验（扩散模型）
后验（目标分布）	证据（常数）

Classifier guidance

A Nichol et al., ICML 2021

$$\text{贝叶斯公式} \quad p(x | c) = \frac{p(c | x) p(x)}{p(c)} \Rightarrow \nabla_x \log p(x | c) = \nabla_x \log p(x) + \nabla_x \log p(c | x)$$

温度

$$\tilde{\epsilon}_{\theta, \phi}(x_t, t, c) := \epsilon_{\theta}(x_t) - s \sigma_t \nabla_{x_t} \log p_{\phi}(c | x_t, t)$$

采样方向 预训练扩散模型 预训练分类器

迭代版本：引入不同噪声层次下的指引，并作泰勒展开近似

Tips: 其实不仅是分类任务，扩散模型的采样过程、或者是更新过程都可以根据贝叶斯公式将后验（目标分布）分为两部分：先验和似然函数。然后温度也就是参数系数可以控制两者之间的权重。

Classifier free guidance

Ho and Salimans, Arxiv preprint 2022

分类器指引	$\nabla_x \log p_s(x c) = \nabla_x \log p(x) + s \nabla_x \log p(c x)$
贝叶斯公式 $p(c x) = \frac{p(x c) p(c)}{p(x)}$ $\Rightarrow \nabla_x \log p(c x) = \nabla_x \log p(x c) - \nabla_x \log p(x)$	
带入分类器指引 采样公式	
无分类器指引	$\nabla_x \log p_s(x c) = (1 - s) \nabla_x \log p(x) + s \nabla_x \log p(x c)$ 采样方向 无条件评分函数 条件评分函数

Classifier free guidance

Ho and Salimans, Arxiv preprint 2022

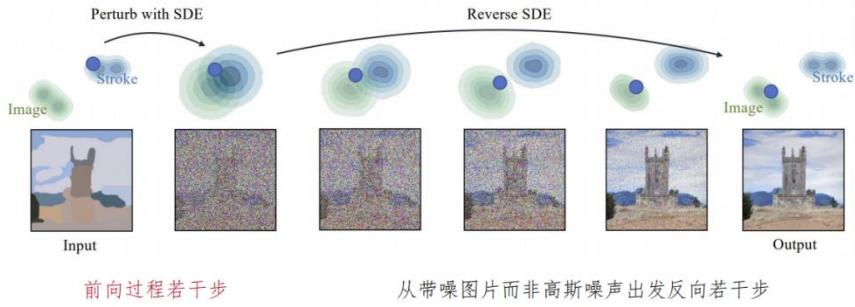
分类器指引	$\nabla_x \log p_s(x c) = \nabla_x \log p(x) + s \nabla_x \log p(c x)$
无分类器指引	
无分类器指引	$\nabla_x \log p_s(x c) = (1 - s) \nabla_x \log p(x) + s \nabla_x \log p(x c)$ 采样方向 无条件评分函数 条件评分函数
<ul style="list-style-type: none"> • 训练两个评分函数模型，共享参数 • 从一个“极端条件模型”采样，即 $s > 1$ • 需要成对数据训练但是参数高效，训练稳定 • 目前对于各类条件生成都非常有效，在文到图生成等任务中应用广泛 	

Tips: 无分类器指引的情况下也可以更新。两个评分函数参数共享，通过是否输入条件和控制 S 来引导生成，使得方法的泛化性能进一步提高。

SDEdit

Meng et al. ICLR 2022

基于预训练模型的图像编辑：零样本采样方法生成目标域的样本



前向过程若干步

从带噪图片而非高斯噪声出发反向若干步

Tips: 可以从带噪的图像或者退化图像开始，而非高斯噪声。

能量指引

Zhao et al, NeurIPS 2022

- 一种加入知识的一般性框架

能量指引

$$\tilde{\epsilon}_{\theta,\phi}(x_t, t, c) := \epsilon_{\theta}(x_t) - s \nabla_{x_t} \epsilon_{\phi}(x_t, t, c)$$

采样方向

预训练扩散模型

预训练能量函数

- 只需能量函数可微
- 分类器指引和无分类器指引是能量指引的特例
- 可以灵活组合各种能量函数

Tips: 这里能量函数指的是先验。

能量指引采样分布

Zhao et al, NeurIPS 2022

- 能量函数定义了如下的概率密度

$$q_{\phi}(x|c) = \frac{1}{Z(\phi)} e^{-\epsilon_{\phi}(x_t, t, c)}, \quad Z(\phi) = \int e^{-\epsilon_{\phi}(x_t, t, c)} dx.$$

- 可以证明，能量指引近似地从如下乘积专家模型中采样

$$p_{\theta,\phi}(x | c) \propto p_{\theta}(x) q_{\phi}(x | c)$$

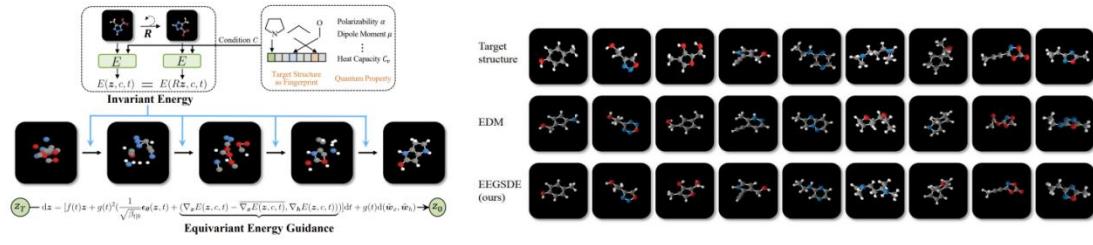
- 贝叶斯公式、RLHF 均为乘积专家模型的特例

Tips: 相当于一个先验专家和特定领域专家合作，概率相乘。

应用二：可控分子生成

Bao & Zhao et al., ICLR 2023

- 几何等变的能量函数指引的扩散模型
- 有效控制官能团、量子性质等合成分子性质，可以同时控制多种属性



Tips: 属性控制

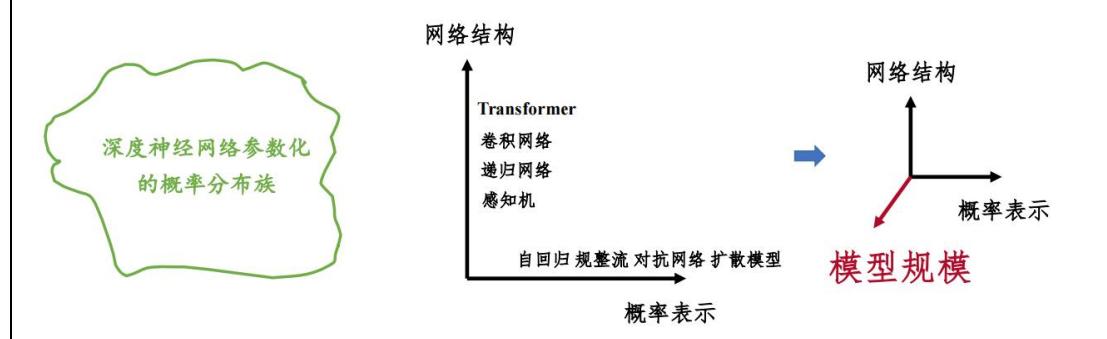
总结：条件扩散模型与 Guidance

- 条件模型
 - 类别半监督：**DPT**
 - 同模态跨域无监督：**SDEdit**、**EGSDE**
- 采样中的指引
 - 指引方式：分类器指引、无分类器指引、能量函数指引
 - 采样分布：条件模型（贝叶斯公式）、乘积专家模型

Tips: 生成模型的先验指引更符合我们特定任务的生成需求，比如超分、去模糊、去噪等等。

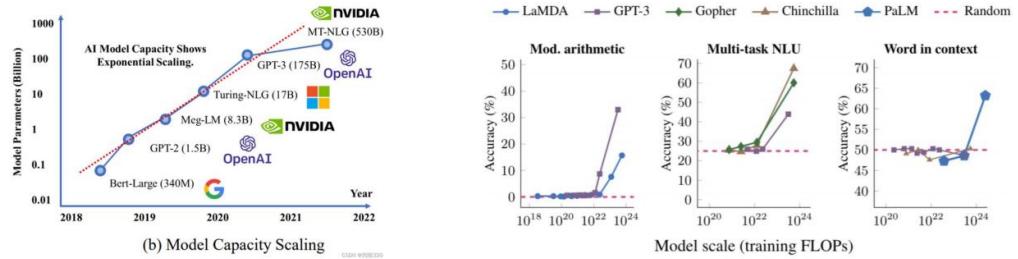
模型规模

- 模型规模是大规模深度生成模型中联合概率分布表示的第三维度



模型规模

- 模型规模是大规模深度生成模型中联合概率分布表示的第三维度



随着模型规模增大，表现显著提升

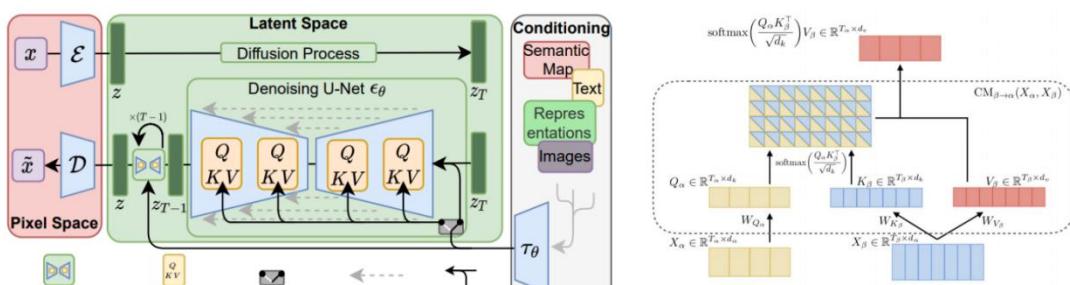
ImageNet：特定域，低噪声，小规模

精准类别标签
约一千万样本



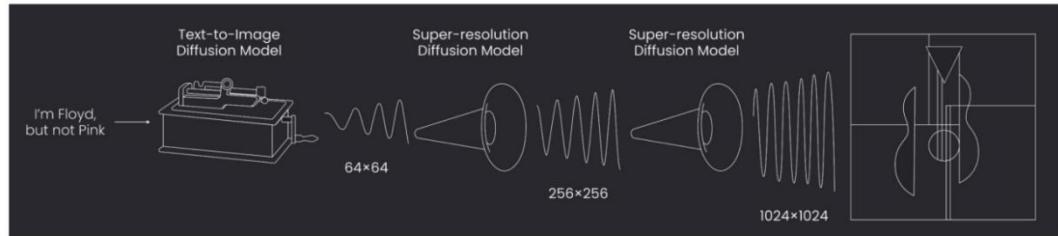
Tips: 就是大模型+大数据+特定域（高质量）=高性能

隐空间扩散模型

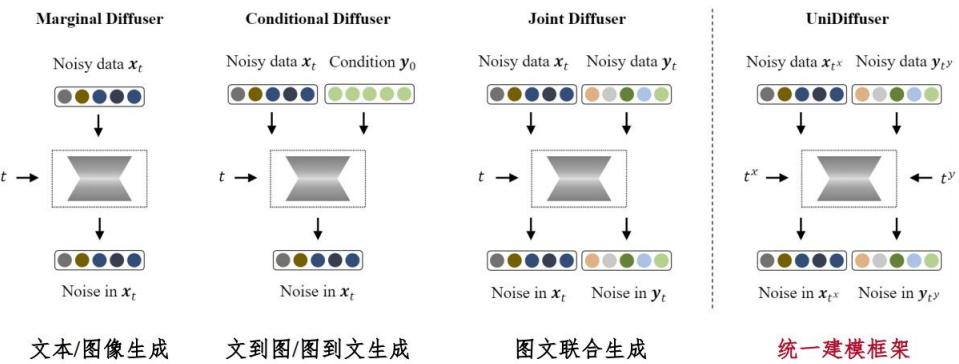


$$\text{训练目标} \quad L_{LDM} := \mathbb{E}_{\mathcal{E}(x), y, \epsilon \sim \mathcal{N}(0, 1), t} \left[\|\epsilon - \epsilon_\theta(z_t, t, \tau_\theta(y))\|_2^2 \right]$$

级联扩散模型



通用多模态扩散模型 UniDiffuser



Bao F et al. One transformer (U-ViT) fits all distributions in multi-modal diffusion. Arxiv 2023.

总结：大规模扩散模型

- 数据
 - 开放域、大规模、高噪声；训练得到强泛化模型
- 训练策略
 - 降维生成再升维：隐空间 vs. 级联
 - 图像、文本的编码器、解码器很重要
 - 任务通用 vs. 任务专用

Tips: 两种训练策略、单任务到多任务通用。

大规模预训练模型赋能 AIGC

- 大规模预训练模型的特点
 - 开放域、强泛化
 - 文本作为主要交互接口
- 下游AIGC的需求与挑战
 - 数据少（利用开放域、强泛化的特点，少样本解决下游任务）
 - 个性化/可控制（加入额外条件控制）

下游任务

- 个性化图像生成
- 图像可控生成与编辑
- 视频可控生成与编辑
- 三维场景生成

Tips: 参考文献

主要工作

• 高效采样算法

- Bao F, Li C, Zhu J, et al. Analytic-DPM: an analytic estimate of the optimal reverse variance in diffusion probabilistic models[J]. **ICLR 2022**.
- Bao F, Li C, Sun J, et al. Estimating the Optimal Covariance with Imperfect Mean in Diffusion Probabilistic Models[J]. **ICML 2022**.
- Lu C, et al. DPM-Solver: A Fast ODE Solver for Diffusion Probabilistic Model Sampling in Around 10 Steps. **NeurIPS 2022**.
- Lu C, et al. DPM-Solver++: Fast Solver for Guided Sampling of Diffusion Probabilistic Models. **Arxiv preprint 2022**.

• 可控采样算法

- Zhao M, Bao F, Li C, Zhu J. EGSDE: Unpaired Image-to-Image Translation via Energy-Guided Stochastic Differential Equations[J]. **NeurIPS 2022**.
- Bao F, Zhao M, Hao Z, Li P, Li C, Zhu J. Equivariant Energy-Guided SDE for Inverse Molecular Design. **ICLR 2023**.
- You Z, et al. Diffusion Models and Semi-Supervised Learners Benefit Mutually with Few Labels. **Arxiv preprint 2023**.

• 多模态大模型

- Bao F et al. All are Worth Words: A ViT Backbone for Diffusion Models. **CVPR 2023**.
- Bao F et al. One transformer (U-ViT) fits all distributions in multi-modal diffusion. **ICML 2023**.



主要工作

• 零样本下游任务

- Xiang C, Bao F, Li C, et al. A Closer Look at Parameter-Efficient Tuning in Diffusion Models[J]. *ArXiv preprint 2023*.
- Wang Z, Lu C, Wang Y, et al. ProlificDreamer: High-Fidelity and Diverse Text-to-3D Generation with Variational Score Distillation[J]. *ArXiv preprint 2023*.
- Zhao M, Wang R, Bao F, et al. ControlVideo: Adding Conditional Control for One Shot Text-to-Video Editing[J]. *ArXiv preprint 2023*.

• 生成模型安全

- Zhao Y, Pang T, Du C, et al. On Evaluating Adversarial Robustness of Large Vision-Language Models[J]. *ArXiv preprint 2023*.

• 生成模型理论

- Zheng C, Wu G, Bao F, et al. Revisiting Discriminative vs. Generative Classifiers: Theory and Implications[J]. *ICML, 2023*.
- Zheng C, Wu G, Li C. Toward Understanding Generative Data Augmentation[J]. *ArXiv preprint 2023*.

• 扩散模型与强化学习

- Lu C, Chen H, Chen J, et al. Contrastive Energy Prediction for Exact Energy-Guided Diffusion Sampling in Offline Reinforcement Learning[J]



开源代码等

• 快速采样算法

- Analytic-DPM: <https://github.com/baofff/Analytic-DPM>
- Analytic-DPM++: <https://github.com/baofff/Extended-Analytic-DPM>
- DPM-Solver(++): <https://github.com/LuChengTHU/dpm-solver>

• 可控生成

- EGSDE: <https://github.com/ML-GSAI/EGSDE>
- EEGSDE: <https://github.com/gracezhao1997/EEGSDE>
- DPT: <https://github.com/ML-GSAI/DPT>

• 多模态大模型

- U-ViT: <https://github.com/baofff/U-ViT>
- Unidiffuer: <https://github.com/thu-ml/unidiffuser>



开源代码等

• 理论

- <https://github.com/ML-GSAI/Revisiting-Dis-vs-Gen-Classifiers>
- <https://github.com/ML-GSAI/Understanding-GDA>

• 安全

- <https://github.com/yunqing-me/attackvlm>

• 项目主页

- Controlvideo: <https://ml.cs.tsinghua.edu.cn/controlvideo/>
- ProlificDreamer: <https://ml.cs.tsinghua.edu.cn/prolificdreamer>
- DPT: <https://github.com/ML-GSAI/DPT-demo>

