学术新星畅谈计算机视觉科研之路：视觉研究已经成熟，跨学科方法成为趋势

AI科技评论 4天前

以下文章来源于微软研究院AI头条 ，作者微软亚洲研究院

微软研究院AI头条
微软亚洲研究院，专注科研23年，盛产黑科技

4月22日，微软亚洲研究院创研论坛 CVPR 2021 论文分享会在线上举行。

如今，计算机视觉在学术界的相关研究已经逐渐进入到一个越来越成熟的阶段，也涌现出了许多学术新秀。因此，本届活动特别邀请了来自卡耐基梅隆大学、哥伦比亚大学、布朗大学、苏黎世联邦理工学院、斯坦福大学、南京理工大学和旷视研究院的7位学术新星，一起畅谈“计算机视觉科研之路”，讨论的内容从研究方向到研究心态，以及未来计算机视觉的探索方向等。

查看此次分享会所有讲者的 PPT 和论文，可访问链接：http://paper.idea.edu.cn/cvpr2021

01:36:
00

1

伟大的研究是如何做出的？

王井东：请举例说明你是如何完成一项工作的。可以从选题、想法、写作等角度详细说明。

朱俊彦：2015-16年，研究 iGAN 的经历令我印象深刻。这项工作是我们用生成模型来解决计算机视觉以及图形学问题的开端。当时的现状是 GAN、VAE、流模型(flow-based model)等生成模型的效果都很差，我们想这些模型目前可能起不到什么作用，但等条件成熟了，其威力也就显现了。

基于这个想法，我们结合传统的计算机视觉算法，例如光流，先让 GAN 生成一个大概的图像，然后再通过光流把修改的效果迁移到原图，实现了用生成模型进行图像编辑的效果。从 iGAN 开始，之后 CycleGAN、Condition GANs 等一批 GAN 的模型也出现了，到最近许多组的 GANs inversion 的优秀工作，都成了研究热点。

这件事给我的启发是：一些技术可能受限于条件不成熟，例如 GPU 的功效，无法发挥其应有的作用。但是，你能看到它的趋势，总有一天，它会变成“可行”的。这时候，如果你提前开展一些工作，就能产生一些影响力。

张祥雨：去年我们有一篇文章被 CVPR 收录为 oral，主要工作是关于密集物体的检测。这项研究我们进行了三年，所以这是一个关于坚持的故事。

该项研究开始于2017年。当时我们的目标检测算法在实际应用中遇到了一个问题：比如在多人通过一个卡口的场景下，由于存在相互遮挡的情况，有些人 AI 模型检测不出来。为了攻克密集场景下的检测问题，我们成立了专门的研发小组。

考虑到 NMS （非极大抑制）后处理是导致密集场景下漏检的主要原因之一，当时我们先尝试的方法是先预测区域里的人数，然后再分别对每个人输出包围框。然而将此方法运用到产品上时，我们发现效果非常差。我们也尝试了许多其他方法，但仍然不起作用。

后来我们意识到，由于密集场景下的训练数据比较少，强行让算法进行学习会遇到严重的样本不均衡问题。此后，我们花了半年多的时间进行数据的采集、标注工作，构建了一万多张包含严重遮挡场景的数据集，即2018年开源的 CrowdHuman 数据集。

有了合适的数据集，算法的准确性（Average Precision，AP）也提高了。但是产品落地却依然困难——FP（误检率）指标相对基线（baseline）大幅度增加。究其原因，我们的解决思路主要停留在替代 NMS 上，我们尝试了很多方法（比如使用微软的Relation Networks），这些方法确实提高了密集场景的检出率，但同时非密集场景的误检率也升高了。如何在提高密集场景准确率的同时，保持稀疏场景的性能不损失，成为一个非常有挑战性的问题。

经过细致地分析，我们发现 NMS 虽然对密集场景不友好，但是在抑制 FP（尤其是稀疏场景的 FP）方面有不可替代的作用。因此，我们转换了研究方向，研究如何避免 NMS 错误地压抑临近的物体。最终在2019年底，经过在多个数据集上反复试验，我们发现了一种非常简单的方法：只需要在检测算法框架中让特征图的每个局部分别预测物体的集合（包含邻域内的所有物体），然后用集合 NMS 取代普通的 NMS，就可以显著提高密集场景的检测性能。这个方法虽然简单，但是却让模型非常鲁棒，同时解决了密集和非密集场景的检测问题。

通过这个小故事，我想告诉大家两件事情：1.学术界和工业界的要求存在差别；2.相信坚持的力量。上面提到的工作，前前后后“熬走了”10多位实习生，很多人试了很多方法，大多数人坚持不住就离职了。只有一位实习生从开始就跟着我们团队，在这项研究中发挥了非常重要的作用。

吴佳俊：我来谈谈在合作过程中学到的一些经验，其实用一个词就可以概括：精益求精。我最早做研究是俊彦带着我一起做的，最近几年和俊彦、舒然也都有合作。在合作的过程中，他们始终思考如何把结果再提高一点。有一次模型生成了一张图片，在我看来，结果已经非常好了，但是俊彦却说，还可以再“完美”一些。不仅是模型，在一些展示工作上，他们对质量也有着很高的要

求，大家可以去他们的论文主页上看看相关的 demo，每一个都做很细致，都力求给读者留下最好的印象。

汤思宇：我今年在 CVPR 上的一个工作，某种程度上是我对未来研究的预期。这个预期在两三年内可能就能验证它是否正确。工作的主要内容是在图像和视频中研究人的行为、动机、姿势等。四年前，大多数研究者的方向都集中于单人和多人的姿势，例如 OpenPose 这一深度学习库。最近，大家开始研究 3D body，例如曲面重建（Surface Reconstruction ）。未来，我们预测人与场景、物体的共同的重建会成为重点。因为，人在现实中经常和环境、物体交互。同时，这也是一个非常困难的研究方向，因为遮挡、人的姿势的多样性等因素的存在。

如何克服这个研究方向上存在的困难？我们认为预设（prior）人体体型非常重要。关于预设，当前用的最多的是 SMPL 模型（simple body model），但 SMPL 在交互方面“差强人意”。因此，基于以上观察，我们提出了一个新观点、新的预设模型，期望模型给出的预测是高准确率和高效的，同时也和 SMPL 模型是兼容的。当然，这项工作也被今年的 CVPR 会议所认可。

对于此，我想表达的是，关于未来的研究方向，你要大胆预测，小心求证，只要不偏离太远，就可能是非常有价值的工作。

孙晨：我来谈谈“如何做有趣的研究”。

我和太太都喜欢研究各种美食，在这个过程中我们发现网上的美食教程是很好的多模态学习数据来源：UP 主们会通过语言把他们做的演示描述出来，这些是我从视频里自监督学习多模态表征的工作（VideoBERT）的训练数据来源。在开展这个工作的过程中，我的另一个乐趣的来源是交流，我对自监督学习的了解很大程度上是通过几年前与做自然语言处理方向的朋友交流了解到的，通过跟有不同学术背景、专长的同事朋友合作可以碰撞出思维的火花，也会学到很多新的知识。最后一个乐趣的来源是探索一些与众不同的方向：我们最近正在开展的工作试图探索视觉信息、多模态模型对于语言理解的影响，这与目前比较流行的 visual grounding、visual question answering 相关，但又有所不同。

宋舒然：同意佳俊刚才的观点，就是与优秀的学者进行合作其实也是快速扩展自己领域见解的方法。我之前和 MIT 同学进行合作的时候也深有感触。更重要的是，这个合作项目对我有很大的教育意义，很大程度上影响了我现在科研的方向。

在项目之前，我的研究方向集中于偏视觉领域，例如物体检测、姿势估计等算法。在开始这个项目之初，我想把之前做过的算法集成到机器人系统中去，让机器人能够抓取任意的物体。但采用传统算法，精度和速度方面都无法达到要求。

后来剖析原因，发现我把问题想简单了，但从另一方面来说，又把问题想复杂了。因为，大多数的时候我们不需要知道物体的类别和物体的姿势，只需要了解到物体的形状就足够了。因此，没必要用到物体检测和姿势估计算法，真正需要的可能是另外一个算法。之后，相似的经历在我进行其他研究时候，例如卫生领域，也会遇到，就是忽略了更核心的问题。

从这个项目学到的经验，直到现在还在不断提醒我：我们要花更多的时间去思考问题的本质，是否可以从新的角度去定义看似经典的问题，从而让其变得更加有意义；我们也不能把自己局限在一个小圈子里，一定要到其他领域去看看，因为这有可能突破研究者固有的认知。

潘金山：我的研究领域主要集中在“去模糊”。针对“去模糊”问题，常用的思路是设计不同的正则化方法。当前已经有很多正则化方法被提出了，那么如何设计出一个更高效的方法？是按照已有的思路？还是另辟蹊径？提出一个想法比较容易，但这个想法是否值得花时间、精力去尝试则需要权衡。例如是否解决了之前的问题；我们的想法发表之后，是否能吸引其他研究者的关注？

在实际解决问题的过程中，我们是这样思考的：当前的正则化方法确实能够解决“去模糊”问题，但是当把其用到不同图像类型的时候，效果并不好，例如针对自然图像的正则化方法并不能有效地解决文本图像去模糊问题。为什么会出现这种问题？

我们从原理上进行了分析，之前的正则化方法的作用是约束解空间，降低问题的病态性，它们大多都是针对特定的图像类型所得出的清晰图像特征的统计规律，没有考虑退化过程。因此，我们就在想能否从退化过程的角度设计出一个正则化方法来克服以上问题？

于是我们对退化过程的原理进行了分析，提出了新的解决方案。由于我们的方法是基于退化过程提出来的，并没有像此前基于统计先验建模的方法那样基于特定图像的统计规律，所以它不依赖于图像类型，可以处理不同场景下的“去模糊”问题。

这给我的启示是：提出有价值的想法，要基于对问题的理解，不要盲目“跟随”之前的研究。另外，也要对现有的方法进行总结，只有总结才能洞察其优缺点，进而针对缺点提出自己的解决方案。

接下来，我也分享一些对写作的感悟。我们做研究，首先要学习别人的论文。这时一定要有自己的判断力，即判断其是否符合高质量论文。如果是高质量论文，那么我们在读懂它的同时，尽量要从作者的角度考虑问题，例如考虑他们是如何构思论文的，我们构思的论文结构和作者构思的论文结构差距在哪？通过不断地比较，找出差距，才能不断进步。

具体到“写”，论文的逻辑结构非常关键。我建议采用一个类似 Coarse-to-fine 的策略，例如我们可以先搭建论文的整体框架结构，然后考虑每一个章节的逻辑，最后再考虑句子与句子之间的承接关系。另外，写每句话的时候都要仔细考虑读者和审稿人的感受，想想他们会不会明白我们所表达的意思，他们会不会有疑问，如果有疑问，我要通过什么措施可以预防。

最后，认真“听导师的话”。强烈建议把导师改过的版本都保存下来，事后仔细对比。通过比较不同版本的差距，我们也能学到很多的写作技巧。慢工出细活，时间也很重要，毕竟高水平的论文大多是靠时间“堆”出来的。

王井东：关于写作，我之前看过一篇文章《你和你的研究》，里面有个观点和金山的观点很契合，都是主张慢工出细活，论文需要打磨。

2

学术新星如何炼成？

> **王井东：**各位都是青年学者，刚脱离学生身份不久。接下来，请各位青年学者分享一下，作为导师，你希望学生应该有什么样的研究状态？

宋舒然：我做学生的时候，更多关注的是问题本身。例如，每天如何把代码顺利运行，算法准确率如何提高等等。其实更重要的是问“为什么”，例如为什么算法可行，为什么参数对结果有巨大影响。

这是一种举一反三的能力。关注问题本身，即回答“如何做”的问题，可能能够发一篇论文，但是如果能够回答“为什么”那么就能够引出一系列的论文，甚至影响一个领域。

因此，我希望同学们能够花一些时间跳出问题本身去探究问题的本质。这也是我从学生到老师身份转换过程中，研究心态上一个很大的变化。

汤思宇：身份转变之前，我是自己思考问题如何解决，但现在我要帮学生进行思考。帮学生思考的过程中，想问题需要更加丰富，角度也要更加多维。例如这个问题解决之后，它的用途在哪里等等。

对学生的期望，我有两点：第一，学生的交流能力和积极性要强，因为交流的越充分，项目就会做得越顺，通过交流我也能更了解细节，从而有针对性地帮助学生；第二，多思考“为什么”，例如经常有学生拿论文中提到的新方法问我，这个方法能否用到那个问题上。其实，他更应该想的是，这个问题的本质是什么，解决问题更好的方式是什么。

王井东：交流真的非常重要。我带过很多学生，在这里分享一个交流过程中经常出现的情况。例如我问学生这个方法为什么有效，一般会回答：结果达到了78分，而对比的方法是75分，所以这个方法有效。其实，这个思维是错误的，这就好比学生的考试成绩，100分的学生不一定比98分更有能力，也可能是运气。因此，只有交流才能看出细节，从而找出问题所在。

张祥雨：我在工业界，对“简单有效”看的比较重，我对学生的要求是对问题要深度思考，但做法要足够简单。只有足够的理解，才能发现更本质的东西，实际应用才能出“奇效”。所以，学生一定尽早要养成独立思考的习惯，形成自己的研究品味，不要把学术研究看成一个军备竞赛。我们做科研的目的是探索认知的边界，是为了促进领域的进步。比“手速”、单纯“盯热门”都是错误思维。

> **王井东：**补充一个观众提问，如何引导和帮助学生问“为什么”？每位同学性格不一，请问各位老师是如何因材施教的？

宋舒然：这两个问题具有高度相关性！有的同学比较“较真”，对于实验结果，总是在思考“为什么”；有的同学则比较“机械”，你让他做什么，他才知道做什么。针对不同的学生，我的方法是：开不同的会，问不同的问题。换句话说，就是有的同学交流少，有的同学进行着重交流。其实在学术界，导师带学生的数量相对少，所以是有可能对某些同学进行“特殊照顾”的。个人认为，这种方法是最开始带学生的时候一个必经的过程。

朱俊彦：一个人的性格很难评判对错、好坏，只有合得来、合不来的区别。如果学生和导师性格合不来，很难强求让学生改变。所以，在面试的时候可以着重进行考察，如果性格真的合不来，我可能就会和学生商量换组。例如，NBA 球队马刺队里的球员都比较无私，有人咨询波波维奇教练如何培养球员无私的性格，教练回答，其实在选人的时候就已经把自私的球员排除掉了。关于如何考察性格，通过聊天就能够洞察，一次不行，就多聊几次。但是，要区别不同性格和不同观点，如果仅仅是学术观点不同，那就可以和学生进行讨论，很激烈的讨论，以及通过做实验进行验证，毕竟老师的观点也不总是正确的。

培养学生最主要的是培养学生的学术能力，而不是论文产量。学生来读博士学位或者硕士学位，我比较在乎的是学生的自身发展。这时候，不能单纯以文章发表来要求学生，而是考察学生在发表了论文之后，学术水平有没有提高，换句话说，主要看：学生下次发表高质量论文比之前发表高质量文章的概率会不会更大一些；就算下次写一个“没那么热门”的研究论文，但是论文本身的质量会不会更高一些，有没有进步。

我日常指导过程中，总是要求同学在投完文章之后，进行经验总结，哪些地方需要改进，哪些优点继续保持。不仅是自己总结，还要和组内同学相互交流，相互吸收经验教训，不要等问题发生到自己身上才进行总结。另外，组内同学也会互相帮助审核代码，如果连自己组内同学都无法使用，那么代码质量肯定不合格，其他研究人员更无法使用。

吴佳俊：一般而言，当学生的时候，你在和其他同学合作时，可能会对自己的任务比较清晰，因为在某种程度上只需要解决他所承担的某一问题就可以了。这其实是典型的学生思维。但如果是一个博士生，你对自己的要求不光是一个工程师，还要是一个好的研究者。这意味着需要思考问题的 How 和 Why，以及弄清楚如何解决这个问题，对领域的贡献是什么。有些个人规划要提前想清楚，毕业之后是去业界担任工程师角色，还是去大学担任教职等等。规划清晰之后，才能在读书期间做到有的放矢。

潘金山：作为导师，肯定希望自己的学生能够做出有影响力的工作，并且得到同行的认可。当我真正成为一名导师的时候，我觉得导师更应该因材施教，和学生共同努力。例如，对于一些低年级的同学，他们刚开始做科研，对于研究方向可能把握不准确，这时候我希望同学们能够及时和我沟通，这样我也能及时纠正他们在科研中犯的错误，帮助他们更好的进行科研。对于高年级的同学，正如前面几位老师谈到的那样，要学会独立思考的能力；多问为什么；自己发现问题等等。

汤思宇：关于怎么让学生多问“为什么”，我采取的方式是多问他们几个为什么。例如，学生告诉我要进行某某领域的研究，那么我就会问他：为什么？研究的意义何在？会有什么样的用处？所以，我会是学生的第一个评审，他一定要有把握说服我才可以继续下一步细节的讨论。

孙晨：对于这个问题，我认为表现的友善一些，给学生更多的亲和力，那么他们自然就会放松下来在交流的时候多问“为什么”。然后，交流的过程中，我会把学生或者实习生看作我的同事，一起讨论，也可以互相质疑对方的思路。

王井东：线上的另外一个问题，如果老师没那么多时间进行交流，那学生该怎么办？

关于这个问题我先回答一下，我会鼓励学生主动 push 我。之前我带过一个学生，他每天下午4点都会找我交流，雷打不动的交流了一个多月之后，他对问题的理解变得非常深刻。后期基本上关于那个问题，他的疑问就非常少了，也节省了很多的时间。

张祥雨：鼓励学生随时交流！只要看到我在座位上，可以随时找我交流。另外，我每天都会安排固定的时间和学生一对一交流。我带了20多个实习生，基本上一两个月就会轮一遍。如果有同学在大组会上不愿意交流、发言，那么我在吃饭的时候，就会和他多谈谈，然后解决一些他的疑惑。

多问“Why”，虽然是个非常好的科研习惯，但在实际操作过程中，学生的知识水平不同，所问问题的深度也不同。至于原因，有可能是阅读、学习积累的还不够。所以，各位同学也不用特别着急，等功底深厚了，自然就可以问出高质量的“为什么”，以及自己解决“为什么”。

王井东：有观众提问，一个博士生毕业后想继续做研究，是应该经过短暂的博士后过程，还是直接去学校担任教职？这两种选择哪一种更有利于职业发展？

朱俊彦：我当时选择博士后，一方面是因为我还没有准备好找教职，还需要一些积累。另一方面，我之前申请 MIT 博士被拒了，所以当时想看看有没有机会在 MIT 做个博士后。

在一年半的 MIT 博士后时间，我学到了很多东西，认识了很多新的朋友。从长远规划来看，博士后经历还是不错的，毕竟找教职是一个短期行为，找到了教职不是可以直接领退休金的，还是要继续做研究，而博士后期间的积累在长远上看对职业发展有帮助。因为博士后会让你在短时间内（1-2年）有机会去新的学校认识新的朋友，更有机会与你欣赏的导师合作，从他们身上学习新的技能和知识。

我觉得，博士后其实是最好的学术时光，不用像学生那样上课和做助教，不用像教授那样教课和找项目资金，合作的对象也更加自由（因为教授需要招生，带学生，还要对学生负责到底），不用像研究员那样担心产品和担心招实习生。每天的所有时间就是做科研和讨论问题。唯一的缺点是工资相对较低。

王井东：孙晨在工业界有过一段经历，现在重回学术界是基于什么考虑？

孙晨：还是选择能够让你开心的事情。我在谷歌待了5年时间，很幸运能够和很多优秀的、不同领域的同事进行合作，打开了我的视野。当然，也经常需要在有趣的研究和有用的研究之间权衡。我的方式是把这两部分工作分开，在一段时间内集中处理一个工作（产品的发布或者科研、投论文）。

我重回学术界，一方面是因为对校园生活、带学生感兴趣，另一方面是基于“前沿研究”的考虑。因为在业界，公司会有自己的需求，虽然可能和学术界有很大程度的重合，但是经过发展，两者的“科技树”可能会朝着不同的方向发展。我感觉学术界在选题上的自由度会更好一些。

3

多元化的计算机视觉发展

王井东：大家对计算机视觉有何预见？

宋舒然：计算机视觉已经进入到了一个越来越成熟的阶段。未来，计算机视觉涉及的不仅仅是视觉，还有可能是听觉、触觉。我目前研究机器人视觉，希望看到与其他学科的交互以及更紧密的合作。

吴佳俊：计算机视觉里面的识别任务已经比较成熟，未来，一方面是和其他领域进行结合，另一方面就是认知部分（cognition）。那什么是认知？它对视觉、感知有何帮助？也是需要回答的问题。对智能的讨论设计到它的科学层面，也涉及工程层面。两者如何更好的结合，也是未来的一个思考方向。

潘金山：我来谈谈底层视觉相关的研究方向。首先，底层视觉和中高层视觉之间有什么关联？高层视觉都是底层特征优化结合得出的，如果高层视觉做的好，会不会反馈给底层视觉？能否借鉴

类脑的工作机制，进行解决底层视觉的问题？另外，随着工业界的需求增多，在画质增强方面，在处理真实复杂场景的时候，应该如何解决？

朱俊彦：最近我的兴趣集中在三个点的关系：模型、人类，以及数据。我读博士的时候一直认为生成模型的作用就是帮助人类创造数据。最近有了新的理解，我认为模型和人类之间可以进行交互，而数据是模型和人类的交互界面 (interface)。换句话说，就是人类如何设计模型，例如人类有 2D、3D 相关的设计软件，那么有没有可能出现一个新的软件是用来设计和创造新的模型的呢？

孙晨：计算机视觉领域的研究者，在跨学科问题研究时，可能会从自己的角度思考问题，而忽视了对方的真正需求，比如我在跟布朗大学做机器人领域的教授聊天时就感觉我们感兴趣的、认为重要的工作并不完全一致。另外，我对认知科学以及它与多模态机器学习的关系也很感兴趣，所以下一步也会涉及这方面的工作。

张祥雨：未来几年我仍然看好大数据+大算力这个方向。在大数据方面，OpenAI 的GPT-3、CLIP 模型都展示出了非常巨大的潜力。尤其这两年自监督研究比较热门，它给出了使用无标签数据的一种方式，大大降低了数据的准备成本；当然这里面挑战也很多，例如目前的自监督框架大都基于“一致性关系”，而这种基于变换鲁棒性的框架可能只是利用了无监督数据蕴含的很小一部分信息。所以，除了一致性，我们还能挖掘哪些关系，是非常重要的研究方向。

另一点是多模态数据的应用。现在大型 CV 预训练模型大都使用了很多图像、视频的数据，但是如果能进一步使用文本和图像的混合数据进行训练模型，是不是会取得更好的效果？OpenAI 的 DALLE 做了一个很好的示范。

第二点，大模型，当你有了大数据之后，就需要表示能力、拟合能力更强的模型能够处理这些数据。但是在把模型 scale up 的过程中，我们很快会发现各种各样的瓶颈，首先是算力和通信，算力可以用分布式系统来解决，但通信还是一个难点。未来，我看好一个方向：分布式友好的模型设计，比如给定一个很大的分布式集群和网络拓扑，如何设计模型才能在这个集群上高效地训练和部署？

最后是大算力。现在进入了芯片时代，各种专用的加速器发展非常快。到底新的加速芯片以什么方式支撑大数据有效地应用？这也是非常重要的课题。

汤思宇：我未来几年比较感兴趣人与人、人与环境的交互。这方面的研究应该会往更加细节的角度进行深入。例如，微软 HoloLens 这一混合现实技术的应用，虽然很棒，但也有很大的拓展空间。