

# Matrix

NO.61

2022年4 - 6月

## 科学智能 (AI4Science) 赋能科学发现的第五范式

3D视频会议系统 VirtualCube:  
相隔万里也如近在咫尺般身临其境

渣画质视频秒变清晰,  
“达芬奇”工具集帮你自动搞定

## 01 焦点

科学智能 (AI4Science) 赋能科学发现的第五范式 2

## 02 前沿求索

3D 视频会议系统 VirtualCube：相隔万里也如近在咫尺般身临其境 4

渣画质视频秒变清晰，“达芬奇”工具集帮你自动搞定 7

NaturalSpeech 模型合成语音在 CMOS 测试中首次达到真人语音水平 10

如何在计算机视觉领域做开创性的前沿研究？ 13

非自回归生成研究最新综述，近 200 篇文献揭示挑战和未来方向 16

### 科研第一线

CVPR 2022 | 一键解锁微软亚洲研究院计算机视觉领域前沿进展 19

ICLR 2022 | 微软亚洲研究院深度学习领域最新研究成果一览 19

ACL 2022 | NLP 领域最新热门研究，你一定不能错过 20

WWW 2022 | 一文解读互联网技术国际顶会最新方向 20

## 03 文化故事

Ada Workshop：从“她力量”到“她行动”，共促计算机领域多元发展 21

科学匠人 | “轮椅上的学霸”，用科技推动无障碍发展 24

科学匠人 | 夏炎：是“超级奶爸”，也是科学研究与技术应用的“摆渡人” 27

对话 2021 微软学者：青年从不缺少描摹未来的潜力 30

计算机科研“打工人”一天的隐藏关卡 33

## 04 观点

计算机领域的女生如何规划职业发展道路？ 35

科研中遇到迷茫困惑如何向前一步？如何在科研中发挥女性优势？ 39

## 05 媒体报道

浙大联合微软亚研院发布视频识别新方法，可对视频逐帧识别且无需数据标记，或可用于手语翻译等 42

## 科学智能 (AI4Science) 赋能科学发现的第五范式



Chris Bishop, 微软技术院士、微软研究院科学智能中心负责人

未来十年，深度学习注定将会给自然科学带来变革性的影响。其结果具有潜在的深远意义，可能会极大地提高我们在差异巨大的空间和时间尺度上对自然现象进行建模和预测的能力。这种能力是否代表着科学发现新范式的曙光？

图灵奖获得者、前微软技术院士 Jim Gary 用“四种范式”<sup>[1]</sup> 描述了科学发现的历史演变。第一范式的起源可以追溯到几千年前，它纯粹是经验性的，基于对自然现象的直接观察。虽然在这些观察中，有许多规律是显而易见的，但没有系统性的方法来捕获或表达这些规律。第二范式以自然理论模型为特征，例如 17 世纪的牛顿运动定律，或 19 世纪的麦克斯韦电动力学方程。这些方程由经验观察，归纳推导得出，可以推广到比直接观察更为广泛的情形。虽然这些方程可以在简单场景下解析求解，但直到 20 世纪有了电子计算机的发展，它们才得以在更广泛的情形下求解，从而产生了基于数值计算的第三范式。21 世纪初，计算再次改变了科学，这一次则是通过收集、存储和处理大量数据的能力，催生了数据密集型科学发现的第四范式。机器学习是第四范式中日益重要的组成部分，它能够对大规模实验科学数据进行建模和分析。这四种范式是相辅相成，并存不悖的。

量子物理学的先驱 Paul Dirac (保罗狄拉克) 在 1929 年说过：“大部分物理学以及整个化学所需的数学理论的基本定律已完全为人们所知，而困难在于这些定律的精确应用会导致方程太过复杂而无法求解。”例如，薛定谔方程在亚原子水平上以极高的精度描述了分子和物质的行为，但高精度的数值解只有在由少量原子组成的非常小的系统中才能得到。如果扩展到更大的系统，则意味着越来越模糊的近似性，这导致人们必须在规模和准确性之间作出权衡。即便如此，量子化学计算已经具有很高的实用价值，成了超级计算机的最主要工作负荷之一。

然而，在过去的一两年里，我们看到了深度学习的一个新用途——兼顾科学发现的速度与准确性的强大工具。这种使用机器

学习的新方式与第四范式数据建模截然不同，因为用于训练神经网络的数据来自科学基本方程的数值解，而非经验观察。我们可以将科学方程的数值解看作自然界的模拟器，以较高的计算成本，对众多我们感兴趣的应用进行计算——例如预测天气、模拟星系碰撞、优化聚变反应堆设计，或计算候选药物分子与目标蛋白的结合自由能。然而，从机器学习的角度来看，模拟过程的中间细节可以被视为训练数据，能够用于深度学习仿真器的训练。此类数据是完全标注的，数据的数量仅取决于计算开销。一旦完成训练，仿真器就可以高效执行新的计算，并大大提升计算速度，有时甚至能够达到几个数量级。

科学发现的“第五范式”代表了机器学习和自然科学领域最激动人心的前沿方向之一。虽然这些模拟器要变得足够快、鲁棒、通用并成为业界主流，还有很长的路要走，但它们对现实世界的潜在影响是显而易见的。例如，仅小分子候选药物的数量估计就多达  $10^{60}$  种，而稳定材料的总数则约为  $10^{180}$  种（大约是已知宇宙中原子数量的平方）。找到更有效的方法来探索这些广阔的空间，将增强我们发现新物质的能力——例如更好的治疗疾病的药物、更好的捕获大气二氧化碳的基质、更好的电池材料、能够为氢经济提供动力的新型燃料电池电极，以及无数的其他应用。

“

AI4Science 是一次深植于微软使命的尝试，这将充分利用我们的人工智能能力来开发新的科学发现工具，从而让我们和科学界的其他同仁能够应对人类面临的最重要的一些挑战。微软研究院成立 30 多年来，始终保持着好奇和探索的传统。我相信，跨越地理和科学领域的 AI4Science 团队，将为这一传统做出非凡的贡献。

——Kevin Scott, 微软执行副总裁兼首席技术官

”

今天，我很高兴地宣布，我将领导微软研究院的一个新的全球团队，其成员来自英国、中国、荷兰等多个国家，专注于将第五范式变为现实。我们的科学智能 (AI4Science)<sup>[2]</sup> 团队由机器学习、计算物理、计算化学、分子生物学、软件和其他学科领域的世界级专家组成，他们共同致力于解决该领域中一些最紧迫的挑战。

以 Graphormer<sup>[3]</sup> 模型为例，它是由我的同事，我们中国团队的负责人、微软杰出首席科学家刘铁岩<sup>[4]</sup> 博士领导建立的。这是一个通用的分子建模模型，对分子有强大的表征能力，对新材料设计与药物发现将有很大帮助。最近，Graphormer 在 Open



Catalyst Challenge (公开催化剂挑战赛) 中夺冠, 这是一项旨在通过 AI 模拟催化剂 - 吸附物反应体系的分子动力学竞赛, 通过密度泛函理论 (DFT) 软件模拟了超过 66 万个催化剂 - 吸附物反应系统 (1.44 亿个结构 - 能量框架)。



另一个项目, 是剑桥的团队与诺华<sup>[5]</sup>合作的生成化学 (Generative Chemistry)<sup>[6]</sup>, 我们携手利用 AI 为科学家赋能, 加快突破性药物的发现和研发。正如诺华人工智能创新实验室全球负责人 Iya Khalil<sup>[7]</sup>最近所指出的, 这项工作不再是科学幻想, 而是科学现实:

“人工智能不仅可以从我们过去的实验中学学习, 而且随着实验室中设计和测试的每一次新迭代, 机器学习算法可以识别新的模式, 并为早期药物发现和开发过程提供指导。希望通过这样的方式, 我们可以增强人类科学家的专业知识, 从而更快地设计出更好的分子。”

该团队利用这一平台已经生成了多个非常有前景的早期分子, 这些分子已被合成, 用于进一步的探索。

除了中国和英国的团队外, 我们位于荷兰的团队也在不断发展壮大, 全球知名机器学习专家 Max Welling<sup>[8]</sup>也加盟其中。今天, 我同样高兴地宣布, 我们在阿姆斯特丹的全新实验室将落户正在建设中的阿姆斯特丹科学园 Matrix One<sup>[9]</sup>。这个专门建造的办公空间紧邻阿姆斯特丹大学和阿姆斯特丹自由大学, 我们将通过联合培养博士生等计划, 与这两所大学保持紧密合作。



阿姆斯特丹科学园 Matrix One

我们怀着自豪和兴奋的心情汇聚成一个跨地域的团队, 追随先驱者的脚步, 为科学发现的下一个范式做出自己的贡献, 并在

此过程中给诸多重要的社会挑战带来有益的影响。如果你也和我们的有着同样的激情和抱负, 希望加入我们的团队, 欢迎查看我们的开放职位<sup>[10]</sup>, 也欢迎你与我们的团队成员取得联系。

## 本文作者:

Chris Bishop, 微软技术院士、微软研究院科学智能中心负责人

## 相关链接:

[1] [https://www.microsoft.com/en-us/research/uploads/prod/2009/10/Fourth\\_Paradigm.pdf](https://www.microsoft.com/en-us/research/uploads/prod/2009/10/Fourth_Paradigm.pdf)

[2] <https://www.microsoft.com/en-us/research/lab/microsoft-research-ai4science/>

[3] <https://www.microsoft.com/en-us/research/project/graphormer/>

[4] <https://www.msra.cn/zh-cn/people/tie-yan-liu>

[5] <https://news.microsoft.com/transform/novartis-empowers-scientists-ai-speed-discovery-development-breakthrough-medicines/>

[6] <https://www.microsoft.com/en-us/research/project/generative-chemistry/>

[7] <https://www.linkedin.com/in/iya-khalil-2898663/>

[8] <https://www.microsoft.com/en-us/research/people/maxwelling/>

[9] <https://www.amsterdamsciencepark.nl/news/matrix-one-a-catalyst-for-science-based-business/>

[10] <https://www.microsoft.com/en-us/research/lab/microsoft-research-ai4science/opportunities/?>



## 3D 视频会议系统 VirtualCube：相隔万里也如近在咫尺般身临其境

常言道：“眼睛是心灵的窗户”，眼神交流所传达的信息也可以进一步提升人们的沟通效果。然而，随着视频聊天、视频会议逐渐成为常态，大家不禁要问，我们有多久没有与同事、朋友、家人确认过眼神了？

而微软亚洲研究院的研究项目 3D 视频会议系统 VirtualCube，可以让在线会议的与会者建立自然的眼神交互，沉浸式的体验就像在同一个房间内面对面交流一样。该技术的相关论文被全球虚拟现实学术会议 IEEE Virtual Reality 2022 接收并获得了大会的最佳论文奖 (Best Paper Award – Journal Papers Track)。

在日常交谈中，相互注视和摆动头部等动作都是自然对话的组成部分，尤其是当我们在面对面交流中变换话题、控制发言或转换交流对象时，都会伴随眼神和肢体动作的交流。然而，当前的视频会议技术却存在着一定的缺陷——由于摄像头和屏幕不在同一高度，如果看向屏幕，眼神往往很不自然，如果看向摄像头则无法关注到其他与会者的反应。因此，视频会议缺少了线下交流的真实感和互动感。而且在实际的工作中，我们还会有各种不同的会议场景，比如多人会议、同排而坐协同工作等情况，对于捕捉与会者的侧方视线和动作来说，现有的视频会议系统就更无能为力了。

如果有一个会议系统，可以让人们即使身处不同的地方，也能像在同一房间里一样交流，视线转动就能与同伴建立起眼神沟通，这是否会给远程办公增添一份沉浸式的真实感呢？

### 利用现有的普通硬件设备搭建的 3D 视频系统



为了解决这些问题，微软亚洲研究院提出了创新的 3D 视频会议系统——VirtualCube，它可以在远程视频会议中建立起真人等大的 3D 形象，无论是正面沟通，还是侧方交流，系统都能够正确捕捉到与会者的眼神、动态，建立起眼神和肢体交流。相关论文被全球虚拟现实学术会议 IEEE Virtual Reality 2022 接收并获得了大会的最佳论文奖 (Best Paper Award – Journal Papers Track)。

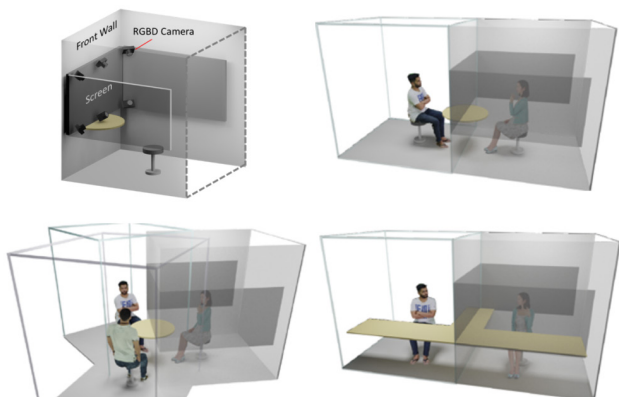


VirtualCube 系统具有三大优势：标准化、简单化，全部使用现有的普通硬件设备。与办公场所中常见的格子间 (Cubicle) 类似，每个 VirtualCube 都提供了一致的物理环境和设备配置：与会者正前方安装有 6 个 Azure Kinect RGBD 摄像头，以捕捉真人的图像和眼神等动作；在与会者的正面和左右两侧还各有一个大尺寸的显示屏，以创造出身临其境的参会感。使用现有的、标准化的硬件能够大大简化用户设备校准的工作量，从而实现 3D 视频系统的快速部署和应用。



6 个 Azure Kinect RGBD 摄像头捕捉人像和眼神等动作

多人、多场景，任意组合。作为在线视频会议的基础构建，VirtualCube 的虚拟会议环境可由多个空间 (Cube) 按照不同的布局组成，以支持不同的会议场景，例如两人的面对面会议、两人并排会议，以及多人的圆桌会议等。



多个空间 (Cube) 可实现任意组合

实时、高质量渲染真人图像。VirtualCube 可以捕捉到参与者的各种细微变化，包括人的皮肤颜色、纹理，面部或衣服上的反射光泽等，并实时渲染生成真人大小的 3D 形象，显示在远程与会者的屏幕中。而且虚拟会议环境的背景也可以根据用户的需求自由选择。



任意变换会议场景，都能身临其境

## V-Cube View 和 V-Cube Assembly 算法双剑合璧，沉浸式会议体验不再是难题

其实业界对 3D 视频会议的研究从未间断过。早在 2000 年，就有人曾提出过与类似混合现实技术有关的畅想。基于这个设想，科研人员一直在探索如何将视频会议以更逼真、更自然的方式呈现，期间也出现了不同的技术路线和解决方案，但都没有达到理想的效果。对此，微软亚洲研究院主管研究员张译中和杨蛟龙表示，过往的研究仍然有很多没有解决的问题：首先，在真实环境下，无论放置怎样的单目摄像设备，即使图像质量再高，与会者也很难形成自然的眼神交流，特别是多人会议的情况；其次，很

多研究针对特定的会议场景进行优化，如两个人面对面的会议或三人的圆桌会议，很难支持不同的会议设置；第三，虽然在影视界我们能够看到一些逼真的虚拟人，但那是需要专业的技术和影视团队长时间打磨和优化才能实现的，仍然需要一定的手工劳动，目前无法进行实时捕捉和实时渲染。

为此，微软亚洲研究院提出了 V-Cube View 和 V-Cube Assembly 两大全新算法，在 VirtualCube 中实现了自动捕捉参与者的手势动作和眼神变化，实时渲染形成高保真图像，让参与者在虚拟会议中体验到真实会议的氛围。

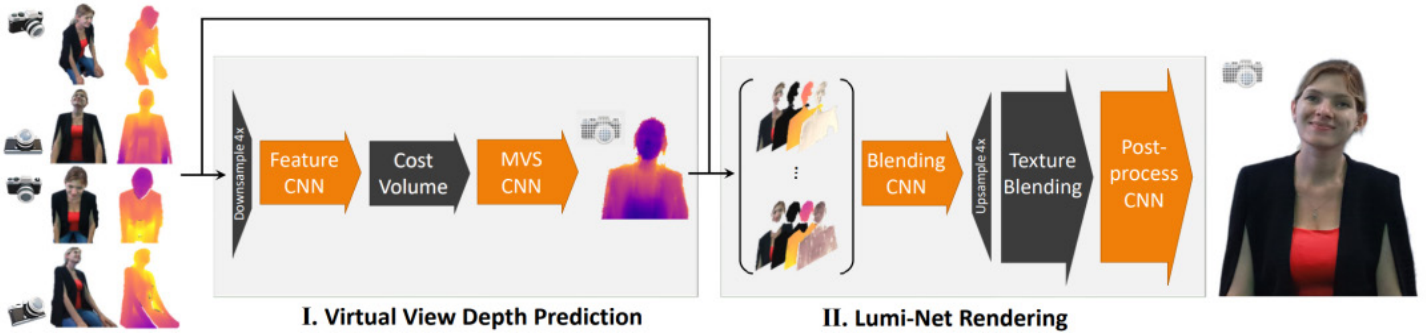


“两个人在交谈且相互注视对方时，对方看到的自己就相当于在自己眼睛的位置放置一个摄像头。但屏幕和摄像头的位置存在高低差，所以当一方注视屏幕中对方的眼睛时，摄像头捕捉到的眼神就会偏离。因此在 VirtualCube 中，我们在与会者正前方的屏幕边缘放置了六个摄像头，通过 V-Cube View 算法合成正确的视点图像，并利用 V-Cube Assembly 确定正确的相对位置，进而给与会者一个沉浸式的会议体验”，张译中介绍道。

基于深度学习的 V-Cube View 算法，通过 VirtualCube 中的六个摄像头的 RGBD 图像作为输入，实时渲染任意目标视点下人的高保真视频。这里的技术挑战是如何同时做到高保真和实时。对此，微软亚洲研究院主管研究员杨蛟龙解释道：“实时渲染高保真的人像，特别是高保真的人脸一直是个具有挑战性的研究课题。传统的三维重建和纹理贴图的做法虽然可以做到实时绘制，却无法重现出真实人脸复杂的材质和在不同视点下外观的变化。为此我们提出了一种新的 Lumi-Net 渲染方法，其核心思想是利用重构的三维几何作为参考来实现一个四维光场的实时渲染，并结合神经网络进行图像增强，从而提高了渲染的质量，特别是人脸区域的高保真度。”

具体而言，V-Cube View 算法分为三步进行。首先，研究员设计了一个神经网络来快速求解目标视点深度图作为人体的几何参考 (geometry proxy)；然后，算法在给定的几何参考下将获取的多视角 RGB 图像 (即光线) 进行融合，实现绘制。在这一步中，研究员受传统的非结构化流明图 (Unstructured Lumigraph) 方法启发，将输入光线与目标像素光线的方向和深度差异作为先验，通过神经网络学习最合适的融合权重。最后，为了进一步提升绘



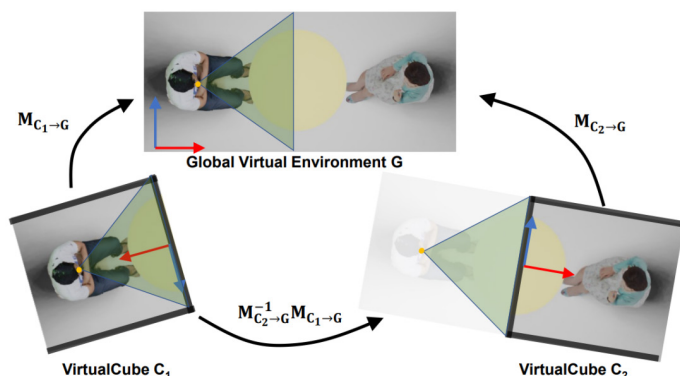


V-Cube View 算法示意图

制质量，研究员们使用了神经网络对上一步的绘制结果进行图像增强。整个算法实现了端到端的训练，并在训练过程中引入了感知损失函数及对抗学习技术，使得算法可以自动学习出最优的神经网络，实现高保真的绘制。而且为了保证绘制的实时性，算法的前两步都在低分辨率图像上执行，这样可以在不损失太多精度的情况下大大降低所需计算量。经过精心设计和优化的 V-Cube View 算法，将实时的三维人物渲染质量提升到了一个新的高度。

另外，为了让 VirtualCube 的使用者拥有和线下交流同样的体验，在将与参会者映射到虚拟环境时，系统还要考虑他们之间的相对位置关系，这时 V-Cube Assembly 算法就发挥了重要的作用。“在整个虚拟会议环境中，V-Cube Assembly 可以被定义为全局坐标系，单个的 VirtualCube 则为局部坐标系。全局坐标系与局部坐标系之间的正确 3D 几何变化，对在视频显示器上正确呈现远程与会者的图像至关重要”，杨蛟龙介绍道。

研究员们首先会在 VirtualCube 中捕捉与会者的 3D 几何体，形成局部坐标系，然后将这些局部坐标系的 3D 几何体数据，投射到全局坐标系，经过 V-Cube Assembly 处理，在全局虚拟会议环境中确定每个 VirtualCube 参与者正确的相对位置，最后再将全局 3D 几何体转换为 VirtualCube 的局部坐标系，映射到 VirtualCube 的屏幕上。

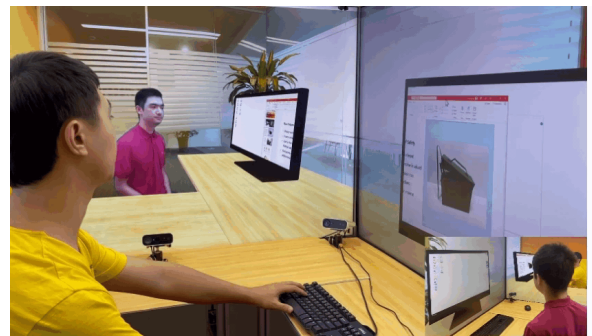


V-Cube Assembly 算法示意图

## 抛砖引玉，畅想未来办公无限可能

VirtualCube 给 3D 视频会议系统提供了一种全新的思路。无论从算法设计、端到端设备部署还是工程调试上，VirtualCube 都证明了利用现有的普通硬件设备就可以实现沉浸式的 3D 视频会议体验。

除了让与会者“共享”同一个物理空间外，研究员们还在探索利用 VirtualCube 系统满足远程办公中的更多协作需求。例如，研究员们展示了这样一种场景：在协同工作时，两位与会者及其电脑桌面都将是视频会议的一部分，因此与会者并排而坐，并且跨屏幕传递自己桌面上的文档和应用程序会让远程协作更加方便。



随着技术的不断精进，未来，大家或许都可以实现身隔万里，却能亲临其境一起办公，自然沟通的体验，而这将极大地提高混合办公的效率。微软亚洲研究院的研究员们也希望 VirtualCube 可以成为一颗探索的种子，给更多研究人员带来启发，在大家共同努力下，找到更好的虚拟空间交互形式，打开未来办公的更多可能。

## 相关链接：

论文链接：

VirtualCube: An Immersive 3D Video Communication System  
<https://arxiv.org/abs/2112.06730>

项目页面：

<https://www.microsoft.com/en-us/research/project/virtualcube/>



## 渣画质视频秒变清晰，“达芬奇”工具集帮你自动搞定

你是否时常“考古”一些老电影、老动漫来回忆旧日时光？你是否也有一些珍贵的录像，带你重温过去的美好？然而，我们已经习惯了高清体验，回头再看曾经的旧影像，画质或许“渣”的让人不忍直视。在这个多媒体内容爆发的时代，人们对视频内容的需求愈加强烈，视频素材的创作、增强及再创作技术也有了大幅提升。尽管利用现有的视频修复工具，视频编辑者也能让视频变得更高清，但其前提条件是需要有超高性能配置的电脑，并付出视频时长几倍甚至几十倍的时间成本，即便如此，结果可能仍不尽如人意。

那么有没有可能在有限的算力，甚至无需联网的情况下，在端设备上实时、高质量地完成视频的高清化、智能插帧呢？微软亚洲研究院的智能视频增强工具集“达芬奇”回答，“我可以！”依托于 400 万高清训练数据和大规模底层视觉预训练方法，“达芬奇”可以实现在端上以较低的计算成本对视频画质进行修复。特别是对于一些实际的生产需求，大规模的底层视觉预训练进一步提升了模型的鲁棒性，使其可以应用于更具挑战性的场景。

1998 年 11 月，微软亚洲研究院在北京成立。时任微软公司 CEO 比尔·盖茨为此特地录制了一段视频，让我们先一起来看看其中的截图。



这段视频对于如今看惯高清视频的我们来说，或许画质有点太渣了。为了解决现有视频增强和修复工具的痛点，充分发挥 AI 技术的优势，微软亚洲研究院多媒体搜索与挖掘组的研究员们将深度学习、机器学习等技术经验，与实际场景和需求结合，推出了一套智能化视频增强工具集——“达芬奇” (DaVinci)，大大降低了用户处理视频素材的门槛，只需一键就可以让视频更清晰、更顺滑。

现在，我们再来看看下面这个被“达芬奇”修复的版本。是不是感觉瞬间拨云见日？



“达·芬奇是文艺复兴时期最杰出的艺术家之一，他将艺术创作和科学相结合，留下了许多不朽之作。因此，我们希望智能视频增强工具集达芬奇可以将 AI 技术应用到多媒体内容处理领域，让视频和图像创作者们更好地挥洒他们的创意，这也是我们将该项目命名为达芬奇的原因，”微软亚洲研究院主管研究员杨欢说。

### “达芬奇”三大技能，将学术概念真正落地应用

据微软亚洲研究院主管研究员傅建龙介绍，学术界很早就开始了对图像视频处理的研究，并在众多方向上做了理论方法的探索，但要将这些停留在概念验证阶段的创新想法真正落地，则需要仔细筛选可行的方向。“经过仔细的研究，我们认为通用图像和视频会议这两大场景中的画质增强具有很大的潜力，其背后主要由图像 / 视频超分辨率、视频插帧、压缩视频超分辨率这三大技术进行支持，有着更好的落地和应用机会，也最有可能让人们体验到当前 AI 技术的优势。”最终，这三项技术被集成到了“达芬奇”工具集中，并通过开源供用户下载使用。

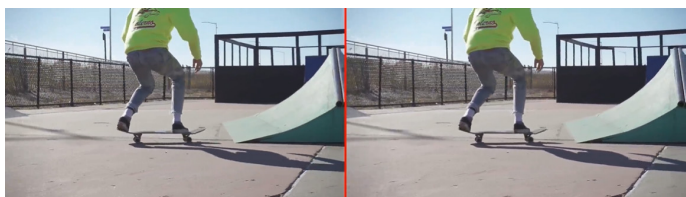
依托于微软亚洲研究院创新的基于 Transformer 的图像 / 视频超分辨率、视频插帧和压缩视频超分辨率技术，“达芬奇”工具集能够实时帮助用户完成不同场景下的视频增强需求。无论是在线还是离线状态，都能生成清晰、连贯的高质量视频，大幅提高视频观感体验。

**视频超分辨率：**是将视频从低分辨率帧序列提升至高分辨率帧序列。对视频来说，最直观的感受就是让我们看到的画面更加高清，视频内容中的细节更加动人，从而满足人们对视频清晰度日益提高的需求，也能更好地适应硬件分辨率的提升。比如将 480P 的老旧视频变成 2K/4K 的高清版本，无论是小屏还是大屏播放，画质都毫无压力。



视频超分辨率结果对比  
(左: 传统 Bicubic 算法, 右: “达芬奇”工具集提供的算法)

**视频插帧:** 是在两个边缘帧之间合成不存在的帧。当前视频的主流帧率是 24 帧 / 秒, 也就是一秒钟播放 24 幅画面。随着视频处理设备以及显示设备性能的提升, 原来的帧率已不能满足大众的需求。尤其是在体育赛事或游戏中, 若能将帧率提升到 60 帧 / 秒甚至 120 帧 / 秒, 那么则可以让画面更丝滑, 也能减少由于帧率不足而导致的眩晕。其实, 这项技术可以应用于许多场景, 包括慢动作视频、帧速率转换等。



视频插帧结果对比  
(左: 传统帧交错算法, 右: “达芬奇”工具集提供的算法)

**压缩视频超分辨率:** 是指从压缩的低分辨率视频帧恢复到高分辨率的视频帧。为了保证视频在互联网上的高传输速率, 或在有限的网络条件下尽可能传输流畅的视频, 互联网或用户设备上的大多数视频都是以压缩格式存储和传输的。然而, 视频压缩会造成质量的损失, 导致终端用户观看时视频好像被打上了马赛克, 特别是在运动较大的场景中, 尤为明显。压缩视频超分辨率就是为了修复这种损失, 使得视频画质更好。



压缩视频超分辨率结果对比  
(左: 传统 Bicubic 算法, 右: “达芬奇”工具集提供的算法)

## 千种设备万般需求, “达芬奇” 创新设计全部搞定

通常, 一项技术在学术论文中所呈现的是其在理想情况下可以达到的上限成果, 而当转化为落地应用的工具时, 该技术就要应对各种下限问题。比如, 我们无法预估用户会使用视频增强工

具来处理哪些类型的视频素材, 它可能是手机记录的儿时影像、DV 拍摄的大好河山, 也可能是胶片电影、MP4 中保存的怀旧金曲 MV, 又或者是朋友分享的压缩后的 4K 电影。因此, 模型需要有足够的鲁棒性才能处理不同的需求。

另外, 用户的部署环境也未可知, 尽管大部分设备可能是手机、笔记本电脑、台式电脑等, 但不同设备的内存、CPU、显卡性能也不尽相同。同时, 研究员们还要考虑计算的功耗问题, 对于手机等移动设备, 电量的消耗、计算处理的时长也都需要精巧的优化和设计。此外, 模型从实验室的服务器迁移到终端设备上, 性能也会有一定的降低。如何确保所有设备体验的一致性, 也是“达芬奇”模型设计的一个重要挑战。

“达芬奇”三大技能实现的背后都是以业界领先的底层视觉预训练技术 (Low-level Vision Pre-training) 作为支撑, 并辅以大量的数据训练。针对模型的鲁棒性, 研究员们一方面利用了可收集到的 400 万公开图像和视频数据, 图像、视频内容涵盖风景、建筑、人物等等, 以及不同压缩率和分辨率的众多场景; 另一方面, 为确保训练的数据量和丰富的数据类型, 研究员们还基于已有的数据, 人工合成了更多含有噪声的退化数据, 从而使整个模型训练能覆盖到更多用户实际的应用场景。

为了应对用户部署环境的多样性需求, 研究员们为模型做了轻量化的设计, 并且对网络结构和模型存储也做了特殊优化。例如, 传统的视频处理方法在处理每一帧图像本身时, 还要考虑整个时间序列, 大大增加了计算量。而微软亚洲研究院的研究员们认为, 视频播放就是物体在时间序列上的运动轨迹, 只有在这个轨迹上的内容才对当前的画面增强有所帮助, 其它区域的内容则关联性较低。

由此, 研究员们提出了基于轨迹 Transformer 的视频超分辨率网络 (TTVSR), 这也是此前微软亚洲研究院图像超分辨率纹理变换 Transformer (TTSR) 的扩展性研究成果。针对超分辨率和插帧的计算, 轨迹感知转换器可以把原先时间  $\times$  视频单帧图像的空间复杂度降低到只是对时间序列的计算, 进而简化整个模型的计算复杂度。原来工业模型处理一分钟的视频需要十分钟甚至一小时, 现在利用“达芬奇”可以实时甚至超实时完成高清视频处理。在压缩视频超分辨率任务上, “达芬奇”可以更好地保留高频视觉细节和指导高频纹理的生成, 降低视频压缩后伪影的影响。

杨欢和傅建龙表示, 相比于图片, 视频的内容更加丰富, 除了空间维度还要考虑时间维度, 对计算的需求更高, 因此, 对于视频处理就需要轨迹感知 Transformer 这样一个设计更为精巧的方法。比如针对视频不同帧之间的连续性和相关性, 如果一个人上一帧出现在这个位置, 下一帧可能向左走一点, 那么我们针对这个人的增强和计算就只需要沿着他运动的轨迹进行计算即可, 不需要对整个视频进行搜索计算。

## 业界指标测试均高于现有方法

通过在业界广泛使用的峰值信噪比 (PSNR) 和结构相似性 (SSIM) 两个指标上进行测试,“达芬奇”工具集的表现都优于现有方法。下表展示了轨迹感知视频超分辨率 Transformer(TTVSR) 在最具挑战性的 REDS4 数据集上的测试结果,其中 PSNR 分别比 BasicVSR 和 IconVSR 提高了 0.70db 和 0.45db。

Method	#Params(M)	FLOPs(T)	PSNR/SSIM
DUF	5.8	2.34	28.63/0.8251
RBPN	12.2	8.51	30.09/0.8590
EDVR	20.6	2.95	31.09/0.8800
MuCAN	13.6	>1.07	30.88/0.8750
BasicVSR	6.3	0.33	31.42/0.8909
IconVSR	8.7	0.51	31.67/0.8948
<b>Ours</b>	<b>6.8</b>	<b>0.61</b>	<b>32.12/0.9021</b>

轨迹感知视频超分辨率转换器 (TTVSR) 在 REDS4 数据集上的测试结果

将上述提到的基于 Transformer 的视频超分辨率相关技术应用到视频插帧以及压缩视频超分辨率上,依然可以得到非常不错的结果。例如,在视频插帧的 Vimeo-90K 数据集上,可以带来 0.36db 的 PSNR 提升;应用于压缩视频超分辨率的 REDS4 数据集上,在 CRF25 的压缩率下,可以带来惊人的 1.04db 的 PSNR 提升。

以上数据集的测试都是基于学术界中特定的退化模型开展的,但考虑到实际的使用场景,用户上传的视频并非高质量的标准素材,也没有一个对比的基准。所以,为了更接近用户的真实需求,微软亚洲研究院的研究员们还设计了一个“不需要标准答案”的视频评估方法 CKDN,即业界非基于参考的质量评估,旨在为业界持续探索视频处理方法提供更多参考依据。(论文链接: <https://arxiv.org/abs/2108.07948>)

## 下载工具集可执行文件,感受清晰、丝滑的视频

目前,微软亚洲研究院已将“达芬奇”工具集的部分可执行文件打包发布,GitHub 上的项目主页也将在近期上线,后续研究员们会在主页上陆续发布、更新更多的视频增强工具。专业开发者可以利用工具集,针对自己的场景进行尝试并与各自的业务深度整合或二次开发,而零技术基础的普通用户也可以通过下载、运行工具集的可执行文件,感受“达芬奇”所带来的清晰、丝滑的视频。

### “达芬奇”工具集下载 1:

图像超分辨率

[https://azsjae.blob.core.windows.net/davinci/release/DaVinci\\_ISR\\_General\\_20220622.zip](https://azsjae.blob.core.windows.net/davinci/release/DaVinci_ISR_General_20220622.zip)

### “达芬奇”工具集下载 2:

视频超分辨率

[https://azsjae.blob.core.windows.net/davinci/release/DaVinci\\_VSR\\_Small\\_Face\\_20220622.zip](https://azsjae.blob.core.windows.net/davinci/release/DaVinci_VSR_Small_Face_20220622.zip)

如在使用工具集时发现任何问题,请发邮件到 [davinci@microsoft.com](mailto:davinci@microsoft.com) 或者在下文即将发布的 GitHub 页面提交 issue 联系我们。

## 相关链接:

论文链接:

Learning Texture Transformer Network for Image Super-Resolution

<https://arxiv.org/abs/2006.04139>

Learning Trajectory-Aware Transformer for Video Super-Resolution

<https://arxiv.org/abs/2204.04216>

Learning Conditional Knowledge Distillation for Degraded-Reference Image Quality Assessment

<https://arxiv.org/abs/2108.07948>

更多链接:

“达芬奇”项目 GitHub 页面 (即将上线,敬请期待!)

<https://github.com/microsoft/davinci>

## 相关阅读

[扫描二维码查看文章](#)

### Transformer 在计算机视觉领域走到哪了?

Transformer 模型在自然语言处理 (NLP) 领域已然成为一个新范式,如今越来越多的研究在尝试将 Transformer 模型强大的建模能力应用到计算机视觉 (CV) 领域。那么未来,Transformer 会不会如同在 NLP 领域的应用一样革新 CV 领域?今后的研究思路又有哪些?微软亚洲研究院多媒体搜索与挖掘组的研究员们基于 Vision Transformer 模型在图像和视频理解领域的最新工作,可能会带给你一些新的理解。





# NaturalSpeech 模型合成语音在 CMOS 测试中，首次达到真人语音水平

AI 合成语音如今已经屡见不鲜，然而在用户听来却不能让人产生与真人对话和阅读般的沉浸感。不过，微软亚洲研究院和微软 Azure 语音团队近日联合推出的全新端到端语音合成模型 NaturalSpeech，在 CMOS 测试中首次达到了真人说话水准。这将近一步提升微软 Azure 中合成语音的水平，让所有合成声音都惟妙惟肖。

文本到语音合成 (Text to Speech, TTS) 是一项根据文本生成可懂且自然的语音的计算机技术。近年来，随着深度学习的发展，TTS 在学术界和工业界取得了快速突破并且被广泛应用。在 TTS 的研究和产品上，微软一直有着深厚的积累。

在研究方面，微软曾创新提出了多个 TTS 模型，包括基于 Transformer 的语音合成 (TransformerTTS)、快速语音合成 (FastSpeech 1/2、LightSpeech)、低资源语音合成 (LRSpeech)、定制化语音合成 (AdaSpeech 1/2/3/4)、歌声合成 (HiFiSinger)、立体声合成 (BinauralGrad)、声码器 (HiFiNet、PriorGrad)、文本分析、说话人脸合成等，而且推出了 TTS 领域最详尽的文献综述。同时，微软亚洲研究院还在多个学术会议上 (如 ISCSLP 2021、IJCAI 2021、ICASSP 2022) 举办了语音合成教程，并在 Blizzard 2021 语音合成比赛中推出了 DelightfulTTS，获得了最好成绩。此外，微软还推出了开源语音研究项目 NeuralSpeech 等。

在产品方面，微软在 Azure 认知服务中提供了强大的语音合成功能，开发人员可以借助其中的 Neural TTS 功能将文本转换为逼真的语音，用于众多场景之中，例如语音助手、有声读物、游戏配音、辅助工具等等。利用 Azure Neural TTS，用户既可以直接选择预置的音色，也可以自己录制上传声音样本自定义音色。目前，Azure Neural TTS 支持超过 120 种语言，包括多语言变体或方言，同时该功能也已整合到了多个微软产品中，并且被业界诸多合作伙伴所采用。为了持续推动技术创新，提高服务质量，微软 Azure 语音团队与微软亚洲研究院密切合作，让 TTS 在不同场景下听起来更多样、更悦耳，也更自然。

近日，微软亚洲研究院和微软 Azure 语音团队研发出了全新的端到端 TTS 模型 NaturalSpeech，该模型在广泛使用的 TTS 数据集 (LJSpeech) 上使用 CMOS (Comparative Mean Opinion Score) 作为指标，首次达到了与自然语音无明显差异的优异成绩。这一创新性的科研成果未来也将集成到微软 Azure TTS 服务中供更多用户使用。

## 四大创新设计，让 NaturalSpeech 超越传统 TTS 系统

NaturalSpeech 是一个完全端到端的文本到语音波形生成系统 (见图 1)，能够弥合合成语音与真人声音之间的质量差距。具体而言，该系统利用变分自编码器 (Variational Auto-Encoder, VAE)，将高维语音 ( $x$ ) 压缩成连续的帧级表达  $z$  (记作后验  $q(z|x)$ )，用于对语音波形  $x$  (记作  $p(x|z)$ ) 的重构。相应的先验 (记作  $p(z|y)$ ) 则从文本序列  $y$  中获取。

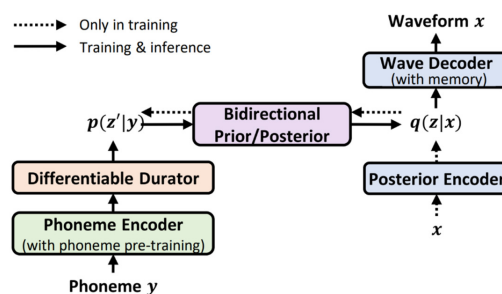


图 1: NaturalSpeech 系统概览

考虑到来自语音的后验比来自文本的先验更加复杂，研究人员们设计了几个模块 (见图 2)，尽可能近似地对后验和先验进行匹配，从而通过  $y \rightarrow p(z|y) \rightarrow p(x|z) \rightarrow x$  实现文本到语音的合成。

- 在音素编码器上利用大规模音素预训练 (phoneme pre-training)，从音素序列中提取更好的表达。
- 利用由时长预测器和上采样层组成的完全可微分的时长模块 (durator)，来改进音素的时长建模。
- 基于流模型 (flow) 的双向先验 / 后验模块 (bidirectional prior/posterior)，可以进一步增强先验  $p(z|y)$  以及降低后验  $q(z|x)$  的复杂性。
- 基于记忆的变分自编码器 (Memory VAE)，可降低重建波形所需的后验复杂性。

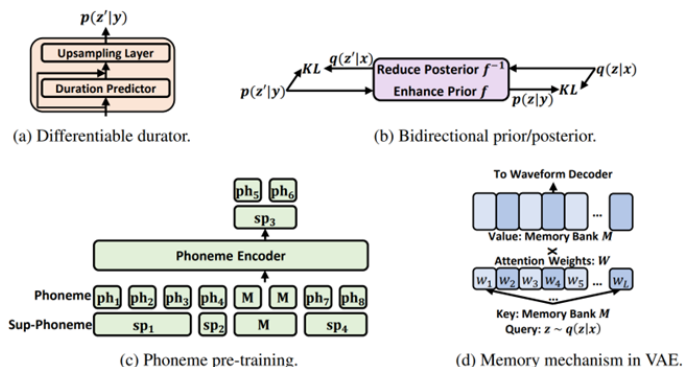


图 2: NaturalSpeech 关键模块

据微软亚洲研究院主管研究员谭旭介绍，与之前的 TTS 系统相比，NaturalSpeech 有以下几大优势：

1. 减少训练和推理的不匹配。先前的级联声学模型 / 声码器系统和显式时长预测都受到了训练推理不匹配的影响。其原因在于声码器使用了真实的梅尔谱以及梅尔谱编码器使用了真实的时长，而推理中使用了相应的预测值。NaturalSpeech 完全端到端文本到波形的生成以及可微时长模块，则能够避免训练推理的不匹配。

2. 缓解了一对多的映射问题。一个文本序列可以对应多个不同的语音表达，例如音高、持续时间、速度、停顿、韵律等方面的变化。以往的研究仅额外预测音高 / 时长，并不能很好地处理一对多的映射问题。NaturalSpeech 中基于记忆的 VAE 和双向先验 / 后验则能降低后验的复杂性并增强先验，有助于缓解一对多的映射问题。

3. 提高表达能力。此前的 TTS 模型往往不足以从音素序列中提取良好的表达以及学习语音中复杂的数据分布。NaturalSpeech 通过大规模音素预训练、带有记忆机制的 VAE、强大的生成模型(如 Flow/VAE/GAN) 可以学习更好的文本表达和语音数据分布。

## 权威评测结果显示：

### NaturalSpeech 合成语音与真人声音不相伯仲

此前的工作通常采用“平均意见分” (Mean Opinion Score, MOS) 来衡量 TTS 质量。在 MOS 评测中，参与者通过听取真人说话录音和 TTS 的合成语音，分别对两种声音的特征进行五分制评分，包括声音质量、发音、语速和清晰度等。但是 MOS 对于区分声音质量的差异不是非常敏感，因为参与者只是对两个系统的每条句子单独打分，没有两两互相比较。而 CMOS (Comparative MOS) 在评测过程中可以对两个系统的句子两两对比并排打分，并且使用七分制来衡量差异，所以对质量差异更加敏感。

因此，在评测 NaturalSpeech 系统和真实录音的质量时，研究员们同时进行了 MOS 和 CMOS 两种测试 (结果如表 1 和

2 所示)。在广泛采用的 LJSpeech 数据集上的实验评估表明，NaturalSpeech 在语句级别与真人录音的对比上实现了 -0.01 CMOS，在 Wilcoxon 符号秩检验中实现了  $p > 0.05$ 。这表明在这一数据集上，NaturalSpeech 首次与真人录音无统计学意义上的显著差异。这个成绩远高于此前在 LJSpeech 数据集上测试的其它 TTS 系统。

Human Recordings	NaturalSpeech	Wilcoxon p-value
4.58 ± 0.13	4.56 ± 0.13	0.7145

表 1: NaturalSpeech 和真人录音之间的 MOS 比较，使用 Wilcoxon 秩和检验 (Wilcoxon rank sum) 来度量 MOS 评估中的 p 值。

Human Recordings	NaturalSpeech	Wilcoxon p-value
0	-0.01	0.6902

表 2: NaturalSpeech 和真人录音之间的 CMOS 比较，使用 Wilcoxon 符号秩检验 (Wilcoxon signed rank test) 来度量 CMOS 评估中的 p 值。

## TTS 发展道阻且长， 需要业界共同打造负责任的 AI

在微软 Azure 认知服务语音首席研发总监赵晟看来，NaturalSpeech 系统首次达到了与真人录音没有显著差异的效果，是 TTS 研究上的一个新的里程碑。从长远角度来讲，虽然借助新模型能够实现更高质量的合成语音，但这并不意味着彻底解决了 TTS 所面临的问题。目前，TTS 仍然存在很多具有挑战性的场景，如充满情感的语音、长篇朗诵、即兴表演的语音等，这些都需要更先进的建模技术来模拟真人语音的表现力和多变性。

随着合成语音质量的不断提升，确保 TTS 能被人们信赖是一个需要攻坚的问题。微软主动采取了一系列措施来预判和降低包括 TTS 在内的人工智能技术所带来的风险。微软致力于依照以人为本的伦理原则推进人工智能的发展，早在 2018 年就发布了“公平、包容、可靠与安全、透明、隐私与保障、负责”6 个负责任的人工智能原则 (Responsible AI Principles)，随后又发布负责任的人工智能标准 (Responsible AI Standards) 将各项原则实施落地，并设置了治理架构确保各团队把各项原则和标准落实到日常工作中。我们正在与全球的研究人员和学术机构合作，继续推进负责任的人工智能的实践和技术。

## Azure AI Neural TTS 的更多功能和声音， 等你来探索

Azure AI Neural TTS 目前共提供 340 多种声音，支持 120 多个语种和方言。此外，Neural TTS 还能帮助企业以多种语言和风格，打造专属的品牌声音。现在，用户可以通过 Neural TTS 试用版来探索更多功能和特色声音。

## 相关链接:

微软 Azure 认知服务 TTS

<https://azure.microsoft.com/en-us/services/cognitive-services/text-to-speech/>

Responsible AI principles from Microsoft

<https://www.microsoft.com/en-us/ai/responsible-ai>

Our approach to responsible AI at Microsoft

<https://www.microsoft.com/en-us/ai/our-approach>

The building blocks of Microsoft's responsible AI program

<https://blogs.microsoft.com/on-the-issues/2021/01/19/microsoft-responsible-ai-program/>

论文链接:

NaturalSpeech: End-to-End Text to Speech Synthesis with Human-Level Quality

<https://arxiv.org/abs/2205.04421>

相关 GitHub 链接:

<https://speechresearch.github.io/naturalspeech/>

<https://speechresearch.github.io/>

<https://github.com/microsoft/neuralspeech>

## 相关阅读

[扫描二维码查看文章](#)

### 语音识别的快速纠错模型 FastCorrect 系列来了!

语音识别支持着许多生活中的常见服务,比如手机端的语音转文字功能、视频网站的字幕自动生成等等。但语音识别模型往往并不完美,需要纠错模型来纠正语音识别中的错误。目前,大部分纠错模型采用了基于注意力机制的自回归模型结构,虽然能够提升语音识别的准确率,但是延迟较高,这也成为了纠错模型在实际应用中的瓶颈。一个直接的做法是利用非自回归模型来提升速度,但是简单利用当前的非自回归模型不能降低错误率。为此,微软亚洲研究院机器学习组与微软 Azure 语音团队合作,推出了 FastCorrect 系列工作,提出了低延迟的纠错模型,相关研究论文已被 NeurIPS 2021 和 EMNLP 2021 收录。



### 应对个性化定制语音合成挑战,微软推出 AdaSpeech 系列研究

提到语音个性化定制,你或许并不陌生。生活中利用明星的声音进行文字播报、语音导航等,都是来自这项重要的文本到语音合成服务。虽其应用广泛,但是现阶段仍面临着很多挑战,例如不同的声学条件对定制效果的影响,需要大量目标说话人的数据及适配参数等等。为此,微软亚洲研究院机器学习组与微软 Azure 语音团队共同合作,推出了 AdaSpeech 1/2/3 系列工作,相关研究论文已分别收录于 ICLR 2021 / ICASSP 2021 / INTERSPEECH 2021 三个顶级学术会议。



### 系统调研 450 篇文献,微软亚洲研究院推出超详尽语音合成综述

近日,微软亚洲研究院的研究员们通过调研了 450 余篇语音合成领域的文献,发表了迄今为止语音合成领域几乎最详尽的综述论文“A Survey on Neural Speech Synthesis”。在文中,研究员们还整理收集了语音合成领域的相关资源如数据集、开源实现、演讲教程等,同时也对语音合成领域未来的研究方向进行了探讨和展望。





## 如何在计算机视觉领域做开创性的前沿研究？

全球人工智能领域最具学术影响力的顶级会议之一 CVPR 于 2022 年 6 月 19-24 日成功举办。为了促进计算机视觉领域前沿研究的分享与交流，推动相关技术进步与发展，近年来，微软亚洲研究院每年都会举办 CVPR 论文分享会。今年的 CVPR 2022 论文分享会于 4 月 23 日圆满落幕。此次分享会由微软智能信息处理西安交通大学实验室主办，微软亚洲研究院承办，中国计算机学会计算机视觉专委会、中国图象图形学会视觉大数据专委会、中国计算机学会计算机女计算机工作者委员会、中国图象图形学会女科技工作者工作委员会协办。

本次活动不仅汇聚了来自清华大学、复旦大学、北京大学、中国科学技术大学、华中科技大学、西安交通大学、东南大学等高校的教授、研究员，还有来自微软亚洲研究院、粤港澳大湾区数字经济研究院等企业研究院的科研人员。20 余位讲者通过网络结构、自监督 + 多模态预训练、多模态应用、3D 视觉、图像和视频生成、检测分割这六大主题，分享了他们在 CVPR 2022 上发表的最新成果。今年，每个主题的论文分享后，还首次设置了主题讨论环节，讲者们就该方向进行了针对性的讨论，探讨整个计算机视觉领域的现状与未来，同时也对观众的问题做出了解答。

此外，CVPR 2022 论文分享会还新增了中场主旨演讲环节，CVPR 2022 程序主席、便利蜂 CTO 和首席科学家华刚博士带来了以“You and Your Research Career with Computer Vision”为题的精彩报告。几位历年视觉顶会最佳论文的斩获者也一同共话“好论文是怎么炼成的”。

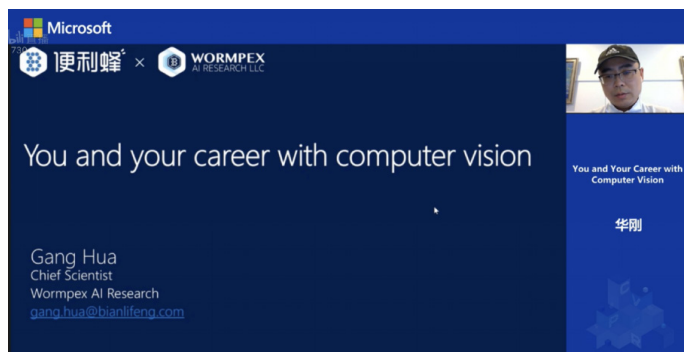
接下来，让我们一起回顾一下 CVPR 2022 论文分享会的精彩内容吧！



### CVPR 2022 主席倾情分享 —— “如何让自己与领域共舞”

华刚博士首先从自己领导便利蜂人工智能研究院的实际科研经验出发，从较高的视角探讨了“计算机视觉的杀手级应用是什么”。华刚博士总结，“它可以数字化物理世界，从社会经济的层面可以给不同的工业、商业领域的信息化做出关键性的贡献。从数字经济、数字生产力的角度来看，它也会对传统工业行业的生产力提高发挥必不可少的关键作用。”

在回顾了过去 20 年的经济发展之后，华刚博士表示，互联网经济是最有活力并产生了巨大财富的一个领域，其背后的高生产效率得益于其自身促成的数字化世界。在这个数字化的世界里，我们可以做各种各样的数字决策，这些数字化决策最后会形成相应的工作任务在物理世界得到执行。未来数字生产力的重点，是将多数传统产业的线下生产、运营的物理过程的信息数字化，形成数字化的映射，并在进一步完成数字化决策后回到物理世界进行相应的工作任务执行。在这个信息数字化的过程中，计算机视觉将发挥巨大且无可替代的作用。



华刚博士为 CVPR 2022 论文分享会作主旨演讲

而后，华刚博士就在计算机视觉的黄金时代里“如何探索自己的计算机视觉研究生涯”进行了分享。他表示，对于一个人的科研生涯而言，第一位是做研究，在研究社群最好的会议和期刊上发表的论文，基本上就代表了你在研究社群中的地位。也就是说，你发表的论文是你在研究领域的形象表征；另一方面则是你对研究社群的服务，例如当审稿人等。科研生涯的评判标准是你对整个领域的影响力，即你给这个研究社群带来了什么改变，包括知识层面和服务层面。而对于“科研生涯”和“职业生涯”的联系，华刚博士认为他们很大程度上是并行的，两者可以互相帮助与促进。

华刚博士接下来还讨论了工业界研究与学术界研究的关系。他指出，两者的主要区别在于经费的来源，在学术界，经费主要来自政府的税收，而在工业界，研究的经费主要来自于公司的利润。那么两者之间的桥梁是什么？华刚博士提到，我们需要建立一条路径，使研究可以产生一些商业上的影响。而对于“工业界是否也可以做基础研究”，华刚博士的回答是肯定的，但他也强调需要建立一个“可以将基础研究转化为商业目标”的架构。做基础研究并不意味着可以很长时间没有结果，需要设置一些里程碑，在完成最终目标的过程中持续产出阶段性的研究成果，将研究项目持续地推进到最终目标。

最后，华刚博士分享了建立自己的主研究社群的重要性，并为广大研究者提出了四点建议：

1. 如果你在工业界做研究，一定要有持久性，它需要你工作之外的额外精力。
2. 无论在公司还是在研究社群，要去理解这个系统是怎么运作的，这样会有助于你在系统内达成自己的目标。
3. 如果在工业界，要思考如何产生商业的影响，你的研究需要设计一条可以产生实际工业影响的路线。
4. 希望大家找到自己的主研究社群，也希望大家把计算机视觉当作自己的主研究社群。

## 视觉顶会最佳论文得主探讨“好论文是如何炼成的”

今年的圆桌论坛邀请到了五位历年计算机视觉顶会最佳论文的斩获者，以“好论文是怎么炼成的？”为主题，分享了他们对于“做好的论文”和“做好的研究”的理解。

巧合的是，这些历年的“最佳论文”研究具有一定的传承性。最佳论文的回顾从 ICCV 99 的 3D 几何研究开始，期间穿插讨论了近两年关于高维空间低秩表达的新书《High-Dimensional Data Analysis with Low-Dimensional Models》，书中所探讨的稀疏低秩表达曾在 10 多年前推动了当时视觉领域最火热的方向，而此后 CVPR 12 的获奖工作，恰好是 3D 几何和低秩表达的完美结合与总结。随着深度学习时代的到来，许多经典工作大量涌现，其中就包括 2015 年的获奖工作 HED，该工作以及前序的 DSN 工作都在尝试探索深度网络的有效训练问题。而这一问题在 2016 年的 ResNet 中得到了几乎完美的解决，CVPR 2017 的最佳论文 DenseNet 则是继 ResNet 之后卷积网络的又一突破。

几位讲者除了分享了自己关于“最佳论文”背后的故事外，还就“如何发现重要的研究问题，找到正确方向”、“一个新领域的开创和发展需要什么条件、准备和时机”、“如何与同行交流，推广

自己的工作”、“如何做好开源”等问题进行了广泛和深刻的探讨。



CVPR 2022 论文分享会现场

其中，屠卓文老师的工作经常很超前，在分享自己“如何发现重要的研究问题，找到正确方向”的经验时，他总结：第一是要善于和导师学习；第二是要手写代码，这会帮助自己对相关内容有更强烈的印象；第三是自己的坚持。这样才会做出最具有代表性与原创性的论文。而马毅老师则特别分享了自己的“盲人摸象”观点——许多不同派别的研究，到最后我们会发现大家都只触摸到了一头大象的一部分。因此，想在自己的领域做出能够青史留名的工作，必须要了解这个领域的发展历史与思想演进。同时，也有许多研究到最后发现只是一个更复杂系统里的局部，是其它领域早已踩过的坑，因此我们不仅要了解自己的领域发生了什么，也要通过历史关注其他领域的问题。针对如何做出真的开创性的工作，他提出：第一，设计一个远大的正确的目标方向；第二，了解历史，知道与这件事情相关的来龙去脉是什么；第三，深耕相关的领域，借深度学习的话就是做学问也要越深越好、越宽越好——越深就是要去了解历史纵深，看到思想的传承，越宽就是要去了解相关领域，由点及面你才能看到大象的整体，也会有自己更独到的思想，而这可能是你一生中最有成就感的事情。

## 六大热门主题，共话领域现状与未来

在本次分享会上，20 多位讲者从六大主题进行了论文分享和研讨。这六大主题分别是：网络结构、自监督 + 多模态预训练、多模态应用、3D 视觉、图像和视频生成，以及检测分割。它们也正是计算机视觉领域近年来最活跃的几个方向。

在网络结构主题中，四位讲者分别带来了 ACMix、SwinV2、CSwin 以及 ReplKNet 的分享，其中两篇论文关于视觉 Transformer，一篇关于卷积网络的“第二次增长”，另一篇则是 CNN 和 Transformer 的融合。值得一提的是，该主题的研讨很有幸汇聚了三位视觉顶级会议的最佳论文奖得主，他们在清华大学鲁继文教授的主持下，就“大模型和小模型在性质上的区别”、“高校如何参与大模型的研究”等问题进行了讨论。



在自监督 + 多模态预训练主题中，三位讲者分别带来了 SimMIM、BEVT 和 HD-VILLA 论文的分析，其中前两篇论文聚焦最近火热的掩码图像建模方法在图像和视频上的应用，最后一篇论文则关注多模态预训练方面的进展。随后讲者和嘉宾们针对“掩码图像建模 (MIM) 预训练能否像掩码语言建模 (MLM) 那样从大数据中获益”等问题分享了自己的观点。

在多模态应用主题中，三位讲者的工作主要涉及将语言和目标对齐、语言和音频对齐，以及语言和多模态知识对齐，完美地覆盖了目前多模态领域的核心概念——对齐，三个工作也在各层面做了非常丰富的探索。在讨论环节，嘉宾们在中科院黄岩老师的主持下，就“多模态未来将产生哪些值得期待的应用场景”等问题进行了分享和讨论。



六大主题论文分享和研讨

在 3D 视觉主题中，三位讲者的论文分别关于动作捕捉、移动端三维重建，以及基于神经辐射场的三维图像生成。前两篇论文是贴合实际应用的研究，而后者则与最近几年火热的神经辐射场相关。讲者和嘉宾们在西安交通大学兰旭光老师的主持下针对“3D 视觉的杀手级应用是什么”以及“NerF 的现状与未来”等问题进行了讨论，并各自分享了自己的观点。

在图像和视频生成主题中，三位讲者分别分享了降噪扩散模型方面的最新成果 VQ-Diffusion，以及基于 Transformer 的图像修复和超分辨率方法。其中降噪扩散模型 (DDPM) 已有取代 GAN 成为图像生成的未来趋势，而 Transformer 则继续在底层视觉问题中攻城掠地。随后，讲者和嘉宾们还针对“现在图像生成有哪些成功或者潜在的应用”、“扩散模型的未来”等问题进行了讨论。

最后一个主题是检测分割，这也是一个广受关注的领域，分享的四个工作其中之二涉及 Transformer 结构，特别是其解码器在检测分割领域的应用；而另外两个则是基于传统检测分割头部网络的工作。其中，来自粤港澳大湾区数字经济研究院的张磊老师克服美国深夜三点的时差，坚持现场演讲。他分享的 DN-DETR 是这个方向的一个重要工作，也是他最近在 COCO 物体检测中取得新纪录的 DINO 方法的重要前序工作。华中科技大学王兴刚老师则分享了他们团队将掩码图像建模应用于 ViT 检测的最新工作，

这也与第二个主题中备受瞩目的掩码图像建模模型相呼应。最后，在主题讨论环节，活动邀请到了中科院张兆翔老师加入讨论，5 位讲者和嘉宾们针对“在检测分割这样的下游任务中，网络结构是否会被 Transformer 一统天下，还是仍然会保持目前一个相对“百花齐放”的局面”等问题进行了精彩讨论和观点分享。

CVPR 2022 分享会聚焦了近期计算机视觉的核心方向与重要研究，为与会者们带来了深度与广度并存的分享，而讲者们也通过彼此的思想交锋获得了研究认知的刷新。

## 相关阅读

[扫描二维码查看文章](#)

### 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

微软亚洲研究院创研论坛 CVPR 2021 论文分享会在线上举行。来自国内外计算机视觉领域的 21 位优秀代表与观众零距离交流，分享了他们在 CVPR 2021 发表的最新研究成果。



### Swin Transformer 迎来 30 亿参数的 v2.0，我们应该拥抱视觉大模型吗？

2021 年，获得 ICCV 最佳论文奖的 Swin Transformer，通过在广泛的视觉问题上证明 Transformer 架构的有效性，加速了计算机视觉领域基本模型架构的变革。2021 年末，微软亚洲研究院的研究员们又进一步提出了 Swin Transformer v2.0 版本，新版本训练了迄今为止最大的稠密视觉模型，并在多个主流视觉任务上大大刷新了记录，相关论文也已被 CVPR 2022 接收。研究员们希望借助 Swin Transformer v2.0 展现视觉大模型的“强悍”能力，呼吁整个领域加大对视觉大模型的投入，并为之提供相应的训练“配方”，从而为视觉领域的科研人员做进一步探索提供便利。





# 非自回归生成研究最新综述，近 200 篇文献揭示挑战和未来方向

近年来，由于并行的快速推理能力，非自回归生成在自然语言处理、语音处理等领域展示出了其特有的优势，并日益成为生成模型的研究热点。为了促进非自回归生成模型的发展，微软亚洲研究院与苏州大学的研究员们共同撰写了综述论文“A Survey on Non-Autoregressive Generation for Neural Machine Translation and Beyond”，回顾了非自回归生成在神经机器翻译以及其他任务中的发展，并对非自回归生成的未来提出了展望。

在如机器翻译、对话生成、语音合成等自然语言、语音等生成任务中，自回归 (auto-regressive, AR) 生成是一种最常采用的生成方法。简单来说，AR 生成指的是用迭代循环的方式来依次生成一句语音或文本。比如，为了生成一句长度为 5 的句子，AR 生成首先会生成第一个词语，然后基于第一个词语生成第二个词语，再基于前二个词语生成第三个词语，以此类推。由于每次新的词语生成都依赖于之前生成的词语，因此自回归的生成方式能够保证生成的准确度。

但显然，这样循环的生成方式效率非常低，尤其是对生成长句子来说则更为明显。为了加速生成过程，非自回归 (non-autoregressive, NAR) 生成被提出，通过一次性并行地生成句子中所有词语的方式，NAR 生成方法极大地提升了生成效率。然而，NAR 生成的准确率并没有得到保证，其性能与自回归生成相比仍有一定差距。因此，如何平衡好 AR 生成与 NAR 生成的优劣，是当下生成任务的研究重点。

## 综述概览

NAR 生成在神经机器翻译 (neural machine translation, NMT) 中首次被提出，此后 NAR 生成便引起了机器学习和自然语言处理领域的广泛关注。如前文所述，虽然 NAR 生成可以显著提升机器翻译的推理生成速度，但与 AR 生成相比，其加速是在牺牲翻译准确性的代价上实现的。近年来，为了弥补 NAR 生成和 AR 生成之间的准确性差距，许多新的模型和算法陆续被提出。

为了促进 NAR 生成模型的发展，微软亚洲研究院与苏州大学的研究员们共同撰写了综述论文“A Survey on Non-Autoregressive Generation for Neural Machine Translation and Beyond”。

在文章中，研究员们给出了一个系统、全面的综述。首先，研究员们从不同方面比较和讨论了各种非自回归翻译 (non-autoregressive translation, NAT) 模型，具体来说就是对 NAT 的工作进行了几组不同的分类，包括数据操作 (data manipulation)、建模方法 (modeling methods)、训练准则 (training criteria)、解码算法 (decoding ways) 以及利用预

训练模型 (benefit from pre-training)。此外，研究员们还简要总结回顾了 NAR 生成在机器翻译之外的其他应用，例如对话生成、文本摘要、语法纠错、语义解析、语音合成和自动语音识别等等。最后，研究员们讨论了 NAR 未来值得继续探索的潜在方向，包括减少对知识蒸馏 (knowledge distillation, KD) 的依赖性、动态解码长度预测、NAR 生成的预训练，以及更广泛的应用。图 1 展示了本篇综述论文的整体结构。

研究员们希望该综述文章可以帮助研究人员更好地了解 NAR 生成的最新进展，启发更先进的 NAR 模型和算法的设计，使行业从业者能够根据其所在领域选择合适的解决方案。

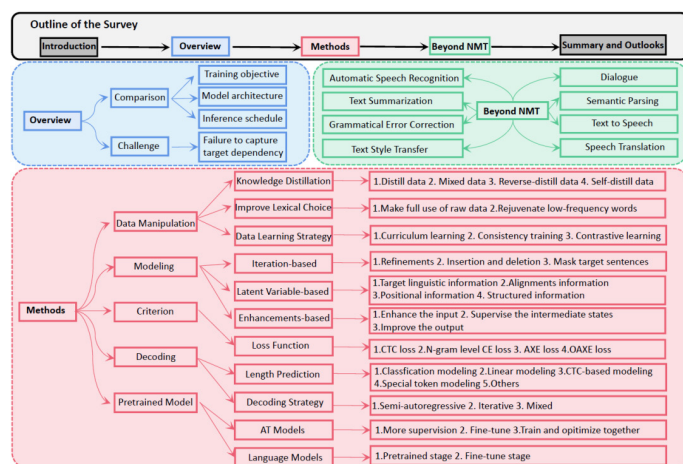


图 1: 非自回归 (NAR) 生成研究综述概览架构图

## NAT 模型面临的主要挑战与解决方案

传统的自回归机器翻译 (autoregressive translation, AT) 模型由编码器和解码器构成，编码器对源语句进行编码后输至解码器，然后解码器根据源语句和上一步预测的目标端语言单词来预测下一个单词，这种逐字的生成方式限制了 AT 模型的解码速度。而为了实现在训练和推理时并行的解码方式，NAT 仅仅依赖源语句信息来生成所有目标单词，摒弃了目标端单词之间的条件依赖。这种方式极大地加速了模型的解码，但也增加了 NAR 模型的训练

难度，造成模型“难以建模目标语言单词之间的条件信息”。

针对该挑战，现有的工作提出了多种解决方案。综述文章对现有工作进行了分类，从数据、模型、损失函数、解码算法、利用预训练模型五个角度对相关方法进行了介绍和比较。其中，数据、模型和损失函数是自回归文本生成模型的三个基本组成部分，这方面的工作旨在研究上述三个方面的传统方法在 NAR 模型上的不足，并进行相应的改进；解码算法和利用预训练模型则是非自回归文本生成模型中区别于 AR 生成的特殊模块，包括目标语句长度预测、非自回归预训练等，这方面的工作旨在设计合理、有效的算法来最大化地提升 NAR 生成模型的效果。这几方面的联系如图 2 所示。

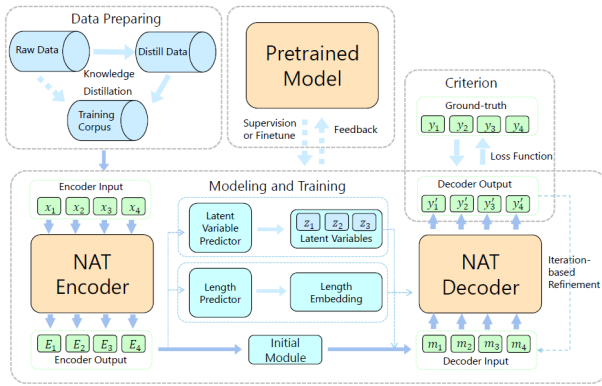


图 2: 非自回归机器翻译模型的主要框架。

其中涉及数据处理、模型改进、训练准则、解码方式、预训练模型的利用等。

具体来说，上述五个方面的改进如下：

1. 数据层面进行的改进，包括利用知识蒸馏来生成数据、设计数据学习算法等。利用预训练 NAR 模型，基于知识蒸馏的方法将训练集中的源语句进行翻译，并将源语句和翻译结果作为 NAR 模型的训练集。这种方式可以减少训练数据的多样性，减轻 NAR 模型的训练难度。需要注意的是，数据层面的方法是通用的方法，例如，基于知识蒸馏的方法被广泛应用在文中介绍的大部分 NAR 生成模型中。

2. 模型层面进行的改进，包括设计迭代式模型、基于隐变量的模型以及增强解码器模型结构等。其中，迭代式模型将原始一次解码的 NAR 模型扩展为多次迭代解码的模型，这样在进行每轮迭代时，上一轮迭代的结果可以作为目标语言端的依赖信息，将一次解码的难度分摊到多次迭代中，从而提升 NAR 模型的效果。与一次解码的 NAR 模型相比，迭代式的模型翻译效果更好，但也牺牲了一部分翻译速度，是属于 AR 模型和 NAR 模型的中间态。

3. 损失函数层面进行的改进，主要针对传统交叉熵损失函数的问题，提出一系列改进方法，包括基于 CTC、n-gram、以及引入顺序信息的损失函数。其中，由于 n-gram 的方法针对传统的

交叉熵损失函数只能提供单词级别的监督信息而无法提供全局信息，研究员们提出了优化预测和目标之间 Bag of N-gram 差异的损失函数，以补充交叉熵损失函数中缺失的全局信息，以更好地对 NAR 模型进行优化。

4. 解码算法层面进行的改进，包括对 NAR 模型的长度预测模块进行改进，以及对传统解码算法的改进。由于 NAR 模型无法像 AR 模型一样隐式地在解码过程中决定目标语句的长度，因此需要在解码过程开始前就对目标语句的长度进行显式预测。这个步骤十分重要，因为目标语句的长度是否匹配直接影响模型最终的翻译效果。因此，类似自回归解码中的 Beam Search，有模型提出了提升长度预测准确率的方法，如多个长度并行解码等。这些方法也被广泛应用在 NAR 模型中。

5. 利用预训练模型的方法，包括利用自回归教师翻译模型的方法，和利用单语大规模预训练语言模型的方法。其中，由于 NAR 模型和 AR 模型结构相似，并且 AR 模型的翻译准确度更高，因此很多方法提出利用预训练的 AR 模型来额外监督 NAR 模型的训练，包括在隐变量层面引入额外监督信息，和基于课程学习的迁移学习方法等。

研究员们将文中讨论的相关论文按照类别列在了表 1 中，供大家查阅。（见第 18 页）

## 关于探索 NAR 的开放性问题 and 未来方向

NAR 除了在 NMT 中的应用之外，还在其它许多的任务中也得到了扩展应用，其中包括文本生成任务，如文本补全、摘要生成、语法纠正、对话、风格变化，语义解析任务、文本语音转化任务、语音翻译任务等等。研究员们在综述文章中给出了一些具体实例的介绍，同时也给出了这些相关工作的实现与资源列表。

为了促进未来 NAR 的发展，研究员们对当前 NAR 产生的问题进行了总结，并对未来可能的方向进行了展望，具体包括：(1) 如何能够摆脱当下 NAR 严重依赖 AR 进行知识蒸馏的技术方案；(2) 如何能够降低迭代式 NAR 模型的计算复杂度以更好地关注纯 NAR 模型；(3) 动态的预测目标端文本的生成长度值得深入探索；(4) 如何像 AR 模型一般将 NAR 模型扩展到多语言多任务的环境中是需要进一步关注的；(5) 如何对 NAR 模型进行更好的预训练。以上这些都是具有研究前景的研究问题。

希望通过本篇综述，在不同领域进行生成任务研究的学者们能够对 NAR 生成有更全面的认识，并且激发创造更加先进的 NAR 模型，以促进 NAR 未来的发展，影响更广阔的生成场景。

Category	Sub-category	Method	Description	Publication
Data Manipulations	Improving KD	MD [49] RDP [20] LRF [48] SDMRT [21]	Add monolingual data, enrich distillation corpus Raw data prior training, improve lexical choice Add reverse-distill data, rejuvenate low-frequency word Self-distillation mixup, pre-rerank and fine-tune training	ACL 2020 ICLR 2020 ACL 2021 ARXIV 2021
	Learning Strategies	Glat [37] latent-Glat [50] PMG [51] MvCR-NAT [52]	Glacing, learn from fragments to whole sentence Introduce glacing strategy to discrete latent variables Multi-granularity, from words, phrases, sentences gradually Consistency-based, masked token and model level consistency	ACL 2021 ACL 2022 ACL# 2021 ARXIV 2021
Modeling	Iteration-based methods	RefineNAT [29] Insertion Transformer [53] Levenshtein [54] CMLM [18] Disco [34] JM-NAT [55] Imputer [36] Rewrite-NAT [26] CMLMC [56]	Denoising autoencoders, iterative refinement Insert tokens each iteration, like balanced binary trees Insert and delete tokens during each iteration Masked language model trained with uniform mask strategy More visible subsets to predict masked tokens Jointly masked strategy, N-gram level masking in decoder Combine conditional masking with CTC Reviewer and locator, locate the error and rewrite CMLM with reveal-position and correction function	EMNLP 2018 ICML 2019 NeurIPS 2019 EMNLP 2019 ICML 2019 ACL 2019 EMNLP 2020 EMNLP 2021 ICLR 2022
	Latent variable-based methods	FT-NAT [16] FlowSeq [57] PNAT [58] SynST [59] LaNAT [60] Reorder-NAT [61] AlignNAT [62] CNAT [63] SNAT [64] Fully-NAT [38]	Fertility predictor, determine the latent alignments Generative flow, a powerful mathematical framework Positional predictor, model the position of target tokens Parse decoder, autoregressively predict a chunked parse tree Delta Posterior, continuous latent variables Reordered the source sentence into a pseudo-translation Aligner module, alignment decomposition strategy Categorical codes, without external syntactic parser Incorporate the explicit syntactic and semantic structures Several tricks to improve the Fully NAT	ICLR 2018 EMNLP 2019 ARXIV 2019 ACL 2019 AAAI 2020 AAAI 2021 EMNLP 2021 NAACL 2021 EACL 2021 ACL# 2021
	Other enhancements-based methods	ENAT [23] NAT-REG [22] LAVA NAT [65] CCAN [66] DSLp [67] DAD [68]	Phrase-table lookup, embedding mapping Similarity and reconstruction regularization Vocabulary attention, reorder prediction labels of a word Context-aware cross-attention, local and global contexts Deep supervision, additional layer-wise predictions Decoder Input Transformation, backward dependency modeling	AAAI 2019 AAAI 2019 ARXIV 2020 COLING 2020 ARXIV 2021 ARXIV 2022
Criterion	Loss function	CTC [36] BoN-Joint [69] AXE-NAT [70] EISL [71] OAXE-NAT [39]	Compute and stores partial log-probability N-gram level loss, minimize the Bag-of-Ngrams difference Aligned cross-entropy, a differentiable dynamic program Compute the n-gram matching different, more robust Order-agnostic cross-entrop, hungarian algorithm	EMNLP 2020 AAAI 2020 ICML 2020 ARXIV 2021 ICML 2021
Decoding	Semi-autoregressive decoding	Semi-NAT [17] RecoverSAT [72]	Generate multi-tokens at one decoding step Recover segment, recover mistakes of multi-tokens	EMNLP 2018 ACL 2020
	Iterative decoding	Mask predicted [18] Easy-first [34] Insert [53] Insert and delete [34]	Mask several tokens with the lowest probability scores Easy-first order, update each position given easier tokens Insert tokens during each iteration Insert or delete tokens during each iteration	EMNLP 2019 ICML 2019 ICML 2019 NeurIPS 2019
	Mixed decoding	Unify [35] Diformer [73]	Unified approach, conditional permutation language modeling Directional transformer, directional embedding and self-attention	COLING 2020 ARXIV 2021
Benefiting from Pre-trained Models	AT models	Imitate-NAT [74] Hint-NART [75] ENGINE [76] EM+ODD [77] FCL-NAT [24] MULTI-TASK NAT [78] TCT-NAT [79]	Imitation learning framework with imitate module Hints from the hidden state, constrain attention distributions Energy-based inference, minimize the AT model's energy Unified framework, dynamically optimize AT and NAT Curriculum learning from better-trained state of AT model Shared encoder, dynamically mix two training loss Task-level curriculum learning, from AT to SAT, then to NAT	ACL 2019 EMNLP 2019 ACL 2020 ICML 2020 AAAI 2020 NAACL 2021 IJCAI 2021
	Pre-trained language models	AB-Net [80] Bert+CRF-NAT [81] CeMAT [40]	Take two different BERT models as the encoder and decoder Employ bert as a backbone, add a CRF Layer Aligned code-switching and masking, dynamic dual-masking	NeurIPS 2020 EACL 2021 ARXIV 2022

表 1: 针对 NAT 模型 5 个方面的研究总结以及具体的相关工作

## 相关链接:

论文链接:

A Survey on Non-Autoregressive Generation for Neural Machine Translation and Beyond

<https://arxiv.org/pdf/2204.09269.pdf>

GitHub 链接:

<https://github.com/LitterBrother-Xiao/Overview-of-Non-autoregressive-Applications>



## 科研第一线



### CVPR 2022 | 一键解锁微软亚洲研究院计算机视觉领域前沿进展！

国际计算机视觉与模式识别会议（CVPR）是人工智能领域最具学术影响力的顶级会议之一。微软亚洲研究院也在4月成功举办了 CVPR 2022 论文分享会。我们为大家精选了基于神经压缩的视频复原、基于神经辐射流形的三维可控图像生成、StyleSwin、SimMIM 学习轨迹感知 Transformer 的视频超分辨率网络等 8 篇微软亚洲研究院被 CVPR 2022 收录的优秀论文，带你探索计算机视觉领域的热点前沿！



扫描二维码查看文章

### ICLR 2022 | 微软亚洲研究院深度学习领域最新研究成果一览

ICLR (International Conference on Learning Representations) 是国际公认的深度学习领域顶级会议之一，众多在人工智能、统计和数据科学领域以及计算机视觉、语音识别、文本理解等重要应用领域极其有影响力的论文都发表在该大会上。我们精选了其中的 6 篇来为大家进行简要介绍，其中研究主题的关键词包括时间序列、策略优化、解耦表示学习、采样方法、强化学习等。



扫描二维码查看文章



## ACL 2022 | NLP 领域最新热门研究，你一定不能错过！

作为自然语言处理领域的国际顶级学术会议，ACL 每年都吸引了大量学者投稿和参会，今年的 ACL 大会是大会采用 ACL Rolling Review 机制后的首次尝试。在此次会议中，微软亚洲研究院有多篇论文入选，本文精选了其中的 6 篇进行简要介绍，论文主题涵盖了：编码器解码器框架、自然语言生成、知识神经元、抽取式文本摘要、预训练语言模型、零样本神经机器翻译等。



扫描二维码查看文章



## WWW 2022 | 一文解读互联网技术国际顶会最新方向

国际万维网会议 (Proceedings of the ACM Web Conference, 简称 WWW) 是互联网技术领域最重要的国际会议之一。今年的 WWW 微软亚洲研究院有多篇论文入选。我们为大家精选了其中的 6 篇进行简要介绍，研究主题关键词包括个性化新闻推荐、图异配性建模、多层推荐推理、日志解析、基于因果学习的可解释推荐、增量推荐算法等。



扫描二维码查看文章

# Ada Workshop: 从“她力量”到“她行动”，共促计算机领域多元发展

倘若在互联网上检索“计算机专业”，与“男孩子读计算机专业有哪些优越性”相对应的，则是“女生能读计算机专业吗”。这体现了一个普遍认知，“计算机”被认为与男性更为相配。

但是，女性真的不适合计算机吗？身边的女性研究员、工程师已经给了我答案，她们都技术扎实、勤勉努力，不断推动整个领域向前进步。与此同时，许多研究也表明，要推动计算机领域的创新与可持续发展，提升性别包容性是重要的一环。异质群体解决问题的表现常常优于同质群体，从业者的多样性也可以很好地激发思想并增强创造力。

因此，悬殊的男女比例并不利于计算机领域的技术进步，我们应当为此有所行动。在这样的视角之下，依托于微软的“多元与包容”文化，微软亚洲研究院自2016年启动 Ada Workshop，并不断拓展其外延，以多元的形式激励与启发心怀科技梦想的高校女生们，为她们提供切实的支持，助其顺利完成从学生到职场人的转换。

回望过去七年，通过 Ada 系列的三十余场活动，数万名学生与计算机领域内的女性榜样在线上与线下面对面，获得了在领域内继续探索的信心。活动的嘉宾们也在回顾、分享自己故事的过程中，收获了“教学相长”带来的领悟。更为重要的是，我们正不断汇聚更多志同道合的伙伴，为计算机领域带来与“多元与包容”相关的积极变化。



2016年第一届 Ada Workshop，微软亚洲研究院常务副院长张冬梅博士举着自拍杆与同学们合影

## 她力量：来自女性榜样，也来自每一个妳

要发起一个面向女学生的全新活动并非易事，突破自己的思维桎梏是第一步。作为一位从校园到职场都很出色的女性，Ada Workshop 发起者之一、我的同事孙丽君坚信女性和男性在工作能力上并无不同，如果做专门的活动来“照顾”女性，反而是一种不公平。直到读到题为《1984年：女性是如何一步步停止编程的？》的文章，并且了解到自己的女实习生找导师时碰壁的经历，她的想法才发生了变化。

女性在计算机发展史上光彩夺目却常被忽视。很少有人知道世界上第一位程序员是一位美丽的女性——Ada Lovelace。至80年代中期，计算机领域内女性的从业比例其实是逐年上升的，但后来，计算机被社会观念塑造为“不适合女性”，领域内的女性比例就开始不断走低，继而形成了今日悬殊的男女比例。

“我们应该为改变现状做些什么”，在微软亚洲研究院副院长潘天佑、学术合作总监马歆的指导与大力支持下以及孙丽君的努力下，2016年12月，第一届 Ada Workshop 落地北京丹棱街5号，120余名女学生以主题“Women and Computing”相聚在微软大厦。有学生兴奋地说，一下子出现这么多同性别且志同道合的朋友，简直是“一夜暴富”。

到今年，我们已经举办了六届 Ada Workshop 年度大会，并每年为她注入新的活力。我们邀请的分享者具有多样的身份，除了微软的研究员、工程师、产品经理，还有来自高校的老师、来自友商的优秀女性。她们的讲述很好地展现了科技女性的力量与灵动，为参与活动的女学生呈现了激励她们前行的生动榜样。正如青海大学党委副书记、校长史元春在演讲中讲到的，“计算机领域不断涌现着新的挑战，我们可以积极应对变化，并在不断肯定自己的过程中，活成自己想要的样子。”

分组午餐与女性榜样直接交流、HR 分享职场干货、Ada 渠道直通微软校招面试……除了主题报告与圆桌讨论，Ada Workshop 不断以更丰富的环节设计帮助学生成长。当疫情袭来，Ada Workshop 也从线下走到线上，通过在线直播普惠全国各地的女学生。

从 Ada Workshop 第一届到第六届年度大会，微软杰出首席科学家、微软亚洲研究院常务副院长、微软 Women Committee 大中华区主席张冬梅博士是其一路发展的见证者。她为“她力量”





史元春教授在 2021 年 Ada Workshop 上分享

总结了三层含义：其一，“她力量”来源于计算机发展史中做出了重要贡献的女性，她们给予了后来者感召的力量、榜样的力量；其二，“她力量”源于当下致力于技术进步、在计算机行业默默耕耘的万千女性，她们就在我们身边，并将力量源源不断地传递给我们；最后，最重要的“她力量”来源于每一位在校女学生——你们的潜力与能力，将为整个计算机领域和行业带来巨大无比的希望与力量。

## 她成长：刷新认知，也找到同行的伙伴

不自信，害怕自己做得不够好，不明白自己的核心竞争力是什么，这是参加 Ada 活动的女学生们最常提起的困惑。台上闪闪发光的分享者们，将“相信自己”的勇气传递给了前来参与活动的女孩子们。

Ada Workshop 带领她们重新认识计算机领域的女性，重新认识自己。“世俗的观念往往认为女生缺乏理性思维，天生与计算机、数学绝缘。但是通过 Ada Workshop，我才知道在科技行业里原来也有许多这样优秀的女性从业者，女生也可以在计算机领域做出丝毫不比男性逊色的成就。”一位北京大学的女孩这样说。



2019 年 Ada Workshop 上，现场参与活动的女学生

“这次活动之于我是一个转折点。不仅开阔了我的眼界，还加深了我对女性在行业中的认识——女性同样能在自己的圈子中产生重要的影响力。微软亚洲研究院对我这个低年级女孩子的照顾和指点，也让我在之后的很多学术会议、技术活动中更有自信，即使与行业大拿、来自世界各地的同学打交道，也毫不怯场。”一位来自重庆邮电大学的女孩在远赴北京参加 Ada Workshop 后这样描述自己的感受。

也有女学生通过 Ada Workshop 与微软产生了连接。“Ada Workshop 是我第一次与微软的研究员和工程师面对面。看到了那么优秀的微软研究员，了解她们的故事，给了我生动的榜样！对，那就是我想成为的人的样子！”毕业后，这位来自清华大学的女孩成为了微软的一员。

在刷新认知的同时，女孩们还找到了有共鸣的伙伴。在活动中，我结识到了许多来自不同学校的计算机专业的女生。通过 Ada 的平台，我们相互交流，不仅对未来的职业发展也有了更加清晰的认识，也更加坚定了继续从事计算机行业的信念。”

这份女孩间的默契也就在之后的线上活动中延续。在微信群中，大家积极交流学业问题、职业规划、计算机专业问题等，甚至找到了科研方向一致、想要深度交流的伙伴。“这个交流群太棒了。”有女孩这样说。虽然无法见面，交流依然无间。

这些个体的改变也体现在数据中。Ada Workshop 举办后发放的调查问卷显示，在评估自己的实践能力、逻辑能力以及是否愿意接受挑战时，参与过 Ada Workshop 的女孩有着明显更高的比例去相信“女性优于男性”或“女性与男性无差别”。

## 她行动：汇聚力量，共促多元发展

推动整个领域发生变化，仅靠我们的力量是远远不够的。自 2016 年 Ada Workshop 启动时，微软就在积极地向外汇聚更多力量，致力于促成一个女性与科技行业更紧密共生的未来。

高校是学生成长的土壤，也是我们最早携手的力量。2017 年起，我们每年都会与高校联合举办 Ada Workshop，得到了全国各大高校的巨大支持。一位中国科学技术大学的女教授在活动后感慨地告诉我，这是她第一次看到这么多计算机及相关专业的女学生坐在一个教室里，她感受到了“女性的独特磁场”，这种感觉“很美好，很珍贵”。

秉持着相同的愿景与目标，许多组织也在与我们共同努力。2021 年，我们首次与中国计算机学会女计算机工作者委员会（简称“CCF 女工委”）共同主办 Ada Workshop。今年，微软携手合作伙伴 Ladies Who Tech、领英、CCF 女工委，联动数字未来女性赋能行动（WDF）推出新计划，我们将以 Ada Workshop X Light Your Way 的形式，将小型沙龙落地到全国各地，为全国各



2021年, Ada Workshop 在中国科学技术大学举办

地为女学生提供与产业界、学术界优秀女性榜样交流的机会。6月, 首场活动已经落地武汉。

作为微软亚洲研究院学术合作总监, 同时也是 CCF 女工委主任的马歆, 促成了微软与 CCF 在助力计算机领域女性发展方面的深度合作。“激励与帮助高校女生是中国计算机学会女工委的重要使命之一, 也是研究院一直在做的事。”马歆说, “我也希望有更多像微软一样致力于推动多元与包容文化的企业与我们一起行动。”



2022年 Ada Workshop 上,  
微软携手 Ladies Who Tech、领英、CCF 女工委推出新计划

基于微软的国际化视野, 我们也在进行更多尝试。通过携手顶尖国际会议 SOSP 2017、ISSTA 2019 举办 Ada Workshop, 参会嘉宾和学生基于一致的研究方向, 进行了更多深度探讨。“我最大的收获就是确定了下一步工作的大概方向。通过 Ada workshop 上的分享, 让我真正感受到从事学术工作是一个螺旋上升的过程, 不能要求自己一下子做到完美。”一位来自中国科学院计算技术研究所的女学生这样讲述她的收获。

七年间, Ada Workshop 与计算机专业的女学生们共同成长, 我的角色也在不断变化——从对“多元与包容”文化不甚了解的局外人, 到负责 Ada Workshop 部分工作的实习生, 又到从同事手

中接过项目的接力棒, 开始独立策划、组织每年的活动, 我感到肩头的责任, 也促使我去思考更多, 努力带来更多改变。

正如微软亚洲研究院院长周礼栋博士曾说过的, “创新是一场灵感孕育、求证、推演、实现的协作竞赛, 而多元、包容的文化有利于我们在新一轮轮竞赛中不断取胜。”这适用于一个企业的发展, 亦适用于整个计算机领域的技术进步。

世界需要科学, 科学需要女性。共促计算机领域多元发展, 我们期待与你一起携手。

## 本文作者:

王靖雯, 微软亚洲研究院学术合作经理

## 相关阅读

[扫描二维码查看文章](#)

### 与她们并肩, 感悟“计算之美”的真实力量

当我们讨论“计算之美”时, 这种美丽到底是什么模样? 也许是像第一位给计算机写程序的人 Ada Lovelace 那样, 为计算机科学留下属于她的烙印; 也许是像中国计算机之母夏培肃那样, 用严谨治学的态度影响了一代又一代的计算机人才; 又或许与正在追逐梦想的每一个妳一样, 即使曾有困惑迷茫, 仍能坚持所爱, 找到属于自己的方向。



### 对话 | 什么样的学生适合做科研? 如何抓住机会实现突围?

“计算机领域的同学要如何开拓视野、抓住机会, 实现自我突破?” “什么样的学生适合做科研?”……在直播的圆桌环节, Ada Camp 特别邀请学界和业界嘉宾就同学们关心的成长方向、道路选择、科研进步等话题进行了深入交流。





## 科学匠人 | “轮椅上的学霸”，用科技推动无障碍发展

本期“科学匠人”栏目的主人公名叫矣晓沅，他用实际行动证明了在轮椅上一样可以书写精彩的人生。从清华校园到微软亚洲研究院，他是如何开启自己的学术生涯？对于未来的研究方向和职业规划他有怎样的思考？科技又将如何推动无障碍领域的发展？让我们一起走近微软亚洲研究院研究员矣晓沅。



在2022年2月19日举行的2021 CCF 颁奖典礼上，矣晓沅从颁奖嘉宾的手中接过了“CCF 优秀博士学位论文奖”的奖杯和证书，这是对计算机领域青年才俊们苦读的最好奖励之一。对矣晓沅来说，这一时刻显得更加弥足珍贵，二十多年求学路上的艰辛、汗水、坚韧与温情都融在手中奖杯和证书中。回顾过往，矣晓沅说，自己是一个普通人，只不过轮椅恰好成为了他生活中的必需品。

### 披荆斩棘艰难求学，心存感恩回馈社会

矣晓沅出生在云南省玉溪市。6岁时，他因一场高烧患上了“不死的癌症”——全身型幼年特发性关节炎，这是一个会慢慢侵蚀身体各个关节并使人逐渐丧失行动能力的疾病。11岁，在ICU昏迷了两个多月的矣晓沅，在苏醒后完全失去了行走能力，自此开启了与轮椅为伴的人生。15岁病情的再次恶化，让他的精神几近崩溃，甚至失去了生活的意志，在母亲的陪伴和父亲的开导下，矣晓沅重新燃起生活的勇气，在学习与读书中找到了新的希望。

“我唯一的武器就是学习，这也是我唯一可以倾尽全力去做好的事，”矣晓沅说。在不断拼搏、发奋学习之后，矣晓沅以云南省理科第16名、玉溪市高考理科榜眼的优异成绩考入了清华大学，被称为“轮椅上的学霸”。然而，由于对新环境和学习节奏的不适应，以及在编程上的零基础，使得矣晓沅的大一期末测试成绩很不理想。不服输的他迅速调整心态，铆足了劲向前追赶。而为了克服手指打字不便的困难，矣晓沅更是见缝插针地练习打字、写代码。到了大三，矣晓沅的成绩已经进入前十名，并在大四时拿到了全

校本科生特等奖学金，顺利保研本院，随后他又继续深造读博。

在“披荆斩棘”的求学成功之路上，家人细致入微的生活照料以及学校、老师、同学们所提供的学习帮助，都让矣晓沅心存感恩，他希望可以尽自己最大的努力回馈社会。因此，矣晓沅不仅热心公益，还作为发起人之一创办了清华大学学生无障碍发展研究协会，并担任第二任会长，持续为残障学生服务、发声。



矣晓沅与母亲在清华大学紫荆操场合影

### 计算机“小白”到科研“达人”，用AI绽放古典文学之美

从大学本科到博士期间，矣晓沅学习的都是计算机专业。谈及选择计算机专业的原因，他坦诚地说，“我在大学之前没有接触过编程，选择计算机专业完全是从我自身的就业角度考虑，编程工作对体力或行动方面的需求不是特别大。”由于清华大学有允许且鼓励本科生进入实验室进行科研实习的传统，矣晓沅也因此开始接触到了科研工作，在探索未知的乐趣中，他对自己的理想有了更清晰的描绘——希望自己能够掌握更多先进的技术，利用技术来造福社会。

今年年初，矣晓沅的博士毕业论文《具有文学表现力的中文古典诗歌自动写作方法研究》获得了2021年度CCF优秀博士学位

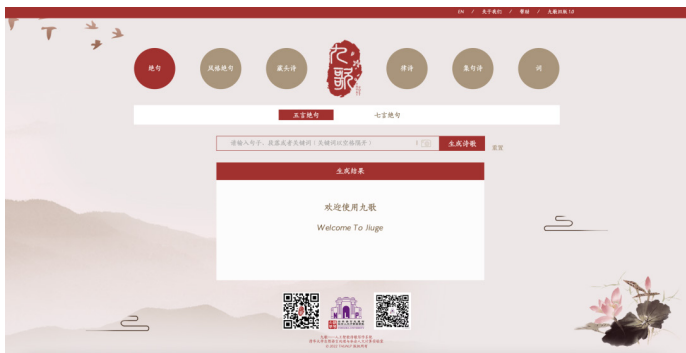


位论文奖。该奖项设立于 2006 年，旨在奖励在计算机科学与技术及其相关领域的基础理论或应用基础研究方面有重要突破，或在关键技术和应用技术方面有重要创新的中国计算机领域博士学位论文的作者，从而鼓励更多计算机科研领域的新鲜血液们做出更多创新。



CCF 理事长梅宏院士(右一)、CCF 奖励委员会主席钱德沛院士(左一)、中国计算机学会女计算机工作者委员会主任、微软亚洲研究院学术合作总监马歆(右二)共同为矣晓沅(左二)颁发 2021 年度 CCF 优秀博士学位论文奖

矣晓沅的这篇获奖论文从文本质量和美学特征上提升了 AI 生成古典诗歌的表现力，使得诗歌的连贯性、扣题性、新颖性和风格化都更胜以往。基于这一创新的自然语言生成技术，矣晓沅和他的导师还集成构建了在线中文古典诗歌自动写作系统——“九歌”，在诗歌创作的可读性、趣味性和情感上，也都给人以意境美的感受。



人工智能诗歌写作系统“九歌”

这项 AI 写诗的科研工作让矣晓沅有机会将自己的兴趣爱好与专业技术进行了完美的结合。“2016 年我读硕士时，正值 AI 热潮。其中，语言智能是 AI 非常重要的体现之一，但让 AI 进行古典文学创作还是一项非常具有挑战性的任务。当时我的导师一直在思考如何将 AI 应用于文学创作领域。而我本身是一个古典文学、古典诗歌爱好者，自己也会创作一些诗歌，所以加入实验室后与导师一拍即合。”

矣晓沅认为研究 AI 在古典文学中的应用有着重要的意义。首先就是从技术上探索和扩展 AI 的自然语言理解和处理能力。例如，一首中文五言绝句有二十个字，假设所有的汉字数量为八千个，那么形成的绝句组合就有  $80^{20}$  种可能，至于七言律诗的组合数量甚至会比宇宙中的原子数还多。如何从海量的数据中生成合辙押韵、意境优美的诗歌，对底层计算能力和算法都有着巨大的挑战。其次，矣晓沅也希望能够以热门的 AI 技术为载体，把中国优秀的诗词文化传承下去，并弘扬光大。

## 入职微软，无障碍软性制度和硬件设施同样重要

2021 年 7 月，矣晓沅从清华大学计算机系博士毕业。离开校园，踏入社会，矣晓沅与其他毕业生面临着同样的考验，求职、就业、初入职场的身份转换等问题扑面而来。除了普通毕业生所关心的就业问题外，工作场所和上下班路途中的无障碍设施情况也是矣晓沅所必须要考虑的。

在求职的过程中，矣晓沅经过筛选，拿到了包括微软亚洲研究院在内 4 家企业的 offer。“在我通过面试的几家公司中，只有微软在无障碍方面有明确的流程，这充分表明了企业支持无障碍建设的态度，毕竟要先有企业文化氛围、制度和流程上的保障，后续才会进一步构建和改进无障碍的相关设施。其实，此前通过一些老师和校友的介绍，我对微软的包容文化就有所耳闻，所以微软是唯一一家我在投简历时就表明了身体状况的公司。拿到 offer 后，公司就主动邀请我到微软大厦进行无障碍设施的体验，并征求我的意见和建议。后来，微软亚太研发集团主席、时任微软亚洲研究院院长洪小文博士也与我交流过公司内无障碍设施的设置与安排，并陪我从楼下一直走到工位，以亲自确认整条路径的无障碍程度，”矣晓沅回忆道。

微软的诚意打动了矣晓沅。正式入职微软亚洲研究院后他发现，当时提出的改进建议已经变为现实。例如，办公楼里的部分推拉门被改造成了自动感应门，还增加了更多的无障碍卫生间等等。微软还将一部原本参与联动的电梯，设置成了专供残障人士使用的独立运行电梯，从而避免了残障人士在多个电梯之间的等待和徘徊。矣晓沅表示，很多他此前并没有提出的建议，也被公司考虑到了，从这些点滴细节都可以看出微软为完善无障碍环境所做的努力。微软对建设无障碍环境的支持，无论是主观的软性制度、文化和氛围，还是客观的硬件设施，都超出了矣晓沅的预期。

“加入微软亚洲研究院之后，我更加充分地感受到了这个计算机领域优秀人才向往之地的魅力，矣晓沅说，“研究院不仅拥有自由、开放的学术氛围，这里所秉持的多元与包容的文化更是让身为残障人士的我倍感安心。之后我也会鼓励其他与我情况类似的学弟学妹们申请微软的工作机会。”



矣晓沅（右二）与社会计算组的同事们在办公室合影

## 将科研成果融入无障碍领域，让残障人士自由独立地生活

入职微软亚洲研究院后，矣晓沅成为了社会计算组的一名研究员，虽然入职时间不长，但他对自己未来的研究方向已经有了初步的思考。“我还是想继续推进自己在语言智能方面的研究。作为残障群体中的一份子，我也希望把自己的研究成果与无障碍领域相关的需求结合起来，真正用技术推动无障碍的发展。”

矣晓沅的理想与微软的无障碍承诺不谋而合——在行业和社会中鼓励无障碍技术的发展，利用无障碍技术为更多残障人士创造就业机会，以及建立一个对残障人士更具包容性的工作环境及场所。微软还意识到如果没有更多残障人士的深入参与，下一代无障碍技术的发展也将会受到限制。



矣晓沅（中间）与同事们一起参加公司活动

具体到研究方向，矣晓沅表示，未来他想从两个方面推动残障人士的就业和发展。一是利用自动生成技术帮助打字和输入困难者更好地利用文字交流，比如，通过技术预测并补全输入内容。可能很多人认为现在语音识别技术已经十分成熟，如果不方便打字输入，直接通过语音转文字也可以实现交流与沟通。但矣晓沅提醒说，“语音转文字的确很方便，但要注意的是，手动输入也是残障人士的权利。就像之前有专门为盲人设计的键盘，我们不能剥夺盲人使用键盘的自然权利。同时，还有一项很重要的一点就是隐私保护，因为并不是所有对话都方便在公共场合通过语音讲出来再转换成文字的，例如身份证号、帐号密码等敏感信息，以及一些私人的聊天记录，这些都是不适宜公开的。”

此外，矣晓沅也想开展针对残障群体的日常文字言论和表述方面的研究。网络环境错综复杂，很多人可能无意中在网上发表了冒犯性言论，对残障群体造成了言语上的伤害。比如“有听力障碍的人”比“聋哑人”更友好，但很多人并不清楚其中的差别。矣晓沅希望通过 AI 技术可以在网络上为不适当的表述做出提示，从而为残障群体创造一个良好的网络环境。矣晓沅在与研究院的同事们分享了这些想法后，也得到了大力支持。

与此同时，矣晓沅还将持续为残障群体服务、发声。“中国现在有 8500 万残障人士，约占中国总人口比例的 6.21%，但由于无障碍设施的不完善，他们在大众的视野中不常见。社会各界可以给予残障群体更高的关注，因为只有更深入地了解，加强无障碍建设的意识，无障碍环境才有可能不断改善。其实身体残障并不是一种疾病，而是一种障碍，这与不同的人会有高矮胖瘦的区别一样，只不过他们因为与其他人的不同导致了某些能力的不足。因此，残障人士不应该被差别对待，既不应该被歧视贬低，也不能被过分拔高。”矣晓沅说，“残障人士最大的渴望就是被社会当成一个健全人来看待，可以自由独立地学习、工作和生活。”

## 相关阅读

[扫描二维码查看文章](#)

### 科学匠人 | 白静：拥抱变化，不断发现计算机科学中的新天地

在计算机领域，研究与产品的关系往往十分微妙。一方面，二者相辅相成、互相推动；另一方面，它们追求的目标又不尽相同——产品需要精确的 KPI 及短期落地目标，研究则更注重宏观的、长期的影响力。微软亚洲研究院首席研究员白静博士在产品与研究领域都深耕多年，她希望自己的研究能成为这两个领域之间融会贯通的纽带。在她看来，让研究成果与产品产生共振，进而推动彼此进化，是研究的价值所在。





## 科学匠人 | 夏炎:是“超级奶爸”,也是科学研究与技术应用的“摆渡人”

科学研究与技术创新的过程总是充满了不确定性,科研人员无法提前计算创新的周期,也无法预料每个灵感所带来的最终结果。若想将一项研究成果落地并通过产品化的方式让更多人感受到前沿技术所带来的便利,研究工程师的参与尤为重要,他们需要全面掌握终端用户的需求,深入了解技术应用的深度与广度,打通各个环节的流程,有效地将算法模型与产品应用连接起来。然而知易行难,这一过程中的艰辛与技术落地时的成就感也只有亲身经历的人才能体会。下面,让我们从微软亚洲研究院研究工程师+“超级奶爸”夏炎的故事中,获取一份十年陈酿的科研成果落地经!

2018年,夏炎的生命中多了一个重要的角色——成为一名父亲,四年多的育儿经历,让他不仅习得全新的语言体系“婴语”,也看见了育儿中的大学问。夏炎发现,育儿与工作中竟然有很多互通的理念,对彼此都有着积极的促进作用。比如通过教练式管理的方法,以身作则,关心孩子的想法和兴趣,激发孩子的潜能;再比如,《如何说孩子才会听,怎么听孩子才会肯说》中的理念,与“倾听员工/团队成员的意见和心声”如出一辙。作为微软亚洲研究院育儿群中的活跃分子,夏炎经常会热情、耐心地与同事们分享自己学到的知识,也因此成为了研究院“享誉盛名”的“超级奶爸”。



如今,拥有首席开发经理与“超级奶爸”双重身份的夏炎,对科研转化和技术落地也有了更深入的思考,而在与众多跨领域专家,包括与儿子这个“儿童专家”的交流,也让他从不同学科中汲取着研究开发的灵感。

### 微软亚洲研究院的“骨灰级实习生”

2008年,还在读研究生的夏炎希望可以找一份实习工作,积累社会经验。很快,与微软亚洲研究院一直保持合作的导师为他做了推荐。凭借自身过硬的履历以及与研究院需求的完美匹配,夏炎顺利开启了他在微软亚洲研究院的实习之旅。而这个实习一做就是两年,早已超过一般实习生的实习周期,夏炎也成为了大家口中的“骨灰级实习生”。

实习期间,夏炎的主要工作是围绕着当时还处于早期发展阶段的互联网搜索技术而展开的。其中一个项目是关于企业的内部搜索,另一个项目则与学术搜索相关。尽管现在大家对搜索技术的应用已经习以为常,但这两个项目在当时都属于非常超前的技术研究与应用。

在跟随研究员们做基于论坛的问答机器人项目时,夏炎第一次接触到了机器学习。问答机器人的主要功能是在同一问题的众多回答中,自动提取最优质的回答进行配对。十几年前,推荐系统、机器学习尚在萌芽期,用户基于搜索引擎输入关键词后,还需花费大量时间阅读搜索结果才能筛选出可用信息,而此时微软亚洲研究院已经将机器学习技术转化到了微软的产品中,大幅提升了搜索的速度和精度。原本专业更偏向于软件工程的夏炎,在经历了将研究算法成功落地于实际场景的项目后,深切地感受到了前沿科技对人们工作生活所带来的巨大影响。两年的实习经历也让夏炎更加系统地了解了如何将研究员论文中的算法转化成对用户有用的工具和产品,为他之后从事其他技术转化项目奠定了基础。



实习期间,夏炎(第三排左三)与其他实习生小伙伴们一起出游

“在研究院实习的两年,我不仅接触到了最前沿的技术,还有机会和很多聪明人共事。无论是其他实习生还是一起合作的研究员,很多都是领域中的大牛,我可以和他们一起做对社会有意义



的技术和产品。对一个在校学生来说，微软亚洲研究院极大地开阔了我的见识和眼界，这也是研究院最吸引我的地方”，夏炎说。

## “既专又广”是做科研转化的必要能力之一

正是这份吸引力让夏炎在研究生毕业后，直接选择加入了微软亚洲研究院。从实习生到正式员工，夏炎并没有感到很大的不同，但是需要从更多做辅助性的工作，逐渐成长为整个项目的负责人。而在一次次将研究成果转化为产品的项目历练中，夏炎愈发意识到科学探索的挑战性和技术转化的重要性。

夏炎曾与微软必应（Bing）产品团队的同事合作，将研究院以图搜图的算法集成到必应搜索中。为了实现这一功能，身为研究工程师的夏炎既要知晓底层索引技术的实现机制，了解中间的运行过程以推荐最合适的内容，还要熟悉最上层的前端交互设计。此外，开发过程中还会涉及到跨团队、跨语言的沟通与协作。在这项技术功能化的过程中，夏炎意识到要想把研究成果成功转化进产品，需要把研究模型、产品功能、交互界面等各个环节全部打通，这对开发人员的综合能力提出了很高的要求。

此后，夏炎又和研究院自然语言计算组的研究员一起对必应的搜索场景进行了优化。当用户在输入诸如“世界上最高的人是谁？”、“全球最富有的人是谁？”这类问题时，搜索引擎可以借助这一功能在海量的数据中自动找出答案并直接呈现出来。在这个项目中，作为科学研究与产品落地的连接人，夏炎不仅要理解自然语言处理（NLP）算法模型的实现原理，还要了解用户对产品的需求。比如，在研究中模型进行计算的耗时在产品端是不可接受的时长。这时工程人员就需要重新解析代码，或者更换另一种编程语言，在实现同样功能的同时，精细地甄别、选择最有价值的特征，减少不必要的计算，并且尽量让模块进行并行处理以达到最短的返回结果时长。只有不断对算法模型进行裁剪、优化、提升性能，产品才能达到上线标准，满足用户的体验。

聊天机器人的落地是夏炎参与的另一项重要工作。要打造一个既可以闲聊又可以提供问答等进阶功能的对话机器人，需要整合多种技术，包括 NLP 领域的语义匹配、文本质量分析、知识图谱、情感分析，以及语音识别、语音合成技术等等。因此，夏炎与各个研究团队都保持着紧密的联系，即时了解各团队的研究方向和成果，以及这些技术发展的最新动态，从而可以在面对产品需求时有备无患。

在对聊天机器人研究的过程中，夏炎和同事们又有了新想法——将 AI 与语言学习相融合，这也是他近几年的工作重点。尽管彼时包括必应词典在内的语言工具可以通过查找释义和例句帮助人们学习单词，但语言学习还有听、说、读、写等更丰富的内容。从技术角度看，语言学习与对话系统、聊天机器人有很强的关联性。于是，夏炎和同事们着手将微软亚洲研究院领先的语音识别、NLP 等 AI 技术引入到了必应词典的功能中。



夏炎（第一排左三）与微软亚洲研究院的同事们合影

2016 年，夏炎和团队将研究院高准确率的语音评测和语音打分等模型结合，开发了个性化英语学习助手“微软小英”。通过模拟日常生活中的对话场景，“微软小英”可以为英语学习者提供沉浸式的口语练习体验。而在解决了英语学习的“听”、“说”难题之后，夏炎和团队又推出了可提供英文写作指导与反馈的微软“爱写作”，通过 NLP 技术完成语法纠错、词语替换、写作评分等功能，从而改善英文写作的准确度和流畅度。

基于在英语学习方面的经验，夏炎和同事们不断探索、尝试，于 2021 年与华东师范大学联合推出了中文写作智能辅导系统“小花狮”。“小花狮”利用现有的中文教学知识的研究成果，结合 AI 技术，在帮助学生们提升中文写作与阅读能力的同时，也辅助老师实施更有针对性的教学。“语言学习并不是纯技术导向的任务，我们需要紧密地与教育专家合作，获取更专业的教学理念和方法，以不断提升学习者的学习效果和使用满意度”，夏炎认为。

从搜索查询到对 NLP、语音技术的探索，从以图搜图，到聊天机器人、AI 语言学习项目的打磨，夏炎对科学研究与技术转化有了更深层的理解。他说“研究工作要既专又广，专的是技术领域，就像有的研究员专注于语音识别、有的专注于自然语言处理或计算机视觉等，广则是指潜在的应用范围，每一项研究工作都是在解决通用任务中的一个核心问题，是为了拿下更泛用问题下技术挑战中最难啃的骨头。工程与产品团队同样也需要又专又广，专的是用户需求和场景，让技术在当前场景中奏效，广则是让技术的适用范围广，解决同一个场景的用户需求时要使用众多不同技术。总之，我们要在具体场景中找到通用研究成果，来解决终端用户的问题，并收集具体场景数据与研究团队一起进行领域研究，使技术更好地满足现实中的场景需求。”

## 跨领域成功合作的关键在跨语言体系的沟通

随着微软亚洲研究院创新技术的成熟，越来越多的技术成果开始在更广泛的行业中应用，科学研究也呈现出跨领域交叉融合的趋势。对此，夏炎表示在跨领域合作中，成功的关键是跨语言体系的沟通。

“所谓跨语言体系指的是合作交流中不同领域人员对专业术语、领域知识的不同认知，”夏炎介绍道。比如，在进行 AI 语言学习项目的研究中，计算机领域的科研人员与教育领域的专家在前期的交流中就遇到了障碍：研究员们会从计算机技术的角度介绍算法的实现、效果、准确率等，更多考虑的是后验概率和物理量；而教育专家们则会从教育理念、学生和教师以及家长使用的不同角度出发，关注技术转化成产品后是否对终端用户有实际的帮助。合作过程中，夏炎和同事们也在逐步摸索如何连通不同的话语体系，“保持开放的心态以及谦虚的学习精神，可以促进更加高效地跨领域沟通。”



微软亚洲研究院 AI 语言学习项目团队  
访问华东师范大学教育学部（夏炎：右二）

升级为“奶爸”后的夏炎，还在自己的跨语言体系中加入了“婴幼儿”的沟通。为了更好地理解孩子的行为与孩子交流，夏炎阅读了大量的科学育儿书籍。例如，育儿领域的权威书籍《儿童及青少年发展心理学》，除了可以让家长更深入地了解孩子的心理，更有效地和孩子沟通外，还可以作为判断其它育儿书籍好坏的标尺；《PET 父母效能训练》让父母与孩子发生冲突时，遵循问题就近原则；《认知天性》可借助心理学理解孩子的所思所想，减少父母不切实际的期望等等。尽管这些育儿理念不会直接影响到项目中模型的设计和功​​能，但它们会潜移默化地帮夏炎做出更符合孩子天性和学习规律的相关产品。



带娃中的夏炎

“在跨领域合作中，研究工程师要成为编码和解码的中间人，通过学习并理解合作领域的知识，一方面将研究员的话语翻译成领域专家的语言体系，另一方面也要把领域专家的需求解释成具体模型的输入输出逻辑，找到大家共同的兴趣点，只有这样才能推动跨领域项目的成功实施”，夏炎说。

## 相关阅读

扫描二维码查看文章

## 她们非同凡响的另一面

“人类经常少年老成、青年迷惑、中年喜欢将他人的成绩与自己相比较，因此觉得受挫，好不容易活到老年还是一个没有成长的笨孩子。我们一直粗糙地活着，而人的一生，便也这样过往了，”女作家三毛曾这样说过。

这些描述是否与你的人生轨迹有些许重合？我们都不约而同地走上了同样的路，忘记了最初的梦想，失去了曾经的兴趣，与那个不一样的自己渐行渐远。不过，在微软亚洲研究院有这样一群女性，她们在看似重复的科学研究和日常琐事中不断探寻内心，始终保持自我，绽放出不一样的烟火。





## 对话 2021 微软学者：青年从不缺少描摹未来的潜力

“微软学者”奖学金是微软亚洲研究院自 1999 年启动的一项面向亚太地区计算机及相关专业和交叉学科优秀博士生的项目，旨在发掘优秀、有潜力的低年级博士生。

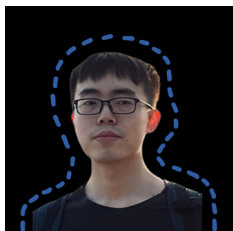
我们与四位 2021 年“微软学者”们聊了聊天，他们的科研生活，或像“探险家”一样，在不同科学领域里穿梭；或像“侦探”一样，不断挖掘现象背后的问题……精彩的背后，他们也曾碰壁，也曾借着灵感的微光向前摸索，靠坚持慢慢积累。期待他们的科研经历与成长感悟，能给予同样在科研之路上探索的你些许鼓励与勇气。

“微软学者”们如何选择与探索自己的研究领域？怎样的科研习惯帮助他们快速成长？申请与获得“微软学者”奖学金的过程中又有怎样的独家体验？

我们与 2021 年的四位微软学者代表展开对话。他们的科研生活，或像“探险家”一样，在不同科学领域里穿梭；或像“侦探”一样，挖掘现象背后的问题……但在精彩的背后，他们也曾碰壁，也曾借着灵感向前摸索，凭着坚持慢慢积累。

一起解锁与他们的聊天记录吧！

### 从“科研小白”到两大领域“双栖选手”



吴齐天 上海交通大学  
导师：严骏驰  
研究兴趣：机器学习、数据挖掘

#### Q：和“微软学者”奖学金的缘分如何开始的？得知自己获奖时是怎样的心情和状态？

**吴齐天：**我在很早之前就听说过“微软学者”奖学金，当时主要是看到有上海交大的学长获得了这个奖项，其实那个时候也没想过我能成为这个奖项的获得者。说实话，得知自己获奖的时候，还是挺吃惊的。我觉得这个奖项对我来说更多的是一份鼓励，它提醒着我要保持一个好的研究品位，做出真正有价值和实际意义的研究。

#### Q：为什么选择在机器学习、数据挖掘领域进行科研探索？

**吴齐天：**对于研究方向的选择，实际上是经历了一个漫长的探索过程。我最开始接触数据挖掘的研究，是本科阶段做的一个关于社交网络微博热度分析的科研项目，后来去企业实习又开启了关于推荐系统的研究。数据挖掘是一个与实际问题紧密接轨的研究领域，我们研究的科学问题一般都是从实际场景中提炼出来的，最后又通过在真实数据上的实验验证回归到实际中。

我后来渐渐又转向了机器学习领域，是因为在做数据挖掘的研究项目中，发现了一些现有方法和范式的不足，而解决这些问题往往需要在底层方法论上做出突破和创新。

我觉得这两个研究领域也是相辅相成的，数据挖掘更多时候从应用出发去寻找一个贴合问题的方法，而机器学习更多时候是从一般化的方法论出发再去寻找合适的应用场景。未来我也会长期同时在两个领域进行探索，做出一些既有理论基础，又能实际落地的东西。

#### Q：科研之路上有什么印象深刻的经历？对于这些经历你有怎样的感悟、启发？

**吴齐天：**印象比较深刻的两段经历分别是刚接触科研的那段时间以及刚转型到机器学习的阶段。

在这两段经历里我都经历了很长时间的挫折。刚接触科研那会儿，我从“科研小白”到中稿第一篇论文，大概经历了 5、6 次的拒稿。而后来从一个领域转到另一个领域，也因为不太适应两者各自的侧重点，中途经历了很多次波折。

不过在第一段经历里反复对论文的修改和精进，也给了我很大的帮助，尤其是对日后我能够独立开展科研起到了积极影响；而第二段经历更像是一种沉淀，我在这一过程中也逐渐培养起了学术鉴赏能力和研究品位。



## 以人为中心，拓展机器学习领域应用



Hiromu Yakura 日本筑波大学  
导师: Masataka Goto  
研究兴趣: 人机交互、机器学习

### Q: 你未来的研究目标是什么?

**Hiromu Yakura:** 作为一名人机交互领域的研究者，我的研究方向是拓展计算机的应用领域，特别是机器学习技术。鉴于我们不能否认机器学习模型存在错误的可能性，为了实现实际应用，我们需要仔细设计模型与人之间的交互。在这种情况下，我认为理解机器学习模型和人类的特征是至关重要的。

如果没有对人类的精确理解（例如，了解人类的优点和缺点），我们就无法确定我们应该将机器学习应用到哪一部分。换句话说，通过积累人类和机器学习之间实际合作的例子，我们将能够使用来自人类和机器学习之间对比的新词汇来描述、认识自己。这就是为什么我把我的研究与其他跨学科领域联系起来，比如行为经济学和文化研究。我希望以这种“以人为中心”的方式探索机器学习的实际应用，从而发现理解人类的新视角。

### Q: 在研究的道路上，是什么激励着你?

**Hiromu Yakura:** 我认为好的科研将带来持久影响，这样的研究有着无限的未来价值，这是我投入科研的重要原因。作为一名程序员，我充分认识到一个应用程序或系统将带来的影响，它使事物变得可扩展和可复制。与此同时，每个应用程序或系统背后，都存在着关于如何使用计算机的抽象观点和想法，这些思想具有更大的潜在影响和价值，因为它可以应用于各种情况，包括未来的情况。在科研道路上，我希望通过探索和形成这些抽象的想法，为科学的长链做出更多贡献。

### Q: 在“微软学者”面试过程中，你有什么印象深刻的故事可以和我们分享吗?

**Hiromu Yakura:** 一位面试官问了我一个概念性问题，是关于我曾经做过的一个项目的。该项目 (Yakura et al., IJCAI 2021) 允许用户在他们的创作过程中利用各种深度感知指标，而不是单纯的比较，面试官针对这一点向我发问：为什么我们一开始可以说这种感知指标“更好”？

这是一个很重要的问题，而我还没有准备好答案。这件事让我意

识到，我们方法的优势不只是使用感知指标，而是允许用户根据自己的喜好使用各种感知指标。这是一个出乎我意料时刻，但它启发了我：什么才是这个项目所提出方法的真正价值所在。

## “阅读”与“质疑”，让我在科研中不断成长



汪庆 清华大学  
导师: 舒继武、陆游游  
研究兴趣: 内存 / 存储系统

### Q: 你的长远研究目标是什么？它可以给系统领域带来哪些变化?

**汪庆:** 从某种层面来看，系统领域的研究如同构造建筑，是兼顾实用性和艺术性的研究。我的长远研究目标是将整座数据中心抽象成一个巨型共享内存机器，进而能够高效地存储、检索和处理数据，以支撑诸多现代应用。目前对于一个具体应用，都需要专门实现一套对应内存系统，然后再去优化它。如果我设想的这个研究目标被实现的话，内存这部分就能通用化，开发者们只需要聚焦于系统的计算部分，这样可以提高数据中心的内存性能和利用率。

有两个方面的因素共同促使我产生这个目标：第一，目前很多应用对内存的需求比较大，比如内存计算、内存数据库等，这是从应用角度驱动我的一个原因。第二，现在出现了很多新的内存介质，也有一些新的网络介质可以把不同机器的内存连在一起，这是从硬件方面驱动我的另一个原因。

### Q: 你有哪些科研习惯，能帮助你在科研中不断成长?

**汪庆:** 一个是阅读的习惯。阅读论文就像阅读新闻，是一个了解并学习别人在做什么的机会。在阅读的过程中，也伴随着延伸思考，我在阅读时会设想“如果是我去解决这个问题，会怎么做？”让自己作为系统设计者去感同身受。

另一个是质疑的习惯。任何系统都不会是完美的，一般只有在某种特定条件或环境下，它才可能表现出优势。所以当我去阅读论文或运行现有系统时，需要像一个“侦探”，去寻找文字或代码背后是否掩藏了一些问题，或者有些东西没有被明说，我一般会抱着质疑的态度去思考：“它到底是不是这样的？是否仅在某种特定条件下才能体现优势？所以它真正的问题在哪里？”

### Q: 申请“微软学者”奖学金是一种怎样的体验?

**汪庆:** 申请“微软学者”奖学金的整个流程我觉得很正式, 从写研究陈述、准备推荐信, 到最后网站提交、面试, 都与论文投稿有些相似。但与之不同的是, 普通论文投稿一般只针对一个研究工作, 而申请“微软学者”奖学金是需要对自己整个博士阶段几年内的研究工作进行梳理和总结, 当成一篇论文展示出来。这个过程帮助我去梳理了自己的研究, 开始思考每项研究工作在我的整个研究框架当中处于什么位置, 这对我的全局观产生了比较正面的影响。

此外, 面试过程中有一件令我印象比较深刻的事情: 一个面试官鼓励我使用编程语言领域的知识去解决计算机系统问题, 因为不同的应用对分布式内存缓存一致性的需求不一样, 面试官启发我通过编程语言的方法, 设计专门的编程模型去支持内存系统。我觉得这是很有意思的方向, 能够通过领域交叉碰撞新的研究想法。总的来说, 整个申请过程既有助于我梳理之前的工作, 也对于我下一步的研究计划带来了启发。

## 多领域交叉, 发现光计算的新大陆



周天贶 清华大学  
导师: 戴琼海、方璐  
研究兴趣: 光智能计算、机器学习

### Q: 光学计算是包涵多个学科的交叉领域, 你认为这个领域有什么独特魅力?

**周天贶:** 是的, 从研究的角度来说, 光学计算是多领域的交叉学科, 它既依赖于物理、化学、材料科学这些基础学科的前沿进展, 也取决于电子工程、计算机科学这些工程学科在系统架构以及算法设计方面的突破。

这个领域的独特魅力是能够在不同的学科间驰骋, 将不同的学科融合到一起, 让他们发生奇妙的化学反应, 在这个过程中, 一些固有的瓶颈就会被打破, 我们看到一些古老而经典的物理概念, 比如光强、相位, 在人工智能的赋能之后, 就能构建起复杂的光学神经网络, 展示出优越的计算性能。我非常期待能参与并见证, 随着不断深入的交叉和融合, 光学计算将改变现有的计算范式, 开启一个新的纪元。

### Q: 学科的交叉性给你带来了什么独特体验或经历?

**周天贶:** 最有趣的经历是和跨学科学者讨论, 看到不同领域间优势互补, 彼此增益。比如我们在进行光学材料设计的时候, 传统的设计算法通常会有尺寸或性能上的不足, 为了实现指标和性能的突破, 我们将人工智能的算法用于材料的设计过程, 实现了更紧凑、更高性能的光学器件。

当前, 学科之间的边界其实正在变得越来越模糊, 一个比较大的趋势是传统的课程和学科划分已经满足不了越来越交叉的研究需要, 因此我们要凭借自己的兴趣多涉猎其他相关学科的知识内容和体系, 我在博士阶段对人工智能和光学方向的课程都有学习。不同学科的熏陶使我们习得不同学科的思维模式, 为灵感的出现打下基础。

### Q: 在获得微软学者这半年多的时间里, 它是否为你的科研探索带来一些变化?

**周天贶:** 获得“微软学者”激励我要努力做出更有影响力的光计算研究, 争取用光计算的方式来解决计算机科学面临的实际问题。相比之前, 我除了注重理论研究之外, 开始更多关注光计算的应用场景。我希望自己研究的内容不仅仅只是停留于发表论文, 而是转化成一项惠及每个人的技术。希望在不久的将来, 能从硬件和软件层面构建一套实用的光计算体系, 在电脑、手机、智能电视中加入一块光学计算模块, 就可以提升这些设备的计算效率, 改变各行各业人们的生活体验。

通过与 2021 微软学者对话, 我们不难发现: 青年从不缺少描摹未来的潜力, 而“微软学者”奖学金的设立, 正是一笔对青年人的投资。过去二十余年, 先后有 400 多名优秀的博士生获得微软学者称号, 其中多位微软学者已经成为学术界中流砥柱或工业界翘楚。

未来, 我们将继续鼎力支持有潜力的学生, 助力他们在科研的汪洋里, 激起一朵朵浪花。

## 相关阅读

[扫描二维码查看文章](#)

微软学者讲座 | 应用计算技术解决生物医学与公共卫生领域未知问题



# 计算机科研“打工人”一天的隐藏关卡

欢迎进入，  
微软亚洲研究院计算机科研“打工人”的世界！

开启“打工人一天的隐藏关卡”！

**计算机科研打工人**

## 实习生

活力 耐心 ·人物关键词

- 活力满满
- 科研蕉绿
- 灵感小生
- 细致

稳重

·人物画像

·人物成分

- 活力: +150%
- 能量: +120%
- 稳重: +50%
- 耐心: +40%
- 细致: +70%

能量

9:00  
始发站  
学生宿舍

9:30  
茶水间

为什么来公司的第一站是茶水间?  
一天的体验结束后即可获得解答哦 >>

10:00  
与微软总部 mentor 交流

竟然还有国外的 mentor ?

11:00  
写论文

14:00  
参加内部 前沿技术讲座

15:00  
头脑风暴

16:00  
继续写论文

17:30  
与国内 mentor 晚餐交流

19:00  
与篮球社小伙伴来一局

除了篮球社实习生还有哪些有趣的活动 ?

21:30  
宿舍提升自我



\*温馨提醒：“?”是打工人世界里的彩蛋哦，你可在体验各角色的一天后收获新的“技能”解答。

### 为什么来公司的第一站是茶水间？

公司每天都会提供丰富的零食、饮料和水果，活力满满才能诞生更多灵感哟！

### 竟然还有国外的mentor?

参与星跃计划的同学将会在国内外研究员/工程师的共同指导下进行科研工作，与来自不同研究背景的资深前辈进行深入交流。

### 除了篮球社，实习生们还有哪些有趣的活动？

我们还有桌游社、羽毛球社、攀岩社、吉他社、密室社等丰富多彩的社团活动等你参加！



计算机科研打工人

# PM

人物关键词: PPT高手, 健身达人, 行动力Max, 细致

人物画像

人物成分:

- 活力: +90%
- 能量: +90%
- 稳重: +90%
- 耐心: +120%
- 细致: +150%



9:45  
茶水间外交

? 什么是茶水间外交?

一天的体验结束后即可获得解答哦 >>

10:00  
外部合作伙伴交流会

11:30  
码PPT

13:30  
项目(1)开会

15:30  
项目(n)开会

16:30  
健身房撸铁

? 工作时间可以撸铁?

17:30  
码PPT

19:00  
晚饭后继续码PPT

22:00  
睡前阅读

? 晚上阅读都有哪些推荐?



**什么是茶水间外交?**

除了定期的会议,我们会经常在茶水间接水或者吃零食时顺便聊一下项目的最新进度。

**工作时间可以撸铁?**

公司就有健身房,有时候我们跑一个CI/CD的时间就先去楼下健个身,回来再继续fix bug。

**晚上阅读都有哪些推荐?**

作为PM,需要大量了解其他领域的知识,尤其是项目相关领域的,因此我们的阅读也会围绕所需要了解的领域展开。

计算机科研打工人

# 研究员

人物关键词: 科研大佬, 全能奶爸, 脑洞大, 细致

人物画像

人物成分:

- 活力: +70%
- 能量: +90%
- 稳重: +100%
- 耐心: +120%
- 细致: +95%



扫描二维码查看  
研究员奶爸的一天

## 计算机领域的女生如何规划职业发展道路？

2022年4月，微软亚洲研究院在线上成功举办了 Ada Workshop 2022，旨在为计算机及相关专业的女学生提供聆听女性榜样成长之路、了解计算机领域最新动向的机会，吸引了两万余名观众在线观看。活动邀请了来自微软和学术界、产业界的优秀女性榜样与男性同盟，为同学们带来了个人职业发展之路上的启迪感悟。其中，在“她力量，耀前行”主题圆桌论坛环节，各位嘉宾针对同学们关心的成长方向、道路选择和科研进步等问题，奉献了精彩的分享与讨论。观众反响热烈，提问踊跃。快来查收这份圆桌论坛的精选内容，希望能为正在迷茫的你坚定选择！



**金芝**  
北京大学教授

### 龚文君

微软（亚洲）互联网工程院  
首席产品主管  
微软（亚洲）互联网工程院  
北京女性员工委员会主席



### 曹婷

微软亚洲研究院  
主管研究员



### 刘铁岩

微软亚洲研究院副院长  
微软杰出首席科学家



### 主持人: 孙丽君

微软亚洲研究院  
资深学术合作经理



2022  
Ada  
Workshop

**孙丽君:** 非常荣幸今天能邀请到来自学界和业界的顶尖教授和研究员们与同学们分享行业前沿，共话职业发展。我的第一个问题想问一下刘铁岩博士，作为一名博士生导师，您一定培养过很多优秀的学生。能否和我们一起介绍一下您培养过的第一位女博士生以及她的成长故事？

**刘铁岩:** 我在微软工作期间确实指导过很多实习生，也作为博士生导师带过许多同学，并且其中有很大的比例是女生。我带过的第一个女生是孙诗昭博士。她是微软亚洲研究院与南开大学的联合培养博士，主要从事机器学习理论和算法方面的研究工作。我印象特别深的是诗昭有着非常强的学习能力，在理论、算法、实践各个方面都有着很均衡的发展。她特别好学，即使遇到自己不懂或者是不熟悉的领域，也会选择非常勇敢地去拥抱它们。

我清晰地记得，她刚进入研究院开始博士学业时，对机器学习知识积累其实并没有那么多，尤其是理论方面。第一年，我们把诗昭送到了北大王立威教授的研究组去提高她的理论修为。在那半年中诗昭表现得非常优秀，从一个理论方面的小白，到可以非常熟练地运用统计学习理论来解决一些很难的问题。她当时也是学术界第一批从机器学习泛化理论的角度，对深层神经网络层数与宽度之间的一种权衡进行理论刻画的学者之一。毕业后，诗昭加入了微软亚洲研究院，继续在科研的道路上向上攀登着。

**孙丽君:** 铁岩的分享让我们看到了一个女生在机器学习领域不断成长的故事。下一个问题我想给到曹婷。我很好奇当年你在澳洲攻读博士学位的时候，你身边计算机领域的男女同学比例大概是怎样的？现在你观察到你们组里的同学们，他们的学校以及周围的环境大致是什么样的？和你当时比有没有一些改变？

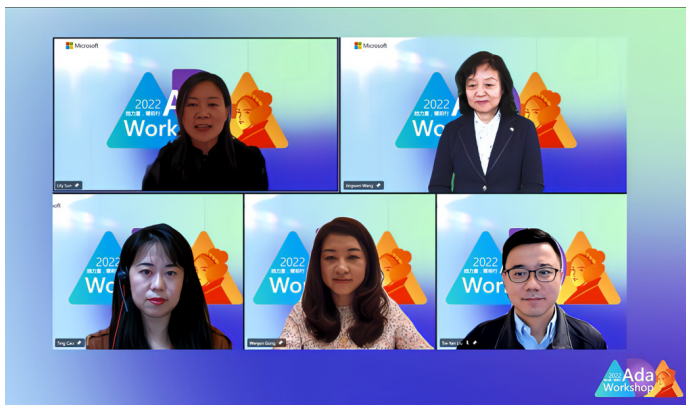
**曹婷:** 我觉得对比还是挺强烈的。我上博士的时候，我们学院整个一楼层就只有我一个女生，总是感到很孤单。但是现在在亚研女生还是很多的，我们组好多时候都是 1:1 的比例，变化确实挺大的。

**孙丽君:** 1:1 的男女比例是微软在过去几年一直努力的方向。但放眼目前国内或国际学术界的环境，尤其是在计算机领域，女生依然是少数派。从学生时代的专业选择，到进入职场之后的发展轨迹，女性可能经常会面临着一些社会偏见，因此在无形之中造成了计算机领域的性别失衡。我相信，包括各位在内的很多个人和组织都已经认识到了这种情况，并且希望去改变，这也是为什么我们过去几年一直

在做 Ada Workshop。接下来想请各位嘉宾分享一下，您所在的组织或群体，在过去 5 年对多元和包容逐步在达成一个什么样的共识？以及您观察到有什么样的举措来推动这个领域的多元和包容？

**金芝：**确实，男女比例不平衡的现象是存在的，对我而言，我是希望尽力去帮助女生的发展。我不管是在北大还是在科学院招生时，在同等条件下我愿意多给女生一些机会。我也看到，有很多优秀的女生都非常踏实，也会积极配合导师推动工作进展，最后也都取得了非常好的成果。比如，我们组之前有两位女生在做深度学习代码，她们有一位是去年毕业的，有一位是今年毕业的，她们都分别获得了 ACM SIGSOFT Distinguished Papers。我认为在软件程序设计领域，女生们也是非常优秀的，她们可以尽情地发挥她们的聪明才智，以及对美好事物的追求。

**曹婷：**多元与包容是微软价值观的一部分，我现在也不断地把它融入到我自己的价值观，作为平时指导我行为的一个准则。我觉得大家可以直观地去理解多元与包容，即我们可以去接受各种各样的人，然后去创造条件把每个人的潜力都发挥出来，不要受制于年龄、外貌、信仰等束缚。这几年我也明显地感受到，通过社会多方的努力，越来越多的声音在鼓励女生们去接触她们之前没有太多机会接触的领域，如计算机。在研究院，我们也推出了一系列的活动，像是 Ada Workshop 和 Ada Camp，帮助女生们去追求和实现她们的梦想。



**孙丽君：**在这方面我也是深有共鸣。对多元和包容这个概念，从一开始接触，然后不断地内化，再到逐渐去影响身边的人，我认为这是一个不断循环、螺旋上升的过程。那么接下来我想请教一下文君，微软（亚洲）互联网工程院的同事们对多元与包容的共识，尤其是在赋能和激励女性成长这一方面是怎样逐步形成的？女性和男性员工，还有领导层，他们分别扮演了怎样的角色呢？

**龚文君：**首先我想说的是微软对于多元与包容文化的认可。包容性和多元化是我们人类社会一个很自然的事情。只有我们去承认这个文化，或者说把这个文化融入到公司的 DNA 中，我们才能做出有包容性的产品，能够实现公司的愿景——予力全球每一人，每一组织，成就不凡。

此外，我们必须承认社会上一些偏见对女性造成的挑战和困扰，因此我们需要做一些事情去消除这些偏见对女性员工造成的影响。在各种各样的活动中，我们看到了很多优秀女性都在承担榜样的角色，不停地告诉大家女性可以做得很好，鼓励更多的女性走得更远。男性员工和领导们其实在承担支持者的角色，与女性一同去塑造一个无差异的环境，那么女性能够很自然而然地在这个环境里有更好的发展。

**孙丽君：**铁岩作为微软亚洲研究院的副院长，曾经也是研究院多元与包容委员会的负责人，请分享一下您观察到研究院在过去几年里，针对多元与包容共识上有什么样的变化以及趋势。

**刘铁岩：**对微软来说，多元与包容是我们企业文化的一个重要组成部分，而且大家对其的认知也是非常多元的。今天我们谈的很多是有关女性的话题，但多元与包容的涵盖面非常广。我们希望的并不是仅仅针对某个特定的群体，而是营造一个整体都非常包容的环境，鼓励大家可以换位思考，可以去共创，所以每一个人在这个环境中都可能是对别人的付出者。因为我们每个人都不一样，所以一旦有了整体的包容环境，我相信整个企业、社会都会向更良性的方向发展。

在过去的几年里，微软亚洲研究院做了很多具体的举措，比如，Ada Workshop 以及 Ada Camp。我们通过这样的形式，向广大的在校学生提供一些职业选择和学术发展方面的建议。公司内部也有很多的活动，比如，我们会有 workshop 或是交流群，帮助女性员工及女实习生之间互相交流。我们也为女性员工提供内部导师，从而帮助她们解决一些职业上的困惑。再比如，在员工招聘或实习生招聘时，我们都会非常有意地提高女性候选人的比例，以及保证在面试的过程中有女性面试官的参与，这样我们对这些候选人就可以有一个更全面的考量。

其次，在我们研究院里也有许多非常优秀的女性领导。我们也会组织很多分享会让这些成功的女性领导者把自己在事业上的经验，以及工作和生活中的心得体会进行分享，从而促进更多女性员工可以有更好的方式来规划自己的职业发展。

我想这些努力有一定的成果。在过去的 5 年里，我们女性员工的比例有大概 40% 的增长，这是一个很高比例的增长。我们也看到研究院文化建设方面的工作氛围也在变得更多元、更多样、更有人文关怀。

**观众提问 1：**请问各位嘉宾，你们认为自己现在的职场角色需要什么样的特质或能力？女生来从事这些角色的话，又有什么特别的优势，或者是面临怎样的挑战？

**金芝：**人走向事业有三个阶段，了解、投入和热爱。我发现大部分女性往往有一个特质，就是干一行爱一行。所以我相信，只要你深入地去了解这个领域，真心地去融入它，就会爱上它，也就



不会觉得工作是负担了。但是，一位女性图灵奖获得者曾提过，尽管她喜欢这项工作，但并不是无时无刻都在想工作，除此之外她依旧有大把的闲暇时光，正是在闲暇时，一些潜意识的思考给她带来了许多灵感。我认为这也是属于女性的特质。

**刘铁岩：**我觉得对于研究员或者是从事科学研究的工作者，最主要的一个特质应该是对世界的好奇和对一些事物的敏感。因为我们只有有这种洞察力，发现一些未知的可能性，才可能比别人更早地去提出一些全新的研究思路，做一些引领性的研究。从这个意义上讲，女性真的是很有优势，因为女性在感知世界方面比大多数男性更敏感、更细腻，也能够对很多事物有一些深入的洞察。我的同事里也有许多优秀的女性研究员。当然，不仅仅是局限在我们公司，整个学术界都有很多非常优秀的女性研究员，她们展示出了强大的洞察力、敏感性和好奇心。我相信，今天很多在线的同学也都有这样的潜质，也希望大家能够多把自己的时间投入到科学研究工作中。

**曹婷：**我非常同意铁岩的观点，我再补充一点就是勇气，我觉得女生可能会因为家庭或社会影响而不敢去挑战，但我们其实不用这样想。我曾经也受这样的思想影响，使得我一开始接触 ACM 编程比赛的时候会觉得自己缺乏经验，不像很多男同学一样从小就接触计算机。但到我真的去参加比赛的时候，我才发现其实这件事并没有那么难。所以我希望大家要有勇气去挑战。

**龚文君：**做一名工程师和 PM，我认为最重要的是两点，一是对客户的同理心，二是要有成长型的思维。同理心能帮助开发出一个真正符合客户需求的产品，同时这也是女生的优势。因为女性往往可以更加善解人意地去理解对方，能够站在对方的角度去思考问题。女生面临的最主要的挑战是社会和环境带来的各种偏见和自信心不足导致的行动力缺乏。在这里就能体现成长型思维的重要性。希望大家可以明白，没有什么事情是不能改变的，永远都有 Plan B，一定要有自信心，要勇敢地去尝试。

**观众提问 2：**如果想为未来职业发展做规划，应该怎样一步一步地明晰学习路径，培养技术积累，以及找到适合自己职业发展的道路？

**刘铁岩：**我觉得其实不管是男性还是女性，职业发展道路都是非常个性化的，就是每个人都有适合自己的道路，需要不断的尝试，所以并没有一个普适的规划。比如，我们作为过来人告诉大家走哪条路比较好，但其实没有那么简单。但是，我们是有一些经验可以分享的，比如用什么样的方法可以让你逐渐找到适合自己的职业方向。

首先就是勇气和自信心，要敢于打破社会上存在的偏见。比如可能有人认为是女性不适合学理工科，更适合人文学科，或者代码能力不够强等。但其实我们仔细观察的话，就会发现有很多代码能力很强、逻辑思维超群、数学功底也很过硬的女性。

其次，因为我是一个人工智能学者，所以我想借用一下我们领域的机器学习方法来说明。比如强化学习，强化学习的基本观点就是要不断地去和环境交互，进行持续的学习和调优，让自己变得强大，然后去最大化长期回报。这个过程中很重要的一个思想就是探索和效用的平衡。举个例子，在职业早期，我们就需要去打开思路，去广泛涉猎，去探索未知。只有通过多和别人交流和请教，进行各种各样的尝试，我们才能获得对世界的充分认知，获得对不同学科的了解。而当我们积累到一定程度的时候，我们就需要学会反思。我们需要根据自己尝试的结果去思考，我们怎样才能更好地聚焦一个相对较小、也更适合自己的领域去做深挖。

而这种思路和我们行业广泛流传的 T 字形人才结构是不谋而合的——既需要拥有广泛的视野，也需要在某一个相对较小的领域有非常深入的探索和挖掘，使自己能够与众不同。这就是一个不断调优的过程，我们要敢于超越自己，甚至否定自己，跳出舒适圈，不断从一个局部最优向全局最优去调整和攀登。我们当然也可以不用一定去成为 T 字形人才，我们可以去成为  $\pi$  字形人才，这都是可能的。所以职业发展是一个非常个性化的过程，只要大家有一个方法论，我相信大家都能找到适合自己的发展方向。

**金芝：**我很认同刚刚铁岩的分享，我也想补充一下我的观点。我认为，每个人都要基于目前已有的才能继续向前奋进，所以需要去思考自己已有什么东西，例如性格、知识面或是环境，这就会使你做出不同的选择。另外，确实每个人的职业发展道路都是非常个性化的，加之计算机领域变幻无穷，你会好奇的问题也多种多样，所以你必须要足够的好奇心、坚持力和专注度，才能找到适合自己的成长道路。

**观众提问 3：**针对计算机相关领域，未来有哪些方向和领域是值得同学们去关注和探索的？

**刘铁岩：**这是个特别好的问题，但我不想给大家一个确定性的答案，因为我觉得预测未来总是不靠谱的。我可以就这个问题和大家做一些深层次的交流。例如有关互联网寒冬或是人工智能领域寒冬的说法，这种担忧在历史上是层出不穷的。人工智能领域也确实经历了几起几落，很多从业者也经常用患得患失来形容自己的心态。

这些讨论背后其实是一个择业观的问题，也涉及到我们对于所在行业一个使命感的问题。为什么这么讲呢？这是因为我们在选择行业的时候，总是有几个不同的选择：可以选择投入到当下最热门的行业里面，享受今时今日的红利；也可以敏感地去洞察刚刚兴起的行业，作为前几波弄潮儿去享受明天的发展机会。当然，我们还可以根据自己的理性判断，根据自己的知识背景，根据自己对行业的认知，去孵化下一个热点。

在这三种选择里，我会认为第三种选择更充分体现了我们的一个使命感，更有可能让我们缔造出一个原来没有的高度，因为你所

孵化的行业方向往往可能是今天没有多少人关注的，是一个冷门的方向。而这就给了我们一个反思和创造的机会。

另外，我想提醒大家的是，我们从事的行业也是分层的。以互联网领域为例，互联网描述的是一个场景，是一个具体的行业，是属于技术分层中比较靠上层的。而计算机技术则是技术分层中更偏底层基础的基石。当社会变化或者行业变迁时，最容易受到影响的就是上层的这些场景。但无论场景怎么变化，对底层技术的需求是相对稳定长久的。所以我从这个角度也很呼吁同学们，你们要加强自己在这些底层技术方面的投入。不管行业如何变化，只要你掌握的技能是基础性的，是本源性的，那么你对社会行业变化的抵抗力就会很强，你的职业就会一直比较稳定、鲁棒地发展。

**观众提问 4：**女性应当如何体现领导力，以及如何在学业道路和职业发展的初期慢慢培养自己的领导力？

**金芝：**我最深的体会就是身体力行。首先就是要做最好的自己，得到大家的尊重和认可，这样大家才能信任和跟随我的决策。当然了，我也要切身地去执行我自己的决策，慢慢地与周围的人建立起信任度。

**曹婷：**首先就是自己要做榜样，这就要求我硬实力和软实力都要过硬，不仅要有扎实的知识储备和科研成果产出，还要与产品部门和高校学者维持一个良好的合作关系。其次就是作为领导要非常关心每个员工的成长，去帮助他们，使得团队能够凝聚成一个整体去发挥其影响力。

**龚文君：**在我看来，领导力很多时候可以转化为影响力。从职业上来看，影响力可以理解为是对商业和产品的影响，而这往往也需要依赖个人对身边人的影响力。通过积极带动身边的人加入进来，我们才能不断地去创造更多更好的产品。这同时也要求我们要不停地鞭策自己，能够跟身边的人共同进步。

**观众提问 5：**我是一位从交互设计跨专业到机器学习的小白，想问一下各位老师，我应该怎样克服自己偶尔的自卑心理和盲从感？

**刘铁岩：**我觉得不管从哪个领域跨到机器学习其实都差不多，因为我们人生会有很多选择的机会。有可能需要去调整方向，有可能进入一个新的环境，而这个自信来源于你的积累，也来源于你的勇气。

有的时候我们会受刻板印象影响而觉得某类人就不适合做某些事情。我们先不去说这种说法是不是过于主观，至少它不科学，因为我们知道每个个体都会有很大的不同。以女性群体为例，人们或许对女性群体会有某种认知，但这些认知其实都只是描述了一

个分布，大家都知道，任何分布都有均值，也有可能有很大的方差。换言之就是千人千面。

所以不管周围的人怎么看待我们，我们每个人都应该活成自己的样子，而不是别人眼中的样子。你要有自信心去突出自己的长处，发挥自己的优势，要成为那个拉大分布方差的人，甚至是通过我们自己的努力去改变整个分布的人。这样我们才会活得更自信。

所以我就想回应这位同学，你要去发掘自己身上的长处，要能够不断勇敢地进行尝试，我相信你一定会做得非常好的。

**曹婷：**其实我自己原来也不是做 AI 相关的，我原来是做高级语言实现的，像 Java 虚拟机等。所以我刚开始转到 AI 的时候，也给了自己一个学习的时间，不断地和身边的同事同学交流，去发现他们身上的亮点，站在巨人的肩膀上。我觉得是要接受这个学习时间，要有耐心，给自己一个逐渐上手学习的过程。

**金芝：**关于如何克服自己的自卑感，我认为最好的方法就是论事不论人。论事不论人指的是不要去想别人批评了我或者别人觉得我的想法太幼稚了，而是将关注点放在自己知识面的扩展和能力的提升上。信心是建立在自己成长的过程中，而不是建立在外界的夸奖上。要学会正确地去认识自己，并把外界的声音和自己的判断去做一个切割。

## Ada Workshop 2022 全场回放视频

扫描二维码观看

回放视频已上线 Bilibili “微软中国视频中心”



## 科研中遇到迷茫困惑如何向前一步？如何在科研中发挥女性优势？

“科研过程中遇到迷茫困惑如何向前一步？”“如何在科研中发挥女性优势？”“如何游刃有余地快速切换研究方向？”在 Ada Workshop 2022 的“女生职业发展专题分享与讨论——研究员专场”中，来自微软亚洲研究院的优秀女性研究员们分享了她们的科研之路，就同学们关心的科研方向、道路选择和成长困惑等问题进行了深入探讨。



**张婷**  
微软亚洲研究院  
主管研究员

**邓攀**  
微软亚洲研究院  
主管研究员



**陈琪**  
微软亚洲研究院  
高级研究员

**主持人：徐勇**  
微软亚洲研究院  
主管研究员



2022  
Ada  
Workshop

**徐勇**：大家都知道，科研并不是一条容易的道路，在这个过程中会遇到许多问题，面临些许困境。第一个问题想问一下三位研究员，当在科研道路上遇到迷茫或者困难时，应该如何去应对？

**张婷**：我觉得针对这个情况可以用一句话来概括，就是静下心来做事情。具体来讲，主要是做到以下三点：第一点是接受自己。在遇到困难或者感到迷茫时，我认为很重要的一点就是要先去接受这一切，去接受自己。你可能需要跟自己谈一谈，你可能会意

识到自己也许并没有想象中的那么聪明，然后需要你以一个乐观积极的心态去面对和接受自己。

第二点是相信自己。虽然你可能并不是最聪明、最有能力的那一个，但是要相信自己付出足够的时间和努力，一定能达到一个不差的结果。因此，在接受自己后，很重要的一个过程就是相信自己，重新塑造自信心。

第三点是要专注自己。这意味着要专注于目前手头正在进行的工作，全身心地投入到每一天的工作进展中，而不要过度在意别人的评价。尽管最终不一定会成功，但无论成功还是失败，只要你付出了精力、时间和努力，你最终都会从这个经历中成长起来，能够更好地认识自己。

**陈琪**：其实我之前也遇到过低谷，然后也干过一些不太成熟的事情，比如换方向之类的。我觉得最主要的一点还是要及时调整自己的心态，让自己尽量不受到外界的干扰，例如身边的同学太优秀、其他人进展的都挺顺利、老师给了一定的压力等等。遇到困境时首先可能需要你静下心来想想，这个进展到底遇到了什么问题，有没有可能解决，应该寻求什么样的帮助。

另外，我认为多增加沟通的机会是很重要的，因为有些时候一个人钻牛角尖并不一定有思路。也许有些同学不太好意思和导师沟通，那你就试着与周边的师兄师姐、师弟师妹们多沟通，甚至可以和其他组的老师沟通。他们可能专注于不同的领域，又或者他们之前也可能遇到过类似的问题，这些沟通有助于你去拓展视角，扩宽知识储备，从而更好地跨过困境。

**邓攀**：在读博或科研过程中遇见低谷是一件挺常见的事情。我认为可以从长期和短期两个方向去解决这个问题。长期要摆正自己的心态。大家刚开始读博的时候，尤其是我们生物学专业，可能都觉得自己能得诺贝尔奖，然后读着读着逐渐发现自己根本就不行。所以首先要做到的就是接受失败，以及接受“失败是一种常态”的事实。一旦接受这种常态后，你再遇到一些失败的结果，内心就能平静一点。

短期则主要有两点，一个是陈琪刚刚说的多和大家交流。因为你和大家聊天就会发现，其实每个人都有过这种挫败和困惑，这并不是你一个人的问题。另一个是可以尝试发展一些别的爱好，不要让自己去钻牛角尖，沉浸在“自己为什么不行”这件事情上。比如，我在读博期间就突然觉得空手道和攀岩让我特别快乐，能够让我不受科研压力的困扰。后来有一个读心理学的朋友跟我说，这就是进入了一种正念的状态。“正念”这个词大家可能都听说过，但其实要想直接达到这个状态是挺难的。所以你可以去找一些自己喜欢的爱好或其他的活动来改善一下自己的想法，从而帮助自



已尽快从这种困惑里走出来。

最后稍微补充一点，其实我觉得如果你做科研的话，压力是会一直伴随着你的。我身边的许多朋友虽然都已经非常成功了，但还是会和我们吐槽自己科研上的压力。所以科研确实是一条相对孤独的道路，面临的压力也很大，你必须相信自己所做的事情，才能够坚持下去。希望大家有这个预期之后，也可以尽早地去调整自己的心态。

**徐勇：**女性角色是否给你们的科研之路带来过什么影响？

比如换研究方向、年龄焦虑、家庭方面的影响等。

**陈琪：**因为我本人博士毕业后就在微软工作，微软给大家提供了一个相对公平的环境，所以我其实没有遇到太多与性别相关的问题。之前在参加活动时，也遇到过有同学担忧关于女性需要平衡工作和生活的问题。我认为，当我们在说女性在这个行业遇到问题的时候，也要考虑到女性在这个行业里其实也有很多优势。

首先，女性的抗压能力普遍较强。其次，女性很擅长去做多线程的工作，可以同时协调好几件事情。另外，女性比较细心，往往能够发觉工作中的一些潜在不足和缺陷。所以其实女生在这个领域是有优势的，不用太担心。

另外，说到家庭和工作的平衡，这其实还是刚才说的专注的问题。Work hard, play harder。比如，我在陪家人的时候我就好好去陪家人去玩，然后在工作的的时候就专心工作。这样才能达到最高效率，也不会受到其他事情的分心。

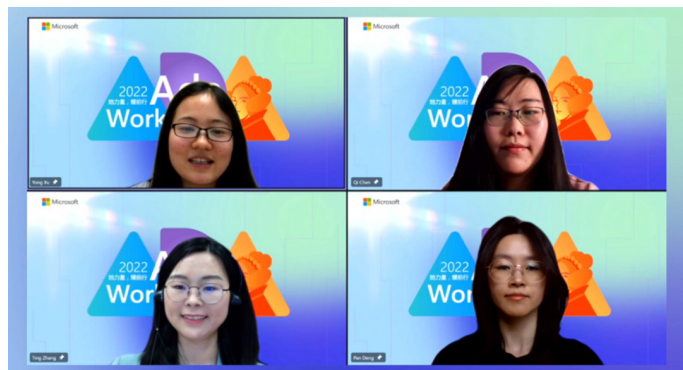
**邓攀：**我想分享一下我自己的经历。我是生物专业出来的，我们专业有一个很有趣的现象。我们本科的时候是女生比男生多的，当时的男女比是 0.72，即 72 个男生和 100 个女生。然后读博的时候其实也是女生多。我之前参加了一些在校博士生的学术交流活动，发现里面有非常多优秀的女生。但是到了开独立实验室的阶段，就发现男女比发生了一个巨大的变化，你很少会见到优秀的女性科学家。为什么 STEM 领域的女性这么少？学术界也有很多的讨论，大家还是会觉得这与社会层面根深蒂固的成见分不开。所以从这个角度考虑，女性在成长过程中或多或少都会受到社会偏见的影响。

就我个人而言，我还是比较幸运的，一路走来没有受过太多性别上的影响。但生活中我也遇到过亲朋好友的“关心”，当我刚打算去读博士时，就有人和我讲，女生为什么要读博士呢，赶紧结婚嫁人就好。所以，我建议女生们尽量亲近适合自我成长的环境，比如，微软有专门鼓励女性获得更多职业机会的一些平台，为女性的成长提供了一个更加健康舒适的环境。

另外，我想说一下我对年龄焦虑的想法，这也是我出国后最大的一个改变。我在国外求学时，发现很多同学都是三四十岁的，他们有些是当过兵再回来读书的，有些是生了小孩又过了好几年才来读书的，还有一些之前是学别的专业甚至是玩乐队的。这些经

历使我开始看淡年龄对我产生的影响。举一个例子，生物领域有一个非常出名的华人女科学家叫庄小威。她本科和博士学的是物理方向，直到博士后才转到了生物方向。她在一次获奖感言中说，她博士后进到实验室时甚至连一些最基础的生物学概念都不是特别清晰，但不可否认的是，她现在在这个领域做得非常出色。所以我认为，年龄完全不是一个限制因素，只是社会和你自己戴上的枷锁。只要你勇于探索，敢于突破这些限制，你还是会有很多不同的可能性的。

**张婷：**我觉得在工作上的表现还是要归结到个人上的，而不是去贴上一个女性的标签，比如，并不是说你细心是因为你是女性，同样也不是说你是女性所以你一定不适合这个领域。尽管大多数女性都会经历生育的过程，从怀孕到哺乳期会需要花费很多的时间和精力，这在短期上确实会对工作和事业产生一定影响，但从长期来看可能不一定，我认为还是要看女性的个人选择。希望大家可以根据自己的优劣势自由地去做一些选择，而少受到社会的刻板偏见影响。



**观众提问 1：**数学背景不好应当如何提升自己？以及数学大概要学到怎样的程度会比较适合从事计算机领域？

**张婷：**关于第一个问题，我觉得首先可能需要判断为什么自己数学背景不好。比如，可能是哪一门学科不太好或者哪一门成绩不太好，那可能需要有针对性地找一些相关的书籍去学习，或者和师兄师姐多沟通交流。

关于第二个问题，如果你是在职场上的话，可能更多的是看你自己在现在所要做方向需要什么样的数学基础知识。比如，你如果要做科研的话，可能还需要一些优化理论的知识，一些矩阵或线性代数的知识等。

我认为，很难在做事情之前做好完全充分的准备。也许更多的是你在实际情况中发现需要什么，然后再努力去补这一块的知识就好。

**观众提问 2：**如果在交流的过程中发现对方盗用了自己的想法，该怎么办？有没有遇到过类似的情况？如何在交流的过程中去推动后续的合作？

**陈琪:**这个事情确实有可能会遇到,但概率比较小。我还是觉得,不要因为这样的原因而不跟别人交流想法,这会限制你自己的发展。

关于第二个问题,我认为还是要自己积极主动。就像我参与了很多项目,但很多项目其实都是我主动找的。当我发现我们组有人在做的项目很有意思的时候,我就会主动去找他/她聊。等别人邀请你去做一个项目其实是比较难的,但你主动的话会相对容易很多,可能别人也正好想多找几个人、多几份力量,所以不要担心会被拒绝。

**张婷:**我想补充一点。除了刚刚陈琪提到的主动去找人交流,我觉得平时也可以多参加一些活动。无论是你感兴趣的学术分享会,还是交流的 Talk,你都可以积极去参与。你可以通过这些活动去建立联系,互相混个脸熟,然后你可以去了解他/她的方向,之后再去找人合作也会更加顺畅。

**观众提问 3:**如何才能游刃有余地在多个研究方向之间快速切换,同时还能高效地完成多项试验任务?

**陈琪:**我之前做过的很多方向大部分其实都是基于兴趣,我会去找相关资料,然后去跟相关的人聊,慢慢地对这个方向有一定的背景了解之后,再去看论文深入掌握。当然,每个方向都不是我学一两天就能完全掌握的,还需要花时间去认真学习并掌握相关研究的历史、现状和未来。所有这些都需要时间的沉淀,但首先最重要的肯定还是兴趣驱动,就是你得对这个东西感兴趣,那样你才会有动力去做这些事情。

另外,当面临研究资料少、一个人做科研比较孤独的情况时,我建议可以寻求合作,可以积极主动地与身边的同事交流你的想法。这样就能有更多的人加入你的研究团队,也可以多些人帮你把关,从而更容易获得一些新的突破。

**观众提问 4:**生物信息这个方向的前景如何?

**邓攀:**现在生物学研究在往偏计算的方向运行,也可以看到越来越多的实验室会把计算工具运用在他们的一些研究项目里。在组学研究兴起之后,包括单细胞组学都会有非常大量的数据产生,这导致我们必须依赖于一些计算工具去处理和理解这些数据。所以我认为如果以生物方向来说的话,偏计算或者信息的一些方法肯定在未来会占到越来越大的比重。不过整体来看,我认为选择工作还是要看自己的兴趣。因为生物信息目前主要还是会做一些工具和 pipeline 方向的工作,如果你不感兴趣的话,即使这个领域未来发展得再好,你在其中也无法获得自我实现感。所以还是要看你自己最想要做哪一个方向,每一个方向都可以创造出不一样的价值。

**观众提问 5:**如何意识到自己所做的事情对自己未来的科研道路是正向的还是可能走偏了?如何看到自己现在的方向是否正确?

**邓攀:**其实我自己是没有这种觉察的。说实话,我读博的决策并不是源于学术追求,只是当时一个从众的选择。等我读到博士之后,我才在想我为什么要读这个博士,以及我到底想做什么。一开始我完全不知道自己未来要往哪里走,在这种情况下,我做的每一步尝试,都是源于兴趣。这些尝试至少并不是一件非常负面的事情,至于是不是绕了弯路,我当时并不能够判断。

我并不认为人生要有一个固定的目标。也许站在现在的这个节点看,你会觉得你某一步走歪了,但是再过五年或者十年,或许你的方向已经发生了变化,当初你感觉走的弯路有可能是帮助你最大的那一段经历。我觉得这就是人生的一个奇妙之处,只有在这种探索之中,你才能真正知道自己想做什么,也能发现许多乐趣。即使大家都不建议你去做,或者短期之内看不到一些未来的曙光,但你还是愿意去做。如果你能找到这样一件事情,其实就已经十分幸福了。

**观众提问 6:**想问一下三位研究员,你们平时如何保持自己的研究水平?

**张婷:**我认为需要不断地学习去保持这个研究水平。首先就是要保持好奇心,经常去看一些新的领域和新的研究方向。这是一个内部的刺激,好奇心的刺激,另外还有一个外部刺激,就是多跟人交流,在互相交流中发现自己的不足之处,驱使自己查漏补缺。

**邓攀:**说一点我最近的想法吧。我最近也在看很多不同的方向,毕竟生物是一个特别大的方向,我们做计算生物也是想开拓一些不同的领域,所以我会看一些发表在高水平期刊(如 Nature、Science 等)上的前沿研究。我发现这些研究的视野非常不一样。这些研究看多了之后,你整个人的研究品位或者一些想法都会发生改变。

**陈琪:**我也认为是要保持好奇心,去关注一些新的领域还有自己感兴趣领域的新工作。也可以多听同事的分享和研究工作的介绍,尝试去挖掘一些新的东西。另外,我认为科研水平也包括写作能力和讲故事的能力,在计算机领域这些能力也很重要,你需要让大家能够接受你的想法。针对这一点,我认为还是要多向高水平的学者学习,或者是和同事们互相学习。举个例子,可能一些同学写论文都是白开水般的介绍,但当你和论文写得好的人多交流后就会发现,论文不能只有介绍怎么做,还需要去思考为什么这么做。你不能只告诉大家这么做是对的,而是要告诉大家为什么这么做是对的。我认为这是很多高水平研究人员的特质,这种思考可以刺激你去不断地修正你的想法,促进你去深思研究的意义,并能够帮助你去发掘更多的研究课题。

# 浙大联合微软亚研院发布视频识别新方法，可对视频逐帧识别且无需数据标记，或可用于手语翻译等

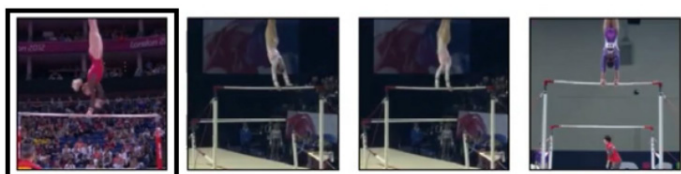
来源: DeepTech 深科技

目前，人工智能在视频理解上已经有广泛应用，例如用深度学习帮助视频分类等任务已取得显著成果。不过当下，有关动作表征学习的各种架构主要为识别视频的全局特征而设计。

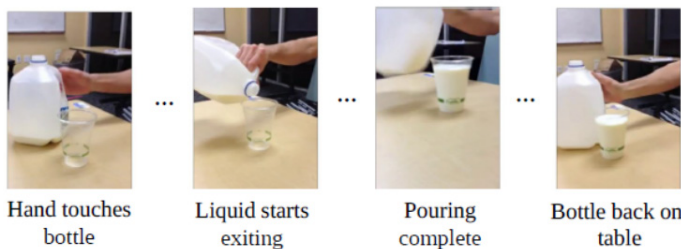
然而在实际应用中，对于视频的逐帧识别也有着强烈的需求，例如有时希望借助人工智能完成视频对齐、手语翻译、机器人模仿学习等操作。这就对算法提出了更高的要求，能对长达数百帧的长视频建模，与此同时，对该长视频进行逐帧表征识别而不仅是全局特征。

最近，浙江大学计算机辅助设计与图形学 (CAD&CG) 国家重点实验室联合微软亚洲研究院共同研发了一种新的名为“对比动作表征学习” (contrastive action representation learning, CARL) 的框架，通过自我监督的形式对长视频等内容的逐帧动作表征进行学习以及识别。而且，该方法并不需要事先对视频进行标记。

为对该方法进行评估，研究人员通过该方法对目前主流的三种视频数据集 FineGym、PennAction 和 Pouring 进行了实验。实验结果证明，通过该方法在各方面的表现皆优于之前的方法技术，特别是下游细粒度动作分类表现尤为明显。相关论文以《基于序列对比学习的长视频逐帧动作表征》为题在 arXiv 上发表。



Query Top-3 Retrieved Results  
(a) Fine-grained frame retrieval on FineGym dataset.



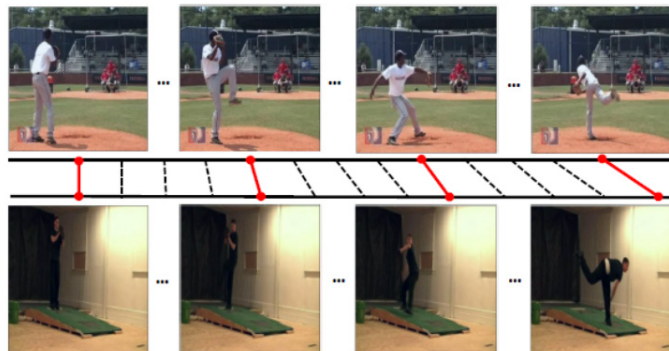
(b) Phase boundary detection on Pouring dataset.

用 CARL 框架对不同数据集中的视频进行逐帧表征学习的实验结果

(a) FineGym 数据集中的细粒度帧检索

(b) Pouring 数据中的相界检测

(来源: arXiv)



(c) Temporal video alignment on PennAction dataset.

用 CARL 框架对不同数据集中的视频进行逐帧表征学习的实验结果

(c) PennAction 数据集中的时间视频对齐

(来源: arXiv)

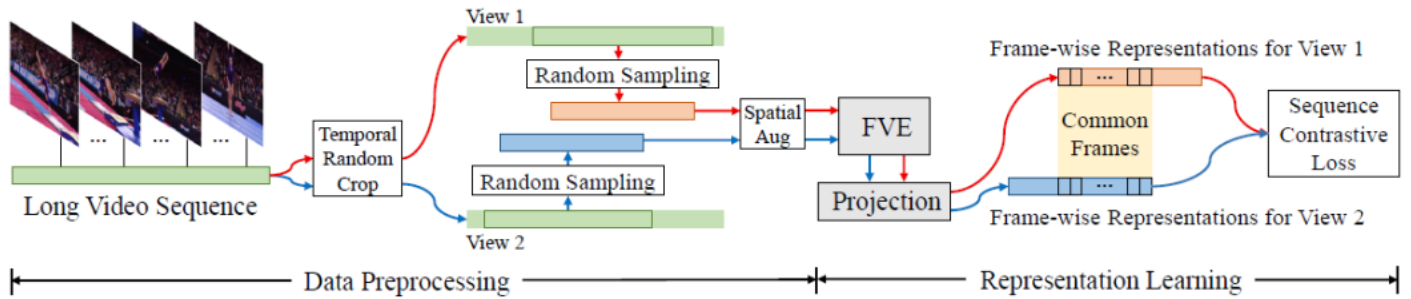
虽然在此之前，也有其他方法通过监督学习对视频进行逐帧地表征学习与识别。但是，这些方法大多需要对视频中的动作边界或阶段边界进行手动标记，在大规模的数据集中进行这一步骤十分耗时，甚至有些不切实际。因此，这些方法很难在现实场景中得到广泛应用。

该团队此次发布的 CARL 框架，并不需要对视频进行标记。该框架受对比表征学习最新进展的启发，通过自我监督的方式对长视频中具有时空上下文相关的信息进行逐帧表征学习。

CARL 框架的工作原理分为数据预处理和表征学习这两个步骤。在数据预处理中，系统首先会通过一系列时空数据增强的方式，构建该视频的两个增强视图。接下来，再输入该增强视图至帧级视频编码器 (frame-level video encoder, FVE) 进行处理，这一步可以提取出其密集表征。另外，FVE 还附加了一个小型映射网，通过该映射网可以生成潜在嵌入的多层感知器。

通常，一段视频中在时间上相邻的两帧可能十分相似。因此，该团队做出合理假设，即两个增强视图的相似性分布应按照高斯分布。而基于该假设，他们通过序列对比损失法来解决该问题，也就是说研究人员对逐帧表征进行了优化。





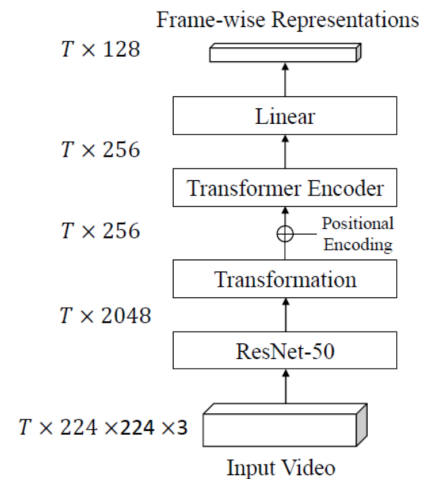
对比动作表征学习 (CARL) 框架的概述 (来源: arXiv)

在数据预处理这一步骤的具体过程是：对具有帧长为  $S$  的训练视频  $V$ ，系统以一系列时空数据增强的方式来创建两个帧长为  $T$  的增强视图。时空数据增强的方式既有时间数据增强，也有空间数据增强。在时间数据增强方式中，系统首先将训练视频进行随机性的剪裁，从而生成两段帧长为  $[T, \alpha T]$  的片段，在这里可以通过  $\alpha$  来控制剪裁的最大长度。

接下来，系统对剪裁后的视频进行随机性的采样，采样帧数为  $T$ ，该采样得出的结果就是两个视频序列，分别为  $V1$  和  $V2$ ，系统对  $T$  的默认值为 240。如果视频的帧数小于  $T$ ，那么在剪裁之前还会有空帧填补这一步骤。之后，还会针对  $V1$  和  $V2$  应用不同种类的空间数据增强方式进行处理。这些方式包括大小调整、水平翻转、高斯模糊等。

而在表征学习这一步骤中，系统引入 FVE 对时间上下文进行建模。FVE 的工作过程具体为：首先通过一个 2D 网络（如 ResNet-50 等）提取出 RGB 视频序列的一个大小为  $T \times 224 \times 224 \times 3$  的空间特征。然后，使用一个转换器将提取出的空间特征投影到一个大小为  $T \times 256$  的中层嵌入上。之后，该嵌入会被编码，并被进一步输入到编码器中进行建模。最后一步采用了线性层，并得出视频的逐帧表征。

该团队还通过将该框架应用于 PennAction、FineGym 和 Pouring 这三个数据集上，来对框架的性能进行评估。结果显示，CARL 框架在这几种数据集上的测试结果都优于此前的最先进方法。



帧级视频编码器 (FVE) 的概述 (来源: arXiv)

## 相关链接：

论文链接：

Frame-wise Action Representations for Long Videos via Sequence Contrastive Learning  
<https://arxiv.org/abs/2203.14957>

GitHub 链接：

[https://github.com/minghchen/CARL\\_code](https://github.com/minghchen/CARL_code)



## 周礼栋

微软亚洲研究院院长

“二十多年来，微软亚洲研究院始终秉承开放、积极的心态，致力于打造自由、平等、可持续的科研协作环境，让分工、协调、合作链环上的每个人都成为新的发现与贡献的核心主体，为各种创造性想法的星星之火提供形成燎原之势的催化剂。

一个创新型组织的成长是不断拓展视野并承担更大社会责任的过程。微软亚洲研究院从创立伊始就持续与国内外计算机科研机构展开深度合作，携手进步，共同发展。在面对当下可持续发展、碳中和、医疗健康等人类社会亟待解决的关键问题时，微软亚洲研究院将守正创新，践行所有有利于激发创新力的原则，大胆接受和改造各种新的范式，与各界伙伴共同推动计算技术的跨界融合发展。”

## 关于微软亚洲研究院

微软亚洲研究院是微软公司在亚太地区设立的研究机构，也是微软在美国本土以外规模最大的一个。从1998年建院至今，通过从世界各地吸纳而来的专家学者们的鼎力合作，微软亚洲研究院已经发展成为世界一流的计算机基础及应用研究机构，致力于推动整个计算机科学领域的前沿技术发展，并将最新研究成果快速转化到微软全球及中国本地的关键产品中，帮助消费者改善计算体验。同时，微软亚洲研究院着眼于下一代革命性技术的研究，助力公司实现长远发展战略和对未来计算的美好构想

微软亚洲研究院一直努力营造奋发、进取的科研环境，倡导对技术进步怀有远大抱负，推崇大胆创新以及富于冒险的极客创新精神。研究院鼓励研究人员树立长远眼光，加强与外界的交流，在第一时间接触世界领先的研究成果，对行业发展、技术趋势、及消费者需求高度敏感、迅速反应。微软亚洲研究院强调成员之间的相互信赖、相互尊重与开放合作，并承诺与高校和科研机构开展持久而有效的合作，促进协作、激发创新、推进教育，变梦想为现实！

## 微软研究院全球布局





微信

知乎



电话：86-10-59178888

网址：<http://www.msra.cn/>

微博：<http://t.sina.com.cn/msra>