

主办 CCF 计算机视觉专业委员会

COMPUTER
VISION
NEWSLETTER

CCCF 计算机视觉 专委会简报

02 2022

总第 32 期



CCF 计算机视觉
专委会

COMPUTER VISION NEWSLETTER



计算机视觉专委会 简报

2022 年第 02 期

总第 32 期

主 办 编委会

CCF 计算机视觉专业委员会



CCF 计算机视觉
专 委 会

/专委动态/

荣誉主编 **王 亮** 中国科学院自动化研究所
主 编 **马占宇** 北京邮电大学
执行主编 **李实英** 上海科技大学
主 编 **毋立芳** 北京工业大学
编 委 **黄 岩** 中国科学院自动化研究所
任传贤 中山大学
杨巨峰 南开大学

/科技前沿/

主 编 **王金甲** 燕山大学
编 委 **储 珺** 南昌航空大学
崔海楠 中国科学院自动化研究所
魏秀参 南京理工大学

/委员风采/

主 编 **余 焯** 合肥工业大学
编 委 **刘海波** 哈尔滨工程大学
赵振兵 华北电力大学

/学术资源/

主 编 **李 策** 兰州理工大学
编 委 **樊 鑫** 大连理工大学
贾 同 东北大学
沈沛意 西安电子科技大学

/海外学者/

主 编 **金 鑫** 北京电子科技学院
编 委 **刘帅奇** 河北大学
张汗灵 湖南大学

/视界专访/

主 编 **张军平** 复旦大学
编 委 **贾熹滨** 北京工业大学
明 悦 北京邮电大学

CONTENTS

简报目录

| 专委动态

- 04 CCF-CV 走进高校系列报告会
- 05 CCF-CV 视界无限系列研讨会
- 11 CVPR 2022 论文分享会、主旨演讲及圆桌论坛
- 13 2022 年度 CCF-CV “中科视拓 Seeta 学术新锐” 评选活动

| 科技前沿

- 14 基于多相机系统的全局式三维建模方法
- 17 对抗学习：消除对抗噪声以提高对抗鲁棒性
- 22 关于构建遥感图像解译数据集的探讨
- 27 ICLR 2022

| 委员风采

- 32 上海交通大学严骏驰副教授访谈
- 35 委员好消息

| 学术资源

- 36 手绘草图领域开源代码
- 39 ADNI 数据集
- 43 好文推荐

| 海外学者

- 46 征文通知

| 视界专访

- 47 同济大学宣国荣教授专访

CCF 计算机视觉
专委会

 CCFCV.CCF.ORG.CN

 CCFCVN@GMail.com

CCF-CV 走进高校系列报告会

第 112 期 华北理工大学

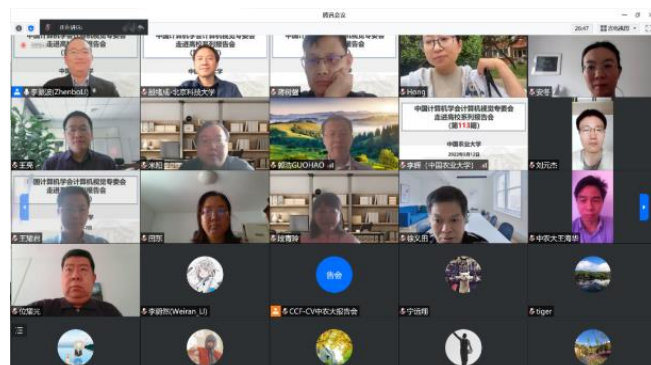


2022 年 4 月 20 日上午，由中国计算机学会计算机视觉专委会（CCF-CV）主办、华北理工大学承办的 CCF-CV 走进高校系列报告会第 112 期活动以线上方式成功举办。本次活动邀请了北京航空航天大学曹先彬教授、中国科学院自动化所王亮研究员、上海交通大学卢策吾教授做特邀报告。华北理工大学人工智能学院李志刚院长、黄晓红副院长担任本次活动执行主席。

活动伊始，华北理工大学副校长李昌存教授致欢迎辞，他首先代表华北理工大学对嘉宾表示了热烈的欢迎和衷心的感谢，然后表示华北理工大学非常荣幸承办第 112 期走进高校报告会，希望在继承往届活动优良传统的基础上，打造多样化的交流平台，共同推动我国相关领域的学术繁荣和产业发展！期待专家们的精彩报告与学术交流分享，并预祝本次报告会取得圆满成功！

最后，华北理工大学黄晓红教授进行了活动总结，再次感谢了三位专家的精彩报告与学术交流分享，同时感谢了线上听众的热情参与和高质量提问，最后再次感谢 CCF-CV 专委会和华北理工大学人工智能学院对活动的大力支持！祝贺本次活动取得了圆满成功！

第 113 期 中国农业大学



2022 年 5 月 12 日下午，由中国计算机学会计算机视觉专委会（CCF-CV）主办、中国农业大学承办的 CCF-CV 走进高校系列报告会第 113 期活动以线上方式成功举办。本次活动邀请了中国科学院自动化所王亮研究员、北京大学彭宇新教授、中科院计算所蒋树强研究员、北京科技大学殷绪成教授做特邀报告。中国农业大学信息与电气工程学院副院长李振波教授担任本次活动执行主席。

会议采取 CCF 计算机视觉专委会 B 站官方账号直播+腾讯会议的方式，B 站人气峰值接近 1000，腾讯会议本校教师参加人数近 100。最后，中国农业大学信息与电气工程学院院长张漫教授进行了活动总结，首先感谢了四位专家的精彩报告与学术交流分享，同时感谢了线上听众的热情参与和高质量提问，最后再次感谢 CCF-CV 专委会、学校和学院对活动的大力支持！祝贺本次活动取得了圆满成功！

责任编辑 毋立芳

第 13 期 视觉 Transformer 与 Attention 的前沿进展与未来趋势

CCF-CV 视界无限系列研讨会

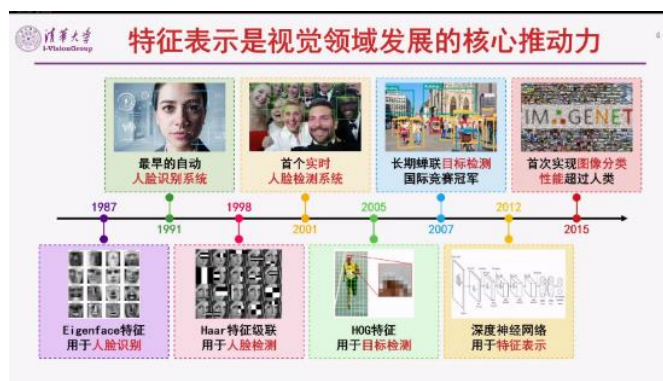


2022年5月17日，由中国计算机学会计算机视觉专委会举办的第13期CCF-CV“视界无限”系列活动——“视觉Transformer与Attention的前沿进展与未来趋势”研讨会在线上举办。研讨会邀请了专委会副主任中科院自动化所王亮研究员致辞，清华大学鲁继文、南京大学王利民、复旦大学张力、南开大学侯淇彬、清华大学国孟昊做主题报告。南开大学程明明教授及以上五位讲者参与了深度研讨。计算机视觉专委会B站公众号对本次会议进行了全程直播，直播人气峰值达到1400+。



2022年5月17日，由中国计算机学会计算机视觉专委会举办的第13期CCF-CV“视界无限”系列活动

——“视觉Transformer与Attention的前沿进展与未来趋势”研讨会在线上举办。研讨会邀请了专委会副主任中科院自动化所王亮研究员致辞，清华大学鲁继文、南京大学王利民、复旦大学张力、南开大学侯淇彬、清华大学国孟昊做主题报告。南开大学程明明教授及以上五位讲者参与了深度研讨。计算机视觉专委会B站公众号对本次会议进行了全程直播，直播人气峰值达到1400+。



鲁继文副教授的报告题目是“视觉基础模型及应用”。基础模型是人工智能领域的研究热点，在计算机视觉和自然语言处理等多个领域中均取得了优异的性能，是视觉监控、自动驾驶、智能终端等重要应用的支撑性技术。鲁继文副教授从模型架构和学习范式两方面回顾视觉基础模型近年来的研究进展，同时介绍清华大学智能视觉实验室在视觉基础模型方面所开展的工作，包括动态稀疏模型、全局滤波模型、球面分形模型、几何敏感模型等，以及它们在目标检测与分割、物体分类与识别、图像与视频检索、三维重建与识别等视觉任务中的应用。



王利民教授的报告题目是“基于注意力机制的视频表征学习与目标跟踪”。视频理解已经成为人工智能研究的热点和难点，其中人体动作识别是视频理解涉及的关键技术之一。王利民教授主要介绍南京大学媒体计算组 (MCG) 在视频动作识别和目标跟踪方面的系列工作。首先，针对视频运动表征与建模，介绍了高效和动态的时序建模模块 (TEINet, TAM, TDN)，在计算效率和建模精度方面取得较优效果。其次，针对视频模型的自监督预训练，介绍了基于掩码自编码器的视频高效学习方法 VideoMAE，验证了 MAE 一种数据高效的 Transformer 自监督训练框架，并且在动作识别主流数据库上取得了优异的识别性能。最后，针对视频目标跟踪技术，介绍了更加简洁的单目标跟踪框架 MixFormer，统一了特征提取和特征融合模块，在 5 个主流跟踪数据集都取得了目前最好的跟踪精度。最后总结和展望了视频理解的发展趋势。



张力研究员的报告题目是“不依赖于 Softmax

的线性复杂度自注意力模型”。Vision transformers 推动了各项视觉识别任务的进展，但是在计算和存储方面都存在二次复杂度。具体来说，传统的自注意力计算需要对特征向量之间的点积缩放后进行 softmax 归一化。保持该 softmax 操作对任何线性化的方法都是一个障碍。张力研究员介绍了不依赖于 softmax 的 Transformer 模型，使用无行归一的高斯核函数来代替之前的点积注意力，并基于此构建低秩的矩阵分解逼近满注意力矩阵。采用 Newton-Raphson 方法来保证近似的鲁棒性并实现近似中涉及的 Moore-Penrose 逆计算。在大规模图像识别数据集 ImageNet 上的实验表明该线性化方法能够显著提升现有 Transformer 模型的计算效率，获得准确性和复杂性之间更优越的权衡。



侯淇彬副教授的报告题目是“如何更好地训练视觉 Transformer”。近年来，视觉 Transformer 推动了视觉领域多项识别任务的发展。在多数视觉任务中，基于 Transformer 的网络结构在性能方面已逐渐优于传统的 CNN 网络。侯淇彬副教授介绍了如何借助 CNN 精准定位目标物体的能力，进一步提升视觉 Transformer 的性能。基于这个动机，提出了 Token Labeling 的训练策略，通过给视觉 Transformer 的每个输出 Token 赋予一个来自 CNN 预测结果的监督信息，可有效提升视觉 Transformer 的识别能力。该工作同样说明基于 Token Labeling 训练策略的预训练模型在语义分割等下游任务中仍有较好的表现。

Panel 实录:

程明明: 非常感谢 5 位老师的精彩分享。Vision Transformer 是目前视觉领域最热点的研究领域之一, 在很多经典的任务上已经取得了成功, 但是我们也能够看到它跟 CNN 相比依然有一些问题或者不足, 比如 Transformer 最开始是针对一维数据设计的, 并不是针对 2D 图像。下面想请各位老师探讨一下与 CNN 相比, Transformer 的优势和劣势分别是什么? 我们先请鲁继文老师谈一谈。

鲁继文: 我觉得 Transformer 这两年这么火或者这么受关注, 根本原因是它刷新了模型性能的 SOTA, 与 CNN 相比, 我觉得它很重要的一点是能够实现多任务的联合统一建模。主要的不足我个人感觉还是对数据的依赖性比较严重。Transformer 很多模型在下游的小数据集上实验还是比较难 work。CNN 发展了有很多年了, 应用范围也很广, 但是它的问题在于面对大规模数据集的时候不适合去训一个大模型, 尤其我们现在这个大模型时代, 大家都希望能够去把大量的数据喂给模型训练, 这也是 Transformer 比 CNN 流行的一个原因, 但它们各有优劣。

张力: Transformer 和它这样一种 Self-attention 的机制已经驱动深度学习快速发展。回顾五年以前, 神经网络就是卷积和池化, 我们以后会觉得这种方式很 naive。那么这种卷积网络的上限并不是深度学习的上限。重新来看这个 Transformer 结构, 可以认为它是一个动态的卷积。那么卷积其实是一个具有很强先验的一种运算方式。Transformer 突破了这个上限, 它通过学习自己乘以自己这样一个模式能够更好地泛化到测试集上, 这也是它比 CNN 更加高效的一个原因。

王利民: CNN 是从 local 到 global 这样一种学习方式, 它局部的一致性会强一点, 全局结构性是依赖于局部的。但是, 像 ViT 这种结构它都是全局的操作, 我觉得他本质上是一个更加灵活的方式。我们可以想象 CNN 是一个箱子来装东西, 他方方正正的。相反, Transformer 可能类似于像一个气球, 这里什么样形状的数据, 你放进去它可能就能跟你贴合的很好。还有就是 Transformer 其实是在提取一些低频的特征, 相反,

Visual Attention Network

Meng-Hao Guo

Tsinghua University
gmh20@mails.tsinghua.edu.cn

May 17, 2022



Meng-Hao Guo (THU) Visual Attention Network May 17, 2022 1 / 22

国孟昊博士的报告题目是“视觉注意力网络”。认知科学中, 注意力机制是指在处理信息时, 人类会选择性的关注重点的部分, 而忽略一些无关的信息。近来, 注意力机制在计算机视觉中得到了广泛的关注, 基于自注意力机制的视觉 Transformer 模型已经在各种视觉任务中取得了巨大的成功。本次报告, 国孟昊主要以计算机视觉中的注意力机制为中心, 重点讨论了三个问题:

- (1) 当我们在谈论视觉注意力的时候, 我们在谈论什么?
- (2) 视觉 Transformer 的成功之道以及它存在的不足之处。
- (3) 一种不同于 Transformer 的新型视觉注意力网络。



在 Panel 环节, 与会嘉宾就“视觉 Transformer 与 Attention 的前沿进展与未来趋势”、“视觉 Transformer 与 CNN 如何共存”、“资源受限的条件下如何进行视觉 Transformer 研究”、“视频相关任务中 Transformer 的现状”等问题展开深入讨论, 并就 B 站直播室观众的问题分享了各自的观点。

CNN 可以抽取一些高频的特征。

程明明：谢谢各位老师的精彩发言。接下来还有一个问题在 B 站直播间出现了好几次，有同学在提问说这种视觉注意力机制在处理点云补全会有一些可能数据量比较大的困难，不知道各位老师对这个有什么看法，鲁老师是不是有做过这方面的工作？

鲁继文：去年我们实验室做了一个点云补全的工作，发表在 ICCV 上。点云补全这个任务，我们为什么可以用 Transformer 来做？很核心的问题是因为我们把这个任务建模成机器翻译的任务。机器翻译，我们都知道从中文到英文，从英文到中文，它实际上是一个转换的系统。点云补全我们把它建模成已知一部分点云，把残缺的点云补出来。实际上，我们也认为考虑到网络的预测能力或者翻译能力，这样去做实际上还是有一些难度的。这种点云数据其实与图像相比或者跟自然语言处理相比，它数据量是很小的，所以还是需要依赖一些已有的模型的架构去做一些初步的或者说做一些特别的表示，表示完之后我们再去做这个点云和点云之间的自注意力，通过 Transformer 去建模，这样可以把那些残缺的点云补全出来。场景点云和物体点云相比，场景点云更复杂，残缺的部分比重更高同时噪音更大，所以目前我们正在看看能不能在这种大场景下能够帮他做一些好的性能提升。

程明明：接下来还有一个问题就是现在学术界出现了很多像 ConvNeXt, VAN 这种比较好的 CNN 网络，那如何看待 CNN 和 ViT 的未来发展趋势？

侯淇彬：我可以先简单聊一下。目前为止大部分视觉任务其实还是比较单一的，就是说他只针对图像数据。但是以人为例，我们做人工智能算法肯定还是希望能够像人脑系统一样去处理不同的信号。目前，Self-attention 机制不管在音频、视频、图像、语言等领域都是有较好表现的，我个人觉得我们需要把不同模型结合起来，以 Transformer 为骨干网络，针对不同任务的结构以 Attention 或者动态路由的形式加入到骨干网络中。比如在处理 2D 图像时，我们可以用国孟昊提出的 visual attention 这样的方法。

鲁继文：现在做视觉技术模型，实际上模型的结合是大势所趋，取长补短。我们经常说 CNN 有 CNN 的优点，Transformer 有 Transformer 的优点，所以取长补短之后一定是有助于提高模型的性能。包括最近谷歌提出的 CNN 模型实际上也都借鉴了 Transformer 的设计思想。我们最近也发现了 Transformer 一个很好的性质。其实 CNN 网络具有比较好的平移不变性，在很多硬件以及优化的很好的情况下，如果进一步把卷积操作改进好，然后再结合 Softmax，可以取得一些更好的性能。

张力：我们今天更多的在谈论注意力机制，但是 Transformer 这个模型除了自注意力机制以外他还有一些其他的特点，比如说他有物体感知的能力。物体感知能力是我们整个视觉最为核心的部分。比方说，我们要做目标检测，最早我们用 selective search, 包括程老师之前也做过 Bing。后来的 RCNN 就能够把这种 proposal 融入到整个深度学习框架当中。通过这种方式，能够产生物体感知。在深度学习框架里面，后来发现这种 Transformer 的 Encoder 能够提升物体的表征能力。然后，去年还是前年有一篇文章，说 Encoder 不仅能够提升我们视觉感知的表示能力，Decoder 也能够重新定义计算机视觉的很多任务，比如目标检测。我们可以通过 Transformer 里面的这种查询机制去做这样的感知。像目标检测、实例分割、pose 等都可以用这种方法来重新定义任务。所以，他除了这种特征表达能力以外，还有一些功能性的优势，能够重新帮助我们怎么去思考。

程明明：还有一个问题是 ViT 主要的成功来自于用 self-attention 这样一种空间的注意力机制，如果说是从注意力机制本身出发的话，各位老师能不能谈一下在未来还有哪些可以进一步挖掘的方向？

国孟昊：我觉得空间注意力可以从两个角度谈这个问题。第一个就是空间的角度，第二个就是注意力这个角度。从空间的角度来看，其实传统方法带给我们的是一种空间的交互方式。比如 CNN 用 3x3, 7x7 这种卷积进行交互，然后就是像 ViT 这种直接全部都是全局的。第二个就是从注意力的角度，我认为其实他是一种聚合

的方式，比如说 VAN 这种也是一种交互方式。从聚合方式的角度来看，其实到底能不能找一种新的信息聚合方式也是非常值得挖掘。所以我认为从空间和注意力两个角度应该是都能继续往下去做的。

程明明：B 站直播间有同学想请各位老师说一说对于视频数据，还有哪些比较值得关注的。

王利民：我觉得视频的问题应该分几个层面来看，现在做视频的工作基本上就在 follow CNN 的解决视频问题思路，核心就是如何把 self-attention 操作从一个空间的操作变成时空的操作，但是这种简单的拓展能处理的视频的长度或者说帧数是比较有限的，所以说这只是一个初级的版本，只能处理几秒钟或者一两秒的视频。这个时候能理解的信息不是全方面的，而是受限的，只能做一些比较简单的运动识别。另外我觉得一个重要的点就是说我们怎么能处理更长的视频，那现在这个确实大家做的比较少，可能还得借鉴 RNN 的那种思路，把一些持续的信息做压缩，通过 Memory 的方式拓展到一个更长的视频乃至一些更高层次的推理。

程明明：还有一个问题是 ViT 训练的过程中需要大量的计算资源，很多学校的学生特别是我们大部分的听众可能没有这么多的计算资源。如果说这些学生也想去这个，有什么建议。比如说，各位老师能不能说一说自己一些比较有代表性的工作大概需要多少计算资源？

鲁继文：计算机视觉对算力的需求比较大，总是感觉算力的增长不能够满足广大同学们对 GPU 需求的增长。我始终坚持一个观点，就是学术界应该做适合学术界干的事，如果硬要去跟工业界 PK 这种数据和算力，我们很难找到自己的位置。所以一定要能够从这种偏 idea 层面去做。比如说我觉得我们去年做了两个视觉基础模型方面的工作，做的还不错。其实最后也就是几台机器就能够把这个事干了。我们说大模型也好，尤其我们谈视觉大模型，其实很多时候是一个系统工程，还有很多东西可以做。很多同学和老师可能关注到的只是模型的基础架构，因为模型架构方面的工作出来之后受关注度比较高。如果没有这个算力，很难找到我们的位置。早期 Deep learning 刚开始出来的时候，那个时候我实验室就一块卡，也做了一个深度度量学习的工作，后来

这个工作的影响力也还不错。所以我坚持一个观点，就是要能够在整个视觉模型的训练、包括优化、推理等里面找到适合我们学术界干的事。

张力：慢慢这两年有一个新兴的方向叫做小型化模型。已经有一些 paper 了，比如说 MobileViT, MobileFormer。都是已经开源的一些 paper，这些模型都是小型化模型，非常小包括我现在那个腾讯的犀牛鸟项目也是在做这个 Transformer 的小型化。我们正在设计一个非常精小的模型，参数量和 FLOPs 都很小。这样一个小模型在 ARM 端去做测试跟 MobileNetV3 去比，速度啊，latency 等都有优势。

程明明：王老师是做视频的，视频是不是对卡的需求更大一些？

王利民：我感觉未来的视觉或者说整个 AI 的发展的可能会逐渐收敛成一个刚才说的基础模型或者大模型，包括跟下游任务的适配。就是说当我们做基础模型时可能得依赖于一些更大的设备。但是当我们没有这样的设备或者平台的时候，可以优先考虑做一些下游的具体任务。关于具体任务，其实视频里面还有很多小的任务大家没有太关注到。因为做小的任务可以用现在的 ViT，比如小样本训练。这一块其实对计算资源需求不是太大。所以说下游任务我觉得是一个很重要的方向。另外一个就是还是需要往更深入的理解这个角度去发展，可以做一些更精细的任务，比如做一些高效的模型。所以说，基础模型我们可以用现成的，然后我们可以做更多下游的任务，我是这样一个观点。

侯淇彬：关于 Transformer 这一块，我刚才也介绍了一个工作，从参数量的角度考虑，比如说最大的模型是 150M，在最后测试的时候是 512*512 这样图像。对于小模型而言的话，比如 ResNet50 这种量级的模型，其实有 8 个 V100 就可以训练了。ViT 不像以前 CNN 依赖 BN，它用的是 LN，因此 4 卡也可以训练。以 V100 为例，训练 30M 左右的一个模型，大概是需要两到三天的时间。如果是训练更大一点的模型，比如说 150M 这种模型，得采用两台 V100，如果是在 512*512 的图像上微调，需要 4 个节点。如果大家资源不是很多，还是建议像之前各位老师说的那样做一下下游任务，比如

显著性检测、Meta Learning、以及其他的模型独立的学习方法。

国孟昊：我们做的 VAN tiny 和 small，在训练的过程中主要是 8 个 3090 显卡，训练两天左右。如果 8 块 3090 也不一定能拿的出来的话。其实可以像刚刚老师们说的那样做一些轻量级的网络或者说做一些其他任务。就是刚刚鲁老师说的那样，要做一些比较原创的东西，因为实验只是验证 idea，如果洞察力够强的话，我觉得实验稍微弱一点也能做一篇非常好的工作，还是要拼 idea，不要去拼计算资源。

程明明：用你的这个 VAN 模型去做一些语义分割的时候，可能需要多长时间？

国孟昊：其实如果去做语义分割的话，VAN 用一张卡就能行，语义分割就比较不耗卡了，包括 Base 这种级别的其实就两张卡就够了，12 个小时之内就能训完，对资源没有要求那么高。

程明明：好的，两张卡和 12 个小时，对部分同学来说已经是是可以接受的了。看看大家还有什么更多的问题吗？咱们在报告的过程中，很多同学提出了问题，部分问题讲者老师已经在 B 站上留言解决了，我们就不再重复。有同学针对王老师这边有一个提问，请问王老师讲讲如何进行有效的长时和短时的特征融合？

王利民：我感觉这个问题还是非常好的。目前还是受限于硬件资源的限制，做常识性的基本上就两个思路：一个思路就是动态采样到一些有一定相关性的片段，然后对他们做一些动态模型来融合，这是一个思路；另外一个思路是不能够一次性的把它放到 memory 里面去，这时候可能得做一个 memory bank，需要把特征存储起来，然后用它来辅助当前的一些短时的信号做决策，这是两种思路。

程明明：好的，有同学在问说 decoder 的这种 query 机制是不是可以做语义分割？

张力：这个思路已经有论文了，比如在全景分割里。

程明明：有同学想请王老师分享一下在做视频训练的过程中对显卡的需求。

王利民：好的，我给大家就举一个例子吧，这个确实消耗很大。我们用自监督学习的方法训练一个视频的 ViT，当时是跟腾讯合作的。我们跑的这个帧数有两个版本，一个是 16 帧的，就输入 16 帧，然后跑一个 Base 的模型，训练可能大概要 2 天到 3 天，然后用的是 32 张 V100。所以说这个确实比较大，然后到后来我们投完稿之后，我们跑了一个 ViT 的大模型，跑的也是 16 帧，用了 128 块卡，然后跑了 10 天，但是那个数据量是到了 50 万左右，所以说数据量扩了一倍。这个开销确实比较大，所以说只能跟企业合作。

程明明：好的谢谢王老师。还有一个很重要的问题，也是大家很关注的问题，大家在训练的过程中，调参还是挺难的，各位老师们在调参方面有什么样的经验可以跟大家分享吗？我看国博这边笑的比较开心，要不然你分享一下？

国孟昊：现在一些训练方法其实都已经写好了，比如说 timm。根据他们的那些默认配置其实不怎么需要调。我一般调的时候就只调一个东西，就是 Drop path rate。

程明明：再次感谢所有的讲者。大家有什么问题的话欢迎联系讲者，谢谢大家，我们今天的活动到此结束。

责任编辑 杨巨峰

CVPR 2022 论文分享会、主旨演讲及圆桌论坛

4月23日，CVPR 2022 论文分享会准时线上举办！为了让更多感兴趣的朋友参与到精彩的分享会中，同时响应疫情防控政策，今年的 CVPR 论文分享会在 B 站“微软中国视频中心”全程进行了直播。

今年，我们很荣幸邀请到本届 CVPR 会议的程序主席华刚博士为大家带来“你和你的计算机视觉研究生涯”的主旨演讲，华刚博士将分享他对于学术人生的理解，以及作为领域的一份子，如何让自己的职业生涯与计算机视觉这一蓬勃的领域共舞。

论文是研究者工作的集萃，与此同时，做出能加深或改变领域理解和工程实践的工作，也是每位研究者的追求。顶级会议最佳论文就是这样的存在！今年的圆桌论坛以“好论文是怎么炼成的？”为主题，马毅、屠卓文、戴玉超、张祥和黄高这几位历年视觉顶会最佳论文的斩获者，分享了他们对于“做好的论文”和“做好的研究”的理解，共话计算机视觉领域的科研淬炼！

华刚：

华刚博士是便利蜂 CTO 兼首席科学家，IEEE Fellow, IAPR Fellow 和 ACM 杰出科学家。华刚博士是计算机视觉和模式识别国际著名专家，其主要研究方向包括计算机视觉、模式识别、机器学习与机器人；当前的技术实践研究重点是人工智能技术在实体零售数字化决策和运营的应用。华刚博士曾任微软研究院首席研究员、斯蒂文森理工学院副教授、IBM 研究院和诺基亚研究院高级研究员。2015 年因其在图像视频人脸识别所作出的贡献而获得国际模式识别

联合会青年生物特征杰出研究员奖。他在国际期刊和顶级会议上发表论文 200 余篇，并拥有 20 多项美国和国际专利，同时还有 20 多项美国和国际专利申请。华刚博士是 2019 年和 2022 年 CVPR 程序主席，以及 2025 年 ICCV 大会主席，并且是 T-PAMI 和 IJCV 的现任编委，并曾任 CVIU 的副主编。

马毅：

马毅教授是 IEEE Fellow, ACM Fellow, SIAM Fellow, 现任加州大学伯克利分校教授。他 1995 年本科毕业于清华大学自动化专业；1997 年获 UCB 电子工程与计算机科学硕士学位；2000 年获 UCB 数学硕士学位和电子工程与计算机科学博士学位。博士毕业后，马毅先在 UIUC 任教，后于 2009 年至 2014 年间任微软亚洲研究院计算机视觉组主任及首席研究员，2014 年后在上海科技大学信息科学与技术学院任教。

研究方向为计算机视觉、高维数据分析、智能系统。是推动 3D 视觉进展的重要学者，其论文 Euclidean Reconstruction and Reprojection up to Subgroups 荣获 1999 年 ICCV 最佳论文奖（马尔奖），其论文 A Unified Algebraic Approach to 2-D and 3-D Motion Segmentation 荣获 2004 年 ECCV 时间检验-荣誉奖。在压缩感知、稀疏低质表达领域发表了一系列经典论文，包括 PAMI2008 上将稀疏表达用于人脸识别的论文等，是该领域的主要发起人和推动者。

屠卓文:

加州大学圣地亚哥分校认知科学系（联属计算机科学与工程系）教授，IEEE Fellow。曾任教于加州大学洛杉矶分校，2011 初年至 2012 年底间在微软亚洲研究院工作。在俄亥俄州立大学获得博士学位，在清华大学获得硕士学位。

发表过的系列重要的工作包括：早期的如 DDMCMC, Image Parsing, PBT, Auto-Context, ChnFtrs 近期的如 DSN, HED, ResNeXt, 以及 GDL (CVPR 2007) 的对抗式生成模型(能量模式)等等。Image Parsing 论文荣获 2003 年 ICCV 最佳论文奖(马尔奖), HED 荣获 2015 年 ICCV 最佳论文提名奖(马尔奖-荣誉奖)。

戴玉越:

教授、博士生导师，是国家级青年人才，院长助理，陕西省信息获取与处理重点实验室主任及国际联合研究中心副主任。于 2005 年、2008 年和 2012 年分别获得西北工业大学学士、硕士和博士学位。2012 年至 2014 年在澳大利亚国立大学从事博士后研究工作，2014 年至 2017 年在澳大利亚国立大学工程研究院担任 ARC DECRA 学者。担任 CVPR、ICCV、ACM-MM 等国际顶级会议领域主席。

主要研究方向涵盖底层视觉的众多问题，包括智能视觉感知、三维场景重建、多视角几何、显著性检测、无人驾驶等。关于 3D 重构论文荣获 CVPR 2012 最佳论文奖(大陆高校首次)，关于 RGB-D 显著性检测论文荣获 CVPR 2020 最佳论文奖提名。

张祥雨:

2017 年博士毕业于西安交通大学，期间参加微软亚洲研究院联合培养博士生项目，师从孙剑博士和何恺明博士。研究方向包括深度卷积网络设计，深度模型的裁剪与加速等。曾在 CVPR/ ICCV/ ECCV/

NeurIPS/ ICLR/ TPAMI 等顶级会议期刊上发表论文四十余篇，并多次获得顶级视觉竞赛如 ImageNet/ COCO 冠军。入选福布斯中国 U30 和 AI 2000 计算机视觉全球最具影响力学者榜单，以及智源青年科学家。

代表作包括业界广泛使用的神经网络架构 ResNets/ ShuffleNets/ RepVGG 系列，神经网络架构搜索算法 SPOS，模型压缩裁剪算法 CP/ MetaPruning 等，其中 ResNet 获 CVPR 2016 最佳论文奖。Google Scholar 引用数 150,000+。

黄高:

清华大学自动化系副教授，博士生导师。获国家优青、达摩院青橙奖、世界人工智能大会 SAIL 先锋奖、中国百篇最具影响国际学术论文和中国人工智能学会自然科学一等奖等荣誉，入选 MIT TR35 (亚太区)、北京智源学者、AI 2000 人工智能最具影响力学者。目前在 NeurIPS, ICML, CVPR 等国际顶级会议和期刊发表学术论文 70 余篇，被引 30000 余次，最高单篇引用超过 2 万次。

研究方向为神经网络结构设计，特别是动态神经网络和高效深度学习方法。其论文 DenseNet 荣获 2017 年 CVPR 最佳论文奖，是视觉的主流基本网络架构。还发表过多篇广受关注的工作，包括目前广泛使用的 Stochastic Depth (Drop Path)正则训练方法, SnapShot 集成方法等，是动态神经网络这一研究方向的主要推动者之一。

为了让更多感兴趣的朋友参与到精彩的分享会中，同时响应疫情防控政策，今年的 CVPR 论文分享会在 B 站“微软中国视频中心”全程进行了直播。感谢大家关注！

责任编辑 黄岩

2022 年度 CCF-CV"中科视拓 Seeta 学术新锐"

评选活动正式启动

为切实推动中国计算机视觉领域的科技进步，鼓励创新性研究，促进青年人才成长，由 CCF-CV 青年工作组提议，于 2018 年起评选中国计算机学会计算机视觉专委会“中科视拓 Seeta 学术新锐”（以下简称“中科视拓 Seeta 学术新锐”）。

本活动每年开展一次，由 CCF-CV 组织评选，中科视拓 Seeta 提供赞助支持，每年最终入选者不超过 3 人。欢迎计算机视觉相关领域的优秀研究生积极报名，CCF-CV“中科视拓 Seeta 学术新锐”将成为您专业领域启航的强大助推力量之源。

一、申报条件

1. 候选人在申请时系在计算机视觉及相关领域从事相关研究的研究生（含硕士研究生与博士研究生），尤其是具备一定科研潜力的低年级博士生。

2. 科研工作取得了突出的成绩。如：成果发表在计算机视觉领域顶级期刊或学术会议上，在权威学术竞赛中取得优异成绩，或为重大科研项目的完成与成果转化做出了重要贡献。

3. 候选人需提供支持同一主题的不超过 3 篇代表作。

4. 候选人需获得导师同意，且由两名 CCF-CV 专委会委员推荐（推荐人不包含导师），方可成为正式候选人。

二、申报方法

1. 每位候选人需提交申请书 1 份，以及由两名 CCF-CV 专委会委员填写的《CCF-CV 学术新锐推荐表》及其电子版（可从“阅读原文”中下载），提交至电子邮箱 ccfcvaward@sina.com。

2. 每位 CCF-CV 专委会委员同一年最多推荐两位候选人。

三、报名时间

即日起--2022 年 7 月 15 日。

四、其他未尽事宜

详见下方“阅读原文”中《CCF-CV 学术新锐评选条例》。

责任编辑 任传贤

专题综述

关于构建遥感图像解译数据集的探讨

武汉大学 夏桂松 龙洋

一、引言

近年来，遥感图像解译技术发展迅速，大幅提升了人类对地物信息的感知能力。随着遥感图像获取能力的不断增强，积累的遥感图像数据越来越多，实际应用中对海量遥感图像的自动化、智能化解译需求也越来越大，而遥感图像基准数据集是研发和测试相关解译算法的前提。因此，如何构建可靠的遥感图像解译基准数据集，为训练、测试和筛选实用的解译算法提供数据支撑，是推动遥感图像自动化、智能化解译发展的关键之一。

通过遥感图像解译算法提取图像中有价值的信息，是实现遥感图像理解和应用的基础。然而遥感图像复杂的光谱和结构等属性特征，为遥感图像内容的解译带来了严峻挑战。近年来，以深度学习为代表的驱动方法已经成为人工解译的重要替代方法，在实现大规模遥感图像自动解译和内容理解方面展现出巨大的潜力。然而，由于遥感图像解译标准数据集的缺乏，遥感图像解译算法的发展仍然面临着诸多挑战：接收的遥感图像数据量越来越大，但其中大部分数据未被赋予有价值的标注信息，数据难以被有效利用；缺乏有代表性的、具有精确语义标注信息的大规模遥感图像数据集，遥感图像解译算法的发展因此受到极大限制；遥感图像解译算法的泛化能力不足，难以满足复杂现实场景的应用需求；缺乏可靠的算法公共测试平台，难以对不同的解译算法进行系统评价和公平比较。

因此，建立可靠的遥感图像数据库构建原则与方法，构建面向实际应用场景的大规模遥感图像标注数据库，发展遥感图像解译算法测试与评价的公共平台，进而推动遥感图像解译算法面向实用化、自动化方向发展，是

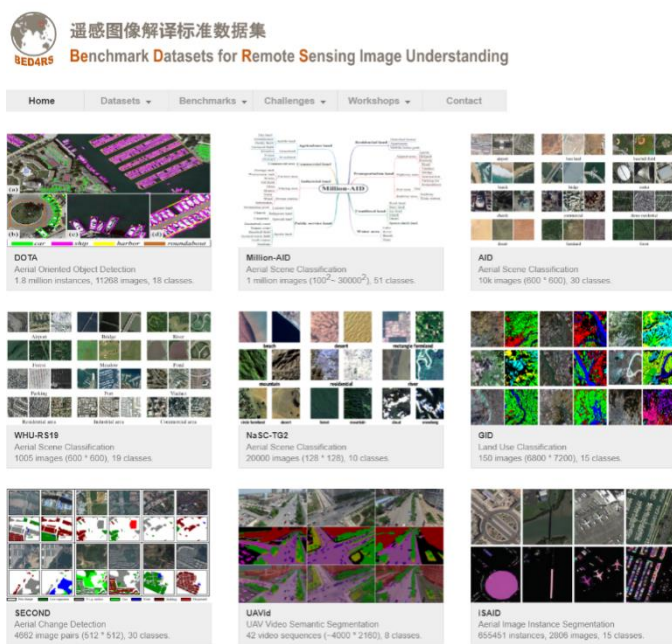


图1 大规模遥感图像解译数据库

应对上述挑战的一种有效途径。研究团队目前已构建了面向遥感图像智能解译的大规模标准数据库，包括面向遥感图像场景分类、目标检测、语义分割和变化检测等任务的精确标注数据集，如图1所示。

二、遥感图像解译数据集构建准则

数据驱动的遥感图像解译算法的性能严重依赖于训练数据集中语义标注信息的规模和质量。构建大规模、高质量图像解译数据集的主要挑战在于数据集构建的效率和质量控制。本文探讨构建遥感图像解译数据集的基本准则和方法，以期为大规模和实用化遥感图像解译数据集的高效构建提供参考指导。

一方面，遥感图像解译数据集的构建应该面向实际

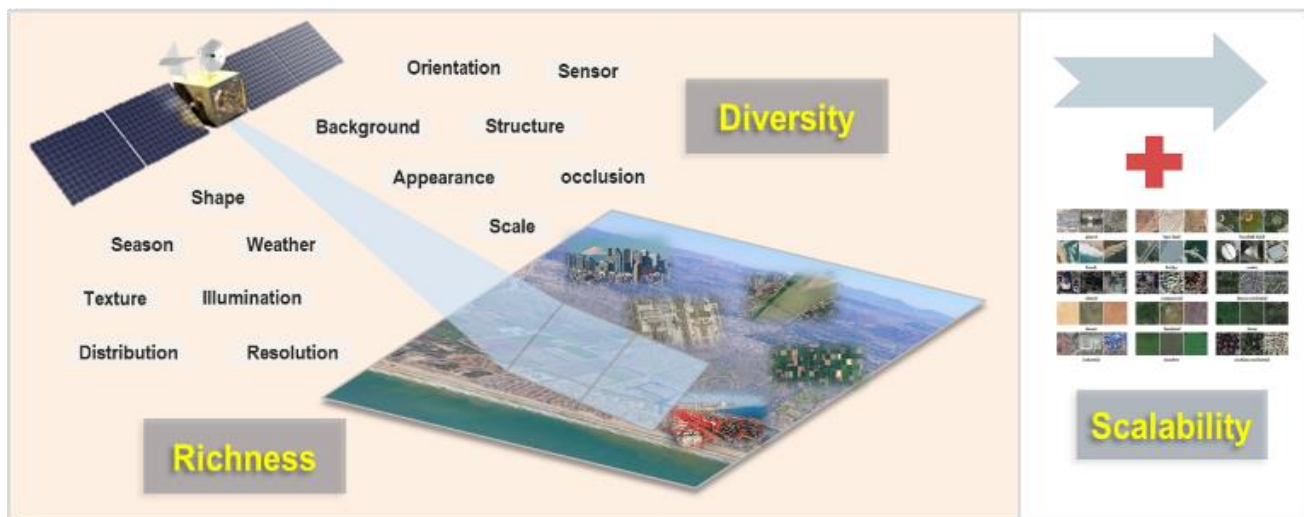


图2 遥感图像解译数据集构建基本原则: 多样性 (Diversity), 丰富度 (Richness), 和扩展性 (Scalability)

应用需求, 而非面向解译算法的特性。事实上, 遥感图像解译数据集应该针对实际应用中解译算法的训练、测试和筛选而构建。因此, 可靠的基准数据集对于全面评估所设计的解译算法的性能, 进而挑选出实用的解译算法至关重要。另一方面, 遥感图像解译数据集的标注应由应用部门而非由算法开发人员来完成。算法开发人员对算法属性比较熟悉, 在标注过程中会不可避免地带来个人偏好, 从而导致所标注的解译数据集偏向于算法特性。相对而言, 应用人员对遥感图像解译的现实应用场景有更深入的理解, 因此更熟悉解译任务中存在的具体挑战。因此, 后者标注的数据集更有利于增强算法的实用性。基于以上认识, 本文提出基于多样性 (Diversity), 丰富度 (Richness) 和可扩展性 (Scalability) 的遥感图像解译数据集构建准则, 如图2所示。

2.1. 多样性

数据集中的图像样本能够反映感兴趣地物在光谱、几何、形状、纹理等方面的属性特征, 可认为该数据集具有良好的多样性。从类内多样性的角度来看, 数据集中的每个标注样本应能从不同方面反映同一类别地物的不同属性特征, 而非地物内容和图像样本的简单重复。因此, 多样性较强的标注样本能够更全面地表征感兴趣地物在现实世界中的分布模式, 进而为训练具有更强特征表达能力的遥感图像解译算法提供可靠保障。

另一方面, 在构建遥感图像解译数据集时还应考虑不同类别地物之间的相似性。为此, 遥感图像数据集中

可以包含具有高度语义重叠和相似特征的细粒度类别, 从而使解译模型学习到区分不同类别地物的本质特征。增强感兴趣地物类内多样性和类间相似性, 是丰富地物特征的一种有效途径, 这对于构建具有较强多样性的遥感图像解译数据集至关重要, 进而增强所构建的解译数据集的实用性。

2.2. 丰富度

除了多样性, 遥感图像解译数据集的丰富度也是学习具有较强稳定性解译算法的重要保障。具体来说, 遥感图像解译数据集, 应具有丰富的图像样本、包含丰富的地物信息。为此, 构建遥感图解译数据集时, 需要采集不同天气、不同季节、不同光照、不同成像条件、不同传感器、不同时间和空间等条件下的遥感图像, 体现感兴趣地物在平移、视角、对象姿态和外观、空间分辨率、光照、背景、遮挡、时空属性等方面的特征差异。

此外, 遥感图像以俯视视角拍摄, 具有地理覆盖范围大、包含地物丰富、背景信息复杂等特点。面对这种情况, 解译数据应包含具有多样化特征的图像场景, 如在几何形状、结构特征、纹理属性等方面的多样性。从这一角度来看, 数据集应该包含大尺度的图像和足够多的标注样本, 以体现地物特征的分布模式。现实中经常会由于图像和样本的不足导致解译模型学习出现过拟合的现象, 这一问题在数据驱动的解译算法(如卷积神经网络)中尤为显著。因此, 基于以上考虑建立具有较强丰富度的遥感图像解译数据集, 能够使得所构建的解译模

型具有更强的泛化能力。

2.3. 可扩展性

可扩展性可用于描述已构建的解译数据集的扩展应用能力。遥感图像解译应用日益广泛，现实中对解译数据集的需求通常会随着时间和应用场景的变化而变化。例如，随着土地覆盖和土地利用的变化，可能需要将新的土地利用类别与数据集已构建的类别体系区分开来。因此，构建的解译数据集需要有充足的类别扩展空间，以包含新的地物类别，同时对不同地物类别之间的关系进行有效组织。因此，考虑到实际应用场景和应用需求的变化，解译数据集需要具有较好的可扩展性。

值得注意的是，现实中每天接收的遥感图像类型多样且规模巨大，迫切需要采用高效的方式为其赋予标注信息，从而发挥应用价值。因此，对遥感图像及其标注信息进行合理的组织、保存和维护对于数据集的可扩展性同样具有重要意义。例如能够将新标注的图像无缝地集成于已构建的数据集中，是数据集应用具有可扩展性的重要体现。因此具有良好可扩展性的解译数据集可以有效适应现实应用场景的需求变化。

三、遥感图像解译数据库构建方法

3.1. 数据库图像语义标注

根据标注过程是否有人工参与以及人工参与的程度，遥感图像数据集的语义标注方法可分为三种类型，即人工标注、自动标注和交互式标注。

人工标注：人工标注的过程是一种完全监督的标注模式，其优势在于具有较高的标注精度，因而许多遥感图像数据集采用人工标注的方式构建。无论是自然图像还是遥感图像，对图像中的内容进行标注的方式都是相似的，并且为了提高图像标注效率，现阶段已开发了許多面向不同解译任务的标注工具。因此，针对自然图像开发的图像标注工具可以进一步引入到遥感图像中，为构建高质量的遥感图像解译数据集提供基础。

自动化标注：遥感图像所包含的地物内容复杂，对于缺乏领域知识的标注者来说，很难对其语义内容进行精确标注。此外，手工标注的方式容易因标注者在领域知识和标注技能等方面的差异产生偏差，因而采用自动

关于构建遥感图像解译数据集的探讨。自动化标注的方法可以降低手工标注的难度，并进一步提高标注效率。自动化标注的方法通过构建一定数量的初始化样本来训练一个解译模型，然后将待标注样本输入到建立的标注模型中进行内容解译，最终将解译结果作为标注信息。由于遥感图像具有地理范围大、内容复杂的特点，可以采用迭代学习或增量学习来过滤噪声标注，提高标注模型的泛化能力。

交互式标注：在遥感大数据时代，现实应用中对遥感图像标注质量和效率需求不断提高，基于人机交互的半自动化标注是一种更为实用的标注方案。在该方案中，可以利用已有的标注数据构建初始标注模型，然后对未标注的遥感图像进行自动化标注。通常，通过使用主动学习策略和设置约束条件，利用解译模型筛选出难标注的图像，再采用人工进行标注，并将人工标注的信息反馈给自动标注模型，最后通过迭代学习的方式来进一步优化标注模型的性能。随着交互标注和迭代学习过程的进行，需要标注的图像数量将大大减少，从而大幅减轻标注难度并减少标注工作量。因此交互式标注的整体性能主要取决于标注者参与交互式标注的时间。

3.2. 图像语义标注质量控制

高质量的遥感图像数据集标注信息对于开发有效的解译算法及其性能评价十分重要，因此需要采用可靠的策略来控制数据集图像标注质量。

规则和样本：建立明确的标注规则是构建高质量遥感图像解译数据集的基础，否则不同的标注者将根据各自认知和偏好进行语义信息标注，从而对标注质量和标注信息的标准化管理造成影响。对于遥感图像标注而言，需要由具有领域知识的图像解译专家，建立合理的标注规则和示例，以建立良好的数据标注基础。

标注人员培训：通过对标注人员的培训，提高标注队伍的专业素质，进而为数据集标注质量提供保证。具体地，可以为每个标注员分配待标注数据，并要求其按规则进行标注，最后剔除未能通过测试评估的标注员。

多阶段标注：一系列复杂的标注操作容易引起标注者的疲劳并导致错误标注。为消除这种影响，可以设计多阶段的图像标注策略，以减小复杂标注任务的难度。通过这种处理，每位标注员只需关注整个标注工程中的

一个简单步骤，从而有效地降低标注错误率。

多重标注：采用多个标注者对同一个对象进行标注，并将不同的标注结果合并，可以有效提升数据集的标注精度。然而该方法的不足之处在于一个标注对象需要多个标注人员进行重复标注，因而标注效率较低。

标注审查：邀请相关标注者开展同行评审并对标注结果的质量进行评级，还可由领域专家进行进一步的评审。通过对标注流程中不同层次和步骤的标注结果进行严格的监督审查，可以实现对整体标注结果的质量控制。

四、大规模遥感图像解译标准数据库

为了促进遥感图像解译算法的研究和发展，研究团队基于所提出的遥感图像解译数据集构建准则和方法，建立了大规模遥感图像解译标准数据库，包含面向场景分类、目标检测、语义分割和变化检测等不同解译任务的大规模遥感图像标注数据集，以期为相关研究人员提供可靠的数据基础和算法测试标准。数据库访问地址：<https://captain-whu.github.io/BED4RS>

4.1. 遥感图像场景分类数据集：AID/Million-AID

AID 数据集包含 30 个场景类别，共 1 万幅场景实例。在 AID 基础上，研究团队建立了半自动化的遥感图像场景标注方案，并构建了百万级遥感图像场景分类数据集 Million-AID，共包含 51 个土地利用类别。与已有的遥感图像场景分类数据集相比，研究团队构建的场景分类数据集地理分布范围更广、场景样本丰富、数据集规模大。不同的场景类别采用层次化的类别网络进行语义关系组织，每个语义类别包含 2000~45000 幅场景图像，能满足数据驱动的解译模型构建与优化需求。

4.2. 遥感图像目标检测数据集：DOTA

DOTA 是一个大规模遥感影像目标检测数据集，目前有 DOTA-V1.0/V1.5/V2.0 三个版本。DOTA 数据集包含来自 GF-2、JL-1 和 Google Earth 等卫星和平台的高分辨率遥感图像，图像尺寸变化大（800~13000），目标类别丰富，标注样本规模大。每一个目标对象采用四边形边界框进行标注，可以精确表征目标的位置和方向。DOTA-V1.0 包含 2806 幅遥感图像，涵盖 15 个常见对象的类别共 188282 个标注实例；DOTA-V1.5 在

关于构建遥感图像解译数据集的探讨 DOTA-V1.0 基础上进行扩展，包含 403318 个标注实例；DOTA-V2.0 包含 11268 幅遥感图像，涵盖 18 个语义类别共 1793658 个标注实例。此外，研究团队为 DOTA 数据集提供了开放的算法测试与性能评估平台。

4.3. 遥感图像语义分割数据集：GID

GID 是一个用于遥感图像语义分割的大规模数据集，包含覆盖中国 60 多个不同城市的高分二号卫星图像，覆盖面积超过 50000 平方公里。GID 数据集由两部分组成：广域分类集和精细分类集。广域分类集包含 5 个语义类别共 150 景像素级标注的高分二号图像，精细分类集包含 15 类共 3 万个场景级标注样本和 150 景像素级标注的高分二号图像。GID 具有覆盖范围大、分布区域广和空间分辨率高等优点，相比现有的土地覆盖分类数据集能更好地满足广域制图的应用需求。

4.4. 遥感图像变化检测数据集：SECOND

大规模语义变化检测数据集 SECOND 包含 4662 对大小为 512x512 的遥感图像。各图像主要分布于成都、上海以及杭州等城市。所有图像像素都被精确地标记为 30 类语义变化类别，即低矮植被、树木、无植被土地、水体、建筑和运动场等地物类别之间的变化。与传统的语义变化检测数据集不同，SECOND 包含同一地物类别之间的变化（建筑物的拆除重建），可以为研究遥感地物语义类别之间的变化提供更好的数据基础。

五、未来发展方向展望

现阶段，研究团队所构建的上述遥感图像解译数据集已经集成到业界最大遥感影像样本库 LuoJiaSET 中，并成功应用于全球首个遥感影像智能解译专用深度学习框架 LuoJiaNET 的构建。在未来的工作中，研究团队将致力于构建面向遥感图像解译的数据集和解译算法在线公共测试平台，并建立遥感图像智能解译算法库。我们相信，遥感图像解译智能化、自动化的发展趋势不可阻挡，未来将会有更多面向真实遥感应用场景的数据集和算法不断涌现，因此应该鼓励更多的遥感图像解译数据集和解译框架在社区内共享，以促进遥感图像智能解译和应用的发展繁荣。

责任编辑 储瑾

参考文献

- [1] C. Toth and G. Józków. Remote sensing platforms and sensors: A survey. *ISPRS J. Photogrammetry Remote Sens.*, 2016, 115: 22–36.
- [2] T.-Z. Xiang, G.-S. Xia, and L. Zhang. Mini-unmanned aerial vehicle-based remote sensing: Techniques, applications, and prospects. *IEEE Geosci. Remote Sen. Mag.*, 2019, 7(3): 29–63.
- [3] G.-S. Xia, J. Hu, F. Hu, et al. AID: A benchmark data set for performance evaluation of aerial scene classification. *IEEE Trans. Geosci. Remote Sens.*, 2017, 55(7): 3965–3981.
- [4] G. Cheng, J. Han, and X. Lu. Remote sensing image scene classification: Benchmark and state of the art. *Proc. IEEE*, 2017, 105(10): 1865–1883.
- [5] Y. Long, G.-S. Xia, S. Li, et al. On creating benchmark dataset for aerial image interpretation: reviews, guidances and Million-AID. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 2021, 14: 4205–4230.
- [6] G.-S. Xia, X. Bai, J. Ding, et al. DOTA: A large-scale dataset for object detection in aerial images. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018: 3974–3983.
- [7] J. Ding, N. Xue, G.-S. Xia, et al. Object detection in aerial images: a large-scale benchmark and challenges. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2022: 1–18.
- [8] M. D. Hossain and D. Chen. Segmentation for object-based image analysis: A review of algorithms and challenges from remote sensing perspective. *ISPRS J. Photogrammetry Remote Sens.*, 2019, 150: 115–134.
- [9] X.-Y. Tong, G.-S. Xia, Q. Lu, et al. Land-cover classification with high-resolution remote sensing images using transferable deep models. *Remote Sens. Environ.* 2020, 237: 111322.
- [10] K. Yang, X.-Y. Tong, G.-S. Xia, et al. Hidden path selection network for semantic segmentation of remote sensing images. *IEEE Trans. Geosci. Remote Sens.*, 2022: 1–13.
- [11] X. Zhu, C. Vondrick, C. C. Fowlkes, et al. Do we need more training data? *Int. J. Comput. Vis.*, 2016, 119(1): 76–92.
- [12] K. Yang., G.-S. Xia, Z. Liu, et al, Asymmetric siamese networks for semantic change detection in aerial images. *IEEE Trans. Geosci. Remote Sens.*, 2022, 60: 1–18.
- [13] D. Acuna, H. Ling, A. Kar, et al. Efficient interactive annotation of segmentation datasets with polygon-rnn++. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018: 859–868.



夏桂松

武汉大学计算机学院教授、博士生导师。研究方向: 计算机视觉、模式识别、机器学习及应用。
主页: <http://www.captain-whu.com/xia.html>
Email: guisong.xia@whu.edu.cn



龙洋

武汉大学测绘遥感信息工程国家重点实验室在读博士生。主要研究方向为遥感图像理解。
Email: longyang@whu.edu.cn

专题综述

对抗学习：消除对抗噪声以提高对抗鲁棒性

西安电子科技大学 周大为 王楠楠

随着人工智能领域的快速发展，神经网络得到了广泛的应用。但是，最近研究发现神经网络对不易察觉但具有对抗性的微小扰动（即对抗噪声）具有明显的脆弱性。为了缓解对抗噪声的影响，对抗防御的研究受到了越来越多的关注。基于预处理的对抗防御是主要的防御类别之一，这种方法期望通过消除对抗噪声来提高神经网络的对抗鲁棒性。然而，基于预处理的防御可能会受到扰动放大效应的影响。此外，生成对抗噪声的对抗攻击算法是在不断演变的，基于已知类型的对抗噪声训练的防御模型通常不能很好地泛化到未知类型的对抗噪声上。为了解决这些问题，使用高层的类激活特征来消除对抗噪声和通过学习攻击不变表征来消除对抗噪声是值得探索的基于预处理的对抗防御策略。

一、神经网络对抗噪声的脆弱性

人工智能处在现代数据驱动科学的前沿，机器学习是人工智能的核心部分。神经网络是机器学习中重要的部分，其已经被广泛应用于商业、卫生和国防等领域。然而，大多数神经网络都存在严重的漏洞。他们很容易被微小的、人类无法察觉但经过精心设计的噪声（即对抗噪声）误导^{[1][2]}。

以分类任务为例，如图 1 所示，在自然样本上添加对抗噪声后生成的对抗样本在人类视觉上与自然样本没有明显的区别，但是他们会使神经网络产生错误的预测结果。神经网络的这种脆弱性对许多安全敏感型应用（比如人脸识别、自动驾驶）构成了严重风险。



图 1 自然样本 (Natural) 与不同的对抗样本。两者在人类视觉上相似，但是神经网络对对抗样本会做出错误的预测。下标“N”表示相应的攻击是非目标攻击。

为了提升神经网络对对抗噪声的鲁棒性（即对抗鲁棒性），针对对抗防御的研究得到了越来越多的关注。基于预处理的对抗防御是主要类别之一。已有研究表明，对输入样本中的对抗噪声进行消除^[10]，可以增强目标模型的对抗鲁棒性。但是，如何设计有效的方法缓解扰动放大效应的影响仍然需要进一步研究。此外，对抗攻击算法是多样且在不断发展的，基于已知对抗噪声训练的防御模型如何对未知对抗噪声具有较好的泛化性也值得深入探索。

二、基于类激活特征的对抗防御

扰动放大效应是指微小残留的对抗噪声在神经网络内部层中被逐渐放大，并最终导致网络输出错误的预测。类激活映射技术^[11]为解决这个问题提供了一种潜在的方法。给定一个分类网络，类激活映射技术可以通过将输出层的类别权重投影回最后的卷积层特征，并对加权特征进行线性求和来识别输入图像每个区域对最终预测的重要性。如图 2 所示，虽然对抗噪声在像素级是不易察觉的，但自然样本的类激活图和对

抗样本的类激活图之间存在明显差异。此外，加权特征位于网络的高层，在该层中，残余对抗噪声导致了较大的扰动。这促使我们可以在类激活特征空间中设计一种防御方法来抑制扰动放大效应。

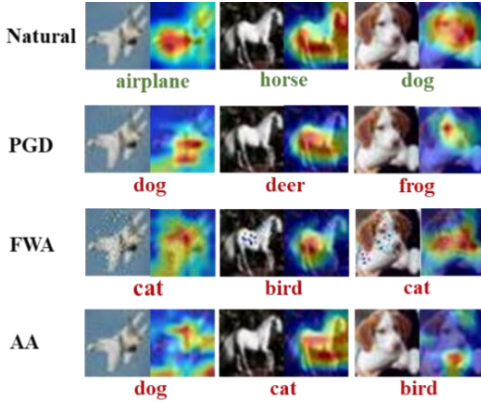


图 2 自然样本和对抗样本的类激活图。虽然对抗噪声在像素级是不易察觉的，但自然样本和对抗样本的类激活图之间存在明显差异。

我们设计了一种通过利用类激活特征来消除对抗噪声的防御方法。这种方法在类激活特征空间中，以自监督的对抗训练方式训练去噪模型，而不需要额外类型的对抗样本和真实的类别标签。具体来说，首先通过最大限度地破坏自然样本的类激活特征来生成对抗噪声并制作对抗样本。类激活特征的差异使得神经网络对对抗样本和自然样本做出不同的预测结果。我们将这种攻击称为基于类激活特征的攻击，优化目标为：

$$\max_{\tilde{x}} \Delta(x, \tilde{x}), \text{ subject to: } \|x - \tilde{x}\|_{\infty} \leq \epsilon,$$

其中， $\Delta(x, \tilde{x}) = \delta(\phi_x, \phi_{\tilde{x}})$ ， $\delta(\cdot)$ 为2范数度量， ϕ_x 和 $\phi_{\tilde{x}}$ 表示自然样本和对抗样本的类激活特征， ϵ 为扰动边界。

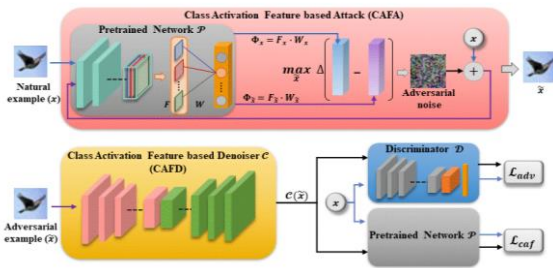


图 3 基于类激活特征的对抗防御方法的框架图。

基于生成的对抗样本，我们训练一个去噪模型，即基于类激活特征的去噪器(CAFD, class activation

feature-based denoiser)。此方法没有直接使用像素级损失函数来训练防御模型，而是最小化自然样本和对抗样本的类激活特征之间的距离。另外，引入一种基于RaGAN^[12]的图像鉴别器来增强去噪后图像的纹理细节，使其更接近自然样本。方法的整体框架如图3所示。

在SVHN^[13]和CIFAR-10^[14]两个数据集上的评估能够说明上述方法的有效性。使用三种具有代表性的攻击来生成对抗样本，分别为PGD^[3]，AA^[4]和FWA^[7]。如图4所示，基于类激活特征的对抗防御方法能够消除对抗噪声，恢复对抗样本的类激活图。定量分析的结果如表1所示，相比于之前的JPEG^[17]、TVM^[17]、APE-G^[18]和HGD^[19]方法，对抗准确率有所提升。

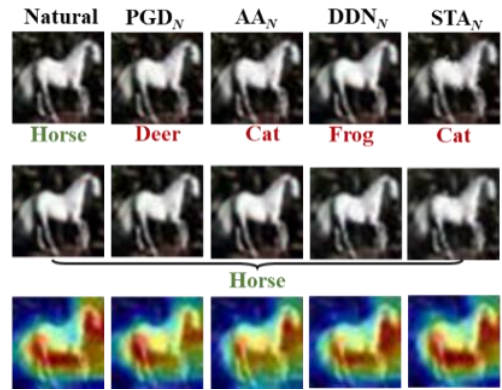


图 4 对抗防御方法针对不同的对抗攻击的去噪效果。(第一行：自然样本和对抗样本；第二行：去噪后的样本；第三行：去噪后的样本的类激活图)。

表 1 防御不同对抗攻击生成的对抗样本的对抗准确率。APE-G 和 HGD 使用 DDN_N生成的对抗样本作为对抗性训练数据。

Dataset	Defense	NONE	PGD _T	PGD _N	AA _N	FWA _N
SVHN	JPEG	90.22	13.40	4.33	2.44	6.93
	TVM	89.99	24.83	5.47	4.16	3.39
	APE-G	89.60	16.60	6.80	18.08	13.43
	HGD	89.88	55.00	42.65	37.75	32.56
	CAFD	92.35	89.37	85.36	86.43	41.93
CIFAR-10	JPEG	86.68	49.66	48.59	41.16	14.73
	TVM	90.35	43.29	31.21	33.54	9.190
	APE-G	91.82	37.69	21.92	23.35	23.34
	HGD	92.36	68.82	53.13	54.27	53.13
	CAFD	91.10	89.42	87.21	88.20	64.41

三、基于攻击不变特征的对抗防御

专注于有限训练数据中已知类型的对抗样本可能导致对抗防御模型过拟合于给定类型的对抗噪声，并缺

乏针对未知类型对抗噪声的通用性或有效性。现实世界中存在广泛甚至未知的对抗攻击类型，这促使我们设计一种可以处理不同类型或未知的对抗样本的防御方法。

认知科学中的一些研究对解决这个问题具有一定的启发。具体来说，一些研究表明^{[20][21]}，即使人脸显示出不同甚至未曾见过的表情，我们也能够识别人脸身份，因为我们的大脑善于提取不变的面部特征。同样，如图 5 所示，我们人类也无法轻易区分自然样本和对抗样本，因为我们只关注表示语义分类信息的不变特征，而忽略了对抗噪声。对抗样本保留了不变特征，以使在人类视觉上无法预先识别对抗样本。我们将这种不变特征命名为攻击不变特征。

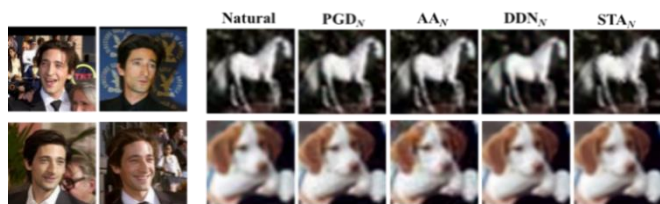


图 5 同一身份不同表情的人脸图像和不同的对抗样本。

基于攻击不变特征，我们设计了一种对抗防御方法 (AIFD, attack-invariant feature-based defense) 来消除对抗噪声。这种方法基于自动编码器的框架，将对抗噪声消除分为学习攻击不变特征和从攻击不变特征中恢复自然样本。方法的整体框架如图 6 所示。

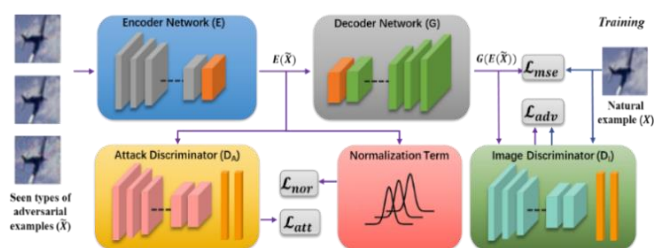


图 6 基于攻击不变特征的对抗防御整体框架图。

具体来说，我们以对抗性特征学习的方式引入了一对编码器和鉴别器，用于将攻击不变特征从对抗样本中分离出来。鉴别器用于从编码的攻击不变特征空间中区分特定于攻击的信息(例如，攻击类型标签)，而编码器旨在提取出使鉴别器无法区分的特征。通过迭代优化，在编码的过程中可以去除特定于攻击的信息，并保留不变特征。编码器和鉴别器对应的损失函数分别是：

$$\mathcal{L}_{D_A} = -\frac{1}{K} \sum_{k=1}^K Y_k^p \cdot \log(\sigma(D_A(E(\tilde{X}_k)))) ,$$

$$\mathcal{L}_{att} = -\frac{1}{K} \sum_{k=1}^K Y_{\zeta}^p \cdot \log(\sigma(D_A(E(\tilde{X}_k)))) ,$$

其中， $\sigma(\cdot)$ 表示 softmax函数， \tilde{X}_k 表示第k种对抗攻击生成的对抗样本， $Y_k^p = [y_{k1}^p, y_{k1}^p, \dots, y_{kN}^p]^T$ 表示第k种攻击的攻击特定标签， $y_{kn}^p = [\xi_1, \xi_2, \dots, \xi_K]^T$ 表示 one-hot 向量且 ξ_i 在 $i = k$ 时等于 1，其他时候为 0。 $Y_{\zeta}^p = [y_{\zeta 1}^p, y_{\zeta 2}^p, \dots, y_{\zeta N}^p]^T$ 是攻击混淆标签， $y_{\zeta n}^p = [1/K, 1/K, \dots, 1/K]^T$ 是一个 k 维常数向量。K 为使用的对抗攻击的类别数。

另外，训练过程中使用的对抗样本往往带有偏差，因为现实世界中广泛存在的攻击类型非常多样。偏差问题可能会使学习到的编码器适用于训练过程中使用的攻击或类似类型的攻击，但对某些明显不同的未知类型的攻击具有较差的泛化能力。为了解决偏差问题，我们在攻击不变特征的编码空间中添加一归一化项，以将每种类型的攻击的特征分布与多元高斯先验分布^{[15][16]}相匹配。通过这种设计，学习到的攻击不变特征有望推广到未知类型的攻击。该项的损失函数为

$$\mathcal{L}_{nor} = JSD(P_1, \dots, P_K) = -\frac{1}{K} \sum_{k=1}^K K \mathcal{L}(P_k || \mathcal{N}).$$

此外，为了提高从攻击不变特征中恢复的样本的质量，引入图像鉴别器和像素级距离度量作为约束来帮助训练解码器。通过联合优化编码器和解码器，防御模型可以有效地消除多种类型的对抗噪声，提高对抗准确率。

如图 7 和表 2 所示，在 MNIST^[9]和 CIFAR-10 两个数据集上的定性和定量分析表明，针对六种对抗攻击 PGD, AA, FWA, CW^[5], DDN^[6], STA^[8]。该方法能够有效去除对抗噪声。相比于之前的方法 APE-G 和 HGD，对抗准确率有所提升。

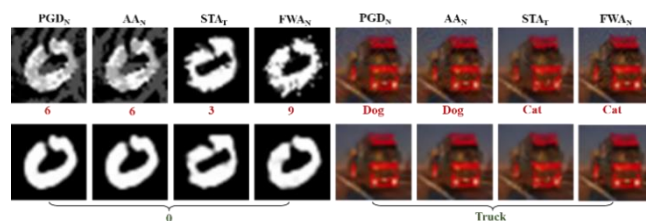


图 7 对抗防御方法针对不同的对抗攻击的去噪效果。(第一行：对抗样本；第二行：去噪后的样本)。下标“T”表示相应的攻击是目标攻击。

表 2 防御不同对抗攻击生成的对抗样本的对抗准确率。APE-G 和 HGD 使用 PGD_N 生成的对抗样本作为对抗性训练数据

Dataset	Defense	NONE	PGD _T	PGD _N	AA _N	FWA _N	CW _N	DDN _N	STAN	STAT
MNIST	APE-G	98.43	96.80	91.24	87.60	66.05	97.66	97.85	83.43	78.60
	HGD	98.64	98.70	98.11	97.57	49.57	98.33	98.46	78.68	63.59
	AIFD	98.84	98.71	98.15	97.62	74.21	98.55	98.72	90.92	86.27
CIFAR-10	APE-G	76.92	60.91	55.62	39.91	57.21	76.82	67.61	52.81	63.54
	HGD	89.59	76.97	60.56	57.66	62.33	86.74	83.98	57.11	68.03
	AIFD	91.79	79.57	61.34	61.06	75.83	88.53	85.36	63.19	76.38

责任编辑 崔海楠

参考文献

- [1] Ian J. Goodfellow, Jonathon Shlens, Christian Szegedy. Explaining and Harnessing Adversarial Examples. ICLR 2015.
- [2] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, Rob Fergus. Intriguing Properties of Neural Networks. ICLR 2014.
- [3] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, Adrian Vladu. Towards Deep Learning Models Resistant to Adversarial Attacks. ICLR 2018.
- [4] Francesco Croce, Matthias Hein. Reliable Avaluation of Adversarial Robustness with an Ensemble of Diverse Parameter-free Attacks. ICML 2020.
- [5] Nicholas Carlini, David Wagner. Towards Evaluating the Robustness of Neural Networks. SSP 2017.
- [6] Jérôme Rony, Luiz G. Hafemann, Luiz S. Oliveira, Ismail Ben Ayed, Robert Sabourin, Eric Granger. Decoupling Direction and Norm for Efficient Gradient-based L2 Adversarial Attacks and Defenses. CVPR 2019.
- [7] Kaiwen Wu, Allen Houze Wang, Yaoliang Yu. Stronger and Faster Wasserstein Adversarial Attacks. ICML 2020.
- [8] Chaowei Xiao, Jun-Yan Zhu, Bo Li, Warren He, Mingyan Liu, Dawn Song. Spatially Transformed Adversarial Examples. ICLR 2018.
- [9] LeCun, Yann, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86, no. 11 (1998): 2278-2324.
- [10] Guoqing Jin, Shiwei Shen, Dongming Zhang, Feng Dai, Yongdong Zhang. APE-GAN: Adversarial Perturbation Elimination with GAN. ICASSP 2019.
- [11] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, Antonio Torralba. Learning Deep Features for Discriminative Localization. CVPR 2016.
- [12] Alexia Jolicoeur-Martineau. The relativistic discriminator: A Key Element Missing from Standard GAN. arXiv preprint, arXiv:1807.00734, 2018.
- [13] Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y Ng. Reading digits in natural images with unsupervised feature learning. 2011.

- [14] Alex Krizhevsky, Geoffrey Hinton, et al. Learning Multiple Layers of Features from Tiny Images. 2009.
- [15] Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, Ian Goodfellow, Brendan Frey. Adversarial Autoencoders. arXiv preprint arXiv:1511.05644, 2015.
- [16] Diederik P Kingma, Max Welling. Auto-encoding Variational Bayes. ICLR 2014.
- [17] Chuan Guo, Mayank Rana, Moustapha Cissé, Laurens van der Maaten. Countering Adversarial Images Using Input Transformations. ICLR 2018.
- [18] Guoqing Jin, Shiwei Shen, Dongming Zhang, Feng Dai, Yongdong Zhang. APE-GAN: Adversarial Perturbation Elimination with GAN. ICASSP 2019.
- [19] Fangzhou Liao, Ming Liang, Yinpeng Dong, Tianyu Pang, Xiaolin Hu, Jun Zhu. Defense Against Adversarial Attacks Using High-level Representation Guided Denoiser. CVPR 2018.
- [20] Mortimer Mishkin, Leslie Gail Ungerleider. Contribution of Striate Inputs to the Visuospatial Functions of Parieto-occipital Cortex in Monkeys. Behavioural brain research, 6(1):57–77, 1982.
- [21] Nancy Kanwisher, Josh McDermott, Marvin M. Chun. The Fusiform Face Area: a Module in Human Extrastriate Cortex Specialized for Face Perception. Journal of neuroscience, 17(11):4302–4311, 1997.



周大为

西安电子科技大学在读博士生。主要研究方向是：对抗机器学习及其应用。
Email: dwzhou.xidian@gmail.com



王楠楠

教授，博士生导师，西安电子科技大学，综合业务网理论及关键技术国家重点实验室副主任。近年来从事计算机视觉和统计机器学习方面的研究，在图像跨域重建与可信识别方面进行了深入研究，内容包括跨模态生成、视频理解与分析、底层视觉处理、对抗学习等。
Email: nnwang@xidian.edu.cn

热点追踪

基于多相机系统的全局式三维建模算法

中科院自动化研究所 崔海楠 申抒含

一、摘要

为了充分感知周围三维场景，机器人和自动驾驶汽车等智能体通常会在头顶或者周围安装多相机系统。该系统通过多个相机同步获取 360 度场景数据，然后进行自动三维建模，完成场景三维感知。传统三维建模流程需要利用标定场对多相机系统提前进行离线标定，以解算多个相机之间的相对旋转和相对平移关系；然后，基于该相对位姿关系，再利用增量式从运动恢复结构(SfM)方法对多相机视频数据进行三维建模。然而，多相机系统的离线标定算法通常需要单独的标定场地，以保证多相机之间有足够的共视区域，不仅设计复杂、计算困难而且操作也不方便，一旦更换相机位置或者传感器就需要重新执行离线标定流程；同时，传统的增量式建模方法也无法满足大规模场景三维建模的效率需求。

本文提出基于多相机模型约束的全局式运动平均方法，以解决大规模场景三维感知问题。首先将相机分为参考相机和非参考相机两种类别，从而在计算过程中可以自动地标定多相机之间的相对位姿关系。然后，提出基于多相机模型的旋转平均和平移平均方法，全局式地解决相机的姿态和位置标定问题。该方法不仅操作简单，避免了离线标定的复杂流程，而且极大地提高了场景三维建模的效率和鲁棒性。相关成果已被 AAAI2022 录用为口头报告。

二、多相机模型

在多相机系统采集数据的过程中，由于多个相机是刚性固定的，所以它们之间的相对位姿在运动中是不会发生改变的，即拥有固定的相对旋转和相对平移关系。

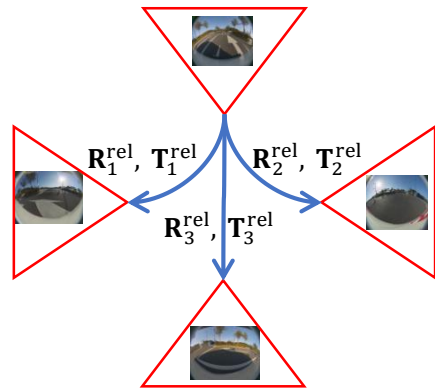


图 1 多相机刚体模型示例

本文将多个相机（假设相机数目为 N ）在同一个位置拍摄时的模型定义为一个刚体模型，每个刚体模型中包括 1 个参考相机和 $N-1$ 个非参考相机，那么原多相机系统标定问题就转化为求取非参考相机到参考相机之间的相对旋转 R_j^{rel} 和相对平移 T_j^{rel} 。如图 1 所示，多相机系统中包含了 4 个相机，假设前向相机为参考相机，待标定的相对位姿为前向相机与左侧相机、右侧相机和后向相机之间的相对位姿。每个相机的绝对位姿 (R_j, C_j) ，可通过多相机模型内部的相对位姿与所在刚体模型的参考相机绝对位姿 (R_j^{ref}, C_j^{ref}) 计算：

$$R_j = R_j^{rel} R_j^{ref},$$

$$C_j = R_j^{refT} T_j^{rel} + C_j^{ref}.$$

传统 SfM 系统优化参数包括：相机内参和外参，通常采用 5 自由度相机内参模型（包括焦距，两个主点和两个畸变），6 自由度相机外参（相机旋转矩阵和位置）。假设 N 个相机拍摄，每个相机采集 M 幅图像，总参数

表 1 本文方法 MRA 和 MTA 与传统方法的精度对比评测（与真实值对比的误差均值）。其中 RRA、IRA 和 MRA 是全局旋转平均算法，误差评测单位为角度；LUD，BATA 和 MTA 是全局平移平均算法，误差评测单位为米。

KITTI	00	01	02	03	04	05	06	07	08	09	10
RRA ^[1]	1.7	2.8	88.8	0.7	0.4	1.8	57.8	2.9	0.8	1.4	1.2
IRA ^[2]	0.8	1.6	5.0	36.7	0.3	1.3	0.5	0.5	0.7	1.1	0.7
MRA	0.8	1.1	2.3	0.7	0.3	0.8	0.5	0.6	0.8	0.6	0.8
LUD ^[3]	38.3	41.6	247.0	7.0	66.9	26.4	71.4	13.3	32.2	49.9	40.3
BATA ^[4]	36.8	29.3	259.9	10.0	87.8	19.7	65.1	8.9	24.8	38.5	26.0
MTA	1.9	23.2	10.1	1.9	13.3	1.2	1.0	2.2	3.1	3.5	1.2

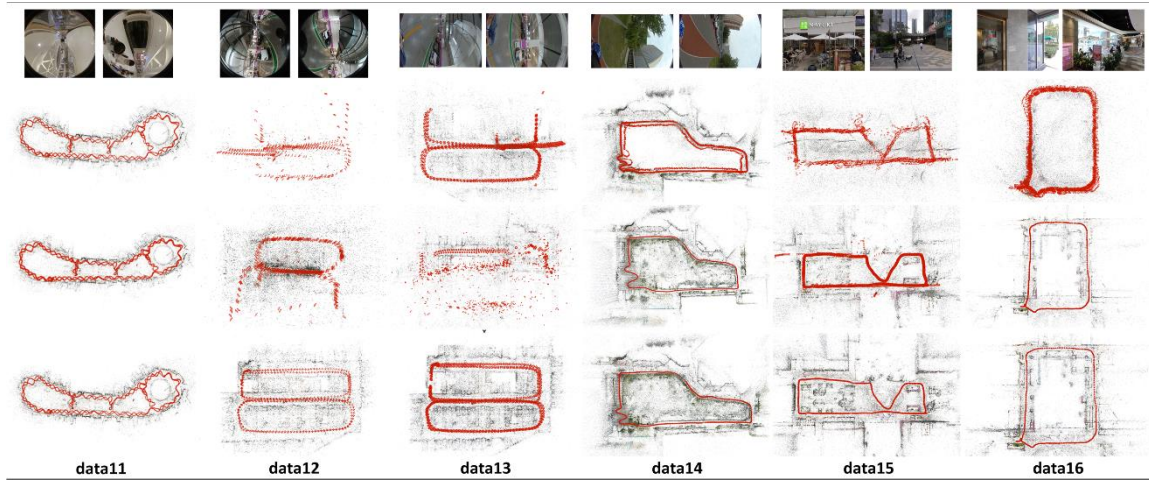


图 2 从上到下依次是 LUD、BATA 和 MTA 建模结果。数据集为 Insta360 多相机采集，红色点表示相机位置。

数目为 $5N+6MN$ 。经过参考相机和非参考相机的分类，本文只需要估计参考相机外参和多相机系统内部的相对旋转和相对平移，因此优化参数数目降为 $5N+6M+6(N-1)$ 。考虑到大规模场景中拍摄图像数目要远大于相机的数目，那么基于多相机模型的建模系统参数数目约等于传统建模系统参数数目的 $1/N$ 。参数的减少不仅可以极大地提高三维建模中捆绑调整优化效率，而且可以提高场景建模的鲁棒性。

三、多相机系统全局运动平均

全局式运动平均流程包括：特征点检测与匹配，摄像机旋转平均，摄像机平移平均，三角化和捆绑调整。基于多相机模型，本文在特征点匹配，摄像机旋转平均和平移平均以及捆绑调整中做了针对性的改变和升级。

对于特征点匹配，由于多相机系统通常都是视频采集数据，所以本文采用的策略是序列式匹配和回环匹配。

序列式匹配只对相邻视频帧进行匹配，而回环匹配采用基于图像检索的方式对图像的相似图像进行匹配，实验中采用的是序列式相邻 20 帧图像和检索 50 幅相似图像进行匹配。由于人造场景中可能存在大量的重复纹理，回环匹配中可能存在大量错误匹配。因此在图像匹配完成以后，本文利用多相机模型去衡量检索图像的可靠性：当两幅图像是回环匹配，那么它们所在的多相机刚体也需要有足够多的匹配，即不仅在某一个方向上有匹配，还需要在 360 度的方向有足够的匹配，才可以确定是回环匹配。在特征点匹配完成后，构造场景图，其中图中每个节点表示一幅图像，当两幅图像之间有足够多的匹配点时连接一条边，同时利用本质矩阵分解得到这条边上的相对旋转和相对平移。其中相机相对位姿 (R_{ij}, t_{ij}) 和绝对位姿 (R_i, C_i) 的关系如下：

$$R_{ij} = R_j R_i^T$$

$$\lambda_{ij} \mathbf{t}_{ij} = \mathbf{R}_j (\mathbf{C}_i - \mathbf{C}_j)$$

基于多相机模型的参考相机和非参考相机的分类，原场景中边转化为四种不同种类的连接边，即参考相机-非参考相机，参考相机-参考相机，非参考相机-参考相机以及非参考相机-非参考相机。每一种连接边上的相对位姿和绝对位姿的关系约束需要重新计算，即公式左侧为本质矩阵分解得到的相对几何关系保持不动，右侧相机的绝对位姿需要利用多相机模型的相对位姿和所在刚体的参考相机位姿表示，从而公式右侧均为要估计的参考相机位姿和多相机模型的相对位姿。根据转化后的场景图中的相对旋转与绝对旋转的关系，将相机旋转矩阵转化为李代数空间 $SO(3)$ 进行基于 $L1$ 误差范数的线性求解，然后再利用 $L2$ 范数进行非线性迭代加权求解，得到所有参考相机的绝对姿态和多相机模型的相对旋转矩阵。基于该结果，所有非参考相机的旋转矩阵可以通过第 2 节中介绍的多相机模型计算得到。这里我们将这种方式称为基于多相机模型的全局旋转平均算法 MRA。同理，根据转化后的场景图中的相对平移与

绝对位置的关系，我们将等式左右相减，基于 $L1$ 误差范数估计多相机模型的相对平移和参考相机的绝对位置，进而获得非参考相机的绝对位姿，这种方式被称为基于多相机模型的全局平移平均算法 MTA。

本文对多组多相机模型进行测试，包括双目相机 (KITTI 数据)，利用 Insta360 采集的双鱼眼和六鱼眼数据。表 1 展示了我们的方法与传统方法对 KITTI 数据的建模精度测试，可以看出我们方法在精度和鲁棒性上较传统方法有了很大提升。图 2 展示了不同方法在其他多相机系统采集数据中的结果，可以看出本文方法鲁棒性更强，更多实验对比结果请参考原文 MMA: Multi-camera Based Global Motion Averaging (aaai.org)。

通过多相机模型理论分析和对大量数据测试，本文方法对于多相机系统内的相机数目和分布均不敏感，且鲁棒性较传统系统有了很大提升。对于机器人和自动驾驶汽车等主流应用，我们的方法可以更加快速鲁棒地进行场景三维感知。

责任编辑 王金甲

参考文献

- [1] RRA: Chatterjee, A.; and Govindu, V. M. 2017. Robust relative rotation averaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 40(4): 958–972
- [2] IRA: o, X.; Zhu, L.; Xie, Z.; Liu, H.; and Shen, S. 2021. Incremental Rotation Averaging. *International Journal of Computer Vision (IJCV)*, 129: 1202–1216.
- [3] LUD: Ozyesil, O.; and Singer, A. 2015. Robust camera location estimation by convex programming. In *CVPR*. IEEE
- [4] BATA: Zhuang, B.; Cheong, L.-F.; and Lee, G. H. 2018. Baseline desensitizing in translation averaging. In *CVPR*. IEEE



崔海楠

中科院自动化研究所副研究员。主要研究方向为基于图像的大规模场景三维重建。
Email: hncui@nlpr.ia.ac.cn



申抒含

中科院自动化研究所研究员。主要研究方向为三维重建，机器人三维感知和语义建模。
Email: shshen@nlpr.ia.ac.cn

顶会观察

ICLR 2022

华东师范大学计算机科学与技术学院副研究员 钱鸿

国际学习表征大会(International Conference on Learning Representations, ICLR)是深度学习与表示学习方向的顶级会议之一。ICLR 由深度学习权威学者、图灵奖得主 Yoshua Bengio 和 Yann LeCun 牵头创办, 尽管 ICLR 从 2013 年才开始第一届会议, 还很“年轻”, 但它已得到学术界和产业界研究者的广泛认可。受疫情影响, 本届 ICLR 大会于 2022 年 4 月 25 日至 29 日以线上会议的形式举办, 包括 4 天的正会和 1 天的 Workshop。

一、会议亮点

线上会议形式: 由于今年是线上会议, 为了尽可能的保证会议效果, ICLR 为注册者们提供与其他参会者进行互动的功能, 例如向演讲者提问、参加现场海报环节、与其他与会者聊天和交流等。与 2021 年类似, 2022 年 ICLR 会议的线上海报环节使用 GatherTown 系统, 通过 Rocket.Chat 实现参会人员与作者的交流。2022 年会议结束后, 官方网站将提供所有视频及报告 PPT。ICLR 的相关动态会在其官方博客(<https://blog.iclr.cc>)同步。

Open Review 评审机制: ICLR 得到众多学者好评的原因之一是其推行的 Open Review 制度。现有的论文评审制度主要分为单盲(Single-Blind Review)、双盲(Double-Blind Review)及开放评审(Open Review)等多种形式。传统的单盲或双盲制度在评审时会隐藏掉作者和评审人的信息, 投稿论文、审稿意见、Rebuttal 内容在评审期间一般不会对外公开。而 Open Review 则不同, 所有提交的论文都会公开, 并且接受同行的评价

及提问 (Open Peer Review)。在初审意见公开后, 论文作者可以根据论文评审意见进行公开的 Rebuttal。

博客 Track: 随着机器学习会议中提交和发表的论文数量剧增, 跟踪某领域的最新进展变得更具挑战性。与此同时, 博客文章逐渐成为了一种流行且实用的探讨科学的方式。博客文章通过灵活的途径来推动关于学术工作的新见解, 或者局限性的公开透明讨论, 从而为科学技术发展提供了巨大的价值。然而, 由于博客并不像学术论文那样被认可, 只有少数研究人员通过维持一个活跃的博客来展示他们的研究工作和提高认可度, 因此, 期望所有研究人员通过博客展示他们的工作是不现实的。

ICLR 为了激励研究人员通过博客来回馈以往的工作和总结他们的成果, 在 2022 年开辟了博客投稿 Track。讨论以往工作的博客可以增加历史工作的认可度, 激励研究人员提交更高质量的工作以获得博客的宣传, 博客文章会发现历史工作的“技巧”以提升历史工作的可复现性和透明度, 并且优质的博客内容也可以帮助非相关专业领域的研究人员快速了解有关工作。经过会议收集和认证的博客会获得更高的可信度, 方便研究者收集和访问, 这是 ICLR 于 2022 年开辟这一新的投稿赛道的初衷。由于是第一次收集博客投稿, 此次博客 Track 限制了博客内容为探讨过去在 ICLR 上发表的论文, 并且要求所述工作不能与作者存在利益冲突。经评委评审以确认博客的内容是否阐述清晰并且带来了新的理论或者理解, 最终通过评审的博客会发布在 ICLR 官网或者 ICLR 的 GitHub 页面上。

二、录用情况

ICLR 2022 共收到了 3391 篇有效投稿，最终接收了 1095 篇论文，接收率为 32.3%。相较于 2021 年，2022 年 ICLR 的投稿量提升 13.1%(394 篇)，录用率也有所上升，录用论文数量较 2021 年提升 27.3%(235 篇)。录用的论文中包括 54 篇 Oral(1.59%)，176 篇 Spotlight(5.19%)，7 篇论文因其卓越的条理性、洞察力、创造力和潜在的持久性影响力获得杰出论文奖。

录用论文的研究热点主要包括强化学习、深度学习、图神经网络、表示学习、自监督学习、联邦学习、鲁棒/泛化性、对比学习等。杰出论文奖中，华人一作论文共有 3 篇，此外，还有 3 篇论文获得了荣誉提名。国内上榜高校包括清华大学、中国人民大学、浙江大学和重庆大学四所。值得一提的是，张钹院士、朱军教授等学者合作的论文获得了本届大会杰出论文奖。

博客 Track 最终接收 21 篇博客，其中有一篇是 ICLR 官方特邀特斯拉人工智能和自动驾驶视觉总监 Andrej Karpathy 撰写的反向传播 33 年进展综述。Andrej Karpathy 复现了 Yann LeCun 等人 1989 年的论文 Backpropagation Applied to Handwritten Zip Code Recognition，并探究了当年 Yann LeCun 等人遇到的难点。有趣的是，Karpathy 还展望了下一个 33 年，即 2055 年的研究者将如何看待今天的深度学习研究。这种跨越时空的对话蕴含着机器学习社区多年来的努力和对未来的美好期望^[1]。

三、邀请报告

ICLR 2022 的特邀报告(Invited Talk)共收到了来自各领域主席以及公众的 60 多个提名，随后本届高级程序主席 Yan Liu 和程序主席们以研究重要性与多样性为评价标准，从中评选并邀请了 8 位专家学者在本次大会上报告。本次特邀报告的内容涵盖了从深度强化学习，可解释性，计算机视觉，AI 伦理，ML 系统等深度学习的基础研究到蛋白质结构预测，神经科学，全球健康与流行病学等各科学领域中深度学习的应用突破。

From Reinforcement Learning to AI. 麦吉尔大学计算机学院副教授、DeepMind 蒙特利尔负责人

Doina Precup 报告并讨论了强化学习如何构建知识并以此发展通用 AI。近年来，强化学习在从简单游戏到复杂控制任务等多个领域上都取得了巨大成功。然而，强化学习可以超越特定的任务，为交互学习式的通用 AI 提供基础，使其能在不断的交互中构建知识，实现目标。在本报告中，Precup 探讨了奖励之于指向目标的重要性，以及强化学习用于学习程序性和预测性通用知识的方式。进一步地，报告还概述了这一领域最新取得的进展，并指出了一些重要的开放性问题。

Beyond Interpretability: Developing a Language to Shape Our Relationships with AI. Google Brain 研究科学家 Been Kim 描述并探讨了一种新的与 AI 进行可解释交互的模式。随着 AI 技术逐渐深入人们的生活，如何与这类新的“同事”相处成为了一个重要的问题。AI 可解释性的研究目标就是去设计人类与 AI 的关系，通常需要通过制作工具从 AI 模型生成解释。然而，Kim 认为还应当将 AI 作为单独的科学对象，与人类隔离研究。这样做不仅为人们制作的工具提供了准则，而且对于将人们与 AI 的工作关系提升到一个新的层次也是必要的。一条可行的道路是开发一种能够让人们学习人工智能并从中获得灵感的“语言”。从人类知识角度而言，这种语言的设计应当基于 AI 的自然特性；从 AI 知识的角度而言，这种语言的设计应当拓宽人类的知识边界。最终，这种语言将会塑造人们的思维，也会塑造未来人们创造的 AI。

Do You See What I See? Large-Scale Learning from Multimodal Videos. 法国国立计算机及自动化研究院(INRIA)研究总监、Google 研究科学家 Cordelia Schmid 介绍了大规模多模态视频表征学习的最新进展。Schmid 介绍了将 Bert 模型用于多模态数据学习的视频和文本联合模型 VideoBert，它实现了零样本预测和视频描述生成的 SOTA。另外，Schmid 团队提出了一种视频问答方法，该方法依赖于教学视频和带有文本问答模块的跨模态监督训练。这一方法实现了在没有任何监督信息情况下进行视频问答(即零样本 VQA)的 SOTA 并在预训练和视频问答数据集微调方面取得了具有竞争力的结果。最后，讲者介绍了一个最新的 VideoCC 数

数据集, 该数据集将图像描述转换为视频, 并达到了多模态零样本视频和音频检索以及视频描述生成的 SOTA。

Representation Learning in the Global South: Societal Considerations-Fairness, Safety and Privacy. 夸梅恩克鲁玛科技大学(KNUST)全球健康系主任 John H. Amuasi 在报告中指出并探讨了表征学习甚至整个机器学习领域中的隐性歧视问题。目前大部分 AI 的引用都依赖于从数据中习得知识, 但许多数据中却可能存在着隐性歧视问题。Amuasi 认为许多数据集缺乏多样性, 不能有效收集众多低或中低收入国家(LMICs)的信息。讲者以非洲为例展现了数据或算法中可能存在的机会不均等、性别或种族的隐形歧视、医疗偏差、隐私泄露等问题。最后, 讲者针对性地提出了许多解决方案, 并介绍了表征学习在 LMICs 医疗领域的应用。

Accelerating AI Systems: Let the Data Flow! 斯坦福大学电子工程与计算机科学系教授、SambaNova Systems 的联合发起人兼首席技术专家 Kunle Olukotun 报告了基于数据流概念设计的软硬件架构如何应对现代 ML 技术发展提出的难题。随着摩尔定律的速度逐渐衰减以及机器学习技术的日渐火热, 后摩尔时代的计算性能提升必须依赖于为机器学习和数据处理应用专门设计的硬件加速器。在未来, 这些应用的特点将包括 TB 级别的模型、数据的普遍稀疏以及不规则的控制流, 这将对传统 CPU 和 GPU 的能力产生严重挑战。在本报告中, Olukotun 将展示如何使用可重构的数据流加速器(RDAs)来提高具有这些特性的广泛数据密集型应用的性能。进一步的, Olukotun 还介绍了 SambaNova Systems 通过应用 RDA 技术开发的可重构数据流单元(RDUs)如何在许多具有挑战性的机器学习任务上实现创纪录的性能表现。

Leveraging AI for Science. DeepMind 首席科学家 Pushmeet Kohli 在报告中介绍了 DeepMind 近年来如何利用 AI 技术解决多个科学领域中极具挑战性的重要问题。现代科学的难点之一在于理解收集到的关于世界的大量信息。无论是大型强子对撞机还是大规模基因组计划, 任何个人都不可能理解所有的信息。DeepMind 的研究哲学为利用基于学习的方法来发现

智能的关键能力, 其以此准则来选择合适的科学问题, 发展通用且具备可解释性的 AI 解决方案。其后, 讲者通过分析结构生物学、基因组学、量子化学乃至纯数学等领域的几个成功案例来展现 AI 技术推进多个学科发展的实现逻辑与细节。

Petascale Connectomics and Beyond. 普林斯顿大学神经科学所与计算机科学系教授 H. Sebastian Seung 在报告中带来了神经连接组研究的最新进展。神经连接组是一种极具挑战性的大脑建模方法, 其将大脑连接表示为一个有向图, 其中节点是神经元, 边是突触。在上个世纪的 80 年代, 神经科学家通过分析电子显微镜图像手工重建了线虫的神经连接组, 然而随着研究对象的愈发复杂, 大脑图像的数据已经发展到了 TB 甚至是 PB 级别, 此时手工重建已经不具备可能性。因此, 科学家们开始应用卷积神经网络对电子显微镜图像进行分析。在本报告中, Seung 展示了目前 AI 技术如何帮助构建神经连接组模型并指出了其中存在的问题, 目前取得的成果表明连接组信息的爆炸性增长揭示了神经系统的固有结构, 这将有利于发展大脑的学习理论。

"Affordances" for Machine Learning. 澳大利亚国立大学社会学院高级讲师 Jenny L. Davis 在报告中介绍了一个针对机器学习的可供性分析框架。可供性(功能可见性)指的是一项技术的自身特征能够直接启发用户使用功能的总和, Davis 展示了一个可供性框架对于 ML 系统的分析和设计价值。具体而言, 讲者描述并应用了可供性机制和条件框架, 该框架模拟了一项 ML 技术请求、需求、鼓励、阻止、拒绝以及允许技术性和社会性结果的方式。讲者通过研究对 ML 系统进行批判性分析并设想性重制的案例, 展现了机制和条件框架不仅揭示了技术选择具有深刻的社会性, 而且还揭示了潜在的选择原理与选择对象。这种方法用语境中系统的特殊性取代了不确定性以及一般性主张, 使得从业者具备批判性思考的能力, 能同时考虑掌握的权利及相应的责任。

四、 热点论文

2022 年度共有以下 7 篇论文获杰出论文奖, 另有 3 篇获杰出论文提名^[2]。

Analytic-DPM: An Analytic Estimate of the Optimal Reverse Variance in Diffusion Probabilistic Models. 获奖理由: 扩散概率模型 (Diffusion Probabilistic Model, DPM) 是一类强大的生成模型, 是机器学习中一个快速发展的话题。本工作旨在解决 DPM 模型的固有局限性, 这种局限性为 DPM 中最优反向方差的计算缓慢且昂贵。作者首先给出了一个令人惊讶的结果, 即 DPM 的最优反向方差和相应的最优 KL 散度都有其得分函数的解析形式。之后他们提出了新颖而优雅的免训练推理框架: Analytic-DPM, 它使用蒙特卡罗方法和预训练的基于得分模型来估计方差和 KL 散度的分析形式。这篇论文在理论贡献(表明 DPM 的最优反向方差和 KL 散度都具有解析形式)和实际益处(提出适用于各种 DPM 模型的免训练推理)方面都很重要, 并且影响未来对 DPM 的研究。

Hyperparameter Tuning with Renyi Differential Privacy. 获奖理由: 本工作对学习算法差分隐私分析的一个重要盲点提供了新的见解, 即学习算法在数据上进行多次运行以调优超参数。作者指出, 在某些情况下, 部分数据可能会扭曲最优超参数, 从而泄露私人信息。此外, 作者在 Renyi 差分隐私框架下为超参数搜索过程提供了隐私保障。该工作考虑了学习算法的日常使用及其对社会隐私的影响, 并提出了解决方案。这项工作将为差分隐私机器学习算法的后续工作提供基础。

Learning Strides in Convolutional Neural Networks. 获奖理由: 本文讨论了使用卷积网络的研究者都面临的一个重要问题, 即以原则性的方式设置 stride, 这种根据原则性的方法忽略了可能的实验和试错。作者提出了一种新颖的、非常聪明的、可以用来学习 stride 的数学公式, 并展示了一种实用方法, 该方法在综合基准中实现了 SOTA 结果。文中主要思想是 DiffStride, 这是第一个具有可学习 stride 的下采样层, 它允许学习傅里叶域中裁剪掩码的大小, 以适合可微编程的方式有效地调整大小。该工作提出了一种可能成为常用工具箱以及深度学习课程一部分的方法。

Expressiveness and Approximation Properties of Graph Neural Networks. 获奖理由: 这篇理论比较

强的论文展示了如何将有关不同图神经网络(GNN)架构的表达性和可分离性的问题进行简化(有时通过检查它们在张量语言中的计算来大大简化), 其中这些问题与常见的组合概念有关, 例如树宽(Treewidth)。特别地, 本文提出通过 Weisfeiler-Leman (WL) 检验, 可以很容易地得到 GNN 分离力(Separation Power)的边界, 该检验已成为衡量 GNN 分离力的标准。该框架对通过 GNN 研究函数的逼近性有一定的指导意义。该工作通过提供描述、比较和分析 GNN 架构的通用框架, 有可能对未来的研究产生重大影响。此外, 该项工作提供了一个工具箱, GNN 架构设计人员可以使用该工具箱分析 GNN 的分离能力, 而无需了解 WL 测试的复杂性。

Comparing Distributions by Measuring Differences that Affect Decision Making. 获奖理由: 该研究提出了一类新的差异性(Discrepancy), 可以根据决策任务的最佳损失比较两个概率分布。通过适当地选择决策任务, 该方法泛化了 Jensen-Shannon 散度和最大平均差异族。与各种基准上的竞争基线相比, 该方法实现了卓越的测试性能, 并且具有广阔的应用前景, 可用于了解气候变化对不同社会和经济活动的影响、评估样本质量以及选择针对不同决策任务的特征。评审委员会认为该论文具有非凡的实验意义, 因为该方法允许用户在通过决策损失比较分布时直接指定其偏好, 这意味着实际应用将有更高的可解释性。

Neural Collapse Under MSE Loss: Proximity to and Dynamics on the Central Path. 获奖理由: 该研究对当今深度网络训练范式中普遍存在的 Neural Collapse 现象提出了新的理论见解。在 Neural Collapse 期间, 最后一层特征塌陷到类均值, 分类器和类均值都塌陷到相同的 Simplex Equiangular Tight Frame, 分类器行为塌陷到最近类均值决策规则。该研究没有采用在数学上难以分析的交叉熵损失, 而是提出一种新的均方误差损失分解, 以便分析神经塌陷下损失的每个组成部分, 这反过来又形成了一种新的中心路径(Central Path)理论构造, 其中线性分类器在整个动态过程中对特征激活保持 MSE 最优。最后, 通过探究沿中心路径的重归一化(Renormalized)梯度流, 研究者推

导出预测神经崩溃的精确动态。该研究为理解深度网络的实验训练动态提供了新颖且极具启发性的理论见解。

Bootstrapped Meta-Learning. 获奖理由：元学习具有增强人工智能的潜力，但元优化一直是释放这种潜力的巨大挑战，该研究为元学习开辟了一个新方向。受 TD 学习的启发，研究者提出一种从自身或其他更新规则引导元学习器的方法。该研究进行了透彻的理论分析和多项实验，在 Atari ALE 基准测试中为无模型智能体实现了新的 SOTA，并在多任务元学习中提升了性能和效率。

除了获奖论文外，还有 3 篇论文获杰出论文奖荣誉提名，分别是：(1) Understanding over-squashing and bottlenecks on graphs via curvature. (2) Efficiently Modeling Long Sequences with Structured State Spaces. (3) PiCO: Contrastive Label Disamb-

iguation for Partial Label Learning.

五、总结展望

2022 年 ICLR 中强化学习、深度学习、图神经网络、表示学习、自监督学习、联邦学习等领域保持高热度，元学习热度有所下降。ICLR 以其 Open Review 机制而出名，这一机制可以使得审稿过程更加透明化，方便了作者和审稿人之间的交流。

与此同时，初次设置的博客 Track 也带来了更多的可能与看点。这不仅构建了一个可以讨论过去论文、展示研究者工作的平台，还使得博客这一形式逐渐步入研究者的视野并被认可。通过众多高质量的博客，初涉某一领域的研究者可以快速了解领域进展，同一领域的研究者也可以展示他们独到的见解。希望这一新的赛道可以给科研工作者带来全新的体验与更多的收获。

责任编辑 魏秀参

参考文献

- [1] <https://mp.weixin.qq.com/s/JPgI7NFEcixbLbKHiTZ5A>
- [2] <https://iclr.cc/Conferences/2022>



钱鸿

华东师范大学计算机科学与技术学院副研究员，上海市“晨光学者”。主要研究方向为机器学习、零阶优化、强化学习等。

Email: yehj@nju.edu.cn

上海交通大学严骏驰副教授访谈

2022年6月2日,《CCF-CV专委简报》在线采访了上海交通大学严骏驰副教授。下面是采访实录。

严老师,您好!首先,请您分享一下您的个人学习和研究经历。

我于2008年毕业于北京科技大学自动化系,2011年硕士毕业于上海交通大学模式识别与智能系统专业,师从计算机视觉专家刘允才教授,亦得到施鹏飞教授的指导。毕业加入IBM中国研究院后,有幸在杨小康院长的引荐和指导下,2012-2015年在交大攻读信息与通信工程专业博士(定向委培),师从机器学习专家查宏远教授。博士论文主要研究图匹配这一NP难问题的算法设计,先后获得ACM中国优博提名和获CCF优博。

在IBM中国研究院的7年时间,从一线技术人员到团队负责人(9级首席专家),见证和参与了深度学习等新一代人工智能技术的落地浪潮,在2017年带领技术团队将基于深度学习和图像处理的面板缺陷检测技术率先在TCL华星光电落地。期间亦前往IBM美国沃森中心师从IBM“深蓝”的算法设计者Murray Campbell博士,进行机器学习与量子计算相关研究。

在2018年4月,我离开IBM,同时也结束了华东师范大学的近三年委培博后生涯,加入交大计算机系任特聘副教授,主要从事机器学习的研究。最近的兴趣则是在组合优化问题的AI求解,以及量子机器学习方面的交叉研究,旨在通过释放GPU乃至QPU的算力,

最终实现复杂的(例如NP难)决策问题的快速求解。

您致力于机器学习驱动的智能感知与决策研究,能否介绍一下您及您团队在此方面主要贡献?

从感知、预测与决策三方面总结过去的研究内容和成果,分别对应三项代表性工作:

感知:有向视觉目标的精细检测的系列深度学习模型。其中R3Det检测器获得PaperDigest根据引用统计数据评选的AAAI21最具影响力论文;发表的系列论文形成开源项目AlphaRotate,同时也成为视觉知名开源库OpenMMLab中旋转检测板块mmRotate主体。

预测:时序点过程机器学习框架和系列相关工作,并在IJCAI、SIGKDD等发起Tutorial进行技术推广;在IEEE TNNLS发起专刊,在交大开设点过程建模课程。

决策:机器学习求解图论和组合问题的系列相关工作,特别是图匹配问题的求解,获得吴文俊人工智能优秀青奖等荣誉。发布了开源Github代码库ThinkMatch。

近期则多聚焦在复杂决策问题的机器学习求解,包括量子线路亲和的图神经网络学习(SIGKDD22)、芯片布局布线辅助设计(NeurIPS21)、自动驾驶感知决策(Waymo比赛获得佳绩)等。这些工作的核心理念旨在研究更好的机器学习软硬件范式:一方面,将感知和决策、乃至控制在整体层面加以建模和学习,并充分利用信号处理、数理方程与机器学习的现有方法与技术;另一方面则考虑硬件计算架构,探索GPU、乃至在量子

计算亲和的机器学习决策范式，试图更好地适配现有和将来可能的硬件计算模型，提升决策收益和求解效率。

作为科技创新 2030 新一代人工智能重大(青年)项目负责人，能否介绍一下您的项目情况和感受？

我所承担的三年期的青年项目属于指南里的高级机器学习开放项目范畴，属于比较自由的理论研究。而我给自己定的题目是《组合优化问题的机器学习求解》，旨在通过机器学习技术，另辟蹊径，对组合优化这类复杂度高的运筹决策问题进行智能求解。这里的主要理念是通过数据驱动的机器学习方法，去替代传统人工设计的算法。同时利用机器学习、特别是深度学习充分释放 GPU 算力的优势，在求解速度方面也获得提升。

所提出的新型双层优化框架，以及在图匹配、图编辑距离、汉密尔顿回路等问题上一些具体研究成果已经在 ICML、NeurIPS、TPAMI 等发表。并进入了新一代人工智能重大项目年度进展的评选。

我深感机会的来之不易和身上承担的重任。作为项目负责人，除了做好日常的事务性组织协调和总结工作，同时也要促使自己多思考，打磨自己的学术理念和品位。个人而言，在前期研究基础上，我更愿意推动团队往前沿、有挑战的科学问题进行探索，并冒一定的风险。例如正在从事的量子机器学习、特别是量子图神经网络，以及 EDA 芯片设计的机器学习等内容。这些方面的研究，更能满足我的内在好奇心，形成持久驱动力。

您在 IBM 工作期间从事了多个智能感知、预测、决策项目的预研与落地，能否介绍下您的这些成果呢？另外，您在产业化方面是否有经验可以分享？

IBM 在 2010 年左右主推智慧城市，这与 2012 年以后深度学习的崛起有一定的巧合，同时也不能不说 IBM 对智能产品和解决方案的前瞻性。

身在 IBM 中国研究院这样一个市场驱动 (in-market research) 的研究院，使得我有机会主导了一些人工智能的工业项目，在 7 年期间的相关工作经历，

涉及感知、预测和辅助决策的各个方面。

由于 IBM 2B 的企业基因，我们多在行业解决方案上进行客户项目的研发和交付，并将一些功能进行抽象并入更标准的产品。在感知方面，从事了包括韩国 SK 能源、TCL 华星光电等客户在内的基于机器学习的异常、缺陷检测项目，特别是在精细视觉目标检测方面进行了较多的积累，为后续回归高校后的相关创新奠定了基础。在预测方面，从事了很多包括香港水务署在内的预防性维护保养项目的研发和落地。这期间也正好积累了在时序点过程方面的研究基础和成果。在决策方面，得益于 IBM 的 ILOG、CPLEX 等求解器与预测性人工智能模块结合应用，潜移默化中使得自己形成了在机器学习求解智能决策方面的研究理念。

在产业化方面，目前更多关注自动驾驶方面的研究，涉及前期感知、决策乃至控制一体化的研究的进一步深化。近期除了在 Waymo 比赛和论文上有所斩获，也正与相关自动驾驶企业深度合作，实现实车的大规模辅助驾驶的落地运营。这也和我自动化专业出身的背景较为一致，让我有一种回到老本行的如鱼得水的感觉。事实上，我的本科毕业论文就是做自主移动机器人的感知、决策与运动控制，后发表在 SCI 期刊 Optical Engineering 上。

能否介绍一下您的研究团队，以及您是如何管理和带领您的团队的？

作为一名青年教师，我的研究团队比较扁平化，主要是各个年级的博士生和硕士生组成，以及一些高年级的本科生作为后备力量。令人欣喜的是，目前各个年级段的博士生和硕士生同学都取得了很不错的成绩，这里列举几位博士生同学：杨学、李龙元、汪润中分别在视觉感知、时序预测、图论决策方面取得了代表性的成果。而另一位博士生吴齐天同学则在图学习、因果学习等方面获得了外界的认可，获得了众多学术荣誉。

事实上，在本科生的培养方面，我也愿意投入较多的精力。一方面是为自己的团队甄选人才，另一方面也

是尽快让他们获得一定的独立科研能力。一个成功的例子是早期带领两名本科生分别在其大四学年和直博一年级发表了第一作者 IEEE TPAMI 论文。做的就是我博士论文图匹配问题求解的延续。

我和学生一直是比较平等的关系，本着双赢的理念将适合的题目交给有兴趣或者有能力的个体同学，并鼓励同学之间做有一定差别的研究，培养学生独立科研的能力、学术品位和自信。从管理而言，除了做一些聚餐、郊游等简单团建，更多是从学术理念、研究方向和学术观点的角度来塑造实验室的认同感和共同愿景，也尊重彼此的不同，打造团队形成凝聚力和张力的兼而有之。

您担任中国图象图形学会视觉大数据专委会副秘书长、多个 SCI 期刊的编委和学术会议的大会主席、领域主席，请问您是如何协调科研和学术服务工作的时间？您是如何看待这些服务和兼职工作？

学术服务有助于更好的与同行进行交流，向业内资深人士学习。当会议的领域主席需要深入阅读所负责的每篇论文和审稿意见，并和审稿人讨论碰撞。尤其是在疫情时代，更是一种很好的远程交流机会。事实上，欣赏、质疑、思辨同行的工作，本身也是科研的一部分。

您围绕感知与决策的机器学习，发表第一作者/通讯作者 TPAMI/IJCV/IKDE 论文 14 篇，第一作者 CCF-A 类论文 12 篇。您是如何做到论文高质量产出的？您在论文撰写和发表方面有什么经验和建议？

这些论文，更多是我从 IBM，到华师大、再到交大期间的合作者和优秀学生的愉快科研合作的见证。我的实验室发展到现在，在高水平论文发表这件事情上，一些学生、特别是早期阶段的研究生和想出国深造的本科生，可能比我自己具有更强烈的意愿。

相对而言，我个人目前把更多的精力投入到芯片辅助设计、量子机器学习求解复杂问题等方面的实质突破。总体而言，营造实验室的严肃而又活泼的学风，保持研究的品质和平常心，不失为一个可持续的方案。

您的博士生获得了 MSRA 奖学金（全球 11 人）、百度奖学金（全球 10 人）、CCF-CV 新锐奖（全国 3 人）、交大学术之星提名等荣誉，这与您的培养是分不开的，能否介绍一下在学生培养方面的经验呢？

我也试图用自己读博士的经历来启发我的学生。和他们一起找一个点作为突破，做出深度和特色。在此基础上，和目前的热点技术做一定的交叉，保持自己的特色的同时又与时俱进。这些奖学金只是开始，我也衷心希望他们不久都能脱离我的指导，展翅高飞，独当一面。

如果吐露研究工作者心声，最想说的的是什么？

人工智能给了我们的研究人员无限的可能，也正与其他学科进行更全面和深入的交叉。我相信在感知、预测、决策、控制乃至科学发现的各个方面，都将是人工智能研究者和从业人员的舞台。

责任编辑 赵振兵 余烨



严骏驰

严骏驰，上海交通大学博士生导师，长聘轨副教授。交大学生创新中心人工智能基地主任，科技创新 2030 新一代人工智能、自然科学基金、教育部 AI 课程资源建设等项目负责人。主要研究兴趣为机器学习，特别是组合优化问题的机器学习求解和量子机器学习。曾任 IBM 中国研究院认知物联首席专家。发表 CCF-A 类论文 100 余篇，谷歌学术引用 6400 余次。任 ICML/NeurIPS/CVPR/AAAI 等会议领域主席，Pattern Recognition 编委。获 CCF 优秀博士论文、吴文俊人工智能优秀青年奖等。曾指导博士生获百度/MSRA 奖学金及 CCF-CV 新锐奖。

委员好消息

2022年4月12日获悉，由CCF-CV专委会委员**夏桂松**（武汉大学）、**白翔**（华中科技大学）和**卢孝强**（中国科学院西安光学精密机械研究所）等合作完成的论文AID: A Benchmark Data Set for Performance Evaluation of Aerial Scene Classification（发表于《IEEE Transactions on Geoscience and Remote Sensing》2017年55卷第7期）获IEEE地球科学与遥感学会2021年度最高影响力论文奖。该文的主要贡献是收集并标注了上万张航空场景图像，构建了航空影像数据AID（Aerial Image Dataset）。

2022年4月14日，爱思唯尔(Elsevier)正式发布了2021“中国高被引学者”(Highly Cited Chinese Researchers)榜单，共4701名各学科最具全球影响力的中国学者入选，CCF-CV专委会44位委员上榜：北京大学**林宙辰**、**彭宇新**，北京邮电大学**邓伟洪**，重庆邮电大学**高新波**，大连理工大学**卢湖川**、**王栋**，复旦大学**姜育刚**，国防科技大学**刘丽**，哈尔滨工业大学**徐勇**、**左旺孟**，华中科技大学**白翔**、**尤新革**，杭州电子科技大学**俞俊**，江西财经大学**方玉明**，南京大学**王利民**、**吴建鑫**，南京航空航天大学**谭晓阳**，南京理工大学**李泽超**、**唐金辉**，南开大学**程明明**，清华大学**郭振华**、**鲁继文**，上海交通大学**卢策吾**、**马超**，天津理工大学**陈胜勇**，同济大学**张林**，武汉大学**荆晓远**、**夏桂松**，西安电子科技大学**邓成**、**董伟生**，西安交通大学**孙剑**，西北工业大学**程焱**、**韩军伟**、**聂飞平**、**王琦**，厦门大学**纪荣嵘**，中山大学**林惊**、**郑伟诗**，中科院西安光机所**卢孝强**，中科院信工所**操晓春**（现中山大学），中科院深圳先研院**乔宇**，中科院

自动化所**雷震**、**王亮**、**张天柱**（现中国科学技术大学）。

2022年4月12日，CCF高级会员评审会（上半年）在北京举行，最终评选出339位高级会员，CCF-CV专委会9位委员获评高级会员：北京航空航天大学**徐迈**、北京交通大学**丛润民**、电子科技大学**姬艳丽**、广东工业大学**费伦科**、哈尔滨工业大学（深圳）**何震宇**、河北大学**刘帅奇**、合肥工业大学**贾伟**、华北电力大学**张轲**、中科院自动化所**董晶**。

2022年4月24日，北京图象图形学学会公布了优秀博士学位论文及优秀导师评选结果，CCF-CV专委会委员中国科学院自动化研究所**赫然**等指导完成《基于受限样本的图像转换方法研究》、北京大学**彭宇新**指导完成《基于知识迁移的跨媒体检索与推理方法研究》获优秀博士学位论文及优秀导师，北京航空航天大学**刘祥龙**等指导完成《面向图像识别的对抗机器学习方法的研究》获优秀博士学位论文提名及优秀导师提名。

2022年4月28日，百度人才智库(TIC)、百度学术、天津大学复杂管理系统实验室、中国科大金院商业智能中心联合发布了全球首份AI华人青年学者榜单，榜单从全球范围内评选出150名在AI领域贡献出众的华人青年。CCF-CV专委会**白璐**（中央财经大学）、**赵洲**（浙江大学）、**郭裕兰**（国防科技大学）、**黄岩**（中国科学院自动化研究所）、**潘金山**（南京理工大学）、**任文琦**（中国科学院信息工程研究所）、**王利民**（南京大学）、**朱翔昱**（中国科学院自动化研究所）等8位委员入选。

责任编辑 刘海波

手绘草图领域开源代码

西安电子科技大学 汪思远 朱光明 张亮

手绘草图作为人际间交流的方式和艺术表达的形式，在绘制过程中，不使用任何绘画技巧进行绘制。草图不同于传统图像：草图由笔划的时间序列组成，而传统图像则是对现实世界的静态复制。与常规图像相比，草图具有极高的抽象性，多样性，稀疏性和不变性，这使得草图建模具有独特性和挑战性。手绘草图领域任务主要分为识别，检索，生成，局部分析和简化抽象任务。本文重点介绍识别，检索和生成任务中的研究成果。

1、Multigraph Transformer

工作：该文针对于手绘草图的特点，将其表示为图的形式，设计了针对图的 transformer 网络架构对手绘草图进行特征提取，并进行识别。该方法处理手绘草图时，共分为四个阶段，标记阶段，输入处理阶段，特征提取阶段和分类阶段。

标记阶段对手绘草图的点集序列中的数量进行统一化并添加标识位，保留手绘草图中的笔画顺序。对于手绘草图中每一个点(x,y),通过标记阶段添加标识位后，每个点转换为一个四维向量(x,y,f,p)。在四维向量中，f 标识当前点在手绘草图中的状态，可取三个状态， f_1 表示当前点是一个笔画的非终止点， f_2 表示当前点是一个笔画的中止点， f_3 表示当前点为因手绘草图绘制点的数量不够，扩充而来的点，p 表示当前点在整个手绘草图输入序列中的位置。输入处理阶段通过线性变换对标记阶段数据进行升维。特征提取阶段通过 multigraph multihead attention 机制对手绘草图进行特征提取，multihead attention 机制结构图如图 1。

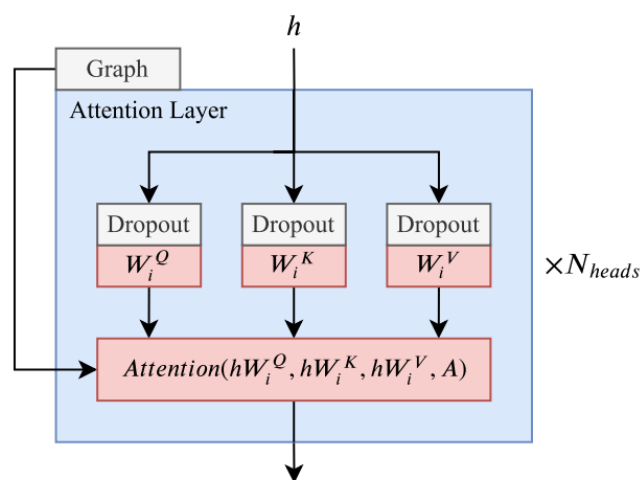


图 1 multihead attention 机制示意图

其中 multigraph 根据对手绘草图中笔画间和笔画内点集序列的不同连接构成对同一手绘草图的多图生成。对于笔画内的连接，通过 k-hop 进行区分并连接，即连接笔画的内的点的 k-近邻点，如图 2 中，(a)表示原始的标绘，(b)表示 1-hop 连接，(c)表示 2-hop 连接。对于笔画间的点的连接，直接连接笔画的结束点和下一笔画的开始点。

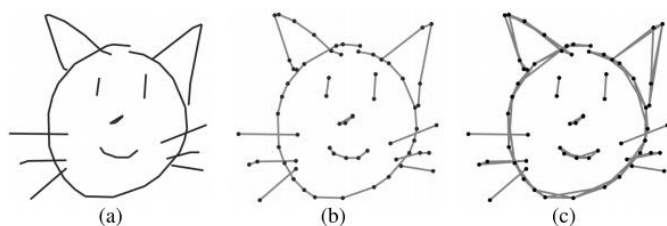


图 2 手绘草图 multi-graph 形式

分类阶段根据提取的特征直接进行分类。该文在当前手绘草图识别领域内应用最多的数据集 QuickDraw 上进行了大量实验，这些实验证明了 Multigraph Transformer 在手绘草图识别任务中的有效性。Multigraph Transformer 的网络结构图如图 3 所示。

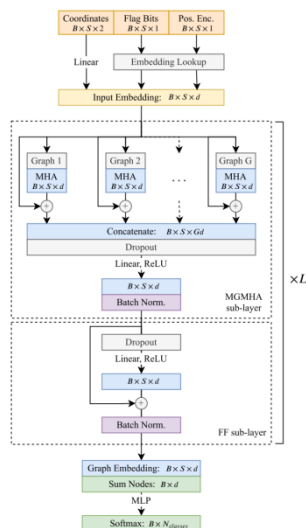


图 3 Multigraph Transformer 结构图

更多有关 Multigraph Transformer 的详细内容可参考发布该方法的论文“Multigraph Transformer for Free-hand Sketch Recognition”。

论文地址: <https://ieeexplore.ieee.org/document/9397867>

代码地址: https://github.com/PengBoXiangShang/multigraph_transformer

2、Modality-Aware Triplet Hard Mining (MATHM)

工作: 目前手绘草图检索方面大多数的做法是使用神经网络分别对手绘草图和真实图像进行特征提取，进而进行特征匹配，使得同一类别两个模态的提取到的特征具有相同特点。该文采用度量性学习的方法针对多模态（本问题是手绘草图和真实图像组成的二模态问题）中进行特征提取时类内特征的紧凑性和类间特征的差异性，提出了三种度量的方式，跨模态度量，单模态度量和联合度量。在训练阶段上述三部分通过可训练的权重进行合并，得到最终的特征。因为在训练时同时考虑了三种度量方式，最终训练的特征结果如图 4 所示。

在图 4 中可以看到基准的网络训练的特征结果只考虑提高，同一类物体图像和手绘草图特征的关联性，降低不同类图像和草图的相似性。而 MATHM 则考虑了同一类手绘草图内部特征的关联性，同一类图像，草图特征的相似性，同时增大其和其他类的差异性。图 4 中不同颜色待变不同类物体，星号表示手绘草图，圆圈表示图像，绿色的线表示模态间联系，蓝色的线表示模态内联系。通过 MATHM 取得的特征，可以更好的进行处理 zero-shot 问题。

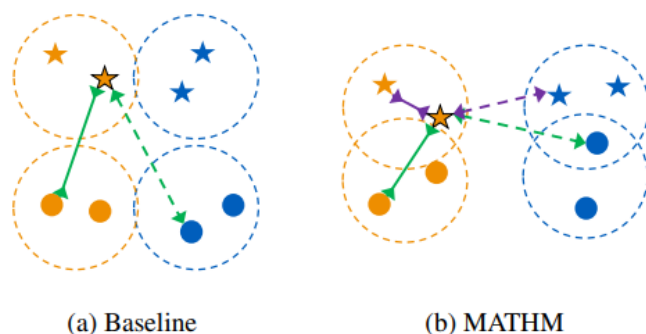


图 4 基准网络特征和 MATHM 特征对比

该文提出的方法同时考虑了样本模态间的差异和类别差异，取得了较好的效果。Modality-Aware Triplet Hard Mining 的网络结构图如图 5 所示。更多有关 Modality-Aware Triplet Hard Mining 的详细内容可参考发布该方法的论文“Modality-Aware Triplet Hard Mining for Zero-shot Sketch-Based Image Retrieval”。

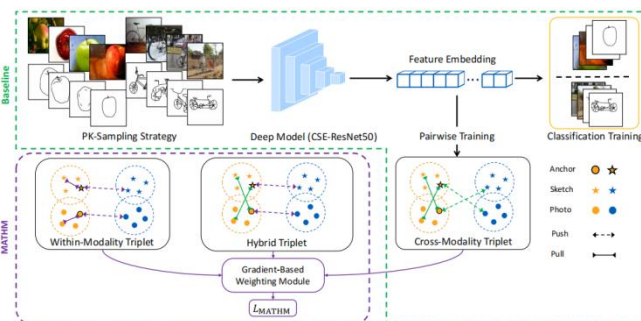


图 5 MATHM 结构图

论文地址: <https://arxiv.org/abs/2112.07966.pdf>

代码地址: <https://github.com/huangzongheng/MATHM>

3、Semantic Scene Completion

工作: 三维语义场景补全任务需要预测三维空间中的像素表示, 同时也需要对每个像素的类别进行分类。在三维语义场景补全任务中, 场景的分辨率成为制约补全效果的瓶颈。该文通过草图场景中的各个物体的结构信息和标准 RGB 图像中场景的细节信息对三维语义场景进行补全, 有效的去除了低分辨率对于三维场景补全效果的影响。该文提出的三维场景补全框架包含三维草图视觉模块和 RGB 图像特征提取模块。三维草图视觉模块使用 Conditional Variational Autoencoder(CVAE)对三维草图场景进行重建, 对三维草图场景的分布进行建模, 提取三维场景中的结构信息。RGB 图像提取特征模块提取三维场景中的细节信息, 通过融合 2 个模块提取的 2D/2.5D 的特征, 进行三维语义场景重建。在三个公共的基准数据集, NYU Depth V2, NYUCAD 和 SUNCG 上均取得当前最好的效果。其网络结构图如图 6 所示。

更多有关 Taming Transformers 的详细内容可参考论文“3D Sketch-aware Semantic Scene Completion via Semi-supervised Structure Prior”。

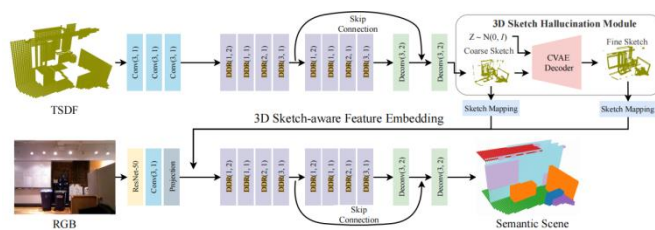


图 6 Semantic Scene Completion 网络结构图

论文地址: https://openaccess.thecvf.com/content_CVPR_2020/html/Chen_3D_Sketch-Aware_Semantic_Scene_Completion_via_Semi-Supervised_Structure_Prior_CVPR_2020_paper.html

代码地址: <https://charlesCXK.github.io>

责任编辑 李策 贾同



汪思远

博士研究生, 西安电子科技大学计算机科学与技术学院, 研究方向为计算机视觉。



朱光明

副教授, 博士生导师, 西安电子科技大学计算机科学与技术学院从事教学与科研工作。研究方向为计算机视觉与图像处理、手语识别。

个人主页: https://faculty.xidian.edu.cn/ZGM/zh_CN/index.htm



张亮

教授, 博士生导师, 西安电子科技大学计算机科学与技术学院从事教学与科研工作。研究方向为计算机视觉与图像处理、三维场景语义分割。

个人主页: https://faculty.xidian.edu.cn/ZL_16/zh_CN/index.htm

ADNI 数据集

东北大学 贾同 贾娜娜

ADNI (Alzheimer's Disease Neuroimaging Initiative) 数据集是目前研究阿尔茨海默症的权威数据中心, 在 2004 年由美国国家卫生研究所和国家老年问题研究所共同资助创建而成, 致力于收集阿尔茨海默症病人数据, 跟踪病人的发病过程, 发掘发病过程的变化与起因, 以揭示阿尔茨海默症的发病原理, 寻找对应的治愈方案。

ADNI 数据集包括从轻度认知障碍 (MCI) 到阿尔茨海默症 (AD) 共 1721 个病例。ADNI 数据目前分为四个阶段, ADNI-1、ADNI-GO、ADNI-2 和 ADNI-3, 其中 ADNI-1 与 ADNI-GO 为基线数据, ADNI-2 与 ADNI-3 主要为后续跟踪数据和新加入的模态数据。该数据集是一个多模态数据集, 数据集包括以下几部分: Clinical Data (临床数据); MR Image Data (磁共振成像); PET Image Data (正电子发射计算机断层扫描); Genetic Data (遗传数据); Biospecimen Data (生物样本数据)。ADNI 的关键目标是提供将遗传学与影像学 and 临床数据相结合的机会, 以帮助调查疾病的机制。

1、ADNI 阶段概述

ADNI 的数据目前分为四个阶段, ADNI-GO、ADNI-1、ADNI-2 和 ADNI-3, 其中 ADNI-GO 与 ADNI-1 为基线数据, ADNI-2 与 ADNI-3 主要为后续跟踪数据和新加入的模态数据(如新型示踪剂下的 PET)等。

ADNI 1

a) 开发改进的方法, 为获取老年痴呆症(AD)、轻度认

知障碍(MCI)和对照患者(CN)的纵向、多位点 MRI 和 PET 数据制定统一标准

b) 构建可访问的数据库, 描述大脑结构和代谢的纵向变化, 以及临床/认知和生物标志物数据

c) 在涉及这些患者的试验中, 开发出能够最大限度地确定治疗效果的方法

d) 测试基于临床和生物标记数据的一系列假设

ADNI GO

a) 通过招募 200 名 AD 症状最轻微的患者(EMCI)来定义和描述 AD 之前的 MCI 阶段

b) 采集 ADNI1 和 EMCI 分组内的 CN 和 LMCI 受试者的 F18 淀粉样蛋白成像。FDG 和 PET 的采集将与 F18 淀粉样蛋白成像的采集联合进行

c) 建立一个全国性的 F18 淀粉样蛋白成像网络, 测试关于脑淀粉样蛋白积累的流行程度和严重程度及其与临床状态、MRI、FDG-PET、CSF 和 ADNI1 血浆生物标志物的变化的关系的假设

d) 在基线、第 3 个月、第 6 个月和第 12 个月收集所有新入组的受试者的 3T MRI

e) 对来自 ADNI1 的约 500 名 LMCI 和认知正常的受试者进行纵向临床/认知和 1.5T MRI 研究, 为期 2 年

f) 收集和分析所有新入组的 EMCI 和随访受试者的血液和脑脊液生物标志物

g) 采集血液样本进行 DNA 和 RNA 提取。新入组的受试者还将采集细胞永生化和 APOE 基因分型的样本

ADNI 2

- 随着病理的发展，确定阿尔茨海默氏病(AD)的临床、影像学、遗传和生化生物标志物特征之间的关系
- 为 AD 的神经科学提供信息，识别可用于临床试验的诊断和预后标记和结果度量，并帮助开发最有效的临床试验方案
- 制定统一标准，获取 AD、MCI 和老年人对照患者纵向多位点 MRI 和 PET 数据对 550 名新入组的受试者进行纵向临床、认知、MRI、PET (18F-Florbetapir 和 FDG)和血液和脑脊液生物标志物研究，并对 ADNI1 和 ADNI 的约 700 名受试者继续研究 5 年
- 采集血液样本进行 DNA 和 RNA 提取，新入组的受试者还将采集细胞永生化和 APOE 基因分型的样本
- 通过参加尸检的 ADNI1、ADNI GO 和 ADNI2 参与者的神经病理学检查，验证临床诊断、影像学和生物标志物替代物

ADNI 3

- 认知和相关生物标志物的纵向变化
- 认知衰退的预测
- 通过将结果与标准的临床测量和病理学相关联，验证基线和纵向获得的生物标志物测量结果
- 确定注意认知下降和 AV-1451 PET (tau PET)的最佳结果措施，认知下降的预测因子，认知正常参与者(二级临床前 AD 试验)、MCI 患者(前驱症状 AD 试验)和 AD 引起的早期痴呆患者的临床试验的纳入/排除标准
- 确定在 AD 大脑和基因中发现的其他已知疾病蛋白以及新发现的基因、蛋白质和分析物的作用，这些分析物提供了关于 AD 发病/诊断的有用信息

2、ADNI 数据集数据类型概述

Clinical Data (临床数据)

ADNI 临床数据集包含每个受试者的临床信息，包括招募、人口统计、体检和认知评估数据。全套临床数据可以作为逗号分隔值(CSV)文件批量下载。

下表总结了收集的临床数据表格。左侧为人口统计、神经学检查、筛选实验室、生命体征、认知评估、生物样本采集、药物、诊断总结、腰椎穿刺。

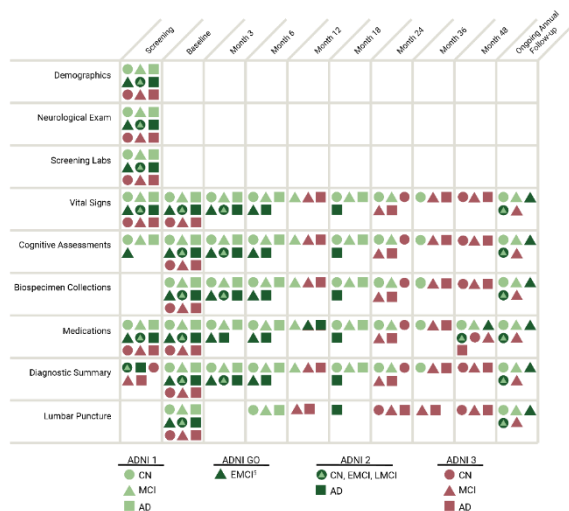


图 1 临床数据表格

MRI Image Data (磁共振成像)

磁共振成像是断层成像的一种，它利用磁共振现象从人体中获得电磁信号，并重建出人体信息。磁共振成像技术与其它断层成像技术(如 CT)有一些共同点，比如它们都可以显示某种物理量(如密度)在空间中的分布；同时也有它自身的特色，磁共振成像可以得到任何方向的断层图像，三维图像，甚至可以得到空间-波谱分布的四维图像。

磁共振成像无骨性伪影，可随意作直接的多方向(横断、冠状、矢状或任何角度)切层，对颅脑、脊柱和脊髓等的解剖和病变的显示，尤优于 CT，磁共振成像借其“流空效应”，可不用血管造影剂，显示血管结构，故在“无损伤”地显示血管(微小血管除外)，以及对肿块、淋巴结和血管结构之间的相互鉴别方面，有独到之处。磁共振成像有高于 CT 数倍的软组织分辨能力，它能敏感地检出组织成分中水含量的变化，故常可比 CT 更有效和早期地发现病变。通过磁共振血流成像技术的研究获得的进展，使在活体上测定血流量和血流门控的使用，使磁共振成像能清楚地、全面地显示心脏、心肌、心包以及心内的其他细小结构，为无损地检查和诊断各种获得性与先天性心脏疾患(包括冠心病等)，以及心脏功能的检查，提供了可靠的方法。又实现了磁共振成像

和局部频谱学的结合（即 MRI 与 MRS 的结合），以及除氢质子以外的其他原子核如氟、钠、磷等的磁共振成像，这些成就将能更有效地提高磁共振成像诊断的特异性，也开阔了它的临床用途。

下表概述了整个 ADNI 研究中收集的 MRI 数据。

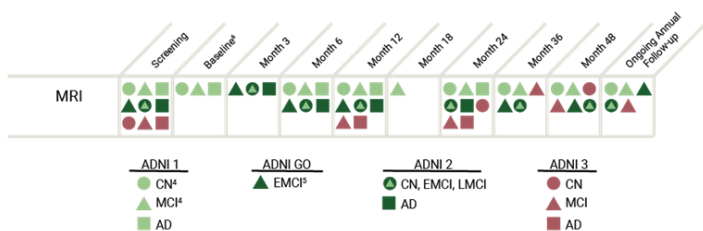


图 2 MRI 数据汇总

PET Image Data (正电子发射计算机断层扫描)

正电子发射计算机断层扫描全称为：正电子发射型计算机断层显像（Positron Emission Computed Tomography, 简称 PET），是核医学领域比较先进的临床检查影像技术。

正常范围 PET 特别适用于在没有形态学改变之前，早期诊断疾病，发现亚临床病变以及评价治疗效果。目前，PET 在肿瘤、冠心病和脑部疾病这三大类疾病的诊疗中尤其显示出重要的价值。

ADNI 数据集中的 PET 数据包括原始、预处理和后处理图像文件、PIB(ADNI1)、FDG(ADNI1/GO/2)、FLORBETAPIR (ADNI GO /2/3)、FLORBETABEN(ADNI3)和 TAU 成像(ADNI3)。

这些图像的收集对于满足 ADNI 开发生物标志物以跟踪阿尔茨海默病的进展和潜在病理学变化目标至关重要。下表概述了整个 ADNI 研究中收集的 PET 数据。

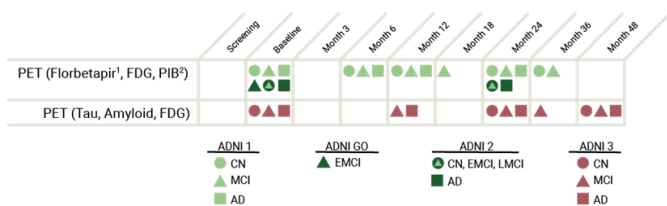


图 3 PET 数据汇总

ADNI 1/GO: ADNI1 患者最初接受的是 PIB32

扫描，而不是 Florbetapir 扫描。然而，由于处理时间限制，该协议在研究结束前进行了修改，包括采用 Florbetapir 扫描而不是 PIB 扫描。因此，PIB 仅用于 ADNI 1 中的一小部分参与者（约 100 名）。对于继续进入 ADNI GO 然后进入 ADNI2 的 ADNI1 CN 和 MCI 参与者，初始 PET 扫描发生在最后一次成功 Florbetapir 之日起两年后和 ADNI GO 下的 FDG PET 扫描。在 ADNI1 中，只有一部分参与者接受了 FDG。

ADNI 2: Florbetapir (以前的 AV-45) 1 PET 和 FDG PET 成像对所有新入组的参与者分别进行了两天（至少 12 小时的时间间隔）。在基线的临床评估前两周或后两周内和基线后 24 个月进行扫描。ADNI 2 名受试者进行了最多 3 次 florbetapir 扫描和最多 2 次 FDG 扫描，每次以 2 年的时间间隔获得。

ADNI 3: 将在所有参与者的初次 ADNI 3 就诊时对其进行 Tau 和淀粉样蛋白成像。随后的 Tau 成像计划取决于淀粉样蛋白状态：80% 的淀粉样蛋白阳性和 20% 的淀粉样蛋白阴性的参与者将进行 3 次额外的 tau 扫描，而其余 80% 的淀粉样蛋白阴性和 20% 的淀粉样蛋白阳性参与者将进行 1 次额外的 tau 扫描。见图表)。淀粉样蛋白 PET 成像将使用 florbetapir (从 ADNI2 继续的参与者) 或 florbetaben (新注册的参与者) 每 2 年进行一次。FDG-PET 成像仅在基线时对 MCI 和 AD 参与者进行 (见图 3)。

Genetic Data (遗传数据)

遗传因素在阿尔茨海默症中起重要作用。因此，ADNI 的一个关键目标是为研究人员提供将遗传学与成像和临床数据相结合的机会，以帮助研究疾病的发病机制。已经为 ADNI 1、ADNI GO、ADNI 2 受试者生成了基因分型和测序数据供 ADNI 研究人员使用。

ADNI 数据库可用基因分型数据有 ApoE 和 TOMM40。ApoE、TOMM40 和血管风险因素在淀粉样蛋白产生和清除过程中起重要作用。同时大量临床研究发现，淀粉样蛋白的沉积过程与认知功能的进行性变化也密切相关。

ApoE (载脂蛋白 E) 是血浆中重要的载脂蛋白之一, 参与体内脂质代谢和胆固醇代谢。ApoE 蛋白功能: 胆固醇转运和再分布、免疫调节, 细胞增殖分化, 载脂蛋白 APOE 基因型分类神经细胞修复等。ApoE4 可以通过加速脑血管病变, 来推进阿尔茨海默疾病过程, 而且这一过程与 β 淀粉样蛋白的沉积无关, 抑制这一过程可以缓解阿尔茨海默病的进展。

风险基因 TOMM40 (线粒体外膜 40 的反式) 的基因分型由 Polymorphic DNA Technologies 进行。TOMM40 定位于 19q13, 是外膜转移酶形成的复合体, 是在蛋白进入线粒体的重要入口通道形成的亚基, 它对于蛋白质导入线粒体必不可少, 与 AD 密切相关。

Biospecimen Data (生物样本数据)

ADNI 的目标之一是收集参与者的生物样本, 包括血液、尿液和脑脊液(CSF)。生物标志物或生物样本是生物状态的物质、测量值或指示物。生物标志物可能在临床症状出现之前就已经存在。ADNI 使用各种生物标志物来帮助预测阿尔茨海默病的发作。ADNI 数据集所收集的生物标志物样本主要有以下几类。

- 同型半胱氨酸
- 异前列腺素的种类
- 脑脊液 tau 蛋白、BACE 水平和酶活性
- 血浆 $A\beta_{40}$ 和 $A\beta_{42}$
- 基于正在进行的多重免疫测定研究和质谱 MRM 研究的其他可用的 CSF 和血浆



贾同

东北大学信息科学与工程学院教授、博士生导师, 智能感知与机器人研究所所长。研究方向为计算机视觉, 模式识别与机器学习等。电子邮箱: jiatong@ise.neu.cn



贾娜娜

博士研究生, 东北大学信息科学与工程学院, 研究方向为医学影像处理
电子邮箱: 2010284@stu.neu.edu.cn

曲线表示在 AD 过程中 $A\beta$ 、Tau 介导的神经元损伤和功能障碍、脑部结构、记忆力、临床功能这五种生物标志物从正常到异常的变化 (从正常认知到痴呆)。

- 1) β -淀粉样蛋白 ($A\beta$) 在脑脊液中或通过淀粉样蛋白 PET 成像测量
- 2) 由脑脊液中测量的 tau 蛋白或通过 FDG-PET 测量的突触功能障碍指示的神经退行性变
- 3) 脑萎缩, 主要在内侧颞叶, 通过结构 MRI 测量
- 4) 记忆力减退, 通过认知测试测量
- 5) 临床功能, 由认知测试测量的一般认知能力下降指示

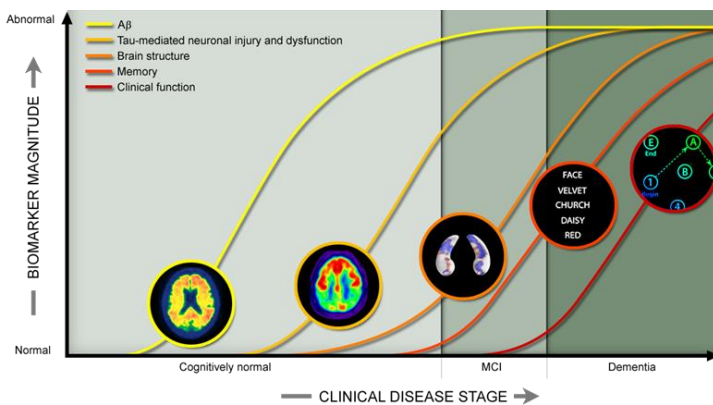


图 4 生物标志物在临床疾病中的阶段变化

3、数据集地址

ADNI 数据集地址为: adni.loni.usc.edu

责任编辑 沈沛意 李策

好文推荐

西安电子科技大学 MIRACLE 团队在“深度度量学习”方向的最新成果发表在 IEEE TPAMI 上。

论文: Xinyi Xu, Zhengyang Wang, Cheng Deng, Hao Yuan, and Shuiwang Ji, Towards Improved and Interpretable Deep Metric Learning via Attentive Grouping, IEEE Transactions on Pattern Analysis and Machine Intelligence.

Paper available on ArXiv:

<https://arxiv.org/abs/2011.08877>.

Code available on Github:

<https://github.com/XinyiXuXD/DGML-master>.

深度度量学习 (DML) 旨在学习一个神经网络, 将数据从高维的原始空间投影到低维的度量空间中。其优化目标是使度量空间中的同类样本尽可能拉近, 而异类样本尽可能推远。深度度量学习因其能够学习到有效的特征表示而具有广泛的应用, 如人脸验证、图像检索、以及细粒度分类等。

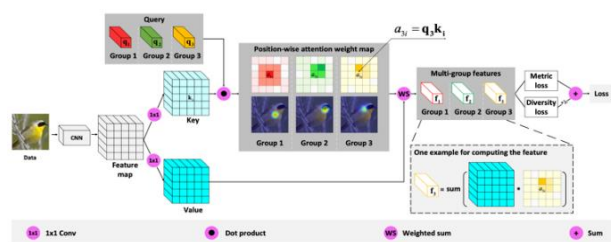


图 1 IDGML 整体框架

现有深度度量学习方法因其只考虑单一的特征空间嵌入而遭受维度饱和问题, 即性能的提升随着特征维度的增加而饱和。为此, 西安电子科技大学 MIRACLE 团队提出一种基于注意力机制的可解释深度分组度量学习算法 (Interpretable Deep Grouping Metric Learning Algorithm based on the Attention, IDGML), 旨在学习样本的多组特征表示, 并挖掘特征

之间的组内和组间关系, 从而提升特征表示能力并赋予模型可解释性。图 1 给出了 IDGML 的框架。如图所示, 样本图像输入模型后, 经过一个卷积神经网络后得到样本的三维特征图。基于该特征图, IDGML 利用基于注意力机制的分组模块生成样本的多组特征表示。最后通过组间和组内关系的约束对度量网络进行训练。IDGML 能够根据每组特征所编码的语义信息, 有效地捕捉样本的多种特征, 如鸟的颈部特征和嘴部特征。

IDGML 的核心组件是 A-分组模块。A-分组模块为每个组计算一个可学习的查询向量, 并通过计算查询与键矩阵的空间位置之间的相互注意力分数来获得不同位置的重要性分数。该查询是完全可训练的。特别地, 当 IDGML 的注意力方法与度量损失相结合时, 可以捕获分布在重要位置上的信息。此外, 当它与多样性损失相结合时, 可以捕获特定于组的信息。

IDGML 在三个基线数据集与现有的 DML 方法进行了实验对比。定量结果表明, IDGML 在所有的数据集上均优于现有的方法, 可以显著地提升所学特征的判别力和可解释性。定性结果表明 IDGML 能够捕捉到输入图像的多个重要位置上的特征信息, 如图二所示, 组 1 关注腹部, 组 2 关注嘴部, 组 3 关注颈部。

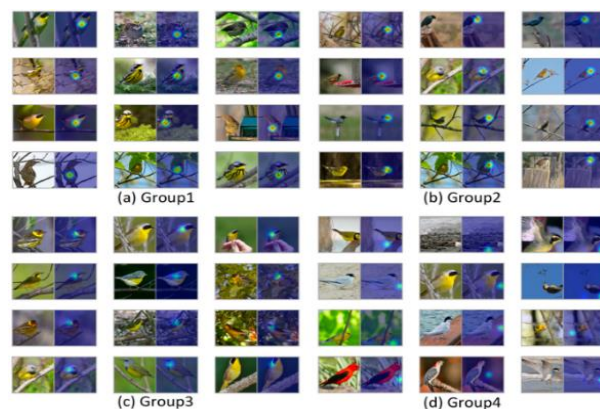


图 2 CUB-200-2011 数据集上的可视化结果

责任编辑 樊鑫 贾同

好文推荐

大连理工大学团队提出“SCI: 快速、灵活、鲁棒的低光照图像增强方法”，该成果发表在 CVPR 2022。

论文: Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo, Toward Fast, Flexible, and Robust Low-light Image Enhancement, CVPR 2022

低光照图像增强任务是图像处理领域的经典任务，在学术界和工业界得到了广泛的关注。现有的低光照图像增强技术主要基于构建数据驱动的深度网络，但通常情况下由于网络模型复杂，导致计算效率和推理速度偏低，并且对训练数据分布的依赖性导致其在未知场景下的性能缺乏保障。针对上述问题，大连理工大学团队从学习策略角度，构建了一种快速、灵活和鲁棒的低光照图像增强方法(SCI)，SCI 的主要贡献如下：

1) 开发了一个自校准的模块用于权重共享的光照学习，以赋予每个阶段的结果之间的收敛，提高曝光稳定性并大大减少计算负担，这是第一个利用学习过程加速低光图像增强算法的工作。

2) 定义了无监督训练损失来约束各阶段在自校准模块作用下的输出，赋予对不同场景的适应能力。并且属性分析表示 SCI 具有操作不敏感的适用性和模型无关的一般性，是现有研究中没有发现的。

3) 大量实验验证了 SCI 的优越性，在暗面检测和夜间语义分割方面也进行了进一步的应用。SCI 在基于网络的低光图像增强领域重新定义了视觉质量、计算效率和下游任务性能的峰值点。

SCI 算法流程图如图 1 所示。该模型主要包含三个核心组件，分别为权重共享的光照学习、自校准模块和无监督损失函数。该模型首先建立了一个具有权重共享的级联光照学习过程，考虑到级联模式的计算负担，构建了自校准模块，该模块可以实现每个阶段结果之间的收敛，产生仅使用单个块进行推理的增益，这样大大降低了级联模块所带来的计算成本，最后定义了无监督的损失函数，用于提升适应一般场景的模型能力。在训练阶段，SCI 由光照估计和自校准模块组成。将自校准模块映射加入到原始低光输入中作为下一阶段光照估计的输入。需要注意的是，这两个模块在整个训练过程中分别是共享参数。在测试阶段，只使用单个光照估计模块。

SCI 在图像质量和推理速度方面均取得了突破，为低光照图像增强任务的解决提供了一种新的视角，即如何在有限资源下赋予网络模型更强的刻画能力，该种视角相信也能够为其他相关视觉增强领域带来启发。

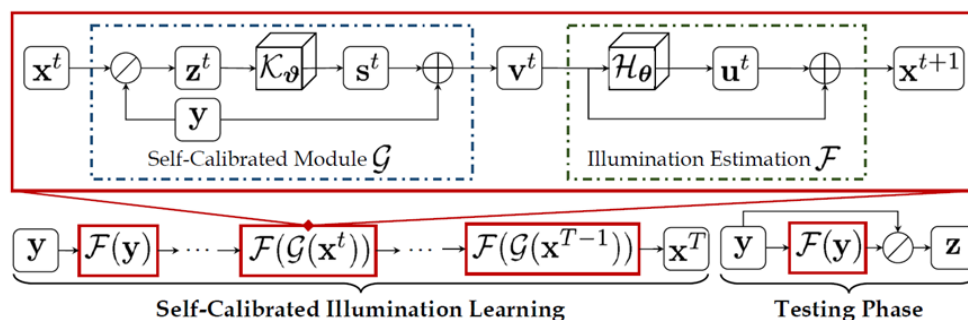


图 1 所提算法流程图

责任编辑 贾同 沈冲意

好文推荐

上海交通大学“Transferable Interactiveness Knowledge for Human-Object Interaction Detection”最新成果发表在 IEEE TPAMI 2022。

论文: Yong-Lu Li, Xinpeng Liu, Xiaoqian Wu, Xijie Huang, Liang Xu, and Cewu Lu. Transferable Interactiveness Knowledge for Human-Object Interaction Detection, IEEE TPAMI, 44: 3870-3882, 2022

人物交互(human-object interaction, HOI)检测是指从静止图像中提取人与物的位置,并同时推断交互类。HOI 作为视觉关系的子任务,与对人体和物体的理解密切相关。该研究可以促进行为理解,模仿学习等相关任务的发展。近年来,深度神经网络技术在这一领域取得了令人瞩目的成果。

在本文中,作者探索了互动知识,该知识可用于表

明人与物体是否存在交互作用。所发现的互动知识可跨 HOI 数据集学习,并缓解不同 HOI 类别设置之间的差距。论文的核心思想是利用交互式网络从多个 HOI 数据集中学习一般的交互式知识,并在推理中进行 HOI 类前的非交互抑制。由于互动性的泛化,该交互式网络是一个可转移知识的学习者,可与任何 HOI 检测模型配合使用,以获得理想的结果。作者利用人的实例和身体的部分特征来学习层次范式中的交互作用,即实例级和身体部分级的交互作用。在此基础上,提出了一致性任务来引导学习,从而提取更深层的交互视觉线索。作者在 HICO-DET、V-COCO 和一个新构建的 HAKE-HOI 数据集上广泛评估了所提方法。通过学习的互动性,所提方法优于目前最先进的 HOI 检测方法,验证了其有效性和灵活性。方法流程图可见图(1),论文方法代码地址:<https://github.com/DirtyHarryLYL/Transferable-Interactiveness-Network>

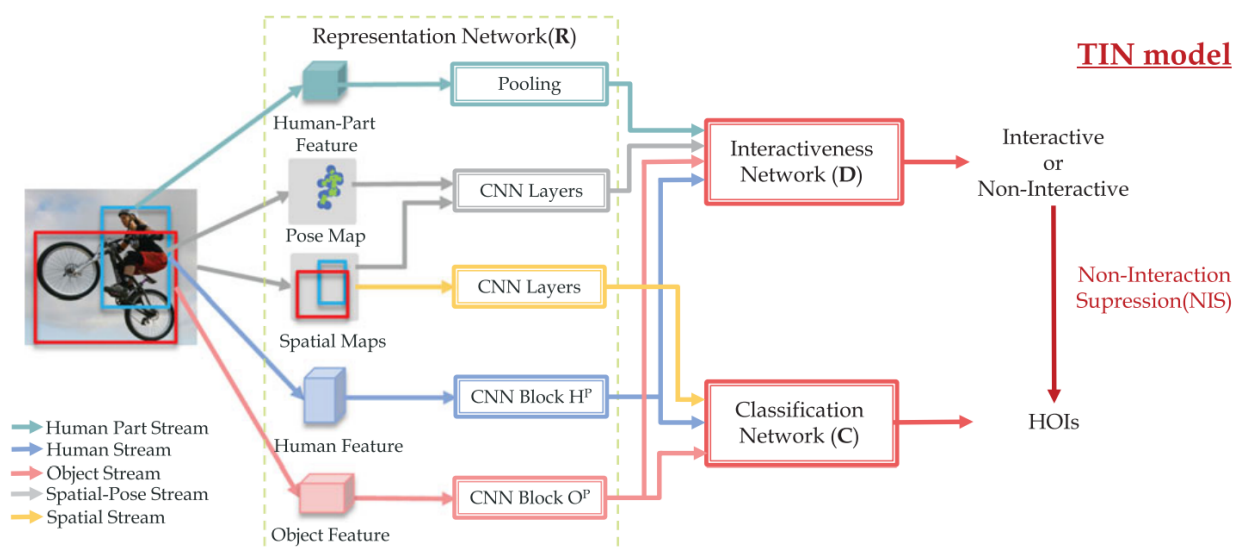


图 1 方法流程图

责任编辑 樊鑫 李策

征文通知

1 会议征文

计算机视觉领域相关国内外会议的征文通知如表 1 所示。同时，可继续关注每个会议举办的 workshop 或 special session。

2 期刊征文

计算机视觉领域近期相关期刊专刊的征文通知如表 2 所示，包括 IEEE Transactions on Multimedia, IEEE Journal of Selected Topics in Signal Processing 和 Physical Communication。

3 会议简介

中国模式识别与计算机视觉学术会议 PRCV (Chinese Conference on Pattern Recognition and Computer Vision)，由中国人工智能学会 (CAAI)、中

国计算机学会 (CCF)、中国自动化学会 (CAA) 和中国图象图形学学会 (CSIG) 联合主办，定位国内顶级的模式识别和计算机视觉领域学术盛。

第五届 PRCV 将于 2022 年 10 月 14 日至 17 日在深圳举行，由南方科技大学和深圳职业技术学院共同承办，香港浸会大学、香港中文大学（深圳）、哈尔滨工业大学（深圳）和中国科学院深圳先进技术研究所联合承办。本届会议旨在促进 PRCV 和湾区学者交流融合、聚焦前沿理论，提高学术交流氛围和质量、技术赋能产业，吸引科技类企业和投资类企业参与、联合粤港澳，提供一个全国科研团队和粤港澳企业近距离交流科研平台和机会。现向广大科技工作者公开征集高质量、原创性的优秀论文。会议论文集将由 Springer 出版社出版，并被 EI 和 CPCI-S (ISTP) 检索。

责任编辑：刘帅奇

表 1 计算机视觉领域相关国内外会议

会议名称	会议时间	会议地点	截稿日期	会议网站
MMSP 2022	2022.09.26-28	Shanghai, China	2022.07.01	https://attend.ieee.org/mmmsp-2022/
ISSPIT 2022	2022.10.13-14	Windsor, Canada	2022.07.22	https://www.isspit.org/isspit/2020/
AAAI 2022	2023.02.07-14	Washington, DC, USA	2022.08.15	https://aaai.org/Conferences/AAAI-23/
DCC 2022	2023.03.21-24	Cliff Lodge, UT	2022.09.08	https://www.cs.brandeis.edu/~dcc/

表 2 计算机视觉领域相关国内外期刊专刊

期刊名称	专刊题目	投稿网址	截稿日期
TMM	Perceiving Humans Under Occlusions	https://signalprocessingsociety.org/blog/ieee-tmm-special-issue-perceiving-humans-under-occlusions	2022.07.15
JSTSP	Biometrics at a distance in the Deep Learning era	https://signalprocessingsociety.org/blog/ieee-jstsp-special-issue-biometrics-distance-deep-learning-era	2022.07.31
PC	Special Issue on Intelligent Reflecting Surface Enhanced Dynamic Spectrum Access	https://research.com/special-issue/intelligent-reflecting-surface-enhanced-dynamic-spectrum-access	2022.07.31
TMM	Point Cloud Processing and Understanding	https://signalprocessingsociety.org/blog/ieee-tmm-special-issue-point-cloud-processing-and-understanding	2022.09.15

心底无私视界宽 - 宣国荣教授专访

本栏目标期望从计算机视觉及相关领域的老前辈那里获取一些历史回忆，从而使本领域的研究人员和爱好者能够了解计算机视觉在中国的发展历程以及老前辈们的贡献，让专委会积累一些历史资料。同时，也希望基于他们的经验和视角来探讨计算机视觉及相关领域的发展现状、优势与不足，以及分享他们在教书育人方面的成功经验。

本次专访的是同济大学的宣国荣教授。宣老师虽然已经 87 岁高龄了，但仍然笔耕不辍、热衷科研，至今仍有科技论文发表。

我是负责本次专访的主要采访人，复旦大学张军平。因疫情原因，本次采访通过微信交流完成，相关问题由《CCF-CV 专委简报》视界专访组提供。为能更好地帮助我们回顾本次采访，我们采用了问答加书面回顾的形式来表述。以下是宣国荣教授的简介和专访内容。

张军平 (采访者, 后缩写为张): 您当年西迁去西交大从事科研工作，是怎么考虑的？

宣国荣 (后缩写为宣): 我 1957 年从交通大学毕业后，留校任教。经国务院批准，1959 年交通大学正式分开为西安交大和上海交大两个学校。我积极响应祖国号召，投身西迁行列。在西安交通大学，我参加了万百五教授和胡保生教授等领衔的“自动控制”新专业的筹建工作。西安交大自动控制专业的成立意义重大，引领学科从“电气化”向“自动化”飞跃。此外，作为年轻教师，我有幸参与全国自动化会战。通过参加兰州化肥厂集中

控制、陕西兴平化肥厂计算机控制等项目，我认识到计算机控制的重要性。早期的这些科研实践，促使我对计算机控制理论深入研究，以及对理论与实践相结合的探索。步入八十年代，我洞察到提升控制最优化的紧迫性，在西安交大创建计算机控制教研室。将以反馈控制为主要特征的自动化，向计算机控制最优化提升。

早在七十年代初，我开始关注智能机器人的动向，对于智能技术的兴趣与日俱增，跃跃欲试。正巧，1974 年电子工业部（当时称四机部）下达给西安交大一项集成电路封装攻关任务：“集成电路引线自动键合机”计算机控制（机械部分是由电子工业部在陕西的西北机器厂负责）。该任务需要采用计算机视觉精确对准和控制，这是国内空白。国际上称该技术为具有视觉的精密机器人。学校把这个光荣而艰巨的重任交给我。我怀着对机器人和计算机视觉的极大兴趣，孜孜不倦，查阅大量英、日语资料，设计方案。作为第一负责人，我带领团队，不懈努力，终于用集成电路自行研制出控制计算机，圆满完成了攻关任务。这个具备计算机视觉的系统，最终通过机电结合的严格检测，各项指标全部合格，正式通过了四机部的审核验收。该科研成果在 1978 年自动化学会年会上发表，受到同行的好评。论文“微程序座标旋转计算机（用于集成电路自动键合）”刊登在《自动化学报》1979 年第 5 卷第 3 期。在此基础上，我们进一步研发智能视觉系统并开辟了更广泛的应用。其中，“微机视觉与汉字识别”系统荣获 1987 年电子工业部科技进步一等奖。



图 1 1988 年宣国荣教授（左 3）和郑南宁博士（右 2）在新加坡国立大学访问



图 2：郑南宁院士 2021 年来上海同济宣国荣教授家探望交谈。

张：能否给我们介绍下西交大人机所成立的背景故事？

宣：1986 年，我与 1985 年从日本庆应大学获得博士学位回国的、我过去的研究生郑南宁一起，提议西安交大成立“人工智能与机器人研究所”。开始校内有不同意见，认为机器人已经包含智能，就用“机器人研究所”。但我们认为工业机器人缺乏对环境的感知能力，例如视觉等。我们坚持名称要加上人工智能。由于得到学校许多老专家，包括当时副校长汪应洛、无线电系胡保生、沈尚贤、蒋大宗，以及机械系老教授等鼎力支持，最后学校同意。1986 年西安交大正式成立“人工智能与机器人研究所”（简称人机所），我任首任所长。郑南宁任副所长。这是国内首创用该名称的研究所。1988 年起，经过所长郑南宁的不懈努力，“人工智能与机器人研究所”得到很大的发展，对于西安交大有重要意义。

以上是背景材料。

回顾从七十年代起，我对于模式识别和计算机视觉，具有极大兴趣。1982 年至 1983 年，我以访问学者身份，公派去美国匹兹堡大学模式识别与图像处理实验室学习交流。回国后，在西安交大开设模式识别课程，自行编写了模式识别教材，开展计算机视觉研究。我们开发出国内领先的“计算机视觉系统”硬件，提供国内许多单位使用。我们的“计算机视觉系统”是许多展览会上最受欢迎的展品，我们的“CV-1 微型计算机视觉系统”获得 1985 年全国微机展览会一等奖。学校早先把香港理工大学送的国内首台“TRS-80”微机调给我们，后来还给我们会讲话、会走路、多自由度、可编程的简单的“HERO-1”智能机器人。一时间，在西安交大校内外，掀起了智能机器人的热潮。

1976 年，我代表西安交大参加中科院北京自动化研究所陈贻运研究员主持召开的座谈会，大约 7 人。我们发起正式成立“中国自动化学会模式识别专业委员会”。在模式识别专业委员会第一届会议上，我被推选为专业委员会委员。此外，1978 年，我还代表西安交大，受中国计算机学会委托，和许多同行联络，包括浙江大学计算机系何志均、上海交大图像所施鹏飞、哈尔滨工业大学计算机系李仲荣教授等，大约 8 人，在浙江计算机学会支持下，发起成立“中国计算机学会人工智能与模式识别专业委员会”，我被推选为组长。通过学会开展学术交流活动，我受益匪浅。

张：您觉得您当年研究的计算机视觉中存在的问题，现在来看还有哪些没有解决的？为什么？

宣：我认识有限，以下意见仅供参考。“无监督机器学习”最大似然估计收敛问题，1977 年 Dempster 提出 EM (Expectation Maximization) 算法，或称最大期望算法，用数学作了严格证明，确定 EM 方法是收敛的、有解的。“Dempster A., Laird N. & Rubin. D., Maximum likelihood from Incomplete Data via the EM Algorithm, Journal of the Royal Statistical

Society, series B, vol. 39-1, 1977, pp.1-38”。Duda 等著“模式分类” (R.O. Duda, P.E. Hart, and D.G. Stork, Pattern Classification, New York: John Wiley & Sons, 2001.), 是模式识别广为应用非常好的教学参考书。虽然 Duda 书中的 EM 算法有单独的介绍, 但是, 在讨论“无监督机器学习”最大似然估计解时, 却只字未提 EM 算法。特别是, Duda 书中对于无监督机器学习最大似然估计的简单例子 (见书中 10.4.1 和 10.4.2 节), 是否有解, 表示怀疑。值得商榷。

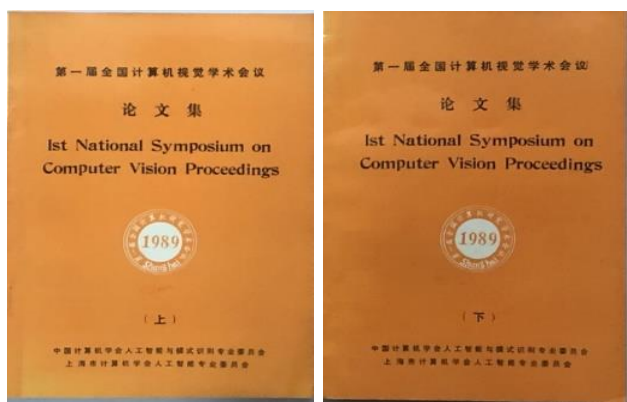


图3 第一届全国计算机视觉学术会议论文集

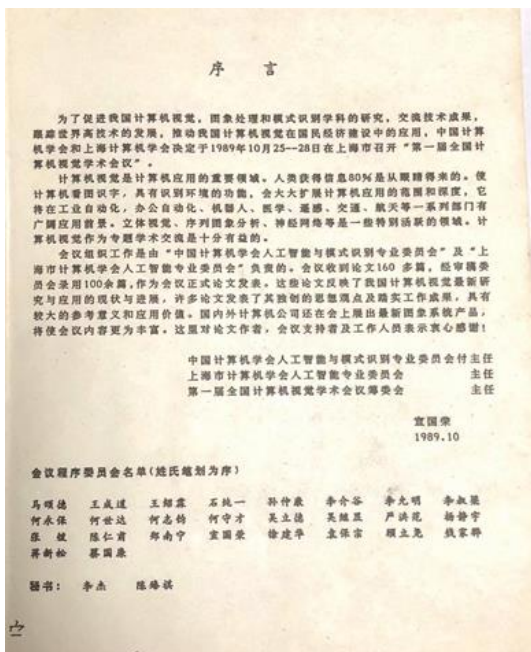


图4 第一届全国计算机视觉学术会议论文集序言

此外, 计算机视觉算法还不十分成熟。记得原来中科院北京自动化研究所的研究员鲍诚志曾介绍他七十年代参加人工智能国际会议时候说, 日本专家现场演示

了一个“人脸识别”系统, 几个日本人能顺利识别出来。结果一个外国人头伸进去, 机器识别了半天, 结果显示“不是人”。现在“人脸识别”系统虽然有飞速的进步, 但是把它普遍应用, 特别是用在提款方面, 还需慎重。

张: 您作为一名优秀的导师培养了一批批包括院士在内的优秀人才, 在您和学生相处的经历中有令您记忆犹新的故事吗? 您有哪些学生培养的经验和大家分享一下? 作为您自己的体验和培养学生的经验, 导师和学生应该建立一种怎样的关系, 才有助于发挥学生的潜能? 有和学生之前印象深刻的故事给我们分享一下吗?

宣: 培养研究生要努力争取良好的培养环境。1987 来同济大学成立计算机系时, 下属成立人工智能研究室和实验室, 在努力完成国家科研任务的同时, 积极开展人工智能方面博士点的建设。1993 年, 我领衔申请, 获得国家学位委员会批准, 授予同济大学“模式识别与智能系统”博士点。这是同济大学电类博士点零的突破, 从此带动同济大学的人工智能、计算机、自动控制等多学科快速发展。我从“控制科学与工程”专业博士生导师转为“模式识别与智能系统”专业博士生导师, 在研究生培养方向有质的飞跃。

我的许多学生有许多突出的成就, 有的成为院士, 这归功于他们本人努力、国家的兴旺和大家的支持, 我是非常高兴的, 但这不能算是我的功劳。我的作用仅是在研究生学习期间, 启发学生对新技术的兴趣和向往, 引领他们进入科学殿堂, 仅此而已。

年轻研究生的闯劲和指导教师的工作经验, 教学相长、互相促进, 是黄金组合。面对飞速发展的科学技术, 我自己也是小学生, 必须努力学习。

有个台湾的教授, 告诉我他怎样带研究生: “中等研究生可以放手, 优秀的研究生你就向他学, 后进的研究生的论文你就帮他做”, 虽然我不能完全苟同, 但似乎也有点道理。

我完成的国家任务项目, 例如“交通图像处理与语音合成系统”, “印刷电路表面缺陷检查图像系统”,

“图纸识别计算机输入系统”，“汉字计算机自动识别”，“虚拟现实的灵镜系统”等。还获得八五国家重点科技攻关“重大成果奖”。通过这些任务，研究生的科研能力得到了锻炼和提高。许多毕业的研究生现在已经成为企业或学校的技术骨干，不少已经是单位的组织者或领导者。

九十年代，我与上海市交通研究所合作，用计算机视觉原理，在上海某交通路口，开发交通流图像识别系统。该系统获得上海市公安局科技成果奖。在上海召开的 IEEE 第六届国际智能交通学术会议上交流。参加该



图 5 施云庆(左)宣国荣(右)在数字水印 IWDW06 国际会议



图 6 宣国荣等数字水印国际会议 IWDW07 论文获奖项目的博士研究生朱江（2002 年毕业），现在已是“海康威视”公司的副总裁。此外，我们为上海市人代会制

作了具有图像识别功能的“自动选票机”，获得国家发明专利。参加该项目的博士研究生中，有现任上海交大信息中心主任、博士生导师管海兵（1999 年毕业）。

每年教师节或重要节日，我指导过的研究生，包括西安交大人工智能与机器人研究所、同济大学人工智能研究室的研究生，从全国甚至世界各地，通过微信、电子邮件，电话等，发来祝贺信息，听到他们的好消息，十分高兴。因为他们不仅是学习知识取得学位的学生，而且是和我一起完成国家任务共同奋斗的合作伙伴，而现在更是人工智能方面的高手了。

张:能分享一些您早年与日本 CSK 合作开拓中国动画市场的故事吗？您是怎样促成这样的合作，还建立一种樱花树下的动人友谊？

宣:同济大学计算机学院和所在的计算机系与日本 CSK 公司合作，开发动画计算机软件制作，开拓国内动画市场。由于校方快速创造许多有利条件，日方为了表示感谢，向同济大学赠送 100 棵樱花树苗。1997 年在同济大学建校九十周年时，学校领导与公司负责人在同济大学爱校路上，举行了樱花树赠送和纪念石碑的落成仪式。前几年，大家在网上可以看到同济大学计算机系总支书记毛广山的文章，生动地介绍了我校与日本 CSK 公司合作的故事和过程，引起广泛的关注。经过学校多年的精心护理和师生的爱护，每逢三月春天来到，同济大学四平校区主干道两旁，美丽的樱花树竞相开放。除了吸引广大同济大学的师生前来欣赏外，还向全市市民开放，得到大家的欢迎和赞许。在纪念石碑上，有红字刻写的“樱花为证，中日友好，世代为继”大字。这象征着高校改革的成果，是对外合作开放的见证。同济大学的樱花给大家带来快乐，我是非常高兴的。

另外，同济大学成立计算机系，有些事情属于巧遇。1986 年初，我遇到同济大学校长江景波，他急切希望我来同济大学筹建计算机系。经过许多波折，西安交大同意我来同济。我和原来同济大学电气工程系党总支和计算机教研室的老师们多方交换意见，反复考虑，作了



图7 1990年西安交大胡保生教授被聘任为同济大学兼职教授。左起：王月娟，柴佩琪，徐麟生（同济人事处长），吴启迪（同济校长），庞瑶华，胡保生，陈辉堂，宣国荣，肖蕴诗。

全面细致的安排，终于在1987年5月同济大学八十周年校庆时，正式成立同济大学计算机科学与程系。我任系主任，叶大兴、沈仰先、臧德滋等任副主任。另外，值得一提的是，1990年，西安交大的胡保生教授也被聘任到同济大学任自动控制方面的兼职教授。

张：作为计算机学科的前辈，您当年开始学科建设的经历有没有可以和大家分享？对计算机学科的昨天、今天和明天能谈谈您的看法吗？

宣：我在担任中国计算机学会人工智能与模式识别专业委员会副主任期间，受中国计算机学会委托，1989年在上海召开“全国首届计算机视觉学术会议”。会议得到上海计算机学会、上海的许多专家教授的支持，如上海交大的李介谷、施鹏飞教授，复旦大学的何永保、吴立德、徐建华教授等的支持，出版了论文集上下册。此外上海市计算机学会，也新开设了人工智能专业委员会，推举我为第一任主任。1985年我已是西安交大自动控制专业的博士生导师。1993年，我领衔申请并获得国家学位委员会批准，授予当时我所在同济大学的“模式识别与智能系统”博士点。1990年我获中华人民共和国人事部颁发的“中青年有突出贡献专家”证书，享受

国务院政府特殊津贴。我们指导的“模式识别与智能系统”专业的博士和硕士们，在国家各种岗位上做出了不少成绩。其中，还有好几位获得我国政府批准的千人计划从美国回来的博士。

张：您认为做为计算机领域的教育工作者，应该以什么样的态度对待基础教学，成为一名优秀的教师？当年您有没有值得回忆的一些校企合作和合作者？

宣：我从学生时代开始，就向前辈学习了许多宝贵的思想。例如不要安于现状，要虚心学习，勇于实践，不怕失败，贵在坚持。另外，数学是计算机视觉的最重要的基础学科之一。我对数学学习和实践有巨大的兴趣和热情。要运用最先进的科学技术，需提高计算机软件和硬件的动手能力，攻破难关。现在我还是能熟练地用MATLAB编程来解决实际问题。当然，还要注重交流：我在美国匹兹堡大学访问期间，相识不少知名人士，例如北方交大的袁保宗教授、沈阳自动化研究所王天然研究员等。现在的美国新泽西理工学院的施云庆教授，成为我近十多年来“无损图像数据隐藏”科学研究课题的合作伙伴。他也常来中国作技术交流，给我的研究生们作报告，还和我共同申请国家自然科学基金课题，发表专利和论文。中科院北京自动化所副所长黄泰翼和中科院沈阳自动化所所长蒋新松院士，也分别接受我的邀请，来校作语音识别和机器人的精彩报告，掌声不断。值得一提的是，黄泰翼和蒋新松都是我在交通大学读本科时的同班同学。当时，黄泰翼是团支书，我是班长。

另外，中科院北京自动化所的胡启恒、马颂德研究员等，都对我帮助很大。美国匹兹堡大学的著名教授施增玮教授和李景崇教授，分别被邀请来同济大学作专题报告，参观同济大学人工智能实验室，并且作了宝贵的指导。

在无损数据隐藏方面紧跟技术发展前沿，我与美国新泽西州理工学院的施云庆教授合作，在国际杂志和国际会议上发表多篇论文，并且申请取得数项中国和美国专利。与国内伙伴“深圳宝嘉电子设备公司”合作，完

成多项任务，同时在深圳取得了开发高技术产品的许可。随着电子商务的发展，电子票据的使用越来越频繁。无损数据隐藏和生物特征识别结合，可以得到更可靠、更为安全的电子支付。同济大学计算机系“人工智能研究室”2005年与“深圳宝嘉电子设备公司”合作，将我们无损图像数据隐藏的理论研究用于实际，研制成“指纹身份认证终端”、型号为SQR-17B无损数据隐藏的系统，从接收到的图像中，无损提取用户帐号，与数据库用户注册的指纹比较，以认证远程用户的身份后确认用户帐号。该技术已经在东北某地用于养老金发放试点。



图 8 中科院北京自动化研究所前副所长黄泰翼（右 2）1996 年交大 90 年校庆期间，应邀访问西安交大，与阔别 30 年的老同学和师长合影。宣国荣（左 1）



图 9 1952 年宣国荣在交通大学与同班同学合影，图中宣国荣（后左 2），黄泰翼（后右 2）



图 10 2006 年李景崇教授（左）访问同济大学，宣国荣（右）



图 11 宣国荣 2008 年在美国佛州 TAMPA 市国际模式识别会议（ICPR08）宣读论文

我 2003 年退休，但仍然关注科技的发展。我 2003 年申请并得到国家自然科学基金会的批准，2004 年开始执行项目“无损图像数据隐藏的理论及关键技术研究”，2006 年完成。并且，退休后在国际会议和杂志上还发表了几十篇论文，2008 年还参加了在美国佛州举办的国际模式识别会议（ICPR08）并宣读了论文。2018 年和 2019 年也仍有论文发表在国际期刊上。

责任编辑 张军平 明悦 贾熹滨



宣国荣

宣国荣，男，1935年出生于上海市。1952年考入交通大学，就读工业企业电气化专业，电力工程系。因学生工作在校工作一年，1957年毕业。1957年至1987年先后在上海交通大学和西安交通大学任教。1985年提升为教授，当年即被批准为“控制科学与工程”专业博士生导师。在西安交通大学期间，曾任信息与控制工程系计算机控制教研室主任；创立西安交通大学人工智能与机器人研究所，任所长。1982年至1983年以访问学者身份赴美国匹兹堡大学模式识别与图像处理实验室交流。1987年创建同济大学计算机系，任系主任。1990任同济大学计算机学院院长。1993年，领衔申请并获得国家学位委员会批准授予同济大学“模式识别与智能系统”博士点，成为“模式识别与智能系统”博士生导师。1987年获电子工业部科技进步一等奖，1996年获八五国家重点科技攻关“重大成果奖”。

COMPUTER VISION NEWSLETTER

02 2022
总第 32 期



计算机视觉专委会简报



CCF 计算机视觉
专委会