
新一代通用视觉技术体系“书生”问世

作者：writer 来源：爱科学

本文原地址：<https://www.iikx.com/news/progress/16651.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

新一代通用视觉技术体系“书生”问世。11月17日，上海人工智能实验室联合商汤科技、香港中文大学、上海交通大学共同发布新一代通用视觉技术体系书生（INTERN），该技术体系旨在系统化解解决当下人工智能视觉领域中存在的任务通用、场景泛化和数据效率等一系列瓶颈问题。目前，相应技术报告《INTERN：一种通向通用视觉的新学习范式》（《INTERN: A New Learning Paradigm Towards General Vision》）已在arXiv平台发布。

书生作为中国古代读书人的经典形象，代表着一个通过不断学习、不断成长进而拥有各方面能力的人格化角色。而将全新的通用视觉技术体系命名为书生，据发布者介绍，意在体现其如同书生一般的特质，可通过持续学习、举一反三，逐步实现通用视觉领域的融会贯通，最终实现灵活高效的模型部署。

任务通用和数据学习效率是制约当前人工智能发展的瓶颈。当前发展通用视觉的核心，是提升模型的通用泛化能力和学习过程中的数据效率。上海人工智能实验室主任助理乔宇表示，面向未来，书生通用视觉技术将实现以一个模型完成成百上千种任务，体系化解决人工智能发展中数据、泛化、认知和安全等诸多瓶颈问题。

通用视觉技术体系书生（INTERN）由7个模块组成，包括通用视觉数据系统、通用视觉网络结构、通用视觉评测基准三个基础设施模块，以及区分上下游的四个训练阶段模块。上述技术报告显示，一个书生基模型即可全面覆盖分类、目标检测、语义分割、深度估计四大视觉核心任务。在ImageNet等26个最具代表性的下游场景中，书生模型展现了极强的通用性，显著提升了这些视觉场景中长尾小样本设定下的性能。

同时，相较于OpenAI 今年发布的CLIP（当前最强开源模型），书生在准确率和数据使用效率上均有大幅提升。具体而言，基于同样的下游场景数据，书生在分类、目标检测、语义分割及深度估计四大任务26个数据集上的平均错误率分别降低了40.2%、47.3%、34.8%和9.4%。此外，书生在数据效率方面的提升尤为令人瞩目：只需要1/10的下游数据，就能超过CLIP基于完整下游数据的准确度，例如在花卉种类识别FLOWER任务上，每一类只需两个训练样本，就能实现99.7%的准确率。

商汤科技研究院院长王晓刚介绍说，书生通用视觉技术体系是商汤在通用智能技术发展趋势下前瞻性布局的一次尝试，也是SenseCore商汤AI大装置背景下的一次新技术路径探索。他表示：‘书生’承载了让人工智能参与处理多种复杂任务、适用多种场景和模态、有效进行小数据和非监督学习并最终具备接近人的通用视觉智能的期盼。希望这套技术体系能够帮助业界更好地探索和应用通用视觉AI技术，促进AI规模化落地。

根据计划，基于书生的通用视觉开源平台OpenGVLab将在2022年年初正式开源，向学术界和产业界公开预训练模型及其使用范式、数据系统和评测基准等。OpenGVLab将与上海人工智能实验室此前发布的OpenMMLab、OpenDILab一道，共同构筑开源体系OpenXLab，助力通用人工智能的基础研究和生态构建。（来源：中国科学报赵广立）

相关论文信息：<https://doi.org/10.26434/chemrxiv-2022-08687>

版权声明：凡本网注明来源：中国科学报、科学网、科学新闻杂志的所有作品，网站转载，请在正文上方注明来源和作者，且不得对内容作实质性改动；微信公众号、头条号等新媒体平台，转载请联系授权。邮箱：shouquan@stimes.cn。

作者：王晓刚等 来源：Arxiv

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](https://www.iikx.com)转发