About Me

- **谢伟迪**，上海交通大学长聘轨副教授，上海人工智能实验室青年科学家，牛津大学视觉几何组访问研究员 (Visiting Researcher at Oxford VGG)，国家自然科学基金优秀青年科学基金获得者（海外），上海市（海外）高层次人才计划获得者。
- 博士毕业于牛津大学视觉几何组。首批 Google-DeepMind 全额奖学金获得者，China Oxford Scholarship Fund (Magdalen Award) 奖学金获得者，牛津大学工程系杰出奖 (Oxford Excellence Award) 获得者
- 科技部科技创新 2030 — “新一代人工智能”重大项目青年项目负责人，2022 年上海市领军人才（海外）获得者，阿里巴巴创新研究计划（Alibaba Innovative Research, AIR）主持人
- 主要研究**计算机视觉，多模态自监督学习，AI4Science, AIGC**。发表论文超 45 篇，Google Scholar 累计引用 9000 次，H 指数 34、i10 因子 65。开源多个领域标准数据集合，包括 VGGFace2, Voxceleb, VGGSound 等; 获得多个国际顶级会议研讨会的最佳论文奖和最佳海报奖，最佳期刊论文奖以及最高被引用作者 (Taylor & Francis Biannual Best Article)；担任计算机视觉和人工智能领域的旗舰会议 CVPR，ECCV，NeurIPS Area Chair。
- 更多细节，请移步个人主页: `https://weidixie.github.io`
- **招本科实习同学，硕士，博士**

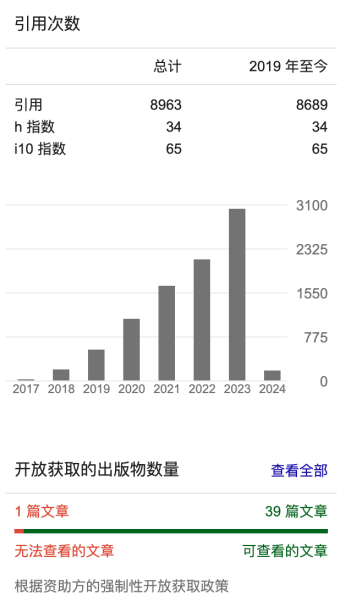
Weidi Xie

已关注

Shanghai Jiao Tong University | VGG, University of Oxford
在 robots.ox.ac.uk 的电子邮件经过验证 - 首页

Computer Vision Machine Learning Medical Image Analysis

标题	引用次数	年份
VGGFace2: A Dataset for Recognising Faces Across Pose and Age Q Cao, L Shen, W Xie, OM Parkhi, A Zisserman IEEE International Conference on Automatic Face & Gesture Recognition (FG …	2845	2018
Microscopy Cell Counting with Fully Convolutional Regression Networks W Xie, JA Noble, A Zisserman MICCAI Workshop	583 *	2017
Voxceleb: Large-scale Speaker Verification in The Wild A Nagrani*, JS Chung*, W Xie*, A Zisserman Computer Speech & Language 60, 101027	567	2020
Video Representation Learning by Dense Predictive Coding T Han, W Xie, A Zisserman ICCV Workshop	399	2019
Self-supervised Co-training for Video Representation Learning T Han, W Xie, A Zisserman Conference on Neural Information Processing Systems (NeurIPS), 2020	377	2020



On-going Research Topics

Traditionally, computer vision research has mainly focused on solving individual task with supervised learning, for example, classifying images, detecting and tracking objects, recognizing human actions, *etc.* However, real-world problems are often complex, open-ended, infinitely fine-grained, the requirement for human annotations quickly becomes unsustainable and infeasible. **As a computer scientist, my long-term ambition is to develop intelligent agents (machines) that can perceive the world at the same level as humans do.**

To be specific, consider a question on the Harry Potter movie, "what does Harry trick Lucius into doing ?" To answer such question, an intelligent agent should be able to extract information from various sources, for instance, images, languages, and audios, to understand when, where, and what actions are being done by whom, to maintain long-term memory (a two-hour movie can have more than 180k frames), to infer relationships between characters and objects, and eventually to reason about the events. My research thus focuses on the following topics:

(I) Open-world Representation Learning refers to the process of training a system to understand visual scene, beyond the seen categories at training time. This is a critical ability for developing foundation models.

References:

- PromptDet: Expand Your Detector Vocabulary with Uncurated Images. In: ECCV2022
- Prompting Visual-Language Models for Efficient Video Understanding. In ECCV2022
- ReCo: Retrieve and Co-segment for Zero-shot Transfer. In: NeurIPS2022
- Learning Open-vocabulary Semantic Segmentation Models From Natural Language Supervision. In: CVPR2023
- OvarNet: Towards Open-vocabulary Object Attribute Recognition. In: CVPR2023
- Towards Open-Vocabulary Video Instance Segmentation. In: ICCV2023

(II) Multi-modal Representation Learning refers to a new paradigm for acquiring effective visual representation from multimodal signals, for example, videos. There is almost an infinite supply available in videos (from Youtube etc.), image level proxy tasks can be used at the frame level; and, there are plenty of additional proxy losses that can be employed from the temporal information. In this area, we are one of the pioneers, and have proposed a number of influential works that are widely used as baselines for various tasks.

References:

- Video Representation Learning by Dense Predictive Coding. In: ICCVW2019
- Self-supervised Co-training for Video Representation Learning. In: NeurIPS2020
- MAST: A Memory-Augmented Self-Supervised Tracker. In: CVPR2020
- Self-supervised Video Object Segmentation by Motion Grouping. In: ICCV2021
- Segmenting Moving Objects via an Object-Centric Layered Representation. In: NeurIPS2022
- AutoAD: Movie Description in Context. In: CVPR2023

(III) AI4Science – Towards Building MedGPT. For a human physician, he/she is expected to see a limited number of patients in the lifetime, each of them with a unique body mass, blood pressure, family history, and so on —a huge variety of features I track in my mental model. Each human has countless variables relevant to their health, but as a human doctor working with a limited session window, he/she will only be able to focus on the several factors that tend to be the most important historically. In contrast, for AIs, they can tirelessly process countless features of every patient, give deep, vast insights, as an example, ChatGPT, Med-PaLM2 have passed the U.S. Medical Licensing Exam. I'm keen to contribute part of my research in revolutionising the medical community !

References:

- MedKLIP: Medical Knowledge Enhanced Language-Image Pre-Training. In: ICCV2023
- Knowledge-enhanced Pre-training for Auto-diagnosis of Chest Radiology Images. In Nature Communications
- PMC-CLIP: Contrastive Language-Image Pre-training using Biomedical Documents. In: MICCAI2023
- PMC-VQA: Visual Instruction Tuning for Medical Visual Question Answering. On Arxiv.

(IV) AIGC: Generating Coherent Videos based on Long Stories. This is what I am interested recently, only some papers under review. **A LOT TO BE DONE HERE !**

References:

- Guiding Text-to-Image Diffusion Model Towards Grounded Generation. In: ICCV2023.
- Intelligent Grimm - Open-ended Visual Storytelling via Latent Diffusion Models. On Arxiv.

To Students

- 你需要在本专业排名至少前 10%
- 你需要对计算机视觉，自监督学习，AI4Science，AIGC 有兴趣，热爱探索未知
- 你需要有极强的自我驱动力，能够应对时常出现的压力和竞争，追求极致的完美
- 你需要能够突破已有学术研究格局，拒绝低质量 paper 发表，宁缺毋滥
- 我们有友好开放的实验室环境，活泼开朗的学长学姐，导师细致耐心的科研指导
- 我们提供与国际顶级研究机构，实验室合作机会，甚至访学机会
- 感兴趣同学，请将简历，成绩单发邮件: weidi@sjtu.edu.cn
- 如果实验室对你提交的内容感兴趣，会尽快联系您，并组织面试