Xinyu Zhang

xyzhang0717@gmail.com; (+61) 0416007483;
Homepage Google Scholar Github LinkedIn

About Me

I am a *Research Fellow* at Australian Institute for Machine Learning (AIML), the University of Adelaide, founded by Centre for Augmented Reasoning (CAR).
I am interested in Machine Learning and Computer Vision, especially in generative models, foundation model pre-training, self-supervised / unsupervised learning and multi-modal perception.

Research Interest

My research focuses on designing machine learning algorithms to *understand and depict* the real large-scale unstructured data, and *generate and create* the synthetic data to simulate the real world.
Specifically, my research topics center on:

- Generative AI models: Image/Video generation/editing
- Foundation model pre-training: Foundation and human-centric pre-training
- Machine learning: Self-supervised / un-supervised / semi-supervised learning
- Multi-modal perception: Object/Attribute detection/recognition; Image/Text-to-image retrieval

Professional Employments

Research Fellow Feb 2024 – Present

AIML, The University of Adelaide
Advisor: Prof. Anton van den Hengel and A/Prof. Lingqiao Liu

Senior Research Scientist Sep 2021 – Feb 2024

Baidu Inc.
Advisor: Chief Scientist Dr. Jingdong Wang

Research Fellow Oct 2020 – Sep 2021

AIML, The University of Adelaide
Medical Project: Machine learning for predicting mortality and morbidity of newborn using unstructured data
Advisor: Prof. Javen Qinfeng Shi

Education

Joint-Training (Visiting) Ph.D. Computer Science Oct 2018 – Oct 2020

AIML, The University of Adelaide
Supervisor: Prof. Chunhua Shen and Prof. Anton van den Hengel

Doctor of Philosophy, Ph.D. Computer Science and Engineering Sep 2015 – Sep 2021

Tongji University
Supervisor: Prof. Chunhua Shen and Prof. Mingyu You

Publications

Conference: Conference on Neural Information Processing Systems (NeurIPS)×2; Conference on Computer Vision and Pattern Recognition (CVPR)×2; International Conference on Computer Vision (ICCV)×2; Association for the Advancement of Artificial Intelligence (AAAI)×1; International Joint Conference on Artificial Intelligence (IJCAI)×1; European Computer Vision Association (ECCV)×1
Journal: IEEE Transactions on Image Processing (TIP)×2; IEEE Transactions on Intelligent Transportation Systems (TITS)×2; Transactions on Machine Learning Research (TMLR)×1; Pattern Recognition (PR)×1
Google Scholar: https://scholar.google.com/citations?user=PSzJxD8AAAAJ Citation: 919 Core A*: 13

*: Equal contribution †: Corresponding Author

Conference Papers

1. Implicit Sample Extension for Unsupervised Person Re-Identification
   Xinyu Zhang*, Dongdong Li*, Zhigang Wang, Jian Wang, Errui Ding, Javen Qinfeng Shi, Zhaoxiang Zhang, Jingdong Wang†.
   *In IEEE Conference on Computer Vision and Pattern Recognition (**CVPR**)*, 2022
2. Diverse Knowledge Distillation for End-to-End Person Search.
   Xinyu Zhang, Xinlong Wang, Jia-Wang Bian, Chunhua Shen, Mingyu You†.
   *The Thirty-Fifth AAAI Conference on Artificial Intelligence (**AAAI**)*, 2021
3. Self-Training with Progressive Augmentation for Unsupervised Cross-Domain Person Re-identification.
   Xinyu Zhang*, Jiewei Cao*, Chunhua Shen†, Mingyu You.
   *In IEEE International Conference on Computer Vision (**ICCV**)*, 2019
4. Evaluation of Text-to-Video Generation Models: A Dynamics Perspective.
   Mingxiang Liao*, Hannan Lu*, **Xinyu Zhang***, Fang Wan, Tianyu Wang, Yuzhong Zhao, Wangmeng Zuo, Qixiang Ye†, Jingdong Wang.
   *Advances in Neural Information Processing Systems (**NeurIPS**)*, 2024
5. HAP: Structure-Aware Masked Image Modeling for Human-Centric Perception.
   Junkun Yuan*, **Xinyu Zhang***†, Hao Zhou, Jian Wang, Zhongwei Qiu, Zhiyin Shao, Shaofeng Zhang, Sifan Long, Kun Kuang†, Kun Yao, Junyu Han, Errui Ding, Lanfen Lin, Fei Wu, Jingdong Wang†.
   *Advances in Neural Information Processing Systems (**NeurIPS**)*, 2023
6. Unified Pre-training with Pseudo Texts for Text-To-Image Person Re-identification.
   Zhiyin Shao*, **Xinyu Zhang***, Changxing Ding†, Jian Wang, Jingdong Wang.
   *In IEEE International Conference on Computer Vision (**ICCV**)*, 2023
7. Learning Granularity-Unified Representations for Text-to-Image Person Re-identification.
   Zhiyin Shao, **Xinyu Zhang**, Meng Fang, Zhifeng Lin, Jian Wang, Changxing Ding†.
   *ACM Multimedia (**ACM MM**)*, 2022
8. VRP-SAM: SAM with Visual Reference Prompt.
   Yanpeng Sun, Jiahui Chen, Shan Zhang, **Xinyu Zhang**, Qiang Chen, Gang Zhang, Errui Ding, Jingdong Wang, Zechao Li†.
   *In IEEE Conference on Computer Vision and Pattern Recognition (**CVPR**)*, 2024.
9. Self-Guided Hard Negative Generation for Unsupervised Person Re-Identification.
   Dongdong Li, Zhigang Wang, Jian Wang, **Xinyu Zhang**, Errui Ding, Jingdong Wang, Zhaoxiang Zhang†.
   *In International Joint Conferences on Artificial Intelligence (**IJCAI**)*, 2022
10. UFO: Unified Feature Optimization.
    Teng Xi, Yifan Sun, Deli Yu, Bi Li, Nan Peng, Gang Zhang, **Xinyu Zhang**, Zhigang Wang, Jinwen Chen, Jian Wang, Lufei Liu, Haocheng Feng, Junyu Han, Jingtuo Liu, Errui Ding, Jingdong Wang.
    *In European Conference on Computer Vision (**ECCV**)*, 2022

Journal Papers

11. CAE v2: Context Autoencoder with CLIP Target Alignment
    Xinyu Zhang*, Jiahui Chen*, Junkun Yuan, Qiang Chen, Jian Wang, Xiaodi Wang, Shumin Han, Xiaokang Chen, Jimin Pi, Kun Yao, Junyu Han, Errui Ding, Jingdong Wang†.
    *In Transactions on Machine Learning Research (**TMLR**)*, 2023
12. Part-Guided Attention Learning for Vehicle Instance Retrieval.
    Xinyu Zhang*, Rufeng Zhang*, Jiewei Cao, Dong Gong, Mingyu You†, Chunhua Shen.
    *In IEEE Transactions on Intelligent Transportation Systems, (**TITS**)*, 2020
13. Context-aware prompt learning for test-time vision recognition with frozen vision-language model.
    Junhui Yin, **Xinyu Zhang**, Lin Wu, Jun Guo, Xiaojie Wang.
    *Pattern Recognition (**PR**)*, 2025
14. A Real-Time Memory Updating Strategy for Unsupervised Person Re-Identification.
    Junhui Yin, **Xinyu Zhang**, Zhanyu Ma†, Jun Guo, Yifan Liu.
    *In IEEE Transactions on Image Processing (**TIP**)*, 2021
15. STAT: Multi-object tracking based on spatio-temporal topological constraints.
    Junjie Zhang, Mingyan Wang, Haoran Jiang, **Xinyu Zhang**, Chenggang Yan, Dan Zeng†.
    *In IEEE Transactions on Image Processing (**TIP**)*, 2023

16. An Extended Filtered Channel Framework for Pedestrian Detection.
Mingyu You, Yubin Zhang, Chunhua Shen, **Xinyu Zhang**. *In IEEE Transactions on Intelligent Transportation Systems, (**TITS**)*, 2018

Preprints and in Submission

17. Training-Free Motion-Guided Video Generation with Enhanced Temporal Consistency Using Motion Consistency Loss.
Xinyu Zhang, Zicheng Duan, Dong Gong, Lingqiao Liu.
In arXiv preprint, arXiv:2501.07563.

18. A Flexible and Versatile Stable Diffusion Model for Enhancing Object Detection.
Xinyu Zhang, Xiaodi Wang, Dong Gong, Gang Zhang, Errui Ding, Jingdong Wang, Lingqiao Liu.
In submission to CVPR 2025.

19. Add-SD: Rational Generation without Manual Reference.
Lingfeng Yang*, **Xinyu Zhang***, Xiang Li, Jinwen Chen, Kun Yao, Gang Zhang, Errui Ding, Lingqiao Liu, Jingdong Wang, Jian Yang†.
In arXiv preprint, arXiv:2407.21016.

20. Are Image Distributions Indistinguishable to Humans Indistinguishable to Classifiers?.
Zebin You, **Xinyu Zhang**, Hanzhong Guo, Jingdong Wang, Chongxuan Li†.
In arXiv preprint, arXiv:2405.18029.

21. LW-DETR: A Transformer Replacement to YOLO for Real-Time Detection.
Qiang Chen*, Xiangbo Su*, **Xinyu Zhang***, Jian Wang, Jiahui Chen, Yunpeng Shen, Chuchu Han, Ziliang Chen, Weixiang Xu, Fanrong Li, Shan Zhang, Kun Yao, Errui Ding, Gang Zhang, Jingdong Wang†.
In arXiv preprint, arXiv:2406.03459.

22. OVLW-DETR: Open-Vocabulary Light-Weighted Detection Transformer.
Yu Wang, Xiangbo Su, Qiang Chen, Xinyu Zhang, Teng Xi, Kun Yao, Errui Ding, Gang Zhang, Jingdong Wang. *In arXiv preprint, arXiv:2407.10655.*

23. Memorizing Comprehensively to Learn Adaptively: Unsupervised Cross-Domain Person Re-ID with Multi-level Memory.
Xinyu Zhang, Dong Gong, Jiewei Cao, Chunhua Shen.
In arXiv preprint, arXiv:2001.04123.

PATENTS

1. Training method of target re-identification model and target re-identification method.
Xinyu Zhang, Jian Wang, Haocheng Feng
Chinese Patent. CN115578613A Grant date: March. 2024.

2. Training method, device, equipment and medium of perception model.
Xinyu Zhang, Jian Wang, Kun Yao, Errui Ding, Jingdong Wang
Chinese Patent. CN116629315A Grant date: Feb. 2024.

3. Image generation, training of large model, image processing method and device, equipment and medium.
Xinyu Zhang, Xiaodi Wang, Haixiao Yue, Gang Zhang, Kun Yao, Errui Ding
Chinese Patent pending. CN118506131A

4. Progressive-enhancement self-learning unsupervised cross-domain pedestrian re-identification method.
Xinyu Zhang, Zhigang Wang, Jian Wang, Hao Sun, Errui Ding
Chinese Patent. CN111598004A Grant date: Dec. 2023.

5. Pedestrian re-recognition model training method and device, electronic equipment and storage medium.
Chunhua Shen, **Xinyu Zhang**, Zhengrong Li
Chinese Patent. CN115273148A Grant date: Sep. 2023.

6. Training method of pedestrian re-recognition model, and pedestrian re-recognition method and device.
Xinyu Zhang, Zhigang Wang, Jian Wang, Hao Sun, Errui Ding
Chinese Patent. CN114724090A Grant date: Aug. 2022.

Academic Services

Area Chair

- ICCV 2025
- IJCAI (Senior Program Committee) 2025

Conference / Journal Reviewer / Program Committee

- ICLR, NeurIPS, ICCV, CVPR, ECCV, AAAI, IJCAI
- TPAMI, IJCV, TIP, TOMM, TNNLS, TMM, PR, Neurocomputing
- Award panel member in Sydney AI meetup 2024

Invited Talks

- "Evaluation of Text-to-Video Generation Models: A Dynamics Perspective", at The Thirty-Eighth Annual Conference on Neural Information Processing Systems, 2024
- "VRP-SAM: SAM with Visual Reference Prompt and Industry Application", at Document and Multi-modal Intelligence Frontier Seminar, 2024
- "Self-supervised pre-training on human-centric perception tasks", VIS Baidu, 2023
- "Diverse Knowledge Distillation for End-to-End Person Search", at Beijing Jiaotong University, 2021
- "Research on Person Re-ID and Person Search", at Tongji University, 2021

Teaching Experience

Guest Lecturer The University of Adelaide

Introduction to Statistical Machine Learning (COMP SCI 7314) Semester 2, 2020
Introduction to Statistical Machine Learning (COMP SCI 7314) Semester 2, 2019

Teaching Assistant Tongji University

Machine Vision (2080214) Semester 1, 2016
Pattern Recognition (2080387) Semester 2, 2015

Honors and Awards

- Outstanding Collaboration Team, Baidu 2023
- Outstanding Employee Award, Baidu 2022 & 2023
- Outstanding Graduates, Tongji University 2021
- ICCV 2019 Student Travel Award 2019
- Second prize of China Graduate Mathematical Contest in Modeling, China 2018
- Social Activities Scholarship, Tongji University 2018 & 2019
- Outstanding Student, Tongji University 2017

Supervision

(university's names are in brief)

- Zicheng Duan (Ph.D. in AIML, with A/Prof. Lingqiao Liu): Image & video generation 2024-Present
- Mingxiang Liao (Ph.D. in UCAS): Video generation 2023-2024
- Zebin You (Ph.D. in RUC): Image generation 2023-2024
- Lingfeng Yang (Ph.D. in NJUST): Image generation 2023-2024
- Yanpeng Sun (Ph.D. in NJUST): MLLM & Segmentation 2022-2024
- Hannan Lu (Ph.D. in HIT): Video generation & Segmentation 2022-Present
- Junkun Yuan (Ph.D.in ZJU, now Tencent): Self-supervised pre-training 2021-2023
- Zhiyin Shao (MPhil. in SCUT, now Baidu): Human-centric pre-training 2023
- Jiahui Chen (MPhil. in BUAA, now Kuaishou): Object detection 2022
- Dongdong Li (MPhil. in CAS): Person Re-Identification 2021-2022
- Junhui Yin (Ph.D. in BUAA, now Postdoc in BUAA): Person Re-Identification 2022

Selected Research Projects

Multi-concept Data Generation 2023 – 2024

Build algorithms and engines to synthesise multi-concept data and evaluate data quality comprehensively.
My role: Project lead
Application: The algorithm and data engine has been extensively used in the WenXin Yige project.
Research outcoms: NeurIPS $\times$1, In submission$\times$ 2

Foundation model pre-training 2022 – 2024

Train foundation model with self-supervised pre-training, especially masked image modeling with CLIP, on large-scale unlabelled dataset to improve various downstream vision tasks, *e.g.*, object detection, semantic segmentation and object recognition.
My role: Core research developer
Application: The industry pre-trained model has been used in the smart city project, including ERNIE project, AI camera and Zhuohua sport.
Research outcoms: TMLR $\times$1, ECCV$\times$ 1, ICCV$\times$ 1

Human-centric AI 2022 – 2024

Train human-centric model with large-scale unlabelled human data for downstream human-centric perception tasks, *e.g.*, pose estimation, person re-identification, attribute and action recognition.
My role: Core research developer
Application: The industry pre-trained model has been used in the Cixi smart city project.
Research outcoms: NeurIPS $\times$1, ICCV$\times$ 1, ACMMMM$\times$ 1, CVPR$\times$ 1

Neonatal Medical Project 2020 – 2021

Develop machine learning to predict neonatal mortality, morbidity and stay length in newborn intensive care.
My role: Core contributor
Application: Assistant doctor's decision.
Funding outcomes: Successfully applied for the second-stage bigger funding

Referees

Prof. Chunhua Shen
Chair Professor
Zhejiang University
(Previous Professor at The University of Adelaide)
Email: chhshen@gmail.com
Associate Prof. Lingqiao Liu
Associate Professor, Discipline Lead
School of Computer and Mathematical Sciences, The University of Adelaide
Email: lingqiao.liu@adelaide.edu.au
Prof. Javen Qinfeng Shi
Professor, Director of Causal AI Group
School of Computer and Mathematical Sciences, The University of Adelaide
Email: javen.shi@adelaide.edu.au
Dr. Jingdong Wang
Chief Scientist
Baidu Inc.
Email: welleast@gmail.com