JIANYANG GU

gu.1220@osu.edu | vimar-gu.github.io | google scholar

RESEARCH INTEREST

I work on machine learning, computer vision, multimodal understanding, and their applications in scientific discovery. I am particularly interested in (i) data-centric machine learning, including enriching and condensing data; (ii) developing and interpreting multimodal foundation models for science.

CURRENT POSITION

The Ohio State University　　　　　　　　　　　　　　　　　　Columbus, OH, USA
Postdoctoral Scholar　　　　　　　　　　　　　　　　　　　　　　*Aug. 2024 – Now*
- Mentored by Dr. Wei-Lun Chao, Dr. Yu Su, and Dr. Tanya Berger-Wolf

EDUCATION

Zhejiang University　　　　　　　　　　　　　　　　　　　　　Hangzhou, China
Ph.D. in Control Science and Engineering　　　　　　　　　　　*Sep. 2019 – Jun. 2024*
- Supervised by Dr. Wei Jiang

Zhejiang University　　　　　　　　　　　　　　　　　　　　　Hangzhou, China
B.Eng. in Control Science and Engineering　　　　　　　　　　　*Sep. 2015 – Jun. 2019*

National University of Singapore　　　　　　　　　　　　　　　Singapore
Visiting Scholar in School of Computing　　　　　　　　　　　　*Sep. 2022 – Oct. 2023*
- Supervised by Dr. Yang You

WORK EXPERIENCE

OPPO Research Intern　　　　　　　　　　　　　　　　　　　Nov. 2021 – Jun. 2022
- Topic: generalizable object re-identification structure.

Alibaba Research Intern　　　　　　　　　　　　　　　　　　Jun. 2020 – Apr. 2021
- Topic: unsupervised domain adaptive person re-identification.

Yitu Tech. CI Intern　　　　　　　　　　　　　　　　　　　May. 2018 – Aug. 2018
- Topic: automatic test pipeline for products.

RESEARCH PROJECTS

Interpretable and Explainable AI for Science
- Build computational tools around biological knowledge bases to analyze image data.
- Develop multimodal foundation models for general biology questions on images.
- Design explainable and trustworthy models to advance biology development as well as public use.

Data-Centric Efficient Training
- Reduce storage and computational data burden by synthesizing small surrogate datasets.
- Enhance the representativeness and diversity of generative diffusion techniques for data generation.
- Conduct more efficient training by selecting informative samples on both online and offline schemes.

Unsupervised Object Re-identification System (Ph.D. Project)
- Construct a network architecture generalizable for incoming unseen tasks.
- Design an unsupervised domain adaptation method to improve the accuracy on new tasks.
- Incorporate continual learning techniques to reduce the catastrophic forgetting on previous tasks.

PUBLICATIONS

(* Equal contribution; <u>   </u> Mentored student; † Corresponding author)

Preprints and Papers in Submission

[p.0] <u>Connor Kilrain</u>, David Carlyn, Julia Chae, Sara Beery, Wei-Lun Chao, and **Jianyang Gu**[†]. Finer-Personalization Rank: Fine-Grained Retrieval Examines Identity Preservation for Personalized Generation. *arXiv*, 2512.19026 (2025).

[p.1] <u>Qinsi Wang</u>, Saeed Vahidian, Hancheng Ye, **Jianyang Gu**, Jianyi Zhang, Yiran Chen. CoreInfer: Accelerating Large Language Model Inference with Semantics-Inspired Adaptive Sparse Activation. *arXiv*, 2410.18311 (2024).

[p.2] Vyacheslav Kungurtsev, <u>Yuanfang Peng</u>, **Jianyang Gu**, Saeed Vahidian, Anthony Quinn, Fadwa Idlahcen, Yiran Chen. Dataset Distillation from First Principles: Integrating Core Information Extraction and Purposeful Learning. *arXiv*, 2409.01410 (2024).

[p.3] **Jianyang Gu**, Hao Luo, Kai Wang, Wei Jiang, Yang You, Jian Zhao. Color Prompting for Data-Free Continual Unsupervised Domain Adaptive Person Re-Identification. *arXiv*, 2308.10716 (2023).

Conference Proceedings

[c.0] <u>Lin Zhao</u>, Xinru Jiang, Xi Xiao, Qihui Fan, Lei Lu, Yanzhi Wang, Xue Lin, Octavia Camps, Pu Zhao, **Jianyang Gu**[†]. HierAmp: Coarse-to-Fine Autoregressive Amplification for Generative Dataset Distillation. *CVPR* (2026).

[c.1] <u>Yu Hu</u>*, **Jianyang Gu***, Hao Liu, Yue Cao, Jozsef Hamari, Zheng Liu, Mohsen Zardadi. AVION: Aerial Vision–Language Instruction from Offline Teacher to Prompt-Tuned Network. *CVPR* (2026).

[c.2] <u>Ziheng Zhang</u>, <u>Xinyue Ma</u>, Arpita Chowdhury, Elizabeth G Campolongo, Matthew J Thompson, Net Zhang, Samuel Stevens, Hilmar Lapp, Tanya Berger-Wolf, Yu Su, Wei-Lun Chao, **Jianyang Gu**[†]. BioCAP: Exploiting Synthetic Captions Beyond Labels in Biological Foundation Models. *ICLR* (2026).

[c.3] **Jianyang Gu**, Haonan Wang, Ruoxi Jia, Saeed Vahidian, Vyacheslav Kungurtsev, Wei Jiang, Yiran Chen. CONCORD: Concept-Informed Diffusion for Dataset Distillation. *WACV* (2026).

[c.4] **Jianyang Gu**, Samuel Stevens, Elizabeth G Campolongo, Matthew J Thompson, Net Zhang, Jiaman Wu, Andrei Kopanev, Zheda Mai, Alexander E White, James Balhoff, Wasila Dahdul, Daniel Rubenstein, Hilmar Lapp, Tanya Berger-Wolf, Wei-Lun Chao, Yu Su. BioCLIP 2: Emergent Properties from Scaling Hierarchical Contrastive Learning. *NeurIPS (Spotlight)* (2025).

[c.5] <u>Ziheng Zhang</u>*, **Jianyang Gu***, Arpita Chowdhury, Zheda Mai, David Carlyn, Tanya Berger-Wolf, Yu Su, Wei-Lun Chao. Finer-CAM: Spotting the Difference Reveals Finer Details for Visual Explanation. *CVPR* (2025).

[c.6] <u>Arpita Chowdhury</u>, Dipanjyoti Paul, Zheda Mai, **Jianyang Gu**, Ziheng Zhang, Kazi Sajeed Mehrab, Elizabeth G Campolongo, Daniel Rubenstein, Charles V Stewart, Anuj Karpatne, Tanya Berger-Wolf, Yu Su, Wei-Lun Chao. Prompt-CAM: Making Vision Transformers Interpretable for Fine-Grained Analysis. *CVPR* (2025).

[c.7] <u>Lin Zhao</u>, Yushu Wu, Xinru Jiang, **Jianyang Gu**, Yanzhi Wang, Xiaolin Xu, Pu Zhao, Xue Lin. Taming Diffusion for Dataset Distillation with High Representativeness. *ICML* (2025).

[c.8] Saeed Vahidian*, <u>Mingyu Wang</u>*, **Jianyang Gu***, Vyacheslav Kungurtsev, Wei Jiang, Yiran Chen. Group Distributionally Robust Dataset Distillation with Risk Minimization. *ICLR* (2025).

[c.9] **Jianyang Gu**, Saeed Vahidian, Vyacheslav Kungurtsev, Haonan Wang, Wei Jiang, Yang You, Yiran Chen. Efficient Dataset Distillation via Minimax Diffusion. *CVPR*, 15793–15803 (2024).

[c.10] **Jianyang Gu**, Kai Wang, Wei Jiang, Yang You. Summarizing Stream Data for Memory-Restricted Online Continual Learning. *AAAI*, 12217–12225 (2024).

[c.11] Ziheng Qin, Kai Wang, Zangwei Zheng, **Jianyang Gu**, Xiangyu Peng, Daquan Zhou, Yang You. InfoBatch: Lossless Training Speed Up by Unbiased Dynamic Data Pruning. *ICLR (Oral)* (2024).

[c.12] <u>Yanqing Liu</u>*, **Jianyang Gu***, Kai Wang, Zheng Zhu, Wei Jiang, Yang You. DREAM: Efficient Dataset Distillation by Representative Matching. *ICCV*, 17314–17324 (2023).

[c.13] Daquan Zhou*, Kai Wang*, **Jianyang Gu***, Xiangyu Peng, Dongze Lian, Yifan Zhang, Yang You, Jiashi Feng. Dataset Quantization. *ICCV*, 17205–17216 (2023).

[c.14] **Jianyang Gu**, Kai Wang, Hao Luo, Chen Chen, Wei Jiang, Yuqiang Fang, Shanghang Zhang, Yang You, Jian Zhao. MSINet: Twins Contrastive Search of Multi-Scale Interaction for Object ReID. *CVPR*, 19243–19253 (2023).

Journal Articles

[j.0] <u>Yao Lu</u>, Xuguang Chen, **Jianyang Gu**, Yuchen Zhang, Qi Xuan, Zhaowei Zhu. Dataset distillation with pre-trained models: A contrastive approach. *Neurocomputing*, 132015 (2025).

[j.1] Haojie Liu, **Jianyang Gu**, Zhiyong Li, Mingyu Wang, QM Jonathan Wu, Wei Jiang. CoMix: Collaborative Mixed Learning via Style Fuzzy Normalization for Visible–Infrared Person Re-Identification. *IEEE TSMCS*, 55 (11), 8572-8586 (2025).

[j.2] Haojie Liu, Zhiyong Li, **Jianyang Gu**, Mingyu Wang, QM Jonathan Wu, Wei Jiang. Stochastic Style Perturbation Modeling for Visible-Infrared Person Re-Identification with Severely Modality Imbalance. *Neural Networks*, 108206 (2025).

[j.3] Xiao Pan, Hao Luo, Weihua Chen, Fan Wang, Hao Li, Wei Jiang, Jianming Zhang, **Jianyang Gu**, Peike Li. Dynamic Gradient Reactivation for Backward Compatible Person Re-identification. *PR*, 146, 110000 (2024).

[j.4] **Jianyang Gu**, Weihua Chen, Hao Luo, Fan Wang, Hao Li, Wei Jiang, Weijie Mao. Multi-view Evolutionary Training for Unsupervised Domain Adaptive Re-identification. *IEEE TIFS* 17, 344-356 (2022).

[j.5] Ran Wei, **Jianyang Gu**, Shuting He, and Wei Jiang. Transformer-Based Domain-Specific Representation for Unsupervised Domain Adaptive Vehicle Re-Identification. *IEEE TITS*, 14 (2), 1-21 (2022).

[j.6] Xiao Pan, Hao Luo, Wei Jiang, Jianming Zhang, **Jianyang Gu**, Peike Li. SFGN: Representing the sequence with one super frame for video person re-identification. *KnoSys*, 249, 108884 (2022).

[j.7] Haonan Xie, Hao Luo, **Jianyang Gu**, Wei Jiang. Unsupervised Domain Adaptive Person Re-Identification via Intermediate Domains. *Applied Science*, 12 (14), 6990 (2022).

[j.8] Hao Wu, **Jianyang Gu**, Xiaojin Fan, He Li, Lidong Xie, Jian Zhao. 3D-Guided Frontal Face Generation for Pose-Invariant Recognition. *ACM TIST*, 14 (2), 1-21 (2022).

[j.9] **Jianyang Gu**, Wei Jiang, Hao Luo, Hongyan Yu. An efficient global representation constrained by Angular Triplet loss for vehicle re-identification. *Pattern Anal Applic* 24, 367–379 (2021).

[j.10] Hao Luo, Wei Jiang, Youzhi Gu, Fuxu Liu, Xingyu Liao, Shenqi Lai, **Jianyang Gu**. A strong baseline and batch normalization neck for deep person re-identification. *IEEE TMM* 22(10), 2597-2609 (2019).

Workshop Papers

[w.0] Zhenyang Feng, Zihe Wang, **Jianyang Gu**, Saul Ibaven Bueno, Tomasz Frelek, Advikaa Ramesh, Jingyan Bai, Lemeng Wang, Zanming Huang, Jinsu Yoo, Tai-Yu Pan, Arpita Chowdhury, Michelle Ramirez, Elizabeth G. Campolongo, Matthew J. Thompson, Christopher G. Lawrence, Sydne Record, Neil Rosser, Anuj Karpatne, Daniel Rubenstein, Hilmar Lapp, Charles V. Stewart, Tanya Berger-Wolf, Yu Su, Wei-Lun Chao. Static Segmentation by Tracking: A Frustratingly Label-Efficient Approach to Fine-Grained Segmentation. *CVPR CV4Animals Workshop (Oral)* (2025).

[w.1] Yao Lu, **Jianyang Gu**, Xuguang Chen, Saeed Vahidian, Qi Xuan. Exploring the Impact of Dataset Bias on Dataset Distillation. *CVPR DDCV Workshop*, 7656–7663 (2024).

[w.2] Kai Wang*, **Jianyang Gu***, Daquan Zhou, Zheng Zhu, Wei Jiang, Yang You. DiM: Distilling Dataset into Generative Model. *ECCV DD Challenge Workshop* (2024).

[w.3] **Jianyang Gu**, Hao Luo, Weihua Chen, Yiqi Jiang, Yuqi Zhang, Shuting He, Fan Wang, Hao Li, Wei Jiang. 1st Place Solution to VisDA-2020: Bias Elimination for Domain Adaptive Pedestrian Re-identification. *ECCV VisDA Challenge* (2021).

[w.4] Hao Luo, Weihua Chen, Xianzhe Xu, **Jianyang Gu**, Yuqi Zhang, Chong Liu, Yiqi Jiang, Shuting He, Fan Wang, Hao Li. An Empirical Study of Vehicle Re-Identification on the AI City Challenge. *CVPR AICity Challenge*, 4095-4102 (2021).

[w.5] Wei Li*, Shimin Chen*, **Jianyang Gu***, Ning Wang, Chen Chen, Yandong Guo. MV-TAL: Multi-view temporal action localization in naturalistic driving. *CVPR ActivityNet Challenge*, 3242-3248 (2022).

AWARDS & HONORS

- **NeurIPS Scholar Award** 2025
- **OSU CSE Research Staff Award** 2025
- **AAAI Scholarship** 2024
- **Third Place, ActivityNet Temporal Action Localization Challenge in CVPR Workshop** 2022
- **Third Place, SoccerNet Challenge 2022 Action Spotting in CVPR Workshop** 2022
- **First Place, AICity Challenge 2021 Track 2 in CVPR Workshop** 2021

- **Alibaba Annual Outstanding Research Intern** 2020
- **Second Prize, National AI Challenge 2020 Person Re-Identification Track** 2020
- **First Place, Visual Domain Adaptation Challenge 2020 in ECCV Workshop** 2020
- **Annual Merit Graduate Student** 2020
- **First Place, Robocup Montreal** 2018

PRESENTATIONS & TALKS

Future of AI Algorithms and Applications Panel
- NAIRR Annual Meeting, Arlington, VA   Mar, 2026

BioCLIP 2 and BioCAP: Extending Species Classification to Rich Ecological Understanding
- Imageomics Seminar, The Ohio State University, Columbus, OH   Feb, 2026

BioCLIP 2: From Fine-Grained Classification to Biologically Meaningful Representation
- Australian National University, Advanced Computer Vision Course, remote   Aug, 2025

Dataset Distillation Progresses and Future Possibilities
- Brown University, remote   Mar, 2024
- Meta Reality Labs, remote   Feb, 2024

MSINet: Twins-Contrastive Search of Multi-Scale Interaction for Object ReID
- TerraSense, remote   2022

Bias Elimination for Domain Adaptive Pedestrian Re-identification
- ECCV VisDA Challenge, remote   2020

FUNDING

NAIRR Pilot: Fine-grained biological trait-aware foundation model   2025-2026
- Award amount   OpenAI credit $3,000
- Role   Primary proposal co-author

ADADEMIC SERVICE

Workshop Organization
- Lead Organizer   Third Workshop on Imageomics @ NeurIPS2025
- Co-organizer   Second Workshop on Imageomics @ AAAI2025
- Co-organizer   Anomaly Detection in Scientific Domains Workshop @ AAAI2025
- PC Member   First Workshop on Dataset Distillation @ CVPR2024

Conference Area Chair
- ICLR 2026

Conference Reviewer
- CVPR, ICCV, ECCV, ICLR, ICML, NeurIPS, AAAI, ACMMM, WACV, ACCV

Journal Reviewer
- IEEE TPAMI, PNAS, PR, CVIU, IEEE TCSVT

OUTREACH ACTIVITIES

- **Panelist**, Student AI Club of Ohio State University   Sept. 2024
- **Chair**, Student AI Association of Zhejiang University   Aug. 2020 - Jun. 2021