JIANYANG GU

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

RESEARCH INTEREST

My research advances **machine learning for scientific discovery** through learning and exposing concepts from scientific observations and investigating how concepts in data shape model learning and behavior. I envision **concepts as a human-centered computational interface** between data and models. Scientists can derive observation-grounded relationships between concepts and use the produced insights to guide further data and model development, forming a **reinforcing cycle between scientific understanding and machine learning**.

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

Massachusetts Institute of Technology — Boston, MA, USA
Visiting Postdoctoral Scholar — *Jul. 2026 – Now*

- Mentored by Dr. Sara Beery

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

Concept Learning and Extraction from Scientific Observations

- Develop foundation models that capture meaningful scientific concepts.
- Reveal reliable, interpretable visual evidence underlying model decisions.
- Enrich foundation model training through explicit concept injection.

Concept Influence and Use in Data-Centric Machine Learning

- Understand data values through condensing surrogates that preserve original characteristics.
- Leverage visual concepts to understand model training and guide future data curation.
- Strengthen the use of visual evidence in multimodal model reasoning.

PUBLICATIONS

(* Equal contribution; ____ Mentored student; † Corresponding author)

Preprints and Papers in Submission

[p.0] Ziheng Zhang, Neha Hulkund, Colin Lee, Zheda Mai, Hui Shen, Ying Ding, Tanya Berger-Wolf, Sara Beery, Wei-Lun Chao, and **Jianyang Gu**†. Mind the Tail: Data Selection Through Sparse Concepts. *arXiv*, 2610.21007 (2026).

[p.1] Connor Kilrain, David Carlyn, Julia Chae, Sara Beery, Wei-Lun Chao, **Jianyang Gu**†. Finer-Personalization Rank: Fine-Grained Retrieval Examines Identity Preservation for Personalized Generation. *arXiv*, 2512.19026 (2025).

[p.2] Vyacheslav Kungurtsev, Yuanfang Peng, **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] Anirudh Potlapally, Brijesh Nanda Brunda Yogananda, **Jianyang Gu**, Maanav Jajodia, Luke Meyers, Wei-Lun Chao, Hari Subramoni. Resolve, Clear, or Defer: Training-Free Active Prototype Triage for Ecological Detection Pipelines. *WACV* (2027).

[c.1] David Carlyn, **Jianyang Gu**, Wei-Lun Chao. ModelEye: A Generative Lens for Understanding Vision Foundation Models. *WACV* (2027).

[c.2] Dongyao Zhu, Zhen Wang, Xi Xiao, Han Jiang, Saeed Vahidian, Wei-Lun Chao, Tanya Berger-Wolf, Yu Su, Raju Vatsavai, **Jianyang Gu**†. Leveraging Latent Visual Reasoning in Silence. *NeurIPS* (2026).

[c.3] Yue Cao, Mohsen Zardadi, Yu Hu, Yanshuo Fan, Jozsef Hamari, Zheng Liu, **Jianyang Gu**†. LCD^3: Layout-Conditioned Diffusion for Dataset Distillation in Object Detection. *NeurIPS* (2026).

[c.4] Yu Hu, **Jianyang Gu**, Yue Cao, Hao Liu, Kangnan Wang, Jozsef Hamari, Zheng Liu, Mohsen Zardadi. RAVEN-Bench: A Paired EO-IR Video QA Benchmark for Aerial Multimodal Understanding. *NeurIPS* (2026).

[c.5] Mridul Khurana, Amin Karimi Monsefi, Justin Lee, Medha Sawhney, David Carlyn, Julia Chae, **Jianyang Gu**, Rajiv Ramnath, Sara Beery, Wei-Lun Chao, Anuj Karpatne, Cheng Zhang. TaxaAdapter: Vision Taxonomy Models are Key to Fine-grained Image Generation over the Tree of Life. *NeurIPS* (2026).

[c.6] Zihang Fu, Fanxiao Li, **Jianyang Gu**, Haonan Wang, Preslav Nakov, Bryan Hooi, Min-Yen Kan, Jiaying Wu. Better with Experience: Self-Evolving LLM Agents for Evidence-Grounded Health Community Notes. *EMNLP Findings* (2026).

[c.7] Jieyuan Liu, **Jianyang Gu**, Shijie Chen, Jefferson Chen, Zhen Wang. Lost at the End: Primacy Bias in Multimodal Retrieval-Augmented Question Answering. *EMNLP* (2026).

[c.8] Yue Cao, **Jianyang Gu**, Vyacheslav Kungurtsev, Yu Hu, Jozsef Hamari, Zheng Liu, Mohsen Zardadi. Structural Assessment for Understanding and Guiding Dataset Distillation in Discrete Token Space. *ECCV* (2026).

[c.9] Xi Xiao, Chen Liu, Chih-Ting Liao, Yunbei Zhang, Qizhen Lan, Yuxiang Wei, Lin Zhao, Janet Wang, **Jianyang Gu**†, Muchao Ye, Tianyang Wang†, Hao Xu. Staying Vigilant: Mitigating Visual Laziness in MLLMs via Information-Theoretic Alignment. *ECCV* (2026).

[c.10] Sooyoung Jeon, Hongjie Tian, Lemeng Wang, Zheda Mai, Vidhi Bakshi, Jiacheng Hou, Ping Zhang, Arpita Chowdhury, **Jianyang Gu**, Wei-Lun Chao. Lessons and Open Questions from a Unified Study of Camera-Trap Species Recognition Over Time. *ECCV* (2026).

[c.11] Qinsi Wang, Saeed Vahidian, Hancheng Ye, **Jianyang Gu**, Jianyi Zhang, Yiran Chen. SparseInfer: Accelerating Large Language Model Inference with Semantics-Inspired Adaptive Sparse Activation. *ICML* (2026).

[c.12] Lin Zhao, 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.13] Yu Hu*, **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.14] Ziheng Zhang, Xinyue Ma, 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.15] **Jianyang Gu**, Haonan Wang, Ruoxi Jia, Saeed Vahidian, Vyacheslav Kungurtsev, Wei Jiang, Yiran Chen. CONCORD: Concept-Informed Diffusion for Dataset Distillation. *WACV* (2026).

[c.16] **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.17] Ziheng Zhang*, **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.18] Arpita Chowdhury, 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.19] Lin Zhao, 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.20] Saeed Vahidian*, Mingyu Wang*, **Jianyang Gu***, Vyacheslav Kungurtsev, Wei Jiang, Yiran Chen. Group Distributionally Robust Dataset Distillation with Risk Minimization. *ICLR* (2025).

[c.21] **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.22] **Jianyang Gu**, Kai Wang, Wei Jiang, Yang You. Summarizing Stream Data for Memory-Restricted Online Continual Learning. *AAAI*, 12217–12225 (2024).

[c.23] 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.24] Yanqing Liu*, **Jianyang Gu***, Kai Wang, Zheng Zhu, Wei Jiang, Yang You. DREAM: Efficient Dataset Distillation by Representative Matching. *ICCV*, 17314–17324 (2023).

[c.25] Daquan Zhou*, Kai Wang*, **Jianyang Gu***, Xiangyu Peng, Dongze Lian, Yifan Zhang, Yang You, Jiashi Feng. Dataset Quantization. *ICCV*, 17205–17216 (2023).

[c.26] **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] Yao Lu, 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] Zhiyuan Tao, Srikumar Sastry, Matthew J Thompson, Elizabeth G Campolongo, Net Zhang, Ziheng Zhang, Hilmar Lapp, Yu Su, Tanya Berger-Wolf, Nathan Jacobs, Wei-Lun Chao, **Jianyang Gu**†. Beyond Flat Labels: Level-Restricted Contrastive Learning for Hierarchical Fine-Grained Vision Classification. *CVPR FGVC Workshop* (2026).

[w.1] David Carlyn, **Jianyang Gu**, Wei-Lun Chao. Simple Localized Counterfactuals for Visual Explanation. *CVPR HOW Workshop* (2026).

[w.2] 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.3] 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.4] Kai Wang*, **Jianyang Gu***, Daquan Zhou, Zheng Zhu, Wei Jiang, Yang You. DiM: Distilling Dataset into Generative Model. *ECCV DD Challenge Workshop* (2024).

[w.5] **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.6] 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.7] 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 PDA Travel 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

From Samples to Concepts: Toward Concept-Centric Data Representation

- Boston University, AIR Seminar, Boston, MA Oct, 2026

From Pixels to the Planet: Building Scalable and Grounded AI for Science

- Voxel51, remote Aug, 2026
- ESIP Semantic Technologies Committee, remote Jul, 2026
- University of Arizona, Tucson, AZ Feb, 2026

- Texas A&M University, guest lecture, remote Feb, 2026

NAIRR Overview / User Panel

- NSF ACCESS Regional Workshop, Columbus, OH Mar, 2026

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

- NAIRR Annual Meeting, Arlington, VA Mar, 2026
- Australian National University, guest lecture, 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

Conference Program Committee

- NAIRR Annual Meeting 2027
- Imageomics Conference 2027

Workshop Organization

- Co-organizer FloraPalooza 2026
- Co-organizer CV4Ecology Workshop @ ECCV2026
- Co-organizer Designing for Discovery Workshop @ Imageomics2026 & Evolution2026
- 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 2027, NeurIPS 2026, ICLR 2026

Conference Reviewer

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

Journal Reviewer

- IEEE TPAMI, IEEE TVCG, IEEE TMM, PNAS, PR, CVIU, IEEE TCSVT

Outreach Activities

- **Co-organizer**, AI Spy in COSI Science Festival in Columbus, OH — May. 2026
- **Mentor**, Interdisciplinary Career Pathways in Bio + Computer Science at OSU — Apr. 2026
- **Panelist**, Student AI Club of The Ohio State University — Sept. 2024
- **Chair**, Student AI Association of Zhejiang University — Aug. 2020 - Jun. 2021