Zhiyuan Tao

tao.623@buckeyemail.osu.edu | (614) 772-0489 | zhiyuantao623.github.io/personal_website

Research Interests

Multimodal representation learning; vision-language learning; compositional and structured reasoning; interpretable representations; controlled evaluation of foundation models.

Education

The Ohio State University — Columbus, OH
M.S. in Electrical and Computer Engineering (Thesis) · GPA: 3.6/4.0 · Advisor: Wei-Lun Chao — *Aug. 2024 – May 2026*

Nanjing University of Posts and Telecommunications — Nanjing, China
B.S. in Information Security · GPA: 3.3/5.0 (Top 20%) — *Sept. 2020 – June 2024*

Publications & Manuscripts

Peer-Reviewed

[1] **Beyond Flat Labels: Level-Restricted Contrastive Learning for Hierarchical Fine-Grained Vision Classification.**
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.
CVPR 2026 Workshop on Fine-Grained Visual Categorization (FGVC13), ***Oral Presentation****. arXiv:2606.21838*

Under Review & In Preparation

[2] **Balanced Candidate Pools, Not Level Labels: A Controlled Evaluation of Multi-Granularity Vision–Language Retrieval.**
Zhiyuan Tao, Qirui Wang.
Submitted to WACV 2027, Datasets Track.

[3] **Input-Aligned Grid-State Handoffs Enable Visual Operations to Compose.**
Zhiyuan Tao.
Under review at ICLR 2027.

[4] **Harder Is Not Less Safe: Independent Task Demand Does Not Drive Multitask Jailbreaks in Vision-Language Models.**
Zhiyuan Tao.
Manuscript in preparation for NAACL.

Research Experience

Graduate Researcher, The Ohio State University — 2024 – 2026
Advisor: Prof. Wei-Lun Chao · *Collaborator: Dr. Jianyang Gu* — *First author, CVPR 2026 FGVC Oral [1]*

- Investigated whether hierarchical consistency in vision-language representations requires explicit parent–child constraints or can emerge from image–text alignment alone.
- Conceived and implemented **level-restricted contrastive learning**, which prevents valid labels at different semantic levels from acting as false negatives and balances optimization across granularities.
- Built on BioCLIP and trained on TreeOfLife-10M; improved hierarchical consistency in Euclidean and hyperbolic spaces, with a **30.47% relative gain** in zero-shot average accuracy across levels over the prior SOTA (RCME) on iNat21.

- Led method development, implementation, experimental analysis, and contribution design; presented the work orally at the CVPR 2026 FGVC Workshop.

Independent Research 2026

Multi-Granularity Vision–Language Learning *First author, WACV 2027 submission [2]*

- Independently conceived and led, from formulation to writing, a controlled study of whether multi-granularity vision–language learning benefits from true semantic-level identities or from the candidate-pool structure induced by level restriction.
- Constructed **VG-MG50K**, a 50,000-image Visual Genome benchmark with four strictly nested captions per image, controlling semantic-level identity, candidate-pool balance, and training objective while preserving image–caption membership.
- Showed across ten paired seeds that balanced random partitions match true semantic levels in retrieval (equivalent within ± 0.30 mean recall) once candidate-pool structure is matched, locating the restricted-softmax gain in balanced per-image pool construction.
- Developed a **membership-only sigmoid objective** that uses only image–caption membership (no level labels), reaching the strongest controlled result: 30.41 vs. 30.00 mean recall for level-restricted softmax.

Compositional Visual Execution *Sole author, ICLR 2027 submission [3]*

- Investigated why learned visual operations that are individually executable fail when composed into unseen programs, using controlled grid-to-grid visual reasoning tasks.
- Developed an *input-aligned* execution strategy that returns each intermediate grid through the visual interface consumed by the next operation, instead of keeping task state implicit in latent activations.
- Under parameter-matched training, raised exact accuracy on held-out primitive pairs from **43.5% to 88.2%**, persisting on longer programs.
- Isolated the interface effect in a frozen executor: changing only the state handoff raised held-out-pair accuracy from **29.2% to 87.6%**.
- Showed with probes and boundary controls that latent states encode the intermediate grid yet fail to support composition; extended results to grounded-object operations and demonstration-recovered programs.

Research Assistant, William & Mary 2026 – Present

Advisor: Prof. Qingyun Wang *Manuscript in preparation [4]*

- Formulated a controlled test of whether increasing the difficulty of an independent benign task makes vision-language models less safe under a fixed harmful request, via a benign-Task1 / harmful-Task2 framework.
- Built the full inference and evaluation pipeline across three VLMs, two safety benchmarks, and multiple benign task families.
- Found that benign-task performance can drop by up to **82.5 points** while attack success changes by at most **2.8 points**; matched protocol controls revealed that response formatting alone shifts ASR by **14–15 points** on VLJailbreakBench.

Additional Research & Teaching

- **Contributor, TreeOfLife-200M Dataset.** Contributed to data-processing infrastructure for the TreeOfLife-200M biodiversity dataset; acknowledged by the dataset authors.
- **Teaching Assistant, ECE 6001, The Ohio State University.** Assisted with course instruction and student support.

Technical Skills

Programming	Python, C/C++, Bash
Machine Learning	PyTorch, Hugging Face Transformers, OpenCLIP, NumPy
Research & Systems	Linux, Slurm, Git, LaTeX, Matplotlib