

Wenhao Zhang

Ph.D. Candidate in Data Science, University of Virginia
wenhaozhang0066.github.io · (434) 437-8911 · bdu8us@virginia.edu

RESEARCH INTERESTS

Vision foundation models and multimodal representation learning; post-training and preference optimization of multimodal LLMs; data-efficient fine-tuning; trustworthy evaluation of VLMs and LLM agents.

EDUCATION

University of Virginia

Ph.D. in Data Science (Candidate since Apr 2026). Advisor: Prof. Sheng Li Aug 2025 – May 2029 (Expected)

Xidian University

M.E. in Artificial Intelligence – GPA: 90.42/100 Aug 2022 – June 2025

Xidian University

B.E. in Computer Science – GPA: 3.7/4.0 Aug 2018 – June 2022

RESEARCH EXPERIENCE

University of Virginia — Graduate Research Assistant Charlottesville, VA · Aug 2025 – Present

- Showed that accuracy alone does not measure evidence use: in a lymph-node metastasis study, most public pathology VLMs barely changed when given the whole slide, an annotated lesion, or no image at all.
- Built **CleanSlide**, a TCGA-based whole-slide VQA benchmark of 149K multiple-choice questions over 9,985 slides with **patient- and tissue-source-disjoint splits**, each question audited for option shortcuts, stem leakage, cross-split duplication, and blind solvability (github.com/wenhaozhang0066/CleanSlide).
- Designed **Pair-DPO**, a preference loss over counterfactual pairs of **real slides with opposite, verified findings** for the same question and source, which cancels the question-attributable signal and leaves image evidence as the source of preference; used it to train a 7B pathology VLM (frozen CONCH tile encoder, LongNet slide aggregator, Qwen2.5-7B-Instruct with a rank-16 LoRA adapter).
- Achieved a **15.29%** gain from image evidence on CleanSlide, versus 2.81% for the best published model, and 57.6% and 59.0% on the external CPTAC and BCNB cohorts, **outperforming all evaluated models by 9.7% and 3.4%** (first author, ICLR 2027 submission).
- Audited patient- and institution-level leakage in whole-slide multimodal benchmarks: **92.3–100% case-level train–test overlap** on TCGA-derived benchmarks, both levels linearly decodable from foundation-model features, and reported accuracies concentrated on the most contaminated ones (first author, **WACV 2027**).
- Built and open-sourced **SlideTwin** (PyPI: `slidetwin`), a whole-slide image viewer simulator where pathologists and AI agents read under the same microscope rules via one session API (web UI, REST, MCP, gym-style Python Env), with a virtual clock, replayable pixel logs, and protocol checks on every report; in one replay, an open VLM reported a micro-metastasis after viewing only 0.4% of the tissue at 10× or higher.
- Audited sampled decisions in self-evolving agents (AgentDojo, Agent-SafetyBench, Terminal-Bench 2): accept tests misjudge 13–37% of clearly separated pairs, selection picks a worse candidate in up to 54% of runs, and reported gains are off by 5+ points in up to 78% of cases (first author, NAACL 2027 submission).
- Built a three-level LLM image-description pipeline over 15 public pathology datasets, yielding Balance9k; fine-tuning DINOv3 on its 9K curated samples rivals million-scale fine-tuning, and **Cell-Guided Attention** with organ-agnostic learning improves cross-organ generalization by ~30%.
- Leading a survey of pathology VLMs centered on an evidence-graded paradigm–task suitability matrix and maintaining **Awesome-Pathology-VLMs**, an open index of pathology VLMs, datasets, and benchmarks.

Xidian University — Graduate Research Assistant Xi'an, Shaanxi · Aug 2022 – June 2025

- Proposed **PedMix**, a region-based cross-modality augmentation, and a **Modality Feature Transfer** module for visible-infrared person re-ID, improving mAP by 4.67% over the baseline (Scientific Reports, 2025).
- Designed a lightweight directional-aware network built on non-learned operators for image classification (Artificial Intelligence Review, 2024); three Chinese invention patents filed.

- Built a patch-wise regression pipeline over 3D volumetric medical image data, replacing the convolutional stages of a U-Net with Swin Transformer blocks to predict subject age from local image features.

TECHNICAL SKILLS

Programming: Python, C++, R, Bash, LaTeX

Deep Learning & Computer Vision: PyTorch, Hugging Face, timm, NumPy, scikit-image; ViT, DINOv3, Swin, CNNs, U-Net, vision-language models, cross-attention; self-supervised and data-efficient fine-tuning

Post-Training: supervised fine-tuning, DPO and pairwise preference optimization, LoRA / parameter-efficient fine-tuning, multi-seed and step-matched controlled evaluation, multi-GPU training

Agents & Evaluation: REST / MCP / gym-style environments, vLLM tool calling, leakage audits, paired tests

Tools: Git, Linux, Docker, Slurm, Weights & Biases

PUBLICATIONS

W. Zhang, Z. Zhou, S. Zhang, et al. “Contamination, Prior, or Evidence? Decomposing and Training Evidence Use in Whole-Slide Vision–Language Models.” Under review, ICLR 2027.

W. Zhang, S. Zhang, D. Qi, et al. “Too Small to Tell: Auditing Sampled Decisions in Self-Evolving Agent Harnesses.” Under review, NAACL 2027.

Y. Yang, G. Sun, P. Wang, **W. Zhang**, et al. “Scores High Alone, Ranks Low in Context: Full-Context Knockout Frame Selection with an MLLM for Long-Video Question Answering.” Under review, ICLR 2027.

P. Wang, H. Yuan, Y. Yang, **W. Zhang**, S. Li. “TabMATE: A Multi-Agent Framework for Tabular Anomaly Detection.” Under review, ICLR 2027.

W. Zhang, Z. Zhou, J. Kang, S. Li. “Auditing Data Leakage in Whole-Slide Image Multimodal Benchmarks.” WACV 2027.

L. Yang, D. Qi, **W. Zhang**, et al. “PixelDense: Dense Prediction as Representation Alignment for Pixel Diffusion.” NeurIPS 2026.

Y. Yang, Y. Li, D. Qi, L. Yang, T. Wang, **W. Zhang**, et al. “HFS: Holistic Query-Aware Frame Selection for Efficient Video Understanding.” ACM MM 2026.

H. Yuan, M. Hu, **W. Zhang**, et al. “Reducing Token Redundancy in LVLMs: A Systematic Review of Token Pruning Methods.” ACL 2026 (Long Papers).

T. Wang, L. Yang, Y. Liu, H. Yuan, **W. Zhang**, et al. “Evaluating Spatial World Modeling in Video Generators via 3D Camera Trajectory Generation.” ICML 2026 Workshop on Foundations of Deep Generative Models.

Y. Guo, **W. Zhang**, L. Jiao, et al. “Visible-Infrared Person Re-Identification with Region-Based Augmentation and Cross-Modality Attention.” Scientific Reports, 2025.

Y. Guo, **W. Zhang**, Y. Gao, et al. “Revisiting Non-Learned Operators Based Deep Learning: A Lightweight Directional-Aware Network for Classification.” Artificial Intelligence Review, 2024.

W. Zhang, et al. “Vision–Language Models for Computational Pathology: A Survey of Paradigms, Granularities, and Benchmarks.” Manuscript in preparation, 2026.

W. Zhang, Z. Zhou, S. Li. “LANTERN: Learning a Data-Efficient Pathology Foundation Model through Multimodal Disentangled Representations.” Manuscript in preparation, 2025.

ACADEMIC SERVICE

Program Co-Chair: 1st Workshop on Big Data for Computational Pathology and Multi-Omics: From Pixels to Biomarkers (P2B), in conjunction with IEEE BigData 2026, Phoenix, AZ.

Reviewer: CVPR 2026, ACM MM 2026, ACL Rolling Review (ARR) May 2026