

MINGYI HUANG

Beijing, China | Columbus, Ohio, U.S. | +1-(614)-371-8148 | huang.5749@osu.edu | hmy02.github.io

Research Interest

My research interests center on building reliable machine learning systems for dynamic environments. My prior work focused on continual learning, and I am now working on time-series anomaly detection.

Education

The Ohio State University

PhD in Computer Science and Engineering
Supervised by Prof. John Paparrizos

Expected June 2029
Columbus, Ohio, U.S.

Beijing University of Technology

BS in Statistics, Minor in Computer Science and Technology
Major GPA: 3.93/4.0 (Rank 1/26); Minor GPA: 4.0/4.0
Awards: CUMCM 2021 Provincial Second Prize; BJUT Scholarship (2022, 2023); BJUT Faculty of Science Scholarship (2021)

June 2024
Beijing, China

Publications

- [1] **Huang, M.**, Liu, Q., Boniol, P., Paparrizos, J. (2026). *HYDRA: A Multi-Level Hierarchy-Driven Approach for Robust Anomaly Detection in Time Series*. Proceedings of the ACM on Management of Data, 4(3), Article 197.
- [2] **Huang, M.**, Liu, Q., Boniol, P., Paparrizos, J. (2026). *GlassboxAD: An Interactive System for Dissecting Hierarchical Time-Series Anomaly Detection*. SIGMOD Companion '26.
- [3] Boniol, P., Krishna, A., Bruel, M., Liu, Q., **Huang, M.**, Palpanas, T., Tsay, R., Elmore, A., Franklin, M., Paparrizos, J. (2025). *VUS: Effective and Efficient Accuracy Measures for Time-Series Anomaly Detection*. The VLDB Journal, 34(3), 32.
- [4] Boniol, P., Liu, Q., **Huang, M.**, Palpanas, T., Paparrizos, J. (2024). *Dive into Time-Series Anomaly Detection: A Decade Review*. arXiv preprint arXiv:2412.20512.
- [5] Wang, L., Xie, J., Zhang, X., **Huang, M.**, Su, H., & Zhu, J. (2023). *Hierarchical Decomposition of Prompt-Based Continual Learning: Rethinking Obscured Sub-optimality*. NeurIPS 2023 Spotlight.
- [6] **Huang, M.**, Wu, L., Wang, Y. (in submission). *Generalized OLS Calibration for Expensive Computer Models*.

Research Experience

OSU Computer Science and Engineering Department

PhD Student

Aug. 2024 – Present
Columbus, Ohio, U.S.

- Time-series anomaly detection with hierarchical, scalable, and interpretable frameworks.
- Worked on time-series anomaly detection research, focusing on scalable methods for long and high-frequency sequences in both univariate and multivariate settings.
- Designed and implemented a hierarchical reference-based detection pipeline that captures anomalies across multiple temporal resolutions while reducing sensitivity to normalization and hyperparameter tuning.
- Built end-to-end experimentation workflows and benchmarked against strong baselines on large-scale TSAD suites under multiple evaluation protocols.
- Developed interactive visualizations and tools to diagnose model behavior, interpret anomaly scores across levels, and support transparent analysis of detection results.

BUPT School of Artificial Intelligence

Research Assistant

Mar. 2024 – Aug. 2024
Beijing, China

- Physics-guided 3D reconstruction and wild inverse rendering with polarization and neural fields.
- Simplified a polarization-based imaging model under strong illumination by focusing on diffuse reflection, enabling robust surface normal recovery.
- Developed a hybrid pipeline combining classical photometric stereo constraints with Neural Radiance Fields for outdoor inverse rendering without a darkroom setup.
- Introduced a controllable flashlight-like lighting constraint to better handle specular highlights and complex geometry, improving fine-detail reconstruction.
- Unified insights from two related projects into a single framework to improve robustness and detail quality in real-world 3D reconstruction.

BJUT Institute of Applied Probability and Statistics

Undergraduate Researcher

Oct. 2022 – May 2024
Beijing, China

- Developed a method called: Generalized Ordinary Least Squares Calibration (GOLS Calibration).
- Used Gaussian Processes as a surrogate model to mimic outputs of complex computer experiments.
- Proved mathematically that additional variance term in the loss function can reduce fluctuation while maintaining accuracy.
- Reproduced other methods, such as KO and GALS calibration, using data from the MATLAB FEM toolbox.
- Applied GOLS to stochastic computer simulation.

Tsinghua Statistical Artificial Intelligence & Learning Group

Mar. 2023 – Aug. 2023

Undergraduate Researcher

Beijing, China

- Optimized task-identity inference and within-task prediction for continual learning models.
- Compiled core research papers in continual learning, made an open-source paper list, and published it on GitHub.
- Compared our method against four advanced prompt-based continual learning methods on existing benchmark.
- Conducted parameter tuning, recorded experiment results, visualized outcomes, and improved code execution.

Work Experience

JD, Research Institute for Consumption and Industrial

June 2022 – Aug. 2022

Data Analyst Intern

Beijing, China

- Coded a web crawler to extract social media post titles and corresponding gifting recommendations for Qixi (Chinese Valentine's Day) customer analysis.
- Screened, compared, and analyzed different consumption characteristics of normal and long-distance relationships.
- Created a consumption trend chart to visualize gift-purchasing dates among various age groups.
- Extracted, organized, and provided reference market data to a treadmill brand client.
- Applied LSTM to predict fitness consumer market trends.
- Collected home appliance sales data under rural replacement policies and compared 2008 and recent data using gray forecasting, ARIMA, LSTM, and related methods.

GuoTai Asset Management, Active Management Department

June 2023 – Aug. 2023

Quant Intern

Beijing, China

- Built a deep-learning stock prediction model combining candlestick-chart representations, convolutional filtering, and LSTM components.
- Improved the model through hyperparameter optimization and data normalization.
- Incorporated embedding-based coding, an attention layer, and dummy variables into a multi-task learning framework to enhance information coefficient (IC).
- Generated alpha factors from model results, conducted factor backtesting, and calculated metrics such as annualized return, which reached up to 28%.
- Mined analytical research reports, constructed attribute sets, and used a FinBERT pretrained model to encode text for cross-validated excess-information training in XGBoost.
- Studied the impact of institutional survey types and predicted survey-count trends based on ARIMA to further optimize stock-screening strategies.

Project Experience

Forecasting the Difficulty of Words in the Wordle Game

Feb. 2023

Contest Team Leader

- Predicted player counts in the Wordle game using an ARIMA model improved by a Markov chain.
- Fitted a Gaussian regression model to the distribution of user attempt frequencies for specific words.
- Categorized words by difficulty using K-means and validated the clustering results with AdaBoost.

Chemical Composition Identification of Ancient Glass Relics

Sept. 2022

Contest Team Leader

- Performed descriptive statistics on texture, category, and color to determine information gain on relic weathering.
- Conducted normality tests on chemical compositions to predict pre-weathering composition.
- Reduced 14 chemical composition attributes to 3 attributes using random forests.
- Used K-means to classify relics by composition and studied relationships among relic types using correlation matrices.

Water Sharing Strategy Solving Severe Drought in Five U.S. States

Feb. 2022

Contest Team Leader

- Estimated minimum-level water storage by fitting historical water level and storage data.
- Developed a model accounting for the water needs of five states drawing from two lakes.

- Maximized economic benefit while considering state GDP, industrial water use, and historical stability of water usage.

Optimizing Raw Material Supply Chains for Manufacturers

Sept. 2021

Contest Team Leader

- Applied time-series analysis to predict sixth-year material supply accuracy using data from the previous five years.
- Designed a rating system for supplier volume and delivery accuracy to identify key suppliers.
- Used MINLP to optimize transportation, storage, and ordering strategies in a redesigned supply-chain model.

Technical Skills

Programming: Python, Java, C, C++, R, SPSS, MATLAB, SQL, HTML

Tools: PyTorch, TensorFlow, Git, L^AT_EX