

# FAN YANG

+1 412-726-2154 | <https://ivan-fan.github.io> | [yang.7007@osu.edu](mailto:yang.7007@osu.edu)

## PERSONAL PROFILE

---

I'm a third-year Ph.D. student in Computer Science & Engineering at The Ohio State University, affiliated with the Datum Lab advised by Prof. John Paparrizos. My research focuses on **time-series analysis**, with emphasis on **efficient representation learning**, **multivariate time-series similarity search**, and **intermittent-demand forecasting**. These methods are designed for scalable, modular deployment and are validated on public benchmarks and large real-world supply chain data.

Currently, I am also exploring research directions in **pre-trained foundation models** and applications of **LLMs for time-series**, including multimodal cues when external signals improve prediction fidelity. Previously, my research spanned areas of **automatic speech recognition** and **multimodal learning** under the advice of Prof. Bhiksha Raj and Prof. Rita Singh at Carnegie Mellon University.

## EDUCATION

---

### The Ohio State University

Ph.D. in Computer Science and Engineering, GPA: 4.00 / 4.00

Columbus, OH  
Aug 2023 - Present

- **Primary Advisor:** Prof. John Paparrizos
- **Area:** Symbolic Time-series Approximation, Multivariate Time-series Similarity Search, and Intermittent Demand Forecasting
- **Core course:** Machine Learning and Statistical Pattern Recognition, Algorithms, Network Science, Software Security

### Carnegie Mellon University

M.S. in Electrical and Computer Engineering, GPA: 4.00 / 4.00

Pittsburgh, PA  
Aug 2021 - May 2023

- **Primary Advisor:** Prof. Bhiksha Raj
- **Area:** A Closer Look at Reinforcement Learning-based Automatic Speech Recognition
- **Core course:** Machine Learning, Deep Learning, ML for Signal Processing, Computer Vision, Advanced Natural Language Processing, (Bio) medical image analysis, Convex Optimization

### Nanjing University of Posts and Telecommunications

B.E. in Communication Engineering, GPA: 4.30 / 5.00 (91.39 / 100)

Nanjing, China  
Sept 2016 - June 2020

- **Core course:** Advanced Mathematics, Linear Algebra, Probability and Stochastic Process, Signals and Systems, Digital Signal Processing, Communication Principles, Digital Circuits and Logic Designs

## PUBLICATION

---

### Pre-print

- Paparrizos, J., **Yang, F.**, Li, H. Bridging the Gap: A Decade Review of Time-Series Clustering Methods. *arXiv preprint arXiv:2412.20582*. 2024. [\[Paper\]](#)
- Paparrizos, J., Li, H., **Yang, F.**, Wu, K., d'Hondt, JE., Papapetrou, O. A Survey on Time-Series Distance Measures. *arXiv preprint arXiv:2412.20574*. [\[Paper\]](#)

### Conference

- **Yang, F.**, Paparrizos, J. SPARTAN: Data-Adaptive Symbolic Time-Series Approximation. *Proceedings of the ACM on Management of Data (SIGMOD)*, 2025. [\[Paper\]](#)
- **Yang, F.**, Paparrizos, J. SAIL: A Voyage to Symbolic Approximation Solutions for Time-Series Analysis. *Proceedings of the VLDB Endowment*, 2025. [\[Paper\]](#)
- d'Hondt, JE\*, Li, H\*, **Yang, F\***, Papapetrou, O., Paparrizos, J., A Structured Study of Multivariate Time-Series Distance Measures. *Proceedings of the ACM on Management of Data (SIGMOD)*, 2025. [\[Paper\]](#)
- Li, X., Wang, J., Xu X., Yang M., **Yang F.**, Zhao Y., Singh R., Raj B. Towards noise-tolerant speech-referring video object segmentation: Bridging speech and text. *EMNLP*, 2023. [\[Paper\]](#)

### Journal

- **Yang, F.**, Yang, M., Li, X., Wu, X., Raj, B., Singh R. A Closer Look at Reinforcement Learning-based Automatic Speech Recognition. *Computer Speech & Language*, 2024. [\[Paper\]](#)
- Jiang, Y., **Yang, F.**, Bian, Z., Lu, C., Xia, S. Mask removal: Face inpainting via attributes. *Multimedia Tools and Applications*, 1-13, 2022. [\[Paper\]](#)

## RESEARCH PROJECT

---

### SPARTAN: Data-Adaptive Symbolic Time-series Approximation

Advisor: Prof. John Paparrizos | The Ohio State University

Columbus, OH

May 2024 - Oct 2024

- Proposed a novel data-adaptive time-series symbolic approximation method (**SPARTAN**) that constructs non-overlapping and uncorrelated latent dimensions and intelligently allocates the encoding budget according to the importance of each dimension.
- Proposed a novel symbolic distance **PCA-MINDIST** with SPARTAN that **lower-bounds Euclidean distance** on the original space.
- Demonstrated the superior representation power of SPARTAN against **7** state-of-the-art methods across **4** analytical tasks, including **classification, clustering, indexing and anomaly detection**, while delivering up to a **2x** speedup to the most accurate rival.

### A Structured Study of Multivariate Time-Series Distance Measures

Advisor: Prof. John Paparrizos, Prof. Odysseas Papapetrou | The Ohio State University & TU/e

Columbus, OH

April 2024 - Oct 2024

- Formalized **three key axes**—normalization, temporal model, and channel-dependency model—establishing a standard guideline for structured studies of multivariate time-series (MTS) distance measures, along with a proposed taxonomy of **8** major categories.
- Implemented, properly parameterized, and evaluated **30** MTS measures across **30** datasets with **13** normalization techniques—conducting the most comprehensive study in this area—assessing their accuracy and scalability across 3 downstream tasks
- Developed an **easy-to-use library** implementing all evaluated MTS measures as an important byproduct of this study, providing clear guidelines for effectively selecting the appropriate MTS distance measure and preprocessing technique in downstream tasks.

### A Closer Look at Reinforcement Learning-based Automatic Speech Recognition

Advisor: Prof. Bhiksha Raj, Prof. Rita Singh | Carnegie Mellon University

Pittsburgh, PA

Nov 2022 - May 2023

- Proposed a pure reinforcement learning (RL) setting wherein the speech recognition models are trained solely on **human feedback**, e.g., word error rate (WER), and demonstrated its feasibility through extensive experiments on the real-world speech corpus.
- Devised a **unified RL formulation** for diverse sampling strategies, effectively bridging the gap between action and reward with a new RL training method aligned with Connectionist Temporal Classification (CTC) loss (**CTC-PG**) in speech recognition context.
- Validated the robustness of the proposed methods across diverse reward designs in both clean and noisy environments.

## EXPERIENCE

---

### Intermittent Time-series Forecasting in Real-world Supply Chain Data

Summer Intern Scientist | Mentors: Dr. Sitan Yang and Raunak Singh | Keystone.AI

New York, NY

Jun 2025 - Aug 2025

- Contributed to the development of an **in-house time-series forecasting library** with full support for model training, evaluation, and analysis on large-scale supply chain datasets.
- Proposed a novel plug-and-play **temporal positional encoding** layer for sequence modeling in time-series forecasting, explicitly encoding **interdemand temporal pattern** to address event sparsity; achieved a **12% WAPE** reduction on internal real-world supply chain data with strong generalization to public benchmarks.
- Designed an end-to-end time-series forecasting pipeline leveraging **pre-order signals** from upstream data, achieving up to **25% WAPE improvement** for short-horizon forecasts and **10% on average** across 1–8 week horizons.

### Object Detection and Classification in Surveillance System

Research Assistant | Advisor: Prof. Marios Savvides | CyLab, Carnegie Mellon University & Oosto Co.

Pittsburgh, PA

Jan 2022 - May 2022

- Proposed a novel **two-stage object detection pipeline**, which combines human pose estimation, hand region proposal, and hand-held object classification to alleviate the false positive issues in complex environments.
- Devised an adaptation mechanism in hand region proposal generation with improved detection accuracy by **17%**.

## TECHNICAL SKILLS

---

**Programming Languages:** Python, C/C++, MATLAB, Verilog

**Tools & Frameworks:** PyTorch, TensorFlow, OpenCV, SQL, Docker, Git

## ACADEMIC SERVICE

---

**Reviewer:** (Conference) EMNLP, and (Journal) Data Mining and Knowledge Discovery, Computer Speech & Language

**Teaching Assistant:**

- OSU CSE 5243 Intro to Data Mining. Supervisor: Prof. John Paparrizos / Prof. Ning Xia
- CMU ECE 18-786 Intro to Deep Learning. Supervisor: Prof. Aswin Sankaranarayanan

*Aug 2024 - Present*

*Jan 2023 - May 2023*