Yunfan Zhang

530 W 120th St, New York, NY 10024 | (919) 564-9552 | yz4244@columbia.edu | www.yunfanzhang.me

EDUCATION

Columbia University   New York, NY
Ph.D. Student in Computer Science   *September 2022 – May 2027 (Expected)*
Duke University   Durham, NC
B.S. in Computer Science, Minor in History   *August 2016 – December 2020*

RESEARCH EXPERIENCE

Columbia University   New York, NY
Research Associate   *August 2021 – Present*

- Worked with Prof. Ethan Katz-Bassett at Systems and Networking Lab.
- Conducted internet measurement research focusing on user behavior, performance, reliability, and security.
- Developed Large Language Model (LLM) techniques for addressing Internet measurement challenges, including Autonomous Systems classification, rDNS inference, and device identification.
- Devised novel methods to infer user activity patterns utilizing public internet services and datasets, such as Google Public DNS, M-Lab Internet Speed Test Dataset, Censys Universal Internet Dataset, and DNS root server logs (DITL).
- Implemented a Gradient Boosting-based ML model for predicting user activity patterns from DNS behavior and performed text analysis on terabyte-scale datasets using Google BigQuery. Developed high-performance networking code in Golang, deployed and maintained Internet measurement software across 20+ geographically distributed vantage points.

Duke University, Intelligent Interactive Internet of Things Lab   Durham, NC
Research Associate   *May 2019 – July 2021*

- Advised by Prof. Maria Gorlatova and Dr. Guohao Lan.
- Conducted research in computer vision, applied machine learning, and mobile computing, with a focus of leveraging edge-based computation offloading techniques to enhance the performance of mobile computer vision and AR applications.
- Proposed a deep learning based inpainting and denoising technique for consumer RGB-Depth cameras. Designed an edge-based computation offloading system to achieve real-time (>30 FPS) latency and throughput on mobile devices. Utilized refined RGB-Depth data to improve user experience in mobile AR applications.
- Prototyped computer vision and image processing algorithms with PyTorch, OpenCV, and SciPy. Implemented AR demos with Magic Leap, Unity, and ARCore. Conducted performance benchmarking and optimizations for the lab's prototype applications.

PUBLICATIONS

- Who Squats IPv4 Addresses?
  Loqman Salamatian, Todd Arnold, Ítalo Cunha, Jiangchen Zhu, **Yunfan Zhang**, Ethan Katz-Bassett, Matt Calder
  To Appear In *ACM SIGCOMM Computer Communication Review, 2023 (CCR 2023)*
- InDepth: Real-time Depth Inpainting for Mobile Augmented Reality
  Yunfan Zhang, Tim Scargill, Ashutosh Vaishnav, Gopika Premsankar, Mario Di Francesco, Maria Gorlatova
  Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, Mar. 2022 (IMWUT/UbiComp 2022)
- Optimal Network Protocol Selection for Competing Flows via Online Learning
  Xiaoxi Zhang, Siqi Chen, **Yunfan Zhang**, Youngbin Im, Maria Gorlatova, Sangtae Ha, Carlee Joe-Wong
  IEEE Transactions on Mobile Computing, 2022 (TMC 2022)
- Edge-assisted Collaborative Image Recognition for Mobile Augmented Reality
  Guohao Lan, Zida Liu, **Yunfan Zhang**, Tim Scargill, Jovan Stojkovic, Carlee Joe-Wong, Maria Gorlatova
  ACM Transactions on Sensor Networks, Vol. 18, No. 1, Feb. 2022 (TOSN 2022)
- Towards Identifying Networks with Internet Clients Using Public Data
  Weifan Jiang, Tao Luo, Thomas Koch, **Yunfan Zhang**, Ethan Katz-Bassett, Matt Calder
  21st ACM Internet Measurement Conference (IMC 2021)
- Towards a Traffic Map of the Internet: Connecting the Dots between Popular Services and Users
  Thomas Koch, Weifan Jiang, Tao Luo, Petros Gigis, **Yunfan Zhang**, Kevin Vermeulen, Emile Aben, Matt Calder, Ethan Katz-Bassett, Lefteris Manassakis, Georgios Smaragdakis, Narseo Vallina-Rodriguez
  20th ACM Workshop on Hot Topics in Networks (HotNets 2021)
- CollabAR: Edge-assisted Collaborative Image Recognition for Mobile Augmented Reality
  Zida Liu, Guohao Lan, Jovan Stojkovic, **Yunfan Zhang**, Carlee Joe-Wong, Maria Gorlatova
  19th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN 2020)
- Edge-based Provisioning of Holographic Content for Contextual and Personalized Augmented Reality
  Michael Glushakov, **Yunfan Zhang**, Yuqi Han, Tim Scargill, Guohao Lan, Maria Gorlatova
  2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops 2020)

SELECTED PROJECTS

Duke Electric Vehicles Team   Durham, NC
Software Developer and Data Analyst   *October 2016 – August 2019*
- Designed and built the world's most efficient electric vehicle (27,482 MPGe, current Guinness world record holder) and the world's most fuel-efficient vehicle (14,573 MPGe, current Guinness world record holder) with the team.
- Developed a real-time vehicle telemetry and data analysis system with a custom-designed Bluetooth module (for data transmission), an Android app (as vehicle dashboard), a Python web backend (for data analysis), and a JavaScript web frontend (for data visualization). Reconstructed vehicle dynamics parameters such as speed, acceleration, and elevation from noisy sensor readings using LOWESS regression. Built a driving strategy planner using dynamic programming to improve driving efficiency.
- Contributed to the club's autonomous driving project. Designed the vehicle's localization system by implementing Extended Kalman Filter based sensor fusion using data from ORB-SLAM based visual odometry, RTK GPS, and IMU.
- Contributed to the vehicle's trapezoidal motor control software. Implemented drivers for vehicle sensors such as IMU and magnetic motor encoder.

KuzoClass   Durham, NC
Lead Software Developer, Cofounder   *February 2019 – August 2019*
- Co-founded KuzoClass, an online education website focused on teaching business-building knowledge and expertise to entrepreneurs. Served as the lead of software engineering until the company received $250K investments with a $5M valuation.
- Responsible for developing the alpha version of the company's web application. Developed the website backend with Python and PostgreSQL, with support for complex features such as accepting online payment, video streaming, and instant messaging. Supervised and contributed to the company's frontend development in JavaScript and React.

Duke University, Laboratory for Unconventional Conflict Analysis & Simulation (LUCAS)   Durham, NC
Software Developer   *September 2017 – May 2018*
- Accelerated complex NLP and simulation workloads by implementing multithreaded task queues and in-memory caches. Used cProfile to debug and optimize memory usage to avoid memory leaks and out of memory errors when running large NLP queries.
- Served as a software developer to the lab's data collection and intelligence analysis platform. Integrated NLP models and simulation algorithms developed by other contributors into the lab's platform.

INTERNSHIP EXPERIENCE

Cloudflare   New York, NY
Research Intern   *June 2023 – September 2023*
- Developed a prototype bot traffic detection system using Language Models.
- Reduced the bot detection error on test datasets by over 80%.
- Implemented data engineering pipelines with ClickHouse SQL. Performed data engineering on O(100 Billion) rows of database records.

Red Hat   Durham, NC
Software Engineering Intern   *May 2018 – August 2018*
- Worked on Ansible and Ansible Tower, the company's distributed server management and orchestration tool.
- Developed new features related to DSL parser, OAuth, user management, auditing, and command line tools.
- Identified and resolved challenging bugs such as race conditions and deadlocks during the software's distributed job execution.
- Reviewed and moderated community issues and contributions for Ansible GitHub repository (a top 10 Python project on GitHub by popularity, with over 40K stars and 4K contributors) and Ansible Tower GitHub repository (over 8K stars and 200 contributors)

TEACHING EXPERIENCE

CSEE 4119: Computer Networks, Columbia University   New York, NY
Teaching Assistant   *September 2023 – December 2023*
CS 356: Computer Network Architecture, Duke University   Durham, NC
Teaching Assistant   *September 2019 – April 2020*

HONORS AND AWARDS
- Dean's List with Distinction, Spring 2019, Fall 2019

As a member of Duke Electric Vehicles Team
- Guinness World Records, Most Efficient Electric Vehicle, 27,482 MPGe, July 2019
- Guinness World Records, Most Fuel-Efficient Vehicle, 14,573 MPGe, July 2018
- Shell Eco-Marathon Americas 2018, Technical Innovation Award, First place in Hydrogen Track and Battery-Electric Track
- Shell Eco-Marathon Americas 2017, First place in Battery-Electric Track

TECHNICAL SKILLS
Research, Machine Learning and Deep Learning, Computer Vision, Multimodal Learning, NLP, Python (OpenCV, PyTorch, SciPy, NumPy, JAX, Web Development), Golang, Java, Android, C++ (OpenCV, NS3, Embedded Development), C# (Unity), JavaScript