

Shuning Jiang

Phone: (380) 895-3570
Email: jiang.2126@osu.edu
Website: www.jiangsn.com

Ph.D. candidate with a focus on machine learning and large-scale data processing, bringing a strong foundation from academic research and hands-on industry experience. Seeking a full-time software engineering role to apply robust problem-solving skills to build impactful, data-driven products.

Education

- **Ph.D. Student, The Ohio State University, Computer Science and Engineering** 08/2019 - current
 - Cumulative GPA: 3.52
 - Research Assistant: Interactive Visual Computing Lab
- **B.S., University of Electronic Science and Technology of China, Software Engineering** 09/2015 - 07/2019
 - Thesis: Automatic Vehicle Detection and License Plate Recognition System

Skills

- **Programming Languages:** Python, C/C++, JavaScript, HTML/CSS, Java, R
- **Framework/Libraries:** TensorFlow, Keras, PyTorch, OpenCV, Qt, OpenGL
- **Tools:** Git, Linux, MySQL, Nginx, Selenium, Unity

Internships

- **Software Engineering Intern | Google LLC | Python, C++, SQL, Internal AI Infrastructures** 05/2025 – 08/2025
 - Designed a C++ data processing pipeline that augments the modality of both datasets and core ML inference systems. This work established a generalizable workflow for evaluating the vision capabilities of all current and future LLMs.
 - Pioneered an end-to-end methodology to build clean and challenging datasets using a dual-sourcing approach with model ensembling and LLM-based judging. This created a repeatable blueprint for the team's future data refinement.
 - Conducted sophisticated statistical analyses to identify key drivers of model accuracy. The resulting insights on model behavior and ensembling strategies significantly advanced the team's ability to guide future model development.

Selected Publications

- *(To be published)* **A Rigorous Behavior Assessment of CNNs' Graphical Perception Using a Sampling Regime,** [S. Jiang](#), W. Chao, D. Haehn, H. Pfister, J. Chen, *IEEE Visualization*, 2025
- **Enhancing Tobacco Product Information Extraction from Online Stores Using Large Language Models,** [S. Jiang](#), S. Ma, J. Chen, C. Shang, *Society for Research on Nicotine and Tobacco*, 2025
- **Use of Machine Learning Tools in Evidence Synthesis of Tobacco Use Among Sexual and Gender Diverse Populations: Algorithm Development and Validation,** S. Ma, [S. Jiang](#), O. Yang, X. Zhang, Y. Fu, Y. Zhang, M. Ling, J. Chen, C. Shang, *JMIR Formative Research*, 2024

Selected Research Projects

- **Online vape shop scraping | Python, BeautifulSoup, Selenium, PyTorch, spaCy** 02/2021 – present
 - Scraped online vape products using various tools and techniques, such as Selenium for structured HTML content, YOLO and OCR for images, and spaCy and LLM for unstructured textual information.
 - Identified 134 new terms and phrases of flavor descriptions using spaCy and LLM.
- **Pathologists' gaze analysis | JavaScript, OpenSeadragon, Python, PyTorch** 09/2022 – 2025
 - Engineered a toolkit for tumor annotations while simultaneously tracking pathologists' eye movements. Visualized pathologists' eye gaze data and analyzed the gaze pattern. Removed noisy data based on pattern analysis.
 - Designed an algorithm that extracts pseudo ground truths from gaze data and demonstrated the feasibility of training neural networks using pseudo ground truth alone while maintaining comparable accuracy

- **Quantifying image complexity** | *JavaScript, Node.js, PHP, MySQL, Nginx, Python* 06/2023 – 03/2025
 - Built a full-stack online experiment platform using jsPsych and MySQL. Used Nginx for load balancing to support tens of thousands of concurrent requests and easily scale up.
 - Calculated images' complexity scores using TrueSkill algorithm. Categorized the factors that make images complex such as shape/color diversity and clutters.
- **Brain imaging annotation toolkit** | *C++, Qt, OpenGL, OpenVR* 08/2019 – 02/2023
 - Developed a toolkit on top of Vaa3D for evaluating participants' annotation speed and pattern in virtual reality or desktop environments.
 - Compared accuracy in both environments and found that annotation in VR environment is 12.9% more accurate.
- **Similarity-based pathology image retrieval** | *Python, TensorFlow, OpenSeadragon, D3.js* 08/2020 – 09/2021
 - Benchmarked multiple image feature extraction algorithms and retrieved similar images in the database.
 - Developed a web application enabling users to select custom areas of whole-slide images and receive recommendations of similar images.
- **Graphical perception with CNNs** | *Python, Keras* 08/2019 – 02/2023
 - Evaluated convolutional neural networks' performance on graphical perception tasks and fairly compared it to human performance. Quantified the effect of data sampling on CNNs' accuracy.
 - Revealed that CNNs have no human-like chart preference, cannot extrapolate, and are not robust to short bars.
- **Automatic vehicles detection and recognition system** | *Python, OpenCV, Keras* 01/2019 – 04/2019
 - Implemented a system utilizing convolutional neural networks to automatically detecting and re-identifying vehicles based on license plate, vehicle model, and color.