Zion Sheng

Address: 2728 Brown Ave #D, 27705, Durham NC, USA | Cell: 713-413-7498 | Email: zion.sheng@duke.edu
Website: https://zs144.github.io | GitHub: https://github.com/zs144

EDUCATION

UC Berkeley Extension — Remote
Advanced Bioscience Program | CGPA: 4/4 — July 2025 – Now
Courses: Introductory Biology & Lab. Coming: Cell & Molecular Biology, Genetics, Modern-Omics: Technologies & Applications

Duke University — Durham, USA
Master of Engineering in Electrical & Computer Engineering | CGPA: 3.56/4 | PGPA: 3.75/4 — Aug 2022 – May 2024
Note: I transferred from Software Engineering track (the 1st semester) back to AI/ML track.

Central University of Finance and Economics (CUFE) — Beijing, China
Bachelor of Economics | CGPA: 3.82/4 — Sep 2017 – Jun 2021

University of California, Los Angeles (UCLA) — Los Angeles, USA
Undergraduate Summer Session Program (Math, Statistics) | CGPA: 3.75/4 — 2019, 2020 Summer

RESEARCH INTERESTS

Computational Biomedicine, Spatial Transcriptomics, Multi-omics, Single-cell Omics, Multi-modal Learning, Graph Learning, Digital Pathology, Computer Vision, Medical Image Analysis, Uncertainty Analysis.

RESEARCH EXPERIENCE

Duke University, Zhang Lab — Durham, NC
Research Technician II, Advisor: Dr. Yi Zhang (Duke Neurosurgery) — May 2025 – now

- Developing a novel **graph-based** model for integrating **image-based** cell segmentation with **high-resolution spatial transcriptomics** to perform subcellular spot-level cell type inference.
- Built **SegOST**: an image-based cell segmentation and annotation pipeline that finetunes **StarDist** for various cell populations in **Visium HD** samples. Accurately segmented dorsal root ganglion (DRG) neurons and satellite glial cells from in-house data, which generates high-quality, spatially resolved **single-cell** gene expression profiles that elucidate the pain-related **neuro-immune microenvironment** in diabetic patients.
- Applied **MetaTiME** to annotate **cell states** on human and mouse breast cancer **scRNA** data that helps to examine the expression pattern of circadian genes in tumor hubs.
- Created several tools for **omics data visualization**, **subcellular-spot-to-cell aggregation**, spatial transcriptomics **synthetic data generation**. Maintained the lab Wiki and wrote numerous tutorials.

Duke University, Lafata Lab — Durham, NC
Part-time Research Assistant, Advisor: Dr. Kyle Lafata (Duke Radiation Oncology) — Jan 2024 – July 2025
Project Topic: Adapt **Vision-Language Model** to the diagnosis of head and neck squamous cell carcinoma (HNSCC)

- Preprocessed HNSCC whole slide images (WSI) from the TCIA-CPTAC database using **QuPath** and **OpenSlide**. Finetuned pathology **visual-language foundational models** (VLFM) to predict **tumor grade** via **multiple instance learning**.
- Conducted **interpretable model inference** on tumor grade prediction by demonstrating the correlation between patch-level attention scores and text-based conceptual features embedded by the VLFM text encoder, as well as morphological and graphical features calculated by **CellProfiler**.
- Collaborated with pathologists to validate the high-attentive regions identified by the model and evaluate how they align with the functional **tissue domains**.
- The first-author abstract was orally presented at AAPM 2025 Annual Meeting & Exhibition (**AAPM25**) [2].

University of Missouri, Malof Lab — Columbia, MO
Research Technician, Advisor: Dr. Jordan Malof (U of Missouri EECS) — May 2024 – May 2025
Project Topic: Investigate the need to use in-domain images for self-supervised pre-training in remote sensing

- Trained **self-supervised learning** (SSL) ViT models via **Masked Autoencoder** (MAE) on large-scale image datasets (ImageNet, SSL4EO, and self-curated GeoNet. Each with 1M+ images) via **multi-GPU parallel training**. Further finetuned the pretrained models for various downstream applications in remote sensing to test the model's transferability, covering classification, object detection, semantic segmentation, road delineation, etc.
- The paper has been published on ArXiv in 2025 [3]. The abstract that introduces our self-curated 1M+ satellite image database, **GeoNet**, has been presented at AGU 2024 Annual Meeting (**AGU24**) [4].

Duke University, Applied Machine Learning Lab — Durham, USA
Research Intern, Advisor: Dr. Leslie Collins (Duke ECE) — Aug 2023 – May 2024
Project Topic: Finetune **Segment Anything Model** (SAM) to segment clouds in multi-band HD satellite images

- Preprocessed satellite cloud images, created prompting boxes, and finetuned SAM on cloud images. Benchmarked the segmentation performance and compared finetuned SAM with a light-weight U-Net baseline trained from scratch.

Project Topic: Review traditional and modern methods in the algorithmic design of EEG-based **Brain Computer Interface** (BCI)

- Conducted a comprehensive literature review on the P300 brain-computer interface (BCI) research, and summarized a standard workflow for P300 spellers from **EEG** time series parsing to character prediction.
- Prototyped a CNN model to process P300 EEG data and classify target signals, which reaches on-par performance compared to the traditional non-DL method (SWLDA) in simulated online experiments.

SELECTED PUBLICATIONS AND PROJECTS

Publications and Presentations

1. Markakis PJ, Gowdy TM, Malof J, **Sheng Z**, Lancellotti B, Collins L, Davitt A, and Bradbury K. "Estimating Global, High-resolution Onsite Building Sector Greenhouse Gas Emissions.", NPJ Urban Sustainability 2025 (Under review).
2. **Sheng Z**, Li X, Russ MK, Lafata KJ, "Attention-Based Multiple Instance Learning of Head and Neck Cancer Grading on Digital Pathology Using Vision-Language Foundational Models.", AAPM 2025. (abstract, presentation)
3. Lahrichi S, **Sheng Z**, Xia S, Bradbury K, Malof J. Is self-supervised pre-training on satellite imagery better than Imagenet? A systematic study with Sentinel-2. arXiv preprint arXiv:2502.10669. 2025. (preprint)
4. Lahrichi S, **Sheng Z**, Xia S, Bradbury K, Malof J. GeoNet: A Global Dataset and Foundation Model for Deep Learning on Optical Satellite Imagery. AGU 2024. (poster)

Other Projects

Intra-retinal cyst segmentation in optical coherence tomography images via transfer learning (link)

- Applied different pre-trained models (Swin Transformer, DeepLabv3 Plus, and MobileNetv3) to perform **cyst segmentation** on retinal OCT images via transfer learning. Achieved a competitive performance compared with the other existing methods, with an average testing Dice score of **0.41** on Duke OCT dataset and **0.81** on UMN OCT dataset.

Comparative analysis of text summarization: A focused case study on CNN/DailyMail dataset (link)

- Used CNN/DailyMail as a benchmark data set, evaluated the **text summarization** performance of multiple text summarizers across extractive and abstractive regimes (TF-IDF, BERT, BART, and T5-small). Our experiment showed that the pre-tuned BART model achieved the highest ROUGE score among all candidates.

INDUSTRY EXPERIENCE

ClimateTrace — Durham, NC

Researcher, Advisor: Dr. Jordan Malof (U of Missouri EECS), Dr. Kyle Bradbury (Duke ECE) — May 2024 – May 2025

Project Topic: Develop the high spatial-temporal resolution GHG emission estimates for the global building sector

- Developed an **in-house Python visualization package** to visualize main data products in our high-resolution greenhouse gas emission inventory. The package has also proven helpful in data sanity checks and comparison with other inventories.
- Led the **uncertainty analysis** of our emission estimate. Designed and managed the uncertainty estimation data pipeline.
- Collaborated with the team to ship data products to the headquarters every month. Performed data quality check and located/fixed bugs in the data generation process.
- The teamwork has led to a paper submitted to **NPJ Urban Sustainability** in 2025 [1].

DiDi — Beijing, China

Data Analyst Intern, Data Operations Department — Dec 2020 – Apr 2021

- Built the automatic workflow that encapsulates data querying, key business indicators computing, and analysis result updating.
- Composed and delivered the daily market report in **GitBook** and the data dashboard in **Tableau**.
- Carried out **cluster analysis** based on the service data of users and drivers, and grouped them by their preferences for different services. The teamwork helped improve the effectiveness of coupon distribution/usage by **34%**.

TEACHING EXPERIENCE

Teaching Assistant, Duke University Graduate School

- ECE 580: Introduction to Machine Learning — Spring 2024
- DECISION 523: Fraud Analysis with R — Fall 2023
- QM 523: Fraud Analysis with R (online) — Fall 2023
- DECISION 563: Programming for Data Analysis with Python — Summer 2023

HONORS AND PRIZES

- Sigma Xi Associate Member — 2025
- AAPM 2025 Science Council Session Finalist Recognition — 2025

TECHNICAL SKILLS & ONLINE LEARNING

- Skills: ML/DL, Python, R, Unix, Docker, C/C++, Data Visualization.
- Online learning (Coursera): Probabilistic Machine Learning, Introductory Human Physiology, Deep Learning Specialization, Mathematics for Machine Learning, Build a Modern Computer from First Principles, Algorithms (Java).