Gavin Simmons

San Diego, CA | gasimmons@ucsd.edu | (949) 566-3107 | https://www.linkedin.com/in/gavinsimmons03 | https://github.com/gasimmons

SUMMARY

A soon-to-graduate aspiring **Software Engineer** with a solid background in backend **data management**, specializing in the extraction, storage, and optimization of large-scale data stores. Eager to contribute technical skills in **software development** and **deep data mining** to build scalable data architectures while possibly further developing expertise in **machine learning, cloud computing, and distributed systems.**

EDUCATION

University of California, San Diego, San Diego, CA — *June 2025*
Bachelor of Science in Computer Science - Bioinformatics — *GPA*: 3.76/4.0
Relevant Coursework: Advanced Data Structures; Algorithm Design, Machine Learning Algorithms

SKILLS

Languages: Python, C++, Java, JavaScript, C, SQL, HTML
Frameworks: Django, Node.js, React.js, PyTorch
Developer tools: Git, AWS, Jira, MySQL, Apollo GraphQL, MongoDB, Docker

WORK EXPERIENCE

UCSD Computer Science and Engineering Department, San Diego, CA
Undergraduate ML/CV Researcher — *September 2023 – Present*

- Implemented PyTorch-based image segmentation using UNet and YoloV8 models, trained on datasets with 80% training, 10% validation, and 10% testing sets. Enhanced each model's recall rate to over 85% through hyperparameter tuning and architecture optimization.
- Develops data analysis pipelines by integrating Python scripts for data manipulation and visualization using Seaborn, Matplotlib, and Pandas. Used Bash scripts for workflow automation and tools like SAM tools and Bedtools for read alignments and variant calling in sequencing projects. Automated complex analyses with Snakemake, reducing processing time by about 10 hours per genome and ensuring reproducible results.

Kargo, San Francisco, CA — *June 2024 – September 2024*
Backend Software Engineer Intern

- Developed a robust data parsing framework using Python to automatically process and upload customer documents (PDFs, CSVs, and Excels) to a MySQL database, reducing manual processing time and ensuring reliable data integration through an exception handling mechanism.
- Enhanced the Django admin site by integrating Google OAuth for secure user authentication, engineered custom tools for managing user permissions, and designed dynamic pages that display database information with links to detailed insights on the dashboard. Improved infrastructure by optimizing server configurations and deploying scalable AWS Cloud environment solutions, which enhanced web application performance and ensured consistent uptime
- Developed the GraphQL API in Node.js to enable secure, scalable data exchange between systems, utilizing both MySQL for relational data management and MongoDB for flexible, document-based storage. Implemented JWT for authentication and optimized database queries to enhance performance and security.

PROJECTS

- **IDEA (Interactive Differential Expression Analysis):** A python package for differential expression analysis of gene expression data, providing a user-friendly and efficient tool for researchers to identify differentially expressed genes between conditions or groups, with comparable functionality and performance up to 78% to the R package DeSeq2, leveraging popular libraries such as NumPy, Pandas, SciPy, and Statsmodels.
- **Tower Debugging Tool:** Developed a python-based tool that when called, retrieves a grid image video and a track file from AWS S3 buckets. Using OpenCV and FFmpeg for video reconstruction, inserting black frames for missing timestamps and drawing bounding boxes on the images. The reconstructed video is uploaded back to an S3 bucket for the team to use in debugging.
- **Avocado Tree Segmentation Model:** Developed an avocado tree semantic segmentation model using a U-Net architecture, achieving a IoU of .74. With a manually labelled dataset, the training pipeline included an early stopping and cosine annealing learning rate scheduler to prevent overfitting and optimize convergence. The production version is deployed on Google AI Platform to enable efficient batch processing of large image datasets for segmentation.