Steven Walton

Email: resume@swalton.ai
Website: swalton.ai | gScholar: he4JY7wAAAAJ | Linkedin: sjwalton | GitHub: stevenwalton

EDUCATION

Ph.D.	Computer Science	University of Oregon	June 2025
M.S.	Computer Science	University of Oregon	June 2023
B.S.	Space Physics	Embry-Riddle Aeronautical University	Dec 2014

EXPERIENCE

Shi Labs @ GaTech/UIUC/UO *Researcher* Sept 2020 - Current

- Researched and developed state-of-the-art (SOTA) Generative Models, including Generative Adversarial Networks (GANs), Vision Transformers (ViTs), Diffusion Models, and Normalizing Flows.
- Developed novel attention mechanisms to efficiently incorporate local and global information which led to reduced computational burdens while increasing performance.
- Led and collaborated on the development of state of the art Vision Transformers and generative models, leading to publications in top tier venues and workshops.
- Maintained and administered lab's computational infrastructure, including procurement.

University of Oregon *Graduate Researcher* Sept 2018 - June 2025

- Researched and developed neural architectures which could performance metrics *while* reducing the number of model parameters and training costs.
- Innovated new analysis techniques to understanding model behavior and training dynamics, which led to developing more robust and computationally efficient AI models.
- Researched High Performance Computing and *in situ* visualization methods for Computer Graphics.
- Developed software for the open source VTK-m project.
- Developed course work and taught undergraduate and graduate students for several classes, including machine learning.

NVIDIA *Metropolis Intern* Sept 2023 - March 2024

- Significantly improved the generalizability of ReIdentificationNet by implementing advanced training techniques and architectural modifications, significantly enhancing performance on customer data.
- Led analysis of Identification models to uncover low performance regions and utilized this information to vastly improve robustness for both small and large ReIdentification models.
- Conducted profile analysis of models and implemented optimizations to reduced computational requirements for customers, significantly increasing model throughput.
- Optimized deep learning models through pruning and quantization, improving efficiency and reducing inference latency while adapting them for seamless deployment with TensorRT compilation.
- Developed models for synthetic data generation, enhancing model robustness and generalization, through the use of both GANs and Diffusion Models.
- Developed tooling to improve team productivity when training models and onboarding to new machines, leveraging my expertise in Linux, Bash Scripting, and familiarity with HPC environments.

Picsart AI Research (PAIR) *Ph.D. Research Intern* June 2021 - Nov 2022

- Researched and developed methods for style based transfer of text, enhancing creativity for the application's commercial use.
- Developed advanced distillation techniques for optimizing model inference without sacrificing performance.
- Investigated the integration of generative models into existing software pipelines, extending the capabilities of Picsart's early AI-driven tools.
- Administered compute infrastructure by developing and implementing tooling such as SLURM scripts and establishing operational standards for ensuring reliability and efficient machine utilization.
- Collaborated in procurement discussions for new machine acquisitions, leveraging my HPC experience to access technical requirements for the business's newly founded AI division, directly engaging with sellers.

Lawrence Livermore Nat. Lab. (LLNL) *Comp. Sci. Intern* June 2020 - Sept 2020

- Developed machine learning software to analyze noisy X-ray images, contributing to projects focused on identifying geometry and material composition of varying objects.
- Collaborated closely with imaging scientists (customer), utilizing my physics background to enhance model accuracy, interpret complex data, and develop robust analysis frameworks.
- Engaged in cross-functional meetings and discussions, developing proper proxy experiments to ensure the proper mission objectives were achieved despite the sensitive nature of the project.

Lawrence Livermore Nat. Lab. (LLNL) *Comp. Sci. Intern* June 2019 - Sept 2019

- Investigated the integration of machine learning techniques into the ALPINE Ascent HPC software suite, optimizing in situ data processing for AI driven data interpolation.
- Researched solutions to overcome HPC data constraints within highly parallel supercomputing environments.
- Provided critical input in technical presentations, leading to subsequent research opportunities and funding within the lab.

Oak Ridge Nat. Lab. (ORNL) *ASTRO Intern* June 2018 - Aug 2018

- Integrated ADIOS2's data management framework into ALPINE Ascent HPC software, enabling efficient data handling in computational environments.
- Integrated streaming into ALPILE Ascent, enabling users to perform visualization and analysis tasks *in situ*, reducing locking operations in highly parallel environments and enabling higher machine utilization.
- Research and development led to 'Visualization as a Service' paradigm, allowing visualization and analysis to be performed in real time and off-node through arbitrary network connections (i.e. infiniband, ethernet, wireless, over LAN or WAN), extending the scope and flexibility of data processing.

Gloyer-Taylor Lab. LLC (GTL) *Engineer & Lead Scientist* July 2015 - May 2018

- Led and secured a NASA STTR Phase I contract (valued \$150k) that led to Phase II funding (valued \$750k), providing essential funding that led to the growth of the company.
- Developed cutting-edge radiation shielding methods capable of generating auxiliary power for satellite operations, minimizing mass expenditures and extending mission capabilities.
- Developed AI optimization algorithms leading to the innovation of novel material and geometric designs for energy producing radiation shielding.

- Spearheaded conversion of acoustic dynamic simulation models leading to over a 100x reduction in computation time, enabling more accurate and complex simulations to be run faster and reducing costs.
- Built and designed computational and physical testing frameworks, including HPC clusters, to facilitate computation critical for day-to-day operations.
- Built open source libraries (H5Easy) and documentation necessary for management of scientific data.

TECHNICAL SKILLS

- **Generative AI**: Diffusion Models, GANs, Normalizing Flows
- **Deep Learning**: ViTs, Transformer Models, Efficient Deep Learning, Neural Architecture Design, Distillation, Pruning, Quantization.
- **Programming**: Python, PyTorch, C, C++, TensorRT, Linux, Bash Scripting, SLURM, OpenMPI, OpenMP.
- **Engineering**: Computational Simulation, Soldering, EagleCAD, circuit design, mechanical engineering, basic lathe operation, milling, 3D Printing, FreeCAD, prototyping.

PUBLICATIONS

- **S. Walton** *Smaller, Faster, Cheaper: Architectural Designs for Efficient Machine Learning* Ph.D. Thesis 2025
- **S. Walton**, A. Hassani, X. Xu, Z. Wang, H. Shi *Efficient Image Generation with Variadic Attention Heads* (StyleNAT) eLVM @ CVPR 2025
- **S. Walton**, V. Klyukin, M. Artemev, D. Derkach, N. Orlov, H. Shi *Distilling Normalizing Flows* eLVM @ CVPR 2025
- A. Hassani, **S. Walton**, *et. al Generalized Neighborhood Attention: Multi-dimensional Sparse Attention at the Speed of Light*
- J Roberts, **S Walton**, *et. al ZeroBench: An Impossible Visual Benchmark for Contemporary Large Multimodal Models*
- N. Kennamer, **S Walton**, A. Ihler *Design Amortization for Bayesian Optimal Experimental Design* AAAI 2023
- A. Hassani, **S Walton**, J Li, S Li, H Shi *Neighborhood Attention Transformer* CVPR 2023
- **S Walton** *Isomorphism, Normalizing Flows, and Density Estimation: Preserving Relationships Between Data*
- J Jain, A Singh, N Orlov, Z Huang, J Li, **S Walton**, H Shi *Semask: Semantically Masked Transformers for Semantic Segmentation* NIVT @ ICCV 2023
- J Li, A Hassani, **S Walton**, H Shi *ConvMLP: Hierarchical Convolutional MLPs for Vision* WFM @ CVPR 2023
- **S Walton**, A Hassani, N Shah, A Abuduweili, J Li, H Shi *Escaping the Big Data Paradigm with Compact Transformers*
- D Pugmire, **S Walton**, *et. al Visualization as a Service for Scientific Data* SMC 2020
- **S Walton** *DATUM: Dotted Attention Temporal Upscaling Method*

Awards

- Outstanding Reviewer CVPR 2025

TEACHING

CS 445/545: Modeling and Simulation Winter 2025

- Helped students with mathematical modeling and simulating physics.
- Help students understand HPC and optimization techniques used in computational environments.

CS 451/551: Database Processing Fall 2024

- Helped students understand the nature of databases.
- Helped students with complex topics such as threading and deadlocking.
- Developed testing scripts, autograders, and helped develop new version of the course.

CS 472/572: Machine Learning Spring 2024

- Helped students understand Neural Networks and Machine Learning principles.
- Developed autograders and submission portals.

CS 472/572: Machine Learning Winter 2022 & Winter 2023

- Lectured and developed new course material, modernizing course offering.
- Taught students fundaments of Machine Learning and advanced topics such as GANs, Diffusion Models, LLMs/GPTs.
- Helped students develop projects that showcase their learning and to prototype products.

CS 414/514: Advanced Data Structors Winter 2021

- Helped students understand mathematics behind algorithms and data structures.

CS 314: Computer Organization Fall 2020

- Helped students understand machine level programming (assembly), optimization, memory management, and translation between C and x86-64 systems.
- Directed student labs and developed coursework

CS 322: Introduction to Software Engineering Fall 2018

- Taught students fudamentals of software engineering, API development, testing, database management, client/server communication, and version control systems (git).
- Directed labs, developed coursework and lecture materials.