

# Tony Tran

[tonyusername1@gmail.com](mailto:tonyusername1@gmail.com) • (346) 932-2540 • [tonyvt0.github.io/](https://github.com/tonyvt0) • [linkedin.com/in/thtran37](https://linkedin.com/in/thtran37) • [github.com/TonyVT0](https://github.com/TonyVT0)

## EDUCATION

### University of Houston, Cullen College of Engineering | Houston, TX

*M.S. in Engineering Data Science and AI, GPA: 3.67/4.00*

May 2026

**Thesis:** [ELASTIC: Efficient Once For All Iterative Search for Object Detection on Microcontrollers](#)

*B.S. in Computer Engineering Technology, GPA: 3.84/4.00*

May 2024

*Minor in Technology Leadership and Innovation Management*

Engineering in Training (FE/EIT) License in Electrical and Computer (Texas: 81224)

June 2024

## EXPERIENCE

### University of Houston | Houston, TX

*Graduate Research Assistant – [Network Autonomous and Intelligent Learning](#) (NAIL) Lab*

Dec. 2023 - Present

- Develop Efficient Neural Architecture Search Algorithms for Edge AI and Embedded Systems
- Reformulated Robot Angle Estimation as a Classification Task (**21 Bins**), Applied **4** Unique Domain Randomization Techniques, Trained Robust DNN (**88% Accuracy**) on **28k** Synthetic Training Images and Validated on **28k** Test Images
- Streamlined Model Development Pipelines through Prompt Engineering and Distributed Training on HPC Clusters and the Lambda Cloud, Reducing End-To-End Experimentation Time by **3x**

### Teaching Assistant

Aug. 2024 – Present

- Lead Weekly Lab Sessions/Office Hours for **30+** Students per Class, Responsible for Grading Homeworks, Tests, Quizzes
- Designed **Six** New Lab Assignments Ranging from Deep Learning Models in PyTorch, Model Compression Techniques (Pruning/Quantization), and CodeGen (Generating C/C++ Code for Edge Models)
- Taught AI in Robotics (SP2026), Embedded Systems (FA2024, SP2025), Sensor Applications (FA2024, SP2025), Introduction to C++ Programming (FA2024)

### Research Mentor, UH RE-PST Program

Summer 2025

- Mentored **Two** End-to-End Embedded Machine Learning Projects, Building Cats vs. Dogs Image Classification and Keyword-Spotting Audio Recognition Demos on **Two** Different Microcontrollers in **<6 Weeks**
- Designed an Intensive AI in Embedded Systems Hands-On Curriculum Containing (**15+** Labs) and Guiding **Two** Undergraduate Researchers with No Prior Experience

## SELECTED PROJECTS

### [ELASTIC](#) – Master’s Thesis, [Paper](#), [Code](#) | Python, PyTorch (+DDP, AMP), Embedded C Aug. 2024 - May 2025

- Designed Lightweight DNNs in under **32 KiB Memory**, Reducing Energy Consumption by **3.5x**, Requiring **1.6x** Less Power Usage, and Lowering Latency by **2.4x** utilizing Proposed Constrained Iterative OFA NAS Framework for Detection
- Deployed Tiny DNNs in under **1/4 MiB Flash** and Achieving **3ms** Inference Time on MAX78000/02 Platforms for Digit and Face Detection Demos
- Map Specialized Neural Network Architectures to Data Memory in Hardware-Accelerated CNN Inference Accelerator

### [Otto Sorting Bin](#) – Capstone Project, 1<sup>st</sup> Place, [Poster](#) | Jetson Nano, YOLO, OpenCV

Aug. 2023 - May 2024

- Finetuned YOLOv8 (**90%** Accuracy) Object Detection Model using TACO Detection Subset (**1.5k** Images) and **>500** Custom-Labeled Images for Robust Trash and Recyclable Detection Management on Nvidia Jetson Nano
- Engineered a Low-Cost T-Bot Sorting System and Constrained Object Priority Algorithm (**>85%** Physical Pickup Success) through Mechanical-Algorithmic Co-Design

### [TrashDet](#) – Accepted to WasteVision2026, [Paper](#) | Python (+OOP), PyTorch, YOLO

Nov. 2025 - Dec. 2025

- Achieved State-Of-The-Art Performance (**19.5 mAP**) on the TACO Dataset (**1.5k** Images), Outperforming Prior Waste Detectors with **2.8x** Less Model Size
- Derived Deployable TinyML Models down to **1.1M** Parameters, Maintaining Real-Time Performance at **37 FPS** for Real Time Waste Detection

## Traffic Severity Modeling | Python, Scikit-Learn, Matplotlib, Seaborn, Pandas, NumPy

**Aug. 2025 - Dec. 2025**

- Interpreted over **7.7 million** Traffic Records with over **40** Variables and Built Accident Severity Predictive Model (**85%** Accuracy) using Logistic Regression, Linear Discriminant Analysis, SVM, Naïve Bayes, Random Forest, and LightGBM
- Employed PCA and K-Means Clustering to Profile **Four** Data-Driven Crash Categories and Improve Interpretability

## ADMM Pruning – Code | Python (+OOP), PyTorch, Matplotlib, NumPy, Pandas, Excel

**Aug. 2025 - Dec. 2025**

- Cut Size of Mobile DNNs in **Half**, Retaining **97%** Predictive Performance, and Increasing Accuracy by **3.5x** Post Pruning
- Sped Up Model Convergence by **36%** and Discovered **One** Architecture Failure Mode under High Complexity

## 1-Lipchitz Layers Beyond Classification – Code | Python (+OOP), PyTorch, Matplotlib

**Aug. 2024 - Dec. 2024**

- Extended 1-Lipchitz Robustness Theory to Object Detection, Evaluating **Three** Layer Designs on Tinier-SSD Detectors using the SVHN Detection Dataset (**>40k Images**)
- Benchmarked Robustness, Achieving **10%** Higher Accuracy, and Retaining **85%** Predictive Performance (Compared to **53%** Against Baseline)
- Quantified Cost of Robustness (**6x** Training Time, **20x** Parameters), Informing Production-Scale Trade-Offs

## PUBLICATIONS

---

### Peer-Reviewed Publications

**[1]** T. Tran and B. Hu, “TrashDet: Iterative Neural Architecture Search for Efficient Waste Detection,” in Proceedings of the *2026 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)*, Tucson, AZ, USA, Mar. 2026.

**[2]** R.R. Suganda, T. Tran, M. Pan, L. Fan, Q. Lin, and B. Hu, “Distributed Perception Aware Safe Leader Follower System via Control Barriers Methods”, *2025 IEEE International Conference on Robotics and Automation (ICRA)*, Atlanta, GA, May 2025. (**Acceptance Rate: 38.67%**).

**[3]** T. Tran and B. Hu, “FACETS: Efficient Once-For-All Object Detection via Constrained Iterative Search”, *2025 IEEE International Conference on Robotics and Automation (ICRA) Late Breaking Session*, 2025.

### Publications Under Review / To Appear

**[4]** T. Tran, Q. Lin, and B. Hu, “ELASTIC: Efficient Once for All Iterative Search for Object Detection on Microcontrollers”, *Under review at IEEE Transactions on Computers – Briefs*, 2025.

## AWARDS AND HONORS

---

- Summa Cum Laude (Awarded by UH **2024**)
- University Honors (Awarded by The Honors College **2024**)
- Dean’s List (**8 Semesters 2020-2024**),
- **7x** Engineering Project Awards (**5x 1<sup>st</sup>** Place, **2x 2<sup>nd</sup>** Place)

## TECHNOLOGIES

---

- **Languages:** Python, C/C++ (Embedded), Java, HTML, MATLAB, Verilog
- **Machine/Deep Learning:** PyTorch, TensorFlow, Scikit-Learn, XGBoost, LightGBM
- **Computer Vision:** Classification, Object Detection, Segmentation; Ultralytics, Detectron2, OpenCV, Roboflow
- **EdgeAI/TinyML:** Neural Architecture Search, Once-For-All, Pruning, Quantization, ONNX, TFLite, TinyEngine
- **Data Libraries:** SciPy, NumPy, Pandas
- **Visualization:** Matplotlib, Seaborn, Tableau, Canva, Draw.io, Lucid Chart, PowerPoint
- **Hardware:** MAX78000/02 (ARM Cortex-M4F + AI CNN Hardware Accelerator), Jetson Nano (ARM Cortex-A57 + NVIDIA Maxwell GPU), TM4C123GXL (ARM Cortex-M4F), Arduino Uno (ATmega328P), STM32F746G (ARM Cortex-M7)
- **Tools:** Git, Linux, Bash
- **AI Tools:** ChatGPT, Gemini, DeepSeek, Cursor, Claude, NotebookLM