Orestis Zambounis

me@orestisz.com · Swiss Citizen

GitHub LinkedIn
Deep Learning, Computer Vision,
Robotics, Software Engineering

Experience

2023 - Present · **Senior ML Engineer / Tech Lead** · QSC (acq. by Acuity Brands) · Zurich, CH · Remote

- **Promoted to ML Tech Lead in Dez. 2024**, leading inference strategy, architecture, optimization and team of 3.
- Converted single-process architecture to a stage-parallel pipeline, **doubling** throughput and improving scalability.
- Ported vision ML models to TensorRT and Dali, **tripling** speed and reducing VRAM usage by **15%.**
- Increased system speed by **30%** on resource-constrained hardware through batched inference implementation.
- Led CV/ML prototyping in detection, tracking, embeddings, and VLMs with state-of-the-art methods.
- Technologies: **Python**, **PyTorch**, **TensorRT**, **ONNX**, **Weights & Biases**, **Grafana**, **ROS**, **Docker**, **GCP**.

2021 - 2023 (2 yrs) · **Machine Learning Engineer** · Seervision (ETHZ Spin-off, acq. by QSC) · Zurich, CH · Remote

- Optimized real-time detection pipeline, reduced latency by **24%**, VRAM by **45%**, and increased accuracy by **10%**.
- Designed, prototyped, tuned, and deployed a face recognition system with a false-positive rate below **5%**.
- Drove real-time inference optimization efforts, **tripling** the number of supported clients per hardware unit.
- Collaborated with the product team to prototype and experiment with CV/ML systems for novel user experiences.
- Technologies: **Python**, **C++**, **PyTorch**, **TensorFlow**, **OpenCV**, **CUDA**, **ROS**, **Docker**, **GitLab CI/CD**, **GCP**.

2020 - 2021 (6 mos) · **ML Infrastructure Engineer** · benshi.ai (funded by BMGF) · Barcelona, ES · Hybrid

- Built large-scale ML infrastructure and data pipelines for production, from ingestion to deployment.
- Technologies: **Python**, **Pandas**, **PySpark**, **Databricks**, **MLflow**, **Docker**, **Kubernetes**, **Azure**, **GitHub Actions**.

2019 - 2020 (1 yr 3 mos) · **Full-Stack Machine Learning Engineer** · Self-employed

- Developed a CNN-based face predictor with an **18%** accuracy improvement, optimized for low-latency inference.
- Developed full-stack application with cross-platform frontend and microservice-based cloud architecture.
- Technologies: **Python**, **TensorFlow**, **scikit-learn**, **Flask**, **React**, **PostgreSQL**, **AWS**.

2016 - 2017 (1 yr) · **Control Systems Engineer, Intern** · Rapyuta Robotics (ETHZ Spin-off) · Tokyo, JP · On-site

- Achieved a **55x speedup** of NumPy-heavy simulation iterations and open-sourced the Python package PyJet.
- Designed energy estimators using a Kalman Filter, enhanced tracking controller and performed sensor tests.
- Technologies: **Python**, **C++**, **NumPy**, **SciPy**, **ROS**.

Education

2017 - 2019 (2 yrs) · **MSc Robotics, Systems & Control** · 5.25/6.0 · ETH · Zurich, CH

- Developed an online deep learning architecture for object instance prediction, pose estimation, and tracking.
- Showed that an additional depth input channel improved the segmentation accuracy of Mask R-CNN by **31%**.
- Technologies: **Python**, **CUDA C/C++**, **TensorFlow**, **Keras**, **Caffe2**, **OpenCV**.

2012 - 2016 (3 yrs 6 mos) · **BSc Mechanical Engineering** · 5.51/6.0 · ETH · Zurich, CH

- Developed balancing algorithms for a 6DoF omnicopter using non-linear control methods.
- Technologies: **C++**, **MATLAB**, **Simulink**.

Projects

2023 - Present · **Beachin' Rentals** Self-service kiosk built using **Flask**, **Stripe**, **Shopify**, **RasPi**, **RS-485**.

2023 - 2024 · **Trap the Cat** Mobile app built with **JavaScript**, **CapacitorJS** and **Firebase**, with **100k+** downloads.

2022 · **Machine Dreams** Experimental fusion of **AI** and digital art using **GANs**, creating surreal NFT artworks.

2022 · **Anti CryptoPunks** NFT project built on the **Polygon** blockchain, with **6 ETH** traded.

2015 · **PyJet** Python library converting **Python/NumPy** operations to **C++**, achieving a **55x** speedup.