Orestis Zambounis

Swiss Citizen / Resident | info@orestis.ch | +41786373591

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

Experience

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

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

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

Full-Stack Machine Learning Engineer · Self-employed   (Feb 2019 - Mar 2020)
- 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**.

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

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

BSc Mechanical Engineering · 5.51/6.0 · ETH · Zurich, CH   (2012 - 2016)
- Developed balancing algorithms for a 6DoF omnicopter using non-linear control methods.
- Technologies: **C++**, **MATLAB**, **Simulink**.

Projects

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

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

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.   (2022)

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