Orestis Zambounis

me@orestisz.com · Swiss Citizen

github.com/orestis-z, linkedin.com/in/orestis-z

Deep Learning, Computer Vision, Robotics,
Systems & Control, Distributed Systems

Experience

2023 - Present (1 yr 6 mos) · **Senior Machine Learning Engineer** · (QSC, acq. by Acuity Brands) · Zurich, CH · Remote

- Ported vision ML models to ONNX and TensorRT, **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.
- Redesigned ML architecture for modularity and flexibility, and led efforts to clean up technical debt.
- Co-managed and mentored the ML team, integrated teams, enforced best practices, and led the hiring process.
- 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) · **MLOps Engineer** · benshi.ai (funded by BMGF) · Barcelona, ES · Hybrid

- Built and maintained scalable data pipelines for ML models in production, from data 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** · Freelancer · Remote

- 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

2018 - 2019 (6 mos) · **Imperial College London** · Master's Thesis · London, UK

- Developed an online multi-task deep learning architecture for object instance prediction, pose estimation, and multi-person tracking.
- Trained the Siamese network for visual cue matching on MOT dataset using Mask R-CNN outputs.
- Technologies: **Python**, **CUDA C/C++**, **Caffe2**.

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

- Showed that an additional depth input channel improved the segmentation accuracy of Mask R-CNN by **31%**; submitted paper to CoRL.
- Designed a time-efficient training strategy using data augmentation, synthetic RGB-D and real-world data.
- Technologies: **Python**, **TensorFlow**, **Keras**, **OpenCV**.

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

- Graduated **top 5%** of the class.
- Developed balancing algorithms for a 6DoF omnicopter using non-linear control methods.
- Technologies: **C++**, **MATLAB**, **Simulink**.