



Ties Robroek

PERSONAL DETAILS

<i>Birthdate</i>	January 4, 1995
<i>Nationality</i>	Dutch
<i>Address</i>	Gearhalsvej 6, 4. tv, 2500 Valby
<i>Phone</i>	+45 93 83 42 37
<i>Mail</i>	ties@trobroek.com
<i>LinkedIn</i>	https://www.linkedin.com/in/ties-robroek-0b5683b9/
<i>GitHub</i>	https://github.com/Sipondo

EDUCATION

PhD — Resource-Aware Data Science

2021-2024

Supervisor: Pınar Tözün

IT University of Copenhagen (Denmark)

Resource-Aware Data Science. Looking for ways to reduce resource consumption and hardware requirements of Deep Learning training workloads. Initial focus on benchmarking and understanding the training of a variety of Deep Learning models. Currently investigating ways of accelerating convergence for Time-To-Accuracy (TTA) focused on training workloads.

<https://en.itu.dk/-/media/EN/Research/PhD-Programme/PhD-defences/2024/PhD-thesis-Temporary-Version-TR-pdf.pdf>

MSc. Data Science

2017-2020

Supervisor: Tom Claassen, Fabian Gieseke

Radboud University Nijmegen (the Netherlands)

Specialisation in Computing Science. Focus on GPU computing, Machine Learning and Neural Networks. My thesis involves accelerating Nearest Neighbour Field search with GPUs. The entire Machine Learning pipeline has been addressed and sped up considerably, with two different techniques having been explored and implemented. Final implementations are capable of processing huge images accurately in short timespans, over 100x faster than the original processor code.

BSc. Computing Science

2013-2017

Supervisor: Tom Heskes, Fabian Gieseke

Radboud University Nijmegen (the Netherlands)

Includes a Bachelor Honours Degree & Thesis (2014-2016). For my thesis I sped up the Machine Learning model BCCD by translating it from MATLAB to C accelerated by CUDA. The end product achieved a massive acceleration compared to the original MATLAB implementation on large datasets.

PUBLICATIONS

TensorSocket: Shared Data Loading for Deep Learning Training

TBD

Ties Robroek, Neil Kimn Nielsen, Pınar Tözün

Under review

Training multiple models on the same dataset leads to data loading redundancies. TensorSocket unifies data loading when training multiple models on the same dataset, reducing CPU load, disk load, and memory movements. Furthermore, TensorSocket operates virtually overhead-free, allows

for varying batch sizes between models, and requires little to no code modifications. Available over at: <https://github.com/Resource-Aware-Data-systems-RAD/tensorsocket>

Modyn: Data-Centric Machine Learning Pipeline Orchestration 2025
Maximillian Böther, Ties Robroek, Viktor Gsteiger, Robin Holzinger, Xianzhe Ma, Pınar Tözün, Ana Klimovic
SIGMOD 2025

Modyn is a data-centric end-to-end machine learning platform continuously training models on dynamic datasets. As new data comes in, Modyn's triggering policies and data selection policies offer fine-grained control over when and on what data to retrain. The system is actively being developed and available over at: <https://github.com/Resource-Aware-Data-systems-RAD/radt>

Towards A Modular End-To-End Machine Learning Benchmarking Framework 2025
Robert Bayer, Ties Robroek, Pınar Tözün
TDIS (Collocated with EuroSys)

We introduce a modular end-to-end ML benchmarking framework designed to address these gaps. Our framework emphasizes modularity and reusability by enabling reusable pipeline stages, facilitating flexible benchmarking across diverse ML workflows.

An Analysis of Collocation on GPUs for Deep Learning Training 2024
Ties Robroek, Aaron Duane, Ehsan Yousefzadeh-Asl-Miandoab, Pınar Tözün
EuroMLSys 2024 (Collocated with EuroSys)

GPU collocation potentially increases GPU utilisation by training multiple models at the same time on the same GPU. We investigate the three ways of collocating model training on GPUs (naïve, MPS, MIG) with a series of experiments and formulate actionable user guidelines for efficient training using GPU collocation.

Data Management and Visualization for Benchmarking Deep Learning Training Systems 2023
Ties Robroek, Aaron Duane, Ehsan Yousefzadeh-Asl-Miandoab, Pınar Tözün
DEEM 2023 (Collocated with SIGMOD)
Best Presentation Award

Resource consumption of deep learning training is growing rapidly and we should investigate and report this consumption. We present an easily extendable framework that allows for plug-and-play minimal-code tracking of resource consumption of deep learning training, including visualisation. The system is actively being developed and available over at: <https://github.com/eth-easl/modyn>

Profiling and Monitoring Deep Learning Training Tasks 2023
Ehsan Yousefzadeh-Asl-Miandoab, Ties Robroek, Pınar Tözün
EuroMLSys 2023 (Collocated with EuroSys)

We dive into popular tools for profiling and monitoring deep learning training and find that their information is descriptive but the selection of metrics requires care. Monitoring tools exhibit negligible impact on training performance and profiling tools offer more detailed insights.

Approximate Nearest-Neighbour Fields via Massively-Parallel Propagation-Assisted K-D Trees 2020
Cosmin Eugen Oancea, Ties Robroek, and Fabian Gieseke
IEEE International Conference on Big Data 2020

A novel parallel implementation for one of the state-of-the-art methods for the computation of nearest neighbour fields that yields valuable computational savings over corresponding multi-core implementations.

WORK EXPERIENCE

Data Scientist, Team Lead Data Science

2020-2021

Hobbi, Full-time

Choosing, deploying and maintaining a visualisation environment for the entire company. Running small-scale machine learning for customer segmentation and advanced business intelligence. Developing, maintaining and deploying a Chrome Extension for customer service agents.

Data Science Student Assistant

2019

A.P. Møller-Mærsk (Denmark), Part-time

Developing a visualisation environment for multiple forecasting products. Working in a Data Science team specialised in forecasting supply and demand, it was my responsibility to design and test a powerful interface to provide insight to the researchers and the customers regarding the efficacy of the forecasting.

Data Science Intern

2018-2019

Forecast (Denmark), Full-time

Developing and implementing a Machine Learning system that combines supervised and unsupervised learning to recommend labels for tasks using word embeddings and NLP. I introduced biweekly presentations where the AI team shows their progress to the rest of the developers, improving company perception of AI.

Teaching Assistant New Devices Lab

2015-2018

Radboud University Nijmegen (the Netherlands), Part-time

Intensive class where students study microcontrollers and modern technologies. My tasks included running the lab for 8 hours per week and helping students with their projects.

Teaching Assistant Languages and Automata

2014-2018

Radboud University Nijmegen (the Netherlands), Part-time

Giving a weekly class on formal grammars, languages and automata. Included teaching, grading homework, and explaining homework assignments. I was responsible for two groups instead of one from 2015 to 2017.

Teaching Assistant Mathematical Structures

2015-2018

Radboud University Nijmegen (the Netherlands), Part-time

Giving an intensive weekly maths class for Computing Science students. Included teaching, grading homework, and explaining homework assignments.