



# AI Webinar Series on Deep Learning for CINI AIIS Labs

June 29th/July 3rd 2020

## Abstract

The goal of this webinar series is to explore the fundamentals of deep learning by building and training neural networks, optimizing data loading and performance through mixed-precision and parallelization, and deploying your trained model in production for inference. You will learn how to design, train, optimize, profile and deploy a deep neural network using NVIDIA technologies. Each session includes a Q/A part for interactions with attendees.

## Target

- PhD students, postdoc, researchers working with DNN and members of the CINI consortium.

## How to follow the webinars

- [NVIDIA Webex](#)

## Webinar Material

- [Shared Folder](#)

## Agenda

### Monday June 29th (from 9.00 to 10.20 CEST)

- **Session 0: Linear Regression in Pytorch** (20' + 20' q/a) - *Christian Hundt*
- **Session 1: Convolutional Neural Networks** (20' + 20' q/a) - *Christian Hundt*

### Wednesday July 1st (from 9.00 to 10.20 CEST)

- **Session 2: Efficient Data Loading using DALI** (20' + 20' q/a) - *Giuseppe Fiameni*
- **Session 3: Mixed Precision Training using Apex** (20' + 20' q/a) - *Paul Graham*

### Thursday July 2nd (from 9.00 to 11.00 CEST)

- **Session 4: Multi-GPU Training using Horovod** (20' + 20' q/a) - *Gunter Roeth*
- **Session 5: Deploying Models with TensorRT** (20' + 20' q/a) - *Niki Loppi*
- **Session 6: Profiling with NVTX** (20' + 20' q/a) - *Giuseppe Fiameni*

## Useful links

- **APEX**: <https://github.com/NVIDIA/apex> (Automatic Mixed Precision)
- **DALI**: <https://github.com/NVIDIA/DALI> (Data Loading Library)
- **HOROVOD**: <https://github.com/horovod/horovod> (Parallelization)
- **NVTX**:  
[https://docs.nvidia.com/gameworks/content/gameworkslibrary/nvtx/nvidia\\_tools\\_extension\\_library\\_nvtx.htm](https://docs.nvidia.com/gameworks/content/gameworkslibrary/nvtx/nvidia_tools_extension_library_nvtx.htm)
- **Deep Learning Examples**: <https://github.com/NVIDIA/DeepLearningExamples>

## Presenters



**Dr. Christian Hundt** has received his diploma degree in theoretical physics at the Johannes Gutenberg University (JGU) in Mainz, Germany. In his PhD thesis he investigated the parallelization of time series data mining algorithms on massively parallel architectures. As postdoctoral researcher at the Parallel and Distributed Architectures group a, he focused on the efficient parallelization of various biomedical applications such as context-aware metagenomic classification, gene set enrichment analysis and deep semantic image segmentation of thorax MRIs. In his current position as Deep Learning Solution Architect he coordinates the technical aspects of the NVIDIA AI Technology Center collaboration in Luxembourg as resident scientist.



**Dr. Niki Loppi** is a Deep Learning Solution Architect responsible for the NVAITC collaboration in Finland. Prior to joining NVIDIA, Niki worked as a researcher in the Department of Aeronautics at Imperial College London where he also obtained his PhD degree in Computational Fluid Dynamics. Specifically, his research focused on the development of high-order accurate numerical methods for solving industrial-scale fluid flow problems using modern GPU architectures.



**Gunter Roeth** joined NVIDIA as a Solution Architect in October 2014 having previously worked at Cray, HP, Sun Microsystems and most recently BULL. He has a Master in geophysics from the Institut de Physique du Globe (IPG) in Paris and has completed a PhD in seismology on the use of neural networks (artificial intelligence) for interpreting geophysical data.



**Paul Graham** joined NVIDIA in Dec 2018 as a Senior Solutions Architect, where he has responsibility for supporting customers and partners in delivering accelerated solutions to the Higher Education, High Performance Computing and AI communities in the UK. Previously he spent 20 years working at EPCC, the supercomputing centre at the University of Edinburgh, where he worked on a broad range of projects porting and optimising code. He also was the coordinator of SHAPE (SME HPC Adoption Programme in Europe), supported as part of the PRACE collaboration of supercomputing centres across Europe. At NVIDIA Paul is an advocate for using accelerated computing in HPC, and the emerging use of AI as a powerful new tool for researchers.



**Dr. Giuseppe Fiameni**, diploma in Computer Science and Ph.D. in Computer Engineering from the University of Modena and Reggio Emilia. He has been working in the field of HPC for more than 10 years as team leader at Cineca, the largest supercomputing center in Italy. Over the years he has been involved in several R&D projects for the design and implementation of data intensive applications employing machine and deep learning methods. He has represented Italy in differ EU boards to help define the European strategy in terms of digital infrastructure for the public and private research.

## NVIDIA Artificial Intelligence Technology Centre (NVAITC) in a nutshell

The NVIDIA AI Technology Centre (NVAITC) is a program to enable and accelerate AI research projects in Italy, focusing expertise and resources on specific research projects. The program is a national-level collaboration centred around project-based collaborations with institutions within the CINI network and fosters the collaboration of the Italian Computing Facility, CINECA. It aims at enabling academic institutions at all levels to conduct their research more efficiently by collaborating into research projects, training students, nurturing startups and spreading adoption of the latest AI technology throughout Italy. Example areas of contribution include

- DL/ML frameworks adoption
- Software development and environments set-up
- Performance optimization and tuning
- Support for efficient data loading, mixed-precision training, inference
- Multi-GPU Scaling
- Research participation
- Dissemination of results

In order to participate to the program and receive support, interested PIs can submit a proposal using this template [1] via email to [segreteria.aiis@consorzio-cini.it](mailto:segreteria.aiis@consorzio-cini.it). The local lead engineer can be contacted for any input, suggestion, or advice before submitting. PI is typically contacted back for clarification and SoW settlement after submission. Proposals are reviewed with the help of fellow NVAITC engineers on a first-come first-serve basis. Review takes a couple of weeks.

Evaluation is based on NVAITC criteria (target publication, technology stack and computing scale), rules of engagement (compact timeline, no compute, no funding, etc) and shared realistic expectations (an agreed-upon SoW). This '**call for proposals**' remains open as long as the program has the capacity to handle projects. Please refer to this page for any further information: <https://www.consorzio-cini.it/index.php/en/labaiis-home/labaiis-nvaitc>.