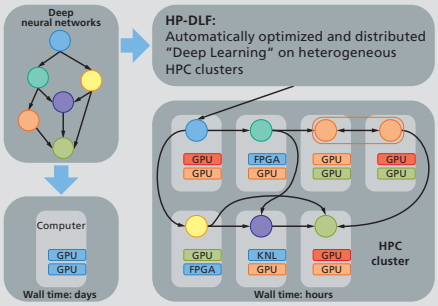
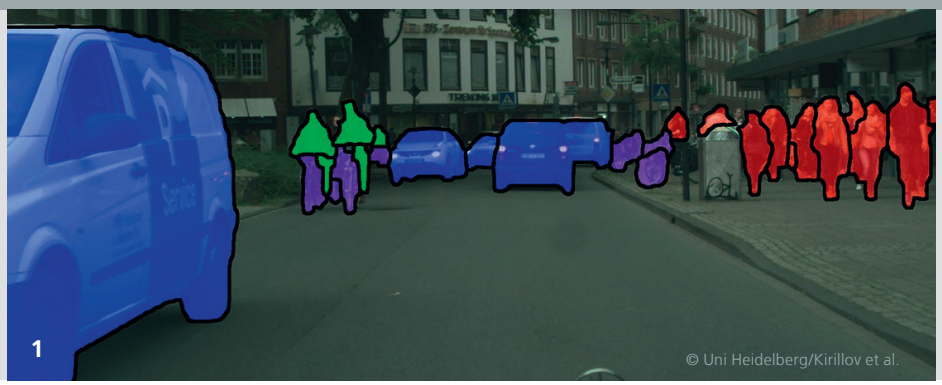
Fraunhofer
ITWM

FRAUNHOFER INSTITUTE FOR INDUSTRIAL MATHEMATICS ITWM





1 *Deep Learning without memory limits*

HP-DLF – HIGH PERFORMANCE DEEP LEARNING FRAMEWORK

The ambition of the HP-DLF crew is to provide a softwareframework that removes the memory limits during the training of very large deep neural networks.

Based on the developing ONNX standard for the description of DNN´s, we will support multiple parallelization layers within one framework

- Model Parallelism
- Batch Parallelism
- Hyperparameter Selection

HP-DLF will provide multiple optimization strategies and support hybrid hardware systems (CPU, GPU, FPGA).

The HP-DLF Framework builds on top of Fraunhofer´s GPI-Space Technology for Memory Driven Computing.

Based on the developing ONNX standard for the description of DNN´s the HP-DLF has the high speed distributed application independent memory layer keeps training data and manages data transport between nodes.

Fraunhofer-Institut für Techno- und Wirtschaftsmathematik ITWM

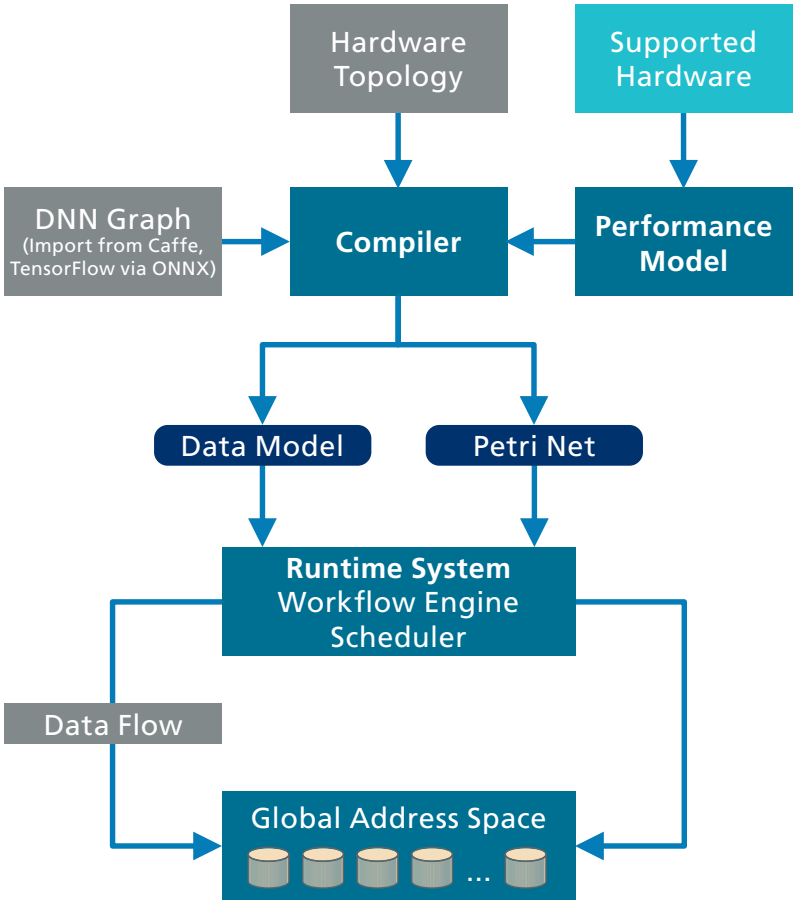
Fraunhofer-Platz 1
67663 Kaiserslautern
Germany

Contact
Dr.-Ing. Janis Keuper
Phone +49 631 31600-47 15
janis.keuper@itwm.fraunhofer.de

Dr. Peter Labus
Phone +49 631 31600-49 82
peter.labus@itwm.fraunhofer.de

www.itwm.fraunhofer.de

hpc_flyer_HP-DLF_EN



The HP-DLF project diagramm

In order to train a neural network the user has to provide an ONNX file – the topology of the DNN – as Input. By using the open standard ONNX, HP-DLF can serve as a HPC-back-end for all major deep learning frameworks.

With this information, the compiler generates a Petri net which encodes the execution of the optimization loop in a task and data parallel fashion. The parallelism is automatically extracted and task are scheduled to free resources. This enables the distribution on the level of single layers. HP-DLF may be used to train very large neural networks without memory restrictions.

Examples of such networks are often found in life and physical science, e. g. a 3D U-Net used for semantic segmentation of medical images.

SPONSORED BY THE

Federal Ministry of Education and Research