

Course Description

Implement neural networks on cloud and edge platforms using the Vitis™ AI development platform.

The emphasis of this course is on:

- Illustrating the Vitis AI tool flow, including optimization and compilation
- Exploring the architectural features of the Deep Learning Processor Unit (DPU)
- Utilizing the Vitis AI Library to optimize pre-processing and post-processing functions
- Creating a custom platform and application
- Deploying a design

What's New for 3.5

- Focus on TensorFlow 2, PyTorch, and some ONNX
- All labs have been updated to the latest software versions

Level – AI 3

Course Details

- 2 days ILT

Course Part Number – AI-INFER

Who Should Attend? – Software and hardware developers, AI/ML engineers, data scientists, and anyone who needs to accelerate their software applications using AMD devices

Prerequisites

- Basic knowledge of machine learning concepts
 - Neural Networks Explained - Machine Learning Tutorial for Beginners: www.youtube.com/watch?v=GvQwE2OhL8I
 - How Convolutional Neural Networks Work: www.youtube.com/watch?v=FmpDlaiMieA
- Deep learning frameworks (TensorFlow and PyTorch)
- Comfort with the C/C++/Python/make programming languages
- Software development flow

Software Tools

- Vitis AI development environment 3.5
- Vivado™ Design Suite 2023.1

Hardware (Optional)

- Alveo™ accelerator cards and adaptive SoCs
- Zynq™ UltraScale+™ MPSoC ZCU104 board*

* Check with your local Authorized Training Provider for the specifics of the in-class lab board or other customizations.

After completing this comprehensive training, you will have the necessary skills to:

- Describe machine learning solutions from the perspective of the Vitis AI development tools
- Enumerate the supported frameworks and models for cloud and edge applications
- Implement neural networks on cloud and edge platforms using the Vitis AI development platform
- Describe the proper Vitis AI tool flow
- Optimize a DPU for edge applications, leveraging the device architecture
- Enumerate the APIs included with the AMD AI Library

- Create a custom hardware overlay based on application requirements
- Create a custom application using a custom hardware overlay and deploy the design

Course Outline

Day 1

Vitis AI Environment Overview

- **Introduction to the Vitis AI Development Environment**
Describes the Vitis AI development environment, which consists of the Vitis AI development kit, for AI inference on AMD adaptive computing hardware platforms, including both edge devices and Alveo accelerator cards. {Lecture}
- **Frameworks Supported by the Vitis AI Development Environment**
Discusses the support for many common machine learning frameworks, including TensorFlow and PyTorch. {Lecture}

ML Concepts

- **Overview of ML Concepts**
Overview of ML concepts such as DNN algorithms, models, inference and training, and frameworks. {Lecture}

Vitis AI Environment Toolchain

- **AI Optimizer**
Describes the optimization of a trained model that can prune a model up to 90%.
This topic is for advanced users and will be covered in detail in the Advanced ML training course. {Lecture}
- **AI Quantizer and AI Compiler**
Describes the AI quantizer, which supports model quantization, calibration, and fine tuning. Also describes the AI compiler tool flow.
With these tools, deep learning algorithms can deploy in the Deep Learning Processor Unit (DPU), which is an efficient hardware platform running on an AMD FPGA or SoC. {Lecture, Lab}

Profiler

- **AI Profiler**
Describes the AI profiler, which provides layer-by-layer analysis to help with bottlenecks. Also covers debugging the DPU running result. {Lecture}

Deep Learning Processor Unit (DPU)

- **Introduction to the Deep Learning Processor Unit (DPU)**
Describes the Deep Learning Processor Unit (DPU) and its variants for edge and cloud applications. {Lecture}
- **DPUCZDX8G Architecture Overview**
Overview of the DPUCZDX8G architecture, supported CNN operations, DPU data flow, and design considerations. {Lecture}

Day 2

AI Libraries

- **Vitis AI Library**
Reviews the Vitis AI Library, which is a set of high-level libraries and APIs built for efficient AI inference with the DPU. It provides an easy-to-use and unified interface for encapsulating many efficient and high-quality neural networks. {Lecture, Lab}
Note that the edge flow version of the lab is not available in the On-Demand curriculum because an evaluation board is required for the entirety of the lab.

Custom Hardware and Application Development

- **DPU Edge Hardware Platform Creation Using the Vivado Design Suite**
Illustrates the steps to build a Vivado Design Suite project, add the DPUCZDX8G IP, and run the design on a target board. {Lab}
- **DPU Edge Kernel Creation Using the Vitis Environment Flow**
Illustrates the steps to build a Vitis unified software platform project that adds the DPU as the kernel (hardware accelerator) and to run the design on a target board. {Lab}
- **Creating a Vitis Embedded Acceleration Platform (Edge)**
Describes the Vitis embedded acceleration platform, which provides product developers an environment for creating embedded software and accelerated applications on heterogeneous platforms based on FPGAs, Zynq SoCs, and Alveo data center cards. {Lecture}
- **Custom Edge DPU Application Creation**
Illustrates the steps to create a custom application, including building the hardware and Linux image, optimizing the trained model, and using the optimized model to accelerate a design. {Lab}