



第6章

学習済みモデル、モデル圧縮、量子化ツール、ソフトウェア・スタックから専用のIPコアまでオールインワン

ロボットにAI搭載

深層学習推論処理向けエッジFPGA
開発プラットフォーム **Vitis AI**

イントロダクション

1

2

3

4

5

6

本章では、FPGAにおける深層学習推論処理を高速化するための開発環境「Vitis AI(Xilinx)」を紹介합니다。

FPGAは深層学習の推論環境として注目されてきましたが、開発コストが高いという問題がありました。図1に示すVitis AIはさまざまなタスクに対応可能な各種学習済みモデルや、モデルの圧縮、量子化ツール、こうした環境を効率的に活用するためのソフトウェア・スタック、深層学習推論処理向けのIPコアから構成されています。そのため、FPGA開発経験の少ないユーザでも容易に利用可能なものとしてまとまっています。

本稿では、このVitis AIについて概説するとともに、写真1に示す高性能な小型FPGAボードであるUltra96-V2(アヴネット)と組み合わせ、実際に深層学習推論処理を行う方法を紹介합니다。本ボードは自律移動ロボットのように、多数のセンサ、アクチュエータの統合と、処理のリアルタイム性が求められる用途などにも適しています。

Vitis AIの特徴

- Vitis AIはFPGA上における深層学習推論処理向け開発プラットフォーム

XilinxのFPGAはエッジ・デバイスに搭載されるも

ユーザ・アプリケーション

深層学習フレームワーク	Caffe	PyTorch	TensorFlow
Vitis AIモデル	Model Zoo	カスタム・モデル	
Vitis AI開発キット	AI Compiler AI Quantizer AI Optimizer		
	AI Profiler AIライブラリ		
	Xilinx Runtimeライブラリ(XRT)		
Overlay	Deep Learning Processing Unit(DPU)		

図1 Vitis AIスタックは著名な深層学習フレームの利用を前提にした変換ツールやソフトウェア・スタック、FPGA上で動作するアクセラレータIPコアから構成されている[文献(1)]

本章ではVitis AIを紹介後、写真1に示すFPGAボードを使って深層学習推論処理を行う

のとデータ・センタ向けのアクセラレータ・カードへ搭載されるものとに大別されます。Vitis AIはこの両方をサポートしています。ベースとなる複雑なFPGAハードウェアを抽象化することにより、FPGAの開発経験が少ないユーザでも容易に深層学習推論アプリケーションを開発できるようになっています。

Vitis AIは2020年1月にバージョン1.0がリリースされ、2020年10月現在の最新バージョンは1.2.1となっており、継続的に更新されています。ここでは最新バージョンを基準に解説しますが、実機動作を確認するときは、環境構築の容易なバージョン1.1を利用しています。なお、Vitis AIには公式のオンライン・ドキュメントが存在します。より詳細な情報については文献(1)に記述されています。

■ **本開発プラットフォームの構成**

図1に示したのはVitis AIの全体構成です。CaffeやPyTorch、TensorFlowといった著名な深層学習フレームワークに対応しています。これらを利用して構築した学習済み深層学習モデルや、モデルを変換しFPGA上で利用可能にするための各種ツールも備えています。また、アプリケーション開発をサポートするツールやソフトウェア・スタック、深層学習推論処理向けに特化したソフトIPコアであるDPU(Deep-Learning Processor Unit)が用意されています。以降

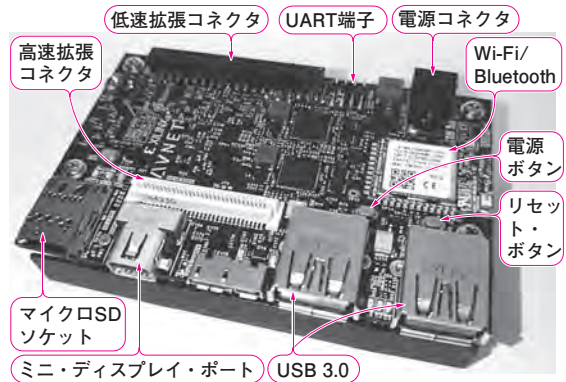


写真1 本章で使用するFPGAボードUltra96-V2(アヴネット)プロセッシング・システムとプログラマブル・ロジックをSoCで統合したデバイスZynq UltraScale+ MPSoC(Xilinx)を搭載している。ラズベリー・パイと同程度のサイズ(85×54mm)にまとまっている