

# FPGA 人工知能の ポテンシャルを探る

第2回 FPGA向けニューラル・ネットワークBNNの動作をまずはPCで確かめる

ご購入はこちら

鈴木 量三朗

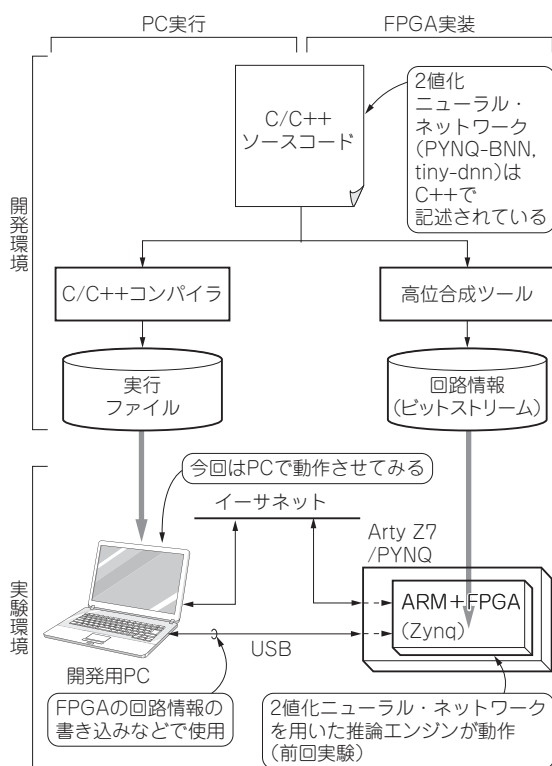


図1 FPGAで有効といわれる2値化ニューラル・ネットワークをPCで動かして動作メカニズムを確かめる

今回はPCで動作させてみることでニューラル・ネットワークを2値化する意味を考える

本連載ではFPGAで威力を発揮すると言われているニューラル・ネットワークBNN (Binarized Neural Networks) のポテンシャルを探ります。

今回は、FPGAベースのニューラル・ネットワークを扱う前段階として、C++で記述されたソースコードをPCで動かし、ニューラル・ネットワークを2値化する意味合いや動作メカニズムについて考えます(図1)。

## FPGA向けニューラル・ネットワークBNNのC++ソースコード

オーソドックスなニューラル・ネットワークの実装では計算に小数を使います。精度が重要視されるので、倍精度の浮動小数点数が使われることがあります。

逆に、精度が要求されないアプリケーションもあります。その場合は0と1、つまり2値でネットワークを組むことができます。並列性の高い処理が要求される分野ですので、2値化のネットワークを柔軟に組むことができるFPGAが研究のプラットフォームとして注目されています。

ここでは実際にC++で記述されたニューラル・ネットワークのソースコードを見てみます。

### ● 背景…最近のFPGAはC/C++でも記述できる

FPGA上のアプリケーションにおいて、C/C++などの今までソフトウェア開発で慣れ親しんできた言語を使って開発することが実用になってきました。

C++で書かれたソースコードを高位合成ツール (Vivado-HLSなど) と呼ばれる一種のコンパイラによって、FPGAに読み込ませる回路情報 (ビットストリーム) へと変換します。

使用時にはARMプロセッサからメモリのように見えるレジスタ (メモリ・マップドI/O) へパラメータを書き込むことで動作させます。

### ● コンパクトで可読性が高いtiny-dnnがベース

第1回で動作させたPYNQ-BNN<sup>(2)</sup>は、C++で書かれたtiny-dnn<sup>(3)</sup>がベースになっています。

tiny-dnnの特徴はコンパクトであり、抽象度の高いC++でソースコードの可読性が高くなるように設計されており、学習・推論の両方に使うことができます。