

C言語で高性能人工知能回路づくりに挑戦!

ピッタリ! ディープ・ニューラル・ネットワーク×FPGA入門

第1回 FPGAがディープ・ニューラル・ネットワークに向く理由 中原 啓貴

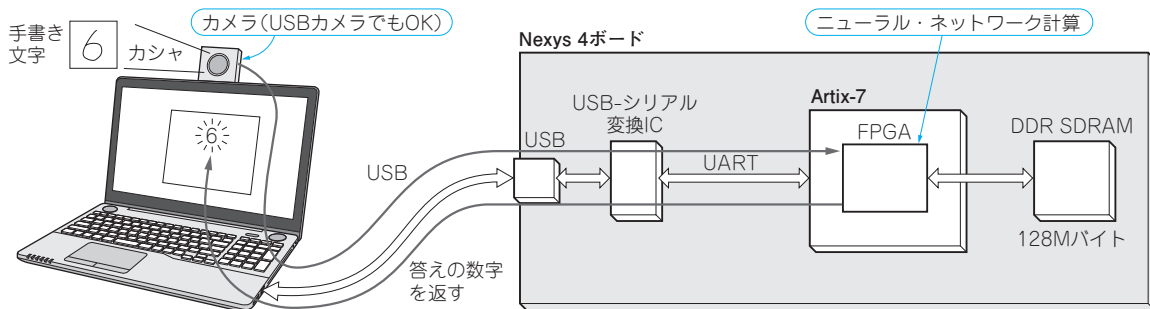


図1 実験…計算量が多いが高性能な注目の「ディープ・ニューラル・ネットワーク」をGPUを使わずにFPGAに構成して人工知能的画像認識(手書き数字認識)を行ってみる

人の脳細胞を模擬するニューロンを2次元上に並べたものを特徴マップといいます。特徴マップを多層にして構成したディープ・ニューラル・ネットワークが、画像認識や音声認識で目覚ましい成果を上げています。ディープ・ニューラル・ネットワークは、与えられたデータに対して深層学習という機械学習で学習することができ、十分なデータセットとネットワークの構成さえ用意できれば誰でも実現可能です。

ディープ・ニューラル・ネットワークの学習や予測の計算量は膨大なため(例えば積和演算10億回/sとかほしくなる)、ラズベリー・パイなどのCPUボードでは力不足である場合が多く、GPUボードがよく使われます。ディープ・ニューラル・ネットワークで行う認識処理は、認識さえできてしまえばいいので、浮動小数点演算のような高精度な計算は必要なく、整数演算でも十分であることが報告されています⁽¹⁾。従って、GPUではなく、FPGA上に整数演算による専用回路を作り込めば、GPUを使うよりもコンパクトかつ低消費電力でディープ・ニューラル・ネットワークを実現できる可能性があります。

本稿ではまず、ディープ・ニューラル・ネットワークの基礎を説明し、FPGAでディープ・ニューラル・ネットワークを構成し、画像の手書き文字(数字)認識を行ってみます(図1)。

FPGAボードには、Artix-7(ザイリンクス)搭載

Nexys 4 DDR (Digilent社)を使います。次の手順で実験を行います。

- (1) ディープ・ニューラル・ネットワークの一種であるLeNet-5をMNISTという手書き数字認識ベンチマークを使って学習
- (2) ネットワークのパラメータとネットワークを模擬するC言語のソースを出力
- (3) 無償版開発環境Vivado HLSでC言語記述からFPGA用回路データを作成(論理合成&配置配線)
- (4) FPGAに書き込んで手書き文字認識を実行

脳からヒントを得たスゴイ処理… ニューラル・ネットワークの基本メカニズム

● 脳の神経細胞ニューロンを模擬する基本回路

脳の神経細胞(ニューロン)を模擬する回路の基本構成を図2に示します。○はニューロンを示します。

N 個の n ビット信号 X_i は、他のニューロンからの入力を表します。各信号に重み W_i (n ビット)を乗算したものの総和を取った値 U を求め、この U にバイアス(W_0)を加算したものを内部信号といいます。内部信号の値によってニューロンは発火(ON/OFF動作のイメージ)します。発火を模擬する活性化関数 $f(U)$ を通すことで、ニューロンの動作を模擬できます。

活性化関数は、シグモイド関数(図3)が使われることが多いですが、ニューロンを学習するには活性化関