

量子化DNNエンジン搭載・低消費電力AIチップを試作

エッジコンピューティングの普及拡大を推進

[横浜発、2020年3月17日] 株式会社ソシオネクスト (Socionext Inc.) は、国立研究開発法人新エネルギー・産業技術総合開発機構 (以下、「NEDO」) 委託事業「進化型・低消費電力AI エッジ LSI の研究開発」の取組みの一部として、量子化ディープニューラルネットワーク (DNN) 技術を組み込んだテストチップを試作し、その動作および性能を確認しました。本テストチップはディープラーニング推論処理を高速かつ低消費電力で実行するために最適化された「量子化 DNN エンジン」を搭載しており、高度な AI 処理を小型・省電力のエッジコンピューティング機器単体で実行することが可能になります。

「IoT」の普及に伴い、エッジコンピューティングの重要性にこれまで以上に注目が集まっています。様々なアプリケーションにおいて、画像認識や解析など AI を活用した処理への要求が高まっていますが、これらの機能をエッジコンピューティング機器に実装する場合、従来の汎用 GPU をベースとしたシステムでは消費電力の増加にともなう発熱対策が必要となり、コスト増大や機器の大型化、また処理性能自体が高度化する要求を満たすことができなくなっている、という課題が明らかになっています。

「量子化 DNN エンジン」

今回当社は、AI 処理の性能を向上すると同時に消費電力を削減するための対策として、ディープラーニングを実行する上で必要なパラメータやアクティベーションを低ビット化する「量子化 DNN 技術」を使った独自のアーキテクチャを開発しました。一般的な 8 ビットに加え、1 ビット (Binary)、2 ビット (Ternary) などの低ビット化技術と独自のパラメータ圧縮技術を組み合わせることで、少ない演算資源で大量の計算処理が実行でき、かつデータ量を大幅に削減することが可能になります。

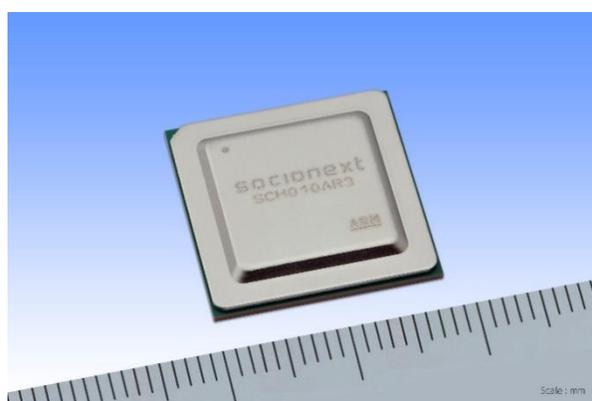
また、高効率なデータ供給を行うオンチップメモリ回路技術も新しく開発しました。多くのディープラーニング処理ではこれまで大容量・広帯域のオンチップメモリや外部メモリを必要としましたが、本技術によりメモリの使用量についても大幅な削減が可能になります。

当社はこれら新技術を採用した「量子化 DNN エンジン」を搭載したテストチップの設計と試作を行い、その動作と性能を確認しました。「YOLO v3」による 30fps での検出と 5W 以下の消費電力を達成しており、これは汎用 GPU と比較して 10 倍以上の電力効率を実現しています。

また同チップには高性能・低消費電力の Arm Cortex-A53 クアッドコア CPU を搭載しており、一般的な AI アクセラレーターとは異なり、1 チップで AI 処理全体を実行することができます。

さらに、ディープラーニングソフトウェア開発環境として、TensorFlow をベースフレームワークとした独自の低ビット Quantization Aware Training (量子化を考慮した学習) や Post Training Quantization (学習後の量子化) を実行できる環境の開発も行いました。今回のテストチップと組み合わせることで、様々なニューラルネットワークに最適な量子化技術を適用して高精度な処理を実行できるようになります。具体的には、先進ドライバー支援システム (ADAS)・監視カメラ・ファクトリーオートメーションなどの用途でのコンピュータービジョン処理を小型・省電力のエッジコンピューティング機器単体で実行することが可能になります。

ソシオネクストは今後、今回のテストチップの評価を通じて各種回路のチューニングや性能の最適化などを行い、NEDO 委託プロジェクトの最終成果物としての AI チップの完成に向けた研究開発をパートナー各社と共同で継続していきます。



量子化 DNN エンジン搭載 AI テストチップ

([クリックで拡大](#))

<NEDO 委託事業>

この成果は、国立研究開発法人新エネルギー・産業技術総合研究機構 (NEDO) の委託業務の結果得られたものです。

「高効率・高速処理を可能とする AI チップ・次世代コンピューティングの技術開発 /

革新的 AI エッジコンピューティング技術の開発 / 進化型・低消費電力 AI エッジ LSI の研究開発」

https://www.nedo.go.jp/koubo/IT3_100047.html

ソシオネクストについて

株式会社ソシオネクスト(Socionext Inc.)は、SoC(System-on-Chip)の設計・開発および販売を事業とするグローバル企業です。コンシューマ、オートモーティブおよびインダストリアル分野における世界トップレベルの技術を核に、今日のさまざまなアプリケーションの進化を支えます。長年培った技術力と経験、さらに豊富な IP ラインナップをベースに卓越したソリューションを提供し、人々の豊かな体験 = “better quality of experience”の実現に貢献します。2015年に設立された株式会社ソシオネクストは横浜市に本社を置き、日本国内、アジア、米国およびヨーロッパの各拠点において製品開発および販売活動をグローバルに展開しています。詳しくは www.socionext.com/jp をご覧ください。

記載されている会社名、製品名などの固有名詞は、各社の商標または登録商標です。プレスリリースに記載された内容は発表日現在のものです。その後予告なしに変更されることがあります。あらかじめご了承ください。