



2017年12月5日

世界最速級のディープラーニング推論エンジン『SoftNeuro™』を製品化 ～ディープラーニングの実用化を加速～

【概要】

株式会社モルフォ（本社：東京都千代田区、代表取締役社長：平賀 督基、以下 モルフォ）は本日、世界最速級^{※1}のディープラーニング^{※2}推論エンジン『SoftNeuro™』を製品化いたしました。同エンジンは単体でライセンス提供するだけでなく、モルフォの既存の画像向け認識エンジン『Morpho Deep Recognizer™』等に組み込むことで、それら既存エンジンの更なる高速化も実現します。

人工知能は近年、ディープラーニング技術の開発が進み、様々な分野でサービスや製品への組込等、実用化が進んでいます。一方、実用化にあたり、エッジ側環境で推論処理を実行する機会が増え、動作環境や、推論速度の制約といった課題が明らかになってきました。そこでモルフォでは、様々なディープラーニング・フレームワークを用いて行われた学習結果を活用して、多くの環境上で動作する高速な推論エンジン『SoftNeuro™』を開発しました。

『SoftNeuro™』は、汎用的な推論エンジンであるため、画像認識だけでなく、音声認識やテキスト解析等にも利用可能です。そのため、画像認識に限らず、ディープラーニングを用いた様々な推論エンジンの開発者向けのライセンス提供も行う予定です。

尚、『SoftNeuro™』は2017年12月6日～8日に開催される「国際画像機器展 2017」のモルフォブースにて展示・デモを行う予定です。

【ディープラーニング推論エンジン『SoftNeuro™』の特徴】

1. 世界最速級（2017年12月5日現在、モルフォ調べ）^{※1}

ディープラーニングによる推論を実用化するにあたり、推論にかかる時間が遅いなど、処理時間が課題になることも少なくありません。

『SoftNeuro™』は、後述の特徴も備えながら、CPUのみの処理において世界最速級の処理速度を実現しています。この高速性は、様々な最適化（ニューラルネットワーク、メモリ使用方法、等）を各プラットフォーム向けに最適化することで実現しています。

（詳細は、※1をご参照ください。）

2. マルチフレームワーク対応

ディープラーニングの学習を行うフレームワークは現在、Caffe、Keras、TensorFlow™等、オープンソースを中心に多数存在しています。

『SoftNeuro™』は、これら主要なフレームワークの学習結果を活用して、高速な処理を実現します（図2）。これまで利用してきたフレームワークの学習資産を無駄にすること無く、推論処理部分の高速化（第1項）や、マルチプラットフォーム対応（第3項）を実現することが可能です。なお、『SoftNeuro™』が対応するフレームワーク及びレイヤーは順次拡大予定です。

3. マルチプラットフォーム対応

ディープラーニングを用いて推論が行われる場所は、クラウド上のサーバーに限らず、スマートフォンや自動車、FA 機器等、多様化しつつあります。それらの動作環境では、CPU や OS といったプラットフォームが異なり、移植や最適化といった作業が必要となっていました。

『SoftNeuro™』は、現時点では CPU のみで推論を実行可能ですが、順次多様なプラットフォームに応じた最適化（GPU、DSP の活用等）も行う予定です（図2）。このマルチプラットフォーム対応により、学習結果を広範な動作プラットフォームへ展開することが容易になるだけでなく、動作プラットフォームの変更等に柔軟に対応可能となります。

4. セキュアファイルフォーマット対応

ディープラーニングを用いて推論が行われる場所は、クラウド上のサーバーに限らず、スマートフォンや自動車、FA 機器等、多様化しつつあります。それらの動作を実現するために、学習済みネットワークが様々な場所に複製されることにより、学習時のノウハウや結果（独自のネットワーク構造やウェイトパラメータ等）が漏洩するリスクが増大します。

『SoftNeuro™』は、学習済みネットワークを暗号化する機能を有するため、学習時のノウハウや結果の漏洩リスクの軽減を実現します。

※1 世界最速級（2017年12月5日現在、モルフォ調べ）

推論処理速度の測定は、モルフォにて入手可能な主要な推論エンジンを比較対象とし（表1）、2つのCPU環境（ARM、Intel）上で実施しています（GPUは利用していません）。

それぞれの環境上の推論処理速度測定結果は、図1-1、図1-2のとおりです。

表1. 推論処理速度測定条件

項目	内容
処理速度測定に使用した学習済みモデル （画像の1,000クラス分類）	➤ MobileNet ➤ ResNet50 ➤ VGG16
推論対象画像サイズ	224x224ピクセル

図1-1. 推論処理速度測定結果（CPU: ARM）

測定環境：Snapdragon 835

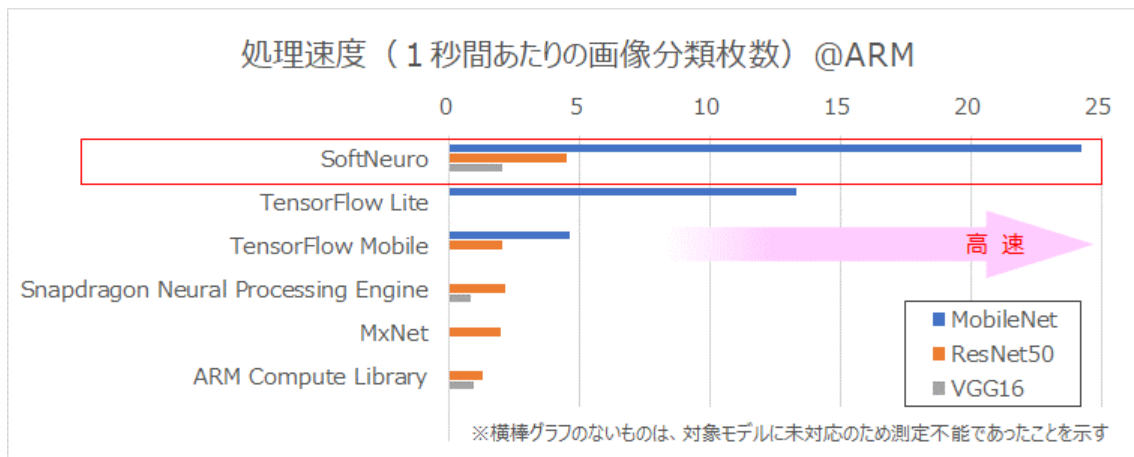
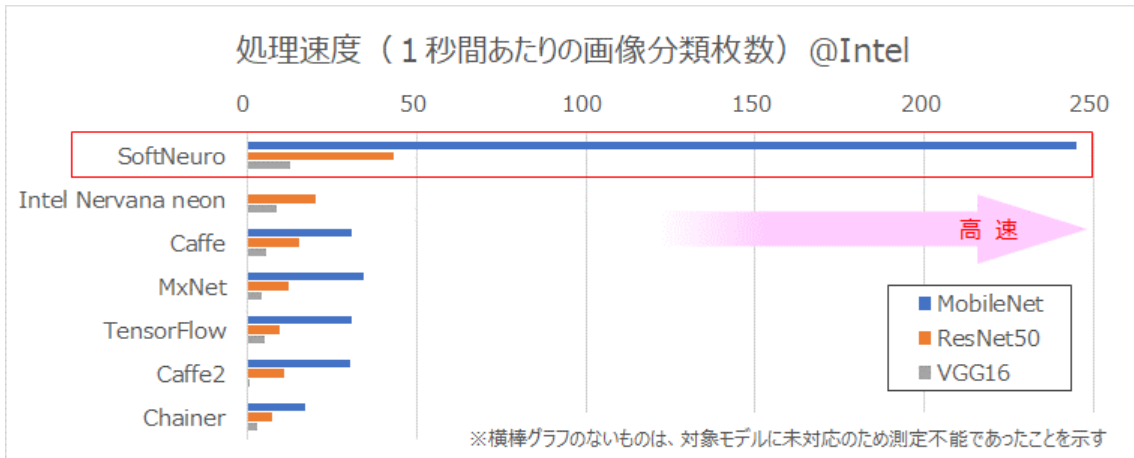


図1-2. 推論処理速度測定結果 (CPU: Intel)

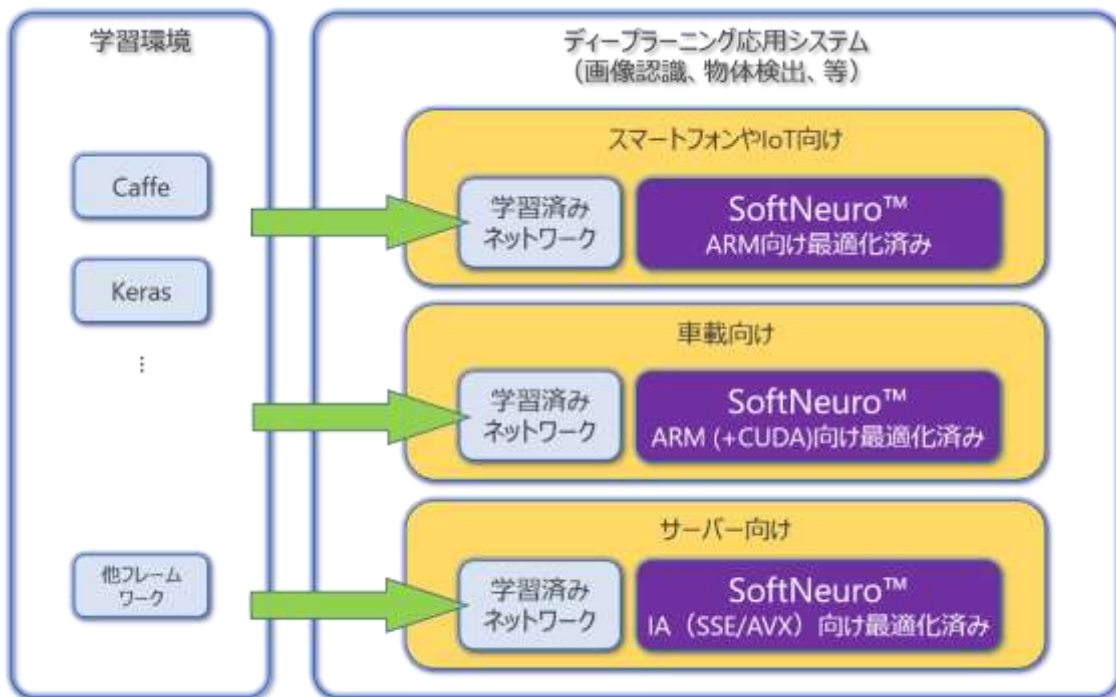
測定環境 : Intel Core i7-6700K CPU @ 4.00GHz



※2 ディープラーニング

人間が特徴を定義づけて識別する従来の機械学習とは異なり、人工知能が学習データから特徴を抽出するものです。主に画像認識・音声認識・言語処理に適用され、マーケティング、セキュリティ、自動翻訳、自動運転などさまざまな分野への活用が期待されている、注目の技術です。

図2. 『SoftNeuro™』の位置づけ



【株式会社モルフォ】

モルフォは、「画像処理技術」の研究開発主導型企業です。高度な「画像処理技術」を組み込みソフトウェアとして、国内外のスマートフォン市場、放送局やコンテンツ配信会社を中心にグローバルに展開しています。また、カメラが捉えた画像情報をデバイスやクラウドに集約し解析する、人工知能（AI）を駆使した「画像認識技術」を、車載機器、Factory Automation、医療分野などへ提供しています。様々なイノベーションの実現をモルフォのイメージング・テクノロジーで幅広くサポートしてまいります。

所在地：東京都千代田区西神田3丁目8番1号 千代田ファーストビル東館12階

代表者：代表取締役社長 平賀 督基（まさき）、【博士（理学）】

設立：2004年5月26日

資本金：1,519,070千円（2017年10月31日現在）

事業内容：画像処理技術の研究開発および製品開発ならびにライセンスング

ホームページ：<http://www.morphoinc.com/>

Facebook：<https://www.facebook.com/morphoinc>

【お問合せ先】

株式会社モルフォ 広報担当

TEL：03-6673-4332（受付：平日10時～13時、14時～18時）

お問合せ URL：<http://www.morphoinc.com/contact>

* TensorFlow, the TensorFlow logo and any related marks are trademarks of Google Inc.

* Intel, Intel Core は、アメリカ合衆国および/またはその他の国における Intel Corporation またはその子会社の商標です。

* モルフォ、Morpho およびモルフォロゴは株式会社モルフォの登録商標または商標です。

* その他の会社名および製品名は、それぞれの所有企業の商標または登録商標である可能性があります。

* Caffe、Keras、TensorFlow™は、ディープラーニングの学習や推論を行うフレームワークです。

* ARMとは、ARM Ltd が開発・提供している、スマートフォンなどの組込機器等に広く用いられているCPUです。

* CUDAとは、NVIDIA Corporation が開発・提供している、GPU 向けの汎用並列コンピューティングプラットフォームおよびプログラミングモデルです。

* IAとは、「Intel Architecture」の略称であり、Intel Corporation のマイクロプロセッサの基本設計の総称です。

* SSE/AVXとは、Intel Corporation が開発したCPUのSIMD拡張命令セットです。