



光の並列計算と ニューラルネットワークプロセッサ

金沢大学
国立研究開発法人科学技術振興機構 (JST) さきがけ
砂田 哲

1. はじめに

昨今の人工知能 (AI)・機械学習の急速な進展により、コンピューティングの需要が爆発的に増大するとともにその要求水準も劇的に変化している。半導体電子回路の技術進歩を表すムーアの法則がそろそろ終焉を迎えつつあると言われる昨今において、ポストムーアの時代を支える新しいコンピューティング技術の出現が望まれている。その中の一つに、先端的なフォトニクス技術に基づくニューラルネットワークプロセッサがある。光特有の自由度 (空間的な広がり, 波長, 偏光等) を生かした並列演算により, 高効率かつ高速・低遅延なニューラルネットワーク演算が可能になると期待され, 現在では世界的に開発が進められている。本稿では, AI アクセラレーションを目的とした最近の光プロセッサの動向について概説し, 筆者の研究例についても紹介する。

2. AI はこのまま持続的に発展できるか?

現在, 我々はコンピュータを日常的に使い, 物体を認識したり, 音声から文字や画像に変換したり, 翻訳もできる。また最近では病状の診断から車の自動運転まで高度かつ多様なタスクをコンピュータが担ってくれている。この進化を可能にした技術が機械学習の一分野であるニューラルネットワークそして深層学習である。深層学習の凄まじい進化を支える要因の一つが, 今日まで続くコンピュータの計算能力の増加と利用可能なデータ量の急増にある。最近では Graphics Processing Unit (GPU) がグラフィックス処理以外に利用され, Google の Tensor Processing Unit (TPU) など AI アクセラレータの開発にも拍車がかかっている。

深層学習などニューラルネットワークに必要な計算量は, 指数関数的に増加しており, それを支えるべく多くのコンピューティングリソース, エネルギーが大量に投入されている。画像認識分野で深層学習ブームが起こるきっかけとなった AlexNet が登場する 2012 年頃までは, ニューラルネットワーク処理に必要な計算量は, ハードウェア性能向上 (2 年で 2 倍程度) とほぼ同じスケールで増加していったが, 2012 年以降の計算量は 3, 4 ヶ月で 2 倍のペースで急増している^{1,2)}。一般に, 大きな深層学習モデルになるほど, 多くのデータを学習できて高度なタスクが可能となる。すなわち, 高度なタスクの実行には, より大きなモデル, そして更に大きなコンピューティングパワーが必要となる。急増する AI 計算のニーズに, それを支えるハードウェア技術がこのまま追従できるか懸念されている²⁾。

また, 大規模な深層学習は環境に負荷を与えるまでになっている。例えば最先端の自然言語処理用の大規模モデルの学習時に排出される CO₂ の量は, 人の一生涯分の自動車の運転で排出される CO₂ 排出量の 5 倍になっていることが報告されている³⁾。コンピューティングニーズの増加に伴う環境負荷は今後ますます増加していくかもしれない。