TSUBAME 共同利用　令和 4 年度　学術利用　成果報告書

人工画像データ及び事前学習モデルのリポジトリ化による AI ハブ構築
AI Hub of Synthetic Image Datasets and Pre-trained Models

谷村　勇輔

産業技術総合研究所　デジタルアーキテクチャ研究センター

邦文抄録（300 字程度）

大規模学習を高速に実行可能にする次世代システムの設計に向けて、アクセラレータの効果的な利用や大規模並列実行における通信の最適化、ストレージ I/O の効率化等に関する手法の研究開発を行っており、TSUBAME にてその評価検証を行った。特に、学習中に tar ファイルから直接画像を読む方法により inode 数を 1/1000 に削減した。

英文抄録（100 words 程度）

Toward the design of next-generation systems that enable high-speed execution of large-scale learning, we have been conducting research and development of methods for effective use of accelerators, optimization of communication in large-scale parallel execution, and efficiency improvement of storage I/O, etc. We evaluated and verified these methods in TSUBAME. In particular, the number of inodes was reduced to 1/1000 by reading images directly from tar files during training.

Keywords: 深層学習、事前学習、Vision Transformer、人工画像、大規模データ

背景と目的

深層学習技術の中で格段に高い精度を発揮しているのが超巨大な transformer の事前学習である。本課題はこのような巨大な transformer の事前学習に必要になる膨大なデータを数式から人工的に作った画像で代替する革新的な基盤技術を提案する。既に ImageNet-21k に関しては同等のデータ量で同じ精度を達成しており、数式から無限に生成できる人工画像で自然画像と同程度の事前学習効果が得られるという事実は、深層学習分野に革命をもたらすものであると予想される。

概要

本課題で用いる計算モデルは深層ニューラルネットの一種である vision transformer である。深層ニューラルネットはこの 20 年間で目覚ましい進歩を遂げてきた。図1に初期の LeNet や LSTM などの深層ニューラルネットから現在の Vision Transformer に至るまでの代表的な深層ニューラルネットの変遷を示す。画像処理分野では LeNet が畳み込みニューラルネットを採用することで精度を向上させ、AlexNet は GPU を用いることで高速にこれを処理できるようにし、ResNet ではスキップ接続とバッチ正規化により学習を容易にした。また、MobileNet の squeeze and excite 機構や EfficientNet の neural architecture search を用いて効率的なニューラルネットが設計できるようになり、精度を維持しつつ小型化できるようになった。一方、自然言語処理分野では LSTM などの再帰的ニューラルネットが長く用いられていたが、2017 年に登場した transformer により大幅な性能向上が実現され、現在では transformer が主流なニューラルネットとなっている。この transformer を画像処理分野で使えるようにしたものが本課題で用いる vision transformer である。LeNet から EfficientNet に至るまで画像処理分野で用いられてきたニューラルネットは、全て畳み込みニューラルネットであったが、vision transformer は畳み込みのような機構を予め人手で組み込むことなくデータからそのような構造をも学習することができるため、大量のデータがある場合には優位になる。ただし、最も大きい実画像データセット JFT を Google が非公開としてい

ることが画像処理分野全体の発展にとって大きな障壁となっている。本課題では、大規模学習を高速に実行可能にする次世代システムの設計に向けて、アクセラレータの効果的な利用や大規模並列実行における通信の最適化、ストレージ I/O の効率化等に関する手法の研究開発を行っており、TSUBAME にてその評価検証を行った。特に、学習中に tar ファイルから直接画像を読む方法により inode 数を 1/1000 に削減した。

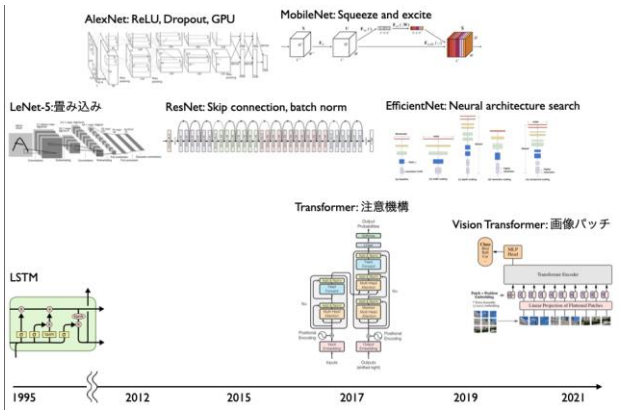

図1 主要な深層ニューラルネットモデルの変遷

結果および考察

JFT-300M 規模に相当する FractalDB-100k（1 億画像）及び FractalDB-300k（3 億画像）のデータセットを生成する必要があるが、Fractal 画像を生成するにはディスプレイに接続されている GPU を用いるのが最も高速であるため、研究室のサーバで全ての画像を生成し TSUBAME に転送することにした。ただし、FractalDB-300k を転送のために tar ファイルに固めるのに 2 週間ほどかかった。さらに、PyTorch の標準の DataLoader から読み込む場合、3 億画像をファイルシステムからアクセスすることになり、I/O がボトルネックとなる。Webdatasets を用いることで、少数の巨大な tar ファイルに画像を固めて、tar から直接読むことができる。ただし、Webdatasets を用いた場合、ノード間シャッフルは起きず、ノード内のシャッフルのみが可能である。このことが学習にどのような影響が出るかを ImageNet-1k の事前学習を用いて検証したところ、学習曲線が一致したため、問題ないと判断した。

以上の様々な問題により、本実施では当初計画を縮小し、モデルサイズは ViT-Large ではなく ViT-Base までとした。データセットは FractalDB-100k と FractalDB-300k を中心に事前学習し、余ったノードでより小規模なデータセットの学習を行った。また、FractalDB-100k では 10 epoch、FractalDB-300k では 3 epoch だけの学習を行った。これらのデータセットでの事前学習は初めてであったことと、事前の FractalDB-50k までの profile では Webdatasets を用いなくとも I/O の時間がほとんど影響しないという計測結果が出ていたため、安全をとって FractalDB-100k や FractalDB-300k の事前学習を Webdatasets を使わずにまずは行った。しかし、3 時間待っても FractralDB-100k の事前学習すら全く始まらず最初の画像読込に膨大な時間がかかっていた。後に torch.profile の計測の仕方が間違っていることが判明し FractalDB-10k などでも I/O の影響は非常に大きかったことが分かった。本実施中にはこのことは判明していなかったが、Webdatasets から FractalDB-100k/300k を読み込む方式に切替えて事前学習を行った。

まとめ、今後の課題

少なくとも ViT-Tiny/Small/Base の 3 種類のモデルと FractalDB-10k/21k/50k と ImageNet-21k の 4 種類のデータセットの全ての組み合わせにおける事前学習を十分な epoch まで完了することができた。これまでの FractalDB を用いた ViT の事前学習の研究例では、ImageNet-1k での事前学習と CIFAR10 での fine-tuning にとどまっていたが、本チャレンジではその 20 倍程度の規模となる ImageNet-21k による事前学習と ImageNet-1k による fine-tuning を行うことができた。Fine-tuning は ImageNet-1k だけでなく、CIFAR100 や CIFAR10 でも行った。興味深い結果としては、CIFAR10 や CIFAR100 などでは FractralDB-21k で事前学習したものが ImageNet-21k で事前学習したものよりも精度が低かったのに対して、最も困難とされる ImageNet-1k の fine-tuning では FractalDB-21k の方が ImageNet-21k よりも高精度になった。さらに、ViT-Tiny や ViT-Small などの小さいモデルでは ImageNet-21k の方が FractalDB-21k よりも精度が

よくなったが、最も大きい ViT-Base では FractalDB-21k の方が ImageNet-1k の fine-tuning において高い精度を達成した。つまり、fine-tuning タスクが困難であるほど、また、ViT の規模が大きいほど FractalDB の ImageNet に対する優位性は向上するという興味深い結果となった。