

イノベーション創発に資する人工知能基盤技術の創出と統合化
2019 年度採択研究代表者

2020 年度 年次報告書

篠田 浩一

東京工業大学 情報理工学院
教授

社会インフラ映像処理のための高速・省資源深層学習アルゴリズム基盤

§ 1. 研究成果の概要

機械学習と高性能計算の研究者が密に連携した Co-Design の枠組みを活用して現在より 10 万倍速い超高速な深層学習基盤を構築し、その上で大量の高精細映像の実時間解析を実現して安心・安全な社会作りに貢献することを目的としている。

機械学習においては、二次最適化、自己教師付き学習、ニューラルモデル構造探索、グラフニューラルネットワークなどの研究を行った。クロネッカー因子分解を用いた二次最適化の高速近似解法はこれまでその有効性が経験的にしか示されてこなかったが、層の数が無限大の極限では厳密な自然勾配法と同じ収束率をもつこと、深層学習のような超高次元モデルでもノイズを加えながら勾配法を実行することで大域的最適解に収束することを理論的に証明した。さらに、グラフニューラルネットワークが多層になると性能が劣化する問題の理論的解明を行い、その問題を回避したマルチスケール GCN 法を提案した。

また、「富岳」における 10 万ノード規模の学習アルゴリズムの超並列化に向けて、3D CNN におけるモデル並列化を推進するとともに、言語モデルや脳シミュレーションのアプリについても「富岳」の評価環境を用いた並列処理性能の予備評価を進めた。一方で、物体検出や領域分割などの応用に向けて FPGA 実装を行い、物体検出応用に関して CPU や GPU よりも高速な処理を実現した。学習後のモデルを FPGA に効率よくポータリングするための Co-Design 環境を開発した。

さらに、新型コロナウイルスの感染リスクの評価を目的として 2021 年 1 月 4 日に国立競技場で開催された Jリーグルヴァンカップ決勝戦の観戦映像データを取得した。また、これまでに作成してきた学習モデルを用いて取得したデータを処理することで顔が識別できない程度の解像度の画像でも人の領域を抽出できることが分かった。

§ 2. 研究実施体制

(1) 篠田グループ

- ① 研究代表者: 篠田 浩一 (東京工業大学情報理工学院 教授)
- ② 研究項目
 - ・社会インフラ映像処理のための高速・省資源深層学習アルゴリズム基盤

(2) 松岡グループ

- ① 主たる共同研究者: 松岡 聡 (理化学研究所計算科学研究センター センター長)
- ② 研究項目
 - ・10 万ノード超並列処理と資源スケジューリング

(3) 大西グループ

- ① 主たる共同研究者: 大西 正輝 (産業技術総合研究所人工知能研究センター 研究チーム長)

② 研究項目

- ・実社会応用における評価

(4)横田グループ

① 主たる共同研究者:横田 理央 (東京工業大学情報理工学院 准教授)

② 研究項目

- ・確率的勾配降下法に代わる二次最適化

(5)村田グループ

① 主たる共同研究者:村田 剛志 (東京工業大学情報理工学院 准教授)

② 研究項目

- ・グラフ深層ニューラルネットによる知識処理との融合

(6)中原グループ

① 主たる共同研究者:中原 啓貴 (東京工業大学工学院 准教授)

② 研究項目

- ・FPGA 実装向け深層学習回路

(7)鈴木グループ

① 主たる共同研究者:鈴木 大慈 (東京大学大学院情報理工学系研究科 准教授)

② 研究項目

- ・深層学習理論(汎化誤差、圧縮率、収束率)

【代表的な原著論文情報】

1. Understanding Approximate Fisher Information for Fast Convergence of Natural Gradient Descent in Wide Neural Networks, Ryo Karakida and Kazuki Osawa, Advances in Neural Information Processing Systems (NeurIPS 2020)
2. Yosuke Oyama, Naoya Maruyama, Nikoli Dryden, Erin McCarthy, Peter Harrington, Jan Balewski, Satoshi Matsuoka, Peter Nugent, Brian Van Essen, "The Case for Strong Scaling in Deep Learning: Training Large 3D CNNs with Hybrid Parallelism", Special Section on Parallel and Distributed Computing Techniques for AI, ML, and DL, In IEEE Transactions on Parallel and Distributed Systems (TPDS), Volume 32, Issue 7, Dec 2020
3. Scalable and Practical Natural Gradient for Large-Scale Deep Learning, Kazuki Osawa, Yohei Tsuji, Yuichiro Ueno, Akira Naruse, Chuan-Sheng Foo, Rio Yokota, IEEE Transactions on Pattern Analysis and Machine Intelligence 2020;PP:10. 1109/TPAMI. 2020. 3004354.
4. Hiroki Nakahara, Zhiqiang Que, Wayne Luk, "High-Throughput Convolutional Neural Network on an FPGA by Customized JPEG Compression," FCCM 2020, pp. 1-9.

5. Taiji Suzuki: Generalization bound of globally optimal non-convex neural network training: Transportation map estimation by infinite dimensional Langevin dynamics. NeurIPS2020, pp. 19224–19237, 2020. (spotlight)