

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. 研究実施体制

### (1) 篠田グループ(契約単位ごと)

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

・知識の構造を活用した高速な深層学習アルゴリズム

### (2) 松岡グループ(契約単位ごと)

① 主たる共同研究者: 松岡 聡(東京工業大学学術国際情報センター、教授)

② 研究項目

・ノード間の通信処理を削減するための高並列アルゴリズムと資源スケジューリングによる全体最適化

### (3) 村田グループ(契約単位ごと)

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

・ノーリアルタイム認識・解析のための Deep Net 構造のコンパクト化アルゴリズム

### (2) 横田グループ(契約単位ごと)

① 主たる共同研究者: 横田 理央(東京工業大学学術国際情報センター、准教授)

② 研究項目

・個々の計算ノードにおける計算量を削減するための行列構造化アルゴリズム

## § 2. 研究実施の概要

この研究期間内では、まず、共同研究基盤の構築を行った。今後のベンチマークとなる、国際ワークショップ TRECVID の映像からのマルチメディアイベント検出(MED)のタスクについて、映像データ、開発ツール、および、評価環境の整備を行った。そして、それらを東工大のスーパーコンピュータ TSUBAME 上に展開し、研究参加者全員が同一のプラットフォームで研究開発を行える環境を構築した。また各グループで以下に述べる要素技術の開発を行った。

篠田グループでは、MED のためのプラットフォームとして、長・短期記憶(LSTM)を用いた再帰型ニューラルネットワーク(RNN)を実装し、その評価を行った。また、イベントを構成するコンセプト間の因果関係をモデル化するために、コネクショニスト時系列分類法(CTC)と n グラム言語モデルを用いた手法を開発した。人間の動作認識のタスクでその評価を行い、従来法を上回る性能を得た。その成果を国際会議に投稿する予定である。

松岡グループでは、並列学習アルゴリズムの性能モデルを作成した。確率的最急降下法(SGD)を用いた畳み込みニューラルネットワーク(CNN)の学習において、ネットワーク構造とハードウェア性能を与えて学習の制御パラメータの最適値を予測するモデルを作成した。また、並列学習アルゴリズムにおいて、単精度浮動小数点数(Float)型データの仮数部をノード間の通信時のみ 8bit に量子化する方式を提案し、認識精度を低下させることなく 2 倍以上の高速化を実現した。

村田グループでは、深層ニューラルネットワーク(Deep Net)のサイズ圧縮を行う既存手法を調査した。まず、局所的な間引きを行う手法を実装し、評価した。30~40%の間引きでは精度の劣化は小さく、向上する場合もあった。また、結合重みのクラスタリングを用いて大域的な圧縮を行う手法の多くは、圧縮後の Deep Net の使用に際し特別なハードウェアを必要とし、汎用性に乏しいことがわかった。今後、より汎用性の高い手法の実現を目指す。

横田グループでは、深層学習においてノード内の演算量を低減するための、低ランク近似を用いた行列構造化アルゴリズムを開発し、メモリ消費量、演算量を各々2分の1に低減した(右図)。その成果を国際会議に投稿した(2017年5月に採録決定)。今後、モデルの種類やバツ

チ学習手法を検討し、メモリ消費量と演算量の各々をさらに5分の1まで低減することを目指す。

