

論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Improving Deep Learning Efficiency Using Reservoir Computing Inspiration
著者(和文)	LOPEZ GARCIA-ARIASANGEL
Author(English)	Ángel López García-Arias
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第12710号, 授与年月日:2024年3月26日, 学位の種別:課程博士, 審査員:本村 真人,一色 剛,高橋 篤司,佐々木 広,原 祐子,藤木 大地
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第12710号, Conferred date:2024/3/26, Degree Type:Course doctor, Examiner:,,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	審査の要旨
Type(English)	Exam Summary

(博士課程)

論文審査の要旨及び審査員

報告番号	甲第	号	学位申請者氏名	LOPEZ GARCIA-ARIAS ANGEL	
論文審査 審査員		氏名	職名	氏名	職名
	主査	本村 真人	教授	原 祐子	准教授
	審査員	高橋 篤司	教授	藤木 大地	准教授
		一色 剛	教授		
	佐々木 広	准教授			

論文審査の要旨 (2000 字程度)

本論文は、Improving Deep Learning Efficiency Using Reservoir Computing Inspiration (リザーバー計算に着想を得た深層学習の効率向上)と題し、英文8章から構成されている。

第1章 Introduction (序論)では、本研究で取り組む問題の紹介が行われている。10年以上にわたり深層学習は商業アプリケーションを支える技術として成熟してきたが、現在の深層学習が要求する高い計算資源とメモリ資源の消費量は、これを処理する巨大データセンターのリソース消費量が世界的な環境問題にインパクトを与えるまでに大きくなっておりその解決が求められている、と述べている。

第2章 Background (研究背景)では、この論文の基盤となる基本的な研究について紹介されている。具体的には、リザーバー計算、宝くじ仮説、ビジョン用の再帰型ニューラルネットワーク、およびスパースニューラルエンジンアクセラレータについて解説されている。

第3章 Efficient Design Inspiration From Reservoir Computing (リザーバー計算に基づく高効率設計手法に着想)では、リザーバー計算の潜在的な低コスト化の代替手段として、セルオートマトンをベースにした実装が述べられている。この実装を使用して、リザーバーコンピューティングが新奇技術から実用化技術へ進む際のスケラビリティの課題を説明している。さらに、この章ではリザーバー計算を深層学習技術で改善するのではなく、逆にリザーバーコンピューティングのキーエレメントを深層学習に持ち込んで効率を向上させるという、新たな研究方向を提案している。

第4章 Hidden-Fold Networks: Random Recurrent Residual Networks Contain Stronger Lottery Tickets (隠れ多重ネットワーク: ランダム再帰型・残差型ネットワークにおけるより強力な宝くじの発見)では、リザーバーコンピューティングにインスパイアされた初の深層学習モデルが提案されている。強い宝くじ仮説と組み合わせ、ランダムな固定重みを持つスパースな再帰型ニューラルネットワークが構築されている。このモデルは、接続パターン (スーパーマスク) の学習のみで、画像分類の競争力のある精度と最大50倍の圧縮率を達成し、オンチップSRAM (約2MB) に収まるモデルを提供している。

第5章 Multicoated Supermasks: Scalar Supermasks Enhance Strong Lottery Tickets (多層重ね合わせスーパーマスク: スカラスーパーマスクによる強い宝くじの強化)では、強い宝くじ仮説の改良が提案されている。複数のマスクを組み合わせ、重み学習モデルと同等の精度を持つスカラスーパーマスクが構築されている。この方法を、従来段階的に行われていた枝刈り、量子化、学習を同時に行う新たな学習手法として提案している。

第6章 The Ternary Strong Lottery Ticket Hypothesis: A Novel Framework for Weight Randomness and Quantization (三値強い宝くじ仮説: 重みのランダム性と量子化のための新しい枠組み)では、接続性、重みの符号、重みの大きさを三つの独立した基本的なスーパーマスクとして見ること、強い宝くじ仮説を任意の接続性を使えるように拡張している。これにより、学習コストの削減とリザーバー計算類似性の向上の可能性が広がる。さらに、三つのマスクを組み合わせること、隠れ多重ネットワークを高精度で高圧縮のモデルに強化できることを示している。

第7章 WhiteDwarf: 40-nm Strong Lottery Ticket Accelerator With Triple Sparsity Exploitation (WhiteDwarf: 三種類のスパース性を利用した40nm強い宝くじアクセラレータ)では、提案したアルゴリズムのためのスパースなニューラルアクセラレータのアーキテクチャを提案し、1実チップとして設計・製造した内容を述べている。三種類のパラメータスパース性を利用することで、モデルのメモリサイズをチップ内で最大330倍、チップ外で最大2000倍に圧縮可能なことが示されている。さらに、乗算器なしの演算やクロックゲーティングなどの回路技術を駆使して消費電力をさらに削減している。

以上を要するに、本論文はランダム固定ネットワークと小さな学習可能部分を持つリザバー計算の考え方に触発され、深層学習におけるニューロン間の接続性、重みの符号、重みの大きさに関するランダム固定可能性を解析し、その中から枝刈と量子化を統合した新たな深層学習手法により深層学習モデルが大幅に小型化可能なことを提示したものであり、学術的および工学的貢献は大きい。よって、審査員は本論文が博士（工学）の学位論文として十分に価値があるものと認める。

注意：「論文審査の要旨及び審査員」は、東工大リサーチポジトリ(T2R2)にてインターネット公表されますので、公表可能な範囲の内容で作成してください。