

論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Neural Network Hardware Codesign via Nimble FPGA Implementations and Synchronized Spiking in Edge-AI
著者(和文)	BartelsJim
Author(English)	Jim Bartels
出典(和文)	学位:博士(学術), 学位授与機関:東京工業大学, 報告番号:甲第12922号, 授与年月日:2024年9月20日, 学位の種別:課程博士, 審査員:伊藤 浩之,岡田 健一,徳田 崇,原 祐子,白根 篤史,Indiveri Giacomo
Citation(English)	Degree:Doctor (Academic), Conferring organization: Tokyo Institute of Technology, Report number:甲第12922号, Conferred date:2024/9/20, 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		Bartels Jim	
論文審査 審査員		氏名	職名		氏名	職名
	主査	伊藤 浩之	教授	審査員	原 祐子	准教授 情報通信系
	審査員	岡田 健一	教授		Giacomo Indiveri	教授 ETH Zurich
		徳田 崇	教授			
		白根 篤史	准教授			

論文審査の要旨 (2000 字程度)

本論文は「Neural Network Hardware Codesign via Nimble FPGA Implementations and Synchronized Spiking in Edge-AI (エッジ AI における軽量な FPGA 実装と同期スパイクによるニューラルネットワークハードウェア協調設計)」と題し、英文 9 章からなっている。

第 1 章「Introduction (序論)」では、エッジ AI の役割と重要性を述べ、エッジ AI の性能を改善するための 2 つのハードウェア協調設計戦略として、ウシの行動推定を題材としたアプリケーション指向の手法と、FPGA とミックスドシグナル・ニューロモर्फック・プロセッサ (DYNAP-SE2) 用スパイクニューラルネットワーク (SNN) 設計における同期の利用に着目したダイナミクス指向の手法を提案している。

第 2 章「Edge-AI for Precision Livestock Farming: A Study on Cow Behavior Estimation (精密畜産のためのエッジ AI : ウシの行動推定に関する研究)」では、AI を利用した精密畜産の可能性と制約を概説し、ウシ行動推定のための 2 つの低消費電力設計手法を述べている。

第 3 章「A 216 μ W, 87% Accurate Cow Behavior Classifying Decision Tree on FPGA With Interpolated Arctan2 (補間 Arctan2 を用いた 216 μ W, 87% 精度の FPGA 上のウシ行動分類用決定木)」では、ウシの行動分類のために首の向きを特徴量とした精度 86.8%、消費電力 216 μ W の決定木を提案し、その FPGA 実装について詳述している。

第 4 章「TinyCowNet: Memory- and Power-Minimized RNNs Implementable on Tiny Edge Devices for Lifelong Cow Behavior Distribution Estimation (TinyCowNet : 生涯のウシ行動分布推定のための小さなエッジデバイスで実装可能なメモリと電力を最小限に抑えた RNN)」では、リカレントニューラルネットワーク (RNN) の新たな量子化スキームと、提案する RNN アーキテクチャ最適化手法によって、2.043kB の小さなメモリサイズで、95.5% という高い推定精度を達成している。

第 5 章「An Integer-Only Resource-Minimized RNN on FPGA for Low-Frequency Sensors in Edge-AI (エッジ AI における低周波センサのための FPGA 上の整数型リソース最小化 RNN)」では、RNN の新たな整数ベース演算法や量子化手法によりリソースを最小限に抑えたハードウェアアーキテクチャを紹介している。これにより、消費電力 340 μ W \sim 3.81mW、消費エネルギー 542 μ J \sim 44.3 μ J を達成し、エッジ AI をアクセシビリティと効率性の面で強化するために RNN ハードウェア・ソフトウェアを協調設計する有効性を実証している。

第 6 章「An Experimental Study of Synchronization Phenomena in Chaotic Spiking Oscillators Towards Physical Reservoirs (物理的リザーバに向けたカオススパイク発振器における同期現象の実験的研究)」では、デュアル・トランジスタ・スパイク発振器のリングネットワークにおける同期現象の実験的研究について説明し、電源電圧等を変化させた場合の影響を調べ、同期を利用した物理的リザーバを提案している。

第 7 章「A Seizure-Encoding Spiking Neural Network Using Partial Synchronization on a Mixed-Signal Neuromorphic Processor (ミックスドシグナル・ニューロモर्फック・プロセッサ上の部分同期を用いた発作符号化スパイクニューラルネットワーク)」では、DYNAP-SE2 上の指数関数的リーク・インテグレート・アンド・ファイヤニューロンのネットワークを用いて頭蓋内脳波データから発作を符号化するために同期を利用する手法を提案し、さらに、ミックスドシグナルハードウェア上の固有のダイナミクスを考慮して SNN を符号化することの重要性を示している。

第 8 章「NimbleSNN: A 1.34 μ J/Image 4639 Logic Element Single-Spike Accelerator on FPGA Using Synchronization (NimbleSNN : 同期を用いた FPGA 上の 1.34 μ J/画像 4639 論理素子シングルスパイクアクセラレータ)」では、SNN におけるタイム・トゥ・ファースト・スパイクコーディングやニューロンにおける同期を活用した NimbleSNN ハードウェアアーキテクチャを提案している。これにより、FPGA 上で、論理素子数の削減と、MNIST 画像 1 枚あたり 65.8 μ s の推論時間、1.34 μ J の消費エネルギーを達成している。

第 9 章「Conclusion, Discussion and Future Work (結論, 考察, および今後の課題)」では、本論文で得られた成果をまとめ、将来の研究展望としてニューラルネットワークのハードウェア設計のための学際的アプローチの必要性を論じ、本論文を締めくくっている。

以上を要するに、本論文は、エッジ AI における性能を改善するための協調設計戦略を論じたもので、学術上、産業上貢献するところが大きい。よって我々は、本論文が博士(学術)の学位論文として、十分に価値あるものと認める。