

論文 / 著書情報
Article / Book Information

題目(和文)	GPU を用いたコンピュータビジョンの高速化と最適化の研究
Title(English)	GPU-Based Acceleration and Optimization Research on Computer Vision
著者(和文)	WANGCHENYU
Author(English)	Chenyu Wang
出典(和文)	学位:博士(理学), 学位授与機関:東京工業大学, 報告番号:甲第12513号, 授与年月日:2023年9月22日, 学位の種別:課程博士, 審査員:遠藤 敏夫,坂本 龍一,高邊 賢史,脇田 建,佐藤 育郎
Citation(English)	Degree:Doctor (Science), Conferring organization: Tokyo Institute of Technology, Report number:甲第12513号, Conferred date:2023/9/22, 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	WANG CHENYU	
論文審査 審査員		氏 名	職 名		氏 名	職 名
	主査	遠藤 敏夫	教授	審査員	佐藤 育郎	特任准教授
	審査員	坂本 龍一	准教授			
		高邊 賢史	准教授			
		脇田 建	准教授			

論文審査の要旨（2000 字程度）

本論文は「GPU-Based Acceleration and Optimization Research on Computer Vision (GPU を用いたコンピュータビジョンの高速化と最適化の研究)」と題し、コンピュータビジョンのモデル改善のために、高い並列性を持つ GPU を利用した高性能実装について提案・評価を行い、演算速度と精度のトレードオフの改善を実証している。本論文は以下のような英文 5 章で構成されている。

第 1 章「Introduction」では、本研究の動機として、現在のコンピュータビジョンモデルは大量の計算を必要とし、それがリアルタイムの応用における大きな障壁となっているという背景を述べている。演算速度と精度の間には一般的にはトレードオフがあり、その改善のために高い並列性を持つ GPU を有効に活用できるようモデルやアルゴリズムを改良するという方針を示している。そして本論文では主に Single Shot Multibox Detector と Swin Transformer を取り上げ、改善の対象とすることを述べている。

第 2 章「Background」では、本論文を理解するうえで必要な背景知識を解説している。まずコンピュータビジョンモデルの基本的な解説を行い、またニューラルネットワーク(NN)処理との関係を説明している。この部分には、機械学習の基本的なアルゴリズムと深層学習のネットワーク構造についての詳細な説明が含まれている。さらに、本研究で活用するハードウェア要素である GPU アーキテクチャについて解説し、その高い並列性を効率的に利用することが本研究の目的の達成に必要であることを述べている。

第 3 章「Speed-up Single Shot Detector on GPU with CUDA」では、物体検出タスクに広く採用されている Single Shot MultiBox Detector (SSD)モデルの高速な実装およびその評価について述べている。既存の SSD 実装では、最も多くの計算量を必要とする NN 処理部分のみを GPU 対応していたが、本章の詳細な性能分析を通してそれ以外の部分が実行時間の比較的大きな割合を占めていることを示している。そこで本章では全ての処理を GPU にて実行する実装方式を提案している。特に、後処理部分に含まれる Non-Maximum-Suppression 演算において、高い並列性を利用するソート処理を実装している。性能評価により、提案した実装は既存実装に比較して後処理部分で 53%、物体検出処理全体で 23%高速であることを示している。検出精度の低下は 0.5%以下と微小であり、提案技術により演算速度と精度の良好なトレードオフを達成していることを示している。

第 4 章「Optimization for the Swin Transformer」では、多彩なコンピュータビジョンタスクにおいて優れた成果を上げている Swin Transformer モデルについて、さらなるモデル改良による演算速度と精度のトレードオフ改善を提案し、評価している。まず関連技術の総括を行うとともに、Swin Transformer の課題として、広域的に分散した情報の関連性を捉えるにあたり、画一的なグリッドサイズが表現上の自由度を制約していることを示している。それを改善するために、Pyramid Swin Transformer と呼ばれるモデルを提案している。この新しいモデルでは、統一されたスケールでサイズの異なるウインドウを使用し、複数の Window-based Multi-head Self-attention (W-MSA)計算を実行するという方針を採用している。この方式により、広域の関連性をより良く抽出することが可能な一方で、既存方式よりも計算量が増加する。この方式の実用性を評価するために、画像分類、物体検出などの 4 つのタスクにおいて計算量と精度の評価を詳細に行い、Swin Transformer を含む種々の既存手法と比較して、トレードオフを改善できることを実証している。さらに GPU 上での計算速度を向上させるために、W-MSA 計算処理の複数の並列化手法を提案し、予備的な評価を行っている。

第5章「Conclusion & Future Work」では、本研究の結論をまとめ、将来の研究の可能性を述べている。また、本研究の限界とそれを克服するために可能な手段についても考察し、新たな研究の道筋を示している。

以上のように、本研究はコンピュータビジョンモデルを改良するために、GPUを活用して計算速度と精度のトレードオフを改善する方針を示している。その方針に基づき SSD モデルの実装の改善および Pyramid Swin Transformer モデルの提案を行い、トレードオフの評価を通してその有用性を実証しており、理学上の貢献するところ大である。よって本論文は博士（理学）の学位論文として十分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、東工大リサーチリポジトリ(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。