JAIST Repository
https://dspace.jaist.ac.jp/

Title	画像分類モデル保護手法のFPGA実装と検証 [課題研究報告書]
Author(s)	小俣, 直史
Citation	
Issue Date	2025-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/19796
Rights	
Description	Supervisor: 井口 寧, 先端科学技術研究科, 修士 (情報科学)

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

画像分類モデル保護手法の FPGA 実装と検証

2310054 小俣 直史

近年，社会における深層ニューラルネットワーク（Deep Neural Network: DNN）と，その活用が進んでいる．DNN モデルはデータセットの学習によって生成され，入力データから推論を実行し，出力データを生成するための”層構造アーキテクチャ”と，各層に含まれる重み・バイアス パラメータから構成される．特に，高性能な DNN モデルの学習には膨大なデータセットと計算資源，専門知識が必要とされ，学習済みの DNN モデルは知的財産に値する．しかし，組み込みコンピュータや FPGA 等のエッジデバイスをターゲットとし，サイドチャネル攻撃等の不正アクセスによって DNN モデルを不正取得される可能性が存在する．

このような経路によって取得されてしまった DNN モデルの更なる不正な利用・配布を防ぐため，エッジデバイス上の DNN モデルを保護する手法が提案されてきた．Manaar Alam らによって提案された NN-Lock アルゴリズム [1] は組み込みコンピュータを対象とした保護手法の 1 つである．DNN モデルのうち，ノード間接続の強さを示す係数重みパラメータ (Weight Parameter) を AES アルゴリズムによって全て暗号化し，推論時に正しい利用者が正しい暗号鍵を使用することで一時的に復号化する．一方で，窃取者等の不正な利用者が誤った暗号鍵を使用した場合，暗号化されたパラメータの値の変化が後続の全ての層に伝播し，最終的な推論結果が暗号化前のモデルと異なるものになる．これにより，正しい復号化キーを持たない攻撃者が DNN モデルを窃取しても，重みパラメータが暗号化されているため DNN モデルの推論精度が大幅に低減し，利用することは困難となり，DNN モデルの不正な利用・配布を防ぐことができる．

近年，FPGA（Field-Programmable Gate Array）を自動車の先進運転支援システムのプロセッサとして使用し，搭載した DNN モデルによって推論処理を実行する事例が増加傾向にある．本研究では，NN-Lock アルゴリズムを FPGA 上のハードウェア化された DNN モデルに初めて適用する．具体的には，重みパラメータが既に NN-Lock アルゴリズムによって暗号化された DNN モデルと NN-Lock アルゴリズムの復号化機構を FPGA 上に構築する手法を提案する．その際，NN-Lock アルゴリズムによって暗号化された重みパラメータを FPGA 上のメモリ BRAM に書込む直前 (BRAM 書込前 NN-Lock) と DNN 推論を実行する直前 (DNN 実行前 NN-Lock)，いずれのタイミングで重みパラメータを復号化するかという”NN-Lock の適用箇所”と DNN，NN-Lock モジュールの処理並列数を変更して実装する．あわせて，NN-Lock の FPGA 実装に適した DNN 推論演算モジュール単体の実装手法も提案する．

その上で，提案手法の運転支援システムとしての活用を想定し，提案手法の実装・評価から得られた回路量と推論処理時間より，実際に流通する FPGA ボードやカメラモジュールの性能を参考に，回路規模と推論処理時間の要求

要件を考察する．また，NN-Lock 適用箇所における重みパラメータの窃取耐性についても，FPGA を対象とした代表的な攻撃手法を複数取り上げ考察する．以上の観点から，幅広いハードウェアを対象に，運転支援システムとしての活用の想定下で，FPGA 上で動作する DNN モデルに NN-Lock を適用する際に必要な情報を提供する．

参考文献

[1] Manaar Alam, Sayandeep Saha, Debdeep Mukhopadhyay, and Sandip Kundu. NN-Lock: A Lightweight Authorization to Prevent IP Threats of Deep Learning Models, ACM Journal on Emerging Technologies in Computing Systems, Volume 18, Issue 3, Article 51 (2022)