暗号化画像を用いたプライバシー保護を考慮した深層学習

Privacy-Preserving Deep Learning with Encrypted Images

貴家　仁志（Hitoshi Kiya, Ph. D.）
東京都立大学　システムデザイン学部　特別先導教授
(Leading Professor, Tokyo Metropolitan University, Faculty of System Design)
IEEE 電子情報通信学会
受賞：電子情報通信学会論文賞 (2008 年).　電気通信普及財団テレコムシステム技術賞（2011 年）他
著書：ディジタル信号処理, コロナ社 (2014 年)　他
研究専門分野：信号処理　コンピュータビジョン　情報セキュリティ

あらまし

視覚情報を保護した形式で画像データを、深層ニューラルネットワーク（DNN）モデルに活用する新しい方法について研究を行った。高性能な DNN モデルの有効活用には、膨大なデータを必要とし、かつ画像データの多くには生体画像に代表されるように個人情報が含まれる。そのため、DNN モデルのための十分な量の良質なデータを確保することは容易ではない。本研究では、学習可能な画像暗号化という新しい暗号化を提案して、それを直接 DNN モデルに適用することによって、上述の課題の解決を試みた。また暗号化画像の利用がモデル性能に及ぼす影響と、使用する暗号化の安全性の観点から提案法の有効性を評価した。

１．研究の目的

本研究の目的は、プライバシー保護を考慮した新しい深層学習法を構築するために、深層学習に直接適用可能な画像暗号化法、すなわち学習可能な画像暗号化法を提案して、関連する研究分野に貢献することである。本研究では、特に DNN モデルの性能劣化を生じさせない画像暗号化法の開発と最先端の DNN モデルの特徴を活用したプライバシー保護法の考察に焦点をあて、目的の達成を目指す。

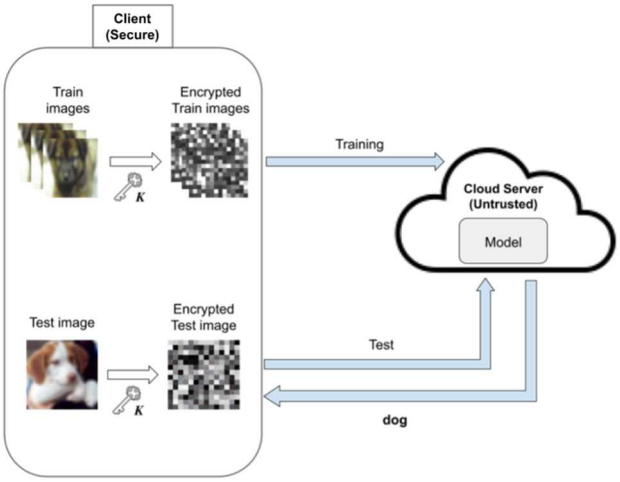
２．研究の背景

急速な発展を続ける DNN は、様々な分野において有効性の確認や実用化が進んでいる。しかし、最先端の DNN モデルは多くのパラメータを有し、それらの決定には膨大な量の学習用データの確保が必須である。しかし、画像の多くは、個人、時間、撮影場所の特定などに関係する個人情報を含んでいるため、プライバシーを保護し高性能な DNN に基づくシステムを実行することとは容易ではない[1]。さらに、膨大なメモリコストと計算コストの理由から、クラウド環境においてシステムを実行する必要性が生じるが、プライバシー保護と不正利用回避の観点からクラウド環境は、一般に信頼できる環境ではない[2]。このことがプライバシー保護を考慮した深層学習複数をより困難にし、DNN の活用に大きな制約を与えている。

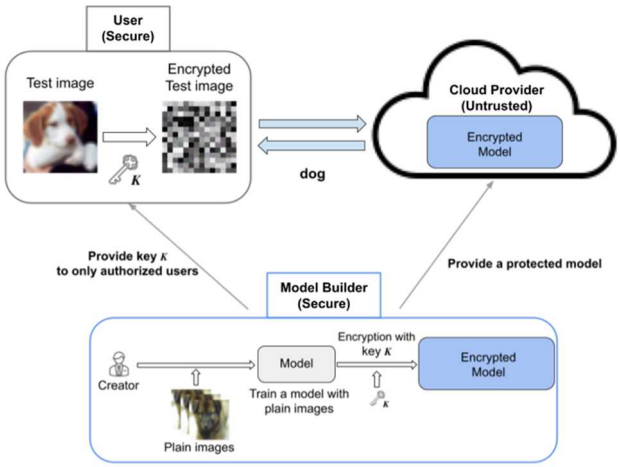
これらの背景から、本研究では、学習可能な画像暗号化という新しい暗号化を提案して、それを直接 DNN モデルに適用することによって、上述の課題の解決を目指す[3]。図１は、プライバシー保護を考慮した DNN のための二つのシナリオである。シナリオ１では、モデルの学習及び各テスト画像に対するモデルによる推定を、共に信頼できない環境（クラウド）で行う。一方、シナリオ２では、モデルの学習は信頼できる環境で行い、モデルによる推定のみをクラウド環境で行う。先行研究の多くは、シナリオ１を想定して進められてきた。本研究では、シナリオ２の想定を考察した場合には、より高性能な DNN モデルが構築可能であることを示す。

暗号化画像を用いたプライバシー保護を考慮した深層学習

Privacy-Preserving Deep Learning with Encrypted Images



(a) シナリオ１



(b)シナリオ２

図１　プライバシー保護の二つのシナリオ

３．研究の方法と結果

画像分類問題を例にして、提案法を説明する。提案法は、高性能な画像分類モデルの一つである Vision Transformer (ViT) [4] への適用を前提に説明される。

３．１　プライバシー保護のための要求条件

図１(b)は、プライバシー保護を考慮した画像分類問題のシナリオの一つである。このシナリオでは、クラウドプロバイダーを不正利用やデータを流出させる可能性があると位置づけ、テスト画像の視覚情報を保護した形式で、画像分類を実行する。さらに、モデルの学習に使用した画像も、モデルから視覚情報の復元を困難とすることを目指す。

このシナリオを実現するために、以下の要求条件をできる限り満たす必要がある。

(a) 正しい鍵を持たない攻撃者は、暗号化画像から視覚情報の復元が困難であること。

(b)モデル学習に使用した画像の視覚情報をモデルから復元することが困難であること。

(c)暗号化画像を使用した場合も、モデル本来の性能の低下がないこと。

(d)鍵の更新が容易で、かつ複数のクライアントに対して異なる暗号化鍵の提供が可能であること。

提案法では、上述の要求条件を考慮している。特に、条件(b)及び(c)を満たす手法の提案は初めてであり、当該分野に対して大きな貢献がある。さらに、暗号文攻撃（Ciphertext-only attack: COA）の仮定の下で、暗号化画像の安全性評価を行った。

３．２　テスト画像の暗号化

図２に示すように、プライバシー保護のために、テスト画像と学習済みのモデルの両方を連動させた形式で暗号化する[1]。最初に、テスト画像の暗号化法の概要を以下に示す。

(a) 画像 x を N個の隣接するブロック（ブロックサイズ $B\times B$）に分割する。

(b) 整数 1 から N を要素とする長さ N の乱数列を生成して、その乱数列に従いブロックの位置をランダムに入れ換える。

(c)サイズ $L\times L$ のランダム直交行列を生成して、それを用いて各ブロック内の $L=B\times B$ 個の画素値を変換する。

ここで、暗号化の鍵は、(b)及び(c)で使用されるランダムな整数列と行列である。それらの生成には幾つかの方法がある。図２に暗号化された画像例を示す。上述のブロック分割に基づく暗号化法によって、画像の視覚情報が保護されているのが確認される。

暗号化画像を用いたプライバシー保護を考慮した深層学習

Privacy-Preserving Deep Learning with Encrypted Images

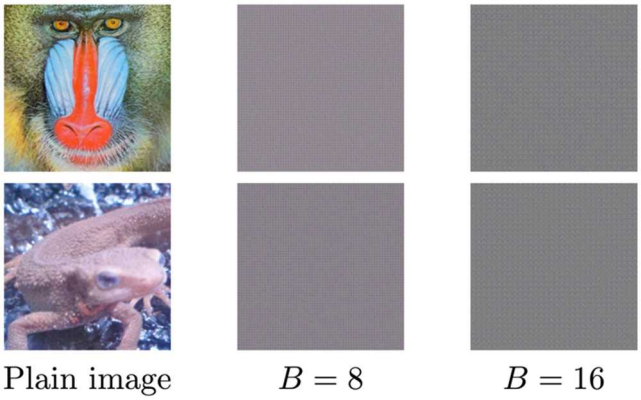


図２　画像の暗号化例

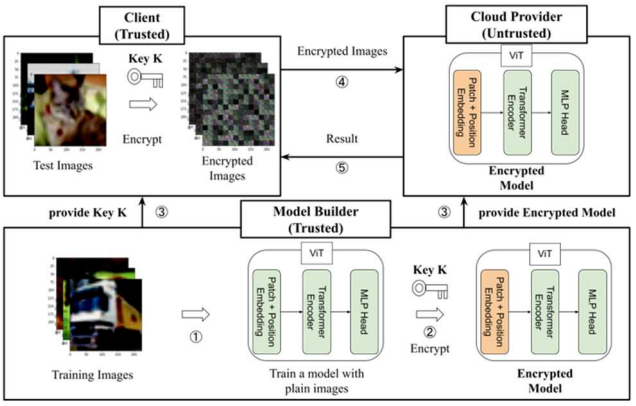


図３　提案法の概要

３．３　モデルの暗号化

図２に示すように、モデルの暗号化は、モデルを学習した後、テスト画像の暗号化で使用される鍵と連動するように実行される[1]。

ViT は、パッチ[*1]という単位でデータをブロック単位で処理し、二つの埋め込み構造（patch embedding, position embedding）を持つという特徴を有する。このパッチ処理と埋め込み構造は、パッチサイズと暗号化のためのブロックサイズを一致させたとき、先のブロック分割に基づく画像暗号化と高い親和性を持ち、本研究において重要な役割を果たす。

ViT モデルは、図３に示すように、三つの処置部からなる。その処理部の一つの埋め込みにランダム線形変換を施すことによって、モデルを暗号化する。モデルの暗号化の目的は二つである。第一は、モデルを暗号化されたテスト画像に対応させるためであり、他はモデルから学習使用した画像情報を復元されるのを防御するためである。詳細は割愛するが、テスト画像の暗号化の手順(b)及び(c)を用いて、パッチと位置情報をそれぞれ変換する。したがって、モデル暗号化の鍵は、テスト画像の暗号化のための鍵と共通である。

モデルの学習とモデルの暗号化は独立の処理である。したがって、モデルの再学習の処理を鍵の更新のために必要としない。さらに、一つのモデルから複数のクライアントに異なる鍵で暗号化されたモデルを提供することも容易である。

３．４　実験結果

CIFAR-10 データセットを用いて 10 クラスの画像分類実験を実行し、提案法の有効性を確認した。CIFAR-10 は60,000枚の画像からなり、モデル学習用に50,000枚、テスト用に 10,000 枚が使用される。また各クラスは 6,000 枚である。また、CIFAR-10 の各画像は、サイズ 3×32×32 であるが、ViT に入力するために 3×224×224 にリサイズして使用された。ViT のパッチサイズと暗号化のためのブロックサイズは、共に 16×16 である。

表１に実験結果を示す。ここで、ベースラインとは、暗号化せずに画像分類を実行した結果である。表から、提案法は、ベースラインと同じ分類精度を達成できていることがわかる。

参考までに暗号化画像を用いた先行研究の結果と比較している。先行研究の方法では、暗号化の影響が回避できず、ベースラインと比べ精度の低下が生じる。一方、提案法は暗号化の影響を回避可能であり、かつ最先端高性能モデルの ViT に適用可能であるため、より高い分類精度が達成できている（従来法は ResNet-20 を使用）。

暗号化画像を用いたプライバシー保護を考慮した深層学習

Privacy-Preserving Deep Learning with Encrypted Images

表１　画像分類精度の評価結果

モデル	暗号化手法	分類精度
ViT	ベースライン	99.03
	提案手法	**99.03**
ResNet-20 [5]	ベースライン	91.55
	Tanaka [6]	87.02
	Pixel-based [7]	86.66
	GAN-based [8]	82.55

４．将来展望

DNN のプライバシー保護を目的とする本研究は、図 1(b)のシナリオのもとで進められた。同図(a)及び(b)以外にも、DNN のさらなる発展のために解決されるべき重要なシナリオが存在する。その一つが連合学習である［10, 11］。

高性能な DNN モデルの学習には、充分な学習データが必須となる。一方、データの多くには個人情報や公表できない情報を含むため、データの確保は容易ではない。特に、複数の異なる機関からデータを集めたい場合には、問題は複雑であり、連合学習がその解決策の一つして期待されている。しかし、幾つかの未解決な急務な課題がある。今回の暗号化画像に関する研究で得られた知見をさらに発展することによって、連合学習の課題解決に応用できると考えている。

おわりに

本研究では、画像分類問題を例にして、画像の視覚情報を保護した形式で実行可能な DNN の開発を目指した。ViT の埋め込み構造が、画像のブロック分割に基づく暗号化と高い親和性を持つことを発見して、それ課題解決に応用した。暗号化鍵を持たない攻撃者は、暗号化画像やモデルから視覚情報の復元が困難である。さらに、暗号化画像を使用した場合も、モデル本来の性能の低下がなく、かつ鍵の更新が容易であるという特徴を開発した手法は持つ。

用語解説

*1 機械学習では、最適なパラメータを発見するためにも、関数の最小値を探す「勾配降下法」と呼ばれる手法が広く使われる。勾配降下法では、一般に学習するデータセットをパッチと呼ばれるいくつかのグループに分ける。

参考文献

[1] H. Kiya, A. P. M. Maung, Y. Kinoshita, S. Imaizumi, and S. Shiota, “An overview of compressible and learnable image transformation with secret key and its applications,” APSIPA Transactions on Signal and Information Processing, vol. 11, no. 1, e11, 2022.

[2] T. Chuman, W. Sirichotedumrong, and H. Kiya, “Encryption-then-compression systems using grayscale-based image encryption for jpeg images,” IEEE Transactions on Information Forensics and Security, vol. 14, no. 6, pp. 1515–1525, 2019.

・・・・・・・・・・・・・・・・・

[3] H. Kiya, R. Iijima, A. P. M. Maung, and Y. Kinoshita, “Image and model transformation with secret key for vision transformer,” IEICE transactions on Information and Systems, vol.106, no.1, pp.2-11, 2023.

[4] A. Dosovitskiy et al., “An image is worth 16x16 words:Transformers for image recognition at scale,” in 9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7. Open Review.net, 2021.

[5] H. Ito, Y. Kinoshita, M. Aprilpyone, and H. Kiya, “Image to perturbation: An image transformation network for generating visually protected images for privacy-preserving deep neural networks,” IEEE Access, vol. 9, pp. 64629–64638, 2021.

[6] H. Kiya, R. Iijima, A. P. M. Maung, and Y. Kinoshita, “Image and model transformation with secret key for vision transformer,” IEICE

transactions on information and systems, vol.106, no.1, pp.2-11, 2023.
[7] M. Tanaka, “Learnable image encryption,” in 2018 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), pp. 1–2, 2018.
[8] W. Sirichotedumrong, T. Maekawa, Y. Kinoshita, and H. Kiya, “Privacy-preserving deep neural networks with pixel-based image encryption considering data augmentation in the encrypted domain,” in 2019 IEEE International Conference on Image Processing (ICIP), pp. 674–678, 2019.
[9] W. Sirichotedumrong and H. Kiya, “A ganbased image transformation scheme for privacy preserving deep neural networks,” in 2020 28th European Signal Processing Conference (EUSIPCO), pp. 745–749, 2021.
[10] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, “Federated learning: Strategies for improving communication efficiency,” arXiv, 2016. [Online]. Available: https://arxiv.org/abs/1610.05492
[11] T. Nagamori, and H. Kiya,” Combined Use of Federated Learning and Image Encryption for Privacy-Preserving Image Classification with Vision Transformer,” arXiv 2023. [Online]. Available: https://arxiv.org/abs/2301.09255

この研究は、令和元年度ＳＣＡＴ研究助成の対象として採用され、令和２～４年度に実施されたものです。