Analysis and Evaluation of Secure and Efficient Medical Image Encryption using Deep Learning Methods

DS 7406 Machine Learning Systems
Joseph Choi
Karolina Naranjo-Velasco
Midterm report

1. Introduction

Computer vision and image processing have been going through revolutionary development with the help of deep learning methods since 2012, with the introduction of AlexNet [1] for classifying/localizing objects in an image from the ImageNet [2] data set, which contains millions of images with 1,000 object labels. The computer vision community has been observing fast developing techniques of deep learning methods every year which outperforms previous year's state-of-the-art methods for the ImageNet task since 2012 [2] and reaching near human-level classification performance with an error rate of less than 5%. The deep learning methods have been applied and expanded in many other fields, including medical fields [3], material science [4], and others.

The promising performance of deep learning methods for the medical application resulted in more than thousands of communication over the network (e.g., sharing data for a collaborative effort to achieve a richer data set [5], sharing promising models to each other, among others, on the Internet of Medical Things [6]. For example, federated learning [5] is a deep learning scheme where it is based on a collaborative effort among multiple parties/institutions to train a deep learning model on a richer and more diverse data set. Federated learning usually requires sharing the data [5] or training information (e.g., gradients) over the network.

Despite the promising performance of deep learning methods in the application of medical fields, the privacy of the patients' sensitive information is at high risk as the data becomes increasingly demanding. Even though there are regulations/standards for the privacy of patient's information, such as the Health Insurance Portability and Accountability Act (HIPAA) [7] and the AI bill of rights [8], the privacy of patients' information is still at risk. Stata reported that over a million of new patients' sensitive information was breached every year [9].

The community also realizes the importance of the security of patients' information when applying deep learning methods. We have noticed two general trends in securing patients' information. First, there are a series of efforts to develop methods or the architecture of the deep learning models to train directly from the encrypted images. Thus, "no one" needs to see the decrypted data, including the "machine". Second, some efforts to encrypt medical images by applying deep learning methods. In this paper, we focus on evaluating and analyzing the security and efficiency of deep learning methods for medical image encryption by analyzing the trends and comparing the performance of the landmark methods on the benchmark dataset (section 2) with some standard metrics (section 4).

2. Related Works:

2.1. Medical Image Encryption (encryption algorithms, hardware-level encryption)

2.2. ML/DL in data privacy

2.3. State our scope (image encryption by DL only)

2. Data:

The deep learning methods are very sensitive to the parameter initialization and hyperparameter setup. Thus, it is sometimes difficult to reproduce the result. Also, the result of proposed deep learning methods might vary from one data set to another. Thus, the benchmark dataset with benchmark metrics is very important to fairly compare and evaluate different deep learning methods. In the paper, the NIH Chest X-Ray dataset [10] and BraTS MRI data set [11] has been used for the benchmark data set.

2.1. NIH Chest X-Ray dataset

108,948 de-identified frontal-view images of chest X-rays of 32,717 unique patients with the tex-mined eight disease labels. This dataset comprises 112,120, 1024 x 1024 pixel frontal-view X-ray images with disease labels and "No finding" labels.

2.2. BraTS MRI dataset

It contains medical image data of 65 multi-contrast MR scans from glioma patients, including high-grade and low-grade gliomas. The data for each patient includes four MRI image modalities:
a) T1: native image, axial 2D acquisitions, 1-6 mm slice thickness
b) T1c: contrast-enhanced, 3D acquisition, 1mm isotropic voxel size.
c) T2: axial 2D acquisition, 2-6mmm slice thickness
d) FLAIR: axial, coronal, sagittal 2D acquisitions, 2-6 mm slice thickness.

The labeled tumors are divided into three nested regions:

- an enhanced tumor region (ET)
- a region composed of enhanced tumor and necrosis (TC)
- an entire region composed of all tumor tissues (WT)

3. Deep learning to encrypt medical image data

The Advanced Encryption Standards (AES) [16] is a U.S. federal standard for encryption algorithms for patient information. The AES has an outstanding performance in the speed of encryption/decryption and security. However, its performance can be degraded significantly by knowing the general pattern of how the private key has been generated. The encryption algorithms with known forms and the process of algorithms allow an attacker to hack the system.

Deep learning is known for a "black box" model [12] as it contains a series of non-linear transformations of an input image to output (e.g., classification of disease, the likelihood of 5-year

survival, etc.). Each non-linear transformation comprises a series of trainable parameters, and the deep learning models have over 100,000 parameters in common. Thus, the attacker has to figure out the specific architecture of the model (e.g. how deep learning is designed: number of layers, number of poolings, etc.) and specific set of parameter values (which are commonly over 100,000 parameters) of the model to decrypt the encrypted images, which is merely impossible. The tendency of "uncrackable" nature of deep learning methods motivates the use of deep learning methods for image encryption.

The deep generative models are known as strong parametrization capability. For example, the autoencoder (AE) [13] and generative adversarial network (GAN) [14] are most common deep generative models where it learns to map the high-dimensional image data (e.g., $R^{225 \times 225}$) to low-dimensional representation (e.g., R^d) where $d << 225 \times 225$. The use of deep generative models allows efficient representation of the medical image, which could further lead to increased performance of (1) efficiency in data storage and (2) efficiency in data sharing over the network.

The security and efficiency of deep learning methods motivate us to use deep learning methods for image encryption methods. We plan to analyze and evaluate some of the methods that we have identified through literature survey, which are listed below:

- Cycle-GAN [15,20]
- Back-propagation based image encryption [17]
- Invertible encryption network [18]
- Image scrambling via adversarial autoencoders [19]

4. Metrics

The methods in scope are evaluated and compared using benchmark metrics. The encrypted image must be decrypted without much loss of information as the physicians or deep learning models have to utilize image features to perform the diagnosis of patients. Thus, reconstruction metrics are one of the critical metrics to evaluate encryption algorithms for the application of the medical image. Also, the memory efficiency of the encryption is an important property that could lead to increased performance during a transfer of data over the network or efficient data storage.

4.1. Reconstruction

We evaluate the loss of information (or reconstruction error) by average mean squared error (mse) or pixel-wise differences,

$$mse = \frac{1}{N \times M} \sum_i^N \quad \sum_j^M \quad \sqrt{x_{i,j} - x'_{i,j}},$$

where *N* and *M* are the width and height of the image, *i* and *j* are the spatial pixel location of the image, *x* and *x'* are original image and decrypted image, respectively.

4.2. Memory efficiency

The memory efficiency is simply measured by the physical space the encrypted image occupies in kilobyte (kb).

References

[1] A. Krizhevsky, et al., ImageNet classification with deep convolutional neural networks, *In Proceedings of the 25th International Conference on Neural Information Processing Systems* **2012**, *1*, 1097-1105.
[2] J. Deng, et al., ImageNet: A large-scale hierarchical image database, *In 2009 IEEE Conference on Computer*

Vision and Pattern Recognition **2009,** 248-255.
[3] G. Litjens, et al., A survey on deep learning in medical image analysis, *Medical Image Analysis* **2017**, *42*, 60-88.
[4] K. Choudhary, et al., Recent advances and applications of deep learning methods in materials science, *npj Computational Materials* **2022**, *8.1*, 1-26.
[5] J. Konecny, et al., Federated learning: strategies for improving communication efficiency, *arXiv* **2016**, arXiv:1610.05492.
[6] S. Vishnu, et al., Internet of medical things (IoMT) - An overview, *2020 5th International Conference on Devices, Circuits and Systems* **2020**, 101-104.
[7] Health Insurance Portability and Accountability Act. Pub. L. No .104-191, § 264, 110 Stat. 1936.
[8] Office of Science and Technology Policy (**2022),** Blueprint for an AI bill of rights, *The White House*, https://www.whitehouse.gov/ostp/ai-bill-of-rights/
[9] Statistica, Number of U.S. residents affected by health data breaches from 2014 to 2020, *Statista* **2020**, https://www.statista.com/statistics/798564/number-of-us-residents-affected-by-data-breaches/
[10] X. Wang, et al., ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* **2017**, 2097-2106. https://www.nih.gov/news-events/news-releases/nih-clinical-center-provides-one-largest-publicly-available-chest-x-ray-datasets-scientific-community
[11] B. H. Menze, et al., The multimodal brain tumor image segmentation benchmark (BRATS), *IEEE Transactions on Medical Imaging* **2015**, *34*, 10, 1993-2024.
[12] I. Goodfellow, et al., Deep learning, MIT Press, **2016**
[13] J. Schmidhuber, et al., Deep learning in neural networks: An overview, *Neural Networks* **2015**, *61*, 85-117.
[14] I. Goodfellow, et al., Generative adversarial networks, *Communications of the ACM* **2020**, *63.11*, 139-144.
[15] Y. Ding, et al., DeepEDN: A deep learning-based image encryption and decryption network for internet of medical things, *IEEE Internet of Things Journal **2022**, 8*, 3, 1504-1518.
[16] F. P. Miller, et al. *Advanced Encryption Standard*, **2009**, *Alpha Press*.
[17] Y. Gao, et al. An improved image processing based on deep learning backpropagation technique, *Complexity* **2022,** *2022*.
[18] F. Wang, et al., Invertible encryption network for optical image cryptosystem, *Optics and Lasers in Engineering* **2022**, *149*, 106784
[19] Z. Bao, et al. Image scrambling adversarial autoencoder based on the asymmetric encryption, *Multimedia Tools and Applications* **2021**, *80*, 28265-28301.
[20] Z. Bao, et al., Research on the avalanche effect of image encryption based on the Cycle-GAN, *Applied Optics* **2021**, *18*, 5320-5334.