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

Joseph B Choi[1, *], Karolina Naranjo-Velasco[1, *]

1 School of Data Science, University of Virginia, Charlottesville, United States.

* Authors contributed equally

***Abstract*— Over 2 million of patient's private information is leaked every year and the emphasis has put more than ever on secure system for medical data. Deep learning's nature of being a "black-box" model and its outbreaking encoding performance motivates to apply deep neural networks for the medical image encryption-decryption task. In this paper, the trends of applying deep learning methods for the encryption-decryption task is discussed along with suggestion of the benchmark data set and metrics.**

Keywords—Deep learning, medical image encryption-decryption, generative adversarial network, data privacy.

I. INTRODUCTION

Computer vision and image processing have been going through revolutionary development with the help of deep learning (DL) 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 DL methods which outperforms previous year's state-of-the-art methods for the ImageNet task since 2012 [2] and reaches near human-level classification performance with an error rate of less than 5%. Therefore, DL methods have been applied and expanded in many other fields, including medical fields [3], material science [4], and others.

The promising performance of DL 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 based on a collaborative effort among multiple parties/institutions to train a DL 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 DL methods in the medical field, the privacy of the patient's sensitive information is at high risk as the data becomes increasingly demanding. Even though there are regulations/standards for data privacy, such as the Health Insurance Portability and Accountability Act (HIPAA) [7] and the AI bill of rights [8], the privacy of patients' data is still at risk. Stata reported that over a million new patients' sensitive information was breached yearly [9].

The community also realizes the importance of protecting patients' information when applying DL methods. We have noticed two general trends in securing patients' information. First, there are a series of efforts to develop methods or DL architectures to train directly from the encrypted images. Thus, "no one" needs to see the decrypted data, including the "machine." Second, medical encryption on medical images with DL methods has been prominent in the literature. This paper evaluates and analyzes the security and efficiency of DL methods for medical image encryption. We will identify the trends and compare the landmark methods' performance on the benchmark dataset (*section III*) with some essential metrics (*section V*).

II. RELATED WORKS

A. Image Encryption

Digital image security seeks to protect users' sensitive data, such as medical images. Some scholars have explained how deep learning (DL) applications are an alternative to digital image encryption [10]. Three traditional types of algorithms of image steganography, image encryption, and image authentication have been applied with deep learning methods [11]. Image steganography consists of distinguishing "innocent" images that attackers send; convolutional neural network (CNN) classifiers are crucial in detecting this type of image. This method, in turn, has the characteristic of higher undetectability. Image authentication is oriented to check the image identity or integrity through, for example, image watermark methods.

Image encryption—the scope of this survey—consists of a process of encrypting and decrypting a plain text image through a key. It aims to maintain the privacy of the images until the correct key is found. Authors such as Bao and Xue [11] and Huang, Yap, Chiu, and Sun [12] have pointed out how this method has become relevant in deep learning applications.

According to some scholars, there are at least six types of image encryption: image compression, resolution

improvements, detection and classification, image key generation, end-to-end encryption, and cryptanalysis. Image compression is characterized by compressing and reconstructing the plaintext image; autoencoders for compression are used in this method. Image resolution uses the ghost imaging technique to transmit and encrypt images using CNN. Next, image key generation uses a Generative adversarial Network (GAN) to generate a private key.

Finally, end-to-end image encryption and cryptanalysis have become prominent in the literature because they use more advanced methods for privacy attacks. On the one hand, deep learning methods of end-to-end image encryption apply Cycle-GANs for optimal results. On the other hand, cryptanalysis uses transforms, such as Fourier, to convert the plaintext image into ciphertext and encrypted image, making the images less exposed to attackers.

B. ML/DL in data privacy/security

Machine Learning (ML) models are designed to learn data characteristics, especially in large datasets, and perform tasks that are difficult to perform by hand. As an example of ML, generative models transform each given layer of input data to create a more complex representation [13]. Since the execution of more complex tasks drives ML, Machine Learning as a Service (MLaaS) is an alternative to facilitate this type of outsourcing of complex tasks. However, this increases the risks of information leakage and, thus, exposes users' privacy.

In generative models, for instance, Generative Adversarial Networks (GAN) are used to learn information and create a copy that can potentially be used by attackers [13].

In order to protect the information of MLaaS users, image encryption becomes essential. First, encryption techniques ensure that access to data is key-protected. Second, it guarantees that the generative models of the attackers' tasks can be decrypted with DL technologies.

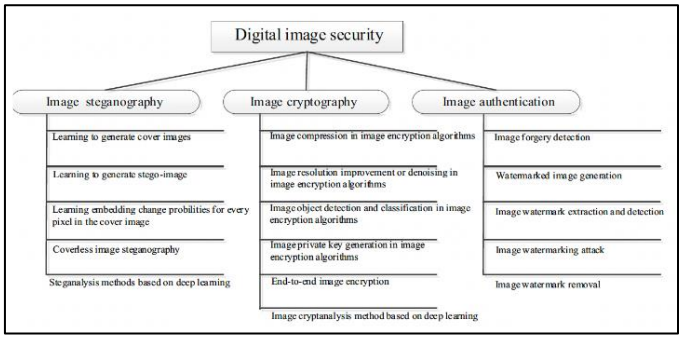

Figure 1. Structure of digital image security [11]

C. Scope of the Survey

We limit our scope on end-to-end image encryption using deep learning methods as there are many other surveys that reviews general methods for image cryptography (Figure 1). We believe this is the first review paper discussing and evaluating the trends of deep learning methods for the medical image encryption-decryption task with the best of our knowledge. There are lines of research to make neural network models trainable on the encrypted images, but this is out of our scope as we are interested in how deep learning methods are applied for image encryption. In a similar vein, there are attempts to use deep learning methods as a component of conventional non-deep learning methods (e.g., key generation by deep learning), which are also outside of the scope of this survey.

III. DATA SET

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 [14] and BraTS MRI data set [15] has been used for the benchmark data set.

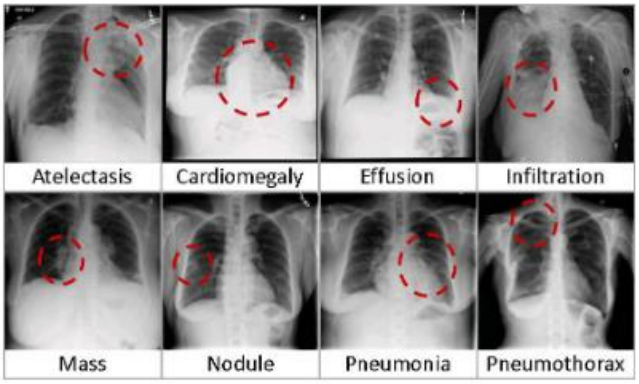

Figure 2. Sample of NIH Chest X-Ray data set [14]

A. NIH Chest X-Ray data set

108,948 de-identified frontal-view images of chest X-rays of 32,717 unique patients with the text-mined eight disease labels (Figure 2). This dataset comprises 112,120, 1024 x 1024-pixel frontal-view X-ray images with disease labels and “No finding” labels.

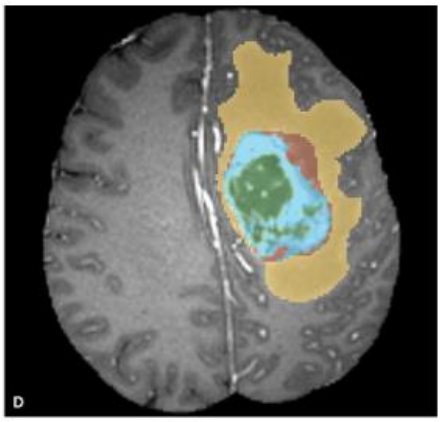

Figure 3. slice of BraTS MRI data set. (blue) enhancing tumor. (green) cystic/necrotic core. (yellow) edema. (red) non-enhancing solid core [15]

B. *BraTS MRI data set*

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 (Figure 3).

- Enhanced tumor region (ET)
- Region composed of enhanced tumor and necrosis (TC)
- Entire region composed of all tumor tissues (WT)

IV. 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 [17] 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 hidden layers, number of pooling layers, 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 security and efficiency of deep learning methods motivate us to use deep learning methods for image encryption methods. The deep learning-based image encryption-decryption is new field, and there are two mainstreams on how to apply deep learning methods for the image encryption-decryption tasks, namely deep generative models (*section IV.A*) and deep neural networks with articulated loss functions (*section IV.B*). The general concept and idea of how each of the methods are applied for the image encryption-decryption task is applied are summarized.

A. *Deep Generative Models*

The deep generative models are known as strong parametrization capability. For example, the autoencoder (AE) [18] and generative adversarial network (GAN) [19] are most common deep generative models where it learns to map the high-dimensional image data (e.g., $\mathbb{R}^{225}x\mathbb{R}^{225}$) to low-dimensional representation (e.g., $\mathbb{R}^{d}$, $d \ll 225 \times 225$). 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.

However, direct application of the deep generative model does not give a valid encryption-decryption algorithms due to the nature of the generative model. The intended goal for deep generative models is to learn the underlying distribution that can generate samples which is alike to the training data set. Many generative models allow a stochasticity in the generation process to allow diversity in the generation process. However, this phenomenon is not desirable when deep generative models are used for the encryption-decryption algorithm. The decrypted image has to be the same as the plain-text instead of many similar looking plain-text images. Removing the stochasticity of the deep generative models are not trivial as the concept of the stochasticity is part of the foundation to the deep generative models. Thus, it is not straightforward to naively apply existing deep generative models for the encryption-decryption task.

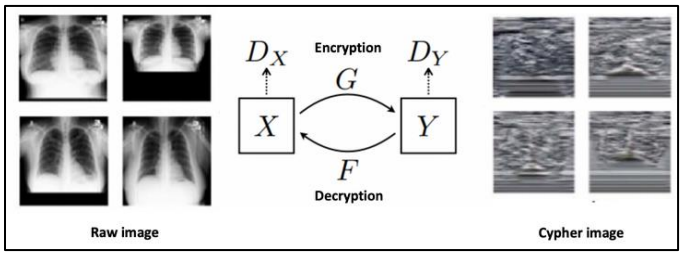

Figure 4. Concept of how Cycle GAN works. G and F are generators acts like a encryption and decryption model, respectively.

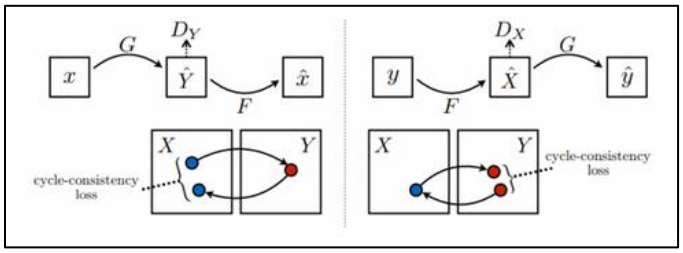

Figure 5. Schematic diagram of cycle-consistency loss [20].

The Cycle Generative Adversarial Network (CycleGAN) [20] is a specific type of the deep generative model that can be directly applied for the image encryption-decryption task. The CycleGAN was first introduced in 2017 for the image-to-image translation, which means the neural network model learns a bijective mapping between two spaces (target domain and input domain) with two generators and discriminators with the conventional adversarial loss (*equation 2*) and cycle consistency loss (*equation 3*), as described below and depicted in Figure 5:

$$L(G, F, D_x, D_y) = L_{adv}(G, D_y, X, Y) + L_{adv}(F, D_x, Y, X) + \lambda L_{cyc}(G, F, X, Y) \quad (1)$$

$$L_{adv}(G, D_y, X, Y) = \frac{1}{m}\sum(1 - D_y(G(x))^2 \quad (2)$$

$$L_{cyc}(G,F,X,Y) = \frac{1}{m}\left[\left(F\left(G(x_i)\right) - x_i\right) + \left(G\left(F(y_i)\right) - y_i\right) \right] \quad (3)$$

, where G (and F) is generators from $S_{plain-text} \rightarrow S_{cipher-text}$ ($S_{cipher-text} \rightarrow S_{plain-text}$), X and Y are image samples from plain-text and cipher-text, respectively, D_y and D_x are discriminators for the cipher-text and plain-text, respectively (Figure 4).

The DeepEDN proposed by Ding et al. [21] is the very first application of the CycleGAN for the medical image encryption-decryption task. The DeepEDN used the CycleGAN as it is from the reference paper [ref], but still achieved significant performance (Table 1). For example, DeepEDN enhanced the execution efficiency by about 3 times compared to state-of-the-art non-deep learning method. However, despite great performance of the CycleGAN method for the image encryption-decryption task, Bao et al. [22] pointed out a weak avalanche effect of the CycleGAN. The avalanche effect means a small change in the plain-text causes a drastic change in the cipher-text. Such a weak avalanche effect is problematic for encryption-decryption task as the correlation between the plain-text and cipher-text must be minimized. Thus, Bao et al. [22] combined CycleGAN and conventional encryption method by applying diffusion and confusion process on the cipher-text image to achieve strong avalanche effect while maintaining the performance of the DeepEDN.

On the other hand, Bao et al [23] tried to replace existing encryption concept called image scrambling using a method called adversarial autoencoder. The autoencoder contains encoder (maps $S_{plain-text} \rightarrow S_{scrambled}$) and decoder (maps $S_{scrambled} \rightarrow S_{plain-text}$). Bao et al. [23] used adversarial loss (similar to *equation 2*) with reconstruction loss (e.g., rmse). We believe the adversarial autoencoder method would have less hostility as it mimics the conventional scrambling method but does not enhance the performance over the CycleGAN method.

B. Deep Neural Networks with articulated loss fuctions

The CycleGAN method for the encryption-decryption task is great as it outperforms state-of-the-art non-Deep Learning methods as a secure and efficient system. However, there are still too many sophisticated deep learning methods proven to be working well for other computer vision task. Limiting ourselves to only using CycleGAN might lose potential opportunity of using other deep learning methods for improved performance. There are lines of research by utilizing promising neural network architectures by articulating the loss function for the encryption-decryption task. Such approach for applying deep neural network usually contains two neural network models (called encryption and decryption neural network models) and the loss function is commonly different from each other.

The FEDResNet by Zhu et al. [24] is one of the very first works where they approach for the encryption-decryption task. The FEDResNet uses popular ResNet [25] as the baseline architecture, includes many tricks to make their network more robust and stable. For example, short-cut connection from ResNet connection and long-range skip connection (similar to the U-Net skip connection [26]) have been applied to make the training process easier. Also, the dilated convolution [27] has been applied to make the receptive field larger. The loss function is articulated by minimizing the information of the cipher-text and also minimizing the information loss during encryption-decryption process as follow:

$$Loss = L_{en} + L_{dc} \quad (4)$$

$$L_{en} = \sum_{i=1}^{N} \frac{[p_i \log p_i + (1-p_i)\log(1-p_i)]}{N} \quad (5)$$

$$L_{dc} = MSE(x,y) + \alpha[1 - SSIM(x,y)] + \beta CSD(x,y) \quad (6)$$

$$CSD(a,b) = \sum_{i=1}^{255} \frac{(a_i - b_i)^2}{a_i + b_i + \tau} \quad (7)$$

, where p_i is pixel intensity at location i, N is the total number of pixels in the image, α and β are weighting coefficients, $MSE(\cdot)$ and $SSIM(\cdot)$ are expressed in *equation 9* and *equation 10*, a and b are the value of the ith bins of image histograms, and τ is a positive constant to avoid numerical error. The FEDResNet adds additional layer of security by creating additional keys by using conventional method called chaotic sequence generation.

In a similar vein, Wang et al. [28] enhanced the conventional chaotic sequence generation algorithm by remove the pattern of the sequence in the ciphertext by further encoding with the neural network, specifically VNet [29]. Applying a VNet to remove the pattern of the chaotic sequence in the cipher-text is straightforward, but we cannot ignore the benefit it brings by making the one of the most popular encryption algorithms even stronger.

Wang et al. [30] applied single invertible neural network for the encryption-decryption task instead of two different neural networks (e.g., encryption and decryption neural networks). The invertible neural networks train its model to map plain-text image to pseudo-random cipher-text images using reconstruction error (e.g., rmse). Intuitively speaking, the invertible neural networks have a switch for forward pass ($S_{plain-text} \rightarrow S_{cipher-text}$) and inverse pass ($S_{cipher-text} \rightarrow S_{plain-text}$) once the model has been trained. Thus, the forward pass operates as a encryption network and inverse pass operates as a decryption network. Although the authors did not compare the performance with non-invertible neural networks, but it is often more stable to train a single model compared to train multiple models at once.

V. METRICS

The deep learning methods in scope are compared evaluated and compared using common metrics for evaluating the encryption methods. It was very challenging to come up with such metrics as there doesn't seem to have any benchmark metrics to evaluate the deep learning encryption methods. Thus, the list of the metrics has been collected that is believed to be the benchmark to evaluate the deep learning-based encryption methods, such as ciphertext security measures (*section V.A*), information loss (*section V.B*), efficiency (*section V.C*), and key security analysis (*section V.D*).

A. Ciphertext security measures

Encrypted image should be resistant to any statistical attack and should hide as much information as possible in gaining any statistical measure from the encrypted images that might be able to correlate to the plain-text images.

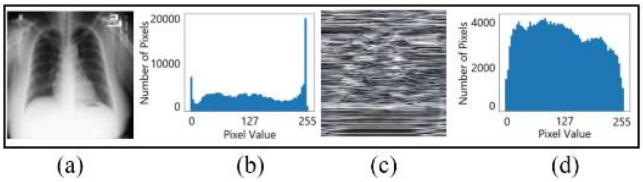

Figure 6. Histogram analysis. (a) plain-text. (b) histogram of plain-text. (c) cipher-text. (d) histogram of cipher-text [21].

The histogram analysis is a qualitative measure to compare the histogram of the plain-text and cipher-text images. Good, encrypted images should not reveal any information of the plain-text through statistical measure from the cipher-text image. For example, figure 6 compares the histogram of the plain-text and cipher-text, and no statistical measure from the cipher-text can be correlated to the plain-text image.

The information entropy (*equation 8*) is an effective measure to quantify how algorithm is resilient to the statistical attacks. The information entropy is a statistical measure that quantifies to what extent the information is spread out over the possible values. Thus, in the scope of the measuring the effectiveness of the encryption algorithms, high entropy measure is idea as it means the image data is similar to the random noise, which indicates extracting a statical measure to correlate to the plain-text image is impossible.

$$entropy = -\sum_{l=0}^{N} p(l) \cdot \log_2 p(l) \tag{8}$$

where N is the number of pixel intensities and $p(l)$ is the probability that intensity l appears in the image. Thus, 8.00 is the maximal entropy value for the *8-bit* (256 intensity value) grayscale image.

B. Information loss

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.

The root mean squared error is very common choice to measure the differences between the two images to evaluate the loss of information (or reconstruction error) by comparing the pixel-level differences.

$$rmse = \frac{1}{n \times M} \sum_{i}^{N} \sum_{j}^{M} \sqrt{x_{i,j} - x'_{i,j}} \tag{9}$$

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.

The structural similarity index measure (SSIM) to measure the structural differences between two images. The SSIM is a perception-based measure that stems on the idea where nearby pixels have strong interdependencies whereas the RMSE measures the absolute error. The SSIM is expressed as,

$$SSIM(x,y) = l(x,y)^{\alpha} \cdot c(x,y)^{\beta} \cdot s(x,y)^{\gamma} \tag{10}$$

$$l(x,y) = \frac{2 \cdot \mu_x \cdot \mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \tag{11}$$

$$c(x,y) = \frac{2 \cdot \sigma_x \cdot \sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \tag{12}$$

$$s(x,y) = \frac{\sigma_{xy} + c_3}{\sigma_x \cdot \sigma_y + c_3} \tag{13}$$

where $l(\cdot), c(\cdot)$, and $s(\cdot)$ measures luminance, contrast, and structure, respectively; μ, and σ is the pixel mean and variance, respectively; c_1 and c_2 are small constants to avoid a numerical error. The SSIM measures in the range between 0 and 1, where 1 indicates two images are perfectly matching and 0 indicates two images does not have any similarity.

C. Efficiency

The encryption-decryption algorithms are commonly applied as a real-time operation. Thus, the amount of time required (*execution efficiency)* to decrypt and encrypt the plain-text image should be measured as part of the benchmark metrics. In addition, the physical size of the cipher-text (*storage efficiency)* should also be measured as the size of the cipher-test is directly related on how much information can be transferred over the network; for example, smaller size of the cipher-text would be transferred to another party over the network much faster compared to the larger size of the cipher-text.

The execution efficiency should measure how many cases an algorithm can encrypt within one second. On the other hand, the storage efficiency could be measured in how much Kilobytes (KB) does the cipher-text contains.

D. Key security analysis

The private key is a critical part to analyze how secure a system is. There are too many measures to analyze the key security, but we chose some of the very essential and baseline measures of the key security.

The brute-force attack is the most rudimentary attack to the secure system where it tries every combination of the answers and cracks the system. It is commonly known that the key space has to be at least 2^{100} in order to be robust against the brute-force attacks.

Two different cipher-text images encrypted with different private key should produce different cipher-text to ensure the encryption algorithm is sensitive to the private-key. Thus, it is a common practice to use SSIM (*equation 10*) to measure the structural similarity between two different cipher-text images; we should expect low SSIM measure if the encryption algorithm is sensitive to the private-key.

VI. Evaluate deep learning encryption methods

The evaluation of different deep learning-based encryption-decryption methods are summarized in the Table 1 based on the metrics discussed in *section V*. We should note that the comparison is not an "apple-to-apple" comparison as the authors did not use a specific benchmark data set. Moreover, not any one of the reviewed literatures provide a public repository for their methods. Thus, our evaluation was very limited to only reflect on selection of metrics on selection of data set that each of the authors decide to include on their papers. In addition, for those of the paper that use the same data set, it is highly likely that two different papers do not compare their result on the same test data set. It is even uncertain if the authors kept set of images for the test data set or not. Thus, the comparing different deep learning methods for encryption-decryption tasks based on the Table 1 might be biased.

All of the deep learning methods satisfy more than enough for the ciphertext security measure with the uniform histogram and information entropy with 7.91, which indicates cipher-text is alike random noise. Generative based approach tends to lose more information than deep neural network with articulated loss function methods (e.g., SSIM of 0.93 for generative models in average compared to SSIM of 0.97 for FEDResNet). However, the generative-based models tend to be much faster to encode and decode compared to the articulation of loss function methods (e.g., 14.28 images per second for generative models versus 8.26 images per second for loss function methods), but still outperforms non-deep learning methods (5 images per second for the state-of-the-art non-deep learning method). The private key is very secure for all of the deep learning methods as the key space is easily over million scales, which is far beyond the minimal requirement of 2^{100}.

VII. Discussion and conclusion

The deep learning methods for the (medical) image encryption-decryption task are a very recent effort (starting from 2020) and a continuously growing field. The powerful encoding performance and "black-box" nature of the deep learning methods motivate the community for (medical) image encryption-decryption tasks where digital data security is critical, such as patients' private information.

Even though there is only a handful of research to review for our scope, we successfully reviewed and identified two mainstreams of the research with deep generative methods and deep neural networks with articulated loss functions.

The most critical research gap for the deep learning-based encryption-decryption task is the absence of benchmark data sets and metrics. As discussed in *section VI* and Table 1, each of the papers we reviewed selects its own data set and metrics. In order to fill in this research gap, we recommended benchmark data set in *section III* and a set of benchmark metrics that is essential as a baseline in *section V*. On top of the benchmark data set, the community should set test data sets apart to promote fair comparison among different deep learning methods.

Medical imaging is unlike any other natural image as it is directly related to the patient's health. For example, we should avoid having physicians misdiagnose patients' diseases due to information loss during the encryption-decryption process. Thus, we should emphasize what the SSIM of 0.9752 means as a quantitative measure. We believe there should be a study to correlate the performance of the physician's diagnosis based on different SSIM measures.

VIII. Future works

We have not selected the where to submit it yet, but we plan to submit it for poster or workshop paper.

Paper	Method	Data set	Ciphertext security		Information loss		Efficiency		Key Security	
			histogram analysis	information entropy	RMSE	SSIM	Execution efficiency	Storage efficiency	Key space	similarity of encrypted
Ding et al. [ref]	CycleGAN	NIH ChesXRay, BraTS	good	7.96	NA	0.9	14.28	NA	(2^32)^2M	0.07
Bao et al. [ref]	CycleGAN	Satelite images	good	7.99	NA	0.94	NA	NA	(2^32)^16M + (2^8)^196K	0.0048
Zhu et al. [ref]	FEDResNet	NIH ChesXRay	good	7.99	NA	0.976	1.34	NA	2^524K	0.009
Want et al. [ref]	V-NET	Lena, MRI slice	good	7.99	NA	NA	NA	NA	10^112	NA
Bao et al. [ref]	adversarial AE	Corel-1000	good	7.91	NA	0.9115	NA	NA	(2^32)^60M	0.0068
Wang et al. [ref]	Invertible NN	natural image	good	7.98	NA	NA	8.26	NA	(2^32)^5M	NA

Table 1. Summary of evaluating deep learning based (medical) image encryption

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. Liu *et al*., "Privacy and Security Issues in Deep Learning: A Survey," in *IEEE Access*, vol. 9, pp. 4566-4593, 2021, doi: 10.1109/ACCESS.2020.3045078.

[11] Bao, Z., and Ru, X. "Survey on deep learning applications in digital image security." *Optical Engineering* 60.12, 2022, https://doi.org/10.1117/1.OE.60.12.120901

[12] Q. -X. Huang, W. L. Yap, M. -Y. Chiu and H. -M. Sun, "Privacy-Preserving Deep Learning With Learnable Image Encryption on Medical Images," in IEEE Access, vol. 10, pp. 66345-66355, 2021, doi: 10.1109/ACCESS.2022.3185206.

[13] E. De Cristofaro, "A Critical Overview of Privacy in Machine Learning," in *IEEE Security & Privacy*, vol. 19, no. 4, pp. 19-27, July-Aug. 2021, doi: 10.1109/MSEC.2021.3076443.

[14] 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

[15] B. H. Menze, et al., The multimodal brain tumor image segmentation benchmark (BRATS), IEEE Transactions on Medical Imaging 2015, 34, 10, 1993-2024.

[16] F. P. Miller, et al. Advanced Encryption Standard, 2009, Alpha Press.

[17] I. Goodfellow, et al., Deep learning, MIT Press, **2016**

[18] D. H. Ballard, Modular learning in neural networks, *Association for the Advancement of Artificial Intelligence* **1987**, *6*, 279-284.

[19] I. Goodfellow, et al., Generative adversarial networks, Communications of the ACM 2020, 63.11, 139-144.

[20] J. Y. Zhu, et al., Unpaired image-to-image translation using cycle-consistent adversarial networks, *Proceedings of the IEEE international conference on computer vision* **2017**, 2223-2232.

[21] 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.

[22] Z. Bao, et al., Research on the avalanche effect of image encryption based on the Cycle-GAN, Applied Optics 2021, 18, 5320-5334.

[23] Z. Bao, et al. Image scrambling adversarial autoencoder based on the asymmetric encryption, Multimedia Tools and Applications 2021, 80, 28265-28301.

[24] L. Zhu, et al., FEDResNet: a flexible image encryption and decryption sheme based on end-to-end iamge diffusion with dilate ResNet, *Applied Optics.* **2022**, 9124-9134.

[25] K. He, et al., Deep residual learning for image recognition, *Proceedings of the IEEE conference on computer vision and pattern recognition* **2016**, 770-778.

[26] O. Ronneberger, et al., U-net: convolutional networks for biomedical image segmentation, *International Conference on Medical image computing and computer-assisted intervention* **2015**, *Springer, Cham*, 234-241.

[27] F. Yu, et al., Multi-scale context aggregation by dilated convolutions, *arXiv* **2015**, *arXiv:1511.07122*

[28] X. Wang, et al., A new V-Net convolutional neural network nased on four-dimensional hyperchaotic system for medical image encryption, *Security and communication networks* **2022**.

[29] F. Milletari, et al., V-net: fully convolutional neural networks for volumetric medical image segmentation, *2016 fourth international conference on 3D vision (3DV)* **2016**, *IEEE*, 565-571.

[30] F. Wang, et al., Invertible encryption network for optical image cryptosystem, Optics and Lasers in Engineering 2022, 149, 106784