DLEDNet: A Deep Learning-based Image Encryption and Decryption Network for Internet of Medical Things

Yi Ding, *Member, IEEE*, Guozheng Wu, Dajiang Chen, *Member, IEEE*, Ning Zhang, *Senior Member, IEEE*, Linpeng Gong, Mingsheng Cao, *Member, IEEE*, and Zhiguang Qin, *Member, IEEE*

***Abstract*—With the rapid development of Internet of Medical Things (IoMT) technology, many medical imaging equipments are connected to the medical information network to facilitate the process of diagnosing and treating for doctors. As medical image contains sensitive information, it is of importance yet very challenging to encrypt the medical image to safeguard the privacy or security of the patient. In this work, a deep learning based encryption and decryption network (DLEDNet) is proposed to fulfill the process of encrypting and decrypting the medical image. Specifically, in DLEDNet, the Cycle-GAN network is employed as the main learning network to transfer the medical image from its original domain to the target domain. Target domain is regarded as a "Hidden Factors" to guide the learning model to realize the encryption process. The encrypted image is restored to the original (plaintext) image through a reconstruction network to achieve an image decryption process. In order to realize the data mining process direct from the privacy-protected environment, a ROI (Region of interest)-mining-network is proposed to extract the interested object directly from the encrypted image. The proposed DLEDNet is evaluated on the chest X-ray dataset. Extensive experimental results show that the proposed method can achieve a high level of security with a good performance in efficiency.**

***Index Terms*—Image encryption, Deep learning, Medical image**

I. Introduction

The Internet of Medical Things (IoMT) is an interdisciplinary filed which adopts the Internet of Things (IoT) technologies in the domain of medicine [1]–[3]. With the development of IoMT, many medical imaging equipments are widely connected and used to facilitate the process of diagnosing and treating for doctors, e.g., the brain magnetic resonance imaging (MRI) for brain tumor diagnosis and the computed tomography (CT) of lung for lung nodule detection. In IoMT, medical images are usually managed by a system called Picture Archiving and Communication Systems (PACS) [4]. When a patient is scanned by the medical imaging equipment, the generated medical images will be firstly stored into the PACS. When the doctor begins to examine the patient, the PACS will retrieve the needed images from the database and transfer the images to the doctors' workstation which works with the patient information from the Hospital Information System (HIS). Although the PACS and HIS works in an intranet environment, there are still some serious security issues for storing, transferring, reviewing medical images, which preserve sensitive privacy information of patients. If an attacker, either an internal or an external attacker, has the ability to intrude the PACS or HIS, it becomes much easy to eavesdrop these medical images, resulting in severe privacy information leak of patients [5]–[7].

To safeguard the IoMT systems and protect the patients' privacy, encryption and decryption approaches can be adopted on medical images, e.g., Data Encryption Standard (DES), Advanced Encryption Standard (AES) and the Hash Function [8], [9]. In addition, image encryption based on chaotic systems are also employed in the literature [10]. However, these methods are hard to achieve a good balance between the security performance and the encryption efficiency. Deep learning also draw great attentions to tackle the problem, which refers to multi-layer neural networks to extract a hierarchy of features from raw input images. The Convolutional Neural Networks (CNNs) [11], [12] has demonstrated the significant advantages in computer vision tasks [13]–[16] as well as in image domain transfer [17], [18]. Transferring the image from one domain onto another can be considered as a problem of texture transfer where the goal is to learn the mapping relationship between an input image and an output image from a set of aligned image pairs. One of the most popular image-to-image translation method is the Cycle-Consistent Adversarial Networks [19], which introduces two cycle consistency losses that transform the image from one domain to the other, and then reconstruct back to the original image. In fact, the deep learning algorithm has also been adopted to solve the problem of image denoising [20]. Image noise refers to the interference information existing in image data, resulting in some useful image information becomes invisible. The image denoising process can be regarded as the image restoration [21].

Inspired by the above works, in this work, a deep learning

Corresponding authors: Dajiang Chen (djchen@uestc.edu.cn) and Ning Zhang (ning.zhang@tamucc.edu)

Yi Ding is with the Network and Data Security Key Laboratory of Sichuan Province, University of Electronic Science and Technology of China, Chengdu, Sichuan, 610054 China; he is also with Institute of Electronic and Information Engineering of UESTC in Guangdong, Dongwan, Guangdong, China (e-mail: yi.ding@uestc.edu.cn).

Dajiang Chen, Linpeng Gong, Mingsheng Cao, and Zhiguang Qin are with the Network and Data Security Key Laboratory of Sichuan Province, University of Electronic Science and Technology of China, Chengdu, Sichuan, 610054 China (e-mail: djchen@uestc.edu.cn; glpglp@std.uestc.edu.cn; cms@uestc.edu.cn; qinzg@uestc.edu.cn).

Guozheng Wu is with National Natural Science Foundation of China, Beijing, China (e-mail:wugz@nsfc.gov.cn).

Ning Zhang is with the Department of Computing Science, Texas A&M University-Corpus Christi, Corpus Christi, TX 78412, USA (e-mail: ning.zhang@tamucc.edu).

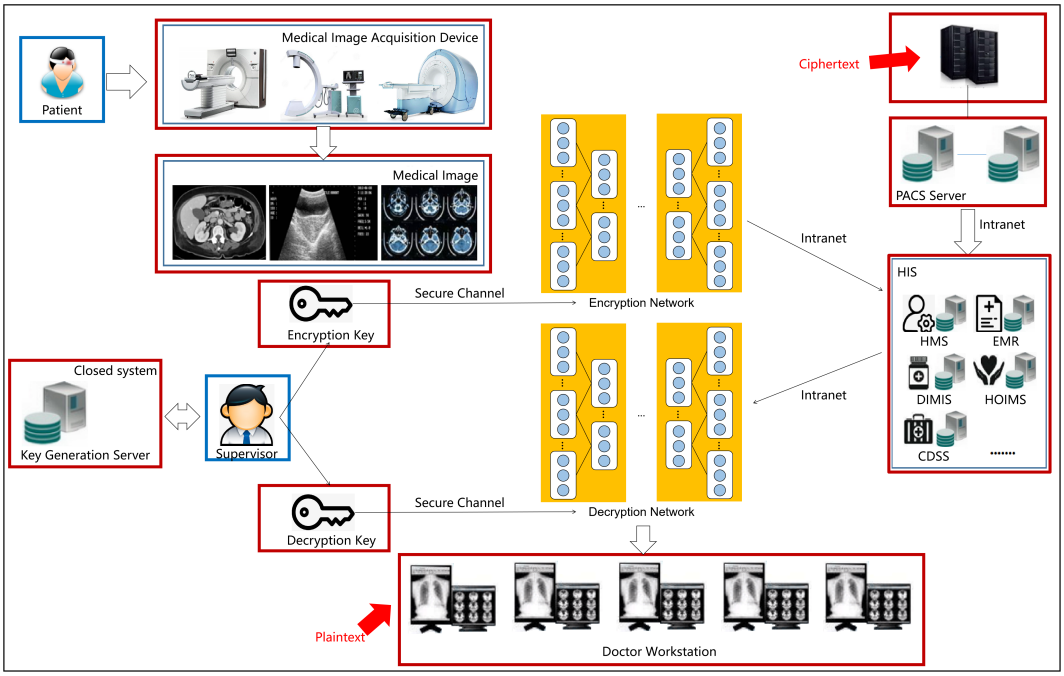

Fig. 1. The architecture of DLEDNet.

based image encryption and decryption network (DLEDNet) is proposed by leveraging deep learning techniques for image-to-image translation and image denoising. The novel idea is based on the following two important insights: (1) If the medical image can be transferred into other image domain which is greatly different from the original one, this medical image can be regarded as encrypted; and (2) the medical image decryption process can be implemented in the manner of image denoising or image reconstruction. In DLEDNet, the Cycle-Gan network is employed as the main learning network to implement the image-to-image translation. There are two domains in the encryption process: the original medical image domain and the target domain, where the target domain is regarded as a "Hidden Factors" to guide the learning model to realize the encryption process. For the encryption network, it consists of a generation network and a discriminator network. The former will generate the image similar to the target domain, while the latter will promote the generation network to generate the same images as the target domain by identifying the generated images. Therefore, after processing using the encryption network, the original medical image can be converted into the target domain and becomes the ciphertext. The decryption process is similar to traditional encryption-decryption methods, which is the inverse operation of the encryption process. In practice, a reconstruction network, which is actually a decryption procedure, is used to restore the encrypted image to the original one. In DELDNet, the parameters of generation network is regarded as the privacy key for encryption while the parameters of reconstruction network is regarded as the privacy key for decryption. Moreover, DELDNet adopts the unsupervised learning method to train the learning network and it doesn't need a lot of labeled samples. It overcomes the dataset problem in network training and is beneficial to the application of deep learning algorithm in cryptography filed.

Based on DLEDNet, the current PACS system is improved by employing a key generation server. As shown in Fig.1, the key generation server was in charge of training the encryption network and the decryption network as mentioned before for later use. The PACS system can call the encryption network to encrypt the medical image obtained from the medical imaging equipment and then store these ciphertext images into the image database. When a doctor wants to review the medical image, the HIS system will retrieve the specified ciphertext image from the PACS server and adopts the decryption network to decrypt the ciphertext image to the original one before transmitting to the doctor's workstation. The encryption network and the decryption network will be transferred with the secure channel in the hospital information system. Moreover, in most cases, the doctor only needs the interested organ or tissue instead of the whole medical image. In order to effectively obtain the desired information, a ROI-mining-network is proposed to directly extract the ROI (organ or tissue) from the encrypted medical image without decryption. Since the deep learning algorithm can extract the useful information from the noisy images, the proposed ROI-mining-network can be implemented by adopting the deep learning algorithm. More specific, when inputting an encrypted medical image into the ROI-mining-network, the interested segmentation object can be directly extracted without revealing other parts of the patient's information.

In a nutshell, the main contributions of this work are summarized as follows:

1. An novel medical image encryption and decryption network, DLEDNet, is developed to realize the encipherment process by applying the deep learning in the field of image-

to-image translation. The proposed encryption method is with the large key space, one-time pad, and be sensitive to key change. To the best of our knowledge, this work is the first work to attempt to adopt the deep learning method in the area of medical image encryption.

2. A ROI-mining-network is proposed to directly extract the interested segmentation region from the encrypted medical image instead of decrypting the ciphertext image firstly. From to the experiments, it can be found that the proposed approach can realize the data mining process directly from the privacy-protected environment.

3. Extensive experiments are conducted on the chest Xray dataset to evaluate the proposed DLEDNet. The results demonstrate that the medical image can be transmitted with a high level of security and efficiency, compared with existing medical image encryption methods. Moreover, the proposed encryption algorithm can resist various attacks, even if the attacker has known the complete process for key generation.

II. Related Work

A. Medical Image Encryption

In the literature, there are many approaches proposed for image encryption [22]–[24]. Lima et al [22] propose a novel scheme for encryption of medical images, based on the cosine number transform, which is a mathematical tool requiring modular arithmetic only. This property avoids rounding-off errors and allows that the image recovered after the encryption/decryption process is identical to the corresponding original image. The proposed scheme is flexible and can be applied to images complying with the DICOM standard, which is frequently employed in medical applications. Natsheh et al [23] propose a simple and effective encryption approach for for multi-frame DICOM medical images. It can reduce the encryption and decryption time of these images by using Advanced Encryption Standard (AES). Mukhedkar et al [24] demonstrate that image encryption can be done using Blowfish Algorithm as it is faster and LSB technique is used for image hiding.

Chaotic maps have the characteristics of pseudo-randomness, ergodicity and initial value sensitivity. Chaotic sequences generated by chaotic maps have good characteristics of security keys. Chaotic cryptography has gradually become a new research direction of cryptography. Medical image encryption can also be performed using chaotic maps. Kanso et al [25] propose a novel full and selective chaos-based image encryption scheme suitable for medical image encryption. The proposed approach consists of several rounds, where each round has two phases: a shuffling phase and a masking phase. Both phases are block-based and use chaotic cat maps to shuffle and mask an input image. Chong et al [26] present a novel chaos-based medical image encryption scheme which introduces a substitution mechanism in the permutation process through a bit-level shuffling algorithm in order to improve the efficiency. In order to protect image effectively, Yu et al [27] present an image encryption algorithm based on wavelet function and four-dimension chaotic system. The algorithm firstly uses Wavelet function chaotic maps to scramble the image pixel location, and then use four-dimension chaotic system disturbs image pixel value.

B. Image-to-Image Transfer by Generative Adversarial Networks

Since the seminal work by Goodfellow et al [28] in 2014, a series of GAN-based methods have been proposed for a wide variety of applications. The original GAN can learn a generator to capture the distribution of real data by introducing an adversarial discriminator that evolves to discriminate the difference between the real data and the fake data [28]. GAN-based methods can produce state-of-the-art results in many applications such as image generation [29], image segmentation [30], image super-resolution [31], and image-to-image translation [37] [38].

Yi et al [40] use a conditional generative adversarial network to learn a mapping from input to output images. It is proved that this approach is effective on synthesizing photos from label maps, reconstructing objects from edge maps, and colorizing images. DualGAN [39] mechanism enables image translators to be trained from two sets of unlabeled images. Taking two sets of unlabeled images as the input, DualGAN simultaneously learns two reliable image translators from one domain to the other and hence can facilitate a wide variety of image-to-image translation tasks. CycleGAN proposed in [19] is a framework to perform image translation with unpaired training data. To achieve this goal, it trains two sets of GAN models at the same time, mapping from class A to class B and from class B to class A, respectively. The loss is formulated based on the combined mapping that maps images to the same class. The key to GANs' success is the idea of an adversarial loss that forces the generated images to be distinguishable from target images. The adversarial loss is used to learn the mapping of the original images to the "target domain images" which represents the Image-to-Image translation.

Regardless of their merits, these algorithms are difficult to achieve a good balance between the security and the efficiency. On the one hand, as there is plenty of information in a single medical image with high correlation among these information, when encrypting the medical image, the block encryption algorithm is with low efficiency and cannot meet the real-time requirement. On the other hand, the chaotic system usually adopts the one-dimensional chaotic map to generate pseudorandom sequences. Consequently, the chaotic system tends to be easy to analyze and predict through a nonlinear prediction method based on phase-space reconstruction [40]. The deep learning algorithm has been used in the security field [41]–[43]. However, there is no work on the medical image encryption and decryption.

In this paper, deep learning techniques are used to encrypt and decrypt medical images, in which parameters of the deep learning network model are regarded as the encryption and decryption keys. Due to the large key space and the complex model structure, the proposed method can achieve a high level of security with a high efficiency.

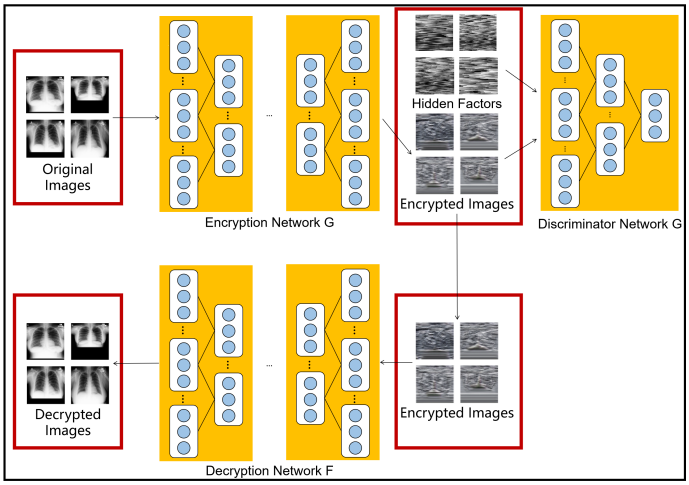

Fig. 2. The overall framework of DLEDNet.

III. Encryption and Decryption Network

A. Architecture of DLEDNet

As shown in Fig.2, DLEDNet mainly consists of three sub-networks: the encryption network G, the discriminator network D, and the decryption network F. The encryption network G is used to encrypt the original input images, the decryption network F is responsible for restoring the encrypted images to the original one (decrypting the image), and the discriminator network D is mainly designed for improving the encryption performance of the encryption network by distinguishing between the generated encrypted generated encryption images and the images in target domain (Hiding factors). In deep learning methods, loss function is usually used to train the model. The overall loss L of the proposed model is given as follows:

$$L = L_G + L_D + L_{reconstruction} \tag{1}$$

Where the L_G indicates the loss of the encryption network G, L_D indicates the loss of the discriminator network D, and $L_{reconstruction}$ indicates the loss of the decryption network F.

1) Encryption Network and Decryption Network: Fig.**??** shows the explicit structure structure of the encryption network G, which is used to transform the input medical images into in target domain for medical image encryption. The G network begins with an initial convolution stage to spatially compress and encode the images, and the useful features extracted in this stage will be used for the following transformation. Then there are nine nine residual blocks similar to the residual block of [44], and combine with identical layout with identical layout to construct the content and manifold features. The output images are reconstructed by two up-convolution blocks which contain a strided convolutional layer and the stride is set to 2. Finally, the prediction is exported by a 7×7 convolution kernel. In addition, the structure of decryption network F is the same as the encryption network G.

The proposed model includes two mappings $G : X \rightarrow Y$ and $F : Y \rightarrow X$. The goal of mapping function G is to learn how to transform the original medical images X into the images Y in target domain, and trick the discriminator network D. When the discriminator network D cannot successfully distinguish whether an image is generated by the encryption network G or a real ciphertext image domain Y, it means that the encryption network G convert the original patient image domain X into a ciphertext image domain Y successfully. The loss L_G of the encrypted network G is:

$$L_G = min_G(E_{x \sim pdata(x)} log(1 - D(G(x))), \tag{2}$$

where G represents an encryption network, and D represents the discriminator network. The loss of G is to minimize the success rate of the discriminator network D in detecting the ciphertext generated by the encryption network G. In addition to the encryption, another goal of the proposed method is to ensure that the restored image reserves the texture information of the original one even it is encrypted. As shown in Fig.2, for each image x from domain X, the reconstruction loss measures the difference between $G(x)$ and the original image, i.e., $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$. The reconstruction loss L defined as:

$$\begin{aligned} L_{reconstruction} &= E_{x \sim p_{data(x)}} ||Y - X||_1 \\ &= E_{x \sim p_{data(x)}} \sum_{i=1}^{n} |y_i - x_i| \\ &= E_{x \sim p_{data(x)}} (|y_1 - x_1| + ... + |y_i - x_i|) \end{aligned} \tag{3}$$

2) Discriminator Network: The discriminator network D is used to check whether the output image of encryption network belongs to the target domain. For the discriminator network D, after initial convolutional layers, the network employs two strided convolutional blocks to reduce the resolution of image and encode essential local features for subsequential discrimination. Then, a feature construction block and a 3×3 convolutional layer are used to obtain the final result. Besides, for each convolutional layer, the Leaky ReLU (LReLU) with $\alpha = 0.2$ is adopted and followed with a batch normalization (BN) layer.

The training of the discriminator network D is to classify the images and determine whether it comes from the ciphertext domain Y or is generated by the encryption network G. The encryption network G attempts to generate an images $G(x)$ similar to the image in domain Y, while the discriminator network D aims to find the difference between translated samples from $G(x)$ and real samples in y. The loss L_D of the discriminator network D can be regarded as maximizing of the classification accuracy of the discriminator network D, which is opposite to the goal of the encryption network G is described as follows:

$$L_D = E_{x \sim pdata(x)} log D(x) + E_{x \sim pdata(x)} log(1 - D(G(x))) \tag{4}$$

Where G represents the encrypted network, and D represents the discriminator network. The L_D and L_G in the GAN network form a adversarial relationship. When the two networks reach an equilibrium state, the discriminator network D can achieve 50% classification accuracy for both the ciphertext image generated and the real ciphertext domain image Y. In other words, the ciphertext image generated by the encryption

network G is very similar to the real ciphertext domain Y so that the discriminator network D cannot distinguish them.

3) The key generation process: In DLEDNet, the final parameters of network G can be considered as the privacy key for encryption while the parameters of network F are regarded as the privacy key for decryption. The process of generating the privacy key is as follows: For encryption, the parameters for each convolutional layer are firstly randomly initialized as follows:

$$W_n = random[w_{n,1}, w_{n,2}, ..., w_{n,j}, ...], \tag{5}$$

where w_n is the n^{th} convolutional layer and $w_{n,j}$ is the j-th parameter of one convolutional layer. Therefore, the privacy key W for encryption is actually composed of all the parameters of each convolutional layer, and is defined as follows:

$$W = consist[W_1, W_2, ..., W_n, ...] \tag{6}$$

When training the encryption network, the privacy key for encryption is continuously updated and refined with different input images through forward propagation training process. The adversarial loss L_{gan} is calculated to measure the difference between the predicted result and the target in "Hidden Factors", thereby guiding the network to train and update the privacy key for encryption.

Except for the forward propagation, the back-propagation algorithm (BP) is also employed to pass loss of the entire network between the convolutional layers. It is actually a gradient descent, which can further update the parameters in each layer to achieve better performance. The gradient descent can be described as:

$$\begin{aligned}
\theta_j &= \theta_j - \alpha \vee J(\theta) \\
&= \theta_j - \alpha \frac{\delta}{\theta_j} J(\theta) \\
&= \theta_j - \alpha \frac{\delta}{\theta_j} \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^i) - y^i)^2 \\
&= \theta_j - \alpha \frac{1}{2m} \sum_{i=1}^{m} \frac{\delta}{\theta_j} (h_\theta(x^i) - y^i)^2 \\
&= \theta_j - \alpha \frac{1}{2m} \sum_{i=1}^{m} 2 \frac{\delta}{\theta_j} (h_\theta(x^i) - y^i)(\frac{\delta}{\theta_j}(h_\theta(x^i) - y^i)) \\
&= \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^i) - y^i) \times (\sum_{i=0}^{n} \frac{\delta}{\theta_i} \theta_i x_i - \frac{\delta}{\theta_i} y^i) \\
&= \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^i) - y^i) \\
&\times (\frac{\delta}{\theta_i}(\theta_0 x_0 + \theta_1 x_1 + ... + \theta_n x_n) - \frac{\delta}{\theta_i} y^i)
\end{aligned} \tag{7}$$

Where θ_j is the value of parameter θ in the j^{th} training epoch. The α is the learning rate and $\vee J(\theta)$ means the gradient that pass back to the convolution layer θ in the j^{th} training epoch.

The generation process of the privacy key for decryption is similar to the process of generating the privacy key for encryption, except that the initial input of the decryption network becomes the predicted result of the encryption network. In addition, the loss of decryption network is the reconstruction loss, which is given in Equ. (8).

$$\begin{aligned}
L_{reconstruction} &= E_{x \sim p_{data(x)}} ||F(P(X)) - O(X)||_1 \\
&= E_{x \sim p_{data(x)}} \sum_{i=1}^{n} |F(P(x_i)) - O(x_i)| \\
&= E_{x \sim p_{data(x)}} (|F(P(x_0)) - O(x_0)| \\
&+ ... + |F(P(x_i)) - O(x_i)|)
\end{aligned} \tag{8}$$

where $F()$ is the decryption network, $P(x)$ is the pixel x in the predicted image, and $O(x)$ is the corresponding position pixel x in the original image. The encryption network G and the decryption network F are trained in an alternative manner. When the loss becomes stable, the final parameters (privacy keys) for encryption and decryption network can be obtained. The complete privacy key generation process is shown in Fig.3.

After obtaining the key, the patient's medical image can be encrypted by the encryption network G, and then decrypted by the decryption network F. The proposed medical image encryption/decryption algorithm is given in Alg. 8.

Algorithm 1 Image Encryption/Decryption.

Initialization: Digitize the 255×255 image into a 255×255 matrix X_*^0.And then enter it into our 21-layer(L_c) encryption/decryption model.

1: **while** L< L_c **do**
2: **for all** element (X_1^L, X_2^L,...) in matrix X^L **do**
3: Each pre-trained 3×3 convolution kernel W_*^L in L^{th} layer sequentially traverses the image matrix and multiplies it with the corresponding elements of the matrix($W_*^L \times X_*^L$).
4: Add the obtained nine $W_* X_*$ to get a new predicted value X_*^{L+1} in (L+1)th.
5: Collect all X^{L+1} and combine them into a new matrix to form the next-level feature matrix.
6: **end for**;
7: $L = L + 1$;
8: **end while**

Output: Convert the last layer of matrix X^{L_c} into an image to get the final encrypted/decrypted image.

Since the GAN model is highly nonlinear and randomly initialized, and the parameters of the learning network can be totally different at different training times. In other words, the GAN network is unstable, which is its weakness when used for computer vision tasks. However, this instability has advantages for cryptography. By utilizing this instability, the proposed deep learning based encryption method can be regarded as an one-Time Pad (OTP) method. Specifically, the parameters of encryption network are totally different after training the network at different times. Overall, due to the depth and complex structure of the learning / encryption network, the proposed framework has higher security.

B. ROI Mining Network in Ciphertext Environments

Although various methods have achieved a good performance in protecting the image privacy, it is still very challeng-

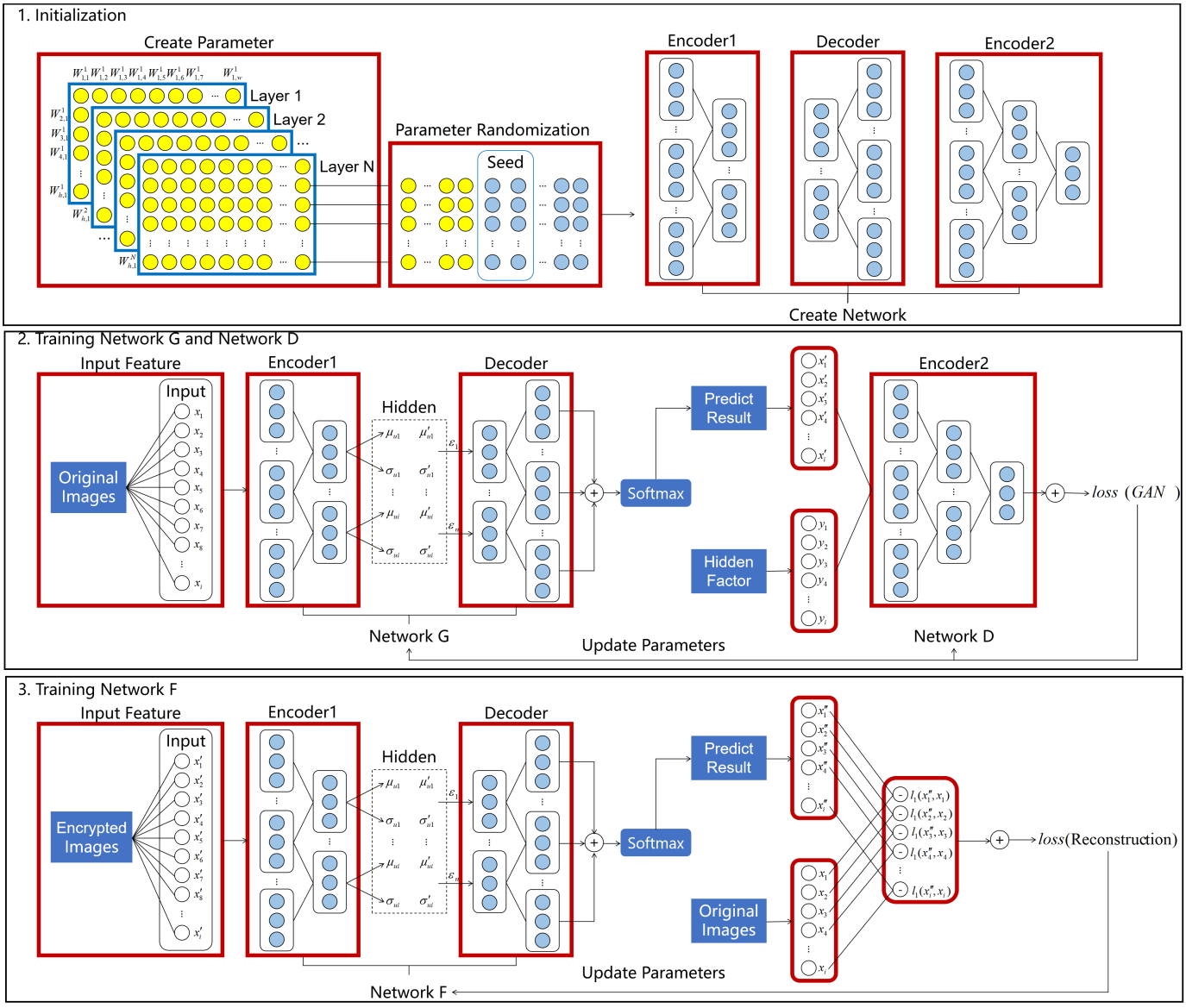

Fig. 3. The key generation process.

ing to directly obtain the effective information in a ciphertext environment, such as extracting the desired ROI from the encrypted medical image. In DLEDNet, a ROI-mining-network is proposed to segment the region of interest from the encrypted medical image. An overview of the network structure of the ROI-mining-network is shown in Fig. 8, in which the backbone is the ResNet-50 [32]. In order to extract useful texture features in a ciphertext environment, a deeper network structure is adopted to learn semantic features to accurately segment the specific target. The input encrypted image will be processed with 5 blocks, and each block has a down-sampling convolution. In the first block, since the convolutional kernel size is set to 3×3, each convolution operation can learn the local information from the input image. As the depth of the network increases, more abstracted semantic information can be obtained. Finally, by combining the output results from each convolution layer, the final prediction results can be obtained.

Each block in ResNet-50 has two sub-blocks. One is the identity (ID) block in which the stride of each convolution layer is 1. The identity block is mainly used to extract abstract features through multi-layer convolution. Because the dimensions of the input and output are the same, these feature maps can be connected serially. The other basic block is Conv Block where the dimensions of input and output are different and it is used to change the dimension of the feature vector and to resize feature size through a strided convolutional layer. The CNN-based neural network commonly converts the image into a small feature map with many channels. However, with the increasing of the network layers, there will be a huge number of output channels and parameters, resulting in increased computational complexity and reduced network efficiency. Therefore, it is necessary to reduce the dimension of Conv Block before processing with the Identity Block.

In DLEDNet, the proposed ROI-mining-network is used to implement the medical image segmentation task in the ciphertext environment. The medical image segmentation is a

critical step in medical image analysis. Its purpose is to extract useful features and segment the doctors' interested objects. The segmentation results can provide a reliable basis for clinical diagnosis and pathological research. When training the ROI-mining-network, the encrypted medical image is first used as the input of the network. Then, the pixel-level segmentation labels in the corresponding medical image are adopted to supervise the training process. Finally, the model parameters are updated by the mean square error (MSE). The loss function of this segmentation model is described as:

$$L_{segment} = \frac{1}{N}\sum_{i=0}^{N}(ground_i - predict_i)^2 \tag{9}$$

where the $ground_i$ represents the value of the i^{th} pixel in the label and the $predict_i$ is the predicted value of the i^{th} pixel in the predicted result. N represents the total number of pixels in this image. The final training result is a high-quality splitter that can segment the medical images without decryption.

The usage of the ROI-mining-network is of great significance for medical image security. It can implement data mining in an untrusted environment to securely extract specific objects, which is also beneficial for protecting the privacy patient. This network can further improve the security of medical image analysis and can be widely used in many medical applications.

C. Adversary Model

In DLEDNet, the most important factors of key generation process include the structure of the model and the chosen hidden factors. If the network structure or hidden factors leaks, the attacker can train a similar encryption network by imitating the privacy key generation process so as to crack the ciphertext image. This kind of attack is called as the imitation learning attack. This paper proposes three possible adversary models for imitation learning attack: the hidden factors leakage, the network architecture leakage, and both the hidden factors and the network architecture leakage.

1) Hidden Factors Leakage: Hidden factors leakage means that the attacker has known the hidden factors used for the encryption, and tries to employ the same hidden factors to train the attacking network with several different network architectures to decrypt the ciphertext image. There are two encryption and decryption networks with different network structures: the encryption/decryption network A and encryption/decryption network B. These two encryption and decryption networks are trained with the same hiding factor. If the decryption network B is able to recover the image encrypted by the encryption network A, it means that the attacker can crack the secure key by imitation learning attack.

2) Network Architecture Leakage: Network architecture leakage assumes that only the architecture of the encryption and decryption network is leaked, and the hidden factors remains confidential. In this adversary model, the attacker can decrypt the encrypted image by training the same network structure without knowing the hidden factors. The attacker can employ different hidden factors to train the same network structure to construct different decryption networks. If the attacker is able to recover the encrypted ciphertext image, the attack is successful.

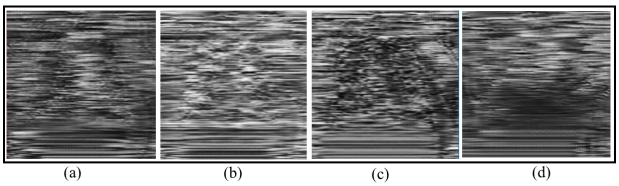

Fig. 4. The same image is encrypted with the key obtained from four networks.

3) Both Hidden Factors and Network Architecture Leakage: The strongest adversary model is that both the network architecture and hidden factors are leaked. In such a scenario, the attacker can train the network with the same network structure and hidden factor adopted for training the encryption/decryption network. To prevent such attacks, after each training of the network, the parameters of the encryption/decryption network representing the actual private key must be completely different. It means that the proposed encryption algorithm should be similar to the OTP and can be regarded as a chaotic encryption algorithm.

IV. Security Analysis

In DLEDNet, both encryption and decryption network are constructed with 24 layers and the number of the parameters for each network is 2,757,936. The explicit specification of the network are shown in Table I. And for the ROI-mining-network, a deeper resnet-50 architecture is adopted. The network structure of the ROI-mining-network is given in Table II. The dataset is the chest x-rays [45]. The proposed method is running on the Nvidia GTX 2080Ti. When training the network, it takes around 10 mins for each epoch of the model.

A. Key Security Analysis

The ideal encryption scheme has the following characteristics: 1) the key space is large enough so that it can effectively resist the exhaustive attack under the premise of the existing computing power; 2) the key generated for each time should be different, i.e., the key generation should be uniform at random; and 3) the encrypted image must be highly sensitive to the key. The security of the key will be analyzed from these three characteristics in the following sections.

1) Key Space Analysis: The size of the key space determines the difficulty of an attacker using an exhaustive attack. In this work, the key space of the proposed encryption algorithm is the number of parameters for the deep learning network, with a total of 2,757,936 parameters in the experiments. Each parameter or key is a floating point number between 0 and 1, which is 32 bits in the computer and can be expressed as a decimal number with 10 significant digits. Therefore, the key space of the encryption model can be expressed as the $(10^{10})^{2757936}$. It becomes very hard for attackers to break down and can effectively resist attacks.

TABLE I
THE STRUCTURE OF ENCRYPTED NETWORK AND DECRYPTION NETWORK.

Convolution Layer Name	Number	Size	Input Channels	Output Channels	Parameters	Total Parameters
Down Convolution1	1	7×7	3	32	4704	4704
Down Convolution2	1	3×3	32	64	18432	23136
Down Convolution3	1	3×3	64	128	73728	95864
Residual Blocks	18	3×3	128	128	2564208	2661072
Up Convolution1	1	3×3	128	64	73728	2734800
Up Convolution2	1	3×3	64	32	18432	2753232
Up Convolution3	1	7×7	32	3	4704	2757936

TABLE II
THE STRUCTURE OF ROI-MINING-NETWORK.

Convolution Layer Name	Number	Size	Input Channels	Output Channels	Parameters	Total Parameters
Block 1	2	7×7	3	64	4704	4704
Block 2	3	3×3	64	256	18432	23136
Block 3	12	3×3	256	512	73728	95864
Block 4	18	3×3	512	1024	2564208	2661072
Block 5	1	3×3	1024	2048	73728	2734800

2) Key Randomness Analysis: The encryption network is trained four times with the same settings. Accordingly, the parameters of these four networks are adopted as encryption keys, i.e., Key A, Key B, Key C and Key D, respectively. The same image is encrypted with these four keys, and the encrypted images are shown in Fig. 11. Fig. 4(a), Fig. 4(b), Fig. 4(c), and Fig. 4(d) are the results obtained by encrypting the same original image from four networks. It is clear that these four images are different. The similarity among these four encrypted images (SSIM) are calculated, and the result can be found in Table III. The SSIM index between different images is mostly lower than 0.1 which indicates that the similarity between different images is very low.

According to the experiment, it can be found that since parameters of the neural network are randomly initialized, the privacy keys for the medical image encryption network are totally different after every training. These difference results in different encrypted images which are processed with different encryption networks. The idea behind this is the training of the deep learning network is not stable. Different initialized parameters can lead to various final parameters in different training. It can be demonstrated that the proposed method is similar to OTP and can be regarded as one type of OTP method.

TABLE III
SSIM BETWEEN TWO ENCRYPTED IMAGES.

Image	A	B	C	D
a	1	0.07	0.11	0.09
b	0.07	1	0.08	0.04
c	0.11	0.08	1	0.05
d	0.09	0.04	0.05	1

3) Key Sensitivity Analysis: Unlike traditional encryption algorithms, the error in deep learning models will be propagated among layers. In the convolution process, the l^{th} pixel in the N^{th} layer feature map is passed to a neighboring pixel of the $(N+1)^{th}$ layer via a 3×3 convolution kernel. When a feature point is erroneous, it will be passed to the 3×3 feature points in the next layer. As the depth of the convolutional network increases, the error of feature points will increase with two pixels for each laye. In the up-sampling process, this error increases exponentially with the superposition of the deconvolution operation. The experiment assumes the attacker knows the most privacy keys. And only about 5% of key parameters are modified which is regarded as the unknown part. Then, the encrypted image is input to the network with new parameters, the network cannot decrypt the ciphertext image to the original one. It means that even if only 5% of the parameters is changed, the privacy key cannot encrypt or decrypt the medical image correctly. In other words, it becomes very hard for attackers to guess at least 95% of the right key parameters in a key space with $(10^{10})^{2757936}$ so as to break the proposed algorithm.

B. Ciphertext Security Analysis

1) Histogram Analysis: To evaluate the performance of the proposed encryption network, The original image is shown in Fig. 5(a) and the encrypted image is shown in Fig. 5(c). Through the experiment, it can be found that the pixel distribution of the original image and the encrypted image is very different. In Fig. 5, the pixel histogram of the original chest X-ray image has a total of 57600*(240 * 240) pixels (Fig. 5(b)), in which more than 30,000 pixels have a value of 0, and more than 5000 pixels have a value of 255. The pixel distribution of original image is relatively concentrated. However, the distribution of encrypted medical images (Fig. 5(d)) is more uniform, which is helpful for mitigating the statistical analysis.

2) Entropy Analysis: The information entropy of the encrypted image is regarded as the effective quantitative measurement for algorithms to against statistical attacks. The image information entropy represents the statistical feature of the grayscale distribution of the image. In an ideal case, the encrypted image should be similar with random noise, the grayscale distribution tends to be uniform, and the expected

value should be 8. The information entropy formula is defined as follows:

$$Entropy = -\sum_{l=0}^{N} p(l) log_2(p(l)) \tag{10}$$

where N is the number of gray levels of the pixel value and $p(l)$ is the probability that the pixel value l appears. The entropy metric is calculated on the encrypted medical image, and the results are shown in Table IV. It is clear that the image encrypted by the proposed method is close to the ideal value of 8 on information entropy. Experiments show that the images encrypted by the proposed method has the ability to resist the statistical attacks.

TABLE IV
EVALUATION OF THE ENTROPY EFFECT OF OUR NETWORK.

Image Id	**1**	**2**	**3**	**4**	**5**
Entropy	7.96	7.96	7.95	7.94	7.95
Image Id	**6**	**7**	**8**	**9**	**10**
Entropy	7.97	7.95	7.96	7.96	7.95

C. Security Analysis under Different Adversary Models

The experiments are conducted to validate whether an attacker can generate a key under three different adversary models.

1) 'Hidden Factors Leakage: In this experiment, four different network structures are considered, namely network A, network B, network C and network D. The training conditions are kept the same. The network structure of these four networks are shown in Table V.

TABLE V
THE NETWORK MODEL OF THE DIFFERENT ARCHITECTURES.

Convolution Layer	**Net. A**	**Net. B**	**Net. C**	**Net. D**
Down Convolution1	1	1	1	1
Down Convolution2	1	1	1	1
Down Convolution3	1	1	1	1
Residual Blocks	18	15	12	9
Up Convolution1	1	1	1	1
Up Convolution2	1	1	1	1
Up Convolution3	1	1	1	1

The original image is encrypted by using the trained network A. The ciphertext image is then decrypted by the decryption network obtained from network A, network B, network C and network D respectively to restore the original image. As shown in Fig.6, the original image (Fig.6(a)) encrypted by network A (the encrypted image is shown in Fig.6(b)), can only be correctly decrypted by the decryption network A as shown in Fig.6(c). While the image decrypted by network B, network C and network D are visually unrecognizable and the result is shown in Fig.6(d), Fig.6(e), Fig.6(f), respectively. Experiments show that even if the attacker knows the hidden factors, the "attack network" trained with different network structure still cannot be used to decrypt the ciphertext image.

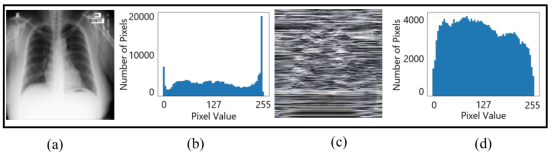

Fig. 5. Pixel distribution of the original image and the encrypted image.

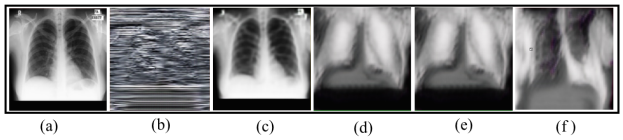

Fig. 6. The decryption performance for different networks.

2) Network Architecture Leakage: In this experiment, different hidden factors are adopted to train the encryption network with the same network structure. All training conditions are kept the same. As shown in Fig.7(a) and Fig.7(b), two different domain images ("Hidden Factors A" and "Hidden Factors B") are chosen as hidden factors to train the network with the same architecture. The Fig.7(c) is original image, Fig.7(d) is the image generated by the encrypted network which is trained by "Hidden Factors A", and Fig.7(e) presents the result of decrypting ciphertext image through the decryption network trained with "Hidden Factors B". From the experiment, it can be found that the image generated by the encrypted network which is trained by "Hidden Factors A" cannot be decrypted by the network trained by "Hidden Factors B". Therefore, it can be proven that the "attack network" with the same architecture trained by different hidden factors, cannot be used to decrypt the ciphertex image with each other. That is, even if attackers obtain the network architecture, they cannot train the decryption network to decrypt the encrypted image without knowing the hidden factors.

3) Both Hidden Factors and Network Architecture Leakage: In this experiment, the network is trained with four times under the same hidden factors and training conditions to get the networks A, B, C and D, respectively. The experiment evaluates the decryption performance for these four networks on the same ciphertext image to verify whether the parameters generated for each network are different. As shown in Fig.11, the gray value distribution of the image decrypted by the decryption key B, the decryption key C, and the decryption key D is completely different from the image decrypted by the decryption key A. It can be clearly found that under the same training condition, the encrypted medical image encrypted by one network, cannot be decrypted by adopting the parameters in other network. Even if the model parameters are trained with the same network architecture and the same hidden factors, they cannot be used to decrypt the image with each other. Experiments show that even if both the network architecture and the hidden factors are leaked, and training the network under the same training conditions, the parameters of each network are totally different, i.e., the secure keys are different. It can be proven that DLEDNet is secure even if the network architecture and hidden factors are revealed.

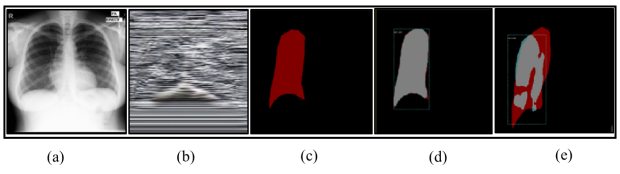

Fig. 7. The mutual decryption performance between networks under different hidden factors training.

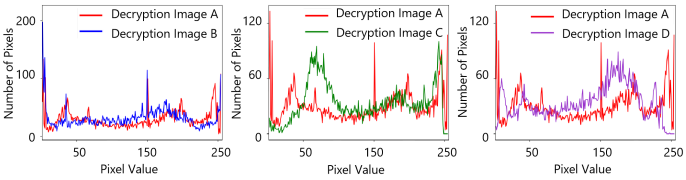

Fig. 8. The decryption performance for these four networks on the same ciphertext image.

D. Security Analysis under Different Attack Models

1) Ciphertext Only Attack: In this type of attack, the attacker has access to a string of ciphertext, but cannot access to the corresponding plaintext.

In DLEDNet, the key space of the encryption model can be expressed as $(10^{10})^{2757936}$ and it is very hard for attacker to break down. At the same time, the privacy key generated with multiple iterations and diffusions is complex. Therefore, it is difficult to crack the ciphertext through ciphertext only attacks.

2) Known Plaintext Attack: In this type of attack, the attacker knows a string of plaintext, and the corresponding ciphertext. The attacker will try to decrypt the rest of the ciphertext by using these known information.

In traditional sequential pixel visiting pattern methods, concrete encryption factors are generally retrieved as equivalent keys for recovering the received ciphertexts. Taking XOR encryption as an example, the masks calculated directly from plaintext and ciphertext are sufficient to decode the ciphertext. Typically, masks sequentially correspond to the plain pixels and the retrieved masks by plaintext attack can be directly used to decrypt other ciphertexts. However, the proposed algorithm adopted the non-sequential encryption mechanism. Without the knowledge of the pixel visiting pattern, the privacy key cannot be obtained by the attacker, thus making plaintext attack infeasible. The proposed algorithm adopts the iteration and diffusion procedures to generate the privacy key. These kinds of producers can significantly promote the security performance and thus provide additional immunity of the cipher against known plaintext attack.

3) Chosen Plaintext Attack: In this type of attack, the attacker can access the encryption device, choose a string of plaintext and construct its corresponding ciphertext string.

Generally, an attacker can observe the change of the ciphertext image by making small changes to the plaintext image, such as changing the value of only one pixel of the ciphertext, so as to obtain the connection between the plaintext image and the ciphertext image. This type of attack is called as differential attack which is a kind of chosen plaintext attack method. If a small change in the plaintext image can cause a huge change in the ciphertext image, this differential attack method usually fails to take effect. It indicates that the encryption algorithm can resist this chosen plaintext attack method. Here, the Number of Pixel Change Rate (NPCR) is adopted to measure the degree of image changing. NPCR refers to the rate of pixels change which indicates the ratio of different pixel values at the same position between two plaintext/ciphertext images. The definition of NPCR is as follows:

$$NPCR = \frac{\sum_{i=0}^{W}\sum_{j=0}^{H} D(i,j)}{W \times H} \tag{11}$$

where W and H are the width and height of the image, respectively. T1 and T2 represent a ciphertext image obtained by encrypting two different plaintext images, respectively. If $T_1(i,j) = T_2(i,j)$, $D(i,j) = 1$. If $T_1(i,j) \neq T_2(i,j)$, $D(i,j) = 0$. In the experiment, there is only about 1% different pixels between these two plaintext images. Both the original plaintext image and the plaintext image with 1% pixel value changed are input to the proposed encryption model. Then, the NPCR is used to compare the differences between these two encrypted images. The calculated average NPCR value is 94.21%, which means that the information of the plaintext image is well diffused into the ciphertext image. Since DLEDNet has good diffusion performance and is with highly sensitive to the plaintext, it achieves a good performance to resist the chosen plaintext attack like the differential attack.

4) Chosen Ciphertext Attack: In this type of attack, the attacker can access the decryption device, choose a string of ciphertext and construct its corresponding plaintext string.

Since the structure of our decryption model is exactly the same as the encryption model, the experiment for chosen ciphertext attack is similar to that in chosen plaintext attack. In this experiment, the input of the decryption network is the ciphertext image and the NPCR is used to calculate the different between two decrypted images. According to the experiment, it is found that when the input ciphertext image changes slightly (just 1% pixels changed), the average NPCR value between two decrypted images is 94.87%. It means that if the input ciphertext image changes slightly, the decrypted image will change dramatically. This demonstrates that the proposed algorithm has good diffusion performance and is also highly sensitive to the ciphertext. It is effective to resist the chosen ciphertext attack.

V. Experiment

A. Performance of Encryption and Decryption

As shown in Fig.9, the results of the proposed method for medical image encryption and decryption are presented in a visual way. It can be seen that the ciphertext image generated by the encryption network G, is totally different from the original medical image and the pathology information cannot be observed. In addition, the image in the third row is decrypted from the encrypted one through the decryption network F, can recover the detailed information of the original image and restore to the original one. In order to evaluate the

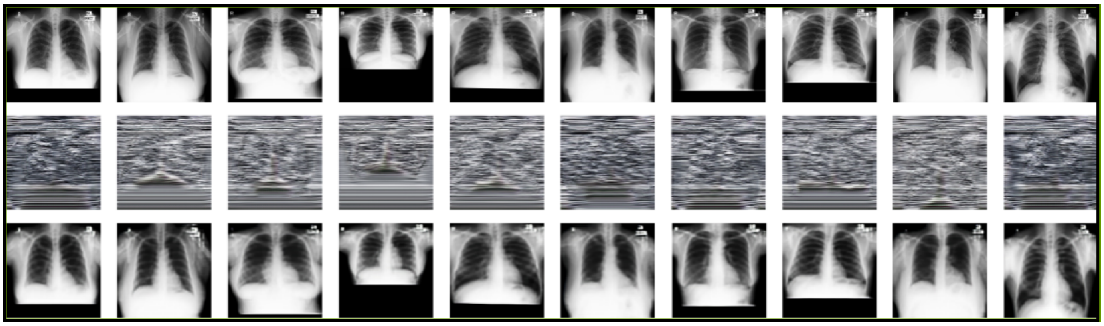

Fig. 9. The encryption and decryption performance of the proposed method, in which, the images are original image, encryption image and decryption image from the top to bottom, respectively.

effectiveness of the decryption network, the Peak Signal to Noise Ratio (PSNR) and structural similarity index (SSIM) are employed as evaluation metrics.

The quantitative measure of the decryption error is PSNR, which is based on the root mean square error (RMSE) between the decrypted data and ground truth. It can be represented as:

$$PSNR = 20log_{10}\frac{255}{RMSE} \tag{12}$$

To further evaluate the performance of encryption and decryption, the SSIM is used as another metric.

$$SSIM(x,y) = [l(x,y)]^{\alpha}[c(x,y)]^{\beta}[s(x,y)]^{\gamma} \tag{13}$$

where $l(x,y)$ is the brightness comparison, $c(x,y)$ is the contrast comparison, and $s(x,y)$ is the structure comparison. The closer the SSIM is to 1, the more resemblance the two images are. And if this value approaches to 0, the two images are completely different. In an ideal case, the SSIM between the encrypted image and the original image is equal to 0, and the SSIM between the decrypted image and the original image is equal to 1. As shown in the second and third rows of Table VI, the SSIM between the encrypted image and the original image is close to 0, and the SSIM between the decrypted image and the original image is close to 1.

For most of medical image processing tasks, the image can be compressed to one-half size of the original one to reduce the storage consumption and does not affect the doctor's diagnosis. In order to ensure that the decrypted image do not affect the doctor's diagnosis, the performance of the reconstructed image decrypted by the decryption network, is also compared with the one-half compressed image. According to the experiment, it is demonstrated that the performance of the reconstructed image is equivalent to that through directly compressing the original image to half and then restoring it. In Table VI, from line 3 to line 6, 2X means that the original image is compressed to one-half and then restored. At this level, human can accurately identify the patient's organ contours and bone information from reconstructed images.

B. Performance of ROI-Mining-Network

Direct extraction of interested information under ciphertext conditions is of great significance for medical image security and also for the data mining with privacy protection.

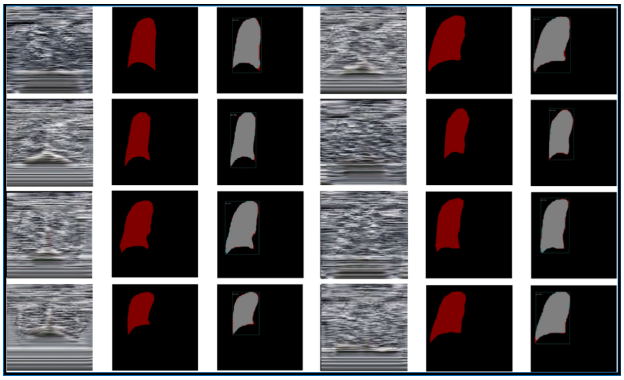

Fig. 10. The performance of ROI-Mining-Network.

The proposed ROI-mining-network can segment the patient's interested organ tissue from the ciphertext image without decrypting the image firstly. The proposed network has the ability to realize the data mining from the privacy environment by extracting the ROI from the encrypted image directly. In order to evaluate the proposed ROI-mining-network, the well-known evaluation metric Dice score is adopted in here and is defined as:

$$Dice(GT, AT) = \frac{GT \cap AT}{(|GT| + |AT|)/2} \tag{14}$$

The GT represents the ground truth and the AT represents the model predictions. Fig.10 shows the performance of the proposed ROI-mining-network on the patient's left lung. It can be clearly seen that the prediction (grey ones) obtained from the model is almost as the same as the ground truth (red ones). In addition, the original medical images are also adopted as the experiment data for training the same ROI-mining-network, which is mainly used as the comparison. Under the same training conditions, the DICE of the segmentation network for plaintext is 0.967, while the DICE of the segmentation network for the ciphertext image is 0.962. It can be proven that the ROI-mining-network can achieve a good segmentation performance on both plaintext and ciphertext images.

As mentioned before, the privacy keys of the network are totally different when training the network at different times even if all the conditions are the same. Therefore, the attacker cannot obtain the same ROI-mining-network even if employing the same ciphertext image for training. The

TABLE VI
EVALUATION OF THE SSIM AND PSNR.

Image Id	1	2	3	4	5	6	7	8	9	10
SSIM(Encrypted)	0.93	0.88	0.90	0.94	0.93	0.91	0.91	0.93	0.91	0.89
SSIM(Decrypted)	0.01	0.02	0.01	0.01	0.02	0.02	0.01	0.02	0.01	0.01
SSIM(2X)	0.90	0.92	0.90	0.92	0.89	0.91	0.88	0.90	0.91	0.90
PSNR	37.43	35.34	36.01	38.03	35.76	35.87	36.13	37.17	35.88	35.74
PSNR(2X)	35.48	35.74	35.03	35.28	34.87	36.73	34.75	34.61	36.17	34.80

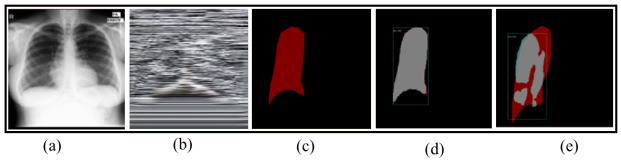

Fig. 11. Attack experiment for the proposed ROI-Mining-Network.

experiment can be found in Fig.11. In this experiment, Fig.11 (a) is the original image and Fig.11 (b) is the ciphertext image of Fig.11 (a). Fig.11 (c) is the ground truth for the right lung segmentation. Fig.11 (d) is the correct extraction result segmented by the ROI-mining-network. Fig.11 (e) is the error extraction result segmented by the attacker.

Fig. 12. The efficiency comparison between our method and other existing methods.

C. Efficiency

To evaluate the efficiency of the proposed network, the running speed of encryption and decryption process on different resolution medical images is evaluated. For 256*256 resolution, the proposed network can encrypt or decrypt 14.28 medical images per second, while the speed is 3.65 images/second for encrypting or decrypting 512*512 resolution image. This encryption/decryption speed can basically meet the efficiency requirement in clinical practice. In addition, some chaotic encryption algorithms have been adopted as the comparison method for evaluating the efficiency. For instance, Zhou et al [32] introduce a simple chaotic system, which employs a combination of two existing one-dimension (1D) chaotic maps (seed maps). Liao et al [33] introduce a novel image encryption algorithm based on self-adaptive wave transmission. Wu et al [35] introduce a wheel-switch chaotic system for image encryption. In [36], the two-dimensional logistic map with complicated basin structures and attractors is firstly used for image encryption. This method can encrypt an intelligible image into a random-like one both from the point of view of the statistical and the human visual system.

Fig. 12 shows the comparison among aforementioned five chaotic encryption algorithms and the proposed method. The FPS represents the number of images that can be encrypted/decrypted in one second. It can be found that our methods achieve the fastest encryption speed both on 512×512 resolution and 256×256 resolution images. Although the number of keys in our method is greater than the number of keys used in chaotic encryption methods, the processing time of our method is still with higher efficiency.

VI. CONCLUSION

In this paper, a novel medical image encryption and decryption method (namely DLEDNet) is proposed by leveraging deep learning techniques, which is one of the early attempts to adopt the concept of "deep learning" for medical image encryption. The Cycle-GAN network is adopted as the learning network to encrypt and decrypt the medical image. A target domain is used to guide the learning model to realize the encryption process. The reconstruction network can decrypt the encrypted image to the original image (plaintext). Moreover, a ROI-mining-network is proposed to directly extract the ROI from the encrypted medical image, with which DLEDNet can segment the interested organ or tissue in the ciphertext environment without decrypting the medical image. We conduct experiments on the chest X-ray datasets, the results show that the proposed algorithm can protect the medical image with a high security level and can encrypt/decrypt the image in a more efficient way, compared with other state-of-the-art counterpart medical image encryption methods.

REFERENCES

[1] A. Gatouillat, Y. Badr, B. Massot, E. Sejdić, "Internet of Medical Things: A Review of Recent Contributions Dealing With Cyber-Physical Systems in Medicine," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3810-3822, Oct. 2018.

[2] N Zhang, P Yang, J Ren, et. al., "Synergy of big data and 5g wireless networks: opportunities, approaches, and challenges," *IEEE Wireless Communications*, vol. 25, no.1, pp. 12-18, 2018.

[3] D. Chen, N. Zhang, Z. Qin, X. Mao, Z. Qin, X. Shen, and X. Y. Li, "S2M: A Lightweight Acoustic Fingerprints-Based Wireless Device Authentication Protocol," *IEEE Internet of Things Journal*, vol. 4, no. 1, pp. 88-100, Feb. 2017.

[4] B. Liu, H. Huang, "Picture archiving and communication systems and electronic medical records for the healthcare enterprise," *Biomedical Information Technology*, Academic Press, pp. 105-164, 2020.

[5] K. Zhang, J. Ni, K. Yang, X. Liang, J. Ren, X. S. Shen, "Security and privacy in smart city applications: Challenges and solutions," *IEEE Communications Magazine*, vol. 55, no. 1, pp. 122-129, Jan. 2017.

[6] D. Chen, N. Zhang, et al., "Channel precoding based message authentication in wireless networks: Challenges and solutions", *IEEE Network*, vol. 33, no. 1, pp. 99-105, 2018.

[7] D. Chen, N. Zhang, N. Cheng, K. Zhang, Z. Qin, X. Shen, Physical Layer based Message Authentication with Secure Channel Codes, IEEE Transactions on Dependable and Secure Computing, DOI: 10.1109/TDSC.2018.2846258.

[8] Y. Zhang, W. Liu, S. Cao, Z. Zhai, X. Nie, and W. Dai, "Digital image encryption algorithm based on chaos and improved DES," *in Proc. 2009 IEEE International Conference on Systems, Man and Cybernetics*, USA, pp. 474-479, Mar. 2009.

[9] K. Chang, Y. Chen, C. Hsieh, C. Huang and C. Chang, "Embedded a Low Area 32-bit AES for Image Encryption/decryption Application," *in Proc. 2009 IEEE International Symposium on Circuits and Systems*, Taipei, pp. 1922-1925, Apr. 2009.

[10] M. Preishuber, T. Hutter, S. Katzenbeisser and A. Uhl, "Depreciating Motivation and Empirical Security Analysis of Chaos-Based Image and Video Encryption," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 9, pp. 2137-2150, Sept. 2018.

[11] Y. LeCun, B. Boser, J. Denker, et. al., "Backpropagation Applied to Handwritten Zip Code Recognition," *Neural Computation*, vol. 1, no. 4, pp. 541-551, Dec. 1989.

[12] L. Ale, N. Zhang, H. Wu, et. al., "Online Proactive Caching in Mobile Edge Computing Using Bidirectional Deep Recurrent Neural Network, IEEE Internet of Things Journal, vol. 6, no. 3, pp. 5520-5530, 2019.

[13] H. Chen, Z. Qin, Y. Ding, et.al., "Brain tumor segmentation with deep convolutional symmetric neural networ," *Neurocomputing*, DOI:10.1016/j.neucom.2019.01.111.

[14] Y. Ding, C. Luo, et.al., "High-order correlation detecting in features for diagnosis of Alzheimer's disease and mild cognitive impairment," *Biomedical Signal Processing and Control*, vol. 53, Sept. 2019.

[15] K. H. Jin, M. T. McCann, E. Froustey, et.al., "Deep Convolutional Neural Network for Inverse Problems in Imaging," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4509-4522, Sept. 2017.

[16] C. Dong, C. C. Loy, K. He and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295-307, Feb. 2016.

[17] P. Isola, J. Zhu, T. Zhou and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," *in Proc. IEEE CVPR2017*, USA, pp. 5967-5976, June. 2017.

[18] A. Cherian and A. Sullivan, "Sem-GAN: Semantically-Consistent Image-to-Image Translation," *in Proc. IEEE WACV2019*, USA, pp. 1797-1806, Oct. 2019.

[19] J. Zhu, T. Park, P. Isola and A. A. Efros, "Unpaired Image-to-Image Translation using Cycle-Consistent adversarial networks," *in Proc. IEEE ICCV2017*, Italy, pp. 2242-2251, Oct. 2017.

[20] K. Zhang, W. Zuo and L. Zhang, "FFDNet: Toward a Fast and Flexible Solution for CNN-Based Image Denoising," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4608-4622, Sept. 2018.

[21] K. Zhang, W. Zuo, Y. Chen, et.al., "Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142-3155, July. 2017.

[22] J. Lima, E. Neto. "Audio encryption based on the cosine number transform," *Kluwer Academic Publishers*, Oct. 2016, DOI:10.1007/s11042-015-2755-6.

[23] Q. N. Natsheh, B. Li and A. G. Gale, "Security of Multi-frame DICOM Images Using XOR Encryption Approach," *Procedia Computer Science*, vol. 90, pp. 175-181, July. 2016.

[24] M. Mukhedkar, P. Powar and P. Gaikwad, "Secure non real time image encryption algorithm development using cryptography & steganography," *in Proc. IEEE INDICON 2015*, India, pp. 1-6, July. 2015.

[25] A. Kanso and M. Ghebleh, "An efficient and robust image encryption scheme for medical applications," *Communications in Nonlinear Science and Numerical Simulation*, vol. 24, pp. 98-116, Jan. 2015.

[26] C. Fu, W. Meng, Y. Zhan, et.al., "An efficient and secure medical image protection scheme based on chaotic maps," *Computers in Biology and Medicine*, vol. 43, no. 8, pp. 1000-1010, May. 2013.

[27] W. Yu, C. Chi, X. Wei and X. Yang, "Image encryption algorithm based on high-dimensional chaotic systems," *in Proc. 2010 International Conference on Intelligent Control and Information Procession*, China, pp. 463-467, Nov. 2010.

[28] I. Goodfellow, J. Pouget-Abadie, M. Mirza, et.al., "Generative adversarial nets," *in Proc. NIPS2015*, Canada, pp. 2672-2680, Dec. 2015.

[29] J. Bao, D. Chen, F. Wen, H. Li and G. Hua, "CVAE-GAN: Fine-Grained Image Generation through Asymmetric Training," *in Proc. IEEE ICCV2017*, Italy, pp. 2764-2773, Oct. 2017.

[30] Y. Li and L. Shen, "cC-GAN: A Robust Transfer-learning Framework for HEp-2 Specimen Image Segmentation," *IEEE Access*, vol. 6, pp. 14048-14058, Oct. 2018.

[31] W. Liu, X. Liu, H. Ma and P. Cheng, "Beyond Human-level License Plate Super-resolution with Progressive Vehicle Search and Domain Priori GAN," *in Proc. the 25th ACM International Conference on Multimedia*, USA, pp. 1618-1626, Oct. 2017.

[32] Y. Zhou, L. Bao and C. L. P. Chen, "A new 1D chaotic system for image encryption," *Signal Process*, vol. 97, pp. 172-182, Oct. 2014.

[33] X. Liao, S. Lai and Q. Zhou, "A novel image encryption algorithm based on self-adaptive wave transmission," *Signal Process*, vol. 90, pp. 2714-2722, Mar. 2010.

[34] Z. Hua, Y. Zhou, C. Pun and C. L. P. Chen, "2D Sine Logistic modulation map for image encryption," *Information Sciences*, vol. 297, pp. 80-94, Feb. 2015.

[35] Y. Wu, J. Noonan and S. Agaian, "A wheel-switch chaotic system for image encryption," *in Proc. 2011 International Conference on System Science and Engineering*, Macao, pp. 23-27, May 2011.

[36] Y. Wu, G. Yang, H. Jin and J. Noonan, "Image encryption using the two-dimensional logistic chaotic map," *Journal of Electronic Imaging*, vol. 21, pp. 3014-3022, Jan. 2012.

[37] A. Cherian and A. Sullivan, "Sem-GAN: Semantically-Consistent Image-to-Image Translation," *in Proc. IEEE WACV2019*, USA, pp. 1797-1806, Oct. 2019.

[38] N. Wang, W. Zha, J. Li and X. Gao, "Back projection: An effective postprocessing method for GAN-based face sketch synthesis," *Pattern Recognition Letters*, vol. 107, pp. 59-65, June. 2018.

[39] Z. Yi, H. Zhang, P. Tan and M. Gong, "DualGAN: Unsupervised Dual Learning for Image-to-Image Translation," *in Proc. IEEE ICCV2017*, Venice, Mar. 2017, pp. 2868-2876.

[40] D. Arroyo, R. Rhouma, G. Alvarez, et.al., "On the security of a new image encryption scheme based on chaotic map lattices," *Chaos (Woodbury, N.Y.)*, vol. 18, pp. 1-8, Aug. 2008.

[41] F. Jiang, Y. Fu, B. Gupta, et.al., "Deep Learning based Multi-channel intelligent attack detection for Data Security," *IEEE Transactions on Sustainable Computing*, DOI: 10.1109/TSUSC.2018.2793284.

[42] D. Chen, N. Zhang, et. al., "An LDPC code based physical layer message authentication scheme with prefect security", *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 4, pp. 748-761, 2018.

[43] A. Ferdowsi and W. Saad, "Deep Learning for signal authentication and security in massive Internet of Things Systems," *IEEE Transactions on Communications*, vol. 67, no. 2, pp. 1371-1387, Feb. 2019.

[44] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," *in Proc. IEEE CVPR2016*, USA, pp. 770-778, Sept. 2016.

[45] S. Jaeger, S. Candemir, S. Antani, et.al., "Two public chest X-ray datasets for computer-aided screening of pulmonary diseases," *Quantitative imaging in medicine and surgery*, vol. 4, pp. 475-7, Dec. 2014.