IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 26, 2019, accepted December 7, 2019, date of publication December 12, 2019,
date of current version December 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2959031

Deep Learning for Improving the Robustness of Image Encryption

JING CHEN[1], XIAO-WEI LI[1], AND QIONG-HUA WANG[2]
[1]School of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China
[2]School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing 100191, China

Corresponding author: Qiong-Hua Wang (qionghua@buaa.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 61975138 and Grant 61705146.

ABSTRACT In this paper, we propose a method to increase the robustness of 2D/3D optical image encryption using the dilated deep convolutional neural network (CNN). In order to solve the problem that encrypted images suffer from some attacks in practical application, we utilize a fast and effective CNN denoiser based on the principle of deep learning. The CNN improves the robustness of the algorithm by improving the resolution of the reconstructed images. Besides, CNN has a high performance against blur and occlusion attacks. We introduce the pixel scrambling method to enhance the security level of the encryption by the private key of pixel scrambling operation. The proposed method can not only realize the encryption of a two-dimensional image but also implement three-dimensional image encryption by combining the integral imaging technology. Double random phase encoding in the fractional Fourier domain is selected for experimental verification, and the results show the capability for robustness, noise immunity, and security of the proposed method.

INDEX TERMS Optical image encryption, integral imaging, fractional Fourier transform, convolutional neural network.

I. INTRODUCTION

With the rapid development of multimedia technology, it is of considerable significance to take efficient and high-security measures to protect the information at the same time. Because of the widespread use of images, image encryption becomes a focus of information security research [1]–[10]. In recent years, more and more image encryption researchers attach importance to the optical information processing technology due to its characteristics of high parallelism, substantial parameter freedom, high latitude, great data capacity, and high security. Furthermore, many mature and capable optical image encryption technologies have emerged. In [11], P. Refregier and B. Javidi proposed the double random phase encoding (DRPE) method based on 4f system. Then many scholars have extended it to its fractional Fourier transform (FrFT) domain, Fresnel transform domain, gyrator transform domain, and have proposed the corresponding image encryption scheme [12]–[16]. In [17], Unnikrishnan *et al.* have proposed an optical image encryption method based on random phase encoding in the FrFT domain. FrFT not only has the excellent properties of traditional Fourier transform but also provides additional keys, the scale factors, and the fraction orders, for the encryption system, which makes the encryption system obtain high security.

In practice, encryption systems suffer from various attacks, the most common of which is noise attack, and the quality of the decrypted image will be reduced. Thus, researchers have put forward numerous solutions to this problem [18]–[24] such as filtering based methods, spare models, nonlocal self-similarity models (NSS), and Markov random field models. Among them, the NSS models are considered as the state-of-the-art methods. And some typical methods include block-matching and three-dimensional (BM3D) filtering, learned sparse coding, nonlocally centralized sparse representation and weighted nuclear norm minimization. In recent, some developed convolutional neural network (CNN) algorithms like image restoration CNN provide exceptional performance in denoising [25]–[29]. Despite the high denoising quality of these methods, they inevitably have two major shortcomings. One of them is involving complex optimization problems and time-consuming in the test process. And the other one is that these models are usually non-convex with multiple artificial hypothesis parameters.

The associate editor coordinating the review of this manuscript and approving it for publication was Mohammad Shorif Uddin.

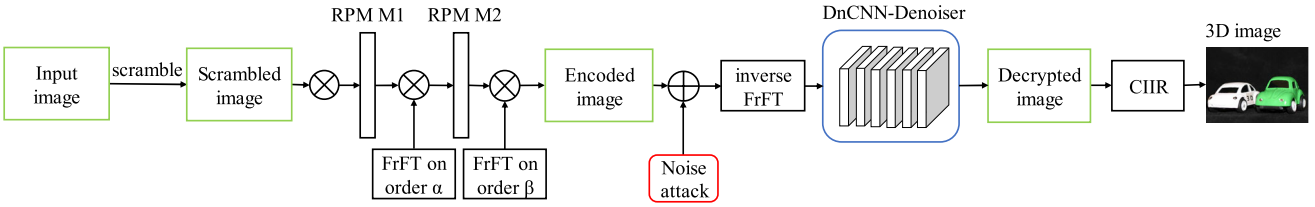

FIGURE 1. Schematic of the proposed encryption method.

To overcome these disadvantages, researchers propose some discriminative learning methods like multi-layer perceptron (MLP). In [30], Zhang *et al.* regard image denoising as a maximum a posteriori problem from the Bayesian perspective and use the deep CNN to learn the prior as a denoiser.

Due to the wide application of three-dimensional (3D) imaging and display, more attention has been paid to the research of 3D image encryption. In [31], Lippmann proposed integral imaging (II). We can obtain a two-dimensional (2D) image array named elemental image (EI) array which contains different perspective information of a 3D scene by using a lenslet array and charge-coupled device (CCD) sensor. And the 3D image can be recovered from the EIs based on computational integral imaging reconstruction (CIIR) algorithm. Because of the distributed memory characteristic of EI, 3D image encryption owns high robustness.

The key contributions of the proposed method are summarized as follow: we introduce an effective CNN denoiser into the encryption to separate the noise from the noised decrypted image; the deep learning method not only can against noise attack but also against blur and occlusion attacks; the CNN uses residual learning algorithm to enhance the accuracy and performance of the model; parametric rectified linear unit (ReLU) and batch normalization (BNorm) method are introduced to improve the speed and the performance of CNN; we use dilated filter to increase the size of receptive field while reducing the network depth as much as possible; the scrambling algorithm is added in the encryption process in order to further improve the security of the encryption system; the encryption system can realize not only 2D but also 3D image.

II. THEORETICAL ANALYSIS OF THE PROPOSED METHOD

The schematic of the proposed encryption method is illustrated in Fig. 1. It can be divided into three parts: encryption process, CNN denoising, and the reconstruction process. In the first process, the input image is preprocessed via pixel position scrambling technology. The rules of dividing, numbering, and scrambling are used as private keys that can provide higher security for the encryption system. Then, the double random phase encoding in the fractional Fourier domain is implemented on the scrambled image to get the encoded image. The decryption is the inverse process of the encryption process.

In the second process, CNN denoising is implemented. The denoising can be considered as an inverse problem whose purpose is to recover the latent clean image x from the noisy image y who follows an image degradation model $y = x + v$, where v is additive white Gaussian noise with standard deviation σ. We learn a denoising convolutional neural network utilizes residual learning approach, whose output is the residual image $\hat{v}$ instead of the clean image x. Thus, we can get the approximated clean image as

$$\hat{x} = y - \hat{v}. \tag{1}$$

In the third process, the CIIR algorithm is added to reconstruct the 3D image. It needs to be emphasized that the input image of 3D image encryption is the EI obtained by lenslet array and CCD sensor.

Some key technologies involved in our proposed encryption method, as well as some specific parameter designs, are described in the following parts of this section.

A. OPTICAL IMAGE ENCRYPTION BASED ON FrFT

The FrFT is an extension of the traditional FT on the order of the continuous FT whose time-frequency analysis characteristic breaks the limitation of the traditional FT. We define the αth-order FrFT of the input image $f(x)$ in the one-dimensional (1D) case as follows

$$F^{\alpha}\{f(x)\} = \int_{-\infty}^{+\infty} B_{\alpha}(x, x') f(x')\, dx'. \tag{2}$$

The transform kernel can be expressed as follows

$$B_{\alpha}(x, x') = A_{\theta} \exp\left[i\pi\left(x^2 \cot\theta - 2xx' \csc\theta + x'^2 \cot\theta\right)\right], \tag{3}$$

$$A_{\theta} = \frac{\exp\left(-\frac{i\pi\, sgn(\sin\theta)}{4} + \frac{i\theta}{2}\right)}{\sqrt{|\sin\theta|}}, \tag{4}$$

where

$$\theta = \frac{\alpha\pi}{2}, \tag{5}$$

and i is the imaginary unit. Where $\alpha \neq 0$ or ± 2.

Here is a brief description of the encryption process: At first, scramble the input image $f(x)$ to obtain $C(x)$. Then, multiply $C(x)$ with the first random phase mask (RPM) ($M1 = \exp[i2\pi n_1(x)]$) and implement FrFT on order α.

Later, multiply the second RPM ($M2 = \exp[i2\pi n_2(x)]$) and carry out the FrFT on an order β to get the encoded image

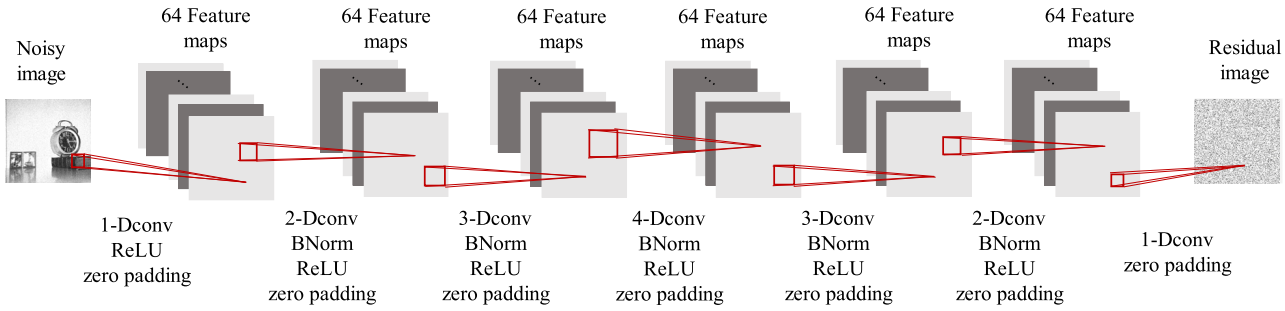

FIGURE 2. Structure of the CNN denoiser.

as follow

$$\xi(x) = F^{\beta}\left\{F^{\alpha}\left\{C(x)\exp\left[i2\pi n_1(x)\right]\right\}\exp\left[i2\pi n_2(x)\right]\right\}, \tag{6}$$

where $n_1(x)$ and $n_2(x)$ are two independent random matrices.

The decryption is the inverse process of the encryption process. We can obtain the $C(x)$ and perform inverse pixel scrambling transformation on it to get the original image $f(x)$. The $C(x)$ is as

$$C(x) = F^{-\alpha}\left\{F^{-\beta}\left\{\xi(x)\exp\left[-i2\pi n_2(x)\right]\right\} \times \exp\left[-i2\pi n_1(x)\right]\right\}. \tag{7}$$

B. DECRYPTED IMAGE DENOISING

The structure of CNN denoiser we proposed in this paper is shown in Fig. 2. We set the depth of the network to 7. The first layer contains ''Dilated Convolution (Dconv) + ReLU''. The middle five layers include ''Dconv + BNorm + ReLU'', and each layer produces 64 feature maps. The last layer has ''Dconv''. The red square frame in every layer means the kernel of the dilated convolution, and the sizes of them are 3×3, 5×5, 7×7, 9×9, 7×7, 5×5 and 3×3. Zero padding is implemented before Dconv to ensure that each feature map and input image are the same sizes.

1) RESIDUAL LEARNING STRATEGY

The depth is an essential parameter of CNN, to a certain extent, better results can be obtained by adding more network layers. However, with the increase of depth, the rate of accuracy will quickly reach saturation and then decline rapidly. Hence, we introduce residual learning strategy into CNN.

Assuming that the network is designed to represent the mapping $H(x) = F(x) + x$ which can be converted to learn a residual function $F(x) = H(x) - x$. So the original function $\hat{x}$ is as

$$\hat{x} = H(x) - F(x). \tag{8}$$

We suggest that it may be more convenient to optimize the residual map than the original reference map. Furthermore, if a recognition map is optimized, it is easier to make its residual value approach zero than to use a stack of nonlinear combinations to fit an identity map.

As verified by experiments in [32], the deep residual learning framework significantly enhances the accuracy and performance of the model. Just because of this, we utilize the residual learning approach to denoise the decrypted images.

2) DILATED CONVOLUTION

Deep CNN has some fatal defects, especially the design of the pooling layer. The pooling layer added into the network will lose information and reduce the accuracy. However, without the pooling layer, the receptive field will become smaller, and the global features will not be learned. If we simply remove the pooling layer and expand the convolution kernel, it will lead to an increase in the burden of computation. Consider these reasons, we use the dilated convolution to achieve tradeoffs between the network depth and the size of the receptive field. A dilated filter with dilation factor s can be regarded as a sparse filter of size $(2s+1)\times(2s+1)$, in which only the components of 9 fixed positions can be non-zeros.

There are two potential problems in the dilated convolution. The first one is the gridding effect. Assuming that we repeatedly use several 3×3 convolution kernels with dilation factor 2, the problem of discontinuity of convolution kernels will occur, and the continuity of information will be lost. The second one is that long-ranged information might be not relevant. In order to solve the problems mentioned above, a solution is proposed [33] and named as Hybrid Dilated Convolution design structure.

In the proposed CNN, we set the dilation factors of 3×3 Dconv in every layer as 1, 2, 3, 4, 3, 2 and 1 in turn. So that the equivalent size of the receptive field of each layer is 3×3, 7×7, 13×13, 21×21, 27×27, 31×31 and 33×33. If we use the traditional 3×3 convolution filter to achieve the same size as the receptive field, we need to design a network in depth of 16. In other words, the increase of the receptive field size is linearly related to the depth of the network if we use traditional convolution, while it increases exponentially in utilizing dilated convolution.

3) BATCH NORMALIZATION

When we train the neural network, the standardized input can improve the training speed. Hence, we combine batch normalization with a convolutional neural network.

Suppose there are N training samples in the current batch processing, and each batch has d dimension, which means the input is $x = \{x_1, x_2, \cdots, x_d\}$. Then, each dimension of x is normalized by

$$\hat{x}_k = \frac{x_k - E(x_k)}{\sqrt{Var(x_k)}}, \tag{9}$$

where $E(x_k)$ is an expectation of x_k, and $Var(x_k)$ is the variance of x_k. However, it is not simple to input the data to the next layer after normalization because it will affect the features learned in the current layer. Therefore, it is necessary to add a transformation and reconstruction, we introduce two parameters γ_k and η_k to scale and shift the normalized value as

$$y_k = \gamma_k \hat{x}_k + \eta_k, \tag{10}$$

where $\gamma_k = \sqrt{Var(x_k)}$ and $\eta_k = E(x_k)$. In such formulation, the convolution network gets the benefits of batch normalization such as higher robustness, faster training speed, better performance, and lower sensitivity to initialization.

4) TRAINING

It is generally accepted that CNN performs better with the more extensive training dataset. So that, we choose Waterloo Exploration Database (WED) which collects 4744 images as our training dataset.

The loss function is a standard to judge the network performance, the smaller the loss function is, the better performance the network has. We set the loss function as

$$L(\Theta) = \frac{1}{2N} \sum_{i=1}^{N} \|f(y_i, \Theta) - (y_i - x_i)\|^2, \tag{11}$$

where N denotes N pairs training data, y_i is the i-th input of noised image and x_i is the i-th corresponding clean image. The network parameters Θ is set by Adam solver, and the other hyper-parameters of Adam choose their default values.

In the network training, the mini-batch size is set as 64. The simulation experiments are implemented in Matlab (R2015b) with the Matconvnet toolbox on a PC with Intel(R) Core(TM) i5-4460 CPU 3.2 GHz and NVIDIA GeForce GTX 1050 Ti.

C. 3D IMAGE COMPUTATIONAL RECONSTRUCTION

The input image of 3D image encryption is the EIs, and the setup for EIs generation is illustrated in Fig. 3. The 3D objects (“Cars”) are recorded as EIs by a CCD camera through a lenslet array. The EIs are a series of 2D elemental images, and each EI contains different perspective information of the 3D objects. We realize the encryption of 3D images by reducing the dimension of 3D to 2D. In other words, we implement the proposed encryption method on each 2D EI and recover the 3D image from the encrypted EIs. Because of the distributed memory characteristic of EIs, 3D image encryption obtains high robustness against data loss attacks.

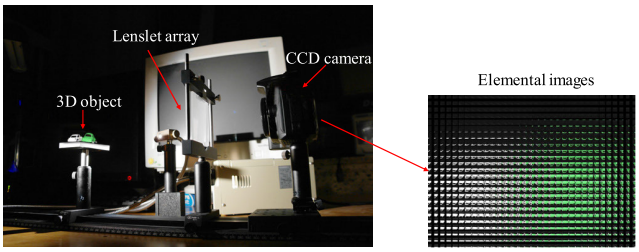

FIGURE 3. The setup for EIs generation.

The 3D image can be digitally reconstructed by the CIIR algorithm. Assume that the reconstructed 3D image and each EI have the same resolution, the reconstructed 3D images at the depth z can be formulated as follow

$$\Re(x, y, z) = \frac{1}{\psi(x, y)} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} E_{m,n}(x - m\frac{M \times p}{c_x \times \gamma}, y - n\frac{N \times p}{c_y \times \gamma}), \tag{12}$$

where $\Re(x, y, z)$ means the pixel intensity of the reconstructed 3D image at the distance z, $\psi(x, y)$ is the superimposed number matrix, M and N represent that each EI has M columns and N rows, $E_{m,n}$ denotes the (m,n)th monospectral elemental image, c_x and c_y are the size of imaging sensor, p is the pitch of each micro-lens, γ represents the magnification factor and $\gamma = z/g$, where g is the focal length.

III. EXPERIMENT RESULTS AND PERFORMANCE ANALYSIS

In order to show the feasibility and the performance of the proposed method, several experiments are conducted. The simulation parameters of 2D image tests are described as follows. The grayscale images of “Clock” and “House” with a size of 256 × 256 pixels are used as input images respectively to be encrypted, as shown in Figs. 4(a) and (e). It is divided into 16384-pixel blocks with the size of 2 × 2 pixels, the scrambled image is obtained after pixel scrambling process, and present as noise distribution, the input image content cannot be recognized directly from them, as shown in Figs. 4(b) and (f). Then the random function is used to generate two random phase masks on the MATLAB simulation platform. The fractional orders are given by $\alpha_x = 1.5$, $\alpha_y = 1.2$, $\beta_x = 1.4$, $\beta_y = 1.6$, α_x and β_x are for the x-direction while α_y and β_y are for the y-direction. The encrypted images are shown in Figs. 4(c) and (g). Figures 3(d) and (h) present the decrypted images using correct keys, we can see that when there is no noise or distortion, the decryption results are almost the same as the original images.

To better simulate the real situation of noise attack on the encryption system, we attacked the encrypted image after the encryption process artificially. The manually introduced noise is the Gaussian noise with different variances.

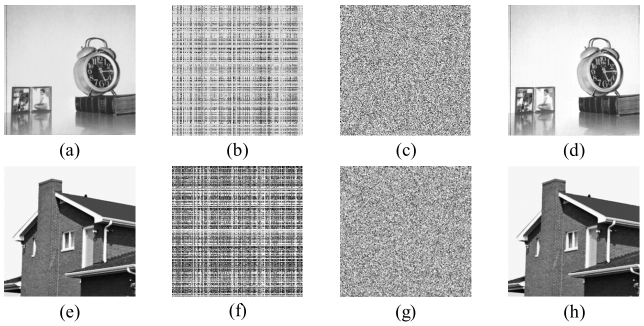

FIGURE 4. Encryption and decryption results: (a) and (e) input images, (b) and (f) scrambled images, (c) and (g) encrypted images, (d) and (h) decryption results using correct keys.

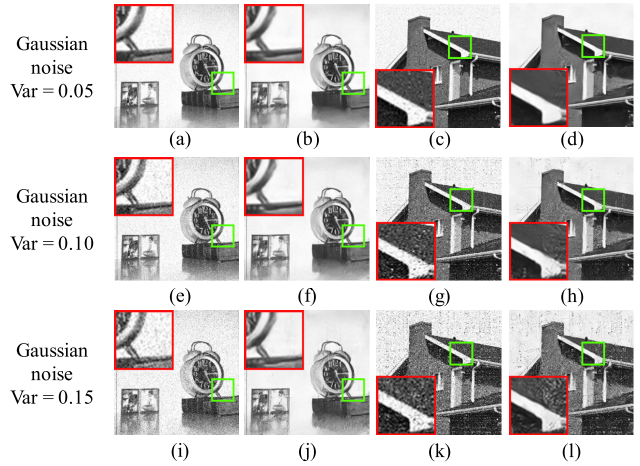

FIGURE 5. Output results with Gaussian noise attack: (a)-(d) decryption images with Gaussian noise variance of 0.05, (e)-(h) decryption images with Gaussian noise variance of 0.10, (i)-(l) decryption images with Gaussian noise variance of 0.15, (a), (c), (e), (g), (i) and (k) decryption images without the CNN Denoiser, (b), (d), (f), (h), (j) and (l) reconstructed images using the CNN Denoiser.

Fig. 5 shows the decryption results without the CNN Denoiser and the reconstructed results using the CNN Denoiser. And we compare the proposed methods with the state-of-the-art algorithms, one is model-based optimization method BM3D in [22] and the other is discriminative learning method MLP in [22]. The results are illustrated in Fig. 7.

We use the peak signal-to-noise ratio (PSNR) and mean structural similarity (SSIM) as indicators to evaluate the quality of the reconstructed images. PSNR works based on the error between the corresponding pixel of the reconstructed image compared with the attacked image. PSNR can be defined as

$$PSNR = 10\log_{10}\frac{(2^n-1)^2}{MSE}, \tag{13}$$

where n is the number of bits per pixel, the grayscale image is 8, that is, the gray image of the pixel of 256, and the mean square error (MSE) is calculated as

$$MSE = \frac{1}{\mathrm{H}\times\mathrm{W}}\sum_{i=1}^{H}\sum_{j=1}^{W}(X(i,j)-Y(i,j))^2, \tag{14}$$

where $H\times W$ represents the size of each of the images, $X(i,j)$ and $Y(i,j)$ mean the attacked and reconstructed images respectively.

Generally, the larger the PSNR value, the smaller the distortion and the better the reconstruction effect. Sometimes, however, the image quality reflected by PSNR does not match the actual human subjective visual perception. When white noise is added to the high, middle and low frequency area of the same image respectively, the quality of the image which is noised in the high frequency area is better than that of the other two, but the PSNR values of the three are the same. Therefore, we also calculated the SSIM to better evaluate the image quality.

SSIM is an index used to measure the similarity between two images, its value ranges from 0 to 1. SSIM is highly consistent with human subjectivity. The SSIM algorithm divides the whole spatial area into blocks. One SSIM value reflects the quality of all pixels in a block, and all SSIM values constitute a SSIM map which reflects the overall quality of the image. The mean SSIM with a high value has more white pixels in its SSIM map, corresponds to a high similarity between the attacked image and the reconstructed image. SSIM can be defined as

$$SSIM(x,y) = \frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{\left(\mu_x^2+\mu_y^2+c_1\right)+\left(\sigma_x^2+\sigma_y^2+c_2\right)}, \tag{15}$$

$$c_1 = (K_1L)^2, \quad c_2 = (K_2L)^2, \tag{16}$$

where μ_x and μ_y represent the mean of x and y, σ_x and σ_y denote the variance of x and y, σ_{xy} is the covariance, L is the dynamic range of the images, and it defaults to 255, K_1 and K_2 are two constants and the default values are 0.01 and 0.03 respectively.

The calculated PSNR and mean SSIM values of the result images by different denoised methods are recorded in Table 1 and Table 2. And the SSIM maps of each image in Fig. 5 are shown in Fig. 6. From Table 1, we can calculate the PSNR values of the reconstructed image "Clock" and "House" are increased 8.9512% and 15.5536% on an average respectively by the proposed CNN Denoiser, while the increment rates by the BM3D are 4.6066% and 10.8944%, and the rates by the MLP are 2.9600% and 12.6195%. We noticed that the PSNR value of the noised decrypted image is the highest in the first column in Table 1, which is caused by the fact that PSNR is not completely consistent with human subjective visual perception. So, we need to evaluate the image quality in combination with the mean SSIM values. The mean SSIM values of the reconstructed image "Clock" and "House" recorded in Table 2 are improved 70.6588% and 45.6067% on an average respectively, while the rates by the BM3D are 56.6344% and 34.7428%, and the rates by the MLP are 24.1649% and 20.8622%. Moreover, from the comparison of the corresponding SSIM map, we can directly see the difference in image quality before and after reconstruction, it is evident that the method we proposed has a good effect.

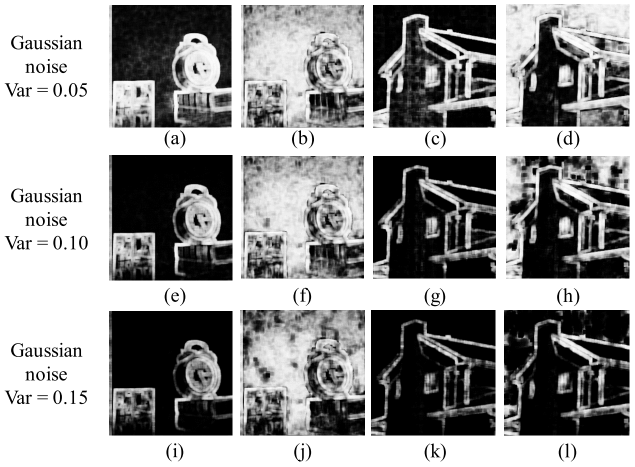

FIGURE 6. SSIM map of each image in FIGURE 5.

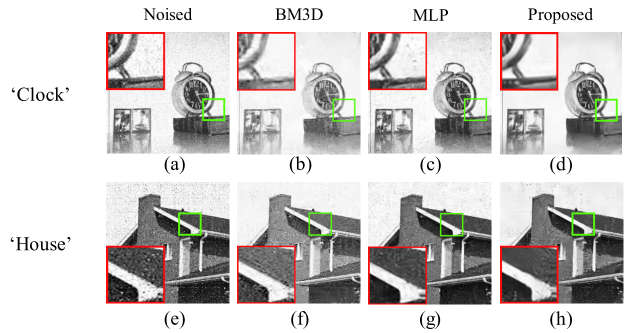

FIGURE 7. Denoised results with Gaussian noise attack by different methods: (a) and (e) decryption images with Gaussian noise variance of 0.10, (b) and (f) denoised by BM3D, (c) and (g) denoised by MLP, (d) and (h) denoised by the proposed CNN Denoiser.

TABLE 1. The PSNR values of result image by different denoising methods.

Variance	0.05		0.10		0.15	
	'Clock'	'House'	'Clock'	'House'	'Clock'	'House'
Noised	**29.5142**	24.3669	23.9238	18.9525	20.7410	16.0191
BM3D	28.0321	26.7814	25.7293	21.8373	23.8347	17.1844
MLP	26.6664	25.3505	26.0198	**23.2613**	23.6885	18.2149
Proposed	28.1825	**27.1840**	**27.3020**	22.8624	**25.3344**	**18.5214**

TABLE 2. The mean SSIM values of result image by different denoising methods.

Variance	0.05		0.10		0.15	
	'Clock'	'House'	'Clock'	'House'	'Clock'	'House'
Noised	0.6870	0.5636	0.4687	0.3695	0.3531	0.2801
BM3D	0.8391	0.7544	0.7862	0.5532	0.7380	0.3271
MLP	0.7011	0.6260	0.6579	0.5144	0.5114	0.3259
Proposed	**0.8885**	**0.7769**	**0.8729**	**0.6118**	**0.8135**	**0.3778**

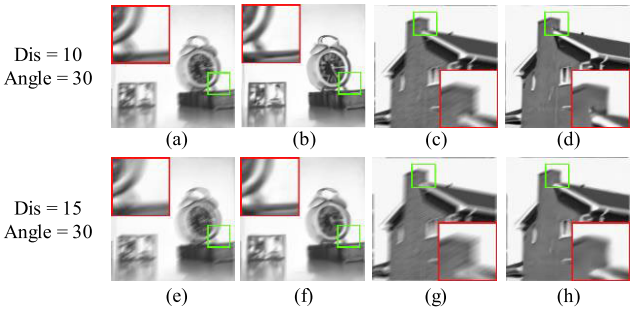

FIGURE 8. Output results with blur attack with an angle at 30°: (a)-(d) decryption images with blur distance of 10, (e)-(h) decryption images with blur distance of 15, (a), (c), (e) and (g) decryption images without the proposed method, (b), (d), (f) and (h) decryption images using the proposed method.

TABLE 3. The PSNR values of each image under different noise attacks.

Noise	'Salt&Pepper'		'Rayleigh'		'uniform'	
	'Clock'	'House'	'Clock'	'House'	'Clock'	'House'
Noised	23.6682	18.9124	24.5107	19.4509	**28.3600**	23.0991
Proposed	**27.2147**	**22.9339**	**27.4266**	**23.6980**	28.2124	**26.8724**

TABLE 4. The mean SSIM values of each image under different noise attacks.

Noise	'Salt&Pepper'		'Rayleigh'		'uniform'	
	'Clock'	'House'	'Clock'	'House'	'Clock'	'House'
Noised	0.4582	0.3716	0.4913	0.3880	0.6452	0.5206
Proposed	**0.8722**	**0.6036**	**0.8768**	**0.6451**	**0.8897**	**0.7677**

Besides the Gaussian noise attack, we also carry out other noise attacks, Salt & Pepper noise, Rayleigh noise, and uniform noise attacks. Moreover, we calculate the PSNR and the mean SSIM values. The experimental results are recorded in Table 3 and Table 4, and specifically, indicate the high quality of the reconstructed images of the proposed method. Besides the noise attack experiments, blur and occlusion attacks are implemented. We choose motion blur with the angle at 30° and different distances (Dis = 10 and Dis = 15), and the output results are shown in Fig. 8. We can see from the contrast images that the proposed method also has a good effect in image deblurring. The experimental results of the proposed method against occlusion attacks are shown in Fig. 9. Even if 40% and 60% pixels in the encrypted images are lost, the proposed method can still decrypt the input images from them. From the above, the proposed encryption method has high robustness against some classical types of attacks.

Here, the security of the proposed method is analyzed experimentally. The security keys include the pixel scrambling rule, the random phase mask: $M1$ and $M2$, and

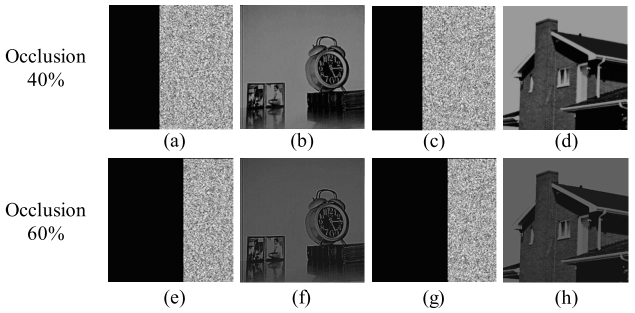

FIGURE 9. Output results with occlusion attack: (a) and (c) 40% pixels from encrypted images are occluded, (b) and (d) decryption images from (a) and (c) using the proposed method, (e) and (g) 60% pixels from encrypted are occluded. (f) and (h) decryption images from (e) and (g) using the proposed method.

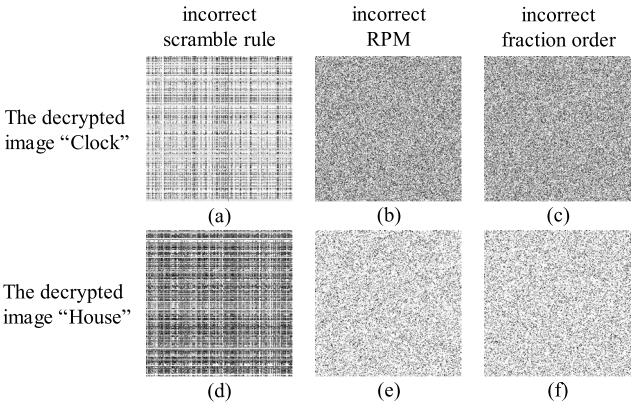

FIGURE 10. Decryption results with incorrect keys: (a)-(c) the decrypted images "Clock", (d)-(f) the decrypted images "House", (a) and (d) with incorrect scramble rule, (b) and (e) with randomly generated different RPMs, (c) and (f) with incorrect fraction order (0.5, 0.5, 0.5, 0.5).

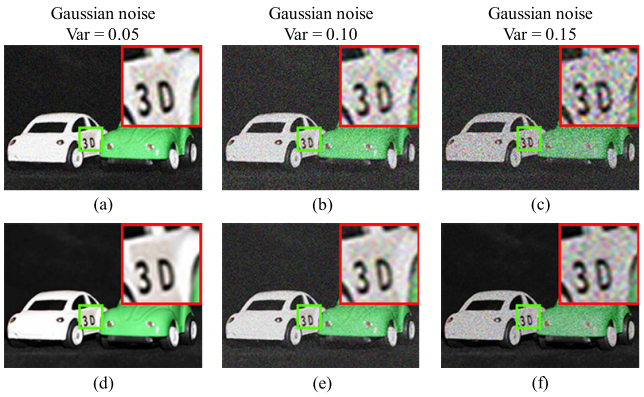

FIGURE 11. Output results with Gaussian noise attack: (a) and (d) decryption images with Gaussian noise variance of 0.05, (b) and (e) decryption images with Gaussian noise variance of 0.10, (c) and (f) decryption images with Gaussian noise variance of 0.15, (a)-(c) decryption images without the CNN Denoiser, (d)-(f) reconstructed images by using the CNN Denoiser.

the fraction order of FrFT: α_x, α_y, β_x, β_y. We simulate the decryption results of these keys in the case of errors respectively. Figs. 10(a) and (d) represent the decrypted image with a wrong scrambling rule which is generated

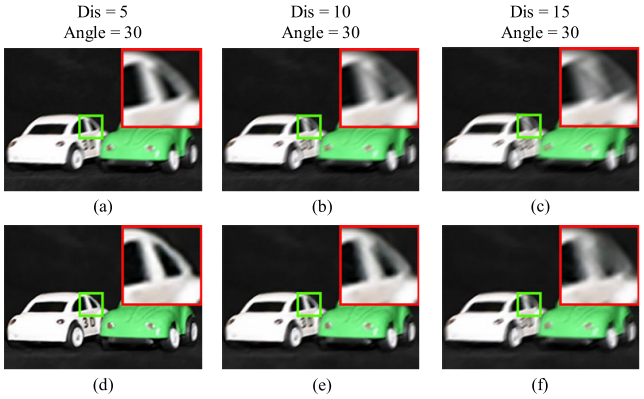

FIGURE 12. Output results with blur attack with an angle at 30°: (a) and (d) decryption images with blur distance of 5, (b) and (e) decryption images with blur distance of 10, (c) and (f) decryption images with blur distance of 15, (a)-(c) decryption images without the proposed method, (d)-(f) decryption images using the proposed method.

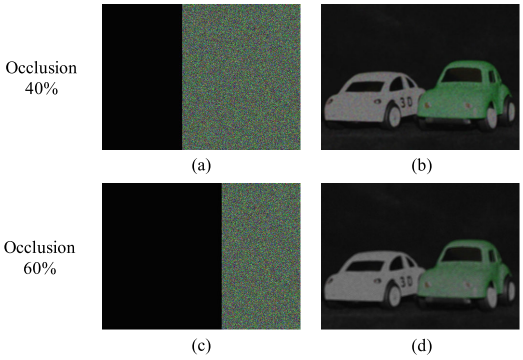

FIGURE 13. Output results with occlusion attack: (a) 40% pixels from encrypted images are occluded, (b) decryption images from (a) using the proposed method, (c) 60% pixels from encrypted are occluded. (d) decryption images from (c) using the proposed method.

randomly. The decrypted results of using the incorrect RPMs $M3 = \exp[i2\pi n_3(x)]$ and $M4 = \exp[i2\pi n_4(x)]$ are shown in Figs. 10(b) and (e), where $n_3(x)$ and $n_4(x)$ are the two independent random matrices different from $n_1(x)$ and $n_2(x)$. Figs. 10(c) and (d) are the results with incorrect fraction order (0.5, 0.5, 0.5, 0.5). It is obvious that when any of the keys is wrong, the correct decrypted images cannot be obtained.

For the 3D image, the denoised results are shown in Fig. 11. To make the conclusion more reliable, we implement some Gaussian noise attacks with different noise intensity (variance of 0.05, 0.10 and 0.15). In Table 5. We can calculate the PSNR values of the reconstructed image are increased 27.4885% on an average by the proposed method. In addition, blur attacks with angle at 30° and different distances (Dis = 5, Dis = 10 and Dis = 15) are carried out and shown in Fig. 12. The PSNR values are improved 3.9056% on an average according the records in Table 6. Moreover, we also accomplish the occlusion attacks with 40% and 60% pixels occluded, and Fig. 13 represents the feasibility of the proposed method against the occlusion attack.

TABLE 5. The PSNR values of each image in Figure 11.

Variance	0.05			0.10			0.15		
channel	R	G	B	R	G	B	R	G	B
Noised	18.1725	19.9512	19.1599	14.5707	15.4417	14.9583	13.7407	14.4732	13.9887
Proposed	**20.7697**	**24.3925**	**22.9074**	**18.2794**	**21.2526**	**19.1138**	**18.1374**	**20.2834**	**19.0291**

TABLE 6. The PSNR values of each image in Figure 12.

Distance	5			10			15		
Channel	R	G	B	R	G	B	R	G	B
Blurred	20.0235	23.1342	21.7265	19.3524	21.9504	20.7355	18.7222	20.9065	19.8554
Proposed	**20.5495**	**24.0891**	**22.5609**	**20.0028**	**23.1193**	**21.3478**	**19.3478**	**21.9692**	**20.7212**

From the above results, we can see that the proposed method is useful to improve the quality of the decrypted image after noise attacking. The proposed method can not only deal with Gaussian noise attacks but also against other kinds of noise attacks. Furthermore, the proposed method utilizes pixel scramble to enhance the security level by the private key of this operation, and it takes advantage of the FrFT to enhance the key space of the encryption system.

IV. CONCLUSION

In this paper, a deep learning method to improve the robustness of 2D/3D image encryption is proposed. We adopt DRPE in the fractional Fourier domain, and introduce the pixel position scrambling method to increase the security of the encryption system. Aiming at the actual problem of noise attacking in encryption, we utilize CNN with a residual learning approach to restoring the original image from the attacked decrypted image. Meanwhile, the batch normalization and dilated convolution are utilized in CNN to improve the performance. Experimental results show that the proposed method is effective against noise, blur and occlusion attacks. Since color images and other media types like video attack more attention in image processing, we will carry out relevant research in the future.

REFERENCES

[1] O. Matoba and B. Javidi, "Encrypted optical memory system using three-dimensional keys in the Fresnel domain," *Opt. Lett.*, vol. 24, no. 11, pp. 762–764, 1999.
[2] W. Chen, "Hierarchically optical double-image correlation using 3D phase retrieval algorithm in fractional Fourier transform domain," *Opt. Commun.*, vol. 427, pp. 374–381, 2018.
[3] G. Situ and J. Zhang, "Double random-phase encoding in the Fresnel domain," *Opt. Lett.*, vol. 29, no. 14, pp. 1584–1586, Jul. 2004.
[4] X. Li, Y. Wang, and Q.-H. Wang, "Modified integral imaging reconstruction and encryption using an improved SR reconstruction algorithm," *Opt. Lasers Eng.*, vol. 112, pp. 162–169, Jan. 2019.
[5] Y. Qin, Z. Wang, and Q. Pan, "Optical color-image encryption in the diffractive-imaging shceme," *Opt. Laser Eng.*, vol. 77, pp. 191–202, 2016.
[6] X. Wang, X. Zhu, and Y. Zhang, "An image encryption algorithm based on Josephus traversing and mixed chaotic map," *IEEE Access*, vol. 6, pp. 23733–23746, 2018.
[7] W. He, X. Peng, and X. Meng, "Collision in optical image encryption based on interference and a method for avoiding this security leak," *Opt. Laser Technol.*, vol. 47, pp. 31–36, 2013.
[8] X. Li, D. Xiao, and Q. Wang, "Error-free holographic frames encryption with CA pixel-permutation encoding algorithm," *Opt. Laser Eng.*, vol. 100, pp. 200–207, 2018.
[9] Z. Hua, B. Xu, F. Jin, and H. Huang, "Image encryption using josephus problem and filtering diffusion," *IEEE Access*, vol. 7, pp. 8660–8674, 2019.
[10] Y. Luo, S. Tang, X. Qin, L. Cao, F. Jiang, and J. Liu, "A double-image encryption scheme based on amplitude-phase encoding and discrete complex random transformation," *IEEE Access*, vol. 6, pp. 77740–77753, 2018.
[11] P. Refregier and B. Javidi, "Optical-image encryption based on input plane and Fourier plane random encoding," *Opt. Lett.*, vol. 20, no. 7, pp. 767–769, 1995.
[12] N. Zhou, H. Jiang, L. Gong, and X. Xie, "Double-image compression and encryption algorithm based on co-sparse representation and random pixel exchanging," *Opt. Lasers Eng.*, vol. 110, pp. 72–79, Nov. 2018.
[13] L. Sui, M. Xin, and A. Tian, "Multiple-image encryption based on phase mask multiplexing in fractional Fourier transform domain," *Opt. Lett.*, vol. 38, no. 11, pp. 1996–1998, 2013.
[14] M. Abuturab, "An asymmetric single-channel color image encryption based on Hartley transform and gyrator transform," *Opt. Laser Eng.*, vol. 69, pp. 49–57, 2015.
[15] W. Chen, "Optical multiple-image encryption using three-dimensional space," *IEEE Photon. J.*, vol. 8, no. 2, Apr. 2016, Art. no. 6900608.
[16] A. Muhammad, "Multiple information encryption by user-image-based gyrator transform hologram," *Opt. Laser Eng.*, vol. 92, pp. 76–84, 2017.
[17] G. Unnikrishnan, J. Joseph, and K. Singh, "Optical encryption by double-random phase encoding in the fractional Fourier domain," *Opt. Lett.*, vol. 25, no. 12, pp. 887–889, 2000.
[18] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 2005, pp. 60–65.
[19] J.-L. Starck, E. J. Candes, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Electron. Packag. Manuf.*, vol. 11, no. 6, pp. 670–684, Jun. 2002.
[20] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.
[21] M. Elad and M. Aharon, "Image denoising via learned dictionaries and sparse representation," *IEEE Trans. Image Process*, vol. 15, no. 12, pp. 3736–3745, 2006.
[22] H. Burger, C. Schuler, and S. Harmeling, "Image denoising: Can plain neural networks compete with BM3D?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2392–2399.
[23] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, nos. 1–4, pp. 259–268, 1992.

[24] F. Liu, G. Hu, C. Chen, W. Chen, and C. Song, "Significant dynamic range and precision improvements for FMF mode-coupling measurements by utilizing adaptive wavelet threshold denoising," *Opt. Commun.*, vol. 426, pp. 287–294, 2018.
[25] X. Wang and Z. Li, "A color image encryption algorithm based on Hopfield chaotic neural network," *Opt. Laser Eng.*, vol. 115, pp. 107–118, 2019.
[26] Y. Tai, J. Yang, X. Liu, and C. Xu, "MemNet: A persistent memory network for image restoration," in *Proc. IEEE Inter. Conf. Comput. Vis.*, Oct. 2017, pp. 4549–4557.
[27] C. Schuler, H. Burger, S. Harmeling, and B. Scholkopf, "A machine learning approach for non-blind image deconvolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1067–1074.
[28] C. Dong, C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, vol. 8692, 2014, pp. 184–199.
[29] C. Liu, Z. Shang, and A. Qin, "A multiscale image denoising algorithm based on dilated residual convolution network," 2018, *arXiv:1812.09131*. [Online]. Available: https://arxiv.org/abs/1812.09131
[30] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2808–2817.
[31] G. Lippmann, "Reversible test prints. Integral photographies," *Acad. Sci.*, vol. 146, pp. 446–451, 1908.
[32] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian Denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.
[33] K. He, X. Zhang, and S. Ren, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

JING CHEN received the B.S. degree in data and computer science from Sun Yat-sen University, Guangzhou, China, in 2017. She is currently pursuing the M.S. degree in optics with Sichuan University. She is responsible for the gesture interaction part of the key research and development project in the lab. Her current research interests include the optical image encryption, the 3D integral imaging, the deep learning, and the convolutional neural networks.

XIAO-WEI LI received the M.S. and Ph.D. degrees in information and communications, engineering from Pukyong National University, South Korea, in 2011 and 2014, respectively. From 2014 to 2015, he was a Researcher with the Collage of Computer Engineering, Yonsei University, South Korea. He is currently an Associate Professor with the School of Electronics and Information Engineering, Sichuan University, China. His research interests include 3D integral imaging, holography, optical encryption, and image watermarking. He published approximately 50 articles cited by Science Citation Index. As the first author, he has published approximately 30 SCI-index articles, and the impact factor of half of articles are greater than three.

QIONG-HUA WANG received the M.S. and Ph.D. degrees from UESTC in 1995 and 2001, respectively. She was a Professor with the School of Electronics and Information Engineering, Sichuan University, from 2004 to 2018. She was a Postdoctoral Research Fellow with the School of Optics/CREOL, University of Central Florida, from 2001 to 2004. She worked at the University of Electronic Science and Technology of China (UESTC) from 1995 to 2001. She is currently a Professor of optics with the School of Instrumentation and Optoelectronic Engineering, Beihang University. She published approximately 200 articles cited by Science Citation Index and authored two books. She holds five U.S. patents and 90 Chinese patents. Her research interests include optics and optoelectronics, especially display technologies. She is a Fellow of the Society for Information Display. She is an Associate Editor of *Optics Express* and the *Journal of the Society for Information Display*.