Measure the Psychometric Functions of Deep Learning Models in Encrypted Image Recognition Tasks

Yirui Yao

Xi'an Jiaotong-Liverpool University, Yirui.Yao20@student.xjtlu.edu.cn

Pengjing Xu*

Xi'an Jiaotong-Liverpool University, Pengjing.Xu@xjtlu.edu.cn

The research aims at applying the convolutional neural network (CNN) including LeNet5, AlexNet, and Visual Geometry Group (VGG) with 16 weight layers to directly classify among 4 categories of fully encrypted images that were encrypted by various cryptographic algorithms involving Advanced Encryption Standard (AES), Blowfish, Data Encryption Standard (DES), and Triple DES (TDES) into without decryption to establish a secure image querying technique. The investigation was implemented with three concrete tasks. Firstly, used CNN models to recognize enciphered images with different encryption algorithms or cryptographic keys. Secondly, applied CNN models to classify enciphered images with simulated inference of Gaussian noise. Thirdly, employed CNN models to identify encrypted images with simulated inference of the reduction of contrast ratios.

The results achieved secure image recognition and proved the underlying capability of the CNN models to recognize encrypted images even with the interference of Gaussian noise and lower contrast ratio, which mostly are impossible for human beings with normal vision to distinguish. Furthermore, the study initially showed that the increased cryptographic strengths of encrypted images usually caused an implicit impact on the accuracy of the three CNN models. Conversely, the variations in the severity of Gaussian noise on enciphered images and the contrast ratio of encrypted images could have explicit impacts on the accuracy of the models. The results also reflected that LeNet-5 is the most suitable CNN model for recognizing enciphered images with different encryption strengths and recognizing enciphered images with Gaussian noise. Moreover, all three CNN models could be suitable when analyzing encrypted images with reduced contrast ratio according to the degree of reduction of the contrast ratio.

CCS CONCEPTS • Neural networks • Computer vision • Symmetric cryptography

Additional Keywords and Phrases: Convolutional Neural Network, Encrypted Images, Gaussian Noise, Contrast Ratio

1 INTRODUCTION

Wang *et al.* [1] introduced that an increasing number of private images are stored in the cloud as the development of cloud computing. Besides storing data, many different services like data analytics and image querying have been available to users, where a desired object is specified and then the relevant and similar images can be queried. Currently, encrypting private images like medical images in the cloud to protect sensitive information is necessary, but the decryption risks revealing the private information of encrypted images [1], and the occurrences of noise and the variations of contrast ratio may unexpectedly happen to images during storage and transmissions, the two interferences probably cause images to be comparatively indistinguishable after image decryption.

To address these challenges, the study aims to apply CNN models, one of the model types for deep learning to directly recognize completely encrypted images without decryption, for Lidkea *et al.* [2] insisted that CNN models can identify widely diverse objects into specific classes, ignoring the context of the image itself, and Lindsay [3] considered that the model architecture of the CNN models contributes them to be suitable candidates for simulating

* Corresponding author.

the visual nervous system. Moreover, the psychometric function was applied to evaluate model performance. Such functions originally correlate response levels of the subject to physical stimulus levels of the subject and provide the essential data for psychophysics, with the abscissa of the function being the stimulus intensity and the ordinate measuring the response level [4]. Evaluating the performance of each CNN model by utilizing psychometric functions assists in selecting the best-suited model for the three tasks, which contain classifying enciphered image datasets that are encrypted by various encryption methods (algorithms and encryption keys), identifying image datasets with identical encryption schemes but different severity of noise, and recognizing datasets with same cryptographic methods while diverse contrast ratios. Introducing the interference of noise and the variation of contrast ratios to the last two tasks was to simulate real-world environmental factors that potentially affect the accuracies of the image classification. The best CNN model among all models used in the tasks would maximally maintain its performance under the influences of cryptographic strength, Gaussian noise, and contrast ratio.

Securer image recognition is one of the main significances of the proposed work since recognizing enciphered images without decryption is normally more confidential than the traditional process that first decrypts encrypted images and then classifies them. Furthermore, the most suitable model for each of the tasks helps humans identify encrypted images that have been damaged to varying degrees due to their robustness under some interference.

2 LITERATURE REVIEW

The section will summarize related research on the encrypted image classification from a relatively earlier point in time to the present.

Previously, Wang *et al.* [1] employed an ML-ELM to classify enciphered images without decryption to efficiently and securely recognize if an enciphered image included the required object. Their proposed framework showed that photos were encrypted by AES or DES before they were stored in the cloud as their privacy should not be leaked to unauthorized entities and cloud service providers. Resembling an image retrieval system, the framework would categorize all encrypted images into two categories to determine without decryption whether the enciphered image contained a queried object if a photo query is specified. Thus, queried images can be efficiently queried from numerous encrypted images without the leakage of privacy information during retrieval in the cloud. The testing accuracy of ML-ELM under DES encryption was 90.44% and AES was 79.83%, which demonstrated the performance of the model can still be improved, so Wang *et al.* [1] suggested customizing an encryption method for ML-ELM.

In 2020, Lidkea *et al.* [2] proposed an image classification framework based on CNN for recognizing encrypted images of different types of vehicles from the intelligent transportation system (ITS) by merely partially decrypting them to ensure the confidentiality of sensitive information such as license and plate numbers. The classification accuracy was 86.8% leveraging less than 6,250 images with AES operated in the Output Feedback (OFB) mode. In addition, the experiment results indicated that compared with a completely decrypted system, the proposed partial decryption classification scheme emerged with up to 18% decrement in average computational complexity. The results also proved that the identification of road vehicles is feasible by utilizing the partially encrypted dataset [2].

Alzamily *et al.* [5] published that they classified encrypted images by a CNN model named ResNet50 without decoding in 2022. The dataset was Modified National Institute of Standards and Technology (MNIST), originally consisting of 70,000 images with 60,000 for training and 10,000 for testing. During the evaluation, ResNet50 accomplished 99.75% accuracy, 94.12% recall, 94.23% precision, and 94.70% F1-score on the testing set. The accuracy was higher than in previous studies. Therefore, the research evidenced that ResNet50 can directly identify encrypted images without decryption. Whereas this research only involved AES, to avoid any contingency, other

cryptographic algorithms such as DES and Blowfish can be employed to explore the performance of the ResNet50 model under their encryption [5].

3 METHODOLOGY

3.1 Psychometric Function

A psychometric function is a mathematical function mapping from the stimulus to the response level, which can typically be a statistical criterion like the percentage of correct samples in some projects. Usually, the psychometric function is presented in a sigmoid shape, for the variation of the response is not instantaneous as they are variable, so the same response for each stimulus is generally impossible [6], [7]. Specifically, the psychophysical threshold and the slope are two typical characteristics of the psychometric function. Firstly, the psychophysical threshold in a psychometric function can be defined by arbitrarily selecting a specific performance level and then referring to its corresponding stimulus level as the threshold. Such a threshold can be determined whether the responses of a psychometric function going from 0 to 100 percent which is shown in the left subplot of Figure 1 as determining if a center square has an intensity difference τ, compared to s_0, the background intensity of square, or the responses going from 50 to 100 percent that displayed in the right subplot of Figure 1 if an increment τ is detected as present. Secondly, the slope of the psychometric function is known as the rate of variability of perception, which reflects the change of the responses for a single stimulus level [7].

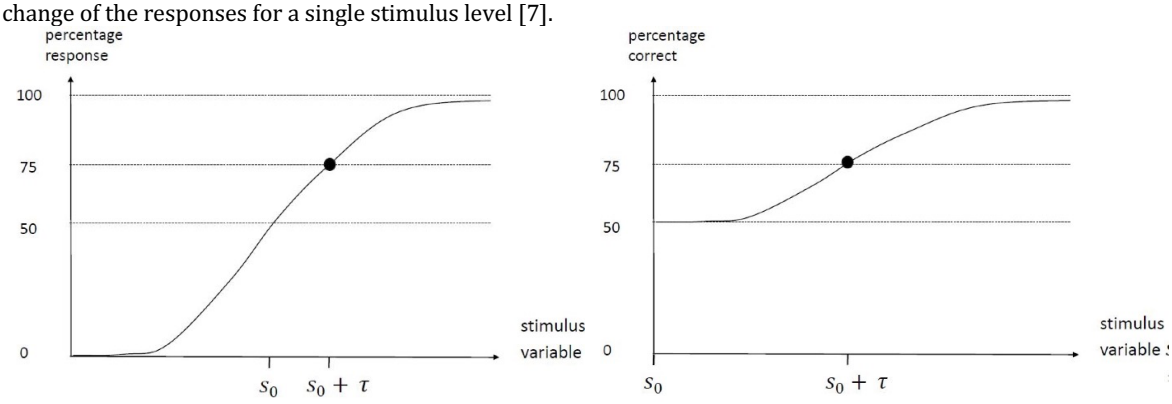

Figure 1: The Responses of a Psychometric Function [6].

Langer [6] reflected that in the experiments about psychophysics, the reasons why a gradual change in the response level were diverse sources of uncertainty the subjects confront during executions, including noise in the stimulus or display, limited resolution of the vision devices, and incorrect operations of the subjects may affect the response level. Likewise, in classifying encrypted image classification tasks, the subject can be various CNN models, the stimulus, which is the explanatory variable of the function, can be noise, contrast ratio, or different cryptographic algorithms and encryption keys. The recognition accuracy, the response variable of the function, can work as the response level.

3.2 CNN Models

Typical CNN models involving LeNet-5, AlexNet, and the VGG network with 16 weight layers (VGG16) were used in the encrypted image identification tasks. Commonly, the CNN model contains an input plane, convolutional layers, sampling layers, and fully connected layers. Firstly, the input plane receives ultimately size normalized and centered images. Subsequently, each unit in a current layer receives inputs from a set of units positioned in a relatively small neighborhood in the former layer. Moreover, units in a layer are organized in several planes within which all the units share an identical set of weights. The set of outputs of the units in such a plane is named a feature map. Secondly, an integrated convolutional layer consists of multiple feature maps with distinguished weight vectors, thereby multiple features can be extracted at each position. The kernel in the convolution layer is the set of connection weights applied by the units in the feature map. Thirdly, the subsampling layer, which executes a local averaging and a subsampling, so that reducing the resolution of the feature map and weakening the sensitivity of the output to distortions and shifts. [8]. Importantly, the input size of each CNN model was customized and consistent in enciphered image classification tasks. The elaborations of the three CNN models are listed as follows. The customized dimension of the input image is 90×90×3.

3.2.1 LeNet-5.

LeNet-5 contains 7 layers except for the input layer and the flattened layer, all of which comprise trainable weights. The first convolutional layer, with 6 feature maps. Every unit in each feature map has 25 inputs which are connected to a neighborhood in the input layer with both the width and height are 5 pixels, such unit is titled the receptive field of the unit. The width and height of the feature maps are both 86 pixels, which ensures every unit of the neighborhood to respectively correspond to a unit in the input layer. The first subsampling layer has 6 feature maps of size which is multiplied by a 43-pixel width and a 43-pixel height. Each unit in every feature map is connected to a neighborhood with a 2-pixel height and a 2-pixel width in the corresponding feature map of the first convolutional layer. The receptive field with the area 2×2 is non-overlapping, thus the number of columns and rows of feature maps output from the first subsampling layer is half of the number of columns and rows of feature maps in the first convolutional layer. Furthermore, the convolutional layer is equipped with a sigmoidal activation function. The second convolutional layer has 16 feature maps with a size of 39×39. Every unit in each feature map is connected to multiple 5×5 neighborhoods at the same locations in a subset of feature maps of the first subsampling layer. The second subsampling layer includes 16 feature maps of size which is multiplied by a 19-pixel width and a 19-pixel height. Each unit per feature map is connected to a 2×2 neighborhood in the corresponding feature map in the second convolutional layer. The first fully connected layer contains 120 feature maps. Each unit in each feature map is connected to a 19×19 neighborhood on all 16 feature maps of the second subsampling layer. The second fully connected layer comprises 84 units and is fully connected to the first fully connected layer. Eventually, the output layer is composed of 4 units since the dataset that will be specified in the section titled Dataset has 4 classes, so one unit for each category, with 84 inputs each. All the feature maps in convolutional layers and fully connected layers except the output layer are passed through hyperbolic tangent (tanh) activation functions, and the activation function of the output layer is a SoftMax [8].

3.2.2 AlexNet.

The CNN model comprises 8 layers with weights, the first 5 of them are convolutional layers and the remaining 3 layers are fully connected layers. The output of the last fully connected layer is fed to a 4-way SoftMax that yields a

distribution over the 4 category labels. The first convolutional layer filters the input image with 96 kernels of each kernel size $11 \times 11 \times 3$ with a stride of 4 pixels. The second convolutional layer filters the output of the first convolutional layer with 256 kernels of size that are multiplied by the 5-pixel height, 5-pixel width, and 48-pixel channels. The third convolutional layer has 384 kernels of size which are multiplied by the 3-pixel width, 3-pixel height, and 256-pixel channels connected to the outputs of the second pooling layer. The fourth convolutional layer has 384 kernels with the size 3×3×192, and the fifth convolutional layer has 256 kernels with the same size as the fourth convolutional layer. Each of the two fully connected layers has 4096 neurons. Additionally, except for the output layer, the activation functions of all convolutional and fully connected layers are Rectified Linear Units (ReLUs), which can be expressed as

$$f(x) = max(0, x)$$

where f indicates the output function of a neuron in AlexNet and x is the input of f [9]. Innovatively, a technique named dropout that aided reduce the severity of overfitting was applied to AlexNet. The main idea of dropout is randomly eliminating units for each training batch. The quantity of units to drop is controlled by a hyperparameter of dropout called the dropout rate, which was set as 0.5 in the experiment to demonstrate every hidden unit is randomly removed from the network with a probability of 50%. [10].

3.2.3 VGG16.

The VGG16 that was imported from Keras Application Programming Interface (API) [11] in the experiments has been pre-trained. The input image is passed through a stack of convolutional layers whose filters with receptive field whose width and height both are 3 pixels or 1 pixel. Max pooling layers are also performed with a 2×2 pooling size and stride 2 for each pooling layer [12].

However, in this investigation, the customized dimension of the input image was different from the required input size of 224×224×3. Consequently, to implement transfer learning of the pre-trained model to all the tasks of the investigation, the required input size of the model was initially defined to 90×90×3, and then a flattened layer and 3 fully connected layers were added based on the requirement of the API document. The first two fully connected layers have 4096 channels each, and the third includes 4 channels one channel per class. The activation functions of all hidden layers are ReLUs [12].

3.3 Encryption Algorithms

Images were encrypted through encryption algorithms comprising AES, DES, Blowfish, and TDES in the research. These algorithms were also ranked based on their encryption strength. A high avalanche effect indicates a high degree of diffusion, which is one of the metrics for evaluating the cryptographic strength of an enciphering algorithm. Entropy is also a metric that indirectly reflects the encryption strength of the encryption algorithm as such a metric demonstrates randomness in the information, and with high randomness, the relationship between key and ciphertext becomes sophisticated for an attacker to guess [13]. Consequently, Patil *et al.* [13] concluded that AES is the best-suited algorithm if encryption strength is of the highest priority in the application after comparing the cryptographic strengths of the remaining algorithms. Furthermore, Blowfish performs more confidential than TDES, and TDES is more secure than DES. Generally, the longer the secret key of a cryptographic algorithm contributes to the cryptographic strength. Take AES and DES as exemplifications to explain their work mechanisms.

3.3.1 AES.

This algorithm can use cryptographic keys of 16, 24, and 32 bytes in length to encipher and decipher data in blocks of 16 bytes. The length of the input block, the output block, and the State of the AES algorithm is 128 bits, which is equivalent to 16 bytes. The number of rounds to be executed of the AES algorithm depends on the key length. AES with the 16-byte encryption key corresponds to 10 rounds, AES with the 24-byte encryption key needs to perform 12 rounds, and 14 rounds to be performed during the execution of the AES with the 32-byte secret key [14].

3.3.2 DES.

The 64 bits of the input block of the flow of encryption by DES to be encrypted are first permuted, defined as the initial permutation (IP). After IP, permuted input consists of a 32-bit block L_0 and a 32-bit block R_0 is formed. Utilize the blocks to iterations, which can be expressed as

$$L_n = R_{n-1}$$
$$R_n = L_{n-1} \oplus f(R_{n-1}, K_n)$$

where n ranges from 1 to 16 and K_n is an output 48-bit block extracted from the 64-bit input block, which is

$$K_n = KS(n, Key)$$

where n also ranges between 1 and 16, Key is a 64-bit block, and KS is named the key schedule function that determines K_n within a certain iteration. Thus, the preoutput blocks are L_{16} and R_{16}. Finally, the blocks go through the inversed IP to output the ciphertext [15].

3.4 Noise and Contrast Ratio

The tasks introduced Gaussian noise to interfere with the recognition of enciphered images. Wang *et al.* [16] illustrated that Gaussian noise is defined by its Probability Density Function (PDF). The universal PDF of the univariate Gaussian noise is

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < +\infty$$

where μ represents the mean value and σ^2 denotes the variance, so σ stands for the standard deviation, which determines the error severity in terms of Cheng *et al.* [17]. Specifically, a larger standard deviation value contributes to more severe errors of Gaussian noise, and vice versa. In the tasks, μ was set as 0, and σ was set from 0 to 0.35. Particularly, to simulate real-world noise interference, which is commonly irregular and unpredictable, so random seed was not set in this study although the application of fixed random seed enables the experimental conditions to be more controllable, especially in the comparison of different interference to weaken the role of instability factors.

Contrast ratio, the ratio of luminance between the brightest and darkest instances of an image [18], is another interference to the task. In the tasks, the contrast ratio of enciphered images was decreased with varying degrees to explore the relationship between contrast ratio and accuracy of image identification. A function of Open Source Computer Vision Library (OpenCV) called convertScaleAbs achieved different levels of reduction of contrast ratio.

3.5 Experiment Setup

During the experiment, the employed hardware that supported the proceeding of the experiment was a machine with a processor titled Intel(R) Core(TM) i9-10920X CPU @ 3.50GHz 3.50 GHz, a graphics card called NVIDIA GeForce RTX 3090, and a RAM with executing memory is 64.0 GB (63.7 GB usable). The leveraged software contained a platform named Anaconda and PyCharm Community Edition worked as a web-based Integrated Development Environment (IDE) to implement and execute Python programs.

4 IMPLEMENTATION

The implementation contains the descriptions of the dataset in usage and the major steps of the tasks.

4.1 Dataset

The dataset used in this research was titled Brain Tumor Magnetic Resonance Imaging (MRI) dataset [19], which has four classes involving glioma, meningioma, pituitary, and no tumor. Figure 2 shows the sample image for each category. There are 5712 images in the training set and 1311 in the testing set. The image counts based on different class labels in the training and testing set are listed in Table 1, which also reflects that the quantity of training images in 4 various categories is comparatively balanced.

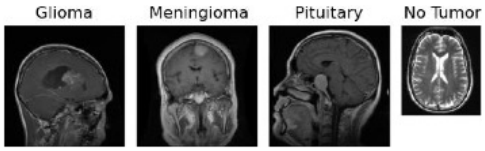

Figure 2: Mixing Data on the State Column-by-column [5].

Table 1: Number of Image in Each Category

Category	*Number of Images in Training Set*	*Number of Images in Testing Set*
Pituitary	1457	300
No Tumor	1595	405
Glioma	1321	300
Meningioma	1339	306

4.2 Data Preprocessing

Generally, the training and testing sets were respectively preprocessed during the data preprocessing process. Firstly, the preprocessing of the training set was composed by filtering images that were exceptionally loaded, eliminating blurred images, data augmentation by rotating every image at a 45-degree angle in the training dataset, removing similar images with setting the similarity of 0.999 as the threshold, reshaping rest of the images in the training set into the size of 128×128. Secondly, the preprocessing of the testing dataset consisted of deleting damaged images in the testing dataset, and reshaping the rest of the images into the size of 128×128. Subsequently, for images in the training dataset and testing set, respectively added Gaussian noise to each image, reduced the contrast ratios of every image, and encrypted each image with a certain cryptographic algorithm and a key. Crucially, all the keys generated during the experiments were only used for image encryption. Finally, labeled every image in the datasets according to the class of the image. New datasets would be generated during data preprocessing.

Innovatively, loaded the encrypted images that were in the Joint Photographic Experts Group (JPEG) form in the local by a built-in function of Python instead of applying two Python packages OpenCV and Python Image Library (PIL) as the packages raised errors when loading such images. The loaded image data was in the form of bytes. Then, convert the bytes into Numerical Python (NumPy) arrays for model training and testing. After converting the bytes to NumPy arrays through a function in a Python package NumPy titled frombuffer, multiple sub-arrays were nested in each array, thereby an extra procedure was to uniform the lengths of sub-arrays of every array.

After data preprocessing, 20 training sets and 167 testing sets would be generated. Compared to other training and testing sets that were filled with enciphered images, a training dataset and a testing dataset entirely contained unencrypted images for blank control to help evaluate the performances of CNN models in various conditions.

4.3 Model Training and Testing

Consistently, the loss function applied was Sparse Categorical Crossentropy, and the used optimizer was the Adaptive Moment Estimation (Adam) optimizer. All CNN models were trained employing a batch size of 96 images over 50 epochs. Applying K-fold Cross Validation with the value K was set to 5. During the cross-validation process, the training dataset would be divided into a smaller training set and a validation set with a constant ratio. Lidkea *et al.* [2] concluded that validation datasets are to validate whether the model performed well at recognizing new images that have been unseen by the trained model. Testing sets were the external data for the final evaluation of the model performance. The implementation of the three tasks in this project is elaborated as follows.

Initially, to explore the relationship between the cryptographic strengths and classified accuracies. The three CNN models were respectively trained by all the training sets to derive 60 different versions of trained CNN models, 20 versions for LeNet, 20 for AlexNet, and 20 for VGG16. Then the model was used to test corresponding testing datasets enciphered by the same algorithm and the same length of encryption key. For instance, if AlexNet is trained in a training set that is encrypted by AES with a secret key length of 16 bytes, the trained model should classify the testing dataset which was also encrypted by AES with the 16-byte key.

Furthermore, to measure the influences of different severity of noise on encrypted images, firstly the three CNN models recognized the encrypted images in the training set without adding noise. Afterward, the trained models identified encrypted images in the testing set that were enciphered by the same algorithm as the training encrypted images. For example, if VGG16 is trained on the training dataset which is enciphered by Blowfish with a cryptographic key length of 50 bytes. The trained VGG16 would then be employed to classify multiple testing datasets encrypted by Blowfish with a 50-byte key and different strengths of Gaussian noise.

Similarly, the relationship between the contrast ratios of the enciphered images and the recognized accuracies can be implemented by transferring the trained model of the training dataset without any contrast ratio reduction to a set of testing datasets with various contrast ratios. The training set and testing sets should share the same cryptographic algorithms and encryption keys.

4.4 Results

All three tasks proved the ability of CNN to classify encrypted images. The experiment results of the first tasks were visualized in Figure 3, Figure 4, and Figure 5, where the first tick label in their horizontal axis indicates datasets of unencrypted images. Each of the other labels consists of the abbreviation of the encryption algorithm and the length of the secret key with the unit of byte. For instance, the tick label des08 indicates the images were enciphered by DES and an 8-byte key. The three figures demonstrate that the classified accuracies of CNN models on enciphered images would not be distinctly affected by the increment of encryption strength. One of the possible reasons was that frombuffer helped interpret some patterns or features which is compatible with CNN to learn by converting encrypted image data in the form of bytes into NumPy array despite the interpretation might be so limited that could not be affected by the change of cryptographic strength. Nevertheless, all the testing accuracies of CNN models identified on unencrypted images were more than 90%, which performed much better than CNN models classified encrypted images, whose testing accuracies were around 40%, for the unenciphered images could be read by OpenCV while enciphered counterparts could not, and frombuffer might inevitably cause the loss of features or patterns of enciphered images. Particularly, the lower limit of testing accuracies from LeNet-5 was the highest among the other two CNN models, which was 43.4%, so LeNet-5 was probably more suitable for analyzing encrypted images with different encryption schemes.

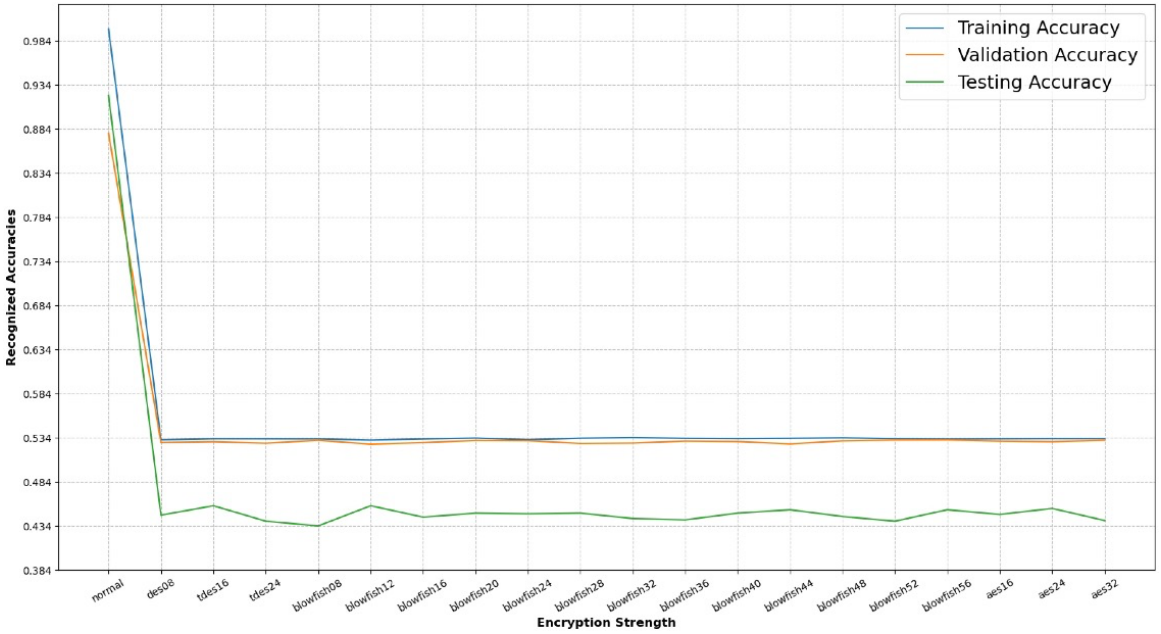

Figure 3: The Relationship between Encryption Strengths and Classified Accuracies (Trained and Tested by LeNet-5)

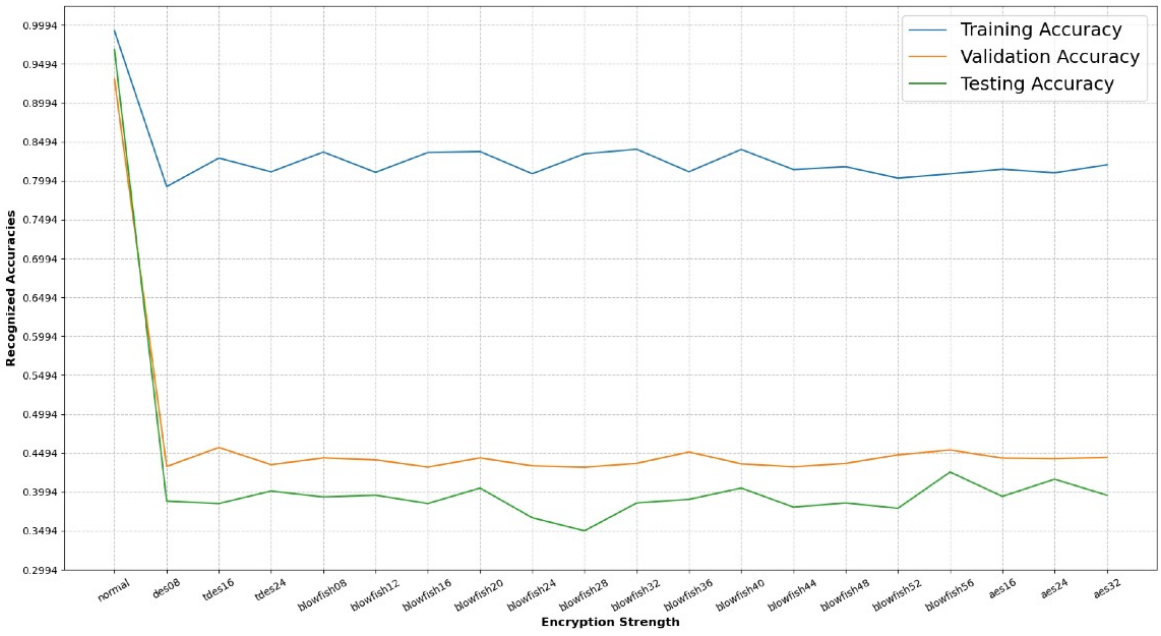

Figure 4: The Relationship between Cryptographic Strengths and Classified Accuracies (Trained and Tested by AlexNet)

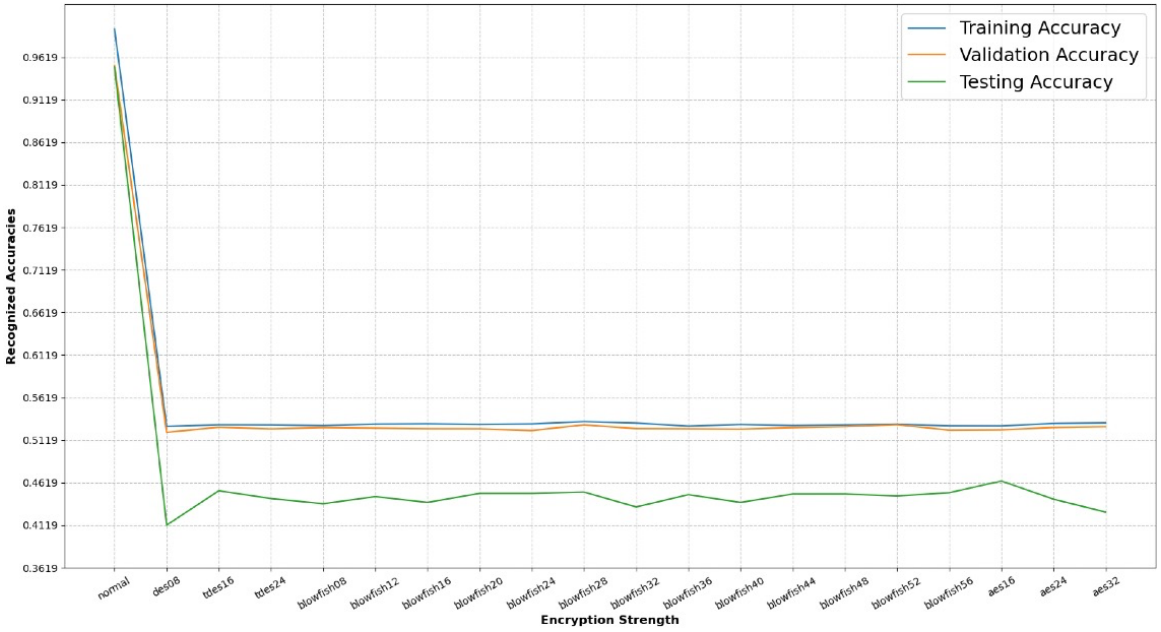

Figure 5: The Relationship between Cryptographic Strengths and Classified Accuracies (Trained and Tested by VGG16)

Moreover, the results of the second task are shown in Figure 6, Figure 7, and Figure 8. Figure 6, Figure 7, and Figure 8 imply that the rising strength of Gaussian noise generally causes obvious reductions in testing accuracies of the three CNN models despite several fluctuations in the accuracies. Mostly, LeNet-5 attained higher testing accuracies than VGG16 and AlexNet. Hence, LeNet-5 was the most suited model in classifying encrypted images through DES with an 8-byte cryptographic key, AES with a 32-byte key, and Blowfish with a 40-byte key given that Gaussian noise was added with the standard deviation ranged from 0.05 to 0.35.

Thirdly, Figure 9, Figure 10, and Figure 11 visualizes the results of the third task. Mostly, the recognized accuracy from LeNet-5 and VGG16 grew as the contrast ratio uplifted. While AlexNet first illustrated accuracy declinations when the alpha was from 0.05 to 0.20 in Figure 9, and 0.05 to 0.15 in Figure 10 and Figure 11. When the alpha exceeds these ranges, the accuracy of AlexNet uplifted except for several fluctuations. The shape of the psychometric functions of LeNet-5 and VGG16 are like sigmoid. Commonly, in Figure 9, given the encrypted images enciphered by AES with a 32-byte key, VGG16 performed better in such encrypted images when the alpha was lower than approximately 0.91 and higher than roughly 0.125. LeNet-5 accomplished the best testing accuracies when the alpha was more than 0.91. Otherwise, AlexNet was the most suited CNN model. Additionally, in Figure 10, when the enciphered images were encrypted by Blowfish with a 40-byte key, LeNet-5 reached testing accuracies higher than the other two models when the alpha was higher than roughly 0.8. Conversely, VGG16 attained accuracies virtually higher than the other models. Ultimately, in Figure 11, when the alpha was lower than around 0.71 and higher than about 0.09, VGG16 was the best-suited model recognizing the encrypted images enciphered by DES with an 8-byte key, while LeNet-5 was the most suitable if the alpha was larger than about 0.705. Otherwise, AlexNet reached the highest testing accuracy.

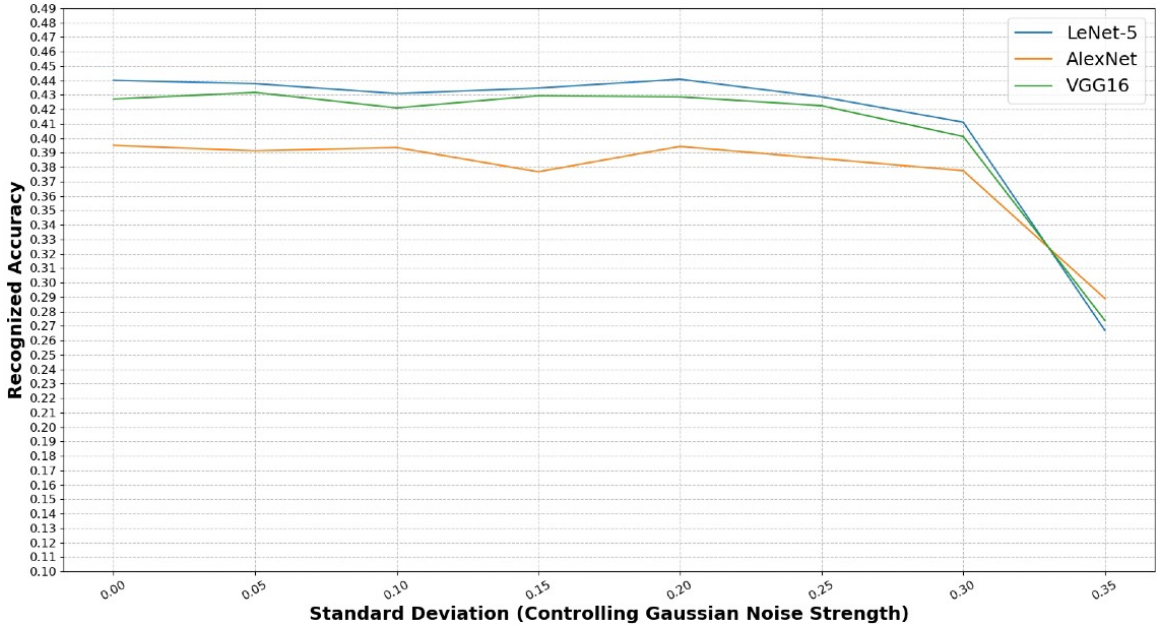

Figure 6: The Relationship between Various Strengths of Gaussian Noise and Classified Accuracies (Enciphered by AES32)

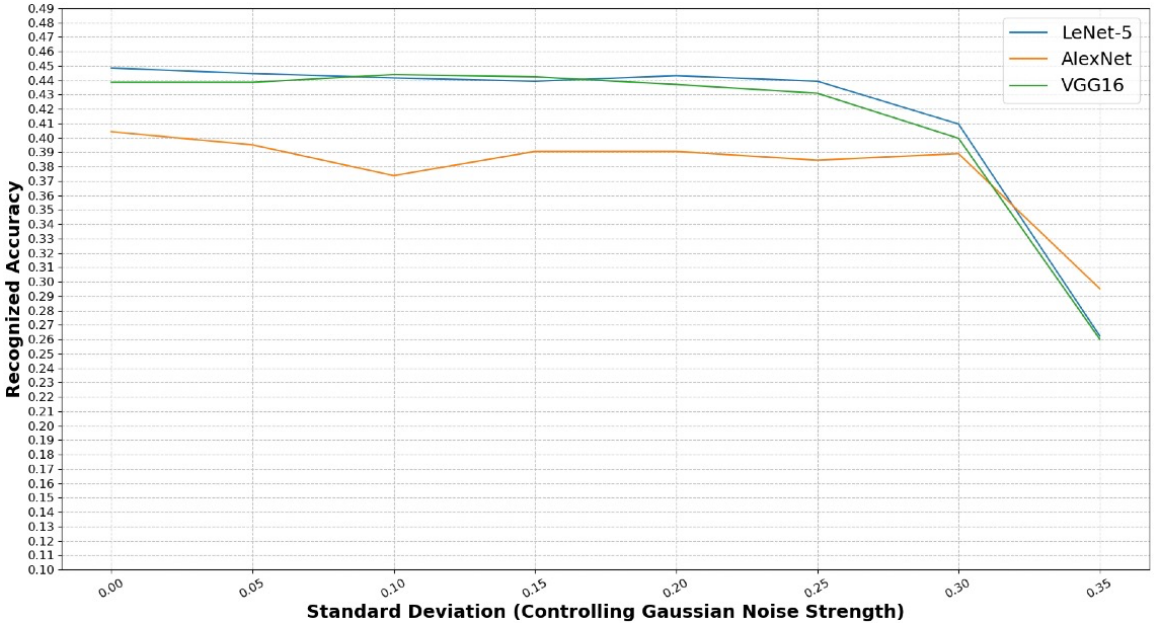

Figure 7: The Relationship between Various Strengths of Gaussian Noise and Classified Accuracies (Encrypted by Blowfish40)

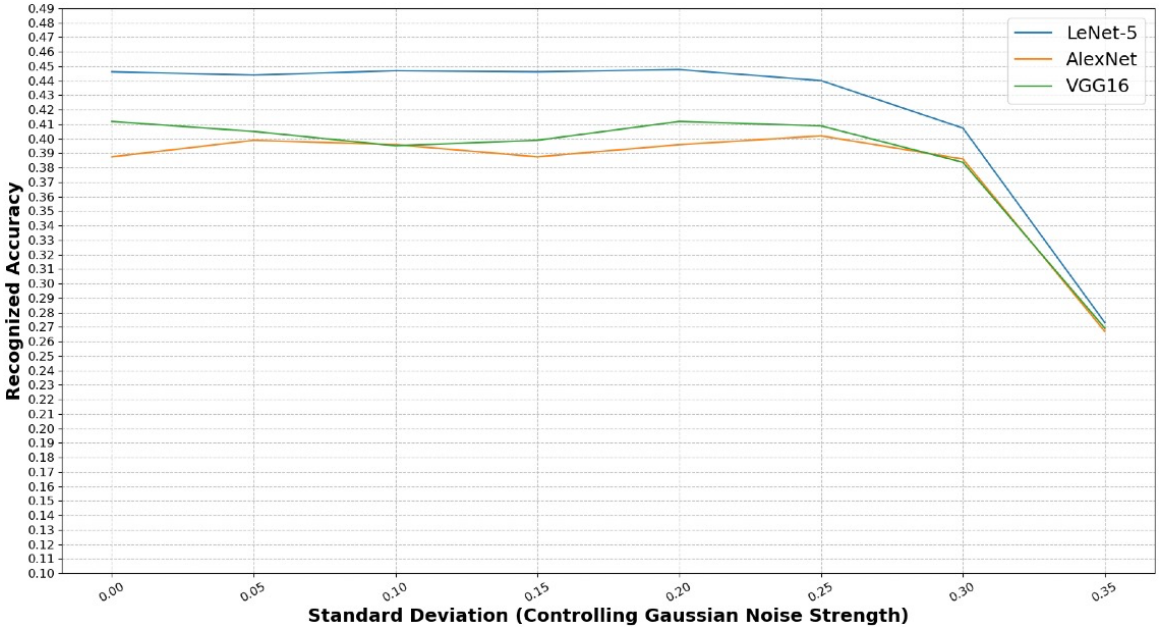

Figure 8: The Relationship between Various Strengths of Gaussian Noise and Classified Accuracies (Enciphered by DES08)

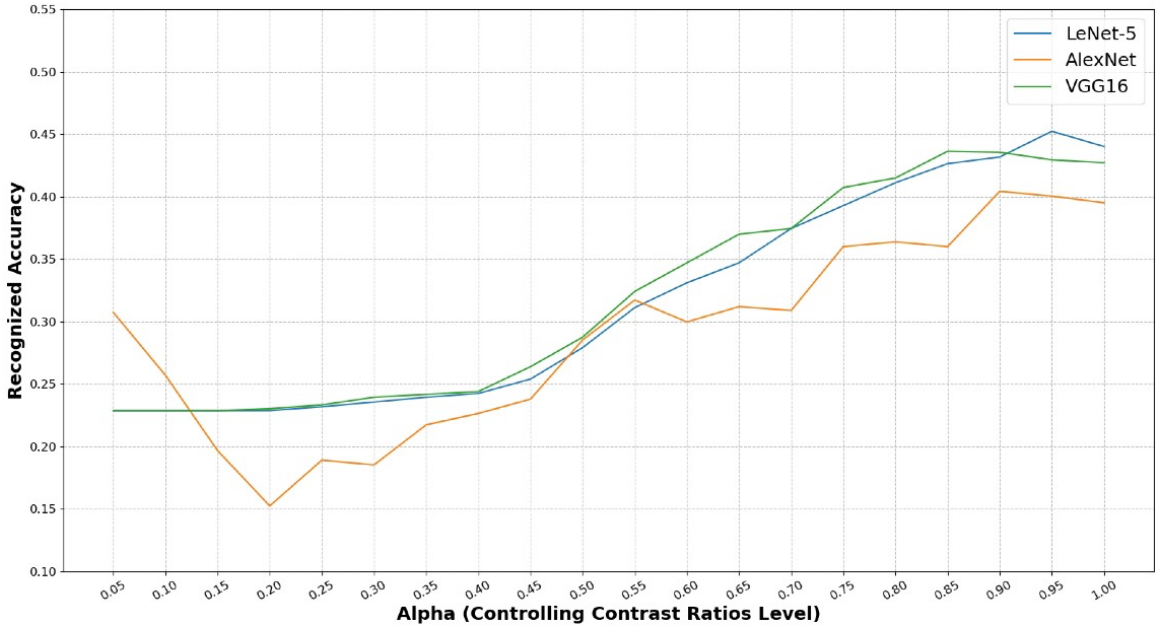

Figure 9: The Relationship between Different Contrast Ratios and Classified Accuracies (Encrypted by AES32)

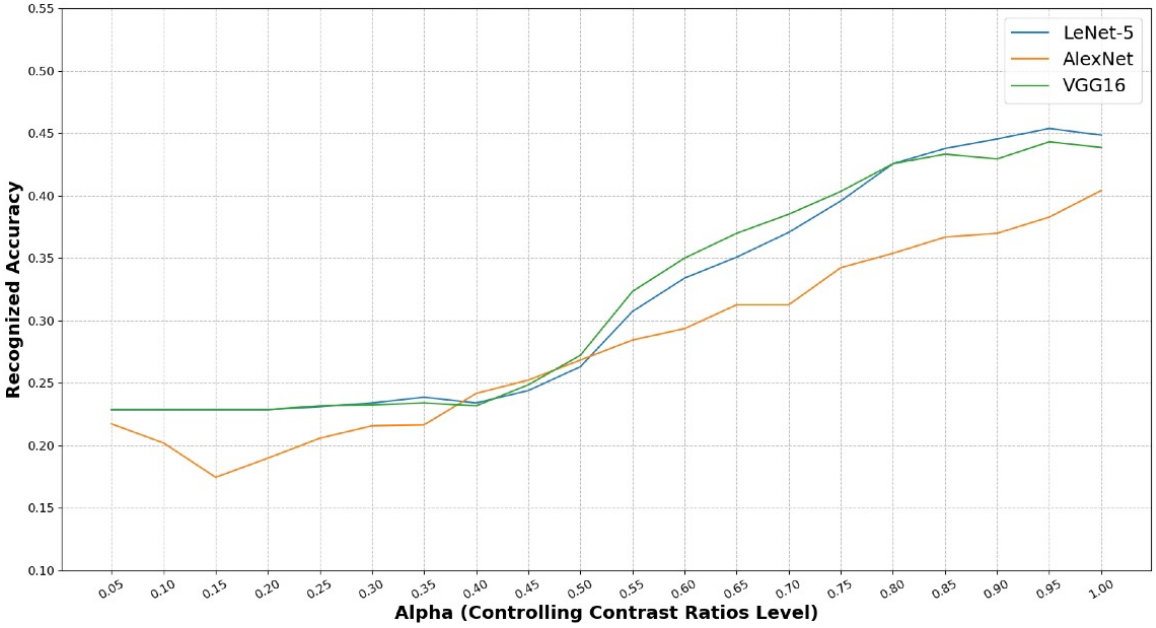

Figure 10: The Relationship between Different Contrast Ratios and Classified Accuracies (Enciphered by Blowfish40)

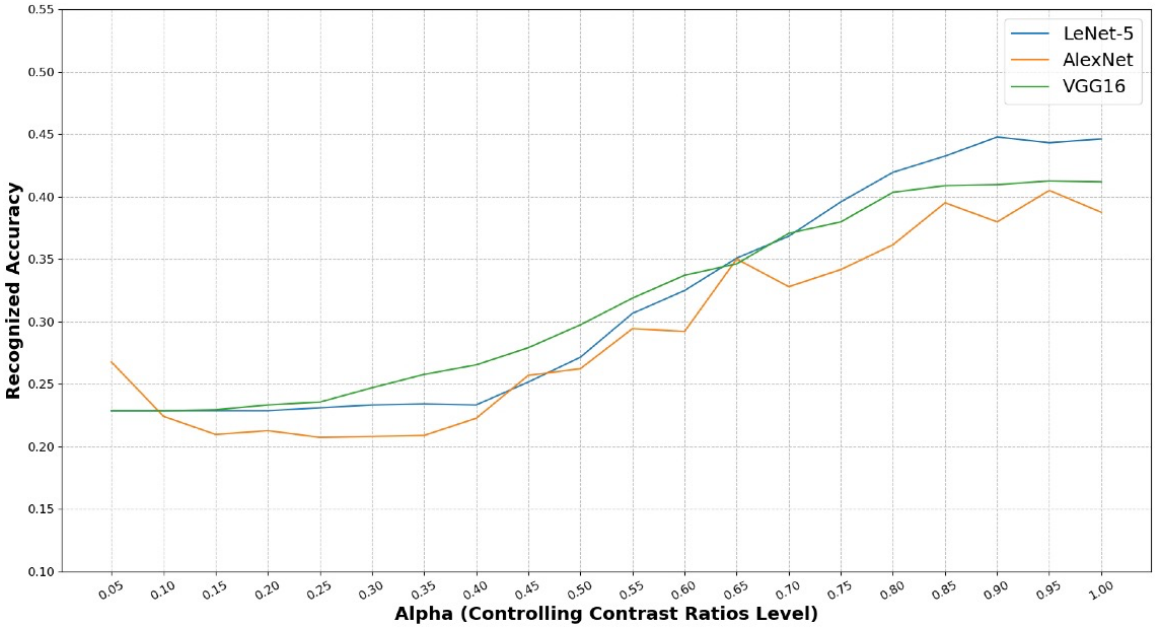

Figure 11: The Relationship between Different Contrast Ratios and Classified Accuracies (Encrypted by DES08)

5 DISCUSSION

One of the strengths of the research was when confronted with four-classification problems of encrypted images. CNN models performed better than human beings with normal vision because encrypted images are probably indistinguishable from humans, thereby they would be obliged to guess which class the image belongs to, and their average guess probability might approach 25%, while the lowest testing accuracy without adding noise and decreasing contrast ratio among the three CNN models was 34.94% in Figure 6. Therefore, even when adding Gaussian noise into enciphered images or reducing the contrast ratio of encrypted images, most of the time the CNN models were still recognized better than the naked eyes. For example, in Figure 11, LeNet-5 identified encrypted images by DES with an 8-byte encryption key with the testing accuracy higher than 25% if the alpha was higher than approximately 0.45. Secondly, the research ensures image recognition with privacy preservation as no decryption procedure is included in all the tasks. Thirdly, the lab experiments of the study helped solve constraints in diverse fields. Economically, environmentally, and sustainably, classifying encrypted images without decryption assisted in avoiding the usage of computational resources in the process of decrypting images, which reduces the cost and saves electricity to some degree. Ethically and socially, privacy preservation is a major importance of the study, privacy protection can mitigate the public panic about the leakage of private information. Lastly, the project allows medical image analysis that assists in medical diagnosis.

Nevertheless, the three CNN models failed to perform exceptional classification on encrypted images compared to unenciphered image identification, which probably resulted from the function frombuffer that converted bytes into NumPy arrays that image patterns and features might lose during the conversion. Unlike decryption can ultimately preserve the two of each image. For future development, a more suitable function for datatype conversion should be customized to substitute frombuffer to encourage more efficient image retrieval. In addition, more CNN networks such as ResNet50 and the Vision Transformer (ViT) model [20] can be tried in the future within reasonable computational resources and cost constraints.

6 CONCLUSION

The investigation applied LeNet-5, AlexNet, and VGG16 to recognize encrypted images by AES, Blowfish, TDES, and DES and different lengths of cryptographic keys of each encryption algorithm. There were 3 specific tasks, identifying enciphered image datasets that are encrypted by various cryptographic schemes, recognizing image datasets with the same encryption techniques but different severity of noise, and classifying datasets with the same encryption methods while diverse contrast ratios. Gaussian noise and contrast ratio in the tasks were to stimulate real-world interference from the environment. The psychometric function recorded the accuracies of the three CNN models under various scenarios such as diverse encryption methods, Gaussian noise, and contrast ratios.

In terms of the results illustrated by multiple psychometrical functions, the variation of cryptographic strengths usually does not measurably affect the recognized accuracy of the three CNN models, while the growth of Gaussian noise severity and the decrement of the contrast ratio influenced noticeably the classification of the models. Furthermore, the experiment results proved the suitability of each CNN model in certain scenarios based on concrete variables and their values. Specifically, LeNet-5 was the most suitable CNN model for recognizing enciphered images with various cryptographic strengths and enciphered images with Gaussian noise. Moreover, all three CNN models could be suited when identifying encrypted images with a decreased contrast ratio based on a certain value of the contrast ratio.

The study allowed a secure way of recognizing encrypted images and exploring the performances of the three CNN models given distinguished conditions. Mostly, the performances of CNN models were superior to the naked

eyes of human beings with normal vision. To develop the efficiency of such image querying, a more optimal function that can process data with the type of bytes and more models will be necessary.

REFERENCES

[1] Weiru Wang, Chi-Man Vong, Yilong Yang, and Pak-Kin Wong. 2017. Encrypted image classification based on multilayer extreme learning machine. In Multidim Syst Sign Process, Vol. 28. 851–865. DOI: https://doi.org/10.1007/s11045-016-0408-1

[2] Viktor M. Lidkea, Radu Muresan, and Arafat Al-dweik. 2020. Convolutional neural network framework for encrypted image classification in cloud-based ITS. In IEEE Open Journal of Intelligent Transportation Systems, Vol. 1. 35–50. DOI: https://doi.org/10.1109/OJITS.2020.2996063

[3] Grace W. Lindsay. 2021. Convolutional neural networks as a model of the visual system: past, present, and future. In J Cogn Neurosci, Vol. 33. 2017–2031. DOI: https://doi.org/10.1162/jocn_a_01544

[4] Stanley A. Klein. 2001. Measuring, estimating, and understanding the psychometric function: A commentary. In Perception & Psychophysics, Vol. 63. 1421–1455. DOI: https://doi.org/10.3758/BF03194552

[5] Jawad Yousef Ibrahim Alzamily, Syaiba Balqish Ariffin, and Samy S. Abu Naser. 2022. Classification of encrypted images using deep learning - ResNet50. In Journal of Theoretical and Applied Information Technology, Vol. 100. 6610–6620. Retrieved from http://www.jatit.org/volumes/Vol100No21/28Vol100No21.pdf

[6] Michael Langer. 2018. Lecture 13 – Psychophysics. Retrieved from https://www.cim.mcgill.ca/~langer/546/13-psychophysics-notes.pdf

[7] Cheryl Olman (Ed.). 2022. Introduction to sensation and perception. Psychometric functions. University of Minnesota Libraries Publishing, Minneapolis

[8] Yann Lecun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. Gradient-Based Learning Applied to Document Recognition. In Proceedings of the IEEE, Vol. 86. 2278–2324. DOI: https://doi.org/10.1109/5.726791

[9] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In Advances in Neural Information Processing Systems, Vol. 25. Retrieved from https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf

[10] Christian Garbin, Xingquan Zhu, and Oge Marques. 2020. Dropout vs. batch normalization: an empirical study of their impact to deep learning. In Multimedia Tools and Applications, Vol. 79. 12777–12815. DOI: https://doi.org/10.1007/s11042-019-08453-9

[11] Keras. 2020. Keras: Deep Learning for humans. Retrieved from https://keras.io.

[12] Karen Simonyan and Andrew Zisserman. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556v6. Retrieved from https://arxiv.org/abs/1409.1556

[13] Priyadarshini Patil, Prashant Narayankar, Narayan D G, and Meena S M. 2015. A Comprehensive Evaluation of Cryptographic Algorithms: DES, 3DES, AES, RSA and Blowfish. In International Conference on Information Security & Privacy (ICISP2015), Nagpur, India, 617–624. https://doi.org/10.1016/j.procs.2016.02.108

[14] Morris J. Dworkin, Elaine Barker, James R. Nechvatal, James Foti, Lawrence E. Bassham, E. Roback, James F. Dray Jr. 2001. Gaithersburg, MD DOI: https://doi.org/10.6028/NIST.FIPS.197

[15] National Institute of Standards and Technology. Data Encryption Standard. Retrieved from https://csrc.nist.gov/publications/detail/fips/46/1/archive/1988-01-22

[16] Shuihua Wang, M. Emre Celebi, Yu-Dong Zhang, Xiang Yu, Siyuan Lu, Xujing Yao, Qinghua Zhou, Martínez-García Miguel, Yingli Tian, Juan M Gorriz, Ivan Tyukin. 2021. In Information Fusion, Vol. 76. 376–421. DOI: https://doi.org/10.1016/j.inffus.2021.07.001

[17] Bowen Cheng, Ross Girshick, Piotr Dollár, Alexander C. Berg, and Alexander Kirillov. 2021. Boundary IoU: Improving Object-Centric Image Segmentation Evaluation. arXiv:2103.16562v1. Retrieved from https://arxiv.org/abs/2103.16562

[18] Frédéric Dufaux, Patrick Le Callet, Rafał K. Mantiuk, and Marta Mrak (Ed.). 2016. High Dynamic Range Video from Acquisition to Display and Applications. Academic Press, Chicago. DOI: https://doi.org/10.1016/B978-0-08-100412-8.00020-6

[19] Masoud Nickparvar. 2021. Brain Tumor MRI Dataset. Retrieved from https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset

[20] Khalid Salama. 2021. Image classification with Vision Transformer. (January 2021). Retrieved August 14, 2024 from https://keras.io/examples/vision/image_classification_with_vision_transformer/