DOI: https://dx.doi.org/10.21123/bsj.2022.6550

Retrieving Encrypted Images Using Convolution Neural Network and Fully Homomorphic Encryption

Emad M. Alsaedi *   *Alaa kadhim Farhan*

Computer Sciences Department, University of Technology, Baghdad, Iraq.
*Corresponding author: cs.19.71@grad.uotechnology.edu.iq.
E-mail addresses: Alaa.K.Farhan@uotechnology.edu.iq.

Received 14/9/2021, Revised 30/1/2022, Accepted 1/2/2022, Published Online First 20/7/2022, Published 1/2/2023

Abstract:

A content-based image retrieval (CBIR) is a technique used to retrieve images from an image database. However, the CBIR process suffers from less accuracy to retrieve images from an extensive image database and ensure the privacy of images. This paper aims to address the issues of accuracy utilizing deep learning techniques as the CNN method. Also, it provides the necessary privacy for images using fully homomorphic encryption methods by Cheon, Kim, Kim, and Song (CKKS). To achieve these aims, a system has been proposed, namely RCNN_CKKS, that includes two parts. The first part (offline processing) extracts automated high-level features based on a flatting layer in a convolutional neural network (CNN) and then stores these features in a new dataset. In the second part (online processing), the client sends the encrypted image to the server, which depends on the CNN model trained to extract features of the sent image. Next, the extracted features are compared with the stored features using a Hamming distance method to retrieve all similar images. Finally, the server encrypts all retrieved images and sends them to the client. Deep-learning results on plain images were 97.94% for classification and 98.94% for retriever images. At the same time, the NIST test was used to check the security of CKKS when applied to Canadian Institute for Advanced Research (CIFAR-10) dataset. Through these results, researchers conclude that deep learning is an effective method for image retrieval and that a CKKS method is appropriate for image privacy protection.

Keywords: CKKS, CNN, Content based image retrieval, Homomorphic encryption, Random forest.

Introduction:

CBIR is the most practical approach for meaningful image searching since the number of digital images on digital resources has expanded exponentially. Most conventional image searches are done using metadata of the image. To retrieve this metadata, text query techniques are used. However, all images must have correct metadata to retrieve better-related images. They are choosing the proper metadata for an image as one of the issues impacting search accuracy. As a result, query by image (QBI) is a better option. The CBIR system is discovered to be employed for image processing inquiries from numerous research investigations. The CBIR system takes a sample image as input and searches a vast library of photos for similar images using low-level features[1]. In CBIR, image retrieval is classified into high-level and low-level features. Initially, single color, shape, and texture elements were used, resulting in good retrieval results due to the availability of diverse visual qualities. Additionally, machine learning techniques are used, which have a high degree of efficiency in automatically extracting low-level information from images[2].Fully homomorphic encryption (FHE) allows the model to compute an unbounded computation over ciphertext and decrypt it over plain text. Equation (1) illustrates a function *f*() that performs arithmetic operations such as (addition or/and multiplication) on the plaintext and is equivalent to the ciphertext[3].

$$f\big(E(m)\big) = E\big(\dot{f}(m)\big) \ldots\ldots\ldots \quad 1$$

Cryptography is a branch of mathematics that uses complex algorithms to ensure the confidentiality of information while it is transmitted and stored. Since the fundamental Diffie Hellman paper was introduced in 1976, several new public-key cryptosystems have been created. Most of them are based on two hard mathematical problems: the factorization and the discrete logarithm problems (e.g., RSA, ElGamal cryptosystem, ECC, and many others).Even though these cryptosystems are very safe and contain considerable keyspace, they have been regarded as expensive and slow[4,5]. Armknecht and Sadeghi developed an algebraically homomorphic approach to cryptography in 2008 [6], while Gentry extended foundational work on fully homomorphic schemes in 2009 [7]. The same year Gentry modified fully homomorphic encryption utilizing ideal lattices[8]. Van Dijk et al. developed completely homomorphic encryption over integers in 2010[9]. Gentry et al. (2013) also developed homomorphic encryption based on error learning[10]. Fig.1 illustrates the approach of completely homomorphic cryptosystems.

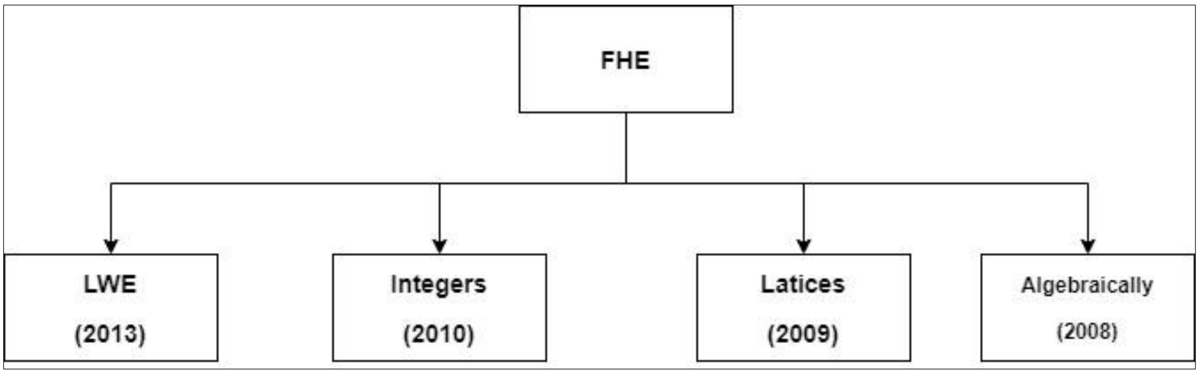


Figure 1. Approach of the FHE

The current state-of-the-art in approximation homomorphic computations for real and complex numbers is the Cheon-Kim-Kim-Song (CKKS) HE scheme. The CKKS system can already be applied to practical use, e.g., in machine learning[11].

Traditional learning algorithms are dependent on hand-designed features. Deep learning, one of the most powerful techniques of machine learning, may transcend this dependence. There are two steps to deep learning: training to improve the model accuracy and inference to utilize the model for analysis such as classification or prediction (see the next section for more details). In recent years, deep learning has been applied in many fields, including big data analytics and applications such as pattern identification, speech recognition, and computer vision. Deep learning poses privacy problems, especially when using a sophisticated cloud infrastructure and a collaborative approach. Concerns about privacy arise from sensitive input data used in training or inference and from sharing the learned model. The training algorithm efficiency may be improved by using a strong remote server or cloud [12]. However, both the users and the server may have privacy issues when using these environments. An attacker with complete knowledge of the training process may access model parameters, posing privacy concerns and shifting the problem from data privacy to model privacy[13]. The leak of sensitive data among participants should also be considered in the case of collaborative learning. Sensitive data leakage between users and external infrastructure, especially via the Internet, should also be considered [12].

Through related works [14–20]The researchers concluded that , there are two significant issues relating to CBIR. Firstly, ensure users' privacy and increase data confidentiality while trade-off exits between security and efficiency. Approaches based on lightweight encryption, such as permutation or substitution, are efficient but insecure; methods based on a heavyweight algorithm are secure but impractical computing costs. Second, a trade-off exists between retrieval accuracy and retrieval efficiency. While many approaches for image retrieval include low-level features like color, texture, and form, the retrieval accuracy usually serves the practical applications due to the "semantic gap" between visual features and the scope of human semantics. A fully homomorphic encryption algorithm was used in this paper because it has been demonstrated through experiments that it is a highly efficient algorithm in preserving data privacy. For cloud computing applications, homomorphic encryption became a popular and powerful cryptographic. Also, it presents security analysis against all the known attacks with respect to the message expansion and homomorphic operations[21]. Still, it has a high cost in the encryption and decryption processes, so the researcher divided the algorithmic load between the client and the server to solve this problem. Another

issue addressed in this research is the balance between accuracy and efficiency, achieved by combining deep learning techniques with Random Forest to reduce and choose the right features. This combination resulted in reducing the processing time and increasing accuracy. The motivation of the proposed work consists of building a secure AI model to balance the accuracy of image retrieval, reduce the processing time of retrieval operations, and preserve the privacy of data transmitted through an insecure medium.
The following are the major contributions in this paper:

1. Trade-off between security and accuracy of the CBIR.
2. Proposing an effective method for image retrieval based on deep learning techniques.
3. Encrypting the retrieved and sent images using FHE-CKKS algorithms to ensure their security.
4. Improving the classification CNN's accuracy by training it on augmentation images.
5. Using the Random Forest method to select the best features that increase accuracy and decrease retrieval time.

The rest of this paper is organized as follows. In Section 2, the researchers give related work in homomorphic encryption, image retrieval, and deep learning. In Section 3, the researchers explain in brief materials and methods used in this paper. The researchers present them propose in Section 4. Experimental results and performance analysis are given in Section 5. Conclusions and future work are given in Section 6.

Related Works:

A large amount of research has been published in the fields of the CBIR and homomorphic encryption.

In Zhihao Cao et al[16]., the dimension reduction and image retrieval by convolutional neural networks were used to extract high-level image features. In addition, multiline core component analysis was used to reduce feature dimensions that were too large and strongly correlated. For efficiency, features are binary encoded after feature reduction.

Manisha and et., al[17]. suggested a strategy employing LNDP for local features. This method converts every pixel in an image into a binary pattern depending on its nearby pixels. Thus, both LBP and LNDP extract information using local pixel intensity.

Selvam and et., al[18]. suggested a method for combining the genetic algorithm (GA) with the HARP aggregation algorithm to increase system retrieval accuracy while using less processing time also to recover the relevant image and possible resolution utilizing CBIR.

Kuo et al. [22] introduced deep convolutional neural networks in their paper for image retrieval. It uses DL to train the weights of a NN, resulting in high-level image feature extraction.

Hsin et al. [23] proposed CNN as an aggregate of ensemble models for image retrieval. This image classifier combines AlexNet and Network in Network (NIN), which are both particularly effectively deep learning networks, to achieve image feature extraction. It computes weighted average feature vectors for image retrieval.

Umer et al. [24] developed an efficient content-based image retrieval CBIR system capable of retrieving correct images semantically. They proposed a hybrid features descriptor consisting of color and texture features for this purpose.

According to Xia et al[15], block and pixel permutation are used to offer a privacy-protected LBP extraction approach in the ciphertext domain in the suggested method. The security of these methods is compromised, but they are efficient.

Fathala et al.'s,[25] paper is essentially dependent on two procedures to retrieval techniques. First, extracting an image features with the histogram, then using statistical features (mean, standard deviation). In this instance, the T-test is used to examine the relationship between a lot of different images.

Challa et al. [26] suggested a modified Reed-Muller Code-based symmetric key fully homomorphic encryption that enables both (MOD 2) additive and multiplication operations limitlessly. The proof of security provides a mathematical analysis and difficulty level. It also analyzes the security of message expansion and homomorphic operations against all known attacks and vulnerabilities.

Syed et al. [27] suggested the use of homomorphic encryption, which allows for the training of deep learning and classical machine learning models while maintaining data privacy and security. The proposed methodology is being evaluated in smart grid applications such as fault diagnosis and localization and load forecasting. The results for fault localization reveal that the proposed privacy-preserving deep learning model's classification accuracy, while utilizing homomorphic encryption, is comparable to the model's classification accuracy on plain data.

Lou et al. [28] described a deep neural network that can process encrypted data using a Shift-accumulation-based LHE-enabled network. They develop ReLU activations and max poolings using the binary operation-friendly Leveled Fast Homomorphic Encryption over Torus (LTFHE) encryption method. Instead of expensive LTFHE multiplications, they use cheaper LTFHE shifts to accelerate inferences.

Obla et al. [29] introduced a methodical approach to generating higher education-friendly activation employment for CNNs. They began by evaluating commonly utilized functions such as linear correcting units (ReLU) and Sigmoid to find the qualities in a good activation function that contributes to performance. Then, they compare the polynomial approximation methods and determine the best range of approximation for polynomial activation. They also proposed a novel weighted polynomial approximation method for distributing the batch adjustment layer's output. Finally, they demonstrated the efficacy of their strategy employing a variety of datasets, including MNIST, FMNIST, and CIFAR-10.

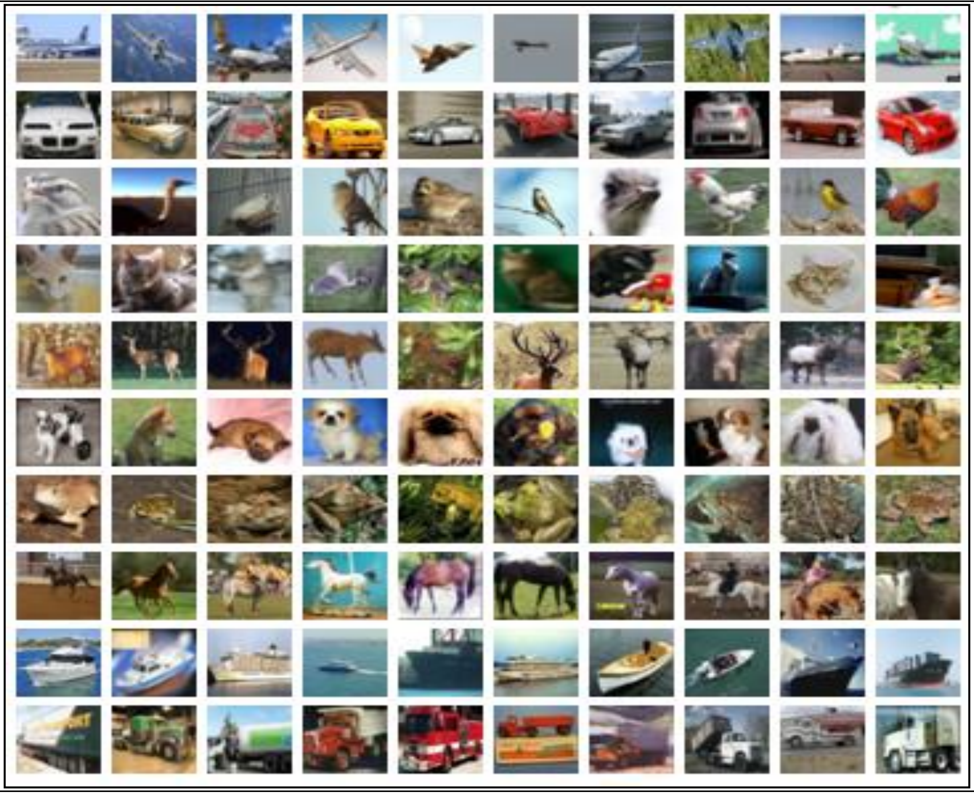
Owusu et al. [30] developed a new framework, MSCryptoNet, that enables MSCryptoNet models to be executed, converted, and scaled in a privacy-preserving manner. Sigmoid and rectified linear units activation functions are approximated using low degree polynomials in homomorphic encryption systems.

Clet et al. [31] comprehensively covered the three most common homomorphic cryptosystems, BFV, CKKS, and TFHE, concerning the training phase of feed-forward neural networks that have been successfully completed on the MNIST dataset.

Materials and Methods:

The CIFAR-10 dataset contains various images used to train machine learning algorithms and computer vision. Therefore, this dataset is increasingly utilized for machine learning research. The CIFAR-10 dataset has 60,000 images, all of which are 32 × 32 in size and are in the PNG format, and they belong to 10 different classes as illustrated in Table 1. For example, an airplane, a vehicle, a bird, a cat, a deer, a dog, a frog, a horse, a ship, and a truck are classes. A total of 50000 training pictures and 10000 test images were taken from the CIFAR-10 dataset, with the first set being used for training and the second set being used for testing and evaluating the proposed model. There are precisely 1000 photos from each class in the test batch that are randomly picked. Randomly arranged 5000 images from each class are included in each training batch.

Table 1. Samples of CIFAR-10 Dataset for 10 Classes

Class	Images
Airplane	
Automobile	
Bird	
Cat	
Deer	
Dog	
Frog	
Horse	
Ship	
Truck	

Deep Learning

Deep learning attempts to extract complicated features from high-dimensional data and use them to construct a model that connects inputs to outputs (such as classes). Deep learning architectures are typically built as multi-layered networks, with higher-level characteristics computed as nonlinear functions of lower-level features. The most prevalent type of deep learning architecture is a layer neural network [32]. The structure of deep learning is represented by many levels. This section will present these layers and how they affect homomorphic encryption.

Convolutional Neural Network (CNN):

CNN is commonly used for image classification and is defined by a convolution layer whose function is to learn the features derived from the dataset. The convolutional layer is N x N in size and will perform dot product multiplication between neighborhood values. As a result, the convolutional layer only contains addition and multiplication functions. This layer does not need to be changed as it can be used for homomorphic encryption data[33].

Activation Layer:

The activation layer is a non-linear feature that applies a mathematical procedure to the output of the convolution layer. Because these tasks are not linear, the difficulty increases significantly when utilized to assess Homomorphic Encrypted (HE) data. As a result, designers must develop a substitute element that only requires multiplication and addition[29].

Pooling Layer:

This sample layer's purpose is to minimize the data size. Pooling can be classified into several types, such as maximum and average pooling, mean pooling, and so on. One would not use the max-pooling option in HE, but average pooling is a solution that is used in HE, because average pooling determines the number of values using two operations that are allowed in HE[28].

Fully Connected Layer:

It is described as a "Fully Connected Layer" since each neuron is connected to the neuron in the previous layer. There is only a dot product operation in this layer, which consists of multiplication and addition functions. As a result, it can be employed over encrypted data[34].

Dropout Layer:

This was done to avoid overfitting. Researchers often get excellent classification results for using machine learning model when training, suggesting bias in the training set[35].

Homomorphic Encryption (HE)

Various tools are employed to protect privacy, such as differential privacy techniques and homomorphic encryption. HE is a type of encryption that allows various kinds of calculations to be performed on ciphertexts to produce an encrypted output. HE is divided into three categories [36].

Partially Homomorphic Encryption (PHE): This provides only one encrypted data process, which is either addition or multiplication.

Somewhat Homomorphic Encryption (SWHE): This provides more than one process, such as multiplication and addition, but the number of operations is limited.

Fully Homomorphic Encryption (FHE): This provides multiple multiplication and addition processes without restriction on the number of functions.

HE schemes include four stages. [37]***:***

The Key Generation (KeyGen): In this stage, security parameters are generated. In an asymmetric type, a single key is generated, while in an asymmetric type, a pair of secret and public keys are generated.

The Encryption Algorithm (Enc): This stage encrypts the plaintext inputs message, m ϵ M, with the encryption key. The ciphertext is generated by c = Enc(m), where c ∊ C, C is the ciphertext space.

The Decryption Algorithm (Dec): In this stage, the original message is recovered by decrypting ciphertext c using the decryption key ((c) = m).

The Evaluation Algorithm (Eval): This stage performs the evaluations of the ciphertexts ($c1$, $c2$),

($c1$, $c2$) = $Eval$ {($m1$, $m2$)}, without revealing the messages ($m1$, $m2$).

CKKS Homomorphic Encryption Scheme

The Cheon-Kim-Kim-Song (CKKS) scheme is a leveled homomorphic encryption method that depends on the RLWE problem's difficulty for security. Unlike other HE systems, CKKS allows precise approximation arithmetic on real and complex numbers. The CKKS method interprets decryption noise as a mistake in the calculation of real values. It excels for applications like machine learning where most computations are approximated. With the use of bootstrapping technique as mentioned [38], the CKKS scheme becomes a FHE (fully homomorphic encryption) scheme.

Proposed system

As demonstrated in Fig. 2, this protocol has two major phases: the offline phase is implemented on the server side, and the online phase is implemented on both the server and the client. The offline phase is referred to as generation. The CNN model phase includes training stages, which consist of three steps performed on plaintext training data to produce a classifier, which is then used with plaintext testing data to produce a trained model. The online phase consists of eight steps on the server side and four steps on the client side, as shown in Fig. 2. The steps on the server side include first, creating the keys, second, sending them to the client's side, third, receiving the encrypted image, fourth, decrypting the encrypted image, fifth, inference of the decrypted image based on the trainer's model, and sixth, sending the decrypted image.

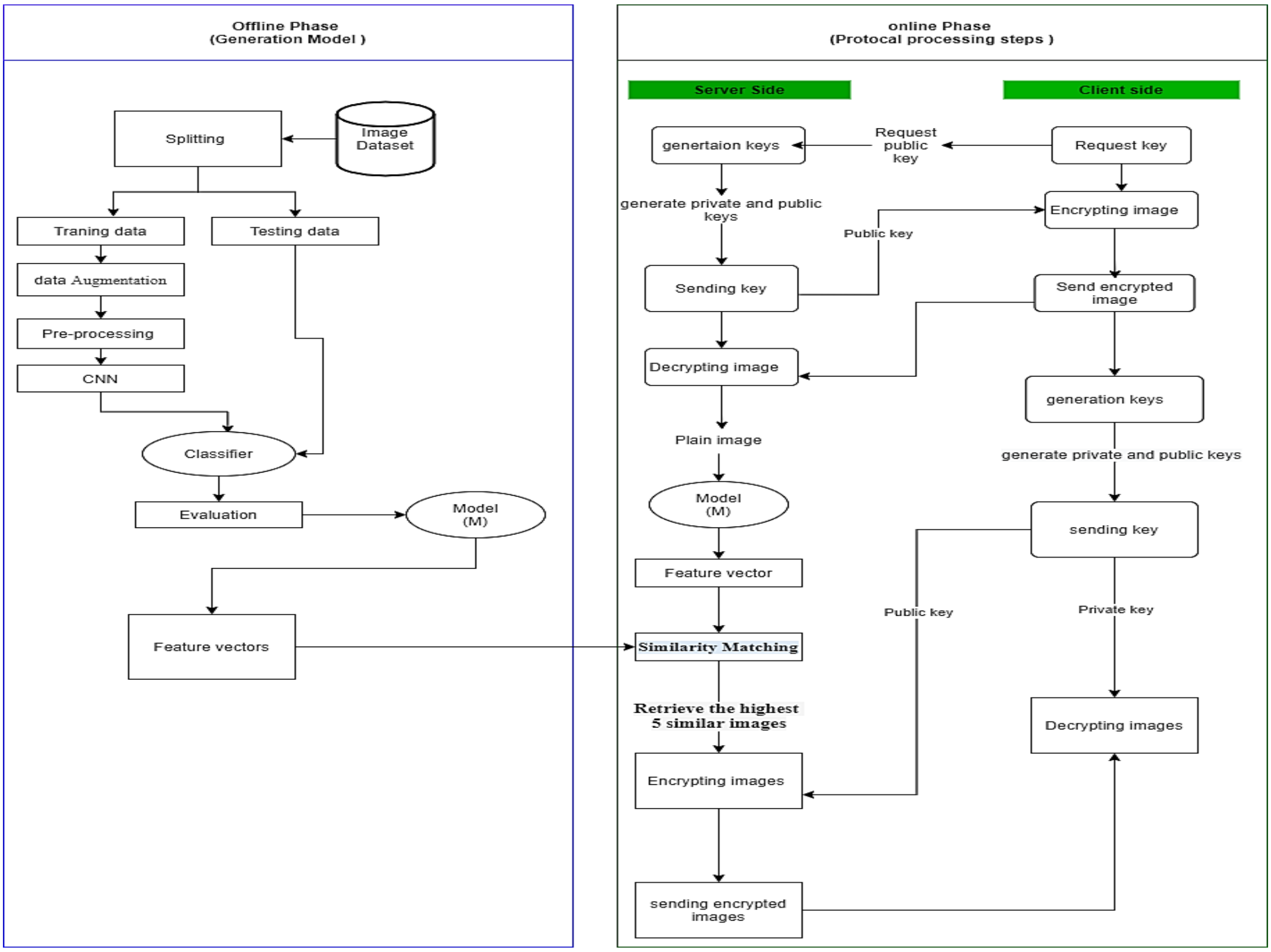


Figure 2. RCNN_CKKS protocol which consists of offline and online phases

Offline phase

This part has two major stages: the first contains a trained model (M), and the second uses the trained model to extract the features for each image in the training dataset.

Generation CNN model phase

This phase consists of several procedures that are necessary to create the classification model. The steps used in this section are explained in the following section. The major procedures for obtaining the trainer model appear in Fig. 2. The image datasets used in this article include CIFAR-10 (Canadian Institute for Advanced Research), a collection of images used to train machine learning and computer vision algorithms. CIFAR-10 is a popular dataset for machine learning research. It contains 60,000 images in the PNG format with a size of 32x32 colors in ten different classes: airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck. This dataset is divided into two parts, the training dataset and the test dataset. The training dataset consists of 50,000 images for training the network, which accounts for 80% of the total dataset. The test data set, on the other hand, comprises 10,000 images for testing the network and accounts for 20% of the entire dataset. Ten useful augmentations are employed in the proposal system: rotating, horizontal shifting, vertical shifting, and flipping. The rotation process is applied to the original training images at an angle of 15 degrees, which generates a new 50,000 images using the Bilinear interpolation method. An image can be shifted horizontally and retain the same dimensions without distorting it by using horizontal shift augmentation. The process of shifting all the pixels vertically while maintaining the image dimensions is referred to as vertical shift augmentation. The neighborhood of each pixel is thresholded, and the result is a binary number with LBP. Other augmentation methods, which use

Gaussian blurring to generate new images, are currently under development. Furthermore, two noise-generating techniques are employed. These are Gaussian and Salt-and-pepper noise. A total of 450,000 training images are produced as a result. The structure of the proposed CNN contains ten layers as shown in Fig. 3.

Convolutional layer1: In this layer, a 3×3 filter slides over the 32 × 32 image to create a feature map that summarizes the presence of detected features in the original input image.

Convolutional layer2: This layer repeats the steps of the convolutional layer1 on the 32×32 image with filter 2 and depth 32.

Max Pooling layer1: After applying two convolutional and ReLU layers the max pooling layer is applied to extract maximum features from the image using a filter with size 2×2, and then uses dropout with 20% to prevent the network from overfitting.

Convolutional layer3: This layer repeats the steps of the convolutional layer2 on the 16×16 image with filter 3 and depth 64, to obtain robust features where minimum features cancel in max pooling layer3.

Convolutional layer4: This layer repeats the same steps of the convolutional layer3 on the 16×16 image with filter 4 and depth 64.

Max Pooling layer2: The max pooling layer is applied after convolutional layer4 and extracts the maximum features from the image using a filter with size 2×2 and then uses dropout with 30% to prevent the network from overfitting.

Convolutional layer5: This layer repeats the steps of the convolutional layer4 on the 8×8 image with filter 5 and depth 128.

Convolutional layer6: This layer repeats the steps of the convolutional layer5 on the 8×8 image with filter 6 and depth 128.

Max Pooling layer3: The max pooling layer is applied after convolutional layer8 and extracts the maximum features from the image using a filter with size 2×2. After max pooling layer3, the image becomes 4×4. It uses dropout with 40% to prevent the network from overfitting.

Flatten layer: The results from the convolutional and max pooling layers produce higher-level features of the input image.

Fully connected layer1: Single feature vectors from the flatten layer become input nodes for the full connection layer. The classifier's purpose is to assign class labels to images based on the training dataset.

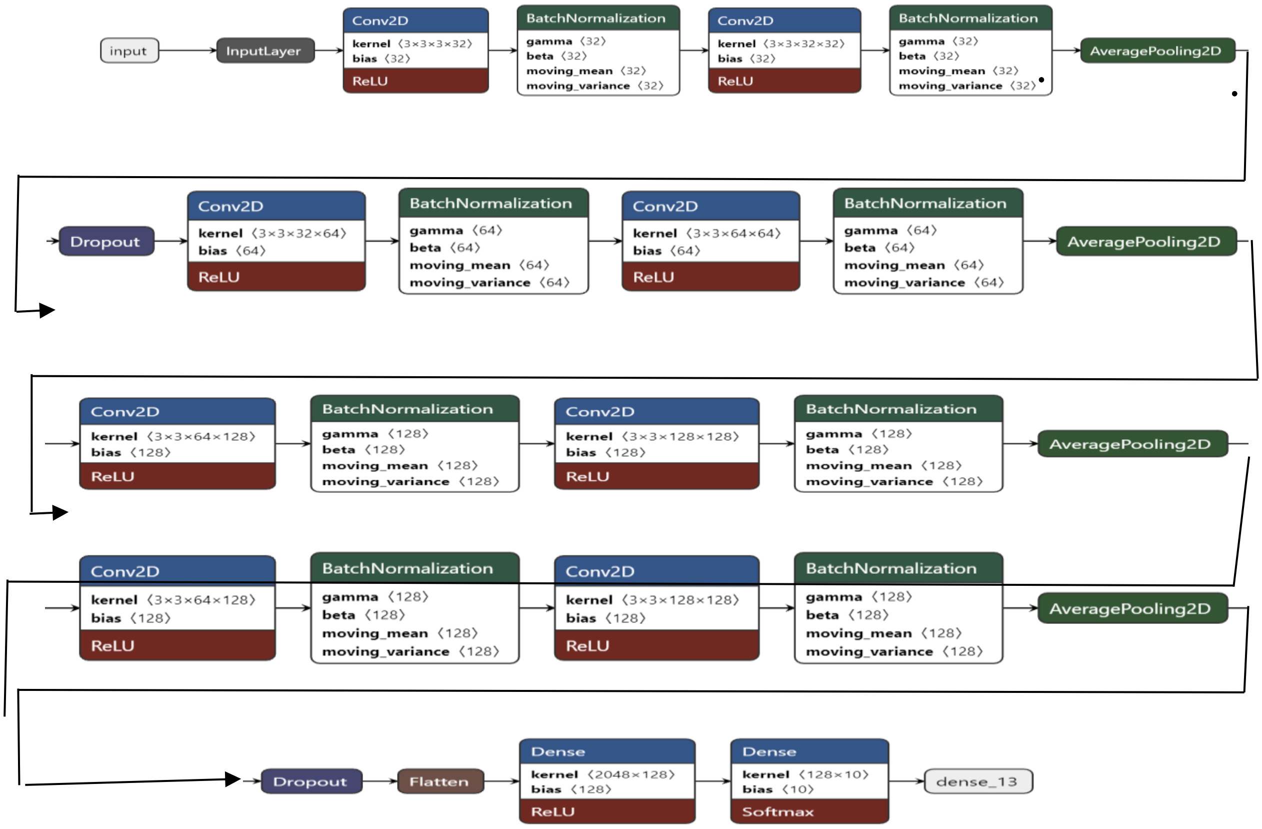


Figure 3. The proposed CNN architecture as drawn by Netron software

Parameters of the algorithms

The proposed system includes three main algorithms: CKKS algorithm used for encryption, CNN algorithm used for predict and extract high-level features, and finally random forest algorithm used to choose the best features. All these algorithms include a set of parameters that raise the efficiency of the proposed system illustrated in tables 2, 3 and 4:

Table 2. CNN algorithmic parameters

Parameter Name	Parameter value	Description
Kernel	3x3	This filter has been used for features extracted from images.
Padding	Same	As an image is processed by the kernel of a convolutional neural network (CNN), a concept known as padding is used to describe the number of pixels added to it.
Dropout	0.2,0.3,0.4	During training, specific neurons are dropped out of evaluation. Therefore, throughout the network, this study employed three distinct values.
Pooling (maxpooling)	2x2	The max-pooling has been used.
Optimizers(RMS prop)	learning_rate=0.01	Optimizers can change weights and the learning rate of your neural network to minimize losses. In this paper, the RMS prop method has been used.
Loss function	Categorical crossentropy	
Epochs	100	The cifar-10 dataset is processed through the neural network 100 times.
batch_size	100	In a single batch, there are 100 images for training purposes.
weight_decay	0.0001	

Table 3. CKKS algorithmic parameters

Parameter name	Parameter value	Description
Polynomial degree	8 bits	All of these parameters are described in the section below
cipher modulus	600 bits	
big modulus	1200 bits	
Scaling factors	100 bits	

Table 4. Random forest algorithmic parameters

Parameter name	Parameter value	Description
n_estimators	30	The forest's tree count.
criterion	Gini, entropy	The ability to evaluate the quality of a split. "gini" for Gini impurity and "entropy" for information gain is supported criteria.
max_depth	10	a tree's maximum depth
Min samples split	2	The smallest number of samples needed to break apart an internal node into its parts
Min samples leaf	1	There is a minimum number of samples that must be taken at each leaf node.

Feature extraction phase

CNN is proposed as a feature extraction method by utilizing the output of model flattening, which converts all the resulting 2-dimensional arrays into a single long continuous linear vector. It extracts 2048 features for each training image and stores these feature vectors as a matrix for all the training images.

Online processing phase

This phase begins when the client requests the public key from the server to encrypt the image and later sends the encrypted image to the server. In this phase using CKKS as the FHE algorithm, the client initially takes a message (M) and convert it to cipher vectors ([⟨ct1,pk⟩]%q,[⟨ct2,pk⟩]%q), where ct1 and ct2 are ciphertexts represented in polynomial form, pk represents the public key, and q refers to the ciphertext modulus. This phase is implemented on the server and client sides as follows:

The initial security parameter λ step (server-side, client-side): this step carries out on the server-side and the client-side, and there are many security parameters:

The polynomial degree N (a power of two): Degree d of polynomial that determines the quotient ring R.

- The ciphertext modulus q:
- The large modulus Q. For both schemes, a larger modulus q allows us to perform more homomorphic operations before the noise gets

too large for testing. Use 41 values of q, anywhere from 40 bits to 1200 bits. However, q and Q are upper bounded based on the security parameter λ and polynomial degree[36].

- The scaling factor Δ. A more significant value of the scaling factor Δ yields more precision but allows fewer homomorphic operations. More specifically, for a plaintext polynomial $(x) \in Rq$ where (x) = Encode (z, Δ), after that compute the error in the result of decoding (x) + (x) for some accumulated error. In practice, also use this estimate to choose Δ. For example, choosing Δ = 215 gives us a final decoded message with 4 bits of precision. If 6 bits of precision have been chosen, choose Δ = 217. Certain operations must increase the size of the scaling factor if the size of our decoded slots becomes small[37].

Generating keys step (server side, client side): Both public (pk) and private keys (sk) are generated on the server side and the client, and each side uses its keys to accomplish the encryption and decryption operations.:

Encrypting image (server side, client side): After receiving the public key from the client, the client performs the image encryption procedures. It resizes the image to scale 32x32 using the bilinear method, resizes the image used in order to be suitable for the model that was produced on the server side, and then it reads the image pixel by pixel and encodes each using CKKS encoding. In this step, every pixel in the image is converted into a polynomial with degree 8. Also, the server side encrypts the retrieved images in the same manner, but with the client's public key.

Decrypting image (server side, client side): The encrypted image was created on the client side and depends on the public key sent by the server. This image is sent to the server to retrieve the top five similar images to this image. This process is done by decrypting the image with the private key generated on the server side. The output of the decryption process is in a polynomial form, so a decoding of the output is used in order to retrieve the real values of the pixels. The client also utilizes the private key to open images received from the server in the same way that the server does.

Similarity matching (server side only): This step of the image retrieval process, by using the model (M), extracts the features of the sent image based on the flatten layer. The Hamming distance approach is used to determine the similarity of the five most similar images to the sent image using the features extracted and the feature vectors stored in the training phase.

Results:

There are many results from this proposed model illustrated as follows:

Augmentation data result

Fig.4 shows an example of the four augmentations used, the rotating, horizontal shifting, vertical shifting, and flipping. There are 50,000 new photos, as indicated in Table 5, when the rotation procedure is applied to the original training images, and the lost value is compensated for using Bilinear Interpolation on the new images. The rotation was applied at a 15-degree angle. The table above shows that 50,000 new images have been generated by applying horizontal shift to every image in training data in the x-direction, right-to-left one byte. Also, this Table shows that vertical shift augmentation and horizontal flipping are a feature of this augmentation method. These methods generate 100,000 new images. Also, four additional methods were used, namely Gaussian blurring, Local Binary Pattern, Gaussian noise, and Salt-and-pepper noise, to generate new data to become the new data set for training

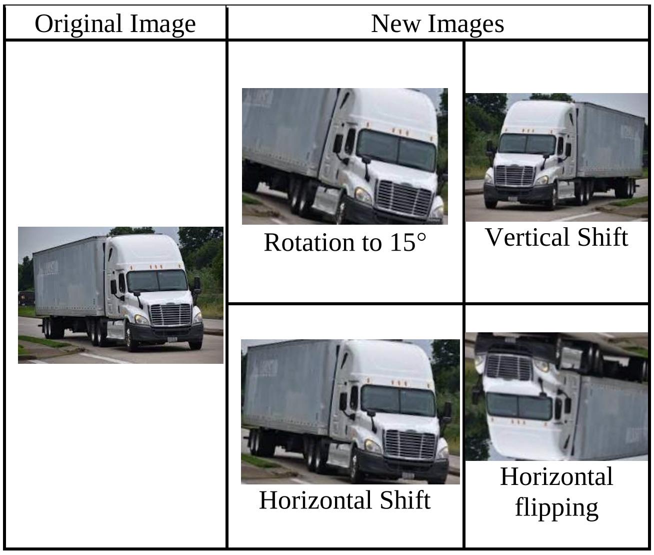


Figure 4. Result of augmentation process

Table 5. Dataset size after applied data augmentation

Dataset Type	No. of images
Original training dataset	50,000
Rotation dataset	50,000
Horizontal shift dataset	50,000
Vertical shift dataset	50,000
Gaussian blurring	50,000
Horizontal flipping dataset	50,000
Local Binary Pattern	50,000
Gaussian noise	5,0000
Salt-and-pepper noise	50,000
Total	450,000

CNN Layers

The proposed model has been implemented using ten CNN layers. Each image in the training set passes through all these layers. This network consists of six convolutional layers, three max-pooling layers, and one fully connected layer. Fig. 3 shows the results of the input images that have passed through the network layers.

Training Results

The model has been trained on 450,000 images (450,000 R, 450,000 G, 450,000 B) training images using optimization method RMSprop with an initial learning rate of 0.001.

The dataset goes through 100 epochs to enhance the images. In every single epoch, the weights are changed to get the image closer to the desired image. Table 6 illustrates the model accuracy and loss with the corresponding hyperparameters through the training stage. This Table shows that the number of times the training and validation samples were repeated is 100 times, each time 64 batches are taken. At first, an initial learning rate was 0.01, and after several epochs, the learning rate reached (0.0003).

Table 6. Training results

Iteration	Batch size	Learning rate		Loss function	Optimizer	Epochs	ETA	VAL-Loss	VAL-Accuracy	Loss	Accuracy
		Initial value	Last value								
100	64	0.001	0.0003	Categorical cross entropy	RMS prop	100	605s	0.3324	0.9843	0.3276	0.9794

Testing Results

The 10,000 testing images which represent 20% of the CIFAR-10 dataset have been input into the model architecture and went through all its layers. By using the saved parameters including the weights that the network reached and multiplying them by those weights, the testing images were classified into ten classes. Table 7 illustrates the results of testing accuracy, the time estimate for testing, and the loss value.

Table 7. Performance of testing

ETA	Loss	Accuracy
35s 4ms	0.330	98.87

The predicted classes of the test images have been compared with the actual test image classes to evaluate the trained network through the confusion matrix, which shows classification accuracy for the CIFAR-10 dataset, and through the calculation of recall, precision, and F1-score values as shown in Table 8. Also, Table 9 shows the classification result for each class.

Table 8. The Confusion matrix of CNN

	Airplane	Automobile	Bird	Cat	Deer	Dog	Frog	Horse	Ship	Truck
Airplane	997	1	0	0	0	1	1	0	0	0
Automobile	1	994	0	0	0	1	0	1	1	2
Bird	3	2	980	5	1	2	1	0	5	1
Cat	0	0	1	999	0	0	0	0	0	0
Deer	2	1	0	2	984	1	4	3	1	2
Dog	1	1	1	2	1	991	1	0	2	0
Frog	3	2	1	1	1	1	983	2	1	5
Horse	0	0	0	0	0	0	0	998	0	2
Ship	1	0	0	1	1	0	2	2	992	1
Truck	6	1	1	3	5	8	1	2	4	969

Table 9. Performance of classification for each class

Class	Precision	Recall	F1-score	Support
0	0.97	0.99	0.96	1000
1	0.99	0.98	0.95	1000
2	0.98	0.99	0.99	1000
3	0.97	0.96	0.98	1000
4	0.98	0.95	0.99	1000
5	0.96	0.98	0.96	1000
6	0.98	0.98	0.94	1000
7	0.99	0.96	0.98	1000
8	0.99	0.95	0.99	1000
9	0.98	0.99	0.98	1000

CNN Model Analysis

Machine learning relies heavily on visualization. Practitioners and academics frequently employ visualization to monitor learned parameters and output metrics to train and improve their models. Along with graph visualization, there is also a module for monitoring the distribution of tensors and images and sounds in the dashboard component TensorBoard. Fig.5 shows changes in loss and accuracy after every epoch. When an entire dataset is passed through a neural network, both forward and backward propagation – It is essential to understand loss and accuracy as training progresses and at what point these metrics are steady. Understanding this scaler graph will help prevent overfitting.Through this Figure, researchers notice a convergence in the increase between loess and accuracy for each training and validation samples. This indicator leads us to conclude that the proposed network in this research does not have a problem with overfitting.

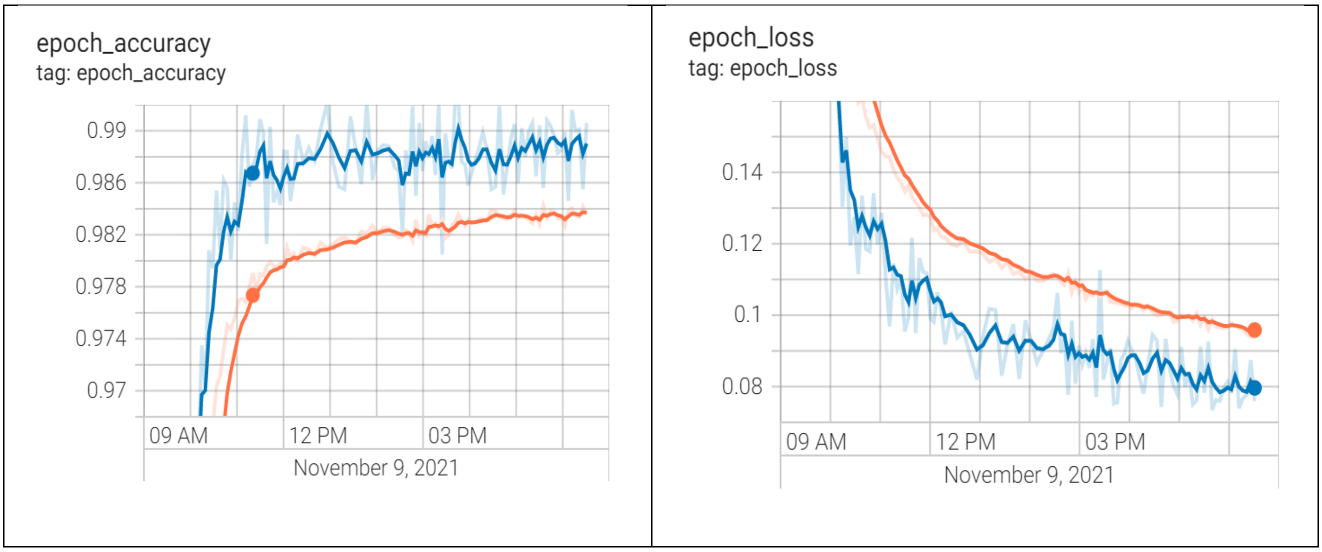


Figure 5. Analysis of the CNN model

NIST Tests Results

This part presents the result NIST for the CKKS algorithms on the CIFAR dataset. In this test, the results were converted from the cipher data, which is in the polynomial format, to binary, and then the NIST measurements were tested. Table 10 illustrates the results of this test.

Table 10. NIST test results of the CKKS algorithm

#	Test	CKKS Algorithm	Pass
1	Run	0.583633	True
2	serial	0.172892	True
3	random excursion variant	0.721127	True
4	random excursion	0.699500	True
5	non overlapping template matching	0.329871	True
6	Frequency Monobit	0.172104	True
7	Maurer’s universal statistical	0.813554	True
8	the longest run of ones in a block	0.999260	True
9	Linear complexity	0.942996	True
10	Frequency test within a Block	0.817504	True
11	Discrete Fourier transform	0.788236	True
12	Cumulative sums	0.138377	True

Timing test

Table 11 displays the time results for each encryption algorithm and deep learning method as well as the image retrieval time. All trials were performed on a computer with a dual-core processor, a clock speed of 2.7 GHz, and a memory capacity of 4 GB with pre-installed Windows 7.

Table 11.Time results

CKKS		CNN model (seconds)		Retrieve each image (seconds)
Encryption each image (seconds)	Decryption each image (seconds)	Training	testing	
23	21	46800	3670	0.06

Comparison with Previous Studies

Several methods have been proposed to enhance the retrieved images. Table 12 illustrates the results of image classification, while Table 13 shows the retrieval performance for CIFAR-10 datasets in mean average precision for different research projects.

Table 12. Image classification accuracy on CIFAR-10

	Image classification accuracy
Kua et. al [21]	96.9
Hsin et al [22]	90.19
RCNN_CKKS proposed system	98.87

Table 13. Image retrieval map on CIFAR-10

	MAP
Kua t al [21]	0.707
Hsin et al [22]	0.867
Umer et al [23]	0.913
RCNN_CKKS proposed system	0.967

Researchers can notice from these Tables that the proposed method performed better. The system analysis proposed by the researchers is depicted in the above section using TensorBoard analysis. Choose the best CNN architecture and parameters based on the results of this analysis. In addition, the random forest method provided valuable features for calculating the distance between stored vectors and input vector features.

Discussion:

Related work studied in this research included results that are less than this research. The presented results showed that the CNN suggested in this research was distinguished from other research with higher results in the classification and retrieval process. This research remarked that the flatting layer provided features used by retrieving similar images because this layer contains properties that depend on trained weights in the training phase. The Random Forest method has been used to choose the best features that give better results in terms of accuracy and time. In addition, the proposed data augmented method led to an increase in dataset samples used to overcome the problems of overfitting and increase the accuracy of training and noticed through previous research that the security of the data sent through the network is not taken care of. Most of the applications that need image retrieval are applications that need to maintain the privacy of the data of the sending person, so proposed a protocol that maintains data privacy.

Conclusions:

This paper has presented an effective content-based image retrieval CBIR system capable of semantically retrieving the correct images with high retrieval performance. More than one algorithm was used together to achieve the highest results. In the encryption part, the CKKS algorithm was used, which is one of the fully homomorphic encryption algorithms. In building a deep learning model, the algorithm CNN used and the random forest has been used to extract the best features, and hamming distance method is used to calculate the distance between the stored and input factors. The researchers propose a method for image retrieval based on CNN developed by taking advantage of the flatten layer, which extracts 2028 image features stored in one feature vector. Next, the researchers apply the random forest algorithm after this layer as features selection to produce 600 features that contribute increased accuracy compared to the previous research. A secure protocol was developed to preserve the data communicated through an insecure connection between the client and the server by using CKKS to provide the highest security. It also overcomes the overfitting issue and improves the msodel's accuracy by increasing the number of training images to use eight different augmentation methods. The researchers note that the CKKS approach is slow and requires more cipher image space, yet it is a powerful encryption method. Results for classification and retrieval were 97.94 and 98.94 percent, respectively. CKKS's safety was also evaluated using the NIST test.

The first contribution achieved in this paper is the trade-off between the accuracy of image retrieval while providing a safe environment for transferring images because of the importance of ensuring the security of applications used in the image retrieval field. Another contribution presented is building CNN model with high classification accuracy, which helped extract the best characteristics from the images. The high accuracy of classification comes through a vital contribution that includes increasing the number of dataset samples by adding a set of expected effects on the images, as previously shown. The last contribution presented by researchers includes using the random forest algorithm to select the best features extracted from the flatten layer. These features are used to calculate the distance between each stored image's features and the input images features, which increases image retrieval accuracy and reduces its time.

Many practical applications can benefit from this proposal, such as building a safe search engine and health fields, because they are the areas that require a safe environment when communicating between the patient and the hospital

One of the futures works is to construct a deep learning model for image retrieval performed on encrypted images by utilizing the capabilities of fully homomorphic encryption algorithms, which include the ability to perform addition and multiplication operations on encrypted data. Another future proposal is to retrieve multimedia by combining deep learning, blockchain, and fully homomorphic encryption algorithms.

Authors' declaration:

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are mine ours. Besides, the Figures and images, which are not mine ours, have been given the permission for re-publication attached with the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee in University of Technology.

Authors' contributions statement:

A.K and E.M. proposed this idea, A.K suggested the general outline of the proposal. Then E.M carried out each part, extracted the results, and discussed it with A.K to suggest improvements.

References:

1. Rout NK, Atulkar M, Ahirwal MK. A review on content-based image retrieval system: Present trends and future challenges. Int J Comput Vis Robot. 2021; 11(5): 461-485. doi:10.1504/IJCVR.2021.117578
2. Murala S, Maheshwari RP, Balasubramanian R. Local tetra patterns: A new feature descriptor for content-based image retrieval. IEEE Trans Image Process. 2012; 21(5): 2874-2886. doi:10.1109/TIP.2012.2188809
3. Onoufriou G, Mayfield P, Leontidis G. Fully Homomorphically Encrypted Deep Learning as a Service. Mach Learn Knowl Extr. 2021; 3(4): 819-834. doi:10.3390/make3040041.
4. Denning DER. Cryptography and Data Security. 1982; Addison-Wesley Longman Publishing Co., Inc. Boston,MA United States. http://portal.acm.org/citation.cfm?id=SERIES11430.539308
5. Yassein HR, Al-Saidi NMG, Farhan AK. A new NTRU cryptosystem outperforms three highly secured NTRU-analog systems through an innovational algebraic structure. J Discret Math Sci Cryptogr. 2020;(June). doi:10.1080/09720529.2020.1741218
6. Armknecht F, Katzenbeisser S, Peter A. Group homomorphic encryption: Characterizations, impossibility results, and applications. Des Codes, Cryptogr. 2013;67(2):209-232. doi:10.1007/s10623-011-9601-2
7. Gentry C. A Fully Homomorphic Encryption Scheme. PhD [dissertation]Stanford Univ. 2009;(September). http://cs.au.dk/~stm/local-cache/gentry-thesis.pdf
8. Plantard T, Susilo W, Zhang Z. Fully homomorphic encryption using hidden ideal lattice. IEEE Trans Inf Forensics Secur. 2013; 8(12): 2127-2137. doi:10.1109/TIFS.2013.2287732
9. Chung H, Kim M, Al Badawi A, Aung KMM, Veeravalli B. Homomorphic comparison for point numbers with user-controllable precision and its applications. Symmetry (Basel). 2020; 12(5): 1-22. doi:10.3390/SYM12050788
10. Pedrouzo-Ulloa A, Troncoso-Pastoriza JR, Gama N, Georgieva M, Pérez-González F. Revisiting multivariate ring learning with errors and its applications on lattice-based cryptography. Mathematics. 2021; 9(8): 1-42. doi:10.3390/math9080858
11. Liu J, Wang C, Tu Z, Wang XA, Lin C, Li Z. Secure KNN Classification Scheme Based on Homomorphic Encryption for Cyberspace. Secur Commun Networks. 2021; 2021: 1-12. doi:10.1155/2021/8759922
12. Amine Boulemtafes, Abdelouahid Derhab YC. A review of privacy-preserving techniques for deep learning. Neurocomputing, Elsevier. 2020; 384(4): 21-45. doi:10.1016/j.neucom.2019.11.041
13. Kadhim AF, Kamal ZA. Generating dynamic S-BOX based on Particle Swarm Optimization and Chaos Theory for AES. Iraqi J Sci. 2018; 59(3): 1733-1745. doi:10.24996/IJS.2018.59.3C.18
14. Xu Y, Zhao X, Gong J. A Large-Scale Secure Image Retrieval Method in Cloud Environment. IEEE Access. 2019; 7: 160082-160090. doi:10.1109/ACCESS.2019.2951175
15. Xia Z, Ma X, Shen Z, Sun X, Xiong NN. Secure Image LBP Feature Extraction in Cloud- based Smart Campus. IEEE Access. 2018; PP(c): 1. doi:10.1109/ACCESS.2018.2845456
16. Cao Z, Mu S, Xu Y, Dong M. Image retrieval method based on CNN and dimension reduction. In 2018. Int. Conf. Secur. Pattern Anal. Cybern, SPAC 2018; 2018: 441-445. doi:10.1109/SPAC46244.2018.8965601
17. Verma M, Raman B. Local neighborhood difference pattern : A new feature descriptor for natural and texture image retrieval. Multimed Tools Appl Springer Sci. 2018; 77: 11843–11866. doi:10.1007/s11042-017-4834-3
18. Selvam S, Kannan ST. A New Architecture for Image Retrieval Optimization with HARP Algorithm. Asian

J Comput Sci Technol. 2017; 6(1): 1-5.

19. Du A, Wang L, Cheng S, Ao N. A privacy-protected image retrieval scheme for fast and secure image search. Symmetry. 2020; 12(2):1-17. doi:https://doi.org/10.3390/sym12020282.

20. Khokher A, Talwar R. Content-based image retrieval at the end of the early years. IEEE Trans Pattern Anal Mach Intell. 2000;22(12):1349-1380.

21. He Q, He H. A novel method to enhance sustainable systems security in cloud computing based on the combination of encryption and data mining. Sustain. 2021;13(1):1-17. doi:10.3390/su13010101

22. Kuo CH, Chou YH, Chang PC. Using deep convolutional neural networks for image retrieval. In: Visual Information Processing and Communication. IS&T Int. Symp. Electron. Imaging Sci Technol. 2016: 1-6. doi:10.2352/ISSN.2470-1173.2016.2.VIPC-231

23. Huang HK, Chiu CF, Kuo CH, Wu YC, Chu NNY, Chang PC. Mixture of deep CNN-based ensemble model for image retrieval. 2016 IEEE 5th Glob Conf Consum Electron GCCE 2016. 2016; (2): 5-6. doi:10.1109/GCCE.2016.7800375

24. Khan UA, Javed A, Ashraf R. An effective hybrid framework for content based image retrieval (CBIR). Multimed Tools Appl. 2021; 80(17): 26911-26937. doi:10.1007/s11042-021-10530-x

25. Ali F, Mohammed AH. Content Based Image Retrieval (CBIR) by statistical methods. Baghdad Sci J. 2020;17:694-700. doi:10.21123/bsj.2020.17.2(SI).0694

26. Challa RK, Gunta VK. A modified symmetric key fully homomorphic encryption scheme based on Read-Muller Code. Baghdad Sci J. 2021; 18(2): 899-906. doi:10.21123/bsj.2021.18.2(Suppl.).0899

27. Syed D, Refaat SS, Bouhali O. Privacy preservation of data-driven models in smart grids using homomorphic encryption. Inf. 2020;11(7):1-17. doi:10.3390/info11070357

28. Lou Q, Jiang L. SHE: A fast and accurate deep neural network for encrypted data. Adv Neural Inf Process Syst. 2019; 32(NeurIPS):1-9.

29. Obla S, Gong X, Aloufi A, Hu P, Takabi D. Effective Activation Functions for Homomorphic Evaluation of Deep Neural Networks. IEEE Access. 2020; 8: 153098-153112. doi:10.1109/ACCESS.2020.3017436

30. Kwabena OA, Qin Z, Qin Z, Zhuang T. MSCryptoNet: Multi-Scheme Privacy-Preserving Deep Learning in Cloud Computing. IEEE Access. 2019; 7: 29344-29354. doi:10.1109/ACCESS.2019.2901219

31. Clet P-E, Stan O, Zuber M. BFV, CKKS, TFHE: Which One Is the Best for a Secure Neural Network Evaluation in the Cloud? Springer International Publishing; 2021. doi:10.1007/978-3-030-81645-2_16

32. Zhang Q, Zhang M, Chen T, Sun Z, Ma Y, Yu B. Recent advances in convolutional neural network acceleration. Neurocomputing. 2019; 323: 37-51. doi:10.1016/j.neucom.2018.09.038

33. Tzelepi M, Tefas A. Deep convolutional learning for Content Based Image Retrieval. Neurocomputing. 2018; 275: 2467-2478. doi:10.1016/j.neucom.2017.11.022

34. Bologna G. A Simple Convolutional Neural Network with Rule Extraction. Appl Sci. 2019; 9(12): 2411. doi:10.3390/app9122411

35. Zaid Khalaf Hussien BND. Anomaly Detection Approach Based on Deep Neural Network and Dropout. Baghdad Sci J. 2020; 17:701-709.

36. Will MA, Ko RKL. A Guide to Homomorphic Encryption. Elsevier Inc. 2015. doi:10.1016/B978-0-12-801595-7.00005-7

37. Shrestha R, Kim S, Integration of IoT with Blockchain and Homomorphic Encryption: Challenging Issues and Opportunities. Adv Comput. 2019; 115. 1st ed. Elsevier Inc. doi:10.1016/bs.adcom.2019.06.002

38. Hee CJ, Andrey K, Miran K, Yongsoo S. Homomorphic Encryption for Arithmetic of Approximate Numbers. In: Inte Conf T Appl Crypt and Inf S Springer. 2017: 409–437. doi:10.1007/978-3-319-78381-9_14

استرجاع الصور المشفرة باستخدام الشبكة العصبية الالتفافية والتشفير المتماثل الكامل

عماد محمد عبود علاء كاظم فرحان

قسم علوم الحاسوب ، الجامعة التكنلوجية ، بغداد، العراق

الخلاصة:

استرجاع الصور المستند إلى المحتوى (CBIR) هو تقنية تستخدم لاسترداد الصور من قاعدة بيانات الصور. ومع ذلك، فإن عملية CBIRتعاني من دقة أقل في استرداد الصور من قاعدة بيانات صور واسعة النطاق وضمان خصوصية الصور. تهدف هذه الورقة إلى معالجة قضايا الدقة باستخدام تقنيات التعلم العميق كطريقة .CNN أيضًا، توفير الخصوصية اللازمة للصور باستخدام طرق تشفير متماثلة تمامًا بواسطة Cheon و Kim و Kim و .Song (CKKS) ولتحقيق هذه الأهداف تم اقتراح نظام RCNN_CKKS يتضمن جزأين. يستخرج الجزء الأول (المعالجة دون اتصال بالإنترنت–) لاستخراج الخصائص العالية المستوى استنادًا إلى طبقة التسطيح في شبكة عصبية تلافيفية (CNN) ثم يخزن هذه الميزات في مجموعة بيانات جديدة. في الجزء الثاني (المعالجة عبر الإنترنت) ، يرسل العميل الصورة المشفرة إلى الخادم ، والتي تعتمد على نموذج CNN المدرب لاستخراج ميزات الصورة المرسلة. بعد ذلك، تتم مقارنة الميزات المستخرجة مع الميزات المخزنة باستخدام طريقة Hamming Distance لاسترداد جميع الصور المتشابهة. أخيرًا، يقوم الخادم بتشفير جميع الصور المسترجعة وإرسالها إلى العميل. كانت نتائج التعلم العميق على الصور العادية 97.94٪ للتصنيف و98.94٪ للصور المسترجعة. في الوقت نفسه، تم استخدام اختبار NIST للتحقق من أمان CKKS عند تطبيقه على مجموعة بيانات المعهد الكندي للأبحاث المتقدمة .(CIFAR-10) من خلال هذه النتائج، استنتج الباحثون أن التعلم العميق هو وسيلة فعالة لاستعادة الصور وأن طريقة CKKS مناسبة لحماية خصوصية الصورة .

الكلمات المفتاحية: CKKS، CNN، استرجاع الصور على أساس المحتوى، تشفير متماثل الشكل، غابة عشوائية