mathematics

Article

Searchable Encrypted Image Retrieval Based on Multi-Feature Adaptive Late-Fusion

Wentao Ma, Jiaohua Qin*, Xuyu Xiang, Yun Tan and Zhibin He

College of Computer Science and Information Technology, Central South University of Forestry & Technology, Changsha 410004, China; mawentaow@163.com (W.M.); xyuxiang@163.com (X.X.); tantanyun@hotmail.com (Y.T.); hzb919086716@163.com (Z.H.)

* Correspondence: qinjiaohua@csuft.edu.cn

Received: 27 April 2020; Accepted: 12 June 2020; Published: 22 June 2020

Abstract: Recently, searchable encrypted image retrieval in a cloud environment has been widely studied. However, the inappropriate encryption mechanism and single feature description make it hard to achieve the expected effects. Therefore, a major challenge of encrypted image retrieval is how to extract and fuse multiple efficient features to improve performance. Towards this end, this paper proposes a searchable encrypted image retrieval based on multi-feature adaptive late-fusion in a cloud environment. Firstly, the image encryption is completed by designing the encryption function in an RGB color channel, bit plane and pixel position of the image. Secondly, the encrypted images are uploaded to the cloud server and the convolutional neural network (CNN) is fine-tuned to build a semantic feature extractor. Then, low-level features and semantic features are extracted. Finally, the similarity score curves of each feature are calculated, and adaptive late-fusion is performed by the area under the curve. A large number of experiments on public dateset are used to validate the effectiveness of our method.

Keywords: searchable encryption; multi-feature adaptive late-fusion; convolutional neural network

1. Introduction

With the rapid development of mobile Internet and intelligent imaging devices, image-based multimedia data is exploding. In the face of the dramatic growth of images, Content-Based-Image-Retrieval (CBIR) is a promising technology that can quickly help us find images of interest [1–3]. However, offline storage space is limited, so how to store, process and manage data safely and efficiently is a challenge. As a novel emerging service mode, cloud computing provides users with convenient, low-cost computing and storage [4]. Nevertheless, any technical service has two sides. On the one hand, we can enjoy the convenience of cloud servers. On the other hand, the cloud server is not entirely trustworthy to the data owner. For example, malicious cloud server operators or external hackers illegally obtain valuable sensitive data and use it for business transactions [5].

To solve privacy-preserving, the existing searchable encryption method is to encrypt the data before outsourcing to the cloud server [6]. However, this encrypted search mechanism increases the computational burden of data owners and query users, which is not feasible when actually processing large-scale image storage. Thus, the researchers proposed outsourcing computationally intensive steps to cloud servers to ease the burden on data owners and query users. In addition, with the development of Artificial Intelligence (AI), CBIR has been further developed into a multi-level retrieval scheme for low-level features and semantic features. For low-level features, image retrieval and searchable encryption are key information to extract images, such as SIFT, SURF, Harris and MPEG of local key points [7–11]. For semantic features, deep learning is widely used for its efficient feature description and accurate retrieval results [12]. In this paper, we adopt the fine-tuned convolutional neural network

(CNN) model to extract features. In general, features generated by deep learning can simulate human perception to a certain extent through convolution, pooling and other operations, so they have better feature description than low-level features. Deep learning works by mimicking the firing of human neurons: some nerve cells in the brain respond only when there are edges in a particular direction [13]. For example, when we look very closely at a picture of a human face, only some of the neurons in our brain are activated, and we can only see the pixel level of a person's face. When we pull apart a little bit, neurons in other parts of the brain are activated. We can also observe the lines of the face to the pattern to the local to the whole face, which is a step by step process to obtain high-level semantic features. Some classical deep learning network models mainly include VGG [14], ResNet [15] and DenseNet [16]. Some researchers have proposed a lightweight model, which has high image classification accuracy and greatly reduces the amount of parameter calculation [17]. In these models, DenseNet has an extremely high feature utilization rate, narrow network structure and fewer parameters than other models. Ref. [2] shows that in image retrieval, both SIFT based low-level features and CNN based semantic features can achieve better results, but these methods adopted single feature as descriptors and there is still a room for improvement in performance. Therefore, to break the upper limit of accuracy brought by a single feature, researchers proposed many effective fusion schemes, such as feature-usion [18,19], index-level fusion [20] and score-level fusion [21]. For the development of feature fusion, this paper proposed a searchable encrypted image retrieval method based on multi-feature adaptive late-fusion in cloud environment: the similarity score curve of good features is "L-shaped", while the score curve of bad features is gradually and slowly decreasing. The system framework of this paper is shown in Figure 1. Firstly, the image owner encrypts the images by the designed encryption function and uploads the encrypted image to the cloud server. Secondly, DenseNet model is fine-tuned with encrypted image on edge computing platform. Finally, feature extraction, similarity score calculation and adaptive late-fusion are completed on the cloud server. The main work of this paper are as follows:

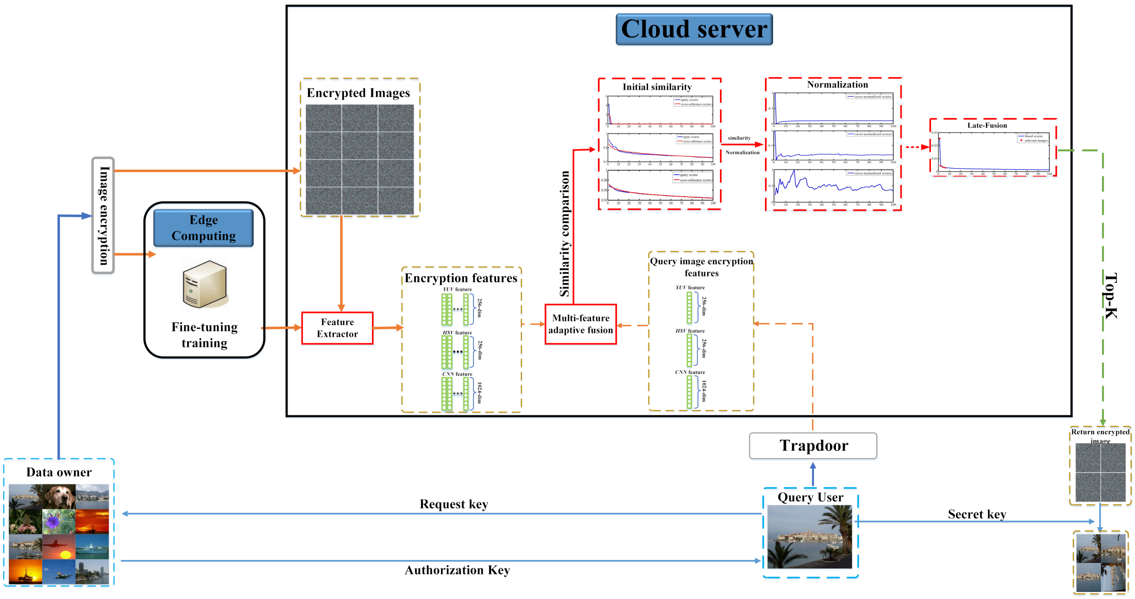

Figure 1. The system framework of our method.

(1) To improve the effectiveness of encryption method. Image encryption is to protect color and texture information. In this paper, RGB color channel replacement is adopted to protect the color information, the bit-plane sequence is converted to 8-bit binary random scrambling and Zig-Zag scan scrambling the pixel position are used to protect the texture information. Combined with the three manners to complete the image encryption, which is more secure and effective than the standard encryption method.

(2) Low-level features and semantic features. In this paper, the pre-trained DenseNet model is fine-tuned by using encrypted images on the edge computing platform, and then semantic features and low-level features are extracted on the cloud server, which greatly reduces the computing burden of the data owner.

(3) Adaptive late-fusion of multi-features. The feature similarity score is calculated on the cloud server and sorted into curves. Then the validity of each feature is negatively correlated with the area under the score curve for effective fusion.

The rest of this paper is structured as follows. The related work is described in Section 2, and the Section 3 introduces the methods and techniques. The features used in the experiment and the proposed adaptive late-fusion are described in Section 4. Section 5 analyzes the experimental results. Finally, the conclusion is drawn in Section 6.

2. Related Works

2.1. Feature Encrypted Image Retrieval

Image retrieval has been widely studied. However, image data cannot be outsourced directly to the cloud server for security. Therefore, Searchable-Encryption-Scheme (SSE) was proposed to solve the problem of privacy-preserving. This method allows the data owner to encrypt the image and upload it to the cloud server, while still maintaining the image retrieval attributes [22]. Lu et al. [23] first proposed a CBIR method for privacy-preserving based on encrypted image. In this approach, visual words generated by clustering are used to represent image information. Meanwhile, visual words are encrypted by the minimum hash algorithm to ensure the privacy of features is not disclosed. Then, the Jaccard distance between visual words is calculated. In another study, Lu et al. used bit plane randomization, random projection and random meta coding to encrypt and protect features [24]. This method combines signal processing technology with cryptography technology and can effectively protect image. However, in practice due to the large amount of computation, it's not suitable for large-scale image retrieval.

In order to improve the performance of above methods, the researchers also proposed many improvement methods. Xia et al. [8] proposed CBIR method based on Scaling Invariant Feature Transform (SIFT) and Earth Movers Distance (EMD) privacy-preserving. In essence, this method adopted linear transformation to solve EMD problem and privacy-preserving. Weng et al. [25] proposed a multimedia retrieval method for privacy-preserving, that's, image owners adopted robust hashing and partial encryption to protect image features. The server adopted the unencrypted hash to retrieve similar images, while the query user adopted plain text hash. Xia et al. [26] proposed an encrypted image retrieval method using four MPEG descriptors to represent image features. The K-means is adopted to ensure that the image information is not leaked, and the local sensitive hash manner is used to improve the efficiency. In addition, there are several other research contributions that can also be applied to feature encrypted image retrieval. Tan et al. [27] proposed the robust non-blind watermarking schemes in YCbCr color space based on channel coding. The watermarking scheme based on channel coding proposed in this method can achieve nearly accurate watermarking recovery under various attacks. Luo et al. [28] proposed a coverless information hiding method based on deep learning in order to improve the security of encrypted information. This method adopts DCT to generate a robust hash sequence with feature sequence, DC and location.

2.2. Searchable Encrypted Image Retrieval

By outsourcing CBIR to the cloud server, we not only want to preserve the data privacy, but also make full use of the computing power of the cloud server. The feature encrypted image retrieval has made gratifying performance in outsourcing CBIR. Although these methods solve the problem of limited offline storage space to some extent, they have a common disadvantage: the amount of computation required by the data owner is large. That is, the data owner performs feature extraction, feature encryption, index construction and other operations. Therefore, to reduce the computing cost of data owners, Ferreira et al. [29] proposed an Image Encryption Scheme Content-based Image Retrieval (IES-CBIR), which encrypts by random scrambling rows and columns of the pixel matrix, and then uploads the encrypted image to the cloud server. Then the cloud server extracts HSV (Hue-Saturation-Value, which represents the color histogram features of the image.) features from the encrypted image, and finally matches the similarity through the hamming distance between feature vectors. The advantage of the IES-CBIR is that it significantly reduces the amount of computation for the image owner and puts more computing overhead on the cloud server. However, this method adopted a color histogram as the feature vector, ignoring the texture information [30] and semantic feature information [12].

2.3. Feature Fusion Image Retrieval

The results show that feature fusion is an effective image retrieval method. Zhang et al. [31] proposed a specific query fusion based on undirected graphs to fuse multiple retrieves. In this method, the graph satisfying the neighborhood relation is connected, and the edge weight is measured by the Jaccard similarity, then reordered the images by link analysis. But graph-fusion was susceptible to Outliers, based on which Liu et al. [32] proposed a graph-fusion with strong robustness to Outliers. In terms of Index-level fusion, Zhang et al. proposed "Semantic Aware Co-indexing" [33], which combined SIFT local features with semantic features to improve the discrimination of indexes.

However, when multiple features are fused, it is not known which features are effective and which features are bad for a given query. Thus, it is very important to distinguish the validity of features and fuse them by means of adaptive query. Based on, this paper proposes a searchable encrypted image retrieval based on multi-feature adaptive late-fusion in cloud environment. This method has two advantages: first, our method extracts encrypted image features, similarity calculations, and CBIR services on cloud servers, greatly reducing the computing costs of data owners. Second, according to the similarity score curve of good features, it's L-shaped, while the similarity score curve of inefficient features decreases slowly. The validity of each feature is negatively correlated with the area under the similarity score curve, which is used for weight allocation and fusion.

3. Technical Overview and Programme Framework

Before describing our method in detail, we briefly explain the terminology, security model, and design goals. To improve the performance of the existing methods, we proposed a searchable encrypted image retrieval method based on multi-feature adaptive late-fusion in the cloud environment, as shown in Figure 1. The framework is mainly composed of three parts: one is executed by the image owner on the edge computing platform, including image encryption and fine-tuning DenseNet model. Second, it's done by the cloud server, including encrypted image storage and CBIR services (feature extraction of encrypted images, similarity calculation, match retrieval and adaptive late-fusion). Third, it's performed by the query user, including constructing trapdoor, requesting authorization and issuing query command to the cloud server.

3.1. Entity Model of System

The system framework of our approach involves three different types of entities: image owner, cloud servers, and query user, as shown in Figure 1.

Image Owner. The image owner has a dataset containing a large number of images and the dataset is denoted as $M = \{m_i\}_{i=1}^{n}$ with a corresponding identity set $INDEX = \{index_i\}_{i=1}^{n}$, where n represents the number of images in the dataset. To save computing costs and flexibility, the image owner outsources the image to the server and encrypts the outsourced image to prevent privacy. The generated encrypted image set is represented as $E = \{e_i\}_{i=1}^{n}$. The identifying index of the set remains the same as before encryption, where m_i and e_i represent the i image of the original dataset and the encrypted dataset, respectively.

Cloud Server. The image owner stores the encrypted image dataset $E = \{e_i\}_{i=1}^{n}$ in the cloud server, solving the problem of limited offline storage space. Meanwhile, the feature extractor trained by the edge computing platform is uploaded to the cloud server for extract the semantic features of the encrypted images. Once a search request is received from the query user, the server finds the most similar ones from the encrypted image dataset and returns them to the query user. In our method, the cloud server completes the data storage and CBIR service, reducing the computation of the data owner to some extent.

Query User. Query users expect to search for images of interest from encrypted image set. To protect the query image, the query user encrypts the query image to generate the trapdoor and transmits the trapdoor to the server. The user needs to request authorization from the image data owner to obtain a trusted authorization and key. Then the cloud server returns the retrieved Top-k similar images to the user, who decrypts it with the key to get the image content.

3.2. Security Model of System

In this paper, we will protect the privacy of image owners and query users. Similar to the SSE [13,15,25,29], on the one hand, cloud servers are considered Honest-But-Curious (HBC). The cloud server will execute the relevant protocol rules correctly, but will also save and analyze the sensitive information. On the other hand, we assume that both the image owner and the authorized user are fully trusted, neither will disclose any information to the cloud server. If the above hypothesis is true, in this case the encrypted image e_i and e_j belong to the same category, and they have a high similarity to the same query image e_{iq} and e_{jq}, then the cloud server will infer that e_i and e_j are similar. This paper does not consider the information leakage caused by this situation. The abov- mentioned information leakage and access mode are a more advanced privacy leakage.

The cloud server can access all the data in RAM and can potentially classify the interested parts according to the traces of the query users, but this form of information leakage has a relatively little impact. If the leakage of access mode is considered, there will be a large computing and communication overhead. Therefore, based on the above two points, we don't consider the privacy disclosure of access mode.

3.3. Design Goals of System

In order to query encrypted images safely, effectively and with high precision under the above security model, our method has the following design goals.

Security. The image dataset uploaded to the cloud server and the query image are encrypted to ensure the security of the communication. The cloud server doesn't need to know anything about the image, including its content, image features, and trapdoor construction mechanism.

Efficiency and Accuracy. Feature extraction and other steps are performed on the cloud server, which makes full use of the computing power of the cloud server and greatly improves the efficiency. Meanwhile, semantic features are used to improve accuracy.

4. Our Method

We have briefly described the technical outline and scheme framework. In this section, details of the proposed framework are introduced, mainly including five parts: key generation, image encryption, fine-tuning DenseNet model, multi-feature adaptive late-fusion retrieval service, and construction of trapdoor, as shown in Figure 2.

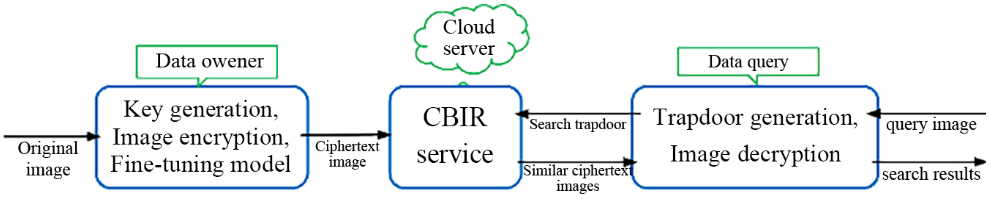

Figure 2. The details of our method.

4.1. Method Definition

Before introducing the method in detail, we first define the image privacy-preserving in this paper. Generally speaking, the image content includes color information and texture information, which together constitute the content of the image. In essence, privacy-preserving is to prevent unauthorized users from recognizing the content of the image, where the color information of images is determined by the color pixel of different channels, while the texture information is determined by the relative position of the pixel and the color change between adjacent pixels.

Based on the above definition and the research of IES-CBIR, we designed the image encryption method, which is a way to encrypt both color and texture information. In view of the strong correlation between color and texture information in image classification and retrieval, RGB color channel replacement is adopted to protect the color information, the bit-plane sequence is converted to 8-bit binary random scrambling and Zig-Zag scan scrambling the pixel position are used to protect the texture information. Through three encryption manners, searchable encrypted image retrieval based on multi-feature adaptive late-fusion can be carried out on the cloud server without the intervention of the data owner.

4.2. Key Generation

By definition, the essence of encryption is to protect the color and texture information of images. This manner protects the color information by replacing the RGB color channel with random number keys. There are two steps to protect the texture information of images. Firstly, the image pixel sequence is converted to 8-bit binary random scrambling, then the position of pixels is scrambled by Zig-Zag scanning. Therefore, the generated key group can be represented as: $K = \{(RandNum, \{RGB_r\}_{r=0}^{5}, key_1), key_2\}$.

Where $RandNum$ is a random number generator, which can randomly generate an integer. In the color information of protected image, there are six combinations of RGB channel permutation, and the range of random integers dividing is 6. There are $ImgSize$ pixels in each channel, and the pixels in corresponding positions between channels are randomly permutated. Therefore, a random integer generator $RandNum$ and a key $\{RGB_r\}_{r=0}^{5}$ are used to determine the encryption method of color information. The formula is as follows.

$$rgb^* \leftarrow (RandNum, \{RGB_r\}_{r=0}^{5}) \tag{1}$$

To encrypt the texture information of images, in addition to converting the color channel pixel sequence to 8-bit binary random scrambling, we also need to conduct Zig-Zag scanning scrambling for the pixel position. The pixels of each color channel are randomly scrambled by using the key key_1 to 8-bit binary.

$$randperm^* \leftarrow (key_1) \tag{2}$$

The Zig-Zag scanning scrambling encryption of pixel position is to scan and scramble the two-dimensional pixel matrix of RGB color channel.

$$zig^* \leftarrow (key_2) \tag{3}$$

RGB channel replacement encryption is not very secure, but the image sequence changes by the color channel pixel replacement. Therefore, converting the image to 8-bit binary random scrambling can greatly improve the security of the image. Finally, image encryption is accomplished by Zig-Zag scanning scrambling. The key groups generated by the above three encryption manners are mainly used to authorize query users. When the image owner receives the authorization request from the query users, the key groups are sent to the query users.

4.3. Encryption Algorithm

In part B, which introduces key generation, we have outlined the main steps of the encryption manner: RGB channel scrambling, image sequence conversion to 8-bit binary random scrambling, image pixel matrix Zig-Zag scanning scrambling. In this section, we describe the three encryption manners in detail, and then define an algorithm to describe the complete encryption process.

Color Channel Replacement Encryption. In the key generation method, there are six RGB channel scrambling methods, namely $\{RGB_r\}_{r=0}^{5}$.The unencrypted RGB color channel is represented by C. The three channel pixel matrices are C^R, C^G and C^B respectively. $c_{ij} = (c_{ij}^R, c_{ij}^G, c_{ij}^B)$ denotes the pixels at the position of the three-channel pixel matrix (i, j), where $(i, j) \in ImgSize$, $ImgSize$ is the number of pixels each color channel of an image. The corresponding encrypted RGB color channel is C'. The three channel pixel matrices are C^{R^*}, C^{G^*} and C^{B^*} , respectively. The pixel in the three-channel position (i, j) is $c'_{(ij)} = (c_{ij}^{R'}, c_{ij}^{G'}, c_{ij}^{B'})$. The scrambled and encrypted image of the color channel pixels is represented as m'. Define the Algorithm 1: ChannelEnc.

Algorithm 1 ChannelEnc

```
Input: image $m$ and the key $rgb^*$.
Output: Color encrypted image $m'$.
Generate the key $rgb^*$ by Equation(1).
Separate the pixel matrix $C^R, C^G$ and $C^B$ of the image $m$.
for $\forall\ c_{ij} \in \{C^R, C^G, C^B\}$.
    $c'_{ij} \leftarrow rgb^*(c_{ij})$.
end for.
Output encrypted image $m'$.
```

Image Sequence Encryption. After the color information is encrypted, the image sequence in RGB channel is converted to 8-bit binary random scrambling. $C' = \{C^{R'}, C^{G'}, C^{B'}\}$ denotes the set of three color channel pixel matrices in RGB of image m', $c'_{(ij)} = (c_{ij}^{R'}, c_{ij}^{G'}, c_{ij}^{B'})$ denotes the $C^{R'}$, $C^{G'}$ and $C^{B'}$ pixels in the position of three color channel pixel matrices in (i, j), where $(i, j) \in ImgSize$ and $ImgSize$ is the number of pixels in each color channel. Then all the pixels are converted to an 8-bit binary, such as $c_{ij}^{R'} = (bit_1, bit_2, ..., bit_8)$. Finally, random scrambling operation is performed for each 8-bit binary sequence.

The random scrambling generated binary $(bit_5, bit_7, ..., bit_2) \leftarrow randperm^*(c_{ij}^{R'})$, $C'(randperm^*) = \{C^{R'}(randperm^*), C^{G'}(randperm^*), C^{B'}(randperm^*)\}$ represents the set of 8-bit binary random scrambled pixel matrices for RGB. The sequence encrypted image is represented as m''. We define Algorithm 2: SequenceEnc.

Algorithm 2 SequenceEnc

```
Input: image $m'$ and the key $randperm^*$.
Output: Sequence encrypted image $m''$.
1. Generate the key $randperm^*$ by Equation(2).
2. Separate the pixel matrix $C^{R'}, C^{G'}$ and $C^{B'}$ of the image $m'$.
3. for $\forall\ c'_{ij} \in \{C^{R'}, C^{G'}, C^{B'}\}$.
4.     $c'_{ij}$=$(bit_1, bit_2, ..., bit_8)$, all pixels are converted to 8-bit binary.
5.     $randperm^*(c'_{ij}) \rightarrow (bit_5, bit_7, ..., bit_2)$, 8-bit binary sequence generated by random scrambling.
6.     $c''_{ij} \leftarrow (bit_5, bit_7, ..., bit_2)$, restore the scrambled 8-bit binary sequence to pixel $c''_{ij}$.
7. end for.
8. Output encrypted image $m''$.
```

Pixel Matrix Scrambling Encryption. To further improve the security of encryption, while encrypting RGB color information and image sequence texture information, Zig-Zag scanning is also used to scan and scramble the two-dimensional pixel matrix of RGB channel. By Zig-Zag scanning, $P''_{(ij)} = (P^{R''}, P^{G''}, P^{B''})$ represents the set of RGB channel pixel vectors of image m'', $P^{R''} = [1, 2, ..., ImgSize]$, $P^{G''} = [1, 2, ..., ImgSize]$, $P^{B''} = [1, 2, ..., ImgSize]$ denotes one-dimensional pixel vectors of three channels, where $ImgSize$ is the number of pixels in each color channel. Then the set of P'' vectors is restored to the pixel matrix $ImgSize$ in the order of RGB channel. The encrypted image e is scanned by Zig-Zag. So we define Algorithm 3: ScramblingEnc, which is described as follows:

Algorithm 3 ScramblingEnc

```
Input: image $m''$ and the key $zig^*$.
Output: encryption image $e$.
1. Generate the key $zig^*$ by Equation(3).
2. Separate the pixel matrix $C^{R''}, C^{G''}$ and $C^{B''}$ of the image $m''$.
3. for $\forall\ c''_{ij} \in \{C^{R''}, C^{G''}, C^{B''}\}$.
4.     $P'' = \{P^{R''}, P^{G''}, P^{G''}\} \leftarrow \{C^{R''}, C^{G''}, C^{B''}\}$, by Zig-Zag scanning three - channel pixel matrix into one-dimensional pixel vector set.
5.     The one-dimensional pixel vector of RGB channel is respectively restored to the $ImgSize$-size image.
6. end for.
7. Output encrypted image $e$.
```

Three encryption manners are described in detail above: RGB channel scrambling, image sequence conversion to 8-bit binary random scrambling, and Zig-Zag scan scrambling of two-dimensional pixel matrix. A complete image encryption Algorithm 4: ImgEnc is defined as follows:

Algorithm 4 ImgEnc

```
Input: image dataset $K$ and the key groups $K = \{(RandNum, \{RGB_r\}_{r=0}^{5}, key_1), key_2\}$.
Output: encryption image dataset $E$.
1. for $\forall m_i \in M$ do.
2.     $m' = ChannelEnc(m_i, rgb^*)$.
3.     $m'' = SequenceEnc(m', randperm^*)$.
4.     $e = ScramblingEnc(m'', zig^*)$.
5. end for.
6. Output encryption image dataset $E$.
```

4.4. Encryption Image Feature Extraction

In this paper, we will use HSV, YUV low-level features and the semantic features extracted by fine-tuning DenseNet model.

[HSV]. For each image, 1000-dim HSV color histogram features are extracted, and the similarity score curve is normalized by l_2. The values of H, S and V are set as 20, 10 and 5 respectively.

[YUV]. For YUV (image color spatial feature, which represents luma and chroma of the image), first convert the color space of the encrypted image from RGB to YUV. Then the encrypted image segmentation into Y, U and V, then statistics each channel respectively the number of pixels within the range of 0-255. Finally, the statistical features of the three channels are added together to form a YUV color histogram feature of 256-dim.

[CNN]. To improve the performance of encrypted image features, we refer to the advantages of deep learning and feature fusion in image retrieval, and apply multi-feature adaptive late-fusion to searchable encryption. Therefore, we use encrypted image dataset to fine-tuning the pre-trained DenseNet model [16], so that the model has better generalization for encrypted images. The fine-tuned DenseNet121 model was used to extract the semantic features of 1024-dim, and l_2 was used to normalize the score curve.

4.5. CBIR Service

The advantage of our method is that feature extraction, similarity calculation, match retrieval and adaptive late-fusion are all completed on the cloud server. In the stage of computing feature similarity, we will adopt the "product rule". It is assumed that K features should be fused, q images are taken as queries, and d images are in the dataset.

The similarity score of each feature $f(i)$ between query image q and dataset d is denoted as $s_{d,q}^{(i)}$, where $i = 1, ..., K$. Suppose the weight of query image q about each feature $f(i)$ is $w_q^{(i)}$ and its sum is 1. The similarity function is defined as:

$$sim(q,d) = \prod_{i=1}^{K} (s_{d,q}^{(i)})^{w_q^{(i)}}, where \sum_{i=1}^{K} w_q^{(i)} = 1 \tag{4}$$

For each fusion feature weight $w_q^{(i)}$, we determined it by the area under the sorting similarity score curve. Before determining the area under the curve, we need to normalize the maximum and minimum values of the similarity score curve.

$$\tilde{s}_q^{(i)} = \frac{s_q^{(i)} - min(s_q^{(i)})}{max(s^{(i)})_q - min(s_q^{(i)})} \tag{5}$$

The maximum and minimum normalized similarity score curves can estimate the effectiveness of eatures, that's, the area is negatively correlated with the effectiveness of features. Thus, the area under each feature $s_{d,q}^{(i)}$ similarity score curve is $(Area)_i$. The weight distribution of adaptive late-fusion is determined by Equation (6).

$$w_q^{(i)} = \frac{\frac{1}{(Area)_i}}{\sum_{k=1}^{K} \frac{1}{(Area)_i}} \tag{6}$$

where $(Area)_i$ denotes the area under the score curve after normalization of feature $f(i)$. The adaptive weight of feature effectiveness calculated by Equation (6) is applied to Equation (4). Finally, the multi-feature adaptive late-fusion was carried out and the Top-k images were returned as the query results.

5. Experimental Results and Performance Evaluation

In this section, we describe the experimental results and performance analysis of our method on public dataset and compare it with other methods. The experiments were conducted on Windows 64, AMD Ryzen5 2600X CPU@3.60GHz, 16GB of RAM and NVIDIA GeForce GTX-2060ti GPUs. First, a brief introduction to the dataset. Holidays consists of 1491 individual holiday images, 500 of which are query images. Average Precision (AP) is used to evaluate the retrieval performance of each query, which is calculated from the area under the Precision recall curve. APs are average for all query images, resulting in average accuracy (mAP).

5.1. The Effectiveness of Image Encryption

Firstly, the performance of our encryption manner is evaluated from the effectiveness of image encryption. Our encryption method is accomplished by RGB channel replacement, image sequence conversion to 8-bit binary random scrambling and Zig-Zag scanning scrambling. The encryption effects are shown in Figure 3.

As shown in Figure 3b, the RGB channel replacement completely changes the color information, but the texture information doesn't change, so the pixel position needs to be encrypted. As can be seen from Figure 3c, the image sequence is converted to 8-bit binary random scrambling, which is blurred compared with the original image, but the contour of the image content can still be distinguished. In Figure 3d, the texture information changes greatly by Zig-Zag scanning scrambling of the pixel position. Finally, as shown in Figure 3e, the image can be well protected by using three encryption manners.

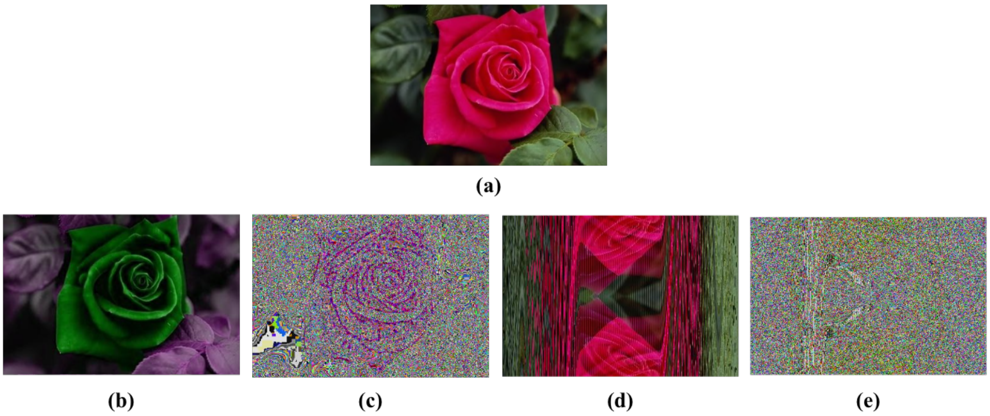

Figure 3. Encryption effect. (**a**) Original image. (**b**) ChannelEnc image. (**c**) SequenceEnc image. (**d**) Zig-ZagEnc image. (**e**) ImgEnc image.

5.2. Individual Feature Retrieval Performance

In the above steps, we extracted HSV, YUV and semantic features. The retrieval accuracy of each feature is shown in Table 1.

Table 1. Retrieval performance of individual feature.

Datasets	IES-CBIR [29]	YUV	HSV	CNN
Holidays, mAP(%)	54.56	18.49	25.89	61.45

It shows that the low-level feature HSV has achieved 25.89% of mAP and YUV has achieved 61.45%. In semantic features, the features extracted from the fine-tunning DenseNet121 model achieved 61.45% of the mAP. CNN achieved the best performance on Holidays.

5.3. Multi-Feature Adaptive Fusion

Multi-feature fusion was performed in the Holidays dataset, and the comparison results are shown in Figure 4. The results show that the performance of our method is better than IES-CBIR. Experiments were conducted on three features fusion: "HSV+YUV", "HSV+CNN" and "HSV+YUV+CNN". More detailed experimental results are shown in Table 2.

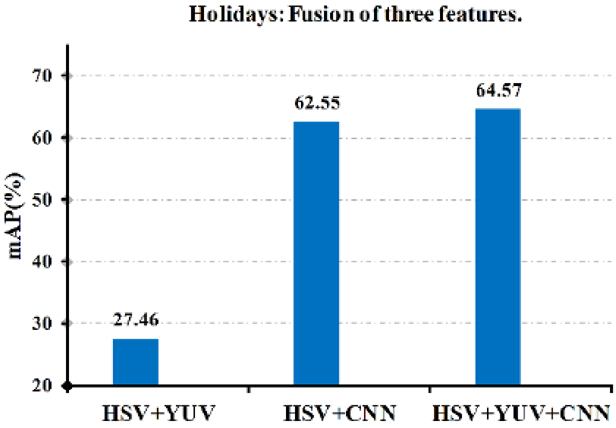

Figure 4. Comparison of the precision by three fusion on Holidays datasets.

Table 2. Comparison with different fusion precision on Holidays.

Feature Combinations	mAP (%)
HSV+YUV	27.46
YUV+CNN	62.55
HSV+CNN	63.57
HSV+YUV+CNN	64.57

5.4. Comparison with Adaptive Weights and Fine-Tuning Weights

To further illustrate the performance of our method, we compare our results with fine-tuning global weights. The fusion results are shown in Table 3. In Table 2, the experimental results show that the performance of our method is better than IES-CBIR [29]. In our approach, we also tried fine-tuning the global weight.

Table 3. Comparison of mAP on Holidays by different parameter tuning.

Feature Combinations	Holidays, mAP(%)	
	Global	Ours
HSV+YUV	26.34	27.46
YUV+CNN	62.27	62.55
HSV+CNN	63.19	63.57
HSV+YUV+CNN	63.83	64.57

For each feature, we assign a global weight $w_q^{(i)}$, Then, we manually tune the $w_q^{(i)}$, where $i = 1, 2, ..., K$. When fusing three features, we use a step of 0.1 for manual tuning. The results are shown in Table 3, our method is superior. Our method can be adaptive to determine the weight, compared with the global tuning resulting in competitive results.

5.5. Security analysis

In the above, we demonstrated that cloud server is a Honest-But-Curious model. In other words, cloud servers can not only correctly perform related protocol tasks, but also attack and analyze the sensitive data information of users. In this section, we will analyze the security of our method from three aspects: image privacy security, feature privacy security and decryption restore image.

Image privacy security. The images in Holidays dataset are all high-resolution and the $ImgSize$ images are encrypted by our method. The computational complexity of replacing any row or column in the RGB channel is $(ImgSize * 6)!$, and the image sequence is converted to 8-bit binary random scrambling, so the total security intensity is $log_2\{(ImgSize * 6)! * 8!\}$ Compared with the IES-CBIR of safety intensity of $3 * log_2\{101!\}$ proposed by Ferreira et al. [29], our method has higher safety.

Features privacy security. The cloud server extracts the 1024-dim semantic features of encrypted images by the fine-tuning DenseNet model. Our encryption method is proved to be secure against the Chipertext-only Attack (COA) model [11,29]. These features are accessible to the cloud server, but it needs computational complexity 1024! to distinguish the correlation among features.

Theorem 1. *In COA model, our scheme can calculate the security of an HBC Probability-Polynomial-Time (PPT) opponent, and the security intensity depends on the image size.*

Proof. *S* simulates an images dataset and the corresponding image index *Index*. Simulator ***S*** have known the number of dataset and the size of each image. However, ***S*** can only fill images with randomly generated pixels. As described in Section 4, the image is composed of color information and texture information. We preserved the color information by *RGB* color channel replacement, the pixel sequence was converted into 8-bit binary random scrambling and *Zig-Zag* scanning scrambling was performed on the pixel position to preserve the texture information. More specifically, the images of size *ImgSize* are encrypted by *RGB* channel replacement (computational complexity is (*ImgSize**6)!) and pixel location scrambling. Then the security intensity of image encryption (*ImgEnc*) can be calculated as:

$$SecImg = log_2\{(ImgSize*6)! * 8\}bit \tag{7}$$

The cloud server has extracted the 1024-dim semantic features of encrypted images by the fine-tuning DenseNet model. The semantic features of the encrypted image can be obtained by the adversary. However, due to the particularity of CNN model, the extracted features are difficult to be deciphered. In addition, even though the simulator ***S*** can analyze semantic features, it needs the computational complexity of 1024! to distinguish the correlation among features. □

Decrypt restore image security. In searchable encryption, PSNR can not only measure the performance of image encryption, but also measure the quality of image decryption restore. Therefore, to verify whether our encryption manner affects the decryption and restoration of images, we calculated the PSNR between images.

In Table 4, we calculated the original image and RGB channel replacement encryption image, sequence conversion to 8-bit binary random scrambling encryption image, Zig-Zag scanning scrambling encryption image, combined encryption image of PSNR on Holidays dataset. Therefore, our encryption manner meets the requirements of image decryption and secure recovery.

Table 4. Comparison of PSNR on Holidays by different encryption manners.

Encryption Method	Holidays, PSNR (dB)
ChannelEnc image	39.06
SequenceEnc image	33.51
ScramblingEnc image	32.71
ImgEnc image	32.39

6. Conclusions

This paper proposes a searchable encrypted image retrieval based on multi-feature adaptive late-fusion in cloud environment. To improve the effectiveness of encryption manner, we adopt RGB channel replacement, image sequence is converted into 8-bit binary random scrambling, and Zig-Zag scanning scrambling is used to protect the color and texture information. In addition, we extract HSV, YUV and semantic features of ciphertext images from in the cloud server, and then conduct adaptive late-fusion of multi-features to improve the retrieval performance. A large number of experiments on public datasets show that our proposed searchable encryption mechanism has better performance than the existing methods.

In our approach, we not only want better retrieval accuracy and encryption performance, but also need to make full use of the computing power of the cloud server. In the experiment, although the implementation of multi-feature adaptive late-fusion in the cloud server reduces the computing overhead of the data owner, the fine-tuning of the pre-training model is also a large computing overhead. Therefore, in future work, we will explore a CNN model suitable for searchable encryption and an accurate weight learning method.

Author Contributions: Supervision, J.Q. and X.X.; Validation, Y.T. and Z.H.; Writing—original draft, W.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant 61772561; in part by the Key Research and Development Plan of Hunan Province under Grant 2018NK2012; in part by the Science Research Projects of Hunan Provincial Education Department under Grant 18A174; in part by the Degree & Postgraduate Education Reform Project of Hunan Province under Grant 2019JGYB154; in part by the Postgraduate Education and Teaching Reform Project of Central South University of Forestry & Technology under Grant 2019JG013. in part by the Natural Science Foundation of Hunan Province under Grant 2020JJ4140; and in part by the Natural Science Foundation of Hunan Province under Grant 2020JJ4141.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Li, H.; Qin, J.; Xiang, X.; Pan, L.; Ma, W.; Xiong, N. An efficient image matching algorithm based on adaptive threshold and RANSAC. *IEEE Access* **2018**, *6*, 66963–66971. [CrossRef]
2. Zheng, L.; Yang, Y.; Tian, Q. SIFT Meets CNN: A decade survey of instance retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 1224–1244. [CrossRef] [PubMed]
3. Guo, J.; Prasetyo, H.; Wang, N. Effective image retrieval system using dot-diffused block truncation coding features. *IEEE Trans. Multimed.* **2015**, *17*, 1576–1590. [CrossRef]
4. Xia, Z.; Zhang, L.; Liu, D. Attribute-based access control scheme with efficient revocation in cloud computing. *China Commun.* **2016**, *13*, 92–99. [CrossRef]
5. Barona, R.; Anita, E. A survey on data breach challenges in cloud computing security: Issues and Threats. In Proceedings of the International Conference on Circuit, Power and Computing Technologies, Kollam, India, 20–21 April 2017; pp. 1–8.
6. Mahajan, P.; Setty, S.; Lee, S.; Clement, A.; Alvisi, L.; Dahlin, M.; Walfish, M. Depot: Cloud storage with minimal trust. *ACM Trans. Comput. Syst.* **2011**, *29*, 1–38. [CrossRef]
7. Hsu, C.; Lu, C.; Pei, S. Image Feature Extraction in Encrypted Domain With Privacy-Preserving SIFT. *IEEE Trans. Image Process.* **2012**, *21*, 4593–4607.
8. Xia, Z.; Zhu, Y.; Sun, X.; Qin, Z.; Ren, K. Towards Privacy-preserving Content-based Image Retrieval in Cloud Computing. *IEEE Trans. Cloud Comput.* **2018**, *6*, 276–286. [CrossRef]
9. Bai, Y.; Zhuo, L.; Cheng, B. Surf feature extraction in encrypted domain. In Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), Chengdu, China, 14–18 July 2014; pp. 1–6.
10. Qin, J.; Li, H.; Xiang, X.; Tan, Y.; Pan, W.; Ma, W.; Xiong, N. An Encrypted Image Retrieval Method Based on Harris Corner Optimization and LSH in Cloud Computing. *IEEE Access* **2019**, *7*, 24626–24633. [CrossRef]
11. Xia, Z.; Xiong, N.; Vasilakos, V.; Sun, X. EPCBIR: An efficient and privacy-preserving content-based image retrieval scheme in cloud computing. *Inf. Sci.* **2017**, *387*, 195–204. [CrossRef]

12. Wan, J.; Wang, D.; Hoi, C.; Wu, P.; Zhu, J.; Zhang, Y.; Li, Ji. Deep learning for content-based image retrieval: A comprehensive study. In Proceedings of the ACM International Conference on Multimedia, Orlando, FL, USA, 10–13 November 2014; pp. 157–166.
13. Hinton, G.; Salakhutdinov, R. Reducing the dimensionality of data with neural networks. *Science* **2016**, *313*, 504–507. [CrossRef]
14. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–10.
15. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
16. Huang, G.; Liu, Z.; Maaten, L.; Weinberger, K. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269.
17. Qin, J.; Pan, W.; Xiang, X.; Tan, Y.; Hou, G. A biological image classification method based on improved CNN. *Ecol. Inform.* **2020**, *28*, 1–8. [CrossRef]
18. Douze, M.; Ramisa, A.; Schmid, C. Combining attributes and Fisher vectors for efficient image retrieval. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 20–25 June 2011; pp. 1–10.
19. Liu, P.; Guo, J.; Wu, C. Fusion of Deep Learning and Compressed Domain features for Content Based Image Retrieval. *IEEE Trans. Image Process.* **2017**, *26*, 5706–5717. [CrossRef]
20. Zheng, L.; Wang, S.; Tian, Q. Coupled Binary Embedding for Large-Scale Image Retrieval. *IEEE Trans. Image Process.* **2014**, *23*, 3368–3380. [CrossRef] [PubMed]
21. Zheng, L.; Wang, S.; Tian, L.; He, F.; Liu, Z.; Tian Q. Query-adaptive late fusion for image search and person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1741–1750.
22. Weng, L.; Amsaleg, L.; Morton, A.; Marchand-Maillet, S. A privacy-preserving framework for large-scale content-based information retrieval. *IEEE Trans. Inf. Forensics Secur.* **2014**, *10*, 152–167. [CrossRef]
23. Lu, W.; Swaminathan, A.; Varna, A. Enabling search over encrypted multimedia databases. *Int. Soc. Opt. Eng.* **2009**, *7254*, 725418.
24. Lu, W.; Varna, A.; Swaminathan, A.; Wu, M. Secure image retrieval through feature protection. In Proceedings of the IEEE International Conference on Acoustics, Taipei, Taiwan, 19–24 April 2009; pp. 1533–1536.
25. Weng, L.; Amsaleg, L.; Furon, T. Privacy-Preserving Outsourced Media Search. *IEEE Trans. Knowl. Data Eng.* **2016**, *28*, 2738–2751. [CrossRef]
26. Xia, Z.; Wang, X.; Zhang, L.; Qin, Z.; Sun, X.; Ren, K. A privacy-preserving and copy-deterrence content-based image retrieval scheme in cloud computing. *IEEE Trans. Inf. Forensics Secur.* **2016**, *11*, 2594–2608. [CrossRef]
27. Tan, Y.; Qin, J.; Xiang, X.; Ma, W.; Pan, W.; Xiong, N. A Robust Watermarking Scheme in YCbCr Color Space Based on Channel Coding. *IEEE Access.* **2019**, *7*, 25026–25036. [CrossRef]
28. Luo, Y.; Qin, J.; Xiang, X.; Tan, Y.; Liu, Q.; Xiang, L. Coverless real-time image information hiding based on image block matching and Dense Convolutional Network. *J. Real-Time Image Process.* **2020**, *17*, 125–135. [CrossRef]
29. Ferreira, B.; Rodrigues, J.; Leitao, J.; Domingos, H. Practical privacy-preserving content-based retrieval in cloud image repositories. *IEEE Trans. Cloud Comput.* **2017**, *13*, 1–14. [CrossRef]
30. Zhang, J.; Marszalek, M.; Lazebnik, S.; Schmid, C. Local features and kernels for classification of texture and object categories: A comprehensive study. *Int. J. Comput. Vis.* **2007**, *73*, 213–238. [CrossRef]
31. Zhang, S.; Yang, M.; Cour, T.; Yu, K.; Metaxas, D. Query Specific Rank Fusion for Image Retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 803–815. [CrossRef] [PubMed]

32. Liu, Z.; Wang, S.; Zheng, L.; Tian, Q. Robust Image-graph: Rank-level feature fusion for image search. *IEEE Trans. Image Process.* **2017**, *26*, 3128–3141. [CrossRef] [PubMed]
33. Zhang, S.; Yang, M.; Wang, X.; Lin, Y.; Tian, Q. Semantic-aware co-indexing for image retrieval. *IEEE Int. Conf. Comput. Vis.* **2013**, *37*, 1673–1680.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).