Expert Systems With Applications 203 (2022) 117508

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A privacy-preserving content-based image retrieval method based on deep learning in cloud computing

Wentao Ma [a], Tongqing Zhou [a], Jiaohua Qin [b,*], Xuyu Xiang [b], Yun Tan [b], Zhiping Cai [a,*]

[a] College of Computer, National University of Defense Technology, Changsha, Hunan, 410073, China
[b] College of Computer Science and Information Technology, Central South University of Forestry & Technology, Changsha, Hunan, 410000, China

ARTICLE INFO

Keywords:
Image retrieval
Privacy-preserving
Deep convolutional network
Edge computing
CBIR

ABSTRACT

Privacy-preserving Content-Based Image Retrieval (CBIR) method is a promising technology to achieve data confidentiality and searchability in cloud-assisted multimedia (i.e., image or video) data environment. However, inappropriate feature-preserving mechanisms and inefficient ciphertext descriptors resulted in lower performance than expected. Therefore, how to design encryption techniques with high security and how to extract effective features from ciphertext images still hinder privacy-preserving CBIR. For this goal, we propose a privacy-preserving image retrieval based on deep convolutional network features. First, a novel hybrid encryption technique is designed to encrypt images and an improved DenseNet model is fine-tuned by using the encrypted images to construct a feature extractor. The encrypted images and fine-tuning feature extractor are then uploaded to cloud server. Meanwhile, secure CBIR service is executed in the cloud server. We conduct experiments on two public benchmark datasets for performance evaluation in terms of mAP and accuracy. As demonstrated in the experimental results, the proposed method can achieve superior result compared with the existing methods, improving the performance on the two metrics by relatively 1.9% and 10%, respectively. Furthermore, the computational cost and parameters of depthwise separable convolution adopted by the improved DenseNet model are 8 to 9 times smaller than that of standard convolutions of the original DenseNet at only a small reduction in accuracy.

1. Introduction

With the rapid popularization of mobile intelligent terminal devices, more and more multimedia data (e.g., images or videos) are produced. As a common practice, individuals or enterprises outsource their image data to cloud server in order to get rid of cumbersome storage and management. In this context, CBIR methods are widely used for data owners to attain retrieval accuracy and efficiency (Gkelios et al., 2021; Zheng et al., 2018). Moreover, authorized query users can obtain images from the cloud server without maintaining communication with the data owners (Anju & Shreelekshmi, 2022; Gu et al., 2020; Song et al., 2022; Xia et al., 2019, 2019, 2020, 2016, 2015). It is worth noting that while we enjoy the convenience of cloud computing, from another perspective, data owners do not fully trust cloud server (Barona & Anita, 2017; Li et al., 2020). In particular, finding a secure and efficient mechanism to manage large-scale images remains one of the major challenges facing CBIR technology today. To protect privacy, traditional privacy-preserving CBIR, namely, feature-encryption CBIR based schemes, proposes to extract features from plaintext images and encrypt the features using the designed encryption techniques. The image owners then upload both encrypted features and encrypted images to the cloud server for storage and future management (Qin et al., 2020; Xia et al., 2016, 2015). Yet, these methods will bring a lot of extra computing burdens to "weak" image data owners and obviously cannot scale well to manage CBIR service of large image libraries. As a result, many researchers have proposed that the computing burdens should be left to the cloud server for local computation efficiency (Anju & Shreelekshmi, 2022; Ferreira et al., 2019; Gu et al., 2020; Li et al., 2020; Song et al., 2022; Tang et al., 2021; Wang et al., 2020; Xia et al., 2019, 2019, 2020, 2017), namely, image-encryption CBIR based schemes. Among them, Anju and Shreelekshmi (2022), Xia et al. (2017) adopt MPEG-7 descriptors for image feature representation, while the proposed IES-CBIR in Ferreira et al. (2019) adopts Hue-Saturation-Value (HSV) descriptor in image feature characterization. However, we point out that most low-level features only focus on extracting local key information, which can hardly discover the semantic correlations required by the privacy-preserving CBIR processes.

* Corresponding authors.
E-mail addresses: wtma@nudt.edu.cn (W. Ma), zhoutongqing@nudt.edu.cn (T. Zhou), qinjiaohua@163.com (J. Qin), xyuxiang@163.com (X. Xiang), tantanyun@hotmail.com (Y. Tan), zpcai@nudt.edu.cn (Z. Cai).

https://doi.org/10.1016/j.eswa.2022.117508
Received 17 November 2021; Received in revised form 19 March 2022; Accepted 3 May 2022
Available online 13 May 2022

To tackle this issue, with the advent of deep convolutional neural network (CNN), CBIR based on low-level features has been gradually replaced by high-level semantic features (Gkelios et al., 2021; Hussain et al., 2021; Öztürk, 2020; Pan et al., 2021) and fused features (Ma et al., 2020; Qin et al., 2020). In general, the CNN model can achieve better results in image classification and retrieval (Anju & Shreelekshmi, 2022; Huang et al., 2019; Ma et al., 2019; Pan et al., 2021), because it can infer human perception by different convolution templates, pooling and other operations. Among the typical CNN models (e.g., AlexNet, GoogLeNet, VGG and ResNet), DenseNet is featured with extremely high feature utilization rate, stronger feature expression ability, fewer parameters and less computation complexity (Huang et al., 2019; Pan et al., 2021). Yet, endowing privacy preservation property on complex semantic features is still an open problem, which needs dedicated design towards effective and security CBIR service.

In view of the above analysis, we propose a privacy-preserving image retrieval based on CNN features, which accommodate information sensitivity in the data preprocessing, storage and searchability of cloud computing. Specifically, we design a hybrid encryption technique, including ChannelEnc, SequenceEnc, and PositionEnc, that can protect both color and texture information of images to prevent untrusted cloud servers from disclosing sensitive information of the data. Meanwhile, the cloud server can extract semantic features from encrypted images by utilizing an improved DenseNet model and then perform feature similarity matching to return all similar retrieval results. The main contributions of our work are in the following three aspects:

- Better hybrid encryption technique. We propose a privacy-preserving mechanism with the color information and texture information preserved jointly. Specifically, the color and texture features of the image are preserved by random substitution of color channels and random scrambling of pixel bit plane sequences, which guarantees stronger security than general encryption technique.
- High retrieval performance. We propose an improved DenseNet model, which replaces a part of the structure of the DenseNet network with an inverted residual block and then adopts this model to extract efficient semantic features from image ciphertexts. Specifically, the encrypted images are leveraged to fine-tune the improved DenseNet model and the semantic features of ciphertext images are extracted on cloud server. Secure CBIR service are then provided leveraging the computation advantages in the cloud environments.
- More competitive. Extensive experimental results on two public benchmark datasets and formal security analysis show that our approach outperforms typical existing methods by a clear margin.

In this section, we discuss basic solutions of CBIR in cloud environment. Specifically, we propose a privacy-preserving CBIR framework based on CNN features and attempt to tackle the problems in the above scheme. After a brief review of the related work in Section 2, we introduce the technical outline and system model overview in Section 3. Section 4 introduces secure data storage and search. In Section 5, the experimental results and formal security analysis of our approach will be given, and the conclusion of our work will be summarized in Section 6.

2. Related work

In the cloud-assisted multimedia data environment, CBIR service has been extensively explored in feature extraction, feature matching, feature fusion, index construction, etc. In this section, we mainly focus on two aspects, namely, feature-encryption CBIR based schemes and image-encryption CBIR based schemes. The kernel difference is that feature extraction is performed before or after the encrypted images are uploaded and we detailedly discuss these schemes in the following.

2.1. Feature-encryption CBIR based schemes

To retrieve similar images quickly from a large number of images, some promising CBIR techniques have been proposed (Amato et al., 2020; Gkelios et al., 2021; Zheng et al., 2018). However, the images always contain rich sensitive information and directly uploading unencrypted images to the cloud is unsafe. For efficient ciphertext image retrieval services and data privacy-preserving requirements, sensitive data must be encrypted before being outsourced to cloud server. Privacy-preserving CBIR allows data owners to upload encrypted images data to the cloud environment while completing secure privacy-preserving image retrieval services in the cloud server (Li et al., 2020; Lu et al., 2009a, 2009b, 2014; Weng et al., 2016; Xia et al., 2016, 2015). To our knowledge, Lu et al. (2009a) is the first to propose privacy-preserving CBIR based on feature encryption. This method adopts a Bag-of-Words (BoW) model to represent the feature information of images while protecting the privacy of visual words through min-hash and order-preserving methods. To improve the security of feature encryption, in their next work Lu et al. (2009b) adopts three techniques to protect image features, including bitplane randomizations, random unary encoding and random projections. Although this method can effectively preserve the privacy of images, it will lead to a low retrieval accuracy than expected. The above two methods are compared with the homomorphic encryption in another work by Lu et al. (2014), which shows that the proposed method requires less computing, memory and communication resources.

Extensive researches have shown that efficient image feature descriptors can improve retrieval performance. Xia et al. (2015) proposes a privacy-preserving CBIR based on Scale-Invariant Feature Transformation (SIFT) and Earth Mover's Distance (EMD). SIFT and EMD are adopted to represent the features and similarity measures of images, respectively. The calculation of the EMD is needed to construct and solve the linear program problem. A linear transformation had been performed on this problem to protect the sensitive parameters. Weng et al. (2016) presents a privacy-preserving multimedia retrieval method, which adopts robust hashing to encrypt image feature information. However, this approach does not presume whether owner data is accessible to the query users or not, which is somewhat debatable for the privacy guarantee in a more strict scenario (i.e., the database is confidential). Xia et al. (2016) investigates a privacy-preserving CBIR with multi-MPEG descriptor feature information. At the same time, to balance security and efficiency, this method adopts k-means to encrypt information and uses Locality Sensitive Hashing (LSH) to improve efficiency. Cheng et al. (2019) proposes a surveillance video privacy-preserving based on person Re-identification (Re-ID). This method integrates the CNN model and represents high precision and efficient image features by binary form. However, this method ignores the optimal relationship between the dimension of feature index and the number of cloud server, which greatly reduces the users experience, and the efficiency of secure person Re-ID in a real-world environment is not high. Qin et al. (2019) extracts Speeded-Up Robust Feature (SURF) from images and designed chaotic encryption to preserve features ans the LSH is also utilized to improve retrieval efficiencies. Li et al. (2020) provides a similarity search for encrypted images in secure cloud computing. This method leverages feature descriptors extracted by the CNN model and K-means clustering based on Affinity Propagation clustering respectively to improve search accuracy and efficiency. Meanwhile, a limited key-leakage k-Nearest Neighbor is designed to protect image privacy.

2.2. Image-encryption CBIR based schemes

For CBIR service outsourced to cloud server, we need to strictly protect data privacy in cloud environment. In above analysis, the feature encryption CBIR has achieved gratifying performance and this approach alleviates problem of insufficient storage space on client-side.

However, operations such as feature extraction and feature encryption will impose significant computational costs on data owners. To reduce the burdens, researchers propose several image-encryption based privacy-preserving CBIR (Anju & Shreelekshmi, 2022; Ferreira et al., 2019; Gu et al., 2020; Ma et al., 2020; Qin et al., 2020, 2020, 2019; Song et al., 2022; Tang et al., 2021; Wang et al., 2020; Xia et al., 2019, 2019, 2020, 2017). In these methods, the image owners only need to encrypt images, while other operations such as feature extraction, index construction and CBIR service are outsourced to cloud server.

Xia et al. (2017) investigates a scheme that supports CBIR over encrypted images without revealing sensitive information to cloud server. In this method, secure k-nearest neighbors is adopted to protect image feature representation vector and position-sensitive hashing is leveraged to improve the search efficiency. Ferreira et al. (2019) proposes the IES-CBIR, which not only supports encrypted images storage and CBIR service but also can counter Honest-But-Curious (HBC) cloud server. In cloud environment, HSV feature descriptors of the encrypted images are extracted and similarity matching search is performed by Hamming distance between the features representations. IES-CBIR can significantly reduce the amount of calculation for image owners, with cloud server bearing more computing costs. Xia et al. (2019) extracts global Local-Binary-Pattern (LBP) histograms from encrypted images. The method achieves better accuracy in privacy-preserving face recognition but poor performance in terms of image retrieval. Moreover, Xia et al. (2019) presents a scheme that extracted features from both spatial and Discrete Cosine Transform (DCT) domains, namely, AC-coefficients and color histograms (ACCH). Specifically, this method calculates and connects the AC-coefficients histogram of the encrypted Y component and two-color histograms of the U, V components as feature representation vector. Wang et al. (2020) designs an outsourced privacy-preserving CBIR based on BoW model, AES encryption, block permutation and random mapping. This scheme can greatly reduce the computing burdens of the data owners by outsourcing feature extraction, index construction and search operations to cloud server. To tackle the problem of leaking values and orders of similarity scores, Song et al. (2022) investigates an efficient threshold-based encrypted image retrieval method that leverages CNN model to extract image feature representations and allows users to specify thresholds for embedding tokens. Tang et al. (2021) proposes an outsourcing secure JPEG image retrieval scheme. This format-compatible method can achieve perfect retrieval performance and get strong security without causing file size expansion. Anju and Shreelekshmi (2022) provides a new and faster secure CBIR method, which clusters global image features, namely, MPEG-7 visual descriptors and adopts Asymmetric Scalar-product Preserving (ASPE) to ensure privacy-preserving ranked search and secure index updation. Although many efforts have been spent in privacy-preserving CBIR, the challenging problem of secure CBIR, namely, ‘‘semantic gap for image-encryption feature’’, still exists between low-level and high-level features. Ma et al. (2020) and Qin et al. (2020) provide a searchable encrypted image retrieval method based on multi-feature adaptive late-fusion in cloud environment. For the former, this method extracts CNN features, HSV features and YUV features of encrypted images, then fuses them in an adaptive manner to improve retrieval accuracy. For the latter, it extracts low-level features (e.g., BoW, Edge Histogram Descriptor (EHD)) and high-level semantic features of encrypted images, then complete secure retrieval of feature fusion. While these two methods adopt K-nearest neighbor and Logistic encryption technique to protect the privacy of fusion features and images, respectively.

As discussed above, to improve retrieval accuracy and encryption technique security, our work proposes a privacy-preserving image retrieval based on CNN features. Unlike existing methods (Ferreira et al., 2019; Tang et al., 2021; Xia et al., 2020, 2017), e.g., ES-CBIR, EPCBIR and OPPR, which only leverages a single encryption technique or low-level features of image encryption. We first adopt hybrid encryption technique to encrypt images then fine-tune the improved DenseNet model using encrypted images on the edge computing platform to build feature extractors. Meanwhile, the encrypted images and deep convolutional feature extractor are uploaded to cloud server. Finally, privacy-preserving matching search is completed in cloud server to greatly reduce the computing burdens of data owners.

3. Problem formulation

We propose a framework of privacy-preserving image retrieval based on CNN features, which supports privacy-preserving, outsourced storage and CBIR service. As shown in Fig. 1, the system model consists of three modules: the edge computing module; the cloud server module and the query user module.

3.1. Methods terminology

Image owner: The image owner encrypts $M = \{m_i\}_{i=1}^{n}$ composed of n images and the encrypted images are denoted as $E = \{e_i\}_{i=1}^{n}$. Where m_i and e_i respectively represent the ith image in dataset M and encrypted images dataset E, the index remains as $INDEX = \{Index_i\}_{i=1}^{n}$.

Cloud server: The encrypted images dataset $E = \{e_i\}_{i=1}^{n}$ is stored on cloud server by the data owners, which tackles limitation of insufficient offline storage space. Meanwhile, the semantic feature representation of encrypted images is extracted by the improved DenseNet extractor with fine-tuning. In our work, the storage of encrypted images and CBIR service are all completed in cloud server.

Image owner: For privacy-preserving of the query image, the users are authorized to generate a trapdoor TD by the retrieving mechanism before uploading the query image. Besides, the inquiring users need to send an authorization request for the ‘decryption information’ key to the data owners.

3.2. Adversary model

Similar to Searchable Symmetric Encryption (SSE) (Ferreira et al., 2019; Gu et al., 2020; Lu et al., 2009a; Ma et al., 2020; Qin et al., 2020, 2019; Song et al., 2022; Tang et al., 2021; Wang et al., 2020; Xia et al., 2019, 2019, 2020, 2017), the goal of our work is not to leak data information about the data owners and the query users. It is important to note that cloud server is assumed to be HBC, which will correctly enforce the relevant standards and will also analyze/track sensitive data information under the protocol settings. It is generally assumed that neither the data owners nor the authorized query users will disclose any information to third parties in our work. Based on the above assumptions, the encrypted images dataset e_i and e_j should be of the same class, while the query images e_{iq} and e_{jq} will have a high similarity to the two types of encrypted images dataset. For cloud server, the two query images are considered to be similar. This type of disclosure is common in data managed by cloud server, which have access to all information in RAM and analyze content of interest based on search traces of the query users. However, this set of disclosures is almost negligible. Furthermore, it would impose unnecessary computational and communication costs on system to consider the privacy disclosure of this form in data matching retrieval process. Therefore, our approach ignores information leakage from this mechanism.

4. The proposed method

In this section, we will introduce the details of our method from four parts: key generation, image encryption, improved DenseNet model fine-tuning framework and secure CBIR service.

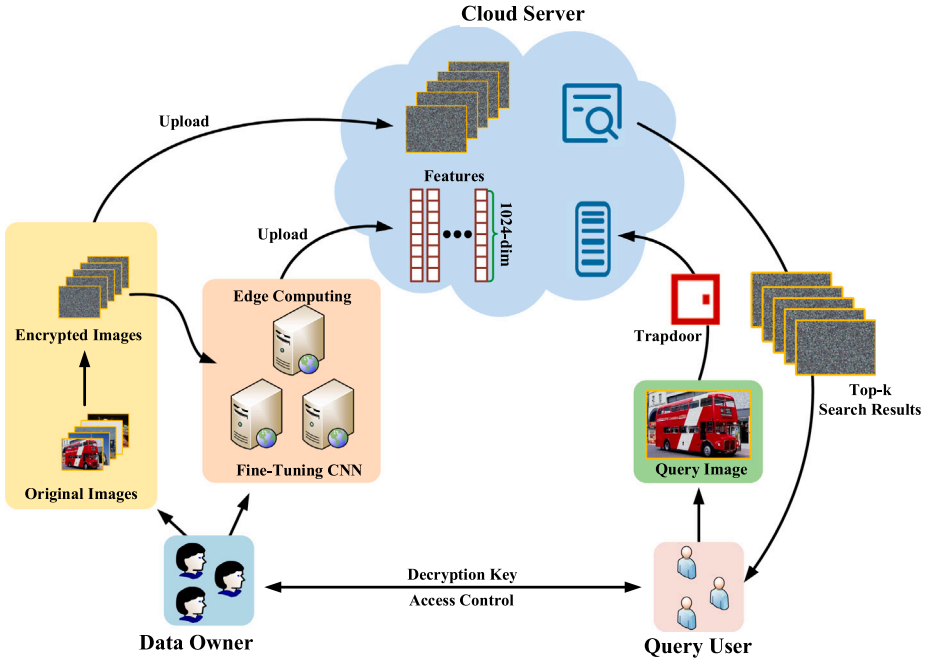


Fig. 1. Overview of our privacy-preserving CBIR framework. The system framework is composed of three modules: edge computing module for fine-tuning improved DenseNet model, cloud server module for complete CBIR service and query user module.

4.1. Key generation

In our work, a hybrid encryption technique is proposed: the color information of images is preserved by randomly replacing RGB color channels, while the texture information is preserved by conversion of pixel sequence and position. Hence, the key groups can be denoted as: $K = \{(RandNum, \{RGB_r\}_{r=0}^{5}, key_1), key_2\}$. And sequence of each channel has $ImgSize$ pixels, the corresponding pixels of spatial position between channels are replaced by random. The $RandNum$ and $\{RGB_r\}_{r=0}^{5}$ are leveraged to preserve the image color information of parameter variables. Color channel encryption function can be defined as:

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

The privacy-preserving of image texture information is combined with pixel bit-plane conversion to binary random scrambling and $Zig-Zag$ scan scrambling. Each pixel sequence by key_1 is converted to binary and random scrambling. According to the description, key_1 can be written as:

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

In essence, $Zig-Zag$ encryption is to scramble the pixel matrix of image, which is defined as:

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

Eqs. (1)–(3) are the encryption techniques defined in our work to constitute the hybrid encryption. Combining these three encryption techniques to complete image encryption, which can greatly improve the security compared with other existing encryption techniques.

4.2. Image encryption

In Section 4.1, we briefly present the three steps of image encryption. In this section, we will go into more detail about the three encryption techniques and define an integrated hybrid encryption algorithm.

ChannelEnc. There are $\{RGB_r\}_{r=0}^{5}$ scrambling ways for RGB channel in key generation. C represents the RGB channels of original image, while the pixel matrices of three channels are C^R, C^G and C^B respectively. Specifically, $c_{ij} = (c_{ij}^R, c_{ij}^G, c_{ij}^B)$ denotes pixel value at the pixel matrix (i, j) for each color channel, where $(i, j) \in ImgSize$, and the number of pixels for the color channel is represented by $ImgSize$. The encrypted images by random replacement of color channel is represented by C'. Hence, the pixel matrices of encrypted channel are $C^{R'}$, $C^{G'}$ and $C^{B'}$ respectively, while the pixels at matrix $c'_{ij} = (c_{ij}^{R'}, c_{ij}^{G'}, c_{ij}^{B'})$ are represented by (i, j). m' indicates image of the encrypted color channel.

The image color privacy-preserving effect by RGB channel encryption is shown in Fig. 2, where the image color information can be preserved in this technique. The quantitative details will be explained in Section 5.

SequenceEnc. The pixel values of image are converted to binary random scrambling encryption. $C' = \{C^{R'}, C^{G'}, C^{B'}\}$ denotes channel pixel matrix set of the color encrypted image m', and $c'_{ij} = (c_{ij}^{R'}, c_{ij}^{G'}, c_{ij}^{B'})$ represents the (i, j) position of $C^{R'}$, $C^{G'}$ and $C^{B'}$ pixel matrix respectively, where $(i, j) \in ImgSize$. Therefore, $c_{ij}^{R'} = (bit_1, bit_2, \ldots, bit_8)$, $c_{ij}^{G'} = (bit_1, bit_2, \ldots, bit_8)$ and $c_{ij}^{B'} = (bit_1, bit_2, \ldots, bit_8)$ can be adopted to denote that the image three-channel sequence is converted to a binary sequence. While $(bit_3, bit_5, \ldots, bit_8) \leftarrow randperm^*(c_{ij}^{R'})$, $(bit_1, bit_2, \ldots, bit_8) \leftarrow randperm^*(c_{ij}^{G'})$ and $(bit_2, bit_5, \ldots, bit_7) \leftarrow randperm^*(c_{ij}^{B'})$ represent randomly generated 8-bit binary sequence. $C'(randperm^*) = \{C^{R'}(randperm^*), C^{G'}(randperm^*), C^{B'}(randperm^*)\}$ represents the conversion of a sequence to a set of 8-bit binary randomly scrambled pixel matrices. The image encrypted by sequence is denoted as m''.

The image texture privacy-preserving effect by sequence encryption is shown in Fig. 3, where the image texture information can be preserved in this technique. The quantitative details will be explained in Section 5.

PositionEnc. We also employ $Zig-Zag$ scanning to scramble and encrypt the pixel matrix. $P'' = \{P^{R''}, P^{G''}, P^{B''}\}$ represents a set of pixels for color channel of the m'' by $Zig-Zag$ scanning. Therefore, $P^{R''} = [1, 2, \ldots, ImgSize]$, $P^{G''} = [1, 2, \ldots, ImgSize]$ and $P^{B''} = [1, 2, \ldots, ImgSize]$ respectively represent the vector of color channel pixels, where $ImgSize$ represents the number in each channel of the image m''. e represents an image that is encrypted in three techniques.

(a) Original image (b) ChannelEnc image

Fig. 2. Example image from the Corel10k dataset. The (a) represents the original image and the (b) denotes the qualitative visualization result of RGB color channel encrypted image.

In light of the above analysis, three encryption techniques are described: ChannelEnc, SequenceEnc and PositionEnc. Now we will define a complete image encryption HybridEnc, as shown in **Algorithm** 1.

The image texture privacy-preserving effect by pixel position encryption is shown in Fig. 4, where the image texture information can be preserved in this technique. The quantitative details will be explained in Section 5.

Algorithm 1 : HybridEnc

Input: image dataset M and the key groups $K = \{(RandNum, \{RGB_r\}_{r=0}^{5}, key_1), key_2\}$.
Output: encrypted images dataset E.

1: for $\forall m_i \in M$ do.
2: $m' = ChannelEnc(\forall c_{ij} \in \{C^R, C^G, C^B\}, c'_{ij} \leftarrow rgb^*(c_{ij}))$.
3: $m'' = SequenceEnc(\forall c'_{ij} \in \{C^{R'}, C^{G'}, C^{B'}\}, c'_{ij} = (bit_1, bit_2, ..., bit_8), randperm^*(c'_{ij}) \rightarrow (bit_5, bit_7, ..., bit_2), c''_{ij} \leftarrow (bit_5, bit_7, ..., bit_2)$.
4: $e = Zig\text{-}ZagEnc(\forall\{C^{R''}, C^{G''}, C^{B''}\}, P'' = \{P^{R''}, P^{G''}, P^{B''}\} \leftarrow \{C^{R''}, C^{G''}, C^{B''}\})$, by $Zig\text{-}Zag$ scanning the pixel matrix into a 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 images dataset E.

4.3. Improved DenseNet model

As investigated in Huang et al. (2019), Ma et al. (2020) and Pan et al. (2021), DenseNet is a typical CNN architecture with some dense block structures, namely, each layer is connected with other layers and each layer is a map of the previous layer's features as input, which has achieved promising performance in image retrieval tasks. Although these dense block structures enhance feature representation leading to superior results, they also add many parameters and huge computational overhead to the model. To reduce the parameters and computational cost, we propose an improved DenseNet structure that replaces a part of DenseNet model by introducing a lightweight module. As shown in Fig. 5, our CNN architecture consists of two parts, including image processing and improved DenseNet model. For image processing, the proposed hybrid encryption technique is used to encrypt images as input of the improved DenseNet model. The detailed encryption steps are shown in **Algorithm** 1. In terms of improved DenseNet model, inspired by Pan et al. (2021) and Sandler et al. (2018), we introduced a lightweight module, namely, inverted residual block. Specifically, different from original DenseNet block, the inverted residual block consists of depthwise separable convolution.

Assuming that the depthwise separable convolution takes an $h_i \times w_i \times d_i$ input feature tensor L_i, and applies convolutional kernel $K \in \mathcal{R}^{k\times k\times d_i\times d_j}$ to produce an $h_i \times w_i \times d_i$ output tensor L_j. The calculation cost V_{cost} of depthwise separable convolution layers is:

$$V_{cost} = k \times k \times h_i \times w_i \times d_i + h_i \times w_i \times d_i \times d_j \tag{4}$$

While the calculation cost U_{cost} of standard convolution layers is:

$$U_{cost} = k \times k \times h_i \times w_i \times d_i \times d_j \tag{5}$$

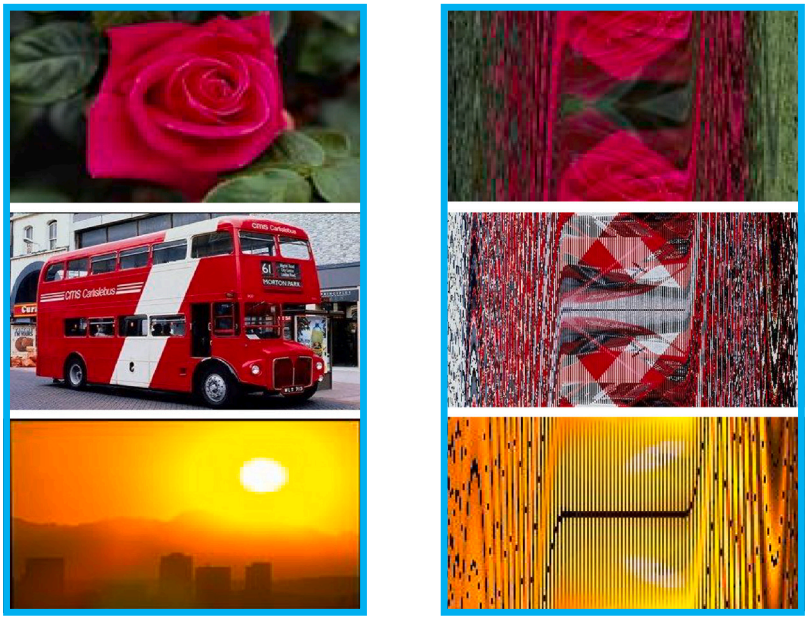
Depthwise separable convolutions are a drop-in replacement for standard convolutional layers. Intuitively speaking, the calculation cost ratio C_{ratio} of depthwise separable convolution V_{cost} to standard convolution U_{cost} is shown in following equation:

$$C_{ratio} = \frac{V_{cost}}{U_{cost}} = \frac{1}{d_j} + \frac{1}{k^2} = \frac{d_j + k^2}{d_j k^2} \tag{6}$$

Furthermore, the parameters of depthwise separable convolution and standard convolution are $V_{paras} = k \times k \times h_i + 1 \times 1 \times h_i \times d_j$ and $U_{paras} = k\times k\times h_i\times d_j$, respectively. As a result, the parameters cost ratio of V_{paras} to U_{paras} is $P_{ratio} = (d_j+k^2)/d_j k^2$. Compared with the traditional convolutional layer, the depthwise separable convolution effectively reduces computation and parameters by almost a factor of k^2. Following the Pan et al. (2021) and Sandler et al. (2018), in our experiments we adopt $k = 3$, namely, 3×3 depthwise separable convolution so the computational cost and parameters are $8 \sim 9$ times smaller than standard convolution with only a small reduction in accuracy.

(a) Original image (b) SequenceEnc image

Fig. 3. Example image from the Corel10k dataset. The (a) represents the original image and the (b) denotes the qualitative visualization result of sequence encrypted image.



(a) Original image (b) PositionEnc image

Fig. 4. Example image from the Corel10k dataset. The (a) represents the original image and the (b) denotes the qualitative visualization result of pixel position encrypted image.

4.4. Retrieval service

The main task of our work is to reduce the computational burdens of data owners, that is, extraction of encrypted images semantic feature representation and image similarity matching retrieval are completed in cloud environment. After trapdoor f_q is generated, the cloud server will conduct complete CBIR service. Then calculate the Euclidean distance between the query image feature f_q and all the encrypted image features stored in cloud server, which can be defined as:

$$sim(e_q, E) = \|f_q - f_i\|^2, \ where \ \ i = 1, 2, \ldots, n \tag{7}$$

Where e_q and f_q respectively denote the encrypted query image and its feature representation, f_i is semantic feature representation of the ith image in the encrypted image dataset E, and n is the number of all encrypted images stored in cloud environment. According to Eq. (7), the corresponding Top-k images are returned.

5. Experimental results and analysis

In this section, we will experiment on two public benchmark datasets (e.g., Holidays, Corel10k) and compare the results with other existing privacy-preserving CBIR methods. This section includes seven

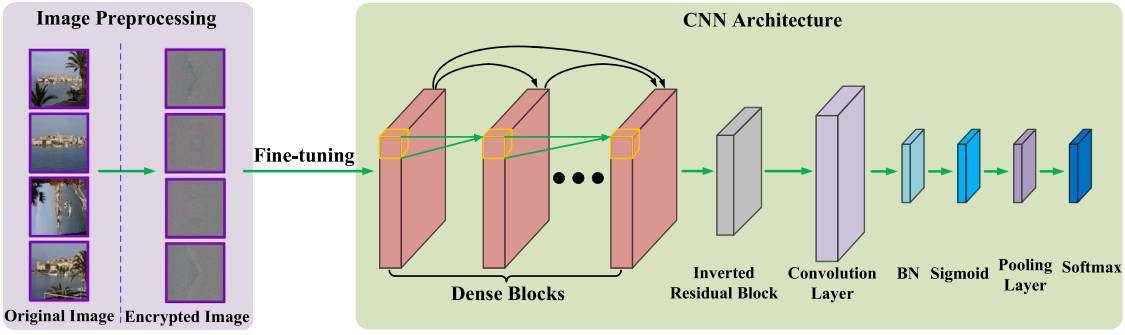


Fig. 5. The architecture of our model consists of two parts, including image processing and an improved DenseNet model.

parts: 5.1. implement details and datasets; 5.2. presenting image encryption effect; 5.3. demonstrating the comparisons with other approaches; 5.4. showing the search efficiency on two datasets; 5.5. comparison performance of different fine-tuning models; 5.6. demonstrating the generalization performance of fine-tuning model; 5.7. security analysis of hybrid encryption technique.

5.1. Implementation details and datasets

Initialize DenseNet201 as the backbone network for feature representations, the entire network is trained with Stochastic Gradient Descent (SGD) on Windows-64, AMD Ryzen5 2600X CPU@3.60 GHz, 16 GB of RAM and one NVIDIA GeForce RTX-2080ti GPU. The learning rate is initialized at 1×10^{-2}, and then decrease it to 1×10^{-3} and 1×10^{-4} at 100 epochs and 150 epochs, and stop at 200 epochs. The Momentum and batchsize are set to 0.9 and 64, respectively. Then, a brief introduction to Holidays (Jegou et al., 2008; Xia et al., 2017) and Corel10k (Wang et al., 2001; Xia et al., 2016).

[Holidays] The dataset consists of 1491 images from 500 categories of similar images. Each image category has one query, totaling 500 query images. In this paper, we adopt Average Precision (AP) as a metric to evaluate model performance. Moreover, since retrieval datasets typically have multiple query images, their respective APs are averaged to produce the final performance evaluation, namely, mean Average Precision (mAP).

[Corel10k] Corel10k is the another benchmark dataset for image retrieval performance test, which contains 10000 images of 100 objects and each category has 100 different image views. In this work, we employ average precision of the search results to evaluate the performance of our method.

5.2. Encryption effect

In this paper, a hybrid encryption function is designed, which includes ChannelEnc: color channel random replacement, SequenceEnc: bit-plane sequence conversion to binary random scrambling, and PositionEnc: pixel position $Zig-Zag$ scanning scrambling. Fig. 6 shows an example of our hybrid encryption technique on Corel10k dataset. Fig. 6(a) represents the original image, while Fig. 6(b) demonstrates the visualization result by our hybrid encryption.

To examine whether quantization processes affect the image restoration, we respectively calculate the Peak Signal to Noise Ratio (PSNR) between four groups of images on two datasets, including Original image & ChannelEnc image, Original image & SequenceEnc image, Original image & PositionEnc image and Original image & HybridEnc image. As we can be seen from the PSNR in Table 1, on Holidays, the proposed hybrid encryption technique approximates ACCH (Xia et al., 2019), while on Corel10k the PSNR of encrypted image by three techniques (hybrid encryption) is better than that of encrypted image by ones. Furthermore, it is worth noting that our encryption technique is better than PartEnc (Xu et al., 2017), either alone or combined.

Table 1
Comparison of perceptual quality (PSNR) on Corel10k and Holidays by different encryption techniques, e.g., PCBIR-CD (Xia et al., 2016), PartEnc (Xu et al., 2017), ACCH (Xia et al., 2019), SCBIR (Anju & Shreelekshmi, 2022), ChannelEnc image, SequenceEnc image, PositionEnc image, HybridEnc image and Restoration image. Here, '–' denotes that no experimental results with same settings are available.

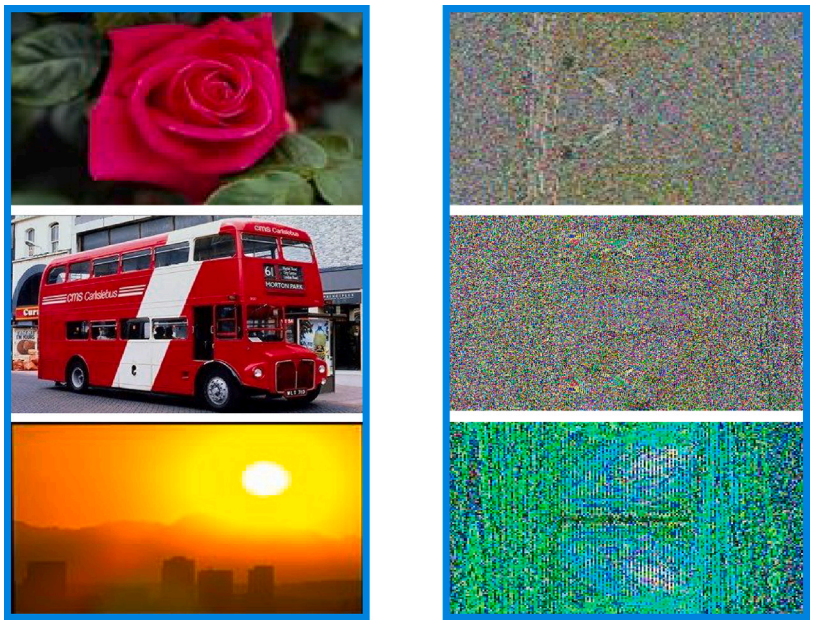
Encryption methods	PSNR (dB)	
	Holidays	Corel10k
PCBIR-CD	–	36.02
PartEnc	–	8.9547
ACCH	33.87	–
SCBIR	–	45.98
ChannelEnc image	39.06	35.83
SequenceEnc image	33.51	34.49
PositionEnc image	32.71	32.49
HybridEnc image	**32.39**	**31.98**
Restoration image	31.25	31.16

Therefore, compared with PartEnc (Xu et al., 2017) and ACCH (Xia et al., 2019), the quantization of our hybrid encryption technique is acceptable and the encrypted image can be restored after decryption.

5.3. Comparison with state-of-the-arts

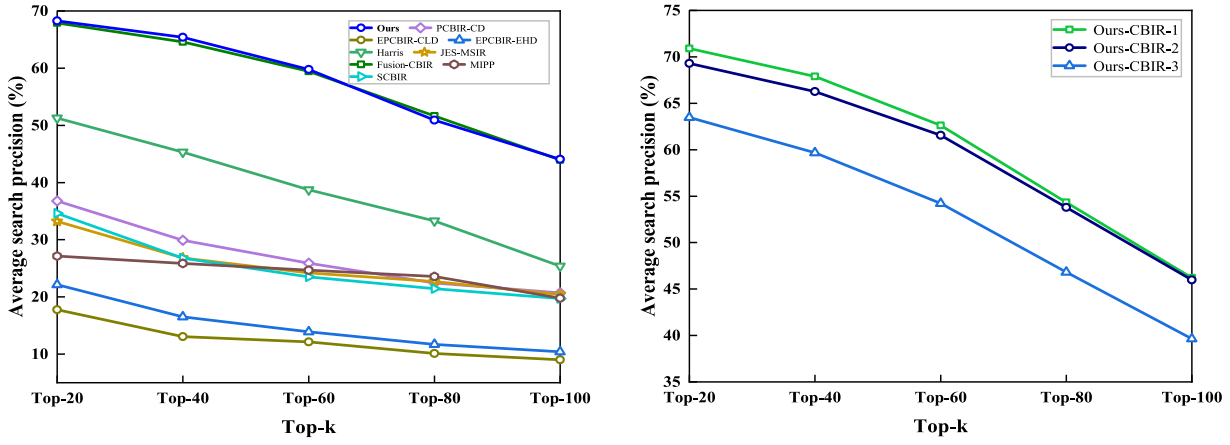
As shown in Fig. 7 and Table 2, to demonstrate the performance of our method, we compare our approach with state-of-the-art privacy-preserving CBIR methods on two public benchmark datasets, specifically PCBIR-CD (Xia et al., 2016), EPCBIR (Xia et al., 2017), Harris (Qin et al., 2019), JES-MSIR (Gu et al., 2020), Fusion-CBIR (Ma et al., 2020), MIPP (Shen et al., 2020) and SCBIR (Anju & Shreelekshmi, 2022); PKHE (Lowe, 2004), PartEnc (Xu et al., 2017), IES-CBIR (Ferreira et al., 2019), ACCH (Xia et al., 2019), LBP-BOW (Xia et al., 2020) and BOEW (Xia et al., 2022).

On Corel10k, we adopt the standard performance evaluation indicator "precision", which is defined as $P_k = k'/k$, where k and k' represent the number of all retrieval results by current query and the number of images similar to query in the results, respectively. In experiment, we divide Corel10k into training set and test set, employ the training set for fine-tuning the improved DenseNet model and then randomly select images from the test set as a query. Fig. 7(a) shows average search precision on Corel10k, it can be seen that our approach is superior to SCBIR (Anju & Shreelekshmi, 2022) and other methods for each Top-k ($k = 20, 40, 60, 80, 100$). In particular, when the retrieval image is Top-100, the performance of our method is 44.06%, which is a significant improvement compared to Harris (Qin et al., 2019) and PCBIR-CD (Xia et al., 2016). We believe the major gain comes from deep convolution feature, which enhances feature representation of the encrypted image. As shown in Fig. 7(b), to avoid uneven distribution of average search precision, we divided Corel10k dataset into three parts with different sizes image collection in the experiment: 1~3.3k, 1~6.6k, and 1~10k, randomly select 150, 300, 500 as query images, while Ours-CBIR-1, Ours-CBIR-2, Ours-CBIR-3 respectively represent their corresponding



(a) Original image (b) Hybrid encrypted image

Fig. 6. Example image from the Corel10k dataset. The (a) represents the original image and the (b) denotes the qualitative visualization result of our hybrid encrypted image.



(a) Top-k search precision.

(b) The Top-k search precision of our method on different sizes image collection of Corel10k.

Fig. 7. The (a) represents average search precision of our approach compared with PCBIR-CD (Xia et al., 2016), EPCBIR (Xia et al., 2017), Harris (Qin et al., 2019), JES-MSIR (Gu et al., 2020), Fusion-CBIR (Ma et al., 2020), MIPP (Shen et al., 2020) and SCBIR (Anju & Shreelekshmi, 2022) on Corel10k. While the (b) denotes the Top-k search precision of our method on different sizes image collection of Corel10k.

results, which indicates that the overall retrieval quality of our method is superior to SCBIR (Anju & Shreelekshmi, 2022) and other methods.

On Holidays, we adopt mAP as the evaluation metric for retrieval results of each query, which is superior to BOEW (Xia et al., 2022) and other methods. However, due to the small number of images in Holidays, the dataset needed to be data augmentation and then the improved DenseNet model fine-tuned. The dataset is increased to 8946 images by processing the images with brightness transformation, contrast transformation, vertical flip, vertical flip, and vertical&vertical flip, respectively. Table 2 shows the comparison with state-of-the-art methods, we can see that the semantic feature representations can yield superior performance in privacy-preserving CBIR. In the experiment, PKHE (Lowe, 2004), IES-CBIR (Ferreira et al., 2019) and BOEW (Xia et al., 2022) respectively leverage SIFT and color (e.g., YUV, HSV and RGB) low-level features, which could not express the semantic representation information of encrypted images. Notely, Fusion-CBIR (Ma et al., 2020) also adopts CNN semantic feature representation, our search efficiency is about 27 times higher than fusion-CBIR with only a small reduction in mAP. Moreover, even though the CNN feature representation dimension extracted by our method is high, while its efficiency is better to other methods.

5.4. Search efficiency

In this section, the efficiency of our method will be demonstrated by image encryption, feature extraction and search time.

Table 2
Comparison of mAP with PKHE (Lowe, 2004), PartEnc (Xu et al., 2017), IES-CBIR (Ferreira et al., 2019), ACCH (Xia et al., 2019), LBP-BOW (Xia et al., 2020), Fusion-CBIR (Ma et al., 2020) and BOEW (Xia et al., 2022) on Holidays. Here, "–" denotes that no experimental results with same settings are available.

Methods	Descriptor	mAP (%)	Search efficiency (ms)
PKHE	SIFT	57.90	–
PartEnc	YUV color	56.04	20
IES-CBIR	HSV color	54.564	–
ACCH	YUV color	52.938	6000
LBP-BOW	LBP features	51.595	143.0
Fusion-CBIR	CNN features & HSV	61.45	358.2
BOEW-YUV	YUV color	62.641	70.4
BOEW-RGB	RGB color	61.11	70.4
Ours	CNN features	**63.01**	**13.20**

Table 3
Feature extraction time of our method compared with IES-CBIR (Ferreira et al., 2019), JES-MSIR (Gu et al., 2020), LBP-BOW (Xia et al., 2020), DLHEIR (Pan et al., 2021) and BOEW (Xia et al., 2022) on Corel10k and Holidays. Here, "–" denotes that no experimental results with same settings are available.

Dataset/Method	**Ours**	IES-CBIR	JES-MSIR	LBP-BOW	DLHEIR	BOEW
Corel10k, time (s)	**67.2**	–	901	–	128.25	–
Holidays, time (s)	**428.8**	7050.5	–	3575.60	–	8284.66

Image encryption. Images are encrypted by ChannelEnc, SequenceEnc and PositionEnc, with time complexity of $O(3 * ImgSize)$, $O(ImgSize)$ and $O(ImgSize)$ respectively. Experiments on Corel10k and Holidays show that the average encryption time per image is 1.5 s and 24.58 s, respectively. Due to the particularity of the Holidays dataset images (relatively high resolution of each image and only 3.84M for the first image), the encryption process is longer than that of Corel10k.

Feature extraction. In this paper, we extract semantic feature representation of encrypted images in cloud server. Moreover, the feature representation is 1024-dim, which is larger than that of the previous other methods (e.g., IES-CBIR, Ferreira et al., 2019, JES-MSIR, Gu et al., 2020, LBP-BOW, Xia et al., 2020, DLHEIR, Pan et al., 2021 and BOEW, Xia et al., 2022), while the feature extraction time is shorter. Notably, we found loading the fine-tuned model to be more time-consuming in our experiments, and considering that the feature extraction is done in cloud server, we do not calculate the time consumption of this process. The efficiency of feature extraction is shown in Table 3.

Search time. Feature extraction of encrypted images dataset $E = \{e_i\}_{i=1}^{n}$ and similarity matching search of query image e_q are completed within cloud server, then the most similar Top-k candidate images are returned to the query users. In order to compare the retrieval efficiency with existing method, we divide Corel10k into five image collections of 2k, 4k, 6k, 8k, and 10k, while set Top-k candidates is 100. In Fig. 8, the semantic feature representation is 1024-dim, while feature dimensions of existing other methods (e.g., PCBIR Xia et al., 2016, EPCBIR Xia et al., 2017, MIPP Shen et al., 2020 and CBIRSH Qin et al., 2020) are low, resulting in a longer retrieval time than theirs. It is worth noting that the CBIR service of our approach is completed in cloud server. The supercomputing power of the cloud server can compensate for the inefficient retrieval.

Table 4 shows the average search time of our method and LBP-BOW (Xia et al., 2020) on Holidays, due to the particularity of the dataset, we set the Top-k = 100. Moreover, no experimental results with same settings are available in IES-CBIR (Ferreira et al., 2019), JES-MSIR (Gu et al., 2020), DLHEIR (Pan et al., 2021) and BOEW (Xia et al., 2022), so we only list the retrieval efficiency of our method and LBP-BOW. Especially, the retrieval time consumption of our method is similar to that of LBP-BOW on small sizes image collection of Holidays, with the increase of image collection, our time consumption is tardily improved, while that of LBP-BOW increases sharply.

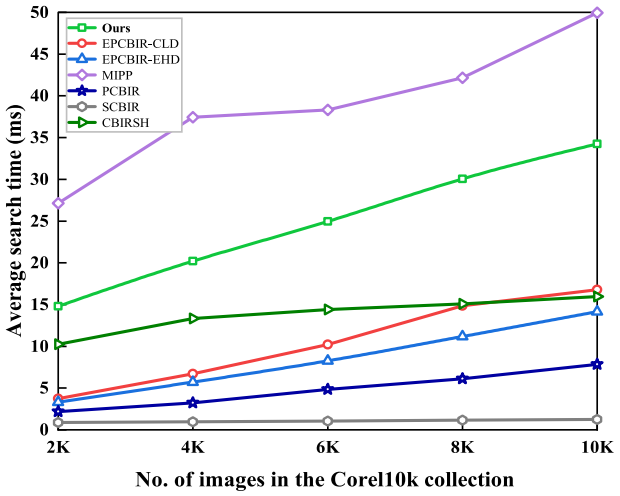


Fig. 8. Average search time (ms) of our method compared with PCBIR (Xia et al., 2016), EPCBIR (Xia et al., 2017), MIPP (Shen et al., 2020), CBIRSH (Qin et al., 2020) and SCBIR (Anju & Shreelekshmi, 2022) for different sizes of image collection on Corel10k.

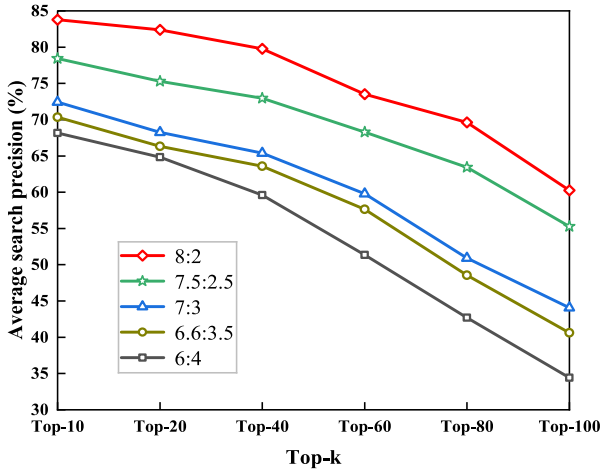


Fig. 9. Comparison of average search precision with different proportions (e.g., 6 : 4, 6.5 : 3.5, 7 : 3, 7.5 : 2.5 and 8 : 2) on Corel10k.

5.5. Comparison performance of different fine-tuning models

According to task of our work, we design a novel DenseNet model by fine-tuning and optimizing parameters, namely, improved DenseNet. In experiments, we leverage two public benchmark datasets Corel10k and Holidays with images of 10000 and 8946 (data augmentation), respectively. Then we respectively divide two encrypted images datasets into a training set and test set at the ratios of 6 : 4, 6.5 : 3.5, 7 : 3, 7.5 : 2.5 and 8 : 2 for fine-tuning training.

The experimental results are demonstrated in Figs. 9 and 10, we test different proportions average search precision and mAP on Corel10k and Holidays, respectively. It is notable that with the proportion increases, the precision is gradually improved. As can be seen from Fig. 9, by different proportions 6 : 4, 6.5 : 3.5, 7 : 3, 7.5 : 2.5 and 8 : 2 when Top-k = 100 boosted performance in the in average search precision by 34.43%, 40.61%, 44.06%, 55.27% and 60.24%, respectively. Similar results can be observed on Holidays, as we can see from Fig. 10 that the mAP is enhanced from 50.4% to 53.71%, 63.01%, 70.08%, and 79.24% through the increase of proportion. Furthermore, we carry out experiments on Holidays with different proportions under four encryption techniques, including ChannelEnc

Table 4
Average search time (ms) of our method compared with LBP-BOW (Xia et al., 2020) for different sizes of image collection on Holidays.

Methods	Feature	No. of images on Holidays dataset (10^2)											
		1	2	3	4	5	6	7	8	9	10	12	14.91
LBP-BOW	LBP features	10.17	19.63	28.34	37.26	47.69	56.23	63.26	75.19	84.43	95.69	115.62	142.37
Ours	CNN features	8.43	8.79	9.50	9.71	9.93	10.22	10.59	10.82	11.22	11.48	12.04	13.20

Table 5
Comparison of parameters, FLOPs, mAP and average search precision with different encryption techniques on Holidays and Corel10k.

Methods	Parameters	MFLOPs	Holidays, mAP (%)	Corel10k, Top-k (%))				
				Top-20	Top-40	Top-60	Top-80	Top-100
DenseNet Model (ChannelEnc)	183.71M	164.68	90.45	91.25	85.62	78.46	72.33	68.59
DenseNet Model (SequenceEnc)			73.69	78.51	72.39	65.37	58.92	52.76
DenseNet Model (PositionEnc)			73.27	77.9	73.64	66.18	56.72	50.97
DenseNet Model (HybridEnc)			61.45	67.89	64.61	59.46	51.64	43.94
Ours Model (ChannelEnc)	176.54M	158.21	92.31	92.39	86.04	77.89	72.05	69.47
Ours Model (SequenceEnc)			72.54	78.23	71.96	64.73	59.36	52.71
Ours Model (PositionEnc)			74.82	78.54	72.24	65.52	55.49	49.89
Ours Model (HybridEnc)			**63.01**	**68.26**	**65.39**	**59.78**	**50.92**	**44.06**

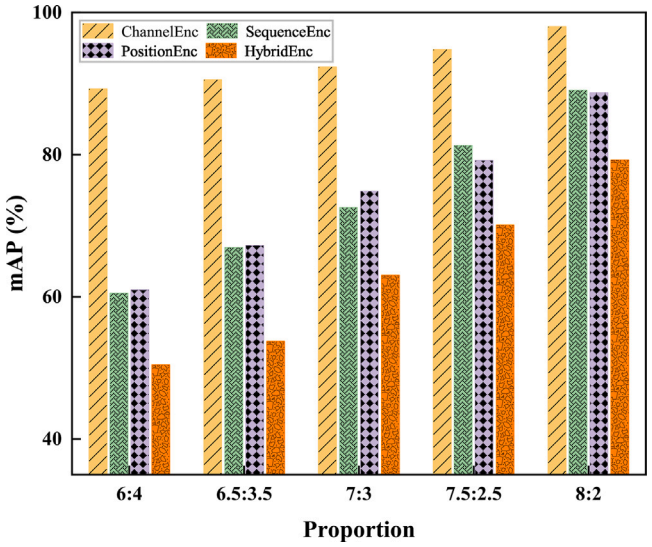


Fig. 10. Comparison of mAP with different proportions on Holidays. The ChannelEnc, SequenceEnc, PositionEnc, and HybridEnc indicate the four encryption techniques, respectively.

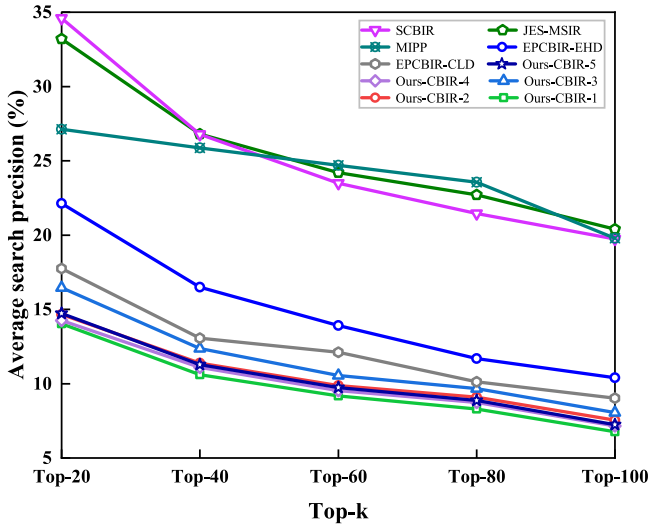


Fig. 11. Comparison of average search precision by "Holidays-model" with SCBIR (Anju & Shreelekshmi, 2022), JES-MSIR (Gu et al., 2020), MIPP (Shen et al., 2020) and EPCBIR (Xia et al., 2017) on Corel10k. Ours-CBIR-1, Ours-CBIR-2, Ours-CBIR-3, Ours-CBIR-4, and Ours-CBIR-5 represent fine-tuning training models at 6 : 4, 6.5 : 3.5, 7 : 3, 7.5 : 2.5 and 8 : 2 proportions on Holidays dataset, respectively.

image, SequenceEnc image, PositionEnc image and HybridEnc image. Fig. 10 shows that regardless of the encryption technique, the mAP improves as the proportion increases. Meanwhile, it is important to note that ChannelEnc achieves best performance, obtaining 89.23, 90.53, 92.31, 94.76 and 97.96 in mAP on 6 : 4, 6.5 : 3.5, 7 : 3, 7.5 : 2.5 and 8 : 2, respectively. By contrast, the SequenceEnc and PositionEnc lead to moderate performance, which yields 72.54 and 74.82 in mAP on 7 : 3, respectively. Although the HybridEnc leads to inferior performance in mAP on all proportions, it can better protect the image information, namely, accuracy-security trade-off of our method. However, research shows that with the proportion increases, the fine-tuning model has superior result for current dataset, but its generalization is greatly weakened. Therefore, we discuss the generalization performance of fine-tuning model on different datasets in Section 5.6.

Moreover, we compare the parameters, FLOPs, mAP and average search precision of our model with the original DenseNet model in different encryption techniques. As we can see from Table 5, the mAP and average search precision of the two models have the same trend. Floating point operations per second (FLOPs) is an indicator to measure computer performance, which is widely adopted in CNN to measure the computation cost of models, such as Sandler et al. (2018) and Pan et al. (2021). In Table 5, our method is better than the original DenseNet in terms of both parameters and FLOPs.

5.6. Generalization performance of fine-tuning model

In Section 5.5, we conduct experiments on various fine-tuning models, and the results demonstrate that the performance gets better and better as the proportion increases. We expect the fine-tuning model to have better generalization performance to other datasets. To verify the generalization, we conduct across-datasets experiment adopting fine-tuning models on two encrypted datasets, namely, the fine-tuning model for Holidays ("Holidays-model") extracts features from Corel10k and calculates average search precision, while the fine-tuning model for Corel10k ("Corel10k-model") extracts features from Holidays and calculates mAP.

As shown in Fig. 11, Ours-CBIR-1, Ours-CBIR-2, Ours-CBIR-3, Ours-CBIR-4, and Ours-CBIR-5 represent the search results of models fine-tuning by Holidays in different proportions on Corel10k, respectively. While EPCBIR-CLD and EPCBIR-EHD denote the performance of EPCBIR (Xia et al., 2017) on CLD and EHD features, respectively.

Compared with Figs. 7 and 9, the generalization performance of the five fine-tuning models by Holidays is significantly reduced across-datasets and is inferior to existing approaches (e.g., SCBIR Anju & Shreelekshmi, 2022, JES-MSIR Gu et al., 2020, MIPP Shen et al., 2020 and EPCBIR Xia et al., 2017). It is worth noting that the generalization of Ours-CBIR-3 ["Holidays-model" (7 : 3)] approximates EPCBIR-CLD. In addition, we also conduct the retrieval experiments on Holidays by five models with Corel10k fine-tuning, and the results showed that the mAP is greatly reduced. This subsection aims to verify generalization of the fine-tuning model, so we will not list experimental data compared with IES-CBIR (Ferreira et al., 2019). Meanwhile, to balance average search precision and generalization, we employ the 7 : 3 fine-tuning model for both datasets in our experiments.

5.7. Security analysis

In above, we present a model based on HBC cloud server, that is, assuming that the cloud server is HBC and it will correctly implement relevant standards under the protocol settings, which also analyze and track sensitive data information. In this section, we will verify the security of our method from five aspects: security of image content, security of image features, security of query image, possible information disclosure and image decryption & recovery.

The privacy security of image content. We leverage color channel random replacement (computational complexity is $(ImgSize * 6)!$) and pixel position scrambling (security intensity is $\log_2\{(ImgSize * 6)! * 8!\}$) to encrypt the images of size $ImgSize$. Compared to security intensity with $3 * \log_2\{101!\}$ of IES-CBIR (Ferreira et al., 2019), our method has higher safety. In Fig. 6, we illustrate original and encrypted images by our hybrid encryption manner.

The privacy security of image features. The cloud server adopts a fine-tuning improved DenseNet model to extract semantic feature representation (1024-dim) of encrypted images stored in cloud environment. Moreover, our encryption technique has been proven effective in the Ciphertext-only Attack (COA) security test (Ferreira et al., 2019; Xia et al., 2017). It is worth noting that the cloud server has access to all the feature representation, but analyzing correlations between the feature representation requires significant computational complexity ($\{1024!\}$).

The privacy security of query image. In order to protect privacy of query images, query users need to upload encrypted query images to cloud server for complete CBIR service. Therefore, security intensity of the query image is $\log_2\{(ImgSize * 6)! * 8!\}$.

The leakage of similarity information. The encrypted images datasets e_i and e_j belong to the same class, while query images e_{iq} and e_{jq} will have a high similarity score with two encrypted image datasets. For cloud server, two query images are considered similar. This type of information disclosure is common for cloud server management data, and our work hardly considers leaks from this mechanism.

The decryption and restoration of image. In privacy-preserving CBIR, PSNR is used to measure performance of image encryption and recovery effect of decrypted images. To verify effect of our hybrid encryption technique on image decryption and recovery, we perform PSNR experiments on Corel10k and Holidays. It is easy to see from Table 1 that our encryption technique outperforms existing methods (PartEnc Xu et al., 2017, ACCH Xia et al., 2019) in image decryption and recovery, which meets the security requirements of encryption system.

6. Conclusion

This paper proposes a privacy-preserving content-based image retrieval based on the deep CNN features by cloud environment. On the one hand, we extract semantic feature representation from encrypted images using fine-tuned improved DenseNet model feature extractor on the cloud server. On the other hand, we adopt a hybrid encryption technique, including ChannelEnc, SequenceEnc, and PositionEnc to protect the privacy and security of images data. Extensive experiments on two public benchmark datasets have demonstrated that our approach consistently outperforms existing methods.

In this work, we not only expect to achieve better performance but also to reduce computational burdens of the data owners. By designing a secure CBIR service in cloud server, this paper reduces the computing cost of data owners to some extent. Nevertheless, fine-tuning on edge computing platform is also time-consuming and the generalization performance of fine-tuning model on other datasets decreases significantly. As a result, we will explore an efficient privacy-preserving CBIR framework with superior generalization performance in future work.

CRediT authorship contribution statement

Wentao Ma: Conceived and designed the study, Performed the experiments, Wrote the paper, Reviewed and edited the manuscript, Read and approved the manuscript. **Tongqing Zhou:** Performed the experiments, Wrote the paper, Reviewed and edited the manuscript, Read and approved the manuscript. **Jiaohua Qin:** Conceived and designed the study, Reviewed and edited the manuscript, Read and approved the manuscript. **Xuyu Xiang:** Wrote the paper, Reviewed and edited the manuscript, Read and approved the manuscript. **Yun Tan:** Performed the experiments, Reviewed and edited the manuscript, Read and approved the manuscript. **Zhiping Cai:** Conceived and designed the study, Reviewed and edited the manuscript, Read and approved the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China under Grant 61772561, 62002392, 62072465 and 62172155; in part by the Key Research and Development Plan of Hunan Province under Grant 2019SK2022; in part by the Postgraduate Excellent teaching team Project of Hunan Province under Grant [2019]370-133; in part by the Natural Science Foundation of Hunan Province under Grant 2020JJ4141 and 2020JJ4140; in part by the Postgraduate Research and Innovation Project of Hunan Province under Grant CX20210080.

References

Amato, G., Carrara, F., Falchi, F., Gennaro, C., & Vadicamo, L. (2020). Large-scale instance-level image retrieval. *Information Processing & Management, 57*, Article 102100.

Anju, J., & Shreelekshmi, R. (2022). A faster secure content-based image retrieval using clustering for cloud. *Expert Systems with Applications, 189*, Article 116070.

Barona, R., & Anita, E. A. M. (2017). A survey on data breach challenges in cloud computing security: Issues and threats. In *2017 international conference on circuit, power and computing technologies* (pp. 1–8). IEEE.

Cheng, H., Wang, H., Liu, X., Fang, Y., Wang, M., & Zhang, X. (2019). Person re-identification over encrypted outsourced surveillance videos. *IEEE Transactions on Dependable and Secure Computing*.

Ferreira, B., Rodrigues, J., Leitao, J., & Domingos, H. (2019). Practical privacy-preserving content-based retrieval in cloud image repositories. *IEEE Transactions on Cloud Computing, 7*, 784–798.

Gkelios, S., Sophokleous, A., Plakias, S., Boutalis, Y., & Chatzichristofis, S. A. (2021). Deep convolutional features for image retrieval. *Expert Systems with Applications, 177*, Article 114940.

Gu, Q., Xia, Z., & Sun, X. (2020). MSPPIR: Multi-source privacy-preserving image retrieval in cloud computing. arXiv preprint arXiv:2007.12416.

Huang, G., Liu, Z., Pleiss, G., Van Der Maaten, L., & Weinberger, K. (2019). Convolutional networks with dense connectivity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, http://dx.doi.org/10.1109/TPAMI.2019.2918284, 1–1.

Hussain, S., Zia, M. A., & Arshad, W. (2021). Additive deep feature optimization for semantic image retrieval. *Expert Systems with Applications*, *170*, Article 114545.

Jegou, H., Douze, M., & Schmid, C. (2008). Hamming embedding and weak geometric consistency for large scale image search. In *European conference on computer vision* (pp. 304–317). Springer.

Li, Y., Ma, J., Miao, Y., Wang, Y., Liu, X., & Choo, K. -K. R. (2020). Similarity search for encrypted images in secure cloud computing. *IEEE Transactions on Cloud Computing*, 1–1.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*, 91–110.

Lu, W., Swaminathan, A., Varna, A. L., & Wu, M. (2009). Enabling search over encrypted multimedia databases. In *Media forensics and security*: *Vol. 7254*, International Society for Optics and Photonics, Article 725418.

Lu, W., Varna, A. L., Swaminathan, A., & Wu, M. (2009). Secure image retrieval through feature protection. In *2009 IEEE international conference on acoustics, speech and signal processing* (pp. 1533–1536). IEEE.

Lu, W., Varna, A. L., & Wu, M. (2014). Confidentiality-preserving image search: A comparative study between homomorphic encryption and distance-preserving randomization. *IEEE Access*, *2*, 125–141.

Ma, W., Qin, J., Xiang, X., Tan, Y., & He, Z. (2020). Searchable encrypted image retrieval based on multi-feature adaptive late-fusion. *Mathematics*, *8*, 1019.

Ma, W., Qin, J., Xiang, X., Tan, Y., Luo, Y., & Xiong, N. N. (2019). Adaptive median filtering algorithm based on divide and conquer and its application in CAPTCHA recognition. *Computers, Materials & Continua*, *58*, 665–677.

Öztürk, Ş. (2020). Stacked auto-encoder based tagging with deep features for content-based medical image retrieval. *Expert Systems with Applications*, *161*, Article 113693.

Pan, W., Wang, M., Qin, J., & Zhou, Z. (2021). Improved CNN-based hashing for encrypted image retrieval. *Security and Communication Networks*, *2021*.

Qin, J., Cao, Y., Xiang, X., Tan, Y., Xiang, L., & Zhang, J. (2020). An encrypted image retrieval method based on SimHash in cloud computing. *Computers Materials & Continua*, 11.

Qin, J., Chen, J., Xiang, X., Tan, Y., Ma, W., & Wang, J. (2020). A privacy-preserving image retrieval method based on deep learning and adaptive weighted fusion. *Journal of Real-Time Image Processing*, *17*, 161–173.

Qin, J., Li, H., Xiang, X., Tan, Y., Pan, W., Ma, W., & Xiong, N. N. (2019). An encrypted image retrieval method based on Harris corner optimization and LSH in cloud computing. *IEEE Access*, *7*, 24626–24633.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. -C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In *IEEE conference on computer vision and pattern recognition* (pp. 4510–4520).

Shen, M., Cheng, G., Zhu, L., Du, X., & Hu, J. (2020). Content-based multi-source encrypted image retrieval in clouds with privacy preservation. *Future Generation Computer Systems*, *109*, 621–632.

Song, L., Miao, Y., Weng, J., Choo, K. -K. R., Liu, X., & Deng, R. H. (2022). Privacy-preserving threshold-based image retrieval in cloud-assisted Internet of Things. *IEEE Internet of Things Journal*.

Tang, J., Xia, Z., Wang, L., Yuan, C., & Zhao, X. (2021). OPPR: An outsourcing privacy-preserving JPEG image retrieval scheme with local histograms in cloud environment. *Journal on Big Data*, *3*, 21.

Wang, J. Z., Li, J., & Wiederhold, G. (2001). SIMPLIcity: Semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*, 947–963.

Wang, H., Xia, Z., Fei, J., & Xiao, F. (2020). An AES-based secure image retrieval scheme using random mapping and BOW in cloud computing. *IEEE Access*, *8*, 61138–61147.

Weng, L., Amsaleg, L., & Furon, T. (2016). Privacy-preserving outsourced media search. *IEEE Transactions on Knowledge and Data Engineering*, *28*, 2738–2751.

Xia, Z., Jiang, L., Liu, D., Lu, L., & Jeon, B. (2022). BOEW: A content-based image retrieval scheme using bag-of-encrypted-words in cloud computing. *IEEE Transactions on Services Computing*.

Xia, Z., Jiang, L., Ma, X., Yang, W., Ji, P., & Xiong, N. N. (2019). A privacy-preserving outsourcing scheme for image local binary pattern in secure industrial Internet of Things. *IEEE Transactions on Industrial Informatics*, *16*, 629–638.

Xia, Z., Lu, L., Qiu, T., Shim, H., Chen, X., & Jeon, B. (2019). A privacy-preserving image retrieval based on AC-coefficients and color histograms in cloud environment. *Computers Materials & Continua*, *58*, 27–43.

Xia, Z., Wang, L., Tang, J., Xiong, N. N., & Weng, J. (2020). A privacy-preserving image retrieval scheme using secure local binary pattern in cloud computing. *IEEE Transactions on Network Science and Engineering*, *8*, 318–330.

Xia, Z., Wang, X., Zhang, L., Qin, Z., Sun, X., & Ren, K. (2016). A privacy-preserving and copy-deterrence content-based image retrieval scheme in cloud computing. *IEEE Transactions on Information Forensics and Security*, *11*, 2594–2608.

Xia, Z., Xiong, N. N., Vasilakos, A. V., & Sun, X. (2017). EPCBIR: An efficient and privacy-preserving content-based image retrieval scheme in cloud computing. *Information Sciences*, *387*, 195–204.

Xia, Z., Zhu, Y., Sun, X., Qin, Z., & Ren, K. (2015). Towards privacy-preserving content-based image retrieval in cloud computing. *IEEE Transactions on Cloud Computing*, *6*, 276–286.

Xu, Y., Gong, J., Xiong, L., Xu, Z., Wang, J., & Shi, Y.-q. (2017). A privacy-preserving content-based image retrieval method in cloud environment. *Journal of Visual Communication and Image Representation*, *43*, 164–172.

Zheng, L., Yang, Y., & Tian, Q. (2018). SIFT meets CNN: A decade survey of instance retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*, 1224–1244.