

ORIGINAL RESEARCH

Open access

Visual Cryptography Meets Remote Sensing Intelligence: A Secure and Efficient Recognition Framework

David Thompson^{1*}, Sarah Mitchell¹, Rachel Adams², James Brown³, Michael Lee⁴

Abstract

With the emergence of novel data collection platforms driven by the Internet of Things (IoT), this study expands beyond the conventional boundaries of traditional remote sensing technologies. The deep integration of remote sensing and computer vision has transitioned into industrial applications, enabling the utilization of Artificial Intelligence to address challenges like automated information extraction and image analysis. Nonetheless, due to the intricate architecture of IoT and the absence of standardized security protocols, remote sensing instruments are highly susceptible to privacy breaches during data transmission. Because traditional cryptographic approaches depend heavily on computational complexity, it is critical to develop security frameworks tailored for resource-constrained IoT devices. Visual Cryptography (VC) serves as an image-based threshold technique that allows direct decoding by the human visual system simply by overlaying encrypted shares. This "stacking-to-see" attribute, combined with basic Boolean decryption processes, positions VC as an optimal candidate for privacy-preserving identification within extensive IoT-based remote sensing imagery. In this investigation, the secure and productive dissemination of high-resolution remote sensing imagery is accomplished via meaningful VC. By distributing the error between the encryption block and the source block to neighboring blocks, the fidelity degradation in the reconstructed images is minimized. Furthermore, by adjusting a pre-trained model using large-scale datasets, the recognition accuracy of small, encrypted remote sensing datasets is enhanced. The empirical findings demonstrate that the introduced lightweight, privacy-preserving recognition system maintains superior identification performance while strengthening data security.

Keywords Artificial intelligence, Remote sensing, Internet of Things (IoT), Visual Cryptography (VC)

*Correspondence:

David Thompson
david.thompson@utoronto.ca

¹ Department of Remote Sensing and Visual Cryptography, Faculty of Engineering, University of Toronto, Toronto, Canada

² Department of Privacy-Preserving Image Recognition, Faculty of Engineering, University of British Columbia, Vancouver, Canada

³ Department of Secure Remote Sensing Intelligence, Faculty of Engineering, McGill University, Montreal, Canada

⁴ Department of Limited Visual Cryptography, Faculty of Engineering, University of Guelph, Guelph, Canada

Introduction

Driven by the technological advancements of the IoT, we are currently capable of acquiring vast volumes of remote-sensing data featuring high spectral, spatial, and temporal resolutions; this data has been effectively deployed across diverse domains, including environmental monitoring, land-use surveying, disaster response management, and mineral exploration [1]. Nonetheless, securing remote sensing data presents numerous hurdles due to massive data capacities, intricate data configurations, limited labeled samples, and non-convex optimization problems. Computational intelligence methods, inspired by biological cognitive systems, offer viable remedies to these difficulties [2, 3].

Artificial intelligence (AI), primarily propelled by deep learning (DL), has outperformed human capabilities in image processing, speech recognition, and language translation since 2013 [4]. The widespread utilization of big data and cloud computing has accelerated the expansion of AI. Massive spatial-temporal datasets gathered via IoT sensor networks serve as the foundational element of AI [5], offering crucial backing for decision support systems (DSS) and operational services within remote sensing applications [6]. The profound convergence of remote sensing and computer vision has proven highly effective for automated feature extraction and image analysis.

With the introduction of modern data collection platforms, we are advancing beyond classical remote sensing paradigms such as aerial mapping, geographic information systems, and Earth observation. Interlinked devices across these integrated platforms are capable of continuous data transmission [7, 8]. Intellectual sensors can now be mounted on minimized platforms. The operational mobility and cost efficiencies of these modern platforms render them highly suitable for continuous monitoring deployments compared to traditional airborne vehicles [9, 10]. Nonetheless, owing to procurement expenses, remote sensing datasets remain typically constrained in size. Training a neural network (NN) classifier directly on such tasks frequently results in overfitting, given that tens of thousands of parameters must be optimized within the model.

The widespread adoption of AI has intensified societal anxieties regarding the privacy exposure of remote-sensing data. Because of the intricate layout of the IoT and the lack of standardized protection frameworks, remote sensing nodes are highly vulnerable to privacy violations during data sharing [11, 12]. To thwart the exposure of confidential details—such as coastal island topographies, water coloration, and maritime vessels—in remote sensing imagery, it is vital to encrypt and secure these images prior to uploading them to public cloud infrastructures [13, 14]. Sensor-layer protection methodologies primarily target threats directed at resource-constrained, compact, and low-power sensors; such devices cannot support conventional security mechanisms that demand substantial computing and storage capacities to defend against intrusions [15]. These computational resources and transmission bandwidths are highly restricted within the IoT [16]. Constructing dependable AI systems and strengthening their privacy attributes has surfaced as a vital domain in technology governance. While a substantial body of literature dedicated to data security has been published, the majority concentrates on text and numerical data, which differ significantly from sensing imagery in terms of volume and scale [17]. For instance, encrypting every single pixel is not mandatory to guarantee the secure transmission of an image [18].

To resolve these difficulties, this manuscript introduces a lightweight privacy protection framework for remote sensing imagery. This system securely distributes vast volumes of remote sensing data in an open environment using the secret-sharing characteristics of Visual Cryptography (VC). By utilizing basic Boolean operations, the proposed framework bypasses the intensive computational overhead and key management dependencies characteristic of traditional encryption, thereby facilitating secure transmission over resource-constrained devices. Additionally, to mitigate the drop in recognition performance caused by noise interference, a fine-tuned network architecture is deployed to enhance identification accuracy on the encrypted data [19]. The primary contributions of this research are outlined below:

- (i) We realize the secure and productive transmission of high-resolution remote sensing imagery via meaningful VC.
- (ii) We enhance the visual quality of the decrypted imagery by applying a restricted error diffusion methodology.
- (iii) We optimize the recognition capabilities of lossy remote sensing imagery by fine-tuning a model anchored in metric learning.

Related Work

Intelligence technology for remote sensing image

Deep learning has established itself as the cornerstone of AI and has captured increasing interest within the remote sensing sector. Chen *et al.* [20] executed two-dimensional convolutions across image planes and one-dimensional convolutions along the spectral axis to isolate spatial and spectral image characteristics, respectively, yielding superior

classification accuracy compared to the random forest approach. Hu *et al.* [21] utilized a DL-centric methodology to evaluate Lidar (Laser Radar) point cloud data, capturing insights during human-computer interaction and subsequently implementing them for automated point cloud processing. This automated analysis achieved an accuracy rate of 95%, thereby substantially boosting operational throughput.

To balance the dependency of DL on massive data volumes with the requirements of copyright protection and data confidentiality in remote sensing, investigators have introduced various strategies. Remote sensing imagery, particularly high-resolution satellite data tied to public safety, demands stringent confidentiality; conversely, DL models require the open sharing of such information to effectively learn underlying features. To reconcile this tension, Wadii *et al.* [22] integrated Paillier homomorphic encryption to bolster the defense of satellite imagery. The empirical outcomes indicate that this approach can successfully execute privacy-preserving DL on ultra-high-resolution remote sensing data. Nevertheless, this framework remains dependent on classical public-key cryptography, leaving open questions regarding efficient key management. To tackle the problem of secure remote-sensing image storage across open cloud infrastructures, Qi *et al.* [23] introduced an image-sharing framework based on weighted cloud service allocation. This method achieves secure and productive storage of massive imagery by leveraging the Chinese Remainder Theorem.

Visual cryptography scheme

Ensuring the confidentiality of captured big data has long been a primary focus of remote sensing technology. Conventional approaches protect sensitive information by means of encryption. However, cryptographic operations are typically computationally intensive, making them prohibitively expensive for resource-constrained IoT terminals. Furthermore, complex encryption algorithms delay application response times and degrade overall performance across IoT systems.

In contrast to text-oriented secret-sharing frameworks that divide a sum into two distinct addends, Visual Cryptography (VC) [24, 25] divides secret pixels into multiple subpixels. It represents a specialized branch of secret sharing designed explicitly for image security [26]. In the baseline VC Scheme (VCS), a single secret pixel is transformed into multiple pixel blocks, which triggers pixel expansion. To address this, Zhang *et al.* [27] developed an alternative VCS based on a cover model that eliminates size expansion. Their approach achieves nearly lossless reconstruction through vertical expansion and contrast adjustment; however, it produces eight or more shares, which introduces a heavy transmission overhead for the encrypted images. By leveraging secret sharing, VCS bypasses the throughput degradation typically seen during the secure transmission of large-scale data, leading to its widespread adoption across multiple domains [27–29].

Rather than generating noisy shares, the Extended VCS (EVCS) [28] utilizes meaningful cover images to prevent attracting the attention of malicious actors. Nevertheless, it remains constrained by pixel expansion. Lee *et al.* [29] introduced a generalized access structure generation algorithm to resolve pixel expansion in conventional EVCCS. This approach first builds a meaningless stock image via a simulated annealing algorithm and subsequently distributes the black pixels of the source image evenly across the stock utilizing a sampling algorithm. These optimization steps, however, render the encryption process more intricate than traditional VCS.

Materials and Methods

The core methodology behind this privacy-preserving recognition framework for remote sensing imagery relies on two components: (i) an enhanced EVCS developed for the secure and rapid transmission of remote sensing data, and (ii) a metric learning approach engineered to maximize the recognition accuracy of encrypted remote sensing imagery. The following sections detail the enhanced EVCS.

The list of notations

Table 1 details the list of notations utilized throughout this paper.

Table 1. The list of notations used in this document.

Notation	Term
	The primary cover image
	The secondary cover image
	The encryption matrices corresponding to pixels in the secret and cover images, respectively
	Threshold function
	The (i)-th pixel in the (j)-th encryption matrix, where
	The number of permitted grey levels
	The pixel block located at the (i)-th row and (j)-th column, where
	The grey intensity value of the color block
	The number of black pixels within the restricted color block
	The total number of pixels in the secret block
	The quantization error associated with the pixel
	Optimal arrays designed to improve diffusion smoothness

Limited block encryption for expansion-free EVCS

Traditional VC is restricted to handling black-and-white binary data. To accommodate various image types, we introduce a pre-processing stage to support different encryption modes: colour, grey, and black-and-white. Among these, colour images represent the most frequent format. For colour data, we isolate each individual colour channel—such as (R, G, B)—and apply a dithering operation to every channel to transform each pixel into a binary black-and-white format [30, 31]. Grey-scale imagery is processed in an identical manner to single-channel images following the dithering stage.

Algorithm 1 outlines the procedure for encrypting different image modes via VC. Once the target data is converted into a black-and-white format, the resulting halftoned images are encrypted pixel by pixel using VCS. We then restrict the available sharing blocks to guarantee that the colour block distribution of the remote sensing images satisfies existing EVCS security parameters both before and after encryption. Following block encryption, the error introduced between the encrypted block and the source block is diffused to neighboring blocks to boost final image quality.

```

Data: A  $M \times N$  secret image  $S$ 
Result: share images  $H_1, H_2$  held by participants  $P_1, P_2$ 
1 if  $S$  is a color or gray image then
2   |  $BW \leftarrow \text{PreProcess}(S)$ 
3 end
4 if  $CI_i$  is a color or gray image, for  $i \in [0, 1]$  then
5   |  $CI_i \leftarrow \text{Resize}(CI_i, S)$ 
6   |  $BW_{CI_i} \leftarrow \text{PreProcess}(CI_i)$ 
7 end
8 foreach block  $b_{Bc1}, b_{Bc2}, b_s$  in the  $x$ -row and  $y$ -column of
    $BW, BW_{CI1}, BW_{CI2}$  do
9   |  $b_{Bi} \leftarrow \lfloor l_L / (l_b + 1) \times (T - \sum B_g / 255) \rfloor$ 
10  |  $c_b^{B_{sij}} \leftarrow [\max(0, b_{Bc1ij} + b_{Bc2ij} - s_{B_{sij}}), \min(c_b^{B_{s1ij}}, c_b^{B_{s2ij}})]$ 
11  | Rearrange  $S_{S_x}^{c1c2}$  into  $H_{ix,y}$ , for  $i \in [0, 1]$ 
12  | Difuse the error between  $b_{Bcix,y}$  and  $H_{ix,y}$  to adjacent pixels, for  $i$ 
   |  $\in [0, 1]$ 
13 end

```

Algorithm 1. A limited error diffusion algorithm for expansion-free EVCS

When a pixel at a given position in the first cover image (C1) is white (0) and the corresponding pixel in the second cover image (C2) is black (1), the candidate encryption matrices are structured as follows:

$$S_p^{p_{c1}p_{c2}} = S_0^{01} \quad (1)$$

Because the black and white pixels within the encrypted blocks no longer possess identical contrast, they become indistinguishable. The total number of black pixels in the encrypted white and black blocks is two and three, respectively:

$$S_p^{p_{c1}p_{c2}} = S_0^{p_{c1}p_{c2}}, \text{OR}(p_{c1}, p_{c2}) = 0 \quad (2)$$

$$S_p^{p_{c1}p_{c2}} = S_0^{p_{c1}p_{c2}}, \text{AND}(p_{c1}, p_{c2}) = 1 \quad (3)$$

Consequently, black blocks exhibit higher contrast than white blocks, taking advantage of the low-pass filtering characteristics inherent in the human visual system. The resulting encrypted subpixels $Sc1$ and $Sc2$ remain meaningful, though they are four times larger than the original source:

$$S_p^{p_{c1}p_{c2}} = \begin{bmatrix} \begin{bmatrix} s_{11} & s_{12} \\ s_{13} & s_{14} \end{bmatrix} \\ \begin{bmatrix} s_{21} & s_{22} \\ s_{23} & s_{24} \end{bmatrix} \end{bmatrix} \quad (4)$$

It should be noted that while the encryption matrix for S_0^{00} deviates from other colour blocks, the left-white and right-black block structures still appear within the matrix S_1^{10} for Sc2. Because the sharing blocks corresponding to a secret pixel can appear entirely on the opposite side, an adversary cannot deduce the source pixel from an isolated share.

$$\text{OR}(s_{1i}, s_{2i}) = \tau_c(p_s), i < 4 \quad (5)$$

$$\sum s_{1i} = \tau_c(p_{c1}) \quad (6)$$

$$\sum s_{2i} = \tau_c(p_{c2}) \quad (7)$$

Like traditional VCS, EVCS operates on an image pixel by pixel, which triggers pixel expansion. To enforce the required limitation, we can process equal-sized secret blocks simultaneously. If error diffusion (ED) is applied directly to a colour block, the resulting blocks may contain anywhere from zero to four black pixels. However, the original EVCS mandates that preprocessed cover images contain strictly two or three black patches. As a result, if a sharing block contains zero, one, or four black pixels, existing general access structures cannot be recycled. In addition to binary imagery, threshold functions can be established to yield ternary or quaternary images via:

$$b_{Bi} = L \lfloor l_L / (l_b + 1) \times (T - \sum g_B / 255) \rfloor \quad (8)$$

where L denotes the permissible grey level, which typically takes a value of [2, 3] or [3, 4] for EVCS. g_B represents the grey-scale value of the colour block. l_L indicates the count of allowable grey values (which is set to two), while l_b denotes the total pixels within the colour block (which is set to four). Eq. 8 ensures that the number of black pixels in the pre-processed sharing colour blocks does not surpass the black pixel count of the corresponding superimposed secret blocks, thereby permitting block-by-block encryption.

$$c_b^{B_{sij}} = \left[\max(0, b_{Bc1ij} + b_{Bc2ij} - s_{Bij}), \min(c_b^{B_{c1ij}}, c_b^{B_{c2ij}}) \right] \quad (9)$$

$$0 \leq i < M/m, 0 \leq j \leq N/n \quad (10)$$

where $c_b^{B_{sij}}$ represents the value of the secret block B_{sij} , $c_b^{B_{c1ij}}$ and $c_b^{B_{c2ij}}$ correspond to the values of the cover blocks B_{c1ij} and B_{c2ij} , which match the dimensions of B_{sij} . M and N represent the overall dimensions of the secret and cover images, while m and n define the dimensions of the colour blocks.

To preserve the security of VC, the human eye can still extract meaningful data from decrypted images even in the absence of completely white blocks, provided there is a contrast differential between black and white colour blocks.

As illustrated in **Figure 1**, halftoning is utilized to emulate continuous tones by modulating the frequency of a constrained set of colours. The visual appearance of the quantized image closely matches the original when observed from an appropriate distance. This halftone approach leverages the low-pass filtering and colour perception characteristics of the human eye to generate optimized images on monochrome, binary, or four-colour printing systems. The eye naturally blends spatially adjacent image segments into continuous tones. Error diffusion [31] is a widely adopted halftoning technique formulated as follows:

$$p_{x,y} += \sum_{ij} e_{x,y} \times c_{x-i,y-j} \quad (11)$$

where $e(x, y)$ represents the localized quantisation error, typically derived from a threshold function truncated by the system. The coefficients w_{ij} comprise an optimal array designed to smooth the diffusion process, defined as [7/16, 3/16, 5/16, 1/16]. By accounting for the grey values of neighboring pixels, the resulting image is smoother and exhibits fewer noise artifacts compared to random or jittered dithering techniques.

$$qe = p_{x,y} - T_{x,y} \quad (12)$$

$$T_{x,y} = 0, \text{ if } p_{x,y} < \tau \quad (13)$$

$$T_{x,y} = 1, \text{ if } p_{x,y} \geq \tau \quad (14)$$

$$c = \begin{bmatrix} p_{old} & p & 7/16 \\ 3/16 & 5/16 & 1/16 \end{bmatrix} \quad (15)$$

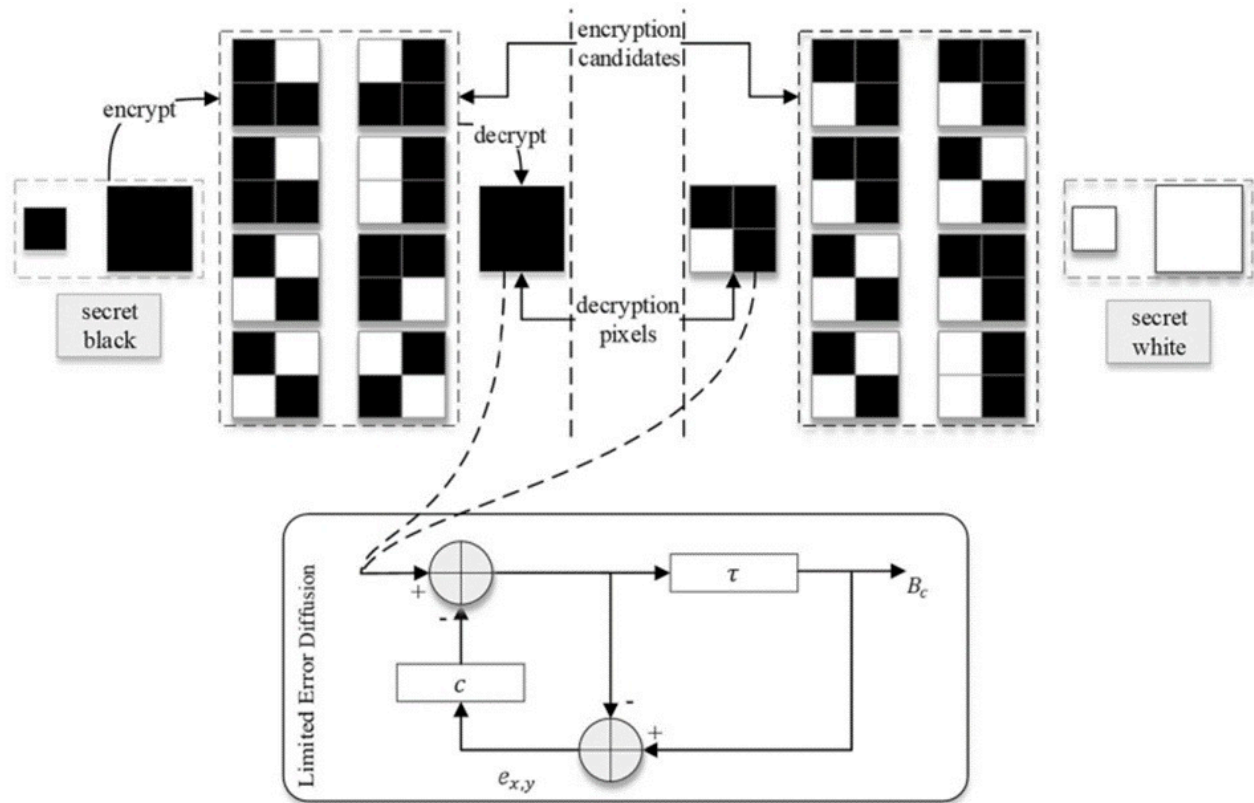


Figure 1. The limited error-diffusion algorithm.

High-precision recognition model for encrypted datasets

As established in the preceding section, ensuring the security of shared images prevents us from completely restoring pure white colour blocks. A white block is decrypted into a block containing three black pixels, which decreases its contrast by 3/4. This reduction in contrast degrades the classification performance of conventional models operating on decrypted images. To overcome this limitation, we next outline a high-accuracy recognition method tailored for lossy datasets utilizing transfer learning [32].

In deep learning, numerous frameworks update and tune model parameters based on the loss gradients calculated from paired samples [33, 34]. During the training of a neural network model, two samples are selected at random, their features are extracted by the network, and the spatial distance between them is computed. If the pair belongs to the same category, their distance should approach zero; if they belong to separate categories, the distance should be maximized. This core concept underpins a variety of loss functions. Contrastive Loss is defined as follows:

$$L(W, Y, X_1, X_2) = \frac{(1 - Y)}{2} \times (D_w)^2 + \frac{Y}{2} \times \max(0, m - D_w)^2 \quad (16)$$

where W represents the network weights and Y denotes the paired sample labels. If the samples share the same class, $Y=0$; if they belong to different classes, $Y=1$. D_w signifies the Euclidean distance within the latent embedding space. When $Y=0$, parameters are optimized to minimize the distance between X_1 and X_2 . When $Y=1$ and the distance between X_1 and X_2 exceeds the margin m , no optimization occurs; otherwise, the distance is driven up to meet the margin m .

Triplet Loss [35] aims to draw features sharing the same label closer together in the embedding space while pushing features with different labels further apart. This loss function simultaneously samples a positive and a negative instance during training, both anchored to a reference baseline sample. This selection yields a trio of samples, termed a Triplet. When the distance separating the negative pair from the positive pair exceeds a specified threshold, the loss drops to 0, concluding the training iteration for that set; otherwise, gradient iteration continues.

$$L = \max (d(a, p) - d(a, n) + m, 0) \quad (17)$$

where a represents the anchor sample, p is a positive sample belonging to the same category as a , and n is a negative sample belonging to a different category than a .

The Triplet approach minimizes interclass distance while simultaneously accentuating intraclass differences. Negative samples located too close to the anchor are penalized based on their calculated Euclidean distance. Because it explicitly balances the distances among positive, negative, and anchor samples, Triplet Loss frequently yields superior training outcomes compared to Contrastive Loss.

As depicted in **Figure 2**, we present a high-precision recognition network engineered for lossy encrypted datasets leveraging metric learning [36]. We first construct an encrypted dataset from the source remote sensing images using our limited EVCS. Following the protocol of the Triplet architecture, we select two positive image samples and one negative image sample. The two positive samples consist of a standard image and an encrypted image belonging to the same category. The negative samples are drawn with equal probability from either the encrypted or standard datasets. We employ a pre-trained ResNet model as the core feature extractor, topped with a fully-connected layer to execute the final classification of the remote sensing images.

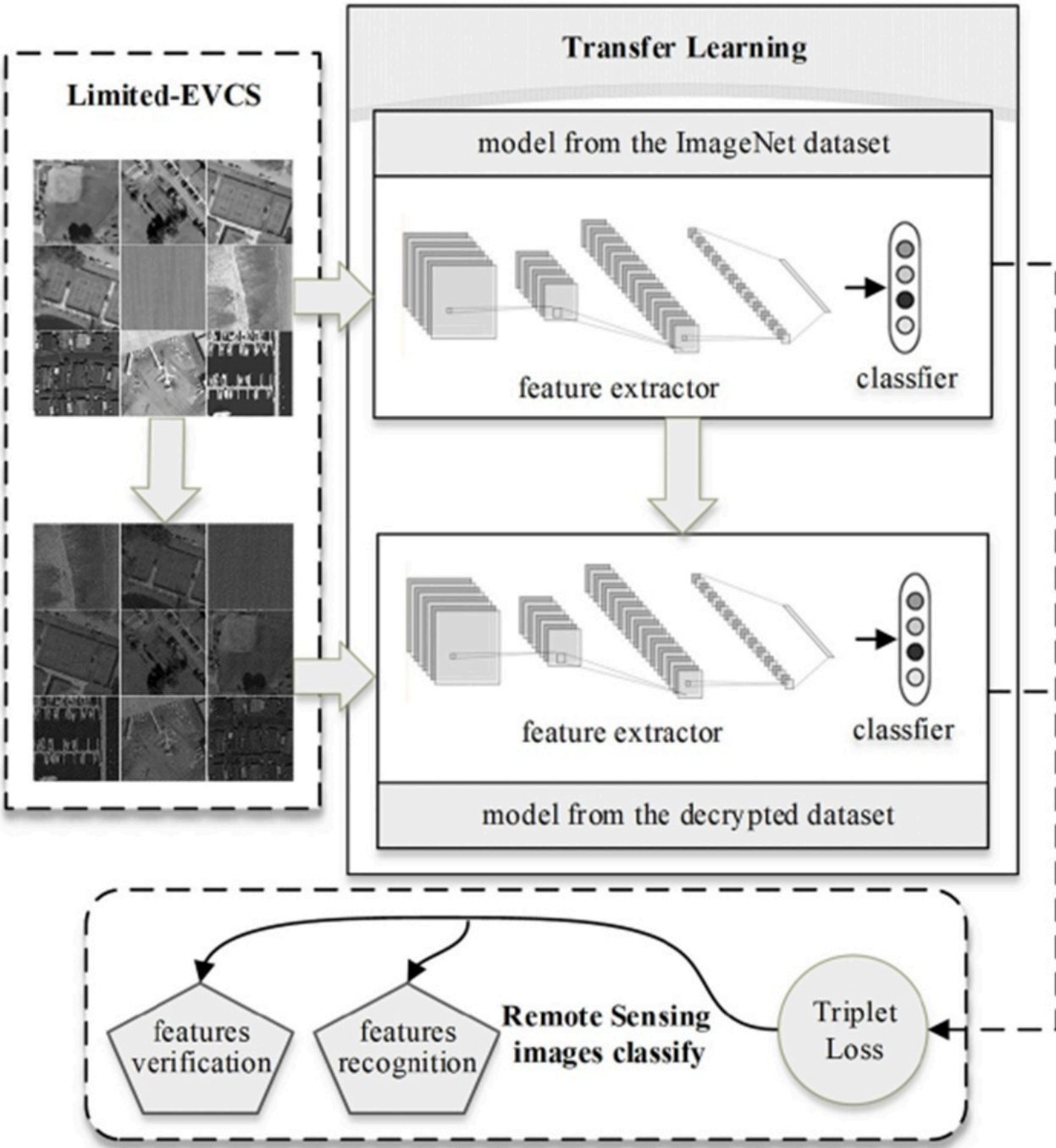


Figure 2. A high-precision recognition model for encrypted remote sensing datasets.

Experiments and Discussion

In this section, the proposed privacy-preserving framework is implemented to evaluate its empirical performance in classifying remote sensing imagery.

Table 2 demonstrates the structural quality of the reconstructed images generated by our proposed method in comparison to the techniques of Zhang, Kang, and Ren, utilizing 21 distinct image classes from the UCMerced [37] dataset. Quantitative evaluation metrics selected for this experiment comprise Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) [38]. The results indicate that our methodology maintains superior image contrast via the limited black encryption protocol. While Zhang's technique maintains a constant pixel size throughout the encryption phase by processing individual pixel blocks sequentially, it sacrifices a degree of contrast within the block, yielding lower reconstructed image quality than our approach. Kang's approach enhances image fidelity through visual information pixel synchronization and halftone processing driven by a similar error diffusion mechanism. The decrypted images from Kang's method achieve a PSNR between 30 and 31. However, because Kang's method relies on an unconventional halftoning strategy to diffuse errors during three-channel pixel synchronization, its encryption process introduces a heavy computational overhead, which runs counter to the efficiency objectives of a VCS. Conversely, Ren's framework yields the lowest decryption quality. A primary limitation of that scheme is that it handles the entire encryption block as a macro-pixel during dithering. This process effectively downsamples the secret image, degrading image quality during the pre-processing stage. Furthermore, the systematic noise introduced by VC during encryption further degrades the reconstruction fidelity of Ren's method.

Table 2. The average encryption performance of our method, Wang's method, and Zhang's method in encrypting 21 classes in the UCMerced dataset.

Class	Ren's [39] (SSIM)	Ren's [39] (PSNR)	Zhang's [27] (PSNR)	Zhang's [27] (SSIM)	Kang's [40] (SSIM)	Kang's [40] (PSNR)	Ours (SSIM)	Ours (PSNR)
Agricultural	1.34	22.74	26.70	6.72	1.89	24.16	7.01	30.31
Aeroplane	1.33	21.46	26.58	6.88	2.27	24.44	6.30	32.69
Baseball diamond	1.97	21.73	25.22	2.75	1.78	23.02	4.50	32.90
Beach	1.78	22.42	25.78	4.88	2.13	24.01	9.02	32.30
Buildings	1.50	22.20	25.93	4.32	2.38	24.55	6.74	32.20
Chaparral	1.59	22.03	26.40	2.58	1.24	24.20	8.80	32.29
Dense residential	1.29	22.65	26.72	2.15	1.38	23.65	4.69	31.71
Forest	1.47	22.78	26.84	2.14	1.99	23.08	7.95	32.58
Freeway	1.56	21.16	26.53	2.06	1.61	23.43	7.20	30.61
Golfcourse	1.87	21.96	26.93	3.69	1.59	24.06	7.49	31.62
Harbour	1.05	22.62	26.19	6.34	2.29	23.38	9.70	31.72
Intersection	1.89	21.69	25.01	6.86	2.53	23.65	7.50	32.19
Medium residential	1.93	21.72	25.26	6.37	1.01	23.51	9.45	30.26
Mobile home park	1.45	22.38	25.65	5.09	1.10	23.82	7.89	31.38
Overpass	1.81	22.38	25.72	3.16	2.34	24.93	8.83	32.30
Parking lot	1.79	22.81	26.49	5.50	1.25	24.94	8.85	31.51

River	1.71	21.83	25.82	5.53	2.36	24.72	9.98	32.70
Runway	1.13	22.43	25.83	6.42	1.76	23.44	8.09	31.67
Sparse residential	1.75	21.69	26.40	5.87	1.24	24.73	4.57	31.31
Storage tanks	1.66	22.35	26.94	3.37	1.38	24.68	8.31	30.63
Tennis court	1.19	21.90	26.82	6.93	2.19	24.49	9.20	30.47

Figure 3 illustrates a comparative analysis between our approach and traditional cryptographic frameworks across varying quantities of processed images. The processing latency of the proposed encryption algorithm is benchmarked against RSA, TLS, and HTTPS protocols. The encryption efficiency of our method outperforms RSA and TLS, which are widely deployed for text-based security. Because RSA relies on asymmetric cryptography, it exhibits the longest processing latency. The TLS protocol limits asymmetric operations to the initial handshake phase and recycles the symmetric channel for subsequent data transmission, positioning its performance between RSA and our approach. HTTPS mirrors the performance curve of TLS, given that its security architecture is built directly on the TLS protocol. Conversely, our framework demonstrates the lowest overall latency; this operational efficiency is crucial when processing large-scale imagery on resource-constrained devices.

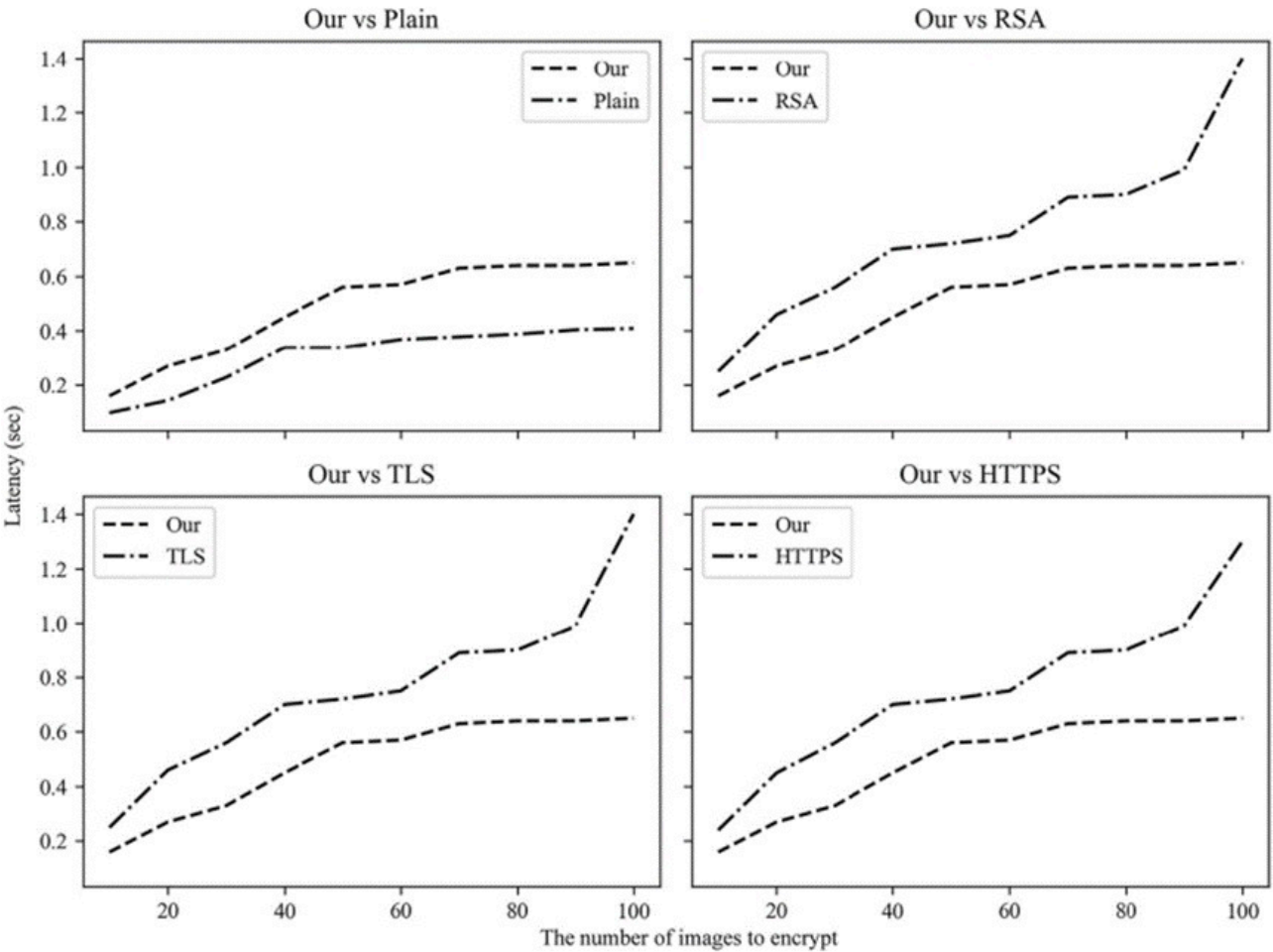


Figure 3. Comparison of latency among our and classic methods for encrypting different numbers of remote sensing images.

To evaluate the performance of the model when fine-tuned via transfer learning from extensive datasets, ResNet18 [41] is implemented as the core backbone architecture. Network adaptation to the remote sensing classification task can be executed via two distinct pathways. The first strategy entails initializing and training the network architecture entirely from scratch; while this pathway can yield high recognition capabilities, it demands substantial computational time. The second pathway involves freezing the weights of the lower structural layers while training the upper layers using an optimization function with weight decay. The underlying ResNet architecture already possesses feature representations of generic imagery—including remote sensing features—mined from the ImageNet dataset. Consequently, we utilize the lower layers of ResNet as a static feature extractor and optimize the higher-level layers using remote sensing data. As illustrated by the ROC curves in Figure 4, the number of frozen layers serves as a tunable parameter; setting this parameter below or above a threshold of six layers causes a notable degradation in classification performance.

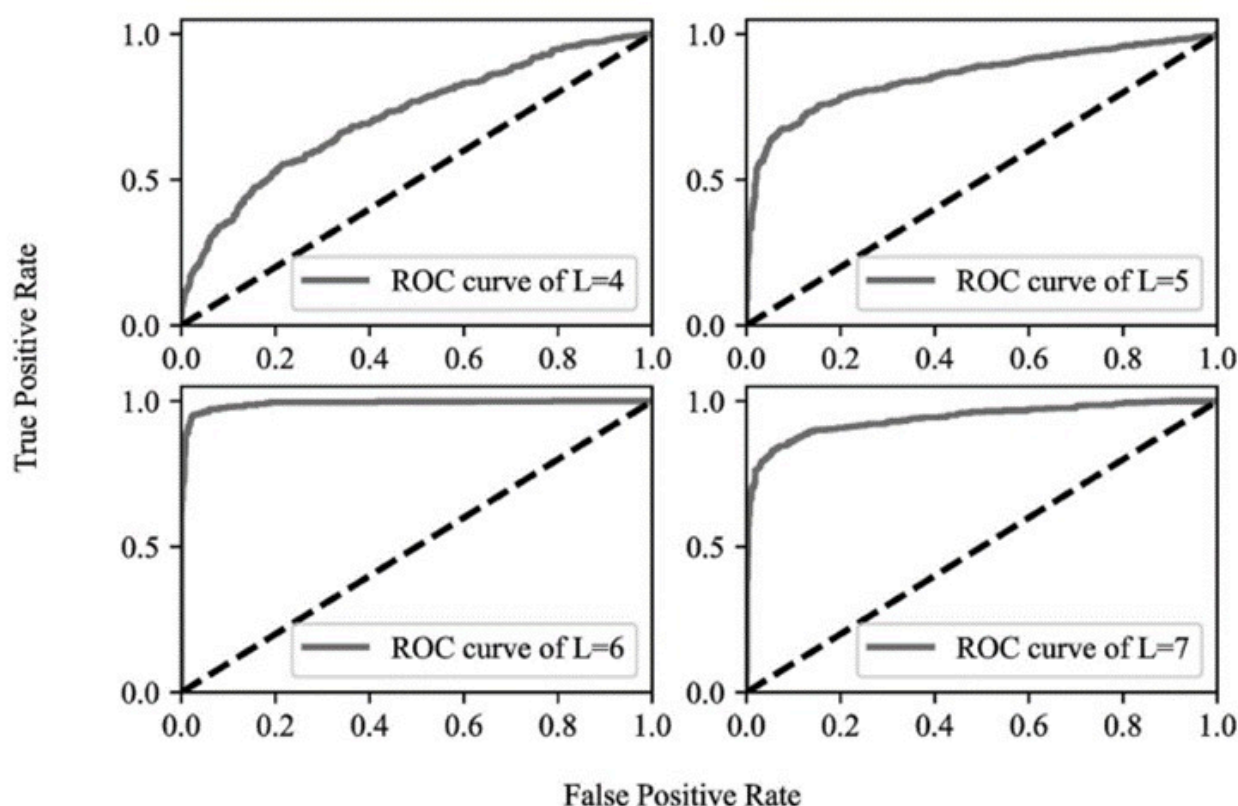


Figure 4. The ROC curves of the UCMerced dataset from encrypted images.

Table 3 Details the recognition performance of various neural-network architectures across standard and encrypted datasets. The RSSCN7 [42] dataset contains 400 images per class, whereas the UCMerced dataset is constrained to 100 images per class. Due to this difference, the cumulative evaluation metrics for RSSCN7 are approximately 0.2 higher than those of UCMerced, emphasizing the critical role of data scale in deep learning performance. Model 1 configurations (e.g., Plain1) display performance metrics on unaltered datasets, whereas Model 2 configurations (e.g., Plain2) present performance on encrypted data. Because our privacy-preserving framework introduces a trade-off in image fidelity, a baseline drop in recognition accuracy is observed. However, when integrated with metric learning and transfer learning, the identification performance is restored to 0.95, narrowing the performance gap to just 0.2 relative to the unaltered datasets. The "Plain" designation represents the baseline ResNet model. By constructing a convolutional neural network spanning 16–19 layers, VGGNet [43] demonstrated that increasing network depth systematically minimizes classification error. This strong scalability makes VGGNet highly adaptable for transfer across diverse datasets. Nevertheless,

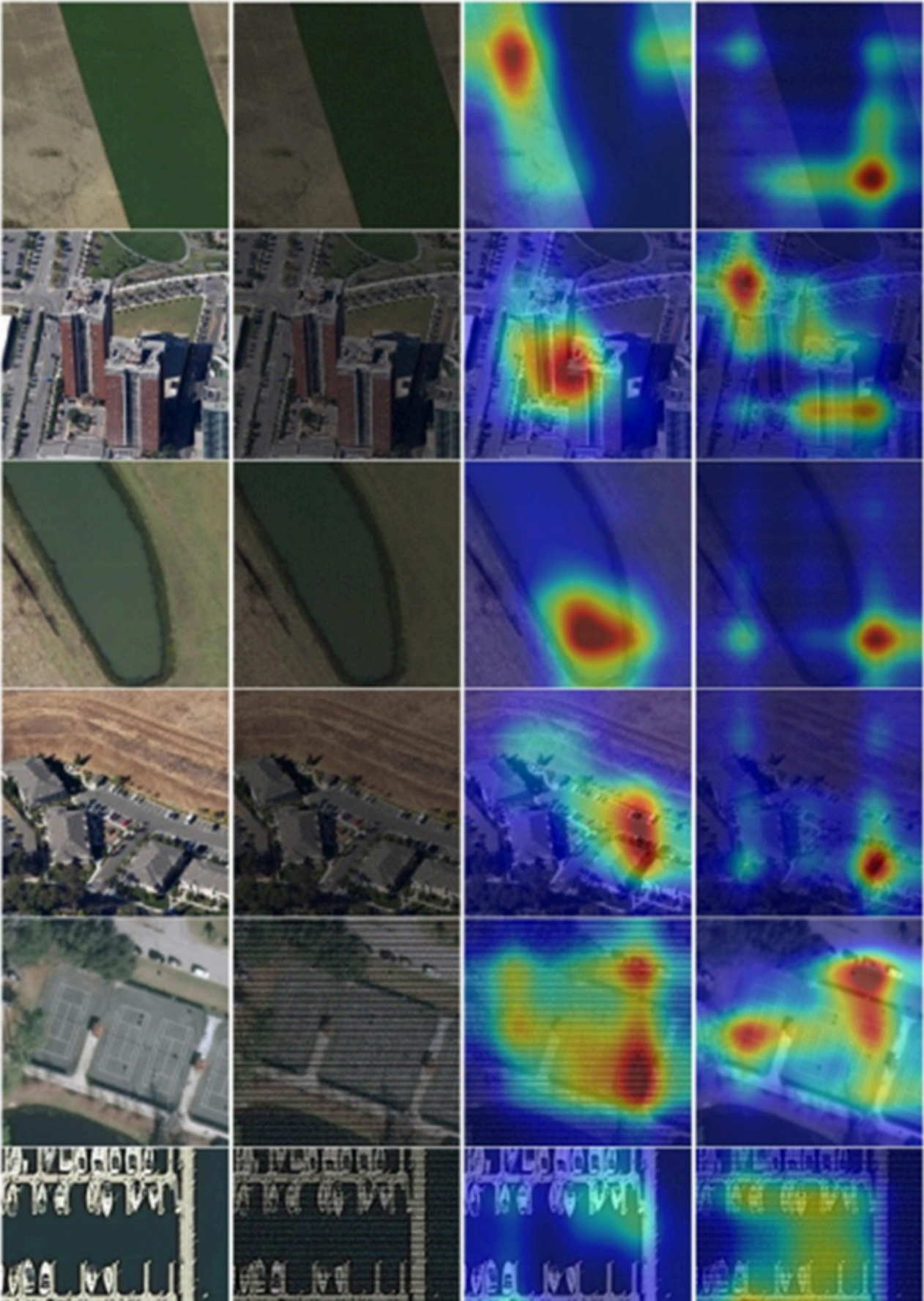
compared to VGGNet, ResNet achieves superior classification performance with a smaller parameter footprint due to its residual connections. Accordingly, ResNet18 is selected as the backbone for the subsequent tuning process. "Fine-tune" identifies models optimized specifically for the target dataset, while "Triplet" denotes models trained utilizing Triplet Loss as the primary objective function. Because the pairing mechanics of the Triplet framework effectively scale up the operational sample volume, this configuration yields the highest recognition performance.

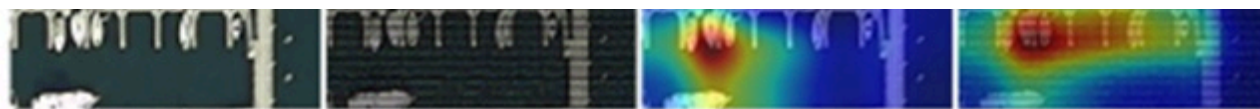
Table 3. Comparison of the recognition performance of different models on the normal (1) and encrypted (2) UCMerced and RSSCN7 datasets, respectively.

Model	RSSCN7 F1	RSSCN7 Recall	RSSCN7 Precision	RSSCN7 Accuracy	UCMerced F1	UCMerced Recall	UCMerced Precision	UCMerced Accuracy
Plain (1)	0.93	0.93	0.95	0.94	0.92	0.93	0.93	0.93
VGGNet (1)	0.91	0.92	0.92	0.91	0.91	0.92	0.91	0.91
Fine- tuned (1)	0.95	0.95	0.95	0.96	0.94	0.95	0.94	0.95
Triplet (1)	0.97	0.98	0.98	0.98	0.97	0.97	0.98	0.98
Plain (2)	0.92	0.91	0.92	0.92	0.90	0.91	0.91	0.91
VGGNet (2)	0.90	0.89	0.90	0.89	0.90	0.89	0.89	0.90
Fine- tuned (2)	0.93	0.93	0.93	0.92	0.91	0.92	0.93	0.92
Triplet (2)	0.95	0.94	0.93	0.96	0.95	0.95	0.96	0.95

Note: The first column is for NN-based recognition models. Plain 1 represents a recognition model trained with normal datasets, while Plain 2 represents a model trained with encrypted datasets.

The recognition performance on encrypted imagery is further demonstrated in [Figure 5](#). The regions of operational attention within the remote-sensing classification network are visualized before and after encryption using the Grad-CAM++ [44] algorithm. Decrypted images exhibit a darker visual tone because blocks containing zero, one, or two black pixels are explicitly omitted to satisfy security conditions.





source data. As a result, the recovered images exhibit global noise artifacts. Crucially, however, this noise does not disrupt the focus regions of the recognition model. It can be inferred that the neural network successfully extracts the correct feature representations across both standard and decrypted samples, and the highlighted focus areas align closely with human visual attention mechanisms. This explains why the fine-tuned model retains strong recognition capabilities despite operating on lossy datasets.

Conclusion

Interlinked devices operating across unified IoT platforms facilitate continuous data transmission. Intelligent sensors can now be integrated into compact, mobile IoT units whose operational cost efficiencies make them optimal for continuous deployment. To prevent the exposure of confidential features—such as coastal topographies, water coloration, and maritime vessels—in remote sensing imagery, these datasets must be encrypted and secured before transmission to public cloud infrastructures. While numerous methodologies have targeted security improvements, many lack practical efficiency.

To resolve these security vulnerabilities during remote-sensing image distribution, this study investigates the viability of Visual Cryptography (VC) for preserving data privacy. Basic VCS establishes an efficient mechanism for decentralized image storage, mitigating the centralized data-breach risks inherent in cloud environments while matching the capabilities of resource-constrained hardware. Conventional image privacy VCS methods frequently compromise these design benefits. By executing pixel-by-pixel encryption and distributing subpixels from a single column of the selection matrices to designated recipients, we realize an optimal VCS that avoids pixel expansion while preserving the computation-free decoding advantage of VC. Furthermore, distributing the error between the encrypted block and the source block to adjacent regions minimizes image degradation. By fine-tuning a pre-trained network using large-scale source datasets, we overcome the performance losses typically associated with training small remote sensing datasets under encryption. The empirical results demonstrate that this lightweight, privacy-preserving framework maintains high recognition accuracy while robustly securing the data.

Future research will focus on exploring data augmentation methodologies to further improve classification accuracy on encrypted remote sensing imagery. While expert-curated augmentation strategies can boost model performance, heuristic approaches often introduce significant variance in final model execution, potentially failing to deliver targeted improvements. Thus, it is essential to examine data subsets at a higher granularity. Secondly, we aim to refine the visual reconstruction quality of the VC-based encryption framework; although the proposed method achieves theoretically optimal contrast, alternative Boolean operations could be adapted to yield more natural decrypted images. Finally, while this paper focuses on remote sensing classification, deep learning is also extensively utilized for semantic segmentation and object detection. These tasks introduce greater complexity than classification, making the demand for lightweight encryption even more pressing. For instance, object detection requires both item classification and bounding box localization. Effectively and securely transmitting non-image data coordinates, such as region proposals and bounding fields, will be our next avenue of research.

Acknowledgements

This work is supported in part by the National Natural Science Foundation of China under Grants (62250410365, 62071084), the Guangdong Basic and Applied Basic Research Foundation of China (2022A1515011542), and the Guangzhou Science and technology program of China (202201010606).

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 10 Apr 2024

Revised: 11 May 2024

Accepted: 30 Jun 2024

Published online: 15 September 2024

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

- Alkhelaiwi M, Boulila W, Ahmad J, Koubaa A, Driss M. An efficient approach based on privacy-preserving deep learning for satellite image classification. *Remote Sens*. 2021;13(11):2221.
<https://doi.org/10.3390/rs13112221>.
- Zhang L, Zhang L, Du B. Deep learning for remote sensing data: a technical tutorial on the state of the art. *IEEE Geosci Remote Sens Mag*. 2016;4(2):22-40.
<https://doi.org/10.1109/MGRS.2016.2540798>.
- Dash A, Ye J, Wang G. A review of generative adversarial networks (GANs) and its applications in a wide variety of disciplines—from medical to remote sensing. *arXiv*. 2021:arXiv:2110.01442.
<https://doi.org/10.48550/arXiv.2110.01442>.
- Shafiq M, Gu Z. Deep residual learning for image recognition: a survey. *Appl Sci (Basel)*. 2022;12(18):8972.
<https://doi.org/10.3390/app12188972>.
- Shafiq M, Gu Z, Cheikhrouhou O, Alhakami W, Hamam H. The rise of “Internet of Things”: review and open research issues related to detection and prevention of IoT-based security attacks. *Wirel Commun Mob Comput*. 2022:8669348.
- Dargan S, Kumar M, Ayyagari MR, Kumar G. A survey of deep learning and its applications: a new paradigm to machine learning. *Arch Comput Methods Eng*. 2020;27(4):1071-92.
<https://doi.org/10.1007/s11831-019-09344-w>.

Wang D, Wu M, He Y, Pang L, Xu Q, Zhang R. An HAP and UAVs collaboration framework for uplink secure rate maximization in NOMA-enabled IoT networks. *Remote Sens.* 2022;14(18):4501.

<https://doi.org/10.3390/rs14184501>.

Fraga-Lamas P, Ramos L, Mondéjar-Guerra V, Fernández-Caramés TM. A review on IoT deep learning UAV systems for autonomous obstacle detection and collision avoidance. *Remote Sens.* 2019;11(18):2144.

<https://doi.org/10.3390/rs11182144>.

Hussain T, Hussain D, Hussain I, AlSalman H, Hussain S, Ullah SS, et al. Internet of Things with deep learning-based face recognition approach for authentication in control medical systems. *Comput Math Methods Med.* 2022;2022:5137513.

Shafiq M, Gu Z, Nazir S, Yadav R. Analyzing IoT attack feature association with threat actors. *Wirel Commun Mob Comput.* 2022;7143054.

Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. TensorFlow: a system for large-scale machine learning. In: *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation.* 2016. p. 265-83.

Xu S, Mu X, Chai D, Zhang X. Remote sensing image scene classification based on generative adversarial networks. *Remote Sens Lett.* 2018;9(7):617-26.

<https://doi.org/10.1080/2150704X.2018.1453173>.

Akhtar N, Mian A. Threat of adversarial attacks on deep learning in computer vision: a survey. *IEEE Access.* 2018;6:14410-30.

<https://doi.org/10.1109/ACCESS.2018.2807385>.

Gu Z, Hu W, Zhang C, Lu H, Yin L, Wang L. Gradient shielding: towards understanding vulnerability of deep neural networks. *IEEE Trans Netw Sci Eng.* 2021;8(2):921-32.

<https://doi.org/10.1109/TNSE.2020.2996738>.

Gu Z, Li H, Khan S, Deng L, Du X, Guizani M. IEPSPB: a cost-efficient image encryption algorithm based on parallel chaotic system for green IoT. *IEEE Trans Green Commun Netw.* 2022;6(1):89-106.

<https://doi.org/10.1109/TGCN.2021.3095707>.

Korala H, Georgakopoulos D, Jayaraman PP, Yavari A. Managing time-sensitive IoT applications via dynamic application task distribution and adaptation. *Remote Sens.* 2021;13(20):4148.

<https://doi.org/10.3390/rs13204148>.

Taheri M, Mozaffari S, Keshavarzi P. Face authentication in encrypted domain based on correlation filters. *Multimed Tools Appl.* 2018;77(13):17043-67.

Thanh TM, Tanaka K. An image zero-watermarking algorithm based on the encryption of visual map feature with watermark information. *Multimed Tools Appl.* 2017;76(11):13455-71.

<https://doi.org/10.1007/s11042-016-3750-2>.

Zhang K, Zuo W, Chen Y, Meng D, Zhang L. Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising. *IEEE Trans Image Process.* 2017;26(7):3142-55.

<https://doi.org/10.1109/TIP.2017.2662206>.

Cheng G, Wang Y, Xu S, Wang H, Xiang S, Pan C. Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network. *IEEE Trans Geosci Remote Sens.* 2017;55(6):3322-37.

<https://doi.org/10.1109/TGRS.2017.2669341>.

Hu X, Yuan Y. Deep-learning-based classification for DTM extraction from ALS point cloud. *Remote Sens.* 2016;8(9):730.

<https://doi.org/10.3390/rs8090730>.

Boulila W, Khazri Khelifi M, Ammar A, Koubaa A, Benjdira B, Farah IR. A hybrid privacy-preserving deep learning approach for object classification in very high-resolution satellite images. *Remote Sens.* 2022;14(18):4631.
<https://doi.org/10.3390/rs14184631>.

He Q, Yu S, Xu H, Liu J, Huang D, Liu G, et al. A weighted threshold secret sharing scheme for remote sensing images based on Chinese remainder theorem. *Comput Mater Contin.* 2019;58(2):349-61.
<https://doi.org/10.32604/cmc.2019.03703>.

Ibrahim DR, Teh JS, Abdullah R. An overview of visual cryptography techniques. *Multimed Tools Appl.* 2021;80(21-23):31927-52.
<https://doi.org/10.1007/s11042-021-11229-9>.

Weir J, Yan W. A comprehensive study of visual cryptography. In: Shi YQ, editor. *Transactions on Data Hiding and Multimedia Security V*. Berlin: Springer Berlin Heidelberg; 2010. p. 70-105.

Naor M, Shamir A. Visual cryptography. In: De Santis A, editor. *Advances in Cryptology—EUROCRYPT'94*. Berlin: Springer Berlin Heidelberg; 1995. p. 1-12.

Zhang D, Zhu H, Liu S, Wei X. A high-quality and printer-friendly visual cryptography scheme. *J Vis Commun Image Represent.* 2021;78:103186.
<https://doi.org/10.1016/j.jvcir.2021.103186>.

Wang L, Yan B, Yang HM, Pan JS. Flip extended visual cryptography for gray-scale and color cover images. *Symmetry (Basel).* 2021;13(1):65.
<https://doi.org/10.3390/sym13010065>.

Lee KH, Chiu PL. An extended visual cryptography algorithm for general access structures. *IEEE Trans Inf Forensics Secur.* 2012;7(1):219-29.
<https://doi.org/10.1109/TIFS.2011.2167611>.

Duda J. Embedding grayscale halftone pictures in QR codes using correction trees. *arXiv.* 2012;arXiv:1211.1572.

Chu HK, Chang CS, Lee RR, Mitra NJ. Halftone QR codes. *ACM Trans Graph.* 2013;32(6):1-8.
<https://doi.org/10.1145/2508363.2508408>.

Raghu M, Zhang C, Kleinberg J, Bengio S. Transfusion: understanding transfer learning for medical imaging. In: *Advances in Neural Information Processing Systems*. 2019;32:3347-57.

Pang H, Xuan Q, Xie M, Liu C, Li Z. Research on target tracking algorithm based on Siamese neural network. *Mob Inf Syst.* 2021;6645629.

Norouzi M, Fleet DJ, Salakhutdinov RR. Hamming distance metric learning. In: *Advances in Neural Information Processing Systems*. 2012. p. 1061-9.

Schroff F, Kalenichenko D, Philbin J. FaceNet: a unified embedding for face recognition and clustering. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston: IEEE; 2015. p. 815-23.

Moutafis P, Leng M, Kakadiaris IA. An overview and empirical comparison of distance metric learning methods. *IEEE Trans Cybern.* 2017;47(3):612-25.
<https://doi.org/10.1109/TCYB.2016.2521767>.

Yang Y, Newsam S. Bag-of-visual-words and spatial extensions for land-use classification. In: *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*. New York: ACM; 2010. p. 270-9.

Hore A, Ziou D. Image quality metrics: PSNR vs. SSIM. In: 2010 20th International Conference on Pattern Recognition. Istanbul: IEEE; 2010. p. 2366-9.

Ren L. A novel raster map exchange scheme based on visual cryptography. *Adv Multimed*. 2021:3287774.

Kang I, Arce GR, Lee HK. Color extended visual cryptography using error diffusion. *IEEE Trans Image Process*. 2011;20(1):132-45.
<https://doi.org/10.1109/TIP.2010.2056376>.

Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, Inception-ResNet and the impact of residual connections on learning. In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. San Francisco: AAAI Press; 2017. p. 4278-84.

Zou Q, Ni L, Zhang T, Wang Q. Deep learning based feature selection for remote sensing scene classification. *IEEE Geosci Remote Sens Lett*. 2015;12(11):2321-5.
<https://doi.org/10.1109/LGRS.2015.2475299>.

Liu S, Deng W. Very deep convolutional neural network based image classification using small training sample size. In: 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR). 2015. p. 730-4.

Chattopadhyay A, Sarkar A, Howlader P, Balasubramanian VN. Grad-CAM++: improved visual explanations for deep convolutional networks. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). 2018. p. 839-47.
<https://doi.org/10.1109/WACV.2018.00097>.