Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Supplementary materials for

Ziyi ZHOU, Chengyue WANG, Kexun YAN, Hui SHI, Xin PANG, 2024. Reversible data hiding in encrypted images based on additive secret sharing and additive joint coding using an intelligent predictor. *Front Inform Technol Electron Eng*, 25(9):1250-1265. https://doi.org/10.1631/FITEE.2300750

1 Parameters of the prediction module

The input feature map has dimensions of C_in × H_in × W_in, where C_in represents the number of input channels, and H_in and W_in represent the height and width of the input feature map, respectively. In the residual block with three convolutional layers, the first convolutional layer utilizes C_out1 convolutional kernels of size 3×3 with padding of 2 and a stride of 1, where C_out1 is the number of output channels for the first convolutional layer. Then, the LeakyReLU activation function is applied. The second convolutional layer uses C_out2 convolutional kernels of size 3×3, where C_out2 is the number of output channels for the second convolutional layer, with padding of 2 and a stride of 1. The third convolutional layer serves as a skip connection layer and introduces an additional 1×1 convolutional layer to adjust the channel number, allowing information to be directly passed to subsequent layers, effectively alleviating optimization issues when training deep networks. The input feature map is transformed using a 1×1 convolution to match the dimensions of the feature map output by the second convolutional layer. The transformed input feature map is then added to the output of convolutional layer 3 and is subjected to the LeakyReLU activation function. Next, the input is passed to a residual block with two convolutional layers. Its structure is similar to the residual block with three convolutional layers, but does not include the skip connection layer (convolutional layer 3). The purpose of the residual group blocks is to handle features in a more in-depth manner by stacking two residual blocks with different convolutional layers and capturing higher-level image information.

2 Threshold derivation

In order to achieve higher compression results and free up more space, threshold control is used to select encoding methods. The threshold is derived as follows, and important symbols are represented in Table S1.

Assumptions:

Number of differing bits recorded in bit-plane comparison encoding: $db\_num = 3$.

Number of bits used in prediction error encoding: $bit\_pred = 4$.

Premises:

Pixel values range from 0 to 255.

Maximum number of bits used in bit-plane comparison encoding: $bit\_cmp\_max = 24$.

Maximum number of bits used in prediction error encoding: $bit\_pred\_max = 8$ (7 bits for error value, 1 bit for sign).

Compression Effectiveness Comparison:

Length of encoding produced by bit-plane comparison: $len\_cmp = db\_num + db \times 3$.

Length of encoding produced by prediction error: $len\_pred = bit\_pred \times 3$.

Threshold Derivation:

Assuming that prediction error encoding is used when the prediction error for each pixel is less than T, and bit-plane comparison encoding is used when the number of bits used for bit-plane comparison encoding is *bit_cmp*.

If the number of bits used in bit-plane comparison encoding is greater than the number of bits used in prediction error encoding—i.e., $db\_num + db \times 3 > bit\_pred \times 3$—we choose prediction error encoding. If the number of bits used in bit-plane comparison encoding is less than or equal to the number of bits used in prediction error encoding—i.e., $db\_num + db \times 3 <= bit\_pred \times 3$—we choose bit-plane comparison encoding.

Therefore, we need to find the minimum *bit_cmp* value that satisfies $db\_num + db \times 3 > bit\_pred \times 3$.

Solving the Inequality:

$db\_num + db \times 3 > bit\_pred \times 3$

$(1 + db) > bit\_pred$

$1 > bit\_pred - db$

$bit\_pred - 1 < db.$

Maximum bit_cmp Value:

The maximum value for db is $bit\_pred - 1$ because $db\_num = 3$. Therefore, the maximum *bit_cmp* value is $(bit\_pred - 1) \times 3 + 3 = 12$. Considering $bit\_pred\_max = 8$, $bit\_pred = 4$, and removing the sign bit, the maximum value for *max*e is (111)2 = (7)10.

Results:

The derived threshold is: $T = 12$, $\tau = 7$.

When $maxe \le \tau \&\& (3 + 3 \times maxd) > T$, the encoding method based on the prediction error is used.

When $maxe \le \tau \&\& (3 + 3 \times maxd) \le T$ P$maxd \le \tau \&\& maxe > \tau$, the encoding method based on the bit-plane comparison is used.

Table S1 Symbol representation

Representation	Symbol
Block pixel representation of the cover image	R, spp_1, spp_2, spp_3
Block pixel representation of the original image	R, sp_1, sp_2, sp_3
Highest differing bit in bit-plane comparison	*maxd*
Maximum prediction error within a block	*maxe*
Threshold for prediction error	T
Threshold for pixel values	τ
Number of bits used in bit-plane comparison encoding	*bit_cmp*
Number of differing bits recorded in bit-plane comparison encoding	*db_num*
Number of bits required to record differing bits in bit-plane comparison encoding	*db*
Number of bits used in prediction error encoding	*bit_pred*
Prediction error value	e_φ
The original image	C
The cross, circle, triangle, square sets	C_1, C_2, C_3, C_4
The cover images representing the true values of the preserved sets C_1, C_2, C_3 and C_4	C_{cross}, C_{circle}, $C_{triangle}$, C_{square}
The shared encrypted cover image generated from the original image	C_share_n
non-overlapping blocks obtained by shuffling the original image C and the cover image (C_{cross}, C_{circle}, $C_{triangle}$, C_{square})	C', C'_{cross}, C'_{circle}, $C'_{triangle}$, C'_{square}
The shared encrypted cover generated from predicted images C'_{cross}, C'_{circle}, $C'_{triangle}$, C'_{square}	$cross\_share_n$, $circle\_share_n$, $triangle\_share_n$, $square\_share_n$

3 Correlation coefficient of the shared encrypted images

Table S2 Vertical correlation coefficient (cor_v), horizontal correlation coefficient (cor_h), and diagonal correlation coefficient (cor_d) of the shared encrypted images preserving true values of the cross set

		Airplane	*Baboon*	*Barbara*	*Boat*	*Jetplane*	*Lena*	*Pepper*	*Man*	*Tiffany*
original image	*cor_v*	0.9463	0.8634	0.8933	0.9454	0.9668	0.9736	0.9808	0.9448	0.9370
	cor_h	0.9458	0.7545	0.9583	0.9748	0.9646	0.9868	0.9831	0.9558	0.9578
	cor_d	0.8961	0.7195	0.8815	0.9264	0.9371	0.9609	0.9657	0.9194	0.9141
$cross\_share_1'$	*cor_v*	0.3751	0.5114	0.5275	0.5009	0.3526	0.5126	0.5196	0.5835	0.4914
	cor_h	-0.3340	-0.4198	-0.4178	-0.3974	-0.3026	-0.4110	-0.4165	-0.4678	-0.4025
	cor_d	-0.3374	-0.4209	-0.4214	-0.4026	-0.3100	-0.4164	-0.4195	-0.4704	-0.4065
$cross\_share_2'$	*cor_v*	0.5910	0.6332	0.6379	0.6280	0.5795	0.6329	0.6346	0.6556	0.6267
	cor_h	-0.4770	-0.5136	-0.5149	-0.5013	-0.4601	-0.5104	-0.5132	-0.5351	-0.5064
	cor_d	-0.4803	-0.5170	-0.5191	-0.5060	-0.4651	-0.5142	-0.5168	-0.5390	-0.5091
$cross\_share_3'$	*cor_v*	-0.1867	-0.3148	-0.0089	0.1097	-0.1222	0.0942	0.0504	-0.1830	-0.2294
	cor_h	-0.2220	-0.0323	-0.0735	-0.1484	-0.1988	-0.2093	-0.1790	0.0157	-0.0792
	cor_d	-0.1555	-0.0573	-0.1277	-0.1965	-0.2190	-0.2345	-0.2327	-0.1786	-0.1145

4 Training specifics

The proposed intelligent predictors are trained on a computational setup consisting of an Intel Core i5-12400F CPU clocked at 2.5 GHz and an NVIDIA 3060 GPU. To construct the training dataset, we randomly selected 1000 images from the widely used ImageNet dataset. All images were converted to grayscale and sized to dimensions of 512×512 pixels. We applied the preprocessing method described in Section 3.1 to the input images of all prediction models. This preprocessing step augmented the training dataset to include a total of 4000 images.

For optimization, we employed the backpropagation technique along with the Adam optimizer. These methods allowed us to iteratively refine the models' performance. We performed several training iterations to enhance the predictive capabilities of the proposed ResNet-based predictor. The training process involved adjusting the model's internal parameters to minimize the prediction errors, utilizing the training dataset. The computational resources of the Intel Core i5-12400F CPU and NVIDIA 3060 GPU enabled efficient training of the models, leveraging their respective capabilities.

By adhering to these training procedures and employing powerful computational resources, we were able to develop and optimize the proposed ResNet-based predictor. These models are designed to achieve superior performance in pixel prediction tasks, leveraging the spatial attention mechanism. The training process ensured that the models were well-suited for the intended objectives and were capable of delivering accurate predictions for various image inputs. These advancements contribute to the field of image compression and demonstrate the potential for further improvements in this area.

The definition of training loss for a batch size of 4 is as follows:

$$Loss = \frac{1}{n}\sum_{i=1}^{n}(I_i^{out} - I_i^{label}). \tag{S1}$$

5 Conclusions and future directions

We present a novel approach for reversible data hiding in encrypted data, integrating intelligent prediction and additive secret sharing. Our proposed method encompasses several key components, including the training of an intelligent predictor, encrypted predictions, additive encryption, and joint encoding for embedding. The

method exhibits noteworthy advantages, such as efficient hiding, robust security, adaptive joint encoding, and lossless recovery.

Our method offers several advantages that contribute to its effectiveness: (1) Efficient Hiding. The incorporation of an intelligent predictor in the data hiding process enhances the efficiency of concealing information within encrypted data. The predictor ensures precise predictions, optimizing the hiding mechanism. (2) Robust Security. Utilizing additive secret sharing for encryption introduces a robust security layer. The requirement for multiple shares for image reconstruction enhances the overall security of the encrypted data, providing a balanced trade-off between security and efficiency. (3) Adaptive Joint Encoding. The adaptive joint encoding technique employed during the embedding process maximizes capacity while minimizing any potential degradation in image quality. This adaptability is crucial for achieving efficient data hiding in various scenarios. (4) Lossless Recovery. A key feature of our method is the capability for lossless recovery, even in scenarios where certain shares are lost. The intelligent predictor plays a pivotal role in accurately filling missing shares during the recovery process, ensuring the integrity of the original data.

While our proposed method showcases promising features, there are avenues for future improvements and exploration: (1) Predictor Enhancement. To further enhance the accuracy of predictions and embedding, future work will focus on improving the intelligent predictor. This may involve refining the neural network architecture, incorporating advanced training techniques, and exploring additional features for improved prediction capabilities. (2) Enhanced Embedding Strategies. Research efforts will be directed towards optimizing the embedding process to achieve increased capacity without compromising image quality. Exploring advanced joint encoding methods and considering alternative data hiding strategies will be part of this endeavor. (3) Robustness Evaluation. A comprehensive evaluation of the method's robustness in diverse scenarios, including different image types, encryption strengths, and prediction challenges, will be conducted. This will provide insights into the versatility and reliability of the proposed approach. (4) Security Analysis. Future work will involve a more extensive security analysis, including vulnerability assessments and testing against potential attacks. This rigorous examination will further validate the robustness and reliability of our reversible data hiding scheme. (5) Real-world Applications. Exploring real-world applications and assessing factors such as scalability, computational efficiency, and applicability to diverse use cases will be crucial for the practical implementation of the proposed method. This includes considering the method's suitability for various scenarios and its potential impact on real-world settings.

In conclusion, our proposed method signifies a significant stride towards efficient and secure reversible data hiding in encrypted data. The outlined future directions aim to refine and extend the capabilities of the method, making valuable contributions to the field. We anticipate that the ongoing enhancements and explorations will further solidify the method's effectiveness, ensuring its relevance and impact in the dynamic landscape of reversible data hiding.