# Enhancing the steganographic resistance of hidden information to active attacks

Yurii Yaremchuk[1,*,†], Olha Saliieva[1,†], Vasyl Karpinets[1,†], Andrii Nikolaienko[1,†] and Nataliia Kunanets[2,†]

[1] *Vinnytsia National Technical University, 95 Khmelnytsky Highway str., 21000 Vinnytsia, Ukraine*
[2] *Lviv Polytechnic National University, 12 Bandera str., 79013 Lviv, Ukraine*

## Abstract

With the rapid development of modern information technologies and systems, the number and complexity of threats aimed at breaching the security of confidential information transmitted secretly in multimedia files is increasing. At the same time, active attacks aimed at steganographic information protection systems are of great importance, as they differ from passive attacks in that the attacker not only tries to detect hidden information but can also modify the image to remove or distort it. Given the importance of this issue, the paper proposes to increase the resistance of hidden information in images to color gamut-changing attacks. To achieve this goal, the hiding algorithm has been improved based on the use of matrix filters to find the most suitable image areas and hide bits of information in them by correlating the average brightness of segments in blocks. The proposed algorithm includes many variable parameters, which makes it possible to adapt it to a wide range of images. In addition, flexible parameter settings enable improved information embedding and extraction accuracy. To validate the effectiveness of the enhanced algorithm, color-change attacks were conducted on an image containing embedded hidden information. Various steganalysis methods were also applied, including visual steganalysis, which reveals the least significant bit of the image; RS steganalysis, which estimates the approximate size of the hidden data; and a method analyzing the distribution of image elements on a plane to detect patterns, structures, or anomalies within the image. The obtained results demonstrate the robustness of the proposed method of information hiding to the selected steganalysis algorithms and to color gamut-changing attacks.

## Keywords

steganography, concealment of information, steganalysis, active attack, color gamut



## 1. Introduction

With the rapid advancement of technology and communications, it is becoming increasingly important to protect confidential information from unauthorized access. After all, transmitting a secret message over an unprotected network channel poses a serious threat to its security. Among the methods for solving this problem are steganographic methods, which involve the open transport of a secret message embedded in a container in advance [1]. However, there are many attacks aimed at breaking into steganographic systems, including attacks that focus on changing the color gamut in images, which can lead to the disclosure of hidden information, violation of its confidentiality and integrity. Therefore, this paper presents a study related to increasing the resistance of hidden information in images to these attacks. Thus, in this paper, based on the analysis of known steganographic methods, it is necessary to improve the algorithm for hiding information using matrix filters and test it to prove its resistance to steganalysis and color change attacks.

In recent years, numerous studies have been conducted on methods of hiding data in images based on the use of algorithms for finding acceptable zones for hiding information. For example, in their study [2], the authors developed the Bald Eagle Search Optimal Pixel Selection with Chaotic Encryption (BESOPS-CE) steganographic image method based on the method of searching for optimal pixels with chaotic encryption. The presented method effectively hides the secret image in encrypted form under the cover image. In this paper [3], a steganographic scheme based on graph wavelet transformation using graph signal processing (GSP) is proposed, which improves the visual quality of a stego image. The authors of the paper [4] presented a new steganographic approach in which the algorithm of an extreme learning machine is modified to create a supervised mathematical model. This algorithm is first trained on a part of the image and then tested in regression mode, which allows choosing the optimal place

*CPITS-II 2024: Workshop on Cybersecurity Providing in Information and Telecommunication Systems II, October 26, 2024, Kyiv, Ukraine*
* Corresponding author.
† These authors contributed equally.
yurevyar@vntu.net (Y. Yaremchuk);
salieva8257@gmail.com (O. Saliieva);
karpinets@gmail.com (V. Karpinets);
andrey.nikolaienko.0@gmail.com (A. Nikolaienko);
nek.lviv@gmail.com (N. Kunanets)

0000-0002-6303-7703 (Y. Yaremchuk);
0000-0003-2388-7321 (O. Saliieva);
0000-0001-8148-2002 (V. Karpinets);
0009-0007-0178-9444 (A. Nikolaienko);
0000-0003-3007-2462 (N. Kunanets)

to embed a message with the best values of the predicted evaluation metrics. In this research [5], a new method was proposed to increase the possibility of embedding secret information in an image based on the edge region. The new approach combines the Kenny and Pruitt edge detection methods using a binary operation, and the secret message is hidden using the least significant bit (LSB) method.

An interesting idea is that of the authors of that paper [6], who proposed an algorithm for encrypting private images in combination with a new tent multi-dynamic piecewise connected mapping lattice (TMDPCML) for spatiotemporal chaotic systems. This algorithm extracts private image information and uses distributed nonlinear diffusion. Study [7] presents a method for detecting steganographic changes in images using convolutional neural networks. In [8], a new stenographic swarm optimization technique with encryption for digital image protection, called CSOES-DIS, was developed. The proposed model applies the method of double chaotic digital image encryption, and then the embedding process is implemented. The authors of this paper [9] proposed a new algorithm for encrypting and hiding triple images by combining a 2D chaotic system, compressive sensing (CS), and 3D discrete cosine transform (DCT).

Data hiding methods that use areas of rapid changes in an image to hide information have many advantages, including high capacity, adaptability to different types of images and data formats, and the ability to use both the spatial and frequency characteristics of an image. However, these methods are not very resistant to color gamut attacks, which limits their effectiveness. In this regard, the paper proposes an improvement that uses luminance correlation and matrix filters to effectively hide information in images while providing a high level of protection.

## 2. Main body

Most algorithms that use areas of rapid changes in the image to hide information use the method of hiding information in one or two last bits of pixels of different channels. These methods are called LSB and 2-LSB according to the number of least significant bits [10]. However, these algorithms cannot withstand color gamut attacks, so it is worth improving an algorithm that will use luminance correlation to embed bits of information in the image, which will help to withstand color gamut attacks by using luminance in areas that are most difficult to change.

To increase the resistance to the above attacks, it is proposed to use matrix filters to detect areas of sharp transitions, which are expressed in the red and green channels, while information is hidden in the blue channel.

The proposed algorithm applies a filter using the discrete Laplace operator [11], which can be represented in two forms:

$$\nabla^2 f = f(i+1,j) + f(i-1,j) + \\ +f(i,j+1) + f(i,j-1) - 4 * f(i,j), \quad (1)$$

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad (2)$$

where $i, j$ are indices in the cell of the selected pixel to determine $\nabla^2 f$.

The next stage of the algorithm is to divide the resulting matrix $M$ and image $P$ into blocks of size $n$ x $n$, with the number of blocks being $r$. For each block of the matrix $M$ a set of pixel indices is defined by $N_i$, exceeding the threshold value. This process begins by calculating the average value between the blocks:

$$h_i(x,y) = \frac{rm_i(x,y) + gm_i(x,y)}{2}, \\ i = 1 \dots r, x = 1 \dots n, y = 1 \dots n, \quad (3)$$

where $h_i$ is $i^{\text{th}}$ block with the average value between the pixels of the blocks; $rm_i$ is $i^{\text{th}}$ block from the red channel of the matrix $M$; $gm_i$ is $i^{\text{th}}$ block from the green channel of the matrix $M$; $n$ is the size of the block.

Next, the set of indices is defined as follows $N_i$:

$$N_i = \{(x,y) | h_i(x,y) > T\}, \quad (4)$$

where $T$ is threshold value.

$N_i = \{(x_1, y_1), (x_2, y_2), \dots, (x_k, y_k)\}, k = 1 \dots n_i$;

$n_i$ is the length of the $i^{\text{th}}$ variable $N_i$.

The next step is to split $N_i$ into indices corresponding to the lower and upper parts of the block:

$$D_i = \{N_i[j] | N_i[j][1] > \frac{n}{2}, j = 1 \dots n_i\}, \quad (5)$$

$$T_i = \{N_i[j] | N_i[j][1] \leq \frac{n}{2}, j = 1 \dots n_i\}, \quad (6)$$

where $D_i = \{(x_1, y_1), (x_2, y_2), \dots, (x_k, y_k)\}, k = 1 \dots d_i$—is the set of lowercase indices of the part of the block from $N_i$; $T_i = \{(x_1, y_1), (x_2, y_2), \dots, (x_k, y_k)\}, k = 1 \dots t_i$ is the set of upper indices of the part of the block from $N_i$; $N_i[j]$ is value $(x_j, y_j)$ from $i^{\text{th}}$ block, and $N_i[j]$ [1] is value of $x_j$ from the set of indices from $N_i$ $i^{\text{th}}$ block.

The last stage of preparation before embedding the information is to calculate the average brightness of the pixels in the blue channel of the image $P$ for the pixels with the lower and upper block indices:

$$pd_i = \frac{\sum_{j=0}^{d_i} bp_i(D_i[j])}{d_i}, \quad (7)$$

$$pt_i = \frac{\sum_{j=0}^{t_i} bp_i(T_i[j])}{t_i}, \quad (8)$$

where $pd_i$ is the average brightness of the lower part of the *i-th* block by indices $D_i$; $pt_i$ is the average brightness of the upper part of the $i^{\text{th}}$ block by indices $T_i$; $D_i[j]$ is indices $(x_j, y_j)$ of the lower part of the $i^{\text{th}}$ block; $T_i[j]$ is indices $(x_j, y_j)$ the upper part of the $i^{\text{th}}$ block; $bp_i$ is a block from the blue image channel $P$, accordingly $bp_i(D_i[j])$ and $bp_i(T_i[j])$ are pixels from the top and bottom part of *i-th* block for the specified indices $T_i[j]$ and $D_i[j]$.

Thus, the preparation stage for hiding information in the image takes place.

Next, the embedding process is examined. Assuming a sequence of bits $W$ exists, to hide one bit of information, the average brightness of the pixels at the calculated indices in the lower and upper parts of the blue channel block of the image must satisfy the conditions for embedding the bit. Moreover, the smaller the embedding dimension, the more intensively the pixel brightness will change, and vice versa.

Example of conditions for hiding a single bit:

$$if\ w_i = 1, \left((0 * \delta) + \varepsilon \leq pd_i \leq \left((0+1) * \delta\right) - \varepsilon\right) \vee, \\ \left((2 * \delta) + \varepsilon \leq pd_i \leq \left((2+1) * \delta\right) - \varepsilon\right) \vee \dots \vee,$$

$\left((c * \delta) + \varepsilon \le pd_i \le ((c+1) * \delta) - \varepsilon\right) \wedge,$

$\left((1 * \delta) + \varepsilon \le pt_i \le ((1+1) * \delta) - \varepsilon\right) \vee,$

$\left((3 * \delta) + \varepsilon \le pt_i \le ((3+1) * \delta) - \varepsilon\right) \vee \ldots \vee,$

$\left((v * \delta) + \varepsilon \le pt_i \le ((v+1) * \delta) - \varepsilon\right),$

$v = 1, 3, 5 \ldots q-1, c = 0, 2, 4 \ldots q-2, \delta = \frac{256}{q},$

where $q$ is embedding dimension; $\varepsilon$ is brightness change coefficient ($0 < \varepsilon < \frac{\delta}{2}$).

At the same time, the conditions for hiding the zero bit will be as follows:

$if\ w_i = 0, \left((0 * \delta) + \varepsilon \le pt_i \le ((0+1) * \delta) - \varepsilon\right) \vee,$

$\left((2 * \delta) + \varepsilon \le pt_i \le ((2+1) * \delta) - \varepsilon\right) \vee \ldots \vee,$

$\left((c * \delta) + \varepsilon \le pt_i \le ((c+1) * \delta) - \varepsilon\right) \wedge,$

$\left((1 * \delta) + \varepsilon \le pd_i \le ((1+1) * \delta) - \varepsilon\right) \vee,$

$\left((3 * \delta) + \varepsilon \le pd_i \le ((3+1) * \delta) - \varepsilon\right) \vee \ldots \vee,$

$\left((v * \delta) + \varepsilon \le pd_i \le ((v+1) * \delta) - \varepsilon\right),$

$v = 1, 3, 5 \ldots q-1, c = 0, 2, 4 \ldots q-2, \delta = \frac{256}{q}.$

If the condition for hiding a bit of information is not met, the total brightness is correlated by increasing or decreasing by a unit block of pixels whose indices belong to the lower or upper threshold of the blue channel block. This happens until the conditions for hiding the selected bit of information are met. This is how information in the image is hidden.

We will develop algorithms that implement an improved method of hiding information.

Fig. 1 shows a general algorithm for embedding one bit of information in one block.

After determining the luminance values, a check is performed to determine whether the bit has been embedded. If this condition is not met, then the correlation value for the lower and upper parts of the block is calculated, then the brightness for these parts of the block is correlated and the brightness value is calculated. This process is repeated cyclically until the desired values are selected and the condition is fully met.

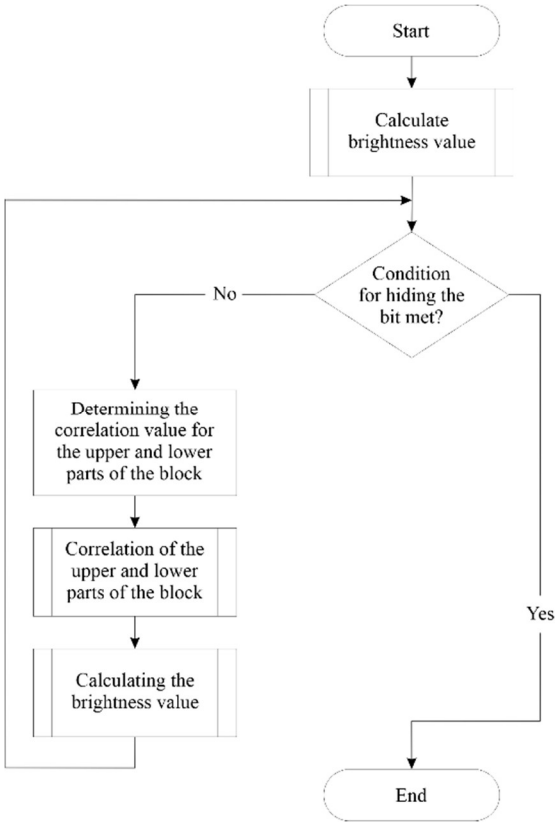


**Figure 1:** Flowchart with a general algorithm for hiding one bit per block

At the same time, the general algorithm for extracting a bit of information from a block is shown in Fig. 2.

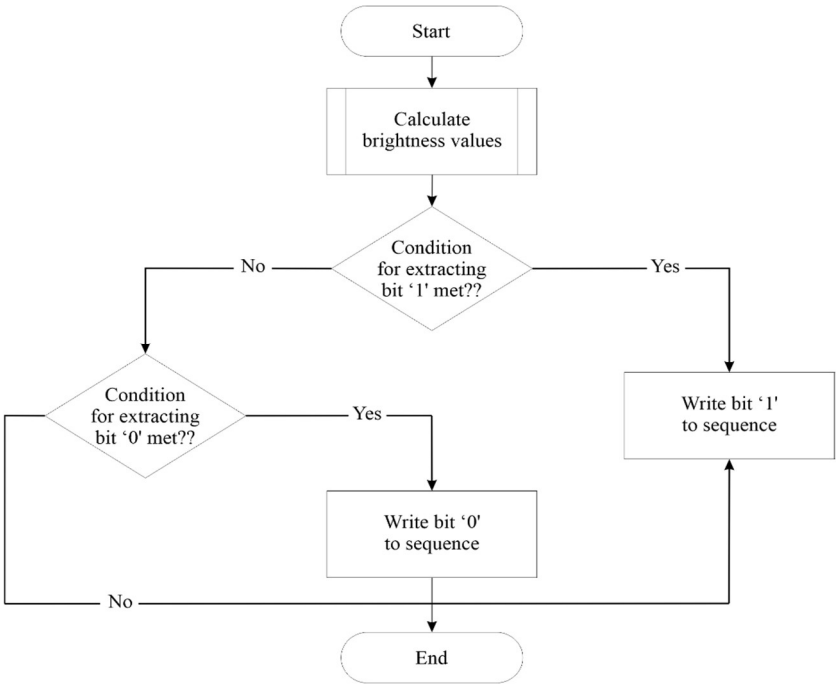


**Figure 2:** Flowchart with a general algorithm for extracting a bit of information

## 3. Experimental research

The security of the proposed algorithm was analyzed by performing color change attacks on the image in which the hidden information was embedded.

To determine the effectiveness of the algorithm's protection against such attacks, the values of $m^p$, are calculated, reflecting the correspondence between the embedded hidden information $w$ and the extracted hidden information $w*$ obtained:

$$m^p = \frac{n^q}{M} 100\%, \quad (9)$$

where $n^q$ is the number of equal bits in both messages; $M$ is the size of the hidden message in bits.

Value of $m^p$ indicates the success of extracting hidden information from the image.

More than 100 different images were selected for the experiment, most of which are detailed full-color images with a sufficient number of contrast transitions and monotonous areas.

Fig. 3 shows an example of one of the images after attacks that altered the color scheme: making it brighter, more contrasted, and darker, closer to grayscale.

**Figure 3:** Images with hidden information after attacks: change of color scheme to brighter, more contrasty, and darker, to shades of grey

Let's examine the impact of color gamut attacks on images with an embedded message (Table 1).

**Table 1**

The result of calculating the value $m^p$ for an image that is under attack

| Attack | $m^p$ |
|---|---|
| Change the color scheme to a brighter one | 98,223 |
| Change the color scheme to a more contrasting and darker one | 92,164 |
| Change the color scheme to shades of grey | 84,857 |

After analyzing the data obtained, it can be concluded that when the color gamut shifts toward grey, the percentage of lost data in the image increases; however, the algorithm remains quite resistant to this attack.

Next, a visual method will be tested that shows the most significant bit of the image. The test result is shown in Fig. 4, with the results from the original image displayed on the left and the modified image on the right.

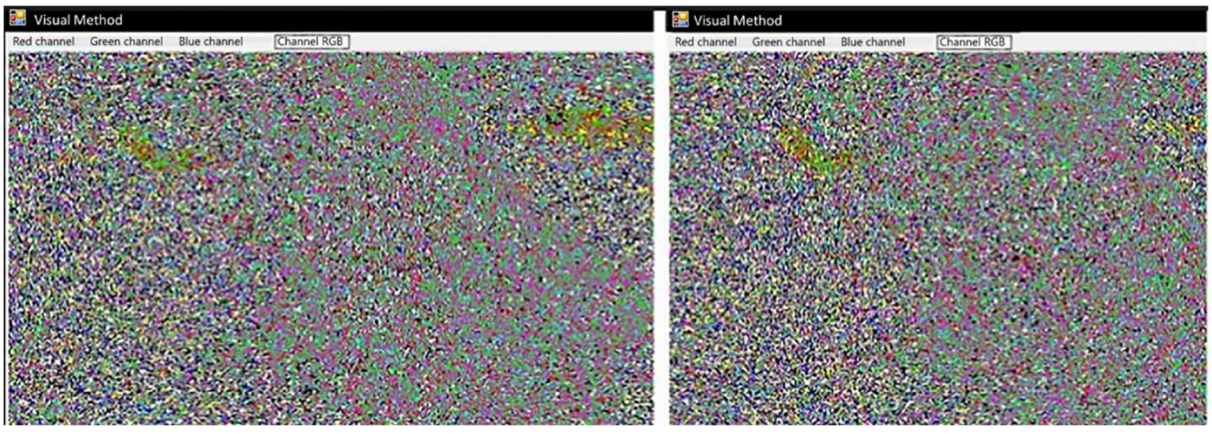


**Figure 4: Comparison** of the least significant bits in the original and hidden images

After analyzing the data, it can be concluded that the least significant bits from the original image differ significantly from those of the container image; however, the container does not exhibit significant distortions or artifacts that would indicate the presence of hidden data.

The following testing will be carried out using the RS steganalysis method, which is effective for detecting the presence of a message in an image and estimating the approximate size of the hidden data. The results of the method on the original image are shown in Fig. 5, and on the modified image—in Fig. 6.


**Figure 5:** The result of steganalysis of the original image using the RS method


**Figure 6:** The result of steganalysis of a modified image using the RS method

Thus, minor changes in different image channels were detected, but these values cannot indicate the presence or absence of hidden information in the container or the original.

Let's conduct additional testing of the algorithm, focusing on the analysis of the distribution of image elements on the plane (Fig. 7).

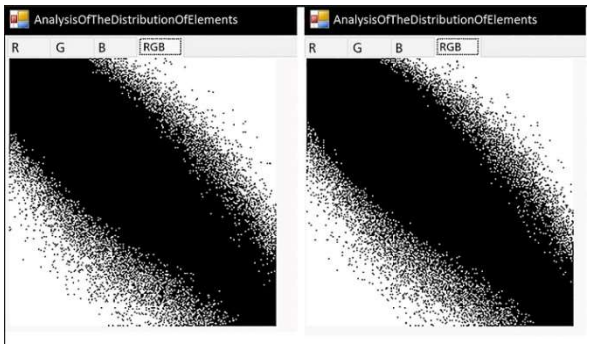


**Figure 7:** The result of steganalysis using the method of analyzing the distribution of image elements on the plane

If the points are randomly distributed over the entire area, the sequence elements are independent, which is typical for containers with embedded data. If the container is not filled, then an uneven distribution of points on the field will be observed. Since the container did not show a chaotic arrangement of points, this method was unable to recognize that the container contained hidden information.

Thus, according to the results of the study, it can be assumed that the proposed method of hiding information is resistant to the described steganalysis algorithms and to color change attacks.

## 4. Conclusions

In this work, the information-hiding algorithm is improved by using matrix filters to identify the most suitable areas in the image and embed information bits into them. This is achieved by varying the brightness of segments in image blocks, which are consistently changed to hide bits of information. The main feature of the proposed algorithm is its high resistance to color gamut attacks, achieved by embedding bits by changing the average brightness of the selected hiding zones. Due to the large number of variable parameters, another significant advantage of the improved algorithm can be identified, namely the flexibility of settings, which allows the algorithm to be adapted to different types of images.

In addition, the study presents the algorithm of the program in the form of flowcharts, which is an important step in testing the improved algorithm for resistance to active attacks. The results of the testing indicate high resistance to color gamut attacks on bright colors and above-average resistance to grey gamut attacks. Also, during the testing, resistance to steganalysis by visual methods, RS method, and the method of analyzing the distribution of the image on the plane was revealed.

Thus, based on the proposed improvement of the algorithm, the steganographic resistance of hidden information in images to active attacks was increased.

## References

[1] S. Buchyk, et al., Improvement of Steganographic Methods based on the Analysis of Image Color Models, in: Workshop on Cybersecurity Providing in Information and Telecommunication Systems, CPITS, vol. 2923 (2021) 117–124.

[2] A. Bahaddad, K. Almarhabi, S. Abdel-Khalek, Image Steganography Technique based on Bald Eagle Search Optimal Pixel Selection with Chaotic Encryption, Alexandria Eng. J. 7(5) (2023) 41–54. doi: 10.1016/j.aej.2023.05.051.

[3] V. K. Sharma, D. K. Srivastava, P. Mathur, Efficient Image Steganography using Graph Signal Processing, IET Image Process. 12(6) (2018) 1065–1071. doi: 10.1049/iet-ipr.2017.0965.

[4] H. A. Atee, et. al., Extreme Learning Machine based Optimal Embedding Location Finder for Image Steganography, PLoS ONE 12(2) (2017). doi: 10.1371/journal.pone.0170329.

[5] N. A. Mohsin, H. A. Alameen, A Hybrid Method for Payload Enhancement in Image Steganography Based on Edge Area Detection, Cybernetics and Information Technologies 21(3) (2021) 97–107. doi: 10.2478/cait-2021-0032.

[6] X. Wang, J. Yang, A Privacy Image Encryption Algorithm based on Piecewise Coupled Map Lattice

with Multi Dynamic Coupling Coefficient, Inf. Sci. 569 (2021) 217–240. doi: 10.1016/j.ins.2021.04.013.

[7] D. Denysiuk, et. al., Method for Detecting Steganographic Changes in Images using Machine Learning, in: 13th International Conference on Dependable Systems, Services and Technologies (DESSERT) (2023) 1–6. doi: 10.1109/DESSERT61349.2023.10416453.

[8] A. A. Eshmawi, et al., Competitive Swarm Optimization with Encryption based Steganography for Digital Image Security, Comput. Mater. Contin. 72(2) (2022) 4173–4184. doi: 10.32604/cmc.2022.028008.

[9] X. Wang, C. Liu, D. Jiang, A Novel Triple-Image Encryption and Hiding Algorithm based on Chaos, Compressive Sensing and 3D DCT, Inf. Sci. 574 (2021) 505–527. doi: 10.1016/j.ins.2021.06.032.

[10] J. T. Tarigan, D. I. Putra, Implementation of Steganography Modified Least Significant Bit using the Columnar Transposition Cipher and Caesar Cipher Algorithm in Image Insertion, Journal of Physics: Conference Series 1898(1) (2021). doi: 10.1088/1742-6596/1898/1/012003.

[11] Discrete Laplacian, Medium (2024). URL: https://hilbert-cantor.medium.com/discrete-laplacian-8a5dde7ff001