This paper reports a comparative analysis of accuracy in the detection of steganograms formed according to adaptive steganographic methods, using steganography detectors based on common and specialized types of artificial neural networks. The results of the review of modern convolutional neural networks applied for the tasks of digital image stegoanalysis have established that the accuracy of operating the steganography detectors based on these networks is significantly compromised when processing image packets characterized by a significant variability of statistical parameters.

The performance accuracy of steganography detectors based on the modern statistical model of container images maxSRMd2 has been investigated, as well as on the latest convolutional and «hybrid» artificial neural networks, in particular, GB-Ras and ASSAF networks, when detecting steganograms formed according to the adaptive steganographic methods HUGO and MiPOD. It was established that the use of the statistical model maxSRMd2 makes it possible to significantly (up to 30 %) improve the accuracy of steganogram detection in the case of analyzing those images that are characterized by a high level of natural noise. It was found that the use of the ASSAF network makes it possible to significantly (up to 35 %) reduce an error of steganogram detection compared to current steganography detectors based on the GB-Ras network and the maxSRMd2 statistical model. It was determined that the high accuracy of the ASSAF network-based steganography detector is maintained even in the most difficult case of image processing with high noise and poor filling of the container image with stegodata (less than 10 %).

The results reported here are of theoretical interest for designing high-precision steganography detectors capable of working under conditions of high variability in image parameters

Keywords: stegoanalysis, digital images, convolutional neural networks, autoencoders

UDC 004.[056.5+932.2]
DOI: 10.15587/1729-4061.2022.251350

ANALYZING THE ACCURACY OF DETECTING STEGANOGRAMS FORMED BY ADAPTIVE STEGANOGRAPHIC METHODS WHEN USING ARTIFICIAL NEURAL NETWORKS

Dmytro Progonov
Corresponding author
PhD, Associate Professor*
E-mail: progonov@gmail.com
Mariia Yarysh*
*Department of Information Security
National Technical University of Ukraine
«Igor Sikorsky Kyiv Polytechnic Institute»
Peremohy ave., 37, Kyiv, Ukraine, 03056

Received date 30.11.2021
Accepted date 11.01.2022
Published date 28.02.2022

How to Cite: *Progonov, D., Yarysh, M. (2022). Analyzing the accuracy of detecting steganograms formed by adaptive steganographic methods when using artificial neural networks. Eastern-European Journal of Enterprise Technologies, 1 (9 (115)), 45–55. doi: https://doi.org/10.15587/1729-4061.2022.251350*

1. Introduction

Special attention is paid to ensuring reliable protection of critical information infrastructure (CII) by both state institutions and private organizations. CII elements are widely used for receiving, processing, storing, and transmitting restricted information (RI) in both local and global computing networks [1].

One of the security threats to the RI circulating in the CII computing networks is the unauthorized transfer of confidential data, in particular using steganographic communication systems (SCS) [1]. The peculiarity of SCS is the concealment (embedding) of messages (stegodata) to container files circulating on the network and the subsequent transfer of modified files (steganograms). This makes it possible to overcome the existing systems of counteraction to RI leaks, as well as to form hidden channels of communication between intruders during attacks on the CII of organizations and institutions [1, 2].

A significant number of modern SCSs are based on the use of multimedia data, in particular, digital images (DI) as container files [2]. Particular attention is paid by intruders to the development of adaptive steganographic methods (ASM) aimed at minimizing distortions of statistical and spectral parameters of container images (CCs) when embedding stegodata. This significantly complicates the detection of formed steganograms and requires the use of computationally-complex methods of statistical stegoanalysis.

To improve the accuracy in detecting the steganograms formed according to ASM, it was proposed to use artificial neural networks (ANN) [3]. That has made it possible to slightly reduce the computational complexity of steganography detector (SD) configuration compared to statistical methods of stegoanalysis while maintaining a fixed accuracy of steganogram detection. However, a significant impact on the accuracy of SD operation based on ANN has a priori data on the used steganographic method (a zero-day problem), as well as the variability of statistical and spectral parameters of the processed images (a domain mismatch problem) [4]. Therefore, an important and relevant task is the development of high-precision methods of DI stegoanalysis, capable

of ensuring high accuracy of steganogram detection even under conditions of the limited a priori data on the used steganographic method (SM) and a significant variation in the parameters of the images under study.

Resolving this issue requires a comprehensive study into the effectiveness of current methods of steganogram detection. Despite a significant body of research on the accuracy of SD operation, based on the use of convolutional neural networks, the analysis of the effectiveness of the use of specialized types of ANN is currently not paid enough attention in the literature. Therefore, it is of interest to study the detection accuracy of steganograms formed according to ASM, when using SDs based on both common and specialized types of ANNs.

2. Literature review and problem statement

Current methods of steganography can significantly reduce the level of distortions (demasking features) in an image container when hiding messages, compared to common steganographic methods, in particular nsF5, JSteg, OutGuess, YASS [5, 6]. This is achieved through the use of special methods to minimize distortions in an image container when embedding stegodata [7, 8], in particular adaptive steganographic methods. These methods are based on the detection of CI pixels, the brightness changes of which do not lead to significant distortions of the statistical parameters of the container, and their subsequent use to hide individual stegobits. Therefore, an overview of the features of the latest ASMs, which have a negative impact on the accuracy of the work of modern steganography detectors, is of interest.

A significant amount of current SMs for DI are based on minimizing the function of estimate distortion estimation $D(\mathbf{X}, \mathbf{Y})$ of the $\mathbf{X}$ container image when forming a steganogram $\mathbf{Y}$ [7]:

$$D(\mathbf{X},\mathbf{Y})=\sum_i \rho_i(\mathbf{X},\mathbf{Y})\to \min, |\mathbf{M}|=\text{const}, \quad (1)$$

where $\rho_i(\cdot)$ is a function for assessing changes in the statistical parameters of CI when embedding the i-th stegobit; $|\mathbf{M}|$ is the size of stegodata (in bits). Representing the process of steganogram formation as a solution to the optimization problem (1) makes it possible to adaptively hide individual stegobits taking into consideration the peculiarities of CI statistical and spectral parameters.

The paper examines the current adaptive steganographic methods HUGO [7] and MiPOD [8]. These methods are based on hiding messages in the CI spatial domain by changing the brightness values of individual pixels in the container. HUGO steganographic method is based on solving the following optimization problem when embedding stegodata to CI [7]:

$$\min_{\pi_i} \mathrm{E}_\pi[D]=\sum_{Y\in Y}\pi(Y)\cdot D(Y), H(\pi)=|M|, \quad (2)$$

$$H(\pi)=-\sum_{y\in \mathrm{Y}}\pi(y)\cdot\log(\pi(y)), \quad (3)$$

where $\mathbf{Y}\in\Upsilon$ is the steganogram $\mathbf{Y}$ from a set of all possible steganograms Υ; π is the probabilistic distribution regarding the choice of the steganogram $\mathbf{Y}$ from the Υ set of steganograms; $\mathrm{E}_\pi[D]$ is the averaging operator of the values of the function $D(\mathbf{X}, \mathbf{Y})$ regarding the distribution of π; $H(\pi)$ is the function for determining the entropy of the π distribution.

Paper [7] shows that the optimal type of the distribution of π to solve an optimization problem (3) is the Gibbs distribution:

$$\pi_\lambda(y)=\exp(-\lambda D(y))/Z(\lambda), \quad (4)$$

where $Z(\lambda)=\sum_{y\in Y}\exp(-\lambda D(y))$ is the normalizing constant. The values of the scalar parameter λ ($\lambda>0$) are determined by solving equation (4) [7]. Subject to additivity of the function $D(\cdot)$ (the independence of individual distortions of CI when embedding individual stegobits), expression (4) can be represented as follows [7]:

$$\pi_\lambda(y)=\prod_i \pi_\lambda(y_i)=\frac{\prod_i \exp(-\lambda\rho_i(y_i))}{\sum_{t\in\Im}\exp(-\lambda\rho_t(y_t))}. \quad (5)$$

As a function $\rho_i(\cdot)$, work [9] proposed using the local potentials $V_C(\cdot)$. A value of the $V_C(\cdot)$ function depends on the degree of correlation of the brightness values of adjacent pixels in the image in a given neighborhood (click) C of the current pixel. This correlation can be evaluated using a contiguity matrix $\mathrm{C}_{k,l}(\mathbf{X})$:

$$\mathrm{C}_{k,l}(\mathbf{X})=\sum_i\sum_j\left[\mathrm{x}_{i,j}=k\right]_I\left[\mathrm{x}_{i,j+1}=l\right]_I, \quad (6)$$

where $x_{i,j}$ is the brightness value of a CI pixel with coordinates (i,j). Consider an example of calculating the matrix of contiguity $\mathrm{C}_{k,l}(\mathbf{X})$ (6) in the case of processing the CI by strings and scanning pixels from left to right [7]:

$$\mathbf{A}^{\rightarrow}_{k,l}(\mathbf{X})=\frac{1}{N(M-2)}\sum_{i,j}\left[\left(\mathbf{D}^{\rightarrow}_{i,j},\mathbf{D}^{\rightarrow}_{i,j+1}\right)(\mathbf{X})=(k,l)\right]_I, \quad (7)$$

$$\left(\mathbf{D}^{\rightarrow}_{i,j},\mathbf{D}^{\rightarrow}_{i,j+1}\right)(\mathbf{X})=(k,l)\Leftrightarrow \mathbf{D}^{\rightarrow}_{i,j}(\mathbf{X})=$$
$$=k\wedge\mathbf{D}^{\rightarrow}_{i,j+1}(\mathbf{X})=l, \mathbf{D}^{\rightarrow}_{i,j}(\mathbf{X})=\mathbf{X}_{i,j+1}-\mathbf{X}_{i,j}. \quad (8)$$

If the brightness values of a CI pixel when embedding a separate stegobit change at (±1), the normalized contiguity matrix $\mathbf{A}^{\rightarrow}_{k,l}(\mathbf{X})$ coincides with the corresponding matrix $\mathbf{A}^{\rightarrow}_{k,l}(\mathbf{Y})$ for the formed steganogram [7]:

$$\left|\mathbf{A}^{\rightarrow}_{k,l}(\mathbf{Y})-\mathbf{A}^{\rightarrow}_{k,l}(\mathbf{X})\right|=\sum_{c\in\mathrm{C}}\mathbf{H}^{\vec{(k,l)}}_{c}(\mathbf{Y}), \quad (9)$$

$$\mathbf{H}^{\vec{(k,l)}}_{c}(\mathbf{Y})=\frac{1}{N(M-2)}\left|\begin{matrix}\left[\left(\mathbf{D}^{\rightarrow}_{i,j},\mathbf{D}^{\rightarrow}_{i,j+1}\right)(\mathbf{Y})=(k,l)\right]_I-\\-\left[\left(\mathbf{D}^{\rightarrow}_{i,j},\mathbf{D}^{\rightarrow}_{i,j+1}\right)(\mathbf{X})=(k,l)\right]_I\end{matrix}\right|, \quad (10)$$

for all possible horizontal clicks of the current pixel $C^{\rightarrow}=\left\{c: c=\left\{(i,j),(i,j+1),(i,j+2)\right\}\right\}$. Similarly, it is possible to determine the contiguity matrices for other clicks C, namely, $\mathbf{A}^{\leftarrow}_{k,l}(\mathbf{Y})$, $\mathbf{A}^{\uparrow}_{k,l}(\mathbf{Y})$ and $\mathbf{A}^{\downarrow}_{k,l}(\mathbf{Y})$.

The procedure for embedding stegodata to CI according to the HUGO method can be represented as a solution to the following optimization problem [7]:

$$D(\mathbf{Y})=\sum_{c\in\mathrm{C}}\sum_{k,l}w_{k,l}\mathbf{H}^{(k,l)}_{c}(\mathbf{Y}), \quad (11)$$

where $C=C^{\rightarrow}\cup C^{\leftarrow}\cup C^{\uparrow}\cup C^{\downarrow}$ is a set of all possible clicks from three elements; $w_{k,l}>0$ is the weight factor.

An alternative approach to the choice of the function $D(\mathbf{X}, \mathbf{Y})$ (1) is based on the representation of the process of steganogram formation as a solution to the problem of multicritical optimization, namely minimization of both the degree of distortion of CI with stegodata and the probability of detecting steganograms when using common types of statistical SDs [10]. An example of steganographic methods based on the use of a given approach is the latest MiPOD method [8].

Hiding messages according to the MiPOD method is carried out in several stages [8]. At the first stage, the CI is treated using a denoising filter F_{dn} to reduce the impact of CI:

$$\mathbf{r} = \mathbf{X} - F_{dn}(\mathbf{X}), \tag{12}$$

where the container image $\mathbf{X}$ is represented by «expanding» over the columns.

After that, one performs an assessment of variance σ_l^2 in the values of the obtained remains $\mathbf{r}$ using the method of maximum likelihood:

$$\mathbf{r}_l = \mathbf{G}\mathbf{a}_l + \xi_l, \tag{13}$$

where $\mathbf{r}_l$ corresponds to the values of the resulting remains $\mathbf{r}$ in the neighborhood of the l-th pixel the size of $p \times p$ elements; $\mathbf{G}_{p^2 \times p}$ is the matrix of mixing model parameters; a_{px1} is the vector of model parameters; $\xi_{p^2 \times 1}$ is the noise vector.

In the second stage, an assessment of the value of variance in σ_l^2 is carried out using the following expression [8]:

$$\sigma_l^2 = \left\| \mathbf{P}_{\mathbf{G}}^{\perp} \mathbf{r}_l \right\|^2 / \left(p^2 - q \right), \tag{14}$$

where $\mathbf{P}_{\mathbf{G}}^{\perp} = \mathbf{I}_l - \mathbf{G}\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{G}^T$ the operator of the orthogonal projection of remains $\mathbf{r}_l$ onto the$(p^2–q)$ space created by the eigenvectors of matrix $\mathbf{G}$; $\mathbf{I}_{l \times l}$ is a unit matrix of $l \times l$ elements.

In the third step, the probability β_l of using the l-th CI pixel when embedding messages is calculated. For common types of statistical SDs, the β_l probability value is chosen to minimize the coefficient of discrepancy ζ^2 between the distributions of CI and the formed steganograms:

$$\zeta^2 = 2\sum\nolimits_{l=1}^{M \cdot N} \beta_l^2 \sigma_l^{-4}, \tag{15}$$

subject to a fixed amount of hidden message:

$$R = \sum\nolimits_{l=1}^{M \cdot N} H(\beta_l), \tag{16}$$

where $H(\beta_l) = -2\beta_l - (1 - 2\beta_l)\log(1 - \beta_l)$ is the function for determining ternary entropy; R is the degree of filling CI with stegodata.

The optimal β_l values, which minimize the value of expression (15), can be determined using the Lagrange multiplier method to solve the following $(l+1)$ equations:

$$\beta_l \sigma_l^{-4} = \frac{1}{2\lambda} \ln\left(\frac{1 - 2\beta_l}{\beta_l} \right), \; l \in [1; M \cdot N], \tag{17}$$

$$R = \sum\nolimits_{l=1}^{M \cdot N} H(\beta_l). \tag{18}$$

Substituting the obtained β_l values in the function $\rho(\cdot)$, we obtain:

$$\rho_l = \ln(1/\beta_l - 2). \tag{19}$$

At the last stage, the message $\mathbf{M}$ is processed using trellis codes, and one further embeds data by changing the brightness of CI pixels. The choice of pixels to hide individual stegobits is carried out taking into consideration the weight coefficients (19) by minimizing the overall level of distortion of the image container.

The use of normal distribution to model the parameters of CI noise components in the MiPOD method makes it possible to take into consideration the non-stationarity of the distribution of natural noises of actual images in the formation of steganograms. That also makes it possible to obtain an analytical expression to assess the probability of detecting formed steganograms based on the value of the parameter ζ^2 (15) [8].

It is worth noting that the HUGO and MiPOD methods considered here are characterized by extremely small CI distortions in the process of hiding messages, compared to common types of SMs [8]. As a result, the effectiveness of the use of modern SDs based on statistical models of the image-container is significantly reduced, therefore, it is of interest to use the latest convolutional neural networks (CNN) to improve the detection accuracy of steganograms formed according to these methods.

A current approach to building an SD is to determine the demasking features of steganograms (the CI parameters that change the most due to the hiding of messages) and their subsequent use to configure a two-class (binary) classifier. Spectral, statistical, and structural parameters of DI [11] are widely used as demasking features. Classification of DI using the resulting vectors is carried out using common types of binary classifiers, in particular, the method of support vectors, ensemble classifiers, etc. [12].

An example of an SD based on the use of a given approach is steganography detectors based on cover rich models (CRM), proposed in work [13]. To analyze changes in the degree of correlation of brightness values of adjacent pixels of the investigated image, caused by the concealment of messages, the mathematical apparatus of Markov chains is used in these SDs. That has made it possible to significantly improve the accuracy of steganogram detection in comparison with signature steganography detectors and to perform stegoanalysis under conditions of the limited a priori data on the SM used.

Despite the high accuracy of the SD based on CRM, the limitation of their practical application is high computational complexity. This is due to the need to use a significant number of DI parameters to ensure high accuracy of steganogram detection (for example, 34,671 parameters for the maxSRM model [10]). To overcome this limitation, approaches were proposed based on reducing the dimensionality of vectors, in particular using the Karhunen-Loève transform [14, 15]. That has made it possible to reduce the dimensionality of vectors used at a controlled loss of accuracy of SD performance.

A further stage in the development of SD was the introduction of methods of preliminary processing of the investigated DIs in order to strengthen weak distortions caused by the concealment of messages [16]. An example of these methods is reported in [5], proposed to reduce the level of DI distortion caused by repeated lossy JPEG compression. A given method involves the decompressing of an image, discarding (cropping) the first four rows and columns of the recovered image, and subsequent JPEG compression using the original quant table. That has made it possible to significantly reduce the distortions of DI caused by lossy JPEG compression, but its effectiveness in reducing other types of distortions when using this method is low.

To overcome this limitation, methods for reducing the impact (suppression of context) of the CI were proposed. A striking example of a given approach is the group of statistical models SRM [17], based on the use of the ensemble of high frequencies filters (HFF). This approach has made it possible to significantly improve the accuracy of steganogram detection compared to the SDs considered and was an impetus for the development of the latest statistical SDs, in particular PSRM [18], GFR [19], and so on. However, the limitation of the practical application of a given approach is the need for pre-selection of HFF to minimize the error of steganogram detection, which requires the use of a priori data on SM. This greatly complicates the reconfiguration (adaptation) of SD to identify the latest ASMs, for which a priori data on the features of hiding messages are limited.

In order to overcome this limitation, work [3] proposed using ANNs. The use of ANN in image processing of SD makes it possible to combine the stages of preprocessing of the image under study and subsequent calculation of statistical parameters of the processed DIs within a single neural network. That makes it possible to combine the constituent parts (layers) of the ANN and adapt them in the learning process to minimize the error of steganogram detection. ANN is configured using the method of error backpropagation (differences between the received and specified output values) to correct the ANN parameters during its adjustment [20, 21].

One of the first ANNs for the tasks of stegoanalysis of DI is the SCAE model (Stacked Convolutional Autoencoder), proposed in [3]. The peculiarity of a given network is the use of an ensemble of auto-encoding networks (AEN) to identify differences between the statistical parameters of the CI and steganograms. The use of AEN in the SCAE model has made it possible to reduce the computational complexity of SD configuration since the adjustment of individual AENs can be carried out without the need to attract pre-marked pairs of container images and steganograms.

The SCAE model [3] launched a wide class of ANN-based stegoanalysis methods. One of the first ANN-based steganography detectors is the Qian-Net model [22], proposed to detect steganograms with data embedded in the spatial domain of DI. A given neural network has made it possible to achieve the accuracy of steganogram detection comparable to the use of CRM-based SD, with a decrease in the computational complexity of the configuration procedure. That has led to a rapid increase in interest in the use of the latest types of CNN, in particular multilayer (deep) CNN, in the field of stegoanalysis of DI, namely: the emergence of such models as Xu-Net [23], Ye-Net [24], Yedroudj-Net [25], SR-Net [26], and Zhu-Net [27].

Modern ANN-based steganography detectors make it possible to achieve the accuracy of steganogram detection, which is comparable with the use of steganography detectors based on CRM. This is achieved due to the following features of ANN-based steganography detectors:

– the use of an HFF ensemble to reduce the impact of CI and enhance distortions caused by the concealment of messages;

– applying a sequence of convolutional layers (CL) of the network;

– the use of a re-sampling operation (pooling) to reduce the dimensionality of the resulting attribute vectors;

– the classification of processed attributes using a fully connected layer of artificial neurons.

One of the first ANNs developed for use in the tasks of DI stegoanalysis is the convolutional neural network Xu-Net [23]. The peculiarity of a given network is the introduction of additional stages of processing statistical parameters of DI, in particular, the use of absolute values of these characteristics in the functions of activating intermediate layers of ANN.

A significant increase in the detection accuracy of steganograms formed according to common SMs, when using the Xu-Net model compared to common SDs, was the impetus for further research into the use of convolutional neural networks in the tasks of DI stegoanalysis. An example is the convolutional neural network Yedroudj-Net [25]. A given network is based on the use of 30 high-frequency filters from the SRM statistical model during preprocessing of DI, as well as the normalization of the calculated statistical parameters of the processed images at the output of the convolutional layers of the network. However, the Yedroudj-Net network does not use methods to estimate the likelihood of pixel changes in CI in the process of hiding messages according to common SMs, which reduces the detection accuracy of steganograms formed using the latest SMs, based on synchronization of pixel brightness changes [28].

One of the first «universal» CNN for the detection of steganograms with data embedded in both spatial and spectral regions of CI is the SR-Net neural network [26]. The peculiarity of a given network is the decrease in the use of heuristic methods of CL configuration, which were inherent in the considered CNN. This is achieved by initializing the convolutional layers of the network using an HFF from the SRM model [17] and further updating the CL settings based on the results of CNN configuration. A given approach has made it possible to significantly improve the accuracy of detection of steganograms formed according to a number of newest SMs, compared to the SDs considered. However, that increase in accuracy was achieved due to a significant complication of the network structure, which led to a significant increase in the computational complexity of its configuration.

To overcome the above limitations of the SR-Net network, the newest Zhu-Net neural network was proposed in [29], based on the use of specialized Depthwise Separable Convolution (DSC) functions. The peculiarity of the DSC convolution is the processing of DI in two stages: the use of the convolution operator with the specified HFF for the processing of individual channels of DI color and the subsequent merging of convolution results for pixels that have the same spatial coordinates. An additional increase in the accuracy of steganogram detection in the Zhu-Net network is achieved by processing the results of DSC convolution on several scales using the spatial pyramid pooling (SPP) method [30]. These features of the Zhu-Net network have made it possible to build one of the most effective modern SDs.

One of the promising CNN for the tasks of digital image stegoanalysis is the GBRAS-Net network proposed in [31]. A given network is aimed at further improving the convolutional Zhu-Net [32], in particular, reducing the computational complexity of adjusting the SD while maintaining a fixed accuracy of detecting steganograms. This is achieved through the preliminary selection of an HFF for the processing of the examined images, as well as methods for reducing the dimensionality of vectors of DI statistical parameters, in particular the use of Average Pooling and Batch Normalization of the obtained values. To reduce the negative impact of reducing the signal adjustment of CNN parameters on the accuracy of network performance, direct connections between individual layers of artificial neurons are used when setting up GB-RAS.

Based on the results of their review of modern steganography detectors based on convolutional neural networks,

one can conclude that the high accuracy of steganogram detection when using these SDs is achieved through the comprehensive use of specialized methods of DI preprocessing, in particular, DSC convolution, as well as methods for mapping the acquired statistical attributes of images onto a space of lower dimensionality. That makes it possible to significantly reduce the error of steganogram detection in comparison with common types of SDs, in particular, based on the use of statistical models of DI, while maintaining the relatively low computational complexity of the steganography detector configuration. Nevertheless, ANN-based steganography detectors are characterized by a significant dependence on the accuracy of steganogram detection on the statistical and spectral parameters of the examined images, which reduces the operational efficiency of these SDs when used in actual RI leak counteraction systems. Therefore, it is of interest to use specialized types of artificial neural networks, in particular «hybrid» networks, in the tasks of DI stegoanalysis to overcome those limitations of SDs based on CNN.

One example of a modern «hybrid» ANN is the ASSAF network [32]. A given network consists of a denoising autoencoder (DAE) and a dual (Siamese) neural network. The use of DAE makes it possible to ensure high accuracy of estimation of CI statistical parameters without the need to use a significant amount of CLs for preprocessing the examined images, which is characteristic of SD based on CNN. A given autoencoder consists of a network encoder, with the help of which vectors of DI statistical parameters are mapped onto a space of lower dimensionality, and a corresponding network-decoder that evaluates the initial type of CI based on the acquired image characteristics. The difference between DAE and common types of auto-encoding networks is changes in the network configuration procedure – the use of «noisy» images (steganograms) and CIs as «expected» network output data. The second part of the ASSAF network, namely the dual neural network, consists of a CL sequence to calculate and compare the statistical features of the examined image and the obtained assessment of the initial type of CI (DAE performance results).

Despite the emergence of the latest types of CNN for the tasks of DI stegoanalysis, the operational accuracy of steganography detectors based on them significantly depends on the presence of a priori data on the used steganographic method and statistical parameters of the examined images [26, 27, 33]. That predetermines the relevance of the task of devising high-precision methods for DI stegoanalysis, capable of ensuring high accuracy of steganogram detection under conditions of the limited a priori data on the used steganographic method and a significant variation in the parameters of the examined images while maintaining relatively low computational complexity of adjustment.

To tackle this issue, a number of methods were proposed aimed at using pre-configured CNN [33], an ensemble of several artificial neural networks [3, 34], special types of layers of artificial neurons [30], and so on. These methods are aimed at overcoming only certain limitations of existing CNN for the tasks of DI stegoanalysis, in particular, increasing the accuracy of work on new image packages, reducing the computational complexity of SD configuration methods, etc. Therefore, it is of interest to study the effectiveness of the use of specialized types of ANNs in the construction of a steganography detector, in particular, the use of auto-encoding networks [29]. However, the scientific literature lacks information on the comparative analysis of the operational accuracy of SD, configured using common and special types of ANNs. This predetermines the relevance of the task to estimate the detection accuracy of steganograms formed according to the latest ASMs when using SDs based on the common and special types of artificial neural networks.

3. The aim and objectives of the study

The purpose of this study is to assess the operational accuracy of steganography detectors based on the use of common and specialized types of artificial neural networks to identify steganograms formed according to the latest adaptive steganographic methods. This study's results could make it possible to compile recommendations on the choice of ANN architecture in order to improve the operational accuracy of modern steganography detectors.

To accomplish the aim, the following tasks have been set:

– to investigate changes in the accuracy of steganogram detection when using current steganography detectors based on statistical models of container images when processing image packets, characterized by a significant variation in statistical parameters;

– to conduct a comparative analysis of the performance accuracy of steganography detectors based on the convolutional neural network GB-RAS and the «hybrid» ASSAF network to identify steganograms formed using modern adaptive steganographic methods.

4. The study materials and methods

The object of our research is the process of steganogram detection in the processing, storage, and transmission of digital images in information and communication systems. The subject of this study is artificial neural networks used in the construction of modern SDs to identify steganograms formed using adaptive steganographic methods.

The current paper tests a hypothesis on the effectiveness of the use of special types of ANNs, in particular «hybrid» neural networks, when designing SDs in order to overcome the limitations of modern steganography detectors, namely a significant dependence of the accuracy of steganogram detection on the statistical and spectral parameters of the examined images.

During the research, optimization theory methods were used to solve single-criteria optimization problems to minimize changes in the statistical parameters of container images when embedding messages according to adaptive steganographic methods. To study the effectiveness of the use of ANNs in the tasks of DI stegoanalysis, elements of the theory of artificial neural networks were used, in particular, methods for constructing convolutional and special types of neural networks. The analysis of statistical parameters of the examined images, in particular the degree of correlation of brightness values of adjacent pixels of container images and steganograms, was carried out using the mathematical apparatus of Markov chains. Setting up and investigating the performance accuracy of steganography detectors involved methods of pattern recognition theory.

The construction of dependence plots for estimating the operational accuracy of steganography detectors, based on the use of common and specialized types of artificial neural networks, to identify steganograms formed according to the

latest adaptive steganographic methods, was carried out using the software package MATLAB (USA).

We analyzed the effectiveness of using steganography detectors based on the considered networks GB-Ras [31] and ASSAF [32] using steganograms formed according to the adaptive steganographic methods HUGO [7] and MiPOD [8]. The degree of CI filling with stegodata varied in the following limits – from 3 % to 5 % in increments of 2 %, from 5 % to 10 % in increments of 5 %, from 10 % to 50 % in increments of 10 %.

The study was conducted using the following digital image test packets:

– ALASKA packet [35]: widely used to evaluate the effectiveness of modern SDs. This packet consists of 80,000 images obtained using 40 common camera models, including smartphones, tablets, and digital cameras;

– VISION packet [36]: proposed in the field of DI expertise, in particular, assessment of the parameters of natural noise and identification of the source of images (the models of camera used). This packet consists of 34,427 images and 1,914 videos from the social networks Facebook, YouTube, and Instagram, obtained using 35 models of mobile devices (smartphones) – Apple (USA), Samsung (South Korea), Huawei (China), LG (South Korea), Sony (Japan), and others.

During the research, for each of the test packets, samples of 10,000 images were pseudo-randomly generated. The images were cropped to provide for the same size of 512×512 pixels. The change in the color system of test images for grayscale representation was carried out using the standard function «rgb2gray» from the mathematical software MATLAB (USA).

Steganography detectors were tested according to the cross-validation procedure when the test image packet was divided into a training sample (70 %) and a control sample (30 %). The adjustment of the SDs based on the GB-RAS and ASSAF networks was carried out on the samples of CI and steganograms formed according to the HUGO and MiPOD steganographic methods, involving a variation in the degree of filling CI with stegodata [9].

According to the recommendations given in [32], the ASSAF neural network was configured in several stages. In the first phase, the test image packet was pseudo-randomly divided into three parts – 4,000 images to adjust the DAE, 4,000 images to adjust the dual neural network, and 2,000 images were used to assess the accuracy of the customized network. At the second stage, the DAE was adjusted using container image pairs and their corresponding steganograms, formed according to the HUGO and MiPOD steganographic methods. The degrees of filling CI with stegodata for both methods were equal to 20 % and 40 %. Formed steganograms were submitted to the input of the DAE, and the corresponding CIs acted as the expected output data. We compared container images and the images acquired during DAE operation using binary cross-entropy and rms deviation. At the third stage, the dual neural network was configured using images pre-processed using the configured DAE.

For comparison, the case of configuring an SD using the current statistical model maxSRMd2 [10] was considered. A given model belongs to the group of statistical models SRM [17], based on the preliminary processing of DI using a set of HFFs to reduce the impact of DI. To assess the degree of correlation of brightness values of adjacent DI pixels in maxSRMd2 model, the first- and second-order Markov chains' mathematical apparatus was employed [10].

To assess the detection accuracy of the formed steganograms when using SD, the general error of image classification P_E was applied [12]:

$$P_E = \min_{P_{FA}} \frac{1}{2}\left(P_{FA} + P_{MD}\left(P_{FA}\right)\right), \quad (20)$$

where P_{FA} and P_{SD} correspond to the probabilities of error of the first (classification of the container image as a steganogram) and the second (classification of the steganogram as a container image) orders. To obtain the average values of the accuracy of SD operation, the breakdown of the DI packet into a training sample and a control sample was carried out in a pseudo-random way 10 times.

5. Investigating the accuracy of steganogram detection when using steganography detectors based on artificial neural networks

5. 1. Results of studying the accuracy of steganogram detection when using statistical steganography detectors

We analyzed the performance accuracy of a steganography detector based on the maxSRMd2 statistical model using steganograms that were formed according to the steganographic methods HUGO [7] and MiPOD [8] when varying the degree of filling CI with stegodata.

The values of a classification error P_E of steganograms formed according to the HUGO and MiPOD steganographic methods, when using SDs based on the maxSRMd2 statistical model, are given in Table 1.

The use of the examined steganography detector makes it possible to ensure the high detection accuracy of steganograms formed according to the examined ASM in the area of average (up to 20 %) and strong (up to 50 %) degree of filling CI with stegodata when processing images from the VISION packet (Table 1).

Table 1

PE values for steganograms formed according to a HUGO method when using a steganography detector based on the maxSRMd2 model

Embedding method	Image packet	The degree to which the container image is filled with stegodata, %						
		3	5	10	20	30	40	50
HUGO	ALASKA	48.64	47.33	44.44	38.90	35.04	31.71	28.59
	VISION	26.54	17.99	9.19	4.48	2.88	2.13	1.64
MiPOD	ALASKA	49.13	48.39	45.84	41.03	36.52	32.80	29.50
	VISION	22.67	14.79	7.31	3.66	2.49	1.90	1.43

Images from this packet are characterized by a high level of natural noise, which increases the efficiency of using the HFF ensemble in the maxSRMd2 statistical model to identify weak changes in CI caused by the concealment of messages. On the other hand, the processing of high-quality images from the ALASKA packet when using a given model leads to a significant increase in the error of classification of steganograms compared to the case of processing DI from the VISION packet (Table 1). This is due to the relatively small level of natural noise of images, as a result of which the use of the HFF ensemble leads to a decrease in the influence of both natural noises and distortions caused by the concealment of messages.

It is worth noting that the examined image packets make it possible to evaluate the achievable accuracy of steganogram detection when using modern statistical steganography detectors in the case of processing high-quality images (ALASKA packet), as well as images formed using common types of digital cameras (VISION packet). However, the performance accuracy of the statistical steganography detector considered may decrease slightly in the case of processing DI circulating on social networks and messaging services. This is due to the use of additional methods to improve the visual quality of DI in these systems. In addition, the accuracy of steganogram detection significantly depends on the size of the images studied [5] – increasing the size of the DI leads to an increase in the accuracy of determining the statistical parameters of the examined images and, accordingly, an increase in the accuracy of steganogram detection. The sizes of the test DIs from the considered ALASKA and VISION packets are equal to 512×512 pixels, which corresponds to the standard image sizes for the tasks of DI stegoanalysis. Accordingly, the processing of images with a larger size would further reduce the error of steganogram classification.

Taking into consideration the relatively low accuracy of steganogram detection, in particular in the area of poor filling the CI with stegodata (less than 10 %), when using a steganography detector based on the maxSRMd2 model, it is of interest to study the effectiveness of the use of modern ANNs to improve the operational accuracy of steganography detectors.

5. 2. Results of studying the accuracy of steganogram detection when using artificial neural networks

We analyzed the performance accuracy of SDs based on the latest GB-RAS and ASSAF neural networks in the same way as the procedure for the steganography detector based on the maxSRMd2 model. The values of classification error P_E of the detectors formed according to HUGO and MiPOD methods when using SDs based on the GB-RAS and ASSAF neural networks are given in Tables 2, 3.

The use of steganograms with an average degree of filling (20 %) in the training sample of DI for GB-RAS makes it possible to insignificantly (up to 2 %) reduce the P_E level for the HUGO embedding method. On the other hand, the use of DAE as part of the ASSAF network can significantly improve the detection accuracy of steganograms formed from CIs from the ALASKA packet. The P_E steganogram classification error rate is 9.6 %, which is higher than the results for SDs based on maxSRMd2 models and the GB-RAS network.

The resulting classification error values for the MiPOD steganographic method (Table 3) are similar to the results obtained earlier for the HUGO method (Table 2). The use of the «hybrid» ASSAF network makes it possible to significantly (up to 37 %) improve the accuracy of steganogram detection compared to the SDs considered in the most difficult case of poor filling the CI with stegodata (less than 10 %).

Table 2

P_E values for HUGO method-based steganograms when using steganography detectors based on the GB-RAS and ASSAF models

Steganography detector	Training image sample	The degree to which the container image is filled with stegodata, %						
		3	5	10	20	30	40	50
ALASKA image packet								
GB-RAS	HUGO (20 %)	49.78	49.54	48.95	48.02	47.00	46.49	45.74
	HUGO (40 %)	49.72	49.49	48.79	47.33	46.03	44.85	43.84
	MiPOD (20 %)	49.86	49.63	49.15	47.99	46.96	45.98	45.07
	MiPOD (40 %)	49.73	49.51	48.73	47.47	46.27	44.98	43.67
ASSAF	HUGO (20 %)	11.90	11.99	11.70	10.98	10.56	10.21	9.63
	HUGO (40 %)	26.06	25.94	25.22	24.15	23.34	22.26	21.66
	MiPOD (20 %)	13.37	13.34	13.02	12.72	11.94	11.22	11.08
	MiPOD (40 %)	27.47	27.26	26.66	25.38	24.95	23.73	22.00
VISION image packet								
GB-RAS	HUGO (20 %)	49.95	49.89	49.74	49.24	48.87	48.52	48.21
	HUGO (40 %)	49.86	47.74	49.31	48.17	46.08	43.03	38.38
	MiPOD (20 %)	49.97	49.90	49.68	48.61	46.95	45.36	43.97
	MiPOD (40 %)	49.74	49.61	48.96	47.64	46.61	45.58	44.69
ASSAF	HUGO (20 %)	17.32	17.05	16.70	16.02	15.30	14.78	13.95
	HUGO (40 %)	16.54	16.78	16.22	15.49	14.68	14.04	13.26
	MiPOD (20 %)	50.00	50.00	50.00	18.12	50.02	18.23	49.98
	MiPOD (40 %)	50.02	50.02	50.02	17.11	50.00	17.08	50.02

Table 3

P_E values for steganograms formed according to the MiPOD method when using steganography detectors based on the GB-RAS and ASSAF models

Steganography detector	Training image sample	The degree to which the container image is filled with stegodata, %						
		3	5	10	20	30	40	50
ALASKA image packet								
GB-RAS	HUGO (20 %)	49.68	49.43	48.91	47.96	47.19	46.51	46.02
	HUGO (40 %)	49.60	49.32	48.57	47.42	46.16	45.08	44.20
	MiPOD (20 %)	49.81	49.61	49.15	48.11	46.96	46.10	45.10
	MiPOD (40 %)	49.78	49.52	48.88	47.61	46.35	45.05	43.65
ASSAF	HUGO (20 %)	11.99	12.18	12.09	12.06	12.03	11.96	12.14
	HUGO (40 %)	26.33	26.27	26.33	26.21	26.34	26.45	26.43
	MiPOD (20 %)	13.52	13.53	13.59	13.59	13.56	13.47	13.52
	MiPOD (40 %)	27.67	27.65	27.74	27.85	27.67	27.91	27.58
VISION image packet								
GB-RAS	HUGO (20 %)	49.92	49.89	49.75	49.29	48.91	48.60	48.35
	HUGO (40 %)	49.81	49.69	48.20	47.67	45.31	41.70	36.68
	MiPOD (20 %)	49.89	49.85	49.62	48.52	46.74	44.72	42.51
	MiPOD (40 %)	49.59	49.31	48.61	47.05	45.49	44.10	42.90
ASSAF	HUGO (20 %)	17.41	17.31	17.46	17.52	17.46	17.44	17.35
	HUGO (40 %)	16.76	16.90	16.88	16.84	16.72	16.52	16.40
	MiPOD (20 %)	50.00	50.00	50.00	18.12	50.02	18.23	49.98
	MiPOD (40 %)	50.02	50.02	50.02	17.11	50.00	17.08	50.02

At the same time, the obtained results confirm the conclusions drawn earlier, regarding the reduction of the performance accuracy of steganography detectors based on the use of the HFF ensemble, when processing DI with a small level of natural noise. It is worth noting the high level of classification errors for a steganography detector based on the GB-RAS network, based on the use of the HFF ensemble in the initial convolution layers. At the same time, the use of special methods of preprocessing images, in particular, a denoising autoencoder for the ASSAF network, makes it possible to significantly (up to 35 %) reduce the error of steganogram classification even in the most difficult case of poor filling the CI with stegodata (less than 10 %, Tables 2, 3).

Similar to the steganography detector based o the statistical model maxSRMd2, the obtained results (Tables 2, 3) make it possible to evaluate the performance accuracy of SD based on ANN when processing images, which are characterized by a low level of natural noise (ALASKA packet) or the use of additional methods of noise data filtration (VISION packet). That corresponds to the case of processing DI circulating in local computing networks CII (images obtained using digital cameras). As a result, the obtained values of the performance accuracy of SD based on the examined types of ANN may decrease in the case of processing images circulating in data exchange services, the peculiarity of which is the use of additional methods to improve the visual quality of images.

6. Discussion of results of studying the accuracy of steganogram detection when using modern steganography detectors

The effectiveness of the use of the latest convolutional network GB-RAS and the «hybrid» network ASSAF to identify steganograms formed according to ASM was studied in this paper. Based on the results of the analysis of the obtained data, it was established that the use of the GB-RAS network leads to a significant increase in the values of a steganogram classification error P_E compared to the statistical SDs considered (Table 1). Our results (Tables 1–3) can be explained by the relatively small level of DI interference from the ALASKA packet, as well as the processing of grayscale images, which reduces the efficiency of using specialized convolution methods in a given network. In the case of processing DI from the VISION packet, the accuracy of detecting steganograms using the GB-RAS network is significantly improved, especially in the case of strong filling the DI with stegodata (more than 30 %). This is due to a significant increase in the level of natural interference with the test images from the VISION packet compared to the corresponding images from the ALASKA packet.

Unlike the GB-RAS network, the maxSRMd2 statistical model makes it possible to significantly increase (up to 15 %) the accuracy of detecting steganograms (Tables 1–3). This is due to the use of an extended set of HFF for DI preprocessing in the maxSRMd2 model, which makes it possible to increase the ratio of stegodata/container compared to the case of GB-RAS network usage. These conclusions are confirmed by the test results for the VISION packet – the use of the maxSRMd2 model allows minimizing the error of detecting steganograms (up to 1.6 %) compared to the SDs considered due to a significant increase in the computational complexity of the SD setting.

The largest increase in the accuracy of steganogram detection was obtained when using the «hybrid» ASSAF network (Tables 2, 3). These results (Tables 2, 3) are explained by the high efficiency of using specialized methods of preprocessing of the investigated DI using a denoising autoencoder.

It is worth noting that the results for the GB-RAS convoluted network were previously unexpected. The use of a steganography detector based on this network did not significantly improve the detection accuracy of steganograms compared to modern steganography detectors based on CRM (Tables 2, 3). This indicates the limitation of the use of the existing approach to the construction of steganography detectors using CNN, in particular, overcoming the problem of differences between the statistical characteristics of the training and control image samples (a domain mismatch problem). The use of CRM in the construction of SD can reduce the negative impact of this phenomenon by using an ensemble of high-frequency filters [5, 17, 18].

The use of a denoising autoencoder as part of the «hybrid» ASSAF network has made it possible to reduce the negative impact of changes in the statistical parameters of DI when using the latest VISION packet (Tables 1, 2). This is confirmed by a significant increase (up to 15 %) in the accuracy of steganogram detection when using an ASSAF-based steganography detector compared to both the GB-RAS convolutional network and the modern maxSRMd2 complex statistical model. This result is due to the performance peculiarities of the denoising autoencoder [20, 21] – the use of space of relatively small dimensionality, in which changes in the attribute vectors of the studied image caused both by the concealment of stegodata and the variability in the DI parameters, are insignificant. However, the practical use of steganography detectors based on denoising autoencoders may be limited due to the need to configure a given model using a large sample of images, which is a computationally complex procedure.

The detected limitation of the use of the convolutional network GB-RAS can be overcome by using an ensemble of several networks configured on DI samples with different levels of natural noise [3]. In addition, one of the approaches to reducing the negative impact of a domain mismatch problem is to increase the number of HFFs used in the input convolution layers of this network. Improving the performance accuracy of steganography detectors based on denoising autoencoders can be achieved through the use of methods for transferring learning results [20, 21] – the use of a pre-configured autoencoder for SD operation on new image samples.

Worth noting is the high accuracy of the estimate of the error of classification of steganograms when using the examined SDs, which is due to the use of powerful digital image packets (about 20,000 processed images). This is confirmed by small variance values of classification error values not exceeding 0.05 for all considered cases.

One of the limitations of our study is the use of test image packets circulating in local (private) information and communication networks CII. These images are characterized by relatively small levels of natural noise or the presence of «typical» distortions caused by the use of methods for reducing this noise when forming images in digital cameras. Accordingly, the obtained estimates of the performance accuracy of SDs may be slightly reduced in the case of processing of DIs received after the complex application of several methods to improve the visual quality of images as well as lossy compression. That could lead to an additional increase in the degree of variability of DI statistical and spectral parameters. In addition, the limitation of the current study is the consideration only of the case of processing images of fixed size (512×512 pixels), which is widely used in research into the field of digital image stegoanalysis.

Further work should be aimed at expanding the list of test packets of images used to analyze the performance accuracy of SDs. In particular, of interest is to analyze the performance accuracy of ANN-based steganography detectors on DI packets, characterized by the use of an extended range of pixel brightness values (High Dynamic Range, HDR), as well as the large sizes of DIs (more than 8 Megapixels). In addition, insufficient attention has been paid to assessing the duration of SD setting and subsequent processing of images, as well as the development of methods for reducing this duration, which is of particular interest to increase the effectiveness of the system to counteract RI leakage.

7. Conclusions

1. It has been established that the use of a steganography detector based on the maxSRMd2 statistical model makes it possible to significantly (up to 30 %) improve the accuracy of steganogram detection in the case of analysis of digital images characterized by a high level of natural noise, compared to the case of image processing with relatively low levels of natural noise. Our result can be explained by the high efficiency of using an ensemble of high-frequency filters to reduce the effects of interference in the images under study. Thus, the use of high-frequency filter ensembles for preprocessing of the studied images makes it possible to ensure high operational accuracy of modern steganography detectors on image packets, characterized by significant variability in the statistical and spectral parameters. This approach is expedient for practical application in the case of using ensembles of filters relative to a small size (up to 7–10 filters) to ensure a compromise between the accuracy of steganogram detection and the computational complexity of setting up a steganography detector (the selection of elements of a given ensemble).

2. It has been found that the performance accuracy of SDs based on the GB-RAS artificial neural network is significantly (to 15 %) decreases in the processing of real DIs compared to the case of using the maxSRMd2 model. Our results can be explained by the relatively low noise level for DIs from the ALASKA packet, as well as the processing of grayscale images, which reduces the efficiency of using specialized convolution methods in a given network. It has been established that the use of the «hybrid» ASSAF network makes it possible to significantly (up to 37 %) improve the accuraфcy of steganogram detection compared to the steganography detectors considered in the most difficult case of poor filling the CI with stegodata (less than 10 %). The obtained results can be explained by the high efficiency of using specialized methods of preprocessing of the investigated DIs, namely the use of a denoising autoencoder. The detected effectiveness of the use of autoencoders in the tasks of steganogram detection is of theoretical and practical interest for designing high-precision steganography detectors capable of working under conditions of considerable variability in the statistical parameters of the examined images. The practical use of «hybrid» and auto-encoding networks would improve the accuracy of steganogram detection in the processing of images with a significant level of natural noise with an insignificant increase in the computational complexity of the steganography detector setting.

References

1. Yaacoub, J.-P. A., Salman, O., Noura, H. N., Kaaniche, N., Chehab, A., Malli, M. (2020). Cyber-physical systems security: Limitations, issues and future trends. Microprocessors and Microsystems, 77, 103201. doi: https://doi.org/10.1016/j.micpro.2020.103201
2. Kopeytsev, V. (2020). Steganography in attacks on industrial enterprises. Kaspersky Lab. Available at: https://ics-cert.kaspersky.com/media/KASPERSKY_Steganography_in_attacks_EN.pdf
3. Tan, S., Li, B. (2014). Stacked convolutional auto-encoders for steganalysis of digital images. Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific. doi: https://doi.org/10.1109/apsipa.2014.7041565
4. Progonov, D. (2021). Performance Analysis of Stego Images Detection Using Shallow Denoising Autoencoders. 2021 IEEE 8th International Conference on Problems of Infocommunications. Science and Technology. Kharkiv.
5. Fridrich, J. (2009). Steganography in Digital Media: Principles, Algorithms, and Applications. Cambridge University Press. doi: https://doi.org/10.1017/cbo9781139192903
6. Konakhovych, H. F., Prohonov, D. O., Puzyrenko, O. Yu. (2018). Kompiuterna stehanohrafichna obrobka y analiz multymediinykh danykh. Kyiv: Tsentr uchbovoi literatury, 558.
7. Filler, T., Fridrich, J. (2010). Gibbs Construction in Steganography. IEEE Transactions on Information Forensics and Security, 5 (4), 705–720. doi: https://doi.org/10.1109/tifs.2010.2077629
8. Sedighi, V., Cogranne, R., Fridrich, J. (2016). Content-Adaptive Steganography by Minimizing Statistical Detectability. IEEE Transactions on Information Forensics and Security, 11 (2), 221–234. doi: https://doi.org/10.1109/tifs.2015.2486744
9. Filler, T., Fridrich, J. (2011). Design of adaptive steganographic schemes for digital images. Media Watermarking, Security, and Forensics III. doi: https://doi.org/10.1117/12.872192
10. Denemark, T., Sedighi, V., Holub, V., Cogranne, R., Fridrich, J. (2014). Selection-channel-aware rich model for Steganalysis of digital images. 2014 IEEE International Workshop on Information Forensics and Security (WIFS). doi: https://doi.org/10.1109/wifs.2014.7084302
11. Avcibas, I., Memon, N., Sankur, B. (2003). Steganalysis using image quality metrics. IEEE Transactions on Image Processing, 12 (2), 221–229. doi: https://doi.org/10.1109/tip.2002.807363
12. Kodovsky, J., Fridrich, J., Holub, V. (2012). Ensemble Classifiers for Steganalysis of Digital Media. IEEE Transactions on Information Forensics and Security, 7 (2), 432–444. doi: https://doi.org/10.1109/tifs.2011.2175919
13. Pevny, T., Bas, P., Fridrich, J. (2010). Steganalysis by Subtractive Pixel Adjacency Matrix. IEEE Transactions on Information Forensics and Security, 5 (2), 215–224. doi: https://doi.org/10.1109/tifs.2010.2045842
14. Belhumeur, P. N., Hespanha, J. P., Kriegman, D. J. (1997). Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19 (7), 711–720. doi: https://doi.org/10.1109/34.598228
15. Murphy, K. P. (2012). Machine Learning: A Probabilistic Perspective. Cambridge: The MIT Press, 1104.
16. Miche, Y., Bas, P., Lendasse, A. (2010). Using multiple re-embeddings for quantitative steganalysis and image reliability estimation. TKK reports in information and computer science. Aalto University School of Science and Technology, 19. Available at: http://lib.tkk.fi/Reports/2010/isbn9789526032504.pdf
17. Fridrich, J., Kodovsky, J. (2012). Rich Models for Steganalysis of Digital Images. IEEE Transactions on Information Forensics and Security, 7 (3), 868–882. doi: https://doi.org/10.1109/tifs.2012.2190402
18. Holub, V., Fridrich, J. (2013). Random Projections of Residuals for Digital Image Steganalysis. IEEE Transactions on Information Forensics and Security, 8 (12), 1996–2006. doi: https://doi.org/10.1109/tifs.2013.2286682
19. Song, X., Liu, F., Yang, C., Luo, X., Zhang, Y. (2015). Steganalysis of Adaptive JPEG Steganography Using 2D Gabor Filters. Proceedings of the 3rd ACM Workshop on Information Hiding and Multimedia Security. doi: https://doi.org/10.1145/2756601.2756608
20. Goodfellow, I., Bengio, Y., Courville, A. (2016). Deep Learning. Cambridge: The MIT Press, 800.
21. Aggarwal, C. C. (2018). Neural Networks and Deep Learning: A Textbook. Springer, 497. doi: https://doi.org/10.1007/978-3-319-94463-0
22. Qian, Y., Dong, J., Wang, W., Tan, T. (2015). Deep learning for steganalysis via convolutional neural networks. Media Watermarking, Security, and Forensics 2015. doi: https://doi.org/10.1117/12.2083479
23. Xu, G., Wu, H.-Z., Shi, Y.-Q. (2016). Structural Design of Convolutional Neural Networks for Steganalysis. IEEE Signal Processing Letters, 23 (5), 708–712. doi: https://doi.org/10.1109/lsp.2016.2548421
24. Ye, J., Ni, J., Yi, Y. (2017). Deep Learning Hierarchical Representations for Image Steganalysis. IEEE Transactions on Information Forensics and Security, 12 (11), 2545–2557. doi: https://doi.org/10.1109/tifs.2017.2710946
25. Yedroudj, M., Comby, F., Chaumont, M. (2018). Yedroudj-Net: An Efficient CNN for Spatial Steganalysis. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). doi: https://doi.org/10.1109/icassp.2018.8461438
26. Boroumand, M., Chen, M., Fridrich, J. (2019). Deep Residual Network for Steganalysis of Digital Images. IEEE Transactions on Information Forensics and Security, 14 (5), 1181–1193. doi: https://doi.org/10.1109/tifs.2018.2871749

27. Zhang, R., Zhu, F., Liu, J., Liu, G. (2018) Efficient feature learning and multi-size image steganalysis based on CNN. arXiv.org. Available at: https://arxiv.org/abs/1807.11428
28. Ker, A. D., Bas, P., Böhme, R., Cogranne, R., Craver, S., Filler, T. et. al. (2013). Moving steganography and steganalysis from the laboratory into the real world. Proceedings of the First ACM Workshop on Information Hiding and Multimedia Security – IH&MMSec '13. doi: https://doi.org/10.1145/2482513.2482965
29. Bas, P., Filler, T., Pevný, T. (2011). «Break Our Steganographic System»: The Ins and Outs of Organizing BOSS. Lecture Notes in Computer Science, 59–70. doi: https://doi.org/10.1007/978-3-642-24178-9_5
30. He, K., Zhang, X., Ren, S., Sun, J. (2015). Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37 (9), 1904–1916. doi: https://doi.org/10.1109/tpami.2015.2389824
31. Reinel, T.-S., Brayan, A.-A. H., Alejandro, B.-O. M., Alejandro, M.-R., Daniel, A.-G., Alejandro, A.-G. J. et. al. (2021). GBRAS-Net: A Convolutional Neural Network Architecture for Spatial Image Steganalysis. IEEE Access, 9, 14340–14350. doi: https://doi.org/10.1109/access.2021.3052494
32. Cohen, A., Cohen, A., Nissim, N. (2020). ASSAF: Advanced and Slim StegAnalysis Detection Framework for JPEG images based on deep convolutional denoising autoencoder and Siamese networks. Neural Networks, 131, 64–77. doi: https://doi.org/10.1016/j.neunet.2020.07.022
33. Butora, J., Yousfi, Y., Fridrich, J. (2021). How to Pretrain for Steganalysis. Proceedings of the 2021 ACM Workshop on Information Hiding and Multimedia Security. doi: https://doi.org/10.1145/3437880.3460395
34. Reinel, T.-S., Raul, R.-P., Gustavo, I. (2019). Deep Learning Applied to Steganalysis of Digital Images: A Systematic Review. IEEE Access, 7, 68970–68990. doi: https://doi.org/10.1109/access.2019.2918086
35. Cogranne, R., Giboulot, Q., Bas, P. (2019). The ALASKA Steganalysis Challenge. Proceedings of the ACM Workshop on Information Hiding and Multimedia Security. doi: https://doi.org/10.1145/3335203.3335726
36. Shullani, D., Fontani, M., Iuliani, M., Shaya, O. A., Piva, A. (2017). VISION: a video and image dataset for source identification. EURASIP Journal on Information Security, 2017 (1). doi: https://doi.org/10.1186/s13635-017-0067-2