*mathematics*

MDPI

*Article*

# High-Pass-Kernel-Driven Content-Adaptive Image Steganalysis Using Deep Learning

**Saurabh Agarwal [1,2], Hyenki Kim [2] and Ki-Hyun Jung [2,*]**

[1] Amity School of Engineering and Technology, Amity University Uttar Pradesh, Noida 201313, India; saurabhnsit2510@gmail.com
[2] Department of Software Convergence, Andong National University, Andong 36729, Republic of Korea; hkkim@anu.ac.kr
* Correspondence: kingjung@anu.ac.kr; Tel.: +82-54-820-7968; Fax: +82-54-820-6257

**Abstract:** Digital images cannot be excluded as part of a popular choice of information representation. Covert information can be easily hidden using images. Several schemes are available to hide covert information and are known as steganography schemes. Steganalysis schemes are applied on stego-images to assess the strength of steganography schemes. In this paper, a new steganalysis scheme is presented to detect stego-images. Predefined kernels guide the set of a conventional convolutional layer, and the tight cohesion provides completely guided training. The learning rate of convolutional layers with predefined kernels is higher than the global learning rate. The higher learning rate of the convolutional layers with predefined kernels assures adaptability according to network training, while still maintaining the basic attributes of high-pass kernels. The Leaky ReLU layer is exploited against the ReLU layer to boost the detection performance. Transfer learning is applied to enhance detection performance. The deep network weights are initialized using the weights of the trained network from high-payload stego-images. The strength of the proposed scheme is verified on the HILL, Mi-POD, S-UNIWARD, and WOW content-adaptive steganography schemes. The results are overwhelming and better than the existing steganalysis schemes.

**Keywords:** digital image steganography; image steganalysis; convolutional neural network; image classification; image forensic

**MSC:** 90-04



**Citation:** Agarwal, S.; Kim, H.; Jung, K.-H. High-Pass-Kernel-Driven Content-Adaptive Image Steganalysis Using Deep Learning. *Mathematics* **2023**, *11*, 4322. https://doi.org/10.3390/math11204322

Academic Editor: Xiang Li

Received: 4 September 2023
Revised: 10 October 2023
Accepted: 11 October 2023
Published: 17 October 2023



## 1. Introduction

Images are the popular information representation choice because they are easier to interpret than text and require less storage than video. Due to their popularity, images are also used for covert communication. Steganography schemes can hide the secret message precisely so that it cannot be detected through the naked eye or by some steganalysis schemes [1]. Most of the content-adaptive steganography schemes involve only one unit change in decimal pixel value. The locations of modified pixels are generally from dense regions. In this paper, four notable content-adaptive steganography schemes, HILL [2], Mi-POD [3], S-UNIWARD [4], and WOW [5], are analyzed. An image BOWS2 dataset [6] is considered in the first row of Figure 1 to study the behavior of the schemes. Residual arrays are created by taking the difference between cover and stego-images [7]. It is apparent from the residual arrays that the number of modified elements is also amplified as the payload increases. However, the variation in the HILL, Mi-POD, S-UNIWARD, and WOW steganography schemes cannot be evaluated directly from residual arrays.

A computer-generated (CG) image (Figure 2a) and it's residual to HILL, Mi-POD, S-UNIWARD, and WOW stego-images are used to discriminate the behavior of these schemes. In the given CG image, the difference between adjacent pixel values is 1 in the smooth area and 255 in the edge region. Notably, the residual images on the second

and third rows show that the HILL residual array has the least densely modified region compared to other schemes. Mostly, modified pixels are dispersed in the whole image, making the detection of HILL stego-images challenging. However, steganography artifacts are apparent in the WOW scheme. This behavior is also reflected in the result analysis. The effect is noticeable even on a low payload, i.e., 0.2 bpp (Figure 2c). The detection of the HILL stego-image is more challenging than other schemes.

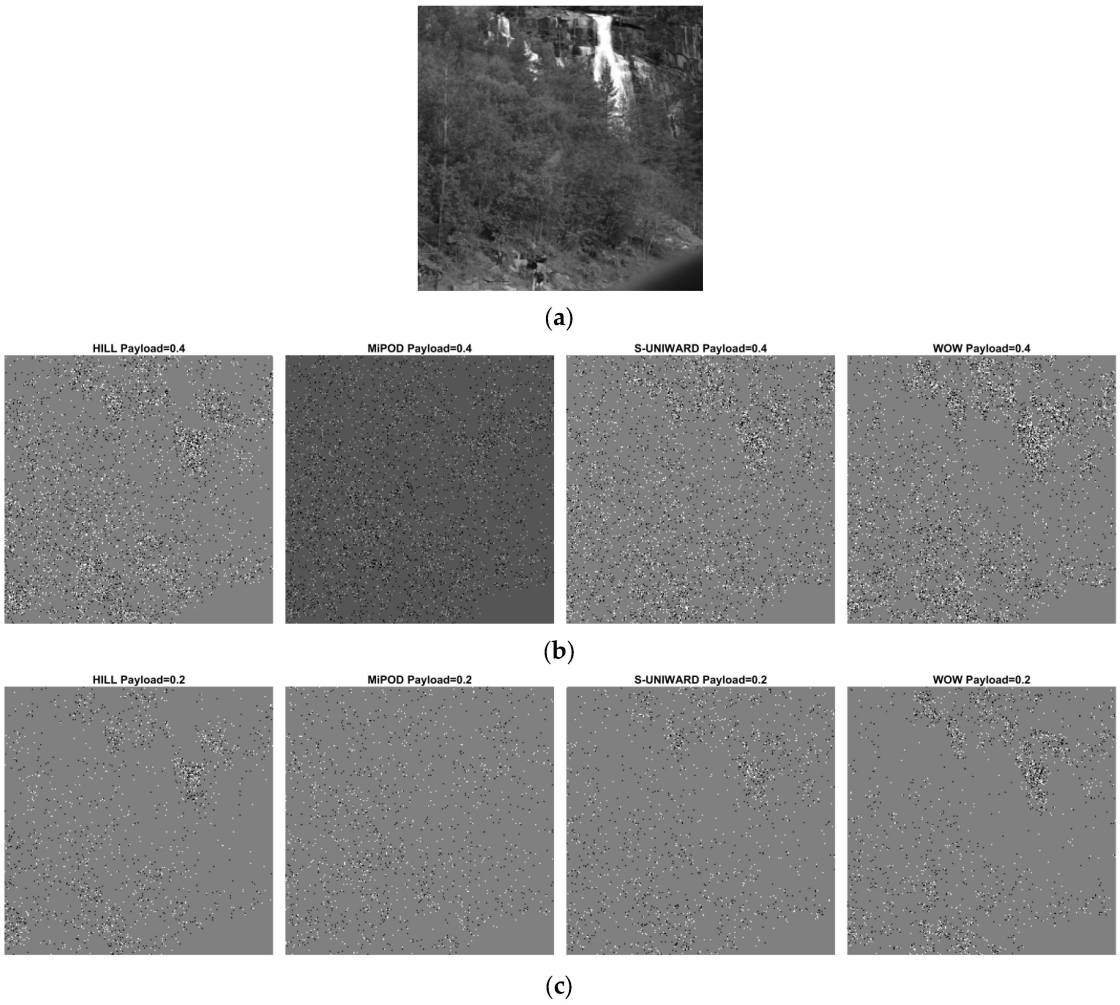



**Figure 1.** (**a**) Cover image; (**b**) residual arrays between cover and stego-images of payload 0.4 bpp; (**c**) residual arrays between cover and stego-images of payload 0.2 bpp.

Existing steganalysis schemes can be categorized broadly based on manual and automatic feature extraction. Manual feature extraction-based schemes highlight the anomalies of stego-images, mainly using co-occurrence features and texture features. Manual schemes are still worthy as they obtain satisfactory results with low computational costs, such as texture classification [8], image forensics [9], and emotion classification using short texts [10]. The convolutional neural network is utilized for automated feature extraction. First, some of the relevant manual schemes are discussed; then, automated schemes are discussed.

Xiong et al. [11] exploited DCT and DFT domains to extract co-occurrence and texture features. The feature vector dimension is also reduced by considering the symmetry in a generalized Gaussian distribution. Xiong et al. applied their method to identify numerous steganography schemes such as LSBM, HUGO, etc. Fridrich and Kodovsky [12] created diverse feature sets by considering linear and nonlinear high-pass kernels of different attributes. Multiple quantization factors are contemplated while forming the residual arrays. Further, fourth-order co-occurrence features of dimension 34,671 are extracted from

residual arrays. The proposed high-pass kernels are popularly known as SRM kernels. They are also used in many deep-learning steganalysis schemes. Tang et al. [13] especially designed a technique for WOW steganography. First, the most probable region of possible embedding is found in stego-images. Then, the scheme [12] is applied to these regions. Denemark et al. [14] introduced the maxSRM to reduce the large feature dimension of the SRM [12]. In the maxSRM, the highest probability of embedding changes is considered out of four residual probabilities. It provides a small feature vector of size 12,753. This scheme is verified using the S-UNIWARD, S-UNIGARD, and WOW steganography techniques. Xu et al. [15] proposed a local correlation pattern operator analogous to the maxSRM [14] to detect S-UNIWARD, HUGO, and LSBMR steganography techniques. The dimension of the feature vector generated by the local correlation pattern is 35,322. Li et al. [16] proposed a local texture pattern to extract features from high-pass filtered images. The size of the feature vector is enormous, i.e., 59,049. The size of the feature vector is decreased by using a PCA two times. Li et al.'s technique is applied to the HUGO and WOW steganography schemes. Li et al. [17] proposed a steganalysis technique for the HILL, S-UNIWARD, Mi-POD, and CMD-HILL [18] steganography schemes. The modified LBP named the TLBP and second-order co-occurrence features are merged to improve the results. Wang et al. [19] improved the TLBP [17] by including the rotation-invariant uniform pattern mapping. A feature separability analysis is also performed before merging the features. The Discrete Fourier Transform domain is exploited for additional SRM features. Wang et al. detected HILL, Mi-POD, and S-UNIWARD stego-images using their proposed scheme. Ge et al. [20] suggested a scheme for detecting HILL, CMD-HILL, and Mi-POD steganography techniques. The threshold LBP and co-occurrence features are extracted from high-pass filtered and non-negative matrix factorization arrays.

Steganalysis schemes based on deep learning are discussed here. Deep learning is effective in various applications, such as natural language processing [21], image and text matching [22], and image classification. Image steganalysis is a two-class classification between the cover and the stego-image. Various deep-learning-based steganalysis schemes have also been presented by different researchers that give promising results. Qian et al. [23,24] started using a deep network to detect HUGO, S-UNIWARD, and WOW steganography techniques. The proposed deep network uses convolutional, ReLU, batch normalization, and pooling layers. Preprocessing is performed on the images using a $5 \times 5$ high-pass kernel. Nevertheless, there is a lot of scope for improvement in the deep network. Xu et al. [25] applied the same filter as Qian for preprocessing. Xu et al. proposed a deep network with a new absolute layer and other conventional layers to detect the HILL and S-UNIWARD steganography schemes. Wu et al. [26,27] analyzed their scheme on HILL, Mi-POD, S-UNIWARD, and WOW stego-images. The preprocessing is performed using Qian's filter. The proposed deep network exploited numerous residual matrices. Ye et al. [28] utilized the residual of thirty SRM kernel [12] arrays in the proposed CNN. A truncated linear unit (TLU) is used in place of a ReLU. The technique is verified using HILL, S-UNIWARD, and WOW. Boroumand et al. [29] proposed a deep network called the SRNet-steganalysis residual network. Multiple steganography techniques, HILL, J-UNIWARD, S-UNIWARD, UED-JC [30], and WOW, were detected. Three kinds of sets of layers are proposed based on skip connections and pooling layers. Statistical moments are recovered after the training of the deep network. Yedroudj et al. [31] modified the [25,28] deep networks to improve the results by using some measures, including the global average pooling layer. The scheme is analyzed using the S-UNIWARD and WOW steganography schemes. Wu et al. [32] considered the shared normalization for better results on HILL, HUGO, S-UNIWARD, and WOW stego-images. Twenty SRM kernels were utilized for preprocessing. Zhang et al. [33] also used the SRM kernels for preprocessing. The proposed deep network followed the bottleneck approach and spatial pyramid pooling with other conventional layers. The scheme was applied to HILL, S-UNIWARD, and WOW steganography techniques. Fu et al. [34] also applied a similar approach without using SRM kernels. Xiang et al. [35] alleged improvements in detecting S-UNIWARD and WOW steganography techniques by

altering the organization of the layers. However, thirty SRM kernels were also used for filtering. Wang et al. [36] utilized the network parameters of low-embedded stego-images for initialization. Spatial and DCT domains are considered to detect S-UNIWARD and WOW steganography techniques. The preprocessing is carried out by SRM kernels.

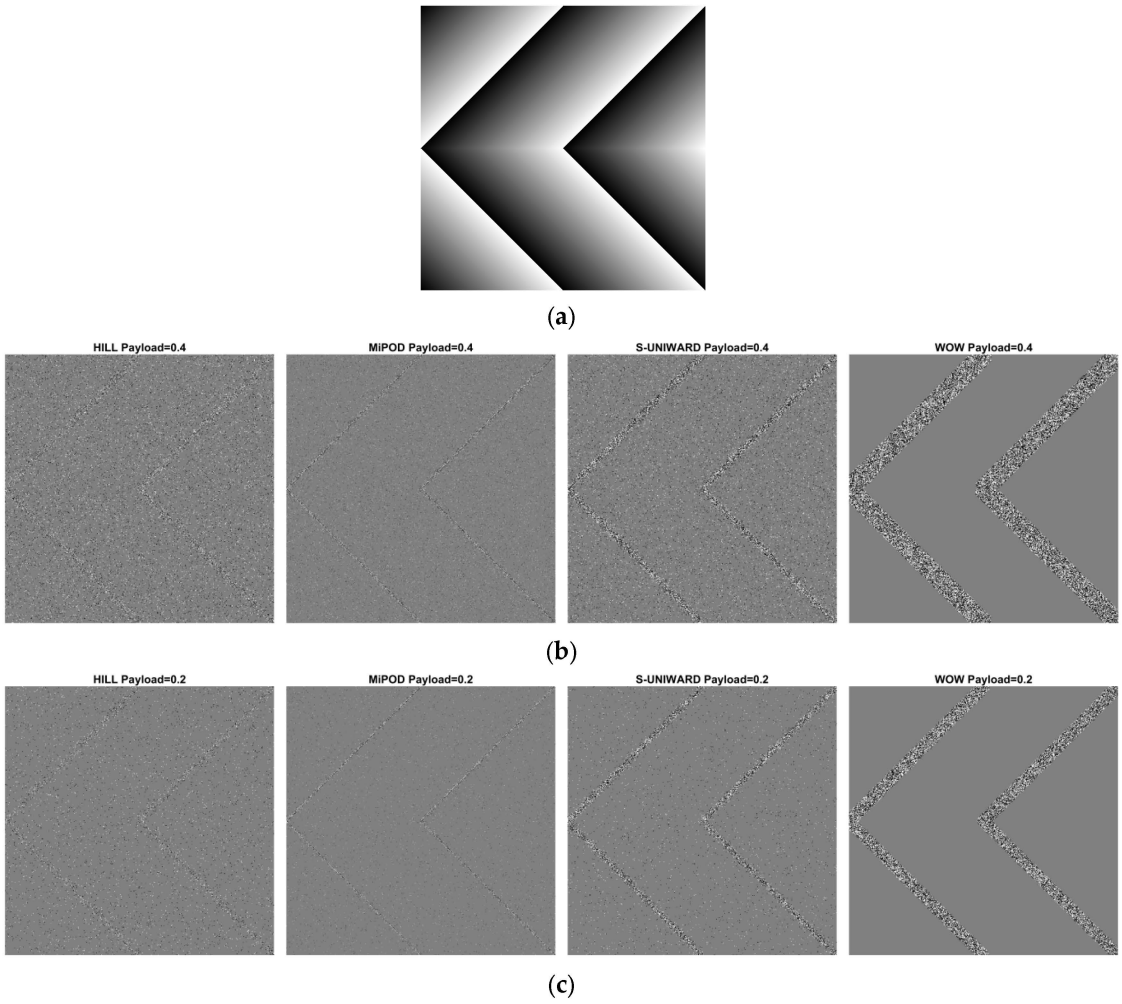



**Figure 2.** (**a**) Cover CG image; (**b**) residual arrays between cover and stego-images of payload 0.4 bpp; (**c**) residual arrays between cover and stego-images of payload 0.2 bpp.

The usage of SRM kernels is the first step in many deep learning based steganalysis schemes and other image processing applications [37,38]. This paper also exploits the SRM kernels to alleviate the weak stego-noise. The novel contributions to improve the classification of the cover and stego-image are highlighted below.

- The proposed scheme amplifies the stego-noise using multiple convolutional layers with predefined kernels. Thirty high-pass SRM kernels and one constant linear kernel are utilized as predefined kernels.
- A layer-specific learning rate higher than the global one is considered in predefined kernels based on convolutional layers. A learning rate variation in the layers boosts the performance of the proposed CNN.
- Each convolutional layer with predefined kernels is followed by three conventional convolutional layers to direct the network correctly.
- The weights of the deep network for low-payload stego-images are initialized using the weights of the trained network from high-payload stego-images.
- To maximize the activation of all neurons within the network and thus improve detection performance, the Leaky ReLU layer is preferred over the standard ReLU

layer. This strategic choice in activation functions enhances the network's ability to capture and process relevant features for more effective detection.

- A single global average pooling layer is employed exclusively to extract unprocessed information, and no intermediate pooling layers are integrated between the network layers.
- An experimental analysis is performed to detect HILL, Mi-POD, S-UNIWARD, and WOW stego-images with payloads of 0.2 bpp, 0.3 bpp, and 0.4 bpp.
- The proposed scheme's detection accuracy is better than the Ye-Net, SRNet, Yedroudj-Net, and Zhu-Net schemes.

The proposed scheme is elaborated in the subsequent sections. The experimental results are displayed in Section 3. The conclusion is presented in Section 4.

## 2. Proposed Scheme

Image steganalysis is necessary as an image is the most famous representation of information. Image steganalysis schemes evaluate the performance of steganography schemes and are used to detect stego-images. Image steganalysis is challenging due to there being only one unit of change in some pixel intensities. There is also no visual distortion in a stego-image. In this paper, a deep network is proposed to address the challenges of image steganalysis effectively. The detection of stego-images is directly related to the payload and steganography scheme used. As shown in Figure 2, the steganography scheme is influential in hiding the content. In Figure 3, different payloads are analyzed using covariance plots. The detection of the stego-image becomes easier as the payload is increased. The covariance plot is created using the standard deviation of cover and stego-images on ten thousand images of the BOWS2 dataset [6]. BOWS2 dataset images are resized to 256 × 256 pixels. Stego-images are created using 256 × 256 cover images after applying a particular steganography scheme with a specific payload. The curves of the cover and HILL stego-images with different payloads are more coupled than other schemes. The WOW has the most negligible coupling. A common deep network can be suggested as the variance represents the essential attribute of an image, and its pattern in the covariance plot is similar.

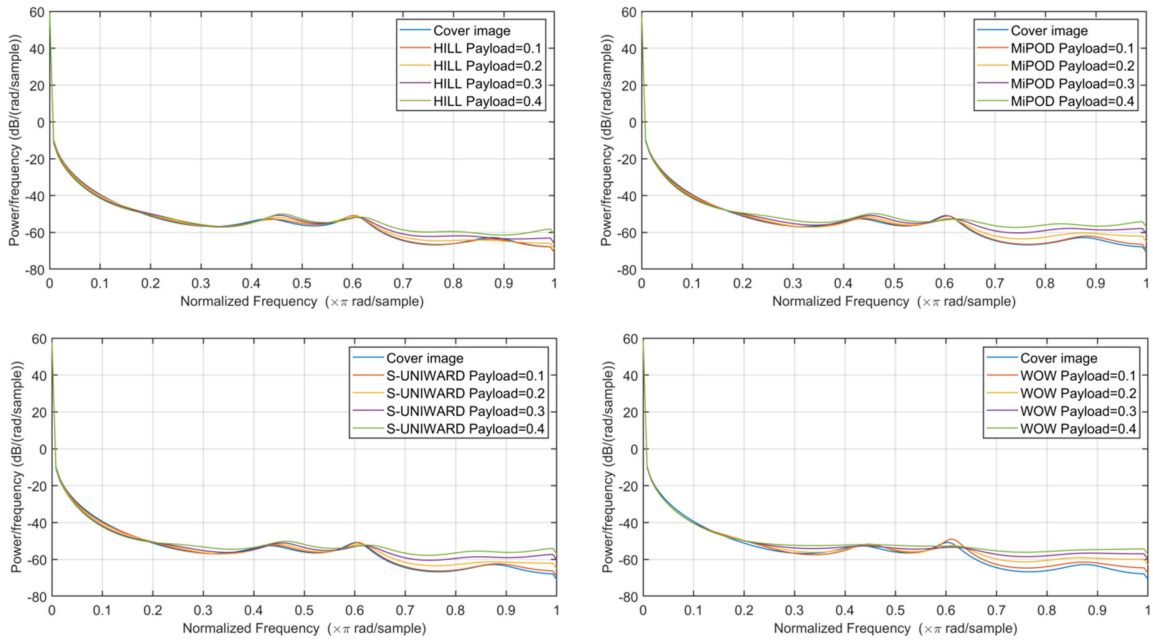



**Figure 3.** Covariance plots of the standard deviation of cover and stego-images.

The steganalysis is performed using the proposed convolutional neural network, as exhibited in Figure 4. The convolutional layer, batch normalization layer, and Leaky rectifier

unit layer are abbreviated as the Conv Layer, BN Layer, and Leaky ReLU Layer, respectively, in the layout diagram. The first convolutional layer is non-trainable, i.e., the learning rate is zero, and the number of predefined kernels is thirty-one. Thirty are SRM kernels, and one is a constant linear kernel out of the thirty-one, each with a size of $5 \times 5$. A leaky ReLU Layer is employed instead of the usual ReLU layer. The leaky ReLU layer activates all the neurons to obtain better results from the trained deep network. The Leaky ReLU layer carried out a threshold process in which every value smaller than zero is multiplied by a constant positive real number $(k)$. The values equal to or greater than zero stay intact. The thresholding process of a Leaky ReLU can be defined as follows:

$$L(p) = \begin{cases} p, & p \geq 0 \\ k * p, & p < 0 \end{cases}$$

where $p$ is the input value in our experimental analysis, and $k = 0.01$ is found suitable.

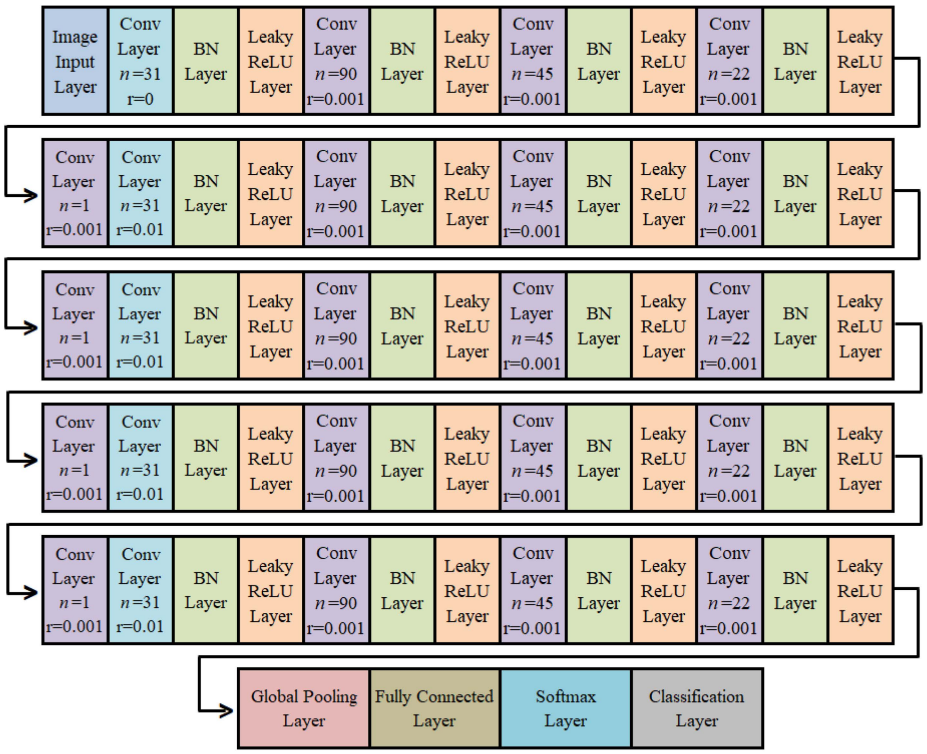



**Figure 4.** Layout of the proposed deep neural network.

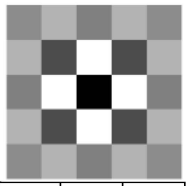
However, the values smaller than zero are substituted by zero, and other values stay intact in the general ReLU layer threading process. The thresholding operation for a ReLU can be described in the following manner:

$$R(p) = \begin{cases} p, & p \geq 0 \\ 0, & p < 0 \end{cases}$$

Traditionally, the max pooling layer is utilized to reduce the computational cost by pooling. Nonetheless, no max pooling layer is exploited to maintain the flow of information between the layers for better results. The second convolutional layer contains $n$ = 90 kernels, followed by the BN Layer and the Leaky ReLU Layer. The third convolutional layer contains $n$ = 45 kernels, followed by the BN Layer and Leaky ReLU Layer. The fourth convolutional layer contains $n$ = 22 kernels, followed by the BN Layer and Leaky ReLU Layer. To initialize with predefined thirty-one kernels, a convolutional layer (fifth) is needed, containing only one kernel. In the sixth convolutional layer, the thirty-one predefined kernels with high learning rates, i.e., r = 0.01, are considered.

The global learning rate r = 0.001 is applied to all except for the convolutional layer with predefined kernels. The cohesion of three convolutional layers with the one convolutional layer with predefined kernels improves the detection accuracy sufficiently. Previously, SRM kernels [12] were found effective in many previous schemes [28,31–33,35,36]. The order of the convolutional layers also creates a bottleneck scenario, which was also found effective in previous schemes [33]. The kernel weights are initialized using the Glorot and Bengio [39] scheme, except for five convolutional layers with predefined kernels. Before the fully connected layer, the global average pooling layer [31,40] is utilized for better classification accuracy. There is a generous improvement in performance after using the global average pooling layer. The RMSprop optimizer is used out of three popular optimizers, i.e., Adam, RMSprop, and SGD. Each class's batch size of 10 images, i.e., cover and stego, is used during network training.

To see the effect of training on the predefined kernels, one of the thirty-one predefined kernels is displayed in the first row in Figure 5. Let us assume the name of this kernel is K. The kernel K exists in the set of predefined kernels of convolutional layers 2, 15, 28, 41, and 54. The updated weight after the training of kernel K is shown in the second and third rows. The variances of the predefined kernel K after training for layers 2, 15, 28, 41, and 54 are 4.98, 4.18, 3.49, 3.51, and 3.59, respectively. The learning rate of the convolutional layers with predefined kernels is higher than the global learning rate. The higher learning rate of the convolutional layers with predefined kernels assures their adaptability in terms of network training by maintaining the primary attributes of high-pass kernels.



| | | | | |
|---|---|---|---|---|
| −1 | 2 | −2 | 2 | −1 |
| 2 | −6 | 8 | −6 | 2 |
| −2 | 8 | −12 | 8 | −2 |
| 2 | −6 | 8 | −6 | 2 |
| −1 | 2 | −2 | 2 | −1 |

| | | | | |
|---|---|---|---|---|
| −0.740 | 1.528 | −1.538 | 1.546 | −0.756 |
| 1.534 | −4.950 | 6.745 | −4.939 | 1.547 |
| −1.521 | 6.746 | −10.345 | 6.792 | −1.540 |
| 1.541 | −4.942 | 6.782 | −5.033 | 1.559 |
| −0.763 | 1.532 | −1.567 | 1.536 | −0.782 |

| | | | | |
|---|---|---|---|---|
| −0.401 | 0.973 | −0.893 | 0.954 | −0.437 |
| 1.020 | −3.914 | 5.652 | −4.026 | 0.901 |
| −0.894 | 5.704 | −9.456 | 5.632 | −0.923 |
| 0.958 | −3.986 | 5.715 | −3.858 | 0.979 |
| −0.410 | 0.989 | −0.917 | 1.001 | −0.360 |

| | | | | |
|---|---|---|---|---|
| −0.416 | 1.078 | −1.049 | 1.075 | −0.418 |
| 1.077 | −3.968 | 5.695 | −4.011 | 1.035 |
| −1.023 | 5.678 | −9.404 | 5.710 | −1.0329 |
| 1.026 | −3.956 | 5.693 | −3.989 | 1.040 |
| −0.405 | 1.031 | −1.035 | 1.052 | −0.442 |

| | | | | |
|---|---|---|---|---|
| −0.307 | 0.958 | −0.936 | 0.984 | −0.286 |
| 0.930 | −4.191 | 5.932 | −4.157 | 0.997 |
| −0.902 | 5.885 | −9.619 | 5.809 | −0.882 |
| 0.964 | −4.116 | 5.899 | −4.136 | 0.958 |
| −0.300 | 0.969 | −0.922 | 0.968 | −0.330 |

**Figure 5.** SRM kernel and updated predefined kernel weights after training.

In Figure 6, a single constant linear kernel is incorporated within a set of predefined kernels. This particular kernel serves to retain unprocessed information and is complemented by the inclusion of thirty high-pass kernels, which focus on extracting details from the input data of the preceding layers.

| | | | | |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 1 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 |

**Figure 6.** Constant linear kernels.

Transfer learning is used to enhance the deep network's robustness. The weights of the deep network for the classification of low-payload stego-images and cover images are initialized using the weights of the trained network from high-payload stego-images. As illustrated in Figure 7, the proposed deep network (Figure 4) follows two steps. In the first step, the pre-training is performed using high-payload stego-images. In the second step, the network is fine-tuned on low-payload images using the weight initialization from the pre-training network weights. The transfer learning approach improves detection accuracy substantially, especially for low-payload stego-images.

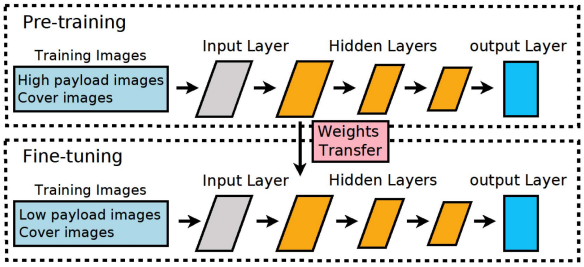



**Figure 7.** Layout of transfer learning approach.

## 3. Experimental Results

Images are popularly used for covert communication. Image steganography schemes can hide a message without any visual distortion. There is a very insignificant change in stage image statistics compared to cover image statistics when using content-adaptive steganography schemes. The experimental analysis is performed on a system of configuration Intel i7-10700K CPU @ 3.80 GHz, NVIDIA GeForce RTX 3070 GPU, and 32 GB RAM. Generally, the BOSSBase [41] and BOWS2 [6] datasets are used to assess the steganalysis schemes' robustness. As prepared in most previous schemes, images of size 512 × 512 from both datasets are converted into 256 × 256. The experimental outputs of the proposed scheme are displayed with other state-of-the-art techniques. Ye-Net [28], SRNet [29], Yedroudj-Net [31], and Zhu-Net [33] schemes are taken into consideration for the fair evaluation of the proposed scheme. Two types of datasets are created for the experimental analysis. The proposed scheme classifies images into two classes, i.e., the cover and stego class. Stego-images are formed by applying the HILL, Mi-POD, S-UNIWARD, and WOW steganography schemes. The stego-images are produced using payloads of 0.2 bpp, 0.3 bpp, and 0.4 bpp. The results are displayed in terms of detection accuracy.

In the first scenario, fifty percent of the images of the BOSSBase dataset cover and stego-images are used for training, and the remaining fifty percent of images are taken for testing. The experimental results are displayed in Tables 1–4 for scenario one. The Leaky ReLU has a significant effect on detection accuracy in comparison to the ReLU. There is an advantage of more than 2% in detection accuracy, as seen in Table 1. The Leaky ReLU is also effective in generative adversarial-based steganography [42] and image resolution enhancement [43].

**Table 1.** Detection accuracy, ReLU vs. Leaky ReLU comparison.

| Layer/Payload (bpp) | HILL | | | Mi-POD | | | S-UNIWARD | | | WOW | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.2 | 0.3 | 0.4 | 0.2 | 0.3 | 0.4 | 0.2 | 0.3 | 0.4 | 0.2 | 0.3 | 0.4 |
| ReLU | 58.32 | 62.95 | 67.34 | 61.19 | 69.53 | 71.60 | 63.60 | 72.81 | 77.77 | 69.64 | 76.69 | 79.14 |
| Leaky ReLU | 60.49 | 65.98 | 71.18 | 64.54 | 72.96 | 74.58 | 66.95 | 76.48 | 81.43 | 72.92 | 79.80 | 82.69 |

The proposed scheme involves using 31 kernels within the convolutional layers as part of a predefined set. Among these 31 kernels, 30 are SRM kernels, while the remaining 1 is a constant linear kernel. It is worth noting that only 30 SRM kernels were utilized in many prior schemes. The noteworthy aspect of this proposed approach is that it has resulted

in a discernible enhancement in detection accuracy compared to the conventional use of 30 kernels. The specific details of this improvement, including the performance metrics and results, can be found in Table 2. This table provides a clear visual representation of how the inclusion of the additional kernel has positively impacted the detection accuracy, highlighting the effectiveness of the proposed scheme.

**Table 2.** Constant linear kernel effect.

| Number of Kernels/Payload (bpp) | HILL | | | Mi-POD | | | S-UNIWARD | | | WOW | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **0.2** | **0.3** | **0.4** | **0.2** | **0.3** | **0.4** | **0.2** | **0.3** | **0.4** | **0.2** | **0.3** | **0.4** |
| 30 | 59.34 | 64.33 | 69.76 | 63.58 | 71.72 | 73.31 | 65.81 | 74.64 | 79.72 | 71.53 | 78.05 | 81.37 |
| 31 | 60.49 | 65.98 | 71.18 | 64.54 | 72.96 | 74.58 | 66.95 | 76.48 | 81.43 | 72.92 | 79.80 | 82.69 |

Researchers leverage the global average pooling (GAP) layer to enhance performance in steganalysis and various other applications. Including the GAP layer has proven to be instrumental in improving detection performance, substantiated by the results presented in Table 3. In our proposed scheme, the utilization of pooling layers is notably minimal, with the sole exception being the GAP layer. This deliberate choice aims to harness and retain the maximum available information, ensuring that valuable data are not lost during processing.

**Table 3.** GAP layer effect.

| GAP Layer Effect/Payload (bpp) | HILL | | | Mi-POD | | | S-UNIWARD | | | WOW | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **0.2** | **0.3** | **0.4** | **0.2** | **0.3** | **0.4** | **0.2** | **0.3** | **0.4** | **0.2** | **0.3** | **0.4** |
| Without GAP | 59.59 | 64.14 | 69.83 | 63.45 | 71.35 | 73.41 | 65.28 | 74.87 | 79.80 | 71.53 | 78.45 | 80.63 |
| GAP | 60.49 | 65.98 | 71.18 | 64.54 | 72.96 | 74.58 | 66.95 | 76.48 | 81.43 | 72.92 | 79.80 | 82.69 |

Iterative learning is also carried out using S-UNIWARD with a payload of 0.4 bpp. Seven iterations of network training are applied to the same training set. However, the detection accuracy improved only for the training set, i.e., for seen images. Furthermore, for the testing set, the detection accuracy did not increase. Overfitting may be the reason for a reduction in testing accuracy. In Figure 8, on the first row, an image is processed by one of the SRM kernels (Figure 8 first row). The effect of seven iterations of layer fifteen can be seen in the second row.

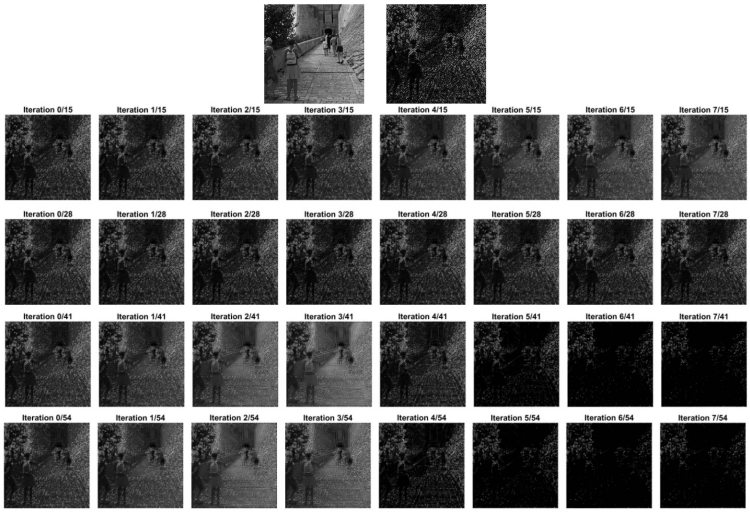



**Figure 8.** Iterative learning effects.

Similarly, the effect of seven iterations on layers 28, 41, and 54 is displayed in the third, fourth, and fifth rows, respectively. However, iterative learning is found beneficial for the 0.2 bpp low payloads. The network is fine-tuned using the previously trained network weights on the same training set. There is an improvement in testing accuracy while considering three iterations. After that, there is a reduction in the testing accuracy.

In Table 4, the experimental outputs of the proposed scheme are compared with the Ye-Net [28], SRNet [29], Yedroudj-Net [31], and Zhu-Net [33] techniques. Zhu-Net used a bottleneck approach and obtained the best results compared to previous schemes. The proposed method provides the best detection performance with multiple improvements in the deep network architecture.

**Table 4.** Performance comparison for scenario 1 with other schemes.

| Steganalysis Scheme/Payload (bpp) | HILL | | | Mi-POD | | | S-UNIWARD | | | WOW | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **0.2** | **0.3** | **0.4** | **0.2** | **0.3** | **0.4** | **0.2** | **0.3** | **0.4** | **0.2** | **0.3** | **0.4** |
| Ye-Net | 52.00 | 56.65 | 61.39 | 55.70 | 60.79 | 63.73 | 57.87 | 65.38 | 71.34 | 66.10 | 69.22 | 75.61 |
| SRNet | 53.15 | 60.62 | 65.07 | 56.41 | 64.25 | 69.92 | 62.97 | 71.31 | 75.73 | 69.50 | 75.80 | 80.36 |
| Yedroudj-Net | 51.79 | 56.39 | 64.96 | 55.86 | 60.47 | 68.43 | 57.61 | 66.07 | 70.40 | 67.32 | 71.72 | 77.15 |
| Zhu-Net | 59.47 | 64.88 | 69.08 | 64.36 | 70.00 | 72.46 | 66.42 | 73.88 | 78.48 | 70.99 | 74.31 | 81.64 |
| Proposed Scheme | 60.49 | 65.98 | 71.18 | 64.54 | 72.96 | 74.58 | 66.95 | 76.48 | 81.43 | 72.92 | 79.80 | 82.69 |

Several previous schemes discuss results with a single dataset and two datasets. Earlier in this paper, results were discussed using only images of the BOSSBase dataset from scenario 1. In the second scenario, two datasets are considered: fifty percent of the images in both the BOSSBase and BOWS2 datasets are used for training, and the remaining fifty percent are taken for testing. The proposed scheme also provides the best results for scenario 2, as illustrated in Table 5. As the results show, the detection accuracy of the proposed and previous methods is highest for WOW and lowest for HILL. This fact can also be visualized in the computer-generated cover–stego residual array images in Figure 2.

**Table 5.** Performance comparison for scenario 2 with other schemes.

| Steganalysis Scheme/Payload (bpp) | HILL | | | Mi-POD | | | S-UNIWARD | | | WOW | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **0.2** | **0.3** | **0.4** | **0.2** | **0.3** | **0.4** | **0.2** | **0.3** | **0.4** | **0.2** | **0.3** | **0.4** |
| Ye-Net | 52.85 | 58.68 | 62.98 | 57.81 | 62.97 | 66.04 | 59.89 | 67.76 | 72.64 | 68.44 | 73.86 | 78.44 |
| SRNet | 56.87 | 63.42 | 66.59 | 59.06 | 66.05 | 71.41 | 64.23 | 73.33 | 77.55 | 73.08 | 79.23 | 83.93 |
| Yedroudj-Net | 54.57 | 60.27 | 65.54 | 57.91 | 62.44 | 70.80 | 60.05 | 68.22 | 72.71 | 69.54 | 76.65 | 81.32 |
| Zhu-Net | 61.44 | 66.54 | 73.12 | 65.87 | 71.73 | 73.82 | 69.18 | 77.41 | 80.89 | 74.63 | 80.99 | 87.46 |
| Proposed Scheme | 62.26 | 69.69 | 74.74 | 67.53 | 73.10 | 76.18 | 70.90 | 79.08 | 82.01 | 77.36 | 82.28 | 88.13 |

## 4. Conclusions

Image steganalysis is a significant part of image forensics. Image steganalysis serves two purposes: it can verify the robustness of image steganography techniques and avert the misuse of stego-images. In this paper, a new steganalysis scheme using a convolutional neural network was proposed. A series of convolutional layers were directed by a convolutional layer with customized kernels, where firm cohesion was offered fully in the directed training. The learning rate of the convolutional layers with predefined kernels was made higher than the global learning rate to adjust the weights according to the network training rapidly. The proposed deep network was fine-tuned using the transfer learning approach. The proposed scheme was applied to stego-images generated from the HILL, Mi-POD, S-UNIWARD, and WOW content-adaptive steganography schemes with multiple

payloads, and the results were superior to the existing steganalysis schemes. Although the suggested scheme performed well with the specified steganography techniques, its ability to generalize to different methods should be evaluated. The effectiveness of deep learning models is significantly influenced by the quality and diversity of their training data, making it essential to incorporate a more extensive and diverse dataset while keeping the computational cost in check.

**Author Contributions:** Methodology, S.A. and H.K.; software, S.A.; writing—original draft, S.A.; writing—review & editing, H.K. and K.-H.J.; supervision, K.-H.J.; project administration, K.-H.J.; funding acquisition, K.-H.J. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Brain Pool program, funded by the Ministry of Science and ICT through the National Research Foundation of Korea (2019H1D3A1A01101687), and the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education (2021R1I1A3049788).

**Data Availability Statement:** The datasets used in this paper are publicly available and their links are provided in the reference section.

**Acknowledgments:** We thank the anonymous reviewers for their valuable suggestions that improved the quality of this article.

**Conflicts of Interest:** This manuscript is the author's original work and has not been published or submitted simultaneously elsewhere. There is no copyright issue in any of the figures.

## References

1. Hong, E.; Lim, K.; Oh, T.-W.; Jang, H. Lightweight image steganalysis with block-wise pruning. *Sci. Rep.* **2023**, *13*, 16148. [CrossRef]
2. Li, B.; Wang, M.; Huang, J.; Li, X. A new cost function for spatial image steganography. In Proceedings of the 2014 IEEE International Conference on Image Processing, ICIP 2014, Paris, France, 27–30 October 2014; pp. 4206–4210.
3. Sedighi, V.; Cogranne, R.; Fridrich, J. Content-Adaptive Steganography by Minimizing Statistical Detectability. *IEEE Trans. Inf. Forensics Secur.* **2016**, *11*, 221–234. [CrossRef]
4. Holub, V.; Fridrich, J.; Denemark, T. Universal distortion function for steganography in an arbitrary domain. *EURASIP J. Inf. Secur.* **2014**, *2014*, 1. [CrossRef]
5. Holub, V.; Fridrich, J. Designing steganographic distortion using directional filters. In Proceedings of the WIFS 2012—Proceedings of the 2012 IEEE International Workshop on Information Forensics and Security, Costa Adeje, Spain, 2–5 December 2012; pp. 234–239.
6. Bas, P.; Furon, T. Break Our Watermarking System. 2008. Available online: http://bows2.ec-lille.fr/ (accessed on 12 November 2020).
7. Agarwal, S.; Kim, C.; Jung, K.-H. Steganalysis of Context-Aware Image Steganography Techniques Using Convolutional Neural Network. *Appl. Sci.* **2022**, *12*, 10793. [CrossRef]
8. Arya, R.; Vimina, E.R. Local Triangular Coded Pattern: A Texture Descriptor for Image Classification. *IETE J. Res.* **2023**, *69*, 3267–3278. [CrossRef]
9. Agarwal, S.; Jung, K.-H. Median filtering detection using optimal multi-direction threshold on higher-order difference pixels. *Multimed. Tools Appl.* **2023**, *82*, 30875–30893. [CrossRef]
10. Liu, X.; Shi, T.; Zhou, G.; Liu, M.; Yin, Z.; Yin, L.; Zheng, W. Emotion classification for short texts: An improved multi-label method. *Humanit. Soc. Sci. Commun.* **2023**, *10*, 306. [CrossRef]
11. Xiong, G.; Ping, X.; Zhang, T.; Hou, X. Image textural features for steganalysis of spatial domain steganography. *J. Electron. Imaging* **2012**, *21*, 033015. [CrossRef]
12. Fridrich, J.; Kodovsky, J. Rich Models for Steganalysis of Digital Images. *IEEE Trans. Inf. Forensics Secur.* **2012**, *7*, 868–882. [CrossRef]
13. Tang, W.; Li, H.; Luo, W.; Huang, J. Adaptive steganalysis against WOW embedding algorithm. In Proceedings of the 2nd ACM Workshop on Information Hiding and Multimedia Security—IH&MMSec'14, Salzburg, Austria, 11–13 June 2014; pp. 91–96.
14. Denemark, T.; Sedighi, V.; Holub, V.; Cogranne, R.; Fridrich, J. Selection-channel-aware rich model for Steganalysis of digital images. In Proceedings of the 2014 IEEE International Workshop on Information Forensics and Security, WIFS 2014, Atlanta, GA, USA, 3–5 December 2014; pp. 48–53.
15. Xu, X.; Dong, J.; Wang, W.; Tan, T. Local correlation pattern for image steganalysis. In Proceedings of the 2015 IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP), Chengdu, China, 12–15 July 2015; pp. 468–472.
16. Li, F.; Zhang, X.; Cheng, H.; Yu, J. Digital image steganalysis based on local textural features and double dimensionality reduction. *Secur. Commun. Netw.* **2016**, *9*, 729–736. [CrossRef]
17. Li, B.; Li, Z.; Zhou, S.; Tan, S.; Zhang, X. New Steganalytic Features for Spatial Image Steganography Based on Derivative Filters and Threshold LBP Operator. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 1242–1257. [CrossRef]

18. Li, B.; Ming Wang, M.; Li, X.; Tan, S.; Huang, J. A Strategy of Clustering Modification Directions in Spatial Image Steganography. *IEEE Trans. Inf. Forensics Secur.* **2015**, *10*, 1905–1917. [CrossRef]
19. Wang, P.; Liu, F.; Yang, C. Towards feature representation for steganalysis of spatial steganography. *Signal Process.* **2020**, *169*, 107422. [CrossRef]
20. Ge, H.; Hu, D.; Xu, H.; Li, M.; Zheng, S. New Steganalytic Features for Spatial Image Steganography Based on Non-Negative Matrix Factorization. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: New York, NY, USA, 2020; Volume 12022 LNCS, pp. 337–351.
21. Yang, S.; Li, Q.; Li, W.; Li, X.; Liu, A.-A. Dual-Level Representation Enhancement on Characteristic and Context for Image-Text Retrieval. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 8037–8050. [CrossRef]
22. Wang, Y.; Su, Y.; Li, W.; Xiao, J.; Li, X.; Liu, A.-A. Dual-Path Rare Content Enhancement Network for Image and Text Matching. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *33*, 6144–6158. [CrossRef]
23. Qian, Y.; Dong, J.; Wang, W.; Tan, T. Deep learning for steganalysis via convolutional neural networks. In Proceedings of the Media Watermarking, Security, and Forensics 2015, San Francisco, CA, USA, 4 March 2015; p. 94090J.
24. Qian, Y.; Dong, J.; Wang, W.; Tan, T. Learning and transferring representations for image steganalysis using convolutional neural network. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; p. 7532860.
25. Xu, G.; Wu, H.-Z.; Shi, Y.-Q. Structural Design of Convolutional Neural Networks for Steganalysis. *IEEE Signal Process. Lett.* **2016**, *23*, 708–712. [CrossRef]
26. Wu, S.; Zhong, S.H.; Liu, Y. Steganalysis via deep residual network. In Proceedings of the International Conference on Parallel and Distributed Systems—ICPADS, Wuhan, China, 13–16 December 2016; pp. 1233–1236.
27. Wu, S.; Zhong, S.; Liu, Y. Deep residual learning for image steganalysis. *Multimed. Tools Appl.* **2017**, *77*, 10437–10453. [CrossRef]
28. Ye, J.; Ni, J.; Yi, Y. Deep Learning Hierarchical Representations for Image Steganalysis. *IEEE Trans. Inf. Forensics Secur.* **2017**, *12*, 2545–2557. [CrossRef]
29. Boroumand, M.; Chen, M.; Fridrich, J. Deep Residual Network for Steganalysis of Digital Images. *IEEE Trans. Inf. Forensics Secur.* **2019**, *14*, 1181–1193. [CrossRef]
30. Guo, L.; Ni, J.; Shi, Y. Uniform Embedding for Efficient JPEG Steganography. *IEEE Trans. Inf. Forensics Secur.* **2014**, *9*, 814–825. [CrossRef]
31. Yedroudj, M.; Comby, F.; Chaumont, M. Yedrouj-Net: An efficient CNN for spatial steganalysis. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 2092–2096. [CrossRef]
32. Wu, S.; Zhong, S.; Liu, Y. A Novel Convolutional Neural Network for Image Steganalysis with Shared Normalization. *IEEE Trans. Multimed.* **2020**, *22*, 256–270. [CrossRef]
33. Zhang, R.; Zhu, F.; Liu, J.; Liu, G. Depth-Wise Separable Convolutions and Multi-Level Pooling for an Efficient Spatial CNN-Based Steganalysis. *IEEE Trans. Inf. Forensics Secur.* **2020**, *15*, 1138–1150. [CrossRef]
34. Fu, T.; Chen, L.; Fu, Z.; Yu, K.; Wang, Y. CCNet: CNN model with channel attention and convolutional pooling mechanism for spatial image steganalysis. *J. Vis. Commun. Image Represent.* **2022**, *88*, 103633. [CrossRef]
35. Xiang, Z.; Sang, J.; Zhang, Q.; Cai, B.; Xia, X.; Wu, W. A New Convolutional Neural Network-Based Steganalysis Method for Content-Adaptive Image Steganography in the Spatial Domain. *IEEE Access* **2020**, *8*, 47013–47020. [CrossRef]
36. Wang, Z.; Chen, M.; Yang, Y.; Lei, M.; Dong, Z. Joint multi-domain feature learning for image steganalysis based on CNN. *Eurasip J. Image Video Process.* **2020**, *2020*, 28. [CrossRef]
37. Atamna, M.; Tkachenko, I.; Miguet, S. Improving Generalization in Facial Manipulation Detection Using Image Noise Residuals and Temporal Features. In Proceedings of the 2023 IEEE International Conference on Image Processing (ICIP), Kuala Lumpur, Malaysia, 8–11 October 2023; pp. 3424–3428.
38. Tan, S.; Chen, B.; Zeng, J.; Li, B.; Huang, J. Hybrid deep-learning framework for object-based forgery detection in video. *Signal Process. Image Commun.* **2022**, *105*, 116695. [CrossRef]
39. Xavier Glorot, Y.B. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the 13th International Conference on Artificial Intelligence and Statistics, Sardinia, Italy, 13–15 May 2010; pp. 249–256.
40. Xu, G. Deep convolutional neural network to detect J-UNIWARD. In Proceedings of the IH&MMSec 17: Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security, June 2017.
41. Bas, P.; Filler, T.; Pevný, T. "Break Our Steganographic System": The Ins and Outs of Organizing BOSS. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin/Heidelberg, Germany, 2011; Volume 6958 LNCS, pp. 59–70.
42. Zhang, R.; Dong, S.; Liu, J. Invisible steganography via generative adversarial networks. *Multimed. Tools Appl.* **2019**, *78*, 8559–8575. [CrossRef]
43. Jebadurai, J.; Jebadurai, I.J.; Paulraj, G.J.L.; Samuel, N.E. Learning Based Resolution Enhancement of Digital Images. *Int. J. Eng. Adv. Technol.* **2019**, *8*, 3026–3030. [CrossRef]