applied sciences

MDPI

Article

An Ensemble Transfer Learning Model for Detecting Stego Images

Dina Yousif Mikhail [1], Roojwan Sc Hawezi [1] and Shahab Wahhab Kareem [1,2,*]

[1] Information System Engineering Department, Technical Engineering College, Erbil Polytechnic University, Erbil 44001, Iraq
[2] Department of Information Technology, College of Engineering and Computer Science, Lebanese French University, Erbil 44001, Iraq
* Correspondence: shahab.kareem@epu.edu.iq

Abstract: As internet traffic grows daily, so does the need to protect it. Network security protects data from unauthorized access and ensures their confidentiality and integrity. Steganography is the practice and study of concealing communications by inserting them into seemingly unrelated data streams (cover media). Investigating and adapting machine learning models in digital image steganalysis is becoming more popular. It has been demonstrated that steganography techniques used within such a framework perform more securely than do techniques using hand-crafted pieces. This work was carried out to investigate and examine machine learning methods' critical contributions and beneficial roles. Machine learning is a field of artificial intelligence (AI) that provides the ability to learn without being explicitly programmed. Steganalysis is considered a classification problem that can be addressed by employing machine learning techniques and recent deep learning tools. The proposed ensemble model had four models (convolution neural networks (CNNs), Inception, AlexNet, and Resnet50), and after evaluating each model, the system voted on the best model for detecting stego images. Since active steganalysis is a classification problem that may be solved using active deep learning tools and modern machine learning methods, this paper's major goal was to analyze deep learning algorithms' vital roles and main contributions. The evaluation shows how to successfully detect images that contain a steganography algorithm that hides data in images. Thus, it suggests which algorithms work best, which need improvement, and which are easier to identify.

Keywords: deep learning; transfer learning; steganography; feature extraction; ensemble model; steganalysis; stego images

Citation: Mikhail, D.Y.; Hawezi, R.S.; Kareem, S.W. An Ensemble Transfer Learning Model for Detecting Stego Images. *Appl. Sci.* **2023**, *13*, 7021. https://doi.org/10.3390/app13127021

Academic Editor: Byung-Gyu Kim

Received: 28 February 2023
Revised: 30 April 2023
Accepted: 9 June 2023
Published: 11 June 2023

1. Introduction

As the quantity of traffic being moved and communicated over the internet in various formats, such as movies and photographs, increases daily, there is a rising concern about the security of the massive amount of data transferred over the internet, including passwords and personal, professional, and financial information, as well as social security numbers and other sensitive data. The need to keep this information safe is rising as well. Network security has become an integral part of today's sophisticated communication infrastructure to keep information private and prevent tampering [1,2].

Using machine learning to analyze and improve digital images, the practice of steganalysis is gaining in popularity. Steganography methods implemented inside such a framework are more secure than those employing manually constructed pricing. Steganalysis refers to the process of identifying hidden messages using steganography. In the world of cryptography, this is analogous to the practice of cryptanalysis. As a result, learning stenography is essential. A communication that has been intercepted and decrypted is present whereas cryptanalysis begins with a collection of suspicious data files without knowing which files, if any, hold a payload. The first step in steganalysis, which effectively

requires a forensic statistician, is to identify the data files that are most likely to have been tampered with in a massive, often exhaustive, collection [3].

There is a constant, growing requirement to safeguard the escalating volume of sensitive information exchanged and communicated online daily in various formats, including movies and photographs. The need for secure networks to protect sensitive data from prying eyes and malicious actors has made such security an integral aspect of today's communication infrastructure [4–8].

Cryptography, which is the practice of transforming plain text into encrypted text via an algorithm, has been widely used for decades to safeguard sensitive information. To read an original message, a recipient must first convert encrypted text into plain text (Smid and Branstad, 1988) [4]. To keep sensitive data safe, encryption techniques such as the advanced encryption standard (AES) (NIST-FIPS, 2001) and the data encryption standard (DES) (NIST-FIPS, 1977) are commonly utilized (Yegireddi and Kumar, 2016) [5]. The fact that encrypted communication can be read by anyone is seen as a weakness in cryptography, which translates secret messages into human-readable forms. Therefore, hackers on the internet may use the heat, tries, and attempts strategy to decipher code. Because of this shortcoming in cryptography, steganography was brought into the field of data protection to circumvent this problem by disguising the fact that a communication was taking place.

The motivation for detecting image steganography using deep learning arises from the increasing use of steganography in various applications, including data hiding and digital watermarking. Steganography involves hiding data within an image, and it can be used for both legitimate and malicious purposes, such as covert communications or concealing sensitive information. With the rise of deep learning techniques, there is an opportunity to develop more accurate and efficient steganalysis methods that can automatically detect hidden data within images. Deep learning models can learn complex features and patterns from large datasets, which can be used to identify steganography content in images.

The motivation for developing such methods is to improve the security and privacy of digital communication and prevent the malicious use of steganography. By detecting steganographic content in images, it is possible to identify potential threats and take appropriate measures to protect sensitive information. Additionally, deep-learning-based steganalysis can offer a better performance than can traditional steganalysis methods, which may require manual feature extraction and may not be as accurate.

The paper's organization is as follows. The introduction and problem statement are set out in the first section, and a literature review is provided in the second section. The third section discusses the proposed methods. The results and discussion are presented in the fourth section. The conclusion comprises the last section.

2. Literature Review

Article [3] discussed how to combine trained CDNs in a multimodal framework, and it examined their detection accuracy. The framework detects each classification modality independently and combines their estimations to create a universal image steganography detector. Six of the latest CDN-based image steganography detection techniques—GNCNN, IGNCNN, XuNet, YeNet, YedroudjNet, and the improved IGNCNN—were trained on stego images generated using WOW, SUNIWARED, and HILL steganography algorithms with payloads of 0.2, 0.3, and 0.4 bits per pixel. Due to the projected similarities between the image steganography systems, the detection accuracy decreased slightly. However, the multimodal image steganography detection based on the improved IGNCNN universal image steganography detection performed best compared to the other five examined detectors [3].

Article [9] discussed detecting steganography-modified JPEG images, and it analyzed image steganography detection using shallow and deep learning methods. Three common stenographic algorithms—JPEG universal wavelet relative distortion (J-Uniward), nsF5, and uniform embedding revisited distortion (UERD) at two density levels—hid information

in BOSS database photos. DCTR and GFR were the best feature spaces validated. At 0.4, the nsF5 algorithm detected bpnzac density with 99.9% accuracy, but the J-Uniward algorithm was barely detectable at 0.1 (a maximum of 56.3%). The study concluded that ensemble classifiers were a promising alternative to deep-learning-based detection [9].

In [10], the authors presented a deep-learning-based approach for steganography detection in digital images. The authors began by describing the importance of steganography detection in the field of digital forensics and highlighted the challenges associated with it. They then proposed a CNN-based model for detecting the presence of hidden data in digital images. The proposed model took as input the pixel values of an image and learned to identify the presence of steganography through a series of convolutional, pooling, and fully connected layers. The authors evaluated the performance of the proposed model on a dataset of stego images and showed that it outperformed existing steganography detection techniques in terms of accuracy, precision, and recall. The results of the study suggested that deep-learning-based approaches can be effective for steganography detection in digital images and can help improve the accuracy and reliability of forensic investigations.

In [11], the authors began by describing the importance of steganalysis in digital forensics and highlighted the limitations of traditional steganalysis techniques. They then proposed a deep-learning-based model for steganalysis that used a combination of convolutional and fully connected neural networks. The proposed model was trained and evaluated on a dataset of stego images that contained spatially embedded hidden information. The authors showed that the proposed model outperformed existing steganalysis techniques in terms of accuracy and sensitivity to different types of spatial image steganography. They also performed a sensitivity analysis of the proposed model to evaluate the impacts of different hyper parameters and architecture choices on the model's performance. The results of the study suggested that deep-learning-based approaches can be highly effective for the steganalysis of spatially embedded hidden information in digital images, and that careful selection of hyper parameters and architecture choices can further improve the performance of a model [11].

The authors of [12] began by describing the importance of steganalysis in digital forensics and highlighted the limitations of traditional steganalysis techniques. They then proposed a deep-learning-based model for steganalysis that used a combination of non-local blocks and multi-channel convolutional networks to identify the presence of hidden information in an image. The proposed model was trained and evaluated on a dataset of stego images that contained spatially embedded hidden information. The authors showed that the proposed model outperformed existing steganalysis techniques in terms of accuracy, precision, and recall. They also showed that the proposed model could be used to localize the regions of an image that contained hidden information. The results of the study suggested that deep-learning-based approaches can be highly effective for the steganalysis of spatially embedded hidden information in digital images and that the proposed model can help improve the accuracy and efficiency of forensic investigations [12].

Article [13] presented a deep-learning-based approach for detecting steganography in color images. The authors began by describing the importance of steganalysis in digital forensics and highlighted the challenges associated with detecting hidden information in color images. They then proposed a multi-frequency residual convolutional neural network (MRF-CNN) for steganalysis that extracted features from different frequency components of an image and learned to identify the presence of hidden information. The proposed model was trained and evaluated on a dataset of stego color images and compared with existing steganalysis techniques. The authors showed that the proposed MRF-CNN model outperformed existing steganalysis techniques in terms of accuracy, precision, and recall. They also showed that the proposed model could be used to localize the regions of an image that contained hidden information. The results of the study suggested that deep-learning-based approaches, such as the proposed MRF-CNN model, can be highly effective for the steg analysis of images and can help improve the accuracy and efficiency of forensic investigations.

The authors of [14] began by describing the importance of hand movement identification in various applications, such as prosthetics and rehabilitation. They then proposed a machine-learning-based approach for identifying hand movements that involved the use of electromyography (EMG) signals recorded from muscles in the arm. The proposed approach used a combination of feature extraction techniques and classification algorithms to identify hand movements. The authors evaluated the proposed approach on a dataset of EMG signals recorded from multiple subjects and showed that the proposed approach achieved high accuracy in identifying hand movements. They also compared the performance of the proposed approach with that of existing approaches and showed that the proposed approach outperformed them. The results of the study suggested that machine-learning-based approaches can be highly effective for identifying hand movements and can have important applications in prosthetics and rehabilitation [14].

In [15], the authors presented a novel deep neural network for identifying contextual steganography methods. The suggested method employed a high-boost filter to reduce high-frequency noise while keeping low-frequency information intact. Thirty high-pass SRM filters were applied to the high-boost image, resulting in thirty high-boost SRM-filtered photos. The suggested CNN used two skip connections to simultaneously gather data from a large number of connections. Despite the standard ReLU layer, a cropped version was investigated. The convolutional neural network (CNN) was built using a bottleneck strategy for maximum efficiency. For comprehensive data persistence, only one layer of global average pooling was used. To further enhance the detection accuracy, SVM was used in place of the softmax classifier. Compared to state-of-the-art methods, the proposed method performed better in terms of detection accuracy and computational cost in the experiments. The HILL, S-UNIWARD, and WOW context-aware steganography algorithms were tested on the BOWS2 and BOSS base datasets, validating the suggested scheme [15].

An ensemble classifier was trained using rich features that detected hidden messages in images for 10 years. Recently, studies such as the one conducted by Xu et al. have shown that well-designed convolutional neural networks (CNN) can perform similarly to two-step machine learning algorithms.

This research proposed a CNN that outperformed the state of the art in error probability. The proposal was a creative combination of essential bricks from several studies and followed prior proposals. The CNN used a pre-processing filter bank, a Truncation activation function, five convolutional layers with batch normalization and a scale layer, and an adequately large, fully connected section. The CNN was trained using an enhanced database.

Our CNN was tested against the S-UNIWARD and WOW embedding algorithms, and it was compared to an ensemble classifier, a rich model, and two other CNN steganalysis methods.

An ensemble classifier trained with rich features was used for approximately 10 years to detect a concealed message in an image, and it showed that well-designed convolutional neural networks (CNN) can perform as well as two-step machine learning techniques can. This research proposed a CNN with a lower error of probability than that of the current state of the art. The proposal continued what had been offered recently and cleverly combined essential bricks from many articles. The CNN used a pre-processing filter bank, a truncation activation function, five convolutional layers with a batch normalization associated with a scale layer, and an appropriately large, fully connected section. An augmented database had also been utilized to train a CNN. The proposed CNN was experimentally evaluated against the S-UNIWARD and WOW embedding techniques, and it was compared to three additional methods: an ensemble classifier, a rich model, and two CNN steganalysis methods [16]. In [17], the authors examined how deep learning could improve web image prediction accuracy and performance. The researchers trained 36 CNN models on the same dataset using convolutional neural networks (e.g., ImageNet). Using a "real-world" binary image categorization application, they evaluated the pre-trained models. Eurasian

lynx (*Lynx lynx*) camera trap images from Croatia were used to classify wildlife photos. According to their analysis, the dataset was extremely uneven in terms of the percentage of shots that showed lynxes, and image quality varied greatly. Several steganography techniques can hide information in JPEG images by altering the discrete cosine transform (DCT) coefficients, even though most of these algorithms work in the spatial domain. To further reduce the likelihood of discovery, some algorithms employ content addictiveness to primarily embed data in less predictable places, making it harder to notice changes. We focused on these alterations since they are the most difficult to identify. Authors of past studies have selected nsF5 [18], JPEG universal wavelet relative distortion (J-Uniward) [19], and uniform embedding revisited distortion (UERD) [20] for their analyses.

3. Methodology

The procedure for the proposed method is shown below in Figure 1. The first step involves collecting images and mixing them between a normal image and a stego image. The second step pre-processes images to clean the low-quality images that serve as the system's entry point. In the final stage of pre-processing, images are segmented and then exposed to a scanning algorithm that extracts the features to create a dataset. The resulting dataset is then considered an input to the proposed categorization scheme. The automatic classification of detecting stego images is displayed in the flowchart in Figure 1.

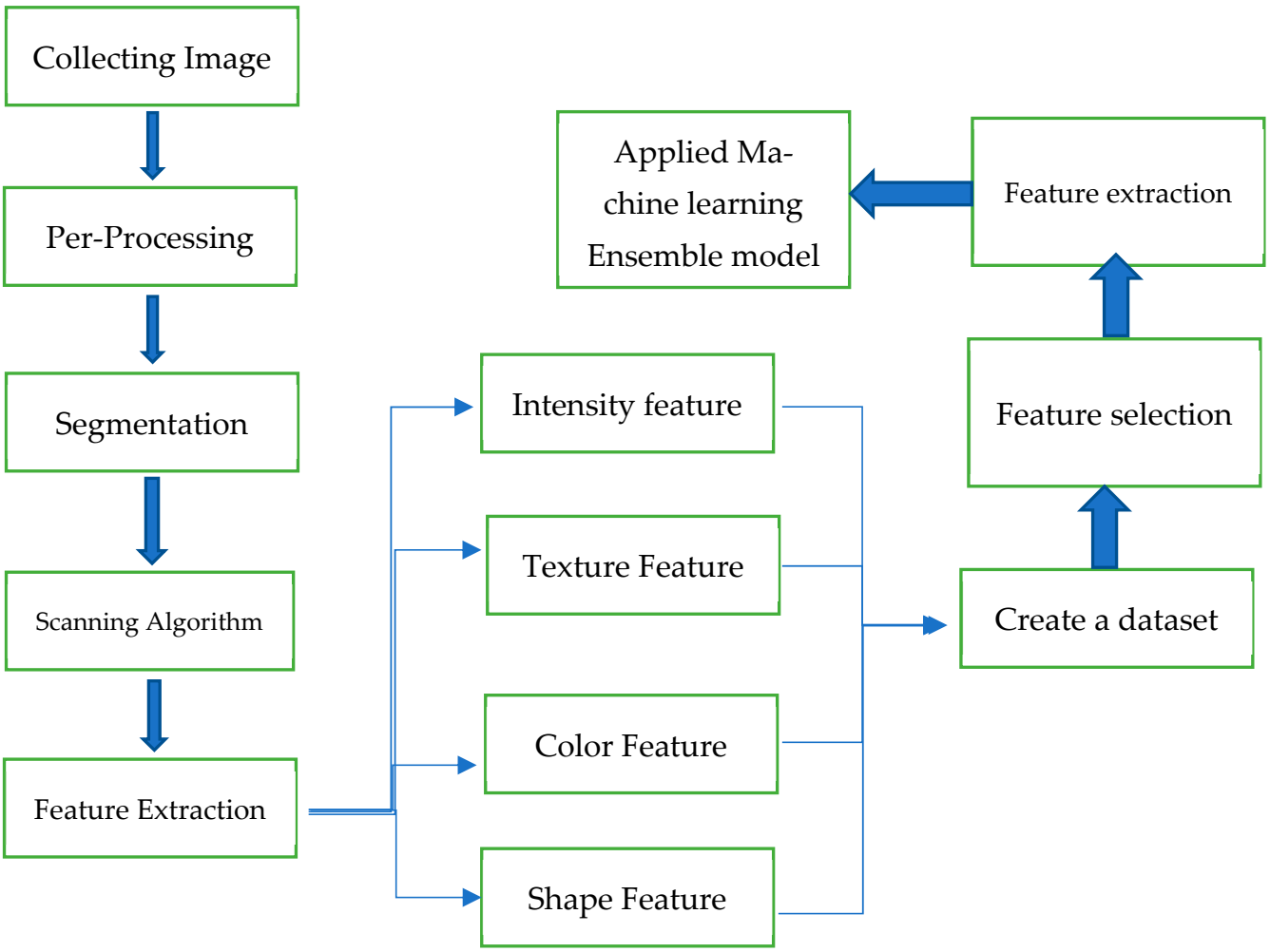

Figure 1. Procedure of the proposed method.

The procedure begins by applying two efficient steganography methods (edge-adaptive and HUGO). Figure 2 shows the steps of the edge-adaptive technique for hiding information in the images. Edge-adaptive steganography is a method that embeds secret information into digital images by exploiting the edges and textures in the image. The technique is based on the observation that the human visual system is less sensitive to changes in the high-frequency regions of an image, which typically correspond to edges and textures. The general flowchart for edge-adaptive steganography involves several steps, such as image pre-processing, feature extraction, the embedding of secret information, and post-processing. The exact details of each step may vary depending on the specific implementation of the technique. The edge-adaptive method uses paired pixels to conceal information. If there is strong evidence of steganography, as indicated by large absolute differences between these pairings, [21] then it can be concluded that steganography was used. The method begins by dividing the cover image into blocks and computing the edge information of each block using the Sobel operator. The authors proposed the use of edge information to improve the embedding process and make it less detectable via steganalysis methods.

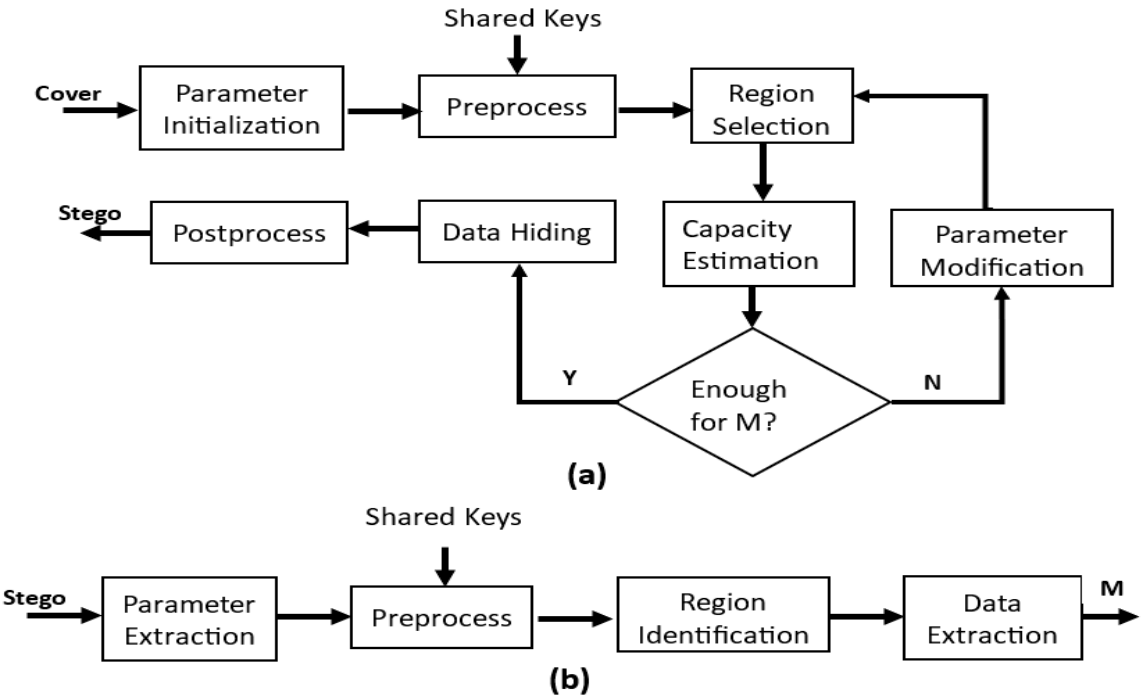

Figure 2. Edge adaptive image steganography (**a**) data embedding and (**b**) data extraction [21].

Specifically, we proposed selecting a block with edge information similar to that of the secret information block in which the secret information would be embedded. The embedding process itself was based on LSB (least significant bit) matching, where the LSBs of a selected block are modified to hide secret information. We also proposed the use of a dynamic quantization scheme to ensure that the embedding process did not degrade the perceptual quality of the cover image. We evaluated the proposed method using various metrics, including embedding capacity, perceptual quality, and security. The experimental results showed that the proposed method outperformed existing LSB-based steganography methods in terms of both embedding capacity and security while maintaining good perceptual quality. Overall, the proposed method improved an existing LSB-based steganography technique by incorporating the edge information of the cover image and using dynamic quantization to improve the embedding quality. The method was shown to have better performance than existing methods had in terms of embedding capacity and security while maintaining good perceptual quality.

The HUGO method was the first known steganography method to employ syndrome trellis codes [22]. This technique employs the difference between four neighbors (pixels) as a feature set to execute secret embedding with minimal distortion as shown in Figure 3. The obtained image enhancement step is often performed as part of the pre-processing stage and serves to make the initial image more suitable for further calculation. The next step is to convert the image into grayscale or to have only black and white tones because black includes information with an intensity value of "0", white contains information with an intensity value of "255", and grayscale images only carry intensity levels of between 0 and 255. By separating an image into its parts, lines, circles, and forms can be isolated. This is the process of disassembling an image. An image is either a collection of segments that cover the entire image or a group of contours drawn from the image. A region's pixels represent a computed property such as color, intensity, or texture. The same traits differ greatly between neighboring places. Segmentation is used to identify objects and boundaries in pictures (such as lines, curves, etc.). Image segmentation involves labeling each pixel in a picture to share visual features. Morphology, a prominent image processing method, changes shapes. By adding a structuring characteristic, morphological procedures create a similar-sized output image. Morphological operations compare each output pixel's value to its input image neighbors. The image structuring element's size and form determine how many pixels are added or subtracted from the image's objects. Before applying the suggested approach to an image, scanning algorithms usually scan the entire image. After pre-processing, the image is separated into binaries by scanning it vertically and horizontally. The sum of the rows must be larger than one to detect the topmost cell region pixel and to record an index when scanning an image vertically from top to bottom. After finding the index, scanning continues until the total of all rows is not zero, indicating the object's end. Algorithms repeat horizontal scanning to close structures. This improves the algorithm by speeding up processing. The classifier's performance is greatly affected by the quality of feature selection. Features must characterize each image subtype and be distinct from one another for a valid classification. All computed features can be partitioned into form features, intensity features, and texture features for easier comparison and assessment. To classify an item, we use its "feature vector", an n-dimensional vector containing a set of values reflecting various attributes, to feed to a classification algorithm, and data are extracted from images using feature extraction to produce a dataset of 22 features. For this study, 22 features were employed, including 5 co-occurrence matrices, the mean, the STD, the skewness, and the kurtosis, as well as 6 color moments and 7 moment-invariant properties. After collecting the data and storing them in a CSV file, the following phase was to employ a machine learning or deep learning technique to obtain the highest possible precision.

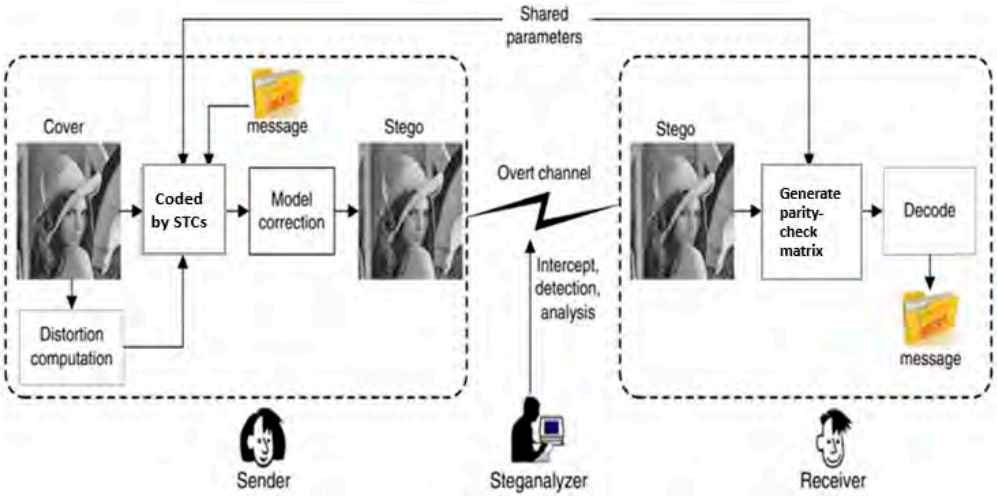

Figure 3. Covert communication based on HUGO steganography [23].

After collecting the data from the feature extraction, the authors selected 22 features that were highly correlated. Four models were applied as a classification method in the next step. In this study, we used the following techniques: CNN, AlexNet, Resnet50, and Inception; the figures below show the specifications used for each network (Figures 4–7). After collecting the evaluation for each model, the system voted for the best model. Deep learning ensemble models are used to increase a system's performance. Ensembles reduce prediction variance by integrating numerous model outputs, improving accuracy, reducing overfitting, and improving model generalization. Several models trained on distinct data subsets can capture diverse data features and increase system generalization, and ensemble makes a model more resilient to noise and outliers. Errors and outliers can be decreased by merging model outputs. The ensemble optimizes computing resources and enables the efficient and concurrent training of numerous smaller models instead of that of a single large model. It is a strong deep-learning technique that improves model performance, generalization, robustness, and computing efficiency.

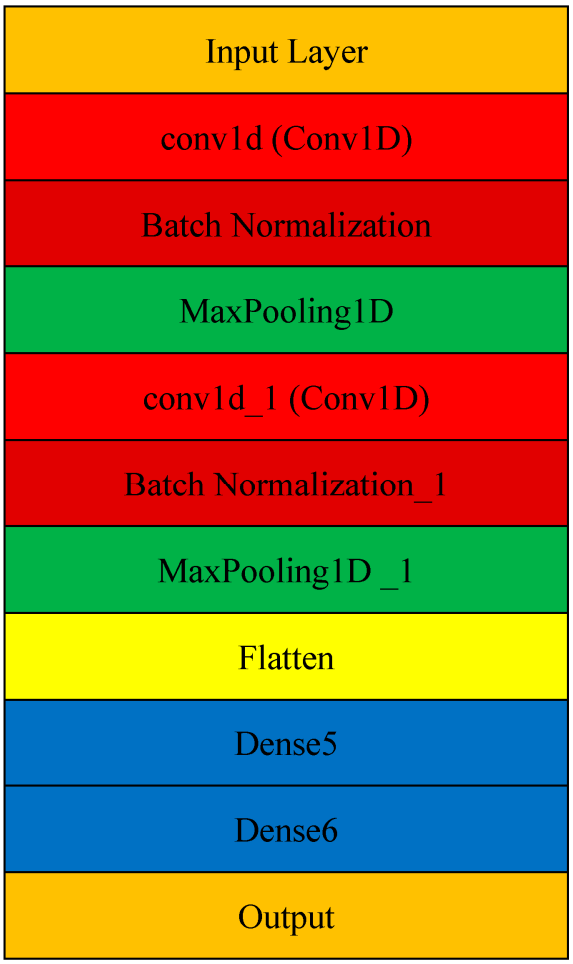

Figure 4. Diagram of a 1D CNN model.

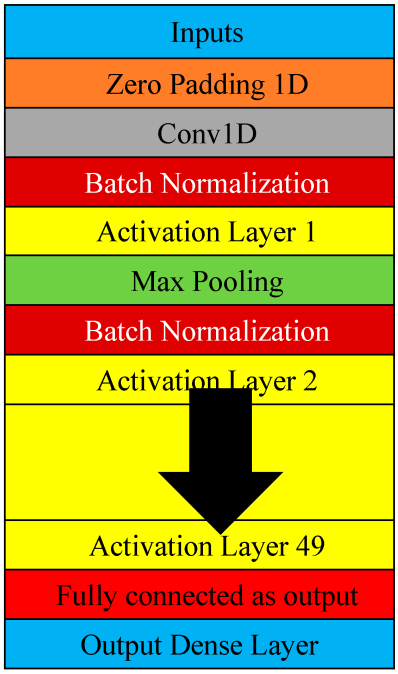

Figure 5. Diagram of an AlexNet model.

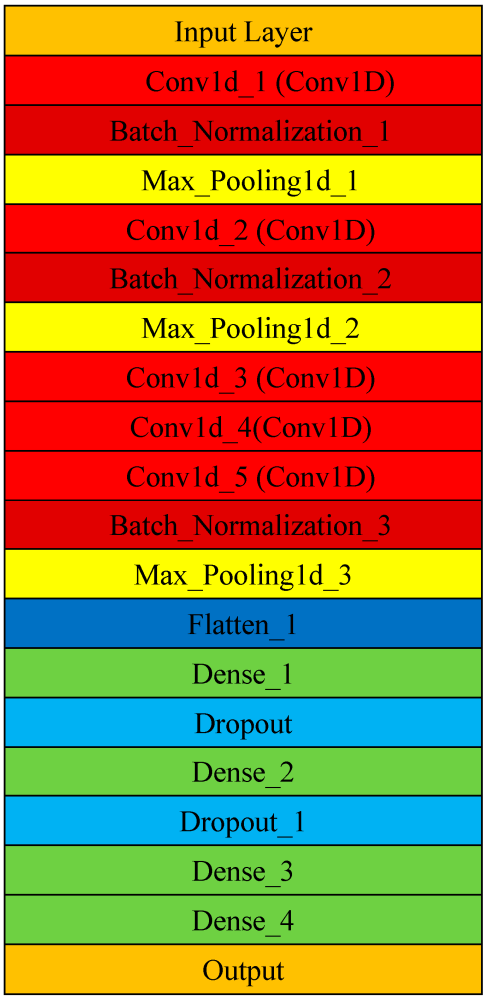

Figure 6. Diagram of a ResNet-50 model.

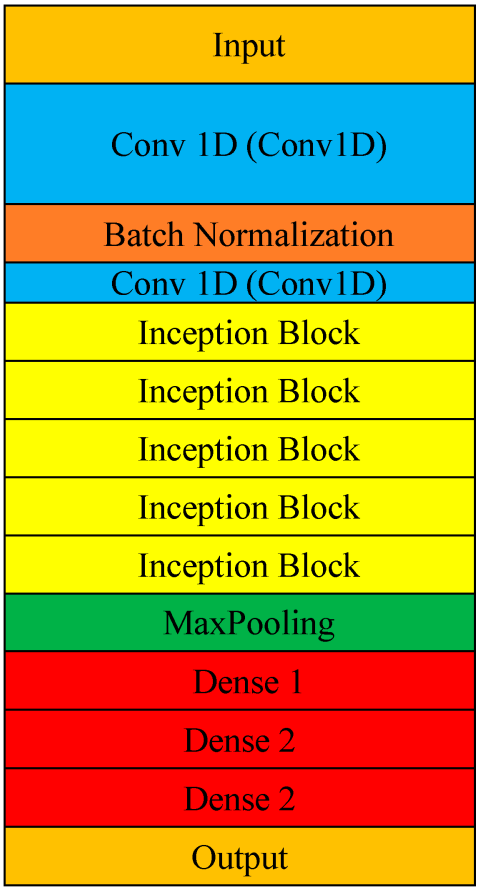

Figure 7. Diagram of an Inception model.

4. Results and Analysis

This study performed tests on a system with two Intel Xeon Silver processors, 256 GB RAM (though only 64 GB was used during training), and two Tesla V100Ss for the PCIe with double-precision 8.2 teraFLOPS cores. The fully linked convolutional layer weight decays were 0 and 0.01. The linearly decaying learning rate was 0.01–0.00001.

The standard BOSS base 1.01 picture dataset [23] was used for the tests. There are 10,000 512 × 512 grayscale cover images in BOSS base 1.01. Each image in the collection is broken into four smaller images, each of which is 256 pixels on the longest side. As part of the training process, the authors used a set of four sub-images (one for each image in the training set). Consequently, there were 40,000 images available for training for each DL-based classification modality for the steganography detection. Each dataset, totaling 40,000 individual images, was split in half, with each half containing 20,000 individual pictures. Each image steganography approach resulted in the exposure of half of the training data to the method, with the other half being used for the clear image for training. Each scenario was selected to expose the dataset to a unique steganography method, with a unique payload being expressed in bits per pixel (bpp).

This article discusses the edge-adaptive and HUGO steganography methods, which can conceal payloads of 0.2, 0.3, and 0.4 bpp. The dataset was used to train a distinct DL-based classification modality for image steganography detection and contained 20,000 stego images, together with the first half of the 20,000 clean photos. In the context of multimodal DL-based image steganography detection, this amounted to a four-pronged approach. Standard measures were used to evaluate the quality of the proposed models. The first was accuracy, which measures how accurately the dataset as a whole has been labeled. The second statistic, "precision", measures how close a classifier comes to correctly labeling a set of outcomes [24–30]. The percentage of a class's results that are detected by a

model is measured by a metric called recall. An F1 score, which is the harmonic mean of the precision and the recall, was the fourth statistic we considered. Lastly, we compared our method to other current DL-based methods for steganography identification using detection error rates as our metric. Table 1 shows the accuracy of each model based on the ppb. As shown in Table 1, the CNN model achieved a better result than did other models as it had a lower error detection rate. Table 1 is an evaluation metric for error detection in steganography, which is the practice of hiding information within another piece of data. This table compares the performance of various methods for detecting errors in steganography when different levels of payload (information to be hidden) are used.

Table 1. Error detection evaluation metrics.

Method	Payload bpp = 0.1	Payload bpp = 0.2	Payload bpp = 0.3	Payload bpp = 0.4
CNN	0.3624	0.3324	0.2501	0.1425
AlexNet	0.3925	0.3525	0.2845	0.2445
ResNet-50	0.3984	0.3484	0.2925	0.2531
Inception	0.3934	0.3434	0.2754	0.2465
VGG16Stego [11]	-	0.3428	-	0.2354
ANN [31]	0.3724	0.3547	0.3375	0.3245

The table lists different methods for error detection (CNN, AlexNet, ResNet-50, Inception, VGG16Stego, and ANN). For each method, the table shows the error detection rates (EDRs) for different payload levels (0.1, 0.2, 0.3, and 0.4 bits per pixel (bpp)). The EDR is a measure of the effectiveness of a method at detecting errors introduced during the process of steganography. A higher EDR indicates that the method is better at detecting errors while a lower EDR indicates that the method is less effective at detecting errors.

All of the methods listed in the table are different types of neural networks that are commonly used in steganography for error detection. CNN, AlexNet, ResNet-50, and Inception are types of convolutional neural networks while VGG16Stego and ANN are different types of artificial neural networks.

The results in the table show that all of the methods had higher EDRs when the payload was set to a lower level (0.1 or 0.2 bpp) and lower EDRs when the payload was set to a higher level (0.3 or 0.4 bpp). This was likely because higher payload levels introduce more errors into the steganography process, making them harder to detect.

Overall, the results suggested that ResNet-50 and Inception are the most effective methods for error detection, with consistently higher EDRs across all payload levels. However, VGG16Stego was also effective at lower payload levels (e.g., 0.2 bpp) while ANN had high EDRs across all payload levels.

Table 2 shows the accuracy, precision, recall, and F1 score of each model. Based on the results, the CNN model achieved a better result than other methods did.

Table 2. Evaluation metrics for the four proposed models.

Method	Accuracy	Precision	Recall	F1 Score
CNN	86%	84%	87%	84%
AlexNet	76%	75%	79%	79%
ResNet-50	75%	74%	78%	78%
Inception	76%	73%	72%	74%
VGG16Stego [11]	82 %	-	-	-
ANN [31]	75%	-	-	-

Based on the evaluation metrics, the CNN model achieved the best result compared to the other models. To separate the valuable characteristics from the noise in the raw signal, a CNN model was employed. By using the provided CNN, we were able to achieve optimal results for both accuracy and speed. It followed that the chosen CNN characteristics would

serve as good features for a method, which was in line with the current results. It was not necessary to complete a denoising step before implementing the proposed method.

Table 2 shows the evaluation metrics for the four proposed models (CNN, AlexNet, ResNet-50, and Inception) for image steganalysis. The evaluation metrics used were accuracy, precision, recall, and the F1 score.

Accuracy refers to the percentage of correctly classified steganography images out of the total number of images. Precision is the ratio of true positive results to the total number of positive results, which indicates the ability of a model to correctly identify steganography images. Recall is the ratio of true positive results to the total number of actual positive results, which indicates the ability of a model to detect all steganography images. The F1 score is the harmonic mean of the precision and the recall, and it provides a balance between the two metrics.

The results showed that the CNN model had the highest accuracy, recall, and F1 score, and AlexNet, ResNet-50, and Inception had lower accuracy and recall scores. However, the precision of the CNN model was slightly lower than that of the other models. This means that the CNN model was better at correctly identifying the steganography images and detecting all of the steganography images, but it may also have had a higher false positive rate than the other models had. The following reasons may be why the CNN model was the most successful of the four methods.

CNNs are very good at capturing spatial features from images, and this ability can also be useful for capturing sequential features from text data. In this case, the numerical representations of the text data could be treated as sequences of one-dimensional images, and a CNN could learn to capture relevant features from these sequences.

Pre-trained CNN models are already trained on large-scale image classification tasks and have learned to recognize a wide range of features that are useful for many different image-related tasks, including feature extraction from text data. The pre-trained CNN model could be fine-tuned using the text data to learn to recognize features that were specific to the text domain.

CNNs are capable of learning features hierarchically, which means they can learn simple features such as edges and corners in lower layers and more complex features such as textures and shapes in higher layers. This makes them very effective at feature extraction as they can learn to recognize both local and global features from the numerical representations of the text data.

Pre-trained CNN models are available in different architectures and can be fine-tuned to a specific task. This allows flexibility in choosing the best model architecture for a given task and can lead to improved performance.

Overall, the results suggested that the CNN model was the most effective for steganalysis among the four proposed models.

5. Conclusions

The study evaluated the performance of four deep learning models, namely, CNN, AlexNet, ResNet-50, and Inception, for detecting steganography at different payload bitrates using the BOSS base 1.01 dataset. The evaluation metrics used were accuracy, precision, recall, F1 score, and error detection rate. The results showed that the CNN model performed better than the other models did at lower payload bitrates while ResNet-50 and Inception were more robust at higher payload bitrates. However, the CNN model achieved the best results overall based on the evaluation metrics used. It is important to note that the results may not necessarily be generalized to other scenarios.

In conclusion, the proposed method for detecting stego images involved collecting images and mixing them between precise and stego images in the first step, followed by pre-processing to clean the low-quality images, segmenting the images, and extracting features to create a dataset. The resulting dataset was then used as an input for the proposed categorization scheme, which employed four classification models, namely, CNN, AlexNet, Resnet50, and Inception. The authors used 22 features, including co-occurrence matrices,

means, STDs, skewness, kurtosis values, color moments, and moment-invariant properties, for feature extraction. Evaluations of each model were performed, and the system voted for the best model. The proposed method has potential applications in digital forensics for identifying steganography in images. It is a useful tool that can be used by law enforcement agencies to detect steganography and prevent the misuse of sensitive information.

Author Contributions: Conceptualization, D.Y.M., R.S.H. and S.W.K.; methodology D.Y.M., R.S.H. and S.W.K.; software, S.W.K.; validation, D.Y.M., R.S.H. and S.W.K.; formal analysis, D.Y.M., R.S.H. and S.W.K.; investigation, D.Y.M., R.S.H. and S.W.K.; resources, S.W.K.; data curation, S.W.K.; writing—original draft preparation, S.W.K.; writing—review and editing, D.Y.M. and R.S.H.; visualization, D.Y.M., R.S.H. and S.W.K.; supervision, S.W.K.; project administration, R.S.H.; funding acquisition, D.Y.M. and R.S.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Butora, J.; Yousfi, Y.; Fridrich, J. How to pretrain for steganalysis. In Proceedings of the 2021 ACM Workshop on Information Hiding and Multimedia Security, Virtual, 21–25 June 2021; pp. 143–148.
2. Chaumont, M. Deep learning in steganography and steganalysis. In *Digital Media Steganography*; Academic Press: Cambridge, MA, USA, 2020; pp. 321–349.
3. Elshafey, M.A.; Amein, A.S.; Badran, K.S. Universal Image Steganography Detection using Multimodal Deep Learning Framework. *J. Inf. Hiding Multim. Signal Process.* **2021**, *12*, 152–161.
4. Smid, M.E.; Branstad, D.K. Data Encryption Standard: Past and future. *Proc. IEEE* **1988**, *76*, 550–559. [CrossRef]
5. Yegireddi, R.; Kumar, R.K. A survey on conventional encryption algorithms of Cryptography. In Proceedings of the 2016 International Conference on ICT in Business Industry & Government (ICTBIG), Indore, India, 18–19 November 2016; pp. 1–4.
6. Ozcan, S.; Mustacoglu, A.F. Transfer learning effects on image steganalysis with pre-trained deep residual neural network model. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; pp. 2280–2287.
7. Reinel, T.-S.; Brayan, A.-A.H.; Alejandro, B.-O.M.; Alejandro, M.-R.; Daniel, A.-G.; Alejandro, A.-G.J.; Buenaventura, B.-J.A.; Simon, O.-A.; Gustavo, I.; Raul, R.-P. GBRAS-Net: A convolutional neural network architecture for spatial image steganalysis. *IEEE Access* **2021**, *9*, 14340–14350. [CrossRef]
8. Selvaraj, A.; Ezhilarasan, A.; Wellington, S.L.J.; Sam, A.R. Digital image steganalysis: A survey on paradigm shift from machine learning to deep learning based techniques. *IET Image Process.* **2021**, *15*, 504–522. [CrossRef]
9. Płachta, M.; Krzemień, M.; Szczypiorski, K.; Janicki, A. Detection of Image Steganography Using Deep Learning and Ensemble Classifiers. *Electronics* **2022**, *11*, 1565. [CrossRef]
10. Reinel, T.S.; Raul, R.P.; Gustavo, I. Deep Learning Applied to Steganalysis of Digital Images: A Systematic Review. *IEEE Access* **2019**, *7*, 68970–68990. [CrossRef]
11. Tabares-Soto, R.; Arteaga-Arteaga, H.B.; Mora-Rubio, A.; Bravo-Ortíz, M.A.; Arias-Garzón, D.; Alzate-Grisales, J.A.; Orozco-Arias, S.; Isaza, G.; Ramos-Pollán, R. Sensitivity of deep learning applied to spatial image steganalysis. *PeerJ Comput. Sci.* **2021**, *7*, e616. [CrossRef] [PubMed]
12. Han, X.; Zhang, T. Spatial Steganalysis Based on Non-Local Block and Multi-Channel Convolutional Networks. *IEEE Access* **2022**, *10*, 87241–87253. [CrossRef]
13. Lin, J.; Yang, Y. Multi-Frequency Residual Convolutional Neural Network for Steganalysis of Color Images. *IEEE Access* **2021**, *9*, 141938–141950. [CrossRef]
14. Mora-Rubio, A.; Alzate-Grisales, J.A.; Arias-Garzón, D.; Buriticá, J.I.P.; Varón, C.F.J.; Bravo-Ortiz, M.A.; Arteaga-Arteaga, H.B.; Hassaballah, M.; Orozco-Arias, S.; Isaza, G.; et al. Multi-subject identification of hand movements using machine learning. In *Sustainable Smart Cities and Territories*; Corchado, J.M., Trabelsi, S., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 117–128.
15. Agarwal, S.; Kim, C.; Jung, K.-H. Steganalysis of Context-Aware Image Steganography Techniques Using Convolutional Neural Network. *Appl. Sci.* **2022**, *12*, 10793. [CrossRef]
16. Yedroudj, M.; Comby, F.; Chaumont, M. Yedroudj-Net: An Efficient Cnn for Spatial Steganalysis. In Proceedings of the IEEE ICASSP' 2018, Calgary, AB, Canada, 15–20 April 2018.

17. Stančić, A.; Vyroubal, V.; Slijepčević, V. Classification Efficiency of Pre-Trained Deep CNN Models on Camera Trap Images. *J. Imaging* **2022**, *8*, 20. [CrossRef] [PubMed]
18. Fridrich, J.; Pevný, T.; Kodovský, J. Statistically undetectable JPEG steganography: Dead ends, challenges, and opportunities. In Proceedings of the the 9th ACM Multimedia & SecurityWorkshop, Dallas, TX, USA, 20–21 September 2007; Association for Computing Machinery: New York, NY, USA, 2007; pp. 3–14.
19. Holub, V.; Fridrich, J.; Denemark, T. Universal distortion function for steganography in an arbitrary domain. *EURASIP J. Inf. Secur.* **2014**, *2014*, 1. [CrossRef]
20. Guo, L.; Ni, J.; Su, W.; Tang, C.; Shi, Y.Q. Using Statistical Image Model for JPEG Steganography: Uniform Embedding Revisited. *IEEE Trans. Inf. Forensics Secur.* **2015**, *10*, 2669–2680. [CrossRef]
21. Luo, W.; Huang, F.; Huang, J. Edge adaptive image steganography based on lsb matching revisited. *IEEE Trans. Inf. Forensics Secur.* **2010**, *5*, 201–214.
22. Filler, T.; Judas, J.; Fridrich, J. Minimizing additive distortion in steganography using syndrometrellis codes. *IEEE Trans. Inf. Forensics Secur.* **2011**, *6*, 920–935. [CrossRef]
23. Luo, X.; Song, X.; Li, X.; Zhang, W.; Lu, J.; Yang, C.; Liu, F. Steganalysis of HUGO steganography based on parameter recognition of syndrome-trellis-codes. *Multimedia Tools Appl.* **2015**, *75*, 13557–13583. [CrossRef]
24. Break Our Steganographic System Base Webpage (BossBase). Available online: http://agents.fel.cvut.cz/boss/ (accessed on 18 January 2022).
25. Ismael, S.H.; Kareem, S.W.; Almukhtar, F.H. Medical Image Classification Using Different Machine Learning Algorithms. *AL-Rafidain J. Comput. Sci. Math.* **2020**, *14*, 133–145. [CrossRef]
26. Pibre, L.; Pasquet, J.; Ienco, D.; Chaumont, M. Deep learning is a good steganalysis tool when embedding key is reused for different images, even if there is a cover source-mismatch. *Electron. Imaging* **2016**, *2016*, 1–11. [CrossRef]
27. Hussain, Z.S.; Danha, N.Y.; Muheden, K.M.; Kareem, S.W. Wind Speed Prediction for Duhok City Applied Recurrent Neural Network. *Int. J. Intell. Syst. Appl. Eng.* **2022**, *10*, 180–188.
28. 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, 8–12 February 2015.
29. Alattar, A.M.; Memon, N.D.; Heitzenrater, C.D. (Eds.) *International Society for Optics and Photonics*; SPIE: Bellingham, WA, USA, 2015; Volume 9409, pp. 171–180.
30. Awla, H.Q.; Kareem, S.W.; Mohammed, A.S. Bayesian Network Structure Discovery Using Antlion Optimization Algorithm. *Int. J. Syst. Innov.* **2022**, *7*, 46–65.
31. Mohamed, N.; Rabie, T.; Kamel, I.; Alnajjar, K. Detecting Secret Messages in Images Using Neural Networks. In Proceedings of the 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), Toronto, ON, Canada, 21–24 April 2021; pp. 1–6.