Proceeding Paper

Deep Learning-Based Coverless Image Steganography on Medical Images Shared via Cloud

Ambika 1,*, Virupakshappa 1, and Deepak S. Uplaonkar 2

1 Department of Computer Science & Engineering, Sharnbasva University, Kalaburagi 585103, Karnataka, India; virupakshi.108@gmail.com
2 Department of Computer Engineering, International Institute of Information Technology, Pune 411057, Maharashtra, India; uplaonkar@gmail.com
* Correspondence: ambika.umashetty@gmail.com

Abstract: Coverless image steganography is an approach for creating images with intrinsic colour and texture information that contain hidden secret information. Recently, generative adversarial networks' (GANs) deep learning transformers have been used to generate secret hidden images. Although it has been proven that this approach is resistant to steganalysis attacks, it modifies critical information in the images which makes the images not suitable for applications like disease diagnosis from medical images shared over cloud. The colour and textural modification introduced by GANs affects the feature vector which is extracted from certain image regions and used for disease diagnosis. To solve this problem, this work proposes an attention-guided GAN which transforms images only in certain regions and retains the originality of images in certain regions. Due to this, there is not much distortion to features and disease classification accuracy.

Keywords: steganography; generative adversarial networks (GANs); medical images; texture; hidden image

1. Introduction

Steganography is the technique of masking sensitive information inside an image and communicating it to the intended party [1]. Traditional image steganographic techniques modified the least significant bits of pixels or modified the frequency components in the discrete cosine transform of images to hide secret information. But these techniques are very weak against steganalysis attacks. Attackers employ various statistical correlation [2–4] and machine learning [5–7] techniques and deep learning techniques [8–10] to decipher the secret information hidden in images. Steganography without embedding (SWE) [11] is a recent technique for hiding information in images by modifying characteristics like colour, texture, edge, contours, pixel information, etc. This makes steganalysis difficult, enabling the secure transfer of secret information. Different SWE methods have been designed by changing the pixel intensity, colour, texture, edge, and contours. The secret information is mapped back to secret information. During the retrieval stage, the image characteristics are mapped back to secret information. SWE uses image hashes, texture synthesis mapping, and bag of words. However, these techniques have constrained steganography abilities and require a large database of images. Recently, deep learning transformers have been used to transform a cover image to a secret information image, and in particular GANs these provide better performance. GANs generate images close to natural images. It becomes difficult for attackers to decipher any information. Though GANs generate images close to natural ones, their use in medical images can distort discriminative features. Medical images like CT and MRI images have discriminative information at certain regions which is important in computer-aided disease diagnosis. GAN transformation does not...
consider these special regions and applies its transformation rules at a global level over the entire image. This distorts the features at the significant regions of the image carrying discriminative information. As a result, when these medical images are transferred to cloud-based disease diagnosis systems, their accuracy is reduced. This work addresses the problem of GANs distorting the discriminative features at certain specific regions of images and proposes a novel technique called attention vector-guided GAN transformation (AVG-GAN).

The proposed AVG-GAN provides an attention vector that contains information about where the transformation for secret information mapping should be performed. The same attention vector is required in the reverse operation of recovering secret information from the hidden secret image. The discriminative ability of the secret hidden image differs from the original image by a substantial factor. As a result, the AVG-GAN is more suitable for categorisation utility-preserving operations and computer-aided disease diagnosis. The hidden information in an image could contain copyright information, allowing data theft of images at the cloud end to be discovered. The following lists this work’s significant contributions:

(i) A novel attention vector-guided GAN for transformation of cover image without distorting specialised regions in the image;
(ii) Classification utility-preserving transformation without reducing the accuracy of disease diagnosis by a large factor in computer-aided disease diagnosis application.

2. Related Works

It has been speculated that SWE-based steganography, which creates a connection between sensitive messages and cover images, can be used to avoid machine learning-based steganography detection. Currently, SWE can be carried out by selecting and synthesising cover images. By combining images with secret information and collecting many images, an image library utilising a cover selection method is created. It is possible to map every message or group of pictures. Cover synthesis generates a new cover image based on secret information. By combining images with secret information and collecting many images, an image library utilising a cover selection method is created. It is possible to map every message or group of pictures. Cover synthesis generates a new cover image based on secret information. Zhou et al. [12] outlined a method for building a database of hashes derived from strong hashing techniques used to index pictures. The secret binary data consist of several segments. According to the techniques retrieved from the database, each segment is assigned an hash integer which is equal to its value. Zhou et al. proposed a bag-of-words (BOW) model as another way to estimate SWE [13]. Visual words are extracted from an image set using the BOW model; this allows text information keywords to be associated with image terms. An associated set of sub-images can be identified in accordance with the relationship in the mapping. Stego images with sub-images allow for secret communication. In [14], a robust image hash was proposed to link secret data to images. By matching local image hash values with secret data segments, data segments are created. Steganography capacity has been improved because of robust image hashing, making carrier images more resistant to attack. A steganography-based cover has two obvious disadvantages: it has inadequate steganography capabilities and demands a substantial local image database. It is possible to create texture images with secret information by combining methods from based on a synthesis approach. Otori et al. [15] used dotted patterns to encode then hide them in a similarly textured image to maintain acceptable quality. Xu et al. [16] presented stego texture, a technique for creating highly detailed textures from images or text messages. By applying a reversible mathematical function to a given message, one may generate a stego texture instantaneously, allowing it to be deciphered. Wu and Wang [17] created a new texture image by inverse texture reconstruction that retained the original look on a local level, but with an arbitrary size. Using texture, steganography hides message synthesis. In spite of their state-of-the-art status, cover synthesis-based schemes have one common weakness: the SWE schemes use special images (such as texture images), and sending them in a huge quantity will alert guards. Hayes et al. [18] designed a GAN-based steganography that is similarly efficient to previous steganography techniques, along with
steganalysis tools for detecting hidden data. In the experiments, the trained steganography service failed 50% of the time to distinguish the messages.

Volkhonskiy et al. [19] discussed using SGANs for steganography due to the resistance of generated images against detection as well as their authenticity. Using secure cover images to deceive steganography analysis is possible when a cover image container is trained. Because of this LSB match-embedding method, SGANs have become steganography containers. When embedding schemes differ, SGANs may need to be retrained. This work does not provide any information on how resistant steganography algorithms are to detection. Tang et al. [20] presented an automated learning framework using GANs with adversarial subnetworks. This framework calculates the embedding change probability for every pixel in the spatial cover image. The discriminator D compares the stego images produced by the generator G—which uses embedding distortions based on change probabilities—with cover images. Nevertheless, these methods perform no better than a few well-known recently developed ones. Hu et al. [21] performed steganography using deep convolutional generative networks. The data to be hidden are converted into a noise vector. Based on the noise vector, a carrier image is then generated. If the parameters of the extractor network are compromised, the method cannot learn the hidden information. Jiang et al. [22] addressed issues such as low steganographic capacity, low information recovery accuracy, as well as a lack of natural showing. An adversarial learning-based GAN was used to solve these problems. Despite the method’s higher accuracy, it lacks a mechanism to prevent information leakage. Ke et al. [23] proposed the Kerckhoffs’ principle for generative steganography. The authors used Kirchhoff’s concept to protect confidential data. The secret information cannot be retrieved without the key. However, the method has a limited steganographic capability.

3. Attention Vector Guided GAN Steganographic Technique

The GAN comprises two networks [24]: a generating network and a discriminative network. The discriminative network looks to see if the sample was legitimately formed by the generative network or if it was constructed with the goal of misleading it. The generative network generates a sample that comes close to being real by virtue of the competition between these two networks. Because they can adjust to complex distributions, the use of GAN is used to produce synthetic data.

According to the objective function of GAN,

\[ L_{GAN} = E_{\tau \sim P_r}[D(\tau)] - E_{\tau \sim P_r}[D(\tau)] + \lambda E_{\tau \sim P_r}[||\nabla_{\tau} D(\tau)||_2^2] \]

\( P_r \) denotes the distribution across \( V \). \( P_g \) denotes the distribution over which the generator generates data. \( P_s \) represents the uniform samples over \( P_r \) and \( P_g \).

Figure 1 shows the conventional architecture of the GAN-based steganographic approach. Random noise is created out of the hidden secret information.
In compliance with the noise vector, the generator (G) synthesises or updates the input cover image. The discriminator determines whether hidden embedding is present in an image. The decoder or extractor separates the stegano image with the noise vector. Also, the noise vector is used to reconstruct the hidden information.

The GAN stenographic approach applies global transformation to the entire image. This generates distortions in significant portions of the image that contain discriminative features. This paper presents AVG-GAN as a solution to this problem. The architecture of the proposed solution is given in Figure 2. The input image is split into grids, and a binary grid matrix mapping the significant regions as 1 and insignificant regions as 0 is created. The grid matrix of the image, an attention vector, and a key are given as the input into the AVG-GAN encoder. AVG-GAN encoder generates the stegno image which is uploaded to cloud. The attention vector is encrypted with the key using the AES encryption algorithm, and the encrypted attention vector is uploaded to cloud.

![Figure 2. Proposed Architecture.](image)

When a user requests secret information, it offers the key to download and decrypt the encrypted attention vector. The AVG-GAN decoder receives the decrypted attention vector. The hidden information mapped in the stegno image is obtained via the decoder.

Figure 3 illustrates the GAN encoder’s architecture. The transformation module is given the input image grids and attention vector matrix. In the significant regions, the transformation module applies a null mask. The transformed image is then sent to the generator network for processing. The generator network has been trained to output the null mask as the input null mask. The generator network generates the output stegno image using the transformed image and the secret text as input. The null mask of the output stegno image is substituted with image grid parts based on the attention matrix. The generated stegno image is then shared through the cloud.

Figure 4 illustrates the GAN decoder’s architecture. The stegno image is divided into grids, and the salient areas are designated as null masks based on the attention vector matrix to generate the transformed image. The transformed image is then sent to the discriminator network, which decodes it to generate the secret text. Because there is no distortion, the significant regions are preserved in the image.
4. Result and Discussion

Three medical image datasets—a brain tumour dataset from Kaggle [25], a glaucoma dataset [26], and an ultrasound ovarian cancer dataset [27]—are used to test the performance of the proposed approach. Each image falls into one of two categories: diseased or normal. The images in the brain tumour image dataset fall into one of two categories: tumour or normal. The images in the dataset for glaucoma images are divided into two categories: glaucoma and healthy. The images in the ultrasound dataset for ovarian cancer are divided into two categories: cancer and normal. Sample images from the database are shown in Table 1.
Table 1. Sample images.

<table>
<thead>
<tr>
<th>Tumour</th>
<th>Normal</th>
<th>Glaucoma</th>
<th>Healthy</th>
<th>Cancer</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Brain Image" /></td>
<td><img src="image2" alt="Normal Image" /></td>
<td><img src="image3" alt="Glaucoma Image" /></td>
<td><img src="image4" alt="Healthy Image" /></td>
<td><img src="image5" alt="Cancer Image" /></td>
<td><img src="image6" alt="Normal Image" /></td>
</tr>
</tbody>
</table>

Comparative analysis is carried out using various performance parameters between the proposed approach and the GAN-based steganography method (SteganoGAN [28]), High_Capacity_Information_Hiding_with_Generative_Adversarial_Network_ (HCISNet [29]), and Compressed_sensing_based_ enhanced_embedding_capacity_image_steganography_(CSIS [30]). Reed_Solomon_Bits_per_pixel_(RS_BPP), Peak_signal_to_noise_ratio_(PSNR), weighted_peak_signal_to_noise_ratio_(WPSNR), accuracy and structural_similarity_(SSIM) are used to compare the performance of the methods [31–34].

Table 2 presents the findings of measuring PSNR, RS-BPP, and WPSNR for several categories of medical images.

Table 2. Comparison of obtained results.

<table>
<thead>
<tr>
<th>Brain Image Dataset</th>
<th>Glaucoma Image Dataset</th>
<th>Ovarian Image Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR</td>
<td>RS-BPP</td>
<td>WPSNR</td>
</tr>
<tr>
<td>Proposed</td>
<td>39.96</td>
<td>6.28</td>
</tr>
<tr>
<td>SteganoGAN</td>
<td>36.46</td>
<td>4.33</td>
</tr>
<tr>
<td>HCISNet</td>
<td>38.87</td>
<td>5.67</td>
</tr>
<tr>
<td>CSIS</td>
<td>33.80</td>
<td>2.06</td>
</tr>
</tbody>
</table>

For the datasets on brain tumours, glaucoma, and ovarian images, the average PSNR in the proposed solution is at least 2.7% higher, 1.9% higher, and 2.7% higher, respectively. The PSNR has increased in the proposed solution because of two reasons: the generated stego image is nearly identical to the original and showcases significant regions in the stego image as in the original image. The SSIM is at least 2% higher in the proposed solution compared to existing works for the same reason.

For all three datasets, the embedding capacity is calculated as a percentage of the total image size that can be hidden. The result is shown in Figure 5. The proposed method has at least 2% less embedding capacity because there are less than 5% significant areas. The embedding capacity of the proposed approach decreases as the percentage of significant regions in an image increases. Figure 6 shows the embedding capacity reduction for various percentages of significant regions. Future work may involve compensating with secret information compression or effective coding.

The stego images in the cloud are taken from the cloud for three datasets, and an SVM machine learning classifier (Mode 1) is trained to classify the images into two classes. In a similar way, the SVM classifier is trained to classify the original images without GAN transformation (Mode 2) into two classes. The performance of the two modes of the classifier is measured in terms of accuracy, precision, recall, and Mathew’s correlation.
coefficient. While accuracy, precision, and recall are standard performance metrics for measuring classifier performance, Mathew's correlation coefficient is calculated as

\[
MCC = \frac{(TP \times TN) - (FN \times FP)}{\sqrt{(TP + FN)(TP + FP)(TN + FN)(TN + FP)}}
\]

The value of it ranges from −1 to +1. The higher the value of it, the better the classifier’s performance is.

The results for all datasets are given in Table 3. From the results of the brain tumour dataset, it can be seen that there is only a 1% difference in accuracy between Mode 1 and Mode 2 in the proposed solution. But the deviation is more than 5% in the other methods. From the results of the glaucoma dataset, it can be seen that there is only a 2% difference in accuracy between Mode 1 and Mode 2 in the proposed solution, but the deviation is more than 8% in other methods. From the results of the ovarian cancer dataset, it can be seen that there is only a 1% difference in accuracy between Mode 1 and Mode 2 in the proposed solution, but the deviation is more than 4% in the other methods.

The deviation in accuracy between Mode 1 and Mode 2 is higher for glaucoma as the significant information is widespread.

The MCC is plotted for all three datasets in Mode 1 and Mode 2 in Figure 7. The MCC is higher in the proposed method in comparison with state-of-the-art methods, demonstrating the least distortion to discriminative features in significant regions of the stegno image.
Table 3. Difference between the classification modes.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Mode 1</th>
<th>Mode 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brain tumour</td>
<td>Accuracy</td>
<td>Precision</td>
</tr>
<tr>
<td>Proposed</td>
<td>96</td>
<td>97</td>
</tr>
<tr>
<td>SteganoGAN</td>
<td>94</td>
<td>93</td>
</tr>
<tr>
<td>HCISNet</td>
<td>93</td>
<td>92</td>
</tr>
<tr>
<td>CSIS</td>
<td>92</td>
<td>93</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Glaucoma</th>
<th>Mode 1</th>
<th>Mode 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>94</td>
<td>95</td>
</tr>
<tr>
<td>SteganoGAN</td>
<td>91</td>
<td>93</td>
</tr>
<tr>
<td>HCISNet</td>
<td>90</td>
<td>92</td>
</tr>
<tr>
<td>CSIS</td>
<td>89</td>
<td>91</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ovarian</th>
<th>Mode 1</th>
<th>Mode 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>92</td>
<td>91</td>
</tr>
<tr>
<td>SteganoGAN</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>HCISNet</td>
<td>89</td>
<td>90</td>
</tr>
<tr>
<td>CSIS</td>
<td>87</td>
<td>88</td>
</tr>
</tbody>
</table>

Figure 7. Comparison of MCC.

5. Conclusions

This paper proposes a novel attention vector-guided GAN transformation for coverless image steganography. The solution was successful in resolving the issue of loss of the discriminative information in sensitive regions of images, which is essential for applications such as disease classification. The method was better suited for preserving utility in medical image sharing applications. The distortion loss has been prevented with little degradation in the quality of the secret information image. The reduction in embedding capacity was less than 2% in order to preserve features in key areas. Future work will include testing the proposed solution for other image classification applications.

Author Contributions: Conceptualisation, A. and V.; methodology, A. and V.; software, A. and D.S.U.; validation, A., V. and D.S.U.; formal analysis, V.; investigation, V.; resources, D.S.U.; data curation, V.; writing—original draft preparation, A.; writing—review and editing, V. and D.S.U.; visualisation, A.; supervision, V.; project administration, A.; 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: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.
Acknowledgments: Authors would like to thank the reviewers for the constructive comments, which improved overall quality of this paper.

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

References
3. Holub, V.; Fridrich, J.; Denemark, T. Universal distortion function for steganography in an arbitrary domain. EURASIP J. Inf. Secur. 2014, 1. [CrossRef]


**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.