A Secure Autoencoder–Based Steganography Method Using Garsia–Wachs Huffman Coding and Uniform Gradient Discriminative Learning

R. Padma
Department of Computer Science and Systems Engineering, GITAM School of Computer Science and Engineering, GITAM University, Bengaluru, India
pramacha@gitam.in (corresponding author)

Vamsidhar Yendapalli
Department of Computer Science and Systems Engineering, GITAM School of Computer Science and Engineering, GITAM University, Bengaluru, India
vyendapa@gitam.edu

Received: 13 April 2026 | Revised: 29 April 2026 | Accepted: 9 May 2026

 DOI: https://doi.org/10.48084/etasr.19311

ABSTRACT

With increasing concerns about security threats during the transmission of valuable data, the role played by steganography is of prime importance. Although the adoption of deep learning for optimization is increasing, it has been observed that it may also lead to degraded performance and reduced data quality. Hence, this manuscript presents an innovative yet simplified computational model for steganography, termed the Garsia–Wachs Huffman Coding and Uniform Gradient Discriminative Autoencoder (GWHC-UGDA). Unlike frequently adopted deep learning models, the proposed model uses an autoencoder to improve image quality and to increase extraction accuracy. The system uses Mutual Normalized Histogram analysis as well as Garsia–Wachs Optimal Huffman Coding for separately processing cover and secret information. The system also contributes to enhancing robustness and minimizing extraction error using a Disentangled and Gradient-Discriminative Log-Likelihood Autoencoder (DGDLLA). Implemented in Python using a standard dataset, the model achieves 21–31% higher extraction accuracy with up to 60% reduction in extraction error in contrast to baseline models.

Keywords-steganography; deep learning; autoencoder; extraction; Huffman coding; histogram analysis

I. INTRODUCTION

Researchers have investigated adversarial example techniques from deep learning to improve resistance against steganalysis due to their effectiveness in misleading detection models. Introducing perturbations directly into stego images can hinder unauthorized extraction but may also affect legitimate recovery. Therefore, adversarial perturbations are applied to cover images prior to embedding to enhance security while preserving recoverability.

Content-adaptive adversarial steganography was proposed in [1], which applied adversarial perturbations generated by adversarial example methods to cover images to deceive steganalysis systems. By employing this generation method, the drawbacks arising from prevailing cover enhancement methods were mitigated. In addition, image texture information based on a hybrid texture descriptor, along with image semantic information, was used for segmentation to select regions of interest, enhancing security, reducing the missed detection rate, and improving Peak Signal-to-Noise Ratio (PSNR). Despite improvements in detection performance and PSNR, extraction error was not addressed.

In general, feature construction and classifier design in steganalysis are often performed independently, making joint optimization difficult. To address this limitation, an Advanced General Convolutional Neural Network (AG-Net) [2] was proposed to bridge the gap between feature design and classifier learning by introducing a confrontation module for extracting and comparing features of cover and stego images. A correlation between adjacent confrontation modules was then established based on feature comparisons from previous modules. Although improvements in detection performance and PSNR were achieved, discrepancies between cover and stego images accumulated in mid and high-level feature representations. Finally, a softmax layer was used for stego image classification. Despite improved detection accuracy, extraction error was still not adequately addressed.

Another deep learning-based approach aimed at reducing message detection rate and execution time was presented in [3]. However, deep learning methods generally present trade-offs in

terms of performance and complexity. To address these limitations, an ensemble-based classifier was proposed in [4], which improved classification accuracy and significantly enhanced PSNR.

Though deep learning has brought a significant paradigm shift in steganography, there is no consensus on the use of deep neural networks in reversible steganography. In [5], reversible steganography improves rate–distortion performance using prediction-based modeling and adaptive embedding guided by prediction accuracy. A study involving different training configurations for predictive analysis in steganography was presented in [6]; however, it did not consider Bit Error Rate (BER) in steganalysis. A deep learning-driven feature-based method focusing on BER was presented in [7].

A hybrid approach combining code-based cryptography with chaotic maps for pseudo-random bit generation was proposed in [8], enhancing the unpredictability and security of generated sequences. The integration of chaotic dynamics improves randomness characteristics, making it suitable for secure communication applications. Nevertheless, this method primarily focuses on cryptographic key generation and does not directly address data embedding or extraction accuracy in steganographic systems.

A multi-image steganography framework using Least Significant Bit (LSB) substitution and Discrete Wavelet Transform (DWT) was proposed in [9]. While this technique improves embedding efficiency, it is still limited in the effective and robust extraction of hidden information from images.

A steganographic technique using YCbCr color space conversion was proposed in [10]. Nonetheless, there is no systematic evaluation of distortion measures and extraction performance.

A reversible data hiding scheme based on intelligent image interpolation was proposed in [11], enabling effective embedding with perfect recovery. Although this approach enhances data payload and reversibility, it introduces computational complexity and lacks sufficient evaluation of metrics such as Structural Similarity Index Measure (SSIM) and extraction precision.

An advanced steganography system using Integer Wavelet Transform (IWT) and Hamming coding was introduced in [12]. This system improves resistance and error correction capability but does not address low-distortion embedding.

In addition to embedding-oriented steganography methods, detection-oriented steganalysis approaches have also been widely investigated. A machine learning-based steganalysis process was described in [13], which uses feature extraction methods for robust detection of stego images. Although this methodology significantly enhances detection capability, its emphasis is more on analysis than on secure data embedding or extraction.

A machine learning-based steganalysis process was described in [14], which proposed a novel blind steganalysis technique using third-order SPAM features together with ensemble classifiers for improved detection. This process can accurately detect hidden information but does not improve embedding and extraction quality.

Authors in [15] proposed a color image steganography technique by exploiting the Hue channel of the HSV color model to increase invisibility. The algorithm ensures image invisibility through reduced distortion but pays little attention to extraction errors. Authors in [16] explored blind image steganalysis through feature-based classification. Their results revealed difficulties in detecting hidden images, but did not consider data embedding and image quality preservation.

From a broader security perspective, recent research has also explored hybrid cryptography and steganography-based communication systems. Authors in [17] presented a machine learning-based system for key generation and encryption to improve confidentiality and robustness. Notably, this approach focuses solely on cryptographic properties without considering embedding distortion or extraction performance in steganographic operations. In another study, authors in [18] proposed an autoencoder-based hybrid cryptosteganography system to improve secure image transmission efficiency. Still, there is no thorough evaluation of image quality and extraction performance. In general, authors [19] discuss emerging cybersecurity threats in intelligent systems and the need for secure communication channels. Yet, these approaches do not jointly address embedding and extraction.

In addition to conventional steganography methods, recent deep learning-based approaches such as Generative Adversarial Network (GAN)-based, U-Net-based, and transformer-based models have also been investigated. GAN-based steganography [20] provides strong embedding capability through adversarial training but often suffers from training instability and high computational complexity. U-Net-based steganography models [21] achieve improved image reconstruction quality due to encoder–decoder skip connections but require a large number of parameters and increased training time. Transformer-based approaches [22] offer enhanced feature representation and global context learning but involve high computational and memory costs.

Although significant progress has been made in steganography, most existing methods primarily focus on embedding efficiency, attack resistance, or cryptographic security. Limited attention has been given to achieving a balanced improvement in image quality, structural similarity, extraction accuracy, and reduction of extraction error simultaneously. In addition, many deep learning-based approaches increase computational complexity without ensuring effective performance in both hiding and revealing stages. Therefore, there remains a research gap in designing a simplified yet robust steganographic framework that can preserve image quality while ensuring accurate and low-error secret data extraction. The proposed Garsia–Wachs Huffman Coding and Uniform Gradient Discriminative Autoencoder (GWHC-UGDA) model addresses this gap through optimized preprocessing and a uniform gradient discriminative autoencoder-based secure embedding framework.

The main contributions of the proposed GWHC-UGDA framework are summarized as follows: (i) a hybrid preprocessing approach combining Mutual Normalized Histogram Equalization and Garsia–Wachs Optimal Huffman Coding is introduced to enhance feature representation and embedding efficiency; (ii) an autoencoder-based architecture is designed to perform simultaneous hiding and revealing of secret data, thereby reducing extraction error; (iii) the proposed model achieves improved performance in terms of reduced BER and enhanced PSNR, ensuring improved image quality and extraction accuracy; and (iv) compared to computationally intensive deep learning models such as GAN-based and transformer-based approaches, the proposed framework provides a lightweight and efficient solution with balanced performance.

II. PROPOSED METHOD

For the cover image (I_c) construction, Mutual Normalized Histogram Equalization is employed to enhance image contrast and feature representation (F). For the secret text message (I_s) processing, Garsia–Wachs Optimal Huffman Coding is utilized as a lossless compression technique for efficient data representation and encoding. In the first phase of the proposed framework, these two preprocessing strategies are applied to the cover image and secret text, respectively. The overall steganographic pipeline is then completed using a Disentangled and Gradient-Discriminative Log-Likelihood Autoencoder (DGDLLA)-based model, which performs joint embedding and reconstruction within an encoder–decoder architecture.

The proposed method uses the correlation between pixel intensity and the mean intensity of its adjacent pixels to enhance overall image contrast. This association is obtained by constructing a mutual histogram. Multiple neighboring pixels are used to construct histograms based on grouped pixel intensity values. Initially, cover images are acquired as input, and normalization is performed to rescale pixel intensity values and improve signal representation. Let CI represent the cover image prior to preprocessing; the normalization function is expressed in (1):

$$\mathrm{CI_n}(p,q) = [\mathrm{CI}(p,q) - \mathrm{PI_{min}}]\frac{\mathrm{N_{max}}-\mathrm{N_{min}}}{\mathrm{PI_{max}}-\mathrm{PI_{min}}} + \mathrm{N_{min}} \qquad (1)$$

From (1), $\mathrm{CI_n}$, $\mathrm{PI_{min}}$, and $\mathrm{PI_{max}}$ denote the normalized cover image, and the minimum and maximum pixel intensity values of the original cover image CI, respectively, at position (p,q). In addition, $\mathrm{N_{min}}$ and $\mathrm{N_{max}}$ represent the minimum and maximum pixel intensity values of the normalized image $\mathrm{CI_n}$, respectively.

Let $x(p,q)$ represent the pixel intensity grey value at position (p,q), where $p \in \{1,2,3,..,M\}$ and $q \in \{1,2,3,...,N\}$. The pixel intensity value with respect to its adjacent window is formulated as follows:

$$y(p,q) = \mathrm{CI_n}(p,q)\left[\frac{1}{W*W}\sum_{m=-l}^{l}\sum_{n=-l}^{l} x(p+m,q+n)\right] \qquad (2)$$

Based on (2), $x(p,q)$ from the first image and $y(p,q)$ from the adjacent image are used as features to construct the mutual histogram. In addition, let $\mathrm{Occ}(i,j)$ represent the occurrence of the pair (p,q), where $x(p,q) = i$ and $y(p,q) = j$. Using (i,j) and $\mathrm{Occ}(i,j)$, the mutual histogram is constructed. The mutual histogram considers adjacent pixel information, capturing relationships that are often ignored in traditional histogram models. The mutual histogram is defined as:

$$\mathrm{PCI} = \mathrm{MH} = \{\mathrm{Occ}(i,j)\} \qquad (3)$$

In (3), $\mathrm{Occ}(i,j)$ represents the grey-level pair occurrences of $x(p,q)$ and $y(p,q)$ at the same position (p,q). The values of i and j lie in the ranges $[0, K-1]$ and $[0, K]$, respectively.

Following this, secret text message reconstruction is performed using Garsia–Wachs Optimal Huffman Coding. Given a secret text message $\mathrm{SM} = \{\mathrm{SM_1}, \mathrm{SM_2}, \ldots, \mathrm{SM_n}\}$ of size n and corresponding weights ' $\mathrm{We} = \{\mathrm{We_1}, \mathrm{We_2}, \ldots, \mathrm{We_n}\}$, where $\mathrm{We_i} = \mathrm{Weight}(\mathrm{SM_i})$, a code $\mathrm{C(We)} = (\mathrm{C_1}, \mathrm{C_2}, .., \mathrm{C_n})$ is generated for each secret text message SM. The optimization objective is defined as:

$$\mathrm{PSM} = \mathrm{Le}\big(\mathrm{C(We)}\big) = \sum_{i=1}^{n} \mathrm{We_i}\mathrm{Length}(\mathrm{C_i}) \qquad (4)$$

To improve computational efficiency, Garsia–Wachs optimality is employed over the Huffman coding results. This approach reduces computational complexity by limiting operations to relevant weight configurations associated with the secret text message representation. The search space is partitioned into subregions based on weight distributions, where the optimization process is selectively applied. As a result, separate formulations for cover image and secret text message construction are achieved, leading to improved PSNR and more efficient preprocessing of both cover images and secret text messages.

To enhance imperceptibility and reduce extraction error, a novel DGDLLA-based steganography model is introduced. Figure 1 shows the structure of the proposed model, which operates using an encoder–decoder framework. The encoder first splits the preprocessed cover image into two equidistant horizontal regions and extracts feature representations. Simultaneously, the preprocessed secret text message is transformed into a texture representation using a uniform distribution guided by the Gradient Log-Likelihood function. The disentanglement process preserves image structure, while the discriminative loss enforces uniform feature distribution for secure embedding. During decoding, both representations are jointly reconstructed to minimize extraction error and improve reconstruction accuracy. This formulation enables simultaneous hiding and revealing with improved robustness and reduced distortion.

The proposed model is based on an autoencoder framework, which is widely used for feature extraction and reconstruction in image processing and steganography applications. The proposed model consists of two main components: an encoder (hidden network) and a decoder (reveal network). The encoder is responsible for embedding by jointly processing the preprocessed cover image and the preprocessed secret text image, and encoding them into a shared latent representation. The decoder reconstructs the cover image and retrieves the hidden secret message from this latent space.

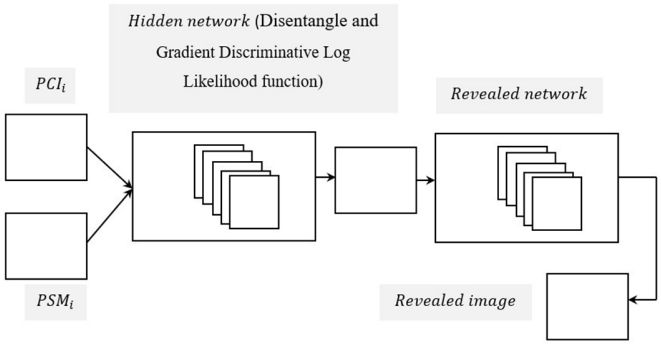

Fig. 1. Structure of the DGDLLA-based steganography model.

Specifically, the encoder processes the preprocessed cover image (PCI) by splitting it into two equidistant horizontal regions (left and right) to extract spatial feature representations. In parallel, the preprocessed secret message (PSM) is transformed into a texture representation and embedded within the latent space. A conventional decoding module is then used to recover key image features from the learned representation.

During embedding, the encoder maps the inputs into a secret tensor, where the cover image representation and secret message representation are jointly encoded. The decoder takes this latent tensor as input and reconstructs both the cover image and the embedded secret message. To ensure statistical consistency, the secret feature vector is constrained using a uniform distribution via the Gradient Log-Likelihood function. Finally, the model is trained using combined loss functions to simultaneously ensure high-quality image reconstruction and minimal extraction error.

The encoder Enc is trained to disentangle an image P into the preprocessed cover image feature PCI and the preprocessed secret message feature PSM as:

$$(\mathrm{PCI_i}, \mathrm{PSM_i}) = \mathrm{Enc(P)} \quad (5)$$

where P is sampled from the BOSSBase v1.0.1 dataset. The cover image is split into two equidistant horizontal parts: horizontal left (HL) and horizontal right (HR). This is expressed as:

$$\mathrm{SCI_i} = \sum_{i=1}^{n} \mathrm{HL(PCI_i)} + \mathrm{HR(PCI_i)} \quad (6)$$

$$\begin{cases} \mathrm{Height} = \sum_{i=1}^{n} \mathrm{PCI_i.Shape[0]} \\ \mathrm{Width} = \sum_{i=1}^{n} \mathrm{PCI_i.Shape[1]} \end{cases} \quad (7)$$

$$\begin{cases} \mathrm{HL(PCI_i)} = \sum_{i=1}^{n} \mathrm{PCI_i\,[Height/2]} \\ \mathrm{HR(PCI_i)} = \sum_{i=1}^{n} \mathrm{PCI_i\,[Width/2]} \end{cases} \quad (8)$$

From (6), the preprocessed cover image (PCI) is divided into two equidistant horizontal regions, namely the horizontal left portion $\mathrm{HL(PCI_i)}$ and the horizontal right portion $\mathrm{HR(PCI_i)}$, respectively. The corresponding dimensions are defined using statistical width and height functions as described in (7) and (8).

Next, a uniform distribution constraint is applied to the secret representation ($\mathrm{PSM_i}$) using a uniform discriminative loss function (L), as given below:

$$\mathrm{SSM_i} = \mathrm{Loss_{Enc,Dis}} = \mathrm{logistic}\left(\mathrm{DD(PSM_i)}\right) \quad (9)$$

From (9), the logistic activation function is used to squash negative values to zero. Here, $\mathrm{PSM_i}$ conforms to the same distribution as a sampled $\mathrm{PSM_j}$.

To further minimize extraction error and reconstruction loss, the gradient of the log-likelihood function is computed as given below:

$$\frac{\partial \mathrm{Log(w)}}{\partial \mathrm{w}} = \sum_{i=1}^{n} \varphi(\mathrm{p_i}, \mathrm{q_i}) - \mathrm{Exp}_{\mathrm{prob(q|p_i,w)}} \varphi(\mathrm{p_i}, \mathrm{q}) \quad (10)$$

From (10), the gradient of the log-likelihood is obtained by optimizing the training data with respect to parameter w, where the expectation term $\mathrm{Exp}_{\mathrm{prob(q|p_i,w)}}$ represents the probabilistic model of q conditioned on $\mathrm{p_i}$ and w.

The overall optimization objective of the proposed GWHC-UGDA model is defined using a combined loss function:

$$\mathrm{L_{total}} = \lambda_1 \mathrm{L_{recon}} + \lambda_2 \mathrm{L_{BER}} + \lambda_3 \mathrm{L_{dis}} \quad (11)$$

Here, $\mathrm{L_{recon}}$ denotes the reconstruction loss, which minimizes the difference between the original cover image and the generated stego image to preserve image quality and structural similarity. $\mathrm{L_{BER}}$ represents the BER loss, which reduces errors during the extraction of the hidden message and improves extraction accuracy. $\mathrm{L_{dis}}$ corresponds to the discriminative loss that enforces uniform feature distribution and enhances the security of embedding through the DGDLLA framework. The parameters λ_1, λ_2, and λ_3 are weighting factors that balance the contribution of each loss component during the training process.

Finally, the stego image (FSI) is obtained by combining the disentangled cover representation ($\mathrm{SCI_i}$) and the secret representation ($\mathrm{SSM_i}$) as:

$$\mathrm{FSI} = \mathrm{SCI_i} \cdot \mathrm{SSM_i} \quad (12)$$

From (12), it is observed that the encoder and decoder networks are symmetric within the proposed autoencoder framework based on Disentangled and Gradient-Discriminative Log-Likelihood learning. The equivalent network structure employed in the proposed work contributes to secure image decryption performance. The system is designed to learn both hiding and revealing simultaneously and reduce extraction error; thus, two distinct functions are applied. First, with the preprocessed cover image acquired as input, the image is subjected to a disentanglement function to split it into two equidistant horizontal parts. Second, the preprocessed text message is processed using the Uniform Gradient Discriminative Log-Likelihood function. Finally, both functions are merged to produce the final output. Hence, the method performs both hiding and revealing concurrently. This integrated formulation not only minimizes extraction error but also improves extraction accuracy significantly.

III. RESULTS AND DISCUSSION

This section presents the experimental results of the proposed GWHC-UGDA-based secure steganography method, and compares it with existing Content-Adaptive Adversarial Steganography (CAAS) [1] and Advanced General

Convolutional Neural Network (AG-Net) [2]. The proposed model was implemented in Python and evaluated using the BOSSBase v1.0.1 dataset [23], a standard benchmark dataset widely used in image steganography and steganalysis research. The dataset contains 10,000 grayscale images with a resolution of 512 × 512 pixels, collected from seven different digital cameras to ensure diversity in content and acquisition conditions.

For evaluation, the dataset was split into 8,000 training samples and 2,000 testing samples. This partition enables effective learning of embedding and extraction features. The same dataset split, training configuration, and experimental settings were applied to all methods to ensure fair comparison and reproducibility. The model was trained using an autoencoder-based framework for 100 epochs with a batch size of 32, employing the Adam optimizer with a learning rate of 0.001. A loss function combining reconstruction loss and discriminative loss was used to preserve image quality while minimizing extraction error. Reconstruction loss ensures structural similarity between the original and stego images, whereas discriminative loss enforces uniform feature distribution to improve embedding security.

The performance of the proposed method was evaluated using BER, PSNR, extraction accuracy, and extraction error. BER is a standard metric used to measure extraction reliability in steganographic systems. Figure 2 shows the BER comparison among GWHC-UGDA, CAAS, and AG-Net. In Figure 2, the blue line represents GWHC-UGDA, the orange line represents CAAS, and the green line represents AG-Net. The results show that GWHC-UGDA achieves a significantly lower BER compared to CAAS and AG-Net, with reductions of approximately 22% and 30%, respectively. This improvement is attributed to the modeling of correlations between pixel intensities and the mean intensity of neighboring pixels, which enhances feature consistency and reduces reconstruction errors.

Figure 3 presents the PSNR comparison among the three methods. PSNR is widely used to evaluate image quality and imperceptibility in steganographic systems. To evaluate imperceptibility, simulations were conducted using 10 sample images of varying sizes measured in KB. The results indicate that PSNR values vary across images, however, the proposed GWHC-UGDA consistently achieves higher PSNR compared to CAAS and AG-Net.

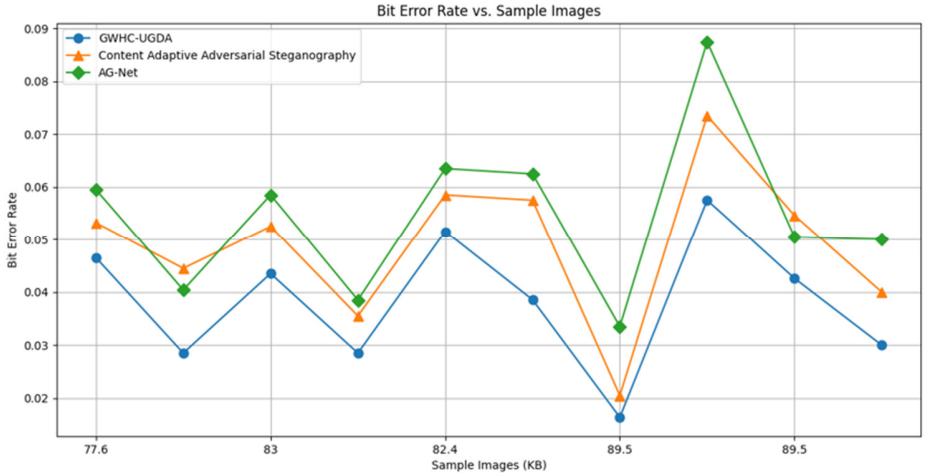

Fig. 2. Analysis of BER.

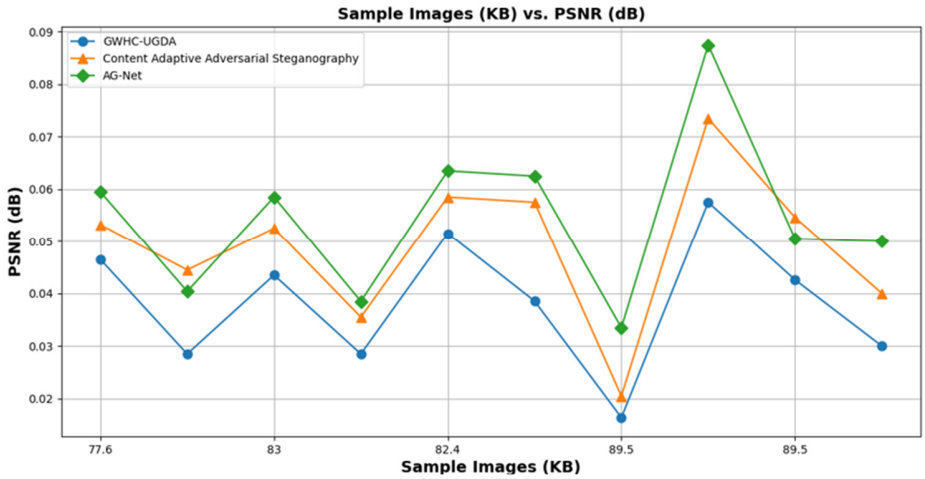

Fig. 3. Analysis of PSNR.

For example, for a 106 KB image, the PSNR values were 41.35 dB for GWHC-UGDA, 39.45 dB for CAAS, and 38.20 dB for AG-Net. This improvement is due to the separate preprocessing of cover images and secret text messages, where Mutual Normalized Histogram Equalization is applied for cover image construction and Garsia–Wachs Optimal Huffman Coding is used for secret text representation. The resulting preprocessed data improve embedding quality and reconstruction performance. Overall, the proposed method improves PSNR by approximately 5% over CAAS and 16% over AG-Net.

Figure 4 illustrates the extraction error comparison among the three methods, which is considered a critical performance metric for evaluating secure steganography systems. In the simulation, 13 samples were used for the first iteration. The proposed GWHC-UGDA method produced only one erroneous sample, whereas CAAS and AG-Net produced two and three erroneous samples, respectively. Consequently, the overall extraction error rates were 7.69%, 15.38%, and 23.07% for GWHC-UGDA, CAAS, and AG-Net, respectively, demonstrating the effectiveness of the proposed method in minimizing extraction error.

This improvement is attributed to the proposed DGDLLA-based steganography framework, which enhances the imperceptibility of hidden data by generating high-quality reconstructed images while simultaneously reducing extraction errors. Furthermore, the proposed model enables end-to-end steganography by jointly performing hiding and revealing operations within a unified architecture. As a result, the GWHC-UGDA method reduces extraction error by 43% and 60% compared to CAAS and AG-Net, respectively.

Finally, Figure 5 illustrates the extraction accuracy obtained using the three methods. The x-axis represents 10 test images of different sizes, whereas the y-axis denotes extraction accuracy for each method. As shown in the figure, the proposed GWHC-UGDA method achieves higher extraction accuracy compared to CAAS and AG-Net. This is confirmed by the simulation results using 13 sample images, where 11 images were correctly extracted using GWHC-UGDA, compared to 10 images for CAAS and 9 images for AG-Net. The performance improvement is attributed to the splitting of the preprocessed cover image into two equidistant horizontal regions (left and right) and the application of a uniform distribution constraint on the preprocessed secret text using the Gradient Log-Likelihood function.

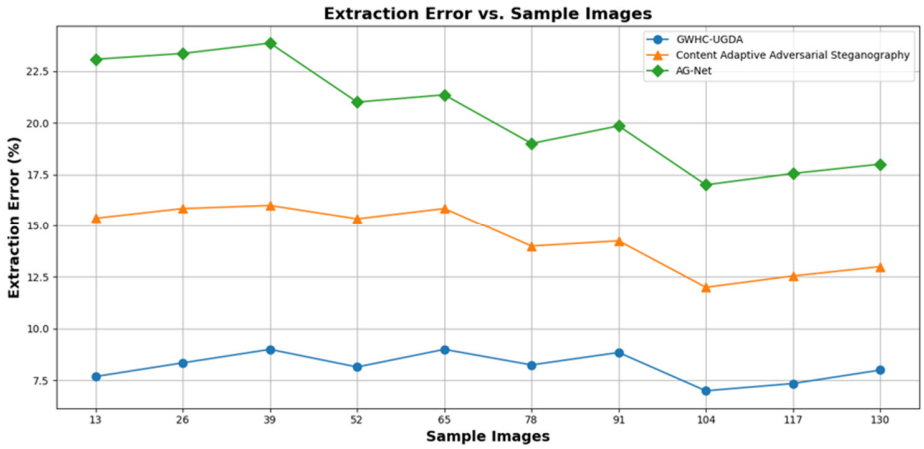

Fig. 4. Analysis of extraction error.

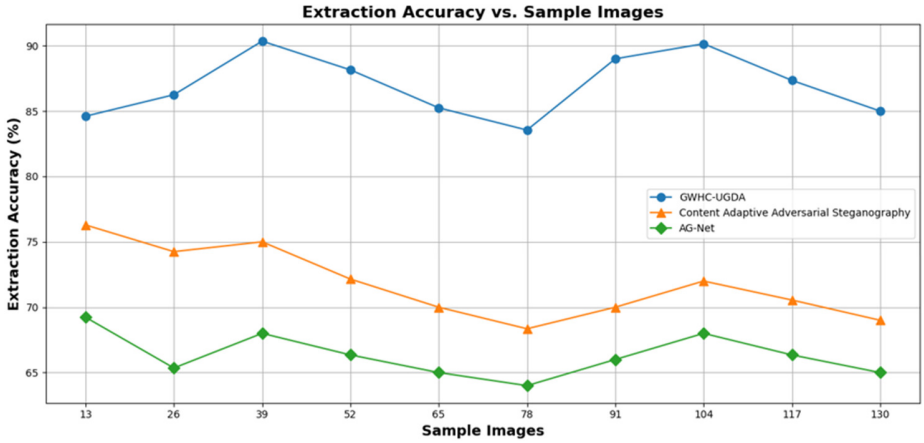

Fig. 5. Analysis of extraction accuracy.

Overall, the proposed method achieves approximately 21% and 31% higher extraction accuracy compared to CAAS and AG-Net, respectively. In addition, Mean Squared Error (MSE) is used to measure pixel-wise distortion between the original and stego images, whereas Structural Similarity Index (SSIM) evaluates perceptual quality based on luminance, contrast, and structural information. The experimental results demonstrate that the proposed GWHC-UGDA method outperforms CAAS and AG-Net in terms of both accuracy and image quality.

Table I presents a comparative performance analysis of the proposed GWHC-UGDA method with existing steganography approaches. The improved performance is attributed to mutual histogram-based preprocessing and the Uniform Gradient Discriminative Autoencoder, which jointly reduce embedding distortion and preserve image structure.

TABLE I. COMPARATIVE PERFORMANCE ANALYSIS OF THE PROPOSED GWHC-UGDA WITH EXISTING STEGANOGRAPHY METHODS

SI No	Methods	PSNR (dB)	MSE	SSIM	BER	Extraction accuracy
1	CASS	39.45	0.72	0.941	Higher	Lower
2	AG-Net	38.20	0.91	0.923	Higher	Lower
3	IWT+ Hamming code	40.10	0.65	0.952	–	85%
4	HSV-based method	40.85	0.58	0.961	–	88%
5	GWHC-UGDA (proposed)	41.35	0.48	0.972	Lowest	Highest

The security of the proposed GWHC-UGDA technique is evaluated based on histogram preservation, robustness against steganalysis attacks, and accuracy of hidden message extraction. It is observed that there is minimal difference between the histograms of the cover image and the corresponding stego image, indicating negligible statistical distortion introduced during embedding. This makes the stego images difficult to detect using conventional statistical analysis methods. In addition, the proposed scheme achieves low BER and low extraction error, enabling reliable recovery of the hidden message. The use of uniform gradient discriminative learning contributes to balanced feature distributions, improving robustness against steganalysis attacks. Compared with recent deep learning-based steganography models, the proposed approach achieves improved image quality preservation and reduced extraction error while maintaining lower computational complexity, thereby providing a more efficient and reliable secure embedding framework.

To further evaluate the security of the proposed framework, a qualitative steganalysis assessment is performed. The histogram comparison between the cover and stego images shows minimal deviation, indicating that the embedding process preserves statistical properties and reduces the risk of detection. The proposed GWHC-UGDA model demonstrates strong resistance to conventional steganalysis techniques due to the uniform feature distribution achieved through the discriminative autoencoder. Compared to existing deep learning-based steganography approaches, such as GAN-based methods [21], which may introduce detectable artifacts under adversarial training, the proposed model maintains consistent embedding patterns with lower distortion. These observations confirm that the proposed method provides improved robustness against steganalysis while preserving image quality and extraction reliability.

The bit-per-pixel (bpp) ratio is defined as the ratio between the total number of embedded secret bits and the total number of pixels in the cover image. It can be mathematically expressed as:

$$\text{bpp} = \frac{\text{Number of embedded bits}}{\text{Total number of pixels}} \quad (13)$$

In all experiments, the proposed method maintains a payload in the range of approximately 0.4–0.5 bpp. This demonstrates that the model achieves a good balance between embedding capacity and image quality, ensuring that the stego images preserve high visual imperceptibility while still carrying sufficient hidden information.

IV. CONCLUSION

A systematic method, namely the Garsia–Wachs Huffman Coding and Uniform Gradient Discriminative Autoencoder (GWHC-UGDA), has been presented in this paper. The proposed work focuses on secure steganography using image samples and employs Mutual Normalized Histogram Equalization and Garsia–Wachs Optimal Huffman Coding for preprocessing, together with a Disentangled and Gradient-Discriminative Log-Likelihood Autoencoder (DGDLLA)-based framework for secure steganography.

In the proposed approach, the steganographic process is organized into two stages. First, preprocessing is performed using Mutual Normalized Histogram Equalization for cover image enhancement and Garsia–Wachs Optimal Huffman Coding for efficient representation of the secret text message. Second, the preprocessed outputs are processed through a jointly designed encoder–decoder architecture based on the DGDLLA framework, enabling simultaneous hiding and revealing of information.

This integrated design supports parallel processing of embedding and extraction, reducing extraction error and improving overall reconstruction accuracy. Furthermore, the use of discriminative learning and logistic activation contributes to improved feature consistency and enhanced robustness of the embedding process. Experiments conducted on the BOSSBase v1.0.1 dataset demonstrate that the proposed model achieves higher Peak Signal-to-Noise Ratio (PSNR) and extraction accuracy, while significantly reducing extraction error and Bit Error Rate (BER), compared to baseline methods.

DECLARATION OF COMPETING INTERESTS

The author declares no conflict of interest.

ACKNOWLEDGMENT

The authors declare that no external funding was received for this work.

DATA AVAILABILITY

The BOSSBase v1.0.1 dataset used in this study is publicly available [23]. Additional data supporting the findings of this

study are available from the corresponding author upon reasonable request.

AI USE AND DECLARATION OF GENERATIVE AI USE

The authors confirm that this work was carried out independently and that all content is original.

REFERENCES

[1] V. K. Sharma, R. N. Mir, and R. K. Rout, "Towards secured image steganography based on content-adaptive adversarial perturbation," *Computers and Electrical Engineering*, vol. 105, Jan. 2023, Art. no. 108484, https://doi.org/10.1016/j.compeleceng.2022.108484.

[2] H. Zhang, F. Liu, Z. Song, X. Zhang, and Y. Zhao, "AG-Net: An Advanced General CNN Model for Steganalysis," *IEEE Access*, vol. 10, pp. 44116–44122, 2022, https://doi.org/10.1109/ACCESS.2022.3150276.

[3] A. A. Mawgoud, M. H. N. Taha, A. Abu-Talleb, and A. Kotb, "A deep learning based steganography integration framework for ad-hoc cloud computing data security augmentation using the V-BOINC system," *Journal of Cloud Computing*, vol. 11, no. 1, Dec. 2022, Art. no. 97, https://doi.org/10.1186/s13677-022-00339-w.

[4] M. Płachta, M. Krzemień, K. Szczypiorski, and A. Janicki, "Detection of Image Steganography Using Deep Learning and Ensemble Classifiers," *Electronics*, vol. 11, no. 10, May 2022, Art. no. 1565, https://doi.org/10.3390/electronics11101565.

[5] A. Kouhi and M. H. Sedaaghi, "Prediction error distribution with dynamic asymmetry for reversible data hiding," *Expert Systems with Applications*, vol. 184, Dec. 2021, Art. no. 115475, https://doi.org/10.1016/j.eswa.2021.115475.

[6] C.-C. Chang, X. Wang, S. Chen, I. Echizen, V. Sanchez, and C.-T. Li, "Deep Learning for Predictive Analytics in Reversible Steganography," *IEEE Access*, vol. 11, pp. 3494–3510, 2023, https://doi.org/10.1109/ACCESS.2023.3233976.

[7] Y. Li, B. Ling, D. Hu, S. Zheng, and G. Zhang, "A Deep Learning Driven Feature Based Steganalysis Approach," *Intelligent Automation & Soft Computing*, vol. 37, no. 2, pp. 2213–2225, June 2023, https://doi.org/10.32604/iasc.2023.029983.

[8] T. K. Alshekly, E. A. AlBahrani, and L. B. Ayed, "Code-Based Cryptography and Chaotic Maps as Pseudo-Random Bit Generator," *Engineering, Technology & Applied Science Research*, vol. 15, no. 6, pp. 29041–29048, Dec. 2025, https://doi.org/10.48084/etasr.13718.

[9] A. Gutub and F. Al-Shaarani, "Efficient Implementation of Multi-image Secret Hiding Based on LSB and DWT Steganography Comparisons," *Arabian Journal for Science and Engineering*, vol. 45, no. 4, pp. 2631–2644, Apr. 2020, https://doi.org/10.1007/s13369-020-04413-w.

[10] S. Alharthi and A. Gutub, "Adjusting image stego practicality via YCbCr color space formation," *Journal of Engineering Research*, vol. 14, no. 1, pp. 756–764, Mar. 2026, https://doi.org/10.1016/j.jer.2025.07.008.

[11] F. S. Hassan and A. Gutub, "Efficient reversible data hiding multimedia technique based on smart image interpolation," *Multimedia Tools and Applications*, vol. 79, no. 39, pp. 30087–30109, Oct. 2020, https://doi.org/10.1007/s11042-020-09513-1.

[12] Z. Saeidi, A. Yazdi, S. Mashhadi, M. Hadian, and A. Gutub, "High performance image steganography integrating IWT and Hamming code within secret sharing," *IET Image Processing*, vol. 18, no. 1, pp. 129–139, Jan. 2024, https://doi.org/10.1049/ipr2.12938.

[13] A. Aljarf, H. Zamzami, and A. Gutub, "Integrating machine learning and features extraction for practical reliable color images steganalysis classification," *Soft Computing*, vol. 27, no. 19, pp. 13877–13888, Oct. 2023, https://doi.org/10.1007/s00500-023-09042-7.

[14] J. Hemalatha, M. Sekar, C. Kumar, A. Gutub, and A. K. Sahu, "Towards improving the performance of blind image steganalyzer using third-order SPAM features and ensemble classifier," *Journal of Information Security and Applications*, vol. 76, Aug. 2023, Art. no. 103541, https://doi.org/10.1016/j.jisa.2023.103541.

[15] F. S. Hassan and A. Gutub, "Improving data hiding within colour images using hue component of HSV colour space," *CAAI Transactions on Intelligence Technology*, vol. 7, no. 1, pp. 56–68, Mar. 2022, https://doi.org/10.1049/cit2.12053.

[16] A. Aljarf, H. Zamzami, and A. Gutub, "Is blind image steganalysis practical using feature-based classification?," *Multimedia Tools and Applications*, vol. 83, no. 2, pp. 4579–4612, Jan. 2024, https://doi.org/10.1007/s11042-023-15682-6.

[17] A. Saini and R. Sehrawat, "Enhancing Data Security through Machine Learning-based Key Generation and Encryption," *Engineering, Technology & Applied Science Research*, vol. 14, no. 3, pp. 14148–14154, June 2024, https://doi.org/10.48084/etasr.7181.

[18] Abhishek and R. Sehrawat, "Sustainable Image-based Encryption Using Cryptography and Steganography with Autoencoder," in *Proceedings of 4th International Conference on ICT for Digital, Smart, and Sustainable Development*, Delhi, India, 2024, pp. 205–214, https://doi.org/10.1007/978-981-97-7831-7_14.

[19] U. Rawat, Abhishek, H. Singh, and A. Ur Rehman, "Cybersecurity Challenges and Risks in AGI Development and Deployment," in *Artificial General Intelligence (AGI) Security: Smart Applications and Sustainable Technologies*, S. El Hajjami, K. Kaushik, and I. U. Khan, Eds. Singapore: Springer Nature, 2025, pp. 291–314, https://doi.org/10.1007/978-981-97-3222-7_14.

[20] J. Zhu, R. Kaplan, J. Johnson, and L. Fei-Fei, "HiDDeN: Hiding Data With Deep Networks," in *15th European Conference on Computer Vision*, Munich, Germany, 2018, pp. 682–697, https://doi.org/10.1007/978-3-030-01267-0_40.

[21] K. A. Zhang, A. Cuesta-Infante, L. Xu, and K. Veeramachaneni, "SteganoGAN: High Capacity Image Steganography with GANs." arXiv, Jan. 30, 2019, https://doi.org/10.48550/arXiv.1901.03892.

[22] Z. Wang, M. Zhou, B. Liu, and T. Li, "Deep Image Steganography Using Transformer and Recursive Permutation," *Entropy*, vol. 24, no. 7, June 2022, Art. no. 878, https://doi.org/10.3390/e24070878.

[23] "Steganography datasets and tools." DDE Lab, Binghamton University. https://dde.binghamton.edu/download/.