SE-Attn StegaVAE: a lightweight dual branch based variational auto-encoder with multi-objective loss for image steganography

Amerah Alabrah

Department of Information Systems, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia

ABSTRACT

Information security is crucial with the increasing data and internet usage. In data communication, Imaging data is one of the most frequently used data types for communication, especially in intelligence operations and in law enforcement use cases. Steganography is a method for hiding secrets in cover images, followed by complex image encryption and decryption methods. Previously, many Deep Learning (DL) methods have been proposed on steganography and achieved competitive results. However, there is still a need for lightweight, computationally less expensive DL models. Therefore, a computationally inexpensive, lightweight dual-branch based Variational Auto Encoder (VAE) model is proposed, namely SE-Attn StegaVAE. In this model, the Squeeze and Excitation (SE) block, Attention, and skipping connections are effectively used, whereas a sequentially optimal way is adapted to hide the secret in attention-oriented cover samples. Furthermore, a multi-objective loss function is proposed to penalize the model to hide secrets effectively and reconstruct them without any loss of information. In this study, three experiments have been performed with three different bit embeddings (2, 4, and 8) on the DIV2K dataset utilizing the SE-Attn StegaVAE model. Two-fold validation-based testing results outperformed as compared to State-of-the-Art (SOTA) methods and proven to be more error-free, computationally less expensive, with competitive similarity scores.

INTRODUCTION

Internet usage is increasing day by day, while in parallel, it is creating a lot of content. It created unprecedented access to data, which sometimes raises privacy concerns (*Hu et al., 2023*). Therefore, it is an urgent need to make our information pass through secure channels (*Zhang, Zhong & Wang, 2022*). For communication, audio, video, and image-type data is spread over the internet, while the most frequently used data over the internet is image-type data. Therefore, data should be secured for secure communication *via* images (*AbdelRazik et al., 2024*). For secure imaging communication, steganography is one of the key methods that is used to hide data (*Hu et al., 2023*). In steganography, a secret

message is communicated by hiding it in a cover image that is not viewable by the human eye (Zhang, Zhong & Wang, 2022). Steganography is the way to hide secret information in ongoing carrier data, which is then broadcast on public channels. The receiver can decrypt the carrier data and can perform steganalysis (Li et al., 2021). Steganalysis is of great importance in various fields, including military, political, medical, and others. Considering the need and significance of steganography, notable progress has been made recently in the field of neural steganography. In neural steganography, autoencoders and invertible methods have been used to create lossless and high-fidelity embeddings. For instance, Liu, Tang & Zheng (2022) proposed a lossless strategy that leverages an Invertible Neural Network (INN) to extract the exact data while hiding secret information with strong visual similarity, but these INN methods face practical trade-offs in payload size, robustness, and security. Furthermore, the deterministic nature of INNs and reliance on clean inversions also make them vulnerable to adversarial steganalysis and signal degradation.

Likewise, convolutional neural network (CNN) and transformer-based methods have been developed recently, such as the ResNet-based autoencoder method, which has been proposed by S.H.O. Researchers (Hashemi, Majidi & Khorashadizadeh, 2022). It embeds images with each other over 40 decibels (db) PSNR and 0.98 SSIM, showing high imperceptibility at high payload. These deep CNN and transformer-based models suffer from high computational complexity due to deep architectures, making them less efficient for real-time or resource-constrained applications. However, a few of the recent methods also adopted easy-to-hard stages based on gradual approaches to make the steganalysis more robust. For instance, another study (Liu, Zhang & Zhang, 2025) introduced a curriculum-based learning, namely STCL, mainly tailored for image steganography. It used an elementary teacher to an advanced teacher model-based approach. It achieved promising results by improving structual similarity index measure (SSIM), peak signal-to-noise ratio (PSNR), decoding strategy, and resistance to steganalysis via a structured training framework by framing it as easy-to-hard samples. However, its reliance on a teacher–student structure poses risks—if the teacher model is compromised, the overall approach is affected—and the dual-model dependence increases computational complexity.

An enhanced steganography method has been developed featuring a down-up structure for efficiency and long skipping connections preserving spatial information (Chen et al., 2022). It further incorporated the non-activated feature fusion mechanism for quality and attention mechanisms, aiming at imperceptibility. All these components are guided by a hybrid loss function aiming to have high-quality images and strong security while doing secret transmission. A lightweight steganography scheme uses graphical embedding and encryption to hide the data in images (Das, Durafe & Patidar, 2023). This method involved an Xception-like network to perform image-based steganalysis. Both statistical and machine-learning evaluations performed, achieved 2.55% more accuracy statistically and 50% in machine-learning analysis, which also outperformed existing state-of-the-art (SOTA) methods. However, despite its claims of being lightweight with a hybrid loss-based CNN framework, the reliance on an Xception-like network introduces high computational cost and extensive training requirements. Furthermore, effective data hiding and reconstruction still demand optimized latent space design and efficient decoding structures.

Another study relies on these deep learning (DL)-based image steganography methods, and proposes a style-based image steganography method by authors in *Hu et al. (2022)*, performing a novel method disguising secret message embeddings as image stylization. It hides the secret in the latent space while making a style transformation. Hiding data in the latent space during style transformation may compromise message robustness, making it vulnerable to distortions such as compression and resizing, while the stylization process itself introduces additional computational overhead. While earlier work primarily emphasized perceptual metrics, our method shifts the focus toward unexplored pixel-level numerical accuracy. Although it achieves comparable PSNR and SSIM to prior state-of-the-art approaches, it significantly reduces Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), leading to improved reconstruction fidelity, an essential requirement in sensitive domains such as medicine and law. Moreover, the proposed framework is lightweight compared to complex DL architectures, leveraging a Variational Auto-Encoder (VAE)-like design for efficiency. Furthermore, the aforementioned studies highlighted the complexity and challenges of decoding or reconstructing the secret from the cover, underscoring its critical importance. In the proposed study, this reconstruction problem is explicitly addressed and alleviated to a considerable extent. In a nutshell, this article contributed to the following:

- Proposed a dual-branch VAE based lightweight model SE-Attn StegaVAE that jointly learns to embed secret into compact latent space and sequentially fuse them in cover samples using an attention module.
- Integrated Attention, Squeeze and Excitation block, and an up-sampling strategy which ensures the secret features are encoded optimally for robust recovery of secret.
- Designed a multi-objective loss to cover pixel-wise, perceptual, gradient, and frequency domains to balance imperceptibility with secret reconstruction fidelity.
- SOTA comparison showed that the proposed method achieved competitive or superior secret reconstruction quality compared to previous SOTA methods, while maintaining acceptable cover image fidelity without requiring complex adversarial model training or complex architectures.

The rest of the article is divided into four sections: a section on related work covering recent advances in proposed methods for steganography and a subsequent proposed method section. The next section is the results and discussion based on the proposed methods and baseline methods. Lastly, the conclusion contains the proposed methods-based insights in terms of strengths and weaknesses and provides future directions.

RELATED WORK

Recently, many methods have been proposed to develop steganography methodologies to hide secrets more strongly and securely. These methods include various kinds of machine learning (ML) strategies, including neural networks, generative neural networks, and transformer-based approaches. It has recently been observed that secrets may be hidden successfully, but Joint Photographic Experts Group (JPEG) compression is problematic,

which is causing challenges while extracting secrets from the secret-containing image. Therefore, an invertible neural network is proposed, namely a robust image hiding network (*Jin et al., 2024*), which introduces a classifier that selects the extractor based on the compression degree. It further included the denoising module to enhance the robustness of the proposed method. The experimental results evaluated on the COCO, ImageNet, and DIV2K datasets showcase a 10 dB improvement in secret image recovery. However, as discussed in the introduction, INNs face trade-offs in payload, robustness, and security, and are vulnerable to adversarial attacks.

It is also observed that existing steganography methods face challenges in terms of security, capacity, invisibility, and especially with style transfers. Therefore, a dense invertible neural network is proposed (*Feng et al., 2024*) to provide a solution against these challenges, but INN has its own limitations too. It used a pre-trained style transfer network to hide the secret in cover images, which led to an impactful reduction in secret detection. Secondly, it introduces the feature reconstruction module and a semantic reconstruction loss. It further increases the robustness of the proposed method by adding a visual center and a self-calibration neck to improve the features' receptive field and to remove any redundancy, if any. Experimental results show that the proposed methods hide multiple secrets securely and extract them efficiently. Another problem in image steganography is with low hidden space, due to which secret images face noise and distortion. Therefore, an enhancement module has been added before and after the extraction process of an invertible neural network (*Yang et al., 2024*). It further claims that it is an overlooked common problem in steganography methods that rounding error is ignored. To overcome this non-differentiable nature of rounding distortion, a gradient approximation is performed. Experimental results prove the robustness and practicality of the proposed method. A different style of architecture, as a concealing and revealing network, is developed by authors (*Cheng et al., 2025*). It includes embedding secret images in concealing networks and the recovery of embedded images *via* revealing networks. It follows a down-up structure that utilizes the compressed intermediate feature maps and reduces the system's computational costs. It mainly included long skip connections, non-activated feature fusion, attention mechanism, and a hybrid loss function, which comprises pixel-domain loss and helps in preserving the structural and visual quality of hidden images. Although this approach is effective, it is computationally expensive.

DL relies on loss function-based proxies to adapt custom networks to produce desired outcomes. However, they have used fixed weight loss in addition to adding any component. Therefore, a two-stage curriculum-based learning loss scheduler network adapted to dynamically handle multinomial losses (*Zhang & Zhang, 2025*). In phase one, the initial focus was on learning information embedding from the original image *via* adversarial training and then shifting the learning focus to improve the decoding precision and, thirdly, generate the steganographic images to resist steganalysis. In phase 2, the loss is dynamically controlled to ensure balanced training. Experimental results evaluated on ALASKA2, VOC2012, and ImageNet datasets demonstrate the improved performance of steganography quality, decoding accuracy, and security. It is mentioned that generative networks face problems with information loss, hindering the capture of fine-grained image

feature information. Therefore, a better cost function needs to be designed while considering better probability maps. A U-Net generator architecture allowing focus on texture-enriched features is proposed (*Wang et al., 2024*), an extra input stream allowing enhancement in images is added in the generator, which enables you to learn the structural properties of images. It further added skip connections to facilitate information sharing between layers. Extensive experimental results demonstrated that the proposed approach effectively learns embeddings of probability maps and gains promising security against steganalysis attacks. Likewise, another generative artificial intelligence (AI) method, namely multi-dilated generative adversarial network (GAN), has been proposed (*Zhao & Huang, 2021*) which aims to improve the quality of hidden images. It improved the discriminator part of the SteganoGAN method by incorporating multiple convolutions and expansion of convolutional filters by varying rates. It allows more discriminative features to be included while composing secrets in cover images. Experimental results show that the improved discrimination strategy, SteganoGAN method, maintains a high steganographic capacity of 4.4 bits per pixel (BPP).

Pixel Value Differencing (PVD), as reflected in the name, it embeds data based on pixel value differences. It offers high capacity and imperceptibility computed by SSIM. Authors *Andono & Setiadi (2023)* optimize this trade-off using adaptive quantization and machine learning to adjust the embedding according to the given cover sample characteristics. However, with all these given improvements, PVD still shows inconsistent behavior of performance across image types, and adding preprocessing makes them more complex, which limits their scalability compared with DL approaches. Another study (*Rustad et al., 2022*) improved inverted least significant bit (LSB) steganography through adaptive pattern selection and reduced the embedding errors. It further achieved higher imperceptibility with PSNR 52-57 dB, SSIM roundly 0.999 at 1 BPP. The approach is lightweight and effective for a simple data hiding approach, but its applicability is mostly limited to bit-level embeddings. Therefore, it is less suitable for complex secret contents such as full image reconstruction. A recent review (*Setiadi et al., 2023*) claims that the evolution of digital image steganography, tracing its shifts from statistical and adaptive methods to AI-driven methods such as GANs, CNNs, VAEs, *etc*. Howevers, its key contributions are classifying steganographic methods by different downstream tasks such as security, imperceptibility, payload, robustness, and analyzing the assessment tools and datasets.

Another recent method claims that modifying the pixel values in natural images makes them detectable by steganalysis methods. However, aligning them with generative AI provides a more robust method, providing a strong embedding of hidden information based on an art-style image generation. In this proposed scheme (*Li et al., 2024*), an encoder-decoder approach is applied, which simultaneously generates the styled image containing the secret. It follows adversarial training, which enhances imperceptibility and makes the stego image indistinguishable from plain style transferred images. However, the presence of original cover style images makes it difficult to detect, whereas experimental results show that the approach withstands existing methods and achieves high results.

Traditional steganography methods face imperceptibility and security issues in textured images. A deep, dense invertible neural network proposed (*Duan et al., 2024*) ensures high imperceptibility and security by adding dense connections in the forward embedding process and designing a straightforward series of connections to develop the reverse extraction process too. Its non-invertible component utilizes a modified U-Net, which enables the high-quality secret image to be recovered by extracting the deep and fine-grained information of secret images. The experimental results show that an average 1.79 dB improvement of PSNR is achieved as compared to SOTA methods by evaluating DIV2K, COCO, and ImageNet datasets. Considering all the above-discussed recent studies on image steganography methods, a lot of improvements and robust invertible strategies have been adopted. However, their pixel-wise numerical accuracy, to reduce the errors, MAE, and RMSE, remains unexplored. To fill this gap, a lightweight, error-reduced, multi-objective loss-based Variational Auto-Encoder-based steganography method is proposed. It provides curriculum insights, attention, and skip connections in its architecture where necessary and outperforms SOTA methods to reduce the error in secret embedding and secret recovery.

METHODOLOGY

In the proposed methodology, a VAE-based encoder-decoder approach, SE-Attn StegaVAE, is followed by a multi-objective loss to reduce the error rates of secret embedding and recovery methods while maintaining the PSNR and SSIM scores. Recently, a survey (*Setiadi, Ghosal & Sahu, 2025*) highlighted that a paradigm shift moving beyond traditional fidelity measures (PSNR and SSIM) towards perceptual and semantic metrics is move, such as learned perceptual image patch similarity (LPIPS), fréchet inception distance (FID), and BERTScore that better reflect human perception. However, most current image steganography and image synthesis or similarity-related works (*Bai et al., 2025*) continue to support PSNR, SSIM, and RMSE for comparison. The current study follows the same convention, acknowledging the importance of integrating perceptual metrics in future evaluations. The proposed method utilizes a VAE-based backbone architecture that preserves the latent information as a latent space embedding. Furthermore, utilized the Squeeze Excitation (SE) blocks for channel-wise recalibration, which made the re-weighted adaptive features. It also includes attention blocks for focusing on cover image features. To preserve the multi-scale details belonging to cover and secret images, skipping connections are used. To blend the secret and cover images, residual fusion is performed. These multiple blocks and combinations are used in an effective way to hide the secret information in source or cover samples securely. In contrast to previous studies, SE-Attn StegaVAE was specifically designed as a VAE backbone to ensure robust latent space representations. SE blocks improve the channel-wise feature discrimination of fine details, whereas attention preserves the spatial details present in the given sample. Together with skipping and residual connections, fusion improves the reconstruction quality and the robustness beyond the conventional autoencoder-based approaches. All the building blocks of the proposed architecture (SE-Attn StegaVAE) are shown in Fig. 1.
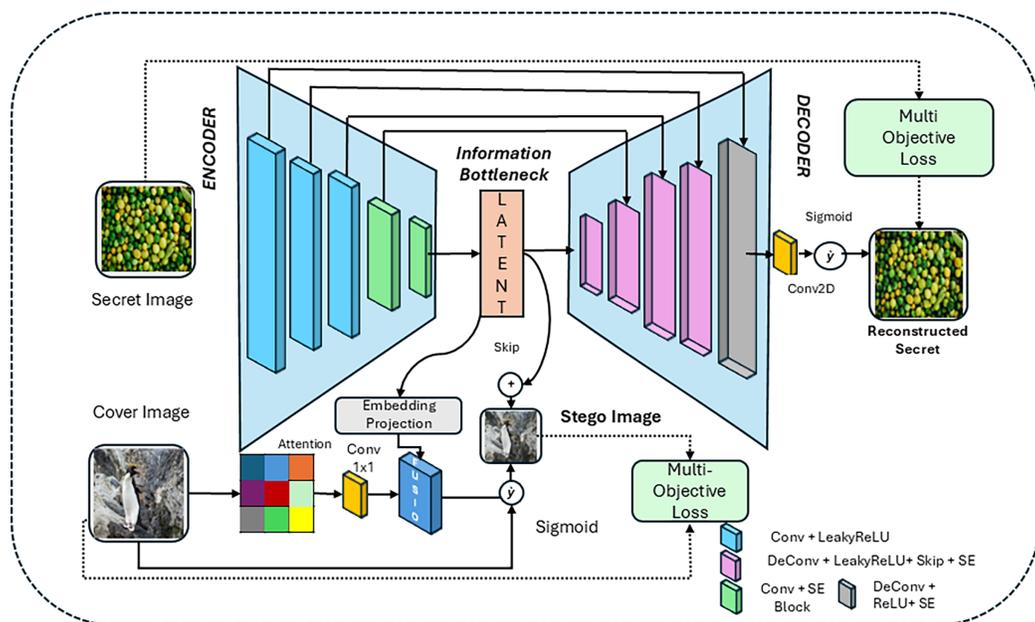
Figure 1 Training architecture of proposed SE-Attn StegaVAE model of steganography.
Full-size ⬚ DOI: 10.7717/peerj-cs.3425/fig-1

In Fig. 1, we can see that the VAE baseline is followed by having an Encoder-Decoder-like architecture with an information bottleneck by taking the secret image directly as input, whereas the cover or sample image is taken as attention and then propagated as an embedding project of the secret image and added into a fusion module. The fused information is passed from the sigmoid, having a stego image-like representation, whereas the skip connection is used to add a latent representation into the stego image. The resultant stego image is formed at this stage.

Secret encoder with SE blocks

In the proposed SE-Attn StegaVAE model, SE blocks are used in both the encoder and decoder parts. In the encoder part, it particularly focused on multi-scale latent feature representations. Utilizing the channel-wise representation learning of the SE block helps to provide an improved feature representation. The secret image $S \in \mathbb{R}^{C \times H \times W}$ is passed through a series of down-sampling convolutional layers. The multiple features-based representation is shown in Eq. (1).

$$F_{\text{Enc}} = f_n(S). \tag{1}$$

f_n represents feature representations where n indicates five number of representations. These five convolutional layers, with Leaky-rectified linear unit (LeakyReLU) activations are applied with 16, 32, 64, 128, and 256 number of features, respectively. The output features are represented as F_{Enc}. These features are further passed from the SE block to improve the representations and shown in Eq. (2).

$$F_{\text{SE}} = SE(F_{\text{Enc}}). \tag{2}$$

The SE block processed the features received from the encoder and delivered output features as F_{SE}. In the SE block, the features are processed *via* global average adaptive pooling and then passed through multiple fully-connected (FC) and LeakyReLU layers.

$$Z = g(F_{SE}). \tag{3}$$

In the SE block, the features passed through FC layers and sigmoid function are applied, and the output is represented as Z, shown in Eq. (3) not 23. Here, Z is known as the latent representation, which is used to be hidden in the cover image, discussed in coming section.

Skip connected decoder enriched with SE blocks

The decoder part of the model utilized the latent representations and skipping connections from the encoder of the model. It followed the 5-step sequence of transposed convolutional layers for up-sampling layers. After each up-sampling, the SE block is added to fuse these up-sampled features with skip connection-based encoder representations.

$$U_n = \text{Up}(Z). \tag{4}$$

The up-sampled sequences are represented as U_n, whereas n represents five sequences of up-sampling layers, and each of them passed through SE blocks. However, they have used mainly the Z embeddings as starting input features in the first up-sampling layer, as shown in Eq. (4).

$$S = \sigma(\text{Conv}(U_n)). \tag{5}$$

Subsequently, a sigmoid function (σ) is utilized for the final up-sampled features, followed by a convolutional layer as represented in Eq. (5).

Attention added cover-image fusion

All processes in the encoder and decoder sections are made to hide the secret information and to decode it in a secure way without distorting or losing information. To hide the secret information in a cover image, the attention-based features fusion is performed by utilizing the latent Z information. The projected embeddings from latent representations are used and fed to up-sample to match with cover image size as shown in Eq. (6).

$$E = \text{Up}(\text{Proj}(Z)). \tag{6}$$

By keeping secret image embeddings, an attention encoded block is directly applied on the cover image and highlights the salient regions of the cover image, followed by a sigmoid function (σ) as shown in Eq. (7).

$$A = \sigma(h(C)). \tag{7}$$

In Eq. (7), A represents the output of attention maps of cover image C, whereas h is the highlighted attention block applied on cover image, followed by the sigmoid function (σ) to get the final output. The latent representations E and attention maps-based features (A) are passed through a feature fusion block as represented in Eq. (7).

$$F_{\text{fuse}} = \text{Conv}[A, E]. \tag{8}$$

Table 1 Block by block architecture of proposed SE-Attn StegaVAE model.

Stage/Block	Layer type & parameters	Input shape	Output shape
Encoder (Secret image)	Conv2D (3 → 64, k = 4, s = 2, p = 1) + ReLU	(B, 3, 256, 256)	(B, 64, 128, 128)
	Conv2D (64 → 128, k = 4, s = 2, p = 1) + ReLU	(B, 64, 128, 128)	(B, 128, 64, 64)
	Conv2D (128 → 256, k = 4, s = 2, p = 1) + ReLU	(B, 128, 64, 64)	(B, 256, 32, 32)
	Conv2D (256 → 512, k = 4, s = 2, p = 1) + ReLU	(B, 256, 32, 32)	(B, 512, 16, 16)
	SEBlock (channel recalibration)	(B, 512, 16, 16)	(B, 512, 16, 16)
	Conv2D (512 → 1,024, k = 1) → Latent	(B, 512, 16, 16)	(B, 1,024, 16, 16)
Decoder (Secret recon)	DeConv2D (1,024 → 256, k = 4, s = 2, p = 1) + ReLU	(B, 1,024, 16, 16)	(B, 256, 32, 32)
	Skip + SEBlock	concat (256 + 256)	(B, 512, 32, 32)
	DeConv2D (512 → 128, k = 4, s = 2, p = 1) + ReLU	(B, 512, 32, 32)	(B, 128, 64, 64)
	Skip + SEBlock	concat (128 +128)	(B, 256, 64, 64)
	DeConv2D (256 → 64, k = 4, s = 2, p = 1) + ReLU	(B, 256, 64, 64)	(B, 64, 128, 128)
	Skip + SEBlock	concat (64 + 64)	(B, 128, 128, 128)
	DeConv2D (128 → 32, k = 4, s = 2, p = 1) + ReLU	(B, 128, 128, 128)	(B, 32, 256, 256)
	SEBlock	(B, 32, 256, 256)	(B, 32, 256, 256)
	Conv2D (32 → 3, k = 3, p = 1) + Sigmoid	(B, 32, 256, 256)	(B, 3, 256, 256) (Reconstructed Secret)
Fusion (Cover + Secret)	AttentionBlock (3 → 3, convs + sigmoid mask)	(B, 3, 256, 256)	(B, 3, 256, 256)
	Conv2D (1,024 → 16, k = 1) + Bilinear Upsampling	(B, 1,024, 16, 16)	(B, 16, 256, 256)
	Concatenate [cover_attn, latent_emb] + Conv2D (19 → 3)	(B, 19, 256, 256)	(B, 3, 256, 256)
	Residual Addition + Sigmoid	(B, 3, 256, 256)	(B, 3, 256, 256) (Stego Image)

After fusing both these features, the secret is well hidden in the cover sample, and a new stego image is generated, whereas residual connections are used while fusing the cover image; these fused feature representations are shown in Eq. (9).

$$\hat{C} = \sigma(C + F_{\text{fuse}}). \tag{9}$$

The final stego image $\hat{C}$ is the combination of the cover image (C) and fused features of cover and secret images, followed by a residual connection. The full architecture of the proposed model with layer details is described below in Table 1.

In given configurations of the model, the three core stages of the proposed model are explained. Comprising these core stages: the encoder compresses the secret image, the decoder reconstructs it, and the fusion stage embeds the secret into the cover image to produce the stego output. For a less noisy and lossless reconstruction, SE blocks and attentions were added. As a result, a VAE-style backbone provides latent representations, SE blocks enhance channel-wise feature discrimination, and attention preserves spatial details. These components, combined with skip and residual connections, improved the reconstruction quality and robustness beyond conventional autoencoders.

Multi-objective loss

There are different modules added in the proposed SE-Attn StegaVAE model through which multiple parameters flow in the final gradient. These multiple objectives give a final secure stego image and a final distortion of free secret decoding from the decoder part of the model. In multi-objective loss, the predicted and original images of cover and secret are

Alabrah (2026), *PeerJ Comput. Sci.*, DOI 10.7717/peerj-cs.3425

9/19

Table 2 Loss function parameter values for cover and secret images.

Parameter	Cover (value)	Secret (value)
α	0.005	1
β	10	2
γ	20	10
δ	0.2	0.1
ρ	0.05	1

processed from a combined loss of cover and secret images. The combined loss of cover and secret images is mathematically represented in Eq. (10).

$$L_{\text{combined}} = \text{MultiObjectiveLoss}(L_{\text{cover}} + L_{\text{secret}}). \tag{10}$$

The predicted images are stego and reconstructed secret images, whereas the stego image is a mimic-like cover image hiding the secret information, and from the stego image, the secret is reconstructed. Therefore, two losses have been used in total, one for the cover and one for the secret image. Each of the losses, part of the combined loss, is used and propagated backward as the sum of both. Each of the loss components is further composed of five components, covering multiple objectives.

$$L = \alpha L_1^{\text{pixel}} + \beta(1 - \text{SSIM}) + \gamma L_{1,\text{VGG}}^{\text{perceptual}} + \delta L_{1,\text{edge}}^{\text{gradient}} + \rho L_{1,\text{FFT}}^{\text{freq}}. \tag{11}$$

The final loss function represented in Eq. (11) contains five different components which are traversed from different points of the proposed model. All five components are controlled with the help of α, β, γ, δ, and ρ weighting parameters. Based on performance, these hyperparameters are tuned, whereas their optimal values for cover and secret images are shown in Table 2.

These values are finalized to make the final loss function stable, having all loss components in a finite range. There are three different experiments conducted based on image bit values, images are preprocessed as 8-bit, having a 0–255 range, 4-bit, having a 0–15 range, and 2-bit, having a 0–3 range of values in pixels. These three different pixel range-based training progress in terms of training and validation losses are shown in Fig. 2.

The training progress of all three embedding-based models is shown in Fig. 2. In sections a, d, and g, loss progress is shown over 50 epochs. The training and validation losses continue to decrease, indicating the validity of the proposed multi-objective loss approach. The fact of loss proxy to improve the model is shown in PSNR and SSIM curves, which are shown in other sections, such as in the b and c sections of the Figure, the 8-bit embeddings based PSNR and SSIM progress is shown. The SSIM reached up to 0.95, whereas PSNR went up to 30 with only 50 epochs. The further training of the model keeps the model improving but till 50, there is significant improvement which has been achieved. The PSNR and SSIM training of 4 and 2-bit embedding-based models, the PSNR reached up to 29, whereas SSIM reached more than 90, as shown in sections e and f of Fig. 2. For 2-bit embeddings, the performance slightly gets down as 22 PSNR and 0.75 SSIM, it depends upon the bit range, as their low variation in pixel values, therefore, their shown
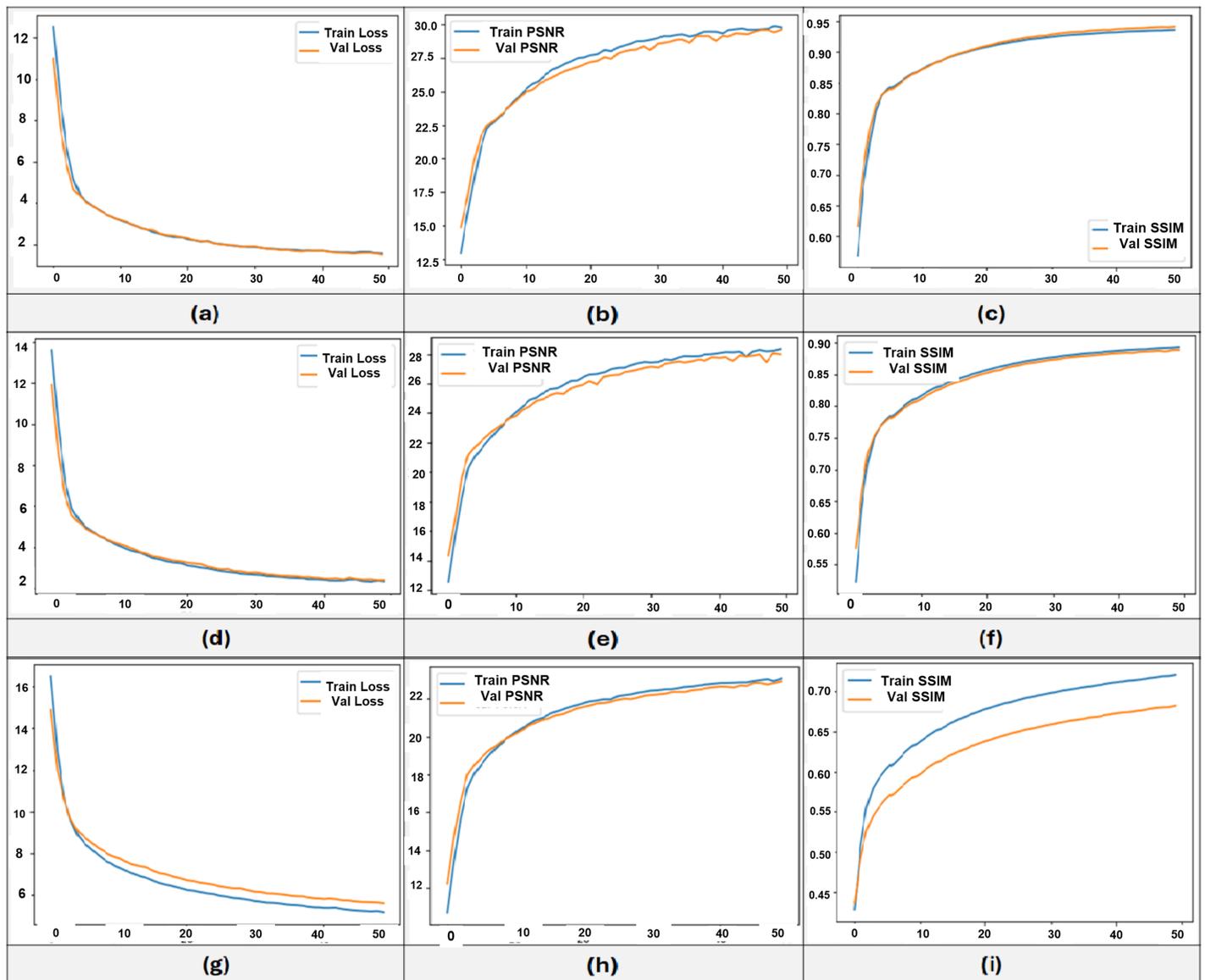
Figure 2 Training and validation curves of SE-Attn StegaVAE model training on three experiments of 8-bit (A–C), 4-bit (D–F), and 2-bit (G–I). Full-size ☑ DOI: 10.7717/peerj-cs.3425/fig-2

difference is also low that's why their scores are slightly lower, but qualitatively, the results are confident as discussed in the qualitative analysis of 'Conclusion'.

RESULTS AND DISCUSSION

The VAE-based baseline architecture is followed while proposing the SE-Attn StegaVAE model. The proposed model achieved SOTA results while comparing with previous methods. To prove the robustness of the proposed model, the DIV2k dataset was used for training and testing. However, images are processed and saved in 8, 4, and 2 bits while performing, training, and testing of the model. There are four evaluation metrics, namely PSNR, SSIM, RMSE, and MAE are calculated against all three bits of images.

Table 3 Dataset description regarding training and testing samples of secret and cover images, their resolution and value ranges depending on number of bits.

Set	Number of samples	Resolution	Number of bits	Range of values
Training secret images	400	$256 \times 256 \times 3$	8, 4, 2	0–255, 0–15, 0–3
Training cover images	400	$256 \times 256 \times 3$	8, 4, 2	0–255, 0–15, 0–3
Testing secret images	100	$256 \times 256 \times 3$	8, 4, 2	0–255, 0–15, 0–3
Testing cover images	100	$256 \times 256 \times 3$	8, 4, 2	0–255, 0–15, 0–3

Dataset description

The DIV2K dataset was originally introduced by Agustsson and Timofte (ETH Zürich) as part of the CVPR 2017 Challenge on Single Image Super Resolution, *Timofte et al. (2017, 2018)*, *Ignatov et al. (2018)* accessible *via* Kaggle from the provided link (https://www.kaggle.com/datasets/joe1995/div2k-dataset) is utilized in this study to propose a new method for image steganography. The dataset contains a diverse range of images; therefore, the proposed method will be robust once provided with a good trade-off of metrics. Dataset detail is given in Table 3 regarding training and testing samples of cover and secret images, their resolution, and their range of bits.

The table presented shows that half of the training samples from 800 in the DIV2K dataset are used as secrets and half as cover samples; similarly, the 100 samples are taken for testing purposes, from which 50 are used as cover and 50 as secret images. The selection of these samples in training and testing is totally random, but keeps the same samples while performing bitwise experiments.

Qualitative analysis

The performance of models in terms of stego image generation and deconstruction is quite promising. It reduces the error and improves the overall performance of the encryption and decryption. As three experiments were conducted on datasets having 8, 4, and 2-bit images, therefore a visual exampled corresponding to each experiment is shown in Fig. 3.

The 8-bit embedding-based model testing results are shown in row 1, having stego-cover scores of 0.96 structural similarity SSIM, and 32.98 PSNR, whereas very low errors of 0.018 MSE and 0.029 RMSE. Similarly, in secret and reconstructed secret, the SSIM is 0.91 and 31.2 dB PSNR, and low error of 0.0192 MAE and 0.0264 RMSE scores. In 4-bit embedding-based testing, 0.93 SSIM and 31.0 dB PSNR are achieved in stego-cover, but lower results are achieved in secret recovery with 0.776 SSIM and 28.1 dB PSNR scores, having a lower error of 0.028 MAE and 0.03 RMSE. In 2-bit embeddings, the model shows lower PSNR and SSIM as compared to 8-bit and 4-bit embeddings. the stego-cover shows 0.78 SSIM and 23.3 dB PSNR, and same errors are very less, likewise in secret reconstruction, the lower scores are showing due to less variability in pixel range values but having lower MAE and RMSE as same of 8- and 4-bit indicates that model showing promising results on all embeddings based settings.

Number of bits	Cover	Secret	Stego	Re-constructed	Stego-Cover Scores	Secret-Reconstructed Scores
8-bit Testing					SSIM = 0.9633 PSNR = 32.9 dB MAE = 0.0184 RMSE = 0.029	SSIM = 0.9119 PSNR = 31.6 dB MAE = 0.0192 RMSE = 0.0264
4-bit Testing					SSIM = 0.9322 PSNR = 31.0 dB MAE = 0.020 RMSE = 0.028	SSIM = 0.7776 PSNR = 28.1 dB MAE = 0.028 RMSE = 0.03
2-bit Testing					SSIM = 0.7876 PSNR = 23.3 dB MAE = 0.058 RMSE = 0.068	SSIM = 0.7918 PSNR = 23.3 dB MAE = 0.032 RMSE = 0.054

Figure 3 Original secret, cover, and predicted stego, reconstructed secret images from three different embeddings (Row 1 = 8 bit. Row 2 = 4 bit, Row 3 = 2 bit) and their corresponding achieved scores.
Full-size 🖼 DOI: 10.7717/peerj-cs.3425/fig-3

Table 4 Testing set results of Cover/Stego samples from DIV2K dataset on three different bits-based embeddings (8, 4 and 2).

Bit	Fold	SSIM	PSNR	MAE	RMSE
8	Fold 1	0.9373	30.21	3.82	5.58
	Fold 2	0.9455	30.47	3.90	5.37
	Average	0.9414	30.34	3.86	5.47
4	Fold 1	0.8972	28.47	7.35	9.64
	Fold 2	0.8990	28.78	7.04	9.29
	Average	0.8981	28.62	7.19	9.46
2	Fold 1	0.7305	23.27	15.52	17.49
	Fold 2	0.7222	23.30	15.57	17.51
	Average	0.7235	23.29	15.54	17.50

Quantitative analysis and metrics trade-off

For qualitative analysis, we have used the SSIM for structural analysis, PSNR for noise or distortion level analysis, MAE is used to quantify the difference between predicted and original images whereas RMSE is squared to errors where we can estimate the larger errors to make corresponding corrections in loss or weights of the model. There are three different experiments based on image bit value that are performed, and their corresponding results are shown in Table 4 as cover/stego-based testing evaluation. Likewise, the secret and recovery of the testing data are shown in Table 5.

Table 4 shows the cover *vs.* stego images quality for three different bits (2, 4, and 8 bits per pixel). There is a clear trend in results: increasing bit depth significantly improves the

Table 5 Testing set results of secret/recovery samples achieved on three different number of bits-based images.

Bit	Fold	SSIM	PSNR	MAE	RMSE
8	Fold 1	0.9666	33.24	3.82	5.56
	Fold 2	0.9657	33.54	3.89	5.40
	Average	0.9662	33.39	3.86	5.48
4	Fold 1	0.9256	28.47	6.04	8.22
	Fold 2	0.8990	28.78	6.48	8.74
	Average	0.9123	28.63	6.26	8.48
2	Fold 1	0.8858	24.99	8.99	14.38
	Fold 2	0.8784	24.74	9.28	14.78
	Average	0.8821	24.87	9.14	14.58

results, with higher similarities in cover and stego images. For 8-bit embeddings, the method achieved an average SSIM of 0.9414 and PSNR of 30.34 dB, indicating high visual fidelity and minimal perceptual differences from the original cover. At 4-bit, the SSIM drops moderately to 0.8981 and the PSNR of 23.29 dB. Look at 2-bit settings, which show a noticeable distortion, with SSIM of 0.7235 and PSNR of 23.29 dB. The increase in MAE and RMSE at lower depths reflects larger pixel-level errors due to the limited embedding capacity. Overall, these results demonstrate that the proposed method effectively maintains cover imperceptibility at higher capacities while highlighting the trade-off between payload and visual quality when embedding constraints are tight. In a nutshell, the increasing embedding capacity from 2-bit to 8-bit significantly improved stego cover fidelity, with SSIM increasing from 0.72 to 0.94 and PSNR rising from 23 to 30 dB. It demonstrated that the proposed model effectively adapts to different payload constraints, but embedding larger payloads at lower bit depths introduces visible distortion as reflected by higher MAE and RMSE.

Table 5 summarizes the secret and reconstructed secret samples from stego image quality for 2-bit, 4-bit, and 8-bit embeddings. We observe that higher embedding bit depths consistently improve secret recovery performance. For the 8-bit testing, the proposed method achieved an average SSIM of 0.9662 and PSNR of 33.39 dB, with a low MAE of 3.86 and RMSE of 5.48. It demonstrates that high fidelity recovers with a minimal pixel error. When embedded at 4-bit, the secret SSIM drops moderately to 0.9123 and the PSNR of 28.63 dB. reflecting a little bit higher distortion. At 2-bit with lower capacity, secret recovery becomes more challenging, with average SSIM of 0.8821 and PSNR of 24.87 dB, which is higher compared to cover/stego recovery, which was discussed above in Table 4. Though it becomes more challenging to recover when we don't have enough pixel capacity, still results are promising if we look at qualitative results. Overall, Table 5 results confirm that the proposed methods maintain robust secret reconstruction at higher capacities and highlight the trade-off between payload and reconstruction quality at lower embedding levels.

Table 6 Comparison of testing results of cover/stego achieved by the proposed method with SOTA methods.

Method	SSIM	PSNR (dB)	RMSE	MAE
CAIS (*Zheng et al., 2023*)	0.969	39.57	4.25	3.08
HiDDeN (*Zhu et al., 2018*)	0.969	35.21	6.82	6.98
Baluja (*Baluja, 2017*)	0.964	36.77	5.02	3.79
Proposed	0.9414	30.34	3.86	5.47

Table 7 Comparison of the proposed method with existing methods on secret recovery performance.

Method	SSIM	PSNR (dB)	RMSE	MAE
CAIS (*Zheng et al., 2023*)	0.965	36.10	5.80	4.36
HiDDeN (*Zhu et al., 2018*)	0.969	36.43	5.50	6.02
Baluja (*Baluja, 2017*)	0.937	35.88	6.11	4.68
Proposed	0.9662	33.39	3.86	5.48

SOTA methods comparison

After qualitative analysis and metrics trade-off, the proposed method results are compared with SOTA methods that used the same DIV2k dataset. Considering the computational cost, the results of SOTA methods are taken from their reported results instead of implementing all of them on our end. The comparison with SOTA methods is shown for Cover/Stego and Secret/Reconstructed images in Tables 6 and 7.

In Table 6, cover and stego images-based comparison is shown with SOTA methods. It looks that the proposed method did not improve the results in terms of hiding secret as compared to SOTA methods, but achieved competitive results. If we focus on error rates, the RMSE score is lower than all compared SOTA methods. We can say that the model achieved competitive results as compared to SOTA results with slight distortion. However, reconstruction or recovery of the proposed method is better than SOTA methods, as shown in Table 7.

While PSNR is commonly used to assess visual fidelity, it is sensitive to global variations and may not fully capture perceptual quality. Our method achieves lower PSNR compared to CAIS and Baluja but consistently yields smaller RMSE values (Tables 6 and 7). This suggests that the proposed model reduces pixel-wise reconstruction error, which benefits accurate secret recovery. Similarly, although the SSIM score is slightly lower for cover–stego comparison, the secret recovery performance (Table 7) demonstrates competitive SSIM and significantly reduced RMSE, highlighting that the method prioritizes reliable recovery without compromising overall imperceptibility. These results show that RMSE and MAE provide complementary insights alongside PSNR and SSIM, and together they confirm the robustness of the proposed approach. It is worth noting that the proposed method achieved competitive results using a lightweight VAE-style model and with very low training of 50 epochs only; therefore, the model is very light and robust as compared to SOTA methods. It further used a multi-objective loss function that covers the gradient, pixel-wise, structural, and frequency domain features to train the model. While reported

results are based on 50 epochs only and with some constrained hyperparameter values, it can be further improved by tuning the architecture and parameters. In calculating the testing results of the proposed method, the 2-fold scheme is applied in which 50/50 samples are taken as secret and cover samples, whereas their average is reported as a hundred samples.

CONCLUSION

In internet communication, the data is vulnerable and very much in the public domain. Therefore, secure communication and the utilization of secure communication are badly needed. For data encryption and decryption, a lot of information security strategies have been proposed previously. Likewise, steganography is one method to hide the secret in cover samples. For law enforcement and intelligence operations, steganography is still utilized to this day. To improve steganography, a lot of methods have been previously proposed, which have problems of computationally expensive and time-consuming adversarial models training; likewise, they have problems in image reconstruction. To tackle challenges, this study utilized a dual-branch based lightweight VAE-based model, namely SE-Attn StegaVAE. Furthermore, a multi-objective covering loss function is proposed, which utilizes five components covering spatial, gradient, pixel-wise, structural, and frequency features. Same loss function applied with different weighting parameters for cover/stego and secret/reconstructed images. Three different bit embeddings (2, 4, and 8) have been used to prove the capability of the proposed model to hide and reconstruct the given secret images. Two-fold testing was performed on the testing set of the DIV2K dataset; the achieved results outperformed SOTA methods with competitive scores and proved the computationally less expensive and lightweight model.

LIMITATIONS AND FUTURE WORK

Despite the promising results, the proposed method has some limitations. First, when operating at low bit depths, there is a noticeable drop in performance, as reduced embedding capacity constrains the ability to preserve both imperceptibility and recovery quality. Second, the model can be sensitive to the content and distribution of cover images—textures with fine-grained or high-frequency details may introduce reconstruction artifacts or reduce robustness. Using more datasets and following the VAE-style architecture, a more robust model could be developed to improve the steganography results. Furthermore, adding more important components to the loss function could penalize the model to perform more robustly. The dataset used for experimental purposes in this study contains a limited number of samples for both training and testing. While this provided sufficient scope to validate the proposed approach, leveraging a larger dataset could further enhance model training and yield more confident and generalizable results. Finally, the current approach has not been extensively validated under adversarial steganalysis or compression scenarios, which remain important directions for future work.

Alabrah (2026), *PeerJ Comput. Sci.*, DOI 10.7717/peerj-cs.3425

16/19

ADDITIONAL INFORMATION AND DECLARATIONS

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Amerah Alabrah conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The DIV2K dataset used in this study are available at Kaggle: https://www.kaggle.com/datasets/joe1995/div2k-dataset, (Agustsson and Timofte, ETH Zürich, CVPR 2017 Challenge).

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj-cs.3425#supplemental-information.

REFERENCES

AbdelRazik HF, Saad S, Salah El-Sayed A, Mahmoud AM. 2024. Image steganography: a comparative and practical study. *International Journal of Intelligent Computing and Information Sciences* **24(2)**:41–57.

Andono PN, Setiadi DRIM. 2023. Quantization selection based on characteristic of cover image for PVD steganography to optimize imperceptibility and capacity. *Multimedia Tools and Applications* **82(3)**:3561–3580 DOI 10.1007/s11042-022-13393-y.

Bai H, Che T, Zhang J, Li S. 2025. BMENet: 3D brain MRI enhancement against various degradations with coarse to fine network. In: *2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI)*. Piscataway: IEEE.

Baluja S. 2017. Hiding images in plain sight: deep steganography. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R, eds. *Advances in Neural Information Processing Systems*. Vol. 30. Red Hook: Curran Associates, Inc.

Chen F, Xing Q, Sun B, Yan X, Cheng J. 2022. An enhanced steganography network for concealing and protecting secret image data. *Entropy* **24(9)**:1203 DOI 10.3390/e24091203.

Cheng Y, Zhou J, Chen J, Yin Z, Zhang X. 2025. RFNNS: robust fixed neural network steganography with popular deep generative models. ArXiv DOI 10.48550/arXiv.2505.04116.

Das D, Durafe A, Patidar V. 2023. An efficient lightweight LSB steganography with deep learning steganalysis. In: *Computational Intelligence in Image and Video Processing*. Boca Raton, Florida: Chapman and Hall/CRC, 131–154.

Duan D, Shen S, Yu S, Yuan Y, Zhou Q, Lv H, Lin H. 2024. DenseJIN: dense depth image steganography model with joint invertible and noninvertible mechanisms. *IEEE Transactions on Circuits and Systems for Video Technology* **35**:1631–1647 DOI 10.1109/TCSVT.2024.3476421.

Feng W, Liu Y, Ye W, Tan D. 2024. Secure multiple image steganography based on invertible neural network and neural style transfer. In: *Sixteenth International Conference on Digital Image Processing (ICDIP 2024)*. Vol. 13274, SPIE, 446–456.

Hashemi SHO, Majidi M-H, Khorashadizadeh S. 2022. Color image steganography using deep convolutional autoencoders based on resnet architecture. ArXiv DOI 10.48550/arXiv.2211.09409.

Hu K, Huang Z, Wang X, Wang X. 2023. StegaEdge: learning edge-guidance steganography. *The Visual Computer* **39(8)**:3319–3331 DOI 10.1007/s00371-023-02974-z.

Hu D, Zhang Y, Yu C, Wang J, Wang Y. 2022. Image steganography based on style transfer. ArXiv DOI 10.48550/arXiv.2203.04500.

Ignatov A, Timofte R, Van Vu T, Minh Luu T, Pham T, Van Nguyen C, Kim Y, Choi J-S, Kim M, Huang J, Ran J, Xing C, Zhou X, Zhu P, Geng M, Li Y, Agustsson E, Gu S, Van Gool L, de Stoutz E, Kobyshev N, Nie K, Zhao Y, Li G, Tong T, Gao Q, Hanwen L, Navarrete Michelini P, Dan Z, Fengshuo H, Hui Z, Wang X, Deng L, Meng R, Qin J, Shi Y, Wen W, Lin L, Feng R, Wu S, Dong C, Qiao Y, Vasu S, Thekke Madam N, Kandula P, Rajagopalan AN, Liu J, Jung C. 2018. Pirm challenge on perceptual image enhancement on smartphones: report. In: *European Conference on Computer Vision (ECCV) Workshops*.

Jin X, Pan C, Cheng Z, Dong Y, Jiang Q. 2024. RIHINNet: a robust image hiding method against JPEG compression based on invertible neural network. *IET Image Processing* **18(13)**:3801–3814 DOI 10.1049/ipr2.13209.

Li L, Zhang X, Chen K, Feng G, Wu D, Zhang W. 2024. Image steganography and style transformation based on generative adversarial network. *Mathematics* **12(4)**:615 DOI 10.3390/math12040615.

Li P, Li Y, Wang H, Liu C. 2021. Research on steganalysis of digital image based on deep learning. In: *2021 4th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE)*. Los Alamitos, CA, USA: IEEE Computer Society, 528–534.

Liu L, Tang L, Zheng W. 2022. Lossless image steganography based on invertible neural networks. *Entropy* **24(12)**:1762 DOI 10.3390/e24121762.

Liu F, Zhang T, Zhang C. 2025. STCL: curriculum learning strategies for deep learning image steganography models. ArXiv DOI 10.48550/arXiv.2504.17609.

Rustad S, Moses Setiadi DRI, Syukur A, Andono PN. 2022. Inverted LSB image steganography using adaptive pattern to improve imperceptibility. *Journal of King Saud University—Computer and Information Sciences* **34(6, Part B)**:3559–3568 DOI 10.1016/j.jksuci.2020.12.017.

Setiadi DRIM, Ghosal SK, Sahu AK. 2025. AI-powered steganography: advances in image, linguistic, and 3D mesh data hiding—a survey. *Journal of Future Artificial Intelligence and Technologies* **2(1)**:1–23 DOI 10.62411/faith.3048-3719-76.

Setiadi DRIM, Rustad S, Andono PN, Shidik GF. 2023. Digital image steganography survey and investigation (goal, assessment, method, development, and dataset). *Signal Processing* **206(3)**:108908 DOI 10.1016/j.sigpro.2022.108908.

Timofte R, Agustsson E, Van Gool L, Yang M-H, Zhang L. 2017. NTIRE 2017 challenge on single image super-resolution: methods and results. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. Piscataway: IEEE.

Timofte R, Gu S, Wu J, Van Gool L. 2018. NTIRE 2018 challenge on single image super-resolution: methods and results. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. Piscataway: IEEE.

Wang D, Yang G, Chen J, Ding X. 2024. GAN-based adaptive cost learning for enhanced image steganography security. *Expert Systems with Applications* **249(7553)**:123471 DOI 10.1016/j.eswa.2024.123471.

Yang H, Xu Y, Liu X, Ma X. 2024. PRIS: practical robust invertible network for image steganography. *Engineering Applications of Artificial Intelligence* **133(7253)**:108419 DOI 10.1016/j.engappai.2024.108419.

Zhang FLT, Zhang C. 2025. TSCL: multi-party loss balancing scheme for deep learning image steganography based on curriculum learning. ArXiv DOI 10.48550/arXiv.2504.18348.

Zhang Y-Q, Zhong K, Wang X-Y. 2022. High-capacity image steganography based on discrete hadamard transform. *IEEE Access* **10**:65141–65155 DOI 10.1109/access.2022.3181179.

Zhao X, Huang H. 2021. Research on image steganography based on multiple expansion generation adversarial network. In: *2021 IEEE 3rd International Conference on Frontiers Technology of Information and Computer (ICFTIC)*. Piscataway: IEEE, 361–366.

Zheng Z, Hu Y, Bin Y, Xu X, Yang Y, Shen HT. 2023. Composition-aware image steganography through adversarial self-generated supervision. *IEEE Transactions on Neural Networks and Learning Systems* **34(11)**:9451–9465 DOI 10.1109/tnnls.2022.3175627.

Zhu J, Kaplan R, Johnson J, Fei-Fei L. 2018. Hidden: hiding data with deep networks. In: *Proceedings of the European Conference on Computer Vision (ECCV)*.