Implicit Steganography Beyond the Constraints of Modality

Sojeong Song[†,1], **Seoyun Yang**[†,1], **Chang D. Yoo**[‡,1], **and Junmo Kim**[‡,1]

[1] Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea
akqjq4985@kaist.ac.kr, seoyun.yang@navercorp.com

Abstract. Cross-modal steganography is committed to hiding secret information of one modality in another modality. Despite the advancement in the field of steganography by the introduction of deep learning, cross-modal steganography still remains to be a challenge to the field. The incompatibility between different modalities not only complicate the hiding process but also results in increased vulnerability to detection. To rectify these limitations, we present INRSteg, an innovative cross-modal steganography framework based on Implicit Neural Representations (INRs). We introduce a novel network allocating framework with a masked parameter update which facilitates hiding multiple data and enables cross modality across image, audio, video and 3D shape. Moreover, we eliminate the necessity of training a deep neural network and therefore substantially reduce the memory and computational cost and avoid domain adaptation issues. To the best of our knowledge, in the field of steganography, this is the first to introduce diverse modalities to both the secret and cover data. Detailed experiments in extreme modality settings demonstrate the flexibility, security, and robustness of INRSteg.

Keywords: Steganography · Implicit neural representations · Cross-modal steganography

1 Introduction

Nowadays, an unprecedented volume of data is shared and stored across various digital platforms and the volume is bound to intensify even further in the future. This proliferation and accessibility of data contributes to convenience in our lives and enables innovation and advancement that weren't achievable in the past; however, it also brings a concerning challenge to the security of information. Steganography aims to hide the secret information inside a cover data resulting in a stego data, also called the container. Unlike cryptography,

[†]: Both authors have equally contributed.
[‡]: Corresponding author.

which investigates in hiding the data of interest in a coded form, steganography operates with imperceptible concealment so that the existence of the secret data is undetectable. For example, individuals can upload their stego data to the cloud and securely trade their private keys with specific parties, and companies can leverage steganography to exchange confidential information, such as new product designs or sensitive meeting recordings, while including ownership details in the data. From a government perspective, steganography can covertly transmit information to foreign spies by inserting confidential national security information into seemingly ordinary cover data.

Up to now, image is the most commonly used modality of cover data, as it is less sensitive to human perception and is comparatively simple to embed secret information. However, image steganography tends to have a limit on payload capacity, and is susceptible to detection. Deep learning methods [1, 30] have increased the security, but suffer from unstable extraction and high computational costs [20]. Other modalities such as audio, video and 3D shape also undergo similar limitations [11,29] and struggle with trade-offs between capacity and security. Moreover, for cross-modal steganography, where the modality of the secret and the cover data differs [5, 6, 24], these limitations intensify due to the difficulties that emerge when the dimension or capacity of the secret data exceed that of the cover data or when the secret data contains temporal information.

Implicit neural representations (INRs) have recently attracted extensive attention as an alternative data representation, using continuous and differentiable functions to represent data via parameters of a neural network. The concept of INRs offers a more flexible and expressive approach than the conventional notion of discrete representations. Multiple modalities, as audio, image, video and 3d shape, are all presentable using a single neural network in various resolutions [3]. Moreover, INRs benefit in capacity by reducing the memory cost according to the network design [9, 15].

In this paper, we present INRSteg, the first cross-modal steganography framework available across a diverse range of modalities, including images, audio, video, and 3D shapes, for both the secret and cover data. By transforming the representation of all data to an implicit form, our framework takes advantage of the flexibility and high capacity inherent in INRs. The flexibility to all modalities enables cross-modal tasks that remained largely unexplored. Along with the versatility across the modalities, our framework is also capable of hiding multiple secret information into one cover data. In Sec. 4, we conduct the hiding of both an audio and an video into a single image, a task which was previously considered highly challenging due to the limited capacity of images. Our framework therefore further enables expansion and intriguing adaptation of steganography into a wider range of fields. In the medical field, for example, private information as an endoscopy video and an audio recording of the heart sound can be hidden under an image of a patient.

To further enhance the security, layer-wise permutation of the INR network using a private key is additionally executed. Therefore, the secret INRs' network parameters can be well distributed within the stego INR's network. Revealing

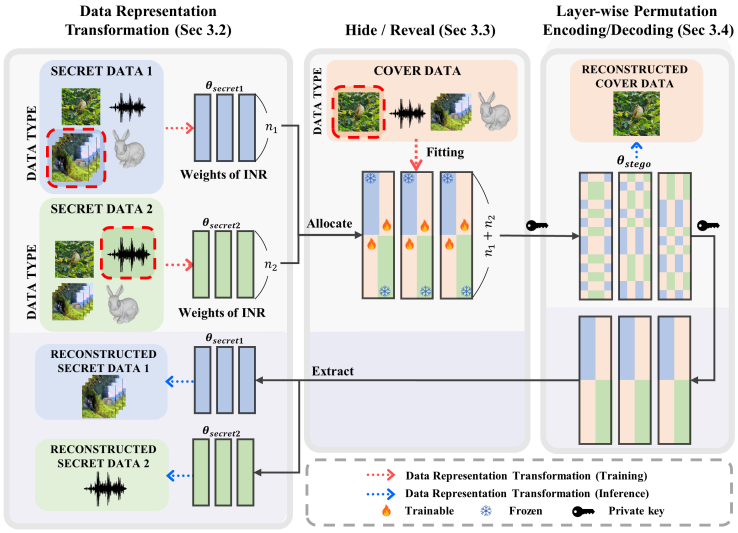

Fig. 1: A general framework of INRSteg for hiding two types of secret data. All four data types can be transformed into INRs, and the red box shows which of the four data types is selected in this case. After the representation transformation of each secret data, the weights of the two INRs are allocated in a new MLP network. With the weight freezing technique, the weights from the secret data are fixed while the rest are fitted on the cover data. The existence of the secret data is concealed by permutation encoding via private key, which does not affect the reconstruction performance. The private key is then used to decode the permuted stego network, and the secret data are revealed by separating each MLP network.

the secret data is executed by re-permuting and disassembling the weights of the stego INR. This procedure is not only simple but also guarantees a lossless revealing process as all neural representations of the secret data are fully recoverable.

Additionally, our framework does not require any deep learning models to be trained for the hiding and revealing process, such as GAN [16, 25] or diffusion models [27]. This excludes the need for any training data and therefore our framework can prevent any biases that may arise from using training datasets [12] and avoids domain shift issues, allowing datasets of all modalities to be applied without any additional pre-processing. Moreover, by eliminating the use of deep networks, our work substantially reduces both the computational and memory costs associated with training and deploying models which have become a significant global challenge.

The proposed framework, INRSteg, achieves flexible cross-modal steganography for multiple secret data by converting the data representation to neural networks. At the same time, we reduce computational and memory costs by

avoiding the use of deep neural networks. For intra-modal steganography, we demonstrate that INRSteg outperforms prior image steganography models such as DeepMIH [4] and DeepSteg [1]. We show INRSteg is robust in the real-world scenarios by applying quantization operation. These are our main contributions, and the details will be described in further sections:

- Introduce an Implicit Neural Representation (INR) steganography framework that flexibly adopts to a large range of modalities and further enables all cross-modal steganography.
- Demonstrate a neural network allocating procedure that enables cost-effective hiding and revealing of multiple secret data at once.
- Comprehensive investigation on distortion, security and capacity shows that INRSteg achieves state-of-the-art cross-modal steganography.
- Figure out that INRSteg is the most efficient compared to the previous steganography models and robust in real-world scenarios especially for quantization operation.

2 Related work

2.1 Steganography

Steganography is a technique that focuses on hiding confidential information by adeptly embedding data within other publicly accessible data, effectively concealing the very existence of the hidden message. When one is aware of this concealed data, it can be retrieved by a revealing process with an appropriate private key [17]. A prominent approach in historical methods is the Least Significant Bit (LSB) based method, which takes advantage of the imperceptibility of small changes to the least significant bits of the data [7,28]. With the rise of deep neural networks, the field of steganography has rapidly evolved to overcome the challenge of security and capacity. Deepsteg [1], which is an early work of image steganography, attempts to use an encoder and decoder to determine where to place the secret information. CRoSS [27] utilizes diffusion models to only reveal the secret information when the accurate prompt is given. To enhance the capacity aspect while ensuring the security, invertible neural networks (INNs) have been actively adopted. In the image steganography field, HiNet [8] incorporates INNs with wavelet domain to minimize distortion, and DeepMIH [4] progresses to hide multiple images into one. LF-VSN [14] further attempts to hide seven videos into one utilizing the INN.

Notwithstanding their innovations, a recurrent limitation across these previous methods is the trade-off between capacity and security. If the cover data is damaged in any way during the embedding process, it becomes susceptible to detection by steganalysis algorithms [23, 26]. This vulnerability poses a significant limitation, especially when attempting to hide large volumes of data such as videos or 3d shapes. INRSteg cannot be detected by traditional steganlalysis tools and therefore ensures security while maintaining high capacity.

2.2 Cross-modal steganography

The evolution of steganography has also expanded to the concealment of information across different data modalities, such as embedding audio data within videos, which is called cross-modal steganography. [24] embeds video information into audio signals by leveraging the human visual system's inability to recognize high-frequency flickering, allowing for the hidden video to be decoded by recording the sound from a speaker and processing it. [6] utilizes a joint deep neural network architecture comprising two sub-models, one for embedding the audio into an image and another for decoding the image to retrieve the original audio. These methods achieve cross-modal steganography using deep learning techniques, but are still restricted on specified modalities.

Recent advancements in Implicit Neural Representations (INRs) have paved the way for a new paradigm in steganography [5, 10]. INRs have been leveraged to represent multiple modalities of data simultaneously. StegaNeRF [10] hides data of various modalities at viewpoints when rendering NeRF. [5] conceals audio, video, and 3D shape data within a cover image, leveraging RGB values as weights of the INRs. While these recent works represent a paradigm shift in steganography techniques, the capacity and security trade-off still remains an issue and the modality of the cover data is set to be singular. Our proposed method in this paper seeks to address and overcome this trade-off while maintaining the advantages of INRs and further eliminating the modality constraint on the cover data, demonstrating a significant advancement in the steganography research.

2.3 Implicit neural representations

Implicit neural representations (INRs) is a rapidly emerging field where data is expressed in a continuous and differentiable manner through neural networks. This representation [18] is resolution-agnostic [19] and is capable of compressing the data effectively [2]. These advantages highlight various applications of INRs that show superior performance than that of the array-based representation. For example, neural radiance field (NeRF) [13] constructs a 5d scene representation of a 3d object through view-dependent synthesis. [3] proposes a general framework that feeds a modulated INR directly on the downstream models and shows competitive performance on multiple modalities. [9] employs meta-learning coupled with pruning techniques to efficiently scale implicit neural representations for large datasets.

3 Method

3.1 Framework

Our research focuses on INR steganography, where the data representations are transformed into Implicit Neural Representations (INRs), such as SIREN [18]. Figure 1 shows the general framework of our methodology, which is segmented into three sub-stages. In the *Data Representation Transformation* stage, secret

data are transformed into their corresponding Implicit Neural Representations, represented as θ_{secret}. Subsequently, in the *Hide/Reveal* stage, θ_{stego} is fitted to represent the cover data, x_{cover}, while preserving the parameters of θ_{secret}. The *Layer-wise Permutation Encoding/Decoding* stage introduces a layer-wise permutation mechanism, designed to avoid detection of the secret data's presence and location.

3.2 Data Representation Transformation

In the first stage, we transform the secret data into Implicit Neural Representations (INRs) by fitting multi-layer perceptrons (MLP) with sine activation functions for each data. This network was first introduced by [18] as SIREN (Sinusoidal representations for neural networks) and is widely used in the INR research field. The structure of the MLP network has n hidden layers of size D and the output is represented as follows:

$$\begin{aligned} \mathbf{y} = \mathbf{W}^{(n)} \left(g_{n-1} \circ \cdots \circ g_1 \circ g_0\right)(\mathbf{x}_0) + \mathbf{b}^{(n)}, \\ \text{where } \mathbf{x}_{i+1} = g_i\left(\mathbf{x}_i\right) = \sigma\left(\mathbf{W}^{(i)}\mathbf{x}_i + \mathbf{b}^{(i)}\right), \end{aligned} \tag{1}$$

where $\circ$ is a function composition of fully-connected layers. In the equation, $\mathbf{x_i}$ is the input of the i^{th} layer for $i \in \{0, 1, 2, ..., n\}$, $\mathbf{y} \in \mathbb{R}^O$ is the output value corresponding to $\mathbf{x_0} \in \mathbb{R}^I$, and $g_i \colon \mathbb{R}^D \to \mathbb{R}^D$. For our experiments, input and output dimensions (I, O) are $(1, 1)$, $(2, 3)$, $(3, 1)$, $(3, 3)$ for audio, image, 3D shapes, and video, respectively. If the input or output dimension of the secret INRs exceeds that of the cover INR, our method employs a strategic reduction in the number of hidden layers within the secret INR. Additionally, as the dimension of the hidden layer D is fixed for each cover modality, the secret INRs' hidden layer dimensions are set accordingly.

In general, given a set S of inputs $\mathbf{x}\in\mathbb{R}^I$ and the corresponding outputs $\mathbf{y}\in\mathbb{R}^O$, the loss function is given by

$$\mathcal{L}_{recon}(\theta) = \sum_{(\mathbf{x},\mathbf{y})\in\mathcal{S}} \left\| f_\theta\left(\mathbf{x}\right) - \mathbf{y} \right\|_2^2 \tag{2}$$

, where f_θ is the neural representation of the data. For example, given a 2D image, the set $\mathcal{S}$ consists of 2D coordinates $\mathbf{x}\in\mathbb{R}^2$ and the corresponding RGB values $\mathbf{y}\in\mathbb{R}^3$. INRs enable a variety of data types to be expressed in this unified network architecture. This benefits the overall security paradigm of our steganography framework by hiding the underlying data types.

3.3 Hide and Reveal

In this section, we explain our main idea associated with the hiding process. The secret INRs, θ_{secret}, undergo a process of allocation and fitting to the cover data, x_{cover}. Within the weight space, only the diagonal blocks of weight matrices are occupied by the parameters of θ_{secret}, so that the non-diagonal parameters can be fitted to x_{cover}, while ensuring the diagonal block's weight values to remain unaltered.

First, we explain when the number of secret data, N, is two without losing generality for hiding multiple data. The number of input nodes, output nodes, hidden layers, and the dimension of the hidden layers of $\theta_{secret1}$ and $\theta_{secret2}$ are each set to (I_1, O_1, n_1, D_1), (I_2, O_2, n_2, D_2), respectively. And (I, O, n, D) for the stego INR network, θ_{stego}, where $D_1 + D_2$ is set to equal D. Now, we substitute the parameters of the initialized θ_{stego} by concatenating $\theta_{secret1}$ and $\theta_{secret2}$ as shown in Figure 2. Note that the initialized parameters of θ_{stego} are omitted for clarity. After the substitution, the parameter of θ_{stego} is as follows:

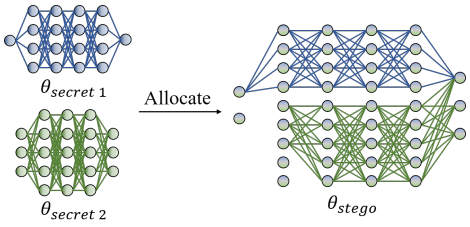

Fig. 2: An example of allocating two secret INRs, $\theta_{secret1}$ and $\theta_{secret2}$, into the stego INR, θ_{stego}. For $\theta_{secret1}$, $\theta_{secret2}$, and θ_{stego}, $(I_1 = 1,\ O_1 = 1,\ n_1 = 4,\ D_1 = 4)$, $(I_2 = 3,\ O_2 = 3,\ n_2 = 3,\ D_2 = 5)$, and $(I = 2,\ O = 3,\ n = 4,\ D = 9)$, respectively.

$$\mathbf{W}^{(i)} = \begin{pmatrix} \mathbf{W}_1^{(i)} & \mathbf{W}_{01}^{(i)} \\ \mathbf{W}_{02}^{(i)} & \mathbf{W}_2^{(i)} \end{pmatrix}, \mathbf{b}^{(i)} = \begin{pmatrix} \mathbf{b}_1^{(i)} \\ \mathbf{b}_2^{(i)} \end{pmatrix}. \tag{3}$$

Here, $\mathbf{W}^{(i)}$, $\mathbf{W}_1^{(i)}$, $\mathbf{W}_2^{(i)}$ and $\mathbf{b}^{(i)}$, $\mathbf{b}_1^{(i)}$, $\mathbf{b}_2^{(i)}$ are the i^{th} layer's weight matrices and bias vectors of θ_{stego}, $\theta_{secret1}$ and $\theta_{secret2}$, where $i \in \{0, ..., n\}$. If $\theta_{secret1}$ or $\theta_{secret2}$ have fewer hidden layers compared to θ_{stego}, the parameters of θ_{stego} are not substituted. $\mathbf{W}_{01}^{(i)}$ and $\mathbf{W}_{02}^{(i)}$ are the remaining parameters of θ_{stego}.

Now, θ_{stego} is fitted to x_{cover} through a selective parameter update using a binary mask, so that the parameters of $\theta_{secret1}$ and $\theta_{secret2}$ are preserved while the remaining weights get updated. After successfully updating the weight space, the reconstructed discrete representation depicts x_{cover} and all secret data can be retrieved from θ_{stego}.

For $N > 2$, the hiding process is expanded in the same manner, so we further explain when $N = 1$. When there is only one secret data, the parameters of $\theta_{secret1}$ can be freely allocated in θ_{stego}. Suppose that $\theta_{secret1}$ is allocated in the middle of θ_{stego}, then the weight space of θ_{stego} is as follows:

$$\mathbf{W}^{(i)} = \begin{pmatrix} \mathbf{W}_{01}^{(i)} & \mathbf{W}_{02}^{(i)} & \mathbf{W}_{03}^{(i)} \\ \mathbf{W}_{04}^{(i)} & \mathbf{W}_{1}^{(i)} & \mathbf{W}_{05}^{(i)} \\ \mathbf{W}_{06}^{(i)} & \mathbf{W}_{07}^{(i)} & \mathbf{W}_{08}^{(i)} \end{pmatrix}, \mathbf{b}^{(i)} = \begin{pmatrix} \mathbf{b}_{01}^{(i)} \\ \mathbf{b}_{1}^{(i)} \\ \mathbf{b}_{02}^{(i)} \end{pmatrix}, \tag{4}$$

where $\mathbf{W}^{(i)}$, $\mathbf{W}_1^{(i)}$ and $\mathbf{b}^{(i)}$, $\mathbf{b}_1^{(i)}$ are the i^{th} layer's weight matrices and bias vectors of θ_{stego} and $\theta_{secret1}$, for $i \in \{0, ..., n\}$. $\mathbf{W}_{01}^{(i)}, \mathbf{W}_{02}^{(i)}, \cdots, \mathbf{W}_{08}^{(i)}, \mathbf{b}_{01}^{(i)}$, and $\mathbf{b}_{02}^{(i)}$ are the remaining parameters of θ_{stego}. This network then goes through the same processes as when there are multiple secret data.

The inherent flexibility of our approach is highlighted by its capability to insert diverse data types, in varying amounts, at arbitrary desired positions. This versatility addresses and mitigates capacity constraints.

3.4 Layer-wise Permutation Encoding and Decoding

To enhance security of our method, we strategically permute the nodes within each layer of the stego network. This layer-wise permutation allows the parameters of the secret INRs to be distributed throughout θ_{stego}, making the parameters indistinguishable and thus undetectable. In this section, we demonstrate how the permutation is executed via one 128-bit private key ρ, utilizing a cryptographic key derivation function (KDF), which derives one or more secret keys from a master key. First, the private key is randomly selected and is employed to generate n secret keys for the n hidden layers of θ_{stego}. Subsequently, the nodes of each layer are permuted according to the secret key assigned to each layer. The total number of possible permutation variations per INR is $(\mathbf{D}!)^n$, where D is the dimension of the hidden layer. This shows that the security implications of this permutation intensify, as the possible combinations of permutation grow exponentially with the number of hidden layers in the θ_{stego}.

It is crucial to emphasize that the overall functionality of the neural network remains unchanged throughout this operation. This operation leverages the inherent property of permutation invariance exhibited by Multi-Layer Perceptrons (MLPs) within a single layer. As INRs are also MLP networks, all permuted networks are identical to the original network. Therefore, θ_{stego} is invariant to the layer-wise permutation, whereas extracting the secret INR networks becomes impossible without the private key.

4 Experiment

We explore the performance of INRSteg focusing on its resilience to distortion, capacity for information hiding, and security against detection. Distortion measures both the indistinguishability between the cover data and the reconstructed data from stego INR, and the performance drop during the hiding and revealing stage of the secret data. Capacity refers to the amount of hidden information that can be privately embedded within the cover data. Security indicates the extent to which the secret data can avoid detection when subjected to steganalysis.

To thoroughly evaluate the distortion, we employ several metrics tailored to different modalities, which are organized in Appendix A, as well as visualization. In terms of capacity, we present various cross-modal steganography experiments and compare the model size with other previous steganography models. For security, we compute the detection accuracy using image steganalysis models, SiaStegNet [26] and XuNet [22], and visualize weight distribution to show the effects of permutation operation.

In our groundbreaking research, we highlight three pivotal aspects of our experimental evaluation of INRSteg, a novel steganography method that sets new benchmarks in data hiding techniques:

- **Cross-Modal Steganography** We initiated pioneering experiments in cross-modal steganography, demonstrating our method's unique capability to privately embed audio and video data within images. Comprehensive ablation

study demonstrates that INRSteg is generally applicable across all data modalities achieving high-quality concealment(Sec. 4.1).

- **Intra-Modal Steganography** We conduct experiments of hiding data within the same modality, specifically embedding images within images. This approach allowed us to conduct a comparative analysis of distortion and security metrics against those reported in existing literature, showcasing the advanced capabilities of INRSteg. For ablation study, we conduct multi-image steganography experiments varying the amount of secret data(Sec. 4.2).
- **Efficiency and Robustness Analysis** To show the efficiency of INRSteg, we perform computational cost comparisons with five existing steganography models. We test the robustness of INRSteg by simulating a challenging scenario where network weights undergo quantization. It demonstrates that our method maintains high performance and robustness even under harsh conditions, which ensures secure and reliable steganography across a wide range of applications(Sec. 4.3).

Table 1: Results of Cross-modal steganography experiment. VoxCeleb2 and ImageNet dataset are used for secret data and cover data, respectively. "Secret/Revealed" indicates the recovery performance of secret data. "Cover/Stego" indicates the reconstruction performance from stego INR.

Secret/Revealed				Cover/Stego	
VoxCeleb2				ImageNet	
Audio		Video		Image	
SNR ↑	MAE ↓	PSNR ↑	APD ↓	PSNR ↑	RMSE ↓
37.558(±5.4)	0.314(±0.3)	41.448(±3.1)	1.349(±0.5)	39.258(±4.4)	2.948(±1.5)

4.1 Cross-modal Steganography

ImageNet and VoxCeleb2 For cross-modal steganography experiment, we conduct a pioneering steganography experiment that embeds the VoxCeleb2 dataset within images from ImageNet dataset. VoxCeleb2 dataset has both video and audio files, thereby enabling simultaneous playback of visual and auditory data. Using our INRSteg method, we leverage images from ImageNet dataset as cover data to hide video and audio data at the same time. Tab. 1 shows the distortion performance for secret/revealed and cover/stego performance demonstrating the efficacy of our approach in preserving the content of cover and the embedded data. For the secret and the revealed secret data pair, as our framework goes through lossless recovery in terms of INR, the secret/revealed secret performance is solely dependent on the data representation transformation phase. Moreover, Figure 3 shows visualization of the reconstructed data, recovery of video and audio secret data. There is no noticeable quality degradation,

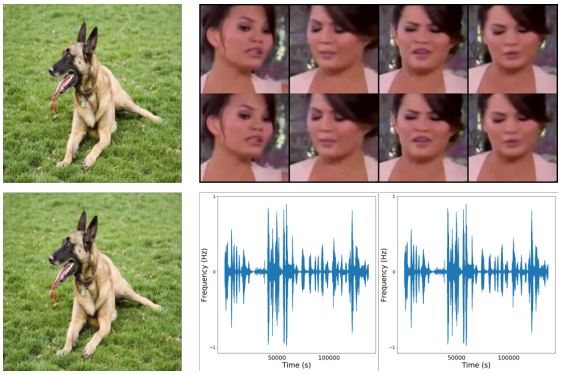

Fig. 3: Visualization results of cross-modal steganography. Can you identify which image is the cover and the reconstructed cover data? For the video and audio examples, can you determine which are the secret and revealed secret data? For images, upper one is cover data and below is reconstructed cover data. For video, upper row is secret data and below is revealed secret data. For audio, the left is secret data and the right is revealed secret data.

which implies successful results of cross-modal steganography. Notably, cross-modal steganography is unprecedented, representing a significant improvement in the field of steganography. More reconstruction results using other modalities can be seen in Appendix C, showing successful multi-data steganography with almost no visible difference for both 3D shape and video. Additionally, there are visualization results of weight space after steganography and comparison results of weight distribution after permutation operation in Appendix D, demonstrating the security of INRSteg.

Ablation study For ablation study, we try to demonstrate that INRSteg is generally applicable across all data modalities for steganography. Therefore, we design experiments that cover all combinations of data modalities where cover data type is image, audio, or video and secret data type is image, audio, 3D shape, or video. The distortion performance of all single-data cross-modal steganography tasks can be seen in Tab. 2, demonstrating achievement of exceptional performance for all cases. For additional reconstruction results, we show them in Appendix C. Recovery performance results for secret data are in Appendix E. A deeper analysis of the results presented in Tab. 2 reveals two key factors that influence improved cover/stego performance: the modality of the secret and cover data, and the capacity of the neural network utilized for the stego INR. We observe that when the modalities of the secret and cover data match, the stego INR demonstrates enhanced accuracy in fitting to the cover data. This phenomenon is attributed to the intrinsic properties of INRs, where the weight spaces of neural networks exhibit similarities when representing with same modality data [21]. Moreover, leveraging the capabilities of INRs, employing a larger neural network for the cover data fitting process enhances the quality of cover data reconstruc-

tion. This can be shown in Tab. 2, where the performance of hiding 3D shapes in images outperforms hiding images in images. This property allows flexible control of the network size for secret and cover data based on the required recovery quality and resource space constraints in the real world. An additional ablation study about the effectiveness of the padding ratio is proposed in Appendix B.

Table 2: Cover/Stego performance of single-data steganography. The best results are **highlighted** for each cover modality.

	Cover type					
	Image		Audio		Video	
Secret type	PSNR ↑	RMSE ↓	SNR ↑	MAE ↓	PSNR ↑	APD ↓
Image	62.34	0.204	32.63	0.599	35.26	3.315
Audio	63.16	0.179	**41.42**	**0.268**	34.69	3.574
Video	56.18	0.408	27.20	0.931	**41.26**	**1.532**
3D shape	**92.95**	**0.000**	23.12	1.403	38.54	2.244

4.2 Intra-modal Steganography

Image steganography In this section, we compare INRSteg to existing image-to-image steganography methods to show that our framework also excels in intra-modal tasks. Tab. 3 compares the performance of INRSteg with other deep learning based image-to-image steganography methods, DeepMIH [4] and an improved DeepSteg [1]. INRSteg outperforms on all metrics showing that our framework achieves state-of-the-art performance. For hiding two secret images into one image, the results in Appendix F, and Figure 4 further visualizes the cover/stego and secret/revealed secret image pairs for two images into one steganography. We continue to compare our results with DeepMIH using difference maps each enhanced by 10, 20, and 30 times. For INRSteg, nearly no visual differences exist between the original and the revealed images even when the difference map is enhanced 30 times. For DeepMIH, on the other hand, we can notice obvious differences for all difference maps when comparing cover/stego and secret1,2/revealed secret 1,2 pairs.

Image steganalysis To examine security, we adopt two image steganalysis tools, SiaStegNet [26] and XuNet [22]. Tab. 4 presents the detection accuracy of our framework and other models. The security is higher as the detection accuracy is closer to 50%. Unlike other steganography models, the detection accuracy of INRSteg is 50%, meaning that the steganalysis tool completely fails to distinguish the stego data from the cover data. This shows that our framework is undetectable using existing steganalysis tools and assures perfect security. This is because our framework does not directly edit the data in the discrete

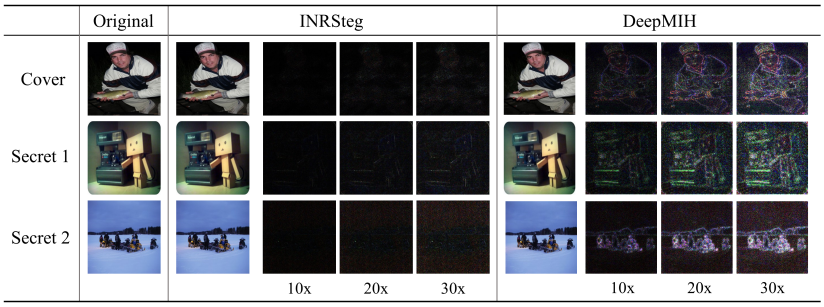

Fig. 4: A comparison of difference maps (enhanced by 10x, 20x, and 30x) between the original images and the revealed images. Columns 2-6 are results of INRSteg and columns 7-10 are results of DeepMIH [4].

representation state, but performs manipulation after transforming the representation into a neural network, which is unnoticeable when re-transformed to the discrete representation. Moreover, we conduct steganalysis analysis on multi-images steganography. The results in Appendix F.

Multi-image steganography In Tab. 5, we conduct experiments for multi-image steganography hiding more than two images when the size of stego INR is fixed, so that one may have to decrease the network size of each hidden data in order to fit the memory capacity. The results show that high recovery performance maintains even when increasing the number of hidden data. For more experiments with other modalities, please refer to Appendix G.

Table 3: Performance comparison of image-to-image steganography.

	Metric	MAE ↓	PSNR ↑	SSIM ↑	RMSE ↓
DeepSteg	cover	2.576	37.48	0.947	3.417
	secret	3.570	33.34	0.957	3.57
DeepMIH	cover	2.839	36.05	0.932	4.276
	secret	3.604	32.97	0.9550	5.746
INRSteg	cover	**0.153**	**62.34**	**0.999**	**0.204**
	secret	**1.084**	**45.84**	**0.988**	**1.326**

Table 4: Steganalysis comparison of image-to-image steganography.

	Accuracy (%)	
	SiaStegNet	XuNet
DeepSteg	92.87	75.83
DeepMIH	90.70	90.82
INRSteg	**50.00**	**50.00**

4.3 Efficiency and Robustness Analysis

Model Size Comparison In our efficiency analysis, we compare model sizes to evaluate the computational cost and efficiency of our proposed INRSteg relative to existing steganography models. As described in Tab. 6, our INRSteg achieves a remarkably compact architecture, consisting of only 400,000 param-

eters. Compared to other existing steganography models, this advantage is notable. DeepSteg [1], with 42.58 million parameters, is approximately 107.46 times larger than our model. DeepMIH [4] model, with 12.42 million parameters, has 31.34 times more parameters, and LF-VSN [14], at 7.40 million parameters, is 18.68 times larger. Especially, CRoSS [27] contains 1,066.24 million parameters, corresponding to 2690.54 times the size of the model. This comparative analysis highlights the efficiency of our INRSteg, demonstrating its capability to achieve robust steganography performance with a significantly reduced computational space. The small model not only promotes faster training and inference times, but also opens up opportunities for deployment on devices with limited computational resources such as smartphones.

Table 5: Distortion performance of hiding multiple images (> 2) into one.

	Metric	PSNR ↑	RMSE ↓
3 images	cover	62.31	0.204
	secret 1,2,3	43.67	1.642
4 images	cover	62.58	0.153
	secret 1,2,3,4	41.86	2.059

Table 6: Model Size Comparison.

Model	#Params	Ratio to Ours
DeepSteg [1]	42.58M	107.46×
HiNet [8]	4.05M	10.22×
DeepMIH [4]	12.42M	31.34×
CRoSS [27]	1,066.24M	2690.54×
LF-VSN [14]	7.40M	18.68×
Ours	0.40M	1×

Robustness We explore the robustness of INRSteg under harsh conditions in real-world scenarios. In particular, we apply quantization operation to store model parameters from float32 to int8 format. This process aims to simulate scenarios where data precision must be reduced due to storage or transmission constraints. After quantization, we recover the parameters from int8 to float32 and evaluate the performance metrics of the recovered secret images. Tab. 7 describes the results of performance comparison before and after quantization across three cover types: image, audio, and video. Notably, while PSNR values decrease after quantization, this decrease is not as significant as might be expected given the harsh nature of the quantization process. This indicates that INRSteg has a robust ability to maintain secret data quality despite the data compression and precision reductions. Furthermore, something to crucially note in Tab. 7 is that the SSIM value hardly changes after quantization. Since SSIM measures the perceptual quality, it is a more representative metric that evaluates image quality. Therefore, the results indicate that the visual content of the secret images remain almost completely. The reconstruction results are in Appendix H. Additionally, we show preserving secret data when using another model compression method, pruning, in Appendix I.

5 Limitations and Discussion

We would like to highlight the potential negative impacts that steganography frameworks may result in the ethical aspect. Users must be cautious not to utilize

Table 7: Performance Metrics Before and After Quantization for Different Cover Types. Pre-Q denotes metrics before quantization, Post-Q denotes metrics after quantization. Higher values are better for PSNR and SSIM, while lower values are preferred for RMSE and MAE.

	PSNR ↑		RMSE ↓		SSIM ↑		MAE ↓	
Cover type	Pre-Q	Post-Q	Pre-Q	Post-Q	Pre-Q	Post-Q	Pre-Q	Post-Q
Image	40.0766	33.7483	2.5245	5.2375	0.9821	0.9554	1.7085	3.519
Audio	40.0766	33.5438	2.5245	5.3623	0.9821	0.9529	1.7085	3.723
Video	40.0766	33.8276	2.5245	5.1899	0.9821	0.9583	1.7085	3.417

our work in unethical situations, where one may hide inappropriate information. Additional limitations are described in Appendix J.

6 Conclusion

In conclusion, this paper introduces INRSteg, an innovative framework designed for the concealment of multiple cross-modal data through the utilization of the weight space inherent in Implicit Neural Representations (INR). Our extensive experimentation demonstrates that INRSteg surpasses existing steganography techniques in key areas, namely distortion evaluation, data capacity, and security measures, across both intra and cross-modal steganography scenarios. A notable advantage of INRSteg is its deployment of smaller model sizes, which directly contributes to reduced energy consumption, making it exceptionally suited for practical real-world applications. Furthermore, our analysis confirms the robustness of INRSteg, even when subjected to the rigorous demands of quantization operations. We believe that INRSteg establishes a new benchmark for achieving an optimal balance between efficiency and performance within the steganography domain, paving the way for future advancements in secure and efficient steganography techniques.

Acknowledgements

This work was partly supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.RS-2022-II220184, Development and Study of AI Technologies to Inexpensively Conform to Evolving Policy on Ethics) and Institute for Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No. 2021-0-01381, Development of Causal AI through Video Understanding and Reinforcement Learning, and Its Applications to Real Environments).

References

1. Baluja, S.: Hiding images in plain sight: Deep steganography. In: Advances in Neural Information Processing Systems (2017)
2. Dupont, E., Goliński, A., Alizadeh, M., Teh, Y.W., Doucet, A.: Coin: Compression with implicit neural representations. arXiv preprint arXiv:2103.03123 (2021)
3. Dupont, E., Kim, H., Eslami, S., Rezende, D., Rosenbaum, D.: From data to functa: Your data point is a function and you can treat it like one. arXiv preprint arXiv:2201.12204 (2022)
4. Guan, Z., Jing, J., Deng, X., Xu, M., Jiang, L., Zhang, Z., Li, Y.: Deepmih: Deep invertible network for multiple image hiding. IEEE Transactions on Pattern Analysis and Machine Intelligence (2023)
5. Han, G., Lee, D.J., Hur, J., Choi, J., Kim, J.: Deep cross-modal steganography using neural representations. In: 2023 IEEE International Conference on Image Processing (ICIP) (2023)
6. Huu, Q.P., Dinh, T.H., Tran, N.N., Van, T.P., Minh, T.T.: Deep neural networks based invisible steganography for audio-into-image algorithm. In: 2019 IEEE 8th Global Conference on Consumer Electronics (GCCE) (2019)
7. Jain, Y.K., Ahirwal, R.: A novel image steganography method with adaptive number of least significant bits modification based on private stego-keys. In: International Journal of Computer Science and Security (IJCSS) (2010)
8. Jing, J., Deng, X., Xu, M., Wang, J., Guan, Z.: Hinet: Deep image hiding by invertible network. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2021)
9. Lee, J., Tack, J., Lee, N., Shin, J.: Meta-learning sparse implicit neural representations. Advances in Neural Information Processing Systems (2021)
10. Li, C., Feng, B.Y., Fan, Z., Pan, P., Wang, Z.: Steganerf: Embedding invisible information within neural radiance fields. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (2023)
11. Liu, Y., Liu, S., Wang, Y., Zhao, H., Liu, S.: Video steganography: A review. Neurocomputing (2019)
12. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., Galstyan, A.: A survey on bias and fairness in machine learning. ACM Comput. Surv. (2021)
13. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM (2021)
14. Mou, C., Xu, Y., Song, J., Zhao, C., Ghanem, B., Zhang, J.: Large-capacity and flexible video steganography via invertible neural network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2023)
15. Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S.: Deepsdf: Learning continuous signed distance functions for shape representation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (2019)
16. Shi, H., Dong, J., Wang, W., Qian, Y., Zhang, X.: Ssgan: Secure steganography based on generative adversarial networks. In: Advances in Multimedia Information Processing – PCM 2017 (2018)
17. Simmons, G.J.: The subliminal channel and digital signatures. In: Advances in Cryptology (1985)
18. Sitzmann, V., Martel, J., Bergman, A., Lindell, D., Wetzstein, G.: Implicit neural representations with periodic activation functions. Advances in neural information processing systems (2020)

19. Strümpler, Y., Postels, J., Yang, R., Gool, L.V., Tombari, F.: Implicit neural representations for image compression. In: European Conference on Computer Vision (2022)
20. Subramanian, N., Elharrouss, O., Al-Maadeed, S., Bouridane, A.: Image steganography: A review of the recent advances. IEEE Access (2021)
21. Tancik, M., Mildenhall, B., Wang, T., Schmidt, D., Srinivasan, P.P., Barron, J.T., Ng, R.: Learned initializations for optimizing coordinate-based neural representations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021)
22. Xu, G.: Deep convolutional neural network to detect j-uniward. In: Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security (2017)
23. Xu, G., Wu, H.Z., Shi, Y.Q.: Structural design of convolutional neural networks for steganalysis. IEEE Signal Processing Letters (2016)
24. Yang, H., Ouyang, H., Koltun, V., Chen, Q.: Hiding video in audio via reversible generative models. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV) (2019)
25. Yang, J., Liu, K., Kang, X., Wong, E.K., Shi, Y.Q.: Spatial image steganography based on generative adversarial network. arXiv preprint arXiv:1804.07939 (2018)
26. You, W., Zhang, H., Zhao, X.: A siamese cnn for image steganalysis. IEEE Transactions on Information Forensics and Security (2021)
27. Yu, J., Zhang, X., Xu, Y., Zhang, J.: Cross: Diffusion model makes controllable, robust and secure image steganography. Advances in Neural Information Processing Systems (2024)
28. Zhang, H., Geng, G., Xiong, C.: Image steganography using pixel-value differencing. In: 2009 Second International Symposium on Electronic Commerce and Security (2009)
29. Zhou, H., Zhang, W., Chen, K., Li, W., Yu, N.: Three-dimensional mesh steganography and steganalysis: A review. IEEE Transactions on Visualization and Computer Graphics (2022)
30. Zhu, J., Kaplan, R., Johnson, J., Fei-Fei, L.: Hidden: Hiding data with deep networks. In: Proceedings of the European conference on computer vision (ECCV) (2018)