Machine Learning for Human Intelligence
2025, Vol. 3, Issue 2

Research Article

Diffusion-Based Generative Coverless Steganography for Robust Face Recognition

Faheem Mazhar[1,*] and Haroon Ahmad[2]

[1]Computer Science Department, NFC Institute of Engineering and Technology, Multan, Pakistan; Email: faheemmazharbalouch@gmail.com
[2]Department of Computer Science, Air University, Islamabad, Pakistan; Email: haroon@aumc.edu.pk
[*]Corresponding author: Faheem Mazhar (faheemmazharbalouch@gmail.com)

Article History

Academic Editor:
Dr. Nabeel Asghar
Submitted: May 21, 2025
Revised: July 26, 2025
Accepted: September 01, 2025

Keywords:
Coverless image steganography (CIS), Contrastive Image Synthesis, Joint Source-Channel Coding (JSCC), peak signal-to-noise ratio (PSNR), Diffusion Model.

Abstract

Conventional image steganography centers on embedding one image within another to evade detection by unauthorized parties. Coverless image steganography (CIS) improves imperceptibility by omitting the use of a cover image. Recent studies have employed text prompts as keys in Contrastive Image Synthesis via diffusion models. The swift advancement of generative models has initiated a novel approach in steganography known as generative steganography (GS). It facilitates message-to-picture creation without requiring a carrier image. This internationally recognized biometric facial recognition technique is extensively utilized in numerous identity verification systems. This research offers a novel coverless steganography framework for face recognition photos based on a diffusion model, aimed at enhancing personal privacy protection and ensuring the secure transmission and sharing of sensitive information without compromising user experience. We propose a Coverless Semantic Steganography Communication system utilizing a Generative Diffusion Model to conceal hidden images within generated stego images. The semantically associated private and public keys allow the legitimate receiver to accurately decode hidden images, while the eavesdropper, lacking the entire and accurate key pairs, is unable to access them. Simulation outcomes illustrate the efficacy of the plug-and-play architecture across several Joint Source-Channel Coding (JSCC) frameworks. The comparative results under various eavesdropping risks indicate that, at a Signal-to-Noise Ratio (SNR) of 2.03 dB, the peak signal-to-noise ratio (PSNR) for the legitimate receiver exceeds that of the eavesdropper by 4.14 dB.

1 Introduction

Steganography is an extensively researched subject that seeks to conceal messages such as sounds, images, and text within a single container image in an imperceptible manner. In its reverse operation, only receivers equipped with an accurate revealing network can reconstruct secret information from the container, which visually resembles the host. In image steganography, conventional techniques frequently employ adaptive encoding based on distortion costs formulated by humans or neural networks,

necessitating established norms and expertise. Steganography, a significant aspect of information concealment, concentrates on embedding sensitive data inside the redundant areas of digital covers to provide unnoticeable transfer and secure storage [1].

In addition to cryptography-based encryption methods, researchers investigate covert communications for SemCom, which seek to protect against eavesdroppers by obscuring the communication patterns between lawful transmitters and receivers. The authors integrated multi-agent reinforcement learning to facilitate collaboration among devices and jammers in identifying susceptible eavesdroppers, consequently formulating solutions that collectively optimize semantic information transfer and power regulation [2]. The authors developed a covert SemCom framework for Unmanned Aerial Vehicle (UAV) scenarios by the combined optimization of flight trajectory and transmission power. [3] presented a covert SemCom system that accommodates several modalities, including text, pictures, and audio. It utilizes a power control mechanism that guarantees the efficacy of clandestine communication while also attaining superior semantic decoding quality. The internet revolution has greatly enhanced communication while also presenting issues in protecting messages transferred online. Steganography is a widely utilized method for concealing information within a container in an inconspicuous manner. Consequently, only authorized recipients can extract the information from the steganographic material [4].

Image steganography, a subset of this discipline, focuses on concealing hidden messages within images, providing a significant level of security and privacy. It is applicable in various domains, such as picture compression, secure communication, and cloud computing. Conventional cover-based picture steganography techniques conceal the hidden message within a cover image by modifying its statistical characteristics. Upon the disclosure of the cover image, the concealed message can be readily identified by steganalysis [5]. Conversely, coverless image steganography (CIS) seeks to encode or map the confidential message directly into a stego image, rather than altering a cover image. Consequently, it exhibits superior imperceptibility relative to cover-based approaches. Nonetheless, three obstacles emerge when endeavoring to execute generative steganography utilizing the diffusion model [6].

Diffusion models fail to generate high-quality images when Gaussian noise is replaced with confidential data in the spatial domain, resulting in a disruption of the Gaussian noise distribution. The aggregation of slight inaccuracies in the spatial domain across the forward and backward processes of the diffusion model leads to a decrease in the extraction precision of confidential data [8]. To maintain the input distribution of generative models, many generative steganography techniques employ reject sampling to transform secret data into a Gaussian distribution, thereafter generating stego pictures with pre-trained generative models [9]. This facilitates secure communication between the sender and recipient by solely disseminating the StegoDiffusion model and the concealing technique. The principal contributions of our GSD plan are as follows:

- We believe we are the pioneers in exploring the problem of generative steganography utilizing DDIM (GSD).
- We suggest concealing secret messages within the frequency domain of Gaussian noise, therefore mitigating the effects of cumulative errors in the time domain on the concealed data.
- We introduce a diffusion model specifically tailored for steganography, termed StegoDiffusion. It facilitates a bidirectional linkage between stego pictures and stego latents.
- In real applications, GSD demonstrates advantages over current approaches, particularly in attaining enhanced extraction accuracy with equivalent payloads.

2 Related Work

Steganography Based on Cover Objects. In steganography, images are frequently favored over text as carriers due to their capacity to convey substantial amounts of information. Recent years have witnessed substantial progress in the domain of image steganography. Wang et al. [10] developed two approaches to update the embedding cost of quantized DCT coefficients, thereby improving JPEG

steganography by examining the similarity of natural image content. In contrast, Yin et al. [11] have introduced a separable fine-tuning network design that is resilient to rounding operations and can significantly mitigate the deterioration of image quality while diminishing the decline in steganalysis efficiency. Li et al. [12] introduced a steganography technique utilizing an artificial immune system, enhancing security and ensuring precise retrieval of concealed data. A prevalent drawback of these steganography techniques is the necessity to perpetually counteract steganalysis programs. The emergence of coverless steganography presented a more secure resolution to this issue.

Traditional picture steganography entails concealing hidden information by altering the cover image. Data embedding can be categorized into spatial domain embedding and frequency domain embedding techniques, based on the data domain. Spatial domain embedding steganography was the initial method devised, characterized by the direct modification of pixel values in the carrier image to conceal a secret message. Common spatial domain embedding steganographic algorithms encompass the Least Significant Bit (LSB) steganography algorithm [13], the Least Significant Bit Match (LSBM) steganography algorithm [14], and the random modulation steganography algorithm. Steganographic techniques that employ frequency domain embedding frequently utilize JPEG images as carriers. JPEG compression utilizes the Discrete Cosine Transform (DCT) method to incorporate a concealed message into the DCT coefficients. The initial algorithm, JSteg, substitutes the least significant bit in DCT coefficients. F5 [15] seeks uniform distribution, while nsF5 mitigates histogram-based detection. Outguess utilizes a dual-round approach to synchronize DCT coefficient histograms between cover and stego pictures. Nevertheless, altering the cover image ultimately results in distortion, which can be readily identified by steganalysis.

The diffusion model, first introduced by Sohl-Dickstein et al. in 2015, has gained significant popularity due to its strong generative powers in areas including picture production, restoration, and translation. Its adaptability enhances and modifies digital photos proficiently. Nonetheless, the model's principal limitation is the extended duration of training and inference, necessitating exploration of optimization methodologies. The latent diffusion model (LDM) improves efficiency by facilitating high-resolution synthesis for various conditional inputs, such as text or bounding boxes [16]. The text inversion technique enhances model controllability by recreating user-defined ideas from a limited number of images [17]. Methods such as masked picture editing and DreamBooth facilitate individualized content generation, whereas cue-based image editing systems permit prompt-driven image alteration. The research of Huang et al. investigates sophisticated techniques for regulating diffusion models [18]. The prospects for diffusion-based, coverless image steganography appear favorable owing to the model's swift advancement and robust generative capabilities.

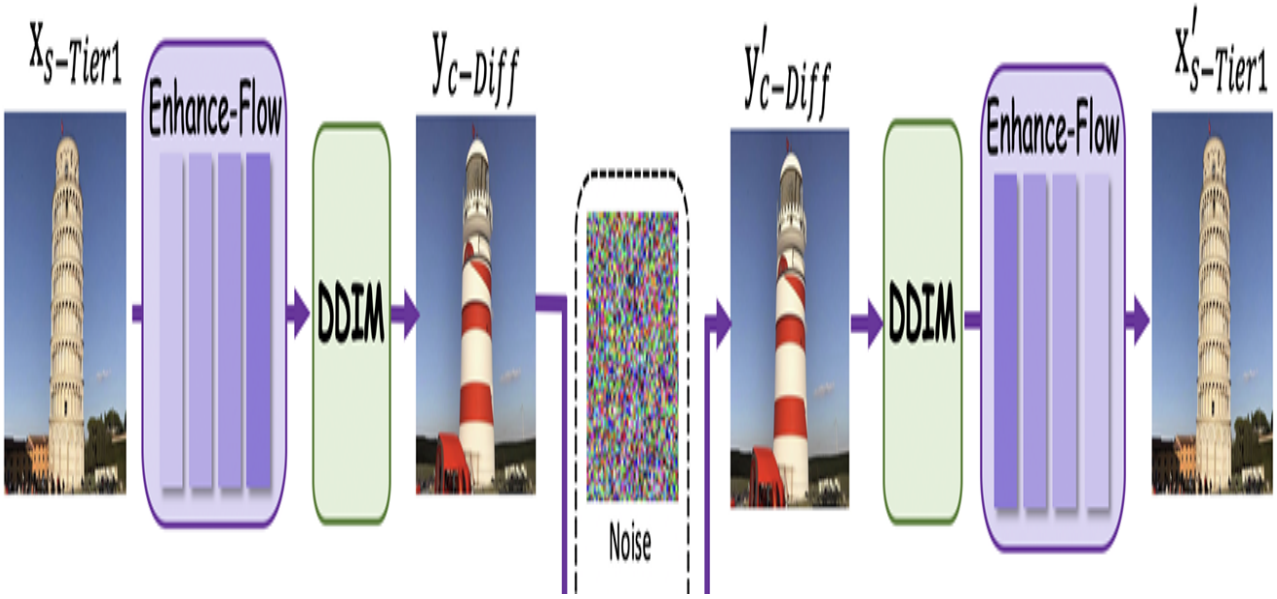

Figure 1: Comparative analysis of the local, optimal, and suggested models.

3 Proposed Methodology

Two fundamental processing mechanisms: the concealment process and the disclosure process as shown in Figure 1. The concealed image is what I wish to obscure. To accurately regulate this process, the FaceParsing model is employed to extract the mask from the concealed image. The mask, together with the concealed picture, is integrated into the steganographic image via a concealment process [19]. The transmission of steganographic images over the Internet may lead to a decline in image quality due to multiple circumstances, culminating in a compromised steganographic image. Nonetheless, the disclosure procedure can still be integrated with a mask to get the restored image from the compromised steganographic image, preserving the semantic integrity of the content. This paper proposes a framework architecture that emphasizes several essential attributes:

- Utilizing a mask allows for exact regulation of the resulting image's content, hence mitigating the influence of external factors such as the background. This guarantees that steganographic images possess both substantial content and high visual quality.
- Concealment techniques are engineered to be challenging to identify visually, even when confronted with steganographic instruments.
- The revealing process can still produce semantically compatible restored images from a damaged steganographic image, even if it differs somewhat from the original [20].

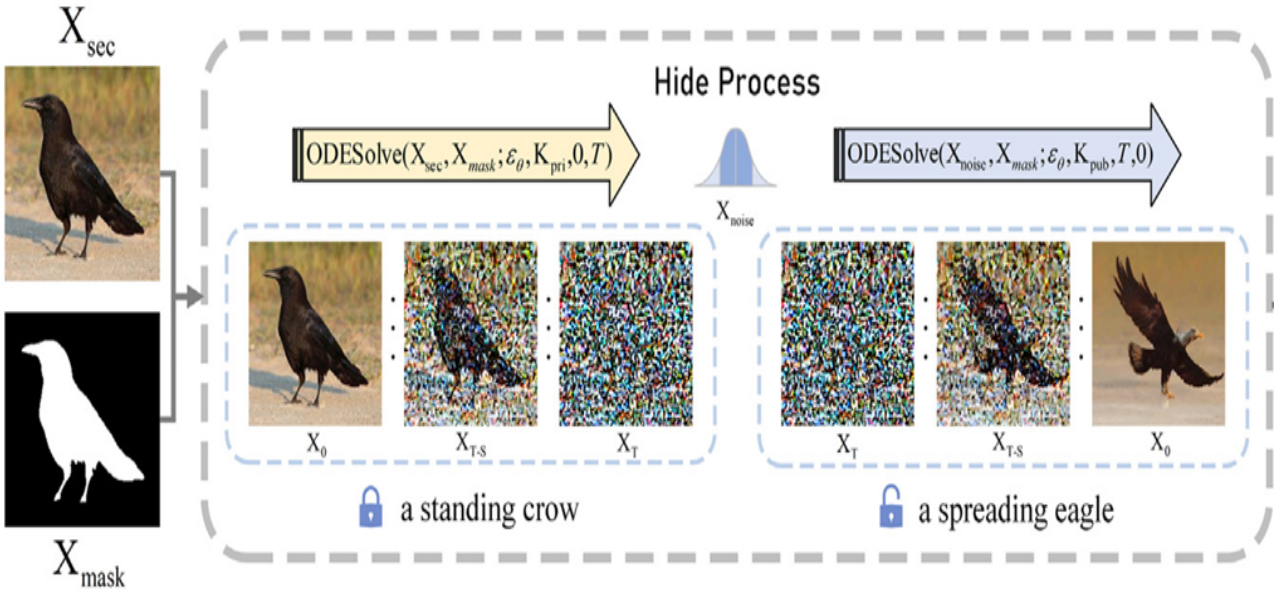

Figure 2: Diffusion model that use deterministic inference.

The forward process in DDIM[36] is delineated by the subsequent equation 1. The reverse sampling procedure of DDIM is delineated by the subsequent the equation 2.

$$x_t = \sqrt{\alpha_t} x_{t-1} + \sqrt{1-\alpha_t}\epsilon, \quad \epsilon \sim \mathcal{N}(0,1) \tag{1}$$

$$x_s \& = \sqrt{\bar{\alpha}_s} f_\theta(x_t, t) + \sqrt{1-\bar{\alpha}_s - \sigma_s^2}\epsilon_\theta(x_t, t) + \sigma_s \epsilon, f_\theta(x_t, t) \& = \frac{x_t - \sqrt{1-\bar{\alpha}_t}\epsilon_\theta(x_t, t)}{\sqrt{\bar{\alpha}_t}} \tag{2}$$

The diffusion model employs deterministic DDIM, which lowers the model's complexity while enhancing its predictability and controllability. Pre-trained noise estimators are utilized to meticulously regulate the noise removal process, consequently enhancing the quality and efficiency of the overall image creation as given in Figure 2. This technology enables the efficient acceleration of the image-generating process while preserving the excellent quality of the produced photos. The equation delineating the sampling procedure utilizing the pre-trained noise estimator. DDIM Inversion delineates

the process in which the original image x0 is transformed into a latent code xT, which is then returned to the original image, resulting in an output image designated as x0, nearly equivalent to x0 [21, 22].

4 Results & Discussion

Moreover, our steganographic photos facilitate the smooth alteration of character attributes, including gender, age, and facial hair, with a significant level of precision. Regarding controllability (shown in Figure 6), our method may execute steganography in some regions while preserving the integrity of other parts. Utilizing the private key to accurately maintain the semantic information of the confidential image, exhibiting exceptional fidelity. Among these, PSNR and SSIM indicate that a higher score for both metrics correlates with superior quality of the reconstructed image [25].

Simultaneously, a lower score in LPIPS, FID, and LDM indicates that the produced image closely resembles the genuine image in visual perception, exhibiting greater similarity in visual content and style. This significantly diminishes the likelihood of being recognized as holding steganographic information. The results indicate that the procedure is markedly superior to alternative methods on the baseline. Additionally, to ascertain the applicability of this strategy in real-world scenarios. Choose two widely utilized facial recognition platforms, Face++ and Aliyun API, as target recognition models as given in Table 1. Figure 8 illustrates the outcomes of comparison experiments conducted on the Stego240 dataset, displaying the confidence scores of facial recognition achieved by various methods across the two models. The experimental results indicate that the face recognition rate for both the recovered image and the secret image exceeds 96% on Face++, achieving the greatest confidence level. This further substantiates that our methodology exhibits exceptional adaptability and great performance across various real-world applications. To assess the efficacy and feasibility of the pro-

Table 1: The produced images using stego were assessed for their resemblance to natural imagery.

Methods	BRISQUE↓	NIQE↓	PIQE↓
Baluja	19.43	4.70	15.25
HiNet	18.01	4.88	13.71
HiDDen	17.78	4.94	11.28
WengNet	17.84	5.09	9.54
Cross	10.11	5.15	6.10
Ours	9.85	5.16	5.81

posed method, three facial recognition models—Deep Face, FaceNet, and ArcFace—were employed for systematic comparison and analysis. The labeling accuracy and facial feature matching on the high-quality public dataset CelebA-HQ, together with selected recovered image datasets, were meticulously assessed. Figure 7 displays a sequence of test photographs. The test photos comprise various representations of the same individual from the CelebA dataset. The distances among each test image, the original image, and the recovered image are calculated, and the verification outcomes are presented. The visual comparison in the picture reveals that the restoration matching results align with the verification outcomes of the original image, exhibiting a minimal distance. This illustrates that the restored image data may adequately substitute the original picture data for facial recognition, and also confirms the superior recovery quality of the approach from a lateral perspective.

To further illustrate the efficacy of the approach, real-world degradation was also evaluated. To emulate the effects of network transmission, experiments were performed to transmit and capture container images on the screen using the WeChat network. As illustrated in Figure 10, under this intricate deterioration situation, all alternative approaches either fail completely or exhibit considerable color distortion. Conversely, the approach effectively discloses the overarching content of the concealed image while preserving substantial semantic coherence with the private key. Once more demonstrating the preeminence of the technique. In both extreme scenarios, the method attains the maximum confidence levels of 93.99% (WeChat) and 87.29% (Shoot) on Face++. The proposed method has also achieved superior reconstruction quality relative to the most recent techniques. The experimental

results comprehensively confirm the efficacy and resilience of the technique across diverse experimental and real-world settings.

5 Conclusion

The rapid progress of generative models has given rise to a new method in steganography termed generative steganography (GS). It enables the generation of images from messages without the need for a carrier image. This globally acknowledged biometric facial recognition method is widely employed in many identity verification systems. This study presents an innovative coverless steganography framework for facial recognition images via a diffusion model, designed to improve personal privacy protection and facilitate the secure transmission and sharing of sensitive information without detracting from user experience. We present a Coverless Semantic Steganography Communication system that uses a Generative Diffusion Model to embed concealed images into created stego images.

References

[1] H. Ullah, M. U. Haq, S. Khattak, G. Z. Khan, and Z. Mahmood, "A robust face recognition method for occluded and low-resolution images," in *Proc. Int. Conf. Appl. Eng. Math. (ICAEM)*, Aug. 2019, pp. 86–91.

[2] W. Junqing, P. Changgen, T. Weijie, and W. Zhenqiang, "FaceEncAuth: Face recognition privacy security scheme based on facenet and small algorithms," *J. Comput. Eng. Appl.*, vol. 58, no. 11, p. 93, 2022.

[3] R. Pena, F. A. Ferreira, F. Caroli, L. J. S. Silva, and H. Lopes, "Globo face stream: A system for video meta-data generation in an entertainment industry setting," in *Proc. ICEIS*, 2020, pp. 350–358.

[4] C. P. Sumathi, T. Santanam, and G. Umamaheswari, "A study of various steganographic techniques used for information hiding," *Int. J. Comput. Sci. Eng. Surv.*, vol. 4, no. 6, pp. 9–25, Dec. 2013. [Online]. Available: https://ar5iv.org/pdf/1401.5561

[5] Y. Xu, C. Mou, Y. Hu, J. Xie, and J. Zhang, "Robust invertible image steganography," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 7875–7884.

[6] S. Mathur, N. Gupta, and D. Garg, "Secure image steganography with blockchain for copyright protection," *J. Cybersecur. Digit. Forensics*, vol. 11, no. 1, pp. 101–115, 2023.

[7] D. Ariatmanto and F. Ernawan, "Adaptive scaling factors based on the impact of selected DCT coefficients for image watermarking," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 3, pp. 605–614, Mar. 2022.

[8] C.-C. Chang and C.-T. Li, "Algebraic secret sharing using privacy homomorphisms for IoT-based healthcare systems," *Math. Biosci. Eng.*, vol. 16, no. 5, pp. 3367–3381, 2019.

[9] G.-D. Su, C.-C. Chang, and C.-C. Lin, "High-precision authentication scheme based on matrix encoding for AMBTC-compressed images," *Symmetry*, vol. 11, no. 8, p. 996, Aug. 2019.

[10] I. J. Kadhim, P. Premaratne, P. J. Vial, and B. Halloran, "Comprehensive survey of image steganography: Techniques, evaluations, and trends in future research," *Neurocomputing*, vol. 335, pp. 299–326, Mar. 2019.

[11] H. Olewi, M. Msallam, S. Salim, and H. Al-Behadili, "Enhanced security through integrated Morse code encryption and LSB steganography in digital communications," *Traitement du Signal*, vol. 1, no. 1, pp. 519–524, 2024.

[12] J. Wang, X. Chen, J. Ni, N. Mao, and Y. Shi, "Multiple histograms-based reversible data hiding: Framework and realization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 8, pp. 2313–2328, Aug. 2020.

[13] S. Kamil Khudhair, M. Sahu, K. R. Raghunandan, and A. Sahu, "Secure reversible data hiding using block-wise histogram shifting," *Electronics*, vol. 12, no. 5, p. 1222, Mar. 2023.

[14] G.-D. Su, Y. Liu, and C.-C. Chang, "A square lattice oriented reversible information hiding scheme with reversibility and adaptivity for dual images," *J. Vis. Commun. Image Represent.*, vol. 64, Oct. 2019, Art. no. 102618.

[15] C.-C. Chang, C.-T. Li, and K. Chen, "Privacy-preserving reversible information hiding based on arithmetic of quadratic residues," *IEEE Access*, vol. 7, pp. 54117–54132, 2019.

[16] S. Baluja, "Hiding images in plain sight: Deep steganography," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.

[17] S.-P. Lu, R. Wang, T. Zhong, and P. L. Rosin, "Large-capacity image steganography based on invertible neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10816–10825.

[18] H. Yang, Y. Xu, X. Liu, and X. Ma, "PRIS: Practical robust invertible network for image steganography," *Eng. Appl. Artif. Intell.*, vol. 133, Jul. 2024, Art. no. 108419.

[19] J. Zhu, R. Kaplan, J. Johnson, and L. Fei-Fei, "Hidden: Hiding data with deep networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 657–672.

[20] J. Qin, Y. Luo, X. Xiang, Y. Tan, and H. Huang, "Coverless image steganography: A survey," *IEEE Access*, vol. 7, pp. 171372–171394, 2019.

[21] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2223–2232.

[22] Z. Zhou, H. Sun, R. Harit, X. Chen, and X. Sun, "Coverless image steganography without embedding," in *Proc. 1st Int. Conf. Cloud Comput. Secur. (ICCCS)*, Nanjing, China. Cham, Switzerland: Springer, 2015, pp. 123–132.

[23] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 6840–6851.

[24] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," 2020, arXiv:2011.13456.

[25] A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte, and L. Van Gool, "RePaint: Inpainting using denoising diffusion probabilistic models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 11461–11471.