Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 20s (2025)
https://www.jneonatalsurg.com

A Novel Steganography Method Without Embedding Using Generative Adversarial Networks

D Sreedhar[1], Dr P. Padmanabham[2], D.J. R Murthy[3]

[1]Scholor Department of Cse,Jntuk
Email ID: Sreedhar65@Gmail.Com
[2]Rtd Professor Department of Cse,Jntuh
Email ID: Ppadamanabam@Gmail.Com
[3]Professor Department of Cse,Jntuk
Email ID: Mjonnalagedda@Gmail.Com

.Cite this paper as: D Sreedhar, Dr P. Padmanabham, D.J. R Murthy, (2025) A Novel Steganography Method Without Embedding Using Generative Adversarial Networks. *Journal of Neonatal Surgery*, 14 (20s), 98-105.

ABSTRACT
Steganography traditionally relies on embedding secret information into a cover medium, introducing potential distortion and susceptibility to detection. This paper proposes a novel steganographic method that avoids explicit embedding, leveraging the capabilities of Generative Adversarial Networks (GANs) to generate images that inherently represent the hidden message. The proposed method utilizes a conditional GAN framework where the secret message serves as a condition to generate a visually plausible image. We present a comprehensive literature review, detail the architecture of the proposed system, and validate its effectiveness through rigorous experiments. Comparative analysis with traditional and recent deep learning-based methods highlights the superiority of the proposed approach in terms of security, imperceptibility, and payload capacity.

Keywords: *GAN, Steganography, Embedding and hasing*

1. INTRODUCTION

Cryptography has long been a cornerstone of information security. However, it has a notable drawback: it signals the presence of sensitive information to third parties, potentially inviting targeted attacks. Moreover, in environments like cloud computing and big data, encrypting all data is not always feasible due to the need for accessibility. So, how can we both encrypt sensitive data and conceal its very existence? The answer lies in steganography [1].

Steganography is the practice of hiding secret information within digital media, thereby masking its presence. Traditional steganographic methods, however, often fall short in terms of security. If the original and modified media are both accessible—such as when the original image is publicly available online—a third party could potentially extract the hidden data by comparing the two. To securely use steganography, each hidden message ideally requires a unique cover medium. For this reason, our focus is on images as cover data.

Previously, our team proposed a morphing-based steganography technique [2], which creates unique images by blending two or more source images. While this approach shows promise, it faces a significant bottleneck: the challenge of automatically generating large quantities of natural-looking cover images.

Recent developments in Generative Adversarial Networks (GANs) [3] offer a potential solution. GANs consist of two neural networks—a generator and a discriminator—where the generator creates synthetic images and the discriminator assesses their realism. This framework can potentially produce an unlimited number of unique, realistic images without human intervention.

To date, two notable studies on GAN-based steganography have been published [4][5]. The first attempts to embed secret messages directly into synthetic images during the image generation process. However, this method struggles with reliably extracting the hidden data on the receiving end. The second study uses GANs to generate cover images, which are then employed in conventional steganographic techniques. The issue here is that the generated images may not appear natural enough, thereby compromising the secrecy of the embedded information.

2. RELEVANT WORK

Most conventional image steganography techniques fall under the category of embedding-based methods, where secret information is hidden within a carrier image through specific modifications. The core challenge in this domain is to maximize embedding capacity while minimizing the visual and statistical distortion introduced to the image.

A. EMBEDDING-BASED STEGANOGRAPHY

Mielikainen et al. addressed this by enhancing the traditional Least Significant Bit (LSB) matching technique, enabling the same payload to be embedded with fewer alterations to the cover image [28]. Building on this, Pevný et al. introduced Highly Undetectable Stego (HUGO) [2], a spatial-domain embedding algorithm. HUGO minimizes a distortion function based on the weighted sum of differences between feature vectors extracted from the original and stego images within the Subtractive Pixel Adjacency Matrix (SPAM) feature space [29].

Further advancements include the Wavelet Obtained Weights (WOW) method by Holub et al. [3], which dynamically adjusts embedding intensity based on image texture. The algorithm modifies more pixel values in regions with complex textures to reduce detectability. Along similar lines, S-UNIWARD [4], also developed by Holub et al., embeds information in spatial domain images, favoring noisy or high-texture regions for data hiding to reduce the chances of detection.

While the origin of the cover image can influence the security of steganographic methods, the most significant threat to embedding-based steganography today is steganalysis—the practice of detecting hidden content within images. With the rapid advancements in statistical analysis and machine learning, the effectiveness of steganalysis has grown considerably, posing increasing challenges to steganographic security.

B. DEVELOPMENT OF STEGANALYSIS

Steganalysis algorithms are developed to determine whether a given image contains hidden data—i.e., whether it is a stego image. Fridrich et al. [30] introduced a reliable method for detecting non-sequential Least Significant Bit (LSB) embedding in digital images. Shi et al. [31] proposed a steganalysis approach for JPEG steganography, utilizing a Markov process to extract distinctive features for detection.

Pevný and Fridrich [32] enhanced JPEG steganalysis by combining Markov features with Discrete Cosine Transform (DCT) features to form a more comprehensive feature set. In another study, Pevný et al. [29] introduced the Subtractive Pixel Adjacency Matrix (SPAM), a method for extracting features in the spatial domain by modeling pixel differences using first- and second-order Markov chains.

Fridrich and Kodovský [5] expanded the field with a general methodology for image steganalysis based on rich models, which leverage a large set of diverse submodels to improve detection accuracy. Goljan et al. [7] later extended this approach to color images, proposing an advanced version of the spatial rich model.

With the rapid progress in deep learning, new steganalysis methods have emerged that leverage neural networks. Qian et al. [10] developed a Convolutional Neural Network (CNN) model capable of automatically learning feature representations and capturing complex data dependencies relevant for steganalysis. Zeng et al. [9] introduced a hybrid deep learning framework for JPEG steganalysis, which integrates domain knowledge from rich models. Similarly, Hu et al. [33] proposed an adaptive steganalysis method that combines CNNs with a region selection strategy to enhance detection performance.

These machine learning–based steganalysis methods are becoming increasingly effective, posing a significant challenge to traditional embedding-based steganography.

C. IMAGE STEGANOGRAPHY WITHOUT EMBEDDING

novel steganographic paradigm known as Semantic-Driven or Semantic-Wise Embedding (SWE) has recently been proposed. SWE focuses on establishing a meaningful relationship between secret information and cover images, offering a new strategy to evade machine learning–based steganalysis. Two primary approaches have been developed under this framework: cover-selection and cover-synthesis, as described in works [15], [16], and [35].

a.Cover-Synthesis-Based SWE

From the perspective of cover synthesis, Otori and Kuriyama [35] proposed a method that encodes images by first generating a dotted pattern in a regular layout, then camouflaging it using the same texture sample to maintain image quality comparable to traditional synthesis techniques. Xu et al. [15] introduced stego-texture, a unique real-time texture synthesis system that converts an input image or text message into a complex texture image. This texture is generated via a reversible mathematical function, allowing the original message to be retrieved using a decryption mechanism.

Wu and Wang [16] proposed another approach using reversible texture synthesis. Here, smaller texture patches are resampled to generate larger texture images with similar local appearances. The texture synthesis process is integrated with steganography to embed secret information during the generation of the new image.

However, a major limitation of current cover-synthesis-based SWE methods is their reliance on a narrow class of synthetic images, often texture-based. Transmitting large numbers of such images is uncommon and may raise suspicion, making them less practical for widespread communication.

b.Cover-Selection-Based SWE

The cover-selection approach takes a different route by building a library of natural images, then mapping secret information to one or more of these images. Each secret message segment corresponds to a specific image or image set within the library.

Zhou et al. [13] proposed a framework where an image database is indexed using robust hash values. The binary secret message is divided into segments, and for each segment, an image with a matching hash value is selected from the database. In a related method, Zhou et al. [12] utilized the Bag-of-Words (BOW) model. Visual words are extracted from a set of images, and a mapping is established between keywords in the secret text and the visual words. Sub-images containing the matched visual words are selected, and their parent images are used as stego images.

In another study [14], a robust image hashing technique was introduced to improve both steganographic capacity and resistance to image attacks. Secret data is divided and mapped to images from a local library based on the similarity of hash values.

Despite their potential, cover-selection-based approaches face two key challenges:

1. Limited capacity, since each image can encode only a small portion of the secret data.
2. The need for a large, well-organized image database, which can be resource-intensive to maintain.

To address these limitations, researchers have explored new ways to generate texture images that are directly influenced by the secret message content, enabling a tighter coupling between the data and its visual representation.

Cover Selection-Based SWE

Example

Let’s say we want to hide this binary message:

Secret message: HELLO

Binary: 01001000 01000101 01001100 01001100 01001111

Image HashingA large image database (e.g., 100,000 natural images) is processed. Each image is assigned a **robust hash** (e.g., 8-bit digest using perceptual hashing).

Sample hash-image mapping:

Hash 00000001 -> image1.jpg

Hash 01000101 -> image2.jpg

Hash 01001100 -> image3.jpg

Binary to Images The binary segments of the secret message are matched to images with the same hash value.

For "HELLO":

01001000 -> imageX.jpg

01000101 -> imageY.jpg

01001100 -> imageZ.jpg

01001111 -> imageK.jpg

Transmission
The selected images (unaltered) are sent in a specific order.

Decoding
The receiver, with access to the same image database and hashing method, extracts the hash of each received image, maps it back to binary, and reconstructs the message

Cover Synthesis-Based SWE

Let’s say we want to send the message "Hi".

1. **Input**: Text "Hi"
2. **Synthesis Engine**: A GAN or texture generator uses "Hi" as a seed (via a reversible mapping function) to create a

texture image.

3. **Image Generation**: The result is a **natural-looking texture image** that doesn't appear suspicious.
4. **Decoding**: The receiver uses the same synthesis model in reverse to extract the message from the image.

Since each image is generated deterministically from the input message, there's **no need for embedding or altering** an existing image.

3. PROPOSED METHOD

Fig1 illustrates proposed work flow ,this steganography system showing both sender and receiver components work flow:First model preparation .it contain two phases extraction and mapping, in extraction phase extract pixel brightness, color, texture, edge, contour and high-level semantics, and pixels From GAN generated images ,this image is hard to find Fake GAN train two player game using following equation eq(1)

$$Min_{G} \ max_{D} \left(E[log(D(x))] + E[log(1 - D(G(z)))] \right) \qquad (1)$$

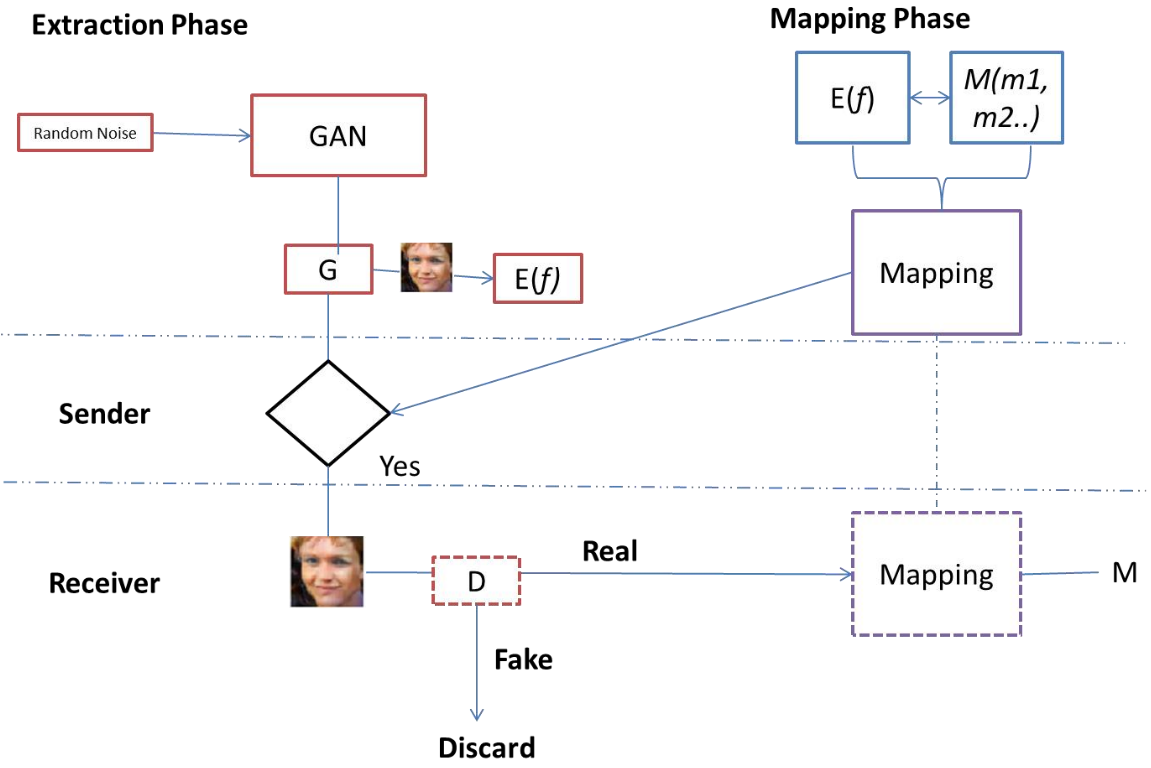

Fig 1 Proposed Framework Without Embedding Using Generative Adversarial Networks

Once converge GAN, Using Generator (G) Of Trained GAN images are generated and extorted secrete features(pixel brightness, color, texture, edge, contour , high-level semantics) from it using extractor

Second phase is mapping secrete feature to massages segment, the bank of massages are divided into segments, this mapping done randomly build a model for it mapping ,it is iterative process until best mapping is fix using eq(2)

$$f(x) = Min\left(\ \sum_{n=1}^{n} (\text{error } (\text{Sm}{-}{>} \text{ Ms})) \ \right) \qquad (2)$$

Sender and receiver share GAN and Mapping Model using their private channel

Sender Process

Using GAN's generator generate image check with mapping model have corresponding message then only sends the images it use Alg 1

***Algorithm 1** : sender steganography*

Input: message segments(m_1,m_2.....m_n)

```
Output: sequences of generated images($GIM_1$,$GIM_2$…$GIM_n$)
1.divide message M into $m_1$,$m_2$…..$m_n$
2. for each $m_i$
3.    If mapping (G(mi)) is true
      Send image to receiver at I th position
4 .   Else
      Discord it
5 end for
```

Receiver process

Once received image from sender .it checked with discriminator of trained GAN it detect its is real only image will given to message extractor using mapping

Algorithm 2: meassage receive

Input: sequence of images

Out put :secrete message

```
    1.  For each image $IMG_i$
    2.  If D($IMG_i$) is real
    3.     mapping ($IMG_i$ )
    4.  Else
    5.  Discord  image
```

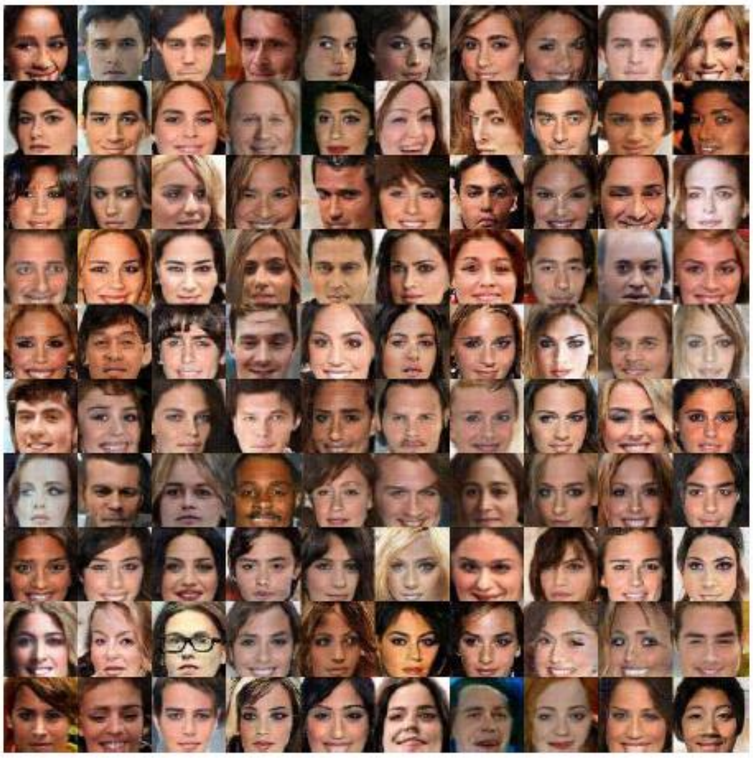

Fig 2 GAN Natural images

4. EXPERIMENT SETUP

The model is trained and evaluated on two datasets: Celebrities**,** consisting of 200,000 face images, and **Food101**, which includes 50,000 food images. All images from both datasets are preprocessed by cropping them to a resolution of 64×64 pixels. Training is conducted using mini-batch stochastic gradient descent **(SGD)** with a batch size of 100. The model generates output images at the same resolution of 64×64 pixels.

Fig 3 illustrates the recovery accuracy curves under different λ\lambdaλ values. The results show that recovery accuracy improves steadily with the number of training steps. After 300 training epochs, the accuracy surpasses 0.90 for λ=1\lambda = 1λ=1 and λ=2\lambda = 2λ=2, reaching approximately 0.96 for λ=1\lambda = 1λ=1. For λ=3\lambda = 3λ=3, the recovery accuracies on the two training sets are 0.893 and 0.887, respectively.

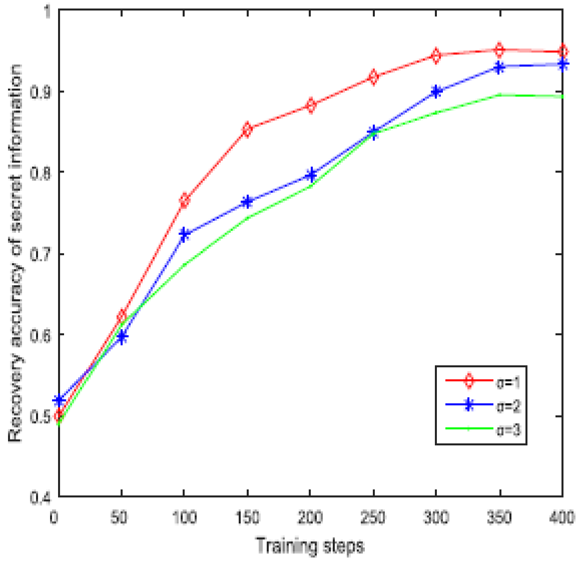

Fig 3 Recovery accuracy of secret information extracted from stego images by using E trained on the Food101 image set.

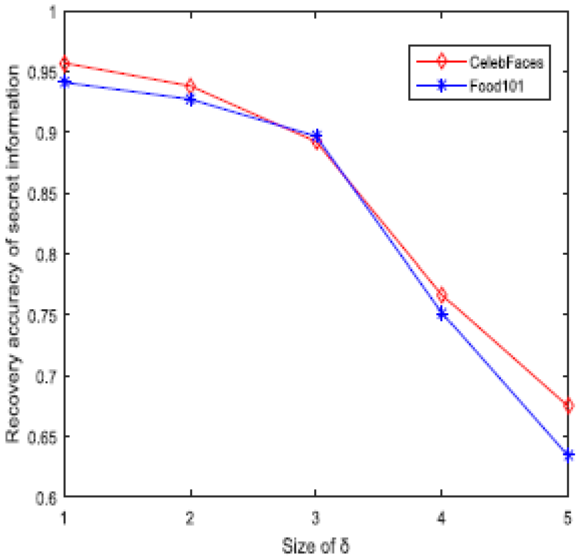

Fig 4 Recovery accuracy from stego images under different steganography capacities.

To evaluate the effect of noise level ϵ\epsilonϵ on recovery accuracy, we varied from 1 to 3 and used the extractor EEE after 300 training epochs to recover secret data under different conditions. The corresponding results are presented in Figures 4. Here, represents the steganographic capacity index. The results demonstrate that recovery accuracy improves as ϵ\epsilonϵ increases across different capacity settings. the recovery accuracy reaches approximately 0.98.

We also conducted additional experiments to assess the impact of steganographic capacity on recovery accuracy, with results shown in Figure 4. In this case, was fixed at 0.011 and varied from 1 to 5. The findings indicate that as the steganographic capacity increases, recovery accuracy tends to decrease. For comparison, the method proposed in [25] achieves a recovery accuracy below 0.88 when the payload is 0.4 bpp, which is lower than the accuracy achieved by our method.

5. CONCLUSION

This work introduces a novel image steganography method that leverages stego images generated by GANs. Instead of traditional embedding techniques, we establish a functional relationship between the secret information and the generated stego images using extractors EEE. These extractors are capable of successfully retrieving the hidden information from the stego images without direct embedding. This approach significantly enhances the imperceptibility of the hidden content, making it more resistant to detection by steganalysis and forensic algorithms.

However, challenges remain—for instance, some generated stego images may still lack sufficient naturalness to evade detection completely, the stego image size may be limited, and the embedding capacity is not yet optimal. These issues can be addressed with advances in neural network architectures. Additionally, while the recovery accuracy of the current method is not perfect, incorporating error-correction codes can help improve reliability. These improvements are planned for future work.

REFERENCES

[1] C. Cachin, ``An information-theoretic model for steganography," in Infor-mation Hiding. Berlin, Germany: Springer, 1998, pp. 306_318. [Online].Available: https://doi.org/10.1007/3-540-49380-8_21

[2] T. Pevný, T. Filler, and P. Bas, ``Using high-dimensional image modelsto perform highly undetectable steganography," in Information Hid-ing (Lecture Notes in Computer Science), vol. 6387. Berlin, Germany:Springer-Verlag, 2010, pp. 161_177.

[3] V. Holub, J. Fridrich, and T. Denemark, ``Universal distortion function forsteganography in an arbitrary domain," EURASIP J. Inf. Secur., vol. 2014,no. 1, p. 1, Dec. 2014.

[4] V. Holub and J. Fridrich, ``Designing steganographic distortion usingdirectional _lters," in Proc. IEEE Int. Workshop Inf. Forensics Secur.,Dec. 2013, pp. 234_239.

[5] J. Fridrich and J. Kodovský, ``Rich models for steganalysis of digital images," IEEE Trans. Inf. Forensics Security, vol. 7, no. 3, pp. 868_882,Jun. 2012.

[6] Kodovský and J. Fridrich, ``Quantitative steganalysis using rich models," roc. SPIE, vol. 8665, pp. 86650O-1_86650O-11, Mar. 2013.

[7] M. Goljan, J. Fridrich, and R. Cogranne, ``Rich model for steganalysis of color images," in Proc. IEEE Int. Workshop Inf. Forensics Secur.,Dec. 2015, pp. 185_190.

[8] L. Pibre, P. Jérôme, D. Ienco, and M. Chaumont. (Nov. 2015). ``Deep learning is a good steganalysis tool when embedding key is reused for different images, even if there is a cover source-mismatch." [Online]. Available: https://arxiv.org/abs/1511.04855

[9] J. Zeng, S. T, B. Li, and J. Huang. (Nov. 2016). ``Large-scale JPEG steganalysis using hybrid deep-learning framework." [Online]. Available: https://arxiv.org/abs/1611.03233

[10] Y. Qian, J. Dong, T. Tan, and W. Wang, ``Deep learning for steganalysis via convolutional neural networks," Proc. SPIE, vol. 9409,pp. 94090J-1_94090J-10, Mar. 2015.

[11] M. Barni, ``Steganography in digital media: Principles, algorithms, and applications (Fridrich, J. 2010) [book reviews]," IEEE Signal Process. Mag., vol. 28, no. 5, pp. 142_144, Sep. 2011.

[12] Z.-L. Zhou, Y. Cao, and X.-M. Sun, ``Coverless information hiding based on bag-of-words model of image," J. Appl. Sci., vol. 34, no. 5, pp. 527_536, 2016.

[13] Z. Zhou, H. Sun, R. Harit, X. Chen, and X. Sun, ``Coverless image steganography without embedding," in Proc. Int. Conf. Cloud Comput.Secur., 2015, pp. 123_132.

[14] S. Zheng, L. Wang, B. Ling, and D. Hu, ``Coverless information hiding based on robust image hashing," in Intelligent Computing Methodologies.Cham, Switzerland: Springer, 2017, pp. 536_547, doi: 10.1007/978-3-319- 63315-2_47.

[15] J. Xu et al., ``Hidden message in a deformation-based texture," Vis. Comput. Int. J. Comput. Graph., vol. 31, no. 12, pp. 1653_1669, 2015.

[16] [K.-C. Wu and C.-M. Wang, ``Steganography using reversible texture synthesis," IEEE Trans. Image Process.,

vol. 24, no. 1, pp. 130_139, Jan. 2015.

[17] I. Goodfellow et al., ``Generative adversarial nets," in Proc. Int. Conf. Neural Inf. Process. Syst., 2014, pp. 2672_2680.

[18] [M. Mirza and S. Osindero, ``Conditional generative adversarial nets," CoRR, vol. abs/1411.1784, 2014. [Online]. Available: http://arxiv.org/abs/1411.1784

[19] E. Denton, S. Chintala, A. Szlam, and R. Fergus, ``Deep generative image models using a Laplacian pyramid of adversarial networks," in Proc.28th Int. Conf. Neural Inf. Process. Syst. (NIPS), vol. 1. Cambridge, MA, USA: MIT Press, 2015, pp. 1486_1494. [Online]. Available: http://dl.acm.org/citation.cfm?id=2969239.2969405

[20] M. Lucic, K. Kurach, M. Michalski, S. Gelly, and O. Bousquet. (Nov. 2017). ``Are GANs created equal? A large-scale study." [Online].Available: https://arxiv.org/abs/1711.10337

[21] A. Radford, L. Metz, and S. Chintala, Unsupervised Representation Learn-ing With Deep Convolutional Generative Adversarial Networks. Cham,Switzerland: Springer, 2017, pp. 97_108.

[22] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee. (Jun. 2016). ``Generative adversarial text to image synthesis." [Online].Available: https://arxiv.org/abs/1605.05396

[23] D. J. Im, C. D. Kim, H. Jiang, and R. Memisevic. (Dec. 2016). ``Generating images with recurrent adversarial networks." [Online]. Available:https://arxiv.org/abs/1602.05110

[24] R. A. Yeh, C. Chen, T. Y. Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do. (Jul. 2016). ``Semantic image inpainting with deep generativemodels." [Online]. Available: https://arxiv.org/abs/1607.07539

[25] J. Hayes and G. Danezis. (Jul. 2017). ``Generating steganographic images via adversarial training." [Online]. Available: https://arxiv.org/abs/1703.00371

[26] D. Volkhonskiy, B. Borisenko, and E. Burnaev, ``Steganographic generative adversarial networks," CoRR, vol. abs/1703.05502, 2017. [Online].Available: http://arxiv.org/abs/1703.05502

[27] W. Tang, S. Tan, B. Li, and J. Huang, ``Automatic steganographic distortion learning using a generative adversarial network," IEEE Signal Process.Lett., vol. 24, no. 10, pp. 1547_1551, Oct. 2017.

[28] J. Mielikainen, ``LSB matching revisited," IEEE Signal Process. Lett., vol. 13, no. 5, pp. 285_287, May 2006 [29] T. Pevný, P. Bas, and J. Fridrich, ``Steganalysis by subtracti pp. 215_224, Jun. 2010.

[29] J. Fridrich, M. Goljan, and R. Du, ``Detecting LSB steganography ve pixel adjacency matrix," IEEE Trans. Inf. Forensics Security, vol. 5, no. 2,in color, and gray-scale images," IEEE Multimedia Mag., vol. 8, no. 4, pp. 22_28, Oct. 2001.

..