This item was submitted to [Loughborough's Research Repository](#) by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

Camouflage generative adversarial network: Coverless full-image-to-image hiding

PLEASE CITE THE PUBLISHED VERSION

<https://doi.org/10.1109/SMC42975.2020.9283054>

PUBLISHER

IEEE

VERSION

AM (Accepted Manuscript)

PUBLISHER STATEMENT

Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

LICENCE

All Rights Reserved

REPOSITORY RECORD

Liu, Xiyao, Ziping Ma, Xingbei Guo, Jialu Hou, Gerald Schaefer, Lei Wang, Victoria Wang, and Hui Fang. 2020. "Camouflage Generative Adversarial Network: Coverless Full-image-to-image Hiding". Loughborough University. <https://hdl.handle.net/2134/13387352.v1>.

Camouflage Generative Adversarial Network: A Coverless Full-image-to-image Hiding

Xiyao Liu, Ziping Ma, Xingbei Guo and Jialu Hou
School of Computer Science and Engineering
Central South University
Changsha, China
{lxyzoewx,qdmaziping,guoxingbei,csuhjl}@csu.edu.cn

Gerald Schaefer
Department of Computer Science
Loughborough University
Loughborough, U.K.

Lei Wang
School of Computer Science and Engineering
textitCentral South University
Changsha, China
wanglei@csu.edu.cn

Victoria Wang
Institute for Criminal Justice Studies
University of Portsmouth
Portsmouth, U.K.
victoria.wang@port.ac.uk

Hui Fang
Department of Computer Science
Loughborough University
Loughborough, U.K.
h.fang@lboro.ac.uk

Abstract—Image hiding, one of the most important data hiding techniques, is widely used to enhance cybersecurity when transmitting multimedia data. In recent years, deep learning-based image hiding algorithms have been designed to improve the embedding capacity whilst maintaining sufficient imperceptibility to malicious eavesdroppers. These methods can hide a full-size secret image into a cover image, thus allowing full-image-to-image hiding. However, these methods suffer from a trade-off challenge to balance the possibility of detection from the container image against the recovery quality of secret image. In this paper, we propose Camouflage Generative Adversarial Network (Cam-GAN), a novel two-stage coverless full-image-to-image hiding method named, to tackle this problem. Our method offers a hiding solution through image synthesis to avoid using a modified cover image as the image hiding container and thus enhancing both image hiding imperceptibility and recovery quality of secret images. Our experimental results demonstrate that Cam-GAN outperforms state-of-the-art full-image-to-image hiding algorithms on both aspects.

Index Terms—image hiding; deep learning; generative adversarial network; image synthesis.

I. INTRODUCTION

Image hiding [1] is an important data hiding technology that allows to conceal a secret message into images when using public communication channels in order to transmit sensitive information, with many image hiding algorithms have been developed to enhance data security [2], [3]. The design of these algorithms focusses mainly on optimising a trade-off between increasing the hidden capacity to embed secret images and preserving their imperceptibility to a third party.

Conventional image hiding techniques, such as LSB [4], HUGO [5], WOW [6], and S-UNIWARD [7], exploit the less significant components from either the image spatial space or frequency space to embed a secret message to achieve

imperceptibility. These conventional image hiding methods unavoidably leave traces of the modification which might be detected by some advanced steganalysis tools.

Coverless image hiding via texture synthesis has become one of the key mechanisms to solve this problem [8]–[11]. These methods design mapping functions based on convolutional neural networks to transfer secret images into another synthesised texture form. Consequently, the embedding stage in the classical image hiding pipeline is not required, making hidden messages much harder to be detected from container images [8].

However, insufficient hidden capacity is one of the major issues of both conventional image hiding techniques and state-of-the-art coverless methods. None of these methods is capable of hiding a full-size secret image. This problem hinders further development of image hiding techniques to satisfy the increasing demand of protecting privacy and security of large-size multimedia data.

Recently, several advanced deep learning-based methods have reached the capacity of a full-size secret image, thus supporting full-image-to-image hiding [12]–[15]. These approaches minimise both the error between cover and container image and the reconstruction error of the secret image using convolutional neural networks. However, an optimal balance between imperceptibility of the container image and reconstruction quality of the secret image is difficult to achieve in these end-to-end deep learning image hiding methods.

In this paper, we propose a two-stage coverless full-image-to-image hiding algorithm named Camouflage Generative Adversarial Network (Cam-GAN) by exploiting both the advantage of coverless image hiding and deep learning-based techniques to achieve full-image-to-image hiding with insignificant reconstruction error of the secret image. The key contributions of Cam-GAN are as follows: (i) A coverless image hiding structure based on image synthesis is deployed to address the error trade-off for better imperceptibility and

This research is supported by the National Natural Science Foundation of China (61602527), Natural Science Foundation of Hunan Province, China (2020JJ4746, 2017JJ3416, 2018JJ2548), and Mobile Health Ministry of Education-China Mobile Joint Laboratory.

Fig. 1. The overall framework of the proposed method. The two-stage Cam-GAN hides a secret image into a container image and recovers the hidden image with insignificant quality loss.

higher reconstruction quality of the secret image. (ii) A cycle consistent GAN is designed to ensure full-image size hidden capacity by synthesising realistic container images. To our knowledge, it is the first coverless image hiding algorithm that can achieve full-image size hidden capacity. (iii) Another refining GAN is introduced together with the cycle consistent GAN as the image hiding decoder to further improve the reconstruction quality of the secret image.

The remainder of the paper is organised as follows. In Section II, we explain the details of our proposed Cam-GAN method. Section III presents experimental results that demonstrate the superiority of our approach. Finally, conclusions are drawn in Section IV.

II. CAM-GAN NETWORK

As illustrated in Fig. 1, we propose a two-stage generative adversarial network (GAN) model for image hiding. The idea is mainly inspired by recent advancements of GANs [16], [17]. Our proposed framework comprises two components: an image hiding encoder and an image hiding decoder. The image encoder is composed of a generator G_1 in Cam-GAN Stage I which synthesises a container image to hide secret images via a texture synthesis function. The image decoder is composed of the other generator, G_2 , in the Cam-GAN Stage I network, and the generator G_3 in the Cam-GAN Stage II network to recover the secret image. In the following, we explain the details of our proposed architecture and its loss functions.

A. Cam-GAN Architecture

In Cam-GAN Stage I, an asymmetric cycle consistent GAN is trained to hide a secret image into a container image via an encoder generator G_1 , and then recover the secret image with insignificant quality loss via a decoder generator G_2 . Here, G_2 is an approximate inverse function of the generator G_1 . The Arnold transformation [18] is applied to scramble the secret image before feeding it into G_1 to conceal the global structure of the secret image. The network architectures of G_1 and G_2 are illustrated in Fig. 2. They comprise four convolutional layers, nine residual network (ResNet) blocks and four deconvolutional layers. This architecture has been successfully deployed in various image transfer tasks [17]. Furthermore, a discriminator D_1 is used to make the container image hard to be discriminated from the gallery of images. This discriminator is a CNN with four convolutional layers and one fully connected layer as illustrated in Fig. 3. The output of the discriminator is a possibility value indicating whether the input image belongs to the image gallery or a synthesised image. The training stage performs adversarial learning as the generators learn a mapping function to synthesise more realistic images so that these images can fool the discriminator.

In Cam-GAN Stage II, a refining GAN is trained to improve the quality of the recovered secret image. The generator G_3 learns the mapping between the recovered and original secret images to improve the recovery quality, while a discriminator D_2 is employed to further enhance the reconstruction quality.

Fig. 2. Network architecture of generators G1 and G2

Fig. 3. Network architecture of discriminator D1.

As illustrated in Fig. 4, G_3 uses a U-Net architecture which contains nine convolutional layers and nine deconvolutional layers and shortcut connections between the convolutional and deconvolutional layers to reuse the feature representations generated from different layers via various convolutional kernels for better recovery quality. The discriminator D_2 shares a similar architecture with the discriminator D_1 (Fig. 3) although the first layer has a different input size.

B. Cam-GAN Loss Function

In Cam-GAN Stage I, there are two terms in our loss function: a reconstruction loss term L_1 defined as

$$L_1(G_1, G_2, x) = E_{x \sim p(x)} [\|G_1(G_2(x)) - x\|_1], \quad (1)$$

and a discriminator loss term L_2 defined as

$$\begin{aligned} L_2(G_1, D_1, x, y) &= E_{x \sim p(x)} [\log(1 - D_1(G_1(x)))] \\ &+ E_{y \sim p(y)} [\log D_1(y)], \end{aligned} \quad (2)$$

leading to

$$G_1^*, G_2^*, D_1^* = \arg \min_{G_1, G_2} \max_{D_1} [\lambda_1 L_1(G_1, G_2) + L_2(G_1, D_1)]. \quad (3)$$

Here, x represents the secret image, y one of the gallery images, G_1 and G_2 denote the two generators to encode and decode the secret image, D_1 is the discriminator to make the synthesised container image more realistic, and λ_1 is a weight to balance the reconstruction and discriminator terms. With the iterative update of D_1 , G_1 and G_2 , the encoder and decoder generators can be converged to hide secret images with texture synthesis.

In Cam-GAN Stage II, we also have a loss function with two terms, a reconstruction term L_3 defined as

$$L_3(G_3, x') = E_{x' \sim p(x')} [\|G_3(x') - x\|_1], \quad (4)$$

and a discriminator term L_4 defined as

$$\begin{aligned} L_4(G_3, D_2, x', x) &= E_{x' \sim p(x')} [\log(1 - D_1(G_1(x')))] \\ &+ E_{x \sim p(x)} [\log D_1(x)], \end{aligned} \quad (5)$$

Fig. 4. Network architecture of generator G3.

giving

$$G_3^*, D_2^* = \arg \min_{G_3} \max_{D_2} [\lambda_2 L_3(G_3) + L_4(G_3, D_2)], \quad (6)$$

where x represents the secret image, x' the recovered secret image from the first stage, G_3 is the generator to refine the quality of the secret image, D_2 the discriminator to enhance the recovered secret image with higher quality, and λ_2 is a weight to balance the reconstruction and discriminator terms.

C. Cam-GAN Implementation

For the implementation, the weight parameters in the loss functions, λ_1 and λ_2 , are set to 10 and 100, respectively, while as training parameters we employ a batch size of 1. The number of epochs of the asymmetric cycle consistent GAN is 300 to 400, while the number of epochs of the refining GAN is fixed to 300. Both learning rates are set to 0.0002 with a linear decay after 150 epochs. For different secret images, we use individual keys to randomly choose the asymmetric cycle consistent GAN models trained after different epochs (randomly selected from 300 to 400) to further enhance the security.

III. EXPERIMENTAL RESULTS

A. Experimental Setup

In our experiments, we first evaluate the image quality of the encoder and decoder of Cam-GAN both subjectively and objectively. We then test the performances of Cam-GAN against steganalysis. To ensure a fair comparison, we use a several image hiding algorithms which have full-image-to-image hidden capacity, including [12] and [13] as benchmark methods, because both image quality and hiding undetectability are directly affected by the size of the hidden message. Further, we compare the hidden capacity of Cam-GAN with [8], [10], [11] to demonstrate superiority compared with state-of-the-art coverless image hiding algorithms.

Our dataset includes 400 gallery images of paintings [17] and 1200 secret images to hide. The image size of both gallery and secret images is 256×256 pixels. The secret images contain 400 facades [19], 400 faces [20] and 400 aerial images [17]; 800 of them are used for Cam-GAN training, while the other 400 are used for testing.

B. Image Quality Evaluation

We first evaluate the performance of the Cam-GAN image hiding encoder and decoder visually, and then provide quantitative comparison results of the reconstruction performance against the benchmark methods in Table I. Note that the proposed method is a coverless image hiding technique, and thus there is no cover image needed in the hiding process.

Fig. 5 and Fig. 6 illustrate that our proposed Cam-GAN can generate realistic container images. In addition, it yields the best reconstruction quality of secret images. As shown in Fig. 5, the residual cover images of both [12] and [13] clearly show patterns from the secret images that can be easily detected by visual analysis. In contrast, the synthesised painting images from Cam-GAN are visually similar to real paintings. The significant advantage of our proposed method is that the container image can better fool both human perception and machine steganalysis without modifying cover images as in the other methods. As illustrated in Fig. 6, for Cam-GAN, the residuals between recovered and original secret images are much more insignificant compared to the other algorithms, demonstrates Cam-GAN to outperform them in terms of recovery quality for the secret image.

To further demonstrate the effectiveness of our proposed Cam-GAN, an objective image quality comparison is performed based on pixel errors, peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM).

PSNR is defined as

$$PSNR = 10 \log_{10} \left(\frac{255^2 HW}{\sum_{i=1}^H \sum_{j=1}^W (I(i, j) - I'(i, j))^2} \right), \quad (7)$$

where I is the original image, I' the recovered image, H and W are the image dimensions, and (i, j) indicate pixel coordinates.

SSIM is calculated as

$$SSIM(x, y) = \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (8)$$

TABLE I
PERFORMANCE COMPARISON OF FULL-IMAGE-TO-IMAGE HIDING METHODS.

	Baluja [12]		Wu <i>et al.</i> [13]		Cam-GAN	
	Cover image	Secret image	Cover image	Secret image	Cover image	Secret image
Pixel Error	9.83	8.39	8.92	7.85	N/A	2.85
PSNR	26.41	27.51	24.81	27.45	N/A	36.23
SSIM	0.91	0.89	0.94	0.95	N/A	0.97

TABLE II
COMPARISON OF HIDDEN CAPACITY.

method	absolute capacity (bit/image)	image size	relative capacity (bit/pixel)
Hu <i>et al.</i> [8]	300	64×64	0.073
Zhang <i>et al.</i> [10]	18.3-135.4	64×64	0.004-0.033
Ke <i>et al.</i> [11]	4	28×28	0.005
Cam-GAN	256×256×24	256×256	24

where μ_x, μ_y are the averages of x and y , σ_x^2, σ_y^2 are the variances and σ_{xy} are covariance for x and y respectively. C_1 and C_2 are balancing constants.

Lower pixel errors, a higher PSNR and a higher SSIM on the secret image (cover image) indicate better recovery quality (image hiding imperceptibility).

The obtained results are shown in Table I from where we can see that the average pixel errors of each channel of the cover images are 9.83 and 8.92 for [12] and [13], respectively. Our proposed Cam-GAN completely avoids using a modified cover images by synthesising new container images, which ensures higher security. Moreover, the average pixel error of the recovered secret image is 2.85 for Cam-GAN, which is significantly lower than for the other two methods, while we also obtain better results in terms of both PSNR and SSIM, clearly demonstrating superior recovery quality.

C. Steganalysis

Avoiding detection from automatic steganalysis tools is another important performance measure of image hiding methods. Following [12], we use a publicly available steganalysis tool, StegExpose [21], for this purpose. We perform detection on a dataset of 400 clean images from [17] and 400 container images from Cam-GAN, [12] and [13] (i.e., 800 images for each method).

A hypothesis test is used to evaluate the efficacy of the proposed method due to the following two reasons: (1) visual comparison based on the receiver operating characteristic (ROC) curve is subjective; it is hard to judge which algorithm performs better if the ROC curves from two methods have many overlapping regions; (2) the curves become unstable when the test sample size is relatively small since the area

under the curve (AUC) can change significantly when adding or removing samples. A significance level p from the hypothesis test can provide an objective comparison by calculating a reliable statistical indicator. We assume that the null hypothesis H_0 is that the ROC curve generated from an algorithm has no significant difference to that obtained from random guessing. Once a p -value is higher than 0.05, this indicates there is no statistically significant difference to the null hypothesis, while a higher p -value indicates better imperceptibility. The obtained p -values are 0.098 for Cam-GAN, 0.0056 for [12], and 0.033 for [13]. Thus, for the two benchmark methods, the p -values are below 0.05 and thus indicate strong evidence against the null hypothesis. Cam-GAN achieves the best result with a p -value larger than 0.05 thus obeying the null hypothesis that there is no statistical difference between our algorithm and random guessing, confirming the hiding imperceptibility capability of our proposed algorithm,

D. Hidden Capacity

Finally, we further compare our Cam-GAN and three coverless image hiding algorithms via image synthesis [8], [10], [11] in terms of hidden capacity as shown in Table II. From there, we can find that Cam-GAN enlarges the hidden capacity significantly compared with these state-of-the-art coverless image hiding algorithms and only our approach yields a capacity 24 bits per pixel (bpp) to enable full-image-to-image hiding.

IV. CONCLUSIONS

In this paper, we have proposed Cam-GAN, a novel coverless full-image-to-image hiding algorithm. To our best knowledge, it is the first coverless image hiding algorithm that can ensure full-size image hiding capacity. Extensive experimental results have demonstrated that our proposed method also outperforms state-of-the-art full-image-to-image hiding algorithms. On one hand, Cam-GAN can generate realistic container images that are difficult to distinguish from gallery images in order to hide the secret image. In this manner, image hiding imperceptibility is significantly enhanced. On the other hand, recovery of the secret image is accompanied by only insignificant image quality loss via a two-stage adversarial learning network. In future work, we will investigate the use of the proposed method in real-world applications.

original secret images

recovered secret images of [12]

residual secret images ($\times 5$) of [12]

recovered secret images of [13]

residual secret images ($\times 5$) of [13]

recovered secret images of Cam-GAN

residual secret images ($\times 5$) of Cam-GAN

original secret images

original cover images

modified cover (container) images of [12]

residual cover images ($\times 5$) of [12]

modified cover (container) images of [13]

residual cover images ($\times 5$) of [13]

synthesed container images of Cam-GAN

Fig. 6. Comparison of image hiding decoder.

Fig. 5. Comparison of image hiding encoder.

REFERENCES

- [1] J. Fridrich, *Steganography in digital media: principles, algorithms, and applications*. Cambridge University Press, 2009.
- [2] A. Cheddad, J. Condell, K. Curran, and P. Mc Kevitt, "Digital image steganography: Survey and analysis of current methods," *Signal Processing*, vol. 90, no. 3, pp. 727–752, 2010.
- [3] V. Holub and J. Fridrich, "Digital image steganography using universal distortion," in *1st ACM Workshop on Information Hiding and Multimedia Security*, 2013, pp. 59–68.
- [4] J. Mielikainen, "LSB matching revisited," *IEEE Signal Processing Letters*, vol. 13, no. 5, pp. 285–287, 2006.
- [5] T. Pevný, T. Filler, and P. Bas, "Using high-dimensional image models to perform highly undetectable steganography," in *International Workshop on Information Hiding*. Springer, 2010, pp. 161–177.
- [6] V. Holub and J. Fridrich, "Designing steganographic distortion using directional filters," in *IEEE International Workshop on Information Forensics and Security*. IEEE, 2012, pp. 234–239.
- [7] V. Holub, J. Fridrich, and T. Denemark, "Universal distortion function for steganography in an arbitrary domain," *EURASIP Journal on Information Security*, vol. 2014, no. 1, p. 1, 2014.
- [8] D. Hu, L. Wang, W. Jiang, S. Zheng, and B. Li, "A novel image steganography method via deep convolutional generative adversarial networks," *IEEE Access*, vol. 6, pp. 38 303–38 314, 2018.
- [9] C. Li, Y. Jiang, and M. Cheslyar, "Embedding image through generated intermediate medium using deep convolutional generative adversarial network," *Computers, Materials & Continua*, vol. 56, no. 2, pp. 313–324, 2018.
- [10] Z. Zhang, J. Liu, Y. Ke, Y. Lei, J. Li, M. Zhang, and X. Yang, "Generative steganography by sampling," *IEEE Access*, vol. 7, pp. 118 586–118 597, 2019.
- [11] Y. Ke, M.-q. Zhang, J. Liu, T.-t. Su, and X.-y. Yang, "Generative steganography with Kerckhoffs' principle," *Multimedia Tools and Applications*, vol. 78, no. 10, pp. 13 805–13 818, 2019.
- [12] S. Baluja, "Hiding images in plain sight: Deep steganography," in *Advances in Neural Information Processing Systems*, 2017, pp. 2069–2079.
- [13] P. Wu, Y. Yang, and X. Li, "Stegnet: Mega image steganography capacity with deep convolutional network," *Future Internet*, vol. 10, no. 6, p. 54, 2018.
- [14] —, "Image-into-image steganography using deep convolutional network," in *Pacific Rim Conference on Multimedia*. Springer, 2018, pp. 792–802.
- [15] R. Zhang, S. Dong, and J. Liu, "Invisible steganography via generative adversarial networks," *Multimedia Tools and Applications*, vol. 78, no. 7, pp. 8559–8575, 2019.
- [16] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, 2014, pp. 2672–2680.
- [17] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *IEEE International Conference on Computer Vision*, 2017, pp. 2223–2232.
- [18] W. Chen, C. Quan, and C. Tay, "Optical color image encryption based on Arnold transform and interference method," *Optics Communications*, vol. 282, no. 18, pp. 3680–3685, 2009.
- [19] R. Tyleček and R. Šára, "Spatial pattern templates for recognition of objects with regular structure," in *German Conference on Pattern Recognition*. Springer, 2013, pp. 364–374.
- [20] S. Yang, P. Luo, C. C. Loy, and X. Tang, "WIDER FACE: A face detection benchmark," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.
- [21] B. Boehm, "Stegexpose - a tool for detecting LSB steganography," *arXiv preprint arXiv:1410.6656*, 2014.