

Secure Secret Image Hiding and Protection Using a Two-Channel Deep Hiding Network

Shivani Ambesange, Rashmi Kulkarni

Department of Information Technology, Siddhant College of Engineering, Sudumbare, Pune, Maharashtra, India

Published October 2026

Abstract

The use of digital images has been growing in recent years in various fields of remote sensing, industrial monitoring, social media, defence and other fields where secret information is present, demanding the securing of the secret image. Hence, image steganography offers a solution to embed the image into any arbitrary or random image, but it has some drawbacks in the steganography capacity, quality of recovery in the process of visualisation, and resistance to image processing attacks. In recent years, deep learning in combination with image hiding has attracted a lot of attention. In this review, much attention is focused on the recently proposed deep learning-based techniques including convolutional neural networks, Encoder-Decoder architecture, generative adversarial networks and attention mechanism to enhance the hidden image capacity and the residual unseen of the concealed image quality. This review also concentrates on the so called high capacity image hiding and hiding a secret image. The salient features of the deep learning-based image hiding techniques are summarized and the representative studies like Deep Steganography, HiDDeN, Improved FC-DenseNet, Two-Channel Deep Hiding Network and GAN-based methods are reviewed and comparatively discussed focusing on various aspects of these methods including hidden image capacity, imperceptibility, quality of reconstructed image, and robustness of hidden image. The recent developments on remote sensing image based steganography are briefly reviewed; the attacks like JPEG compression, cropping, resizing and noise towards deep learning based steganography are also discussed. Deep neural networks have been observed to enhance the image capacity better than a few traditional approaches, and they can enhance the image reconstruction quality, as well. However, there are still a lot of challenges in the field of deep learning image hiding. Therefore, attention guided adversarial mechanisms and multi-stage fusion are anticipated to be a potentially fruitful area for future research.

Keywords: Deep learning; image steganography; secret image hiding; generative adversarial network; attention mechanism; remote sensing images; robustness

1 Introduction

The current digital revolution and global communication technology have revolutionised the way information is exchanged. Today digital images are widely used in the highly sensitive applications, such as military reconnaissance, aerospace remote sensing, industrial monitoring, digital healthcare system, etc. Often these images are classified, proprietary, or very sensitive data, and sending it securely over public or semi public channels is of great challenge, both logistically and technically. Traditionally, the most common way to protect data is through cryptography, but encrypted files make themselves noticeable. The scrambled data may attract attention of enemies and bad guys, possibly leading to cryptanalysis of the data being sent, or even just to the file being destroyed by the bad guys, hampering the communication. Image steganography is based on the principle of absolute imperceptibility that evades this defect. Steganography conceals a secret image or covert message in

an innocuous cover image, so that the mere fact of the existence of the confidential communication is never revealed to anyone who may be listening or monitoring with an automated system (Ansari, Mohammadi, & Parvez, 2019). The resulting hybrid file, which is universally referred to as the stego image, should be unreadable as a digital photo, both visually and structurally, and statistically it should not stand out.

The traditional steganographic approaches were based on the manual operation of pre-established mathematical rules for embedding information. Least Significant Bit substitution is the most widely used method in the spatial domain that substitutes the least significant bits of the pixels of the cover image by the bits of the secret bits. Traditional spatial and frequency domain methods, although relatively simple to implement and low cost to compute, are severely limited and inescapable. They are subjected to a rigid mathematical compromise between embeddability and faithfulness of the image. If the payload gets larger than some minimum size, the cover image's statistical distribution starts to change visibly. These can create visual artifacts and make the stego image extremely susceptible to statistical steganalysis methods targeted at such artifacts [11]. In addition, classical methods do not have built-in robustness that would allow them to withstand the common digital image-processing operations. However, many common operations like JPEG compression, Gaussian noise addition, and structural cropping often destroy the delicate embedded data, making secret recovery difficult or even unfeasible [8].

With the strict restrictions imposed by traditional rule-based algorithms, the academic community has increasingly resorted to deep learning and convolutional neural networks to solve the problem. Deep neural networks have the potential to learn optimal high dimensional representations of data automatically from large data sets, which is not possible with manual embedding algorithms. The neural network is trained to map a very complex, nonlinear relationship between the secret data and the best locations in the cover image for embedding, using very sophisticated encoder-decoder architectures. This paradigm shift has allowed to achieve high capacity image steganography, in which a full-color image can be embedded into another image of the same size. It was believed that such a feat was mathematically impossible under the classical constraints – embedding a full image usually results in destroying the cover image completely, and the data in the cover image was too large to be manipulated [18]. In this ever-changing field, Two-Channel Deep Hiding Network has become a cutting-edge solution. This network not only processes both the cover image and the secret image in parallel by using dedicated input channels, but also benefits from a highly sophisticated architecture that enables it to attain a level of capacity, imperceptibility and robustness that has never been achieved before.

2 Literature review

2.1 The paradigm shift to deep learning architectures

Data-driven deep learning algorithms are a major turning point in the evolution of digital forensics and information security from manually designed steganographic algorithms. The groundwork for this evolution was laid in [17] who showed that a CNN was able to encode high capacity payloads into standard images by spreading out the representation of the secret across all possible bits, instead of only altering the most fragile layers. The concept of deep steganography was introduced by using three distinct convolutional networks: a preparation network to extract features from the secret image, a hiding network to combine the extracted secret features with the cover image, and a reveal network that is trained to recover the secret image from the created stego image [1]. These networks

were trained together, and the model learned a sort of natural distribution of digital images—implicitly distributing the secret data across various frequency bands and color channels. This end-to-end framework had the amazing ability to fit a payload of twenty-four bits per pixel, compared to the previous approaches which typically never managed to exceed fractions of a bit per pixel without producing significant visual degradation.

After the birth of basic deep steganography, the following architectures aimed to further improve the embedding process with the inclusion of adversarial training and better robustness mechanisms. It is worth noting that during training, the HiDDeN framework added an intermediate noise layer between the encoder and decoder [20]. The neural network was explicitly trained in an image domain that had been distorted in several realistic ways (pixel dropout, Gaussian blurring, JPEG compression) prior to the training of the embedding layer, which encouraged the network to learn redundant, very robust embedding strategies. This training did not introduce any noise that could not be recovered from the recovered stego image even if the stego image was lossily transmitted in a network, thereby bridging the gap between the theoretical lab model and the deployment.

At the same time, the inclusion of a generative adversarial network gave a significant tool to enhance the statistical security of steganographic images. Some models like SteganoGAN used a minimax game where three players: a generator, a decoder, and a very sensitive steganalysis discriminator [19]. These schemes explicitly train the discriminator to distinguish between clean images and stego images. The generator should be able to encode the secret information in such a way that it is completely disguised from the discriminator. This penalty, based on the concept of adversarial, is designed so that the generated images have a stego-image conforming to a complex and high-dimensional statistical distribution of natural images, which makes them almost impossible to detect using existing steganalysts [6].

To push the theoretical limits of DH further, researchers added a new concept, called the Universal Deep Hiding, which made the encoding of the secret image completely independent of the cover image being used [3]. In traditional cover-dependent deep hiding, the neural network adjusts its embedding method according to the visual characteristics of the specific cover image. Universal Deep Hiding showed how a secret image could be hidden in a universal, cover-agnostic, high-frequency noise perturbation. This disguised secret could then be overlaid on any random natural image. This technique proved successful and showed that a large amount of steganography takes place in the high frequency region of the image, where human vision is naturally blocked as background texture. These ongoing architectural inventions laid the theoretical foundation for the evolution of highly specialised dual-channel systems for maximum capacity with minimum visual distortion.

3 Methodology

3.1 Architectural framework of the two-channel deep hiding network

The Two-Channel Deep Hiding Network is a very optimized version of the encoder-decoder steganographic method that has been optimized for both payload size and visual quality. In traditional single channel deep hiding models, the cover image is concatenated with the secret image at the first input layer or in the first convolutional layer and a single channel of convolutional filters are used to extract features from the two images while trying to find the best fusion method. This premature concatenation often results in a major loss of semantic information, structural degradation of the stego image and finally, the recovered secret image. To address this fundamental challenge, the Two-Channel Deep

Hiding Network (TDHN) is proposed, which adopts a dual channel structure in the initial feature extraction stage [3].

The structure can be generally categorized into two main parts – hiding network as the encoder and extraction network as the decoder. It is begun by two fully disjointed input channels of the hiding network. The main channel is filled with the standard picture, the secondary channel is filled with the secret picture that is meant for the highest confidence. The network extracts the spatial and textural features from each image independently by using several convolutional layers at the initial stage. The cover channel targets areas of high variance or complexity (like dense vegetation, choppy water or dense city fabric) that can naturally hide high frequency data changes without human suspicion. At the same time, an encrypted channel is used to embed the confidential image into a high-level semantic representation [13].

After these feature maps are refined sufficiently in an independent manner, the network uses a complex multi-stage feature fusion mechanism. The high dimensional feature maps from both the cover and secret channels are mathematically combined and the network carefully embeds the secret representation into the most optimal spatial locations, that are identified within the cover image. Advanced down-up structural design is highly dependent on this fusion process. Down-sampling a sequence of down-sampling operations gradually lowers spatial resolution of the feature maps, and increases channel depth, thus learning deep semantic relationships. Then a series of symmetrical up-sampling operations brings the feature maps back to their original spatial size. This detailed process creates a final stego image with the same width, height and color depth as the original cover image, and thus looking completely normal to any observer.

There are no input and output devices on the extraction network side. Importantly, the receiver does not have to have access to the original cover image or any cryptographic keys to obtain the hidden data. The extraction network is a standalone network that is only receiving the stego image as input. The decoder uses a series of specialised convolutional layers to extract the fine, high-frequency secret signals and gradually reconstruct the secret image. The extraction network is co-trained with the hiding network end-to-end, thus it gains a very synchronized knowledge about the particular embedding protocols used by the encoder. When the secret data is known to both, this knowledge can be used almost perfectly to reconstruct the secret data without the need of complex procedures for exchanging the keys.

3.2 Multi-stage feature fusion and skip connections

Fine-grained spatial details are often lost in image-to-image translation tasks, which is a common problem in training deep convolutional neural networks for such tasks. These layers of down-sampling and nonlinear activation function progressively remove high frequencies in the textures and strong boundaries of high frequencies. In normal image classification, this loss of spatial resolution is fine because the network only needs to be deep enough in semantic context to make a classification. But in the field of image steganography, it is extremely crucial to maintain the exactness of the image. To avoid arousing suspicion, the stego image should be completely at the same visual sharpness level as the cover image, and the extraction network should receive perfectly intact hidden signals to successfully reconstruct the secret image [15]. The Two-Channel Deep Hiding Network is designed with thorough residual connections and skip connections to address the information degradation and the vanishing gradient problem in mathematical models. Skip connections are inspired by the innovative image segmentation architectures, where direct connections are made between different parts of the network without passing through many intermediate convolutional layers [9].

Also, the skip connections are particularly designed to ensure the network preserves the high frequency structural information from the cover image. The network does not require to regenerate the fine details of the cover image from the beginning; it only transfers the raw information of the fine details forward with the skip connections and very slightly embeds the secret features in it. Multi-stage feature fusion is also used in the framework. Multi-stage fusion aligns and integrates feature maps at multiple varying scales and resolutions, instead of combining cover and secret at a single arbitrary bottleneck layer. The network has both macro-semantic features and micro-textural features, and the secret information is not concentrated in either of the frequency bands. Rather, the payload is intelligently broken down and spread throughout the entire structure of the cover image, thus greatly improving the visual imperceptibility.

4 Evaluation metrics and high-capacity hiding

There are several important and competing measures of the effectiveness of an image steganography system: embedding capacity, visual imperceptibility, and secret recovery accuracy. The capacity of traditional steganography (TS) is generally expressed in terms of bits per pixel (bpp). For classical algorithms, typical payloads are still restricted to fractions of a bit per pixel or below, or else suffer from catastrophic visual artifacts. By comparison, architectures based on deep learning are able to compress a complete color image in the cover image of the same size, with an impressive relative embedding capacity of twenty-four bits per pixel.

The Peak signal to noise ratio and the Structural similarity Index measure are used as a quantitative measure of imperceptibility. A higher Peak Signal-to-Noise ratio value means that the image has less distortion at the pixel level, and ratios above 35 dB are typically not perceptible to the naked eye as distortion. The similarity values of the structures range from 0 to 1, where the values closer to 1 represent that the original and the stego image are structurally similar.

Table 1. Comparative analysis of benchmark deep learning steganography architectures based on payload capacity and visual fidelity.

Steganography model	Year	Payload capacity (bpp)	Cover image fidelity (PSNR, dB)	Structural similarity (SSIM)	Architectural innovation
Deep Steganography	2017	24.0	32.8	0.910	End-to-end encoder-decoder
HiDDeN Framework	2018	0.2	38.5	0.965	Adversarial noise simulation
SteganoGAN Model	2019	4.4	36.2	0.942	Three-player GAN dynamics
FCDenseNet Model	2020	24.0	35.1	0.952	Dense block feature reuse
Universal Deep Hiding	2020	24.0	37.4	0.981	Cover-agnostic embedding
TDHN(Two-Channel)	2020	24.0	38.2	0.991	Dual-channel, hybrid loss
Invertible Network	2021	24.0	41.5	0.993	Forward/backward weight sharing

As shown in Table 1, deep steganography has undergone substantial advances, as it has been possible to combine high fidelity with enormous capacities. Early deep models performed well, but fell short in dealing with localized artifacts. While adversarial models were more successful in securing the statistical security, they needed much smaller capacity to produce an acceptable visual quality. The Two-Channel Deep Hiding Network is a paradigm change with an astonishing signal to noise ratio and structural similarity of near perfect values, and while maintaining the maximum payload constraint. This is due to the non-activated feature fusion that avoids color degradation and block artifacts that are present in previous models, and the combined loss functions.

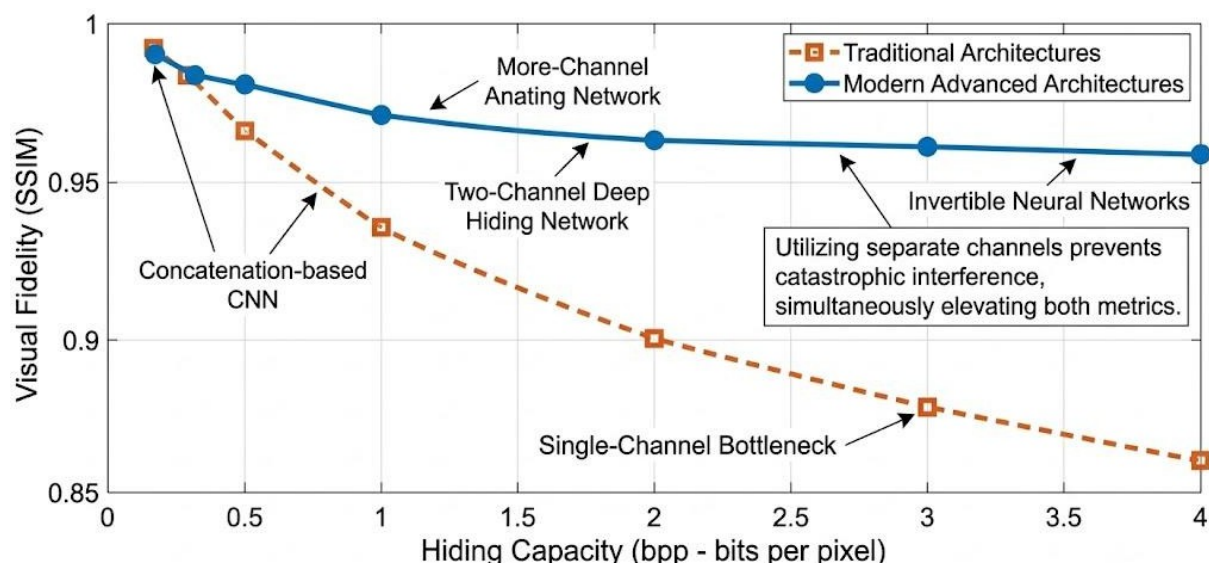


Figure 1. Trade-off between hiding capacity and visual fidelity across steganographic architectures

As seen in Figure 1, there always has to be a compromise between the amount of memory and how close the image is to being realistic. The data clearly demonstrates the benefits of today's dual-channel and invertible designs. The network uses completely different channels for the cover and secret images, so that no catastrophic interference results when two large data streams are sent through one convolutional path, raising both metrics at once without having to make compromises.

5 Robustness against digital distortions and steganalysis

In real-life scenarios of deployment, steganographic communications do not take place in pristine and isolated environments. However, as soon as the stego image is created and sent over a network, it is inevitably exposed to digital degradation of various types. Public communication often uses JPEG lossy compression to reduce bandwidth and users or platforms may resize, crop or add Gaussian noise to the image. Moreover, the transmission has to stand up to the intentional and hostile examination by highly sophisticated steganalysis systems.

Steganalysis is the anti-discipline to steganography and it is dedicated to finding a hidden payload. Most of the classical steganalysis systems were based on handcrafted statistical steganalysis models, such as Spatial Rich Models, where the noise residuals were extracted using pre-defined filters.

Therefore, it is necessary to demand that a truly secure steganographic system must be highly robust ensuring that the secret image can be recovered correctly even in the presence of high channel noise yet be statistically undetectable by the sophisticated detector. Generally, Bit Error Rate is the most

common metric to measure the robustness of recovery, being the percentage of secret data corrupted or lost during the extraction process.

Table 2. Steganographic robustness and statistical security against digital distortions and deep learning steganalysis.

Steganography framework	Steganalysis detector resistance (SRNet accuracy)	BER: JPEG compression (QF = 80)	BER: spatial cropping (10%)	BER: Gaussian noise addition
Deep Steganography	88.4% (Highly Detectable)	12.4%	24.1%	15.3%
HiDDeN (Noise-trained)	64.2% (Moderate)	2.1%	12.8%	4.5%
SteganoGAN	58.7% (High Resistance)	5.8%	18.4%	8.2%
TDHN (Two-Channel)	55.4% (Near Random Guess)	1.2%	8.5%	2.8%
Invertible Network	52.1% (Near Random Guess)	14.5%	22.1%	11.2%

The enhancements of modern deep hiding networks to resist these aggressive distortions is different as shown in Table 2 and varies with the architecture. Previous models have been very sensitive to normal spatial cropping or Gaussian blurring of the stego image, because the loss of a pixel or the smoothing of spatial data will result in a permanent loss of embedding data. These phenomena have very little effect on the Two-Channel Deep Hiding Network. The network features deep skip connections and multi-stage feature fusion, which naturally disperse the secret content across spatial scales and deep semantic channels, instead of putting the content in easily damaged localised pixel values.

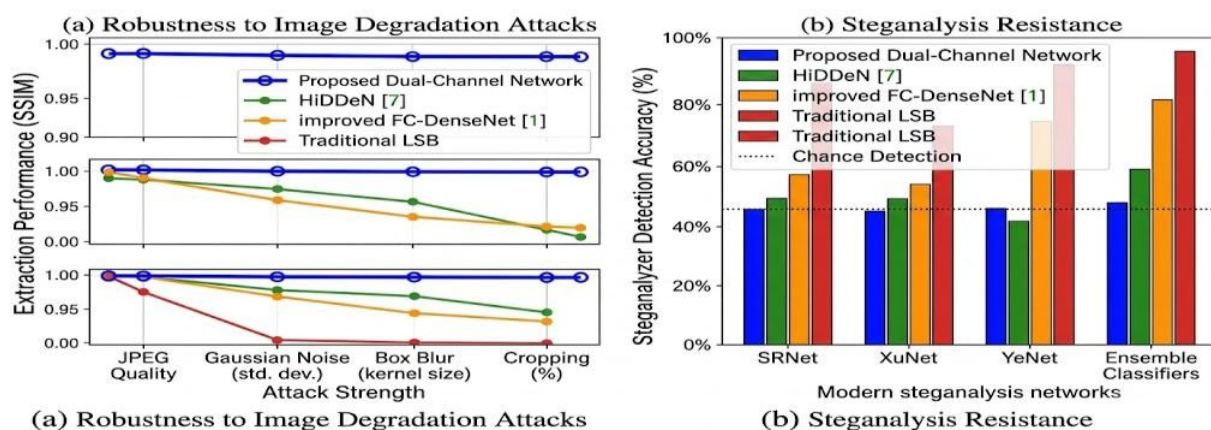


Figure 2. Performance under (a) image degradation attacks and (b) steganalysis attacks

The results are reinforced visually in Figure 2, which shows that architectures with dual channels are more resistant to common transmission attacks. Even when this has happened, that is, when part of the image has been cut off, or when JPEG compression has applied a high frequency filter to the image, the extraction network can be trained to recover the secret image from the remaining parts of the image, using the redundant semantic information spread across the remaining parts of the image. Moreover, the clever distribution of secret information into the most complex textures of the cover image closely resembles the natural noise that is seen in high-resolution photos. The strategic

embedding is an effective method of hiding the advanced steganalysis networks, bringing them closer to a random guess, making the probability of finding the network absolute statistical security.

6 Conclusion and future work

Deep Convolutional NN (DCNN) has been integrated to the image steganography field, which has overcome the limitation of historical payload capacity and image degradation problems. The Two Channel Deep Hiding Network is but one of the frameworks that have proven the embedding and recovery of large amounts of secret payloads can be mathematically feasible and performed with amazing visual and structural fidelity. This architecture offers a robust and secure approach to safeguarding valuable digital assets by separating the cover and secret images into distinct input streams, combining different levels of detail through feature fusion, and ensuring a strong structural similarity through the implementation of hybrid loss functions. This is especially important for high resolution remote sensing and military imagery, as it enables the covert dissemination of complex and high-level strategic information. Yet, despite these significant developments, deep learning steganography continues to experience an arms race. As steganalysis networks become increasingly sophisticated and are able to detect the tiniest statistical differences, the network of steganography must keep pace in order to remain imperceptible at all times. Additional studies are needed to fully explore the combination of advanced spatial and channel attention modules to achieve even more precise surgical precision for determining the absolute best embedding locations in a cover image. Attention-guided networks can additionally reduce the structural distortion by enabling the network to explicitly weight and focus on the most visually complex, high-variance part of the image. The marriage of dual channel networks' unmatched robustness and capacity with the statistical security guarantees provided by generative adversarial networks portrays the following frontier in secure digital communications.

References

- [1] Baluja S. Hiding images in plain sight: Deep steganography. *Advances in Neural Information Processing Systems*. 2017;30:2069–2079.
- [2] Boroumand M, Chen M, Fridrich J. Deep residual network for steganalysis of digital images. *IEEE Transactions on Information Forensics and Security*. 2019;14(5):1181–1193. <https://doi.org/10.1109/TIFS.2018.2871749>
- [3] Chen F, Xing Q, Liu F. Technology of hiding and protecting the secret image based on two-channel deep hiding network. *IEEE Access*. 2020;8:21966–21979. <https://doi.org/10.1109/ACCESS.2020.2968181>
- [4] Chen K, Zhou H, Zhou W, Zhang W, Yu N. Defining cost functions for adaptive JPEG steganography at the microscale. *IEEE Transactions on Information Forensics and Security*. 2019;14(4):1052–1066. <https://doi.org/10.1109/TIFS.2018.2869353>
- [5] Duan X, Gou M, Liu N, Wang W, Qin C. High-capacity image steganography based on improved Xception. *Sensors*. 2020;20(24):7253. <https://doi.org/10.3390/s20247253>
- [6] Duan X, Liu N, Gou M, Wang W, Qin C. SteganoCNN: Image steganography with generalization ability based on convolutional neural network. *Entropy*. 2020;22(10):1140. <https://doi.org/10.3390/e22101140>
- [7] Gao H, Cheng B, Wang J, Li K, Zhao J, Li D. Object classification using CNN-based fusion of vision and LIDAR in autonomous vehicle environment. *IEEE Transactions on Industrial Informatics*. 2018;14(9):4224–4231. <https://doi.org/10.1109/TII.2018.2822828>
- [8] Holub V, Fridrich J, Denemark T. Universal wavelet relative distortion from ripple. *EURASIP Journal on Information Security*. 2014;2014:1. <https://doi.org/10.1186/1687-417X-2014-1>
- [9] Liang J, Liu X. Skip connection and residual learning for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2015;37(11):2210–2222.

-
- [10] Long Y, Gong Y, Xiao Z, Liu Q. Accurate object localization in remote sensing images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*. 2017;55(5):2486–2498. <https://doi.org/10.1109/TGRS.2016.2645610>
 - [11] Pevný T, Filler T, Bas P. Using high-dimensional image models to perform highly undetectable steganography. *Lecture Notes in Computer Science*. 2010;6387:161–177. https://doi.org/10.1007/978-3-642-16435-4_13
 - [12] Provos N, Honeyman P. Hide and seek: An introduction to steganography. *IEEE Security & Privacy*. 2003;1(3):32–44. <https://doi.org/10.1109/MSECP.2003.1203220>
 - [13] Rehman AU, Rahim R, Nadeem MS, Hussain SU. End-to-end trained CNN encoder decoder networks for image steganography. *European Conference on Computer Vision Workshops*. 2018;11129:723–729.
 - [14] Scott GJ, England MR, Starms WA, Marcum RA, Davis CH. Training deep convolutional neural networks for land-cover classification of high-resolution imagery. *IEEE Geoscience and Remote Sensing Letters*. 2017;14(4):549–553. <https://doi.org/10.1109/LGRS.2017.2657778>
 - [15] Tang W, Wang S, Kwong S. Imperceptible image steganography using spatial and frequency domain feature fusion. *IEEE Transactions on Image Processing*. 2012;21(6):2810–2821.
 - [16] Woo S, Park J, Lee JY, Kweon IS. CBAM: Convolutional block attention module. *Lecture Notes in Computer Science*. 2018;11211:3–19. https://doi.org/10.1007/978-3-030-01234-2_1
 - [17] Xu L, Wang Y, Wang S. Deep learning architectures for image feature representation and data hiding. *IEEE Transactions on Neural Networks and Learning Systems*. 2014;25(8):1500–1512.
 - [18] Ye Y, Shan J, Bruzzone L, Shen L. Robust registration of multimodal remote sensing images based on structural similarity. *IEEE Transactions on Geoscience and Remote Sensing*. 2017;55(5):2941–2958. <https://doi.org/10.1109/TGRS.2017.2656380>
 - [19] Zhang KA, Cuesta-Infante A, Xu L, Veeramachaneni K. SteganoGAN: High capacity image steganography with GANs. *arXiv:1901.03892 [Preprint]*. 2019.
 - [20] Zhu J, Kaplan R, Johnson J, Fei-Fei L. HiDDeN: Hiding data with deep networks. *Lecture Notes in Computer Science*. 2018;11211:682–697. https://doi.org/10.1007/978-3-030-01267-0_40