A Survey on Deep Convolutional Neural Networks for Image Steganography and Steganalysis

Israr Hussain, Jishen Zeng, Xinhong, Shunquan Tan*

Shenzhen Key Laboratory of Media Security, College of Electronic & Information Engineering, Shenzhen University, Shenzhen 518060

israrhusain@szu.edu.cn, jishenzeng@hotmail.com, 2156130112@email.szu.edu.cn, tansq@szu.edu.cn,

*Corresponding author: Shunquan Tan

Received August 10, 2019; revised October 12, 2019; accepted December 3, 2019; published March 31, 2020

Abstract

Steganalysis & steganography have witnessed immense progress over the past few years by the advancement of deep convolutional neural networks (DCNN). In this paper, we analyzed current research states from the latest image steganography and steganalysis frameworks based on deep learning. Our objective is to provide for future researchers the work being done on deep learning-based image steganography & steganalysis and highlights the strengths and weakness of existing up-to-date techniques. The result of this study opens new approaches for upcoming research and may serve as source of hypothesis for further significant research on deep learning-based image steganography and steganalysis. Finally, technical challenges of current methods and several promising directions on deep learning steganography and steganalysis are suggested to illustrate how these challenges can be transferred into prolific future research avenues.

Keywords: Steganalysis, Steganography, Deep learning, Convolutional Neural Network

This work was supported in part by NSFC (Grant 61772349, U1636202, 61872244, 61572329) and in part by Guangdong R & D Program in Key Areas (2019B010139003), Guangdong Basic and Applied Basic Research Foundation (Grant 2019B151502001) and Shenzhen R&D Program (Grant GJHZ20180928155814437, JCYJ20180305124325555).

http://doi.org/10.3837/tiis.2020.03.017 **ISSN : 1976-7277**

1. Introduction

Nowadays, cyber security threats become a challenge task around protecting confidential information over the open networks. There is a risk that confidential information can be access by attackers, being conferred, revealed, revised and affecting its accessibility. Obviously, the information that transmit over an insecure channel can be easily manipulated therefore steganography becomes one of the most popular research areas because of easy multimedia communication through various low-cost devices such as smart-phones, IP cameras & Social Media Apps like Facebook, Instagram, WhatsApp, WeChat, QQ, Twitter. The ability to secure the information from adversaries during transmission over internet that are opposed to be conferred, is crucial in a world of cyber warfare emerging risks [1]. We do believe that secure steganographic algorithms [2], [3], might be helpful to tackle organized cyber-crimes and new secure steganographic schemes should be set to train and prepare our security institutions into overcoming upcoming threats and problems that may occur in IT World. Steganalysis & steganography play hide and seek game [4]. Because of the advancement in deep learning steganalytic algorithms [5], [6], [7], [8], [29], [30] the task of designing more reliable & robust steganographic framework becoming more and more imperative. Image steganography and steganalysis received a lot of attention from law enforcement agencies and social media due to easy of multimedia communication through the internet. Image steganography is the most prominent type of carrier because of fast and easy to send confidential information. Image steganography frameworks mainly classified into two categories spatial domain frameworks & transform domain frameworks. In spatial domain frameworks, we directly deal with image pixel values. The pixel values are modified to achieved desired goal whereas transform domain frameworks work on transform domain coefficient that are obtained. Recently most of the steganographic algorithms accessible on the internet can conceal the secret information in digital images are save it in different image formats such as PNG, BMP, TIFF, and the lossy JPEG.

Early steganographic algorithms try to increase the invisibility of secure information by decreasing the quantity of embedding changes in images [9], [10]. However, it's not enough to assurance the security performance since the strong interrelation in natural image. At present high dimensional feature based on complicated correlation of image neighborhood achieved significant improvements to model's accuracy such as Spatial Rich Model (SRM) / SRMQ1 [24], Projected Spatial Rich Model [27]. Rich model features belong to high dimensional steganalysis features, which imitate image distortion introduce by embedding a secret message into an image. SRM high dimensional features achieved high level of detection accuracy for the adaptive steganography embedding methods. However, these high dimensional features imposed large computational cost on feature extraction and related classification training. Relatively few researchers have tried to reduce the dimension of steganalysis features [76], [77] and careful research in this area is still needed. An alternative approach to overcome these problems, an existing steganalytic methods assimilate selection-channel-information [78], [79] into the steganalytic features to improve detection performance of the algorithms. Selection-channel-aware based steganalysis approaches effectively detecting content-adaptive steganography schemes. Yang et al. [30] and Ye et al. [28] proposed methods to integrate the

selection channel information in their networks. But designed this kind of features take a lot of time and efforts. Moreover, it will reliably become tough with the continues development of more secure deep learning-based steganographic framework. To insure the best security and invisibility of secret message some approches [26] used object detection methods such as Faster R-CNN [37], R-FCNN [42] and SSD [82] to select a complex texture region of the image, which is suitable for hiding information. After obtaining the most complex texture area, steganographic algoritms WOW, HUGO & S-Uniward is used to embed secret information in the selected area of the image to obtain stego image. Fater R-CNN introduce the RPN (region proposal network) based on Fast-R-CNN replacing the slow search slective search algotitms. Faster CNN applies the RPN network to accelerate the speed generation of independent region proposals. Along with the advancement of deep learning based steganography, deep learning steganalysis framework have also been proposed and achieved series of successful approaches [5], [6], [7], [28], [35], [36]. After many successful and approved studies there are still plenty of practical applications and opportunities ahead that need to study further for real world applications using deep learning-based techniques.

In this paper, we present a comprehensive review of current deep learning-based image steganography and steganalysis frameworks. We examined tactical foundation of the techniques to analyze their performances, strengths and limitations. We also provided comparison between approved steganography and steganalysis studies and embedding algorithms to illustrate how these challenges can be transformed into prolific future research avenues. Current research trends and directions are suggested by current state of our knowledge.

The remaining paper is organized as follows. In Section II, we provide comprehensive review of existing up to date deep learning based steganalysis approaches and highlights their pros and cons. Comparison were also made for different steganalysis approaches depend on deep learning tricks and domain knowledge. In Section III, we briefly discuss deep learning-based image steganography, GANs and adversarial examples based steganographic techniques and discuss their pros and cons. Section IV present existing challenges and outline future research directions for researcher. Finally, conclusion of this paper is presented in Section V.

2. Image Steganalysis Based on Deep Learning

Deep learning frameworks achieved excellent performance in many fields [75], [80]. Researchers in image steganography and steganalysis also demonstrated to explore the capability of deep learning algorithms in various key areas of multimedia security. The design of steganalyzers based on convolutional neural networks, especially deep learning has achieved amazing performance. Absolutely, a deep learning based steganalyzer allows us to automatic feature extraction and classification steps in a distinctive network architecture, beyond any prior feature selection. Inspired by successful approaches based on CNNs, it's has fascinated the attention of many scholars and made great development. After all, recent development on deep leaning steganalysis is still facing numerous challenges. Therefore, its indeed of studying these issues by applying deep learning techniques. Several issues related to existing deep learning-based algorithms are presented in table **Table 1**.

2.1 Spatial Steganalysis Methods

In spatial domain steganography the secret message is hide inside the image by applying some manipulation over different pixel values of the cover image, which effect the statistical characteristic of an image. This section analyzed various deep learning-based approaches in spatial domain steganalysis and highlighted the contribution of individual work, comparison were made of different machine learning approaches.

In 2014 Tan & Li et al. [20], proposed first CNN structure for steganalysis of digital images in spatial domain. Although proposed model is performing better than the SPAM [27], but still inferior to SRM [24]. The network is not enough deeper with only three convolutional layers and quite slow because of too large fully connected layer. Their proposed network acquired error rate of 48% with random parameter initialization against detecting HUGO [13] at embedding rate of 0.4bpp. At the beginning of 2015 Qian et al. [5], designed a customized CNN network for steganalysis which obtained comparable performance to Spatial Rich Model [24] and additionally justify that transfer learning is favorable to detect steganographic frameworks with low embedding rate. The role of high pass filter in their network is to conceal the image content and improve the SNR between the stego signal and cover image. The author noted that without high pass filter network not converging well. One year later, Pibre et al. [38], examine Qian's work and obtained better detection accuracy in the scenario of reusing the same embedding key for different images. Experiment results demonstrated that proposed network is able to decrease the detection performance 16% in comparison with the SRM-base steganalyzers. But the network achieved worse performance when considering a unlike key for each embedding. Salomon et al. [40] proposed another deep learning framework for steganalysis. Correspond to Qian et al. [5], saloman uses only two convolutional layers in their network & increase the number of activation map in each convolutional layer and removed the pooling layer that is disadvantage for the subsequent steganalysis operation because of smoothing the noise. In 2016, Xu-NetV1 et al. [35], proposed the first deep learning framework achieved competitive performance compared with SRM [24]. In their proposed network, they used an absolute ABS activation layer for feature map generated from first convolutional layer. It can learn more resultful features that might be helpful to avoid overfitting problems. They also used BN (batch normalization) and pooling layers in their network and achieved better accuracy to SRM [24]. Xu-NetV2 et al. [22], the author extended his previous work [35] by adding one more group of layers called "group 6" at the end of the convolutional module in their network and increase the max pooling kernel size of last two pooling layers from 5×5 to 7×7. The aim of proposed network is to introduce ensemble learning rather than develop CNN frameworks. Another successful approach by Wu et al. [41], introduce novel normalization technique called "share normalization" in their network to share statistic properties during training and testing of the network. This way can train the network effectively by squeeze the weak stego signal of the image. Compare with SRM and maxSRM the proposed network shows obvious performance on different steganographic algorithms at various embedding rates. Proposed algorithm also achieved superior performance compared with Xu-Net et al. [22] & Qian Net et al. [5]. In 2017 Ye-Net et al. [28], firstly integrate the truncation approach into the design of steganalysis CNN networks. By applying selection-channel-information & data augmentation techniques, proposed framework achieved superior performance than classic SRM on re-sample and cropped image dataset. The network also specify the significance of using larger amount of data samples for deeper networks and the benefits of alternative adaptive optimizer specially AdaDelta gradient decent variant. In 2018 ReSt-Net et al. [43] explore another success approach by mean of diverse activation modules & parallel subnet-based CNN for spatial steganalysis. Their architecture

consist of (DAMs) diverse activation modules, which activate the convolution outputs differently and then combine their outputs for the following layers. The network used more sub-nets with less quantity of filters therefore it is more productive than increase the number of filters for preprocessing layers. To stimulate the training process, pre-trained the subnets separately. The network perform better than Xu-Net, Ye-Net (without SCA). In 2017, Sedighi-Net et al**.** [44], implement specific CNN layer to replicate rich steganalytic frameworks but could not be able of achieved state-of-the-art performance. Special activation function Mean-shifted Gaussian have been used in their network. It performs slightly better than PSRM [27]. Yedroudj-Net et al.[23], proposed CNN framework that outperforms in term of the error probability. Experiments were performed to show its supremacy with other state-of the art framework like Xu-net [35], & Ye-Net [28] in its not informed version and to Ensemble Classifier fed by the Spatial Rich Model [33]. In 2018, Zhang et al. [10], proposed an adequate feature learning & multi-size image steganalysis framework based on CNN called Zhu-net and the proposed network achieved better detection performance compared to Yedrouudj-Net [23]. Zhu-Net offer three improvement in Yedroudj-Net that are the renovate the kernel filters of pre-processinglayer, secondly replace the first two convolutional blocks with two module of depth-wise detachable convolutions that can extract the spatial and channel correlation of residuals to increase SNR and obviously improve the accuracy, finally replace the gobal pooling with spatial pyramid pooling to deal with arbitrary images. Recently in 2019, LU JICANG et al. [45], proposed an improved steganalysis framework based on feature selection & pre-classification techniques. First the author apply k-means algorithm to image dataset to extract images with different texture and complexities then optimal features for each cluster are selected for final decision which might improve overall performance of the stenanalysis schemes. Another milestone in the same year by Zeng et al. [46], proposed a color image steganalyzer called wider separate-then-reunion network (WISER-Net). They split the color image into their corresponding bands then initialize every band with 30 high pass SRM filter. They claim that the aggregation in normal convolution is one kind of “linear collusion attacks” which is the process of convex combination of input color bands. It retains complex correlation pattern whereas reduce uncorrelated noise. In this section we discussed all deep learning frameworks based on spatial domain. We analyzed that the accuracy of deep learning frameworks can be further improve through proper network design, fusion and learning strategy, deep learning tricks and incorporating prior domain knowledge into the CNN architecture. Researcher in steganalysis achieved many successful results by applying deep learning techniques in their networks. However, many challenges remain & they are currently not addressed very well. Some of the existing challenges that need further study are presented in **Table 1**. Evolution of deep learning based steganalysis framework are presented in **Fig. 1**.

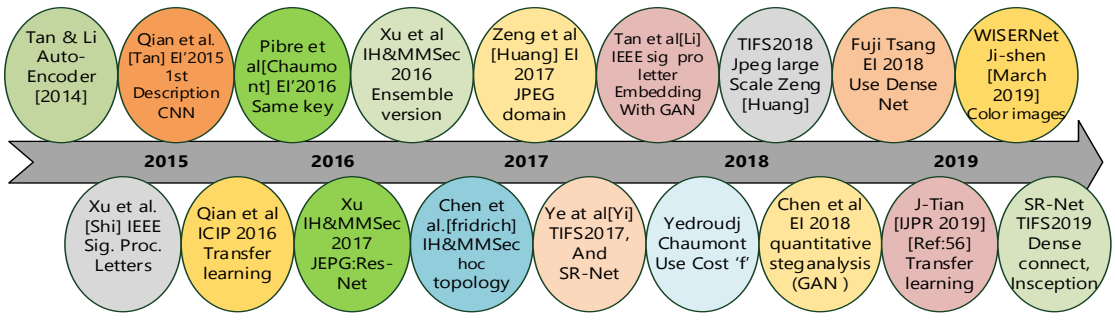


Fig. 1. Evolutionary development of deep learning based steganalysis framework

Table 1. Pros and Cons of Different Steganalysis Algorithms in JPEG and Spatial Domain

Algorithms	Pros	Cons	Activation Functions
Tan-Net [20]	First CNN Structure for Steganalysis. The resemblance between SRM and CNN. Comparable to SPAM.	Not enough deeper only 3 Conv layers & slower since FC layers is too lager. Avg pooling is better than max pooling but slower.	Sigmoid
Qian-Net [5]	Resemblance b/w steganalysis & GCNN. Better than SPAM and introduce Gaussian Activation function and HPF layer.	Performance is worse than SRM. Not enough Deeper. In the absence of high pass filter their network not converging.	ReLU &TanH
Xu-Net-V1[35]	Generate noise residuals enhance the detection performance of the CNNs. Used HPF layer same as in Qian net. Used 5 groups of Conv layers and 5×5 avg pooling.	Not deeper enough, So the performance is not good enough.	ReLU &TanH
Xu-Net-V2 [22]	7 × 7 Pooling and 6 groups. Ensemble of sub-models. Studying strategies of ensemble learning.	Long training time about three days to run all the experiments.	ReLU & TanH
Ye-Net [28]	First CNN with selection channel information. Introduce AUG importance in CNNs for steganalysis.	Deep enough but Slow training	ReLU & TLU
Sedighi-Net [44]	Performs slightly better than PSRM. Featured based steganalysis. Histogram layer.	Poor network performance due to some limitations in modeling but proof of concept.	ReLU
ReSt-Net [43]	Better than Xu-Net and Ye-Net without SCA. Gabor, SRM linear and SRM nonlinear filters were used.	Used wider Structure. Deeper & slower	ReLU, Sigmoid TanH
Zeng-Net [7]	Hybrid deep learning model for large JPEG image steganalysis. Used 5 × 5 kernels and less parameters than Xu-net. Got better performance than DCTR, PHARM, Xu-Net in term of accuracy.	Quantization and Truncation is not learn-able. Without truncation the CNNs experienced slow convergence.	Quantization / Truncation
Chen-Net [47]	Modified Xu-Net by porting the concept of JPEG phase-awareness and proposed P-Net and V-Net. Used Katalyst Kernel and two directional Gabor filter. Implement a new phase-split layer.	P-Net gives a slightly better performance but longer training time and more complex.	ReLU& TanH
Xu-Net-V3 [21]	Decompress to spatial domain (without rounding). Deeper network with 20 layers, Res-Net structure, 4×4 DCT pre-processing (as in Zeng).	Much deeper and more complicated. Quantization is no need when the net-work is enough deep.	ReLU
Yang-Net [48]	Used 44 DCT preprocessing (as Zeng-Net), also used BN. Performed better than Xu-Net. 32 layers Dense Net Structure.	Deeper but Slower.	ReLU
SR-Net [8]	Deep residual architecture. Minimize the use of heuristic & externally enforced element. Work well for both JPEG and spatial domain. Selection channel as 2nd channel.	Dense connection doesn't provide satisfactory. The Network is very slow as compare to Xu-Net.	ReLU
Zhu-Net [55]	Used two separable blocks to replace traditional convolutional layers. Spatial pyramid to deal with arbitrary sized images.	Larger Conv kernel size may Loss a lot of details that can lead you to under-fitting.	ReLU & TLU
Wu-Net [49]	Larger depth proves to be efficient to capture the statistic of images. Used residual learning and residual learning in deep residual network (DRN) retain the stego signal from secret messages.	BN was not correctly used, Deeper but Slower.	ReLU
Yedroudj-Net [23]	Used 5 groups of Conv layers same as in Xu-Net1. Used BN & ABS layers. 30 filters bank for preprocessing layer same as in Ye-Net. It's better than Xu-Net and Ye-Net	Used 3 FC layers increase model complexity and make it slower conversion.	ReLU
Pitfalls-Net [38]	Detects steganographic algorithms both in spatial and frequency domain with low payloads. 64 filters of size 509×509 are used in Conv2.	Results can be obtained only on dataset provided by the authors.	TanH
Mo-Chen-Net [36]	Proposed deep learning regressor for quantitative steganalysis both spatial and JPEG domain. Robust payload estimator	Large Fully connected layers, excessive number of parameters it become prone to over-fitting.	TanH

WISER-Net [46]	Steganalysis framework for JPEG color images. Introduce channel-wise convolution in their network. Obtained superior performance.	Complex course source miss-matching problem.	ReLU

2.2 JPEG Image Steganalysis

JPEG is most prominent and extensively used image file format to save & transmitting digital images over the world wide web. Because it can be compressed to one eight of its original size & still contain good visible quality. In JPEG domain methods, secret messages are inserted by modifying coefficient values after transformation such as DCT, DWT and DFT. There exist many JPEG steganalysis deep learning frameworks, some of them are either unreliable or time consuming. Significant research progress was made in JPEG steganalysis by designing appropriate network models or introduce phase aware concept into the CNN networks architectures. It is a big challenge to uncover weak concealed information in a JPEG compressed image. In the following section we will cover several deep learning algorithms that hide information in jpeg domain & will also highlight pros and cons.

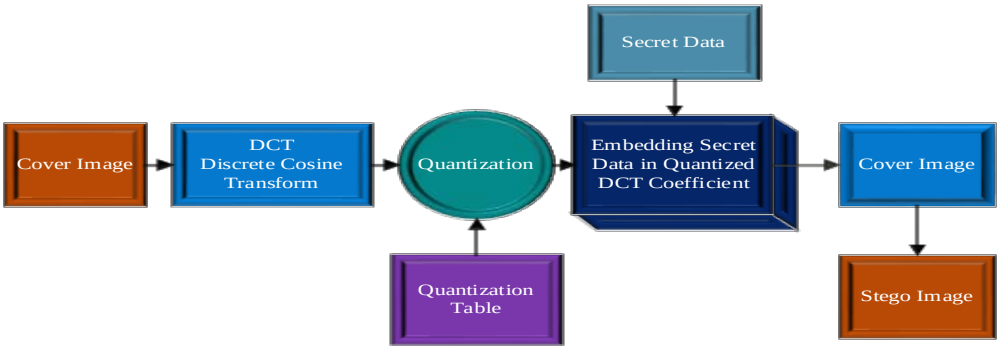


Fig. 2. Block Diagram of Embedding process of JPEG Steganography

In 2017, Zeng et al. [51], proposed the first deep learning based steganalysis framework with a pre-processing block at input encouraged by Rich Models [33]. Proposed network can obtain remarkable performance boost compared to DCTR [39], but still inferior to PHARM. In the same year Xu et al. [21], construct a deep learning neural network with 20-layers for JPEG steganalysis, strongly inspired by Res-Net [52] with shortcut connection tricks [21]. Proposed network replaced pooling layers with convolutional layers also improved the result in terms of accuracy. It cut the error rate to 35% achieved by Zeng et al. [7] for large scale JPEG steganalysis. Later, Chen et al. [47], introduce a deep learning framework with phase-split concept inspired by JPEG compression algorithm. The network is modified from Xu-Net [21] by dividing the activation maps into 64 parallel channels to port jpeg phase aware in their network. In their network they introduced two ways for incorporating phase awareness within the network architecture which is P-Net and V-Net. The experimental results exhibit that the proposed CNN structure is performing superior to SCA-GFR on J-UNIWARD and UED. Yang-Net et al. [48], proposed a deeper 32-layers CNN framework with feature reuse technique by integrating all features from the prior layers as a result improve gradient & flow of information. Bottleneck layers and shared features in

their network further boost feature propagation and reduce the model parameters dramatically. Experiment results shows that the proposed architecture can reduce the error rate 5.67% for 0.2bpnzAC and 4.41% for 0.4bpnzAC, while the number of training parameters in their framework is only 17% of what used by in Xu-NetV3 [21]. In 2017 Wu et al. [49], proposed deep residual framework that has two main differences with existing network. First, proposed network is deeper than the existing networks which prove to be more productive to capture the statistical feature of digital images. Secondly residual learning is used to actively preserve stego signal coming from stego images which is extremely beneficial for fumigate of stego and cover images. One important point that we have noted in this network is the batch normalization (BN) is not used correctly. Huang et al. [2018], proposed a variant of Xu-Net [21] called ResDet to detect adaptive JPEG steganography with close results. Shortcut connections inspired by Xu-net [21] also adapted in their network to overcome gradient vanishing problem. The network performed better than [21], [47] with high embedding rate. In the same year, Zhong et al. [53], proposed another successful framework for steganalysis of jpeg images depend on filter diversity section. In their network the author initiates three ensemble methods intend to increase the diversity between classifier. Another milestone for jpeg image steganalysis who have made significant contribution by Zeng et al. [7], for hybrid deep neural network models. Proposed Network includes two principal phases: The first phase is synthesized phase, analogous to the convolution & quantization truncation phase of SRM [33] and the secondary phase hold a composite neural network, which learn parameters of the network during training process. In their Network they also used three sub-nets inspired by Xu-Net [35]. Proposed framework that first time deploy QT (quantization & truncation) into deep learning based steganalyers. It is less efficient that Xu-NetV3 but give first compelling approach reconcile by Rich models. Its performance is better than DCTR, PHARM, Xu-NetV1 on J-UNIWARD, UED, UERD. Tsang, et al. [54], used a modified Ye-Net [28] without the selection channel-aware part. One of the modifications is the addition of a BN (batch normalization) after each ReLU that might helpful to prevents the network from overfitting and gives a slightly better detection accuracy. Secondly, the author reduces the stride of 9^{th} convolutional layer before classification to one, which made the size of the 16 features before the IP layer to be 7×7 rather than 3×3 as in the original Ye-Net. Mo Chen et al. [36], proposed CNN based regressor for both jpeg and spatial image steganalysis. The design called bucket estimator, starts by training a family of CNN detectors, each detector for a fixed payload and then using their concatenated feature maps as a feature on which a fully connected network (regressor) is train by using the Mean Square Error (MSE) as the loss function, the design come out as the best performer among other natural choices. Best CNN framework for JPEG as well as spatial steganalysis at the end of 2018 is SR-Net that has been proposed with side-channel-information [8]. The network is corresponding to the combination of convolutional blocks beyond the pooling layer immediately after the first convolution block of the Yedrouj- Net [23]. Essential part of SR-Net is noise residual extraction section consist of first seven layers. The Network is very slow as compare to Xu-NetV3 it takes 20 hours to train the network, while SR-Net takes 1 week. The network was proposed in 2018 but the paper have been published in May 2019. HU DONGHUI et al. [2019], proposed a new self-seeking steganalysis framework depend on deep reinforcement leaning & visual attention [25] to recognize JPEG based adaptive steganographic algorithms. Their method change image into AFRs by visual attention scheme after that makes repeated judgment by reinforcement learning to choose AFRs that are more suitable for steganalysis. Researchers

achieved significant improvement in detecting jpeg-based steganography algorithms by applying different deep learning-based strategies in their frameworks. Comparison between

various frameworks based on deep learning tricks and domain knowledge are presented in **Table 2**.

Table 2. Comparison of Different Steganalysis Frameworks based on Deep Learning

		Tan Net [20]	Qian Net [5]	Xu Net-v1 [35]	Xu-Net v2 [22]	Ye-Net [28]	Se-Net [44]	Rest-Net [43]	Zeng-Net [7]	Chen Net [47]	Xu-Net-v3 [21]	Yang-Net [48]	SR-Net [8]	Yedrdj Net [14]	Wu Net [49]	Wiser-Net [46]	Hu Net [25]	Tian Net [50]
Domain Knowl-edge	Conv & fully Connected	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
	Pair-wised training	√	√	√	√	√	√	√	√	√	√	√	√	√			√	
	Special activation function		√															
	Pre-processing layer		√	√	√	√		√	√	√	√	√		√		√	√	√
	Special Layer			√	√	√												√
	Quantization /Trunca-tion								√		√			√			√	
	Selection Channel					√					√							
	Diversity /Ensemble				√			√				√						
	Undecimated DCT								√	√	√	√	√				√	
	JPEG phase Aware									√								
	Channel-Wise Conv															√		
	Data Pre-Classification																√	
	Feature Selection																√	
Deep learning Tricks	Deep		√	√	√	√			√	√	√	√	√	√			√	
	Wide							√	√			√						
	BN			√	√			√	√	√	√	√	√	√		√	√	
	Residual										√		√				√	
	Dense											√						
	Drop out													√				
	Data Augmentation					√								√				
	ABS				√		√	√		√				√		√		
	Transfer Learning																	√

3. Deep Learning Based Image Steganography

Deep neural network approach to embed secret information inside the digital images ensuring the secure steganography. In contrast to previous traditional studies many deep learning-based frameworks have been developed successfully that allow researchers to hide large information inside the natural images without possible eavesdropper [56-59], [72-74]. The primarily application of deep learning to steganography was GAN based steganography [56]. Zhang et al. [58] proposed a GAN based deep learning framework for hiding arbitrary binary data in digital images and their proposed network achieved state-of-the-art payloads of 4.4bpp. Later, Saleema [59] proposed another milestone to refine the embedding image generated by mean of traditional steganography schemes. Volkhonskiys [60] & Shis [34] also concentrated on generating secure cover images for traditional steganography algorithms. In 2018 Steg et al. [32] proposed deep learning framework for image steganography to embed secret information without any involvement of traditional steganographic frameworks. Balujaet al. [33] & Steg et al. [32] both doing the same job. Although, the concealed image is a bit detectable on residual images of the generated embedded images. Furthermore, proposed network uses three sub-networks which requires more computations & GPU memory and it also takes more time to hide the secret information. Recently in 2019 Duan et al. [57] proposed a new reversible steganography framework based on U-Net structure. Their network composed of two networks, hiding network called U-Net and an extraction network. Extraction network subsist of six convolutional layers with filter size of 3×3. These successful approaches by using deep learning for steganography draw the attention of more scholars and achieved great progress but there are still some limitations with exiting methods are presented in **Table 3** that we need to study further using different deep learning approaches.

3.1 Adversarial Examples in Deep Learning Based Steganography

Deep neural networks are surprisingly susceptible to a small perturbation called adversarial examples. Adversarial examples are input to a neural network yield inaccurate output from the networks [19]. At the same time the presence of adversarial examples is generally seen as destructive for neural network, but in another side, it can be adorable for information hiding. Papernot et al. [61] explored that adversarial examples develop for network can be transfer to another network. It also exploits that adversarial examples are vigorous toward image transformation, when adversarial examples are printed or either photographed, the model still miss distinguish the photo. Szegedy & Goodfellow's [19] made a formative work and propose a method for adversarial example construction depend on neural network gradient. Traditional approaches to digital image steganography are only efficient only for appropriate payload of 0.4bpp. Beyond this limit, they produce artifacts that can be easily detected by deep learning based steganalysis framework and, in extreme cases, it can also be detected by the human eyes. Advances in deep learning over the last few decades a new class of image steganographic frameworks based on deep learning are emerging [62- 65]. In 2018, Zhu et al. [66] proposed end to end deep learning framework for data hiding in digital images. Compared to Hayes et al. [67] proposed method uses only convolutional networks which greatly improved the image quality closer to cover image and achieved lower error rate. In the same year Sai Ma et al. [68] has proposed another CNN based method to generate steganographic adversarial examples in order to enhance the steganographic security of exiting algorithms. These adversarial examples increase the detection error of CNN based steganalyzers. While embedding proposed method exploits the gradient feature map to determine the flipping direction of the

pixels, flipping is equivalent to 1. After SaiMa et al. [68] another successful approach by Wu et al. [32], to apply neural networks for image steganography to embed secret information in images aside any influence or involvement of traditional steganography algorithms. The proposed network boost image steganography embedding rate to an average of 23.57bpp (bit per pixel) by modify only around of 0.76% of input cover image. Zhang et al. [64], used a unique scheme to generate adversarial samples for steganographic algorithms, proposed framework first appends the adversarial noise to the cover image to construct a booming enhanced cover images then embed the secret information in it to generate stego image. In this scenario the receiver can recover the original message from cover image successfully. Anyhow, during the construction of adversarial samples modification would be introduce to the image automatically. Recently in 2019 Kevin et al. [58], proposed a novel technique for concealing arbitrary binary data in digital images using GANs (generative adversarial networks) based approach which allows us to enhance the perceptual quality of the images produce by the network. The proposed techniques achieve a relatively payload of 4.4bpp. The key differences between the proposed scheme and Zhu et al. [27] approach is the loss functions used to train the model, the architecture of the model, and how data is presented to the network. Tang et al. [69], proposed CNN-based adversarial steganographic embedding scheme with a new achievement called (ADV-EMB) adversarial embedding, which accomplish the goal of hiding a stego message simultaneously deceive the CNN (convolutional neural network) base steganalyzer. In this section some sophisticated algorithms for steganography based on adversarial examples are analyzed further their pros and cons are presented in **Table 3**.

3.2 GAN Based Steganography

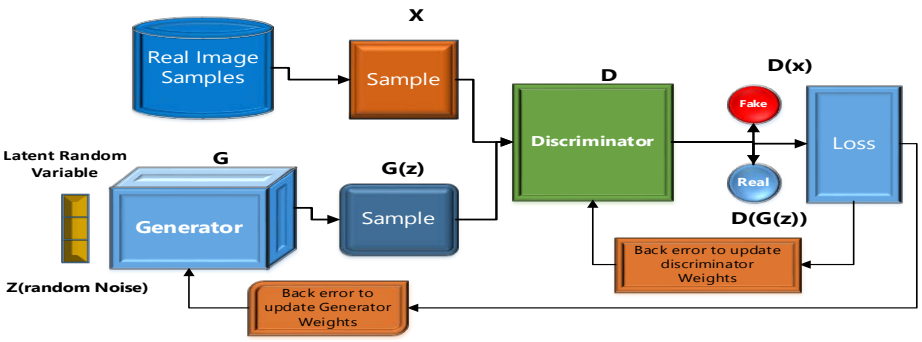


Fig. 3. SGAN Black Diagram

Recently developed GANs (generative adversarial networks) have opened many new approaches to achieve image steganography. In order to apply GANs based approaches to steganographic algorithms researchers proposed steganographic generative adversarial networks (SGANs) as shown in **Fig. 3**. It consists of two type of neural networks, generator network (G) and a discriminator network (D). Generator generates fake images and tries to deceive the discriminator and the discriminator tries to categorize between real and fake data samples. Train them against each other and repeat this and we get better Generator and Discriminator. Volkhonskiy et al [60] first introduce a new adversarial training framework to construct image alike cover image container belonging to deep convolutional generative adversarial networks (GANs). This approach allows us to generate more secure cover images to fool any steganalyzer. But generated images by this method will be drawn attention easily.

To improve the convergence speed, training stability and image quality Shi et al. [34] proposed a generative adversarial network like Volkhonskiy [60], but the steganography by this method is not enough secure. Abadi and Martin [70] also used adversarial training to develop two neural networks to encode a small message that deceive a discriminator network. In order to train the steganographic algorithm and a steganalyzer together Hayes et al. [67] introduce a game between three people, Alice, Bob and Eve. Beyond traditional adversarial learning applications such as image generation tasks. This was the one of the foremost real-life applications of adversarial training. Another successful approach by Tang et al. [71] called (ASDL-GAN) automatic stego distortion learning framework. Generator in their framework automatically find the pixels which are more acceptable for embed a secret message. Furthermore, in their network the discrimination is replaced with Xu-net [35] But this kind of the networks are less secure and low capacity than conventional approaches. Later, Yang et al [72] made three improvement in ASDL-GAN: framework first changed activation function to Tanh- simulator to decrease number of epochs of training. Secondly changed U-NET based generator network and third one is adding SCA to discriminator to boost resisting performance to SCA based steganalysis algorithms. Although, above discussed all GANs based frameworks are embedding based techniques. The proposed frameworks are only focus on adversarial game & at the same time neglect the most important part of the GANs to generate powerful samples. Since GANs network preference is to generate data samples, it looks like each other, it is a very inherent objective to use GANs to construct a semantic stego carrier precisely from a message. After all, recovery of message is an important restriction to steganographic algorithms based on GANs networks. Researchers have made introductory progress in this instinctive idea. Such as Ke et al. [73], first proposed generative steganography network called kerckhoffs principle (GSK) in which they used generator network to generative a secret message instead of concealed the secret information into the cover image, resulting no alterations appeared in the cover image. In [73] the author first introduces the term “generative steganography”. Liu et al. [74] proposed a scheme to design and train the algorithm based on AC-GANs by constructing a database and dictionary. This method effectively resolves the problem of cover modifications in traditional information hiding. Odena et al. [81] introduce a new method for improve training of GANs for image synthesis. In their work the secret message is conceal in the most complex region of the image that need to permeate by a Cardan grille again corrupted stego image is given to the GAN network for stego generation. Later, Liu et al. [74] proposed another practical method called Digital Cardan Grille (DCG) based on generative steganography framework. In the same year Hu et al. [31], proposed a new steganography algorithm based on stego images generated by GCGANs according to secret information. Methods described above made great progress toward GANs based steganography but still have some issues presented in **Table 3** we need to study further.

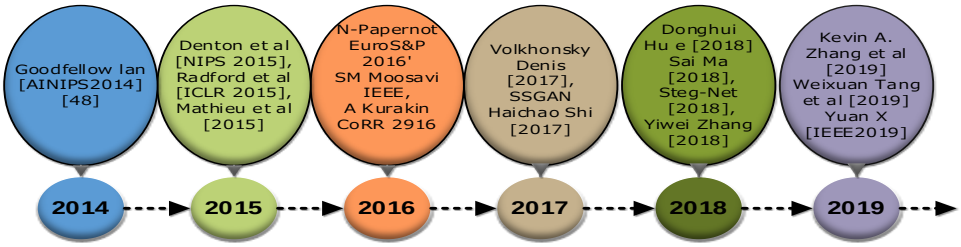


Fig. 4. Development of Deep Learning GANs & Adversarial Examples Based Steganography

Table 3. Pros and Cons of GANs and Adversarial based Steganography Schemes

Approach	Authors	Pros	Cons
GAN Based Steganography	Goodfellow's et al. [19] [2014]	First approach based on images synthesis by mean of GANs and this scheme has been widely used for image generation.	Quite complicated because, first have to generated images then embed the secret information.
	Yan Ke et al. [73] [2017]	Embed messag are constructed by a cover image using generator network instead of embedding message in cover image.	key must send through key channel, which might restrict the applications of GSK.
	Volkhonskiy et al. [60] [2017]	Proposed method not only for credibility of constructed images but also for fighting against to generate more secure steganalysis embedding algorithms.	Constructed images are twisted in semantic order. So, it can easily draw the attention of the attackers or steganalyzers.
	S Haichao Shia et al. [34] [ICLR2017]	Used WGAN instead of DCGAN and faster the network training. Highly secured and improved convergence speed.	Same as in Goodfellow's generated images then embed secret message. It seems high complexity .
	Weixuan Tang et al. [71] [2017]	Proposed automatic steganographic distortion learning algorithm consist of a steganographic generative sub-network & a steganalytic discriminative subnetwork. Achieved good performance.	Low Secure and low capacity than conventional frameworks. TES Sub-network of ASDL-GAN need a long time to pre-train with 1000K iterations.
	Donghui Hu et al. [31] [2018]	First approach exploiting the GAN mechanism to generate stego images without modifications. Highly secured.	The recovery of secret information is not perfect. Size of the stego image is small so, the embedding capacity is not highly satisficed.
	Kevin. et al. [58] [2019]	Embed 4.4bpp which is ten times higher than the others machine learning-based frameworks. It works for various-sized cover images & arbitrary binary data.	Did not present numerical comparison with other deep learning based steganographic schemes. Applicable only for spatial image steganography.
	Jianhua Yang et al. [72] [2018]	Used TanH to fit the optimal embedding simulator. To resist the max SRMd2 selection channel aware-ness are incorporated into the discriminator Faster than ASDL-GAN.	It works only for spatial domain and embedding capacity is not very high.
Adversarial Training	Zhu jiren et al. [66] [2018]	Improved network's quantitative and qualitative performance. Robust to distortion of arbitrary image and highly secured.	Reconstruction of secret message is not absolute, possess an error rate of -0.000005. Embedding capacity is not very high. Its work images with arbitrary size but can not successfully scale to higher relative payloads.
	Sai Ma et al. [68] [2018]	The proposed method does not build new network, it generates adversarial data from steganalyzer, to increase the security of existing methods.	It works only for exiting spatial domain methods not for JPEG.
	Yewei Zhang et al. [64] [2018]	Propose method respectively construct enhanced cover images that can oppose the steganographic algorithms for steganalysis. Highly secure & robustness.	During construction of adversarial examples, traces would be appeared automatically in images.
	Weixuan Tang et al. [65] [2018]	Hiding a stego image simultaneously fool deep learning based steganalyzers. High security and robustness.	The proposed scheme only uses the signs of the gradients. It is indeed either to explore that amplitude of and universal perturbation can also be useful.
	Baluja et al. [62] [2017]	Large capacity and strong invisibility. Proposed method is efficient to train, fast to execute, and produces remarkable adversarial examples,	Stego images produce by this framework is distorted in color and its insecure.
	Hayes J et al. [67] [2017]	Message is embedding into least-cost-location in each of training period. Hide information resulting in weak indivisibly area which make it highly secure.	Stego images produce by this scheme are easily detected by steganalyzers.

4. Existing Challenges & Future Directions

Recent advancements in deep convolutional neural network has made remarkable progress in steganalysis and steganography. Numerous successful approaches to steganography and steganalysis based on deep learning have received widespread attention of researchers and made great progress. Although, this technology still faces many difficulties and challenges remain unsolved. Further study is indeed in order to develop more robust and accurate deep learning algorithms that meet real world applications.

4.1 Challenges in Steganalysis

Breaking Neural Networks with Adversarial Attacks: Neural Network are known to be vulnerable to adversarial examples: inputs that are closed to natural inputs are misclassified by classifiers with very high confident. Currently there is not any satisfied steganography and steganalysis framework for adversarial samples. The presence of adversarial examples is considered as destructive for neural networks, but it may be useful for information hiding frameworks. To account the adversarial problems, research scholars need to develop secure machine learning algorithms that look for outlines and false flags.

Quality of Dataset: Steganalyzers for content adaptive steganography shows worse detection performance when dataset contain images with different content complexities. Deep learning algorithms solve many tasks by extracting useful information from dataset. So, the quality of the training greatly depends on the quality of the dataset input. Steganographic and steganalysis deep framework for digital images are continuously designed and benchmark based on BOSSbase 1.01 dataset. While standardized image dataset is important for advancing these fields, getting results from a single source may not provide fair results and even lead to designs that are over-optimized and highly sub-optimal on other image sources.
Images of Arbitrary Size: At present, the steganographic analysis framework for deep leaning can solve only be fed with input media with fixed size. But there is not unified solution for input media with different input sizes. Some of the algorithms required fixed image size as the input and had low accuracy due to under use of the residual obtained by various type of filters [10]. At present there is no universal deep learning steganographic analysis framework for arbitrary size of input images.

Low Payload steganography: Actually, embedding with low payload is still a challenging task for steganalyzers. Existing steganalyzers are less effective productive to detect stego images with low payload of 0.1bpnzAC. To embed payload of 0.1bpnAC less than 2% image contents are needed to modify. Therefore, it's very hard to notice the statistical properties of the steganographic signal. Selection of training samples and learning approaches also playing a significant impact to improve the detection performance of steganalyzers in case of low payload steganographic signal detection.
Cover Source Mismatch (CSM) or Stego Mismatch: Cover source mismatch problems take place when the steganalysis detector is trained on one dataset and test on different dataset. It's hard to get same source dataset in real life applications. Overfitting problem of the deep leaning-based steganography framework in Cover Source Mismatch (CSM) environment caused by different factors such as:

1) Different JPEG Quality Factors (QF)
2) Different Demosaic algorithms used to convert RAW images into TIFF, PNG, BMP.

3) Different step size in down-sampling.
4) Different camera model / different photos sensors.
5) Different digital processing.
6) Different focal length cameras.
7) Different image resolutions.

The stego mismatch induce through different number of embedding bits embed by different approaches. Cover source mismatch problem has not yet been addressing very well & its worthy of studying this problem further.

Feature Learning: The convolution structure in deep learning frameworks is beneficial to capture equivalences among neighboring pixel values of an image. However, in case of global information CNN usually fuses the information of local area layer with pooling operation or scaling of convolution layer. It's worthwhile to developed algorithms to inaugurate global information and gain more fruitful steganalysis feature learning techniques.

Large Number of Training Samples: Large number of training sets are needed for in-depth learning of small sample size training to achieve good detection results. However, large sample size training is time consuming and laborious, and sometimes large number of samples are difficult to obtain. It is an urgent need to train an effective steganalysis frameworks based on deep learning that work with small number of training samples.

4.2 Challenges in Steganography

Consistency between embedding and extraction: Compared with the adaptive steganography methods, using the generative adversarial networks (GANs) to directly generate stego images enable all the process to be completed in one step. However, due to the inevitable errors in training, the information cannot be extracted correctly. Therefore, it is necessary to ensure the consistency of embedding and extraction for practical application.

Embedding efficiency: For steganography based on generative adversarial network, complex network structures often require more time to complete embedding, and there is a high demand for hardware resources. Considering the application, in order to achieve efficient information steganography in low configuration terminals, it is necessary to study the embedding efficiency in network structure.

4.3 Challenges in GANs Based Deep learning Steganography frameworks

In fact, result generated by GANs can be impressive but it can be challenging to train a stable network. Because training process is innately unstable. There are some common challenges faced during training GANs based models. Below are some of the major issues that researcher might come up while training a GANs models.

Convergence Instability: The model parameters fluctuate, diminish and never converge. Convergence is the major problem about GANs both in term of theory and practical.

Mode Collapse: Many GANs model experience major problem during training GANs network called mode collapse. Mode collapse arise when the generator produces limited varieties of samples from distribution of a real dataset.

Diminished gradient: Vanishing gradient problem is experience during training the network. The discriminator network performing very well but generator network has to faced gradient vanishing problems & it learn nothing. So, unbalance between the generator network & discriminator network make overfitting problems, its highly sensitive to the hyper-parameter selections.

4.4 Future Research Directions

Deep learning-based steganography and steganalysis algorithms offer solutions to many problems that are difficult to solve by traditional methods. However, as listed in previous section many challenges remain, and they are currently not addressed very well. We provided an inclusive survey to highlight current exiting challenges in image steganography and steganalysis also intensify the advantages and disadvantages of exiting up to date techniques for researcher that are associate with the design of deep learning framework for steganalysis & steganography schemes. We also undertaken to overview of image steganography and steganalysis deep learning framework to discussed according to the pixel selection, payload capacity and embedding algorithms to open important research issues in the future works. Therefore, its worthy of studying these challenges apace with machine learning techniques. We also listed potential future research directions and point out the developments in deep learning that could lead to potential solutions.

1) Study more deep learning frameworks and deep learning models with theoretical support, such as RNN and Bayesian neural networks. Then apply them to the confrontation between deep learning steganography and steganalysis.
2) Study the compression and clipping of the current super-large parameter/size deep learning steganographic analysis framework and construct a tiny and concise steganographic analysis framework with good performance.
3) Study the automatic generation of deep learning steganography analysis framework. In the design of deep learning steganographic analysis framework, human factors should be completely excluded.
4) Embedding rate and security of deep learning steganography frameworks are comparable to the minimal distortion embedding framework.

5. CONCLUSION

In this paper, we demonstrate, the role of deep learning approaches based on convolutional neural network for image steganography and steganalysis. Recently steganalysis & steganographic algorithms implemented as Deep Convolutional Neural Networks (DCNNs) achieved remarkable performance. Objective of this study on DL (deep learning) in image steganography and steganalysis are to: 1) encapsulated what has been accomplish up to date; 2) State of art deep learning-based image steganography and steganalysis approached have been assessed and compared. Analyze typical & exclusive problems, approaches & methodologies that researchers have taken to express these challenges; and 3) established some of the suspicious approaches for the future both in terms of applications along with practical innovations. We also discussed important issues and considerations involved in image steganalysis and steganography to illustrate how these challenges can be transformed into prolific future research avenues. It is concluded that tremendous improvement will be achieved if we considered all pros & cons of existing frameworks when deep learning techniques is applied to image steganalysis and steganography.

Reference

[1] A. Kokaj, "Cyber war and terrorism in kosovo," *Academic Journal of Business, Administration, Law and Social Sciences*, 2019.

[2] B. Li, J. He, J. Huang, and Y. Q. Shi, "A survey on image steganography and steganalysis," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 2, no. 2, pp. 142–172, 2011.

[3] J. Fridrich, *Steganography in digital media: principles, algorithms, and applications*, Cambridge University Press, 2009.

[4] A. Laszka and D. Szeszle, "Hide and seek in digital communication: the steganography game," in *Proc. of the 9th Hungarian-Japanese Symposium on Discrete Mathematics and Its Applications*, Fukuoka Japan, June 2-5, 2015,

[5] Y.Qian, J. Dong, W. Wang, "Deep learning for steganalysis via convolutional neural networks," in *Proc. of Media Watermarking, Security, and Forensics International Society for Optics and Photonics*, vol. 9409, pp. 94090J, 2015. Article (CrossRef Link).

[6] X. Xu, Y.Sun, G. Tang, S. Chen, & J. Zhao, "Deep learning on spatial rich model for steganalysis," in *Proc. of International Workshop on Digital Watermarking. Springer*, pp. 564–577, 2016. Article (CrossRef Link).

[7] J. Zeng, S. Tan, B. Li, and J. Huang, "Large-scale JPEG image steganalysis using hybrid deep-learning framework," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 5, pp. 1200–1214, 2018. Article (CrossRef Link).

[8] M. Boroumand, M. Chen, and J. Fridrich, "Deep residual network for steganalysis of digital images," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 5, pp. 1181–1193, 2019. Article (CrossRef Link).

[9] J. Fridrich and D. Soukal, "Matrix embedding for large payloads," *SPIE 6072, Security, Steganography, and Watermarking of Multimedia Contents VIII*, 2006. Article(CrossRef Link).

[10] R. Zhang, F. Zhu, J. Liu, and G. Liu, "Efficient feature learning and multi-size image steganalysis based on CNN," *Multimedia*, Jul 2018.

[11] V. Holub and J. Fridrich, "Designing steganographic distortion using directional filters," in *Proc. of* 2012 *IEEE International workshop on information forensics and security (WIFS). IEEE*, pp. 234–239. 2012. Article (CrossRef Link).

[12] B. Li, M. Wang, J. Huang, and X. Li, "A new cost function for spatial image steganography," in *Proc. of IEEE International Conference on Image Processing (ICIP)*, pp. 4206–4210, 2014. Article (CrossRef Link)

[13] T. Pevny, T. Filler, and P. Bas, "Using high-dimensional image models to perform highly undetectable steganography," in *Proc. of International Workshop on Information Hiding. Springer*, pp. 161–177, 2010. Article (CrossRef Link).

[14] V. Holub and J. Fridrich, "Digital image steganography using universal distortion," in *Proc. of the first ACM workshop on Information hiding and multimedia security*. *ACM*, pp. 59–68, 2013. Article (CrossRef Link).

[15] L. Guo, J. Ni, and Y. Q. Shi, "Uniform embedding for efficient JPEG steganography," *IEEE transactions on Information Forensics and Security*, vol. 9, no. 5, pp. 814–825, 2014. Article (CrossRef Link).

[16] Z. Wang, X. Zhang, and Z. Yin, "Hybrid distortion function for JPEG steganography," *Journal of Electronic Imaging*, vol. 25, no. 5, p. 050501, 2016. Article (CrossRef Link).

[17] V. Holub, "Universal distortion function for steganography in an arbitrary domain," *Eurasip Journal on Information Security*, vol. 2014, p. 1, 2014. Article (CrossRef Link).

[18] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, "*Intriguing properties of neural networks*," 2013.

[19] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," *ICLR*, 2015.

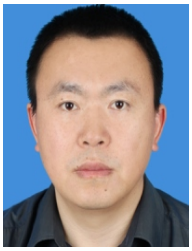
[20] S. Tan and B. Li, “Stacked convolutional auto-encoders for steganalysis of digital images,” in *Proc. of Signal and Information Processing Association Annual Summit and Conference (APSIPA),* 2014 *Asia-Pacific. IEEE*, pp. 1–4, 2014. Article (CrossRef Link).
[21] G. Xu, “Deep convolutional neural network to detect J-UNIWARD,” in *Proc. of the 5th ACM Workshop on Information Hiding and Multimedia Security. ACM*, pp. 67–73, 2017. Article (CrossRef Link).
[22] G. Xu, H.-Z. Wu, and Y. Q. Shi, “Ensemble of CNNs for steganalysis: An empirical study,” in *Proc. of the 4th ACM Workshop on Information Hiding and Multimedia Security. ACM*, pp.103–107, 2016. Article (CrossRef Link).
[23] M. Yedroudj, F. Comby, and M. Chaumont, “Yedroudj-Net: An efficient CNN for spatial steganalysis,” in *Proc. of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, pp. 2092–2096, 2018. Article (CrossRef Link).
[24] J. Fridrich and J. Kodovsky, “Rich models for steganalysis of digital images,” *IEEE Transactions on information Forensics and Security*, vol. 7, no. 3, pp. 868–882, 2012. Article (CrossRef Link).
[25] D. Hu, S. Zhou, Q. Shen, S. Zheng, “Digital image steganalysis based on visual attention and deep reinforcement learning,” *IEEE Access*, 2019. Article (CrossRef Link).
[26] Meng, Ruohan, Steven G. Rice, "A fusion steganographic algorithm based on faster R-CNN,” *Computers, Materials & Continua*, vol. 55, no. 1, 2018.
[27] V. Holub and J. Fridrich, “Random projections of residuals for digital image steganalysis,” *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 12, pp. 1996–2006, 2013. Article (CrossRef Link).
[28] J. Ye, J. Ni, and Y. Yi, “Deep learning hierarchical representations for image steganalysis,” *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 11, pp. 2545–2557, 2017. Article (CrossRef Link).
[29] Jung, Ki-Hyun, "A Study on Machine Learning for Steganalysis,” in *Proc. of the 3rd International Conference on Machine Learning and Soft Computing. ACM*, pp. 12-15, 2019. Article (CrossRef Link).
[30] J. Yang, K. Liu, X. Kang, E. Wong, and Y. Shi, “Steganalysis based on awareness of selection-channel & deep learning,” in *Proc. of International Workshop on Digital Watermarking. Springer*, pp. 263–272, 2017. Article (CrossRef Link).
[31] 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. Article (CrossRef Link).
[32] P. Wu, Y. Yang, and X. Li, “StegNet: Mega image steganography capacity with deep convolutional network,” *Future Internet*, vol. 10 , p. 54, 2018. Article (CrossRef Link).
[33] S. Baluja, “Hiding images in plain sight: Deep steganography,” *Advances in Neural Information Processing Systems*, pp. 2069– 2079, 2017.
[34] H. Shi, J. Dong, W. Wang, Y. Qian, and X. Zhang, “SSGAN: secure steganography based on generative adversarial networks,” in *Proc. of Pacific Rim Conference on Multimedia. Springe*r, pp. 534–544, 2017.Article (CrossRef Link).
[35] G. Xu, H.-Z. Wu, “Structural design of convolutional neural networks for steganalysis,” *IEEE Signal Processing Letters*, vol. 23, pp.708–712, 2016. Article (CrossRef Link).
[36] M. Chen, M. Boroumand, and J. Fridrich, “Deep learning regressors for quantitative steganalysis,” *Electronic Imaging*, vol. 2018, pp. 160-1-160-7(7), 2018. Article (CrossRef Link).
[37] Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. "Faster R-CNN: Towards real-time object detection with region proposal networks,” *Advances in neural information processing systems*, pp. 91-99. 2015.
[38] L. Pibre, J. Pasquet, D. Ienco, and M. Chaumont, “Deep learning is a good steganalysis tool when embedding key is reused for different images, even if there is a cover source mismatch,” *Electronic Imaging*, vol. 2016, no. 8, pp.1–11, 2016. Article (CrossRef Link).
[39] V. Holub and J. Fridrich, "Low-Complexity Features for JPEG Steganalysis Using Undecimated DCT," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 2, pp. 219-228, Feb. 2015. Article (CrossRef Link)

[40] M. Salomon, R. Couturier, C. Guyeux, "Steganalysis via a convolutional neural network using large convolution filters for embedding process with same stego key: A deep learning approach for telemedicine," *European Research in Telemedicine/La Recherche*, vol. 6, no. 2, pp. 79–92, 2017. Article (CrossRef Link).
[41] S. Wu, S.-h. Zhong, and Y. Liu, "A novel convolutional neural network for image steganalysis with shared normalization," *IEEE Transactions on Multimedia*, vol. 22, no. 1, pp. 256-270, 2019. Article (CrossRef Link).
[42] Dai, Jifeng, Yi Li, Kaiming He, and Jian Sun, "R-FCN: Object detection via region-based fully convolutional networks," *Advances in neural information processing systems*, pp. 379-387, 2016.
[43] B. Li, W. Wei, A. Ferreira, and S. Tan, "Rest-Net: Diverse activation modules and parallel subnets-based CNN for spatial image steganalysis," *IEEE Signal Processing Letters*, vol. 25, no. 5, pp. 650–654, 2018. Article (CrossRef Link).
[44] V. Sedighi and J. Fridrich, "Histogram layer, moving convolutional neural networks towards feature-based steganalysis," *Electronic Imaging*, vol. 2017, no. 7, pp. 50–55. Article (CrossRef Link).
[45] J. Lu, G. Zhou, C. Yang, Z. Li, "Steganalysis of content-adaptive steganography based on massive datasets pre-classification and feature selection," *IEEE Access*, vol. 7, pp. 21 702–21711, 2019. Article (CrossRef Link).
[46] J. Zeng, S. Tan, G. Liu, B. Li, and J. Huang, "WISERNET: Wider separate-then-reunion network for steganalysis of color images," *IEEE Transactions on Information Forensics and Security*, 2019. Article (CrossRef Link).
[47] M. Chen, V. Sedighi, M. Boroumand, and J. Fridrich, "Jpeg-phase- aware convolutional neural network for steganalysis of jpeg images," in *Proc. of the 5th ACM Workshop on Information Hiding and Multimedia Security. ACM*, pp. 75–84, 2017. Article (CrossRef Link).
[48] J. Yang, Y.-Q. Shi, E. K. Wong, and X. Kang, "JPEG steganalysis based on Dense-Net," *Multimeida*, *ArXiv abs/1711.09335*, 2017
[49] S. Wu, S. Zhong, "Deep residual learning for image steganalysis," *Multimedia tools & Applications*, vol. 77, no. 9, pp.10 437–10 453, 2018. Article (CrossRef Link).
[50] J. Tian and Y. Li, "Convolutional neural networks for steganalysis via transfer learning," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 33, no. 02, p. 1959006, 2019. Article (CrossRef Link).
[51] J. Zeng, S. Tan, B. Li, and J. Huang, "Pre-training via fitting deep neural network to rich-model features extraction procedure and its effect on deep learning for steganalysis," *Electronic Imaging*, vol. 2017, no. 7, pp. 44–49, 2017. Article (CrossRef Link).
[52] M. Zheng, S.-h. Zhong, S. Wu, and J. Jiang, "Steganographer detection via deep residual network," in *Proc. of 2017 IEEE International Conference on Multimedia and Expo (ICME). IEEE*, pp. 235–240, 2017. Article (CrossRef Link).
[53] K. Zhong, G. Feng, "Deep learning for steganalysis based on filter diversity selection," *Science China Information Sciences*, vol. 61, p. 129105, 2018. Article (CrossRef Link).
[54] C. F. Tsang and J. Fridrich, "Steganalyzing images of arbitrary size with CNNs," *Electronic Imaging*, vol. 2018, no. 7, pp. 121-1-121-8(8), pp. 1–8, 2018. Article (CrossRef Link).
[55] X. Zhang, "Efficient steganographic embedding by exploiting modification direction," *IEEE Communications Letters*, vol. 10, pp. 781–783, 2006. Article (CrossRef Link).
[56] D. Volkhonskiy, B. Borisenko, and E. Burnaev, "Generative adversarial networks for image steganography," *ICLR*, 2017.
[57] X. Duan, K. Jia, B. Li, D. Guo, E. Zhang, "Reversible image steganography scheme based on a u-net structure," *IEEE Access*, vol. 7, pp. 9314–9323, 2019. Article (CrossRef Link).
[58] K. A. Zhang, A. Cuesta-Infante, and K.Veeramachaneni, "Steganogan: Pushing the limits of image steganography," January 2019.
[59] A. Saleema, "A new steganography algorithm using hybrid fuzzy neural networks," *Procedia Technology*, vol. 24, pp. 1566–1574, 2016. Article (CrossRef Link)
[60] D. Volkhonskiy, I. Nazarov, B. Borisenko, and E. Burnaev, "Steganographic generative adversarial networks,"*Multimedia*, 2017.

[61] N. Papernot, P. McDaniel, I. Goodfellow, "Practical black-box attacks against machine learning," in *Proc. of Asia conference on computer and communications security. ACM*, pp. 506–519, 2017. Article (CrossRef Link).
[62] S. Baluja and I. Fischer, "Adversarial transformation networks: Learning to generate adversarial examples," *arXiv preprint arXiv:1703.09387*, 2017.
[63] J. Kos, I. Fischer, and D. Song, "Adversarial examples for generative models," in *Proc. of* 2018 *IEEE Security and Privacy Workshops. IEEE*, pp. 36–42, 2018. Article (CrossRef Link).
[64] Y. Zhang, W. Zhang, K. Chen, J. Liu, Y. Liu, and N. Yu, "Adversarial examples against deep neural network based steganalysis," in *Proc. of the 6th ACM Workshop on Information Hiding and Multimedia Security*, pp. 67–72, 2018. Article (CrossRef Link).
[65] W. Tang, B. Li, S. Tan, M. Barni, and J. Huang, "CNN based adversarial embedding with minimum alteration for image steganography, *IEEE Trans. Inf. Forensics Secur*," 2018.
[66] J. Zhu, R. Kaplan, "Hidden: Hiding data with deep networks," in *Proc. of the European Conference on Computer Vision*, pp. 682-697, 2018. Article (CrossRef Link).
[67] J. Hayes, "Generating steganographic images via adversarial training," *Advances in Neural Information Processing Systems*, pp.1954–1963, 2017.
[68] S. Ma, Q. Guan, X. Zhao, and Y. Liu, "Adaptive spatial steganography based on probability-controlled adversarial examples," 2018.
[69] W. Tang, B. Li, S. Tan, M. Barni, and J. Huang, "CNN-based adversarial embedding for image steganography," *IEEE Transactions on Information Forensics and Security*, 2019.
[70] M. Abadi and D. G. Andersen, "Learning to protect communications with adversarial neural cryptography," in *Proc. of ICLR*, 2017.
[71] W. Tang, S. Tan, B. Li, and J. Huang, "Automatic steganographic distortion learning using a generative adversarial network," *IEEE Signal Processing Letters*, vol. 24, no. 10, pp. 1547–1551, 2017. Article (CrossRef Link)
[72] J. Yang, K. Liu, X. Kang, E. K. Wong, and Y.Q. Shi, "Spatial image steganography based on generative adversarial network," 2018.
[73] Y. Ke, M. Zhang, J. Liu, and X. Yang, "Generative steganography with kerckhoffs' principle based on generative adversarial networks," *Multimedia*, 2017. Article (CrossRef Link).
[74] Z. Zhang, G. Fu, J. Liu, and W. Fu, "Generative information hiding method based on adversarial networks," in *Proc. of International Conference on Computer Engineering and Networks. Springer*, pp. 261–270, 2018. Article (CrossRef Link)
[75] Israr Hussain, Qianhua He, & Zhuliang Chen, "Automatic Fruit Recognition Based on DCNN for Commercial Source Trace System," *International Journal on Computational Science & Applicatons (IJCSA)*, vol. 8, no. 2/3, pp. 1-14, 2018. Article (CrossRef Link).
[76] W. Huang, X.-F. Zhao, D.-G. Feng, and R.-N. Sheng, "Jpeg steganalysis based on feature fusion by principal component analysis," *Ruanjian Xuebao/Journal of Software*, vol. 23, no. 7, pp. 1869–1879, 2012. Article (CrossRef Link)
[77] J. Qin, X. Sun, X. Xiang, "Principal feature selection and fusion method for image steganalysis," *Journal of Electronic Imaging*, vol. 18(3), 033009, 2009.Article (CrossRef Link).
[78] T. Denemark, J. Fridrich, "Improving selection-channel-aware steganalysis features," *Electronic Imaging*, vol. 2016, no. 8, pp. 1–8, 2016. Article (CrossRef Link).
[79] T.Denemark,V.Sedighi, V.Holub, R.Cogranne, and J. Fridrich, "Selection-channel-aware rich model for steganalysis of digital images," in *Proc. of 2014 IEEE International Workshop on Information Forensics and Security. IEEE*, pp. 48–53, 2014. Article (CrossRef Link).
[80] Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C. and Yu, P.S, "A comprehensive survey on graph neural networks," *Journal of Latex Class Files*, 2019.
[81] Odena, A., Olah, C. and Shlens, "Conditional image synthesis with auxiliary classifier gans," in *Proc. of ICML*, Volume 70 , pp. 2642-2651, August 2017.
[82] Liu, Wei, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, "SSD: Single shot multibox detector," in *Proc. of European conference on computer vision*, pp. 21-37, 2016. Article(CrossRef Link).

Israr Hussain, received his B.E in Electronic Engineering from Dawood University Karachi Pakistan in 2013 and M.S in Information and Communication Engineering from South China University of Technology China in 2018. He was awarded Academic Excellent Award by South China university of Technology. He is currently a Ph.D candidate in the Shenzhen Key Laboratory of Media Security, College of Information and Communication Engineering. His current research interests include multimedia security & multimedia forensics and Machine leanring.

Jishen Zeng (S'16) received the B.S degree of electronic information science and technology from Sun Yat-sen University, Guangzhou, China in 2015. He is currently a Ph.D. student in Shenzhen University majoring in information and communication engineering. His current research interests include steganography, steganalysis, multimedia forensics, and deep learning.



Xinghong Qin received the B.S. degree from Northwestern Polytechnical University, Xi'an, China, in 1994, and the M.S. degree from Beihang University, Beijing, China, in 2009. He is currently pursuing the Ph.D. degree in information and communication engineering with Shenzhen University, Shenzhen, China. His current research interests include steganography, steganalysis, multimedia forensics, & deep learning.

Shunquan Tan (M'10–SM'17) received the B.S. degree in computational mathematics and applied software and the Ph.D. degree in computer software and theory from Sun Yat-sen University, Guangzhou, China, in 2002 and 2007, respectively. He was a Visiting Scholar with New Jersey Institute of Technology, Newark, NJ, USA, from 2005 to 2006. He is currently an Associate Professor with College of Computer Science and Software Engineering, Shenzhen University, China, which he joined in 2007. His current research interests include multimedia security, multimedia forensics, and machine learning.