Security and Forensics Exploration of Learning-based Image Coding

Deepayan Bhowmik, Mohamed Elawady and Keiller Nogueira
Division of Computing Science & Mathematics, University of Stirling, Stirling, FK9 4LA, UK
{deepayan.bhowmik, mohamed.elsayed.elawady, keiller.nogueira}@stir.ac.uk

***Abstract*—Advances in media compression indicate significant potential to drive future media coding standards, *e.g.*, Joint Photographic Experts Group's learning-based image coding technologies (JPEG-AI) and MJoint Video Experts Team's (JVET) deep neural networks (DNN) based video coding. These codecs in fact represent a new type of media format. As a dire consequence, traditional media security and forensic techniques will no longer be of use. This paper proposes an initial study on the effectiveness of traditional watermarking on two state-of-the-art learning based image coding. Results indicate that traditional watermarking methods are no longer effective. We also examine the forensic trails of various DNN architectures in the learning based codecs by proposing a residual noise based source identification algorithm that achieved 79% accuracy.**

***Index Terms*—Media forensics, security, learning based image coding, JPEG-AI, DNN, watermarking, source identification.**

I. Introduction

Recent years have seen a huge growth in creative industries, which includes, film, TV, video, radio and photography, advertising, publishing, galleries, libraries, archives and museums (GLAM), music and visual arts. Creative industries contributed significantly worldwide, *e.g.*, more than £111bn to the UK in 2018 and is one of the fastest growing sectors across the economy [1]. Digital rights management, privacy and security, integrity verification, authenticity are some of major challenges in the media content consumption chain, *e.g.*, the loss of income in creative sector due to piracy, spread of fake news, or evidence tampering for fraud purposes. Therefore the need for media security and forensics solutions has been more than ever. This paper explores the effectiveness of the current security and forensic solutions against new technologies such as machine learning based coding standards.

Advances in deep learning show major promises across various applications of which, examples of media compression [2], [3] are of interest in this paper due to its direct contribution in the creative industry. Literature indicates media compression techniques using deep learning have significant potential to drive future media coding standards, *e.g.*, Joint Video Experts Team's (JVET) deep neural networks (DNN) based video coding [4] and Joint Photographic Experts Group's (JPEG) learning-based image coding technologies (JPEG AI) [5]. As a dire consequence, traditional media security techniques including watermarking and other data hiding algorithms may not be of use. Similarly, traditional media forensics, in particular, camera source identification [6] may require a fresh outlook due to the fact that learning based image coding no longer necessarily keep the original sensor noise like artefacts. Therefore, it requires new explorations and approaches in order to develop a holistic solution in secured media distribution and consumption chain.

This paper examines the effectiveness of current solutions, such as, digital watermarking and source identifications against learning based coding and compression. The security and forensic exploration is inspired by a recent JPEG activity named **JPEG AI** on learning based image coding [7] and therefore the scope of this work is restricted to similar image coding and compression. For image security, we evaluate watermarking embedding distortion and robustness performance against two state-of-the-art learning based coding methods. For forensics we made an attempt to identify specific DNN-based compression architectures that are used in this new type of image coding. Unlike camera sensors that produces raw pixels, we argue that DNN-based approaches are new type of imaging sources that learn the visual representations in order to produce highly compressed images. Therefore, it is fair to proclaim that DNN-based compression architecture identification is a new but important aspect of the new age image forensic. To the best knowledge of the authors, no such explorations were carried out in the literature. The main contributions of the work are two fold:

- Watermarking based security evaluation against learning based image coding and compression and
- DNN architecture (as a new imaging source) identification using three deep learning based methods with varying layer depths.

The rest of the paper is organised as follows: Section II outlines existing learning based image coding techniques, Section III describes the watermarking based security evaluation and DNN architecture (source) identification approaches followed by results and discussion in Section IV and concluding remarks in Section V.

II. Learning based Image Coding

With the increasing accessibility to high-quality, easy-to-use cameras and, consequently, in the generation of images and videos, it is essential to compress image data as much as possible, whereas maintaining good quality, in order to improve storage and transmission. Historically and till date JPEG has played a major role in image compression with various standards such as JPEG 1 [8], a format which is still dominant today even after 28 years of its inception; JPEG

2000 [9], largely used in medical imaging, digital cinema and geographic information systems; and more recent others, *e.g.*, JPEG XS [10] for low latency applications, JPEG XL [11], a potential successor of JPEG 1 with significant improvement in efficiency *etc*. Other approaches include intra coding of High Efficiency Video Coding (HEVC) [12] and WebP [13].

Currently, with the resurgence of neural networks and their widespread success in several tasks [14], [15], researchers began to investigate and propose data-driven learning-based compression approaches based on deep learning [16]–[19]. Such methods seek to **learn** the compression transformations/representations based on a large amount of training data and ultimately aim to produce high quality images at a very low bit rate achieving high compression ratio. Understandably for past couple of years JPEG has been engaged in exploring the potential of an end-to-end learning based image coding standard through its activity called JPEG AI. Currently it is evaluating various approaches proposed in the literature. In this paper we seek to understand the security and forensic implications of such compression and coding methods.

Among the deep learning-based compression approaches, two state-of-the-art works have been selected to be further investigated for the watermarking and source compression identification tasks within the scope of this work. The first one, proposed by Mentzer *et al.* [19] and called HiFic, uses a Conditional Generative Adversarial Network [20] to learn the compression transformations. Precisely, in this approach, two networks play a minimax game: one (generative) network seeks to learn effective compression transformations whereas another (discriminative) network tries to distinguish between real and (generated) compressed data. Such technique has three variants (depending on the target bit rate) whose difference is the final compression quality: (i) HiFic-lo, for low quality, (ii) HiFic-mi, for intermediate quality, and (iii) HiFic-hi, for high compression quality.

The second technique [18] proposes an end-to-end trainable model for image compression based on variational autoencoders. This method, optimized using Mean Squared Error (MSE), has two variations: (i) the first one uses a fully **Factorized** (F) prior density model to learn the compression transformations, and (ii) the second defines a **Hyperprior** (H) capable of capturing compression dependencies. These variations can vary in compression quality, from 1 (lowest) to 8 (highest) indicating the quality level.

III. Methodology

In exploring security and forensics of learning based image coding, we *1)* use a commonly known multiplicative watermarking technique and evaluate embedding distortion and robustness performances; and *2)* propose a residual noise based deep learning algorithm for DNN architecture identification and report the accuracy.

A. Watermarking

Digital watermarking is in use for quite sometime to provide media security, especially for copyright protection or digital rights management. Traditionally frequency domain watermarking schemes have shown better performances [21]–[23], and therefore we propose to use a similar approach in measuring the effect of compression over a non-blind watermarking process (as shown in figure Fig. 1). The main steps of this framework are as follow: (1) one of the frequency-based transforms (Discrete Wavelet Transform (DWT) or Discrete Cosine Transformation (DCT)) are used to embed a watermark inside an original image, (2) a compression coding is applied over the watermarked image using JPEG 1 [24] or learning-based coding (HiFic [19]), F-MSE [18]) at different quality levels, (3) the watermark is extracted through the same frequency-based transform used in the embedding process at step 1. The embedding process is defined as follows:

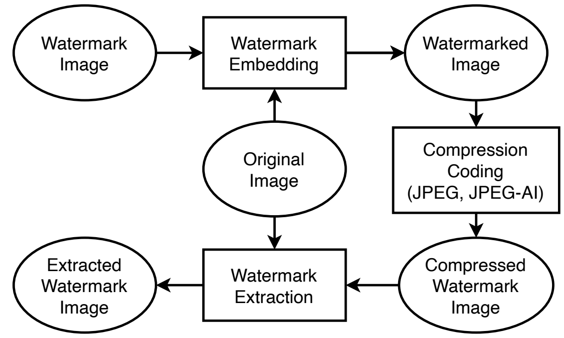

Fig. 1: Image security evaluation using watermarking.

$$\hat{O}_{i,j} = FT(O_{x,y}) \tag{1}$$

$$\hat{D}_{i,j} = \hat{O}_{i,j}(1 + W_{x,y} \cdot G) \tag{2}$$

$$D_{x,y} = FT^{-1}(\hat{D}_{i,j}) \tag{3}$$

where $O_{x,y}$ and $W_{x,y}$ are original and watermark images in spatial domain. $FT(.)$ and $FT^{-1}(.)$ are forward and inverse frequency-based transform functions (either DCT or DWT). The watermarked image is represented in frequency domain as $\hat{D}_{i,j}$ and in spatial domain as $D_{x,y}$. G is the gain factor for watermarking fusion. The compression coding process is computed as follows: $C_{x,y} = H(D_{x,y})$, where $H(.)$ is compression function and $C_{x,y}$ is compressed watermarked image. The extraction process is defined as follows:

$$\hat{C}_{i,i} = FT(C_{x,y}) \tag{4}$$

$$\hat{E}_{i,j} = \frac{\hat{C}_{i,j} - \hat{O}_{i,j}}{\hat{O}_{i,j} \cdot G} \tag{5}$$

$$E_{x,y} = FT^{-1}(\hat{E}_{i,j}) \tag{6}$$

where the extracted watermark image is represented in frequency domain as $\hat{E}_{i,j}$ and in spatial domain as $E_{x,y}$.

B. Source (DNN Architecture) Identification

Due to the increasing use of machine generated image contents [25], camera can no longer be considered as the single original image source. In addition, learning based image coding (as discussed in Section II) relies on large input data to train DNN architectures to learn compression

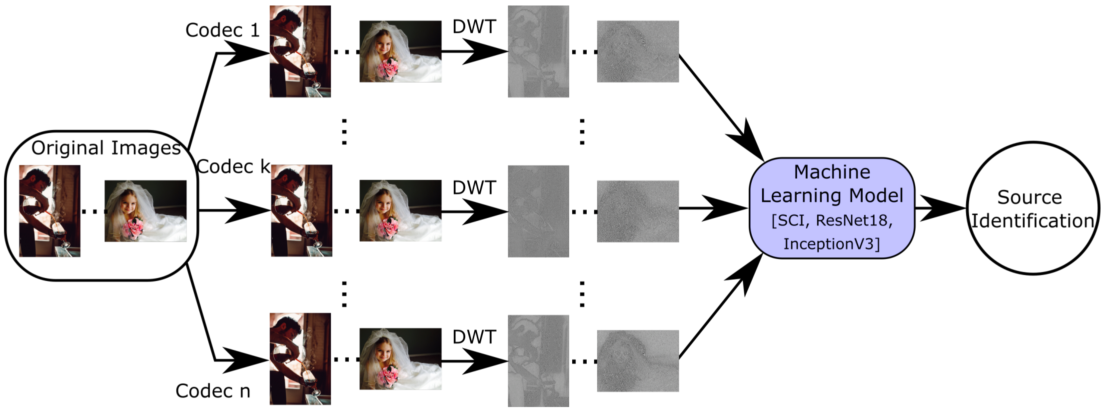

Fig. 2: Overview of the compression source identification pipeline.

transformations/representations and other compression parameters. Therefore, it is more pragmatic to restate the *camera source identification* problem as joint identification of DNN-based image coding architectures along with the original sensors as the new age imaging source(s). We model this DNN source identification as a classification task, in which the DNN architecture of each learning based compression technique represents a class/label. Specifically, given an input image (decoded following a compression), the main objective is to predict which class that image belongs to, *i.e.*, which compression method was applied to that image.

The proposed compression source identification pipeline is composed of three main steps, as depicted in Fig. 2. Firstly, images are compressed using the compression techniques mentioned in Section II. Then, each compressed image, after decoding, is further processed using DWT to extract residual noises. Finally, the high frequency diagonal (or HH) component of the DWT is used as input for a machine learning model, which is responsible for classifying the input data.

This classification process is capable of performing compression source identification given that it associates the input data to a label, which, as explained, is directly related to the compression method applied to that data. Distinct machine learning techniques can be used with the described pipeline. Given the success of neural networks for the image classification task [14], [15], [26], in this work, three deep learning-based approaches were selected to be evaluated for the compression source identification task.

The first approach [27], is a simple Convolutional Network conceived for Source Camera Identification (SCI). This method, referenced hereafter as **SCI**, is composed of only 3 convolutional layers. The motivation for testing this approach was to assess how efficient traditional state-of-the-art methods in SCI are in source compression identification.

The second approach evaluated for source compression identification is the famous and traditional Residual Networks [28]. This technique has several convolutional layers with residual (or skip) connections in order to allow better optimization. Precisely, in this work, we evaluated the **ResNet-18** [28], which has 18 convolutional layers.

Finally, the last technique evaluated for compression source identification was the **InceptionV3** [29]. Such convolutional network, evolution of the traditional GoogleNet [30], is composed of modules that process the input with multiple convolutional filters in parallel. As expected, InceptionV3, which is 48 convolutional layers deep, has several improvements when compared to the original network, including label smoothing, factorized convolutions, and the use of auxiliary classifiers.

IV. Results and Discussion[1]

A. Watermarking

Test images are selected from Kodak lossless true color image suite[2]: original images (kodim01, kodim02, kodim05, kodim21, kodim23) and watermark image (kodim15). Both the embedding distortion and watermark robustness performances are measured using common metrics such as Structural Similarity (SSIM) Index [31] and Peak Signal to Noise Ratio (PSNR) for quantitative comparison. The values of gain factor (empirically chosen) are 0.9 for DCT and 0.4 for DWT. Low frequency components of the original images are used to embed the watermark image. A post-processing step using median filter has been applied on watermark images and its extracted version for noise reduction.

Table I shows the average results of watermarking task with compression methods of different quality levels. The embedding and extraction stages always reach highest watermarking performance without using any compression. With high compression quality, the learning-based methods (especially F-MSE-8) and JPEG have close results. With decreasing quality levels, JPEG better preserves the watermark content at the extraction stage against others. Figure 3 shows that learning-based methods have superior compression ratio over JPEG while maintaining similar SSIM values, but perform poorly in extracting a clear watermark. In summary, handcrafted watermarking methods are not compatible well with learning-based compression models for watermark preservation. Therefore, alternative approaches such as deep learning based watermarking methods [32]–[34] might be well worth to explore.

[1]Source code is available at: https://github.com/mawady/vcip21

[2]http://r0k.us/graphics/kodak/index.html

		Embedding		Extraction	
		SSIM	PSNR	SSIM	PSNR
DWT	No comp	**0.946**	22.869	**0.992**	**43.761**
	JPEG Q90	*0.909*	*22.710*	*0.700*	*21.013*
	JPEG Q70	0.870	22.462	0.605	16.449
	JPEG Q50	0.842	22.271	0.510	13.349
	F-MSE-8 [18]	0.914	**22.876**	0.732	23.234
	F-MSE-4 [18]	0.804	22.518	0.530	14.616
	F-MSE-1 [18]	0.664	21.317	0.360	11.260
	HiFic-hi [19]	0.842	22.521	0.570	17.625
	HiFic-mi [19]	0.796	22.231	0.534	15.697
	HiFic-lo [19]	0.725	21.728	0.442	13.258
DCT	No comp	**0.919**	25.575	**0.956**	**31.781**
	JPEG Q90	0.885	*25.349*	*0.539*	*13.581*
	JPEG Q70	0.845	24.840	0.426	13.340
	JPEG Q50	0.818	24.506	0.368	12.226
	F-MSE-8 [18]	*0.882*	**25.649**	0.544	15.667
	F-MSE-4 [18]	0.768	24.909	0.274	10.976
	F-MSE-1 [18]	0.652	22.995	0.219	10.038
	HiFic-hi [19]	0.825	25.029	0.349	11.090
	HiFic-mi [19]	0.787	24.852	0.324	11.211
	HiFic-lo [19]	0.725	23.990	0.285	10.281

TABLE I: Watermarking embedding distortion and robustness performances against standard JPEG compression and two state-of-the-art learning based image coding. Top three results are highlighted as **bold**, underline, *italic* respectively.

B. Source Identification

For the source identification part, we used three state-of-the-art learning based image coding approaches, F-MSE-8 [18], H-MSE-8 [18] and HiFic-hi [19] as target classes. To train and test our approach, we exploited the JPEG AI dataset[3], which is composed of $5,264$ images for training, 350 for validation, and 40 for test. Although the image resolution of this dataset varies considerably, ranging from 835×628 to 6000×4000 pixels, all images have been resized according to each network's specifications. Data from all sets are processed using the aforementioned compression methods. Images from the training set are used to optimize the networks whereas the validation set is only used to evaluate the convergence of the algorithms. Finally, results are reported based on the test set.

Aside from this, all networks assessed in this work were implemented using PyTorch. During training, all networks used the same set of hyperparameters, which were defined based on convergence analyses. Specifically, learning rate, weight decay, batch size, and number of epochs are 0.01, 0.005, 128, and 50, respectively. All experiments were performed on a 64 bit Intel i7 8700K machine with 3.7GHz of clock, Debian 10, 64GB of RAM memory, and a GeForce GTX 1080 Ti with 12GB of memory under a 10.1 CUDA version.

Table II presents the obtained results. As can be seen through this table, SCI [27] produced the worst results, which indicates that methods proposed for source camera identification are not suitable for source compression identification.

[3]https://jpeg.org/jpegai/dataset.html

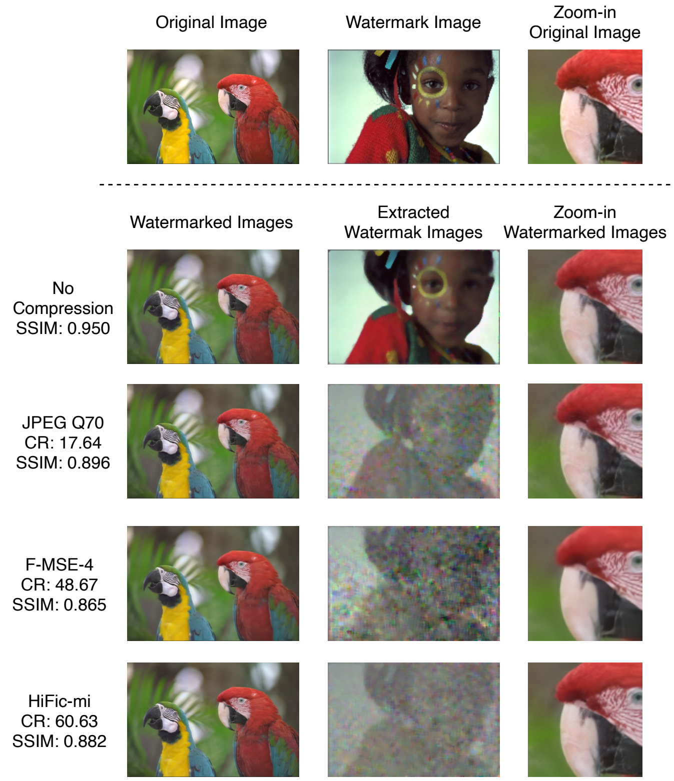

Fig. 3: Representative visual results of DCT watermarking with compression coding options (no compression, JPEG with quality 70, F-MSE-8 [18] and HiFic-mi [19]). Column 1 shows compression ratios (CR) and SSIM values of watermarked images. For similar visual quality at high CR (for both learning based coding), watermarking information is severely affected.

Network →	SCI [27]	ResNet18 [28]	InceptionV3 [29]
Accuracy (%)	34.17	59.17	**79.17**

TABLE II: Source (DNN architecture) identification performance for state-of-the-art learning based image coding.

Aside from this, InceptionV3 [29] yielded the best outcome, followed by ResNet18 [28], an expected outcome, given that deeper networks have more learning capacity and, consequently, tend to produce better results.

V. Conclusions

This paper explores the security and forensic performances of the futuristic learning based image codecs. The image security has been evaluated by using frequency domain multiplicative watermarking techniques and compared against standard JPEG compression and two state-of-the-art learning based codecs. Results indicate that generic watermarking techniques are unable to provide adequate robustness against learning based codecs. The forensic aspects are examined with a residual noise based deep learning source identification algorithm where the DNN-based compression architectures of the learning based codecs are classified with considerable accuracy.

References

[1] Ziga Dernac, "DCMS Sectors Economic Estimates 2018 (provisional): Gross Value Added," Tech. Rep., Department for Digital, Culture, Media and Sport, UK Government, Feb 2020.

[2] Ionut Schiopu and Adrian Munteanu, "Deep-learning-based lossless image coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 7, pp. 1829–1842, 2019.

[3] Dong Liu, Yue Li, Jianping Lin, Houqiang Li, and Feng Wu, "Deep learning-based video coding: A review and a case study," *ACM Computing Surveys (CSUR)*, vol. 53, no. 1, pp. 1–35, 2020.

[4] Jill M Boyce, Jianle Chen, Jens-Rainer Ohm, Gary J Sullivan, Thomas Wiegand, and Yan Ye, "Guest editorial introduction to the special section on the joint call for proposals on video compression with capability beyond hevc," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 5, pp. 1203–1207, 2020.

[5] Joao Ascenso, Pinar Akyazi, Fernando Pereira, and Touradj Ebrahimi, "Learning-based image coding: early solutions reviewing and subjective quality evaluation," in *Optics, Photonics and Digital Technologies for Imaging Applications VI*. International Society for Optics and Photonics, 2020, vol. 11353, p. 113530S.

[6] Francesco Marra, Giovanni Poggi, Carlo Sansone, and Luisa Verdoliva, "Blind PRNU-based image clustering for source identification," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 9, pp. 2197–2211, 2017.

[7] "JPEG learning-based image coding (JPEG AI)," https://jpeg.org/jpegai/, Accessed: 2021-06-27.

[8] Graham Hudson, Alain Léger, Birger Niss, and István Sebestyén, "JPEG at 25: Still going strong," *IEEE MultiMedia*, vol. 24, no. 2, pp. 96–103, 2017.

[9] Charilaos Christopoulos, Athanassios Skodras, and Touradj Ebrahimi, "The JPEG2000 still image coding system: an overview," *IEEE transactions on consumer electronics*, vol. 46, no. 4, pp. 1103–1127, 2000.

[10] Antonin Descampe, Joachim Keinert, Thomas Richter, Siegfried Fößel, and Gaël Rouvroy, "JPEG XS, a new standard for visually lossless low-latency lightweight image compression," in *Applications of Digital Image Processing XL*. International Society for Optics and Photonics, 2017, vol. 10396, p. 103960M.

[11] Jyrki Alakuijala, Ruud van Asseldonk, Sami Boukortt, Martin Bruse, Iulia-Maria Comșa, Moritz Firsching, Thomas Fischbacher, Evgenii Kliuchnikov, Sebastian Gomez, Robert Obryk, et al., "JPEG XL next-generation image compression architecture and coding tools," in *Applications of Digital Image Processing XLII*. International Society for Optics and Photonics, 2019, vol. 11137, p. 111370K.

[12] Gary J Sullivan, Jens-Rainer Ohm, Woo-Jin Han, and Thomas Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Transactions on circuits and systems for video technology*, vol. 22, no. 12, pp. 1649–1668, 2012.

[13] Google, "Webp: Compression techniques," http://developers.google.com/speed/webp/docs/compression, Accessed: 2021-06-26.

[14] Zhihao Wang, Jian Chen, and Steven CH Hoi, "Deep learning for image super-resolution: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[15] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[16] David Minnen, Johannes Ballé, and George D Toderici, "Joint autoregressive and hierarchical priors for learned image compression," in *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds. 2018, vol. 31, Curran Associates, Inc.

[17] Johannes Ballé, "Efficient nonlinear transforms for lossy image compression," in *2018 Picture Coding Symposium (PCS)*. IEEE, 2018, pp. 248–252.

[18] Johannes Ballé, David Minnen, Saurabh Singh, Sung Jin Hwang, and Nick Johnston, "Variational image compression with a scale hyperprior," in *International Conference on Learning Representations*, 2018.

[19] Fabian Mentzer, George D Toderici, Michael Tschannen, and Eirikur Agustsson, "High-fidelity generative image compression," *Advances in Neural Information Processing Systems*, vol. 33, 2020.

[20] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets," in *International Conference on Neural Information Processing Systems*, 2014, p. 2672–2680.

[21] Yuan-Min Li, Deyun Wei, and Lina Zhang, "Double-encrypted watermarking algorithm based on cosine transform and fractional fourier transform in invariant wavelet domain," *Information Sciences*, vol. 551, pp. 205–227, 2021.

[22] Deepayan Bhowmik and Charith Abhayaratne, "Quality scalability aware watermarking for visual content," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5158–5172, 2016.

[23] Ingemar J Cox, Joe Kilian, F Thomson Leighton, and Talal Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1673–1687, 1997.

[24] Gregory K Wallace, "The JPEG still picture compression standard," *IEEE Transactions on Consumer Electronics*, vol. 38, no. 1, pp. xviii–xxxiv, 1992.

[25] Jianmin Bao, Dong Chen, Fang Wen, Houqiang Li, and Gang Hua, "Cvae-gan: fine-grained image generation through asymmetric training," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2745–2754.

[26] Lei Ma, Yu Liu, Xueliang Zhang, Yuanxin Ye, Gaofei Yin, and Brian Alan Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS journal of photogrammetry and remote sensing*, vol. 152, pp. 166–177, 2019.

[27] David Freire-Obregon, Fabio Narducci, Silvio Barra, and Modesto Castrillon-Santana, "Deep learning for source camera identification on mobile devices," *Pattern Recognition Letters*, vol. 126, pp. 86–91, 2019.

[28] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[29] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[30] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[31] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[32] Jiren Zhu, Russell Kaplan, Justin Johnson, and Li Fei-Fei, "Hidden: Hiding data with deep networks," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 657–672.

[33] Seung-Min Mun, Seung-Hun Nam, Haneol Jang, Dongkyu Kim, and Heung-Kyu Lee, "Finding robust domain from attacks: A learning framework for blind watermarking," *Neurocomputing*, vol. 337, pp. 191–202, 2019.

[34] Xin Zhong, Pei-Chi Huang, Spyridon Mastorakis, and Frank Y Shih, "An automated and robust image watermarking scheme based on deep neural networks," *IEEE Transactions on Multimedia*, 2020.