AI-BASED COMPRESSION: A NEW UNINTENDED COUNTER ATTACK ON JPEG-RELATED IMAGE FORENSIC DETECTORS?

Alexandre Berthet, Jean-Luc Dugelay

EURECOM
Digital Security
Biot, France

ABSTRACT

The detection of forged images is an important topic in digital image forensics. There are two main types of forgery: copy-move and splicing. These forgeries are created with image editors that apply JPEG compression by default, when saving the forged images. As a result, the authentic and falsified areas have different compression statistics, including histograms of DCT coefficients that show inconsistencies in the case of double JPEG compression. Therefore, the detection of double JPEG compression (DJPEG-C) is an important topic for JPEG-related image forensic detectors. Since the emergence of deep learning in image processing, AI-based compression methods have been proposed. This paper is the first to consider AI-based compression with digital image analysis tools. The objective is to understand whether AI-based compression can be a new unintended counter-attack for JPEG-related image forensic detectors. To verify our hypothesis, we selected the best detector to date, an AI-based compression method and the Casia v2 database that contains both splicing and copy-move (all publicly available). We focused our experiment on benign post-processing operations: AI-based and JPEG recompressions (with different quality levels). The evaluation is performed using different metrics (average precision, F1 score and accuracy, PSNR, SSIM) to take into account both the impact on detection and image quality. At similar image quality, AI-based recompression achieves a decrease in performance at least twice higher than JPEG, while preserving high visual image quality. Thus, AI-based compression is a new unintended counter-attack, which can no longer be ignored in future studies on image forensic detectors.

Index Terms— Image Forensic Detectors, Double JPEG Detection, AI-based Compression, Counter-Forensics Attack.

1. INTRODUCTION

With the rise of social networking and access to new technologies that make them easier to take, pictures and videos have become commonplace in our daily lives. In parallel to this phenomenon, image editors have developed and are now easy to access and use, leading to potentially malicious modifications. These falsifications can affect various aspects of our society (political, social, etc.). Moreover, they are increasingly difficult to distinguish with the naked eye. Digital Image Forensics (DIF [1]) is a field that provides tools for the blind analysis of images and the localization of certain forgeries. The main manipulations are splicing, where part of an image A is merged into an image B, and cloning, also known as copy and paste, where part of an image is copied onto itself. The localization of these forged regions is done by analyzing the artifacts that result from the process of creating a digital image. This process consists of three stages: acquisition, post-processing and storage. The storage stage includes JPEG compression, which creates artifacts in the image. Notably, blocks of pixels are converted into frequency space by the discrete cosine transform (DCT) during the quantization step. These artifacts have been particularly used in the literature to detect malicious manipulations. Forgeries are created with image editors that often apply further JPEG compression while saving the forged image, leading to double compressed images. As a result, authentic and falsified areas do not have the same compression statistics, as shown by Lukas *et al.* in [2]. In this context, Lin observed inconsistencies in the histograms of DCT coefficients, with missing values and peaks, in the case of double compression [3]; whereas these histograms should follow a Gaussian distribution in the case of single compression. Based on these initial analyses, the detection of double JPEG compression (DJPEG-C) has become an important topic of discussion inside the image forensics community. Most of the methods were based on the analysis of histograms of DCT coefficients. There are different cases of double compression that have been addressed. In fact, double compression artifacts change depending on the quality factors applied. The most likely case is where the 1^{st} quality factor (QF1) is different from the 2^{nd} (QF2). In the case of a similar quantization matrix (i.e. QF1 = QF2), no anomaly exists in the histograms, which makes the detection much harder. As this case is particularly challenging, there are some articles about identical quantization matrix [4]. Similarly, there are two possibilities of applying double compression depending

on the position of DCT blocks: non-aligned (NA-DJPEG-C) [5] or aligned (A-DJPEG-C). As these DCT blocks are of size 8×8, there is only one possibility for the 2^{nd} to be aligned with the 1^{st} one (i.e. 63 out of 64 to be non-aligned). Of course, aligned double JPEG compression is also a case to consider, although it is less common.

The analysis of statistics related to JPEG compression is thus an important topic in digital image forensics. Recently, with the rise of deep learning (DL), some AI-based compression methods have emerged. These solutions were mainly based on auto-encoders, which are composed of two parts: the encoder that reduces the input into a bottleneck containing the main features and the decoder that reconstructs the input from the bottleneck. With the emergence of an innovative compression process, the JPEG organization decided to evaluate these AI-based compression methods to create JPEG-AI[1] as the next image coding standard. The aim of this new compression standard is to provide a better compression for humans and machines. Thus, instead of having a single output (i.e. the reconstructed image), JPEG-AI aims to provide three solutions: the standard reconstruction, an image processing task (e.g. denoising) and a computer vision task (e.g. image classification) [6]. This new compression format is expected to be available in the next few years (estimated in April 2024), as the JPEG-AI proposals will be presented and discussed at the 96^{th} JPEG meeting (July 2022). As a result, the field of DIF could be impacted, and in particular image forensic detectors that are based on JPEG artifacts. This new standard based on deep learning (a trendy field) could become the next democratized compression method. Therefore, we are the first to face both domains to study the impact of AI-based compression on image forensic detectors that are based on JPEG artifacts. The purpose is to determine whether AI-based compression can be a potential unintended attack, in anticipation of the future JPEG-AI standardization.

This section has introduced the topic of DJPEG-C detection, as well as the main objective of our paper. In the section 2, we present the state of the art of detectors based on JPEG artifacts. In the section 3, we detail the process of our method, as well as the models selected for this purpose. The results of our evaluation are presented in section 4. We conclude our work in section 5.

2. RELATED WORKS

Early methods based on compression artifacts, for localizing falsifications, used DJPEG-C detection as a solution for finding forged areas. In [7], A and NA-DJPEG-C have been taken into account with a method based on the derivation of a unified statistical model characterizing the DCT coefficients. The result is a likelihood map of the images indicating if the blocks are double compressed or not, which allows finding the forged areas. In [8], localization of splicing is addressed using NA-DJPEG-C detection, in the case of QF2 higher than QF1, with a region-wise algorithm. With the development of deep learning (DL) in the last decades, deep architectures have been used for digital image forensics, and thus for forgery detection. In particular, convolutional neural networks (CNNs) [9] have been widely used for this task, with some preprocessing before or in the network. In fact, DL methods for digital image forensics require a preprocessing module to extract relevant artifacts that are overshadowed by the image content [10]. In this section, we detail the state-of-the-art methods with their different architectures and preprocessing modules.

Wang *et al.* propose a method [11] based on histograms of DCT coefficients, which are mainly used to detect DJPEG compressed images. As stated in the paper, the artifacts are handcrafted by concatenating the histograms before feeding the network. An interval is set to solve the problem of variable histogram size and reduces the computation with negligible information loss, resulting in a 99×1 vector to feed the network. Their architecture is based on a basic CNN with convolutional layers followed by three fully connected layers for classification. Their model performed well in the case of NA-DJPEG-C, especially when QF2 was higher than QF1 and even for small patches (64×64). Barni *et al.* present the first method [12] based on the CNN that extracts artifacts, thanks to a pre-processing module integrated into the network. In fact, three preprocessing techniques are detailed: i) based on the pixel domain with the subtraction of the image mean (handcrafted); ii) based on the noise domain with the residual noise (handcrafted); iii) with the histograms of DCT coefficients (embedded). The results show that the network based on handcrafted artifacts localizes better when dealing with A-DJPEG-C, while the CNN based on embedded module is the best on NA-DJPEG-C. Furthermore, the CNN based on embedded module is able to work even with some basic processing operations.

Although previous methods have been successful, this was only in specific cases (notably NA-DJPEG-C) and for certain quality factors (e.g. QF2 $>$ QF1). In [13], Park *et al.* propose a solution to detect DJPEG-C in general cases with mixed quality factors to localize splicing and copy-move. First, a new dataset dedicated to DJPEG-C detection is detailed, with the objective of being more realistic. They selected 1,120 quantization tables (QF between 0 and 100) from JPEG images that they extracted from their forensic tool, which guarantees the authenticity of images and is available on a public website to characterize real-world scenarios. To create their dataset, they applied single and double compressions to RAW images by randomly selecting from these 1,120 quantization tables. The method is based on DCT coefficient histograms with an embedded module in the network and quantization tables that are reshaped into vectors and added to the classification part. These quantization tables, contained in the header file, are generally not used for

[1] https://jpeg.org/jpegai/index.html

DJPEG-C detection because the quality factor is fixed, which is not the case here (mixed QFs). In [14], Verma *et al.* follow the same process with the use of the DenseNet architecture, which is fed by histograms of DCT coefficients, whose size has been calculated to be optimal. The results obtained by the state-of-the-art methods for DJPEG-C detection (on dataset from [13]) and the performance for forgery localization (on RAISE [15]) are summarized in the table 1.

Methods	DJPEG-C (Acc.)	Copy-move (F1)	Blurring (F1)
Wang [11]	73.05%		
Barni [12]	83.47%	0.6323	0.6450
Park [13]	92.76%	0.7704	0.7428
Verma [14]	94.49%	0.7992	0.7744

Table 1. Performance of the state-of-the-art methods for DJPEG-C detection (accuracy, on the dataset from [13]) and forgeries localization (F1-score, on RAISE database [15]).

3. PROPOSED FRAMEWORK

The objective of this paper is to provide a first study on the combination of two fields that have never been confronted: digital image forensics and AI-based compression. In particular, we want to analyze the impact of such recompression on JPEG-related image forensic detectors. Recompression can degrade the artifacts used in forgery localization. This can occur when distributing images, whether on social networks or via messaging applications, as they apply compression. Thus, recompression, whether JPEG or AI based, is a non-malicious process, which could unfortunately affect forgery detectors. In contrast, other post-processing operations (e.g. median filtering, Gaussian blur, additional noise, etc.) are applied with the intention to degrade their performance. In this paper, we want to study the impact of a possible unintended counter-attack on JPEG-related image forensic detectors. Thus, this paper mainly focuses on AI-based and JPEG recompressions, which are considered benign. The objective is to select the best methods in each area and to confront them through a framework to evaluate whether AI-based compression can be considered as a new unintended attack on JPEG-related image forensic detectors. Our framework is based on three elements that are publicly available: CAT-Net (detector[2]), HiFiC (compressor[3]) and on the Casia v2 (database[4]).

CAT-Net [16] is a detector capable to localize splicing and copy-move, based on DJPEG-C detection (accuracy of 93.93% on the dataset of [13]). It was evaluated on six databases for forgery detection and for robustness to recompression (with four JPEG QFs) and outperformed several methods in the literature. The CAT-Net analyses the DCT and RGB domains via two streams that process the raw DCT coefficients of the Y-channel with a quantization table and the RGB image respectively. Both streams use the HRNet architecture [17], which maintains high resolution representations, and a fusion step is applied to their outputs to obtain a prediction map. The RGB stream is the HRNet itself, while its first stage is replaced by a JPEG learning artifact module for the DCT stream. We chose CAT-Net over state-of-the-art methods for three main aspects: i) the use of the DCT volume representation, which preserves spatial information (better for localization); ii) the feeding of the network with DCT RAW coefficients (instead of histograms); iii) the pre-training of the DCT stream on DJPEG-C detection.

The literature on AI-based compression is quite recent. Toderici *et al.* published an article (2017) that discusses compression with rational rates based on recurrent neural networks (LSTM, GRU, etc.) with a single learning [18]. Ballé *et al.* have also proposed a work (2018) [19] that presents an end-to-end network to improve the quality of compression, including distortion rates. However, our choice is the HiFiC [20] (High-Fidelity Compression), which is the first AI-based compression method using a GAN (Generative Adversarial Network). Mentzer *et al.* presented three aspects of their method that outperform the state of the art: i) high perceptual fidelity close to the input, with half the bit rate; ii) applicable to high resolution images; iii) optimization of the method with different metrics (PSNR, MS-SSIM, etc.). In addition, they propose three different models with increasing quality: low, medium and high.

Methods from related work (section 2) have been tested on RAISE database, which contains only copy-move. However, Casia v2 database is dedicated to forgery detection with both splicing and copy-move. CASIA v2 database contains 7,200 authentic images and 5,123 forged images of various sizes (320×240 to 800×600). As stated in [21], the ground-truth masks were available through a third party user [22].

4. EXPERIMENTAL RESULTS

Based on these elements, we decided to apply AI-based and JPEG recompressions to Casia V2 database images. All the original images are in JPEG format, leading to high detection performance with CAT-Net. In accordance with our framework, we applied different versions of HiFiC, as well as JPEG compression (N.B. QF from 50 to 80, with a step of 5) on these images. We also included additive white Gaussian noise (AWGN, $sigma = 5.1$), which affects the image quality in the same way as HiFiC-high compression, to give a reference with respect to malicious operations. To evaluate our experiment, we used the same metrics as in [16], based on binary segmentation with true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN). Thus, we used the accuracy ($Acc = \frac{TP+TN}{TP+TN+FP+FN}$) for authentic images,

[2] https://github.com/mjkwon2021/CAT-Net

[3] https://github.com/Justin-Tan/high-fidelity-generative-compression

[4] https://github.com/namtpham/casia2groundtruth

Type of Operations		No	JPEG Compression							HiFiC Compression		AWGN
Parameters			50	55	60	65	70	75	80	Low	High	$\sigma = 5.1$
Objective Quality	PSNR (dB)		**33,7**	**33,9**	34,14	34,41	34,78	35,19	35,81	31,9	**33,75**	**33.81**
	SSIM		**0.915**	**0.92**	0.926	0.932	0.938	0.945	0.954	0.787	**0.901**	**0.861**
Forged Images	Average Precision	0.94	0.45	0.46	0.49	0.58	0.96	0.90	0.84	0.67	0.20	0.3
	F1 score	0.79	0.35	0.33	0.32	0.37	0.40	0.38	0.43	0.17	0.12	0.28
Authentic Images	Accuracy (%)	88.48	83.88	86.44	86.67	90.76	91.24	87.98	91.56	92.11	92.25	80.97

Table 2. Results of forgery localization, with accuracy (%), F1 score and average precision, according to various operations. Objective quality of processed images is furnished (PSNR, SSIM). original - important drop - the hugest drop.

while we calculated the F1 score ($F1 = \frac{2TP}{2TP+FN+FP}$) that emphasizes the positive class for forged images. As accuracy and F1 score depend on a fixed threshold, they also used the average precision (area under the recall-precision curve), which is a threshold-free performance.

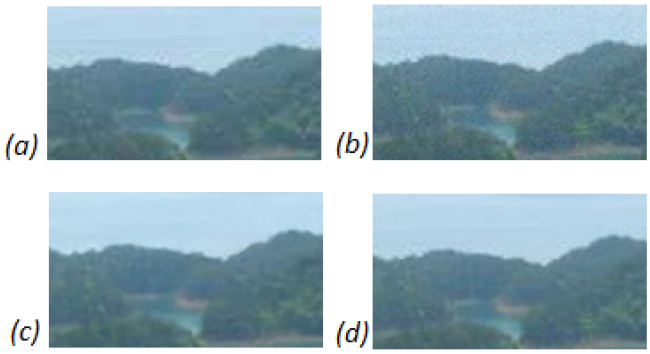

Fig. 1. Comparison of visual image quality of each operation (same objective quality) with a region of an image from CASIA. (a) QF50 (b) AWGN (c) HiFiC-Hi (d) original.

Table 2 shows the results of our experiment on the Casia v2 database. On the one hand, the accuracy is quite high regardless of the compression quality, which means that authentic images are not affected by recompression. On the other hand, according to the F1 score, AI-based compression has a more negative impact than JPEG compression. Average precision gives additional hints on each operation, with various results according to the chosen parameters. Low quality factors have more impact on localization than high factors for JPEG, while the opposite is true for AI-based compression. Overall, if we compare both compressions at equivalent objective image quality (PSNR, SSIM), the localization performance is more impacted (at least twice as much) by AI-based compression than by JPEG or even a malicious operation like AWGN (orange vs. red in the Tab 2). Moreover, HiFiC has been optimized for reducing the bitrate (twice less than JPEG [20]), while preserving a high visual image quality (see Fig. 1). Therefore, HiFiC-high is able to overcome the detector without compromising the image quality.

5. CONCLUSION

This manuscript is the first to address AI-based compression and digital image forensics together. The purpose is to evaluate the impact of AI-based recompression on JPEG-related image forensic detectors. Therefore, in this article we reviewed the literature of such detectors, and we proposed a framework that confronts both fields with their respective best methods to date. Our framework is based on three elements that are publicly available: CAT-Net (detector), HiFiC (compressor) and the Casia v2 (database). We applied HiFiC, JPEG compression and AWGN to 50 images from CASIA to compare their impact on forgery localization. Our result shows that HifiC-High is the most effective operation to lead to a considerable decrease in performance while maintaining high visual image quality. AI-based compression is a new unintended counter-attack for JPEG-related forgery detectors and should be considered in further studies in image forensics.

6. REFERENCES

[1] Judith Redi, Wiem Taktak, and Jean-Luc Dugelay, "Digital image forensics: A booklet for beginners," *Multimedia Tools Appl.*, vol. 51, pp. 133–162, 10 2011.

[2] Jan Lukáš and Jessica Fridrich, "Estimation of primary quantization matrix in double compressed jpeg images," in *Proc. Digital forensic research workshop*, 2003, pp. 5–8.

[3] Zhouchen Lin, Junfeng He, Xiaoou Tang, and Chi-Keung Tang, "Fast, automatic and fine-grained tampered jpeg image detection via dct coefficient analysis," *Pattern Recognition*, vol. 42, no. 11, pp. 2492 – 2501, 2009.

[4] Xiaosa Huang, Shilin Wang, and Gongshen Liu, "Detecting double jpeg compression with same quantization matrix based on dense cnn feature," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 3813–3817.

[5] Tiziano Bianchi and Alessandro Piva, "Detection of nonaligned double jpeg compression based on integer

periodicity maps," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 842–848, 2012.

[6] 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*, Peter Schelkens and Tomasz Kozacki, Eds. International Society for Optics and Photonics, 2020, vol. 11353, pp. 164 – 176, SPIE.

[7] Tiziano Bianchi and Alessandro Piva, "Image forgery localization via block-grained analysis of jpeg artifacts," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 1003–1017, 2012.

[8] M. Barni, A. Costanzo, and L. Sabatini, "Identification of cut amp; paste tampering by means of double-jpeg detection and image segmentation," in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*, 2010, pp. 1687–1690.

[9] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "Imagenet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[10] Alexandre Berthet and Jean-Luc Dugelay, "A review of data preprocessing modules in digital image forensics methods using deep learning," in *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, 2020, pp. 281–284.

[11] Qing Wang and Rong Zhang, "Double jpeg compression forensics based on a convolutional neural network," *EURASIP Journal on Information Security*, vol. 2016, 10 2016.

[12] M. Barni, L. Bondi, N. Bonettini, P. Bestagini, A. Costanzo, M. Maggini, B. Tondi, and S. Tubaro, "Aligned and non-aligned double jpeg detection using convolutional neural networks," *Journal of Visual Communication and Image Representation*, vol. 49, pp. 153–163, Nov 2017.

[13] Jinseok Park, Donghyeon Cho, Wonhyuk Ahn, and Heung-Kyu Lee, "Double jpeg detection in mixed jpeg quality factors using deep convolutional neural network," in *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.

[14] Vinay Verma, Deepak Singh, and Nitin Khanna, "Block-level double jpeg compression detection for image forgery localization," *arXiv preprint arXiv:2003.09393*, 2020.

[15] Duc-Tien Dang-Nguyen, Cecilia Pasquini, Valentina Conotter, and Giulia Boato, "Raise: A raw images dataset for digital image forensics," in *Proceedings of the 6th ACM Multimedia Systems Conference*, New York, NY, USA, 2015, MMSys '15, p. 219–224, Association for Computing Machinery.

[16] Myung-Joon Kwon, Seung-Hun Nam, In-Jae Yu, Heung-Kyu Lee, and Changick Kim, "Learning jpeg compression artifacts for image manipulation detection and localization," *arXiv preprint arXiv:2108.12947*, 2021.

[17] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, et al., "Deep high-resolution representation learning for visual recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 10, pp. 3349–3364, 2020.

[18] George Toderici, Damien Vincent, Nick Johnston, Sung Jin Hwang, David Minnen, Joel Shor, and Michele Covell, "Full resolution image compression with recurrent neural networks," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2017, pp. 5306–5314.

[19] Johannes Ballé, David Minnen, Saurabh Singh, Sung Jin Hwang, and Nick Johnston, "Variational image compression with a scale hyperprior," *arXiv preprint arXiv:1802.01436*, 2018.

[20] Fabian Mentzer, George D Toderici, Michael Tschannen, and Eirikur Agustsson, "High-fidelity generative image compression," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, Eds. 2020, vol. 33, pp. 11913–11924, Curran Associates, Inc.

[21] Jing Dong, Wei Wang, and Tieniu Tan, "Casia image tampering detection evaluation database," in *2013 IEEE China Summit and International Conference on Signal and Information Processing*, 2013, pp. 422–426.

[22] Nam Thanh Pham, Jong-Weon Lee, Goo-Rak Kwon, and Chun-Su Park, "Hybrid image-retrieval method for image-splicing validation," *Symmetry*, vol. 11, no. 1, 2019.