Higher-Order, Adversary-Aware, Double JPEG-Detection via Selected Training on Attacked Samples

Mauro Barni, Ehsan Nowroozi, Benedetta Tondi
Department of Information Engineering and Mathematics, University of Siena
Via Roma 56, 53100 - Siena, ITALY
barni@dii.unisi.it, ehsan.nowroozi@student.unisi.it, benedettatondi@gmail.com

***Abstract*— In this paper we present an adversary-aware double JPEG detector which is capable of detecting the presence of two JPEG compression steps even in the presence of heterogeneous processing and counter-forensic (C-F) attacks. The detector is based on an SVM classifier fed with a large number of features and trained to recognise the traces left by double JPEG detection in the presence of attacks. Since it is not possible to train the SVM on all possible kinds of processing and C-F attacks, a selected set of images, manipulated with a limited number of attacks is added to the training set. The processing tools used for training are chosen among those that proved to be most effective in disabling double JPEG detection. Experimental results prove that training on such a kind of most powerful attacks allows good detection in the presence of a much wider variety of attacks and processing. Good performance are retained over a wide range of compression quality factors.**

I. INTRODUCTION

Image forensic techniques for double JPEG detection in the presence of an adversary is one of the most widely studied topics in adversarial image forensics [1], [2]. Due to the artifacts left by double JPEG (D-JPEG) compression into the histograms of DCT coefficients [3], most studies have focused on detectors based on first order statistics of block DCT coefficients. In turn, a number of powerful attacks which are capable to prevent a correct detection while keeping the mean squared error distortion introduced by the attack limited, have been developed. In [4] and [5], counter-forensic schemes were introduced to remove the artefacts of double compression in the distribution of the first significant digits (FSD) of the DCT coefficients thus making the detector fail [6]. In [7], a universal double JPEG attack against first order based detectors has been proposed, which extends to the DCT domain the universal attack in the pixel domain originally proposed in [8].

In general, it is easy to cope with such attacks by resorting to detectors based on second-order statistics [9], or by properly training the detector with images subject to attacks of the same class of the attack that is going to be used by the adversary [10]. When other kinds of attacks are considered, however, we expect these techniques to fail, all the more that the attacker may decide to combine his attacks with geometric transformations or any other kind of processing capable of impeding a correct detection. In addition, in real applications, the attack is not known in advance, thus impeding to build an ad-hoc detector. This problem is particularly relevant with data driven detectors based on machine learning due the to difficulty of training the detector on all possible attacks.

In order to alleviate the above problem, we propose to build a classifier based on a Support Vector Machine (SVM) accepting as input a large number of features computed both in the pixel and the frequency domain and add to the training set some images which underwent a limited set of attacks. Using a large number of heterogeneous features ensures that the classifier has the necessary degrees of freedom to distinguish images processed in several different ways. The use of features computed in the pixel domain is motivated by the need to cope with geometric attacks that de-synchronize the 8×8 grid at the basis of JPEG compression. With regard to the attacked images used to train the adversary-aware version of the classifier, we include only images processed with the attacks that, when used against a non-aware version of the classifier, result in the worst performance. The rationale behind such a choice is that a detector trained to recognise images subject to this kind of Most Powerful Attacks (MPAs) should also be able to detect double compressed images subject to milder processing.

Experimental results corroborate our expectations showing that, up to a certain extent, the classifier is able to correctly process images that underwent attacks not included in the training set. To cope with the few cases for which this is not the case, we refined the classifier by adding some new attacked image samples. The performance of the final classifier obtained in this way are constantly good across a wide class of attacks and a wide range of quality factors.

The rest of the paper is organised as follows: in Section II we describe the general idea behind the MPA-aware detector; then, we focus on the case of D-JPEG detection and describe our choice of the features for the adversary-aware classification. In Section III, we describe the experimental methodology. The results of the experiments are discussed in Section IV. Conclusions are given in Section V together with some considerations on further research.

II. MPA-AWARE SVM DETECTOR

Our goal is to design an adversary-aware detector that reveals if an images has undergone a double JPEG compression possibly in the presence of other processing or counter-forensic attacks. To do so, we train an SVM detector not only with single and double compressed images but also with a limited number of properly selected examples of attacked

ISBN 978-0-9928626-7-1

images. The idea behind such an approach is illustrated in Fig. 1: training on benign samples leaves wide room for attacks. Adding attacked samples to the training set permits to refine the decision region and make new attacks more difficult. Since, it is not viable to consider all possible kinds of attacks, we train the classifier by including only those attacks that degrade most the performance of an unaware version of the classifier. Hereafter we refer to such attacks as MPAs (Most Powerful Attacks), and the detector trained to recognise them MPA-aware detector. When the analysis is limited to first order statistics and the attack must satisfy a per-pixel distortion constraint, the optimum attack is known and the MPA corresponds to this attack (see [8] and [7] for attacks in the spatial and frequency domain respectively). The optimum attack in the DCT domain has been used in [10] to build an adversary-aware SVM, which was shown to be able to resist to other double JPEG counter-forensic attacks belonging to the same class, namely, first order attacks (e.g., the attack to the FSD coefficients [4], [5]).

The goal of this paper is to overcome the first order statistic limitation inherent in the analysis proposed in [10], and build an adversary-aware detector which is able to work under a wider variety of attacks. As a second goal, we aim at improving the resilience against attacks, like geometrical attacks, for which the visual distortion introduced cannot be measured (and hence constrained) on a per-pixel basis. Figure 2 schematises the detection task addressed in this paper[1]: we let H_0 correspond to the case of single compressed images (in the absence of manipulation), and H_1 to the case in which the image is either compressed twice or compressed, attacked and then compressed again. An attack placed in the middle between the two compression stages may correspond to the application of a processing operation or to a C-F attack to single compressed images, i.e. an attack aimed at erasing single compression traces so to make the image look like an uncompressed one. When the attack occurs after the second compression (* last row in Figure 2), we implicitly assume that it ends up with a JPEG image. This is the case of a C-F attack aiming at making a double compressed image look like a single compressed one. We observe that a three class classification could also be considered to distinguish between single compressed, double compressed, and double compressed and attacked images. However, we opted for a two-class approach since our purpose is to use the presence of double JPEG traces as an indication that the image has been processed in any way after its acquisition. The rationale behind our approach is that most images are stored in JPEG format, and hence any processing is always accompanied by a double JPEG compression.

In order to build a classifier capable to capture different types of dependencies among neighbouring pixels, we need to resort to a large number of features. In this sense, we could adopt the rich models for both spatial and frequency domain

[1]It is worth stressing that, although in this paper we focus on double JPEG detection, the arguments about the MPA-aware classification are general and can be applied to other decision test under adversarial conditions.

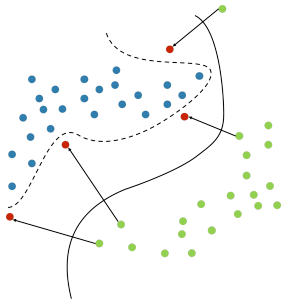

Fig. 1. Rationale behind the design of the adversary-aware classifier. The introduction of a limited number of attacked samples (red dots) permits to narrow the region around legitimate samples (blue) thus making more difficult to camouflage green samples as blue ones.

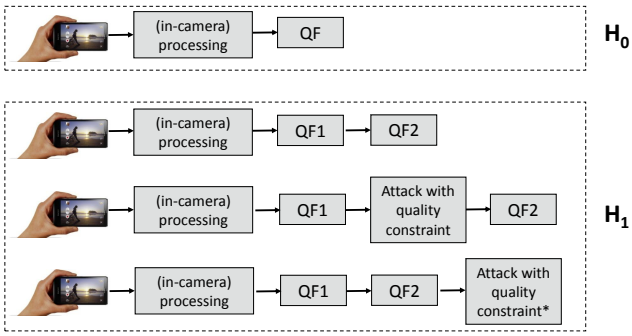

Fig. 2. Adversarial D-JPEG detection task considered in this paper.

described in [11], [12]. However, the huge dimensionality of these models asks for an extremely large training set making the use of standard machine learning techniques (as the SVM) no longer viable (a possibility would be to resort to an ensemble classifier as in [13]). By following a common trend in the literature, then, we select only some higher-order features (both in spatial and frequency domain) to build a model which is rich enough to capture the artefacts introduced by D-JPEG compression under various attacking conditions. To be specific, we selected the Substractive Pixel Adjancency Model (SPAM) features [14] for the pixel (spatial) domain and the CC-PEV features [15] for the DCT domain. The SPAM features are extracted by computing the first order residual (difference array) in all the directions (horizontal, vertical, diagonal), then truncating the values at T ($T = 3$) and finally computing the second-order co-occurrences. In the CC-PEV model, we considered the global histogram and individual histograms for 5 DCT modes at low frequencies, total variation and blockiness (capturing the inter-block dependence) and transition probability matrix from difference arrays (capturing the inter-block dependencies). The final feature space dimensionality is 960 (686 for SPAM and 274 for CC-PEV without calibration).

III. EXPERIMENTAL METHODOLOGY

To perform our experiments we started with gray-scale images in uncompressed format. Part of them was used for training ($\mathcal{S}_{tr}$) and part for testing ($\mathcal{S}_t$). We built the images

ISBN 978-0-9928626-7-1 © EURASIP 2017

to be used for training and testing according to the following procedure (we refer to Fig. 2): for the first class (H_0), the images were single compressed with quality factor QF; for the second class (H_1), the double compressed images were built by compressing the images first with various $QF1$s and then with $QF2$. The attacked images were obtained by first compressing the same images with the $QF1$s, then attacking them with various processing and/or counter-forensics, and finally re-compressing them with $QF2$. For a meaningful analysis, we let $QF = QF2$. In summary, for each image, we built a single compressed version with $QF2$, many double compressed versions with second quality factor $QF2$ (one for each $QF1$), and the same number of attacked versions for each attack.

To compress and attack the images we used the Matlab image processing toolbox. Specifically, we considered geometric transformations (resizing, zooming, rotation, cropping, mirroring and seam-carving), filtering operations (median filtering, blurring, denoising), histogram enhancement and editing (copy-move). For resizing and rotation, we considered the bilinear (BIL), bicubic (BIC) and nearest-neighbor (NN) interpolation methods. Specifically, we considered a resizing scaling factor of 0.9, a zooming factor of 1.2, and a rotation angle of 5 degrees. Cropping was carried out by considering a 440×440 area both aligned and non-aligned with the 8×8 JPEG grid of the image. As to seam carving, the number of vertical seams to be removed was chosen in such a way that the final image has approximately the same size of a resized image with resizing factor equal to 0.9. With regard to filtering, In order to limit the visual degradation of the attacked images, we considered a 3×3 window size for the median filter, and a 3×3 Gaussian smoothing kernel with variance $\sigma^2 = 1$ in the blurring operation. For denoising operation, we considered the wavelet-based filter proposed in [16] with $\sigma^2 = 10$. Histogram enhancement was performed by using contrast-limited adaptive histogram equalisation (CLAHE). Finally, in the copy-move operation, a random part of the image of size 256×256 was copied and pasted into a different part of the image

Regarding counter-forensic, we considered the anti-forensic JPEG algorithm described in [17], which removes the blocking artefacts of JPEG compression by applying a median filter followed by the addition of a Gaussian noise (*dithering*). Such a scheme is known to be quite an effective C-F attack; however, its impact on the attacked image is perceptually significant (especially when the attack is applied in the DCT domain) [18]. To limit visual degradation, we considered a 3x3 median filter and a small variance σ^2 for the noise which is related to the variance of the image and ranges from 1 to 2. It is worth pointing out that, the C-F scheme in [17] aims at erasing the traces of single compression, and then corresponds to a case in which the attack is placed between the two compression steps (last row in Fig. 2). Finally, we also consider the universal counter-forensic schemes of double compression proposed in [10], that is, the MPA against first order based detector.

To build the classifier, we used the (960-dimensional) features extracted from the images to fed an SVM with Gaussian kernel. The kernel parameters are chosen by 5-fold cross validation. In the unaware case, we trained the SVM with single and double compressed images and then tested it on single, double and attacked images In the adversary-aware case, we trained the SVM also with examples of attacked images. To chose the attacks for aware training, we considered the attacks leading to the worse classification accuracy in the unaware case. In all our experiments we set $QF2 = 85$ and considered several values of $QF1 < QF2$. We have checked that similar considerations hold for a different choice of $QF2$, with some obvious differences in the numerical values expressing the performance of the detector.

TABLE I

AUC VALUES OF THE UNAWARE SVM CLASSIFIER.

Attack	D-JPEG	1st order MPA	Attack in [17]	wavelet denoise	median filtering	copyMove 256x256
AUC	1	0.98	0.43	0.8	0.62	0.99
Attack	CLAHE	resize BIC0.9	resize BIL0.9	resize NN0.9	rotation BIC5	rotation NN5
AUC	0.98	0.49	0.53	0.58	0.56	0.58
Attack	zoom 1.2	crop align	crop no align	mirror	blur	seam-carving
AUC	0.64	0.72	0.70	0.58	0.86	0.97

IV. EXPERIMENTAL RESULTS

In our experiments, camera-native (uncompressed) images were taken from the RAISE dataset [19]. We also used uncompressed images from the Dresden Image Database [20] for additional testing. Specifically, the 2000 images in RAISE-2K were split as follows: 1400 images were selected to build the training set (plus other 300 images used for setting the kernel parameters, i.e., internal cross validation) and 300 images for the test set. These images have a large size (4288×2848), then to fasten the feature computation, we sub-sampled them down to a size of 1072×770. Larger images of size 2144×1424 were also considered for testing.

A. Choice of the attacks for MPA-adversary aware training

We trained the unaware SVM classifier with the images in $\mathcal{S}_{tr}$ single compressed with $QF = 85$ and double compressed with $(QF1, QF2) = \{(50, 85), (65, 85), (70, 85), (75, 85), (80, 85)\}$. Accordingly, the training set contained 1200 single compressed images for H_0 and 7000 (5×1200) images for H_1. Table I shows the performance of the unaware SVM. The Area Under the Curve (AUC) of the ROC curve for the classification single vs double and single vs attacked images (for all the considered attacks) is given. The results for rotation BIL are not reported being always very similar to those of the BIC case. The unaware SVM is able to correctly classify single and double compressed images with great accuracy (AUC = 100%). Not surprisingly, the unaware SVM is also able to counter the double JPEG anti-forensic technique in [10]. Indeed, since the scheme is limited to first order statistics of the DCT coefficients, it leaves traces on higher order statistics. However, the unaware SVM fails to classify the attacked samples for almost all the manipulations. The most harmful attacks are the geometrical attacks and the

ISBN 978-0-9928626-7-1 © EURASIP 2017

TABLE II

AUC VALUES OF THE MPA-AWARE CLASSIFIER.

Attack	D-JPEG	1st order MPA	Attack in [17]	wavelet denoise	median filtering	copyMove 256x256
AUC	0.99	0.98	0.96	0.90	0.98	0.99
Attack	CLAHE	resize BIC0.9	resize BIL0.9	resize NN0.9	rotation BIC5	rotation NN5
AUC	0.92	0.92	0.95	0.80	0.91	0.81
Attack	zoom 1.2	crop align	crop no align	mirror	blur	seam-carving
AUC	0.97	0.92	0.92	0.99	0.98	0.95

counter-forensic attack in [17], in which case the unaware detector completely fails.

B. Generalisation capabilities of the MPA-aware detector

Based on the above results, we built the MPA example images by considering the resizing attack (with bicubic interpolation) and the anti-forensic attack by Stamm et al. [17], and re-trained the SVM by adding examples of images attacked in this way. In fact, these kinds of manipulations alter the image in a completely different way and training the classifier to recognise one of the two processing does not help with respect to the other. Accordingly, the images of the training set $\mathcal{S}_{tr}$ were compressed with $QF1 = \{50, 65, 70, 75\}$, attacked (resized and attacked with the anti-forensic scheme in [17]) and then re-compressed with $QF2 = 85$.[2] Then, we added these images to the double compressed images as further examples of the H_1 class. The number of images used for the manipulated class, then, raisesd to 16600 (7000 double + 4800 resized and 4800 C-F attacked). Table II shows the results of the test against the SVM trained in such a way. The AUC is above 90% for almost all the processing operations and the counter-forensic attacks, thus confirming the good generalisation capability of the detector. The good performance against non-aligned cropping suggests that the detector is also robust to non-aligned D-JPEG compression, or similarly, to a grid de-calibration attack.

C. Refined MPA-aware detector

From the results in Table II, we see that when a geometrical operation like resizing and rotation is performed by using a nearest neighbour interpolation, the performance of the classification degrade. Then, we refined the MPA-aware detector by adding also some examples of such kind of manipulation in the H_1 class (thus adding further 4800 attacked samples). Table III shows the results of the refined detector. As expected, the performance with respect to resizing with NN interpolation improves. The performance with respect to rotation with NN interpolation also improves. Finally, the performance with respect to the other processing and attacks remain good. In order to get more insight into the impact that the QFs have on the performance, Table IV and V show the AUC of the refined MPA-aware detector for the QF pairs $(65, 85)$ and $(80, 85)$ respectively. Not surprisingly, the $(80, 85)$ case leads to worse results. To check that

[2]Notice that we did not consider the pair $(80 - 85)$ for the training, as we found experimentally that including attacked images with such a small difference between the QFs slightly reduces the performance of the classifier.

TABLE III

AUC VALUES OF THE REFINED MPA-AWARE CLASSIFIER.

Attack	D-JPEG	1st order MPA	Attack in [17]	wavelet denoise	median filtering	copyMove 256x256
AUC	0.99	0.98	0.96	0.91	0.98	0.99
Attack	CLAHE	resize BIC0.9	resize BIL0.9	resize NN0.9	rotation BIC5	rotation NN5
AUC	0.91	0.92	0.94	0.92	0.92	0.91
Attack	zoom 1.2	crop align	crop no align	mirror	blur	seam-carving
AUC	0.97	0.92	0.93	0.99	0.98	0.95

TABLE IV

AUC OF THE REFINED MPA-AWARE CLASSIFIER FOR THE PAIR $(65, 85)$.

Attack	D-JPEG	1st order MPA	Attack in [17]	wavelet denoise	median filtering	copyMove 256x256
AUC	0.99	0.99	0.98	0.96	0.99	0.99
Attack	CLAHE	resize BIC0.9	resize BIL0.9	resize NN0.9	rotation BIC5	rotation NN5
AUC	0.96	0.96	0.97	0.95	0.97	0.95
Attack	zoom 1.2	crop align	crop no align	mirror	blur	seam-carving
AUC	0.98	0.97	0.96	0.99	0.99	0.95

TABLE V

AUC OF THE REFINED MPA-AWARE CLASSIFIER FOR THE PAIR $(80, 85)$.

Attack	D-JPEG	1st order MPA	Attack in [17]	wavelet denoise	median filtering	copyMove 256x256
AUC	0.99	0.98	0.92	0.79	0.94	0.96
Attack	CLAHE	resize BIC0.9	resize BIL0.9	resize NN0.9	rotation BIC5	rotation NN5
AUC	0.87	0.83	0.87	0.78	0.86	0.86
Attack	zoom 1.2	crop align	crop no align	mirror	blur	seam-carving
AUC	0.89	0.85	0.84	0.99	0.91	0.93

the classification results are not affected by the size of the images, Table VI shows the results we obtained by testing the detector on the larger versions of the images of size 2144×1424. A well known problem with forensic tools

TABLE VI

AUC VALUES OF THE REFINED CLASSIFIER FOR LARGER IMAGES.

Attack	D-JPEG	1st order MPA	Attack in [17]	wavelet denoise	median filtering	copyMove 256x256
AUC	0.99	0.98	0.96	0.91	0.90	0.98
Attack	CLAHE	resize BIC0.9	resize BIL0.9	resize NN0.9	rotation BIC5	rotation NN5
AUC	0.91	0.91	0.93	0.93	0.91	0.90
Attack	zoom 1.2	crop align	crop no align	mirror	blur	seam-carving
AUC	0.97	0.93	0.93	0.99	0.99	0.95

TABLE VII

AUC OF THE REFINED MPA-AWARE CLASSIFIER TESTED ON A MIXTURE OF IMAGED FROM RAISE-2K ($\approx 30\%$) AND DRESDEN ($\approx 70\%$).

Attack	D-JPEG	wavelet denoise	resize BIC0.9	resize NN0.9	rotation BIC5	rotation NN5
AUC	0.92	0.87	0.91	0.89	0.92	0.89

based on machine learning, is that they may be affected by the problem of database mismatch [21], that is, the classifier has poor performance when tested with images coming from a different dataset with respect to training. To verify that our system is not affected by this problem, we tested the refined MPA-aware classifier trained on RAISE-2K dataset using images from the Dresden database. Starting from the 752 uncompressed images made available in that database, with

ISBN 978-0-9928626-7-1 © EURASIP 2017

size 1936×1296, we generated single compressed, double compressed and attacked images according to the same procedure described so far. Table VII shows the performance of the classifier when tested on a mix of images taken from both RAISE-2K and Dresden datasets for the most dangerous attacks among those considered in Table III. As we can see the results do not differ much from those obtained using images from RAISE-2K dataset only.

V. CONCLUDING REMARKS

We presented an adversary-aware double JPEG detector able to work even in the presence of heterogenous processing and C-F attacks. We observe that we have conducted our tests in a rather controlled scenario: training and testing with the same $QF2$ and $QF1 < QF2$. As regards the value of $QF2$, the performance may decrease in case of mismatch between training and testing. However, $QF2$ is generally known to the defender (since it can be derived from the JPEG bitstream or reliably estimated from the image), then many versions of the detector can be trained and used for different values of $QF2$. Regarding $QF1$, performance degrade when the detector is tested with $QF1 > QF2$. To get good performance in this case, examples of images for which $QF1 > QF2$ must be included in the training set, even of the overall accuracy may decrease a bit (about 2% of the AUC value according to some preliminary tests we carried out[3]). In practice, better generalisation capabilities can be achieved at the price of a reduction of performance. It is worth observing that the proposed approach is not meant for localization, and the performance are expected to decrease on small images (or patches). As future work, we plan to pass from detection to localisation and extend the idea to Convolutional Neural Networks (CNNs). Exploring the link of our approach with one-class classifiers is another interesting research direction.

ACKNOWLEDGMENT

This work has been partially supported by a research sponsored by DARPA and Air Force Research Laboratory (AFRL) under agreement number FA8750-16-2-0173. The U.S. Government is authorised to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA and Air Force Research Laboratory (AFRL) or the U.S. Government.

REFERENCES

[1] R. Böhme and M. Kirchner, *Counter-Forensics: Attacking Image Forensics*. New York, NY: Springer New York, 2013, pp. 327–366. [Online]. Available: http://dx.doi.org/10.1007/978-1-4614-0757-7_12

[2] M. Barni and F. Pérez-Gonzàlez, "Coping with the enemy: Advances in adversary-aware signal processing," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 8682–8686.

[3]The results are referred to the MPA-aware detector in Section IV-B

[3] A. C. Popescu and H. Farid, "Statistical tools for digital forensics," in *Proceedings of the 6th International Conference on Information Hiding*, ser. IH'04. Berlin, Heidelberg: Springer-Verlag, 2004, pp. 128–147. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-30114-1_10

[4] S. Milani, M. Tagliasacchi, and S. Tubaro, "Antiforensics attacks to Benford's law for the detection of double compressed images," in *Proceedings of ICASSP 2013, IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 3053–3057.

[5] C. Pasquini, P. Comesaña-Alfaro, F. Pérez-González, and G. Boato, "Transportation-theoretic image counterforensics to first significant digit histogram forensics," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 2699–2703.

[6] B. Li, Y. Q. Shi, and J. Huang, "Detecting doubly compressed JPEG images by using mode based first digit features," in *Multimedia Signal Processing, 2008 IEEE 10th Workshop on*. IEEE, 2008, pp. 730–735.

[7] M. Barni, M. Fontani, and B. Tondi, "Universal counterforensics of multiple compressed JPEG images," in *International Workshop on Digital Watermarking*. Springer, 2014, pp. 31–46.

[8] ——, "A universal technique to hide traces of histogram-based image manipulations," in *Proceedings of the on Multimedia and Security*, ser. MMSec '12. New York, NY, USA: ACM, 2012, pp. 97–104. [Online]. Available: http://doi.acm.org/10.1145/2361407.2361424

[9] C. Chen, Y. Q. Shi, and W. Su, "A machine learning based scheme for double JPEG compression detection," in *19th International Conference on Pattern Recognition, 2008. ICPR 2008.* IEEE, 2008, pp. 1–4.

[10] M. Barni, Z. Chen, and B. Tondi, "Adversary-aware, data-driven detection of double JPEG compression: How to make counter-forensics harder," in *2016 IEEE International Workshop on Information Forensics and Security (WIFS)*. IEEE, 2016, pp. 1–6.

[11] 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.

[12] J. Kodovskỳ and J. Fridrich, "Steganalysis of JPEG images using rich models," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2012, pp. 83 030A–83 030A.

[13] J. Kodovsky, J. Fridrich, and V. Holub, "Ensemble classifiers for steganalysis of digital media," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 432–444, 2012.

[14] T. Pevny, P. Bas, and J. Fridrich, "Steganalysis by subtractive pixel adjacency matrix," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 2, pp. 215–224, June 2010.

[15] T. Pevny and J. Fridrich, "Merging Markov and DCT features for multi-class JPEG steganalysis," *Proc. SPIE*, vol. 6505, pp. 650 503–650 503–13, 2007.

[16] M. K. Mihcak, I. Kozintsev, and K. Ramchandran, "Spatially adaptive statistical modeling of Wavelet image coefficients and its application to denoising," in *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1999.*, vol. 6. IEEE, 1999, pp. 3253–3256.

[17] M. C. Stamm, S. K. Tjoa, W. S. Lin, and K. R. Liu, "Undetectable image tampering through JPEG compression anti-forensics," in *2010 17th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2010, pp. 2109–2112.

[18] G. Valenzise, M. Tagliasacchi, and S. Tubaro, "The cost of JPEG compression anti-forensics," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2011, pp. 1884–1887.

[19] D.-T. Dang-Nguyen, C. Pasquini, V. Conotter, and G. Boato, "Raise: A raw images dataset for digital image forensics," in *Proceedings of the 6th ACM Multimedia Systems Conference*, ser. MMSys '15. New York, NY, USA: ACM, 2015, pp. 219–224. [Online]. Available: http://doi.acm.org/10.1145/2713168.2713194

[20] T. Gloe and R. Böhme, "The 'Dresden Image Database' for benchmarking digital image forensics," in *Proceedings of the 2010 ACM Symposium on Applied Computing*, ser. SAC '10. New York, NY, USA: ACM, 2010, pp. 1584–1590. [Online]. Available: http://doi.acm.org/10.1145/1774088.1774427

[21] J. Kodovskỳ, V. Sedighi, and J. Fridrich, "Study of cover source mismatch in steganalysis and ways to mitigate its impact," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2014, pp. 90 280J–90 280J.