On the Impact of AI-based Compression on Deep Learning-based Source Social Network Identification

Alexandre Berthet, Chiara Galdi, Jean-Luc Dugelay
dept. of Digital Security, *EURECOM*
Sophia Antipolis, France
{berthet, galdi, dugelay}@eurecom.fr

***Abstract*—Recognition of the social network of origin of an image is a relatively recent topic that is part of the techniques that fall under the umbrella of digital image forensics. It consists of the classification of images according to the social network on which they were posted. In contrast with other topics of digital image forensics, there are no works addressing counter forensic for source social network identification. Thus, we analyse the impact of image manipulations on its performances. We focus our study on AI-based compression, which tends to become the new compression solution with the upcoming standard JPEG AI. To conduct a fair analysis, we compare the AI-based compression with the conventional legacy JPEG compression, and also include three other manipulations: median filtering, Gaussian blurring, and additional white Gaussian noise, which are often used to assess the robustness of digital image forensic methods. We define two sets of parameters based on the resulting image quality in terms of structural similarity, which correspond respectively to attacks with strong and limited image degradation. In the context of strong downgrade of the image quality, all the manipulations lead to similar decrease in performance, while for attacks that preserve image quality, AI-based compression is able to reach a drop in identification rate twice higher than the other manipulations.**

***Index Terms*—Artificial Intelligence, AI compression, Deep Learning, Source Social Network, Digital Image Forensics**

I. Introduction

Since the beginning of the century, with the rise of the internet and remote communication, social media and messaging applications are more and more present in our daily lives. They are useful to communicate, to get information or to have a look at the news. However, this constant use of social networks such as Facebook or Twitter has also led to new types of cybercrime, such as cyberbullying or revenge porn. These crimes are committed through these platforms, on which it is often difficult to trace who originally uploaded such content. In this context, identifying the source social network of an image could be relevant. Source social network identification (SSNI) is an area of digital image forensics, whose goal is to find from which social network the images were downloaded. The literature is already vast, and different methods have been proposed, which are mostly based on deep learning architectures.

Most of these methods base their classification on specific image artefacts such as noise residuals or photo response non-uniformity (PRNU) [1], or discrete cosine transform (DCT) features. In fact, social networks apply different processing manipulations when users upload images. These manipulations consist mostly in resizing and JPEG compression, which leave traces on the uploaded images. Each social network has its own processing algorithm, which results in unique traces from one social network to another. In other areas of digital image forensics, such as source camera recognition, methods are often evaluated according to their robustness to manipulations (i.e. post-processing operations). Most of these manipulations are used as attacks to hide artefacts essential for identification. In the case of social network identification, JPEG compression could be one of the key attacks as it is expected to impact DCT-based features as well as indirectly impact the extraction of the distinctive noise pattern.

The JPEG organization [2] announced the upcoming new compression standard that will be based on deep learning architectures: JPEG AI[1]. This has been followed by the proposal of various AI-based compression methods, which mainly exploit auto-encoders [3]. The new standard will aim at providing better compression for both human and machine points of view. The proposed architectures all consist of an encoder, a bottleneck and three outputs: the (usual) decoder and two others for image processing and computer vision tasks. The proposals have been presented and discussed at the 96^{th} JPEG meeting (July 2022), and the JPEG AI standard should be available in the next few years (April 2024).

Studies on the impact of non-AI image coding on the performance of AI-based computer vision solutions such as detection and classification have recently received attention [4] [5] [6], although rarely in the context of digital image forensics. Soon AI-based techniques such as compression will be in common use and may have an impact in various fields. For instance, in [7], Bousnina *et al.* analyse the impact of different image AI coding techniques on face recognition. In [8], Berthet and Dugelay consider whether AI-based compression can be a new unintended counter-attack for JPEG-related image tampering detectors. Robustness assessment is important in digital image forensic tasks, including SSNI, as social network (SN) classification is based on artefacts too. Applying manipulations could mask or even delete traces

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

that are unique to each social network. Thus, the classifiers could be fooled and their performances affected. Despite its relevance, there is no work to date addressing the topic of post-processing operations as an attack on SSNI. The objective of this paper is to evaluate the impact of such manipulations. The work is particularly focused on AI-based compression, which will be the next standard of the JPEG organization and will likely be democratized in the next few years to essential areas of image processing. The contributions of this paper are as follows:

- For the first time, the impact of AI-based compression on SSNI is studied;
- The impact of AI-based compression on the performance of SSNI is compared to traditional JPEG compression, as well as to three other manipulations. Two groups of manipulations are created based on the resulting image quality in terms of SSIM: corresponding to *limited* and *strong* visual image quality degradation.

The extensive testing on three databases dedicated to social networks. The methods and the data are publicly available, and their links are provided, which makes our evaluation fully reproducible.

II. Related Works

A. Challenges in Digital Image Forensics

For a long time, digital image forensic methods have used mathematical statistics to extract appropriate features for classification. With the emergence of deep learning, their architecture has shifted, in particular to convolutional neural networks (CNNs), which are rather known for their effectiveness for image analysis. Along with CNNs, pre-processing modules have often been used to highlight the essential artefacts [9]. Despite their recent novelty, SSNI already achieves high classification performances. These results mainly come from the analysis of artefacts related to social network processing algorithms, which often include resizing, cropping, and compression operations. These processes often create artefacts that can be identified through some image analysis operations, such as noise residuals and DCT features analysis. In what follows, we describe state-of-the-art methods that analyse image artefacts for source digital camera recognition – which is a fundamental step that enable many tasks of digital image forensics – as well as for SSNI and we focus in particular on methods that exploit PRNU and DCT features. In addition, we report the evaluation results that were obtained on three different databases for social network identification: Ucid Social, Social Public[2], and IPLAB[3], which all share three classes (Flickr, Facebook, and Twitter).

B. Photo Response Non-Uniformity

Source digital camera identification (SDCI) is a fundamental part of digital image forensics (DIF). What was discovered for SDCI, enabled the development of many other image forensic solutions for various tasks, including SSNI. That is why in this Section we discuss about the origin of PRNU extraction – which is one of the more reliable image characteristics for DIF – and its current open challenges in the era of deep learning. The sensor noise is the result of three main components: pixel defects, fixed pattern noise (FPN), and photo response non-uniformity (PRNU). It is a distinctive pattern due to imperfections in the silicon wafer during the sensor manufacturing, different even among cameras of the same model. These imperfections imply that the pixels have different sensitivities to light. A distinctive pattern can be extracted by analysing the image in the frequency domain and by selecting those frequencies that are more likely to be associated with the sensor noise. The method was first presented by Lukas et al. in [1] for the purpose of identifying the source digital camera of an image. Several works in the field of media forensics have been exploring the difficulties that arise when attempting to apply existing methods on more complex real world scenarios. In [10], Lorch *et al.* propose a robust camera model identification method based on sparse Gaussian processes. Many methods assume that the image under analysis originates from a given set of known camera models. In practice, however, a photo can come from an unknown camera model, or its appearance could have been altered by unknown post-processing. In such a case, forensic detectors are prone to fail silently. This work propose a way to mitigate silent failures by using a rejection mechanism for unknown examples.

C. Challenges in Source Social Network Identification

The work by Karunakar *et al.* [11], addresses the problem of classifying whether an image was uploaded or not on a social network. The authors propose a deep learning based approach that learns the unique traces from the images transformed into the discrete cosine and wavelet domains, to investigate whether the image under test originates directly from a camera or from a specific social network site.

Another open issue in SSNI, is addressed in [12]. In this work the authors address the problem of tracking multiple sharing on SN, by extracting specific traces left by each SN within the image file, due to the process each of them applies, to perform a multi-class classification. Innovative strategies, based on deep learning, are proposed and satisfactory results are achieved in recovering till triple up-downloads. In [13], Verde *et al.*, address the same issue by proposing a supervised framework for the reconstruction of image sharing chains on SN platforms. The system is structured as a cascade of backtracking blocks, each of them tracing back one step of the sharing chain at a time. Blocks are designed as ensembles of classifiers trained to analyse the input image independently from one another by leveraging different feature representations that describe both content and container of the media object. Individual decisions are then properly combined by a late fusion strategy. Results highlight the advantages of employing multiple clues, which allow accurately tracing back up to three steps along the sharing chain.

[2] http://lci.micc.unifi.it/labd/2015/01/trustworthiness-and-social-forensic/

[3] https://iplab.dmi.unict.it/popularitychallenge/

In [14], Amerini *et al.* based their method on the analysis of DCT-based features. The pipeline is a CNN of two convolutional layers followed by three fully connected layers plus a preprocessing module, which is dedicated to extracting DCT-based features. First, patches of size 64×64 pixel are cropped from the image and DCT coefficients are extracted from 8×8 blocks. Then, histograms consisting of 101 bins (range -50 to +50 for the coefficients) are computed for the first nine spatial frequencies of the blocks. Finally, a vector of size 909×1 is computed from these histograms and fed to the network.

The same authors have proposed, in [15], to use the PRNU to recognize the source social network. Classification is performed through a CNN of four convolution layers, followed by two fully connected layers. First, the residual noise is obtained using a PRNU extraction module, and then small patches of size 64×64 pixel are cropped from the image and fed into the network. The PRNU fingerprint is computed from a group of images by applying a minimum variance estimator from their noise residuals. Finally, the residual noise is obtained from the original I_i images and their denoised versions I_i^{den}, and the reference noise pattern of a device $\hat{K}$ can be obtained by averaging the residual noise over a set of images captured by that device (see equations 1 and 2).

$$W_i = I_i - I_i^{den} \tag{1}$$

$$\hat{K} = \frac{\sum_{i=1}^{M} W_i I_i}{\sum_{i=1}^{M} (I_i)^2} \tag{2}$$

In [16], the proposed method improves the recognition performances by combining both DCT and Noise-Residual features. The proposed architecture is a two-stream CNN. The DCT-based subnetwork is fed by a vector of size 1980, of normalised DCT coefficients. The Noise-Residual-based subnetwork is the CNN-based noise extractor *Noiseprint* [17], which is particularly well known and has already proven its effectiveness in digital image forensic tasks. *Noiseprint* is a pre-trained network able to provide the residual noise from a single image. The two-stream neural network [16] achieved better performances in most cases against the DCT-based network [14] (see the confusion matric in Tab. I, the results of [15] are not reported in the table as it was not assessed on all databases, and did not perform as well as the others. However, more details can be found in [16].)

For further details on image forensics and on image forensics on social media platforms, extensive reviews are available in [18] and [19].

III. Proposed Framework

With this work we aim to provide a first evaluation of the impact of AI-based image compression on SSNI. Therefore, we set up a framework that brings together a social network identification method, an AI-based compression model, and three databases. Moreover, all the elements of our framework are publicly available and reproducible. The proposed framework is illustrated in Fig. 1.

TABLE I: Image classification confusion matrix: comparison between [16] and [14] on the *UCID Social*, *Social Public*, and *IPLab* datasets.

UCID Social	[16]	[14]	[16]	[14]	[16]	[14]
	Facebook		*Flickr*		*Twitter*	
Facebook	**98.20%**	97.37%	0.20%	0.00%	1.40%	2.63%
Flickr	0.00%	0.00%	**100%**	**100%**	0.00%	0.00%
Twitter	0.00%	0.00%	0.00%	0.00%	**100%**	**100%**

Social Public	[16]	[14]	[16]	[14]	[16]	[14]
	Facebook		*Flickr*		*Twitter*	
Facebook	**91.21%**	88.24%	0.00%	0.00%	8.79%	11.76%
Flickr	0.00%	0.99%	**98.15%**	97.03%	1.85%	1.98%
Twitter	0.10%	0.00%	1.23%	0.00%	98.67%	**100%**

IPLab	[16]	[14]	[16]	[14]	[16]	[14]
	Facebook		*Flickr*		*Twitter*	
Facebook	**97.86%**	96.01%	2.14%	3.99%	0.00%	0.00%
Flickr	2.45%	1.68%	**97.55%**	97.06%	3.28%	1.26%
Twitter	0.00%	0.00%	0.00%	1.26%	**100%**	98.74%

As for the SSNI method, we selected the one by Berthet *et al.* [16][4]. This method combines the analysis of two domains (noise residuals and DCT), which are particularly affected by the application of image manipulations.

In terms of AI-based compression, the literature is vast, with many open source solutions. However, the JPEG organization has reviewed early learning-based compression solutions to summarize recent progress. We selected the HiFiC [20] (High-Fidelity Compression[5]) method, as it was the first to use a generative adversarial network (GAN), which achieves high perceptual fidelity for the reconstructed images. Mentzer *et al.* [20] explained that HiFiC was trained using several quality measures (PSNR, MS-SSIM, etc.) while maintaining a low bit rate (half that of state-of-the-art methods). Moreover, this AI-based method can compress images with three different quality levels: low, medium, and high. To conduct a thorough analysis, we compared AI-compression with the legacy JPEG compression. We also included three other manipulations in our study: additional white Gaussian noise (AWGN), Gaussian blur (GB), and median filtering (MF), which are often used in literature for robustness analysis of digital image forensics [21]. Since there are different image quality levels for HiFiC compression, we selected one that can be considered as a strong attack, that is *Low*, which means with a strong decrease of the image quality, and the other one, which has limited visual degradation of the image, that is *High*. This selection is based on the resulting image quality of both compression levels, measured with structural image similarity (SSIM). SSIM is used, instead of other image quality measures such as PSNR, because in our experiments it was found to be more suitable to identify the different image quality levels for all manipulations. Also, in [22], Horè and Ziou show that SSIM and PSNR are closely related and that SSIM is more sensitive than PSNR to JPEG compression, which is the main focus of this study. In order to have a fair comparison, we selected two sets of parameters for the other manipulations as well, which correspond to equal image quality in terms of SSIM. SSIM is

[4]https://github.com/francescotescari/social

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

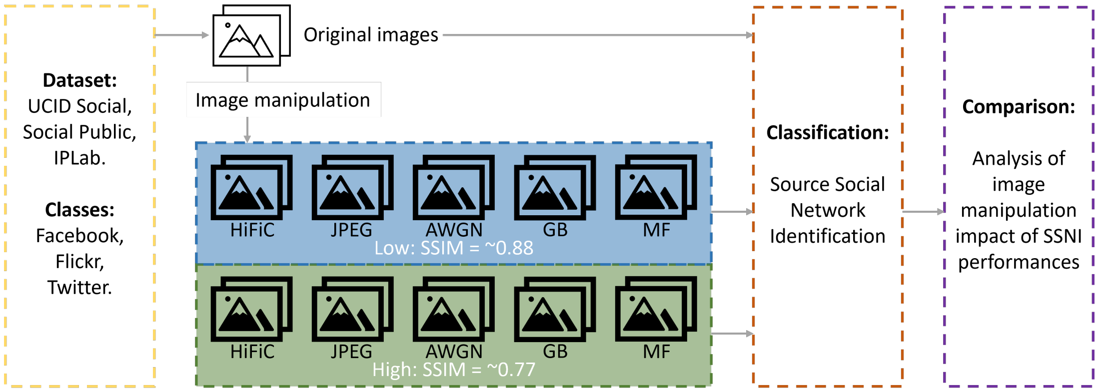


Fig. 1: Block diagram of the proposed framework for image manipulation impact analysis.

TABLE II: Detailed values of SSIM for the two groups of quality level for each manipulation. First group = strong attacks; Second group = moderate attacks. "QF" = JPEG quality factor; "σ" = variance; "size" = kernel size.

Manipulations	HiFiC	JPEG	AWGN	GB	MF
First group	*High*	*QF = 25*	σ = 2.5	*size = 3*	*size = 3*
SSIM	0.88	0.89	0.88	0.9	0.87
Second group	*Low*	*QF = 10*	σ = 6	*size = 7*	*size = 5*
SSIM	0.78	0.79	0.77	0.77	0.75

defined as follows:

$$SSIM(I,J) = [l(I,J)]^{\alpha} \dot{[} c(I,J)]^{\beta} \dot{[} s(I,J)]^{\gamma} \tag{3}$$

Where $l(I,J) = \frac{2\mu_I \mu_J + C_1}{\mu_I^2 + \mu_J^2 + C_1}$, $c(I,J) = \frac{2\sigma I \sigma_J + C_2}{\sigma_I^2 + \sigma_J^2 + C_2}$ and $s(I,J) = \frac{\sigma_{IJ} + C_3}{\sigma_I + \sigma_J + C_3}$ with μ_I, μ_J, σ_I, σ_J and σ_{IJ} are the mean of I and J, the standard deviation of I and J, and the covariance of I and J, respectively. The luminance index compares the average brightness of the pixels in the reference and distorted images. The contrast index and the structural index, juxtapose the standard deviation and the covariance of the pixel values in the reference and distorted images respectively.

For our evaluation, we chose three publicly available databases, which also allow comparison with the state of the art: *UCID Social*, *Social Public*[6], and *IPLab*[7]; and in particular we selected three classes they have in common: *Facebook*, *Flickr*, and *Twitter*. *UCID Social* is obtained from the database of uncompressed color images (UCID) [23], which contains 1,338 images that were first compressed with 10 different JPEG quality factors (50 to 95 with a step of 5), and then uploaded to the desired social networks. This way of creating a social network database leads to a *controlled environment dataset*, where images are first selected, then uploaded to a social network, and finally downloaded. *IPLab* is also a *controlled environment dataset* that was generated from 240 images, uploaded and then downloaded to 8 different social networks. In contrast, *Social Public* is an *uncontrolled environment dataset*, as it is composed of images directly from social networks. It contains 3000 images downloaded from three different social networks.

[6]http://lci.micc.unifi.it/labd/2015/01/trustworthiness-and-social-forensic/
[7]https://iplab.dmi.unict.it/popularitychallenge/

In order to obtain the classification for an image, all the 64×64-pixel non-overlapping patches from the image are classified using the trained CNN. Then the majority-voted class is considered as the predicted class for the entire image. Finally, since the images are processed by different SNs, and the processing includes resizing in some cases, the final number of extracted patches is going to be different for each class, which is not ideal for training neural networks as this can introduce bias towards the most represented classes. Thus, an infinite-loop generator for each class, with reshuffling, is created. Then, the three-class generators are interleaved, by taking a sample from each class in turn, to create a random patch dataset generator of *perfectly balanced* batches of training data. The same class-balance issue can be found for the validation and test set, as the metrics used (e.g. recall and accuracy) can output misleading results if the validation classes are unbalanced. Therefore, validation and test sets are balanced by defining specific class weights [16].

IV. Experimental Results

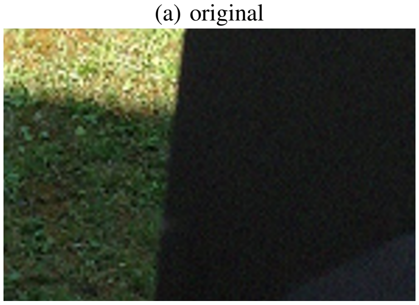
Based on the elements from our proposed framework, we have applied AI-based and legacy-JPEG compression, as well as the other manipulations, to images from the selected databases. As we explained earlier, based on the resulting images quality obtained with HiFiC-High and -Low, we have selected two set of parameters for each manipulation and created two groups that match the image quality obtained for AI-compression (see Tab. II). Some manipulations, have less visual impact than others (see Fig. 2). The first group corresponds to attacks with limited visual degradation and an average SSIM of 0.88, while the second group corresponds to strong attacks resulting in image degradation with an average SSIM of 0.77.

The results in terms of classification accuracy are reported in Tab. III, according to the database, the social network, and the image manipulation used. In order to compare all post-processing operations and their impact on recognition

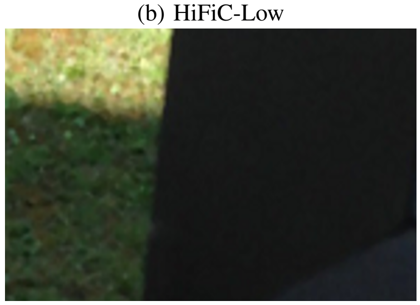
(a) original

(b) HiFiC-Low

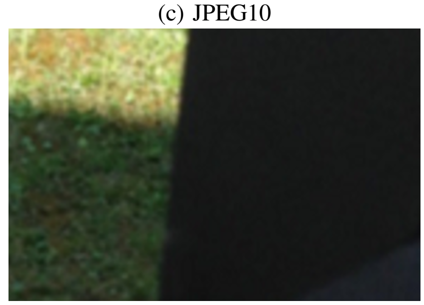
(c) JPEG10



(d) AWGN6



(e) MF5



(f) GB7

Fig. 2: Comparison of visual image quality for each manipulation of an image from IPLAB - flickr. (a) original (b) HiFiC-Low (c) JPEG QF = 10 (d) AWGN $\sigma = 6$ (e) MF kernel size = 5 (f) GB kernel size = 7. Please, refer to the HD version online.

performance, we also report $Diff\_acc$, which is the average difference of recognition accuracy between the classification accuracy obtained on original images Acc_o and on the manipulated images Acc_m (see Eq. 4) over the K social networks.

$$Diff\_acc(j) = \frac{\sum_{i=1}^{K} Acc_o(i) - Acc_m(i,j)}{K} \tag{4}$$

Where $i = 1, 2, 3, ..., N$ is the image index and $j = 1, 2, 3, ..., M$ indicates one of the considered M manipulations. This metric shows the drop of SSNI accuracy when a manipulation is applied on the images. Moreover, as we selected multiple databases, different in terms of composition of environment (i.e. *controlled* and *uncontrolled*), we also report the average $Diff\_acc$ over the different D databases, which corresponds to the average drop of performance per manipulation along the three databases.

$$AVG(Diff\_acc(j)) = \frac{\sum_{d=1}^{D} Diff\_acc(j,d)}{D} \tag{5}$$

When considering strong attacks (i.e. second group) and an average SSIM across datasets of 0.77, similar results are obtained with the different image manipulations. In fact, except for the AWGN ($\sigma = 6$), all the post-processing operations have a $AVG(Diff\_acc)$ percentage of about 24 points. In contrast, for the first group, the decrease of performance is not the same for each manipulation. When considering attacks with limited visual degradation of the image (first group), HiFiC-High is leading to the largest decrease in performance. In fact, the $AVG(Diff\_acc)$ percentage is of ~ 20 points for HiFiC-High against ~ 10 for the other manipulations (about twice higher). Interestingly, in terms of $Diff\_acc$, it is on the IPLab database and with MF that the largest average drop of performance is achieved. The results obtained show that in average, over all databases and all image manipulations, AI-based compression is able to strongly interfere with SSNI even when it applies a moderate image compression with very little image quality degradation, which is very likely to be used in real world scenarios. And thus, AI-based compression, which is not a malicious attack, poses a serious threat to deep learning-based SSNI systems.

V. Conclusion

This paper is the first to address the possible impact of AI-based techniques for compression in the field of SSNI. Of course, the purpose of the paper is broader and is to draw attention to the fact that soon AI-based techniques such as compression will be in common use and may have an impact in various fields [7], not only the one analysed in this paper. In a context of strong quality downgrade, all the operations lead to a similar decrease in performance, while HiFiC-High reached the highest decrease in performance for attacks with limited degradation. Therefore, AI-based compression should be included to the group of post-processing operations to assess the robustness of digital image forensic methods, and notably for SSNI.

Concerning future extension of this work, we aim at improving the robustness of SSNI and to investigate the more complex scenario when an image or video undergoes multiple uploading/downloading to SNs or multiple image post-processing manipulations.

TABLE III: Image source social network classification. This table reports the comparison of the different performance values obtained by [16] after applying different image manipulations on the *UCID Social*, *Social Public*, and *IPLab* datasets, for each class (*Facebook*, *Flickr* and *Twitter*). First and second attack groups are to be considered separately for the selection of the max values as they represent separated experiments.

UCID Social	**Original**	**HiFiC**		**JPEG**		**AWGN**		**GB**		**MF**	
		High	Low	25	10	2.5	6	3	7	3	5
Facebook	98.20%	76%	73.6%	89.9%	71.3%	86.9%	82.31%	97.3%	70.6%	92.6%	76.4%
Flickr	100%	71.3%	60.1%	89.5%	69.4%	95.9%	71.3%	81.3%	72%	84.1%	72.7%
Twitter	100%	76%	73.3%	90.6%	72%	94.3%	83%	86%	69.5%	89.4%	78.6%
$Diff\_acc$		**24.97**	**30.4**	9.4	28.5	7.03	20.53	11.2	28.7	10.7	23.5
Social Public	**Original**	**HiFiC**		**JPEG**		**AWGN**		**GB**		**MF**	
		High	Low	25	10	2.5	6	3	7	3	5
Facebook	91.21%	80.6%	77.5%	82.8%	74.4%	93.1%	85.6%	83.5%	76.1%	85.4%	75.8%
Flickr	98.15%	70.6%	74%	80.9%	75.4%	93.3%	83.1%	80.8%	86%	91.5%	71.8%
Twitter	98.67%	66.9%	54.6%	78.6%	67.9%	91.2%	84.6%	78.7%	68.7%	81.5%	68.2%
$Diff\_acc$		**23.31**	**27.31**	14.91	23.44	3.48	11.58	15.01	19.08	9.88	24.08
IPLab	**Original**	**HiFiC**		**JPEG**		**AWGN**		**GB**		**MF**	
		High	Low	25	10	2.5	6	3	7	3	5
Facebook	97.86%	70.6%	62%	86.2%	82.1%	98.9%	74.5%	87.6%	75.9%	78.5%	69.4%
Flickr	97.55%	91.4%	85.6%	86.4%	86.2%	73%	68.1%	92.1%	82%	87.2%	66%
Twitter	100%	95%	89.9%	91.8%	72.8%	88%	72.7%	76.7%	55.5%	74.1%	71.8%
$Diff\_acc$		12.80	19.30	10.34	18.1	11.84	26.70	13	27.34	**18.54**	**29.40**
$AVG(Diff\_acc)$		**20.36**	**25.67**	11.55	23.35	7.45	19.6	13.07	25.04	13.04	25.66

References

[1] Jan Lukas, Jessica Fridrich, and Miroslav Goljan, "Digital camera identification from sensor pattern noise," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 205–214, 2006.

[2] ISO/IEC JTC1/SC29/WG1 N100095, "Final call for proposals for jpeg ai," *Online*, 2022.

[3] 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*. SPIE, 2020, vol. 11353, pp. 164–176.

[4] Matt Poyser, Amir Atapour-Abarghouei, and Toby P. Breckon, "On the impact of lossy image and video compression on the performance of deep convolutional neural network architectures," in *2020 25th International Conference on Pattern Recognition (ICPR)*, Los Alamitos, CA, USA, jan 2021, pp. 2830–2837, IEEE Computer Society.

[5] Max Ehrlich, Larry Davis, Ser-Nam Lim, and Abhinav Shrivastava, "Analyzing and mitigating jpeg compression defects in deep learning," in *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2021, pp. 2357–2367.

[6] Sara Mandelli, Nicolò Bonettini, Paolo Bestagini, and Stefano Tubaro, "Training cnns in presence of jpeg compression: Multimedia forensics vs computer vision," in *2020 IEEE International Workshop on Information Forensics and Security (WIFS)*, 2020, pp. 1–6.

[7] Naima Bousnina, João Ascenso, Paulo Lobato Correia, and Fernando Pereira, "Impact of conventional and ai-based image coding on ai-based face recognition performance," in *2022 10th European Workshop on Visual Information Processing (EUVIP)*, 2022, pp. 1–6.

[8] Alexandre Berthet and Jean-Luc Dugelay, "Ai-based compression: A new unintended counter attack on jpeg-related image forensic detectors?," in *2022 IEEE International Conference on Image Processing (ICIP)*, 2022, pp. 3426–3430.

[9] 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.

[10] Benedikt Lorch, Franziska Schirrmacher, Anatol Maier, and Christian Riess, "Reliable camera model identification using sparse gaussian processes," *IEEE Signal Processing Letters*, vol. 28, pp. 912–916, 2021.

[11] Manisha, A.K. Karunakar, and Chang-Tsun Li, "Identification of source social network of digital images using deep neural network," *Pattern Recognition Letters*, vol. 150, pp. 17–25, 2021.

[12] Quoc-Tin Phan, Giulia Boato, Roberto Caldelli, and Irene Amerini, "Tracking multiple image sharing on social networks," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 8266–8270.

[13] Sebastiano Verde, Cecilia Pasquini, Federica Lago, Alessandro Goller, Francesco GB De Natale, Alessandro Piva, and Giulia Boato, "Multi-clue reconstruction of sharing chains for social media images," *arXiv preprint arXiv:2108.02515*, 2021.

[14] Irene Amerini, Tiberio Uricchio, and Roberto Caldelli, "Tracing images back to their social network of origin: A cnn-based approach," in *2017 IEEE Workshop on Information Forensics and Security (WIFS)*, 2017, pp. 1–6.

[15] Roberto Caldelli, Irene Amerini, and Chang Tsun Li, "Prnu-based image classification of origin social network with cnn," in *2018 26th European Signal Processing Conference (EUSIPCO)*, 2018, pp. 1357–1361.

[16] Alexandre Berthet, Francesco Tescari, Chiara Galdi, and Jean-Luc Dugelay, "Two-stream convolutional neural network for image source social network identification," in *2021 International Conference on Cyberworlds (CW)*, 2021, pp. 229–237.

[17] Davide Cozzolino and Luisa Verdoliva, "Noiseprint: A cnn-based camera model fingerprint," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 144–159, 2020.

[18] Judith A. Redi, Wiem Taktak, and Jean-Luc Dugelay, "Digital image forensics: A booklet for beginners," *Multimedia Tools Appl.*, vol. 51, no. 1, pp. 133–162, jan 2011.

[19] Cecilia Pasquini, Irene Amerini, and Giulia Boato, "Media forensics on social media platforms: a survey," *EURASIP Journal on Information Security*, vol. 2021, no. 1, pp. 1–19, 2021.

[20] Fabian Mentzer, George D Toderici, Michael Tschannen, and Eirikur Agustsson, "High-fidelity generative image compression," *Advances in Neural Information Processing Systems*, vol. 33, pp. 11913–11924, 2020.

[21] Belhassen Bayar and Matthew C. Stamm, "Constrained convolutional neural networks: A new approach towards general purpose image manipulation detection," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 11, pp. 2691–2706, 2018.

[22] Alain Horé and Djemel Ziou, "Image quality metrics: Psnr vs. ssim," in *2010 20th International Conference on Pattern Recognition*, 2010, pp. 2366–2369.

[23] Gerald Schaefer, "An uncompressed benchmark image dataset for colour imaging," in *2010 IEEE International Conference on Image Processing*, 2010, pp. 3537–3540.