Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2023.1120000

Exploiting AC Histogram Statistics for Misalignment Estimation in Double JPEG Compressed Images

SEBASTIANO BATTIATO[1], (Senior Member, IEEE), FRANCESCO GUARNERA[2], and GIOVANNI PUGLISI[3]

[1]Department of Mathematics and Computer Science, University of Catania, Catania, Italy (e-mail: battiato@dmi.unict.it)
[2]Department of Mathematics and Computer Science, University of Catania, Catania, Italy (e-mail: francesco.guarnera@dmi.unict.it)
[3]Department of Mathematics and Computer Science, University of Cagliari, Cagliari, Italy (e-mail: puglisi@unica.it)

Corresponding author: Francesco Guarnera (e-mail: francesco.guarnera@unict.it).

The work of Francesco Guarnera has been supported by MUR in the framework of PNRR PE0000013, under project "Future Artificial Intelligence Research – FAIR".

ABSTRACT A critical task in forensics investigation is the recovery of the manipulation history of the image under analysis. To this aim, considering a typical life-cycle of a digital image, the estimation of the misalignment occurred between consecutive JPEG compressions can be considered an useful starting point to localize forgeries and retrieve information about the camera that took the picture through first quantization matrix estimation. In this work, starting from statistics computed from the AC histograms obtained applying a third JPEG compression, an effective and robust deep learning based approach devoted to estimate the aforementioned misalignment has been designed. Finally, to assess the performance of the proposed solution a series of tests has been conducted at varying of patch sizes, quantization matrices, employed datasets, and comparisons with state-of-the-art solutions.

INDEX TERMS Image forensics, non-aligned double JPEG compression, misalignment estimation, DCT coefficient analysis.

I. INTRODUCTION

TODAY a large number of images is acquired by digital devices, manipulated by editing software and uploaded to Social Networks and Instant Messaging platforms. Multiple JPEG compressions ([1]–[3]) are then usually applied on the considered image and part of the data contained in the original picture is actually lost. One of the most common problems in the image forensics field is then the reconstruction of the history of an image ([4]–[6]). Data related to the camera that took the picture ([7], [8]), together with the reconstruction of further processing, could provide useful hints about the reliability of the visual document under analysis. At first, the investigation could start with the detection of the traces left by double JPEG compression [9], followed by the retrieval of information about camera model with First Quantization Estimation (FQE) ([10]–[14]). Finally the source can be identified exploiting Photo Response Non Uniformity (PRNU) analysis ([7], [15], [16]). Multiple JPEG compressions are usually involved in the typical digital image life-cycle. Moreover, based on the presence of misalignment between consecutive quantizations, two different scenarios have been studied in literature: aligned and non-aligned double JPEG compression (see Figure 1). The estimation of these horizontal and vertical shifts is the crucial step to recover the manipulation history of the image under analysis. Such as example, tampered regions generated by conventional manipulations (e.g., splicing, copy-move) are

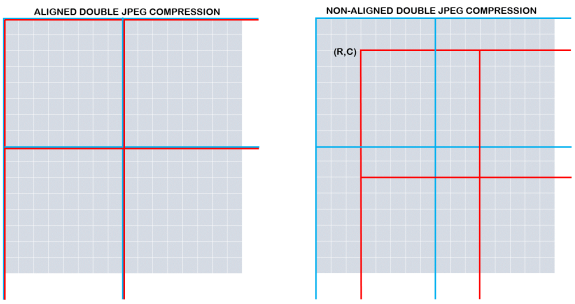


FIGURE 1. Aligned (on the left) and non-aligned (on the right) double JPEG compression scenarios. First compression DCT grid is represented in blue whereas the second compression one in red.

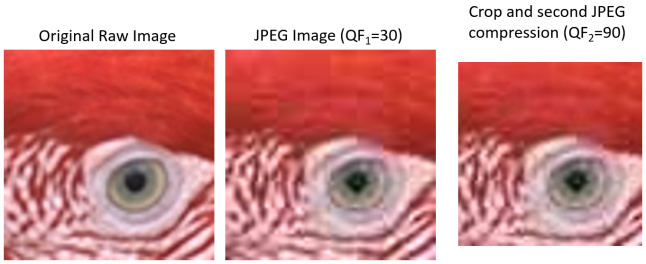


FIGURE 2. Due to the quantization process involved in the JPEG algorithm [24], visible artifacts appear at block boundaries (blocking artifacts) and have been exploited by pixel domain approaches to estimate the misalignment between DCT grids employed in consecutive compressions.

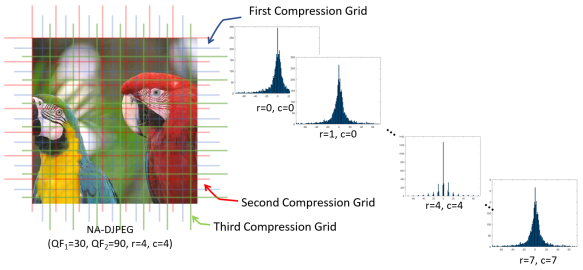


FIGURE 3. Frequency domain approaches general scheme. These methods usually crop the double JPEG compressed image considering all the possible shifts (i.e., $r, c \in \{0, \ldots, 7\}$), perform a further compression and exploit DCT (Discrete Cosine Transform) coefficient analysis to obtain the misalignment estimation.

usually misaligned with respect to the compression grid of the background image ([17], [18]). It is worth noting that also first quantization matrix estimation methods working in NA-DJPEG (non-aligned double JPEG) scenarios ([19]–[21]) usually exploit the information about grid misalignment.

In this work we propose a robust and effective approach able to estimate the misalignment occurred between consecutive JPEG compressions; to be effective, the proposed technique was tested not just on simple scenarios but mainly "on the wild" assessing performances with respect to different involved variabilities. Specifically, statistics computed from AC histograms are exploited as input feature of a CNN designed to perform the misalignment estimation. Experimental results conducted in challenging scenarios (i.e., small patches, custom quantization tables) and comparisons with state-of-the-art solutions, demonstrate the effectiveness of the proposed approach. It is worth noting that a preliminary version of this method, introducing the early ideas and performing preliminary tests, has been presented in our previous solution [22]. In this paper, the main ideas related to the employed features have been better analyzed. Starting from the new analysis, the fusion strategy related to the features computed from different DCT frequencies [23] has been improved employing a deep learning based solution. Moreover, the experimental section has been widely extended studying the performances of the proposed approach at varying of patch sizes (from 64×64 to 256×256), quantization matrices (both standard and customs) and employed datasets (medium and high resolution). Note that in all the aforementioned tests, results achieved by [22] have been always reported as a baseline to further evaluate the effectiveness of the proposed improvements.

The remainder of this paper is organized as follows: Section II reviews state-of-the-art solutions, Section III summarizes the standard JPEG algorithm, Section IV describes the proposed approach. Section V reports datasets and experimental protocol employed in the evaluation phase whereas Section VI describes experimental results and comparisons with state-of-the-art solutions in different scenarios. Finally, Section VII concludes the paper.

II. RELATED WORKS

Misalignment estimation in a NA-DJPEG scenario has been mainly studied in pixel domain. As can be seen in Figure 2, the quantization strategy applied in the JPEG algorithm [24] could generate visible artifacts at block boundaries (i.e., blocking artifacts). These artifacts, have been often exploited to estimate the misalignment between DCT grids employed in consecutive compressions. In [25], blocking artifact characteristics matrix (BACM) has been proposed and the analysis of its symmetry exploited to detect the shifts. Moreover, high-pass filters in pixel domain have been employed in several methods [26], [27] and improved selecting only homogeneous regions [21] to avoid being deceived by the content of the picture. DCT domain based approaches have been also designed to perform misalignment estimation. These methods, usually, crop the input double JPEG compressed image considering all the possible shifts, perform a further compression and exploit DCT coefficient analysis to estimate the misalignment (see Figure 3). In [19], the uniformity of a designed Integer Periodicity Map (IPM) has been studied to perform the shift estimation considering only DC term. The same authors in [17], in a scenario involving tampered regions inside the picture under investigation, estimate the misalignments through an expectation-maximization (EM) algorithm considering every possible grid shift. Also in this case only DC coefficients are exploited to limit the complexity of their solution. To improve the results also in challenging scenarios (e.g., small patches), methods considering all the AC modes, based on the analysis of similarities among DCT histograms related to aligned and non-aligned grids [28] exploiting also properties about their statistics [22] have been proposed.

It is worth noting that both pixel and DCT based state-of-the-art solutions have been usually designed and tested considering big or medium size pictures and standard quantization tables. These design choices could then limit their applicability "on the wild" scenarios where, as example, small image regions could be manipulated or custom matrices employed in the compression pipeline. To overcome these limitations, in this work, we developed a misalignment estimation approach exploiting statistics computed from AC histograms as input feature of a CNN employed as classifier.

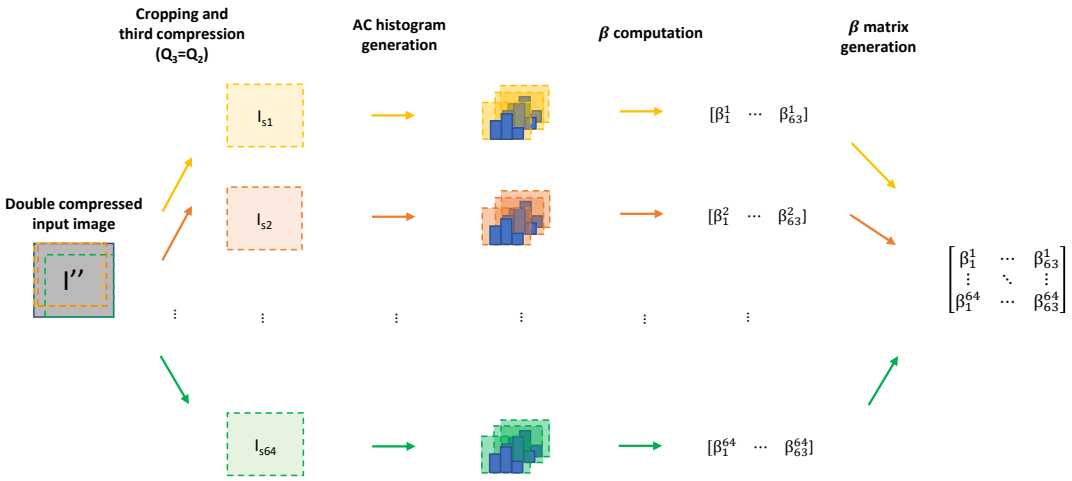


FIGURE 4. Overall scheme of the feature extraction strategy. At first, 64 images I_s are cropped from the double JPEG compressed input I'' considering all the possible shifts. These images are then further compressed with Q_2 as third quantization matrix and 63 AC histograms h_s^{cf} are generated. Considering all the AC frequencies and shifts a 64 $\times$ 63 matrix M_β is built computing β values from the related histograms h_s^{cf}.

III. JPEG NOTATION

Given as input a raw image I together with a 8×8 quantization matrix Q, JPEG compression [24] can be defined as a function $f_Q(I)$ providing as output a JPEG compressed image I'. As first step, the RGB image I is converted to a different color space (YCbCr) and partitioned into 8×8 non-overlapping blocks. Integer Discrete Cosine Transform (DCT) is then applied to each block and a pixel by pixel division by Q is performed. Finally, the results are rounded and encoded by classic entropy based engine. Due to the limited amount of information contained in the subsampled chromatic channels Cb and Cr, only luminance (i.e., Y channel) is considered in this paper. A double JPEG compressed image I'' can be hence obtained applying the aforementioned function two times $I'' = f_{Q_2}(f_{Q_1}(I))$, with Q_1 and Q_2 representing quantization matrices employed in the first and the second compression respectively. Moreover, we refer to QF_i as the quality factor associated to the standard quantization matrix [24] related to the i_{th} compression. Finally, based on the presence of a misalignment between DCT grids employed in the consecutive compressions, A-DJPEG and NA-DJPEG (i.e., aligned and non-aligned) scenarios are defined.

IV. PROPOSED METHOD

The main aim of the proposed solution is the estimation of the misalignment occurred between consecutive JPEG compressions in challenging conditions (e.g., small patches, custom quantization tables). A deep learning based solution has been then built considering statistics computed from AC histograms as input of a CNN designed to perform the misalignment estimation. Note that the designed input features, the main contribution of the proposed solution, are the key factor that permit to cope with variabilities in terms of image resolution and employed quantization matrix. Due to the goodness of the aforementioned features, even a simple classifier based on a CNN with a classical architecture can be employed to perform the classification task. In the following subsections we will describe in details the computed features, the employed datasets, and the neural network architecture with related implementation details.

A. FEATURES

Starting from a double JPEG compressed image I'' a third compression with $(Q_3 = Q_2)$ is performed considering all the possible shift pairs (r, c) with $r, c \in \{0, 1, \ldots, 7\}$. It has been observed that the shapes of DCT histograms built considering first and third block grids aligned usually differ from the ones obtained with grid misalignment. To properly describe the shape of these histograms, robust statistics computed from DCT coefficients can be exploited. Specifically, A zero-centred Laplace distribution [29] has been employed to model AC coefficient distributions:

$$f(x) = \frac{1}{2\beta} \exp\left(-\frac{|x|}{\beta}\right) \tag{1}$$

with β scale parameter obtained by maximum likelihood estimation close form solution computed as follows:

$$\beta = \frac{1}{N} \sum_{k=1}^{N} |x| \tag{2}$$

with N representing the number of involved samples.

More specifically, due to the presence of blocking artifacts (see Figure 2), the third compressed images obtained employing a crop operation with almost all shifts include some discontinuities and anomalies. Corresponding DCT coefficient

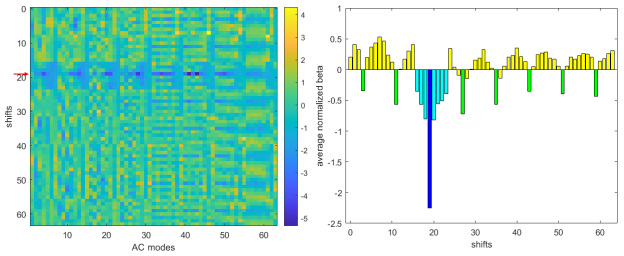


FIGURE 5. M_β **matrix with z-score normalization computed from a double compressed image with** $QF_1 = 70$, $QF_2 = 90$, **and** $s = 19$ **(on the left) and the relative average with respect to DCT modes (on the right). The minimum value usually corresponds to the correct shift (in blue). Moreover, values related to the correct row (in cyan) or column (in green) are often lower than the other ones (in yellow).**

values are then usually larger than ones related to the correct shift (i.e., first and third grid aligned). For any specific AC frequency, β values computed from the DCT histograms at varying of the cropping parameters can be used as powerful features to estimate the grid misalignment. DCT coefficients related to the correct shift usually tend to be closer to zero and the related β value corresponds to a minimum.

Starting from this simple idea, an effective feature extraction strategy can be then designed (see Figure 4). As first step, 64 cropped images I_s with $s = r \times 8 + c$ $(r, c \in \{0, 1, \dots 7\})$ are extracted from the double JPEG compressed picture I''. Later, an additional compression with $Q_3 = Q_2$ is performed and 63 AC histograms h_s^{cf} (where s and cf represent employed shifts and AC location indexes) are built. The shape of each AC histogram h_s^{cf} is associated to a specific β_s^{cf} exploiting eq. (2). Finally, taking into account all the AC positions and shifts, a 64×63 matrix M_β is generated. As can be seen from Figure 5 where an example of normalized M_β (column-wise z-score normalization) is depicted, this matrix can be effectively used to perform misalignment estimation. Specifically, fixed a specific AC frequency, the minimum value corresponds to the correct shift. Note that values related to the correct row or column are often lower than the other ones.

B. DATASETS

One of the main issues of machine learning based approaches have to cope with is the mismatch between training and real case conditions. Both architectural choices and data used for training have to be then carefully considered to achieve a satisfactory degree of generalization. To this aim, UCID (Uncompressed Color Image Dataset) [30] collection has been employed to train the proposed solution. This dataset, originally developed for Content Based Image Retrieval (CBIR), has been often exploited in forensics tasks due to the variety of its content in terms of natural scenes, man-made objects, indoor and outdoor environments. Starting from 1300 UCID images, applying a non-aligned double JPEG compression, training, validation and test sets have been generated according to a 70/15/15 ratio. For each raw input picture the following parameters have been then considered:

1) Crop size $\in \{256 \times 256, 128 \times 128, 64 \times 64\}$
2) First quantization matrix $Q_1 \in \{standard\}$
3) Second quantization matrix $QF_2 \in \{90, 80\}$
4) Misalignment $s \in \{0, 1, 2, 3, ..., 63\}$

where *standard* refers to standard quantization matrices with quality factors ranging from 50 to 95 at step of 5. Considering then all the involved parameters (i.e., crop size, Q_1, Q_2, misalignment) 4992000 $(1300 \times 3 \times 10 \times 2 \times 64)$ non-aligned double JPEG compressed images have been generated.

It is worth noting that further datasets not involved in the design of the proposed solution have been generated to test the generalization capability of the proposed solution as reported in Section VI-B.

C. NETWORK

As reported before, the main aim of the proposed solution is the estimation of the misalignment occurred between consecutive JPEG compressions in challenging conditions. Taking into account the partitioning into 8×8 blocks of the original picture due to the JPEG algorithm, 64 different shifts must be considered (see Figure 1).

Layer	Output Shape	Involved Parameters
Conv2D	(61, 62, 16)	kernel size=(3, 3)
Batch Normalization	(61, 62, 16)	epsilon=0.00001
Activation	(61, 62, 16)	type=relu
Conv2D	(59, 60, 16)	kernel size=(3, 3)
Batch Normalization	(59, 60, 16)	epsilon=0.00001
Activation	(59, 60, 16)	type=relu
Max Pooling 2D	(29, 30, 16)	pool size=(2, 2) strides=(2, 2)
Conv2D	(27, 28, 32)	kernel size=(3, 3)
Batch Normalization	(27, 28, 32)	epsilon=0.00001
Activation	(27, 28, 32)	type=relu
Conv2D	(25, 26, 32)	kernel size=(3, 3)
Batch Normalization	(25, 26, 32)	epsilon=0.00001
Activation	(25, 26, 32)	type=relu
Max Pooling 2D	(12, 13, 32)	pool size=(2, 2) strides=(2, 2)
Conv2D	(10, 11, 64)	kernel size=(3, 3)
Batch Normalization	(10, 11, 64)	epsilon=0.00001
Activation	(10, 11, 64)	type=relu
Dropout	(10, 11, 64)	value=0.5
Flatten	(7040)	
Dense	(64)	activation=softmax

TABLE 1. CNN architecture in terms of employed layers, dimension of the intermediate volumes and related parameter settings.

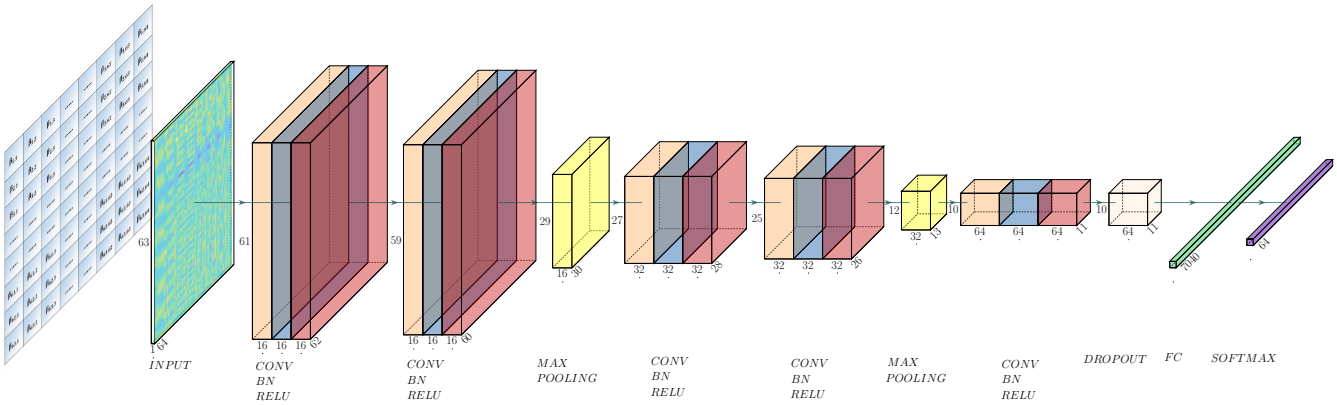


FIGURE 6. Overall architecture of the employed CNN. The input layer is the feature matrix M_β after z-score normalization. A block CB_i composed by the concatenation of a convolutional layer, batch normalization and ReLU activation function is then applied 5 times, with a Max Pooling carried out after CB_2 and CB_4 and a Dropout performed after CB_5. The last two layers consist of a fully connected and a softmax layer with 64 elements (i.e., the overall number of possible shifts).

Although a regression approach could have been also taken into account, a classifier with 64 classes has been employed (each class is associated to a specific shift). To perform the classification task a deep learning based solution has been then exploited. Together with the proper selection of the employed dataset (see Section IV-B) architectural choices have to be carefully taken into account.

Specifically, the input layer of the employed CNN is the feature matrix M_β after z-score normalization. This 64×63 matrix contains useful information about the correct shift; each column refers to a specific AC mode and the minimum values is often related to the misalignment to be estimated (see Section IV-A). A classical CNN architecture based on stacked convolutional layers and a fully connected layer followed by a softmax function has been employed (see Figure 6 and Table 1) to properly combine information related to different AC modes of the input matrix. Batch normalization and ReLU activation function are also exploited in the proposed network. Differently than end-to-end deep learning based solution where the original picture is the input data, the proposed approach has no explicit limit about the size of the image under analysis. The input picture is actually converted into a fixed size feature matrix without performing any cropping or resizing that could hurt the performances of the method destroying JPEG traces.

Moreover, the variability of the first quantization matrix could be a big issue in the design of an end-to-end CNN solution. Also in this case, the input feature employed in the proposed approach (i.e., the matrix M_β) is fairly invariant with respect to the values (i.e., quantization steps) of the quantization matrix. Note that the goodness of the proposed features (i.e., the input of the neural network) allow us to employ a simple CNN with a classical architecture, in terms of involved layers, to perform the classification task.

The proposed approach was implemented in Python language (version 3.9.7) using Keras [31] library. All experiments were done on a NVIDIA Tesla K80 GPU. Network training was performed for 15 epochs, with batch size equal to 512, and with a starting learning rate of 0.01 that decayed of 10% every 5 epochs; Stochastic Gradient Descent (SGD) was employed as optimizer and Categorigal Cross Entropy as loss function. The effectiveness of these design choices has been proven in Section VI.

V. EXPERIMENTAL PROTOCOL

A. TEST DATASETS

In order to evaluate the effectiveness of the proposed solution, two of the most well-known benchmark datasets of uncompressed images employed in forensics scenarios involving multiple JPEG compressions (UCID [30], RAISE [33]) together with an additional collection of high resolution pictures [17] have been considered. Further details about these datasets are provided as follows:

- UCID (Uncompressed Color Image Dataset) [30] collection contains more than 1300 medium resolution (512 $\times$ 384) uncompressed TIFF images with a satisfactory content variability (natural scenes, man-made objects, indoor and outdoor environments).

- RAISE [33] dataset, designed to be employed to evaluate digital forgery detection methods, consists of 8156 high-resolution uncompressed images. These pictures have been acquired with three different cameras (Nikon D90, Nikon D7000, Nikon D40) and assigned to one category (outdoor, indoor, landscape, nature, people, objects, buildings).

- The dataset employed in [17] is composed of 100 high-resolution uncompressed TIFF images, with heterogeneous contents acquired with three different cameras (Nikon D90, Canon EOS 450D, Canon EOS 5D).

QF_1	$QF_2 = 90$				$QF_2 = 80$			
	Our	Puglisi [22]	Bianchi [19]	Yao [21]	Our	Puglisi [22]	Bianchi [19]	Yao [21]
50	**1**	**1**	0.93	0.59	**1**	**1**	0.92	0.15
55	**1**	**1**	0.94	0.53	**1**	**1**	0.93	0.10
60	**1**	**1**	0.94	0.44	**1**	**1**	0.93	0.06
65	**1**	**1**	0.94	0.34	**1**	**1**	0.93	0.04
70	**1**	**1**	0.94	0.24	**1**	0.99	0.93	0.02
75	**1**	**1**	0.94	0.15	**1**	0.95	0.89	0.01
80	**1**	**1**	0.94	0.06	**0.87**	0.28	0.21	0.01
85	**1**	**1**	0.92	0.02	**0.72**	0.01	0.02	0.01
90	**0.94**	0.51	0.06	0.01	**0.26**	0.02	0.01	0.01
95	**0.07**	0.02	0.01	0.01	**0.03**	0.02	0.01	0.01
MEAN	**0.90**	0.85	0.76	0.23	**0.79**	0.63	0.58	0.04

TABLE 2. Accuracy estimation of our method compared to S.O.T.A on standard tables [24] with patch size 256×256 at varying of QF_1 and $QF_2 \in \{80, 90\}$

QF_1	$QF_2 = 90$				$QF_2 = 80$			
	Our	Puglisi [22]	Bianchi [19]	Yao [21]	Our	Puglisi [22]	Bianchi [19]	Yao [21]
50	**1**	**1**	0.93	0.43	**1**	0.99	0.91	0.11
55	**1**	**1**	0.93	0.37	**1**	0.98	0.91	0.08
60	**1**	**1**	0.92	0.31	**1**	0.97	0.89	0.06
65	**1**	**1**	0.91	0.23	**0.99**	0.96	0.87	0.04
70	**1**	0.99	0.91	0.16	**0.99**	0.91	0.83	0.03
75	**1**	0.99	0.91	0.11	**0.95**	0.72	0.44	0.01
80	**1**	0.99	0.87	0.05	**0.55**	0.14	0.02	0.01
85	**0.99**	0.95	0.76	0.02	**0.30**	0.01	0.01	0.01
90	**0.74**	0.23	0.01	0.01	**0.08**	0.01	0.01	0.01
95	**0.03**	0.01	0.01	0.01	**0.02**	0.01	0.01	0.01
MEAN	**0.88**	0.81	0.72	0.17	**0.69**	0.57	0.49	0.04

TABLE 3. Accuracy estimation of our method compared to S.O.T.A on standard tables [24] with patch size 128×128 at varying of QF_1 and $QF_2 \in \{80, 90\}$

QF_1	$QF_2 = 90$				$QF_2 = 80$			
	Our	Puglisi [22]	Bianchi [19]	Yao [21]	Our	Puglisi [22]	Bianchi [19]	Yao [21]
50	**0.99**	0.97	0.81	0.26	**0.97**	0.89	0.61	0.07
55	**0.99**	0.97	0.80	0.22	**0.97**	0.87	0.47	0.06
60	**0.99**	0.96	0.79	0.18	**0.95**	0.82	0.39	0.05
65	**0.99**	0.96	0.75	0.13	**0.93**	0.75	0.17	0.03
70	**0.99**	0.95	0.73	0.1	**0.86**	0.61	0.09	0.02
75	**0.99**	0.94	0.57	0.07	**0.69**	0.35	0.02	0.01
80	**0.97**	0.86	0.19	0.04	**0.19**	0.05	0.01	0.01
85	**0.92**	0.69	0.05	0.02	**0.06**	0.01	0.01	0.01
90	**0.30**	0.08	0.01	0.01	**0.03**	0.01	0.01	0.01
95	**0.02**	0.01	0.01	0.01	**0.02**	0.01	0.01	0.01
MEAN	**0.82**	0.74	0.47	0.10	**0.57**	0.43	0.18	0.03

TABLE 4. Accuracy estimation of our method compared to S.O.T.A on standard tables [24] with patch size 64×64 at varying of QF_1 and $QF_2 \in \{80, 90\}$

Crop Size	$QF_2 = 90$				$QF_2 = 80$			
	Our	Puglisi [22]	Bianchi [19]	Yao [21]	Our	Puglisi [22]	Bianchi [19]	Yao [21]
64×64	**0.43**	0.35	0.25	0.05	**0.24**	0.13	0.05	0.02
128×128	**0.49**	0.44	0.42	0.10	**0.35**	0.25	0.26	0.03
256×256	**0.57**	0.49	0.49	0.14	**0.40**	0.32	0.32	0.03
MEAN	**0.50**	0.43	0.39	0.10	**0.33**	0.23	0.21	0.02

TABLE 5. Accuracy estimation of our method compared to S.O.T.A on custom tables [32] at varying of crop size

B. TESTING PROTOCOL

Starting from the datasets described in the previous section, two sets of tests to compare the proposed solution to state-of-the-art approaches have been considered:

- A first series of tests (see Subsection VI-A) has been conducted to assess the performance of the proposed solution at varying of image resolution and quantization matrices related to first and second JPEG compressions. Note that, in these tests, a subset of images belonging to UCID [33] dataset and standard quantization tables are employed.

- A second series of tests (see Subsection VI-B) has been performed to verify the effectiveness of the proposed solution in scenarios with a considerable mismatch with respect to training conditions. Such as example, several tests have been conducted with patches extracted from high resolution images (RAISE [33] and [17]), employing custom quantization matrices [32] and patch sizes not considered in the training phase.

VI. EXPERIMENTAL RESULTS

A. COMPARISON

To assess the performances of the proposed approach, a first series of tests has been performed on several datasets built considering UCID collection [30]. Specifically, a set of test datasets have been generated starting from images not employed in training and validation sets (see Section IV-B), considering 3 patch sizes (64×64, 128×128, 256×256), standard first quantization matrices (*standard*), all the 64 shifts and two quality factors ($QF_2 \in \{90, 80\}$) for the final compression. Considering all the involved parameters, 768.000 ($200 \times 3 \times 10 \times 64 \times 2$) non-aligned double JPEG compressed images have been employed.

Comparisons with state-of-the-art solutions have been performed considering the original code provided by the authors ([19], [22]) or reimplementing the method described in the related paper [21]. As far as we know the proposed method is the first deep-learning based approach for the task of misaligment estimation of NA-DJPEG images. A first analysis has been conducted by employing standard quantization tables in the first compression. As can be observed from Tables 2, 3 and 4, all the considered approaches decrease their performances at increasing of QF_1 with a significant drop in the $QF_1 > QF_2$ scenario. Moreover, pixel domain based approach [21], due to the limited number of homogeneous blocks to be selected in small patches, is deceived by the image content and does not work properly. Both the proposed approach and our preliminary solution [22], exploiting the information contained in all the AC modes (M_β matrix), perform better than [19] that consider only DC component in the related algorithm. A considerable gain is achieved in challenging conditions such as small patches (64×64) and stronger final quantization. Finally, the proposed approach, employing a CNN to extract the useful information from the feature matrix M_β outperforms [22] where a simple voting strategy was considered.

B. GENERALIZATION

To work properly in real conditions, design choices about input features and architecture have been carefully considered (see Section IV). To further confirm the effectiveness of the proposed approach, the robustness with respect to scenarios unseen in the training phase has been then analysed.

Note that in forensics scenarios involving JPEG algorithm, the variability of the employed quantization tables has to be taken into account. Although lots of state-of-the-art solutions have been designed considering only standard quantization matrices, as reported in [32], custom tables are often used in the compression pipeline. Methods involving JPEG algorithm should then verify their effectiveness also with quantization matrices different with respect to the standard ones.

To this aim several tests have been then conducted with custom tables. Specifically, we considered a similar scenario as described before for standard quantization matrices (i.e., UCID dataset, 3 crop sizes, 2 QF_2, 64 shifts) but employing custom quantization matrices from [32] in the first compression. For each UCID image (unseen in training phase), 10 quantization tables are randomly selected from [32] as Q_1 and two quality factors ($QF_2 \in \{80, 90\}$) as Q_2; therefore also in this investigation 768.000 ($200 \times 3 \times 10 \times 64 \times 2$) images have been tested. Differently than quantization matrices related to quality factors where all the tables can be derived from a base one, custom matrices show a high variability more difficult to cope with. As reported in Table 5 our approach outperforms considerably the other solutions also in this challenging scenario. Finally, note that custom quantization tables employed in this test have not been seen in the training phase of the proposed CNN.

QF_1	$QF_2 = 90$				$QF_2 = 80$			
	Our	Puglisi [22]	Bianchi [19]	Yao [21]	Our	Puglisi [22]	Bianchi [19]	Yao [21]
50	**0.97**	0.94	0.82	0.57	**0.95**	0.94	0.81	0.07
55	**0.97**	0.95	0.83	0.50	**0.95**	0.93	0.81	0.05
60	**0.97**	0.95	0.83	0.40	**0.96**	0.94	0.82	0.04
65	**0.98**	0.96	0.85	0.27	**0.96**	0.93	0.82	0.03
70	**0.98**	0.96	0.86	0.17	**0.96**	0.91	0.82	0.03
75	**0.99**	0.97	0.87	0.11	**0.91**	0.67	0.79	0.02
80	**0.99**	0.98	0.86	0.06	**0.60**	0.06	0.30	0.01
85	**0.99**	0.96	0.86	0.03	**0.66**	0.01	0.03	0.01
90	**0.66**	0.22	0.10	0.01	**0.28**	0.01	0.02	0.01
95	**0.08**	0.01	0.02	0.01	**0.04**	0.01	0.02	0.01
MEAN	**0.86**	0.79	0.69	0.21	**0.72**	0.54	0.52	0.03

TABLE 6. Accuracy estimation of our method compared to S.O.T.A with standard tables [24], patch size 256 × 256 on RAISE collection [33] at varying of QF_1 and $QF_2 \in \{80, 90\}$

.

QF_1	$QF_2 = 90$				$QF_2 = 80$			
	Our	Puglisi [22]	Bianchi [19]	Yao [21]	Our	Puglisi [22]	Bianchi [19]	Yao [21]
50	**0.95**	0.92	0.80	0.47	**0.92**	0.89	0.78	0.08
55	**0.95**	0.92	0.81	0.40	**0.92**	0.89	0.78	0.07
60	**0.96**	0.92	0.81	0.32	**0.92**	0.87	0.78	0.05
65	**0.96**	0.93	0.81	0.23	**0.92**	0.86	0.75	0.04
70	**0.97**	0.93	0.81	0.15	**0.90**	0.75	0.73	0.03
75	**0.96**	0.93	0.82	0.10	**0.77**	0.42	0.50	0.02
80	**0.97**	0.93	0.75	0.05	**0.31**	0.03	0.03	0.01
85	**0.95**	0.87	0.69	0.03	**0.30**	0.01	0.02	0.01
90	**0.42**	0.11	0.02	0.01	**0.08**	0.01	0.02	0.01
95	**0.03**	0.01	0.02	0.01	**0.02**	0.01	**0.02**	0.01
MEAN	**0.81**	0.75	0.63	0.18	**0.61**	0.47	0.44	0.03

TABLE 7. Accuracy estimation of our method compared to S.O.T.A on standard tables [24], patch size 128 × 128 on RAISE collection [33] at varying of QF_1 and $QF_2 \in \{80, 90\}$

.

QF_1	$QF_2 = 90$				$QF_2 = 80$			
	Our	Puglisi [22]	Bianchi [19]	Yao [21]	Our	Puglisi [22]	Bianchi [19]	Yao [21]
50	**0.92**	0.85	0.72	0.32	**0.86**	0.75	0.60	0.05
55	**0.92**	0.85	0.70	0.25	**0.86**	0.73	0.47	0.04
60	**0.92**	0.85	0.70	0.21	**0.85**	0.68	0.41	0.03
65	**0.92**	0.86	0.66	0.15	**0.83**	0.60	0.20	0.02
70	**0.92**	0.85	0.64	0.10	**0.71**	0.43	0.12	0.02
75	**0.93**	0.85	0.52	0.07	**0.41**	0.17	0.02	0.01
80	**0.92**	0.79	0.19	0.04	**0.09**	0.02	0.02	0.01
85	**0.83**	0.60	0.06	0.02	**0.06**	0.01	0.02	0.01
90	**0.16**	0.04	0.02	0.01	**0.03**	0.01	0.02	0.01
95	**0.01**	**0.01**	**0.01**	**0.01**	**0.02**	0.01	**0.02**	0.01
MEAN	**0.74**	0.66	0.42	0.12	**0.47**	0.34	0.19	0.02

TABLE 8. Accuracy estimation of our method compared to S.O.T.A on standard tables [24], patch size 64 × 64 on RAISE collection [33] at varying of QF_1 and $QF_2 \in \{80, 90\}$

.

QF_1	crop size = 96 × 96				crop size = 192 × 192			
	Our	Puglisi [22]	Bianchi [19]	Yao [21]	Our	Puglisi [22]	Bianchi [19]	Yao [21]
50	**1**	0.99	0.96	0.40	**1**	**1**	0.97	0.57
55	**1**	0.99	0.96	0.32	**1**	**1**	0.98	0.51
60	**1**	0.99	0.95	0.26	**1**	**1**	0.96	0.42
65	**1**	0.99	0.94	0.19	**1**	**1**	0.97	0.32
70	**1**	0.98	0.94	0.13	**1**	**1**	0.97	0.22
75	**1**	0.98	0.92	0.07	**1**	**1**	0.97	0.13
80	**1**	0.96	0.80	0.04	**1**	**1**	0.97	0.06
85	**1**	0.87	0.44	0.02	**1**	0.99	0.95	0.02
90	**0.57**	0.14	0.02	0.01	**0.89**	0.38	0.03	0.01
95	**0.03**	0.01	0.02	0.01	**0.07**	0.02	0.02	0.01
MEAN	**0.86**	0.79	0.69	0.14	**0.90**	0.84	0.78	0.23

TABLE 9. Accuracy estimation of our method compared to S.O.T.A standard tables [24], $QF_2 = 90$, on UCID collection [30] considering path sizes not seen in the training phase at varying of QF_1.

QF_1	$QF_2 = 90$				
	Our	[22]	[19]	[21]	[27]
50	**1**	0.98	0.91	0.48	0.19
55	**1**	0.98	0.94	0.41	0.18
60	**1**	0.98	0.94	0.29	0.17
65	**1**	0.99	0.94	0.19	0.15
70	**1**	0.99	0.94	0.11	0.13
75	**1**	0.99	0.95	0.06	0.11
80	**1**	0.98	0.93	0.04	0.08
85	**1**	0.92	0.89	0.02	0.05
90	**0.43**	0.1	0.02	0.01	0.02
95	**0.03**	0.01	0.02	0.01	0.01
MEAN	**0.85**	0.79	0.75	0.16	0.11

TABLE 10. Accuracy estimation of our method compared to S.O.T.A on standard tables [24], patch size 128 × 128 on [17] at varying of QF_1 and $QF_2 = 90$

In the testing phase of a machine learning based approach, the employed dataset is a critical aspects. The proposed solution has been trained on UCID collection (512 × 384 pixel size), then further tests performed on datasets with a different resolution could provide useful hints about performances in unseen conditions; to this aim, RAISE collection [33] has been selected. This collection consists of high-resolution uncompressed pictures captured in different scenes (indoor, outdoor, etc.) employing also different cameras. Starting from 200 randomly selected RAISE images, the same strategy of previous section was used: every image was cropped with size $\in \{256 \times 256, 128 \times 128, 64 \times 64\}$ and JPEG compressed with $Q_1 \in$ *standard*; then a second compression was performed with $QF_2 \in \{90, 80\}$ and misalignment $s \in \{0, 1, 2, 3, ..., 63\}$, generating 768.000 ($200 \times 3 \times 10 \times 64 \times 2$) test images. As reported in Tables 6, 7 and 8 the proposed solution achieves satisfactory results also in a scenario with a considerable resolution mismatch between datasets employed in the training (UCID) and test set (RAISE) respectively. Note that lower accuracy achieved by all the considered approaches with respect to the similar scenario employing UCID dataset (see Tables 2, 3, 4) depends on the available information contained in a fixed size patch extracted from datasets with different resolution images. Also these tests confirm the gap obtained with respect to the state-of-the-art.

As described in Section IV, the employed CNN takes as input the normalized matrix of β (M_β). This 64×63 matrix can be computed from images of different resolution without modifying the employed architecture. Differently than end-to-end deep learning based solution with the original image as input data, the proposes approach has no explicit limit about the resolution of the image under analysis. However, due to the mismatch between training and testing conditions a series of experiments have been conducted with patch sizes unseen by the obtained model. Specifically, images from UCID dataset has been considered, with $Q_1 \in$ *standard*, crop size $\in \{96 \times 96, 192 \times 192\}$, $Q_2 = 90$, and 64 shifts. As reported in Table 9, the proposed solution is able to work properly also with image sizes not seen in the training phase.

A final test has been conducted starting from high resolution images collected in [17]. Specifically, every image in [17] was cropped with size 128×128 and JPEG compressed with standard quantization tables ($QF_1 \in$ *standard*, $QF_2 = 90$) considering all the misalignments. An additional approach based on high pass filtering in pixel domain [27] has been considered in this test employing the original code provided by the authors. Note that results in Table 10 further confirm the effectiveness of the proposed solution able to exploit the information contained in small patches better than the other state-of-the-art approaches. Note that methods based on high-pass filtering in pixel domain ([21], [27]) often designed to work with the whole image, are not able to provide satisfactory results in the analysed scenarios.

VII. CONCLUSION

Information about the misalignment occurred between consecutive JPEG compressions can be useful exploited in forensics investigation to retrieve useful hints about the manipulation history of the image under analysis. In this work, statistics extracted from AC histograms obtained applying a third JPEG compression have been exploited as input of a deep learning based approach devoted to estimate the aforementioned misalignment. Moreover, the goodness of the proposed features, i.e., the main contribution of the proposed solution, has been confirmed by both the simplicity of the employed network architecture and the achieved results. A series of tests and comparisons with state-of-the-art solutions, performed at varying of patch sizes, quantization matrices (both standard and custom), employed datasets (medium and high resolution), demonstrates the effectiveness of the proposed solution. Finally, future works will be devoted to exploit the estimated misalignment information to localize forgeries and to estimate the first quantization matrix.

REFERENCES

[1] O. Giudice, A. Paratore, M. Moltisanti, and S. Battiato, *A classification engine for image ballistics of social data*. Springer International Publishing, 2017, pp. 625–636 (LNCS, volume 10 485).

[2] M. Moltisanti, A. Paratore, S. Battiato, and L. Saravo, "Image manipulation on facebook for forensics evidence," in *Proc. of International Conference on Image Analysis and Processing*. Springer, 2015, pp. 506–517.

[3] S. Verde, C. Pasquini, F. Lago, A. Goller, F. De Natale, A. Piva, and G. Boato, "Multi-clue reconstruction of sharing chains for social media images," *IEEE Transactions on Multimedia*, pp. 1–15, 2023.

[4] Z. Fan and R. De Queiroz, "Identification of bitmap compression history: JPEG detection and quantizer estimation," *IEEE Trans. on Image Processing*, vol. 12, no. 2, pp. 230–235, 2003.

[5] H. Farid, "Digital image ballistics from JPEG quantization: A followup study," *Department of Computer Science, Dartmouth College, Tech. Rep. TR2008-638*, 2008.

[6] S. Battiato and G. Messina, "Digital forgery estimation into DCT domain: a critical analysis," in *Proc. of the First ACM workshop on Multimedia in forensics*. 6: 25, 2010, pp. 389–399.

[7] A. Piva, "An overview on image forensics," *ISRN Signal Processing*, vol. 2013, p. 22, 2013.

[8] L. Verdoliva, "Media forensics and deepfakes: an overview," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 5, pp. 910–932, 2020.

[9] O. Giudice, F. Guarnera, A. Paratore, and S. Battiato, "1-D DCT domain analysis for JPEG double compression detection," in *Proc. of International Conference on Image Analysis and Processing*. Springer, 2019, pp. 716–726 (LNCS, volume 11 752).

[10] S. Battiato, O. Giudice, F. Guarnera, and G. Puglisi, "First quantization estimation by a robust data exploitation strategy of DCT coefficients," *IEEE Access*, vol. 9, pp. 73 110–73 120, 2021.

[11] E. Kee, M. K. Johnson, and H. Farid, "Digital image authentication from JPEG headers," *IEEE Trans. on Information Forensics and Security*, vol. 6, no. 3, pp. 1066–1075, 2011.

[12] S. Battiato, O. Giudice, F. Guarnera, and G. Puglisi, "Estimating previous quantization factors on multiple JPEG compressed images," *EURASIP Journal on Information Security*, vol. 2021, no. 8, pp. 1–11, 2021.

[13] F. Galvan, G. Puglisi, A. R. Bruna, and S. Battiato, "First quantization matrix estimation from double compressed JPEG images," *IEEE Trans. on Information Forensics and Security*, vol. 9, no. 8, pp. 1299–1310, 2014.

[14] S. Battiato, O. Giudice, F. Guarnera, and G. Puglisi, "CNN-based first quantization estimation of double compressed JPEG images," *Journal of Visual Communication and Image Representation*, vol. 89, p. 103635, 2022.

[15] M. C. Stamm, M. Wu, and K. J. R. Liu, "Information forensics: An overview of the first decade," *IEEE Access*, vol. 1, pp. 167–200, 2013.

[16] M. Barni, P. Campisi, E. J. Delp, G. Doërr, J. Fridrich, N. Memon, F. Pérez-González, A. Rocha, L. Verdoliva, and M. Wu, "Information forensics and security: A quarter-century-long journey," *IEEE Signal Processing Magazine*, vol. 40, no. 5, pp. 67–79, 2023.

[17] T. Bianchi and A. 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.

[18] L. Wu, X. Kong, B. Wang, and S. Shang, "Image tampering localization via estimating the non-aligned double JPEG compression," in *Media Watermarking, Security, and Forensics 2013*, A. M. Alattar, N. D. Memon, and C. D. Heitzenrater, Eds., vol. 8665, International Society for Optics and Photonics. SPIE, 2013, pp. 260 – 266. [Online]. Available: https://doi.org/10.1117/12.2003695

[19] T. Bianchi and A. 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.

[20] N. Dalmia and M. Okade, "Robust first quantization matrix estimation based on filtering of recompression artifacts for non-aligned double compressed JPEG images," *Signal Processing: Image Communication*, vol. 61, pp. 9–20, 2018.

[21] H. Yao, H. Wei, C. Qin, and X. Zhang, "An improved first quantization matrix estimation for nonaligned double compressed jpeg images," *Signal Processing*, vol. 170, p. 107430, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165168419304827

[22] G. Puglisi and S. Battiato, "Misalignment estimation in non-aligned double JPEG scenario based on AC histogram analysis," in *Pattern Recognition, Computer Vision, and Image Processing. ICPR 2022 International Workshops and Challenges*, J.-J. Rousseau and B. Kapralos, Eds. Cham: Springer Nature Switzerland, 2023, pp. 338–346.

[23] S. Battiato, M. Mancuso, A. Bosco, and M. Guarnera, "Psychovisual and statistical optimization of quantization tables for DCT compression engines," in *Proc. of 11th International Conference on Image Analysis and Processing, 2001*. IEEE, 2001, pp. 602–606.

[24] G. K. Wallace, "The JPEG still picture compression standard," *Communications of the ACM*, vol. 34, no. 4, pp. 30–44, 1991.

[25] W. Luo, Z. Qu, J. Huang, and G. Qiu, "A novel method for detecting cropped and recompressed image block," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, vol. 2, 2007, pp. II–217–II–220.

[26] A. R. Bruna, G. Messina, and S. Battiato, "Crop detection through blocking artefacts analysis," in *Image Analysis and Processing – ICIAP 2011*, G. Maino and G. L. Foresti, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 650–659.

[27] N. Dalmia and M. Okade, "A novel technique for misalignment parameter estimation in double compressed jpeg images," in *2016 Visual Communications and Image Processing (VCIP)*, 2016, pp. 1–4.

[28] G. Puglisi and S. Battiato, "A robust misalignment estimation approach in non-aligned double JPEG compression scenario," in *2022 IEEE International Conference on Image Processing (ICIP)*, 2022, pp. 3096–3100.

[29] E. Y. Lam and J. W. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Trans. on Image Processing*, vol. 9, no. 10, pp. 1661–1666, 2000.

[30] G. Schaefer and M. Stich, "UCID: An uncompressed color image database," in *Storage and Retrieval Methods and Applications for Multimedia 2004*, vol. 5307. International Society for Optics and Photonics, 2003, pp. 472–480.

[31] F. Chollet *et al.*, "Keras," https://keras.io, 2015.

[32] J. Park, D. Cho, W. Ahn, and H. Lee, "Double JPEG detection in mixed JPEG quality factors using deep convolutional neural network," in *The European Conference on Computer Vision (ECCV)*, September 2018.

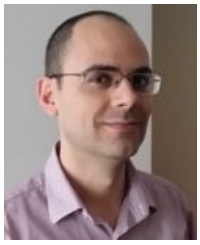
[33] D. Dang-Nguyen, C. Pasquini, V. Conotter, and G. Boato, "Raise: a raw images dataset for digital image forensics," in *Proc. of the 6th ACM Multimedia Systems Conference*, 2015, pp. 219–224.

ACKNOWLEDGMENT

The work of Francesco Guarnera has been supported by MUR in the framework of PNRR PE0000013, under project "Future Artificial Intelligence Research – FAIR".

SEBASTIANO BATTIATO (M'04–SM'06) received the degree (summa cum laude) in computer science from the University of Catania, in 1995, and the Ph.D. degree in computer science and applied mathematics from the University of Naples, in 1999. From 1999 to 2003, he was the Leader of the "Imaging" Team, STMicroelectronics, Catania. He joined the Department of Mathematics and Computer Science, University of Catania, as an Assistant Professor, an Associate Professor, and a Full Professor, in 2004, 2011, and 2016, respectively. He has been Chairman of the Undergraduate Program in Computer Science (2012-2017), and Rector's delegate for Education: postgraduates and Phd (2013-2016). He is currently the Scientific Coordinator of the PhD Program in Computer Science and Deputy Rector for Strategic Planning and Information Systems at the University of Catania. Prof. Battiato is involved in research and directorship of the IPLab research lab (http://iplab.dmi.unict.it). He coordinates IPLab's participation on large scale projects funded by national and international funding bodies, as well as by private companies. Prof. Battiato has participated as principal investigator in many international and national research projects He has supervised about 15 phd students and 3 postdocs. Prof. Battiato has edited 6 books and co-authored about 300 papers in international journals, conference proceedings and book chapters and has also been involved as "guest editor" of several special issues published in International Journals. His current research interests include computer vision, imaging technology, and multimedia forensics. Prof. Battiato has been a regular member of numerous international conference committees. He was a recipient of the 2017 PAMI Mark Everingham Prize for the series of annual ICVSS schools and the 2011 Best Associate Editor Award of the IEEE Transactions on Circuits and Systems for Video Technology. He has been the Chair of several international events, including MMFORwild 2020-2022, IMPROVE 2021-2024, INTELLYSIS 2020-2024, SIGMAP 2019-2020ICIAP 2017,VINEPA 2016, ACIVS 2015, VAAM2014-2015-2016, VISAPP2012-2015, IWCV2012, ECCV2012, ICIAP 2011, ACM MiFor 2010-2011, and SPIE EI Digital Photography 2011-2012-2013. He is an Associate Editor of the SPIE Journal of Electronic Imaging . He is the Director (and Co-Founder) of the International Computer Vision Summer School (ICVSS) and Director (and Co-Founder) of IFOSS - International Forensics Summer School.

FRANCESCO GUARNERA received the bachelor degree (summa cum laude) in computer science, in 2009, and the master degree (summa cum laude) in computer science in 2018 from University of Catania. From 2009 to 2016 he was a developer/analyst/project manager of web applications. In 2022 he obtained the title Doctor of Philosophy (Ph.D.) with a thesis entitled "Advanced Methods for Image Forensics: First Quantization Estimation and Document Authentication" at the Department of Mathematics and Computer Science, University of Catania. From April 2022 to February he was a research fellow at the Department of Pharmaceutical Sciences of the University of Catania studies with the grant entitled "Modeling using Agent Based Modeling (ABM) methodologies by In Silico Trials for drug evaluation" joining Medical Imaging research projects focused on the analysis of magnetic resonances of the brain.He is currently a researcher at the Department of Mathematics and Computer Science, University of Catania. His research interests include image processing, multimedia forensics, medical imaging and deep learning. He coauthored more than 20 papers in international journals and conference proceedings. Since 2017 he has been a member of the IPLab research group of the University of Catania. He participated at the International Computer Vision Summer School (ICVSS) in the 2019 edition. From 2022 he is part of the committee of the International Forensics Summer School (IFOSS).



GIOVANNI PUGLISI Giovanni Puglisi received the M.S. degree in computer science engineering (summa cum laude) from Catania University, Catania, Italy, in 2005, and the Ph.D. degree in computer science in 2009. From 2009 to 2014, he worked at the University of Catania, Italy, as post-doc researcher. He joined the Department of Mathematics and Computer Science, University of Cagliari, as associate professor in 2014. His research interests include image/video enhancement and processing, camera imaging technology and multimedia forensics. He edited one book, coauthored more than 50 papers in international journals, conference proceedings and book chapters. He is a co-inventor of several patents and serves as reviewer different international journals and conferences.