Forensics of front camera acquisitions (related works and project status)

Joy Battocchio   Davide Guidolin   Alessio Belli   Alberto Casagrande

I. INTRODUCTION

Source identification is one of the most important tasks in digital image forensics. In fact, the ability to reliably associate an image with its acquisition device may be crucial both during investigations and before a court of law. The key assumption of forensics source identification is that acquisition devices leave traces in the acquired content, and that instances of these traces are specific to the respective (class of) device(s). This kind of traces is present in the so-called device fingerprint. A major impulse to research in this field came with the seminal work of Lukas et al. [2] showing that reliable device identification is possible based on the camera photo-response non uniformity (PRNU) pattern. Each camera is characterized by its unique PRNU pattern, which can be regarded as a sort of camera fingerprint. All photos taken by a given camera carry traces of its fingerprint which, under suitable hypotheses, can be retrieved, enabling reliable device identification and brand and model identification as well. However, due to spatial transformations, filtering, AI post-processing and more in general in-device or out-device post-processing, PRNU is today more than ever compromised. Even for media coming from the exact same device, it is really difficult that we can compare the two PRNUs of these ones straightforwardly, but we need to resynchronize one with respect to the other. In fact, Mandelli et al.'s [4] work is based on the assumption that the images are already geometrically synchronized. Its main contribution is using a 2 channel CNN network that for every device d $\in$ DT is fed using noise residuals coming from device d, paired with PRNU K_d (coherent pair) and with PRNU K_d^- coming from a different device (non-coherent pair). The CNN is able to learn a similarity measure Cs for the source identification task. This method proves to be generally faster than PCE, in particular when a large amount of potential provenance devices is investigated, requiring much less query image content to obtain enhanced attribution accuracy.

II. RELATED WORKS

Of all the related works cited in the main paper, it has been decided to focus on three:

- **Kharrazi et al.** "Blind source camera identification" [1]
- **Lukáš et al.** "Digital camera identification from sensor pattern noise" [2]
- **Bondi et al.** "First Steps Toward Camera Model Identification with Convolutional Neural Networks" [3]

The choice is motivated by the influence that these three works had on the Mandelli's one. They all represent important milestones in the solution of the source identification problem. For this reason it has been decided to present them in temporal order of publication, because each one of them gives an additional contribution and it's logically dependent on the previous ones.

A. Kharrazi et al.

The authors of "Blind source camera identification" propose to extract handcrafted features from images in order to use them as input for a SVM classifier.
The chosed features are the ones that bring more evidence of the CFA configuration and the color processing carried out by the camera, which should be unique for each sensor and therefore useful for their identification.
In particular they are:

- **Average pixel value**. Values in RGB channels of an image should average to gray assuming that the images has enough color variations.
- **RGB pairs correlation**. Capture the fact that depending on the camera structure, the correlation between different color bands could varies.
- **Neighbor distribution Center of mass**. Calculate the number of pixel neighbors for each pixel value, where the pixel neighbors are defined as all pixels having a difference in value of 1 or -1, from the pixel value in question. This is calculated for each color band.
- **RGB pairs energy ratio**. It is used in the process of white point correction.
- **Wavelet domain statistic**. Decompose each color band of the image using separable quadratic mirror filters and then calculate the mean for each of the 3 resulting sub-bands.

In addition to these features, different cameras produce images of different quality, so we can extract the Image Quality Metrics as features to aid in distinguishing between cameras. As it is written above, all these features are then used by a SVM to classify the right source camera.

B. Lukáš et al.

In the "Digital camera identification from sensor pattern noise" paper, the authors proposed a new method for the camera identification problem that is based on the sensor's

pattern noise. The basic idea is to compute for each camera a pattern noise that can act as a unique fingerprint. Then, if you want to identify the camera that shoot a specific photo, the paper suggests to use a correlation filter in order to compare the noise residual of the image with the fingerprint of all the cameras inside the dataset.

In the image acquisition process, there are many sources of imperfections and noise that can enter into various phases. Among these noises, we can identify two types of noise: *shot noise* (a random component) and *pattern noise*. The latter is a deterministic component that has the characteristic of remaining approximately the same if different images of the same scene are taken. This is why this type of noise can be used for the camera identification problem. One of the dominant parts of the pattern noise is the photo-response nonuniformity noise (PRNU). It is caused by the inhomogenity of silicon wafers and imperfections during the sensor manufacturing process. Given the character and these origins, it is unlikely that different sensors exhibit correlated noise patterns.

The algorithm proposed by the authors of the paper states that if you want to verify whether a specific image p was taken by the camera C, we first have to compute the fingerprint (reference pattern) P_C and then establish the presence of this reference pattern in p using the correlation. It is possible to obtain an approximation of the PRNU noise by averaging multiple images. However, if you want to speed up the process, you can suppress the scene content from the image by applying a denoising filter F and then averaging the noise residuals $n^{(k)}$

$$n^{(k)} = p^{(k)} - F(p^{(k)})$$

The authors experimented with several denoising filters and they decided to use the wavelet-based one. To decide if an image p was taken by the camera C they calculated the correlation ρ_C between the noise residual and the camera reference pattern P_C

$$\rho_C(p) = corr(n, P_C) = \frac{(n - \bar{n}) * (P_C - \bar{P}_C)}{||n - \bar{n}|| ||P_C - \bar{P}_C||}$$

C. Bondi et al.

About 10 years after the PRNU-based method for digital camera identification, in 2017, the authors of "First Steps Toward Camera Model Identification with Convolutional Neural Networks" investigated a novel approach to solve the camera model identification problem, based on Convolutional Neural Networks. The choice was fully motivated by other successful deep-learning algorithms applied to forensics tasks. In particular, they made use of a *Convolutional Neural Network* (CNN) which learns the features characterizing each camera model directly from the acquired images, rather than using handcrafted features. In addition, they also used linear *Support Vector Machines* (SVMs) for classification.

As shown in Fig. 1, given a set of training and validation labeled samples coming from N known camera models, the training proceeds as follows: for each image I with the corresponding label L (camera model), K non overlapping patches are extracted, where each patch P_k inherits the same
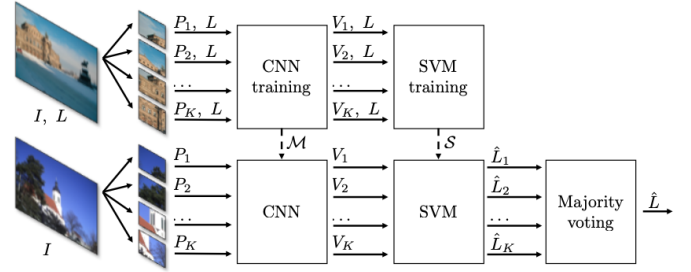

Fig. 1. CNN - SVM architecture — training and evaluation pipelines

label L of the source image. For each patch, the CNN extracts a feature vector V_k of 128 elements. The resulting feature vectors are then used to train a battery of $N \cdot (N-1)/2$ linear binary SVM classifiers S. During evaluation, the battery S of linear SVMs assigns a label $\hat{L}_k$ to each patch. The predicted label $\hat{L}$ for image I is obtained through majority voting on $\hat{L}_k$, $k \in [1, K]$.

The advantages of the proposed CNN are mainly two:

- It learns a feature extraction methodology that generalizes well on a set of unknown camera models;
- The resulting feature vectors have only 128 elements, allowing to characterize camera models in a space with reduced dimensionality and consequently enabling the use of simple classifiers like SVMs.
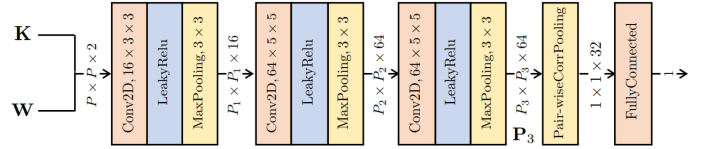
III. CNN-BASED FAST SOURCE DEVICE IDENTIFICATION

Fig. 2. Pair-Wise Correlation Network (PCN) architecture

In the "*CNN-based fast source device identification*" [4] paper the authors propose a CNN-based approach to address the problem of source device identification based on PRNU. In particular, their goal is to replace the PCE test with an alternative method based on CNN. Their approach differs from [3] because they train an end-to-end network instead of using a CNN followed by an SVM.

The core idea of this paper is to use a 2-channel CNN to compare the estimated device PRNU K_d with the noise residual W obtained from the given test image I.

First of all they estimate the device PRNU [2] K_d from a set of flat-field images taken from the same device and they extract the noise residual W from a set of images depicting natural scenes. Then they crop the central PxP pixel region from both K_d and W and they feed the CNN with these 2 image patches. Since the method is independent from the used CNN they experiment different pre-trained networks and they propose a simple but effective architecture that is composed
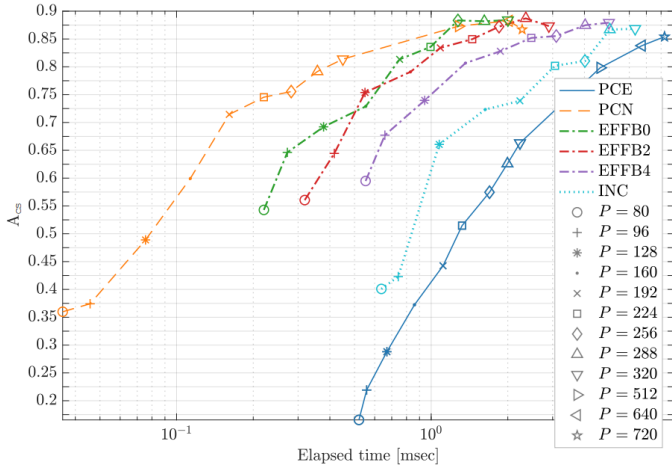
Fig. 3. Accuracy A_cs as a function of time [milliseconds] and crop size P.

Fig. 4. Area under the ROC curve AUC_os as a function of crop size P.

by 3 convolutional layers, a pair-wise correlation pooling layer and a fully connected layer (Fig. 2). The pair-wise correlation pooling layer computes the correlation between adjacent pairs of features, while the fully connected layer outputs an identification score C_s that is high for coherent pairs and low for non-coherent ones.

The network can be used in both a closed-set scenario and in an open-set scenario. In the closed-set scenario we have a finite set of devices and we have to identify the source of the image from that set. In the open-set scenario we have a test image and a candidate device and we have to tell if that device shot the given image.

The authors shows that in both the open and closed set scenario the proposed approach outperforms the PCE test both in accuracy and inference time. In particular, regarding the closed set scenario, we can see from Fig. 3 that the proposed Pair-wise Correlation Net (PCN) is faster than all the other architectures and it outperforms PCE in accuracy, however it is outperformed by the other networks. It is important to notice also that all the methods perform better when the patch size is bigger. Regarding the open set scenario, the area under the ROC curve related to the patch size is shown in Fig. 4 and it's clear that PCE is worst than all the proposed architectures. Finally, in Fig. 5, we can see the accuracy as a function of crop size when the dataset is double-JPEG compressed. The PCN outperforms PCE even after double compression and we can notice that the performances are better when the network is trained on data where the noise residual is extracted from double compressed images.

A. Final comment

In [1] are described the bases of using sensor-specific informations to identify the source of an image, however it was not clear which of the many analized handcrafted features were the more important for this task. After this work it is clear that the direction is to gain a better understanding on the relevant component of the camera fingerprint.
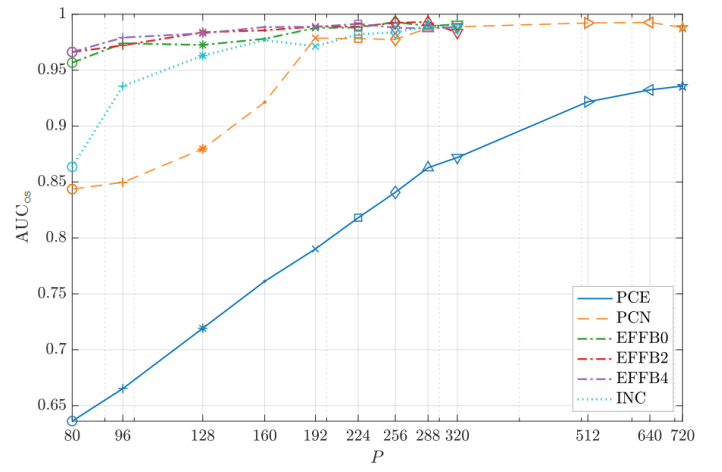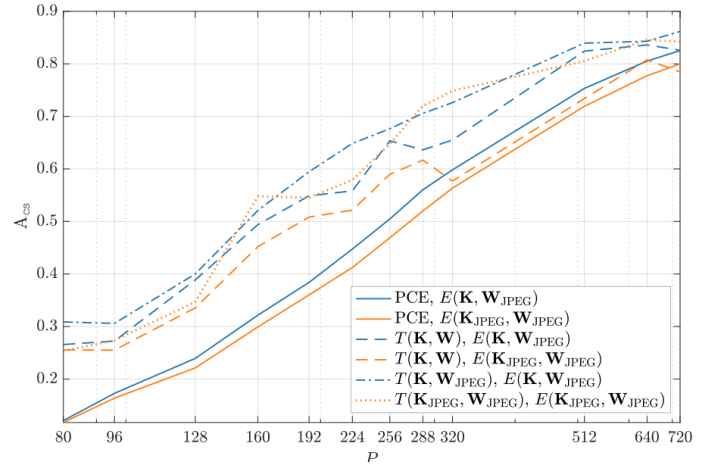The next milestone [2] analizes and categorizes the sources

Fig. 5. Accuracy A_cs as a function of crop size P, considering double-JPEG compressed dataset with quality 90 and PCN network. $T()$ and $E()$ refer to training and evaluation datasets, respectively.

of deterministic noise, where PRNU results to be the most relevant for solving our problem. They also describe the way to extract the PRNU from the images, the same method is used in the reference paper [4]. Lastly in [3] CNN are used to extract automatically the relevant features. After the great success of NN they have been used broadly for all kind of tasks, and many advantages are still to be discovered. In this example CNN are used without considering what has been learned about PRNU in the previous works; the reason for this could be that the authors wanted to find a way to learn automatically what is relevant to identify the source without rely on human knowledge (approach that has been successful in other applications), or just because it is to be considered as a first and incomplete try to use CNN in image forensics. Anyway all these groups helped Mandelli et al. [4] to describe and implement a method that relies on all the relevant aspects of their discoveries and outperforms them.

REFERENCES

[1] M. Kharrazi, H. T. Sencar and N. Memon, "Blind source camera identification," 2004 International Conference on Image Processing, 2004.

[2] J. Lukas, J. Fridrich, and M. Goljan, "Digital camera identification from sensor pattern noise".

[3] L. Bondi, L. Baroffio, D. Guera, P. Bestagini, E. Delp, and S. Tubaro , "First Steps Toward Camera Model Identification with Convolutional Neural Networks".

[4] S. Mandelli, D. Cozzolino, P. Bestagini, L. Verdoliva and S. Tubaro, "CNN-Based Fast Source Device Identification"