

Received 17 March 2026, accepted 27 April 2026, date of publication 4 May 2026, date of current version 7 May 2026.

Digital Object Identifier 10.1109/ACCESS.2026.3689977

RESEARCH ARTICLE

When CNNs Truly Outperform PCE on Source Camera Attribution?

ANDREA MONTIBELLER¹, (Member, IEEE), MIGUEL MASCIOPINTO²,
FERNANDO PÉREZ-GONZÁLEZ³, (Fellow, IEEE), AND GIULIA BOATO¹, (Member, IEEE)

¹Department of Information Engineering and Computer Science, University of Trento, 38123 Trento, Italy

²Galician Research and Development Center in Advanced Telecommunications (GRADIANT), 36214 Vigo, Spain

³atlanTTic Research Center, Signal Processing in Communications Group, Universidade de Vigo, 36310 Vigo, Spain

Corresponding author: Andrea Montibeller (andrea.montibeller@unitn.it)

This work was supported in part by FEDER and Xunta de Galicia through the “Grupos de Referencia Competitiva” under Grant ED431C 2021/47, in part by FEDER and MCIN/AEI through FELDSPAR under Grant TED2021-130624B-C21, in part by the “NextGenerationEU/PRTR” through TRUFFLES, and in part by FF4ALL and SERICS through the MUR National Recovery and Resilience Plan funded by European Union-NextGenerationEU under Grant PE00000014.

ABSTRACT Source camera attribution is a central problem in multimedia forensics, traditionally addressed using the Peak-to-Correlation Energy (PCE) statistic. A recent study by Mandelli et al. claimed that trainable CNN-based detectors substantially outperform the PCE. In this work, we show that this conclusion is affected by a methodological issue: the evaluation setup adopted for the PCE places it at an inherent disadvantage, leading to an unfair performance comparison. After clarifying the correct use of the PCE detector and re-establishing a fair baseline, we introduce two improvements—a weighted cross-entropy loss to mitigate class imbalance, and a finetuning procedure that better reflects real-world scenarios—to enhance the CNN-based approaches. With these enhancements, some of the architectures proposed by Mandelli et al. indeed surpass a properly implemented PCE across a broader range of operating conditions. The gains are particularly pronounced for medium and large test image crop sizes.

INDEX TERMS Convolutional neural networks (CNNs), digital image forensics, fingerprint detection, image source attribution, photo response non-uniformity (PRNU), peak to correlation energy (PCE).

I. INTRODUCTION

Of all the features proposed for multimedia forensics since its inception, the Photo Response Non-Uniformity (PRNU) stands out as a remarkable success story [1], [2], and is customarily used in criminal investigations involving digital images due to its effectiveness and reliability. For this reason, research on PRNU continues to be a focal point in advancing forensic techniques.

The PRNU is a noise-like pattern inherent to the manufacturing process of the silicon wafer of all digital imaging sensors [2]. This distinctive pattern is widely utilized for forensic camera source identification and attribution tasks, owing to its unique and resilient nature inherent in all images captured using a digital imaging sensor device. Thus,

The associate editor coordinating the review of this manuscript and approving it for publication was Amin Zehtabian^{id}.

the PRNU acts as a device-specific fingerprint and can be estimated from a set of images taken with the target device. This estimate can be subsequently used to detect the presence of the device PRNU in a test image, thereby deciding whether the test image can be attributed to the target camera device [2].

Over the past eighteen years (see Table 1), several enhancements to the PRNU estimation and detection method proposed in [2] (still regarded as the *de facto* baseline) have been suggested [3], [4], [5], [6]. Furthermore, in [7], [8], and [9] the theoretical model has been revised and improved with a focus on increasing both the performance and the computational efficiency of PRNU-based source attribution methods.

For most of this period, the state-of-the-art (SoA) PRNU detector for source attribution has involved the evaluation of the Peak-to-Correlation Energy (PCE) statistic, which is a form of normalized cross-correlation [10], [11]. However,

since the assumptions on the PRNU that justify the use of the PCE do not exactly hold [7], [8], it is reasonable to consider more sophisticated detectors. In [12], Mandelli et al. proposed five convolutional neural network (CNN) architectures to replace the prevalent PCE-based detector: three EfficientNet variants (EffB0, EffB2, and EffB4), InceptionNet (Inc), and the Pair-wise Correlation Network (Pcn). This is reasonable in view of the fact that other CNN-based approaches have been successfully applied to related problems, such as camera model identification [13] or filtering for PRNU extraction [14]. Furthermore, all five CNN architectures in [12] allegedly outperform the PCE.

In this paper, we revisit the CNN-based PRNU detectors for forensic camera attribution proposed in [12]. Our first contribution is critical: a misinterpretation of the standard PCE detector [16] in [12] is identified, leading to an inconsistent and unfair comparison with CNN-based methods. By clarifying the correct PCE formulation, we demonstrate how the results reported in [12] need to be reconsidered in order to establish the genuine performance gains offered by deep learning architectures presented in [12]. Building on the corrected baseline, we propose two additional key improvements that reveal when CNNs genuinely outperform the PCE: 1) a weighted cross-entropy loss to mitigate class imbalance during training, ultimately achieving CNN performance comparable to that of the PCE; 2) a cost-effective finetuning strategy to reflect a realistic deployment scenario, and show that under these conditions, CNN-based methods outperform the PCE across an even wider range of operating settings.

We note that, although end-to-end deep architectures that jointly subsume forensic fingerprint extraction and detection have been proposed for device source attribution [15], [17], we do not consider them here, as recent evidence suggests that their apparent performance gains may result from dataset biases rather than genuine modeling advantages [18]. Moreover, their lack of explainability raises concerns for forensic use. Unlike PRNU-based detectors, whose behavior is grounded in a well-established and extensively studied physical characteristic of the imaging sensor, end-to-end deep architectures offer no comparable interpretability for expert testimony. Furthermore, purely data-driven models are highly susceptible to dataset leakage, where a network may learn spurious correlations based on image content rather than genuine sensor-level fingerprints.

We also note that, beyond device attribution, there is a related forensic task aimed at identifying the model or brand of the camera that captured a given image. Several works addressing this model/brand identification problem have likewise proposed end-to-end deep architectures. However, because these approaches do not rely on device-specific fingerprints (most notably the PRNU) they fall outside the scope of this paper.

The remainder of this paper is organized as follows. Section II introduces our mathematical notation and defines the PRNU [11]. Section III analyzes the misinterpretation

of the PCE implementation in [12] and [16] and presents our improvements to their CNN-based detectors. Section IV describes the experimental setup and Section V reports the results. Finally, Section VI provides our ablation study, and Section VII concludes the paper.

II. PRELIMINARIES

Matrices are denoted by boldface uppercase letters. For matrices $\mathbf{A}, \mathbf{B} \in \mathcal{M}_{M \times N}(\mathbb{R})$, the element in the i -th row and j -th column of $\mathbf{A}$ is denoted by $[\mathbf{A}]_{i,j}$. The Hadamard (i.e., element-wise) product is written as $\mathbf{A} \circ \mathbf{B}$. The element-wise p -th power of $\mathbf{A}$, for $p \in \mathbb{Z}$, is denoted by $\mathbf{A}^{\circ p}$, where its elements are given by $[\mathbf{A}^{\circ p}]_{i,j} = ([\mathbf{A}]_{i,j})^p$. Finally, for a given set $\mathcal{S}$, its cardinality is denoted by $|\mathcal{S}|$.

A. PHOTO RESPONSE NON-UNIFORMITY

In the classical PRNU detector [11], the PCE is evaluated on a pair composed of an estimated PRNU fingerprint $\hat{\mathbf{K}}$ of size $M \times N$ pixels (i.e. the native resolution of the imaging sensor) and a test residual $\mathbf{W}$ of size $M_t \times N_t$. This residual is obtained through a denoising operation on the test image $\mathbf{Y}$, while the PRNU fingerprint can be estimated from a set $\{\mathbf{Y}^{(l)}\}_{l=1}^L$ of L images available from the target camera sensor device as $\hat{\mathbf{K}} = \left(\sum_{l=1}^L \mathbf{Y}^{(l)} \circ \mathbf{W}^{(l)} \right) \circ \left(\sum_{l=1}^L (\mathbf{Y}^{(l)})^{\circ 2} \right)^{\circ -1}$ [11].

The source attribution problem can be posed as the following binary hypothesis test:

- H_0 : $\mathbf{W}$ and $\hat{\mathbf{K}}$ do not contain the same PRNU,
- H_1 : $\mathbf{W}$ and $\hat{\mathbf{K}}$ contain the same PRNU.

In case the test image $\mathbf{Y}$ is cropped to size $M_t \times N_t$, $M_t \leq M, N_t \leq N$ there is a potential misalignment of residual $\mathbf{W}$ with respect to $\hat{\mathbf{K}}$. To account for this, one may define a zero-padded residual $\mathbf{W}^{(0)}$ such that $[\mathbf{W}^{(0)}]_{m,n} = [\mathbf{W}]_{m,n}$ for all $(m,n) \in \{0, \dots, M_t - 1\} \times \{0, \dots, N_t - 1\}$, and $[\mathbf{W}^{(0)}]_{m,n} = 0$ whenever $m \in \{M_t, \dots, M - 1\}$ or $n \in \{N_t, \dots, N - 1\}$.

Then, the (s_1, s_2) th element of the circular cross-correlation matrix between the estimated PRNU and the possibly cropped residual is defined as

$$[\mathbf{C}]_{s_1, s_2} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} [\Delta_{(s_1, s_2)}(\mathbf{W}^{(0)}) \circ \hat{\mathbf{K}}]_{m,n}, \quad (1)$$

where operator $\Delta_{(s_1, s_2)}$ applies a right circular shift of (s_1, s_2) pixels with $\mathcal{S} = \{(s_1, s_2) \mid 0 \leq s_1 < M, 0 \leq s_2 < N\}$ the set of considered shifts.

Due to the possible misalignment caused by cropping, it is necessary to search for the maximum of the cross-correlation between the estimated PRNU and the residual over the set of all candidate shifts $\mathcal{Q} = \{(q_1, q_2) \mid 0 \leq q_1 \leq M - M_t, 0 \leq q_2 \leq N - N_t\}$. Let $\mathbf{C}_{\mathcal{Q}}^+ \doteq \max_{(q_1, q_2) \in \mathcal{Q}} [\mathbf{C}]_{q_1, q_2}$. Then the PCE over $\mathcal{Q}$ is defined as [4] and [10]:

$$\text{PCE}_{\mathcal{Q}} \doteq \frac{\text{sgn}(\mathbf{C}_{\mathcal{Q}}^+) \cdot (\mathbf{C}_{\mathcal{Q}}^+)^2}{\frac{1}{M \cdot N - |\mathcal{A}|} \sum_{(s_1, s_2), (s_1, s_2) \notin \mathcal{A}} ([\mathbf{C}]_{s_1, s_2})^2}, \quad (2)$$

TABLE 1. Summary of state-of-the-art (SoA) approaches for source camera attribution.

Authors	Year	Core Methodology	Forensic Focus
Goljan & Fridrich [2]	2008	PCE Statistics	Baseline Detection
Li [6]	2010	Enhanced Sensor Pattern Noise (SPN)	Noise Enhancement
Kang et al. [4]	2011	Phase-only Correlation	Cropping Robustness
Lawgaly et al. [5]	2014	Weighted Averaging	Noise Estimation
Bondi et al. [13]	2016	CNN	Model Identification
Pérez-González et al. [9]	2016	Sequential Detection	Fast Camera Fingerprinting
Masciopinto & Pérez-González [8]	2018	Revised PRNU Model	PRNU Theoretical Modeling
Kirchner & Johnson [14]	2019	SPN-CNN	CNN-based PRNU Extraction
Mandelli et al. [12]	2020	CNN (EffB, Pcn, Inc)	CNN-based Camera Source Attribution
Manisha et al. [15]	2022	End-to-End DL	End-to-End Camera Source Attribution
Montibeller & Pérez-González [3]	2023	Adaptive Radial Correction Inversion	PRNU-based Camera Source Attribution
Fernández-Menduiña et al. [7]	2023	PRNU Emphasis	Multiplicative Model Generalized
This Work	2026	CNN + PCE	Fair Benchmark and Improvements

where $\mathcal{A}$ is a small squared region around the position of the maximum of $[\mathbf{C}]_{q_1, q_2}$, and $\text{sgn}(x)$ is -1 when $x < 0$, and 1 when $x \geq 0$.

Crucially, notice that when the test image has not been cropped, there is only one candidate shift, i.e. $\mathcal{Q} = (0, 0)$, so $\mathbf{C}_{\mathcal{Q}}^+$ is directly $[\mathbf{C}]_{0,0}$. Thus, under no cropping, $\text{PCE}_{\mathcal{Q}}$ becomes $\text{PCE}_{(0,0)}$.

In the CNN-based approach from [12], the CNNs are trained and tested on pairs $(\hat{\mathbf{K}}, \mathbf{W})$ of the same size obtained from the camera devices of the dataset, and assuming they are geometrically aligned. A different CNN model is trained for each considered image size. This CNN outputs an identification score which is directly related to the probability that the test residual $\mathbf{W}$ contains the same PRNU as $\hat{\mathbf{K}}$. This score plays an equivalent role to that of the PCE in the source attribution hypothesis test.

III. REVISITING AND IMPROVING CNN-BASED PRNU DETECTION

In this section, we revisit the CNN-based PRNU detectors from [12], demonstrating how a misinterpretation of the PCE method [10] led to an overestimation of CNN detectors performance advantage. We then propose two modifications: a weighted cross-entropy loss and a finetuning strategy; that enable the CNN detectors to genuinely outperform the PCE-based method across a significantly wider range of operating conditions. Henceforth, we will use the subscript w to denote models trained with the weighted cross-entropy loss and the superscript F for finetuned models. The presence of both indicates that a model incorporates both improvements.

A. PROPERLY ACCOUNTING FOR ALIGNMENTS

The distribution of the cross-correlation values in (1) under H_0 hypothesis has been modeled by a zero-mean Gaussian distribution [10]. Therefore, when there is only one candidate shift in $\mathcal{Q}$, the probability of false positive can be approximated as $P_{F_0} \approx Q(\sqrt{\tau})$, where $Q(x) \doteq \frac{1}{2\pi} \int_x^\infty e^{-x^2/2} dx$, and τ is the PCE threshold to decide between H_0 and H_1 . Under

the reasonable assumption that the distribution of the PCE is independent and identically distributed for every possible candidate shift in $\mathcal{Q}$, the probability of false positive is

$$P_{FP} = 1 - (1 - P_{F_0})^{|\mathcal{Q}|} \approx |\mathcal{Q}| \cdot P_{F_0}. \quad (3)$$

where the approximation holds whenever $P_{F_0} \ll 1$, as it is sought in practice. It is thus evident that P_{FP} approximately grows linearly with the number of possible candidate shifts. In the case of cropping of the test image from $M \times N$ pixels to $M_t \times N_t$, we have $|\mathcal{Q}| = (M - M_t + 1) \cdot (N - N_t + 1)$.

In [12], all of the proposed models clearly outperform the PCE-based detector. However, when we tried to replicate the results using the authors' code published in [16] and [19], we found an unexpected low performance of the PCE-based detector. Analyzing the code in [16] and [19], we came to the conclusion that their implementation of the PCE-based detector takes the maximum over a set of $M \times N$ candidate shifts, that is, it computes $\text{PCE}_{\mathcal{S}}$.

We claim that using $\text{PCE}_{\mathcal{S}}$ instead of $\text{PCE}_{(0,0)}$ puts the PCE-based detector at an unfair disadvantage, because the CNN models proposed in [12] are trained and tested assuming perfect alignment under H_1 . As we have seen, carrying out the maximization of the PCE over $\mathcal{S}$ roughly amounts to increasing its FPR by a factor of $M \times N$, which is very large considering the sizes of the images in the dataset. In any case, notice that even if non-aligned cropping were considered, the correct set of candidate shifts would be $\mathcal{Q}$ instead of $\mathcal{S}$, since shifts like $(M_t + 1, N_t + 1)$ are unfeasible.

When, for a fairer comparison, the PCE-based detector is set to operate under the same conditions as the CNN-based ones (i.e. under the aligned case), the performance advantages of the latter become much more modest, if any. Moreover, note that in the cropped, non-aligned case, the CNN-based detector requires testing all possible candidate shifts individually, whereas the PCE may exploit computational savings leveraging the use of the Fast Fourier Transform [10].

B. WEIGHTED CROSS-ENTROPY LOSS

While the CNN models in [12] were trained using a standard sigmoid cross-entropy loss, we propose a weighted cross-entropy loss function to improve the true positive rate (TPR) without increasing the false positive rate (FPR). In fact, the weighted cross-entropy loss is often employed to mitigate class imbalance by penalizing errors on the minority class more heavily during training [20], [21], [22]. While training batches are balanced, following the method in [12], a fundamental imbalance exists in the underlying data population. Specifically, the set of potential negative (H_0) samples is vastly larger than that of positive (H_1) samples, as any mismatched pair of a camera fingerprint $\hat{\mathbf{K}}$ and a noise residual $\mathbf{W}$ constitutes an H_0 sample.

The weighted cross-entropy loss is defined as:

$$L = -\frac{1}{N} \sum_{i=1}^N w \cdot y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i), \quad (4)$$

where N is the number of samples in which the loss is evaluated (commonly the batch size), y_i is the class label of the i -th sample (0 for negative class, 1 for positive class), p_i is the predicted probability of the i -th sample from the CNN model output, and w is a hyperparameter that weighs the positive class in the loss function.

In our case, w is determined by maximizing the validation accuracy via a grid search. We begin by searching the interval $[0.0, 2.0]$ with a step size of 0.1. If the maximum accuracy occurs at the upper bound of this interval, the search is incrementally extended up to a maximum of $w = 4.0$ until the optimum is found.

As an example, Fig. 1 illustrates the validation accuracy as a function of w for the EffB4 model from [12] across several central crop sizes ($P \in \{80, 192, 256\}$). The optimal w values for each CNN architecture from [12] are reported in Table 2.

TABLE 2. Hyperparameter w maximizing the weighted cross-entropy computation for each of the CNN-based detectors proposed in [12].

P	w				
	EffB0	EffB2	EffB4	Inc	Pcn
80	2.5	2.5	2.5	2.	0.25
192	3.75	3.5	3.	2.5	3.25
256	1.75	1.25	1.5	1.25	1.75

C. FINETUNING WITH THE TARGET DEVICE

In practical camera attribution scenarios, a forensic analyst often encounters devices that were not seen during the training of a CNN-based PRNU detector. Since it is reasonable to expect (and will be confirmed experimentally in Sect. V) that such a detector performs better on devices included in the training set, we propose finetuning it for each new device. Formally, let $\mathcal{D}_t$ denote the set of devices used to train the CNN-based PRNU detector, and let d represent the new device. The finetuning process then uses pairs $(\hat{\mathbf{K}}, \mathbf{W})$ from the target device d in combination with the

initial training data from $\mathcal{D}_t$, so that the updated training and validation sets include images from $\mathcal{D}_t \cup \{d\}$. Specifically, all new positive-class samples (i.e., pairs under hypothesis H_1) are generated exclusively from a finetuning image set $\mathcal{I}_2^f$ known to originate from d , together with its estimated PRNU $\hat{\mathbf{K}}$. Since estimating $\hat{\mathbf{K}}$ for any device d inherently requires multiple images, we assume that, in practice, collecting a small number of additional samples to finetune the CNN-based PRNU detector for d is always feasible.

The proposed finetuning stage is carried out by setting the maximum number of epochs to 60 and the learning rate to 10^{-4} , while the other hyperparameters retain the same values as in pre-training. This process yields a version of the pre-trained CNN-based PRNU detectors with weights specifically adapted to the single target device d , used during finetuning. Moreover, while retraining the entire network from scratch with an augmented training set (i.e., $\mathcal{D}_t \cup \{d\}$) could potentially yield even better results, we avoid this approach due to its prohibitive computational cost. As a result, the average number of epochs required to reach the training stop condition is much lower than that needed to obtain the initial pretrained model (additional details are provided in Sect. V-C).

IV. EXPERIMENTAL SET-UP

We use the same dataset, i.e. devices from VISION and Dresden datasets [23], [24], as well as the same training, validation and test partitions as in [12]. While these datasets reflect specific camera-sensor characteristics and compression pipelines, their use provides a standardized framework based on well-established benchmarks for camera source attribution. This is essential for ensuring fair comparison with [12] and for guaranteeing the full reproducibility of our experimental results.

The dataset used in [12] consists of a collection of devices $\mathcal{D}$, originally divided into two disjoint subsets of camera devices: $\mathcal{D}_1$ and $\mathcal{D}_2$. As illustrated in Fig. 2, images from devices in $\mathcal{D}_1$ were partitioned into training ($\mathcal{I}_1^t$), validation ($\mathcal{I}_1^v$), and part of the test ($\mathcal{I}_1^e$) sets. The remaining portion of the test set consisted of images $\mathcal{I}_2$ from devices in $\mathcal{D}_2$, which were never seen during training. Consequently, the full test set is defined as $\mathcal{I}_e = \mathcal{I}_1^e \cup \mathcal{I}_2$ for all experiments, except those assessing the performance gain from finetuning, as explained next.

To implement and evaluate the proposed finetuning strategy (see Sect. III-C), we further partitioned $\mathcal{I}_2$ into two equally sized subsets: a finetuning set $\mathcal{I}_2^f$ and a test set $\mathcal{I}_2^e$, resulting in a 50/50 split. Based on this partitioning, the evaluation test set is defined exclusively by $\mathcal{I}_2^e$ when measuring the performance gain from finetuning.

Following the pipeline of Fig. 3, we perform device-wise finetuning implementing the exact protocol proposed in Section III-C. Let $d_i \in \mathcal{D}_2$, for $i = 1, \dots, |\mathcal{D}_2|$, denote a target device on which we finetune the pre-trained CNN-based PRNU detector, and $\mathcal{I}_{2,i}^f \subset \mathcal{I}_2^f$ be the subset of

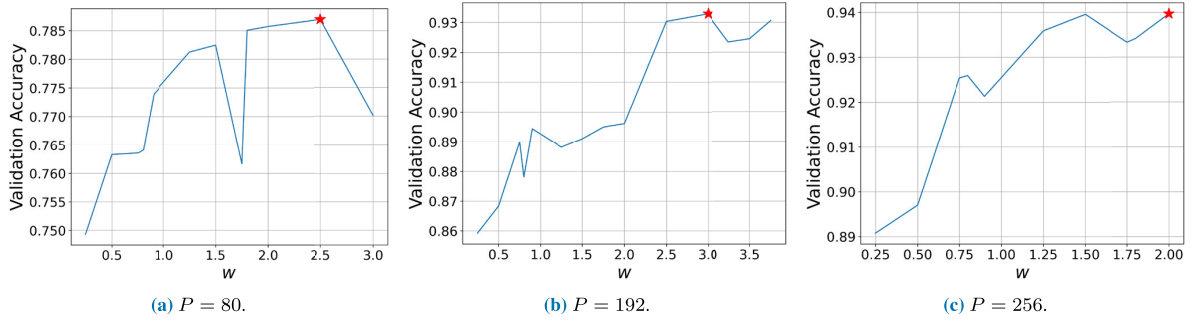


FIGURE 1. Validation accuracy vs. weighted cross-entropy parameter w for EffB4 detector. The values of w that achieved the highest validation accuracy are marked with a red star $\star$.

finetuning images corresponding to d_i . During this process, images from $\mathcal{I}_{2,i}^f$ are combined with those from the original training set $\mathcal{I}_1^f$.

The detector finetuned on d_i is then evaluated on $\mathcal{I}_2^e$: images $\mathcal{I}_{2,i}^e$ serve as H_1 examples, while the set $\mathcal{I}_2^e \setminus \mathcal{I}_{2,i}^e$ provides H_0 examples. This finetuning and evaluation procedure is repeated independently for each $d_i \in \mathcal{D}_2$, and the resulting per-device ROC curves are averaged to produce the results reported in Sect. V.

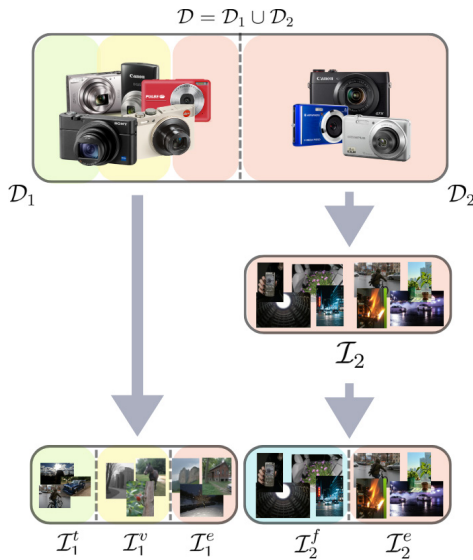


FIGURE 2. Partition of the device dataset $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$ used in the experiments. Image subsets $\mathcal{I}_1^f$, $\mathcal{I}_1^v$, and $\mathcal{I}_1^e$ represent the training, validation, and partial test sets, respectively. Following [12], the full test set $\mathcal{I}_e$ is defined as $\mathcal{I}_e = \mathcal{I}_1^e \cup \mathcal{I}_2^e$. In the finetuning scenario, models are trained using $\mathcal{I}_2^f$ and evaluated on $\mathcal{I}_1^e \cup \mathcal{I}_2^e$.

The rest of the experimental setup (including training and testing hyper-parameters) is identical to the one described in [12].¹ The only differences lying in the following two aspects. First, the treatment of the Olympus devices from [24] that, while they were considered as a single device in [12], are treated as five separate devices in [24], so we adopt this approach when constructing $\mathcal{D}_1$. This results in $|\mathcal{D}_1| = 40$,

¹We extend our gratitude to Dr. Sara Mandelli for providing us with the specific image training, validation and testing partitions for each device in $\mathcal{D}$.

$|\mathcal{D}_2| = 51$, and $|\mathcal{D}| = |\mathcal{D}_1| + |\mathcal{D}_2| = 91$. Second, in contrast with [16], we disabled CPU parallelization for data loading. We found this step was necessary to eliminate non-determinism in the data sampling process and ensure the full reproducibility of our results.

The pairs $(\hat{\mathbf{K}}, \mathbf{W})$ are obtained in exactly the same way as explained in [12]. We report the performance of the PRNU detectors for test image sizes $P \times P$ pixels, with $P \in \{80, 192, 256\}$, applying the same central image cropping as in [12], therefore evaluating the performance of the PRNU detectors for different image resolutions on the geometrically aligned case. In the CNNs training phase, data augmentation by 90° rotation and horizontal flipping is randomly applied on each training pair to increase the number of training samples as in [16].

Models' performance is evaluated using several key metrics, including Receiver Operating Characteristic (ROC) curves. We plot the True Positive Rate (TPR) against the False Positive Rate (FPR) using either a linear or a semi-logarithmic scale. The latter is selected to enhance the visualization of performance differences at low FPRs (≤ 0.01), which is standard for practical source attribution algorithms [2], [3]. As a summary measure, we also report the Area Under the Curve (AUC). Finally, to compare each architecture against the baseline, we identify the cross-over FPR, which we define as the FPR value above which that architecture achieves a higher TPR than the $\text{PCE}_{(0,0)}$ detector.

V. EXPERIMENTAL RESULTS

In [12], five different CNN architectures are compared against the non-aligned PCE-based detector PCE_S . However, as we have argued in Sec. III-A, the comparison should have been conducted against $\text{PCE}_{(0,0)}$.

Throughout the subsequent discussion, we replicate the experiments from [12] including our improved architectures and the aligned PCE detector, $\text{PCE}_{(0,0)}$, in the performance evaluation. Our refined models demonstrate significant gains in terms of ROC curves, AUC, and FPR cross-over points. In addition, in our ablation study (described in Section VI) we discuss how the fine-tuning device cardinality $|\mathcal{I}_{2,i}^f|$, the image pixel activity, and the filters used for PRNU estimation can impact the overall performance of the CNN-based detector.

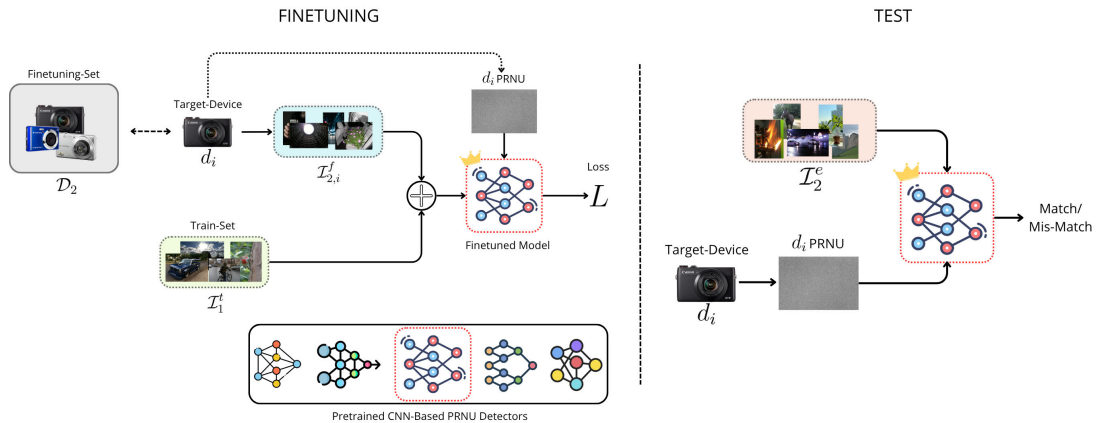


FIGURE 3. Flowchart of the finetuning and evaluation pipelines. Pretrained detectors (i.e. EffB0, EffB2, EffB4, Inc, and Pcn) are finetuned independently for each target device d_i and tested afterwards. Evaluation is performed on $\mathcal{I}_2^e$ to assess the finetuning impact, using images from d_i for the matching scenario (H_1) and images from other devices in $\mathcal{D}_2$ for the non-matching scenario (H_0).

We recall that the superscript F and subscript w denote the application of the finetuning strategy and the weighted cross-entropy loss function, respectively. For instance, EffB4 $_w^F$ refers to the refined EffB4 model trained using both the weighted cross-entropy loss specified in (4) and the finetuning strategy detailed in Sec. III-C.

A. CNN-BASED PRNU DETECTION VS. $PCE_{(0,0)}$

This section compares the performance of the CNN-based PRNU detectors and the non-aligned PCE detector PCE_S from [12] against the aligned PCE detector $PCE_{(0,0)}$. The evaluation is based on ROC curves and AUC metrics.

Figure 4 presents the ROC curves with logarithmic x-axis, for the CNN-based detectors from [12] alongside the PCE-based detectors (i.e., $PCE_{(0,0)}$ and PCE_S). The evaluation was performed on the $\mathcal{I}_e$ test set using images of size $P \times P$, where $P \in \{80, 192, 256\}$. We note that, as did in [12], the evaluation of detectors on $P \in \{80, 192, 256\}$ serves as a rigorous assessment of robustness and generalization capabilities under data constrained conditions where only fragments of an image may be available during multimedia forensic analysis (e.g. camera source attribution, forgery detection and localization).

It is important to note that $\mathcal{I}_e$ is the set of devices used for testing in [12]. Consequently, our ROC curves for the CNNs and the PCE_S detector on this set mirror the findings originally reported in that work. However, as illustrated in Fig. 4, the performance advantage of the CNN-based detectors is substantially diminished when benchmarked against the correctly implemented aligned detector, $PCE_{(0,0)}$, instead of the non-aligned PCE_S .

This performance gap is particularly evident for smaller images ($P = 80$), where $PCE_{(0,0)}$ is superior in both ROC and AUC metrics. Conversely, for larger images ($P \geq 192$), some CNNs do achieve a higher AUC than $PCE_{(0,0)}$. Nevertheless, in the critical low-FPR region corresponding to $FPR \leq 0.01$, the $PCE_{(0,0)}$ detector consistently yields a higher TPR than any method from [12].

Furthermore, we stress that the ROC curves for the non-aligned detector PCE_S , retained in Fig. 4 for reference, do not represent a fair benchmark against any CNN architectures, as detailed in Sec. III-A. Consequently, when compared with respect to $PCE_{(0,0)}$, performance improvements for CNN-based detectors are restricted to operating regimes with $FPR \gg 0.01$, which are generally outside the practical range for source attribution tasks [2], [3].

Finally, to analyze the CNN detector's generalization performance, we introduce a second experiment in Fig. 5. Here, for $P = 80$, we measure the corresponding performance improvement (resp. degradation) of EffB4 (the best performing CNN-based detectors in Fig. 4), when evaluating on images $\mathcal{I}_1^e$ from devices previously encountered (resp. not encountered $\mathcal{I}_2^e$) during training, in terms of both ROC and AUC. The figure clearly shows that EffB4 performs much better on images $\mathcal{I}_1^e$ from previously seen devices, and underperforms on unseen ones from $\mathcal{I}_2^e$, based on both ROC and AUC metrics. As discussed in Sec. III-C, the scenario of always testing on images $\mathcal{I}_1^e$ from previously seen devices is impractical in the real world, as it would require retraining the model from scratch on the test set. However, we will demonstrate that this underlying concept can be effectively leveraged through the finetuning strategy proposed in this paper.

B. IMPROVING CNN-BASED PRNU DETECTION

This section evaluates the performance gains obtained by applying the enhancements from Sec. III to the CNN-based detectors of [12]. The improvements are measured in terms of ROC, AUC, and cross-over FPR.

First, Fig. 6 shows the ROC curves and AUC scores on the test set $\mathcal{I}_2^e$ for the EffB4 architecture—the top-performing CNN in Sect. V-A—both with and without the improvements proposed in Sect. III-B and Sect. III-C. Next, Tables 3 and 4 detail the FPR cross-over points for all CNNs relative to the $PCE_{(0,0)}$ detector under two distinct experimental

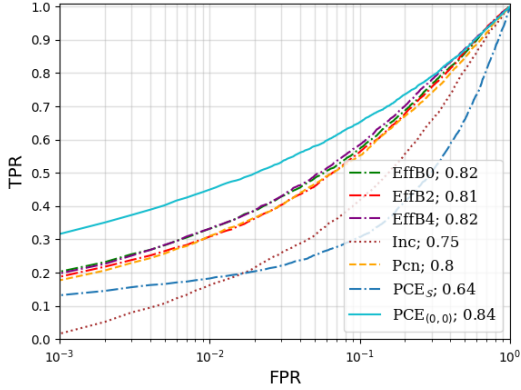
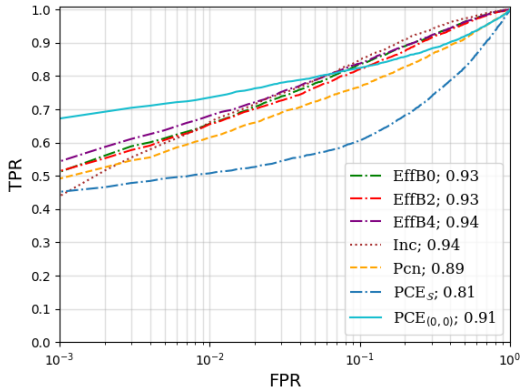
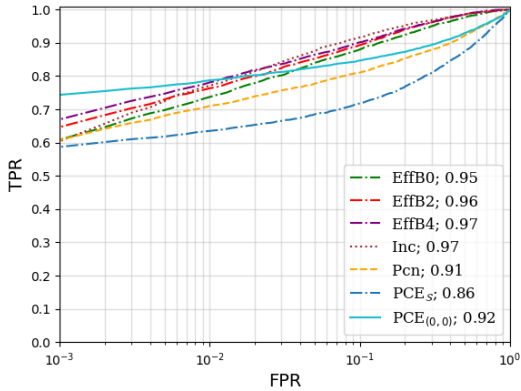
(a) $P = 80$.(b) $P = 192$.(c) $P = 256$

FIGURE 4. ROC curves of CNN-based detectors in [12] and detectors PCE_S , $PCE_{(0,0)}$ when tested on $\mathcal{I}_e$. Decimals values in the legends refers to the corresponding AUCs.

conditions. Specifically, Table 3 reports the results on the $\mathcal{I}_1^e$ set when implementing only the weighted cross-entropy loss. Meanwhile, Table 4 shows the results on the $\mathcal{I}_2^e$ set when using the weighted loss in combination with our finetuning strategy.

We recall that the subscript w and superscript F denote the use of the weighted cross-entropy loss and the finetuning strategy, respectively.

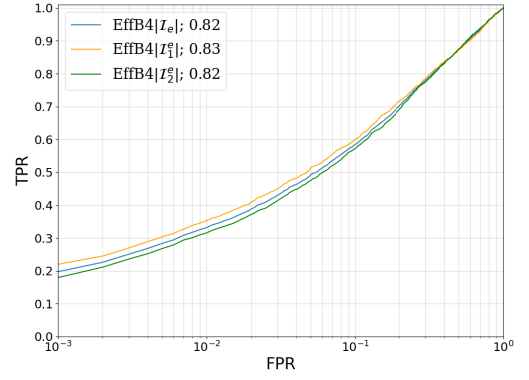


FIGURE 5. ROC curves using log-scale x-axis for the detector EffB4 when tested respectively on $\mathcal{I}_e$, $\mathcal{I}_1^e$ and $\mathcal{I}_2^e$ for image size lengths $P = 80$. The decimal numbers in the legend indicate the corresponding AUC.

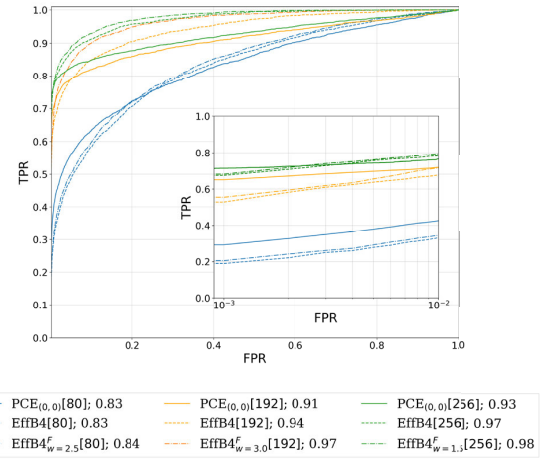


FIGURE 6. ROC curves for detectors $PCE_{(0,0)}$, EffB4, and EffB4^F when tested respectively on $\mathcal{I}_2^e$ for different image side lengths P (indicated within brackets). The decimal numbers in the legend indicate the corresponding AUC. The inset box contains a zoomed-in view for FPRs < 0.01 using a log-scale for the x-axis.

As shown in Fig. 6, our enhanced EffB4^F model achieves a substantial performance improvement in terms of ROC and AUC metrics compared to both the standard EffB4 from [12] and the aligned $PCE_{(0,0)}$ detector. Specifically, EffB4^F obtains the highest AUC across all image sizes. Moreover, while its performance margin over $PCE_{(0,0)}$ is limited for smaller images ($P = 80$) in the critical region of $FPR < 0.01$, its ROC curve becomes clearly superior for larger images ($P = 192$ and $P = 256$). This finding is further corroborated by the results in Tables 3 and 4, which present the FPR cross-over points for our improved CNNs compared to the original versions from [12], evaluated on the $\mathcal{I}_1^e$ and $\mathcal{I}_2^e$ test sets, respectively.

Table 3 shows that applying the weighted cross-entropy loss yields consistent improvements for all architectures from [12] when evaluated on the $\mathcal{I}_1^e$ test set. This enhancement produces TPRs comparable to the $PCE_{(0,0)}$ detector but at substantially lower FPRs compared to those obtained by the CNNs of [12]. These gains are further validated in Table 4, which demonstrates that combining the weighted loss with the proposed finetuning on $\mathcal{I}_2^e$ leads to significant

improvements for all CNNs across all image sizes P . On the other hand, while notable gains are evident even for $P = 80$, the benefits are more pronounced for larger images. For $P = 192$, the FPR cross-over point consistently approaches the critical value of 0.01, and for $P = 256$, several models (EffB0_w^F, EffB2_w^F, EffB4_w^F, and Inc_w^F) achieve a cross-over at an FPR ≤ 0.01 .

The only exception to these trends is the performance of the Pcn detectors. This is attributable to Pcn shallower architecture; with fewer convolutional layers, their capacity for improvement from our proposed optimizations is inherently limited.

TABLE 3. Model cross-over FPRs with respect to $PCE_{(0,0)}$, using $\mathcal{I}_1^e$. Best results are in bold.

P	80	192	256
EffB0	0.72	0.06	0.03
EffB0 _w	0.2	0.02	0.008
EffB2	0.62	0.09	0.01
EffB2 _w	0.43	0.01	0.006
EffB4	0.88	0.04	0.01
EffB4 _w	0.32	0.009	0.008
Inc	0.95	0.03	0.006
Inc _w	0.49	0.009	0.006
Pcn	0.95	0.61	0.58
Pcn _w	0.96	0.49	0.55

TABLE 4. Networks cross-over FPRs with respect to $PCE_{(0,0)}$, using $\mathcal{I}_2^e$. Best results are in bold.

P	80	192	256
EffB0	0.61	0.1	0.05
EffB0 _w	0.31	0.05	0.02
EffB0 _w ^F	0.47	0.05	0.02
EffB0 _w ^F	0.28	0.02	0.007
EffB2	0.50	0.13	0.03
EffB2 _w	0.46	0.06	0.03
EffB2 _w ^F	0.45	0.07	0.03
EffB2 _w ^F	0.31	0.02	0.01
EffB4	0.32	0.08	0.02
EffB4 _w	0.25	0.02	0.01
EffB4 _w ^F	0.23	0.02	0.006
EffB4 _w ^F	0.21	0.01	0.005
Inc	1.	0.12	0.03
Inc _w	0.62	0.05	0.03
Inc _w ^F	0.88	0.05	0.01
Inc _w ^F	0.56	0.02	0.01
Pcn	0.77	0.63	0.76
Pcn _w	0.80	0.62	0.79
Pcn _w ^F	0.76	0.6	0.66
Pcn _w ^F	0.78	0.51	0.65

C. FINETUNING COMPUTATIONAL COST IMPROVEMENT

In this section, we measure the gain in terms of computational cost obtained when finetuning a detector on a new device

using the pipeline proposed in Sec. III-C, with respect to retraining the models from scratch. Table 5 quantifies these computational savings by reporting the average number of epochs ($\forall d_i \in \mathcal{D}_2$) required to reach the stopping condition. The table contrasts the models from [12] (denoted Pcn, EffB0, etc.), which were trained from scratch on the $\mathcal{D}_1 \cup \{d_i\}$ dataset, with our proposed enhanced detectors (Pcn_w^F, EffB0_w^F, etc.). Our models were pre-trained on $\mathcal{D}_1$ using the weighted loss in (4) and then, as shown in Fig. 3, finetuned independently on the specific device $d_i \in \mathcal{D}_2$. The comparison reveals a significant reduction in training epochs for all finetuned models. Moreover, this finetuning process enhances performance stability across devices by mitigating biases associated with the original training set composition.

TABLE 5. Average number of epochs (Avg. epochs) to reach the training stop condition, and average savings when using finetuning with respect to training from scratch. Best results are in bold.

Detectors	$P=80$		$P=192$		$P=256$	
	Avg. epochs	Savings	Avg. epochs	Savings	Avg. epochs	Savings
Pcn	161	73%	162	76%	209	84%
Pcn _w ^F	43.7		39.05		34.25	
EffB0	209	92%	250	76%	215	80%
EffB0 _w ^F	16.17		60.97		43.16	
EffB2	170	69%	206	80%	289	85%
EffB2 _w ^F	53.19		41.61		44.47	
EffB4	265	86%	297	96%	409	89%
EffB4 _w ^F	37.8		13.11		44.48	
Inc	191	81%	251	88%	294	90%
Inc _w ^F	35.75		30.17		29.64	

VI. ABLATION STUDIES

In this section, we investigate three key design choices for the solutions proposed in this paper. First, we provide insights into the role played by the number of finetuning images used by each device $|\mathcal{I}_{2,i}^f|$. Second, we illustrate how CNN-based detectors achieve higher output stability in the presence of images with high pixel activity, specifically regarding texture and luminance. Finally, we evaluate the influence of employing a denoising filter during testing that differs from the one on which the best-performing CNN-based detector, the EffB4_{w=3.0}^F, was trained.

A. FINETUNING DEVICE CARDINALITY

In Fig. 7, we evaluate EffB4_{w=3.0}^F on $\mathcal{I}_2^e$ with $P = 192$ using a variable number s of finetuning images per device $|\mathcal{I}_{2,i}^f|$. In fact, as shown in Table 6, not all devices provide the same number of finetuning images. Specifically, over half of the devices are represented by approximately 123 images, while others range between 75 and 200. Our goal is to quantify the impact that s have on the final detection accuracy. To do so, we set $s \in \{10, 25, 50, 75, 100, \text{ALL}\}$, where ALL refers to the same results presented in Fig. 6.

From Fig. 7, we can observe that maximizing $|\mathcal{I}_{2,i}^f|$ (e.g. $s = \text{ALL}$) is not the sole factor for success. While EffB4_{w=3.0,s=ALL}^F outperforms most of the other configurations in terms of AUC, it performs worse for FPR < 0.01 than those using $s \in \{50, 75, 100\}$, and achieves lower AUC when compared with EffB4_{w=3.0,s=75}^F (which

corresponds to the minimum $|\mathcal{I}_{2,i}^f|$ reported in Table 6) and $\text{EffB4}_{w=3.0, s=100}^F$. These results suggest that dataset balance during fine-tuning may play a role in improving model performance at low FPR (which is the core focus of our study), but further experimentation is necessary to validate this conjecture.

TABLE 6. Cardinality $|\mathcal{I}_{2,i}^f|$ across the 51 finetuning devices.

Device	Count	Device	Count
D09	109	Nikon CoolPixS710 1	98
D10	89	Nikon CoolPixS710 2	92
D14	105	Nikon CoolPixS710 3	80
D15	114	Nikon CoolPixS710 4	89
D18	102	Nikon D200 1	190
D26	75	Nikon D70 1	95
D27	127	Nikon D70s 1	95
D29	112	Panasonic DMC FZ50 1	208
D30	136	Panasonic DMC FZ50 2	126
D32	118	Pentax OptioA40 1	79
D33	78	Pentax OptioA40 2	78
D35	77	Pentax OptioA40 3	78
Canon Ixus70 1	97	Praktica DCZ5 9 1	100
Canon Ixus70 2	93	Praktica DCZ5 9 2	103
Casio EX Z150 1	95	Praktica DCZ5 9 3	103
Casio EX Z150 2	94	Praktica DCZ5 9 4	100
Casio EX Z150 3	94	Ricoh GX100 1	79
Casio EX Z150 4	91	Ricoh GX100 2	83
FujiFilm FinePixJ50 1	103	Ricoh GX100 3	83
FujiFilm FinePixJ50 2	108	Ricoh GX100 4	79
Rollei RCP 7325XS 1	97	Sony DSC T77 0	91
Rollei RCP 7325XS 2	99	Sony DSC T77 1	86
Samsung NV15 1	107	Sony DSC T77 2	95
Samsung NV15 2	107	Sony DSC T77 3	91
Sony DSC H50 0	142	Sony DSC W170 0	98
Sony DSC H50 1	129		
Total Residual Count		5197	

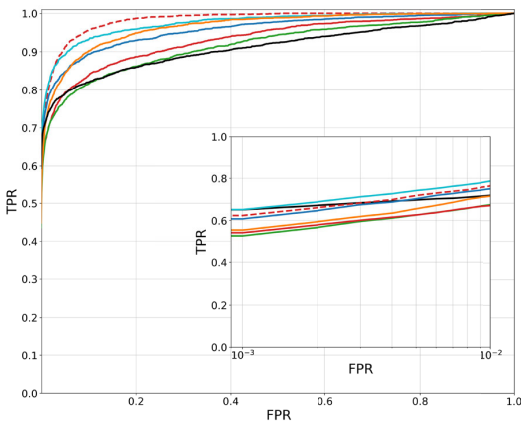


FIGURE 7. Performance evaluation of $\text{EffB4}_{w=3.0}^F$ across different finetuning sample sizes s .

B. IMPACT OF IMAGE PIXEL ACTIVITY

We evaluate the performance degradation of CNN-based and PCE-based detectors as a function of image pixel activity. We define pixel activity as the presence of textures and variations of luminance within each of the analyzed

192×192 image crops. To quantify this activity, we employ the normalized standard deviation, defined as $\sigma(\mathcal{I}_{2,i}^e)/\bar{\mathcal{I}}_{2,i}^e$, where $\mathcal{I}_{2,i}^e$ is the test image crop, $\bar{\mathcal{I}}_{2,i}^e$ is its average pixel value, and $\sigma(\mathcal{I}_{2,i}^e)$ its standard deviation. Additionally, we measure the normalized average gradient magnitude (AGM) as $\text{AGM}(\mathcal{I}_{2,i}^e)/\bar{\mathcal{I}}_{2,i}^e$, where AGM is computed by first applying Sobel filters in the horizontal and vertical directions, then calculating the gradient magnitude from these responses, and finally taking the standard deviation of the resulting magnitudes. In this experiment, we compared the proposed $\text{EffB4}_{w=3.0}^F$ with $\text{PCE}_{(0,0)}$.

As observed in Fig. 8, and consistent with existing literature [2], [3], our findings indicate that the PCE is significantly more sensitive to pixel activity, such as highly textured and luminance image crops. In contrast, $\text{EffB4}_{w=3.0}^F$ demonstrates superior robustness to these variations. This stability likely stems from the deep convolutional layers of the EfficientNet architecture, which effectively filter the device PRNUs and potentially prioritize more resilient forensic artifacts. Moreover, these results suggest that CNN-based detectors are expected to perform better for images characterized by high pixel activity, while PCE-based detectors may be more suitable in the opposite cases, a direction we intend to explore in future work.

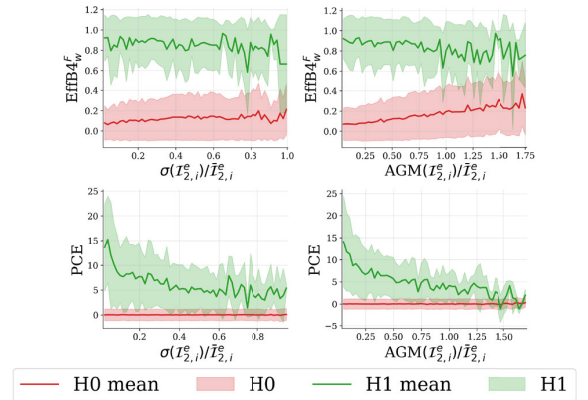


FIGURE 8. Analysis of detector sensitivity to image pixel activity.

C. IMPACT OF NOISE CHARACTERISTICS

Motivated by the impact that image pixel activity has on CNN-based detectors, we investigated whether these models are robust to alterations in the input noise characteristics. To do so, we evaluated $\text{EffB4}_{w=3.0}^F$, fine-tuned on residuals extracted from $\mathcal{I}_2^f$ using the pipeline described in Sect. II, on noise residuals estimated from $\mathcal{I}_2^e$ using the Dual-Tree Complex Wavelet Transform (DTCWT) denoising filter proposed in [25]. This filter has been successfully applied in prior work for source camera attribution [7], [26], providing a validated framework for evaluating model sensitivity to noise variations. By leveraging this approach, we directly test whether the learned CNN representations remain stable when the underlying noise residuals are altered.

As illustrated in Fig. 9, the specific filter employed during the finetuning and training phase significantly influences the model predictive accuracy. For $\text{EffB4}_{w=3,0}^F$, which was trained on residuals as described in Sect. II, and tested on $\mathcal{I}_2^c$ using the DTCWT denoising filter, the corresponding ROC reflects the same results obtained by the non-finetuned detector EffB4. Consequently, maintaining consistency between the filtering techniques used in both the training and inference stages is paramount for ensuring the reliability of these PRNU detectors.

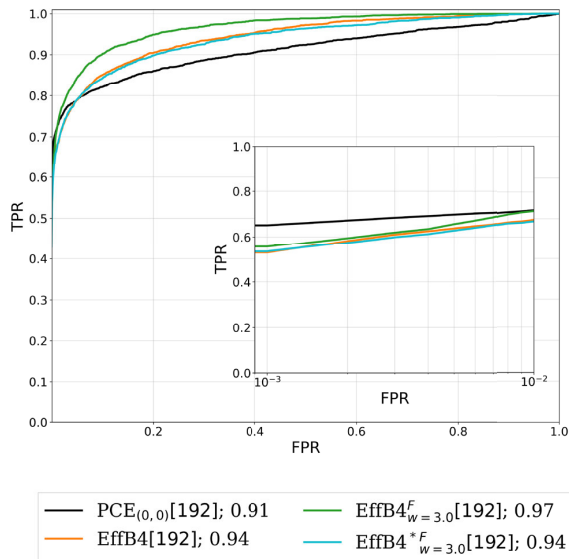


FIGURE 9. ROC curves illustrating the impact of using inconsistent denoising filters between training and testing.

VII. CONCLUSION

In this paper, we revisited the CNN-based approach to camera source attribution introduced in [12] and clarified a fundamental methodological issue affecting its comparison with PCE-based detectors. We showed that the evaluation setup used for the PCE in previous literature implicitly assumed spatial misalignment of test crops, which led to a substantially higher number of correlation tests and an inflated false positive rate. When the comparison is conducted under the correct aligned setting, the PCE remains competitive and outperforms CNN based methods in low FPR operating regions, particularly for small crop sizes ($P = 80$).

Building on this corrected baseline, we introduced two enhancements: fine tuning and a weighted cross entropy loss, which allow some of the CNN architectures from [12] to surpass a properly implemented PCE over a broader range of operating conditions. Specifically, the best performing architecture, the $\text{EffB4}_{w=3,0}^F$, outperforms the other detectors in terms of AUC, reaching values of 0.98 for $P = 256$, 0.97 for $P = 192$, and 0.84 for $P = 80$. Furthermore, the $\text{EffB4}_{w=3,0}^F$ model consistently improved upon the standard EffB4 in terms of FPR cross-over, reaching cross over points as low as 0.008 for $P = 256$, 0.009 for $P = 192$, and 0.32 for

$P = 80$, demonstrating superior performance in the critical low FPR region (< 0.01) required for forensic applications.

Despite these improvements, achieving performance superior to the PCE at extremely low false positive rates for the smallest crops ($P = 80$) remains an open challenge. Future research should focus on optimizing training dynamics for small scale forensic traces to bridge this remaining gap.

REFERENCES

- [1] 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 Process. Mag.*, vol. 40, no. 5, pp. 67–79, Jul. 2023.
- [2] M. Chen, J. Fridrich, M. Goljan, and J. Lukas, "Determining image origin and integrity using sensor noise," *IEEE Trans. Inf. Forensics Security*, vol. 3, no. 1, pp. 74–90, Jan. 2008.
- [3] A. Montibeller and F. Pérez-González, "An adaptive method for camera attribution under complex radial distortion corrections," *IEEE Trans. Inf. Forensics Security*, vol. 19, pp. 385–400, 2024.
- [4] X. Kang, Y. Li, Z. Qu, and J. Huang, "Enhancing source camera identification performance with a camera reference phase sensor pattern noise," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 2, pp. 393–402, Apr. 2012.
- [5] A. Lawgaly, F. Khelifi, and A. Bouridane, "Weighted averaging-based sensor pattern noise estimation for source camera identification," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 5357–5361.
- [6] C.-T. Li, "Source camera identification using enhanced sensor pattern noise," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 2, pp. 280–287, Jun. 2010.
- [7] S. Fernández-Mendiña, F. Pérez-González, and M. Masciopinto, "Source camera attribution via PRNU emphasis: Towards a generalized multiplicative model," *Signal Process., Image Commun.*, vol. 114, May 2023, Art. no. 116944.
- [8] M. Masciopinto and F. Perez-Gonzalez, "Putting the PRNU model in reverse gear: Findings with synthetic signals," in *Proc. 26th Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2018, pp. 1352–1356.
- [9] F. Pérez-González, M. Masciopinto, I. González-Iglesias, and P. Comesaña, "Fast sequential forensic detection of camera fingerprint," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 3902–3906.
- [10] M. Goljan and J. Fridrich, "Camera identification from cropped and scaled images," *Proc. SPIE*, vol. 6819, pp. 154–166, 2008.
- [11] M. Goljan, J. Fridrich, and T. Filler, "Large scale test of sensor fingerprint camera identification," *Proc. SPIE*, vol. 7254, pp. 170–181, 2009.
- [12] S. Mandelli, D. Cozzolino, P. Bestagini, L. Verdoliva, and S. Tubaro, "CNN-based fast source device identification," *IEEE Signal Process. Lett.*, vol. 27, pp. 1285–1289, 2020.
- [13] L. Bondi, L. Baroffio, D. Güera, P. Bestagini, E. J. Delp, and S. Tubaro, "First steps toward camera model identification with convolutional neural networks," *IEEE Signal Process. Lett.*, vol. 24, no. 3, pp. 259–263, Mar. 2017.
- [14] M. Kirchner and C. Johnson, "SPN-CNN: Boosting sensor-based source camera attribution with deep learning," in *Proc. IEEE Int. Workshop Inf. Forensics Security (WIFS)*, Dec. 2019, pp. 1–6.
- [15] Manisha, C.-T. Li, X. Lin, and K. A. Kotegar, "Beyond PRNU: Learning robust device-specific fingerprint for source camera identification," *Sensors*, vol. 22, no. 20, p. 7871, Oct. 2022.
- [16] *CNN-fast-SDI*, Image and Sound Processing Lab, GitHub Code Repository, 2024.
- [17] J. Lu, C. Li, X. Huang, C. Cui, and M. Emam, "Source camera identification algorithm based on multi-scale feature fusion," *Comput., Mater. Continua*, vol. 80, no. 2, pp. 3047–3065, 2024.
- [18] A. A. Mostafa, F. Pérez-González, and M. Masciopinto, "Exploring the pitfalls of black boxes in media forensics: A case study in source camera identification," *IEEE Access*, vol. 13, pp. 76528–76547, 2025.
- [19] *PRNU-Python*, Image and Sound Processing Lab, GitHub Code Repository, 2024.
- [20] M. Kukar and I. Kononenko, "Cost-sensitive learning with neural networks," in *Proc. ECAI*, 1998, pp. 445–449.
- [21] C. Elkan, "The foundations of cost-sensitive learning," in *Proc. Int. Joint Conf. Artif. Intell.*, 2001, pp. 973–978.

- [22] Y. S. Aurelio, G. M. de Almeida, C. L. de Castro, and A. P. Braga, "Learning from imbalanced data sets with weighted cross-entropy function," *Neural Process. Lett.*, vol. 50, no. 2, pp. 1937–1949, Oct. 2019.
- [23] D. Shullani, M. Fontani, M. Iuliani, O. A. Shaya, and A. Piva, "VISION: A video and image dataset for source identification," *EURASIP J. Inf. Secur.*, vol. 2017, no. 1, pp. 1–16, Dec. 2017.
- [24] T. Gloe and R. Böhme, "The 'dresden image database' for benchmarking digital image forensics," in *Proc. ACM Symp. Appl. Comput.*, Mar. 2010, pp. 1584–1590.
- [25] I. Selesnick, R. G. Baraniuk, and N. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Process. Mag.*, vol. 22, no. 6, pp. 123–151, 2005.
- [26] H. Zeng, Y. Wan, K. Deng, and A. Peng, "Source camera identification with dual-tree complex wavelet transform," *IEEE Access*, vol. 8, pp. 18874–18883, 2020.



ANDREA MONTIBELLER (Member, IEEE) received the degree in science (information and communication engineering) and the Ph.D. degree in information and communication technology from the University of Trento, Trento, Italy, in 2020 and 2024, respectively. In 2018, he was a Visiting Researcher with the Department of Signal Theory and Communications, University of Vigo, Spain. In 2023, he was a Research Scholar with the Department of Electrical and Computer

Engineering, Drexel University, Philadelphia. He is currently an Assistant Professor with the University of Trento and the Co-Founder and Chief AI Officer of Truebees srl. He is actively involved in fundamental research activities related to the Extended Partnership SERICS. His research interests include AI generated image and video forgery detection and localization and camera source attribution.



MIGUEL MASCIOPINTO received the M.Sc. degree in telecommunications engineering from the University of Vigo, Spain, in 2013.

He was a Researcher with the Signal Theory and Communications Department, University of Vigo, from 2013 to 2024. In May 2024, he joined Galician Research and Development Center in Advanced Telecommunications, Multimodal Information Department, as a Senior Research Engineer in identity and forensics research and development line. He has actively participated as a Research Engineer with European REWIND and NIFTY projects. Additionally, he contributed as a Research and Development Engineer and a Software Developer to the AIDFORK technology transfer project. He was involved in national and regional projects contributing his research expertise to the development of novel technologies for detecting image manipulations in digital onboarding processes and video injection attacks. Currently, he leads the image and video analysis and detection lines in the FAIR project, which focuses on identifying multimedia content generated by artificial intelligence systems. He is the co-author of four papers in international conference proceedings, two articles in international journals, and one book chapter. His research interests include information forensics and security.



FERNANDO PÉREZ-GONZÁLEZ (Fellow, IEEE) received the degree in telecommunication engineering from the University of Santiago, Santiago, Spain, in 1990, and the Ph.D. degree in telecommunications engineering from the University of Vigo, Vigo, Spain, in 1993.

From 2007 to 2010, he was the Program Manager of Spanish National Research and Development Plan on Electronic and Communication Technologies, Ministry of Science and Innovation.

From 2009 to 2011, he was the Prince of Asturias Endowed Chair of Information Science and Technology, The University of New Mexico, Albuquerque, NM, USA. From 2007 to 2014, he was the Executive Director of Galician Research and Development Center in Advanced Telecommunications. He is currently a Professor with the School of Telecommunication Engineering, University of Vigo. He has been a Principal Investigator with the Signal Processing in Communications Group, University of Vigo, which participated in several European projects, including CERTIMARK, ECRYPT, REWIND, NIFTY, WITDOM, UNCOVER, and TRUMPET. He has co-authored more than 70 articles in leading international journals, 190 peer-reviewed conference papers, and several international patents. His research interests include the areas of digital communications, adaptive algorithms, privacy enhancing technologies, and information forensics and security.

Dr. Pérez-González is a fellow of AAIA and AIIA. He was an Associate Editor of IEEE SIGNAL PROCESSING LETTERS, from 2005 to 2009, and IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, from 2006 to 2010; a Senior Area Editor of IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, from 2019 to 2022; and the Editor-in-Chief of *EURASIP Journal on Information Security*, from 2017 to 2022. He serves as an Associate Editor for IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY.



GIULIA BOATO (Member, IEEE) received the degree in mathematics and the Ph.D. degree in information and communication technology from the University of Trento, Trento, Italy, in 2002 and 2005, respectively.

Since 2012, she has been leading the group within the Media Laboratory, focusing on multimedia forensics. She is currently a Full Professor with the University of Trento and the Co-Founder of Truebees. She has published more than 170 papers in international conferences and journals. Her research interests include image and signal processing, with particular attention to multimedia data protection and authentication, data hiding, and digital forensics, but also intelligent multidimensional data management and analysis.

Dr. Boato was a Guest Editor for the Special Issue on Event-Based Media Processing and Analysis of *Image and Vision Computing* (Elsevier) and the Special Issue on Facing Emerging Challenges in Multimedia Forensics of *EURASIP Journal on Information Security*. She served as an Associate Editor for IEEE TRANSACTIONS ON IMAGE PROCESSING, *Journal of Electronic Imaging*, an international journal *Signal Processing: Image Communication* (Elsevier), and *Frontiers in Signal Processing*.

...