New Approaches Based on PRNU-CNN for Image Camera Source Attribution in Forensic Investigations

Giorgio De Magistris[1], Rafał Grycuk[2], Lorenzo Mandelli[1] and Rafał Scherer[2]

[1]Department of Computer, Control and Management Engineering, Sapienza University of Rome
[2]Department of Computational Intelligence, Czestochowa University of Technology

Abstract
Digital image forensics currently mainly uses PRNU noise as a fingerprint to attribute an image to a particular camera. However PRNU is usually extracted manually using Maximum Likelihood estimation from multiple images from the same source device. In this paper we show that the PRNU can be learned in a data driven fashion using a ResNet based neural network. We also show that it is possible to train a neural network for camera attribution directly on the residual noise, that contains both the PRNU and a random component. We show that both approaches are valid as we obtained results comparable with the state of the art.

Keywords
PRNU, CNN, Image Classification, Digital Camera Identification, Forensic, Pattern Noise.

1. Introduction

Camera source identification consists in the attribution of an image to the digital camera from which it was originally captured by using only image features and no external information. Camera source attribution plays a pivotal role in forensic investigations, particularly in the realm of digital imagery and video analysis. The source information holds significant value for several reasons, making it a critical aspect of modern forensic examinations.

One of the primary reasons camera source attribution is crucial is its role in authenticating evidence. In any criminal investigation, the authenticity of evidence is paramount. By determining the camera source, the courts can verify whether an image or video is an original capture or if it has been tampered with or manipulated. This verification process is crucial for establishing the chain of custody and ensuring the evidence presented in court is reliable and admissible.

Camera source attribution also helps in determining the integrity of images. With the advent of sophisticated photo and video editing software, the risk of forged or altered visuals has increased. However, each camera model possesses unique characteristics that act as digital fingerprints.

Furthermore, camera source attribution aids investigators in linking suspects to crime scenes. Surveillance cameras, smartphones, and other digital devices often capture photographs and videos that serve as crucial evidence in criminal investigations. By determining the camera source, investigators can establish a connection between a suspect and a specific crime scene or event. This information becomes vital in establishing a suspect's presence at a particular location and time, strengthening the case against them.

Camera source attribution also aids in tracking cyber criminal activity, particularly in cases involving child exploitation, cyberbullying, or online harassment. By identifying the camera source, law enforcement can trace the origin of illegal or harmful content, leading to the identification and apprehension of offenders. This proactive approach helps protect potential victims and curtail criminal activities.

Moreover, standardized camera source attribution practices facilitate cooperation among law enforcement agencies. Criminal activities often transcend jurisdictional boundaries, and evidence may be collected by different agencies. By following consistent attribution practices, professionals can seamlessly exchange and analyze visual evidence, enhancing the overall effectiveness of criminal investigations.

The rest of the paper is structured as follows: section 2 reviews relevant related works, section 3 describes the dataset, section 4 describes the proposed method, and in particular section 4.1 describes a convolutional networks that is trained to classify the source directly from the residual noise, that contains both the PRNU and a random component, while section 4.2 introduces a different convolutional network that is trained on PRNU isolated. The results are presented in section 5 and conclusions are drawn in section 6.

SYSYEM 2024: 10th Scholar's Yearly Symposium of Technology, Engineering and Mathematics, Rome, December 2-5, 2024
demagistris@diag.uniroma1.it (G. D. Magistris); rafal.grycuk@pcz.pl (R. Grycuk); mandelli@diag.uniroma1.it (L. Mandelli); rafal.scherer@pcz.pl (R. Scherer)
0000-0002-3076-4509 (G. D. Magistris); 0009-0008-8447-0830 (L. Mandelli)

2. Related works

Camera attribution techniques are based on the analysis of the sensor pattern noise (SPN) that is introduced by the acquisition device. This signal however is the sum of two components: a random component that depends on different factors in the image acquisition process and a deterministic component, that depends on intrinsic properties of the image sensor. This second component, called Photo Response Non-Uniformity Noise (PRNU), should be approximately the same in different images acquired by the same device and can be used as a fingerprint of the device itself. The PRNU component of an image can be estimated from multiple images coming from the same device as follows. First the image signal I is separated from the residual noise W using a low-pass filter f:

$$W = I - f(I) \tag{1}$$

Then the deterministic component K is separated from the random component by averaging the residual noise of multiple images [1] or using a more sophisticated minimum variance estimator like in [2, 3]:

$$K = \frac{\sum_{i=1}^{N} W_i I_i}{\sum_{i=1}^{N} I_i^2} \tag{2}$$

where N is the number of images used to estimate the PRNU, I_i is an input image and W_i is the residual noise obtained through high-pass filtering. Once the deterministic component K_c of the camera C is known, the attribution of a new input image I_p is computed thresholding the correlation between the residual noise of the input image $W_p = I_p - f(I_p)$ and the PRNU of the sensor (K_c).

$$\delta(I_p \in C) = \eta(corr(k_c, W_p)) \tag{3}$$

where δ is the Dirac function and η is a thresholding function.

In the last few years deep learning has revolutionized the field of computer vision, and in particular many state of the art approaches in computer vision tasks such as classification, segmentation, etc. are based on convolutional neural networks[4, 5, 6, 7, 8]. End-to-end deep learning approaches have been successfully applied also to image source identification, formulating the attribution as a classification task. Some approaches apply convolutional neural networks directly to raw images [9, 10, 11], however often to this approach is preferred the application of a domain transformation before processing [12, 13, 14, 15]. While the authors of [3] extract the PRNU manually and then use a convolutional network for the classification.

In this paper we present two approaches: the first consists in a simple convolutional network applied directly to the residual noise while the second uses a ResNet based CNN to extract the PRNU from a single image and then the same convolutional network for classification. With both approaches we obtain results comparable with the state of the art.

3. Dataset

We tested the proposed method on the Vision dataset [16], that contains labelled images acquired by common devices. In particular the dataset contains *flat* and *natural* images for each device, where natural images represent common scenes while flat images represent homogeneus backgrounds, without edges, and can be used to extract the PRNU. Some samples from the dataset are shown in figure 1. We used flat images to compute the residual noise and the PRNU target. We cropped all images, keeping only the top left corner of the image with size 256×256.

4. Method

In this section we describe the proposed method, and in particular we describe the two approaches mentioned in section 1: in section 4.1 the input of the network is the residual noise, with both deterministic and random components, while in section 4.2 the input of the classification network is the PRNU extracted by a second neural network. The two approaches share the the same convolutional neural network for classification described in section 4.3.

4.1. Classification from Residual Noise

In this section we describe in detail the process we used to isolate the residual noise from the input signal. We used a method based on wavelet decomposition. This method is based on the assumption that the wavelet coefficients are modeled as iid Gaussian variables with zero mean and variance given by a deterministic, unknown spatially varying variance field. Given the variance field, the image wavelet coefficients without noise are estimated with a Minimum Mean Squared Error procedure. The extraction of the residual noise then consists in the estimation of the variance field and the estimation of the clean coefficients using the variance field. This process can be summarized in the following steps:

Step 1. Calculate the fourth-level wavelet decomposition of the noise image. The following steps will be taken with $h(i, j)$ as an example and the same steps will be taken for other subbands. Denote the vertical, horizontal and diagonal subbands as $v(i, j), h(i, j), d(i, j)$, where (i, j) runs through an index set J that depends on the decomposition level.

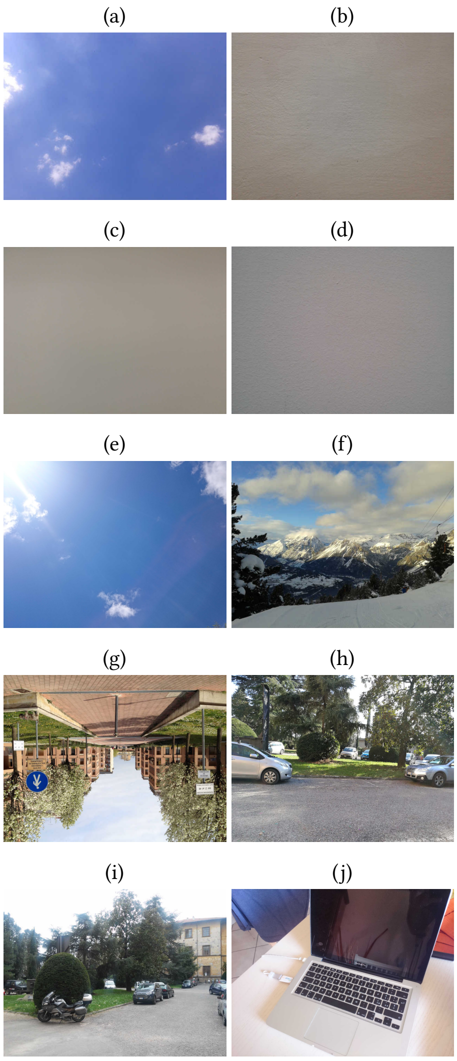

Figure 1: Sample images: 1a, 1b, 1c, 1d, 1e is the flat image from Samsung GalaxyS3Mini, Apple iPhone4s, Huawei P9, LG D290, Sony XperiaZ1Compact and 1f, 1g, 1h, 1i, 1j is the corresponding nat image.

Step 2. Then use MAP estimation to estimate the local covariance of the original noise-free image for each wavelet coefficient for 4 sizes of a square $W \times W$ neighborhood N, for $W \in \{3, 5, 7, 9\}$.

$$\hat{\sigma}_W^2 = max(0, \frac{1}{W^2} \sum_{(i,j)\in N} h^2(i,j) - \sigma_0^2), (i,j) \in J \quad (4)$$

Take the minimum of the 4 variances as the final estimate,

$$\hat{\sigma}^2(i,j) = min\{\sigma_3^2(i,j), \sigma_5^2(i,j), \sigma_7^2(i,j), \sigma_9^2(i,j)\}, (i,j) \in J \quad (5)$$

Step 3. Use Winner filter to denoise the wavelet coefficient.

$$h_{den}(i,j) = h(i,j)\frac{\hat{\sigma}^2(i,j)}{\hat{\sigma}^2(i,j) + \sigma_0^2} \quad (6)$$

and similar for $v(i,j)$ and $d(i,j)$, $(i,j) \in J$

Step 4. Repeat Steps 1–3 for each level and each color channel. The denoised image is obtained by applying the inverse wavelet transform to the denoised wavelet coefficients.

For further details we refer the reader to [17, 18, 15, 19, 20]. The isolated residual noise is the input of the convolutional neural network described in section 4.3, that predicts the source camera.

4.2. Classification from PRNU

The most common method for extracting PRNU is MLE, but MLE requires multiple images from the same device to extract the PRNU noise from a single image. In this paper we propose a novel approach based on deep learning to extract the PRNU component from the residual noise. According to [21], when the original mapping is closer to an identity mapping, the residual mapping is easier to be optimized. Therefore we propose Resnet-based CNN, and in particular the CSI-CNN architecture[22] shown in Figure 2.

In order for the model to learn the specific target PRNU, the input data is residual noise, the target is PRNU noise extracted with MLE and the loss function is mean squared error(MSE). The PRNU extracted from the Resnet-based CNN is then used to train the Convolutional classifier described in the next section. The complete workflow is illustrated in figure 3.

4.3. Convolutional Classifier

As we already mentioned in section 4, the final block of the two classification method is the same Convolutional Network, which architecture is described in this section. As shown in figure 4, the network is composed of two convolutional blocks and two dense layers. Each convolutional block contains two convolutional layers with Relu activation, followed by a Max Pooling layer. The input channel of the first convolutional layer is 1, the kernel size is 3 × 3 and the stride size is 2 × 2 for all convolutional layers. The pooling window size is 2 × 2. The output of the last dense layer is K-dimensional vector encoding the probability distribution of the target cameras. The loss function is the standard cross-entropy function for classification.

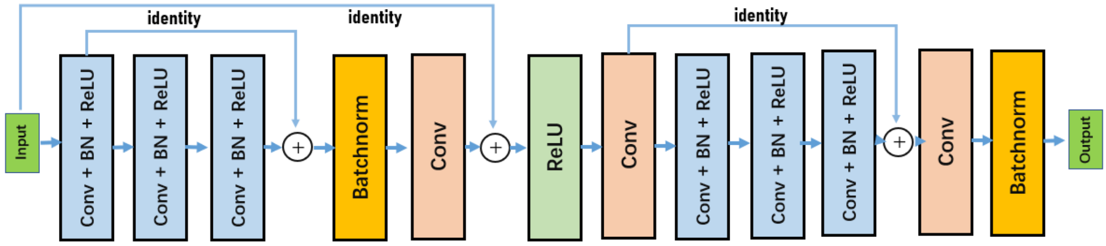

Figure 2: CSI-CNN architechture of the PRNU generate model.

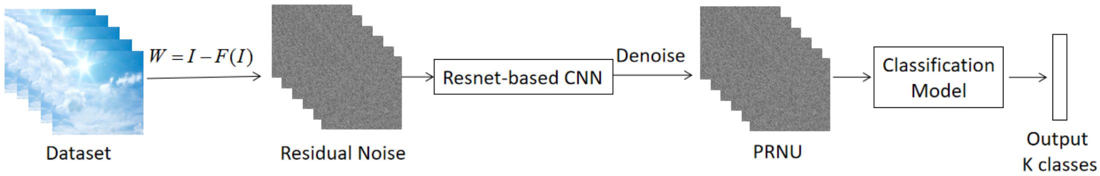

Figure 3: The pipeline we proposed for image source forensics. I is the input image, F the wavelet based high pass filter, W is the residual noise noise. The final result is a probability distribution over the K classes representing the acquisition devices.

5. Results

Our classifier is trained to recognize ten camera classes (K = 10): Samsung GalaxyS3Mini, Apple iPhone4s, Apple iPhone5c, Huawei P9, LG D290, Lenovo P70A, Sony XperiaZ1Compact, Microsoft Lumia640LTE, Wiko Ridge4G, Xiaomi RedmiNote3, two of these devices are from the same brand. In total, there are 2194 images. As described in the previous section, we tested two configurations: in the first the input of the classifier is the residual noise while in the second the ResNet based neural network is trained to predict the PRNU from raw images and then the obtained PRNU is used to train the classifier. Figure 5 shows the learning curve of the ResNet based neural network. Figure 6 shows the confusion matrices for both configurations. With both approaches we reached state of the art results, as shown in table 1. I

6. Conclusion

In this paper we presented two novel deep learning approaches for camera attribution from raw images, obtaining results comparable with the state of the art. Moreover we showed that the PRNU, that is the fingerprint of the image sensor, can be isolated from the input image using a data driven approach. We also showed that a neural network can be trained to classify the target source camera directly from the residual noise, that contains both the PRNU and a random component, obtaining even better results.

While significant progress has been made in addressing image source attribution, there are still areas that warrant further exploration and development. A potential avenue for future research and improvements of this work could be the application of similar techniques to

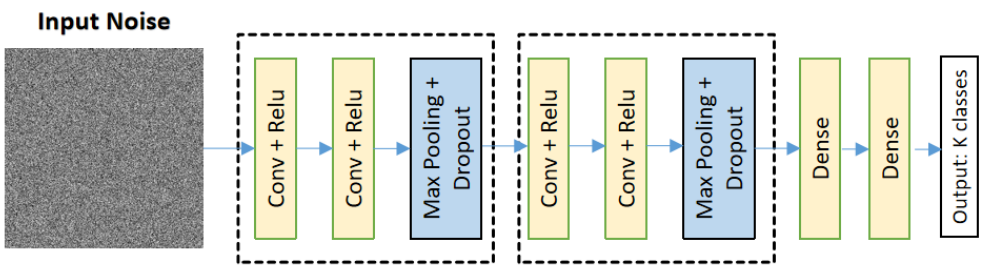

Figure 4: Structural details of the classification model.

Table 1
Result Comparison

Model	Dataset	Target	Labels	Level	Accuracy
Classification from learned PRNU (ours)	VISION	Camara	10	Patch	**90.79%**
Classification From Residual Noise (ours)	VISION	Camera	10	Patch	**92.41%**
Shallow CNN[23]	VISION	Camera	35	Patch	80.77%
DenseNet-40[23]	VISION	Camera	35	Patch	87.96%
Roberto, C et al[3]	VISION	Social Network	3	Patch	79.48%
Roberto, C et al[3]	VISION	Social Network	3	Image	89.83%
Bondi, L et al[24]	Dresden	Camera	18	Patch	72.90%

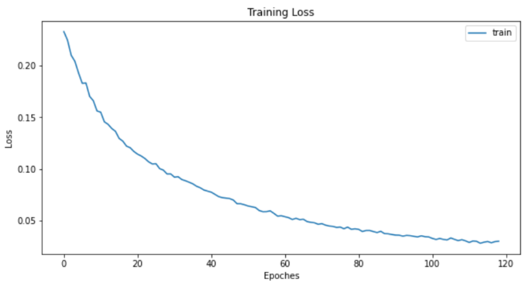

Figure 5: Learning curve for the PRNU generation.

compressed images. This is particularly important because the popular social media platforms compress the uploaded content, hindering the source attribution. Another important line of research could be the application of these models to AI generated content. The advent of AI-generated images indeed, driven by advancements in machine learning and deep learning algorithms, has significant implications across various domains. While these technologies offer exciting possibilities and creative opportunities, they also raise important ethical, legal, and societal concerns.

Figure 6: On the left the confusion matrix for the classification from residual noise, on the right the confusion matrix for the classification from the learned PRNU.

7. Acknowledgment

This work was partially supported by the Hermes-WIRED project within the Large Research Projects grant framework 2020 funded by Sapienza University of Rome.

References

[1] J. Lukas, J. Fridrich, M. Goljan, Digital camera identification from sensor pattern noise, IEEE Transactions on Information Forensics and Security 1 (2006) 205–214.

[2] M. Chen, J. Fridrich, M. Goljan, J. Lukas, Determining image origin and integrity using sensor noise, IEEE Transactions on Information Forensics and Security 3 (2008) 74–90. doi:10.1109/TIFS.2007.916285.

[3] R. Caldelli, I. Amerini, C. T. Li, Prnu-based image classification of origin social network with cnn, in: 2018 26th European Signal Processing Conference (EUSIPCO), IEEE, 2018, pp. 1357–1361.

[4] G. De Magistris, R. Caprari, G. Castro, S. Russo, L. Iocchi, D. Nardi, C. Napoli, Vision-based holistic scene understanding for context-aware human-robot interaction 13196 LNAI (2022) 310–

325. doi:10.1007/978-3-031-08421-8_21.

[5] R. Brociek, G. De Magistris, F. Cardia, F. Coppa, S. Russo, Contagion prevention of covid-19 by means of touch detection for retail stores, in: 2021 Scholar's Yearly Symposium of Technology, Engineering and Mathematics, SYSTEM 2021, 2021, pp. 89–94.

[6] C. Napoli, C. Napoli, V. Ponzi, A. Puglisi, S. Russo, I. E. Tibermacine, Exploiting robots as healthcare resources for epidemics management and support caregivers, in: CEUR Workshop Proceedings, volume 3686, 2024, p. 1 – 10.

[7] E. Iacobelli, S. Russo, C. Napoli, A machine learning based real-time application for engagement detection, in: CEUR Workshop Proceedings, volume 3695, 2023, p. 75 – 84.

[8] N. Brandizzi, A. Fanti, R. Gallotta, S. Russo, L. Iocchi, D. Nardi, C. Napoli, Unsupervised pose estimation by means of an innovative vision transformer, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), volume 13589 LNAI, 2023, p. 3 – 20. doi:10.1007/978-3-031-23480-4_1.

[9] Y. Chen, Y. Huang, X. Ding, Camera model identification with residual neural network, in: 2017 IEEE International Conference on Image Processing (ICIP), IEEE, 2017, pp. 4337–4341.

[10] V. Ponzi, S. Russo, A. Wajda, C. Napoli, A comparative study of machine learning approaches for autism detection in children from imaging data, in: CEUR Workshop Proceedings, volume 3398, 2022, p. 9 – 15.

[11] X. Ding, Y. Chen, Z. Tang, Y. Huang, Camera identification based on domain knowledge-driven deep multi-task learning, IEEE Access 7 (2019) 25878–25890.

[12] S. Falciglia, F. Betello, S. Russo, C. Napoli, Learning visual stimulus-evoked eeg manifold for neural image classification, Neurocomputing 588 (2024). doi:10.1016/j.neucom.2024.127654.

[13] S. Russo, I. E. Tibermacine, A. Tibermacine, D. Chebana, A. Nahili, J. Starczewscki, C. Napoli, Analyzing eeg patterns in young adults exposed to different acrophobia levels: a vr study, Frontiers in Human Neuroscience 18 (2024). doi:10.3389/fnhum.2024.1348154.

[14] C. Napoli, F. Bonanno, G. Capizzi, An hybrid neuro-wavelet approach for long-term prediction of solar wind, in: Proceedings of the International Astronomical Union, volume 6, 2010, p. 153 – 155. doi:10.1017/S174392131100679X.

[15] M. Wozniak, C. Napoli, E. Tramontana, G. Capizzi, G. Lo Sciuto, R. Nowicki, J. Starczewski, A multiscale image compressor with rbfnn and discrete wavelet decomposition, in: Proceedings of the International Joint Conference on Neural Networks, volume 2015-September, Institute of Electrical and Electronics Engineers Inc., 2015. doi:10.1109/IJCNN.2015.7280461.

[16] D. Shullani, M. Fontani, M. Iuliani, O. A. Shaya, A. Piva, Vision: a video and image dataset for source identification, EURASIP Journal on Information Security 2017 (2017) 1–16.

[17] M. K. Mihcak, I. Kozintsev, K. Ramchandran, Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising, in: 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No. 99CH36258), volume 6, IEEE, 1999, pp. 3253–3256.

[18] C. Napoli, G. Pappalardo, E. Tramontana, Improving files availability for bittorrent using a diffusion model, in: Proceedings of the Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises, WETICE, 2014, p. 191 – 196. doi:10.1109/WETICE.2014.65.

[19] G. Lo Sciuto, G. Capizzi, R. Shikler, C. Napoli, Organic solar cells defects classification by using a new feature extraction algorithm and an ebnn with an innovative pruning algorithm, International Journal of Intelligent Systems 36 (2021) 2443 – 2464. doi:10.1002/int.22386.

[20] C. Napoli, G. Pappalardo, E. Tramontana, An agent-driven semantical identifier using radial basis neural networks and reinforcement learning, in: CEUR Workshop Proceedings, volume 1260, 2014. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84919742629&partnerID=40&md5=c3ee8a3fa1716b39215326edfc67d955.

[21] K. Zhang, W. Zuo, Y. Chen, D. Meng, L. Zhang, Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising, IEEE transactions on image processing 26 (2017) 3142–3155.

[22] Y. Wang, Q. Sun, D. Rong, S. Li, L. D. Xu, Image source identification using convolutional neural networks in iot environment, Wireless Communications and Mobile Computing 2021 (2021).

[23] F. Marra, D. Gragnaniello, L. Verdoliva, On the vulnerability of deep learning to adversarial attacks for camera model identification, Signal Processing: Image Communication 65 (2018) 240–248.

[24] L. Bondi, L. Baroffio, D. Güera, P. Bestagini, E. J. Delp, S. Tubaro, First steps toward camera model identification with convolutional neural networks, IEEE Signal Processing Letters 24 (2016) 259–263.