IMAGE ANONYMIZATION DETECTION WITH DEEP HANDCRAFTED FEATURES

Nicolò Bonettini[1], *David Güera*[2], *Luca Bondi*[1], *Paolo Bestagini*[1], *Edward J. Delp*[2], *Stefano Tubaro*[1]

[1]Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano - Milano, Italy
[2]School of Electrical and Computer Engineering, Purdue University - West Lafayette, IN, USA

ABSTRACT

In recent years, the number of images shared online has continuously grown. The forensics community has kept the pace by developing techniques to both reliably extract information from these images, but also to remove it. In particular, the latest developments in image anonymization methods exposes an attack vector when used by skilled ill-intentioned image producers that may want to elude prosecution. We present an approach to detect whether or not an image has undergone a laundering process, i.e., it has been tampered with so that its unique characterizing features have been changed to avoid detection. We focus on photo response non uniformity (PRNU) noise unique to every imaging sensor, we consider that an image has been "laundered" when we detect the absence of PRNU from an image. We propose a per image preprocessing pipeline that generates information-rich features later used as input of fine-tuned convolutional neural networks (CNNs). We study the performance of the proposed approach using various CNN architectures and blind anonymization techniques, and show its effectiveness under several training and testing scenarios. Our results also show that CNN models trained with the proposed feature are capable of generalizing over unseen devices and are robust against non-geometric transformations.

***Index Terms*—** Image forensics, PRNU, camera attribution, deep learning

1. INTRODUCTION

In the last decade, the massive adoption of the smartphone and the global popularization of social networking websites has led to unprecedented rates of social digital media sharing. In 2010 alone, Facebook reported storing more than 260 billion images, with users uploading one billion new images each week [1]. In 2016, Instagram had more than 400 million active monthly users who shared over 40 billion images, with an average of 3.5 billion daily "likes" for more than 80 million images shared daily on the site [2].

The forensic community [3] has developed a series of multimedia source attribution techniques in the pixel domain that have enabled one to detect which device has been used to acquire an image with precise results [4, 5]. These tools do not using any exchangeable image file format (Exif) information which can easily be modified or removed.

This set of tools have been successfully used to establish forensic evidence in major international legal proceedings and have recently passed the United States Judiciary *Daubert* standard [6]. The *Daubert* standard is a rule of evidence regarding the admissibility of expert witnesses' testimony in United States Federal Courts and has been thoroughly vetted by the FBI Crime Laboratory and the US National Institute of Justice (NIJ) [7].

The image and video source attribution problem consists in detecting which device has been used to acquire a specific image or video, thus tracing back the digital media to its owner [8, 9]. The most promising approaches exploit the photo response non uniformity (PRNU) noise [5, 10]. This is a multiplicative noise pattern characterizing each imaging sensor, which is inevitably injected into every acquired image or video. By estimating a noise fingerprint from an image, it is possible to compare it with the PRNUs of known camera devices, thus determining which device "took the picture". PRNU-based approaches can also be used on scaled and cropped images or videos [11].

Forensic techniques that focus on removing traces of PRNU from images have been well studied in the literature. We can distinguish two major approaches: the first group requires knowledge of the PRNU pattern to be deleted, whereas the other major group does not require access to the real PRNU to remove it from a given image.

Sensor fingerprint removal based on knowledge of the underlying PRNU was first suggested by Lukáš et al. in [8]. This approach assumes a known PRNU fingerprint estimate of a particular imaging sensor is latent in any given image intensity acquired by the same sensor. Hence, the removal of the PRNU fingerprint (the anonymization of the image source) can be achieved by subtraction of the fingerprint estimate from the image intensity.

More recently, the approach presented by Bonettini et al. in [12] explores the possibilities offered by CNNs in terms of camera device anonymization based on the knowledge of the reference PRNU. An image-wise anonymization step is part of a CNN-based noise extractor. An autoencoder fully-convolutional neural network is trained as an anonymization function via back-propagation, exploiting the possibilities offered by a CNN-based denoising method introduced by Zhang et al. [13].

Other image anonymization methods work by blindly modifying pixel values and scrambling their positions in order to make the underlying PRNU unrecognizable. Dirik et al. [14] propose to anonymize images by applying seam-carving to change pixel locations and more recently Entrieri and Kirchner [15] compare patch-based methods to shuffle small image blocks. Mandelli et al. [16] investigate parallel and fast inpainting techniques as methods for image anonymization.

As discussed in the Malicious AI report [17], one should always reflect on the dual-use nature of their work, allowing misuse to influence research priorities and norms.

This material is based on research sponsored by DARPA and Air Force Research Laboratory (AFRL) under agreement number FA8750-16-2-0173. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA and Air Force Research Laboratory (AFRL) or the U.S. Government.

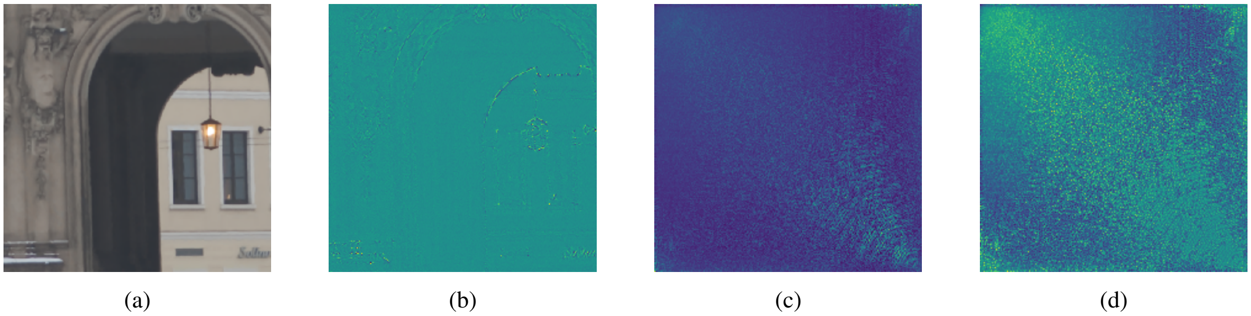


Fig. 1: (a) RGB image anonymized using [15]. (b) Wavelet noise extracted from the image. (c) DFT_2 of the signal. (d) Final feature after Wiener filtering.

It is not far-fetched to imagine malevolent agents using the above forensic tools to anonymize images showing sensitive or illegal content. As an example, an illegal content producer could anonymize its images to avoid being linked to them in case that its image acquisition equipment were to be seized while in his possession. Hence, it is of paramount importance to know when an image has undergone any process to attempt to remove its underlying PRNU.

For this reason, in this paper we focus on the problem of image anonymization detection. This is, given an image, understand whether someone edited it in order to hide PRNU traces to avoid device identification. The method we propose is based on the use of a CNN. However, rather then feeding the network with images in the pixel domain, we show that applying the proposed pre-processing technique greatly improves anonymization detection performance. The proposed solution is able to generalize to different kinds of anonymization methods never seen during training, thus showing that using specific domain knowledge can help pure data-driven techniques.

2. BACKGROUND AND PROBLEM STATEMENT

Photo response non uniformity (PRNU) is a characterizing feature of the sensor present in digital image acquisition devices that is manifested as a noise term in acquired images [10]. It is due to device-level anomalies in the semiconductors used to manufacture the imaging sensor of digital cameras. Due to the physical origin of these anomalies, PRNU is a unique feature of each individual camera. It also makes it impractical for the image acquisition pipeline in cameras to effectively compensate for PRNU so artifacts are present in the digital images that are produced with the device. PRNU is also robust to lossy compression, which makes it suitable as a robust feature for camera identification [10].

Being such a powerful forensic footprint, PRNU has been greatly investigated in the literature under various scenarios. In the following, we report the formal definition of typical PRNU-based forensic problems in order to highlight the goal of this paper.

Device attribution. Given an image $\mathbf{I}$ and a generic denoising function $D(\cdot)$, we can compute the noise residual $\mathbf{W} = \mathbf{I} - D(\mathbf{I})$. From a series of noise residuals obtained from images shot with the same device, it is possible to estimate the PRNU fingerprint $\mathbf{K}$ by applying a weighted average operation [10]. Given the PRNU fingerprint $\mathbf{K}$ of a camera, we can bind an image $\mathbf{I}$ to that camera if:

$$\mathrm{NCC}(\mathbf{W}, \mathbf{K} \otimes \mathbf{I}) > \tau, \tag{1}$$

where NCC is the normalized cross-correlation function, $\otimes$ denotes the Hadamard (element-wise) product and τ is a threshold set in order to bound false-detection probability below a confidence value α.

Blind device anonymization. Given an image $\mathbf{I}$, a blind anonymization function $A(\cdot)$ is a function that generates an anonymized version of $\mathbf{I}$, namely $\hat{\mathbf{I}} = A(\mathbf{I})$. The anonymization process ensures that:

$$\mathrm{NCC}(\hat{\mathbf{W}}, \mathbf{K} \otimes \hat{\mathbf{I}}) < \tau, \tag{2}$$

where $\hat{\mathbf{W}} = \mathbf{I} - D(\hat{\mathbf{I}})$. We call it blind since it is not required any prior knowledge on the reference PRNU $\mathbf{K}_I$ for computing $\hat{\mathbf{I}}$.

Anonymization detection. We define anonymization detection problem as a two-class classification problem, where $\mathcal{C}_0$ is the class of original images and $\mathcal{C}_1$ is the class of the anonymized images. Given an image $\mathbf{I}$ under investigation, an anonymization detector is an operator $M(\cdot)$ such that:

$$\hat{y} = M(\mathbf{I}), \tag{3}$$

where $\hat{y} \in \{0, 1\}$ represents a label that assigns $\mathbf{I}$ to $\mathcal{C}_0$ or $\mathcal{C}_1$. The goal of this paper is to design an operator $M(\cdot)$.

3. PRNU REMOVAL DETECTION

Given an image under analysis $\mathbf{I}$, our goal is to detect whether its PRNU traces have been removed or not. We propose the following method to determine this: (i) we pre-process the image in order to extract a feature that exposes salient anonymization information; (ii) we add the proposed feature to a CNN that identifies whether the analyzed image has been anonymized or not.

Feature extraction. Despite the well-known capabilities of CNNs to work directly in the pixel domain, they can yield better performance when coupled with domain specific knowledge of the problem to be tackled. Using domain knowledge as indicated in [18, 19], we leverage the efforts of the forensics community and propose a preprocessing approach to extract the residual noise left in the input image and then shift it into the Fourier domain. Due to the noisy nature of the PRNU pattern and the subtlety of the traces left by the anonymization techniques, if pixel domain information were to be input into the neural network, this would lead to poorly trained models incapable of generalization or overfitted on image features instead of PRNU removal detection. However, our proposed preprocessing leads to a boost in the detection accuracy of our trained neural network model while generalizing to unseen camera models, as shown in the Results section.

Our feature extraction method is defined as follows. Let us consider a grayscale $H \times W$ image $\mathbf{I}$. By means of the Wavelet denoising function $D_{\mathrm{w}}(\cdot)$ proposed in [20] and often used for PRNU estimation, we compute the noise residual $\mathbf{W}$ from the image as:

$$\mathbf{W} = \mathbf{I} - D_{\mathrm{w}}(\mathbf{I}). \tag{4}$$

We then compute the magnitude of the 2D Discrete Fourier Transform (DFT_2) as:

$$\mathbf{W}_{\mathrm{F}} = |\mathrm{DFT}_2(\mathbf{W})|, \tag{5}$$

and we then Wiener filter this signal, following the method described in [8]:

$$[\mathbf{\Phi}]_{ij} = [\mathbf{W}_{\mathrm{F}}]_{ij} \cdot \frac{\sigma_s^2}{[\mathbf{S}_{\mathrm{W}}]_{ij} + \sigma_s^2}, \tag{6}$$

with

$$\sigma_s^2 = \delta \cdot \sigma_{\mathrm{W}}^2, \tag{7}$$

where $i = 1, \dots, H$, $j = 1, \dots, W$, σ_{W}^2 is the variance of $\mathbf{W}_{\mathrm{F}}$ and $\mathbf{S}_{\mathrm{W}}$ is the matrix containing the variance of the energy of $\mathbf{W}_{\mathrm{F}}$ computed over a 3×3 moving window. The parameter δ must be chosen depending on the input signals in order to drive the Wiener filtering operation (we set $\delta = 0.77$ in our experiments). Without loss of generality, we iterate this procedure over each of the three channels of a RGB image. The result is a $H \times W \times 3$ feature $\mathbf{\Phi} = \Phi(\mathbf{I})$ we use as network input. Fig. 1 visually displays all the feature extraction steps.

Model. Due to the formulation of our problem, we can make use of transfer learning, which is known to increase the performance of detection models in limited data settings [21]. This means that we can use CNN models pretrained on large image datasets and leverage their learnt filters. After selecting a suitable architecture, we use transfer learning on a model trained on ImageNet by replacing its last fully-connected layer with a $(n_l, 1)$ fully-connected layer, where n_l is the number of input features of the original last layer. Additionally, we perform a sigmoid over the network output to bind it to $[0, 1]$. Thus, giving as input a batch of B samples $\mathbf{X}$ of size $B \times H \times W \times 3$ and its label tensor $\mathbf{y}$ of size $B \times 1 \in \{0, 1\}$, the network output $\hat{\mathbf{y}}$ of size $B \times 1 \in [0, 1]$ is a tensor of scalars representing the likelihood of each sample to be anonymized. We choose Binary Cross Entropy as a loss function between the target and the output:

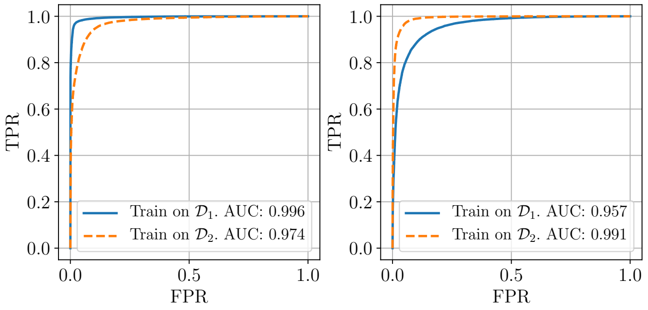
$$\mathscr{L}_b = -\left[y_b \cdot \log \hat{y}_b + (1 - y_b) \cdot \log\left(1 - \hat{y}_b\right)\right], \tag{8}$$

where $b \in \{1, ..., B\}$ is the sample in-batch index. We train each model using Adam optimizer [22] until reaching a validation plateau.

In our experiment, we consider ResNet [23] as CNN model. AlexNet [24] and VGG [25] were also considered, but underperformed ResNet, probably due to the consistently higher number of parameters.

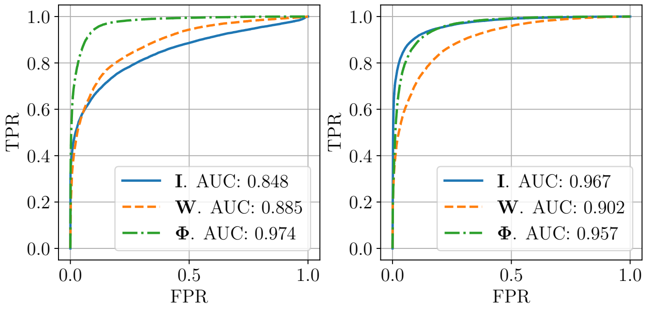
4. RESULTS

In this section we describe the dataset, the training strategy, and the different kinds of analysis we performed to highlight different aspects of the proposed solution.



(a) Test on $\mathcal{T}_1$ (b) Test on $\mathcal{T}_2$

Fig. 2: ROC curves for two different testing dataset.



(a) Train on $\mathcal{D}_2$, test on $\mathcal{T}_1$ (b) Train on $\mathcal{D}_1$, test on $\mathcal{T}_2$

Fig. 3: ROC curves with various preprocessing approaches.

Dataset. Starting from the 600 original images from the Dresden database [26] used in [16], we applied to each image the two anonymization procedures described in [16] and [27], thus obtaining a corpus of $600 \times 3 = 1800$ images (i.e., the original images, and the two sets of anonymzed ones). To the best of our knowledge, these are the most recent methods for blind device anonymization known in literature.

In order to validate the proposed approach on various kinds of anonymized images, as well as to test cross-dataset generalization, starting from the above images we define different sets of data. Specifically, we define three training dataset:

- $\mathcal{D}_1$ composed by 300 original images and 300 images anonymized following [16],
- $\mathcal{D}_2$ composed by the same 300 original images and 300 images anonymized following [27],
- $\mathcal{D}_3$ composed by the same 300 original images, the first 150 images anonymized by [16] and the last 150 images anonymized by [27].

Following the same strategy, we construct two testing sets with the remaining images:

- $\mathcal{T}_1$ composed by 300 original images and 300 images anonymized following [16],
- $\mathcal{T}_2$ composed by the same 300 original images and 300 images anonymized following [27].

All the images are RGB images and have been central-cropped to size 512×512 pixels. To increase the number of available samples, we split each image in 224×224 blocks with an overlapping stride of 32×32. During training, we use 70% of each training set for training, and the remaining 30% for validation. In total, we have 42000 samples in each training set, 18000 in each validation set and 60000 in each test set.

Cross dataset results. During the test phase we examined the models trained on $\mathcal{D}_1$, $\mathcal{D}_2$, $\mathcal{D}_3$ and we test them against $\mathcal{T}_1$ and $\mathcal{T}_2$. We perform our test in very challenging scenarios, including training on a specific anonymization method and testing on the other

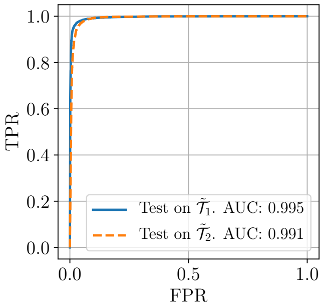


Fig. 4: ROC for Leave One Out strategy. Train on $\tilde{\mathcal{D}}_3$ test on $\tilde{\mathcal{T}}_1$and $\tilde{\mathcal{T}}_2$.

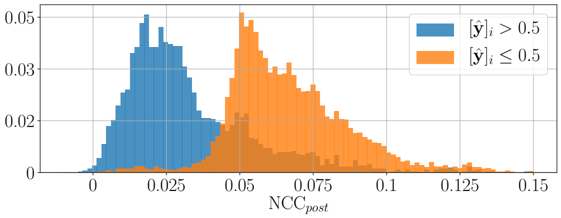


Fig. 5: Distribution of NCC_{post} values.

one. The test procedure consists in freezing the network weights and predicting $\hat{y}$ given as input the feature $\mathbf{\Phi}$. By thresholding $\hat{y}$ with different thresholds, we compute true positive rate (TPR) and false positive rate (FPR) w.r.t. the true label y and plot them as a Receiver Operating Characteristic (ROC) curve. We show in Fig. 2 the ROC curves for the most challenging training and testing conditions. The Area Under Curve (AUC) on same-type dataset are very high and we are able to reach comparable AUC even in the worst case scenario in which we train on $\mathcal{D}_1$ and test on $\mathcal{T}_2$ and viceversa. This shows that the proposed method generalizes over different anonymization methods, and does not simply overfit to recognize one.

Fig. 3 shows the network performances with different input preprocessing (image $\mathbf{I}$ or the noise residual $\mathbf{W}$ or the feature $\mathbf{\Phi}$). The designed feature $\mathbf{\Phi}$ is the one that shows the best results regardless the arduous testing scenarios.

Leave One Out. One might be concerned the proposed CNN model learns to recognize just the PRNU of the devices used during training. To verify the model performance with unseen camera devices, we design a Leave One Out testing procedure. We modify our training dataset $\mathcal{D}_3$ by removing all the images acquired with device `Nikon_D200_0`, creating $\tilde{\mathcal{D}}_3$. Similarly, we remove all the images from all the devices except `Nikon_D200_0` from $\mathcal{T}_1$ and $\mathcal{T}_2$ and we add to them the images removed from $\mathcal{D}_3$. These two new mono-device datasets are known as $\tilde{\mathcal{T}}_1$ and $\tilde{\mathcal{T}}_2$. Fig. 4 shows the ROC for training on $\tilde{\mathcal{D}}_3$ and testing on $\tilde{\mathcal{T}}_1$ and $\tilde{\mathcal{T}}_2$. We can assess that our model is capable of discriminating between original and anonymized images even when it is tested against a device which was not present in the training set.

Robustness to transformations. As it is clear from the literature [28], PRNU can be somehow corrupted by other editing operations. We are therefore interested in studying whether our method recognizes these situations or not, and we design a proper testing strategy.

Table 1: Transformation table with their parameters set. Parameter for JPEG is the quality factor, for Gamma is the exponent, for Brightness ranges from 0 (black image) to 1 (original image), for Contrast ranges from 0 (solid gray image) to 1 (original image).

Transformation	**Parameters set**
JPEG compression	$70, 75, 80, 85, 90$
Gamma correction	$0.5, 0.6, 0.7, 0.8, 0.9$
Brightness correction	$0.5, 0.6, 0.7, 0.8, 0.9$
Contrast correction	$0.5, 0.6, 0.7, 0.8, 0.9$

In the test phase, before extracting the feature matrix $\mathbf{\Phi}$ from the image $\mathbf{I}$, we compute:

$$\mathrm{NCC}_{pre} = \mathrm{NCC}(\mathbf{W}, \mathbf{IK}), \tag{9}$$

where $\mathbf{W}$ is the noise residual obtained from $\mathbf{I}$ and $\mathbf{K}$ is the reference PRNU. NCC_{pre} gives us a baseline metric of correlation between an image and its noise fingerprint. Then we modify the image with one of the available transformations in Table 1, randomly selecting it and a parameter from its set, and we compute:

$$\mathrm{NCC}_{post} = \mathrm{NCC}(\mathbf{W}_t, \mathbf{I}_t\mathbf{K}), \tag{10}$$

where $\mathbf{I}_t$ and $\mathbf{W}_t$ denote the transformed image and the noise residual obtained from it, respectively. NCC_{post} gives us a measure of the degradation of the fingerprint introduced by the transformation. It is worth noting that all the transformations in Table 1 are non-geometric transformations, hence we do not need to transform the reference PRNU $\mathbf{K}$ too. After computing these two metrics, we compute the network output $\hat{\mathbf{y}}$ from $\mathbf{\Phi}$. Fig. 5 shows the distribution of NCC_{post} values. For the sake of visualization, we show only the samples with $\mathrm{NCC}_{post} > 0.05$. We can clearly distinguish two value distributions, the left one with $[\hat{\mathbf{y}}]_i > 0.5$ and the right one with $[\hat{\mathbf{y}}]_i \leq 0.5$. This shows that our method is robust to transformations: when the NCC value is low the network classifies the sample as anonymized, whereas when the NCC value is high, the network classifies the sample as original. More in general this result highlight the fact that we are not simply learning to discriminate the artifacts left by the two considered anonymization methods.

5. CONCLUSIONS

In this paper, we considered the problem of detecting image anonymization by means of PRNU removal. We propose a new feature starting from RGB images, and use this feature as input to a convolutional neural network. We select two different blind anonymization techniques and perform network finetuning on images processed with these techniques. Results show the effectiveness of our method comparing to several different preprocessing pipelines. Our model is capable of generalizing over unseen devices and it is robust against non-geometric transformations.

Despite the undoubted capability of convolutional neural networks, pure data-driven approach was not sufficient for solving the problem well enough. Forensics domain knowledge allowed us to carefully design a preprocessing pipeline for feature extraction to ease network training, thus showing that model-based and data-driven methods can benefit from one another.

6. REFERENCES

[1] D. Beaver, S. Kumar, H. C. Li, J. Sobel, P. Vajgel, et al., "Finding a needle in haystack: Facebook's photo storage," *Proceedings of the USENIX Conference on Operating Systems Design and Implementation*, vol. 10, no. 10, pp. 1–8, Oct. 2010, Vancouver, BC, Canada.

[2] S. Alhabash and M. Ma, "A tale of four platforms: Motivations and uses of facebook, twitter, instagram, and snapchat among college students?," *Social Media + Society*, vol. 3, no. 1, Feb. 2017.

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

[4] S. Bayram, H. Sencar, N. Memon, and I. Avcibas, "Source camera identification based on CFA interpolation," *Proceedings of the IEEE International Conference on Image Processing*, pp. 69–72, September 2005, Genova, Italy.

[5] M. Chen, J. Fridrich, M. Goljan, and J. Lukas, "Source digital camcorder identification using sensor photo-response nonuniformity," *Proceedings of the SPIE Electronic Imaging - Security, Steganography, and Watermarking of Multimedia Contents*, Mar. 2007, San Jose, CA.

[6] M. Goljan, M. Chen, P. Comesaña, and J. Fridrich, "Effect of compression on sensor-fingerprint based camera identification," *Proceedings of the IS&T Electronic Imaging*, vol. 2016, no. 8, pp. 1–10, Jan. 2016, Burlingame, CA.

[7] P. C. Giannelli, "Daubert and forensic science: The pitfalls of law enforcement control of scientific research," *University of Illinois Law Review*, p. 53, 2011.

[8] J. Lukáš, J. Fridrich, and M. Goljan, "Digital camera identification from sensor pattern noise," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 205–214, June 2006.

[9] M. Kirchner and T. Gloe, "Forensic camera model identification," in *Handbook of Digital Forensics of Multimedia Data and Devices*, pp. 329–374. John Wiley & Sons, Ltd, Chichester, UK, 2015.

[10] M. Chen, J. Fridrich, M. Goljan, and J. Lukáš, "Determining image origin and integrity using sensor noise," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 1, pp. 74–90, Mar. 2008.

[11] M. Goljan and J. Fridrich, "Camera identification from cropped and scaled images," *Proceedings of SPIE Security, Forensics, Steganography, and Watermarking of Multimedia Contents*, vol. 6819, pp. 6819 – 6819 – 13, March 2008.

[12] N. Bonettini, L. Bondi, D. Güera, S. Mandelli, P. Bestagini, S. Tubaro, and E. J. Delp, "Fooling prnu-based detectors through convolutional neural networks," *Proceedings of the IEEE European Signal Processing Conference*, Sept. 2018, Rome, Italy.

[13] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, July 2017.

[14] A. E. Dirik, H. T. Sencar, and N. Memon, "Analysis of seam-carving-based anonymization of images against prnu noise pattern-based source attribution," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2277–2290, Dec. 2014.

[15] J. Entrieri and M. Kirchner, "Patch-based desynchronization of digital camera sensor fingerprints," *Proceedings of the IS&T Electronic Imaging*, vol. 2016, no. 8, pp. 1–9, Jan. 2016, Burlingame, CA.

[16] S. Mandelli, L. Bondi, S. Lameri, V. Lipari, P. Bestagini, and S. Tubaro, "Inpainting-based camera anonymization," *Proccedings of the IEEE International Conference on Image Processing*, pp. 1522–1526, Sept. 2017, Beijing, China.

[17] M. Brundage et al., "The malicious use of artificial intelligence: Forecasting, prevention, and mitigation," *arXiv:1802.07228*, Feb. 2018.

[18] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, Aug 2013.

[19] T. Teng, A. Tan, and J. M. Zurada, "Self-organizing neural networks integrating domain knowledge and reinforcement learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 5, pp. 889–902, May 2015.

[20] M. K. Mihcak, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Processing Letters (SPL)*, vol. 6, pp. 300–303, 1999.

[21] K. Weiss, T. M. Khoshgoftaar, and D Wang, "A survey of transfer learning," *Journal of Big Data*, vol. 3, no. 1, pp. 9, May 2016.

[22] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *Proceedings of the International Conference on Learning Representations*, vol. arXiv, no. 1412.6980, May 2015, San Diego, CA.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, June 2016, Las Vegas, NV.

[24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, pp. 1097–1105, Dec. 2012, Lake Tahoe, NV.

[25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Proceedings of the International Conference on Learning Representations*, vol. arXiv, no. 1409.1556, May 2015, San Diego, CA.

[26] T. Gloe and R. Böhme, "The Dresden image database for benchmarking digital image forensics," *Journal of Digital Forensic Practice*, vol. 3, pp. 150–159, Dec. 2010.

[27] M. Kirchner, "Fast and reliable resampling detection by spectral analysis of fixed linear predictor residue," *Proceedings of the ACM Workshop on Multimedia and Security*, Sept. 2008, Oxford, UK.

[28] K. Rosenfeld and H. T Sencar, "A study of the robustness of PRNU-based camera identification," *Proceedings of the IS&T/SPIE Electronic Imaging*, Jan. 2009, San Jose, CA.