

RESEARCH ARTICLE

Open-Set Source Camera Device Identification of Digital Images Using Deep Learning

MANISHA¹, CHANG-TSUN LI², (Senior Member, IEEE),
AND KARUNAKAR A. KOTEGAR¹, (Senior Member, IEEE)

¹Department of Data Science and Computer Applications, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104, India

²School of Information Technology, Deakin University, Geelong, VIC 3216, Australia

Corresponding author: Chang-Tsun Li (changtsun.li@deakin.edu.au)

ABSTRACT Source camera identification plays an important role in forensics investigations on images. It is a forensic problem of linking an image in question to the camera used to capture it. Several source identification techniques have been developed in the literature since this may be a facilitating tool that help to trace back the images to the camera device held by the accused in various forensic applications. However, one of the key disadvantages is that the existing techniques fail if the image in question was taken by a new camera that is not used in the training process. Under a real-world forensic scenario, it is not possible to presume that each image being analyzed comes from one of the cameras used to train the source identification system. To address this issue, we propose a data-driven system based on convolutional neural network to identify the source camera device in an open-set scenario. The experimental results on various sets of cameras show that it is possible to leverage the data-driven model as the feature extractor paired with an open-set classifier to trace back the images to the open-set cameras. The results show that the proposed system outperforms the state-of-the-art techniques in identifying the exact device that are never seen before with considerably high accuracy and is resilient to unknown post-processing applied by the social network platforms. Moreover, the experimental results demonstrate the good generalization capability of the proposed system in extracting the source information, making it more suitable for open-set scenarios.

INDEX TERMS Deep learning, image forensics, open-set problem, social network, source camera identification.

I. BACKGROUND

Source camera identification (SCI) is a forensic issue that involves tracing an image back to the camera that captured it. This could be a useful tool in forensic applications to help identify potential suspects, especially in relation to cyber-crime (e.g., identifying the offender of child pornography or fake news creation). Thus, the issue needs to be tackled with extreme care as a wrongly determined source could lead to the unjust imprisonment of an innocent individual. To tackle this issue, forensic researchers developed numerous solutions based on the hardware or software-related traces left on the image during the acquisition process. In addition to this, every day, a large amount of images is shared among users through social networks such as Facebook, WhatsApp,

The associate editor coordinating the review of this manuscript and approving it for publication was Chuan Li.

and Instagram. Each social network has its proprietary compression standards, that are applied while uploading and downloading images. Such scenarios are challenging as any post-processing can suppress the intrinsic source information and makes the SCI task even more difficult and challenging for forensic analysts.

Over the last decade, a significant effort has been made to improve the performance of SCI techniques. These can be broadly split into two categories: (i) feature-based approaches relying on Photo Response Non-Uniformity (PRNU), a unique noise imprinted on images due to manufacturing defects in the camera sensor; and (ii) Convolutional Neural Network (CNN) based data-driven approaches that learn the intrinsic source information automatically. In the feature-based source identification category, several approaches have been developed to identify the source cam-

era by extracting the PRNU fingerprint from the images. Lukas et al. [1] firstly provided a method for extracting the PRNU from images to identify the source camera. The noise residuals from various images taken with a specific camera are extracted using a denoising filter and used their average as the reference pattern. The PRNU of a test image is then correlated with the reference pattern using Normalized Cross Correlation (NCC) to identify if the image is taken with the same camera. Later, the SCI performance was improved by eliminating the undesirable artifacts to enhance the quality and reliability of PRNU [2], [3], [4], [5], [6]. Various other studies have also used statistical features extracted from the PRNU fingerprint and trained the machine learning classifiers to map images to their source [7], [8]. However, all these techniques rely on PRNU which will be suppressed when images are shared through social networks [9] and may become a major bottleneck in achieving SCI.

In recent years, various attempts have been made to identify source cameras using deep learning. Bondi et al. [10] and Tuama et al. [11] used CNN to successfully identify the source camera model. Yao et al. [12] developed a CNN-based robust multi-classifier and attained a identification accuracy of nearly 100% over 25 camera models. Freire-Obregon et al. [13] proposed a deep learning based method for mobile device identification and found that performance degrades when multiple cameras of the same model are taken into account. The source identification technique proposed by Huang et al. [14] suggests that deeper CNN can achieve improved classification accuracy. By modifying AlexNet and including a Local Binary Pattern (LBP) pre-processing layer, Wang et al. [15] created a framework that allows the CNN to focus more on intrinsic camera-specific information. Following that, the residual network [16], [17] and Siamese network [18] are proposed to perform brand, model, and device identification. Sameer et al. [19] presented a deep learning technique to identify the source camera model of images shared through Facebook. Further, a data-driven approach was developed in [20] to compare PRNUs by using CNN for source identification. Although CNN-based approaches have been proven to be effective in identifying the camera model, they are incapable of accurately distinguishing various cameras of the same model. The individual device identification accuracy is far from forensically satisfactory (e.g., the data-driven approach proposed in [18] is able to achieve a classification accuracy of only 64.80% for three Apple iPhone 5c devices). Moreover, these CNN-based techniques suffer highly when the test images are subjected to unknown post-processing applied by social networks. The literature suggests that identifying cameras at the device-level is more challenging than identifying them by brand and model as the devices of the same model employ common in-camera processing techniques. Though there are limited works on source device identification, it is crucial to identify the exact device since it enables more accurate traceability.

Another practical concern in this study is linking the images created by a new camera that is not used in the training

process. As outlined above, whether it is a feature-based approach or a CNN-based approach, the dataset used for training includes a finite closed set of cameras. Source camera identification in the closed-set scenario is the problem of associating an image in question to a camera within a known set of cameras. In particular, the SCI system is constructed with just a limited number of cameras for verifying such techniques, and each image in the testing phase is attributed to one of the cameras used to train the system. In this scenario, the forensic investigator is confident that the image in question belongs to one of the cameras used to train the SCI system. Nevertheless, due to the rapid development of technology, new camera models are released to market rapidly making it difficult to maintain the camera database up to date. As a result, it is not possible to keep the SCI system trained on all existing camera models available in the market leading to the open-set problem. In the practical forensic scenario, the image in question can belong to either a set of known cameras used to train the SCI system or unknown cameras (i.e., cameras that were not available during the training time of the SCI system). Therefore, even if the image in question was not taken with any of the known cameras available during the training time of the data-driven model, the closed-set SCI system will incorrectly map the image to one of those known devices. Under such circumstances, the existing CNN-based classifier needs to be re-trained from scratch each time when the forensic analyst is confronted with the new camera under investigation. This might be computationally expensive and time-consuming and may become a major impediment when SCI systems involve the training of deeper CNNs.

The conventional techniques considered the open-set SCI task as the $(N + 1)$ classification problem, in which images of N known cameras and one class of all unknown cameras are used to train the classifier [21], [22], [23], [24]. Moreover, these classification task does not provide the forensic analyst with information about the actual camera in the open-set that captured the image. Huang et al. [23] developed a unknown detection method based on K -nearest neighbours and performed $(K + 1)$ classification. Although it can distinguish between images of known camera models, it is only able to identify images of unknown models but cannot determine the exact source camera model in the open-set. Junior et al., [24] proposed training protocols to properly estimate the parameters to perform open-set source identification by considering several algorithms. Mayer and Stamm [25] developed a CNN-based technique to extract the camera model information from image patches. Then a similarity score is measured between two patches to determine if they were both captured with the same camera. In [26], Bayar and Stamm proposed a constrained CNN to address the open-set SCI problem. Though the method in [26] achieved a high identification rate in open-set camera model identification, as we will show in Section III-D, it is by no means capable of identifying individual camera devices (different devices of the same model). Wang et al. [27] developed a data clustering method to differentiate between images from known camera and

images from unknown camera. However, to perform open-set source identification, the method needs to be re-trained on new camera model if it identifies that the image was captured with unknown camera. Thus, to address the open-set SCI problem, we propose a new robust data-driven system that is capable of effectively mapping an image to an individual camera that the SCI system has not seen during the training phase. In summary, our main contributions are:

- 1) We propose a CNN-based robust data-driven system that serves as the feature extractor for camera device identification without resorting to any hand-crafted features such as PRNU fingerprint.
- 2) To address the open-set camera device identification of digital images, we train the CNN on the images taken with a known set of cameras and exploit it as the feature extractor to extract the source information from the images taken with new set of cameras. Later, we train a classifier on the extracted features to map the images to the respective source categories in the open-set.
- 3) We show that the proposed data-driven system is robust to the post-processing applied by social network platforms.

The remainder of the paper is structured as follows. Section II presents the proposed approach to solve the open-set SCI problem. Experimental results and discussions are given in Section III. Finally, conclusions are drawn in Section IV.

II. METHODOLOGY

In this section, we present a data-driven approach for individual camera device identification of digital images in the open-set scenario. There are two major modules in the proposed approach: (i) a feature extractor, and (ii) a classifier. The overall architecture is shown in Fig. 1.

For the closed-set SCI evaluation, the feature extractor module in the proposed architecture computes a feature vector $f_{closed-set}$ from original images (images taken directly from the camera) $I_{closed-set}$ taken from known-set of cameras. It will be trained to capture the intrinsic source information from images. The feature vector $f_{closed-set}$ extracted by the trained feature extractor (F_E) that serves as the source information will be fed to the classifier module (M_C) to learn the mapping between feature vector $f_{closed-set}$ and the respective source camera in the closed-set ($C_{closed-set}$).

For the open-set SCI evaluation, original images from cameras ($C_{open-set}$) that are not considered for closed-set evaluation (i.e., $\forall c \in C_{open-set}, c \notin C_{closed-set}$) are split into training and testing sets. The trained feature extractor (F_E) that has learned to extract the source information in closed-set scenario is employed to extract the feature vectors $f_{open-set}$ from images in the training set $\{I_{open-set}\}_{train}$. The feature extraction is done by a single pass through the images in the $\{I_{open-set}\}_{train}$. At the training, the classifier M_o learns how to associate the images to the open-set cameras based on the features $f_{open-set}$. At the testing time, the trained classifier

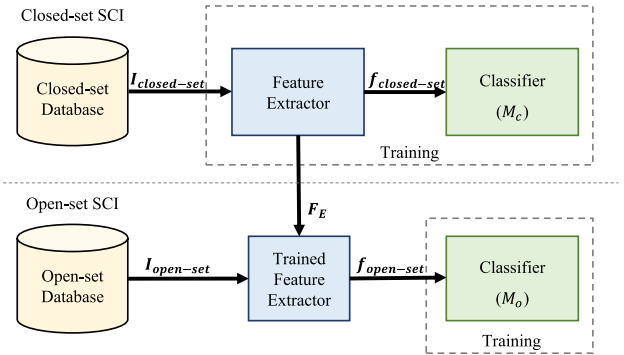


FIGURE 1. Proposed architecture for closed-set and open-set SCI.

M_o is used to estimate the source cameras of images in the testing set $\{I_{open-set}\}_{test}$ based on the feature vector extracted by learned feature extractor F_E .

Once the feature extractor is trained on closed-set cameras, it can be deployed to extract the source information f from any image in question. The classifier trained on the features extracted from images captured with suspect cameras is used to map the feature vector f to the authentic source camera.

A. FEATURE EXTRACTOR

The first phase in our approach is to build a feature extractor to learn the extraction of source information from the images. To do this, we employ residual network (ResNet) [28] as the feature extractor. The depth of the CNN is one of the crucial parameters that greatly affects the performance of the SCI task. However, the traditional scaling method does not improve the performance after a certain number of layers. It starts to have an adverse effect by lowering the CNN performance. With deeper CNN one can extract richer features, but it does not always lead to improved accuracy. This degradation in the performance issue is addressed by introducing the ResNet model [28]. It enables the training of a deeper CNN by using skip connections. These skip connections receive features from one layer and add to the layer that is deeper in the network and enable learning of both low-level and high-level features. Thus, we employ ResNet model that facilitates learning richer features by simultaneously extracting low-level and high-level features. The ResNet has different variants such as ResNet18, ResNet50, and ResNet101, each with a different number of layers. With more number of layers, the ResNet model can combine features from multiple layers, which aids in capturing features with strong discriminative capability. Thus, to have a strong feature representation that can effectively distinguish different camera devices we use the ResNet with 101 layers (ResNet101). Also, it was empirically found that the ResNet101 variant achieves improved performance over other variants of the ResNet. We extract the features from the *Global Average Pooling* (GAP) layer of the ResNet101 model, which results in a 2048 dimensional feature vector from each image. The GAP layer has no trainable parameters to optimize, which

avoids the overfitting of the model to the training data. Thus, the features extracted at this layer helps to enhance the performance of the proposed system in extracting the discriminative features.

B. CLASSIFIER

The second phase is to map an image to its source camera based on the features extracted by the feature extractor module. To do this, we train a multi-layer perceptron (MLP). The input layer of the proposed MLP receives a 2048 dimensional feature vector that serves as the intrinsic source information from the *Global Average Pool* (GAP) layer of the ResNet101 model. Two *dense layers* with 1024 and 64 neurons, respectively, are used in succession with the *Rectified Linear Unit* (ReLU) activation function to add non-linearity for each layer. To avoid overfitting of the model to the training data, a *dropout layer* with a probability of 0.5 is utilized, which de-activates some of the neurons in the layer and forces the learning to be independent in each iteration. Finally, we employ an output *dense layer* with the number of neurons equal to the number of cameras (D) used to train the classifier followed by the softmax function to predict the class probability.

C. TRAINING PROTOCOL

Formally, open-set SCI is the classification problem of linking an image in question to its exact source camera, where the source camera may belong to either known cameras (available during training time) or unknown cameras (not available during training time). To tackle open-set source camera device identification, we investigate the feasibility of using the CNN model trained on images taken with known set of cameras to capture camera specific features. We consider the open-set source camera identification as the scenario where the forensic analyst wants to trace back an image I to its source camera which is not the part of the training set used to train the feature extractor. In general, the proposed data-driven system (d) can be thought of as the composition of two functions, $f(\cdot)$ and $m(\cdot)$ as represented in equation 1,

$$d(I) = m(f(I)) \quad (1)$$

where $f(\cdot)$ is the feature extractor and $m(\cdot)$ is the classifier that differentiates between open-set cameras ($C_{open-set}$). Specifically, we define $f(\cdot)$ as the ResNet101 model (F_E , i.e., we use the terms $f(\cdot)$ and F_E interchangeably throughout the paper) and $m(\cdot)$ as the open-set MLP classifier that we discussed in Section II-A and Section II-B, respectively. The feature extractor potentially learns from $I_{closed-set}$ which represents the images taken with closed-set cameras. The classifier m learns from the features extracted by the feature extractor from training images taken with cameras in $C_{open-set}$. To accomplish this, we partition the dataset into two disjoint sets of cameras, i.e., known-set of cameras ($C_{closed-set}$) and unknown-set of cameras ($C_{open-set}$).

- **Closed-set data** ($C_{closed-set}$): images captured with known set of cameras available at training time of

TABLE 1. Dataset details.

Dataset	Camera model	No. of devices	Total no. of image patches	
Daxing	Vivo X9	3	1444	
	Xiaomi 4A	5	2304	
VISION	iPhone 5c	3	1523	
	iPhone 4s	2	761	
	iPhone 5	2	856	
	VISION-8	Samsung Galaxy S3 Mini	2	710
		OnePlus A3000	1	574
		OnePlus A3003	1	472
		Huawei P8	1	532
		Huawei Honor 5c	1	542

TABLE 2. Performance of the proposed system in a closed-set scenario.

Experiment	Cameras Used	Accuracy (%)	Average Precision (%)	Average Recall (%)
1	3 Vivo X9	97.08	97.24	97.08
2	5 xiaomi 4A	94.57	94.58	94.57

the proposed data-driven system. We split the images taken from the closed-set cameras into the training $\{I_{closed-set}\}_{train}$ and test sets $\{I_{closed-set}\}_{test}$. The training set is used to tune the feature extractor to learn camera-specific information from images. The feasibility of the feature extractor in mapping images to the known-set of cameras $c \in C_{closed-set}$ is evaluated using the test set.

- **Open-set data** ($C_{open-set}$): images captured with cameras that are not available at training time of the feature extractor. In a practical scenario, the forensic analyst may be confronted with an image that was not taken with any of the cameras used during the training process of the CNN model. The image in question may belong to either the closed-set cameras that were used to train the CNN model or a new camera that was not available during the training process (open-set). Since the open-set data were unavailable or unknown during the training of the CNN-based feature extractor, we treat them as ‘unknown’ from the point-of-view of the feature extractor. Even though the new cameras are known to the trained feature extractor or classifier when it is deployed to perform source identification in an open-set scenario, they were not known when the CNN model was built to learn feature extraction. Therefore, in the proposed work, we consider the open-set/unknown data as images captured with new cameras that were assumed to be not available during the training time to mimic the real-world forensic scenario.

In this approach, we first build and train the feature extractor F_E (i.e., ResNet101) using the training data from each of the cameras in the closed-set ($\{I_{closed-set}\}_{train}$). The closed-set classifier M_c is trained to map the images to the cameras in $C_{closed-set}$ based on the features $f_{closed-set}$ extracted by the feature extractor. For open-set evaluation, the proposed system relies on the CNN trained in a supervised manner on a known set of cameras. We retain the feature extractor F_E learned while training in the closed-set scenario,

but set aside the classifier M_c . We replace M_c with a new MLP classifier M_o designed to associate the images to the open-set cameras $C_{open-set}$. We extract features $f_{open-set}$ from the training images in the open-set $\{I_{open-set}\}_{train}$ using the feature extractor F_E . These features are fed to the open-set classifier M_o to learn mapping images to the cameras in the open-set. After the open-set classifier is trained to differentiate between cameras in the open-set it will be deployed to perform source identification of the images in the test set $\{I_{open-set}\}_{test}$.

The proposed method solves the limitation of incorrectly mapping the image to one of the cameras in the closed-set by the existing techniques. With the combination of feature extractor $f(\cdot)$ (trained on closed-set data) and classifier $m(\cdot)$ (trained on source information extracted from open-set data), we provide the solution d for source camera identification in the open-set scenario. During the investigation, the forensic analyst will have access to the image in question I , the suspect camera C , and the trained feature extractor (F_E) capable of extracting the camera-specific feature from the image in question. To map an image to a suspect camera based on closed-set approach, the forensic analyst must re-train the SCI system on the sample images taken with the suspect camera. On the contrary, in the open-set scenario, the proposed feature extractor (F_E) can be directly employed to extract source information without requiring to re-train it on the sample images taken by the suspect camera in order to attest that the image in question is taken by the suspect camera.

III. EXPERIMENTAL RESULTS AND DISCUSSION

A. EXPERIMENTAL SETUP

The original images in the Daxing dataset [29] are used to build the feature extractor in the closed-set scenario, whereas, the original and social network images from the VISION dataset [30] are used for the evaluation of the open-set scenario. The entire dataset is split into train and test sets, with 75% of images from each device being randomly selected for training and remaining images (25%) from each device being used for testing. Because the number of images for each camera is limited, we split each image into two equal sized patches to produce a larger dataset for the training process. Each patch is downsized to 224×224 pixels to meet the input size requirement of the ResNet101 model. Details of the camera device and the images used are provided in Table 1. We train the ResNet101 and MLP for 20 and 1500 epochs, respectively. The optimum set of parameters are obtained by minimizing the *categorical cross-entropy loss function* using *Adam* optimizer with the learning rate and mini-batch size set to 0.0001 and 12, respectively. Experiments are performed in MATLAB2021a using a HP EliteDesk 800 G4 Workstation with 32GB RAM and NVIDIA GeForce GTX 1080 GPU.

B. CLOSED-SET EVALUATION

First, we create the closed-set of cameras ($C_{closed-set}$) by considering the different camera devices of the same brand and

model to build the feature extractor. We train the ResNet101 model using the original images taken from cameras in $C_{closed-set}$ (as given in Table 2) to capture the source information for the exact device identification. Next, to verify the ability of the features captured by the feature extractor in identifying the source camera, we train the MLP classifier (M_c) using the 2048 dimensional features ($f_{closed-set}$) taken by the GAP layer of the trained ResNet101 model. In Table 2, we report the performance achieved by the proposed system on camera device identification in terms of classification accuracy, average precision and average recall achieved on the testing set. It can be noticed from the results that the proposed system is able to effectively trace back the images to their respective source cameras. It demonstrates the good ability of the proposed feature extractor in capturing the source information capable of distinguishing devices of the same model.

C. OPEN-SET EVALUATION

We conduct a set of experiments to evaluate the performance of the proposed system in identifying the exact source of original/social network images in an open-set scenario. For these experiments, we employ the ResNet101 model trained previously on the original images taken with (i) three devices of Vivo X9, and (ii) five devices of the Xiaomi 4A model as the feature extractor (F_E). Given the fact that in a real-world forensic investigation, the number of suspect cameras belonging to the same model is usually less, we choose a smaller number of devices for open-set evaluation to mimic the practical scenario (as reported in Table 3). Our approach proceeds by first extracting the features using the feature extractor (F_E) and training the MLP classifier (M_o) based on the features extracted from the training images $\{I_{open-set}\}_{train}$. Once the classifier is trained, it is used to predict the source camera of original test images $\{I_{open-set}\}_{test}$. Furthermore, we evaluate the robustness in two phases: (i) without prior knowledge of social network images (here, the MLP classifier (M_o) does not receive features extracted from the social network images for training), and (ii) with prior knowledge of social network images (here, the MLP classifier (M_o) receives features extracted from the social networks images for training). We evaluate the performance of the proposed system on original test images and all three post-processed versions of the test images (shared through Facebook in high quality: FBH, and low quality: FBL, and WhatsApp: WA). The identification accuracies achieved without prior knowledge of the social media images are reported in Table 3. From this table, we can observe that the proposed data-driven system has the ability to identify the source camera in the open-set scenario. In particular, the proposed feature extractor trained only on three device of Vivo X9 model is able to distinguish between eight cameras (VISION-8) with 88.97% identification accuracy and achieved greater than 90% identification accuracy in differentiating different devices of the same model. This demonstrates the good generalization capability of the proposed system in distinguishing camera devices that are never

TABLE 3. Identification accuracy achieved on original and social network images in the open-set scenario.

Feature Extractor	Cameras Used for Testing	Identification Accuracy (%)						
		Original	Without prior knowledge			With prior knowledge		
			FBH	FBL	WA	FBH	FBL	WA
3 Vivo X9	3 iPhone 5c	92.81	91.18	90.20	92.48	90.85	90.85	
	2 iPhone 4s	94.89	93.32	91.48	92.05	94.32	93.18	
	2 iPhone 5	99.02	97.55	97.06	97.06	98.53	99.02	
	2 Samsung Galaxy S3 Mini	92.76	90.79	88.82	88.82	91.45	90.79	
	2 iPhone 4s, 3 iPhone 5c, 2 iPhone 5	85.88	80.84	80.36	80.84	85.55	84.42	
	VISION-8*	88.97	86.76	86.76	86.53	87.99	87.75	
5 Xiaomi 4A	3 iPhone 5c	93.14	91.18	90.52	91.18	92.81	92.81	
	2 iPhone 4s	94.89	93.18	92.61	92.61	94.32	93.18	
	2 iPhone 5	99.51	96.08	96.08	96.08	98.04	97.06	
	2 Samsung Galaxy S3 Mini	92.11	86.16	84.87	88.82	88.18	87.50	
	2 iPhone 4s, 3 iPhone 5c, 2 iPhone 5	85.39	83.77	78.57	78.73	84.25	80.52	
	VISION-8*	87.25	85.05	84.31	84.80	86.27	85.54	

TABLE 4. Training time evaluation.

Cameras used	Training time in minutes	
	Time involved in re-training the entire model (Training ResNet and MLP)	Time involved in training only the classifier (Freezing Feature extractor + Training MLP)
2 Samsung Galaxy S3 Mini	17.51	0.49
2 iPhone 4s	19.64	0.50
2 iPhone 5	24.55	0.53
3 iPhone 5c	34.42	0.78
2 iPhone 4s, 2 iPhone 5, 3 iPhone 5c (7 cameras)	69.77	1.35
VISION-8 (8 cameras)	116.09	2.85

TABLE 5. Identification accuracy achieved in closed-set scenario by the state-of-the-art techniques trained on three devices of Vivo X9 model.

Method	Identification Accuracy (%)
Bondi et al., [10]	33.33
Yao et al., [12]	96.78
Obregon et al., [13]	96.49
Wang et al., [15]	96.78
Bayar and Stamm [26]	36.00
Chen et al., [16]	96.78
Proposed Method	97.08

seen before. Moreover, these results demonstrate that a CNN can learn camera-specific features to identify the unknown cameras even from images taken with cameras that are not used while training the feature extractor. Furthermore, albeit the images shared through social networks were not part of training the proposed classifier, the identification accuracy achieved on test images shared through social networks is close to the performance of the proposed system on original test images. Thus, the results show that the impact of unknown post-processing applied by the social network on the proposed system is very minimal.

We perform an additional experiment to see how prior knowledge of social media images benefits the proposed system in identifying the source camera of the degraded images. To do this, we train the MLP classifier (M_o) with features extracted from images shared through social networks, and the resulting accuracies on the test images are reported in Table 3. The results show that the prior knowledge yields marginally improved accuracies on shared images, which

TABLE 6. Comparison of proposed method with state-of-the-art methods.

Method	Device	Classification Accuracy (%)			
		Orig	FBH	FBL	WA
Bondi et al., [10]	3 iPhone 5c	33.33	33.33	33.33	33.33
	2 SG S3 Mini	50.00	50.00	50.00	50.00
Yao et al., [12]	3 iPhone 5c	38.89	38.89	38.89	37.58
	2 SG S3 Mini	58.55	57.24	55.92	57.89
Obregon et al., [13]	3 iPhone 5c	57.84	56.54	55.88	55.88
	2 SG S3 Mini	73.03	71.05	69.08	71.71
Wang et al., [15]	3 iPhone 5c	44.44	43.46	41.50	42.81
	2 SG S3 Mini	59.21	57.89	57.89	58.55
Bayar and Stamm [26]	3 iPhone 5c	36.60	36.37	37.25	37.25
	2 SG S3 Mini	48.33	47.78	46.67	47.22
Chen et al., [16]	3 iPhone 5c	67.32	66.67	66.33	66.01
	2 SG S3 Mini	74.32	72.37	71.71	67.11
ResNet101 + SVM	3 iPhone 5c	91.50	90.52	89.87	90.85
ResNet101 + KNN	3 iPhone 5c	82.03	81.37	81.37	80.72
ResNet50 + MLP	3 iPhone 5c	90.20	89.54	89.22	89.87
ResNet18 + MLP	3 iPhone 5c	86.27	84.64	84.31	84.97
Proposed Method (ResNet101+MLP)	3 iPhone 5c	92.81	91.18	90.20	90.20
	2 SG S3 Mini	94.74	94.74	92.11	91.45

suggests that the proposed feature extractor has good capability to learn the source information from the degraded images. It is worth highlighting that the proposed feature extractor is trained on original images taken with closed-set cameras. Therefore, even if it is trained on original images, it still has the good generalization ability to identify the exact device of the social network images taken with cameras that are never seen before with considerably high accuracy.

Further, we have calculated the time involved in re-training the entire model (ResNet101 + MLP classifier) from scratch and training only the classifier by freezing the trained

feature extractor (ResNet101). It is observed that on average re-training the entire model takes around 20 minutes for two cameras and 116.09 minutes for eight cameras. Whereas employing the ResNet101 model as a feature extractor and training the MLP classifier takes only 0.5 minutes for two cameras and 2.85 minutes for eight cameras. Also, when there is more number of cameras involved, training time might be seen to increase significantly. Therefore, re-training the deep CNN is not feasible in the real-world forensic scenario.

D. PERFORMANCE COMPARISON WITH STATE-OF-THE-ART

Various CNN frameworks are used in the literature to perform source camera identification. Here, we employ the state-of-the-art CNNs [10], [12], [13], [15], [16], and [26] as the feature extractor to investigate their feasibility for an open-set scenario. For this comparative analysis, we consider original images taken from three devices of Vivo X9 model to train the feature extractor. We compare the performance of the proposed method with the state-of-the-art techniques in the closed-set scenario. The identification accuracy achieved on the test images taken from three devices of the Vivo X9 model are reported in Table 5. We can notice that the state-of-the-art data-driven systems [12], [13], [15], [16] have performed equally well in distinguishing different devices of the same model in closed-set scenario.

For the open-set evaluation, we consider the state-of-the-art feature extractors trained on three Vivo X9 devices to perform source identification. We summarize the achieved results in terms of classification accuracy on original (Orig)/social network images in Table 6. It can be observed that the state-of-the-art CNNs that have shown excellent performance in closed-set SCI in the literature (as seen in Table 5) report a significant drop in the classification accuracy in open-set scenario. In particular, the feature extractors [12], [13], [15], [16] that achieved a classification accuracy of over 96% in closed-set source identification were unable to trace back images to the unknown cameras. For example, the performance of the CNN developed by Yao et al. [12] that had achieved 96.78% accuracy in closed-set scenario dropped to 38.89% in the case of distinguishing three devices of the iPhone 5c model in the open-set scenario. Noticeably, the residual network with five convolutional blocks proposed by Chen et al. [16] specifically to perform exact device identification is ineffective in differentiating devices of the same model in the open-set scenario. Further, the CNN framework proposed by Bayar and Stamm [26] specifically to address the camera model identification in open-set scenario performed poorly in the case of exact device identification. The possible reasons for the inferior performance of the state-of-the-art CNNs in open-set scenario is the use of shallow CNN, which may not be sufficient to learn generalized information to perform source identification in the open-set scenario. In particular the forensic features learned by those CNNs are overfitting to the cameras used to train the feature extractor

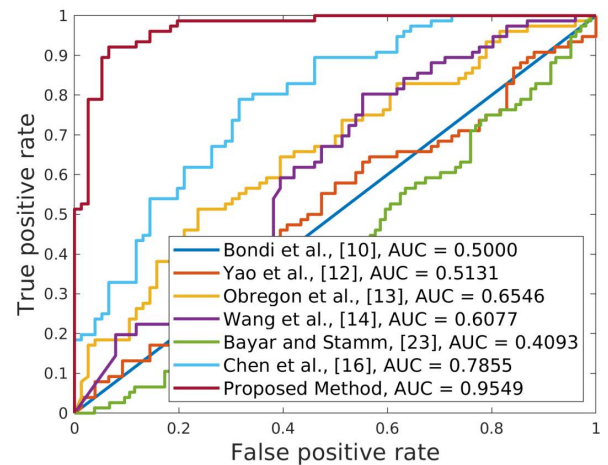


FIGURE 2. ROC for the proposed method and state-of-the-art.

and hence unable to distinguish between camera devices used in the open-set scenario.

By leveraging the powerful learning capability of the residual neural network we were able to learn the intrinsic source information to trace back the images to the cameras in the open-set scenario. To identify the most suitable ResNet variant for the open-set SCI problem, we analyse the behaviour of other variants such as ResNet18 and ResNet50 paired with MLP classifier. The identification accuracy achieved by different variants of residual network in differentiating three devices of the iPhone 5c model are reported in Table 6. We can notice that the ResNet101 variant has outperformed the other variants in terms of source identification accuracy, further demonstrating the vital role that deeper CNN plays in learning the most discriminating characteristics for source camera identification. To further confirm the effectiveness of the proposed system, we use ResNet101 as the baseline and train the machine learning classifiers such as SVM and KNN using the features extracted by the ResNet101. Specifically, it shows which combination of the feature extractor and classifier enables to achieve high performance in open-set scenario. Noticeably, our feature extractor (ResNet101) associated with the MLP classifier outperforms the choice of other classifiers. The more comprehensive picture of the feasibility of the source information extracted by various feature extractors is illustrated in Figure 2 by plotting Receiver Operating Characteristic (ROC) curves for the Samsung Galaxy (SG) S3 Mini device. The proposed system surpassed the state-of-the-art CNNs by a large margin, with an area under curve of 0.9549. The existing techniques are adversely affected due to the presence of unknown camera devices which greatly hinders their classification capability. This clearly shows that the forensic feature extraction learned by existing CNNs developed for closed-set SCI is not appealing for open-set scenario.

IV. CONCLUSION

In this work, we propose a CNN-based data-driven system to perform open-set SCI on original images and those shared

through social networks. To do this, we train the ResNet101 model on images taken with the closed-set of cameras to learn to capture the source information from images. For the evaluation of the open-set scenario, we exploit the feature extractor trained on a known set of cameras and train the MLP classifier on the extracted features to perform source identification on both original and social network images. The comparative analysis shows that our proposed system has a good generalization to unseen cameras and is resilient to the unknown compression applied to the images by social network platforms. Testing on various sets of cameras confirms the effectiveness of the proposed system in extracting the source information to identify the source camera of a new image under investigation.

Some of the aspects of future research in this regard would be to investigate the performance on the doctored images to test the robustness of the proposed method. Further, images that have been altered using a phone other than the source camera may contain traces that are not unique to the source camera, opening the door to the counter-forensic investigation. The encouraging results obtained motivate us to extend the proposed work to perform a detailed analysis of the influence of such images on the identification results of the proposed method.

REFERENCES

- [1] J. Lukáš, J. Fridrich, and M. Goljan, "Digital camera identification from sensor pattern noise," *IEEE Trans. Inf. Forensics Security*, vol. 1, no. 2, pp. 205–214, Jun. 2006.
- [2] 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.
- [3] B. Gupta and M. Tiwari, "Improving performance of source-camera identification by suppressing peaks and eliminating low-frequency defects of reference SPN," *IEEE Signal Process. Lett.*, vol. 25, no. 9, pp. 1340–1343, Sep. 2018.
- [4] X. Lin and C. T. Li, "Enhancing sensor pattern noise via filtering distortion removal," *IEEE Signal Process. Lett.*, vol. 23, no. 3, pp. 381–385, Mar. 2016.
- [5] Q. Rao and J. Wang, "Suppressing random artifacts in reference sensor pattern noise via decorrelation," *IEEE Signal Process. Lett.*, vol. 24, no. 6, pp. 809–813, Jun. 2017.
- [6] A. Mehrish, A. V. Subramanyam, and S. Emmanuel, "Sensor pattern noise estimation using probabilistically estimated RAW values," *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 693–697, May 2016.
- [7] K. R. Akshatha, A. K. Karunakar, H. Anitha, U. Raghavendra, and D. Shetty, "Digital camera identification using PRNU: A feature based approach," *Digit. Invest.*, vol. 19, pp. 69–77, Dec. 2016.
- [8] A. Tuama, F. Comby, and M. Chaumont, "Camera model identification based machine learning approach with high order statistics features," in *Proc. 24th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2016, pp. 1183–1187.
- [9] A. K. Karunakar and C.-T. Li, "Identification of source social network of digital images using deep neural network," *Pattern Recognit. Lett.*, vol. 150, pp. 17–25, Oct. 2021.
- [10] 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.
- [11] A. Tuama, F. Comby, and M. Chaumont, "Camera model identification with the use of deep convolutional neural networks," in *Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, Dec. 2016, pp. 1–6.
- [12] H. Yao, T. Qiao, M. Xu, and N. Zheng, "Robust multi-classifier for camera model identification based on convolution neural network," *IEEE Access*, vol. 6, pp. 24973–24982, 2018.
- [13] D. Freire-Obregón, F. Narducci, S. Barra, and M. Castrillón-Santana, "Deep learning for source camera identification on mobile devices," *Pattern Recognit. Lett.*, vol. 126, pp. 86–91, Sep. 2019.
- [14] N. Huang, J. He, N. Zhu, X. Xuan, G. Liu, and C. Chang, "Identification of the source camera of images based on convolutional neural network," *Digit. Invest.*, vol. 26, pp. 72–80, Sep. 2018.
- [15] B. Wang, J. Yin, S. Tan, Y. Li, and M. Li, "Source camera model identification based on convolutional neural networks with local binary patterns coding," *Signal Process., Image Commun.*, vol. 68, pp. 162–168, Oct. 2018.
- [16] Y. Chen, Y. Huang, and X. Ding, "Camera model identification with residual neural network," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 4337–4341.
- [17] X. Ding, Y. Chen, Z. Tang, and Y. Huang, "Camera identification based on domain knowledge-driven deep multi-task learning," *IEEE Access*, vol. 7, pp. 25878–25890, 2019.
- [18] V. U. Sameer and R. Naskar, "Deep Siamese network for limited labels classification in source camera identification," *Multimedia Tools Appl.*, vol. 79, nos. 37–38, pp. 28079–28104, Oct. 2020.
- [19] V. U. Sameer, I. Dali, and R. Naskar, "A deep learning based digital forensic solution to blind source identification of Facebook images," in *Proc. Int. Conf. Inf. Syst. Secur. Cham, Switzerland: Springer*, 2018, pp. 291–303.
- [20] 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.
- [21] T. Gloe, "Feature-based forensic camera model identification," in *Transactions on Data Hiding and Multimedia Security* (Lecture Notes in Computer Science). Berlin, Germany: Springer, 2012, pp. 42–62.
- [22] F. D. O. Costa, E. Silva, M. Eckmann, W. J. Scheirer, and A. Rocha, "Open set source camera attribution and device linking," *Pattern Recognit. Lett.*, vol. 39, pp. 92–101, Apr. 2014.
- [23] Y. Huang, J. Zhang, and H. Huang, "Camera model identification with unknown models," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 12, pp. 2692–2704, Dec. 2015.
- [24] P. R. Mendes Junior, L. Bondi, P. Bestagini, S. Tubaro, and A. Rocha, "An in-depth study on open-set camera model identification," *IEEE Access*, vol. 7, pp. 180713–180726, 2019.
- [25] O. Mayer and M. C. Stamm, "Learned forensic source similarity for unknown camera models," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 2012–2016.
- [26] B. Bayar and M. C. Stamm, "Towards open set camera model identification using a deep learning framework," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 2007–2011.
- [27] B. Wang, Y. Wang, J. Hou, Y. Li, and Y. Guo, "Open-set source camera identification based on envelope of data clustering optimization (EDCO)," *Comput. Secur.*, vol. 113, Feb. 2022, Art. no. 102571.
- [28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [29] H. Tian, Y. Xiao, G. Cao, Y. Zhang, Z. Xu, and Y. Zhao, "Daxing smartphone identification dataset," *IEEE Access*, vol. 7, pp. 101046–101053, 2019.
- [30] 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.



MANISHA received the B.E. and M.Tech. degrees in electronics and communication engineering from Visvesvaraya Technological University, Belagavi, Karnataka, India, in 2015 and 2017, respectively. She is currently pursuing the joint Ph.D. degree with the Department of Data Science and Computer Applications, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, India, and the School of Information Technology, Deakin University, Australia. Her research interests include digital image forensics and deep learning.



CHANG-TSUN LI (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from the National Defence University (NDU), Taiwan, in 1987, the M.Sc. degree in computer science from the U.S. Naval Postgraduate School, USA, in 1992, and the Ph.D. degree in computer science from the University of Warwick, U.K., in 1998. He was an Associate Professor with the Department of Electrical Engineering, NDU, from 1998 to 2002, and a Visiting Professor with

the Department of Computer Science, U.S. Naval Postgraduate School, in 2001. He was a Professor with the Department of Computer Science, University of Warwick, until January 2017, and a Professor with Charles Sturt University, Australia, from January 2017 to February 2019. He is currently a Professor with the School of Information Technology, Deakin University, Australia. His research interests include multimedia forensics and security, biometrics, data mining, machine learning, data analytics, computer vision, image processing, pattern recognition, bioinformatics, and content-based image retrieval. The outcomes of his multimedia forensics research have been translated into award-winning commercial products protected by a series of international patents and have been used by a number of police forces and courts of law around the world. He is currently the *EURASIP Journal of Image and Video Processing* (JIVP) and an Associate Editor of *IET Biometrics*. He is currently the Chair of Computational Forensics Technical Committee (TC-6) of the International Association of Pattern Recognition (IAPR). He contributed to the organization of many international conferences and workshops and also served as member of the international program committees for several international conferences. He is also actively contributing keynote speeches and talks at various international events and institutions.



KARUNAKAR A. KOTEGAR (Senior Member, IEEE) received the B.Sc. and M.C.A. degrees from Karnataka University, Karnataka, India, in 1995 and 1998, respectively, and the Ph.D. degree from the Manipal Academy of Higher Education, Karnataka, in 2009. He is currently the Director of the International Collaborations of Manipal Academy of Higher Education (MAHE), Manipal, India, and a Professor with the Department of Data Science and Computer Applications, Manipal

Institute of Technology, MAHE, Karnataka. His research interests include image/video processing and communication, scalable video coding, media aware network elements, multi-view video coding, scalable video over peer-to-peer networks, error resilient and concealment for scalable video, stereo vision, and image and video forensics.

...