IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 6, 2021, accepted August 25, 2021, date of publication September 3, 2021, date of current version October 6, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3110297

Ensembling Shallow Siamese Neural Network Architectures for Printed Documents Verification in Data-Scarcity Scenarios

ANSELMO FERREIRA, (Member, IEEE), NISCHAY PURNEKAR, AND MAURO BARNI, (Fellow, IEEE)
Department of Information Engineering and Mathematics, University of Siena, 53100 Siena, Italy
Corresponding author: Anselmo Ferreira (anselmo.ferreira@gmail.com)

This work was supported in part by European Union Marie Sklodowska-Curie Project PrintOut under Grant 892757, and in part by the Defense Advanced Research Projects Agency (DARPA) and Air Force Research Laboratory (AFRL) under Agreement FA8750-20-2-1004.

ABSTRACT The popularity of printing devices has multiplied the diffusion of printed documents, raising concerns regarding the security and integrity of their content. The same device that prints reliable contracts, newspapers, and others, can also be used for malicious purposes, such as printing fake money, forging fake contracts, and produce illegal packaging, thus calling for the development of image forensics techniques to pinpoint criminal printed materials and trace back to their origin. Despite some recent advances, previous works model such a problem as a big data-focused closed-set classification problem. In this work, we address the source linking problem of printed color documents by treating it as a verification problem. Specifically, we aim at deciding if two documents have been printed by the same printer or not. To achieve this goal, and to cope with the data scarcity deriving from the difficulty of gathering massive amounts of printed and scanned documents, we propose to use an ensemble of Siamese Neural Networks, with unique architectures expressly designed to work with a small training dataset. As a further unique feature, the proposed approach is suited to work in an open set scenario, where the printers used to produce the documents analyzed at the test time are not included in the training set. Results obtained under both open and closed set conditions, with a thorough comparison with available baseline methods, showed classification performance higher than 97% in the closed set scenario and higher than 86% in the open set case, highlighting the practicality of such approaches in real-world scenarios.

INDEX TERMS Digital image forensics, laser printer forensics, printer source attribution, Siamese networks.

I. INTRODUCTION

Despite the tremendous efforts to replace the number of printed documents with their digital counterparts, printed documents are still common and can be found everywhere. Generating printed documents is very cheap, with printing devices being more accessible and easy to use than ever. This accessibility and ease of use plays a major role in the production of vast amounts of printed documents, as we see every day. From advertisements to currencies, books to newspapers, magazines to contracts, and product packaging, there is always a printing technology involved. With the diffusion of staff-less and cashless stores in big cities [1], there will be only printed data available (such as the QR-CODES) for purchases and interaction with clients, making printing and scanning technologies even more important in a near future.

Notwithstanding the wide diffusion of printed information, the lack of regulation and forensic procedures on printed documents allow counterfeiters and other criminals to use printing technology for malicious purposes. For example, fake currency can be printed and distributed in a neighborhood, thus harming the local economy; pedophiles can print and distribute child porn to avoid the controls over the Internet of security agencies; deceivers can fake badges to access restricted areas, hitting up the organization and the security of meetings and other kinds of events.

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

Additionally, the analysis of printed documents related to other criminal activities like corruption or terrorism may help to trace back to crime perpetrators.

In addition to the previous issues, modern printing and scanning technologies have made counterfeiting easier and more profitable than ever, as counterfeiters can perfectly copy and print packages of fake products that are visually identical to the original ones. Acknowledging such a problem, the International Chamber of Commerce has raised an alarm of 3.7 trillion loss due to counterfeiting and piracy, with 5.4 million jobs at risk by 2022 [2]. Finally, counterfeiting has also a significant impact on health: according to the World Health Organization, up to half of the malaria medications could be fake [3].

As an answer to the above needs, several works have focused on present Printer Source Linking solutions using Computer Vision and Machine Learning to pinpoint the ownership of printed texts [4]–[23], color images [24]–[34] or both [35]–[37]. In particular, the approaches based on Convolutional Neural Networks (CNNs) and Deep Learning (DL), in general, [19], [21], [37] have allowed significant progress in this research area, in special due to their ability of learning from the data itself, artefacts specific to given printers. CNN-based solutions usually have better performance than previous image/noise description approaches for printer attribution.

When applied in real-life scenarios, however, CNN-based approaches have the following significant limitations: (i) as far as we know, there is no study in the literature considering the fact that the source of an unknown document is not included in the training set. Such an open-set setting is the most probable one in the real world and the lack of methods suitable to work in such a scenario limits strongly the practical usability of CNN techniques proposed so far; and (ii) DL models require a huge amount of data (printed and scanned documents) for training, especially when very deep networks with a huge amount of parameters are used. However, the dataset generation process is not only expensive (given that printer inks/toners are among the most expensive liquids in the world [38]), but it is also a very time-consuming task, requiring dedicated personnel and very strict procedures.

In this paper, we propose to deal with the previous limitations by treating the source linking problem as a *verification* problem. According to such an approach, any new sample is compared against a reference document, producing a score indicating how similar the unknown sample and the reference template are. Given that the network is trained to decide whether two printed documents have been generated by the same device, the samples analyzed at test time do not need to been generated by one of the printers used during training, thus opening the way to the use of the network in an open-set scenario. Indeed, a similar approach has been conveniently adopted in biometric recognition [39], where one person is allowed to access a system if his/her biometric traits are similar to anyone in the dataset of enrolled users, even when the system has not been trained to directly ''recognize'' the biometric traits of the users. However, In contrast to the biometrics scenario, where visually-relevant traits are compared, the answer to the question: ''have two documents containing different or even the same content, been generated by the same printer?'', requires the analysis of subtle, texture-like patterns produced by different printers, possibly neglecting the semantic of the images. The answer to such a question can be useful in several forensics and anti-counterfeiting applications. As a first example, we may consider a situation where a printed child pornography picture is found and someone is suspected to have printed it. The investigators can print other pictures with the printer of the suspect person and detect if the criminal document has been printed by the same printer.

Anti-counterfeiting provides another scenario fitting our solution. Assuming that authentic packages or labels can be printed only by a pool of authorized printers. If, in a suspected counterfeited merchandise, the printing artefacts of the barcode (or any other pattern) printed on the package do not match the patterns present in the original (templates) barcodes printed by an authorized printer, the product is classified as being counterfeited. Finally, the manipulation of printed documents is another interesting application of our research. As the verification network is trained to find similarities, or dissimilarities, in printed patterns, if a document (*e.g.*, a contract) is counterfeited by replacing some pages with fake ones printed by a different printer, then the fake document can be revealed by verifying that all the image pairs of the documents have been printed by the same printer.

In order to develop a printed document verification system, expressly thought to cope with the data scarcity problem typical of such a scenario, we propose to use an ensemble of Siamese Neural Networks (SNNs) with novel architectures characterized by shallow-but-wide topologies. In particular, we tested the performance of two different ensembles: one with diverse SNN topologies and another consisting of networks having the same topology, but with some tweaks in the SNNs parameters. The results we got in the closed set and, most importantly, in the open set scenarios built based on a challenging realistic dataset, demonstrate the promising performance of the proposed solution.

In summary, the contributions of this paper are:

1) We propose the use of Siamese Networks for printed documents attribution, treating it as a *verification* problem. This approach marks a significant difference concerning state-of-the-art solutions, which mostly treat such a problem as a *classification* one.
2) To deal with the availability of small training datasets, we propose several shallow-and-wide SNN architectures, that can be trained also on small amounts of data.
3) We propose two different ensemble strategies with and without diversity in the network topologies, achieving different, and somewhat complementary results.
4) As far as we know, this is the first study dealing with open set source linking of printing devices.

Last but not least, we compare all the proposed solutions against several baseline techniques, in both open and closed set conditions.

The remaining of this paper is organized as follows: in Section II, we discuss some progress of the related work in detail. Section III describes our printer verification solution composed of SNNs ensembles. Section IV details the experimental procedure adopted to validate the solutions for the printer verification problem. Section V reports the results of the experiments we carried out and, finally, Section VI concludes this paper highlighting future research directions and open problems.

II. RELATED WORK

Several previous works have investigated artifacts that printers leave in the generated documents in order to identify their source. Although several other works have worked on physical, microscopical, and chemical techniques [40], most of them are destructive methods that require specialized staff and expensive material [35]. Therefore, we focus our discussion on methods based on machine learning and computer vision, as they are cheaper, faster, and simpler. For the task of pinpointing the owner of printer documents, most of the surveys in the literature [41]–[44] divide them into the following branches: (i) approaches focused on text documents source linking; (ii) approaches for colored documents source linking; and (iii) approaches for any kind of document. Several other techniques have been proposed to identify among different printing technologies (*e.g.,* inkjet, laser, *etc.*) [45]–[49], however, in this section, we focus on linking the laser printer source of any kind of documents.

The literature has reported two possible clues that can be used to pinpoint the source of a text document. The first one is the *extrinsic signatures*. These signatures are introduced by the printing process in order to explicitly input the printer biometrics in any printed document. Such signatures are investigated by *active forensic methods*, and some examples are the embedding code sequences in electrophotographic halftone images [50] and also machine identification codes [51]. Approaches based on extrinsic signatures are usually expensive because they require expensive modifications of the printer. The main disadvantages of techniques based on extrinsic signature are the fact that some of these signatures can be hidden from the printed material [52] and that they are not adopted as a standard for all printer brands.

The second branch of investigation analyses *intrinsic signatures* through *passive forensic methods*. Intrinsic signatures are unintentionally injected into the printed document due to electromechanical printer devices imperfections. The most investigated intrinsic signatures are banding, jitter, and skewed jitters [25]. As the artifacts investigated by such a branch of research are difficult to erase, the literature has been mostly focused on such signatures, which are also the focus of our proposed solution.

In the case of text documents, intrinsic signatures are not so explicit given that the printed data is very small, and such artifacts are often visible only employing a microscope [35]. Due to this limitation, literature solutions have been focused on character extraction and texture description of such images [27]. For example, in one of the first works in this regard, Ali *et al.* [4] extracted "I" letters from printed text, using the raw pixel values in a multi-class machine learning algorithm to discriminate different printers intrinsic artifacts. Given an unknown document, majority voting of individual "I" letters classification defines the printer source of the document. Inspired by this solution, several works have followed a similar path. Some examples are the Gray Level Co-occurrence Matrices (GLCMs) features from Mikkilineni *et al.* [5]–[7], [12] and Ferreira *et al.* [35], features from Distance Transform from Deng *et al.* [9], features from Discrete Cosine Transform from Jiang *et al.* [11], statistics of GLCMs, residual noise and Wavelet Transform features from Tsai and Liu [13], Tsai *et al.* [15], [18] and Elkasrawi and Shafait [14], *ad-hoc* texture descriptors from Joshi and Khanna [20], [23], SURF and ORB features from Kumar *et al.* [53], and geometric distortions signatures from Jain *et al.* [22]. Finally, the use of deep neural networks from Ferreira *et al.* [19] and their extension from Joshi *et al.* [21] proved their ability to learn better the features from the data itself when sufficient data is used for training.

In the case of colored documents, which is the focus of the research presented in this paper, intrinsic artifacts are more frequently found as more areas in the paper are printed. One of the first works in this direction comes from Ali *et al.* [24], who extracted from image patches Fourier transform features to discriminate color laser printers through different banding frequencies. Eid *et al.* [25] looked for jitter artifacts with Gabor filtering and Discrete Fourier Transform. Choi *et al.* [27] looked for specific banding artifacts by calculating noise features from Discrete Wavelet Transform sub-bands in RGB and CMYK channels. Later, the same authors estimated noise with Wiener filtering and Gray Level Co-occurrence Matrix statistics [28]. Tsai *et al.* [30] extracted features from different Discrete Wavelet Transform sub-bands, adding to the pipeline feature selection. Other important techniques for color documents source attribution involve describing geometric distortions [26], [33] and halftone texture descriptors [29], [32], [34].

The last branch of approaches is those that can deal with any kind of document, being it a color-image or a text document. For example, the work of Ferreira *et al.* [35] proposed texture descriptor approaches based on GLCMs, together with a Convolutional Texture Convolutional Filter approach to be applied on frames, which are areas of the printed paper with sufficient printed material. Tsai *et al.* [36] applied several filters to the documents, extracting features from GLCMs, Discrete Wavelet Transform, spatial filters, Gabor filter, Wiener filter, Gray Level Co-occurrence Matrices features, and fractal features. In recent work, Nguyen *et al.* [54] printed eight different patterns and investigated the printer attribution using shape descriptor indexes features on Support Vectors Machines and Random

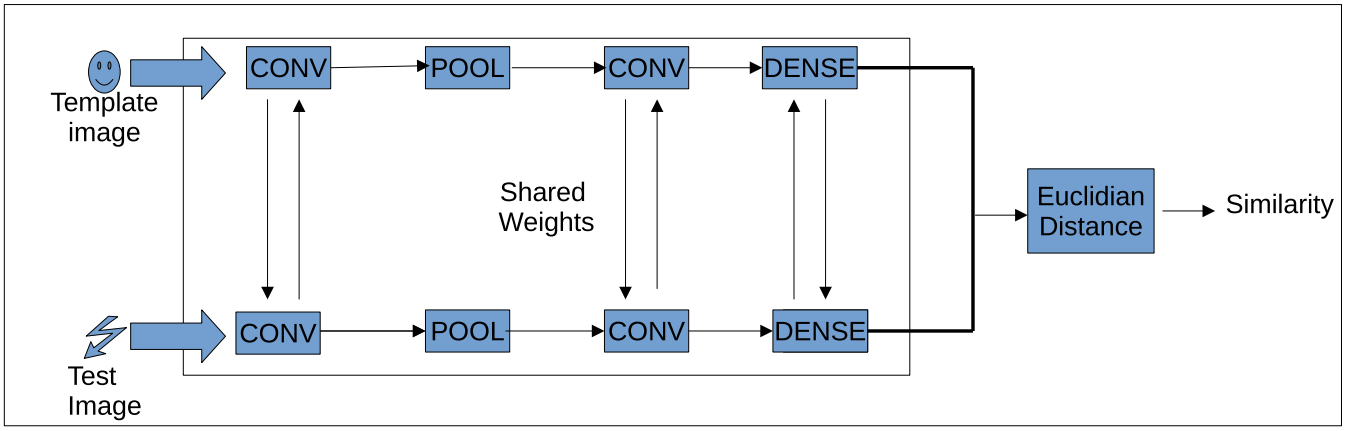


FIGURE 1. General pipeline of a Siamese network, where trainable layers share weights among two neural networks in order to evaluate the similarity of the input image pair.

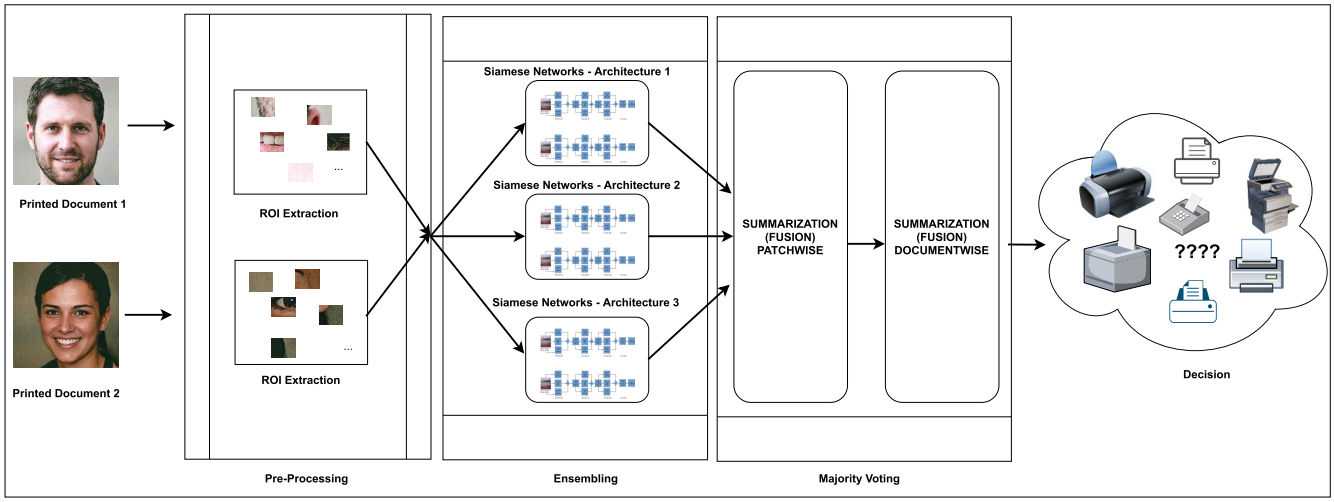


FIGURE 2. The proposed printed document verification pipeline.

Forest classifiers to perform printer attribution. Finally, Bibi *et al.* [37] reached state of the art results by using convolutional neural networks.

Notwithstanding the progress made in the last 20 years, there are still several gaps and limitations that should be explored, including: (i) solutions proposed so far treat laser printer attribution as a *classification* problem, neglecting the possibility to treat it as a *verification* problem; (ii) the datasets considered in the research are quite outdated, without professional printers being considered; and (iii) as far as we know, the open set laser printer attribution has not been explored in the literature. In the next section, we detail our solution to tackle such limitations.

III. PROPOSED METHOD

In this paper, we propose to treat the printed document source linking problem as a *verification* problem. To do so, we train an ensemble of Siamese Neural Networks on a dataset of printed images. Siamese networks can learn similarities from pairs of images to minimize the distance of similar pairs by sharing the same weights in the two networks. The general pipeline of a Siamese neural network is reported in Figure 1.

Siamese Networks have the following advantages over common CNNs for our specific problem: (i) they learn to find similarities or dissimilarities between different documents according to the printing sources, learning to detect if the patterns introduced during the printing process originate from the same printer or not; (ii) Siamese Neural Networks are also known for their one-shot learning capabilities [55]–[57], not requiring many samples for successful learning; and (iii) since they learn to match - or unmatch - similar pairs no matter their class labels, they can be used in an open set modality to also classify samples produced by printers that do not belong to the training set.

The pipeline of the proposed method is illustrated in Figure 2. During the training stage, patches of interest are extracted from a dataset of known printers, and then several Siamese Networks are trained so to let them learn to minimize the distances between patches from the same printers and maximize them when the patches have been generated

by different printers. After the models have been trained, we ensemble them in two manners: (i) *without diversity*, where we tweak only some parameters of a unique SNN architecture; and (ii) *with diversity*, where completely diverse SNN topologies are considered. During testing, unknown test patterns are compared by these networks with a template generated by a printer that does not necessarily belong to the training dataset. Each pair of patches is classified according to a majority voting procedure. Finally, the second round of majority voting, this time merging the results obtained on every single patch, is applied to decide on the source of the analyzed images. In the following subsections, we detail each of the above steps of our proposed ensemble. For reproducibility and open access purposes, the source code of our approach can be found at GitHub.[1]

A. PRE-PROCESSING (ROI EXTRACTION)

Most of the solutions in the literature consider source linking in specific regions of the printed material, where the artifacts investigated are expected to be more visible: Ali *et al.* [24] searched for letters "I", Kee and Farid [8] performed their investigation by focusing only on letters "E", and Ferreira *et al.* [35] searched for artifacts on areas with sufficient printed material.

In this paper, we follow a path similar to that adopted in [58], focusing on the top-n high energy patches of printed documents after Canny filtering binarization. Selecting pairs of patches in this way has a threefold advantage: (i) more relevant information about printer sources is acquired, as the top-n high energy patches are usually noisy areas containing more information about printing patterns (edges/banding with high energy); and (ii) they can generate significant training data of different printers from a relatively small dataset of printed documents, as one printed document is now represented by several high energy patches; (iii) The binarization and edge extraction applied by Canny filtering helps to remove the semantic content of the images, given that all the edges are equally weighted thanks to binarization.

The first step for the detection of high-energy patches consists of the application of a Canny filter. Such a filter initially applies a Gaussian filter to images. The Gaussian filter F of size $(2k+1) \times (2k+1)$ is defined as follows:

$$F(i,j) = \frac{1}{2\pi\sigma^2} \times \mathrm{e}^{\left(\frac{(i-(k+1))^2 + (j-(k+1))^2}{2\sigma^2}\right)}, \quad (1)$$

where σ is the standard deviation of the Gaussian distribution.

After Gaussian lowpass filtering, the smoothed image is filtered with a Sobel kernel in both horizontal and vertical directions to get first the derivative in the horizontal (G_x) and vertical (G_y) directions. The derivatives are computed by applying to the images two 3×3 kernels, defined as follows:

$$G_x = \begin{vmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{vmatrix}, \quad G_y = \begin{vmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{vmatrix}. \quad (2)$$

[1] https://github.com/anselmoferreira/siamese4PrintedVerification

After such convolutions, the magnitude G and the direction θ of the gradient are computed:

$$G = \sqrt{G_x^2 + G_y^2}, \qquad \theta = atan2(G_y, G_x). \quad (3)$$

Then, the edges are processed by two final steps. The first one, commonly called *non-maximum suppression*, eliminates thin edges. To do that, the strength of each edge pixel is compared with the edge strength of the neighboring pixels in the positive and negative θ gradient directions calculated in Equation 3. If the edge strength G of the current pixel is the largest, such a value is preserved, otherwise, it is set to zero. The second step uses one low and one high threshold value to filter edges. The following comparisons are then made using the two thresholds: (i) if a pixel gradient is higher than the upper threshold, the pixel is an edge detected; (ii) if a pixel gradient value is below the lower threshold, then it is rejected; and (iii) if the pixel gradient is between the thresholds interval, then it is accepted only if it is connected to an accepted edge (a pixel whose gradient is higher than the upper threshold, or one that passed condition (i)). The low and high thresholds are automatically generated and are image-dependent so, in our system, we used the median intensity of the gray-level image $\tilde{I}$ to generate the intervals of thresholds as follows:

$$low = max(0, ((1-\delta) \times \tilde{I})) \quad (4)$$

$$high = min(255, ((1+\delta) \times \tilde{I})) \quad (5)$$

where δ is used to vary the thresholds using a percentage of the median $\tilde{I}$. In our system, we used $\delta = 0.33$, a value that works well in many object detection applications [59].

After the above operations have been completed, we divide the resulting binary image into squared blocks of varying sizes. We choose squared blocks of 28×28, 64×64, 224×224, depending on the size of the input accepted by the SNNs used in our system. Then, we apply a one-level Discrete Wavelet Decomposition (DWT) using Daubechies family to calculating the energy E of each image patch by considering the horizontal detail (H), vertical detail (V), and diagonal detail (D) sub-bands of the DWT as follows:

$$E = \frac{\sum_{i=1}^{i=N}\sum_{j=1}^{j=N} H(i,j)^2}{M^2} + \frac{\sum_{i=1}^{i=N}\sum_{j=1}^{j=N} V(i,j)^2}{M^2} + \frac{\sum_{i=1}^{i=N}\sum_{j=1}^{j=N} D(i,j)^2}{M^2}, \quad (6)$$

where N is the number of values in the sub-bands of the DWT and M is the size of the squared patches.

We chose to compute the energy in the DWT domain because DWT sub-bands contain useful information about edges in different directions, as noted by several previous works [13], [27], [30], [36]. After the energy is calculated for all patches, we select only the n patches with the largest energy of each printed image to train and test the Siamese

networks. The areas with high energy contain enough information for printer source linking based on noisy/banding artifacts.

When we apply such a procedure to digital images, we also tend to find areas with strong sharp edges. However, for the case of printed documents, flat areas can also be selected. This happens because, by ranking the energy of binarized Canny filtered images, the algorithm searches areas with *many edges*, instead of extracting areas with the *sharpest edges*. This is useful when the noise of the printer is hidden in the background or flat areas. As the halftone printing patterns and printing imperfections from printers can introduce several new edges in the form of noise in the background, the proposed patch extractor can help in the forensic analysis by isolating such areas for further investigation. To exemplify such an effect, in Figure 3, we show the background of a digital and printed version of the same image. In the printed version, more halftoning and noisy artifacts are visible, which can be used for printer attribution.

(a) digital (b) printed

FIGURE 3. The same background portion in the digital and printed versions of the same image. Our patch extractor can, in specific cases, extract background patches as they contain some intrinsic artefacts useful for the printer source verification task.

B. APPLYING SIAMESE NETWORKS ARCHITECTURES TO HIGH ENERGY PATCHES

In our method, we take further advantage of SNNs characteristics by training them on a small training set. Inspired by previous works on steganalysis [60], [61] we propose to use shallow-but-wide architectures, where multiple and diverse filters are applied to the same outputs of the previous layers, with their final outputs fused and forwarded through the network. Specifically, we propose three novel SNNs architectures based on this idea as discussed in the following paragraphs.

1) MULTI-CONVOLUTION SUMMATION SNN (MCS−SNN)

Our first architecture is inspired to the the Residual Convolutional Neural network, also known as RESNET. Proposed by He *et al.* [62], RESNET uses a residual module consisting of two convolutional layers, where the output of the second layer is summed to the input of the first one through a bottleneck operation.

The MCS−SNN model is a shallow-but-wide SNN that uses *multiple convolutions* modules, where different feature maps are built in parallel, and further fused and forwarded to the rest of the network. Part of the topology of one sister MCS−SNN is illustrated in Figure 4. Every input is

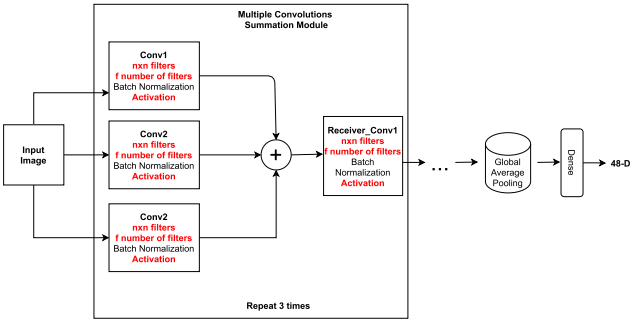


FIGURE 4. Pipeline of one Siamese network with MCS−SNN architecture. The parameters in red are modified in order to create similar SNNs with different performances while parameters in black are fixed. After the summation of the feature maps, batch normalization and max pooling are applied.

processed by three different convolutional layers, then the output maps are summed together and forwarded to an output convolutional layer. This procedure is repeated three times, and then a global average pooling is applied to build a pair of 48-dimensional vectors indicating the similarity or dissimilarity of a pair of printed image patches. Concerning RESNET, MCS−SNN is wide, simpler, and shallower. In addition, we neither apply bottleneck nor skip connections, making such a network more effective and fast for printer documents verification in data scarcity scenarios.

Similar to the other architectures that will be discussed later, the MCS−SNN is trained by applying the *Contrastive Loss* function. Such a loss is more effective than the common Binary Cross Entropy as the goal of SNNs is not to classify sample pairs, but to differentiate between them. Essentially, such a loss evaluates how well the SNNs are distinguishing image pairs using the following formula:

$$CL = Y_p * D_p^2 + (1 - Y_p) * max(m - D_p, 0)^2, \qquad (7)$$

where Y_p is the label of the training pairs images (0 if the pair comes from different printers and 1 otherwise) and D_p is the Euclidian Distance between the image pairs given by the 48-D output vectors of the SNNs. Finally, the margin m defines a radius around the sample space in such a way that dissimilar pairs only contribute to the loss function if D_p is within the margin. In particular, we set $m = 1$, so that only distances between 0 and 1 will contribute to the loss.

By training the networks with such a loss, pairs of images coming from the same printer, have lower distances, while images generated by different sources have large distances. Then, according to a threshold value T (we set it to 0.5), we verify if the pair of documents have been printed by the same printer or not.

We instantiated four different networks with this specific architecture, whose details are given below:

1) MCS−1: it contains the same number of filters (32) in all layers, but in the multiple convolution summation modules we apply 3×3, 5×5 and 7×7 convolutions.

The receiver convolutional layer has 3× 3 filters, and we apply the LeakyRELU [63] activation in all layers.

2) MCS−2: it contains the same number of filters (32) in all layers, but in the multiple convolution summation modules we apply 3× 3, 5 × 5 and 7 × 7 convolutions. In the multiple convolutions summation modules, we have different activations for each layer inside the multiple convolutions summation modules: LeakyRELU, RELU, and sigmoid. The receiver convolutional layers have 3× 3, 5 × 5 and 7 × 7 convolutions through the different modules, and we apply the LeakyRELU [63] activation to their outputs.
3) MCS−3: it is somewhat similar to MCS−1The differences are: (i) the receiver convolutional layers have 3× 3, 5 × 5 and 7 × 7 convolutions through the different modules, and they also have 32, 64, and 96 filters.
4) MCS−4: it contains the same number of filters (32) in all the multiple convolutions summation modules layers, and we apply 3× 3, 5 × 5 and 7 × 7 convolutions in every multiple convolutions summation module. However, the activation is fixed to be only LeakyRELU [63], RELU, or Sigmoid for each module. The receiver convolutional layers have 3× 3, 5 × 5 and 7 × 7 convolutions, and they also have 32, 64, and 96 filters, all of them activated by LeakyRELU [63].

All the networks (as all the others that will be presented in the remaining of this section) work on $64 \times 64 \times 3$ top-energy input patches, extracted as explained in Section III-A.

2) MULTI-CONVOLUTION INCEPTION SNN (MCI−SNN)

The second shallow but wide set of Siamese Networks is based on the Inception modules proposed by Szegedy *et al.* [64]. The SNN with Inception modules uses modules of parallel convolutional layers with filters of different sizes. The output feature maps are concatenated, as illustrated in Figure 5, and passed through the rest of the SNN. These SNNs are somewhat similar to the MCS−SNNs, with the difference of performing concatenation of feature maps instead of summation.

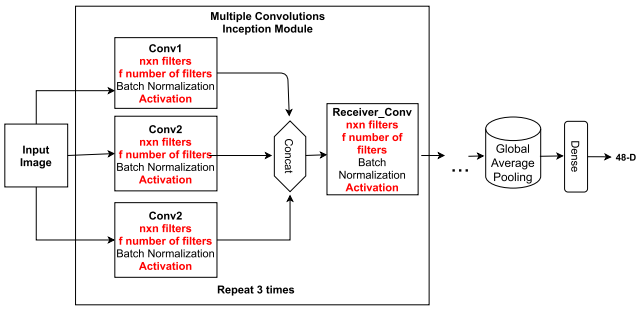


FIGURE 5. Pipeline of one siamese sister of the MCI−SNN class. Parameters in red can be modified to create an ensemble of similar SNNs while the parameters in black are fixed. After the inception operation, batch normalization and max pooling are applied.

As detailed below, we also created four networks with this specific architecture:

1) MCI−1: it contains the same number of filters (32) in all layers but in the multiple convolutions inception modules we apply 3× 3, 5 × 5 and 7 × 7 convolutions. The receiver convolutional layer has 3× 3 filters, and we apply the LeakyRELU [63] activation in all layers.
2) MCI−2: it contains the same number of filters (32) in all layers but in the multiple convolutions inception modules we apply 3× 3, 5 × 5 and 7 × 7 convolutions. In the multiple convolutions inception modules we also have different activations for each layer in a given multiple convolutions inception module: LeakyRELU [63], RELU, and sigmoid. The receiver convolutional layers have 3× 3, 5 × 5 and 7 × 7 convolutions through the different modules, and we apply the LeakyRELU activation to their outputs.
3) MCI−3: it is somewhat similar to MCI-1. The differences are: (i) the receiver convolutional layers have 3× 3, 5 × 5 and 7 × 7 convolutions through the different modules, and they also have 32, 64, and 96 filters.
4) MCI−4: it contains the same number of filters (32) in all the multiple convolutions inception modules layers, and we apply 3× 3, 5 × 5 and 7 × 7 convolutions in every module. However, we let the activation be LeakyRELU [63], RELU or Sigmoid. The receiver convolutional layers have 3× 3, 5 × 5, or 7 × 7 convolutions, and they also have 32, 64, and 96 filters, all of them activated by LeakyRELU [63].

3) SHALLOW XCeption SNN (SX−SNN)

The last set of Siamese Neural networks is inspired by the XCeption architecture [65]. Such architectures have in common point-wise and depth-wise convolutions. Firstly, ordinary 1×1 convolutions are performed, then channel-wise spatial convolution is applied (depth-wise convolution).

This architecture evolved from the INCEPTION−v3 [64] network with two differences: (i) no ReLU is applied after depth-wise (channel-wise) convolutions; and (ii) the operations in the XCeption architecture are applied in inverse order. Such adaptations have been found to be faster to train than than INCEPTION−v3 (having less parameters), and to provide a higher accuracy on Imagenet [65], [66].

Point-wise convolutions are a special kind of convolution, with a size of $1 \times 1 \times N$. 1×1 convolutions perform convolutions over the channels of an input image, an operation also called channel-wise pooling or feature map pooling. Such filters contain one single parameter, or weight, for each input channel and are then well-suited to summarize the input feature maps. After the point-wise convolutions, another convolution is applied to each channel of the feature map created previously. The result of such operations are added to the output of another ordinary convolution carried out in parallel through a bottleneck operation such as the one in RESNET [62] CNN.

Our XCeption-based Siamese networks follow the original pipeline of the XCeption network, which is composed of flows (or modules), which are pieces of networks with their

specific layers, bottlenecks, separable convolutions, max-pooling, activation, inputs, and outputs. One of these flows (the middle flow) is repeated several times, and in the end, all the flows are stacked in such a way to form a unique CNN. The main differences between XCeption and our architecture are (i) the size of the input images is $64 \times 64 \times 3$; and (ii) we control the depth of the SNN in our ensemble, by proposing several variations of shallow XCeptions as follows:

1) SX-1: one input flow, one middle flow, and one final flow.
2) SX-2: one input flow, three middle flows, and one final flow.
3) SX-3: one input flow, five middle flows, and one final flow.
4) SX-4: one input flow, seven middle flows, and one final flow.

As with the other SNNs, the output of the `SX-SNN` networks undergoes global average pooling, then the fully connected layers output a pair of 48-D of vectors, whose distances are minimized or maximized according to the labels of the input images. Training is carried out by adopting the contrastive loss shown in Equation 7.

C. ENSEMBLES OF DIVERSE AND HOMOGENEOUS SNNs

Ensembles is a popular approach in machine learning according to which several complementary models are applied trusting that even the weakest models can help to classify some samples that are commonly mistaken by the others. The power of ensembling has been validated in several works in diverse applications [67]–[70] in the literature.

In this paper, we ensemble different versions of the SNN architectures described in the previous section. In particular, we apply a two-fold majority voting fusion scheme, as illustrated in the rightmost part of Figure 2. In the first fusion stage (patch-wise fusion), we apply several SNN models to classify each pair of patches as coming from the same printer or not. The decision on each pair is made by looking at the most frequent class given by the SNN ensemble. The second round (document-wise fusion) considers the decisions made on all the patch pairs to decide if a pair of documents comes from the same printer or not. Such a fusion, also known as late fusion [71] has commonly been applied in several previous works in the printer source attribution literature [4], [8], [19], [35].

To enrich the ensembling analysis and performance, we investigate two different ensembling methods in the patch-wise classification. In the first one, which we call *Homogeneous* ensemble, we apply the same architecture (either *MCS*, *MCI*, or *SX*) with their unique parameters and topologies and evaluate the performance of the ensemble when they are combined three by three (we chose an odd number to avoid ties in the majority voting). By Evaluating hundreds of unique combinations of three classifiers with the same architecture but trained with different weight optimization methods, we found that that the best performance is achieved with the following combination of SNNs:

1) `MCS-2`: It contains the same number of filters (32) in all layers, but in the multiple convolutions summation modules we apply $3\times$ 3, 5×5 and 7×7 convolutions. In the multiple convolutions summation modules, we also have different activations for each layer in a given multiple convolutions summation module: LeakyRELU, RELU, and sigmoid. The receiver convolutional layers have $3\times$ 3, 5×5 and 7×7 convolutions through the different modules, and we apply the LeakyRELU activation on their outputs. Such an SNN is trained with the ADAMAX weight optimizer [72].
2) `MCS-3`: It contains the same number of filters (32) in all layers, but in the multiple convolutions summation modules we apply $3\times$ 3, 5×5 and 7×7 convolutions. The receiver convolutional layers have $3\times$ 3, 5×5 and 7×7 convolutions through the different modules, and they also have 32, 64, and 96 filters. Such an SNN is trained with the RMSPROP optimizer [73].
3) `MCS-4`: It contains the same number of filters (32) in all the multiple convolutions summation modules layers, and we apply $3\times$ 3, 5×5 and 7×7 convolutions in every module. However, we keep the activation as being only LeakyRELU, RELU, or Sigmoid for each module. The receiver convolutional layers have $3\times$ 3, 5×5 and 7×7 convolutions, and they also have 32, 64, and 96 filters, all of them activated by LeakyRELU. Such an SNN is trained with the NADAM optimizer [74].

Throughout the rest of the paper, we refer to the above ensemble as `HOM-ENS`.

In the second patch-wise verification ensemble method, we fuse the output of different SNN architectures. By selecting three unique models belonging to the *MCS*, *MCI*, and *SX* classes, we investigated a total of 4096 unique diverse ensembles with all possible combinations of weight optimization methods. Based on our experiments, we selected an ensemble containing the following models:

1) `MCS-3`: it contains the same number of filters (32) in all layers, but in the multiple convolutions summation modules we apply $3\times$ 3, 5×5 and 7×7 convolutions. The receiver convolutional layers have $3\times$ 3, 5×5 and 7×7 convolutions through the different modules, and they also have 32, 64, and 96 filters. Such an SNN is trained with the RMSPROP optimizer [73].
2) `MCI-3`: its an SNN similar to `MCS-3`, but with inception modules instead of summation operations. Such an SNN is trained with the ADAM optimizer [72].
3) `SX-4`: its a shallow-based XCeption architecture with seven blocks in the middle flow (instead of eight from the original architecture). Such an SNN is trained with the RMSPROP optimizer [73].

We name such a heterogeneous ensemble as `HET-ENS` as it consists of different sister Siamese networks specifically thought for printed documents verification.

TABLE 1. The first version of the VIPPrint dataset [58] used for the closed set experiments.

VIPPrint Dataset: Closed Set Experiments Printers					
ID	*Brand*	*Model*	*Resolution*	*Type*	*#Images*
#1	Epson	WorkForce WF-7715	4800 x 2400 dpi	Laser	200
#2	Kyocera	Color Laser	600 x 600 dpi	Laser	200
#3	Kyocera	TaskAlfa 3551	600 x 600 dpi	Laser	200
#4	Kyocera	TaskAlfa 3551	600 x 600 dpi	Laser	200
#5	Samsung	Multiexpress X3280NR	600 x 600 dpi	Laser	200
#6	HP	Color LaserJet Pro rfp-r479fdw	600 x 600 dpi	Laser	200
#7	HP	Color LaserJet rfp-r377dw	600 x 600 dpi	Laser	200
#8	OKI	C612 LaserColor	1200 x 600 dpi	Laser	200

IV. EXPERIMENTAL SETUP

To assess the effectiveness of the proposed SNNs and their ensembles for the problem of close and open set printed documents verification, we performed a series of experiments further explained in this section. In the following subsections, we discuss the dataset, methodology, baselines, and metrics considered.

A. DATASET

Our work has been validated on the second version of the VIPPrint dataset [58]. Such a dataset has a challenging nature as it contains pictures of human faces printed by several modern and professional color laser printers. We consider the original set of eight printers presented in such a dataset which are further described in Table 1. The first set of experiments aims at analyzing the behavior of Siamese Networks for printed documents verification in a closed set environment.

In the second set of experiments, we make one step further, addressing the capability of the proposed architectures for open set source linking. To this aim, we printed the same digital images of the VIPPrint dataset with four new printers, contained in a second version of the dataset [75], and we tested the performance of a system trained on the first printers on these new printers. The printers we used are professional and large-scale printers such as those used in print shops. From the analysis of Table 2, we can appreciate the difficulties presented by this new scenario: (i) there are printer brands that were not present in the training set (such as CANON and Ricoh printers); and (ii) new printing resolutions are also present which were not represented in the training set (*e.g.*, 1200 $\times$ 1200 *dpi* and 2400 $\times$ 2400 *dpi*). For the acquisition process, we used the same scanner, a TaskAlfa 3551 multi-functional printer scanner with 600 x 600 *dpi* resolution at the highest possible sharpness. The images are saved in a lossless compression configuration and used to feed and evaluate the accuracy of the proposed and baseline classifiers.

B. METHODOLOGY

To train the SNNs we selected 800 printed images (100 from each printer) and extracted the top-10 high energy patches from all of them (see Section III-A). For each block in the training set, we semi-randomly[2] selected another block from the same printer and another block from a different printer, thus building two pairs of similar and dissimilar blocks. In Figure 6, we show some examples of image patches coming from the same printer or different printers. This procedure results in about 16,000 pairs of images to train the SNNs. In a testing scenario, we used other 800 documents printed by the same printers, but containing different printed patterns (*i.e.,* different faces). Then, we randomly selected (with a fixed random seed as stated previously) 400 pairs of documents for testing. As in a real-world scenario, there will be more pairs of documents coming from different printers, so we investigate the closed set behavior considering 357 pairs of documents coming from different printers and 43 pairs of documents coming from the same printer.

TABLE 2. New printers in the second version of the VIPPrint dataset [75] used for the open set experiments.

VIPPrint Dataset: Open Set Experiments Printers					
ID	*Brand*	*Model*	*Resolution*	*Type*	*#Images*
#9	OKI	831	1200 x 600 dpi	Laser	200
#10	Ricoh	Aficio MP C3002	1200 x 1200 dpi	Laser	200
#11	CANON	C600	2400 x 2400 dpi	Laser	200
#12	Kyocera	P5021 CDN	1200 x 1200 dpi	Laser	200

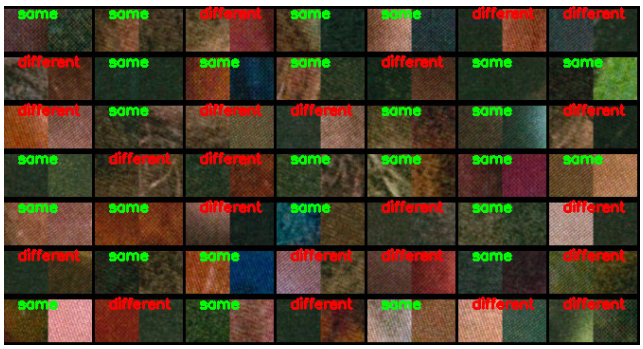


FIGURE 6. Some examples of top energy image patches coming from different (in green) and the same printer (in red). From the figure, it can be noticed that the SNNs learn specific printer artifacts for verification regardless of whether the patterns in the pair are similar or not.

[2]For sake of reproducibility, the same random seed is chosen so the same pairs of blocks are used for all SNNs that use the same input block size. All random operations reported in the remaining of this paper also use the same methodology.

We opted for such a random choice of the pairs to avoid any bias coming from a manual selection.

In the open set scenario, we would like to analyze the behavior of the SNNs when facing pairs of documents coming from unknown printers. For such a task, we initially considered pairs of documents coming from the same unknown printers. In such a testing scenario, we printed the same 200 images in the original dataset with the four new unknown printers, obtaining 100 semi-random pairs of documents from the same unknown printers for models evaluation. In a second set experiment, we also select semi-random pairs of documents from all possible combinations of different printers, evaluating the probability that documents coming from different printers are judged as being printed by the same printer.

C. BASELINES

We compare our approach against nine baseline approaches, some of them based on Siamese Networks others relying on different technologies. The first baseline SNN is inspired by one CNN applied on the `MNIST` dataset [76]. This is the shallower network we have analyzed since it contains only two convolutional layers, followed by max pooling. Finally, we adopted a dropout training strategy after the max-pooling layer. This network, similarly to ours, is trained from scratch on a dataset of top-10 energy patches.

The second set of SNNs includes IMAGENET pretrained CNNs: (i) `RESNET-50` [62]; (ii) `EFFICIENTNET-B0` and `EFFICIENTNET-B7` [77], which are the shallower and deeper versions of the same network; (iii) `MOBILENET` [78]; (iv) `NASNET-MOBILE` [79]; and `DENSENET-201` [80]. This second set of approaches are initialized with IMAGENET weights, and fine tuned on top-10 energy patches with data augmentation affine image operations.

Finally, specifically for the closed set scenario, we consider a handcrafted feature approach, used together with K-Nearest Neighborhood classifiers. According to this approach, which we call `LBP-KNN`, pairs of test images are described through Local Binary Patterns [81], and features from each image are classified individually by a KNN classifier. The KNN classifier is previously trained with LBP features from the 8 closed set classes. If the features from a pair of images under investigation lie in the same cluster, we define them as coming from the same printer and different printers otherwise. We performed a grid search on the number of possible neighbors in the KNN classifier, selecting them on the set of possible neighbors.

D. METRICS

In this section, we present the metrics used to validate our approaches against the baselines. For a better understanding of the metrics shown in the following subsections and also the discussion present in the remaining of this paper, we define the *positive* class as the class with pairs of documents coming from the same printer, and the *negative* one as the class with documents printed by different printers.

1) SPECIFICITY

For binary classification problems, the *specificity*, also known as true negative rate, indicates the percentage of correctly classified negative samples and is calculated as:

$$SPECIFICITY = \frac{TN}{TN + FP}, \tag{8}$$

where TN (True Negatives) represents the number of samples correctly classified as negatives, and FP (False positives) is the number of negative samples wrongly labeled as positive. In our binary classification problem, the sensitivity metric measures how many pairs of documents created by different printers were correctly detected as such.

2) RECALL

For binary classification problems, the *recall*, also known as true positive rate, indicates the percentage of correctly classified positive samples and is calculated as

$$RECALL = \frac{TP}{TP + FN}, \tag{9}$$

where TP (True Positives) represents the number of samples correctly classified as positives, and FN (False Negatives) is the number of positive samples wrongly labeled as negative. In our binary classification problem, the Recall metric measures how many pairs of documents created by the same printer were correctly detected as such.

3) PRECISION

With this metric, we want to know how many samples classified as positive are indeed positive. That is

$$PRECISION = \frac{TP}{TP + FP}, \tag{10}$$

where FP (False Positives) is the number of samples incorrectly classified as being positive (from the same printer).

4) F-MEASURE

This is the most important metric for unbalanced problems and thus will be used to rank our results. The F-measure (F) calculates the harmonic mean of precision and recall as follows:

$$F = 2 \times \frac{Precision \times Recall}{Precision + Recall}, \tag{11}$$

such a metric, like all the others considered, is normalized between 0 and 1 (1 being the best value).

5) GEOMETRIC MEAN

In the binary classification case, Geometric Mean can combine Sensitivity and Specificity into a single score that balances both concepts in the following manner:

$$GMEAN = \sqrt{Sensitivity \times Recall}. \tag{12}$$

The Geometric mean combines the true negative rate and the true positive rate at one specific threshold of the SNNs

TABLE 3. Closed Set experiments details and results. Approaches are trained using the ordinary Binary Cross Entropy (BCE) or the Contrastive Loss (CL) and we report the losses with the best results. The comparative study uses metrics specifically thought of unbalanced problems, namely: (i) the Balanced Accuracy (BACC); (ii) G-Mean (GMEAN); and (iii) f-measure (F1). The performance of the proposed approaches are boldfaced while the best results are highlighted in yellow. Results are sorted according to the F1 measure.

APPROACH	*Test Results-Closed Set*									
	DETAILS				*METRICS*					
	Input Shape	*Pre-Trained*	*Loss*	*Data Augmentation*	*BACC*	*GMEAN*	*F1*	*Precision*	*Specificity*	*Recall*
HOM-ENS	**64X64X3**	**no**	**CL**	**no**	**0.98**	**0.98**	**0.86**	**0.75**	**0.96**	**1.00**
MCS-3	**64X64X3**	**no**	**CL**	**no**	**0.98**	**0.98**	**0.85**	**0.74**	**0.96**	**1.00**
HET-ENS	**64X64X3**	**no**	**CL**	**no**	**0.97**	**0.97**	**0.84**	**0.72**	**0.95**	**1.00**
MNIST [76]	28X28X1	no	CL	no	0.95	0.95	0.71	0.55	0.90	1.00
LBP+KNN-224 [82]	224X224X3	-	-	-	0.64	0.54	0.44	0.86	0.99	0.30
LBP+KNN-64 [82]	64X64X3	-	-	-	0.60	0.45	0.33	0.81	0.99	0.20
DENSENET-201 [81]	224X224X3	yes	BCE	yes	0.64	0.62	0.30	0.22	0.79	0.48
RESNET-50 [77]	224X224X3	yes	CL	yes	0.55	0.54	0.20	0.13	0.66	0.44
NASNET-MOBILE [80]	224X224X3	yes	BCE	yes	0.52	0.47	0.17	0.12	0.74	0.30
MOBILENET [79]	224X224X3	yes	BCE	yes	0.49	0.48	0.16	0.10	0.56	0.41
EFFICIENTNET-B0 [78]	224X224X3	yes	BCE	yes	0.46	0.35	0.10	0.08	0.77	0.16
EFFICIENTNET-B7 [78]	224X224X3	yes	BCE	yes	0.46	0.30	0.09	0.07	0.82	0.11

similarity score (in our case, we use 0.5 as threshold according to the contrastive loss discussed in Section III-B1). This formula has the beneficial property of averaging out both scores penalizing unbalanced pairs.

6) BALANCED ACCURACY

The balanced accuracy gives the quality of detection based on the mean performance on the positive and negatives classification, or

$$BACC = \frac{Specificity + Recall}{2}. \tag{13}$$

The balanced accuracy is the most recommended metric when considering unbalanced testing scenarios and was validated in several works in the literature where such an environment is found [82], [83].

V. EXPERIMENTS

In this section, we discuss the results we have got by validating our proposed method against the baselines detailed in the previous section. We first consider the closed set scenario, then we pass on the more challenging case of open set verification.

A. CLOSED SET PRINTED DOCUMENTS VERIFICATION

We start validating our approaches in a closed set scenario with the methodology explained in Section IV-B. From the results in Table 3, we see that deeper networks in a Siamese setup do not work properly for printer verification. In particular, we highlight the bad results of EFFICIENTNET-B0 and EFFICIENTNET-B7 for such a problem. Such networks were applied successfully in some digital image forensic works [84], [85], however, in the Siamese setup such networks provide bad performance due to their deep structure and complexity which would require very large training sets. Training printed documents data-driven models on large datasets, however, is a very expensive and time-consuming task, thus suggesting the use of fewer complex models.

Our second set of experiments included some SNNs of interest as they are less complex being suitable to run also on mobile devices: the MOBILENET and NASNET-MOBILE based SNNs. Both these networks do not show good results being designed and pre-trained for other applications. In Table 3 their results are similar or even worse than random classifiers, even though they have been pre-trained and their generalization capabilities were supposed to increase due to data augmentation. In the same regard, results in Table 3 show that IMAGENET-based classifiers that are better than random guessers are those that include specific modules such as the dense modules (DENSENET-201) and those at the basis of our MCS-based approaches (residual modules from RESNET-50). Results in Table 3 show that the best IMAGENET-based SNN does not help in the identification of the source of the printed documents, with a specificity of 0.79 and a recall equal to 0.48. Such results further prove the non-transferability of the IMAGENET dataset weights to the printer verification problem.

The results we got highlight the better results of non-pre-trained methods. Our specific LBP-inspired approaches used as baselines (LBP+KNN-224 and LBP+KNN-64) outperform the IMAGENET SNNs by a large margin, but with a performance that are still unsatisfactory. We found that using 224 $\times$ 224 blocks for this task is a little better, with an almost perfect specificity (0.99) and a small false alarm probability (0.86 precision). However, such an approach has the following drawbacks: (i) it is an easy approach to be attacked given its small recall; and (ii) given the fact that it uses nearest neighborhoods in a known number of clusters, it cannot be used for the open set case.

We finish the analysis of the baselines by discussing the results of the shallower and simpler methods used in the experiments. The MNIST SNN showed a better balance between precision and recall, resulting in the highest F1 measure (0.71) among the baselines. Such better performance comes from the fact that this SNN is trained from scratch on our dataset and because of its simplicity that makes it easier to train on a small dataset involved. However, such an approach also has an unacceptable false alarm rate (0.55 precision).

TABLE 4. The confusion matrix of the best individual proposed SNN approach (MCS-3).

	Different	Same
Different	342	15
Same	0	43

Finally, the top results in Table 3 highlight the promising performance of our shallow but wide SNNs and their ensembles. Our best individual SNN (MCS−3) shows an almost perfect accuracy of 0.98, classifying correctly all the documents from the same printer and misclassifying only 4% of documents from different printers. However, as the problem is highly unbalanced (there are more documents from different printers than from the same), such an approach shows the fourth better precision (0.74). For sake of completeness, in Table 4 we report the confusion matrix of MCS−3.

Some interesting findings were discovered by analyzing the semi-randomly generated pairs of testing documents and their misclassification/classification performance by the best proposed individual method. First, all of the misclassified documents come from the same pairs of printers: the Kyocera Color Laser (printer #2 in Table 1) and one individual sample of the Kyocera Task Alfa 3551 (printer #3 in Table 1). Such difficulty and discriminating such printers are in line with previous work in the same dataset [58]. Second, the semi-random pairs selected for testing contained five pairs of the same picture, but only one was incorrectly classified (the one that contained the pair of printers discussed before). Such problems are only partially alleviated with the ensembles containing twins Siamese networks (HOM−ENS) which minimized partially such a confusion of these two printers (exactly one document is now classified correctly), and no improvement was done with the diverse ensemble (HET−ENS). As such a scenario is easier and the individual best classifier got already an almost perfect result, the real importance of such ensembles will be further highlighted in the open set scenario discussed in the following subsections.

B. OPEN SET PRINTED DOCUMENTS VERIFICATION

We start the performance investigation in the open set scenario by considering the case of documents produced by the same printer, analyzing the recall or true positive rate. Then, we analyze the false alarm rate obtained by feeding the networks with documents coming from different printers. We show and discuss such results in the next subsections.

1) OPEN SET RECALL ANALYSIS

We start by evaluating in Table 5 the performance of some of the approaches presented previously when tested on *pairs coming from the same but unknown printers* (*i.e*, printers from Table 2).

The first issue we notice is the bad performance of SNNs based on IMAGENET, with the best result being achieved by RESNET−50. These networks face several difficulties when applied to the problem herein: (i) they were trained and structured for another kind of application; (ii) they are too deep for the amount of training data we can rely on; and (iii) they were never validated in an open set scenario. As we will see, these networks always predict any pair of documents as coming from different printers.

For the proposed methods, we see that the multiple convolution summation SNN in its third version (MCS−3) classifies almost perfectly all the pairs coming from the OKI (printer #9) and CANON (#11). The CANON printer is a professional printer usually employed by print shops, making these results very more interesting as it is very likely that this kind of printer is used by counterfeiters. The images coming from the other two printers were more difficult to classify: the KYOCERA P5021 (printer #12) is some months old, and the Ricoh is a brand that was not included in the training set, with a resolution (1200×1220) that was neither used during training. This last printer presented the hardest challenge for all the best approaches. The proposed ensembles without diversity (HOM−ENS) helped just a little, exactly as happened before in the closed set, improving the detection accuracy of Kyocera and RICOH.

For the last baseline, the results of MNIST seem to be the best one in this scenario. The minimum recall is 0.79 for the Ricoh printer. However, such a simple structure got only 0.01 better result than our approaches, and, as we will show later, MNIST approach suffers a lot when dealing with documents produced by different printers.

2) OPEN SET SPECIFICITY ANALYSIS

In this final set of experiments, we fed the various systems with documents produced by different printers. We did so by choosing all possible pairs of the four new printers in Table 2.

The results we got, reported in Table 6, confirm the biased behavior of IMAGENET-based SNNs. For this family of SNNs, the NASNET−MOBILE showed the best result of all approaches, with a 0.95 mean specificity, highlighting the fact that, according to its worse results on Table 5, this SNN tends to always say that any pair of documents come from different printers. All other related CNNs (DENSENET−201, RESNET−50,MOBILENET, EFFICIENTNET−B0 and EFFICIENTNET−B7) follow a similar path.

The results in Table 6 also show how the shallower MNIST SNNs can fail in classifying pairs of documents coming from different printers, having the second-worst mean specificity (0.41).

Finally, the results in Table 6 highlight even better the promising results allowed by the ensembles, since they boost significantly the results of the best individual MCS−3 SNN. In particular, the ensemble with diversity (HET−ENS) with three different SNN architectures achieves the best result for three pairs of printers, achieving a mean specificity of 0.94, with the smallest false alarm of 0.6. Such result improves by 11% the best individual MCS−3 SNN. The ensemble obtained with homogeneous networks (HOM−ENS) also outperforms by 5% the best individual approach, with a clear improvement for the *OKI − CANON* and *RICOH − KYOCERA* pairs.

TABLE 5. Open Set experiments results in terms of recall. Results are sorted by the mean recall, proposed approaches labels and results are boldfaced and the best results are highlighted in yellow.

APPROACH	Test Results-Open Set (same printer) Recall				
	OKI	RICOH	CANON	KYOCERA	MEAN
MNIST [76]	0.92	0.79	0.82	0.83	0.84
HOM-ENS	**0.98**	**0.75**	**0.99**	**0.62**	**0.83**
MCS-3	**0.98**	**0.74**	**0.99**	**0.61**	**0.83**
HET-ENS	**0.97**	**0.75**	**0.85**	**0.58**	**0.78**
RESNET-50 [77]	0.36	0.37	0.58	0.48	0.44
DENSENET-201 [81]	0.04	0.42	0.03	0.44	0.23
EFFICIENTNET-B0 [78]	0.35	0.13	0.14	0.04	0.16
NASNET-MOBILE [80]	0.48	0.01	0.04	0.04	0.14
EFFICIENTNET-B7 [78]	0.07	0.05	0.26	0.17	0.13
MOBILENET [79]	0.05	0.11	0.09	0.15	0.1

TABLE 6. Open Set results considering all possible pairs of unknown printers (specificity metric). Results are sorted by the mean specificity, proposed approaches labels and results are boldfaced with the best results highlighted in yellow.

APPROACH	Test Results-Open Set (different printers) Specificity						
	OKI-RICOH	OKI-CANON	OKI-KYOCERA	RICOH-CANON	RICOH-KYOCERA	CANON-KYOCERA	MEAN
NASNET-MOBILE [80]	0.97	0.86	0.99	0.98	0.96	0.94	0.95
HET-ENS	0.96	0.94	0.99	0.99	0.74	1.0	0.94
HOM-ENS	0.87	0.81	0.98	0.99	0.72	0.95	0.88
MCS-3	0.95	0.44	1.00	0.99	0.65	1.0	0.83
EFFICIENTNET-B0 [78]	0.60	0.63	0.66	0.77	0.87	0.83	0.72
EFFICIENTNET-B7 [78]	0.94	0.76	0.89	0.75	0.89	0.72	0.82
MOBILENET [79]	0.94	0.81	0.91	0.67	0.87	0.62	0.80
DENSENET-201 [81]	0.87	0.93	0.86	0.91	0.52	0.87	0.82
MNIST [76]	0.36	0.16	0.72	0.27	0.40	0.59	0.41
RESNET-50 [77]	0.62	0.41	0.47	0.41	0.52	0.39	0.47

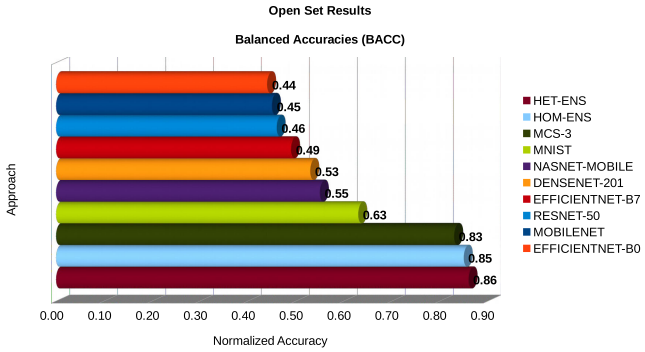


FIGURE 7. Summary of the performance of all the approaches in the open set case, by sorting balanced accuracies.

3) OPEN SET FINAL ANALYSIS

To give a better global view of the results we have got in the open set scenario, we calculated the Balanced Accuracies (BACC) using the previous specificities and recalls, sorting such performance metric in Figure 7. It can be seen that both proposed ensembles are better positioned among all the other approaches, with the MNIST being the best competitor. The fact that shallow networks are better positioned both in the open and close-set cases, is due to the data scarcity scenarios we face in this application.

In particular, it can be noticed that the ensembles, being them homogeneous (HOM−ENS) or heterogeneous (HET−ENS) outperform the best individual approach (MCS−3) by 2% and 3% respectively.

In general, the results discussed so far highlight the use of the diverse ensemble (HET−ENS), achieving the best final result considering both closed and open sets. In the closed set, the diverse ensemble showed a competitive performance with both individual MCS−3 and HOM−ENS. In the open set, it showed a reasonable 78% recall but a very good 94% specificity. In summary, the HOM−ENS showed to be the best when the printer is known, but it yields a 17% false acceptance and 12% false alarm. Our proposed HET−ENS has a similar final result of HOM−ENS (it is only 1% better), but its true power comes from minimizing the false alarms, having only 6% error when innocent pairs come to test being, thus, our best proposed method in general.

VI. CONCLUSION AND FUTURE WORK

The accessibility and popularity of printing and scanning devices is a double-edged sword: while they can generate increasingly good and reliable printed information such as currency, advertisements, attention signs, they can also be used to generate illegal content such as fake currency and counterfeited products. Several efforts have been made in the digital forensics literature to detect the source of unreliable printed documents, however, most of them have been tested on simple datasets (with common printers), use black-box approaches that require massive amounts of training data, and do not consider that printers outside the training dataset could be used by the counterfeiter.

In this paper, we propose to treat printer source attribution as a verification problem (instead of a classification problem).

By using siamese networks, we can make the best out of a small training dataset. With Siamese networks, documents coming from unknown printers can be evaluated, given that the networks learn similarities instead of labels. We propose to better exploit such networks by creating three shallow but wide siamese networks. We further propose four variations of each network and improve even more their verification accuracy by building ensembles with and without diversity. Our best approach using ensemble of heterogeneous Siamese Networks showed a promising mean 0.92 balanced accuracy when considering a challenging unbalanced closed set scenario and an open set scenario with four unknown printers at the same time, defeating other baseline approaches (both deep and shallow ones) by a large margin.

Notwithstanding these promising results, further work can be done to better exploit the verification scenario in the printer source attribution. First of all, the obvious search for more data and their effect on the verification task should be studied. Second, other losses, such as the triplet loss [86] could be used to improve the performance of the proposed networks. Finally, other shallow but wide architectures could be designed and their performance could also be evaluated to enrich the ensemble architecture.

REFERENCES

[1] R. McGregor. (2017). *Staffless Stores*. Accessed: May 17, 2020. [Online]. Available: https://www.lsnglobal.com/news/article/21083/staffless-stores

[2] *The Economic Impacts of Counterfeiting and Piracy*, Frontier Econ., London, U.K., 2016.

[3] *A Study on Public Health and Socioeconomic Impact of Substandard and Falsified Medical Products*, World Health Org., Geneva, Switzerland, 2017.

[4] G. N. Ali, A. K. Mikkilineni, E. J. Delp, J. P. Allebach, P.-J. Chiang, and G. T. Chiu, "Application of principal components analysis and Gaussian mixture models to printer identification," in *Proc. Int. Conf. Digit. Printing Technol.*, 2004, pp. 301–305.

[5] A. K. Mikkilineni, P.-J. Chiang, G. N. Ali, G. T.-C. Chiu, J. P. Allebach, and E. J. Delp, "Printer identification based on texture features," in *Proc. Int. Conf. Digit. Printing Technol.*, Dec. 2004, pp. 306–311.

[6] A. K. Mikkilineni, P.-J. Chiang, G. N. Ali, G. T. C. Chiu, J. P. Allebach, and E. J. Delp, III, "Printer identification based on graylevel co-occurrence features for security and forensic applications," *Proc. SPIE*, vol. 5681, pp. 430–440, Mar. 2005.

[7] A. K. Mikkilineni, O. Arslan, P.-J. Chiang, R. M. Kumontoy, J. P. Allebach, G. Chiu, and E. J. Delp, "Printer forensics using SVM techniques," in *Proc. Int. Conf. Digit. Printing Technol.*, Jan. 2005, pp. 223–226.

[8] E. Kee and H. Farid, "Printer profiling for forensics and ballistics," in *Proc. 10th ACM Workshop Multimedia Secur.* New York, NY, USA: Association for Computing Machinery, 2008, pp. 3–10.

[9] W. Deng, Q. Chen, F. Yuan, and Y. Yan, "Printer identification based on distance transform," in *Proc. Int. Conf. Intell. Netw. Intell. Syst.*, Nov. 2008, pp. 565–568.

[10] Y. Wu, X. Kong, X. You, and Y. Guo, "Printer forensics based on page document's geometric distortion," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 2909–2912.

[11] W. Jiang, A. T. S. Ho, H. Treharne, and Y. Q. Shi, "A novel multi-size block Benford's law scheme for printer identification," in *Advances in Multimedia Information Processing—PCM 2010*, G. Qiu, K. M. Lam, H. Kiya, X.-Y. Xue, C.-C. J. Kuo, and M. S. Lew, Eds. Berlin, Germany: Springer, 2010, pp. 643–652.

[12] A. K. Mikkilineni, N. Khanna, and E. J. Delp, "Forensic printer detection using intrinsic signatures," *Proc. SPIE*, vol. 7880, pp. 278–288, Feb. 2011.

[13] M. Tsai and J. Liu, "Digital forensics for printed source identification," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2013, pp. 2347–2350.

[14] S. Elkasrawi and F. Shafait, "Printer identification using supervised learning for document forgery detection," in *Proc. IAPR Int. Workshop Document Anal. Syst.*, Apr. 2014, pp. 146–150.

[15] M. J. Tsai, J. S. Yin, I. Yuadi, and J. Liu, "Digital forensics of printed source identification for Chinese characters," *Multimedia Tools Appl.*, vol. 73, no. 3, pp. 2129–2155, Dec. 2014.

[16] J. Hao, X. Kong, and S. Shang, "Printer identification using page geometric distortion on text lines," in *Proc. IEEE China Summit Int. Conf. Signal Inf. Process. (ChinaSIP)*, Jul. 2015, pp. 856–860.

[17] S. Shang, X. Kong, and X. You, "Document forgery detection using distortion mutation of geometric parameters in characters," *Proc. SPIE*, vol. 24, no. 2, pp. 1–10, 2015.

[18] M.-J. Tsai, C.-L. Hsu, J.-S. Yin, and I. Yuadi, "Japanese character based printed source identification," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2015, pp. 2800–2803.

[19] A. Ferreira, L. Bondi, L. Baroffio, P. Bestagini, J. Huang, J. A. D. Santos, S. Tubaro, and A. Rocha, "Data-driven feature characterization techniques for laser printer attribution," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 8, pp. 1860–1873, Aug. 2017.

[20] S. Joshi and N. Khanna, "Single classifier-based passive system for source printer classification using local texture features," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 7, pp. 1603–1614, Jul. 2018.

[21] S. Joshi, M. Lomba, V. Goyal, and N. Khanna, "Augmented data and improved noise residual-based CNN for printer source identification," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 2002–2006.

[22] H. Jain, S. Joshi, G. Gupta, and N. Khanna, "Passive classification of source printer using text-line-level geometric distortion signatures from scanned images of printed documents," *Multimedia Tools Appl.*, vol. 79, nos. 11–12, pp. 7377–7400, Mar. 2020.

[23] S. Joshi and N. Khanna, "Source printer classification using printer specific local texture descriptor," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 160–171, 2020.

[24] G. N. Ali, A. K. Mikkilineni, P.-J. Chiang, J. P. Allebach, G. T. Chiu, and E. J. Delp, "Intrinsic and extrinsic signatures for information hiding and secure printing with electrophotographic devices," in *Proc. Int. Conf. Digit. Printing Technol.*, Sep. 2003, pp. 1–4.

[25] A. H. Eid, M. N. Ahmed, and E. E. Rippetoe, "EP printer jitter characterization using 2D Gabor filter and spectral analysis," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2008, pp. 1860–1863.

[26] O. Bulan, J. Mao, and G. Sharma, "Geometric distortion signatures for printer identification," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2009, pp. 1401–1404.

[27] J.-H. Choi, D.-H. Im, H.-Y. Lee, J.-T. Oh, J.-H. Ryu, and H.-K. Lee, "Color laser printer identification by analyzing statistical features on discrete wavelet transform," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 1505–1508.

[28] J.-H. Choi, H.-K. Lee, H.-Y. Lee, and Y.-H. Suh, "Color laser printer forensics with noise texture analysis," in *Proc. ACM Workshop Multimedia Secur.* New York, NY, USA: Association for Computing Machinery, 2010, pp. 19–24.

[29] S.-J. Ryu, H.-Y. Lee, D.-H. Im, J.-H. Choi, and H.-K. Lee, "Electrophotographic printer identification by halftone texture analysis," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2010, pp. 1846–1849.

[30] M.-J. Tsai, J. Liu, C.-S. Wang, and C.-H. Chuang, "Source color laser printer identification using discrete wavelet transform and feature selection algorithms," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2011, pp. 2633–2636.

[31] J.-H. Choi, H.-Y. Lee, and H.-K. Lee, "Color laser printer forensic based on noisy feature and support vector machine classifier," *Multimedia Tools Appl.*, vol. 67, no. 2, pp. 363–382, Nov. 2013.

[32] D. Kim and H.-K. Lee, "Color laser printer identification using photographed halftone images," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, 2014, pp. 795–799.

[33] H. Wu, X. Kong, and S. Shang, "A printer forensics method using halftone dot arrangement model," in *Proc. IEEE China Summit Int. Conf. Signal Inf. Process. (ChinaSIP)*, Jul. 2015, pp. 861–865.

[34] D. G. Kim and H.-K. Lee, "Colour laser printer identification using halftone texture fingerprint," *Electron. Lett.*, vol. 51, no. 13, pp. 981–983, 2015.

[35] A. Ferreira, L. C. Navarro, G. Pinheiro, J. A. D. Santos, and A. Rocha, "Laser printer attribution: Exploring new features and beyond," *Forensic Sci. Int.*, vol. 247, pp. 105–125, Feb. 2015.

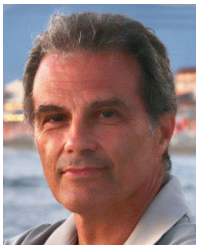
[36] M.-J. Tsai, M. Yuadi, Y.-H. Tao, and J.-S. Yin, "Source identification for printed documents," in *Proc. IEEE Int. Conf. Collaboration Internet Comput. (CIC)*, Oct. 2017, pp. 54–58.
[37] M. Bibi, A. Hamid, M. Moetesum, and I. Siddiqi, "Document forgery detection using printer source identification—A text-independent approach," in *Proc. Int. Conf. Document Anal. Recognit. Workshops (ICDARW)*, vol. 8, Sep. 2019, pp. 7–12.
[38] A. Melvin. (2016). *The 10 Most Expensive Liquids in the World*. Accessed: May 17, 2020. [Online]. Available: https://beyondtype1.org/the-10-most-expensive-liquids-in-the-world/
[39] R. Guha, "A report on automatic face recognition: Traditional to modern deep learning techniques," in *Proc. Int. Conf. Converg. Technol. (I2CT)*, Apr. 2021, pp. 1–6.
[40] J. E. Girard, *Criminalistics: Forensic Science, Crime, and Terrorism*, 3rd ed. Burlington, MA, USA: Jones & Bartlett, 2013.
[41] N. Khanna, A. K. Mikkilineni, A. F. Martone, G. N. Ali, G. T.-C. Chiu, J. P. Allebach, and E. J. Delp, "A survey of forensic characterization methods for physical devices," *Digit. Invest.*, vol. 3, pp. 17–28, Sep. 2006.
[42] N. Khanna, A. K. Mikkilineni, G. T.-C. Chiu, J. P. Allebach, and E. J. Delp, "Survey of scanner and printer forensics at Purdue University," in *Computational Forensics*, S. N. Srihari and K. Franke, Eds. Berlin, Germany: Springer, 2008, pp. 22–34.
[43] P.-J. Chiang, N. Khanna, A. K. Mikkilineni, M. V. O. Segovia, S. Suh, J. P. Allebach, G. T.-C. Chiu, and E. J. Delp, "Printer and scanner forensics," *IEEE Signal Process. Mag.*, vol. 26, no. 2, pp. 72–83, Mar. 2009.
[44] M. U. Devi, C. R. Rao, and A. Agarwal, "A survey of image processing techniques for identification of printing technology in document forensic perspective," *Int. J. Comput. Appl.*, vol. 1, no. 1, pp. 9–15, 2010.
[45] J. Oliver and J. Chen, "Use of signature analysis to discriminate digital printing technologies," in *Proc. Int. Conf. Digit. Printing Technol.*, Jan. 2002, pp. 218–222.
[46] C. H. Lampert, L. Mei, and T. M. Breuel, "Printing technique classification for document counterfeit detection," in *Proc. Int. Conf. Comput. Intell. Secur.*, vol. 1, Guangzhou, China, 2006, pp. 639–644.
[47] C. Schulze, M. Schreyer, A. Stahl, and T. Breuel, "Using DCT features for printing technique and copy detection," in *Advances in Digital Forensics V*, G. Peterson and S. Shenoi, Eds. Berlin, Germany: Springer, 2009, pp. 95–106.
[48] M. Schreyer, C. Schulze, A. Stahl, and W. Effelsberg, "Intelligent printing technique recognition and photocopy detection for forensic document examination," in *Proc. Informat., Fachwissenschaftlicher Informat.-Kongress*, Bonn, Germany, Jan. 2009, pp. 39–42.
[49] A. Roy, B. Halder, and U. Garain, "Authentication of currency notes through printing technique verification," in *Proc. Indian Conf. Comput. Vis., Graph. Image Process. (ICVGIP)*. New York, NY, USA: Association for Computing Machinery, 2010, pp. 383–390.
[50] P. J. Chiang, J. P. Allebach, and G. T.-C. Chiu, "Extrinsic signature embedding and detection in electrophotographic halftoned images through exposure modulation," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 3, pp. 946–959, Sep. 2011.
[51] J. van Beusekom, F. Shafait, and T. M. Breuel, "Automatic authentication of color laser print-outs using machine identification codes," *Pattern Anal. Appl.*, vol. 16, no. 4, pp. 663–678, Nov. 2013.
[52] T. Richter, S. Escher, D. Schönfeld, and T. Strufe, "Forensic analysis and anonymisation of printed documents," in *Proc. ACM Workshop Inf. Hiding Multimedia Secur.* New York, NY, USA: Association for Computing Machinery, Jun. 2018, pp. 127–138.
[53] M. Kumar, S. Gupta, and N. Mohan, "A computational approach for printed document forensics using SURF and ORB features," *Soft Comput.*, vol. 24, no. 17, pp. 13197–13208, Sep. 2020.
[54] Q.-T. Nguyen, A. Mai, L. Chagas, and N. Reverdy-Bruas, "Microscopic printing analysis and application for classification of source printer," *Comput. Secur.*, vol. 108, Sep. 2021, Art. no. 102320.
[55] L. Torres, N. Monteiro, J. Oliveira, J. Arrais, and B. Ribeiro, "Exploring a Siamese neural network architecture for one-shot drug discovery," in *Proc. Int. Conf. Bioinf. Bioeng. (BIBE)*, Oct. 2020, pp. 168–175.
[56] M. E. Hossain, A. Islam, and M. S. Islam, "A proficient model to classify Bangladeshi bank notes for automatic vending machine using a tiny dataset with one-shot learning & Siamese networks," in *Proc. Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT)*, 2020, pp. 1–4.
[57] A. Ullah, K. Muhammad, K. Haydarov, I. U. Haq, M. Lee, and S. W. Baik, "One-shot learning for surveillance anomaly recognition using Siamese 3D CNN," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–8.
[58] A. Ferreira, E. Nowroozi, and M. Barni, "VIPPrint: Validating synthetic image detection and source linking methods on a large scale dataset of printed documents," *J. Imag.*, vol. 7, no. 3, p. 50, Mar. 2021.
[59] A. Rosebrock. (2015). *Zero-Parameter, Automatic Canny Edge Detection With Python and OpenCV*. [Online]. Available: https://www.pyimagesearch.com/2015/04/06/zero-parameter-automatic-canny-edge-detection-with-python-and-opencv/
[60] B. Li, W. Wei, A. Ferreira, and S. Tan, "ReST-Net: Diverse activation modules and parallel subnets-based CNN for spatial image steganalysis," *IEEE Signal Process. Lett.*, vol. 25, no. 5, pp. 650–654, May 2018.
[61] J. Fridrich and J. Kodovský, "Rich models for steganalysis of digital images," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 3, pp. 868–882, Jun. 2012.
[62] 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.
[63] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.
[64] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
[65] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1800–1807.
[66] *The Basic Model of Deep Learning-Xception*. Accessed: Jul. 7, 2021. [Online]. Available: https://www.programmersought.com/article/66246112992/
[67] F. A. Anifowose, J. Labadin, and A. Abdulraheem, "Ensemble machine learning: An untapped modeling paradigm for petroleum reservoir characterization," *J. Petroleum Sci. Eng.*, vol. 151, pp. 480–487, Mar. 2017.
[68] F. Anifowose, J. Labadin, and A. Abdulraheem, "Towards an improved ensemble learning model of artificial neural networks," in *Artificial Intelligence: Concepts, Methodologies, Tools, and Applications*. Hershey, PA, USA: IGI Global, Jan. 2017, ch. 4.
[69] S. Carta, A. Corriga, A. Ferreira, A. S. Podda, and D. R. Recupero, "A multi-layer and multi-ensemble stock trader using deep learning and deep reinforcement learning," *Int. J. Speech Technol.*, vol. 51, no. 2, pp. 889–905, Sep. 2020.
[70] A. Ferreira, H. Chen, B. Li, and J. Huang, "An inception-based data-driven ensemble approach to camera model identification," in *Proc. IEEE Int. Workshop Inf. Forensics Secur.*, Dec. 2018, pp. 1–7.
[71] G. Iyengar and H. J. Nock, "Discriminative model fusion for semantic concept detection and annotation in video," in *Proc. 11th ACM Int. Conf. Multimedia (MULTIMEDIA)*. New York, NY, USA: Association for Computing Machinery, 2003, pp. 255–258.
[72] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, Dec. 2014, pp. 1–15.
[73] Y. N. Dauphin, H. De Vries, and Y. Bengio, "Equilibrated adaptive learning rates for non-convex optimization," in *Proc. 28th Int. Conf. Neural Inf. Process. Syst. (NIPS)*, vol. 1. Cambridge, MA, USA: MIT Press, 2015, pp. 1504–1512.
[74] T. Dozat, "Incorporating Nesterov momentum into Adam," in *Proc. Int. Conf. Learn. Represent.*, 2016, pp. 2013–2016.
[75] A. Ferreira, E. Nowroozi, and M. Barni, "VIPPrint: A large scale dataset for colored printed documents authentication and source linking, version 1.0," Zenodo, Switzerland, Jan. 2021. [Online]. Available: https://zenodo.org/record/4454971#.YThjAvwzZH4, doi: 10.5281/zenodo.4454971.
[76] L. Deng, "The MNIST database of handwritten digit images for machine learning research," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 141–142, Nov. 2012.
[77] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," *Proc. Mach. Learn. Res.*, vol. 97, pp. 6105–6114, Jun. 2019.
[78] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*. [Online]. Available: http://arxiv.org/abs/1704.04861
[79] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8697–8710.
[80] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.

[81] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognit.*, vol. 29, no. 1, pp. 51–59, 1996.
[82] A. Ferreira, S. C. Felipussi, C. Alfaro, P. Fonseca, J. E. Vargas-Munoz, J. A. D. Santos, and A. Rocha, "Behavior knowledge space-based fusion for copy–move forgery detection," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4729–4742, Oct. 2016.
[83] A. Ferreira, S. C. Felipussi, R. Pires, S. Avila, G. Santos, J. Lambert, J. Huang, and A. Rocha, "Eyes in the skies: A data-driven fusion approach to identifying drug crops from remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4773–4786, Dec. 2019.
[84] A. A. Pokroy and A. D. Egorov, "EfficientNets for deepfake detection: Comparison of pretrained models," in *Proc. IEEE Conf. Russian Young Res. Electr. Electron. Eng. (ElConRus)*, Jan. 2021, pp. 598–600.
[85] Y. Yousfi, J. Butora, E. Khvedchenya, and J. Fridrich, "ImageNet pretrained CNNs for JPEG steganalysis," in *Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, Dec. 2020, pp. 1–6.
[86] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *J. Mach. Learn. Res.*, vol. 10, pp. 207–244, Jun. 2009.

ANSELMO FERREIRA (Member, IEEE) received the Ph.D. degree (Hons.) in computer science from State University of Campinas, Brazil, in 2016.

Since 2016, he has been a Postdoctoral Fellow in several research institutes and universities in Brazil, China, and Italy, researching and developing machine learning solutions focused on industry, forensics, and market applications. He currently holds a Marie Skowlodska Curie Postdoctoral Fellowship at the University of Siena, Siena, Italy, in the European Union Project PrintOut, researching and developing machine learning and computer vision solutions for printed document forensics and package anticounterfeiting. His main research interests include computer vision, multimedia forensics, and big data analysis. He is also an Elected Member of the IEEE Information Forensics Technical Committee, where he is part of the technical directions subcommittee, where he acts identifying future trends and directions in information forensics. He has also being working as a reviewer for dozens of conferences and journals.

NISCHAY PURNEKAR is currently pursuing the M.Sc. degree in electronics and telecommunication engineering with the Department of Information Engineering and Mathematics, University of Siena, Siena, Italy. He is also a part of the Research Project Printout, where he researches and develops machine learning solutions for anti-counterfeiting. His main research interests include computer vision-based anti-counterfeiting solutions, machine learning, and deep learning.



MAURO BARNI (Fellow, IEEE) received the Ph.D. degree in informatics and telecommunications from the University of Florence, in 1995. During the last two decades, he has been studying the application of image processing techniques to copyright protection and authentication of multimedia, and the possibility of processing encrypted signals without decrypting them. Lately, he has been working on theoretical and practical aspects of adversarial signal processing and machine learning. He is the author of about 350 articles with an H-number equal to 64 (Google Scholar), and holds five patents in the field of digital watermarking and image authentication. He was a recipient of the Individual Technical Achievement Award of EURASIP, in 2016. He was the Chairman of the IEEE Information Forensic and Security Technical Committee, from 2010 to 2011. He was the Technical Program Chair of ICASSP 2014. He was the Editor-in-Chief of the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, from 2015 to 2017. He was a Funding Editor of the *EURASIP Journal on Information Security*. He has been serving as an associate editor for many journals, including several IEEE TRANSACTIONS. He was appointed as a Distinguished Lecturer of IEEE SPS, for the years 2013–2014.