CDP-Sim: Similarity metric learning to identify the fake Copy Detection Patterns

Hédi Zeghidi, Carlos F Crispim-Junior, Iuliia Tkachenko

CDP-Sim: Similarity metric learning to identify the fake Copy Detection Patterns

Hédi Zeghidi, Carlos Crispim-Junior and Iuliia Tkachenko
Univ Lyon, Univ Lyon 2, CNRS, INSA Lyon, UCBL, LIRIS, UMR 5205, F-69676 Bron, France
carlos.crispim-junior@liris.cnrs.fr, iuliia.tkachenko@liris.cnrs.fr

***Abstract*—Due to development and broad availability of high-quality printing and scanning devices, the number of counterfeited products and documents is dramatically increasing. Therefore, different security elements have been suggested to prevent this socioeconomic plague. One of the most promising and cheap solutions is the use of Copy Detection Pattern (CDP), a maximum entropy image, generated using a secret key. This pattern takes full advantage of information loss principle during printing-and-digitization process to detect copies. Such an unpredictable pattern is highly sensitive to distortions occurring inevitably during production (printing), verification (digitization) and reproduction (duplication) processes. Initially, the detection of counterfeited CDP was devoted to evaluating the level of information loss using Pearson correlation. However, the security of CDP based authentication system was shown to be vulnerable to estimation attacks based on neural network that can infer a CDP after scanning. In this paper, we study how to increase the performance of a detector using a similarity metric learning approach.**

***Index Terms*—copy detection pattern, similarity metric learning, fake detection**

I. Introduction

The number of counterfeits increases each year due to the accessibility of editing software and printing and digitization devices. Counterfeits have a significant impact on our health or safety and due to growing number of fakes, people lose the trust on medicines and product quality. The majority of existing security elements (holograms, watermarks, sticky labels, specific inks) cannot be easily verified by the common users.

One of the promising solutions is the use of printable unclonable codes [10], [17] that has easy integration and verification processes. Copy Detection Pattern (CDP) [10] is the most commonly used printable unclonable code. It is a black-and-white maximum entropy image generated with a secret key, illustrated in Fig. 1.a.

(a) (b)

Fig. 1. Example of CDP from Indigo dataset [3]: a) an original random binary image before printing and b) its degraded version by P&D.

The security of CDP is based on the information loss principle [10], which applies to each printing and digitization process. The stochastic nature of Print-and-Digitization (P&D) process [9] impacts both the structure and the image quality of any CDP as illustrated in Fig. 1.b. Thanks to the stochastic nature of P&D process, the CDP was supposed to be unclonable. Nevertheless, recently high quality fakes, known as estimation attacks, were generated using neural networks [14], [19], [21]. The classical detectors based on Pearson correlation have become ineffective against estimation attack. It was shown that the detector can be improved using some pre-processing techniques [7]. However, more recent estimation attacks [3] are very precise and the known similarity metrics (Pearson correlation, Hamming distance and Jaccard metric) cannot separate them from original CDP, even with pre-processing techniques.

As shown in [11], the noise altering a CDP is difficult to characterize as each printer and scanner has its own characteristics [5], [8]. Therefore, in this paper, we explore the possibility to detect the fakes using the differences of printer signatures and the metric learning approach [6]. The contributions of this paper are the following:

- We propose the metric learning approach based on Siamese neural network to identify the printer used for CDP production. The idea behind this task is to extract printer forensic features that can precisely separate the original CDP (printed using a known printer) from counterfeits (printed by unknown printer).
- We explore the different configurations of input and outputs layers for similarity metric learning.
- We propose a similarity based detector that can efficiently distinguish the originals from seen and unseen fakes.

The rest of the paper is organized as follows. We start by describing the related work on estimation attacks and detectors in Section II. The proposed similarity metric learning approach (CDP-Sim), as well as the proposed detector are presented in Section II. The experimental results are shown in Section IV. Finally, in Section V we summarize our conclusions and discuss the future work.

II. Security aspects of CDP

In this section, we discuss the CDP security aspects focusing both on new challenges brought by the deep learning based attacks and on the proposed countermeasures to detect the fake CDP produced by these estimation attacks.

A. Estimation attacks

The CDP authentication systems based on Pearson correlation were robust against duplication attack and naive image processing (Otsu thresholding, unsharp masking) attacks [15], [16]. Nevertheless, the estimation attacks based on deep learning approach have shaken up the CDP based authentication systems. During the last 4 years a big amount of estimation attacks have been developed using different deep learning architectures. The first architectures used for estimation attacks were bottleneck deep neural network [14], selectional auto-encoders [20], super resolution generative adversarial network [19], among others. The majority of fakes produced by these estimations attacks can be detected using image processing techniques that improve the detector based on Pearson correlation values [7].

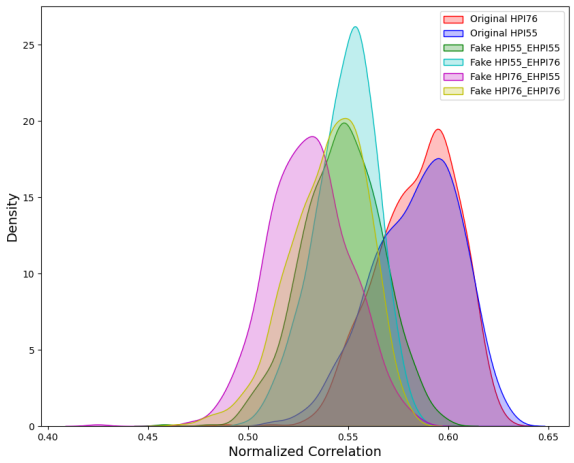

Fig. 2. Correlation values calculated to whole samples of Indigo dataset.

However, a novel publicly available dataset, named Indigo [3], was published recently. On this dataset, the classical similarity metrics underperform. Fig. 2 illustrates the distribution of normalized correlation values calculated between the template CDP and printed version of CDP (both originals and fakes). One can note that it is impossible to identify an authentication threshold that can efficiently separate the original CDP (red and blue curves in Fig. 2) from fake CDP (illustrated using magenta, yellow, cyan and green curves in Fig. 2). Therefore, it is urgent to find novel detectors to counter high quality fakes.

B. Authentication detectors

The high accessibility of deep learning models adds new challenges to CDP security system as the classical detectors that were used to identify the duplicated CDP [10] are inefficient against estimation attacks. Therefore, novel detectors were proposed [1], [12], [18].

In the first approach, the authors [12] proposed a machine learning based authentication system that uses only the CDP templates. It can identify the originals and accurately locate the anomalies in the fake CDP.

The second detector uses a mathematical model of P&D process based on local statistics to train a one-class classifier to identify the originals and reject the fakes [18]. The proposed detector works well while using for authentication uniquely the symbols with low probability of bit-error on the training set.

Another approach consists to simulate the P&D process using an encoder-decoder architecture [1]. The authors believe that these synthetic CDP generated using the proposed model can allow them to build a new classifier capable of detecting unseen fakes.

These results show us that the construction of powerful detector is a challenging problem that is still an open issue. In this paper, we want to explore another approach for detector construction trying to characterize the printer used to produce authentic CDP samples.

III. Proposed Authentication Approach

The studied pipeline is illustrated in Fig. 3. The manufacture produces a genuine physical object with CDP. Some printed samples are fed to the detector to learn the printer characteristics (the deep learning model and detector are detailed bellow). This detector is accessible to a final user to authenticate the physical object. The verification stage consists in digitization of physical object and calculation of the similarity between captured image and reference detector dataset (a set of images used as reference for similarity comparison).

All the forgeries are created between printing and digitization processes. An opponent has access to the printed genuine physical object. He should digitize the object and forge the CDP using estimation attack. After the forging process, the fake CDP is printed by opponent device and sent to the market. The forged and authentic physical objects will be analyzed by the same verification system. The aim of authentication detector is to identify the difference between genuine and fake physical objects and reject the fakes.

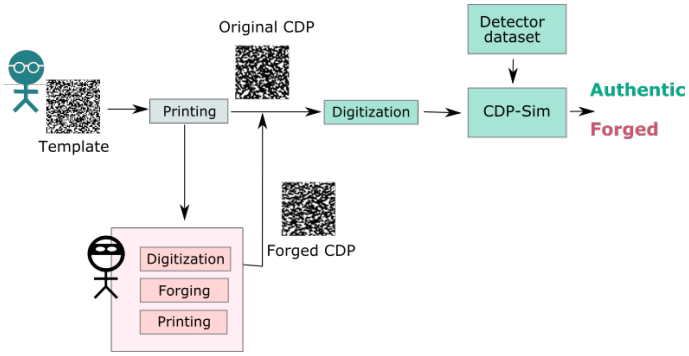

Fig. 3. Studied pipeline depicting an authentic and an opponent production channel.

In this paper, we propose CDP-Sim, a method to learn a non-linear, similarity metric learning distance that can separate original and fake examples of CDP. The proposed method is composed of a Siamese neural network (SNN) and a CDP

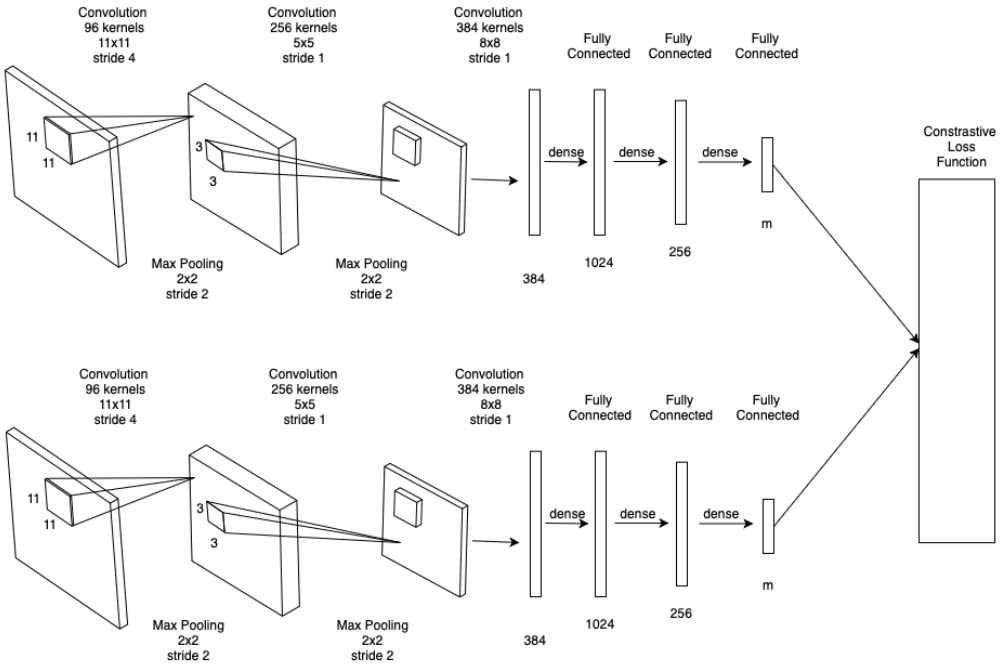

Fig. 4. Adopted Siamese Neural Network architecture [13].

binary detector. The SNN model [2], [4] learns to minimize the distance between original examples of CDP, and to maximize their distance with respect to fakes. The CDP binary detector uses the learned SNN module to classify CDP images into original and fakes, based on the measured distance between an unknown example and a set of reference, positive examples.

A. Siamese neural network

An SNN architecture can be viewed as two identical, parallel models sharing the same weight set [4]. The SNN models project each input into a feature vector of fixed length. The feature vectors can be then compared using a distance metric to measure their similarity using as a basis the projected space learned by the SNN. In practice, we may use the same neural network model for both inputs, hence enforcing the shared weights aspect. The CDP-SIM method follows the described approach and adopts the architecture [13] as its basis. The adopted architecture is the following (Fig. 4): The first convolution layer is composed of 96 squared kernels of 11x11 with a stride of 4. This layer is followed by a ReLU (Rectified Linear Unit) activation function and max pooling layer (kernel size of 2, stride of 2). The second convolution layer is composed of 256 squared kernels of size 5 and stride of 1, and it is also followed by a ReLU activation layer and max pooling layer with the same hyperparameters as the previous max pooling layer. The third convolution layer adopts 384 squared kernels of size 8, with a stride of 1, and it is also followed by a ReLU activation layer. The next layers are fully connected layers of dimensions 1024, 256, and m, respectively. The term m defines the size of the output feature vector. We study the influence of this therm in the experimental section. Fig. 4 illustrates the overall architecture of the deep neural network adopted and the dimensions of the output volume of each layer. We adopted the contrastive divergence loss [4], as the pair-wise loss function to train CDP-Sim.

B. Detector

The SNN architecture described measures the distance between two samples of CDP. In this section, we describe the detection algorithm we adopt to classify CDP examples in true or false class. The proposed detection algorithm is inspired on nearest neighbor algorithm. During training phase, we randomly choose N positive examples of CDP from our training dataset. The selected examples constitute the reference subset D.

During test time, we compute the distance between a tested example I and all examples $D_i, i = 1, \ldots N$ in the reference set using the learned SNN model. Each reference example classifies the new example as true/fake, using a pre-defined threshold th. The new example I is considered authentic if more than $N/2$ reference examples vote as such. Otherwise, the test CDP image I is rejected as a fake.

IV. Experimental results

In this section we will discuss the dataset preparation, the accuracy of the CDP-Sim as well as the detector results in the simulated real-world use case.

A. Dataset preparation

In this work, we use the most recent publicly available dataset - Indigo 1x1 [3], that contains originals and fake CDP patterns. The fake CDP were obtained using machine learning based estimation attack which has an accuracy score

of 94% [3].
Indigo dataset consists of 720 CDP templates of size 684×684 pixels that were printed by two printers (HP55 and HP76) and digitized by the same scanner. Then the printed CDP were used for estimation attack. The estimated CDP were printed again by the same printers, that gives 4 types of fakes (F55/55, F55/76, F76/55, F76/76).
In this work, we use only a subset from the Indigo dataset: 719 images per printer (HP55 and HP76), 719 fakes F55/55 and 719 fakes F55/76. This subset is used for two use cases: known-fakes use case and real-world use case, as illustrated in Table I.
The known-fake use case simulates the situation when the manufacture has information about the originals and the fakes that could be found in the market. We will use this use case in Section IV-C to identify best configuration of the adopted architecture.
The real-world use case is used in the CDP-Sim detector evaluation. We simulate the scenario, when a manufacture train the detector using original CDP and possible fakes that s/he creates by himself. Nevertheless, the tests are done using the known and unknown fakes as in real life the manufacture cannot predict all the possible fake types.

Scenario	Training dataset		Testing dataset	
	Original	Fake	Original	Fake
Known-fakes use case	HP55	F55/55	HP55	F55/55
Real-world use case	HP55	F55/55	HP 55	F55/55 F55/76 HP76

TABLE I
USED SUBSET OF INDIGO DATASET FOR TWO STUDIED SCENARIO.

The dataset is split into training (60%), validation (20%) and testing (20%) sets: 431 originals and 431 fakes for training, 144 originals and 144 fakes for validation and the same quantity for test.
To train the SNN model, we construct pairs of images to maximize the distance between original and fakes examples (dissimilar images), and to minimize the distance between the examples of the original class (similar images). We constructed the CDP pairs for training and validation using the following rules:

- the first image of the pair is always randomly picked from the original class,
- the second image is 50% of the time randomly selected from the class of originals (but cannot be the same CDP), and the 50% of the time from class of fakes.

For CDP-Sim accuracy evaluation (see Section IV-C), the CDP pairs used during testing were created following the same rules.
In the case of real-world use case (see Section IV-D), for each CDP image in the test set we create N pairs with images from the detector dataset.

B. Model training and evaluation

All model weights have been initialized randomly using Xavier uniform distribution. We used Adam algorithm for model weights optimization with a learning rate of 5×10^{-4}. All SNN models are trained for 100 epochs on the training set.
To evaluate the variation of model accuracy across training, we have used a fixed threshold, set to 0.5. By the end of the training we have computed the best threshold using the validation set, and defined it as the authentication threshold th.

C. SNN evaluation

To evaluate the proposed network, we have used the known-fake use case dataset. We have tested the network models with different input and output sizes to identify the best configuration for our problem. Table II presents the accuracy of the SNN models for different input and output sizes.

Output size	Input size		
	$171 \times 171 \times 16$	$684 \times 684 \times 3$	$171 \times 171 \times 3$
2×1	64%	**94%**	89%
32×1	61%	**95%**	90%
64×1	68%	**96%**	89%
128×1	62%	**91%**	86%

TABLE II
SNN ACCURACY FOR DIFFERENT INPUT AND OUTPUT SIZES.

One can note that the best accuracy is obtained while the input is of size $684 \times 684 \times 3$. We use this input configuration for the detector implementation. The results for different sizes of output are quite similar, thus, we have decided to keep the smallest output size.

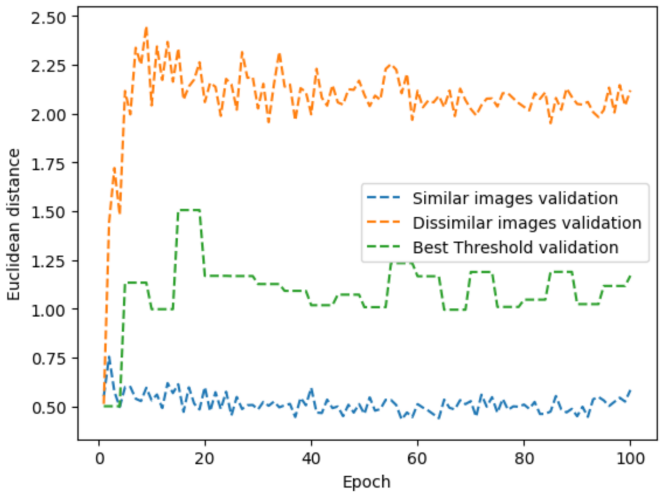

Fig. 5. Optimization of authentication threshold th in validation dataset.

The authentication threshold is defined on the validation set. We can note in Fig. 5 that the best authentication threshold for our dataset is nearly 1. We will use this authentication threshold value in the detector implementation.

D. Proposed detector

We evaluate the performance of the proposed detector on real-world use case. The detector is trained uniquely on originals (HP55) and fakes produced using these originals (F55/55).

This scenario simulates the situation where a manufacturer tries to produce the fakes using known estimation attacks, but using the known production devices (printer and scanner), which is the hardest case for detector. During the test phase, the CDP images may come from 4 different sets:

- Originals - Known HP Indigo printer (HP55)
- Fakes - Known fakes used to train the detector (F55/55)
- Fakes - Unknown fakes (F55/76)
- Fakes - Unknown HP Indigo printer (HP76)

In the studied scenario, the CDP printed using HP76 are considered as fakes, as they were printed using unknown (non-authorized) printer.

This real-world use case is a challenging task for the detector based on Pearson correlation. As shown in Fig. 6, the distributions of correlation values overlaps, what complicates the choice of an authentication threshold.

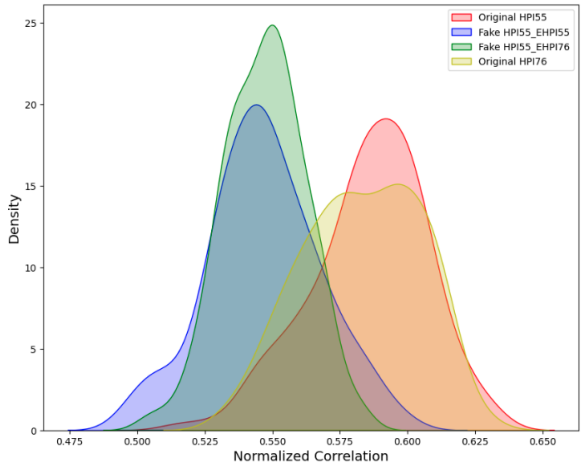

Fig. 6. Distribution of correlation values of 4 labels data set.

We have simulated the detector based on the correlation: the training and validation dataset were used to find the threshold that separates optimally the originals (HP55) from fakes (F55/55). Than, this threshold was used for test dataset to detect the originals and fakes. The confusion matrix in Table III illustrates the results of this state-of-the-art detector.

	Predicted	
Actual	Original	Fake
$HP55$	82%	18%
$F55/55$	17%	83%
$HP76$	76%	24%
$F55/76$	12%	88%

TABLE III
CONFUSION MATRIX OF SAMPLES AUTHENTICATION COMPUTED BASED ON PEARSON CORRELATION METRIC.

We note that this detector can identify fakes that come from estimation attacks, but it accepts nearly 15% of fake samples. We can also observe that the majority of CDP printed by unknown printer are detected as original CDP, even though we would expect them to be detected as fakes. This observation indicates that if the opponent guesses the CDP structure, the fake codes will be accepted as originals. We can note that 18% of originals (HP55) are considered as fakes. This value of false positives is quite high for a CDP detector.

To address this limitation we propose to use the CDP-Sim detector to better separate the seen originals (printed using known printer) from unseen originals and fakes. The training process was performed as described in Section IV-A. In test phase, we evaluated three setups with $N = \{5, 15, 25\}$. It is worth to mention that in all three setups the results are very similar. We have hence chosen to comment only on the results obtained for $N = 5$.

Fig. 7 illustrates the distribution of mean Euclidean distances calculated between the feature vector of CDP images from the detector dataset and the feature vector of test CDP samples. We can note an almost perfect separation between the originals (represent the similar pairs) and fakes (represent dissimilar pairs).

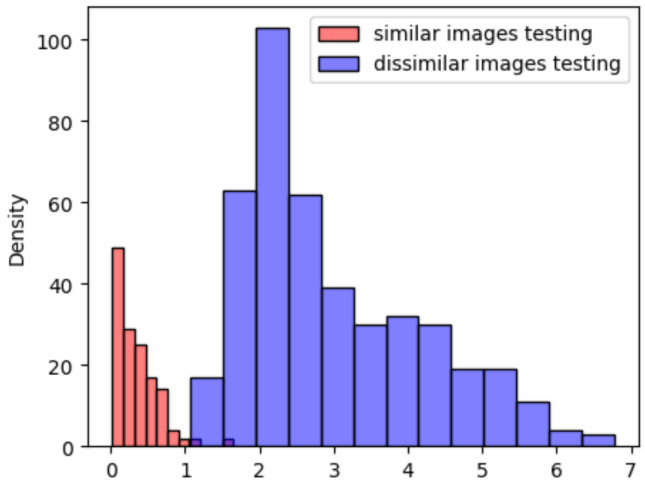

Fig. 7. Distribution of Euclidean distances on test dataset.

For the calculation of CDP-Sim accuracy, we have used the optimal authentication threshold defined on the validation set. In this particular test, the optimal authentication threshold is equal to $th = 1.009$.

	Predicted	
Actual	Original	Fake
$HP55$	97%	3%
$F55/55$	0%	100%
$HP76$	0%	100%
$F55/76$	0%	100%

TABLE IV
CONFUSION MATRIX OF DETECTOR RESULTS ON THE TEST SET.

The detection results are presented in Table IV. We note that the optimal threshold choice allows us to accurately separate the samples between original and fake classes. The accuracy of the proposed CDP-Sim detector is 99%, which means that the proposed detector can be a good solution to detect fakes created by estimation attacks. It is worth to mention that the

presented results are obtained for one particular use case.
The main drawback of the proposed solution is that we do not consider the CDP template in the authentication stage. By consequence, any CDP printed on the same printer will be considered as authentic by our CDP-Sim detector. Nevertheless, it is quite unreal that the opponent will have the possibility to print the fake CDP in the same printing company. Another direction of work will be to acquire another dataset that is larger and depicts real-world condition such as CDP captured by smartphones.

V. Conclusions

In this paper, we investigated a similarity metric learning approach on the printer forensic task to separate the profiles of authentic and counterfeited printing devices. We proposed a novel authentication method named CDP-Sim that is able to distinguish the fake CDP produced by high quality estimation attacks. We have simulated a real-world use case, and the proposed detector can successfully detect all the fakes produced by estimation attacks, and even CDP printed by an unknown printer.
In future work, we plan to evaluate the performance of proposed detector in other datasets and to consider the CDP templates in training stage of the proposed model to increase its robustness in less favorable setting. For instance, the situation where the forger has an access to the known printer.

Acknowledgment

This work was funded by the project *FakeNets* supported by the Computer Science Federation of Lyon (Féderation d'Informatique de Lyon).

References

[1] Y. Belousov, B. Pulfer, R. Chaban, J. Tutt, O. Taran, T. Holotyak, and S. Voloshynovskiy. Digital twins of physical printing-imaging channel. In *2022 IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 1–6. IEEE, 2022.

[2] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah. Signature verification using a "siamese" time delay neural network. In J. Cowan, G. Tesauro, and J. Alspector, editors, *Advances in Neural Information Processing Systems*, volume 6. Morgan-Kaufmann, 1993.

[3] R. Chaban, O. Taran, J. Tutt, T. Holotyak, S. Bonev, and S. Voloshynovskiy. Machine learning attack on copy detection patterns: are 1x1 patterns cloneable? In *2021 IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 1–6. IEEE, 2021.

[4] S. Duffner, C. Garcia, K. Idrissi, and A. Baskurt. Similarity Metric Learning. In *Multi-faceted Deep Learning - Models and Data*. 2021.

[5] A. Ferreira, L. Bondi, L. Baroffio, P. Bestagini, J. Huang, J. A Dos Santos, S. Tubaro, and A. Rocha. Data-driven feature characterization techniques for laser printer attribution. *IEEE Transactions on Information Forensics and Security*, 12(8):1860–1873, 2017.

[6] A. Ferreira, N. Purnekar, and M. Barni. Ensembling shallow siamese neural network architectures for printed documents verification in data-scarcity scenarios. *IEEE Access*, 9:133924–133939, 2021.

[7] E. Khermaza, I. Tkachenko, and J. Picard. Can copy detection patterns be copied? evaluating the performance of attacks and highlighting the role of the detector. In *2021 IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 1–6. IEEE, 2021.

[8] L. C Navarro, A. KW Navarro, A. Rocha, and R. Dahab. Connecting the dots: Toward accountable machine-learning printer attribution methods. *Journal of Visual Communication and Image Representation*, 53:257–272, 2018.

[9] A. T. Phan Ho, B. A. Mai Hoang, W. Sawaya, and P. Bas. Document authentication using graphical codes: Reliable performance analysis and channel optimization. *EURASIP Journal on Information Security*, 2014:1–17, 2014.

[10] J. Picard. Digital authentication with copy-detection patterns. In *Electronic Imaging 2004*, pages 176–183. International Society for Optics and Photonics, 2004.

[11] J. Picard. On the security of copy detectable images. In *NIP & Digital Fabrication Conference*, volume 2008, pages 796–798. Society for Imaging Science and Technology, 2008.

[12] B. Pulfer, Y. Belousov, J. Tutt, R. Chaban, O. Taran, T. Holotyak, and S. Voloshynovskiy. Anomaly localization for copy detection patterns through print estimations. In *2022 IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 1–6. IEEE, 2022.

[13] S. J. Rao, Y. Wang, and G. Cottrell. A deep siamese neural network learns the human-perceived similarity structure of facial expressions without explicit categories. *Cognitive Science*, 2016.

[14] O. Taran, S. Bonev, and S. Voloshynovskiy. Clonability of anti-counterfeiting printable graphical codes: a machine learning approach. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2482–2486. IEEE, 2019.

[15] I. Tkachenko and C. Destruel. Exploitation of redundancy for pattern estimation of copy-sensitive two level qr code. In *2018 IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 1–6. IEEE, 2018.

[16] I. Tkachenko, C. Destruel, O. Strauss, and W. Puech. Sensitivity of different correlation measures to print-and-scan process. In *Electronic Imaging*, number 7, pages 121–127, 2017.

[17] I. Tkachenko, W. Puech, C. Destruel, O. Strauss, J-M. Gaudin, and C. Guichard. Two-level QR code for private message sharing and document authentication. *IEEE Transactions on Information Forensics and Security*, 11(3):571–583, 2016.

[18] J. Tutt, O. Taran, R. Chaban, B. Pulfer, Y. Belousov, T. Holotyak, and S. Voloshynovskiy. Mathematical model of printing-imaging channel for blind detection of fake copy detection patterns. In *2022 IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 1–6. IEEE, 2022.

[19] R. Yadav, I. Tkachenko, A. Trémeau, and T. Fournel. Copy sensitive graphical code estimation: Physical vs numerical resolution. In *2019 IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 1–6. IEEE, 2019.

[20] R. Yadav, I. Tkachenko, A. Trémeau, and T. Fournel. Estimation of copy-sensitive codes using a neural approach. In *Proceedings of the ACM Workshop on Information Hiding and Multimedia Security*, pages 77–82, 2019.

[21] R. Yadav, I. Tkachenko, A. Trémeau, and T. Fournel. Estimation of copy-sensitive codes using a neuronal approach. In *ACM workshop on Information hiding and multimedia security*, Paris, France, June 2019.