2023-08-07

An Investigation into the Application of the Meijering Filter for Document Recapture Detection

John Magee
Technological University Dublin, b00149241@mytudublin.ie

Stephen Sheridan PhD
Technological University Dublin, stephen.sheridan@tudublin.ie

Christina Thorpe PhD
Technological University Dublin, christina.thorpe@tudublin.ie

An Investigation into the Application of the Meijering Filter for Document Recapture Detection

John Magee[1], Stephen Sheridan[1], and Christina Thorpe[1]

[1]School of Informatics and Cybersecurity, Technological University Dublin, Ireland

***Abstract*—The proliferation of mobile devices allows financial institutions to offer remote customer services, such as remote account opening. Bad actors can easily manipulate identity document images using modern image processing software. This represents a low-cost, high-risk threat to modern financial systems, opening these institutions to fraud through crimes related to identity theft. In this paper we describe our exploratory research into the application of biomedical imaging algorithm, the Meijering filter, to the domain of document recapture detection in the mobile channel. We perform a statistical analysis to compare different types of recaptured documents and train a support vector machine classifier on the raw histogram data generated using the Meijering filter. The results show that there is potential in biomedical imaging algorithms, such as the Meijering filter, as a form of texture analysis that help identify recaptured documents.**

***Index Terms*—Image processing, image filtering, identity document images, document recapture detection, biomedical imaging algorithms, the Meijering filter.**

I. Introduction

Traditional onboarding and Know Your Customer (KYC) channels for financial institutions are slow, inefficient, and costly [1]. As a result, remote customer onboarding services and electronic Know Your Customer (eKYC) services are seen as a viable alternative to traditional methods and help reduce costs and friction experienced by customers signing up for services. Financial institutions, including retail banking institutions, are increasingly including eKYC services within mobile banking apps, including the ability to open a new account remotely. The perception is that mobile devices are more secure than shared PCs and they allow for a more personal eKYC journey. However, the downside of such services is the ease at which a bad actor can use modern digital imaging software to manipulate identity documents, exposing financial institutions to fraud [2]. The consequences of fraud are significant; in 2022 the UK National Crime Agency reported that money laundering cost the UK economy in the region of "hundreds of billion pounds per year" [1].

The objective of this research is to develop a new method of identity document recapture detection that is explainable, simple to implement and transparent. As described in the next section, the state-of-the-art mechanisms rely on modern machine learning techniques such as Convolutional Neural Networks (CNNs). While CNNs have proven to be extremely useful in this domain, they lack transparency because it is not clear how they make decisions as their reasoning is learned, not programmed; this makes CNNs black boxes. Real world eKYC situations demand more than this, and GDPR explicitly demands that any automated AI system must be transparent and explainable [2]. CNNs, or any other complex neural network architecture cannot meet this legal requirement, therefore any solution based solely on this technology puts an onus on the institution to explain their decision making.

Motivated by the risk of identity theft and fraud can result in serious reputational damage to financial institutions, in this paper we introduce our exploratory research demonstrating the potential application of the biomedical imaging algorithm, the Meijering filter, to the domain of recaptured identity document detection.

The structure of this paper is as follows: Section 2 presents a brief overview of the related work in this area; Section 3 introduces the data set used in this research, the statistical analysis, the support vector machine classifier and an analysis of the classification results; Section 4 provides a discussion about the limitations of this exploratory research, a conclusion and the scope of future work. The main contribution of this work is to demonstrate that imaging filters developed in the biomedical domain have potential to be reused for new applications.

II. Related Work

A. Data Sets

The Mobile Identity Document Video (MIDV) data set is a collection of data sets released by Smart Engines [3] consisting of recaptured identity documents. Released in 2018, 2019 and 2020 respectively, these data sets are referred to as MIDV-500 [3], MIDV-2019 [4] and MIDV-2020 [5]. The main purpose of the MIDV data sets is to study the effects that mobile captures of identity documents have on document processing technologies such as Optical Character Recognition (OCR) and machine readable zone (MRZ) processing.

Kumar et al. [6] describe an algorithm that can produce synthetic passports, driver's licenses and Visa stickers and made 15,000 images available publicly. The goal of this

[1]https://www.parliament.uk/business/lords/media-centre/house-of-lords-media-notices/2022/november-2022/the-government-must-take-the-fight-to-the-fraudsters-by-slowing-down-faster-payments-and-prosecuting-corporates-for-failure-to-prevent-fraud/

[2]A good AI & GDPR primer is available at https://www.linklaters.com/en/insights/blogs/digilinks/ai-and-the-gdpr-regulating-the-minds-of-machines

[3]https://smartengines.com/

data set is to study the detection of forged character sets, a leading indicator of forged documents. Special character sets are used in official documents that are readily identifiable by humans, but not by machines. Work in forgery detection include Beusekom et al. [7] who developed a technique of forgery detection to check for text-line rotation and misalignments. Shang et al. [8] developed a forgery detection technique based on character distortions introduced due to printing on common printers. James et al. [9] approach the problem as a graph comparison problem. Kumar et al. have produced a very useful tool as it allows development of new synthetic data sets.

Soares et al. [10] publicly released the Brazilian Identity Document data set (BID) data set, consisting of synthetic identity documents generated based on templates from genuine documents. The purpose of this data set is to advance the domain of document segmentation and Optical Character Recognition (OCR) research. The researchers intentionally add variation to their data set to improve the robustness of any future work. The data set from Soares et al. is used as the basis for this research.

Some identity document data sets exist on Kaggle, such as the 'Identity Card' data set [4], the 'Identity Image' data set [5] and the 'identitydocuments' data set [6]. The 'Identity Card' data set contains 19 images of passports, but these are extremely low resolution of open source sample images. The 'Identity Image' data set does not contain any images of identity documents. The 'identitydocuments' data set contains images of identity documents, including many images from the BID data set [10].

The BID data set from Soares et al. [10] is used as the base data set in this research because it best represents the structure of an official identity document.

B. Document Recapture and Forgery Detection

Berenguel et al. [11] developed a system using image acquisition from a mobile device for counterfeit detection. Their research focused on Spanish identity documents and counterfeit bank notes. Identity documents are captured using a mobile device, cropped and uploaded for assessment. The disadvantage of their solution is that it requires model generation per document type, increasing complexity as the number of document types increase.

Yang et al. [12] developed a Convolutional Neural Network (CNN) based solution to detect recaptured images. Their research is not focused on recaptured identity documents, rather they reused images from existing data sets [13,14] consisting of general photographs. Their data set consists of 10,000 genuine and 10,000 recaptured images. All images are 512 x 512 pixels in size, which is considerably low resolution when compared to the capabilities of modern mobile cameras. Their contribution is to add a Laplician filter into the CNN as an enhancement layer and they report an classification accuracy of 99.74%.

[4]https://www.kaggle.com/datasets/omrastogi/identity-card-dataset
[5]https://www.kaggle.com/datasets/sdfasfas/identity-image-dataset
[6]https://www.kaggle.com/datasets/sbardelatti/identitydocuments

Berenguel et al. [15] proposed a Counterfeit Recurrent Comparator (CRC) network design to spot the difference between genuine identity documents and counterfeit documents. The CRC design is based on research on the human perception system [16]. In this work, the researchers reused a data set from their own work [17] that consists of identity documents and counterfeit bank notes. The document is split into 224 x 224 patches and the CRC network iterates over all the patches until the complete document is assessed. The CRC network performance is compared to that of PeleeNet [18] and they achieve a mean AUC score of 0.984, outperforming PeleeNet.

Chen et al. [19] proposed a Siamese network design to detect recaptured documents. The network components typically perform the feature extraction process, allowing for direct comparison of the extracted features using a loss function. They created a database of 320 captured and 2627 recaptured document images based on generated synthetic data. Chen et al. achieved 6.92% APCER and 8.51% BPCER by their proposed network.

C. The Meijering Filter

Introduced by Meijering et al. [20], this technique was designed to assist the analysis of neurite growth of fluorescence images. Conceptually, this can be considered a form of texture detection where the neurite is the textured object being detected against the background. Texture detection is a technique previously used for recapture detection [17,21]. The downside of these approaches is that they do not target identity documents specifically, or they rely on neural networks and hand-crafted feature extraction techniques for the best results. Hand crafted feature extraction techniques may overfit the training data and neural networks are black boxes as their reasoning is learned, not programmed, meaning their reasoning cannot be relied upon under GDPR. The Meijering filter is used in this research as a feature detection technique for recaptured documents. The premise is that the Meijering filter will be able to detect residual features due to the texture of the document substrate, in this case, printer paper.

III. Experiment Design

A. The Data Set

We introduce a new data set of recaptured images for use in this research that will be publicly released in parallel with this paper. This data set contains 104 images from the BID data set that are printed and recaptured using iPhone8 and iPhone12 mobile devices. Each document image was printed using a HP Envy Inkjet 4500 and Lexmark XC4140 laser printer and recaptured twice, once just as a raw paper print out and again as a paper print out with a plastic covering to mimic the surface of a genuine identity document. The BID data set contains document images in landscape and portrait form. When printed, the landscape images consume the entire A4 page whereas the portrait one consumes approximately half an A4 page i.e. A5. The size of the portrait images has the added benefit of being approximately the same size as passport documents when held open. For this reason, only portrait document images are selected for inclusion in this data

set. Documents are selected incremental based on the BID file naming scheme, starting at document 00000001. Some document images contain too much noise and are excluded from the data set, documents 0000053, 0000094 and 0000154 are good examples of this. This data set contains a total of 832 recaptured documents. The dimensions of each document image is 1344 by 848 pixels and they are in the JPEG format (JFIF). Throughout this paper, these recaptured images are collectively referred to as the printed recaptured images.

B. Screen Recaptured Documents

This research is at the exploratory phase and it is necessary at this point to have a ground truth data set as a reference in order to determine the applicability of our research and measure how well, or poorly, our ideas are performing. It was necessary to generate a synthetic data set to represent the ground truth because no ground truth data sets are available for reuse. The synthetic data set was created using the same BID images and mobile devices used in the creation of the data set described in Section III-A. The synthetic data set is generated from BID recaptured images displayed on a Dell monitor. While images recaptured on a screen are not the same as those of a real recaptured identity document, it aids this research because the surface texture of screen recaptured images will be different to printed recaptures, allowing this research to analyse the differences in surface texture and we believe that a similar effect would be seen in real recaptured identity documents. The images in the synthetic data set are collectively referred to in this paper as the screen recaptured images and, in the absence of any real data, act as a proxy for genuine recaptured documents in this research.

C. Data Processing

In this research work, Python scripts were used to process all of the recaptured images and the data processing pipeline is represented in Figure 1. The Meijering filter is used as the feature extraction process on all the recaptured images. The Meijering filter was applied to each image using the out-of-the-box implementation from Scikit-Image using only the default parameters [7]. The Meijering filter implementation can only be applied to a single channel, therefore images are loaded as greyscale images. After applying the Meijering filter to all the images in the data set, it was observed that screen recaptured images contained unique artifacts close to the edges of the image, indicating where the identity document image transitioned to the screen background. These artifacts, while interesting in their own right, are not what we expect to see in genuine recaptured documents, therefore all images generated by the Meijering filter are cropped to remove these artifacts by removing the bordering 50 pixels from each side of the image. An example of the border artifact is shown in Figure 2. All images are stored using the viridis colour space as this best highlighted the features extracted by the Meijering filter. Histograms are generated from the grayscale representations of the cropped images, as shown in part B of Figure 1. In this instance, the histogram represents the binned distribution of the grayscale intensity value in the image, the grayscale value ranging from 0 to 255. Each intensity value indicates the frequency of this value in the image. A bin value of 1 is used to provide the highest possible resolution for the histogram. Each histogram represents a distribution that will be used to measure the different between screen recaptured images and printed recaptured images. Image processing scripts are written in Python, version 3.8.5, Scikit-Image version 0.19.3 and OpenCV 4.4.0.42.

[7]https://scikit-image.org/docs/dev/api/skimage.filters.html

D. Chi Squared Test

The Chi Squared test is a statistical method used to measure differences in frequencies of categorical variables [22]. It also applies to binned distributions of categorical variables. A histogram is a distribution of binned intensity values, therefore the Chi Squared statistical test is used here to measure the frequency differences across each bin from the two distributions.

The Chi Squared test is used to compare the histograms of the screen recaptured and printed recaptured images. The histograms represent 256 individual intensity values for each image and the Chi Squared test is applied to these distributions directly. A Chi Squared test is performed for each screen recaptured document against each printed recapture, resulting in 416 Chi Squared test results. All 416 Chi Squared results are statistically significant at $p<0.001$ for each device.

E. Student t-test

The Student t-test is a parametric statistical method used to measure the difference of the means between two different distributions [22]. The main idea is that if the samples are from the same underlying group, then their mean values will be similar. Any deviation from this can indicate that their samples are not from the same underlying group.

Eight independent Student t-tests are used to compare the distribution of the median pixel values from the screen recaptured images against the printed recaptured images for each mobile device. In this way, the screen recaptured images are compared to the printed (inkjet and laser) and the plastic covered printed images (inkjet and laser) for each device. 104 screen median pixel values are compared to the 416 median pixel values for the recaptured images. The results of the Student t-tests show a statistically significant difference exists between the median pixel values for the screen recaptured images and the printed recaptured images at $p<0.001$.

F. Mann-Whitney U Test

The Mann-Whitney U statistical test is a non-parametric equivalent of the t-test, comparing the central tendency of different groups, assuming that each group is independent.

Eight Mann-Whitney U test were used to compare the median pixel values of the screen recaptured images against the printed recaptured images for the same mobile device. In this way, the screen recaptured images are compared to the printed (inkjet and laser) and the plastic covered printed images (inkjet and laser) for each device. 104 screen recaptured median pixel

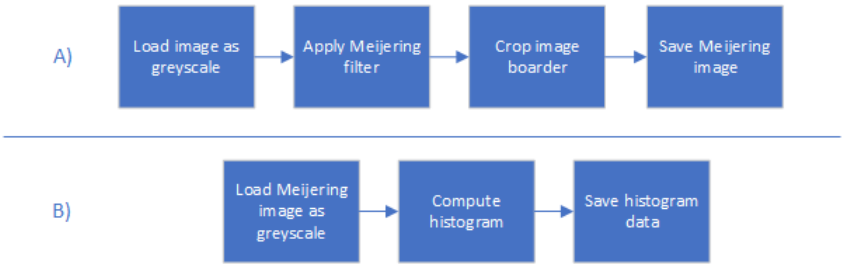

Fig. 1. Shows the image processing pipelines applied to all the images in the data set. Part A represents the initial data processing while part B represents the histogram data generation.

Fig. 2. Fig 2.A shows the boarder artifacts visible in screen recaptured images generated by the Meijering filter. Fig 2.B is the same image after cropping. Images are shown in the viridis colour space, which is the default for matplotlib.

values are compared to the 416 median pixel values for the recaptured images. The results of the Mann-whitney tests show a statistically significant difference exists between the median pixel values for the screen recaptured images and the printed recaptured images.

G. Building a Classifier

The statistical results in the previous section are encouraging. To add more confidence to these results, a machine learning classification model was trained to see how it would perform using the data generated from the Meijering filter. We decided to use a simple classification algorithm that minimises the number of hyper-parameters and for this reason, the Support Vector Machine (SVM) algorithm was chosen. Previous research has shown the SVM classifier to be useful in this domain [21,23]. Other benefits of the SVM algorithm is that of explanability and it is considered one of the best out of the box classifiers [24,25]. No effort was made to clean the data set before training.

The SVM classifier was trained to distinguish between the screen and printed recaptured images. The input features for the SVM are the histogram intensity values generated from the Meijering filtered images and each input is assigned a class label. The SVM is being used as a binary classifier, therefore a class value of 0 was used as a label for screen recaptured images and a class value of 1 was used as the label for printed recaptured images.

In order to generate sufficient classification accuracy scores, twenty different seeded tests were run, each using stratified 10 fold cross validation, resulting in 200 accuracy scores for each test. These results allow statistical analysis of the accuracy score distributions. The SVM was trained using Python 3.8.5 and scikit-learn 1.2.2.

H. Results

Across all the 200 tests for each recapture type, the iPhone8 produced a mean APCER 15.45% and mean BPCER 24.40% while the iPhone12 produced a mean APCER 29.35% and mean BPCER 24.05%. The iPhone8 produced a mean AUC 0.803 while the iPhone12 produced a mean AUC 0.732.

IV. Discussion

A. The Data Set

The BID data set [10] was used as the source data set in this research and it was chosen as it best represents an identity document data set. However, the data set is not without issues. This data set was released to help advance the domain of document segmentation and OCR and, as a result, the researchers intentionally added variation to the document structure and textual features. An example of the impact of this variation is seen in Figure 3.

This variation is not representative of a true identity document where consistency is a key security feature. They also intentionally censors the facial images on the documents that obscures legitimate variation that will exist in identity documents. Several documents in the BID data set are also of extremely low quality. Images 0000053, 0000094 and 0000154 are examples of this and are excluded from this research. Another added inconsistency is different coloured backgrounds, variation that is not representative of true identity documents. Despite this, due to the limited data sets in this

Fig. 3. Artifacts added to the BID data set to assist with document segmentation result in features extracted by the Meijering filter that would not exist otherwise. Image A is an example where the authors added bold borders around the biographic textual fields. Image B, obtained by applying the Meijering filter to image A, clearly shows that the artifacts added in Image A result in distinct bright lines where the bold borders were introduced.

domain, these limitations are accepted and discussed here to highlight the potential impact on the results of this research.

As part of this exploratory research, it was necessary to generate a data set to represent "genuine" document recaptures in order to measure the effectiveness of the Meijering filter as a feature extraction technique. This was prepared by recapturing the original BID documents from a computer monitor. This is problematic as these screen recaptures are obviously not genuine document recaptures, but given the lack of data in this domain, no other option existed. This is a common issue in this domain of research and needs to be taken into account when considering the results and how they can be applied to real-world scenarios. However, this was justified because after a manual evaluation, the interesting artifacts generated by the Meijering filter on paper document images were not present in the screen recaptures, further adding to the confidence that the Meijering filter can be used to detect document recaptures. However, it is noted that screen recapture introduces unique artifacts that had an impact on the SVM classification results and we cannot rule out that other screen recaptures would produce different results. More research is required to ascertain conclusively if the monitor recaptures aren't unduly influencing the results of this research.

V. Conclusion

The objective of this research was to show that medical imaging algorithms can be useful in the domain of identity document recapture detection. We have demonstrated that the Meijering filter can produce unique features present in printed recaptured identity documents that are not present in the same documents captured via screen capture. We then used a support vector machine, trained on the raw histogram data of the recaptured documents after applying the Meijering filter. The iPhone8 produced a mean APCER 15.45% and mean BPCER 24.40% while the iPhone12 produced a mean APCER 29.35% and mean BPCER 24.05%. Initial assessment is that these results do not compare well when compared to results published by Chen et al. The iPhone8 produced a mean AUC 0.803 while the iPhone12 produced a mean AUC 0.732 and this also compares poorly with the results published by Berenguel et al. After further consideration, taking into account the simplified approach to transform and classify the data, as well as the small test data set used in each test, these results demonstrate that document recapture detection using the Meijering filter is worth further investigation. We have shown that it is possible to distinguish between screen recaptured documents (a proxy for the genuine cohort) and physical printed documents, and this is very encouraging. The limited number of samples in the test data set also contribute to the error rates that seem high because a single mis-classification will result in a minimum error rate of 9%.

This exploratory research has demonstrated that the Meijering filter can be used successfully to identify recaptured documents under the right conditions, but further work is required to ascertain if these conditions can be extrapolated into real-world scenarios with real-world data.

Conflict of Interest

The authors have no conflict of interest.

Contributions

This collaborative research paper involved three authors: supervisors Dr Christina Thorpe and Dr Stephen Sheridan contributed equally to conceptualisation, methodology, analysis, writing, reviewing, and editing. John Magee played a significant role in dataset generation, conducting experiments, analysis and interpretation of data, and took the lead in writing the original draft.

References

[1] R. Soltani, U. Trang Nguyen, and A. An, "A new approach to client onboarding using self-sovereign identity and distributed ledger," in *2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*. IEEE, 2018, pp. 1129–1136.

[2] H. B. Macit and A. Koyun, "Tamper detection and recovery on RGB images," in *Artificial Intelligence and Applied Mathematics in Engineering Problems*, D. J. Hemanth and U. Kose, Eds. Springer International Publishing, 2020, pp. 972–981.

[3] V. V. Arlazarov, K. Bulatov, T. Chernov, and V. L. Arlazarov, "MIDV-500: A dataset for identity documents analysis and recognition on mobile devices in video stream," vol. 43, no. 5, 2019.

[4] K. Bulatov, D. Matalov, and V. V. Arlazarov, "MIDV-2019: Challenges of the modern mobile-based document OCR," p. 64.

[5] K. Bulatov, E. Emelianova, D. Tropin, N. Skoryukina, Y. Chernyshova, A. Sheshkus, S. Usilin, Z. Ming, J.-C. Burie, M. M. Luqman, and V. V. Arlazarov, "MIDV-2020: A comprehensive benchmark dataset for identity document analysis," vol. 46, no. 2, pp. 252–270, 2022.

[6] T. Kumar, M. Turab, S. Talpur, R. Brennan, and M. Bendechache, "Forged character detection datasets: Passports, driving licences and visa stickers," vol. 13, no. 2, pp. 21–35, 2022.
[7] J. van Beusekom, F. Shafait, and T. M. Breuel, "Text-line examination for document forgery detection," vol. 16, no. 2, pp. 189–207, 2013.
[8] S. Shang, X. Kong, and X. You, "Document forgery detection using distortion mutation of geometric parameters in characters," *Journal of Electronic Imaging*, vol. 24, 2015.
[9] H. James, O. Gupta, and D. Raviv, "Ocr graph features for manipulation detection in documents," *ArXiv*, vol. abs/2009.05158, 2020.
[10] D. S. Soares, R. B. Das Neves Junior, and B. L. D. Bezerra, "BID dataset: a challenge dataset for document processing tasks," in *Anais Estendidos da Conference on Graphics, Patterns and Images (SIBRAPI Estendido 2020)*. Sociedade Brasileira de Computação, pp. 143–146.
[11] A. Berenguel, O. R. Terrades, J. Llados, and C. Canero, "E-counterfeit: A mobile-server platform for document counterfeit detection," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2017, pp. 15–20.
[12] P. Yang, R. Ni, and Y. Zhao, "Recapture image forensics based on laplacian convolutional neural networks," in *Digital Forensics and Watermarking*, Y. Q. Shi, H. J. Kim, F. Perez-Gonzalez, and F. Liu, Eds. Springer International Publishing, 2017, vol. 10082, pp. 119–128, series Title: Lecture Notes in Computer Science.
[13] H. Cao and A. C. Kot, "Identification of recaptured photographs on LCD screens," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, pp. 1790–1793.
[14] R. Li, R. Ni, and Y. Zhao, "An effective detection method based on physical traits of recaptured images on LCD screens," in *Digital-Forensics and Watermarking*, Y.-Q. Shi, H. J. Kim, F. Pérez-González, and I. Echizen, Eds. Springer International Publishing, vol. 9569, pp. 107–116, series Title: Lecture Notes in Computer Science.
[15] A. Berenguel, O. Ramos Terrades, J. Llados Canet, and C. Canero Morales, "Recurrent comparator with attention models to detect counterfeit documents," in *2019 International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, pp. 1332–1337.
[16] P. Shyam, S. Gupta, and A. Dukkipati, "Attentive recurrent comparators," in *Proceedings of the 34th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, D. Precup and Y. W. Teh, Eds., vol. 70. PMLR, 06–11 Aug 2017, pp. 3173–3181.
[17] A. B. Centeno, O. R. Terrades, J. L. i. Canet, and C. C. Morales, "Evaluation of texture descriptors for validation of counterfeit documents," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, pp. 1237–1242.
[18] R. J. Wang, X. Li, and C. X. Ling, "Pelee: A real-time object detection system on mobile devices," in *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds., vol. 31. Curran Associates, Inc., 2018.
[19] C. Chen, S. Zhang, F. Lan, and J. Huang, "Domain-agnostic document authentication against practical recapturing attacks," vol. 17, pp. 2890–2905.
[20] E. Meijering, M. Jacob, J.-C. Sarria, P. Steiner, H. Hirling, and M. Unser, "Design and validation of a tool for neurite tracing and analysis in fluorescence microscopy images," vol. 58A, no. 2, pp. 167–176.
[21] X. Hou, T. Zhang, G. Xiong, Y. Zhang, and X. Ping, "Image resampling detection based on texture classification," vol. 72, no. 2, pp. 1681–1708, 2014.
[22] A. P. Field, J. Miles, and Z. Field, *Discovering statistics using R*. Sage, 2012, OCLC: ocn760970657.
[23] V. Lohweg, J. L. Hoffmann, H. Dörksen, R. Hildebrand, E. Gillich, J. Hofmann, and J. Schaede, "Banknote authentication with mobile devices," 2013, p. 866507.
[24] C. M. Bishop, *Pattern recognition and machine learning*, ser. Information science and statistics. Springer, 2006.
[25] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: with Applications in R*, ser. Springer Texts in Statistics. Springer US, 2021.

John Magee, BSc, MSc graduated from the Dublin Institute of Technology in 1997 with a B.Sc. (Hons) in Applied Science (Computer Science and Software Engineering) and received a Masters Degree in Data Science from Technological University Dublin in 2021. John has worked as an IT professional since 1997 and is currently the Director of Professional Services for EMEA in Daon.

Stephen Sheridan, BSc, PhD is lecturer in TU Dublin's faculty of Digital and Data in the school of Informatics and Cybersecurity. His current research focuses on the application of Artificial Intelligence and Machine Learning to solve problems in the Cybersecurity domain and in particular how these techniques can be used to detect patterns in network traffic that are often a sign of malicious command and control activity.

Christina Thorpe, BSc, PhD graduated with a B.Sc. (Hons) in Computer Science from University College Dublin (UCD) in 2005 and a Ph.D. in Computer Science from UCD in 2011. She was a postdoctoral research fellow in the Performance Engineering Lab in UCD from 2011 - 2018. Dr Thorpe has lectured in Computer Science and Cybersecurity since 2011 and is currently the Head of Cybersecurity in Technological University Dublin.