FORGED CHARACTER DETECTION DATASETS: PASSPORTS, DRIVING LICENCES AND VISA STICKERS

Teerath Kumar[1], Muhammad Turab[2], Shahnawaz Talpur[2],
Rob Brennan[1] and Malika Bendechache[1]

[1]CRT AI and ADAPT, School of Computing, Dublin City University, Ireland
[2]Department of Computer Systems Engineering, Mehran University of Engineering and Technology, Jamshoro, Pakistan

ABSTRACT

Forged documents specifically passport, driving licence and VISA stickers are used for fraud purposes including robbery, theft and many more. So detecting forged characters from documents is a significantly important and challenging task in digital forensic imaging. Forged characters detection has two big challenges. First challenge is, data for forged characters detection is extremely difficult to get due to several reasons including limited access of data, unlabeled data or work is done on private data. Second challenge is, deep learning (DL) algorithms require labeled data, which poses a further challenge as getting labeled is tedious, time-consuming, expensive and requires domain expertise. To end these issues, in this paper we propose a novel algorithm, which generates the three datasets namely forged characters detection for passport (FCD-P), forged characters detection for driving licence (FCD-D) and forged characters detection for VISA stickers (FCD-V). To the best of our knowledge, we are the first to release these datasets. The proposed algorithm starts by reading plain document images, simulates forging simulation tasks on five different countries' passports, driving licences and VISA stickers. Then it keeps the bounding boxes as a track of the forged characters as a labeling process. Furthermore, considering the real world scenario, we performed the selected data augmentation accordingly. Regarding the stats of datasets, each dataset consists of 15000 images having size of 950 x 550 of each. For further research purpose we release our algorithm code [1] and, datasets i.e. FCD-P [2] , FCD-D [3] and FCD-V [4].

KEYWORDS

Character detection dataset, Deep learning forgery, Forged character detection

1. INTRODUCTION

Personal documents are very important as they have very critical information that could be used for identity theft or fraud. They can be forged very easily with online available tools [1]. Forged documents are used for illegal and fraudulent purposes including theft, robbery, terrorism [1]. Among those documents, passport, driving licence and VISA stickers are the daily use documents for personal identification, driving and traveling, respectively. DL algorithms have shown remarkable performance in a wide variety of domains [2, 3, 4, 5, 6, 7, 8], so these DL algorithms can forge documents very easily using convolutional neural networks from color, font size and

[1] https://github.com/turab45/Dataset-forged-characters-detection-on-driving-licences-and-passports
[2] https://www.kaggle.com/turabbajeer/forged-characters-detection-on-passports
[3] https://www.kaggle.com/turabbajeer/forged-characters-detection-on-driving-licence
[4] https://www.kaggle.com/turabbajeer/forged-character-detection-dataset-on-visa-cards

DOI: 10.5121/ijaia.2022.13202

font type prospective, but there are two key disadvantages - 1) they are computationally very expensive. 2) Finding the bounding box of the randomly forged character is extremely difficult. Apart from these issues, data is very private. Finding the dataset for forged character detection is a challenging task due to many reasons i.e., labeling issues, limited access and many more. To fill this gap, we propose an algorithm which generates the three datasets by simulating the character forgery task, using five different countries' passports, driving licences and VISA stickers. These datasets are prepared using selected data augmentations considering the real world scenarios, and these datasets can be used for forged characters detection tasks.

In [9], we proposed the forged character dataset generation algorithm and provided two open datasets. In this work, we extend the algorithm by exploring new data augmentation approaches including conventional data augmentations i.e. rotation, scaling etc, and different color space data augmentations i.e. blur, contrast, saturation and brightness, based on real world scenarios. Furthermore, we also introduce another dataset named forged characters detection for VISA stickers (FCD-V) for forged character detection simulation.

Previously several methods have been used for forged characters detection [9,10] from document plain text. Algorithm [10] automatically detects forged characters from document images by measuring distance between feature vectors of Hu moments. The algorithm calculates possible conception errors by considering principal inertia axis, horizontal axis and character size; further character is classified as real or fake based on the score system. The algorithm in [11] detects whether the character is real or fake with the help of geometric parameter distortion mutation, for a single character algorithm estimated distortion parameters based on translation and rotation distortion. This algorithm [12] detects characters from an ID card using a traditional image processing method consisting of four stages: pre-processing, text-area extraction, segmentation, and recognition. The work in [13] proposed a method for signature forgery detection with verification using convolution neural network (CNN) and it showed impressive performance. In [14], authors assume whenever a fraudster uses a toolbox to modify information in a document and imitates the font from other documents having the same font types. There are high chances of error in font type. Based on this clue, the author proposed a method for automatic forgery detection based on font features using Conditional Random Field as measurement of probability. Finally, based on the probability that a certain character is classified as real or fake by comparing its probability with neighbouring characters. The work in [15] proposed a novel approach for forgery detection based on classification using uniform local binary patterns (LBP) to get unique textures that are common in the forged area. Apart from that, it combines many descriptors from neighbouring areas to learn contextual information and then support vector machines (SVM) is used for classification of certain areas. The work in [16] proposed a dataset for forgery detection for payslips. The dataset has 477 corrupted payslips with 6000 characters forged. In this work [17], researchers proposed a new dataset of 1969 images of receipts having OCT results of each. A few characters are forged in the receipt for fraud detection purposes. The work [18] describes the ICPR2018 fraud detection contest and methods used for that contest by different participants are thoroughly discussed. In [19], the author proposed a novel method for exploring the relationship between positive and negative coffiereint of discrete cosine transform to know the effect of distortion caused by forging. For that purpose, first R,G and B channels are fused to generate RGBF (RGB fusion), then histogram is calculated and from histogram, feature vector is generated. Finally the feature vector is passed to the neural network for classification of altered text images. The work in [20] proposed a novel method based on Fourier spectrum-based for forged character detection from video images. The work in [21], author discussed image forgery detection techniques and furthermore two important techniques for forgery based on pixels are discussed. First author discussed copy-move forgery technique and then the second was fast-copy-move detection. The work in [22] proposed the novel method based on mis-alignment and mis-rotation of text in documents and the author assumes that mis-alignment and mis-rotation are the important clues for forged character detection. In this work [23], the author explored a new approach for handwritten signature verification exploiting the regularity of length and curvature of a signature. Content of the signature at different angles is exploited to create histogram and then histogram is passed to the classifier. In [24], author new approach for forged international mobile equipment identity (IMEI) for smart mobile devices using R, G and B color components. The work in [25] explored spectral density and variation of hand writing images through clustering using low and high frequency amplitude. The extracted features that are invariant to scaling and rotation, are passed to the neural network for classification. In [26], the author compares the forged Urdu signatures done by people who are unaware of Urdu and of specific religion and region. In this work [27], the author proposed a new approach for the authenticity verification of seal impression printed on documents based on seal overlay metric. This work [28] presents a new dataset for copy-move forgery detection and consists of 260 forged images. Each image set has an original image, two masked images and a forged image. Further to increase dataset size, 5 different data augmentations are applied including translation, scaling, rotation, distortion and combination of these. This work [29] investigated the handwriting forgery detection, forgery done by either human or machine. It was found that the reason for forged handwriting is traced or copied slowly, so forged handwriting seems crumpled. In [30] the author deals with freehand forgeries of signatures on bank checks. This work [31] proposed a novel approach for detection of forged characters by exploiting different inks using image processing. First document is segmented into objects. Each object has nine features based on RGB channels. Distance between close pairs is calculated using the root mean square. Then an anomaly detection test is performed to get outliers away. In all the above mentioned works, the datasets are private or algorithms are evaluated on private datasets or the work is not done for passport, driving licence and VISA stickers. To contribute to this gap in this research work, we propose a new algorithm that simulates the forged character for passports, driving licences and VISA stickers. Furthermore, we release three synthesis forgerd character datasets i.e. FCD–P FCD-D and FCD-V.

2. DESCRIPTION

Each released dataset consists of 15000 RGB images of dimension 900 x 550 each. First we get a plain background image of passport or driving licence or VISA sticker, first two are taken online of five different countries including Australia, Canada, Ireland, Pakistan and USA, as driving licence and passport images are shown in **figure 2** and **figure 3,** respectively, and VISA stickers are taken of five different countries, China, Canada, Japan, South Korea and USA, as shown in **figure 4**. The sources of the image acquisition are described in table 1. First, we remove unwanted text and images on the passport or driving licence using an online website to make those plain, then we apply the proposed algorithm. The proposed algorithm consists of 4 steps, as shown in **figure 1** and described in **algorithm 1**. 4 steps are described as following:

2.1. Fill the document with data

First we read a plain background image of passport or driving licence or VISA stickers using the PIL library in python, then read a csv data file, data [5] was taken from kaggle platform which consists of five attributes including first name, last name, email, gender and age. We get a single record from a data file, and adjust it on a plain background image.

[5] https://www.kaggle.com/avkash/5feature30kcsv/version/1 (accessed on 1/17/2022)

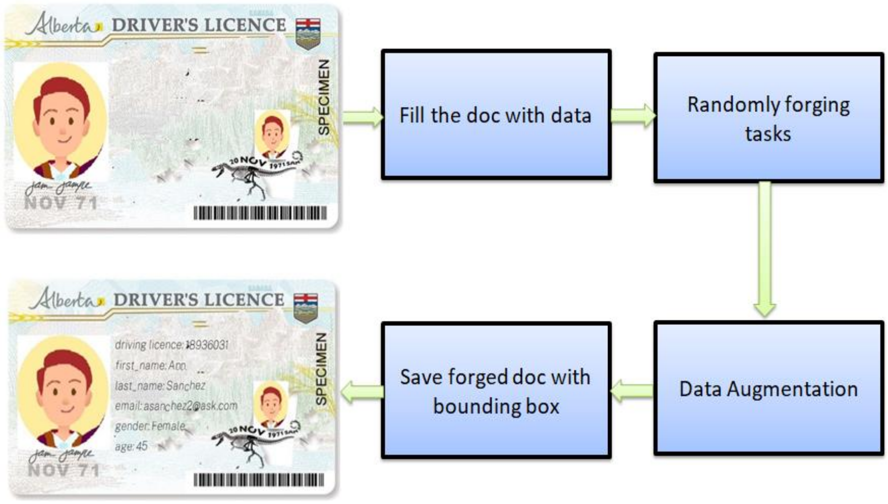

Figure 1. Proposed Algorithm Workflow

```
Forged_Character_for_document_generation(text_font, c_font, data, info):
    /*  Fill the doc with given data and return bouning boxes */
    1.  margin_top = 10 // margin top of the image
    2.  margin_bottom = 30 // margin bottom of the image
    3.  color = (0,0,0) // Color of the text
    4.  size = text_font.getSize(title_str) // Font size
    5.  x, y = (info[0] - size[0])/2, margin_top  // value of x and y axis to start printing text
    6.  dy= (info[1] - margin_top - margin_bottom - size[1] - 30)/4, margin_top
    7.  y = margin_top + size[1] + 110
    8.  for key,val in data.items() do
            a.  x = 10
            b.  size_key = text_font.getSize(key)
            c.  x = (info[0] - size[0])/2 - 180
            d.  draw text
            e.  x = x + size_key[0] + 10
       return positions
    9.  save_docs(font_styles, background, csv_file):
         /* Read record from given csv file then print on the background image with given font style and save it */
    10. data_aug(image, technique):
        /* Apply given data augmentation technique (Rotation, Scaling, Shearing and Translation) on the image
        and return it */
    11. color_aug(image, technique):
        /* Apply given color augmentation technique (Blur, Contrast, Saturation and Brightness) on the image and
        return it */
    12. save_tempered_docs (path):   /* Save the tempered doc in the given path */
```

Algorithm 1. Proposed algorithm

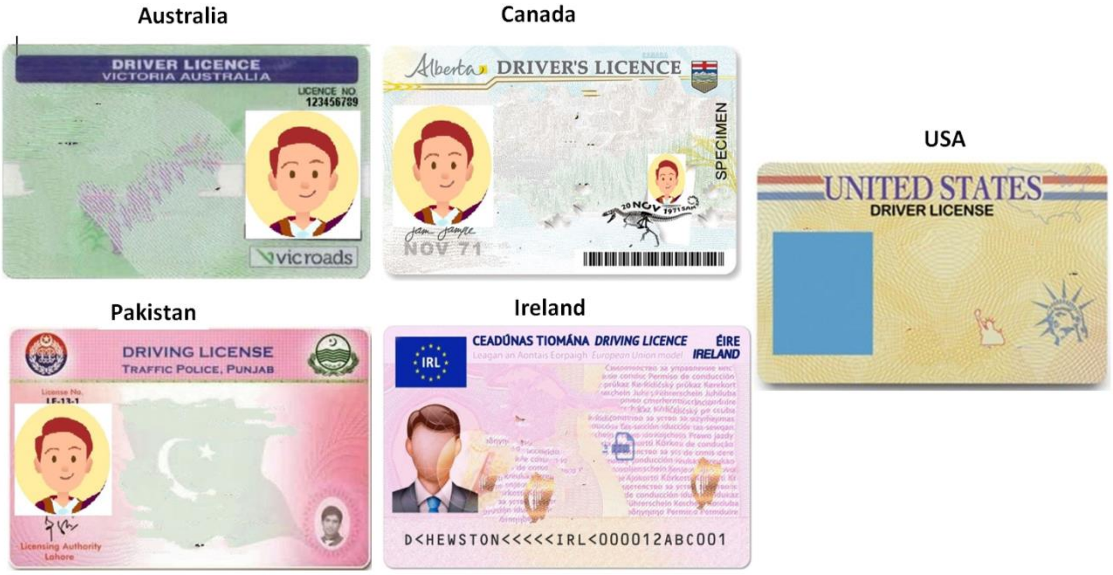

Figure 2. Plain Driving Licences

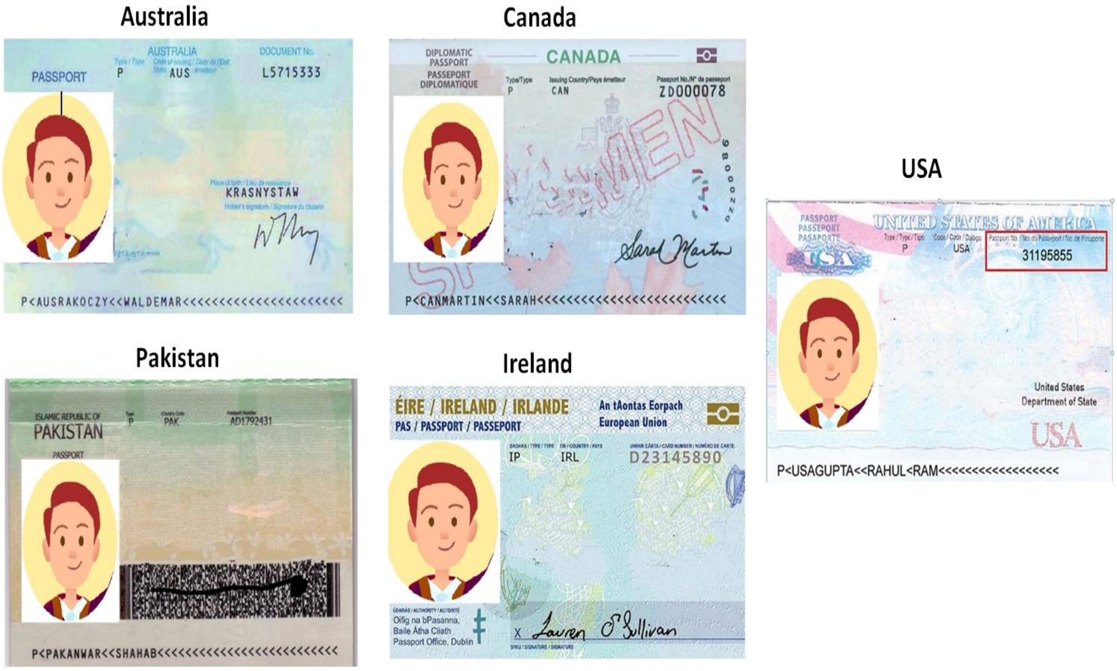

Figure 3. Plain Passports

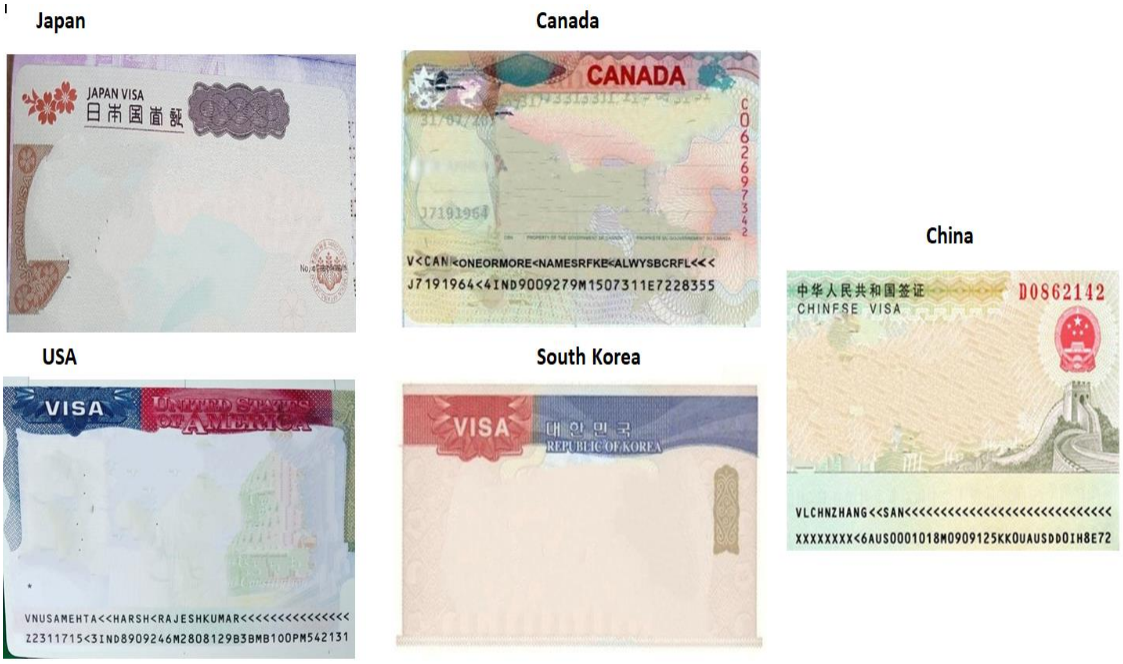
Japan
Canada
China
USA
South Korea
JAPAN VISA
CANADA
V<CAN<ONEORMORE<NAMESRFKE<ALWYSBCRFL<<<
J7191964<4IND9009279M1507311E7228355
中华人民共和国签证
CHINESE VISA
D0862142
VLCHNZHANG<<SAN<<<<<<<<<<<<<<<<<<<<<<<<<<<<
XXXXXXXX<6AUS0001018M0909125KKOUAUSDDOIH8E72
VISA
UNITED STATES OF AMERICA
VNUSAMEHTA<<HARSH<RAJESHKUMAR<<<<<<<<<<<<<<<
Z2311715<3IND8909246M2808129B3BMB100PM542131
VISA
대한민국
REPUBLIC OF KOREA

Figure 4. Plain VISA stickers

Table 1. Source of passports, driving licences and VISA stickers sample images

Passport	Australia	Canada	Ireland	Pakistan	USA
Driving licence	Australia	Canada	Ireland	Pakistan	USA
VISA Stickers	China	Canada	South Korea	Japan	USA

2.2. Randomly forging tasks

When documents are forged, there are two possibilities, either the forged character is not aligned with other characters properly [22] or the forged character has a little bit of noise in the background of the character area. While considering these possibilities, we randomly pick any character and change its location either up or down as shown in **figure 5 (A)** where in the email attribute the first ‘s’ of the census is moved down, or add a little bit of uniform noise in the character background as shown in **figure 5 (B)** where driving licence, first name, last name and email have a noise in one character of each.

2.3. Data augmentation

In the real world, documents are not always placed as straight as shown in the input column of **figure 6 and figure 7**, documents can be placed at any angle or stretch due to the camera settings. To meet the real world scenario, we perform two augmentations namely rotation and shearing for the driving licences and passports image as shown in columns rotation and shearing of each **figure 6 and figure 7**. The bounding boxes of forged characters are rotated using the below formula.

$$\begin{bmatrix} \mathit{alpha} & \mathit{beta} & (1\text{-}\mathit{alpha})*x\text{-}\mathit{beta}*y \\ \text{-}\mathit{beta} & \mathit{alpha} & \mathit{beta}*x+(1\text{-}\mathit{alpha})*y \end{bmatrix}$$

where
*alpha = scale * cos (theta)*
*beta = scale * sin (theta)*
and theta is the rotation angle
For this case, we use scale=1

Furthermore, for VISA stickers, we explored many other augmentations considering the real world scenarios i.e. lightning conditions, documents are not captured as a whole in the image etc. Conventional augmentations including rotation, shearing, translation as shown in **figure 8 (a)**, and color augmentations including brightness, saturation, blurriness and contrast as shown in **figure 8 (b)**, are explored.

2.4. Save forged document with bounding box

In this step we describe the motivation of recording the bounding box of forged character, as unlabeled data can be found online on a large amount, but to get the data labelled is extremely difficult, time consuming, tedious and expensive task [32]. To mitigate that issue, our proposed

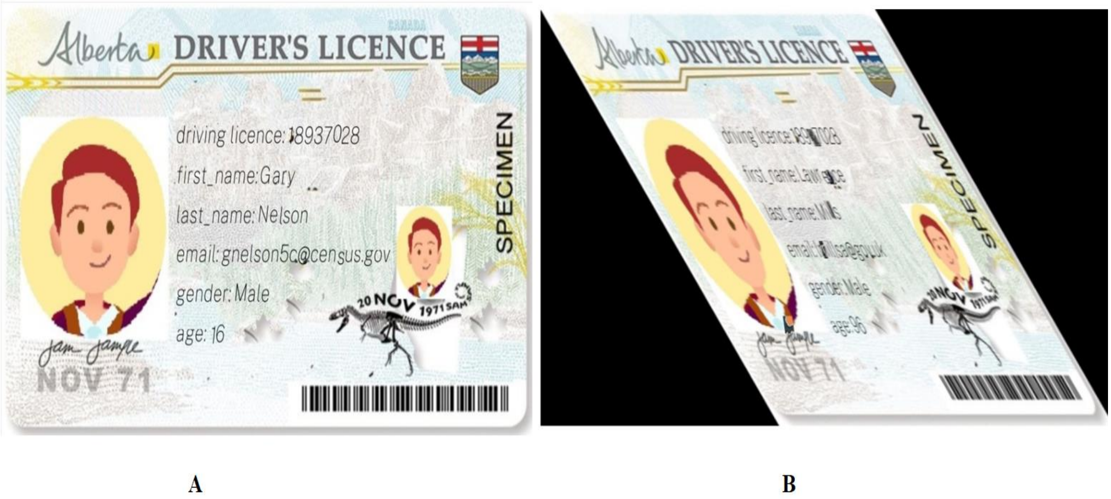

Figure 5. Forging Tasks

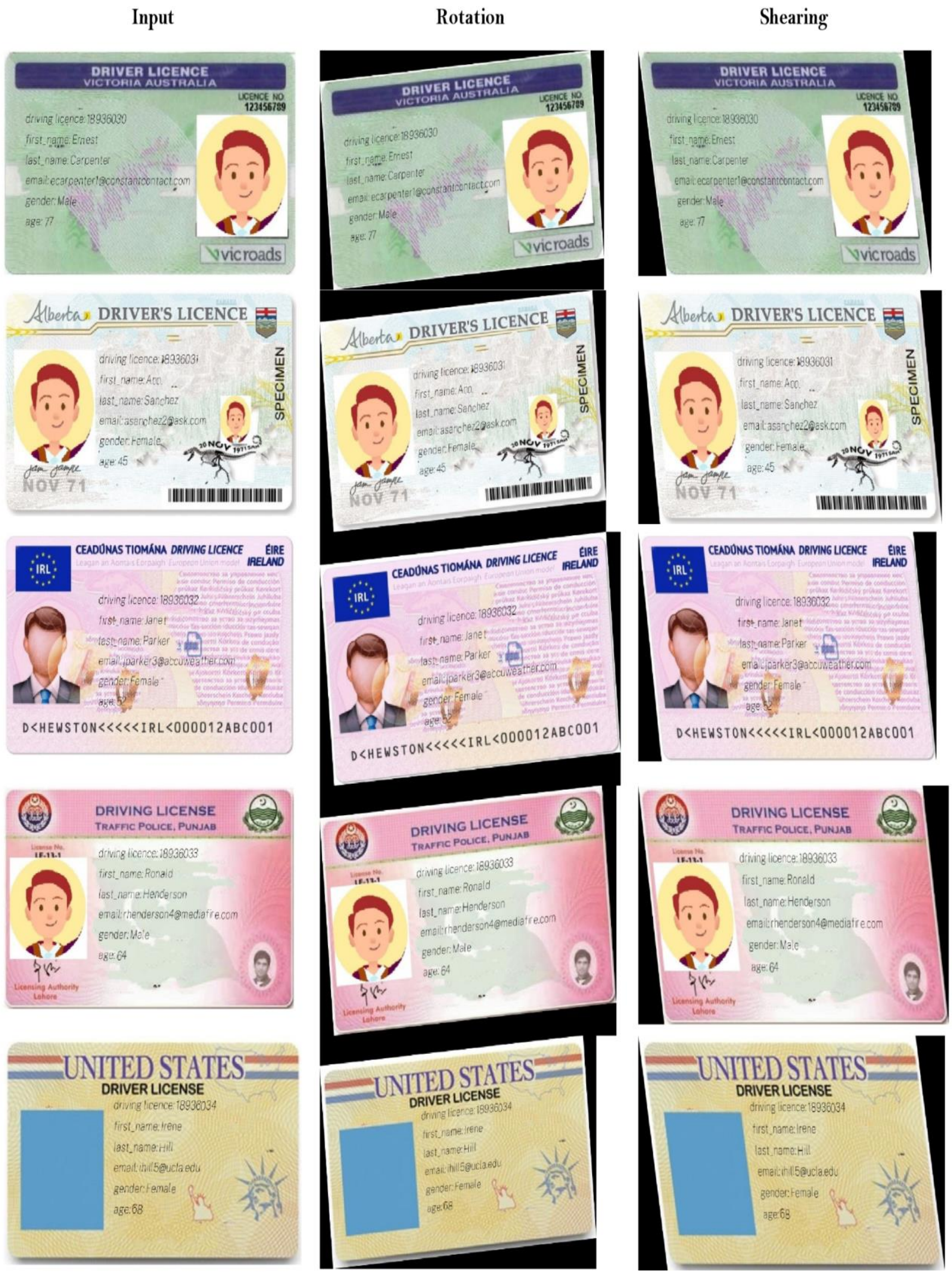

Figure 6. Driving licence with applied data augmentation Right to left, input, rotation augmentation and shearing augmentation.

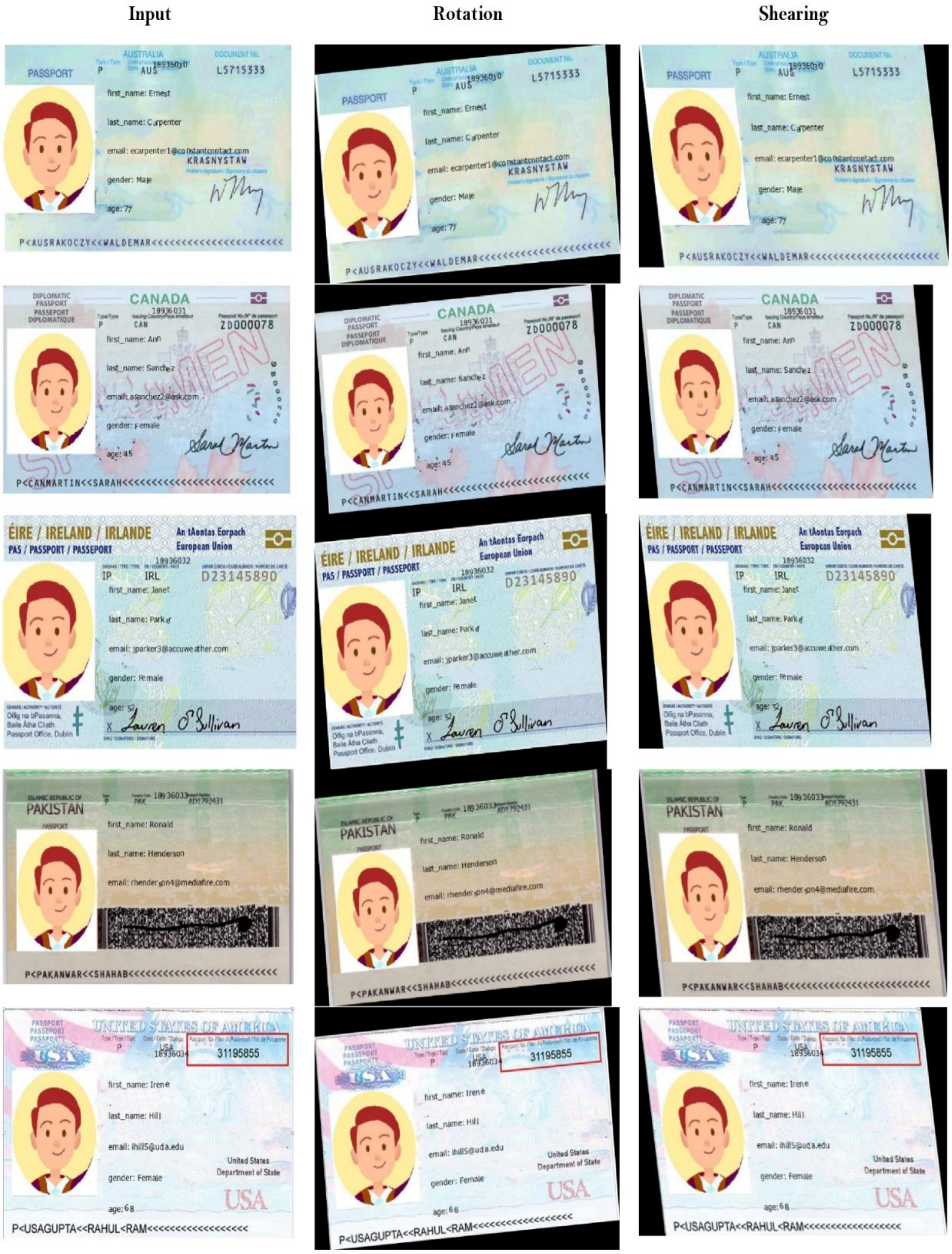

Figure 7. Passport with applied data augmentation Right to left, input, rotation augmentation and shearing augmentation.

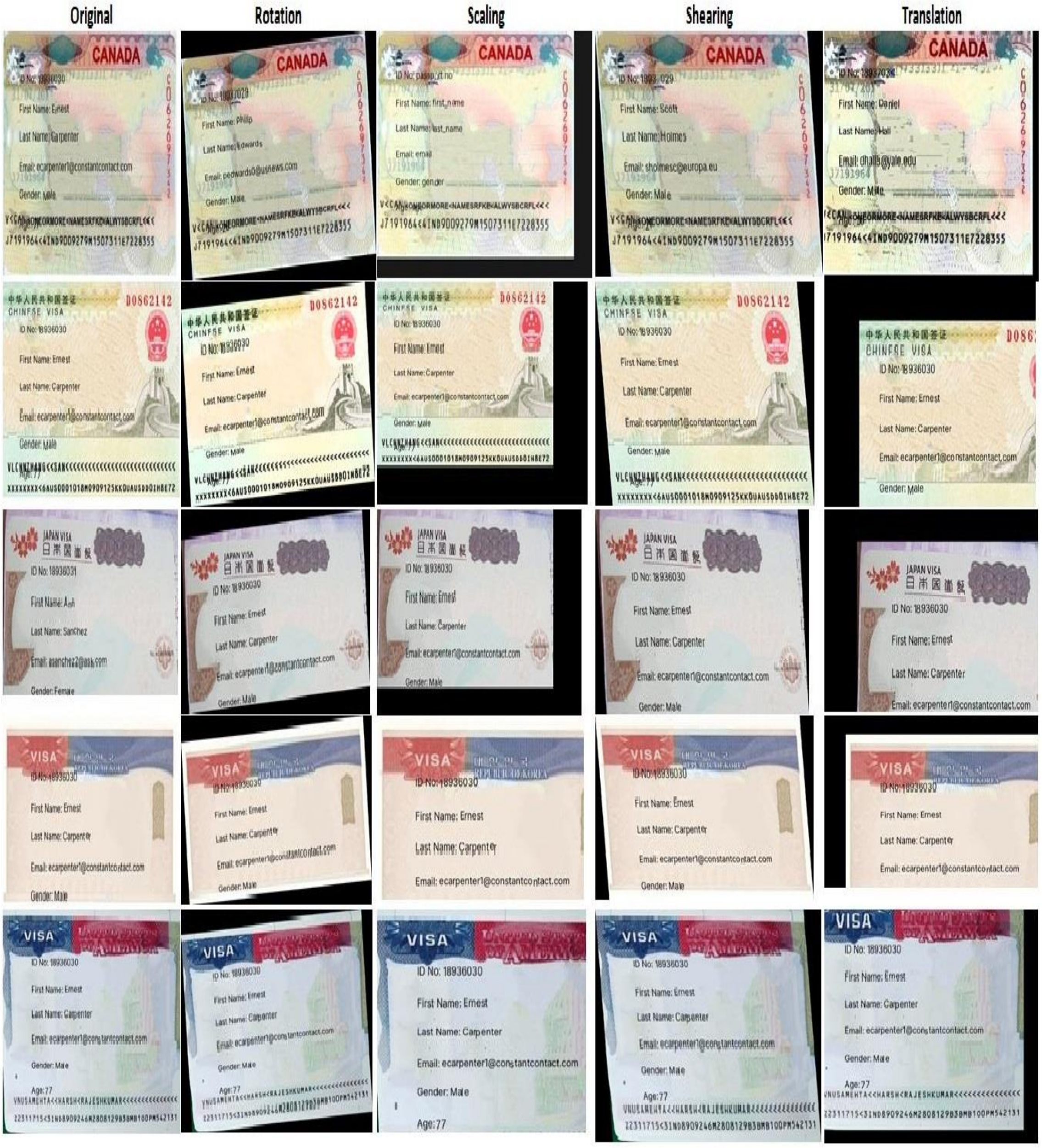

Figure 8 (a). VISA stickers with applied data augmentation Right to left, input, rotation, scaling, shearing and translation augmentation.

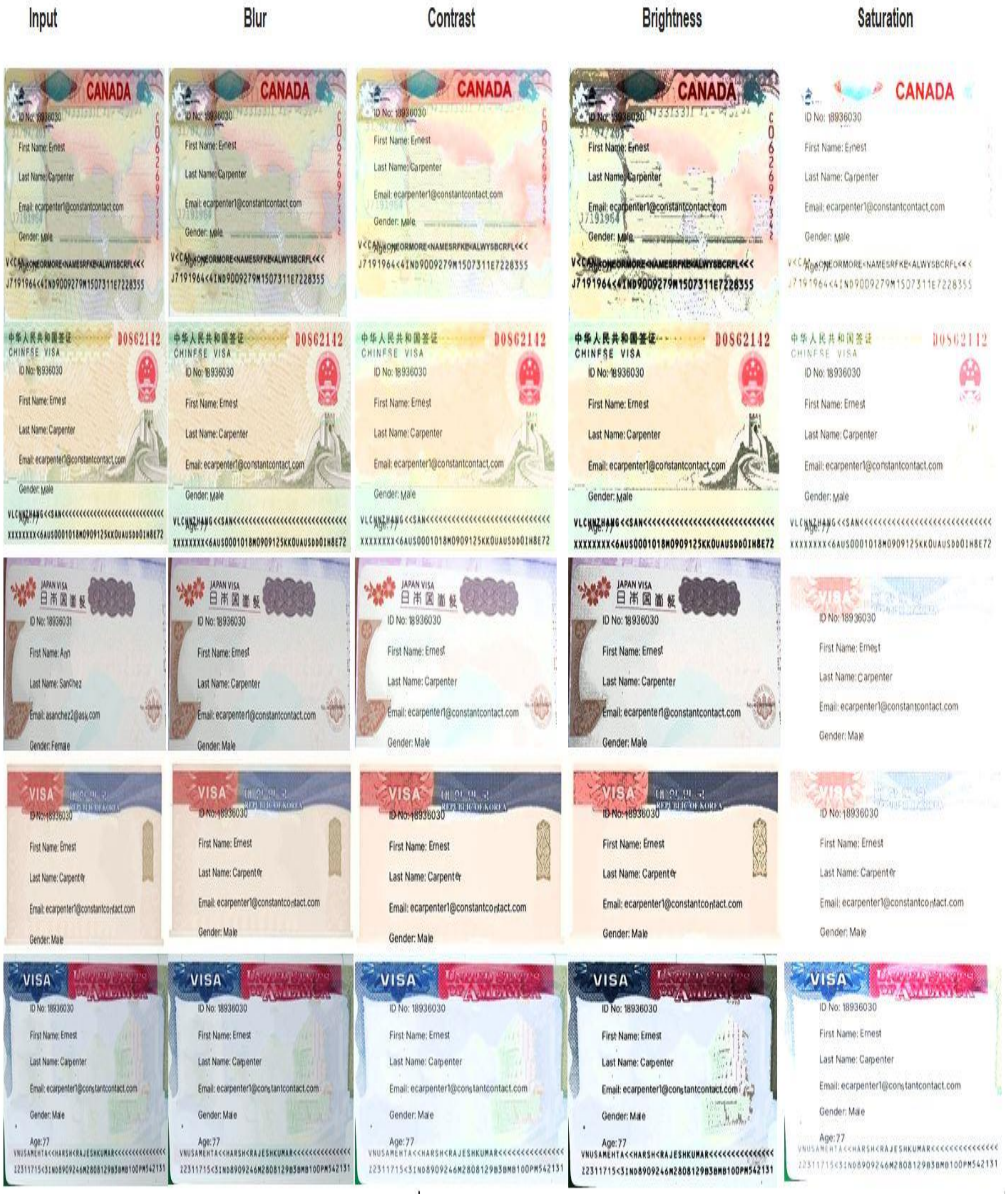

Figure 8 (b). VISA stickers with applied color data augmentation Right to left, input, blur. contrast, brightness and saturation augmentation.

algorithm records the bounding boxes of the forged characters and saves them in the json files so that the research community can use it for training networks for detection. Above four steps process is described for one document. To synthesise more data, we repeat this process to generate 3000 documents for each country document, finally we save forged document images with their bounding boxes in a json format file.

3. LIBRARIES / PACKAGE USED

We used multiple libraries to forge the character in the passport and driving licences. Libraries/packages including PIL [6] library to perform operation of image reading, setting and drawing font with its position, numpy [7] to deal rotation and other mathematical operations, random [8] to perform randomness for position and random noise adding, os [9] library to deal with file listing and path handling, json [10] library to deal with json file as annotation and cv [11] library to deal with augmentations.

4. CONCLUSIONS

This paper addresses the gap of forged character detection of documents i.e. passport, driving licence and VISA stickers, due to datasets unavailability. To fill this gap, this paper presents a new algorithm of synthesising data considering real world scenarios and releases three new datasets named forged characters detection for passport (FCD-P), forged characters detection for driving licences (FCD-D) and forged characters detection for VISA stickers (FCD-V), using five different countries' passport, driving licences and VISA stickers. This research work opens new challenges for forged character detection on passports, driving licences and VISA stickers. Finally, we release our code and datasets for the research purpose of the research community. Possible future work is to increase the datasets by including more countries' passports, driving licences and VISA stickers, and we will apply state-of-the-art detection algorithms on these datasets.

ACKNOWLEDGEMENTS

This publication has emanated from research [conducted with the financial support of/supported in part by a grant from Science Foundation Ireland under Grant number 18/CRT/6223 and is supported by the ADAPT Centre for Digital Content Technology which is funded under the SFI Research Centres Programme (Grant 13/RC/2106/_P2), Lero SFI Centre for Software (Grant 13/RC/2094/_P2) and is co-funded under the European Regional Development Fund. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

[6] https://pillow.readthedocs.io/en/stable/ (accessed on 1/17/2022)

[7] https://numpy.org/ (accessed on 1/17/2022)

[8] https://docs.python.org/3/library/random.html (accessed on 1/17/2022)

[9] https://docs.python.org/3/library/os.html (accessed on 1/17/2022)

[10] https://docs.python.org/3/library/json.html (accessed on 1/17/2022)

[11] https://pypi.org/project/opencv-python/ (accessed on 1/17/2022)

REFERENCES

[1] Fake identity brits warned that their lives are in danger, Online Available:https://www.independent.co.uk/news/world/middle-east/fake-identity-brits-warned-that-their-lives-are-in-danger-1905971.html .

[2] Wu, L., Zhang, C., Liu, J., Han, J., Liu, J., Ding, E., & Bai, X. (2019, October). Editing text in the wild. In *Proceedings of the 27th ACM international conference on multimedia* (pp. 1500-1508).

[3] Yang, Q., Huang, J., & Lin, W. (2020). Swaptext: Image based texts transfer in scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 14700-14709).

[4] Roy, P., Bhattacharya, S., Ghosh, S., & Pal, U. (2020). STEFANN: scene text editor using font adaptive neural network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 13228-13237).

[5] Korshunov, P., & Marcel, S. (2018). Deepfakes: a new threat to face recognition? assessment and detection. *arXiv preprint arXiv:1812.08685*.

[6] Zhao, L., Chen, C., & Huang, J. (2021). Deep Learning-based Forgery Attack on Document Images. *arXiv preprint arXiv:2102.00653*.

[7] Adadi, A. (2021). A survey on data-efficient algorithms in big data era. *Journal of Big Data*, *8*(1), 1-54.

[8] Aiman, A., Shen, Y., Bendechache, M., Inayat, I., & Kumar, T. (2021). AUDD: Audio Urdu Digits Dataset for Automatic Audio Urdu Digit Recognition. *Applied Sciences*, *11*(19), 8842.

[9] Kumar, T., Turab, M., Talpur, S Brennan, R., Bendechache, M. (2022). Detection Datasets: Forged Characters for Passport and Driving Licence. 6th International Conference on Artificial Intelligence, Soft Computing and Applications (AISCA 2022), (pp. 45-54)

[10] Bertrand, R., Gomez-Krämer, P., Terrades, O. R., Franco, P., & Ogier, J. M. (2013, August). A system based on intrinsic features for fraudulent document detection. In *2013 12th International conference on document analysis and recognition* (pp. 106-110). IEEE.

[11] Shang, S., Kong, X., & You, X. (2015). Document forgery detection using distortion mutation of geometric parameters in characters. *Journal of Electronic Imaging*, *24*(2), 023008.

[12] Ryan, M., & Hanafiah, N. (2015). An examination of character recognition on ID card using template matching approach. *Procedia Computer Science*, *59*, 520-529.

[13] Poddar, J., Parikh, V., & Bharti, S. K. (2020). Offline signature recognition and forgery detection using deep learning. *Procedia Computer Science*, *170*, 610-617.

[14] Bertrand, R., Terrades, O. R., Gomez-Krämer, P., Franco, P., & Ogier, J. M. (2015, August). A conditional random field model for font forgery detection. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*(pp. 576-580). IEEE.

[15] Cruz, F., Sidere, N., Coustaty, M., d'Andecy, V. P., & Ogier, J. M. (2017, November). Local binary patterns for document forgery detection. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)* (Vol. 1, pp. 1223-1228). IEEE.

[16] Sidere, N., Cruz, F., Coustaty, M., & Ogier, J. M. (2017, September). A dataset for forgery detection and spotting in document images. In *2017 Seventh International Conference on Emerging Security Technologies (EST)* (pp. 26-31). IEEE.

[17] Artaud, C., Doucet, A., Ogier, J. M., & d'Andecy, V. P. (2017, November). Receipt Dataset for Fraud Detection. In *First International Workshop on Computational Document Forensics*.

[18] Artaud, C., Sidère, N., Doucet, A., Ogier, J. M., & Yooz, V. P. D. A. (2018, August). Find it! fraud detection contest report. In *2018 24th International Conference on Pattern Recognition (ICPR)* (pp. 13-18). IEEE.

[19] Nandanwar, L., Shivakumara, P., Pal, U., Lu, T., Lopresti, D., Seraogi, B., & Chaudhuri, B. B. (2021). A new method for detecting altered text in document images. *International Journal of Pattern Recognition and Artificial Intelligence*, *35*(12), 2160010

[20] Nandanwar, L., Shivakumara, P., Mondal, P., Raghunandan, K. S., Pal, U., Lu, T., & Lopresti, D. (2021). Forged text detection in video, scene, and document images. *IET Image Processing*, *14*(17), 4744-4755.

[21] Deshpande, P., & Kanikar, P. (2012). Pixel based digital image forgery detection techniques. *International Journal of Engineering Research and Applications (IJERA)*, *2*(3), 539-543.

[22] Van Beusekom, J., Shafait, F., & Breuel, T. M. (2013). Text-line examination for document forgery detection. *International Journal on Document Analysis and Recognition (IJDAR)*, *16*(2), 189-207.

[23] Wilkinson, T. S., & Goodman, J. W. (1991, February). Slope histogram detection of forged handwritten signatures. In *High-Speed Inspection Architectures, Barcoding, and Character Recognition* (Vol. 1384, pp. 293-304). SPIE.

[24] Shivakumara, P., Basavaraja, V., Gowda, H. S., Guru, D. S., Pal, U., & Lu, T. (2018, August). A new RGB based fusion for forged IMEI number detection in mobile images. In *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)* (pp. 386-391). IEEE

[25] Kundu, S., Shivakumara, P., Grouver, A., Pal, U., Lu, T., & Blumenstein, M. (2019, November). A new forged handwriting detection method based on Fourier spectral density and variation. In *Asian Conference on Pattern Recognition* (pp. 136-150). Springer, Cham.

[26] Hashmi, S. A. A., Upadhyay, S., & Kumar, R. (2021). Comparative Study of Forged Urdu Signatures Done By Persons Not Familiar To Language Belongs To Region of Sikkim And Kashmir. *Journal of Punjab Academy of Forensic Medicine & Toxicology*, *21*(1).

[27] Lee, J., Kong, S. G., Lee, Y. S., Moon, K. W., Jeon, O. Y., Han, J. H., ... & Seo, J. S. (2012). Forged seal detection based on the seal overlay metric. *Forensic science international*, *214*(1-3), 200-206.

[28] Tralic, D., Zupancic, I., Grgic, S., & Grgic, M. (2013, September). CoMoFoD—New database for copy-move forgery detection. In *Proceedings ELMAR-2013* (pp. 49-54). IEEE.

[29] Cha, S. H., & Tappert, C. C. (2002, August). Automatic detection of handwriting forgery. In *Proceedings Eighth International Workshop on Frontiers in Handwriting Recognition* (pp. 264-267). IEEE.

[30] Nagel, R. N., & Rosenfeld, A. (1977). Computer detection of freehand forgeries. *IEEE Transactions on Computers*, *26*(09), 895-905.

[31] Megahed, A., Fadl, S. M., Han, Q., & Li, Q. (2017, November). Handwriting forgery detection based on ink colour features. In *2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)* (pp. 141-144). IEEE.

[32] Zhu, X., & Goldberg, A. B. (2009). Introduction to semi-supervised learning. *Synthesis lectures on artificial intelligence and machine learning*, *3*(1), 1-130.

AUTHORS

Teerath kumar received his Bachelor's degree in Computer Science with distinction from National University of Computer and Emerging Science (NUCES), Islamabad, Pakistan, in 2018. Currently, he is pursuing PhD from Dublin City University, Ireland. His research interests include advanced data augmentation, deep learning for medical imaging, generative adversarial networks and semi-supervised learning.

Muhammad Turab is an undergraduate final year student at Computer Systems Engineering MUET, Jamshoro. He has done 60+ projects with java and python, all projects can be found on GitHub. His research interests include deep learning, computer vision and data augmentation for medical imaging.

Shahnawaz Talpur is the chairman of Computer Systems Engineering Department at Muet Jamshoro. He has done his masters from MUET and PhD from Beijing Institute of Technology, China. His research interests include high performance computing, computer architecture and big data.

R. Brennan is an Assistant Professor in the School of Computing, Dublin City University, founding Chair of the DCU MA in Data Protection and Privacy Law and a Funded investigator in the Science Foundation Ireland ADAPT Centre for Digital Content Technology which is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund, His main research interests are data protection, data value, data quality, data privacy, data/AI governance and semantics.

M. Bendechache is an Assistant Professor in the School of Computing at Dublin City University, Ireland. She obtained her Ph.D. degree from University College Dublin, Ireland in 2018. Malika's research interests span the areas of Big data Analytics, Machine Learning, Data Governance, Cloud Computing, Blockchain, Security, and Privacy. She is an academic member and a Funded Investigator of ADAPT and Lero research centres.