

Article

Document Liveness Challenge Dataset (DLC-2021)

Dmitry V. Polevoy^{1,2,3,*}, Irina V. Sigareva^{1,4}, Daria M. Ershova^{1,5}, Vladimir V. Arlazarov^{1,2},
Dmitry P. Nikolaev^{1,6}, Zuheng Ming⁷, Muhammad Muzzamil Luqman⁷ and Jean-Christophe Burie⁷

¹ Smart Engines Service LLC, 117312 Moscow, Russia; iv.sigareva@smartengines.com (I.V.S.);

d.ershova@smartengines.com (D.M.E.); vva@smartengines.com (V.V.A.); dimonstr@iitp.ru (D.P.N.)

² Federal Research Center "Computer Science and Control" RAS, 119333 Moscow, Russia

³ National University of Science and Technology MISIS, 119049 Moscow, Russia

⁴ Moscow Institute of Physics and Technology, 141701 Dolgoprudny, Russia

⁵ Faculty of Mechanics and Mathematics, Lomonosov Moscow State University, 119991 Moscow, Russia

⁶ Institute for Information Transmission Problems (Kharkevich Institute) RAS, 127051 Moscow, Russia

⁷ L3i Laboratory, La Rochelle University, 17042 La Rochelle, France; zuheng.ming@univ-lr.fr (Z.M.);

muhammad_muzzamil.luqman@univ-lr.fr (M.M.L.); jcburie@univ-lr.fr (J.-C.B.)

* Correspondence: dvpsun@gmail.com

Abstract: Various government and commercial services, including, but not limited to, e-government, fintech, banking, and sharing economy services, widely use smartphones to simplify service access and user authorization. Many organizations involved in these areas use identity document analysis systems in order to improve user personal-data-input processes. The tasks of such systems are not only ID document data recognition and extraction but also fraud prevention by detecting document forgery or by checking whether the document is genuine. Modern systems of this kind are often expected to operate in unconstrained environments. A significant amount of research has been published on the topic of mobile ID document analysis, but the main difficulty for such research is the lack of public datasets due to the fact that the subject is protected by security requirements. In this paper, we present the DLC-2021 dataset, which consists of 1424 video clips captured in a wide range of real-world conditions, focused on tasks relating to ID document forensics. The novelty of the dataset is that it contains shots from video with color laminated mock ID documents, color unlaminated copies, grayscale unlaminated copies, and screen recaptures of the documents. The proposed dataset complies with the GDPR because it contains images of synthetic IDs with generated owner photos and artificial personal information. For the presented dataset, benchmark baselines are provided for tasks such as screen recapture detection and glare detection. The data presented are openly available in Zenodo.



Citation: Polevoy, D.V.; Sigareva, I.V.; Ershova, D.M.; Arlazarov, V.V.; Nikolaev, D.P.; Ming, Z.; Luqman, M.M.; Burie, J.-C. Document Liveness Challenge Dataset (DLC-2021). *J. Imaging* **2022**, *8*, 181. <https://doi.org/10.3390/jimaging8070181>

Academic Editors: Georgina Cosma and Ioannis Pratikakis

Received: 7 December 2021

Accepted: 24 June 2022

Published: 28 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: document analysis; document recognition; identity documents; mobile recognition; liveness detection; document anti-fraud; document forgery detection; screen recapture detection; open data

1. Introduction

The growing popularity of mobile services increases the risk of financial and other losses from fraudulent user actions. To reduce the number of illegal actions and comply with the law when using mobile services, it is often required for users to present their identity documents. In the case of remote access via a mobile device, this means receiving and analyzing identity (ID) document images. ID document recognition systems [1,2] are widely used to obtain and check users' personal information in many applications. At the same time, despite a large number of publications on the topic of ID document recognition, due to legal and ethical restrictions, researchers are constrained by [3] a lack of open datasets that can be used to reproduce and compare results. The absence of open datasets for ID document fraud prevention research inspired us to create a new dataset, called DLC-2021 [4–6]. It can be used to establish an evaluation methodology and set up

baselines for document image recapture detection, document photocopy detection, and document lamination detection methods.

2. Overview

The GDPR [7] and other local laws prohibit the creation of datasets with real ID images. Thus, researchers began to use artificially generated ID document images for open dataset creation [8–12]. As far as we know, printed mock documents are used only in MIDV family datasets, and MIDV-500 was the first [8]. This dataset contained 500 video clips of 50 identity documents, with 10 clips per document type. The identity documents were of different types, and were mostly “sample” or “specimen” documents that could be found in WikiMedia and were distributed under public copyright licenses. The conditions represented in MIDV-500 thus had some diversity regarding the background and the positioning of the document in relation to the mobile capturing process; however, they did not include variation in lighting conditions, or significant projective distortions. MIDV-2019 [9] was later published as an extension of MIDV-500. It contained video clips captured with very low lighting conditions and with higher projective distortions. The dataset was also supplemented with photos and scanned images of the same document types to represent the typical input for server-side identity document analysis systems. MIDV-2020 [10] was published recently to provide variability in the text fields, faces, and signatures, while retaining the realism of the dataset. The MIDV-2020 dataset consists of 1000 different physical documents (100 documents per type), all with unique, artificially generated faces, signatures, and text field data. Each physical document was photographed and scanned, and for each a video clip was captured using a smartphone. The ground truth includes ideal text field values, and the geometrical position of documents and faces in each photo, scan, and video clip frame (with 10 frames-per-second annotation). MIDV-LAIT [11] contains video for ID documents with textual fields in Perso-Arabic, Thai, and Indian scripts.

When using mobile-based ID document recognition systems, the most technically simple and accessible attack methods are different types of rebroadcast attacks [13]. For the DLC-2021 dataset we shot mock documents from the MIDV-2020 collection as originals (Figure 1a) and modeled those types of attacks that remain realistic when using mock documents: capturing a color printed copy of a document without lamination (Figure 1b), capturing a gray printed unlaminated copy of a document (Figure 1c) and capturing a displayed image of a document (Figure 1d).



Figure 1. Types of video in DLC-2021 dataset: (a) original document, (b) unlaminated color copy, (c) unlaminated gray copy, and (d) document recaptured from screen.

Thus, all images in the MIDV family of datasets [8–11] can be considered as images of genuine documents and can be used as negative samples for document fraud detectors.

There are many studies on engineering- [14–20] and neural-network-based methods [13,21–24] for screen recapture detection, but most of them are focused on the analysis of natural images publicly available in the following datasets: NTU-ROSE [15], ICL-COMMSP [16,18], and BJTU-IIS [17]. These datasets are captured as photos by high-resolution DSLR cameras.

Document-specific methods for detecting document recapture are based on the latest advances in deep learning. The algorithm proposed in [25] takes advantage of both metric learning and image forensic techniques. The authors considered practical domain generalization problems, such as the variations in printing/imaging devices, substrates, recapturing channels, and document types with a private dataset. The texture and reflectance characteristics of the bronzing region are used as discriminative features to detect a recaptured certificate document in [26]. The dataset used in the study is available upon request.

Thus, for research in the field of document recapture prevention, new specialized open datasets captured with smartphones are required.

3. DLC-2021 Dataset Description

The set of 10 ID document types for DLC-2021 (Table 1) coincides with the set of document types in the MIDV-2020 dataset.

Table 1. Document types featured in DLC-2021.

#	Document Type Code	Document Type
1	alb_id	ID Card of Albania
2	aze_passport	Passport of Azerbaijan
3	esp_id	ID Card of Spain
4	est_id	ID Card of Estonia
5	fin_id	ID Card of Finland
6	grc_passport	Passport of Greece
7	lva_passport	Passport of Latvia
8	rus_internalpassport	Internal Passport of Russia
9	srb_passport	Passport of Serbia
10	svk_id	ID Card of Slovakia

For each type of document, eight examples of physical documents were taken. For selected physical documents, color and gray paper hard copies were made by printing without lamination. All color copies and some of the gray copies were cut to fit the original document page shape.

While preparing the DLC-2021 dataset, we focused on video capture. On the one hand, the video stream allows for analysis changes in time, and this provides much more information for assessing liveliness. On the other hand, video frames usually contain compression artifacts that can significantly affect the performance of analysis algorithms.

In general, DLC-2021 follows the structure of the MIDV-2020 folder and files, except for clip names. In DLC-2021, the two-digit document template number clip name is extended with a two-letter video type code (Table 2) and four-digit serial number.

Table 2. Types of video.

Video Type Code	Description
cc	unlaminated color copy
cg	unlaminated gray copy
or	“original” laminated documents from MIDV-2020 collection
re	video recapture for document on device screen

An Apple iPhone XR and Samsung S10 were used for video capturing, as in MIDV-2020. Video clips were shot with a wide-angle camera (Table 3) using a standard smartphone camera application.

Table 3. Smartphone camera specification.

Characteristic	iPhone XR [27]	Samsung S10 [28]
focal length (equivalent)	26 mm	26 mm
lens aperture	$f/1.8$	$f/1.5-2.4$
sensor size	1/2.55"	1/2.55"
sensor pixel size	1.4 μ	1.4 μ
sensor pixel count	12 MP	12 MP

To make videos more varied, we used two different frame resolutions (1080 \times 1920, 2160 \times 3840) and two different frame rates (30, 60 fps) for shooting video clips. Table 4 summarizes the number of video clips by type.

Table 4. Number of video clips by type.

Smartphone	Resolution	FPS	Video Type				Total
			or	cc	cg	re	
Samsung S10	3840 \times 2160	30	140	283	121	200	744
iPhone XR	3840 \times 2160	60	70	201	51	200	522
Samsung S10	1920 \times 1080	30	40		39		79
iPhone XR	1920 \times 1080	30	40		39		79
Total:			290	484	250	400	1424

Each clip was shot vertically and was at least five seconds long. Frames were extracted at 10 frames per second using ffmpeg version n4.4 with default parameters, and for the first 50 extracted frames the document position was manually annotated. The annotation file for each clip followed the MIDV-2020 JSON format [10] and was readable with VGG Image Annotator (v2.0.11) [29].

3.1. Paper Document Shooting

We captured video with the “original” documents and printed copies under different lighting conditions, such as natural daylight, bright light with deep shadows, artificial light, colored light, low light, and flashlight. The color characteristics of document images varied significantly under different lighting and capture conditions (Figure 2).

Low or uneven lighting and white balance correction algorithms inappropriate for the lighting conditions dramatically affect color reproduction and complicate the process of distinguishing color documents from gray copies (Figure 3) without specialized color correction algorithms, such as that in [30].

To achieve greater realism of the video, various document occlusions were made on some of the clips (Figure 4), such as holding the document with fingers, and a brightly colored object in the document area.



Figure 2. Color variation of original document images for a variety of capture conditions, type grc_passport.



Figure 3. Color variation in gray copies of document images for a variety of capture conditions, type grc_passport.



Figure 4. Overlapping variation of document images, type grc_passport.

In the task of detecting gray copies, such partial occlusion can create additional difficulties, as it can lead to an increase in the color diversity of pixels in the document area.

Since ID documents are used regularly, manufacturers protect them from dirt, creases, and other damage by using a special protective coating or lamination. Such a coating can preserve the integrity of documents for a long time from various environmental influences and also significantly complicate attempts to change the content of the document, for example, such as replacing a photo. However, laminated documents can easily introduce reflection and saturation phenomenon, especially when a strong illuminant such as a flash, a fluorescent lamp, or even the sun lights the document during the video-capturing process. Figure 5 shows some images extracted from a video captured with a smartphone.



Figure 5. Reflections caused by the lighting condition on the surface of laminated documents, type grc_passport.

Strong reflections on the smooth surface of the laminated documents can partially or totally hide the content of the document, making it impossible to analyze the pictures or to extract the text. In addition, the shape and the size of the area of reflection may vary depending on the orientation of the document relative to the smartphone lens. On the one hand, these variations are challenging for detection, segmentation, and recognition algorithms. On the other hand, the analysis of the shape and consistency of changes in highlights and scene geometry can serve as an important indicator of the liveliness of a document. For example, exploiting the camera flashlight during the capture process creates semi-controlled lighting conditions in which laminated and unlaminated documents in some cases can be differentiated more robustly (Figure 6).



Figure 6. Frames with reflections caused by a flashlight for a laminated document, clip est_id/04.or0004 (top), and unlaminated document, clip est_id/04.cc0009 (bottom).

3.2. Screens Shooting

For screen recapture, we used two office desktops and two notebook LCD monitors. Figure 7 shows samples from the template image and video for original and screen-recaptured cases.



Figure 7. Zones from geometry-normalized images for template (column 1), original (column 2) and recaptured documents (column 3–5), type srb_passport.

It should be noted that the documents themselves may have a complex textured page background, for example, when using document-protection technologies such as guilloche. Another interesting case is textured scene objects, or even the LCD screen behind the document. In such cases, moiré and other recapture artifacts can also occur outside the document zone when the original document is captured with a digital camera.

4. Experimental Baselines

While the main goal of the paper is to present a document liveness challenge dataset, DLC-2021, in order to provide a baseline for future research involving the dataset, in the

following sections several benchmarks using DLC-2021 will be presented. As a baseline method we chose Convolutional Neural Networks (CNNs) in view of the fact that CNNs show state-of-the-art results in image classification tasks. In our experiments we used the Keras (2.6.0) library from the Tensorflow (2.8.0) [31] framework and Scikit-learn (1.2.0) library [32]. Scripts, instructions, and pre-trained models to reproduce our experiments can be downloaded from [4].

4.1. Screen Recapture Detection

For screen recapture detection, we used a classification CNN model based on ResNet-50 architecture [33] pre-trained on ImageNet weights from TensorFlow Model Garden. We froze the first 49 layers and reduced the number of the last softmax layer outputs to 2. For learning, we used the binary cross-entropy loss function and Adam [34] optimizer with a constant learning rate ($lr = 0.1$).

The screen recapture detector classifies 224×224 patches cut from the center of the document on the original frame. Negative samples are collected from MIDV-2020 images. To collect positive samples, we cropped and manually labeled patches from DLC-2021 recaptured images of Spanish IDs, Latvian passports, and internal passports from Russia. The training set consisted of 19,543 positive and 25,980 negative samples.

The validation dataset contained 11,009 positive and 16,264 negative samples formed from original document images and recaptured images for other DLC-2021 document types. Table 5 shows results from the validation dataset for CNN-based and Scikit Dummy Classifier detectors with different strategies: “constant” (generates constant prediction), “stratified” (generates predictions with respect to the balance of training set classes), and “uniform” (generates predictions uniformly at random). Results for “stratified” and “uniform” strategies were averaged over 10 runs with different seed values, and the standard deviation values are shown in the table.

Table 5. Performance comparison between CNN-based and Scikit Dummy Classifier detectors for a screen recapture per-frame detection task.

Metrics	CNN	Dummy Classifier			
		const = false	const = true	Stratified	Uniform
accuracy	89.67%	59.63%	40.37%	$51.28 \pm 0.38\%$	$50.16 \pm 0.31\%$
precision	85.89%	–	40.37%	$40.29 \pm 0.45\%$	$40.53 \pm 0.29\%$
recall	89.03%	0.00%	100.00%	$42.93 \pm 0.59\%$	$50.20 \pm 0.34\%$

Most of the false-positive (FP) errors were caused by documents having complex textured backgrounds and compression artifacts, as shown in Figure 8.



Figure 8. FP error samples for CNN-based screen recapture detector.

4.2. Unlaminated Color Copy Detection

The presence of glare is the most evident feature of laminated documents. An unlaminated color copy detector classifies projective undistorted images by frame markup and scaled-down document images. The ResNet-50-based CNN detector showed a steady trend of overfitting, so a more simple architecture, as presented in Table 6, was used.

Table 6. CNN-based unlaminated color copy detector architecture (layers).

#	Type	Parameters	Output Size	Activation Function
1	Conv	8 filters 3×3 , stride 1×1 , no padding	$74 \times 74 \times 8$	relu
2	Conv	16 filters 3×3 , stride 1×1 , no padding	$72 \times 72 \times 16$	relu
3	MaxPool	pooling 2×2 , no padding	$36 \times 36 \times 16$	
4	Conv	16 filters 3×3 , strides 2×2 , padding 2×2	$34 \times 34 \times 16$	relu
5	Conv	24 filters 3×3 , stride 1×1 , padding 1×1	$16 \times 16 \times 24$	relu
6	Conv	32 filters 2×2 , stride 1×1 , padding 1×1	$15 \times 15 \times 32$	relu
7	MaxPool	pooling 2×2 , no padding	$7 \times 7 \times 32$	
8	Conv	32 filters 2×2 , stride 1×1 , no padding	$6 \times 6 \times 32$	relu
9	Conv	12 filters 2×2 , stride 1×1 , no padding	$5 \times 5 \times 12$	relu
10	Flatten		$1 \times 1 \times 300$	
11	Dropout	dropout rate = 0.4	$1 \times 1 \times 300$	
12	Fully Connected	2 outputs	$1 \times 1 \times 2$	softmax

The CNN-based detector was trained on gray images scaled down to 76×76 with a binary cross-entropy loss function and Adam optimizer (learning rate = 0.05). Early stopping and data augmentation (brightness distortion with range [0.9, 1.1]) were used to avoid overfitting.

The training dataset was collected from manually labeled MIDV-500 and MIDV-2020 images and contained 29,564 positive and 7544 negative samples. The validation dataset was collected from manually labeled DLC-2021 clip images (or and cc types) and contained 34,607 positive and 3388 negative samples. Table 7 shows the results from the validation dataset for CNN-based and Scikit Dummy Classifier detectors.

Table 7. Performance comparison between CNN-based and Scikit Dummy Classifier detectors for an unlaminated color copy per-frame detection task.

Metrics	CNN	Dummy Classifier			
		const = false	const = true	Stratified	Uniform
accuracy	83.61%	8.92%	91.08%	$74.39 \pm 0.19\%$	$49.98 \pm 0.21\%$
precision	96.01%	–	91.08%	$91.10 \pm 0.07\%$	$91.07 \pm 0.15\%$
recall	85.56%	0.00%	100.00%	$79.67 \pm 0.20\%$	$49.99 \pm 0.17\%$

4.3. Gray Copy Detection

Projective undistorted document images were used for classification. Positive samples in the training set were collected from gray copy clips of Azerbaijani passports, Finnish ID cards, and Serbian passports. Negative samples in the training set were obtained from MIDV-2020. The training set contained 3492 positive and 1000 negative samples. The validation set contained copied grey clips for all other types of documents and original document clips from DLC-2021 (10473 positive and 16264 negative samples).

All experiments with ResNet-50-like models (similar to Section 4.1) and more simple CNN models (similar to Section 4.2) failed. Models either did not train at all or were overfitted. One reason for this result is that CNNs are sensitive to intensity gradient features but ignore color features. Since the development of a more sophisticated CNN architecture is beyond the scope of this article, as a simple baseline, we examined the Scikit Dummy Classifier detector on the validation dataset, as shown in Table 8.

Table 8. Performance of the Scikit Dummy Classifier in a gray copy per-frame detection task.

Metrics	Dummy Classifier			
	const = false	const = true	Stratified	Uniform
accuracy	60.83%	39.17%	44.06 ± 0.25%	50.09 ± 0.26%
precision	–	39.17%	39.22 ± 0.15%	39.26 ± 0.25%
recall	0.00%	100.00%	77.82 ± 0.32%	50.11 ± 0.35%

5. Conclusions

In this paper, we presented the DLC-2021 dataset containing video clips of mock “real” identity documents from the MIDV-2020 collection and three types of popular rebroadcast attacks: capturing a color printed copy of a document without lamination, capturing a gray printed un laminated copy of a document and capturing a displayed image of a document. Video was captured using modern smartphones with different video quality, and a wide range of different real-world capturing conditions were simulated. Selected video frames were accompanied by the geometric markup of the outer borders of the document.

Using mock documents from the MIDV-2020 collection as targets for shooting DLC-2021 video makes it easy to use field values and document geometry markup from MIDV-2020 templates. The prepared open dataset can be used for other ID-recognition tasks:

- Document detection and localization in the image [35–37];
- Document type identification [35,37];
- Document layout analysis;
- Detection of faces in document images [38] and the choice of the best photo of the document owner [39];
- Integration of the recognition results [40];
- Video frame quality assessment [41] and the choice of the best frame [42].

As the videos were captured with two different smartphones, the DLC-2021 dataset can be used for sensor noise (PRNU)-based method analysis.

In the future, we plan to expand the DLC dataset with more screen types and devices for shooting, as well as increase the variety of document types.

Regarding ethical AI, the published dataset has no potential to affect the privacy of individuals regarding personal data, since all documents are synthetic mock-ups and comply with the GDPR.

The authors believe that the provided dataset will serve as a valuable resource for ID document recognition and ID document fraud prevention, and lead to more high-quality scientific publications in the field of ID document analysis, as well as in the general field of computer vision.

Author Contributions: All authors contributed to the study conceptualization and design. The first draft of the manuscript was written by D.V.P., supervised by D.P.N., V.V.A., M.M.L. and J.-C.B. Data curation I.V.S., D.V.P. and Z.M. The experiments with CNN detectors were conducted by D.M.E. All authors provided critical feedback, made amendments and helped shape the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented are openly available under Creative Commons Attribution–ShareAlike 2.5 Generic License (CC BY-SA 2.5) in Zenodo at doi:10.5281/zenodo.6466767, doi:10.5281/zenodo.6466770, doi:10.5281/zenodo.6466764.

Acknowledgments: All source images for MIDV-2020 dataset were obtained from Wikimedia Commons. Author attributions for each source images are listed in the original MIDV-500 description table (<ftp://smartengines.com/midv-500/documents.pdf>). Face images by Generated Photos (<https://generated.photos>).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bulatov, K.; Arlazarov, V.V.; Chernov, T.; Slavin, O.; Nikolaev, D. Smart IDReader: Document recognition in video stream. In Proceedings of the 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, Japan, 9–15 November 2017; Volume 6, pp. 39–44.
2. Attivissimo, F.; Giaquinto, N.; Scarpetta, M.; Spadavecchia, M. An Automatic Reader of Identity Documents. In Proceedings of the 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), Bari, Italy, 6–9 October 2019; pp. 3525–3530.
3. Centeno, A.B.; Terrades, O.R.; Canet, J.L.; Morales, C.C. Identity Document and banknote security forensics: A survey. *arXiv* **2019**, arXiv:1910.08993.
4. Polevoy, D.V.; Sigareva, I.V.; Ershova, D.M.; Arlazarov, V.V.; Nikolaev, D.P.; Ming, Z.; Luqman, M.M.; Burie, J.C. Document Liveness Challenge (DLC-2021)—Part 1 (or, cg) [Data set]. *Zenodo* **2022**. [[CrossRef](#)]
5. Polevoy, D.V.; Sigareva, I.V.; Ershova, D.M.; Arlazarov, V.V.; Nikolaev, D.P.; Ming, Z.; Luqman, M.M.; Burie, J.C. Document Liveness Challenge (DLC-2021)—Part 2 (re) [Data set]. *Zenodo* **2022**. [[CrossRef](#)]
6. Polevoy, D.V.; Sigareva, I.V.; Ershova, D.M.; Arlazarov, V.V.; Nikolaev, D.P.; Ming, Z.; Luqman, M.M.; Burie, J.C. Document Liveness Challenge (DLC-2021)—Part 3 (cc) [Data set]. *Zenodo* **2022**. [[CrossRef](#)]
7. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation). Available online: <https://eur-lex.europa.eu/eli/reg/2016/679/oj> (accessed on 19 September 2021).
8. Arlazarov, V.V.; Bulatov, K.; Chernov, T.; Arlazarov, V.L. MIDV-500: A Dataset for Identity Document Analysis and Recognition on Mobile Devices in Video Stream. *Comput. Opt.* **2019**, *43*, 818–824. [[CrossRef](#)]
9. Bulatov, K.; Matalov, D.; Arlazarov, V.V. MIDV-2019: Challenges of the Modern Mobile-Based Document OCR. In Proceedings of the 12th International Conference on Machine Vision (ICMV'19), Amsterdam, The Netherlands, 16–18 November 2019; Volume 11433, pp. 1–6. [[CrossRef](#)]
10. Bulatov, K.B.; Emelyanova, E.V.; Tropin, D.V.; Skoryukina, N.S.; Chernyshova, Y.S.; Sheshkus, A.V.; Usilin, S.A.; Ming, Z.; Burie, J.C.; Luqman, M.M.; et al. MIDV-2020: A Comprehensive Benchmark Dataset for Identity Document Analysis. *arXiv* **2021**, arXiv:2107.00396.
11. Chernyshova, Y.S.; Emelianova, E.V.; Sheshkus, A.V.; Arlazarov, V.V. MIDV-LAIT: A challenging dataset for recognition of IDs with Perso-Arabic, Thai, and Indian scripts. In *Lecture Notes in Computer Science (LNCS)*, 2nd ed.; Lladós, J., Lopresti, D., Eds.; Springer: New York, NY, USA, 2021; Volume 12822, pp. 258–272. [[CrossRef](#)]
12. De Sá Soares, A.; das Neves Junior, R.B.; Bezerra, B.L.D. BID Dataset: A challenge dataset for document processing tasks. In Proceedings of the Anais Estendidos do XXXIII Conference on Graphics, Patterns and Images (SBC), Virtual Conference, 7–10 November 2020; pp. 143–146.
13. Agarwal, S.; Fan, W.; Farid, H. A diverse large-scale dataset for evaluating rebroadcast attacks. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 1997–2001.
14. Yu, H.; Ng, T.T.; Sun, Q. Recaptured photo detection using specularly distribution. In Proceedings of the 2008 15th IEEE International Conference on Image Processing, San Diego, CA, USA, 12–15 October 2008; pp. 3140–3143.
15. Cao, H.; Kot, A.C. Identification of recaptured photographs on LCD screens. In Proceedings of the 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, Dallas, TX, USA, 14–19 March 2010; pp. 1790–1793.
16. Muammar, H.; Dragotti, P.L. An investigation into aliasing in images recaptured from an LCD monitor using a digital camera. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 2242–2246.
17. Li, R.; Ni, R.; Zhao, Y. An effective detection method based on physical traits of recaptured images on LCD screens. In Proceedings of the International Workshop on Digital Watermarking, Tokyo, Japan, 7–10 October 2015; pp. 107–116.
18. Thongkamwitoon, T.; Muammar, H.; Dragotti, P.L. An image recapture detection algorithm based on learning dictionaries of edge profiles. *IEEE Trans. Inf. Forensics Secur.* **2015**, *10*, 953–968. [[CrossRef](#)]
19. Mahdian, B.; Novozámský, A.; Saic, S. Identification of aliasing-based patterns in re-captured LCD screens. In Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 27–30 September 2015; pp. 616–620.
20. Wang, K. A simple and effective image-statistics-based approach to detecting recaptured images from LCD screens. *Digit. Investig.* **2017**, *23*, 75–87. [[CrossRef](#)]
21. Yang, P.; Ni, R.; Zhao, Y. Recapture image forensics based on Laplacian convolutional neural networks. In Proceedings of the International Workshop on Digital Watermarking, Beijing, China 17–19 September 2016; pp. 119–128.
22. Choi, H.Y.; Jang, H.U.; Son, J.; Kim, D.; Lee, H.K. Content recapture detection based on convolutional neural networks. In Proceedings of the International Conference on Information Science and Applications, Quito, Ecuador, 23–27 November 2017; pp. 339–346.
23. Li, H.; Wang, S.; Kot, A.C. Image recapture detection with convolutional and recurrent neural networks. *Electron. Imaging* **2017**, *2017*, 87–91. [[CrossRef](#)]

24. Wang, J.; Wu, G.; Li, J.; Jha, S.K. A new method estimating linear gaussian filter kernel by image PRNU noise. *J. Inf. Secur. Appl.* **2019**, *44*, 1–11. [[CrossRef](#)]
25. Chen, C.; Zhang, S.; Lan, F.; Huang, J. Domain generalization for document authentication against practical recapturing attacks. *arXiv* **2021**, arXiv:2101.01404.
26. Yan, J.; Chen, C. Cross-Domain Recaptured Document Detection with Texture and Reflectance Characteristics. In Proceedings of the 2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Tokyo, Japan, 14–17 December 2021; pp. 1708–1715.
27. Wikipedia. iPhone XR. Available online: https://en.wikipedia.org/wiki/IPhone_XR (accessed on 15 February 2022).
28. Wikipedia. Samsung Galaxy S10. Available online: https://en.wikipedia.org/wiki/Samsung_Galaxy_S10 (accessed on 15 February 2022).
29. Dutta, A.; Zisserman, A. The VIA annotation software for images, audio and video. In Proceedings of the 27th ACM International Conference on Multimedia, Nice, France, 21–25 October 2019; pp. 2276–2279. [[CrossRef](#)]
30. Polevoy, D.; Panfilova, E.; Ershov, E.; Nikolaev, D. Color correction of the document owner’s photograph image during recognition on mobile device. In Proceedings of the Thirteenth International Conference on Machine Vision—International Society for Optics and Photonics, Online, 8–14 November 2021; Volume 11605, p. 1160510.
31. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. TensorFlow: A system for large-scale machine learning. In Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation (OSDI’16), Savannah, GA, USA, 2–4 November 2016; pp. 265–283.
32. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
33. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778. [[CrossRef](#)]
34. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
35. Skoryukina, N.; Shemyakina, J.; Arlazarov, V.L.; Faradzhev, I. Document localization algorithms based on feature points and straight lines. In Proceedings of the Tenth International Conference on Machine Vision, Vienna, Austria, 13–14 November 2017; pp. 1–8. [[CrossRef](#)]
36. Puybureau, E.; Géraud, T. Real-time document detection in smartphone videos. In Proceedings of the 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 1498–1502.
37. Skoryukina, N.; Arlazarov, V.V.; Nikolaev, D.P. Fast method of ID documents location and type identification for mobile and server application. In Proceedings of the International Conference on Document Analysis and Recognition (ICDAR’19), Sydney, NSW, Australia, 20–25 September 2019; pp. 850–857. [[CrossRef](#)]
38. Bakkali, S.; Luqman, M.M.; Ming, Z.; Burie, J.C. Face detection in camera captured images of identity documents under challenging conditions. In Proceedings of the 2019 International Conference on Document Analysis and Recognition Workshops (ICDARW), Los Alamitos, CA, USA, 22–25 September 2019; Volume 4, pp. 55–60.
39. Polevoy, D.V.; Aliev, M.A.; Nikolaev, D.P. Choosing the best image of the document owner’s photograph in the video stream on the mobile device. In Proceedings of the 13th International Conference on Machine Vision (ICMV 2020), Rome, Italy, 2–6 November 2021; Volume 11605, pp. 1–9. [[CrossRef](#)]
40. Bulatov, K.B. A Method to Reduce Errors of String Recognition Based on Combination of Several Recognition Results with Per-Character Alternatives. *Bull. South Ural. State Univ. Ser. Math. Model. Program. Comput. Softw.* **2019**, *12*, 74–88. [[CrossRef](#)]
41. Chernov, T.; Razumnuy, N.; Kozharinov, A.; Nikolaev, D.; Arlazarov, V. Image quality assessment for video stream recognition systems. In Proceedings of the Tenth International Conference on Machine Vision (ICMV 2017), Vienna, Austria, 13–15 November 2017; Volume 10696, p. 10696. [[CrossRef](#)]
42. Aliev, M.A.; Kunina, I.A.; Kazbekov, A.V.; Arlazarov, V.L. Algorithm for choosing the best frame in a video stream in the task of identity document recognition. *Comput. Opt.* **2021**, *45*, 101–109. [[CrossRef](#)]