IDNet: A Novel Identity Document Dataset via Few-Shot and Quality-Driven Synthetic Data Generation

Lulu Xie*[a], Yancheng Wang*[a], Hong Guan*[a], Soham Nag[a], Rajeev Goel[a], Niranjan Swamy[a],
Yingzhen Yang[a], Chaowei Xiao[c], Jonathan Prisby[b], Ross Maciejewski[a], Jia Zou[a]

Department of Homeland Security–Center of Accelerated Operational Excellence, Arizona State University[a]
Department of Homeland Security–Science and Technology Directorate [b]
University of Wisconsin, Madison [c]

2024 IEEE International Conference on Big Data (BigData) | 979-8-3503-6248-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/BigData62323.2024.10825017

Abstract—**Effective fraud detection and analysis of government-issued identity documents, such as passports, driver's licenses, and identity cards, are essential in thwarting identity theft and bolstering security on online platforms. The accuracy of training fraud detection and analysis tools depends on the availability of extensive and diverse identity document datasets. However, current publicly available benchmark datasets for identity document analysis, including MIDV-500, MIDV-2020, and FMIDV, fall short in several aspects: they offer a limited number of samples of ten European country document types, cover insufficient varieties of fraud patterns, and seldom include alterations in critical personal identifying fields such as portrait images, limiting their utility in training models capable of detecting realistic frauds while preserving privacy. In response to these shortcomings, our research introduces a new benchmark dataset, IDNet, designed to advance privacy-preserving fraud detection efforts, synthesized by integrating the generative models and a Bayesian optimization approach. The IDNet dataset comprises $837,060$ images of synthetically generated identity documents, totaling approximately 490 gigabytes, categorized into 20 types from 10 U.S. states and 10 European countries, which is the largest identity document dataset publicly available today. We evaluated the fidelity and utility of IDNet to demonstrate the effectiveness of our unique synthetic data generation method. We also presented two use cases of the dataset, illustrating how it can aid in training privacy-preserving fraud detection methods, and facilitating the generation of camera and video capturing of identity documents.**

I. Introduction

The surge in digital platforms offering remote identity proofing has escalated concerns regarding the forgery of identity documents, including passports, driver's licenses, and identity cards. The Financial Crimes Enforcement Network (FinCEN) reported that in 2021, around 1.6 million Bank Secrecy Act (BSA) reports—constituting 42% of all reports filed that year—were related to identity fraud, highlighting $212 billion in suspicious transactions [1]. This issue poses risks in various sectors, including finance, healthcare, travel, retail, government, telecommunications, and gambling [2]. According to a recent industry analysis [2], fraudulent techniques have evolved from simple forgeries (e.g., name alterations) to advanced use of generative AI/ML techniques to create deceptive images (e.g., face morphing [3]). Since most remote identity validation services on digital platforms rely on images captured in white light rather than multi-spectral imaging techniques like near-infrared and ultra-violet light, this paper focuses on synthesizing datasets captured under white light conditions.

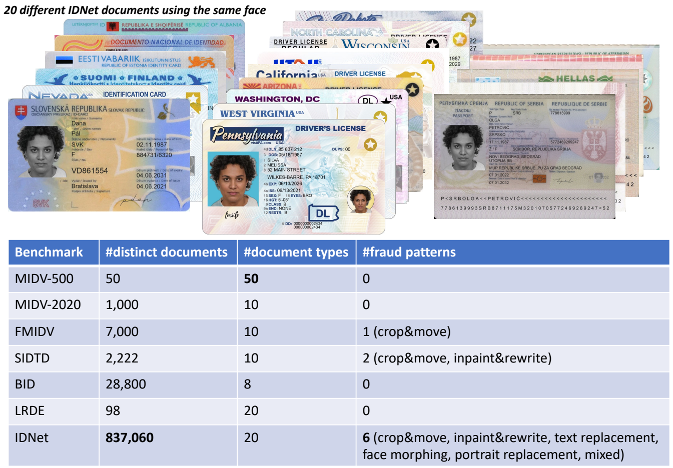
Despite the availability of several public datasets for identity document analysis, focusing on images taken in white light, such as MIDV-500 [4], MIDV-2019 [5], MIDV-2020 [6], FMIDV [7], and SIDTD [8], our examination has uncovered significant limitations in these resources.

- **Limited number of distinct complete samples:** Most existing datasets contain less than $1,000$ distinct identity documents. While these datasets may help develop tools for simple tasks such as optical character recognition (OCR), they are insufficient for training and testing AI/ML models for complicated tasks such as fraud detection. Although the BID dataset contains $28,800$ distinct identity documents, the portrait photos are blurred, which makes it unsuitable for critical tasks such as detecting face morphing and portrait substitution, where clear images are essential for accurate model performance.
- **Insufficient fraud patterns**: Only a few publicly available datasets, FMIDV [7] and SIDTD [8], which build upon the MIDV dataset, contain identity documents with fraudulent alterations. FMIDV presents a sole Copy-and-Move fraud pattern, where guilloche patterns are replicated and repositioned among documents. Conversely, SIDTD employs basic Crop-and-Move with inpainting techniques to simulate fraudulent activity. Nevertheless, fraud techniques such as face morphing, portrait substitution, and the intricate alteration of textual data remain unrepresented in these public collections. Crucially, as privacy issues take center stage in identity document management, the introduction of complex fraud patterns that intersect with extensive personal identifier information (PII), like portrait photos, ghost images, dates of birth, names, and addresses, is imperative for honing privacy-centric fraud detection methodologies. (If fraud patterns were not intruding upon PII fields, redacting these fields could help preserve privacy during model training, which is less challenging.) The creation and availability of a new benchmark dataset containing representative and challenging fraud patterns are pivotal for enhancing the precision and confidentiality aspects of fraud detection in complex scenarios.
- **Non-scalable synthetic data generation methodology**: The synthesis of identity documents has several challenges. (1)

* These authors made equal contributions

979-8-3503-6248-0/24/$31.00 ©2024 IEEE

The multi-modal challenge: An identity document consists of image and text segments; (2) **The semantic constraints**: Cross-segment constraints widely exist, e.g., the facial image must fit the fields such as sex, age, eye color, ethnicity group, etc., and the issue date must match the expiration date; and (3) **The unavailability of source datasets**: Due to privacy regulation, it is challenging to obtain a sufficiently large collection of real-world identity documents to train synthetic data generation models. Therefore, a few-shot synthetic data generation pipeline that only takes a few samples downloaded from government websites is highly desirable.


20 different IDNet documents using the same face

Benchmark	#distinct documents	#document types	#fraud patterns
MIDV-500	50	**50**	0
MIDV-2020	1,000	10	0
FMIDV	7,000	10	1 (crop&move)
SIDTD	2,222	10	2 (crop&move, inpaint&rewrite)
BID	28,800	8	0
LRDE	98	20	0
IDNet	**837,060**	20	**6** (crop&move, inpaint&rewrite, text replacement, face morphing, portrait replacement, mixed)

Fig. 1: Overview of IDNet. The face images and the text information in the ID cards are 100% artificially generated.

To address these limitations and challenges, this paper has made the following unique and significant contributions:

- **We created and open-sourced a novel synthetic identity document dataset called IDNet (Sec. IV)**, which contains $837,060$ documents from 20 types, with its overview illustrated in Fig. 1. Each document type has $5,979$ distinct non-fraud document samples with different facial images. For each such document sample, we generate six forged samples with different fraud patterns. The first two fraud patterns are included in existing datasets [8], which are (1) Crop-and-Move: Cropping a random field from one identity document and moving it to another document; and (2) Inpaint-and-Rewrite: Inpainting a random field and replacing the text in the field by using a different font style or size. We then created four popular patterns that haven't been implemented in any publicly available datasets: (1) face morphing that merges two distinct faces into one face; (2) portrait substitution by a disqualified portrait; (3) direct alterations in text fields, including random changes to text content, font, and background color schemes without inpainting, and (4) various combinations of these fraud patterns. The selection of these specific fraud patterns is informed by their prevalence in real-world fraud instances [2] and their intersection with personal identifier information (PII) fields, thereby posing a substantial challenge for research in privacy-preserving fraud detection. In total, we have $41,853$ document samples for each document type. All datasets are publicly available in Zenodo[1]. To our knowledge, IDNet is the most comprehensive public dataset of identity documents.
- **We designed a few-shot AI-assisted synthetic data generation pipeline for multi-modal identity documents (Sec. III).** The pipeline uses Stable Diffusion 2.0 [9] to remove portrait photos and other PII information from a few publicly available sample identity documents (e.g., released by the Department of Motor Vehicles (DMV)) to create templates of different types of identity documents. All portrait photos used to fill the templates are artificially generated using AI [10]. The pipeline includes a large language model (LLM) [11], ChatGPT-3.5-turbo, to generate data to fill in text fields, with constraints (e.g., DOB and sex should match with the photo) satisfied. The filling process adopts a divide-and-conquer approach that automatically searches for hyper-parameters, such as font size, style, color, and coordinates, for each image segment, guided by a Bayesian optimizer that maximizes the target quality metric. Our approach prioritizes the generation of identity documents that, while not intended for illicit use, are sufficiently authentic to support research demands.
- **We provide a quality framework to evaluate the generated data (Sec. VI)**. We evaluated the document fidelity (similarity to real-world documents) and the utilities of different fraud detection tasks. We also presented two use cases, showcasing how new opportunities and challenges arise with our IDNet dataset. The first use case compares standard privacy-preserving fraud detection techniques, such as *masking*, which involves obscuring sensitive information; and *Pixel-DP* [12], where pixel-level perturbation based on differential privacy (DP) is applied to entire images. We observed a notable reduction in the effectiveness of fraud detection when employing the existing privacy-preserving methods we tested, which illuminates the inherent challenges in designing privacy-preserving algorithms to balance accuracy and privacy. We further showed that IDNet can be used as a foundation to efficiently create a large-scale synthetic identity document dataset within various camera/video capturing environments, e.g., captured by different mobile devices, with different indoor/outdoor backgrounds.

II. Background and Related Works

A. Identity Documents

As illustrated in Fig. 2, an identity document usually contains (1) security features, (2) PII information, and (3) other information, which are explained as follows:

Security Features that we explore in this work focus on those that are amenable to digital capture and analysis under standard white lighting conditions, including barcodes, watermarks, micro-printing, guilloche (also known as rainbow printing), distinct color schemes, unique text font, barcodes, and the machine-readable zones (MRZ) [13] [14].

Personally Identifiable Information (PII) includes but are not limited to the portrait photo, signature, barcode, family name, given name, DOB, customer identifier, cardholder address, and ghost image. First, when generating the IDNet datasets, we must

[1] https://zenodo.org/search?q=IDNet&l=list&p=1&s=10&sort=bestmatch

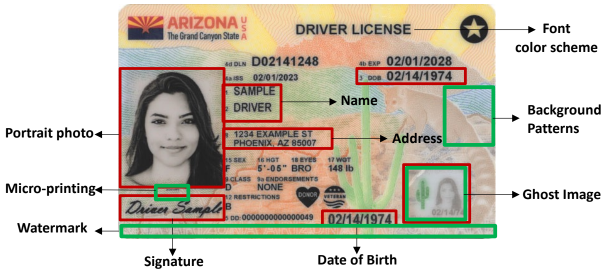


Fig. 2: Overview of identity documents – Arizona Drivers' License (downloaded from DMV website) as an example. Examples of PII and security features are highlighted in red and green respectively.

not disclose any PII information from the real world. Second, given the need to develop new privacy-preserving methods that prevent fraud detection or other analysis processes from disclosing PII information, a primary goal of designing our novel identity document benchmark dataset is to facilitate such analysis by providing fraud patterns that overlap with PII fields and pose challenges for privacy-preserving fraud analysis.

Other Information includes but is not limited to date of issue, date of expiry, document discriminator, endorsement, restrictions, date of first issue, separate expiry, name suffix, weight, height, sex, etc. These fields do not contain any personal information and can be disclosed during the analysis process.

B. Existing Public Identity Document Datasets

Many existing publicly available document datasets are designed for recognizing, classifying, and restoring information from documents captured as videos or photos using mobile devices. However, the identity documents used for producing these videos or photos are very limited. The SmartDoc dataset [15] contains a training set of 10 document samples captured as video clips. Each training sample also contains an image of the document used to produce the sample, which is considered the ground truth for comparison to the document restored from the video. In addition, SmartDoc offers a testing set of 37 document capture samples. The LRDE identity document image database [16] comprises 100 videos for a dozen different types of visas and passports from various countries using different backgrounds and smartphones. MIDV-500 [4] contains 500 video clips of 50 identity documents, including 17 ID cards, 14 passports, 13 driver's licenses, and 6 other identity documents. Each of the 50 document were used in 5 different backgrounds to generate 10 video clips using two mobile phone devices, targeting simple analysis tasks such as face detection, optical character recognition (OCR), and document type classification. MIDV-2019 [5] extended the MIDV-500 dataset to include four more videos with distorted identity documents and different lighting conditions for each identity document type. MIDV-2020 [6] increased the number of unique document samples to $1,000$ (100 unique documents for each of 10 document types).

Datasets featuring fraudulent identity documents remain scarce. FMIDV [7] addresses this gap by introducing seven forged IDs for each sample in the MIDV-2020 dataset, focusing on the guilloche-pattern fraud. To generate a forged ID, they randomly selected a few blocks only containing guilloche patterns from one ID, and copied these blocks to random locations in the blank area of another ID. FMIDV is limited to the single guilloche-related copy-and-move fraud pattern and overlooks many popular identity document fraud patterns. SIDTD [8] is the most recent extension of the MIDV-2020 dataset. It used crop-and-move and inpainting techniques to create simple frauds, containing 1222 fraud documents. Yet the MIDV family has a limited number of distinct document samples, ranging from 50 to 1000, which poses challenges for AI/ML applications in achieving high accuracy.

There exists a separate class of identity document benchmark datasets for detecting presentation attacks [17]–[19] distinguishing identity documents directly captured by phones from those photos capturing identity documents on the screen and printed on papers. For example, the KID34K dataset [17] is manually collected for classifying the photos of physical IDs from the photos of digital IDs displayed on screen or printed on paper. It used 37 Korean registration cards and 45 distinct driver's licenses, belonging to 46 non-existing people. They further used (1) 12 different smartphones to take $13,746$ different photos of these 82 ID cards, labeled as the genuine class; (2) eight displays to display these ID cards on different screens, which are further captured using smartphones as $13,729$ images, labeled as the screen class; and (3) two printers to print these ID cards on papers, which are then captured using smartphones as $7,187$ images, labeled as the print class. These datasets did not contain any fraud patterns that alternate the ID cards, covered only a few types of IDs, and were orthogonal to our work. Our purpose is to automatically synthesize a large number of identity document samples belonging to diverse non-existent people as well as fraud samples. Our dataset can be further displayed on screens, or printed out on papers, and then captured as photos to augment existing datasets.

C. Existing Synthetic Identity Document Generation Methods

Most of these publicly available datasets are created manually, except that FMIDV and SIDTD used the inpaint-and-rewrite, and the crop-and-move techniques to generate fraud samples, and Benalcazar et al [19] proposed two synthetic identity document generation approaches. The first approach is to generate information and fill the information into an identity document template. The second approach is to train a StyleGAN2-ADA [20] model to generate identity documents. Although StyleGAN2-ADA is well-known for its capability of generating data with a limited number of examples [20], Benalcazar et al [19] used more than $2,000$ images in each class to train the model, due to the learning challenges brought by the multi-modal nature of the images [19]. Yet the accuracy achieved on the synthetic datasets is suboptimal as demonstrated in the paper. In this paper, we proposed an AI-assisted and quality-driven methodology to create a large-scale synthetic identity document dataset, which only requires one training sample. In addition, we also demonstrate that our

IDNet dataset outperformed the existing datasets on multiple quality metrics while facilitating several critical use cases.

III. Methodology

A. System Overview

Our synthetic data generation targets several unique goals: (1) **Few-shot generation.** The generation process should work well if only one real-world identity document (e.g., a sample Arizona drivers' license downloaded from the website of Arizona MVD) is available for each type of document (e.g., Arizona drivers' licenses). (2) **Quality-driven generation.** The generation process should be guided and calibrated by data quality. The quality metrics used for calibration should be customizable by users. (3) **Privacy-preserving generation** (100% synthesis). Except for the input of one real-world sample for each type of document, the data generation pipeline should not access private data or AI/ML models trained on private data [2], or use any private data to fine-tune AI/ML models pre-trained on public data. In addition, the generated data should not disclose personal information in the training sample. (4) **Cost-effective generation.** Given that users may frequently update the input document, the attribute constraints, and the quality metrics used for calibrating the generation process, it is important to automate the synthetic data generation pipeline, and we want to minimize the expensive data labeling, data preprocessing, and training overheads.

To meet the goals, our synthetic identity document generation pipeline starts by generating document templates and metadata. Then, it leverages a quality-driven Bayesian optimization algorithm to finetune the hyper-parameters that determine how to fit the artificially generated facial images and metadata into the template, such as font style, font size, field coordinates, image background color, etc, and use the fine-tuned parameters to generate the final identity document. In addition, we apply various alternations to those documents to create fraud samples.

B. Automatic Template Generation Using Diffusion Model.

To create the IDNet dataset, it is essential first to acquire a template for each type of identity document. However, high-quality, blank templates of real-world documents are generally unavailable. The existing template-based approach requires manual extraction of the template from an identity document sample using PhotoShop [19]. To automate this process, we utilize image generative models, such as diffusion models [22] [23] [24], to produce our ID templates by erasing the content from actual identity documents, employing the Stable Diffusion version 2.0 from Hugging Face [3], which is based on the Latent Diffusion Model [23]. This model is adept at editing masked areas of an input image in accordance with text prompts. For our purposes, we mask all customizable information on the IDs and direct Stable Diffusion with the prompt ”remove all texts/photos in the masked areas.” Consequently, the model adeptly eliminates customized data from the document and replenishes the masked sections with appropriate backgrounds. Using this approach, we created templates for 20 different types of real-world IDs (See Sec. IV for a list of the types).

C. Constraint-Aware Metadata Generation

We synthesize the metadata information to be filled in the identity document templates, which fall into two categories:
Portrait Images. We used a well-known public face image dataset for academic research[4], which was widely used, e.g., it was also used for generating the MIDV benchmark family. All faces in the dataset are synthetically generated. The dataset consists of $10,000$ synthetic human face images with metadata such as face landmarks, sex, emotion, ethnicity, eye color, age, facial expression, etc.
Text Fields. Different types of identity documents have different text fields. For example, the United States driver's license usually includes first and last names, date of birth, issue date, expiration date, sex, eye color, height, weight, driver's license class, number, discriminator, and address. In addition, different types of documents from different countries may use different languages. We implemented a tool to generate such information. Different from existing tools such as Faker [25], which generate information randomly. We allow the users to specify a type and a value distribution for each attribute. In addition, our tool is aware of user-specified constraints.

Each different type of identity document is associated with a set of constraints falling into two categories (1) Single-attribute constraints, e.g., some types of documents (e.g., US Drivers' Licence) are only issued to adults, so DOB should follow the rule, and the face portrait has constraints on age, facial expression, head pose, and wearings; (2) Cross-attribute constraints, e.g., the sex, ethnicity group, age, and eye color should match the portrait photo; the name should match the sex and ethnicity group; and the difference between the expiration and the insuring date should be a fixed value. To represent such semantic constraints, we allow users to provide a graph to represent semantic constraints, in which each node represents a metadata attribute, e.g., portrait.age, portrait.emotion, portrait.headpose, portrait.wearing, date_of_birth, first_name, last_name, eye_color, issue_date, expiration_date, etc., and each edge represents a filter predicate to express the constraints that are defined over the two attribute nodes connected by the edge, e.g., expiration_date - issue_date $= 5$. A cycle edge that starts and ends with the same node is also allowed, e.g., portrait.age $>= 18$, portrait.emotion.happiness < 0.8, portrait.headpose.roll > 9, date_of_birth < 1981. In our experiments, the graph is manually created by domain experts with the assistance of LLM and retrieval augmented generation (RAG) by retrieving such constraints from the vector database (implemented in Faiss [26]) that stores the embedding vectors of chunks of identity document design standards such as ICAO 9303 standard [27]. For example, using RAG helps us discover constraints such as ”the validity of an Arizona driver's license is 5 years for people above 65 years old”.

[2] We assume publicly available LLMs such as ChatGPT from OpenAI is subject to privacy regulation and is free from private information [21].

[3] https://huggingface.co/stabilityai/stable-diffusion-2

[4] https://generated.photos/

By applying the semantic constraints, among the $10,000$ synthetic portrait photos, only $5,979$ photos are qualified. Given a portrait photo, we randomly generate other metadata information for a non-existing person following the user-customizable distributions on each attribute while ensuring that all constraints specified for the corresponding document type are met by verifying the sampled information using the rules obtained by traversing the graph.

D. Quality-driven Hyper-Parameter Calibration

All generated synthetic information needs to be added to the corresponding template to generate the final document. This process involved many hyper-parameters, e.g., font size, font style, font color, interval spaces between two characters and two words, and the positions of each text field value. Manual tuning of these parameters is tedious and time-consuming, significantly increasing the cost of generating identity documents.

To address the problem, we proposed a novel quality-driven hyper-parameter tuning algorithm that automatically optimizes the hyper-parameters to maximize the quality of the generated document using Bayesian optimization [28]. In our implementation, we used the Structural Similarity Index (SSIM) [29] between the generated image and the ground-truth image (which was used to extract the template) as our quality metric. SSIM is a popular metric that quantifies the visual similarity between two images by considering luminance, contrast, and structure. SSIM provides a robust measure for evaluating the quality of the generated image, ensuring it closely matches the original in terms of appearance and detail.

Every text field has multiple parameters and each parameter has many possible values. For example, 1714 font styles are publicly available in Google Fonts, while font sizes, colors, interval spaces, and field positions are all numerical numbers, enlarging the search space.

To overcome the challenge, we adopt the idea of divide and conquer and Bayesian optimization. We co-partition the ground-truth image and its extracted template into segments by text fields, with each field being a segment. For each segment, we finetune the hyper-parameters to maximize the SSIM between the generated and ground-truth segments using the Bayesian optimization algorithm. We use the segments from the real-world sample, from which we extract the template (Sec. III-B), as ground truths. In the optimization process, for each text field, it first trains a surrogate model that predicts the resulting SSIM value for a candidate combination of hyper-parameters by sampling and evaluating combinations of hyper-parameters repeatedly. After that, it selects a combination of parameters estimated to maximize the SSIM metric using the surrogate model, generates the document by filling in the metadata information of the ground-truth segment using the selected parameters, measures the resulting SSIM, and updates the surrogate model accordingly. We repeat the process until it converges or reaches the maximum number of iterations.

E. Document Generation

Once we calibrated the hyper-parameters, we use them to fill in the generated metadata (Sec. III-C) to the text fields of the template. The parameters of portrait photos (e.g., size, background color, and bounding box) are more static than text fields. They are directly extracted from the real-world image while generating the template. In addition, many types of identity documents encompass the ghost image (serving both as a security feature and additional PII) and the signature of the ID holder: (1) The generation process for the ghost image commences with the removal of the portrait photo's background, followed by its conversion into a single-channel image retaining only the luminosity level data. This transformation effectively shifts the image to grayscale, closely mirroring the visual characteristics of a genuine ID's ghost image. The resultant ghost image is then seamlessly integrated into the template. (2) A specialized randomizer algorithm was developed for signature generation. It first hashes the ID holder name into a reasonably compact string, which is subsequently applied to the template using a randomly selected font style, emulating various handwriting styles. This dual process of name hashing and font randomization ensures diverse representations of individuals' handwriting styles across the dataset.

F. Fraud Pattern Generation.

A pivotal component of our IDNet dataset involved forged identity documents. First, we incorporated both fraud patterns provided by the SIDTD dataset [8]: (1) Inpaint-and-Rewrite Fraud Pattern. One text field is randomly selected from all the available fields on the ID. A realistic mask is then applied to this specific region containing the field, while the font style for the replacement text is chosen randomly from the fonts available (2) Crop-and-Move Fraud Pattern. In this method, a field is selected randomly from one ID and then cropped and replaced with the field of another ID. In both cases, the fields are selected randomly, with a 95% probability that the same field is chosen in both IDs and a 5% probability that different PII fields are selected, following SIDTD.

Additionally, we provide unique fraud patterns as follows: **Face-Morphing Fraud** [3], [30], which morphed the characteristics from two faces into one face, recently emerged as a notable threat [2]. This type of fraud leverages the natural variations in human facial features over time, creating opportunities for identity deception. It operates on the premise that an individual's facial characteristics can significantly alter from those documented on their official identification. This variance enables an attacker (referred to as Person A) to misuse the identification of another individual (Person B), provided there is a sufficient resemblance between their facial features [3], [30]. We adopted cutting-edge image fusion methods to integrate this complex fraud pattern into our analysis, including Image Warp and Cross Dissolve [31]. To achieve high-quality morphing as suggested by the NIST face morphing report [32], only the face area of the facial images was averaged after alignment and feature warping. In addition, the face area was adjusted to the face color histogram of the first input facial image. For each of the 5979 qualified artificially generated photos, we morphed it with another randomly selected photo of the same ethnicity and sex. We set the blending factor as 0.5 in the face morphing process following the NIST face morphing implementation [32], which suggests that both faces contribute equally to the morphed face.

Portrait Substitution Fraud. Based on industrial studies [2], a significant portion of digital identity document attacks on online platforms in 2023 are at low forging costs. Portrait substitution is one example, which uses a disqualified photo to replace the original photo. To implement this type of fraud, for each identity document, we uniformly sampled one photo from the $4,021$ portrait photos identified as disqualified (Sec. III-C). We then used that sampled photo to replace the original photo and produce a forged identity document.

More Complicated Text-Field Replacement Fraud. This is a more complicated alternative to the inpaint-and-rewrite fraud. Replacing text fields often not only leads to subtle alterations in the text content, font styles, and sizes, as covered in the inpaint-and-rewrite fraud, but also affects the background colors of PII fields. Therefore, we randomly changed the text content, font style, font size, and the color, contrast, and saturation of PII fields, such as first and last names, sex, DOB, expiration data, etc. We further classified this fraud into two levels (1) easy-level, where the replacing text fields' information (e.g., name, DOB) does not match the portrait photo's sex and age, which is relatively easier to detect; and (2) hard-level, where the forged information is consistent with the rest of the identity document information. The ratio of easy-level fraud was determined to be around 65%, and the rest were all hard-level frauds, consistent with a recent industry survey [2].

Mixed Fraud Pattern. It mixed fraud patterns through the following steps. For each non-fraud sample, the text replacement fraud was first applied to some of the text fields. Subsequently, a random variable was sampled to decide whether to apply a face morphing or portrait substitution fraud to the portrait photo of the document. This dataset can be used to test the integration of specialized fraud detection algorithms.

IV. The IDNet Dataset

Leveraging the cutting-edge AI-assisted pipeline, outlined in Sec. III, we have developed an identity document dataset, named *IDNet*, which is entirely synthetically generated and devoid of any private information. This dataset encompasses a total of $837,060$ identity documents, spanning across 20 different document types, including driver's licenses or ID card from 10 US states including Arizona (AZ), California (CA), Nevada (NV), North Carolina (NC), Pennsylvania (PA), South Dakota (SD), Utah (UT), Washington D.C. (DC), West Virginia (WV), and Wisconsin (WI), and passports or ID card 10 European countries including Albania (ALB), Azerbaijan (AZE), Estonia (EST), Finland (FIN), Greece (GRC), Latvia (LVA), Russia (RUS), Serbia (SRB), Slovakia (SVK), and Spain (ESP). The templates of identity documents from US states and European countries are extracted from sample documents downloaded from US state local government websites and the MIDV-family benchmark respectively.

Each document type has $5,979$ unique document samples. Each sample comprises one authentic copy alongside six fraudulent variations, including face morphing, portrait substitution, text alteration, a combination of these frauds, and two fraud patterns (inpaint-and-rewrite and crop-and-move) from the SIDTD benchmark, as detailed in Sec. III-F. IDNet stands as the largest publicly accessible identity document dataset to date, as depicted in Fig. 1.

The IDNet dataset is released on Zenodo 1 and has been downloaded for 2800+ times since it was released in Feb 2024 until Nov 2024. It includes identity document image files along with JSON files that describe the metadata of each document. The metadata for each positive sample (w/o fraud patterns) includes the document identifier, face image ID, name, sex, date-of-birth (DOB), etc., and the corresponding hyper-parameter information. The metadata for each fraud identity document contains additional information, such as the fraud type and fraud parameters. For portrait substitution, we recorded the identifier of the original face, and the identifier of the new face. For face morphing [3], we documented the ID and the morphing weight of each morphed face.

V. Ethical Considerations

As described in Sec. III and Sec. IV, all the data used in our dataset comes either from public source (e.g., portrait images and samples) or synthetically generated. The dataset is available on Zenodo and only for research purpose.

VI. Evaluation

Our evaluation focuses on two questions: R1. The quality of the IDNet dataset; and R2. The efficiency and effectiveness of our few-shot and quality-driven synthetic datasets.

A. Evaluation Metrics

To evaluate the quality of the synthetic dataset, we consider a comprehensive set of metrics as follows:

Fidelity. It is defined as the similarity between the generated and the original dataset. It is measured as the SSIM between the original real-world identity document and the document generated by our synthetic data generation pipeline using the template, and the portrait photo and text values extracted from the original document.

Task Utility. Given an AI/ML model learned on IDNet for a learning task, the testing accuracy of the model on real-world dataset defines the utility of the task. In this work, we have considered measuring the utility of IDNet for two different learning tasks: detection of the face morphing fraud, and the fraud patterns shared with SIDTD (i.e., inpaint-and-rewrite, and crop-and-move). We chose these tasks because of the availability of external benchmark datasets for these frauds, which made the measurement possible.

We also evaluated other metrics such as diversity and stealthiness, which are omitted from this paper due to space limitations. You can find more details in our extended version [33].

B. Evaluation of IDNet Data Quality

1) Fidelity: We compared the fidelity of the generated images of 20 types of identity documents with hyper-parameters selected by Bayesian optimization (Sec. III-D) and hand-tuned by a domain expert. For the Bayesian optimization algorithm, the maximum number of iterations is set to 1000. As illustrated in Fig. 3, our hyper-parameter tuning method improved the SSIM by up to 12%. With Bayesian optimization, for these identity document types from U.S. states, the measured SSIM ranges from 0.935 (for SD) to 0.980 (for PA and CA). However, the fidelity for the identity document types from European countries is significantly lower, ranging from 0.801 (for EST) and 0.889 (for ALB), which is because of the inconsistent color schemes, RGB (red, green, blue) vs. CMYK (cyan, magenta, yellow, black), between our generated images and the images from the MIDV-family.

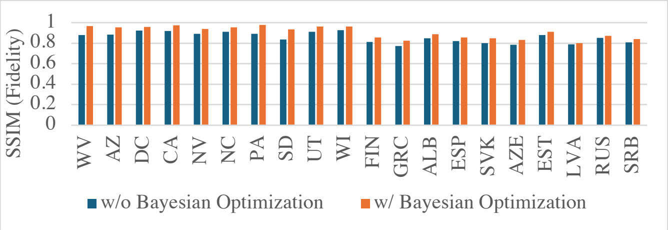


Fig. 3: Comparison of Fidelity w/o (expert hand-tuning) and w/ Hyper-Parameter Tuning using Bayesian Optimization.

2) Utility: We measured the utility of IDNet for three different fraud detection tasks, described as follows:

Detection of Face Morphing. IDNet used a publicly available synthetic portrait photo dataset [10]. We evaluated its quality using the utility of the face morphing task. Leveraging the availability of real-world face morphing data, we evaluated the utility of the artificially generated portrait photos dataset [10] used in our IDNet dataset from two perspectives, which are (1) training models on synthetic data and then assessing their performance on real-world data, and (2) analyzing whether models maintain their relative performance rankings when trained on both synthetic and real-world data [34].

TABLE I: Utility (detection accuracy) evaluation results of face morphing detection models trained on IDNet and real-world datasets.

Training Dataset	Models	Test Datasets (%)			
		IDNet (synthetic)	FRLL (real-world)	FERET (real-world)	FRGC (real-world)
IDNet (synthetic)	SPL-MAD	92.8	87.3	74.8	77.3
	PW-MAD	97.6	89.7	78.0	80.5
	Inception	98.2	90.7	78.4	81.6
	MixFaceNet-S	98.7	92.8	79.1	82.0
FRLL (real-world)	SPL-MAD	84.1	90.5	74.2	77.5
	PW-MAD	87.3	97.8	76.0	80.0
	Inception	89.5	96.8	77.9	81.2
	MixFaceNet-S	89.7	98.3	78.4	81.3
FERET (real-world)	SPL-MAD	87.5	87.5	80.9	77.9
	PW-MAD	92.3	93.2	84.3	82.3
	Inception	93.5	94.3	84.5	83.0
	MixFaceNet-S	94.0	94.6	85.2	83.9
FRGC (real-world)	SPL-MAD	86.9	88.9	75.1	82.1
	PW-MAD	91.9	92.8	79.2	85.6
	Inception	92.5	93.5	79.8	85.8
	MixFaceNet-S	92.8	93.3	80.5	87.2

We evaluated how models trained on IDNet transfer to real-world datasets in terms of face morphing detection accuracy. This study assumes that models trained on synthetic datasets will be directly applied to real-world applications. For comparison, we also tested how models trained on real-world datasets transfer to other real-world datasets.

In our evaluation, we used three real-world face morphing detection datasets [35], FRLL [36], FERET [37], and FRGC [38]. We used IDNet's morphed face dataset as mentioned in Sec. III-F, which applied FaceMorpher [35] to morph each IDNet's qualified portrait photo with another photo randomly selected from the same dataset using a morph percentage of 0.5. We split each morphed face dataset into a training set containing 80% of the data and a test set containing 20% of the data. All the facial images are cropped using the MTCNN [39] with an extension rate of 5% to include the whole face following [40]. We implemented four face morphing detection methods, PW-MAD [40], Inception [41], MixFaceNet-S [42], and SPL-MAD [43]. The evaluation results on real-world datasets FRLL, FERET, and FRGC, using face morphing detection models trained on IDNet's face morphing dataset, are shown in Table I. It illustrates that despite the performance drops due to differences in background, brightness, etc., the models trained on IDNet maintained acceptable accuracy on the real-world datasets compared with models trained on real-world datasets. In addition, models trained on IDNet achieved similar domain-transfer performance with models trained on some real-world datasets.

TABLE II: Utility (detection accuracy) of the models trained on IDNet and SIDTD for detecting inpaint-and-rewrite and copy-and-move frauds for Finland ID cards.

Training Dataset	Models	Test Datasets	
		IDNet-Finland	SIDTD-Finland
IDNet	EfficientNet-b3	99.9	96.3
	ResNet18	99.4	93.7
	ResNet50	99.7	95.0
SIDTD	EfficientNet-b3	67.2	100.0
	ResNet18	67.2	95.5
	ResNet50	67.2	100.0

TABLE III: Utility (detection accuracy) of the models trained on IDNet and SIDTD for detecting inpaint-and-rewrite and copy-and-move frauds for Greece passports.

Training Dataset	Models	Test Datasets	
		IDNet-Greece	SIDTD-Greece
IDNet	EfficientNet-b3	99.4	95.0
	ResNet18	98.4	96.8
	ResNet50	100.0	92.3
SIDTD	EfficientNet-b3	67.7	100.0
	ResNet18	67.7	100.0
	ResNet50	67.7	95.2

Detection of Inpaint-and-Rewrite, and Crop-and-Move. To evaluate the synthetic fraud data generated from the text-field fraud replacement pattern, we focus on the inpaint-and-rewrite pattern and the crop-and-move pattern, since they were also used in SIDTD [8] and the utility of the fraud detection models trained on IDNet can be evaluated on SIDTD, and vice versa. We selected two European identification documents, the Finland ID and the Greece passport, for evaluation, as both document templates have been used in IDNet and SIDTD [8]. (The results on identity documents from other European countries shared by the two datasets are similar.) For each document type, our IDNet dataset contains 5979 positive samples (w/ fraud) for each fraud pattern, and 5979 negative samples (w/o fraud). However, SIDTD only contains 100 negative samples for each of 10 types, and 1078 and 144 positive samples of the inpaint-and-rewrite and the crop-and-move patterns for all 10 types, respectively. Particularly, there are 112 and 10 positive samples of the inpaint-and-rewrite fraud and the copy-and-move fraud for the Finland ID card, and 109 and 12 positive samples of the inpaint-and-rewrite and the copy-and-move fraud, for the Greece passport.

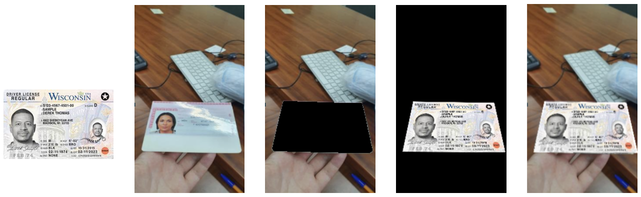
Tables II and Table III illustrate the utility comparison of fraud detection models (with three popular architectures: EfficientNet, ResNet18, and ResNet50) trained on IDNet and SIDTD. We can interpret that when the model is trained on IDNet datasets and tested against the corresponding SIDTD datasets, the model successfully captures the fraud patterns of SIDTD. However, in the reverse scenario, the model fails to capture the underlying fraud patterns.

C. *Cost and Efficiency of IDNet Data Generation*

We measured the time and costs spent at every pipeline stage, as illustrated in Tab. IV. The pipeline ran on a system with dual Intel Xeon Gold 6226 CPUs at 2.70GHz with 24 cores each, four Nvidia GeForce 2080 Ti GPUs, and 196 GB memory to produce the identity documents. (We estimate its cost to be $2 per hour based on the costs of AWS EC2 on-demand instances of similar capabilities [44].) The stable diffusion 2.0 model is free and open-sourced. The ChatGPT-3.5-turbo API costs $3 and $6 for 1M input and output tokens (according to OpenAI pricing in Aug 2024). All generated documents are archived using the free Zenodo service. As a result, the operational cost for producing each identity document is lower than $0.0001 with a latency of 0.14 second. It demonstrated the cost-effectiveness of the proposed pipeline.

VII. Use Cases

A. *ID replacement in user-defined backgrounds*



(a) ID (b) Background (c) Masked BG (d)Transformed ID (e) Final image

Fig. 4: ID replacement process examples. The face images and the text information in the ID cards are 100% artificially generated.

One use case of the IDNet dataset is to generate mobile documents (i.e. camera captured pictures and videos that contain ID documents) with user-specified parameters, such as camera model, indoor/outdoor environment, background objects, and lighting condition. In this section, we present a simple technique to create such a dataset by replacing an ID document in an existing picture with an identity document from IDNet as shown in Figure 4(e). This necessitates ensuring that the inserted ID card appears natural and consistent within the context of the original image, which involves addressing several technical challenges: accurate detection and localization of the existing ID card, precise segmentation to isolate the ID card from the background, alignment of the new ID card to match the perspective of the original, and blending to ensure a smooth transition between the new ID card and the background. To meet these challenges, we employ a series of advanced models and image processing techniques. Grounding DINO [45], a state-of-the-art model that combines object detection and language grounding, is employed to accurately identify and localize the ID card within the background image. Following this, in Figure 4(c) the Segment Anything Model (SAM) [46] is employed to segment the detected ID card from the background. SAM uses a promptable interface to identify and provide a precise mask of the ID card, facilitating accurate alignment of the new ID image. In Figure 4(d), a perspective transformation is then applied to align the new ID image with the segmented area of the original ID card. This transformation adjusts the new ID image to fit naturally within the segmented region, ensuring correct perspective, alignment and proportion. Finally, an image blending process using Laplacian pyramids merges the transformed ID image with the background image seamlessly.

B. *Privacy-Preserving Fraud Detection*

In releasing our novel dataset for identity document analysis and fraud detection, it is important to consider the privacy concerns. The dataset's intrinsic value lies not only in its comprehensive coverage and potential to revolutionize fraud detection but also in the sensitive nature of the information it encompasses (e.g., the photo identity, name, and so on). To responsibly harness this value while safeguarding individual privacy, the design and deployment of a privacy-preserving algorithm is not just beneficial but imperative. Thus, in this section, our goal is to evaluate the performance of our dataset when it meets with the current privacy-preserving algorithm.

To achieve this goal, we selected two standard privacy-preserving algorithms: Masking [47]and PixelDP [12]. The algorithm details and settings are described as follows:

Masking. As shown in Figure5 (b), we applied masks to those regions that contain sensitive information, *i.e.* zeroing the pixel values. While this approach, which is also termed redaction, has been widely used [47], it leads to information loss and may disable analysis tasks that rely on the redacted information.

PixelDP [12] is a robust privacy-preserving method via adding noise sampled from a specific distribution to the input or intermediate features. We focused on the input-level PixelDP, directly applying Gaussian noise to the input examples.

TABLE IV: Breakdown of Time and Monetary Costs of the Proposed IDNet Pipeline.

	Template	Metadata	Portrait preprocess	Parameter tuning	Filling	Total	Avg (per doc)
Time cost (seconds)	231	555	13,153	4,040	103,249	**121, 228**	**0.14**
Monetary cost ($)	0.12	0.3	7.3	2.2	59.5	**69.5**	**0.00008**

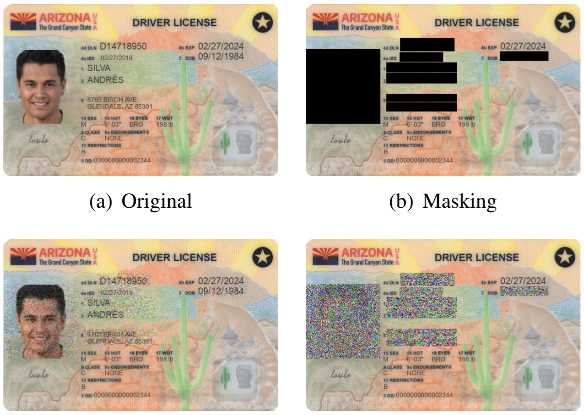
TABLE V: Fraud detection on Arizona Driver Licenses.

Task	Privacy-Preserving Method			
	None	Masking	PixelDP (L=0.1)	PixelDP (L=1.0)
Face Morphing	98.1	50.0	88.3	52.7
Portrait substitution	88.8	50.0	91.3	89.1
Mixed Fraud Face	88.8	50.0	84.3	70.4
Text Replacement	100.0	50.0	100.0	50.0

Specifically, We set $\epsilon = 1.0$ and $\delta = 0.05$, which provides $(1.0, 0.05)$-DP guarantee. The perturbations are sampled from a Gaussian distribution with zero mean and standard deviation $\sigma = \sqrt{2\ln(\frac{1.25}{\delta})}\Delta_{p,2}L/\epsilon$, where we use L_2-norm (*i.e.* $p = 2$) and $\Delta_{2,2} = 1$. With larger L, the perturbations will be more significant, and thus distort more information.

We evaluated these algorithms on four fraud ID detection tasks, including face morphing detection, portrait substitution detection, mixed fraud face detection, and text replacement detection. For the first three tasks in the face regions, we used a MixFaceNet-S [48] model as the detector, with $embedding\_size = 128$, $width\_scale = 1.0$, $gdw\_size = 1024$, and shuffling disabled. For text replacement detection, we used a simple convolutional network with 5 convolutional layers and 2 fully-connected layers. The validation dataset consists of 1000 positive and negative examples.

The results on Arizona Driver License are shown Table V, and the example images are illustrated in Figure 5. The results on other document types are similar as shown in our extended version of the paper [33]. Directly masking almost disabled the fraud detection capability, although it completely preserved the privacy of the sensitive information. That's because the helpful representations for fraud detection were also completely erased by masking, and the detector learned nothing from the training examples. Meanwhile, PixelDP with smaller L barely degraded the fraud detection performance (and even slightly improved it, serving as a kind of data augmentation), while the sensitive information were still highly recognizable. With larger L, the sensitive information werebetter concealed, at the cost of a nonnegligible fraud detection performance drop, since the helpful representations for fraud detection were also distorted (except for portrait substitution detection, which is a relatively easier task.) From the results of two baseline privacy-preserving algorithms, Masking and PixelDP, we observed a significant gap in their capabilities in balancing utility and privacy. They cannot achieve satisfactory fraud detection performance while protecting sensitive information from being recognized. How to simultaneously conceal sensitive information and keep helpful representations for fraud detection is yet an unsolved conundrum. Our released dataset can serve as a new benchmark, bringing new challenges for privacy-preserving algorithms.



(a) Original (b) Masking

(c) PixelDP ($L = 0.1$) (d) PixelDP ($L = 1.0$)

Fig. 5: Arizona Driver License examples. The face images and the text information in the ID cards are 100% artificially generated.

VIII. Conclusion

In this study, we introduce IDNet, a vast and comprehensive benchmark dataset comprising $837,060$ synthetic identity documents across 20 distinct types, aimed at facilitating research in identity document analysis and privacy-enhanced fraud detection. Combining quality-driven hyper-parameter tuning using Bayesian optimization and recent breakthroughs in generative AI, we developed a cost-effective methodology for generating representative document templates and identities. Additionally, we meticulously crafted a set of representative fraud patterns designed to intersect with personal identifier information. This intersection presents significant challenges for the development of privacy-preserving fraud detection mechanisms, as evidenced by our detailed evaluations. In addition, the IDNet dataset can be used to generate mobile identity document images captured using different mobile devices and lighting conditions.

To the best of our knowledge, IDNet stands as the most extensive publicly accessible synthetic dataset for identity document benchmarks to date. Looking ahead, we plan to expand IDNet by extending its applications and enhancing the explainability of the quality evaluation framework.

IX. Acknowledgements and Disclaimer

This material is based upon work supported by the U.S. Department of Homeland Security under Grant Award Number 17STQAC00001-07-00.

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.

References

[1] "Fincen issues analysis of identity-related suspicious activities." https://www.fincen.gov/news/news-releases/fincen-issues-analysis-identity-related-suspicious-activity.

[2] "Identity fraud report 2024." https://onfido.com/landing/identity-fraud-report/, 2023.

[3] S. Venkatesh, R. Ramachandra, K. Raja, and C. Busch, "Face morphing attack generation and detection: A comprehensive survey," *IEEE transactions on technology and society*, vol. 2, no. 3, pp. 128–145, 2021.

[4] V. V. Arlazarov, K. B. Bulatov, T. S. Chernov, and V. L. Arlazarov, "Midv-500: a dataset for identity document analysis and recognition on mobile devices in video stream," *Computer Optics*, vol. 43, no. 5, pp. 818–824, 2019.

[5] K. Bulatov, D. Matalov, and V. V. Arlazarov, "Midv-2019: challenges of the modern mobile-based document ocr," in *Twelfth International Conference on Machine Vision (ICMV 2019)*, vol. 11433, pp. 717–722, SPIE, 2020.

[6] B. K. Bulatovich, E. E. Vladimirovna, T. D. Vyacheslavovich, S. N. Sergeevna, C. Y. Sergeevna, M. Zuheng, B. Jean-Christophe, and L. M. Muzzamil, "Midv-2020: a comprehensive benchmark dataset for identity document analysis," *Computer Optics*, vol. 46, no. 2, pp. 252–270, 2022.

[7] M. Al-Ghadi, Z. Ming, P. Gomez-Krämer, J.-C. Burie, M. Coustaty, and N. Sidere, "Guilloche detection for id authentication: A dataset and baselines," in *IEEE 25th International Workshop on Multimedia Signal Processing*, 2023.

[8] C. Boned, M. Talarmain, N. Ghanmi, G. Chiron, S. Biswas, A. M. Awal, and O. R. Terrades, "Synthetic dataset of id and travel document," *arXiv preprint arXiv:2401.01858*, 2024.

[9] L. Yang, Z. Zhang, Y. Song, S. Hong, R. Xu, Y. Zhao, W. Zhang, B. Cui, and M.-H. Yang, "Diffusion models: A comprehensive survey of methods and applications," *ACM Computing Surveys*, vol. 56, no. 4, pp. 1–39, 2023.

[10] "Academic dataset by generated photos." https://generated.photos/datasets.

[11] J. Wei, Y. Tay, R. Bommasani, C. Raffel, B. Zoph, S. Borgeaud, D. Yogatama, M. Bosma, D. Zhou, D. Metzler, *et al.*, "Emergent abilities of large language models," *arXiv preprint arXiv:2206.07682*, 2022.

[12] M. Lecuyer, V. Atlidakis, R. Geambasu, D. Hsu, and S. Jana, "Certified robustness to adversarial examples with differential privacy," in *2019 IEEE symposium on security and privacy (SP)*, pp. 656–672, IEEE, 2019.

[13] "Aamva dl and id card design standard (2020)." https://www.aamva.org/assets/best-practices,-guides,-standards,-manuals,-whitepapers/aamva-dl-id-card-design-standard-(2020), 2020.

[14] "Anti-counterfeiting technology guide." https://euipo.europa.eu/tunnel-web/secure/webdav/guest/document_library/observatory/documents/reports/2021_Anti_Counterfeiting_Technology_Guide/2021_Anti_Counterfeiting_Technology_Guide_en.pdf, 2021.

[15] J. Chazalon, P. Gomez-Krämer, J.-C. Burie, M. Coustaty, S. Eskenazi, M. Luqman, N. Nayef, M. Rusiñol, N. Sidere, and J.-M. Ogier, "Smartdoc 2017 video capture: Mobile document acquisition in video mode," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, vol. 4, pp. 11–16, IEEE, 2017.

[16] M. Ô. V. Ngoc, J. Fabrizio, and T. Géraud, "Saliency-based detection of identy documents captured by smartphones," in *2018 13th IAPR international workshop on document analysis systems (DAS)*, pp. 387–392, IEEE, 2018.

[17] E.-J. Park, S.-Y. Back, J. Kim, and S. S. Woo, "Kid34k: A dataset for online identity card fraud detection," in *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pp. 5381–5385, 2023.

[18] R. Mudgalgundurao, P. Schuch, K. Raja, R. Ramachandra, and N. Damer, "Pixel-wise supervision for presentation attack detection on identity document cards," *IET Biometrics*, vol. 11, no. 5, pp. 383–395, 2022.

[19] D. Benalcazar, J. E. Tapia, S. Gonzalez, and C. Busch, "Synthetic id card image generation for improving presentation attack detection," *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 1814–1824, 2023.

[20] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila, "Training generative adversarial networks with limited data," *Advances in neural information processing systems*, vol. 33, pp. 12104–12114, 2020.

[21] "Open-ai privacy policy." https://openai.com/policies/row-privacy-policy/.

[22] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.

[23] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022.

[24] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," *arXiv preprint arXiv:2010.02502*, 2020.

[25] "Faker the perfect python package to generate fake data." https://medium.com/@mohamedmeqlad9/faker-the-perfect-python-package-to-generate-fake-data-6f43fa168e86.

[26] M. Douze, A. Guzhva, C. Deng, J. Johnson, G. Szilvasy, P.-E. Mazaré, M. Lomeli, L. Hosseini, and H. Jégou, "The faiss library," *arXiv preprint arXiv:2401.08281*, 2024.

[27] "Icao 9303: Machine readable passports." https://www.icao.int/publications/pages/publication.aspx?docnum=9303.

[28] P. I. Frazier, "A tutorial on bayesian optimization," *arXiv preprint arXiv:1807.02811*, 2018.

[29] A. Hore and D. Ziou, "Image quality metrics: Psnr vs. ssim," in *2010 20th international conference on pattern recognition*, pp. 2366–2369, IEEE, 2010.

[30] P. Korshunov and T. Ebrahimi, "Using face morphing to protect privacy," in *2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance*, pp. 208–213, IEEE, 2013.

[31] G. Wolberg, "Image morphing: a survey," *The visual computer*, vol. 14, no. 8-9, pp. 360–372, 1998.

[32] M. Ngan, P. Grother, K. Hanaoka, and J. Kuo, "face recognition vendor test (frvt): Morph-performance of automated face morph detection," *National Institute of Technology (NIST)*, 2022.

[33] H. Guan, Y. Wang, L. Xie, S. Nag, R. Goel, N. E. N. Swamy, Y. Yang, C. Xiao, J. Prisby, R. Maciejewski, *et al.*, "Idnet: A novel dataset for identity document analysis and fraud detection," *arXiv preprint arXiv:2408.01690*, 2024.

[34] J. Jordon, L. Szpruch, F. Houssiau, M. Bottarelli, G. Cherubin, C. Maple, S. N. Cohen, and A. Weller, "Synthetic data–what, why and how?," *arXiv preprint arXiv:2205.03257*, 2022.

[35] E. Sarkar, P. Korshunov, L. Colbois, and S. Marcel, "Are gan-based morphs threatening face recognition?," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2959–2963, IEEE, 2022.

[36] "Frll-morphs dataset." https://figshare.com/articles/dataset/Face_Research_Lab_London_Set/5047666.

[37] E. Sarkar, P. Korshunov, L. Colbois, and S. Marcel, "Vulnerability analysis of face morphing attacks from landmarks and generative adversarial networks," *arXiv preprint arXiv:2012.05344*, 2020.

[38] "Frgc-morphs dataset." https://www.nist.gov/programs-projects/face-recognition-grand-challenge-frgc.

[39] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE signal processing letters*, vol. 23, no. 10, pp. 1499–1503, 2016.

[40] N. Damer, N. Spiller, M. Fang, F. Boutros, F. Kirchbuchner, and A. Kuijper, "Pw-mad: Pixel-wise supervision for generalized face morphing attack detection," in *Advances in Visual Computing: 16th International Symposium, ISVC 2021, Virtual Event, October 4-6, 2021, Proceedings, Part I*, pp. 291–304, Springer, 2021.

[41] R. Ramachandra, S. Venkatesh, K. Raja, and C. Busch, "Detecting face morphing attacks with collaborative representation of steerable features," in *Proceedings of 3rd International Conference on Computer Vision and Image Processing: CVIP 2018, Volume 1*, pp. 255–265, Springer, 2019.

[42] F. Boutros, N. Damer, M. Fang, F. Kirchbuchner, and A. Kuijper, "Mixfacenets: Extremely efficient face recognition networks," in *2021 IEEE International Joint Conference on Biometrics (IJCB)*, pp. 1–8, IEEE, 2021.

[43] M. Fang, F. Boutros, and N. Damer, "Unsupervised face morphing attack detection via self-paced anomaly detection," in *2022 IEEE International Joint Conference on Biometrics (IJCB)*, pp. 1–11, IEEE, 2022.

[44] "Aws pricing." https://aws.amazon.com/ec2/pricing/on-demand/.

[45] S. Liu, Z. Zeng, T. Ren, F. Li, H. Zhang, J. Yang, C. Li, J. Yang, H. Su, J. Zhu, *et al.*, "Grounding dino: Marrying dino with grounded pre-training for open-set object detection," *arXiv preprint arXiv:2303.05499*, 2023.

[46] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, *et al.*, "Segment anything," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4015–4026, 2023.

[47] "Dod 5220.22-m – the secure wiping standard to get rid of data." https://www.bitraser.com/blog/dod-wiping-the-secure-wiping-standard-to-get-rid-of-data/.

[48] M. Tan and Q. V. Le, "Mixconv: Mixed depthwise convolutional kernels," *arXiv preprint arXiv:1907.09595*, 2019.