ONOT: a High-Quality ICAO-compliant Synthetic Mugshot Dataset

ONOT: a High-Quality ICAO-compliant Synthetic Mugshot Dataset

Nicolò Di Domenico, Guido Borghi, Annalisa Franco, Davide Maltoni
Department of Computer Science and Engineering. University of Bologna, Italy
{nicolo.didomenico, guido.borghi, annalisa.franco, davide.maltoni}@unibo.it

Fig. 1: Samples of the ONOT dataset compliant with the ISO/IEC 39794-5 standard and ICAO guidelines. The dataset exhibits a great inter-class variety, in terms of, among others, genders, ethnicity, age and face-specific traits.

***Abstract*— Nowadays, state-of-the-art AI-based generative models represent a viable solution to overcome privacy issues and biases in the collection of datasets containing personal information, such as faces. Following this intuition, in this paper we introduce ONOT[1], a synthetic dataset specifically focused on the generation of high-quality faces in adherence to the requirements of the ISO/IEC 39794-5 standards that, following the guidelines of the International Civil Aviation Organization (ICAO), defines the interchange formats of face images in electronic Machine-Readable Travel Documents (eMRTD). The strictly controlled and varied mugshot images included in ONOT are useful in research fields related to the analysis of face images in eMRTD, such as Morphing Attack Detection and Face Quality Assessment. The dataset is publicly released[2], in combination with the generation procedure details in order to improve the reproducibility and enable future extensions.**

I. INTRODUCTION

The rapid advancement of Artificial Intelligence (AI) has introduced a new era of unprecedented opportunities and challenges. Among the various applications of AI, face-based systems have gathered significant attention due to their potential to enhance effectiveness in fields ranging from security and surveillance (*e.g.*, Face Recognition [50], [34], Morphing Attack Detection [38], [5]) to human-computer interaction (*e.g.*, Facial Expression Recognition [30], [43], Facial Landmark Detection [48], [21]).

However, the adoption of these technologies has raised critical concerns, particularly related to privacy infringement and inherent biases [10]. For instance, algorithms for face image analysis have heavily relied on large-scale datasets [41], [42], [51] containing images of individuals' faces. While essential for training robust models, the acquisition and the release of these datasets have become increasingly problematic: the utilization of real facial images raises significant privacy concerns related, among others, to unauthorized face recognition, compromising privacy and personal security.

In this scenario, synthetic data generated through novel generative methods emerges as a promising solution to address these pressing issues [1]. Indeed, synthetic facial data offers a way to mitigate these privacy risks: by employing generative methods such as Generative Adversarial Networks (GANs) [22], Variational Autoencoders (VAEs) [29] and Diffusion Models [39], it is possible to generate highly realistic facial images that do not directly correspond to any real individual's identity, thus granting anonymity.

Furthermore, the creation of traditional face-based datasets has perpetuated biases that exist in society [26]. Biases related to ethnicity, gender, age, and other demographic factors have been - inadvertently or not - embedded in these datasets, leading to biased AI models [36]. Synthetic data presents an opportunity to counteract these biases: by carefully controlling the attributes of synthetic faces, it is possible to contrast the underrepresentation of specific groups, ultimately leading to fairer face recognition technologies.

[1] *One, No one and One hundred Thousand* (L. Pirandello, 1926)
[2] https://miatbiolab.csr.unibo.it/icao-synthetic-dataset

Therefore, in this paper, we introduce ONOT, a novel dataset of synthetic faces, meticulously crafted in adherence to the principles outlined in the ISO/IEC 19794-5 standard [23], successively modified by ISO/IEC 39794-5 [24], *i.e.* the reference standard in the context of face verification in electronic Machine-Readable Travel Documents (eMRTD). The standard describes the specific requirements for enrollment images imposing strict quality criteria to be fulfilled to enable effective automatic face verification: a summary of these principles is reported in Table I. This ISO standard has been designed starting from the guidelines initially provided by the International Civil Aviation Organization (ICAO) for passport photographs [46] (in the following, this standard is referred also as ISO/ICAO).

More specifically, we take this standard as the inspiring principle for our generation process, which is aimed at the creation of high-quality and well-controlled images with specific characteristics including, among others, frontal face pose with uniform background and illumination, neutral expression, and the absence of shadows (see Fig. 1). Then, we aim to create a synthetic dataset combining AI-based generative procedures, in terms of facial likeness and realism, together with the strict ISO/ICAO requirements. These unique features enable the use of the ONOT dataset for a variety of vision-based tasks related to the analysis of identity documents or, in general, in which there is the need for high-quality and standard frontal images, including the development of methods for Morphing Attack Detection [45] or Face Quality Assessment [20], for which ad-hoc public synthetic datasets are generally not available.

Summarizing, the ONOT dataset offers several key advantages and features:

- **ISO/ICAO compliance**: the dataset is a pioneering example of synthetic data specifically designed to meet ISO/ICAO standard requirements, and its compliance has been validated using a commercial SDK. To the best of our knowledge, this is the first synthetic dataset of its category in the literature.
- **Facial realism**: ONOT dataset presents high quality and realism in the generated faces, thanks to the use of a state-of-the-art generative method. The dataset comprises a collection of several subjects, including for each at least one ISO/ICAO compliant image and multiple additional samples. Each facial attribute is provided in dataset annotation.
- **Identity check**: rigorous verification procedures ensure both intra-subject consistency (all images of the same subject share the same identity) and inter-subject consistency (each subject presents a distinct and unique identity with respect to all the other subjects).
- **Reproducibility**: this dataset is highly reproducible and expandable, as it provides comprehensive documentation regarding the model, and the image generation and selection procedures. Indeed, we release the prompts used for each generation, fostering transparency and encouraging further research and data contributions.

No	Description of the test
1	Unique and valid face
2	Face fully included in image frame
Geometric tests	
3	Eye distance
4	Horizontal/vertical position
5	Head image width/height ratio
Photographic tests	
6	Face is correctly focused
7	Sharpness of the image
8	Face saturation
9	Image color conformance
10	Shadows over the face
11	Glasses with dark colored lenses or glare
12	Cluttered background
Pose and facial attributes tests	
13	Gaze direction
14	Mouth expression
15	Correct position of shoulders
16	Both eyes visible and open
17	Eyes color
18	Eyes occluded by glasses or hair
19	Presence of glasses
20	Glasses' frames too heavy
21	Presence of hat/cap on head

TABLE I: Tests carried out by the commercial ICAO SDK to decide whether an image is ISO/ICAO compliant. As reported, tests range from geometric and photographic to pose- and attribute-related aspects.

II. RELATED WORK

Due to the spread and efficacy of new AI-based generative algorithms, several datasets that contain synthetic faces are available in the literature [9]. The large majority of face-based synthetic datasets have been collected specifically for the Face Recognition task, and then they are created to have a high number of images, identities, head poses, neglecting specific standard requirements.

SynFace [37] addresses the challenges in collecting large-scale real-world training data for face recognition, especially considering label noise and privacy issues. The work identifies the performance gap between face recognition models trained with synthetic and real face images as poor intra-class variations and the domain gap between synthetic and real images. The synthesis is based on the DiscoFaceGAN [17] model and regards mostly frontal-view images, but the identity preservation between subjects is not evaluated.

In [3] DigiFace-1M, a large-scale synthetic dataset, is presented. The dataset is designed to address the scarcity, the biases and the label noise of diverse datasets for training face recognition models. DigiFace-1M, created through the framework presented in [47], provides a comprehensive set of facial images with varied attributes, including ethnicity, age, and facial expressions. The dataset is particularly notable for its scale (1 million images), but unfortunately, the level of realism seems to be limited.

The USynthFace dataset [8], generated through DiscoFaceGAN [17], includes synthetic face images with variabil-

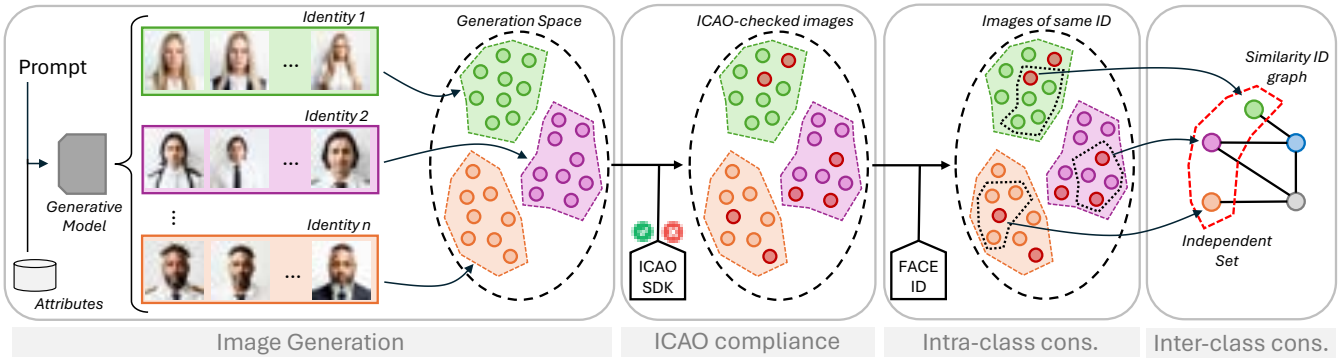


Fig. 2: Steps for the generation of the ONOT dataset. Starting from the initial image generation procedure, we apply a commercial SDK to verify if the generated images are compliant with the ISO/ICAO standard. The following steps regard the verification of the intra-class consistency, *i.e.* all images of the same subject share the same identity and the inter-class consistency, *i.e.* each subject presents a unique identity with respect to all the other generated subjects.

ity in identities, poses, illuminations, and expressions. In the paper, the authors particularly emphasize the use of synthetic data for training models in an unsupervised manner.

Differently, SFace [7] is created using a StyleGAN2-ADA [28] generative model under class-conditional settings, which generated 634k synthetic images, equally distributed across 10k classes. The main limitations are due to the limited variation in the same class and the demographic bias inherited from StyleGAN2. These limitations have been specifically addressed in [33] with GANDiffFace framework, based on a combined use of GAN and diffusion models.

Recently, a variety of synthetic datasets have been generated through the use of diffusion models [35], [18]: these datasets, in particular, are focused on identity preservation and diversification through inversion of pre-trained face recognition models (ID3PM [25]), style variation combined with subject consistency (DCFace [6]) and the use of authentic embeddings obtained from the authentic training datasets to enhance the realism of generated images (IDiff-Face [6]). In the context of Face Morphing, a synthetic dataset is proposed in [15]. The SMDD dataset contains 30k morphing attack and 50 bona fide samples. The morphing attack detection models [38] trained on SMDD demonstrated high performance even when tested against unknown attack types and morphing techniques, indicating its robustness and generalizability. Unfortunately, we found that these images do not pass the ISO standard checks (see Table I), resulting often in morphed or bona fide images with low-quality or visible artifacts.

In summary, the literature demonstrates significant progress in the development and utilization of synthetic facial image datasets. On the one hand, these datasets are increasingly being recognized as valuable tools for addressing the challenges of privacy, biases, and data availability in face recognition research. On the other hand, these datasets, often explicitly created only for the face recognition task in uncontrolled scenarios, tend to disregard the consistency of synthesized identities and the need for standard images that are used in document-related tasks.

III. DATASET GENERATION

A representation of the generation is provided in Figure 2.

A. *Image generation*

In this step, the goal is to generate facial images with a high level of realism and quality, compliant with the requirements of the ISO/ICAO standard.

To start the image generation process, 15k initial identities, here referred to as pseudo-classes (since, at this step, it is not guaranteed that different generated images correspond to different real identities), are defined through a random seed that, among others, contains information about the identity. Each pseudo-class is defined by the combination of a prompt and the initial seed. For each pseudo-class we generate 64 images, using a fixed negative prompt, a random positive prompt and increasing the initial seed by 1. The generation is based on a fine-tuned version of Stable Diffusion 1.5 [39], namely Realistic Vision 5.1. The model is served using Stable Diffusion Web UI [2]. Each image has a resolution of 512×512 and is generated using the DPM++ SDE Karras sampler [32], [27] with 25 steps. To generate the 15k identities (64 images per identity, for a total of 960k generated images), we employ $32\times$A100 64GB Nvidia GPUs for 14 hours in total.

Positive prompts are generated by randomizing values inserted into a predetermined template. Specifically, to emulate the characteristics of an official eMRTD picture, we engineer the prompt to obtain images able to pass the tests listed in Table I reflecting the ISO/ICAO requirements. The main aspects we explicitly control are related to the neutral expression upright frontal pose, bright background and uniform lighting. These desired attributes are assigned a higher weight than the rest of the prompt, given their importance in the context of this dataset, as detailed in the prompt template reported in Table II (in which the negative prompt and two samples of positive prompts are also reported). Properties such as gender and face traits are chosen following a weighted selection algorithm, and the probabilities are set as follows: 48% for male/female, 4%

Negative prompt	(deformed iris, deformed pupils, semi-realistic, CGI, 3D, render, sketch, cartoon, drawing, anime:1.4), text, close up, cropped, out of frame, worst quality, low quality, jpeg artifacts, ugly, duplicate, morbid, mutilated, extra fingers, mutated hands, poorly drawn hands, poorly drawn face, mutation, deformed, blurry, dehydrated, bad anatomy, bad proportions, extra limbs, cloned face, disfigured, gross proportions, malformed limbs, missing arms, missing legs, extra arms, extra legs, fused fingers, too many fingers, long neck, hair in front of the eyes, hat, (shadows), (three-quarter pose), (face in profile:1.1)
Prompt template	RAW front photo, face portrait photo of ({years} years old:1.1), {ethnicity} ({gender}:1.1), {hair color} hair, ({hair style} hair style:1.1), ({traits}:1.1), neutral expression, wearing dress, (white background:1.4), head horizontally aligned, (uniform lighting:1.4), top of the hair visible, (passport photo:1.1)
Prompt	RAW front photo, face portrait photo of (81 years old:1.1), African (female:1.1), black wavy hair, (braids hair style:1.1), (glasses and freckles:1.1), neutral expression, wearing dress, (white background:1.4), head horizontally aligned, (uniform lighting:1.4), top of the hair visible, (passport photo:1.1)

TABLE II: The negative prompt used for generating the images, the template of the positive prompt and one example of prompt of a subject, as detailed in Section III-A. The extensive negative prompt ensures that the images have a natural look, with realistic facial traits. The words within parentheses are assigned a greater weight by the model.

for non-binary; 23% for moles, freckles, moles and freckles; 2% for the other combinations of facial traits and attributes, included the presence of the glasses. A comprehensive list of these properties is given in Table III, which also includes the file naming convention used for the dataset. To further improve the variability of the dataset, we also include details in the prompt about the hair color (*e.g.* blonde, brown) and style (*e.g.* curly, straight, bold, with fringe), glasses type (*e.g.* round lenses, metal glasses) and gender-specific traits (*e.g.* beard), sampled through a uniform probability distribution.

B. ISO/IEC 39794-5 compliance

In this step, we aim to verify if each generated image fulfills the ISO/ICAO quality requirements. This ISO/IEC 19794-5 standard [23], recently modified by ISO/IEC 39794-5 [24], has been introduced to establish uniform guidelines and specifications for the exchange of biometric data, specifically facial images, between different systems and organizations. It was developed to address the need for interoperability and consistency in the field of biometrics, especially in applications based on identity verification and authentication [40]. Then, the standard promotes compatibility between various biometric systems and helps prevent data inconsistencies and errors when using automated facial recognition technology [12].

The compliance verification procedure is carried out through a commercial SDK[3]. Specifically, this SDK verifies the presence of scene constraints (such as pose, and expression), photographic properties (*e.g.* lighting, positioning, and camera focus), as well as digital image attributes (*e.g.* image resolution, and image size). A comprehensive list of tested features is reported in Table I. Upon the completion of the validation, pseudo-classes that do not contain at least one ISO/ICAO-compliant image are discarded.

C. Intra-class consistency

In this step, we aim to verify if the 64 images grouped in each pseudo-class belong or not to the same identity. Therefore, a face recognition pipeline is applied to each image. In particular, we detect faces employing the MTCNN [49] face detector and align them following the SphereFace [31] protocol. Any image that does not contain a face, or has more than one face, is discarded. Finally, for each remaining image, we extract its ArcFace [16] embedding.

We start by defining the similarity of faces i and j as the cosine distance of their respective embeddings $\mathbf{e}_i$, $\mathbf{e}_j$:

$$D_C\left(\mathbf{e}_i, \mathbf{e}_j\right) = 1 - \frac{\mathbf{e}_i \cdot \mathbf{e}_j}{\|\mathbf{e}_i\|_2 \|\mathbf{e}_j\|_2} \tag{1}$$

A cosine distance D_C closer to 0 (or more specifically, under a given threshold t) means that the two images' embeddings are similar. Therefore, if $D_C\left(\mathbf{e}_i, \mathbf{e}_j\right) \leq t$, we can conclude that i and j have the same facial identity.

[3] https://www.correlance.com/cms/en/home

Field	Description	Values
m	Model name	S - Stable Diffusion
x{8}	Seed	00000000 - 99999999
Gg	Gender	GM - male GF - female GN - non-binary
Aaa	Age	A18-A99 years
Eee	Ethnicity	EEA - European/American EAF - African EIA - Indian-Asian EAS - East-Asian EME - Middle Eastern
Ttt	Face traits	T00 - none T01 - moles T02 - scars T03 - freckles T10 - glasses T11 - glasses and moles T12 - glasses and scars T13 - glasses and freckles T14 - moles and freckles T15 - glasses, moles, and freckles T25 - freckles and scars
Innnn	Image number	0001 - 9999
Fff	Image format	F00 - digital F01 - Print&Scan (P&S)

TABLE III: The file naming scheme used to save the images to disk, which allows to understand the variety of the generated elements included in the ONOT datasets.

Fig. 3: In addition to ISO/ICAO-compliant samples, other images are generated for each identity. As shown, intra-class variance is present in terms of different head and body poses and facial traits.

The first step is to ensure that images within the same pseudo-class are consistent, *i.e.* all images are similar enough to all other images of the same class. Moreover, we require that each pseudo-class must contain at least one ISO/ICAO-compliant image.

More formally, given a set I of n images and a subset $C \subseteq I$ of ISO/ICAO-compliant images in the same given pseudo-class, we find the largest subset $V \subseteq I$ so that:

$$\forall \{i,j\} \in V \quad D_C(\mathbf{e}_i, \mathbf{e}_j) \leq t \wedge V \cap C \neq \emptyset \tag{2}$$

To find the above-mentioned subset, we start by constructing a similarity matrix $S \in \mathbb{R}^{n \times n}$, which is defined as

$$S_{i,j} = \begin{cases} 1 & \text{if } i \neq j \wedge D_C(\mathbf{e}_i, \mathbf{e}_j) \leq t \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

The symmetric binary matrix S can be interpreted as an adjacency matrix of an unweighted undirected similarity graph G, where each node represents an identity and each edge indicates that two faces are similar enough. Then, the set V that satisfies Equation 2 is found as the largest maximal clique in G that contains at least one ISO/ICAO-compliant image; all other images that are not part of such clique are discarded. To enumerate all maximal cliques we employ the Bron-Kerbosh algorithm [11] with pivoting [44], [14]. Despite having a worst-case time complexity of $\mathcal{O}\left(3^{V/3}\right)$, the running time of the algorithm remains practical given that the graph G contains at most 64 nodes.

The result of this procedure is shown in Figure 3, in which for each line we report the images in the same pseudo-class that we include in the dataset in addition to the ISO/ICAO compliant ones.

D. Inter-class consistency

As the prompts for the different pseudo-classes may generate subjects that are too similar to each other, the next step is to select the pseudo-classes that contain faces that are all dissimilar enough.

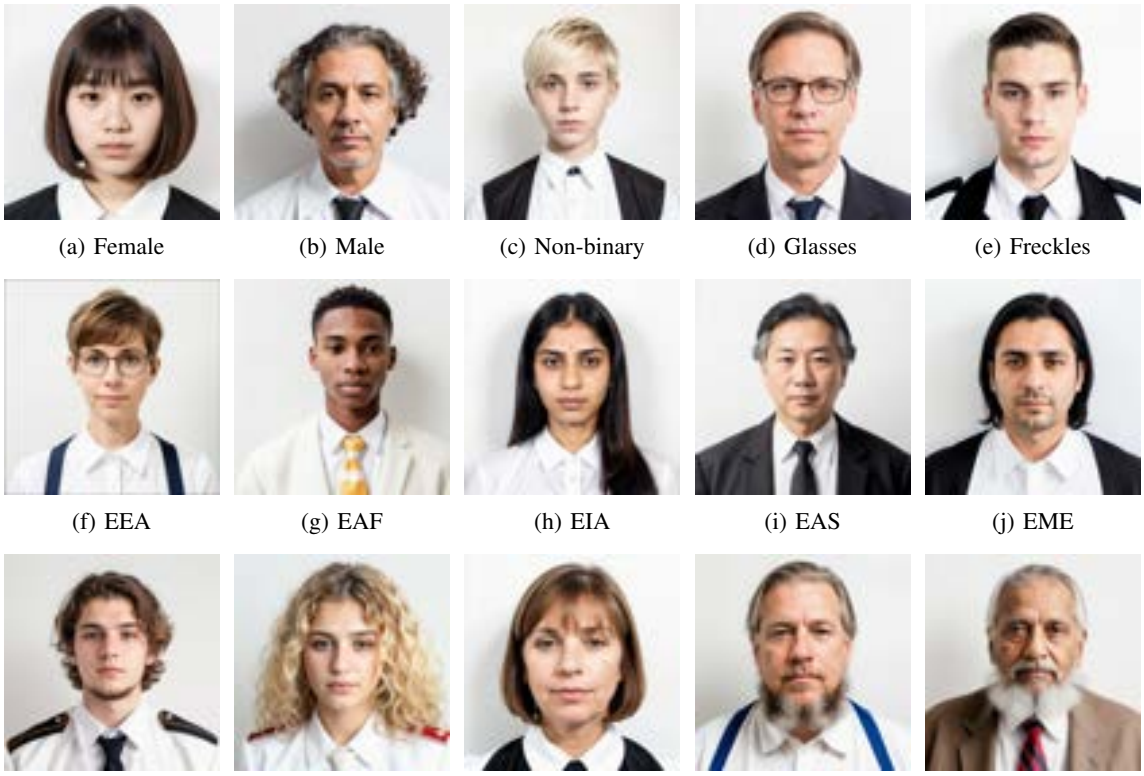
More formally, given the set of all n pseudo-classes P, we want to find a subset of classes $Q \subseteq P$ so that:

$$\forall \{i,j\} \in Q, i \neq j \quad D_C(\mathbf{e}_i, \mathbf{e}_j) > t \tag{4}$$

Note that after this step we can refer to the elements of Q as proper classes because each one contains exactly only one homogeneous identity, and different classes represent different identities. To find Q, we compute the similarity matrix $S \in \mathbb{R}^{n \times n}$ as defined by Equation 3; each cell represents the comparison of the ISO/ICAO-compliant images' embeddings across all pseudo-classes. As before, S can be interpreted as an adjacency matrix of an unweighted undirected similarity graph G, where each node represents a pseudo-class and each edge indicates that two pseudo-classes are similar enough. Then, we note that finding Q consists of computing the maximum independent set in G. All pseudo-classes not part of the found subset of nodes are discarded.

E. ICAO and identity consistency test statistics

As mentioned, the initial dataset generation includes 15k different pseudo-classes. After the first ISO/ICAO compliance test, 4032 identities survive: this reduction (-73%)



(a) Female (b) Male (c) Non-binary (d) Glasses (e) Freckles

(f) EEA (g) EAF (h) EIA (i) EAS (j) EME

(k) 18 y.o. (l) 30 y.o. (m) 50 y.o. (n) 70 y.o. (o) 80 y.o.

Fig. 4: Samples of the subject variability included in the ONOT dataset. Different genders, ethnicities, ages and facial traits are included in the dataset, enhancing the variability of the dataset. The naming convention is reported in Table III.

indicates a certain complexity in controlling specific face characteristics, as further analyzed in the next section.

The following inter- and intra-class consistency tests are based on a given threshold (t) for the face verification system [16]: in our case, we employ three distinct thresholds, experimentally determined by the execution of a set of 20k impostor face verification attempts on a separate real face dataset. The identified thresholds correspond to the FMR_{100}, FMR_{1000}, FMR_{10000}, and are respectively 0.597, 0.493, and 0.413. We obtained this way three image subsets; note that as the threshold values increase, the number of images within each class grows, while the count of distinct classes decreases. Specifically, after the identity consistency test 55, 125 and 255 distinct identities remain, for the three thresholds, respectively. These numbers reveal the challenges of generating faces that combine strict ICAO-compliant requirements and identity-based checks.

We observe these distinct subsets correspond to different working scenarios: for instance, the use of a low threshold implies a high level of similarity across different identities and a lower intra-class variability, representing a challenging benchmark for face analysis tasks since the resulting dataset will include several cases of look-alike subjects. Vice versa, a higher threshold implies the presence of more distinct identities, but a higher level of intra-class variability, making it suitable, for instance, to improve the robustness of FRSs to typical variations of face appearance.

(a) Digital (b) P&S

Fig. 5: Visual samples of the application of the P&S operation (see Sect. III-F) on two original images.

The ONOT dataset is released including the index annotation files needed to reproduce the three subsets.

F. Print&Scan Generation

Finally, for each image available in the ONOT dataset, we simulate the print and scan process (P&S) through the method described in [19]. We include the P&S operation since it is typical in procedures related to the issuance of

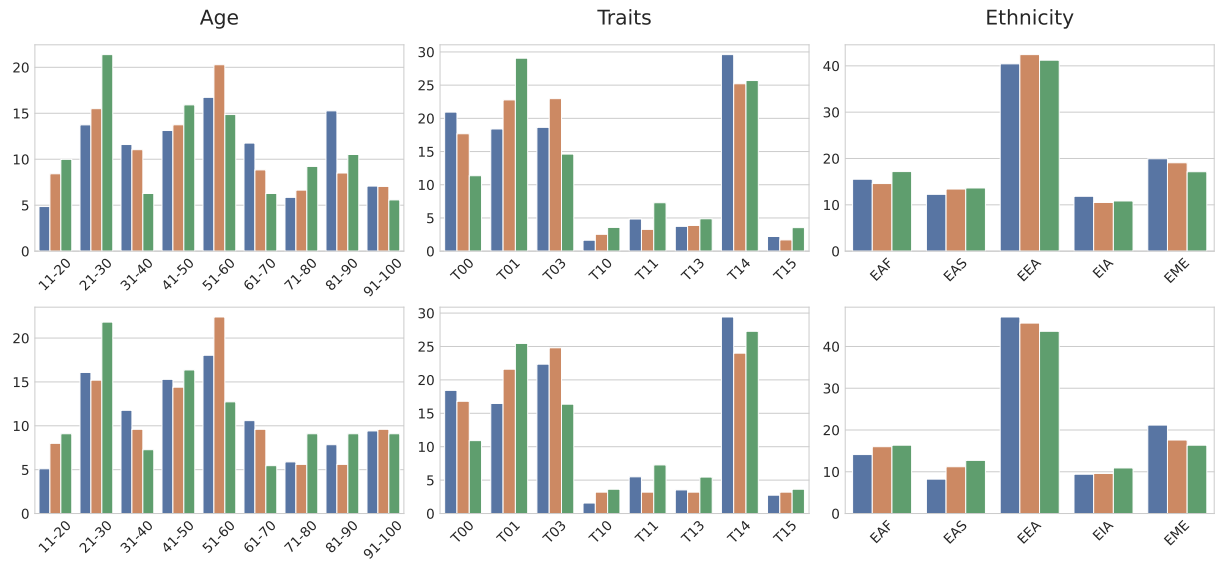


Fig. 6: Age, traits, and ethnicity distributions of ONOT images. The blue, orange, and green bars respectively denote images belonging to subsets defined by the three thresholds defined in Section III-E. The bottom three plots include only ISO/ICAO-compliant images, while the top three include all images. The naming convention is reported in Table III.

electronic identity documents [4]: such processes commonly entail the submission of a passport-sized photograph, which is later scanned and compressed for storage in the document's chip. P&S images are released together with the original version of the ONOT dataset. Figure 5 provides a visual representation of the results of these operations.

IV. DATASET STATISTICS

Some examples representing the variability of the dataset are depicted in Figure 4. In addition, main dataset statistics are illustrated in Figure 6, in which the first line shows plots computed on all the dataset images, while the second row the plots computed only on the ISO/ICAO compliant images.

We observe that the majority of the identities that pass the ICAO and identity consistency tests are in the range of $[21-60]$ years old, with the main peaks located in $[21-30]$ and $[51-60]$. The second plots reveal that the most common trait is represented by the combination of freckles and moles (T14), followed by no specific attributes (T00), freckles (T01) and moles (T03): these percentages follow the distribution detailed at the beginning of this section. Thus, this indicates that generating faces with specific attributes does not significantly influence compliance with the ICAO test. Noticeably, the third plots reveal the presence of a significant ethnic bias toward caucasian (EEA) subjects, which comprises more than 40% of the dataset regardless of the chosen subset. To investigate this behavior, we evaluate the proportion of images grouped by ethnicity relative to the total number of images before and after the checks detailed in Section III-B. In particular, the proportion of images that depict a caucasian subject initially accounts for 20.4% of the dataset (since 5 different ethnicities are considered). This proportion significantly increases to 45.6% after the ICAO and identity consistency tests. Conversely, the East-Asian (EAS) and Indian-Asian (EIA) ethnicities experience a substantial reduction in representation, decreasing from 20.4% and 19.8% to 6.0% and 7.1% respectively. Finally, Middle Eastern (EME) and African (EAF) ethnicities exhibit minimal variation in representation in the dataset.

These observations indicate a potential bias in the employed face verification, which is less able to discriminate identities in no-caucasian ethnicities, as suggested also in the literature [13]. With respect to the commercial ICAO SDK exploited, the slight difference in the distribution of the two rows denotes that the software is more robust, having a more uniform behavior across all ethnicities. Besides, these values can also indicate a complexity in the generation of images of a specific ethnicity, due to, for instance, an underrepresentation bias in data used for training the generative model.

Moreover, we plot the scores' distribution for the tests reported in Table I in Figure 7. These scores are output by the commercial ICAO SDK validation tool and each produced score is in the range $[0, 1]$. Following the official guideline of the SDK, an image is considered ISO/ICAO-compliant if all tests return a score greater or equal to 0.5 (red line in the plot). Results indicate that two tests, specifically "shadows over face" and "saturation", pose notable challenges for the images in the dataset. This highlights the difficulty faced by the image generator in controlling lighting conditions and saturation of the generated images, therefore significantly impacting the number of images that pass the ISO/ICAO compliance checks.

V. DISCUSSION AND FUTURE WORK

Despite the high quality of the generated images, the prompt-based generation is complex, especially when strict quality requirements have to be fulfilled. Indeed, there are specific facial characteristics particularly challenging to ac-

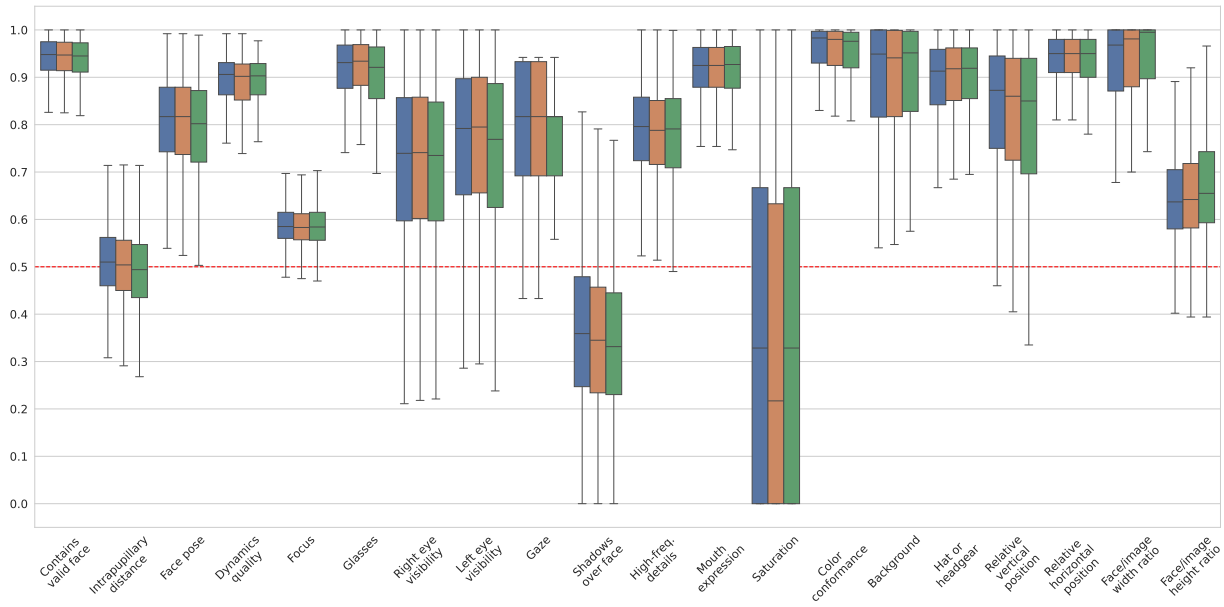


Fig. 7: Scores distribution of the tests performed during ISO/ICAO compliance check. An image is considered compliant if all checks return a score greater than 0.5 (dashed red line). The blue, orange, and green boxes respectively denote images belonging to subsets defined in Section III-D. Each box represents the quartiles of each score distribution, while the whiskers encompass the whole values' range.

Fig. 8: Failure cases of the generation procedure.

curately control during the generation such as, as previously mentioned, uniform lighting and skin color and the presence of shadows. We hypothesize this is due to the nature of the images used for training the generative model, mainly belonging to unconstrained real-world scenarios with images including, for instance, flashes and glares. Furthermore, we observe the presence of a limited number of images presenting some generation artifacts, depicted in Figure 8, that are difficult to detect through automatic face verification or face quality analysis systems. For instance, the first two images are correctly detected as faces by the used face detector [49], while in the remaining images only a single face is detected.

Another critical aspect is the variability in generated identities: indeed, starting from 15k pseudo-classes, only 255 identities survive the selection procedures (ICAO compliance and identity consistency tests). In consideration of this difficulty in controlling generated identities, we replicated a similar generation experiment by forcing in the positive prompt the generation of specific identities associated with well-known individuals (*e.g.* actors, politicians, and the like). In this case, starting from the same number of initial pseudo-classes, there was an observed increase of 37% in surviving identities, denoting the complexity of generating anonymous identities from seed with respect to the generation of images of known identities. Moreover, we observe a certain complexity also in generating multiple images of the same identity: employing the same prompt with slightly different seeds does not guarantee a constant identity across the generated images, thus requiring intra-class consistency tests.

Finally, another significant challenge is posed by the ISO/ICAO compliance verification tool: as many of these tools are commercial and closed-source, the precise reasons for a particular image failing a specific quality test are difficult to determine; therefore, engineering prompts that maximize the number of images that pass the ISO/ICAO compliance checks proves to be arduous.

As future work, we plan to increment the number of generated identities, and the release of a novel set of morphed images created starting from ONOT bona fide subjects, using multiple morphing algorithms. Another research topic should regard the possibility of constraining the generation through not only the input prompt, but also exploiting additional multi-modal information sources (*e.g.* models, images) that control specific elements (*e.g.* head pose, identity, age).

ACKNOWLEDGMENTS

This project received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 883356. Disclaimer: this text reflects only the author's views, and the Commission is not liable for any use that may be made of the information contained therein.

We thank Andrea Pilzer, NVIDIA AI Technology Center, EMEA, for his support. We also acknowledge the CINECA award under the ISCRA initiative, for the availability of high-performance computing resources and support.

REFERENCES

[1] N. C. Abay, Y. Zhou, M. Kantarcioglu, B. Thuraisingham, and L. Sweeney. Privacy preserving synthetic data release using deep learning. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2018, Dublin, Ireland, September 10–14, 2018, Proceedings, Part I 18*, pages 510–526. Springer, 2019.

[2] AUTOMATIC1111. Stable Diffusion Web UI, 2022.

[3] G. Bae, M. de La Gorce, T. Baltrušaitis, C. Hewitt, D. Chen, J. Valentin, R. Cipolla, and J. Shen. Digiface-1m: 1 million digital face images for face recognition. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3526–3535, 2023.

[4] G. Borghi, N. Di Domenico, A. Franco, M. Ferrara, and D. Maltoni. Revelio: A modular and effective framework for reproducible training and evaluation of morphing attack detectors. *IEEE Access*, 2023.

[5] G. Borghi, E. Pancisi, M. Ferrara, and D. Maltoni. A double siamese framework for differential morphing attack detection. *Sensors*, 21(10):3466, 2021.

[6] F. Boutros, J. H. Grebe, A. Kuijper, and N. Damer. Idiff-face: Synthetic-based face recognition through fizzy identity-conditioned diffusion model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19650–19661, 2023.

[7] F. Boutros, M. Huber, P. Siebke, T. Rieber, and N. Damer. Sface: Privacy-friendly and accurate face recognition using synthetic data. In *2022 IEEE International Joint Conference on Biometrics (IJCB)*, pages 1–11. IEEE, 2022.

[8] F. Boutros, M. Klemt, M. Fang, A. Kuijper, and N. Damer. Unsupervised face recognition using unlabeled synthetic data. In *2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG)*, pages 1–8. IEEE, 2023.

[9] F. Boutros, V. Struc, J. Fierrez, and N. Damer. Synthetic data for face recognition: Current state and future prospects. *Image and Vision Computing*, page 104688, 2023.

[10] K. W. Bowyer. Face recognition technology: security versus privacy. *IEEE Technology and society magazine*, 23(1):9–19, 2004.

[11] C. Bron and J. Kerbosch. Algorithm 457: finding all cliques of an undirected graph. *Communications of the ACM*, 16(9):575–577, 1973.

[12] C. Busch. Standards for biometric presentation attack detection. In *Handbook of Biometric Anti-Spoofing: Presentation Attack Detection and Vulnerability Assessment*, pages 571–583. Springer, 2023.

[13] J. G. Cavazos, P. J. Phillips, C. D. Castillo, and A. J. O'Toole. Accuracy comparison across face recognition algorithms: Where are we on measuring race bias? *IEEE transactions on biometrics, behavior, and identity science*, 3(1):101–111, 2020.

[14] F. Cazals and C. Karande. A note on the problem of reporting maximal cliques. *Theoretical computer science*, 407(1-3):564–568, 2008.

[15] N. Damer, C. A. F. López, M. Fang, N. Spiller, M. V. Pham, and F. Boutros. Privacy-friendly synthetic data for the development of face morphing attack detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1606–1617, 2022.

[16] J. Deng, J. Guo, N. Xue, and S. Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4690–4699, 2019.

[17] Y. Deng, J. Yang, D. Chen, F. Wen, and X. Tong. Disentangled and controllable face image generation via 3d imitative-contrastive learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5154–5163, 2020.

[18] P. Dhariwal and A. Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

[19] M. Ferrara, A. Franco, and D. Maltoni. Face morphing detection in the presence of printing/scanning and heterogeneous image sources. *IET Biometrics*, 10(3):290–303, 2021.

[20] A. Franco, A. Magnani, D. Maltoni, D. Maio, L. Odorisio, and A. De Maria. Face image quality assessment in electronic id documents. *IEEE Access*, 10:77744–77758, 2022.

[21] E. Frigieri, G. Borghi, R. Vezzani, and R. Cucchiara. Fast and accurate facial landmark localization in depth images for in-car applications. In *Image Analysis and Processing-ICIAP 2017: 19th International Conference, Catania, Italy, September 11-15, 2017, Proceedings, Part I 19*, pages 539–549. Springer, 2017.

[22] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

[23] ISO/IEC 19794-5 — Information technology — Biometric data interchange formats — Part 5: Face image data. Standard, International Organization for Standardization, 2011.

[24] ISO/IEC 39794-5 — Information technology — Extensible biometric data interchange formats — Part 5: Face image data. Standard, International Organization for Standardization, 2019.

[25] M. Kansy, A. Raël, G. Mignone, J. Naruniec, C. Schroers, M. Gross, and R. M. Weber. Controllable inversion of black-box face recognition models via diffusion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3167–3177, 2023.

[26] K. Karkkainen and J. Joo. Fairface: Face attribute dataset for balanced race, gender, and age for bias measurement and mitigation. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 1548–1558, 2021.

[27] T. Karras, M. Aittala, T. Aila, and S. Laine. Elucidating the design space of diffusion-based generative models. *Advances in Neural Information Processing Systems*, 35:26565–26577, 2022.

[28] 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*, 33:12104–12114, 2020.

[29] D. P. Kingma, M. Welling, et al. An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning*, 12(4):307–392, 2019.

[30] S. Li and W. Deng. Deep facial expression recognition: A survey. *IEEE transactions on affective computing*, 13(3):1195–1215, 2020.

[31] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song. Sphereface: Deep hypersphere embedding for face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 212–220, 2017.

[32] C. Lu, Y. Zhou, F. Bao, J. Chen, C. Li, and J. Zhu. Dpm-solver++: Fast solver for guided sampling of diffusion probabilistic models. *arXiv preprint arXiv:2211.01095*, 2022.

[33] P. Melzi, C. Rathgeb, R. Tolosana, R. Vera-Rodriguez, D. Lawatsch, F. Domin, and M. Schaubert. Gandiffface: Controllable generation of synthetic datasets for face recognition with realistic variations. *arXiv preprint arXiv:2305.19962*, 2023.

[34] P. Melzi, R. Tolosana, R. Vera-Rodriguez, M. Kim, C. Rathgeb, X. Liu, I. DeAndres-Tame, A. Morales, J. Fierrez, J. Ortega-Garcia, et al. Frcsyn-ongoing: Benchmarking and comprehensive evaluation of real and synthetic data to improve face recognition systems. *Information Fusion*, page 102322, 2024.

[35] A. Q. Nichol and P. Dhariwal. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*, pages 8162–8171. PMLR, 2021.

[36] E. Ntoutsi, P. Fafalios, U. Gadiraju, V. Iosifidis, W. Nejdl, M.-E. Vidal, S. Ruggieri, F. Turini, S. Papadopoulos, E. Krasanakis, et al. Bias in data-driven artificial intelligence systems—an introductory survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(3):e1356, 2020.

[37] H. Qiu, B. Yu, D. Gong, Z. Li, W. Liu, and D. Tao. Synface: Face recognition with synthetic data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10880–10890, 2021.

[38] K. Raja, M. Ferrara, A. Franco, L. Spreeuwers, I. Batskos, F. de Wit, M. Gomez-Barrero, U. Scherhag, D. Fischer, S. K. Venkatesh, et al. Morphing attack detection-database, evaluation platform, and benchmarking. *IEEE transactions on information forensics and security*, 16:4336–4351, 2020.

[39] 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 (CVPR)*, pages 10684–10695, June 2022.

[40] T. Schlett, C. Rathgeb, O. Henniger, J. Galbally, J. Fierrez, and C. Busch. Face image quality assessment: A literature survey. *ACM Computing Surveys (CSUR)*, 54(10s):1–49, 2022.

[41] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015.

[42] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Web-scale training for face identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2746–2754, 2015.

[43] Y. Tian, T. Kanade, and J. F. Cohn. Facial expression recognition. *Handbook of face recognition*, pages 487–519, 2011.

[44] E. Tomita, A. Tanaka, and H. Takahashi. The worst-case time complexity for generating all maximal cliques and computational experiments. *Theoretical computer science*, 363(1):28–42, 2006.

[45] S. Venkatesh, R. Ramachandra, K. Raja, and C. Busch. Face morphing attack generation and detection: A comprehensive survey. *IEEE*

transactions on technology and society, 2(3):128–145, 2021.

[46] A. Wolf. ICAO: Portrait quality (reference facial images for MRTD), version 1.0. standard. *International Civil Aviation Organization*, 2018.

[47] E. Wood, T. Baltrušaitis, C. Hewitt, S. Dziadzio, T. J. Cashman, and J. Shotton. Fake it till you make it: face analysis in the wild using synthetic data alone. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3681–3691, 2021.

[48] Y. Wu and Q. Ji. Facial landmark detection: A literature survey. *International Journal of Computer Vision*, 127:115–142, 2019.

[49] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE signal processing letters*, 23(10):1499–1503, 2016.

[50] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM computing surveys (CSUR)*, 35(4):399–458, 2003.

[51] Z. Zhu, G. Huang, J. Deng, Y. Ye, J. Huang, X. Chen, J. Zhu, T. Yang, J. Lu, D. Du, et al. Webface260m: A benchmark unveiling the power of million-scale deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10492–10502, 2021.