Synthetic Data in Human Analysis: A Survey

Indu Joshi, Marcel Grimmer, Christian Rathgeb, Christoph Busch, Francois Bremond, Antitza Dantcheva

▶ **To cite this version:**

Indu Joshi, Marcel Grimmer, Christian Rathgeb, Christoph Busch, Francois Bremond, et al.. Synthetic Data in Human Analysis: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, pp.1-20. 10.1109/TPAMI.2024.3362821 . hal-04519951

HAL Id: hal-04519951
https://hal.science/hal-04519951

Submitted on 25 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Synthetic Data in Human Analysis: A Survey

Indu Joshi, Marcel Grimmer, Christian Rathgeb, Christoph Busch, Francois Bremond, Antitza Dantcheva

Abstract—Deep neural networks have become prevalent in human analysis, boosting the performance of applications, such as biometric recognition, action recognition, as well as person re-identification. However, the performance of such networks scales with the available training data. In human analysis, the demand for large-scale datasets poses a severe challenge, as data collection is tedious, time-expensive, costly and must comply with data protection laws. Current research investigates the generation of *synthetic data* as an efficient and privacy-ensuring alternative to collecting real data in the field. This survey introduces the basic definitions and methodologies, essential when generating and employing synthetic data for human analysis. We summarise current state-of-the-art methods and the main benefits of using synthetic data. We also provide an overview of publicly available synthetic datasets and generation models. Finally, we discuss limitations, as well as open research problems in this field. This survey is intended for researchers and practitioners in the field of human analysis.

Index Terms—Human Analysis, Deep Neural Networks, Synthetic Data, Survey

1 Introduction

Deep neural networks (DNNs) have witnessed remarkable advancement in the past decade, leading to mature and robust algorithms in visual perception, natural language processing, and robotic control [1], among others. Such advancement has been fuelled by the development of *algorithms* to train DNNs, the availability of *large-scale* training *datasets*, as well as the progress in *computational power*.

DNN techniques have been designed for, among other applications, *human analysis*, aiming to recognize human characteristics, behaviour, as well as interactions with the physical world. In this context, human analysis ranges from the unique authentication of single individuals, the classification of human attributes or actions to the evaluation of crowd-based data. Despite the immense benefit of processing human data, *lack of annotated training data* still hinders DNNs from unfolding their full potential. In addition, the implementation of data protection laws, such as the *European general data protection regulation (GDPR)*, defines strict rules for processing data that can reveal identity information, thus violating the data subjects' informational self-determination. According to article 9 of the GDPR, biometric data is considered as *sensitive data*, and processing without explicit consent of the data subjects is imposed with fines of up to 20 million Euro or 4% of the firm's worldwide annual revenue from the preceding financial year (article 83).

One solution to overcome challenges related to limited training data and data protection has to do with creating large-scale *synthetic datasets*. Progress of deep generative models has allowed for the generation of highly realistic synthetic human images - challenging to distinguish from real data by both humans, and computer vision algorithms [2][3] (see Figure 1). While generative models have been able to produce highly realistic synthetic samples, we note that they are prone to leak information from training datasets. This is specifically of concern when human data is involved, and hence identity leaks risk at infringing personal privacy rights. In this context, current research indicates that identity leaks in deep generative networks become less likely, in case the complexity of the training dataset exceeds the complexity of the model architecture [4]. The main reason for identity leaks stems from generative model overfitting to the training dataset, with the consequence of specific units in the network revealing information of single data subjects - a concept referred to as *generative adversarial network (GAN) memorization*.

(a) StyleGAN2 [5]

(b) SFinGe [6]

(c) SpoofGAN [7]

(d) SURREAL [8]

(e) ElderSim [9]

Fig. 1: Synthetic images generated for human analysis, namely (a) 2D face image generation, (b) fingerprint image generation, (c) fingerprint presentation attack detection, (d) 2D pose estimation, (e) and elderly action recognition

- *I. Joshi, F. Bremond, and A. Dantcheva are with the STARS team of Inria, Sophia Antipolis - Méditerranée and Université Côte d'Azur, France. E-mail: indu.joshi@inria.fr, francois.bremond@inria.fr, antitza.dantcheva@inria.fr*
- *Marcel Grimmer is with the NBL - Norwegian Biometrics Laboratory, Norwegian University of Science and Technology, Norway. E-mail: marceg@ntnu.no*
- *C. Rathgeb and C. Busch are with the da/sec - Biometrics and Internet-Security Research Group, Hochschule Darstadt, Germany. E-mail: christian.rathgeb@h-da.de, christoph.busch@h-da.de*

1.1 Domains of application

Synthetic data boosts the performance of many data-driven models in human analysis [10] [11] [12]. In this context, a number of training schemes have been introduced including *data replacement* and *data enrichment*. The motivation for replacing real samples with synthetic data (*i.e.*, *synthetic training*) has to do with alleviating privacy concerns. In contrast, the combination of synthetic and real data (*i.e.*, *augmented training*) mainly aims at reducing biases achieved by re-balancing according to observed soft characteristics. Another optimization scheme aims at initializing model weights based on synthetic data with subsequent fine-tuning on a small subset of real data, referred to as *model initialization*. Finally, domain translation techniques are utilized to close the synthetic vs real domain gap (*domain adaptation*), thereby increasing the realism of synthetic datasets while preserving fine-grained annotations.

Deviating from synthetic data employed for model training, *synthetic evaluation datasets* have been utilized to benchmark the performance of existing algorithms, pre-trained models, and systems. This field of research is fuelled by the increasing representativeness of synthetically generated samples, which allows interference with systems and observed outcomes similar to those expected by real evaluation datasets. The preparation of large-scale testing databases intends to detect weaknesses in the human analysis pipeline without requiring expensive data collection initiatives. Apart from the cost factor, real data from specific (demographic) subgroups may not be accessible, so synthetic samples could balance underrepresented categories.

1.2 Structure of paper

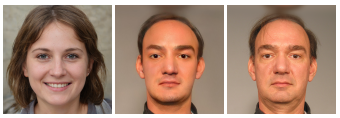
Given the increasing popularity of synthetic data, the main contribution of this survey is to revisit current research in human analysis, illustrating applications, benefits, and open challenges to accelerate future research. We introduce basic *terminology* and *scope* in Section 2, followed by Section 2.3, which provides an overview of the main *benefits* associated to synthetic data. Section 3 elaborates on *techniques for generating synthetic data*, followed by the most prominent *application scenarios* presented in Section 4. Section 3.5 summarises *synthetic datasets* and *data generation tools* that are publicly available across human analysis domains. Finally, in Section 5 we discuss *open challenges* identified in the literature analysis with promising new DNN concepts outlined in Section 6.

2 Synthetic Data in Human Analysis

The vast progress of deep generative networks has brought to the fore highly realistic synthetic data beneficial in automated human-centred analysis. To avoid ambiguity throughout this survey, we proceed to establish terminology of basic concepts, as utilized in this overview article.

2.1 Synthetic data

In general, *synthetic data* can be defined as *digital information generated by computer algorithms to approximate information collected or measured in the real world* [13]. Synthetic data stems generally from *traditional modelling* or *deep generative models*.



(a) F-S (b) S-S (c) Real

Fig. 2: Example images of a fully-synthetic (F-S), semi-synthetic (S-S), as well as real samples. The S-S face image (b) was generated with InterFaceGAN [20] by editing the age of the real face image depicted on the right side [21]. The F-S sample (a) was randomly generated with StyleGAN2 [5].

While traditional modelling generates real-world patterns based on prior expert knowledge through the *formulation of mathematical models*, deep generative models are designed to *automatically* learn patterns from the training dataset. In the last decade, deep generative models have outperformed traditional modelling techniques, *w.r.t.* quality and generalizability of the synthetic samples [14] [15]. In this survey, we refer to *generative models* in the context of both mathematical modelling and deep generative models.

Synthetic data samples can be *fully-synthetic*, as well as *semi-synthetic*. Fully-synthetic samples are generated without representing an underlying real-world object [16], generally by generative models, random sampling from a learned distribution [17][18]. At the same time, semi-synthetic samples constitute representations of real subjects, whose semantics have been manipulated [19]. For example, in human analysis, predicting the future appearance of a real face is considered semi-synthetic, as the image maintains the identity information, while altering the age. In contrast, fingerprint images synthesized by GANs based on random noise vectors are defined as fully-synthetic. An example image for each class is demonstrated in Figure 2.

In computer vision, real-world information is represented either at *sample* or *feature* level. In particular, we refer to data samples as the analogue or digital representation of human characteristics before feature extraction. According to the harmonic biometric vocabulary of ISO/IEC 2382-37:2017 [22], a feature vector is composed of *numbers or labels extracted from the data sample*. Specifically, feature vectors are treated as compressed sample representations, often encapsulating information, optimised for a specific downstream task, such as biometric recognition. In practice, generative models can either focus on generating "*synthetic samples*" [2] or "*synthetic features*"[23], depending on the target application.

2.2 Data replacement versus Data enrichment

While deep neural networks have achieved remarkable results in various computer vision tasks, it is still challenging to unleash their full potential due to the limited availability of large-scale datasets. The generation of synthetic samples can improve scalability and diversity, motivated by the following: Firstly, existing datasets being enriched with synthetic samples can increase dataset diversity. In this context, *data enrichment (DE)* imparts balancing of the proportions of soft characteristics in order to reduce dataset biases [24]. Note that in this survey, data enrichment signifies minor *data perturbations* such as image cropping, colour transformation,

as well as noise injection [25]. Due to a plethora of data augmentation techniques, distinction between *synthetic* and *augmented* samples is often challenging. Therefore, we refer to augmented samples as semi-synthetic, given that the original sample is at hand. In addition, we here denote weak supervision learning as a type of DE, as both synthetic and real samples are jointly employed for model training (see Section 4.5.3).

Secondly, *data replacement (DR)* refers to the replacement of real data with synthetic data [26]. This is instigated by *privacy* concerns in human analysis, where identity information can be linked with the corresponding sample.

Training human analysis models on domain-adapted synthetic datasets is considered a sub-category of DR, as only high-level information from a small subset of real data is being utilised (see Section 4.5.2). In contrast, the initialisation or fine-tuning of model weights with synthetic data is defined as a sub-category of DE due to the active involvement of real data that remains part of the training process (see Section 4.6).

Figure 4 introduces DNN-related training, evaluation, and attack mechanisms in which synthetic data has been employed including the following.

- **Augmented Training** refers to learning human analysis models or classifiers from a mixed training dataset that includes both real and synthetic data samples.
- **Weakly-Supervised Learning** signifies combined training with weak labels (real data) and accurate annotations (synthetic data).
- **Model Initialisation** denotes initial training on synthetic data with subsequent fine-tuning on real data towards reduction of the *synthetic versus real* domain gap.
- **Consistency Regularisation** denotes the utilisation of semi-synthetic data to enforce the consistency of model predictions for similar training samples.
- **Synthetic Training** signifies the training of models or classifiers on datasets composed of synthetic data only.
- **Unsupervised Domain Adaptation** denotes the employment of models trained on synthetic data to domain adaptation techniques (*e.g.,* Cycle-GAN), aiming to close the gap between *synthetic* versus *real* domain.
- **Synthetic Performance Evaluation** refers to assessing synthetic datasets generated to test the scalability and performance of systems, algorithms, or pre-trained models.
- **Digital Perturbation Attacks** describe either fully-synthetic or semi-semi-synthetic data generated to maliciously interfere with automated human analysis systems (*e.g.*, presentation attacks in biometric systems [27])) or deceive the human perception in recognizing individuals (*e.g.*, Deepfakes [28]).

Motivated by the above, synthetic data has enabled a number of applications, listed in Table 1 and elaborated on in Section 4. Further, Figure 4 summarises application scenarios derived from the forthcoming literature survey.

2.3 Benefits of synthetic data

Synthetic data can impart a performance boost to human analysis models, augment controllability and scalability, and mitigate privacy concerns. We here outline such benefits, whereas Section 4 revisits relevant works.

Performance boost. One ample application of synthetic data has been towards boosting the performance of human analysis models. Table 1 demonstrates such boost by comparing the associated performance before and after the use of synthetic data in several domains such as action recognition, crowd counting, face recognition, pose estimation, and gender classification. Moreover, Table 1 shows that synthetic evaluation datasets, including controlled labels, are exploited to evaluate the performance of new algorithms and pre-trained models. In human analysis, the high fidelity of evaluation datasets has been mainly fuelled by the remarkable progress in the domain of conditional image synthesis, which enables the generation of *synthetic mated samples* by manipulating single image semantics.

Controllability and scalability. The advances in generative models have enabled the generation of synthetic data, incorporating fine-grained control over semantics. Consequently, synthetic datasets can be created to balance important factors of variation (*e.g.*, the proportion of images pertained to male and female subjects), reducing biases caused by the unequal class distributions often observed in real-world datasets. Further, the employment of image synthesis models enables the generation of large-scale synthetic datasets, a factor known to correlate with the performance of DNNs.

Mitigating privacy concerns. Finally, fully-synthetic datasets reduce privacy concerns related to the distribution and processing of sensitive human data. Despite known incidents of information leaks of GANs [4], [29], [30], the reconstruction of training samples remains a challenge, as opposed to real data processing. We note that such *information leakage* is of concern and a set of related countermeasures have been identified, such as the concepts of *differential privacy* [31] and *precision reduction* [30]. While due to legal and privacy concerns, large-scale biometric datasets, such as MegaFace [32], have been withdrawn from public channels, we envision that large-scale synthetic datasets will be availed for DNN training and evaluation.

2.4 Human analysis

This survey defines *human analysis* as the analysis of human characteristics, behaviour, and interaction with the physical world. Such analysis has a myriad of applications, summarised in Figure 4. To elaborate, we note the following applications.

- **Biometric recognition** refers to the automated recognition of individuals based on their biological and behavioural characteristics [22].
- **Emotion Classification** refers to the process of classifying human emotion [33].
- **Soft-biometric classification** aims at automated classification of human characteristics in pre-defined categories, such as demographic, anthropometric or behavioural groups [34].
- **Presentation attack detection** (PAD) refers to the automated determination of a presentation to the

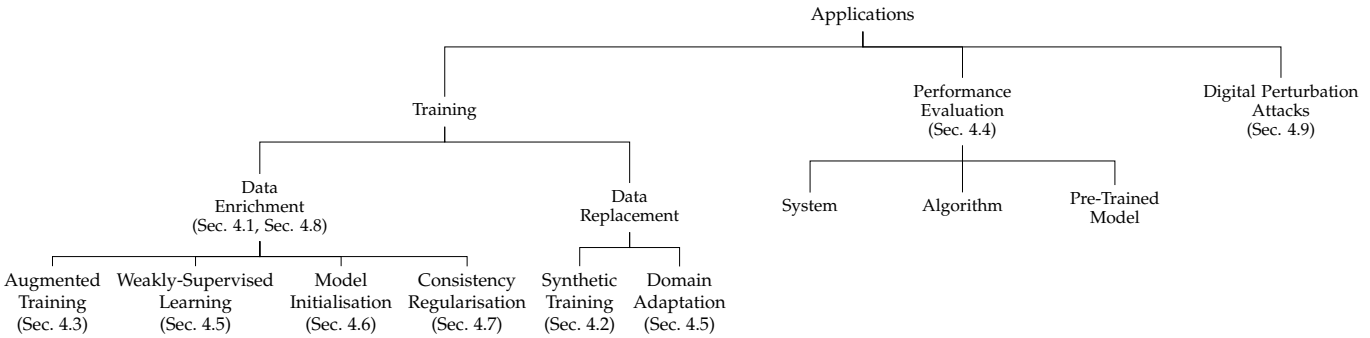


Fig. 3: Application types of synthetic data in human analysis.

biometric data capture subsystem to interfere with the operation of the biometric system [35].

- **Person Re-Identification** is the task of identifying an individual captured in images and videos acquired from different cameras or camera angles [36] [37]
- **Human interaction recognition** is the task of analysing human interactions of at least two individuals who are interrelated to each other (*e.g.*, handshaking) [38].
- **People detection/counting** denotes the detection or counting of individuals within a given image or video [39][40].
- **Semantic segmentation** signifies the pixel-based image classification with the goal of tracking human bodies [8] or body parts [11] in a given image or video.
- **(3D) Pose estimation** quantifies the transformation of the human body [41] or head [42] from a reference pose, given an image or a 3D scan [43]. In this context, **pose tracking** refers to the temporal pose estimation within video sequences [44].
- **Optical Flow Estimation** refers to tracking and visualizing the 2D motion of humans in videos by tracking human-specific features [45], [46].
- **Action recognition** focuses on recognizing activity of individual(s) from a series of observations from data subjects and their environment [47].
- **Anomaly detection** refers to classifiers trained to detect human behaviours, interactions, or movements deviating from normality [48].
- **Medical analysis** refers to the automated analysis of data collected in medical applications with the greater goal of restoring and maintaining human health. In this survey, synthetic data in medical applications is considered out-of-scope, and interested readers are referred to the work of Chen *et al.* [49].

3 How can synthetic datasets be generated?

Initial approaches for synthetic data generation generally exploit *mathematical modelling*, *3D rendering tools* or *perturbations using classical and hand-crafted* means. However, the success of deep neural networks in image generation has catapulted *dynamic perturbations* and *deep neural networks* as primary generation models. We proceed to provide details on such generation methods for synthetic data, leaving the reader with a selected choice of generation tools and openly available synthetic datasets.

3.1 Mathematical modelling

Mathematical modelling constitutes an early approach for generating human data aimed at approximating the distribution of real human data through mathematical modelling. Sampling from the approximated model can then be used to generate synthetic samples and exploit such in downstream human analysis tasks. Approximation of the mathematical model pertaining to the human data requires domain expertise and a careful understanding of model parameters. A popular mathematical modelling-based synthetic fingerprint generation (SFinGe) method is proposed by Cappelli *et al.* [6]. The authors exploited domain expertise to define a fingerprint orientation model characterized by the number and location of the fingerprint cores and deltas. The synthetic fingerprint generation starts from initializing the locations of core and deltas, followed by ridge orientation and density generation. Subsequently, the authors applied space-invariant linear filtering to obtain a binarized good quality fingerprint image. Lastly, domain-specific noise was introduced to simulate realistic greyscale fingerprint images. Approaches exploiting mathematical modelling using domain knowledge for synthetic data generation include finger vein recognition [73], hand shape recognition [74], face recognition [75], and iris recognition [23].

3.2 3D rendering tools

Several studies exploit 3D modelling to create mathematical representations of the three-dimensional surface of the object of interest. Subsequently, a 3D rendering tool is exploited to render images corresponding to a 3D model. Han *et al.* [52] argued that the generation of synthetic samples in 3D space allows for the incorporation of extreme changes in illumination, viewpoint, occlusion, scale, and background. Additionally, rendering engines allow precise control over environmental conditions such as pose variations, lighting, and object geometry, leading to accurate annotations, which are often acquired for a real dataset. Most popular 3D rendering tools include Blender[1], Maya[2], 3ds Max[3], Cinema

1. https://www.blender.org/
2. https://www.autodesk.fr/products/maya/overview
3. https://www.autodesk.com/products/3ds-max/overview

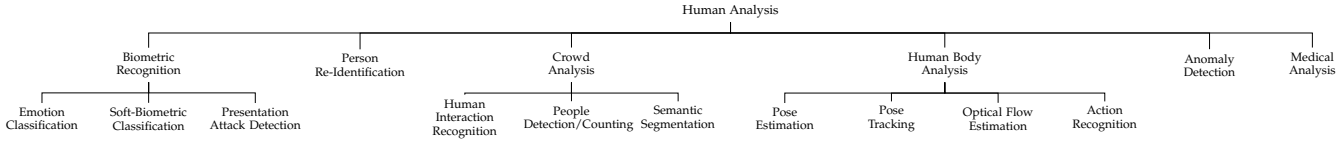


Fig. 4: Application domains in human analysis

TABLE 1: Performance of human analysis models trained or evaluated with and w/o synthetic data. Numbers given in %, except for MPJPE, which is reported in in *mm* (DE=data enrichment, DR=data replacement, EER=equal error rate, MAE=mean absolute error, MSE=mean square error, FNMR=false non-match rate, MPJPE=mean per joint position error, U=Illumination, E=Expression, P=Pose). For precise definitions of the metrics used, we refer to the referenced studies.

Reference	Application Domain	Application Type	Metric	w/o synthetic data	DE	DR
Aranjuelo *et al.* [39]	People detection	Augmented Training	Average Precision (↑)	70	82	-
Wang *et al.* [40]	People counting	Synthetic Training	MAE (↓)	275.5	-	225.9
Yadav *et al.* [15]	Iris PAD	Augmented Training	EER (↓)	25.18	18.52	-
Grosz and Jain [7]	Fingerprint PAD	Augmented and Synthetic Training	Accuracy (↑)	99.52	100	36.53
Bird *et al.* [50]	Speaker recognition	Model Initialization	Average Accuracy (↑)	95.48	99.35	-
Tapia *et al.* [51]	Gender classification from periocular images	Evaluation	Accuracy (↑)	82.76	-	91.9
Han *et al.* [52]	Face Detection	Evaluation	Average Precision (↑)	64	74.5	
Basak *et al.* [53]	Head pose estimation	Domain Adaptation	MAE (↓)	6.34	-	5.13
Bird *et al.* [54]	Speaker recognition	Model Initialization	Accuracy (↑)	96.58	98.83	
Dou *et al.* [10]	Gait recognition	Augmented Training	Rank-1 Accuracy (↑)	95.0	96.4	-
Piplani *et al.* [55]	Passthrough authentication	Augmented Training	Accuracy (↑)	90.8	95	-
Gouiaa *et al.* [56]	Posture recognition	Augmented Training	Accuracy (↑)	94.58	99	-
Ruiz *et al.* [57]	Signature recognition	Augmented Training	EER (↓)	11.11	4.9	-
Kim *et al.* [58]	Face recognition	Synthetic Training	Average Accuracy (↑)	94.62	-	91.21
Chen *et al.* [59]	Emotion Classification	Augmented Training	Accuracy (↑)	58.6	64.5	-
Melo*et al.* [60]	Signature Recognition	Synthetic Training	EER (↓)	10.26	-	9.74
Öz *et al.* [61]	Eye Segmentation	Augmented Training	mIoU (↑)	73	75.4	-
Wang *et al.* [62]	People Counting	Model Initialization	MSE (↓)	14.3	13	-
Irtem *et al.* [12]	Fingerprint Classification	Augmented and Synthetic Training	Classification accuracy (↑)	91.9	95.53	69.47
Engelsma *et al.* [17]	Fingerprint Recognition	Model Initialization	True acceptance rate (↑)	73.37	87.03	-
Bozorgtabar *et al.* [33]	Expression Recognition	Domain Adaptation	Accuracy (↑)	70.15	-	72.1
Qiu *et al.* [26]	Face Recognition	Augmented and Synthetic Training	Accuracy (↑)	91.22	95.78	91.97
Kortylewski *et al.* [24]	Face Recognition	Model Initialization	Accuracy (↑)	91.2	93.3	88.9
Colbois *et al.* [63]	Face Recognition	Evaluation	False non-match rates U/E/P (↓)	11/3/55	-	12/25/51
Marriott *et al.* [64]	Pose-invariant Face Recognition	Augmented Training	Accuracy (↑)	93.59	95.29	-
Wood *et al.* [11]	Face Segmentation	Synthetic Training	F_1 score (↑)	91.6	-	92
Ahmed *et al.* [65]	Facial Expression Classification	Augmented Training	Accuracy (↑)	92.95	96.24	-
Niinuma *et al.* [66]	Facial Expression Classification	Synthetic Training	Inter-rater reliability (↑)	48.9	-	52.5
Ranjan *et al.* [45]	Optical Flow Estimation	Evaluation	Motion compensated intensity (↓)	158.3	–	71.5
Zhu *et al.* [67]	Pose tracking	Synthetic Training	MPJPE (↓)	-	-	76.41
Cai *et al.* [68]	Pose Estimation	Augmented Training	Procrustes-aligned MPJPE (↓)	65.7	57.9	61.7
Varol *et al.* [69]	Action Recognition	Augmented Training	Accuracy 0°/45°/90° (↑)	88.8/78.2/57.3	90.5/83.3/68	-
Hatay *et al.* [70]	(Phone) Action Recognition	Model Initialization	Accuracy (↑)	95.83	96.67	-
Souza *et al.* [71]	Action Recognition	Augmented Training	Accuracy (↑)	93.3	92.7	-
Varol *et al.* [8]	Human Body Segmentation	Model Initialization	Accuracy (↑)	58.54	67.72	56.51
Priesnitz *et al.* [72]	Contactless Fingerprint Recognition	Evaluation	Average EER (↓)	30.93	-	3.55

Fig. 5: Aranjuelo *et al.* [39] utilized 3ds Max software to virtually render humans (right) on a real scene (left) for detection of moving subjects.

4D[4], Unity[5], and Unreal Engine[6]. Despite the precise control over subject and environmental-related attributes, it is still an open challenge to close the domain gap between synthetic and real data in terms of visual quality and approximation of intricate details [76], [77].

Aranjuelo *et al.* [39] virtually rendered humans on real scenes for application in the detection of individuals (see Figure 5). Similarly, Öz *et al.* [61] used a 3D rendering tool to generate synthetic eye images and exploit the generated samples to learn eye region segmentation (see Figure 6).

4. https://www.maxon.net/en/cinema-4d
5. https://www.unity3D.com
6. https://www.unrealengine.com

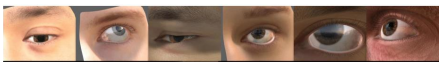


Fig. 6: Öz *et al.* [61] generated synthetic eye images employing UnityEyes [81], a 3D rendering tool. The synthetic data is used for learning eye region segmentation.

Recently, 3D rendering tools employed in video games have become a valuable source for collecting synthetic data, aiming to improve performances across different human analysis tasks. Among others, Zhu et al. [67] and Cai et al. [68] extracted training data from *NBA2K2019* and *GTA-V*, in order to achieve state-of-the-art performances in 3D human body reconstruction. Other studies exploiting 3D rendering tools for generating synthetic data spanned applications in re-identification of individuals [78], face recognition [79], [52], and gait recognition [80].

3.3 Input perturbations

Perturbations of a given input are widely used to generate synthetic data. Such perturbations are either introduced by noise using classical and hand-crafted methods or by a learning-based approach. We proceed to provide a brief discussion on both approach types.

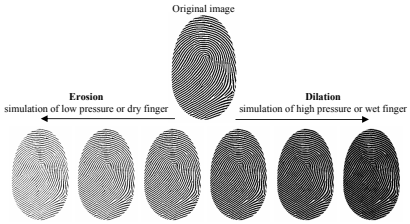


Fig. 7: Cappelli *et al.* [6] exploited morphological operations such as erosion and dilation to vary ridge thickness while generating multiple impressions of a fingerprint image.

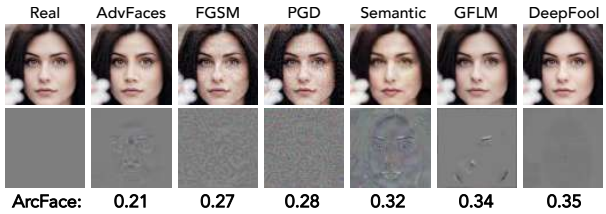


Fig. 8: Jain *et al.* [87] dynamically perturb a face image using six different adversarial training mechanisms (top row). The corresponding perturbations are provided in the bottom row. The authors demonstrate that synthetic faces generated using dynamic perturbations can increase face comparison score (obtained using ArcFace) in non-mated comparison trials.

3.3.1 Perturbations using classical and hand-crafted methods

Classical and hand-crafted methods can perturb a given input to either introduce variations in the available data or simulate cases that are difficult to capture otherwise. Most prominent classical and hand-crafted methods include Gaussian blurring, image blending, colour jittering, horizontal and vertical flipping, rotation, translation, as well as affine transformations. Some studies utilize morphological operations such as erosion and dilation to generate synthetic data samples. Following this direction of synthetic data generation, Ibsen *et al.* [82] exploited image processing techniques to synthetically blend tattoos on human faces. Similarly, Cappelli *et al.* [6] generated synthetic multiple impressions from a given input fingerprint using morphological operations (see Figure 7). Other studies that generate synthetic data using classical methods have been instrumental in fingerprint recognition [83] [84], iris recognition [85] , and re-identification of individuals [86].

3.3.2 Dynamic perturbations

A dynamic perturbation is defined as an input-specific perturbation introduced through an adversarial training mechanism such that a learning-based human analysis model is likely to make an erroneous prediction [87]. Training a human analysis model with the synthetic data generated using dynamic perturbations is beneficial for regularization and improvement of robustness. Following this approach, several studies generated synthetic data using adversarial training. Jain *et al.* [87] generated synthetic non-mated facial images using dynamic perturbations that obtain high comparison scores (see Figure 8). Other studies in human analysis exploiting dynamic perturbations include applications in re-identification of individuals [88], face recognition [89], iris recognition [90], and fingerprint recognition [91].

3.4 Deep neural networks

Deep neural networks (DNNs) represent state-of-the-art architectures for generating synthetic data for among others, applications in human analysis. By revisiting related literature, we identify following categories for doing so.

3.4.1 Sequence-based neural networks

Originally, a recurrent neural network (RNN) is a DNN designed to process time-series, as well as sequential or variable-length input data. Such models are designed for applications, where input data samples depend on the previous data samples, as RNNs are aimed at capturing dependencies between data samples. Towards capturing long-range dependencies, state-of-the-art RNNs exploit long short-term memory (LSTM) and gated recurrent units (GRU) [28] to store information from previous inputs or states and generate the subsequent output of the input sequence. An LSTM comprises three gates: input, output and forget gate, while a GRU incorporates a reset and an update gate. These gates determine the most informative part of the input to make a prediction in the future. Additionally, in the realm of computer vision, vision transformers (ViTs) [92] have been proposed as an effective architecture for image data processing, leveraging self-attention mechanisms to capture global patterns in high-resolution images.

One of the applications exploiting RNN to generate synthetic data is the contribution of Bird *et al.* [54], where a character-level RNN is exploited to generate audio sentences for speaker identification. In addition, RNNs are employed for generating deep fakes, where these architecture render continuous realistic flow in audio or video [28].

3.4.2 Auto-Encoders

Auto-Encoder (AE) based generative models constitute a pair of encoder and decoder networks. While the encoder network learns an efficient representation of the input, the decoder network generates an output corresponding to the given latent vector provided as output by the encoder network. These models generate synthetic data by learning the joint distribution of the latent space and the training data. Such models are generally regularized by imposing a prior distribution on the latent space to facilitate generation during inference [93]. Prominent auto-encoder architectures for synthetic data generation include variational auto-encoder [93], adversarial autoencoder [94] and Wasserstein auto-encoder [95], which includes a Gaussian prior. However, the Gaussian prior is simplistic and might fail to capture complex latent distributions. To alleviate this limitation, rich classes of distributional priors have been explored [96], [97]. Several research efforts have attempted to learn disentangled representations in the latent space of the VAE [98], [99]. Such a factored representation is beneficial in interpolating the latent space, leading to the generation of diverse samples and plausible modification in input data. Despite offering interpretable inference, stable training, and an efficient sampling procedure, the generation quality of

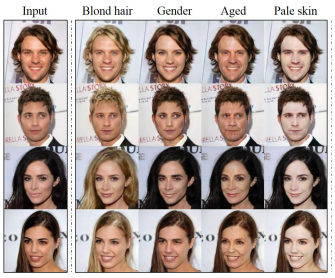


Fig. 9: Choi *et al.* [14] proposed StarGAN, a generative adversarial network that is able to alter attributes of a given face image. The generated synthetic faces have been commonly used as deepfakes.

VAEs is not as impressive as the one achieved by GANs [97]. Next, we discuss the most widely used state-of-the-art deep generative framework, namely GANs [100].

3.4.3 Generative adversarial networks (GANs)

Goodfellow *et al.* in their seminal work [100] proposed a framework incorporating two networks, a generator and a discriminator. The generator learns a distribution of training samples, whereas the discriminator network is aimed at classifying whether the input samples stem from the training set or are generated by the generator (real or fake). Both networks are trained in an adversarial manner (zero-sum game), and the framework targets to facilitate improved approximation of true distribution by the generative model [101]. Hence, the name *generative adversarial network*. GANs are broadly categorized as *noise to image translation* or *image to image translation* networks. The former aim at upscaling a randomly sampled noise vector to a realistic image, whereas the latter are trained to transform a given image to another image. Although achieving photorealistic and high-resolution image quality, GANs suffer from training instability and mode collapse, constraining the diversity by generating synthetic samples close to the average of a training dataset.

Prominent noise to image translation GANs include DCGAN [102] and Wasserstein GAN [103], whereas frequently empolyed image to image translation GANs include pix2pix [104] and Cycle-GAN [105]. Several studies in human analysis exploited GANs to generate synthetic data [106], [55], [107], [108], [33]. One such study includes the contribution of Cao and Jain [109]. The authors generated synthetic fingerprints using noise to image translation GAN. Similarly, Choi *et al.* [14] proposed an image to image translation GAN to modify attributes in facial images (see Figure 9).

3.4.4 Scene Graphs

Deng *et al.* [110] argued the necessity of understanding the relationship between different objects in a scene to generate synthetic data with multiple objects in a scene. The authors proposed to represent a multi-object scene as a tree-structured probabilistic scene graph that is trained with variational inference. Scene graphs are additionally utilized to generate moving objects [111]. For details on scene graph-based generative models, the readers are referred to [112].

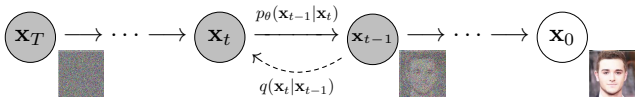


Fig. 10: Ho *et al.* [113] proposed to generate synthetic data using a DPM that uses a directed graphical model at its core.

3.4.5 Diffusion Probabilistic Models

The most recent generative models for synthetic data generation include diffusion probabilistic models (DPMs) [113], aiming to overcome the diversity constraints of GANs. A DPM is parameterized through a Markov chain that exploits variational inference to train the model towards generating realistic data samples after a finite time. Transitions among these chains are learnt via iteratively introducing noise into the training set until the signal is destroyed. The overall goal of noise injection has the goal to allow the model to learn to reverse the diffusion process, and eventually to learn to generate realistic data samples from a given noise vector (see Figure 10). In addition, DPMs are exploited for conditional data generation [114], as well as image-to-image translation [115].

3.5 Open-Source Availability

Finally, we provide an overview of synthetic datasets and synthetic data generation tools available for public usage. We emphasize the importance of sharing datasets and tools within the research community for improved reproducibility of results. As such, Table 2 presents publicly-available datasets comprised of synthetic data only. Further, Table 3 introduces synthetic data generation tools to enable new researchers in the field of human analysis to build custom-generated datasets tailored to their needs.

In summary, we find that DNNs and 3D rendering tools are frequently used techniques for generating synthetic data for human analysis. However, the major challenge in synthetic data generation remains related to ensuring diversity, representation and preventing identity leakage (discussed in Section 5).

4 How can synthetic data be utilized?

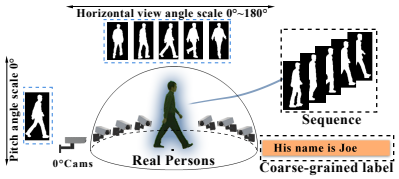
Synthetic data is frequently used to *simulate complex scenarios* for which the data collection is particularly challenging, to *overcome privacy issues* observed for collection of real human analysis datasets, *increase the size and diversity of training datasets*, as well as to *mitigate bias* in real training datasets. Furthermore, looking at the challenge in collecting large-scale datasets, synthetic data is widely used in *scalability analysis* of systems. Additionally, as obtaining annotations can be both time-consuming and expensive, *synthetic data, whose annotations can be automatically derived* is prominently used. With *consistency regularization* techniques, synthetic data is used to learn generalizable models. Synthetic data can also be employed to produce *presentation attacks* on human authentication systems. We proceed to provide details on different uses of synthetic data.

TABLE 2: Publicly available synthetic datasets.

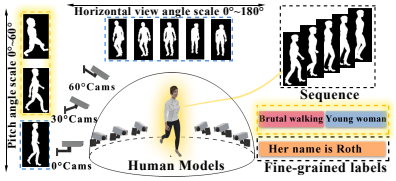
Reference	Name	Application Domain	Year	Data Type	Dataset Size
Wood *et al.* [11]	Microsoft Face Synthetics	Landmark localization, Face parsing	2021	Images	100,000
Falkenberg *et al.* [116]	HDA-SynChildFaces	Child-based Face Recognition	2023	Images	188,832 Frames
Varol *et al.* [8]	SURREAL	Human Pose Estimation	2017	Video Frames	6,000,000
Fabbri *et al.* [44]	Joint Track Auto (JTA)	Human Pose Tracking	2018	Videos	512
Barbosa *et al.* [117]	SOMASet	Person Re-identification	2017	Images	100,000
Varol *et al.* [69]	SURREACT	Action Recognition	2021	Videos	106,000
Da *et al.* [118]	Mixamo Kinetics	Action Recognition	2020	Videos	36,195
Ariz *et al.* [119]	UPNA Synthetic Head Pose Database	Head Pose Estimation	2016	Videos	120
Roitberg*et al.* [120]	Sims4Action	Action Recognition	2021	Videos	625.6 minutes
Hwang*et al.* [9]	KIST SynADL	Elderly Action Recognition	2020	Videos	462,000
Ranjan*et al.* [45]	MHOF	Multi-Human Optical Flow	2020	Video Frames	111,312 Frames

TABLE 3: Publicly available synthetic data generation models (MM=Mathematical Modelling).

Reference	Generation tool	Year	Method
Drozdowski *et al.* [23]	Synthetic Iris Code Generator	2017	MM
Li *et al.* [76]	3D Face Model Generation (FLAME)	2019	3D MM, DNN
Feng *et al.* [121]	3D Face Model Encoder (FLAME)	2021	DNN
Chan *et al.* [122]	3D-Aware Face Image Generation	2022	DNN
Karras *et al.* [2]	Face Image Generation (StyleGAN3)	2021	DNN
Maltoni *et al.* [6]	Fingerprint Image Generator (SFinGe)	2009	MM
Priesnitz *et al.* [72]	Contactless Fingerprint Image Generator	2022	MM
Sun *et al.* [37]	Person Re-Identification (PersonX)	2019	3D MM
Hwang *et al.* [9]	Elderly Action Recognition	2020	3D MM
Zhu *et al.* [67]	3D Pose Estimation	2020	DNN



(a) Real dataset collection.



(b) Synthetic data generation.

Fig. 11: Dou *et al.* [10] discussed the limitations of existing real databases for video-based gait recognition in capturing complex scenarios. For instance, the authors discussed that (a) real datasets are only acquired with a single camera pitch angle and (b) on the other hand, the synthetic dataset is generated with a diverse range of camera pitch angles. Thus, synthetic data can be used to simulate complex scenarios, which are otherwise difficult to acquire for a real dataset for human analysis.

4.1 Simulating complex scenarios

Although suffering from the real-vs-synthetic domain gap, synthetic datasets for augmented training or model initialization of DNNs are employed to improve associated robustness towards complex scenarios across various applications, where collection of real data is particularly difficult. Dou *et al.* [10] argued that existing real databases for video-based gait recognition do not possess examples of complex scenarios that can be crucial for obtaining satisfactory performance in real-world applications. For instance, real datasets are captured under ideal settings with only a single camera pitch angle (see Figure 11). Specifically in the OU-MVLP dataset [123] for gait recognition subjects only walk twice without the change of bag or clothing, with only one subject appearing per video frame. However, real-world scenarios naturally include multiple walking individuals. Towards bridging this gap, the authors generated approximately one million synthetic silhouette sequences of 11,000 subjects. The resulting synthetic dataset VersatileGait comprises of gait sequences with a diverse range of camera pitch angles and fine-grained annotations of attributes. Furthermore, to promote the design of multi-person gait recognition algorithms, the authors also generated multi-person walking scenarios with up to three people walking simultaneously.

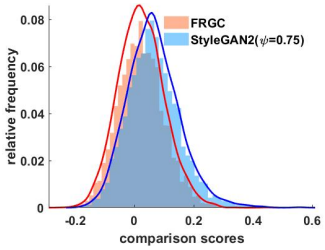


Fig. 12: Zhang *et al.* [127] compared the distribution of non-mated comparison scores of real (FRGC-V2 face database [21]) versus synthetic faces generated. Only minor differences in non-mated comparison scores corresponding to synthetic data were observed, as seen in real data. These results illustrated the potential of synthetic data being utilized instead of real data, alleviating privacy issues.

Similarly, Aranjuelo *et al.* [39] argued that existing real datasets for human detection do not exploit omnidirectional cameras to capture a 360° view in surveillance videos. To take advantage of the 360° view, the authors proposed the subject detection model to be trained with synthetic data. Other applications, exploiting synthetic data to simulate complex scenarios include the contributions of Lai *et al.* [124] for generating synthetic skilled forgery attacks, Tabassi *et al.* [125] for simulating altered fingerprints and the contributions of Arifoglu and Bouchachia [126] for simulation of (abnormal) behaviour observed for dementia patients.

4.2 Addressing privacy concerns

Data collection is often governed by strict rules to preserve the privacy of individuals. For settings in which

Fig. 13: Feng *et al.* [79] generated synthetic faces with 11 yaw rotations (ranging from −50° to 50° over a step size of 5°) and 5 pitch rotations (ranging from −30° to 30° over a step size of 5°). Therefore, augmenting the synthetic dataset with the real training set increased the size of the training set. Furthermore, synthetic data provided face images with diverse pose variations. As a result, facial landmark detection performance improved *w.r.t.* variations in facial poses.

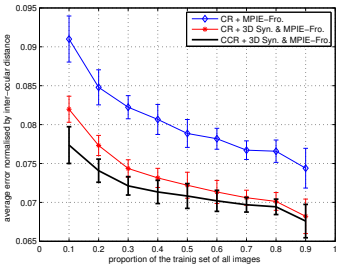


Fig. 14: Feng *et al.* [79] selected $44,820$ images from the Multi-PIE [131] as the training set and augmented it with $8,965$ synthetic 2D face images. The authors demonstrated that the face detection error of the cascaded regression (CR) based method significantly decreased, when trained with augmented data (plot in red) compared to when the landmark detection model was trained on only real faces (plot in blue). Motivated by this observation, the augmented data was used to train the proposed method based on cascaded collaborative regression (CCR, plot in black) to achieve the best face detection performance.

data collection is challenging due to time, cost, or privacy constraints, the generation of synthetic data can be seen as an efficient and privacy-preserving alternative [127]. However, a challenge with these applications has been to ensure that synthetic data has a similar distribution (for instance, distribution of minutiae in fingerprints [109], or distribution of sample quality scores [109], [127]) as the real data. Many studies demonstrated that synthetic data with similar characteristics to the real data can be generated and used, rather than the privacy-constrained real data. One such study includes the generation of $50,000$ synthetic face images each using StyleGAN [18] and StyleGAN2 [5] for face recognition applications in face recognition systems at the Schengen border [127]. The authors demonstrated that realistic face images with image quality scores similar to real faces can be generated. In addition, the authors compared face recognition performance of models trained on synthetic and real data and reported only minor differences, see Figure 12. Similar to Zhang *et al.* [127], Bozkir *et al.* [128] and Hillerström *et al.* [129] proposed to generate synthetic data for applications implying gaze estimation and finger vein recognition, respectively, in order to circumvent privacy issues, occurring when publicly sharing human data. However, the privacy-preserving property of synthetic datasets is closely related to the underlying generation model as recent studies have shown that DPMs can leak information from their training datasets [130] by learning to copy individual samples. Likewise, concerns over identity leakage have been raised for GANs [29].

4.3 Increasing the size and diversity of training dataset

Training DNNs requires a tremendous amount of data. At the same time, datasets in human analysis have often very limited samples. However, training with smaller datasets may lead to poor generalization onto real-world test examples. Therefore, several studies in human analysis advocated augmentation through synthetic data. Augmentation with synthetic data improves diversity by introducing more variations in training data, as well as increases the size of the training set. Training with a more extensive and diverse set leads to improved training and generalizability of the trained model on the test data.

Feng *et al.* [79] discussed the limited availability of annotated datasets to train a facial landmark detection model. The authors generated $8,965$ synthetic 2D face images to address this limitation with 11 different yaw rotations and five pitch rotations (see Figure 13). The authors augmented the training set for landmark detection and found that the face detection error reduced significantly after training on the augmented dataset (see Figure 14). Similarly, Masi *et al.* [132] augmented the training set of face images using augmentations that introduce variations in pose and shape. The authors demonstrated that rank-1 face recognition accuracy on the IJB-A dataset [133] improved from 94.6% to 96.2% after augmentation with synthetic samples. Several other studies additionally advocated augmenting the training set with synthetic data. Some of these studies include applications in human posture recognition [56], brain-based authentication [55], face photo-sketch recognition [134] , and cross spectral face recognition [135].

We observe that due to the simplicity and low computational requirements, perturbation of training samples to generate semi-synthetic data remains the most frequently used method towards increasing size and diversity of a training dataset.

4.4 Assessing scalability of systems

Evaluation of scalability of large-scale systems such as the Aadhar database maintained by the unique identification authority of India requires assessment of a system's performance for a colossal number of enrollees, sometimes up to a billion (Aadhar has 1.32 billion enrollments till 31 October 2021[7]). Scalability analysis of automated systems is crucial to assess whether these can be deployed for large-scale real-world applications. However, the collection of such large-scale datasets pertaining to humans is often challenging. To address this problem, researchers proposed to generate large-scale synthetic data. Such synthetic data is instrumental for scalability analyses of human analysis systems.

7. https://uidai.gov.in/

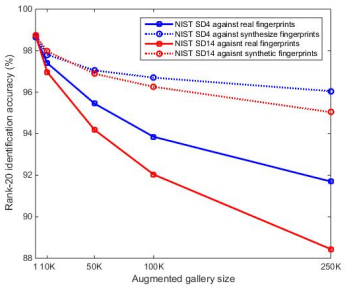


Fig. 15: Cao and Jain [109] generated synthetic fingerprints and augmented the gallery of standard real fingerprint databases to assess the scalability of state-of-the-art fingerprint search algorithms [136]. The authors found that the rank-20 identification accuracies dropped as the gallery was augmented with synthetic fingerprints. These results signified the usefulness of synthetic data, in order to assess the scalability of systems.

We note that the scalability can either be evaluated with system-relevant metrics (*e.g.*, throughput rate) or metrics that reflect the employed algorithms' performance or pre-trained models. According to the work of Sumi *et al.* [137], synthetic evaluation datasets have to comply with following three criteria.

1) **Privacy.** There shall not be a link between a synthetic sample to one of the individuals contained in the training dataset.
2) **Precision.** The performance of a pre-trained model evaluated with synthetic data shall be equal to the performance reported based on real data.
3) **Universality.** The precision shall be consistent across the evaluation of different pre-trained models.

Wilson *et al.* [138] demonstrated that the identification performance of a fingerprint recognition system drops linearly with the increase of enrolment records in the gallery. This observation motivated Cao and Jain [109] to generate 10 million synthetic rolled fingerprints using I-WGAN [139] in order to evaluate the scalability of fingerprint search algorithms. Similar to the trend observed for real data [138], the authors found that the rank-20 accuracy on NIST SD4[8] accuracy drops from 98.7% to 96.1% after the gallery was augmented with 250K synthetic fingerprints generated by the authors. Related to that, the report NIST SD14 [140] indicated that the rank-20 accuracy drops from 98.7% to 95.0% (see Figure 15).

Recently, Colbois *et al.* [63] analysed the verification accuracy and privacy of synthetic face images generated with StyleGAN2 [5] and InterFaceGAN [20]. The authors introduced a synthetic version of the Multi-PIE dataset [131] (Synth-Multi-PIE), representing the same factors of variation. The precision was assessed following the evaluation protocol of [131], identifying only minor performance differences between Synth-Multi-PIE and Multi-PIE. Similar studies on scalability analysis using synthetic data have been conducted for hand-shape biometrics recognition [141], face recognition [75], and iris verification [23]. However, as also noted in Section 5, synthetic data may not always be truly representative of real data. Hence, assessing scalability using synthetic data may suffer from reliability issues.

8. https://www.nist.gov/srd/nist-special-database-4

Fig. 16: Costa *et al.* [118] introduced a large-scale synthetic human action recognition dataset to promote the design of unsupervised domain adaptation methods for minimizing the cost and human effort in acquiring a large annotated dataset for human action recognition.

4.5 Providing annotated data for supervision

4.5.1 *Supervised Learning*

Numerous applications can be formulated as a supervised learning problem, for which annotation might be challenging to obtain. For such applications, representative synthetic samples and their annotations are generated in order to train models in supervised learning manner [83], [142], [8], [143], [144], [145], [146], [82], [147], [148]. Feng *et al.* [79] argued that manually annotated facial landmarks are often inaccurate for occl5uded facial regions. The annotations of synthetic faces generated from a 3D model are correct for all different pose variations as these are direct projections to 2D from 3D. Therefore, the authors used a synthetic dataset to obtain reliable and consistent annotations for various image variations. Some applications have exploited synthetic data to learn a transformation from distorted to clean samples. Associated to this direction, Dieckmann *et al.* [149] proposed to learn the pre-aligning of fingerprint images through horizontally and vertically translated and rotated synthetic fingerprints. Likewise, Zhang *et al.* [150], and Joshi *et al.* [151] utilized synthetic data to learn blind inpainting of face images, and enhancement of fingerprints, respectively.

4.5.2 *Unsupervised domain adaptation*

Supervised DNNs require a massive amount of manually annotated training data. However, collection, and particularly annotation of such is tedious, time-consuming and expensive. Furthermore, many human analysis applications require annotations by domain experts [152], or reliable annotations cannot be obtained for the real data [153]. To address this challenge, researchers proposed to train models on a synthetic training dataset whose annotations can be computationally acquired. However, a huge gap in model performance was observed between real and synthetic data due to the visible *domain shift* (see Figure 16). Researchers adapted models to unannotated real-world datasets, in order to reduce the performance gap between real and synthetic data. An important application of unsupervised domain adaptation of human analysis models includes the contributions of Wang *et al.* [40] [62]. The authors exploited

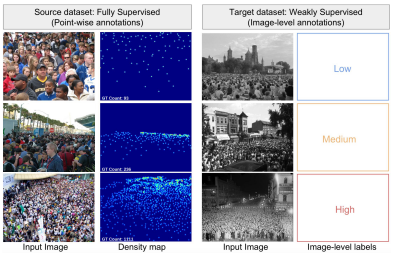


Fig. 17: Sindagi *et al.* [156] studied domain adaptation of a crowd counting model. While the source had pixel-level annotations, the target data was annotated on image level and only provided weak supervision.

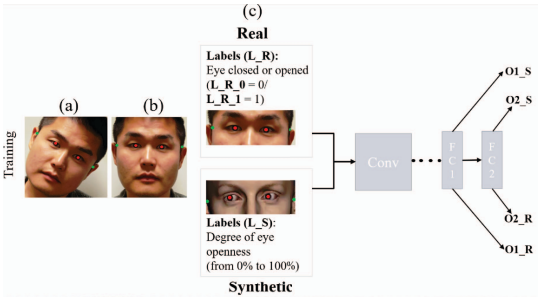


Fig. 18: Mequanint *et al.* [157] proposed weakly supervised learning in an eye-closeness estimation model. The model exploited synthetic annotated data that provided a degree of openness of eyes, whereas the real data only provided weak supervision, whether the eye is open or closed. Thus, annotated synthetic data can be used to enable learning in weakly supervised learning.

$15,212$ synthetic labelled crowd scene images containing more than $7,000,000$ subjects for the purpose of training a model for pixel-level understanding in a crowd. However, instead of directly using the synthetic data, the authors firstly translated synthetic images into realistic images using a GAN. This was beneficial in reducing the domain gap between synthetic and real data. Next, the model was trained on translated images instead of actual real images. The authors reported that the structural similarity index measure (SSIM) value improved from 0.554 to 0.660 after exploiting the synthetic crowd counting dataset.

Joshi *et al.* [152] highlighted the dependence of state-of-the-art fingerprint segmentation models on annotated data as a means to obtain satisfactory performance on a newly introduced fingerprint capture device. To mitigate this limitation, the authors only used synthetic data (source domain) annotations to learn fingerprint segmentation. To adapt the model to a new fingerprint capture device (target domain), the authors aligned the source and target domain features using recurrent adversarial learning. Extending the theme of unsupervised domain adaptation, Bondi *et al.* [153] argued that annotations of thermal infrared videos were often erroneous and therefore proposed to train the detection and tracking model on a synthetic dataset, adapting subsequently to a real dataset. Several applications spanning areas such as face recognition [154], person re-identification [36], human action recognition [118] and head pose estimation [155] successfully exploited synthetic data to eliminate the need for annotations of real data through unsupervised domain adaptation.

4.5.3 Weakly supervised learning

Synthetic annotated data has been utilized in weakly supervised learning (see Figure 17) aiming to introduce a higher degree of supervision. For instance, Mequanint *et al.* [157] highlighted the unavailability of annotated data for training an eye-openness estimation model (see Figure 18). To alleviate this issue, the authors generated 1.3 million annotated synthetic eye images with varying levels of eye openness to enable supervised learning. Furthermore, to counter the domain shift between real and synthetic eye images, the authors exploited weak supervision (eyes simply open or closed). It was demonstrated that the classification (open/close) accuracy improves from 96.30% to 100% after utilizing synthetic data. Deviating from the above, Zhang *et al.* [158] generated weakly labelled face images (labels as bounding box and class) using a deep convolutional generative adversarial network (DCGAN) [102] and used a limited amount of fully annotated real data (labels as landmark vector, bounding box and class). A weakly supervised learning framework was used to train the facial landmark detection model, which improved the average error distance for landmark detection on the labelled face parts in the wild (LFPW) dataset [159] from 4.25 to 3.12 after utilizing synthetic faces.

Synthetic data offers a remarkable substitute for an array of applications where annotations are not available. However, as we note in Section 5, such datasets are often not made publicly available, leading to unfair comparison among different baselines, as well as rendering reproducibility challenging.

4.6 Model initialization

DNNs impart a large number of parameters and, therefore, require a large amount of training data to avoid overfitting, *e.g.*, ImageNet incorporates approximately 1.2 million annotated images. However, in human analysis often only limited annotated training sets are publicly available, including hundreds or thousands of images. Therefore, once again synthetic data is advantageous in alleviating the need for a large amount of training data required for training data-hungry DNNs. It is common practice to generate annotated synthetic datasets and use such to pre-train deep models, which are then fine-tuned with annotated real data. A number of studies demonstrated that such pre-training with synthetic datasets leads to better performance than training directly on the real dataset. In one of the recent studies, Engelsma *et al.* [17] demonstrated that performance gain was observed by a DNN-based fingerprint recognition model (DeepPrint) [160] that was pre-trained on synthetic fingerprints and fine-tuned on real fingerprints. The authors generated 525K synthetic fingerprints for pre-training DeepPrint and fine-tuned it on 25K fingerprints from the NIST SD302 database [161]. The authors then assessed the fingerprint recognition performance of DeepPrint on NIST SD4 database[9], with and without pre-training with synthetic

9. https://www.nist.gov/srd/nist-special-database-4

data. The authors observed that the true acceptance rate (TAR) @ false acceptance rate (FAR)=0.01% improves from 73.37% to 87.03%, when pre-trained with synthetically generated fingerprints.

Similarly, Wang *et al.* [40] trained a pixel-level crowd understanding model on large-scale synthetic data ($15,212$ images of $7,625,843$ subjects) and fine-tuned it on labelled real data. The mean square error decreased by 14.1% after pre-training on synthetic data was noted, compared to the performance of the crowd counting model pre-trained on ImageNet dataset [162]. Similar trends were observed for other applications analyzing human data including speech recognition [50], hand shape recognition [163], head pose estimation [53], eye gaze tracking [164], re-identification of individuals [117] and face recognition [165].

4.7 Enforcing consistency regularization

4.7.1 Contrastive Learning

Contrastive learning is a learning paradigm that ensures that representations of similar samples must be close, whereas representations of dissimilar samples are far apart in the latent space. Various studies exploited synthetically augmented data to generate similar samples for a given input. Subsequently, using contrastive learning, the model was encouraged to have similar representations for the original and the augmented input samples. Ryoo *et al.* [166] introduced different low resolution (LR) transformations into videos and trained an activity recognition model such that the images obtained from the same scene, pertaining to different pixel values due to LR transformation, shared a common embedding (see Figure 19). The authors demonstrated that the classification accuracy under low-resolution constraints improves from 31.50% to 37.70% after using synthetic data. Neto *et al.* [167] applied augmentation techniques to generate synthetically masked faces. Contrastive learning brought then representations of masked and unmasked faces of the same data subject close to each other. The authors demonstrated that the model trained using the synthetically masked images outperformed existing standard face recognition systems on masked face recognition. Several other applications in speaker recognition [168], face recognition [169], person re-identification [106], [170] and electrocardiogram (ECG) based authentication [171], proposed to generate synthetic data for exploiting contrastive learning.

4.7.2 Self-supervision

Self-supervision is an unsupervised learning paradigm, through which a model can be regularized by introducing an auxiliary task. Several approaches in human analysis have introduced transformations to an input data to generate synthetic labelled data for training the auxiliary task in a supervised manner. For example Zhou *et al.* [172] proposed rotate-and-render, an augmentation technique that rotates faces back and forth in $3D$ space and subsequently renders them back in $2D$ (see Figure 20). Such augmentation strategy ensured consistency regularization, while training face recognition models. As a result, $TAR@FAR = 0.001$ on the IJB-A dataset improved from 80.00% to 82.48% after

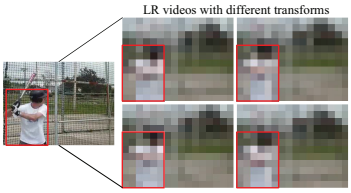


Fig. 19: Ryoo *et al.* [166] exploited different low-resolution transformations towards synthetically generating videos with the same scene and different pixel values due to changes in resolution. Later, a Siamese network was trained that ensures that the feature representations of the original and augmented video frames were similar. Thus, synthetic data can be used to ensure consistency among representations learnt by a human analysis model.

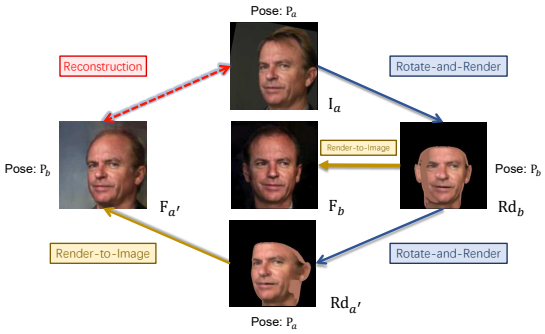


Fig. 20: Zhou *et al.* [172] proposed rotate-and-render augmentation that given a face image generates a synthetic face image with a varying pose. Later, the synthetic face was rendered back to the original pose. Such an augmentation ensured self-supervision in face recognition models. As a result, the model learned to preserve consistency in identity information while varying facial poses.

introducing self-supervision through the proposed augmentation strategy. Other applications utilizing synthetic data for self-supervision include deepfake detection [173], facial expression recognition [174], face recognition [175] and sleep recognition [176].

4.7.3 Few-shot learning

Few-shot learning is characterized by learning with a limited number of samples. Specifically, in order to compensate for limited availability of data and to promote the *learning to learn* paradigm, augmentation strategies simulate challenging real-world scenarios and ensure consistency in prediction for real and augmented input sample. Ge *et al.* [177] proposed in this context a knowledge distillation framework to improve face recognition performance under limited data and low resolution constraints. The face recognition model was trained on high-resolution face images, serving as teacher network. The authors then synthetically generated low-resolution face images and trained the student model such that the output of the student model on the synthetic low-resolution face was close to the output of the teacher model on the real high-quality face image (see Figure 21). The associated performance of face verification on the UMDFace dataset [178] improved from

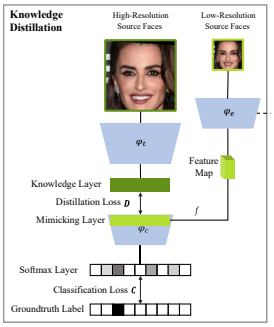


Fig. 21: Ge *et al.* [177] proposed a knowledge distillation framework for few-shot face recognition in the wild. The authors exploited consistency regularization among the output of the teacher model for the high-quality input and output of the student model for synthetic low-quality face images. Therefore, synthetic data can be used to enforce consistency regularization for improved performance of the human analysis model in few-shot learning scenarios.

67.59% to 73.58% after knowledge distillation compared to training the student model directly on synthetic faces. Thus, consistency regularization between real and synthetic data improved the face recognition performance with few-shot learning. Other studies utilizing synthetic data for few-shot learning in human analysis include applications in attribute-based person search [179], deepfake detection [180], login authentication [181], signature verification [182], and gaze estimation [134].

4.8 Mitigating dataset bias and ensuring fairness

Human datasets often contain demographic bias *w.r.t.* attributes such as ethnicity, gender, or age [184]. In addition, collected datasets might be biased to a certain group of labels [66]. Synthetic data is able to balance and unbias datasets beneficial in training and designing fair and unbiased human analysis models. Georgopoulos *et al.* [183] exploited an attribute-transfer based approach to balance underrepresented demographic groups in training datasets. Attributes such as skin tone, gender, and age were transferred into given training samples (see Figure 22) towards creation of an unbiased training dataset. In the related study the accuracy of face recognition on dark-skinned women over 60 years old characterized by true positive rate (TPR) improved by 20% on the UNCW dataset [185] after training on the training set augmented with synthetic faces.

Similarly, Niinuma *et al.* [66] discussed that real datasets employed for facial action detection are not balanced *w.r.t.* action unit (AU) intensity labels. To address this limitation, the authors generated a balanced training set using GANimation [186]. The generated balanced training dataset was used to train the facial action detector, with the related inter-rater reliability score of AU intensity level estimation, improving from 48.90% to 52.50% after training the model on synthetic data, as opposed to training on real data. Several other studies in face recognition [24] [108] confirmed the ability of synthetic data to train unbiased and fair models.

Again, as the representation ability of synthetic samples can be questionable (see Section 5), using synthetic samples for under-represented classes or demographics can affect the reliability of models trained on such datasets and may adversely impact related deployment in real-world.

4.9 Inducing digital perturbation attacks

Synthetic data is particularly instrumental in creating novel attacks on biometric systems. One prominent study introduced DeepMasterPrints [27], which aimed to generate one masterprint, namely a synthetic fingerprint that was designed to impersonate a set of fingerprints and falsely match with a large number of non-mated enrollees in the enrolment database (see Figure 23). This presentation attack for fingerprint recognition systems employed a GAN, where the latent input variables in the generator network were obtained using a covariance matrix adaptation evolution technique. The associated false match rate (FMR) of 0.1% increased via DeepMasterPrints to 8.61% on the NIST 9 dataset [140], as well as to 22.50% on the FingerPass DB7 dataset [187]. Additional attacks facilitated by synthetic data include those in iris recognition [15], [188], [189], face recognition [190] and fingerprint recognition [191].

A related direction has to do with digital human creation [2], [18], [192], as well as with manipulation of human faces [193], [194], [195]. Specifically, a face image of a target individual being superimposed on a video of a source individual has been widely accepted and referred to as *deepfake* (see Figure 24). Deepfakes entail several challenges and threats, given that (a) such manipulations can fabricate animations of subjects involved in actions that have not taken place and (b) such manipulated data can be circumvented nowadays rapidly via social media. Deepfakes are considered in human analysis as digital perturbation attacks, attracting large interest by their own right, with overview articles focusing on deepfake creation and detection [28], [196], [197], as well as adversarial attacks and defences in images, graphs, and text [198]. We note that similarly morphing attacks can be introduced using synthetic data [199]. A morphing attack is characterized by a synthetic image for which the authentication system is compelled to match with two contributing subjects instead of one. A morphed image is usually generated by aligning and blending images of two different contributors. For a comprehensive survey on published morphing attacks and associated detection methods, we refer to related overview articles [200], [201].

4.10 Learning by synthesis

A machine learning model can be categorized as a discriminative or generative model. The former learns a conditional distribution $p(y|x;\theta)$, where y denotes the output y for the input sample x and θ signifies model parameters. A generative model learns the joint distribution $p(x,y)$ and hence learns the distribution of data by learning to generate synthetic data. Such model is able to generalize on new and unseen test examples. A related seminal work [202] presented a hierarchical generative model, which jointly synthesizes eye images in a top-down approach, while estimating eye gaze in a bottom-up approach. A further generative modelling-based approach includes a relativistic average standard GAN (RaSGAN) [203] by Yadav *et al.*. RaSGAN was trained to generate synthetic iris images,

Fig. 22: Georgopoulos *et al.* [183] improved intra-class diversity in the training set by transferring demographic attributes (left to right): age, skin tone and gender. The authors demonstrated that training with diverse synthetic samples of the same subject is instrumental in mitigating demographic bias observed in face recognition models.

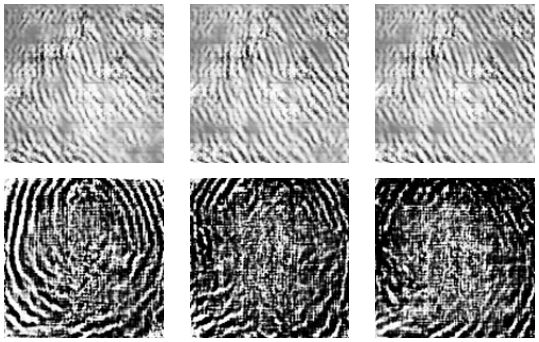


Fig. 23: DeepMasterPrints [27] by Bontrager *et al.* constitutes a synthetic fingerprint *masterprint* aimed at presentation attacks on fingerprint recognition systems. The top and bottom illustrate the masterprints for the rolled fingerprints and fingerprints acquired using a capacitive capture device. The first, second and third columns represent the masterprint to achieve a false match rate (FMR) of 0.01%, 0.1%, and 1%, respectively.

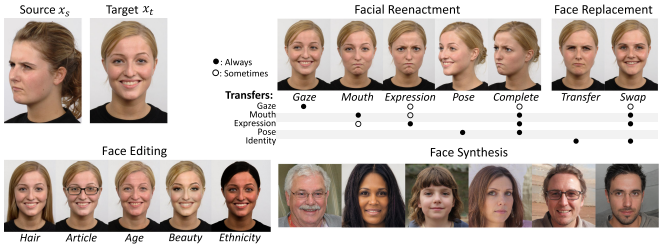


Fig. 24: Synthetic videos called Deepfakes with varying attributes, can be generated with th induce an attack to ruin the public perception of an individual [28].

demonstrating the ability of its discriminator to generalize better on new and unseen presentation attacks. Several approaches learning to synthesize data for improved model performance were proposed for re-identification of individuals [204] [106] and face recognition [205] [206].

5 Open Challenges and Discussion

We discussed benefits and means to generate and use synthetic datasets, placing emphasis on synthetic datasets being instrumental in mitigating challenges associated to real datasets. Despite related advances, there are a number of open research problems in this expanding field.

1) *Identity leakage*. Studies that advocate using synthetic data for alleviating the privacy issues related to human data frequently do not conduct supporting experiments to show that there is no identity leakage from the training dataset [29], [130]. Such an assessment is critical to address privacy concerns related to sharing data for applications in human analysis. For instance, Engelsma *et al.* [17] computed comparison scores between training samples and the synthetically generated fingerprints. Only 0.04% of the training samples obtained comparison scores above a threshold, and all such samples were removed from the synthetic dataset before introducing it in the public domain. Similar practices need to be adopted by the research community working in human analysis to mitigate any identity leakage.
2) *Lack of diversity*. The development of synthetic datasets in human analysis, generally speaking requires the generation of mated and non-mated samples. Recently, Grimmer *et al.* [207] emphasised the challenge of approximating the full intra-identity variation of real datasets. Mated samples were obtained through minor manipulations of various semantic attributes in a given sample. However, the generated datasets still lacked diversity compared to real-world datasets. Another challenge has to do with creating synthetic datasets balanced *w.r.t.* demographics. Generative models are often trained on biased datasets, thus lowering the generation quality of synthetic samples from underrepresented classes. We note that the current working draft of ISO/IEC 19795-10 [208] aims at quantifying the biometric system performance variation across demographic groups, hence providing a standardized and consistent evaluation framework to assess the diversity of synthetic datasets.
3) *Representation ability*. Numerous scientific work, particularly in biometrics [127], [109], have observed that while the generated synthetic data appears realistic, its characteristics represent notable differences from real biometric samples. Such observations question the representation ability of generated synthetic data and motivate the design of representative synthetic data generation methods. For instance, synthetic videos (deepfakes) frequently incorporate artefacts *e.g.*, in the eye or lip region. In addition, characteristics/semantics in synthetic data differs from those in real samples. For instance, Gottschlich and Huckemann [209] demonstrated the distribution of minutiae in synthetic fingerprints generated by SFinGe [6] was different from the

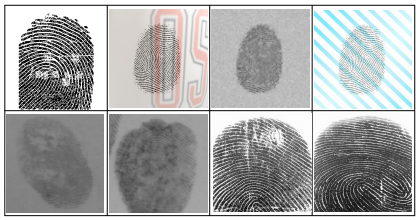


Fig. 25: First row: synthetically distorted training samples. Second row: real testing samples for fingerprint enhancement algorithms, as used in [84]. The performance of the fingerprint enhancement model was directly dependent on how well the synthetic data modelled the noise observed in real fingerprints. Therefore, synthetically distorted training data must be publicly available to ensure fair comparison among different fingerprint enhancement algorithms.

one observed for real fingerprints. Therefore, the representation ability of synthetic data needs to be carefully validated before exploiting it for real-world applications.

4) *Lack of comparison*. While state-of-the-art works in human analysis have gradually exploited synthetic data, related datasets are often not shared publicly. This is crucial, as the performance of human analysis models is directly dependent on how well synthetic data aligns with the testing dataset (see Figure 25). In the case of Figure 25, the fingerprint enhancement performance is dependent not only on the enhancement model but also on how carefully curated synthetic training data is. Therefore, to foster reproducibility and ensure a fair comparison among different methods, there is a need to share synthetic datasets publicly.

6 Conclusions and Future Applications

A review of the human analysis literature suggests that research in synthetic data is on the rise. This expansion is due to the large number of associated benefits in settings including enrichment and replacement of existing real datasets. In this article, we revisited methods that have been developed for generation and exploitation of synthetic data in human analysis. In particular, we discussed techniques for generating semi-synthetic and fully synthetic data.

In addition, we provided examples of related applications, elaborating on simulation of complex scenarios, mitigating bias and privacy concerns, increasing the size and diversity of training datasets, assessing scalability of systems, providing additional data for supervision, pre-training and fine-tuning of DNNs, enforcing consistency regularization, as well as adversarial attacks. Finally, we discussed open research problems in synthetic data research. We believe that synthetic data has the ability to mitigate issues related to privacy, scalability, and generalization of unseen data. Although currently synthetic data is abundantly utilized in human analysis, we believe that additional research directions including active learning, knowledge distillation and source-free domain adaptation will benefit in future from synthetic data. Furthermore, upcoming synthetic data generation mechanisms such as scene graphs and diffusion models will be exploited in future to generate data for applications in human analysis.

Acknowledgments

This research work has been supported by the French Government, by the National Research Agency (ANR) under Grant ANR-18-CE92-0024, project RESPECT, as well as by the German Federal Ministry of Education and Research and the Hessian Ministry of Higher Education, Research, Science and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE.

References

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[2] T. Karras, M. Aittala, S. Laine, E. Härkönen, J. Hellsten, J. Lehtinen, and T. Aila, "Alias-free generative adversarial networks," Advances in Neural Information Processing Systems, vol. 34, 2021.

[3] R. Gal, D. C. Hochberg, A. Bermano, and D. Cohen-Or, "Swagan: A style-based wavelet-driven generative model," ACM Trans. on Graphics, vol. 40, no. 4, pp. 1–11, 2021.

[4] Q. Feng, C. Guo, F. Benitez-Quiroz, and A. M. Martinez, "When do gans replicate? on the choice of dataset size," in Proc. of the IEEE/CVF Intl. Conf. on Computer Vision, pp. 6701–6710, 2021.

[5] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of stylegan," in Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, pp. 8110–8119, 2020.

[6] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, "Synthetic fingerprint generation," Handbook of Fingerprint Recognition, pp. 271–302, 2009.

[7] S. A. Grosz and A. K. Jain, "Spoofgan: Synthetic fingerprint spoof images," arXiv preprint arXiv:2204.06498, 2022.

[8] G. Varol, J. Romero, X. Martin, N. Mahmood, M. J. Black, I. Laptev, and C. Schmid, "Learning from synthetic humans," in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, pp. 109–117, 2017.

[9] H. Hwang, C. Jang, G. Park, J. Cho, and I.-J. Kim, "Eldersim: A synthetic data generation platform for human action recognition in eldercare applications," arXiv preprint arXiv:2010.14742, 2020.

[10] H. Dou, W. Zhang, P. Zhang, Y. Zhao, S. Li, Z. Qin, F. Wu, L. Dong, and X. Li, "Versatilegait: A large-scale synthetic gait dataset with fine-grained attributes and complicated scenarios," arXiv preprint arXiv:2101.01394, 2021.

[11] 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 Proc. of the IEEE/CVF Intl. Conf. on Computer Vision, pp. 3681–3691, 2021.

[12] P. İrtem, E. İrtem, and N. Erdoğmuş, "Impact of variations in synthetic training data on fingerprint classification," in Intl. Conf. of the Biometrics Special Interest Group, pp. 1–4, IEEE, 2019.

[13] F. K. Dankar and M. Ibrahim, "Fake it till you make it: Guidelines for effective synthetic data generation," Applied Sciences, vol. 11, no. 5, p. 2158, 2021.

[14] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, pp. 8789–8797, 2018.

[15] S. Yadav, C. Chen, and A. Ross, "Synthesizing iris images using rasgan with application in presentation attack detection," in Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops, pp. 0–0, 2019.

[16] S. I. Nikolenko, Synthetic data for deep learning, vol. 174. Springer, 2021.

[17] J. J. Engelsma, S. A. Grosz, and A. K. Jain, "Printsgan: Synthetic fingerprint generator," arXiv preprint arXiv:2201.03674, 2022.

[18] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in Proc of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, pp. 4401–4410, 2019.

[19] Y. Alaluf, O. Patashnik, and D. Cohen-Or, "Only a matter of style: Age transformation using a style-based regression model," ACM Trans. on Graphics (TOG), vol. 40, no. 4, pp. 1–12, 2021.

[20] Y. Shen, J. Gu, X. Tang, and B. Zhou, "Interpreting the latent space of gans for semantic face editing," in Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, pp. 9243–9252, 2020.

[21] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in IEEE computer society Conf. on Computer Vision and Pattern Recognition, vol. 1, pp. 947–954, 2005.

[22] ISO/IEC JTC1 SC37 Biometrics, ISO/IEC 2382-37:2017 Information Technology - Vocabulary - Part 37: Biometrics. Intl. Organization for Standardization, 2017.

[23] P. Drozdowski, C. Rathgeb, and C. Busch, "Sic-gen: A synthetic iris-code generator," in Intl. Conf. of the Biometrics Special Interest Group, pp. 1–6, IEEE, 2017.

[24] A. Kortylewski, B. Egger, A. Schneider, T. Gerig, A. Morel-Forster, and T. Vetter, "Analyzing and reducing the damage of dataset bias to face recognition with synthetic data," in Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops, pp. 0–0, 2019.

[25] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," Journal of big data, vol. 6, no. 1, pp. 1–48, 2019.

[26] H. Qiu, B. Yu, D. Gong, Z. Li, W. Liu, and D. Tao, "Synface: Face recognition with synthetic data," in Proc. of the IEEE/CVF Intl. Conf. on Computer Vision, pp. 10880–10890, 2021.

[27] P. Bontrager, A. Roy, J. Togelius, N. Memon, and A. Ross, "Deepmasterprints: Generating masterprints for dictionary attacks via latent variable evolution," in IEEE 9th Intl. Conf. on Biometrics Theory, Applications and Systems, pp. 1–9, 2018.

[28] Y. Mirsky and W. Lee, "The creation and detection of deepfakes: A survey," ACM Computing Surveys, vol. 54, no. 1, pp. 1–41, 2021.

[29] P. Tinsley, A. Czajka, and P. Flynn, "This face does not exist... but it might be yours! identity leakage in generative models," in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 1320–1328, 2021.

[30] M. Fredrikson, S. Jha, and T. Ristenpart, "Model inversion attacks that exploit confidence information and basic countermeasures," in Proc. of the 22nd ACM SIGSAC Conf. on Computer and Communications Security, pp. 1322–1333, 2015.

[31] C. Xu, J. Ren, D. Zhang, Y. Zhang, Z. Qin, and K. Ren, "Ganobfuscator: Mitigating information leakage under gan via differential privacy," IEEE Trans. on Information Forensics and Security, vol. 14, no. 9, pp. 2358–2371, 2019.

[32] I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller, and E. Brossard, "The megaface benchmark: 1 million faces for recognition at scale," in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, pp. 4873–4882, 2016.

[33] B. Bozorgtabar, M. S. Rad, H. K. Ekenel, and J.-P. Thiran, "Using photorealistic face synthesis and domain adaptation to improve facial expression analysis," in 14th IEEE Intl. Conf. on Automatic Face & Gesture Recognition, pp. 1–8, 2019.

[34] A. Dantcheva, P. Elia, and A. Ross, "What else does your biometric data reveal? a survey on soft biometrics," IEEE Trans. on Information Forensics and Security, vol. 11, no. 3, pp. 441–467, 2016.

[35] ISO/IEC JTC1 SC37 Biometrics, ISO/IEC 30107-1. Information Technology - Biometric Presentation Attack Detection - Part 1: Framework. Intl. Organization for Standardization, 2016.

[36] S. Bak, P. Carr, and J.-F. Lalonde, "Domain adaptation through synthesis for unsupervised person re-identification," in Proc. of the European Conf. on Computer Vision, pp. 189–205, 2018.

[37] X. Sun and L. Zheng, "Dissecting person re-identification from the viewpoint of viewpoint," in Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, 2019.

[38] X. Shu, J. Tang, G.-J. Qi, W. Liu, and J. Yang, "Hierarchical long short-term concurrent memory for human interaction recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 3, pp. 1110–1118, 2019.

[39] N. Aranjuelo, S. García, E. Loyo, L. Unzueta, and O. Otaegui, "Key strategies for synthetic data generation for training intelligent systems based on people detection from omnidirectional cameras," Computers & Electrical Engineering, vol. 92, p. 107105, 2021.

[40] Q. Wang, J. Gao, W. Lin, and Y. Yuan, "Pixel-wise crowd understanding via synthetic data," Intl. Journal of Computer Vision, vol. 129, no. 1, pp. 225–245, 2021.

[41] C. Ménier, E. Boyer, and B. Raffin, "3d skeleton-based body pose recovery," in 3rd Intl. Symposium on 3D Data Processing, Visualization, and Transmission, pp. 389–396, IEEE, 2006.

[42] T. Hempel, A. A. Abdelrahman, and A. Al-Hamadi, "6d rotation representation for unconstrained head pose estimation," arXiv preprint arXiv:2202.12555, 2022.

[43] P. Patel, C.-H. P. Huang, J. Tesch, D. T. Hoffmann, S. Tripathi, and M. J. Black, "Agora: Avatars in geography optimized for regression analysis," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13468–13478, 2021.

[44] M. Fabbri, F. Lanzi, S. Calderara, A. Palazzi, R. Vezzani, and R. Cucchiara, "Learning to detect and track visible and occluded body joints in a virtual world," in Proc. of the European Conf. on Computer Vision, pp. 430–446, 2018.

[45] A. Ranjan, D. T. Hoffmann, D. Tzionas, S. Tang, J. Romero, and M. J. Black, "Learning multi-human optical flow," Intl. Journal of Computer Vision, vol. 128, pp. 873–890, 2020.

[46] D. T. Hoffmann, D. Tzionas, M. J. Black, and S. Tang, "Learning to train with synthetic humans," in Pattern Recognition: 41st DAGM German Conference, DAGM GCPR 2019, Dortmund, Germany, September 10–13, 2019, Proceedings 41, pp. 609–623, 2019.

[47] A. Roitberg, D. Schneider, A. Djamal, C. Seibold, S. Reiß, and R. Stiefelhagen, "Let's play for action: Recognizing activities of daily living by learning from life simulation video games," in IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems, pp. 8563–8569, IEEE, 2021.

[48] J. Kolberg, M. Grimmer, M. Gomez-Barrero, and C. Busch, "Anomaly detection with convolutional autoencoders for fingerprint presentation attack detection," IEEE Trans. on Biometrics, Behavior, and Identity Science, vol. 3, no. 2, pp. 190–202, 2021.

[49] R. J. Chen, M. Y. Lu, T. Y. Chen, D. F. Williamson, and F. Mahmood, "Synthetic data in machine learning for medicine and healthcare," Nature Biomedical Engineering, vol. 5, no. 6, pp. 493–497, 2021.

[50] J. J. Bird, D. R. Faria, A. Ekárt, C. Premebida, and P. P. Ayrosa, "Lstm and gpt-2 synthetic speech transfer learning for speaker recognition to overcome data scarcity," arXiv preprint arXiv:2007.00659, 2020.

[51] J. E. Tapia and C. Arellano, "Soft-biometrics encoding conditional gan for synthesis of nir periocular images," Future Generation Computer Systems, vol. 97, pp. 503–511, 2019.

[52] J. Han, S. Karaoglu, H.-A. Le, and T. Gevers, "Improving face detection performance with 3d-rendered synthetic data," arXiv preprint arXiv:1812.07363, 2018.

[53] S. Basak, P. Corcoran, F. Khan, R. Mcdonnell, and M. Schukat, "Learning 3d head pose from synthetic data: A semi-supervised approach," IEEE Access, vol. 9, pp. 37557–37573, 2021.

[54] J. J. Bird, D. R. Faria, C. Premebida, A. Ekárt, and P. P. Ayrosa, "Overcoming data scarcity in speaker identification: Dataset augmentation with synthetic mfccs via character-level rnn," in IEEE Intl. Conf. on Autonomous Robot Systems and Competitions, pp. 146–151, IEEE, 2020.

[55] T. Piplani, N. Merill, and J. Chuang, "Faking it, making it: Fooling and improving brain-based authentication with generative adversarial networks," in IEEE 9th Intl. Conf. on Biometrics Theory, Applications and Systems, pp. 1–7, IEEE, 2018.

[56] R. Gouiaa and J. Meunier, "Learning cast shadow appearance for human posture recognition," Pattern Recognition Letters, vol. 97, pp. 54–60, 2017.

[57] V. Ruiz, I. Linares, A. Sanchez, and J. F. Velez, "Off-line handwritten signature verification using compositional synthetic generation of signatures and siamese neural networks," Neurocomputing, vol. 374, pp. 30–41, 2020.

[58] M. Kim, F. Liu, A. Jain, and X. Liu, "Dcface: Synthetic face generation with dual condition diffusion model," in Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, pp. 12715–12725, 2023.

[59] G. Chen, Y. Zhu, Z. Hong, and Z. Yang, "Emotionalgan: generating ecg to enhance emotion state classification," in Proc. of

the Intl. Conf. on Artificial Intelligence and Computer Science, pp. 309–313, 2019.
[60] V. K. Melo, B. L. D. Bezerra, D. Impedovo, G. Pirlo, and A. Lundgren, "Deep learning approach to generate offline handwritten signatures based on online samples," IET Biometrics, vol. 8, no. 3, pp. 215–220, 2019.
[61] M. Öz, T. Danisman, M. Günay, E. Z. Sanal, Ö. Duman, and J. W. Ledet, "The use of synthetic data to facilitate eye segmentation using deeplabv3+," Annals of Emerging Technologies in Computing, vol. 5, no. 3, pp. 1–10, 2021.
[62] Q. Wang, J. Gao, W. Lin, and Y. Yuan, "Learning from synthetic data for crowd counting in the wild," in Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, pp. 8198–8207, 2019.
[63] L. Colbois, T. de Freitas Pereira, and S. Marcel, "On the use of automatically generated synthetic image datasets for benchmarking face recognition," in 2021 IEEE Intl. Joint Conf. on Biometrics (IJCB), pp. 1–8, IEEE, 2021.
[64] R. T. Marriott, S. Romdhani, and L. Chen, "A 3d gan for improved large-pose facial recognition," in Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, pp. 13445–13455, 2021.
[65] T. U. Ahmed, S. Hossain, M. S. Hossain, R. ul Islam, and K. Andersson, "Facial expression recognition using convolutional neural network with data augmentation," in 2019 Joint 8th Intl. Conf. on Informatics, Electronics & Vision (ICIEV) and 2019 3rd Intl. Conf. on Imaging, Vision & Pattern Recognition (icIVPR), pp. 336–341, IEEE, 2019.
[66] K. Niinuma, I. O. Ertugrul, J. F. Cohn, and L. A. Jeni, "Synthetic expressions are better than real for learning to detect facial actions," in Proc. of the IEEE/CVF Winter Conf. on Applications of Computer Vision, pp. 1248–1257, 2021.
[67] L. Zhu, K. Rematas, B. Curless, S. M. Seitz, and I. Kemelmacher-Shlizerman, "Reconstructing nba players," in 16th European Conf. on Computer Vision, pp. 177–194, Springer, 2020.
[68] Z. Cai, M. Zhang, J. Ren, C. Wei, D. Ren, Z. Lin, H. Zhao, L. Yang, C. C. Loy, and Z. Liu, "Playing for 3d human recovery," arXiv preprint arXiv:2110.07588, 2021.
[69] G. Varol, I. Laptev, C. Schmid, and A. Zisserman, "Synthetic humans for action recognition from unseen viewpoints," Intl. Journal of Computer Vision, vol. 129, no. 7, pp. 2264–2287, 2021.
[70] E. Hatay, J. Ma, H. Sun, J. Fang, Z. Gao, and H. Yu, "Learning to detect phone-related pedestrian distracted behaviors with synthetic data," in Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, pp. 2981–2989, 2021.
[71] C. R. de Souza12, A. Gaidon, Y. Cabon, and A. M. López, "Procedural generation of videos to train deep action recognition networks," 2017.
[72] J. Priesnitz, C. Rathgeb, N. Buchmann, and C. Busch, "Syncolfinger: Synthetic contactless fingerprint generator," Pattern Recognition Letters, 2022.
[73] F. Hillerström, A. Kumar, and R. N. J. Veldhuis, "Generating and analyzing synthetic finger vein images," in BIOSIG 2014 (A. Brömme and C. Busch, eds.), (Bonn), pp. 121–132, Gesellschaft für Informatik e.V., 2014.
[74] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," Computer Vision and Image Understanding, vol. 61, no. 1, pp. 38–59, 1995.
[75] M. Osadchy, Y. Wang, O. Dunkelman, S. Gibson, J. Hernandez-Castro, and C. Solomon, "Genface: Improving cyber security using realistic synthetic face generation," in Intl. Conf. on Cyber Security Cryptography and Machine Learning, pp. 19–33, Springer, 2017.
[76] T. Li, T. Bolkart, M. J. Black, H. Li, and J. Romero, "Learning a model of facial shape and expression from 4d scans," ACM Trans. Graph., vol. 36, no. 6, pp. 194–1, 2017.
[77] A. A. Osman, T. Bolkart, and M. J. Black, "Star: Sparse trained articulated human body regressor," in 16th Proc. of the European Conf. of Computer Vision, pp. 598–613, Springer, 2020.
[78] F. Wan, Y. Wu, X. Qian, Y. Chen, and Y. Fu, "When person re-identification meets changing clothes," in Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops, pp. 830–831, 2020.
[79] Z.-H. Feng, G. Hu, J. Kittler, W. Christmas, and X.-J. Wu, "Cascaded collaborative regression for robust facial landmark detection trained using a mixture of synthetic and real images with dynamic weighting," IEEE Trans. on Image Processing, vol. 24, no. 11, pp. 3425–3440, 2015.
[80] C. C. Charalambous and A. A. Bharath, "A data augmentation methodology for training machine/deep learning gait recognition algorithms," arXiv preprint arXiv:1610.07570, 2016.
[81] E. Wood, T. Baltrušaitis, L.-P. Morency, P. Robinson, and A. Bulling, "Learning an appearance-based gaze estimator from one million synthesised images," in Proc. of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications, pp. 131–138, 2016.
[82] M. Ibsen, C. Rathgeb, P. Drozdowski, and C. Busch, "Face beneath the ink: Synthetic data and tattoo removal with application to face recognition," arXiv preprint arXiv:2202.05297, 2022.
[83] K. Cao and A. K. Jain, "Latent orientation field estimation via convolutional neural network," in Intl. Conf. on Biometrics, pp. 349–356, 2015.
[84] I. Joshi, A. Utkarsh, R. Kothari, V. K. Kurmi, A. Dantcheva, S. D. Roy, and P. K. Kalra, "Data uncertainty guided noise-aware preprocessing of fingerprints," in Intl. Joint Conf. on Neural Networks, pp. 1–8, 2021.
[85] L. Cardoso, A. Barbosa, F. Silva, A. M. Pinheiro, and H. Proença, "Iris biometrics: Synthesis of degraded ocular images," IEEE Trans. on information forensics and security, vol. 8, no. 7, pp. 1115–1125, 2013.
[86] J. Yin, S. Zhang, J. Xie, Z. Ma, and J. Guo, "Unsupervised person re-identification via simultaneous clustering and mask prediction," Pattern Recognition, p. 108568, 2022.
[87] A. K. Jain, D. Deb, and J. J. Engelsma, "Biometrics: Trust, but verify," IEEE Trans. on Biometrics, Behavior, and Identity Science, 2021.
[88] H. Wang, G. Wang, Y. Li, D. Zhang, and L. Lin, "Transferable, controllable, and inconspicuous adversarial attacks on person re-identification with deep mis-ranking," in Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, pp. 342–351, 2020.
[89] Y. Zhong and W. Deng, "Towards transferable adversarial attack against deep face recognition," IEEE Trans. on Information Forensics and Security, vol. 16, pp. 1452–1466, 2020.
[90] S. Soleymani, A. Dabouei, J. Dawson, and N. M. Nasrabadi, "Adversarial examples to fool iris recognition systems," in Intl. Conf. on Biometrics, pp. 1–8, IEEE, 2019.
[91] S. Marrone and C. Sansone, "On the transferability of adversarial perturbation attacks against fingerprint based authentication systems," Pattern Recognition Letters, vol. 152, pp. 253–259, 2021.
[92] P. Esser, R. Rombach, and B. Ommer, "Taming transformers for high-resolution image synthesis," in Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, pp. 12873–12883, 2021.
[93] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," arXiv preprint arXiv:1312.6114, 2013.
[94] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey, "Adversarial autoencoders," arXiv preprint arXiv:1511.05644, 2015.
[95] I. Tolstikhin, O. Bousquet, S. Gelly, and B. Schoelkopf, "Wasserstein auto-encoders," in Intl. Conf. on Learning Representations, 2018.
[96] A. K. Mondal, H. Asnani, P. Singla, and A. Prathosh, "Flexae: Flexibly learning latent priors for wasserstein auto-encoders," in Uncertainty in Artificial Intelligence, pp. 525–535, PMLR, 2021.
[97] B. Dai and D. Wipf, "Diagnosing and enhancing vae models," in Intl. Conf. on Learning Representations, 2018.
[98] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner, "beta-vae: Learning basic visual concepts with a constrained variational framework," 2016.
[99] H. Kim and A. Mnih, "Disentangling by factorising," in Intl. Conf. on Machine Learning, pp. 2649–2658, 2018.
[100] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," Advances in Neural Information Processing Systems, vol. 27, 2014.
[101] A. Mondal, A. Bhattacharjee, S. Mukherjee, H. Asnani, S. Kannan, and A. Prathosh, "C-mi-gan: Estimation of conditional mutual information using minmax formulation," in Conf. on Uncertainty in Artificial Intelligence, pp. 849–858, PMLR, 2020.
[102] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," arXiv preprint arXiv:1511.06434, 2015.
[103] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in Intl. Conf. on Machine Learning, pp. 214–223, PMLR, 2017.

[104] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, pp. 1125–1134, 2017.

[105] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in Proc. of the IEEE international Conf. on Computer Vision, pp. 2223–2232, 2017.

[106] H. Chen, Y. Wang, B. Lagadec, A. Dantcheva, and F. Bremond, "Joint generative and contrastive learning for unsupervised person re-identification," in Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, pp. 2004–2013, 2021.

[107] D. S. Trigueros, L. Meng, and M. Hartnett, "Generating photo-realistic training data to improve face recognition accuracy," arXiv preprint arXiv:1811.00112, 2018.

[108] Z. Zhai, P. Yang, X. Zhang, M. Huang, H. Cheng, X. Yan, C. Wang, and S. Pu, "Demodalizing face recognition with synthetic samples," in Proc. of the AAAI Conf. on Artificial Intelligence, vol. 35, pp. 3278–3286, 2021.

[109] K. Cao and A. Jain, "Fingerprint synthesis: Evaluating fingerprint search at scale," in Intl. Conf. on Biometrics, pp. 31–38, 2018.

[110] F. Deng, Z. Zhi, D. Lee, and S. Ahn, "Generative scene graph networks," in International Conference on Learning Representations, 2021.

[111] A. Kosiorek, H. Kim, Y. W. Teh, and I. Posner, "Sequential attend, infer, repeat: Generative modelling of moving objects," Advances in Neural Information Processing Systems, vol. 31, 2018.

[112] J. Yuan, T. Chen, B. Li, and X. Xue, "Compositional scene representation learning via reconstruction: A survey," arXiv preprint arXiv:2202.07135, 2022.

[113] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," Advances in Neural Information Processing Systems, vol. 33, pp. 6840–6851, 2020.

[114] C. Meng, Y. He, Y. Song, J. Song, J. Wu, J.-Y. Zhu, and S. Ermon, "Sdedit: Guided image synthesis and editing with stochastic differential equations," in International Conference on Learning Representations, 2022.

[115] C. Saharia, W. Chan, H. Chang, C. Lee, J. Ho, T. Salimans, D. Fleet, and M. Norouzi, "Palette: Image-to-image diffusion models," in ACM SIGGRAPH 2022 Conference Proceedings, pp. 1–10, 2022.

[116] M. Falkenberg, A. B. Ottsen, M. Ibsen, and C. Rathgeb, "Child face recognition at scale: Synthetic data generation and performance benchmark," arXiv preprint arXiv:2304.11685, 2023.

[117] I. B. Barbosa, M. Cristani, B. Caputo, A. Rognhaugen, and T. Theoharis, "Looking beyond appearances: Synthetic training data for deep cnns in re-identification," Computer Vision and Image Understanding, vol. 167, pp. 50–62, 2018.

[118] V. G. T. da Costa, G. Zara, P. Rota, T. Oliveira-Santos, N. Sebe, V. Murino, and E. Ricci, "Dual-head contrastive domain adaptation for video action recognition," in Proc. of the IEEE/CVF Winter Conf. on Applications of Computer Vision, pp. 1181–1190, 2022.

[119] M. Ariz, J. J. Bengoechea, A. Villanueva, and R. Cabeza, "A novel 2d/3d database with automatic face annotation for head tracking and pose estimation," Computer Vision and Image Understanding, vol. 148, pp. 201–210, 2016.

[120] A. Roitberg, D. Schneider, A. Djamal, C. Seibold, S. Reiß, and R. Stiefelhagen, "Let's play for action: Recognizing activities of daily living by learning from life simulation video games," 2021.

[121] Y. Feng, H. Feng, M. J. Black, and T. Bolkart, "Learning an animatable detailed 3d face model from in-the-wild images," ACM Trans. on Graphics, vol. 40, no. 4, pp. 1–13, 2021.

[122] E. R. Chan, C. Z. Lin, M. A. Chan, K. Nagano, B. Pan, S. De Mello, O. Gallo, L. J. Guibas, J. Tremblay, S. Khamis, et al., "Efficient geometry-aware 3d generative adversarial networks," in Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, pp. 16123–16133, 2022.

[123] N. Takemura, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi, "Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition," Trans. on Computer Vision and Applications, vol. 10, no. 1, pp. 1–14, 2018.

[124] S. Lai, L. Jin, L. Lin, Y. Zhu, and H. Mao, "Synsig2vec: Learning representations from synthetic dynamic signatures for real-world verification," in Proc. of the AAAI Conf. on Artificial Intelligence, vol. 34, pp. 735–742, 2020.

[125] E. Tabassi, T. Chugh, D. Deb, and A. K. Jain, "Altered fingerprints: Detection and localization," in IEEE 9th Intl. Conf. on Biometrics Theory, Applications and Systems, pp. 1–9, 2018.

[126] D. Arifoglu and A. Bouchachia, "Abnormal behaviour detection for dementia sufferers via transfer learning and recursive auto-encoders," in IEEE Intl. Conf. on Pervasive Computing and Communications Workshops, pp. 529–534, 2019.

[127] H. Zhang, M. Grimmer, R. Ramachandra, K. Raja, and C. Busch, "On the applicability of synthetic data for face recognition," in IEEE Intl. Workshop on Biometrics and Forensics, pp. 1–6, 2021.

[128] E. Bozkir, A. B. Ünal, M. Akgün, E. Kasneci, and N. Pfeifer, "Privacy preserving gaze estimation using synthetic images via a randomized encoding based framework," in ACM Symposium on Eye Tracking Research and Applications, pp. 1–5, 2020.

[129] F. Hillerström and A. Kumar, "On generation and analysis of synthetic finger-vein images for biometrics identification," Technical Report No. COMP-K-17, 2014.

[130] N. Carlini, J. Hayes, M. Nasr, M. Jagielski, V. Sehwag, F. Tramer, B. Balle, D. Ippolito, and E. Wallace, "Extracting training data from diffusion models," in 32nd USENIX Security Symposium, pp. 5253–5270, 2023.

[131] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-pie," Image and vision computing, vol. 28, no. 5, pp. 807–813, 2010.

[132] I. Masi, A. T. Tran, T. Hassner, G. Sahin, and G. Medioni, "Face-specific data augmentation for unconstrained face recognition," Intl. Journal of Computer Vision, vol. 127, no. 6, pp. 642–667, 2019.

[133] B. F. Klare, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. Allen, P. Grother, A. Mah, and A. K. Jain, "Pushing the frontiers of unconstrained face detection and recognition: Iarpa janus benchmark a," in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, pp. 1931–1939, 2015.

[134] Y. Yu, G. Liu, and J.-M. Odobez, "Improving few-shot user-specific gaze adaptation via gaze redirection synthesis," in Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, pp. 11937–11946, 2019.

[135] D. Anghelone, C. Chen, P. Faure, A. Ross, and A. Dantcheva, "Explainable thermal to visible face recognition using latent-guided generative adversarial network," in 16th IEEE Intl. Conf. on Automatic Face and Gesture Recognition, pp. 1–8, IEEE, 2021.

[136] K. Cao and A. K. Jain, "Fingerprint indexing and matching: An integrated approach," in IEEE Intl. Joint Conf. on Biometrics, pp. 437–445, 2017.

[137] K. Sumi, C. Liu, and T. Matsuyama, "Study on synthetic face database for performance evaluation," in Intl. Conf. on Biometrics, pp. 598–604, Springer, 2006.

[138] C. L. Wilson, C. I. Watson, M. D. Garris, A. Hicklin, et al., Studies of fingerprint matching using the NIST verification test bed (VTB). US Department of Commerce, Technology Administration, National Institute of . . . , 2003.

[139] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," Adv. in Neural Information Processing Systems, vol. 30, 2017.

[140] C. Watson, "Nist special database 14-mated fingerprint card pairs 2," National Institute of Standards and Technology, 1993.

[141] A. Morales, M. A. Ferrer, R. Cappelli, D. Maltoni, J. Fierrez, and J. Ortega-Garcia, "Synthesis of large scale hand-shape databases for biometric applications," Pattern Recognition Letters, vol. 68, pp. 183–189, 2015.

[142] E. Richardson, M. Sela, and R. Kimmel, "3d face reconstruction by learning from synthetic data," in 4th Intl. Conf. on 3D vision, pp. 460–469, 2016.

[143] S. Basak, H. Javidnia, F. Khan, R. McDonnell, and M. Schukat, "Methodology for building synthetic datasets with virtual humans," in 31st Irish Signals and Systems Conf., pp. 1–6, 2020.

[144] S. Park, X. Zhang, A. Bulling, and O. Hilliges, "Learning to find eye region landmarks for remote gaze estimation in unconstrained settings," in Proc. of the 2018 ACM Symposium on Eye Tracking Research & Applications, pp. 1–10, 2018.

[145] C. Roberto de Souza, A. Gaidon, Y. Cabon, and A. Manuel Lopez, "Procedural generation of videos to train deep action recognition networks," in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, pp. 4757–4767, 2017.

[146] Y. Xu, Y. Wang, J. Liang, and Y. Jiang, "Augmentation data synthesis via gans: Boosting latent fingerprint reconstruction,"

in IEEE Intl. Conf. on Acoustics, Speech and Signal Processing, pp. 2932–2936, 2020.
[147] D. Anghelone, S. Lannes, V. Strizhkova, P. Faure, C. Chen, and A. Dantcheva, "TFLD: Thermal face and landmark detection for unconstrained cross-spectral face recognition," in Intl. Joint Conf. on Biometrics, 2022.
[148] E. Richardson, M. Sela, R. Or-El, and R. Kimmel, "Learning detailed face reconstruction from a single image," in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, pp. 1259–1268, 2017.
[149] B. Dieckmann, J. Merkle, and C. Rathgeb, "Fingerprint pre-alignment based on deep learning," in Intl. Conf. of the Biometrics Special Interest Group, pp. 1–6, 2019.
[150] S. Zhang, R. He, Z. Sun, and T. Tan, "Multi-task convnet for blind face inpainting with application to face verification," in Intl. Conf. on Biometrics, pp. 1–8, 2016.
[151] I. Joshi, A. Anand, M. Vatsa, R. Singh, S. D. Roy, and P. Kalra, "Latent fingerprint enhancement using generative adversarial networks," in IEEE Winter Conf. on Applications of Computer Vision, pp. 895–903, 2019.
[152] I. Joshi, A. Utkarsh, R. Kothari, V. K. Kurmi, A. Dantcheva, S. D. Roy, and P. K. Kalra, "Sensor-invariant fingerprint roi segmentation using recurrent adversarial learning," in Intl. Joint Conf. on Neural Networks, pp. 1–8, 2021.
[153] E. Bondi, R. Jain, P. Aggrawal, S. Anand, R. Hannaford, A. Kapoor, J. Piavis, S. Shah, L. Joppa, B. Dilkina, et al., "Birdsai: A dataset for detection and tracking in aerial thermal infrared videos," in Proc. of the IEEE/CVF Winter Conf. on Applications of Computer Vision, pp. 1747–1756, 2020.
[154] Y. Zhong, Y. Pei, P. Li, Y. Guo, G. Ma, M. Liu, W. Bai, W. Wu, and H. Zha, "Depth-based 3d face reconstruction and pose estimation using shape-preserving domain adaptation," IEEE Trans. on Biometrics, Behavior, and Identity Science, vol. 3, no. 1, pp. 6–15, 2020.
[155] F. Kuhnke and J. Ostermann, "Deep head pose estimation using synthetic images and partial adversarial domain adaption for continuous label spaces," in Proc. of the IEEE/CVF Intl. Conf. on Computer Vision, pp. 10164–10173, 2019.
[156] V. A. Sindagi and V. M. Patel, "Ha-ccn: Hierarchical attention-based crowd counting network," IEEE Trans. on Image Processing, vol. 29, pp. 323–335, 2019.
[157] E. Mequanint, S. Zhang, B. Forutanpour, Y. Qi, and N. Bi, "Weakly-supervised degree of eye-closeness estimation," in IEEE/CVF Intl. Conf. on Computer Vision Workshop, pp. 4416–4424, 2019.
[158] R. Zhang, C. Mu, M. Xu, L. Xu, and X. Xu, "Facial component-landmark detection with weakly-supervised lr-cnn," IEEE Access, vol. 7, pp. 10263–10277, 2019.
[159] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, and N. Kumar, "Localizing parts of faces using a consensus of exemplars," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 12, pp. 2930–2940, 2013.
[160] J. J. Engelsma, K. Cao, and A. K. Jain, "Learning a fixed-length fingerprint representation," IEEE Trans. on pattern analysis and machine intelligence, vol. 43, no. 6, pp. 1981–1997, 2019.
[161] G. P. Fiumara, P. A. Flanagan, J. D. Grantham, K. Ko, K. Marshall, M. Schwarz, E. Tabassi, B. Woodgate, C. Boehnen, et al., "Nist special database 302: Nail to nail fingerprint challenge," 2019.
[162] J. Deng, "A large-scale hierarchical image database," Proc. of IEEE Computer Vision and Pattern Recognition, 2009, 2009.
[163] J. Svoboda, P. Astolfi, D. Boscaini, J. Masci, and M. Bronstein, "Clustered dynamic graph cnn for biometric 3d hand shape recognition," in 2020 IEEE Intl. Joint Conf. on Biometrics (IJCB), pp. 1–9, IEEE, 2020.
[164] R. Ranjan, S. De Mello, and J. Kautz, "Light-weight head pose invariant gaze tracking," in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition Workshops, pp. 2156–2164, 2018.
[165] A. Kortylewski, A. Schneider, T. Gerig, B. Egger, A. Morel-Forster, and T. Vetter, "Training deep face recognition systems with synthetic data," arXiv preprint arXiv:1802.05891, 2018.
[166] M. Ryoo, K. Kim, and H. Yang, "Extreme low resolution activity recognition with multi-siamese embedding learning," in Proc. of the AAAI Conf. on Artificial Intelligence, vol. 32, 2018.
[167] P. C. Neto, F. Boutros, J. R. Pinto, N. Darner, A. F. Sequeira, and J. S. Cardoso, "Focusface: Multi-task contrastive learning for masked face recognition," in 16th IEEE Intl. Conf. on Automatic Face and Gesture Recognition, pp. 01–08, 2021.
[168] J. Huh, H. S. Heo, J. Kang, S. Watanabe, and J. S. Chung, "Augmentation adversarial training for self-supervised speaker recognition," arXiv preprint arXiv:2007.12085, 2020.
[169] Y.-L. Lee, M.-Y. Tseng, Y.-C. Luo, D.-R. Yu, and W.-C. Chiu, "Learning face recognition unsupervisedly by disentanglement and self-augmentation," in IEEE Intl. Conf. on Robotics and Automation, pp. 3018–3024, 2020.
[170] H. Chen, Y. Wang, B. Lagadec, A. Dantcheva, and F. Bremond, "Learning invariance from generated variance for unsupervised person re-identification," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022.
[171] J. R. Pinto and J. S. Cardoso, "Self-learning with stochastic triplet loss," in Intl. Joint Conf. on Neural Networks, pp. 1–8, 2020.
[172] H. Zhou, J. Liu, Z. Liu, Y. Liu, and X. Wang, "Rotate-and-render: Unsupervised photorealistic face rotation from single-view images," in Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, pp. 5911–5920, 2020.
[173] T. Zhao, X. Xu, M. Xu, H. Ding, Y. Xiong, and W. Xia, "Learning self-consistency for deepfake detection," in Proc. of the IEEE/CVF Intl. Conf. on Computer Vision, pp. 15023–15033, 2021.
[174] Y. Li, Y. Gao, B. Chen, Z. Zhang, G. Lu, and D. Zhang, "Self-supervised exclusive-inclusive interactive learning for multi-label facial expression recognition in the wild," IEEE Trans. on Circuits and Systems for Video Technology, 2021.
[175] Y.-J. Ju, G.-H. Lee, J.-H. Hong, and S.-W. Lee, "Complete face recovery gan: Unsupervised joint face rotation and de-occlusion from a single-view image," in Proc. of the IEEE/CVF Winter Conf. on Applications of Computer Vision, pp. 3711–3721, 2022.
[176] A. Zhao, J. Dong, and H. Zhou, "Self-supervised learning from multi-sensor data for sleep recognition," IEEE Access, vol. 8, pp. 93907–93921, 2020.
[177] S. Ge, S. Zhao, X. Gao, and J. Li, "Fewer-shots and lower-resolutions: Towards ultrafast face recognition in the wild," in Proc. of the 27th ACM Intl. Conf. on Multimedia, pp. 229–237, 2019.
[178] A. Bansal, A. Nanduri, C. D. Castillo, R. Ranjan, and R. Chellappa, "Umdfaces: An annotated face dataset for training deep networks," in IEEE Intl. Joint Conf. on Biometrics, pp. 464–473, 2017.
[179] Y.-T. Cao, J. Wang, and D. Tao, "Symbiotic adversarial learning for attribute-based person search," in European Conf. on Computer Vision, pp. 230–247, Springer, 2020.
[180] P. Korshunov and S. Marcel, "Improving generalization of deepfake detection with data farming and few-shot learning," IEEE Trans. on Biometrics, Behavior, and Identity Science, 2022.
[181] J. Solano, L. Tengana, A. Castelblanco, E. Rivera, C. Lopez, and M. Ochoa, "A few-shot practical behavioral biometrics model for login authentication in web applications," in NDSS Workshop on Measurements, Attacks, and Defenses for the Web, 2020.
[182] R. Tolosana, P. Delgado-Santos, A. Perez-Uribe, R. Vera-Rodriguez, J. Fierrez, and A. Morales, "Deepwritesyn: On-line handwriting synthesis via deep short-term representations," in Proc. AAAI Conf. on Artificial Intelligence, 2021.
[183] M. Georgopoulos, J. Oldfield, M. A. Nicolaou, Y. Panagakis, and M. Pantic, "Mitigating demographic bias in facial datasets with style-based multi-attribute transfer," Intl. Journal of Computer Vision, vol. 129, no. 7, pp. 2288–2307, 2021.
[184] P. Drozdowski, C. Rathgeb, A. Dantcheva, N. Damer, and C. Busch, "Demographic bias in biometrics: A survey on an emerging challenge," IEEE Trans. on Technology and Society, vol. 1, no. 2, pp. 89–103, 2020.
[185] K. Ricanek and T. Tesafaye, "Morph: A longitudinal image database of normal adult age-progression," in 7th Intl. Conf. on Automatic Face and Gesture Recognition, pp. 341–345, 2006.
[186] A. Pumarola, A. Agudo, A. M. Martinez, A. Sanfeliu, and F. Moreno-Noguer, "Ganimation: One-shot anatomically consistent facial animation," Intl. Journal of Computer Vision, vol. 128, no. 3, pp. 698–713, 2020.
[187] X. Jia, X. Yang, Y. Zang, N. Zhang, and J. Tian, "A cross-device matching fingerprint database from multi-type sensors," in Proc. of the 21st Intl. Conf. on Pattern Recognition, pp. 3001–3004, 2012.
[188] N. Kohli, D. Yadav, M. Vatsa, R. Singh, and A. Noore, "Synthetic iris presentation attack using idcgan," in IEEE Intl. Joint Conf. on Biometrics, pp. 674–680, 2017.
[189] F. Boutros, N. Damer, K. Raja, R. Ramachandra, F. Kirchbuchner, and A. Kuijper, "Iris and periocular biometrics for head mounted

displays: Segmentation, recognition, and synthetic data generation," Image and Vision Computing, vol. 104, p. 104007, 2020.

[190] H. Zhang, S. Venkatesh, R. Ramachandra, K. Raja, N. Damer, and C. Busch, "Mipgan—generating strong and high quality morphing attacks using identity prior driven gan," IEEE Trans. on Biometrics, Behavior, and Identity Science, vol. 3, no. 3, pp. 365–383, 2021.

[191] R. Bouzaglo and Y. Keller, "Synthesis and reconstruction of fingerprints using generative adversarial networks," arXiv preprint arXiv:2201.06164, 2022.

[192] Y. Wang, P. Bilinski, F. Bremond, and A. Dantcheva, "G3an: Disentangling appearance and motion for video generation," in Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition, pp. 5264–5273, 2020.

[193] J. Thies, M. Zollhöfer, and M. Nießner, "Deferred neural rendering: Image synthesis using neural textures," ACM Trans. on Graphics, vol. 38, no. 4, pp. 1–12, 2019.

[194] A. Siarohin, S. Lathuilière, S. Tulyakov, E. Ricci, and N. Sebe, "First order motion model for image animation," Adv. in Neural Information Processing Systems, vol. 32, 2019.

[195] Y. Wang, D. Yang, F. Bremond, and A. Dantcheva, "Latent image animator: Learning to animate images via latent space navigation," in Proc. of the Intl. Conf. on Learning Representations, 2022.

[196] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "Deepfakes and beyond: A survey of face manipulation and fake detection," Information Fusion, vol. 64, pp. 131–148, 2020.

[197] C. Rathgeb, Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks. Springer Nature, 2021.

[198] H. Xu, Y. Ma, H.-C. Liu, D. Deb, H. Liu, J.-L. Tang, and A. K. Jain, "Adversarial attacks and defenses in images, graphs and text: A review," Intl. Journal of Automation and Computing, vol. 17, no. 2, pp. 151–178, 2020.

[199] 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 Trans. on Information Forensics and Security, vol. 16, pp. 4336–4351, 2020.

[200] S. Venkatesh, R. Ramachandra, K. Raja, and C. Busch, "Face morphing attack generation & detection: A comprehensive survey," IEEE Trans. on Technology and Society, 2021.

[201] U. Scherhag, C. Rathgeb, J. Merkle, R. Breithaupt, and C. Busch, "Face recognition systems under morphing attacks: A survey," IEEE Access, vol. 7, pp. 23012–23026, 2019.

[202] K. Wang, R. Zhao, and Q. Ji, "A hierarchical generative model for eye image synthesis and eye gaze estimation," in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, pp. 440–448, 2018.

[203] S. Yadav, C. Chen, and A. Ross, "Relativistic discriminator: A one-class classifier for generalized iris presentation attack detection," in Proc. of the IEEE/CVF Winter Conf. on Applications of Computer Vision, pp. 2635–2644, 2020.

[204] L. An, Z. Qin, X. Chen, and S. Yang, "Multi-level common space learning for person re-identification," IEEE Trans. on Circuits and Systems for Video Technology, vol. 28, no. 8, pp. 1777–1787, 2017.

[205] Y. Gao, N. Xiong, W. Yu, and H. J. Lee, "Learning identity-aware face features across poses based on deep siamese networks," IEEE Access, vol. 7, pp. 105789–105799, 2019.

[206] C. Fu, X. Wu, Y. Hu, H. Huang, and R. He, "Dvg-face: Dual variational generation for heterogeneous face recognition," IEEE Trans. on Pattern Analysis and Machine Intelligence, 2021.

[207] M. Grimmer, H. Zhang, R. Ramachandra, K. Raja, and C. Busch, "Generation of non-deterministic synthetic face datasets guided by identity priors," arXiv preprint arXiv:2112.03632, 2021.

[208] ISO/IEC JTC1 SC37 Biometrics, ISO/IEC WD 19795-10: Information Technology – Biometric Performance Testing and Reporting – Part 10: Quantifying biometric system performance variation across demographic groups. Intl. Organization for Standardization.

[209] C. Gottschlich and S. Huckemann, "Separating the real from the synthetic: minutiae histograms as fingerprints of fingerprints," IET Biometrics, vol. 3, no. 4, pp. 291–301.

Indu Joshi has obtained Ph.D. from IIT Delhi. She is a recipient of the prestigious Raman-Charpak Fellowship. She has presented her research at several reputed international forums such as BRICS Young Scientist Conclave and has also attended the prestigious Heidelberg Laureate Forum. She is a DAAD Postdoc-Net-AI fellow and her research interests include biometrics, bioinformatics, medical image processing, and uncertainty estimation.

Marcel Grimmer is a PhD candidate of the Norwegian Biometrics Laboratory (NBL) at the Norwegian University of Science and Technology (NTNU). His active research is dedicated to the generation of synthetic images in the context of face recognition, with a particular focus on face image quality assessment.

Christian Rathgeb is a Professor with the Faculty of Computer Science, Hochschule Darmstadt (HDA), Germany. He is also a Principal Investigator with the National Research Center for Applied Cybersecurity (ATHENE). His research interests include pattern recognition, iris and face recognition, the security aspects of biometric systems, secure process design, and privacy enhancing technologies for biometric systems.

Christoph Busch is member of NTNU-Gjøvik, Norway and holds a joint appointment with Hochschule Darmstadt, Germany. Further he lectures at Denmark's DTU since 2007. He was initiator and participated in multiple projects on biometrics (e.g. 3D-Face, FIDELITY and iMARS). He is also PI in ATHENE. Christoph is co-founder of the European Association for Biometrics (EAB). He co-authored more than 600 papers. Furthermore, he is convenor of WG3 in ISO/IEC JTC1 SC37.

François Brémond is a Research Director at INRIA Sophia Antipolis, where he has headed the Inria Stars team since 2013. He obtained his PhD degree in 1997 in video understanding from Nice University. He has supervised 20 PhD theses. He is a co-fonder of Keeneo, Ekinnox and Neosensys, three companies in intelligent video monitoring and business intelligence.

Antitza Dantcheva is a Research Scientist (CRCN) with the STARS team of INRIA Sophia Antipolis, France. Her research is in computer vision and specifically in designing algorithms that seek to learn suitable representations of the human face in interpretation and generation. She received her Ph.D. degree from Telecom ParisTech in signal and image processing 2011 and has co-authored over 70 international publications and 3 patents.