Artificial Humans: an Overview of Photorealistic Synthetic Datasets and Possible Applications

Desislava Nikolova[1], Ivaylo Vladimirov[2] and Zornitsa Terneva[3]

Abstract – **In this scientific paper, an overview of different photorealistic synthetic human datasets is presented. The creation of more and more artificial data is leading to rapid progress in various fields. Synthetic faces and whole bodies are needed during the processes of training and exploitation of applications in the field. The state of the art synthetic human representations are listed, including their applications.**

Keywords – **Artificial Humans, Synthetic Dataset, Photorealistic, Human Face, Human Body, Overview;**

I. Introduction

In recent years the focus of the entertainment industry is heavily concentrated on creating synthetic pictures and videos in order to close the gap between the computer generated and real-life events. Synthetic data is created algorithmically and used as a stand-in for real-world data test datasets, mathematical model validation, and machine learning model training. With the rise of the topic of metaverse, it promises to provide a larger crossover of our digital and physical lives in wealth, sociability, work, commerce, and entertainment, whether in virtual reality (VR), augmented reality (AR), or just on a screen. This vision cannot exist without avatars and the two types are Face and Full body avatars. Both of them require realistic face and full body representations [1, 2].

In this paper some of the best synthetic human datasets are listed. The next section looks deeper into face centred datasets and the other one into the full human body datasets. In the conclusion some of their possible implementations are discussed.

II. Datasets with faces

Machine learning developments, notably Generative Adversarial Networks (GANs) [1], have enabled the creation of extraordinarily photo-realistic faces. One of the topics most discussed in the field of synthetic data is the computer generated human faces. In the recent years the synthetic faces are getting more and more unrecognisable from the real ones. Sometimes average people are unable to dependably distinguish the genuine from the synthetic even after being taught about specific synthesis artifacts. In this section some of the best synthetic datasets of human faces will be listed [2].

[1]Desislava Nikolova is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. E-mail: dnikolova@tu-sofia.bg

[2]Ivaylo Vladimirov is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. E-mail: ivladimirov@tu-sofia.bg

[3]Zornitza Terneva is with the Faculty of Telecommunications at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria. E-mail: zterneva@tu-sofia.bg

a) b)

Fig. 1 Frames of deepfake datasets with faces
a) from FaceForensics++ [2]
b) from Celeb-DF[4]

Face Forensics' initial dataset "**FaceForensics**++: Learning to Detect Manipulated Facial Images" is limited to the face swapping methods (as seen on Fig.1a). Due to the rapid progress of face swapping and forgery detection algorithms, face forensics datasets have a short shelf life. The domain-adversarial quality control method might be a feasible option, as it allows for automated dataset updates using more advanced deepfake methods. The model struggles in circumstances with moderate light changes and continuous motions. In such cases, the revised faces within the same tracklet usually show great consistency and less artifacts. Long-term features become less informative as a result, which can easily lead to poor performance [3].

As a result, the focus of "forgery" research is mostly on face swapping. Given the broader concerns about how imagery is manipulated to influence the political sphere, "forgery" should be investigated further under controlled camera properties, such as controlling body motions, modifying facial expressions, synthesizing realistic talking head recordings, or exchanging faces.

Another dataset created by Face Forensics "**Face Forensics in the Wild**" demonstrates that trained forgery detectors can detect face picture altering methods, even when the results are aesthetically amazing. They provide a new dataset of edited face films that outperforms publicly available forensic datasets in terms of training detectors using domain-specific data [4].

The influence of compression on the detectability of cutting-edge manipulation methods is the subject of this research, which also offers a standard baseline for future research. The image data, training models, and benchmarks are all open source and have been used by other researchers before. Approaches for identifying fakes with little to no training data must be developed as new modification techniques arise on a regular basis. This forensic transfer learning work makes use of the database, in which knowledge of one source manipulation area is transferred to a separate target domain. The dataset and benchmark might be used as a starting point for further research.

978-1-6654-8500-5/22/$31.00 ©2022 European Union

The authors of **CelebDF** propose a new large-scale dataset that will be used to create and test DeepFake detection techniques. The Celeb-DF dataset bridges the gap between DeepFake datasets and the real DeepFake videos that circulate online in terms of visual quality. They provide a complete performance evaluation of current DeepFake detection algorithms using the Celeb-DF dataset, and show that there is still significant space for improvement(as seen on Fig.1b).

The most important challenge for future study is to expand the Celeb-DF dataset and enhance the visual quality of the synthesized films. This requires increasing the present synthesis algorithm's operating efficiency and model structure. In addition, while forgers can increase visual quality in general, they can also use anti-forensic techniques to conceal DeepFake synthesis traces, which are used by detection algorithms. Anti-forensic methods will be included in Celeb-DF in anticipation of such countermeasures becoming available to forgers [5].

Another research, "**SynFace**: Face Identification with Synthetic Data," did a thorough empirical examination and uncovered the following ideas on how to use synthetic face photographs for face recognition efficiently:

1) The recommended identity mixup may be utilized to boost synthetic data's intra-class variability, which improves performance consistently.

2) The depth and width of the training synthetic dataset have an impact on performance, with saturation first appearing on the depth dimension.

3) Position, lighting, and expression all have different effects on performance; for example, adjusting pose and illumination greatly improves performance, but the expression variety in produced face pictures is limited.

4) By using only a small percentage of real-world face pictures, the recommended domain mixup can significantly improve SynFace performance [6].

Umur Aybars Ciftci, Ilke Demir and Lijun Yin present **FakeCatcher**, a fake portrait video detector based on biological signals, in their work "FakeCatcher: Detection of Synthetic Portrait Videos Using Biological Signals." They show that GAN-erated content does not successfully sustain the spatial coherence and temporal consistency of such signals. Based on physiological changes, they build a powerful synthetic video classifier. They also encapsulate those signals in distinct PPG maps, allowing for the creation of a CNN-based classifier that improves accuracy while being independent of any generative model. They test the technique for pairwise separation and authenticity classification of segments and videos in Face Forensics, attaining 99.39 percent pairwise separation accuracy, 96 percent constrained video classification accuracy, and 91.07 percent in the wild video classification accuracy [7].

These results show that FakeCatcher is incredibly accurate in detecting false content, regardless of the source, content, resolution, or video quality, according to these findings.

"Where Do Deep Fakes Look?" is another intriguing subject in the world of synthetic data. "**Gaze Tracking for Synthetic Face Detection**" is answering this question [8].

Eye tracking research can employ synthetic gazes with improved photorealism in virtual/augmented reality applications, avatars, data augmentation, transfer learning, and controlled learning and testing settings. Another objective is to include 3D gaze categorization into false detectors in the future to improve their accuracy. Saccades, fixations, and other gaze movements have been found to carry much more biological information.

The study is the first to examine deep false gazes in depth, as well as the first to offer an approach for constructing a detector based only on holistic eye and gaze attributes (instead of cherry-picking a few). They put the approach to the test on four different datasets, compared it to biological and deep detectors, and conducted ablation experiments. They claim that existing fake detectors include visual, geometric, temporal, metric, and spectral properties because real eyes show natural signals that fake ones have yet to mimic consistently. The signatures are based on authenticity signals rather than noise generation, which is advantageous.

Fig. 2 Creating synthetic dataset with human faces [9]

One state of the art paper is Microsoft's "**Fake it till you make it**: face analysis in the wild using synthetic data alone" where they paid close attention to every detail. They demonstrate that synthetic data alone can be used to conduct face-related computer vision in the wild (as seen on Fig.2). The domain gap between actual and synthetic data has remained a challenge throughout time, particularly in the case of human faces. Researchers have attempted to bridge the gap through data mixing, domain adaption, and domain-adversarial training, but they show that it is possible to synthesis data with minimum domain gaps, allowing models trained on synthetic data to generalize to real-world datasets [9].

It is discussed how to produce training pictures with unparalleled realism and diversity by combining a procedurally generated parametric 3D face model with a vast collection of hand-crafted elements. They use synthetic data to train machine learning systems for tasks like landmark localization and face parsing, demonstrating that synthetic data may both equal real data in accuracy and open up new ways when manual labeling is impractical.

The computer-generated faces are lifelike, diversified, and expressive. The identity is randomised, a random expression is picked, a random texture is applied, and random hair and clothing are attached, starting with the template face. Finally, Cycles, a physically-based path tracing renderer, is used to depict the face in a random environment. Identities were taken from a generative model and sampled. They used a variety of high-quality scan data to train the model.

978-1-6654-8500-5/22/$31.00 ©2022 European Union

III. DATASETS WITH BODIES

Contrary to the synthetic faces, the human bodies are not as photo-realistic yet. Synthetic datasets with whole human bodies are widely discussed topic in the field and are getting more and more important mainly because of the many options it is providing. One of the implementations of these datasets is the action recognition [10].

Fig. 3 Examples of SURREAL dataset with human bodies [11]

HAR (Human Action Recognition) is an important project that has had a lot of scholarly attention in recent decades, and it's been done with a range of data modalities, each with its unique set of characteristics. "**Human Action Recognition** from Various Data Modalities: A Review" examines HAR approaches that use a variety of data modalities. Multi-modality recognition techniques like fusion and co-learning are also researched. Authors further stated, as well as other potential study areas, were also investigated [11].

In one of the first papers on the synthetic data for human action recognition is "Learning from Synthetic Humans". **SURREAL** (Synthetic hUmans foR REAL tasks) is a big dataset that contains synthetically created yet realistic pictures of humans rendered using 3D sequences of human motion capture data (Fig.3). With ground truth posture, depth maps, and segmentation masks, over 6 million frames are created. This study and the SURREAL dataset have opened up a lot of doors for more advanced human analysis utilizing low-cost, large-scale synthetic data [12].

"**HuMoR**: 3D Human Motion Model for Robust Pose Estimation" is a learnt generative model of 3D human motion that can consistently recover position and shape during test-time optimization from 3D, RGB, and RGB-D observations. For both generative and downstream optimization tasks, the approach's essential components have been proved to be generalizable to novel motions and body types. HuMoR surpasses strong learning and optimization-based baselines in both predicting plausible motion and delivering consistent ground plane and contact outputs under high occlusions [13].

HuMoR's static camera and ground plane assumptions are sufficient for indoor scenarios, but in-the-wild operations demand solutions for coping with dynamic cameras and complex terrain. This model should be upgraded to record scene-person interactions for better scene perception.

"**Moulding Humans**: Non-parametric 3D Human Shape Estimation from Single Images" proposed employing a double 2.5D depth map representation to encode a person's 3D shape in a non-parametric way: a "visible" depth map depicts the areas of the surface that can be seen clearly in the image, while a "hidden" depth map depicts the occluded 3D surface. They devised a system that accepts a single image as input and simultaneously creates estimates for both depth maps, yielding a point cloud of the whole 3D surface when combined. The approach can recover detailed surfaces while keeping the output size manageable. The learning process becomes more efficient as a result of this [14].

Another paper on the topic is "**Creating a Large-scale Synthetic Dataset for Human Activity Recognition**". The authors show how a completely constructed dataset may be utilized to train a high-performing classifier in this paper. Furthermore, synthetic data may be used in conjunction with real data to improve performance, which is particularly useful in optical flow processing. These findings provide proof of concept, implying that synthetic data may be utilized to detect human activity. Even if the movies aren't completely realistic, they're able to provide a rich dataset for training by being able to create an infinite number of variations of people in their environment. More realistically created movies may allow for improvements in both the RGB and optical flow streams [15].

The huge Motion and Video dataset "**MoVi**: A Large Multipurpose Motion and Video Dataset" is available online. Motion recordings (optical motion capture, video, and IMU) of 90 male and female actors executing a set of 20 everyday movements and sports motions, as well as one additional self-chosen motion, are included in the dataset. The dataset's distinct sequences comprise synchronized recordings of the three hardware devices. In addition, as part of the AMASS collection, full-body motion capture recordings are accessible as realistic 3D human meshes represented by a rigged body model. Because of its multi-modality, the dataset may be used to solve a variety of problems, including human posture estimation and tracking, body form estimation, human motion prediction and synthesis, action identification, and gait analysis [16].

"**VI-Net**—View-Invariant Quality of Human Movement Assessment" presents QMAR, a multi-view, non-skeleton, non-mocap rehabilitation movement dataset, to evaluate the recommended method's performance, which might be employed in a community comparison research. They demonstrate that the recommended approach is suitable for cross-subject, cross-view, and single-view movement analysis by achieving excellent average rank correlation. OpenPose fails to give sufficiently consistent heat maps when long-term occlusions occur and VI-performance Nets decreases. However, many techniques struggle with long-term occlusions; therefore such failure is to be expected [17].

978-1-6654-8500-5/22/$31.00 ©2022 European Union

Fig. 4 Creating synthetic dataset with human poses [18]

The authors of "**Synthetic Humans for Action Recognition from Unseen Viewpoints**" revealed how to use a simple approach to automatically augment action recognition datasets using synthetic videos (Fig.4). Several characteristics in the synthetic data, such as views and movements, were examined for their importance. The study looks at ways to broaden motions within a particular action category. Action identification improves significantly as a result of unobserved perspectives and one-shot training. However, this approach is limited by the performance of 3D posture estimation, which might fail in crowded contexts. Two potential future directions are action-conditioned generative models for motion sequences and modeling of contextual information for action detection [18].

IV. Conclusion

Enhancing creativity and imagination, traveling the world without moving, overcoming obstacles e.g. disabilities that prevent us from doing something in real life and creating completely new job opportunities are just a few of the many opportunities the metaverse will be able to provide in the future. And one of the most important steps in creating it is human representation and creating realistic human faces and bodies. They can also be used in human action and face recognition and have the potential to be very realistic.

It has to be noted that there are synthetic faces and bodies which have been used to create fake social media profiles and fake news. However most of the created synthetic data is focused mainly on creating avatars and image recognition for better teaching algorithms. The estimate of human postures and behaviour for various activities in Virtual Reality or Augmented Reality might have a wide range of applications. Human fall monitoring is particularly useful for the elderly and for non-traditional VR/AR activities.

In the next steps of this research we are planning to answer the question whether the already existing synthetic datasets can be used in training algorithms for human recognition.

Acknowledgement

This scientific research is part of a contract №222ПД0012-07 for a research project to help doctoral students: "Analysis of algorithms for recognizing activity using synthetic training data" of the Technical University of Sofia, Bulgaria Research Sector.as a PhD student in the Faculty of Telecommunications, Technical University of Sofia, Sofia, Bulgaria.

References

[1] S. Mystakidis, "Metaverse." Encyclopedia 2.1 (2022);
[2] M. Mozumder, "Overview: Technology Roadmap of the Future Trend of Metaverse based on IoT, Blockchain, AI Technique, and Medical Domain Metaverse Activity." 2022 24th International Conference on Advanced Communication Technology (ICACT). IEEE, 2022;
[3] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, Sh. Ozair, A. Courville, Y. Bengio, "Generative Adversarial Networks", Advances in neural information processing systems 27, 2014;
[4] S. Nightingale, S. Agarwal, E. Härkönen, J. Lehtinen, H. Farid, "Synthetic faces: how perceptually convincing are they?", Journal of Vision September, 2021;
[5] A. Rössler, D. Cozzolino, L. Verdoliva, Ch. Riess, J. Thies, M. Nießner, "FaceForensics++: Learning to Detect Manipulated Facial Images", Computer Vision and Pattern Recognition, 2019;
[6] T. Zhou, W. Wang, Zh. Liang, J. Shen, "Face Forensics in the Wild", Computer Vision and Pattern Recognition, 2021;
[7] Y. Li, X. Yang, P. Sun, H. Qi, S. Lyu, "Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics", 2020;
[8] H. Qiu, B. Yu, D. Gong, Z. Li, W. Liu, D. Tao, "SynFace: Face Recognition with Synthetic Data", 2021;
[9] U. A. Ciftci, I. Demir, "FakeCatcher: Detection of Synthetic Portrait Videos using Biological Signals", 2020;
[10] U. A. Ciftci, I. Demir, "Where Do Deep Fakes Look? Synthetic Face Detection via Gaze Tracking", 2021;
[11] E. Wood, T. Baltrušaitis, Ch. Hewitt, S. Dziadzio, M. Johnson, V. Estellers, Th. J. Cashman, J. Shotton, "Fake It Till You Make It: Face analysis in the wild using synthetic data alone", 2021;
[12] D. Nikolova, I. Vladimirov, Z. Terneva, "Human Action Recognition for Pose-based Attention: Methods on the Framework of Image Processing and Deep Learning", ICEST 2021;
[13] Z. Sun, J. Liu, Q. Ke, H. Rahmani, M. Bennamoun, G. Wang, "Human Action Recognition from Various Data Modalities: A Review", 2021;
[14] G. Varol, J. Romero, X. Martin, N. Mahmood, M. J. Black, I. Laptev, C. Schmid, "Learning from Synthetic Humans", 2017;
[15] D. Rempe, T. Birdal, A. Hertzmann, J. Yang, S. Sridhar, L. J. Guibas, "HuMoR: 3D Human Motion Model for Robust Pose Estimation", 2021;
[16] V. Gabeur, J. Franco, X. Martin, C. Schmid, G. Rogez, "Moulding Humans: Non-parametric 3D Human Shape Estimation from Single Images", ICCV 2019;
[17] O. Matthews, K. Ryu, T. Srivastava, "Creating a Large-scale Synthetic Dataset for Human Activity Recognition", 2020;
[18] S. Ghorbani, K. Mahdaviani, A. Thaler, K. Kording, DJ. Cook, G. Blohm, "MoVi: A large multi-purpose human motion and video dataset", PLoS ONE 16(6): e0253157, 2021;
[19] F. Sardari, A. Paiement, S. Hannuna, M. Mirmehdi, "VI-Net—View-Invariant Quality of Human Movement Assessment", Sensors 2020;
[20] G. Varol, I. Laptev, C. Schmid, A. Zisserman, "Synthetic Humans for Action Recognition from Unseen Viewpoints", International Journal of Computer Vision 2021;