μPose: Synthetic Dataset for Human Pose Estimation in Microgravity Environments

Luís Fernando de Souza Cardoso
Virtual Words and Digital Games Group
Ilmenau University of Technology
Ilmenau, Germany
luis.cardoso@tu-ilmenau.de

Tobias Schwandt
Virtual Words and Digital Games Group
Ilmenau University of Technology
Ilmenau, Germany
tobias.schwandt@tu-ilmenau.de

Wolfgang Broll
Virtual Words and Digital Games Group
Ilmenau University of Technology
Ilmenau, Germany
wolfgang.broll@tu-ilmenau.de

***Abstract*—The field of space mission has been growing rapidly in recent years, with a multitude of projects and initiatives being undertaken by various organizations around the world. One critical aspect of such endeavors is the ability to simulate and visualize the complex environments and scenarios that space missions may encounter. This paper presents the challenges of human pose estimation in microgravity environments. Existent frameworks face challenges in estimating human poses in non-ordinary positions, and these difficulties are further compounded by the shortage of pertinent data. To overcome these challenges, we propose a synthetic dataset featuring virtual humans in microgravity, serving as a cost-effective alternative for training pose estimation models. Our methodology employs transfer learning with an HRNET backbone and associative embedding, facilitating the retraining of models for 2D human pose estimation on established datasets like COCO and MMPose. Evaluation against publicly available images of astronauts in the International Space Station (ISS) demonstrates our model's robust generalization, with the best result showing a 20% improvement compared to models trained with COCO dataset. This research demonstrates the viability of synthetic datasets for training models in unique environments. The proposed methodology showcases improved model performance in microgravity scenarios, marking a stride in advancing human pose estimation for space exploration.**

***Index Terms*—synthetic dataset, virtual dataset, pose estimation, transfer learning**

I. Introduction

With upcoming human space missions such as the Lunar Gateway and Artemis missions where crew members will be long-term exposed to a microgravity environment, it becomes crucial to increase our understanding of the impact of this environment on the crew members [1], [2].

One possible way to understand human behavior in space is by analyzing their postures, actions, and movements. In this scenario, human pose estimation is a potential tool to support individuals in space by improving crew procedures and enhancing training for future astronauts.

However, one of the primary challenges in estimating human pose in space is that most state-of-the-art frameworks are not trained to understand humans in non-ordinary poses, such as scenes where people are in different orientations [3].

Training new models for such an environment also poses a challenge. Firstly, there is a limited amount of videos and images publicly available to create annotations for the human body. Additionally, it depends on specific space missions to create the dataset, and astronauts would need to wear sensors to collect the ground truth. Therefore, creating such a dataset is an expensive endeavor.

In this context, this paper proposes a synthetic dataset that creates virtual humans moving in a microgravity environment. The proposed dataset serves as a foundation for transfer learning, enabling the retraining of a model for 2D human pose estimation on the COCO dataset and MMPose. The chosen methodology employs an HRNET backbone and associative embedding in a bottom-up approach.

The outcomes yielded by the retrained model are pitted against publicly available imagery of astronauts in the ISS, offering a robust basis for assessing the model's generalization and performance in a real-world space environment.

The main contributions of the paper are:

- a synthetic dataset of virtual body movement in microgravity environment.
- transfer learning approach to use synthetic data generated by virtual dataset to estimate human poses.
- comparison with public available video footage of astronauts inside the ISS.
- refined model for pose estimation in space environment.

II. Related Works

A. 2D Pose Estimation

The field of 2D pose estimation from RGB cameras has seen extensive researched, with several notable contributions [4]. A deep learning-based approach for 2D pose estimation has emerged as a powerful technique in recent years. Deep learning models, particularly convolutional neural networks (CNNs), have demonstrated remarkable capabilities in understanding complex visual patterns and extracting meaningful features from images or videos.

These models can be trained on large-scale datasets of annotated human poses, allowing them to learn the intricate relationships between body joints and their corresponding visual representations.

Two different approaches can be done in order to estimate human pose, single or multi person pose estimation. While single person pose estimation detects a specific person's pose in an image and when an image contains multiple people, then

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00282

it is cropped until leave only one person [5]. Multi person pose estimation allows the simultaneous detection and tracking of multiple individuals in an image or video sequence. This is particularly beneficial in crowded scenes or scenarios where there are multiple people interacting or performing activities together [4]–[6].

Two different approaches can be done to obtain multi person pose estimation: top-down or bottom-up.

The top-down approach involves detecting and segmenting individuals in the image using object detection techniques such as Faster R-CNN or Mask R-CNN [7]–[10]. Then, for each detected person, a pose estimator is applied individually to predict the joint locations [11], [12].

the bottom-up approach aims to identify all the joints present in the image and then group them into complete poses. In this method, keypoint detection techniques are used, such as algorithms based on part affinity fields or graphical models. Joints are independently detected and then associated to form complete poses [13]–[16]. The bottom-up approach is more suitable for scenes with multiple people and occlusion as it deals with joint detection and grouping in a joint step [6], [12].

B. Datasets for deep learning-based approaches

Datasets are crucial for human pose estimation using deep learning-based approaches. They provide annotated ground truth data for training and evaluating models. These datasets contain images or videos with corresponding annotations that indicate the locations of human joints.

One of the most used datasets for human pose estimation is COCO (Common Objects in Context) [17]. This dataset includes a large collection of images with diverse scenes and multiple people. COCO provides detailed annotations for keypoint locations of 17 body joints, making it suitable for both single-person and multi-person pose estimation. The dataset's size and diversity enable the development of robust and generalizable pose estimation models.

One significant difficulty lies in obtaining a dataset that truly captures the diversity and complexity of human experiences. Ensuring that the dataset encompasses a wide range of demographics, including race, gender, age, and socio-economic backgrounds, requires careful consideration and meticulous effort [18], [19]. Privacy concerns and ethical considerations further complicate the process, as obtaining informed consent and protecting individuals' identities becomes paramount [20], [21].

Annotating and labeling the data accurately is another hurdle, as it often necessitates the involvement of human experts who possess domain knowledge and expertise. This process can be both labor-intensive and time-consuming, particularly when dealing with subjective or nuanced information that requires careful evaluation and categorization [22], [23].

Synthetic datasets offer complete control over data generation, allowing researchers to manipulate various factors such as lighting conditions, camera viewpoints, and background settings. This control enables the creation of highly diverse and challenging training scenarios that can improve the generalization capability of pose estimation models [24], [25].

While synthetic datasets featuring whole human bodies may not yet achieve photorealism, they have become increasingly valuable in supporting tasks such as action recognition, pose estimation, and human tracking [26].

Among synthtetics datasets for human pose estimation, SURREAL dataset [27] was one of the precursors. In this paper authors encompass over 6 million frames and includes accurate posture information, depth maps, and segmentation masks. The groundbreaking research conducted in this study has paved the way for numerous advancements in human analysis, leveraging cost-effective and large-scale synthetic data.

More recent developments highlight the potential of synthetic data as an alternative to real-world data [28], [29]. In PEOPLESANSPEOPLE [28], the authors introduced a human-centric synthetic data generator which offers simulation-ready 3D human assets, parameterized lighting and camera systems, and various labels such as bounding boxes, instance and semantic segmentation, and COCO pose labels. The researchers conducted benchmark experiments using the dataset and a Detectron2 Keypoint R-CNN variant. They discovered that pre-training a network with synthetic data and fine-tuning it with real-world data led to significant improvements in keypoint average precision, surpassing models trained solely with real data or pre-trained on ImageNet.

III. Microgravity Dataset

A. Virtual astronauts

Firstly, ten different humanoid avatars were created using Ready Player Me[1]. These avatars represent five men and five women with diverse skin colors, facial characteristics, and hairstyles. The objective behind this selection was to address one of the challenges identified by [19] and ensure a more inclusive representation across different demographics.

To accurately capture the essence of astronauts aboard the ISS, we dressed our avatars in either a blue or white polo shirt, khaki pants, and white sneakers. These clothing choices closely resemble the uniforms worn by ISS personnel. Figure 1 shows the created avatars.

While there are numerous tools available for recreating animations based on videos or using pre-defined animation tools, we encountered difficulties finding existing movements suitable for microgravity scenarios.

The unique posture that astronauts have in microgravity, coupled with the lack of videos without occlusion and with a fixed camera in high resolution, made it challenging to create or obtain suitable movements based on videos. Therefore, we create basic movements simulating astronauts. Initially, all avatars were imported into Mixamo[2] to standardize their T-pose. Subsequently, we employed the Mixamo add-on for

[1]https://readyplayer.me/, last accessed 19.06.2023
[2]https://www.mixamo.com, last accessed 19.06.2023

Fig. 1: Ten different avatars (five males and five females) with a casual look like astronauts on the ISS.

Fig. 2: ISS interior module scene background with an avatar.

Blender to generate body movements such as raising arms and legs, as well as rotating the torso.

In order to create and execute our scenes, we leveraged the capabilities of Unity. Each avatar in Unity was equipped with a humanoid animator, enabling us to introduce randomized movements.

We divided our animator into four layers: the base layer, legs layer, and right and left arm layers and organized the animations into Blend Trees, where values between 0 and 1 represented the movement magnitude of the body parts. The target magnitude is randomly generated when a collision between the avatar and the environment is detected. The increase and decrease of the value is gradually changed each frame until achieve the assigned value. Furthermore, when the avatar is not in a collision state, we have defined its default pose as having arms closed and legs flexed, similar to the standard neutral body posture in microgravity [30].

To enable collision detection with the environment and predefined targets, we assigned meshes to all avatars. For the avatar's body, we used capsule meshes created with Easy Collider Editor [3] to simplify the simulation process.

When a collision is detected, we analyze which body part collided and the orientation of the avatar's body. Using this information, we trigger one of the animator's layers and apply an impulse force in the direction of the subsequent target. This allows us to simulate realistic reactions to collisions and facilitate the avatar's movement.

B. Scene background, lighting and cameras

The chosen scene background consists of the interior of the ISS laboratory, as illustrated in Figure 2. The background was created using a 3D model available on NASA's website [4].

Multiple target zones were implemented in the scene to enhance the experimental setup. When the avatars reached one of these zones, they received a force impulse for translation or rotation, guiding them towards the next target. These targets were evenly spaced, effectively dividing the environment into seven segments. The primary objective of this segmentation was twofold: to increase the variety of poses performed by the avatars and to allow them sufficient time to return to their neutral pose.

Four cameras were incorporated to capture the avatars' poses. These cameras were equipped with a 15mm focal length and configured with a 65mm ALEXA sensor type. We set their clipping planes to 0.3 for the near plane and 100 for the far plane. The selection of the camera model occured due to its similarity to the internal cameras found on the ISS [31]. Additionally, it required no additional plugins to recreate it within the Unity environment.

C. Data generations

We conducted our simulation with a capture rate of 30fps. Over the course of the experiment, we instantiated varying combinations of avatars, ranging from one to six, every two minutes. Each avatar was assigned distinct targets and orientations to diversify the dataset.

For each avatar, we created 17 annotation points according to the COCO dataset. During the simulations, we collected the projections of each annotation point on each camera, as well as the visibility state of each point and the bounding box of each avatar.

We adhered to the COCO standard for visibility labels, where a value of 0 indicates that the keypoint is not labeled, 1 indicates that it is labeled but not visible, and 2 indicates that it is visible.

We stored the annotations in a JSON format, and for every 30 frames, we saved a JPEG file with a resolution of 1920x1080 pixels.

IV. Transfer Learning

A. Test Set

In order to compare our proposed dataset with real images and videos of the ISS, we generated annotations on the actual footage using the Coco annotator tool [32]. We defined the standard 17 keypoints from the COCO dataset and manually crafted bounding boxes for visible individuals.

[3]https://assetstore.unity.com/packages/tools/level-design/easy-collider-editor-67880, last accessed 01.08.2024

[4]https://nasa3d.arc.nasa.gov/detail/iss-internal, last accessed 19.06.2023

TABLE I: Quantitative results using standard frameworks (expressed in percentages)

Synthetic dataset						
	AP	AP50	AP75	AR	AR50	AR75
Openpose	39.1	57.4	41.8	46.4	66.0	48.2
Openpifpaf	13.6	28.8	11;8	15.0	30.6	13.0
MMPose (HRNET)	**62.6**	**76.8**	**64.1**	**74.0**	**90.2**	**75.5**
MMPose (ViTPose)	36.8	55.7	36.1	52.9	75.4	52.5
Real footage						
	AP	AP50	AP75	AR	AR50	AR75
Openpose	24.8	33.6	24.8	40.7	52.3	41.0
Openpifpaf	32.5	50.6	32.6	51.3	71.9	53.0
MMPose (HRNET)	**50.1**	**65.7**	**52.1**	**55.6**	**69.6**	**57.5**
MMPose (ViTPose)	46.4	60.2	48.5	60.2	74.0	63.0

A total of 77 pictures, each containing at least one person, were annotated. Following the annotation process, we augmented the dataset by mirroring and flipping the images. As a result, our test set comprised a total of 312 images with resolutions ranging from 512×512 to 3000×2000 pixels.

B. Existing Frameworks Evaluation and Selection for Transfer Learning

Initially, we evaluated different 2D pose estimation frameworks to determine the most suitable for our use case. We selected Openpose, Openpifpaf (using Shufflenetv2k16 trained model), and MMPose (utilizing a bottom-up approach with HRNET backbone and associative embedding, and a top-down approach with ViTPose large models). Table I displays our results for mean average precision (mAP) and mean average recall (mAR) on both our synthetic and real sets of images from the ISS.

Our comparison revealed that models employing MMPose exhibited higher values for mAP and mAR when using the bottom-up and top-down approaches, respectively. Given the nature of our use case, which often involves occlusions that obscure most of the astronauts, we decided to adopt the bottom-up approach. Furthermore, training with the bottom-up approach eliminates the need to refine both object detection and keypoint identification independently.

C. Transfer Learning Strategy

For the transfer learning task, we utilized data collected from our dataset. Initially, we randomly selected 90000 images for training purposes, ensuring the exclusion of subsequent frames to generate diverse avatar poses. Further refinement was applied using the Openpose approach, considering an avatar as visible only when at least 5 keypoints were present.

Additionally, we implemented the "iscrowd" flag following the COCO standard. This flag was assigned when the bounding box containing the avatar was smaller than 5% of the image area, ensuring that only images with visible avatars were utilized.

After refining the training images, we were left with 35473 images for transfer learning. All images were resized to 512x512 pixels to maintain the dimensions of the pretrained model.

The standard HRNET model used for the transfer learning task consists of the backbone and 4 extra layers before its head. Given that our transfer learning dataset is entirely composed of virtual images, we experimented with different strategies by freezing varying numbers of extra layers and retaining the weights from the backbone.

Our model underwent training for 1200 epochs, with evaluations conducted every 10 epochs. All parameters from the pretrained configuration file were retained for our transfer learning. The training process was executed on a NVIDIA A100-PCIE-40GB GPU.

Aiming to validate our dataset against other synthetic datasets, we generated a sample of 90,000 images using the PEOPLESANSPEOPLE dataset. These images included human avatars in various orientations, mimicking our approach with our dataset. We applied identical preprocessing steps as those used in our synthetic dataset, resulting in a total of 32,845 images for training.

D. Transfer Learning Results

We began by assessing our results with varying numbers of frozen layers. Upon scrutinizing the training outcomes, it became evident that the number of frozen layers influences the quality of the results. The optimal outcomes for different approaches are consolidated in Table II. The variations in mAP results are approximately 2.5%.

TABLE II: Quantitative results for ablation study (expressed in percentages)

	AP	Epoch
3 Frozen Layers	**60.0**	**1180**
2 Frozen Layers	59.5	590
1 Frozen Layer	59.2	940
0 Frozen Layers	58.5	590

Despite the marginal variation, we chose to employ the optimal approach, which entailed freezing three from four additional layers, for both our quantitative and qualitative analyses.

To conduct a comparative analysis between our dataset and another synthetic dataset, we utilized our generated data from PEOPLESANSPEOPLE with the same configuration parameters that yielded our best result. However, the model's performance degraded slightly, and our best result occurred at epoch 305, which was marginally better than the performance of the pretrained model. The resulting values for mAP and mAR are outlined in Table III.

When comparing the obtained results with standard models, a noticeable divergence is observed between our approach's performance on virtual datasets and real footage. Despite this distinction, our model exhibits an improvement of approximately 20% for both mAP and mAR.

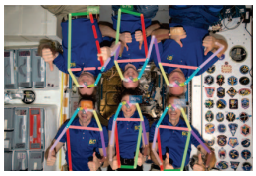
Furthermore, we conducted a qualitative evaluation of our results, as illustrated in Figure 3. This qualitative analysis

TABLE III: Quantitative results after transfer learning (expressed in percentages)

		AP	AP50	AP75	AR	AR50	AR75
Model trained with Our Synthetic Dataset	**Virtual dataset (1920x1080 pixels)**	**82.3**	**87.3**	**83.5**	**91.6**	**97.8**	**92.5**
	Virtual dataset (512x512 pixels)	**86.1**	**96.1**	**89.9**	**90.5**	**99.2**	**93.3**
	Real footage	**60.0**	**79.6**	**64.8**	**66.6**	**84.4**	**70.6**
Model trained with PEOPLESANSPEOPLE Dataset	Virtual dataset (1920x1080 pixels)	24.7	43.8	22.9	74.4	85.2	78.0
	Virtual dataset (512x512 pixels)	15.6	28.8	14.1	65.3	84.3	69.1
	Real footage	52.8	69.6	55.4	59.3	74.4	61.6

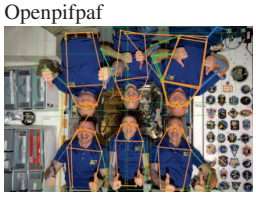
provides insights into the strengths and weaknesses of the trained model.

(a) Pose estimation with OpenPose



(b) Pose estimation with Openpifpaf

(c) Pose estimation with HRNET



(d) Pose estimation with ViTPose

(e) Pose estimation with retrained model

Fig. 3: Comparison pose estimation with different frameworks and retrained model.

While ViTPose and Openpifpaf appear to identify more keypoints, these frameworks prove less efficient in recognizing individuals in microgravity. Upon analyzing the images, it becomes apparent that these frameworks might incorrectly identify more people than are actually present in the scene.

In comparing our retrained model with Openpose and HRNET, our model performed well in identifying accurate keypoints, surpassing the standard HRNET. Additionally, it showed improved capability in understanding individuals in varying orientations compared to Openpose.

V. Conclusion

In conclusion, the utilization of synthetic datasets has emerged as a viable alternative for training pose estimation models in scenarios characterized by a scarcity of available data.

Our dataset played a pivotal role in a transfer learning paradigm, demonstrating a significant enhancement of approximately 20% in both mAP and mAR for the real footage from the ISS. This underlines the efficacy of our synthetic dataset in improving model performance. In comparison to the transfer learning results obtained from the PEOPLESANSPEOPLE synthetic dataset, our dataset achieved a superior improvement of around 13% in both AP and AR when applied to real footage, thereby refuting the hypothesis that simply increasing the amount of data leads to proportional performance gains.

Furthermore, this paper provided a comprehensive comparison among various frameworks and methodologies for pose estimation. It underscored the limitations of standard models in comprehending unconventional poses, as encountered in microgravity scenarios. The critical evaluation of these frameworks sheds light on the challenges that synthetic datasets, like ours, aim to address.

As part of our future endeavors, we envision leveraging our dataset to advance 3D pose estimation. This involves generating data in diverse contexts and utilizing the virtual environment as ground truth, providing a more comprehensive and nuanced understanding of human poses in three-dimensional space. This extension of our work holds promising potential for furthering the capabilities of pose estimation models and enhancing their applicability across a broader spectrum of scenarios.

Acknowledgment

The authors gratefully acknowledge the support of the European Space Agency (ESA) under Contract No. 4000138113/22/NL/GLC/ov. This research was conducted as part of Digital Twins of Humans for Space Operations with XR telepresence, and the authors appreciate the resources and opportunities provided by ESA throughout the duration of this project.

References

[1] T. Y. Kim, "Theoretical study on microgravity and hypogravity simulated by random positioning machine," *Acta Astronautica*, vol. 177, pp. 684–696, 12 2020.

[2] E. De Martino, D. A. Green, D. Ciampi de Andrade, T. Weber, and N. Herssens, "Human movement in simulated hypogravity—Bridging the gap between space research and terrestrial rehabilitation," *Frontiers in Neurology*, vol. 14, 2 2023.

[3] T. Kitamura, H. Teshima, D. Thomas, and H. Kawasaki, "Refining OpenPose with a new sports dataset for robust 2D pose estimation," in *2022 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)*. IEEE, 1 2022, pp. 672–681.

[4] S. Dubey and M. Dixit, "A comprehensive survey on human pose estimation approaches," *Multimedia Systems*, 2 2022.

[5] P. Parekh and A. Patel, "Deep Learning-Based 2D and 3D Human Pose Estimation: A Survey," in *Lecture Notes in Networks and Systems*, vol. 203 LNNS. Springer Science and Business Media Deutschland GmbH, 2021, pp. 541–556.

[6] M. Ben Gamra and M. A. Akhloufi, "A review of deep learning techniques for 2D and 3D human pose estimation," 10 2021.

[7] S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, "Convolutional Pose Machines," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 6 2016, pp. 4724–4732.

[8] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 7 2017, pp. 936–944.

[9] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 6 2017.

[10] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 386–397, 2 2020.

[11] G. Papandreou, T. Zhu, N. Kanazawa, A. Toshev, J. Tompson, C. Bregler, and K. Murphy, "Towards Accurate Multi-person Pose Estimation in the Wild," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 7 2017, pp. 3711–3719.

[12] Y. Chen, Z. Wang, Y. Peng, Z. Zhang, G. Yu, and J. Sun, "Cascaded Pyramid Network for Multi-person Pose Estimation," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, 6 2018, pp. 7103–7112.

[13] E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, and B. Schiele, "DeeperCut: A Deeper, Stronger, and Faster Multi-person Pose Estimation Model," 2016, pp. 34–50.

[14] L. Pishchulin, E. Insafutdinov, S. Tang, B. Andres, M. Andriluka, P. Gehler, and B. Schiele, "DeepCut: Joint Subset Partition and Labeling for Multi Person Pose Estimation," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 6 2016, pp. 4929–4937.

[15] M. Kocabas, S. Karagoz, and E. Akbas, "MultiPoseNet: Fast Multi-Person Pose Estimation Using Pose Residual Network," 2018, pp. 437–453.

[16] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 172–186, 1 2021. [Online]. Available: https://ieeexplore.ieee.org/document/8765346/

[17] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár, "Microsoft COCO: Common Objects in Context," 5 2014.

[18] J. Buolamwini and T. Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," in *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, S. A. Friedler and C. Wilson, Eds. PMLR, 2 2018, pp. 77–91. [Online]. Available: https://proceedings.mlr.press/v81/buolamwini18a.html

[19] J. W. Gichoya, I. Banerjee, A. R. Bhimireddy, J. L. Burns, L. A. Celi, L.-C. Chen, R. Correa, N. Dullerud, M. Ghassemi, S.-C. Huang, P.-C. Kuo, M. P. Lungren, L. J. Palmer, B. J. Price, S. Purkayastha, A. T. Pyrros, L. Oakden-Rayner, C. Okechukwu, L. Seyyed-Kalantari, H. Trivedi, R. Wang, Z. Zaiman, and H. Zhang, "AI recognition of patient race in medical imaging: a modelling study," *The Lancet Digital Health*, vol. 4, no. 6, pp. e406–e414, 6 2022.

[20] B. Lepri, N. Oliver, and A. Pentland, "Ethical machines: The human-centric use of artificial intelligence," *iScience*, vol. 24, no. 3, p. 102249, 3 2021.

[21] X. Yan, Y. Xu, C. Chen, and S. Zhang, "Privacy preserving for AI-based 3D human pose recovery and retargeting," *ISA Transactions*, 4 2023.

[22] J. Wang, S. Tan, X. Zhen, S. Xu, F. Zheng, Z. He, and L. Shao, "Deep 3D human pose estimation: A review," *Computer Vision and Image Understanding*, vol. 210, 9 2021.

[23] L. K. Topham, W. Khan, D. Al-Jumeily, and A. Hussain, "Human Body Pose Estimation for Gait Identification: A Comprehensive Survey of Datasets and Models," *ACM Computing Surveys*, 7 2022.

[24] W. Chen, H. Wang, Y. Li, H. Su, Z. Wang, C. Tu, D. Lischinski, D. Cohen-Or, and B. Chen, "Synthesizing Training Images for Boosting Human 3D Pose Estimation," in *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE, 10 2016, pp. 479–488.

[25] G. Rogez and C. Schmid, "MoCap-guided data augmentation for 3D pose estimation in the wild," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*. Barcelona: Curran Associates Inc., 2016, pp. 3116–3124.

[26] D. Nikolova, I. Vladimirov, and Z. Terneva, "Artificial Humans: an Overview of Photorealistic Synthetic Datasets and Possible Applications," in *2022 57th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST)*. IEEE, 6 2022, pp. 1–4.

[27] G. Varol, J. Romero, X. Martin, N. Mahmood, M. J. Black, I. Laptev, and C. Schmid, "Learning from Synthetic Humans," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 7 2017, pp. 4627–4635.

[28] S. E. Ebadi, Y.-C. Jhang, A. Zook, S. Dhakad, A. Crespi, P. Parisi, S. Borkman, J. Hogins, and S. Ganguly, "PeopleSansPeople: A Synthetic Data Generator for Human-Centric Computer Vision," 12 2021. [Online]. Available: http://arxiv.org/abs/2112.09290

[29] A. Mukhopadhyay, G. Rajshekar Reddy, I. Mukherjee, G. Kumar Gopa, A. Pena-Rios, and P. Biswas, "Generating Synthetic Data for Deep Learning using VR Digital Twin," in *2021 5th International Conference on Cloud and Big Data Computing (ICCBDC)*. New York, NY, USA: ACM, 8 2021, pp. 52–56.

[30] National Aeronautics and Space Administration, "Zero-Gravity Body Posture Influences Acupressure Massage Chair," 2020. [Online]. Available: https://spinoff.nasa.gov/Spinoff2020/cg_5.html

[31] P. Muri, S. Runco, C. Fontanot, and C. Getteau, "The High Definition Earth Viewing (HDEV) payload," in *2017 IEEE Aerospace Conference*. IEEE, 3 2017, pp. 1–7.

[32] J. Brooks, "COCO Annotator," 2019. [Online]. Available: https://github.com/jsbroks/coco-annotator/