Video-Based Gait Analysis for Assessing Alzheimer's Disease and Dementia with Lewy Bodies

Diwei Wang, Chaima Zouaoui, Jinhyeok Jang, Hassen Drira, Hyewon Seo

Video-based gait analysis for assessing Alzheimer's Disease and Dementia with Lewy Bodies

Diwei Wang[1], Chaima Zouaoui[1,2], Jinhyeok Jang[3], Hassen Drira[1], and Hyewon Seo[1]

[1] ICube laboratory, University of Strasbourg, France
{d.wang,hdrira,seo}@unistra.fr
[2] Ecole Polytechnique de Tunisie
chaima.zouaoui@ept.ucar.tn
[3] ETRI, South Korea
jjh6297@etri.kr

Abstract. Dementia with Lewy Bodies (DLB) and Alzheimer's Disease (AD) are two common neurodegenerative diseases among elderly people. Gait analysis plays a significant role in clinical assessments to discriminate these neurological disorders from healthy controls, to grade disease severity, and to further differentiate dementia subtypes. In this paper, we propose a deep-learning based model specifically designed to evaluate gait impairment score for assessing the dementia severity using monocular gait videos. Named MAX-GR, our model estimates the sequence of 3D body skeletons, applies corrections based on spatio-temporal gait features extracted from the input video, and performs classification on the corrected 3D pose sequence to determine the MDS-UPDRS gait scores. Experimental results show that our technique outperforms alternative state-of-the-art methods. The code, demo videos, as well as 3D skeleton dataset is available at https://github.com/lisqzqng/Video-based-gait-analysis-for-dementia.

Keywords: Gait impairment score · Dementia subtypes · Human 3D motion estimation · Geometric deep learning.

1 Introduction and related work

Through many previous studies, it is now well understood that the quantitative gait impairment analysis is an established method for accessing neurodegenerative diseases such as Dementia with Lewy Bodies (DLB) or Alzheimer (AD) and gauging their severity, even in the prodromal phase [20]. In order to facilitate quantitative gait analysis, previous works have often relied on wearable sensors [12,27,6,17] or electronic walkways [19,22]. According to Merory et al.'s study [22], individuals with AD and DLB show comparable spatiotemporal gait characteristics that differ significantly from those of the normal population. Conversely, Mc Ardle et al [19] demonstrate that the two subtype groups exhibit

distinct pathological gait signatures. Another study [18] has shown that the environment where walking takes place has an influence on the characteristics of gait impairment in different types of dementia. However, these studies do not aim to automatically estimate severity scores based on the measurement data.

Numerous efforts have been made to classify or estimate the severity of a patient's condition and even distinguish between different dementia subtypes, by using gait data. Muller [23] employed a decision tree [29] to analyse gait motions of individuals with AD and DLB, and showed that walking speed and the asymmetry in left-to-right step lengths were the two primary factors for distinguishing between dementia subtypes and estimating the disease severity. However, the reliance on wearable sensors equipped with tri-axial accelerometers or electronic walkways in such studies can be cumbersome in terms of wearability, calibration, and may not always be readily accessible.

The progress in deep learning has opened up new possibilities for vision-based severity assessment methods. Albuquerque et al [1] have developed a spatiotemporal deep learning technique by producing a gait representation that combines image features extracted through Convolutional Neural Networks (CNNs) with a temporal encoding based on Long Short-Term Memory (LSTM) networks. Lu et al [16] extract 3D body pose from videos, track them through time, and classify the sequence of 3D poses based on the MDS-UPDRS gait scores by using a temporal convolutional neural network. Similarly, Sabo et al [25] have shown that ST-GCN models operating on 3D joint trajectories outperform alternative models. Motivated by the achievements of previous studies, we adopt a similar approach of extracting 3D pose sequences from gait videos with an aim to enhance pathological gait analysis. However, our work distinguishes itself in that we introduce a new dedicated model for 3D motion estimation from monocular gait videos. Additionally, we employ a geometric deep learning module specially crafted for 3D skeleton-based action recognition. Consequently, our method achieves superior performance compared to numerous state-of-the-art techniques.

2 Method

2.1 Our patient data

The videos of patients undergoing the MDS-UPDRS gait examination at a neurology clinic have been used in our study. The patient walks along a GAITRite (`https://www.gaitrite.com/`) electronic walkway with dimensions of $0.6m \times 8m$, from one end to the other end. Three views have been interchangeably chosen for the RGB camera, without calibration: a side view from the mid-way of the walkway, a front view as the patient walks towards the camera, and a back view as the patient walks away from the camera. In the two latter cases, the distance between the camera and the patient varied from 1 meter to 9-10 meters. The recorded images had a resolution of 480×640 pixels, and the frame rate was set

at 30Hz. A total of 92 sequences have been recorded from 44 subjects, including 41 patients with AD and DLB. In addition, each video has been annotated with the personal data including the height and age of the person, dementia type and the severity, and the gait parameters measured by the GAITRite system, such as walk speed, step lengths of each leg, times of contact, etc. 3D joint positions were not available.

2.2 MAX-GRNet for 3D Pose estimation

The first component of our work is the 3D human motion estimation from the 2D RGB video. Like many others, we base our 3D pose estimator on the SMPL [15] model, thus the estimated poses are represented in the form of a sequence of SMPL pose parameters [15]. Our proposed 3D gait reconstructor, named MAX-GRNet, is illustrated in Fig. 1.

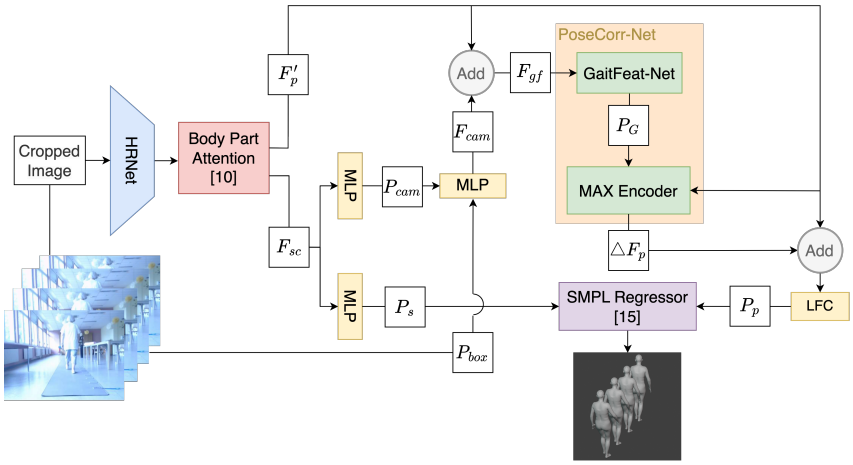

Fig. 1. The architecture of MAX-GRNet, the proposed 3D gait motion reconstructor. $P's$ denote the parameters whereas $F's$ denote the features. LFC stands for locally-connected linear layers. In LFC, pose parameter of each body joint is regressed by an individual linear layer [10].

In our patient video, challenging scenarios can arise due to truncations or reduced visual clarity. To mitigate the impact of truncation and enhance the overall accuracy, we employ the Part Attention mechanism [10], encouraging the model to focus on more credible visual features. As illustrated in Fig. 1, the visual features extracted by the HRNet-W32 [10] at each frame are fed into the Body Part Attention module, to generate joint-specific features. The subject-camera distance, frequently observed to be fairly long in our video (up to more than 10 meters when the patient and the camera is at either ends of the walkway), can also significantly decrease reconstruction accuracy, leading to unrealistic poses

with distorted walking patterns, perturbed foot swings, and illogical step lengths. To tackle this issue, we introduce our Pose Correction Network (PoseCorr-Net), which incorporates several gait characteristics as additional controls and employs the spatio-temporal encoding along with attention mechanism. To address different aspects of the gait motion, we combine average parameters and per-frame parameters as the gait characteristics. The average parameters ensure consistency in the reconstructed walking patterns, while the per-frame parameters accommodate the variations across different gait phases. To regress these parameters, we introduce a GRU-based module named as GaitFeat-Net. The gait parameters P_G are subsequently mapped into a higher-dimensional feature space to enable the fusion with the pose feature F'_P. To further enhance the integration of estimated P_G with F'_P, we introduce a Transformer-based encoder [28] which we refer to as MAX Encoder, to capture the intra-dependencies across different time steps and among the body joints, respectively, before their final merging.

GaitFeat-Net: We feed the pose feature F'_p and camera feature F_{cam} into an one-layer GRU to estimate a number of gait parameters. The effectiveness of gait parameter estimation using a GRU-based network from per-joint 3D position has been previously demonstrated in QuaterNet [24], in the context of generating plausible locomotion given previous poses and locomotive parameters as controls. Differently from Quaternet, our approach uses per-joint feature as the input, instead of per-joint 3D positions. This allows us to capture more detailed information and potentially improve the accuracy of gait parameter estimation. Additionally, we utilize F_{cam} obtained from camera parameters P_{cam} ($[s,t], t \in \mathbb{R}^2$ defined by a weak-perspective camera model) as additional input. As depicted in Fig. 1, we construct a patch of 224×224 pixels from each of the initial video frames based on the bounding box, before feeding the patch sequence into the reconstructor. By incorporating the parameters of the bounding box P_{box}, F_{cam} can effectively capture the position of the patient within the original video frame, and the dynamic information regarding the motion in the video. Based on the available locomotive parameters in our patient data (Sec. 2.1), we have selected gait parameters $P_G = [V, D, \Phi]$, where $V = \overline{\|v\|}$ is the speed amplitude averaged over the sequence, $D = [\overline{l_{left}}, \overline{l_{right}}]$ is the average left/right step lengths, and $\Phi = [cos(\phi_{left}), sin(\phi_{left}), cos(\phi_{right}), sin(\phi_{right})]$ encodes the phase of the left/right gait cycle.

MAXEncoder: By developing this encoder based on a multi-head self-attention mechanism (MSA) [28], we aim to overcome the limitations of recurrent models that struggle to capture long-range relationships in the sequence. Fig. 2 depicts its detailed architecture. We devise an attention-based block that builds upon the MSA variants MSA-T and MSA-S of Spatial-Temporal Encoder (STE) [30]. To incorporate the gait parameters P_G obtained from GaitFeat-Net, and accommodate the joint-wise nature of the pose feature F'_P, we construct different inputs $\tilde{F}^T_{gf}$ and $\tilde{F}^S_{gf}$ for our temporal (TAE) and spatial attention (SAE) blocks. Unlike MSA-S, which focuses on modeling the intra-dependencies within each feature map, the proposed SAE explicitly captures the dependencies among each body joint while leveraging the corresponding gait feature. To merge the

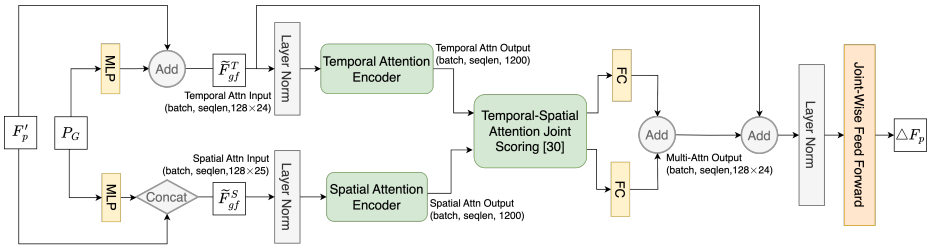

Fig. 2. Architecture of the MAX Encoder.

TAE and SAE blocks, we adopt the parallel connection method as described in [30]. Additionally, we replace the Feed-Foward Network (FFN) in the standard Transformer and the Vision Transforer (ViT) [3] with a Joint-Wise Feed Forward network (JWFF). JWFF employs separate linear layers for each joint, contrary to the global FC layers in FFN. Ablation studies in Sec. 3.3 clearly demonstrate the efficacy of such JWFF approach. The total loss of our proposed MAX-GRNet is:

$$L = w_1 \cdot L_{2D} + w_2 \cdot L_{3D} + w_3 \cdot L_{SMPL} + w_4 \cdot L_{GaitFeat}, \tag{1}$$

where w_i is the weight assigned to each loss, more details on these weight values can be found in Sec. 3.2. L_{SMPL}, L_{3D} and L_{2D} denote respectively the loss term associated with the Euclidean distances calculated on SMPL parameters, 3D joint positions and projected 2D joint positions. $L_{GaitFeat}$ is designed to supervise the gait parameters regressed by the GaitFeat-Net.

2.3 Geometric deep learning for severity assessment

In the second part of our study, we employ the KShapeNet [5], a geometric deep learning model on Kendall's shape space specially developed for skeleton-based human action recognition. It has demonstrated favorable performances on the two large scale skeleton datasets NTU-RGB+D [26] and NTU-RGB+D120 [13] datasets. Initially, skeleton sequences are modeled as trajectories on Kendall's shape space by filtering out the scale and rigid transformations. Next, the sequences are mapped to a linear tangent space and the resulting structured data are fed into a deep learning model, KShapeNet. Notably, it includes a unique layer that learns the optimal rigid and nonrigid transformation to be applied to the 3D skeletons, thereby enhancing the precision of action recognition. This layer is followed by a Conv Block and an LSTM layer that captures the temporal dynamics of the sequences. A subsequent fully connected block generates the corresponding action class as the output. In this work we utilized only the optimization over the rigid transformation layer.

3 Experiments

3.1 Datasets

We utilized Human3.6M motion capture data [2,7] consisting of synchronized 3D joint positions, 2D joint positions, and RGB video frames recorded at 50Hz. We selected 6 subjects to create a set of 79 sub-sequences of walking motions to train and validate the reconstructors. The GPJATK gait dataset [11] was used for the evaluation. It contains Vicon mocap data with video recordings of 4 calibrated RGB cameras. We estimated 3D poses from their videos, which have been time-aligned with the 3D poses computed from their marker data, allowing us to obtain consistent 2D-3D (video and the 3D pose sequence) pairs. To train and test the classifiers, we combine our patient data with a subset of the Toronto Older Adults Gait Archive [21], where we randomly selected 22 walking sequences from 5 subjects.

3.2 Implementation details

3D Pose Estimator: We begin by training the GaitFeat-Net module separately on the gait sequences from Human3.6M, to enhance the stability of the overall MAX-GRNet training. Only $L_{GaitFeat}$ in Eq. 1 has been used in this pre-training stage, with gait parameters computed by analyzing the 3D joint positions. We apply L1 loss for V, D and Φ, with distinct weight assignments: w_V=250, w_D=50, w_Φ=100. During the integral training of MAX-GRNet, we use distinct weights for the losses in Eq. 1, specifically w_{2D}=100, w_{3D}=100, w_θ=60, w_β=0.01, where θ and β denote the pose and shape parameters in SMPL. $L_{GaitFeat}$ is applied only for the initial 20 epochs with weights different from the previous training: w_V=100, w_D=20, w_Φ=50, and the subsequent training continues without any supervision on the GaitFeat-Net. The metadata from the electronic walkway only provides the average locomotive parameters in the patient data such as the duration of the left/right gait cycle. Thus, we apply a fast Fourier transform to obtain the spectrogram from the reconstructed gait, to effectively supervise the estimated Φ. Given the average frequency $\bar{f}$, the phase loss on the patient data is formulated as $L_\Phi = w_{\Phi'} \cdot (\frac{1}{\beta} \cdot A_{\bar{f}} + \beta \cdot \sum_{i=0}^{n} A_{f_i} |f_i - \bar{f}|)$, where $w_{\Phi'}$=0.05 and β=0.2 are weighting coefficients, and A_f denotes the amplitude of f in the spectrogram. The estimation of SMPL parameters on the Human3.6M dataset is performed by using MoSh [14] on the available 3D marker data, and each pose parameter is subsequently represented as a 6D vector [31]. The models were trained on 1 Nvidia RTX 3090 GPU using a batch size of 4 and an Adam optimizer with a learning rate of 5×10^{-5}. Optimal hyperparameters were chosen through grid search.

Motor Severity Assessment: As outlined in Sec. 3.1, we train the classifiers using the estimated skeleton sequences from both healthy and diseased older adults. To obtain the required number of frames for analysis, which has been set to 100 frames, we utilized a sliding window approach. It involved traversing

the original video, extracting subsequences consisting of 100 frames each, and maintaining an overlap of 50 frames between adjacent subsequences. The last frames fewer than 50 frames were not used. We refrained from time-warping the sequences in order to preserve the velocity of the gait motion.

Each subsequence obtained by this way has been assigned the same label as the original video. This sliding window protocol was applied after dividing the dataset into training and testing sets to prevent the occurrence of subsequences from the same video being present in both sets. Note that this increases the total number of sequences used for training and testing, thereby augmenting the limited amount of our patient data.

3.3 Results

Evaluation criteria: The reconstructor model has been validated and evaluated by measuring the 3D per-joint position error respectively on Human3.6M and GPJATK datasets. The classification accuracy has been used to evaluate the overall performance and to compare with other SOTA methods. Different variations of severity assessment tasks has been tested: normal/patient, 3-class diagnosis with normal/Alzheimer/DLB, and 3-class gait scoring (normal-0, moderate-1, and severe-2). All evaluations have used a 10-fold cross-validation scheme. Due to the limited size of the data, we opted to perform a train-test split.

Table 1. Comparison of performance with different model configurations, measured on mean per-joint position error (MPJPE) (*mm*), and classification accuracy (%). Numbers in boldface indicate the top-1 performance, with the top-2 denoted as underlines.

	MPJPE (valid)	MPJPE (test)	Normal /Patient	Normal /AD/DLB	Gait Score (0,1,2)
VIBE [9]	70.40	131.52	<u>94.60</u>	<u>71.37</u>	63.69
Baseline	70.61	103.78	90.11	68.94	62.57
+ MAX Encoder /wo JWFF	70.35	<u>101.83</u>	88.05	59.28	60.20
+ MAX Encoder /with JWFF	68.90	106.40	86.76	58.03	<u>65.24</u>
+ Avg.+ MAX Encoder /wo JWFF	71.13	102.43	85.54	57.74	58.43
+ Avg.+ MAX Encoder /with JWFF	70.38	102.27	87.36	61.58	59.33
+ PoseCorr-Net /wo JWFF	**66.47**	**101.78**	87.38	66.03	60.37
+ PoseCorr-Net /with JWFF (Ours)	<u>67.17</u>	103.77	**96.22**	**75.39**	**65.41**

Ablation study: We assessed six model configurations es well as VIBE model [9] based on the mean per-joint position error (MPJPE) and the classification accuracy. The design of MAX Encoder has been tested with and without joint-wise forward feedback (JWFF). To validate the design of the GaitFeat-Net, we performed three distinct configurations. First, we conducted tests without it, followed by tests using its estimation of only the average parameters V and D (referred to as Avg.), and finally, with the complete estimation P_G. The results

are shown in Table 1. In general, utilizing gait parameters by GaitFeat-Net and subsequently regularizing the 3D pose estimation improves the reconstruction on GPJATK dataset, which predominantly contains regular walking of young and normal people, but reduces the overall classification accuracy. This somewhat aligns with the observations by [19], who pointed out that gait asymmetry and variability are significant factors for differentiating between disease subtypes and the patient group from the normal one. JWFF tends to improve the classification accuracy both in gait scores and dementia subtypes, especially when gait parameters are used with it, indicating its efficacy of extracting informative features from the estimated locomotive parameters.

Table 2. Comparison of classification accuracy with state-of-the-art methods (% accuracy). Numbers in boldface indicate the top-1 accuracy, while the top-2 is denoted with underlines.

	Normal /Patient	Normal /AD/DLB	Gait Score (0,1,2)
VIBE[9] + OF-DDNet[16]	89.60	<u>72.78</u>	60.43
VIBE[9] + KShapeNet	<u>94.60</u>	71.37	63.69
MAX-GRNet + OF-DDNet[16]	89.60	69.92	64.68
MAX-GRNet + ST-GCN [25]	94.24	66.67	**72.31**
MAX-GRNet + FSA-CNN[8]	93.02	72.19	<u>66.74</u>
MAX-GRNet + PoseC3D[4]	92.95	66.59	62.27
Ours (MAX-GRNet + KShapeNet)	**96.22**	**75.39**	65.41

Comparison with the state-of-the-art: We compare our method with two closely related studies, which focus on vision-based gait analysis of parkinsonism severity in dementia [16,25]. Additionally, we include two state-of-the-art models proposed in the vision-based action recognition community [4,8] for further comparison. To evaluate the classifier based on a ST-GCN [25], we evaluate their classifier using solely skeletons as input, excluding the spatio-temporal gait features originally utilized in their work. This is due to the inability to compute these features from the skeleton data reconstructed with MAX-GRNet, as it does not provide the root. Results shown in Table 2 demonstrate that our method achieves favorable performance compared to others, and remains competitive with the CNN-based action recognition approach, which demonstrates superior performance in differentiating dimentia subtypes.

Limitations: Our reconstructor struggles to estimate 3D poses in videos exhibiting severe cases where the patient's gait pattern is highly irregular. The classifier incurs an additional time cost and requires separate processing for the projection of skeleton data onto the tangent space, which hinders its seamless integration with the reconstructor in an end-to-end manner.

4 Conclusion

We have presented a new model aimed at evaluating gait impairment score for assessing the dementia severity using monocular gait videos. Our model features a gait motion reconstructor, which is specifically designed for 3D motion estimation from gait videos based on a gait parameter estimator and a multi-head attention Transformer. Additionally, we employ a geometric deep neural network tailored for the specific task of 3D skeleton-based classification. Our method improves the performance over state-of-the-art techniques in both 3D pose estimation and classification, thus demonstrating significant advancements in the field. In the future, we plan to improve the precision of both 3D pose estimation and classification by effectively leveraging image evidences and gait parameters, respectively.

Prospect of application. In addition to its clear applications in clinical environments, where the computed gait scores from gait videos can be presented to clinicians, our research could also be used in the realm of care robots assisting elderly individuals in residential settings. Specifically, they could identify early signs or variations in the pathological condition through video analysis.

Acknowledgements We would like to thank Dr. Candice Muller and Prof. Dr. Frédéric Blanc at the Robsertsau hospital for sharing patient data and theier valuable expertise. This work has been partially supported by the French national project ArtIC (Artificial Intelligence for Care, ANR-20-THIA-0006) and the binational project "Synthetic Data Generation and Sim-to-Real Adaptive Learning for Real-World Human Daily Activity Recognition of Human-Care Robots (21YS2900)" granted by the ETRI, South Korea. Chaima Zouaoui was supported by the ICube laboratory API project ShaGAI.

References

1. Albuquerque, P., Verlekar, T.T., Correia, P.L., Soares, L.D.: A spatiotemporal deep learning approach for automatic pathological gait classification. Sensors **21**(18), 6202 (2021) 1
2. Catalin, I., Fuxin, L., Cristian, S.: Latent structured models for human pose estimation. In: International Conference on Computer Vision (2011) 3.1
3. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020) 2.2
4. Duan, H., Zhao, Y., Chen, K., Lin, D., Dai, B.: Revisiting skeleton-based action recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2969–2978 (2022) 2, 3.3
5. Friji, R., Drira, H., Chaieb, F., Kchok, H., Kurtek, S.: Geometric deep neural network using rigid and non-rigid transformations for human action recognition. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12611–12620 (2021) 2.3

6. Hsu, W.C., Sugiarto, T., Lin, Y.J., Yang, F.C., Lin, Z.Y., Sun, C.T., Hsu, C.L., Chou, K.N.: Multiple-wearable-sensor-based gait classification and analysis in patients with neurological disorders. Sensors **18**(10), 3397 (2018) 1
7. Ionescu, C., Papava, D., Olaru, V., Sminchisescu, C.: Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. IEEE Transactions on Pattern Analysis and Machine Intelligence (2014) 3.1
8. Jang, J., Kim, D., Park, C., Jang, M., Lee, J., Kim, J.: Etri-activity3d: A large-scale rgb-d dataset for robots to recognize daily activities of the elderly. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 10990–10997. IEEE (2020) 2, 3.3
9. Kocabas, M., Athanasiou, N., Black, M.J.: Vibe: Video inference for human body pose and shape estimation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5253–5263 (2020) 1, 3.3, 2
10. Kocabas, M., Huang, C.H.P., Hilliges, O., Black, M.J.: Pare: Part attention regressor for 3d human body estimation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 11127–11137 (2021) 1, 2.2
11. Kwolek, B., Michalczuk, A., Krzeszowski, T., Switonski, A., Josinski, H., Wojciechowski, K.: Calibrated and synchronized multi-view video and motion capture dataset for evaluation of gait recognition. Multimedia Tools and Applications **78**, 32437–32465 (2019) 3.1
12. Li, H., Mehul, A., Le Kernec, J., Gurbuz, S.Z., Fioranelli, F.: Sequential human gait classification with distributed radar sensor fusion. IEEE Sensors Journal **21**(6), 7590–7603 (2020) 1
13. Liu, J., Shahroudy, A., Perez, M., Wang, G., Duan, L.Y., Kot, A.C.: Ntu rgb+ d 120: A large-scale benchmark for 3d human activity understanding. IEEE transactions on pattern analysis and machine intelligence **42**(10), 2684–2701 (2019) 2.3
14. Loper, M., Mahmood, N., Black, M.J.: Mosh: motion and shape capture from sparse markers. ACM Trans. Graph. **33**(6), 220–1 (2014) 3.2
15. Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., Black, M.J.: Smpl: A skinned multi-person linear model. ACM transactions on graphics (TOG) **34**(6), 1–16 (2015) 2.2
16. Lu, M., Poston, K., Pfefferbaum, A., Sullivan, E.V., Fei-Fei, L., Pohl, K.M., Niebles, J.C., Adeli, E.: Vision-based estimation of mds-updrs gait scores for assessing parkinson's disease motor severity. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 637–647. Springer (2020) 1, 2, 3.3
17. Mannini, A., Trojaniello, D., Cereatti, A., Sabatini, A.M.: A machine learning framework for gait classification using inertial sensors: Application to elderly, post-stroke and huntington's disease patients. Sensors **16**(1), 134 (2016) 1
18. Mc Ardle, R., Del Din, S., Donaghy, P., Galna, B., Thomas, A.J., Rochester, L.: The impact of environment on gait assessment: considerations from real-world gait analysis in dementia subtypes. Sensors **21**(3), 813 (2021) 1
19. Mc Ardle, R., Galna, B., Donaghy, P., Thomas, A., Rochester, L.: Do alzheimer's and lewy body disease have discrete pathological signatures of gait? Alzheimer's & Dementia **15**(10), 1367–1377 (2019) 1, 3.3
20. McKeith, I.G., Ferman, T.J., Thomas, A.J., Blanc, F., Boeve, B.F., Fujishiro, H., Kantarci, K., Muscio, C., O'Brien, J.T., Postuma, R.B., et al.: Research criteria for the diagnosis of prodromal dementia with lewy bodies. Neurology **94**(17), 743–755 (2020) 1

21. Mehdizadeh, S., Nabavi, H., Sabo, A., Arora, T., Iaboni, A., Taati, B.: The toronto older adults gait archive: video and 3d inertial motion capture data of older adults' walking. Scientific data **9**(1), 398 (2022) 3.1
22. Merory, J., Wittwer, J., Rowe, C., Webster, K.: Quantitative gait analysis in patients with dementia with lewy bodies and alzheimer's disease. Gait posture **26**, 414–9 (10 2007). https://doi.org/10.1016/j.gaitpost.2006.10.006 1
23. Muller, C., Perisse, J., Blanc, F., Kiesmann, M., Astier, C., Vogel, T.: Corrélation des troubles de la marche au profil neuropsychologique chez les patients atteints de maladie d'alzheimer et maladie à corps de lewy. Revue Neurologique **174**, S2–S3 (2018) 1
24. Pavllo, D., Grangier, D., Auli, M.: Quaternet: A quaternion-based recurrent model for human motion. In: Proceedings of the British Machine Vision Conference (BMVC) (2018) 2.2
25. Sabo, A., Mehdizadeh, S., Iaboni, A., Taati, B.: Estimating parkinsonism severity in natural gait videos of older adults with dementia. IEEE journal of biomedical and health informatics **26**(5), 2288–2298 (2022) 1, 2, 3.3
26. Shahroudy, A., Liu, J., Ng, T.T., Wang, G.: Ntu rgb+ d: A large scale dataset for 3d human activity analysis. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1010–1019 (2016) 2.3
27. Teufl, W., Taetz, B., Miezal, M., Lorenz, M., Pietschmann, J., Jöllenbeck, T., Fröhlich, M., Bleser, G.: Towards an inertial sensor-based wearable feedback system for patients after total hip arthroplasty: Validity and applicability for gait classification with gait kinematics-based features. Sensors **19**(22), 5006 (2019) 1
28. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017) 2.2, 2.2
29. Von Winterfeldt, D., Edwards, W.: Decision Analysis and Behavioral Research. Cambridge University Press (1986) 1
30. Wan, Z., Li, Z., Tian, M., Liu, J., Yi, S., Li, H.: Encoder-decoder with multi-level attention for 3d human shape and pose estimation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 13033–13042 (2021) 2.2
31. Zhou, Y., Barnes, C., Lu, J., Yang, J., Li, H.: On the continuity of rotation representations in neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5745–5753 (2019) 3.2