Supplementary Material

Bingzhi Duan, Xiaoyue Wan, and Xu Zhao*

Shanghai Jiao Tong University, Shanghai, China
{duanbingzhi,sherrywaan,zhaoxu}@sjtu.edu.cn

Table 1: Reconstruction Network Parameters. The even-numbered layers are active (RELU) layers. Cha, Ker Str, Pad means Channels, Kernel Size, Stride and Padding respectively. ConvT2d denotes the ConvTranspose2d layer.

	Encoder					Decoder				
Layer	Type	Cha	Ker	Str	Pad	Type	Cha	Ker	Str	Pad
1	Conv2d	8	4	2	1	ConvT2d	1	3	1	1
3	Conv2d	4	4	2	1	Conv2d	4	3	1	1
5	Conv2d	4	3	1	1	Conv2d	4	1	1	0
7	Conv2d	4	3	1	1	Conv2d	4	3	1	1
9	Conv2d	4	1	1	0	Conv2d	4	1	1	0
11	Conv2d	4	3	1	1	ConvT2d	8	4	2	1
13	Conv2d	4	1	1	0	ConvT2d	1	4	2	(1,1)
15	Conv2d	1	1	1	0					

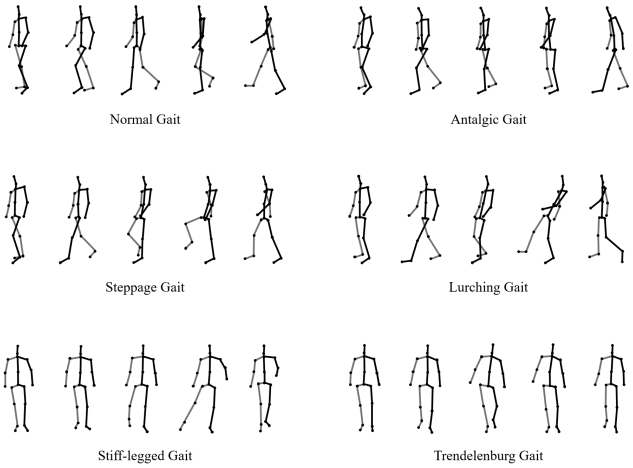
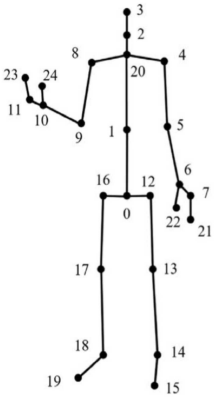


Fig. 1: Pathological gait dataset visualization [2].



0: Spine base
1: Mid-spine
2: Neck
3: Head
4: Left shoulder
5: Left elbow
6: Left wrist
7: Left hand
8: Right shoulder
9: Right elbow
10: Right wrist
11: Right hand
12: Left hip
13: Left knee
14: Left ankle
15: Left foot
16: Right hip
17: Right knee
18: Right ankle
19: Right foot
20: Spine shoulder
21: Left hand tip
22: Left thumb
23: Right hand tip
24: Right thumb

Fig. 2: The original pose data [2]. [0,3,4,5,7,8,9,11,12,13, 14,15,16,17,18,19,20] are chosen to form 17 joints pose.

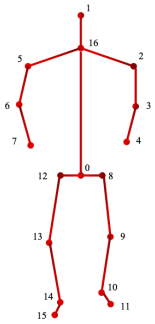


Fig. 3: Human pose structure encoded by 17 joints.

Comparative experiment. To compare with more baseline and prove the generalization capability of FSGait, we select part of AGD-CMU and conduct comparative experiments, as shown in Tab. 2. More information about selected part of AGD-CMU is available at https://github.com/BingzhiDuan/FSGait. The model structure we adopt refers to the supervised learning model VIT [1]. The Encoder structure of Transformer is utilized for gait sequence encoding, followed by an MLP classification head for frame-level binary classification of normal and abnormal, and it is validated on AGD-CMU dataset.

Experiment settings are as followed. Only the normal data in the training set are used in baseline and FSGait training, while Transformer is trained on normal data and additional abnormal data for 100 epochs. We control the number of gait abnormality types in the training set. As shown in Tab. 2, AN represents the number of abnormal gait types in training set. Test set contains 23 abnormal types.

The comparative indicators are shown in Tab. 2. AUC, ACC, and Specificity of the supervised Transformer are slightly higher than ours when the types of anomalies in the training and test sets are the same (AN = 23). However, FSGait significantly outperforms in Precision and Specificity, which represents a stronger ability to distinguish normal samples. When the training set contains fewer anomaly types than the test set (AN = 13), FSGait achieves the best performance across several metrics, including AUC and ACC. In this way, we further demonstrate the effectiveness of FSGait.

To prove the generalization capability, we do comparative experiments by reducing AN from 23 to 13, leaving 10 untrained gait abnormalities in the test set. ACC and AUC of Transformer with supervised learning decrease significantly, while indexes of FSGait remained basically unchanged or even increased. FSGait only needs normal gait for training and has strong generalization capability.

Table 2: Effect Comparisons on AGD-CMU.

Method	Learning Type	AN	AUC	Accuracy	Precision	Sensitivity	Specificity
baseline	UnSupervised	23	0.655	0.706	0.761	0.884	0.184
Transformer	Supervised	23	0.854	0.793	0.843	0.887	0.519
FSGait(ours)	UnSupervised	23	0.852	0.775	0.915	0.771	0.790
baseline	UnSupervised	13	0.615 ↓	0.699 ↓	0.778 ↑	0.856 ↓	0.161 ↓
Transformer	Supervised	13	0.813 ↓	0.758 ↓	0.882 ↑	0.794 ↓	0.638 ↑
FSGait(ours)	UnSupervised	13	**0.854** ↑	**0.778** ↑	**0.927** ↑	**0.774** ↑	**0.789** ↓

References

1. Dosovitskiy, A., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations (2021) 2
2. Jun, K., Lee, Y., Lee, S., Lee, D.W., Kim, M.S.: Pathological gait classification using kinect v2 and gated recurrent neural networks. IEEE Access **8**, 139881–139891 (2020) 1, 2